Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 1004–1009.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1159–1165.

SOLIDS
Structure

The Relaxation Rate of the Magnetic Moment of a Shallow Acceptor Center as a Function of Impurity Concentration in Silicon

T. N. Mamedov[a, *], D. G. Andrianov[b], D. Herlach[c], V. N. Gorelkin[d], K. I. Gritsaĭ[a], O. Kormann[e], J. Major[e, f], A. V. Stoĭkov[a], and U. Zimmermann[c]

[a]*Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia*
[b]*Giredmet State Institute of Research and Design in the Rare-Metals Industry, Moscow, 109017 Russia*
[c]*Paul Scherrer Institut, Villigen PSI, CH-5232, Switzerland*
[d]*Moscow Institute of Physics and Technology, Dolgoprudnyi, Moscow oblast, 141700 Russia*
[e]*Max-Planck-Institut für Metallforschung, Stuttgart, D-70569 Deutschland*
[f]*Universität Stuttgart, Institut für Theoretische und Angewandte Physik, Stuttgart, D-70569 Deutschland*
**e-mail: tmamedov@nu.jinr.ru*

Received August 30, 2000

Abstract—A study is made into the temperature dependence of residual polarization of negative muons in crystalline silicon with the concentration of impurity of the *n*- and *p*-types ranging from 8.7×10^{13} to 4.1×10^{18} cm^{-3}. The measurements are performed in a magnetic field of 1000 G transverse to the muon spin, in the temperature range from 4.2 to 300 K. The form of the temperature dependence of the relaxation rate ν of the magnetic moment of the $_{\mu}Al^0$ acceptor in silicon is determined. For a nondegenerate semiconductor, the relaxation rate depends on temperature as $\nu \propto T^q$ ($q \approx 3$). A variation in the behavior of the temperature dependence and a multiple increase in the relaxation rate are observed in the range of impurity concentration in excess of 10^{18} cm^{-3}. The importance of phonon scattering and spin-exchange scattering of free charge carriers by an acceptor from the standpoint of relaxation of the acceptor magnetic moment is discussed. The constant of hyperfine interaction in an acceptor center formed by an atom of aluminum in silicon is estimated for the first time: $|A_{hf}(\mathrm{Al})/2\pi| \sim 2.5 \times 10^6$ s^{-1}. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Interest in the investigations of impurity centers and various defects in semiconductors never wanes, because these investigations produce detailed data about the distribution of the wave function of these centers in the crystal lattice, about the structure of the impurity conduction band, and so on (see, for example, [1]). In spite of the fact that a wealth of experimental data about various impurities and defects in semiconductors have been accumulated over a period of several decades, shallow acceptor centers in semiconductors with a crystal diamond structure (silicon, germanium, etc.) are still little studied both experimentally and theoretically [1–4].

The main difficulties of theoretical investigations of acceptor impurities in diamond-like semiconductors are caused by the structure of the valence band of these semiconductors. The upper edge of the valence band in silicon and germanium is found at $\mathbf{k} = 0$, and the band is fourfold degenerate at this point.

In the case of shallow acceptor centers in diamond-like semiconductors, the use of the electron paramagnetic resonance (EPR) technique (which played the key part in the investigation of various paramagnetic impurities in semiconductors) is limited because of the high relaxation rate of the magnetic moment of the acceptor and the broadening of the EPR line due to degeneracy of the ground state of acceptor impurity and occasional residual deformations in crystal [5].

The use of negative muons, which was justified theoretically in [6] and developed in the experimental studies [7–10], enables one to extend the possibilities of studying shallow acceptor centers in silicon.

The muonic atom formed as a result of implantation of a negative muon into silicon simulates the behavior of an aluminum atom, which is the acceptor impurity. The behavior of residual polarization of a negative muon in the 1*S*-state of the muonic atom is governed by the state of the electron shell of this atom and by the interaction of the atom and the medium. The results of theoretical treatment of the kinetics of the formation of a muonic atom in silicon [11], which were supported by experimental results, indicate that, in the temperature range below approximately 50 K in the period of time $\lesssim 10^{-9}$ s, an acceptor center (muonic atom) forms in the neutral (paramagnetic) state, $_{\mu}Al^0$. In an external magnetic field

transverse to the muon spin, the relaxation (relaxation rate, λ) and shift of the precession frequency ($\Delta\omega$) of muon spin occur, which are caused by the muon interaction with the magnetic moment of the electron shell of atom. The correlation between the muon polarization parameters and the quantities that characterize the magnetic interactions of the acceptor center (A_{hf} is the constant of hyperfine interaction between the muon magnetic moments and the electron shell in muonic atom, and ν is the relaxation rate of the magnetic moment of electron shell) at $\nu \gg A_{hf}$ is defined by the relations [12, 13]

$$\frac{\Delta\omega}{\omega_0} = -\frac{g\mu_B}{2\mu_B^\mu}\frac{j(j+1)\hbar A_{hf}}{3k_BT} + \frac{A_{hf}^2}{2(\nu^2+\omega_e^2)}, \qquad (1)$$

$$\lambda = \frac{j(j+1)}{3}\left(\frac{A_{hf}^2}{\nu} + \frac{A_{hf}^2}{2(\nu^2+\omega_e^2)}\right), \qquad (2)$$

where $\Delta\omega = \omega(T) - \omega_0$ (ω_0 is the angular precession frequency of muon spin in the diamagnetic state of muonic atom); k_B is the Boltzmann constant; μ_B and μ_B^μ is the Bohr magneton for an electron and muon, respectively; g is the g-factor of the acceptor center; $\omega_e = g\mu_B B/\hbar$ is the angular precession frequency of the magnetic moment of the electron shell of the acceptor center in the external magnetic field B; and T is the temperature in Kelvins. For a shallow acceptor center in silicon, $j = 3/2$ [14].

The range of values of ν accessible for experimental study is defined, according to Eq. (2), by the range of values of λ being measured and by the value of the hyperfine interaction constant. We assume that $0.05 \lesssim \lambda \lesssim 20\ \mu s^{-1}$ and use the estimate of $A_{hf}/2\pi \approx 30\ \mu s^{-1}$ [10] to derive $5\times10^9 \lesssim \nu \lesssim 2\times10^{12}\ s^{-1}$. Therefore, the μ^-SR method enables one to considerably increase the upper limit of the relaxation rate (compared with the capabilities of the EPR method: $\nu \lesssim 10^9\ s^{-1}$ [5]) and study the processes of magnetic relaxation of shallow acceptors in undeformed silicon samples in a wide range of temperature and impurity concentrations.

2. MEASUREMENTS

Given in this paper are the results of investigating the residual polarization of negative muons in crystalline silicon with boron (8.7×10^{13}, 5.5×10^{16}, 1.4×10^{18}, $4.1\times10^{18}\ cm^{-3}$), gallium ($1.1\times10^{15}$, $1.1\times10^{18}\ cm^{-3}$), and arsenic ($8.0\times10^{15}$, $2.0\times10^{17}\ cm^{-3}$) impurities in the temperature range from 4 to 300 K.

The measurements were performed in the Stuttgart LFQ Spectrometer facility [15] in the μE4 muon channel of the proton accelerator at the Paul Scherrer Institute (PSI, Switzerland). The samples to be investigated were cut in the form of cylinders (30 mm in diameter, 15 mm high) from single crystals of silicon and arranged so that the cylinder axis coincided with the muonic beam axis. A transverse magnetic field of 1000 G was induced on the sample by Helmholtz coils. The long-term stability of the magnetic field was at least 10^{-4}. The samples were placed in a cryostat and cooled by a flow of liquid helium vapors, with the temperature stabilized within 0.1 K in the range from 4.2 to 300 K. The electrons, which were caused by decay of muons that stopped in the target, were registered by scintillation counters. The intervals of time between the moment of stopping of muon and the escape of decay electron were stored in the form of histograms (μSR spectra). The time channel width of the spectrometer was 0.625 ns, with a total of 16 000 channels in the spectrum.

3. RESULTS AND DISCUSSION

In addition to the target, some of the muons stop in the cryostat material and in the scintillation counters and make a contribution to the spectrum being observed. Because the lifetime of negative muons in the 1S-state depends on the charge of atomic nucleus, the spectrogram of the time distribution of decay electrons $\mu^- \longrightarrow e^-$ may be represented in the form

$$f(t) = \sum_X N_X e^{-t/\tau_X}[1 + \alpha p_X(t)] + B(t), \qquad (3)$$

where N_X is the number of events in the zero channel of the histogram due to the decay of muons captured by atoms of the element X; τ_X and p_X denote the muon lifetime and the projection of polarization vector of a muon in the 1S-state of the element X onto the direction of observation, respectively; α is the experimentally obtained coefficient of asymmetry of the spatial distribution of decay electrons with due regard for the solid angle of acceptance of electrons; and B is the background of random coincidences. The explicit form of the function $p_X(t)$ depends on the experimental conditions. Here, we treat the results of measurements in an external magnetic field transverse to the muon spin.

The scintillation counters consist of hydrocarbons, and the cryostat is made of brass. Accordingly, components are always present in the spectra that correspond to the capture of muons by carbon and copper.

The experimental data were treated by the least squares method. The lifetimes of negative muons in the 1S-state of atoms (τ_X) were fixed by the average value of the experimental data of [16] (2030, 760, and 163 ns for carbon, silicon, and copper, respectively).

The results of independent measurements in graphite and copper have revealed that (a) the contributions by the counters are described by the polarization function in the form

$$p_C(t) = p_0\cos(\omega t + \phi)$$

with a value of $\alpha p_0 = 0.009 \pm 0.002$; (b) the muons which stopped in the cryostat walls (copper) lose their polarization completely during a time of less than 10^{-9} s,

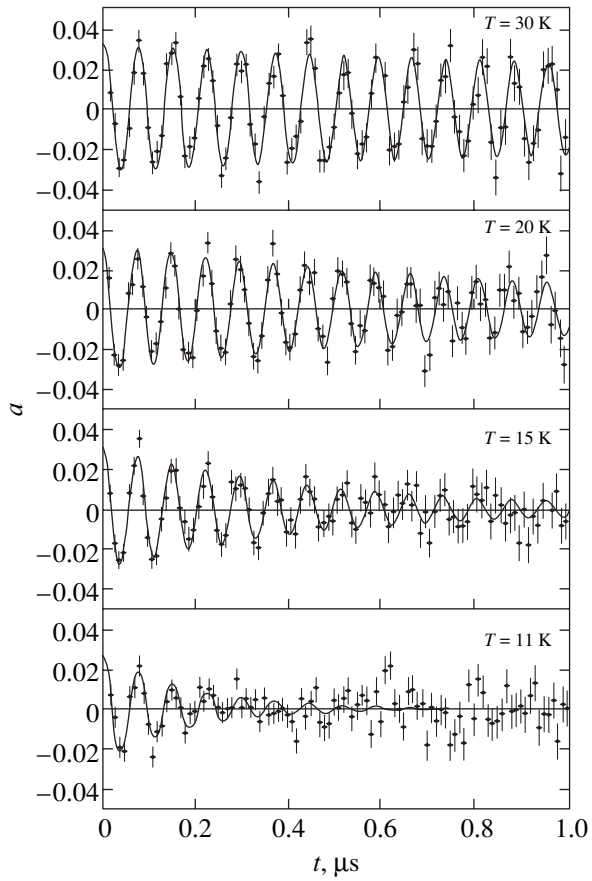


Fig. 1. Experimentally obtained μ^-SR spectra for a silicon sample with a gallium impurity (1.1×10^{15} cm^{-3}). After subtraction of the background, the data are corrected for the decay exponent, $a = \alpha p(t)$.

i.e., $p_{\mathrm{Cu}}(t) = 0$; and (c) the background of random coincidences has the form

$$B(t) = b_0 + \sum_{n=1}^{2} b_n \cos(n\omega_{hf} t + \phi_n),$$

where ω_{hf} is defined by the time structure of the proton beam.

Examples of experimentally obtained μ^-SR spectra for a silicon sample with a gallium impurity (1.1×10^{15} cm^{-3}) at different temperatures are given in Fig. 1. One can see in the figure that the precession of the polarization vector occurs at a frequency close to that of free spin precession of muon and, as the temperature decreases, the damping rate of the precession amplitude (the muon spin relaxation rate) increases.

When treating the experimental data with due regard for the muon spin relaxation at low temperatures, the polarization function for silicon has the form

$$p(t) = p_0 e^{-\lambda t} \cos(\omega t + \phi), \qquad (4)$$

where p_0 is the polarization at the initial moment of time; ω and ϕ denote the frequency and initial phase of muon spin precession in an external transverse magnetic field, respectively; and λ is the muon spin relaxation rate.

The treatment of experimental data resulted in determining the parameters of the polarization function given by Eq. (4) for negative muons that stopped in silicon. Figure 2 gives the temperature dependence of the muon spin relaxation rate and precession frequency shift for two silicon samples with a gallium impurity.

The temperature dependence of the frequency shift for all investigated samples (except for two silicon samples with a high concentration of boron impurity [B] = 1.4×10^{18} cm^{-3} and [B] = 4.1×10^{18} cm^{-3}, the experimental data for which are insufficient to determine the form of the $\Delta\omega/\omega_0$ correlation) does not contradict Curie's law of $1/T$ (see, for example, Fig. 2). This means that the contribution by the second term in expression (1) for the frequency shift is small compared with the contribution by the first, paramagnetic term. We ignored the second term in Eq. (1) to derive from the experimental data on $\Delta\omega/\omega_0$ an estimate of A_{hf}, the constant of hyperfine interaction for a muonic aluminum in silicon (the value of g was taken to be –1.07 [17]). The values of $A_{hf}/2\pi$ for all investigated samples are approximately the same and equal to $\sim 3 \times 10^7$ s^{-1}. With due regard for the ratio of the magnetic moments of muon and atomic nucleus of aluminum, we obtain $A_{hf}(\mathrm{Al})/2\pi \sim -2.5 \times 10^6$ s^{-1} for the constant of hyperfine interaction in the acceptor center formed by an aluminum atom in silicon.

These results, obtained for the temperature dependence of the muon spin relaxation rate, were described by relation (2) assuming that the relaxation rate of the magnetic moment of the acceptor center depends on temperature as

$$\nu(T) = C\tilde{T}^q. \qquad (5)$$

In treating the experimental data, it was assumed that $A_{hf}/2\pi = 3 \times 10^7$ s^{-1}. The obtained values of the parameters C and q for different samples are given in the table, where one can see that, for the first five samples in which the impurity concentration varies by more than three orders of magnitude, the values of the relaxation rate of the magnetic moment of the acceptor impurity are close to one another. However, as the impurity concentration increases further by a factor of approximately five, one observes an increase in $\nu(T = 10\ \mathrm{K})$ by more than an order of magnitude. Figure 3 gives the parameter q as a function of the impurity concentration (n) in silicon in view of the data for samples investigated by us previously [7–10] (for one of the samples, see [10], the aluminum impurity concentration was refined to 2.2×10^{17} cm^{-3}). The results given in Fig. 3 indicate that, in the impurity concentration range of up to approximately 2×10^{17} cm^{-3} for n- and p-type silicon, the value of the

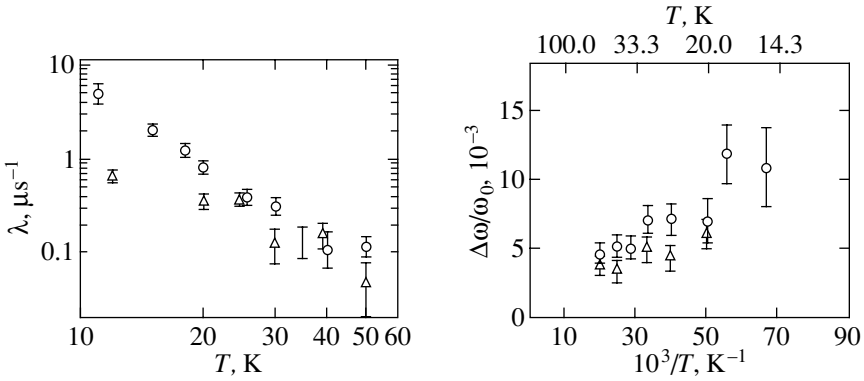


Fig. 2. The temperature dependence of the muon spin relaxation rate (on the left) and precession frequency shift (on the right) for silicon samples with different concentrations of gallium impurity: ○, 1.1×10^{15} cm^{-3}; △, 1.1×10^{18} cm^{-3}.

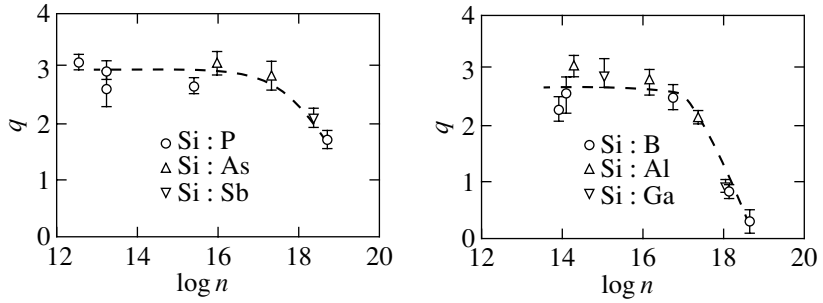


Fig. 3. The dependence of the exponent q in the function $\nu = C\tilde{T}^{q}$ in the impurity concentration n in samples of n-type (on the left) and p-type (on the right) silicon (the broken lines are drawn for graphic demonstration).

power law exponent q is virtually constant and close to three (although some spread of values is observed in p-type silicon). However, in the impurity concentration range above approximately ~10^{18} cm^{-3}, a considerable dependence of q on n is observed. For example, in p-type silicon, the value of q decreases with n by a factor of more than five. Therefore, in the impurity concentration range above 10^{18} cm^{-3}, the relaxation rate of the magnetic moment of the acceptor center ceases to depend on temperature; i.e., both the absolute value of ν and the behavior of its temperature dependence vary. The foregoing facts may serve as evidence of the variation of the mechanism of relaxation of the acceptor center magnetic moment at $n \sim 10^{18}$ cm^{-3}.

The relaxation of the acceptor magnetic moment may be caused by its interaction with the crystal lattice (spin-lattice relaxation) and by spin-exchange scattering of free charge carriers from the acceptor center.

The results of assessing the importance of various phonon processes, performed by Yafet [18] for a perfect silicon crystal, indicate that, in the temperature range of $10 \text{ K} \lesssim T \lesssim 100$ K, the key mechanism of relaxation of the magnetic moment of a shallow acceptor is the Raman phonon scattering and the relaxation rate ν depends on temperature as T^5.

The relaxation of paramagnetic centers in silicon under conditions of spin-exchange scattering of free charge carriers was treated by Semenov [19]. In a nondegenerate semiconductor, the rate of spin-exchange scattering ν_e is usually related to the spin-exchange scattering cross section δ as

$$\nu_e \approx 2\sqrt{\frac{2k_BT}{\pi m^*}}\sigma N, \tag{6}$$

and, in the case of degeneracy,

$$\nu_e \approx \hbar^{-1}N^{2/3}\sigma k_BT, \tag{7}$$

where m^* and N denote the effective mass of free charge carriers and their concentration, respectively.

In a noncompensated semiconductor, the concentration N of free charge carriers at $k_BT \ll E_i$ depends on the impurity concentration n as [20]

$$N \approx (n_0 n)^{1/2}\exp\left(-\frac{E_i}{2k_BT}\right), \tag{8}$$

where $n_0 = 2(m^*k_BT/2\pi\hbar^2)^{3/2}$ is the effective density of states in the conduction band (valence band), and E_i is the impurity ionization energy.

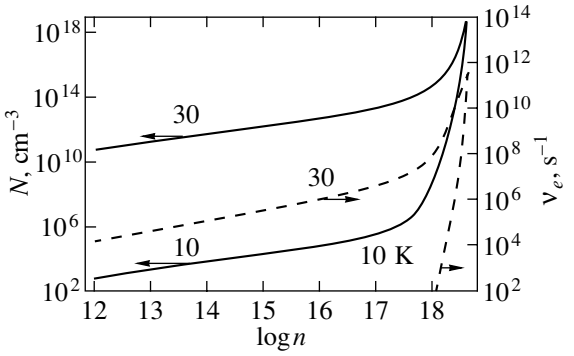


Fig. 4. Estimates of the dependence of the concentration N of free charge carriers and of the rate ν_e of spin-exchange scattering of charge carriers by the acceptor on the concentration n of impurity atoms in silicon.

It is known (see, for example, [21]) that an increase in the impurity concentration in a semiconductor is accompanied by a decrease in the impurity ionization energy. For example, in silicon, E_i decreases from the maximal value (≈45 meV) at $n \approx 10^{17}$ cm^{-3} to zero at $n \approx n_c$ (n_c is the critical concentration of impurity, which corresponds to the dielectric-metal transition; in silicon, $n_c \approx 4 \times 10^{18}$ cm^{-3} [22]). Figure 4 gives the estimates of the correlations $N = f(n)$ and $\nu_e = f(n)$, calculated by formulas (8) and (6), (7) assuming that E_i decreases linearly from 45 meV at $n = 10^{17}$ cm^{-3} to zero at $n = 4 \times 10^{18}$ cm^{-3}, with $m^* = 0.3\, m_e$ and $\sigma = \pi a^2$ (a = 10 Å [23] is a quantity of the order of the Bohr radius of the acceptor). Formula (8) is invalid in the vicinity of n_c, and, at $n = n_c$, the concentration N of charge carriers was taken to be equal to the concentration n of impurity atoms. One can see in the figure that, in the impurity concentration range above ~10^{17} cm^{-3} of interest to us, the concentration N of free charge carriers in silicon increases rapidly to approach the concentration n of impurity atoms. Also observed in this concentration range is a fast increase in the rate of spin-exchange scattering.

A comparison of the predicted values of the rate ν_e of spin-exchange scattering, given in Fig. 4, with the experimentally obtained values of the relaxation rate ν of the acceptor magnetic moment (see table) reveals that, at T = 5 to 30 K and $n \lesssim 2 \times 10^{17}$ cm^{-3}, we have $\nu_e/\nu < 10^{-4}$. When the impurity concentration rises above approximately ~2×10^{17} cm^{-3}, the rate of spin-exchange increases and, at $N \approx n = 4 \times 10^{18}$ cm^{-3} (degenerate silicon), $\nu_e \approx \nu$.

Therefore, the calculation results demonstrate that the spin-exchange scattering does not produce a sizable contribution to the relaxation of the magnetic moment of acceptor center in the impurity concentration range of up to ~2×10^{17} cm^{-3} at $T \lesssim 30$ K. Apparently, the main mechanism of relaxation in this concentration range is the Raman phonon scattering. As was already mentioned in [10], the difference between the temperature dependence of the relaxation rate of the magnetic moment of the acceptor center and the dependence of the form of T^5 [18] may be due to the fact that Yafet [18], who treated a perfect silicon crystal, disregarded the lifetime of phonons, which varies as T^{-2} at temperatures below the Debye temperature and, in a number of cases, it varies even more [24].

As the impurity concentration rises above approximately ~10^{18} cm^{-3}, the relaxation rate comes to depend much less on temperature (a decrease in the parameter q); at the same time, the relaxation rate at T = 10 K increases by more than an order of magnitude in a narrow range of concentration variation. It is known [20] that, at a high concentration of impurity, the phonon spectrum of crystal varies considerably, which may lead to a variation of the spin-lattice relaxation rate. However, in view of the fact that the experimentally observed relaxation rate of the acceptor magnetic moment is virtually constant in a wide range of impurity concentration from ~5×10^{12} to ~2×10^{17} cm^{-3}, one can hardly expect so abrupt a variation of the spin-lat-

The parameters of the temperature dependence of the relaxation rate $\nu(T) = C\tilde{T}^q$ of the magnetic moment of acceptor center, obtained as a result of treatment of the temperature dependence of the muon spin relaxation rate (n is the impurity concentration)

Impurity	n, cm^{-3}	C, 10^6 s^{-1}	q	ν, 10^9 s^{-1}	
				T = 10 K	T = 30 K
B	8.7×10^{13}	120 ± 80	2.24 ± 0.22	21	240
Ga	1.1×10^{15}	18 ± 13	2.87 ± 0.25	13	310
As	8.0×10^{15}	9 ± 6	3.03 ± 0.22	10	270
B	5.5×10^{16}	50 ± 30	2.45 ± 0.21	14	210
As	2.0×10^{17}	20 ± 15	2.80 ± 0.25	13	270
Ga	1.1×10^{18}	$(2.0 \pm 0.5) \times 10^4$	0.90 ± 0.10	158	430
B	1.4×10^{18}	$(2.4 \pm 0.6) \times 10^4$	0.81 ± 0.12	155	380
B	4.1×10^{18}	$(2.5 \pm 1.3) \times 10^5$	0.28 ± 0.21	476	650

tice relaxation rate with an increase in concentration above approximately ~10^{18} cm^{-3}. On the other hand, at $n > 10^{18}$ cm^{-3}, a transition to metallic conduction occurs in silicon [22] and, as follows from the foregoing estimates, the spin-exchange scattering of free charge carriers may become the predominant mechanism of relaxation of the acceptor magnetic moment.

4. CONCLUSION

1. The temperature dependence of the relaxation rate of the magnetic moment of an acceptor center in silicon for different concentrations of impurity has been investigated; it is demonstrated that, at $n < 2 \times 10^{17}$ cm^{-3}, the experimental data are well described by the empirical correlation $\nu(T) = CT^q$ with the value of $q \approx 3$.

2. An estimate of the constant of hyperfine interaction for an acceptor center of aluminum in silicon has been produced for the first time ever: $|A_{hf}(\mathrm{Al})/2\pi| \sim 2.5 \times 10^6$ s^{-1}.

3. The behavior of the temperature dependence of the relaxation rate of the magnetic moment of acceptor center has been found to change in the impurity concentration range corresponding to the semiconductor-metal transition.

4. The results of analysis of the data lead one to conclude that, with the values of the impurity concentration of less than 2×10^{17} cm^{-3}, the spin-lattice interaction is the key mechanism of relaxation of acceptor center in silicon. In the range of higher values of impurity concentration, a significant contribution to relaxation may be made by the spin-exchange scattering of free charge carriers, which is apparently the key mechanism of relaxation of the acceptor magnetic moment in the metal phase.

ACKNOWLEDGMENTS

We are grateful to the management of the Paul Scherrer Institute (Switzerland) for making it possible for us to perform the above-mentioned measurements. These investigations were partly financed by the German Federal Ministry for Education and Research (Bundesministerium für Bildung und Forschung), contracts 03-MA5ST1 and 03-MA5ST2.

REFERENCES

1. G. D. Watkins, Fiz. Tverd. Tela (St. Petersburg) **41**, 826 (1999) [Phys. Solid State **41**, 746 (1999)].
2. A. Baldereschi and N. O. Lipari, Phys. Rev. B **8**, 2697 (1973).
3. A. V. Malyshev, I. A. Merkulov, and A. V. Rodina, Fiz. Tekh. Poluprovodn. (St. Petersburg) **30**, 159 (1996) [Semiconductors **30**, 91 (1996)].
4. T. L. Linnik and V. I. Sheka, Fiz. Tverd. Tela (St. Petersburg) **41**, 1556 (1999) [Phys. Solid State **41**, 1425 (1999)].
5. G. W. Ludwig and H. H. Woodbury, *Electron Spin Resonance in Semiconductors* (Academic, New York, 1962).
6. V. N. Gorelkin and V. P. Smilga, Zh. Éksp. Teor. Fiz. **66**, 1201 (1974) [Sov. Phys. JETP **39**, 586 (1974)].
7. V. N. Gorelkin, V. G. Grebinnik, K. I. Gritsaĭ, *et al.*, Yad. Fiz. **56** (10), 29 (1993) [Phys. At. Nucl. **56**, 1316 (1993)].
8. V. N. Gorelkin, V. G. Grebinnik, K. I. Gritsaĭ, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 539 (1996) [JETP Lett. **63**, 566 (1996)].
9. T. N. Mamedov, V. N. Duginov, D. Gerlach, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 61 (1998) [JETP Lett. **68**, 64 (1998)].
10. T. N. Mamedov, D. G. Andrianov, D. Gerlach, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 637 (2000) [JETP Lett. **71**, 438 (2000)].
11. V. N. Gorelkin, T. N. Mamedov, and D. V. Rubtsov, Hyperfine Interact. C **1**, 191 (1996).
12. A. S. Baturin and V. N. Gorelkin, Physica B (Amsterdam) **289-290**, 578 (2000).
13. V. N. Gorelkin, T. N. Mamedov, and A. S. Baturin, Physica B (Amsterdam) **289-290**, 585 (2000).
14. G. L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors* (Nauka, Moscow, 1972; Wiley, New York, 1975).
15. R. Scheuermann, J. Schmidl, A. Seeger, *et al.*, Hyperfine Interact. **106**, 295 (1997).
16. T. Suzuki, D. F. Measday, and J. P. Roalsvig, Phys. Rev. C **35**, 2212 (1987).
17. H. Neubrand, Phys. Status Solidi B **86**, 269 (1978).
18. Y. Yafet, J. Phys. Chem. Solids **26**, 647 (1965).
19. Yu. G. Semenov, Fiz. Tverd. Tela (Leningrad) **22**, 3190 (1980) [Sov. Phys. Solid State **22**, 1868 (1980)].
20. C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1976; Nauka, Moscow, 1978).
21. V. I. Fistul', *Heavily Doped Semiconductors* (Nauka, Moscow, 1967; Plenum, New York, 1969).
22. P. Ph. Edwards and M. J. Sienko, Phys. Rev. B **17**, 2575 (1978).
23. R. C. Enck and A. Honig, Phys. Rev. **177**, 1182 (1969).
24. I. Ya. Pomeranchuk, *Collection of Scientific Works* (Nauka, Moscow, 1972), Vol. 1.

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 1010–1018.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1166–1175.

SOLIDS
Structure

Phase Transitions in Ferromagnetic $Ni_{2+x}Mn_{1-x}Ga$ Alloys with Regard for the Modulation Order Parameter

V. D. Buchel'nikov[a, *], A. T. Zayak[a], A. N. Vasil'ev[b], D. L. Dalidovich[b], V. G. Shavrov[c, **], T. Takagi[d], and V. V. Khovailo[d]

[a]*Chelyabinsk State University, Chelyabinsk, 454021 Russia*
[b]*Moscow State University, Vorob'evy gory, Moscow, 119899 Russia*
[c]*Institute of Radio Engineering and Electronics, Russian Academy of Sciences, ul. Mokhovaya 18, Moscow, 103907 Russia*
[d]*Institute of Fluid Science, Tohoku University, Sendai, 980-8577 Japan*
**e-mail: buche@cgu.chel.su*
***e-mail: shavrov@mail.cplire.ru*
Received October 12, 2000

Abstract—The phase diagram of ferromagnetic alloys $Ni_{2+x}Mn_{1-x}Ga$ is reconstructed on the basis of temperature dependences of the resistance. It is seen from this diagram that for small x, structural transitions from the cubic to the tetragonal phase are preceded by structural transformations in the cubic phase. In the framework of the phenomenological Landau theory of phase transitions, phase diagrams of the structural and magnetic phase transitions in these alloys are analyzed with regard for the modulation order parameter. It is shown that premartensitic and postmartensitic phase transitions related to the appearance of the modulated structure can occur along with martensitic transformations. The strain and modulation order parameters substantially affect the magnetic phase transitions via the interaction with the magnetic order parameter. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Ferromagnetic Heusler alloys $Ni_{2+x}Mn_{1-x}Ga$ are of major interest in connection with the fact that they can exhibit the effect of huge magnetostriction reaching values up to 6% [1]. This effect is due to a reconstruction of interconnected magnetic and structural domains, which stimulates interest in studying phase transitions in these alloys. Of major interest are the so-called "martensitic transformations," which are first-order phase transitions from the initial highly symmetric phase to the low symmetric phase, which exists at low temperatures [2]. When martensitic transformations are described in the framework of the phenomenological Landau theory, the main order parameters are the components of the strain tensor. In this case, martensitic transformations are described with the help of the density of free elastic energy with regard for the anharmonicity (the proper phase transition) [3]. If martensitic transformations occur in the ferromagnetic matrix, they can be accompanied by changes in the magnetic subsystem due to the magnetoelastic interaction that exists in ferromagnetics. As a result, martensitic transformations cause magnetic phase transitions in ferromagnetics.

At high temperatures, the point symmetry group O_h of the ferromagnetic Heusler alloys $Ni_{2+x}Mn_{1-x}Ga$ has the cubic structure. When the temperature decreases, phase transitions to the tetragonal structure occur in these alloys [4]. Martensitic phase transformations and the related magnetic phase transitions in $Ni_{2+x}Mn_{1-x}Ga$ are studied in [5–12] in detail in the framework of the Landau theory.

Presently, it has been established that martensitic transitions can be accompanied by the appearance of an intermediate modulated phase that is characterized by the set of wave vectors $\mathbf{q}_i$ [13]. In experiments, the presence of the intermediate phase manifests itself in softening of the corresponding phonon modes [13, 14]. Experiments show that in $Ni_{2+x}Mn_{1-x}Ga$, the transition from the initial structure to the modulated phase is a first-order phase transition [15]; however, the jumps of the order parameters and heat capacity are substantially smaller than those of the martensitic transformation. It was shown in [14, 16] that the structural phase transition to the modulated phase can be described in the framework of the Landau theory using free energy, which includes the elastic energy with regard for anharmonicity, terms describing the modulated structure, and terms that account for the interaction of the tetragonal strains with the modulation order parameter. In this case, the tetragonal strains are the secondary order parameters, which are related to the primary parameters describing the modulated structure. Strains appear simultaneously with the modulation at the point of the structural transition to the modulated phase. Their magnitude is determined by the magnitude of the coupling coefficient with the modulated order parameter.

$Ni_{2+x}Mn_{1-x}Ga$ alloys occupy a special place among a large number of solids in which both martensitic

phase transitions without the appearance of the modulated structure and transitions with an intermediate modulated structure occur. X-ray experiments show that the low-temperature martensitic phase of these alloys contains both strong tetragonal distortions of the lattice ($c/a \approx 0.94$) and a modulation of the tetragonal structure with the polarization vector parallel to the $[1\bar{1}0]$ axis and the wave vector oriented along the [110] axis [17, 18]. The first studies revealed only one point ($T = 202$ K) of transition to the martensitic phase with modulation and tetragonal distortions. In later experiments [19], two different points of phase transitions were discovered. The first transition, which is observed at a higher temperature, is the transition from the symmetric cubic structure to the modulated structure with the wave and polarization vectors mentioned above. This intermediate phase, which was called *premartensitic* in [19], has a modulation vector with the module 1/3. This phase has an approximately cubic structure and is accompanied by the appearance of small tetragonal strains. The second phase transition is the transition from the intermediate modulated structure to the actually martensitic phase, which is modulated and has large tetragonal strains. Thus, the phenomenological approach described in [2, 5–12, 14, 16] must be generalized so as to describe two successive phase transitions.

Another reason for which $Ni_{2+x}Mn_{1-x}Ga$ attract much attention is that the structural phase transitions in these alloys occur in the ferromagnetic matrix. It turns out that structural transformations affect not only the direction of the magnetization vector but also its magnitude [5, 9, 11, 12]. Certain experiments show that the magnetization increases in the course of the transition to the modulated phase [20]. This fact must be taken into account in the phenomenological description.

The actual magnetic structure of the tetragonal phase in $Ni_{2+x}Mn_{1-x}Ga$ has not been thoroughly studied. It is only known from the measurements of the magnetic susceptibility that the structural transition to the tetragonal phase is accompanied by the appearance of strong magnetic anisotropy [4]. In [21], a phenomenological model of the inhomogeneous martensitic structure that appears in the alloy Ni_2MnGa below the temperature of the martensitic transition is constructed, and estimates of the static magnetic susceptibility and magnetization of the martensite are obtained.

Recently, theoretical paper [22] was published, which is dedicated to modeling the premartensitic phase transition in Ni_2MnGa by the Monte Carlo method. The authors of [22] conclude that premartensitic effects are caused by the magnetoelastic coupling between the magnetic subsystem and the phonon mode, which softens as the modulation order is formed. The premartensitic transition occurs in the case when this magnetoelastic interaction is sufficiently large. This result suggests that the magnetoelastic interaction must be taken into account in the phenomenological theory. The premartensitic phase obtained in [22] does not contain tetragonal strains. No tetragonal strains were actually observed in the premartensitic phase [13]; however, this can be due to the fact that they are small. Thus, a theoretical study of tetragonal strains caused by the modulation order is necessary. Effects that occur, for example, in the magnetic subsystem and accompany small tetragonal strains in the premartensitic phase could reveal the presence of the strains if they really exist.

Experimental studies presented in [11, 12] made it possible to establish composition dependences of the temperature of the magnetic and martensitic phase transition in alloys $Ni_{2+x}Mn_{1-x}Ga$ for $0 \le x \le 0.2$. It turned out that, as the content of Ni at the expense of Mn increases, the temperature of the structural transition also increases and the temperature of the magnetic transition decreases so that they coincide in the interval $0.17 \le x \le 0.19$. The composition dependences determined from the measurements of the resistance, magnetic susceptibility, and heat capacity coincide within the accuracy of the experimental techniques. This fact makes it possible to use any of these approaches to refine the phase diagram of the alloys $Ni_{2+x}Mn_{1-x}Ga$.

In the present paper, the phase diagram of the alloys $Ni_{2+x}Mn_{1-x}Ga$ in the interval $0 \le x \le 0.09$ is analyzed on the basis of measuring the resistance of polycrystalline samples. Along with the specific features caused by the structural and magnetic phase transitions, anomalies that indicate the existence of premartensitic transformations were observed. In order to explain the results obtained, the phenomenological Landau theory of phase transitions, which takes into account interactions between the magnetic, strain, and modulation order parameters, is extended to the case of cubic ferromagnetics. It is shown that the so-called premartensitic and postmartensitic phase transitions can occur in $Ni_{2+x}Mn_{1-x}Ga$ along with martensitic transformations. Formers are connected with the appearance of the modulated structure. The strain and modulation order parameters cause magnetic phase transitions in ferromagnetics by way of the interaction with the magnetic order parameter. This fact leads to a substantial variety of possible magnetic states in crystals.

2. EXPERIMENT

Polycrystals with the composition described above were prepared by arc melting in the argon atmosphere from high-purity components. To perform transport measurements, samples were cut from massive polycrystals by spark cutting; then, current and potential contacts were welded on those samples. In order to eliminate the thermoelectric voltage, the measurements were performed in a wide temperature range using the four-contact scheme for the alternate current with the frequency $f \sim 1$ kHz; the results were registered using phase-conscious measurements. The temperature was

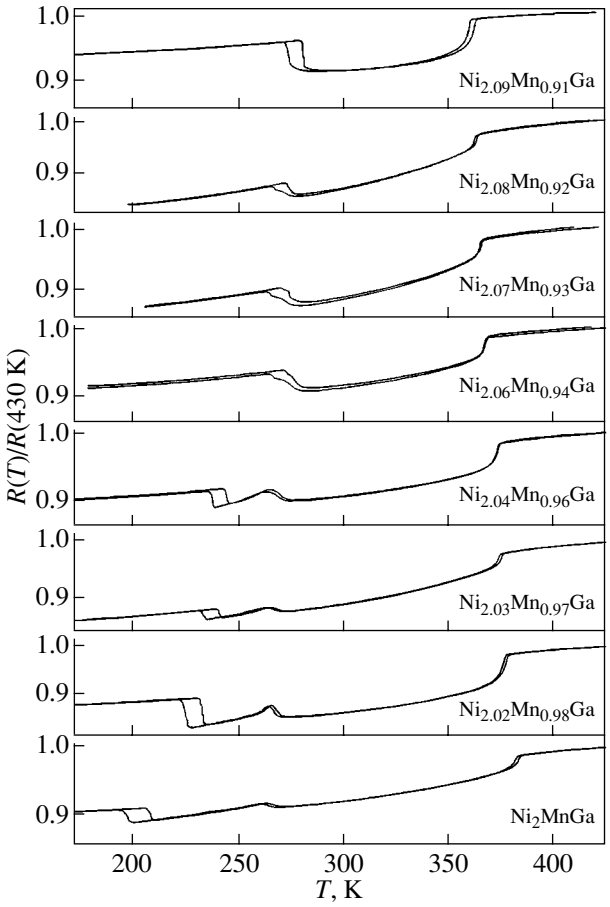


Fig. 1. Temperature dependences of the resistance for $Ni_{2+x}Mn_{1-x}Ga$ polycrystals.

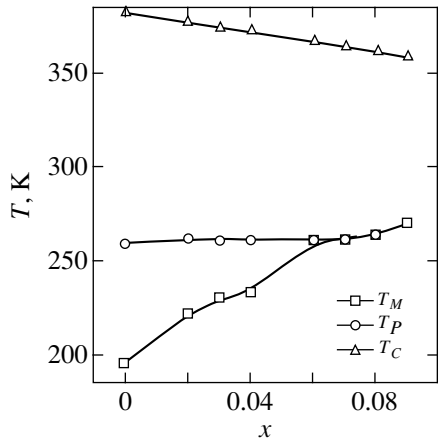


Fig. 2. The *T*–*x* phase diagram of $Ni_{2+x}Mn_{1-x}Ga$ alloys for $0 \le x \le 0.09$ (experimental).

measured by a platinum resistance thermometer that was placed at the distance of ≈0.15 cm from the sample.

The temperature dependences of the resistance of the polycrystals $Ni_{2+x}Mn_{1-x}Ga$ for $0 \le x \le 0.09$ are presented in Fig. 1. In the sample of the stoichiometric Ni_2MnGa, at the Curie point $T_C \approx 380$ K, there is a step in the dependence $R(T)$, which is characteristic of second-order phase transitions. This step is caused by a decrease of charge carriers scattering from magnetic fluctuations. As the temperature decreases further, the resistance behaves like that in metals; however, at $T_P \approx 260$ K, there is a peak, and at $T_M \approx 200$ K another step is observed. The anomaly at T_M has a well-marked hysteresis, which is characteristic of first-order phase transitions. It is well known (see [17]) that the transition from the cubic to the tetragonal phase occurs at this temperature. It seems that the anomaly at T_P corresponds to the premartensitic transformation in the cubic phase when a static modulation wave appears in the crystal [19]. It is difficult to determine from the above measurements whether the anomaly at T_P is a consequence of the first- or second-order phase transition; this is due to the fact that there is a small delay between the temperatures of the sample and the thermometer. It was shown in [11, 12] that the Curie temperature decreases monotonically with increasing x, whereas the temperature of the structural transition increases (Fig. 2). The temperature of the premartensitic transformation in the samples examined is almost independent of the concentration and coincides with the temperature of the structural transition at $x \approx 0.07$–0.09.

3. THEORY

The sequence of phase transitions observed in $Ni_{2+x}Mn_{1-x}Ga$ can be described in the framework of the Landau theory of structural phase transitions for cubic crystals. In such crystals, both modulated and tetragonal states can occur at low temperatures. Distortions of the cubic lattice under structural transitions are described by homogeneous strains written in the form of linear combinations of the strain tensor components e_{ii} [2]:

$$e_1 = \frac{1}{3}(e_{xx} + e_{yy} + e_{zz}), \quad e_2 = \frac{1}{\sqrt{2}}(e_{xx} - e_{yy}),$$
$$e_3 = \frac{1}{\sqrt{6}}(2e_{zz} - e_{xx} - e_{yy}). \tag{1}$$

The strain e_1, which corresponds to the change of the volume, does not violate the symmetry of the lattice. A symmetry violation occurs due to the strains e_2 and e_3, which are responsible for the transition of the lattice from the cubic to the tetragonal phase. This transition is accompanied by softening of the combination of the elastic moduli $C_{11} - C_{12}$. The total expression for the density of the free energy must also include the strains $e_4 = e_{xy}$, $e_5 = e_{yz}$, and $e_6 = e_{zx}$, which lead to a distortion of the elementary cell down to a symmetry lower than the tetragonal one.

In order to describe acoustic phonon modes of the type (1/3, 1/3, 0) in the crystal, it is necessary to note

that six different orientations of the wave modulation vector exist. They can be written in the form

$$\mathbf{k}_1 = \zeta(1, 1, 0), \quad \mathbf{k}_2 = \zeta(0, 1, 1), \quad \mathbf{k}_3 = \zeta(1, 0, 1),$$
$$\mathbf{k}_4 = \zeta(1, \bar{1}, 0), \quad \mathbf{k}_5 = \zeta(0, 1, \bar{1}), \quad \mathbf{k}_6 = \zeta(\bar{1}, 1, 0),$$

where $\zeta = 1/3$. Thus, as a rule, an order parameter should exist which consists of 12 components (six amplitudes and six phases): $\psi_1 \ldots \psi_6$, $\varphi_1 \ldots \varphi_6$, where

$$\psi_j = |\psi_j| \exp(i\varphi_j).$$

The atomic displacements corresponding to each of these order parameters have the form

$$\mathbf{u}_j(\mathbf{r}) = |\psi_j| \mathbf{p}_j \sin(\mathbf{k}_j \cdot \mathbf{r} + \varphi_j),$$

where $\mathbf{p}_1, \ldots, \mathbf{p}_6$ are the unit polarization vectors oriented along the axes $[\bar{1}, 1, 0]$, $[0, \bar{1}, 1]$, $[1, 0, \bar{1}]$, [1, 1, 0], [0, 1, 1], and [1, 0, 1], respectively. The total expression for the density of free energy must be invariant under the spatial transformations of the point symmetry group O_h. It consists of terms of three types:

$$F = F_e(e_j) + F_\psi(\psi_i) + F_{e\psi}(e_j, \psi_i). \tag{2}$$

Here, $F_e(e_j)$ is the density of free energy containing the terms that are responsible for the anharmonicity of the elastic subsystem with respect to the order parameter (e_2, e_3). It has the form (see [2])

$$F_e(e_j) = \frac{1}{2}a(e_2^2 + e_3^2) + \frac{1}{3}be_3(e_3^2 - 3e_2^2) + \frac{1}{4}c(e_2^2 + e_3^2)^2 + \frac{1}{2}(C_{11} + 2C_{12})e_1^2 + \frac{1}{2}C_{44}(e_4^2 + e_5^2 + e_6^2), \tag{3}$$

where the coefficients a, b, and c are linear combinations of the components of the elastic moduli of the second, third, and fourth orders, respectively:

$$a = C_{11} - C_{12}, \quad b = \frac{1}{6\sqrt{6}}(C_{111} - 3C_{112} + 2C_{123}),$$
$$c = \frac{1}{48}(C_{1111} + 6C_{1112} - 3C_{1122} - 8C_{1123}). \tag{4}$$

Since Eq. (3) includes terms of the third order, the phase transition with respect to the order parameter (e_2, e_3) has the first order. When approaching the point of the structural transition to the tetragonal phase, the elastic modulus $a = C_{11} - C_{12}$ tends to zero; in the vicinity of the transition point ($T \longrightarrow T_M$), it can be written as

$$a = a_0(T - T_M)/T_M,$$

where T_M is the temperature of the martensitic transition.

The total expression for $F_\psi(\psi_i)$ can be found in [16, 23]. Here, we consider the simplest case of the modulation that takes into account only one phonon mode 1/3 (1, 1, 0). It is described by the order parameter $\psi = |\psi|\exp(i\varphi)$ (the index is omitted for simplicity). Then, the expression for the density of the modulation part of free energy can be written in the form

$$F_\psi(\psi) = \frac{1}{2}A|\psi|^2 + \frac{1}{4}B|\psi|^4 + \frac{1}{6}C_0|\psi|^6 + \frac{1}{6}C_1[\psi^6 + (\psi^*)^6]. \tag{5}$$

The last term in this formula can be minimized with respect to the phase:

$$[\psi^6 + (\psi^*)^6] = |\psi|^6(\exp(-i6\varphi) + \exp(i6\varphi)) = 2|\psi|^6\cos(6\varphi).$$

The minimum of energy (5) is attained at $\varphi = \pm\pi/6$, $\pm\pi/2$, and $\pm 5\pi/6$ when $C_1 > 0$, and at $\varphi = 0$, $\pm\pi/3$, $\pm 2\pi/3$, and π when $C_1 < 0$. In Eq. (5), we set $C' = C_0 - |C_1|$ and assume that $C' > 0$ for stability. The parameter A depends on T; in the vicinity of the temperature of the transition to the modulated state ($T \longrightarrow T_P$), it can be written in the form

$$A = A_0(T - T_P)/T_P.$$

The energy $F_{e\psi}$ relating the strains e_i to the order parameter that describes the modulation must be invariant under all symmetry operations connected with e_i and ψ_j. When only the phonon mode 1/3(1, 1, 0) is taken into account, $F_{e\psi}$ has the form

$$F_{e\psi}(\psi, e_i) = \left(\frac{1}{\sqrt{3}}D_1e_1 + \frac{2}{\sqrt{6}}D_2e_3 + D_3e_4\right)|\psi|^2. \tag{6}$$

Equations (3), (5), and (6) completely determine the density of the free energy of the cubic crystal and allow for the description of phase transitions from the cubic phase to the modulated and tetragonal phases.

It follows from experiments that structural transformations in $Ni_{2+x}Mn_{1-x}Ga$ occur in the ferromagnetic matrix, which makes it necessary to take into account the influence of the magnetic subsystem on the structural transitions. In these alloys, both in the cubic and tetragonal phases, manganese atoms are the main carriers of the magnetic moment [17]. This makes it possible to describe the $Ni_{2+x}Mn_{1-x}Ga$ alloys by a single-lattice magnetic subsystem with the vector of the macroscopic magnetization **M**. The free energy also includes terms that connect **M** with other parameters of the system. For the sake of convenience, we introduce the dimensionless magnetization vector $\mathbf{m} = \mathbf{M}/M_0$, where M_0 is the saturation magnetization.

The contribution of the magnetic subsystem to the total energy of the cubic ferromagnetic consists of two terms. The first one is of exchange origin. It is necessary to take into account the dependence of the magni-

tude of the magnetization vector **m** on temperature. This term has the form

$$F_{ex}(m) = \frac{1}{2}\alpha(m_x^2 + m_y^2 + m_z^2) + \frac{1}{4}\delta_1(m_x^2 + m_y^2 + m_z^2)^2. \quad (7)$$

Here α and δ_1 are the exchange constants. The parameter of the exchange interaction α depends on temperature; it can be written in the following form in the vicinity of the Curie point:

$$\alpha = \alpha_0(T - T_C)/T_C.$$

The second term represents the energy of magnetic anisotropy of the cubic ferromagnetic; it can be written in the form

$$F_a(m_i) = K_1(m_x^2m_y^2 + m_y^2m_z^2 + m_z^2m_x^2), \quad (8)$$

where K_1 is the first cubic anisotropy constant.

The free energy must also include terms that relate the components of **m** to other order parameters of the system. The first of them relates the magnetization components m_i to the strains e_i; it has the form

$$F_{me}(m_i, e_i) = \frac{1}{\sqrt{3}}B_1e_1\mathbf{m}^2 + B_2\left[\frac{1}{\sqrt{2}}e_2(m_x^2 - m_y^2) + \frac{1}{\sqrt{6}}e_3(3m_z^2 - \mathbf{m}^2)\right] + B_3(e_4m_xm_y + e_5m_ym_z + e_6m_zm_x). \quad (9)$$

This expression is the simplest form of the magnetoelastic energy with the phenomenological magnetoelastic constants B_1, B_2, and B_3.

The second term describes the interaction of the components of the magnetization vector m_i with the modulation order parameter ψ; it can be written in the form

$$F_{m\psi}(m_i, \psi) = \left[\frac{1}{3}N_1\mathbf{m}^2 + N_2\left(m_z^2 - \frac{1}{3}\mathbf{m}^2\right) + N_3m_xm_y\right]|\psi|^2. \quad (10)$$

Here, the coefficients N_i are the coupling parameters of the magnetic and modulation subsystems.

As a result, the final expression for the density of free energy has the form

$$F = F_e(e_i) + F_\psi(|\psi|^2) + F_{e\psi}(e_i, |\psi|^2) + F_{ex}(m) + F_a(m_i) + F_{me}(m_i, e_j) + F_{m\psi}(m_i, |\psi|^2). \quad (11)$$

This equation describes three phase transitions: the Curie point, the transition to the modulated state, and the transition to the tetragonal phase.

The equation for the density of free energy (11) includes variables that are not responsible for phase transitions, i.e., those that are indirect order parameters e_1, e_4, e_5, and e_6. After the minimization of energy with respect to these variables, certain constants involved in (11) are renormalized:

$$B' = B - 2\left(\frac{D_1^2}{3(C_{11} + 2C_{12})} + \frac{D_3^2}{C_{44}}\right), \quad K = K_1 - \frac{B_3^2}{2C_{44}},$$

$$N_1' = \frac{1}{3}N_1 - \frac{D_1B_1}{6(C_{11} + 2C_{12})}, \quad N_3' = N_3 - \frac{B_3D_3}{C_{44}},$$

$$\delta = \delta_1 - \frac{2B_1^2}{3(C_{11} + 2C_{12})}.$$

With regard for this renormalization, the expression for the density of free energy takes the form

$$F = \frac{1}{2}a(e_2^2 + e_3^2) + \frac{1}{3}be_3(e_3^2 - 3e_2^2) + \frac{1}{4}c(e_2^2 + e_3^2)^2 + \frac{1}{2}A|\psi|^2 + \frac{1}{4}B'|\psi|^4 + \frac{1}{6}C'|\psi|^6 + \frac{2}{\sqrt{6}}D_2e_3|\psi|^2 + \frac{1}{2}\alpha\mathbf{m}^2 + \frac{1}{4}\delta\mathbf{m}^4 + K(m_x^2m_y^2 + m_y^2m_z^2 + m_z^2m_x^2) + B_2\left[\frac{1}{\sqrt{2}}e_2(m_x^2 - m_y^2) + \frac{1}{6}e_3(3m_z^2 - \mathbf{m}^2)\right] + \left[N_1'\mathbf{m}^2 + N_2\left(m_z^2 - \frac{1}{3}\mathbf{m}^2\right) + N_3'm_xm_y\right]|\psi|^2. \quad (12)$$

In order to determine all equilibrium states of the alloys under consideration, it is necessary to minimize function (12) with respect to the remaining order parameters e_2, e_3, $|\psi|$, m_1, m_2, and m_3. The system of nonlinear algebraic equations obtained can only be solved numerically. To perform the numerical computation on the basis of the available experimental data [11–13, 15, 17, 19, 20], the following values of the parameters involved in Eq. (12) were used: $a_0 = 10^{11}$ erg/cm^3, $b = 3 \times 10^{11}$ erg/cm^3, $c = 3 \times 10^{12}$ erg/cm^3, $D_2 = 10^3$ erg/cm^3, $A_0 = 10^{23}$ erg/cm^3, $B' = 10^{38}$ erg/cm^3, $C' = 10^{55}$ erg/cm^3, $K = -10^5$ erg/cm^3, $B_2 = 1.5 \times 10^7$ erg/cm^3, $\alpha_0 = -10^9$ erg/cm^3, $\delta = 10^9$ erg/cm^3, $N_1' = 10^3$ erg/cm^3, $N_2 = -10^2$ erg/cm^3, and $N_3' = -10^2$ erg/cm^3. It is seen from Fig. 2 that the temperature of the magnetic, T_C, and structural, T_M, phase transitions depend on x approximately linearly, and the temperature of the premartensitic transformation $T_P = 260$ K is almost independent of the composition of the alloys analyzed. Thus, the composition dependences of these temperatures can be written in the form

$$T_C = T_{C0} - \gamma x, \quad T_M = T_{M0} + \kappa x,$$

where

$$T_{C0} = 390 \text{ K}, \quad \gamma = 175 \text{ K},$$

$$T_{M0} = 200 \text{ K}, \quad \kappa = 700 \text{ K}.$$

The phase T–x diagram of the Heusler alloys $Ni_{2+x}Mn_{1-x}Ga$ calculated for the parameters indicated above is shown in Fig. 3. In is seen that the following phases can be realized as the temperature and the composition vary: the paramagnetic cubic phase (1), the ferromagnetic cubic phase (2) with the magnetization along the axis [111], the ferromagnetic angular quasi-cubic modulated phase (3) with the magnetization in the plane of the type (110), and the ferromagnetic tetragonal modulated phase (4) with the magnetization along the axis [001]. The line $C'C$ is the line of the magnetic second-order phase transition between the paramagnetic and ferromagnetic cubic phases. $T'T$ is the line of the first-order phase transition between the ferromagnetic phases (2) and (3). It is accompanied by the occurrence of modulations of the cubic lattice, which results in a small tetragonal distortion. $N'N$ and $O'O$ are the lines of the loss of stability of phases (2) and (3), respectively. $B'B$ is the line of the martensitic phase transition between the ferromagnetic modulated phases (3) and (4). This transition is accompanied by the occurrence of large tetragonal distortions of the lattice. $G'G$ and $F'F$ are the lines of the loss of stability of phases (3) and (4), respectively. It is seen from Fig. 3 that the martensitic transition is accompanied by a large hysteresis, while the hysteresis of the transition from the ferromagnetic cubic phase to the ferromagnetic quasi-cubic one is very small. These facts are in good agreement with experimental results on measuring the resistance (see Fig. 1). For large x, the hysteresis regions of the premartensitic and martensitic transitions partially intersect. As a result, these transitions partially superimpose on one another and are difficult to distinguish experimentally (Fig. 1).

4. CONCLUSIONS

The experimental and theoretical study of the phase diagram of $Ni_{2+x}Mn_{1-x}Ga$ alloys in the interval $0 \le x \le 0.1$ conducted in this paper made it possible to reveal an important specific feature in the behavior of these materials, consisting in the occurrence of static modulations of the crystal lattice in the cubic and tetragonal phases. The phenomenological theory of phase transitions of $Ni_{2+x}Mn_{1-x}Ga$ alloys made it possible to describe the sequence of two structural transitions: from the cubic lattice to the modulated phase with small tetragonal distortions (the premartensitic transition), and then from the modulated quasi-cubic to the martensitic structure with large tetragonal distortions and modulation. These transitions are accompanied by magnetic orientational transitions.

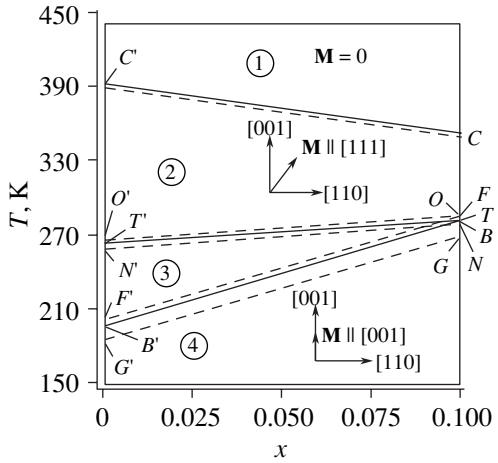


Fig. 3. The T–x phase diagram of alloys $Ni_{2+x}Mn_{1-x}Ga$ for $0 \le x \le 0.1$ (theoretical).

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 99-02-18247; by the Ministry for Education, project no. 97-0-7.0-11; and by the Russian Universities program.

APPENDIX

When constructing the theoretical phase diagram, we assumed that only the elastic moduli a and the constant of exchange interaction α depended on temperature and composition. Variations of other parameters in (12) can lead to a qualitative change of the phase diagram. For example, it is interesting to see how the phase diagram of $Ni_{2+x}Mn_{1-x}Ga$ changes when the parameters A and D_2, which are responsible for the modulation of the crystal lattice, vary.

Figure 4 presents the phase diagram of $Ni_{2+x}Mn_{1-x}Ga$ on the coordinates a and A for $b > 0$, $D_2 > 0$, and $N_{1,2,3} < 0$ in the case when the alloy is in the ferromagnetic state. It is seen from the diagram that in this case the material can be in one of five states. In the phase diagram, OC is the line of the structural and magnetic orientational first-order phase transitions from the high-temperature cubic (austenitic) phase $FC_{[111]}$ to the low-temperature (martensitic) phase $FT_{[001]}$, which is characterized by large tetragonal distortions and no modulation. The phase $FC_{[111]}$ is stable above the line GH, and the phase $FT_{[001]}$ is stable below MI. It is seen by the region of their intersection that this transition is accompanied by a large hysteresis. On the curve OB, the structural and magnetic orientational first-order phase transitions from the symmetric phase $FC_{[111]}$ to the modulated angular phase $FCM_{[uuw]}$ occur. The initial cubic phase is stable to the right of the curve JG, while the modulated one is stable to the left of QF. We note that the phase $FCM_{[uuw]}$ is not strictly cubic. It has

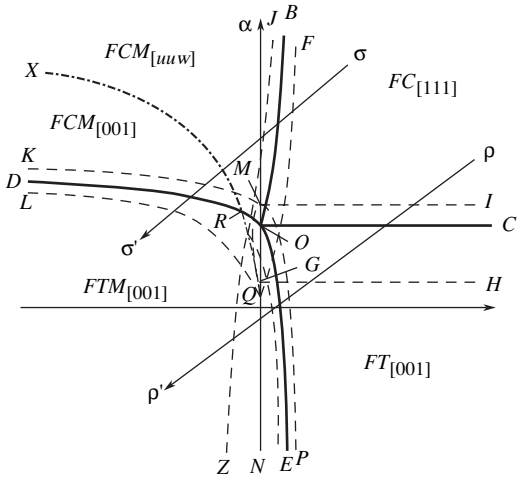


Fig. 4. The phase diagram of the Ni–Mn–Ga alloy on the coordinates $a - A$ for $b > 0$, $D_2 > 0$, and $N_{1,2,3} < 0$. Solid curves are the curves of first-order phase transitions, dot-and-dash curves correspond to the second-order phase transition, and dashed curves are the boundaries of stability regions of the phases; σσ' and ρρ' are thermodynamic paths; *F* denotes the ferromagnetic phase, *M* is the modulated phase, *C* the phase with a cubic lattice, and *T* the phase with a tetragonal lattice. The direction of the magnetization vector of the phases is indicated in square brackets.

small tetragonal distortions that appear along with the modulation. In this phase, the magnitude of tetragonal strains is determined by the parameter D_2. If we set $D_2 = 0$, then there will be no tetragonal strains. On the curve *RX*, the magnetic orientational second-order phase transition from the angular phase $FCM_{[uuw]}$ to the symmetric phase $FCM_{[001]}$ occurs. This phase transition is caused by the interaction of the modulation order parameter with the strain or magnetic order parameters. The existence of this phase transition depends on the magnitude of the interactions and, therefore, on the magnitude of small tetragonal distortions in quasi-cubic phases. *RO* is the curve of the martensitic and, simultaneously, the magnetic orientational first-order phase transition from the angular quasi-cubic phase $FCM_{[uuw]}$ to the symmetric tetragonal phase $FTM_{[001]}$, which has large tetragonal distortions. The existence of this phase transition is also determined by the interaction of the modulation order parameter with the strain or magnetic order parameters. On the curve *RD*, the martensitic phase transition from the symmetric quasi-cubic modulated phase $FCM_{[001]}$ to the tetragonal symmetric phase $FTM_{[001]}$ occurs. It is accompanied by a large jump of tetragonal strains. On this curve, no reorientation of the magnetization vector and, therefore, no magnetic phase transition occur.

The analysis of the equations obtained by the minimization of energy (12) shows that the location of point *R* on the curve *OD*, the phase $FCM_{[001]}$, and the curve of the orientational second-order phase transition *RX* on the phase diagram strongly depend on the value of the parameters D_2 and N_2. In the case of the phase diagram in Fig. 4, we have $D_2 \gg N_2$; thus, their location in the phase diagram is mainly determined by the interaction of the modulation and strain order parameters D_2. It turns out that, depending on the magnitude of this interaction (and, therefore, on the magnitude of the tetragonal distortions, which are determined by this interaction in the modulated phases), the phase $FCM_{[001]}$ can either be present on the phase diagram or not. Correspondingly, the martensitic transition on the curve *RD* can be either accompanied by the magnetic orientational phase transition or not. Figure 4 presents a situation when the parameter D_2 is greater than a certain critical value at which the phase $FCM_{[001]}$ still exists. If D_2 is less than this critical value, then the phase $FCM_{[001]}$ and the transition on the curve *RX* on the phase diagram do not exist, since the tetragonal distortions in the premartensitic phase are not large enough. In this case, the martensitic first-order phase transition from the angular quasi-cubic phase $FCM_{[uuw]}$ to the symmetric tetragonal phase $FTM_{[001]}$ with large tetragonal distortions occurs on the curve *OD*. This phase transition is accompanied by the magnetic orientational first-order phase transition at which a reorientation of the magnetization from the direction [*uuw*] to the direction [001] occurs. Depending on the parameter D_2, the curve *QX* changes its location on the diagram between the curves *QL* and *QF*. As D_2 increases, *QX* moves in the direction of *QF* (clockwise); however, even for very large values of D_2, these curves do not coincide; they rather converge asymptotically. The situation is different when D_2 decreases. In this case, the curve *QX* approaches the curve *QL* (anticlockwise), and they merge at a certain value of D_2. Before they merge, the orientational second-order phase transition on the curve *RX* occurs; on the curve *RD*, the martensitic transition occurs, which is accompanied by jumps of the magnetization modulus and tetragonal strains, but without reorientation of the magnetization. When the curves *QX* and *QL* coincide, the martensitic transition on the curve *OD* is accompanied by simultaneous jumps of the orientation, magnetization modulus, and tetragonal strains. Thus, depending on the parameters of the problem, a structural first-order phase transition occurs on the curve *OD* either from the phase $FCM_{[uuw]}$ or the phase $FCM_{[001]}$ to the phase $FTM_{[001]}$. The phase $FTM_{[001]}$ has both large tetragonal distortions and modulation order. In Fig. 4, the phase $FCM_{[uuw]}$ is stable above the curve *QX*, the phase $FCM_{[001]}$ is stable above *QL* and below *QX*, and the phase $FTM_{[001]}$ is stable below *KM*. At last, the curve *OE* corresponds to the structural first-order phase transition between the martensitic tetragonal phase $FT_{[001]}$ and the modulated tetragonal phase $FTM_{[001]}$. This transition is not accompanied by an orientational magnetic phase transition. The phase $FT_{[001]}$ is stable to the right of the curve *NM*, and the phase $FTM_{[001]}$ is stable to the left of *PM*. It must be noted that there exist equal energy states $FT_{[001]}$, $FT_{[010]}$, $FT_{[100]}$, and so on in the

phase *FT*. All these solutions, except for $FT_{[001]}$, which is connected with the transition to the modulated state $FTM_{[001]}$, are stable to the right of the curve *MZ* and become metastable on the curve *OE*. Thus, the degeneracy of solutions in the modulated phase due to the crystal symmetry is removed.

For $b > 0$ and $D_2 < 0$ in the angular modulated phase on the curve *RX*, the magnetization vector is reoriented to the [110] axis rather than to [001].

It is seen from Fig. 4 that when the modulation order parameter is taken into account, the martensitic transformation is accompanied either by a premartensitic (with the thermodynamic path $\sigma\sigma'$) or a postmartensitic (with the thermodynamic path $\rho\rho'$) phase transition. Thus, the variation of the parameter D_2 determines the properties of the martensitic transformation.

In addition to structural and magnetic orientational phase transitions, the theory suggested in this paper allows for a description of the magnetic phase transition of the type disorder–order (the Curie point). Figure 5 presents the phase diagram on the coordinates $\alpha - A$ for the case when the $\alpha - a$ diagram (for $\psi = 0$, see [5, 9, 11, 12]) contains a second-order phase transition from the paramagnetic cubic phase to the ferromagnetic cubic phase.

It is seen that, in this case, seven equilibrium states can exist. *PC* is a paramagnetic cubic phase. On the curve *OB*, a structural first-order phase transition to the modulated quasi-cubic paramagnetic phase *PCM* occurs, in which small tetragonal distortions appear due to the interaction of the modulation and strain order parameters. Then, on the curve *ML*, a martensitic first-order transition from the phase *PCM* to the tetragonal modulated phase *PTM* with large tetragonal strains occurs. *OM* is the curve of a magnetic second-order phase transition between the modulated quasi-cubic paramagnetic (*PCM*) and ferromagnetic ($FCM_{[001]}$) phases. On the curve *MQ*, a similar transition between the tetragonal modulated phases *PTM* and $FTM_{[001]}$ occurs. *ON* is the curve of an orientational second-order phase transition between the quasi-cubic modulated symmetric ($FCM_{[001]}$) and angular ($FCM_{[uuw]}$) phases. On the curves *MN* and *NP*, a martensitic phase transition from these phases to the tetragonal ferromagnetic phase $FTM_{[001]}$ occurs. *OD* is the curve of structural and orientational phase transitions from the ferromagnetic cubic phase $FC_{[111]}$ to the modulated quasi-cubic angular phase $FCM_{[uuw]}$. Finally, a magnetic second-order phase transition between the paramagnetic, *PC*, and ferromagnetic, *FC*, cubic phases occurs on the curve *OC*. The stability domain of the phase *PC* is bounded by the curves *OC* and *OI*; of the phase *FC* by *OC* and *OH*; of the phase *PCM* by *OK*, *OQ*, and *QR*; of the phase *PTM* by *OF* and *OU*; of the phase $FTM_{[uuw]}$ by *OE*, *ES*, and *OG*; of the phase $FCM_{[001]}$ by *OQ*, *QE*, and *EO*; and of the phase $FTM_{[001]}$ by *OF* and *OT*.

We note that, for the chosen value of the parameter *a*, only premartensitic phase transitions to modulated phases occur in the $\alpha - A$ diagram. However, for other values of *a*, postmartensitic phase transitions emerge in the $\alpha - A$ diagram.

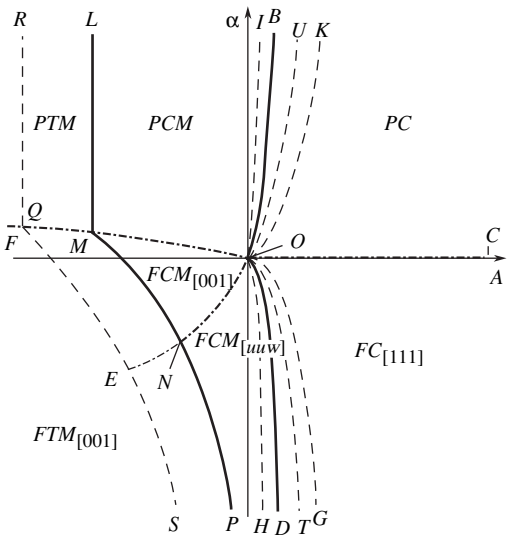


Fig. 5. The phase diagram of the Ni–Mn–Ga alloy on the coordinates $\alpha - A$ for $b > 0$, $D_2 > 0$, and $N_{1,2,3} < 0$. Solid curves are the curves of first-order phase transitions, dot-and-dash curves mark second-order phase transitions, and dashed curves are the boundaries of stability regions of the phases; *P* denotes the paramagnetic phase, *F* the ferromagnetic phase, *M* the phase with modulation, and *C* the phase with a cubic lattice. The direction of the magnetization vector of the phases is indicated in square brackets.

REFERENCES

1. S. J. Murray, M. Marioni, S. M. Allen, *et al.*, Appl. Phys. Lett. **77**, 886 (2000).
2. V. S. Boĭko, R. I. Garber, and A. M. Kosevich, *Reversible Crystal Plasticity* (Nauka, Moscow, 1991).
3. M. F. Fradkin, Phys. Rev. B **50**, 16326 (1994).
4. A. N. Vasil'ev, V. V. Kokorin, Yu. I. Savchenko, and V. A. Chernenko, Zh. Éksp. Teor. Fiz. **98**, 1437 (1990) [Sov. Phys. JETP **71**, 803 (1990)].
5. A. D. Bozhko, A. N. Vasil'ev, V. V. Khovaĭlo, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 212 (1998) [JETP Lett. **67**, 227 (1998)].
6. V. D. Buchel'nikov, A. N. Vasil'ev, I. E. Dikshteĭn, *et al.*, Fiz. Met. Metalloved. **85**, 5 (1998).
7. V. D. Buchel'nikov, A. N. Vasil'ev, I. E. Dikshteĭn, and V. G. Shavrov, Fiz. Met. Metalloved. **85**, 54 (1998).
8. V. D. Buchel'nikov, A. N. Vasil'ev, A. T. Zayak, *et al.*, Vestn. Chelyabinsk. Gos. Univ., Ser. 6: Fiz. **1**, 20 (1998).
9. V. D. Buchel'nikov, A. N. Vasil'ev, I. E. Dikshteĭn, and V. G. Shavrov, Vestn. Chelyabinsk. Gos. Univ., Ser. 6: Fiz. **1**, 5 (1998).
10. V. D. Buchelnikov, V. S. Romanov, and A. T. Zayak, J. Magn. Magn. Mater. **191** (1–2), 203 (1999).
11. A. N. Vasil'ev, A. D. Bozhko, V. V. Khovailo, *et al.*, Phys. Rev. B **59**, 1113 (1999).

12. A. D. Bozhko, A. N. Vasil'ev, V. V. Khovaĭlo, *et al.*, Zh. Éksp. Teor. Fiz. **115**, 1740 (1999) [JETP **88**, 954 (1999)].

13. A. Zheludev, S. M. Shapiro, and P. Wochner, Phys. Rev. B **54**, 15045 (1996).

14. J. A. Krumhansl and R. J. Gooding, Phys. Rev. B **39**, 3047 (1989).

15. A. Planes, E. Obrado, A. Gonzales-Comas, and L. Manosa, Phys. Rev. Lett. **79**, 3926 (1997).

16. R. J. Gooding and J. A. Krumhansl, Phys. Rev. B **38**, 1695 (1988).

17. P. J. Webster, K. P. A. Ziebeck, S. L. Town, and M. S. Peak, Philos. Mag. **49**, 295 (1984).

18. V. V. Kokorin, V. V. Martynov, and V. A. Chernenko, Fiz. Tverd. Tela (Leningrad) **33**, 1250 (1991) [Sov. Phys. Solid State **33**, 708 (1991)].

19. A. Zheludev, S. M. Shapiro, P. Wochner, *et al.*, Phys. Rev. B **51**, 11310 (1995).

20. F. Zuo, X. Su, and K. H. Wu, Phys. Rev. B **58**, 11127 (1998).

21. V. A. L'vov, E. V. Gomonaj, and V. A. Chernenko, J. Phys.: Condens. Matter **10**, 4587 (1998).

22. T. Castan, E. Vives, and P.-A. Lindgard, Phys. Rev. B **60**, 7071 (1999).

23. I. A. Folkins and M. B. Walker, Phys. Rev. B **40**, 255 (1989).

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 1019–1023.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1176–1181.

SOLIDS
Structure

The Influence of Magnetoelastic Interaction on Structural Phase Transitions in Cubic Ferromagnetics

V. D. Buchel'nikov[a, *], A. N. Vasil'ev[b], A. T. Zayak[a, c], and P. Entel[c, **]

[a]*Chelyabinsk State University, Chelyabinsk, 454021 Russia*
[b]*Moscow State University, Vorob'evy gory, Moscow, 119899 Russia*
[c]*Theoretical Low-Temperature Physics Gerhard–Mercator University, 47048, Duisburg, Germany*
**e-mail: buche@cgu.chel.su*
***e-mail: entel@thp.uni-duisburg.de*

Received December 7, 2000

Abstract—In the framework of the Landau theory of phase transitions, the influence of the magnetoelastic interaction on structural transitions in cubic ferromagnetics with a positive first magnetic anisotropy constant is analyzed. It is shown that structural transitions are not accompanied by a reorientation of magnetization in this case. The phase diagrams of such ferromagnetics either contain a termination point of the structural transition or a critical point in which the first-order transition is replaced by a second-order one. Magnetoelastic interaction also leads to the appearance of an interval of the ferromagnetic parameters in which a coupled first-order structural–magnetic transition exists. The phase T–x diagram for Heusler $Ni_{2+x}Mn_{1-x}Ga$ alloys is calculated, which is in good agreement with the experimental phase diagram of these alloys. © *2001 MAIK "Nauka/Interperiodica".*

Magnetoelastic interaction in ferromagnetics makes it possible to initiate structural transitions using a magnetic field. From this viewpoint, of special interest are martensitic transformations in ferromagnetics, which possess the properties of shape memory and huge magnetostriction. For example, such are Heusler $Ni_{2+x}Mn_{1-x}Ga$ alloys, in which a transition from the cubic to the tetragonal phase occurs as the temperature decreases [1–3]. The influence of magnetostriction on structural transitions in cubic ferromagnetics with a negative first magnetic anisotropy constant when the magnetization in the initial phase is oriented along the [111] axis was investigated in [4, 5]. In this case, structural phase transitions are accompanied by the appearance of angular magnetic phases and a reorientation of the magnetization. Experimental results that have been recently obtained in [6] imply that, in the stoichiometric Ni_2MnGa, the first anisotropy constant is positive; i.e., the magnetization in the initial phase is oriented along the [100] axis. In the present paper, we investigate phase diagrams for this case.

In order to construct phase diagrams of a cubic ferromagnetic, we write out an expression for the free energy that involves only the components of the ordering parameters responsible for the structural and magnetic transitions, which is usually used in the Landau theory [3–5, 7]:

$$F = \frac{1}{2}a(e_2^3 + e_3^2) + \frac{1}{3}be_3(e_3 - 3e_2^2) + \frac{1}{4}c(e_2^2 + e_3^2)$$
$$+ B\left[\frac{1}{\sqrt{2}}e_2(m_x^2 - m_y^2) + \frac{1}{\sqrt{6}}e_3(3m_z^2 - \mathbf{m}^2)\right] \quad (1)$$
$$+ K(m_x^2m_y^2 + m_y^2m_z^2 + m_z^2m_x^2) + \frac{1}{2}\alpha\mathbf{m}^2 + \frac{1}{4}\delta\mathbf{m}^4.$$

Here e_i are linear combinations of the components of the strain tensor, $e_2 = (e_{xx} - e_{yy})/\sqrt{2}$ and $e_3 = (2e_{zz} - e_{yy} - e_{xx})/\sqrt{6}$; a, b, and c are linear combinations of the modulus of elasticity of the second, third, and fourth order, $a = c_{11} - c_{12}$, $b = (c_{111} - 3c_{112} + 2c_{123})$, $c = (c_{1111} + 6c_{1112} - 3c_{1112} - 8c_{1123})$; $\mathbf{m} = \mathbf{M}/M_0$ is the dimensionless unit magnetization vector; M_0 is the saturation magnetization; B is the magnetostriction constant; K is the first cubic anisotropy constant; and α and δ are the exchange constants. In the vicinity of the Curie temperature $T = T_C$, the exchange constant can be written in the form

$$\alpha = \alpha_0(T - T_C)/T_C.$$

In the vicinity of the point of the structural phase transition $T = T_M$, the generalized elastic modulus of the second order is written in the form

$$a = a_0(T - T_M)/T_M.$$

In alloys of the $Ni_{2+x}Mn_{1-x}Ga$ type, the temperature of the structural and phase transitions substantially depends on the composition. In this connection, it is inter-

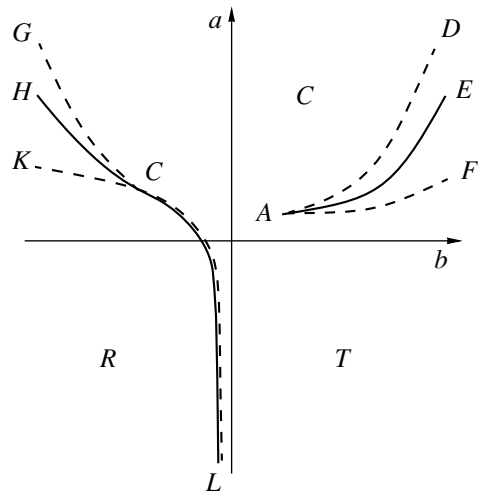


Fig. 1. The phase diagram of a cubic ferromagnetic for $T_M < T_C$ on the coordinates a, b (C is the cubic phase with small tetragonal distortions, T is the tetragonal phase, and R is the orthorhombic phase). For all phases, the magnetization is oriented along the [001] axis. Solid curves correspond to phase transitions, and dashed ones are the curves of the loss of stability of phases.

esting to investigate phase diagrams both for $T_M < T_C$ (compounds close to stoichiometry) and for $T_M \sim T_C$ (compounds with $x = 0.16$–0.20). For definiteness, we assume that the magnetostriction constant $B > 0$, and the elastic modulus of the fourth order $c > 0$.

First, consider the case when $T_M < T_C$, i.e., when the structural transition occurs in the ferromagnetic matrix. Then, we can set $\mathbf{m}^2 = 1$ in (1) and change from the Cartesian components of the magnetization m_x, m_y, and m_z to the polar (θ) and azimuth (φ) angles of this vector. Minimization of the free energy over e_2, e_3, θ, and φ leads to the following equilibrium states.

1. The cubic, C, and tetragonal, T, phases with the magnetization along the [001] axis and strains determined by the equations

$$e_2 = 0, \quad ae_3 + be_2^2 + ce_3^3 + \frac{\sqrt{6}}{9}B = 0 \qquad (2)$$

are stable for $b \le 0$ in the region described by the inequality

$$a \ge \frac{\sqrt{6}}{6}\frac{Bc}{b} + \frac{2}{3}\sqrt{-bB\sqrt{6}}. \qquad (3)$$

For $b \ge 0$, the regions of existence of these phases are separated by the branches of the discriminant curve

$$a^3 - \frac{b^2}{4c}a^2 - \frac{3\sqrt{6}}{2}abB + \frac{9}{2}cB^2 + \frac{\sqrt{6}}{3}\frac{b^3}{3}B = 0 \qquad (4)$$

of the cubic Eq. (2). Inside the region bounded by the branches of curve (4), both phases are stable.

2. The orthorhombic phase R with the magnetization along the [001] axis ($\theta = 0$) and strains determined by the equations

$$4cbe_3^2 - 2b^2e_3 + ab + \frac{\sqrt{6}}{3}Bc = 0,$$
$$e_2 = \pm\sqrt{-e_3^2 - \frac{a}{c} + e_3\frac{2b}{c}} \qquad (5)$$

is stable for

$$a \le \frac{b^2}{4c} - \frac{\sqrt{6}}{3}\frac{Bc}{b}, \quad b < -\left(\frac{16\sqrt{6}}{9}Bc^2\right)^{1/3}, \qquad (6a)$$

and for

$$a \le \frac{\sqrt{6}}{9}\frac{Bc}{b} + \frac{2}{3}\sqrt{-\sqrt{6}bB}, \qquad (6b)$$
$$-\left(\frac{16\sqrt{6}}{9}Bc^2\right)^{1/3} < b \le 0.$$

From symmetry considerations, it follows that in addition to the states considered above, ferromagnetics can exhibit similar phases with the magnetization along the [100] and [010] axes.

The analysis of distortions of the initial cubic lattice in the phases C and T shows that these phases possess the same tetragonal symmetry and differ only in the magnitude of spontaneous strains. In the phase C, these strains are determined by the strains of the cubic lattice due to magnetostriction; in the phase T, they are determined by structural strains when passing to the martensitic state. The curves of phase transitions between the states C, T, and R are determined from the condition of equality of the phase energy.

The phase diagram of the cubic ferromagnetic on the coordinates a, b for $T_M < T_C$ is presented in Fig. 1. Depending on the magnitude of the elastic moduli of the second (a) and third (b) order, the ferromagnetic can exhibit the following structural transformations. For $b > 0$, the first-order phase transition from the phase C to the phase T occurs on the curve AE defined by the equation

$$a = \frac{2}{9}\frac{b^2}{c} + \frac{\sqrt{6}Bc}{b}. \qquad (7)$$

This transition is accompanied by a jump of the strains e_3 and is a martensitic transformation. From the symmetry point of view, it is an isostructural transition, which has a termination point A with the coordinates $(18cB^2)^{1/3}$, $(9\sqrt{6}\,Bc^2)^{1/3}$. To the left of the point A, the transition from the phase C to T is smooth, without any jump in the strain e_3. For $b \le 0$, the martensitic transformation (the first-order structural phase transition) on the curve CH occurs from the cubic phase C to the orthorhombic one R, which is accompanied by a jump of the strains e_2 and e_3. On the curve CL, the second-

order structural phase transition between these two phases occurs. An expression for the curve of the first-order phase transition (*CH*) can be found from the condition of equality of the energy for the phases *C* and *R*. The equation for the curve of the second-order phase transition (*CL*) is determined by the equality in (3). The critical point of the termination of the first-order phase transition (*C*) has the coordinates $(\sqrt{6}/3 - 1/2)(B^2c)^{1/3}$, $-(16\sqrt{6}\,Bc^2/9)^{1/3}$. The coordinates of the points *A* and *C* are determined by the magnitude of the magnetostriction *B*. For $B = 0$, the phase diagram coincides with the diagram of the nonmagnetic cubic crystal [8].

Now consider the phase diagram of a cubic ferromagnetic for the case when the temperature of the martensitic (T_M) and magnetic (T_C) transitions become comparable. In this case, in order to find equilibrium states of the cubic ferromagnetic, the free energy (1) must be minimized over the variables m_x, m_y, m_z, e_2, and e_3. For definiteness, we assume that $b > 0$ and the signs of the other constants remain the same as in the case $T_M < T_C$.

The minimization of (1) yields the following equilibrium states of the ferromagnetic.

1. The cubic paramagnetic phase (*PC*),

$$m_x = m_y = m_z = 0, \quad e_2 = e_3 = 0, \tag{8}$$

is stable for $\alpha \geq 0$, $a \geq 0$.

2. The tetragonal paramagnetic phase (*PT*),

$$m_x = m_y = m_z = 0, \quad e_2 = 0,$$
$$e_3 = -\frac{b + \sqrt{b^2 - 4ac}}{2c}, \tag{9}$$

is stable for

$$\alpha \geq \frac{\sqrt{6}Bb}{3c}, \quad a \leq \frac{b^2}{4c},$$
$$a \geq \frac{b^2}{4c} - \left(\frac{\sqrt{6}}{4}\frac{\alpha}{B}\sqrt{c} - \frac{b}{2\sqrt{c}}\right)^2. \tag{10}$$

3. The cubic ferromagnetic (*FC*) and tetragonal ferromagnetic (*FT*) phases with the magnetization along [001],

$$m_x = m_y = 0, \quad m_z^2 = -\frac{1}{\delta}\left(\alpha + \frac{2\sqrt{6}}{3}Be_3\right), \tag{11}$$

with strains determined by the equations

$$e_2 = 0, \quad ae_3 + be_3^2 + c3e_3^3 + \sqrt{6}Bm_z^2/3 = 0. \tag{12}$$

The phase *FC* is stable for

$$\alpha \leq 0, \quad \alpha \geq \frac{\sqrt{6}b}{54Bc^2}(2\delta b^2 + 12cB^2 - 9\delta ac)$$
$$- \frac{\sqrt{6}}{27Bc^2}(\delta b^2 + 4cB^2 - 3\delta ac)^{3/2}, \tag{13}$$

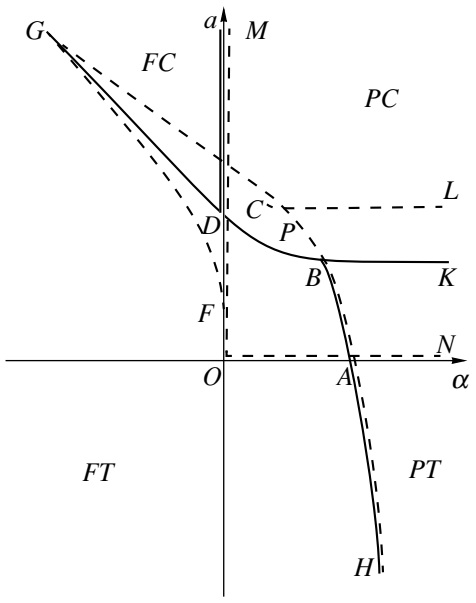


Fig. 2. The phase diagram of a cubic ferromagnetic for $T_M \sim T_C$ on the coordinates *a*, α (*PC* is the paramagnetic cubic phase, *PT* is the paramagnetic tetragonal phase, *FC* is the ferromagnetic cubic phase with small tetragonal distortions, and *FT* is the ferromagnetic tetragonal phase). In the *FC* and *FT* phases, magnetization is oriented along the [001] axis. Solid curves correspond to phase transitions, and dashed ones are the curves of the loss of stability of phases.

and the phase *FT* is stable for

$$\alpha \leq \frac{\sqrt{6}}{54Bc^2}(2\delta b^2 + 12cB^2 - 9\delta ac)$$
$$+ \frac{\sqrt{6}}{27Bc^2}(\delta b^2 + 4cB^2 - 3\delta ac)^{3/2}, \tag{14}$$
$$a \leq \frac{b^2}{4c} - \left(\frac{\sqrt{6}}{4}\frac{\alpha\sqrt{c}}{B} - \frac{b}{2\sqrt{c}}\right)^2 \quad (\alpha \geq 0).$$

The stability region of the phases *FC* and *FT* is also bounded by the inequalities

$$\alpha \geq -\delta,$$
$$a \geq -\frac{3c(\alpha + \delta)^2}{8B^2} + \frac{\sqrt{6}b(\alpha + \delta)}{4B} + \frac{4B^2}{3(\alpha + \delta)}, \tag{15}$$

which follow from the condition $m_z^2 \leq 1$.

From symmetry considerations, it follows that in addition to these states, the crystal can exhibit other equilibrium phases with the energies and regions of stability coinciding with those described above. These are tetragonal paramagnetic phases with strains along the axes [100] and [010], cubic ferromagnetic phases with magnetization along the axes [100] and [010], and tet-

ragonal phases with strains and magnetization along the axes [100] and [010].

As in the case $T_M < T_C$, the analysis of strains of the cubic lattice in the phases *FC* and *FT* determined by Eqs. (12) shows that these phases possess identical tetragonal symmetry. The phases *FC* and *FT* differ in the magnitude of spontaneous strains e_3. In the phase *FC*, they are mainly determined by the magnetostriction, whereas in the phase *FT* by structural distortions that occur while passing to the martensitic state.

The phase diagram of the cubic crystal for the case $T_M \sim T_C$ on the coordinates a, α is presented in Fig. 2. The following phase transitions are possible from the paramagnetic cubic phase *PC*. On the curve *BK* determined by the equation $a = b^2/4c$, the first-order structural phase transition to the tetragonal paramagnetic phase *PT* with large distortions of the lattice occurs (martensitic transformation). On the curve *DM* ($\alpha = 0$), the second-order phase transition to the ferromagnetic cubic phase *FC* with small tetragonal distortions of the lattice occurs. On the curve *DB*, the coupled structural–magnetic first-order phase transition to the ferromagnetic cubic phase *FT* with large tetragonal distortions of the lattice occurs. The equation of this curve follows from the equality of the energy of the phases *PC* and *FT*. In addition to the transition $PT \longrightarrow PC$ along *BK*, the second-order isostructural magnetic phase transition from the paramagnetic tetragonal phase *PT* to the ferromagnetic tetragonal phase *FT* can occur along the curve *BH*. The equation of this curve follows from the second condition of stability of the phase *FT* in (14) when the inequality turns into equality. The first-order isostructural phase transition between the ferromagnetic phases *FC* and *FT* can occur along the curve *GD*. The equation of this curve has the form

$$a = \frac{2}{9}\frac{b^2}{c} + \frac{4}{3}\frac{B^2}{\delta} - \frac{\sqrt{6}Bc\alpha}{6\delta}. \quad (16)$$

This transition is accompanied by a jump of the strain e_3 and is classified as a martensitic transformation. In the a–α diagram, it can have the termination point *G*. This situation occurs in the case when the point *G* is to the right of the stability curve of the phases *FC* and *FT* in (15) (for simplicity, this curve is not shown in the phase diagram). The analysis of inequalities (15) and Eq. (16) in combination with conditions (13) and (14) of the stability of the phases *FC* and *FT* shows that the termination point of the phase transition $FT \longrightarrow FC$ exists only for large values of the magnetoelastic constant $B \sim b^3/c^2$. In this case, the transition between the phases *FT* and *FC* to the left of the point *G* proceeds smoothly, without jumps of the strain e_3.

The region of absolute stability of the phase *PC* is bounded by the lines *OM* and *ON*. For the phase *PT*, this region is bounded by the curves *LC* and *CH*. The phase *FT* is absolutely stable in the region to the left of the curve *GPH*, and the phase *FC* above the curve *GFM*. The points *D* and *B* are critical. At these points, the second-order phase transition curves split into two curves of first-order phase transitions. The coordinates of these points are

$$D\left(\frac{2}{9}\frac{b^2}{c} + \frac{4}{3}\frac{B^2}{\delta}, 0\right), \quad B\left(\frac{2}{9}\frac{b^2}{c}, \frac{4\sqrt{6}}{9}\frac{Bb}{c}\right).$$

In experimental studies of phase transitions in alloys of the $Ni_{2+x}Mn_{1-x}Ga$ type, the phase diagram is usually reconstructed on the temperature–concentration (T–x) coordinates [3, 5]. In order to be able to compare the theoretical results obtained in the present study with the available experimental data, we present the theoretical phase T–x diagram of the cubic ferromagnetic.

It follows from the experimental data [3, 5] that the Curie temperature T_C and the martensitic transformation temperature T_M for $Ni_{2+x}Mn_{1-x}Ga$ alloys vary almost linearly with the concentration. We use this result to construct the theoretical a–α phase diagram (Fig. 2) on the coordinates T–x. For this purpose, we write the coefficients a and α in the equation of the free energy (1) in the form

$$a = a_0(T - T_M)/T_M, \quad \alpha = \alpha_0(T - T_C)/T_C, \quad (17)$$

where $T_M = T_{M0} + \sigma x$, $T_C = T_{C0} - \gamma x$, T_{M0} and T_{C0} are the temperatures of the martensitic and magnetic transition at $x = 0$, and σ and γ are proportionality coefficients. T_{M0}, T_{C0}, σ, and γ are determined experimentally.

In order to calculate the phase T–x diagram, we use the following values of the parameters involved in (1) and (17) (see [3, 5, 6]):

$$a_0 = 10^{11}\ \text{erg/cm}^3, \quad \alpha_0 = -10^9\ \text{erg/cm}^3,$$
$$T_{M0} = 202\ \text{K}, \quad T_{C0} = 375\ \text{K}, \quad \sigma = 700\ \text{K},$$
$$\gamma = 175\ \text{K}, \quad b = 3 \times 10^{11}\ \text{erg/cm}^3, \quad (18)$$
$$c = 3 \times 10^{12}\ \text{erg/cm}^3, \quad B = 1.5 \times 10^7\ \text{erg/cm}^3,$$
$$K = 4 \times 10^4\ \text{erg/cm}^3, \quad \delta = 10^9\ \text{erg/cm}^3.$$

Figure 3 presents the T–x phase diagram of $Ni_{2+x}Mn_{1-x}Ga$ alloys constructed numerically on the basis of Eqs. (17) and (18). Since the magnetoelastic constant B is small, the region in the vicinity of the intersection of the temperatures T_M and T_C cannot be shown on the scale of Fig. 3a. This region is shown in Fig. 3b on a greater scale. The designations in Fig. 3 are the same as in Fig. 2. It is seen from Fig. 3 that, for the given values of the parameters, the region *DB* in which the first-order coupled structural–magnetic phase transition can occur is confined within a very small interval of concentrations around $x \approx 0.19$. This interval strongly depends on the magnitude of the magnetoelastic constant B. As B increases, the interval in which the coupled structural–magnetic phase transition exists

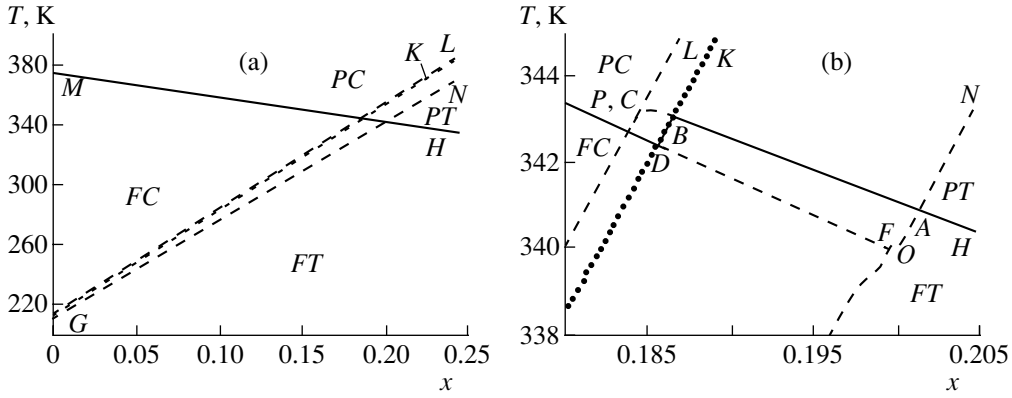


Fig. 3. The theoretical T–x phase diagram of cubic ferromagnetic Heusler $Ni_{2+x}Mn_{1-x}Ga$ alloys: (a) for $0 \le x \le 0.25$ and (b) in the region of the intersection of the temperatures T_M and T_C on a greater scale. The solid curve corresponds to the magnetic phase transition, dot curves correspond to the structural transition, and the dashed curves are the curves of the loss of stability of phases. The other designations coincide with those used in Fig. 2.

also increases. The phase diagram presented in Fig. 3 is in good agreement with the experimental T–x diagram (see Figs. 5 and 4 in [3, 5], respectively).

The analysis of the influence of the magnetoelastic interaction on the phase diagrams of cubic ferromagnetics allows one to draw the following conclusions.

In contrast to the case of the negative cubic anisotropy constant ($K < 0$), the structural phase transitions for a positive first anisotropy constant ($K > 0$) are not accompanied by reorientation of the magnetization. This is due to the fact that already in the cubic phase the magnetoelastic interaction decreases (strictly speaking) the symmetry of the phase down to the tetragonal one. The symmetry of the low-temperature phase is either tetragonal ($b > 0$) or orthorhombic ($b < 0$). Due to the fact that the high-temperature and low-temperature phases contain the same symmetry elements (e.g., axes of the second and fourth orders), no reorientation of magnetization occurs in the case $K > 0$ under structural transitions. When the symmetries of the high-temperature and low-temperature phases are identical ($b > 0$), the curve of the structural phase transition can have a termination point. To the right of this curve, the structural transition is accompanied by a jump of strains and hysteresis, and behaves like a martensitic transformation. To the left of this point, no such transition occurs, the strains vary smoothly from quasi-cubic to tetragonal ones, and there is no hysteresis. For $b < 0$, the symmetries of the high-temperature and low-temperature phases are not identical. In this case, for large $|b|$, the structural transition between phases is a first-order transition (a martensitic transformation); for small $|b|$, it is a second-order transition. Thus, there is a critical point on the phase diagram for $b < 0$; at this point, the transition type changes. The coordinates of the termination point of the structural transition (for $b > 0$) and the critical point (for $b < 0$) are determined by the magnetostriction constant B. The magnetoelastic interaction leads to the situation when first-order coupled structural–magnetic phase transitions occur in a certain interval of the parameters of the cubic ferromagnetic. The magnitude of the interval on the phase diagram in which such transitions occur is determined by the magnitude of the magnetoelastic interaction. The calculated theoretical T–x diagram of the Heusler alloys $Ni_{2+x}Mn_{1-x}Ga$ is in good agreement with the phase diagram obtained experimentally.

REFERENCES

1. P. J. Webster, K. R. A. Ziebeck, S. L. Town, and M. S. Peak, Philos. Mag. B **49**, 295 (1984).
2. D. S. J. Murray, M. Marioni, S. M. Allen, *et al.*, Appl. Phys. Lett. **77**, 886 (2000).
3. A. N. Vasil'ev, A. D. Bozhko, V. V. Khovaĭlo, *et al.*, Phys. Rev. B **59**, 1113 (1999).
4. V. D. Buchelnikov, V. S. Romanov, and A. T. Zayak, J. Magn. Magn. Mater. **191**, 203 (1999).
5. A. D. Bozhko, A. N. Vasil'ev, V. V. Khovaĭlo, *et al.*, Zh. Éksp. Teor. Fiz. **115**, 1740 (1999) [JETP **88**, 954 (1999)].
6. R. Tickle and R. D. James, J. Magn. Magn. Mater. **195**, 627 (1999).
7. M. A. Fradkin, Phys. Rev. B **50**, 16326 (1994).
8. Yu. A. Izyumov and V. N. Syromyatnikov, *Phase Transitions and Symmetry of Crystals* (Nauka, Moscow, 1984).

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 1024–1034.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1182–1194.

SOLIDS
Structure

The Kinetics of the Anomalous State Formation and the Growth of Helium Crystals at High Supersaturation

V. L. Tsymbalenko*
Kurchatov Institute Russian Scientific Center, Moscow, 123182 Russia
**e-mail: vlt@isssph.kiae.ru*
Received November 20, 2000

Abstract—The kinetics of the formation of an anomalous state of a helium crystal with a fast-growing surface are investigated. It is demonstrated experimentally that it is external supersaturation that is the determining factor of formation of an anomalous state. The dependence of the time of formation of anomalous state on temperature and initial supersaturation is measured. The problem of crystal growth with the excitation of the first-sound wave in the container is solved. This solution is used to determine the dependence of the kinetic coefficient of growth of anomalous facets on temperature and initial supersaturation. It is demonstrated that the kinetic coefficient of facet growth decreases on approaching the boundaries of the region of the existence of an anomalous state. The kinetic coefficient of growth of atomically rough surfaces in an anomalous state is determined by the damping of pressure oscillation. It is found that the value of the latter coefficient is three–four times that of the respective value for the facets but is considerably, by an order of magnitude, less than the value of the coefficient of growth of such surfaces in the normal state. Phenomena are treated which accompany the spiral growth of facets, namely, the excitation of oscillations of a screw dislocation during spiral rotation and the emergence of vortex rings in superfluid helium. The effect of these phenomena on the kinetics of facet growth and on the formation of an anomalous state is discussed. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The kinetic and thermodynamic properties of the interface between liquid helium and its solid HCP phase were the subject of intensive investigations following the theoretical prediction by Andreev and Parshin [1] and experimental verification by Keshishev *et al.* [2] of the quantum behavior of facet growth in the atomically rough state. By now, a fairly comprehensive picture has been formed of thermodynamic transitions on the surface and of the kinetics of facet growth. Three roughening transitions have been revealed: at $T_{R1} = 1.3$ K, the faceting of the basal plane (*c*-facets) is observed; then, at $T_{R2} = 0.9$ K, the side planes (*a*-facets) are faceted [3–5]; and, at the temperature $T_{R3} = 0.36$ K, the faceting of $(1\bar{1}01)$-planes (*s*-facets) is observed [6]. The kinetics of crystal growth under conditions of minor deviations from equilibrium ($Dp < 1$ mbar) are defined by the state of the surface. Above the roughening transition in the atomically rough state, the rate of growth is controlled by the scattering of quasiparticles on the surface [7, 8] and by the dissipation in the medium [9]. Below the roughening transition, in accordance with the models of growth existing in classical crystallography [10], the facet growth is associated either with two-dimensional nucleation [10, 11] or with topologically irremovable defects formed on the facet by dislocations (spiral growth, Frank-Read sources [10]). These models (with necessary correction for the case of helium [12]) describe well most of the observed phenomena. The only effect that drops out from the general pattern was the burst-like growth of a dislocation-free crystal facet in the temperature range from 2 mK to 0.25 K, which was observed by Ruutu *et al.* [12].

Another phenomenon, which does not fit the framework of the known concepts either, consists in a sharp acceleration of crystal growth by two–three orders of magnitude (anomalous growth) under conditions of fairly high supersaturation below the second roughening transition [13–16].

The investigation results revealed a number of features of this transition: namely, (1) the acceleration of growth occurs on the *c*- and *a*-facets simultaneously [14]; (2) the crystal changes over to the state with fast-growing facets only in the case of initial supersaturation above a certain value of Dp^*, which decreases monotonically with temperature (phase diagram [15, 17]; and (3) away from the transition boundary, the rate of crystal growth becomes so high that the growth becomes oscillating, with the damping decrement of oscillation decreasing away from the transition boundary [16].

In this paper, we give the results of studies into the effect of the condition of crystal formation on the transition to anomalous state with fast-growing facets, as well as the results of analysis of the kinetics of growth after transition. Section 2 contains a brief description of the experimental procedure. The results of checking the hypothesis which relates the transition to the behavior of crystal nucleation are described in Section 3. The experimental data on the time of emergence of an

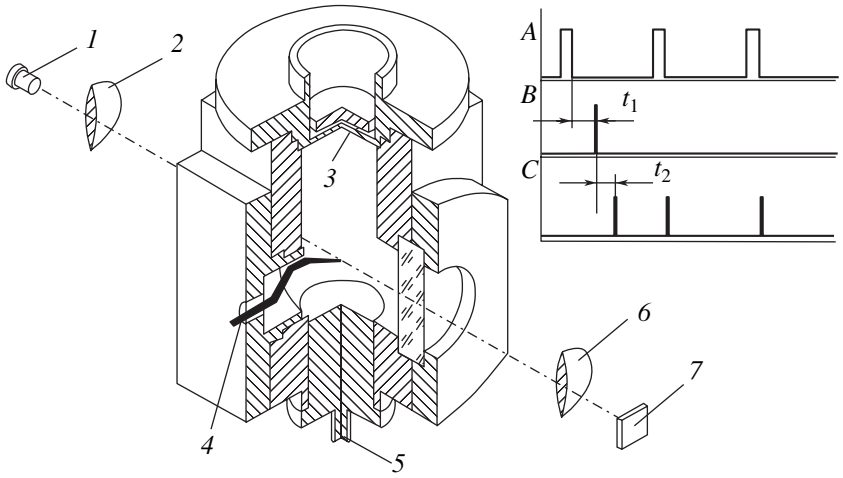


Fig. 1. Experimental scheme: (*1*) pulsed infrared light-emitting diode, (*2*) lens, (*3*) capacitive sensor diaphragm, (*4*) tungsten needle, (*5*) copper refrigerant conduit connected to a ^{3}He bath, (*6*) objective, (*7*) CCD matrix. The optical axis is arranged horizontally. The details of the optical cryostat [19] are not shown. Time diagrams: *A*, frame pulses of the video signal; *B*, high-voltage pulse; *C*, LED flashes.

anomalous state are given in Section 4. The results of analysis of the pattern of crystal growth after transition to anomalous state and in the closing stage of growth are given in Section 5. Section 6 contains a discussion of the experimental data and of the results of checking some hypotheses; also treated in this section are the processes which possibly initiate this transition.

2. EXPERIMENTAL PROCEDURE

The experimental apparatus and the container in which the investigations were performed are described in methodological papers [18–20] and in a number of other papers [9, 21]. The investigations were performed in an optical container (see Fig. 1), which made possible the filming of the crystal in the process of growth. The inner volume of the container was 4 cm^3. The nucleation and subsequent growth of a crystal occurred on a tungsten needle *4* whose point was located on the optical axis of the container. Crystal production was initiated by a high-voltage pulse applied to the needle. The use of this procedure resulted in solving of the following problems:

(1) the effect of the wall is ruled out: the crystal grows almost freely, without touching the container walls;

(2) the emergence of crystal is easily synchronized with the operation of the video system;

(3) the initial supersaturation in helium Dp_0 may be selected in the range from zero to the pressure of spontaneous crystal nucleation on the wall (0–15 mbar);

(4) one can change the conditions of crystal nucleation by varying the pulse amplitude and duration.

This procedure made it possible (see [15, 17]) to construct the phase diagram of the anomalous state and film the fast crystal growth.

The crystals were grown from thermomechanically cleaned helium. The experimental sequence was as follows (see the time diagrams in Fig. 1). First, desired supersaturation was set up in the fluid. Then, a high-voltage pulse *B* was supplied to the needle at the moment of time synchronized with the video signal (series of frame pulses *A*) with the delay t_1. The first LED flash of about 15 μs in duration was delayed by t_2 relative to the high-voltage pulse (see the series *C*). Subsequent flashes of lighting were synchronized with the frame pulses of the video signal that followed with a period of 20 ms. Therefore, the time interval between the first two frames was less than 20 ms (and usually amounted to 13.6 ms), and that between the subsequent frames was equal to the frame pulse period. Because the video picture is transmitted in half-frames, the odd frames are shifted by one line on the vertical relative to the even frames. The crystal growth was recorded in the memory of a signal processor (the first five frames) [20] and, if required, was simultaneously recorded on videotape. The minimal interval between the first and second frames is limited by the duration of the synchronization pulse and cannot be shorter than 2 ms. This time resolution is insufficient to get an idea of the stage of fast crystal growth. In order to solve this problem, we modified the procedure. The experiment in crystal growing at a fixed temperature was repeated approximately 15 times; in so doing, the initial supersaturation in the container, the voltage, and the duration of the high-voltage pulse were identical. It turned out that the curves of pressure variation agreed within the measurement accuracy, and the shape and orientation of crystals photographed with one and the same delay t_2 were similar; i.e., the process of growth was well reproducible. We varied the flash delay with a step equal to the duration of linear sweep of 64 μs to obtain a series of frames

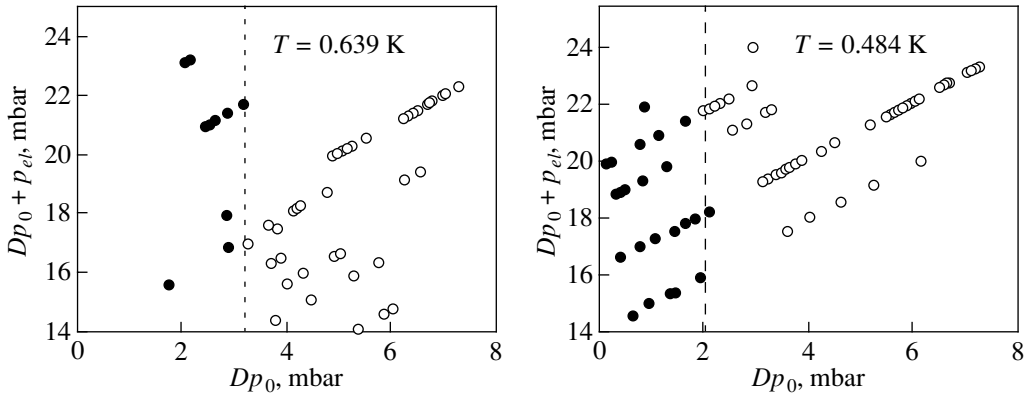


Fig. 2. The emergence of an anomalous state as a function of external hydrostatic supersaturation and of supersaturation at the moment of critical nucleation. Hollow points indicate anomalous growth, and solid points indicate normal growth; Dp_0, initial hydrostatic pressure; p_{el}, addition due to electrostatic field.

which give an idea about the variation of crystal shape in the mode of oscillating growth (see Fig. 7 below). The curve of pressure variation is averaged over 16 measurements. One can see that the crystals produced under identical conditions have one and the same orientation.

The pressure drop accompanying the crystal growth was measured synchronously with video recording by a capacitive pressure cell located on the container. The cell diaphragm is at a distance of approximately 1 cm from the needle. This results in a delay of the order of 30 μs due to the propagation of the first-sound wave in liquid helium. The response of the measuring path to a step variation of pressure consists in a delay of 32 μs and exponential relaxation with a single time constant. Within a single series of experiments, the relaxation constant remained the same. In different series, its value was 54 or 77 μs. For exactly relating the pressure values to the moment of filming the crystal in the stage of fast growth, correction was performed consisting of a time shift by 62 μs and compensation of the transient characteristic.

Note that no special attention was given to the pulse amplitude and duration in previous studies [15–17], because the pulse had only to initiate the guaranteed crystal nucleation. Therefore, the amplitude was taken to be high and identical for all experiments, and the duration was several milliseconds. In [15–17], the beginning of the pulse was recorded, and the voltage was switched off at the moment of filming the first frame.

3. THE EFFECT OF THE CONDITIONS OF CRYSTAL NUCLEATION

These experiments were performed to check the hypothesis that relates the emergence of an anomalous state to the behavior of critical nucleation. Assume that the nucleus emerges in the classical case with an already faceted surface, and in the quantum case its surface is in the atomically rough state, thus ensuring a high growth rate. Indeed, in the classical case, the subcritical nucleus is in a close-to-equilibrium state, which corresponds to the faceted form at the given temperatures. In the quantum case, the nucleus emerges in the final state. Note that the number of states of a nucleus with a disordered surface is larger, because the entropy of such a surface is higher than the entropy of a crystal surface consisting of flat facets. Therefore, during tunneling, the nucleus in most cases will find itself in states with an atomically rough surface. Of course, a state like this is nonequilibrium, and the relaxation of the nucleus to its equilibrium form will proceed simultaneously with its growth. If the relaxation time is long (see the footnote made by Nozieres and Uwaha in [22], p. 403), the crystal surface will remain in the atomically rough state until the termination of its growth. The relaxation to the equilibrium form will manifest itself in the fact that, because of the dynamic emergence of flat portions on the surface, the rate of growth will be lower than the rate of growth of atomically rough surfaces.

If this hypothesis is valid, then the boundary between normal and anomalous growth is the boundary of transition from classical to quantum nucleation. In 1972, Kagan and Lifshitz [23] have called attention to the fact that the boundary is sharp and the transition supersaturation increases with temperature, which agrees qualitatively with the form of the phase diagram in [17]. Then, instead of the external hydrostatic pressure, the main parameter of the process is the sum of this pressure with the electrostatic pressure of the pulse. This fact was noted by Yu.M. Kagan, who suggested an experimental investigation of the effect of the pulse amplitude. If the foregoing hypothesis is valid, the position of the boundary between normal and anomalous growth must be defined by the sum of hydrostatic and electrostatic pressures referred to above.

Figure 2 gives the results of measurements in which these parameters were varied independently at two temperatures. One can see in the graph that it is by the hydrostatic pressure of the medium that the boundary

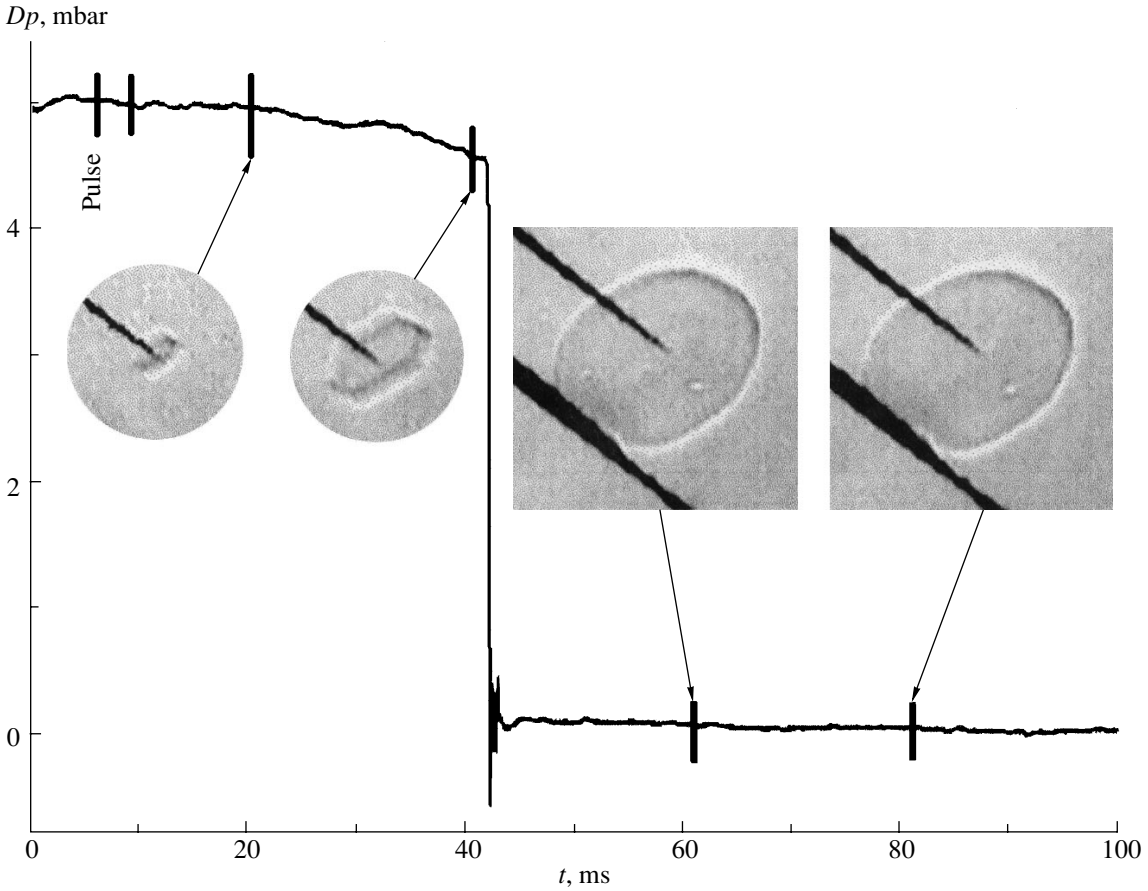


Fig. 3. An example of a record of the pressure drop in the container upon emergence of anomalous state with a delay of 36 ms at $T = 0.639$ K. Prior to the moment of transition, one can see the crystal faceting characteristic of growth below the first and second roughening transitions.

between normal and anomalous growth is defined. This means that the hypothesis relating the effect of anomalous growth to nucleation is invalid.

4. THE KINETICS OF EMERGENCE OF ANOMALOUS STATE

Because, as has now become clear, the anomalous state does not arise simultaneously with crystal, but develops under the effect of external hydrostatic pressure, it is of interest to determine the characteristic times of this process. For this purpose, the pulse duration was reduced to 128 μs (two lines of linear sweep of the video adapter), and the amplitude was selected such as to guarantee the nucleation of crystal during this time. The overall time of pressure recording was approximately 230 ms, and the crystal was photographed during the first 80 ms.

Figure 3 illustrates the crystal growth in the case when the transition to a state with a fast-growing surface occurs 36 ms after the crystal nucleation. One can see that the crystal first grows in the form of a hexagonal plate up to the moment of transition to an anomalous state, after which it grows completely during approximately 170 μs, which is accompanied by a drop of pressure to that of phase equilibrium. This pattern is typical for all observed series with long delay times. Figure 4 gives the curves of pressure variation during crystal growth with initial supersaturation above the transition boundary. One can clearly see that, on approaching the boundary supersaturation Dp^*, the delay time τ_{jump} between the crystal nucleation and the beginning of fast growth increases. The magnitude of pressure jump in the case of fast growth may be much less than that of the initial supersaturation (see Fig. 4). The general tendency toward increasing delay time τ_{jump} as the boundary of anomalous region is approached is illustrated by Fig. 5. Note that, for close values of initial supersaturation Dp_0, there nevertheless exists a considerable scatter of the values of τ_{jump}. For example, at $Dp_0 \approx 4.8$ mbar and $T = 0.639$ K, the value of delay varies from experiment to experiment in the range from 8 to 40 ms. As the temperature decreases, the value of delay with the same initial supersaturation decreases as well. However, one cannot unambiguously infer if it is the temperature decrease that leads to the reduction of the delay time, because, as is seen in

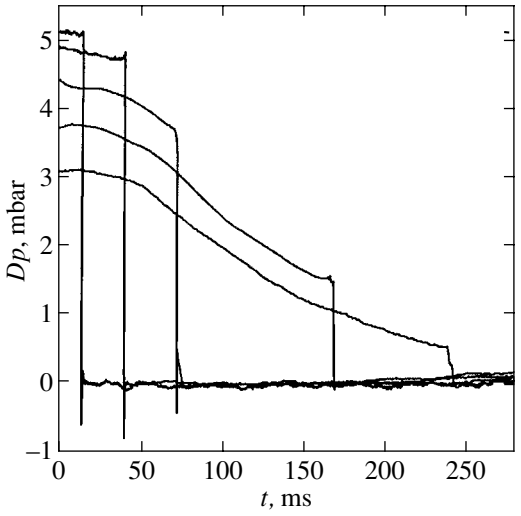


Fig. 4. Records of the pressure drop in the container during the time of crystal growth at $T = 0.639$ K with different values of initial supersaturation. An increase in Dp_0 leads to a reduction of the time of emergence of anomalous state.

Fig. 4, a decrease in the boundary supersaturation Dp^* occurs simultaneously. If we combine the data of Dp^* for two different temperatures with the respective shift of data, it will turn out that the data agree with one another within the scatter. Therefore, it still remains to be determined which factor defines the variation of τ_{jump}, namely, the temperature or the proximity to the transition boundary.

Prior to transition to the anomalous state, the crystal grows rather slowly. Its shape, as follows from Fig. 3, is almost invariable, which is indicative of the proportionality of the rates of growth of the *a*- and *c*-facets. For the case of crystal growth to supersaturation of about 0.5 mbar, this fact was previously observed in [21]. The results of filming demonstrate that the rate of growth of the *c*-facet is lower than the rate of growth of the *a*-facets, with the anisotropy of the rate of growth V_a/V_c being of the order of 2–3.5. As for the case of normal crystal growth, the rates of growth of equivalent *a*-facets differ from one another by almost 50% in some cases, which is indicative of the effect of surface defects on the kinetics of growth. Figure 6 gives the supersaturation dependence of the rate of growth for crystals prior to transition and for crystals growing normally up to the pressure of phase equilibrium. Because, as was already mentioned above, a considerable scatter is observed of both the anisotropy of growth and of the absolute values of the growth rate, the values of the growth rate were averaged as follows. The time dependence of pressure for seven series of measurements was used to determine the variation of the crystal volume in the process of growth. The quantity

$$R = \left(\frac{3}{4\pi}\mathcal{V}_c\right)^{1/3}$$

($\mathcal{V}_c$ is the crystal volume) was taken to be the linear dimension of crystal. We differentiate the function $R(t)$ to obtain the average rate of growth related to the rate of growth of the *a*- and *c*-facets by relation (A.7). The results obtained in crystals in which no transition was observed until the termination of growth are indicated by hollow points. Solid points indicate the results of such treatment of three curves with the longest delay times, as given in Fig. 4. One can see that these correlations agree within the scatter. Therefore, prior to transition, the crystal growth does not differ qualitatively and quantitatively from normal growth at small values of

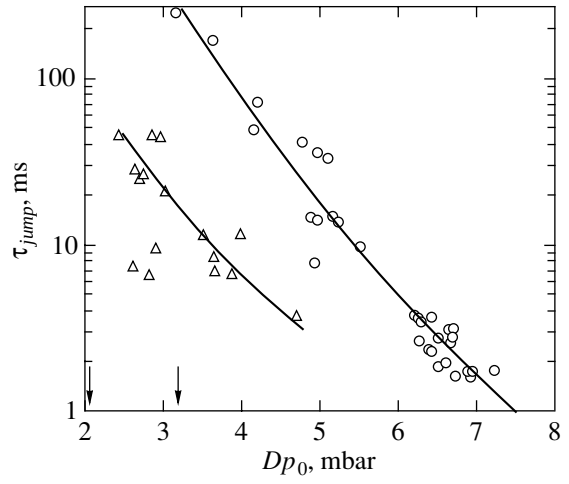


Fig. 5. The time of emergence of anomalous state as a function of initial supersaturation at 0.639 K (circles) and at 0.484 K (triangles). The arrows indicate the position of the boundary of anomalous region at these temperatures.

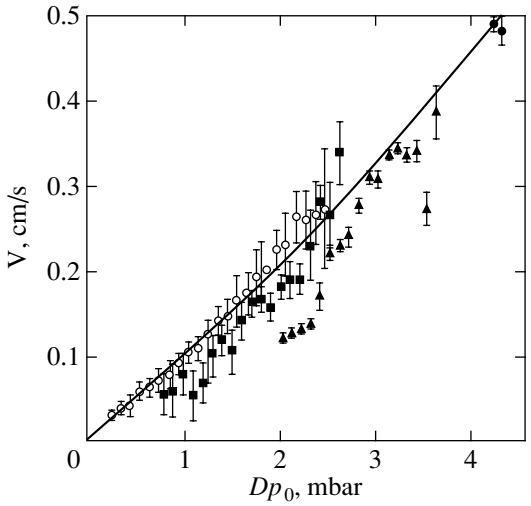


Fig. 6. The averaged rate of crystal growth in the normal state (hollow points) at $T = 0.639$ K and the results of treatment, by the same method (see the text), of three curves of pressure drop given in Fig. 4 (solid points).

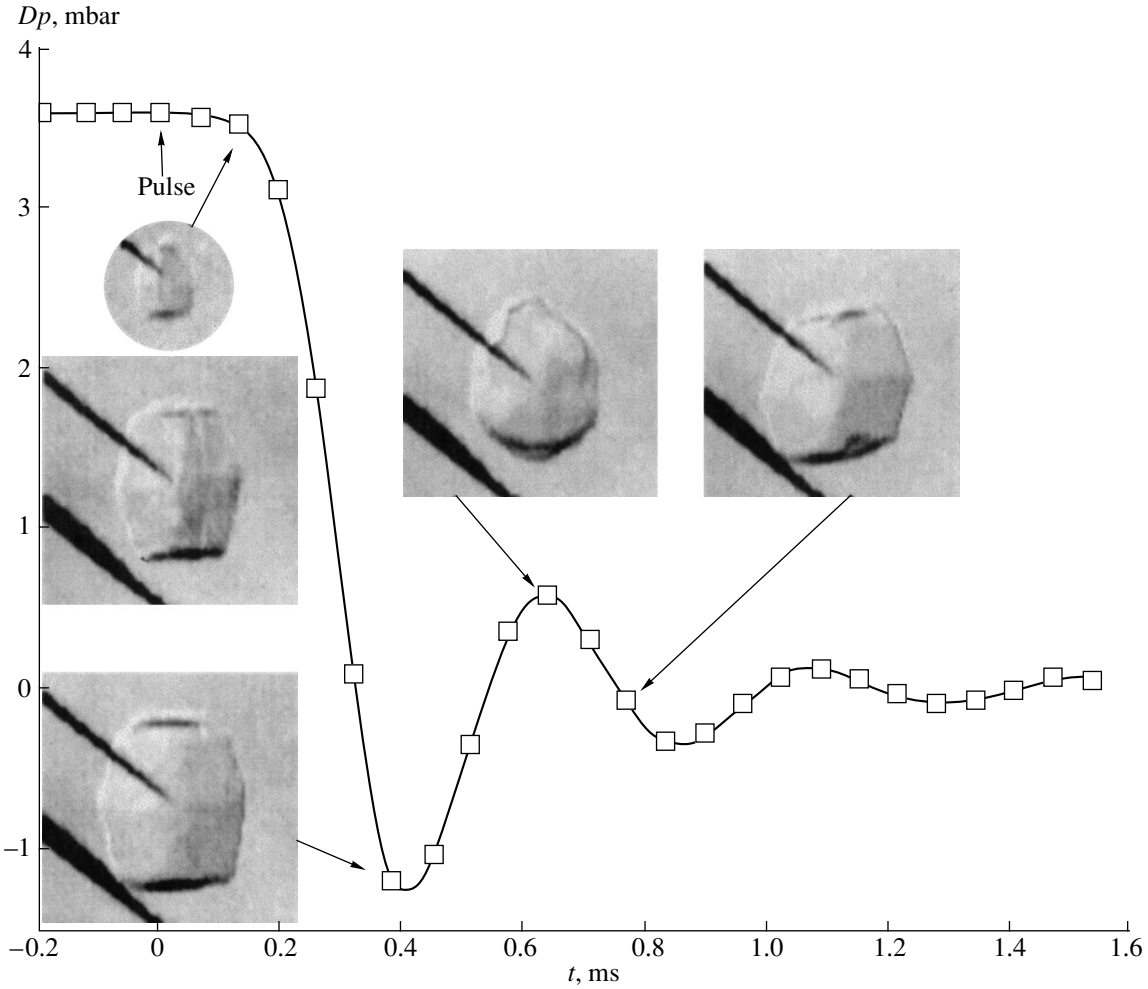


Fig. 7. The oscillating crystal growth at $T = 0.484$ K and $Dp_0 = 3.6$ mbar. One can see the kinetic faceting in the process of growth and the rounding of crystal upon melting.

supersaturation. As observed by a number of researchers [11, 12, 21], such a growth is due to well-known physical mechanisms such as the spiral growth and the growth induced by Frank-Read sources. The effect of defects shows up, for example, in that the equivalent crystal facets grow at different rates, which results in different sizes of the side facets (see Fig. 3).

In summing up the results of these experiments, we will formulate the main features of the process of emergence of an anomalous state of a helium crystal with a fast-growing surface.

(1) The transition to an anomalous state is realized only if the initial supersaturation exceeds the boundary supersaturation corresponding to the phase diagram in [17], $Dp_0 > Dp^*$.

(2) The time of delay between the crystal nucleation and beginning of fast growth increases with approaching the transition boundary.

(3) The values of delay are characterized by considerable scatter.

(4) Prior to the beginning of fast growth, the kinetics of facet growth are defined by the surface defects and do not differ qualitatively and quantitatively from the kinetics at small values (< 1 mbar) of supersaturations.

5. THE KINETICS OF CRYSTAL GROWTH AFTER TRANSITION TO AN ANOMALOUS STATE

After transition to an anomalous state, the crystal growth proceeds so rapidly that a radial acoustic oscillation with a frequency of the order of 2–3 kHz is excited in the container volume [16]. As a result, a pressure gradient now exists in the liquid, and the readings of the sensor located on the wall do not coincide with the values of pressure on the crystal boundary. It is also of importance that the crystal growth and melting are asymmetric processes. During growth, the crystal shape is defined (as is seen in Fig. 7) by the slow-growing facets. During melting, on the contrary, its shape is a result of the melting of surfaces with a high value of the kinetic coefficient of growth (in our case, of the

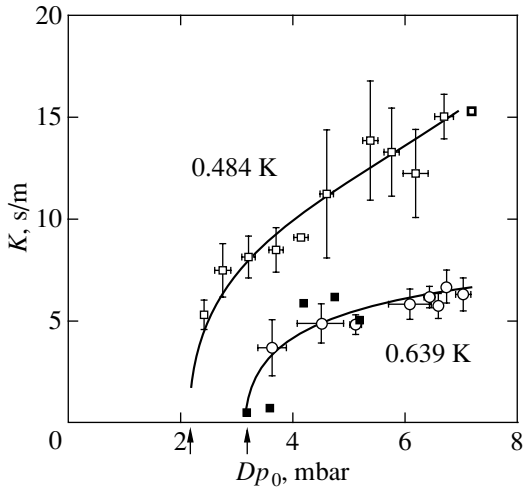


Fig. 8. The dependence of the kinetic coefficient of facet growth on initial supersaturation, determined in the first stage of oscillating crystal growth. The hollow symbols indicate the data obtained with the times of formation of the anomalous state of less than 10 ms. The solid squares indicate the results of treatment of the curves in Fig. 4. The arrows indicate the boundary of the anomalous region for the given temperatures.

atomically rough regions). Therefore, the growth equations will include different kinetic parameters in the stages of crystal growth and melting; also of importance is the ratio between the areas of slowly growing facets and atomically rough regions. Below, we will restrict ourselves to treating two stages of the process of crystal growth, namely, the initial region of fast crystal growth from the moment of nucleation to the beginning of first melting and the end region in which the pressure variation becomes small. In the former case, as is seen in the photographs of the growth process (Fig. 7), the crystal grows in the form of a hexagonal prism and retains its shape, which makes it possible to determine the kinetic coefficient of facet growth in the anomalous state. In the latter case, the damping of oscillation of the crystal-liquid system must be largely defined by the kinetics of growth/melting of the atomically rough regions of the crystal surface, which enables one to estimate the effect of transition on the kinetics of atomically rough surfaces.

