Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 1005–1009.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 1121–1126.
Original English Text Copyright © 2003 by Vekilov, Isaev, Godoniuk.

SOLIDS
Electronic Properties

Electronic Spectrum of the Three-Dimensional Penrose Lattice[¶]

Yu. Kh. Vekilov*, E. I. Isaev, and A. V. Godoniuk
Theoretical Physics Department, Moscow State Institute of Steel and Alloys, Moscow, 119991 Russia
**e-mail: yuri-vekilov@yahoo.com*

Received March 18, 2003

Abstract—The electronic spectrum of the three-dimensional Penrose lattice with “central” decoration by atoms is investigated using the tight binding model with nearest-neighbor interaction. Inverse participation ratios, higher moments of density probabilities, and fractal dimensions of the system are determined. The WFs are critical (they have a power-law dependence on the distance) at all energies in the band and are multifractal measures leading to the entire spectrum of the exponents. The results show that the system is in the critical state of the metal-insulator transition. On critical WFs, the cubic root temperature dependence of the conductivity is obtained. *© 2003 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Measurements of conductivity show that at low temperatures, perfect quasicrystals (QCs) behave similarly to conventional disordered conductors (disordered metals and heavily doped semiconductors in the vicinity of the metal-insulator transition (MIT)), and the possibility of electron localization–delocalization (LD) in QCs now is being actively discussed. It is known that in disordered conductors, the electrons can undergo a transition to the insulating state (Anderson localization) with an increasing degree of disorder. The electron diffusion coefficient takes a finite value in the conducting phase and vanishes in the insulating phase, which is revealed by crossing of the Fermi level at a certain value of energy, called the mobility edge. Localization occurs for a sufficiently strong disorder because of quantum interference effects brought about by randomness of the disorder. At a finite temperature, according to the scaling theory of the Anderson transition with interacting electrons [1], the correction to the conductivity in the region $L < L_T$ and $\xi < L_T$ (where L is the sample size, $L_T = \sqrt{D\hbar/T}$ is the interaction length, and D is the diffusion coefficient) is proportional to $\sqrt{T}$. In the region where $\xi \gg L_T > L$, $\sigma \sim T^{1/3}$. Sufficiently far on the insulating side, the conductivity follows the Mott law for a variable range hopping (VRH) conductivity, $\sigma = \sigma_0 e^{-(T_0/T)^{1/4}}$.

QCs have an extremely high resistivity value with a pronounced negative temperature coefficient and a finite small electronic contribution to the specific heat. In contrast to the conventional disordered conductors, QCs become more insulating with increasing sample quality and annealing of defects. The quality measure of QCs is the resistivity ratio $\mathcal{R} = \rho(4.2\text{ K})/\rho(300\text{ K})$. The more perfect samples have higher $\mathcal{R}$, as well as $\mathcal{R}$ ranges from several units to two hundred and even higher depending on the object and sample quality.

Empirically, R can therefore serve as a parameter to control the MIT. The highest resistivity of all the known QCs occurs in icosahedral *i*-Al-Pd-Re QCs, where the value of resistivity at 4.2 K exceeds 1 Ω cm. Recent experiments for *i*-Al-Pd-Re [2, 3] show that for samples all having different ratios $\mathcal{R}$, a square root temperature dependence of conductivity $\sigma \sim \sqrt{T}$ is ordinarily observed at $T < 20$ K. For samples with $\mathcal{R}$ on the order 20 or higher, this dependence is replaced by the $\sigma \sim T^{1/3}$ law at $T < 5$ K. For samples with high $\mathcal{R}$ (~45 or higher), a VRH conductivity obeying the Mott law, $\sigma = \sigma_0 \exp(-(T_0/T)^{1/4})$, or even the Efros–Shklowski law, $\sigma = \sigma_0 \exp(-(T'_0/T)^{1/2})$, is observed. (The same temperature dependences of σ were obtained by other authors for samples with slightly different values of $\mathcal{R}$ [4, 5].) We thus see that an obvious analogy exists in the behavior of low-temperature conductivity in perfect QCs and in disordered conductors near the MIT, although the reasons for the electron localization in these objects are different. This analogy is also valid at the microscopic level.