5.1. Crystal Growth after Nucleation

We will treat the spherically symmetric problem of crystal growth in a container of radius R_0, assuming that the rate of facet growth V is proportional to supersaturation,

$$V = K\frac{\Delta\rho}{\rho\rho'}Dp = K\left((s-s')\Delta T + \frac{\Delta\rho}{\rho\rho'}p\right), \qquad (1)$$

where ρ is the density, s is the entropy per unit mass, and ΔT and p denote the deviations of temperature and pressure from the line of phase equilibrium. The symbols with primes indicate the solid phase, and those without primes indicate the liquid phase. In the temperature range for which this problem is treated, the entropies of the solid and liquid phases are virtually the same, and the measure of nonequilibrium is equal, to a good accuracy, to p (for more detail, see [24]). The flow equations for an ideally compressible liquid are as follows:

$$\Delta\varphi - \frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} = 0,$$
$$v = \frac{\partial\varphi}{\partial r}, \quad p = -\rho\frac{\partial\varphi}{\partial t}, \qquad (2)$$

where c is the velocity of sound. The boundary conditions

$$v(r_0) = -\frac{\Delta\rho}{\rho}V = -\frac{\Delta\rho}{\rho}K\frac{\Delta\rho}{\rho\rho'}p(r_0) = K\frac{\Delta\rho^2}{\rho\rho'}\frac{\partial\varphi}{\partial t}\bigg|_{r=r_0},$$
$$v(R_0) = \frac{\partial\varphi}{\partial r}\bigg|_{r=R_0} = 0 \qquad (3)$$

allow for the absence of liquid flow on the container wall and for the equation of continuity on the crystal boundary. In this problem, the crystal is treated as a sphere of radius r_0 with the isotropic coefficient of growth K. The initial conditions at the moment of crystal nucleation have the form

$$p(r) = p_0, \quad v(r) = \begin{cases} -K\dfrac{\Delta\rho^2}{\rho^2\rho'}p_0, & r = 0 \\ 0, & r > 0. \end{cases} \qquad (4)$$

For small deviations of the crystal surface from the equilibrium shape, Eqs. (2) with boundary conditions (3) give the oscillating crystal growth with damping defined by the value of the parameter K (see [16]). There is no small parameter in the stage of interest to us, and it is necessary to solve the nonstationary problem of crystal growth simultaneously with that of the propagation of the first-sound wave in the container volume. The numerical calculation results are given in the Appendix.

Figure 8 gives the dependence of average values of the growth coefficient on initial supersaturation at two temperatures. One can see in the graph that the kinetic coefficient of growth increases away from the transition boundary, which agrees qualitatively with the behavior of damping of pressure oscillation described in [16]. Nevertheless, the absolute values of K at $T = 0.484$ K and $Dp_0 = 8$ mbar are almost 40 times less than the respective values for atomically rough surface at the same temperature. The dark squares indicate the values of K for the growth curves with jump delays of 14 to 240 ms relative to the moment of crystal nucleation (Fig. 4). Within the scatter, these points agree with the values determined with a short delay (< 10 ms), for which

the transition to fast growth occurs at almost the same degree of supersaturation at which the crystal nucleation occurs. This means that the kinetic properties of the surface in the stage of fast growth are defined by the initial supersaturation in the liquid rather than by the pressure at which the transition occurs.

5.2. Damping of Oscillation upon Termination of Growth

As was mentioned above, the crystal growth and melting in the general case are due to the kinetics of the entire crystal surface, and one cannot distinguish between the contributions made by the facets and atomically rough surfaces using the correlations $p(t)$ alone. However, as the pressure oscillation is damped, the moment comes when the oscillation amplitude becomes so low that the melting–solidification of only the atomically rough regions is sufficient to compensate for the variation of the liquid density. Below this amplitude of p_{low}, the growth kinetics and, consequently, the damping are defined by the kinetic coefficient of growth of only the atomically rough regions. By measuring the damping of pressure oscillation below this threshold, one can determine the kinetic coefficient of surface growth after a time of the order of 0.5–1 ms following the beginning of fast crystal growth.

The threshold pressure p_{low} is determined from geometric considerations. Because the rate of facet growth is much lower than that for the rough regions, we will assume the facets to be stationary. In this case, the crystal growth is defined by the rough surfaces alone. The boundary pressure corresponds to the situation when, under conditions of maximal melting, the surface will nevertheless still contact the side facet (we take the side facets because, as is seen in Fig. 7, the transverse dimension of crystal is greater than its thickness). We assume the transverse dimension-to-thickness ratio to be $L_a/L_c = r$ to derive that, for the values of $r = 2$ to 4, the p_{low}/Dp_0 ratio is in the range from 0.15 to 0.1. Figure 9a gives a record of pressure oscillation during crystal growth, in which one can clearly see six periods. The values of extreme amplitudes on a semilog scale are given in Fig. 9b. One can see that the damping decrement has one and the same value both above $p_{low} \approx 1.5$ mbar and below this pressure. By the value of damping, one can determine the kinetic coefficient of growth of atomically rough surfaces in an anomalous state, $1/K = 0.023$ m/s. This value is intermediate between the coefficient of growth for anomalous facets ($1/K_{facet} = 0.071$ m/s) and the coefficient of growth for atomically rough surfaces ($1/K_{rough} = 4.5 \times 10^{-4}$ m/s) measured previously by the crystallization-wave method [5] with small deviations from equilibrium.

Therefore, the rate of growth of atomically rough surfaces is higher than the rate of facet growth, which agrees with the observed kinetic faceting of crystal in the stage of growth and with the variation of the crystal shape in the process of melting. The fact that the kinetic

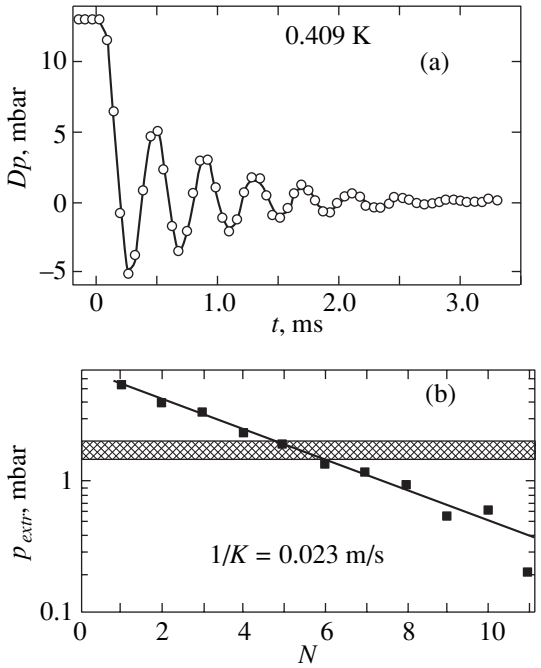


Fig. 9. (a) Pressure oscillation during crystal growth deep in the anomalous region. (b) The decrease of pressure amplitudes with time, N is the number of oscillation half-period of the curve in (a). The supersaturation region is shaded, where a variation of the decrement of oscillation damping should have been observed if the kinetic coefficient of growth of atomically rough surfaces in anomalous state corresponded to the coefficient measured at low values of supersaturation in normal crystals.

coefficient of growth of an anomalous crystal surface is much less than the coefficient of growth of a normal crystal surface may have two interpretations. First, this decrease in the surface mobility may be associated with the variation of the physical mechanism responsible for the motion of elementary surface defects, i.e., steps. With such supersaturations and temperatures, the rate of surface growth reaches the value of approximately 10 m/s, and the rate of motion of the steps, as follows from the results of estimating their mobility by the data of [5, 22, 25], becomes constant approaching the velocity of sound. A similar case, as applied to spiral facet growth, was treated by Ruutu *et al.* [12]. A theoretical analysis of the latter study reveals that, in this "relativistic" situation, the mobility of the steps decreases. The second possibility is associated with the variation of the state of crystal and, as a result, of the kinetics of its surface. In order to distinguish between these options, one must measure the kinetic coefficient of growth under conditions when the viscosity of the medium is fairly high and the velocity of the steps is much lower than the velocity of sound. As is demonstrated by the estimates, in the case of supersaturation of up to 15 mbar, these conditions are valid at a temperature above 0.6 K. In this region, the damping of pressure oscillation is too fast, which prevented the experimental determination of whether the decrement varies at low amplitudes.

6. DISCUSSION OF THE RESULTS

We will sum up the main features of the kinetics of the formation of an anomalous transition. One can see from the experimental results that the process that leads to the emergence of an anomalous state is initiated by external supersaturation. This process has a considerable duration in the vicinity of critical supersaturation Dp^*; however, the process accelerates considerably away from the phase boundary. As for the boundary between the anomalous and normal regions, it is possibly kinetic. In other words, this transition is not observed at $Dp < Dp^*$ only because the time of the formation of a fast state becomes longer than the time of crystal growth. The kinetics of crystal growth before transition do not differ qualitatively and quantitatively from the kinetics of normal crystal growth; i.e., the process leading to transition does not reflect on the rate of facet growth before the moment of transition. This behavior of the phenomenon brings it even closer to the phenomenon of burst-like growth of dislocation-free facets of crystal observed in the temperature range from 0.02 to 0.25 K. An assumption on the generality of the physical mechanisms of these phenomena was made previously in [15] on the basis of the similarity of their phase diagrams.

Another qualitative aspect consists in that the variation of kinetics was not caused by the surface transition alone. This inference was made in [16] on the basis of the results of filming crystals in the mode of oscillating growth. As is seen in Fig. 7, the melting of crystal accompanied by disappearance of facets begins after the growth stage. If the effect was associated with the surface alone, the rate of facet growth in the next stage of growth (that starts upon supersaturation that corresponds in the phase diagram to the region of normal growth) would be two–three orders of magnitude lower, and the oscillation would cease. Because no such situation is observed in the experiment, the reason for high mobility has to do with the volume. This inference is indirectly supported by the previously observed fact that the transition occurs simultaneously on all facets, as well as by the fact that the mobility of the atomically rough surfaces of an anomalous crystal differs from that of the atomically rough surfaces of a normal crystal.

The observation of oscillating crystal growth reveals that the time of existence of an anomalous state after its emergence exceeds 3 ms (see Fig. 9). The results of measurement of pressure variation after the completion of crystal growth, associated with the flow of liquid into the container, have led us to the conclusion that the kinetics of facets become normal in approximately 100 ms. Therefore, the relaxation of crystal to equilibrium state occurs in a period of time less than 100 ms.

It follows from the foregoing that the effect resides essentially in the variation of the bulk properties, is kinetic, and shows up as acceleration of facet growth. Based on the available experimental data, it is not yet possible to define the type of rearrangement of volume nor to determine whether it occurs in crystal or in liquid. No explanation can be found for this phenomenon within the framework of the existing models. Therefore, we will treat the factors which were not included in the classical model of spiral growth. According to this model, it is assumed that the spiral center associated with screw dislocation is stationary. This assumption is valid for the case of crystal growth at high temperatures at which the phonon viscosity is very high. In the case of helium at temperatures close to absolute zero, the viscosity is low and decreases with temperature as T^n, where $n = 3$ to 5 [26–30]. This means that the spiral rotation may bring about the excitation of dislocation oscillation and additional dissipation associated with dislocation friction. This process will cause, on the one hand, the heating of crystal and, on the other hand, the excitation of the dislocation subsystem of crystal. For qualitative estimation of this effect, we will restrict ourselves to the range of temperatures above 0.55 K, where the velocity of individual steps is much less than the velocity of sound up to supersaturation of the order of 15 mbar, the corrections of Ruutu *et al.* [12] are minor, and the classical relations for spiral growth are valid. One can treat the motion of dislocation in Koehler's approximation of the string model [31] to derive an approximate expression for the power transferred to dislocation by the growth spiral,

$$P \approx \frac{\beta^2}{C}\sqrt{\frac{C}{A}}\frac{\eta}{1+2\eta^{3/2}}, \quad \eta = \frac{1}{2\omega}\frac{B}{A}. \tag{5}$$

Here, β is the line energy of a step; ω is the angular velocity of rotation of the growth spiral; and A, B, and C are the string model parameters defined by the expressions

$$\begin{aligned} A\frac{\partial^2\zeta}{\partial t^2} + B\frac{\partial\zeta}{\partial t} - C\frac{\partial^2\zeta}{\partial z^2} &= 0, \\ A = \frac{\rho' b^2}{\pi}, \quad C &= \frac{2Gb^2}{\pi(1-\nu)}, \end{aligned} \tag{6}$$

where ρ' is the solid helium density, b is the Burgers vector, G is the shear modulus, and ν is the Poisson ratio. The numerical estimate for the dislocation concentration $\Lambda = 2 \times 10^5$ cm^{-2} (determined for crystals grown on a needle [21] by the value of threshold supersaturation from which the facet growth starts) indicates that the additional dissipation associated with dislocations is negligibly small. In this temperature range, up to supersaturation of the order of 15 mbar, this contribution does not exceed 1% of the energy of phase transition. Therefore, the energy of dislocation oscillation excited by spiral growth is much less than the characteristic energy of the process of growth. For this reason, the foregoing mechanism will hardly have a considerable effect on the state of a crystal, and it is unlikely that

this phenomenon will bring about the observed transition to the state with a fast-growing surface.

Another assumption of the classical model consists in that the flow of liquid to the surface during spiral growth is treated as laminar and irrotational. This is apparently true of the growth of ordinary crystals, when the rate of liquid flow to the boundary is low. In our case, the mobility of the steps is high, so that their velocity may approach the velocity of sound. The entrainment of liquid by the growth spiral develops a disturbance of laminar flow, this possibly leading to the generation of vortices. A vortex filament attracted toward the solid boundary may in principle produce steps on the surface and, thereby, develop centers of facet growth. Indeed, the flow of liquid in the vicinity of the vortex core causes a reduction of pressure, i.e., promotes the melting of crystal. As follows from estimation, the production of steps at a distance of the order of 10^{-7} cm from the surface occurs without a threshold. However, the numerical estimates of the possibility of generation of vortices by growth spirals are very pessimistic. We assume for the associated mass the values obtained in [22, 32] and derive that the energy and angular momentum of liquid are too low to produce a vortex line issuing from the center of growth spiral and closing on a spiral with the opposite sense of rotation (Frank-Read source). The production of vortex rings by a growth spiral remains an open question, because the flow of liquid in the vicinity of a moving step is potential. Therefore, it is unlikely that the spiral growth produces in a superfluid vortices with a concentration sufficient to cause a variation of the kinetics of facet growth.

7. CONCLUSION

The results of studies into the emergence of an anomalous state of helium crystal with a fast-growing surface, produced by a short electrostatic pulse in metastable superfluid helium, have demonstrated that the emergence of this state requires considerable time whose duration increases on approaching the boundary of the anomalous region. The results of observation of oscillating crystal growth have demonstrated that this phenomenon is not due to the variation of the surface properties alone. Therefore, the anomalous state develops under the effect of external supersaturation, which leads, as one can see from the experimental data, to a variation of the bulk properties of crystal or liquid. It follows from the results of treating the phenomena accompanying the spiral growth (dislocation oscillation and generation of vortices) that these mechanisms will hardly lead to the transition being observed. Moreover, both these mechanisms require the presence of a sufficient number of dislocations. If this effect and the phenomenon of "explosive" facet growth [12] are of the same physical nature, the transition to the state with a fast-growing surface is not at all associated with dislocations, because, in the latter case, it is on the dislocation-free facet that a fast growth is observed. Therefore, the observed phenomenon cannot be explained within the known models by modifying those models and extending them to the case of helium crystal growth in a superfluid.

ACKNOWLEDGMENTS

I am grateful to L.A. Maksimov for numerous discussions in the process of my work and to Yu.M. Kagan for valuable discussion of the results and for ideas suggested during this discussion, as well as to L.B. Dubovskii and S.N. Burmistrov for useful working discussions. Special thanks go to A.Ya. Parshin for thorough critical discussions of the experimental results and their interpretation.

This study received support from the Russian Foundation for Basic Research (project no. 99-02-17289).

APPENDIX

We will treat the spherically symmetric problem of crystal growth in a container. In this approximation, we ignore the nonsphericity of the container (the 1 : 2 ratio of perpendicular dimensions) and replace a real crystal growing in the form of a hexagonal prism with the dimension ratio of 1 : k, where k = 2 to 3, by a sphere of the same volume. The kinetic coefficient of growth is taken to be isotropic. We will change over to a one-dimensional problem using the substitution

$$\varphi(r, t) = \frac{f(r, t)}{r}, \tag{A.1}$$

reduce Eqs.(2)–(4) to the dimensionless form by the replacement

$$R_0 r \longrightarrow r, \quad t\frac{R_0}{c} \longrightarrow t, \quad cv \longrightarrow v, \qquad \rho c^2 p \longrightarrow p, \quad cR_0^2 f \longrightarrow f, \tag{A.2}$$

and introduce the dimensionless parameter λ proportional to the surface mobility,

$$\lambda = cK\frac{\Delta\rho^2}{\rho\rho'} = K \cdot 3.53[\text{ m/s}]. \tag{A.3}$$

Then, Eqs. (2) and boundary and initial conditions (3) and (4) will take the form

$$\frac{\partial^2 f}{\partial t^2} - \frac{\partial^2 f}{\partial r^2} = 0, \quad v = \frac{\partial}{\partial r}\left(\frac{f}{r}\right), \quad p = -\frac{1}{r}\frac{\partial f}{\partial t}, \tag{A.4}$$

$$v(r_0, t) - \frac{\lambda}{r_0}\frac{\partial f}{\partial t}\bigg|_{r = r_0} = 0, \qquad v(1, t) = \frac{\partial}{\partial r}\left(\frac{f}{r}\right) = 0, \tag{A.5}$$

$$p(r, 0) = p_0, \quad v(r, 0) = \begin{cases} -\lambda p_0, & r = 0, \\ 0, & 0 < r < 1. \end{cases} \tag{A.6}$$

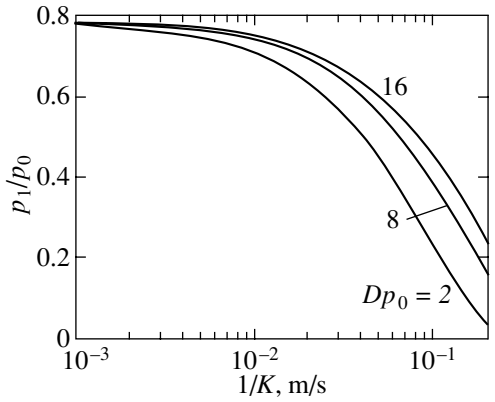


Fig. 10. The results of numerical calculation of the ratio between the amplitude of the first minimum of pressure, measured by a sensor on the container wall, to initial supersaturation, as a function of kinetic coefficient of growth K. The integration constant of the measuring circuit is 54 μs.

The solution of this problem gives the time dependence of pressure on the container wall where the capacitive sensor is located. Note that, in the case of fast growth with a delay (see above), the initial value of the crystal radius was taken to be finite and determined by the pressure at which the fast growth started. Then, the signal was convolved with the transfer function of the measuring system. By way of illustration of the method, Fig. 10 gives the results of calculation of the ratio between the amplitude of the first minimum of pressure to initial supersaturation for three values of initial supersaturation and for the integration constant of the transfer function of 54 μs used in the given series of experiments. The amplitude ratio may be used to determine the kinetic coefficient of growth if $1/K$ = 0.01–0.2 m/s.

Because the real crystal has the anisotropic coefficient of growth, we will give the relations that relate the averaged kinetic coefficient of growth K_0 to the coefficients of facet growth. For this purpose, one must use the crystal shape to determine the ratio between the average growth rates, $V_a/V_c = r$. Then,

$$K_c = K_0 \frac{1}{r^{2/3}} \frac{\pi^{1/3}}{\sqrt{3}} \approx K_0 \frac{0.846}{r^{2/3}}, \qquad K_a \approx K_0 r^{1/3} 0.846. \tag{A.7}$$

REFERENCES

1. A. F. Andreev and A. Ya. Parshin, Zh. Éksp. Teor. Fiz. **75**, 1511 (1978) [Sov. Phys. JETP **48**, 763 (1978)].
2. K. O. Keshishev, A. Ya. Parshin, and A. V. Babkin, Pis'ma Zh. Éksp. Teor. Fiz. **30**, 63 (1979) [JETP Lett. **30**, 56 (1979)].
3. S. Balibar and B. Castaing, J. Phys. Lett. **41**, 329 (1980).
4. J. A. Avron, L. S. Balfour, C. G. Kuper, *et al.*, Phys. Rev. Lett. **45**, 814 (1980).
5. K. O. Keshishev, A. Ya. Parshin, and A. V. Babkin, Zh. Éksp. Teor. Fiz. **80**, 716 (1981) [Sov. Phys. JETP **53**, 362 (1981)].
6. S. Balibar, F. Gallet, and P. E. Wolf, Phys. Rev. Lett. **51**, 1366 (1983).
7. A. F. Andreev and V. G. Knizhnik, Zh. Éksp. Teor. Fiz. **83**, 416 (1982) [Sov. Phys. JETP **56**, 226 (1982)].
8. R. M. Bowley and D. O. Edwards, J. Phys. (Paris) **44**, 723 (1983).
9. L. A. Maksimov and V. L. Tsymbalenko, Zh. Éksp. Teor. Fiz. **114**, 1313 (1998) [JETP **87**, 714 (1998)].
10. A. A. Chernov *et al.*, *Modern Crystallography*, Vol. 3: *Crystal Growth*, Ed. by B. K. Vaĭnshteĭn, A. A. Chernov, and L. A. Shuvalov (Nauka, Moscow, 1980; Springer-Verlag, Berlin, 1984).
11. S. Balibar, F. Gallet, P. Nozieres, *et al.*, J. Phys. (Paris) **46**, 1987 (1985).
12. J. P. Ruutu, P. J. Hakonen, A. V. Babkin, *et al.*, J. Low Temp. Phys. **112**, 117 (1998).
13. V. L. Tsymbalenko, Phys. Lett. A **211**, 177 (1996).
14. V. L. Tsymbalenko, Phys. Lett. A **248**, 267 (1998).
15. V. L. Tsymbalenko, Phys. Lett. A **257**, 209 (1999).
16. V. L. Tsymbalenko, Phys. Lett. A **274**, 223 (2000).
17. V. L. Tsymbalenko, J. Low Temp. Phys. **121**, 53 (2000).
18. V. L. Tsymbalenko, Cryogenics **36**, 65 (1996).
19. V. L. Tsymbalenko, Prib. Tekh. Éksp., No. 4, 161 (1997).
20. V. L. Tsymbalenko, Prib. Tekh. Éksp., No. 2, 78 (1999).
21. V. L. Tsymbalenko, Fiz. Nizk. Temp. **21**, 162 (1995) [Low Temp. Phys. **21**, 120 (1995)].
22. P. Nozieres and M. Uwaha, J. Phys. (Paris) **48**, 389 (1987).
23. Yu. M. Kagan and I. M. Lifshits, Zh. Éksp. Teor. Fiz. **62**, 385 (1972) [Sov. Phys. JETP **35**, 206 (1972)].
24. V. L. Tsymbalenko, J. Low Temp. Phys. **88**, 55 (1992).
25. O. A. Andreeva, K. O. Keshishev, and S. Yu. Osip'yan, Pis'ma Zh. Éksp. Teor. Fiz. **49**, 661 (1989) [JETP Lett. **49**, 759 (1989)].
26. I. Iwasa, S. Wales, and R. Wanner, Solid State Commun. **18**, 853 (1976).
27. K. Araki, I. Iwasa, and H. Suzuki, J. Phys. Soc. Jpn. **46**, 1119 (1979).
28. Y. Hiki and F. Tsuruoka, Phys. Lett. A **56**, 484 (1976); Phys. Rev. B **20**, 2702 (1979).
29. V. L. Tsymbalenko, Zh. Éksp. Teor. Fiz. **74**, 1507 (1978) [Sov. Phys. JETP **47**, 787 (1978)]; Zh. Éksp. Teor. Fiz. **87**, 943 (1984) [Sov. Phys. JETP **60**, 537 (1984)].
30. D. J. Bishop, H. W. Dail, and M. A. Paalanen, Phys. Rev. Lett. **46**, 664 (1981).
31. J. S. Koehler, in *Imperfections in Nearly Perfect Crystals*, Ed. by W. Shockley *et al.* (Wiley, New York, 1952).
32. A. M. Kosevich and Yu. A. Kosevich, Fiz. Nizk. Temp. **7**, 1347 (1981) [Sov. J. Low Temp. Phys. **7**, 655 (1981)].

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 1035–1037.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1195–1197.

SOLIDS
Structure

The Anisotropic Low-Temperature Specific Heat of Organic Superconductor κ-(BEDT–TTF)$_2$Cu(NCS)$_2$ in Magnetic Field¶

A. E. Kovalev[a, b, *], T. Ishiguro[a], J. Yamada[c], S. Takasaki[c], and H. Anzai[c]

[a]*Department of Physics, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan*
Japan Science and Technology Corporation, Kawaguchi 332-0012, Japan
[b]*Institute of Solid State Physics, Chernogolovka, Moscow oblast, 142432 Russia*
[c]*Faculty of Science, Himeji Institute of Technology, Koto, Akoh 678-1297, Japan*
**e-mail: kovalev@issp.ac.ru*

Received February 11, 2001

Abstract—We measured the low-temperature specific heat of the layered organic superconductor κ-(BEDT–TTF)$_2$Cu(NCS)$_2$ for the magnetic field directed along and across the conducting plane and found a difference between the two measurements. Our data indicate the existence of a nodeless superconducting state at zero field and low temperature. The field dependence of the specific heat anisotropy consists of two linear branches with the crossover field equal to the upper critical field perpendicular to the conducting plane. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The problem of superconductivity in low-dimension organic metals continues to attract much attention. A wide discussion on the possibility of an unconventional superconducting state in the κ-(BEDT–TTF)$_2$X compound involves arguments both in agreement and in disagreement with the hypothesis. For κ-(BEDT–TTF)$_2$Cu(NCS)$_2$, there is extensive evidence in favor of the unconventional character of superconductivity: NMR [1], high-frequency conductivity [2], thermal conductivity [3], and specific heat data [4]. As to the penetration depth data, both the unconventional [5] and the conventional [6] behavior have been reported. Recent data of Carrington *et al.* [7] support the presence of low-lying excitations but do not give a definite answer as to their origin.

The previously reported data [8] of the specific heat of κ-(BEDT–TTF)$_2$Cu(NCS)$_2$ under the magnetic field up to 6 T perpendicular to the conducting plane and the temperature region 1.65–4.4 K demonstrated an almost linear field dependence of the specific heat for the field considerably below the perpendicular upper critical field $H_{c2\perp}$. Above $H_{c2\perp}$, the specific heat is field-independent within the experimental error. To obtain more information about the character of the low-temperature superconducting state, we performed specific heat measurements under different orientations of the magnetic field.

¶This article was submitted by the authors in English.

2. EXPERIMENTAL

For the measurements, we used a modification of the standard ac-modulation technique; the experimental details are described in [9] and [10]. One single crystal with the total mass 0.45 mg was used. In addition to the specific heat, the magnetoresistance of the sample was measured. The Dingle temperature extracted from the Shubnikov–de Haas oscillation was about 0.5 K. The specific heat was measured mainly for the magnetic field orientation along and perpendicular to the conducting plane. It was rather difficult to estimate the absolute value of the sample specific heat because of the small sample mass and an involved and significant field dependence of the specific heat of the thermometer. It was previously found that the specific heat of the thermometer is isotropic in the magnetic field. Our experimental setup makes it possible to rotate the sample *in situ*. Calculating the difference between the specific heat measured in the magnetic field parallel and perpendicular to the conducting plane, $C_\perp - C_\parallel = \Delta C$, we obtained a reliable value of the specific heat anisotropy ΔC.

3. RESULTS AND DISCUSSION

In Fig. 1, we plot the low-temperature specific heat anisotropy ΔC of the layered organic superconductor κ-(BEDT–TTF)$_2$Cu(NCS)$_2$ divided by the temperature. We can see that there are two regions for each curve in Fig. 1: one is below $H_{c2\perp}$ (the upper critical field per-

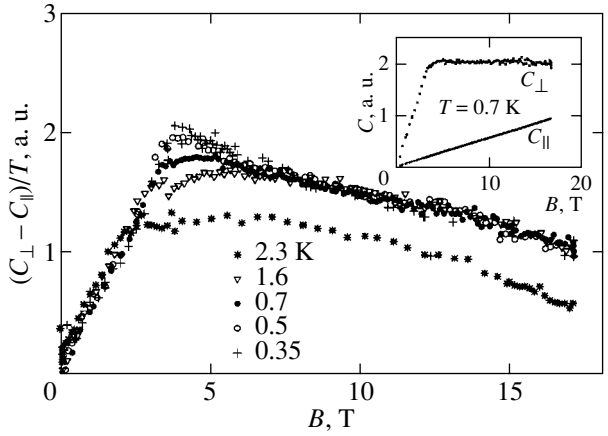


Fig. 1. The anisotropy of the field dependence of the specific heat at low temperatures. Inset: the extracted specific heat perpendicular ($C_\perp$) and parallel ($C_\parallel$) to the conducting layers.

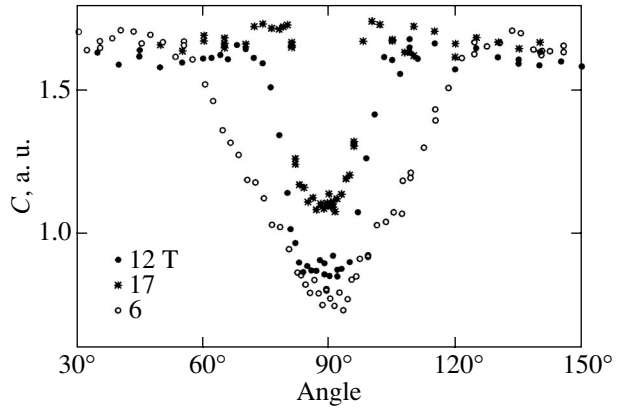


Fig. 2. The angular dependence of the specific heat at 0.5 K and various fields.

pendicular to the conducting layers) and the other is above this field.

We first note that, except in the region near $H_{c2\perp}$, all the curves coincide with each other. Using the data from [8], we can conclude that much below $H_{c2\perp}$,

$$C_\perp = A_1 TH, \tag{1}$$

and at all fields,

$$C_\parallel = A_2 TH, \tag{2}$$

where A_1 and A_2 are some constants. The field dependence of $C_\perp$ coincides with the one reported in [8], although it is different from the one reported in [4], where a more abrupt increase of the electronic density of states in the magnetic fields below 0.03 T was observed.

The linear dependence of the specific heat on the magnetic field follows from the London model. It gives the electronic specific heat expressed as [11]

$$C(H, T) \approx -\frac{T\Phi_0 H}{32\pi^2}\frac{\partial^2}{\partial T^2}\left[\frac{1}{\lambda^2(T)}\frac{\alpha H_{c2}(T)}{H}\right], \tag{3}$$

where Φ_0 is the flux quantum, $H_{c2}(T)$ is the upper critical field, and $\lambda(T)$ is the effective penetration depth in the plane perpendicular to the magnetic field. In the BCS approximation, Eq. (3) leads to the following field dependence of $C_\perp$ (see [12]) at the temperatures $T/T_c \ll 1$:

$$C_\perp \approx C_n \frac{H}{H_{c2\perp}(0)}, \tag{4}$$

with C_n being the electronic specific heat in the normal state. We recall that the London model, which is valid in the region $H_{c1} \ll H \ll H_{c2}$, neglects the Pauli breaking effect and, therefore, H_{c2} in Eq. (4) is actually the orbital upper critical field.

We note that the applicability of the London model does not depend on the superconducting pairing mechanism. The only important condition is a slow spatial variation of the order parameter over the length scale of the penetration depth. This condition is easily satisfied if the penetration depth of the magnetic field is much larger than the coherence length and the magnetic field is much lower than the upper critical field. In our case, the field dependence of the specific heat at low temperatures is almost linear up to the upper critical field.

Considerably above $H_{c2\perp}$, the data for the temperatures below 1.6 K almost coincide with each other. This indicates a BCS-like nodeless low-temperature superconducting state with the final gap in the excitation spectrum of the quasiparticle.

Using Eq. (4), we determined the upper critical field at the zero temperature as $H_{c2\perp}$ (0.5 K) = 4 T. The same value is obtained if we find the crossing point of the two linear branches. This is less than the value of about 6 T reported by Sasaki *et al.* [13] and that about 5 T reported by Belin *et al.* [14], which were determined from the magnetic torque and the heat conduction measurements, respectively. On the other hand, this is very close to the field of irreversibility at low temperatures reported in [13].

We note that the specific heat dependences on the magnetic field parallel to the layers is too strong for the expected Josephson coupling between the layers (see [15]). This problem requires a more detailed investigation of the low-temperature state for this orientation of the magnetic field. At present, we do not have any reasonable explanation of this fact.

The angular dependence of the specific heat is demonstrated in Fig. 2. It is noteworthy that the specific heat is almost independent of the angle in the region ±3° near the direction parallel to the conducting plane. We do not know the reason for this behavior but it can hardly be explained by the misalignment domain struc-

ture within the sample (with the planes inclined with respect to each other): such a misalignment has not been observed up to now [16].

4. CONCLUSION

In summary, we have demonstrated the nodeless superconductivity in κ-(BEDT–TTF)$_2$Cu(NCS)$_2$. The field dependence of the specific heat at the field direction parallel to the plane demonstrates the behavior that seems to be incompatible with the Josephson coupling between the layers. Further investigations are necessary in order to solve this problem. In addition, we estimated the upper critical field perpendicular to the layers to be about 4 T.

ACKNOWLEDGMENTS

The work of A. E. K. was supported by the CREST program of Japan Science and Technology Corporation, Japan, RFBR (project no. 99-02-16119), and the NWO (grant FN4359).

REFERENCES

1. A. Van-Quynh, C. Berthier, H. Mayaffre, *et al.*, Phys. Rev. B **59**, 12064 (1999).
2. J. M. Schrama, E. Rzepniewski, R. S. Edwards, *et al.*, Phys. Rev. Lett. **83**, 3041 (1999).
3. S. Belin, K. Behnia, and A. Deluzet, Phys. Rev. Lett. **75**, 4122 (1995).
4. Y. Nakazawa and K. Kanoda, Physica C (Amsterdam) **282**, 1897 (1997).
5. K. Kanoda, K. Akiba, K. Suzuki, *et al.*, Phys. Rev. Lett. **65**, 1271 (1990); L. P. Le, G. M. Luke, B. J. Sternlieb, *et al.*, Phys. Rev. Lett. **68**, 1923 (1992); D. Achkir, M. Poirier, C. Bourbonnais, *et al.*, Phys. Rev. B **47**, 11595 (1993).
6. D. R. Harshman, R. N. Kleiman, R. C. Haddon, *et al.*, Phys. Rev. Lett. **64**, 1293 (1990); M. Lang, N. Toyota, T. Sasaki, and H. Sato, Phys. Rev. Lett. **69**, 1443 (1992).
7. A. Carrington, I. J. Bonalde, R. Prozorov, *et al.*, Phys. Rev. Lett. **83**, 4172 (1999).
8. A. E. Kovalev, T. Ishiguro, G. Saito, *et al.*, J. Supercond. **12**, 515 (1999).
9. A. E. Kovalev, T. Ishiguro, T. Kondo, *et al.*, Phys. Rev. B **62**, 103 (2000).
10. V. A. Bondarenko, M. A. Tanatar, A. E. Kovalev, *et al.*, Rev. Sci. Instrum. **71**, 3148 (2000).
11. M. E. Reeves, S. E. Stupp, T. A. Friedmann, *et al.*, Phys. Rev. B **40**, 4573 (1989); S. E. Stupp, W. C. Lee, J. Giapintzakis, *et al.*, Phys. Rev. B **45**, 3093 (1992).
12. A. S. Fetter and P. C. Hohenberg, in *Superconductivity*, Ed. by R. D. Parks (Marcel Dekker, New York, 1969), Vol. 2, Appendix E.
13. T. Sasaki, W. Biberacher, K. Neumaier, *et al.*, Phys. Rev. B **57**, 10889 (1998).
14. S. Belin, K. Behnia, and A. Deluzet, Phys. Rev. Lett. **81**, 4728 (1998).
15. L. N. Bulaevskii, M. Ledvij, and V. G. Kogan, Phys. Rev. B **46**, 366 (1992).
16. N. D. Kushch, private communication.

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 1038–1048.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1198–1209.

SOLIDS
Electronic Properties

A Phenomenological Theory of Metastable States in Disordered Ising Magnets

P. N. Timonin*

Research Institute of Physics, Rostov State University, pr. Stachki 194, Rostov-on-Don, 344090 Russia

**e-mail: timonin@icomm.ru*

Received April 6, 2000

Abstract—A mechanism of the formation of an exponentially large number of metastable states in magnetic phases of disordered Ising magnets as a result of condensation of fractal delocalized modes near the localization threshold is suggested. The thermodynamic properties of metastable states are studied in the effective-field approximation in the vicinity of transitions in magnets with zero uniform magnetization in the ground state such as dilute antiferromagnets, spin glasses, and dilute ferromagnets with dipole interaction. These properties are shown to determine the parameters of nonequilibrium processes in the glassy phase, namely, the shape of the hysteresis loop, the thermodynamic values in field-cooled and zero-field-cooled regimes, and the thermoremanent and isothermal remanent magnetization values. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A characteristic feature of disordered magnets is the appearance of many metastable states accompanying magnetic transitions in them. These states cause irreversibility phenomena in the response of magnets to changes in the external magnetic field and temperature, such as the dependence of thermodynamic parameters on the order and rate of these changes, the appearance of hysteresis loops, whose shapes change depending on the amplitude of the applied field, etc. [1–4]. The irreversibility phenomena are inherent in all known types of magnetic disorder in the corresponding inhomogeneous magnetic phases, from dilute magnets with nonmagnetic atom impurities [2, 4] to spin glasses existing in crystalline solid solutions of ferromagnets with antiferromagnets [1, 3].

Evidences for the existence of metastable states, whose number exponentially depends on the number of spins, have been obtained in numerical studies of several models of disordered magnets with short-range exchange [5–8]. A theoretical description of related nonergodicity phenomena in the majority of models also requires the use of numerical methods, e.g., see [1, 4, 9].

The most important analytic results were obtained for the Sherrington–Kirkpatrik spin glass model with an infinite interaction radius. In this model, the nonergodicity region was found to be bounded by finite fields lower than the Almeida–Thouless H_{AT} field [10]. The thermodynamic parameters for field-cooled and zero-field-cooled regimes with subsequently applying it [1] were determined within the replica symmetry breaking scheme [11] and using the concept of the hierarchy of macroscopic relaxation times [12]. This is, however, not an exhaustive account of information about irreversible processes obtained in terms of this model. Although this approach is extensively used in studying other effective-field models, e.g., see [13, 14], it remains unclear how the methods suggested in [11, 12] can be used to describe all variety of irreversible processes related to transitions between various metastable states.

At the same time, a theoretical description of nonergodic effects would be quite obvious and simple if the physical properties of metastable states of disordered magnets, their regions of existence, and points of possible phase transitions between them at $H < H_{AT}$ were known. In infinite-range effective-field models, this information can in principle be obtained by studying (nonaveraged) equations for local magnetic moments. The best known example of using this approach is the Thouless–Anderson–Palmer (TAP) equations in the same Sherrington–Kirkpatrik model [15]. Although the number of solutions to these equations was shown to be exponentially large [16], their explicit form could not be determined by analytic methods. It is not even clear how many of these solutions correspond to thermodynamic potential minima and in what way barriers, which diverge in the thermodynamic limit, appear between them.

In addition, the results of studying TAP equations for a large but finite interaction radius [17] cast doubt on the possibility of using infinite-range models as a good approximation for describing real disordered magnets. In [17], attention was for the first time called to consequences of the qualitative difference between the spectrum of a Gaussian ensemble of random exchange matrices in the Sherrington–Kirkpatrik

1063-7761/01/9206-1038$21.00

model, in which all eigenvectors are delocalized, and the spectrum of short-range matrices with localized states at its tails [18]. According to [17], the condensation of local modes in a three-dimensional spin glass with a fairly large but finite interaction radius can change the character of the transition by excluding the possibility of the condensation of delocalized modes if the critical index of the localization radius $\nu > 2/3$. Although such a situation with the absence of metastable states can hardly arise in most real disordered magnets, these results show that qualitative peculiarities of the spectral characteristics of random exchange should be taken into account.

At the same time, studies of the spectra of various ensembles of short-range random matrices showed that all of them had like structures with localized states near the spectrum boundaries and a fractal structure of eigenvectors near the localization threshold [18–20]. It can therefore be assumed that the universal character of irreversibility phenomena in disordered magnets with different short-range random exchange is related precisely to the similarity of their spectral characteristics. The general mechanism of the appearance of many metastable states in such systems can be virtually simultaneous condensation of a macroscopically large number of almost overlapping and weakly interacting fractal modes, which are close to the localization threshold.

It should be noted that the finiteness of the interaction radius by no means rules out the possibility of a fairly accurate quantitative description of such a transition mechanism at the level of the effective-field approximation. The matter is that a macroscopically large number of condensing modes (order parameter components) substantially decreases the number of noncondensing modes with close eigenvalues (order parameter fluctuations) in comparison with a homogeneous magnet of the same dimensionality. It can therefore be expected that the Ginzburg parameter should be noticeably lower in such a transition than in homogeneous magnets for all ensembles of random short-range exchanges.

Based on these ideas, we make an attempt to construct a phenomenological effective-field theory of disordered Ising magnets with zero magnetization in the ground state; this theory uses heuristic assumptions on the form of fractal eigenvectors of exchange integral matrices. Within this approach, we are able to naturally explain the appearance of multiple metastable states in inhomogeneous magnetic phases, and the use of some simplifying assumptions allows a complete analytic description of the thermodynamic properties of these states in the vicinity of phase transitions to be obtained. These results in turn make it possible to determine the parameters of arbitrary irreversible processes and, in particular, to obtain the first analytic expressions describing the shape of hysteresis loops at an arbitrary field amplitude and the field and temperature dependences of remanent magnetizations in magnets of the types specified above.

2. EFFECTIVE-FIELD APPROXIMATION FOR DISORDERED ISING MAGNETS

The Hamiltonian of a disordered Ising magnet has the form

$$\mathcal{H} = -\frac{1}{2}\sum_{i,j} J_{ij} S_i S_j, \tag{1}$$

where J_{ij} is the matrix of random exchange integrals and $S_i = \pm 1$. In most realistic models, J_{ij} is only nonzero if the distances between lattice sites $|\mathbf{r}_i - \mathbf{r}_j|$ do not exceed several radii of the nearest crystal lattice coordination spheres, and the J_{ij} value itself is bounded,

$$|J_{ij}| < J_{\max}.$$

Here, we will consider precisely such models. The effective-field approximation for Hamiltonian (1) reduces to replacing the J_{ij} matrix by the projector onto its maximum eigenvalue J,

$$J_{ij} \approx J \sum_{\alpha=1}^{N_0} c_i^\alpha c_j^\alpha. \tag{2}$$

Here, c_i^α are normalized eigenvectors of the J_{ij} matrix corresponding to the largest (generally, degenerate in multisublattice antiferromagnets) eigenvalue J and $\alpha = 1, \ldots, N_0$.

As a result, the Hamiltonian becomes a function of the (multicomponent) order parameter,

$$\eta^\alpha = N^{-1/2}\sum_{i=1}^{N} c_i^\alpha S_i,$$

and it only remains to calculate the entropy

$$S(\eta^\alpha) = \ln \mathrm{Tr}\, \delta_{N^{1/2}\eta^\alpha, \Sigma c_i^\alpha S_i}, \tag{3}$$

to obtain the nonequilibrium thermodynamic potential

$$F(\eta^\alpha) = -\frac{NJ}{2}\sum_{\alpha=1}^{N_0} (\eta^\alpha)^2 - TS(\eta^\alpha).$$

Minimizing F with respect to η^α allows us to obtain equilibrium thermodynamic parameter values that correspond to the deepest minimum and metastable state parameters that correspond to shallower minima. In particular, we can find the average spin values

$$\langle S_i \rangle = N^{1/2}\sum_{\alpha=1}^{N_0} c_i^\alpha \eta^\alpha. \tag{4}$$

The results obtained using this approximation may give a qualitative and, in certain instances (at fairly

large interaction radii or weak interactions of various modes), a fairly accurate quantitative description of the thermodynamics of second-order transitions in homogeneous magnets. We therefore believe that it would be natural to apply the effective-field approximation to models of disordered magnets with random exchange. It would then be sufficient to average the results obtained within the scheme described above over random J and c_i^α values. The thermodynamics of disordered magnets would then be qualitatively similar to the thermodynamics of ideal magnetic crystals. In particular, in inhomogeneous magnetic phases, multiple metastable states would be absent.

The most probable reason for the appearance of an exponentially large number of metastable states is the special structure of the spectrum and eigenvectors of the J_{ij} random matrix. Indeed, the eigenstates of the J_{ij} random matrix with the properties described above are localized in the vicinity of the upper and lower spectrum boundaries [18]. The description of a transition in a disordered magnet therefore requires that the maximum eigenvalue in (2) be replaced by a somewhat smaller eigenvalue J at the localization threshold; that is, the eigenvalue which is maximum among the eigenvalues with delocalized eigenvectors. This requirement arises because a macroscopic phase transition can only occur when the delocalized spin configuration undergoes condensation, whereas the condensation of local configurations with large eigenvalues (transition temperatures) that precedes the macroscopic transition corresponds to a special transition to the Griffiths state, which is not accompanied by noticeable anomalies of thermodynamic values [21, 22]. However, note that, according to [17], sequential condensation of local modes in a three-dimensional spin glass with a fairly large interaction radius can make delocalized modes stable, and no macroscopic transition then takes place. In this work, we assume that the interaction radius is sufficiently small for such a transition to nevertheless occur.

At the same time, it appears quite probable that, in most random realizations, the condensation of a single delocalized spin configuration with an eigenvalue closest to the localization threshold is insufficient for stabilizing the new phase. It is likely that, for dimension $d > 2$, the J_{ij} eigenvectors have a fairly loose (fractal) structure near the threshold and almost decompose into separate localization regions only connected by quasi-one-dimensional chains [19, 20]. In other words, there are many sites with the structure resembling that of the percolation cluster [23], where $N(c_i^\alpha)^2 \gg 1$, whereas $N(c_i^\alpha)^2 \ll 1$ on the other sites. This substantially distinguishes modes close to the localization threshold from modes within the continuous spectrum and eigenfunctions of translationally invariant J_{ij} matrices, for which $N(c_i^\alpha)^2 \approx 1$ on almost all lattice sites.

Evidence for the existence of such a fractal structure of modes at the localization threshold were obtained in many numerical studies of various statistical ensembles of short-range random matrices, see [19, 20]. According to (4), the condensation of one such mode results in the appearance of fairly high mean spin values only on a loose fractal structure, which is insufficient for stabilizing modes with lower eigenvalues, which are generally localized on other fractal sets of sites. More exactly, the condensation of the first delocalized mode with vector c_i^0 can only stabilize those modes with $J_\alpha < J$ that noticeably overlap with it; that is, such modes that $Nc_i^0 c_i^\alpha \gg 1$ for most of the sites where $N(c_i^\alpha)^2 \gg 1$. For this reason, the condensation of the mode closest to the localization threshold will, as temperature decreases, be followed by the condensation of another loose mode that almost does not overlap with the first one, next, a mode that almost does not overlap with the first and second modes undergoes condensation, etc.

Such a sequential condensation of nonoverlapping modes with decreasing eigenvalues occurs until fairly large mean spin values appear on almost all lattice sites. Generally, in the intervals between the eigenvalues of modes undergoing condensation, there can exist an arbitrary number of modes that do not condense because they substantially overlap with the already condensed ones. In the effective-field approximation, such modes are order parameter fluctuations, which can be ignored.

The fractal structure of modes undergoing condensation presupposes that their number should diverge as $N \longrightarrow \infty$. Indeed, if we assume that the sets of sites on which the modes under consideration are predominantly localized have fractal dimension $d_f < d$ (that is, the number of sites in these order regions is $N^{d_f/d}$), the number of condensed modes N_0 can be estimated at a value of the order of $N^{1-d_f/d}$.

The suggested mechanism allows us to easily explain the appearance of an exponentially large number of metastable states in inhomogeneous magnetic phases. Indeed, the condensation of one mode in a zero field results in the formation of two stable states related to global spin flip, and each subsequent condensation doubles their number. It follows that the condensation of N_0 modes results in the formation of $2^{N_0} \sim \exp(N^{1-d_f/d}\ln 2)$ stable states. Each such state is related to other via spin flips of independent groups of spins corresponding to fractal modes. Precisely this structure of spin configurations of ground states was observed in recent numerical studies of a three-dimensional spin glass model with binary random exchange [8], which is unambiguous evidence in favor of the mechanism under consideration.

As the eigenvectors of condensing modes c_i^α almost do not overlap (that is, $Nc_i^\alpha c_i^\beta \ll 1$ for most sites at $\alpha \neq \beta$), our analysis can be considerably simplified. The space of such c_i^α can be approximated by a set of nonoverlapping (normalized) e_i^α vectors ($e_i^\alpha e_i^\beta = 0$ for $\alpha \neq \beta$), which coincide with the c_i^α eigenvectors in the regions of their predominant localization [where $N(c_i^\alpha)^2 \gg 1$] and equal zero outside these regions. In the subspace spanned by (clearly, orthogonal) e_i^α vectors, J_{ij} is almost diagonal,

$$J_{ij} = \sum_{\alpha,\beta=1}^{N_0} (J\delta_{\alpha\beta} - J_{\alpha\beta}) e_i^\alpha e_j^\beta.$$

Here, $J_{\alpha\beta}$ ia a small nonnegative definite matrix,

$$|J_{\alpha\beta}| \ll J.$$

It is then simple to determine the effective-field thermodynamic potential, which depends on the multicomponent order parameter

$$l_\alpha = N_\alpha^{-1/2} \sum_i e_i^\alpha S_i, \quad N_\alpha \equiv \sum_i \theta(|e_i^\alpha|)$$

[$\theta(x)$ is the Heaviside step function] and the (quasi-)local magnetizations

$$m_\alpha = N_\alpha^{-1} \sum_i S_i \theta(|e_i^\alpha|).$$

The potential has the form

$$F = -\frac{1}{2} \sum_{\alpha,\beta=1}^{N_0} \sqrt{N_\alpha N_\beta} (J\delta_{\alpha\beta} - J_{\alpha\beta}) l_\alpha l_\beta - T \sum_{\alpha=1}^{N_0} N_\alpha S_\alpha(l_\alpha, m_\alpha) - NHm, \tag{5}$$

$$S_\alpha(l_\alpha, m_\alpha) = N_\alpha^{-1} \ln \mathrm{Tr}_\alpha \delta_{N_\alpha^{1/2} l_\alpha, \Sigma e_i^\alpha S_i} \delta_{N_\alpha m_\alpha, \Sigma S_i \theta(|e_i^\alpha|)}. \tag{6}$$

Here, Tr_α denotes summing over the spin configurations of those sites where $e_i^\alpha \neq 0$. Note that, according to the aforesaid, $N_\alpha \longrightarrow \infty$ as $N \longrightarrow \infty$ and

$$\sum_\alpha N_\alpha \approx N,$$

and the uniform magnetization is

$$m \approx \sum_\alpha \frac{N_\alpha}{N} m_\alpha.$$

Equation (5) includes the dependence on a small random matrix $J_{\alpha\beta}$ and random vectors e_i^α, the form of which is determined by the type of the J_{ij} ensemble of random matrices. In some instances, the form of e_i^α can be specified, as, e.g., in a spin glass with binary random exchange,

$$J_{ij} = \pm J_{\max},$$

where each configuration of bonds includes nonfrustrated d-dimensional clusters (that is, clusters characterized by a unique spin configuration σ_i corresponding to the energy minimum) and

$$J_{ij}\sigma_i \approx 2dJ_{\max}\sigma_i.$$

It follows that delocalized eigenvectors with the largest eigenvalues can approximately be constructed from local spin configurations through combining certain nonfrustrated d-dimensional clusters by branching chains without closed loops, which are also nonfrustrated at any bond configuration [24]. As a result, the form of e_i^α vectors will be

$$e_i^\alpha = N_\alpha^{-1/2} \sigma_i^\alpha, \tag{7}$$

where σ_i^α are the spin configurations constructed as described above on nonfrustrated fractal subsets of sites.

In dilute ferromagnets and antiferromagnets, vectors e_i^α can also be represented in form (7) through combining d-dimensional ferromagnetic (antiferromagnetic) clusters of the percolation cluster of bonds by chains. In a dilute ferromagnet, $\sigma_i^\alpha = 1$, and in a dilute antiferromagnet, $\sigma_i^\alpha = (-1)^{\mathbf{kr}_i}$. Note however that expressions for e_i^α can substantially differ from (7) for some actual bond configurations. We will nevertheless assume that, in most configurations, (7) fairly well approximates e_i^α and can be used to estimate the ensemble-average of the sums

$$u_{n\alpha} = N_\alpha^{n/2-1} \sum_i (e_i^\alpha)^n. \tag{8}$$

We will show that the thermodynamics in the vicinity of the transition to an inhomogeneous magnetic phase weakly depends on the detailed form of random vectors e_i^α and is only determined by several constants of form (8).

The advantage of representing F by (5) is entropy additivity. However, if $J_{\alpha\beta}$ is an arbitrary random

matrix, equation (5) is still difficult to analyze. To simplify it, we can use the observation that the eigenvalues of $J_{\alpha\beta}$ distributed between zero and some $J_0 \ll J$ should condense to zero, because eigenvectors nonoverlapping with all preceding eigenvectors become rarer as the distance to the localization threshold increases. As the mean interval between neighboring eigenvalues is of the order of $1/N$, small eigenvalues of $J_{\alpha\beta}$ may be of the same order. This allows $J_{\alpha\beta}$ to be approximately replaced by the projector to some (random) r_α vector whose properties are determined by the type of the ensemble of random matrices J_{ij},

$$J_{\alpha\beta} = J_0 r_\alpha r_\beta, \quad \sum_{\alpha=1}^{N_0} r_\alpha^2 = 1. \tag{9}$$

Generally, the assumption that a single eigenvalue of the $J_{\alpha\beta}$ matrix is finite, and all the others are of the order of $1/N$, is fairly rough. This assumption leads to the coincidence of the points of condensation of all modes except one and makes the transition sharper, whereas, in reality, the condensation of modes can take place in the whole interval from $T = J$ to $T = J - J_0$. At the same time, such an assumption allows a fairly simple analytic description to be obtained for the thermodynamics of all metastable states. This description qualitatively agrees with experiment and can be used as a starting point for constructing a more accurate theory taking into account the distribution of condensation temperatures for fractal modes.

We will show that the form of r_α in (9) can be found taking into account that the $J_{\alpha\beta}$ matrix determines the type of the ground state of the crystal. As a result, thermodynamic potential (5) takes the form

$$F = -\frac{J}{2}\sum_{\alpha=1}^{N_0} N_\alpha l_\alpha^2 + \frac{J_0}{2}\left(\sum_{\alpha=1}^{N_0}\sqrt{N_\alpha} r_\alpha l_\alpha\right)^2 - T\sum_{\alpha=1}^{N_0} N_\alpha S_\alpha(l_\alpha, m_\alpha) - NHm. \tag{10}$$

The partial entropies (6) can be written as

$$S_\alpha(l_\alpha, m_\alpha) = \ln 2 - \max_{\varphi,\psi}\left[\varphi m_\alpha + \psi l_\alpha - N_\alpha^{-1/2}\sum_i \theta(|e_i^\alpha|)\ln\cosh(\varphi + \psi e_i^\alpha\sqrt{N_\alpha})\right],$$

where the φ_α and ψ_α values corresponding to the maximum are found from the equations

$$m_\alpha = N_\alpha^{-1}\sum_i \tanh(\varphi_\alpha + \psi_\alpha e_i^\alpha\sqrt{N_\alpha}),$$
$$l_\alpha = N_\alpha^{-1}\sum_i e_i^\alpha \tanh(\varphi_\alpha + \psi_\alpha e_i^\alpha\sqrt{N_\alpha}). \tag{11}$$

Differentiating potential (10) with respect to l_α and m_α yields the equation of state

$$J_0 N_\alpha^{-1/2} r_\alpha \sum_\beta N_\beta^{1/2} r_\beta l_\beta - J l_\alpha + T\psi_\alpha = 0,$$
$$T\varphi_\alpha = H. \tag{12}$$

Stable solutions to (11) and (12) corresponding to F minima are determined by the condition of positive definiteness of the matrix

$$G_{\alpha\beta} = \delta_{\alpha\beta}\left\{T\left[1 - \sum_i (e_i^\alpha)^2 \times \tanh^2(\varphi_\alpha + \psi_\alpha e_i^\alpha\sqrt{N_\alpha})\right]^{-1} - J\right\} + J_0 r_\alpha r_\beta. \tag{13}$$

At $H = 0$ and $T = 0$, it follows from (11) and (12) that

$$m_\alpha = N_\alpha^{-1}\sum_i \mathrm{sgn}(e_i^\alpha l_\alpha), \quad |l_\alpha| = N_\alpha^{-1/2}\sum_i |e_i^\alpha|.$$

We therefore have 2^{N_0} stable solutions to (11) and (12) that differ in signs of l_α.

If the ensemble of random J_{ij} matrices is such that, in the majority of realizations, the transition to the state with $m = 0$ occurs, than the energy for states with $m = 0$ can be minimized by setting $\sum_{\alpha=1}^{N_0}\sqrt{N_\alpha} r_\alpha l_\alpha = cm$. At arbitrary signs of l_α, this condition is satisfied by only one r_α vector,

$$r_\alpha = c'\sum_i \mathrm{sgn}(e_i^\alpha)/\sum_i |e_i^\alpha|. \tag{14}$$

Here, c' is the normalization constant.

We will confine our consideration to disordered magnets with $m = 0$ in the ground state, such as dilute antiferromagnets, spin glasses, and dilute ferromagnets with dipole interactions [25]. By virtue of the properties of random vectors e_i^α in dilute magnets and binary spin glasses discussed above, (14) can be written in the form [see (7) and (8)]

$$r_\alpha = u_{1\alpha}\left(\frac{N_\alpha}{\sum_\beta N_\beta u_{1\beta}^2}\right)^{1/2}. \tag{15}$$

To summarize, a study of metastable states in magnets of the types specified above reduces to the search for stable solutions of equations of state (11) and (12) with r_α in form (15). The spin configurations corresponding to the l_α and m_α values obtained are determined by the equation

$$\langle S_i \rangle_T = \sum_\alpha N_\alpha^{-1/2} e_i^\alpha \frac{l_\alpha - u_{1\alpha} m_\alpha}{1 - u_{1\alpha}^2} + \sum_\alpha N_\alpha^{-1} \theta(|e_i^\alpha|) \frac{m_\alpha - u_{1\alpha} l_\alpha}{1 - u_{1\alpha}^2}. \quad (16)$$

Averaging the solutions over disorder reduces to averaging over random e_i^α vectors and over $J_0 \ll J$. Note that the localization threshold J is not a random value and characterizes the whole ensemble of random matrices J_{ij}.

The smallness of $J_0 > 0$ means that the corresponding distribution function should have a fairly narrow bounded support; that is, the possible J_0 values should be smaller than some $\bar{J} > 0$ satisfying the condition $\bar{J} \ll J$. In contrast to the case of the sums (8) of a macroscopic number of terms, there is no reason to consider J_0 fluctuations self-averaging; that is, to assume that $\langle J_0^k \rangle \longrightarrow \langle J_0 \rangle^k$ as $N \longrightarrow \infty$. The thermodynamic parameters of metastable and equilibrium states in an inhomogeneous magnetic phases will therefore generally be non-self-averaging values, being determined by different J_0 values in different samples. Note that the absence of self-averaging equilibrium parameters was also observed in numerical studies of disordered magnetic models [1, 26].

3. THERMODYNAMICS IN THE VICINITY OF THE TRANSITION

In the absence of a field, equations of state (11) and (12) have a unique paramagnetic solution at $T > J$ and many stable solutions at $T < J$. It follows that, at $T = J$ and $H = 0$, the transition from the paramagnetic to inhomogeneous magnetic phase occurs.

Consider the thermodynamics of the model in the vicinity of these transitions. It is determined by the condition

$$l_\alpha, m_\alpha \ll 1. \quad (17)$$

Equations (11), (12), and (15) then give equations for the magnetizations of condensed modes m_α,

$$\tau m_\alpha + \frac{\tau_0 u_{1\alpha}^2 m N}{\sum_\beta N_\beta u_{1\beta}^2} + \frac{u_{4\alpha} m_\alpha^3}{3u_{1\alpha}^2} = \frac{u_{1\alpha}^2 H}{J}, \quad (18)$$

and l_α can be written through m_α,

$$u_{1\alpha} l_\alpha = m_\alpha + \frac{(u_{1\alpha}^2 - 1)H}{J} + \frac{(u_{3\alpha} - u_{1\alpha} u_{4\alpha}) m_\alpha^3}{3u_{1\alpha}^3}. \quad (19)$$

Here, $\tau = 1 - J/T$ and $\tau_0 = J_0/J > 0$.

Gaussian (13) can be rewritten as

$$T^{-1} G_{\alpha\beta} = \left(\tau + \frac{u_{4\alpha} m_\alpha^2}{u_{1\alpha}^2} \right) \delta_{\alpha\beta} + \tau_0 r_\alpha r_\beta.$$

It follows from (17) and (18) that $H \ll J$, $\tau \ll 1$, and $\tau_0 \ll 1$. The last condition also follows from the smallness of τ.

It is natural to assume that the sums of a macroscopic number of terms (8) are self-averaging and can be replaced by the mean values. On the assumption that, for the majority of random realizations, approximation (7) is valid on the majority of sites, we obtain

$$\bar{u}_{4\alpha} = 1, \quad \bar{u}_{3\alpha} = \bar{u}_{1\alpha}.$$

Let us also assume that the $\bar{u}_{1\alpha}^2$ constants are independent of mode numbers α; that is,

$$\bar{u}_{1\alpha}^2 = \bar{u}_1^2 \equiv N_0^{-1} \sum_\alpha \bar{u}_{1\alpha}^2. \quad (20)$$

Generally, this approximation cannot qualitatively change the results. Abandoning it would only lead to some fluctuations in the solutions. At the same time, it allows equations (18) to be substantially simplified and makes them open to analytic treatment.

The $\bar{u}_1^2$ constant can be estimated from equation (7). It follows from (7) and (8) that

$$\bar{u}_1^2 = N_0^{-1} \sum_\alpha (\nu_\alpha^+ - \nu_\alpha^-)^2,$$

where ν_α^+ and ν_α^- are the fractions of positive and negative e_i^α values, and we always have $\bar{u}_1^2 \le 1$. In a dilute dipole ferromagnet, $\bar{u}_1^2 = 1$. In a dilute antiferromagnet, the $\nu_\alpha^+ - \nu_\alpha^-$ difference can be nonzero only because of noncompensated spins on the surface of d-dimensional antiferromagnetic clusters, on which e_i^α are predominantly localized. For this reason, $\nu_\alpha^+ - \nu_\alpha^-$ is of the order of the ratio between the number of sites in d-dimensional clusters and the number of sites on their surface; that is,

$$\bar{u}_1^2 \approx D^{-2},$$

where D is the mean diameter of d-dimensional clusters (in lattice parameter units). Clearly, D is a function of the concentration of antiferromagnetic atoms and tends to infinity when the concentration tends to one.

The $\bar{u}_1^2$ value in a binary spin glass depends on the concentration of ferromagnetic bonds p; $\bar{u}_1^2 = 1$ for $p > 1 - p_c$ and $\bar{u}_1^2 = D^{-2}$ for $p < p_c$, where p_c is the threshold of percolation over bonds on a lattice of magnetic atoms. At $p_c < p < 1 - p_c$, the dependence of $\bar{u}_1^2$ on p can qualitatively be described by the formula

$$\bar{u}_1^2 = \frac{p - p_c + D^{-2}(1 - p - p_c)}{1 - 2p_c}.$$

We will also replace N_α by their average values

$$N_\alpha = N/N_0.$$

The introduction of reduced magnetizations μ_α,

$$\mu_\alpha = m_\alpha/\bar{u}_1,$$

in place of m_α and the use of dimensionless field $h = \bar{u}_1 H/J$ allows (18) to be written as

$$\tau\mu_\alpha + \tau_0\mu + \frac{\mu_\alpha^3}{3} = h, \tag{21}$$

where

$$\mu = N_0^{-1}\sum_\alpha \mu_\alpha.$$

At $\tau > 0$, Eqs. (21) have a single paramagnetic solution with equal $\mu_\alpha = \mu$. Let us denote this solution by μ_0. It satisfies the equation

$$(\tau + \tau_0)\mu_0 + \frac{\mu_0^3}{3} = h. \tag{22}$$

At $\tau < 0$, up to $2^{N_0} - 2$ stable inhomogeneous solutions can exist in addition to μ_0. These solutions have the form

$$\mu_\alpha = \sqrt{-\tau}(\sin\varphi + \sqrt{3}\sigma_\alpha\cos\varphi),$$

where $\sigma_\alpha = \pm 1$, and $\varphi = \varphi(\tau, \tau_0, h, \Delta)$ is the solution to the equation

$$3\tau_0(\sqrt{3}\Delta\cos\varphi + \sin\varphi) - 2\tau\sin 3\varphi = 3h|\tau|^{-1/2}, \tag{23}$$

$$\Delta = \sum_\alpha \sigma_\alpha/N_0. \tag{24}$$

The Δ parameter varies in the interval $(-1, 1)$ and determines the degree of inhomogeneity of the metastable state. The $\Delta = \pm 1$ values correspond to a paramagnetic state with $\mu_\alpha = \mu$. All states with equal Δ have equal uniform magnetization values,

$$\mu = \sqrt{-\tau}(\sin\varphi + \sqrt{3}\Delta\cos\varphi), \tag{25}$$

Edwards–Anderson parameters,

$$q = N^{-1}\sum_i \langle S_i\rangle_T^2 - m^2 \approx N_0^{-1}\sum_\alpha \mu_\alpha^2 - \bar{u}_1^2\mu^2 = 3\tau(\Delta^2 - 1)\cos^2\varphi + (1 - \bar{u}_1^2)\mu^2, \tag{26}$$

and also the same thermodynamic potential

$$\frac{4F}{TN} = (\tau_0 + \bar{u}_1^2\tau)\mu^2 + \tau q - 3h\mu - 4\ln 2. \tag{27}$$

These states are stable at $\tau + \mu_\alpha^2 > 0$, which is equivalent to the inequality

$$|\varphi| < \pi/6. \tag{28}$$

As $\tau_0 > 0$, the left-hand side of (23) is a monotonically increasing function of φ at $|\varphi| < \pi/6$, and there is therefore only one stable solution for φ at a given Δ value, which exists in the interval of fields $h_- < h < h_+$,

$$h_\pm = \sqrt{3}h_{AT}\frac{\Delta}{2} \pm h_c, \tag{29}$$

$$h_{AT} = \sqrt{-3\tau}\tau_0, \quad h_c = \sqrt{-\tau}\left(\frac{\tau_0}{2} - \frac{2\tau}{3}\right). \tag{30}$$

In this interval, solution (23) is fairly well approximated by the quadratic field function

$$\varphi \approx \frac{\pi}{12h_c}\left[2h - \sqrt{3}h_{AT}\Delta - \frac{4(2 - \sqrt{3})\Delta h_{AT}(h_+ - h)(h - h_-)}{4h_c^2 - (2 - \sqrt{3})^2\Delta^2 h_{AT}^2}\right], \tag{31}$$

which gives exact $\varphi(h_\pm) = \pm\pi/6$ and $\varphi(\Delta h_{AT}) = 0$ values.

It follows from stability condition (27) and equalities (23) and (25) that metastable states are stable in the region

$$9(\tau_0\mu - h)^2 < -4\tau^3,$$

which is a band on the μ, h plane. Magnetization is a monotonically increasing function of h and Δ within this band, and the field dependences of magnetization are a family of nonintersecting curves bounded from above and below by the $\mu_0(h)$ curve, as is shown in Fig. 1.

Clearly, the region of the existence of metastable states in Fig. 1 determines the shape of the hysteresis loop, which appears as a response to a slowly varying external field, whose amplitude exceeds

$$h_e = h_c + \frac{\sqrt{3}h_{AT}}{2}. \tag{32}$$

The shape of the loop changes as temperature varies, because, at $-2\tau < 3\tau_0$, only a part of the metastable states are stable at $h = 0$ (see Fig. 1a), whereas, at $3\tau_0 < -2\tau$, all metastable states are stable at $h = 0$ (Fig. 1b). Note that if the field amplitude is smaller than h_e, the shape of the hysteresis loop is determined by the field dependences of magnetization in the corresponding

metastable states. Consider the equations for (dimensionless) magnetic susceptibility $\chi = \partial\mu/\partial h$,

$$\chi^{-1} = \tau_0 - \frac{2\tau\cos 3\varphi}{\cos\varphi - \sqrt{3}\Delta\sin\varphi},$$

and specific entropy S and specific heat capacity C,

$$S = \ln 2 - \frac{q + \bar{u}_1^2\mu^2}{2},$$

$$C = 1 + \chi\left[\frac{3}{2}\frac{\tau_0(1-\Delta^2)}{1-\sqrt{3}\Delta\tan\varphi} - \tau_0 - \tau\right].$$

These equations give an idea of the character of field and temperature dependences of the thermodynamic values in various metastable states. For instance, at the boundaries of the stability region at $h = h_\pm$, q and χ^{-1} have the smallest values,

$$q = \frac{9}{4}\tau(\Delta^2 - 1) - \frac{1}{4}(1 - \bar{u}_1^2)\tau(3\Delta \pm 1)^2, \qquad \chi^{-1} = \tau_0, \tag{33}$$

and the magnetization, entropy, and heat capacity are given by

$$\mu = \frac{\sqrt{-\tau}}{2}(3\Delta \pm 1), \quad S = \ln 2 + \frac{\tau}{4}(5 \pm 3\Delta), \qquad C = \frac{3}{2}(1 \pm \Delta) - \frac{\tau}{\tau_0}. \tag{34}$$

As $|h|$ approaches h_e given by (32), more and more uniform states with $\Delta \longrightarrow \pm 1$ remain stable and their magnetization tends to $\mu_0(\pm h_e) = \pm 2\sqrt{-\tau}$. The magnetic susceptibilities and heat capacities of these states, however, tend to limits different from the corresponding values in the paramagnetic state,

$$\chi_0^{-1} = \tau + \tau_0 + \mu_0^2, \quad C_0 = \frac{\mu_0^2}{\tau + \tau_0 + \mu_0^2}.$$

In the middle of the stability band at $\varphi = 0$ or $h = \Delta h_{AT}$, we have

$$\mu = \Delta\sqrt{-3\tau}, \quad q = -3\tau(1 - \bar{u}_1^2\Delta^2), \quad \chi^{-1} = \tau_0 - 2\tau,$$

$$S = \ln 2 + \frac{3\tau}{2}, \quad C = \frac{3}{2}\left(1 - \Delta^2\frac{\tau_0}{\tau_0 - 2\tau}\right).$$

As inhomogeneity decreases when $\Delta \longrightarrow \pm 1$ or $h \longrightarrow \pm h_{AT}$, the μ, χ, S, and C values tend to their values for the paramagnetic phase.

The Almeida–Thouless h_{AT} field (30) determines the phase transition point to the paramagnetic phase (accurate to terms of the order of $1/N_0$). To show this, let us find the Δ_{eq} values corresponding to the deepest states.

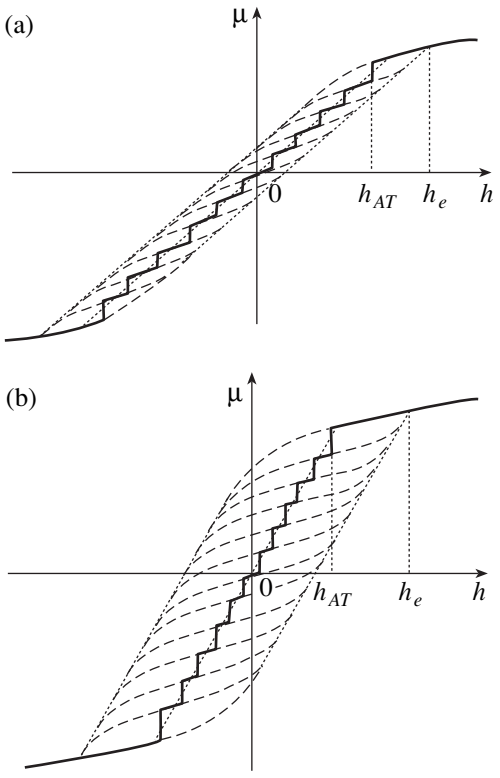


Fig. 1. Field dependences of magnetization of metastable (dashed lines) and stable (solid lines) states in the vicinity of the transition for (a) $2\tau < 3\tau_0$ and (b) $3\tau_0 < -2\tau$.

Differentiating potential F (27) with respect to Δ and using (23), (25), and (26), we obtain

$$\frac{\partial F}{\partial\Delta} = -NT\sin\varphi\cos^3\varphi, \quad \left.\frac{\partial^2 F}{\partial\Delta^2}\right|_{\varphi = 0} > 0.$$

It follows that the smallest potential value is characteristic of states with such Δ for which $\varphi = 0$ at given τ and h. Clearly, solutions to (23) with $\varphi = 0$ exist for $\Delta = h/h_{AT}$, which is only possible if $h^2 < h_{AT}^2$. If $h^2 > h_{AT}^2$, $F(\Delta)$ has no minimum in the domain of definition of $\Delta^2 < 1$, and a minimum is attained at its boundaries at $\Delta_{eq} = \operatorname{sgn} h$. It follows that transitions to the paramagnetic state occur at $h = \pm h_{AT}$.

As Δ is a rational value of the form $2n/N_0 - 1$ [see (24)], Δ cannot exactly equal h/h_{AT} at all $h^2 < h_{AT}^2$. For this reason, Δ_{eq} corresponding to the smallest F value is determined by the condition of a minimum of $|\Delta - h/h_{AT}|$ and can be represented in the form

$$\Delta_{eq} = \sum_{n=1}^{N_0 - 1}\left(\frac{2n}{N_0} - 1\right)\theta(N_0^{-2} - \varepsilon_n^2) + (\operatorname{sgn} h)\theta\left[h^2 - \left(\frac{N_0 - 1}{N_0}\right)^2 h_{AT}^2\right],$$

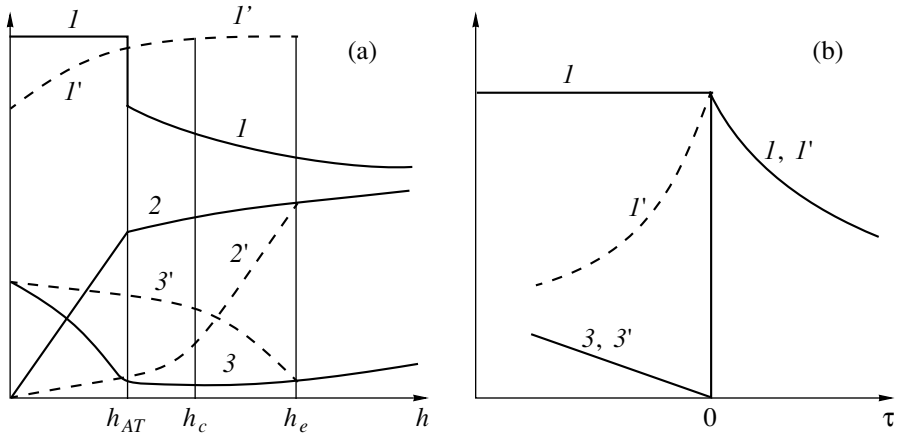


Fig. 2. (a) Field and (b) temperature dependences (at $h = 0$) of thermodynamic parameters in the FC and ZFC modes: (*1*) FC susceptibility, (*1*') ZFC susceptibility, (*2*) FC susceptibility, (*2*') ZFC susceptibility, (*3*) FC Edwards–Anderson parameter, and (*3*') ZFC Edwards–Anderson parameter.

where

$$\varepsilon_n \equiv \frac{h}{h_{AT}} - \frac{2n}{N_0} + 1.$$

It follows that, at $h^2 < h_{AT}^2$, a series of first-order transitions occur between inhomogeneous states in the fields

$$h_n = h_{AT}\left(\frac{2n+1}{N_0} - 1\right).$$

The corresponding φ_{eq} value at $h^2 < h_{AT}^2$ is

$$\varphi_{eq} = \frac{\sqrt{3}\tau_0}{\tau_0 - 2\tau}\sum_{n=1}^{N_0-1}\varepsilon_n\theta(N_0^{-2} - \varepsilon_n^2).$$

Substituting Δ_{eq} and φ_{eq} into (25) and (26) yields equilibrium μ_{eq} and q_{eq} values for $N_0 \longrightarrow \infty$,

$$\mu_{eq} = \frac{h}{\tau_0}\theta(h_{AT}^2 - h^2) + \mu_0\theta(h^2 - h_{AT}^2),$$

$$q_{eq} = -3\tau\left(1 - \frac{h^2}{h_{AT}^2}\right)\theta(h_{AT}^2 - h^2) + (1 - \bar{u}_1^2)\mu_{eq}^2.$$

Differentiating μ_{eq} with respect to h, we obtain

$$\chi_{eq} = \tau_0^{-1}\theta(h_{AT}^2 - h^2) + (\tau + \tau_0 + \mu_0^2)^{-1}\theta(h^2 - h_{AT}^2).$$

The equilibrium entropy is found by differentiating the equilibrium potential, which, accurate to terms of the order of ε_n^2, has the form

$$F_{eq} = F\left(\Delta = \frac{h}{h_{AT}}\right) - TS_{conf},$$

where configurational entropy S_{conf} is determined by the logarithm of the number of states with equal potentials F,

$$S_{conf} = N^{-1}\ln\begin{pmatrix} N_0 \\ N_0(1-\Delta_{eq})/2 \end{pmatrix}.$$

The S_{conf} value is of the order of N_0/N and can be ignored. For $N_0 \longrightarrow \infty$, this gives

$$S_{eq} = \ln 2 + \frac{3\tau}{2}\theta(h_{AT}^2 - h^2) - \frac{\mu_0^2}{2}\theta(h^2 - h_{AT}^2).$$

The equilibrium heat capacity is given by

$$C_{eq} = \frac{3}{2}\theta(h_{AT}^2 - h^2) - \frac{\mu_0^2}{\tau + \tau_0 + \mu_0^2}\theta(h^2 - h_{AT}^2).$$

Note that μ_{eq}, q_{eq}, and S_{eq} are continuous at $h^2 = h_{AT}^2$, whereas χ_{eq} and C_{eq} experience jumps in the transition to the paramagnetic phase.

Note also that the mean equilibrium parameters are, generally, unobservable values because of the presence of macroscopic barriers separating metastable states. Experimental values close to these parameters can, probably, be obtained by cooling the system in a low constant field (field cooled mode, FC) at temperatures close to the transition temperature [1, 3], when the barriers between metastable states are comparatively low, and the system has sufficient time to attain the deepest state (or a state close to it) if cooled fairly slowly. When the system is cooled in the absence of a field to some temperature at which the field is switched on (zero field cooled mode, ZFC), the observed values differ from equilibrium, because the system then reaches one of the states with $\Delta = 0$ and remains in this state at $h < h_c$ (30) (see Fig. 1), and the results of measurements corre-

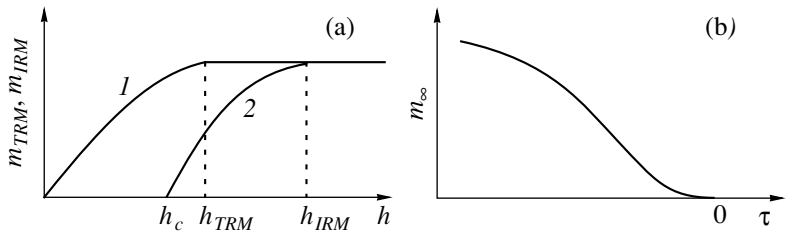


Fig. 3. (a) Field dependences of m_{TRM} and m_{IRM} and (b) temperature dependence of saturation magnetization.

spond to this state. Their values are obtained from the general formulas at $\Delta = 0$ and $\varphi = \pi h/6h_c$[see (31)].

If $h_c < h < h_e$, the system relaxes to a state at the boundary of stability region (on the lower branch of the hysteresis loop) with

$$\Delta_{ZFC} = \frac{2}{\sqrt{3}}\frac{h - h_c}{h_{AT}}.$$

Substituting $\Delta = \Delta_{ZFC}$ into (33) and (34) (with sign +) gives thermodynamic parameter values to which the values observed in the ZFC mode relax at $h_c < h < h_e$. Lastly, at $h > h_e$, the ZFC parameters correspond to the paramagnetic state. The field and temperature dependences of the thermodynamic parameters in the FC and ZFC modes are shown in Fig. 2.

Similarly, proceeding from the intervals of the existence of metastable states (see Fig. 1) and their parameters, we can find values determined in other arbitrary nonequilibrium processes in inhomogeneous magnetic phases. For instance, we can find the thermoremanent magnetization (m_{TRM}), which persists when the field is switched off after an FC process, and isothermal remanent magnetization (m_{IRM}), which persists after cooling in zero field and subsequent switching on of a field for a finite time interval (longer than the time of intravalley relaxation) [1, 3]. For instance, it is clear that m_{IRM} is nonzero only if $h > h_c$, and the expression for this value can be obtained from (25) at

$$\Delta_{IRM} = \min\left(1, \frac{2h_c}{\sqrt{3}h_{AT}}, \frac{2}{\sqrt{3}}\frac{h - h_c}{h_{AT}}\right),$$

$$\varphi_{IRM} \equiv \varphi(h = 0, \Delta_{IRM}) \approx -\frac{\sqrt{3}\pi h_{AT}\Delta_{IRM}}{12h_c}.$$

The m_{TRM} value can also be obtained from (25) by substituting

$$\Delta_{TRM} = \min\left(1, \frac{2h_c}{\sqrt{3}h_{AT}}, \frac{h}{h_{AT}}\right),$$

$$\varphi_{TRM} \equiv \varphi(h = 0, \Delta_{TRM}) \approx -\frac{\sqrt{3}\pi h_{AT}\Delta_{TRM}}{12h_c}.$$

The field dependences of m_{TRM} and m_{IRM} are shown in Fig. 3a. At

$$h > h_{TRM} = \min\left(\frac{2h_c}{\sqrt{3}}, h_{AT}\right)$$

m_{TRM} becomes field-independent, and at

$$h > h_{IRM} = \min(2h_c, h_e)$$

m_{IRM} also becomes constant and coincides with m_{TRM}. The saturation magnetization in strong fields is given by

$$\mu_\infty = \frac{2}{3\tau_0}\theta(3\tau_0 + 2\tau)(-\tau)^{3/2}$$
$$+ [-3(\tau_0 + \tau)]^{1/2}\theta(-3\tau_0 - 2\tau)$$

for both values. Its temperature dependence is shown in Fig. 3b.

4. CONCLUSION

The most important result of this work is a complete, although qualitative, description of the properties of all metastable states in magnetic phases of disordered Ising magnets with $m = 0$ in the ground state. We also showed how these properties are related to the parameters of slow nonequilibrium processes. The results given in Fig. 1 allow the description of an arbitrary fairly slow nonequilibrium process with all sequences of heating, cooling, and field variation imaginable. Qualitative agreement of the parameters of some such processes obtained in this work with experiment and numerical calculations performed in [1–4] is evidence that the approximations used in (7), (9), (15), and (20) are justified and that the condensation of a macroscopic number of fractal modes existing near the localization threshold can indeed be the physical mechanism of the appearance of an exponentially large number of metastable states. Note once more that the structure of the set of ground states observed in numerical calculations of the 3*d* spin glass model with binary exchange [8] also lends support to the suggested mechanism.

It should also be stressed that all results obtained within the framework of the accepted approximations are expressed through the statistic characteristics of the

matrix of random exchange interactions and contain no phenomenological parameters. The suggested approach can therefore be a starting point for constructing a more accurate quantitative theory of the properties of metastable states of disordered magnets. Such a theory should rely on detailed studies of the properties of the eigenvectors of random matrices J_{ij} near the localization threshold, which, are of necessity described at the level of phenomenological assumptions in this work. The goal of such a quantitative theory may, in particular, be testing the universality of the properties of all disordered magnets with $m = 0$ in the ground state, the results for which, obtained using the accepted approximations, only differ by $\bar{u}_1^2$ constant values.

ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research (project no. 98-02-18069).

REFERENCES

1. K. Binder and A. P. Young, Rev. Mod. Phys. **58**, 801 (1986).
2. R. B. Stinchcomb, in *Phase Transitions and Critical Phenomena*, Ed. by C. Domb and J. L. Lebowitz (Academic, London, 1983), Vol. 7.
3. C. Y. Huang, J. Magn. Magn. Mater. **51**, 1 (1985).
4. C. Ro, G. S. Grest, and C. M. Soukoulis, Phys. Rev. B **31**, 1682 (1985).
5. P. Sibani, E-print archives, cond-mat/9706232 (1997).
6. S. Bastea and P. M. Duxbury, E-print archives, cond-mat/9801108 (1998).
7. A. K. Hartmann, E-print archives, cond-mat/9902120 (1999).
8. G. Hed, A. K. Hartmann, D. Stauffer, and E. Domany, E-print archives, cond-mat/0007356 (2000).
9. A. A. Likhachev, E-print archives, cond-mat/0007504 (2000).
10. J. R. L. de Almeida and D. J. Thouless, J. Phys. A **11**, 983 (1978).
11. G. Parisi, J. Phys. A **13**, L115 (1980); **13**, 1101 (1980); **13**, 1807 (1980).
12. H. Sompolinsky, Phys. Rev. Lett. **47**, 935 (1981).
13. C. de Dominicis and Y. Y. Goldschmidt, Phys. Rev. B **41**, 2184 (1990).
14. Crisanti and H.-J. Sommers, Z. Phys. B **87**, 341 (1992).
15. D. J. Thouless, P. W. Anderson, and R. G. Palmer, Philos. Mag. **35**, 593 (1977).
16. A. J. Bray and M. A. Moore, J. Phys. C **13**, L469 (1980).
17. L. B. Ioffe and M. V. Feĭgel'man, Zh. Éksp. Teor. Fiz. **89**, 654 (1985) [Sov. Phys. JETP **62**, 376 (1985)].
18. I. M. Lifshitz, S. A. Gredescul, and L. A. Pastur, *Introduction to the Theory of Disordered Systems* (Wiley, New York, 1988).
19. M. Janssen, Int. J. Mod. Phys. B **8**, 943 (1994); Phys. Rep. **295**, 1 (1998).
20. F. Evers and A. D. Mirlin, E-print archives, cond-mat/0001086 (2000).
21. R. B. Griffiths, Phys. Rev. Lett. **23**, 17 (1969).
22. A. J. Bray, Phys. Rev. Lett. **59**, 586 (1987).
23. D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor and Francis, London, 1992).
24. A. Coniglio, F. di Liberto, G. Montroy, and F. Peruggi, Phys. Rev. B **44**, 12605 (1991).
25. M. J. Stephen and A. Aharony, J. Phys. C **14**, 1665 (1981).
26. S. Wiseman and E. Domany, E-print archives, cond-mat/9802095 (1998).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 1049–1059.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1210–1221.

SOLIDS
Electronic Properties

On the Confinement Potential Formation in a Two-Electron Quantum Dot

M. Dineykhan[a], S. A. Zhaugasheva[a], and R. G. Nazmitdinov[b, *]

[a]*Al-Farabi State University of Kazakhstan, ul. Vinogradova 95, Almaty, 480012 Kazakhstan*

[b]*Bogoliubov Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia*
**e-mail: rashid@thsun1.jinr.ru*

Received August 17, 2000

Abstract—A model of a quantum dot for two interacting electrons is proposed and analyzed. The properties of the ambient determining the form of the confinement potential for electrons are simulated using the electrostatic field of the image charge. Analytic expressions for the eigenvalues of each subsystem are derived taking into account the external magnetic field and using the representation of the system Hamiltonian as the sum of the Hamiltonians of the center of mass and of relative motion on the basis of the method of oscillator representation [M. Dineykhan and G. V. Efimov, Element. Chast. At. Yadra **26**, 651 (1995); M. Dineykhan, G. V. Efimov, G. Ganbold, and S. N. Nedelko, *Oscillator Representation in Quantum Physics*, Lecture Notes in Physics, Vol. 26, Springer, Berlin (1995)]. The relative motion of electrons is responsible for a confinement potential which differs from the parabolic confinement potential and is a function of the electron effective mass as well as the characteristics of the image charge. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Advances in modern technology make it possible to create semiconducting nanostructures (quantum dots) [1–3] in which a finite number of electrons are "captured" in a bounded volume. A quantum dot is associated with a quantum well formed at the interface between two finite-size semiconductors (such as GaAs and GaAlAs) owing to different positions of the forbidden gaps on the energy scale in these semiconductors. Besides, external contacts which make it possible to control the properties of a quantum well also play an important role. The presence of discrete energy levels and even the manifestations of the shell structure which was predicted [4, 5] and experimentally observed [6] for quantum dots give grounds for treating them as artificial atoms. The possibility of monitoring and controlling the properties of quantum dots attracts considerable attention to these objects, which may be used, among other things, as a new elemental basis for future generations of computers [2].

Two-electron quantum dots are the simplest example of such systems which can be used to trace the peculiarities of even more intricate complexes.

Using the hypothesis that the effective electron confinement potential in a quantum dot corresponds to the parabolic confinement potential, it is possible to describe [7–9] the typical features of transport processes [10] and spin oscillations in the ground state of a quantum dot in a magnetic field [11]. However, electron correlations may strongly affect the form of the confinement potential under certain experimental conditions. For example, the description of experiments involving photoemission [12] in a quantum dot requires the introduction of anharmonic corrections [13] to the parabolic confinement potential. A natural question arises concerning the mechanism of the confinement potential formation in a quantum dot.

What must the conditions be for the formation of, say, a parabolic confinement potential in a quantum dot? Which parameters of the system may lead to the suppression of anharmonism or, on the contrary, give rise to these effects?

The present work aims at analyzing the mechanism of formation of the confinement potential in a two-electron quantum dot in a constant external magnetic field. The main components responsible for the formation of the confinement potential in a quantum dot in the so-called Coulomb-blockade regime include the external voltage applied to the layered nanostructure and the properties of contacts having various geometries and connecting the quantum dot with the ambient [2, 3]. We proceed from the fact that quantum-mechanical effects play a significant role in the description of the mechanism of quantum dot formation. Further, we assume that a quantum well is homogeneous in its dielectric properties, while the system on the whole is inhomogeneous, and the continuity conditions for the tangential derivatives of potentials must hold. These assumptions lead to the introduction of an effective positive image charge which is associated with external factors.

This approach is well known in analyzing the properties of insulators in electrostatics [14]. Thus, we assume that the image potential, which is created, for one, due to a considerable difference in the permittivi-

ties of the layers forming the quantum dot (e.g., a vacuum and a semiconductor or a semiconductor and an insulator), plays a significant role in the formation of the confinement potential.[1]

We will consider a three-body Coulomb system in an external magnetic field, which is formed by two electrons and an image charge. It should be noted that in our formulation of the problem, the image charge may also be associated with an impurity in the quantum dot.

Our analysis is based on the oscillator representation method [16], which was successfully applied for calculating the energy spectrum of the system controlled by the Coulomb and power potentials as well as by the Coulomb potential and the Yukawa potential [17]. The material layout is as follows. In Section 2, the model Hamiltonian of the three-body system, which can be divided into the Hamiltonian of the center-of-mass system and the Hamiltonian of relative motion, is considered. In Section 3, the Hamiltonian of relative motion is analyzed. Examples of calculation of the energy spectrum of a two-electron system in the 2D case on the basis of the results obtained in Section 3 are given in Section 4. The main results are summarized in the Conclusion. Some technical details of the calculations based on the oscillator representation method are given in the Appendix.

2. FORMULATION OF THE PROBLEM

A quantum dot containing a few electrons may be treated as a model of an atom in which the effective positive image charge plays the role of the nucleus. Our task is to determine the confinement potential for the electrons proceeding from the Coulomb interaction between the electrons and the image charge in the quantum-mechanical formalism. For this purpose, we consider a three-body system with the Coulomb interaction in an constant external magnetic field. Let m_1, m_2, and m_3 be the masses and $-Z_1e$, $-Z_2e$, and $-Z_3e$ the charges of the particles. The Hamiltonian of the system can be written in the form

$$H = \frac{1}{2}\sum_{j=1}^{3}\frac{1}{m_j}\left[\mathbf{P}_j + \frac{e}{c}\mathbf{A}(r_j)\right]^2 + \frac{1}{4\pi\varepsilon\varepsilon_0}\frac{Z_1Z_2e^2}{|\mathbf{r}_1-\mathbf{r}_2|} - \frac{1}{4\pi\varepsilon\varepsilon_0}\frac{Z_1Z_3e^2}{|\mathbf{r}_1-\mathbf{r}_3|} - \frac{1}{4\pi\varepsilon\varepsilon_0}\frac{Z_3Z_2e^2}{|\mathbf{r}_2-\mathbf{r}_3|}. \quad (1)$$

Here, ε and ε_0 are the relative and absolute permittivities and $\mathbf{A}(\mathbf{r})$ is the vector potential, which is defined in the standard manner:

$$\mathbf{A}(\mathbf{r}) = \frac{1}{2}[\mathbf{B}\times\mathbf{r}], \quad (2)$$

where $\mathbf{B}$ is the external magnetic field strength. We introduce the Jacobi coordinates $\{\mathbf{x}, \mathbf{y}\}$ and the center-of-mass coordinates $\mathbf{R}$ of the system:

$$\mathbf{r}_1 = \frac{m_2}{m_1+m_2}\mathbf{x} + \frac{m_3}{m_1+m_2+m_3}\mathbf{y} + \mathbf{R},$$
$$\mathbf{r}_2 = -\frac{m_1}{m_1+m_2}\mathbf{x} + \frac{m_3}{m_1+m_2+m_3}\mathbf{y} + \mathbf{R}, \quad (3)$$
$$\mathbf{r}_3 = -\frac{m_1+m_2}{m_1+m_2+m_3}\mathbf{y} + \mathbf{R}.$$

The confinement potential in quantum dots is characterized by a strong blocking along one of the three coordinate axes (x, y, or z). As a result, low-lying quantum excitations are determined by the properties of the confinement potential along the remaining two axes. Consequently, quantum dots can be regarded from the geometrical point of view as effective 2D systems. The external magnetic field may be oriented, for example, in a plane perpendicular to the plane of the quantum dot. We assume that these planes intersect along a straight line directed strictly along x; i.e., $\mathbf{A}(y) = 0$. Taking into account this assumption, we can divide Hamiltonian (1) into two parts: the Hamiltonian of the center-of-mass system

$$H_{cm} = \frac{1}{2}\mathbf{P}_Q^2 + \frac{\hbar^2}{4}\frac{m^*}{m_t}\omega_c^2\rho_Q^2 + \frac{m^*}{2m_t}\hbar\omega_c L_z \quad (4)$$

and the Hamiltonian of relative motion

$$H_{rm} = \frac{1}{2M}\mathbf{P}_x^2 + \frac{1}{2\mu}\mathbf{P}_y^2 + m^*\frac{\rho_x^2}{16}\omega_c^2 + \frac{e^2}{4\pi\varepsilon\varepsilon_0}\frac{Z_1Z_2}{|\mathbf{x}|} - \frac{e^2}{4\pi\varepsilon\varepsilon_0}\frac{Z_1Z_3}{|\mathbf{y}+\mathbf{x}/2|} - \frac{e^2}{4\pi\varepsilon\varepsilon_0}\frac{Z_2Z_3}{|\mathbf{y}-\mathbf{x}/2|} + \frac{1}{2}\hbar\omega_c L_{x_z}. \quad (5)$$

Here, $\omega_c = eB/cm^*$ is the cyclotron frequency, $m_1 = m_2 = m^*$ is the effective electron mass, and the following notation has been introduced:

$$M = \frac{1}{2}m^* = \frac{m_1m_2}{m_1+m_2}, \quad \mu = \frac{(m_1+m_2)m_3}{m_t},$$
$$m_t = m_1+m_2+m_3, \quad \mathbf{Q} = \frac{\sqrt{m_t}}{\hbar}\mathbf{R}, \quad (6)$$
$$\rho_Q^2 = Q_1^2 + Q_2^2.$$

Due to the presence of the magnetic field, each Hamiltonian contains the components of the angular momentum $\mathbf{L} = -i\hbar[\mathbf{r}\times\Delta_r]$ along the z axis in the intrinsic coordinate systems. Accordingly, operators L_{x_z} and L_z are the angular momentum components in the coordi-

[1] See, for example, the discussion on the role of image potentials in the formation of excitons in nanostructures of the type of superlattices and quantum wells [15].

nate systems of relative motion ($\mathbf{r} \equiv \mathbf{x}$) and of the center of mass ($\mathbf{r} \equiv \mathbf{Q}$).

It should be noted that the Coulomb interaction makes zero contribution to the Hamiltonian (4) describing the motion of the center of mass. Solutions for the center-of-mass Hamiltonian in the presence of a parabolic confinement potential were considered for the first time by Fock and are known in the literature and the Fock–Darwin levels [18]. However, in the case under investigation, in which the existence of a confinement potential (e.g., parabolic confinement potential) is not presumed *a priori*, the eigenvalues of the center-of-mass Hamiltonian have the form

$$E_{NM} = (2N + |M| + 1)\sqrt{\frac{m^*}{2m_t}}\hbar\omega_c + \frac{m^*}{2m_t}M\hbar\omega_c, \quad (7)$$

where $N = 0, 1, 2, \ldots$ is the radial quantum number and $M = 0, \pm1, \pm2, \ldots$ is the angular momentum quantum number. In contrast to the Fock–Darwin solutions, which are determined by the parabolic confinement potential and cyclotron frequency, the energy spectrum of the motion of the center of mass in our model is determined by the kinetic energy of electrons, the cyclotron frequency and depends on the ratio of the electron effective mass and the sum of the masses of the system components. In order to analyze the Hamiltonian of relative motion, we go over to new variables

$$\mathbf{r} = \frac{\sqrt{M}}{\hbar}\mathbf{x}, \quad \boldsymbol{\zeta} = \frac{\sqrt{\mu}}{\hbar}\mathbf{y}. \quad (8)$$

Further, we assume that the image charge Z_3 depends not only on the effective electric charge Q, but also on the ratio of the permittivities of the media:

$$Z_3 = \sqrt{\frac{\varepsilon}{\varepsilon'}}\frac{|\varepsilon - \varepsilon'|}{\varepsilon + \varepsilon'}Q,$$

where ε and ε' are the relative dielectric constants, e.g., of a semiconductor and an insulator. The corresponding Schrödinger equation for relative motion has the form

$$\left\{\frac{1}{2}\mathbf{P}_r^2 + \frac{1}{2}\mathbf{P}_\zeta^2 + \frac{\hbar^2}{8}\omega_c^2\rho_r^2 + \frac{\hbar}{a^*\sqrt{2m^*}}\frac{Z_1Z_2}{r} + \frac{1}{2}\hbar\omega_c L_{r_z} - \frac{\hbar\sqrt{2}f}{a^*\sqrt{m^*}}\frac{Z_1Z_3}{|\boldsymbol{\zeta} + f\mathbf{r}|} - \frac{\hbar\sqrt{2}f}{a^*\sqrt{m^*}}\frac{Z_2Z_3}{|\boldsymbol{\zeta} - f\mathbf{r}|} - E\right\}\Psi(\mathbf{r}, \boldsymbol{\zeta}) = 0, \quad (9)$$

where $f = \sqrt{m_3/m_t}$, $a^* = a_B\varepsilon m_e/m^*$ is the effective radius of the system and a_B is the Bohr radius. In the next section, we will find the solutions of the Schrödinger equation (9).

3. ANALYSIS OF RELATIVE MOTION

3.1. Energy Spectrum of the Intrinsic System

In the two-center adiabatic approximation [19], the wave function of the relative motion of a three-body Coulomb system can be represented in the form

$$\Psi(\mathbf{r}, \boldsymbol{\zeta}) = \chi(\mathbf{r})\Phi(r, \boldsymbol{\zeta}), \quad (10)$$

where $\Phi(r, \boldsymbol{\zeta})$ is the wave function of the intrinsic system. In the two-center approximation, variable r is considered as an external parameter; i.e., r = const. We present vector $\boldsymbol{\zeta}$ in the cylindrical system of coordinates: $\boldsymbol{\zeta} = \{\rho, z, \varphi\}$. In this case, the wave function $\Phi(r, \boldsymbol{\zeta})$ assumes the form

$$\Phi(r, \boldsymbol{\zeta}) = \frac{e^{im\varphi}}{\sqrt{\pi}}\tilde{\Phi}_m(r; \rho, z). \quad (11)$$

Here, φ is the azimuthal angle and m is the magnetic quantum number. Taking into account expression (11) and simplifying the Schrödinger equation (9), we obtain

$$\left\{-\frac{\hbar^2}{2}\left[\frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} - \frac{m^2}{\rho^2} + \frac{\partial^2}{\partial z^2}\right] - \frac{Z_1Z_3\lambda}{\sqrt{\zeta^2 + 2frz + f^2r^2}} - \frac{Z_2Z_3\lambda}{\sqrt{\zeta^2 - 2frz + f^2r^2}}\right\}\tilde{\Phi}_m(r; \rho, z) = E_r(r)\tilde{\Phi}_m(r; \rho, z), \quad (12)$$

where $E_r(r)$ is the eigenvalue of the Hamiltonian of the intrinsic system and parameter λ is defined as

$$\lambda = \frac{\hbar\sqrt{2}f}{a^*\sqrt{m^*}}.$$

Carrying out the substitution of variables

$$\rho = 2\sqrt{\rho_1\rho_2}, \quad z = (\rho_1 - \rho_2) \quad (13)$$

and going over to the parabolic system of coordinates in Eq. (12), after relevant calculations, we obtain

$$\left\{-\frac{\hbar^2}{2}\left[\rho_1\frac{\partial^2}{\partial\rho_1^2} + \frac{\partial}{\partial\rho_1} + \rho_2\frac{\partial^2}{\partial\rho_2^2} + \frac{\partial}{\partial\rho_2} - \frac{m^2}{4\rho_1} - \frac{m^2}{4\rho_2}\right] - (\rho_1 + \rho_2)E_r - \frac{Z_1Z_3\lambda(\rho_1 + \rho_2)}{\sqrt{(\rho_1 + \rho_2)^2 + 2fr(\rho_1 - \rho_2) + f^2r^2}} - \frac{Z_2Z_3\lambda(\rho_1 + \rho_2)}{\sqrt{(\rho_1 + \rho_2)^2 - 2fr(\rho_1 - \rho_2) + f^2r^2}}\right\}\tilde{\Phi}_m(r; \rho_1, \rho_2) = 0. \quad (14)$$

Before defining the energy spectrum and the wave function of the Schrödinger equation (14) using the oscillator representation method [16], it is appropriate to note that this method is based on the ideas and methods of the quantum theory of a scalar field. However, a considerable difference between quantum field theory

and quantum mechanics is that in the former case, the quantized fields in the form of a set of an infinite number of oscillators for the ground state of a vacuum preserve their oscillatory nature in the quantum-field interaction. In quantum mechanics, the behavior of the eigenfunctions for most potentials differs from the Gaussian behavior of the oscillator wave function. For this reason, while applying the methods and ideas of quantum field theory for solving quantum-mechanical problems, the variables in the initial radial Schrödinger equation should be changed so that the sought wave function would display the Gaussian behavior at large distances, and the transformed equation should be identified with the radial Schrödinger equation in a space with a large dimension. It should be noted that a similar idea was discussed for the first time by Fock while solving the problem of the spectrum of the hydrogen atom with the help of transformation to the four-dimensional momentum space [20].

Following Fock [21], we will assume that the asymptotic behavior of the wave function of the intrinsic system is of the Coulomb type. In accordance with what has been said above, we change the variables as follows (see [16] for details):

$$\rho_k = q_k^2, \quad \tilde{\Phi}_m = q_1^{|m|} q_2^{|m|} \Psi_m(q_1^2, q_2^2), \quad k = 1, 2. \tag{15}$$

Using the atomic system of units ($\hbar = e = c = 1$), we obtain from (14)

$$\left\{ -\frac{1}{2}\sum\left[\frac{\partial^2}{\partial q_j^2} + \frac{d-1}{q_j}\frac{\partial}{\partial q_j}\right] - \frac{4Z_1 Z_3 \lambda (q_1^2 + q_2^2)}{\sqrt{(q_1^2 + q_2^2)^2 + 2fr(q_1^2 - q_2^2) + f^2 r^2}} - 4E(q_1^2 + q_2^2) - \frac{4Z_2 Z_3 \lambda (q_1^2 + q_2^2)}{\sqrt{(q_1^2 + q_2^2)^2 - 2fr(q_1^2 - q_2^2) + f^2 r^2}} \right\} \Psi_m(q_1^2, q_2^2) = 0, \tag{16}$$

where d is the dimension of the auxiliary space, which is given by

$$d = 2 + 2|m|. \tag{17}$$

As a result of the change of variables, we obtain a modified Schrödinger equation in the d-dimensional auxiliary space R^d. It follows from Eqs. (16) and (17) that the magnetic quantum number m appears in the definition of the dimension d of the space. This approach makes it possible to determine all the characteristics we are interested in, including the spectrum and the wave function, by solving the modified Schrödinger equation for the ground state only in the d-dimensional auxiliary space R^d. The wave function $\Psi_m(q_1^2, q_2^2)$ of the ground state in R^d is a function of variables q_1^2 and q_2^2 only. For this reason, we identify the operator

$$\frac{\partial^2}{\partial q_k^2} + \frac{d-1}{q_k}\frac{\partial}{\partial q_k} \equiv \Delta_{q_k}, \quad k = 1, 2 \tag{18}$$

with the Laplacian Δ_{q_k} in the auxiliary space R^d, which acts on the wave function of the ground state, which is a function of radius q_k only. Proceeding from the modified Schrödinger equation

$$H\Psi_m(q_1, q_2) = \varepsilon(E_r)\Psi_m(q_1, q_2), \tag{19}$$

in accordance with Eq. (16), we find that the energy spectrum in R^d is equal to zero:

$$\varepsilon(E_r) = 0. \tag{20}$$

We will consider this relation as the condition for determining the energy spectrum E_r of the Hamiltonian (12). Following the oscillator representation method, we write the canonical variables in terms of the creation and annihilation operators in the R^d space:

$$q_j^{(k)} = \frac{a_j^k + a_j^{k+}}{\sqrt{2\omega_k}}, \quad P_j^{(k)} = \sqrt{\frac{\omega_k}{2}}\frac{a_j^k - a_j^{k+}}{i}, \quad k = 1, 2, \quad j = 1, \ldots, d, \quad [a_i^k, a_j^{k+}] = \delta_{i,j}, \tag{21}$$

where ω_k is the oscillator frequency, which is yet unknown. Substituting expressions (21) into Eq. (16) and ordering in the creation and annihilation operators, we obtain

$$H = H_0 + \varepsilon_0(E_r) + H_I. \tag{22}$$

Here, H_0 is the Hamiltonian of two uncoupled oscillators,

$$H_0 = \omega_1(a_j^+(1)a_j(1)) + \omega_2(a_j^+(2)a_j(2)), \tag{23}$$

and $\varepsilon_0(E_r)$ is the ground-state energy in the zeroth approximation of the oscillator representation [16, 22], which has the form

$$\varepsilon_0(E_r) = \frac{d}{4}\omega_1 + \frac{d}{4}\omega_2 - 2\frac{dE_r}{\omega_1} - 2\frac{dE_r}{\omega_2} - 4(\omega_1\omega_2)^{d/2} \times \int_0^\infty\int_0^\infty \frac{d\beta_1 d\beta_2}{\Gamma^2(d/2)}\left[\frac{Z_1 Z_3 \lambda (\beta_1\beta_2)^{d/2-1}(\beta_1 + \beta_2)}{\sqrt{(\beta_1 + \beta_2)^2 + 2fr(\beta_1 - \beta_2) + f^2 r^2}} + \frac{Z_2 Z_3 \lambda (\beta_1\beta_2)^{d/2-1}(\beta_1 + \beta_2)}{\sqrt{(\beta_1 + \beta_2)^2 - 2fr(\beta_1 - \beta_2) + f^2 r^2}}\right] \times \exp(-\omega_1\beta_1 - \omega_2\beta_2). \tag{24}$$

The interaction Hamiltonian H_I can also be represented in normal form in the creation and annihilation opera-

tors. It contains no quadratic terms in the canonical variables:

$$H_I = -\frac{\partial}{\partial\beta}\int_0^\infty \frac{dt}{\sqrt{\pi t}}\int_{-\infty}^{\infty}\frac{d\tau}{\sqrt{\pi}}e^{-\tau^2}\int_{-\infty}^{\infty}\left(\frac{d\eta_1}{\sqrt{\pi}}\right)^d\int_{-\infty}^{\infty}\left(\frac{d\eta_2}{\sqrt{\pi}}\right)^d$$

$$\times\exp[-(fr)^2 t-\eta_1^2-\eta_2^2]\left\{\exp\left\{-\eta_1^2\frac{\mu_+}{\omega_1}-\eta_2^2\frac{\mu_-}{\omega_2}\right\}\right.$$

$$\times F(2i\sqrt{\mu_+}(\eta_1 q_1), 2i\sqrt{\mu_-}(\eta_2 q_2)) \qquad (25)$$

$$+\exp\left\{-\eta_1^2\frac{\mu_-}{\omega_1}-\eta_2^2\frac{\mu_+}{\omega_2}\right\}$$

$$\left.\times F(2i\sqrt{\mu_-}(\eta_1 q_1), 2i\sqrt{\mu_+}(\eta_1 q_2))\right\}\Bigg|_{\beta=0},$$

where we have introduced the following notation:

$$F(y_1, y_2) =: e_2^{-y_1} : : e_2^{-y_2} : + : e_2^{-y_2} :\left(1+\frac{1}{2}:y_1^2:\right)$$

$$+ : e_2^{-y_1} :\left(1+\frac{1}{2}:y_2^2:\right),$$

$$\mu_\pm = \beta \pm 2rft + 2i\sqrt{t}\tau.$$

Here, :…: symbolizes normal ordering, and the following notation has been used: $e_2^x = e^x - 1 - x - x^2/2$ (see also [17]). Some details of the representation of the Hamiltonian in normal form are given in the Appendix.

The contribution of the interaction Hamiltonian H_I is considered as a small perturbation. In quantum field theory, after representing the canonical variables in terms of the creation and annihilation operators and representing the interaction Hamiltonian in normal form, we find that the requirement of the absence of second-degree field operators in the interaction Hamiltonian is essentially equivalent to renormalizations of the coupling constant and the wave function [23]. Moreover, such a procedure makes it possible to take into account the main quantum contribution through the renormalization of masses and through the energy of the vacuum. In other words, all quadratic forms are completely included in the Hamiltonian of a free oscillator. This requirement makes it possible to formulate, in accordance with the oscillator representation, the conditions

$$\frac{\partial\varepsilon_0}{\partial\omega_1} = 0, \quad \frac{\partial\varepsilon_0(E)}{\partial\omega_2} = 0 \qquad (26)$$

for determining the frequencies ω_1 and ω_2 of uncoupled oscillators, which determine the main quantum contribution. Taking into account Eq. (24), we can use Eqs. (20) and (26) for calculating the energy E_r of the intrinsic system as a function of parameter r. Since we are interested in the ground state, we will not consider radial perturbations here. It was verified more than once by using the oscillator representation for various potentials [17, 22, 24] that the first-order correction associated with the interaction Hamiltonian is identically equal to zero, while the second-order correction is smaller than one percent. For this reason, we will confine our analysis to the zeroth approximation only.

In this approximation, energy E_r is given by

$$E_r = \frac{\omega_1\omega_2}{8}-\frac{2}{d}\frac{(\omega_1\omega_2)^{d/2+1}}{\omega_1+\omega_2}\int_0^\infty\int_0^\infty\frac{d\beta_1 d\beta_2}{\Gamma^2(d/2)}$$

$$\times(\beta_1\beta_2)^{d/2-1}\left[\frac{Z_1 Z_3\lambda(\beta_1+\beta_2)}{\sqrt{(\beta_1+\beta_2)^2+2fr(\beta_1-\beta_2)+f^2r^2}}\right. \qquad (27)$$

$$\left.+\frac{Z_2 Z_3\lambda(\beta_1+\beta_2)}{\sqrt{(\beta_1+\beta_2)^2-2fr(\beta_1-\beta_2)+f^2r^2}}\right]$$

$$\times\exp\{-\omega_1\beta_1-\omega_2\beta_2\}.$$

We introduce new parameters

$$\omega_+ = \frac{\omega_1+\omega_2}{2}, \quad \omega_- = \frac{\omega_1-\omega_2}{2}, \qquad (28)$$

which can be determined from Eqs. (26) taking into account expressions (24) and (27) as functions of quantity r.

It should be noted that in the approach we used here, the violation of spherical symmetry is controlled by parameter ω_-.

3.2. Structure of Confinement Potential

Let us now consider the problem of the confinement potential. Taking into account Eq. (12), substituting expression (10) into Eq. (9), and carrying out transformations (in the ordinary system of units), we obtain

$$\left[\frac{1}{2}\mathbf{P}_r^2+\frac{\hbar^2}{8}\omega_c^2\rho_r^2+\frac{\hbar}{a^*\sqrt{2m^*}}\frac{Z_1Z_2}{r}-E\right.$$

$$\left.+\frac{\hbar\omega_c}{2}Lr_z+V_c(r)\right]\chi(\mathbf{r}) = 0, \qquad (29)$$

where E is the energy spectrum of the initial system and the quantity $V_c(r)$ is precisely the required confinement potential:

$$V_c(r) = E_r(r)+\frac{1}{4}\left(\frac{1}{\omega_+}\frac{\partial}{\partial r}\omega_+\right)^2. \qquad (30)$$

The first term in this expression, $E_r(r)$, is the potential (27) created by the electrostatic field of the image charge, while the second term is associated with the relative motion of particles and is determined by averaging the total Hamil-

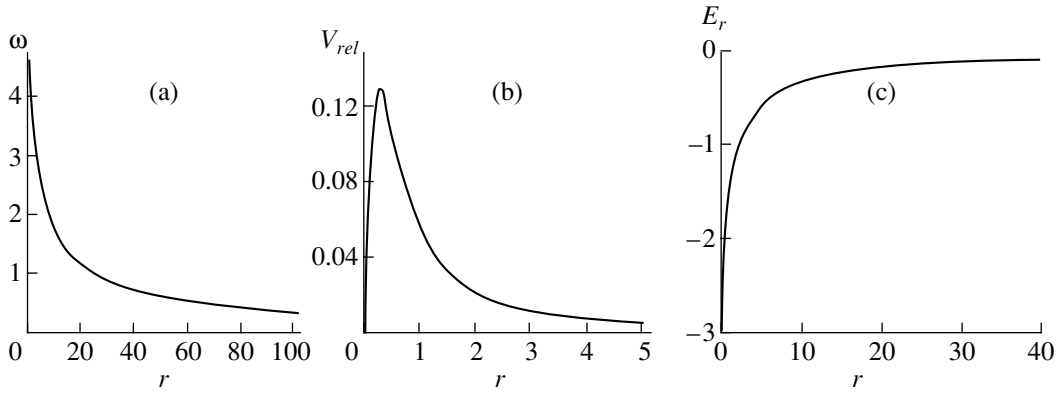


Fig. 1. Dependences of (a) the oscillator frequency, (b) the potential created by the relative motion of electrons, and (c) the potential created by the electrostatic field of the image charge on the separation between two electrons.

tonian (9) over the wave function $\Phi(r, \zeta)$ of the intrinsic system. Obviously, the confinement potential obtained contains various solutions depending on the form of Coulomb interaction and on the magnitude of the magnetic field. Here, we confine our analysis only to the spherically symmetric solution $\omega_- = 0$, i.e., $\omega_1 = \omega_2 \equiv \omega$. In this case, we have

$$V_c(r) = \frac{\omega^2}{8} + \frac{1}{4}\left(\frac{1}{\omega}\frac{\partial\omega}{\partial r}\right)^2 - \frac{\lambda\omega}{2}(Z_1Z_3 + Z_2Z_3)\left[2\frac{1-e^{-fr\omega}}{fr\omega} - e^{-fr\omega}\right], \quad (31)$$

where ω is defined by the equation

$$\omega - 2\lambda(Z_1Z_3 + Z_2Z_3)(1 + fr\omega)e^{-fr\omega} = 0. \quad (32)$$

Thus, the confinement potential in our model of a two-electron quantum dot differs from the parabolic confinement potential and is defined by expression (31) and Eq. (32).

Considering the limit $r \ll 1$ and expanding potential V_c into a Taylor series in variable r, we obtain

$$V_c(r) = -\frac{\hbar\omega_0}{8} + \hbar^2\left(\frac{1}{24} + \frac{f^2}{4}\right)f^2\omega_0^2r^2 - \left(\frac{1}{48} + \frac{f^2}{4}\right)\hbar^{3/2}\omega_0^{3/2}r^3 + \hbar^3\frac{\omega_0^3f^4}{60}r^4 + O(r^5), \quad (33)$$

where

$$\omega_0 = \left(4Z_3\frac{f\sqrt{2\hbar}}{a^*\sqrt{m^*}}\right)^2. \quad (34)$$

Confining ourselves only to the second power in r in Eq. (33), we obtain the parabolic confinement potential with the confinement frequency ω_0. It follows from Eqs. (31), (33), and (34) that the properties of the potential depend on the image charge and the effective electron mass. This dependence is included in the final expression for the energy spectrum through parameters ω_0 and f. If the mass of the image charge is created by all the electrons of the ambient medium or by heavy ions, parameter $f \approx 1$ ($f = 1/\sqrt{3}$ for $m_1 = m_2 = m_3$). The characteristic length of the quantum dot (well) [8, 9] formed at the interface between two media,

$$l_0 \equiv \sqrt{\frac{\hbar}{m^*\omega_0}} = \frac{a^*}{4Z_3f\sqrt{2}},$$

is also a function of the image charge.

Using specific values of parameters in the case of a quantum dot (e.g., $Z_1 = Z_2 = Z_3 = 1$ and $m_1 = m_2 = m_3$), we can determine the dependence of the oscillator frequency ω on the value of r appearing in Eq. (32). Accordingly, this allows us to determine the dependence of the potential

$$E_r(r) = \frac{\omega^2}{8} - \lambda\omega\left[2\frac{1-e^{fr\omega}}{fr\omega} - e^{-fr\omega}\right] \quad (35)$$

created by the electrostatic field of the image charge, as well as of the potential

$$V_{\rm rel} = \frac{1}{4}\left(\frac{1}{\omega}\frac{\partial\omega}{\partial r}\right)^2 \quad (36)$$

associated with the relative motion of electrons, on parameter r. It can be seen from Fig. 1a that with increasing r, the oscillator frequency ω decreases smoothly from 4.619 to 0. Potential $V_{\rm rel}$ is equal to zero at the origin, increases over small distances, and then decreases rapidly upon a further increase in r (Fig. 1b). Potential $E_r(r)$ is finite for $r = 0$; i.e., it has no singularity, while for $r = \infty$, it decreases as a Coulomb potential. The value of potential $V_{\rm rel}$ is an order of magnitude smaller than the absolute value of $E_r(r)$. Thus, the main contribution to the confinement potential is determined by the interaction of the electrons in the electrostatic field of the image charge.

4. ENERGY SPECTRUM

In this section, the oscillator representation [16] is used for calculating the energy spectrum of a two-electron system with potential $V_c(r)$ (33). For illustration, we consider the case when $z = 0$, i.e., a 2D system which may serve as a model of a quantum dot. According to Eqs. (29) and (33), the Hamiltonian of the relative motion of the two-electron system has the form

$$H = -\frac{\hbar^2}{2}\left[\frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} - \frac{m^2}{\rho^2}\right] + \frac{\hbar^2}{2}\Omega^2\rho^2 + \frac{k\sqrt{\hbar\omega_0}}{\rho} - \hbar^2\sqrt{\hbar}W\rho^3 + \hbar^3 G\rho^4 - \frac{\hbar\omega_0}{8} + \frac{1}{2}m\hbar\omega_c, \quad (37)$$

where $m = 0, \pm1, \ldots$ is the magnetic quantum number and

$$\omega_r^2 = \frac{\lambda^2\omega_0^2}{12}, \quad W = \frac{\lambda^3}{48}\omega_0^2\sqrt{\omega_0}, \quad \Omega = \sqrt{\omega_r^2 + \frac{\omega_c^2}{4}}, \quad G = \frac{\lambda^4\omega_0^3}{160}, \quad k = \frac{l_0}{a^*}. \quad (38)$$

The Schrödinger equation for Hamiltonian (37) assumes the form

$$\left[\frac{1}{2}P_\rho^2 + \frac{\hbar}{2}\Omega^2\rho^2 + \frac{k\sqrt{\hbar\omega_0}}{\rho} - \hbar^{3/2}W\rho^3 + \hbar^3 G\rho^4\right]\Psi_m = U_m\Psi_m. \quad (39)$$

Here, U_m is the energy parameter,

$$U_m = E_m + \frac{\hbar\omega_0}{8} - \frac{m}{2}\hbar\omega_c. \quad (40)$$

We will first consider a purely parabolic potential; i.e., $W = 0$ and $G = 0$. In this case, we obtain the following expression for the energy spectrum (see [9] for details):

$$E_m = \hbar\omega_0\left\{-\frac{1}{8} + t\frac{m}{2} + x^2(1 + |m|)\sqrt{\frac{f^2}{2}\left(\frac{1}{6} + f^2\right) + \frac{t^2}{4}} + \frac{3kx}{2}\left[\frac{f^2}{2}\left(\frac{1}{6} + f^2\right) + \frac{t^2}{4}\right]^{1/4} \times \frac{\Gamma(|m| + 1/2)}{\Gamma(1 + |m|)} + \frac{t}{4}[1 - (-1)^m]\frac{m^*}{m_e}g^*\right\}. \quad (41)$$

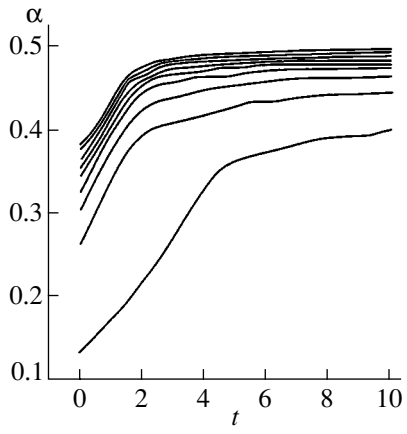


Fig. 2. Dependence of parameter α on the relative external magnetic field strength t for various values of the magnetic quantum number m. The lower level corresponds to $m = -1$, the next to $m = -2$, and so on.

Parameter x can be determined from the equation

$$x^4 + x^3 k\left[\frac{f^2}{2}\left(\frac{1}{6} + f^2\right) + \frac{t^2}{4}\right]^{-1/4} \times \frac{\Gamma(|m| + 1/2)}{\Gamma(2 + |m|)} - 1 = 0. \quad (42)$$

Here, $t = \omega_c/\omega_0$ is the relative magnetic field strength and g^* is the effective Lande factor. Expression (41) also includes the contribution from the Zeeman effect associated with the spin interaction between the two electrons in a magnetic field. Expressions (41) and (42) allow us to determine the ground states of a quantum dot as functions of its size $k = l_0/a^*$ and the relative magnetic field strength t.

Let us now calculate the energy spectrum of Hamiltonian (39). In this case, the change of variables is presented as follows:

$$\rho = q^{2\alpha}, \quad \Psi_m = q^{2\alpha|m|}\Phi_m(q), \quad (43)$$

where parameter α is associated with the behavior of the wave function at large distances. The potential contains anharmonic terms, and we will determine parameter α using the results obtained in [22]. For large values of ρ, the asymptotic form of the wave function is determined by the anharmonic term $G\rho^4$ with $\alpha = 1/3$. For small values of G and W, the true wave function is closer to the Gaussian wave function and, hence, $\alpha = 1/2$. This limit corresponds to a parabolic confinement potential. Thus, parameter α, which was initially regarded as a variational parameter in the minimization of the ground-state energy in the zeroth approximation [22], may change in the interval $1/3 \le \alpha \le 1/2$. Figure 2 shows the dependence of parameter α on the magnetic field strength $t = \omega_c/\omega_0$ for states with $m = -1, -2, -3, \ldots$. The results of analysis show that for states with a small absolute value of the magnetic quan-

tum number m, parameter $\alpha < 1/2$. As the magnetic field increases, parameter α asymptotically tends to the limit $\alpha = 1/2$ corresponding to a Gaussian wave function. It should be noted that in the oscillator representation, one can avoid the summation of series in the perturbation theory; i.e., the Dyson phenomenon problem [25] can be bypassed successfully owing to the introduction of parameter α.

After some transformations of Eq. (39), we derive the modified Schrödinger equation

$$\left\{-\frac{\hbar^2}{2}\left[\frac{\partial^2}{\partial q^2}+\frac{d-1}{q}\frac{\partial}{\partial q}\right]+4k\sqrt{\hbar\omega_0}\alpha^2 q^{2(\alpha-1)}\right.$$
$$+2\alpha^2\hbar^2\Omega^2 q^{2(4\alpha-1)}-4\hbar^{3/2}\alpha^2 W q^{2(5\alpha-1)} \qquad (44)$$
$$\left.+4\alpha^2\hbar^3 G q^{2(6\alpha-1)}-4\alpha^2 U_m q^{2(2\alpha-1)}\right\}\Phi_m(q^2) = 0,$$

where $d = 2 + 4\alpha|m|$. The ground-state energy in the zeroth approximation is given by

$$\varepsilon_0(U_m) = \frac{d\omega\hbar}{4}-\frac{4\alpha^2 U_m}{(\omega\hbar)^{2\alpha-1}}\frac{\Gamma(d/2+2\alpha-1)}{\Gamma(d/2)}$$
$$+\frac{4k\sqrt{\hbar\omega_0}\alpha^2}{(\omega\hbar)^{\alpha-1}}\frac{\Gamma(d/2+\alpha-1)}{\Gamma(d/2)}$$
$$-\frac{4\hbar^{3/2}\alpha^2 W}{(\omega\hbar)^{5\alpha-1}}\frac{\Gamma(d/2+5\alpha-1)}{\Gamma(d/2)} \qquad (45)$$
$$+\frac{2\alpha^2\hbar^2\Omega^2}{(\omega\hbar)^{4\alpha-1}}\frac{\Gamma(d/2+4\alpha-1)}{\Gamma(d/2)}$$
$$+\frac{4\alpha^2\hbar^3 G}{(\omega\hbar)^{6\alpha-1}}\frac{\Gamma(d/2+6\alpha-1)}{\Gamma(d/2)},$$

and for the interaction Hamiltonian we have

$$H_I = \int_0^\infty d\tau\int_{-\infty}^{\infty}\left(\frac{d\eta}{\sqrt{\pi}}\right)^d e^{-\eta^2(1+\tau)} : e_2^{-2i\sqrt{x\hbar\omega}(q\eta)} :$$
$$\times\left[\frac{4k\sqrt{\hbar\omega_0}\alpha^2}{(\omega\hbar)^{\alpha-1}}\frac{\tau^{-\alpha}}{\Gamma(1-\alpha)}\right.$$
$$-\frac{4\hbar^2\sqrt{\hbar}\alpha^2 W}{(\omega\hbar)^{5\alpha-1}}\frac{\tau^{-5\alpha}}{\Gamma(1-5\alpha)}-\frac{4U_m\alpha^2}{(\omega\hbar)^{2\alpha-1}}\frac{\tau^{-2\alpha}}{\Gamma(1-2\alpha)} \qquad (46)$$
$$\left.+\frac{2\alpha^2\hbar^2\Omega^2}{(\omega\hbar)^{4\alpha-1}}\frac{\tau^{-4\alpha}}{\Gamma(1-4\alpha)}+\frac{4\alpha^2\hbar^3 G}{(\omega\hbar)^{6\alpha-1}}\frac{\tau^{-6\alpha}}{\Gamma(1-6\alpha)}\right].$$

Equation

$$\frac{\partial\varepsilon_0(U_m)}{\partial\omega} = 0 \qquad (47)$$

is used for determining parameter ω as a function of energy U_m as well as other parameters of the potential.