In the theory of Anderson localization in disordered conductors, one is interested in the effect of a random potential on quantum-mechanical WFs (WFs). When the randomness is weak, the WFs are extended throughout the entire system (metallic side of the MIT), whereas at sufficiently high disorder, all WFs become localized (insulator side of the MIT). In the vicinity of the critical point of the MIT, the WFs are neither extended nor exponentially decaying; as numerical cal-

---

[¶] This article was submitted by the authors in English.

culations show, they display scaling behavior and decrease with distance following a power law ("critical" WFs) [6–8].

Discussions of the problem of localization of electronic states in QCs began immediately after their discovery (see, e.g., [9]). Highly resistive QCs are usually attributed to the existence of a deep pseudogap in the density of electronic states (DOS) at the Fermi level (DOS at the Fermi level in QCs is low but finite) and to the tendency of the electrons at the Fermi level to be localized. However, the presence of a pseudogap is not sufficient to explain the high value of resistivity; the main reasons for this are seen in the low electron mobility, which is obviously caused by the specific symmetry of QCs. From the general standpoint, one can conclude that due to the self-invariant structure of QCs, the WFs must be critical. The critical behavior of the WFs in QCs has been well established in the cases of one- and two-dimensional QCs [10, 11–13]. However, for three-dimensional systems (icosahedral QCs), the first publications were controversial [14–16], and even some recent publications contain the claim that in the three-dimensional case, the critical nature of WFs may be lost to some extent [17, 18]. At the same time, other numerical investigations of the electron spectra of low-order periodic approximants of icosahedral QCs show that most of the WFs are still critical, although the electron spectrum does not come into contact with a hierarchical gap structure typical of the Cantor set of measure zero in one-dimensional QCs [19–21]. Thus, the problem exists and more information on the electron spectra and WFs is required in order to judge the electron localization in three-dimensional (icosahedral) QCs.

In this paper, we present the results of a numerical investigation of the scaling behavior of the electron spectrum and WFs of the three-dimensional Penrose lattice. The main information needed to characterize the LD transition in QCs is obtained. The inverse participation numbers (the second moments of the density probabilities) and the generalized inverse participation numbers (higher moments) are obtained. Fractal dimensions of the spectrum are obtained and the critical behavior of the WFs is studied. The results are important for understanding the electron LD transition in icosahedral QCs. This work is a continuation of the previous studies [19, 20], where the singularities of the electron spectrum of icosahedral QCs and the effect of small perturbations on it have been studied using tight-binding and level-statistic methods. In [19], singularities of the electron spectrum were analyzed, and it was shown that the spectrum is not Cantorian, but contains a singular part. In [20], we studied the influence of chemical disorder and phasons on the electron spectrum by changing on-site energies and transfer integrals.

This paper is organized as follows. In Section 2, we consider the main model approximations and calculation technique. In Section 3, the results of investigation of the scaling behavior of the electron spectrum are discussed. Section 4 contains our conclusions.

2. MODEL APPROXIMATIONS AND CALCULATION TECHNIQUES

The electronic spectrum of the three-dimensional Penrose lattice (the Amman–Kramer network) treated as a structural limit of a sequence of periodic cubic approximants with increasing period has been studied in the tight-binding approximation (TBA). The first five cubic approximants to the icosahedral QCs (1/1, 2/1, 3/2, 5/3, and 8/5) were investigated. We considered the central decoration of approximants with "atoms" of one type, namely, atoms with one *s*-orbital per atom located at rhombohedral centers. The unit cells of these approximants contained 32, 136, 576, 2440, and 10 330 atoms, respectively. The projection technique for construction of approximants was described previously [19]. To minimize the number of adjustable parameters of the model, we used a Hamiltonian with constant hopping integrals between nearest neighbors (atoms).