Taking into account Eqs. (47) and (45) and using notation (38), we obtain, after appropriate transformations, the following expression for the energy spectrum:

$$\frac{E_m}{\hbar\omega_0} = \min_\alpha\left\{-\frac{1}{8}+\frac{tm}{2}+\frac{tm^*}{4m_e}[1-(-1)^m]g^*\right.$$
$$+\frac{z^2}{4\alpha}\frac{\sqrt{\Gamma(2+2\alpha|m|)\Gamma(4\alpha+2\alpha|m|)}}{\Gamma(2\alpha+2\alpha|m|)}\sqrt{\frac{f^2}{3}+t^2}$$
$$+\frac{3kz\sqrt{\alpha}}{2}\frac{\Gamma(\alpha+2|m|)}{\Gamma(2\alpha+2\alpha|m|)}$$
$$\times\left[\frac{\Gamma(4\alpha+2\alpha|m|)}{\Gamma(2+2\alpha|m|)}\left(\frac{f^2}{3}+t^2\right)\right]^{1/4}$$
$$+\frac{f^3}{96z^3\alpha^{3/2}}\frac{\Gamma(5\alpha+2\alpha|m|)}{\Gamma(2+2\alpha m)} \qquad (48)$$
$$\times\left[\frac{\Gamma(2+2\alpha|m|)}{\Gamma(4\alpha+2\alpha|m|)}\left(\frac{f^2}{3}+t^2\right)^{-1}\right]^{3/4}$$
$$-\frac{f^4}{160z^4\alpha^2}\frac{\Gamma(6\alpha+2\alpha|m|)}{\Gamma(2+2\alpha|m|)}$$
$$\left.\times\frac{\Gamma(2+2\alpha|m|)}{\Gamma(4\alpha+2\alpha|m|)}\left(\frac{f^2}{3}+t^2\right)^{-1}\right\}.$$

Parameter z is determined from the equation

$$z^4+4k\alpha^{3/2}z^3\frac{\Gamma(\alpha+2\alpha|m|)}{\Gamma(2+2\alpha|m|)}$$
$$\times\left[\frac{\Gamma(2+2\alpha|m|)}{\Gamma(4\alpha+2\alpha|m|)}\left(\frac{f^2}{3}+t^2\right)^{-1}\right]^{1/4}-1$$
$$+\frac{f^3}{4z\sqrt{\alpha}}\frac{\Gamma(5\alpha+2\alpha|m|)}{\Gamma(2\alpha+2\alpha|m|)} \qquad (49)$$
$$\times\left[\frac{\Gamma(2+2\alpha|m|)}{\Gamma(2\alpha+2\alpha|m|)}\left(\frac{f^2}{3}+t^2\right)^{-1}\right]^{5/4}$$
$$-\frac{f^4}{z^2 10\alpha}\frac{\Gamma(6\alpha+2\alpha|m|)}{\Gamma(2\alpha+2\alpha|m|)}$$
$$\times\left[\frac{\Gamma(2+2\alpha|m|)}{\Gamma(4\alpha+2\alpha|m|)}\left(\frac{f^2}{3}+t^2\right)^{-1}\right]^{3/2} = 0.$$

Obviously, the energy spectrum (48) differs from the energy spectrum (41) determined by the parabolic nature of the confinement potential. It follows from Fig. 3 that the dependences of the energy spectrum on the size $k = l/a^*$ of the quantum dot and the relative magnetic field strength $t = \omega_c/\omega_0$ for parabolic and quasi-para-

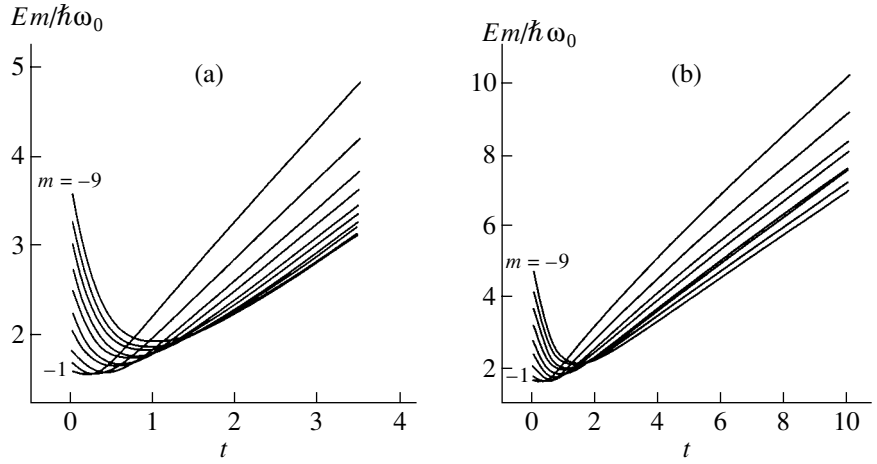


Fig. 3. Energy spectra of the Hamiltonian of the relative motion taking into account the Zeeman effect as functions of the relative strength t of the external magnetic field for (a) parabolic and (b) quasi-parabolic potentials calculated by using formulas (41) and (48), respectively.

bolic potentials are similar. However, singlet–triplet transition in weak magnetic fields in the case of a quasi-parabolic potential occur at higher values of the magnetic field (cf. Figs. 3a and 3b). We expect that the confinement potential in weak magnetic fields will differ from the parabolic confinement potential. At the same time, an analysis of expressions (48) and (49) shows that in strong magnetic field, $\omega_c \geqslant \omega_0$ (i.e., in the limit $t \longrightarrow \infty$), the contribution of the potential (33) associated with the quasi-parabolic structure is less noticeable ($z \sim 1$). Thus, for strong magnetic fields, the hypothesis on the parabolic nature of the confinement potential for a two-electron system is obviously quite justified. In our calculations, we used the following parameters typical of GaAs: effective mass $m^* = 0.067m_e$ and $g^* = -0.44$.

5. CONCLUSION

Proceeding from the assumption concerning the existence of an image charge which may be due to a large difference between the permittivities of the layers forming a quantum dot or the presence of an impurity in a semiconductor, we have formulated the model of a two-electron quantum dot. The model makes it possible to determine consistently the effective confinement potential for electrons, which is a function of the electron effective mass and the characteristics of the image charge.

Using the oscillator representation, we have analytically calculated the energy spectrum of a quantum dot for various values of the magnetic field. The division of the total Hamiltonian into the Hamiltonian of the center-of-mass motion and the Hamiltonian of relative motion gives two types of solutions. The energy spectrum of the center-of-mass Hamiltonian is harmonic and determined by the kinetic motion of the electrons, the cyclotron frequency, and the ratio of the electron effective mass to the sum of the masses of the two electrons and the image charge. It should be emphasized that in contrast to the Fock–Darwin case, this solution is independent of the confinement potential for electrons. The second solution determines the confinement energy. In the model proposed by us, the confinement potential is completely determined by the interaction of the electrons in the field of the image charge. The results of analysis lead to the conclusion that the confinement potential in a quantum dot may differ significantly from the parabolic potential (especially for small values of the magnetic field), which can be verified from an analysis of the spin oscillations of the ground state of a two-electron system in magnetic fields. The departure of the confinement potential from the parabolic confinement potential does not contradict the Kohn theorem [26], which is valid in the case of the electron–electron interaction, which is determined exclusively by the relative distance.

ACKNOWLEDGMENTS

This research was partly financed by the Russian Foundation for Basic Research (project no. 00-02-17194).

APPENDIX

An important element of calculations in the oscillator approximation [16] is the representation of canonical variables in normal form. We will give here the details of this representation for various potentials. Let us consider the quantity

$$I = \frac{q^2}{\sqrt{q^4 + 2\gamma x q^2 + \gamma^2}} = -\frac{\partial}{\partial\beta}\int_0^\infty \frac{dt}{\sqrt{\pi t}} \exp[-\beta q^2 - t(q^4 + 2\gamma x q^2 + \gamma^2)]\Big|_{\beta = 0}, \quad \text{(A.1)}$$

where q_j is a vector in the auxiliary space R^d. Taking into account the relations (q_j, $\eta_j \in R^d$)

$$\exp(-tq^4) = \int_{-\infty}^{\infty} \frac{d\tau}{\sqrt{\pi}} \exp(-\tau^2 - 2i\sqrt{t}\tau q^2),$$
$$\exp(-q^2\kappa) = \int_{-\infty}^{\infty} \left(\frac{d\eta}{\sqrt{\pi}}\right)^d \exp[-\eta^2 - 2i\sqrt{\kappa}(q\eta)], \tag{A.2}$$

substituting them into Eq. (A.1), and using expressions (21) for the canonical variable q_j, we carry out normal ordering over the creation (a_j^+) and annihilation (a_j) operators. This gives

$$I = -\frac{\partial}{\partial\beta}\int_0^{\infty} \frac{dt}{\sqrt{\pi t}} \exp(-\gamma^2 t) \int_{-\infty}^{\infty} \frac{d\tau}{\sqrt{\pi}} \exp(-\tau^2)$$
$$\times \int_{-\infty}^{\infty} \left(\frac{d\eta}{\sqrt{\pi}}\right)^d \exp\left[-\bar{\eta}^2\left(1 + \frac{\kappa}{\omega}\right)\right] : \exp[-2i\sqrt{\kappa}(q\eta)] :\Big|_{\beta = 0}, \tag{A.3}$$

where

$$\kappa = \beta + 2\gamma x t + 2i\sqrt{t}. \tag{A.4}$$

Using representation (A.3), we derive the expressions (24) for the ground-state energy $\varepsilon_0(E)$ and (25) for the interaction Hamiltonian H_I.

While analyzing Hamiltonian (44) with a quasi-parabolic potential, we must present in normal form the quantity $q^{2\tau}$, where τ may assume any value. For this purpose, we will use the relation

$$q^{2\tau} = \int_0^{\infty} \frac{dx}{\Gamma(-\tau)} x^{-1-\tau} \exp(-xq^2)$$
$$= \int_0^{\infty} \frac{dx}{\Gamma(-\tau)} x^{-1-\tau} \int_{-\infty}^{\infty} \left(\frac{d\eta}{\sqrt{\pi}}\right)^d$$
$$\times \exp\left[-\eta^2\left(1 + \frac{x}{\omega}\right)\right] : \exp[-2i\sqrt{x}(q\eta)] :$$
$$= \frac{1}{\omega^\tau}\frac{\Gamma(d/2 + \tau)}{\Gamma(d/2)} + : q^2 : \frac{\tau}{\omega^{\tau-1}}\frac{\Gamma(d/2 + \tau)}{\Gamma(d/2 + 1)}$$
$$+ \frac{1}{\omega^\tau}\int_0^{\infty} \frac{dx}{\Gamma(-\tau)} x^{-1-\tau} \int_{-\infty}^{\infty} \left(\frac{d\eta}{\sqrt{\pi}}\right)^d$$
$$\times \exp[-\eta^2(1 + x)] : \exp_2[-2i\sqrt{x\omega}(q\eta)] :. \tag{A.5}$$

We will now give the details of calculating integrals of the following form:

$$J = \int_0^{\infty}\int_0^{\infty} \frac{d\beta_1 d\beta_2}{\Gamma^2(d/2)}$$
$$\times \frac{(\beta_1\beta_2)^{d/2-1}(\beta_1 + \beta_2)\exp(-\omega_1\beta_1 - \omega_2\beta_2)}{\sqrt{\gamma^2 - 2\gamma(\beta_1 - \beta_2) + (\beta_1 + \beta_2)^2}}. \tag{A.6}$$

First of all, we carry out the substitution of variables:

$$s = \frac{\beta_1 - \beta_2}{\sqrt{2}}, \quad t = \frac{\beta_1 + \beta_2}{\sqrt{2}},$$
$$\beta_1 = \frac{s + t}{\sqrt{2}}, \quad \beta_2 = \frac{t - s}{\sqrt{2}}. \tag{A.7}$$

In this case, integral (A.6) can be written in the form

$$J = \frac{\gamma^d}{2^{d-1}}\int_0^{\infty} dt \int_{-1}^{1} dx \frac{(1 - x^2)^{d/2-1} t^d}{\Gamma^2(d/2)}$$
$$\times \frac{\exp[-\omega_+ t\gamma - \omega_- x t\gamma]}{\sqrt{1 - 2xt + t^2}}, \tag{A.8}$$

where

$$\omega_+ = \frac{\omega_1 + \omega_2}{2}, \quad \omega_- = \frac{\omega_1 - \omega_2}{2},$$
$$d = 2 + 2|m|.$$

We will consider the case when $\omega_- = 0$. In order to calculate integral (A.8), we will use the following relations [27]:

$$\frac{1}{\sqrt{1 - 2xt + t^2}} = \sum_{k=0}^{\infty} \begin{cases} t^k P_k(x), & |t| \le 1 \\ t^{-1-k} P_k(x), & |t| \ge 1, \end{cases}$$
$$\int_{-1}^{1} dx x^{2j} P_{2k}(x) = \frac{\Gamma(j + 1)}{\Gamma(1 + j - k)}\frac{\Gamma(1/2 + j)}{\Gamma(k + 3/2 + j)}. \tag{A.9}$$

Here, $P_{2k}(x)$ is a Legendre polynomial. Taking into account these representations, we obtain from (A.8)

$$J = \frac{\gamma^{2(|m|+1)}}{2^{2|m|+1}} \sum_{j=0}^{|m|} \frac{(-1)^j}{|m|!(|m| - j)!}$$
$$\times \sum_{k=0}^{j} \left\{ \int_0^1 dt \exp(-\omega_+ t\gamma)(t^{2|m|+2+2k} - t^{2|m|-2k+1}) \right.$$
$$\left. + \frac{\Gamma(2|m| - 2k + 2)}{(\gamma\omega_+)^{2(|m|+1-k)}} \right\} \frac{\Gamma(1/2 + j)}{\Gamma(1 + j - k)!\Gamma(k + 3/2 + j)}. \tag{A.10}$$

The next integral can be calculated exactly:

$$\int_0^1 dt t^n e^{-At} = (-1)^n \frac{\partial^n}{\partial A^n} \int_0^1 dt e^{-At} = (-1)^n \frac{\partial^n}{\partial A^n} \frac{1-e^{-A}}{A}, \quad n = 0, 1, \ldots \tag{A.11}$$

Using these relations, we arrive at expression (31) and (A.6)–(A.11) for the confinement potential (33).

REFERENCES

1. T. Chakraborty, Comments Condens. Matter Phys. **16**, 35 (1992); M. A. Kastner, Phys. Today **46**, 24 (1993).
2. R. Turton, *The Quantum Dot. A Journey into Future Microelectronics* (Oxford Univ. Press, New York, 1995).
3. L. Jacak, P. Hawrylak, and A. Wojs, *Quantum Dots* (Springer-Verlag, Berlin, 1997).
4. M. Macucci, K. Hess, and G. J. Iafrate, Phys. Rev. B **48**, 17354 (1993); J. Appl. Phys. **77**, 3267 (1995).
5. W. D. Heiss and R. G. Nazmitdinov, Phys. Lett. A **222**, 309 (1996); Phys. Rev. B **55**, 16310 (1997); Pis'ma Zh. Éksp. Teor. Fiz. **68**, 870 (1998) [JETP Lett. **68**, 915 (1998)].
6. S. Tarucha, D. G. Austing, T. Honda, *et al.*, Phys. Rev. Lett. **77**, 3613 (1996).
7. M. Maksym and T. Chakraborty, Phys. Rev. Lett. **65**, 108 (1990); Phys. Rev. B **45**, 1947 (1992).
8. U. Merkt, J. Huser, and M. Wagner, Phys. Rev. B **43**, 7320 (1991); M. Wagner, U. Merkt, and A. V. Chaplik, Phys. Rev. B **45**, 1951 (1992).
9. M. Dineykhan and R. G. Nazmitdinov, Phys. Rev. B **55**, 13707 (1997); J. Phys.: Condens. Matter **11**, L83 (1999).
10. Bo Su, V. J. Goldman, and J. E. Cunningham, Phys. Rev. B **46**, 7644 (1992).
11. R. C. Ashoori, H. L. Stormer, J. S. Weiner, *et al.*, Phys. Rev. Lett. **71**, 613 (1993); R. C. Ashoori, Nature (London) **379**, 413 (1996).
12. T. Demel, D. Heitmann, P. Grambow, and K. Ploog, Phys. Rev. Lett. **64**, 788 (1990).
13. D. Phannkuche and R. R. Gerhardts, Phys. Rev. B **43**, 12098 (1991); **44**, 13132 (1991).
14. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, New York, 1984).
15. N. A. Gippius, V. D. Kulakovskiĭ, and S. G. Tikhodeev, Usp. Fiz. Nauk **167**, 558 (1997) [Phys. Usp. **40**, 534 (1997)]; E. A. Mulyarov and S. G. Tikhodeev, Zh. Éksp. Teor. Fiz. **111**, 274 (1997) [JETP **84**, 151 (1997)].
16. M. Dineykhan and G. V. Efimov, Fiz. Élem. Chastits At. Yadra **26**, 651 (1995) [Phys. Part. Nucl. **26**, 275 (1995)]; M. Dineykhan, G. V. Efimov, G. Ganbold, and S. N. Nedelko, in *Lecture Notes in Physics*, Vol. 26: *Oscillator Representation in Quantum Physics* (Springer-Verlag, Berlin, 1995).
17. M. Dineykhan and R. G. Nazmitdinov, Yad. Fiz. **62**, 143 (1999) [Phys. At. Nucl. **62**, 138 (1999)].
18. V. Fock, Z. Phys. **47**, 446 (1928); C. G. Darwin, Proc. Cambridge Phil. Soc. **27**, 86 (1930).
19. I. V. Komarov, L. I. Ponomarev, and S. Yu. Slavyanov, *Spheroidal and Coulomb Spheroidal Functions* (Nauka, Moscow, 1976); S. I. Vinitskiĭ and L. I. Ponomarev, Fiz. Élem. Chastits At. Yadra **13**, 1336 (1982) [Sov. J. Part. Nucl. **13**, 557 (1982)].
20. V. A. Fock, *The Principles of Quantum Mechanics* (Nauka, Moscow, 1976; Mir, Moscow, 1978).
21. V. A. Fok, Izv. Akad. Nauk SSSR, Ser. Fiz. **18**, 161 (1954).
22. M. Dineykhan and G. V. Efimov, Rep. Math. Phys. **36**, 287 (1995); Yad. Fiz. **59**, 862 (1996) [Phys. At. Nucl. **59**, 824 (1996)].
23. E. S. Fradkin, Nucl. Phys. **49**, 624 (1963); K. Hayashi, M. Hirayama, T. Muta, *et al.*, Fortschr. Phys. **15**, 625 (1967); A. Salam, *Nonpolynomial Lagrangians. Renormalization and Gravity* (Gordon and Breach, New York, 1971).
24. M. Dineykhan, Z. Phys. D **41**, 77 (1997).
25. F. J. Dyson, Phys. Rev. **85**, 631 (1952).
26. W. Kohn, Phys. Rev. **123**, 1242 (1961).
27. I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Fizmatgiz, Moscow, 1962; Academic, New York, 1980).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 1060–1071.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1222–1234.

SOLIDS
Electronic Properties

Optical Size Resonances in Nanostructures

O. N. Gadomsky* and T. T. Idiatullov
Ul'yanovsk State University, Ul'yanovsk, 432700 Russia
**@e-mail: qed-group@mail.ru*
Received November 10, 2000

Abstract—The existence of optical size resonances in atomic nanostructures is proved. The properties of optical size resonances strongly depend on the interatomic distances and on the polarization of an external radiation field. The properties of linear and nonlinear size resonances are considered in the case of two-dimensional nanostructures. The linear optical size resonances are described based on a closed system of equations for dipole oscillators and nonlocal field equations taking into account the dipole–dipole interactions of atoms in the radiation field. Using a stationary solution to these equations, it is demonstrated that two isotropic atoms with definite intrinsic frequencies form an anisotropic system in the radiation field, possessing two or four size resonances depending on whether the component atoms are identical or different. The nanostructure composed of two different atoms possesses two size resonances with positive dispersion and two other resonances with negative dispersion. The frequencies of the size resonances significantly differ from the intrinsic frequencies of isolated atoms entering into the nanostructure. By changing the angle of incidence of the external wave, it is possible to excite various size resonances. The properties of nonlinear optical size resonances excited by an intense radiation field were theoretically and numerically studied using the modified Bloch equations and nonlocal field equations. Dispersion relationships for the nonlinear resonances were derived and the inversion properties of atoms in the nanostructure were studied for various polarizations of the external optical wave. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The optical properties of nanostructures obtained using modern technologies, including aggregates [1, 2] and dimers [3, 4] on the clean surface of solids, ultrathin films [5, 6], quantum wires [7] and quantum dots [8], near-surface atomic probe [9, 10], surface clusters [11–13], etc., can serve as a basis for the development of unique optical methods of investigation and for the creation of unique optical devices. This paper is devoted to a theoretical study of the interaction of two-dimensional nanostructures with optical radiation. An example of such a nanostructure is offered by a dimer on the solid surface [4]. In our opinion, a complete description of the optical properties of such nanostructures provides a basis for the knowledge of optical properties of more complicated nanostructures.

Previously [14, 15], the near-field effect was predicted based on the description of dielectrics as a discrete-continuous medium in contrast to the traditional concept of a continuous dielectric [10, 16]. In [14, 15] and in the subsequent papers [5, 6, 17], it was shown that the structural factor related to a discrete distribution of atoms near the observation point significantly affects the reflection and transmission properties of the surface, the index of refraction of the subsurface layer, the optical properties of thin films, and the lifetime of an excited atom at the surface. In [18], the near-field effect was used as a basis for recording quantum information on individual atoms of a diatomic quantum computer with the aid of intense quasiresonance radiation at a variable angle of incidence of an external electromagnetic wave. The readout of the recorded quantum information can be performed with the aid of probing radiation, based on a solution obtained in [19] for a system of equations for dipole oscillators.

In this paper, we will demonstrate that the internal interaction between atoms in the radiation field may significantly change the spectral properties of nanostructures exposed to both weak and strong optical fields. It will be shown that two identical or different atoms possessing isolated resonances may acquire the so-called size resonances in the field of optical radiation, the frequencies of which strongly depend on the interatomic distance and on the orientation of the nanostructure axis relative to the direction of propagation of the external wave.

2. LINEAR OPTICAL SIZE RESONANCES IN DIATOMIC NANOSTRUCTURES

Let the origin of a coordinate system coincide with atom 1 of a small diatomic object (nanoobject). Then the vectors $\mathbf{r}_1(0, 0, 0)$ and $\mathbf{r}_2(0, R, 0)$ determine the positions of two atoms constituting this object, the linear size of which is assumed to be considerably smaller compared to the external radiation wavelength: $k_0 R \ll 1$, where $k_0 = \omega/c$, ω is the circular frequency of the external radiation, and c is the speed of light in vacuum.

We assume the positions of atoms in the object to be fixed, for example, on a substrate, and the substrate

effect to be taken into account in the corresponding values of the intrinsic frequencies of atoms and of the widths of their energy levels.

The electric field strength of a light wave at any observation point $\mathbf{r}$ at a tine instant t is determined by the following equation:

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_I(\mathbf{r}, t) + \sum_{j=1}^{2} \mathrm{rot}\,\mathrm{rot}\frac{\mathbf{p}_j(t - R_j/c)}{R_j}, \quad (1)$$

where $\mathbf{E}_I(\mathbf{r}, t)$ is the electric field strength of the external wave, $R_j = |\mathbf{r} - \mathbf{r}_j|$, and $\mathbf{p}_j$ is the induced dipole moment of the jth atom.

The dipole moments of atoms can be represented in the following form:

$$\mathbf{p}_j = e(\mathbf{u}_j - i\mathbf{v}_j)\exp(-i\omega t) + \text{c.c.}, \quad (2)$$

where e is the electron charge; $\mathbf{u}_j$ and $\mathbf{v}_j$ ($j = 1, 2$) are some vector functions of time, the meaning of which will be explained below. The equations of motion for the dipole moments (2) are as follows:

$$\begin{aligned}\ddot{\mathbf{p}}_1 + \omega_{01}^2\mathbf{p}_1 + \gamma_{01}\dot{\mathbf{p}}_1 + \gamma_{02}\mathbf{p}_2 &= \frac{e^2}{m}\mathbf{E}(\mathbf{r}_1, t),\\ \ddot{\mathbf{p}}_2 + \omega_{02}^2\mathbf{p}_2 + \gamma_{02}\dot{\mathbf{p}}_2 + \gamma_{01}\mathbf{p}_1 &= \frac{e^2}{m}\mathbf{E}(\mathbf{r}_2, t),\end{aligned} \quad (3)$$

where ω_{01} and ω_{02} are the intrinsic frequencies of atoms 1 and 2, respectively; m is the electron mass; $\mathbf{E}(\mathbf{r}_j, t)$ is the field strength at the site of the jth atom; and $\gamma_{0j} = 2e^2\omega_{0j}^2/3mc^3$. The terms proportional to $\dot{\mathbf{p}}_j$ in Eqs. (3) take into account a radiation-induced retardation in the diatomic system [20].

Let the field $\mathbf{E}(\mathbf{r}_j, t)$ in Eqs. (3) have the form

$$\mathbf{E}(\mathbf{r}_j, t) = \mathbf{E}_0(\mathbf{r}_j, t)\exp(-i\omega t) + \text{c.c.} \quad (4)$$

In addition, we assume that the functions $\mathbf{u}_j$ and $\mathbf{v}_j$ are slowly varying with time, so that

$$|\dot{\mathbf{u}}_j| \ll \omega|\mathbf{u}_j|, \quad |\dot{\mathbf{v}}_j| \ll \omega|\mathbf{v}_j|. \quad (5)$$

Upon substituting expressions (2) into Eqs. (3), conditions (5) allow the terms proportional to $\ddot{\mathbf{u}}_j$ and $\ddot{\mathbf{v}}_j$ to be rejected. Let us introduce the notation

$$\mathbf{u}_1 - i\dot{\mathbf{v}}_1 = \mathbf{X}_1, \quad \mathbf{u}_2 - i\dot{\mathbf{v}}_2 = \mathbf{X}_2.$$

Omitting the factor $\exp(-i\omega t)$, we obtain the following set of equations from system (3):

$$\begin{aligned}&2\dot{\mathbf{X}}_1(-i\omega + \gamma_{01}) + \mathbf{X}_1(-\omega^2 + \omega_{01}^2 - i\omega\gamma_{01})\\ &+ \gamma_{02}[\dot{\mathbf{X}}_2] - \gamma_{02}i\omega[\mathbf{X}_2] = \frac{e}{m}\mathbf{E}(\mathbf{r}_1, t),\\ &2\dot{\mathbf{X}}_2(-i\omega + \gamma_{02}) + \mathbf{X}_2(-\omega^2 + \omega_{02}^2 - i\omega\gamma_{02})\\ &+ \gamma_{01}[\dot{\mathbf{X}}_1] - \gamma_{01}i\omega[\mathbf{X}_1] = \frac{e}{m}\mathbf{E}(\mathbf{r}_2, t),\end{aligned} \quad (6)$$

where the symbol $[\ldots]$ indicates that the corresponding value is determined at the time instant $t - R/c$. Substituting expressions for the fields $\mathbf{E}_0(\mathbf{r}_1, t)$ and $\mathbf{E}_0(\mathbf{r}_2, t)$ depending on the dipole moments of atoms 1 and 2 into these equations, we obtain a self-consistent system of equations determining the unknown functions $\mathbf{X}_1$ and $\mathbf{X}_2$ at an arbitrary moment of time for a diatomic nanoobject excited by pulsed or continuous fields.

2.1. Stationary Regime of Excitation of a Small Object

In the case of continuous irradiation of atoms of the small object studied, we will assume that

$$\dot{\mathbf{X}}_1 = \dot{\mathbf{X}}_2 = 0 \quad (7)$$

and that the fields $\mathbf{E}_0(\mathbf{r}_1, t)$ and $\mathbf{E}_0(\mathbf{r}_2, t)$ are independent of time. Taking this into account in Eqs. (6), we obtain the following system of equations:

$$\begin{aligned}&\mathbf{X}_1(-\omega^2 + \omega_{01}^2 - i\omega\gamma_{01})\\ &- \gamma_{02}i\omega\mathbf{X}_2\exp(ik_0R) = \frac{e}{m}\mathbf{E}(\mathbf{r}_1),\\ &\mathbf{X}_2(-\omega^2 + \omega_{02}^2 - i\omega\gamma_{02})\\ &- \gamma_{01}i\omega\mathbf{X}_1\exp(ik_0R) = \frac{e}{m}\mathbf{E}(\mathbf{r}_2).\end{aligned} \quad (8)$$

According to Eq. (1),

$$\begin{aligned}E_0^y(1) &= E_{0I}^y + 2GeX_2^y\exp(ik_0R),\\ E_0^\beta(1) &= E_{0I}^\beta - FeX_2^\beta\exp(ik_0R),\\ E_0^y(2) &= E_{0I}^y\exp(i\mathbf{k}_0\cdot\mathbf{R}) + 2GeX_1^y\exp(ik_0R),\\ E_0^\beta(2) &= E_{0I}^\beta\exp(i\mathbf{k}_0\cdot\mathbf{R}) - FeX_1^\beta\exp(ik_0R),\end{aligned} \quad (9)$$

where $\beta = x, z$; $E_0^y(j) = E_0^y(\mathbf{r}_j)$; $\gamma = x, y, z$; the quantities

$$G = \frac{1}{R^3} - i\frac{k_0}{R^2}, \quad F = G - \frac{k_0^2}{R} \quad (10)$$

determine contributions of the Coulomb and retarding polarization fields in Eqs. (8).

It is convenient to solve Eqs. (8) with an allowance for relationships (9) and (10) separately for the y and β components of the dipole moments. Taking into account that $p_{0j}^y = eX_j^y$, we obtain the following formulas:

$$\begin{aligned}p_{01}^y &= \alpha_{\text{eff}}^y(1)E_{0I}^y, \quad p_{02}^y = \alpha_{\text{eff}}^y(2)E_{0I}^y,\\ p_{01}^y &= \alpha_{\text{eff}}^\beta(1)E_{0I}^\beta, \quad p_{02}^\beta = \alpha_{\text{eff}}^\beta(2)E_{0I}^\beta,\end{aligned} \quad (11)$$

where α_{eff} are the effective polarizabilities taking into account the mutual influence of atoms 1 and 2 in the diatomic object studied. These quantities are expressed as

$$\alpha_{\text{eff}}^{y}(1) = \alpha_1 \frac{1 + \alpha_2\left(2G + \frac{2i}{3}k_0k_{02}^2\right)\exp(i\mathbf{k}_0\cdot\mathbf{R})\exp(ik_0R)}{1 - \alpha_1\alpha_2\left(2G + \frac{2i}{3}k_0k_{01}^2\right)\left(2G + \frac{2i}{3}k_0k_{02}^2\right)\exp(2ik_0R)},$$

$$\alpha_{\text{eff}}^{y}(2) = \alpha_2 \frac{\exp(i\mathbf{k}_0\cdot\mathbf{R}) + \alpha_1\left(2G + \frac{2i}{3}k_0k_{01}^2\right)\exp(ik_0R)}{1 - \alpha_1\alpha_2\left(2G + \frac{2i}{3}k_0k_{01}^2\right)\left(2G + \frac{2i}{3}k_0k_{02}^2\right)\exp(2ik_0R)},$$

$$\alpha_{\text{eff}}^{\beta}(1) = \alpha_1 \frac{1 - \alpha_2\left(F - \frac{2i}{3}k_0k_{02}^2\right)\exp(i\mathbf{k}_0\cdot\mathbf{R})\exp(ik_0R)}{1 - \alpha_1\alpha_2\left(F - \frac{2i}{3}k_0k_{01}^2\right)\left(F - \frac{2i}{3}k_0k_{02}^2\right)\exp(2ik_0R)}, \quad (12)$$

$$\alpha_{\text{eff}}^{\beta}(2) = \alpha_2 \frac{\exp(i\mathbf{k}_0\cdot\mathbf{R}) - \alpha_1\left(F - \frac{2i}{3}k_0k_{01}^2\right)\exp(ik_0R)}{1 - \alpha_1\alpha_2\left(F - \frac{2i}{3}k_0k_{01}^2\right)\left(F - \frac{2i}{3}k_0k_{02}^2\right)\exp(2ik_0R)},$$

where $k_{01} = \omega_{01}/c$, $k_{02} = \omega_{02}/c$; α_1 and α_2 are the isotropic polarizabilities of atoms 1 and 2 determined without allowing for the mutual influence. For atoms 1 and 2 with the isolated resonances ω_{01} and ω_{02}, the isotropic polarizabilities are given by the formula

$$\alpha_j = \frac{e^2}{m}\frac{1}{\omega_{0j}^2 - \omega^2 - i\omega\gamma_{0j}}. \quad (13)$$

Let us consider two different atoms possessing isolated resonances with the wavenumbers $k_{01} = 89000$ cm^{-1} and $k_{02} = 92000$ cm^{-1}. An analysis of formulas (12) shows that the size resonances in a diatomic object with $R = 1$ nm correspond to $k'_{01} = 86988.152$ cm^{-1}, $k'_{02} = 93904.548$ cm^{-1}, $k'_{03} = 88324.489$ cm^{-1}, and $k'_{04} = 92648.712$ cm^{-1}. The full width at half-maximum (FWHM) of these size resonances corresponds to $\Delta k'_{01} = 0.00281$ cm^{-1}, $\Delta k'_{02} = 0.00007$ cm^{-1}, $\Delta k'_{03} = 0.00045$ cm^{-1}, and $\Delta k'_{04} = 0.00273$ cm^{-1}. The values of y and β components of the effective polarizabilities of atoms 1 and 2 in the diatomic nanoobject differ from one another, which indicates the presence of an induced anisotropy in the nanostructure studied. Moreover, the size resonances k'_{02} and k'_{03} possess a negative dispersion, while the dispersion of other size resonances is positive. In the case of two identical atoms with $\alpha_1 = \alpha_2 = \alpha$ and $\omega_{01} = \omega_{02} = \omega_0$, expressions (12) yield

$$\alpha_{\text{eff}}^{y} = \frac{\alpha}{1 - \alpha\left(2G + \frac{2i}{3}k_0k_{01}^2\right)\exp(ik_0R)},$$

$$\alpha_{\text{eff}}^{\beta} = \frac{\alpha}{1 + \alpha\left(F + \frac{2i}{3}k_0k_{01}^2\right)\exp(ik_0R)}, \quad (14)$$

where it is assumed that $\mathbf{k}_0 \cdot \mathbf{R} = 0$, which means that the incident wave is directed along the normal to the diatomic object axis. A numerical analysis shows that the effective polarizabilities of atoms 1 and 2 in the diatomic nanoobject studied for various frequencies of the external field are virtually independent of the angle between vectors $\mathbf{k}_0$ and $\mathbf{R}$. A consistent interaction of atoms in the object studied leads to the appearance of two size resonances with a positive dispersion and the wavenumbers 85775 and 95569 cm^{-1} for $k_{01} = \omega/c = 89000$ cm^{-1}.

2.2. Optical Field Outside a Small Object

Let us calculate the field at an observation point r in the wave zone where $k_0|\mathbf{r} - \mathbf{r}_1| \gg 1$ and $k_0|\mathbf{r} - \mathbf{r}_2| \gg 1$. For such observation points, the major role in Eq. (1) belongs to the terms proportional to $1/R_j$. The electric

and magnetic field strengths at the observation point **r** are given by the expressions [16]:

$$\mathbf{E}(\mathbf{r}, t) = -\frac{\mathbf{n}_1 \times [\mathbf{n}_1 \times \mathbf{p}_1]}{c^2 R_1} - \frac{\mathbf{n}_2 \times [\mathbf{n}_2 \times \mathbf{p}_2]}{c^2 R_2},$$
$$\mathbf{H}(\mathbf{r}, t) = \frac{\mathbf{p}_1 \times \mathbf{n}_1}{c^2 R_1} + \frac{\mathbf{p}_2 \times \mathbf{n}_2}{c^2 R_2}. \quad (15)$$

The induced dipole moments of atoms 1 and 2 in the small diatomic object studied are determined by formulas (15) at the time instants $t - R_1/c$ and $t - R_2/c$. The expressions for these moments are as follows:

$$p_j^\gamma = -\omega^2 p_{0j}^y \exp(ik_0 R_j) \exp(-i\omega t) + \text{c.c.}, \quad (16)$$

where γ indicates the component (x, y, or z) of the induced dipole moment of the jth atom and $R_j = |\mathbf{r} - \mathbf{r}_j|$. The field strength at the observation point **r** can be determined by substituting expressions (15) into the energy conservation law written in terms of the Poynting vector and a unit normal vector for the surface of a sphere with the radius R_1. In the wave zone, however, the E and H values are approximately equal and the intensity of dipole radiation of a diatomic nanoobject is given be the formula

$$dI = \frac{\omega^4 e^2}{4\pi c^3} |\mathbf{X}_1 \exp(-ik_0 R_1) + \mathbf{X}_2 \exp(-ik_0 R_2)|^2 \sin^2\theta_0 d\Omega_0, \quad (17)$$

where θ_0 is the angle between vectors $\mathbf{X}_1$ and $\mathbf{R}_1$, $d\Omega_0$ is the element of a solid angle in the direction of vector $\mathbf{R}_1$, and $\mathbf{X}_1$ and $\mathbf{X}_2$ are vectors determined for the effective polarizabilities.

A numerical analysis of formula (17) showed that the direction of maximum dipole radiation is determined by the $\sin^2\theta_0$ law for various values of amplitude of an external field propagating along the z axis. For the intrinsic resonance frequency, when $\mathbf{k}_0 = (0, 0, 89000\ \text{cm}^{-1})$ and $\mathbf{E}_{0I} = (1/\sqrt{2}, 1/\sqrt{2}, 0)$ CGSE units, the $I_{\max}$ value at $\theta_0 = \pi/2$ is 1.64×10^{-13} erg/s. For the size resonance frequency, where $k_0 = (0, 0, 93904\ \text{cm}^{-1})$, the same $\mathbf{E}_{0I}$ at $\theta_0 = \pi/2$ yields $I_{\max} = 1.62$ erg/s. For the other size resonance with $k_0 = (0, 0, 88324\ \text{cm}^{-1})$, we obtain $I_{\max} = 0.28$ erg/s.

Thus, by changing the external field frequency so as to make it coinciding with one of the size resonance frequencies of the nanoobject studied, we increase the dipole radiation intensity by approximately 12 orders of magnitude.

3. NONLINEAR INTERACTION OF OPTICAL RADIATION WITH A DIATOMIC NANOOBJECT

Now let us consider the properties of a nanoobject consisting of two different atoms in the field of an intense wave. The Hamiltonian of the system is

$$H = \frac{1}{2}\hbar\omega_{01}\hat{I} + \frac{1}{2}\hbar\omega_{02}\hat{I} + \frac{1}{2}\hbar\omega_{01}\sigma_{31} + \frac{1}{2}\hbar\omega_{02}\sigma_{32} - \mathbf{d}'_{01} \cdot \mathbf{E}'_1\sigma_{11} + \mathbf{d}''_{01} \cdot \mathbf{E}_1\sigma_{21} - \mathbf{d}'_{02} \cdot \mathbf{E}'_2\sigma_{12} + \mathbf{d}''_{02} \cdot \mathbf{E}_2\sigma_{22}, \quad (18)$$

where $\hat{I}$ is the unitary operator; ω_{01} and ω_{02} are the frequencies of intrinsic transitions in atoms 1 and 2, respectively; $\mathbf{d}_1$ and $\mathbf{d}_2$ are the dipole moment operators of the atoms

$$\mathbf{d}_1 = \mathbf{d}'_{0j}\sigma_{11} - \mathbf{d}''_{01}\sigma_{21}, \quad \mathbf{d}_2 = \mathbf{d}'_{0j}\sigma_{12} - \mathbf{d}''_{02}\sigma_{22} \quad (19)$$

determined by the real ($\mathbf{d}'_{01}$, $\mathbf{d}'_{02}$) and imaginary ($\mathbf{d}''_{01}$, $\mathbf{d}''_{02}$) parts of the complex dipole transition moments of atoms 1 and 2; and $\sigma_{\alpha j}$ are the effective spin operators of the jth atom ($j = 1, 2$; $\alpha = 1, 2, 3$) obeying the relationships

$$[\sigma_{1j}, \sigma_{2k}] = 2i\sigma_{3j}\delta_{jk}, \quad [\sigma_{2j}, \sigma_{3k}] = 2i\sigma_{1j}\delta_{jk}, \quad [\sigma_{3j}, \sigma_{1k}] = 2i\sigma_{2j}\delta_{jk}, \quad (20)$$

where δ_{jk} is the Kronecker delta symbol. The electric field strengths E_1 and E_2 at the sites of atoms 1 and 2 satisfy Eqs. (1), where the induced dipole moments $\mathbf{p}_1$ and $\mathbf{p}_2$ of the atoms depend nonlinearly on these fields. For the same arrangement of atoms in the system as in Section 2, the field components $\mathbf{E}_1$ and $\mathbf{E}_2$ are described by expressions (9) with the corresponding P_j^γ substituted for eX_j^α, where

$$\mathbf{P}_j = \mathbf{d}'_{0j}\langle\sigma_{1j}\rangle - \mathbf{d}''_{0j}\langle\sigma_{2j}\rangle, \quad (21)$$

and $\langle\sigma_{\alpha j}\rangle$ are the average values of the corresponding operators $\sigma_{\alpha j}$. Omitting the symbol $\langle\ldots\rangle$ and using the commutation relationships (20), we obtain the following equations of motion:

$$\dot{\sigma}_{1j} = -\omega_{0j}\sigma_{2j} + \frac{2}{\hbar}\mathbf{d}''_{0j} \cdot \mathbf{E}_j\sigma_{3j},$$
$$\dot{\sigma}_{2j} = \omega_{0j}\sigma_{1j} + \frac{2}{\hbar}\mathbf{d}'_{0j} \cdot \mathbf{E}_j\sigma_{3j}, \quad (22)$$
$$\dot{\sigma}_{3j} = -\frac{2}{\hbar}\mathbf{d}''_{0j} \cdot \mathbf{E}_j\sigma_{1j} - \frac{2}{\hbar}\mathbf{d}'_{0j} \cdot \mathbf{E}_j\sigma_{2j},$$

where

$$\mathbf{E}_j = (\mathbf{e}' - i\mathbf{e}'')(E'_{0j} - iE''_{0j}) \exp(-i\omega t) + \text{c.c.}, \quad (23)$$

$\mathbf{e}' - i\mathbf{e}''$ is the complex polarization vector of the field on atoms of the object, $\mathbf{E}'_{0j} - i\mathbf{E}''_{0j}$ is the complex field amplitude, and ω is the external field frequency. We introduce the following notation:

$$\chi_{1j} = a_{1j}\cos\omega t - a_{2j}\sin\omega t,$$
$$\chi_{2j} = a_{3j}\cos\omega t - a_{4j}\sin\omega t, \tag{24}$$

where

$$a_{1j} = \frac{2}{\hbar}\{\mathbf{d}'_{0j}\cdot\mathbf{e}'E'_{0j} - \mathbf{d}'_{0j}\cdot\mathbf{e}''E''_{0j}\},$$
$$a_{2j} = \frac{2}{\hbar}\{(\mathbf{d}'_{0j}\cdot\mathbf{e}'E''_{0j}) + \mathbf{d}'_{0j}\cdot\mathbf{e}''E'_{0j}\},$$
$$a_{3j} = \frac{2}{\hbar}\{\mathbf{d}''_{0j}\cdot\mathbf{e}'E'_{0j} - \mathbf{d}''_{0j}\cdot\mathbf{e}''E''_{0j}\}, \tag{25}$$
$$a_{4j} = \frac{2}{\hbar}\{\mathbf{d}''_{0j}\cdot\mathbf{e}'E''_{0j} + \mathbf{d}''_{0j}\cdot\mathbf{e}''E'_{0j}\}.$$

In addition, we introduce the following transformation:

$$\sigma_{1j} = u_j\cos\omega t - v_j\sin\omega t,$$
$$\sigma_{2j} = u_j\sin\omega t + v_j\cos\omega t, \quad \sigma_{3j} = w_j. \tag{26}$$

Using Eqs. (22) with notations (24) and (25) and taking into account the conditions

$$a_{3j} = -a_{2j}, \quad a_{4j} = a_{1j}, \tag{27}$$

we obtain a system of equations

$$\dot{u}_j = v_j(-\omega_{0j} + \omega) - a_{2j}w_j,$$
$$\dot{v}_j = u_j(\omega_{0j} - \omega) + a_{1j}w_j, \tag{28}$$
$$\dot{w}_j = a_{2j}u_j - a_{1j}v_j.$$

Conditions (27) imply that, for $(\mathbf{e}')^2 + (\mathbf{e}'')^2 = 1$,

$$\mathbf{d}''_{0j}\cdot\mathbf{e}'' = \mathbf{d}'_{0j}\cdot\mathbf{e}', \quad \mathbf{d}'_{0j}\cdot\mathbf{e}'' = \mathbf{d}''_{0j}\cdot\mathbf{e}' = 0, \tag{29}$$

which shows that the induced dipole moments of the nanoobject are parallel to the complex polarization vector inside the object.

In the general case, Eqs. (28) must be supplemented with relaxation terms containing the characteristic times of phase (T'_{2j}) and energy (T'_{1j}) relaxation of the jth atom in the nanoobject. Finally, we obtain a set of equations

$$\dot{u}_j = -\Delta_j v_j - a_{2j}w_j - \frac{u_j}{T'_{2j}},$$
$$\dot{v}_j = \Delta_j u_j + a_{1j}w_j - \frac{v_j}{T'_{2j}}, \tag{30}$$
$$\dot{w}_j = a_{2j}u_j - a_{1j}v_j - \frac{w_j - w_{0j}}{T_{1j}},$$

where $\Delta_j = \omega_{0j} - \omega$ is the detuning of the jth atom from resonance and ω_{0j} is the initial value of inversion of the jth atom.

Let us separate the variables u_j and v_j in the values a_{1j} and a_{2j} entering into Eqs. (28) and (30). Taking into account Eqs. (1), we obtain the following expression for the field at the site of atom 1:

$$\mathbf{E}_1 = \mathbf{y}_0(E_1^y(1) + 2G\exp(ik_0R)\exp(-i\omega t)p_{02}^y)$$
$$+ \mathbf{x}_0(E_I^x(1) - F\exp(ik_0R)\exp(-i\omega t)p_{02}^x) \tag{31}$$
$$+ \mathbf{z}_0(E_I^z(1) - F\exp(ik_0R)\exp(-i\omega t)p_{02}^z) + \text{c.c.},$$

where $\mathbf{x}_0$, $\mathbf{y}_0$, $\mathbf{z}_0$ are the unit vectors of the corresponding coordinate axes and

$$\mathbf{p}_{02} = \frac{1}{2}(u_2 - iv_2)(\mathbf{d}'_{02} - i\mathbf{d}''_{02}),$$
$$\mathbf{E}_I(1) = (\mathbf{e}'_I - i\mathbf{e}''_I)E_{0I}(1)\exp(-i\omega t). \tag{32}$$

Note that the complex polarization vector $\mathbf{e}'_I - i\mathbf{e}''_I$ of the external wave in our analysis may, in the general case, not coincide with the polarization vector $\mathbf{e}' - i\mathbf{e}''$ of the field inside the nanoobject. An analogous expression can be obtained for the field E_2 at the site of atom 2. Substituting $\mathbf{E}_I(1) \longrightarrow \mathbf{E}_I(2)$ and $\mathbf{p}_{02} \longrightarrow \mathbf{p}_{01}$, we obtain

$$\mathbf{E}_2 = \mathbf{y}_0(E_{0I}\exp(i\mathbf{k}_0\cdot\mathbf{R})\exp(-i\omega t)(e'_{Iy} - ie''_{Iy})$$
$$+ 2G\exp(ik_0R)\exp(-i\omega t)p_{01}^y)$$
$$+ \mathbf{x}_0(E_{0I}\exp(i\mathbf{k}_0\cdot\mathbf{R})\exp(-i\omega t)(e'_{Ix} - ie''_{Ix}) \tag{33}$$
$$- F\exp(ik_0R)\exp(-i\omega t)p_{01}^x)$$
$$+ \mathbf{z}_0(E_{0I}\exp(i\mathbf{k}_0\cdot\mathbf{R})\exp(-i\omega t)(e'_{Iz} - ie''_{Iz})$$
$$- F\exp(ik_0R)\exp(-i\omega t)p_{01}^z) + \text{c.c.},$$

where E_{0I} is the real amplitude of the external wave and $\mathbf{k}_0$ is the wavevector of this wave with the modulus $k_0 = \omega/c$. Expressions (31) and (33) are valid for diatomic systems the dimensions of which are either smaller than or comparable with the wavelength of the external radiation. These expressions can be transformed in the following manner. Let us separate a negative-frequency part proportional to $\exp(-i\omega t)$ and take a scalar product of both parts of these expressions by $(\mathbf{e}' + i\mathbf{e}'')$. Taking into account that $\mathbf{k}_0\cdot\mathbf{R} = 0$, $k_0R \longrightarrow 0$ and neglecting the retarding interaction of atoms, we may put $F = G = 1/R^3$ in Eqs. (31) and (33). Eventually, we arrive at the equations

$$E'_{0j} = AE_{0I} + Bu_j, \quad E''_{0j} = CE_{0I} + Bv_j, \tag{34}$$

determining the quantities a_{1j} and a_{2j}. For $k \neq j$, the coefficients in (34) are given by the formulas

$$A = \mathbf{e}' \cdot \mathbf{e}'_I + \mathbf{e}'' \cdot \mathbf{e}''_I, \quad C = \mathbf{e}' \cdot \mathbf{e}''_I - \mathbf{e}'' \cdot \mathbf{e}'_I,$$
$$B = d_{0k} G \Big\{ [(e'_y)^2 + (e''_y)^2] - \frac{1}{2}[(e'_x)^2 + (e''_x)^2] - \frac{1}{2}[(e'_z)^2 + (e''_z)^2] \Big\}, \tag{35}$$

where d_{0k} is the modulus of the dipole transition moment of the kth atom.

The field dependence in the equations of motion (28) and (30) is determined by the values

$$a_{1j} = \frac{2}{\hbar}\mathbf{d}'_{0j} \cdot \mathbf{e}'(AE_{0I} + Bu_k),$$
$$a_{2j} = \frac{2}{\hbar}\mathbf{d}'_{0j} \cdot \mathbf{e}'(CE_{0I} + Bv_k), \tag{36}$$

which strongly depend on the interatomic distance and the orientation of dipole moments relative to the polarization vector of the external wave. Expressions (30) and (34) form a closed system of equations for determining the field and atomic variables of the diatomic nanoobject studied. Below, we will obtain some partial solution to these equations.

3.1. Stationary Solution

Let us consider a stationary solution to Eqs. (30) and (34) for the conditions $\dot{u}_j = \dot{v}_j = \dot{w}_j = 0$. Equation (34), with an allowance for (36), leads to a system of nonlinear algebraic equations. The u_j and v_j values are given by the formulas

$$u_j = -\frac{\Delta_j a_{1j} T'^2_{2j} + a_{2j} T'_{2j}}{D_j} w_{0j},$$
$$v_j = \frac{a_{1j} T'_{2j} - \Delta_j a_{2j} T'^2_{2j}}{D_j} w_{0j}, \tag{37}$$
$$D_j = 1 + \Delta_j^2 T'^2_{2j} + (a_{1j}^2 + a_{2j}^2) T_{1j} T'_{2j}.$$

Substituting expressions (36) into these relationships, we obtain a closed system of equations for the unknowns u_1, v_1 and u_2, v_2, which can be solved numerically for various values of the parameters B and C determined by the direction of field polarization inside the nanoobject. Using the u_j and v_j determined from these equations, we may find the inversions ω_1 and ω_2 of the atoms by the formula

$$w_j = \frac{1 + \Delta_j^2 T'^2_{2j}}{D_j} w_{0j}.$$

In the case of a nanoobject comprising atoms 1 and 2 exposed to a weak field such that $(a_{1j}^2 + a_{2j}^2) T_{1j} T'_{2j} \ll 1$, we may use expressions (37) for solving Eqs. (12).

3.2. Nonstationary Solution for $T_{1j} \longrightarrow \infty$, $T'_{2j} \longrightarrow \infty$

Let us solve Eqs. (28) (subscript j omitted) on assuming that the light pulse duration is relatively short $\tau \ll T_1$, T'_2 and we may neglect the relaxation processes. Moreover, we will also assume that the values a_1 and a_2 remain constant (time-independent) during the time interval τ.

Let us write Eqs. (28) in the matrix form:

$$\frac{d}{dt}\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \hat{M}\begin{pmatrix} u \\ v \\ w \end{pmatrix}, \tag{38}$$

where $\hat{M}$ is the matrix

$$\hat{M} = \begin{pmatrix} 0 & -\Delta & -a_2 \\ \Delta & 0 & a_1 \\ a_2 & -a_1 & 0 \end{pmatrix}. \tag{39}$$

A solution to Eq. (38) is obtained with the aid of the transformation

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \exp(\hat{M}t)\begin{pmatrix} u_0 \\ v_0 \\ w_0 \end{pmatrix}, \tag{40}$$

where u_0, v_0, and w_0 are the initial values of u, v, and w, respectively.

Using the roots λ_i of the characteristic equation $|\lambda\hat{I} - \hat{M}| = 0$, we may represent the operator $\exp(\hat{M}t)$ in the following form [21]:

$$\exp(\hat{M}t) = \sum_j \prod_{i \neq j} \frac{\hat{M} - \lambda_j \hat{I}}{\lambda_i - \lambda_j} \exp(\lambda_i t). \tag{41}$$

Using expressions of the roots of the characteristic equation,

$$\lambda_1 = 0, \quad \lambda_{2,3} = \pm i\sqrt{\Delta^2 + a_1^2 + a_2^2} \equiv \pm i\Omega, \tag{42}$$

we can transform Eq. (41) to

$$\exp(\hat{M}t) = \hat{M}^2 \frac{1}{\Omega^2}(1 - \cos\Omega t) + \hat{M}\frac{1}{\Omega}\sin\Omega t + \hat{I}, \tag{43}$$

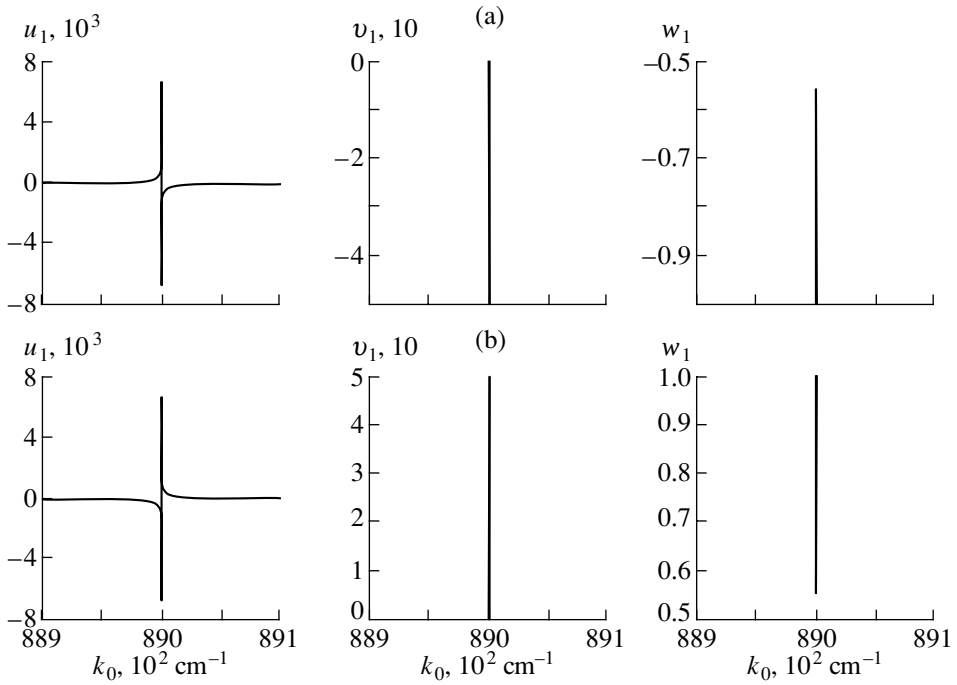


Fig. 1. Nonlinear resonance in an isolated two-level atom (atom 1): (a) frequency dependences of u_1, v_1, and w_1 for the initial inversion $w_{01} = -1$ (atom in the ground state); (b) the same for $w_{01} = 1$ (atom in the excited state). Numerical calculations were performed for $k_{01} = 89000$ cm^{-1}, $d_{01} = 7 \times 10^{-18}$ CGSE units, $T_{11} = T'_{21} = 2.24 \times 10^{-8}$ s, $E_{0I} = 3.32 \times 10^{-3}$ CGSE units, $k_0 = \omega/c$, $f_{01} = 1$. The external wave is linearly polarized, with the polarization vector $\mathbf{e}'_I$ being parallel to the vector $\mathbf{d}_{01} = \mathbf{d}'_{01}$; $d''_{01} = 0$.

where according to (39),

$$\hat{M}^2 = \begin{pmatrix} -\Delta^2 - a_2^2 & a_1 a_2 & -\Delta a_1 \\ a_1 a_2 & -\Delta^2 - a_1^2 & -\Delta a_2 \\ -\Delta a_1 & -\Delta a_2 & -a_1^2 - a_2^2 \end{pmatrix}. \qquad (44)$$

Using Eq. (43), we may study the inversion of the atom-observer. Substituting matrices (39) and (44) into (43), we obtain an expression for the inversion of a two-level atom in the field of a single rectangular pulse:

$$w = \frac{1}{\Omega^2}(1 - \cos\Omega t) \times [-\Delta a_1 u_0 - \Delta a_2 v_0 - (a_1^2 + a_2^2) w_0] + \frac{1}{\Omega}\sin\Omega t (a_2 u_0 - a_1 v_0) + w_0. \qquad (45)$$

For $a_2 = 0$ (i.e., for a real amplitude and polarization of the field acting upon the atom), expression (45) coincides with the formula for inversion following from the Rabi solution [22].

A solution to the modified Bloch optical equations (38) for the time $t > \tau$ (pulsed radiation switched off) can be readily obtained by putting $a_1 = a_2 = 0$ in matrices (39) and (44).

Let us consider another property of solution (40). If the light pulse acting upon the two-level atom is not rectangular, the interval τ can be divided into smaller intervals during which the values E'_0, E''_0 and a_1, a_2 are constant. Then the solution to Eq. (38) can be presented in the following form:

$$\begin{pmatrix} u(t) \\ v(t) \\ w(t) \end{pmatrix} = \exp\left(\int_{-\infty}^{t} \hat{M}(t')dt'\right) \begin{pmatrix} u_0 \\ v_0 \\ w_0 \end{pmatrix}. \qquad (46)$$

Let us also determine the area of the light pulse in the modified Bloch equations using the values a_1 and a_2 characterizing the polarization and amplitude of the radiation pulse. Using the roots (42) with $\Delta = 0$, we may determine the pulse area as

$$\theta(t) = \int_{-\infty}^{t} \sqrt{a_1^2(t') + a_2^2(t')}\,dt'. \qquad (47)$$

The mutual influence of atoms in the nanoobject exposed to intensive pulsed radiation can be determined using solution (40) for variables u_1, v_1 and u_2, v_2 determined at various time instants with the aid of expressions (36).

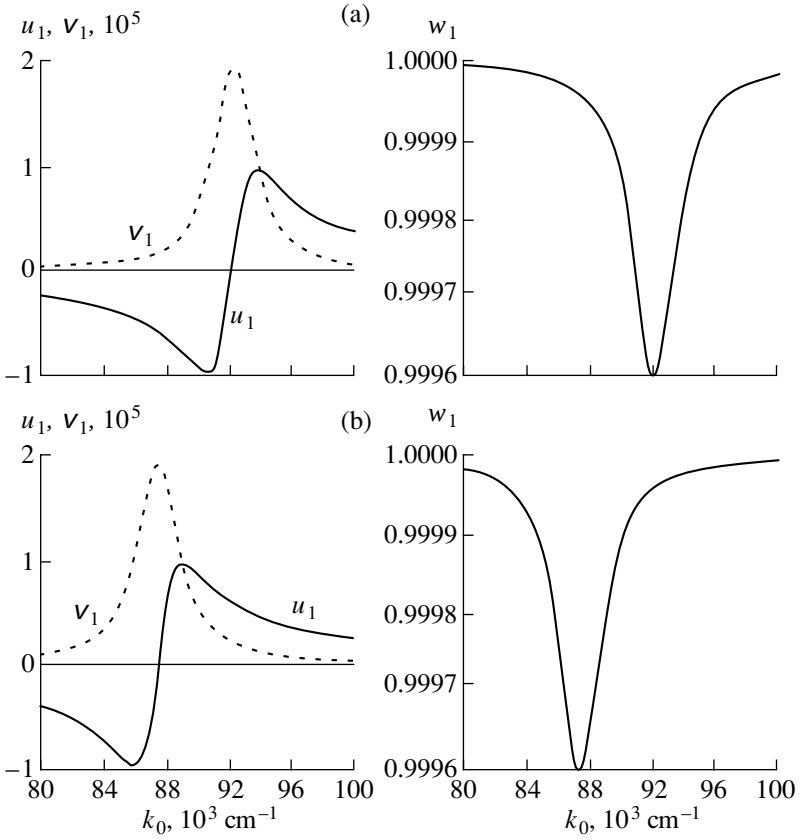


Fig. 2. Nonlinear optical size resonances in a diatomic nanoobject composed of two identical atoms, for various polarizations of the external radiation: (a) frequency dependences of u_1, v_1, and w_1 of atom 1 for the linear polarization with vector $\mathbf{e}'_{0I}$ directed along the object axis; (b) the same for $\mathbf{e}'_{0I} \perp \mathbf{R}$. Numerical calculations were performed for $d_{01} = d_{02} = d_0 = 7 \times 10^{-18}$ CGSE units, $T_{11} = T_{12} = 2.24 \times 10^{-8}$ s, $E_{0I} = 3.4 \times 10^{-3}$ CGSE units, $f_{01} = f_{02} = f_0 = 1$, $T'_{21} = T'_{22} = 2.1 \times 10^{-14}$ s; $C = 0$, $A = 1$, $w_{01} = w_{02} = 1$, $B = d_0G$ (a) and $-(1/2)d_0G$ (b).

4. NONLINEAR OPTICAL SIZE RESONANCES IN A NANOOBJECT COMPOSED OF TWO IDENTICAL ATOMS

Let us numerically investigate the stationary solution (37) for two identical atoms with the intrinsic frequencies $\omega_{01} = \omega_{02}$. For this purpose, we will compare the spectral properties of the system to the analogous properties of isolated atoms in the field of intense radiation.

Figure 1 shows the characteristic plots of the quantities u_1, v_1, and w_1, for example, for atom 1 not interacting with atom 2 and excited only by the field of an external wave. The dipole transition moment of this atom d_{01} is determined using a formula for the oscillator strength f_{01} [23] corresponding to the frequency ω_{01} or the wavenumber $k_{01} = \omega_{01}/c = 89000$ cm^{-1}. The energy relaxation time T_{11} is equal to the lifetime of the excited state of atom 1 and the phase relaxation time is $T'_{21} = T_{11}$. The field amplitude E_{0I} was selected such that the value $(2d_{01}/\hbar)^2 T_{11} T'_{21}$ would be comparable with unity. As the field E_{0I} grows, the maximum u_1, v_1, and w_1 values and the resonance width significantly increase. Depending on the initially selected value of inversion ω_{01}, we may obtain either positive (Fig. 1a) or negative (Fig. 1b) dispersion.

Figure 2 shows the results of another numerical experiment, in which the optical properties of a diatomic nanostructure with the interatomic distance $R = 1$ nm were studied in the field of intense radiation with an allowance for the mutual influence of atoms 1 and 2. The behavior of atoms in this object is described by functions (37) in combination with relationships (36) depending on the dipole orientation in the radiation field. Let us assume that the polarization vectors of the external wave and the internal field inside the nanoobject coincide, being parallel to the induced dipole moments of atoms. In this case, according to formulas (35), we have for the linearly polarized wave $A = 1$, $C = 0$, and $B = d_0G$ (for $\mathbf{e}'_{0I} \parallel \mathbf{R}$) or $B = -d_0G/2$ (for

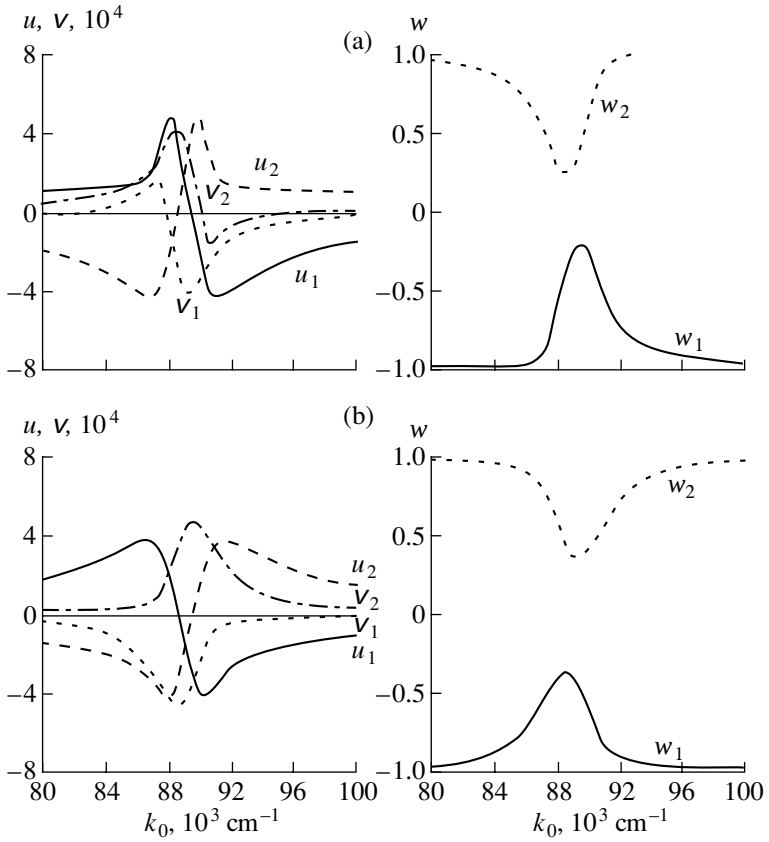


Fig. 3. Nonlinear optical size resonances in a diatomic nanoobject composed of two identical atoms for various polarizations of the external radiation and various initial inversions of atoms 1 and 2: (a) polarization $\mathbf{e}'_{0I} \parallel \mathbf{R}$; (b) polarization $\mathbf{e}'_{0I} \perp \mathbf{R}$ (for a linearly polarized wave with $\mathbf{e}'_{0I} \parallel 0x$). Numerical calculations were performed for $f_0 = 1$, $E_{0I} = 3.4 \times 10^{-3}$ CGSE units, $w_{01} = -1$, $w_{02} = 1$.

$\mathbf{e}_{0I}$ parallel to the x axis). As is seen in Fig. 2, the diatomic nanoobject composed of identical atoms exhibits two nonlinear size resonances at $k_1 = 92160$ cm^{-1} and $k_2 = 87500$ cm^{-1} corresponding to two directions of polarization of the external wave: $\mathbf{e}'_{0I} \parallel \mathbf{R}$ and $\mathbf{e}'_{0I} \perp \mathbf{R}$ (instead of a single resonance k_{01} observed for the isolated atoms).

The properties of nonlinear size resonances differ from those of the linear size resonances corresponding to solution (14). These differences are as follows.

1. The positions of nonlinear size resonances strongly depend on the external field amplitude. As the field amplitude E_{0I} varies from 3.4 to 0.034 CGSE units, the nonlinear resonance at $\mathbf{e}'_{0I} \parallel \mathbf{R}$ shifts from $k_1 = 90590$ cm^{-1} to $k_1 = 92160$ cm^{-1} (Fig. 2a). An analogous situation takes place for another size resonance corresponding to the polarization $\mathbf{e}'_{0I} \perp \mathbf{R}$ (Fig. 2b).

2. The width of the nonlinear size resonances is considerably (10^3 times) greater than the width of the linear size resonances and strongly depends on the field amplitude E_{0I}.

3. The width and position of the nonlinear size resonance strongly depends on the transition oscillator strength. For example, a 10-fold decrease in the oscillator strength f_0 leads to an approximately 10-fold narrowing of the nonlinear resonance and shifts the nonlinear resonance maximum to $k_1 = 89160$ cm^{-1} instead of $k_1 = 90590$ cm^{-1} for $f_0 = 1$ and $E_{0I} = 3.4$ CGSE units.

4. A decrease in the oscillator strength and the external field amplitude leads to a dispersion dependence for u_j and v_j typical of the isolated resonance (at $k_{01} = 89000$ cm^{-1}) of noninteracting atoms.

5. The dispersion relationships of u_1, v_1 and u_2, v_2 for atoms 1 and 2 of the nanoobject differ if their initial inversions are not the same. Figure 3 shows these relationships calculated for particular conditions of the numerical experiment. Here, the nonlinear size resonances occur near the intrinsic resonance (at $k_{01} = 89000$ cm^{-1}).

6. The minima of inversion correspond to the size resonances occurring on the left and on the right of the intrinsic resonance at $k_{01} = 89000$ cm^{-1}, depending on the external wave polarization and amplitude.

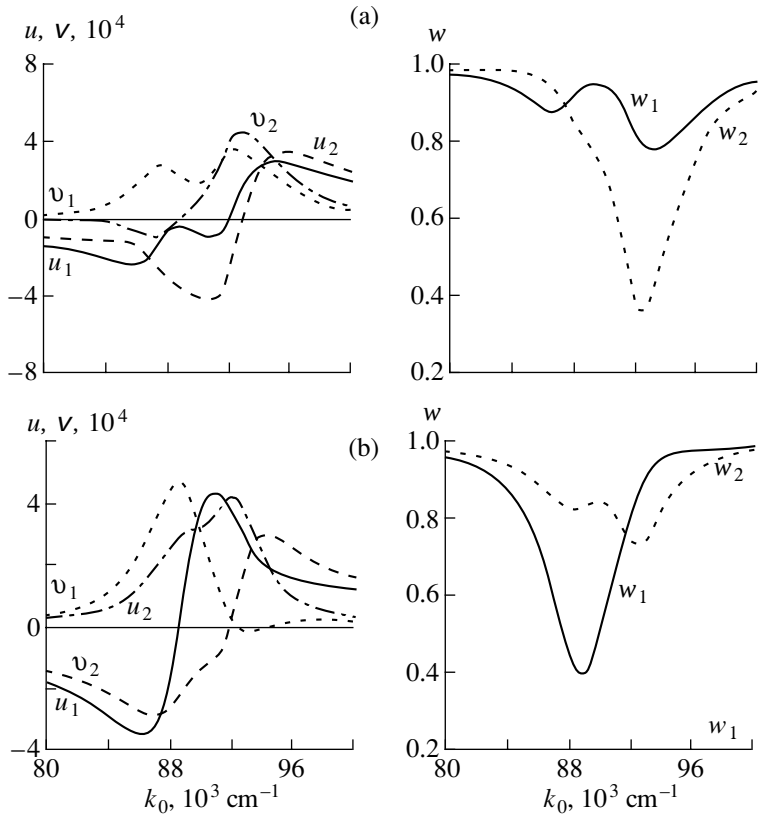


Fig. 4. Nonlinear optical size resonances in a diatomic nanoobject composed of two different atoms for two different polarizations of the external radiation relative to the object axis and a fixed external field amplitude: (a) polarization $\mathbf{e}'_{0I} \parallel \mathbf{R}$; (b) polarization $\mathbf{e}'_{0I} \parallel x$. Numerical calculations were performed for $f_{01} = f_{02} = 1$, $d_{01} = 7.06 \times 10^{-18}$ CGSE units, $d_{02} = 6.95 \times 10^{-18}$ CGSE units, $T_{11} = T_{12} = 2.24 \times 10^{-8}$ s, $E_{0I} = 3.4 \times 10^{-3}$ CGSE units, $T'_{21} = T'_{22} = 2.1 \times 10^{-14}$ s; $C = 0$, $A = 1$, $w_{01} = w_{02} = 1$, $B_1 = 7.07 \times 10^3$, $B_2 = 6.09 \times 10^3$ (a); $B_1 = -3.54 \times 10^3$; $B_2 = -3.47 \times 10^3$ (b).

5. NONLINEAR OPTICAL SIZE RESONANCES IN A NANOOBJECT COMPOSED OF TWO DIFFERENT ATOMS

Consider a nanoobject to be composed of different atoms 1 and 2 possessing isolated resonances with the wavenumbers $k_{01} = 89000$ cm^{-1} and $k_{02} = 92000$ cm^{-1}. In Section 2, we have studied the properties of linear optical resonances in an analogous diatomic object exposed to a low-intensity radiation field. Let us consider properties of the same object in a stationary radiation field, taking into account the inversion of atoms described by solution (37). Figures 4 and 5 present the results of numerical experiments obtained for various polarizations of the external field and various initial inversions.

As is seen from Fig. 4, the case of $\mathbf{e}_{0I} \parallel \mathbf{R}$ (when the linearly polarized external wave has the wavevector $\mathbf{k}_{0I} \perp \mathbf{R}$ and the real vector $\mathbf{e}'_{0I}$ is directed along the axis of the diatomic nanoobject with the size $R = 1$ nm) is characterized by the appearance of two pronounced maxima of v_1 and v_2 for the wavenumbers 87 500 and 93000 cm^{-1}. For the other polarization (Fig. 4, $\mathbf{e}_{0I} \parallel \mathbf{R}$), the maxima of v_1 and v_2 are observed at 88500 and 92500 cm^{-1}.

Thus, exposed to an intense radiation field, a two-level nanoobject composed of two different atoms exhibits (as well as in the linear case) four size resonances. The shapes and positions (on the wavelength scale) of these resonances strongly depend on the external field amplitude. The inversions of the two atoms significantly differ for the same frequencies of the external field. As is seen from Fig. 5, the spectral properties of such a nanoobject significantly depend on the initial inversions of both atoms.

6. CONCLUSION

Thus, we proved the existence of optical size resonances in atomic nanostructures. The size resonance frequencies significantly differ from the intrinsic frequencies of isolated atoms. The proof was presented for diatomic nanoobject representing dimers on a sold surface, with a fixed arrangement relative to the surface

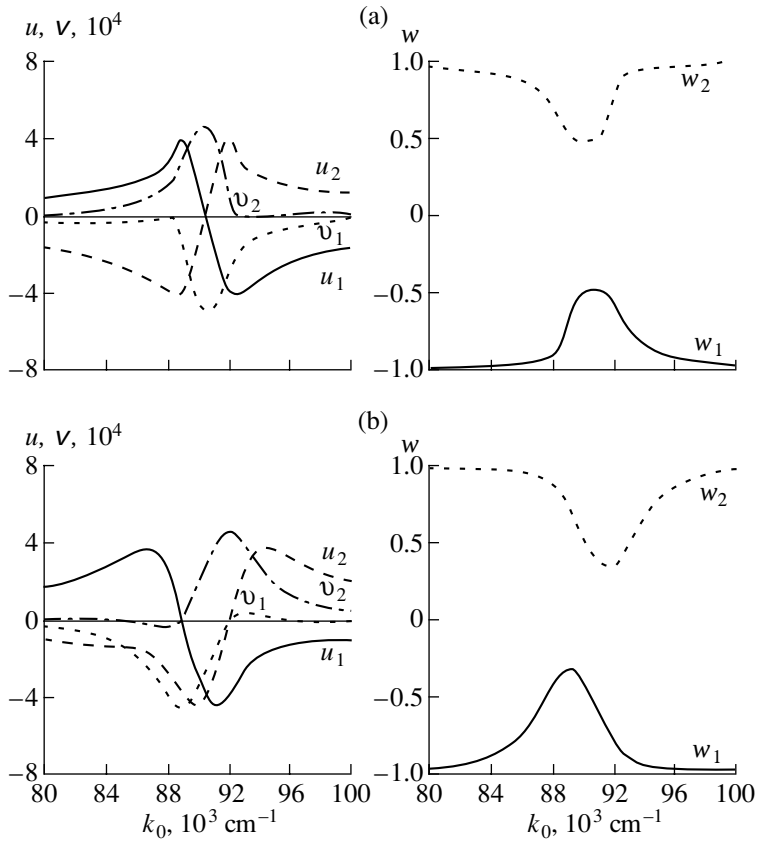


Fig. 5. Nonlinear optical size resonances in a diatomic nanoobject composed of two different atoms for different initial inversions of atoms 1 and 2: (a) polarization $\mathbf{e}'_{0I} \parallel \mathbf{R}$; (b) polarization $\mathbf{e}'_{0I} \parallel x$. Numerical calculations were performed for the same parameters as indicated in caption to Fig. 4, except for $w_{01} = -1$, $w_{02} = 1$.

crystallographic axes and the intrinsic frequencies and polarizabilities of atoms determined by their interaction with the surface. In this system, it is possible to study dimers with the aid of an external optical radiation field polarized in various directions. An example is offered by the experiment [4] with arsenic dimers on a clean GaAs surface. We have already performed numerical calculations showing the presence of optical size resonances in N-atomic nanostructures such as liner atomic chains and fullerenes.

The linear optical size resonances exhibit a high selectivity and the resonance width is determined by the radiation decay time of dipole oscillators. The nonlinear optical size resonances are characterized by a markedly greater width, which is related to a resonance excitation transfer between two closely spaced atoms. The nature of optical size resonances is determined by the electric dipole–dipole interaction of atoms in the radiation field. We separate the static and dynamic interaction components for atoms in a nanostructure. The static component, independent of the radiation field, leads to a change of the intrinsic atomic frequencies and determines the phase relaxation times T'_{21} and T'_{22}. The dynamic part of the interatomic interaction depends on the radiation field and significantly modifies (at small distances) the optical spectrum of the nanostructure, which no longer contains intrinsic frequencies and displays only the size resonances.

The number of resonances in a nanostructure depends on the number of atoms and on whether these atoms are identical or different. A diatomic nanostructure consisting of identical atoms exhibits two size resonances instead of one intrinsic resonance. In the case of different atoms forming a diatomic nanoobject, there appear four size resonances instead of two intrinsic resonances.

The variable field of a nanoobject in the wave zone is strongly dependent on the frequency of the external field. When this frequency coincides with that of one of the size resonances, the field strength in the wave zone increases by approximately 12 orders of magnitude as compared to the value in the same nanostructure observed at the intrinsic resonance frequency.

Various size resonances of a nanoobject can be excited by changing the polarization of the external field. This implies that various optical size resonances can be studied by changing the angle of incidence of an external wave on the nanoobject. The properties of size resonances also strongly depend on the interatomic dis-

tance in the nanostructure, on the electric field strength in the external wave, and on the initial conditions.

We believe that the existence of optical size resonances in nanostructures opens wide possibilities in the study of such objects with aid of optical radiation. This may lead to the development of unique optical methods of investigation (e.g., optical holography of nanostructures) and unique optical devices (e.g., a nanodimensional light polarizer).

ACKNOWLEDGMENTS

This study was supported by the Ministry of Education of the Russian Federation and by the Russian Foundation for Basic Research.

REFERENCES

1. F. C. Spano and J. Knoester, Adv. Magn. Opt. Reson. **18**, 117 (1994).
2. V. Malyshev and P. Moreno, Phys. Rev. A **53**, 416 (1996).
3. *Molecular Interactions*, Ed. by H. Ratajczak and W. Orville-Thomas (Wiley, Chichester, 1981; Mir, Moscow, 1984).
4. V. L. Berkovits and D. Paget, Surf. Sci. **441**, 26 (1999).
5. O. N. Gadomskiĭ and S. V. Sukhov, Kvantovaya Élektron. (Moscow) **25**, 529 (1998).
6. K. V. Krutitsky and S. V. Suhov, J. Phys. B **30**, 5341 (1997).
7. L. T. Canham, Appl. Phys. Lett. **57**, 1046 (1990).
8. A. N. Oraevskiĭ, M. Scully, and V. L. Velichanskiĭ, Kvantovaya Élektron. (Moscow) **25**, 211 (1995).
9. S. K. Sekatskiĭ and V. S. Letokhov, Pis'ma Zh. Éksp. Teor. Fiz. **65**, 441 (1997) [JETP Lett. **65**, 465 (1997)].
10. O. N. Gadomskiĭ, Usp. Fiz. Nauk **170** (11), 1145 (2000).
11. W. A. de Heer, Rev. Mod. Phys. **65**, 612 (1993).
12. M. Brack, Rev. Mod. Phys. **65**, 677 (1993).
13. G. V. Shpatkovskaya, Zh. Éksp. Teor. Fiz. **118**, 87 (2000) [JETP **91**, 76 (2000)].
14. O. N. Gadomskiĭ and K. V. Krutitskiĭ, Zh. Éksp. Teor. Fiz. **106**, 936 (1994) [JETP **79**, 513 (1994)].
15. O. N. Gadomsky and K. V. Krutitsky, J. Opt. Soc. Am. B **13**, 1679 (1996).
16. M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1969; Nauka, Moscow 1973).
17. O. N. Gadomskiĭ and S. V. Sukhov, Opt. Spektrosk. **89**, 287 (2000) [Opt. Spectrosc. **89**, 261 (2000)].
18. O. N. Gadomskiĭ and Yu. Yu. Voronov, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 750 (1999) [JETP Lett. **69**, 804 (1999)].
19. O. N. Gadomskiĭ and Yu. Yu. Voronov, Zh. Prikl. Spektrosk. **66**, 765 (1999).
20. V. M. Fain and Ya. I. Khanin, *Quantum Electronics* (Sov. Radio, Moscow, 1965; MIT Press, Cambridge, 1968).
21. F. R. Gantmacher, *The Theory of Matrices* (Nauka, Moscow, 1967, 3rd ed.; Chelsea, New York, 1959).
22. L. Allen and J. H. Eberly, *Optical Resonance and Two-Level Atoms* (Wiley, New York, 1975; Mir, Moscow, 1978).
23. A. S. Davydov, *Quantum Mechanics* (Fizmatgiz, Moscow, 1963; Pergamon, Oxford, 1976).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 1072–1083.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1235–1249.

SOLIDS
Electronic Properties

Steady-State Phonon-Free Current in Superlattices in a Strong Electric Field Induced by Terahertz Radiation

V. V. Bryksin[a, *] and P. Kleinert[b, **]

[a]*Ioffe Physicotechnical Institute, Russian Academy of Sciences, ul. Politekhnicheskaya 26, St. Petersburg, 194021 Russia*
[b]*Paul-Drude-Institut für Festkörperelektronik, 10117 Berlin, Germany*
**e-mail: val@bryk.ioffe.rssi.ru*
***e-mail: kl@pdi-berlin.de*
Received December 25, 2000

Abstract—The effect of strong terahertz radiation on the current–voltage characteristics in one-dimensional superlattices is studied using the method of two-time nonequilibrium Green's functions. It is shown that the broadening of discrete spectrum lines due to elastic scattering plays a fundamental role in this problem. A new phonon-free type of photon-induced charge transfer in the hopping mode is predicted. Under certain conditions, the constant components of current and electric field may have opposite directions in this regime. The phonon-free current cannot be described in principle in the formalism of the one-time density matrix and reflects the two-time nature of correlation functions. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The nonlinear dynamics of charged particles in semiconducting superlattices (SL) in strong dc and ac electric fields has been studied intensely both theoretically and experimentally [1–23]. Interesting experimental results [1–5] stimulated the application of various theoretical approximations based on the balance equation [6–9] or tunneling [10–12]. Most theoretical publications [13–20] are devoted to the solution of the Boltzmann equation for one-dimensional models in the relaxation-time approximation. Only a few publications [21–23] deal with quantum effects in current transport in SLs in the presence of electromagnetic radiation.

We will consider the current transport along the normal to the SL layers for a low concentration of charge carriers (with Boltzmann's statistics). If the Bloch frequency $\Omega_{dc} = eE_{dc}d/\hbar$ (d is the SL period and E_{dc} is a dc electric field) is higher than the reciprocal relaxation time $1/\tau$, it is convenient to use the Stark ladder representation. The electron transport is carried out through inelastic transitions between the levels of this ladder due to phonon absorption and emission. If an additional ac component of the electric field is also present, an additional channel for current transport due to the absorption and emission of photons is opened. In this case, the photon absorption under certain conditions may be more effective than photon emission; as a result, the constant component of the electric current is directed opposite to the constant component of the field [1].

In the presence of a strong magnetic field directed along the electric field (and across the SL layers), the electron spectrum is discrete in the 3D case also due to the Wannier–Stark and Landau quantization. In this case, the level broadening due to elastic scattering plays a fundamental role. The electron transport cannot be studied using the perturbation theory in elastic scattering. The summation of an infinite set of diagrams required in this case leads to a nonanalytic dependence of the electron state lifetime on the electron–impurity interaction constant. For a continual electron spectrum (e.g., in 3D systems in zero magnetic field) with Stark localization, the perturbation theory in the electron–impurity interaction can be used and the analysis can be confined to the one-time density matrix $f(\mathbf{k}, t)$ ($\mathbf{k}$ is the wave vector) [24]. In the presence of an alternating electric field component with frequency ω_{ac}, the distribution function is periodic $f(\mathbf{k}, t) = f(\mathbf{k}, t + 2\pi/\omega_{ac})$, and the time dependence of current can be analyzed by studying the Fourier components $f_m(\mathbf{k}, \omega_{ac})$ (m is the number of the Fourier component). In all previous publications [21–23], this approach or the Boltzmann equation in the relaxation-time approximation was used. This method effectively leads to the Esaki–Tsu theory [25] taking into account the alternating component of the field. The same result was obtained using two-time Green's functions and the Kadanov–Beym approximation [26–28].

In this paper, we will show that such a semiphenomenological approach is not satisfactory for systems with a discrete spectrum, for which a consistent inclusion of the lifetime of electronic states is of fundamental importance. On the microscopic level, the lifetime is associated with the introduction of an additional time t_- having the meaning of the duration of an act of scattering ($t_- \longrightarrow 0$ in the density matrix approximation). As a result, we obtain the transport equation for the two-time Green's function $f(\mathbf{k}, t, t_-)$, which is the main object of our investigation. It should be noted that the application of the Kadanov–Beym assumption trans-

forms this equation into a kinetic equation of the Boltzmann type for the standard one-time density matrix. The transport equation obtained describes the hopping as well as the band nature of transport in various limiting cases. Besides, it takes into account the effect of an electric field on scattering processes and level broadening due to scattering. It will be proved that the two-time nature of the distribution function leads to a steady current induced by photons. The analysis will be carried out using a 1D model. A more realistic case of the discrete spectrum of 3D systems in a magnetic field is only slightly more complicated technically since it requires the application of the Wigner representation but basically does not differ from the model considered here.

In Section 2, Dyson equations for two-time Green's functions are considered and the symmetry relations for these functions are presented. The expression for the density of states (Green's function G^+) is derived in Section 3. In Section 4, the transport equation for the G^- component is analyzed without using the Kadanov–Beym assumption. The computation of the phonon-free contribution is carried out in Section 5. The results of numerical calculations are given in Section 6.

2. BASIC EQUATIONS

We study the nonlinear transport in a narrow-band SL with an electric field applied only along the z axis, which is directed across the layers. The analysis will be based on the Keldysh technique developed for two-time Green's functions $G^\pm$. The Dyson equations for such Green's functions have the form [28]

$$\begin{aligned}\left[i\hbar\frac{\partial}{\partial t} - \varepsilon(\mathbf{k}) + ie\mathbf{E}(t)\nabla_k\right]G^\pm(\mathbf{k}t|\mathbf{k}'t') \\ = \pm\hbar\int d\mathbf{k}_1\left\{\int_{t'}^{t} dt_1\Sigma^\pm(\mathbf{k}t|\mathbf{k}_1t_1)G^\mp(\mathbf{k}_1t_1|\mathbf{k}'t')\right. \\ + \int_{-\infty}^{t'} dt_1\Sigma^\pm(\mathbf{k}t|\mathbf{k}_1t_1)G^\pm(\mathbf{k}_1t_1|\mathbf{k}'t') \\ \left. - \int_{-\infty}^{t} dt_1\Sigma^\mp(\mathbf{k}t|\mathbf{k}_1t_1)G^\pm(\mathbf{k}_1t_1|\mathbf{k}'t')\right\}.\end{aligned} \tag{1}$$

We will use below the following energy–momentum relation for the SL:

$$\varepsilon(\mathbf{k}) = \frac{\hbar^2k_\perp^2}{2m^*} + \frac{\Delta}{2}[1-\cos(k_zd)], \tag{2}$$

where d is the SL period, m^* is the effective mass, Δ is the width of a miniband, and $\mathbf{k}_\perp$ is the momentum in the xy plane of the layers. Field $\mathbf{E}(t)$ is the sum of a constant and a varying (in time) component which are directed along the z axis. The eigenenergies $\Sigma^\pm$ are considered in the Born approximation and will be given below.

In the presence of an electric field, Green's functions G and energies Σ are nondiagonal in the momentum representation, but obey the translational symmetry relations,

$$G^\pm(\mathbf{k}t|\mathbf{k}'t') = G^\pm(\mathbf{k}|t,t')\delta[\mathbf{k}'-\mathbf{k}-\mathbf{A}(t',t)], \tag{3}$$

with the vector potential $\mathbf{A}(t)$ defined by the relations

$$\frac{d\mathbf{A}(t)}{dt} = \frac{e\mathbf{E}(t)}{\hbar}, \quad \mathbf{A}(t',t) = \mathbf{A}(t') - \mathbf{A}(t).$$

Relation (3) makes it possible to simplify Eq. (1):

$$\begin{aligned}\left[\frac{\partial}{\partial t} + \frac{i}{\hbar}\varepsilon(\mathbf{k}) - \frac{\partial\mathbf{A}(t',t)}{\partial t}\nabla_k\right]G^\pm(\mathbf{k}|t,t') \\ = \mp i\left\{\int_{t'}^{t} dt_1\Sigma^\pm(\mathbf{k}|t,t_1)G^\pm(\mathbf{k}+\mathbf{A}(t_1,t)|t_1,t')\right. \\ + \int_{-\infty}^{t'} dt_1\Sigma^\pm(\mathbf{k}|t,t_1)G^\mp(\mathbf{k}+\mathbf{A}(t_1,t)|t_1,t') \\ \left. - \int_{-\infty}^{t} dt_1\Sigma^\mp(\mathbf{k}|t,t_1)G^\pm(\mathbf{k}+\mathbf{A}(t_1,t)|t_1,t')\right\}.\end{aligned} \tag{4}$$

It is convenient to write the second symmetry relation by introducing a new notation for Green's functions:

$$\tilde{G}(\mathbf{k}|t,t') = G\left(\mathbf{k}-\frac{1}{2}\mathbf{A}(t',t)|t,t'\right). \tag{5}$$

Then condition $G(x,x') = -G^*(x',x)$ for the initial Green's function leads to

$$\tilde{G}^\pm(\mathbf{k}|t,t')^* = -\tilde{G}(\mathbf{k}|t',t), \tag{6}$$

and relations (4) and (5) lead to the following set of Dyson's equations for $\tilde{G}^\pm$:

$$\begin{aligned}\left[\frac{\partial}{\partial t} + \frac{i}{\hbar}\varepsilon\left(\mathbf{k}-\frac{1}{2}\mathbf{A}(t',t)\right) - \frac{1}{2}\frac{\partial\mathbf{A}(t',t)}{\partial t}\nabla_k\right]\tilde{G}^\pm(\mathbf{k}|t,t') \\ = \mp i\left\{\int_{t'}^{t} dt_1\tilde{\Sigma}^\pm\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1,t')|t,t_1\right)\right. \\ \left.\times\tilde{G}^\pm\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1,t)|t_1,t'\right)\right. \\ + \int_{-\infty}^{t'} dt_1\tilde{\Sigma}^\pm\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1,t')|t,t_1\right)\end{aligned} \tag{7}$$

$$\times \tilde{G}^{\mp}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1,t)|t_1,t'\right)$$

$$-\int_{-\infty}^{t} dt_1 \tilde{\Sigma}^{\mp}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1,t')|t,t_1\right)$$

$$\times \tilde{G}^{\pm}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1,t)|t_1,t'\right)\Bigg\}.$$

It is more convenient to calculate $\tilde{G}^{\pm}$ using a slightly different form of Dyson's equations, which can be obtained by summing Eq. (7) with a complex conjugate equation and carrying out the substitution $t \longleftrightarrow t'$ (see Appendix A). For a low electron concentration, we have $\tilde{G}^{-} \ll \tilde{G}^{+}$, where $\tilde{G}^{+}$ describes the density of electron states, while $\tilde{G}^{-}$ plays the role of a two-time density matrix (see below). The roles of $\tilde{G}^{+}$ and $\tilde{G}^{-}$ interchange upon a transition from electrons to holes.

The quantity $\varepsilon(\mathbf{k})$ can be eliminated from the left-hand side of Eq. (7) using the substitution

$$\tilde{G}^{\pm}(\mathbf{k}|t,t') = \mp i g^{\pm}(\mathbf{k}|t,t') \times \exp\left\{-\frac{i}{\hbar}\int_{t'}^{t} d\tau \varepsilon(\mathbf{k}-\mathbf{A}(t,t'|\tau))\right\}, \quad (8)$$

where

$$\mathbf{A}(t,t'|\tau) = \frac{1}{2}[\mathbf{A}(t,\tau)+\mathbf{A}(t',\tau)]. \quad (9)$$

In accordance with Eq. (6), we have the symmetry relation

$$g^{\pm}(\mathbf{k}|t,t') = g^{\pm}(\mathbf{k}|t',t)^{*}. \quad (10)$$

The eigenenergies in the Born approximation can be expressed in terms of Green's functions:

$$\tilde{\Sigma}^{\pm}(\mathbf{k}|t,t') = \sum_{\mathbf{q}} D_{\mathbf{q}}^{\pm}(t'-t)\tilde{G}^{\pm}(\mathbf{k}+\mathbf{q}|t,t'). \quad (11)$$

In the case of elastic scattering at ionized impurities, we can use the simplest model in which function $D_{\mathbf{q}}^{\pm}(t)$ is replaced by a constant scattering parameter U determining the white-noise pair correlator [28]. For scattering at polar optical phonons, we have

$$D_{\mathbf{q}}^{\pm}(t) = \frac{2\pi}{\hbar^2}\frac{|M_{\mathbf{q}}|^2}{\sinh(\hbar\omega_{\mathbf{q}}/2k_BT)} \times \cos\left\{\omega_{\mathbf{q}}\left(t \mp \frac{i\hbar}{2k_BT}\right)\right\}, \quad (12)$$

where $M_{\mathbf{q}}$ is the matrix element of the electron–phonon interaction for phonons with the wave vector $\mathbf{q}$. For the sake of simplicity, we consider the interaction with ordinary bulk phonons disregarding their renormalization due to the SL potential. This is immaterial since we are studying the phonon-free contribution to current under the effect of terahertz radiation.

3. CALCULATION OF DENSITY OF STATES (GREEN'S FUNCTION G^{+})

It was noted above that in the case of a low charge carrier concentration, $G^{+} \gg G^{-}$ since G^{-} is proportional to the electron concentration. In this limit, we can omit the contributions proportional to G^{-} (and Σ^{-}) on the right-hand side of Eq. (7) for G^{+}, which leads to the following closed equation for G^{+}:

$$\left[\frac{\partial}{\partial t}+\frac{i}{\hbar}\varepsilon\left(\mathbf{k}-\frac{1}{2}\mathbf{A}(t',t)\right)-\frac{1}{2}\frac{\partial \mathbf{A}(t',t)}{\partial t}\nabla_k\right]\tilde{G}^{+}(\mathbf{k}|t,t') = -i\int_{t'}^{t} dt_1 \tilde{\Sigma}^{+}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1,t')|t,t_1\right) \times \tilde{G}^{+}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1,t)|t_1,t'\right). \quad (13)$$

Going over to g^{+}, we obtain

$$\left[\frac{\partial}{\partial t}-\frac{1}{2}\frac{\partial \mathbf{A}(t',t)}{\partial t}\nabla_k\right]g^{+}(\mathbf{k}|t,t') = -\sum_{\mathbf{q}}\int_{t'}^{t} dt_1 D_{\mathbf{q}}^{+}(t_1-t)\Phi_{t_1t}(\mathbf{k},\mathbf{q}|t,t') \times g^{+}\left(\mathbf{k}+\mathbf{q}+\frac{1}{2}\mathbf{A}(t_1,t')|t,t_1\right)g^{+}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1,t)|t_1,t'\right). \quad (14)$$

In this equation, the phase factor

$$\Phi_{t_1t}(\mathbf{k},\mathbf{q}|t,t') = \exp\left\{-\frac{i}{\hbar}\int_{t_1}^{t} d\tau[\varepsilon(\mathbf{k}+\mathbf{q}-\mathbf{A}(t,t'|\tau))-\varepsilon(\mathbf{k}-\mathbf{A}(t,t'|\tau))]\right\} \quad (15)$$

has been introduced to describe the effect of the electric field on scattering.

In the range of quantizing electric fields in which we are interested (when $\Omega_{dc}\tau \gg 1$), it is convenient to go over to the Fourier representation in k_z, which corresponds to a transition from the momentum representation to the Stark ladder:

$$g^{\pm}(\mathbf{k}|t,t') = \sum_{l=-\infty}^{\infty} g_l^{\pm}(\mathbf{k}_{\perp}|t,t')\exp(ilk_zd). \quad (16)$$

The convenience of the Stark ladder representation for $\Omega_{dc}\tau \gg 1$ is due to the fact that $g_0^{\pm} \gg g_l^{\pm}$ $(l \neq 0)$ under such conditions. The latter inequality indicates a transition to the hopping transport region over the levels of the Stark ladder; the current in this case is proportional to the probability of transitions between these levels.

In the subsequent analysis, it is convenient to use the time variables

$$T = \frac{t+t'}{2}, \quad t_- = t' - t, \tag{17}$$

where time T describes variations of ω_{ac}^{-1} on a macroscopic scale and t_-, on a microscopic scale (of the order of the interaction time).

Let us now consider the 1D model in which the dependence on $\mathbf{k}_\perp$ is absent. This model was used in [13–20] for studying the transport in SLs. In the following analysis, we assume that the lifetime of electronic states is associated only with elastic scattering. The field dependence of the lifetime can be analyzed using Eq. (14) both in the hopping and in the band modes. For $\Omega_{dc}\tau \ll 1$, it is convenient to use the momentum representation, while for $\Omega_{dc}\tau \gg 1$, we write Eq. (14) in the Stark representation (16). For $g_0 \gg g_l$, we have

$$\left\{\left\{\frac{1}{2}\frac{\partial}{\partial T} - \frac{\partial}{\partial t_-}\right\} + \frac{il}{2}\left[\Omega_{dc} + \Omega_{ac}\cos\left(\omega_{ac}\left(T - \frac{t_-}{2}\right)\right)\right]\right\} g_l^+(T, t_1)$$
$$= U\int_0^{t_-} dt_1 I_l(T, t_-, t_1) g_0^+\left(T + \frac{t_1 - t_-}{2}, t_-\right) \times g_0^+\left(T + \frac{t_1}{2}, t_- - t_1\right), \tag{18}$$

where U is the coupling constant for elastic scattering, $\Omega_{ac} = eE_{ac}d/\hbar$ for $E_{ac}(t) = E_{ac}\cos(\omega_{ac}t)$. The kernel of the integral equation has the form

$$I_l(T, t_-, t_1) = \exp\left[-i\frac{eld}{2\hbar}\int_0^{t_-} d\tau E\left(\tau + T - \frac{t_-}{2}\right)\right] \times \sum_{k_z q_z} \exp(-ik_z d) \times \exp\left\{\frac{i}{\hbar}\int_0^{t_1} d\tau\left[\varepsilon\left(k_z + q_z + \frac{e}{\hbar}\int_0^{\tau} d\tau' E\left(\tau' + T - \frac{t_-}{2}\right)\right) - \varepsilon\left(k_z + \frac{e}{\hbar}\int_0^{\tau} d\tau' E\left(\tau' + T - \frac{t_-}{2}\right)\right)\right]\right\}. \tag{19}$$

This function is periodic in T; i.e.,

$$g_l^{\pm}(T + 2\pi/\omega_{ac}, t_-) = g_l^{\pm}(T, t_-), \tag{20}$$

which allows us to go over to the Fourier representation

$$g_l^{\pm}(T, t_-) = \sum_{m=-\infty}^{\infty} g_{lm}^{\pm}(t_-)\exp(im\omega_{ac}T). \tag{21}$$

The Fourier components satisfy the equation

$$\left[\frac{m}{2}\omega_{ac} + \frac{l}{2}\Omega_{dc} + i\frac{\partial}{\partial t_-}\right] g_{lm}^+(t_-) + \frac{l}{4}\Omega_{ac}\left[g_{lm-1}^+(t_-)\exp\left(-i\frac{\omega_{ac}t_-}{2}\right) + g_{lm+1}^+(t_-)\exp\left(i\frac{\omega_{ac}t_-}{2}\right)\right] = -iU\sum_{m_1 m_2}\int_0^{t_-} dt_1 I_{m_1 m_2}^{lm}(t_-, t_1) g_{0m_1}^+(t_1) g_{0m_2}^+(t_- - t_1) \tag{22}$$

with the kernel

$$I_{m_1 m_2}^{lm}(t_-, t_1) = \frac{\omega_{ac}}{2\pi}\int_0^{2\pi/\omega_{ac}} dT I_l(T, t_-, t_1) \times \exp\left\{i\omega_{ac}\left[m_1\left(T + \frac{t_1 - t_-}{2}\right) + m_2\left(T + \frac{t_1}{2}\right) - mT\right]\right\}. \tag{23}$$

Finally, introducing the quantity

$$h_{0m}(t_-) = g_{0m}^+(t_-)\exp\left(-\frac{i}{2}m\omega_{ac}t_-\right), \tag{24}$$

we obtain the following equation for this quantity:

$$\frac{\partial h_{0m}(t_-)}{\partial t_-} = -U\sum_{m_1 m_2}\int_0^{t_-} dt_1 I_{m_1+m_2-m}(t_-, t_1) \times \exp\{i\omega_{ac}[m_1 t_1 + (m_2 - m)t_-]\} h_{0m_1}(t_1) h_{0m_2}(t_- - t_1) \tag{25}$$

with

$$I_m(t_-, t_1) = \frac{\omega_{ac}}{2\pi}\int_0^{2\pi/\omega_{ac}} dT \exp(im\omega_{ac}T) I_{l=0}(T, t_-, t_1). \tag{26}$$

For narrow minibands ($\Delta \longrightarrow 0$) in the Laplace representation in t_-, we can write Eq. (25) in the form

$$sh_{0m}(s) + U\sum_{m_1} h_{0m_1}(s) h_{0m-m_1}(s + im_1\omega_{ac}) = \delta_{m,0}. \tag{27}$$

In the transition to the Laplace representation, we confined our analysis to the region $t_- > 0$. The solution can be continued to the region $t_- < 0$ using relation (10). It can be seen from Eq. (27) that for $m \neq 0$, all $h_{0m} = 0$. If, however, $m = 0$, Eq. (25) can be presented in compact form for an arbitrary width of a miniband also:

$$\left[s + U\int_0^\infty dt e^{-st} I_0(t) g_{00}^+(t)\right] g_{00}^+(s) = 1 \tag{28}$$

with a kernel

$$I_0(t) = \frac{\omega_{ac}}{2\pi}\int_0^{2\pi/\omega_{ac}} dT J_0^2 \times\left(\frac{\Delta}{\hbar}\left|\sum_{k=-\infty}^{\infty} J_k\left(\frac{\Omega_{ac}}{\omega_{ac}}\right)\frac{\sin(\Omega_{dc}+k\omega_{ac})t/2}{\Omega_{dc}+k\omega_{ac}}\exp(ik\omega_{ac}T)\right|\right), \tag{29}$$

where $J_k(x)$ are Bessel's functions.

Using the integral equation (28), we can find the dependence of the density of states on the frequency ω_{ac} of the ac component of the field and the Bloch frequencies Ω_{dc} and Ω_{ac}. The field dependence of the density of states calculated numerically using Eq. (28) has the form of a main peak and a number of auxiliary peaks displaced relative to it by integral values of Ω_{dc} and ω_{ac}. It is most significant that the density of states has no "tails" in the intervals between the peaks. This can be demonstrated analytically using Eq. (27) in the narrow-band limit. In this case, we obtain the well-known result independent of the field,

$$g_{00}^+(s) = \frac{\sqrt{s^2+4U}-s}{2U}, \tag{30}$$

with a typical nonanalytic dependence on the coupling constant U. A similar result was obtained by us earlier for zero alternating field [29]. After the inverse transition to the time representation,

$$g_{00}^+(t) = \frac{1}{\sqrt{U}t} J_1(2\sqrt{U}t), \tag{31}$$

we obtain an oscillating time dependence decreasing with time according to a power law. Such a slow decrease in time leads to sharp peaks in the energy space of the density of states. Analytic continuation $s \longrightarrow i\omega$ in (30) leads to the following expression for the density of states:

$$g_{00}^+(\omega) = \mathrm{Re}\frac{\sqrt{4U-\omega^2}}{2U}. \tag{32}$$

Such a nonanalytic result cannot be reproduced in perturbation theory. In [29], we compared it with the result of the pole approximation leading to an exponential time dependence $g_{00}^+(t)$ and, hence, to Lorentzian tails for the density of states.

4. CALCULATION OF GREEN'S FUNCTION $\tilde{G}^-$

We will calculate the Green's function $\tilde{G}^-$ using Eq. (A.1) from Appendix A. For a nondegenerate electron gas, we can omit the contribution $\Sigma^- G^-$ on the right-hand side of Eq. (A.1), which leads to the linear equation (A.2) for G^-. The method for solving this equation is similar to that used in Section 3 and will not be repeated here. In contrast to the calculation of the density of states, we must take into account the inelastic scattering responsible for energy dissipation. Below, we will consider the interaction with polar optical phonons without taking into account their dispersion and the coupling constant independent of $\mathbf{q}$. From Eq. (A.3) in Appendix A, we obtain

$$\begin{aligned}\left[\frac{\partial}{\partial T} + il\Omega_{dc} + il\Omega_{ac}\cos(\omega_{ac}T)\cos\left(\frac{1}{2}\omega_{ac}t_-\right)\right] g_l^-(T, t_-) \\ = \int_0^\infty dt_1 \Bigg\{ D^-(t_- - t_1) A_l(t, t'|t'-t_1) g_0^-\left(T - \frac{t_1}{2}, t_- - t_1\right) \\ \times g_0^+\left(T + \frac{t_- - t_1}{2}, t_1\right) - D^+(t_1) B_l(t, t'|t'-t_1) \\ \times g_0^-\left(T - \frac{t_1}{2}, t_- - t_1\right) g_0^+\left(T + \frac{t_- - t_1}{2}, t_1\right) \\ + D^-(t_- + t_1) B_l(t, t'|t-t_1) g_0^-\left(T - \frac{t_1}{2}, t_- + t_1\right) \\ \times g_0^+\left(T - \frac{t_- + t_1}{2}, -t_1\right) - D^+(-t_1) A_l(t, t'|t-t_1) \\ \times g_0^-\left(T - \frac{t_1}{2}, t_- + t_1\right) g_0^+\left(T - \frac{t_- + t_1}{2}, -t_1\right)\Bigg\},\end{aligned} \tag{33}$$

which corresponds to Eq. (18) for g_l^+, where

$$A_l(t, t'|t_1) = \sum_{k_z q_z} \Phi_{t_1 t'}(k_z, q_z|t, t')\exp(-ilk_z d), \tag{34}$$

$$B_l(t, t'|t_1) = \sum_{k_z q_z} \Phi^*_{t_1 t'}(k_z, q_z|t, t')\exp(-ilk_z d). \tag{35}$$

Equation (33) may be used for determining both $g_{l=0}^-$ and the current density depending on $g_{l\neq 0}^-$. Using the

Fourier representation (21), we obtain the following equation for $g^-_{00}(t)$ from Eq. (33):

$$\int_{-\infty}^{\infty} dt_1 [F^-(t_1) g^-_{00}(t_1) g^+_{00}(t-t_1) - F^+(t_1) g^+_{00}(t_1) g^-_{00}(t-t_1)] = 0, \quad (36)$$

where

$$F^{\pm}(t) = D^{\pm}(t) I_0(t) \quad (37)$$

with the field-dependent function $I_0(t)$ (29). The linear integral equation (36) should be solved taking into account the additional condition $g^-_{00}(t = 0) = 1$.

It is now expedient to go over to the Fourier representation in time and to introduce the frequency distribution $f_{00}(\omega)$:

$$g^-_{00}(\omega) = g^+_{00}(\omega) f_{00}(\omega). \quad (38)$$

Then we obtain the following equation for $f_{00}(\omega)$ from Eq. (36):

$$\int_{-\infty}^{\infty} d\omega' g^+_{00}(\omega - \omega') \times [F^-(\omega') f_{00}(\omega - \omega') - F^+(\omega') f_{00}(\omega)] = 0 \quad (39)$$

with the additional constraint

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} g^+_{00}(\omega) f_{00}(\omega) = 1. \quad (40)$$

Equation (39) is the main result of this research. Together with Eq. (40), it determines the frequency (ω) or time (t_-) dependence of the distribution function, which reflects the two-time nature of correlator $g^-_l(T, t_-)$. In the case of elastic scattering, when $F^+(\omega) = F^-(\omega)$, we have a simple solution ($f_{00}(\omega) = 1$); i.e., the two-time nature of the distribution function is immaterial. The Kadanov–Beym approximation corresponds to the case when the trivial solution $f_{00}(\omega) = 1$ is valid for inelastic scattering also (see [26–28]). On the contrary, Eq. (39) has no trivial solution for inelastic scattering (i.e., in the presence of energy dissipation). However, such an approximation can be used if we disregard the dependence of $g^-_l(T, t_-)$ on t_-. In this case, Eq. (33) leads to a standard kinetic equation for the one-time density matrix [22]. If this approximation is used, the current can be determined from the density matrix components $g^-_{l\neq 0}(T, t_- = 0)$. Such an approximation was proposed long ago by Polyanovskii [21] and was used by us [22, 23] for analyzing the cyclotron–Stark–phonon–photon resonance in SLs. However, the two-time approximation proposed here makes it possible to consider completely new physical properties associated with the dependence of the distribution function on t_-. These properties have no analogue when the standard density matrix [22] or Boltzmann equation is used [17, 18, 20]. It will be shown in the next section that the inclusion of the two-time nature of Green's functions is not just of academic interest, but leads to a basically new contribution to current which does not appear when the one-time approximation is used.

5. APPLICATION: PHONON-FREE CURRENT

Here, we will apply the formalism developed above for describing the photon-induced contribution to current. For this purpose, we will use the relation between the steady-state current and the Green's function g^-:

$$j_z = \frac{en}{\hbar} \sum_{k_z} \frac{\partial \varepsilon(k_z)}{\partial k_z} \frac{\omega_{ac}}{2\pi} \int_0^{2\pi/\omega_{ac}} dT g^-(k_z | T, t_- = 0), \quad (41)$$

where n is the electron concentration or, in the Fourier representation (16) and (21),

$$j_z = \frac{en_s\Delta}{2\hbar} \times \frac{1}{2i} [g^-_{l=-1, m=0}(t_- = 0) - g^-_{l=1, m=0}(t_- = 0)], \quad (42)$$

where n_s is the 2D electron density. The quantity $g^-_{l,0}(t_- = 0)$ was calculated in Appendix B. From Eq. (B.8), we obtain (see also [23])

$$g^-_{l,0}(t_- = 0) = \sum_{m=-\infty}^{\infty} \frac{P_{lm} S_{lm}}{il\Omega_{dc}}. \quad (43)$$

The matrix elements

$$S_{lm} = \sum_{k=-\infty}^{\infty} J_{k-m}\left(l\frac{\Omega_{ac}}{\omega_{ac}}\right) J_k\left(l\frac{\Omega_{ac}}{\omega_{ac}}\right) \frac{i\Omega_{dc}}{l\Omega_{dc} + k\omega_{ac}} \quad (44)$$

are equal to $\delta_{m,0}$ in the limit of zero alternating field component ($\Omega_{ac} \longrightarrow 0$). Components $P_{l=\pm 1, m}$ were obtained by applying the Fourier transformation to the right-hand side of Eq. (B.1) and taking into account the two-time nature of the correlation functions. Using the symmetry of matrix elements $P_{-1,-m} = P^*_{1,m}$, $S_{-1,-m} = S^*_{1,m}$, we can reduce the expression for current to the form

$$j_z = \frac{en_s\Delta}{2\hbar\Omega_{dc}} \mathrm{Re} \sum_{m=-\infty}^{\infty} P_{1m} S_{1m}. \quad (45)$$

In [22, 23], we studied the transport in SLs in the interaction with polar optical phonons. These calculations

can be repeated using the new approximation. However, it is much more interesting to consider the phonon-free contribution which emerges only when the two-time approach is used. In the case of elastic scattering from impurities, we obtain

$$P_{l=\pm1,m} = 2U\frac{\omega_{ac}}{2\pi}\int_0^{2\pi/\omega_{ac}} dT\exp(-im\omega_{ac}T) \times\int_0^\infty dt_1 A_{l=\pm1}(t, t|t-t_1) \times\left[g_0^-\left(T-\frac{t_1}{2},-t_1\right)g_0^+\left(T-\frac{t_1}{2},t_1\right) - g_0^+\left(T-\frac{t_1}{2},-t_1\right)g_0^-\left(T-\frac{t_1}{2},t_1\right)\right]. \quad (46)$$

Equations (42)–(46) describe the transport to a SL with an arbitrary width of minibands in a quantizing electric field, when $\Omega_{dc}\tau > 1$. It can be seen from Eq. (46) that the matrix elements $P_{l=\pm1,m} = 0$ when the Kadanov–Beym approximation is used and $g_{00}^+(\omega) = g_{00}^-(\omega)$. This means that the phonon-free contribution to current is equal to zero in the one-time density matrix approximation.

Let us now carry out the Fourier transformation of the right-hand side of Eq. (B.1) in T and retain only the components $g_{00}^{\pm}(\omega)$ with $m = 0$ for a strong terahertz field:

$$P_{l=\pm1,m} = 4iU\int_0^\infty dt_1 A_{l=\pm1,m}(t_1)\mathrm{Im}[g_{00}^-(-t_1)g_{00}^+(t_1)] \quad (47)$$

with the matrix elements

$$A_{l,m}(t_1) = \frac{\omega_{ac}}{2\pi}\int_0^{2\pi/\omega_{ac}} dT\exp(-im\omega_{ac}T) \times J_0\left(\frac{\Delta}{\hbar}\left|\sum_{k=-\infty}^{\infty} J_k\left(\frac{\Omega_{ac}}{\omega_{ac}}\right)\frac{\sin[(\Omega_{dc}+k\omega_{ac})t_1/2]}{\Omega_{dc}+k\omega_{ac}} \times\exp\left[ik\omega_{ac}\left(T-\frac{t_1}{2}\right)\right]\right|\right) \times\int_0^{2\pi}\frac{dk}{2\pi}\exp\left[-ilk+\frac{i\Delta}{2\hbar}\int_0^{-t_1} d\tau\cos[k+A(T,\tau)]\right]. \quad (48)$$

In this equation, the following notation should be used:

$$A(T,\tau) = \Omega_{dc}\tau + \frac{\Omega_{ac}}{\omega_{ac}}\{\sin[\omega_{ac}(T+\tau)]-\sin(\omega_{ac}T)\}. \quad (49)$$

We will confine our subsequent analysis to the limit of narrow minibands. Then the first nonvanishing contribution in Δ from (B.6) has the form

$$\mathrm{Re}P_{1m} = \frac{U\Delta}{2\pi\hbar} \times\sum_{k=-\infty}^{\infty}\int_0^{2\pi} dT\cos(mT)J_k\left(2\frac{\Omega_{ac}}{\omega_{ac}}\cos T\right) \times\int_0^\infty dt_1\frac{\sin[(\Omega_{dc}+(m+k)\omega_{ac}/2)t_1]}{\Omega_{dc}+(m+k)\omega_{ac}/2}\mathrm{Im}g_{00}^-(-t_1)g_{00}^+(t_1). \quad (50)$$

This relation can be simplified using the equality

$$\int_0^{2\pi}\frac{dT}{2\pi}J_{2k}\left(2\frac{\Omega_{ac}}{\omega_{ac}}\cos T\right) = J_k^2\left(\frac{\Omega_{ac}}{\omega_{ac}}\right). \quad (51)$$

In this case, we have

$$\mathrm{Re}P_{10} = -\frac{U\Delta}{8\pi\hbar}\sum_{k=-\infty}^{\infty}\frac{J_k^2(\Omega_{ac}/\omega_{ac})}{\Omega_{dc}+k\omega_{ac}}\int_{-\infty}^{\infty} d\omega g_{00}^-(\omega) \times[g_{00}^+(\omega+\Omega_{dc}+k\omega_{ac})-g_{00}^+(\omega-\Omega_{dc}-k\omega_{ac})]. \quad (52)$$

We can now obtain from (45) the expression for current in the form

$$j_z = \frac{en_sU\Delta^2}{8\hbar^2\Omega_{dc}} \times\left[\mathrm{Re}\sum_{k'=-\infty}^{\infty}J_{k'}^2\left(\frac{\Omega_{ac}}{\omega_{ac}}\right)\frac{i\Omega_{dc}+1/\tau_{ac}}{i(\Omega_{dc}+k'\omega_{ac})+1/\tau_{ac}}\right] \times\sum_{k=-\infty}^{\infty}\frac{J_k^2(\Omega_{ac}/\omega_{ac})}{\Omega_{dc}+k\omega_{ac}}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}g_{00}^+(\omega)g_{00}^+(\omega+\Omega_{dc}+k\omega_{ac}) \times[f_{00}(\omega+\Omega_{dc}+k\omega_{ac})-f_{00}(\omega)]. \quad (53)$$

Here, the relaxation time τ_{ac} removing the divergence at resonance points $\Omega_{dc} = k\omega_{ac}$ has been introduced phenomenologically in the same way as in [23]. This attenuation has a completely different source than the lifetime which can be expressed in terms of g_{00}^+ (see also [22], where τ_{ac} is described on the level of microscopic theory).

In the limit of narrow minibands ($\Delta \longrightarrow 0$), the kinetic equation (39) has an exact analytic solution:

$$f_{00}(\omega) = \frac{2\hbar\sqrt{U}/k_BT}{I_1(2\hbar\sqrt{U}/k_BT)}\exp\left(\frac{\hbar\omega}{k_BT}\right). \quad (54)$$

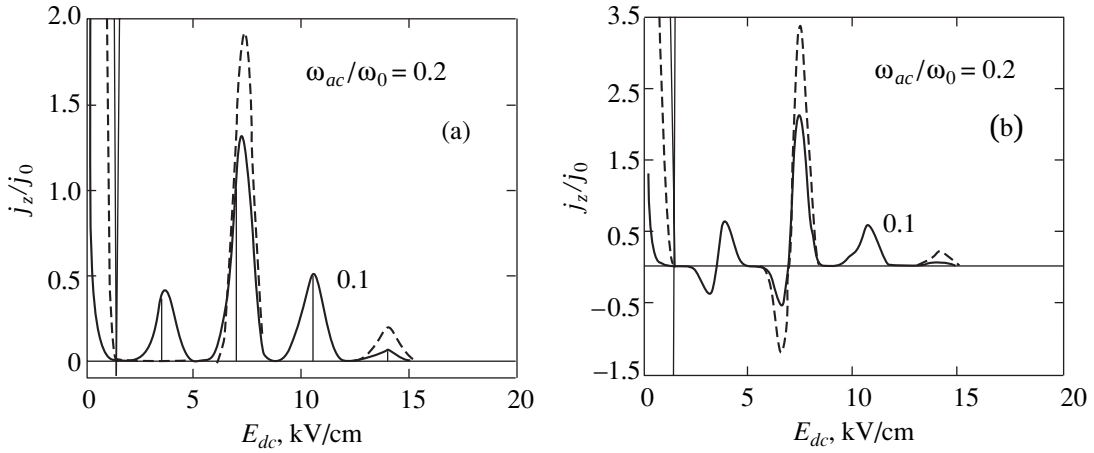


Fig. 1. Dimensionless current density j_z/j_0 as a function of the dc electric field component for E_{ac} = 10 kV/cm, $\sqrt{U}/\omega_0 = 0.01$, T = 4 K and τ_{ac} = 0.5 (a) and 1 ps (b). The optical phonon energy used as the energy scale is $\hbar\omega_0$ = 35 meV. The solid and dashed curves are obtained from Eq. (53) for ω_{ac}/ω_0 = 0.1 and 0.2, respectively. The thin vertical line on the left-hand side of the figure marks the boundary $E_{dc} = 4\hbar\sqrt{U}/ed$ below which our approximation is inapplicable. The thin solid line shows the current for $\Omega_{ac} = 0$. This current vanishes in fields $E_{dc} > 4\hbar\sqrt{U}/ed$. The positions of resonances in Fig. 1a are also marked by vertical lines.

Such a thermodynamically equilibrium distribution function is realized if the charge carrier thermalization time at a level of the Stark ladder is shorter than the time of hopping between the levels (the latter time determines the current in the system). If this condition is violated, the transport becomes of a non-Markovian type and the distribution function should be determined numerically from (39) and (40). It should be noted that inelastic scattering is controlled by the thermalization of carriers at a level of the Stark ladder, occurring in the intervals between the jumps.

The phonon-free current vanishes if we switch off the alternating field ($\Omega_{ac} \longrightarrow 0$) since $g^+_{00}(\omega)g^+_{00}(\omega + \Omega_{dc}) = 0$ for $2\Omega_{dc} > \sqrt{U}$ (see (32)). This means that in a strong electric field, when the transport has the form of jumps over the levels of the Stark ladder, elastic scattering does not lead to the emergence of current (i.e., no electron delocalization takes place). In the previous studies of transport in SLs in the presence of electromagnetic radiation [1–4, 10, 12, 17, 18, 20, 30–32], the following formula derived in [33, 34] was used:

$$j_z(\Omega_{dc}, \Omega_{ac}) = \sum_{k=-\infty}^{\infty} J_k^2\left(\frac{\Omega_{ac}}{\omega_{ac}}\right) j_z^{(dc)}(\Omega_{dc} + k\omega_{ac}). \quad (55)$$

This formula gives $j_z(\Omega_{dc}, \Omega_{ac} = 0) \equiv j_z^{(dc)}(\Omega_{dc}) \neq 0$ which contradicts the result obtained above using two-time Green's functions. It is this fundamental difference from the available results that indicates that the contribution to current obtained above and referred to as phonon-free is new in principle.

The expression (53) for current may be interpreted in the framework of hopping transport between the levels of the Stark ladder with photon emission and absorption. The current–voltage characteristic has the form of a set of sharp peaks having a width of the order of $2\sqrt{U}$ (on the Ω_{dc} scale) and separated from one another by ω_{ac}. The presence of an additional phonon contribution to current leads to the emergence of a background filling the deep valleys between the peaks.

6. NUMERICAL RESULTS AND DISCUSSION

We also numerically calculated the photon-induced current from (53). We confined ourselves to the case when the thermalization at a level of the Stark ladder is quite rapid and the equilibrium distribution (54) can be applied.

Figure 1a shows the current–voltage characteristic (53) for $\omega_{ac}/\omega_0 = 0.1$. We use the $\hbar\omega_0$ energy scale (ω_0 is the optical phonon frequency), although phonons do not participate explicitly in the charge transport, and introduce the constant $j_0 = en_s\Delta^2/16\hbar^2\omega_0$ having the dimensions of current density. It was noted above that the phonon-free current is absent for $\Omega_{ac} \longrightarrow 0$ as well as in a strong field $E_{dc} > 4\hbar\sqrt{U}/ed$. In the region of weak fields $E_{dc} < 4\hbar\sqrt{U}/ed$, the transport is of the band type and not of the hopping type, and the results obtained above are not valid under these conditions. In Fig. 1a, the region of such weak fields is separated by a vertical line. Photon-induced current peaks are centered at points $\Omega_{dc} = k\omega_{ac}$ and marked by vertical segments. The peaks are separated by regions of zero current, in which the densities of the initial and final states do not overlap during hop-

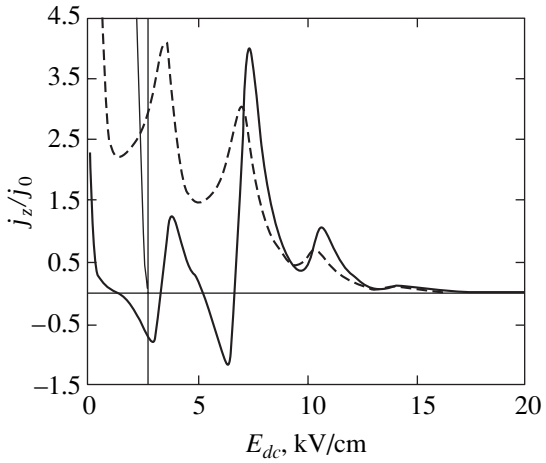


Fig. 2. Dimensionless current density j_z/j_0 as a function of electric field for E_{ac} = 10 kV/cm, $\sqrt{U}/\omega_0 = 0.02$, T = 4 K, and $\omega_{ac}/\omega_0 = 0.1$. The solid and dashed curves were plotted using Eq. (53) for τ_{ac} = 1 ps and 0, respectively.

ping. It should be noted that photon-induced jumps are also realized in the case of localization in disordered systems and can be described using the balance equation [35]. The current peak width depends on the lifetime of electrons on the levels of the Stark ladder.

For $\tau_{ac} \longrightarrow \infty$, expression (53) for current is simplified (see Appendix C):

$$j_z = \frac{\pi e n_s U \Delta^2}{8\hbar^2 \omega_{ac}^2} \frac{J_\nu(\nu') J_{-\nu}(\nu')}{\sin \pi\nu} F(\nu), \qquad (56)$$

where $\nu = \Omega_{dc}/\omega_{ac}$, $\nu' = \Omega_{ac}/\omega_{ac}$, and

$$F(\nu) = \sum_{k=-\infty}^{\infty} \frac{\sinh[\hbar\omega_{ac}(\nu - k)/2k_BT]}{\pi(\nu - k)}$$

$$\times \int_{-\infty}^{\infty} d\omega g_{00}^{+}\left(\omega - \frac{\omega_{ac}}{2}(\nu - k)\right) g_{00}^{+}\left(\omega + \frac{\omega_{ac}}{2}(\nu - k)\right) f_{00}(\omega).$$

Naturally, this expression diverges when ν is equal to an integer. This is demonstrated in Fig. 1b, where τ_{ac} is doubled as compared to Fig. 1a. In this case, regions are formed in which the alternating field induces a direct current flowing against the dc component of the field. Such a negative current is induced as a result of photon absorption and the motion of carriers up the Stark ladder. This clearly distinguishes the phonon-free current from the phonon contribution for which the current is always positive since the processes of phonon emission always have a higher probability as compared to the absorption processes, and an electron on the average moves down the ladder.

The effect of the level width of the Stark ladder on the current–voltage characteristic is illustrated in Fig. 2. The current peaks are well resolved when the condition $8\sqrt{U}/\omega_{ac} < 1$ is satisfied. The solid curve corresponds to the same parameters as in Fig. 1b, but the value of $\sqrt{U}$ is doubled. The level width is close to 2.8 meV. In this case, there are no zero-current regions. In the limit of not very small values of parameter $8\sqrt{U}/\omega_{ac}$, the results in general do not differ considerably from those obtained by the method of one-time density matrix [22, 23]. The dashed curve in Fig. 2 gives the phonon-free contribution to current in the case of a large attenuation ($\tau_{ac} = 0$). In this limit, the current–voltage characteristic for the phonon-free contribution is similar to the electron–phonon resonance [36].

7. CONCLUSION

In this work, the nonlinear transport under the effect of dc and ac electric fields in systems with a discrete electron spectrum is studied for one-dimensional SLs. In such systems, the level broadening of the density of states due to elastic scattering (e.g., from ionized impurities) plays a fundamental role. The Keldysh method is used to derive the system of Dyson equations for the density of states and the two-time electron distribution function. These equations are analyzed in the limit of strong field ($\Omega_{dc}\tau > 1$), when the Wannier–Stark quantization is significant. It is shown that the density of states has the form of a set of narrow and sharp peaks. An analytic solution of the transport equation is obtained in the limit of narrow minibands of the superlattice. We did not use the Kadanov–Beym approximation since the two-time nature of Green's functions plays the major role. The effectiveness of the proposed method is demonstrated for the photon-induced phonon-free hopping transport between the levels of the Stark ladder. This contribution to current cannot be obtained using the method of a one-time density matrix or through the Kadanov–Beym approximation, which is employed in most theoretical publications. An important property of phonon-free current is the possibility of its flowing opposite to the dc electric field accompanied by the absorption of the electromagnetic energy. We believe that the experimental investigations of phonon-free current transport will clarify the nature of two-time transport for which the finiteness of the scattering act duration must be taken into consideration.

Many years ago, phonon-free transport was studied for the localization in disordered systems on the basis of the balance equations for the electron density at a lattice site [35]. The relations obtained in [35] have a structure similar to Eq. (53). The main difference is that the degree of localization in systems with a Stark ladder is associated with the magnitude of electric field: as the field becomes smaller, a transition from the hopping to band transport takes place. The proposed method allows in principle to describe both these modes and a transition from one mode to the other in the region $\Omega_{dc}\tau \approx 1$.

ACKNOWLEDGMENTS

The authors are grateful to Deutsches Zentrum für Luft- und Raumfahrt for financial support.