The Hamiltonian was expressed as

$$H = \sum_i |i\rangle \varepsilon_i \langle i| + \sum_{i \neq i} |i\rangle t_{ij} \langle j|. \qquad (1)$$

If atoms of only one type are present, the diagonal elements ε_i can be omitted. In this case, the Schrödinger equation in the TBA can be written as

$$\sum_{ii} t_{ij} \Psi_j = E_i \Psi_i, \qquad (2)$$

where the transfer integrals are set equal to a nonzero constant $t_{ij} = -1$ only in the case of the nearest-neighbor atoms. The periodic boundary conditions were used to help reduce the size-dependence effects.

We study the localization problem in the TBA by calculating the inverse participation numbers (moments or 2*q*-norms of the wave function) defined by the relation

$$P^{-1} = \|\Psi\|_{2q} = \frac{\sum_j |\Psi_j|^{2q}}{\left(\sum_j |\Psi_j|^2\right)^q}, \qquad (3)$$

from which "participation ratios" and fractal dimensions D_q can be determined. P is called the participation number because it is the measure of the number of sites that contribute to a state of a given energy E_i. The corresponding fraction $p = P/N$ of all the sites is called the participation ratio. The value of p for $q = 2$ is frequently used in the problem of electron localization. The WFs were classified in accordance with their normalization integrals. They are considered delocalized if

$$\int\limits_{|r|<R} |\Psi(r)|^2 dr \sim R^d,$$

where d is the space dimensionality. They are assumed localized when their finite norms exist, and are defined as critical, $\Psi \sim r^{\alpha}$, when they cannot be normalized in an infinite space and are not delocalized. Strongly localized WFs correspond to the case where $\alpha = \infty$, and freely extended WFs correspond to the case where $\alpha = 0$; Ψ can be normalized in the three-dimensional case only for $\alpha \geq 3/2$. For extended states, the moments of the WF depend on the system size as $\|\Psi\|_{2q}^{\text{extended}} \sim N^{1-q}$, as follows from Eq. (3). For exponentially decaying localized functions, we have $\|\Psi\|_{2q}^{\text{exp.loc}} \approx N^0$. We can therefore obtain the exponent of the WFs by analyzing the system size dependence of the moments calculated in a system of a sufficient size. For the relative number of states with moments $\|\Psi\|_{2q} \leq N^{\gamma}$, the integrated distribution function defined as

$$I_{2q}(\gamma) = \frac{1}{N}\sum_{n=1}^{N}\theta(\gamma - \log_N\|\Psi\|_{2q})$$

gives the integrated distribution of the exponent of power-law decay for a specified system if the finite-size correction is negligible. The procedure of finding the exponent α has been described in [19] (see also [13] for two-dimensional Penrose lattice), and we here note that the behavior of the function $\gamma(q, \alpha)$ has been analyzed for the first five approximants. As a result, the "localization" exponent α was found for each approximant under investigation.

From the relation $P_q \sim N^{-D_q(q-1)}$, which applies near the critical point, we obtained the fractal dimensions D_q of the system (here, N is the number of atoms in the unit cell of an approximant).

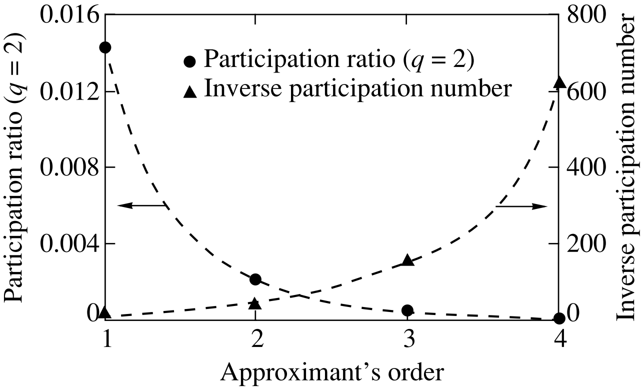


Fig. 1. Inverse participation numbers P^{-1} and the participation ratio (q = 2) for the five rational approximants.

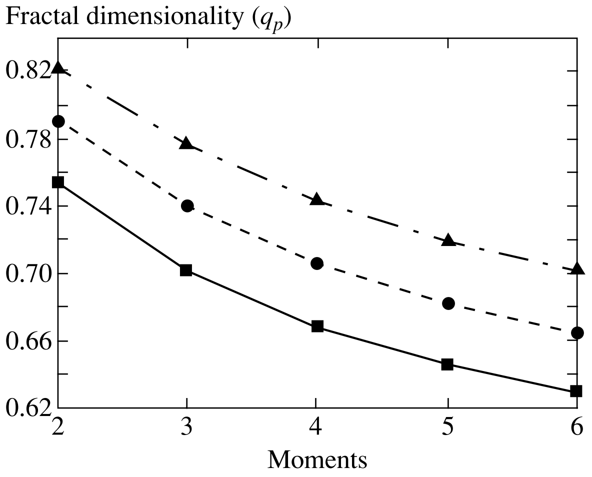


Fig. 2. Fractal dimensionality (q_p) for different moments ($2q$-norms) of the spectrum ($D_{p=0}$ = 3 for all approximants).

3. RESULTS AND DISCUSSION

The results of calculations are presented in Figs. 1–4. The behavior of the inverse participation numbers, participation ratios (Fig. 1), and fractal dimensions (Fig. 2) shows that the electronic states are neither localized nor delocalized in all the considered approximants (the first five approximants were considered). Indeed, the inverse participation numbers P are proportional to N^{γ}, where γ must vanish for localized states because they fit into a sample of a given size, and $\gamma = 1$ for states uniformly extended over the entire sample. Because the calculated value of P does not satisfy both of these limits, we can assume that the WFs, or rather their envelopes, drop as an inverse power of the distance $\Psi \sim r^{-\alpha}$. We see next (Fig. 2) that the calculated dimensions D_q of the system satisfy the inequalities $D_0 > D_1 > D_2 > \ldots$, where for all approximants, D_0 is equal to topological dimension (3) and the dimensions are therefore not simply fractal, but multifractal. The multifractality regime means that the system is in the critical state and the WFs at criticality are multifractal measures leading to the entire spectrum of critical exponents. The spectrum of multifractal dimensions has universal features for states in the vicinity of the MIT. We can therefore conclude that the ground state of three-dimensional perfect QCs is a critical state of the MIT.

The results in Fig. 3 show that the WF is critical at all energies in the band. The dependence $\Psi \sim r^{-\alpha}$ is typical of the critical state of the system. It is known that systems without characteristic intrinsic length scales obey homogeneity laws under rescaling. The absence of length scales means that some observable F shows a