APPENDIX A

Kinetic Equation for G^-

In this Appendix, we present the intermediate results which were used in deriving the kinetic equation for $g_l^-(t, t')$. We sum system of equations (7) and the complex-conjugate system, where the substitution $t \longleftrightarrow t'$ is made:

$$\left[\frac{\partial}{\partial t}+\frac{\partial}{\partial t'}+\frac{i}{\hbar}\left(\varepsilon\left(\mathbf{k}-\frac{1}{2}\mathbf{A}(t', t)\right)\right)-\varepsilon\left(\mathbf{k}-\frac{1}{2}\mathbf{A}(t, t')\right)\right.$$
$$\left.+\frac{1}{2}\left(\frac{\partial}{\partial t}-\frac{\partial}{\partial t'}\right)\mathbf{A}(t', t)\nabla_k\right]\tilde{G}^{\pm}(\mathbf{k}|t, t')$$
$$= \mp i\left\{\int_{t'}^{t} dt_1\left[\tilde{\Sigma}^{\pm}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t')|t, t_1\right)\right.\right.$$
$$\times\tilde{G}^{\pm}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t)|t_1, t'\right)$$
$$\left.-\tilde{\Sigma}^{\pm}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t)|t_1, t'\right)\tilde{G}^{\pm}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t')|t, t_1\right)\right] \quad \text{(A.1)}$$
$$+\int_{-\infty}^{t'} dt_1\left[\tilde{\Sigma}^{\pm}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t')|t, t_1\right)\tilde{G}^{\mp}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t)|t_1, t'\right)\right.$$
$$\left.-\tilde{\Sigma}^{\mp}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t)|t_1, t'\right)\tilde{G}^{\pm}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t')|t, t_1\right)\right]$$
$$-\int_{-\infty}^{t} dt_1\left[\tilde{\Sigma}^{\mp}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t')|t, t_1\right)\tilde{G}^{\pm}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t)|t_1, t'\right)\right.$$
$$\left.\left.-\tilde{\Sigma}^{\pm}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t)|t_1, t'\right)\tilde{G}^{\mp}\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t')|t, t_1\right)\right]\right\}.$$

Using Eq. (8), we can now go over from $G^{\pm}$ to $g^{\pm}$. Taking into account only the main contribution for a nondegenerate electron gas, we obtain from Eq. (A.1)

$$\left[\frac{\partial}{\partial t}+\frac{\partial}{\partial t'}+\frac{1}{2}\left(\frac{\partial}{\partial t}-\frac{\partial}{\partial t'}\right)\mathbf{A}(t', t)\nabla_{\mathbf{k}}\right]g^-(\mathbf{k}|t, t')$$
$$= \sum_{\mathbf{q}}\left\{\int_{-\infty}^{t'} dt_1\left[D_{\mathbf{q}}^-(t_1-t)\Phi_{t_1t}(\mathbf{k}, \mathbf{q}|t, t')\right.\right.$$
$$\times g^-\left(\mathbf{k}+\mathbf{q}+\frac{1}{2}\mathbf{A}(t_1, t')|t, t_1\right)g^+\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t)|t_1, t'\right)$$
$$-D_{\mathbf{q}}^+(t'-t_1)\Phi_{t_1t'}^*(\mathbf{k}, \mathbf{q}|t, t')$$
$$\times g^+\left(\mathbf{k}+\mathbf{q}+\frac{1}{2}\mathbf{A}(t_1, t)|t_1, t'\right)$$
$$\left.\times g^-\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t')|t, t_1\right)\right] \quad \text{(A.2)}$$
$$-\int_{-\infty}^{t} dt_1\left[D_{\mathbf{q}}^+(t_1-t)\Phi_{t_1t}(\mathbf{k}, \mathbf{q}|t, t')\right.$$
$$\times g^+\left(\mathbf{k}+\mathbf{q}+\frac{1}{2}\mathbf{A}(t_1, t')|t, t_1\right)g^-\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t)|t_1, t'\right)$$
$$-D_{\mathbf{q}}^-(t'-t_1)\Phi_{t_1t'}^*(\mathbf{k}, \mathbf{q}|t, t')$$
$$\left.\left.\times g^-\left(\mathbf{k}+\mathbf{q}+\frac{1}{2}\mathbf{A}(t_1, t)|t_1, t'\right)g^+\left(\mathbf{k}+\frac{1}{2}\mathbf{A}(t_1, t')|t, t_1\right)\right]\right\},$$

where the quantity Φ_{t_1t} is defined by relation (15). In the special case of a one-dimensional system, using Eq. (16), we obtain

$$\left[\frac{\partial}{\partial t}+\frac{\partial}{\partial t'}+il\Omega_{dc}\cos\left(\frac{\omega_{ac}}{2}(t+t')\right)\cos\left(\frac{\omega_{ac}}{2}(t-t')\right)\right]$$
$$\times g_l^-(t, t') = \int_{-\infty}^{t'} dt_1[D^-(t_1-t)A_l(t, t'|t_1)g_0^-(t, t_1)g_0^+(t_1, t')$$
$$-D^+(t'-t_1)B_l(t, t'|t_1)g_0^+(t_1, t')g_0^-(t, t_1)] \quad \text{(A.3)}$$
$$-\int_{-\infty}^{t} dt_1[D^+(t_1-t)A_l(t, t'|t_1)g_0^+(t, t_1)g_0^-(t_1, t')$$
$$-D^-(t'-t_1)B_l(t, t'|t_1)g_0^-(t_1, t')g_0^+(t, t_1)] \equiv P_l(t, t'),$$

where the quantities $A_l(t, t'|t_1)$ and $B_l(t, t'|t_1)$ are defined by relations (34) and (35).

APPENDIX B

Solution of Kinetic Equation

In this Appendix, we solve analytically the kinetic equation for Green's function $g_l^-(T, 0)$. Introducing the new variables T and t_- in accordance with relations (17), we obtain from Eq. (A.3)

$$\left[\frac{d}{dT}+i\frac{dA_l(T)}{dT}\right]g_l^-(T, 0)$$
$$= \int_0^{\infty} dt_1\left\{D^-(-t_1)A_l(t, t|t-t_1)g_0^-\left(T-\frac{t_1}{2}, -t_1\right)g_0^+\left(T-\frac{t_1}{2}, t_1\right)\right.$$

$$
\begin{aligned}
&- D^{+}(t_1) B_l(t, t|t-t_1) g_0^{+}\left(T - \frac{t_1}{2}, t_1\right) g_0^{-}\left(T - \frac{t_1}{2}, -t_1\right) \\
&+ D^{-}(t_1) B_l(t, t|t-t_1) g_0^{-}\left(T - \frac{t_1}{2}, t_1\right) g_0^{+}\left(T - \frac{t_1}{2}, -t_1\right) \\
&- D^{+}(-t_1) A_l(t, t|t-t_1) g_0^{+}\left(T - \frac{t_1}{2}, -t_1\right) \\
&\times g_0^{-}\left(T - \frac{t_1}{2}, t_1\right) \Bigg\} \equiv P_l(T),
\end{aligned}
\tag{B.1}
$$

where the right-hand side of the equation is denoted by $P_l(T)$. The quantity $A_l(T)$ is defined by the equation

$$
\frac{dA_l(T)}{dT} = l\Omega_{dc} + l\Omega_{ac}\cos(\omega_{ac}T) \tag{B.2}
$$

and satisfies the relation

$$
A_l\left(T + \frac{2\pi}{\omega_{ac}}\right) = A_l(T) + 2\pi l\frac{\Omega_{dc}}{\omega_{ac}}. \tag{B.3}
$$

The linear differential equation (B.1) can be solved easily,

$$
\begin{aligned}
g_l^{-}(T, 0) &= \exp[-iA_l(T)] \\
&\times \left\{C + \int_0^T dT' P_l(T') \exp[iA_l(T')]\right\},
\end{aligned}
\tag{B.4}
$$

and constant C is defined by the periodicity condition,

$$
\begin{aligned}
C &= \left[1 - \exp\left(-2\pi i l\frac{\Omega_{dc}}{\omega_{ac}}\right)\right]^{-1} \\
&\times \int_{-2\pi/\omega_{ac}}^{0} dT' P_l(T') \exp[iA_l(T')].
\end{aligned}
\tag{B.5}
$$

Integrating with respect to T in (B.4) and (B.5) using the relation

$$
\begin{aligned}
&\exp\left(il\frac{\Omega_{ac}}{\omega_{ac}}\sin(\omega_{ac}T)\right) \\
&= \sum_{k=-\infty}^{\infty} J_k\left(l\frac{\Omega_{ac}}{\omega_{ac}}\right)\exp(ik\omega_{ac}T)
\end{aligned}
\tag{B.6}
$$

and carrying out the Fourier transformation

$$
P_l(T) = \sum_{m=-\infty}^{\infty} P_{lm}\exp(im\omega_{ac}T), \tag{B.7}
$$

we obtain the final result that we employed in Section 5:

$$
\begin{aligned}
g_l^{-}(T, 0) &= -i\sum_{m=-\infty}^{\infty} P_{lm} \sum_{k, k'=-\infty}^{\infty} J_k\left(l\frac{\Omega_{ac}}{\omega_{ac}}\right) J_{k'}\left(l\frac{\Omega_{ac}}{\omega_{ac}}\right) \\
&\times \frac{\exp[i(m + k - k')\omega_{ac}T]}{l\Omega_{dc} + (k + m)\omega_{ac}}.
\end{aligned}
\tag{B.8}
$$

APPENDIX C

Alternative Expression for Current

In this Appendix, we present a form of current density other than (53) for $\sqrt{U}/\omega_{ac} \ll 1$ and $\omega_{ac}\tau_{ac} \longrightarrow \infty$. In this case, expression (53) can be written in the form

$$
j_z = \frac{en_sU\Delta^2}{8\hbar^2}AB, \tag{C.1}
$$

where

$$
\begin{aligned}
A &= \frac{1}{\omega_{ac}}\sum_{k'=-\infty}^{\infty}\frac{J_{k'}^2(\nu')}{\nu + k'} \\
&= \frac{\pi}{\omega_{ac}\sin\pi\nu}J_{\nu}(\nu')J_{-\nu}(\nu').
\end{aligned}
\tag{C.2}
$$

The other (field-dependent) factor in Eq. (C.1) has the form

$$
\begin{aligned}
B &= \frac{1}{\omega_{ac}}\sum_{k=-\infty}^{\infty}\frac{J_k^2(\nu')}{k + \nu}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}g_{00}^{+}(\omega)g_{00}^{+}(\omega + \omega_{ac}(k + \nu)) \\
&\times [f_{00}(\omega + \omega_{ac}(k + \nu)) - f_{00}(\omega)].
\end{aligned}
\tag{C.3}
$$

Since the level width for the density of states is of the order of $\sqrt{U}$ and is larger than ω_{ac}, the index k of Bessel's functions can be replaced by $-\nu$. As a result, we obtain

$$
B \approx J_{\nu}^2(\nu')F(\nu)/\omega_{ac}, \tag{C.4}
$$

$$
\begin{aligned}
F(\nu) &= \sum_{k=-\infty}^{\infty}\frac{1}{\nu - k}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}g_{00}^{+}(\omega)g_{00}^{+}(\omega + \omega_{ac}(k + \nu)) \\
&\times [f_{00}(\omega + \omega_{ac}(k + \nu)) - f_{00}(\omega)] \\
&= 2\sum_{k=-\infty}^{\infty}\frac{\sinh[\hbar\omega_{ac}(\nu - k)/2k_BT]}{\nu - k}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi} \\
&\times g_{00}^{+}\left[\omega - \frac{\omega_{ac}}{2}(\nu - k)\right]g_{00}^{+}\left[\omega + \frac{\omega_{ac}}{2}(\nu - k)\right]f_{00}(\omega).
\end{aligned}
\tag{C.5}
$$

Relations (C.2), (C.4), and (C.5) lead to an expression for current in form (56) if we use relation (54) for $f_{00}(\omega)$.

REFERENCES

1. B. J. Keay, S. Zeuner, S. J. Allen, *et al.*, Phys. Rev. Lett. **75**, 4102 (1995).
2. B. J. Keay, S. J. Allen, J. Galan, *et al.*, Phys. Rev. Lett. **75**, 4098 (1995).
3. S. Zeuner, D. J. Keay, S. J. Allen, *et al.*, Phys. Rev. B **53**, R1717 (1996).
4. K. Unterrainer, D. J. Keay, M. C. Wanke, *et al.*, Phys. Rev. Lett. **76**, 2973 (1996).
5. S. Winnerl, E. Schomburg, J. Grenzer, *et al.*, Phys. Rev. B **56**, 10303 (1997).
6. X. L. Lei, B. Dong, and Y. Q. Chen, Phys. Rev. B **56**, 12120 (1997).
7. J. B. Xia, Phys. Rev. B **58**, 3565 (1998).
8. E. W. S. Caetano, E. A. Mendes, V. N. Freire, *et al.*, Phys. Rev. B **57**, 11872 (1998).
9. X. L. Lei and H. L. Cui, Eur. Phys. J. B **4**, 513 (1998).
10. G. Platero and R. Aguado, Appl. Phys. Lett. **70**, 3546 (1997).
11. R. Aguado and G. Platero, Phys. Rev. B **55**, 12860 (1997).
12. A. Wacker, A. P. Jauho, S. Zeuner, and S. J. Allen, Phys. Rev. B **56**, 13268 (1997).
13. A. A. Ignatov and Yu. A. Romanov, Fiz. Tverd. Tela (Leningrad) **17**, 3388 (1975) [Sov. Phys. Solid State **17**, 2216 (1975)].
14. A. A. Ignatov and Yu. A. Romanov, Phys. Status Solidi B **73**, 327 (1976).
15. V. V. Pavlovich and E. M. Épshteĭn, Fiz. Tekh. Poluprovodn. (Leningrad) **10**, 1196 (1976) [Sov. Phys. Semicond. **10**, 710 (1976)].
16. S. Y. Mensah, J. Phys. C **4**, L325 (1996).
17. A. A. Ignatov, E. Schomburg, J. Grenzer, *et al.*, Z. Phys. B **98**, 187 (1995).
18. O. M. Yevtushenko, Phys. Rev. B **54**, 2578 (1996).
19. H. N. Nazareno and R. A. Masut, Solid State Commun. **101**, 819 (1997).
20. X. G. Zhao, G. A. Georgaki, and Q. Niu, Phys. Rev. B **56**, 3976 (1997).
21. V. M. Polyanovskiĭ, Fiz. Tverd. Tela (Leningrad) **22**, 1975 (1980) [Sov. Phys. Solid State **22**, 1151 (1980)].
22. V. V. Bryksin and P. Kleinert, Phys. Rev. B **59**, 8152 (1999).
23. P. Kleinert and V. V. Bryksin, J. Phys.: Condens. Matter **11**, 2539 (1999).
24. V. V. Bryksin and Yu. A. Firsov, Zh. Éksp. Teor. Fiz. **61**, 2373 (1971) [Sov. Phys. JETP **34**, 1272 (1971)].
25. L. Esaki and R. Tsu, IBM J. Res. Dev. **14**, 61 (1970).
26. H. Haug and A. P. Jauho, *Quantum Kinetics in Transport and Optics of Semiconductors* (Springer-Verlag, Berlin, 1996).
27. P. Lipavsky, V. Spicka, and B. Velicky, Phys. Rev. B **34**, 6933 (1986).
28. V. V. Bryksin and P. Kleinert, J. Phys. A **32**, 2731 (1999).
29. V. V. Bryksin and P. Kleinert, J. Phys. A **33**, 233 (2000).
30. S. Zeuner, B. J. Keay, S. J. Allen, *et al.*, Superlatt. Microstruct. **22**, 149 (1997).
31. A. Wacker and A. P. Jauho, Phys. Status Solidi B **204**, 73 (1997).
32. A. Wacker, S. J. Allen, J. S. Scott, *et al.*, Phys. Status Solidi B **204**, 95 (1997).
33. P. R. Tien and J. P. Gordon, Phys. Rev. **129**, 647 (1963).
34. J. Tucker, IEEE J. Quantum Electron. **QE-15**, 1234 (1979).
35. H. Boettger and V. V. Bryksin, *Hopping Conduction in Solids* (Akademie-Verlag, Berlin, 1985).
36. V. V. Bryksin and P. Kleinert, J. Phys. C **9**, 7403 (1997).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 1084–1089.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1250–1256.

SOLIDS
Electronic Properties

Effect of Disorder on the Transport Properties of the High-T_c Superconductor $Nd_{2-x}Ce_xCuO_{4+\delta}$

A. N. Ignatenkov[a], A. I. Ponomarev[a,*], L. D. Sabirzyanova[a], T. B. Charikova[a], G. I. Kharus[a], N. G. Shelushinina[a], A. A. Ivanov[b], and O. A. Churkin[b]

[a]*Institute of Metal Physics, Ural Division, Russian Academy of Sciences, ul. S. Kovalevskoĭ 12, Yekaterinburg, 620219 Russia*
[b]*Moscow State Institute of Engineering Physics, Kashirskoe sh. 31, Moscow, 115410 Russia*
**e-mail: ponomarev@imp.uran.ru*
Received December 22, 2000

Abstract—The temperature dependences $\rho_{ab}(T)$ of $Nd_{2-x}Ce_xCuO_{4+\delta}$ single crystals with $0 \le x \le 0.20$ are studied and analyzed on the basis of the concepts in the theory of disordered 2D systems. The results are compared with the data obtained for other copper-oxide HTSC. It is found that a transition to the superconducting state in the optimal doping region $0.14 \le x \le 0.18$ occurs only in crystals with a fairly small degree of disorder ($k_F l \ge 2$, where l is the mean free path). This transition is compatible with the weak 2D-localization mode as long as the localization radius is longer than the characteristic size of a Cooper pair. The superconducting transition temperature in the optimal doping region increases monotonically with the parameter $k_F l$ characterizing the degree of disorder in the crystal. The degradation of superconducting properties upon a further increase in the doping level ($x > 0.18$) is apparently associated with a transition from 2D to 3D conductivity in the single crystal. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well known that superconductivity emerges in a wide range of copper-oxide HTSC compounds as a result of doping the initial dielectric matrix with nonisovalent impurities. By way of a simple example, we can mention the doping of the dielectric system La_2CuO_4 with strontium (which leads to the emergence of holes [1]) or of the dielectric Nd_2CuO_4 with cerium, leading to the emergence of electrons [2] in CuO_2 planes. A comparative analysis of phase diagrams $T_c(x)$ for the superconducting transition temperature as a function of the dopant concentration for these two systems is given in the review by Dagotto [3].

As a result of the substitution of Sr^{2+} for La^{3+} (Ce^{4+} for Nd^{3+}) and the migration of extra holes (electrons) to CuO_2 planes, the buffer layer between these planes acquires a potential of charged impurity centers distributed randomly in the lattice. The disorder generated by this random impurity potential hinders the free motion of carriers and may lead to their localization. Consequently, the insulator–superconductor transition is closely related to the insulator–metal transition [4]. This is analogous to the doping of semiconductors with acceptors or donors, the essential difference being that HTSC materials display clearly manifested 2D properties of charge carriers in macroscopic 3D crystals. For this reason, it is natural to use the ideas developed in the physics of disordered 2D systems [5] for studying transport phenomena in these materials. An analysis of the effect of disorder introduced in the course of doping on the transport properties of HTSC (the magnitude and temperature dependence of resistivity, the presence or absence of superconductivity, and the superconducting transition temperature) has been carried out by many authors.

For example, Beschoten *et al.* [6] investigated the emergence of the superconducting transition upon an increase in the doping level of calcium in the insulator $Bi_2Sr_2PrCu_2O_8$. The substitution of Ca^{2+} for Pr^{3+} generates holes and leads to the insulator–superconductor transition in $Bi_2Sr_2(Pr_{1-z}Ca_z)Cu_2O_8$ for $z = z_c = 0.52$. The authors of [6] considered the closeness of this transition to the insulator–metal transition and found that superconductivity in this case coexists with the spatial localization of charge carriers for $z > z_c$.

Mandrus *et al.* [7] studied the degradation of superconductivity in the $Bi_2Sr_2(Ca_{1-x}Y_x)Cu_2O_8$ system upon an increase in yttrium concentration. The authors analyzed the temperature dependence of resistivity in the *ab* plane for $0.2 \le x \le 0.8$ and attributed the disappearance of the superconducting transition for $x_c = 0.45$ directly to the metal–insulator transition due to a decrease in the concentration of holes and their localization in the fluctuations of the random potential created as a result of substitution of trivalent yttrium for bivalent calcium.

Schlenga *et al.* [8] studied in detail the resistance and magnetoresistance for a series of monocrystalline

$La_{2-x}Ba_xCuO_{4-\delta}$ samples with the barium concentration $x = (0.05 \pm 0.01)$–(0.075 ± 0.01) in the region of the insulator–metal transition induced by the substitution of Ba^{2+} for La^{3+}. Analyzing the temperature dependence $\rho_{ab}(T)$, the authors observed a gradual transition from strong to weak localization (and superconductivity) upon an increase in x. The absence of a sharp concentration boundary for the insulator–metal transition was attributed to the 2D-type conductivity in the system.

The dependence of resistivity ρ_{ab} in the normal phase (for fixed temperature $T = 40$ K) on the strontium concentration for a series of $La_{2-x}Sr_xCuO_4$ samples with $x < 0.15$ presented by Karpinska *et al.* [9] led to the conclusion that an insulator–metal transition occurs for $x = 0.05$. For samples with Sr concentration in the immediate vicinity of this transition ($x = 0.048$ and $x = 0.051$), the temperature dependence of resistivity ρ_{ab} was measured in magnetic fields up to 8.5 T at low and ultralow temperatures down to 30 mK. It was found that the normal state of these samples after the destruction of superconductivity by a magnetic field is insulating ($\rho \longrightarrow \infty$ for $T \longrightarrow 0$); i.e., the magnetic field-induced superconductor–insulator transition takes place.

Fukuzumi *et al.* [10] thoroughly investigated the process of replacement of copper in the CuO_2 plane by another element (zinc in the present case). As an isovalent impurity, zinc does not change the initial concentration of carriers but, being a strong scatterer with a short-range potential, causes a disorder in the system due to a difference in ion cores. The temperature dependence of the resistivity was measured in the *ab* plane for $La_{2-x}Sr_xCu_{1-z}Zn_zO_4$ single crystals with $0.1 \le x \le 0.3$ and for $YBa_2(Cu_{1-z}Zn_z)_3O_{7-y}$ single crystals with an oxygen concentration varying from 6.63 to 6.93 ($0.06 \le y \le 0.37$) for various concentrations of Zn ($z \le 0.04$). It was shown that the introduction of even a few percent of Zn increases the residual resistance sharply, decreases T_c, and leads to a superconductor–insulator transition. The lower the hole concentration in the matrix (Sr concentration in the La system or the oxygen concentration in the Y system), the easier the superconductivity degradation; i.e., the superconductor–insulator transition requires a lower concentration of Zn.

In the present work, we study the effect of doping as well as various annealing conditions (oxygen concentration) on the transport properties of $Nd_{2-x}Ce_xCuO_{4+\delta}$ single crystals, separating (whenever possible) these two ways of introducing disorder in the system. A transition from the dielectric to the superconducting state upon a change in the annealing conditions in $Nd_{2-x}Ce_xCuO_{4+\delta}$ was also studied earlier [11–14], but no systematic studies were carried out (to our knowledge) for $Nd_{2-x}Ce_xCuO_{4+\delta}$ single crystals in a wide range of cerium concentration under optimal annealing ($\delta \longrightarrow 0$).

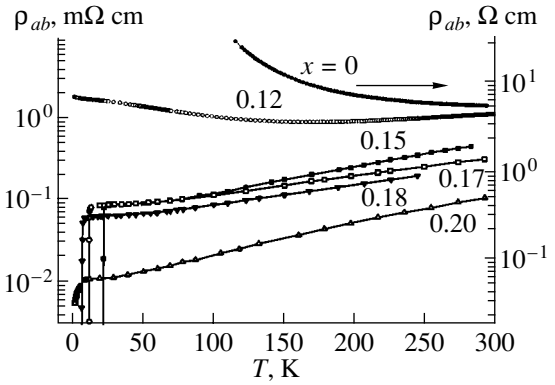


Fig. 1. Temperature dependences of resistivity ρ_{ab} of $Nd_{2-x}Ce_xCuO_{4+\delta}$ samples with various Ce concentrations.

2. EXPERIMENTAL RESULTS AND DISCUSSION

2.1. Doping with Cerium

We measured the resistivity ρ_{ab} in the CuO_2 planes for optimally annealed monocrystalline films of $Nd_{2-x}Ce_xCuO_{4+\delta}$ with $0.12 \le x \le 0.20$. The technique of growth, the conditions of annealing, and other characteristics of these samples are described in [15, 16]. The samples having the highest value of T_c for a given x are regarded as optimally annealed. The temperature dependences $\rho_{ab}(T)$ in the temperature range $T = 1.4$–300 K are presented in Fig. 1. The figure shows for comparison the results obtained for an undoped bulk single crystal of Nd_2CuO_4 at $T = 116$–300 K.

Figure 2a shows the dependence of resistivity on the Ce concentration for our series of the samples. Besides, the available data for the best single crystals of $Nd_{2-x}Ce_xCuO_{4+\delta}$ with $x = 0.14$ [14] and 0.16 [12] are also presented. The values of $\rho_{ab}(T)$ are taken for $T = T_c$ for superconducting samples and at $T = 4.2$ K for nonsuperconducting ones. In the same form, Fig. 2a presents the results obtained by Fournier *et al.* [17] for epitaxial monocrystalline $Pr_{2-x}Ce_xCuO_4$ films with $0.13 \le x \le 0.20$ (dashed curve).

Figure 2b presents the phase diagram $T_c(x)$ for the same samples as in Fig. 2a. Besides, the results obtained in [18] are presented, which are the same for Nd- and Pr-containing systems with $x = 0.15$: $T_c = 24$ K ($\rho_{ab} = 8 \times 10^{-5}$ Ω cm at $T = 50$ K). The $T_c(x)$ dependence for the Nd system is in accord with the phase diagram available for the best polycrystalline samples of $Nd_{2-x}Ce_xCuO_{4+\delta}$ [2, 3]. Note that not only are the dependences $\rho_{ab}(x)$ and $T_c(x)$ in the Nd and Pr systems similar qualitatively, but the numerical values of these quantities are also close in samples with close stoichiometric compositions in oxygen ($\delta \longrightarrow 0$). Slightly lower values of ρ_{ab} for $x < 0.18$, as well as systematically higher values of T_c for the Pr system, are probably

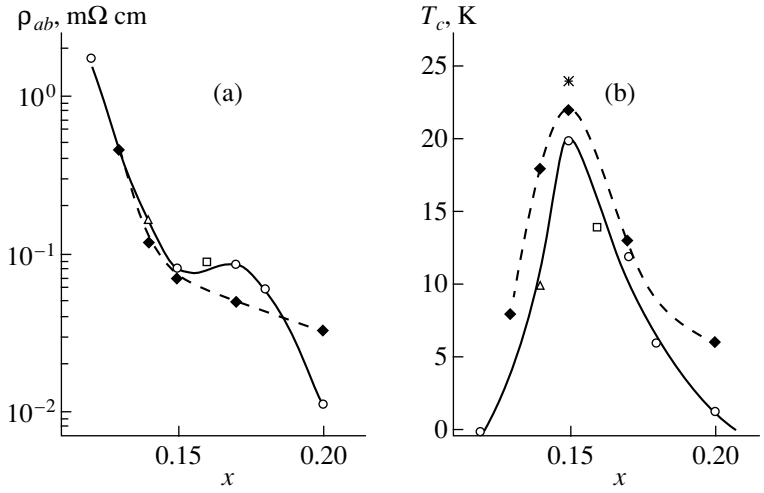


Fig. 2. Dependences of (a) resistivity ρ_{ab} and (b) superconducting transition temperature T_c on the cerium concentration x: our results (light circles), the results obtained in [12] (light squares) and [14] (light triangles) for the Nd system; dark rhombi correspond to the results obtained in [17] for a Pr system, while the asterisk corresponds to the results from [18] for Nd- and Pr systems.

due to a better quality of epitaxially grown films of $Pr_{2-x}Ce_xCuO_4$.

The following three segments can be singled out on the $\rho_{ab}(x)$ curve for $Nd_{2-x}Ce_xCuO_4$ (Fig. 2a): a sharp decrease in the resistivity of samples with a lower ($x <$ 0.12–0.14) doping level ($\rho_{ab} = 26.5\ \Omega$ cm for Nd_2CuO_4 at $T = 116$ K, which is more than four orders of magnitude larger than the value of ρ_{ab} for $x = 0.12$); a comparatively weak dependence $\rho_{ab}(x)$ for samples with $0.14 \le x \le 0.18$, and a noticeably sharper (as compared to $1/x$) decrease in the value of ρ_{ab} in the region with a higher doping level ($x > 0.18$). A decrease in the resistivity by several orders of magnitude upon an increase in the Ce concentration from $x = 0$ to $x = 0.12$–0.14 is undoubtedly associated with the transition from the strong exponential localization mode to the weak logarithmic localization mode for charge carriers in cuprate planes (insulator–metal transition in a disordered 2D system). This is confirmed by the temperature dependence of resistivity. For $x = 0$, we have $\rho_{ab} \propto \exp(-E_A/kT)$ with $E_A \approx 60$ meV [19]. On the other hand, the effects of weak 2D localization (logarithmic increase in ρ_{ab} upon a decrease in temperature and a negative magnetoresistance) are clearly manifested in the sample with $x =$ 0.12 [16].

Indeed, it was mentioned above that the substitution of Ce^{4+} for Nd^{3+} leads, on the one hand, to the emergence of electrons in the CuO_2 plane, and on the other hand, to a disorder generated by the random distribution of Ce ions. As in the case of semiconductors, localization or delocalization of charge carriers is determined by the ratio of the kinetic energy (Fermi energy ε_F) and the potential energy of interaction of electrons with impurities (random potential amplitude γ) [20]. For 2D systems, $\varepsilon_F \propto n_s$, where n_s is the 2D electron concentration, and the amplitude γ of impurity potential fluctuations is proportional to N_s, where $N_s = n_s$ is the surface density of impurities [21]. Obviously, $\gamma \gg \varepsilon_F$ for low electron concentrations, which corresponds to an insulator (strong localization). For high concentrations, $\varepsilon_F \gg \gamma$, the quasi-metallic mode for the mean free path is realized with quantum interference corrections due to the weak localization effects. The condition $\varepsilon_F \gg \gamma$ corresponds to the "flooding" by electrons of the potential relief at the bottom of the band, where conduction is observed (the two-dimensional $pd\sigma$ band formed by the electron states in the CuO_2 plane [22]).

An Anderson-type insulator–"metal" transition takes place in the region $\gamma \approx \varepsilon_F$. The parameter characterizing the extent of disorder may be the 2D conductivity σ_{2D} in dimensionless units: $g = h\sigma_{2D}/e^2$. A transition from strong localization to metal-type conduction with weak localization effects in a 2D system corresponds to the condition $g \approx 1$. In the "metallic" region ($g \gg 1$), we have [5]

$$\sigma_{2D} = \frac{e^2}{h} k_F l, \tag{1}$$

where $k_F = (2\pi n)^{1/2}$ is the Fermi quasimomentum and l is the mean free path.

The $\rho_{ab}(T)$ curve for a sample with $x = 0.12$ (which does not undergo a superconducting transition) clearly displays segments typical of metal-type conduction and corresponding to a semiconductor with the minimum value of $\rho_{ab} \approx 8 \times 10^{-4}\ \Omega$ cm at $T \approx 150$ K (Fig. 1). The estimate

$$\sigma_{2D} = \rho_{ab}^{-1} c \tag{2}$$

of the partial conductivity in a CuO_2 layer, where $c = 6$ Å is the separation between cuprate planes, shows that this value of resistivity corresponds to $k_F l \approx 2$. In actual practice, a superconducting transition is observed only in samples with a high metal-type conductivity $g = k_F l \gg 1$ ($10 \le k_F l \le 25$) for samples with $0.14 \le x \le 0.18$. The data for stoichiometric Nd systems with a Ce concentration between $x = 0.12$ ($k_F l \approx 2$) and $x = 0.14$ ($k_F l \approx 10$) are not available in the literature. In a Pr system whose crystal structure and physical properties are close to those of a Nd system, a clearly manifested superconducting transition with $T_c = 8$ K is observed in a sample with $x = 0.13$ ($k_F l \approx 3.4$) [17] (see Figs. 2a and 2b). Thus, we can assume that doping the stoichiometric samples of Nd or Pr systems with cerium leads to a transition from an Anderson-type insulator to a superconductor for $k_F l > (2–3)$.

A noticeable decrease (to about one-sixth) in the value of ρ_{ab} upon a slight variation of x from 0.18 to 0.20 (see Fig. 2a) is accompanied by a decrease in T_c, which is apparently associated with the second metallization stage [4], i.e., the activation of charge transfer between CuO_2 layers and a transition from the 2D to 3D conduction. This is how Eliashberg interpreted in his review [4] the experimental results obtained by Sreedhar and Ganguly [1] on the "anomalous" degradation of the superconductivity for a high hole concentration ($x > 0.3$) in the $La_{2-x}Sr_xCuO_4$ system: the disappearance of the superconducting transition is accompanied by a decrease in resistivity ρ_{ab} to a fraction of its initial value. The transition from a 2D- to 3D-type of conduction in the La system for $x \approx 0.3$ [23] and in the Nd system for $x \approx 0.18$ [19] is manifested in the change in the temperature dependence of resistivity along the c axis from insulator- to metal-type and in a strong decrease in the anisotropy ρ_c/ρ_{ab}.

2.2. The Role of Nonstoichiometry

We also analyzed the $\rho_{ab}(T)$ dependence in the $Nd_{2-x}Ce_xCuO_{4+\delta}$ system for various concentrations δ of oxygen (different annealing conditions) for a fixed x. Figure 3 presents our results obtained for $x = 0.18$: in the absence of annealing, for an intermediate annealing mode ($T_c = 4$ K), and for optimal annealing in vacuum for 40 min at 800°C ($T_c = 6$ K).

It is well known that $Nd_{2-x}Ce_xCuO_{4+\delta}$ samples which are not subjected to annealing are not superconducting even at a close-to-optimal doping level ($0.14 \le x \le 0.18$). The formation of a phase with zero resistivity requires annealing the samples in nitrogen [11], argon [11, 12, 14], or in vacuum [13] for lowering the oxygen concentration. According to modern concepts [11, 13], the main role of annealing lies in the removal of "excess" nonstoichiometric oxygen ($\delta \longrightarrow 0$). Excess oxygen, occupying the apex oxygen sites, which are vacant in optimally annealed $Nd_{2-x}Ce_xCuO_{4+\delta}$, creates a random impurity potential which localizes electrons supplied by cerium. As excess oxygen atoms are removed during annealing, the disorder in the system decreases and the mean free path increases. The decrease in the extent of disorder must lead to a transition from the strong localization mode to a mode of weak localization of charge carriers in cuprate planes for $k_F l \approx 1$.

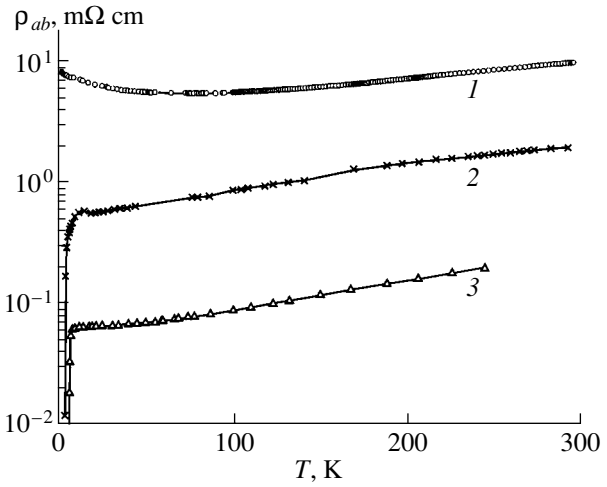


Fig. 3 Temperature dependences of resistivity ρ_{ab} of $Nd_{1.82}Ce_{0.18}CuO_{4+\delta}$ samples for various annealing modes: (*1*) in the absence of annealing ($T_c = 0$), $k_F l = 0.25$; (*2*) at intermediate annealing ($T_c = 4$ K), $k_F l = 2.5$; (*3*) at optimum annealing ($T_c = 6$ K), $k_F l = 25$.

According to Sadovskii [24], superconductivity is observed in disordered systems when the carrier localization radius R_{loc} exceeds the size ξ of a Cooper pair. For small mean free paths, in the so-called "dirty" limit, when $l < \xi_0$ (ξ_0 is the coherence length in the pure substance), we have $\xi = (\xi_0 l)^{1/2}$. For the carrier localization radius in 2D systems under weak localization conditions ($k_F l > 1$), the following estimate is valid [5]:

$$R_{loc} \approx l \exp\left(\frac{\pi}{2} k_F l\right). \qquad (3)$$

As we go over to the weak localization regime, the correlation length ξ increases with l according to a power law, while the localization radius increases exponentially. For this reason, the condition $R_{loc} > \xi$ may be attained even for $k_F l$ of the order of several units.

Indeed, a $Nd_{1.82}Ce_{0.18}CuO_{4+\delta}$ single crystal is superconducting even for $k_F l = 2.5$ (see Fig. 3). In $Nd_{2-x}Ce_xCuO_{4+\delta}$ samples with $x = 0.15$ [14] and 0.18 [12], in which the oxygen concentration decreases as a result of annealing, a superconducting transition was also observed only for $k_F l > 2$. For example, Tanda *et al.* [12] presented the $\rho_{ab}(T)$ dependences for eight samples with $x = 0.18$ at various stages of annealing from samples in the insulator phase to superconducting

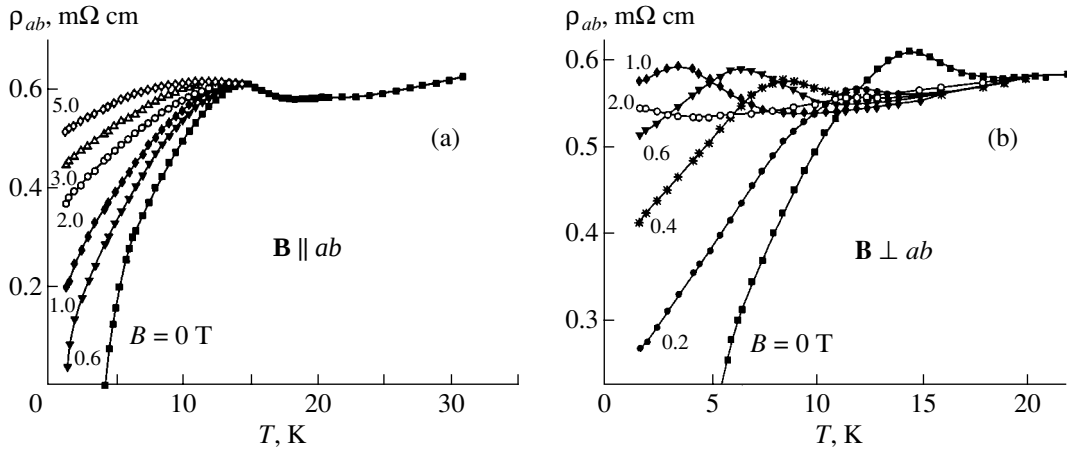


Fig. 4. Temperature dependences of the resistivity ρ_{ab} of a $Nd_{1.82}Ce_{0.18}CuO_{4+\delta}$ sample (sample *2* in Fig. 3) in a magnetic field (a) parallel and (b) perpendicular to the *ab* plane.

samples: $0.1 \le g \le 3.5$. They noted that at the boundary of the disorder-induced transition from the superconductor to the insulator, the resistivity of the sample corresponds to $k_F l \approx 3$ ($R_{2D} = \rho_{ab}/c = 8.5$ kΩ).

Thus, a change in the disorder in the $Nd_{2-x}Ce_xCuO_{4+\delta}$ system both due to doping (variation of x for $\delta \approx$ const), as well as due to annealing (variation of δ for x = const), leads to the emergence of a superconducting transition only in crystals with a high conductivity. In all the publications mentioned above [6–10, 12, 14], it was established experimentally that the maximum surface resistance R_{2D} in the normal phase, from which a transition to the superconducting phase is still possible, is close to the universal theoretical value $R^*_{2D} = h/(2e)^2$ ($\sigma^*_{2D} = 4e^2/h$). This quantity corresponds to the conductivity quantum of a Cooper pair with a charge $2e$ and emerges in the Bose scenario of the transition from a superconductor to an insulator upon an increase in disorder [25]. Tanda *et al.* [12] consider this fact as a serious argument in favor of precisely the Bose scenario of the superconducting transition. However, the conductivity of a Fermi gas above the boundary of the insulator–metal transition ($k_F l \ge 1$) attains approximately the same value. Besides, the temperature and magnetic-field dependences of resistivity ρ_{ab} in the normal phase clearly display the features typical of the weak localization effects for Fermi particles in a disordered system. Weak localization defects will be discussed in greater detail below.

2.3. Effects of Weak 2D Localization

The suppression of superconductivity by a magnetic field allows us to analyze the properties of the normal state of a superconductor below the superconducting transition temperature ($T < T_c$). Experiments on the $Nd_{1.82}Ce_{0.18}CuO_{4+\delta}$ single crystal with $T_c = 6$ K (sample *3* in Fig. 3) in a magnetic field (up to 12 T) perpendicular to CuO_2 planes revealed a logarithmic increase in resistivity ρ_{ab} upon cooling (0.2 K ≤ T ≤ 5 K) as well as a clearly manifested negative magnetoresistance in fields $B > B_{c2}$ (B_{c2} = 3–4 T) [16]. These facts were interpreted by us as manifestations of the effects of weak 2D localization for the metal-type conduction ($k_F l$ = 25) along cuprate planes.

In the present work, we analyzed the magnetoresistance of a $Nd_{1.82}Ce_{0.18}CuO_{4+\delta}$ sample with a lower value of T_c = 4 K (sample *2* in Fig. 3) and a broad ($\Delta T \approx$ 10 K) superconducting transition, which is apparently associated with a larger disorder in the system ($k_F l$ = 2.5). We measured resistivity ρ_{ab} in a magnetic field up to 5 T perpendicular (**B** ⊥ *ab*) and parallel (**B** ∥ *ab*) to a CuO_2 plane in the temperature range 1.4 K ≤ T ≤ 30 K. Owing to a high extent of disorder, the logarithmic increase in ρ_{ab} is observed upon a decrease in temperature even in zero magnetic field. As a result, the $\rho_{ab}(T)$ dependence for $B = 0$ has the shape of a curve with a peak at $T \approx 14$ K (Figs. 4a and 4b).

It can be seen from Fig. 4a that the magnetic field parallel to CuO_2 planes gradually suppresses superconductivity and virtually does not affect the resistance in the normal phase at $T \ge 14$ K. Figure 4b demonstrates, on the one hand, a more rapid degradation of superconductivity by the transverse magnetic field, which is associated with anisotropy of the upper critical field in layered cuprate superconductors [26]. On the other hand, for **B** ⊥ *ab*, a stronger effect of the magnetic field on the segment describing the decrease in ρ_{ab} upon cooling in the normal phase is observed: as the value of B increases, the segment is displaced towards lower temperatures. Such a behavior is attributed to the suppression of the quantum correction to conductivity associated with weak localization [16].

It can be seen from Fig. 4b that the transverse magnetic field B = 2 T suppresses the weak localization effect almost completely, while the longitudinal field

virtually does not influence this effect up to $B = 5$ T. Such an anisotropy of magnetoresistance is typical just of 2D systems [27].

3. CONCLUSION

An analysis of the publications devoted to the study of electrical conductivity of copper-oxide HTSC and our results obtained for $Nd_{2-x}Ce_xCuO_{4+\delta}$ single crystals leads to the following conclusions concerning the effect of disorder on the transport in these systems. In layered HTSC crystals, spatial separation of charge carriers and charged impurities takes place: electrons (in Nd systems) or holes (in La systems) do not remain in the bound state at Ce^{4+} or Sr^{2+} ions, but go over to CuO_2 planes. Charge carriers in a "perfect" crystal are delocalized in the *ab* plane and localized in the *c* direction. The situation is similar to that observed in semiconducting superlattices or multilayer heterostructures with selective doping in barriers; i.e., an HTSC crystals may be regarded as a system of multiple quantum wells (CuO_2 layers) separated by doped layers of Nd or La oxides [28].

For a low concentration of charge carriers, the random potential of ionized impurities or nonstoichiometric defects leads to their strong localization and to the temperature dependence of conductivity typical of insulators. Upon an increase in the charge carrier concentration of charge carriers and the Fermi energy, a transition to the weak localization mode occurs in the vicinity of the universal value $\sigma_{2D} \approx e^2/h$ of 2D conductivity of a CuO_2 layer ($k_F l \approx 1$). In the range of weak localization, a superconducting transition may occur when the localization radius exceeds the characteristic size of a Cooper pair. This condition is usually observed even for small values of parameter $k_F l \geq 2$–3.

ACKNOWLEDGMENTS

This work was supported financially by the Russian Foundation for Basic Research (project no. 00-02-17427) and by the Ministry of Science (State Contract no. 107-1(00)-P).

REFERENCES

1. K. Sreedhar and P. Ganguly, Phys. Rev. B **41**, 371 (1990).
2. H. Takagi, S. Uchida, and Y. Tokura, Phys. Rev. Lett. **62**, 1197 (1989).
3. E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994).
4. G. M. Eliashberg, in *Physical Properties of High Temperature Superconductors*, Ed. by D. M. Ginsberg (World Scientific, Singapore, 1989; Mir, Moscow, 1990), p. 505.
5. P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 293 (1985).
6. B. Beschoten, S. Sadewasser, C. Güntherodt, and C. Quitmann, Phys. Rev. Lett. **77**, 1837 (1996).
7. D. Mandrus, L. Forro, C. Kendziora, and L. Mihaly, Phys. Rev. B **44**, 2418 (1991).
8. K. Schlenga, H. Bach, and K. Westerholt, Physica C (Amsterdam) **221**, 161 (1994).
9. K. Karpinska, A. Malinowski, M. Z. Cieplak, *et al.*, Phys. Rev. Lett. **77**, 3033 (1996).
10. Y. Fukuzumi, K. Mizuhashi, K. Takenaka, and S. Uchida, Phys. Rev. Lett. **76**, 684 (1996).
11. N. A. Fortune, K. Murata, M. Ishibashi, *et al.*, Phys. Rev. B **43**, 12930 (1991); W. Jiang, J. L. Peng, Z. Y. Li, and R. L. Greene, Phys. Rev. B **47**, 8151 (1993).
12. S. Tanda, S. Ohzeki, and T. Nakayama, Phys. Rev. Lett. **69**, 530 (1992).
13. X. Q. Xu, S. N. Mao, Wu Jiang, *et al.*, Phys. Rev. B **53**, 871 (1996).
14. T. Fujita, N. Kikugawa, M. Ito, *et al.*, in *Proceedings of VI Internatonal Conference M2S-HTSC 2P09-83, Houston, 2000.*
15. A. A. Ivanov, S. G. Galkin, A. V. Kuznetsov, *et al.*, Physica C (Amsterdam) **180**, 69 (1991).
16. G. I. Harus, A. N. Ignatenkov, A. I. Ponomarev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 93 (1999) [JETP Lett. **70**, 97 (1999)]; G. I. Harus, A. N. Ignatenkov, A. I. Ponomarev, *et al.*, Zh. Éksp. Teor. Fiz. **116**, 1723 (1999) [JETP **89**, 933 (1999)].
17. P. Fournier, P. Mohanty, E. Maiser, *et al.*, Phys. Rev. Lett. **81**, 4720 (1998).
18. L. Alff, S. Meyer, S. Kleefisch, *et al.*, Phys. Rev. Lett. **83**, 2644 (1999).
19. T. B. Charikova, A. I. Ponomarev, A. N. Ignatenkov, *et al.*, Fiz. Met. Metalloved. **91**, 46 (2001).
20. B. I. Shklovskiĭ and A. L. Efros, in *Electronic Properties of Doped Semiconductors* (Nauka, Moscow, 1979; Springer-Verlag, New York, 1984), Chap. 11.
21. A. L. Efros, Solid State Commun. **70**, 253 (1989).
22. N. M. Plakida, *High Temperature Superconductors* (Mezhdunarodnaya Programma Obrazovaniya, Moscow, 1996).
23. Y. Nakamura and S. Uchida, Phys. Rev. B **47**, 8369 (1993).
24. M. V. Sadovskiĭ, Sverkhprovodimost: Fiz., Khim., Tekh. **8**, 337 (1995).
25. M. P. A. Fisher, G. Grinstein, and S. M. Girvin, Phys. Rev. Lett. **64**, 587 (1990).
26. E. Z. Meĭlikhov, Sverkhprovodimost: Fiz., Khim., Tekh. **4**, 1437 (1991).
27. B. L. Al'tshuler and A. G. Aronov, Pis'ma Zh. Éksp. Teor. Fiz. **64**, 407 (1996) [JETP Lett. **64**, 444 (1996)].
28. V. V. Kapaev and Yu. V. Kopaev, in *Proceedings of the XXXI Conference on Low Temperature Physics, Moscow, 1998*; A. I. Ponomarev, A. N. Ignatenkov, L. D. Sabirzyanova, *et al.*, in *Proceedings of the International Conference on the Physics of Semiconductors, Jerusalem, 1998*, CD-ROM, Sec. 5, Subsec. A, No. 36.

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 1090–1094.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1257–1261.

SOLIDS
Electronic Properties

The Influence of the Chemical Potential Oscillations on the De Haas–Van Alphen Effect in Quasi-Two-Dimensional Compounds¶

P. Grigoriev

Landau Institute for Theoretical Physics, Chernogolovka, Moscow oblast, 142432 Russia
Grenoble High Magnetic Field Laboratory MPI-FKF and CNRS, BP 166, F-38042 Grenoble Cedex 09, France
e-mail: pashag@itp.ac.ru

Received January 16, 2001

Abstract—The de Haas–van Alphen effect in quasi-two-dimensional metals is studied at arbitrary parameters. Oscillations of the chemical potential can substantially change the temperature dependence of harmonic amplitudes that is commonly used to determine the effective electron mass. The processing of the experimental data using the standard Lifshitz–Kosevich formula can therefore lead to substantial errors even in the strong harmonic damping limit. This can explain the difference between the effective electron masses determined from the de Haas–van Alphen effect and the cyclotron resonance measurements. The oscillations of the chemical potential and the deviations from the Lifshitz–Kosevich formula depend on the reservoir density of states that exists in organic metals due to open sheets of the Fermi surface. This dependence can be used to determine the density of electron states on open sheets of the Fermi surface. We present analytical results of the calculations of harmonic amplitudes in some limiting cases that show the importance of the chemical potential oscillations. We also describe a simple algorithm for a numerical calculation of the harmonic amplitudes for arbitrary reservoir density of states, arbitrary warping, spin-splitting, temperature, and Dingle temperature. © *2001 MAIK "Nauka/Interperiodica".*

The quantum magnetization oscillations (or the de Haas–van Alphen (dHvA) effect) were discovered long ago [1] and have been widely used as a powerful tool in studying the Fermi surfaces and single electron properties in metals [2]. In a 3D metal, a good quantitative description of this effect is given by the Lifshitz–Kosevich (LK) formula [3]. In two- or quasi-two-dimensional compounds, deviations from the LK formula are possible for three reasons: the harmonic damping in the 2D case is different, the impurity scattering cannot be described by the usual Dingle law, and the chemical potential also becomes an oscillating function of the magnetic field. The first problem is important only when the harmonic damping is weak and can be easily solved using the 2D harmonic expansion [2]. The second problem concerns the accurate calculation of the density of states (DOS) with electron–electron interactions and the impurity scattering. The electron–electron interactions are not very important if many Landau levels (LLs) are occupied (we consider the case where the Fermi energy ϵ_F is much greater than the LL separation and the temperature). In the 3D case, the impurity scattering adds an imaginary part $i\Gamma(B)$ to the electron spectrum, which means that the electron can leave its quantum state with the probability $w = \Gamma(B)/\pi\hbar$ per second. Assuming this energy level width $\Gamma(B)$ to be independent of the magnetic field B, one obtains the Dingle law of harmonic damping [4]

$$A_l \propto \exp(-2\pi l\Gamma/\hbar\omega_c),$$

where A_l is the amplitude of the harmonic number l and $\omega_c = eB/m^*c$ is the cyclotron frequency. This Dingle law has been proved by many experiments on 3D metals. In the 2D case, this law may be incorrect and the problem of the DOS distribution in 2D metals has not been solved yet, although many theoretical works have been devoted to this subject (for example, [5–7]). The problem is complicated because even the exact calculation of the pointlike impurity scattering is not sufficient because the long-range impurities (and probably, the electron–electron interactions) are also important in the 2D case [8]. The procedure of extracting the DOS distribution from the dHvA measurements was recently proposed in [9]. In the present paper, we focus on the third question: we assume the Dingle law to be valid and consider the influence of the chemical potential oscillations on the harmonic amplitudes of the dHvA oscillations in this approximation. Because we consider the quasi-2D case, the Dingle law is not a bad approximation. We show that the chemical potential oscillations substantially change the temperature and the Dingle temperature dependence of the harmonic amplitudes even in the limit of a strong harmonic damping. Therefore, the estimate of the effective electron mass based on the LK for-

¶ This article was submitted by the author in English.

mula can lead to an error of up to 30%. This can explain the difference between the effective electron masses obtained from the dHvA effect and from the cyclotron resonance measurements (for example, in [10] and [11]). This problem was examined numerically by Harrison *et al.* [12] at zero warping W of the Fermi surface (FS). In this paper, we derive explicit formulas describing the quantum magnetization oscillations at arbitrary parameters. We study the result analytically in some limiting cases. This shows the importance of the chemical potential oscillation effect on harmonic amplitudes.

The energy spectrum of the quasi-two-dimensional electron gas is given by

$$E_{n,k_z,\sigma} = \hbar\omega_c\left(n+\frac{1}{2}\right) + \frac{W}{2}\cos(k_z d) + \sigma\mu_e B, \qquad (1)$$

where W is the warping of a quasi-cylindrical FS. The DOS distribution with the impurity scattering can be written as

$$\rho(E,B) = \rho_0(E,B) + \tilde{\rho}(E,B),$$

where for $E \gg \hbar\omega_c$, the oscillating part of the DOS is [13]

$$\tilde{\rho}(E,B) = \frac{4g}{\hbar\omega_c}\sum_{l=1}^{\infty}(-1)^l\cos\left(2\pi l\frac{E}{\hbar\omega_c}\right) \times J_0\left(\pi l\frac{W}{\hbar\omega_c}\right)\cos\left(2\pi l\frac{\mu_e B}{\hbar\omega_c}\right)\exp\left(-\frac{2\pi l\Gamma}{\hbar\omega_c}\right). \qquad (2)$$

In this formula, $g = B/\Phi_0$ is the LL degeneracy, the factor $\cos(2\pi l\mu_e B/\hbar\omega_c)$ is due to the spin splitting, and the factor $J_0(\pi lW/\hbar\omega_c)$ comes from the finite warping W of the quasi-cylindrical FS. $J_0(x)$ is the zero-order Bessel function. The last factor in (2) is the usual Dingle factor.

The nonoscillating part of the DOS is given by

$$\rho_0(E,B) = \frac{2g}{\hbar\omega_c}(1 + n_R(E)),$$

where $n_R(E)$ is the ratio of the reservoir density of states to the average DOS on the quasi-2D part of the FS. The reservoir DOS occurs in quasi-2D organic metals because of open sheets of the FS. These quasi-one-dimensional states do not directly contribute to the magnetization oscillations because they form the continuous spectrum and the nonoscillating DOS.

If the DOS is known, one can calculate the thermodynamic potential

$$\Omega(\mu,B,T) = -T\int_0^{\infty}\rho(E,B)\ln\left[1+\exp\left(\frac{\mu-E}{T}\right)\right]dE = \Omega_0(\mu,B,T) + \tilde{\Omega}(\mu,B,T), \qquad (3)$$

where $\mu(B)$ is the chemical potential and the oscillating part of the thermodynamic potential is given by [13]

$$\tilde{\Omega} = 2gT\sum_{l=1}^{\infty}\frac{(-1)^l}{l}\cos\left(2\pi l\frac{\mu}{\hbar\omega_c}\right)\frac{\lambda l}{\sinh(\lambda l)} \times J_0\left(\pi l\frac{W}{\hbar\omega_c}\right)\cos\left(2\pi l\frac{\mu_e H}{\hbar\omega_c}\right)\exp\left(-\frac{2\pi l\Gamma}{\hbar\omega_c}\right),$$

where $\lambda \equiv 2\pi^2 T/\hbar\omega_c$. The total particle number is usually constant,

$$N = -\left(\frac{\partial\Omega(\mu,B,T)}{\partial\mu}\right)_{T,B} = \int_0^{\infty}\frac{\rho(E,B)}{1+\exp\left(\frac{E-\mu}{T}\right)}dE = \text{const}.$$

This is an equation for the chemical potential as a function of the magnetic field. Separating the oscillating part of the DOS and substituting

$$N = \int_0^{\infty}\frac{\rho_0(E,B)}{1+\exp\left(\frac{E-\varepsilon_F}{T}\right)}dE$$

(where ε_F is the Fermi energy at zero magnetic field), we obtain

$$\int_0^{\infty}\left(\frac{1}{1+\exp\left(\frac{E-\varepsilon_F}{T}\right)} - \frac{1}{1+\exp\left(\frac{E-\mu}{T}\right)}\right) \times \rho_0(E,B)dE = \int_0^{\infty}\frac{\tilde{\rho}(E,B)}{1+\exp\left(\frac{E-\mu}{T}\right)}dE. \qquad (4)$$

We next use the fact that the relative reservoir DOS $n_R(E)$ does not change appreciably at the scale of T or $|\mu - \varepsilon_F| < \hbar\omega_c/2$ (this is true if many LLs are occupied because $n_R(E)$ changes substantially at the Fermi energy scale). It then follows that $n_R(E) \approx n_R(\varepsilon_F) = \text{const} \equiv n_R$. The left-hand side of (4) can be simplified, and after the insertion of (2), we obtain the equation for the oscillating part $\tilde{\mu}(B)$ of the chemical potential,

$$\tilde{\mu}(B) \equiv \mu(B) - \varepsilon_F = \frac{\hbar\omega_c}{\pi(1+n_R(\varepsilon_F))} \times \sum_{l=1}^{\infty}\frac{(-1)^{l+1}}{l}\sin\left(\frac{2\pi l(\varepsilon_F + \tilde{\mu}(B))}{\hbar\omega_c}\right)\frac{\lambda l}{\sinh(\lambda l)} \times \cos\left(2\pi l\frac{\mu_e H}{\hbar\omega_c}\right)\exp\left(-\frac{2\pi l\Gamma}{\hbar\omega_c}\right)J_0\left(\pi l\frac{W}{\hbar\omega_c}\right). \qquad (5)$$

This nonlinear equation cannot be solved analytically without any approximations, but it determines oscillations of the chemical potential with arbitrary parameters (it is only assumed that $\epsilon_F \gg T, \hbar\omega_c$).

The magnetization oscillations at the constant electron density N = const are given by

$$M = -\frac{d(\Omega + N\mu)}{dB}\bigg|_{N = \text{const}} = -\frac{\partial\Omega}{\partial B}\bigg|_{\mu, N = \text{const}}$$

$$-\left(\frac{\partial\Omega}{\partial\mu}\bigg|_{N, B = \text{const}} + N\right)\frac{d\mu}{dB}\bigg|_{N = \text{const}} = -\frac{\partial\Omega}{\partial B}\bigg|_{\mu, N = \text{const}}.$$

The oscillating part of the magnetization is

$$\tilde{M}(B) = -\frac{\partial\tilde{\Omega}}{\partial B}\bigg|_{\mu, N = \text{const}}$$

$$= \frac{2g}{\pi B}\varepsilon_F\sum_{l=1}^{\infty}\frac{(-1)^{l+1}}{l}\frac{\lambda l}{\sinh\lambda l}$$

$$\times\cos\left(2\pi l\frac{\mu_e H}{\hbar\omega_c}\right)\exp\left(-\frac{2\pi l\Gamma}{\hbar\omega_c}\right) \qquad (6)$$

$$\times\left\{\sin\left(2\pi l\frac{\mu(B)}{\hbar\omega_c}\right)J_0\left(\pi l\frac{W}{\hbar\omega_c}\right)\right.$$

$$\left.+\frac{W}{2\mu}\cos\left(2\pi l\frac{\mu(B)}{\hbar\omega_c}\right)J_1\left(\pi l\frac{W}{\hbar\omega_c}\right)\right\},$$

where $\mu(B)$ is given by Eq. (5) and involves the dependence of the magnetization on the reservoir DOS. Equations (5) and (6) describe the magnetization oscillations at arbitrary parameters. The only approximation used in deriving these formulas is the Dingle law of harmonic damping. In quasi-2D organic metals with the warping $W > T_D$, the Dingle law is believed to be a sufficiently good approximation.

Equations (5) and (6) are a good starting point for numerical calculations. It follows that in the limit $W/\mu \ll 1$, the oscillating parts of the magnetization and the chemical potential are related simply by

$$\tilde{M}(B) = \frac{\varepsilon_F}{B}\frac{2g}{\hbar\omega_c}(1 + n_R)\tilde{\mu}(B). \qquad (7)$$

For zero warping, this was obtained in [9].

Nonlinear equation (5) for $\tilde{\mu}(B)$ can be solved analytically only in some simple approximations. We do this to illustrate the influence of the chemical potential oscillations on the temperature and the Dingle temperature dependence of the harmonic amplitudes. We thus consider zero warping, zero spin splitting, and zero temperature. The sum in the right-hand side of Eq. (5) can then be calculated and we obtain

$$\frac{x}{2} = \frac{1}{(1 + n_R)}\arctan\left(\frac{\sin(y + x)}{\cos(y + x) + e^b}\right), \qquad (8)$$

where $x \equiv 2\pi\tilde{\mu}(B)/\hbar\omega_c$, $y \equiv 2\pi\varepsilon_F/\hbar\omega_c$, and $b \equiv 2\pi\Gamma/\hbar\omega_c$.

For a very large electron reservoir $n_R = \infty$, we have $x = 0$, which implies the case of a fixed chemical potential. In this case, the magnetization is given by [13]

$$\tilde{M}(B) = \frac{2g\varepsilon_F}{\pi B}\arctan\left(\frac{\sin y}{e^b + \cos y}\right). \qquad (9)$$

The temperature dependence of the harmonic amplitudes is given by the LK formula

$$A_l(T) = \frac{2\pi^2 Tl/\hbar\omega_c}{\sinh(2\pi^2 Tl/\hbar\omega_c)}. \qquad (10)$$

It is also possible to solve Eq. (8) analytically at $n_R = 0$ and $n_R = 1$. At zero electron reservoir $n_R = 0$, the solution to this equation is

$$\frac{x}{2} = \pi\frac{\tilde{\mu}(\beta)}{\hbar\omega_c} = \arctan\left(\frac{\sin y}{e^b - \cos y}\right).$$

It describes oscillations of the chemical potential. The magnetization at zero electron reservoir is given by the formula

$$\tilde{M}(B) = \frac{2g\varepsilon_F}{\pi B}\arctan\left(\frac{\sin y}{e^b - \cos y}\right), \qquad (11)$$

which coincides with (9) after the phase shift $y \longrightarrow y + \pi$ and the sign change $\tilde{M} \longrightarrow -\tilde{M}$. This implies that the harmonic damping law

$$A_l \propto \frac{1}{l}\exp(-lb) \qquad (12)$$

does not change and only the sign of all even harmonics is reversed. This symmetry between the cases of the fixed chemical potential μ = const and the constant particle density N = const is a feature of the special exponential law of the harmonic damping. Any finite temperature and the electron reservoir density break this symmetry.

We now consider the intermediate case where $n_R = 1$. Equation (8) then becomes

$$\frac{\sin x}{\cos x} = \frac{\sin(y + x)}{\cos(y + x) + e^b}. \qquad (13)$$

This gives

$$x = \arcsin(e^{-b}\sin y).$$

For the magnetization, we obtain

$$\tilde{M}(y) = \frac{g\varepsilon_F}{\pi B}\arcsin(e^{-b}\sin y). \qquad (14)$$

To determine how the harmonic damping has changed, we must calculate the amplitudes of the first several harmonics of this expression. The amplitude of the first harmonic is

$$A_1(b) = \frac{1}{\pi}\int_{-\pi}^{\pi} \arcsin(e^{-b}\sin y)\sin y\,dy,$$

and after the integration by parts, we obtain

$$A_1(b) = \frac{4}{\pi}\int_0^{\pi/2} \frac{\cos^2 y e^{-b}\,dy}{\sqrt{1-e^{-2b}\sin^2 y}}.$$

This is a superposition of two elliptic integrals,

$$A_1(b) = \frac{4}{\pi}[e^b E(e^{-b}) - 2\sinh b K(e^{-b})]. \qquad (15)$$

For $b \gg 1$, the deviations of $A_1(b)$ from the LK formula are small,

$$A_1(b) = e^{-b} + e^{-3b}/8 + \ldots .$$

In the opposite limit $b \ll 1$, we obtain

$$A_1(b) = \frac{4}{\pi}\left\{1 - b\left(\ln\frac{4}{\sqrt{2b}} - \frac{1}{2}\right) + O(b^2)\right\}. \qquad (16)$$

This is substantially different from the LK dependence $A_1(b) = \exp(-b) \approx 1 - b$. For example, the value $A_1(0)$ is $4/\pi$ times larger than the LK prediction.

A stronger deviation from LK formula (12) can be seen in the amplitudes of the next harmonics. All even harmonics vanish because expression (14) possesses the symmetries $\tilde{M}(\pi - y) = \tilde{M}(y)$ and $\tilde{M}(-y) = \tilde{M}(y)$.

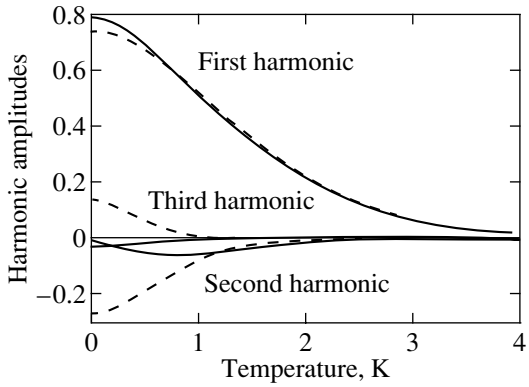
The amplitude of the third harmonic can also be calculated. For $b \gg 1$ and $e^{-b} \ll 1$, we have

$$A_3(b) = -e^{-3b}/12 + O(e^{-5b}).$$

This result is in contrast to the cases where $n_R = 0$ or $n_R = \infty$, where we had $A_3(b) = e^{-3b}/3$. This is not surprising because in the symmetric case $n_R = 1$, the oscillations must be much smoother and more sinusoidal. Therefore, the first harmonic must increase and the higher harmonics must decrease. For $b = 0$, we have

$$A_3(0) = \frac{4}{3\pi}\int_0^{\pi/2} \frac{\cos 3y \cos y\,dy}{\cos y} = -\frac{4}{9\pi}, \qquad (17)$$

which is ~2.35 times less than the LK prediction $A_3(0) = 1/3$ and has the opposite sign. In the case where $n_R = 1$, the first harmonic is therefore increased while the amplitudes of the others are strongly decreased compared to the cases of zero and infinite electron reservoir. The deviation from the LK formula reduces as the warping of the FS increases. The above analysis also shows that at low temperature and low Dingle temperature, the harmonic ratios can give a quantitative estimate of the electron reservoir density that is much more precise than just an observation about the slope of magnetization oscillations.



Temperature dependence of harmonic amplitudes. The solid lines are the numerical results (for $n_R = 1$, $m^* = 2m_0$, $T_D = 0.2$ K, and $W = 1$ K; see text) and the dashed lines are the LK prediction at the same parameters. Their strong deviations are clearly seen, especially for higher harmonics.

To include the correct temperature dependence, warping, and spin-splitting and to consider an arbitrary reservoir density, one can perform numerical calculations based on solving Eq. (5) for the chemical potential and inserting this solution in formula (6) for the magnetization. This can be easily done for arbitrary parameters that are available experimentally. The temperature dependence of the first three harmonic amplitudes is given in the figure for the following set of parameters close to the real experiments on α-(BEDT-TTF)$_2$KHg(SCN)$_4$: the reservoir density $n_R = 1$, the dHvA frequency $F = 700$ T, the effective mass $m^* = 2m_0$, the Dingle temperature $T_D = 0.2$ K, and the warping $W = 1$ K. We see a substantial deviation from the LK dependence. As $T \longrightarrow 0$, the obtained amplitude of the first harmonic is about 1.1 times larger than the LK prediction. If we also let $T_D \longrightarrow 0$ and $W \longrightarrow 0$, their ratio becomes $4/\pi = 1.27$, in agreement with analytical result (16). The second harmonic amplitude is close to zero at $T = 0$. The amplitude of the third harmonic changes its sign at $T \approx 0.8$ K and deviates very strongly from the LK formula. It is damped much more strongly than the LK predictions. At $T = 0$ and $W = 0$, it also coincides with prediction (17).

To conclude, it was shown both analytically and numerically that the oscillations of the chemical potential are essential for the temperature dependence of harmonic amplitudes of dHvA oscillations in quasi-two-dimensional compounds. The accurate determination of the effective electron mass from the dHvA effect should take this effect into account. This can be done by a simple numerical calculation based on Eqs. (5) and (6). The oscillations of the chemical potential depend on the reservoir density of states in accordance with Eq. (5). This

fact may be used for estimating the reservoir density of states in organic metals.

ACKNOWLEDGMENTS

The author thanks A.M. Dyugaev and M.V. Kartsovnik for fruitful discussions. The work was supported by the Russian Foundation for Basic Research, project no. 00-02-17729a.

REFERENCES

1. W. J. de Haas and P. M. van Alphen, Commun. Phys. Lab. Univ. Leiden A **212**, 215 (1930); Proc. R. Acad. Sci. Amsterdam **33**, 1106 (1930).
2. D. Shoenberg, *Magnetic Oscillations in Metals* (Cambridge Univ. Press, Cambridge, 1984).
3. L. M. Lifshitz and A. M. Kosevich, Zh. Éksp. Teor. Fiz. **29**, 730 (1956) [Sov. Phys. JETP **2**, 636 (1956)].
4. R. B. Dingle, Proc. R. Soc. London, Ser. A **211**, 517 (1952).
5. T. Ando and Y. Uemura, J. Phys. Soc. Jpn. **36**, 959 (1974).
6. K. B. Efetov and V. G. Marikhin, Phys. Rev. B **40**, 12126 (1989).
7. T. Maniv and I. Vagner, Phys. Rev. B **38**, 6301 (1988).
8. X. C. Xie, Q. P. Li, and S. Das Sarma, Phys. Rev. B **42**, 7132 (1990).
9. P. Grigoriev and I. Vagner, E-print archives, cond-mat/0009409 (2000).
10. J. Wosnitza, G. W. Crabtree, H. H. Wang, *et al.*, Phys. Rev. Lett. **67**, 263 (1991).
11. S. Hill, Phys. Rev. B **55**, 4931 (1997).
12. N. Harrison, R. Bogaerts, J. Singleton, *et al.*, Phys. Rev. B **54**, 9977 (1996).
13. T. Champel and V. P. Mineev, E-print archives, cond-mat/0006156 (2000).

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 887–894.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1029–1037.

GRAVITATION, ASTROPHYSICS

Correlations between the Arrival Directions of Ultrahigh Energy Cosmic Rays and the Large-Scale Structure of the Universe

A. V. Glushkov* and M. I. Pravdin

Institute of Cosmophysical Investigations and Aeronomy, Siberian Division, Russian Academy of Sciences, Yakutsk, 677891 Russia

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

Received November 23, 2000

Abstract—We present the results of analyzing the arrival directions of cosmic rays with energies $E_0 \geq 4 \times 10^{17}$ eV and zenith angles $\theta \leq 45°$ detected at the Yakutsk setup during 1974–2000. It is shown that increased particle fluxes exceeding the anticipated random distribution levels by (4–5)σ arrive from the galactic plane at $E_0 \approx (2–4) \times 10^{18}$ eV and from the supergalactic plane at $E_0 \geq 8 \times 10^{18}$ eV. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Revealing the sources of ultrahigh energy cosmic rays ($E_0 \geq 10^{17}$ eV) presents a difficult problem in astrophysical research. Investigations in this direction have been conducted for more than four decades all around the world, but the origin of these cosmic rays is still unclear. On the global scale, the cosmic rays do not contradict the concept of isotropic distribution, but in the region of extremely high energies ($E_0 \geq 10^{19}$ eV) there is a weak correlation between the arrival directions and both the galactic (see, e.g., [1, 2]) and the supergalactic planes [3, 4]. It was reported [5, 6] that clusters were observed in the directions of arrival of the primary particles with energies $E_0 \geq 10^{19}$ eV. Moreover, a correlation was found between individual clusters and pulsars situated on the entrance side of the local arm of the Galaxy [5]. There are some indications that cosmic rays of extremely high energies may originate from galaxies with active nuclei [7, 8]. Previously [9], it was demonstrated that particles with energies $E_0 \geq 10^{19}$ eV arriving from the equatorial region of a supergalaxy are characterized by a higher flux (as compared to other directions) and by a time-dependent intensity. Data presented below pour additional light on the problem of the origin of the ultrahigh energy cosmic rays.

2. COSMIC RAY CHARACTERISTICS STUDIED, RESULTS AND DISCUSSION

Below we report on extensive air showers (EASs) with energies $E_0 \geq 4 \times 10^{17}$ eV and zenith angles $\theta \leq 45°$ detected at the Yakutsk setup during 1974–2000. We studied correlations between the cosmic ray arrival directions and the galactic and supergalactic planes. The analysis involved the data for EASs with the directions of arrival determined by not less than five detectors and the axes falling within the central circle with a radius of ≤1700 m. These events provide for a minimum error in determining the main EAS parameters (arrival direction, axis coordinates, E_0, etc.). The primary particle energies were determined using the relationships

$$E_0\ [\mathrm{eV}] = (4.8 \pm 1.6) \times 10^{17} (\rho_{s,\,600}(0°))^{1.0 \pm 0.02}, \quad (1)$$

$$\rho_{s,\,600}(0°)\ [\mathrm{m}^{-2}] = \rho_{s,\,600}(\theta) \exp\left[(\sec\theta - 1)\frac{1020}{\lambda_\rho}\right], \quad (2)$$

$$\lambda_\rho\ [\mathrm{g/cm}^2] = (450 \pm 44) + (32 \pm 15) \log(\rho_{s,\,600}(0°)), \quad (3)$$

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

Figure 1a shows the fraction $r = N(|b_{G(SG)}| \leq 10°)/N_{\mathrm{all}}$ (N_{all} is the total particle number) of the primary particles arriving from equatorial regions ($|b_{G(SG)}| \leq 10°$) of the Galaxy (*G*) and supergalaxy (*SG*) plotted versus the primary energy E_0. Solid and dashed lines indicate the values for an isotropic flux calculated by the Monte Carlo method. The north pole of the supergalaxy has the galactic coordinates $b_G = 6.32°$ and $l_G = 47.37°$ [10]. Figure 1b presents an energy spectrum of the cosmic rays measured at the Yakutsk setup [11].

As is seen in Fig. 1, there is a tendency toward systematic increase in r_{SG} with the energy for $E_0 \geq 5 \times 10^{18}$ eV, although the statistical accuracy is insufficiently high. At first glance, the Galaxy exhibits no excess flux on the disk side. However, there is a slight increase in r_G for $E_0 \leq 4 \times 10^{18}$ eV which (as will be

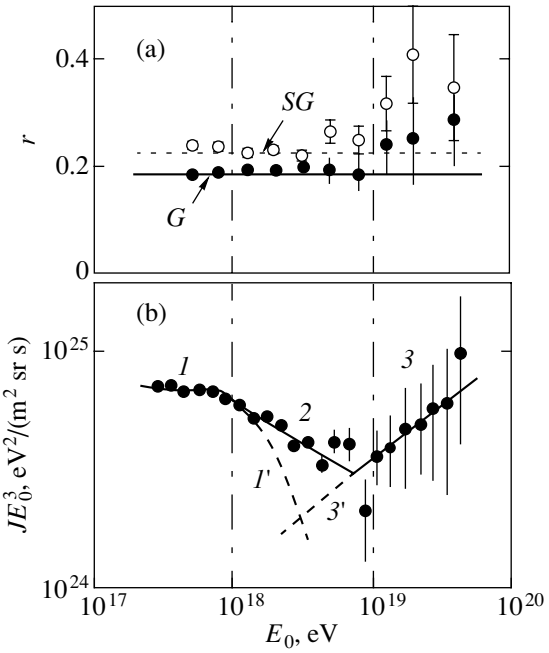


Fig. 1. (a) The fraction r (relative to the total number) of the primary particles arriving from equatorial regions ($|b| \le 10°$) of the Galaxy (*G*) and supergalaxy (*SG*) plotted versus the primary particle energy E_0. Solid and dashed lines indicate the values for an isotropic flux calculated by the Monte Carlo method. (b) An energy spectrum of the cosmic rays $J(E) \propto E^{-\gamma}$ measured on the Yakutsk setup [11]. Solid curves show the approximations according to the power law: (*1*) $E_0 < 10^{18.0}$ eV ($\gamma_1 = -3.05 \pm 0.04$); (*2*) $10^{18} \le E_0 < 10^{19.0}$ eV ($\gamma_2 = -3.34 \pm 0.05$); (*3*) $E_0 \ge 10^{19.0}$ eV ($\gamma_3 = -2.53 \pm 0.25$); (*1'*, *3'*) proposed galactic and supergalactic components, respectively.

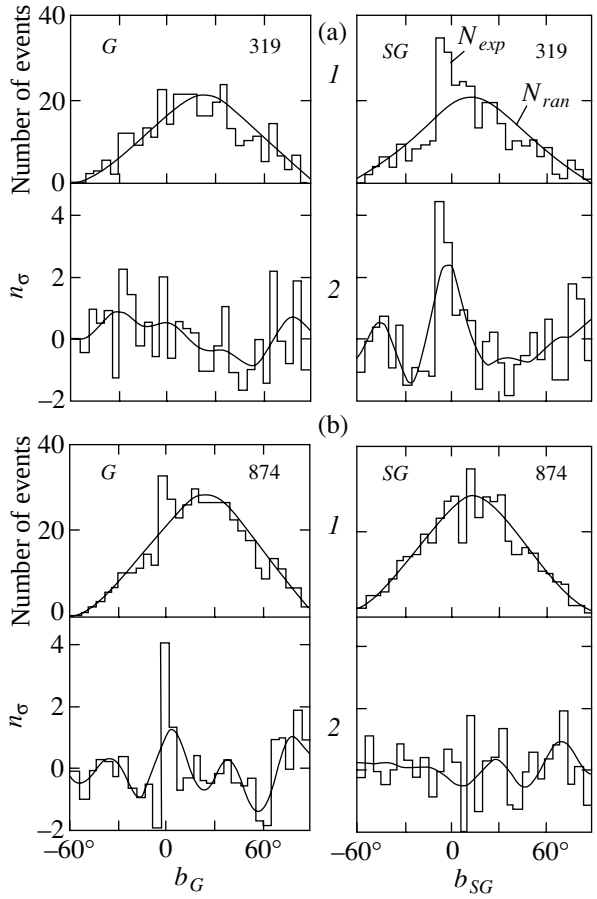


Fig. 2. Data for EAS with (a) $E_0 \ge 8 \times 10^{18}$ eV and (b) $E_0 = (3–4) \times 10^{18}$ eV: (*1*) plots of arrival directions versus latitude in the galactic (*G*) and supergalactic (*SG*) coordinates showing experimental data (N_{exp} histograms), figures indicate the total number of showers) and anticipated random distributions (N_{ran} curves) for isotropic fluxes; (*2*) deviations $n_\sigma = (N_{exp} - N_{ran})/\sqrt{N_{ran}}$ plotted as histograms and averaged curves.

shown below) is nevertheless indicative of a certain role of the Galaxy in the origin of particles with these energies.

Let us consider this situation in more detail. Figure 2a shows the distributions of arrival directions with respect to latitude (at a step of $\Delta b = 5°$) for 319 EASs with $E_0 \ge 8 \times 10^{18}$ eV in the galactic and supergalactic coordinates. The top curves (*1*) show the experimentally observed (N_{exp}) and anticipated random (N_{ran}) distributions; the bottom curves (*2*) present deviations of the number of observed events from that expected, plotted in units of the standard $\sigma = \sqrt{N_{ran}}$, $n_\sigma = (N_{exp} - N_{ran})/\sigma$. Curves *2* in Fig. 2a correspond to the average behavior of n_σ upon smoothening using a Fourier series with five harmonics.

The values of N_{ran} were determined by playing the number of showers randomly distributed over the celestial sphere. The procedure was as follows. For each measured EAS, the real arrival time and azimuth (determined in a horizontal coordinate system) were replaced at random to determine 500 directions in the galactic and supergalactic coordinates. The resulting distributions of random events were normalized to real with respect to the absolute value.

As is seen, the supergalactic plane exhibits a significant excess of events. In the latitude band $\Delta b_{SG} = -10–0°$, there are 65 EASs instead of expected 35, which corresponds to a relative excess of $(65 - 35)/\sqrt{35} \approx 5\sigma$. The galactic plane exhibits no effects in this energy range, except for a weak ($\approx 2\sigma$) positive deviation in the latitude band $\Delta b_G = -5–0°$.

Figure 2b shows distributions of the arrival directions for 874 EASs with $E_0 = (3–4) \times 10^{18}$ eV. These curves exhibit a relative maximum of $r_G$in comparison with the anticipated isotropic flux density (Fig. 1a). The events in the galactic plane ($|b_G| \le 5°$) exhibit a significant peak exceeding the expectation by $(121–87)/\sqrt{87} \approx 3.6\sigma$. The supergalactic plane is not manifested in this energy range.

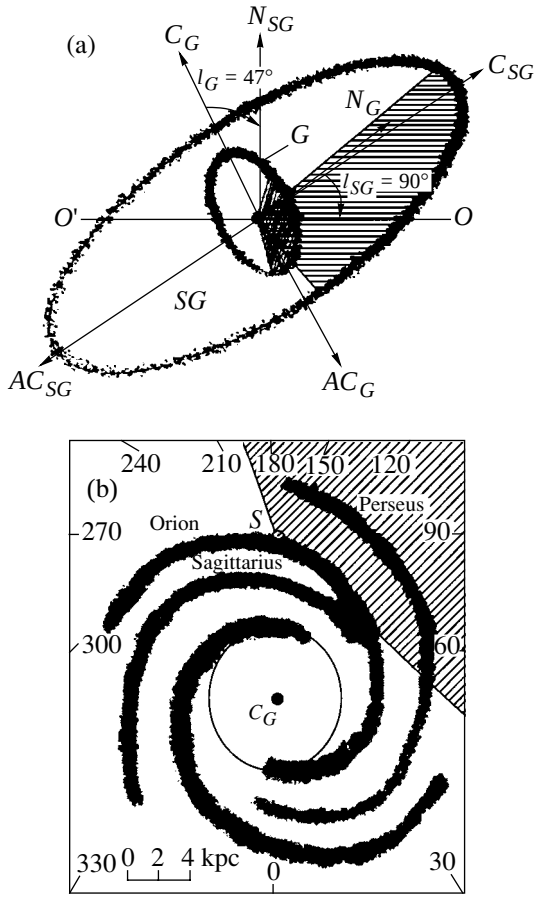


Fig. 3. (a) A schematic diagram showing the mutual arrangement of the galactic (*G*) and supergalactic (*SG*) Milky Ways: $C_G(C_{SG})$, centers AC_G (AC_{SG}), anticenters; $N_G(N_{SG})$, north poles; OO', line of intersection of the galactic and supergalactic planes; S, point of observation; dashed areas indicate visible sectors for Yakutsk EAS setup; (b) a schematic diagram of the galactic spiral [12].

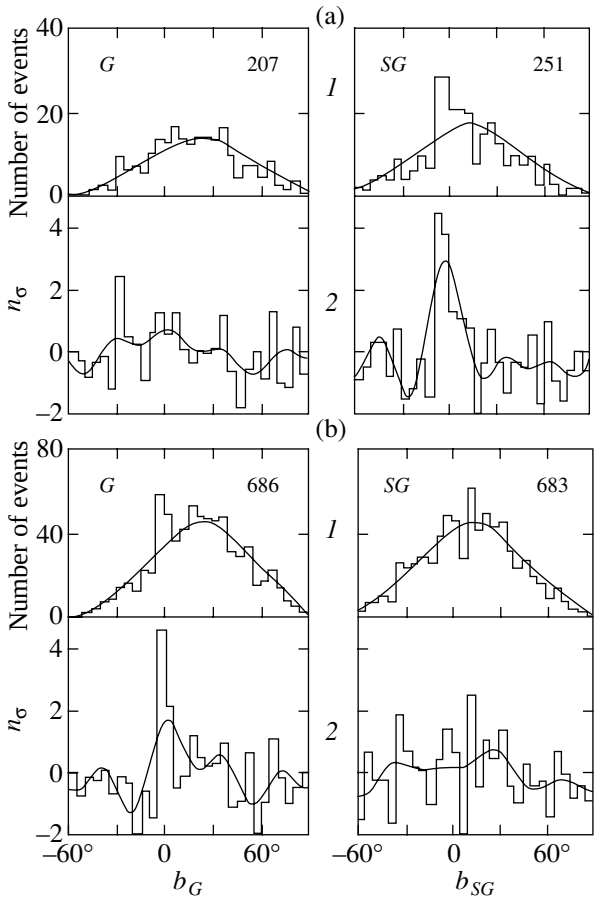


Fig. 4. A diagram of the arrival directions (the same as in Fig. 2, but without events in the latitude band $|b_{SG(G)}| \le 10°$).

The above data reveal the following pattern. In the energy range $E_0 \le (6–8) \times 10^{18}$ eV, the primary particles are apparently mostly generated in the Galaxy. An increase in their fraction in the disk (r_G) can be interpreted as due to a decrease in the degree of directed particle motion "smearing" caused by the galactic magnetic field. A sharp change in shape of the energy spectrum (Fig. 1b) and a strong correlation of the arrival direction with the supergalactic plane for $E_0 \ge (0.8–1) \times 10^{19}$ eV are indicative of the predominantly extragalactic origin of these particles.

The galactic plane is almost perpendicular to the supergalactic plane (Fig. 3a). We may use this circumstance to refine some details in Fig. 2. Figure 4 shows the distributions without events for $|b_{SG(G)}| \le 10°$ in the supergalactic (galactic) plane. As is seen, the distributions in Figs. 2 and 4 are similar, but the latter curves exhibit a stronger correlation in the cases indicated above. The supergalactic latitude band $\Delta b_{SG} = –10–0°$ exhibits 56 EASs (expected, 28) with $E_0 \ge 8 \times 10^{18}$ eV, which corresponds to an excess of $(56 – 28)/\sqrt{28} \approx 5.3\sigma$, while the galactic plane ($|b_G| \le 5°$) shows a peak of 105 EASs for $E_0 = (3–4) \times 10^{18}$ eV with an excess of $(105–68)/\sqrt{68} \approx 4.5\sigma$. Figure 5 presents a diagram of the arrival directions for EASs with $E_0 = (3–4) \times 10^{18}$ eV plotted on the celestial sphere in the galactic (*G*) and supergalactic (*SG*) coordinates.

This increase in correlations may, at first glance, appear rather unexpected—especially for the galactic EASs with $E_0 = (3–4) \times 10^{18}$ eV. What can be the influence of the supergalaxy, which has a markedly greater volume and contains the Galaxy as a part? This is only possible provided that particles in this range also originate from outside the Galaxy.

In this context, it was especially of interest to study the distributions of events in the latitude band with $|b_{SG(G)}| \le 10°$, which are depicted in Figs. 6 and 7. As is seen, all distributions in Fig. 6 are different from those considered above. The first feature to be noted is that the observed patterns are strongly different from those

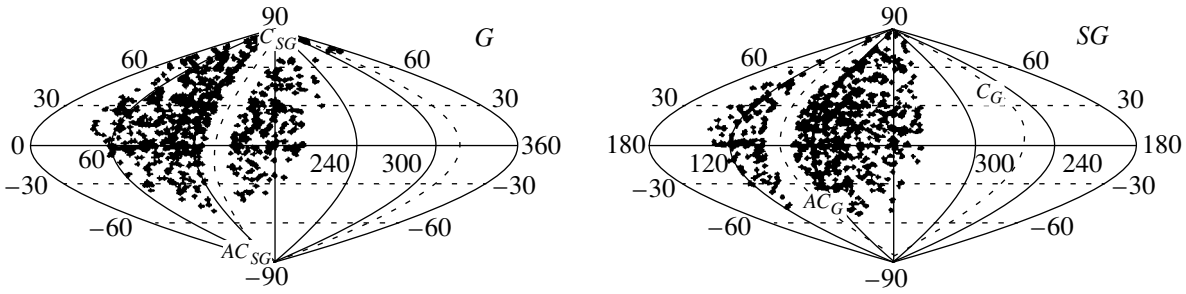


Fig. 5. A diagram of arrival directions for 686 EAS with $E_0 = (3–4) \times 10^{18}$ eV plotted on the celestial sphere in the galactic (*G*) and supergalactic (*SG*) coordinates without events in the latitude band $|b_{SG(G)}| \le 10°$.

anticipated for the random events. For example, the data in Fig. 6a(*G*) give $\chi^2 = 50$ for $n = 30$ degrees of freedom (random probability, $P \approx 10^{-2}$) and in Fig. 6a (*SG*), $\chi^2 = 55$ ($P \approx 3.5 \times 10^{-3}$). According to Fig. 6b we observe for the Galaxy (*G*) $\chi^2 = 77$ ($P < 10^{-5}$) and for the supergalaxy (*SG*) $\chi^2 = 85$ ($P < 10^{-5}$). Secondly, the distributions are not similar to those presented in Fig. 2. This result may seem surprising because the data depicted in Fig. 6 present only a part of the total set depicted in Fig. 2.

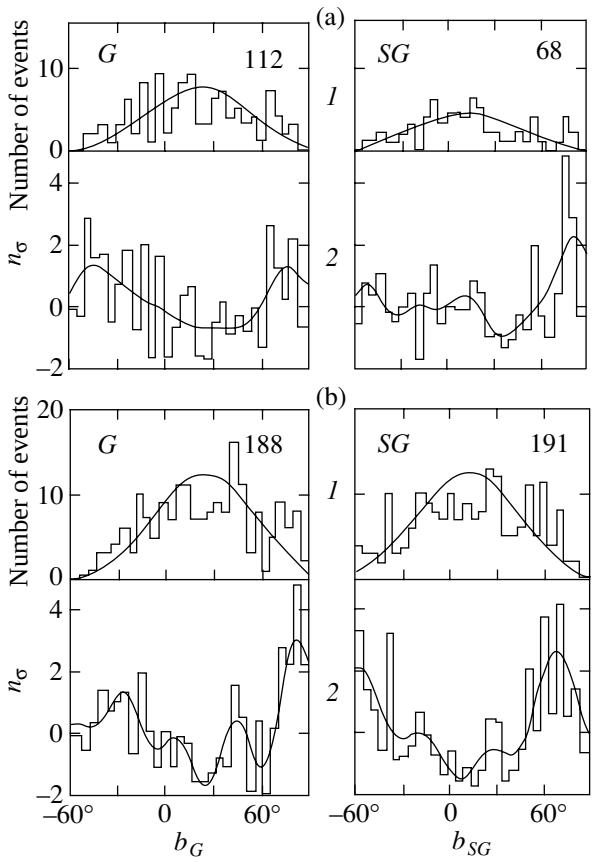


Fig. 6. Distributions of events in the latitude band $|b_{SG(G)}| \le 10°$ (for notations see Fig. 2).

Let us consider the distribution of arrival directions presented in Fig. 7. Since the planes of the supergalaxy and Galaxy are mutually almost perpendicular, the pattern of events in the $|b_{SG}| \le 10°$ band with different b_G (Fig. 6(*G*)) presents in fact a distribution along the supergalactic disk with respect to the latitude l_{SG} and vice versa, the pattern of events in the $|b_G| \le 10°$ band with different b_{SG} (Fig. 6(*SG*)) presents essentially a distribution along the galactic disk with respect to the longitude l_G. The l_{SG} value is counted clockwise from the direction to the supergalaxy center (Fig. 3a).

Figure 8 shows the distributions of arrival directions for EASs from the supergalactic and galactic disks ($|b| \le 10°$) plotted versus the longitude coordinate. The data in Fig. 8a(*SG*) can be interpreted as follows. In the supergalactic disk sector studied ($l_{SG} \approx 0–130°$), the relative intensity of cosmic rays with $E_0 \ge 5 \times 10^{18}$ eV varies in a smooth manner (spline curve *2* in Fig. 8a). An increase in the flux at $l_{SG} \approx 90°$, where the supergalactic and galactic planes intersect, is most probably due to an additional contribution due to the particles arriving from the galactic disk. The peak observed in Fig. 8a (*G*) at $l_G \approx 137°$ also corresponds to the intersection of these planes.

Now let us proceed to the analysis of data in Fig. 8b showing the arrival directions for EASs with $E_0 = (2.5–4) \times 10^{18}$ eV. The first peak in the galactic disk (see the n_τ distribution) at $l_G \approx 70°$ is apparently due to an increased flux of particles originating from the exit of the galactic local arm. This is well illustrated in Fig. 3b, schematically depicting a spiral structure of the Galaxy [12], where a galactic disk sector observed at the Yakutsk setup is shaded. The fourth peak at $l_G \approx 180°$ is related to a high intensity of emission from the anticenter, where the particle absorption in the disk is at minimum. This behavior is also quite possible if the particles arrive from outside the Galaxy. As for the peaks at $l_G \approx 110$ and 155°, these features can be attributed to the other arms of the Galaxy. The magnetic fields of arms in the galactic disk are schematically depicted in Fig. 9e [13], where open and dark circles indicate the field orientation outward and inward rela-

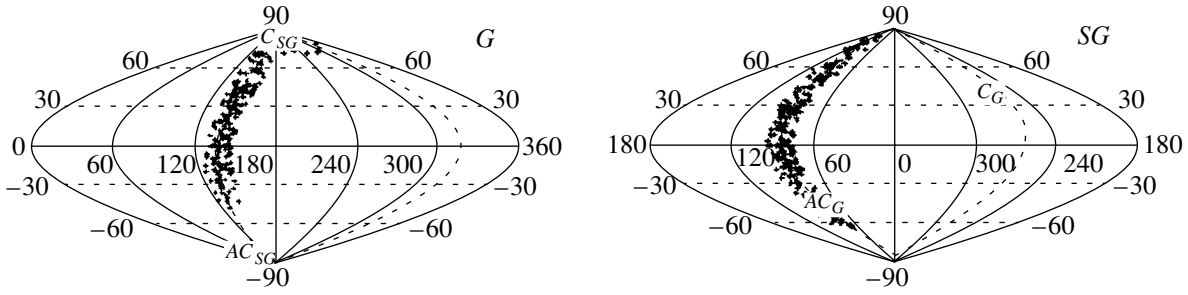


Fig. 7. A diagram of arrival directions for 194 EAS with $E_0 = (3–4) \times 10^{18}$ eV plotted on the celestial sphere in the galactic (*G*) and supergalactic (*SG*) coordinates for events in the latitude band $|b_{SG(G)}| \le 10°$.

tive the figure plane and the field strength is proportional to the circle area.

Figure 9 shows distributions of the *n* values with respect to the longitude l_G for EASs arriving with various energies from the equatorial region ($|b_G| \le 10°$) of the Galaxy. These data refer to events with the zenith angles $\theta \le 60°$ and the shower axes falling inside the entire perimeter of the Yakutsk setup. Smooth curves show the spline functions constructed using a Fourier series with 13 harmonics. Note a special sequence of changes in the histogram fragments. The aforementioned excess of particles arriving from the anticenter with $E_0 = (2.5–4) \times 10^{18}$ eV is also observed at higher energies (Fig. 9a). This flux decreases at $E_0 < 2.5 \times 10^{18}$ eV (Fig. 9c) and completely vanishes in the region of $E_0 \approx (1–1.5) \times 10^{18}$ eV (Fig. 9d). In contrast, the

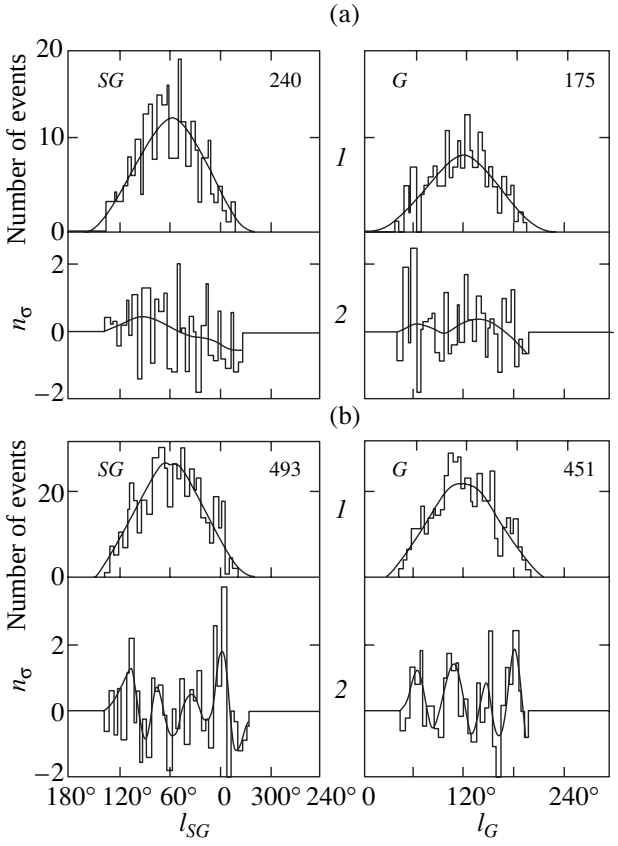


Fig. 8. Distributions of EAS with (a) $E_0 \ge 5 \times 10^{18}$ eV and (b) $E_0 = (2.5–4) \times 10^{18}$ eV versus longitude $l_{SG(G)}$ for cosmic rays arriving from the galactic (*G*) and supergalactic (*SG*) disks ($|b| \le 10°$); (for notations, see Fig. 2).

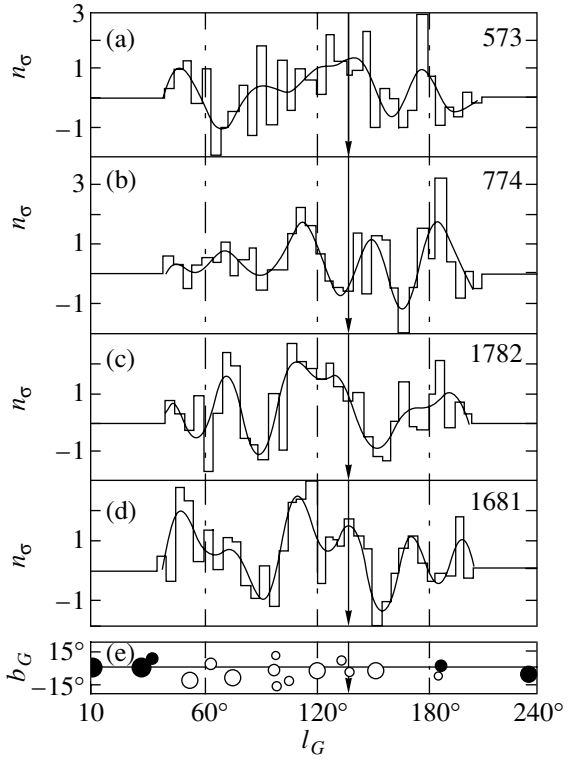


Fig. 9. Distributions of EAS versus longitude l_G for cosmic rays arriving from the galactic disk ($|b| \le 10°$) with various energies: (a) $E_0 \ge 10^{18.6}$ eV; (b) $E_0 = 10^{18.4–18.6}$ eV; (c) $E_0 = 10^{18.2–18.4}$ eV; (d) $E_0 = 10^{18.0–18.2}$ eV. (e) Diagram of arrangement of the galactic magnetic arms [13]: (○) field directed outward; (●) field directed inward; arrows indicate the line of intersection of the galactic and supergalactic planes ($l_G \approx 137°$); figures indicate the total numbers of events.

direction of $l_G \approx 75°$ exhibits a minimum at $E_0 \geq 5 \times 10^{18}$ eV, which changes to maximum when the primary particle energy decreases to $E_0 \approx 2 \times 10^{18}$ eV (Fig. 9c). Pronounced peaks are observed in the sector $\Delta l_G \approx 105$–$145°$ for the events with $E_0 < 4 \times 10^{18}$ eV (Figs. 9b–9d). For $E_0 = (2.5–4) \times 10^{18}$ eV, the excess in the number of observed events (690) over the random distribution (581) in this sector amounts to $(690–581)/\sqrt{581} \approx 4.6\sigma$.

The positions of these peaks correlate with the arrangement of galactic magnetic arms (Fig. 9e). Therefore, it is not excluded that these peaks, as well as the dynamics of their variation for $E_0 < 4 \times 10^{18}$ eV, are related to the activity of these arms. The role of individual galactic arms in the origin of cosmic rays in various energy intervals is probably different.

In the range of energies $E_0 \geq 4 \times 10^{18}$ eV, the contribution of the Galaxy is apparently not as large. The minimum observed in Fig. 9d for $l_G \approx 75°$ is most probably due to a relatively strong absorption of extragalactic particles by the Orion arm (see Fig. 3b) as compared to adjacent parts of the galactic disk. Note also a minimum at $l_G \approx 137°$ in Fig. 9b, which corresponds to the intersection of galactic and supergalactic planes (line *SO* in Fig. 3a). We may suggest that this minimum is related to an increase in the flux of particles from equatorial regions of the supergalaxy, because these particles are more strongly absorbed in the supergalactic disk.

The data presented in Fig. 8b(*SG)* (curves *2*) show that the number of events in the latitude band $|b_{SG}| \leq 10°$ also exhibits a minimum in the direction of intersection with the galactic disc at $l_{SG} \approx 90°$. This minimum can also be explained by the fact that particles arriving from outside the Galaxy exhibit a stronger absorption in the galactic disk then in the adjacent regions. In the direction toward the center of the supergalaxy ($l_{SG} \approx 0°$), we observe a clearly pronounced peak with the observed frequency exceeding the random value by $(46–25)/\sqrt{25} \approx 4\sigma$ in the angular sector $\Delta l_{SG} = 15°$.

Based on the fact that curves *2* in Fig. 8b(*SG*) certainly reveal a supergalactic structure, we must admit that particles with the energies $E_0 = (2.5–4) \times 10^{18}$ eV may arrive from extragalactic sources. In this case, the fraction of these particles may be comparable with the Galaxy contribution. Assuming that this fraction accounts for half of the events, the energy spectrum in Fig. 1b can be represented as a sum of the galactic (*1*') and extragalactic (*3*') components. The former spectrum (*1*') does not contradict the hypothesis [14] that the galactic primary particles may be predominantly neutrons with a spectrum of this very shape. This is indicated by an additional radiation observed [14] at the AGASA setup near the Galaxy center with an excess of 4.1σ over the anticipated isotropic flux. This result was confirmed and refined by the SUGAR group [15]. Unfortunately, the Yakutsk group cannot see the galactic center.

The aforementioned significant correlations between the arrival directions of cosmic rays and the galactic and supergalactic planes were revealed due to a very large statistics of EASs observed on the Yakutsk setup (a total of about 37 000 for $E_0 \geq 10^{18}$ eV) and a very rigid selection of events for the analysis. Unfortunately, no analogous data were reported for the AGASA setup, which is similar to ours with respect to the type of detectors and the method of EAS data processing. Recent report [6] only pointed out that two clusters were observed at $E_0 \geq 4 \times 10^{19}$ eV in the direction close to the supergalactic plane.

At first glance, the results obtained at the Yakutsk and AGASA setups are at variance as far as the possible role of the supergalaxy in the origin of extremely high energy cosmic rays is considered. However, in fact there are no discrepancies. This is confirmed by Fig. 10, showing the distribution of arrival directions for EASs with $E_0 \geq 10^{19}$ eV in the supergalactic coordinates, plotted at a $\Delta b_{SG} = 10°$ step using the data reported by both the Yakutsk and AGASA groups. We obtained a distribution analogous to that measured on the AGASA setup by adding three distributions reported in [6]. The histograms *2* show the values of n_σ representing deviations of the number of observed events from that for isotropic fluxes (analogous to the histograms in Fig. 2b). The number of analyzed events is the same for both setups. In our set, we employed the EASs with $\theta \leq 60°$ detected over the entire circle area (even for the shower axes falling outside the setup perimeter).

As is seen, despite a difference in the initial distributions *1* in Fig. 10, the distributions *2* are similar to one another. Note certain important features. First, both distributions *2* show a $(1.8–2.1)\sigma$ excess in the number of measured events over isotropic fluxes in a latitude band of $|b_{SG}| \leq 10°$. Second, there are dips indicative of -1.5σ deficiency in the events occurring symmetrically relative to the supergalactic plane at $|b_{SG}| \approx 10–20°$. Observed against these minima, the peaks of excess radiation originating from the supergalactic plane become more significant. Some other details in the two distributions *2* are also coinciding to a certain extent, thus showing evidence that these features are not accidental.

A comparison of our data presented in Figs. 2 and 10 shows the inexpediency of using the $|b_{SG}| \approx 10°$ step because this decreases significance of the results. The real experimental accuracy, especially that ensured by a rigid selection of events for the analysis, is not worse than 1–2°.

The fact that the arrival directions of the primary particles with energies $E_0 \geq 8 \times 10^{18}$ eV are correlated with the supergalactic disk orientation suggests that the particles of extragalactic origin must be electrically

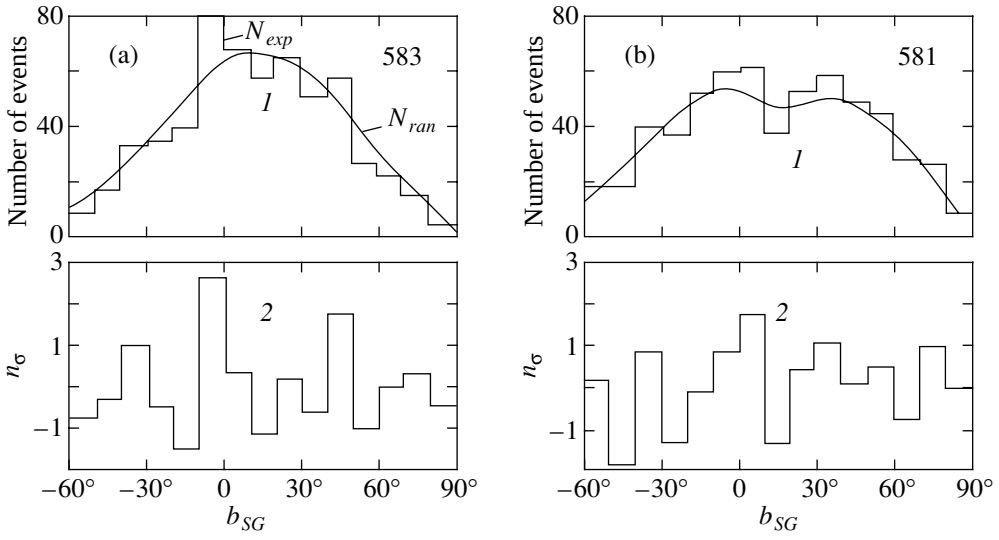


Fig. 10. Data of the (a) Yakutsk and (b) AGASA [6] groups on the distribution of showers with $E_0 \geq 10^{19}$ eV: (*1*) arrival directions versus latitude b_{SG} in the supergalactic coordinates; (*2*) deviations $n_\sigma = (N_{\exp} - N_{\mathrm{ran}})/\sqrt{N_{\mathrm{ran}}}$ (for notations, see Fig. 2).

neutral. Otherwise, particles bearing the electric charge z would move in the magnetic field by a trajectory with a curvature radius (depending on the energy E_0)

$$R = R_0/300Hz. \quad (4)$$

According to this, even the trajectories of protons moving in the galactic magnetic field ($H \approx 3 \times 10^{-6}$ Gs) with energies on the order of 10^{19} eV would have a curvature radius of approximately 3 kpc (which is considerably smaller than the galactic disk radius of ~15 kpc). The intergalactic magnetic fields, albeit weak ($H \approx 9 \times 10^{-10}$ Gs), still give for these protons $R \approx 10$ Mpc, which is also significantly smaller than the supergalactic disk diameter of ~60 Mpc). Under these conditions, the cosmic rays would "forget" their origin and their arrival directions would exhibit no correlations with the Galaxy and supergalaxy structure, which is not the case in the experiment.

The primary particles can hardly be neutrons. Indeed, the neutrons with $E_0 \sim 10^{19}$ eV possess a Lorentz factor on the order of 10^{10} and can cover before decay only a distance on the order of 100 kpc, which is significantly smaller than supergalaxy size. We believe that the ultrahigh energy cosmic rays consist of some other neutral particles. This conclusion is based on a complex analysis of the spatial and temporal structure of EASs observed at the Yakutsk setup [16–18]. The results of these investigations showed that the experimental data for $E_0 \leq (1\text{–}3) \times 10^{18}$ eV agree with calculations according to the QGSJET model [19] assuming a composition of the primary particles varying from a mixture enriched at $E_0 \sim 10^{17}$ eV with heavy nuclei (z = 10–30 fraction accounting for 63 ± 7% [20]) to a mixture enriched with protons at $E_0 \sim 10^{18}$ eV. In the region of $E_0 \geq (3\text{–}5) \times 10^{18}$, the showers develop differently, significantly changing their transverse structure. These changes cannot be described within the framework of the QGSJET model (irrespective of the composition of primary particles, from protons to iron nuclei), thus requiring other concepts concerning the development of extended air showers in the regions of extremely high energies.

3. CONCLUSION

As is seen from the data presented above, cosmic rays with energies $E_0 \approx (3\text{–}5) \times 10^{18}$ eV exhibit correlations in their arrival directions with galactic and supergalactic structure. This behavior gives certain grounds to believe that the primary particles with these energies may be of both galactic and extragalactic origin. As for the energy range $E_0 \geq 8 \times 10^{18}$ eV, there are indications of the predominantly extragalactic origin of these cosmic rays.

The experimental facts presented above, together with the results reported in [16–18], show evidence in favor of a hypothesis that the extragalactic primary radiation may contain a neutral component. On their way to Earth, these neutral particles may "transilluminate" a large-scale structure of the Universe, probably covering a region of space markedly exceeding the volume of a supergalaxy.

ACKNOWLEDGMENTS

This study was made possible due to the Ministry of Science and Technology of the Russian Federation supporting the Yakutsk Complex EAS Setup, included

(Registration No. 01-30) into the "List of Unique Scientific Research and Experimental Setups of National Significance."

REFERENCES

1. J. Szabelsky, J. Wdowczyk, and A. W. Wolfendale, J. Phys. G **12**, 1433 (1986).
2. B. N. Afanasiev, M. N. Dyakonov, T. A. Egorov, *et al.*, in *Proceedings of the 24th International Conference on Cosmic Rays, Rome, 1995*, Vol. 2, p. 756.
3. T. Stanev, P. L. Bierman, J. Lloyd-Evans, *et al.*, Phys. Rev. Lett. **75**, 3056 (1995).
4. A. A. Ivanov, A. D. Krasil'nikov, S. I. Nikol'skiĭ, and M. I. Pravdin, Izv. Ross. Akad. Nauk, Ser. Fiz. **61**, 520 (1997).
5. A. A. Mikhaĭlov, Izv. Ross. Akad. Nauk, Ser. Fiz. **63**, 557 (1999).
6. M. Takeda, N. Hayashida, K. Honda, *et al.*, Astrophys. J. **522**, 225 (1999).
7. A. V. Glushkov, Pis'ma Zh. Éksp. Teor. Fiz. **48**, 513 (1988) [JETP Lett. **48**, 555 (1988)].
8. A. V. Uryson, Zh. Éksp. Teor. Fiz. **116**, 1121 (1999) [JETP **89**, 597 (1999)].
9. A. V. Glushkov and I. E. Sleptsov, Izv. Ross. Akad. Nauk, Ser. Fiz. **65**, 437 (2001).
10. A. Hewitt and G. Burbidge, Astrophys. J., Suppl. Ser. **63**, 246 (1987).
11. M. I. Pravdin, M. N. Dyakonov, A. V. Glushkov, *et al.*, in *Proceedings of the 26th International Cosmic Ray Conference, Salt Lake City, 1999*, Vol. 3, p. 292.
12. Y. M. Georgelin and Y. P. Georgelin, *Cambridge Atlas of Astronomy*, Ed. by J. Audouze and G. Israel (Cambridge Univ. Press, Cambridge, 1988), map, p. 308.
13. A. G. Lyne and F. Graham-Smith, *Pulsar Astronomy* (Cambridge Univ. Press, Cambridge, 1990).
14. N. Hayashida, M. Nagano, D. Nishikawa, *et al.*, Astropart. Phys. **10**, 303 (1999).
15. J. A. Bellido, R. W. Clay, B. R. Dawson, and M. Johnston-Hollitt, E-print archives, astro-ph/0009039 (2000).
16. A. V. Glushkov, V. B. Kosarev, I. T. Makarov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 361 (1998) [JETP Lett. **67**, 383 (1998)].
17. A. V. Glushkov, I. T. Makarov, M. I. Pravdin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 145 (2000) [JETP Lett. **71**, 97 (2000)].
18. A. V. Glushkov, M. I. Pravdin, I. E. Sleptsov, *et al.*, Yad. Fiz. **63**, 1557 (2000) [Phys. At. Nucl. **63**, 1477 (2000)].
19. N. N. Kalmikov, G. B. Khristiansen, S. S. Ostapenko, and A. I. Pavlov, in *Proceedings of the 24th International Cosmic Ray Conference, Rome, 1995*, Vol. 1, p. 123.
20. E. A. Vishnevskaya, N. N. Kalmykov, G. V. Kulikov, *et al.*, Yad. Fiz. **62**, 300 (1999) [Phys. At. Nucl. **62**, 265 (1999)].

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 895–904.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1038–1049.

GRAVITATION, ASTROPHYSICS

The Surfatron Acceleration of Cosmic Rays in the Galactic Plasma

G. N. Kichigin

Institute of Solar-Terrestrial Physics, Siberian Division, Russian Academy of Sciences, Irkutsk, 664033 Russia
e-mail: king@iszf.irk.ru

Received January 4, 2001

Abstract—The optimum conditions for a prolonged holding of charged particles resonantly trapped from the galactic plasma by nonlinear waves and for the acceleration of these particles to high energies by the surfatron mechanism are established. The density of particles trapped by the plasma waves of large amplitude and by the quasitransverse magnetosonic shock waves is estimated. Various reasons leading to possible breakage of the process of surfatron acceleration of cosmic rays in the Galaxy are considered. Within the framework of the surfatron acceleration mechanism, galactic cosmic rays originate predominantly from the interstellar plasma and their energy spectrum is formed in two stages. In the first stage, some of the galactic plasma particles are accelerated from thermal energies to 10^{15} eV/nucleon; in the second stage, the cosmic rays may continue gaining energy up to 10^{19} eV/nucleon and above. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

One of the main problems in any theory explaining the origin of cosmic rays is the mechanism of acceleration of charged particles entering into these rays. As was noted in [1, 2], the most studied acceleration processes can be reduced to mechanisms of two types: (i) acceleration caused by the electric field induced by a variable magnetic field and (ii) acceleration due to collisions with moving inhomogeneities of a magnetic field. However, these mechanisms may provide for an increase in the charged particle energy only up to ~10^{12} eV/nucleon [3, 4]. Explaining the presence of particles with greater energies in the spectrum of cosmic rays requires searching for new acceleration processes.

In recent years, actively discussed as a method of the collective acceleration of charged particles in a weakly magnetized collisionless plasma is the so-called surfatron mechanism (surfing) [2, 4–11], which can provide for a relatively high rate of energy gain by charged particles. According to this mechanism, the particles are trapped by a potential wave running in the plasma across a weak magnetic field and accelerated in the direction along the wave front. The moving positive potential jump can accelerate ions, while the negative jump accelerates electrons.

Below we will consider two types of wave potential perturbations most frequently encountered in a collisionless plasma: (i) a longitudinal plasma wave and (ii) a quasitransverse magnetosonic shock wave (MSSW). A periodic plasma wave containing both positive and negative potential jumps is capable of accelerating both ions and electrons. In contrast, an MSSW comprises a positive potential jump and accelerates only ions.

In the galactic plasma, the stationary plasma waves and MSSWs may appear as a result of anomalous phenomena such as supernova and nova explosions, vigorous processes in unstable stars, flares in quiet stars of the Sun type, and some other analogous pulsed processes. In addition, nonlinear plasma waves moving at a subluminal speed may form as a result of the electromagnetic energy conversion in hybrid resonance layers [4], due to nonlinear plasma oscillations generated by a high-energy branch of the cosmic ray spectrum (wake waves [12]), or upon a strong relativistic magnetic dipole radiation emission from pulsars [2, 4].

A large (in principle, unlimited) duration of the process of particle acceleration by surfing is provided by a particle–wave synchronism stably maintained as a result of the particle outrunning the wave being prevented by deviations along the front in a constant magnetic field [11]. The particle accelerator based on this principle is called Surfatron [7], and the corresponding mechanism is referred to as the surfatron acceleration or surfing. It must be noted that the mechanism of particle acceleration operative in the Surfatron was originally considered by Sagdeev [13] within the framework of an analysis of the ion motion in the MSSW front.

The surfatron acceleration mechanism under consideration takes place in a weakly magnetized plasma. The substance in the Galaxy frequently occurs in the plasma state both in stellar atmospheres and in the interstellar medium [2]. Typical values of the plasma parameters for a stellar atmosphere are offered by those of the upper chromosphere of the Sun (density, $n_0 \sim 10^9$ cm^{-3}; temperature, $T \sim 10$ eV; magnetic field strength, $B_0 \sim 1$ Oe) and the solar wind ($n_0 \sim 10$ cm^{-3}; $T \sim 10$ eV; average magnetic field strength, $B_0 \sim 10^{-4}$ Oe). In the Galactic disk

and its environment, the substance predominantly occurs in the state of so-called hot gas [2] representing a completely ionized plasma with the following typical parameters: $n_0 \approx 3 \times 10^{-3}$ cm^{-3}; $T \sim 100$ eV; magnetic field strength, $B_0 \approx 3 \times 10^{-6}$ Oe.

Estimates of the energy of cosmic rays, obtained in [4] for the charged particles accelerated by means of surfing in the nonlinear plasma waves in the galactic plasma, amount to $\mathscr{E} \sim 10^{17}$–10^{21} eV. This result suggests that the mechanism of surfatron acceleration is a promising approach to explaining the formation of the high-energy branch of the spectrum of cosmic rays. Despite this optimism, however, there are several questions concerning features of the surfatron acceleration of cosmic rays in the Galaxy. What is the source of cosmic rays involved in the surfing process? Whether or not the source power is sufficient to provide for the observed density of cosmic rays? Is there any difference in the mechanism of acceleration for electrons and ions? What is the role of relativistic particle emission for the surfatron acceleration of cosmic rays in the Galaxy? These and some other questions will be considered below.

The paper is organized as follows. In Section 2, we will establish optimum conditions for the surfatron acceleration mechanism and consider the properties of nonlinear waves by which the charged particles are trapped and accelerated. An important characteristic of the surfatron acceleration process—the density of particles trapped by a nonlinear wave—will be estimated in Section 3. Section 4 is devoted to elucidating the possible reasons capable of breaking the particle acceleration process. The energies of cosmic rays gained as a result of the acceleration by surfing in the Galaxy are estimated in Section 5. The main conclusions following from this investigation are summarized in Section 6.

2. CONDITIONS NECESSARY FOR THE SURFATRON ACCELERATION PROCESS

As noted above, consideration will be restricted to analysis of the charged particle acceleration by surfing in the plasma waves and quasitransverse MSSWs. In both cases, we will assume the stationary waves to be plane. The motion of articles will be described either in a laboratory frame (where the plasma is resting as the whole) or in the wave frame. Consider a wave moving in the laboratory frame in the direction opposite to the x axis at a phase velocity u not exceeding the speed of light c. The characteristic conversion factor for the passage from one to another frame is

$$\gamma_f = 1/\sqrt{1-\beta^2}, \quad \beta = u/c.$$

The surfatron acceleration takes place upon trapping of a small group of particles by a potential wave moving in the plasma across a weak magnetic field. As is known [2, 4-11], a necessary condition for the long-term ("eternal") acceleration of the trapped particles is provided by $R \geq 1$, where

$$R = E_0/B,$$

E_0 being the maximum value (amplitude) of the electric field strength in the wave and B, the magnetic field strength at a point (in the wave frame) of the electric field maximum. According to the formulas of the field transformation on the passage from one frame to another, the magnetic field strengths are related as

$$B = \gamma_f B_0,$$

where B_0 is the magnetic field strength in the laboratory frame; the electric field E_0 is the same in both frames of reference.

Let us check whether the condition $R \geq 1$ holds for the longitudinal plasma waves of large amplitude propagating in a galactic plasma. We will consider a plasma with the particle density n_0, the temperature $T \ll mc^2$, and the longitudinal wave frequency

$$\omega = \sqrt{\omega_{pe}^2 + \omega_{ce}^2}$$

for the most typical interstellar medium with

$$\omega_{pe}^2 \gg \omega_{ce}^2,$$

where

$$\omega_{pe} = \sqrt{4\pi n_0 e^2/m}, \quad \omega_{ce} = eB_0/mc,$$

e and m being the electron charge and mass, respectively, and B_0 the magnetic field strength in the resting plasma. In this case, we may ignore the effect of magnetic field on the dispersion properties of the plasma and take $\omega \approx \omega_{pe}$.

The maximum theoretically possible amplitude of the electric field strength in the stationary nonlinear plasma wave is [14]

$$E_{m0} \approx \sqrt{8\pi n_0 mc^2(\gamma_f - 1)}.$$

Taking this expression into account, the parameter R can be readily presented in the following form:

$$R \approx \frac{\omega_{pe}}{\omega_{ce}} \frac{\sqrt{\gamma_f - 1}}{\gamma_f}.$$

This relationship indicates that the necessary condition $R > 1$ is fulfilled if

$$\frac{\gamma_f^2}{\gamma_f - 1} < \frac{\omega_{pe}^2}{\omega_{ce}^2}.$$

The right-hand part of this inequality can be written as

$$\frac{\omega_{pe}^2}{\omega_{ce}^2} \approx \frac{4\pi n_0 T}{\theta_e B_0^2},$$

where $\theta_e = T/mc^2$ is a dimensionless temperature normalized to the electron rest energy. Below we will assume that the plasma pressure in the galactic plasma is on the order of the average magnetic field pressure:

$$n_0 T \sim B_0^2/8\pi.$$

This assumption is acceptable for the Galaxy [1, 2] and can be recognized as valid in the circumsolar plasma (the chromosphere of the Sun, solar corona, and solar wind). Under these conditions, we obtain

$$\frac{\omega_{ce}^2}{\omega_{pe}^2} \sim \theta_e \sim \frac{v_{TE}^2}{c^2} \ll 1,$$

where $v_{TE} = \sqrt{T/m}$ is the thermal velocity of electrons. Now the condition $R > 1$ can be expressed in terms of γ_f in a form more convenient for practical use:

$$\frac{\omega_{pe}^2}{\omega_{ce}^2} \sim \frac{1}{\theta_e} > (\gamma_f - 1) > \frac{\omega_{ce}^2}{\omega_{pe}^2} \sim \theta_e. \quad (1)$$

The temperature of particles in the circumsolar plasma is $T \approx 10$ eV, while that in the interstellar plasma of the galactic disk is $T \approx 100$ eV. Thus, the galactic plasma is characterized by the parameter

$$\theta_e \approx 2 \times (10^{-4}–10^{-5}).$$

Substituting this value into relationship (1), we obtain estimates for the maximum value of the conversion factor

$$\gamma_f \approx 5 \times (10^3–10^4)$$

and for the minimum value of this factor

$$\gamma_f - 1 \approx 2 \times (10^{-4}–10^{-5}).$$

Thus, we arrive at a conclusion that there is a limited, albeit sufficiently broad, range of plasma wave velocities in which the wave may provide for a prolonged ("eternal") acceleration of the trapped particles. As is seen, this regime may be realized for both relativistic and nonrelativistic waves in the galactic plasma, but the rate of acceleration is significantly higher in the latter case. Indeed, for a plasma wave moving in the laboratory frame in the direction opposite to the x axis and the transverse magnetic field B_0 directed in the z axis, the field accelerating particles in the surfing mode will be directed (in the wave frame) along the y axis and equal to

$$E_y = \beta B = \frac{\beta E_0}{R} = \frac{mc}{eR}\frac{\omega_{pe}(\gamma_f - 1)\sqrt{\gamma_f + 1}}{\gamma_f}.$$

This formula indicates that, for a given n_0 and the most optimum value of $R \sim 1$ [10], the accelerating field E_y determining the surfatron acceleration is markedly greater for a relativistic plasma wave ($\gamma_f > 1$) than for a nonrelativistic wave ($\gamma_f \approx 1$).

Let us proceed to determining conditions for a prolonged acceleration of nuclei by surfing in quasiperiodic MSSWs. These waves represent essentially a propagating jump of the magnetic field strength [13, 15]. Usually, both the wave front amplitude and width are related to the magnetic field jump. One of the main MSSW characteristics is the Alfvén Mach number

$$\mathcal{M}_A = u/v_A,$$

where

$$v_A = B_0/\sqrt{4\pi n_0 M}$$

is the Alfvén velocity and M is the ion mass. The MSSWs with a Mach number $\mathcal{M}_A < 3$ are called laminar, while the waves with $\mathcal{M}_A > 3$ are referred to as turbulent.

In a laminar MSSW, the magnetic field and potential profiles are approximately coinciding. A potential jump in the wave is $\varphi_A \approx \mathcal{E}_\mathcal{K}/e$ and the jump width is approximately $d = c/\omega_{pe}$ [13, 15] (here, $\mathcal{E}_\mathcal{K} = Mu^2/2$ is the kinetic energy of a flow incident onto the wave front in the wave frame). Therefore, the electric field strength can be expressed as

$$E_0 \approx \frac{Mu^2\omega_{pe}}{ec} = \frac{\mathcal{M}_A^2 B_0^2}{c\sqrt{4\pi n_0 m}},$$

and the conversion parameter $R = E_0/B$ for a laminar wave is

$$R = \mathcal{M}_A^2(v_A/c)\sqrt{M/m}.$$

Substituting the values of parameters of the circumsolar and interstellar plasma into this formula, one can readily see that $R < 1$ provided that $\mathcal{M}_A < 3$. Thus, laminar MSSWs cannot feature a prolonged acceleration regime.

In the case of a turbulent MSSW, the potential jump amplitude is approximately the same as that for the laminar wave, but the jump width becomes significantly smaller than that of the magnetic field strength. The potential jump takes place at the front end, where the magnetic field strength is almost equal to that behind the front. On the scale of the potential jump (which is on the order of the Debye radius $d \sim v_{TE}/\omega_{pe}$), the magnetic field remains virtually constant and the jump is referred to as isomagnetic [8, 9, 15]. For a turbulent MSSW with a jump width of $d \sim v_{TE}/\omega_{pe}$, the parameter

$$R \approx Mu^2/edB$$

is determined by the relationship

$$R \sim \mathcal{M}_A^2 \frac{c}{v_{TE}}\frac{\omega_{ce}}{\omega_{pe}}.$$

Under the assumptions made above for the galactic plasma,

$$\omega_{ce}/\omega_{pe} \sim v_{TE}/c,$$

we obtain that

$$R \sim \mathcal{M}_A^2.$$

Thus, in a turbulent MSSW ($\mathcal{M}_A > 3$) propagating in a galactic plasma, the regime of prolonged acceleration for ions can be realized provided that the potential jump width is sufficiently small:

$$d < \mathcal{M}_A^2 v_{TE}/\omega_{pe}.$$

An analysis of the data of laboratory and satellite measurements [9, 15] allows us to conclude that such small width of the isomagnetic potential jump is quite possible for MSSWs in practice.

3. CALCULATION OF THE DENSITY OF CHARGED PARTICLES TRAPPED BY A WAVE

Now we will estimate the density of charged particles trapped into the potential well of a wave. Assuming that the particles are trapped from a plasma, the density of such particles in the laboratory frame is determined by the density of particles moving in the plasma at a velocity close to the wave velocity. Now we will separately consider the trapping of plasma electrons and ions, assuming that their velocities are distributed according to the Maxwell law.

First, let us consider the conditions for trapping electrons from the plasma. The density of electrons n_T trapped by a nonlinear plasma wave obeys the relationship

$$n_T \propto n_0 \exp\{-(\gamma_f - 1)/\theta_e\}. \quad (2)$$

As is seen, a considerable proportion of electrons can be trapped by a wave only in a nonrelativistic case, that is, for $\gamma_f - 1 \sim \theta_e$. According to relationships (1), a prolonged acceleration regime in this case is still possible. Thus, a relativistic wave is characterized by a maximum accelerating field at an almost zero density of electrons, while a nonrelativistic wave may trap a large density of electrons but provides for a low acceleration rate. The main conclusion from this analysis is that, provided relationship (1) is satisfied, an electron concentration observed in cosmic rays ($n_e \sim 10^{-12}$ cm^{-3}) can be ensured by the partial resonance trapping of electrons from the galactic plasma into a nonrelativistic plasma wave.

Let us determine the density of trapped ions. Evidently, the proportion of ions trapped from plasma by a nonlinear plasma wave is negligibly small. Indeed, this density is determined by formula (2), where the exponent is M/m times greater than that in the case of electrons because the θ_e value has to be replaced by $\theta_i = T/(Mc^2)$, which is M/m times smaller than θ_e.

In order to find the density of ions trapped from plasma by a quasitransverse MSSW, we might use the results reported in [8], where the n_T value was calculated as a function of the electric field strength at the potential jump, the potential amplitude, the temperature, and the mass of ions incident onto the shock wave front. However, we will restrict the consideration to obtaining rough estimates of the density of trapped ions using the formula

$$n_T \propto n_0 \exp\{-Mu^2/2T\}.$$

As is seen, the n_T/n_0 ratio is determined primarily by the exponent $Mu^2/2T$. For a galactic plasma, this value can be written as

$$\frac{Mu^2}{2T} = \frac{\mathcal{M}_A^2 B_0^2}{8\pi n_0 T} \sim \mathcal{M}_A^2,$$

which shows that the density of ions in a laminar MSSW ($\mathcal{M}_A < 3$) provides for an ion density observed in the cosmic rays ($n_i \sim 10^{-10}$ cm^{-3}). The same conclusion would be obtained based on the results of a stricter analysis [8]. Using the above formula to estimate the density of particles for a turbulent MSSWs ($\mathcal{M}_A > 3$), we establish that the observed ion density can be provided for waves with the Mach number below 5.

In the Galaxy, the wave may trap particles both from the plasma and from galactic cosmic rays. Evidently, the latter implies trapping by relativistic plasma waves. Let us evaluate the density of trapped particles in this case as well. As is known [1, 2], the energy distribution function of cosmic rays in the laboratory frame can be written as

$$f(\gamma) = \frac{K(k-1)}{c\mathscr{E}^k} = \frac{K(k-1)}{c(\gamma m_r c^2)^k},$$

where K is a proportionality factor [particles/(cm^2 sr s)], k is the exponent, $\mathscr{E}$ is the energy, $\gamma = \mathscr{E}/m_r c^2$ is the dimensionless energy, and m_r is the mass of particles (electrons or ions).

Apparently, in the wave frame all particles which move at velocities sufficiently close to zero will be trapped. Based on the results obtained previously [10], the scatter in dimensionless energy in the case of $1 \le R \le 2$ and $eZ\varphi_A/(m_r c^2) > 1$ (Z is the charge number) may be limited by two characteristic quantities: (i) $\Delta\gamma \sim 1$ (rest particle energy) and (ii) $\Delta\gamma \sim \gamma_f$ (energy of particles moving at a wave velocity). In the former case, the dimensionless momentum of trapped particles varies between zero and $p_x = \pm 1$ (accordingly, the dimensionless energy γ varies from 1 to $\sqrt{2}$). By the same token, in the latter case, the momentum varies from zero to $p_x = \pm\beta\gamma_f$ and the energy, from $\gamma = 1$ to γ_f.

In the laboratory frame, the dimensionless energy varies from $\gamma = \gamma_f/2$ to $2.5\gamma_f$ in the former case and from $\gamma = 1$ to γ_f^2 in the latter case. Accordingly, the density of trapped particles in the laboratory frame in the former case is

$$n_{T1}^L \approx \int_{\gamma_f/2}^{2.5\gamma_f} d\gamma \frac{K(k-1)}{c\gamma^k} \approx \frac{K}{c\gamma_f^{(k-1)}},$$

and that in the latter case is

$$n_{T2}^L \approx \int_1^{\gamma_f^2} d\gamma \frac{K(k-1)}{c\gamma^k} \approx \frac{K}{c}.$$

The n_{T2}^L value is comparable with the density of the corresponding component observed in the cosmic rays. Although the probability of trapping such a large number of particles in the wave is evidently small, we will consider this value as limiting.

Thus, the above analysis revealed that the number of electrons trapped by a nonrelativistic plasma wave from the galactic plasma in the regime of prolonged acceleration is sufficient to provide for the density of electrons observed in the cosmic rays. The necessary amount of ions can be trapped in the MSSW fronts at a Mach number not exceeding 5. The regime of prolonged acceleration can be realized only for turbulent MSSWs in the interval $3 < \mathcal{M}_A < 5$. For a relativistic plasma wave, the number of particles (both electrons and ions) trapped from the galactic plasma is negligibly small. In this case, however, the particles can be trapped from cosmic rays. A wave of sufficiently large amplitude can trap a considerable amount of particles from this source.

4. POSSIBLE FACTORS LIMITING THE PARTICLE ACCELERATION TIME

For any process of particle acceleration, an important role is played by the initial stage usually involving the so-called injection problem [1, 2]. We all demonstrate that, in the case of particle acceleration by the surfatron mechanism, the rate of energy gain is so high that the acceleration proceeds in the injectionless mode (i.e., the injection problem is completely eliminated). Let us compare the rate of energy gain by particles involved in the surfing process to the rate of energy loss by accelerated particles for the ionization and collision events. The maximum level of energy losses takes place in the nonrelativistic energy range (for protons in the atomic hydrogen medium, $\mathcal{E}_H \approx 60$ keV [1]). The maximum loss in the gas medium can be expressed as [1]

$$W_I \approx \frac{4\pi e^4 n_0 Z^2}{mv} \approx \frac{Z^2 e^2 \omega_{pe}^2}{v}, \quad (3)$$

where n_0 is the gas density and v is the velocity of ions with the charge number Z (for protons, $v \approx 3 \times 10^8$ cm/s). In a fully ionized gas, the velocity v in formula (3) has to be replaced by v_{TE} [1].

Let us determine the ratio of W_I to the rate W_S of energy gain by surfing. In the nonrelativistic case ($\gamma_f \approx 1$), the latter value obeys the relationship

$$W_S = eZE_y v = eZB_0 u v/c.$$

Taking into account that, for an electron temperature in the range $T_e \sim 10$–100 eV the particle velocity is on the order of $v \sim v_{TE}$, we obtain for the W_I/W_S ratio

$$\frac{W_I}{W_S} = \frac{e^2 Z^2 \omega_{pe}^2}{eZB_0 u v_{TE}^2} \approx \frac{\omega_{pe}^2}{\omega_{ce}^2} \frac{mc^2}{T_e} \frac{e^3 B_0}{m^2 c^4} \frac{cZ}{u}$$

$$\approx \frac{\omega_{pe}^2}{\omega_{ce}^2} \frac{Z}{\theta_e \beta} \frac{B_0}{E_e} \approx \frac{Z}{\beta \theta_e^2} \frac{B_0}{E_e}.$$

Here, $E_e = e/r_0^2$ is the electric field generated by an electron at a distance of r_0 from its center, where $r_0 = e^2/(mc^2)$ is the classical electron radius. Since the E_e value is extremely large ($E_e \sim 10^{16}$ CGSE units), the above ratio for the galactic plasma is significantly lower than unity in the entire range of parameters β, θ, charge numbers Z, and magnetic field strengths B_0.

Now let us analyze the possible factors that can hinder the theoretically unlimited growth in the energy of particles trapped in the potential wave and accelerated by the surfatron mechanism. The first circumstance is the finite size of a region in which the plasma wave may spread in a real case. The second factor is the radiative loss of energy by the accelerated particles. The third reason is the wave damping as a result of the energy loss for the acceleration of charged particles.

Let us consider the particle energy limitation related to finite dimensions of the region of wave spreading. In the surfatron mechanism, a particle is accelerated in the direction perpendicular to the wave propagation, while the particle is trapped and carried by the wave. Thus, in the laboratory frame, the particle performs a two-dimensional motion by (i) drifting with the wave in the propagation direction and (ii) shifting in the transverse direction due to the acceleration by surfing. For a nonrelativistic velocity, the transverse wave size is more important [6, 8, 9]. This is related to the fact that the particle velocity in the transverse direction may rapidly exceed the wave propagation velocity; therefore, a prolonged acceleration will only be possible provided that the transverse size of a region occupied by the wave is greater than the longitudinal size. Here, it would be expedient to recall a conclusion made in [6], according to which restrictions related to the transverse wave size can be removed to a certain extent for a sufficiently high curvature of the potential wave front.

For a relativistic wave, the path traveled by a particle in the transverse direction in the laboratory frame is smaller than the longitudinal distance traveled with the wave. Indeed, for an immobile observer (occurring in the laboratory frame), a wave propagating at a velocity of $u \approx c$ for a time t would cover a distance of $L_x = ut \approx ct$. Passing to the wave frame and assuming that the particle accelerated by surfing has reached a relativistic stage, we obtain for the particle velocity components

$$v_x \approx 0, \quad v_y \approx c.$$

In the wave frame, according to the Lorentz transformation, the corresponding time interval amounts to $t_W = t/\gamma_f$. During the time t_W, the particle in the wave frame would travel in the transverse direction a distance of $L_y \approx ct_W$. According to the Lorentz transformation rules, the transverse path observed in both frames must be the same. This implies that, in the laboratory frame, the particle travels a distance of $L_y \approx ct/\gamma_f = L_x/\gamma_f$, which is γ_f times shorter than L_x.

Now let us estimate the energy of a particle trapped by a nonlinear wave and accelerated due to surfing for a time t during which the wave travels a distance of $L_x = ut$. The rate of the energy gain by the particle must be the same in both frames, but the expression is simpler in the wave frame, where this rate is $eZE_y v_y$ ($E_y = \beta B = \beta\gamma_f B_0$, B_0 being the magnetic field strength in the laboratory frame). As was demonstrated in [10], the velocity v_y of a trapped particle accelerated by surfing approaches within a short time (on the order of a cyclotron period) the speed of light. Putting $v_y \approx c$ in the above expression, we obtained that the energy gained by the particle for the time t is

$$\mathcal{E} \approx eZ\beta\gamma_f B_0 ct. \tag{4}$$

In the case of a nonrelativistic plasma wave and quasitransverse MSSWs propagating in the galactic plasma at a nonrelativistic velocity, the particle acceleration time will be restricted (see above) by the transverse system size $L_y \approx ct$. Using Eq. (4) with $\gamma_f \approx 1$, we obtain for the corresponding particle energy

$$\mathcal{E} \approx eZ\beta B_0 L_y, \tag{5}$$

where $L_y \approx L_x c/u \gg L_x$.

For a relativistic plasma wave ($\beta \approx 1$), the acceleration time is limited by the longitudinal size $L_x = ct$. In this case, Eq. (4) yields

$$\mathcal{E} \approx 10^{21} Z\gamma_f B_0 L_x, \tag{6}$$

where B_0 is expressed in oersteds (Oe), L_x in parsecs (pc), and $\mathcal{E}$ in electronvolts (eV). Here, a particle in the relativistic wave will shift in the transverse (acceleration) direction by a distance that is γ_f times shorter than L_x.

Thus, the maximum energies of particles accelerated by surfing across the nonlinear plasma waves can be limited as a result of a finite size of the region featuring favorable wave propagation conditions. For a limiting energy gained during acceleration in a nonrelativistic wave, the most critical parameter is the transverse size L_y given by formula (5). For a relativistic wave, the maximum energy is limited mostly by the longitudinal size L_x of the possible wave propagation region described by formula (6).

A part of the energy of particles accelerated by surfing up to relativistic and ultrarelativistic velocities is spent for radiation. We will consider the energy lost by accelerated particles for the radiation of three types. The first radiation is related to the surfatron acceleration process as such. As is known [2, 4–11], a particle in the course of surfing moves (in the wave frame) predominantly in a constant homogeneous electric field $E_y = uB/c = \beta\gamma_f B_0$. Indeed, an ideally trapped particle reaching the relativistic or ultrarelativistic acceleration stage is subject in this frame to no forces other than that of the electric field E_y. The constant power of radiation due to the surfatron acceleration of the particle in the constant homogeneous electric field is [10]

$$W_E \approx \frac{Z^4 e^4 E_y^2}{m_r^2 c^3}.$$

Since the radiation power is inversely proportional to the squared particle mass, a significant role in practice is played only by the acceleration of electrons, so that we may put $m_r = m$ and $Z = 1$.

Let us estimate the ratio of the radiation power W_E to the rate W_S of the energy gain for electrons accelerated by surfing in a plasma wave. Assuming the velocity components of trapped electrons to be $v_x = 0$ and $v_y = c$ [10], we obtain $W_S = eE_y c$. Therefore, the radiated to gained power ratio is

$$\frac{W_E}{W_S} = \frac{e^3 E_y}{m^2 c^4} = \beta\gamma_f \frac{B_0}{E_e}.$$

For the parameters β, γ_f and the B_0 values employed, the power radiated by accelerated electrons in the Galaxy is negligibly small as compared to the rate of energy gain. Thus, the radiation from electrons (and the more so from ions) accelerated by surfing in the electric field E_y can be ignored.

The absence of the effect of a magnetic field on the ideally trapped particles accelerated by the surfatron mechanism leads to an important conclusion that this process is not accompanied by synchrotron (magnetic bremsstrahlung) radiation from electrons. As is known [1, 2], the main part of cosmic radiation in the RF range is due to cosmic rays, the major contribution being due to the synchrotron radiation of relativistic electrons moving in the galactic magnetic fields. It was suggested [3] that electrons in the Galaxy, losing energy for the synchrotron radiation, cannot gain energies exceeding 10^{15} eV. It appears that electrons experiencing acceleration by the surfatron mechanism generate, despite the presence of a

magnetic field, no synchrotron radiation and, hence, are not subject to this restriction. Indeed, electrons accelerated by surfing may, in principle, acquire an arbitrarily high energy on moving across the magnetic field with the wave. These electrons produce no synchrotron radiating as long as they are trapped in the wave.

Another possible radiation from particles moving in a plasma medium is related to their retardation as a result of the wave excitation in the plasma [14]. Let us estimate the energy loss for a plasma wave observed from the laboratory frame. As above, we will consider a stationary stage of the surfing process and assume the particle velocity to be close to the speed of light.

According to [14], the power W_R of the radiation of this type for particles with a mass $m_r \gg m$ is described by a formula similar to Eq. (3):

$$W_R = \frac{4\pi n_0 Z^2 e^4}{mv}\ln\Lambda \approx \frac{Z^2 e^2 \omega_{pe}^2}{v}\ln\Lambda.$$

Here $\Lambda = mc^2/\hbar\omega_{pe}$ and it is assumed that the particle velocity obeys the requirements

$$v \gg v_{TE}, \quad v > Ze^2/\hbar \approx 2Z \times 10^8 \text{ cm/s}.$$

Note that the latter formula can be used only for electrons with $v \approx c$ [14].

Let us estimate the ratio of the radiation power W_R to the rate W_S of the particle energy gain during surfing in a plasma wave. For the particles with $v \approx c$, the radiated to gained power ratio is

$$\frac{W_R}{W_S} \approx \frac{Z^2 e^2 \omega_{pe}^2}{ZeE_y c^2}\ln\Lambda \approx \frac{Ze\omega_{pe}^2}{\beta\gamma_f B_0 c^2}\ln\Lambda$$

$$\approx \frac{Z}{\beta\gamma_f}\frac{\omega_{pe}^2}{\omega_{ce}^2}\frac{B_0}{E_e}\ln\Lambda.$$

For the plasma parameters employed, this ratio is negligibly small because $\ln\Lambda < 100$. Thus, the particle retardation as a result of the wave emission during motion in the plasma can be ignored as well.

In the Galaxy, relativistic electrons may lose energy as a result of the scattering on photons, which is referred to as Compton energy losses [1, 2, 16]. The powers of synchrotron and Compton radiation in the Galaxy are proportional to the squared particle energy and exhibit comparable values for electron energies below $\mathscr{E}_m \sim 10^{11}$ eV [1,2, 6]. Above the $\mathscr{E}_m$ level, the Compton losses cease to depend on the energy and remain virtually constant [16].

As above, let us compare the power of Compton radiation to the rate of the particle energy gain due to the surfatron acceleration process. According to [1, 2, 16], for $\mathscr{E} < \mathscr{E}_m$ the rate of Compton energy losses is $\sim 10^{-16}\mathscr{E}^2$ GeV/s ($\mathscr{E}$ is the electron energy in GeV). The ratio of the rate of energy losses as a result of the Compton scattering under conditions typical of the Galaxy ($B_0 \approx 3 \times 10^{-6}$ Oe, $\mathscr{E}_m \sim 10^2$ GeV, $\gamma_f > 1$) to the rate of the electron energy gain due to acceleration is negligibly small. Therefore, we may neglect the Compton energy losses for electrons accelerated by surfing in the relativistic plasma waves.

Thus, we come to a conclusion that the energy losses related to the main three types of radiation generated by particles accelerated according to the surfatron mechanism can be ignored.

Finally, let us evaluate the energy limitation as a result of the wave damping caused by the trapped particles. Since we are speaking of the maximum energies, we will consider the case of particles trapped by a relativistic wave. As was demonstrated in Section 3, this can take place only for particles in the cosmic rays. The relativistic plasma wave damping can be roughly estimated using the relationship

$$\frac{d\langle E^2\rangle}{d\tau} + 8\pi m_r c^2 \left\langle n_T \frac{d\gamma}{d\tau}\right\rangle = 0,$$

where $\tau = eB_0 t/m_r c$ is the dimensionless time; the angle brackets denote averaging with respect to the wavelength $\lambda = u/\omega_{pe}$. Taking into account that, in the relativistic stage ($\beta \approx 1$), the energy in the laboratory frame is

$$\gamma \approx \gamma_f \beta\tau \approx \gamma_f \tau,$$

and, hence,

$$d\gamma/d\tau \approx \gamma_f,$$

we obtain

$$\langle n_T d\gamma/d\tau\rangle \approx \gamma_f \langle n_T\rangle.$$

The wavelength-averaged density $\langle n_T\rangle$ is rather difficult to evaluate because all the trapped particles accelerated by surfing tend to gather at a special point in the potential profile, in which the strengths of electric and magnetic fields are equal $E_0 = B$ (in the wave frame) [10]. In a rough approximation based on a thorough analysis of the surfing process performed in [10], we may take

$$\langle n_T\rangle/n \approx \delta x/\lambda,$$

where n is the density of particles in a negligibly small vicinity δx of the special point. The $\delta x/\lambda$ ratio can be estimated using the relationship

$$\delta x/\lambda \sim (R-1)/R,$$

where $R \approx 1$ (optimum surfing regime) and, hence,

$$\delta x/\lambda \sim (R-1)/R \ll 1.$$

Note that $R \longrightarrow 1$ at least in the course of the wave damping, since we are considering a final stage of the electric field decay in which the particles "fall out" of the potential well [4, 10]. Taking for simplicity

$$(R-1)/R \leq 1/8\pi,$$

we obtain an expression for the limiting energy

$$\mathscr{E}_m \approx E_0^2/n.$$

Since $E_0 \approx B = \gamma_f B_0$, we finally arrive at

$$\mathscr{E}_m \approx (\gamma_f B_0)^2/n.$$

Assuming that the trapped particle density n is comparable with that observed in cosmic rays ($n \sim 10^{-10}$ cm^{-3}) and substituting $B_0 \approx 3 \times 10^{-6}$ Oe and $\gamma_f \sim 10^2$–10^4, we obtain an estimate for the limiting energy $\mathscr{E}_m \sim 10^{15}$–$10^{19}$ eV. This estimate shows that, even when a greater part of the particles from cosmic rays are trapped in a plasma wave, the particles may gain an energy up to 10^{19} with an allowance for the wave damping. Apparently, a much lower fraction of particles are usually trapped and, hence, the wave damping does not prevent particles from acquiring a still greater energy.

We did not consider some other factors that may restrict the time or reduce the efficiency of the acceleration process. These may include the nuclear and photonuclear interactions of the ion component of cosmic rays with an interstellar medium, the nuclear fragmentation processes, the presence of a nonzero angle between the wave propagation direction and the magnetic field vector (oblique wave) [9], the wave front deviation from plane geometry (these factors were partly considered in [6]), the effect of inhomogeneity in the plasma density and magnetic field strength distribution, the instability of the plasma–nonlinear wave–accelerated article system, etc.

5. ESTIMATES OF THE PARTICLE ENERGY GAINED BY SURFING

In order to provide for optimum surfing conditions, the nonlinear plasma waves in the Galaxy must be capable of spreading over sufficiently extended regions with a quasihomogeneous magnetic field. It was suggested [1] that such regions may exist in the spiral arms of the Galaxy. The magnetic field must be quasihomogeneous upon large-scale averaging. The spatial dimensions of magnetic fields in the spiral arms are comparable to the size of these arms, with a thickness of 200–500 pc and a length on the order of 10^3–10^4 pc.

Now let us estimate the energy the particles may acquire as a result of acceleration by the surfatron mechanism in various regions of the Galaxy. First, consider the particles accelerated in a stellar atmosphere (exemplified by the chromosphere of the Sun). The acceleration of protons by surfing in the chromospheric plasma was considered in detail previously [9]. It was established that protons trapped in a turbulent MSSW with an isomagnetic potential jump may acquire an energy on the order of 10 GeV, while the energy gained in a laminar MSSW may be on the order of 10 MeV. By the same token, one may readily check that energies of the same order can be gained by protons accelerated by surfing in interstellar MSSWs propagating in the solar system (i.e., in the solar wind plasma).

Let us consider the acceleration of electrons in a circumsolar plasma by means of surfing in a nonrelativistic plasma wave. Substituting typical parameters of the chromospheric plasma ($\beta_0 \approx 10$ Oe, $\beta \sim 10^{-2}$, $L_y \sim 10^9$ cm) or the solar wind ($\beta_0 \approx 10^{-4}$ Oe, $\beta \sim 10^{-2}$, $L_y \sim 10$ AU $\sim 10^{14}$ cm) into formula (5), we obtain a value on the order of $\mathscr{E} \sim 10^{10}$ eV for the energy of electrons accelerated by surfing in nonrelativistic plasma waves in the vicinity of the Sun.

Thus, charged particles in the circumsolar plasma can be accelerated by surfing up to an energy on the order of 10 GeV/nucleon. As was pointed out in [1], a product of the characteristic scale by the magnetic field entering into formulas (4)–(6) for some other stars may be greater by three orders of magnitude. Therefore, we may conclude that particles accelerated by the surfatron mechanism in a stellar atmosphere may gain an energy of up to 10^{13} eV/nucleon.

Note that a source of the stellar cosmic rays are particles originating deep in the stars and then leaving them to be accelerated in the solar plasma atmosphere from thermal to relativistic energies as a result of surfing across the nonlinear waves generated by flares or other powerful perturbations in the circumsolar plasma. Subsequently, these high-energy particles may be carried out to the interstellar space: electrons, by nonrelativistic plasma waves and ions, by magnetosonic shock waves.

Let us estimate the energies of electrons and ions trapped from interstellar plasma and accelerated in nonrelativistic nonlinear waves ($\gamma_f \approx 1$) propagating in a plasma of the galactic disk. As was demonstrated in Section 3, electrons of the interstellar plasma can be also trapped in this system and accelerated in the plasma waves. Substituting the corresponding parameters ($B_0 \approx 3 \times 10^{-6}$ Oe, $\beta \sim 10^{-2}$, $L_y \sim 100$ pc $\sim 10^{20}$ cm) into formula (5), we estimate the electron energy at $\mathscr{E} \sim 10^{15}$ eV.

As was demonstrated in Section 3, the ions may be trapped in a sufficient amount by MSSWs with the Mach numbers below 5. In a laminar MSSW ($\mathscr{M}_A < 3$), a small proportion of ions continuously incident onto the wave front are also trapped and accelerated to an energy of $\mathscr{E} \approx \mathscr{E}_{\mathscr{H}}(M/m)$ [13]. For example, a shock wave with $\mathscr{M}_A = 2$ and $T \sim 100$ eV in the interstellar plasma accelerates protons up to an energy of

$$\mathscr{E} \approx Mu^2(M/m) \approx \mathscr{M}_A^2 T/(8\pi n_0 T/B_0^2)(M/m)$$
$$\approx \mathscr{M}_A^2 T(M/m) \approx 1 \text{ MeV}.$$

Upon a prolonged acceleration by surfing in a turbulent MSSW (for $3 < \mathscr{M}_A < 5$), a limiting energy of protons in the interstellar plasma can be estimated by formula (5). Taking $\mathscr{M}_A = 5$, $n_0 = 3 \times 10^{-3}$ cm^{-3} and the typical values

of $B_0 = 3 \times 10^{-6}$ Oe, $\beta = \mathcal{M}_A v_A/c \approx 2 \times 10^{-3}$, and $L_y \sim 100$ pc $\approx 3 \times 10^{20}$ cm into formula (5), we estimate the ion energy at $\mathcal{E} \approx 10^{15}$ eV.

Thus, the charged particles trapped by nonrelativistic nonlinear waves from the galactic plasma can be accelerated by surfing in the Galaxy up to an energy of $\mathcal{E} \sim 10^{13}$ eV/nucleon in the stellar atmospheres and up to $\mathcal{E} \sim 10^{15}$ eV/nucleon in the galactic disk.

Acceleration of these particles to still higher energies can be provided only by the relativistic plasma waves ($\gamma_f > 1$). To this end, particles acquiring relativistic energies as described above must be trapped again by relativistic plasma waves and continue gaining energy in the new stage of surfing. Thus, we conclude that cosmic rays may acquire large energies ($\mathcal{E} > 10^{15}$ eV/nucleon) only upon several stages of the surfatron acceleration. We may suggest that the acceleration of particles proceeds in most cases in two stages. In the first stage, ions and electrons are trapped by nonrelativistic waves from the galactic plasma and accelerated by surfing to relativistic energies. In the second stage, these particles continue gaining energy by surfing in a relativistic plasma wave.

We believe that nonlinear plasma waves continuously appear in the Galaxy and most probably possess isotropic directions and a sufficiently broad range of velocities. Under these conditions, particles trapped in and then lost from one wave may be resonantly trapped by another (newly created) nonlinear wave moving in a favorable direction at an appropriate velocity. This acceleration process can be multiply repeated in a "relay" mode.

Note that, on leaving a trapping wave, particles will be held by a magnetic field within a certain region determined by their Larmor radius. Then the particles are either trapped by another appropriate relativistic plasma wave so as to continue gaining the energy, or they lose the previously acquired energy via various channels. Apparently, in the interruptions between acceleration stages, the energy is more rapidly lost by relativistic and ultrarelativistic electrons. Indeed, such particles moving in the galactic magnetic field will intensively lose energy by producing synchrotron radiation.

Let us estimate the energies gained by particles accelerated in the second stage of surfing in a relativistic plasma wave propagating over a distance characteristic of the Galaxy. Substituting a magnetic field strength of $B_0 \approx 3 \times 10^{-6}$ Oe and a characteristic size comparable with the galactic disk dimensions $L_x \approx 300$ pc into formula (6), we obtain for $\gamma_f \approx 10$–10^4 an estimate of $\mathcal{E} \approx 3Z(10^{16}$–$10^{19})$ eV. Therefore, the proposed mechanism of the charged particle acceleration by the surfatron mechanism provides an explanation for the formation of a high-energy part of the spectrum of cosmic rays observed in the Galaxy.

Thus, upon considering the formation of the energy spectrum of cosmic rays we may conclude that the whole spectrum is formed in two stages. In the first stage, the particles of interstellar plasma belonging to a far "tail" part of the Maxwell distribution function are trapped by nonrelativistic waves and accelerated by surfing in an injectionless mode up to an energy of $\mathcal{E} \sim 10^{15}$ eV. In the second stage, these particles are trapped by relativistic plasma waves and continue gaining energy by surfing up to $\mathcal{E} \sim 10^{19}$ eV. Here, a noteworthy fact is that the boundary energy of $\mathcal{E} \sim 10^{15}$ eV separating the two stages falls within the region of inflection in the energy spectrum of cosmic rays, after which the spectrum slope becomes steeper [2].

An important question concerning the shape of the energy spectrum of cosmic rays requires special analysis. However, preliminary conclusions formulated in [4] suggest that the differential energy spectrum corresponding to the surfatron acceleration mechanism is described by a power law with the exponent $k = 3$ close to that observed in experiment.

6. FINAL REMARKS AND PRINCIPAL CONCLUSIONS

The results of this investigation of the process of surfatron acceleration of particles showed that many problems related to the acceleration of cosmic rays in the Galaxy can be solved within the framework of this mechanism. First, the charged particles are trapped by plasma waves directly from the galactic plasma at an amount sufficient to provide for the observed density of cosmic rays in the Galaxy. Second, the same mechanism may account for the injectionless particle acceleration from thermal energies in the plasma to a limiting energy in the cosmic rays. Third, there are no principal differences in the acceleration of various charged particles: both electrons and nuclei are accelerated in a similar manner to the same limiting energies (per unit charge). Since the particles are trapped by a wave in a resonant manner and the acceleration rate is constant, there is no danger for complex nuclei to break in the course of acceleration. Finally, according to the results obtained in [4], the differential energy spectrum of cosmic rays corresponding to the surfatron acceleration mechanism is close to that observed in experiment (a power law with an exponent close to $k \approx 3$).

Thus, summarizing the results of this investigation we may conclude the following:

1. The main source of cosmic rays is represented by the interstellar plasma. The cosmic rays formed in the plasma atmospheres of stars are partly carried away by nonlinear waves into the interstellar medium. Within the framework of the surfatron acceleration mechanism, a problem concerning the primary or secondary origination of the electron component in the cosmic rays [1] is solved in favor of the primary electrons.

2. The high-energy part of the spectrum of cosmic rays in the Galaxy is formed in two stages. In the first stage, a small part of the charged particles of the galactic plasma

are trapped by nonrelativistic nonlinear waves and accelerated by the surfatron mechanism from thermal energies to an energy of $\mathscr{E} \sim 10^{13}$ eV/nucleon in star atmospheres or up to $\mathscr{E} \sim 10^{15}$ eV/nucleon in the galactic disk. In the second stage, these particles can be trapped by relativistic plasma waves and continue gaining energy by surfing up to $\mathscr{E} \sim 10^{16}$–10^{19} eV/nucleon. A remarkable fact is that the boundary energy separating the two stages falls within the region of inflection in the energy spectrum of cosmic rays [2].

3. The maximum energy of particles accelerated by surfing in the nonlinear plasma waves is limited by dimensions of the region of wave propagation. During the surfatron acceleration, the energy losses of relativistic particles related to the known radiation types and the damping of nonlinear weaves as a result of the energy being spent for the acceleration of particles can be ignored in the first approximation.

4. The surfatron acceleration is not accompanied by the synchrotron radiation, which is the most effective channel of energy loss for the accelerated relativistic particles. For this reason, electrons (as well as nuclei) in the Galaxy can be accelerated by the surfatron mechanism up to an energy of $\mathscr{E} \sim 10^{19}$ eV.

REFERENCES

1. V. L. Ginzburg and S. I. Syrovatskii, *The Origin of Cosmic Rays* (Akad. Nauk SSSR, Moscow, 1963; Gordon and Breach, New York, 1969).
2. V. S. Berezinskiĭ, S. V. Bulanov, V. L. Ginzburg, *et al.*, in *The Astropbysics of Cosmic Rays*, Ed. by V. L. Ginzburg, (Nauka, Moscow, 1984).
3. E. G. Berezhko and G. F. Krymskiĭ, Usp. Fiz. Nauk **154**, 49 (1988) [Sov. Phys. Usp. **31**, 27 (1988)].
4. N. S. Erokhin, S. S. Moiseev, and R. Z. Sagdeev, Pis'ma Astron. Zh. **15**, 3 (1989) [Sov. Astron. Lett. **15**, 1 (1989)].
5. B. É. Gribov, R. Z. Sagdeev, V. D. Shapiro, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **42**, 54 (1985) [JETP Lett. **42**, 63 (1985)].
6. S. V. Bulanov and A. S. Sakharov, Pis'ma Zh. Éksp. Teor. Fiz. **44**, 421 (1986) [JETP Lett. **44**, 543 (1986)].
7. T. Katsouleas and J. M. Dawson, Phys. Rev. Lett. **51**, 392 (1983).
8. G. N. Kichigin, Zh. Éksp. Teor. Fiz. **101**, 1487 (1992) [Sov. Phys. JETP **74**, 793 (1992)].
9. G. N. Kichigin, Pis'ma Astron. Zh. **19**, 547 (1993) [Astron. Lett. **19**, 216 (1993)].
10. G. N. Kichigin, Zh. Éksp. Teor. Fiz. **108**, 1342 (1995) [JETP **81**, 736 (1995)].
11. Ya. B. Faĭnberg, Fiz. Plazmy **13**, 607 (1987) [Sov. J. Plasma Phys. **13**, 350 (1987)].
12. B. Breizman, D. L. Fisher, P. Z. Chebotaev, *et al.*, Preprint No. DOE/ET-53088-502 (Institute for Fusion Studies, 1991).
13. R. Z. Sagdeev, in *Reviews of Plasma Physics*, Ed. by M. A. Leontovich (Atomizdat, Moscow, 1964; Consultants Bureau, New York, 1968), Vol. 4, p. 20.
14. *Plasma Electrodynamics*, Ed. by A. I. Akhiezer (Nauka, Moscow, 1974; Pergamon, Oxford, 1975).
15. V. G. Eselevich, Planet. Space Sci. **31**, 615 (1983).
16. V. L. Ginzburg, *Theoretical Physics and Astrophysics* (Nauka, Moscow, 1975; Pergamon, Oxford, 1979).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 905–920.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1050–1066.

GRAVITATION, ASTROPHYSICS

Non-Abelian Stokes Theorems in the Yang–Mills and Gravity Theories¶

D. I. Diakonov[a, b, *] and V. Yu. Petrov[b, **]

[a]*NORDITA, DK-2100 Copenhagen ∅, Denmark*

[b]*St. Petersburg Nuclear Physics Institute, Gatchina, Leningrad region, 188350 Russia*

**e-mail: diakonov@nordita.dk*

***e-mail: victorp@thd.pnpi.spb.ru*

Received February 1, 2001

Abstract—We discuss the interpretation of the non-Abelian Stokes theorem for the Wilson loop in the Yang–Mills theory. For the "gravitational Wilson loops," i.e., holonomies in curved $d = 2, 3, 4$ spaces, we then derive "non-Abelian Stokes theorems" that are similar to our formula in the Yang–Mills theory. In particular, we derive an elegant formula for the holonomy in the case of a constant-curvature background in three dimensions and a formula for small-area loops in any number of dimensions. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

One of the main objects in the Yang–Mills theory and in gravity is the parallel transporter along closed contours, or the holonomy. In the Yang–Mills theory, it is conventionally called the Wilson loop; it can be written as a path-ordered exponential

$$W_r = \frac{1}{d(r)}\mathrm{Tr}\,\mathrm{P}\exp\left(i\oint d\tau \frac{dx^\mu}{d\tau}A^a_\mu T^a\right), \qquad (1)$$

where $x^\mu(\tau)$ with $0 \le \tau \le 1$ parametrizes the closed contour, A^a_μ is the Yang–Mills field (or connection), and T^a are the gauge group generators in a given representation r whose dimension is $d(r)$. For d-dimensional vectors in curved Riemannian spaces, the "gravitational Wilson loop," or holonomy, can also be written as a trace of the path-ordered exponential of the connection given by the Christoffel symbol,

$$W^G_{\mathrm{vector}} = \frac{1}{d}\left[\mathrm{P}\exp\left(-\oint d\tau \frac{dx^\mu}{d\tau}\Gamma_\mu\right)\right]^\kappa_\kappa. \qquad (2)$$

One can also consider parallel transporters of spinors in a curved background: the holonomy is then defined not by the Christoffel symbols, but by the spin connection (see the precise definitions below).

The Yang–Mills Wilson loop is invariant under gauge transformations of the background field A_μ; the gravitational Wilson loop is invariant under general coordinate transformations, or diffeomorphisms, provided the contour is transformed as well.

It is generally believed that, in three and four dimensions, the average of the Wilson loop in pure Yang–Mills quantum theory exhibits the area law behavior for large and simple (e.g., flat rectangular) contours. This must be true not for all representations, but only those with a nonzero "N-ality"; in the simplest case of the $SU(2)$ gauge group, these are the representations with a half-integer spin J.

One of the difficulties in proving the area law for the Wilson loop is that it is a complicated object by itself: it is impossible to compute it analytically in a general non-Abelian background field, not to mention averaging it over an ensemble of configurations.

A decade ago, we suggested a formula for the Wilson loop in a given background belonging to any gauge group and any representation [1]. In this formula, the path ordering along the loop is removed at the price of an additional integration over all gauge transformations of the given non-Abelian background field, or more precisely, over a coset depending on the particular representation in which the Wilson loop is considered. Furthermore, the Wilson loop can be presented in the form of a surface integral [2], see the next section. We call this representation the non-Abelian Stokes theorem. It is quite different from previous interesting statements [3–6] that were also referred to as the "non-Abelian Stokes theorem," but which involved surface ordering. Our formula has no surface ordering. A classification of "non-Abelian Stokes theorems" for arbitrary groups and their representations was recently given by Kondo *et al.* [7] who used the naturally arising techniques of flag manifolds.

Although these formulas do not usually facilitate finding Wilson loops in particular backgrounds, they

¶ This article was submitted by the authors in English.

can be used in averaging Wilson loops over ensembles of Yang–Mills configurations or over different metrics, and in more general settings, see, e.g., [7–11].

The main aim of this paper is to present new formulas for the gravitational holonomies in curved $d = 2, 3, 4$ spaces; these formulas are similar to our non-Abelian Stokes theorem for the Yang–Mills case. We get rid of the path ordering in Eq. (2) and write the holonomies as exponentials of surface integrals. Instead of the path ordering, we have to integrate over certain covariantly unit vectors (for $d = 3$) or covariantly unit (anti)self-dual tensors (for $d = 4$). Remarkably, these formulas put parallel transporters of different spins on the same footing. In particular, holonomies for half-integer spins are presented in terms of the metric tensor (and its derivatives) only, but not in terms of the vierbein or the spin connection.

In addition to a purely theoretical interest, we have a practical motivation in mind. Recently, it was shown, both on the lattice [13, 14] and in the continuum [12, 13], that the $SU(2)$ Yang–Mills partition function in $d = 3$ can be exactly rewritten in terms of local gauge-invariant quantities given by the six components of the dual space metric tensor. This rewriting can be useful in investigating the spectrum and the correlation functions of the theory directly in a gauge-invariant way, but it is insufficient for studying the interactions of external sources because these couple to the Yang–Mills potential and not to gauge-invariant quantities. The present paper demonstrates, however, that a typical source, i.e., the Yang–Mills Wilson loop, can be expressed not only through the potential (or connection) but also through the metric tensor, which is gauge-invariant. Thus, not only the partition function, but also the Wilson loops in the $d = 3$ Yang–Mills theory can be expressed through local gauge-invariant quantities. A detailed formulation of the resulting theory is given elsewhere.

Although the main content of the paper is the non-Abelian Stokes theorems for holonomies in 3 and 4 dimensions, we add three short sections with relevant material. For completeness, we add the Stokes theorem in two dimensions, compute the holonomy in the special case of a constant curvature with a cylinder topology in three dimensions, and give a general formula for the "gravitational Wilson loop" for small loops in any number of dimensions.

2. NON-ABELIAN STOKES THEOREM IN THE YANG–MILLS THEORY

Let τ parametrize the loop defined by the trajectory $x^\mu(\tau)$ and let $A(\tau)$ be the tangent component of the Yang–Mills field along the loop in the fundamental representation of the gauge group,

$$A(\tau) = A_\mu^a t^a \frac{dx^\mu}{d\tau}, \quad \mathrm{Tr}(t^a t^b) = \frac{1}{2}\delta^{ab}.$$

Gauge transformations of $A(\tau)$ are given by

$$A(\tau) \longrightarrow S(\tau)A(\tau)S^{-1}(\tau) + iS(\tau)\frac{d}{d\tau}S^{-1}(\tau). \tag{3}$$

Let H_i be the Cartan subalgebra generators ($i = 1, \ldots, r$, where r is the rank of the gauge group) and the r-dimensional vector $\mathbf{m}$ be the highest weight of the representation r in which the Wilson loop is considered. The formula for the Wilson loop derived in [1] is a path integral over all gauge transformations $S(\tau)$ that are periodic along the contour:

$$W_r = \int DS(\tau) \times \exp\left(i\int d\tau \mathrm{Tr}[m_i H_i (SAS^{-1} + iS\dot{S}^{-1})]\right). \tag{4}$$

We stress that Eq. (4) is manifestly gauge invariant, as is the Wilson loop itself. For example, in the simple case of the $SU(2)$ group, Eq. (4) becomes

$$W_J = \int DS(\tau) \times \exp\left(iJ\int d\tau \mathrm{Tr}[\tau_3 (SAS^\dagger + iS\dot{S}^\dagger)]\right), \tag{5}$$

where τ_3 is the third Pauli matrix and $J = 1/2, 1, 3/2, \ldots$ is the "spin" of the representation of the Wilson loop considered.

The path integrals over all gauge rotations in Eqs. (4) and (5) are not of the Feynman type: they do not contain terms quadratic in the derivatives in τ. Therefore, a certain regularization of these integrals is implied ensuring that $S(\tau)$ is sufficiently smooth. For example, one can introduce quadratic terms in the angular velocities $iS\dot{S}^\dagger$ with small coefficients eventually set equal to zero; see [1] for details. Equation (5) was derived in [1] in two independent ways: (i) by a direct discretization and (ii) by using the standard Feynman representation of path integrals as a sum over all intermediate states, in this case for the axial top supplemented by an action of the "Wess–Zumino" type. Another discretization leading to the same result was recently used by Kondo [7]. A similar formula has been used by Alekseev, Faddeev, and Shatashvili [16] in deriving a formula for group characters to which the Wilson loop is reduced ifor a constant A field (which is the case actually considered in [16]). In [17], Eq. (4) was rederived in an independent way specifically for the fundamental representation of the $SU(N)$ gauge group. Finally, another derivation of a variant of Eq. (5) using lattice regularization was recently given in [18].

The second term in the exponent in Eqs. (4) and (5) is in fact a "Wess–Zumino"-type action, and it can be rewritten not as a line but as a surface integral associated with a closed contour. For simplicity, we consider

the $SU(2)$ gauge group and parametrize the $SU(2)$ matrix S in Eq. (5) by Euler's angles,

$$S = \exp\left(i\frac{\gamma\tau_3}{2}\right)\exp\left(i\frac{\beta\tau_2}{2}\right)\exp\left(i\frac{\alpha\tau_3}{2}\right)$$

$$= \begin{pmatrix} \cos\frac{\beta}{2}\exp\left(i\frac{\alpha+\gamma}{2}\right) & \sin\frac{\beta}{2}\exp\left(-i\frac{\alpha-\gamma}{2}\right) \\ -\sin\frac{\beta}{2}\exp\left(i\frac{\alpha-\gamma}{2}\right) & \cos\frac{\beta}{2}\exp\left(-i\frac{\alpha+\gamma}{2}\right) \end{pmatrix}. \quad (6)$$

The derivation of Eq. (5) implies that $S(\tau)$ is a periodic matrix. This means that $\alpha \pm \gamma$ and β are periodic functions of τ with the period 4π.

The second term in the exponent in Eq. (5), which we denote by Φ, is then

$$\Phi = \int d\tau \mathrm{Tr}(\tau_3 i S\dot{S}^\dagger) = \int d\tau(\dot{\alpha}\cos\beta + \dot{\gamma})$$
$$= \int d\tau[\dot{\alpha}(\cos\beta - 1) + (\dot{\alpha} + \dot{\gamma})] \quad (7)$$
$$= \int d\tau\dot{\alpha}(\cos\beta - 1).$$

The last term is a total derivative and can be actually dropped because $\alpha + \gamma$ is 4π-periodic, and therefore, does not contribute to Eq. (5) even for half-integer representations J. We note that α can be 2π-periodic if γ (which drops from Eq. (7)) is 2π-, 6π-, …-periodic. If $\alpha(1) = \alpha(0) + 2\pi k$, $\alpha(\tau)$ makes k windings. The integration over all possible $\alpha(\tau)$ implied in Eq. (5) can be divided into distinct sectors with different winding numbers k.

Introducing a unit 3-vector

$$n^a = \frac{1}{2}\mathrm{Tr}(S\tau^a S^\dagger\tau_3)$$
$$= (\sin\beta\cos\alpha, \sin\beta\sin\alpha, \cos\beta), \quad (8)$$

we can rewrite Φ as

$$\Phi = \frac{1}{2}\int d\tau d\sigma \epsilon^{abc}\epsilon_{ij} n^a \partial_i n^b \partial_j n^c,$$
$$i, j = \tau, \sigma, \quad (9)$$

where we integrate over any spanning surface for the contour (we call it a "disk"), and $\mathbf{n}$ or α and β are continued to the interior of the disk without singularities. We denote the second coordinate by σ such that $\sigma = 1$ corresponds to the edge of the disk coinciding with the contour and $\sigma = 0$ corresponds to the center of the disk. See [14] for the details on the continuation to the interior of the disk.

We note that if the surface is closed or infinite, the right-hand side of Eq. (9) is the integer topological charge of the $\mathbf{n}$ field on the surface,

$$Q = \frac{1}{8\pi}\int d\sigma d\tau \epsilon^{abc}\epsilon_{ij} n^a \partial_i n^b n^c. \quad (10)$$

Equation (9) can also be rewritten in the form that is invariant under surface reparameterizations. Introducing the invariant surface element

$$d^2 S^{\mu\nu} = d\sigma d\tau\left(\frac{\partial x^\mu}{\partial\tau}\frac{\partial x^\nu}{\partial\sigma} - \frac{\partial x^\nu}{\partial\tau}\frac{\partial x^\mu}{\partial\sigma}\right) = \epsilon^{\mu\nu} d(\mathrm{Area}), \quad (11)$$

we can rewrite Eq. (9) as

$$\Phi = \frac{1}{2}\int d^2 S^{\mu\nu}\epsilon^{abc} n^a \partial_\mu n^b \partial_\nu n^c. \quad (12)$$

For the Wilson loop, we then obtain [1]

$$W_J = \int D\mathbf{n}(\sigma, \tau)$$
$$\times \exp\left[iJ\int d\tau(A^a n^a) + \frac{iJ}{2}\int d^2 S^{\mu\nu}\epsilon^{abs} n^a \partial_\mu n^b \partial_\nu n^c\right]. \quad (13)$$

The interpretation of this formula is obvious: the unit vector $\mathbf{n}$ plays the role of the instant direction of the color "spin" in the color space. However, multiplying its length by J does not guarantee that we deal with a true quantum state from the representation labelled by J; this is achieved only by introducing the "Wess–Zumino" term in Eq. (13) that fixes the representation to which the probe quark of the Wilson loop belongs to be exactly J.

Finally, we can rewrite the exponent in Eq. (13) such that both terms appear to be surface integrals [2],

$$W = \int D\mathbf{n}(\sigma, \tau)$$
$$\times \exp\left[\frac{iJ}{2}\int d^2 S^{\mu\nu}(-F^a_{\mu\nu} n^a + \epsilon^{abc} n^a (D_\mu n)^b (D_\nu n)^c)\right], \quad (14)$$

where

$$D^{ab}_\mu = \partial_\mu\delta^{ab} + \epsilon^{acb} A^c_\alpha$$

is the covariant derivative and

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + \epsilon^{abc} A^b_\mu A^b_\nu$$

is the field strength. Indeed, expanding the exponent in Eq. (14) in powers of A_μ, we observe that the quadratic term cancels while the linear term is a total derivative reproducing the $A^a n^a$ term in Eq. (13); the zero-order term is "Wess–Zumino" term (9) or (7). We note that both terms in Eq. (14) are explicitly gauge invariant. We call Eq. (14) the non-Abelian Stokes theorem. We stress that it is different from the previously proposed Stokes-like representations of the Wilson loop based on ordering elementary surfaces inside the loop [3–6]. For a further discussion of Eq. (14), see [18].

We now briefly discuss gauge groups higher than $SU(2)$: for this purpose, we must return to Eq. (4). It is valid for any group and any representation, however, its surface form depends explicitly on the group representation in which the Wilson loop is considered. Equation (4) says that one can in fact integrate not over all gauge transformations S but only over those that do not commute with the combination of Cartan generators $m_i H_i$, where **m** is the highest weight of a given representation. In the $SU(2)$ case, one has

$$m_i H_i = J\tau_3, \quad J = 1/2, 1, 3/2, \ldots,$$

because $SU(2)$ has the rank 1 and there is only one Cartan generator. In the $SU(2)$ case, one therefore integrates over the coset $SU(2)/U(1)$ for any representation; this coset can be parametrized by the **n** field as described above.

For higher groups, there are several possibilities of taking cosets: a particular coset depends on the representation of the Wilson loop. For example, in the case where the Wilson loop is in the fundamental representation of the $SU(N)$ group, the combination $m_i H_i$ is proportional to one particular generator of the Cartan subalgebra that commutes with the $SU(N-1)\times U(1)$ subgroup. (For $SU(3)$, this generator is the Gell–Mann λ_8 matrix or a permutation of its elements.) For the fundamental representation of the $SU(N)$ group, the appropriate coset is therefore given by

$$SU(N)/SU(N-1)/U(1) = CP^{N-1}.$$

A possible parametrization of this coset is given by a complex N-vector u^α of the unit length, $u^\dagger_\alpha u^\alpha = 1$. To be specific, the Cartan combination in the fundamental representation can always be set equal to

$$m_i H_i = \mathrm{diag}(1, 0, \ldots, 0)$$

by rotating the axes and subtracting the unit matrix. In this basis, u^α is just the first column of the unitary matrix $S^\dagger$ and $u^\dagger_\alpha$ is the first row of S. Unitarity of S implies that

$$u^\dagger_\alpha u^\alpha = 1.$$

In this parametrization, Eq. (4) can be written as

$$W^{SU(N)}_{\mathrm{fund}} = \int DuDu^\dagger \delta(u^\dagger_\alpha u^\alpha - 1) \times \exp i\int d\tau \frac{dx^\mu}{d\tau} u^\dagger_\alpha (i\nabla_\mu)^\alpha_\beta u^\beta, \tag{15}$$

$$(\nabla_\mu)^\alpha_\beta = \partial_\mu \delta^\alpha_\beta - iA^a_\mu (t^a)^\alpha_\beta.$$

Using the identity

$$\epsilon_{ij}\partial_i(u^\dagger \nabla_i u) = \epsilon_{ij}[(\nabla_i u)^\dagger(\nabla_j u) + u^\dagger \nabla_i \nabla_j u] = \epsilon_{ij}\left[-\frac{i}{2}(u^\dagger F_{ij} u) + (\nabla_i u)^\dagger(\nabla_j u)\right], \tag{16}$$

we can present Eq. (15) in a surface form,

$$W^{SU(N)}_{\mathrm{fund}} = \int DuDu^\dagger \delta(|u|^2 - 1) \times \exp\left(i\int dS^{\mu\nu}\left[\frac{1}{2}(u^\dagger F_{\mu\nu} u) + i(\nabla_\mu u)^\dagger(\nabla_\nu u)\right]\right), \tag{17}$$

where $F_{\mu\nu}$ is the field strength in the fundamental representation. Equation (17) was first published in [17], however, with an unexpected overall coefficient 2 in the exponent. Equation (17) presents the non-Abelian Stokes theorem for the Wilson loop in the fundamental representation of $SU(\mathrm{N})$. In the particular case of the $SU(2)$ group, transition to Eq. (14) is achieved by identifying the unit 3-vector

$$n^a = u^\dagger_\alpha (\tau^a)^\alpha_\beta u^\beta,$$

where

$$u^\alpha = \begin{pmatrix} \cos\frac{\beta}{2}\exp\left(-i\frac{\alpha+\gamma}{2}\right) \\ \sin\frac{\beta}{2}\exp\left(i\frac{\alpha-\gamma}{2}\right) \end{pmatrix}, \tag{18}$$

$$2iu^\dagger \partial_\tau u = \dot\alpha(\cos\beta - 1) + (\dot\alpha + \dot\gamma).$$

It must be mentioned that the quantity

$$\int d\sigma d\tau \epsilon_{ij} i\partial_i u^\dagger_\alpha \partial_j u^\alpha = 2\pi Q \tag{19}$$

appearing in Eq. (17) is the topological charge of the 2-dimensional CP^{N-1} model. For closed or infinite surfaces, Q is an integer.

In the case where the Wilson loop is taken in the adjoint representation of the $SU(N)$ gauge group, the combination $m_i H_i$ in Eq. (4) is the highest root. Only group elements of the form $\exp(i\alpha_i H_i)$ commute with this combination (these elements belong to the maximum torus subgroup $U(1)^{N-1}$). In the case of the adjoint representation, one therefore integrates over the flag manifold [19, 7]

$$SU(N)/U(1)^{N-1} = F^{N-1}.$$

3. "GRAVITATIONAL WILSON LOOPS"

An object similar to the Wilson loop of the Yang–Mills theory also exists in gravity theory. It is the parallel transporter of a vector on a Riemannian manifold along a closed contour, also called a holonomy. The holonomy is trivial if the space is flat but becomes a nontrivial functional of the curvature if it is nonzero. In the remaining sections, we present new formulas for the parallel transporters on $d = 2, 3, 4$ Riemannian manifolds.

We first recall some notation from differential geometry. We use [20] as a general reference book. Let

$g_{\mu\nu} = g_{\nu\mu}$ ($\mu, \nu = 1, \ldots, d$) be the covariant metric tensor, with the contravariant tensor $g^{\mu\nu}$ being its inverse, $g_{\mu\nu}g^{\nu\kappa} = \delta_\mu^\kappa$. The determinant of the covariant metric tensor is denoted by g. The Christoffel symbol is defined as

$$\Gamma^\mu_{\nu\kappa} = g^{\mu\lambda}\Gamma_{\lambda,\nu\kappa} = \frac{g^{\mu\lambda}}{2}(\partial_\nu g_{\lambda\kappa} + \partial_\kappa g_{\lambda\nu} - \partial_\lambda g_{\nu\kappa}),$$
$$\Gamma^\kappa_{\nu\kappa} = \frac{\partial_\nu g}{2g}. \tag{20}$$

The action of the covariant derivative on a contravariant vector is defined as

$$(\nabla_\rho)^\kappa_\lambda v^\lambda = (\partial_\rho \delta^\kappa_\lambda + \Gamma^\kappa_{\rho\lambda})v^\lambda. \tag{21}$$

The commutator of two covariant derivatives determines the Riemann tensor,

$$[\nabla_\rho \nabla_\sigma]^\kappa_\lambda = R^\kappa_{\lambda\rho\sigma} = g^{\kappa\kappa'}R_{\kappa'\lambda\rho\sigma}$$
$$= \partial_\rho\Gamma^\kappa_{\sigma\lambda} - \partial_\sigma\Gamma^\kappa_{\rho\lambda} + \Gamma^\kappa_{\rho\tau}\Gamma^\tau_{\sigma\lambda} - \Gamma^\kappa_{\sigma\tau}\Gamma^\tau_{\rho\lambda}. \tag{22}$$

A contraction of the Riemann tensor gives the symmetric Ricci tensor,

$$R_{\lambda\sigma} = R^\kappa_{\lambda\kappa\sigma}, \quad R^\kappa_\rho = R^\kappa_{\lambda\rho\sigma}g^{\lambda\sigma}. \tag{23}$$

Its full contraction is the scalar curvature

$$R = R_{\lambda\sigma}g^{\lambda\sigma} = R^\kappa_\kappa. \tag{24}$$

The parallel transporter of a contravariant vector along a curve $x^\mu(\tau)$ is determined by solving the equation

$$\frac{dx^\mu}{d\tau}(\nabla_\mu)^\kappa_\lambda v^\lambda(\tau) = 0. \tag{25}$$

The solution can be written using the evolution operator

$$v^\kappa(\tau) = [W^G(\tau)]^\kappa_\lambda v^\lambda(0), \tag{26}$$

where $v^\lambda(0)$ is the vector at the starting point of the contour and $v^\lambda(\tau)$ is the parallel-transported vector at the point labelled by τ. The evolution operator can be symbolically written as a path-ordered exponential of the Christoffel symbol,

$$[W^G(\tau)]^\kappa_\lambda = \left[\mathrm{P}\exp\left(-\int_0^\tau d\tau \frac{dx^\mu}{d\tau}\Gamma_\mu\right)\right]^\kappa_\lambda. \tag{27}$$

We define the "gravitational Wilson loop" as the trace of the parallel transporting evolution operator along the closed curve $x^\mu(\tau)$ with $x^\mu(1) = x^\mu(0)$,

$$W^G_{\text{vector}} = \frac{1}{d}[W^G(1)]^\kappa_\kappa. \tag{28}$$

This quantity is diffeomorphism-invariant: the metric tensor is transformed under coordinate changes $x^\mu \longrightarrow x'^\mu(x)$, but if the contour is changed, as

$$x^\mu(\tau) \longrightarrow x'^\mu(x(\tau)),$$

the gravitational Wilson loop or the holonomy remains the same. In this respect, the gravitational holonomy is different from the Yang–Mills loop which is invariant under gauge transformations without changing the contour.

The parallel transporter of a covariant vector is given by the transposed matrix; its trace coincides with that of the matrix used in transporting contravariant vectors.

4. RELATION OF GRAVITY QUANTITIES TO THOSE OF THE YANG–MILLS THEORY

We now show that the "gravitational Wilson loop" is not only analogous to, but directly expressible through, the Yang–Mills Wilson loops of the $SU(2)$ group. For this purpose, we introduce the standard vierbein e^A_μ and its inverse $e^{A\mu}$ such that

$$e^A_\mu e^A_\nu = g_{\mu\nu}, \quad e^A_\mu e^{B\mu} = \delta^{AB},$$
$$e^{A\mu}e^{A\nu} = g^{\mu\nu}, \quad \det e^A_\mu = \sqrt{g}. \tag{29}$$

We decompose the vector experiencing the parallel transport in vierbeins, $v^\lambda = c^A e^{A\lambda}$, with the reciprocal decomposition

$$c^A = e^A_\kappa v^\kappa, \tag{30}$$

and insert this in Eq. (25) defining the parallel transport. We then have

$$0 = \frac{dx^\mu}{d\tau}(\nabla_\mu)^\kappa_\lambda c^A e^{A\lambda}$$
$$= \frac{dx^\mu}{d\tau}[e^{A\kappa}\partial_\mu c^A + c^A(\partial_\mu e^{A\kappa} + \Gamma^\kappa_{\mu\lambda}e^{A\lambda})] \tag{31}$$
$$= \frac{dx^\mu}{d\tau}e^{B\kappa}(\partial_\mu\delta^{BA} + \omega^{BA}_\mu)c^A,$$

where we introduced the spin connection

$$\omega^{AB}_\mu = -\omega^{BA}_\mu = \frac{1}{2}e^{A\kappa}(\partial_\mu e^B_\kappa - \partial_\kappa e^B_\mu)$$
$$-\frac{1}{2}e^{B\kappa}(\partial_\mu e^A_\kappa - \partial_\kappa e^A_\mu) \tag{32}$$
$$-\frac{1}{2}e^A_\kappa e^{B\lambda}e^C_\lambda(\partial_\kappa e^C_\lambda - \partial_\lambda e^C_\kappa)$$

and used the fundamental relations

$$\partial_\mu e^{A\kappa} + \Gamma^\kappa_{\mu\lambda}e^{A\lambda} = -\omega^{AB}_\mu e^{B\kappa}, \tag{33}$$

$$\partial_\mu e^{A\kappa} - \Gamma^\lambda_{\mu\kappa}e^A_\lambda = -\omega^{AB}_\mu e^B_\kappa. \tag{34}$$

One can introduce the $SO(d)$ "field strength"

$$\begin{aligned}\mathcal{F}_{\mu\nu}^{AB} &= [\partial_\mu + \omega_\mu, \partial_\nu + \omega_\nu]^{AB} \\ &= \partial_\mu\omega_\nu^{AB} - \partial_\nu\omega_\mu^{AB} + \omega_\mu^{AC}\omega_\nu^{CB} - \omega_\nu^{AC}\omega_\mu^{CB}\end{aligned} \tag{35}$$

related to the Riemann tensor as

$$\begin{aligned}\mathcal{F}_{\mu\nu}^{AB} e_\kappa^A e_\lambda^B &= -R_{\kappa\lambda\mu\nu}, \\ \mathcal{F}_{\mu\nu}^{AB} &= -R_{\kappa\lambda\mu\nu} e^{A\kappa} e^{B\lambda}, \\ \mathcal{F}_{\mu\nu}^{AB} e^{A\mu} e^{B\nu} &= R.\end{aligned} \tag{36}$$

The above material is common for any number of dimensions. To proceed further, we consider the cases where $d = 3$ and $d = 4$ separately. The case where $d = 2$ is considered in Section 6.

4.1. d = 3

In three dimensions, one can immediately identify the spin connection with the $SU(2)$-valued Yang–Mills field as

$$A_i^c = -\frac{1}{2}\epsilon^{abc}\omega_i^{ab}. \tag{37}$$

Working in three dimensions, we denote the Lorentz indices by $i, j, \ldots = 1, 2, 3$ and the flat triade indices by $a, b, \ldots = 1, 2, 3$. Recalling the generators in the $J = 1$ representation,

$$(T^c)^{ab} = -i\epsilon^{cab}, \quad [T^c T^d] = i\epsilon^{cdf} T^f, \tag{38}$$

we can rewrite the last parenthesis in Eq. (31) as

$$\partial_i\delta^{ab} + \omega_i^{ab} = \partial_i\delta^{ab} - iA_i^c (T^c)^{ab} \equiv (D_i)^{ab}, \tag{39}$$

which is the standard Yang–Mills covariant derivative in the adjoint representation. In the fundamental (spinor) representation, the Yang–Mills covariant derivative is

$$\begin{aligned}(\nabla_i)_\beta^\alpha &= \partial_i\delta_\beta^\alpha - iA_i^c\left(\frac{\sigma^c}{2}\right)_\beta^\alpha = \partial_i\delta_\beta^\alpha + \frac{1}{8}\omega_i^{ab}[\sigma^a\sigma^b]_\beta^\alpha, \\ &\alpha, \beta = 1, 2,\end{aligned} \tag{40}$$

which coincides with the known expression for the covariant derivative in the spinor representation in a curved space.

The standard Yang–Mills field strength is directly related to that in Eq. (35),

$$F_{ij}^a = \partial_i A_j^a - \partial_j A_i^a + \epsilon^{abc} A_i^b A_j^c = -\frac{1}{2}\epsilon^{abc}\mathcal{F}_{ij}^{bc}. \tag{41}$$

It follows from Eq. (36) that

$$\epsilon^{abc} F_{ij}^a e_k^b e_l^c = R_{ijkl}. \tag{42}$$

We next consider the parallel transporter of a 3-vector in a curved space, as defined by Eq. (25). In accordance with Eqs. (31) and (39), solving Eq. (25) is equivalent to solving the Yang–Mills equation for the parallel transporter,

$$\frac{dx^i}{d\tau}(D_i)^{ab} c^b = 0, \tag{43}$$

whose solution is

$$\begin{aligned}c^a(\tau) &= [W_1^{YM}(\tau)]^{ab} c^b(0), \\ [W_1^{YM}(\tau)]^{ab} &= \left[\mathrm{P}\exp\left(i\int d\tau \frac{dx^i}{d\tau} A_i^c T^c\right)\right]^{ab},\end{aligned} \tag{44}$$

where the subscript "1" indicates that the path-ordered exponential is taken in the $J = 1$ representation. The parallel transport of a contravariant vector is therefore given by

$$\begin{aligned}v^k(\tau) &= c^a(\tau) e^{ak}(\tau) \\ &= e^{ak}(\tau)[W_1^{YM}(\tau)]^{ab} e_l^b(0) v^l(0),\end{aligned} \tag{45}$$

which immediately implies the sought relation between the "gravitational" and Yang–Mills parallel transporters,

$$[W_1^G(\tau)]_l^k = e_k^a(\tau)[W_1^{YM}(\tau)]^{ab} e^{bl}(0). \tag{46}$$

The relation becomes especially neat for the Wilson loops, i.e., for the traces of parallel transporters along closed contours. Because the vierbeins take identical values at the end points of a closed contour, $e_k^a(1) = e_k^a(0)$, we obtain

$$W_{\text{vector}}^G = \frac{1}{3}[W_1^G]_k^k = \frac{1}{3}[W_1^{YM}]^{aa} = W_1^{YM}. \tag{47}$$

In a similar way, one can show that the same equation is valid for the gravitational parallel transporter of covariant vectors and, more generally, for parallel transporters of any integer spin J. In this case, the Yang–Mills Wilson loop must be taken in the same representation as the gravitational one,

$$W_J^G = W_J^{YM}. \tag{48}$$

It is understood that the right-hand side of Eq. (48) is expressed through the Yang–Mills field equal to the spin connection in accordance with Eq. (37), while the left-hand side is expressed through the Christoffel symbols, that is, through the metric. It must be stressed that the spin connection is defined via the vierbein, which is not uniquely determined by the metric tensor. The Wilson loop, being a gauge-invariant quantity, is nevertheless uniquely determined by the metric tensor and its derivatives. This is the meaning of Eq. (48).

For a half-integer J, there is no way to define the parallel transporter other than through the spin connection. Nevertheless, as we show in Section 8, where we present the holonomy for any spin in a surface form, the "gravitational Wilson loop" is also expressible through the metric tensor and its derivatives, even for half-integer spins.

4.2. $d = 4$

In four Euclidean dimensions, the rotation group is $SO(4)$, with its algebra isomorphic to that of $SU(2) \times SU(2)$, and therefore, all irreducible representations of $SO(4)$ can be classified by (J_1, J_2), where $J_{1,2} = 0, 1/2, 1, \ldots$ label the representations of the two $SU(2)$ subgroups. For example, the 4-vector representation whose parallel transporter was considered in the beginning of this section transforms in the (1/2, 1/2) representation of $SU(2)$ x $SU(2)$. Because of this, it is convenient to decompose the spin connection ω_μ^{AB} into self-dual and anti-self-dual parts using 't Hooft's η and $\bar\eta$ symbols

$$\eta^{aAB} = \frac{1}{2i}\mathrm{Tr}\sigma^a(\sigma^{A+}\sigma^{B-} - \sigma^{B+}\sigma^{A-}), \qquad \sigma^{A\pm} = (\pm i\sigma, 1), \tag{49}$$

$$\bar\eta^{aAB} = \frac{1}{2i}\mathrm{Tr}\sigma^a(\sigma^{A-}\sigma^{B+} - \sigma^{B-}\sigma^{A+}). \tag{50}$$

We use the capital Latin characters to denote flat 4-dimensional vierbein indices, $A, B, \ldots = 1, 2, 3, 4$, while $a, b, \ldots = 1, 2, 3$; σ^a are the three Pauli matrices. The spin connection ω_μ^{AB} transforms in the 6-dimensional representation of $SO(4)$, which can be decomposed into the sum (1, 0) + (0, 1) of the adjoint representations of the two $SU(2)$ subgroups. We write

$$\omega_\mu^{AB} = -\frac{1}{2}\pi_\mu^a\eta^{aAB} - \frac{1}{2}\rho_\mu^a\bar\eta^{aAB}. \tag{51}$$

The $SO(4)$ "field strength" in Eq. (35) is then decomposed as

$$\mathcal{F}_{\mu\nu}^{AB} = -\frac{1}{2}F_{\mu\nu}^a(\pi)\eta^{aAB} - \frac{1}{2}F_{\mu\nu}^a(\rho)\bar\eta^{aAB}, \tag{52}$$

where

$$F_{\mu\nu}^a(\pi) = \partial_\mu\pi_\nu^a - \partial_\nu\pi_\mu^a + \epsilon^{abc}\pi_\mu^b\pi_\nu^c, \tag{53}$$

$$F_{\mu\nu}^a(\rho) = \partial_\mu\rho_\nu^a - \partial_\nu\rho_\mu^a + \epsilon^{abc}\rho_\mu^b\rho_\nu^c \tag{54}$$

are the usual Yang–Mills field strengths of the $SU(2)$ Yang–Mills potentials π_μ^a and ρ_μ^a. We stress that $6 \times 4 = 24$ variables ω_μ^{AB} equivalent to $2 \times 3 \times 4 = 24$ variables π_μ^a and ρ_μ^a are defined by only $4 \times 4 = 16$ tetrades e_μ^A via Eq. (32), and therefore, not all of them are independent.

Contracting Eq. (36) with the η and $\bar\eta$ symbols, we obtain

$$F_{\mu\nu}^a(\pi) = \frac{1}{2}\eta^{aAB}e^{A\kappa}e^{B\lambda}R_{\kappa\lambda\mu\nu}, \tag{55}$$

$$F_{\mu\nu}^a(\rho) = \frac{1}{2}\bar\eta^{aAB}e^{A\kappa}e^{B\lambda}R_{\kappa\lambda\mu\nu}. \tag{56}$$

We now return to the parallel transporter of a 4-vector. As shown in the beginning of this section, finding this parallel transporter is equivalent to solving the equation

$$\frac{dx^\mu}{d\tau}(\partial_\mu\delta^{AB} + \omega_\mu^{AB})c^B = 0. \tag{57}$$

We represent the 4-vector c^A as a combination of two spinors,

$$c^A = \chi_a^\dagger(\sigma^{A+})_\beta^\alpha\psi^\beta, \quad \chi_\alpha^\dagger\psi^\beta = \frac{1}{2}c^A(\sigma^{A-})_\alpha^\beta, \qquad \alpha, \beta = 1, 2. \tag{58}$$

Inserting this in Eq. (57) and decomposing ω_μ^{AB} as in Eq. (51), we obtain

$$\frac{dx^\mu}{d\tau}\Big\{\partial_\mu[\chi^\dagger\sigma^{A+}\psi] - \frac{1}{2}(\pi_\mu^a\eta^{aAB} + \rho_\mu^a\bar\eta^{aAB})[\chi^\dagger\sigma^{B+}\psi]\Big\} = 0. \tag{59}$$

Using the definition of the η-symbols in Eqs. (49) and (50), it is easy to verify that this equation is satisfied provided the spinors χ and ψ satisfy

$$\frac{dx^\mu}{d\tau}\left[\partial_\mu\delta_\beta^\alpha - i\pi_\mu^a\left(\frac{\sigma^a}{2}\right)_\beta^\alpha\right]\chi^\beta = 0$$
$$\text{or } \frac{dx^\mu}{d\tau}\chi_\alpha^\dagger\left[\overleftarrow{\partial}_\mu\delta_\beta^\alpha + i\pi_\mu^a\left(\frac{\sigma^a}{2}\right)_\beta^\alpha\right] = 0, \tag{60}$$

$$\frac{dx^\mu}{d\tau}\left[\partial_\mu\delta_\beta^\alpha - i\rho_\mu^a\left(\frac{\sigma^a}{2}\right)_\beta^\alpha\right]\psi^\beta = 0. \tag{61}$$

The expressions in square brackets are identical to the Yang–Mills covariant derivatives, with the role of the Yang–Mills potentials played by π_μ^a and ρ_μ^a, respectively. Equations (60) and (61) define the Yang–Mills parallel transporters in the fundamental representation. Their solution can be written as evolution operators,

$$\chi^\alpha(\tau) = [W^\pi(\tau)]_\gamma^\alpha\chi^\gamma(0)$$
$$\text{or } \chi_\alpha^\dagger(\tau) = \chi_\gamma^\dagger(0)[W^{\pi\dagger}(\tau)]_\alpha^\gamma, \tag{62}$$

$$\psi^\beta(\tau) = [W^\rho(\tau)]_\delta^\beta\psi^\delta(0), \tag{63}$$

$$[W^{\pi}(\tau)]^{\alpha}_{\gamma} = \left[\mathrm{P}\exp\left(i\int d\tau \frac{dx^{\mu}}{d\tau}\pi^{a}_{\mu}\frac{\sigma^{a}}{2}\right)\right]^{\alpha}_{\gamma}, \quad (64)$$

$$[W^{\rho}(\tau)]^{\alpha}_{\gamma} = \left[\mathrm{P}\exp\left(i\int d\tau \frac{dx^{\mu}}{d\tau}\rho^{a}_{\mu}\frac{\sigma^{a}}{2}\right)\right]^{\alpha}_{\gamma}. \quad (65)$$

Returning to the 4-vector c^A in Eq. (58), we see that its evolution is determined by

$$\begin{aligned} c^{A}(\tau) &= [W_{\mathrm{vector}}(\tau)]^{AB}c^{B}(0), \\ [W_{\mathrm{vector}}(\tau)]^{AB} &= \frac{1}{2}\mathrm{Tr}[W^{\pi\dagger}(\tau)\sigma^{A+}W^{\rho}(\tau)\sigma^{B-}]. \end{aligned} \quad (66)$$

We now choose a closed contour and take the trace of the evolution operator. The "gravitational Wilson loop" for a 4-vector is then given by

$$\begin{aligned} W^{G}_{\left(\frac{1}{2},\frac{1}{2}\right)} &= \frac{1}{4}e^{A\kappa}(1)[W_{\mathrm{vector}}(1)]^{AB}e^{B}_{\kappa}(0) \\ &= \frac{1}{4}[W_{\mathrm{vector}}(1)]^{AA} = \frac{1}{2}\mathrm{Tr}\,W^{\pi}\cdot\frac{1}{2}\mathrm{Tr}\,W^{\rho}. \end{aligned} \quad (67)$$

Its generalization to the holonomy in an arbitrary representation (J_1, J_2) is obvious,

$$\begin{aligned} W^{G}_{(J_1,J_2)} &= W^{\pi}_{J_1}W^{\rho}_{J_2}, \\ W^{\pi,\rho}_{J} &= \frac{1}{2J+1}\mathrm{Tr}_{(2J+1)}W^{\pi,\rho}. \end{aligned} \quad (68)$$

Thus, the holonomy in the (J_1, J_2) representation in a curved $d = 4$ space is equal to the product of two Yang–Mills Wilson loops, with the role of the Yang–Mills potentials played by the self-dual π^{a}_{μ} and anti-self-dual ρ^{a}_{μ} parts of the spin connection. In Section 9, we show that both W^{π} and W^{ρ} can be written in terms of the metric tensor.

5. SMALL WILSON LOOPS

For small-area contours, the "gravitational Wilson loop" can be expanded in powers of the area. The most straightforward way to do this is to use the path-ordered form of W^G in Eq. (27). We take a square contour of the size $a \times a$ lying in the 12 plane and expand the path-ordered exponential in powers of a. After some simple algebra, we obtain the first nontrivial term of this expansion, which happens to be $O(a^4)$,

$$\begin{aligned} W^{G}_{\mathrm{vector}} &= \frac{1}{d}[W^{G}_{\mathrm{vector}}]^{\kappa}_{\kappa} = 1 + \frac{a^4}{d}R^{\kappa}_{\lambda 12}R^{\lambda}_{\kappa 12} \\ &= 1 - \frac{2(\Delta S)^{\mu\nu}(\Delta S)^{\mu'\nu'}}{4d}R_{\kappa\lambda\mu\nu}R_{\rho\sigma\mu'\nu'}g^{\kappa\rho}g^{\lambda\sigma}, \end{aligned} \quad (69)$$

where $(\Delta S)^{\mu\nu}$ is the surface element lying in the $\mu\nu$ plane. We note that the first correction to the holonomy is negative-definite. We emphasize that the first-order term in ΔS is in general present in the expansion of the parallel transporter; however, it vanishes after taking the trace owing to the identity $R^{\kappa}_{\kappa\mu\nu} \equiv 0$, and therefore, the expansion of the trace starts with the $(\Delta S)^2$ term.

In three dimensions, Eq. (69) can be further simplified because the Riemann tensor is expressed through the Ricci tensor via

$$\begin{aligned} R_{ijkl} &= R_{ik}g_{jl} - R_{il}g_{jk} + R_{jl}g_{ik} \\ &\quad - R_{jk}g_{il} + \frac{R}{2}(g_{il}g_{jk} - g_{ik}g_{jl}). \end{aligned} \quad (70)$$

Because the Riemann tensor is antisymmetric with respect to each pair of subscripts, we can replace

$$\begin{aligned} g^{km}g^{ln} &\longrightarrow \frac{1}{2}(g^{km}g^{ln} - g^{kn}g^{lm}) \\ &= \frac{1}{2g}\epsilon^{kli}\epsilon^{mnj}g_{ij}. \end{aligned} \quad (71)$$

Introducing the dual surface element

$$\Delta S^{pq} = \epsilon^{pqr}\Delta S_r, \quad (72)$$

we have

$$\epsilon^{kli}\epsilon^{pqr}R_{klpq} = -4\left(R^{ir} - \frac{1}{2}Rg^{ir}\right), \quad (73)$$

which as a matter of fact is the Einstein tensor. For the parallel transporter of an arbitrary spin J, the factor 2 in the numerator of Eq. (69) must be replaced by $J(J + 1)$.

Combining all the factors, we obtain

$$\begin{aligned} W^{G}_{J} &= 1 - \frac{2J(J+1)}{3g}\left(R^{ir} - \frac{1}{2}Rg^{ir}\right) \\ &\quad \times g_{ij}\left(R^{js} - \frac{1}{2}Rg^{js}\right)\Delta S_r\Delta S_s. \end{aligned} \quad (74)$$

This is our final expression for the trace of the spin-J parallel transporter for small loops in a curved $d = 3$ space. We note that Eq. (74) is invariant under diffeomorphisms.

6. GRAVITATIONAL WILSON LOOP IN TWO DIMENSIONS

In a curved $d = 2$ space, the trace of the parallel transporter along a closed loop can be computed exactly for any metric and can be presented in the form of a "Stokes theorem". The result is related to the Gauss–Bonnet theorem and is generally known: we present it here for the sake of completeness.

The key observation is that in two dimensions, spin connection (32) has only one component,

$$\omega_i^{ab} = \epsilon^{ab}\omega_i. \tag{75}$$

In this section, all indices take only two values 1, 2. In accordance with Eq. (31), the parallel transporter of a vector is determined by the equation

$$\frac{dc^a}{d\tau} - \frac{dx^i}{d\tau}\omega_i\epsilon^{ab}c^b = 0, \tag{76}$$

which is solved by

$$c^a(\tau) = W^{ab}(\tau)c^b(0),$$
$$W^{ab}(\tau) = \begin{pmatrix} \cos\gamma(\tau) & \sin\gamma(\tau) \\ -\sin\gamma(\tau) & \cos\gamma(\tau) \end{pmatrix}, \tag{77}$$
$$\gamma(\tau) = \int_0^\tau d\tau\frac{dx^i}{d\tau}\omega_i.$$

According to the general theorem in Section 4, the gravitational Wilson loop is equal to the Yang–Mills one, and we obtain

$$W_1^G = \frac{1}{2}W^{aa}(1) = \cos\Phi, \tag{78}$$

where

$$\Phi = \gamma(1) = \int_0^1 d\tau\frac{dx^i}{d\tau}\omega_i = \frac{1}{2}\oint dx^i\epsilon_{ab}\omega_i^{ab}. \tag{79}$$

This formula is not fully satisfactory because the holonomy is expressed through the spin connection and not through the metric. Expressing it through the metric can be achieved if we apply the Stokes theorem and write Eq. (79) in a surface form. We have

$$\Phi = \frac{1}{2}\int dS\epsilon_{ab}\epsilon^{ij}\partial_i\omega_j^{ab}, \tag{80}$$

where dS is the element of the spanning surface for the contour. Introducing the field strength related to the Riemann tensor,

$$F_{ij}^{ab} = \partial_i\omega_j^{ab} - \partial_j\omega_j^{ab} + \omega_i^{ac}\omega_j^{cb} - \omega_j^{ac}\omega_i^{cb}$$
$$= R_{ij}^{kl}e_k^a e_l^b, \tag{81}$$
$$\epsilon_{ab}e_k^a e_l^b = \epsilon_{kl}\sqrt{g},$$

and noticing that the commutator term is zero in two dimensions, we rewrite Eq. (80) as

$$\Phi = \frac{1}{2}\int dS\sqrt{g}R, \quad W_1^G = \cos\Phi, \tag{82}$$

where

$$R = (1/2)\epsilon^{ij}\epsilon_{kl}R_{ij}^{kl}$$

is the scalar curvature. It is gratifying that the holonomy is expressed through the Einstein-Hilbert action, which is known to be a total derivative in two dimensions. Needless to explain, Eq. (82) is diffeomorphism-invariant.

In two dimensions, there is essentially only one component of the Riemann tensor,

$$R_{1212} = \frac{1}{2}Rg \tag{83}$$

(see [20]). Taking this into account, it is easy to verify that for small areas, the expansion of Eq. (82) gives the same result as Eq. (69) written for small loops.

7. AN EXAMPLE OF BIG LOOPS: A CONSTANT-CURVATURE BACKGROUND IN THREE DIMENSIONS

In three dimensions, the Riemann tensor is expressible through the Ricci tensor, see Eq. (70). Therefore, the diffeomorphism-invariant information about curved spaces is fully contained in the three eigenvalues of the symmetric Ricci tensor,

$$R_j^i = \lambda\delta_j^i, \tag{84}$$

with the scalar curvature being the sum of the three,

$$R = \lambda_1 + \lambda_2 + \lambda_3.$$

For example, the de Sitter S^3 space corresponds to

$$\lambda_1 = \lambda_2 = \lambda_3 = R/3 = \text{const}.$$

In this section, we consider another constant-curvature case, namely, the cylinder space $S^2 \times R$ characterized by

$$\lambda_1 = \lambda_2 = R/2 = \text{const}, \quad \lambda_3 = 0.$$

We show that the parallel transporter in these spaces can be computed for any form of the contour and any metric and that the gravitational Wilson loop is given by an elegant formula.

A general metric can be considered as the one induced by six external coordinates $w^A(x_1, x_2, x_3)$,

$$g_{ij} = \partial_i w^A\partial_j w^A, \quad A = 1, \ldots, 6. \tag{85}$$

In the special case of the cylinder space $S^2 \times R$, it is sufficient to use only four external coordinates w^a ($a = 1, 2, 3$) and w^4 subject to the constraint

$$\sum_{a=1}^{3}(w^a)^2 = \frac{2}{R}. \tag{86}$$

An example of such external coordinates is given by

$$w^{1,2,3}(x) = \sqrt{\frac{2}{R}}\frac{x^{1,2,3}}{r}, \quad w^4(x) = \sqrt{\frac{2}{R}}\ln r, \tag{87}$$

leading to the metric tensor

$$g_{ij} = \frac{2}{R}\frac{1}{r^2}\delta_{ij}, \quad \sqrt{g} = \left(\frac{2}{R}\right)^{3/2}\frac{1}{r^3}. \tag{88}$$

A simple calculation using formulas from Section 3 shows that this metric indeed gives a zero eigenvalue of the Ricci tensor with the other two eigenvalues equal to the constant $R/2$. Because the eigenvalues of the Ricci tensor are diffeomorphism-invariant, a general change of coordinates $x^i \longrightarrow y^i(x)$ in Eq. (87) results in the same eigenvalues. Therefore, the most general description of the cylinder spaces $S^2 \times R$ is given by

$$w^a(x) = \sqrt{\frac{2}{R}}\frac{y^a(x)}{|y(x)|}, \quad w^4(x) = \sqrt{\frac{2}{R}}\ln|y(x)|,$$
$$g_{ij} = \frac{2}{R}\frac{\partial_i y^a \partial_i y^a}{y^2}, \tag{89}$$

$$\sqrt{g} = \left(\frac{2}{R}\right)^{\frac{3}{2}}\frac{1}{3!}\epsilon^{ijk}\epsilon_{abc}\frac{\partial_i y^a \partial_j y^b \partial_k y^c}{|y|^3}$$
$$= \frac{1}{2}\sqrt{\frac{R}{2}}\epsilon^{ijk}\epsilon_{abc}\partial_i w^a \partial_j w^b w^c \partial_k w^4, \tag{90}$$

where $y^a(x)$ are three arbitrary functions of the coordinates x^i. We note that g_{ij} is given by the product of two matrices

$$M_i^a = \partial_i y^a/|y|,$$

and hence, $\sqrt{g}$ is itself a determinant (of the matrix M).

Our aim is to calculate the Wilson loop for any contour in any metric (89) corresponding to the cylinder spaces. We use the diffeomorphism invariance of the Wilson loop. If we compute it for a general contour in some metric representing cylinder spaces, the most general case is recovered by diffeomorphisms of both the contour and the metric. We start with the specific metric given by Eqs. (87) and (88).

Given metric tensor (88), we construct a vierbein corresponding to it. This is, of course, not unique but any choice of the vierbein suits us. We choose

$$e_i^a = \sqrt{\frac{2}{R}}\frac{1}{r}\delta_i^a, \quad e_i^a e_j^a = g_{ij}. \tag{91}$$

Given the vierbein, we construct the spin connection (or the Yang–Mills field) from its definition (32) and obtain

$$A_i^a = -\frac{1}{2}\epsilon^{abc}\omega_i^{bc} = \epsilon^{aij}\frac{x^j}{r^2}, \tag{92}$$

which happens to be the field of the Wu–Yang monopole; the scalar curvature R has dropped from the spin connection. According to the theorem in Section 4, the gravitational Wilson loop is equal to the Yang–Mills Wilson loop, provided the Yang–Mills potential A_i^a is the spin connection of the metric under consideration. Therefore, all we have to do is to compute the Wilson loop for a general contour in the field of the Wu-Yang monopole.

This task is easily solvable if we use another invariance, the gauge invariance of the Wilson loop. It is well known that the Wu–Yang monopole in hedgehog gauge (92) can be transformed to the string gauge where the potential has only one nonzero component along the third color axis (plus a Dirac string). In this gauge, the Yang–Mills potential is basically Abelian, and the Wilson loop in any representation J is therefore given by

$$W_J^G = W_J^{YM} = \frac{1}{2J+1}\sum_{m=-J}^{J}\exp(im\Phi),$$
$$\Phi = \oint dx^i A_i^3 = \int dS_i \frac{x^i}{r^3}. \tag{93}$$

In the last equation, we used the normal Stokes theorem for the circulation and also used the fact that in the string gauge, the magnetic field of the monopole is the Coulomb field of a point charge; dS_i is the element of the spanning surface for the contour and is orthogonal to the surface.

Equation (93) is the gravitational Wilson loop for arbitrary contours but in a specific metric given by Eq. (88). To generalize it to the general metric given by (89), it only remains to perform the general coordinate transformation of Eq. (93). To this end, it is convenient to use, instead of dS_i, its dual dS^{ij} such that $dS_i = \epsilon_{ijk}dS^{jk}$. We recall that under a general coordinate transformation $x^i \longrightarrow y^i(x)$, the contravariant vector transforms as

$$V^i \longrightarrow V^k \partial_k y^i$$

and the antisymmetric contravariant tensor transforms as

$$dS^{ij} \longrightarrow dS^{mn}\partial_m y^i \partial_n y^j.$$

The flux in Eq. (93) is therefore given by

$$\Phi = \int dS_i \frac{x^i}{r^3} = \int dS^{ij}\epsilon_{ijk}\frac{x^k}{r^3} \longrightarrow$$
$$\longrightarrow \int dS^{mn}\frac{\epsilon_{ijk}\partial_m y^i \partial_n y^j y^k}{|y|^3}. \tag{94}$$

This equation takes a more symmetric form in terms of external coordinates (89),

$$\Phi = \left(\frac{2}{R}\right)^{\frac{2}{3}}\frac{1}{2}\int dS_k \epsilon_{abc}\epsilon^{ijk}\partial_i w^a \partial_j w^b w^c,$$
$$\sum_{a=1}^{3} w^{a2} = \frac{2}{R}. \tag{95}$$

Equations (93) and (95) are our final result for the gravitational Wilson loop in the cylinder $S^2 \times R$ space of the constant curvature R. The Wilson loop implicitly depends on the metric through Eq. (89). We now make several comments.

(i) The parallel transporter must depend on the metric along the contour but not on the spanning surface for the contour, because this surface can be drawn arbitrarily. This is indeed so despite the surface form of the result because

$$\partial_k(\epsilon_{abc}\epsilon^{ijk}\partial_i w^a \partial_j w^b w^c) = 0. \tag{96}$$

Therefore, the flux in Eq. (95) can be presented as a circulation of a certain vector.

(ii) The flux in Eq. (95) has the form of a well-known expression for the winding number of a mapping $S^2 \mapsto S^2$. For a closed or infinite surface, the winding number is normalized as

$$\frac{1}{8\pi}\left(\frac{2}{R}\right)^{3/2} \times \int dS_k \epsilon_{abc}\epsilon^{ijk}\partial_i w^a \partial_j w^b w^c = Q = \text{integer}. \tag{97}$$

(iii) For small contours, Eqs. (93) and (95) reproduce the result of the previous section. To check this, we rewrite the general small-loop expansion (69) for the specific metric in Eq. (87). We find

$$R_{klpq} = \frac{1}{Rr^6}\epsilon_{klu}x^u \epsilon_{pqv}x^v, \quad g^{ij} = \frac{R}{2}r^2\delta^{ij}. \tag{98}$$

Inserting this in Eq. (69) and then performing a general coordinate transformation $x^i \longrightarrow y^i(x)$, we obtain, after some simple algebra,

$$W_J^G = 1 - \frac{J(J+1)}{6}\left(\frac{\epsilon_{pqu}y^u \partial_i y^p \partial_j y^q \Delta S^{ij}}{|y|^3}\right)^2, \tag{99}$$

which exactly coincides with the expansion of Eq. (93) in the small loop area ΔS up to the second order.

8. THE NON-ABELIAN STOKES THEOREM IN $d = 3$ GRAVITY

In Section 4, we have shown that the gravitational Wilson loop viewed as a functional of the metric is equal to the Yang–Mills Wilson loop viewed as a functional of the Yang–Mills potential, provided this potential is set equal to the spin connection corresponding to the metric in question.

We now present the Yang–Mills Wilson loop in terms of our non-Abelian Stokes formula, see Eq. (14):

$$W_J^G[\text{metric}] = W_J^{YM}[\text{spin connection}] = \int D\mathbf{n}\delta(\mathbf{n}^2 - 1) \times \exp\frac{iJ}{2}\int d^2S^{ij}[-F_{ij}^a n^a + \epsilon^{abc}n^a(D_i n)^b(D_j n)^c]. \tag{100}$$

We next replace the surface element by its dual $dS^{ij} = \epsilon^{ijp}dS_p$ with the aim to rewrite this representation for the Wilson loop in terms of the metric of the curved three-dimensional space. To this end, we first decompose the integration unit vector $\mathbf{n}$ in the dreibein:

$$n^a = m^i e_i^a, \quad n^a n^a = m^i m^j e_i^a e_j^a = m^i m^j g_{ij} = 1. \tag{101}$$

The new 3-vector $\mathbf{m}$ is a covariant unit vector. Because the background metric g_{ij} is fixed, we only change the integration variables from $\mathbf{n}$ to $\mathbf{m}$ as

$$\int D\mathbf{n}\delta(\mathbf{n}^2 - 1)\ldots = \int D\mathbf{m}\sqrt{g}\delta(m^i m^j g_{ij} - 1)\ldots \tag{102}$$

We next use relation (42) of the field strength F_{ij}^a computed from the spin connection

$$A_i^a = (1/2)\epsilon^{abc}\omega_i^{bc}$$

to the Riemann tensor. The first term in the exponent of Eq. (100) becomes

$$\text{first term} = -dS_p\epsilon^{ijp} \times\left(-\frac{1}{2}\right)\epsilon^{abc}m^n e_n^a R_{kij}^l e_l^b e^{ck}. \tag{103}$$

Using

$$\epsilon^{abc}e^{bl}e^{ck} = \frac{1}{\sqrt{g}}\epsilon^{lkm}e_m^a, \quad \sqrt{g} = \det e_i^a, \tag{104}$$

Equation (103) can be continued as

$$\text{first term} = dS_p\epsilon^{ijp}\frac{1}{2\sqrt{g}}R_{ijkl}\epsilon^{klm}g_{mn}m^n. \tag{105}$$

The combination of the covariant Riemann tensor and two antisymmetric epsilon symbols has been encountered in Section 5: in three dimensions, it gives the Einstein tensor, see Eq. (73). We thus obtain

$$\text{first term} = dS_p\sqrt{g}(R\delta_n^p - 2R_n^p)m^n, \tag{106}$$

where R_n^p is the Ricci tensor and $R = R_k^k$ is the scalar curvature.

We now turn to the second term in the exponent in Eq. (100) and again use decomposition (101). We exploit fundamental relation (33) that can be presented as

$$D_j^{bb'} n^{b'} = e_k^b (\nabla_j)_l^k m^l, \tag{107}$$

where

$$D_j^{bb'} = \partial_j \delta^{bb'} + \epsilon^{bcb'} A_j^c$$

is the Yang–Mills covariant derivative and

$$(\nabla_j)_l^k = \partial_j \delta_l^k + \Gamma_{jl}^k$$

is the gravitational covariant derivative. The second term is therefore given by

$$\begin{aligned} \text{second term} &= dS_p \epsilon^{abc} e_k^a e_l^b e_n^c \epsilon^{ijp} m^k (\nabla_i)_{l'}^l \\ &\times m^{l'} (\nabla_j)_{n'}^n m^{n'} \\ &= dS_p \sqrt{g} \epsilon^{ijp} \epsilon_{kln} m^k (\nabla_i m)^l (\nabla_j m)^n. \end{aligned} \tag{108}$$

Gathering Eqs. (102), (106), and (108) together, we finally obtain a non-Abelian Stokes theorem for the gravitational Wilson loop or the trace of the spin-J parallel transporter along a closed contour:

$$\begin{aligned} W_J^G &= \int D\mathbf{m} \sqrt{g} \delta(m^i m^j g_{ij} - 1) \\ &\times \exp i \frac{J}{2} \int dS_k \sqrt{g} [(R\delta_p^k - 2R_p^k) m^p \\ &+ \epsilon^{ijk} \epsilon_{pqr} m^p (\nabla_i m)^q (\nabla_j m)^r]. \end{aligned} \tag{109}$$

Several comments are in order here.

(i) The holonomy, which was defined as a path-ordered exponential, is expressed here by a simple exponential of an integral over the spanning surface for the closed contour. That is why we call our formula a "Stokes theorem." The price to pay is the functional integration over the covariantly unit vector **m** defined on the surface.

(ii) Equation (109) is invariant under diffeomorphisms in the sense that the holonomy remains invariant under a general coordinate transformation

$$x^i \longrightarrow x'^i(x^i)$$

and the appropriate change of the surface.

(iii) The parallel transporter depends only on the contour but must not depend on the spanning surface. The surface integral in Eq. (109) has the form

$$\int dS_k \sqrt{g} V^k, \tag{110}$$

and the condition that it does not depend on the form of the surface is

$$\partial_k (\sqrt{g} V^k) = 0, \tag{111}$$

or equivalently,

$$(\nabla_k)_l^k V^l = 0, \tag{112}$$

because

$$\Gamma_{kl}^k = \Gamma_{lk}^k = \partial_l \ln \sqrt{g}.$$

The verification of Eq. (112) is rather lengthy and we relegate it to the Appendix.

(iv) With condition (112) or equivalently (111) satisfied, the surface integral can be written as

$$\begin{aligned} &\int dS_k \sqrt{g} V^k \\ &= \int dS_k \epsilon^{ijk} \partial_j B_k = -\oint dx^i \dot{B}_i \end{aligned} \tag{113}$$

proving that it depends only on the contour, as it should be. However, the vector field B_i cannot be uniquely determined from the metric tensor and the covariantly unit vector **m**.

(v) The following comment is closely related to the previous one. Parallel transporters of integer spins 1, 2, … are defined via Christoffel's Γ symbols and hence by the metric tensor, while parallel transporters of half-integer spins 1/2, 3/2, … are not: they are defined via the spin connection that is not uniquely constructed from the metric. Nevertheless, it should be expected that the holonomy for half-integer spins, being a diffeomorphism-invariant quantity, can be expressed through the metric only. Equation (109) solves this nontrivial problem: only the metric and its derivatives are involved. The solution is possible only with the holonomy represented in the form of a surface integral, as in Eq. (109). One cannot solve this problem in a contour form because it is not uniquely expressible through the metric. If that were possible, one would be able to write a parallel transporter in terms of the metric along an open contour as well, but that is not so for half-integer spins.

(vi) Equation (109) solves another long-standing problem in the Yang–Mills theory. It was recently shown [12–14] that the $SU(2)$ Yang–Mills partition function in three dimensions can be exactly rewritten in terms of gauge-invariant quantities given by the six components of the dual space metric tensor. The usual argument why this rewriting is not very useful is that external sources couple to the Yang–Mills potential and not to gauge-invariant quantities. However, we now have demonstrated that a typical source—the Yang–Mills Wilson loop—can be expressed not only through the potential but also through the metric tensor, which is gauge-invariant. Thus, not only the partition function, but also the Wilson loops in the $d = 3$ Yang–Mills theory can be expressed through local gauge-invariant quantities.

9. THE NON-ABELIAN STOKES THEOREM IN $d = 4$ GRAVITY

The aim of this section is to express the holonomy $W_{(J_1, J_2)}^G$ in the representation (J_1, J_2) in a curved $d = 4$ space through the metric tensor and its derivatives.

Equation (68) presents the holonomy in terms of the (anti)self-dual parts of the spin connection. The latter is not uniquely determined by the metric, which is not satisfactory. In addition, we would like to eliminate the path-ordering in the Yang–Mills Wilson loops $W^{\pi,\rho}$ entering Eq. (68). Both goals are achieved via the non-Abelian Stokes theorem similar to that of the previous section, which we now derive.

We start by applying representation (14) to the Yang–Mills Wilson loop W^{π},

$$W_J^{\pi} = \int D\mathbf{n}\,\delta(\mathbf{n}^2-1)\exp\Big(i\frac{J}{2}\int dS^{\mu\nu}[-F^a_{\mu\nu}(\pi)n^a + \epsilon^{abc}n^a(D_\mu(\pi)n)^b(D_\nu(\pi)n)^c]\Big), \tag{114}$$

where

$$D^{ab}_\mu(\pi) = \partial_\mu\delta^{ab} + \epsilon^{acb}\pi^c_\mu$$

is the covariant derivative with respect to the self-dual part of the spin connection and $F^a_{\mu\nu}(\pi)$ is the appropriate field strength (53); it is related to the Riemann tensor via Eq. (55). We next introduce the antisymmetric tensor

$$m^{\kappa\lambda} = \frac{1}{2}n^a\eta^{aAB}e^{A\kappa}e^{B\lambda}. \tag{115}$$

The first term in Eq. (114) can be written as $-R_{\kappa\lambda\mu\nu-}m^{\kappa\lambda}$. The tensor $m^{\kappa\lambda}$ has actually only two independent components. To see this, we introduce two covariant projector operators

$$P^+_{\kappa\lambda\mu\nu} = \frac{1}{4}\eta^{aAB}\eta^{aCD}e^A_\kappa e^B_\lambda e^C_\mu e^D_\nu = \frac{1}{4}(g_{\kappa\mu}g_{\lambda\nu}-g_{\kappa\nu}g_{\lambda\mu}+\sqrt{g}\epsilon_{\kappa\lambda\mu\nu}), \tag{116}$$

$$P^-_{\kappa\lambda\mu\nu} = \frac{1}{4}\bar\eta^{aAB}\bar\eta^{aCD}e^A_\kappa e^B_\lambda e^C_\mu e^D_\nu = \frac{1}{4}(g_{\kappa\mu}g_{\lambda\nu}-g_{\kappa\nu}g_{\lambda\mu}-\sqrt{g}\epsilon_{\kappa\lambda\mu\nu}), \tag{117}$$

satisfying the projector conditions

$$P^\pm_{\kappa\lambda\mu\nu}g^{\mu\mu'}g^{\nu\nu'}P^\pm_{\mu'\nu'\rho\sigma} = P^\pm_{\kappa\lambda\rho\sigma}, \tag{118}$$

$$P^\pm_{\kappa\lambda\mu\nu}g^{\mu\mu'}g^{\nu\nu'}P^\mp_{\mu'\nu'\rho\sigma} = 0, \tag{119}$$

$$P^\pm_{\kappa\lambda\mu\nu}g^{\kappa\mu}g^{\lambda\nu} = 3. \tag{120}$$

$P^\pm_{\kappa\lambda\mu\nu}$ are (covariantly) orthogonal projectors, each having three zero and three nonzero eigenvalues. They project a general antisymmetric tensor into (covariantly) self-dual and anti-self-dual parts, respectively. It is easy to verify that the tensor $m^{\kappa\lambda}$ introduced in Eq. (115) is self-dual,

$$P^-_{\kappa\lambda\mu\nu}m^{\kappa\lambda} = 0, \tag{121}$$

and satisfies the normalization condition

$$m^{\kappa\lambda}m_{\kappa\lambda} = P^+_{\kappa\lambda\mu\nu}m^{\kappa\lambda}m^{\mu\nu} = 1, \tag{122}$$

which follows from the normalization $\mathbf{n}^2 = 1$. Therefore, $m^{\kappa\lambda}$ indeed has only two independent degrees of freedom in a given metric. We change the integration variables in Eq. (114) from $\mathbf{n}$ to $m^{\kappa\lambda}$,

$$\int D\mathbf{n}\,\delta(\mathbf{n}^2-1)\ldots = \int Dm^{\kappa\lambda}\sqrt{g}\,\delta(P^-_{\kappa\lambda\mu\nu}m^{\mu\nu})\delta(m^{\kappa\lambda}m_{\kappa\lambda}-1)\ldots \tag{123}$$

We now compute the covariant derivative of $m^{\kappa\lambda}$ as

$$\begin{aligned} m^{\kappa\lambda}{}_{;\mu} &= \partial_\mu m^{\kappa\lambda} + \Gamma^\kappa_{\mu\nu}m^{\nu\lambda} + \Gamma^\lambda_{\mu\nu}m^{\kappa\nu} \\ &= \frac{1}{2}\eta^{aAB}[\partial_\mu n^a e^{A\kappa}e^{B\lambda} + n^a(\partial_\mu e^{A\kappa} + \Gamma^\kappa_{\mu\nu}e^{A\nu})e^{B\lambda} \\ &\quad + n^a e^{A\kappa}(\partial_\mu e^{B\lambda} + \Gamma^\lambda_{\mu\nu}e^{B\nu})] \\ &= \frac{1}{2}\eta^{aAB}[\partial_\mu n^a e^{A\kappa}e^{B\lambda} \\ &\quad - n^a\omega^{AC}_\mu e^{C\kappa}e^{B\lambda} - n^a e^{A\kappa}\omega^{BC}_\mu e^{C\lambda}], \end{aligned} \tag{124}$$

where in the last equation, we have used fundamental relation (33). We now insert the decomposition of the spin connection ω^{AB}_μ into the self-dual and anti-self-dual parts, Eq. (51). Using the relations for the η, $\bar\eta$ symbols,

$$\begin{aligned} \eta^{aAB}\eta^{bAC} &= \delta^{ab}\delta^{BC} + \epsilon^{abc}\eta^{cBC}, \\ \bar\eta^{aAB}\bar\eta^{bAC} &= \delta^{ab}\delta^{BC} + \epsilon^{abc}\bar\eta^{cBC}, \end{aligned} \tag{125}$$

$$\eta^{aAB}\bar\eta^{bAC} = \eta^{aAC}\bar\eta^{bAB}, \tag{126}$$

it is easy to see that only the self-dual piece of ω^{AB}_μ survives in Eq. (124), giving

$$\begin{aligned} m^{\kappa\lambda}{}_{;\mu} &= \frac{1}{2}\eta^{aAB}e^{A\kappa}e^{B\lambda}(\partial_\mu\delta^{ab} + \epsilon^{acb}\pi^c_\mu)n^b \\ &= \frac{1}{2}\eta^{aAB}e^{A\kappa}e^{B\lambda}(D_\mu(\pi)n)^a. \end{aligned} \tag{127}$$

In other words, the gravitational covariant derivative of $m^{\kappa\lambda}$ is expressed through the Yang–Mills covariant derivative of the $\mathbf{n}$ field entering the second term in Eq. (114).

Using consecutively Eqs. (125) and (127), we finally rewrite Eq. (114) in terms of the metric:

$$W^{\pi}_{J_1} = \int Dm^{\kappa\lambda}\sqrt{g}\,\delta(P^{-}_{\kappa\lambda\mu\nu}m^{\mu\nu})\,\delta(m^{\kappa\lambda}m_{\kappa\lambda}-1) \times \exp\left(i\frac{J_1}{2}\int dS^{\mu\nu}\left[-R_{\kappa\lambda\mu\nu}m^{\kappa\lambda} - \frac{1}{2}\sqrt{g}\,\epsilon_{\kappa\rho\sigma\tau}g_{\lambda\lambda'}m^{\kappa\lambda'}m^{\lambda\rho}{}_{;\mu}\,m^{\sigma\tau}{}_{;\nu}\right]\right). \quad (128)$$

Similarly, W^{ρ} is obtained by integrating over the anti-self-dual covariantly unit tensors:

$$W^{\rho}_{J_2} = \int Dm^{\kappa\lambda}\sqrt{g}\,\delta(P^{+}_{\kappa\lambda\mu\nu}m^{\mu\nu})\,\delta(m^{\kappa\lambda}m_{\kappa\lambda}-1) \times \exp\left(i\frac{J_2}{2}\int dS^{\mu\nu}\left[-R_{\kappa\lambda\mu\nu}m^{\kappa\lambda} + \frac{1}{2}\sqrt{g}\,\epsilon_{\kappa\rho\sigma\tau}g_{\lambda\lambda'}m^{\kappa\lambda'}m^{\lambda\rho}{}_{;\mu}\,m^{\sigma\tau}{}_{;\nu}\right]\right). \quad (129)$$

As derived in Section 4.2, the gravitational holonomy in the representation (J_1, J_2) is the product of the two components,

$$W^{G}_{(J_1, J_2)} = W^{\pi}_{J_1}W^{\rho}_{J_2}. \quad (130)$$

Equations (128), (129), and (130) constitute the "non-Abelian Stokes theorem" for the holonomy in a curved $d = 4$ space. It expresses the holonomy via surface integrals over spanning surfaces for the contour and presents the holonomy in terms of the metric tensor and its derivatives only, without referring to the spin connection, even for half-integer representations (J_1, J_2).

10. CONCLUSION

The main results of this paper are the non-Abelian Stokes theorems for holonomies: the Yang–Mills Wilson loop (Eq. (14)) and the traces of parallel transporters in curved $d = 3$ (Eq. (109)) and $d = 4$ (Eqs. (128) and (129)) spaces. In all these cases, the path-ordered exponentials of the connections are replaced by ordinary exponentials of surface integrals, which, however, do not actually depend on the way the surface is spanned on the contour. The price to pay for the removal of path-ordering is high: we obtain functional integrals instead. In the simplest case of the $SU(2)$ Yang–Mills theory, this is an integral over a unit 3-vector **n** "living" on the surface; for the $d = 3$ Riemannian manifold, this is an integral over a covariantly unit 3-vector **m**, and for $d = 4$, one integrates over (anti)self-dual covariantly unit tensors.

In spite of the occurrence of functional integration, we believe that our formulas are aesthetically appealing. Compared to path-ordered exponentials, they are better suited to averaging over quantum ensembles of Yang–Mills fields or over various metrics. We hope that elegant formulas can also be used in more general settings.

In addition to the general non-Abelian Stokes formulas, we have presented holonomy as a surface integral for a specific background, namely, for a constant-curvature $d = 3$ space with the cylinder topology $S^2 \times R$. The "gravitational Wilson loop" is given by a formula for the character whose argument is the winding number of external coordinates, see Section 8.

Parallel transporters of integer spins have a dual description: such a transporter can be defined either as a path-ordered exponential of Christoffel symbols or as a path-ordered exponential of the spin connection in the appropriate representation. In Section 4, we have shown that these representations are equivalent. Even though the spin connection is not uniquely determined by the metric tensor, this equivalence implies that the holonomy written in terms of the spin connection can in fact be expressed through the metric only.

For half-integer spins, the situation is far less trivial because the only way to define the holonomy is via the spin connection, and it is not at all clear beforehand that the holonomy can be uniquely written through the metric tensor and its derivatives. The non-Abelian Stokes theorem proved in this paper demonstrates that this rewriting can be achieved, but only with the holonomy presented in the surface form. Although the surface integral does not depend on the way one draws the surface and can actually be written as an integral along the contour, the contour form is not uniquely defined by the surface one, which reflects the ambiguity in determining the spin connection from the metric.

This finding has an interesting implication for the Yang–Mills theory in three dimensions, which can be identically reformulated as a quantum gravity theory with the partition function written as a functional integral over the metric tensor of the dual space [12, 13, 14]. This metric tensor is local and gauge invariant (in the Yang–Mills sense). However, one might wish to calculate the average of the Wilson loop, which is originally defined by the Yang–Mills potential, but not by the metric tensor. In the "quantum gravity" formulation, the Yang–Mills Wilson loop becomes a parallel transporter in the gravitational sense. It is therefore very important that the Yang–Mills Wilson loop in any representation can be expressed through the gauge-invariant metric tensor. Thus, not only the partition function but also the Wilson loop can be presented in terms of local and gauge-invariant quantities. This subject is described in more detail elsewhere [15].

ACKNOWLEDGMENTS

One of us (V. P.) thanks NORDITA for kind hospitality and the Russian Foundation for Basic Research for partial support (grant no. 97-27-15L).

APPENDIX

Proof that Eq. (109) does not depend on the surface

The path-integral representation for the "gravitational Wilson loop" in Eq. (109) must not depend on the choice of the spanning surface for a given contour, but only on the contour itself. To prove that this is so, we verify Eq. (112),

$$(\nabla_k)_l^k V^l = 0, \tag{131}$$

where

$$V^k = (R\delta_p^k - 2R_p^k)m^p + \epsilon^{ijk}\epsilon_{pqr}m^p(\nabla_i m)^q(\nabla_j m)^r, \qquad m^i m^j g_{ij} = 1. \tag{132}$$

To simplify the notation, we denote covariant derivatives by ";" (see [20]). Explicitly, the covariant derivatives of a scalar, a vector, and a tensor are given by

$$S_{;k} = \partial_k S,$$

$$V^i_{;k} = \partial_k V^i + \Gamma^i_{kl}V^l, \quad V_{i;\,k} = \partial_k V_i - \Gamma^l_{ik}V_l, \qquad T^{ij}_{;k} = \partial_k T^{ij} + \Gamma^i_{kl}T^{lj} + \Gamma^j_{kl}T^{il}, \tag{133}$$

$$T_{ij;\,k} = \partial_k T_{ij} - \Gamma^l_{ik}T_{lj} - \Gamma^l_{jk}T_{il}, \quad \text{etc.}$$

The ordinary derivative of a convolution of two tensors can be written as a sum of covariant derivatives,

$$\partial_k(T^{(1)\ldots i}_{\ldots}T^{(2)\ldots}_{\ldots i}) = T^{(1)\ldots i}_{\ldots;\,k}T^{(2)\ldots}_{\ldots i} + T^{(1)\ldots i}_{\ldots}T^{(2)\ldots}_{\ldots i;\,k}. \tag{134}$$

We apply the covariant derivative to the first term of the vector V^k,

$$\nabla_k[(R\delta_p^k - 2R_p^k)m^p] = (R\delta_p^k - 2R_p^k)_{;k}m^p + (R\delta_p^k - 2R_p^k)m^p_{;k}. \tag{135}$$

The covariant derivative of the Einstein tensor is known to be zero [20, Eq. (92.10)]. Therefore, only the second term survives in Eq. (135).

We next apply the covariant derivative to the second term of V^k as

$$\nabla_k[\epsilon^{ijk}\epsilon_{pqr}m^p(\nabla_i m)^q(\nabla_j m)^r] = \epsilon^{ijk}\epsilon_{pqr}(\nabla_k m)^p(\nabla_i m)^q(\nabla_j m)^r + 2\epsilon^{ijk}\epsilon_{pqr}m^p(\nabla_i m)^q(\nabla_k\nabla_j m)^r. \tag{136}$$

The first term here vanishes for the following reasons. Differentiating the condition that m^i is a covariantly unit vector, we obtain

$$0 = \partial_k(m^i m^j g_{ij}) = 2g_{ij}(\nabla_k m)^i m^j = 2(\nabla_k m)^i m_i, \tag{137}$$

because the covariant derivative of the metric tensor is zero. This implies that the three vectors $(\nabla_{1,2,3}m)^i$ are not linearly independent, because three linearly independent vectors cannot be orthogonal to a given vector (in this case, m_i) in three dimensions. The first term in Eq. (136) is the antisymmetrized product of these three linearly dependent vectors and is therefore zero.

The second term in Eq. (136) contains the commutator of covariant derivatives, equal to

$$\epsilon^{ijk}(\nabla_k\nabla_j m)^r = \frac{1}{2}\epsilon^{ijk}[\nabla_k\nabla_j]^r_s m^s = \frac{1}{2}\epsilon^{ijk}g^{rt}R_{tskj}m^s, \tag{138}$$

where R_{tskj} is the Riemann tensor. Therefore, the second (and the only nonzero) term in Eq. (136) can be written as

$$\epsilon^{ijk}\epsilon_{pqr}g^{rt}R_{tskj}m^p m^s(\nabla_i m)^q. \tag{139}$$

We next use Eq. (70) to express the Riemann tensor through the Ricci and metric tensors and write the product of two epsilon symbols as a determinant made of Kronecker deltas. Performing all convolutions, we obtain that Eq. (139) can be identically rewritten as

$$[g_{qs}(R\delta^i_p - 2R^i_p) - g_{ps}(R\delta^i_q - 2R^i_q)] \times m^p m^s(\nabla_i m)^q. \tag{140}$$

Here, the first term is zero because of Eq. (137) and in the second term, we use

$$q_{ps}m^p m^s = 1.$$

This gives

$$-(R\delta^i_q - 2R^i_q)(\nabla_i m)^q, \tag{141}$$

which cancels exactly with Eq. (135). Thus, $(\nabla_k)_l^k V^l = 0$, q.e.d.

REFERENCES

1. D. Diakonov and V. Petrov, Pis'ma Zh. Éksp. Teor. Fiz. **49**, 284 (1989) [JETP Lett. **49**, 323 (1989)]; Phys. Lett. B **224**, 131 (1989).
2. D. Diakonov and V. Petrov, in *Nonperturbative Approaches to Quantum Chromodynamics: Proceedings of the International ECT* Workshop, Trento, 1995*, Ed. by D. Diakonov, p. 36; hep-th/9606104 (1996).
3. M. B. Halpern, Phys. Rev. D **19**, 517 (1979).
4. I. Ya. Aref'eva, Theor. Math. Phys. **43**, 353 (1980).
5. N. Bralic, Phys. Rev. D **22**, 3090 (1980).
6. Yu. A. Simonov, Yad. Fiz. **50**, 213 (1989) [Sov. J. Nucl. Phys. **50**, 134 (1989)].

7. K.-I. Kondo and Y. Taira, hep-th/9911242.
8. D. Diakonov and V. Petrov, Phys. Lett. **242**, 425 (1990).
9. A. M. Polyakov, Nucl. Phys. (Proc. Suppl.) **68**, 1 (1998); hep-th/9711002 (1997).
10. C. Kortals-Altes and A. Kovner, hep-ph/0004052 (2000).
11. B. Broda, E-print archives, math-ph/0012035.
12. R. Anishetty, S. Cheluvaraja, H. S. Sharatchandza, and M. Matur, Phys. Lett. B **341**, 387 (1993).
13. D. Diakonov and V. Petrov, Zh. Éksp. Teor. Fiz. **118**, 1012 (2000) [JETP **91**, 873 (2000)]; hep-th/9912268.
14. R. Anishetty, S. Cheluvaraja, and H. S. Sharatchandza, Phys. Lett. B **373**, 373 (2000).
15. D. Diakonov and V. Petrov, Phys. Lett. B **493**, 169 (2000); hep-th/0009007.
16. A. Alekseev, L. Faddeev, and S. Shatashvili, J. Geom. Phys. **5**, 391 (1989).
17. F. A. Lunev, Nucl. Phys. B **494**, 433 (1997); hep-th/9609166 (1996).
18. D. Diakonov and V. Petrov, hep-lat/0008004 (2000).
19. A. M. Perelomov, *Generalized Coherent States and Their Applications* (Springer-Verlag, New York, 1986); Phys. Rep. **146**, 135 (1987).
20. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Nauka, Moscow, 1988; Pergamon, Oxford, 1980).

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 921–929.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1067–1076.

GRAVITATION, ASTROPHYSICS

The Formation of Primary Galactic Nuclei during Phase Transitions in the Early Universe

S. G. Rubin[a, b, *], A. S. Sakharov[c, **], and M. Yu. Khlopov[a, b, d, *]**

[a]*Moscow Engineering Physics Institute, Moscow, 115409 Russia*
[b]*Center for Cosmoparticle Physics "Cosmion," Moscow, 125047 Russia*
[c]*Labor für Höchenergiephysik, ETH-Hönggerberg, HPK-Gebäude, CH-8093, Zürich, Switzerland*
[d]*Institute for Applied Mathematics, Russian Academy of Sciences, Moscow, 125047 Russia*
**e-mail: serg.rubin@mtu-net.ru*
***e-mail: sakhas@particle.phys.ethz.ch*
****e-mail: mkhlopov@orc.ru*
Received February 8, 2001

Abstract—A new mechanism describing the formation of protogalaxies is proposed, based on the second-order phase transition in the inflation stage and the domain wall formation upon the end of inflation. The presence of closed domain walls with the size markedly exceeding the cosmological horizon at the instant of their formation and the wall collapse in the postinflation epoch (when the wall size becomes comparable with the cosmological horizon) lead to the formation of massive black hole clusters that can serve as nuclei for the future galaxies. The black hole mass distributions obtained do not contradict the available experimental data. The number of black holes with $M \sim 100$ solar masses ($M_\odot$) and above is comparable with the number of galaxies in the visible Universe. Development of the proposed approach gives grounds for a principally new scenario of galaxy formation in the model of a hot Universe. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In the past decade, investigations into the nature of active galactic nuclei exhibited a considerable progress. Now there is virtually no doubt that the centers of galaxies contain massive black holes [1]. It is the existence of black holes with masses on the order of 10^6–10^8 $M_\odot$ in the galactic nuclei and the accretion of matter onto these holes that is believed to account for the physical nature of their activity. A possible explanation for the formation of such supermassive black holes assumes the collapse of a large number of stars caused by their high concentration at the galaxy center. However, the mechanism of galactic nuclei formation is still unclear. According to Veilleux [2], there are serious grounds to believe that the formation of stars and galaxies proceeded simultaneously. Stiavelli [3] considered a model of galaxy formation around a massive black hole and presented arguments in favor of this model (see also [4]). Each of the two approaches has certain advantages, while not being free of drawbacks.

The problem of the possible "primordial" black hole (PBH) formation is still open. In contrast to the case of "secondary" black holes, which are related to the evolution of stars and stellar systems, there is no convincing astronomic evidence for the existence of PBHs. Restrictions posed by the astronomic data on the PBH concentration offer a unique source of information on the inhomogeneity of the early Universe [5] and on the physical processes accounting for this inhomogeneity [6]. Generally speaking, the PBH mass may be arbitrary, ranging from the Planck value (or even below [7]) up to a mass contained within the contemporary cosmological horizon. However, in most cases the astrophysical effects related to the presence of PBH are restricted to masses much lower than the solar mass. The reason is that the mechanism of PBH formation is usually related to the development of inhomogeneities bounded by the cosmological horizon. The data of observations concerning the distribution of light elements and the spectrum of cosmic microwave radiation pose very rigid restrictions on the magnitude of inhomogeneities existing in the pregalactic stage following the first second of expansion of the Universe. Thus, realistic mechanisms of PBH formation have to be apparently related to very early ($t \ll 1$ s) stages of the evolution of the Universe—when the mass contained within the cosmological horizon and limiting the possible PBH mass was significantly lower than the solar mass. Nevertheless, the actively discussed possibility of a genetic relationship between quasars and active galactic nuclei, on the one hand, and the existence of PBHs with much greater masses, on the other hand [8], becomes a subject for detailed investigations [9–11].

Below we will consider a new mechanism describing the early formation of PBHs, which serve as the nucleation centers in the subsequent formation of galaxies. This mechanism may prove to be free of disad-

vantages inherent in the models based on the concept of a single PBH being a nucleus of the future galaxy.

Previously [11] we proposed a new mechanism of PBH formation that opens the possibility of massive black hole formation in the early Universe. The mechanism is based on the possibility that black holes are formed as a result of a collapse of closed walls formed during a second-order phase transition. The masses of such black holes may vary within broad limits, up to a level on the order of $10^8\ M_\odot$.

Let us assume that a potential of the field system possesses at least two different vacuum states. Then there are two possible distributions of these states in the early Universe. The first possibility is that the Universe contains approximately equal numbers of both states, which is typical of a temperature-controlled phase transition under usual conditions. The alternative possibility corresponds to the case when the two vacuum states form with different probabilities. In this case, there appear islands of less probable vacuum state surrounded by the sea of another, more probable vacuum state. As was recently demonstrated [12], an important condition for this distribution is the existence of valleys in the scalar field potential during inflation. Then the background de Sitter fluctuations in this massless scalar field lead to the formation of islands representing one vacuum in the sea of another vacuum. This phase transition takes place after the end of inflation in the Friedmann–Robertson–Walker Universe. After the phase transition, the two vacuum states are separated by a wall; the size of this wall may be significantly greater as compared to the cosmological horizon at that period of time. At some instant after crossing the horizon, the walls begin to contract because of the surface tension. As a result, provided that friction is absent and the wall does not radiate a considerable part of its energy in the form of scalar waves, almost all energy of this closed wall may be concentrated within a small volume inside the gravitational radius. This is a necessary condition for black hole formation.

The mass spectrum of black holes formed by this mechanism depends on parameters of the scalar field potential determining the direction and size of the potential valley during inflation and the postinflation phase transition. Although we deal here with the so-called pseudo-Nambu–Goldstone field, the proposed mechanism possesses a sufficiently general character.

The presence of massive PBHs is a new factor in the development of gravitational instability in the surrounding matter and may serve as a basis for new scenarios of the formation and evolution of galaxies.

2. THE FORMATION OF CLOSED WALLS OF A COMPLEX FIELD

Now we will describe a mechanism accounting for the appearance of massive walls with the size markedly greater than the horizon at the end of inflation. Let us consider a complex scalar field with the potential

$$V(\varphi) = \lambda(|\varphi|^2 - f^2/2)^2, \tag{1}$$

where $\varphi = re^{i\theta}$. We assume the mass of the radial field component r to be sufficiently large, so that the complex field would occur in the ground state even before the end of inflation. Since the minimum of potential (1) is degenerate, the field has the form

$$\varphi \approx (f/\sqrt{2})e^{i\theta},$$

with the phase θ acquiring the meaning of a massless field.

For the following considerations, it should be noted that, using expression (1) in the inflation period, we ignored the term

$$\delta V(\theta) = \Lambda^4(1 - \cos\theta) \tag{2}$$

reflecting the contribution of instanton effects to the Lagrangian renormalization. Since the parameter Λ appears as a result of the renormalization, its value cannot be large and we may quite reasonably assume that $\Lambda \ll H, f$. The omitted term (2) begins to play a significant role in the postinflation stage, when the Hubble constant sharply decreases with time ($H = 1/2t$ during the radiation dominated epoch).

Let us assume that a part of the Universe occurring inside the contemporary horizon was formed N_U e-folds before the end of inflation. As was demonstrated in [13], the quantum field fluctuations during inflation are rapidly transformed into a classical field component, while the massless field values in the neighboring causality-disconnected regions with the size H^{-1} differ on the average by

$$\delta\theta = H/2\pi f \tag{3}$$

after a single e-fold. In the next time step $\Delta t = H^{-1}$ (i.e., during the next e-fold) each causality-connected domain is divided into $\sim e^3$ causality-disconnected subdomains; the phase in each of the new domains differs by $\sim\delta\theta$ from that in the preceding step. Thus, more and more domains appear with time in which the phase differs significantly from the initial value. More precisely, the probability of finding the phase θ is [14–16]

$$P(\theta, N) = \frac{1}{\sqrt{2\pi}\sigma_N}\exp\left\{-\frac{(\theta_U - \theta)^2}{2\sigma_N^2}\right\},$$
$$\sigma_N = \frac{H}{2\pi f}\sqrt{N_U - N}, \tag{4}$$

where N is the number of e-folds remaining to the end of the inflation period and θ_U is the random phase value at the instant of formation of the causality-connected domain corresponding to a visible part of the contemporary Universe. Without a loss of generality, we may select $0 < \theta_U < \pi$. Below we will demonstrate that a par-

ticular value of the initial phase significantly affects the evolution of the Universe in the postinflation epoch.

The probability of finding a certain phase obeys the Gaussian distribution (4) and, hence, the phase averaged over the entire space equals the random initial phase θ_U. A principally important point is the appearance of domains with the phases $\theta > \pi$. Appearing only after a certain period of time during which the Universe exhibited exponential expansion, these domains turn out to be surrounded by a space with the phase $\theta < \pi$. These very domains lead in what follows to the formation of large-scale structures. Note that the phase fluctuations during the first *e*-folds may, generally speaking, eventually transform into fluctuations of the cosmic microwave radiation, which will lead to imposing restrictions on the scaling parameter *f*. This difficulty can be avoided by taking into account interaction of the field φ with the inflaton field (i.e., by making parameter *f* a variable).

Initially, the potential (1) possessed a $U(1)$ symmetry and the phase θ corresponded to a massless scalar field. Owing to the term (2), the symmetry is broken after the end of the inflation period: the potential of the θ field acquires minima at the points $\theta_{\min} = 0, \pm 2\pi, \pm 4\pi, \ldots$, and the field acquires the mass $m_\theta = 2f/\Lambda^2$. According to the classical equation of motion,

$$\ddot{\theta} + 3H\dot{\theta} + \frac{d\delta V}{d\theta} = 0, \qquad (5)$$

the phase performs decaying oscillations about the potential minimum, the initial values being different in various space domains. Moreover, domains with the initial phase $\pi < \theta < 2\pi$ perform oscillations about the potential minimum at $\theta_{\min} = 2\pi$, whereas the phase in the surrounding space tends to a minimum at the point $\theta_{\min} = 0$. Upon ceasing of the decaying phase oscillations, the system contains domains characterized by the phase $\theta_{\min} = 2\pi$ surrounded by the space with $\theta_{\min} = 0$. Apparently, on moving in any direction from the inside to the outside of the domain, we will unavoidably pass through a point where $\theta = \pi$ because the phase varies continuously. This implies that a closed surface must exist which is characterized by the phase $\theta_{\text{wall}} = \pi$. The size of this surface depends on the moment of domain formation in the inflation period, while the shape of the surface may be arbitrary. A principal point for the subsequent considerations is that the surface is closed.

Thus, we obtained a field configuration connecting various vacuum states of the potential (2). A rigorous classical solution of this problem possessing a translational symmetry in the two directions in space (flat wall) is well known [17]:

$$\theta_{\text{wall}}(x - x_0) = -4\arctan\left[\exp\left(\frac{x - x_0}{d}\right)\right], \qquad (6)$$

where *d* is the wall thickness. Since the thickness of a closed wall is related to microscopic parameters of the theory, whereas the characteristic wall size is *a priori* unlimited, expression (6) is applicable to within a sufficiently high accuracy in the case under consideration. As can be readily shown, the wall possesses an energy concentrated where the phase is $\theta = \pi$ [17]. Thus, we obtained a mechanism providing the formation of domains surrounded by closed walls. The surface energy of a wall depends on the Lagrangian parameters, while the wall size is determined by the time of crossing the phase value equal to π during the inflation period.

3. DOMAIN WALL DECELERATION DURING MOTION THROUGH A PLASMA

The first- and second-order phase transitions lead to the formation of a field walls separating one vacuum of this field from another. One of these mechanisms was described in the preceding section. In turn, the walls are moving at a subluminal velocity and interact with the surrounding plasma. Depending on the character of this interaction and the shape of the field potential, there are two possible situations. In the first case, the plasma particles pass through the wall, falling into a different vacuum and acquiring a certain mass. This situation corresponds to an electroweak phase interaction [18], whereby the corresponding Higgs field is responsible for a mechanism of the fermion mass production. In the opposite case, the particle mass is not changed upon going from one to another vacuum (an example is offered by the case of interaction with an axion wall). In the former case, the interaction with the medium leads to a significant retardation of the domain wall, while in the latter case, the walls are virtually transparent for the medium provided that the parameters are given reasonable values.

All considerations are conveniently conducted in the resting wall frame. The probability of a particle scattering from the plane resting wall is

$$dw = dn(\mathbf{k})2\pi\delta(\varepsilon - \varepsilon')|M|^2\frac{d^3k'}{2\varepsilon V(2\pi)^3\varepsilon'}, \qquad (7)$$

where $dn(\mathbf{k})$ is the distribution of incident particles with respect to momentum and *M* is the matrix element for the particle transition from a state with the energy ε and momentum $\mathbf{k}$ to the state with the energy ε' and momentum $\mathbf{k}'$ upon interaction with the potential $U = U(z)$ describing the plane wall. The pressure produced by incident particles upon the wall is related to the rate of their momentum transfer to the wall,

$$p = \frac{1}{S}\int dw q_z, \quad q_z = k_z' - k_z, \qquad (8)$$

where *S* is the wall area. Let us select the Lagrangian of the particle–wall interaction in the following form representing a classical configuration of the complex field phase:

$$L_{\text{int}} = i\partial_z\theta(z)J_z, \quad J_\mu = \bar{\psi}\gamma_\mu\psi. \qquad (9)$$

Calculating a matrix element for the particle scattering from the wall with the transition from the initial momentum k to final momentum k',

$$M = \langle k'|\int L_{\rm int} d^4x|k\rangle, \tag{10}$$

we obtain

$$|M|^2 = 8(4\pi)^6 S\delta^{(2)}(\mathbf{q}_\parallel)k_z^2 \frac{1}{\cosh^2(k_z d\pi)}. \tag{11}$$

In deriving formula (11), we took into account that the laws of the energy–momentum conservation lead to the following relationships:

$$k_z' = \pm k_z; \quad \mathbf{q}_\parallel \equiv \mathbf{k}_\parallel' - \mathbf{k}_\parallel = 0,$$

according to which a nonzero contribution to the pressure is only due to the reflected particles with $k_z' = -k_z$. Therefore, the pressure of incident particles upon the wall can be written as

$$p = \frac{4}{\pi^2}\int \frac{k_z^2}{\cosh^2(\pi k_z d)}(k_z - k_z')\delta(\varepsilon - \varepsilon') \times \delta(\mathbf{k}_\parallel - \mathbf{k}_\parallel')\frac{dn(k)}{V}\frac{d\mathbf{k}'}{\varepsilon\varepsilon'}. \tag{12}$$

Let us determine the distribution of the incident particles with respect to the transverse momentum $dn(\mathbf{k})$. In the resting plasma frame,

$$dn_0(\mathbf{k}_0) = C\exp\left\{-\frac{E_0(\mathbf{k}_0)}{T}\right\}\frac{d\mathbf{k}_0 V}{(2\pi)^3}. \tag{13}$$

Here and below, the subscript 0 denotes quantities determined in the resting plasma frame. Assuming the plasma temperature T to be significantly greater as compared to the fermion masses and normalizing it to the total particle density,

$$n_{\rm tot} \approx N(g*)T^3, \quad N(g*) \approx 5,$$

we obtain $C \approx 20\pi^2$. In addition, it is evident that

$$dn(\mathbf{k}) = dn_0(\mathbf{k}_0), \tag{14}$$

where the incident particle momentum in the resting wall and plasma frames (in the latter frame, the wall moves at a velocity v) are related as

$$\begin{aligned} \mathbf{k}_{0\parallel} &= \mathbf{k}_\parallel, \\ k_{0z} &= \gamma(k_z + v\varepsilon), \\ E_0 &= \gamma(vk_z + \varepsilon), \\ \gamma &= \frac{1}{\sqrt{1-v^2}}. \end{aligned} \tag{15}$$

Integrating the pressure (12) with resect to the momentum of the incident particle, we obtain

$$p = \mathrm{sgn}(k_z)\frac{32C}{(2\pi)^5}\gamma\int\frac{d\mathbf{k}}{\varepsilon^2}(\varepsilon + vk_z)\frac{k_z^2}{\cosh^2(\pi k_z d)} \times \exp\left\{-\gamma\frac{\varepsilon + vk_z}{T}\right\}. \tag{16}$$

This formula was derived with allowance for the Lorentz invariance of the phase volume d^3k'/ε'. Numerical calculation of the integral in (16) presents no difficulties, but we are interested in analytically estimating the pressure produced by the medium upon the wall. For this purpose, note that the walls are formed at temperatures $T \sim \Lambda$ and the wall thickness is $d = f/2\Lambda^2$. Therefore, there is a large parameter

$$Td \approx f/\Lambda \gg 1,$$

using which we may obtain a sufficiently reliable estimate of the integral. According to (16), the most effective scattering takes place for an incident particle momentum of

$$k_z \sim 1/\pi d \ll T.$$

At the same time, it is evident that

$$k_\parallel \sim \varepsilon \sim \gamma T \gg k_z.$$

Using these relationships, we may estimate the integral in (16). A final expression for the pressure produced by the surrounding medium upon the domain wall is as follows:

$$p \approx \frac{20}{\pi^7}\frac{\gamma}{d^4}. \tag{17}$$

4. CONDITIONS FOR PRIMORDIAL BLACK HOLE FORMATION

After heating of the Universe, the evolution of domains formed with the phase $\theta > \pi$ and which sharply increased in volume during the inflation period proceeds on the background of the Friedmann expansion and is described by the relativistic equation of state. First, an equilibrium state with the "vacuum" phase $\theta = 2\pi$ inside the domain and the $\theta = 0$ phase outside is established at $T \sim \Lambda$. A closed wall corresponding to the phase $\theta = \pi$ is formed in the transition region with a width of $\sim 1/m \sim f/\Lambda^2$, which separates the domain from the surrounding space. The surface energy density on the wall amounts to $\sim f/\Lambda^2$.