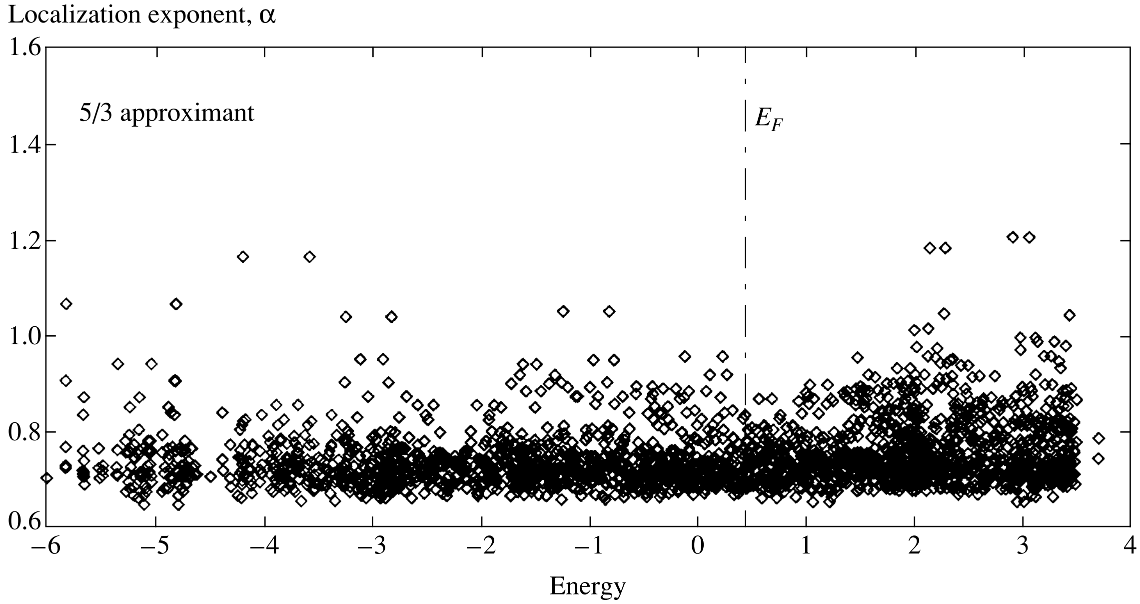


Fig. 3. Distribution of the localization exponent α ($\Psi \sim r^{-\alpha}$) on the energy band. The eigenstates are critical at all energies.

typical homogeneity law $F(sx) = s^kF(x)$, where k is called the homogeneity exponent and s is a real number. This implies that the rescaling of x can be compensated by a rescaling of the observable F. For real-valued functions $F(x)$, the solution of the homogeneity equation is a power-law function $F(x) \sim x^k$. The function $\Psi \sim r^{-\alpha}$ is therefore a solution of the homogeneity equation with the homogeneity exponent $k = \alpha$, and we have a scale-invariant behavior of the system, typical of critical states. If $F(x)$ is a functional of degree q of the observables involved in the definition of $F(x)$ (i.e., moments in our case), $F(x) = F^{[q]}(x)$, then in the simplest situation, $k(q)$ defined by

$$F^{[q]}(sx) = s^{k(q)}F^{[q]}(x)$$

is a linear function of q. If $k(q)$ significantly deviates from linearity, the scaling behavior of $F(x)$ is anomalous and the system therefore demonstrates multifractal behavior [22]. Calculations show that the multifractal behavior of the system becomes pronouncedly apparent for higher order approximants (5/3).

Taking into consideration the behavior of the localization exponent α, we see that α tends to a certain value in the thermodynamic limit ($N \longrightarrow \infty$). It is known from the theory of Anderson transitions that near the transition into the dielectric (metallic) state, the correlation (localization) length ξ tends to infinity. As mentioned above, the theory of Anderson localization for interacting electrons implies that $\sigma \propto T^{1/3}$ at the critical point of the MIT, where $\xi \gg L_T > L$, with $L_T = \sqrt{D\hbar/T}$ being the interaction length [3]. For QCs, the conventional picture of the Anderson localization in disordered systems is not relevant. In QCs, the electronic states can be localized by the quasiperiodic potential itself, and as we have shown (Fig. 3), the WFs in three-dimensional QCs behave as in the critical state of the MIT for conventional disordered systems. As shown in [23], it is possible to obtain the $\sigma \propto T^{1/3}$ dependence by considering a VRH conductivity on the critical WFs. Following the Mott procedure, we define the tunneling integral on the critical WFs as