It must be noted that the process of establishing the equilibrium ("vacuum") phase values may acquire a protracted character. If the stage of coherent phase oscillations about the equilibrium values is sufficiently long, the energy density of these oscillations may become dominating and determine the dust period of

expansion. Let us consider the factors influencing the cosmological evolution of such a wall.

1. First, note that the domain size immediately after the end of inflation markedly exceeds the horizon size in the Friedmann expansion stage. The overall contraction of the closed wall may begin only when the horizon size R_h will be equal to the domain size. Up to this moment, the characteristic domain size increases with the expanding Universe because we assumed that the existing field contribution to the total energy–momentum tensor is small as compared to that of the inflaton field. Accordingly, the field gives also a small contribution to the total energy density of the Universe upon heating, when the energy density of inflaton transforms into the energy density of relativistic particles. Evidently, internal stresses developed in the wall after crossing the horizon initiate processes tending to minimize the wall surface. This implies that the wall tends, first, to acquire a spherical shape and, second, to contract toward the center. For simplicity, below we will consider the motion of closed spherical walls.

2. Since the energy of the surrounding plasma rapidly decreases, the wall energy may become at a certain time instant comparable with the energy of the surrounding medium. Simultaneously, the domain separates from the general expansion process and its radius R_w may become smaller than R_h.

3. The wall energy is proportional to its area at the instant of crossing the horizon. By the moment of maximum contraction, this energy is virtually completely converted into kinetic energy. Should the wall by this moment be localized within the limits of the gravitational radius, a PBH is formed.

4. Contracting under the action of internal forces, the wall moves through the surrounding plasma. The resulting force of friction may, under certain conditions, become significant and lead to a uniform (nonaccelerated) contraction of the wall. In this case, the potential energy of the wall is dissipated in the surrounding medium. Only when the wall would decrease to a certain small size R_f, will the internal forces proportional to the surface curvature dominate and the wall will again contract with acceleration to supply the necessary energy to the center sufficient to form a PBH.

The above considerations show that the energy concentrated in the course of wall contraction can be determined using the condition

$$E \approx 4\pi R^2 \sigma, \quad R = \min(R_h, R_w, R_f), \tag{18}$$

where $\sigma = 4\Lambda^2 f$ is the surface energy density of the wall. A condition of PBH formation is that

$$R_{\min} \sim d < r_g = 2E/m_{pl}^2. \tag{19}$$

It is assumed that the spherical wall contracts until reaching a size on the order of the wall thickness.

Let us determine the values of R_h, R_w, and R_h for a system with Lagrangian (1). Consider a domain with a certain phase appearing N e-folds before the end of inflation. The domain size at the end of inflation period is

$$l_e = H^{-1} e^N, \tag{20}$$

where H is the Hubble constant at that time instant. By the moment of crossing the horizon, the domain will expand to acquire the characteristic size

$$R_h \approx \frac{e^{2N}}{2HN}. \tag{21}$$

Below we assume that a visible part of the Universe is formed $N = N_U = 60$ e-folds before the end of inflation.

The second characteristic size R_w is determined from the condition that the wall energy (18) is equal to the energy of plasma contained in the domain bounded by the closed wall:

$$E_V = \rho \frac{4\pi}{3} R^3,$$

where

$$\rho = \frac{\pi^2}{30} g^* T^4$$

is the plasma density during the radiation dominated epoch. Taking into account that $\sigma = 4\Lambda^2 f$, we obtain the critical wall size corresponding to the domain separating from the general expansion:

$$R = R_{\text{crit}} = \frac{3\sigma}{\rho} = \frac{360\Lambda^2 f}{\pi^2 g^* T^4} \approx \frac{\Lambda^2 f}{T^4}. \tag{22}$$

As is known, the temperature in the Robertson–Walker Universe during the radiation-dominated epoch varies with time as

$$T = \left(\frac{45}{32\pi^2 g^*}\right)^{1/4} \sqrt{\frac{m_{pl}}{t}}. \tag{23}$$

Taking into account that an increase in the wall radius up to the instant of separation from the general expansion is proportional to the scaling factor

$$R(t) = l_e \sqrt{t/t_e}, \tag{24}$$

we arrive at the desired relationship for $R_w(N)$:

$$R_w = (2^8 \pi\sigma)^{-1/3} \left(\frac{m_{pl}}{H}\right)^{2/3} N^{-2/3} \exp\left(\frac{4N}{3}\right). \tag{25}$$

An expression for the characteristic wall size (radius)R_f above which the friction is significant can be obtained by equating the pressure developed by the internal forces $p_{\text{int}} = 2\sigma/R_f$ to that produced by the surrounding

medium on the moving wall. Using relationship (17) for the latter pressure, we obtain

$$R_f = \frac{\pi^7 \sigma d^4}{10\gamma}. \quad (26)$$

The above considerations do not take into account the effect of a gravity field on the wall dynamics. Therefore, the obtained relationships are valid provided that the initial wall size is much greater than the gravitational radius. Generally speaking, the gravitational radius may be comparable with (or even exceed) the wall size for a sufficiently large domain size. In this study, we deal with smaller domains for which the intrinsic gravity field does not affect the wall evolution. Now we proceed to the study of PBH cluster formation in the early evolution stage of the Universe.

5. CORRELATIONS IN THE BLACK HOLE DISTRIBUTION

Previously [11] we have studied a new process involving the formation of uncorrelated PBHs in the Universe. It was demonstrated that a model with reasonably selected parameters readily provides for the formation of 10^{11} massive (10^{30}–10^{40} g each) black holes, which is precisely equal to the number of galaxies in the visible Universe. In that analysis, we did not take into account correlations (inherent in this mechanism) between the formation of a massive black hole and the appearance of smaller black holes surrounding it. This correlation is related primarily to certain features of the above-discussed process of the formation of domains with the phases $\theta > \pi$. Apparently, the appearance of such domains creates prerequisites for the formation of new smaller domains inside.

Let us estimate the mass distribution of these daughter domains. Consider a region with a size on the order of H^{-1} and a phase within $\pi < \theta_0 < \pi + \delta$ (where $\delta = H/2\pi f$ is the average phase jump during the H^{-1} time period) formed during the inflation period as a result of fluctuation in a certain region of space with the phase $\theta < \pi$. During the next *e*-fold, this space domain will separate into e^3 subdomains and some of these will acquire a phase θ_1 in the interval $\pi - \delta < \theta_1 < \pi$. Upon the subsequent phase transition, these domains will be separated by walls from the external region. Similar transitions, with crossing the phase $\theta = \pi$ in the reverse direction, will take place in each subdomain during the next *e*-fold. Thus, a structure of the fractal type appears which reproduces itself in each time step on a decreasing scale.

Let ζ denote the number of subdomains formed in each step, around which a wall may form with time. Apparently, this value obeys the inequality $1 < \zeta \ll e^3$. In the subsequent estimates, we will assume that $\zeta \approx 2$–3. Since each causality-connected domain touches approximately six neighboring domains, we can hardly expect ζ to be greater for a total number of $\sim e^3 \approx 20$. The mass of the future black hole (if it would actually form) is determined by the area of a closed surface with the phase $\theta = \pi$. The ratio of areas of the initial (mother) and daughter domains is readily estimated: the initial area after a single *e*-fold is

$$S_0 \approx e^2 H^{-2},$$

and the daughter subdomain area is

$$S_1 \approx H^{-2}.$$

Therefore, the ratio of masses of the black holes belonging to two sequential generations is

$$M_j/M_{j+1} \approx S_j/S_{j+1} \approx e^2 \quad (27)$$

for their relative number assumed to be

$$N_{j+1}/N_j = \zeta. \quad (28)$$

As is readily seen, the number and mass of black holes appearing upon the *j*th *e*-fold after the initial domain formation are determined by parameters of the largest black hole genetically related to the primary domain in which the phase originally exceeded π. It is evident that

$$N_j \approx \zeta^j, \quad M_j \approx M_0/e^{2j}. \quad (29)$$

Excluding *j* from these relationships, we obtain the desired black hole mass distribution in a cluster:

$$N_{cl}(M) \approx (M_0/M)^{(1/2)\ln\zeta}. \quad (30)$$

The total mass of the cluster can be expressed through the mass M_0 of the largest initial black hole:

$$M_{\text{tot}} = M_0 + \zeta M_1 + \zeta^2 M_2 + \ldots = M_0 + \zeta e^{-2} M_0 + (\zeta e^{-2})^2 M_0 + \ldots = M_0[1 - \zeta/e^2]^{-1}. \quad (31)$$

As is seen, the total mass of the black hole cluster is only one and a half to two times greater than the largest initial black hole mass. The number of daughter black holes depends on the factors considered in the previous section.

6. DISCUSSION OF RESULTS

In the preceding sections, we considered only the principal possibility of the formation of domain walls connecting adjacent vacuum states. We have used the formulas derived above to estimate efficiency of the proposed mechanism of black hole cluster formation. The numerical calculations were performed for the following values of parameters (which are consistent with the observed anisotropy in the cosmic microwave radiation): the Hubble constant at the end of inflation, $H = 10^{13}$ GeV; Lagrangian parameters, $f = 1.77H$ and $\Lambda = 5$ GeV. The initial phase, at which the visible part

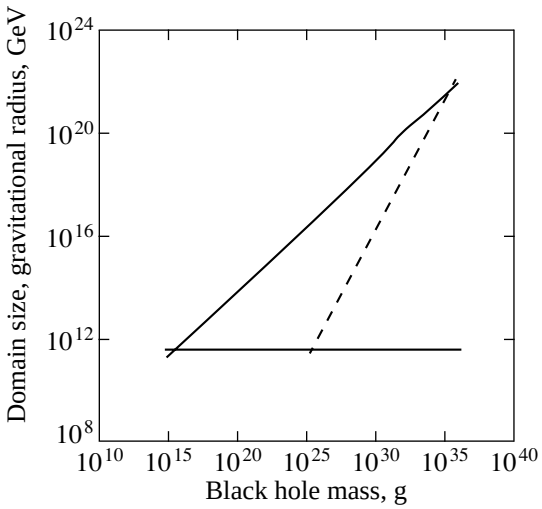


Fig. 1. Plots of the characteristic domain size (inclined solid line) and gravitational radius (dashed line) versus the domain mass. Horizontal line indicates the wall thickness. See the text for numerical values of parameters.

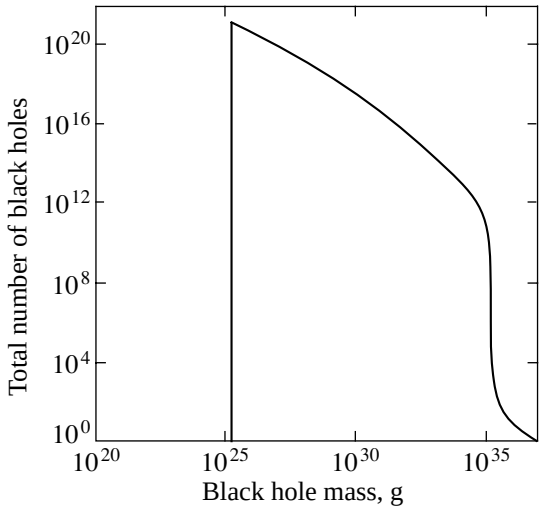


Fig. 2. Primordial black hole mass distribution in the Universe.

of the Universe is formed by the time $t_U \approx 60H^{-1}$ to the end of inflation controls the number of domains and, accordingly, the number of closed walls formed in the postinflation stage. This random value, not related to the Lagrangian parameters, must be selected taking into account the data of observations on the abundance of black holes in the Universe. We will use the numerical value $\theta_U = 0.05\pi$, which ensures a sufficiently large number of massive black holes, while the presence of numerous smaller black holes does not contradict experimental restrictions.

Figure 1 shows the results of numerical calculations constructed in a logarithmic scale. The bottom horizontal line shows the wall thickness. As is seen, the condition of wall existence (the characteristic domain size must exceed the wall thickness) is fulfilled for the domains with masses exceeding 10^{15} g. The domains of lower energies possess (for the parameters selected) dimensions below the wall thickness. This implies that the wall formation is impossible and the domain exhibits only fluctuations in the energy density. During contraction, the wall energy is eventually completely converted into radiation.

The necessary condition for black hole formation as a result of the domain wall collapsing is that the gravitational radius of the domain must be greater than the wall thickness. As is seen from Fig. 1, this condition is fulfilled for black holes with masses $\geq 10^{25}$ g. Therefore, the proposed mechanism of black hole formation leads to a nontrivial situation: massive PBHs exist in the complete absence of black holes possessing masses below this threshold. Note that the most significant observational restrictions concerning the PBH abundance refer to the mass region $\sim 10^{15}$ g (which cannot form for the parameters selected). Since the gravitational radius is proportional to the wall surface area, the plots corresponding to the domain size and its gravitational radius must intersect. This intersection actually takes pace at a wall mass of $\sim 10^{35}$ g. For greater masses, the gravity effects have to be taken from the very beginning, which will limit from above the possible PBH masses. According to formula (26), the friction becomes significant only for supermassive walls, the number of which is negligibly small.

Figure 2 shows the PBH mass distribution calculated for the selected parameters (see also the discussion in [11]). As is seen, the PBH masses fall within the range from 10^{25} to 10^{35} g. The initial phase θ_U was selected so as to provide that the number of massive PBHs ($\sim 10^{35}$ g) was equal to the number of galaxies in the visible part of the Universe. The total mass of black holes amounts to ~1% of the contemporary baryonic contribution.

The results of calculations are sensitive to changes in the parameter Λ and the initial phase θ_U. As the Λ value decreases to ≈ 1 GeV, still greater PBHs appear with a mass of up to $\sim 10^{40}$ g. A change in the initial phase leads to sharp variations in the total number of black holes. As was shown in Section 5, each domain generates a family of subdomains in close vicinity. The total mass of such a cluster is only 1.5–2 times that of the largest initial black hole in this space region. Thus, our calculations confirm the possibility of formation of the clusters of massive PBHs (~100 $M_\odot$ and above) in the earliest stages of evolution of the Universe at a temperature of 1–10 GeV. These clusters represent stable energy density fluctuations around which increased baryonic density may concentrate in the subsequent stages, followed by the evolution into galaxies.

7. CONCLUSION

This paper proposes a new mechanism for the formation of protogalaxies, which is based on the cosmological inferences of the elementary particle models predicting nonequilibrium second-order phase transition in the inflation stage period and the domain wall formation upon the end of inflation. The presence of closed domain walls with the size markedly exceeding the cosmological horizon at the instant of their formation leads to the wall collapse in the postinflation epoch (when the wall size becomes comparable with the cosmological horizon), which results in the formation of massive black hole clusters that can serve as nuclei for the future galaxies. The black hole mass distributions are calculated, which do not contradict the available experimental data. The number of black holes with $M \sim 100\ M_\odot$ and above is comparable with the number of galaxies in the visible Universe. A mechanism of deceleration of the wall motion is considered, and it is shown that this process may affect only the dynamics of collapse of supermassive walls.

Development of the proposed approach gives grounds for a principally new scenario of galaxy formation in the model of a hot Universe. Traditionally, the hot Universe model assumes a homogeneous distribution of matter on all scales, whereas the appearance of observed inhomogeneities is related to the growth of small initial density perturbations. However, an analysis of the cosmological inferences of the theory of elementary particles indicates the possible existence of strongly inhomogeneous primordial structures in the distribution of both the latent mass and baryons. These primordial structures represent a new factor in the theory of galaxy formation. Topological defects such as the cosmological walls and filaments, primordial black holes, archioles in the models of axion cold latent mass [19, 20], and essentially inhomogeneous baryosynthesis (leading to the formation of antimatter domains in the baryon-asymmetric Universe) [12, 21] offer by no means a complete list of possible primary inhomogeneities inferred from the existing elementary particle models.

The proposed approach discloses a number of interesting aspects in this direction. Indeed, this model provides for a possibility of the quantitative analysis of correlations in the formation of massive PBHs and the primary inhomogeneity of the latent mass and baryons. Originally inherent in this mechanism is the inhomogeneous phase distribution, which eventually acquires (similar to what takes place in the invisible axion cosmology) a dynamical sense of the initial amplitude of the coherent oscillations of a scalar field. Irrespective of the efficiency of dissipation of the energy of these oscillations, the regions of closed wall formation must be correlated with the regions of maximum energy density of the latent mass. If these oscillations are not decaying, their energy density may provide for the contemporary latent mass density. Inhomogeneity in the initial amplitude of these oscillations would then imply an inhomogeneity in the initial energy density and, hence, regions of black hole formation will become the regions of increased latent mass density. A qualitatively similar effect (albeit not as pronounced) takes place in the dissipation of coherent oscillations at the expense of particle production. An increase in the oscillation energy density transforms into a local increase in the density of latent mass particles produced in this region.

In the class of spontaneous baryosynthesis models, a change in the phase determines the production of excess baryons. Therefore, in addition to an increase in the latent mass density, the regions of massive PBH formation may be characterized by a higher baryon density. Inside a closed wall, where the phase is $\theta > \pi$, the same mechanism leads to the production of excess antibaryons [12]. However, this antimatter domain will survive only provided that its size is sufficiently large [22]. Thus, development of the proposed approach may lead to a number of interesting scenarios of initial stages in the formation of protogalaxies, depending on the selection of particular elementary particle models and their parameters. This study presents the first step in this direction.

ACKNOWLEDGMENTS.

The authors are grateful to E.D. Zhizhin for fruitful discussion of the results and to A.A. Starobinsky for his interest in this study. The work of S.G.R. and M.Yu.Kh. was partly performed within the framework of the "Cosmoparticle Physics" Project of the State Scientific-Technological Program "Astronomy: Fundamental Space Research" supported by the Cosmion–ETHZ and Epcos–AMS international cooperation (Scientific School Support Grant no. 00-15-96699).

REFERENCES

1. D. E. Rosenberg and J. R. Rutgers, astro-ph/0012023 (2000).
2. S. Veilleux, astro-ph/0012121 (2000).
3. M. Stiavelli, astro-ph/9801021 (1998).
4. M. R. Merrifield, D. A. Forbes, and A. I. Terlevich, astro-ph/0002350 (2000).
5. I. D. Novikov, A. G. Polnarev, A. A. Starobinskiĭ, and Ya. B. Zel'dovich, Astron. Astrophys. **80**, 104 (1979).
6. A. G. Polnarev and M. Yu. Khlopov, Usp. Fiz. Nauk **143**, 369 (1985) [Sov. Phys. Usp. **18**, 213 (1985)].
7. Ya. B. Zel'dovich, Zh. Éksp. Teor. Fiz. **42**, 641 (1962) [Sov. Phys. JETP **15**, 446 (1962)].
8. M. Yu. Khlopov, Soobshch. Spets. Astrofiz. Obs. **6**, 7 (1987).
9. J. Yokoyama, astro-ph/9802357 (1998).
10. Hee Il Kim, Phys. Rev. D **62**, 063504 (2000); astro-ph/9907372 (1999).
11. S. G. Rubin, M. Yu. Khlopov, and A. S. Sakharov, Gravitation Cosmology, Suppl. **6**, 51 (2000).

12. M. Yu. Khlopov, S. G. Rubin, and A. S. Sakharov, Phys. Rev. D **62**, 083505 (2000).
13. A. A. Starobinskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **30**, 719 (1979) [JETP Lett. **30**, 682 (1979)].
14. A. Vilenkin and L. Ford, Phys. Rev. D **26**, 1231 (1982).
15. A. D. Linde, Phys. Lett. B **116**, 335 (1982).
16. A. Starobinsky, Phys. Lett. B **117**, 175 (1982).
17. R. Rajaraman, *Solitons and Instantons* (North-Holland, Amsterdam, 1982).
18. B. H. Liu, L. McLerran, and N. Turok, Phys. Rev. D **46**, 2668 (1992).
19. A. S. Sakharov and M. Yu. Khlopov, Yad. Fiz. **57**, 514 (1994) [Phys. At. Nucl. **57**, 485 (1994)].
20. A. S. Sakharov, D. D. Sokolov, and M. Yu. Khlopov, Yad. Fiz. **59**, 1050 (1996) [Phys. At. Nucl. **59**, 1005 (1996)].
21. M. Yu. Khlopov, *Cosmoparticle Physics* (World Scientific, Singapore, 1999).
22. M. Yu. Khlopov *et al.*, Astropart. Phys. **12**, 367 (2000).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 930–939.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1077–1086.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Tagged-Photon Events in Polarized Deep Inelastic Scattering¶

G. I. Gakh, M. I. Konchatnij, and N. P. Merenkov*
Kharkov Institute of Physics and Technology, National Scientific Center, Kharkov, 61108 Ukraine
**e-mail: merenkov@kipt.kharkov.ua*
Received October 31, 2000

Abstract—Deep inelastic scattering events of a longitudinally polarized electron by a polarized proton with a tagged collinear photon radiated from the initial-state electron are considered. The corresponding cross section is derived in the Born approximation. The model-independent radiative corrections to the Born cross section are also calculated. The obtained result is applied to the elastic scattering. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The idea of using radiative events (events with the emission of an additional tagged photon) in lepton–hadron interaction to expand the experimental possibilities for studying different topics in high-energy physics has become quite attractive recently.

Photon radiation from the initial e^+e^--state in the events with missing energy has been successfully used at LEP in measuring the number of light neutrinos and in searching for new physical signals (for a recent publication, see, e.g., [1]). The possibility to undertake the bottonium spectroscopy studies at *B*-factories by using the hard photon emission from the electron or the positron was considered in [2]. An important physical problem of the total hadronic cross section scanning in the electron-positron annihilation process at low and intermediate energies by means of the initial-state radiative events was extensively discussed in [3].

The initial-state collinear radiation is very important in certain regions of the deep inelastic scattering (DIS) at the HERA kinematic domain. It leads to a reduction of the projectile electron energy and therefore, to a shift of the effective Bjorken variables in the hard scattering process compared to those determined from the actual measurement of the scattered electron alone. That is why the radiative events in the DIS process

$$e^-(k_1) + p(p_1) \longrightarrow e^-(k_2) + \gamma(k) + X(p_x) \qquad (1)$$

must be carefully taken into account [4].

In addition, the measurement of the energy of the photon emitted very close to the incident electron beam direction [5, 6] allows studying the overlap of the kinematical photoproduction region $Q^2 = -(k_1 - k_2)^2 \approx 0$ and the DIS region with small transferred momenta (Q^2 about several GeV^2) within the high-energy HERA experiments. These radiative events can also be used for independently determining the proton structure functions F_1 and F_2 in a single run without lowering the beam energy [5, 7]. The high-precision calculation of the corresponding cross section (taking the radiative corrections (RC) into account) was performed in [8].

In this paper, we investigate the events for deep-inelastic radiative process (1) with a longitudinally polarized electron beam and a polarized proton as a target. As in [8], we suggest that the hard photon is emitted very close to the direction of the incoming electron beam ($\theta_\gamma = \widehat{\mathbf{p}_1\mathbf{k}_1} \le \theta_0$ with $\theta_0 \ll 1$) and the photon detector (PD) measures the energy of all photons inside a narrow cone with the opening angle $2\theta_0$ around the electron beam. The scattered electron 3-momentum is fixed simultaneously.

We consider the longitudinal (along the electron beam direction) and perpendicular (in the plane ($\mathbf{k}_1$, $\mathbf{k}_2$)) polarizations of the proton. In Section 2, we derive the corresponding cross sections in the Born approximation, and in Section 3, we calculate the different RC contributions to the Born cross section. The total radiative correction for different (exclusive and calorimeter) experimental conditions for the scattered electron measurement is given in Section 4. Our results can be applied to the cross section of process (1). We consider the target proton at rest and also the colliding electron-proton beams. In Section 5, we apply the results obtained in Section 4 to describe the quasi-elastic scattering using the relation between the spin-dependent proton structure functions and the proton electromagnetic form factors in this limiting case.

2. BORN APPROXIMATION

The spin-independent part of the DIS cross section for the experimental setup considered here was recently investigated in detail [8]. We now consider the spin-dependent part of the corresponding cross section that is described by the proton structure functions g_1 and g_2. Because the opening angle of the forward PD is very

¶This article was submitted by the authors in English.

small and we consider only the cross section where the tagged photon is integrated over the solid angle covered by the PD, we can apply the quasi-real electron method [9] and parametrize these radiative events using the standard Bjorken variables

$$x = \frac{Q^2}{2p_1(k_1-k_2)}, \quad y = \frac{2p_1(k_1-k_2)}{V}, \qquad V = 2p_1k_1, \tag{2}$$

and the energy fraction of the electron after the initial-state radiation of a collinear photon

$$z = \frac{2p_1(k_1-k)}{V} = \frac{\varepsilon-\omega}{\varepsilon}, \tag{3}$$

where ε is the initial-electron energy and ω is the energy deposited in the PD.

An alternative set of kinematic variables that is specially adapted to the collinear photon radiation is given by the shifted Bjorken variables [5, 10]

$$\hat{Q}^2 = -(k_1-k_2-k)^2, \quad \hat{x} = \frac{\hat{Q}^2}{2p_1(k_1-k_2-k)}, \qquad \hat{y} = \frac{2p_1(k_1-k_2-k)}{2p_1(k_1-k)}. \tag{4}$$

The shifted and standard Bjorken variables are related by

$$\hat{Q}^2 = zQ^2, \quad \hat{x} = \frac{xyz}{z+y-1}, \quad \hat{y} = \frac{z+y-1}{z}. \tag{5}$$

At fixed values of x and y, the lower limit of z can be derived from the constraint on the shifted variable $\hat{x}$,

$$\hat{x} < 1 \longrightarrow z > \frac{1-y}{1-xy}.$$

In the Born approximation, we determine the DIS cross section in radiative process (1) in terms of the contraction of the leptonic and hadronic tensors as[1]

$$\frac{d\sigma}{\hat{y}d\hat{x}d\hat{y}} = \frac{4\pi\alpha^2(\hat{Q}^2)}{\hat{Q}^4} z L^B_{\mu\nu}H_{\mu\nu}, \tag{6}$$

where $\alpha(\hat{Q}^2)$ is the running electromagnetic coupling constant that takes the vacuum polarization effects into account and the Born leptonic current tensor is given by [11]

$$L^B_{\mu\nu} = \frac{\alpha}{4\pi^2}\int_\Omega 2i\varepsilon_{\mu\nu\lambda\rho}q_\lambda(k_{1\rho}R_t + k_{2\rho}R_s)\frac{d^3k}{\omega}, \qquad q = k_1-k_2-k, \tag{7}$$

where Ω covers the solid angle of the PD.

[1] In what follows, we are only interested in the spin-dependent part of the cross section.

For the initial-state collinear radiation considered in this paper, the quantities R_t and R_s can be written as

$$R_t = -\frac{1}{(1-z)t} - \frac{2m^2}{t^2}, \qquad R_s = -\frac{z}{(1-z)t} + \frac{2m^2(1-z)}{t^2}, \qquad t = -2kk_1, \quad q = zk_1 - k_2. \tag{8}$$

In accordance with the quasi-real electron approximation [9], the trivial angular integration of the Born leptonic tensor gives

$$L^B_{\mu\nu} = \frac{\alpha}{2\pi}P(z,L_0)dzi\varepsilon_{\mu\nu\lambda\rho}q_\lambda k_{1\rho}, \quad L_0 = \ln\frac{\varepsilon^2\theta_0^2}{m^2}, \qquad P(z,L_0) = \frac{1+z^2}{1-z^2}L_0 - \frac{2(1-z+z^2)}{1-z}, \tag{9}$$

where m is the electron mass.

We write the spin-dependent part of the hadronic tensor in the right-hand side of Eq. (6) as

$$H_{\mu\nu} = -iM\frac{\varepsilon_{\mu\nu\lambda\rho}q_\lambda}{2p_1q}\left[(g_1+g_2)S_\rho - g_2\frac{Sq}{p_1q}p_{1\rho}\right], \tag{10}$$

where M is the proton mass and S is the proton polarization 4-vector. In writing expressions (7) and (10), we assume that the polarization degree of both the electron and the proton is equal to 1.

Our normalization is such that the proton structure functions g_1 and g_2 are dimensionless and in the limiting case of the elastic scattering ($\hat{x} \longrightarrow 1$) they are expressed in terms of the proton electric (G_E) and magnetic (G_M) form factors as

$$g_1(\hat{x},\hat{Q}^2) \longrightarrow \delta(1-\hat{x}) \times\left[G_MG_E + \frac{\lambda}{1+\lambda}(G_M-G_E)G_M\right], \qquad \lambda = \frac{\hat{Q}^2}{4M^2}, \tag{11}$$

$$g_2(\hat{x},\hat{Q}^2) \longrightarrow -\delta(1-\hat{x})\frac{\lambda}{1+\lambda}(G_M-G_E)G_M, \qquad G_{M,E} = G_{M,E}(\hat{Q}^2).$$

It is convenient to parametrize the proton polarization 4-vector in terms of the 4-momenta of the reaction under study [12],

$$S^\parallel_\mu = \frac{2M^2k_{1\mu} - Vp_{1\mu}}{MV}, \qquad S^\perp_\mu = \frac{up_{1\mu} + Vk_{2\mu} - [2u\tau + V(1-y)]k_{1\mu}}{\sqrt{-uV^2(1-y) - u^2M^2}}, \tag{12}$$

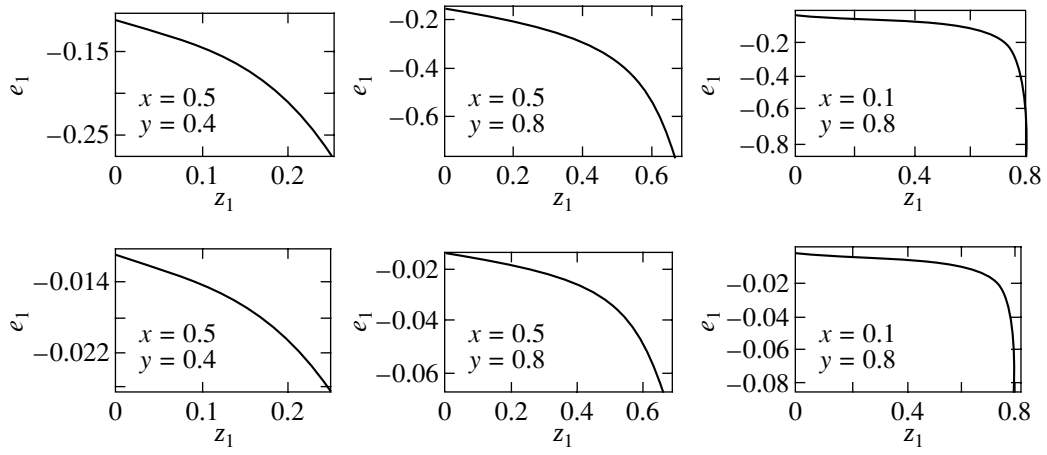


Fig. 1. The dependence of e_1 on the energy fraction of the tagged photon $z_1 = 1 - z$ for different values of x, y, and V. The upper set corresponds to $V = 10$ GeV^2 and the lower one to $V = 100$ GeV^2. The maximum value of z_1 is $y(1 - x)/(1 - xy)$.

where $u = -Q^2$, $\tau = M^2/V$, and we neglect the electron mass. The 4-vector of the longitudinal proton polarization has the respective components

$$S_\mu^\parallel = (0, \mathbf{n}_1), \quad S_\mu^\parallel = \left(-\frac{|\mathbf{p}_1|}{M}, \frac{\mathbf{n}_1 E_1}{M}\right) \tag{13}$$

for the target at rest and the colliding beams. Here, $E_1(\mathbf{p}_1)$ is the proton energy (3-momentum) and $\mathbf{n}_1$ is the unit vector along the initial electron 3-momentum direction. The 4-vector of the perpendicular proton polarization $S_\mu^\perp$ is the same in both these cases,

$$S_\mu^\perp = \left(0, \frac{\mathbf{n}_2 - \mathbf{n}_1(\mathbf{n}_1 \cdot \mathbf{n}_2)}{\sqrt{1 - (\mathbf{n}_1 \cdot \mathbf{n}_2)^2}}\right), \tag{14}$$

where $\mathbf{n}_2$ is the unit vector along the scattered electron 3-momentum direction. It is easy to verify that $S^\parallel S^\perp = 0$.

Using the definitions of the DIS cross section in Eq. (6), leptonic and hadronic tensors (9) and (10), and the parametrization of the proton polarization in Eq. (12), after simple calculations we derive a spin-dependent part of the cross section of process (1) with a tagged collinear photon radiated from the initial state,

$$\frac{d\sigma^B_{\parallel,\perp}}{\hat{y}d\hat{x}d\hat{y}dz} = \frac{\alpha}{2\pi}P(z, L_0)\Sigma_{(\parallel,\perp)}(\hat{x}, \hat{y}, \hat{Q}^2), \tag{15}$$

$$\Sigma_\parallel = \frac{4\pi\alpha^2(\hat{Q}^2)}{\hat{V}\hat{y}}\left(\hat{\tau} - \frac{2 - \hat{y}}{2\hat{x}\hat{y}}\right) \times g_1(\hat{x}, \hat{Q}^2)[1 + e_1\hat{R}(\hat{x}, \hat{Q}^2)], \tag{16}$$

$$\Sigma_\perp = -\frac{4\pi\alpha^2(\hat{Q}^2)}{\hat{V}\hat{y}}\sqrt{\frac{M^2}{\hat{Q}^2}(1 - \hat{y} - \hat{x}\hat{y}\hat{\tau})} \times g_1(\hat{x}, \hat{Q}^2)[1 + e_2\hat{R}(\hat{x}, \hat{Q}^2)], \tag{17}$$

where

$$e_1 = \frac{4\hat{\tau}\hat{x}}{2\hat{x}\hat{y}\hat{\tau} + \hat{y} - 2}, \quad e_2 = \frac{2}{\hat{y}},$$

$$\hat{R} = \frac{g_2(\hat{x}, \hat{Q}^2)}{g_1(\hat{x}, \hat{Q}^2)}, \quad \hat{\tau} = \frac{M^2}{\hat{V}}, \quad \hat{V} = zV.$$

It is helpful to recall that the unpolarized DIS cross section is proportional to $\sigma_T(1 + eR)$, where $R = \sigma_L/\sigma_T$. For the events with the tagged collinear photon [5], we have

$$e = \frac{2(1 - \hat{y})}{1 + (1 - \hat{y})^2}.$$

Because the quantities e_1 and e_2 strongly depend on z, the determination of the proton structure functions g_1 and g_2 is possible by measuring the z-dependence of cross section (15) in a single run without lowering the electron beam energy. The quantity e_1 is proportional to τ and is therefore very small under HERA conditions. Thus, the separation of g_1 and g_2 in the DIS process with the longitudinally polarized proton is possible in experiments with the target at rest and low values of V (up to 20 GeV^2). At HERA, the cross section of this process can be used for measuring the structure function g_1 only. This can be seen in Fig. 1. On the other hand, Fig. 2 shows that the experiments with the tagged photon and the perpendicular proton polarization can be used to measure both g_1 and g_2 in a wide range of energies (provided that Q^2 is not large).

3. RADIATIVE CORRECTIONS

We restrict ourselves to the model-independent QED radiative corrections related to the radiation of the real and virtual photons by leptons. The remaining sources of RC in the same order of the perturbation the-

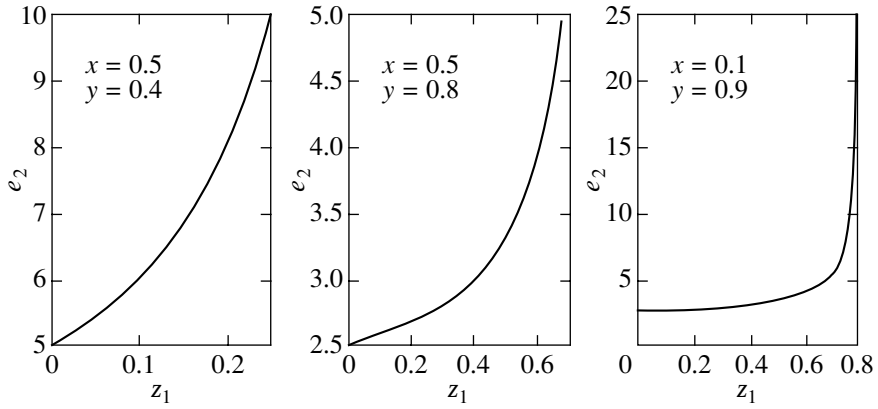


Fig. 2. The quantity e_2 at different values of x and y as a function of $z_1 = 1 - z$.

ory, such as the virtual corrections with a double photon exchange mechanism and the bremsstrahlung of the proton and partons, are more involved and model-dependent. They are not considered here. Our approach to the calculation of the RC is based on taking into account all the essential Feyman diagrams that describe the observed cross section in the chosen approximation. To avoid cumbersome expressions, we retain the terms accompanied by at least one power of large logarithms in the RC. In the considered case, three different types of such logarithms appear,

$$L_0, \quad L_Q = \ln\frac{Q^2}{m^2}, \quad L_\theta = \ln\frac{\theta_0^2}{4}. \tag{18}$$

In the chosen approximation, we additionally neglect the terms of the second order θ_0^2, $m^2/\varepsilon^2\theta_0^2$, and m^2/Q^2 in the cross section.

The total RC to cross section (15) includes the contributions of the virtual and soft photon emission and also the hard photon radiation.

As one can see, we use the standard gauge-invariant expression for the hadronic tensor. The leptonic tensor was calculated in accordance with the QED rules. The complete set of Feynman diagrams for the calculation of the radiative correction caused by the real photon emission is taken account. Taking the loop correction into account involves the gauge invariant method for solving both the infrared and the ultraviolet divergence problems The results obtained in our paper are therefore gauge-invariant. We begin with calculating the virtual and soft corrections.

3.1. Virtual and Soft Corrections

To calculate the virtual- and soft-photon emission corrections, we start from the expression for the one-loop corrected Compton tensor with a heavy photon for the longitudinally polarized electron [13]. For the hard collinear initial-state radiation considered here, this Compton tensor can be written as

$$L_{\mu\nu}^V = \frac{\alpha}{2\pi}\rho L_{\mu\nu}^B + \frac{\alpha^2}{4\pi^3}\int_\Omega i\varepsilon_{\mu\nu\lambda\rho}q_\lambda k_{1\rho}\frac{d^3k}{\omega}$$

$$\times\left[\frac{T}{-t} + \frac{4m^2(1-z+z^2)}{t^2}L_Q\ln z\right],$$

$$T = \frac{1+z^2}{1-z^2}\{2\ln z[l_t - \ln(1-z) - L_Q] - 2F(z)\}$$

$$+ \frac{1+2z-z^2}{2(1-z)}, \tag{19}$$

$$F(z) = \int_1^{1/z}\frac{dx}{x}\ln|1-x|,$$

$$l_t = \ln\frac{-t}{m^2},$$

$$\rho = 4(L_Q - 1)\ln\frac{\delta}{m} - L_Q^2 + 3L_Q + 3\ln z + \frac{\pi^2}{3} - \frac{9}{2},$$

where δ is the fictitious photon mass and the tensor $L_{\mu\nu}^B$ is defined in Eq. (9).

To eliminate the photon mass, we must add the contribution of the additional soft photon emission with an energy less than $\Delta\varepsilon$, $\Delta \ll 1$. This contribution was found in [14] and the corresponding procedure of the photon mass elimination was described in [15]. The result is

$$L_{\mu\nu}^{V+S} = L_{\mu\nu}^V(\rho \longrightarrow \tilde{\rho}),$$

$$\tilde{\rho} = 2(L_Q - 1)\ln\frac{\Delta^2}{Y} + 3L_Q + 3\ln z - \ln^2 Y - \frac{\pi^2}{3} - \frac{9}{2} + 2\mathrm{Li}_2\left(\cos^2\frac{\theta}{2}\right), \quad Y = \frac{\varepsilon^2}{\varepsilon}, \tag{20}$$

where ε_2 is the scattered electron energy and θ is the electron scattering angle ($\theta = \widehat{\mathbf{k}_1\mathbf{k}_2}$).

The angular integration with respect to the hard tagged photon over the solid angle of the PD gives (within the chosen accuracy)

$$L_{\mu\nu}^{V+S} = \left(\frac{\alpha}{2\pi}\right)^2 [\tilde{\rho}P(z, L_0) + G]dz i\varepsilon_{\mu\nu\lambda\rho}q_\lambda k_{1\rho},$$

$$G = \left\{\frac{1+z^2}{1-z^2}[\ln z(L_0 - 2L_Q) - 2F(z)] + \frac{1+2z-z^2}{2(1-z)}\right\}L_0 + \frac{4(1-z+z^2)}{1-z}L_Q\ln z. \tag{21}$$

Using the right-hand side of Eq. (21) instead of $L_{\mu\nu}^B$ in the right side of Eq. (6), we derive the contribution of the virtual and soft corrections to Born cross section (15) as

$$\frac{d\sigma_{\parallel,\perp}^{V+S}}{\hat{y}d\hat{y}d\hat{x}dz} = \left(\frac{\alpha}{2\pi}\right)^2 [\tilde{\rho}P(z, L_0) + G]\Sigma_{\parallel,\perp}(\hat{x}, \hat{y}, \hat{Q}^2), \tag{22}$$

where $\Sigma_{\parallel,\perp}(\hat{x}, \hat{y}, \hat{Q}^2)$ are defined in Eqs. (16) and (17).

3.2. Double Hard Bremsstrahlung

We now consider the emission of an additional hard photon with the 4-momentum $\tilde{k}$ and an energy higher than $\Delta\varepsilon$. To calculate the contribution from the real hard bremsstrahlung, which in our case corresponds to the double hard photon emission with at least one photon seen in the forward PD, we specify three kinematical domains:

(i) both hard photons hit the forward PD—i.e., both are emitted within a narrow cone around the electron beam $\widehat{\mathbf{k}\mathbf{k}_1}$, $\widehat{\tilde{\mathbf{k}}\mathbf{k}_1} \le \theta_0$;

(ii) one hard photon is tagged by the PD and the other is collinear to the outgoing electron momentum $\widehat{\tilde{\mathbf{k}}\mathbf{k}_2} \le \theta_0'$, $\theta_0' \ll 1$;

(iii) an additional photon is emitted at large angles (i.e., outside both narrow cones defined above) with respect to both incoming and outgoing electron momenta.

The contributions of regions (i) and (ii) contain quadratic terms in the large logarithms L_0 and L_Q, whereas region (iii) contains terms of the order L_0L_θ, which can give an even larger numerical contribution if $2\theta_0 > \varepsilon\theta_0/m$.

We refer to the third kinematical region as the semicollinear one. Beyond the leading logarithmic accuracy, the calculation can be performed using the results in [16] for the leptonic current tensor with the longitudinally polarized electron for the collinear as well as semicollinear regions.

The contribution of kinematical region (i), where both hard photons hit the PD and each has an energy higher than $\Delta\varepsilon$, can be written as

$$\frac{d\sigma_{\parallel,\perp}^{i)}}{\hat{y}d\hat{y}d\hat{x}dz} = \left(\frac{\alpha}{2\pi}\right)^2 L_0 \times \left\{\left[\frac{1}{2}P_\theta^{(2)}(z) + \frac{1+z^2}{1-z}\left(\ln z - \frac{3}{2} - 2\ln\Delta\right)\right]L_0 + 7(1-z) - 2(1-z)\ln z + \frac{3+z^2}{2(1-z)}\ln^2 z - 2\frac{3-2z+3z^2}{1-z}\ln\frac{1-z}{\Delta}\right\}\Sigma_{\parallel,\perp}(\hat{x}, \hat{y}, \hat{Q}^2). \tag{23}$$

The double-logarithm terms in the right side of Eq. (23) are the same for the polarized and unpolarized cases, whereas the one-logarithm terms are different. In Eq. (23), we use the notation $P_\theta^{(2)}(z)$ for the Θ part of the second-order electron structure function $D(z, L)$ [17],

$$D(z, L) = \delta(1-z) + \frac{\alpha}{2\pi}P^{(1)}(z)L + \frac{1}{2}\left(\frac{\alpha}{2\pi}\right)^2 P^{(2)}(z)L^2 + \ldots,$$

$$P^{(i)}(z) = P_\theta^{(i)}(z)\Theta(1-z-\Delta) + \delta(1-z)P_\delta^{(i)}, \quad \Delta \longrightarrow 0, \tag{24}$$

$$P_\theta^{(1)}(z) = \frac{1+z^2}{1-z}, \quad P_\delta^{(1)} = \frac{3}{2} + 2\ln\Delta,$$

$$P_\theta^{(2)}(z) = 2\left[\frac{1+z^2}{1-z}\left(2\ln(1-z) - \ln z + \frac{3}{2}\right) + \frac{1}{2}(1+z)\ln z - 1 + z\right].$$

To calculate the contribution of kinematical region (ii), we can use the quasi-real electron method to describe the radiation of both collinear photons. This contribution to the observed cross section depends on the event selection, in other words, on the method of measuring the scattered electron.

For the exclusive event selection, where only the scattered electron is detected, but the photon emitted almost collinearly (i.e., within the opening angle $2\theta_0'$ around the scattered electron momentum) goes unnoticed or is not taken into account in calculating the kinematical variables, we have, in accordance with [9],

$$\frac{d\sigma_{\parallel,\perp}^{ii),\,\mathrm{excl}}}{\hat{y}d\hat{y}d\hat{x}dz} = \frac{\alpha^2}{4\pi^2}P(z, L_0) \times \int\limits_{\Delta/Y}^{y_{1\max}} \frac{dy_1}{1+y_1}\left[\frac{1+(1+y_1)^2}{y_1}(\tilde{L}-1)+y_1\right] \times \Sigma_{\parallel,\perp}(x_s, y_s, Q_s^2), \tag{25}$$

where y_1 is the energy fraction of the photon radiated along the 3-momentum $\mathbf{k}_2$ relative to the scattered electron energy ($y_1 = \tilde{\omega}/\varepsilon_2$) and

$$\tilde{L} = \ln\frac{\varepsilon^2\theta_0'^2}{m^2} + 2\ln Y,$$

$$x_s = \frac{xyz(1+y_1)}{z-(1-y)(1+y_1)},$$

$$y_s = \frac{z-(1-y)(1+y_1)}{z}, \quad Q_s^2 = Q^2 z(1+y_1).$$

The upper integration limit in Eq. (25) can be found from the condition of the inelastic process occurrence $p_x^2 = (M+\mu)^2$, where μ is the pion mass. Taking into account that $q = zk_1 - (1+y_1)k_2$ for kinematics (ii), we obtain

$$y_{1\max} = \frac{2z\varepsilon[M-\varepsilon_2(1-c)] - 2M\varepsilon_2 - \mu^2 - 2M\mu}{2\varepsilon_2[M+z\varepsilon(1-c)]}$$

for the proton target at rest and

$$y_{1\max} = \frac{2z - Y(1+c)}{Y(1+c)}$$

for the HERA collider, where $c = \cos\theta$. In writing this limit for HERA, we neglect the electron energy and the proton mass compared to the proton beam energy. We note that for the exclusive event selection, the parameter θ_0' is purely auxiliary and does not enter the final result when the contribution of region (iii) is added.

From the experimental point of view, a more realistic measurement method is the calorimeter event selection, where the photon and the electron cannot be distinguished inside a narrow cone with the opening angle $2\theta_0'$ along the outgoing electron momentum direction. Therefore, only the sum of the photon and electron energies can be measured if the photon belongs to this cone. In this case, we obtain

$$\frac{d\sigma_{\parallel,\perp}^{ii),\,\mathrm{cal}}}{\hat{y}d\hat{y}d\hat{x}dz} = \frac{\alpha^2}{4\pi^2}P(z, L_0) \times \int\limits_{\Delta/Y}^{\infty} \frac{dy_1}{(1+y_1)^3}\left[\frac{1+(1+y_1)^2}{y_1}(\tilde{L}-1)+y_1\right] \times \Sigma_{\parallel,\perp}(\hat{x}, \hat{y}, \hat{Q}^2) = \frac{\alpha^2}{4\pi^2}P(z, L_0)\left[(\tilde{L}-1)\left(2\ln\frac{Y}{\Delta}-\frac{3}{2}\right)+\frac{1}{2}\right] \times \Sigma_{\parallel,\perp}(\hat{x}, \hat{y}, \hat{Q}^2). \tag{26}$$

For the calorimeter event selection, the parameter θ_0' is physical and the final result depends on it (see below).

To calculate the contribution of region (iii), we can use the quasi-real electron method [9] and write the leptonic tensor in this region (which describes the collinear photon radiation with the energy fraction $1-z$ and the noncollinear photon radiation with the 4-momentum $\tilde{k}$) as

$$L_{\mu\nu}(k_1, k_2, (1-z)k_1, \tilde{k}) = \frac{\alpha}{2\pi}P(z, L_0)\frac{dz}{z}L_{\mu\nu}(zk_1, k_2, \tilde{k}), \tag{27}$$

$$L_{\mu\nu}(zk_1, k_2, \tilde{k}) = \frac{\alpha}{4\pi^2}\frac{d^3\tilde{k}}{\tilde{\omega}}L_{\mu\nu}^{\gamma}(zk_1, k_2, \tilde{k}),$$

$$L_{\mu\nu}^{\gamma}(zk_1, k_2, \tilde{k}) = 2i\varepsilon_{\mu\nu\lambda\rho}\tilde{q}_\lambda\left[\frac{(\tilde{u}+\tilde{t})z}{\tilde{s}\tilde{t}}k_{1\rho} + \frac{\tilde{s}+\tilde{u}}{\tilde{s}\tilde{t}}k_{2\rho}\right],$$

$$\tilde{q} = zk_1 - k_2 - \tilde{k}, \quad \tilde{u} = -2zk_2k_1,$$

$$\tilde{s} = 2\tilde{k}k_2, \quad \tilde{t} = -2z\tilde{k}k_1.$$

In the general case of the noncollinear photon radiation with the 4-momentum k, the contraction of the leptonic tensor $L_{\mu\nu}^{\gamma}(k_1, k_2, k)$ and the hadronic one is given by

$$L_{\mu\nu}^{\gamma}(k_1, k_2, k)H_{\mu\nu}^{\parallel} = -\frac{1}{st}\{(2\tau A_t + q^2 B)g_1 + 2\tau[A_t - x'(u+t)B]g_2\}\frac{x'}{q^2}, \tag{28}$$

$$L^{\gamma}_{\mu\nu}(k_1, k_2, k)H^{\perp}_{\mu\nu} = -\frac{1}{st}\left\{\left[A_s - \frac{uq^2}{V}B - A_t\left(1 - y + \frac{2u\tau}{V}\right)\right]g_1 + \left[A_s + x'(s+u)B + \left(1 - y + \frac{2u\tau}{V}\right)[x'(u+t)B - A_t]\right]g_2\right\} \times \frac{x'}{q^2}\sqrt{\frac{M^2}{Q^2}\left(1 - y + \frac{u\tau}{V}\right)^{-1}},$$

$$A_t = (u+t)^3 + (uq^2 - st)(u+s), \tag{29}$$

$$B = (u+t)\left(2V + \frac{u+t}{x'}\right) + (u+s)\left(2V(1-y) - \frac{u+s}{x'}\right),$$

$$A_s = (u+s)^3 + (uq^2 - st)(u+t), \quad q = k_1 - k_2 - k,$$

$$x' = \frac{-q^2}{2p_1q}, \quad g_{1,2} = g_{1,2}(x', q^2).$$

The contraction of the shifted leptonic tensor $L^{\gamma}_{\mu\nu}(zk_1, k_2, \tilde{k})$ entering the definition of the leptonic tensor in region (iii) and the hadronic tensor can be obtained from Eqs. (28) and (29) by the substitution

$$(k_1, k) \longrightarrow (zk_1, \tilde{k}),$$
$$(s, t, u, q, x') \longrightarrow (\tilde{s}, \tilde{t}, \tilde{u}, \tilde{q}, \tilde{x}), \quad \tilde{x} = \frac{-\tilde{q}^2}{2p_1\tilde{q}}. \tag{30}$$

We use the approach developed in [8] to extract the leading contributions (those proportional to $\ln\theta_0$ and $\ln\theta_0'$) to the respective cross section and to separate the infrared singularities. We write the cross section as

$$\frac{d\sigma^{iii)}_{\parallel,\perp}}{\hat{y}d\hat{x}d\hat{y}dz} = \frac{\alpha^2}{4\pi^2} \times \left\{P(z, L_0)\left[\int_{\Delta}^{x_{1\max}} \frac{dx_1[z^2 + (z-x_1)^2]}{x_1 z(z-x_1)} \times \ln\frac{2(1-c)}{\theta_0^2}\Sigma_{\parallel,\perp}(x_t, y_t, Q_t^2) + \int_{\Delta/Y}^{y_{1\max}} \frac{dy_1[1 + (1+y_1)^2]}{y_1(1+y_1)} \times \ln\frac{2(1-c)}{\theta_0'^2}\Sigma_{\parallel,\perp}(x_s, y_s, Q_s^2)\right] + \frac{1+z^2}{1-z}L_0 Z_{\parallel,\perp}\right\}, \tag{31}$$

where

$$x_t = \frac{xy(z-x_1)}{z-x_1+y-1}, \quad y_t = \frac{z-x_1+y-1}{z-x_1}, \quad Q_t^2 = Q^2(z-x_1).$$

For the proton target at rest, we have

$$x_{1\max} = \frac{2z\varepsilon[M - \varepsilon_2(1-c)] - 2M\varepsilon_2 - \mu^2 - 2\mu M}{2\varepsilon[M - \varepsilon_2(1-c)]},$$

and for the HERA collider conditions,

$$x_{1\max} = z - \frac{Y(1+c)}{2}.$$

The dependence on the infrared auxiliary parameter Δ and on the angles θ_0 and θ_0' is contained in the first two terms on the right-hand side of Eq. (31), whereas the quantities $Z_{\parallel,\perp}$ do not contain the infrared and collinear singularities. They can be written as

$$Z_{\parallel,\perp} = -\frac{2(1-c)}{zQ^2}\int_0^{\infty}\frac{du}{1+u^2}\left\{\int_0^1 \frac{dt_1}{t_1|t_1 - a|} \times \left[\int_0^{x_m}\frac{dx_1}{x_1}\Phi_{\parallel,\perp}(t_1, t_2(t_1, u)) - \int_0^{y_{1m}}\frac{dx_1}{x_1}\Phi_{\parallel,\perp}(a, 0)\right] + \int_0^a \frac{dt_1}{t_1 a}\left[\int_0^{y_{1m}}\frac{dx_1}{x_1}\Phi_{\parallel,\perp}(a, 0) - \int_0^{x_{1m}}\frac{dx_1}{x_1}\Phi_{\parallel,\perp}(a, 0)\right]\right\}, \tag{32}$$

where we use the same notation as in [8], namely,

$$t_{2,1} = \frac{1 - c_{1,2}}{2}, \quad a = \frac{1-c}{2},$$

$$t_2(t_1, u) = \frac{(a-t_1)^2(1+u^2)}{x_+ + u^2 x_-}, \quad c_{1,2} = \cos\theta_{1,2},$$

$$\theta_{1,2} = \widehat{\tilde{\mathbf{k}}\mathbf{k}}_{1,2},$$

$$x_{\pm} = t_1(1-2a) + a \pm 2\sqrt{a(1-a)t_1(1-t_1)}.$$

The quantity $\Phi_{\parallel,\perp}(t_1, t_2)$ is given by

$$\Phi_{\parallel,\perp}(t_1, t_2) = \frac{\alpha^2(\tilde{q}^2)\tilde{x}}{\tilde{Q}^6}G_{\parallel,\perp},$$

$$G_{\parallel} = g_1(2\hat{\tau}\tilde{A}_t + \tilde{q}^2\tilde{B}) + 2g_2\hat{\tau}[\tilde{A}_t - \tilde{x}(\tilde{u} + \tilde{t})\tilde{B}].$$

$$g_{1,2} = g_{1,2}(\tilde{x}, \tilde{q}^2),$$

$$G_{\perp} = \sqrt{\frac{M^2}{\hat{Q}^2}(1 - \hat{y} - \hat{x}\hat{y}\hat{\tau})^{-1}} \tag{33}$$

$$\times\left\{g_1\left[\tilde{A}_s-\frac{u\tilde{q}^2}{V}\tilde{B}-\tilde{A}_t\left(1-\hat{y}+\frac{2u\hat{\tau}}{V}\right)\right]\right.$$

$$+g_2\left[\tilde{A}_s+\tilde{x}(\tilde{s}+\tilde{u})\tilde{B}+\left(1-\hat{y}+\frac{2u\hat{\tau}}{V}\right)\right.$$

$$\left.\left.\times[\tilde{x}(\tilde{u}+\tilde{t})\tilde{B}-\tilde{A}_t]\right]\right\}.$$

For the proton target at rest and the HERA collider, the respective upper integration limits in the right-hand side of Eq. (32) are

$$x_m=\frac{2Mz\varepsilon-2M\varepsilon_2-2z\varepsilon\varepsilon_2(1-c)-\mu^2-2M\mu}{2\varepsilon[M+z\varepsilon(1-c_1)-\varepsilon_2(1-c_2)]},$$

$$x_m=\frac{2z-Y(1+c)}{1+c_1}.$$

4. THE TOTAL RADIATIVE CORRECTION

The total RC to Born cross section (15) is given by the sum of the virtual and soft corrections and the hard-photon emission contribution. The last one is different for the exclusive and calorimeter event selection. In the considered approximation, it is convenient to write this RC as

$$\frac{d\sigma^{RC}_{\parallel,\perp}}{\hat{y}d\hat{x}d\hat{y}dz}=\frac{\alpha^2}{4\pi^2}(\Sigma_{i\parallel,\perp}+\Sigma_{f\parallel,\perp}). \tag{34}$$

The first term Σ_i is independent of the experimental selection rules for the scattered electron and is given by

$$\Sigma_{i\parallel,\perp}=L_0\left\{\frac{1}{2}L_0P^{(2)}_\theta(z)+\frac{1+z^2}{1-z}\right.$$

$$\times\left[5\ln z-2F(z)+\ln^2Y-2\ln z\ln Y-\frac{\pi^2}{3}\right.$$

$$\left.+2\mathrm{Li}_2\left(\frac{1+c}{2}\right)\right]+\frac{3+z^2}{2(1-z)}\ln^2z$$

$$-\frac{2(3-2z+3z^2)}{1-z}\ln(1-z) \tag{35}$$

$$\left.+\frac{3-20z+z^2}{2(1-z)}\right\}\Sigma_{\parallel,\perp}(\hat{x},\hat{y},\hat{Q}^2)+P(z,L_0)\ln\frac{2(1-c)}{\theta_0^2}$$

$$\times\int_0^{u_0}\frac{du}{1-u}P^{(1)}(1-u)\Sigma_{\parallel,\perp}(x_t,y_t,Q_t^2)$$

$$+\frac{1+z^2}{1-z}L_0Z_{\parallel,\perp},\quad u_0=\frac{x_{1\max}}{z},$$

where the quantity $P^{(1)}(x)$ is defined by relations (24) and the quantities x_t, y_t, and Q_t^2 depend on $u=x_1/z$.

On the other hand, the second term in the right-hand side of Eq. (34), denoted by Σ_f, explicitly depends on the event selection rule. It includes the main effect of the scattered-electron radiation. For the exclusive event selection, where only the scattered bare electron is measured and any photon that is collinear to its momentum direction is ignored, this contribution is

$$\Sigma^{\mathrm{excl}}_{f\parallel,\perp}=P(z,L_0)$$

$$\times\int_0^{y_1\max}dy_1\left[(L_Q+\ln Y-1)P^{(1)}\left(\frac{1}{1+y_1}\right)+\frac{y_1}{1+y_1}\right] \tag{36}$$

$$\times\Sigma_{\parallel,\perp}(x_s,y_s,Q_s^2).$$

In this case, as mentioned above, the parameter θ_0' that separates kinematical regions (ii) and (iii) is not physical, and we see that the final result does not contain it. However, the mass singularity that is related to the scattered electron radiation exhibits itself through L_Q in the right-hand side of Eq. (36).

The situation is quite different for the calorimeter event selection, where the detector cannot distinguish between the events involving a bare electron and events where the scattered electron is accompanied by a hard photon emitted within a narrow cone with the opening angle $2\theta_0'$ around the scattered electron momentum direction. For this experimental setup, we derive

$$\Sigma^{\mathrm{cal}}_{f\parallel,\perp}=P(z,L_0)$$

$$\times\left[\ln\frac{2(1-c)}{\theta_0'^2}\int_0^{y_{1\max}}dy_1P^{(1)}\left(\frac{1}{1+y_1}\right)\right. \tag{37}$$

$$\left.\times\Sigma_{\parallel,\perp}(x_s,y_s,Q_s^2)+\frac{1}{2}\Sigma_{\parallel,\perp}(\hat{x},\hat{y},\hat{Q}^2)\right].$$

For the calorimeter setup, the parameter θ_0' defines the event selection rule and is therefore physically meaningful. The final result depends on it. However, the mass singularity due to the photon emission by the final electron is cancelled in accordance with the Kinoshita–Lee–Nauenberg theorem [18]. The absence of the mass singularity clearly indicates that the term containing $\ln\theta_0'$ in the right-hand side of Eq. (37) arises due to the contribution of kinematical region (iii), where the scattered electron and the photon radiated from the final state are well separated. That is why no

question arises as to determining the quantity ε_2 that enters the expression for $y_{1\max}$.

Comparing our analytical results for the RC due to the real and virtual photon emission with similar calculations for the unpolarized case [8], we see that within the leading-log accuracy (double-logarithm terms in our case), these RCs are the same for the spin-dependent and spin-independent parts of the cross section of radiative DIS process (1). The difference appears at the level of the next-to-leading-log accuracy (single logarithmic terms in our case). That is true for the photonic corrections in an arbitrary order of the perturbation theory.

We note that the correction to the usually measured asymmetry, which is the ratio of the spin-dependent part of the cross section to the spin-independent one, is not large because the main factorized contribution due to the virtual and soft photon emission trends to cancellation in this case. If the experimental information about the spin observables is extracted directly from the spin-dependent part of the cross section (see [19] for the corresponding experimental method), this cancellation does not occur and the factorized correction gives the basic contribution.

5. THE CASE OF QUASI-ELASTIC SCATTERING

In the previous sections, we considered the tagged-photon events in the DIS process. These events can be used to measure the spin-dependent proton structure functions g_1 and g_2 in a single run without lowering the electron beam energy. In the quasi-elastic case, where the target proton is scattered elastically,

$$e^-(k_1) + p(p_1) \longrightarrow e^-(k_2) + \gamma(k) + p(p_2), \qquad (38)$$

the tagged-photon events can also be used to measure the proton electromagnetic form factors G_E and G_M. Our final results obtained in Section 4 can then be applied using relation (11) between the spin-dependent proton structure functions g_1 and g_2 and the proton electromagnetic form factors in this limit. In this case, we can therefore use all the formulas in Section 4 with $\Sigma_{\parallel,\perp}$ and $G_{\parallel,\perp}$ entering the definition of $Z_{\parallel,\perp}$ replaced by $\Sigma^{\rm el}_{\parallel,\perp}$ and $G^{\rm el}_{\parallel,\perp}$, respectively,

$$\Sigma^{\rm el}_{\parallel}(x, y, Q^2) = \frac{4\pi\alpha^2(Q^2)}{y(4M^2 + Q^2)}$$

$$\times\left[4\tau\left(\tau + 1 - \frac{1}{y}\right)G_M G_E - \left(1 - \frac{y}{2}\right)(1 + 2\tau)G_M^2\right] \qquad (39)$$

$$\times\,\delta(1 - x),$$

$$\Sigma^{\rm el}_{\perp}(x, y, Q^2) = \frac{8\pi\alpha^2(Q^2)}{y(4M^2 + Q^2)}\sqrt{\frac{M^2}{Q^2}[1 - y(1 + \tau)]}$$

$$\times\left[\left(1 - \frac{y}{2}\right)G_M^2 - (1 + 2\tau)G_M G_E\right]\delta(1 - x), \qquad (40)$$

$$G^{\rm el}_{\parallel,\perp} = \frac{\tilde{Q}^2}{4M^2 + \tilde{Q}^2}$$

$$\times\,(D_{\parallel,\perp}G_M^2 + E_{\parallel,\perp}G_M G_E)\delta(1 - \tilde{x}),$$

$$D_{\parallel} = \bar{B}[\tilde{q}^2 + 2\hat{\tau}(\tilde{u} + \tilde{t})],$$

$$E_{\parallel} = 2\hat{\tau}\left[\left(1 + \frac{4M^2}{\tilde{Q}^2}\right)\tilde{A}_t - \bar{B}(2\hat{V} + \tilde{u} + \tilde{t})\right], \qquad (41)$$

$$D_{\perp} = -K\bar{B}\left[\frac{u\tilde{q}^2}{V} + \tilde{s} + \tilde{u} + (\tilde{u} + \tilde{t})\left(1 - \hat{y} + \frac{2u\hat{\tau}}{V}\right)\right],$$

$$E_{\perp} = K\left\{\left(1 + \frac{4M^2}{\tilde{Q}^2}\right)\left[\tilde{A}_s - \left(1 - \hat{y} + \frac{2u\hat{\tau}}{V}\right)\tilde{A}_t\right]\right.$$

$$\left. + \bar{B}\left[\tilde{s} + \tilde{u}(1 + 4\hat{\tau}) + (\tilde{u} + \tilde{t})\left(1 - \hat{y} + \frac{2u\hat{\tau}}{V}\right)\right]\right\},$$

where

$$\bar{B} = (\tilde{u} + \tilde{t})(2\hat{V} + \tilde{u} + \tilde{t})$$

$$+ (\tilde{u} + \tilde{s})[2\hat{V}(1 - \hat{y}) - \tilde{u} - \tilde{s}],$$

$$K = \sqrt{\frac{M^2}{\hat{Q}^2}(1 - \hat{y} - \hat{x}\hat{y}\hat{\tau})^{-1}}$$

and the form factors in the right-hand side of Eq. (41) depend on $\tilde{q}^2$.

The description of the form factors is a very important test for any theoretical model of strong interactions [20]. The proton magnetic form factor G_M is known with a high accuracy in a wide range of the momentum transfer, while the data about the electric form factor G_E are very poor. The recent experiment at the Jefferson Lab on the measurement of the ratio of the recoil proton polarizations performed by the Hall A Collaboration [21] improves the situation in the region up to $Q^2 \approx 3.5\ \mathrm{GeV}^2$, but the higher momentum transfer region remains unexplored. The use of radiative events (38), with both the polarized and unpolarized proton target, at accelerators with a high-intensity electron beam (for example, CEBAF) can open new possibilities in the measurement of G_E as compared to both the Rosenbluth method [22] and the method based on measuring the recoil proton polarization ratio [23].

REFERENCES

1. M. Acciarri *et al.* (L3 Collab.), Preprint CERN–EP/99–129, Sept. (1999).
2. M. Benayoun, S. I. Eidelman, V. N. Ivanchenko, and Z. K. Silagadze, hep-ph/9910523 (1999).
3. A. B. Arbuzov, E. A. Kuraev, N. P. Merenkov, and L. Trentadue, JHEP 12, 009 (1998); S. Binner, J. H. Kühn, and K. Melnikov, Phys. Lett. B **459**, 279 (1999); M. Konchatnij and N. P. Merenkov, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 769 (1999) [JETP Lett. **69**, 811 (1999)]; S. Spangolo, Eur. Phys. J. C **6**, 637 (1999); G. Cataldi, A. Denig, W. Kluge, and G. Venanzoni, Preprint KLOE MEMO 195, August 13 (1999); V. A. Khoze *et al.*, hep-ph/0003313 (2000).
4. L. Mo and Y. Tsai, Rev. Mod. Phys. **41**, 205 (1969); H. Spiesberger *et al.*, in *Physics at HERA: International Meeting on Fundamental Physics*, Ed. by W. Buchmüller and G. Ingleman (DESY, Hamburg, 1991); A. Akhundov, L. Kalinovskaya, and T. Riemann, Fortschr. Phys. **44**, 373 (1996).
5. M. W. Krasny, W. Placzek, and H. Spiesberger, Z. Phys. C **53**, 687 (1992).
6. S. Jadach, M. Jezabek, and W. Placzek, Phys. Lett. B **248**, 417 (1990); D. Bardin, L. Kalinovskaya, and T. Riemann, Z. Phys. C **76**, 487 (1997); T. Ahmed *et al.* (H1 Collab.), Z. Phys. C **66**, 529 (1995).
7. L. Favart, M. Gruwe, P. Marage, and Z. Zhang, Z. Phys. C **72**, 425 (1996).
8. H. Anlauf, A. B. Arbuzov, E. A. Kuraev, and N. P. Merenkov, JHEP 10, 013 (1998); Phys. Rev. D **59**, 014 003 (1999).
9. V. N. Baier, V. S. Fadin, and V. A. Khoze, Nucl. Phys. B **65**, 381 (1973).
10. E. A. Kuraev, N. P. Merenkov, and V. S. Fadin, Yad. Fiz. **47**, 1593 (1988) [Sov. J. Nucl. Phys. **47**, 1009 (1988)].
11. T. V. Kukhto and N. M. Shumeiko, Nucl. Phys. B **219**, 12 (1983).
12. I. V. Akushevich and N. M. Shumeiko, J. Phys. G **20**, 513 (1994).
13. I. V. Akushevich, A. B. Arbuzov, and E. A. Kuraev, Phys. Lett. B **432**, 222 (1998); G. I. Gakh, M. I. Konchatnij, and N. P. Merenkov, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 328 (2000) [JETP Lett. **71**, 227 (2000)].
14. G. t'Hooft and N. Veltman, Nucl. Phys. B **153**, 365 (1979).
15. E. A. Kuraev, N. P. Merenkov, and V. S. Fadin, Yad. Fiz. **45**, 782 (1987) [Sov. J. Nucl. Phys. **45**, 486 (1987)].
16. M. Konchatnij and N. P. Merenkov, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 845 (1999) [JETP Lett. **69**, 893 (1999)].
17. S. Jadach, M. Skrzypek, and B. F. L. Ward, Phys. Rev. D **47**, 3733 (1993).
18. T. Kinoshita, J. Math. Phys. (N. Y.) **3**, 650 (1962); T. D. Lee and M. Nauenberg, Phys. Rev. **133**, 1549 (1964).
19. N. D. Gagunashvili, Nucl. Instum. Methods Phys. Res. A **343**, 606 (1994); N. D. Gagunashvili *et al.*, Preprint No. E1-96-483, OIYaI (Joint Institute for Nuclear Research, Dubna, 1996).
20. S. J. Brodsky and G. R. Farrar, Phys. Rev. D **11**, 1309 (1975); M. F. Gari and W. Krümpelmann, Z. Phys. A **322**, 339 (1996); G. Holzwarth, Z. Phys. A **356**, 339 (1996); P. L. Chung and F. Coester, Phys. Rev. D **44**, 229 (1991); I. G. Aznauryan, Phys. Lett. B **316**, 391 (1993); P. Kroll, M. Schürmann, and W. Schweiger, Z. Phys. A **338**, 339 (1991); D. H. Lu, A. M. Thomas, and A. G. Williams, Phys. Rev. C **57**, 2628 (1998); A. V. Radushkin, Acta Phys. Pol. B **15**, 40 (1984); S. Capitani *et al.*, Nucl. Phys. B (Proc. Suppl.) **73**, 294 (1999).
21. M. K. Jones, K. A. Aniol, F. T. Baker, *et al.*, Phys. Rev. Lett. **84**, 1398 (2000).
22. M. N. Rosenbluth, Phys. Rev. **79**, 615 (1950).
23. A. I. Akhiezer and M. P. Rekalo, Fiz. Élem. Chastits At. Yadra **4**, 663 (1973) [Sov. J. Phys. Part. Nucl. **4**, 227 (1974)]; R. Arnold, C. Carlson, and F. Gross, Phys. Rev. C **23**, 363 (1981).

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 940–950.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1087–1098.

ATOMS, SPECTRA, RADIATION

The Hanle Effect in Forward Scattering in Excited Media

B. A. Veklenko*
Moscow Power Institute, ul. Krasnokazarmennaya 17, Moscow, 111250 Russia
**e-mail: phf@deans.mpei.ac.ru*
Received May 11, 2000

Abstract—A new method for calculating the density matrix of a quantized electromagnetic field which interacts with the environment in the presence of kinetic processes in the medium is suggested. This method accurately takes into account photon–photon quantum correlators and possesses certain symmetry properties, which substantially facilitate the summation of the Feynman diagrams that arise in the calculations. Forward scattering of resonance radiation by two-level gas atoms in a magnetic field is considered as an application. Insufficiency of a semiclassical description of this coherent process in excited media with the use of the unique tensor index of refraction is shown to follow from quantum electrodynamics. The introduction of one more function depending on the frequency of irradiation and the concentration of excited atoms is necessary. Experiments that simultaneously determine both this additional function and the standard refractive index of the medium are specified.

1. INTRODUCTION

Work [1], which called attention to the special features of the Hanle effect in forward scattering, initiated an unremitting flow of studies [2–11]. The characteristic feature of this effect as distinguished from the standard Hanle effect [12] is the coherence of secondary radiation sources. As a rule, two aspects of the problem are considered: linear scattering with the purpose of diagnosing scattering media [9–11] and nonlinear magnetooptical effects with the purpose of studying their characteristic features [1–8].

In both theoretical and experimental studies, the influence of excited atoms on the processes under study has not been given due attention. It was implicitly and sometimes explicitly [2, 4] assumed that the influence of excited atoms through medium polarization depending on the difference $n_\mu - n_m$ (n_μ and n_m are the concentrations of nonexcited and excited atoms, respectively) can only weaken the observed effects.

In reality, and we call attention to this circumstance, the role played by excited atoms under the conditions of coherence of secondary radiation sources is much more important even in linear scattering processes, to which our consideration will be confined. Consistent quantum electrodynamics, which deals with quantized electromagnetic fields, shows that, in this case, the scattering medium cannot be described with the use of one polarization or refractive index tensor, and we must introduce one more function depending on the electromagnetic field frequency. A similar situation arises in studying the effects of selective reflection of resonance radiation from excited media [13, 14]. It follows that we are dealing with the general properties of the evolution of quantized electromagnetic fields in excited dispersing media. The Hanle effect in forward scattering is convenient in that it allows various characteristics of media, including those that are not described by their indexes of refraction, to be diagnosed with the use of existing experimental units by observing mutually orthogonal scattered light polarizations.

Recall that the index of refraction in a natural way appears in classical optics in describing the deterministic (not fluctuation) radiation component by excluding the dynamic variables of the medium from the Maxwell system of equations. Such a procedure cannot be applied to field operators in quantum electrodynamics in the presence of absorbing media. The index of refraction in quantum electrodynamics is naturally present in the method of Green quantum functions [15]. Such a calculation procedure has already been used in several optical works [16–20]. The problem, however, persists, and its essence is as follows. We cannot obtain closed equations for quantum-averaged values because, in deriving them, the chain of Bogolyubov equations has to be closed by correlator breaks. In the Green function technique, an equivalent procedure is the neglect of $\langle \hat{\alpha}^+_{\mathbf{k}\lambda}\hat{\alpha}^+_{\mathbf{k}\lambda}\hat{\alpha}_{\mathbf{k}\lambda}\hat{\alpha}_{\mathbf{k}\lambda}\rangle$ correlators, where $\hat{\alpha}_{\mathbf{k}\lambda}$ ($\hat{\alpha}^+_{\mathbf{k}\lambda}$) is the operator of annihilation (creation) of a photon in the state with wave vector $\mathbf{k}$ and polarization index λ. Arguments of one or another kind are used to justify the assumption that these correlators play a comparatively unimportant role (all $\mathbf{k}$ and λ are equal) in the final equations. At the same time, the presence of induced radiation processes in excited media causes the appearance of correlators of just this type. Their coefficients are singular because of the presence of δ functions that ensure that the conservation laws be obeyed. For this reason, the possibility of ignoring such correlators is far from obvious.

1063-7761/01/9206-0940$21.00

To retain these correlators and elucidate the role that they play, we use another technique for excluding medium parameters from the calculation scheme. Let ρ be the density matrix of the atoms + electromagnetic field system. After summing over the parameters of medium atoms, the density matrix of an electromagnetic field in a medium, $\rho_{ph} = \mathrm{Sp}_a\rho$, arises in a natural way. The mean field characteristics are found by applying the standard procedure,

$$\langle \hat{\mathscr{E}}^{\nu} \rangle = \mathrm{Sp}_{ph}\hat{\mathscr{E}}^{\nu}\rho_{ph}, \quad \langle \hat{\mathscr{E}}^{\nu}\hat{\mathscr{E}}^{\nu'} \rangle = \mathrm{Sp}_{ph}\hat{\mathscr{E}}^{\nu}\hat{\mathscr{E}}^{\nu'}\rho_{ph},$$

where $\hat{\mathscr{E}}^{\nu}$ is the operator of electromagnetic field strength. This technique does not involve breaking photon–photon correlators.

Below, we describe a new method for deriving a closed equation for ρ_{ph}. This method possesses internal symmetry, which substantially facilitates calculations.

Attempts at finding equations for the density matrices of subsystems (such as the photon field) interacting with the environment were made in [21–23]. The form of the equations and the absence of symmetry relations impeded going beyond perturbation theory. To overcome these difficulties, the method of Γ operators was developed in [13, 14]. Experience shows that familiarizing oneself with the method that uses the abstract Γ space involves serious difficulties. This prompted us to develop another independent method for calculating ρ_{ph} which has similar symmetry properties.

We use the method of matrix Green functions developed in [24] for calculating quantum means. We will show that some of its modification allows ρ_{ph} to be calculated. Such a calculation scheme requires certain nonobvious preliminary constructions and for this reason has remained undemanded.

Consistently taking into account photon–photon correlators has nonobvious consequences, such as the insufficiency of describing the optical properties of excited media by a unique polarization tensor. What is more, the existence of such a tensor would contradict the fundamentals of quantum electrodynamics. This assertion can be illustrated by the example that admits elementary analysis, see below.

2. PRELIMINARY ANALYSIS

Some properties of light passed through a thermally excited gas layer in a magnetic field can be predicted based on the following nonrigorous semiquantitative analysis. Let us use the $\hat{S}$ matrix theory, and let ρ be the density matrix of the medium + field system after switching on the interaction Hamiltonian. The ρ_{ph} density matrix of the photon subsystem can always be written [14] as the sum of two terms,

$$\rho_{ph} = \rho_{ph}^{(c)} + \rho_{ph}^{(n)}. \tag{1}$$

The $\rho_{ph}^{(c)}$ term describes electromagnetic field scattering not accompanied by changes in the initial state of the medium (coherent scattering channel). The $\rho_{ph}^{(n)}$ term describes processes that change the state of the medium (incoherent scattering channel). Sum (1) shows that there is no quantum (described by wave functions) interference phenomena between these channels because medium atomic functions in finite states are orthogonal. Let $\hat{N}_{\mathbf{k}\lambda} = \hat{\alpha}^{+}_{\mathbf{k}\lambda}\hat{\alpha}_{\mathbf{k}\lambda}$ be the operator of the number of photons in the state corresponding to the $(\mathbf{k}, \lambda)$ mode. It follows from (1) that

$$\begin{aligned} \langle \hat{N}_{\mathbf{k}\lambda} \rangle &= \langle \hat{N}_{\mathbf{k}\lambda} \rangle^{(c)} + \langle \hat{N}_{\mathbf{k}\lambda} \rangle^{(n)}, \\ \langle \hat{N}_{\mathbf{k}\lambda} \rangle^{(c)} &= \mathrm{Sp}_{ph}\hat{N}_{\mathbf{k}\lambda}\rho_{ph}^{(c)}, \\ \langle \hat{N}_{\mathbf{k}\lambda} \rangle^{(n)} &= \mathrm{Sp}_{ph}\hat{N}_{\mathbf{k}\lambda}\rho_{ph}^{(n)}. \end{aligned} \tag{2}$$

This shows that, if $(\mathbf{k}, \lambda)$ is the mode to which photons are scattered from the $(\mathbf{k}_0, \lambda_0)$ initial state as a result of their interaction with the medium, the $\langle \hat{N}_{\mathbf{k}\lambda} \rangle$ mean in this mode is representable by the sum of two positive definite terms. This conclusion is of fundamental importance. The positive definiteness of the terms in (2) follows from the possibility in principle of measuring them. A formal proof of this statement can be found in [13].

Next, consider the result of scattering of transverse ($\lambda = 1, 2$) photons on two structurally identical atoms, one of which is not excited and occurs at point $\mathbf{R}_1$, and the other is excited and occurs at point $\mathbf{R}_2$. The atoms will be assumed to be two-level but possessing Zeeman sublevels. Let the scattered light be in quasi-resonance, $|k - \omega_{m\mu}| < k$, with the $\omega_{m\mu}$ frequency of optical transitions in atoms, $\hbar = c = 1$. The resonance condition complicates the analysis and, generally, makes perturbation theory inapplicable. For preliminary analysis purposes, we will proceed as follows. Let us use perturbation theory retaining $\pm i0$ terms, which arise as a consequence of the adiabatic hypothesis, in resonance denominators. In a more complete theory, these terms will be replaced by $\pm i\gamma/2$ (γ is the width of the atomic energy levels) because of the presence of mass operators. The collisional width will always be assumed to exceed the radiation width. The signs of $i0$ and $i\gamma/2$ should coincide because of the causality principle. It follows that taking into account $i0$ terms, which will immediately be replaced by $i\gamma/2$, allows us to predict the structure of the final equations in a more elaborate theory.

We assume that each atom has one valence electron. The spin effects are ignored. In the second quantization representation, the Hamiltonian of the system is written as

$$\hat{H} = \hat{H}^0 + \hat{H}',$$

$$\hat{H}^0 = \sum_j \varepsilon_j \hat{b}_j^{+}\hat{b}_j + \sum_j \varepsilon_j \hat{\beta}_j^{+}\hat{\beta}_j + \sum_{\mathbf{k}\lambda} k\left(\hat{\alpha}^{+}_{\mathbf{k}\lambda}\hat{\alpha}_{\mathbf{k}\lambda} + \frac{1}{2}\right),$$

$$\hat{H}' = -\sum_{ij}\frac{P_{ij}^{\lambda}(\mathbf{k})}{\sqrt{2kV}} \times (e^{i\mathbf{k}\cdot\mathbf{R}_1}\hat{b}_i^+\hat{\alpha}_{\mathbf{k}\lambda}\hat{b}_j + e^{i\mathbf{k}\cdot\mathbf{R}_2}\hat{\beta}_i^+\hat{\alpha}_{\mathbf{k}\lambda}\hat{\beta}_j) + \text{H.c.}$$

Here, $V = L_xL_yL_z$ is the normalization volume, ε_j are the allowed atomic energy values taking into account the presence of a magnetic field, and $\hat{b}_j$ ($\hat{b}_j^+$) and $\hat{\beta}_j$ ($\hat{\beta}_j^+$) are the operators of annihilation (creation) of the first and second atom states corresponding to these energy levels. In this special problem, all operators can be considered belonging to Bose–Einstein fields. Further,

$$P_{ij}^{\lambda}(\mathbf{k}) = \frac{e}{m}\int\psi_i^*(\mathbf{r})\hat{\mathbf{p}}\mathbf{e}_{\mathbf{k}}^{\lambda}e^{i\mathbf{k}\cdot\mathbf{r}}\psi_j(\mathbf{r})d\mathbf{r}, \quad \hat{\mathbf{p}} = -i\nabla_{\mathbf{r}},$$

where ψ_i is the wave function of an atom in the state with energy ε_i, and $\mathbf{e}_{\mathbf{k}}^{\lambda}$ are the unit vectors of linear polarization of photons.

The $\hat{S}$ matrix will be calculated accurate to third-order perturbation theory,

$$\hat{S} = 1 + \hat{S}^{(1)} + \hat{S}^{(2)} + \hat{S}^{(3)}.$$

Standard calculations show that, for the coherent channel [14],

$$\langle\hat{N}_{\mathbf{k}\lambda}\rangle^{(c)} = \langle\hat{S}^{(2)}|\hat{N}_{\mathbf{k}\lambda}|\hat{S}^{(2)}\rangle = 4\pi^2\left|\frac{P_{m\mu}^{\lambda*}(\mathbf{k})P_{m\mu}^{\lambda_0}(\mathbf{k}_0)}{2kV}\right| \times\delta^2(k-k_0)\left[\frac{\exp[i(\mathbf{k}_0-\mathbf{k})\cdot\mathbf{R}_1]}{k-\omega_{m\mu}+i\gamma/2} + \frac{\exp[i(\mathbf{k}_0-\mathbf{k})\cdot\mathbf{R}_2]}{\omega_{m\mu}-k+i\gamma/2}\right]^2 N_0, \tag{3}$$

where N_0 is the number of photons in the scattered mode. Here, it is taken into account that, by virtue of Zeeman splitting, only the ε_m and ε_μ sublevels are in resonance with frequency k, and, for this reason, only these sublevels are retained in (3).

For the incoherent scattering channel, we have

$$\langle\hat{N}_{\mathbf{k}\lambda}\rangle^{(n)} = \langle\hat{S}^{(1)} + \hat{S}^{(3)}|\hat{N}_{\mathbf{k}\lambda}|\hat{S}^{(1)} + \hat{S}^{(3)}\rangle \geq 0. \tag{4}$$

The $\langle\hat{S}^{(1)}|\hat{N}_{\mathbf{k}\lambda}|\hat{S}^{(1)}\rangle$ term is not of interest to us because it does not describe scattering processes. Explicit calculations of (4) with retention of terms fourth-order in charge [as in (3)] yield

$$\langle\hat{S}^{(1)}|\hat{N}_{\mathbf{k}\lambda}|\hat{S}^{(3)}\rangle + \text{c.c.} = -8\pi^3 i \times\left|\frac{P_{m\mu}^{\lambda*}(\mathbf{k})P_{m\mu}^{\lambda_0}\mathbf{k}_0}{2kV}\right|^2 N_0 \times\frac{\delta^2(\varepsilon_\mu-\varepsilon_m-k_0)\delta(k-k_0)}{k-\omega_{m\mu}+i\gamma/2} \times\exp[-i(\mathbf{k}_0-\mathbf{k})\cdot(\mathbf{R}_2-\mathbf{R}_1)] + \text{c.c.} \tag{5}$$

Let us analyze (3)–(5) taking into account that scattering occurs forward. If the incoherent channel is absent, then only coherent scattering channel does, according to (3), participate in the formation of $\langle\hat{N}_{\mathbf{k}\lambda}\rangle$ for the ($\mathbf{k}_0$, λ) scattered mode. This mode differs from the initial ($\mathbf{k}_0$, λ_0) mode by its polarization, which is perpendicular to that of the latter. According to (3), if $k = \omega_{m\mu}$, then the amplitudes of waves scattered on different atoms are algebraically summed. This means that, in the presence of many scatterers and at their low concentration,

$$\langle\hat{N}_{\mathbf{k}\lambda}\rangle^{(c)} \propto n_\mu + n_m. \tag{6}$$

However, if $|k-\omega_{m\mu}| > \gamma/2$, then, again according to (3), the amplitudes of scattered waves are subtracted from each other, and, for extended media, we should expect

$$\langle\hat{N}_{\mathbf{k}\lambda}\rangle^{(c)} \propto n_\mu - n_m. \tag{7}$$

The difference between (6) and (7) is at variance with calculations based on the standard quantum Green function methods with the use of refractive index κ of the medium related to polarization operator Π as $\kappa = 1 - \Pi/k^2$ [15]. Such calculations using the procedure of breaking photon–photon correlators show [20] that $\Pi \propto n_\mu - n_m$ for all frequencies.

It may seem that our analysis is incomplete because of the neglect of the incoherent scattering channel. Can taking this channel into account restore agreement between the theories? At first sight, this is precisely how matters stand, because the sum of (3) and (5) gives a complete result, which only differs from (3) in the sign of $i\gamma/2$ in the denominator in the second term. According to this expression, the amplitudes of scattering of optical waves on excited and nonexcited atoms in the $\mathbf{k} = \mathbf{k}_0$ direction are in antiphase for all frequencies. This implies (7), which seemingly restores agreement between the theories. Such a conclusion would be erroneous. The matter is that (5) is negative at $\mathbf{k} = \mathbf{k}_0$, and this contradicts the positive definiteness of the contribution of the incoherent scattering channel to $\langle\hat{N}_{\mathbf{k}\lambda}\rangle$. The positive definiteness is restored by consistently taking the $\hat{S}^{(3)}$ term in (4) into account. This in turn means that terms sixth-order in charge, which determine the sign of the final result, should be included in (4). Clearly, sixth-order terms in the incoherent channel exceed fourth-order terms in magnitude, and perturbation theory becomes inapplicable. The elementary sum-

mation of the Feynman diagrams resulting in the replacement of $i0$ by $i\gamma/2$ does not change the situation. We must use other additional summation methods.

To summarize, calculations of incoherent channel contributions to resonance scattering when the medium contains excited atoms always require going beyond perturbation theory and taking into account infinite subsequences of Feynman diagrams. At the same time, the positive definiteness of the incoherent scattering channel is evidence of the inequality

$$\langle \hat{N}_{\mathbf{k}_0\lambda}\rangle \geq \langle \hat{N}_{\mathbf{k}_0\lambda}\rangle^{(c)} \propto n_\mu + n_m$$

for

$$k = \omega_{m\mu}, \quad \lambda \neq \lambda_0.$$

The example considered above shows that the process that changes polarization for orthogonal to initial when light passes through an excited resonance medium in a magnetic field cannot be described with the use of standard refractive index and medium polarizability, which depend on $n_\mu - n_m$. It will be shown that the intensity of light passed through a medium with scattering but without polarization changes ($\lambda = \lambda_0$) admits such a description, and the refractive index can then appropriately be used.

It follows that studying the optical properties of excited isotropic resonance media by quantum electrodynamics requires at least two functions depending on $\mathbf{k}$ to be used simultaneously. The absence of one of them would contradict the fundamentals of quantum theory. Explicit equations for these functions are given below.

Note that calculations of these equations requires the use of quantized field theory methods. Our results cannot be obtained in semiclassical theory dealing with nonquantized fields. The usual assertion that, in strong fields, the results of quantum optics using coherent [25] states and those of classical optics should coincide, because perturbation theory series in these theories only differ by vacuum terms, which are of little consequence, is inapplicable because of the inapplicability of perturbation theory.

Consider the results of semiclassical theory in more detail. According to this theory, the light scattering mechanism is the induction of dipole moments of atoms by radiation undergoing scattering and then the induction of radiation by induced and therefore time-depended secondary wave dipole moment.

If a classical electromagnetic wave falls on a two-level atom in the excited state ψ_m, and the strength of this wave is described by the formula

$$\mathscr{E}^\nu = \frac{1}{2}\mathscr{E}_0^\nu(e^{i\omega t} + e^{-i\omega t}),$$

then, according to [26], the induced dipole moment in the units that we use equals

$$d_m^\nu = e\int \psi_m^* r^\nu \psi_\mu d\mathbf{r} = -\frac{e^2}{8\pi}$$

$$\times \sum_{\nu'} \mathscr{E}_0^{\nu'} \langle \mu|r^{\nu'}|m\rangle \langle m|r^\nu|\mu\rangle$$

$$\times\left(\frac{e^{i\omega t}}{\omega_{m\mu} - \omega - i\gamma/2} + \frac{e^{-i\omega t}}{\omega_{m\mu} + \omega - i\gamma/2}\right) + \text{c.c.}$$

If the atom occurs in the ground state ψ_μ, the external field induces the dipole moment

$$d_m^\nu = -\frac{e^2}{8\pi}\sum_{\nu'} \mathscr{E}_0^{\nu'} \langle m|r^{\nu'}|\mu\rangle\langle \mu|r^\nu|m\rangle$$

$$\times\left(\frac{e^{i\omega t}}{\omega_{\mu m} - \omega - i\gamma/2} + \frac{e^{-i\omega t}}{\omega_{\mu m} + \omega - i\gamma/2}\right) + \text{c.c.}$$

If the excited and nonexcited atoms are situated close to each other, then, in the quasi-resonance approximation, $\omega \approx \omega_{m\mu}$, and the resultant dipole moment is given by the sum

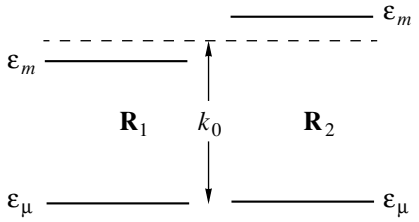
$$d_m^\nu + d_\mu^\nu = -\frac{e^2}{8\pi}\sum_{\nu'} \mathscr{E}_0^{\nu'}\left(\frac{\langle \mu|r^{\nu'}|m\rangle\langle m|r^\nu|\mu\rangle}{\omega_{m\mu} - \omega - i\gamma/2}\right.$$

$$\left. + \frac{\langle \mu|r^{\nu'}|m\rangle^*\langle \mu|r^\nu|m\rangle^*}{-\omega_{m\mu} + \omega + i\gamma/2}\right)e^{i\omega t} + \text{c.c.}$$

We see that all dipole moment components are fully compensated. For this reason, the scattering amplitude in extended media is determined by the $n_\mu - n_m$ difference, and, when $n_m \longrightarrow n_\mu$, scattering disappears according to semiclassical theory.

The theory of the Hanle effect based on the semiclassical approach taking into account the presence of excited atoms is described in [4], where it is shown that, in agreement with the semiclassical analysis performed above, the Hanle effect is determined by the $n_\mu - n_m$ difference not only in the linear approximation but also in the approximation of an arbitrary order nonlinear with respect to field $\mathscr{E}_0$.

Next note that the same semiclassical theory can be used to obtain a different result. Let the energy widths of atoms be zero. A system comprising an excited and a nonexcited atom does not possess dipole scattering. Suppose that an extraneous particle approaches one of the atoms. The energy level of the atom shifts under such perturbation. The resonance frequencies of the excited and nonexcited atoms cease to coincide. This means that waves scattered in antiphase by the excited and nonexcited atoms cease to strictly compensate each other because of different energy denominators in the scattering amplitudes. Dipole scattering appears. What



Scheme of energy levels corresponding to arithmetic summation of the amplitudes of light reflected by excited and nonexcited atoms.

is more, if the scattered light frequency falls within the detuning of the resonance frequencies (see figure), then, because of a change in the sign of one of resonance denominators, waves scattered by the atoms become synphase and are arithmetically summed. A flash of scattered light occurs.

Such an elementary consideration in terms of semiclassical theory shows that the intensity of light scattered in resonance by two atoms one of which is excited should be proportional to the $n_\mu + n_m$ sum. Unfortunately, formally taking into account collisional processes in semiclassical theory by the introduction of a collisional width leads to a different result, see above. This leads us to conclude that, in the semiclassical theory of radiation, widths of atomic levels in the presence of excited media are not always correctly taken into account.

Consistent quantum electrodynamic theory has no such shortcomings. This theory will now be constructed.

3. DENSITY MATRIX OF THE PHOTON SUBSYSTEM

3.1. Basic Propositions of the Theory

Consider a gas comprising two-level, but possessing Zeeman sublevels, atoms each with one valence electron. The transverse electromagnetic field will be assumed to be in quasi-resonance, $|k - \omega_{m\mu}| < k$, with optical atomic transitions. Taking into account translational degrees of freedom, let us write the Schrödinger equation in the form

$$i\frac{\partial\Psi}{\partial t} = \hat{H}\Psi, \quad \hat{H} = \hat{H}_0 + \hat{H}',$$

$$\hat{H}_0 = \hat{H}_a + \hat{H}_{ph}, \qquad \hat{H}_a = \sum_{i\mathbf{p}} \varepsilon_i(p)\hat{b}^+_{i\mathbf{p}}\hat{b}_{i\mathbf{p}},$$

$$\hat{H}_{ph} = \sum_{\mathbf{k}\lambda} k\left(\hat{\alpha}^+_{\mathbf{k}\lambda}\hat{\alpha}_{\mathbf{k}\lambda} + \frac{1}{2}\right),$$

$$\hat{H}' = -\frac{e}{m}\int \hat{\psi}^+(\mathbf{r},\mathbf{R})\hat{\mathbf{p}}\hat{\mathbf{A}}(\mathbf{r})\hat{\psi}(\mathbf{r},\mathbf{R})d\mathbf{r}d\mathbf{R},$$

$$\hat{\psi}(\mathbf{r},\mathbf{R}) = \sum_{i\mathbf{p}} \psi_i(\mathbf{r}-\mathbf{R})\frac{e^{i\mathbf{p}\cdot\mathbf{R}}}{\sqrt{V}}\hat{b}_{i\mathbf{p}},$$

$$\hat{\mathbf{A}}(\mathbf{r}) = \sum_{\mathbf{k}\lambda} \frac{\mathbf{e}^\lambda_{\mathbf{k}}}{\sqrt{2kV}} e^{i\mathbf{k}\cdot\mathbf{r}}\alpha_{\mathbf{k}\lambda} + \text{H.c.}$$

Here, $\varepsilon_i(p) = \varepsilon_i + p^2/2M$, $\mathbf{p}$ is the momentum of the atom, M is the mass, and $\hat{b}_{i\mathbf{p}}$ ($\hat{b}^+_{i\mathbf{p}}$) are the operators of annihilation (creation) of an atom in the state with momentum $\mathbf{p}$ and internal state described by wave function ψ_i. In the absence of thermal degeneracy, all operators can be considered belonging to the Bose–Einstein field.

Hamiltonian $\tilde{H}$ responsible for interaction of medium atoms with the reservoir can be introduced into the Schrödinger equation. We will not do this explicitly, but it will be assumed that the reservoir maintains the concentration of excited atoms at a constant level and broadens their energy levels.

Let us rewrite the Schrödinger equation in the integral form

$$\Psi = \Psi^0 + \mathcal{G}^0_r\hat{H}'\Psi, \quad i\frac{\partial\Psi^0}{\partial t} - \hat{H}^0\Psi^0 = 0, \tag{8}$$

where Ψ^0 is the initial state of the system before switching on Hamiltonian $\hat{H}'$,

$$i\frac{\partial\mathcal{G}^0_r}{\partial t} - \hat{H}^0\mathcal{G}^0_r = I. \tag{9}$$

The solution to (8) can be written in the form

$$\Psi = \Psi^0 + \mathcal{G}_r\hat{H}'\Psi^0, \quad i\frac{\partial\mathcal{G}_r}{\partial t} - \hat{H}\mathcal{G}_r = I. \tag{10}$$

3.2. Green Matrix Functions $\mathcal{G}^0_{ll'}$

Let

$$\mathcal{G}^0_{vac} = \mathcal{G}^0_r - \mathcal{G}^0_a, \quad \mathcal{G}^0_a = \mathcal{G}^{0+}_r.$$

Let us introduce the eigenfunctions and eigenvalues of free field operators

$$\hat{H}_{ph}\Phi_{\mathbf{N}} = \varepsilon_{ph}(\mathbf{N})\Phi_{\mathbf{N}}, \quad \hat{H}_a\chi_{\mathbf{N}} = \varepsilon_a(\mathbf{N})\chi_{\mathbf{N}}.$$

Here, the argument of the photon field is $\mathbf{N} = N_{\mathbf{k}_1\lambda_1}$, $N_{\mathbf{k}_2\lambda_2}$, …, where $N_{\mathbf{k}_j\lambda_j}$ are the occupation numbers of states ($\mathbf{k}_j$, λ_j), and the argument of the atomic field is $\mathbf{N} = N_{i_1\mathbf{p}_1}$, $N_{i_1\mathbf{p}_2}$, …, where $N_{i_j\mathbf{p}_j}$ are the occupation numbers of states (i_j, $\mathbf{p}_j$). Put

$$\begin{aligned} &\mathcal{G}^0_{21} = \mathcal{G}^0_{vac} - iZ\Psi^0\Psi^{0+}, \quad \mathcal{G}^0_{12} = -iZ\Psi^0\Psi^{0+}, \\ &\Psi^0 = \Phi_{\mathbf{N}_0}\exp[-i\varepsilon_{ph}(\mathbf{N}_0)t]\chi_{\mathbf{N}'_0}\exp[-i\varepsilon_a(\mathbf{N}'_0)t], \end{aligned} \tag{11}$$

where the $\mathbf{N}_0$ and $\mathbf{N}_0'$ vectors correspond to the initial states of the system and Z is an arbitrary dimensionless constant. By definition,

$$\mathcal{G}_{11}^0 = \mathcal{G}_{21}^0\theta^> + \mathcal{G}_{12}^0\theta^<, \quad \mathcal{G}_{22}^0 = \mathcal{G}_{21}^0\theta^< + \mathcal{G}_{12}^0\theta^>, \quad (12)$$

where $\theta^> = \theta(t - t')$ and $\theta^< = \theta(t' - t)$, θ is the Heaviside function. It follows from (12) that

$$\mathcal{G}_{11}^0 + \mathcal{G}_{22}^0 = \mathcal{G}_{12}^0 + \mathcal{G}_{21}^0 \quad (13)$$

and

$$\mathcal{G}_{11}^0 - \mathcal{G}_{12}^0 = \mathcal{G}_r^0, \quad \mathcal{G}_{11}^0 - \mathcal{G}_{21}^0 = \mathcal{G}_a^0. \quad (14)$$

The $\mathcal{G}_r^0$ and $\mathcal{G}_a^0$ functions possess retarded and advanced characters. They satisfy (9).

3.3. Matrix Green Functions $\mathcal{G}_{ll'}$

We will define these functions as solutions to the integral equation

$$\mathcal{G}_{ll'} = \mathcal{G}_{ll'}^0 + \mathcal{G}_{ll_1}^0\sigma_{l_1l_1}^{(3)}\hat{H}_{l_1l_2}'\mathcal{G}_{l_2l'}, \quad (15)$$

where $\sigma_{ll'}^{(3)} = \delta_{ll'}(-1)^{l+1}$ and the repeat subscripts imply summation.

Consider the following question. What are the properties of operator $\hat{H}_{ll'}'$ that ensure that the $\mathcal{G}_{ll'}$ function will possess algebraic properties (13), (14) inherent in the $\mathcal{G}_{ll'}^0$ function?

Let us act on Eq. (15) from the left by the matrix [24]

$$\hat{U} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

and, from the right, by the matrix $\hat{U}^{-1} = \hat{U}^+$. Equating the elements of the obtained matrix equation taking into account (13) and (14) yields four equalities:

$$\begin{aligned}\mathcal{G}_{12} + \mathcal{G}_{21} &= \mathcal{G}_{12}^0 + \mathcal{G}_{21}^0 + (\mathcal{G}_{12}^0 + \mathcal{G}_{21}^0)\hat{H}_a'\mathcal{G}_a \\ &+ \mathcal{G}_r^0\hat{H}_r'(\mathcal{G}_{12} + \mathcal{G}_{21}) + \mathcal{G}_r^0(\hat{H}_{21}' - \hat{H}_{12}')\mathcal{G}_a,\end{aligned} \quad (16)$$

$$\mathcal{G}_r = \mathcal{G}_r^0 + \mathcal{G}_r^0\hat{H}_r'\mathcal{G}_r, \quad \mathcal{G}_a = \mathcal{G}_a^0 + \mathcal{G}_a^0\hat{H}_a'\mathcal{G}_a, \quad (17)$$

$$\hat{H}_{11}' - \hat{H}_{22}' = \hat{H}_{21}' - \hat{H}_{12}', \quad (18)$$

where the notation

$$\hat{H}_r' = \hat{H}_{11}' + \hat{H}_{12}', \quad \hat{H}_a' = \hat{H}_{11}' - \hat{H}_{12}' \quad (19)$$

is used. It follows from the equality $\mathcal{G}_a = \mathcal{G}_r^+$ that

$$\hat{H}_a' = \hat{H}_r'^+. \quad (20)$$

Equation (16), in turn, yields

$$\begin{aligned}\mathcal{G}_{12} &= (1 + \mathcal{G}_r\hat{H}_r')\mathcal{G}_{12}^0(1 + \hat{H}_a'\mathcal{G}_a) - \mathcal{G}_r\hat{H}_{12}'\mathcal{G}_a, \\ \mathcal{G}_{21} &= (1 + \mathcal{G}_r\hat{H}_r')\mathcal{G}_{21}^0(1 + \hat{H}_a'\mathcal{G}_a) + \mathcal{G}_r\hat{H}_{21}'\mathcal{G}_a.\end{aligned} \quad (21)$$

Conditions (18)–(20) can be satisfied by requiring that

$$\hat{H}_{ll'}' = \delta_{ll'}\delta(t - t')\hat{H}'. \quad (22)$$

Equations (17) are now equivalent to differential equation (10). It follows from (21) and (11) that

$$\mathcal{G}_{12} = -iZ\Psi\Psi^+,$$

where Ψ satisfies (8).

We have shown that the $\Psi\Psi^+$ product, where Ψ is the solution to (8), can be found by solving integral equation (15) under condition (22). We have

$$\rho_{ph} = iZ^{-1}\langle\mathcal{G}_{12}\rangle, \quad t \longrightarrow t'. \quad (23)$$

Averaging in (23) is performed over atomic variables in both quantum and statistical senses.

3.4. Averaging over an Ensemble of Systems

Let us use the notation $\mathcal{G}_{ll'}|_{Z=0} = \Delta_{ll'}$. Retaining only the first term of the expansion of $\mathcal{G}_{ll'}$ in powers of Z, we obtain

$$\begin{aligned}\mathcal{G}_{ll'} = \Delta_{ll'} - iZ\rho_{ll'}, \quad \mathcal{G}_{ll'}^0 = \Delta_{ll'}^0 - iZ\rho_{ll'}^0, \\ \rho_{ll'}^0 = \Psi^0\Psi^{0+}.\end{aligned} \quad (24)$$

All matrix $\rho_{ll'}^0$ elements are equal. According to (23),

$$\langle\rho_{12}\rangle \xrightarrow[t \to t']{} \rho_{ph}, \quad \langle\Delta_{12}\rangle = 0.$$

Let us solve (15) iteratively setting Z equal to 0. This yields

$$\begin{aligned}\Delta &= \Delta^0 + \Delta^0\sigma^{(3)}\hat{H}'\Delta^0 \\ &+ \Delta^0\sigma^{(3)}\hat{H}'\Delta^0\sigma^{(3)}\hat{H}'\Delta^0 + \ldots\end{aligned} \quad (25)$$

Next, we will differentiate the iterative series with respect to Z and then set $Z = 0$. We arrive at

$$\rho = \rho^0 + \rho^0\sigma^{(3)}\hat{H}'\Delta^0 + \Delta^0\sigma^{(3)}\hat{H}'\rho^0 + \ldots \quad (26)$$

The structure of series (26) is identical to that of series (25) with sequential replacement of each Δ^0 matrix by ρ^0 in such a way that each term of (26) necessarily contains the ρ^0 matrix, but only raised to the power one.

Let us average (25) and (26) after multiplying them by the $\exp[i\hat{H}_a(t-t')]$ operator from the left. Consider this procedure as applied to the $\mathcal{G}^0_{ll'}$ matrix. Because

$$\mathcal{G}^0_r = -i\theta(t-t')\sum_{\mathbf{NN'}}\chi_{\mathbf{N'}}\Phi_{\mathbf{N}}\chi^+_{\mathbf{N'}}\Phi^+_{\mathbf{N}} \times \exp\{-i[\varepsilon_a(\mathbf{N'}) + \varepsilon_{ph}(\mathbf{N})](t-t')\},$$

we have

$$\langle \exp[i\hat{H}_a(t-t')]\mathcal{G}^0_r\rangle = -i\sum_{\mathbf{N}}\Phi_{\mathbf{N}}\Phi^+_{\mathbf{N}}\exp[-i\varepsilon_{ph}(\mathbf{N})(t-t')]\theta(t-t') = \langle\Delta^0_r\rangle.$$

Further, according to (11) and (24),

$$\langle\rho^0_{12}\rangle = \Phi_{\mathbf{N}_0}\Phi^+_{\mathbf{N}_0}\exp[-i\varepsilon_{ph}(\mathbf{N}_0)(t-t')],$$

and, taking into account (14), we obtain

$$\langle\Delta^0_{21}\rangle = \langle\Delta^0_r\rangle - \langle\Delta^0_a\rangle, \quad \langle\Delta^0_{12}\rangle = 0, \quad \langle\Delta^0_a\rangle = \langle\Delta^0_r\rangle^+.$$

The other $\langle\Delta^0_{ll'}\rangle$ elements are determined by equalities (12). These elements are related by Eqs. (13). This means that

$$\Delta^0_{ll'} = \langle\Delta^0_{ll'}\rangle\exp[-i\hat{H}_a(t-t')]I.$$

Let us return to series (26) and average the third term (second-order perturbation theory). We will use the interaction representation

$$\exp(i\hat{H}_a t)\psi(\mathbf{r},\mathbf{R})\exp(-i\hat{H}_a t) = \hat{\psi}(X), \quad X = \{\mathbf{r},\mathbf{R},t\}.$$

Taking into account a stochastic (Gaussian) distribution of atoms over degrees of freedom yields

$$\langle\hat{\psi}^+(X_1)\hat{\mathbf{p}}\hat{\mathbf{A}}\hat{\psi}(X_1)\hat{\psi}^+(X_2)\hat{\mathbf{p}}\hat{\mathbf{A}}\hat{\psi}(X_2)\rangle = -\hat{\mathbf{p}}\hat{\mathbf{A}}G_{21}(X_1,X_2)\hat{\mathbf{p}}\hat{\mathbf{A}}G_{12}(X_2,X_1),$$

where

$$G_{12}(X_2,X_1) = -i\langle\hat{\psi}^+(X_1)\hat{\psi}(X_2)\rangle,$$
$$G_{21}(X_1,X_2) = -i\langle\hat{\psi}(X_1)\hat{\psi}^+(X_2)\rangle.$$

Other possible terms of the performed factorization give zero as the end result. Let us introduce the notation

$$\mathcal{P}_{l_1l_2}(t_1,t_2) = -\left(\frac{e}{m}\right)^2 \times\int\hat{\mathbf{p}}\hat{\mathbf{A}}G_{21}(X_1,X_2)\langle\Delta^0_{l_1l_2}(t_1t_2)\rangle \times\sigma^{(3)}_{l_2l_2}\hat{\mathbf{p}}\hat{\mathbf{A}}G_{12}(X_2,X_1)d\mathbf{r}_1d\mathbf{R}_1d\mathbf{r}_2d\mathbf{R}_2. \tag{27}$$

The expression to be calculated takes the form

$$\langle\Delta^0_{ll_1}\rangle\sigma^{(3)}_{l_1l_1}\hat{\mathcal{P}}_{l_1l_2}\langle\Delta^0_{l_2l'}\rangle.$$

This formula was obtained in [13] by the method of Γ operators. The term describing the second order of the iterative series for the $\rho_{ll'}$ matrix is calculated in a similar way. We eventually have

$$\langle\Delta_{ll'}\rangle = \langle\Delta^0_{ll'}\rangle + \langle\Delta^0_{ll_1}\rangle\sigma^{(3)}_{l_1l_1}\hat{\mathcal{P}}_{l_1l_2}\langle\Delta^0_{l_2l'}\rangle, \tag{28}$$

$$\langle\rho_{ll'}\rangle = \langle\rho^0_{ll'}\rangle + \langle\rho^0_{ll_1}\rangle\sigma^{(3)}_{l_1l_1}\hat{\mathcal{P}}_{l_1l_2}\langle\Delta^0_{l_2l'}\rangle + \langle\Delta^0_{ll_1}\rangle\sigma^{(3)}_{l_1l_1}\hat{\mathcal{P}}^{\rho}_{l_1l_2}\langle\Delta^0_{l_2l'}\rangle + \langle\Delta^0_{ll_1}\rangle\sigma^{(3)}_{l_1l_1}\hat{\mathcal{P}}_{l_1l_2}\langle\rho^0_{l_2l'}\rangle. \tag{29}$$

The odd terms of the iterative series vanish if the medium initially does not possess coherent properties. Equations (28) and (29) found in second-order perturbation theory contain $\langle\Delta^0\rangle$ and $\langle\rho^0\rangle$ as right-hand multipliers. They are replaced by complete functions $\langle\Delta\rangle$ and $\langle\rho\rangle$ as a result of taking into account subsequent perturbation theory terms and the Dyson summation of Feynman diagrams. The derived equations are exact if $\hat{\mathcal{P}}_{ll'}$ and $\hat{\mathcal{P}}^{\rho}_{ll'}$ are understood to be complete expressions for polarization operators. In the simplest, so-called one-loop approximation, we have (27) for $\hat{\mathcal{P}}_{ll'}$, and a similar equation for $\hat{\mathcal{P}}^{\rho}_{ll'}$ in which $\langle\Delta^0\rangle$ is replaced by $\langle\rho^0\rangle$. It follows from the vanishing of $\langle\Delta^0_{12}\rangle$ that $\langle\Delta_{12}\rangle = \hat{\mathcal{P}}_{12} = 0$. Therefore,

$$\langle\Delta_{11}\rangle = \langle\Delta_r\rangle, \quad -\langle\Delta_{22}\rangle = \langle\Delta_a\rangle, \quad \langle\Delta_a\rangle = \langle\Delta_r\rangle^+,$$
$$\hat{\mathcal{P}}_{11} = \hat{\mathcal{P}}_r, \quad \hat{\mathcal{P}}_{22} = \hat{\mathcal{P}}_a, \quad \hat{\mathcal{P}}_a = \hat{\mathcal{P}}^+_r.$$

Equations of the algebraic structure of Eqs. (28) and (29) were considered in [13]. Following the reasoning of [13], we obtain

$$\langle\rho_{12}\rangle = \langle\rho_{12}\rangle^{(c)} + \langle\rho_{12}\rangle^{(n)},$$
$$\langle\rho_{12}\rangle^{(c)} = (1+\langle\Delta_r\rangle\hat{\mathcal{P}}_r)\langle\rho^0_{12}\rangle(1+\hat{\mathcal{P}}_a\langle\Delta_a\rangle),$$
$$\langle\rho_{12}\rangle^{(n)} = -\langle\Delta_r\rangle\hat{\mathcal{P}}^{\rho}_{12}\langle\Delta_a\rangle, \tag{30}$$
$$\langle\Delta_r\rangle = \langle\Delta^0_r\rangle + \langle\Delta^0_r\rangle\hat{\mathcal{P}}_r\langle\Delta_r\rangle,$$
$$\langle\Delta_a\rangle = \langle\Delta^0_a\rangle + \langle\Delta^0_a\rangle\hat{\mathcal{P}}_a\langle\Delta_a\rangle.$$

This system of equations shows that the evolution of a quantum field in a medium is generally governed by two operators, $\hat{\mathcal{P}}_r$ and $\hat{\mathcal{P}}^{\rho}_{12}$. The impossibility of using a unique medium polarization tensor is now made obvious. Coherent scattering processes are described by the $\hat{\mathcal{P}}_r$ operator, and incoherent processes including

induced radiation, by the $\hat{\mathcal{P}}_{12}^{\rho}$ operator. Induced radiation processes are incoherent because, while changing the state of atoms, they disturb the phase of the atom + field wave function and do not interfere with coherent scattering processes [27]. As shown below, the standard refractive index arises as some combination of $\hat{\mathcal{P}}_r$ and $\hat{\mathcal{P}}_{12}^{\rho}$.

Equation (27) is valid if $n\lambda\!\!\!^{-3}\gamma_r/\gamma < 1$, where n is the total concentration of scatterers, $\lambda\!\!\!^{-} = 2\pi/k$, and $\gamma_r(\gamma)$ is the radiation (total) width of excited atom energy levels. Below, in accordance with this equation, we give expanded formulas for $\hat{\mathcal{P}}_r$ and $\hat{\mathcal{P}}_{12}^{\rho}$ in the energy representation

$$f(E) = \int_{-\infty}^{\infty} e^{iEt} f(t)\,dt.$$

We have [13]

$$\hat{\mathcal{P}}_r(E) = \sum_{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2} [\hat{\alpha}_{\mathbf{k}_1\lambda_1} A_r(E - \hat{H}_{ph})\hat{\alpha}^{+}_{\mathbf{k}_2\lambda_2} + \hat{\alpha}^{+}_{\mathbf{k}_1\lambda_1} C_r(E - \hat{H}_{ph})\hat{\alpha}_{\mathbf{k}_2\lambda_2}],$$

$$A_r^{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}(E) = \sum_{m\mu\mathbf{p}} \frac{P_{m\mu}^{\lambda_1}(\mathbf{k}_1)P_{m\mu}^{\lambda_2 *}(\mathbf{k}_2)}{2\sqrt{k_1 k_2}} \times \frac{n_m(\mathbf{p})\delta_L(\mathbf{k}_2, \mathbf{k}_1)}{E + \omega_{m\mu} + (\mathbf{k}_1 + \mathbf{k}_2)\cdot\mathbf{p}/2M + i\gamma/2}, \quad (31)$$

$$C_r^{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}(E) = \sum_{m\mu\mathbf{p}} \frac{P_{m\mu}^{\lambda_1 *}(\mathbf{k}_1)P_{m\mu}^{\lambda_2}(\mathbf{k}_2)}{2\sqrt{k_1 k_2}} \times \frac{n_\mu(\mathbf{p})\delta_L(\mathbf{k}_1, \mathbf{k}_2)}{E - \omega_{m\mu} - (\mathbf{k}_1 + \mathbf{k}_2)\cdot\mathbf{p}/2M + i\gamma/2}.$$

Here, $n_m(n_\mu)$ is the concentration of excited (nonexcited) atoms on Zeemen sublevels, $\gamma = \gamma_m + \gamma_\mu$, and $\gamma_{m,\mu}$ are the collisional widths of the corresponding energy levels. For a plane parallel scattering layer of thickness L with symmetry axis z [13],

$$\delta_L(\mathbf{k}_1, \mathbf{k}_2) = \delta(k_{1x}, k_{2x})\delta(k_{1y}, k_{2y})\theta_L(k_{1z} - k_{2z}),$$

$$\theta_L(q) = \int_0^L e^{-iqz}\frac{dz}{L_z} = \frac{1 - e^{-iqL}}{iqL_z}.$$

Further,

$$\hat{\mathcal{P}}_{12}^{\rho}(E) = \sum_{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2} \Big[\hat{\alpha}^{+}_{\mathbf{k}_1\lambda_1}\int A_{12}(E - E')\langle\rho_{12}^0(E')\rangle\frac{dE'}{2\pi}\hat{\alpha}_{\mathbf{k}_2\lambda_2} \quad (32)$$

$$+ \hat{\alpha}_{\mathbf{k}_1\lambda_1}\int C_{12}(E - E')\langle\rho_{12}^0(E')\rangle\frac{dE'}{2\pi}\hat{\alpha}^{+}_{\mathbf{k}_2\lambda_2}\Big],$$

where

$$A_{12}^{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}(E) = -2\pi\sum_{m\mu\mathbf{p}} \frac{P_{m\mu}^{\lambda_1 *}(\mathbf{k}_1)P_{m\mu}^{\lambda_2}(\mathbf{k}_2)}{2\sqrt{k_1 k_2}} n_m(\mathbf{p}) \times \delta_\gamma\Big(E - \omega_{m\mu} - \frac{\mathbf{p}\cdot(\mathbf{k}_1 + \mathbf{k}_2)}{M}\Big)\delta_L(\mathbf{k}_1, \mathbf{k}_2),$$

$$C_{12}^{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}(E) = -2\pi\sum_{m\mu\mathbf{p}} \frac{P_{m\mu}^{\lambda_1}(\mathbf{k}_1)P_{m\mu}^{\lambda_2 *}(\mathbf{k}_2)}{2\sqrt{k_1 k_2}} n_\mu(\mathbf{p}) \times \delta_\gamma\Big(E + \omega_{m\mu} + \frac{\mathbf{p}\cdot(\mathbf{k}_1 + \mathbf{k}_2)}{M}\Big)\delta_L(\mathbf{k}_2, \mathbf{k}_1),$$

$$\delta_\gamma(E) = \frac{i}{2\pi}\Big[\Big(E + i\frac{\gamma}{2}\Big)^{-1} - \Big(E - i\frac{\gamma}{2}\Big)^{-1}\Big].$$

3.5. Coherent Scattering Channel

Let the scattered flux of photons be in the initial Fock state $(\mathbf{k}_0, \lambda_0)$,

$$\langle\rho_{12}^0(E)\rangle = 2\pi\delta(E - \hat{H}_{ph})|N_0\rangle\langle N_0|.$$

We will be interested in the density of photons that passed through a plane parallel scattering layer of thickness L. The photon density operator will be written as

$$\hat{n}^{\lambda_1\lambda_2}(\mathbf{r}) = \frac{1}{V} \times \sum_{\mathbf{k}_1\mathbf{k}_2} \hat{\alpha}^{+}_{\mathbf{k}_1\lambda_1}\exp[-i(\mathbf{k}_1 - \mathbf{k}_2)\cdot\mathbf{r}]\hat{\alpha}_{\mathbf{k}_2\lambda_2}.$$

In accordance with (30), in the lowest order with respect to the concentration of scattering particles, we have

$$\langle\rho_{12}\rangle^{(c)} = \langle\rho_{12}^0\rangle + \langle\Delta_r^0\rangle\hat{\mathcal{P}}_r\langle\rho_{12}^0\rangle + \langle\rho_{12}^0\rangle\hat{\mathcal{P}}_a\langle\Delta_a^0\rangle. \quad (33)$$

Consider the contribution of the second term in (33) to $\langle\hat{n}^{\lambda_0\lambda_0}(\mathbf{r})\rangle$. Equation (31) yields

$$\langle\hat{n}^{\lambda_0\lambda_0}(\mathbf{r})\rangle^{(c)} = \frac{1}{V}\int \mathrm{Sp}_{ph}\langle\Delta_r^0\rangle \times \sum_{\mathbf{k}_2} \hat{\alpha}^{+}_{\mathbf{k}_0\lambda_0}\exp[-i(\mathbf{k}_0 - \mathbf{k}_2)\cdot\mathbf{r}]\hat{\alpha}_{\mathbf{k}_2\lambda_0} \times [\hat{\alpha}_{\mathbf{k}_0\lambda_0} A_r^{\mathbf{k}_0\lambda_0\mathbf{k}_2\lambda_0}\hat{\alpha}^{+}_{\mathbf{k}_2\lambda_0} + \hat{\alpha}^{+}_{\mathbf{k}_2\lambda_0} C_r^{\mathbf{k}_2\lambda_0\mathbf{k}_0\lambda_0}\hat{\alpha}_{\mathbf{k}_0\lambda_0}]\langle\rho_{12}^0\rangle\frac{dE}{2\pi}. \quad (34)$$

Note that the term with $\mathbf{k}_1 = \mathbf{k}_2 = \mathbf{k}_0$ written separately in the polarization operator makes a zero contribution to the result of calculations because of smallness of its statistical weight when $L_z \longrightarrow \infty$. It follows that no

mode should be treated separately in the summation over $\mathbf{k}_2$ in (34). The technique of the further calculations was described in [13, 14]. According to (33), we eventually have

$$\langle \hat{n}^{\lambda_0\lambda_0}(\mathbf{r})\rangle^{(c)}_{z\to\infty} = \frac{N_0}{V}$$

$$\times\left[1 - 2\pi L\sum_{m\mu\mathbf{p}}\left|\frac{P^{\lambda_0}_{m\mu}(\mathbf{k}_0)}{\sqrt{2k_0}}\right|^2 (n_\mu(\mathbf{p}) + n_m(\mathbf{p}))\right. \qquad (35)$$

$$\left.\times\,\delta_\gamma\left(k_0 - \omega_{m\mu} - \frac{\mathbf{k}_0\cdot\mathbf{p}}{M}\right)\right].$$

Consider the scheme of calculations of $\langle \hat{n}^{\lambda\lambda}(\mathbf{r})\rangle^{(c)}$ for polarization of transmitted light perpendicular to polarization of incident radiation ($\lambda \neq \lambda_0$). According to (30), the diagonal element of this value is given by

$$\langle \rho_{12}\rangle^{(c)} = \Delta^0_r\hat{\mathcal{P}}_r\langle\rho^0_{12}\rangle\hat{\mathcal{P}}_a\langle\Delta^0_a\rangle.$$

As a result,

$$\langle \hat{n}^{\lambda\lambda}(\mathbf{r})\rangle^{(c)}_{z\to\infty}$$

$$= \frac{N_0}{V}L^2\left|\sum_{m\mu\mathbf{p}}\frac{P^{\lambda*}_{m\mu}(\mathbf{k}_0)P^{\lambda_0}_{m\mu}(\mathbf{k}_0)}{2k_0}\frac{(n_\mu(\mathbf{p}) - n_m(\mathbf{p}))(k_0 - \omega_{m\mu} - \mathbf{k}_0\cdot\mathbf{p}/M) - i\gamma(n_\mu(\mathbf{p}) + n_m(\mathbf{p}))/2}{|k_0 - \omega_{m\mu} - \mathbf{k}_0\cdot\mathbf{p}/\mathrm{M}|^2 + \gamma^2/4}\right|^2. \qquad (36)$$

We will discuss this result below. Here, we will only note that, if $|k_0 - \omega_{m\mu} - \mathbf{k}_0\cdot\mathbf{p}/M| > \gamma$, the terms in (36) depend on the $n_\mu - n_m$ difference. At the same time, if $|k_0 - \omega_{m\mu} - \mathbf{k}_0\cdot\mathbf{p}/M| < \gamma$, they depend on the $n_\mu + n_m$ sum.

3.6. Incoherent Scattering Channel

First note that the contribution of the incoherent channel to the diagonal element of $\langle n^{\lambda\lambda}(\mathbf{r})\rangle$ is positive definite at any λ. This follows from the preliminary analysis of the problem given in Section 2, and this was independently proved for (30) in [13]. It follows that Eqs. (35) and (36) determined by the coherent scattering channel estimate the sought values from below. However, for $\lambda = \lambda_0$ the term taking into account the role played by the medium has a negative sign in (35), and the positive definite term from the incoherent channel can compensate it. For this reason, the contribution of the incoherent channel is of special interest.

In accordance with (30) and (32), let us calculate the expression

$$\langle n^{\lambda_0\lambda_0}(\mathbf{r})\rangle^{(n)}_{z\to\infty} = \int \mathrm{Sp}_{ph}\sum_{\mathbf{k}_1}\hat{\alpha}^+_{\mathbf{k}_0\lambda_0}\hat{\alpha}_{\mathbf{k}_1\lambda_0}$$

$$\times\exp[i(\mathbf{k}_1 - \mathbf{k}_0)\cdot\mathbf{r}]\langle\Delta^0_r\rangle\hat{\alpha}^+_{\mathbf{k}_1\lambda_0} \qquad (37)$$

$$\times\int A^{\mathbf{k}_1\lambda_0\mathbf{k}_0\lambda_0}_{12}\langle\rho^0_{12}\rangle\frac{dE'}{2\pi}\hat{\alpha}_{\mathbf{k}_0\lambda_0}\langle\Delta^0_a\rangle\frac{dE}{2\pi V} + \text{c.c.}$$

Note that although the incoherent channel is responsible for the contribution of induced processes to the overall scattering intensity, the term with $\mathbf{k}_1 = \mathbf{k}_2 = \mathbf{k}_0$ in the $\hat{\mathcal{P}}^{\rho}_{12}$ operator makes no contribution because of the smallness of its statistical weight. Formally, this term contains the L_z^{-1} multiplier and vanishes as $V \longrightarrow \infty$. Such a form of calculations of induced radiation is typical of media of finite dimensions. For the same reason, the term with C_{12} contained in $\hat{\mathcal{P}}^{\rho}_{12}$ falls out of calculations.

Subsequent transformations of (37) with ignoring spontaneous radiation give an almost obvious result,

$$\langle \hat{n}^{\lambda_0\lambda_0}(\mathbf{r})\rangle^{(n)} = 4\pi L\frac{N_0}{V}$$

$$\times\sum_{m\mu\mathbf{p}}\left|\frac{P^{\lambda_0}_{m\mu}(\mathbf{k}_0)}{\sqrt{2k_0}}\right|^2 n_m(\mathbf{p})\delta_\gamma\left(k_0 - \omega_{m\mu} - \frac{\mathbf{k}_0\mathbf{p}}{M}\right).$$

The sum of this expression and (35) can be written as

$$\langle \hat{n}^{\lambda_0\lambda_0}(\mathbf{r})\rangle^{(c)} + \langle \hat{n}^{\lambda_0\lambda_0}(\mathbf{r})\rangle^{(n)}$$

$$= \frac{N_0}{V}[1 - 2k_0L\,\mathrm{Im}\,\kappa^{\lambda_0}(\mathbf{k}_0)], \qquad (38)$$

where

$$\kappa^{\lambda_0}(\mathbf{k}_0) = 1 - \frac{1}{k_0}\sum_{m\mu}\int\frac{d\mathbf{p}}{(2\pi)^3}\left|\frac{P^{\lambda_0}_{m\mu}(\mathbf{k}_0)}{\sqrt{2k_0}}\right|^2$$

$$\times\frac{n_\mu(\mathbf{p}) - n_m(\mathbf{p})}{k_0 - \omega_{m\mu} - \mathbf{k}_0\cdot\mathbf{p}/M + i\gamma/2},$$

as $V \longrightarrow \infty$, and $\kappa^{\lambda_0}(\mathbf{k}_0)$ is simply the diagonal element of the standard index of refraction of the medium which can be obtained in terms of the theory described in [15]. Like medium polarizability, this index of refraction depends on the $n_\mu - n_m$ difference and the concentrations of atoms on the lower and upper atomic transition levels. According to (38), the medium becomes less active as $n_m \longrightarrow n_\mu$.

Another conclusion follows from (36). The intensity of light that passes through the medium with a 2π change in its linear polarization direction is determined by the $n_\mu + n_m$ sum under the conditions of resonance with some Zeeman frequency $k = \omega_{m\mu}$ and does not decrease as $n_m \longrightarrow n_\mu$. Taking into account the contribution of the incoherent channel cannot change this result, because this contribution is positive definite.

We arrive at the conclusion that this intensity cannot be described by the refractive index of the medium that contains excited atoms. Its description requires the use of another medium characteristic, which depends, in particular, on the $n_\mu + n_m$ sum.

4. CONCLUSION

The results of this work, in particular, formulas (36) and (38), were obtained in the quasi-resonance approximation for two-level scatterers. It was assumed that the interaction of an excited atom with radiation caused light scattering or induced radiation. Absorption was ignored. It follows that the eventual formula would be substantially different if the medium contained scatterers with an equidistant energy spectrum.

This insufficiency of the description of the optical properties of excited dispersing media in a semiclassical radiation theory is caused by not taking into account the photon–photon correlators. Like the medium refractive index, the new characteristic of the medium that appears in our analysis, which depends on the $n_\mu + n_m$ sum, depends on the Planck constant only through oscillator strengths. This is quite natural, if the essentially quantum two-level object of scattering is taken into account. Note that the dependence of the refractive index on the $n_\mu - n_m$ difference is also a consequence of quantum theory. Such a difference cannot be obtained for light scattering on classical oscillators.

In this work, we did not explicitly take into account the role played by the incoherent channel in the formation of the intensity of light scattered with a $\pi/2$ change in polarization. We only noted that this role was positive definite and required going beyond the scope of perturbation theory. By analogy with studies performed in [14], it can be expected that taking this channel into account would result in an observable frequency broadening of scattered light because of a change in the number of quanta in the system. This effect would be determined by the coefficient of absorption of photons by the medium.

In testing the suggested theory, studying the Hanle effect in a longitudinal magnetic field may be useful. We call attention to works [10, 11], in which an electric discharge in inert gases served as a scatterer. In such experiments, the existence of a noticeable fraction of excited scatterers should be expected. A series of experiments with different discharge currents would be of interest, because the dependences of passed light intensities with different λ on n_μ and n_m are qualitatively different. If, in addition, side scattering intensities of spectral lines related to the $^3S \longleftrightarrow {}^3P$ and $^3P \longleftrightarrow {}^3D$ transitions are measured, for instance, when discharges occur in He, then the relative occupancies of the lower n_μ and upper n_m energy levels, transitions between which are responsible for the Hanle effect, can be determined.

Such an experiment would not only allow us to refine the rules governing the Hanle effect in forward scattering but would also call attention to the general features of constructing quantum electrodynamics for excited dispersing media.

REFERENCES

1. A. Corney, B. P. Kibble, and G. Series, Proc. R. Soc. London, Ser. A **293**, 70 (1966).
2. S. Giraud-Cotton, V. P. Kaftandjian, and L. Klein, Phys. Rev. A **32**, 2211 (1985).
3. B. Lobodziński and W. Gawlik, Phys. Scr. T **T70**, 138 (1997).
4. B. W. Holmes and J. A. P. Griffith, J. Phys. B **28**, 2829 (1995).
5. W. Gawlik, J. Kawalski, R. Neumann, and F. Trager, Phys. Lett. A **48**, 283 (1974).
6. B. Ståhlberg, P. Jungner, T. Fellman, and Å. Lindberg, Appl. Phys. B **B50**, 547 (1990).
7. X. Chen, V. L. Telegli, and A. Weis, Opt. Commun. **74**, 301 (1990).
8. L. M. Barkov, D. A. Melik-Pashaev, and M. S. Zolotarev, Opt. Commun. **70**, 467 (1989).
9. G. Stansel, Phys. Lett. A **47**, 283 (1974).
10. N. G. Lukomskiĭ, V. A. Polishchuk, and M. P. Chaĭka, Opt. Spektrosk. **58**, 474 (1985) [Opt. Spectrosc. **58**, 284 (1985)].
11. N. G. Lukomskiĭ, V. A. Polishchuk, M. P. Chaĭka, and E. N. Amplieva, Opt. Spektrosk. **83**, 420 (1997) [Opt. Spectrosc. **83**, 393 (1997)].
12. M. P. Chaĭka, *Interference of Degenerate Atomic States* (Leningr. Gos. Univ., Leningrad, 1975).
13. B. A. Veklenko, Zh. Éksp. Teor. Fiz. **96**, 457 (1989) [Sov. Phys. JETP **69**, 258 (1989)].
14. B. A. Veklenko, R. B. Gusarov, and Yu. B. Sherkunov, Zh. Éksp. Teor. Fiz. **113**, 521 (1998) [JETP **86**, 289 (1998)].
15. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskii, *Methods of Quantum Field Theory in Statistical Physics* (Gostekhizdat, Moscow, 1962; Prentice-Hall, Englewood Cliffs, 1963).
16. V. Korenmann, Ann. Phys. (N. Y.) **39**, 72 (1962).

17. B. A. Veklenko and G. B. Tkachuk, Opt. Spektrosk. **38**, 1132 (1975) [Opt. Spectrosc. **38**, 653 (1975)].

18. I. B. Levinson, Zh. Éksp. Teor. Fiz. **75**, 234 (1978) [Sov. Phys. JETP **48**, 117 (1978)].

19. H. R. Zaidi, Can. J. Phys. **59**, 737 (1981); **59**, 750 (1981).

20. A. A. Panteleev, V. A. Roslyakov, A. N. Starostin, and M. D. Taran, Zh. Éksp. Teor. Fiz. **97**, 1777 (1990) [Sov. Phys. JETP **70**, 1003 (1990)].

21. S. Fujita, Physika **28**, 281 (1962).

22. I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience, New York, 1962; Mir, Moscow, 1964).

23. M. O. Scally and W. E. Lamb, Phys. Rev. **159**, 208 (1967).

24. L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1515 (1964) [Sov. Phys. JETP **20**, 1018 (1964)].

25. J. R. Klauder and E. C. G. Sudarshan, *Fundamentals of Quantum Optics* (Syracuse University, New York–Amsterdam, 1968; Mir, Moscow, 1970).

26. A. S. Davydov, *Quantum Mechanics* (Fizmatgiz, Moscow, 1963; Pergamon, Oxford, 1976).

27. B. A. Veklenko, Izv. Vyssh. Uchebn. Zaved., Fiz., No. 6, 132 (1987).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 951–959.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1099–1108.

ATOMS, SPECTRA, RADIATION

Resonant Charge Exchange with the *p*-Electron Transition¶

B. M. Smirnov

Institute for High Temperatures, Russian Academy of Sciences, Moscow, 127412 Russia
e-mail: smirnov@orc.ru

Received December 8, 2000

Abstract—The asymptotic resonant charge exchange theory is developed for slow collisions of atoms and ions with valent *p*-electrons. Because of a small rotation angle of the molecular axis in the course of the *p*-electron transition, the resonant charge exchange cross section is not sensitive to the rotational energy of colliding particles, and the cross sections are nearly equal for cases "*a*", "*b*", and "*d*" of the Hund coupling, and also for cases "*c*" and "*e*" of the Hund coupling. The cross sections of the resonant charge exchange process are evaluated under various conditions and for various elements of the periodical table with *p*-electron shells of atoms and ions. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The resonant charge exchange process corresponds to the tunnel transition of a valent electron from the field of one atomic core to another in the course of a collision of an ion with the parent atom. Under real conditions, one can consider nuclei to move along classical trajectories [1–3]: at small collision velocities compared to a typical atomic velocity, the rate of this process can be expressed through the parameters of a molecular ion consisting of the colliding atom and its ion [4, 5]. In particular, for the transiting *s*-electron and structureless cores, with only two electron terms of the molecular ion involved in this process, the probability P_{res} of the electron transition from one core to another after the collision is given by [4]

$$P_{\text{res}} = \sin^2\zeta(\rho), \quad \zeta(\rho) = \int_{-\infty}^{\infty}\frac{\Delta(R)}{2}dt, \tag{1}$$

where $R(t)$ is the distance between the nuclei, t is time, ρ is the impact parameter of the collision, $\zeta(\rho)$ is the exchange phase, and the ion–atom exchange interaction potential is defined as

$$\Delta(R) = \varepsilon_g(R) - \varepsilon_u(R), \tag{2}$$

$\varepsilon_g(R)$ and $\varepsilon_u(R)$ being the energies of the even and odd states of the molecular ion. These states are characterized by different symmetries of the electron reflection with respect to the symmetry plane, which is perpendicular to the line joining the nuclei and bisects it.

In reality, the resonant charge exchange cross section is large compared to a typical atomic value of this dimensionality. This implies that transitions at large separations give the leading contribution to the cross section. This fact is the basis of the asymptotic theory of the resonant charge exchange [6, 7], where the cross section is expanded with respect to a small parameter that is inversely proportional to the typical separation for the electron transition. Within the framework of the asymptotic theory, the cross section of this process is expressed through the collision parameters and the radial wave function parameters of the transiting valent electron in the atom when the electron is located far from the core. In particular, for the *s*-electron transition between structureless cores, the cross section of the resonant charge exchange is [7]

$$\sigma_s = \int_0^{\infty}2\pi\rho d\rho\sin^2\zeta(\rho) = \frac{\pi R_0^2}{2},$$
$$\zeta(R_0) = \frac{e^{-C}}{2} = 0.28, \tag{3}$$

where $C = 0.577$ is the Euler constant and the exchange phase for the free motion of the nuclei $R = \sqrt{v^2t^2 + \rho^2}$ is given by [7]

$$\zeta(\rho) = \int_{-\infty}^{\infty}\frac{\Delta(R)}{2}dt = \frac{1}{v}\sqrt{\frac{\pi\rho}{2\gamma}}\Delta(\rho)$$
$$= \frac{1}{v}\sqrt{\frac{\pi}{2\gamma}}A^2e^{-1/\gamma}\rho^{2/\gamma - 1/2}\exp(-\rho\gamma). \tag{4}$$

Here, v is the collision velocity, $\gamma^2/2$ is the atom ionization potential or the electron binding energy, and A is the asymptotic coefficient of the electron wave function in this atom; we use atomic units where $\hbar = m_e = e = 1$. The asymptotic parameters of the atom are contained in

¶This article was submitted by the author in English.

the normalized radial wave function of the electron at large distances r from the core,

$$\psi(r \longrightarrow \infty) = Ar^{1/\gamma - 1}e^{-r\gamma}. \quad (5)$$

In particular, $A = 2$ for the hydrogen atom [8].

We note that the small asymptotic theory parameter is $1/R_0\gamma$, and because the asymptotic theory involves the information about the electron location far from the core, we can keep only two expansion terms in the expression for the cross section, as we did in Eq. (3). This corresponds to the asymptotic theory accuracy for the transition of a valent s-electron in the range 1–5% [9] for the eV-collision energies.

For a transiting p-electron, the asymptotic theory of the resonant charge exchange process becomes cumbersome [10, 11] because, on the one hand, the electron transfer process is entangled with the rotation of the molecular axis, and on the other hand, the fine splitting of the ion and atom levels can be important in this process. The above formulas for the s-electron transition were used in some recent calculations [12–14] as model ones for the evaluation of the cross section of this process for the transiting p-electron. This leads to an uncontrolled error that is absent in the asymptotic theory. We note that the contemporary computer technique allows us to formulate the asymptotic theory in a simple way and to exhibit various aspects of the process under consideration. In this paper, we present the asymptotic theory for the resonant charge exchange process with the transiting p-electron. We are guided mostly by the ground states of the colliding ion and atom and focus on the cross section averaging over the directions of the initial momenta.

2. RESONANT CHARGE EXCHANGE WITH THE TRANSITING p-ELECTRON FOR LIGHT ATOMS

We first consider the case of a small spin–orbit splitting of atom and ion levels and neglect relativistic interactions. This corresponds to the LS-coupling scheme for the atom. At large separations, the quantum numbers of the molecular ion are the atom quantum numbers LSM_LM_S (the orbital momentum, spin, and their projections on the molecular axis) and the same quantum numbers of the ion $lsmm_s$. The atom orbital and spin momenta L and S are given by the sum of the electron orbital and spin momenta l_e and 1/2 and of the respective momenta of the atomic core l and s. The atom spin S and the spin of the other atom core s are then summed into the total spin I of the molecular ion. The atomic wave function is then expressed through parameters of the core and the valent electron by means of the genealogical or Racah coefficients [15, 16]. The ion–atom exchange interaction potential is then given by [10, 17]

$$\Delta(l_e\mu, lms, LM_LS) = \frac{\bar{I} + 1/2}{2s + 1} n (G_{ls}^{LS})^2 \times \begin{bmatrix} l_e & l & L \\ \mu & m & m + \mu \end{bmatrix} \begin{bmatrix} l_e & l & L \\ \mu & M_L - \mu & M_L \end{bmatrix} \Delta_{l_e\mu}, \quad (6)$$

where n is the number of identical valent electrons of the atom, G_{ls}^{LS} is the genealogical (Racah) coefficient [15, 16], the square brackets are the Clebsch–Gordan coefficients that are responsible for the summation of the electron and ion orbital momenta into the atom orbital momentum, and $\Delta_{l_e\mu}$ is the exchange interaction potential for one valent electron located in the field of structureless cores. We note a weak dependence of the exchange interaction potential on the total spin I of the molecular ion. Indeed, the level splitting corresponding to different total spins of the molecular ion is determined by the exchange of two electrons and varies at large separations R as $\exp(-2\gamma R)$. Therefore, Eq. (6) contains the average spin of the molecular ion. Next, because the exchange interaction potential $\Delta_{l_e\mu}$ decreases as $R^{-\mu}$ with increasing μ, we are restricted by the term with the minimum value of μ in Eq. (6). As a result, for the valent p-electron, we have [10, 17]

$$\Delta_{10}(R) = 3\Delta_0, \quad \Delta_{11}(R) = \frac{6}{R\gamma}\Delta_0, \quad (7)$$

where [7]

$$\Delta_0 = A^2 R^{(2/\gamma) - 1} e^{-R\gamma - (1/\gamma)} \quad (8)$$

is the exchange interaction potential for a valent s-electron [7] with the same asymptotic radial wave function.

Formula (6) allows one to construct the matrix of the exchange interaction potential of an ion and an atom with valent p-electrons. Below, we represent these matrices in the case where the atom and the ion are in the ground electron states. One can be convinced of the identity of the transiting electron and the hole. In accordance with Eq. (6) for atoms of group 3 (with one valent p-electron) and atoms of group 8 (with one valent p-hole) in the periodical table of elements, with the ground states of the atom and the ion given by ^1S and ^2P, the exchange interaction potential of the interacting atom and the ion is given by the matrix

$$\Delta(M_L) = \begin{array}{|c|c|c|} \hline M_L = -1 & M_L = 0 & M_L = +1 \\ \hline \Delta_{11} & \Delta_{10} & \Delta_{11} \\ \hline \end{array}, \quad (9a)$$

where M_L is the orbital momentum projection for the atom (elements of group 3) or the ion (elements of group 8).

For elements of groups 4 and 7 of the periodical table, with the ground electron states of the atom and the ion given by ^3P and ^2P, the matrix of the exchange interaction potential, in accordance with Eq. (6), is

$$\Delta(m, M_L) = \frac{5}{3} \times \begin{array}{|c|c|c|c|} \hline & M_L = -1 & M_L = 0 & M_L = +1 \\ \hline m = -1 & \Delta_{10} & \Delta_{11} & \Delta_{10} \\ \hline m = 0 & \Delta_{11} & 2\Delta_{11} & \Delta_{11} \\ \hline m = 1 & \Delta_{10} & \Delta_{11} & \Delta_{10} \\ \hline \end{array}, \quad (9b)$$

where m and M_L are the projections of the orbital ion and atom momenta.

For elements of groups 5 and 6 of the periodical table with the atom and ion ground states ^4S and ^3P, the matrix of the exchange interaction potential is given by

$$\Delta(m) = \frac{7}{3} \begin{array}{|c|c|c|} \hline m = -1 & m = 0 & m = 1 \\ \hline \Delta_{11} & \Delta_{10} & \Delta_{11} \\ \hline \end{array}. \quad (9c)$$

As the quantization axis, we take the direction on which the projection of the electron momentum is zero. We let θ denote the angle between the quantization and molecular axes. By definition, the exchange interaction potential Δ(θ) of the atom and its ion with valent *p*-electrons is equal to

$$\Delta(\theta) = \frac{1}{3}\sum_M \left|d^1_{M0}(\theta)\right|^2 \Delta_{1M} = \frac{4\pi}{3}\sum_M |Y_{1M}(\theta, \varphi)|^2 \Delta_{1M},$$

where $d^1_{M0}(\theta)$ is the Wigner rotation function [20] and $Y_{1M}(\theta)$ is the spherical function; it follows that $4\pi|Y_{1M}(\theta)|^2$ is the probability of finding a state with the momentum projection M at the angles θ and φ with respect to the molecular axis. The spherical function satisfies the normalization condition

$$\int_{-1}^{1} d\cos\theta |Y_{1M}(\theta)|^2 = \frac{1}{4\pi}.$$

It follows that for groups 3 and 8 of the periodical table of elements, the exchange interaction potential of an atom and a parent ion is given by

$$\Delta(\theta) = \Delta_{10}\cos^2\theta + \Delta_{11}\sin^2\theta. \quad (10a)$$

Matrix (9b) gives the ion–atom exchange interaction potential as a function of the angles between the quantization and molecular axes for elements of groups 4 and 7 of the periodical system,

$$\Delta(\theta) = \frac{5}{3} \times [\Delta_{10}\sin^2\theta_1\sin^2\theta_2 + \Delta_{11}(\cos^2\theta_1 + \cos^2\theta_2)], \quad (10b)$$

where θ_1 and θ_2 are the respective angles between the molecular axis and the quantization axes for the atom and the ion, with zero electron momentum projection on the quantization axis. For groups 4 and 7 of the periodical system, the exchange interaction potential is similar to that for atoms of groups 3 and 8 and is given by

$$\Delta(\theta) = \frac{7}{3}(\Delta_{10}\cos^2\theta + \Delta_{11}\sin^2\theta). \quad (10c)$$

Although we are restricted by the ground states of the ion and the parent atom, this is a general scheme of constructing the ion–atom exchange interaction potential. Being averaged over the total quasimolecule spin I, the exchange interaction potential depends on the ion (m) and the atom (M_L) angular momentum projections on the molecular axis. This corresponds to the *LS*-coupling scheme for atoms and ions, which means neglecting the spin-orbital interaction. Therefore, the above expressions correspond to the hierarchy of the interaction potentials

$$V_{ex} \gg U(R), \quad \Delta(R), \quad (11)$$

where V_{ex} is the typical exchange interaction potential for valent electrons inside the atom or the ion, $U(R)$ is the long-range interaction potential between the atom and the ion at large separations R, and $\Delta(R)$ is the exchange interaction potential between the atom and the ion. Within the framework of the *LS*-coupling scheme for atoms and ions, we assume the excitation energies inside the electron shell to be relatively large. This criterion is fulfilled for light atoms and ions. In the same manner, one can construct the exchange interaction potential matrix for excited states within a given electron shell.

Because the exchange interaction potential is determined by the transition of one electron from the valent electron shell and the transiting electron carries a certain momentum and spin, additional selection rules apply for the one-electron interaction. In particular, for the transition of a *p*-electron, the selection rules are given by

$$|L - l| \le 1, \quad |S - s| \le 1/2. \quad (12)$$

These selection rules follow from the properties of the Clebsch–Gordan coefficients entering Eq. (6). If these conditions are violated, the ion–atom exchange interaction potential is zero at the scale of one-electron interaction potentials. In Table 1, we list the states of atoms and their ions with valent *p*-electrons for which the

Table 1. The ion and parent atom states with valent *p*-electrons. The one-electron transition is forbidden between these states and the exchange interaction potential of the ion and the parent atom is zero

Electron configuration and ion state	Electron configuration and atom state
$p^2(^1D)$	$p^3(^4S)$
$p^2(^1S)$	$p^3(^4S)$
$p^2(^1S)$	$p^3(^2D)$
$p^3(^4S)$	$p^4(^1D)$
$p^3(^4S)$	$p^4(^1S)$
$p^3(^2D)$	$p^4(^1S)$

ion–atom one-electron exchange interaction potential is zero.

When a valent *p*-electron transits from one atomic core to the other in the course of collision, the processes of charge exchange and the electron momentum rotation are entangled. One can partially separate these processes because the charge exchange proceeds in a narrow range of separations R where the molecular axis turns at a small angle of the order $1/\sqrt{R_0\gamma}$. Indeed, the range of distances ΔR between the nuclei where the charge exchange phase ζ varies considerably is $\Delta R \sim 1/\gamma$, which corresponds to the rotation angle $\vartheta \sim vt/R \sim 1/\sqrt{R\gamma} \ll 1$. This fact allows us to simplify the derivation of the resonant charge exchange cross section. Below we find the cross section averaged over angles between the collision impact parameter and the quantization axis on which the orbital momentum projection is zero.

Proceeding in this way, we orientate the quantization axis with respect to the direction of the collision impact parameter as the motionless axis; the average cross section of the resonant charge exchange is then given by

$$\bar{\sigma} = \frac{1}{4\pi}\int_{-1}^{1}\int_{0}^{2\pi}\sigma(\vartheta, \varphi)d\cos\vartheta d\varphi, \qquad (13)$$

where ϑ and φ are the polar angles of the impact parameter direction with respect to the quantization axis direction and $\sigma(\vartheta, \varphi)$ is the charge exchange cross section at a given direction of the collision impact parameter. We first consider the cases where the momentum projection on the molecular axis is conserved in the course of the electron transfer. This corresponds to a small rotational energy and, under condition (11), is determined by the condition

$$\frac{v}{\rho} \ll U(\rho), \Delta(\rho), \qquad (14)$$

which means that the rotational energy v/ρ at the closest approach distance is small compared to the long-range splitting $U(\rho)$ of molecular levels for states with different projections of the orbital momentum on the molecular axis or to the exchange interaction splitting $\Delta(\rho)$. Criterion (14) corresponds to cases "*a*" and "*b*" of the Hund coupling [18, 19]. For the free motion of colliding particles, the current angle θ between the molecular and quantization axes and the angle ϑ between these axes at the closest approach distance are related by

$$\cos\theta = \cos\vartheta\cos\alpha + \sin\vartheta\sin\alpha\cos\varphi, \qquad (15)$$

where α and φ are the polar angles of the molecular axis with respect to its direction at the closest approach distance; we have $\sin\alpha = vt/R$, where v is the collision velocity, t is time, and R is the current distance between the colliding particles.

Using Eqs. (15) and (10), we can represent the exchange phases in the form of an expansion in the small parameter of the theory $1/\rho\gamma$ if criterion (14) is satisfied. For elements of groups 3, 5, 6, and 8 of the periodical system, we have

$$\zeta(\rho, \vartheta, \varphi) = \zeta(\rho, 0) \times\left[\cos^2\vartheta - \frac{1}{\gamma\rho}\cos^2\vartheta + \frac{1}{\gamma\rho}\sin^2\vartheta(2 + \cos^2\varphi)\right]. \qquad (16a)$$

This expression applies to large collision impact parameters, and $\zeta(\rho, 0)$ is the phase of the charge exchange process when the quantization axis has the same direction as the molecular axis at the closest approach distance; $\zeta(\rho, 0) = 3\zeta_0(\rho)$ for elements of groups 3 and 8 and $\zeta(\rho, 0) = 7\zeta_0(\rho)$ for elements of groups 5 and 6, where the charge exchange phase ζ_0 is given by Eq. (4) and is related to the *s*-electron transition with the same asymptotic parameters γ and A. For atoms of groups 4 and 7, the charge exchange phase at large impact parameters of the collision is given by

$$\zeta = 5\zeta_0(\rho)\left\{\sin^2\vartheta_1\sin^2\vartheta_2 + \frac{1}{\gamma\rho}\right. \times[2\cos^2\vartheta_1 + 2\cos^2\vartheta_2 + \sin^2\vartheta_1\cos^2\vartheta_2 + \cos^2\vartheta_1\sin^2\vartheta_2 - \sin^2\vartheta_1\sin^2\vartheta_2(\cos^2\varphi_1 + \cos^2\varphi_2) + \left.\sin 2\vartheta_1\sin 2\vartheta_2\cos\varphi_1\cos\varphi_2]\right\}, \qquad (16b)$$

where ϑ_1, φ_1 and ϑ_2, φ_2 are the respective polar angles of the quantization axes of the atom and the ion relative to the molecular axis at the closest approach distance.

We now determine the average cross section from formula (13) using formula (4) for the cross section at a given angle and the dependence $\zeta \propto \exp(-\gamma\rho)$ for the

exchange phase. For elements of groups 3, 5, 6, and 8 of the periodical system, this gives

$$\frac{\bar{\sigma}}{\sigma_s} = \frac{1}{2\pi} \times \int_0^{1}\int_0^{2\pi} \left[1 + \frac{1}{\gamma R_0}\ln\frac{\zeta(R_0, \vartheta, \varphi)}{\zeta_0(R_0)}\right]^2 d\cos\vartheta d\varphi, \quad (17)$$

where $R_0 = \sqrt{2\sigma_s/\pi}$. The same expression applies to elements of groups 4 and 7 with the integration over four angles ϑ_1, φ_1, ϑ_2, φ_2; here σ_s is the cross section in Eq. (3) for the transiting *s*-electron with the same asymptotic parameters as for the *p*-electron. Table 2 contains these reduced cross sections depending on the parameter $R_0\gamma$. The value of Σ_3 in Table 2 gives $\bar{\sigma}/\sigma_s$ for elements of groups 3 and 8 of the periodical table, the value of Σ_4 is $\bar{\sigma}/\sigma_s$ for elements of groups 4 and 7, and the value of Σ_5 gives $\bar{\sigma}/\sigma_s$ for elements of groups 5 and 6. In addition, this table contains the reduced cross sections Σ_{10} and Σ_{11} corresponding to the respective projections 0 and 1 of the orbital momentum on the impact parameter direction.

We also consider the opposite case to (14), where

$$\frac{v}{\rho} \gg U(\rho), \Delta(\rho). \quad (18)$$

In this case, the exchange phase for elements of groups 3, 5, 6, and 8 is given by

$$\zeta(\rho, \vartheta, \varphi) = \zeta(\rho, 0)\left(\cos^2\vartheta + \frac{2}{\gamma\rho}\sin^2\vartheta\right) \quad (19)$$

instead of (16a). The corresponding reduced average cross section is denoted by Σ_3^d in Table 2. This corresponds to cases "*b*" and "*d*" of the Hund coupling, and according to the data in Table 2, the results for this case practically coincide with case "*a*" of the Hund coupling. The value $\bar{\Sigma}$ in Table 2 is

$$\bar{\Sigma} = \Sigma_{10}/3 + 2\Sigma_{11}/3.$$

Comparing it to the average cross section one testifies to the sensitivity of the cross section to different methods of averaging.

3. RESONANT CHARGE EXCHANGE WITH THE TRANSITING *P*-ELECTRON FOR HEAVY ATOMS

In considering the resonant charge exchange process, we follow the general scheme of classifying the limiting cases of momentum coupling in diatomic molecules. This scheme is given in Table 3 [18, 19]. The cases of the Hund coupling unify different relations between energetic parameters of colliding particles. An important energetic parameter of the quasimolecule consisting of colliding particles is the interaction potential V_e between the orbital angular momentum of the electrons and the molecular axis. This includes the exchange interaction potential V_{ex} inside the atom and the ion due to the Pauli exclusion principle, the splitting of the molecular ion levels due to the long-range interaction $U(R)$, and the exchange interaction potential $\Delta(R)$ between the ion and the atom. Within the framework of Hund schemes, we compare this interaction potential to the relativistic interaction δ_f given by the sum of spin–orbit interactions of the individual electrons and other relativistic interactions and the rotational energy given by $V_r = v\rho/R^2$ for the free motion of colliding particles. For colliding atomic particles, in contrast to a molecule, different types of the Hund coupling can be realized on one classical trajectory of particles. We use the general Nikitin scheme [21–23], which relates the characters of the momentum coupling of colliding atomic particles moving along one trajectory. The problem under consideration is simpler because we are interested in the behavior of colliding particles on the trajectory element where the electron transition occurs. Only one type of the momentum coupling is realized on this part of the trajectory.

Above, we considered the cases where relativistic interactions are negligible and the projection of the

Table 2. The reduced cross sections of the resonant charge exchange

$R_{0\gamma}$	6	8	10	12	14	16
Σ_{10}	1.40	1.29	1.23	1.19	1.16	1.14
Σ_{11}	1.08	0.98	0.94	0.92	0.91	0.91
$\bar{\Sigma}$	1.19	1.08	1.04	1.01	0.99	0.95
Σ_3	1.17	1.09	1.05	1.03	1.02	1.01
Σ_3^d	1.16	1.08	1.04	1.02	1.01	1.00
Σ_4	1.50	1.32	1.23	1.18	1.14	1.12
Σ_5	1.44	1.29	1.22	1.17	1.14	1.11
$\Sigma_{1/2}$	1.18	1.10	1.07	1.05	1.04	1.03
$\Sigma_{3/2}$	1.18	1.10	1.06	1.04	1.03	1.02
$\Sigma_{3/2}^e$	1.16	1.09	1.06	1.04	1.03	1.02

Table 3. The Hund coupling cases

Hund coupling case	Relation
a	$V_e \gg \delta_f \gg V_r$
b	$V_e \gg V_r \gg \delta_f$
c	$\delta_f \gg V_e \gg V_r$
d	$V_r \gg V_e \gg \delta_f$
e	$V_r \gg \delta_f \gg V_e$
e'	$\delta_f \gg V_r \gg V_e$

Table 4. The ground states of atoms with *p*-electron shells in the framework of the *LS* and *jj* coupling schemes and the ion–atom exchange interaction potential Δ for the Hund coupling cases "*c*" and "*e*"

Shell	*J*	*LS*-term	*jj*-shell	Δ
p	1/2	$^2P_{1/2}$	$[1/2]^1$	$\Delta_{1/2}$
p^2	0	$^3P_0$	$[1/2]^2$	$\Delta_{1/2}$
p^3	3/2	$^4S_{3/2}$	$[1/2]^2[3/2]^1$	$\Delta_{3/2}$
p^4	2	$^3P_2$	$[1/2]^1[3/2]^3$	0
p^5	3/2	$^2P_{3/2}$	$[1/2]^2[3/2]^3$	$\Delta_{1/2}$
p^6	0	$^1S_0$	$[1/2]^2[3/2]^4$	$\Delta_{3/2}$

orbital electron momentum on the molecular or the motionless axis is conserved in the course of the electron transfer, which corresponds to cases "*a*", "*b*", and "*d*" of the Hund coupling and is realized for light atomic particles. In what follows, we are guided by heavy atoms and examine the cases where relativistic interactions are important.

In the collisions of heavy atomic particles, the relativistic interactions are dominant, and therefore, the *jj* coupling scheme becomes valid for an individual atomic particle. The quantum numbers of the interacting atom and ion are therefore given by J and M_J, the total atomic electron momentum and its projection on the molecular axis for the atom, and also by the respective ion quantum numbers j and m_j. At large separations, these quantum numbers are related to a molecular ion consisting of the ion and the parent atom. We note that the total momentum J and its projection on a given direction M_J are the quantum numbers of an individual atomic particle in both momentum coupling schemes (*LS* and *jj*), which simplifies the analysis in the general case. Next, taking the relativistic effects into account reduces the atom symmetry. For this reason, on the one hand, the ion–atom exchange interaction potential is expressed through the one-electron exchange interaction potential in a simpler way, and on the other hand, the prohibition of some one-electron transitions strengthens in the presence of relativistic interactions because of a weaker mixing of states in this case. Table 4 contains parameters of the electron shells for the ground electron states of atoms and ions with *p*-electron shells. We note that for the *jj*-coupling, the similarity between the transitions of the *p*-electron and the *p*-hole is lost because of different signs of the spin–orbit interaction potential for the electron and the hole. Hence, the ion–atom exchange interaction potential is different in the cases where the *p*-electron shells of the atom and its ion are replaced by shells consisting of identical *p*-holes. Moreover, for group 6 of the periodical system of elements, the one-electron ion–atom exchange interaction potential is zero if the atom and the ion are found in the ground states. We note that for all the groups in the periodic table of elements with valent *p*-electrons, the ion–atom one-electron exchange interaction potential is not zero for light atoms if atoms and their ions are found in the ground states.

It follows from the data in Table 4 that the ion–atom exchange interaction potential is simpler in the presence of relativistic interactions because of a lower symmetry of atomic particles in this case. For the *LS*-coupling scheme for individual atomic particles, we were restricted by the ground states of atomic particles because of a cumbersome problem, but the presence of relativistic effects simplifies this problem. As a demonstration of this, Table 5 contains the matrix of the exchange interaction potential for elements of group 5.

Table 5. The exchange interaction potential for atoms of group 5 of the periodical system of elements whose atomic electron shell is p^3 and their ions have the electron shell p^2

LS		$^4S_{3/2}$	$^2D_{3/2}$	$^2D_{5/2}$	$^2P_{1/2}$	$^2P_{3/2}$
	jj	$\left[\left(\frac{1}{2}\right)^2\left(\frac{3}{2}\right)\right]_{3/2}$	$\left[\left(\frac{1}{2}\right)\left(\frac{3}{2}\right)^2\right]_{3/2}$	$\left[\left(\frac{1}{2}\right)\left(\frac{3}{2}\right)^2\right]_{5/2}$	$\left[\left(\frac{1}{2}\right)\left(\frac{3}{2}\right)^2\right]_{1/2}$	$\left[\left(\frac{3}{2}\right)^3\right]_{3/2}$
$^3P_0$	$\left[\left(\frac{1}{2}\right)^2\right]_0$	$\Delta_{3/2}(+)$	$0(+)$	$0(+)$	$0(+)$	$0(+)$
$^3P_1$	$\left[\left(\frac{1}{2}\right)\left(\frac{3}{2}\right)\right]_1$	$\Delta_{1/2}(+)$	$\Delta_{3/2}(+)$	$\Delta_{3/2}(+)$	$\Delta_{3/2}(+)$	$0(+)$
$^3P_2$	$\left[\left(\frac{1}{2}\right)\left(\frac{3}{2}\right)\right]_2$	$\Delta_{1/2}(+)$	$\Delta_{3/2}(+)$	$\Delta_{3/2}(+)$	$\Delta_{3/2}(+)$	$0(+)$
$^1D_2$	$\left[\left(\frac{3}{2}\right)^2\right]_2$	$0(0)$	$\Delta_{1/2}(+)$	$\Delta_{1/2}(+)$	$\Delta_{1/2}(+)$	$\Delta_{3/2}(+)$
$^1S_0$	$\left[\left(\frac{3}{2}\right)^2\right]_0$	$0(0)$	$\Delta_{1/2}(0)$	$\Delta_{1/2}(0)$	$\Delta_{1/2}(+)$	$\Delta_{3/2}(+)$

The notation for the ion and atom electron terms used in Table 5 applies to the *LS*- and *jj*-coupling schemes. The values of the exchange interaction potentials are given assuming that the *jj* momentum coupling scheme applies, and it is indicated in parentheses whether this potential is zero (0) or nonzero (+) for the *LS*-coupling scheme. In particular, for the ground atom and ion states, the exchange interaction potential occupies one cell in Table 5, while within the framework of the *LS*-coupling scheme, it is given by matrix (9c).

We note that for the *jj*-coupling scheme, the *p*-electron shell of an atom or an ion is separated into two independent subshells with $j = 1/2$ and $j = 3/2$. Hence, the difference between the numbers of electrons in these subshells for the interacting ion and the atom cannot exceed one. This is the criterion of the one-electron transition replacing Eq. (12) for the *LS*-coupling scheme. If this criterion is not satisfied, the one-electron ion–atom exchange interaction potential is zero; otherwise, it is equal to $\Delta_{1/2}$ or $\Delta_{3/2}$ depending on the momentum of the transiting electron (see Tables 4 and 5).

We now focus on elements of groups 3 or 8 of the periodical table, where one transiting *p*-electron (or *p*-hole) is located in the field of two structureless cores. If the spin–orbit splitting of the electron levels is large compared to the electrostatic ion–atom interaction, the quantum numbers of the molecular ion are jm_j, the total electron momentum and its projection on the molecular axis. The exchange interaction potential Δ_{jm_j} pertaining to the *jj*-coupling scheme for atoms and ions and the exchange interaction potentials Δ_{1m} pertaining to the *LS*-coupling schemes are related by

$$\Delta_{jm_j} = \sum_{\mu} \begin{bmatrix} \frac{1}{2} & 1 & j \\ \sigma & \mu & m_j \end{bmatrix}^2 \Delta_{1\mu}.$$

This follows from the relation between the electron wave functions for the respective states. For the exchange interaction potentials Δ_{jm_j}, where $m_j = \sigma + \mu$ in accordance with the properties of the Clebsch–Gordan coefficients, this gives

$$\begin{aligned} \Delta_{1/2,\,1/2} &= \frac{1}{3}\Delta_{10} + \frac{2}{3}\Delta_{11}, \\ \Delta_{3/2,\,1/2} &= \frac{2}{3}\Delta_{10} + \frac{1}{3}\Delta_{11}, \\ \Delta_{3/2,\,3/2} &= \Delta_{11}, \end{aligned} \tag{20}$$

where Δ_{10} and Δ_{11} are given by Eqs. (7) and (8).

By analogy with the previous operations, if the molecular axis is at the angle θ to the quantization axis on which the angular momentum projection is zero, the exchange interaction potentials are given by

$$\begin{aligned} \Delta_{1/2} &= \frac{1}{3}\Delta_{10} + \frac{2}{3}\Delta_{11}, \\ \Delta_{3/2}(\theta) &= \left(\frac{1}{6} + \frac{1}{2}\cos^2\theta\right)\Delta_{10} \\ &\quad + \left(\frac{1}{3} + \frac{1}{2}\sin^2\theta\right)\Delta_{11}. \end{aligned} \tag{21}$$

This corresponds to elements of groups 3 and 8 of the periodical system and gives the exchange phases for case "*c*" of the Hund coupling,

$$\begin{aligned} \frac{\zeta_{1/2}(\rho, \vartheta)}{\zeta_0(\rho)} &= 1 + \frac{4}{\rho\gamma}, \quad \frac{\zeta_{3/2}(\rho, \vartheta)}{\zeta_0(\rho)} = \frac{1}{2} + \frac{3}{2}\cos^2\vartheta \\ &\quad + \frac{1}{\rho\gamma}\left(\frac{1}{2} + \frac{9}{2}\sin^2\vartheta + \frac{3}{2}\sin^2\vartheta\cos^2\varphi\right), \end{aligned} \tag{22}$$

where $\zeta_0(\rho)$ is the charge exchange phase for the transiting *s*-electron with the same asymptotic parameters *A* and γ as defined in accordance with Eq. (4) and ϑ, φ are the polar angles of the impact parameter direction with respect to the quantization axis. Table 2 contains the reduced cross sections $\Sigma_j = \overline{\sigma}_j/\sigma_s$, where the average cross section $\overline{\sigma}_j$ for a given total momentum is determined by formula (13). As can be seen, the difference of the average cross sections for different total momenta is small compared to the accuracy of determining the cross sections, and we neglect this difference. One can determine the cross sections for case "*e*" of the Hund coupling, where because of a large rotational energy, the momentum projection on the motionless axis is conserved for the state with $j = 3/2$. In Table 2, $\Sigma^e_{3/2}$ is the reduced cross section of the resonant charge exchange for the state with $j = 3/2$ in case "*e*" of the Hund coupling. According to the data in Table 2, the connection between the molecular and motionless axis is not essential for the cross section of this process.

Thus, one can see that the molecular axis rotation gives a small contribution to the resonant charge exchange cross section. That is, the difference between cases "*a*", "*b*", and "*d*" of the Hund coupling, as well as between cases "*c*" and "*e*", is not essential for this process. Next, according to the data in Table 2, the difference between the cross sections for cases "*a*" and "*c*" of the Hund coupling is not significant for atoms of groups 3 and 8 of the periodical system of elements, and it is essential for atoms of groups 4, 5, 6, and 7. Thus, the transition between these coupling cases results from the competition between the splitting $U(R)$ due to a long-range ion–atom interaction, the splitting $\Delta(R)$ due to the exchange interaction, and the fine level splitting δ_f. Tables 6 and 7 contain these values for atoms of groups 3 and 8 of the periodical system of ele-

Table 6. Parameters of the resonant charge exchange process for collisions of atoms of group 3 of the periodical system with their ions at the energy 1 eV in the laboratory coordinate system

	B	Al	Ga	In	Tl
$\bar{\sigma}$, 10^{-14} cm^2	1.1	1.8	2.0	2.2	2.1
γR_0	12	14	14	15	14
$\Delta\sigma/\bar{\sigma}$, %	0.7	0.5	0.4	0.3	0.4
δ_f, cm^{-1}	15	112	826	2213	7793
$U(R_0)$, cm^{-1}	360	350	320	330	390
$\Delta(R_0)$, cm^{-1}	11	5	3	2.5	2

Table 7. Parameters of the resonant charge exchange process for collisions of atoms of group 8 of the periodical system with their ions at the energy 1 eV in the laboratory coordinate system

	Ne	Ar	Kr	Xe
$\bar{\sigma}$, 10^{-15} cm^2	3.3	5.8	7.5	10
γR_0	11	12	13	14
$\Delta\sigma/\bar{\sigma}$, %	0.8	0.5	0.4	0.3
δ_f, cm^{-1}	780	1432	5370	10537
$U(R_0)$, 10^{-3} cm^{-1}	5	4	2	2
$\Delta(R_0)$, cm^{-1}	13	8	5	3

ments. Next, the long-range splitting of levels depends on the atom and ion states. If atoms and ions are found in the ground states, the long-range splitting $U(R)$ of atomic levels for elements of groups 3, 4, 6, and 7 results from the interaction of the ion charge with the atom quadrupole moment and is given by

$$U(R) = \frac{5\langle r^2\rangle}{6R^3}, \tag{23a}$$

where R is the distance between the interacting particles and $\langle r^2\rangle$ is the mean square of the electron orbit radius of the valent electron in the atom. The long-range splitting of ion levels for elements of groups 4 and 7, where the quadrupole momenta of the atom and the ion is not zero, is determined by the interaction of the quadrupole momenta, and the long-range ion–atom interaction potential $V(R)$ is then given by

$$V(R) = \frac{Q_iQ_a}{R^5}, \tag{23b}$$

where Q_a and Q_i are the respective quadrupole momenta of the atom and the ion, which are $\pm 2\langle r^2\rangle/5$ for states with zero orbital momentum projection and $\mp 4\langle r^2\rangle/5$ for states for which the orbital momentum projection on the motionless axis is 1. Expression (23b) relates to elements of groups 4 and 7 of the periodical system, where the quadrupole moment of atoms and ions is not zero. Next, the splitting of ion levels for elements of groups 5 and 8, whose atoms have zero quadrupole moment, is given by

$$U(R) = \frac{12\beta(\langle r^2\rangle)^2}{25R^8}, \tag{23c}$$

where β is the atom polarizability. The value $\Delta\sigma/\bar{\sigma}$ in Tables 6 and 7 characterizes the error in the cross section arising from using only the exponential dependence of the exchange phase $\zeta(\rho) \propto \exp(-\gamma\rho)$, as we did in Table 2.

The data in Tables 6 and 7 show the role of different interactions for the resonant charge exchange process involving real ions and atoms. In particular, it follows from these tables that the long-range splitting of molecular terms is important for elements of group 3 and is negligibly small compared to the exchange interaction potential for molecular ions of rare gases. In addition, in Table 8 we give the average cross sections of the resonant charge exchange processes for elements with valent p-electrons. We note that in accordance with Eqs. (3) and (4), the cross section σ of this process depends on the collision velocity v as [1, 5]

$$\sigma = \frac{\pi}{2\gamma^2}\ln^2\frac{C}{v}, \tag{24}$$

Table 8. The parameters of the cross section of the resonant charge exchange for elements with valent p-electrons of atoms and ions at the collision energy 1 eV

Element	B	C	N	O	F	Ne	Al	Si	P
σ, 10^{-15} cm^2	11	8.6	6.2	6.6	4.9	3.3	18	15	11
$\alpha = -d\ln\sigma/d\ln v$	0.16	0.16	0.16	0.16	0.17	0.18	0.15	0.14	0.14
Element	S	Cl	Ar	Ga	Ge	As	Se	Br	Kr
σ, 10^{-15} cm^2	10	8.0	5.8	20	18	13	13	10	7.5
$\alpha = -d\ln\sigma/d\ln v$	0.15	0.15	0.16	0.14	0.13	0.14	0.14	0.13	0.15
Element	In	Sn	Sb	Te	I	Xe	Tl	Pb	Bi
σ, 10^{-15} cm^2	22	19	17	16	13	10	21	20	22
$\alpha = -d\ln\sigma/d\ln v$	0.14	0.13	0.13	0.13	0.13	0.14	0.14	0.13	0.12

where C is a constant. The first order of the asymptotic theory allows us to represent this relation as

$$-\frac{d\ln\sigma}{d\ln v} = \alpha = \frac{1}{2R_0\gamma}, \text{ or } \frac{\sigma(v)}{\sigma(v_0)} = \left(\frac{v_0}{v}\right)^{\alpha}. \quad (25)$$

Table 8 gives the parameters of this formula at the collision energy 1 eV in the laboratory coordinate system.

4. CONCLUSION

We have developed the asymptotic theory for the resonant charge exchange process in slow collisions of an ion and a parent atom with the transiting *p*-electron. The cross section of this process is not sensitive to the relation between the rotational and other interactions of the colliding particles and inside them, but can depend on the spin–orbit interaction. We have two limiting cases that correspond to cases "*a*", "*b*", and "*d*" and cases "*c*" and "*e*" of the Hund coupling, or to the *LS*- and *jj*-coupling schemes for isolated colliding atomic particles. For elements of groups 3 and 8 of the periodical system, the average cross sections of the resonant charge exchange are nearly equal for these limiting cases of the momentum coupling, while for other groups the difference between the cross sections for different coupling schemes exceeds the accuracy of the evaluation of these cross sections. The accuracy of the asymptotic theory of the resonant charge exchange with the transiting *p*-electron is worse than that in the case of the transiting *s*-electron (1–5% [9]) and is estimated as ~10% at eV energies.

ACKNOWLEDGMENTS

This study was partly supported by the Russian Foundation for Basic Research (project no. 00-02-17090).

REFERENCES

1. L. A. Sena, Zh. Éksp. Teor. Fiz. **9**, 1320 (1939).
2. L. A. Sena, Zh. Éksp. Teor. Fiz. **16**, 734 (1946).
3. L. A. Sena, *Collisions of Electrons and Ions with Atoms* (Gostekhizdat, Leningrad, 1948).
4. O. B. Firsov, Zh. Éksp. Teor. Fiz. **21**, 1001 (1951).
5. Yu. N. Demkov, Uch. Zap. Leningr. Gos. Univ. im. A. A. Zhdanova, Ser. Phys. Nauk, No. 146, 74 (1952).
6. B. M. Smirnov, Zh. Éksp. Teor. Fiz. **46**, 1017 (1964) [Sov. Phys. JETP **19**, 692 (1964)].
7. B. M. Smirnov, Zh. Éksp. Teor. Fiz. **47**, 518 (1965) [Sov. Phys. JETP **20**, 345 (1965)].
8. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1989; Pergamon, New York, 1980).
9. B. M. Smirnov, Phys. Scr. **61**, 595 (2000).
10. B. M. Smirnov, *Asymptotic Methods in Theory of Atomic Collisions* (Atomizdat, Moscow, 1973).
11. E. L. Duman *et al.*, Preprint No. 3532/12 (Kurchatov Institute of Atomic Energy, Moscow, 1982).
12. S. Sakabe and Y. Izawa, At. Data Nucl. Data Tables **49**, 257 (1991).
13. S. Sakabe and Y. Izawa, Phys. Rev. A **45**, 2086 (1992).
14. F. B. M. Copeland and D. S. F. Crothers, At. Data Nucl. Data Tables **72**, 57 (1999).
15. G. Racah, Phys. Rev. **61**, 186 (1942); **62**, 438 (1942).
16. I. I. Sobelman, *Atomic Spectra and Radiative Transitions* (Nauka, Moscow, 1977; Springer-Verlag, Berlin, 1979).
17. E. L. Duman and B. M. Smirnov, Zh. Tekh. Fiz. **40** (1), 91 (1970) [Sov. Phys. Tech. Phys. **15**, 61 (1970)].
18. R. S. Mulliken, Rev. Mod. Phys. **2**, 60 (1930).
19. F. Hund, Z. Phys. **36**, 637 (1936).
20. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Scientific, Singapore, 1988).
21. E. E. Nikitin, Opt. Spectrosc. **22**, 379 (1966).
22. E. E. Nikitin and B. M. Smirnov, Usp. Fiz. Nauk **124** (2), 201 (1978) [Sov. Phys. Usp. **21**, 95 (1978)].
23. E. E. Nikitin and S. Ya. Umanskii, *Theory of Slow Atomic Collisions* (Atomizdat, Moscow, 1979; Springer-Verlag, Berlin, 1984).

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 960–968.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1109–1117.

ATOMS, SPECTRA, RADIATION

Absorption of Electromagnetic Energy by Slow Electrons under Scattering from Coulomb Centers

V. P. Krainov*

Moscow Institute of Physics and Technology, Dolgoprudnyĭ, Moscow oblast, 141700 Russia
**e-mail: krainov@cyberax.ru*

Received January 3, 2001

Abstract—Simple analytical expressions are obtained for the rate of the inverse stimulated bremsstrahlung absorption under electron scattering from a Coulomb center with charge Z in the presence of the electromagnetic field. The initial and final values of electron energy are assumed to be small compared to the Rydberg energy Z^2 (atomic units are used throughout). Single-photon processes of absorption and induced radiation of photon by electron are treated. It is assumed that the electromagnetic field frequency ω is rather low, so that the condition $Z\omega/p^3 \ll 1$, where p is the electron momentum, and the condition $\hbar\omega \ll p^2$ are valid. However, this frequency is assumed to be fairly high compared to the electron–Coulomb center collision frequency: $\omega \gg \nu_{ei}$. The dependences of the rates of photon absorption and induced radiation on the angle θ between the direction of incident electron and the electromagnetic field polarization vector (assumed to be linearly polarized) are obtained. It is demonstrated that, for any angles θ, the rate of photon absorption is higher than the rate of induced radiation and, therefore, the Marcuse effect for slow electrons (electromagnetic field amplification) is absent. It is further demonstrated that a slow electron on the average absorbs double ponderomotive energy per collision with an ion (Coulomb center) in Maxwellian plasma. This agrees both with the known results calculation for fast electrons and with the known results of the calculation based on the classical Boltzmann kinetic equation for plasma. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well known that a free electron cannot really absorb or radiate photons of an external monochromatic electromagnetic field. However, in the presence of a Coulomb center of potential Z/r, such processes are possible (here and throughout, we use, as a rule, the atomic system of units, $e = m_e = \hbar = 1$, and take into account the Planck constant only in Section 2). We assume that the electromagnetic field is rather weak, so that only single-photon processes of photon absorption and induced radiation occur (the first-order perturbation theory with respect to external electromagnetic field). The respective upper constraint on the field strength will be given below. The rate of the photon absorption and induced emission was calculated by Marcuse [1] for fast electrons using the perturbation theory with respect to the Coulomb potential as well. We will write p for the initial electron momentum (prior to scattering) and p' for the final electron momentum (after scattering), the condition of validity of the Born approximation for the Coulomb potential has the form $p \gg Z$, $p' \gg Z$. A detailed description of the corresponding results obtained by Marcuse can be found in monograph [2]. It has been found that, in a certain range of angles θ between the initial electron momentum **p** and the vector of magnetic field strength **E** (in all instances, we will assume the electromagnetic field to be linearly polarized), the rate of photon absorption, w_a, is less than the rate of induced photon radiation, w_e, so that the electromagnetic radiation increases (this is precisely the so-called Marcuse effect). However, in the case of integration with respect to all angles assuming a uniform angular distribution of electrons, this effect disappears and, conversely, an electron extracts the electromagnetic energy from radiation; i.e., the radiation is absorbed. If the rate of absorption of electromagnetic energy is averaged over the electron velocities assuming their Maxwellian distribution, the average energy absorbed by the electron in a single collision with a Coulomb center is $E^2/2\omega^2$, where ω is the electromagnetic field frequency; i.e., the average energy is equal to the double average ponderomotive electron energy $U_p = E^2/4\omega^2$. This value coincides exactly with that yielded by the Boltzmann equation for elastic scattering of classical electrons by multiply charged ions in plasma [3, 4] when treating the high-frequency conductivity of plasma (here, the high frequency implies the validity of the inequality $\omega \gg \nu_{ei}$, where ν_{ei} is the electron-ion collision frequency).

This paper deals with the investigation of the opposite limiting case of slow electrons, $p \ll Z$, $p' \ll Z$, when the quasiclassical approximation is valid. Of course, the photon energy is always assumed to be small compared to the electron energy, i.e., $\hbar\omega \ll p^2$. In addition to this criterion, however, there is the classical dimensionless parameter $\beta = Z\omega/p^3$, which does not contain

the Planck constant. In [5], we treated the limiting case of high frequencies, namely, $Z\omega/p^3 \gg 1$ (which does not contradict the above criterion $\hbar\omega \ll p^2$, because $p \ll Z$). In so doing, the scattering of electrons through large angles is significant.

In practice, however, the inverse limiting case of low radiation frequencies is realized, $Z\omega/p^3 \ll 1$, if one has in mind the frequencies of typical laser light sources and the electron temperatures of multiply charged laser cluster plasma of tens of electron-volts and higher [6]. Therefore, we treat the problem for slow electrons in just the above-identified limiting low-frequency case of $Z\omega/p^3 \ll 1$. The electron scattering through small angles in the process of photon absorption or emission is important.

We will first turn to the solution of this problem within the classical field theory (Section 2). In textbooks on the classical field theory (see, for example, [7]), one can find results only for the electromagnetic energy loss averaged over the angle θ. Therefore, it is of interest to derive simple analytical expressions for energy loss with a fixed value of this angle. Of course, within this classical approach, the rates of induced emission and absorption of a photon are equal, because the Planck constant is zero within the classical field theory.

One could calculate the work of an electron in an external field directly as well and, thereby, determine directly the difference between the rates of photon emission and absorption. For this purpose, however, one must solve a linear differential equation for small perturbation of the coordinate of an electron, scattered from a Coulomb center, by a variable electromagnetic field, which is a fairly complicated problem. Apparently, it would be much simpler to turn to the known results of quantum electrodynamics [8] for spontaneous bremsstrahlung under electron scattering from a Coulomb center, use the well-known rules to change over to induced radiation [2], and perform the necessary simplifications of the results in the quasiclassical limit. In Section 3, this is done for total rates of photon absorption and emission, averaged over the angle θ, and in Section 4, for rates with a fixed value of this angle between the direction of incident electron and the polarization vector of electromagnetic field. At the same time, the known results of Marcuse for fast electrons are given in all instances for comparison.

In conclusion, the results obtained are compared with the known results for the average increase of the electron energy in Maxwellian plasma during a single collision with an ion in the presence of external electromagnetic field. For both fast and slow electrons, this increase in energy is equal to the double average ponderomotive energy, which is in good numerical agreement with the results of the calculation based on the Boltzmann kinetic equation for classical electrons [3, 4].

2. CLASSICAL APPROACH

Within the classical field theory (see formula (67.10) in [7]), we will start from the spectral radiation density. We divide this quantity by $\hbar\omega$ to derive the number of photons of frequency ω emitted spontaneously by an electron to the solid angle $d\Omega$,

$$dN_e = \frac{\omega^3}{2\pi^2 c^3 \hbar}|\mathbf{e}\cdot\mathbf{r}_\omega|^2 d\Omega d\omega. \quad (1)$$

Here, the quantity $\mathbf{r}_\omega$ is the Fourier component of the classical radius vector of an electron during its motion in the field of the attracting Coulomb center of charge Z, $\mathbf{e}$ is the polarization vector of the emitted spontaneous photon, and c is the velocity of light.

Upon transition to induced processes of photon emission or absorption, one must perform the following substitution in the foregoing expression (for more detail, see monograph [2]):

$$d\Omega d\omega \longrightarrow \frac{\pi^2 c^3 E^2}{\hbar\omega^3},$$

where E is the amplitude of strength of the external variable electric field and ω is the field frequency. Then, we derive a simple expression for the number of photons subjected to induced emission or absorption,

$$N_{e,a} = \frac{1}{4\hbar^2}|\mathbf{E}\cdot\mathbf{r}_\omega|^2.$$

We will assume that the electron is scattered in the xy plane. Then, the Fourier component of its radius vector is represented as the expansion in unit vectors in this plane,

$$\mathbf{r}_\omega = x_\omega \mathbf{i}_x + y_\omega \mathbf{i}_y.$$

The expressions for the Fourier components of projections of the electron radius vector are well known [7],

$$x_\omega = \frac{\pi a}{\omega} H_{i\beta}^{(1)'}(i\beta\varepsilon),$$

$$y_\omega = -\frac{\pi a}{\omega}\sqrt{1-\frac{1}{\varepsilon^2}}\,H_{i\beta}^{(1)}(i\beta\varepsilon).$$

Here, the notation

$$\beta = \frac{Z\omega}{p^3} \ll 1, \quad a = \frac{Z}{p^2}, \quad \varepsilon = \sqrt{1+\left(\frac{\rho}{a}\right)^2}$$

is introduced.

The quantity ε is the eccentricity of the hyperbolic path followed by an unperturbed electron in the field of the Coulomb center. The function $H_{i\beta}^{(1)}(x)$ is the cylindrical Hankel function, and ρ is the impact parameter.

Unit vectors along the x, y axes can be represented as

$$\mathbf{i}_x = \frac{\mathbf{p}-\mathbf{p}'}{2p\sin(\vartheta/2)}, \quad \mathbf{i}_y = \frac{\mathbf{p}+\mathbf{p}'}{2p\cos(\vartheta/2)},$$

where ϑ is the electron scattering angle, i.e., the angle between the initial electron momentum $\mathbf{p}$ and its final momentum $\mathbf{p}'$. Then, we derive

$$\mathbf{r}_\omega \cdot \mathbf{E} = \frac{E}{2}\left[x_\omega \frac{\cos\theta' - \cos\theta}{\sin(\vartheta/2)} - y_\omega \frac{\cos\theta' + \cos\theta}{\cos(\vartheta/2)}\right].$$

Here, θ' is the angle between the final electron momentum $\mathbf{p}'$ and the vector of electric field strength $\mathbf{E}$. We use the formulas of spherical trigonometry, which relate the angles θ, θ', and ϑ with one another, to derive

$$\mathbf{r}_\omega \cdot \mathbf{E} = E\left[x_\omega\left(\cos\theta\sin\frac{\vartheta}{2} - \sin\theta\cos\frac{\vartheta}{2}\cos\varphi\right) - y_\omega\left(\cos\theta\cos\frac{\vartheta}{2} - \sin\theta\sin\frac{\vartheta}{2}\cos\varphi\right)\right]. \quad (2)$$

Here, φ is the angle between the projection of the vector $\mathbf{p}'$ on the plane perpendicular to the vector $\mathbf{p}$ and the line of intersection of this plane with the plane of the vectors $\mathbf{p}$ and $\mathbf{E}$.

The averaging over the angle φ is elementary, and Eqs. (1) and (2) yield

$$N_{e,a}(\theta,\vartheta) = \frac{E^2}{4\hbar^2}|x_\omega|^2 \times\left(\cos^2\theta\sin^2\frac{\vartheta}{2} + \frac{1}{2}\sin^2\theta\cos^2\frac{\vartheta}{2}\right) + \frac{E^2}{4\hbar^2}|y_\omega|^2\left(\cos^2\theta\cos^2\frac{\vartheta}{2} + \frac{1}{2}\sin^2\theta\sin^2\frac{\vartheta}{2}\right). \quad (3)$$

The angle ϑ of Coulomb scattering of electron is related to the impact parameter ρ and eccentricity ε by the known simple relations

$$\cot\frac{\vartheta}{2} = \frac{\rho}{a}, \quad \sin\frac{\vartheta}{2} = \frac{1}{\varepsilon}.$$

This enables one to reduce the averaging of rate given by Eq. (3) over the angle ϑ to integration with respect to the impact parameter ρ; for this purpose, one must multiply dimensionless expression (3) by $2\pi n_i p\rho d\rho$ and integrate it with respect to all impact parameters. Here, n_i is the spatial concentration of Coulomb centers, from which the given electron is scattered. From integration with respect to ρ we will then turn to integration with respect to ε, using the relation $\rho d\rho = a^2\varepsilon d\varepsilon$.

As a result, we derive the number of photons subjected to induced emission (absorption) per unit time,

$$N_{e,a}(\theta) = \frac{\pi Z^2 n_i E^2}{2p^3\hbar^2} \times\int_1^\infty \varepsilon d\varepsilon\left\{|x_\omega|^2\left[\frac{1}{\varepsilon^2}\cos^2\theta + \frac{1}{2}\left(1-\frac{1}{\varepsilon^2}\right)\sin^2\theta\right]\cdot + |y_\omega|^2\left[\frac{1}{2\varepsilon^2}\sin^2\theta + \left(1-\frac{1}{\varepsilon^2}\right)\cos^2\theta\right]\right\}. \quad (4)$$

The above-identified Fourier components are simplified in the $\beta \ll 1$ limit,

$$x_\omega = \frac{2a}{\omega}K_1(\beta\varepsilon), \quad y_\omega = \frac{2ia}{\omega}\sqrt{1-\frac{1}{\varepsilon^2}}K_0(\beta\varepsilon).$$

Here, $K_0(x)$ and $K_1(x)$ are Macdonald functions.

We substitute these quantities into Eq. (4) to calculate the arising integrals. In so doing, one must ignore a part of terms in Eq. (4) containing smallness of the order of $\beta \ll 1$. As a result, we derive

$$N_{e,a}(\theta) = \frac{2\pi Z^2 n_i E^2}{p\omega^4\hbar^2} \times\left\{\cos^2\theta\int_\beta^\infty tdtK_0^2(t) + \frac{1}{2}\int_\beta^\infty tdtK_1^2(t) + \left(\cos^2\theta - \frac{1}{2}\sin^2\theta\right)\beta^2\int_\beta^\infty dtK_1^2(t)\frac{1}{t}\right\}.$$

We calculate dimensionless integrals to a desired accuracy with respect to β,

$$\int_\beta^\infty tdtK_0^2(t) = \frac{1}{2}, \quad \int_\beta^\infty tdtK_1^2(t) = \ln\frac{1}{\beta\gamma} - \frac{1}{2},$$

$$\int_\beta^\infty \frac{1}{t}dtK_0^2(t) = \frac{1}{2\beta^2}.$$

Here, $\gamma = 1.781\ldots = \exp C$ and $C = 0.577\ldots$ is Euler's constant. We derive

$$N_{e,a}(\theta) = \frac{2\pi Z^2 n_i E^2}{p\omega^4\hbar^2} \times\left[\cos^2\theta + \frac{1}{2}\left(\ln\frac{2p^3}{\gamma Z\omega} - 1\right)\sin^2\theta\right]. \quad (5)$$

This result is new and is the key result in this section. It gives the angular distribution for photon absorption or induced emission by an electron under scattering from a Coulomb center in the presence of electromagnetic field. Because the logarithm argument is great, the process of photon absorption or emission takes place mainly when the electron moves normally to the field polarization ($\theta = \pi/2$).

The absorption or induced emission taken singly may show up indirectly, for example, in the Boltzmann kinetic equation when treating radiation processes in plasma. Of most interest is, of course, their difference proportional to the electromagnetic wave absorptivity being measured.

We average expression (5) over the angle θ to derive

$$\bar{N}_{e,a} = \frac{2\pi Z^2 n_i E^2}{3p\omega^4\hbar^2}\ln\frac{2p^3}{\gamma Z\omega}. \quad (6)$$

On returning from induced to spontaneous radiation, the substitution [2]

$$E^2 \longrightarrow \frac{8\omega^3\hbar d\omega}{\pi c^3}$$

must be performed; we find from Eq. (6)

$$\bar{N}_{e,a} = d\omega\frac{16Z^2 n_i}{3c^3\hbar p\omega}\ln\frac{2p^3}{\gamma Z\omega}.$$

The energy radiated by an electron per unit time (it does not contain the Planck constant) is obtained from the foregoing formula as a result of multiplication by $\hbar\omega$,

$$dE_\omega = d\omega\frac{16Z^2 n_i}{3c^3 p}\ln\frac{2p^3}{\gamma Z\omega}.$$

We divide this expression by $n_i p$ to derive exactly expression (70.21) from [7] for effective radiation of low frequencies, as is to be expected.

3. QUANTUM APPROACH: TOTAL RATES

In the quantum approach, the rates of photon absorption and induced emission are different from each other. Treated in this section is a simpler problem of calculating the total rate integrated with respect to the angle θ between the initial direction of electron motion and the polarization vector of electromagnetic radiation, assuming that $\beta = Z\omega/p^3 \ll 1$. The cross section of spontaneous photon radiation by an electron with initial momentum p and final momentum p' under scattering from a Coulomb center with charge Z is given by the known relation of quantum electrodynamics [8] (in what follows, we assume the Planck constant to be equal to unity),

$$d\sigma_\omega = \frac{64\pi^2 Z^2}{3c^3}\frac{p}{p(p-p')^2}$$
$$\times\frac{d|F(-x)|^2/dx}{[1-\exp(-2\pi\nu')][\exp(2\pi\nu)-1]}\frac{d\omega}{\omega}.$$

Here, we used the notation

$$\nu = \frac{Z}{p} \gg 1, \quad \nu' = \frac{Z}{p'} \gg 1$$

and introduced a reduced notation for the complete hypergeometric function $F(-x) = F(i\nu, i\nu', 1; -x)$ and its argument

$$x = \frac{4pp'}{(p-p')^2} \gg 1.$$

The transition to induced emission is achieved by the substitution [2]

$$d\omega \longrightarrow \frac{\pi c^3 E^2}{8\omega^3}.$$

Then, we derive the following compact expression for the rate of induced photon emission averaged over all angles:

$$\bar{w}_e = n_i p d\sigma_\omega = \frac{8\pi^3 Z^2 n_i E^2 p'\exp(-2\pi\nu)}{3(p-p')^2\omega^4} \times\frac{d|F(-x)|^2}{dx}. \quad (7)$$

The rate of photon absorption is given by a similar formula in which one needs only to replace the factor $\exp(-2\pi\nu)$ by $\exp(-2\pi\nu')$ (for detailed explanation of this difference, see the discussion of formula (92.8) in [8]).

The hypergeometric function entering expression (7) may be simplified, under condition $x \gg 1$, by transition to a combination of hypergeometric functions with the argument $1/x \ll 1$,

$$F(i\nu, i\nu', 1; -x) \approx \frac{\Gamma(i(\nu'-\nu))}{\Gamma(i\nu')\Gamma(1-i\nu)}x^{-i\nu} + \frac{\Gamma(i(\nu-\nu'))}{\Gamma(i\nu)\Gamma(1-i\nu')}x^{-i\nu'}.$$

Here, $\Gamma(z)$ is the gamma function. Use is also made of the condition $\nu' - \nu \ll 1$ equivalent to the condition of low field frequency $Z\omega/p^3 \ll 1$.

The hypergeometric function is further simplified using Stirling's formula for gamma functions of large arguments $\nu, \nu' \gg 1$. We derive

$$F(i\nu, i\nu', 1; -x) \approx \frac{i}{\pi}x^{i\bar{\nu}}\exp(\pi\bar{\nu})\ln\frac{\sqrt{x}}{\gamma\bar{\nu}}. \quad (8)$$

Here, the notation $\bar{\nu} = (\nu + \nu')/2$ is introduced. As is to be expected, asymptotic representation (8) of the hypergeometric function is symmetric relative to the permutation of its first two indices $\nu \longleftrightarrow \nu'$.

The corrections in Eq. (8) when calculating the rate given by Eq. (7) have a relative smallness $1/\nu^2 \ll 1$ or $(Z\omega/p^3)^2 \ll 1$. Therefore, these corrections do not contain terms linear with respect to the emitted photon frequency ω. This is very important because, when the rate of photon absorption is subtracted from the rate of its induced emission, the main parts of the rates cancel out, and it is the small parts, which are linear with respect to ω, that remain.

We substitute Eq. (8) into (7) to find for the rate of induced photon emission

$$\bar{w}_e = \frac{2\pi Z^2 n_i E^2 \exp(\pi\nu' - \pi\nu)}{3p\omega^4} \ln\frac{2(pp')^{3/2}}{\gamma Z\omega}. \quad (9)$$

Here, $p' \approx p - \omega/p$. We expand Eq. (9) in terms of the small quantity $Z\omega/p^3 \ll 1$ to derive

$$\bar{w}_e = \frac{2\pi Z^2 n_i E^2}{3p\omega^4}\left[\left(1 + \frac{\pi Z\omega}{p^3}\right)\ln\frac{2p^3}{\gamma Z\omega} - \frac{3\omega}{2p^2}\right]. \quad (10)$$

We similarly calculate the rate of photon absorption. Instead of Eq. (9), we derive an expression that differs only by the form of the exponent,

$$\bar{w}_a = \frac{2\pi Z^2 n_i E^2 \exp(\pi\nu - \pi\nu')}{3p\omega^4} \ln\frac{2(pp')^{3/2}}{\gamma Z\omega}. \quad (11)$$

Now $p' \approx p + \omega/p$. Note that here we corrected the misprint made in the respective expression (56) in [5]. We expand Eq. (11) in terms of the small quantity $Z\omega/p^3 \ll 1$ to derive

$$\bar{w}_a = \frac{2\pi Z^2 n_i E^2}{3p\omega^4}\left[\left(1 + \frac{\pi Z\omega}{p^3}\right)\ln\frac{2p^3}{\gamma Z\omega} + \frac{3\omega}{2p^2}\right]. \quad (12)$$

One can see that, if the additions linear with respect to frequency are ignored, expressions (10) and (12) coincide with classical expression (6), as must be the case.

We subtract Eq. (10) from (12) to find the difference defining the rate of inverse stimulated bremsstrahlung absorption, averaged over all angles,

$$\bar{w}_T = \bar{w}_a - \bar{w}_e = \frac{2\pi Z^2 n_i E^2}{p^3\omega^3}. \quad (13)$$

We multiply this expression by the frequency ω to obtain the energy absorbed by an electron per unit time. On averaging this energy over the Maxwellian distribution at some average electron temperature T, we derive (with logarithmic accuracy) the average energy being absorbed in the form

$$\frac{dU}{dt} = \langle\omega\bar{w}_T\rangle = \frac{2\sqrt{2\pi}Z^2 n_i E^2}{3\omega^2 T^{3/2}} \ln\frac{T^{3/2}}{Z\omega}.$$

Because the average number of electron collisions with Coulomb centers (ions) per unit time is given by the known relation of the kinetic theory of plasma [3, 4]

$$\nu_{ei} = \frac{4\sqrt{2\pi}n_i Z^2}{3T^{3/2}} \ln\Lambda,$$

where $\Lambda = T^{3/2}/Z\omega$ is the Coulomb logarithm (in the case of plasma, the Coulomb logarithm contains the plasma frequency instead of $Z\omega$, see [4]), the previous expression may be transcribed in the form

$$\frac{dU}{dt} = \frac{E^2}{2\omega^2}\nu_{ei},$$

and, therefore, the quantity $E^2/2\omega^2$ is the average energy absorbed by a slow electron in a single collision with a Coulomb center. Turning back to the discussion in the Introduction, one can conclude that the foregoing quantity is exactly equal to the respective energy for a fast electron, as well as to the respective energy derived from the solution of the classical Boltzmann kinetic equation for Maxwellian plasma [3, 4].

This result is valid in the high-frequency limit of $\omega \gg \nu_{ei}$, when the time between two successive collisions of an electron with Coulomb centers is much longer than the period of laser field oscillations [4]. In the opposite low-frequency limit of $\omega \ll \nu_{ei}$, ω in the foregoing expression must be replaced by ν_{ei} (see the derivations in [6]). Then we find that the energy absorption is independent of the electromagnetic field frequency and is defined by the expression

$$\frac{dU}{dt} = \frac{16}{3\pi}\frac{E^2}{\nu_{ei}}.$$

In real cases, for example, in the interaction between ultrashort laser pulses and large atomic clusters, both the high-frequency and low-frequency limits with respect to the collision frequency between free electrons and multiply charged ions inside a cluster may be realized.

4. QUANTUM APPROACH: ANGULAR DISTRIBUTIONS

In this section, which is the key section of the paper, we obtained the rates of photon absorption and of induced emission by an electron scattering from a Coulomb center in the presence of the electromagnetic field with a fixed angle θ between the initial direction of the electron momentum $\mathbf{p}$ and the vector of electric field strength $\mathbf{E}$.

The rate of electron transition from the initial state with momentum $\mathbf{p}$ to the final state with momentum $\mathbf{p}'$ in the first order perturbation theory with respect to the field of electromagnetic wave is given by the Fermi golden rule,

$$w(\mathbf{p} \longrightarrow \mathbf{p}') = \frac{\pi}{2}\int|V(\mathbf{p} \longrightarrow \mathbf{p}')|^2 \times \delta\left(\frac{p'^2}{2} - \frac{p^2}{2} \mp \omega\right) n_i \frac{d\mathbf{p}'}{(2\pi)^3}. \quad (14)$$

Here, the perturbation has the form of dipole interaction between an electron and the electromagnetic field,

$$V = E\mathbf{e}\cdot\mathbf{r},$$

where **e** is the unit vector of field polarization and E is the amplitude of the field strength.

Because the matrix element of the coordinate operator is related to the matrix element of the momentum operator by the known relation of quantum mechanics,

$$\mathbf{r}(\mathbf{p} \longrightarrow \mathbf{p}') = i\hat{\mathbf{p}}(\mathbf{p} \longrightarrow \mathbf{p}')/\omega,$$

the transition rate given by Eq. (14) may be written as

$$w(\mathbf{p} \longrightarrow \mathbf{p}') = \frac{E^2 n_i p'}{16\pi^2\omega^2}\int|\mathbf{e}\cdot\hat{\mathbf{p}}(\mathbf{p} \longrightarrow \mathbf{p}')|^2 d\Omega_{\mathbf{p}'}.$$

The expression for the matrix element of the momentum operator is given in [8]. We derive (in the case of induced emission)

$$|\mathbf{e}\cdot\hat{\mathbf{p}}(\mathbf{p} \longrightarrow \mathbf{p}')| = \frac{8\pi^2\sqrt{\nu\nu'}\exp(-\pi\nu)}{\omega(p-p')^2}$$
$$\times\left|\frac{i\nu p(\mathbf{p}'-\mathbf{p})\cdot\mathbf{e}}{1+x}F(-x) + (p'\mathbf{p}\cdot\mathbf{e} - p\mathbf{p}'\cdot\mathbf{e})F'(-x)\right|.$$

Here, as above,

$$\nu = \frac{Z}{p} \gg 1, \quad \nu' = \frac{Z}{p'} \gg 1.$$

In addition, new notation is introduced,

$$x = 2\frac{pp' - \mathbf{p}\cdot\mathbf{p}'}{(p-p')^2} = \frac{4pp'}{(p-p')^2}\sin^2\frac{\vartheta}{2} > 0.$$

As above, the angle between the vectors **p** and **p**' (scattering angle) is indicated by ϑ. We expand the scalar products of the vectors to derive the rate of induced photon emission (14)

$$w_e(\theta) = \frac{4\pi^2 E^2 n_i Z^2\exp(-2\pi\nu)}{\omega^4 p(p-p')^4}$$
$$\times\int\left|\frac{iZ}{1+x}(p'\cos\theta' - p\cos\theta)F(-x)\right.$$
$$\left. + pp'(\cos\theta - \cos\theta')F'(-x)\right|^2 d\Omega_{\mathbf{p}'}.$$

Here, we introduce the angle θ' between the vectors **p**' and **e**. We express the differential of the solid angle in terms of the differential of the variable x,

$$d\Omega_{\mathbf{p}'} = \frac{(p-p')^2}{2pp'}dx\,d\varphi.$$

On expressing the angle θ' in terms of θ and ϑ using the formulas of spherical trigonometry (as in Section 2) and integrating with respect to the angle φ (the latter is determined as in Section 2), we will derive the rate of induced photon emission in the form

$$w_e(\theta) = \frac{4\pi^3 n_i E^2 Z^2\exp(-2\pi\nu)}{\omega^4 p^2 p'(p-p')^2}$$
$$\times\int_0^a dx\left\{\cos^2\theta\left|\frac{iZ}{x}(p'\cos\vartheta - p)F(-x)\right.\right.$$
$$\left. + 2pp'\sin^2\frac{\vartheta}{2}F'(-x)\right|^2 \qquad (15)$$
$$\left. + \frac{p'^2}{2}\sin^2\theta\sin^2\vartheta\left|\frac{iZ}{x}F(-x) - pF'(-x)\right|^2\right\}$$
$$= A\sin^2\theta + B\cos^2\theta.$$

Here,

$$a = \frac{4pp'}{(p-p')^2}.$$

In the Appendix, a new asymptotic representation is derived for the complete hypergeometric function, which is adequate for the parameters in the problem being treated,

$$F(-x) = F(i\nu, i\nu', 1; -x)$$
$$\approx \frac{i}{\pi}\exp(\pi\bar{\nu})x^{-i\bar{\nu}}K_0\left(\frac{2\bar{\nu}}{\sqrt{x}}\right). \qquad (16)$$

Here, as above, $K_0(z)$ is the Macdonald function with zero index, and

$$\bar{\nu} = \frac{\nu+\nu'}{2}.$$

Representation (16) is valid with relative accuracy $(\nu-\nu')^2 \ll 1$, which enables one to include correctly both the main terms in the rate (15) and the corrections linear with respect to frequency ω, which define the difference between the rates of photon absorption and induced emission.

We will first turn to the term A in Eq. (15). We substitute Eq. (16) into (15) to derive

$$\left|\frac{iZ}{x}F(-x) - pF'(-x)\right|^2 = \frac{p^2\exp(2\pi\bar{\nu})}{(\pi x)^2}$$
$$\times\left[\frac{(\nu'-\nu)^2}{4}K_0^2\left(\frac{2\bar{\nu}}{\sqrt{x}}\right) + \frac{\bar{\nu}^2}{x}K_1^2\left(\frac{2\bar{\nu}}{\sqrt{x}}\right)\right].$$

On substituting this expression into Eq. (15), one can readily make sure that the term $K_0^2(2\bar{\nu}/\sqrt{x})$ makes a small contribution to the rate of induced photon emis-

sion, which proportional to ω^2. We ignore this term to derive

$$A = \frac{\pi E^2 n_i Z^2 \exp(2\pi\bar{\nu} - 2\pi\nu)}{\omega^4 p}\bar{\nu}^2 \int_b^\infty dt\left(t - \frac{b^2}{t}\right)K_1^2(\bar{\nu}t).$$

Here, we use the notation

$$b = \frac{p - p'}{\sqrt{pp'}} \ll 1.$$

The integral is calculated as in Section 2. We derive

$$A = \frac{\pi E^2 n_i Z^2 \exp(2\pi\bar{\nu} - 2\pi\nu)}{\omega^4 p}\left(\ln\frac{2}{\gamma b\bar{\nu}} - 1\right).$$

Let us now turn to the calculation of B in Eq. (15). We substitute Eq. (16) into the first term in Eq. (15) to find

$$\left|\frac{iZ}{x}(p'\cos\vartheta - p)F(-x) + 2pp'\sin^2\frac{\vartheta}{2}F'(-x)\right|^2$$

$$= \left(\frac{pp'}{\pi x}\right)^2 \exp(2\pi\bar{\nu})$$

$$\times\left[(\nu' - \nu)^2 K_0^2\left(\frac{2\bar{\nu}}{\sqrt{x}}\right) + (b\bar{\nu})^2 K_1^2\left(\frac{2\bar{\nu}}{\sqrt{x}}\right)\right].$$

On substituting this expression into Eq. (15), we find

$$B = \frac{\pi E^2 n_i Z^4 \exp(2\pi\bar{\nu} - 2\pi\nu)}{\omega^4 p^2 p'}$$

$$\times\int_b^\infty dt\left[2tK_0^2(\bar{\nu}t) + \left(\frac{\omega}{pp't}\right)^2 K_1^2(\bar{\nu}t)\right].$$

The calculation of the integrals the way it was done in Section 2 (here, both integrals introduce comparable contributions) gives

$$B = \frac{\pi E^2 n_i Z^2 \exp(2\pi\bar{\nu} - 2\pi\nu)}{\omega^4 p}.$$

We substitute the obtained results into Eq. (15) and perform expansion in terms of $\omega/p^2 \ll 1$ with an accuracy within terms linear with respect to frequency to find the rate of induced photon emission,

$$w_e(\theta) = \frac{\pi E^2 n_i Z^2}{\omega^4 p}\left(1 + \frac{\pi Z\omega}{p^3}\right)$$
$$\times\left[2\cos^2\theta + \left(\ln\frac{2p^3}{\gamma Z\omega} - 1 - \frac{3\omega}{2p^2}\right)\sin^2\theta\right]. \quad (17)$$

The rate of photon absorption is found similarly,

$$w_a(\theta) = \frac{\pi E^2 n_i Z^2}{\omega^4 p}\left(1 + \frac{\pi Z\omega}{p^3}\right)$$
$$\times\left[2\cos^2\theta + \left(\ln\frac{2p^3}{\gamma Z\omega} - 1 + \frac{3\omega}{2p^2}\right)\sin^2\theta\right]. \quad (18)$$

On averaging over the angle θ, we obtain from Eqs. (17) and (18) the expressions of the previous section (Eqs. (10) and (12), respectively), as is to be expected.

We subtract Eq. (17) from (18) to find the rate of inverse stimulated bremsstrahlung absorption,

$$w_T = w_a - w_e = \frac{3\pi E^2 n_i Z^2}{\omega^3 p^3}\sin^2\theta. \quad (19)$$

One can see that this quantity is always positive; i.e., the Marcuse effect is not observed for slow electrons. We average Eq. (19) over the angle θ to derive expression (13) from the previous section, as is to be expected.

One can further see from Eq. (19) that the rate of inverse stimulated bremsstrahlung absorption is maximum when the initial direction of electron is normal to that of polarization of electromagnetic field.

The results obtained may be used in analyzing the heating of the electronic component of clusters in the field of high-power laser radiation [6, 9, 10].

In conclusion, note that also of interest is the analogous problem in which two laser fields are involved rather than one, with one of those fields being much more intense than the other. The respective problem of absorption and amplification of a weak electromagnetic field in the kinetic theory of plasma was treated by Chichkov and Uryupin [11].

ACKNOWLEDGMENTS

I am grateful to M.V. Fedorov for valuable advice in performing this study. The study received support from the Russian Foundation for Basic Research (project nos. 99-02-17810 and 01-02-16056).

APPENDIX

We will derive the asymptotic representation (16) for the complete hypergeometric function used in this study.

It is advisable that the hypergeometric function $F(i\nu, i\nu', 1; -x)$ with a great argument $x \gg 1$ should be first expanded in terms of hypergeometric functions with a small argument,

$$F(i\nu, i\nu', 1; -x) \approx \frac{\Gamma(i(\nu' - \nu))}{\Gamma(i\nu')\Gamma(1 - i\nu)}x^{-i\nu}$$

$$\times F\left(i\nu, i\nu, 1 + i(\nu - \nu'); -\frac{1}{x}\right)$$

$$+ \frac{\Gamma[i(\nu - \nu')]}{\Gamma(i\nu)\Gamma(1 - i\nu')} x^{-i\nu'} F\left(i\nu', i\nu', 1 + i(\nu' - \nu); -\frac{1}{x}\right).$$

We introduce into treatment, as in Section 2, the small difference $\Delta\nu = (\nu' - \nu)/2 \ll 1$ and the average value of $\bar{\nu} = (\nu' + \nu)/2$ and transcribe this expression ignoring the terms of the order of $(\Delta\nu)^2 \ll 1$,

$$F(i\nu, i\nu', 1; -x) = \frac{\exp(\pi\bar{\nu})}{4\pi\Delta\nu} x^{-i\bar{\nu}}$$

$$\times \left[\sqrt{\frac{\nu'}{\nu}} \exp\left(-2i\Delta\nu \ln\frac{\gamma\bar{\nu}}{\sqrt{x}}\right)\right.$$

$$\times F\left(i\nu, i\nu, 1 + i(\nu - \nu'); -\frac{1}{x}\right) \qquad \text{(A.1)}$$

$$- \sqrt{\frac{\nu}{\nu'}} \exp\left(2i\Delta\nu \ln\frac{\gamma\bar{\nu}}{\sqrt{x}}\right)$$

$$\left.\times F\left(i\nu', i\nu', 1 + i(\nu' - \nu); -\frac{1}{x}\right)\right].$$

We now treat each of the two new hypergeometric functions,

$$F\left(i\nu, i\nu, 1 + i(\nu - \nu'); -\frac{1}{x}\right)$$

$$= 1 + \frac{\nu^2/x}{1![1 + i(\nu - \nu')]} \qquad \text{(A.2)}$$

$$+ \frac{\nu^2(\nu - i)^2/x^2}{2![1 + i(\nu - \nu')][2 + i(\nu - \nu')]} + \cdots$$

With $\nu \gg 1$, this gives, in the zero approximation, the modified Bessel function,

$$F^{(0)}\left(i\nu, i\nu, 1 + i(\nu - \nu'); -\frac{1}{x}\right) = I_0\left(\frac{2\nu}{\sqrt{x}}\right).$$

A first-order correction with respect to $\nu - \nu' \ll 1$ in the denominator in Eq. (A.2) gives the following combination of modified Bessel functions:

$$F^{(1)}\left(i\nu, i\nu, 1 + i(\nu - \nu'); -\frac{1}{x}\right)$$

$$= i(\nu' - \nu)\left[K_0\left(\frac{2\nu}{\sqrt{x}}\right) + \ln\frac{\gamma\nu}{\sqrt{x}} I_0\left(\frac{2\nu}{\sqrt{x}}\right)\right].$$

Here, as in the main text, the quantity γ is the logarithm of Euler's constant. A correction of the next order with respect to $\nu - \nu' = Z\omega/p^3 \ll 1$ is real and, in combination with the included terms, gives a contribution to (A.1) that is quadratic with respect to frequency ω. As was done in all instances above, we ignore such a correction.

We further include first-order corrections with respect to $1/\nu \ll 1$ in the numerator in Eq. (A.2),

$$F^{(2)}\left(i\nu, i\nu, 1 + i(\nu - \nu'); -\frac{1}{x}\right) = -\frac{i}{x} I_2\left(\frac{2\nu}{\sqrt{x}}\right).$$

We collect all first-order corrections to derive the asymptotic representation for the first hypergeometric function entering Eq. (A.1),

$$F\left(i\nu, i\nu, 1 + i(\nu - \nu'); -\frac{1}{x}\right)$$

$$= I_0\left(\frac{2\nu}{\sqrt{x}}\right) - i\frac{\nu}{x} I_2\left(\frac{2\nu}{\sqrt{x}}\right) \qquad \text{(A.3)}$$

$$+ i(\nu - \nu')\left[K_0\left(\frac{2\nu}{\sqrt{x}}\right) + \ln\frac{\gamma\nu}{\sqrt{x}} I_0\left(\frac{2\nu}{\sqrt{x}}\right)\right].$$

The second hypergeometric function in the right-hand part of (A.1) is derived from (A.3) by permutation of indices $\nu \longleftrightarrow \nu'$.

We substitute Eq. (A.3) and the analogous second hypergeometric function in (A.1) to derive, after expansions in terms of $\Delta\nu \ll 1$,

$$F\left(i\nu, i\nu', 1; -\frac{1}{x}\right) = \frac{\exp(\pi\bar{\nu})}{\pi} x^{-i\bar{\nu}}$$

$$\times \left\{\left(iK_0(z) + \frac{1}{\sqrt{x}}\right)\left[\frac{1}{z} I_0(z) - I_1(z) + \frac{z}{2} I_2(z) \ln\frac{2}{\gamma z}\right]\right\}.$$

Here, $z = 2\bar{\nu}/\sqrt{x}$.

We are interested in the following regions of argument and indices of the hypergeometric function:

$$x \gg 1, \quad \nu, \nu' \gg 1,$$
$$|\nu - \nu'| \ll 1, \quad \sqrt{x} \gg \frac{\nu}{\ln\nu}. \qquad \text{(A.4)}$$

Thus, the quantity $\sqrt{x}$ may be both greater and smaller than ν. Indeed, in calculating the integrals of modified Bessel functions, the values of the variable x of the order of ν^2 are important.

In satisfying these conditions, one can ignore the modified Bessel functions I_0, I_1, and I_2 in the preceding expression (they have smallness of the order of $1/\nu \ll 1$); then, we derive the final asymptotic representation for the complete hypergeometric function,

$$F\left(i\nu, i\nu', 1; -\frac{1}{x}\right) \approx i\frac{\exp(\pi\bar{\nu})}{\pi} x^{-i\bar{\nu}} K_0\left(\frac{2\bar{\nu}}{\sqrt{x}}\right), \qquad \text{(A.5)}$$

used in the main text.

REFERENCES

1. D. Marcuse, Bell Syst. Tech. J. **41**, 1557 (1962).
2. M. V. Fedorov, *Atomic and Free Electrons in a Strong Laser Field* (World Scientific, Singapore, 1997).
3. V. P. Silin, Zh. Éksp. Teor. Fiz. **47**, 2254 (1964) [Sov. Phys. JETP **20**, 1510 (1964)].
4. V. P. Silin, in *An Introduction to the Kinetic Theory of Gases* (Nauka, Moscow, 1971), p. 141.
5. V. P. Krainov, J. Phys. B **33**, 1585 (2000).
6. V. P. Krainov and M. B. Smirnov, Usp. Fiz. Nauk **170**, 969 (2000).
7. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Nauka, Moscow, 1988; Pergamon, Oxford, 1975).
8. V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Quantum Electrodynamics* (Nauka, Moscow, 1989; Pergamon, Oxford, 1982).
9. T. Ditmire, T. Donnelly, A. M. Rubenchik, *et al.*, Phys. Rev. A **53**, 3379 (1996).
10. J. Zweiback, T. E. Cowan, R. A. Smith, *et al.*, Phys. Rev. Lett. **85**, 3640 (2000).
11. B. N. Chichkov and S. A. Uryupin, Phys. Rev. E **48**, 4659 (1993).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 969–978.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1118–1128.

PLASMA, GASES

Ionization Formation of Plasma Inhomogeneity by the Near-Zone Field of a Magnetic-Type Source in a Magnetized Plasma

A. V. Kudrin*, L. E. Kurina, and E. Yu. Petrov
Nizhni Novgorod State University, Nizhni Novgorod, 603600 Russia
**e-mail: kud@rf.unn.runnet.ru*
Received November 16, 2000

Abstract—An investigations is made of the steady-state structure of a plasma inhomogeneity arising as a result of high-frequency heating and additional ionization of a background magnetized plasma by the near-zone field of a magnetic-type source (ring electric current). It is assumed that the source axis is parallel to an external magnetic field; the source frequency belongs in the low hybrid band. The main attention is focused on the particular case (important for possible applications) when the characteristic longitudinal and transverse scales of density distribution considerably exceed the corresponding scales of distribution of the electron temperature and of the source field. Simplified equations for the near-zone field of the source, the electron temperature, and the plasma density are written for this particular case. Based on the numerical solution of these equations, steady-state distributions of plasma parameters in the formed plasma inhomogeneity are found. It is demonstrated that a plasma inhomogeneity proves to be markedly extended along the external magnetic field. It is found that, for the values of the source current that are attainable under the conditions of active ionospheric and model laboratory experiments, the maximum plasma density in a nonuniform plasma may appreciably exceed the background value.

1. INTRODUCTION

Nonuniform plasma structures, developed in a magnetized plasma by the fields of electromagnetic sources, are of considerable interest in the context of problems associated with high-frequency heating of plasma, with improvements in the methods of "active" plasma diagnostics, and with the artificial effect on the parameters of ionospheric plasma. The singularities associated with the development of such plasma structures were subjected to both experimental and theoretical investigations (see [1–12] and the references cited in these papers). Particular interest is aroused by the possibilities of developing strong disturbances of plasma as a result of nonlinear ionization effects in the near zone of the sources, i.e., the region which is usually characterized by the maximal values of the electric field intensity.

Of the studies pertaining to the above-mentioned scope of problems, note the theoretical investigation (performed by Mareev and Chugunov [10]) of the steady-state structure of high-frequency discharge in the near-zone field of an electric-type antenna operating in a rarefied neutral gas in the presence of external constant magnetic field. Note, however, that no theoretical models are in fact available at present that would describe the processes of discharge structures forming in the near-zone field of radiators located in the already existing (background) magnetoactive plasma. At the same time, it is such formulation of the problem that is characteristic of a number of laboratory [8, 10, 11] and ionospheric [9] experiments, in which it was demonstrated that an additional ionization of background plasma in the field of an antenna (of both electric and magnetic types) may serve as an efficient method of developing strong disturbances of plasma extended along the external magnetic field. The presence of such nonuniform plasma formations, as is demonstrated in [13], leads to a considerable increase in the power of radiation of the sources in some frequency bands, as well as to a number of other interesting effects [14].

This paper deals with the investigation of the steady-state structure of a nonuniform plasma formation developed in the case of additional ionization of a background magnetized plasma by the near-zone field of a magnetic-type source (ring electric current). We will restrict ourselves to treating the case in which the source frequency ω belongs in the whistler (lower hybrid) band,

$$(\Omega_H \omega_H)^{1/2} \ll \omega < \omega_H \ll \omega_p, \qquad (1)$$

where ω_H and ω_p denote the gyrofrequency and the plasma frequency of electrons, respectively, and Ω_H is the gyrofrequency of ions. Note that the specific nature of nonlinear effects in this range, which is of great interest from the viewpoint of various applications [10, 13], is largely due to the presence of the resonance cones of the sources [10, 15].

1063-7761/01/9206-0969$21.00

2. BASIC EQUATIONS

We will treat the given ring electric current (harmonic in time: ($\propto \exp(i\omega t)$) of the density

$$\mathbf{j}(\mathbf{r}) = \varphi_0 \frac{2I_0}{a^2}\rho \exp\left(-\frac{\rho^2}{a^2}\right)\delta(z), \tag{2}$$

where I_0 is the amplitude of total current, a is the characteristic scale of current distribution over the transverse coordinate ρ (ρ, φ, and z are cylindrical coordinates, and δ is Dirac's delta function). It is assumed that the z axis is directed along the external magnetic field,

$$\mathbf{B}_0 = B_0 \mathbf{z}_0.$$

The maximum value of the density of the given current specified in the form of Eq. (2) is evidently attained at

$$\rho = a_0 = a/\sqrt{2}.$$

Note that the diffusion of current over the cross section is necessary for eliminating the singularity of the field on resonance cones of the source.

A cold magnetized plasma, in which the source is located, is described by the permittivity tensor

$$\hat{\varepsilon} = \begin{pmatrix} \varepsilon & -ig & 0 \\ ig & \varepsilon & 0 \\ 0 & 0 & \eta \end{pmatrix}. \tag{3}$$

In the frequency band given by Eq. (1), the components of the tensor (3) have the form [16]

$$\varepsilon = \frac{\omega_p^2}{\omega_H^2 - \omega^2}\left(1 - i\frac{\nu_e}{\omega}\frac{\omega_H^2 + \omega^2}{\omega_H^2 - \omega^2}\right),$$

$$g = -\frac{\omega_p^2 \omega_H}{(\omega_H^2 - \omega^2)\omega}\left(1 - i\frac{2\nu_e \omega}{\omega_H^2 - \omega^2}\right), \tag{4}$$

$$\eta = -\frac{\omega_p^2}{\omega^2}\left(1 + i\frac{\nu_e}{\omega}\right),$$

provided that the additional conditions $\nu_e \ll \omega$ and $\nu_e \ll \omega_H - \omega$ (ν_e is the effective frequency of electron collisions) are valid.

In what follows, the characteristic source scale a is in all instances taken to be small compared to the electromagnetic wavelength

$$\lambda \sim \frac{2\pi}{k_0 |\varepsilon - g|^{1/2}},$$

which enables one to use the quasi-electrostatic approximation for describing the near-zone field. Within this approximation, the equations for the field in the near zone of a magnetic-type source are written as [10, 15]

$$\mathbf{E} = -\nabla\psi - ik_0\mathbf{A}, \quad \mathbf{B} = \operatorname{rot}\mathbf{A}, \tag{5}$$

$$\operatorname{div}(\hat{\varepsilon}\nabla\psi) = -ik_0 \operatorname{div}(\hat{\varepsilon}\mathbf{A}), \tag{6}$$

where the vector potential $\mathbf{A}$ in its turn satisfies the equation

$$\Delta\mathbf{A} = -(4\pi/c)\mathbf{j}, \tag{7}$$

ψ is the scalar potential, and $k_0 = \omega/c$ is the wave number in free space.

We will use the balance equations for density and energy [17] in order to describe the steady-state distribution of plasma density N and electron temperature T_e in a nonuniform plasma formation arising as a result of heating and additional ionization of background plasma in the source field. The equation for steady-state distribution of N has the form

$$\operatorname{div}\mathbf{\Gamma}_e = \operatorname{div}\mathbf{\Gamma}_i = (\nu_i - \nu_\alpha)N - \alpha N^2 + q_{ext}, \tag{8}$$

where $\mathbf{\Gamma}_e$ and $\mathbf{\Gamma}_i$ denote the flux density of electrons and ions, respectively; ν_i is the rate of ionization of neutron molecules by electron impact; ν_a is the rate of electron attachment; α is the coefficient of electron-ion recombination; and q_{ext} is the intensity of the external source which maintains the equilibrium value of plasma density N_0,

$$q_{ext} = (\alpha_0 N_0 + \nu_{a0} - \nu_{i0})N_0$$

(the subscript zero indicates the background values of the respective quantities).

We will write the equation for steady-state distribution of temperature T_e. We will assume that the following correlations are valid in the nonuniform plasma formation:

$$\nu_{ei} \ll \nu_{en}, \quad \delta_{en}\nu_{en} \ll \omega, \quad \delta_{ei}\nu_{ei} \ll \nu_{in}. \tag{9}$$

Here, ν_{in}, ν_{en}, and ν_{ei} denote the rate of collision of ions and electrons with neutral molecules and ions, respectively ($\nu_e = \nu_{en} + \nu_{ei}$), and $\delta_{en}$and δ_{ei} are average relative fractions of the energy lost by electrons in collisions with neutral molecules and ions. The last inequality in (9) enables one to disregard the heating of ions [17]. We further assume that the characteristic scale of nonuniformity of the field amplitude considerably exceeds the transverse scale of the electron thermal conductivity, and it is possible to disregard the heat flux across the transverse magnetic field in the balance equation for energy. As a result, we have the following equation for the electron temperature [17, 18]:

$$\frac{1}{N}\frac{\partial}{\partial z}\left(\kappa_{e\|}\frac{\partial T_e}{\partial z} + \beta_{e\|}\Gamma_{e\|}\right) - \frac{\Gamma_{e\|}}{N}\frac{\partial T_e}{\partial z} - \delta_{en}\nu_e(T_e - T_{e0}) + \frac{2Q}{3N} = 0, \tag{10}$$

where $\kappa_{e\|}$ is the electron thermal conductivity along the external magnetic field; $\Gamma_{e\|}$ is the longitudinal component of the vector Γ_e; $\beta_{e\|}$ is the component of the ther-

mal force tensor for electrons $\hat{\beta}_e$ ($\beta_{e\|} = (\hat{\beta}_e)_{zz}$); T_{e0} is the background value of the electron temperature, taken to be equal to the neutral temperature; and

$$Q = \frac{3}{2}\nu_e N\tilde{Q}(|\mathbf{E}|)$$

is the time average power of Joule heating. The quantity $\tilde{Q}(|\mathbf{E}|)$ may be represented as

$$\tilde{Q}(|\mathbf{E}|) = \frac{e^2}{3m\omega^2}\left[\frac{(\omega_H^2+\omega^2)\omega^2}{(\omega_H^2-\omega^2)^2}(|E_\rho|^2+|E_\varphi|^2) + 4\frac{\omega_H\omega^3}{(\omega_H^2-\omega^2)^2}\mathrm{Im}(E_\rho E_\varphi^*)+|E_z|^2\right], \tag{11}$$

where e is the absolute value of electron charge and m is the electron mass. In deriving formula (11), we used expressions (4) for the components of the plasma dielectric tensor.

In order to avoid misunderstanding, note that, in changing over to a simplified form of Eq. (10) for the energy balance equation, we ignored some unimportant terms of the order of $(\delta_{en}\nu_e N)^{-1}\mathrm{div}\mathbf{\Gamma}_e$. These terms are small if the inequalities

$$|\nu_i - \nu_a| \ll \delta_{en}\nu_e, \quad \alpha N \ll \delta_{en}\nu_e$$

are satisfied; these inequalities are always valid under conditions of real interest to us, in which a nonuniform plasma formation is developed in the ionosphere or in a laboratory plasma which simulates the situation in the ionosphere.

In the case of weakly ionized plasma treated here, the expressions for the electron and ion flux density may be written as

$$\mathbf{\Gamma}_e = N\hat{b}_e\nabla\phi - \hat{D}_e\nabla N - \hat{D}_e^{(T)}\frac{N}{T_e}\nabla T_e,$$
$$\mathbf{\Gamma}_i = -N\hat{b}_i\nabla\phi - \hat{D}_i\nabla N, \tag{12}$$

where ϕ is the potential of the electric field arising in the plasma due to the presence of nonuniformity; $\hat{D}_e$ and $\hat{b}_e$ are the tensors of diffusion and mobility for electrons, respectively; $\hat{D}_i$ and $\hat{b}_i$ are the respective tensors for ions; and $\hat{D}_e^{(T)}$ is the tensor of thermodiffusion for electrons. In the case of azimuthally symmetric plasma formations, we can restrict ourselves to analyzing only the behavior of the diagonal components of these tensors. The respective components, as well as the quantities $\kappa_{e\|}$ and $\beta_{e\|}$ entering Eq. (10), allow the following representation:

$$D_{e\|} = K_D\frac{T_e}{m\nu_e}, \quad D_{e\perp} = \frac{T_e\nu_e}{m\omega_H^2},$$
$$D_{i\|} = \frac{T_i}{M\nu_{in}}, \quad D_{i\perp} = \frac{T_i}{M\nu_{in}}\left(1+\frac{\Omega_H^2}{\nu_{in}^2}\right)^{-1},$$
$$b_{e\|} = \frac{e}{T_e}D_{e\|}, \quad b_{e\perp} = \frac{e}{T_e}D_{e\perp},$$
$$b_{i\|} = \frac{e}{T_i}D_{i\|}, \quad b_{i\perp} = \frac{e}{T_i}D_{i\perp}, \tag{13}$$
$$D_{e\|}^{(T)} = k_{e\|}^{(T)}D_{e\|}, \quad D_{e\perp}^{(T)} = k_{e\perp}^{(T)}D_{e\perp},$$
$$\kappa_{e\|} = \kappa_0\frac{NT_e}{m\nu_e}, \quad \beta_{e\|} = \beta_0 T_e.$$

Here, $k_{e\|}^{(T)}$ and $k_{e\perp}^{(T)}$ are thermodiffusion relations given by the formulas

$$k_{e\|}^{(T)} = 1 - \frac{T_e}{\nu_e}\frac{d\nu_e}{dT_e}, \quad k_{e\perp}^{(T)} = 1 + \frac{T_e}{\nu_e}\frac{d\nu_e}{dT_e}, \tag{14}$$

K_D, κ_0, and β_0 are coefficients dependent on the form of the function $\nu_e(T_e)$; T_i is the ion temperature ($T_i = T_{e0}$); and M is the ion mass; the symbols "$\|$" and "$\perp$" indicate, as usual, the directions along and across the external magnetic field, respectively.

Equations (5)–(8) and (10), along with relations (11)–(14), enable one to investigate the steady-state structure of a plasma inhomogeneity arising in an undisturbed (background) plasma as a result of nonlinear ionization effects in the near zone of a high-frequency source of the magnetic type. The results of solving these equations are largely defined by the dependence of the quantities appearing in them on temperature and other factors. Therefore, in order to obtain any results, these dependences must be defined concretely.

We will treat ionization effects which occur in the air at typical discharge temperatures of the order of $T_e \sim$ 5–10 eV. In this case, the temperature dependence of quantities appearing in Eqs. (8) and (13) may be approximately described by the following model expressions [17, 19]:

$$\nu_e = 2\times10^{-7}N_nT_e(1+T_e^{1/2})^{-1},$$
$$\nu_i = 2.7\times10^{-8}N_n(T_e/I)^{1/2} \times(1+2T_e/I)\exp(-I/T_e),$$
$$\nu_a = \beta_a(T_e)N_n, \tag{15}$$
$$\alpha = 2.5\times10^{-7}(0.026/T_e)^{1.2} + 1.1\times10^{-7}(0.026/T_e)^{0.7},$$

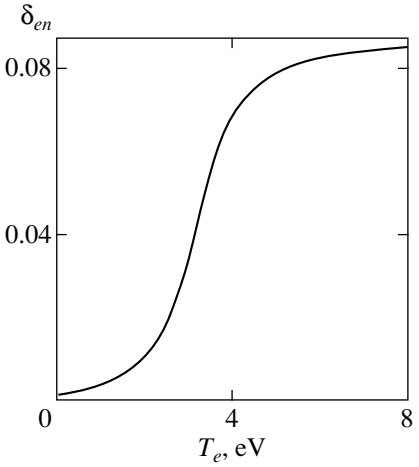


Fig. 1. The model temperature T_e dependence of the average fraction of energy lost by an electron upon collision with neutral molecules in air. It is assumed within the model employed that $\delta_{en} \longrightarrow 0.1$ at $T_e \longrightarrow \infty$.

in which the quantities ν_e, ν_i, and ν_a are expressed in s^{-1}, T_e in eV, α in $cm^3\ s^{-1}$, the concentration of neutral molecules N_n in cm^{-3}; I is the effective ionization potential of neutral molecules, hereinafter taken to be $I = 14$ eV. We will assume the attachment coefficient β_a to be

$$\beta_a = 3 \times 10^{-13}[1 + 10(T_e - T_{e0})]$$
$$\text{at } T_{e0} \le T_e \le 10 \text{ eV}.$$

For the quantity $\delta_{en}(T_e)$, we will use the approximation given in Fig. 1. In the electron temperature range T_e = 1–10 eV, this correlation is in adequate agreement with the experimental data for air available in the literature [20]. With the adopted form of the function $\nu_e(T_e)$, the quantities K_D, κ_0, and β_0 in Eq. (13) may be approximately assumed to be independent of T_e for the values of electron temperature being treated,

$$K_D \approx 1, \quad \kappa_0 \approx 1.5, \quad \beta_0 \approx -1/3.$$

3. FIELD AND ELECTRON TEMPERATURE DISTRIBUTION

The solution of the set of equations (5)–(8) and (10) is a fairly complicated problem. In this paper, we will restrict ourselves to treating a particular case, which is of importance from the standpoint of possible applications, when these equations may be investigated within some simplified approach. We will assume that the characteristic longitudinal and transverse scales $l_{N\|}$ and $l_{N\perp}$ of the plasma density distribution considerably exceed the respective scales of electron temperature distribution ($l_{T\|}$ and $l_{T\perp}$) and of source field distribution ($l_{E\|}$ and $l_{E\perp}$). In view of obvious relations

$$l_{E\|} \sim (-\eta/\varepsilon)^{1/2} a$$

(compare with [10]) and $l_{T\perp} \sim l_{E\perp} \sim a$, the formulated condition is written as

$$l_{N\|} \gg \max\{l_{T\|}, (-\eta/\varepsilon)^{1/2} a\}, \quad l_{N\perp} \gg a. \tag{16}$$

In this case, as a first approximation, one can ignore the space derivatives of the plasma density N in Eqs. (6) and (10).

As a result, Eq. (6) for the potential ψ takes the form

$$\gamma^2 \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\psi}{\partial\rho}\right) - \frac{\partial^2\psi}{\partial z^2} = k_0\gamma_g^2\frac{1}{\rho}\frac{\partial}{\partial\rho}(\rho A_\varphi), \tag{17}$$

where

$$\gamma^2 = -\varepsilon/\eta, \quad \gamma_g^2 = g/\eta;$$

it is taken into account that in the case being treated

$$A_\rho = A_z = 0.$$

Next we ignore the small terms corresponding to the inclusion of collisions in expressions (4) for the components of the dielectric tensor and derive the following formulas for γ and γ_g:

$$\gamma = [(\omega_H/\omega)^2 - 1]^{-1/2}, \quad \gamma_g = \gamma(\omega_H/\omega)^{1/2}. \tag{18}$$

As is seen from Eq. (18), the quantities γ and γ_g in the frequency band given by Eq. (1) do not depend on the plasma density.

As for Eq. (10), the terms $N^{-1}\Gamma_{e\|}\partial T_e/\partial z$ and $N^{-1}\partial(\beta_{e\|}\Gamma_{e\|})/\partial z$ in this equation may be ignored along with the derivatives of density N. As one can readily see, in the case of (16), the maximal contributions by these terms with respect to the term containing the thermal conductivity $\kappa_{e\|}$ are small quantities of the order of $k_{e\|}^{(T)}/\kappa_0$ and $\beta_0 k_{e\|}^{(T)}/\kappa_0$, respectively. Finally, we derive

$$\frac{\partial^2 T_e}{\partial z^2} + \frac{1}{T_e}\left(1 - \frac{T_e}{\nu_e}\frac{\partial\nu_e}{\partial T_e}\right)\left(\frac{\partial T_e}{\partial z}\right)^2 + \frac{m\nu_e^2}{\kappa_0 T_e}$$
$$\times [\tilde{Q}(|\mathbf{E}|) - \delta_{en}(T_e - T_{e0})] = 0. \tag{19}$$

The solution of Eq. (19) must satisfy the following conditions:

$$\frac{\partial T_e}{\partial z} = 0 \text{ at } z = 0;$$
$$T_e \longrightarrow T_{e0} \text{ at } z \longrightarrow \infty. \tag{20}$$

Therefore, the distributions of the near-zone field of a magnetic-type source and of the electron temperature, when conditions (16) are valid, are described to a first approximation by equations which are independent (with due regard for relations (18)) of the plasma density.

We will now turn to seeking the solution to equations for the near-zone field. One can readily see that the Fourier transforms of the field and current,

$$\mathbf{E}(\mathbf{n}) = \int \mathbf{E}(\mathbf{r})\exp(ik_0\mathbf{n}\cdot\mathbf{r})d\mathbf{r},$$

$$\mathbf{j}(\mathbf{n}) = \int \mathbf{j}(\mathbf{r})\exp(ik_0\mathbf{n}\cdot\mathbf{r})d\mathbf{r},$$

are related by

$$\mathbf{E}(\mathbf{n}) = i\frac{4\pi}{\omega(\mathbf{n}\cdot\mathbf{n})}\left[\frac{\mathbf{n}(\mathbf{n}\cdot\hat{\varepsilon}\mathbf{j}(\mathbf{n}))}{(\mathbf{n}\cdot\hat{\varepsilon}\mathbf{n})} - \mathbf{j}(\mathbf{n})\right],$$

whence, after performing inverse Fourier transformation, we have

$$E_\rho = -iE_0\frac{\gamma_g^2}{1+\gamma^2}(F_1^{(1)} + i\gamma^{-1}F_1^{(2)}),$$

$$E_\varphi = E_0F_1^{(1)}, \qquad (21)$$

$$E_z = \mathrm{sgn}(z)E_0\frac{\gamma_g^2}{1+\gamma^2}(F_0^{(1)} - F_0^{(2)}).$$

Here,

$$E_0 = I_0\pi^{1/2}k_0/c,$$

$$F_m^{(1,2)} = \int_0^{2\pi} \xi_{1,2}\,\mathrm{erfc}(\xi_{1,2})\exp(\xi_{1,2}^2 + im\zeta)d\zeta, \qquad (22)$$

$$m = 0, 1,$$

where

$$\xi_1 = \frac{|z| - i\rho\cos\zeta}{a}, \quad \xi_2 = i\frac{\gamma|z| - \rho\cos\zeta}{a},$$

erfc(ξ) is the complementary error function:

$$\mathrm{erfc}(\xi) = \frac{2}{\pi^{1/2}}\int_\xi^\infty \exp(-\tau^2)d\tau.$$

Expressions (21) and (22) make it possible to calculate the distribution of the field components in the near zone of the source given by Eq. (2). Further we substitute the appropriate distributions into formula (11) for the quantity $\tilde{Q}$ to find the solution of Eq. (19) that would satisfy conditions (20). The distribution of the field components and of the electron temperature was calculated numerically for specified values of the parameters

$$k_0a = 0.02, \quad \omega/\omega_H = 0.34,$$

$$\omega_{p0}/\omega_H = 3.5, \quad \nu_{e0}/\omega_H = 8.6\times10^{-5}.$$

The size of the source was taken to be $a = 2$ m. Note that, in this case, the above-identified values of the dimensionless parameters characterizing a background plasma correspond to the conditions of active experiments in the Earth's ionosphere at an altitude of ~150 km (see [9, 17]): $N_0 = 3\times10^5$ cm^{-3} and $B_0 = 0.5$ G (we will further indicate the values of $T_{e0} = 0.07$ eV, $\nu_{e0}/\nu_{in} = 7.6$, and $\nu_{in}/\Omega_H = 0.63$).

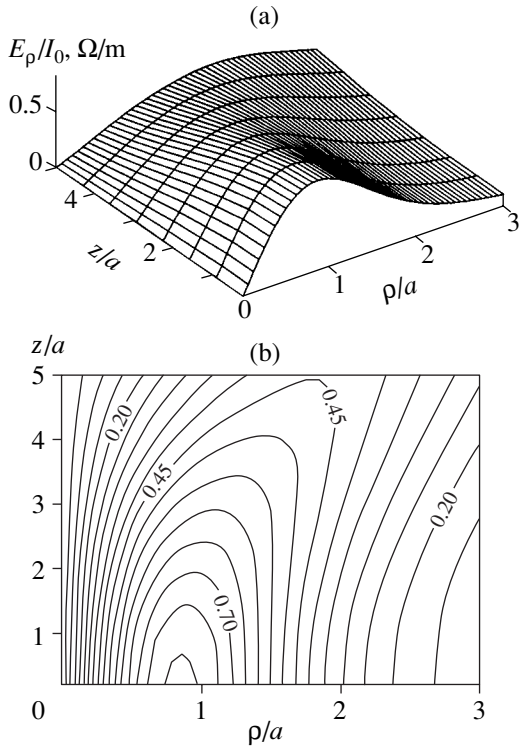


Fig. 2. The space distribution of the radial component of the near-zone electric field of the source given by Eq. (2) at γ = 0.36, $\gamma_g = 0.62$, $k_0a = 0.02$, $a = 2$ m; (a) the dependence of $|E_\rho|$ on the coordinates ρ, z; (b) $|E_\rho|$ level lines.

The results of calculating the field components are given in Figs. 2–4. The dependences of the absolute values of the field components on the coordinates ρ and z are given, as well as the isolines corresponding to these dependences. In addition, Fig. 5 gives analogous graphs for the quantity $\tilde{Q}$ that characterizes the spatial distribution of the power of Joule heating in the plasma. As is demonstrated by the data presented here, the power of Joule heating exhibits local maxima at $\rho = a_0$ (in the $z = 0$ plane) and at the focal point $\rho = 0$, $z = \gamma^{-1}a_0$, which is the vertex of a convergent resonance cone for a ring source with the maximal value of current density at $\rho = a_0$, $z = 0$. A divergent resonance cone originating on the source and a divergent resonance cone with the vertex at the focal point prove to be less pronounced, although the absolute value of the field and the value of $\tilde{Q}$ decrease along the generatrices of these cones at a slower rate than in the other directions (see Figs. 2b, 4b, and 5b).

As for the distribution of the electron temperature $T_e(\rho, z)$, it is characterized by the presence of a maxi-

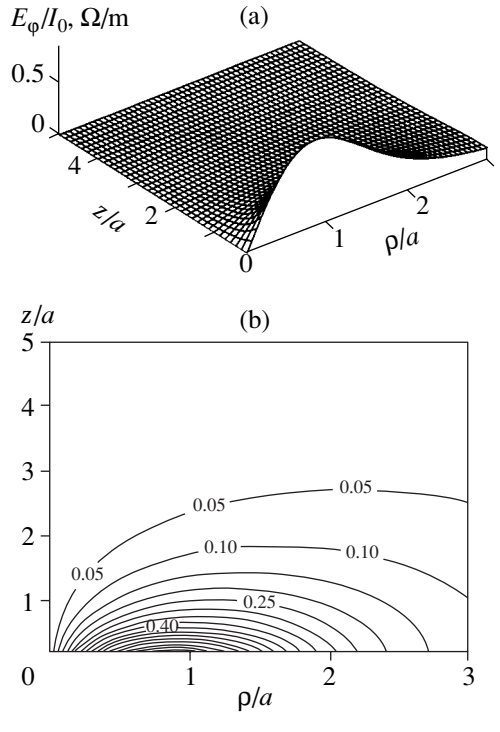


Fig. 3. The space distribution of the azimuth component of the near electric field of the source given by Eq. (2). The values of parameters are as given in Fig. 2; (a) the dependence of $|E_\varphi|$ on the coordinates ρ, *z*; (b) $|E_\varphi|$ level lines.

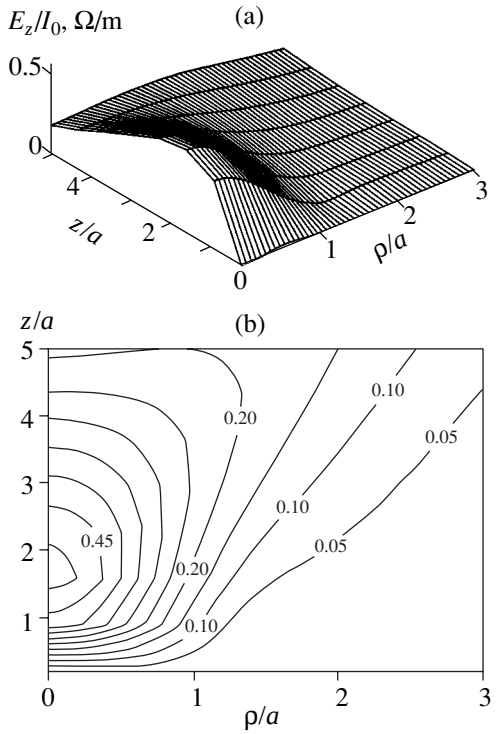


Fig. 4. The spatial distribution of the longitudinal component of the near electric field of the source given by Eq. (2). The values of parameters are as given in Fig. 2; (a) the dependence of $|E_z|$ on the coordinates ρ, *z*; (b) $|E_z|$ level lines.

mum at the origin of the reference frame ρ = 0 and *z* = 0, in spite of the nonmonotonic behavior of the $\tilde{Q}$ function in the neighborhood of this point. This singularity of the T_e profile is attributed to the fact that the temperature distribution is markedly affected by the longitudinal electron heat conduction. The dependence of the maximal electron temperature $T_e(0, 0)$ on the source current I_0 is given in Fig. 6. Figure 7 gives the temperature distribution on the transverse and longitudinal coordinates, which corresponds to the value of current I_0 = 76 A. One can readily see that the transverse scale of temperature distribution is defined by the source size ($l_{T\perp} \sim a_0$), and the longitudinal scale of temperature distribution is defined by the characteristic length of electron heat conduction along the external magnetic field,

$$l_{T\parallel} \sim (T_e/m\nu_e^2\delta_{en})^{1/2}.$$

It must be emphasized that the results obtained are valid only if conditions (16) are satisfied. We will see below that, as the temperature T_e increases, inequalities (16) cease to be valid with the necessary margin. Therefore, for relatively high temperatures (for T_e > 10 eV), the theoretical model being treated needs to be considerably refined.

4. PLASMA DENSITY DISTRIBUTION

An analysis of the results of solving the equation for the electron temperature T_e reveals that the characteristic dimensions $l_{T\parallel}$ and $l_{T\perp}$ of the region with a higher value of T_e satisfy the condition

$$l_{T\parallel}^2/l_{T\perp}^2 \gtrsim D_{e\parallel}/D_{i\perp},$$

which is transformed into the inequality

$$l_{T\parallel}^2/l_{T\perp}^2 \gg D_{e\parallel,0}/D_{i\perp}$$

upon substitution of the background value $D_{e\parallel,0}$ of the diffusion coefficient $D_{e\parallel}$. It is natural to assume that an analogous inequality

$$l_{N\parallel}^2/l_{N\perp}^2 \gg D_{e\parallel}/D_{i\perp} \tag{23}$$

is valid for the characteristic dimensions of a region with a higher plasma density. As we will see below, this assumption is supported by the results of calculating the density distribution.

One can readily demonstrate [21] that, given the validity of condition (23) and of the inequality

$$\omega_H\Omega_H \gg (T_e/T_i)\nu_{en}\nu_{in}(1 + \Omega_H^2/\nu_{in}^2),$$

the density balance equation (8) may be approximately represented as

$$\mathrm{div}_\perp(D_\perp \nabla_\perp N) + \frac{\partial}{\partial z}\left(D_\parallel \frac{\partial N}{\partial z}\right) + \mathrm{div}\left(\hat{D}_e^{(T)} \frac{N}{T_e} \nabla T_e\right) + (\nu_i - \nu_a)N - \alpha N^2 + q_{ext} = 0, \quad (24)$$

where

$$D_\perp = D_{e\perp}\left(1 + \frac{T_i}{T_e}\right), \quad D_\parallel = D_{e\parallel}\left(1 + \frac{T_i}{T_e}\right). \quad (25)$$

Without dwelling on the details of deriving Eq. (24), note that it is derived from the initial equation (8) if we set

$$\nabla\phi = -(T_i/eN)\nabla N$$

in the expression for the flux $\mathbf{\Gamma}_e$ in Eq. (12), which is approximately valid if condition (23) is satisfied (for details, see [18, 21]). We will further take into account the fact that, in the case given by (16), the thermodiffusion term in Eq. (24) admits the representation

$$\mathrm{div}\left(\hat{D}_e^{(T)} \frac{N}{T_e} \nabla T_e\right) \approx N\,\mathrm{div}\left(\hat{D}_e^{(T)} \frac{\nabla T_e}{T_e}\right).$$

In the heated part of the nonuniform plasma formation, the contribution of this term considerably exceeds the contribution made by the proper diffusion terms. In view of this, we will replace the quantities given by (25) by their background values

$$D_{\perp 0} = 2D_{e\perp,0}, \quad D_{\parallel 0} = 2D_{e\parallel,0}.$$

It is obvious that such a replacement, which is exact outside of the heating region, has no appreciable effect on the behavior of the left-hand side of Eq. (24) for the steady-state distribution of plasma density. We derive

$$D_{\perp 0}\Delta_\perp N + D_{\parallel 0}\frac{\partial^2 N}{\partial z^2} + \nu(\mathbf{r})N - \alpha(\mathbf{r})N^2 + q_{ext} = 0, \quad (26)$$

where

$$\nu(\mathbf{r}) = \nu_i - \nu_T - \nu_a, \quad \nu_T = -\mathrm{div}\left(\hat{D}_e^{(T)} \frac{\nabla T_e}{T_e}\right).$$

The concrete form of the functions $\nu(\mathbf{r})$ and $\alpha(\mathbf{r})$ is defined by the dependence on temperature $T_e(\mathbf{r})$ of the quantities entering these functions.

Because the quantities $\nu(\mathbf{r})$ and $\alpha(\mathbf{r})$ differ considerably from their background values ν_0 and α_0 only in the heating region, whose dimensions $l_{T\parallel}$ and $l_{T\perp}$ are small compared to the dimensions $l_{N\parallel}$ and $l_{N\perp}$ of the nonuniform plasma formation, the functions

$$\Delta\nu(\mathbf{r}) = \nu(\mathbf{r}) - \nu_0, \quad \Delta\alpha(\mathbf{r}) = \alpha(\mathbf{r}) - \alpha_0$$

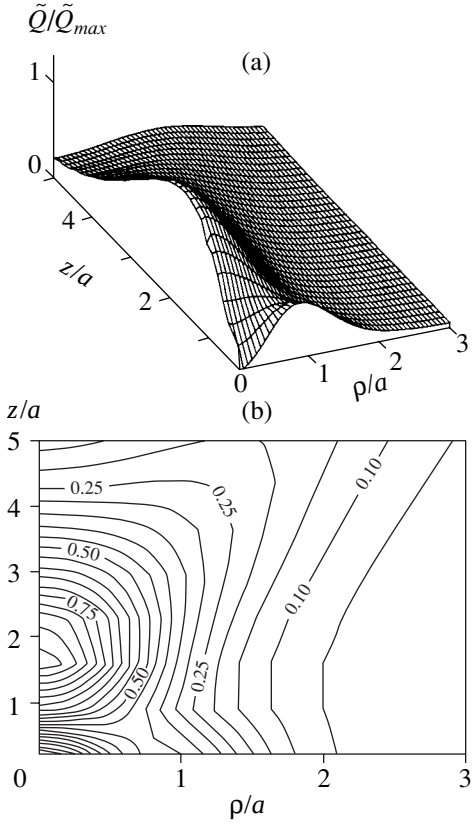


Fig. 5. The spatial distribution of the $\tilde{Q}$ function corresponding to the components of the near electric field given in Figs. 2–4: (a) the dependence of $\tilde{Q}/\tilde{Q}_{\max}$ on the coordinates ρ, z; (b) $\tilde{Q}/\tilde{Q}_{\max}$ level lines.

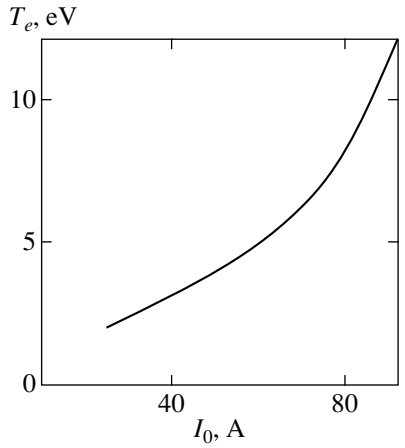


Fig. 6. The maximum electron temperature $T_{e,\max} = T_e(0, 0)$ as a function of the source current I_0.

are characterized by the δ-function-like behavior against the background of the $N(\mathbf{r})$ distribution. Therefore, the above-mentioned functions may be replaced by some model distributions $f_\nu(\mathbf{r})$ and $f_\alpha(\mathbf{r})$ decreasing fairly rapidly away from the point $\mathbf{r} = 0$ and satisfying the conditions

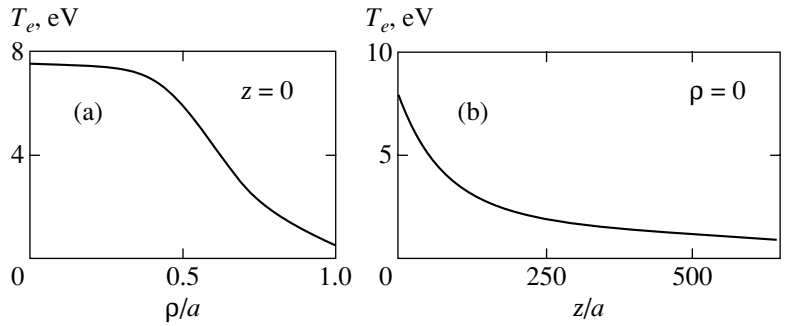


Fig. 7. The distribution of the electron temperature $T_e(\rho, z)$ on the transverse and longitudinal coordinates at $I_0 = 76$ A: (a) $T_e(\rho, 0)$, (b) $T_e(0, z)$.

$$\int f_\nu(\mathbf{r})d\mathbf{r} = \int \Delta\nu(\mathbf{r})d\mathbf{r},$$
$$\int f_\alpha(\mathbf{r})d\mathbf{r} = \int \Delta\alpha(\mathbf{r})d\mathbf{r}. \quad (27)$$

We will select the distributions $f_\nu(\mathbf{r})$ and $f_\alpha(\mathbf{r})$ in the form

$$f_\nu(\mathbf{r}) = \frac{\tilde{\nu} - \nu_0}{\cosh^2[(\rho/\rho_0)^2 + (z/z_0)^2]^{1/2}},$$
$$f_\alpha(\mathbf{r}) = \frac{\tilde{\alpha} - \alpha_0}{\cosh^2[(\rho/\rho_0)^2 + (z/z_0)^2]^{1/2}}, \quad (28)$$

where ρ_0 and z_0 must satisfy the inequalities

$$\rho_0 \ll l_{N\perp}, \quad z_0 \ll l_{N\parallel},$$

and the constants $\tilde{\nu}$ and $\tilde{\alpha}$ are defined by the formulas

$$\tilde{\nu} = \nu_0 + \frac{3}{\pi^3\rho_0^2 z_0}\int \Delta\nu(\mathbf{r})d\mathbf{r},$$
$$\tilde{\alpha} = \alpha_0 + \frac{3}{\pi^3\rho_0^2 z_0}\Delta\alpha(\mathbf{r})d\mathbf{r}, \quad (29)$$

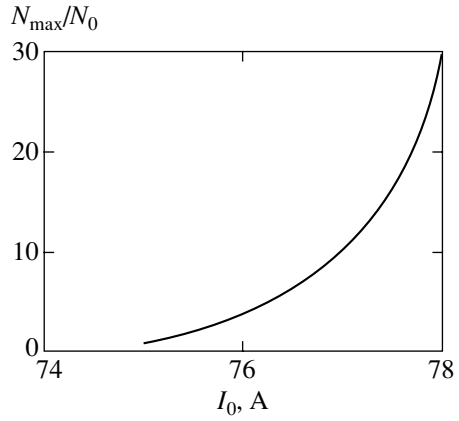


Fig. 8. The maximal plasma density N_{max} as a function of the source current I_0 (N_0 is the density of background plasma).

which follow from (27). In the problem treated here, it is convenient to set

$$\rho_0 = a_0, \quad z_0 = a_0(D_{\parallel 0}/D_{\perp 0})^{1/2}$$

and change in Eq. (26) to new coordinates x, y, and $\tilde{z}$, where

$$\tilde{z} = z(D_{\perp 0}/D_{\parallel 0})^{1/2}.$$

We will finally derive

$$D_{\perp 0}\tilde{\Delta}N + \frac{\tilde{\nu} - \nu_0}{\cosh^2(\tilde{r}/a_0)}N - \frac{\tilde{\alpha} - \alpha_0}{\cosh^2(\tilde{r}/a_0)}N^2$$
$$+ (\nu_{i0} - \nu_{a0})(N - N_0) - \alpha_0(N^2 - N_0^2) = 0, \quad (30)$$

where

$$\tilde{r} = (x^2 + y^2 + \tilde{z}^2)^{1/2}$$

and $\tilde{\Delta}$ denotes the Laplace operator in the new coordinates.

Equation (30) proves to be very convenient for numerical investigation. An analysis of this equation reveals that, for a given value of the source current I_0 (i.e., for a known form of $T_e(\mathbf{r})$ distribution), it has a spatially localized, spherically symmetric (in the x, y, $\tilde{z}$ coordinates) solution of $N(\tilde{r})$ which satisfies the conditions

$$\frac{\partial N}{\partial \tilde{r}} = 0 \text{ for } \tilde{r} = 0;$$
$$N \longrightarrow N_0 \text{ for } \tilde{r} \longrightarrow \infty. \quad (31)$$

Equation (30) was solved numerically for the same values of the source dimensions and of the parameters for the background plasma as those used in the previous section to search for the $T_e(\mathbf{r})$ distribution. The constants $\tilde{\nu}$ and $\tilde{\alpha}$ for each given value of the source current I_0 were calculated by formulas (29) with due regard for the respective $T_e(\mathbf{r})$ distribution. As follows from the performed calculations, the solution of Eq. (30) has a maximum value at $\mathbf{r} = 0$. The dependence of the maxi-

mal plasma density N_{max} on the source current is given in Fig. 8. Figure 9 gives the spatial distribution of density for two cases of $N_{max}/N_0 = 10$ and 30 which correspond to the values of source current $I_0 \approx 77$ and 78 A. One can readily see that the distributions of plasma density obtained satisfy all the restrictions specified above.

Note that the process of formation of density disturbance exhibits a threshold behavior. The value of current $I_0 = 75$ A (in this case, $T_{e,\max} = 7.4$ eV), to which corresponds a relative disturbance of density at the center $N_{max}/N_0 = 1.001$, may be interpreted as the threshold of development of plasma inhomogeneity. As follows from the calculation results, in the case of a slight excess of the threshold value of current, the increase in the ionization rate is largely compensated by an increase in thermodiffusion loss, and the increase in the plasma density in the plasma inhomogeneity is largely due to a decrease in the recombination coefficient ($\partial\alpha/\partial T_e < 0$). In our case, this behavior is observed for I_0 values at which $N_{max}/N_0 < 10$. For these values, the conditions of validity of the approximate description employed by us (see inequalities (16)) are satisfied with a very large margin.[1] As I_0 increases, a sharp increase in the plasma density in the plasma inhomogeneity occurs, which is due to a faster rise of the ionization rate compared with thermodiffusion loss. Although, in this case, the characteristic scales of distribution of plasma density and electron temperature approach each other, conditions (16) are still valid at $N_{max}/N_0 < 50$, though with a smaller margin. As the source current continues to increase, inequalities (16) cease to be valid, and the theoretical model employed becomes inadequate. Nevertheless, as revealed by the foregoing treatment, this model, in spite of the existing restrictions, enables one to describe the steady-state structure of a plasma inhomogeneity with a fairly strong disturbance of density.

Note that, under conditions of increasing transverse dimension a of the source, accompanied by an appreciable decrease in the contribution made by transverse thermodiffusion to the ionization balance, much lower values of the source current I_0 are required to provide for the given disturbance of plasma density N_{max}/N_0. For example, for $a = 3$ m and for the above-identified values of parameters of background plasma, the relative disturbance of density $N_{max}/N_0 = 10$ is attained at $I_0 = 46$ A (instead of $I_0 = 77$ A in the case of $a = 2$ m).

[1] We do not dwell on the analytical calculation of the quantity N_{max}, which may be performed for the given conditions. When model representations (28) are used, the appropriate treatment is performed the way it was done in [22] for a high-frequency discharge in rarefied gas in the absence of external magnetic field.

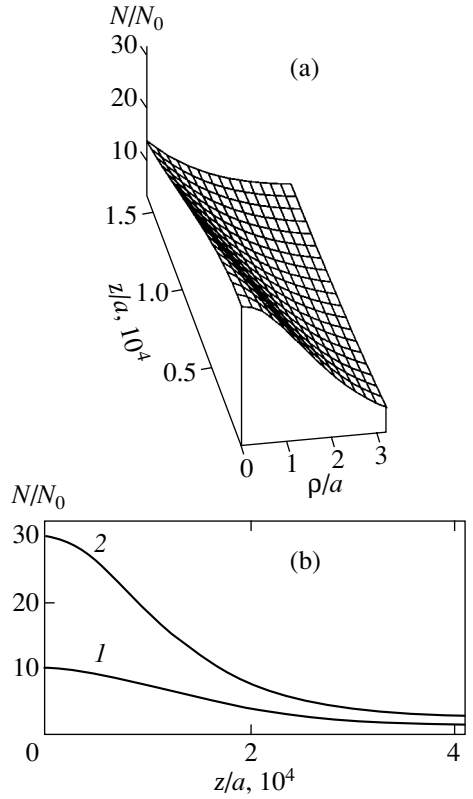


Fig. 9. The spatial distribution of plasma density: (a) the $N(\rho, z)/N_0$ distribution for $N_{max}/N_0 = 30$; (b) the density distribution along the z axis ($\rho = 0$) for $N_{max}/N_0 = 10$ (curve *1*) and for $N_{max}/N_0 = 30$ (curve *2*).

5. CONCLUSION

We have suggested a theoretical model which enables one to investigate the steady-state structure of a plasma inhomogeneity arising upon additional ionization of a magnetized background plasma by the near-zone field of a magnetic-type source, under conditions when characteristic spatial scales of density distribution exceed considerably the size of the heating region. Within this model, it proves possible to study the dependences of all basic characteristics of a steady-state nonuniform plasma formation on the parameters of the source and background plasma. One can see from the treatment results that the characteristics of a nonuniform plasma formation are largely defined by the dependence of the coefficients of transport in a magnetoactive plasma on the electron temperature. It must be emphasized that the presence of nonuniformity of the electron temperature necessitates the inclusion of thermodiffusion in analyzing the distribution of plasma density.

In conclusion, note that the results of our calculations demonstrate that a magnetic-type source placed in a magnetized plasma is capable of maintaining a nonuniform plasma formation markedly extended along the external magnetic field; the density of plasma in this

formation considerably exceeds the background value. This fact is of interest in view of the possibility of appropriate experiments in cosmic and laboratory plasma, as well as from the standpoint of analyzing the operation of some devices designed to produce a dense plasma.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 01-02-16949, 01-02-17390), the programs "Leading Scientific Schools" (project no. 00-15-96734) and "Universities of Russia—Basic Research" (project no. 992852), and the Ministry of Education of the Russian Federation (project no. E00-3.5-227).

REFERENCES

1. R. L. Stenzel, Phys. Fluids **19**, 865 (1976).
2. H. Sugai, M. Maruyama, M. Sato, and S. Takeda, Phys. Fluids **21**, 690 (1978).
3. S. V. Egorov, A. V. Kostrov, and A. V. Tronin, Pis'ma Zh. Éksp. Teor. Fiz. **47**, 86 (1988) [JETP Lett. **47**, 102 (1988)].
4. T. M. Zaboronkova, A. V. Kostrov, A. V. Kudrin, *et al.*, Zh. Éksp. Teor. Fiz. **102**, 1151 (1992) [Sov. Phys. JETP **75**, 625 (1992)].
5. L. E. Kurina, Fiz. Plazmy **24**, 937 (1998) [Plasma Phys. Rep. **24**, 873 (1998)].
6. A. V. Kostrov, A. V. Kudrin, L. E. Kurina, *et al.*, Phys. Scr. **62**, 51 (2000).
7. G. A. Markov, V. A. Mironov, A. M. Sergeev, and I. A. Sokolova, Zh. Éksp. Teor. Fiz. **80**, 2264 (1981) [Sov. Phys. JETP **53**, 1183 (1981)].
8. I. A. Vdovichenko, G. A. Markov, V. A. Mironov, and A. M. Sergeev, Pis'ma Zh. Éksp. Teor. Fiz. **44**, 216 (1986) [JETP Lett. **44**, 275 (1986)].
9. Yu. N. Agafonov, V. S. Bazhanov, V. Ya. Isyakaev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **52**, 1127 (1990) [JETP Lett. **52**, 530 (1990)].
10. E. A. Mareev and Yu. V. Chugunov, *Antennas in Plasma* (Inst. Prikl. Fiz. Akad. Nauk SSSR, Nizhni Novgorod, 1991).
11. G. Yu. Golubyatnikov, S. V. Egorov, B. G. Eremin, *et al.*, Zh. Éksp. Teor. Fiz. **107**, 441 (1995) [JETP **80**, 234 (1995)].
12. A. V. Kudrin, L. E. Kurina, and G. A. Markov, Zh. Éksp. Teor. Fiz. **112**, 1285 (1997) [JETP **85**, 697 (1997)].
13. I. G. Kondrat'ev, A. V. Kudrin, and T. M. Zaboronkova, *Electrodynamics of Density Ducts in Magnetized Plasmas* (Gordon and Breach, Amsterdam, 1999).
14. I. G. Kondrat'ev, A. V. Kudrin, and T. M. Zaboronkova, J. Atmos. Sol.-Terr. Phys. **59**, 2475 (1997).
15. V. I. Karpman, Fiz. Plazmy **12**, 836 (1986) [Sov. J. Plasma Phys. **12**, 480 (1986)].
16. V. L. Ginzburg, *The Propagation of Electromagnetic Waves in Plasmas* (Nauka, Moscow, 1967; Pergamon, Oxford, 1970).
17. A. V. Gurevich and A. B. Shvartsburg, *The Nonlinear Theory of Radio Waves Propagation in the Ionosphere* (Nauka, Moscow, 1973).
18. V. A. Rozhanskiĭ and L. D. Tsendin, *Collisional Transport in a Partially Ionized Plasma* (Énergoatomizdat, Moscow, 1988).
19. Yu. P. Raĭzer, *Gas Discharge Physics* (Nauka, Moscow, 1987; Springer-Verlag, Berlin, 1991).
20. H. S. W. Massey and E. H. S. Burhop, *Electronic and Ionic Impact Phenomena* (Clarendon Press, Oxford, 1952; Inostrannaya Literatura, Moscow, 1958).
21. B. N. Gershman, *Dynamics of the Ionospheric Plasma* (Nauka, Moscow, 1974).
22. V. E. Semenov, Candidate's Dissertation in Physical and Mathematical Sciences (Inst. Prikl. Fiz. Akad. Nauk SSSR, Gor'ki, 1983).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 979–985.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1129–1136.

PLASMA, GASES

Transport Properties of Macroparticles in Dust Plasma Induced by Solar Radiation

O. S. Vaulina, A. P. Nefedov, O. F. Petrov*, 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: ipdustpl@redline.ru*

Received December 14, 2000

Abstract—The paper deals with the results of experimental investigation of the dynamic behavior of macroparticles charged by way of photoemission, under conditions of microgravity. The experimental data have been obtained for bronze particles subjected to solar radiation in a buffer gas at a pressure of 40 Torr (Mir space station). Different procedures for determining the transport properties of macroparticles by analyzing video records of experiments are treated. The velocity distribution, the temperature, the charge, the friction coefficient, and the dust particle diffusion coefficients are found. The results of comparing the experimental and theoretical estimates demonstrate that the dynamic behavior of macroparticles under the conditions of investigations are defined by the process of their ambipolar diffusion. © *2001 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

Photoemission is one of the main mechanisms of charging dust particles under space conditions. Macroparticles several microns in size may acquire a positive charge of the order of 10^2–10^5 times the electron charge and form gas, liquid, or crystal dust structures [1–4]. The phase state of such structures is closely associated with the processes of macroparticle diffusion.

The diffusion is a nonequilibrium process caused by thermal motion of particles, which is one of the main sources of energy loss (dissipation) in plasma-dust systems. Both suspended macroparticles of matter in buffer gases (Brownian motion) and the gas molecules proper or particles of the plasma component (self-diffusion) may diffuse. In the case of plasma-dust clouds consisting of charged macroparticles, ions, and electrons, the transport properties of the system may be affected considerably by combined diffusion transfer of particles of unlike charge (ambipolar diffusion).

Most of the known methods of experimental determination of the diffusion coefficients of ions and electrons are based on the results of indirect measurements of microparticle mobility in external electric fields [5]. Such methods are unfit for use in the diagnostics of particles in plasma because they introduce considerable perturbations in the system being investigated. For determining the transport characteristics of weakly interacting macroparticles, extensive use is made of the methods of correlation spectroscopy of photons, whose range of validity is restricted to the short-range order of interparticle interaction [6, 7]. The simplest and most obvious method of diagnostics for dust systems is the analysis of mean-square shifts or of particle density gradients using video recording of experiments. The main difficulties arising in solving these problems are associated either with the need for correct identification or with the lack of a sufficient number of particles in the volume being analyzed.

This paper describes the results of investigating the transport properties of macroparticles charged by way of photoemission, under conditions of microgravity. The experimental data were obtained for bronze particles subjected to solar radiation (Mir space station, 1997).

2. EXPERIMENT

The experiments were performed aboard the Mir space station. A detailed description of the experimental setup was given in [2]. The main element of the working chamber was a glass ampoule with bronze particles coated with a monolayer of cesium. The particles were placed in a buffer gas (neon) at a pressure P of about 40 Torr. The parameters of particles are given in Table 1.

The ampoule was a cylinder, one end of which was a flat uviol window intended for illumination of the particles by solar radiation (Fig. 1a). The particles in the ampoule were additionally illuminated by a sheet laser beam ("laser knife") whose width did not exceed 200 μm. A semiconductor laser with an operating wavelength of 0.67 μm was used for this purpose. The image was recorded by a CCD camera, the signal from which was recorded on magnetic tape (frame frequency of 25 s^{-1}). The videocamera field of vision was a rectangle of approximately 8×9 mm (Fig. 1); the videocamera was set on the ampoule center, with a depth of focus for the selected diaphragm setting of 16 being approximately 9 mm (see Fig. 1a). The video records were later processed using special computer codes

Table 1. The average radius r_p of bronze particles with the density ρ and the work function W. The limiting charge Z_{max} (7) of the particles and the friction coefficients ν_{+1} (Stokes' formula) and ν_{+2} (free-molecular mode)

r_p, μm	ρ, g/cm^3	W, eV	$Z_{max}(e)$	ν_{+1}, s^{-1}	ν_{+2}, s^{-1}
37.5	8.2	1.5	6.9×10^4	7.7	7.8

enabling one to identify the shifts of individual particles in the videocamera field of vision. In the presence of intense fluxes of solar radiation, the number of particles recorded by the videocamera was defined by the depth of focus of the video system, which made it possible to follow the positions of individual particles during periods of time $t > 5\tau$ (where $\tau = \nu_+^{-1}$ is the deceleration time and ν_+ is the collision frequency of dust particles and buffer gas molecules) sufficient for analysis of the transport characteristics of the system. The number of identified particles was less than 40% of the total number of particles recorded by the videocamera at the initial moment of time.

The first stage of the experiment involved the observation of the behavior of macroparticles under conditions of microgravity in the absence of solar radiation ("dark" mode). During the time of observation (15 to 20 min), the number of particles in the field of vision of the video system did not vary appreciably. The dust concentration n_0 was determined by the particles entering the plane of the laser knife and was in the range from 200 to 400 cm^{-3}.

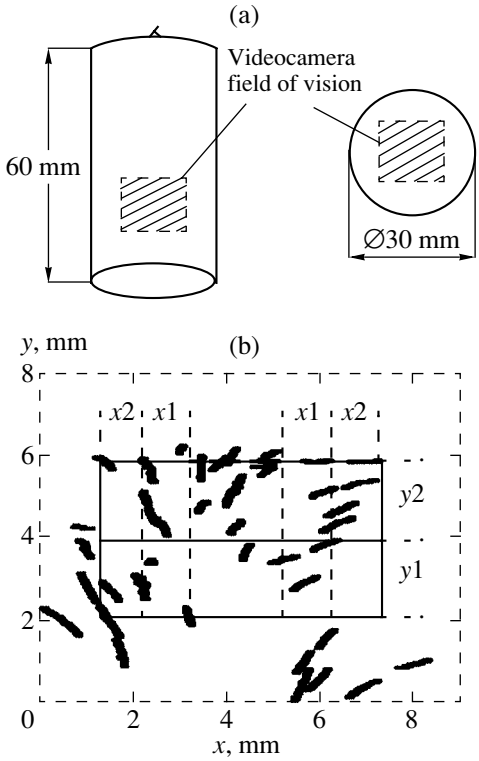


Fig. 1. (a) The geometric dimensions of the working ampoule and (b) the mechanical trajectory of macroparticles after the system is subjected to the effect of radiation.

The second stage of the experiment involved the observation of the behavior of macroparticles under conditions of illumination of the dust cloud by solar radiation. In the initial state, the bronze particles were located on the ampoule walls and, therefore, the experiments were performed in accordance with the following scheme:

(1) dynamic effect (impact) on the system with the illuminator shutters closed;

(2) exposure in darkness for a period of 2–4 s $\gg \tau$ in order to reduce the particle velocity acquired from the initial impulse (impact);

(3) illumination of the ampoule by solar radiation;

(4) relaxation of the particles to the initial state (departure to the walls), after which the illuminator shutters closed.

The time of particle departure to the ampoule walls was 3 to 5 min, this being more than three orders of magnitude less than the time of total diffusion loss of macroparticles at room temperature because of their Brownian motion. The vectors of particle velocity in the initial stage of illumination were directed chaotically. Some time (~1–3 s) after the effect of solar radiation, the particle motion assumed a clearly defined direction towards the ampoule walls. Figure 1b gives the trajectories of 40 particles after the system was subjected to the effect of radiation.

3. DETERMINATION OF TEMPERATURE AND VELOCITY SPECTRA OF DUST PARTICLES

An analysis of video records of the experiments revealed irregular variations of the magnitude and direction of the velocity of individual particles against the background of their combined drift motion. Such irregular fluctuations of the velocity of particles reflect their kinetic temperature, which, for the Maxwellian velocity spectrum (V_x, V_y), may be derived from the estimation of dispersion as [8]

$$T_{x(y)} = m_+\{\langle V_{x(y)}^2\rangle - \langle V_{x(y)}\rangle^2\}, \qquad (1)$$

where m_+ is the mass of a dust particle and $\langle\ \rangle$ describes averaging over the ensemble and in time, assuming that the system is ergodic. The quantity $\langle V_{x(y)}\rangle = V_d^{x(y)}$ is the drift velocity of regular drift of particles, against whose background the thermal motion of particles occurs. Determining the temperature from relation (1) for different samplings of particles (20 to 60) gives $T_x \approx 51$ eV and $T_y \approx 22$ eV with an accuracy of 5%, which is much higher than their room temperature $T \approx 0.03$ eV. In so doing, the recorded values of particle velocity distribution (V_x, V_y) in both the x- and y-direction were close to

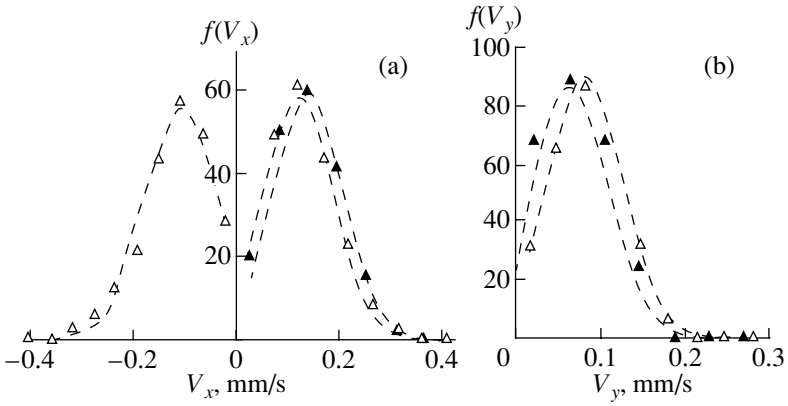


Fig. 2. Experimental spectra of velocities V_x (a) and V_y (b) for particles from various regions indicated in Fig. 1b: regions $x1$, $x2$ (△); regions $x2$, $y2$ (▲). Dashed curves describe the approximation of experimental data by the Maxwellian distribution with temperatures $T_x \approx 51$ eV (a) and $T_y \approx 22$ eV (b).

Maxwellian, with the temperature derived by formula (1). The particle velocity spectra $f(V_x)$ and $f(V_y)$ are given in Figs. 2a and 2b for different regions of the measuring volume, marked in Fig. 1b as $x1$, in Fig. 2a as $x2$, and in Fig. 2b as $y1$ and $y2$. The values of drift velocity of particles for these regions corresponded to

$$V_d^{x1} = 0.011 \text{ cm/s}, \quad V_d^{x2} = 0.0145 \text{ cm/s},$$

$$V_d^{y1} = 0.0072 \text{ cm/s}, \quad V_d^{y2} = 0.0061 \text{ cm/s}.$$

Note that the dust particles in plasma may have other than uniform distribution of irregular kinetic energy over degrees of freedom (i.e., Maxwellian spectra with $T_x \neq T_y$ are possible) and the value of kinetic temperature of macroparticles may considerably exceed that of the temperature of the gas surrounding them. These effects may be associated, for example, with fluctuations of the particle charge or with the spatial inhomogeneity of the parameters of the plasma-dust system [9–13]. The anomalous "heating" of macroparticles was repeatedly observed in experiments in studying dust structures in laboratory plasma [12–15].

4. DIFFUSION OF MACROPARTICLES

Because the plasma-dust system being treated consists of positively charged macroparticles and photoelectrons emitted by them, one can assume that the transport properties of such a system will depend on the ambipolar diffusion of particles. As a result of the considerable difference between the mobility of electrons μ_e and that of dust particles μ_+, the components of such a system will separate in the entire volume of the ampoule and a negative surface charge will arise on the walls. The arising electric field of polarization interferes with further separation of charged components. As a result, electrons and heavy particles may diffuse "together," with some effective coefficient D_a of ambipolar diffusion. The quantity D_a is defined by the diffusion coefficient of a slower component and, in the absence of magnetic and external electric fields, may be written in the form [5, 16]

$$D_a = \{D_e\mu_+ + D_+\mu_e\}/\{\mu_+ + \mu_e\}, \tag{2}$$

where D_e and D_+ are the coefficients of free diffusion of electrons and particles, respectively; in the presence of a Maxwellian velocity spectrum, the latter coefficients obey the relation

$$D_{e(+)} = T_{e(+)}/\nu_{e(+)}m_{e(+)}, \tag{3}$$

where $T_{e(+)}$, $m_{e(+)}$, and $\nu_{e(+)}$ denote the temperature, mass, and collision frequency (coefficient of friction) with buffer gas neutrals for electrons and particles. Because $\mu_e \gg \mu_+$, the coefficient of ambipolar diffusion may be written as

$$D_a \approx D_+ + D_e\mu_+/\mu_e,$$

whence, in view of formula (3), we have

$$D_a \approx (1 + ZT_e/T_+)D_+, \tag{4}$$

where Z is the macroparticle charge in elementary electron charges (e). The temperature T_e of photoelectrons leaving the particle surface depends on the particle material and, in most cases, is in the range from 1 to 2 eV [17, 18].

Note that relations (2)–(4) are valid only for the case of weakly ionized plasma-dust system, when the dissipation is defined by buffer gas neutrals and the collisions of charged components are unimportant. On the other hand, the coefficient of ambipolar diffusion describes polarization effects which are impossible in a rarefied plasma with a low density of charged components. The particle diffusion of such a plasma is defined by coefficients (3).

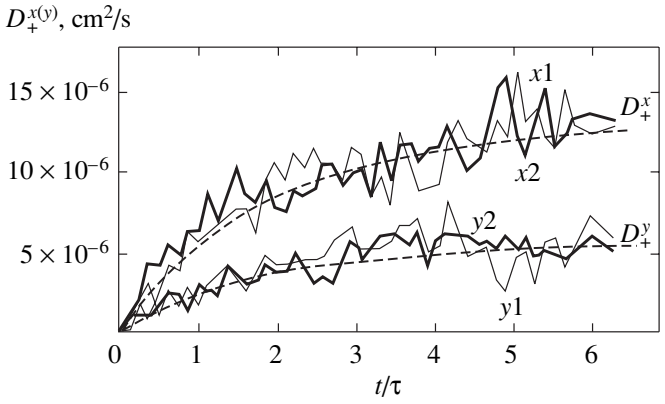


Fig. 3. The experimentally obtained dependence of $D_+^{x(y)}$ and of the results of its approximation (broken lines) by curves (6) on the characteristic parameter t/τ (on the number of deceleration times) for different regions of the measuring volume $x1$, $x2$, $y1$, and $y2$, (Fig. 1b).

5. DETERMINATION OF THE COEFFICIENTS OF FRICTION AND FREE DIFFUSION OF MACROPARTICLES

The direct determination of the diffusion coefficients of macroparticles by treating the video records of experiments is complicated by a number of difficulties, the main one of which is caused by the impossibility of identifying the motion of individual particles during a period of time sufficient for the diagnostics of their parameters. It is this fact (associated with the narrow field of vision bounded by the "laser knife" plane) that served as the reason for the unsuccessful attempt at determining the free diffusion coefficient D_+ of macroparticles in the "dark" mode. The illumination of dust cloud by solar radiation brings about an expansion of the measuring volume, which enables one to follow the motion of a fairly large number of particles for a period of time sufficient for the diagnostics of their parameters.

Given the particle temperature, the free diffusion coefficients D_+ may be estimated using relation (3). However, in spite of the close values of the friction coefficients ν_{+1} (Stokes' formula [19]) and ν_{+2} (free-molecular mode [20]), which are given by different theoretical models (see Table 1), the value of ν_+ depends, to a large measure, on the accuracy of determining the effective size of particles and on the pressure of the surrounding gas. Therefore, the data about the coefficient ν_+ call for an independent experimental verification.

The coefficients of friction ν_+ and diffusion $D_+^{x(y)}$ of macroparticles may be recovered using the results of measuring temperature and velocity $V_d^{x(y)}$ of the regular drift of particles (see Section 3),

$$D_+^{x(y)}(t) = \frac{\langle \Delta r(t)^2 \rangle - (V_d^{x(y)} t)^2}{2t}, \qquad (5)$$

where $\langle \Delta r(t)^2 \rangle$ is the mean-square shift of individual particle in the direction of the x (or y) axis, and the quantity

$$D_+^{x(y)}(t \longrightarrow \infty) = D_+^{x(y)}$$

corresponds to relation (3). The function $D_+^{x(y)}(t)$ for the thermal motion of particles may also be represented as

$$D_+^{x(y)}(t) = D_+^{x(y)}\left(1 - \frac{1 - \exp(-\nu_+ t)}{\nu_+ t}\right). \qquad (6)$$

Figure 3 gives the time dependence of $D_+^{x(y)}$ for different regions of the measuring volume ($x1$, $x2$, $y1$, and $y2$, Fig. 1b) and the results of the closest approximation of these dependences by curves (6) at $\nu_+^{\exp} = 3.1$ s^{-1} for $T_+^x \approx 51$ eV and $T_+^y \approx 22$ eV. Therefore, we can derive

$$D_+^x \approx 1.4 \times 10^{-5} \text{ cm}^2/\text{s}, \quad D_+^y \approx 6.2 \times 10^{-6} \text{ cm}^2/\text{s}$$

for the coefficients of free diffusion of particles.

6. MACROPARTICLE CHARGE

Data on the macroparticle charge Z are required for determining the ambipolar diffusion coefficient D_a from relation (4). The limiting estimate of the dust particle charge $Z_{\max} = 69000$ may be derived from the condition of equality of the surface potential ϕ_S to the quantity $h\nu_{\max} - W$ [2, 3],

$$Z_{\max} = (h\nu_{\max} - W) r_p / e, \qquad (7)$$

where $h\nu_{\max}$ is the maximal quantum energy, which, in our case, corresponds to the wavelength $\lambda_{\min} \approx 0.3$ μm and is defined by the transmission function of the experimental chamber. A more exact estimate of the charge Z may be obtained as a result of analysis of the time dependence of the relative variation of the particle concentration $n(t)/n_0$ [2]. The experimental dependences $n(t)/n_0$ are given in Fig. 4. The initial concentration of macroparticles $n_0 \equiv n(t = 0)$ was close to 1.95×10^2 cm^{-3}.

In the case when the electric field forces acting on an individual macroparticle on the side of other particles are balanced by the friction forces, the dependence $n(t)/n_0$ may be approximated by the following function [2]:

$$n(t)/n_0 = (1 + 3\omega_0^2 t/\nu_+)^{-1}, \qquad (8)$$

where $\omega_0 = \sqrt{(Ze)^2 n_0/m_+}$ is the dust frequency at $t = 0$. Relation (8) enables one to determine ω_0 by way of optimal matching of the experimental and prediction data. This may give the value of the macroparticle charge for the given initial concentration n_0. The results

of approximation of the experimental dependence $n(t)/n_0$ by function (8) is given in Fig. 4 for

$$\nu_+^{\exp} = 3.1\ \text{s}^{-1}, \quad \omega_0 = 0.2\ \text{s}^{-1}.$$

Proceeding from the obtained values of $n_0 = 195\ \text{cm}^{-3}$ and $\omega_0 = 0.2\ \text{s}^{-1}$, the particle charge may be estimated as

$$Z_{\exp} = (\omega_0/e)\sqrt{m_+/n_0} = 4.03 \times 10^4 e.$$

The results of test calculations of the dependence $n(t)/n_0$ by the method of molecular dynamics are given in Fig. 4 for a system of particles with $Z = 40\,300$ at $n_0 = 195\ \text{cm}^{-3}$ and $\nu_+ = 3.1\ \text{s}^{-1}$. The three-dimensional system of equations of motion was solved for a cylindrical ampoule (Fig. 1a) with due regard for F_{br} of the thermal motion of particles under conditions of their initial velocity being equal to zero and absorption of macroparticles on the cylinder walls,

$$m_+ \frac{d^2\mathbf{r}_k}{dt^2} = \sum_j \Phi(r)\Big|_{r=|\mathbf{r}_k-\mathbf{r}_j|} \frac{\mathbf{r}_k - \mathbf{r}_j}{|\mathbf{r}_k - \mathbf{r}_j|} - m_+\nu_+ \frac{d\mathbf{r}_k}{dt} + \mathbf{F}_{br}. \quad (9)$$

Here,

$$\Phi(r) = \frac{(eZ)^2}{r^2}$$

is the parameter of electric interaction and r is the distance between a pair of interaction particles. An analysis of the solution of Eqs. (9) for different parameters of particles has revealed that the dependence $n(t)/n_0$ is defined by the ratio between the eigenfrequency ω_0 and the friction coefficient ν_+. The thermal motion of particles has no appreciable effect on the variation of the concentration $n(t)/n_0$ with their kinetic temperatures of up to 50 eV.

One can readily see (Fig. 4) that, for times $t < 15\tau = (\nu_+^{\exp})^{-1} \approx 5$ s, bronze particles move in a mode close to that given by Eq. (8). A further decrease in the particle concentration at $t > 10$ s leads to a reduction of the forces of interparticle interaction and, accordingly, to agreement between the experimental data and the results of calculations by the method of Brownian dynamics (9) and approximation (8).

Note that the effects of polarization of unlike charges were not included explicitly in any of the numerical problems given by Eqs. (8) and (9). Nevertheless, these effects indirectly influence the value of the equilibrium charge of dust particles. The particle potential ϕ_S reaches a steady-state value as a result of balance between the electron recombination on the particle surface and the photoemission flux from the particle. In so doing, the floating potential ϕ_w of the ampoule wall surface is defined by the photoelectrons escaping from the particles. The electric field arising in the ampoule prevents some of the emitted electrons from leaving the dust cloud. The presence of photoelectrons returning to the surface of macroparticles may cause a considerable reduction of their equilibrium charge relative to the limiting value $Z_{\max}$ given by Eq. (7). For the conditions being treated ($n_0 \approx (1\text{–}3) \times 10^2\ \text{cm}^{-3}$), the determination of the charge from the balance equation gives $Z \approx (3.5\text{–}5) \times 10^4 e$ [2], which corresponds to the value of $Z_{\exp}$ obtained as a result of measurements.

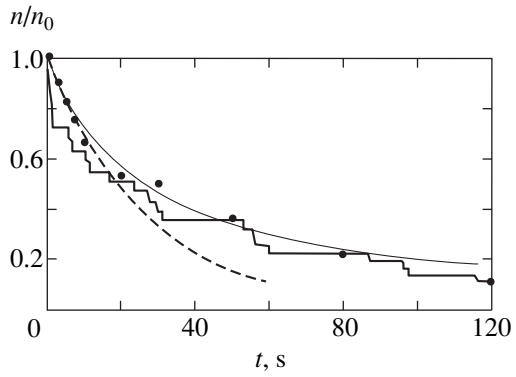


Fig. 4. The relative concentration $n(t)/n_0$ as a function of time t: points, experiment; bold line, method of molecular dynamics; fine line, approximation (8); broken line, region of ambipolar diffusion.

7. EXPERIMENTAL ANALYSIS OF AMBIPOLAR DIFFUSION OF MACROPARTICLES

One can use the measurement results given in Sections 3–6 to determine the ambipolar diffusion coefficient D_a for macroparticles from relation (4) for the known temperature of electrons in the system of dust particles emitting those electrons. The temperature T_e of electrons in the system being treated may differ from the temperature of photoelectrons leaving the particle surface. In the absence of electric fields, the electron energy relaxation time (the time during which the energy of an electron decreases by a factor of approximately 2.78) is defined as

$$\tau_u \approx \tau\delta^{-1},$$

where $\tau = \nu_+^{-1}$, δ^{-1} is the number of effective collisions and the effective relaxation length is $\Lambda_u \approx (0.8\text{–}1)\lambda\delta^{-1/2}$, λ is the mean free path of an electron with an energy equal to the initial energy [5]. For neon, $\delta = 10^{-4}$, and the relaxation distance Λ_u [cm] ≈ 10 to $12/P$ [Torr] under the pressure being treated has a value of approximately 0.25 to 0.3 cm, exceeding the average distance $l_p \approx 0.15$ to 0.18 cm between the dust particles which are the background electron sources. Hence we derive that, in the absence of electric fields, the electron energy loss over distances of about $\sim l_p/2$ will amount to approxi-

Table 2. The results of measurements of the charge Z_{exp}, the friction coefficient ν_+^{exp}, the temperature (T_x, T_y), the free diffusion coefficient (D_+^x, D_+^y), and the ambipolar diffusion coefficient D_a for bronze particles

$Z_{exp}(e)$	ν_+^{exp}, s^{-1}	T_x, eV	T_y, eV	D_+^x, cm^2/s	D_+^y, cm^2/s	D_a, cm^2/s
4.03×10^4	3.1	51	22	1.4×10^{-5}	6.2×10^{-6}	$(1.1–2.1) \times 10^{-2}$

mately 30% of the initial energy of photoelectrons at the dust particle surface. Note that, in the presence of electric fields E, an electron may further acquire an additional stochastic energy ε which is some function of the quantity $E\delta^{-1/2}$ [5]. Therefore, a rigorous determination of the electron temperature in a system of emitting dust particles calls for solving the complete kinetic equation. Therefore, in order to estimate the value of the ambipolar diffusion coefficient D_a, we will assume that the temperature of electrons in the system does not actually differ from the temperature of photoelectrons at the particle surface, whence we find, for $T_e \approx 1–2$ eV, $D_a \approx (1.07–2.14) \times 10^{-2}$ cm^2/s.

Because the loss of charges in the experiment under consideration are associated with their diffusion towards the walls, one can write an additional estimation formula allowing for the average rate of diffusion loss of dust particles in order to check the correctness of determining the ambipolar diffusion coefficient [5],

$$dn/dt = -n\nu_d \equiv -nD_a\Lambda^2, \qquad (10)$$

where ν_d is the frequency of diffusion drifts and Λ is some characteristic scale. For a cylindrical volume of radius R and length $L = 4R$, the quantity $\Lambda \approx R/2 = 0.75$ cm (Fig. 1) to an accuracy of up to the coefficient 2 [5]. A more accurate determination of Λ requires the solution of the diffusion equation for macroparticles with concrete boundary conditions. The frequency of diffusion drifts may be estimated by the rate of variation of the relative particle concentration $n(t)/n_0$ (Fig. 4). The charge polarization effects responsible for ambipolar diffusion of particles show up during comparison of the predicted and experimentally obtained $n(t)/n_0$ curves. The experimentally obtained $n(t)/n_0$ curve at $t < 10$ s agrees well with the exponential solution $n = n_0\exp(-\nu_d t)$ of Eq. (10) with $\nu_d \approx 3\omega^2/\nu_+^{exp} = 0.05$ s^{-1} (Fig. 4), from which we derive the independent estimate $-D_a = \Lambda^2\nu_d \approx 1.97 \times 10^{-2}$ cm^2/s for the coefficient of ambipolar diffusion. This result fully agrees with the estimates given at the beginning of the paragraph. This leads one to assume that the electron temperature in the system of dust particles was close to the photoelectron temperature at the particle surface.

One can readily see that the derived coefficient D_a is much higher than the free diffusion coefficient D_+ and agrees with the theoretical predictions given by Eq. (4) within the accuracy of determining characteristic diffusion distance Λ, the assumption $T_e = 1$ to 2 eV, and the errors in measuring the particle parameters (Z, T, ν_+). Therefore, one can conclude that the process of ambipolar diffusion was observed in the present experiment, i.e., that the density values of charges of both signs were high enough for a substantial space charge to form as a result of the charge separation, this space charge leading to the emergence of an electric field of polarization.

8. CONCLUSION

The objective of this study was to investigate experimentally and theoretically the transport properties of macroparticles charged by way of photoemission under the effect of solar radiation, under conditions of microgravity. In order to solve the problem set, we treated the data (video records) of experiments performed aboard the Mir space station with bronze particles in a buffer gas (neon) at a pressure of 40 Torr.

Different procedures for determining the transport characteristics of macroparticles by analyzing video records of experiments were treated. The velocity distribution, the temperature, the charge, the friction coefficient, and the diffusion coefficients for dust particles have been found (Table 2). The results of comparing the experimental and theoretical estimates have demonstrated that the dynamic behavior of macroparticles under the conditions of our investigations were defined by the process of their ambipolar diffusion. The results of analyzing the experimental results leads one to conclude that the electric interaction between particles has no considerable effect on the transport characteristics of plasma-dust systems with the parameter of interparticle interaction

$$\Gamma = (eZ)^2(4\pi n/3)^{1/3}/T < 50.$$

ACKNOWLEDGMENTS

This study was supported in part by the Russian Foundation for Basic Research (project nos. 00-02-81036 and 01-02-16658) and CRDF (project no. RP2-2250).

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. E. Fortov, A. P. Nefedov, O. S. Vaulina, *et al.*, Zh. Éksp. Teor. Fiz. **114**, 2004 (1998) [JETP **87**, 1087 (1998)].

3. M. Rosenberg and D. A. Mendis, IEEE Trans. Plasma Sci. **23**, 177 (1995).
4. V. N. Tsytovich, Usp. Fiz. Nauk **167**, 57 (1997) [Phys. Usp. **40**, 53 (1997)].
5. Yu. P. Raĭzer, *Gas Discharge Physics* (Nauka, Moscow, 1987; Springer-Verlag, Berlin, 1991).
6. A. P. Nefedov, O. F. Petrov, and S. A. Khrapak, in *Proceedings of the V International Scientific and Technical Conference on Optical Methods for Researching Streams, Moscow, 1999*, p. 235.
7. *Photon Correlation and Light Beating Spectroscopy*, Ed. by H. Z. Cummins and E. R. Pike (Plenum, New York, 1974; Mir, Moscow, 1978).
8. I. Bronshteĭn and K. Semendyaev, *Handbook of Mathematics* (Nauka, Moscow, 1986).
9. O. S. Vaulina, A. P. Nefedov, O. F. Petrov, and S. A. Khrapak, Zh. Éksp. Teor. Fiz. **115**, 2067 (1999) [JETP **88**, 1130 (1999)].
10. O. S. Vaulina, A. P. Nefedov, O. F. Petrov, *et al.*, Zh. Éksp. Teor. Fiz. **118**, 1325 (2000) [JETP **88**, 1130 (1999)].
11. O. S. Vaulina, S. A. Khrapak, A. P. Nefedov, and O. F. Petrov, Phys. Rev. E **60**, 5959 (1999).
12. V. Zhakhovskiĭ, V. Molotkov, A. Nefedov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 392 (1997) [JETP Lett. **66**, 419 (1997)].
13. H. Thomas and G. Morfill, Nature **379**, 806 (1996).
14. A. Melzer, A. Homann, and A. Piel, Phys. Rev. E **53**, 2757 (1996).
15. J. Pieper and J. Goree, Phys. Rev. Lett. **77**, 3137 (1996).
16. D. A. Frank-Kamenetskii, *Course on Plasma Physics* (Atomizdat, Moscow, 1964).
17. C. K. Goertz, Geophys. Rev. **27**, 271 (1989).
18. V. A. Grilikhis, P. P. Orlov, and L. B. Popov, *Solar Energy and Space Flights* (Nauka, Moscow, 1986).
19. D. V. Sivukhin, in *General Course of Physics* (Nauka, Moscow, 1979), Vol. 1, p. 496.
20. I. T. Yakubov and A. G. Khrapak, Sov. Technol. Rev. B **2**, 269 (1989).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 986–990.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1137–1142.

PLASMA, GASES

Microwave Discharge on a Dielectric Surface in Vacuum

M. P. Brizhinev, S. V. Golubev, D. S. Dorozhkina, B. G. Eremin, V. G. Zorin*, A. G. Litvak, I. V. Plotnikov, S. V. Razin, V. E. Semenov, A. V. Strikovskii, and O. N. Tolkacheva

Institute of Applied Physics, Russian Academy of Sciences, Nizhni Novgorod, 603155 Russia
**e-mail: zorin@appl.sci-nnov.ru*

Received December 29, 2000

Abstract—The paper describes the results of investigation of a discharge arising in vacuum on the surface of solid dielectric materials when irradiated by intense (up to 25 MW/cm^2) electromagnetic centimeter wave radiation. When the density of the microwave energy flux exceeds some threshold value depending on the target material, a discharge emerges in the vicinity of the surface. Its emergence is associated with the evaporation of the target material and the breakdown of evaporated matter. The thus forming plasma initially has the form of a thin (on the wavelength scale) layer with the electron density of the order of 10^{16} cm^{-3}. It is demonstrated experimentally that effective generation of multiply charged ions occurs in the plasma. The measured energy distribution of ions in expanding plasma agrees with the predicted distribution obtained in solving the problem on quasineutral expansion into vacuum of a localized bunch of collisionless plasma with cold ions. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Recently, a marked increase of interest has been observed in the investigations of a discharge arising in vacuum in the vicinity of the surface of solids when irradiated by intense electromagnetic radiation. This interest is due, on the one hand, to progress reached in the development of high-power microwave oscillators, which made it possible to investigate the discharge at high values of microwave radiation heretofore inaccessible (of the order of tens of megawatts per square centimeter) and, on the other hand, to the possible practical application of such discharge for modifying the surface of solids [1, 2] and for developing ion sources. It appears of interest to investigate this discharge from the standpoint of high-power electronics, because the development of a discharge at outlet windows and insulators of REB oscillators may restrict the power and duration of microwave pulse. This paper gives the results of investigation of a discharge arising on the surface of dielectric materials when irradiated by powerful quasioptical beams of electromagnetic waves; in particular, the paper contains the first experimental data pertaining to the behavior of the expansion of a multicomponent plasma and to the efficiency of generation of multiply charged ions.

2. EXPERIMENTAL RESULTS AND THEIR DISCUSSION

The experiments were performed using high-power short-pulse microwave carsinotron radiation. Radiation with the frequency of 10 GHz and pulse duration of 40 ns was formed into a quasioptical beam of linearly polarized electromagnetic waves and focused to a vacuum chamber (the intensity in the focal region reached 25 MW/cm^2). The cross-sectional area of the focal spot was 10 cm^2. The pressure in the chamber was maintained at a level of $p \approx 10^{-3}$ to 10^{-5} torr.

When dielectric materials were brought into the focal region of the microwave beam, a discharge occurred on their surface (the discharge was registered by a flash of light), with the radiation intensity at the moment of emergence of the discharge exceeding some threshold value dependent on the target material and independent of the residual gas pressure in the range employed in the experiment ($p < 10^{-3}$ torr). For example, the threshold value of the intensity during the emergence of the discharge was 20 MW/cm^2 on Teflon, 8 MW/cm^2 on glass, and 2 MW/cm^2 on Plexiglas. A photograph of the discharge is given in Fig. 1. The plasma glow was a plurality of filaments extended in the direction of the electric field of the wave. The characteristic transverse dimension of filament was 0.1–0.2 cm, the mean distance between filaments was 0.2–0.4 cm, and the length of filaments was defined by the transverse dimensions of the microwave beam and reached several centimeters (up to 10 cm).

The space-time characteristics of discharge luminescence were investigated using a high-speed electron-optical streak camera. The velocity of discharge propagation on the dielectric surface from the focal spot center in the direction of the electric field of the wave reached 10^8 cm/s, and the velocity of motion of the ionization front toward the incident microwave was $V_z \approx 3 \times 10^7$ cm/s. Figure 2 gives a characteristic optical

scan of a discharge that illustrates its propagation toward microwave radiation.

The development of a discharge at a fairly high power of radiation was accompanied by complete shielding of microwave radiation. Characteristic oscillograms of a transmitted microwave signal are given in Fig. 3. These measurements were performed with the aid of a cooled, fast-neutron, germanium microwave detector and special calorimeters. Given a high oscillator power, the transmission factor through a plasma layer decreased more than 50 times after approximately ten nanoseconds.

The observed development of a discharge may be associated only with the breakdown of evaporating matter of the dielectric target. The intrinsic absorption of microwave radiation by the dielectric is too low even for its appreciable heating. Batanov *et al.* [3] have assumed that, in the case of high intensity of microwave radiation, a secondary-emission discharge arises on the dielectric surface in vacuum, whose electrons bombard the surface to cause an appreciable increase in the electrical conductivity in the thin surface layer (the so-called induced conductivity). It was the absorption of microwave energy in this layer that apparently resulted in its heating and evaporation with subsequent breakdown of the vapors. The amount of evaporated matter in our experiments was estimated by the variation of pressure in the chamber after each discharge and by the recoil momentum acquired by the target during evaporation of matter. The pressure was measured with the aid of an open ionization lamp at several distances from the discharge at the moments of time 2, 3, and 30 ms after the termination of the microwave pulse, when the plasma no longer affected the accuracy of measurement. The pressure increment was $(4–8) \times 10^{-5}$ torr at a background pressure of 4×10^{-4} torr. The estimate of evaporated target matter for the experimental conditions varies from 6×10^{17} to 1.5×10^{18} particles per shot. The mass of evaporated matter was estimated by the recoil momentum (acquired by the target after the microwave shot and measured using a pendulum sensor) and found to agree with the estimate made by the pressure increment.

The expansion of evaporating matter into vacuum proceeds at a speed of the order of the sound velocity corresponding to the evaporation temperature. For a constant evaporation rate, this process is described by self-similar expansion wave [4],

$$N(x, t) = N_0\left(1 - \frac{(\gamma - 1)x}{(\gamma + 1)V_s t}\right)^{2/(\gamma - 1)}, \qquad (1)$$

where $N(x, t)$ describes the spatial distribution of vapor density, x is the distance from the target, t is the time from the beginning of the evaporation process, N_0 is the density of vapors at the target surface, V_s is the velocity of sound in vapors at the target surface, and γ is the adiabatic exponent in vapor treated as ideal gas. The density of the vapor flux from the target is in this case equal to the product $N_0 V_s$. For estimation, we assume that $V_s \approx 10^5$ cm/s to find that the vapor density at the target

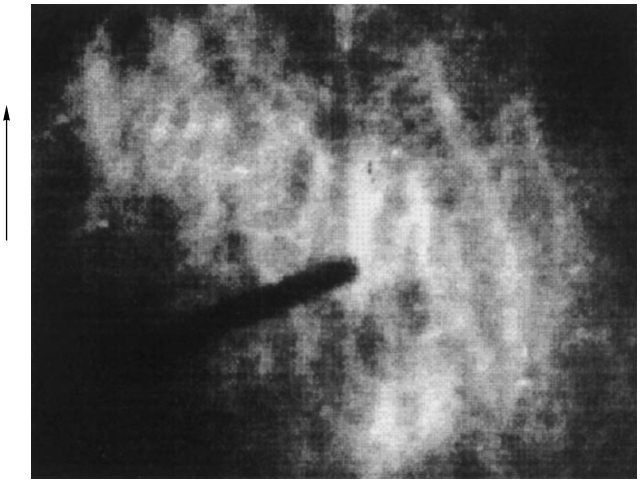


Fig. 1. A photograph of a discharge on Plexiglas, taken along the axis of a microwave beam. The dark strip at the frame center is the shadow of a Langmuir probe, and the arrow on the left indicates the direction of the microwave electric field.

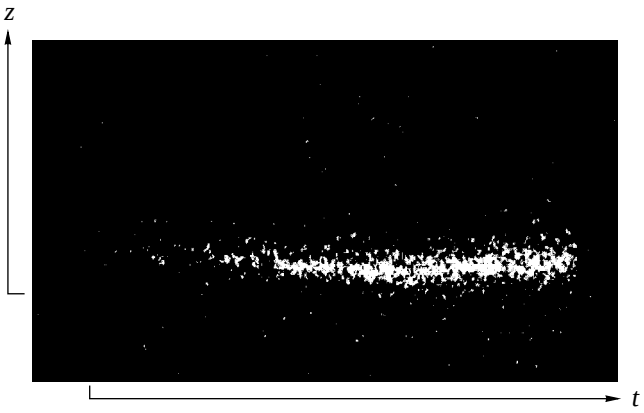


Fig. 2. An optical scan of a discharge. The z axis corresponds to the coordinate perpendicular to the target plane, and the t axis indicates time.

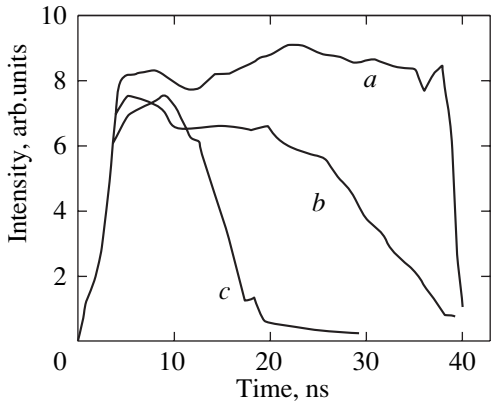


Fig. 3. Oscillograms of (a) incident and transmitted microwave radiation for different values of microwave radiation power; (b) 4 MW/cm^2, (c) 12 MW/cm^2.

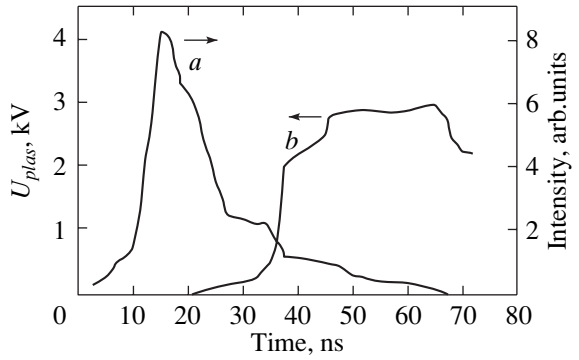


Fig. 4. Oscillograms of (*a*) microwave radiation transmitted through plasma and (*b*) plasma potential.

reached values of $(1–5) \times 10^{19}$ cm^{-3} characteristic of air at atmospheric pressure. The vapors expanded to take up, during the characteristic time of discharge development $\tau_i \approx 10^{-8}$ s, a region of thickness $l_g \approx 5 \times 10^{-3}$ cm. Within this region, the vapor density decreases rapidly away from the target, so that, even at a distance of $l_g/2$, the effective frequency ν of collisions between electrons and vapor molecules comes to be of the order of the cyclic field frequency ω. Estimates indicate that, for the intensity of microwave radiation characteristic of the experiment, the frequency ν_i of ionization by electron impact in an undisturbed field reached values of the order of 10^{10} s^{-1} in dense gas at the dielectric surface and of 10^{11} s^{-1} in the region of its maximum, where $\nu \approx \omega$. The breakdown of vapors in these conditions occurs very rapidly (during a time of less than 1 ns), and the plasma being thus formed proves to be localized initially in a thin layer with a thickness of the order of 10^{-3} cm, in which the collision frequency of electrons is of the order of the field frequency. The plasma density in this layer continues to increase until, because of its screening effect, the microwave field intensity in the vicinity of the target decreases to a level at which the ionization frequency turns out to be of the order of the characteristic frequency ν_{loss} of loss of plasma from the breakdown region. Estimates indicate that the main loss of plasma in the region of maximum of ionization is due to its ambipolar diffusion, and $\nu_{loss} \approx 10^9$ s^{-1}. This value of the ionization frequency in the region being treated is attained with a microwave electric field amplitude of the order of several kV/cm, while the undisturbed value of amplitude is approximately 100 kV/cm. The thin (on the wavelength scale) plasma layer may provide for a corresponding decrease in the field amplitude owing to reflection of incident radiation. One can use the known formulas for the reflection factor Γ of the thin plasma layer,

$$\Gamma \approx -\frac{J}{1+J}, \quad J = \frac{ic}{2}\int \frac{N_e dx}{N_c(\omega - i\nu)}, \qquad (2)$$
$$E = E_0(1+\Gamma),$$

to demonstrate that the desired decrease in the field amplitude occurs if the total number of electrons per unit area of the layer reaches a value of the order of

$$50 c N_c/\omega \approx 2 \times 10^{13} \text{ cm}^{-2}.$$

Here, c is the velocity of light in vacuum, N_e is the electron density, and N_c is the critical plasma density. Integration is performed over the entire thickness of the plasma layer; E and E_0 denote the complex amplitude of electric field in plasma and the amplitude of incident plane wave, respectively.

Therefore, the breakdown of a thin layer of vapor in the vicinity of the target must bring about, during a time of the order of 1 ns, the formation of a plasma layer 10^{-3} cm thick with an electron density of the order of 10^{16} cm^{-3}, which shields the target from incident microwave radiation. Because of shielding, the ionization processes must decelerate considerably, but the plasma layer will expand first because of diffusion and then, when the plasma extends outside of the vapor cloud, its free expansion will begin at the ionic-sound velocity, which is apparently registered by the streak camera as the ionization front motion.

Movable electric probes were used to investigate the plasma potential and the characteristics of plasma expansion. The plasma potential U_{plas} was measured by a solitary probe of high load resistance (several megaohms) placed in the vicinity of the dielectric (a typical oscillogram of the potential is given in Fig. 4). The measurement results demonstrated that the plasma potential increased rapidly after a dense plasma was formed, reached a value of several kilovolts, and was maintained at this level for a long time. The high potential of the plasma points to a high electron temperature. In all probability, this high potential is developed by the plasma on the periphery of the discharge, where the plasma density is low, and the microwave field amplitude at a distance of quarter the wavelength from the dense plasma layer (i.e., at the antinode of the standing wave being formed) may even exceed the incident wave amplitude. Therefore, the electron temperature here is maintained at a high level (of the order of oscillatory energy of electrons, which amounts to several keV).

In addition, the ions escaping from the discharge were subjected to time-of-flight and energy analysis. The measurements were performed using a five-channel ion analyzer enabling one to determine the time that ions with different energies arrived at the analyzer. The analyzer was located at a distance of approximately 3 m from the discharge. The collimator axis of this instrument coincided with the direction of the electric vector in the wave. The analyzing element of the instrument was provided by a capacitor which deflected the ions to an angle defined by their energy. The particles were further delivered to five cylindrical capacitors separating ions of certain energy and were registered by a secondary-emission multiplier (SEM). A typical oscillogram

of the SEM current $I_V(t)$ is given in Fig. 5. The peaks in the oscillogram correspond to ions whose energy per unit elementary charge was defined by channel tuning (the oscillogram in Fig. 5 corresponds to tuning the channel to the energy of singly ionized ions of 200 eV). Assuming that the ions fly from the target to analyzer uniformly for a large part of the track [5], one can further determine their velocity by the time of arrival of ions at the analyzer. Therefore, a quite certain ratio between the ion charge and its mass may be assigned to each peak in the SEM current oscillogram, which enables one to identify ions. The results of such identification are given in Fig. 5. The first (by the time of arrival) two peaks correspond here to ions of atomic and molecular hydrogen, and the next three peaks are associated with carbon ions with the charge numbers +3, +2, and +1, respectively. The results of these measurements lead one to conclude that multiply charged ions are effectively formed in the discharge plasma; the values of density of carbon ions with charges of 1, 2, and 3 are comparable.

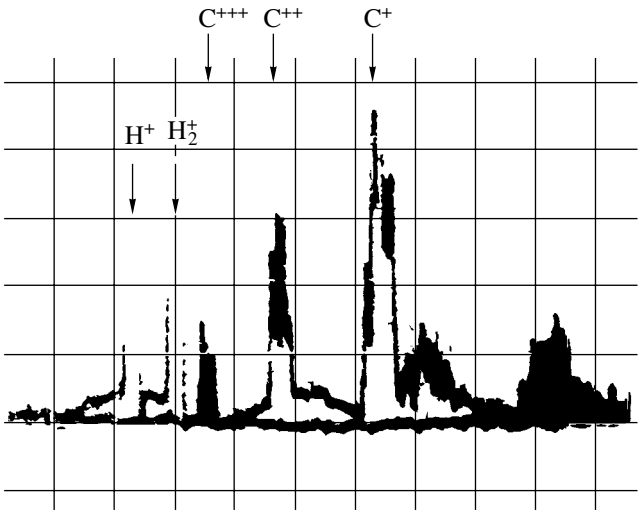


Fig. 5. An oscillogram of the SEM current of ion analyzer. The oscillograph scan, 10 μs per division.

By varying the channel tuning, one could determine the dependence of the time of arrival of ions of each type on their energy W. The measured respective dependence for singly ionized carbon ions is given in Fig. 6 (points). Also given in Fig. 6 for comparison is a prediction curve obtained assuming that ions expand with a constant velocity ($W \propto t^{-2}$). Such a dependence may also be obtained using the solution of the problem on the expansion of a localized plasmoid to vacuum [5]. The agreement between the predicted and experimentally obtained results points to the validity of the initial assumption of the inertial behavior of ion expansion that was used in the identification of ions.

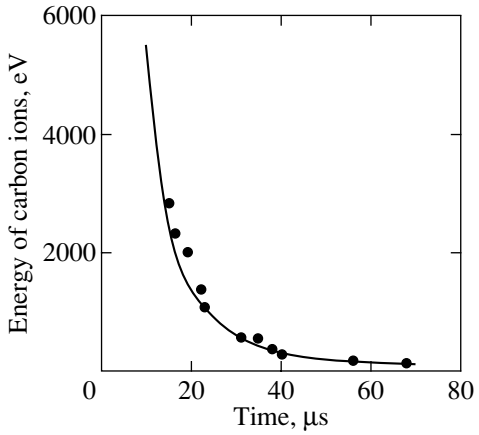


Fig. 6. The time of arrival of C^+ ions at the analyzer as a function of their energy.

Figure 7 gives the energy distribution (points) of C^+ ions registered by the analyzer and a curve corresponding to the dependence proportional to $1/W$. In the energy range $W < 3$ keV, the experimental points coincide well with this curve; in the case of high values of energy, the points lie much lower than the curve. This means that, in the above-identified range, the energy spectrum of ions arriving at the analyzer is inversely proportional to energy; at high values of energy, this spectrum decreases much more abruptly. Such an energy distribution of ions, experimentally recorded away from the plasma source, may be obtained in solving the problem on expansion of a quasineutral bunch of collisionless plasma with cold ions, if the initial velocity distribution of electrons along the collimator axis (i.e., in the direction of the electric vector in incident electromagnetic wave) is defined by the expression

$$f(V) \propto [V_0^2 - V^2]^{-1/2}, \qquad (3)$$

where V_0 is the velocity of electrons of energy $mV_0^2/2 \approx 3$ keV. It is interesting to note that, in accordance with the results of Ignat'ev and Rukhadze [6], it is the forming of just such a distribution function that one should

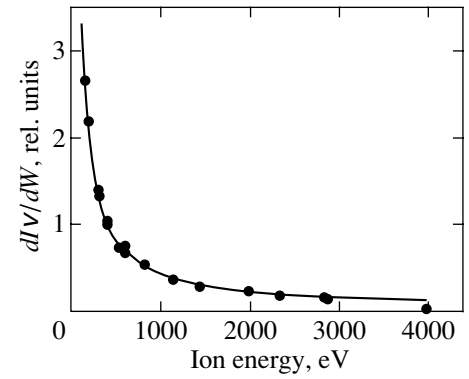


Fig. 7. The energy distribution of C^+ ions.

expect in the case of gas ionization in superstrong microwave fields in which the free oscillation energy of electrons is much higher than 100 eV. In this case, the value of V_0 is defined by the amplitude of the oscillation velocity of electrons. Under experimental conditions, the direction from the discharge plasma to the ion analyzer coincided with the direction of polarization of incident microwave radiation, and the characteristic values of the oscillatory energy of electrons reached several kiloelectron-volts. Therefore, one can assume that the ions registered by the analyzer accelerated as a result of expansion of the plasma formed on the discharge periphery where the collision frequency of electrons is low compared with the field frequency, and the amplitude of electric field is of the order of its amplitude in an incident electromagnetic wave.

3. CONCLUSION

The investigation results have demonstrated that the plasma arising in the vicinity of a dielectric target in vacuum when irradiated by intense electromagnetic radiation is characterized by a number of unique properties that may define the future uses of this plasma. The plasma turns out to be substantially nonequilibrium, with its density reaching the value of 10^{16} cm^{-3}. The volume and shape taken up by the plasma may vary depending on the shape and size of the microwave beam and the target. In the experiments described in this paper, the plasma had the form of a thin disk approximately 10 cm in diameter.

ACKNOWLEDGMENTS

This study received support from the Russian Foundation for Basic Research (project no. 98-02-17052).

REFERENCES

1. G. M. Batanov, V. A. Ivanov, M. E. Konyzhev, and A. A. Letunov, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 163 (1997) [JETP Lett. **66**, 170 (1997)].
2. L. V. Grishin, A. A. Dorofeyuk, I. A. Kossyĭ, *et al.*, Tr. Fiz. Inst. Akad. Nauk SSSR **92**, 82 (1977).
3. G. M. Batanov, V. A. Ivanov, and M. E. Konyzhev, Pis'ma Zh. Éksp. Teor. Fiz. **59**, 655 (1994) [JETP Lett. **59**, 690 (1994)].
4. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1986; Pergamon, New York, 1987).
5. D. S. Dorozhkina and V. E. Semenov, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 543 (1998) [JETP Lett. **67**, 573 (1998)].
6. A. V. Ignat'ev and A. A. Rukhadze, Fiz. Plazmy **9**, 1317 (1983) [Sov. J. Plasma Phys. **9**, 760 (1983)].

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 991–997.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1143–1150.

PLASMA, GASES

Subsonic Non-Steady-State Gas Flows in Channels with Inner Cavities

V. V. Breev, N. B. Rodionov*, and N. M. Efremov**

State Scientific Center, Troitsk Institute for Innovation and Thermonuclear Studies (TRINITI), Troitsk, Moscow oblast, 142190 Russia

**e-mail: nbrodnik@triniti.ru*

***e-mail: efremov@triniti.ru*

Received February 13, 2001

Abstract—A set of gasdynamic equations is given in the general form for matter with an arbitrary equation of state in the case when the entropy equation is used instead of the energy equation. In the ideal gas approximation in view of viscosity, a numerical investigation is performed of non-steady-state two-dimensional flows in a channel with a cavity. The calculation results have demonstrated that, given the flow velocity and the geometry of channel and cavity, pressure pulsations arise that are due to the departure of vortices from the cavity into the main flow. The values of the amplitude and frequency of pressure pulsations are determined. If measures are taken aimed at limiting the departure of vortices from the cavity, for example, a baffle is installed to restrict the interaction between the main flow and gas in the cavity, one can considerably increase the flow velocity in the channel, unaffected by the cavity. Such non-steady-state flows may be realized in MHD-generator channels, resonators of gas flow lasers, gas ducts for ventilation and gas transport systems, mufflers, whistles, etc. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

By now, a large number of experimental and prediction studies have been performed to investigate subsonic steady-state flows in smooth-walled channels of different shapes [1]. Such flows are realized in the subsonic part of a two-dimensional Laval nozzle, in vane channels of turbines and compressors, in chemical reactors for the deposition of thin films, etc. Along with this, the study of non-steady-state flows is of certain scientific and practical interest. Here, one can single out the problem of gas flow past a cavity in a channel. The solution of this problem will help explain the origination of vortices in liquid or gas flows and study in detail the generation of the so-called vortex sound. Numerous technical devices exist in which at least one wall has a closed or open cavity, which may bring about the emergence of the non-steady-state flow mode in the device. Such devices include MHD-generator channels, resonators of gas flow lasers, whistles, gas ducts for ventilation and gas transport systems, mufflers, and others. In a number of cases, for example, in pulse-periodic gas lasers during the imposition of discharge, the flow in the resonator is non-steady-state and characterized by the inleakage of flow into the cavity and by possible emergence of acoustic disturbances in the flow. The treatment of such flows will enable one to estimate the effect of the non-steady-state characteristics of flow or of the geometry of the flow passage, for example, on the parameters of radiation being generated, on the electrode processes in lasers and MHD generators, and so on. The parameters of such systems may be estimated analytically only in simple cases. The numerical calculation of such flows helps in analyzing the dynamics of disturbances in gas and liquid, calculating the parameters of acoustic disturbances, and studying the effect on these parameters of the characteristics of flow in the real geometry of the flow passage.

At present, mathematical simulation is widely used for numerical analysis of hydrogasdynamic problems; fairly extensive literature is available on the subject. The monograph by Godunov *et al.* [2] may be cited as an example. The set of equations employed includes equations of conservation of mass, momentum, and total energy and is closed by the equation of state. The energy equation may be written in several equivalent forms; however, it is usually written in terms of specific energy or temperature. In [3, 4], the equation for entropy (which has a simpler form) was used instead of the energy equation, and the density ρ (or pressure P) and the entropy S were used as the independent thermodynamic variables. This approach is more advisable because, for example, under conditions of adiabatic contractions and expansions, only ρ changes in the variables, while in the case of (ρ, T) both variables change (ρ, S). We will also adhere to this approach. Given in this paper is the set of equations in the variables (ρ, S) and (P, S) in the general case of matter with an arbitrary equation of state. Within the ideal gas model in view of viscosity, numerical simulation was performed of subsonic gas flow in a square channel with an inner cavity, simulating the flow passage of the resonator of a gas laser. It is the objective of this study

to perform a numerical investigation of non-steady-state flows in a channel with closed cavity, which may cause the emergence of acoustic vibrations.

2. EQUATIONS OF GAS DYNAMICS

The set of equations of gas dynamics includes the continuity equation

$$\frac{d\rho}{dt} + \rho \operatorname{div} \mathbf{u} = 0, \tag{1}$$

the equation of motion

$$\frac{d\mathbf{u}}{dt} + \frac{1}{\rho}\nabla P = \frac{1}{\rho}\mathbf{F}, \tag{2}$$

and the energy equation, which may be written in several equivalent forms, with the simplest form being

$$\frac{dS}{dt} = \frac{Q}{\rho T}. \tag{3}$$

Here, ρ is the density; $\mathbf{u}$ is the velocity vector; P is the pressure; T is the temperature; S is the entropy; and $\mathbf{F}$ and Q denote the density of power and thermal effect on the flow due to external sources and dissipation processes (viscosity and thermal conductivity), respectively.

The following arguments may be advanced in favor of the equation of entropy instead of energy: first, the equation for entropy is written in a simpler form, namely, only convective transfer S on space coordinates is observed in the absence of heat sources; and, second, in processes with constant entropy, the set of equations (1)–(3) reduces in fact to the solution of two equations (1) and (2). The model being treated may be especially efficient when used in a weakly compressed liquid, i.e., when the velocity of gas particles is much lower than that of sound. This enables one to study the propagation of hydrodynamic disturbances both in gas and in liquid. This set of equations will be especially simple in the case when the velocity of sound is constant and no source-related terms are present.

We will use the equation defining the velocity of sound as the equation relating the pressure and density,

$$c^2 = \left(\frac{\partial P}{\partial \rho}\right)_S. \tag{4}$$

The velocity of sound c and the quantities appearing in the set of equations when this replacement is made are calculated using the thermodynamic functions. In the general case of matter with an arbitrary equation of state, (ρ, $\mathbf{u}$, S) or (P, $\mathbf{u}$, S) may be selected as independent variables for the set of equations (1)–(4). We will discuss these two cases in more detail. Assume that (ρ, $\mathbf{u}$, S) are selected as independent variables and the internal energy E is the preassigned function of ρ and S: $E = E(\rho, S)$. Then, in view of Eq. (4), we have for the pressure gradient

$$\nabla P = c^2 \nabla \rho + \left(\frac{\partial P}{\partial S}\right)_\rho \nabla S.$$

In view of this, the set of equations (1)–(3) will be written as

$$\begin{gathered} \frac{d\varphi}{dt} + \operatorname{div} \mathbf{u} = 0, \\ \frac{d\mathbf{u}}{dt} + c^2 \nabla \varphi + \frac{1}{\rho}\left(\frac{\partial P}{\partial S}\right)_\rho \nabla S = \frac{1}{\rho}\mathbf{F}, \\ \frac{dS}{dt} = \frac{Q}{\rho T}, \end{gathered} \tag{5}$$

where $d\varphi = d\rho/\rho$. In so doing, the velocity of sound, the temperature, and the derivative of pressure with respect to entropy at constant density are determined from the thermodynamic relations

$$\begin{gathered} c^2 = \frac{\partial}{\partial \rho}\left[\rho^2 \left(\frac{\partial E}{\partial S}\right)_S\right]_S, \quad \left(\frac{\partial P}{\partial S}\right)_\rho = \frac{\partial^2 E}{\partial \rho \partial S}\rho^2, \\ T = \left(\frac{\partial E}{\partial S}\right)_\rho. \end{gathered} \tag{6}$$

In the case of ideal gas, the expression for internal energy has the form

$$E = E_0 \left(\frac{\rho}{\rho_0}\right)^{\gamma - 1} \exp\left(\frac{S - S_0}{C_V}\right), \tag{7}$$

where E_0, S_0, and ρ_0 are some reference values of energy, entropy, and density, respectively; γ is the adiabatic exponent; $C_V = R/(\gamma - 1)$; and R is the gas constant. Then, we use formulas (6) to derive expressions for temperature, velocity of sound, and the partial derivative of pressure with respect to entropy,

$$\begin{gathered} T = T_0 \left(\frac{\rho}{\rho_0}\right)^{\gamma - 1} \exp\left(\frac{S - S_0}{C_V}\right), \quad c^2 = \gamma R T, \\ \frac{1}{\rho}\left(\frac{\partial P}{\partial S}\right)_\rho = (\gamma - 1)T. \end{gathered}$$

The pressure is determined using the equation of state

$$P = \rho R T.$$

The set of equations (5) is written in the variables (φ, $\mathbf{u}$, S), which may be used for calculation in the case of a homogeneous medium. In the case of multicomponent media, it is more convenient to replace ρ by P, because the pressure is continuous on the contact surface. We will write the set of gasdynamic equations in the independent thermodynamic variables (P, $\mathbf{u}$, S) and assume that the enthalpy h is preassigned as a function of P and S. Then, the equation of state will be represented as $\rho = \rho(P, S)$. We differentiate this expression

with respect to time and take into account Eq. (4) defining the velocity of sound to derive

$$\frac{d\rho}{dt} = \frac{1}{c^2}\frac{dP}{dt} + \left(\frac{\partial\rho}{\partial S}\right)_P \frac{dS}{dt}.$$

In view of this, Eqs. (1)–(3) transform to

$$\begin{gathered}\frac{1}{c^2}\frac{dP}{dt} + \rho\,\mathrm{div}\,\mathbf{u} = -\left(\frac{\partial\rho}{\partial S}\right)_P \frac{Q}{\rho T},\\ \frac{d\mathbf{u}}{dt} + \frac{1}{\rho}\nabla P = \frac{1}{\rho}\mathbf{F},\\ \frac{dS}{dt} = \frac{Q}{\rho R}.\end{gathered} \tag{8}$$

The velocity of sound, the temperature, the density, and the derivative of density with respect to entropy at constant pressure are determined from the thermodynamic relations

$$\begin{gathered}c^2 = -\frac{1}{\rho^2}\Big/\left(\frac{\partial^2 h}{\partial P^2}\right)_S, \quad \left(\frac{\partial\rho}{\partial S}\right)_P = -\rho^2\frac{\partial^2 h}{\partial P\partial S},\\ T = \left(\frac{\partial h}{\partial S}\right)_P, \quad \frac{1}{\rho} = \left(\frac{\partial h}{\partial \rho}\right)_S.\end{gathered} \tag{9}$$

In the case of ideal gas, the expression for enthalpy has the form

$$h = h_0\left(\frac{P}{P_0}\right)^{(\gamma-1)/\gamma}\exp\left(\frac{S-S_0}{C_P}\right),$$

where h_0 is some reference value of enthalpy, and $C_P = R\gamma/(\gamma-1)$. Then, formulas (9) yield

$$T = T_0\left(\frac{P}{P_0}\right)^{(\gamma-1)/\gamma}\exp\left(\frac{S-S_0}{C_P}\right), \quad c^2 = \gamma RT,$$

$$\left(\frac{\partial\rho}{\partial S}\right)_P = -\frac{\rho}{R}.$$

The equation of state enables us to determine the density as

$$\rho = P/RT.$$

In Eqs. (5) and (8), the equations of motion and continuity without sources form a set of equations of nonlinear acoustics. Energy equation (3) may be solved independently if the velocity of sound is known at every space point or if a relation is available that relates the velocity of sound to hydrodynamic quantities (for example, the Poisson adiabat).

3. FORMULATION OF THE PROBLEM

We treat, within the framework of the equations given above, a two-dimensional problem of gas flow in a square channel with a closed cavity in a side wall. The singularity of the effects observed in this case consists in that, on the one hand, the cavity is an acoustic resonator with eigenfrequencies ν_i and, on the other hand, vortices may arise and be carried away by the flow of gas or liquid past that cavity. The periodic departure of the formed vortices from the cavern will lead to pressure pulsations. The eigenfrequencies of the acoustic resonator are defined by its characteristic dimensions l_r and velocity of sound c and may be represented in the general form as [4]

$$\nu_i = \frac{c}{l_r}\phi\left(\frac{c}{l_r}\alpha_i\right),$$

where $\phi(c\alpha_i/l_r)$ is some numerical coefficient and α_i is the damping factor for vibrations with the frequency ν_i. The frequency of pressure pulsations, governed by the periodic departure of vortices arising in the cavity, is proportional to the quantity VF/l_r, where F is some coefficient depending on geometry. This system may be treated as a self-oscillatory nonlinear system [4]; i.e., the resonator eigenfrequencies and the formation of a vortex structure in the flow are interrelated. We will treat such a system within the ideal-gas model. For this purpose, we will write the set of gasdynamic equations (5) in view of Eqs. (6) and (7) in the form

$$\begin{gathered}\frac{d\varphi}{dt} + \mathrm{div}\,\mathbf{u} = 0,\\ \frac{d\mathbf{u}}{dt} + c^2\nabla\varphi = \frac{\mu}{\rho}\Delta\mathbf{u} + \frac{1}{3}\frac{\mu}{\rho}\nabla(\mathrm{div}\,\mathbf{u}),\\ \frac{dS}{dt} = \frac{\lambda}{\rho}\frac{\Delta T}{T} + \frac{Q}{\rho T},\end{gathered} \tag{10}$$

where Q *i*s the heat released by the work of viscous forces, μ is the viscosity coefficient, and λ is the thermal conductivity coefficient for gas. We will restrict ourselves to the case in which the gas flow velocity is much lower than the velocity of sound c_0. In this case, one can ignore the heat released under conditions of viscous dissipation. We will likewise ignore the thermal conductivity. The term related to viscosity is determining in a very thin layer. For the characteristic geometric dimensions and gas dynamic parameters treated by us, the thickness of laminar boundary layer was $\delta \approx 0.003$ m [5]. In view of these assumptions, the set of equations (10) will be written as

$$\begin{gathered}\frac{d\varphi}{d\tau} + \frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\beta V_x}{x} = 0,\\ \frac{dV_x}{d\tau} + \frac{\partial\varphi}{\partial x} = \frac{\nu}{c_0}\left(\frac{\partial^2 V_x}{\partial x^2} + \frac{\partial^2 V_x}{\partial y^2}\right)\\ + \frac{\nu}{3c_0}\left(\frac{\partial^2 V_x}{\partial x^2} + \frac{\partial^2 V_y}{\partial y\partial x}\right),\end{gathered} \tag{11}$$

$$\frac{dV_y}{d\tau} + \frac{\partial \varphi}{\partial y} = \frac{\nu}{c_0}\left(\frac{\partial^2 V_y}{\partial x^2} + \frac{\partial^2 V_y}{\partial y^2}\right)$$

$$+ \frac{\nu}{3c_0}\left(\frac{\partial^2 V_y}{\partial y^2} + \frac{\partial^2 V_x}{\partial x \partial y}\right), \quad \frac{\partial S}{\partial t} = 0,$$

where $\beta = 0, 1$ for plane and cylindrical geometry, respectively; V_x and V_y are the components of the velocity vector **u** along the x and y axes, respectively, normalized to the sound velocity c_0; and $\nu = \eta/\rho$ is the kinematic viscosity. The variable $\tau = c_0 t$ is introduced instead of time. The set of equations (11) is used to perform a numerical analysis of the dynamics of nonstationary interaction of gas flow in a square channel with a cavity. We will write Eqs. (11) in a compact matrix form,

$$\frac{d\Phi}{d\tau} + A\Phi + \psi = 0, \tag{12}$$

where

$$\Phi = \begin{Bmatrix} \varphi \\ V_x \\ V_y \end{Bmatrix}, \quad A = \begin{vmatrix} 0 & \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \\ \frac{\partial}{\partial x} & 0 & 0 \\ \frac{\partial}{\partial y} & 0 & 0 \end{vmatrix},$$

$$\Psi = \begin{vmatrix} \frac{\beta V_x}{x} \\ \frac{\nu}{c_0}\Delta V_x + \frac{\nu}{3c_0}\frac{\partial \operatorname{div} \mathbf{V}}{\partial x} \\ \frac{\nu}{c_0}\Delta V_y + \frac{\nu}{3c_0}\frac{\partial \operatorname{div} \mathbf{V}}{\partial y} \end{vmatrix}.$$

The initial conditions are preassigned in the form

$$\Phi(x, y, 0) = \Phi_0(x, y).$$

The form of the boundary conditions depends on the concrete formulation of the problem.

4. NUMERICAL MODEL

The total approximation technique [9] is used to find the numerical solution of Eq. (12). Two problems are solved successively at the step $\Delta\tau$,

$$\frac{\partial \Phi_x}{\partial \tau} + A_x \frac{\partial \Phi_x}{\partial x} + \Psi_x = 0, \tag{13}$$

$$\frac{\partial \Phi_y}{\partial \tau} + A_y \frac{\partial \Phi_y}{\partial y} + \Psi_y = 0, \tag{14}$$

where

$$\Phi_x = \begin{Bmatrix} \varphi \\ V_x \end{Bmatrix}, \quad \Phi_y = \begin{Bmatrix} \varphi \\ V_y \end{Bmatrix},$$

$$A_x = \begin{vmatrix} V_x & 1 \\ 1 & V_x \end{vmatrix}, \quad A_y = \begin{vmatrix} V_y & 1 \\ 1 & V_y \end{vmatrix},$$

$$\Psi_x = \begin{vmatrix} \frac{\beta V_x}{x} \\ \frac{\nu}{c_0}\left(\Delta V_x + \frac{\partial \operatorname{div} V}{3\partial x}\right) \end{vmatrix},$$

$$\Psi_y = \begin{vmatrix} 0 \\ \frac{\nu}{c_0}\left(\Delta V_y + \frac{\partial \operatorname{div} V}{3\partial y}\right) \end{vmatrix}.$$

For Eqs. (13) and (14), use is made of the explicit two-step difference scheme of [6–8] modified for this problem,

$$\tilde{\Phi}^{\tau+\Delta\tau}_{xk\pm1} = \Phi^{\tau}_{xk\pm1} \pm B^{\tau}_{xk\pm1}\frac{\Phi^{\tau}_{xk\pm1} - \Phi^{\tau}_{xk}}{\Delta x}\Delta\tau + \Psi^{\tau}_{xk}\Delta\tau,$$

$$(\tilde{\Phi}^{\tau+\Delta\tau}_{yk\pm1} = \Phi^{\tau}_{yk\pm1} \pm B^{\tau}_{yk\pm1})\frac{\Phi^{\tau}_{yk\pm1} - \Phi^{\tau}_{yk}}{\Delta x}\Delta\tau + \Psi^{\tau}_{yk}\Delta\tau,$$

where

$$B^{\tau}_{xk\pm1} = \begin{vmatrix} \frac{V_{xk\pm1} + V_{xk}}{2} & 1 \\ 1 & \frac{V_{xk\pm1} + V_{xk}}{2} \end{vmatrix},$$

$$B^{\tau}_{yk\pm1} = \begin{vmatrix} \frac{V_{yk\pm1} + V_{yk}}{2} & 1 \\ 1 & \frac{V_{yk\pm1} + V_{yk}}{2} \end{vmatrix},$$

$$\Psi^{\tau}_{xk} = \begin{vmatrix} \frac{\beta V_x}{x} \\ \nu\left(\Delta V_x + \frac{\partial \operatorname{div} \mathbf{V}}{3\partial x}\right)_k \end{vmatrix},$$

$$\Psi^{\tau}_{yk} = \begin{vmatrix} 0 \\ \nu\left(\Delta V_y + \frac{\partial \operatorname{div} \mathbf{V}}{3\partial y}\right)_k \end{vmatrix},$$

$$\tilde{\tilde{\Phi}}_{xk}^{\tau+\Delta\tau} = \Phi_{xk}^{\tau} + A_{xk}^{\tau+\Delta\tau}\frac{\tilde{\Phi}_{xk+1}^{\tau+\Delta\tau} - \tilde{\Phi}_{xk-1}^{\tau+\Delta\tau}}{2\Delta x}\Delta\tau + \tilde{\Psi}_{xk}^{\tau+\Delta\tau}\Delta\tau,$$

$$\tilde{\tilde{\Phi}}_{yk}^{\tau+\Delta\tau} = \Phi_{yk}^{\tau} + A_{yk}^{\tau+\Delta\tau}\frac{\tilde{\Phi}_{yk+1}^{\tau+\Delta\tau} - \tilde{\Phi}_{yk-1}^{\tau+\Delta\tau}}{2\Delta x}\Delta\tau + \tilde{\Psi}_{yk}^{\tau+\Delta\tau}\Delta\tau,$$

$$A_{xk}^{\tau+\Delta\tau} = \begin{vmatrix} V_{xk} & 1 \\ 1 & V_{xk} \end{vmatrix}, \quad A_{yk}^{\tau+\Delta\tau} = \begin{vmatrix} V_{yk} & 1 \\ 1 & V_{yk} \end{vmatrix}.$$

The expressions for $\tilde{\Psi}_{xk}^{\tau+\Delta\tau}$ and $\tilde{\Psi}_{yk}^{\tau+\Delta\tau}$ are similar to those for Ψ_{xk}^{τ} and Ψ_{yk}^{τ}, but are constructed with due regard for the form of $\tilde{\Phi}_{xk\pm1}^{\tau+\Delta\tau}$ and $\tilde{\Phi}_{yk\pm1}^{\tau+\Delta\tau}$.

The foregoing scheme is of the second order of accuracy with respect to the space coordinates and of the first order with respect to time [9]. The conditions of stability of the scheme are valid with a margin.

We will calculate the boundary conditions using Eqs. (1) and (2) written in the characteristic form,

$$\begin{aligned} \frac{d\varphi}{d\tau} \pm \frac{dV_x}{d\tau} + \frac{V_x\beta}{x} &= 0, \quad \frac{dx}{d\tau} = V_x \pm 1, \\ \frac{d\varphi}{d\tau} \pm \frac{dV_y}{d\tau} + \frac{V_y\beta}{y} &= 0, \quad \frac{dy}{d\tau} = V_y \pm 1. \end{aligned} \tag{15}$$

For visualizing the velocity field, we will calculate the quantities

$$\bar{V}_x = \frac{V_x}{\sqrt{V_x^2 + V_y^2}}, \quad \bar{V}_y = \frac{V_y}{\sqrt{V_x^2 + V_y^2}}$$

and construct vector diagrams of the vector field ($\bar{V}_x$, $\bar{V}_y$).

By way of a test, a calculation was performed of one-dimensional flat, cylindrical, and spherical sound pulses on the condition that the velocity of sound is much higher than the rate of displacement of gas particles. Such calculations were performed in accordance with the procedure described in [6]. The results obtained (gasdynamic parameters of pulses) were compared with analytical expressions [10], and the laws of conservation in this case were checked for validity. In order to check the computer codes, the flow rate in the inlet and outlet cross sections was monitored. The calculation results demonstrated that, at $\tau \geq 20$, the values of flow rate agreed within 1%. The viscous steady-state plane-parallel flow in the channel was calculated as well, and a parabolic velocity profile was obtained to a good accuracy [12].

5. NUMERICAL CALCULATION RESULTS

Within the problem set, we will treat the flow of air in a square channel with a cavity in a side wall, which simulates the flow passage of a gas laser. A schematic diagram of such a device is given in Fig. 1. The following geometric dimensions and parameters of flow and

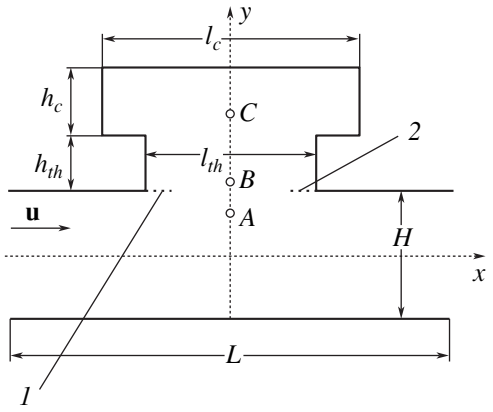


Fig. 1. The channel and cavity scheme: L and H, the channel length and width, respectively; l_c and h_c, the length and width of the resonator cavity, respectively; l_{th} and h_{th}, the length and width of the throat connecting the channel and resonator cavity, respectively. The broken line indicates two baffles (*1*, *2*). The circles indicate the points *A*, *B*, and *C*, where the pressure pulsations are calculated.

cavity were selected for calculation: the channel length along the flow $L = 0.3$ m, the channel width $H = 0.1$ m, the cavity length $l_c = 0.13$ m, the cavity height $h_c = 0.04$ m, and the length and height of the throat connecting the channel and the resonator cavity $l_{th} = 0.1$ m and $h_{th} = 0.03$ m. The pressure P_{out} at the channel outlet was taken to be 0.1 atm with a temperature of 300 K. The inlet velocity was varied from 10 to 100 m/s. The calculations were performed in a square mesh grid. The step in space was assumed to be $\Delta h = 0.002$ m, and the step in time $\Delta\tau = C_C\Delta h$, where C_C is the Courant number. It was assumed that $C_C = 0.9$ [8, 9]. The calculation was performed during a period of time $\tau \approx 0$–0.02 s. In the inlet cross section, the velocity was preassigned, and in the outlet cross section, the pressure. With $t = 0$ in the calculation region, zero velocity was preassigned, and the pressure was assumed to be equal to the outlet pressure. The flow relaxed to a steady state in periods of time of about L/V. The transition process caused pressure pulsations which either decayed or did not decay. It has been assumed that, if the pulsation amplitude continues to decrease over periods of time $\tau \gg L/V$, such a vibrational process is decaying, and conversely, if the amplitude did not change with time, pressure pulsations occur in the flow. Also performed for monitoring were calculations during a period of time considerably exceeding L/V (≈0.6 s), whose results demonstrated the presence or absence of pressure pulsations, depending on the selected parameters.

The calculation results demonstrate that, in the entire range of the treated parameters of the problem, the pressure oscillation with a frequency $\nu \approx 800$ Hz arises during the initial period in the cavity over its entire depth, with the amplitude of this oscillation reaching a value of 8% pressure at the channel outlet.

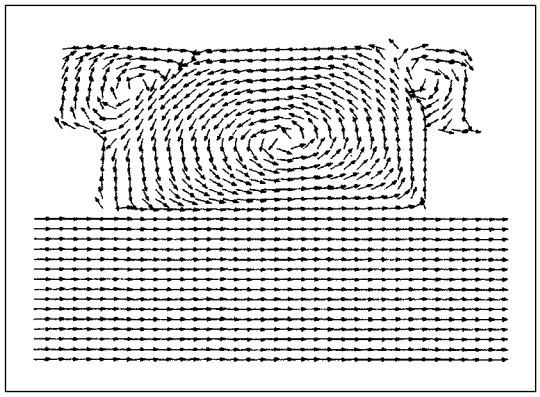


Fig. 2. A fragment of the pattern of velocity field ($\bar{V}_x$, $\bar{V}_y$) in the cavity and channel with the inlet velocity of flow of 20 m/s.

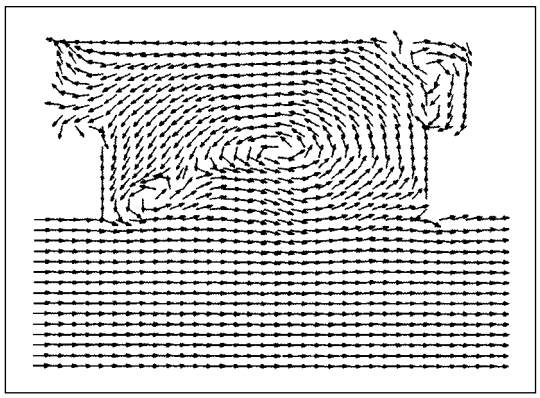


Fig. 3. A fragment of the pattern of velocity field ($\bar{V}_x$, $\bar{V}_y$) in the cavity and channel with the inlet velocity of flow of 60 m/s.

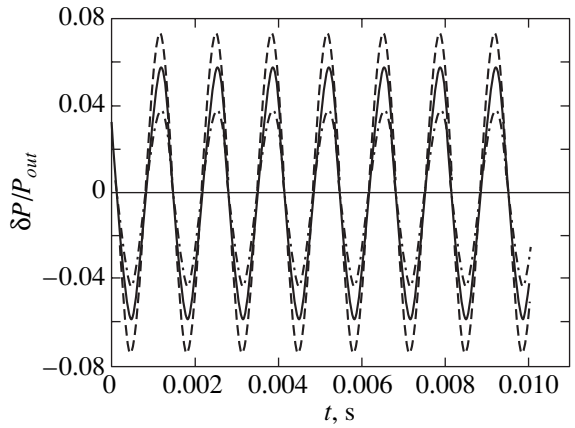


Fig. 4. The pressure pulsations ($\delta P/P_{\text{out}}$) as a function of time, with the inlet velocity of 60 m/s, at the points *A*, *B*, and *C* (Fig. 1). Dot-and-dash curve, at point *A*; solid curve, at point *B*; dashed curve, at point *C*.

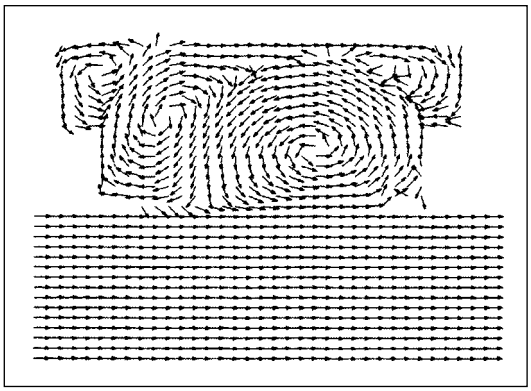


Fig. 5. A fragment of the pattern of velocity field ($\bar{V}_x$, $\bar{V}_y$) in the cavity and channel in the presence of baffles on the side of undisturbed flow. The inlet velocity of flow, 60 m/s.

In the course of time, depending on the velocity of flow in the channel, this oscillation may either decay (at $V \leq$ 40 m/s) or its amplitude reaches some steady-state value, and this oscillation is imparted to the main flow ($V \geq 60$ m/s). Figures 2 and 3 give fragments of the pattern of velocity field ($\bar{V}_x$, $\bar{V}_y$). An analysis of the velocity field in the channel and cavity reveals that a vortex forms in the cavity. With a main flow velocity of up to 40 m/s, this vortex is in the cavity and its center is stationary (Fig. 2). When the velocity increases from 40 to 60 m/s, the vortex moves in the cavity; simultaneously, much smaller vortices are formed in the cavity. The vortices, which form in the vicinity of the flow, are carried periodically one by one into the main flow; therefore, as the velocity increases, the decay of pressure pulsations becomes weaker. When the inlet velocity reaches a value of about 60 m/s, all of the vortices formed in the cavity depart periodically one by one into the main flow. In so doing, the secondary vortex that forms at the left-hand wall of the throat is comparable in magnitude to the initial vortex (Fig. 3). From this moment on, the pressure oscillation ceases to decay. The flow in the channel starts "making noise." The calculation results demonstrate that the period of repetition of the velocity field pattern corresponds to a frequency of approximately 800 Hz. One can assume that the periodic departure of vortices from the cavity maintains the pressure oscillation. Figure 4 gives the patterns of pressure distribution as a function of time at the points *A*, *B*, and *C* (Fig. 1). One can see that the oscillation occurs with a frequency of ≈800 Hz. The calculations performed with a finer mesh grid produced the same results.

We performed calculations of flow as a function of the geometric dimensions of the cavity. It has been found that an increase in the cavity height h_c by approx-

imately 20% causes very little change in the frequency but leads to a decrease in the value of critical velocity at which the pressure pulsations cease to decay. In our case, this critical velocity is 20 m/s. Analogously, a 20% increase in the width of the resonator cavity leads to a decrease in the critical velocity by half.

The calculation results have demonstrated that one can substantially reduce the pressure pulsations if a thin baffle *1* 1 cm in size or less is installed between the cavity and channel on the side of undisturbed flow at the beginning of the cavity (see Fig. 1). As a result, the baffle stabilizes the initial flow in the cavity and prevents the origination of vortices in the vicinity of the main flow in the cavity. Therefore, the velocity at which the pressure oscillation decays increases to 100 m/s (no calculations were performed at higher values of velocity). The pattern of velocity field in this case is similar to that given in Fig. 2. The provision of yet another baffle (shown at *2* in Fig. 1) at the end of the cavity no longer affects the behavior of flow in the channel; however, two conjugate vortices are formed in the cavity (Fig. 5).

Note that, because the cavity dimensions are not small compared with the region under investigation, it would be more correct to treat the entire region as a complex acoustic resonator, when the channel length along the flow L and the channel width H affect the critical velocity and the frequency of pressure oscillation. Therefore, the obtained results are qualitative and, for a concrete device, calculations must be made with due regard for concrete geometry.

This procedure is also suitable for investigating a non-steady-state flow of liquid in a channel with a cavity, if the equation of state for ideal gas is replaced by the respective equation for liquid in the form of Tate's formula [11].

6. CONCLUSION

In this paper, a set of gasdynamic equations is given in the general form with an arbitrary equation of state for the case when the entropy equation is used instead of the energy equation. In the ideal gas approximation in view of viscosity, a numerical investigation is performed of a non-steady-state two-dimensional subsonic flow in a channel with a cavity, simulating the flow passage of the resonator of a gas laser. It has been found that, in such cases, given some velocity of incident flow, pressure pulsations may arise in this flow that are defined by the characteristics of undisturbed flow and by the geometry of the flow passage. The frequency and amplitude of these pulsations have been analyzed. It has been found that a 20% variation of the geometric dimensions of the cavity causes the critical velocity, at which acoustic disturbances occur, to vary by a factor of two. The calculation results have demonstrated that the pressure oscillation is associated with the departure of vortices arising in the cavity and that the frequency of this oscillation coincides with the frequency of departure of vortices from the cavity. The provision of baffles in the closed cavity in the vicinity of the flow may considerably reduce the effect of the cavity on the main flow and increase the critical velocity of flow at which pressure pulsations arise in the channel.

REFERENCES

1. U. G. Pirumov and G. S. Roslyakov, *Gas Flow in Nozzles* (Mosk. Gos. Univ., Moscow, 1978).
2. S. K. Godunov, A. V. Zabrodin, M. Ya. Ivanov, *et al.*, *Numerical Solution of Multidimensional Problems of Gas Dynamics* (Nauka, Moscow, 1976).
3. K. P. Stanyukovich, *Non-Steady-State Flows of Continuous Medium* (Nauka, Moscow, 1971).
4. D. I. Blokhintsev, *Acoustics of a Nonuniformly Moving Medium* (Nauka, Moscow, 1981).
5. M. E. Deĭch, *Technical Gas Dynamics* (Gosénergoizdat, Moscow, 1953).
6. V. V. Breev, L. G. Kurkina, and N. B. Rodionov, Preprint No. 5352/7, Inst. At. Énerg. im. I. V. Kurchatova (Kurchatov Institute of Atomic Energy, Moscow, 1991).
7. V. V. Breev, L. G. Kurkina, and N. B. Rodionov, Preprint No. 5532/16, Inst. At. Énerg. im. I. V. Kurchatova (Kurchatov Institute of Atomic Energy, Moscow, 1992).
8. V. V. Breev and O. I. Pechenova, Preprint No. 4022/12, Inst. At. Énerg. im. I. V. Kurchatova (Kurchatov Institute of Atomic Energy, Moscow, 1984).
9. A. A. Samarskiĭ, *The Theory of Difference Schemes* (Nauka, Moscow, 1977).
10. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1988; Pergamon, New York, 1987).
11. R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves* (Interscience, New York, 1948; Inostrannaya Literatura, Moscow, 1960).
12. H. Schlichting, *Boundary Layer Theory* (McGraw-Hill, New York, 1955; Inostrannaya Literatura, Moscow, 1956).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 92, No. 6, 2001, pp. 998–1003.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 6, 2001, pp. 1151–1158.

PLASMA, GASES

Analysis of Characteristic X-ray Generation Induced by Laser Plasma Electrons Accelerated by an Electric Field

A. A. Erokhin, A. S. Kishinets, Yu. V. Korobkin, I. V. Romanov, V. M. Romanova, A. A. Rupasov*, and A. S. Shikanov

Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 117924 Russia
e-mail: rupasov@mail1.lpi.ac ru

Received February 13, 2001

Abstract—The results of experiments devoted to the study of spectral, spatial, and time characteristics of a spectrally bright point x-ray source based on a vacuum diode with a laser-plasma cathode and a titanium needle anode with a photon energy approximately equal to 4.5 keV are presented. The experimental results revealed a considerable difference between the electron emission from laser plasma in a strong electric field and the explosive electron emission and demonstrated the effectiveness of laser plasma as an electron source. The optimization of the laser radiation power density, the accelerating voltage, and the interelectrode spacing made it possible to create a point x-ray source whose spectral brightness exceeds available sources in the class of small-size pulse x-ray instruments (tubes with explosive cathodes). It has been proved experimentally that the maximum contrast of the characteristic lines of the anode material is attained in the case of an optimal choice of accelerating voltage. The x-ray source has the following parameters: (1) spectral brightness of the *K*-lines of titanium of the order of 10^{21} photons/cm^2 s sr keV; (2) emitting region size of 250 μm; and (3) laser pulse duration less than 20 ns. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The microscopy of stationary and moving objects with high gradients of density and refractive index (which include crystalline structures, high-temperature plasma, and biological media) requires the use of pulsed, spectrally bright, and high-contrast sources of x-ray radiation. Nowadays, x-ray tubes with explosive cathodes, as well as plasmas produced by a high-power laser radiation, are widely used for generating short x-ray pulses. Each of these sources undoubtedly has its own advantages and drawbacks. The broad spectrum of laser plasma radiation, the considerable amount of power required for attaining the high electron temperature, and the displacement of the radiation maximum to the range of a few kiloelectronvolts hamper the creation of a small-size, high-efficiency, and contrast source of x-ray quanta. In x-ray tubes with an explosive electron emission, the lower working voltage threshold is not lower than 50 kV, which deteriorates the contrast of the characteristic line emitted by elements with a small atomic number. Stable explosive electron emission is impossible in the range of accelerating voltages ≤20 kV; this complicates the synchronization of the source with diagnostic equipment. For this reason [1–4], one of the promising trends in creating a small-size, spectrally bright, and contrast x-ray source is the study of a vacuum diode with a laser-plasma cathode, which emits radiation in the wavelength range determined by the anode material. X-ray diodes with laser-plasma cathodes have the following advantages.

(1) The electron temperature, concentration, and expansion velocity of a laser plasma, which exceed the corresponding parameters of a cathode plasma formed during explosive electron emission by at least an order of magnitude (under similar conditions for applied voltage and electrode geometry) determine the high emissivity of a laser-plasma cathode, thus increasing the total discharge current and the spectral brightness of the source.

(2) A high emissivity of a laser-plasma cathode is observed even for low values of working voltage. For example, for $U = (3–4)U_0$ (U_0 is the excitation threshold of the characteristic radiation from the anode material), the ratio of the characteristic radiation intensity to the bremsstrahlung intensity attains its peak value [5]. As the voltage increases, the radiation yield in the *K*-lines attains saturation or even starts to decrease [6]. At the same time, the bremsstrahlung yield only increases with the voltage. Consequently, the maximum contrast of radiation in the *K*-lines is attained by applying the optimum voltage used in our experiments.

(3) The possibility of operating in the point source mode is ensured by a needle-shaped anode.

(4) The control of the accelerating voltage and the use of various materials for the anode make it possible to obtain high-contrast x-ray line emission in a fairly wide spectral range.

(5) Finally, such diodes are distinguished by stability of the output parameters, simplicity of phasing with diagnostic equipment, and low operation costs.

1063-7761/01/9206-0998$21.00

2. ESTIMATION OF THE EMISSIVITY OF A LASER-PLASMA CATHODE

When laser radiation with a low power density (the characteristic fluxes used in our experiments are 10^{10}–10^{12} W/cm^2) acts on a plane target, the characteristics of the formed plasma may be estimated to a high degree of accuracy using self-similarity [7]. For such fluxes, the laser energy is absorbed in the plasma corona owing to the backward bremsstrahlung mechanism. In this case, the electron temperature of the plasma in the absorption region is defined as

$$T\ [\mathrm{eV}] = 5\left(\frac{P\lambda}{10^{10}}\right)^{4/9}\left(\frac{A}{2Z}\right)^{2/9}(\bar{Z}_i L\Lambda)^{2/9},$$

where P is the laser radiation density at the target in W/cm^2, λ is the laser wavelength in micrometers, A and Z are the atomic mass and number of the target, $\bar{Z}_i$ is the average charge of plasma ions, Λ is the Coulomb logarithm, and L is the scale of density inhomogeneity in micrometers, corresponding to the diameter of the focal spot of the laser beam (in the case of a plasma pulse of a nanosecond duration).

Thus, for $P = 10^{10}$ W/cm^2 and $\lambda = 1$ μm, the temperature of the laser plasma at the aluminum target is 30–35 eV. A further expansion of the plasma cloud leads to its adiabatic cooling, and the electron temperature at a distance of several millimeters from the target is $T \le 10$ eV. The total number of particles in the plasma is given by [8]

$$N = \frac{E_L(\gamma - 1)}{2(1 + \bar{Z}_i)kT},$$

where E_L is the laser radiant energy, γ is the adiabatic exponent equal to 1.3–1.4, k is the Boltzmann constant, and the coefficient 1/2 takes into account the kinetic energy of the plasma. For the degree of ionization $\bar{Z}_i$ = 3–4, the total number of electrons in the plasma may be as large as 3×10^{15}.

If we apply voltage U to a spherical plasma cloud, the number of electrons escaping from the cloud due to the acceleration of charges in the external electric field is such that their charge $Q = UR$, where R is the radius of the cloud. Coulomb forces prevent a further emission of electrons from the cloud. The current increases due to an increase in the number of electrons being accelerated:

$$\frac{dQ}{dt} = U\frac{dR}{dt} = Uv,$$

where v is the plasma expansion velocity. This relation shows that for a constant accelerating voltage, the duration of the x-ray pulse is determined by the laser plasma expansion velocity. For $R = 2$ mm (half the anode–cathode distance) and $U = 20$ kV, the charge is $Q = 13.3$ CGS units, which is equivalent to 3×10^{10} electrons. In this connection, only a small fraction (~10^{-5}) of the total number of electrons present in the plasma participate in the generation of x-rays at the anode. It should be noted, however, that the above estimate is rough. In order to construct a more comprehensive model, we must consider a number of factors, which are associated with

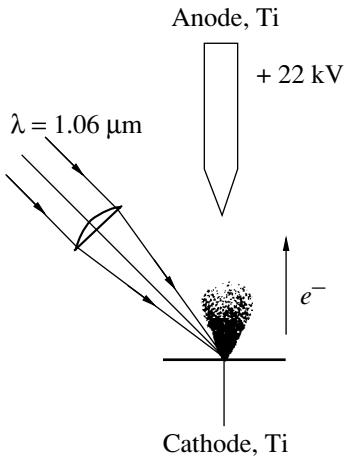


Fig. 1. Geometry of an x-ray source with a titanium anode and a laser-plasma titanium cathode.

(a) the limitations introduced by the characteristics of the electric circuit of the source (the upper bound imposed on the number of electrons participating in the generation of x-rays, by the capacitance of the discharge circuit and the effect of the circuit inductance on the shape and duration of a current pulse);

(b) the functional dependence of the perveance (proportionality factor in the 3/2 law) of the electron flow on the source geometry, the velocity of the current-removal boundary [9], and also with the possible nonmonotonic potential distribution in the interelectrode space [10];

(c) a decrease in the volume charge in the gap between the electrodes due to levelling out of the conductivities of the metal and the laser plasma [11], leading to an increase in the emissivity of the cathode;

(d) recombination during the expansion of laser plasma, which in turn leads to a decrease in the number of electrons being accelerated in the interelectrode gap.

3. EXPERIMENT

The experimental setup is presented in Fig. 1. The discharge circuit of the source [1] possessed a low inductance and consisted of an earthed titanium (aluminum) target serving as a cathode and a titanium anode having the shape of a cone with a tip diameter of 250 μm and a 5-nF capacitor whose charge, corresponding to 10^{-4} C (10^{15} electrons for $U = 20$ kV), limited the number of electrons from the laser plasma, which were accelerated by the electric field. The voltage across the capacitor was maintained by a dc voltage source in the range 3–26 kV. The charging time of the capacitor was

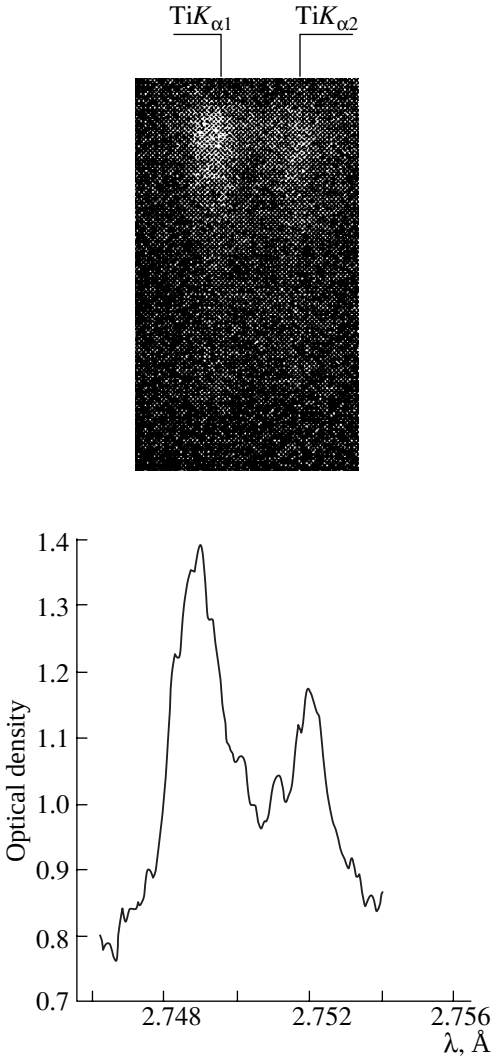


Fig. 2. Spectrogram and densitogram of emission of doublet lines $K_{\alpha1}$ (2.749 Å) and $K_{\alpha2}$ (2.752 Å) by a titanium anode. The spectrogram was obtained by firing 30 laser pulses on crystalline mica spectrograph with double focusing in the VII order of reflection. The radius of the spherical surface of the crystal R = 100 mm; a = 111.5 mm and b = 96.6 mm are the distances from the center of crystal bending to the source and the photographic film, respectively. The backward linear dispersion $D_x = 7.373 \times 10^{-3}$ Å/mm, the working voltage of the source is 22 kV, the laser pulse energy E_L = 0.27 J, the pulse duration τ = 10 ns, the laser beam diameter at the target is 150 μm, and the power density $P = 1.5 \times 10^{11}$ W/cm^2.

determined by a high-resistance resistor of 100 MΩ which excluded the effect of the voltage source on the discharge of the circuit. The current in the diode circuit was measured with the help of the Rogowski loop and independently by the method of voltage drop across the standard low-inductance resistor of 10 mΩ. We used a neodymium laser (λ = 1.06 μm) with the following parameters: pulse duration of 10 and 30 ns and the energy variation in the range 10^{-3}–1 J. The laser beam was focused to the plane anode to form a spot with a radius of approximately 100 μm. The maximum power density of laser radiation at the target was $P = 10^{12}$ W/cm^2. The pressure in the discharge chamber was 10^{-5} torr. Electrons emitted from the laser plasma formed at the cathode were accelerated by the electric field towards the titanium anode, initiating x-ray radiation in the characteristic *K*- and *L*-lines of titanium as well as the x-ray bremsstrahlung with a peak at λ [nm] = 1.86/U (U is measured in kilovolts) [12].

The x-ray radiation intensity was measured with the help of silicon *p–i–n* diodes (Quantrad 100-PIN-250). The diodes facing the titanium anode were protected by filters with a cutoff energy approximately equal to 3.6 keV (100 μm of beryllium and 6 μm of aluminum) and 5.2 keV (200 μm of beryllium and 20 μm of aluminum). In turn, the diode intended for measuring the intensity of radiation emitted by the laser plasma was placed behind the anode of the source. The characteristic radiation and bremsstrahlung emitted by the anode in the direction of the *p–i–n* diode were cut by a specially bent lead screen. The photosensitive surface of the diode was coated with a mylar film. The quantum yield of x-rays in the *K*-lines of the anode ($h\nu(K_\alpha)$ = 4.51 keV and $h\nu(K_\beta)$ = 4.93 keV) was determined from the difference in signals from the diodes facing it from the known value of the response, equal to 0.2 C/J. The spatial distributions of x-ray radiation emitted by the laser plasma (cathode) and the titanium anode were registered by a multiframe pinhole camera with filters 10 μm of beryllium (the cutoff energy $E \approx 1$ keV) and 10 μm of beryllium with 15 μm of titanium ($E \approx 3.8$ keV) on UVVR and DEF x-ray films. X-ray bremsstrahlung was registered on a DEF film (which is more sensitive to hard radiation) mounted behind the UVVR film serving as an extra filter. The pinhole camera was arranged relative to the electrodes in such a way that the magnification of the objects was $\Gamma \approx 4$.

The emission spectrum (Fig. 2) in the region of *K*-lines was recorded on the UVVR film with the help of a crystalline mica spectrograph (with a double interplanar spacing of 19.9 Å) with double focusing. The electron temperature of the cathode plasma was measured from the peak of the spectral distribution with the help of an x-ray spectrograph based on a transmission diffraction grating with a period of 1.4 μm [4]. The dispersion and spectral resolution of such a spectrograph were 20 Å/mm and ~10 Å, respectively.

4. DISCUSSION OF EXPERIMENTAL RESULTS

In our experiments, we studied the spectral, spatial, and time characteristics of a pulsed point x-ray source based on a vacuum diode with a laser-plasma cathode and a needle-shaped anode. In the case of a titanium anode and a laser-radiation power density of 4.2×10^{11} W/cm^2 at the anode, the spectral brightness of the radiation source in the *K*-lines was 1.3×10^{21} photons/cm^2 s sr keV, which corresponds to 1.6×10^{11} x-ray quanta per pulse. The radiation was formed by the doublet $K_{\alpha1}$ (2.749 Å) and

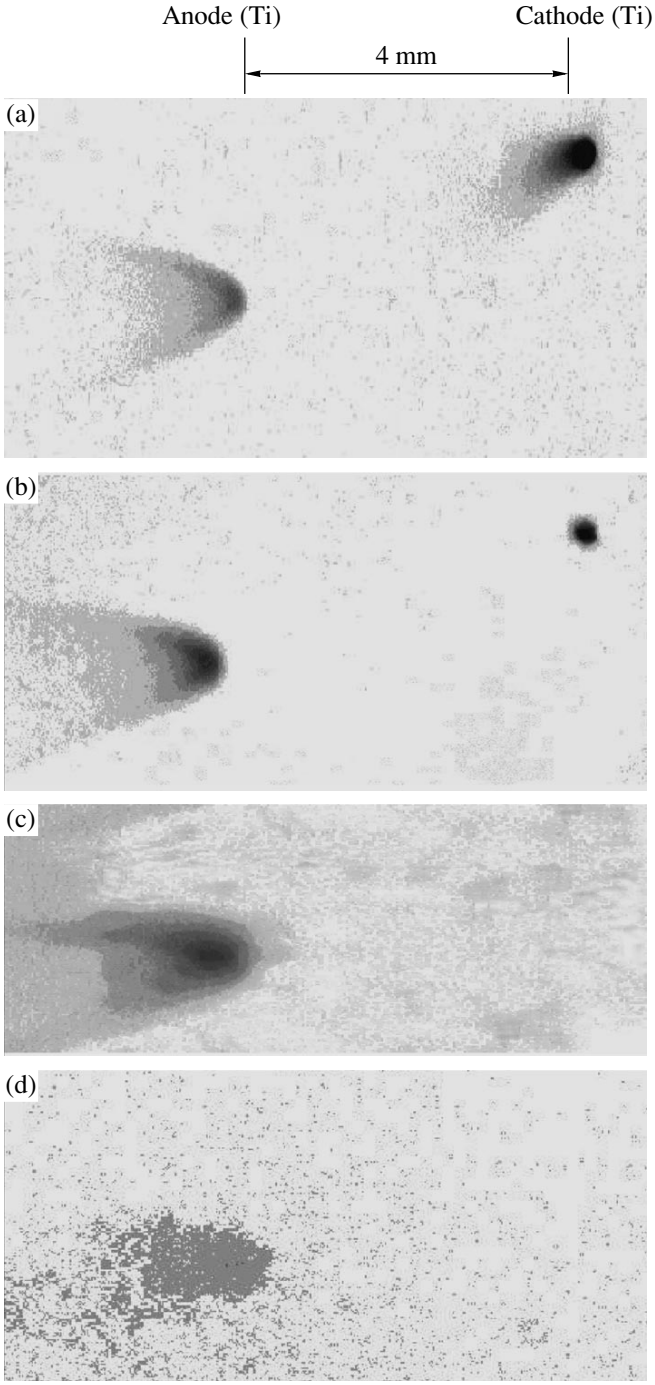


Fig. 3. Spatial distribution of the x-ray radiation emitted by the anode and laser plasma. The working voltage of the source is 22 kV, the laser radiant energy $E_L = 0.33$ J, and the pulse duration is 10 ns. The separation from the source to the pinhole camera aperture is 1 cm and the magnification $\Gamma = 4$. The images were obtained after 15 shots. (a) UVVR film, Be filter of thickness $d = 10$ μm, the diameter D of the camera aperture is 130 μm, the diameter $2r$ of the focal spot of the laser beam at the target is 180 μm, and the density P of laser radiation is 1.3×10^{11} W/cm^2; (b) UVVR film, Be, d = 10 μm, D = 190 μm, $2r$ = 100 μm, $P = 4.2 \times 10^{11}$ W/cm^2; (c, d) Be, d = 10 μm, and Ti, d = 15 μm, D = 130 μm, $2r$ = 100 μm, $P = 4.2 \times 10^{11}$ W/cm^2; UVVR film (c), and DEF film mounted behind the UVVR film (d).

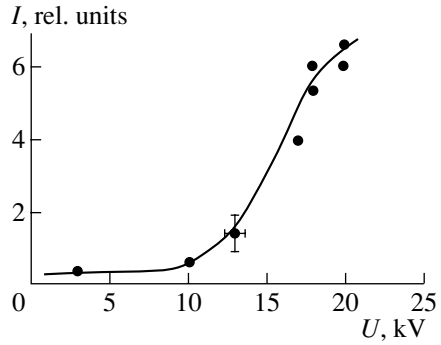


Fig. 4. X-ray radiation yield from a vacuum diode with a laser-plasma cathode as a function of the applied accelerating voltage U ($l = 2.2$ mm, $E_L = 0.2$ J).

$K_{\alpha 2}$ (2.752 Å) and the K_β line (2.514 Å). Microdensitometric measurements proved that bremsstrahlung is negligibly weak as compared to the characteristic radiation in the range 2.4–3 Å (Fig. 2). An analysis of the spatial distribution of x-ray radiation emitted by a titanium cathode and anode revealed that for a power density of laser pulse at the cathode lower than 5×10^{11} W/cm^2, the radiation intensity of the cathode plasma for $\lambda < 3.5$ Å is much lower than the intensity of characteristic radiation and bremsstrahlung emitted by the anode (Fig. 3). The electron temperature $T \approx 70$ eV of the titanium cathode plasma measured with the help of an x-ray spectrograph from the spectral distribution peak for $P = 10^{11}$ W/cm^2 is in good agreement with the model estimates [7]. The x-ray radiation intensity emitted by a source with a needle-shaped anode decreases sharply from the tip of the anode to its bottom (Fig. 3). This allowed us to estimate the source size, which amounted approximately to 250 μm. The obtained images of the discharge gap also confirmed the absence of additional sources of radiation, indicating the absence of the pinch effect.

In order to optimize the diode characteristics, we measured the intensity of x-ray radiation as a function of the laser radiant energy, accelerating voltage, and the distance between the target (cathode) and anode. In these experiments, the laser beam was focused on a plane aluminum target to form a spot of diameter $2r = 130$ μm; the light pulse duration was 30 ns.

The dependence of the x-ray intensity on the accelerating voltage for the electrode spacing $l = 2.2$ mm laser radiation energy $E_L = 0.2$ J is presented in Fig. 4. The laser energy was $Q = 200$ mJ to within ±5%). A strong increase in the x-ray radiation emitted by the anode is observed in the range of accelerating voltages U = 10–20 kV, above which the growth becomes slower. This is in accord with the condition that the maximum contrast of the characteristic radiation lines is attained when the accelerating voltage is three to four times higher than the line excitation thresholds. For this

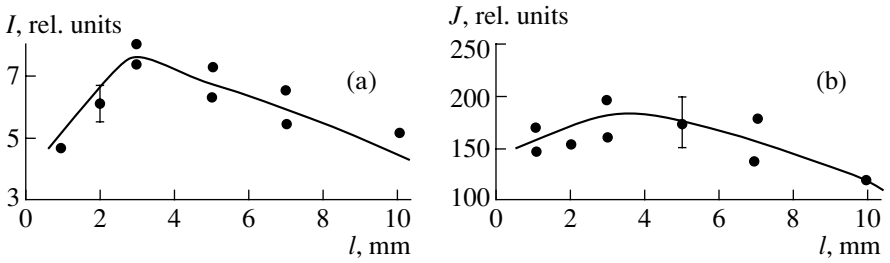


Fig. 5. Dependence of x-ray yield (a) and current in the x-ray source circuit (b) on the electrode spacing l (E_L = 25 mJ, U = 19 kV).

reason, we carried out our subsequent experiments for a voltage of 19 kV.

The dependences of the x-ray intensity and the current pulse amplitude on the electrode spacing l for a fixed laser energy E_L = 25 mJ and voltage U = 19 kV are presented in Fig. 5. It can be seen that the current pulse amplitude in the diode circuit correlates with the intensity of x-ray radiation.

Figure 6 shows the dependence of the x-ray intensity on the laser pulse energy for the anode–cathode spacing l = 2.2 mm and for the accelerating voltage U = 19 kV. The presence of peaks on the dependences presented in Figs. 5 and 6 indicates the extremum of the perveance of the electron flow emitted by the source and the optimal conditions for generating x-rays in given intervals of E_L and l.

In order to study the processes occurring in the electrode gap during the acceleration of electrons from the cathode plasma undergoing hydrodynamic expansion and to estimate the stability of the source operation and the possibility of its synchronization with the complex of the diagnostic equipment, we measured the time shift of the x-ray pulse emitted by the source relative to the radiation emitted by the laser plasma. The dependence of the delay time t_d for the peak of the x-ray diode pulse relative to the peak of plasma radiation on the laser energy for the constant anode–cathode separation l = 2.2 mm and the voltage U = 19 kV is shown in Fig. 7a. For laser radiation energies at the target up to E_L = 300 mJ, the peak of the pulse emitted by the diode is delayed relative to the peak of the x-ray pulse emitted by the laser plasma, this delay becoming smaller upon an increase in E_L. For E_L > 300 mJ, the peak of the pulse under investigation leads the peak of the plasma pulse. In our opinion, this advance is associated with specific experimental conditions (the relation between the electrode spacing and the laser pulse duration). As the energy of laser radiation increases, the velocity of the current-removal boundary increases, and the discharge gap is filled with laser plasma more rapidly, while the action of the laser pulse on the cathode still continues.

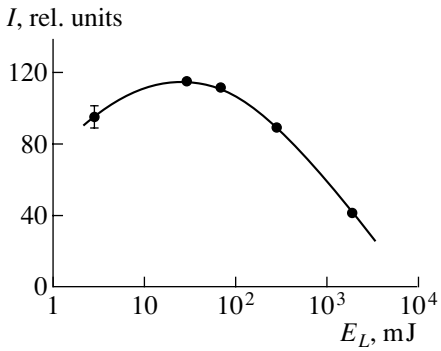


Fig. 6. Dependence of the x-ray yield from the diode on the laser radiant energy (l = 2.2 mm, U = 19 kV).

The time delay t_d between the peaks of x-ray pulses emitted by the diode and the plasma was also measured for various electrode spacings and a constant laser energy. Figure 7b shows the dependence of the delay time t_d for E_L = 25 mJ. This dependence for low laser radiation energies is defined as t_d = 12 [ns/mm]l [mm]. The obtained coefficient of 12 ns/mm is much larger that the value 0.25 ns/mm expected for the acceleration of free electrons by a potential difference of 19 kV. This is explained by the screening of the field applied to the plasma. Electrons reaching the anode are essentially the electrons from the surface layer of the plasma cloud experiencing hydrodynamic expansion. The total number of electrons in a current pulse is estimated by a quantity of the order of 10^{15}, although only a small fraction of these electrons (10^{-4}–10^{-3}) is accelerated to the excitation energy required for generating x-ray quanta in the K-lines of the anode. For l > 7 mm, a two-peak structure of the pulse is observed. This indicates the presence of a group of "fast" electrons leaving the plasma much earlier than the main part of plasma electrons.

The duration of an x-ray pulse from the source under investigation for a fixed power density of laser radiation at the target and a fixed potential at the anode increases with the interelectrode spacing (from 18 ns for l = 1 mm to 28 ns for l = 7 mm) (Fig. 8) and decreases upon an increase in the laser power density for fixed l and U.

The intensity of x-ray radiation depends on the state of the surface of the cathode. Repeated impacts in the same region of the cathode lower the intensity of x-ray radiation.

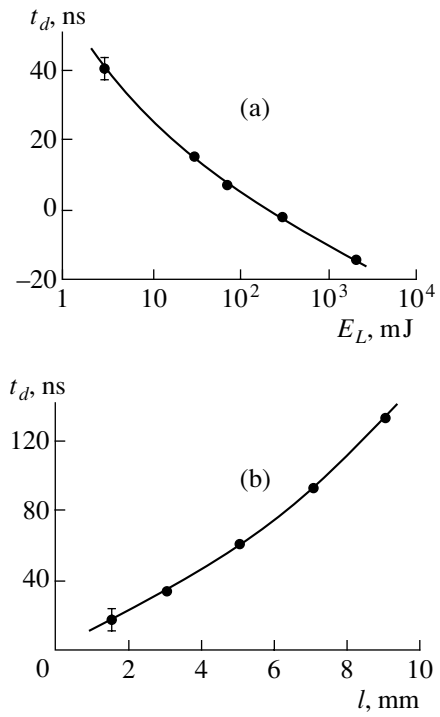


Fig. 7. Dependence of the delay time t_d of the x-ray pulse peak relative to the peak of the x-ray pulse emitted by plasma on (a) the laser energy (l = 2.2 mm, U = 19 kV) and (b) the distance l (E_L = 25 mJ, U = 19 kV).

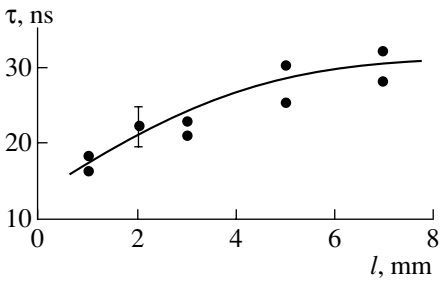


Fig. 8. Dependence of the duration of an x-ray pulse emitted by the vacuum diode on distance l (E_L = 25 mJ, U = 19 kV).

5. CONCLUSION

Thus, our investigations have led to the creation of a pulse point source of x-rays emitting in the range of *K*-lines of Ti with a pulse duration shorter than 20 ns and a spectral brightness of 10^{21} photons/cm^2 s sr keV. The source is superior to the commercially available x-ray tubes with explosive cathodes, having a spectral brightness of 10^{15}–10^{17} photons/cm^2 s sr keV. The parameters of our source demonstrate the prospects of its application for studying rapid processes and in microscopy of moving objects. In turn, the optimization of the parameters of electrical circuit of the source for its operation in the frequency mode makes it possible to apply the source in the x-ray structural analysis of stationary objects and in lithography, where the x-ray radiation dose is one of the applicability criteria. for example, when x-ray pulses repeated at a frequency of 100 Hz, the spectral brightness of such a source in the units adopted for a comparative analysis of continuous-wave sources of x-rays (synchrotrons and x-ray tubes) are of the order of 10^7 photons/mm^2 s mrad^2 (0.1% of the line width), which corresponds to the spectral brightness of tubes with a rotating anode [13]. The study of various operating conditions of the source (for various electrode spacings, power densities of laser radiation, and the voltages across the electrodes) proved that the attainment of optimal characteristics of radiation (spectral brightness, pulse duration, and delay time) is a self-consistence problem which must be solved individually for special purposes.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundations for Basic Research (projects nos. 99-02-18499a and 00-02-16113).

REFERENCES

1. M. V. Dmitriev, Yu. A. Zakharenkov, and A. S. Shikanov, Preprint No. 91, Fiz. Inst. Akad. Nauk SSSR (Lebedev Institute of Physics, Academy of Sciences of USSR, 1989).
2. Yu. V. Korobkin, V. B. Rozanov, A. S. Shikanov, and G. A. Vergunova, Preprint No. 27, Fiz. Inst. Ross. Akad. Nauk (Lebedev Institute of Physics, Russian Academy of Sciences, 1995).
3. Yu. V. Korobkin, V. B. Rozanov, A. S. Shikanov, and G. A. Vergunova, J. Russ. Laser Res. **19**, 101 (1998).
4. Yu. V. Korobkin, I. V. Romanov, A. A. Rupasov, *et al.*, Phys. Scr. **60**, 76 (1999).
5. I. B. Borovskiĭ, *Physical Foundations of X-ray Spectral Investigations* (Nauka, Moscow, 1956).
6. B. L. Henke and M. A. Tesler, in *Advances in X-ray Analysis* (Plenum, New York, 1975), Vol. 18.
7. P. Mora, Phys. Fluids **25**, 1051 (1982).
8. Yu. V. Afanas'ev, N. G. Basov, O. N. Krokhin, *et al.*, Itogi Nauki Tekh., Radiotekh. **17**, 156 (1978).
9. G. A. Mesyats and D. I. Proskurovskiĭ, *Pulse Electric Discharge in Vacuum* (Nauka, Novosibirsk, 1984).
10. A. V. Bolotov, A. V. Kozyrev, A. V. Kolesnikov, *et al.*, Zh. Tekh. Fiz. **61** (1), 40 (1991) [Sov. Phys. Tech. Phys. **36**, 23 (1991)].
11. F. Chen, *Introduction to Plasma Physics* (Plenum, New York, 1984; Mir, Moscow, 1987).
12. *Physical Quantities. Handbook*, Ed. by I. S. Grigor'ev and E. Z. Melikhov (Énergoatomizdat, Moscow, 1991).
13. J. Kirz, O. T. Attwood, B. L. Henke, *et al.*, *X-ray Data Booklet* (Lawrence Berkley Laboratory, University of California at Berkley, 1986), PUB-490 Rev.

Translated by N. Wadhwa