$$I \sim |\Psi|^2 \sim |R^{-\alpha}|^2 \equiv \exp(-2\alpha \ln R). \qquad (4)$$

We then define the conductivity

$$\sigma \propto I \exp(-\Delta E)/(kT),$$

where $\Delta E = 3/(4\pi R^3 N(E_F))$ is the minimal activation

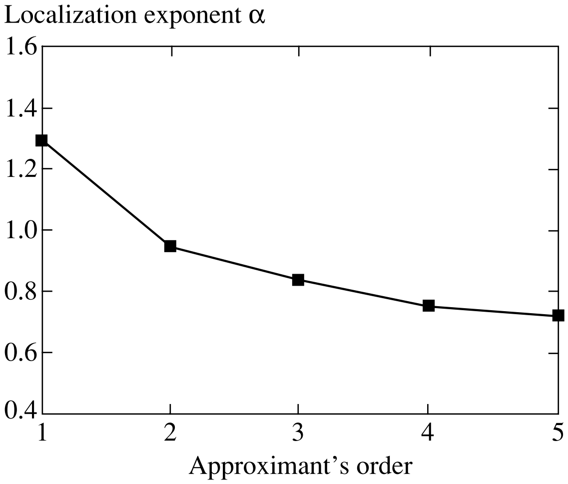


Fig. 4. The localization exponent α (averaged over the band) for different approximants.

energy for hopping over the distance R. The expression

$$\exp(-2\alpha\ln R)\exp(-\Delta E/kT) \tag{5}$$

has a maximum when the exponent $(-2\alpha\ln R - \Delta E/kT)$ has a minimum. Substituting ΔE and finding this minimum, we obtain $\sigma \propto T^{2\alpha/3}$. In order to obtain $\sigma \propto T^{1/3}$, the exponent α must be equal to 0.5. The results of calculations of α (Fig. 4) show that α decreases as the order of the approximant increases. It is difficult to say to what value α tends in the thermodynamic limit, but the tendency is obvious. The results of calculations also show that the value of α depends on the moment number, and the last expression for σ should involve some realization of α. Therefore, the result coincides with the experiment and predictions of the scaling theory of localization for the critical region of the MIT.

4. CONCLUSIONS

The results of investigation of the scaling behavior of the electron spectrum for several periodic approximants (1/1, 2/1, 3/2, 5/3, 8/5) of the three-dimensional Penrose lattice with central decoration are presented. The critical behavior effects are visible even for these low-order approximants. The calculated WFs are critical for all energies in the band and are multifractal measures with the entire spectrum of critical exponents. The electronic states are more localized at the Fermi level than at the bottom of the band. The results show that the background state of perfect regular icosahedral QCs should be the critical state of the LD transition. The nature of electron localization in icosahedral QCs has been discussed previously [9, 19, 20], and it was shown that this localization is unstable under small perturbations (phasons, chemical disorder, and magnetic fields). The critical behavior of WFs can explain the experimentally observed power-law dependence of conductivity, $\sigma \propto T^{1/3}$. Calculating the VRH probability on critical WFs, we immediately obtain the $\sigma \propto T^{2\alpha/3}$ law. To coincide with experiment, the realization of exponent α should be equal to 0.5 in the thermodynamic limit. At the same time, it is impossible to obtain the Mott law on critical WFs for VRH conductivity on the insulating side of the MIT. New ideas are necessary.

We are grateful to the Russian Foundation for Basic Research (project no. 03-02-16970) and the Royal Swedish Academy of Sciences for financial support.

REFERENCES

1. B. L. Altshuler and A. C. Aronov, JETP Lett. **37**, 410 (1983).
2. J. Delahaye and C. Berger, Phys. Rev. B **64**, 094203 (2001).
3. V. Srinivas, M. Rodmar, R. König, *et al.*, Phys. Rev. B **65**, 094206 (2002).
4. Chang-Ren Wang and Shui-Tien Lin, J. Phys. Soc. Jpn. **68**, 3988 (1999).
5. C. R. Wang, T. I. Su, and S. T. Lin, J. Phys. Soc. Jpn. **69**, 3356 (2000).
6. M. Schriber and M. Grussbach, Phys. Rev. Lett. **67**, 607 (1991).
7. M. Grussbach and M. Schriber, Phys. Rev. B **51**, 663 (1995).
8. A. D. Mirlin and F. Evers, Phys. Rev. B **62**, 7920 (2000).
9. S. J. Poon, Adv. Phys. **41**, 303 (1992).
10. P. A. Kalugin, A. Yu. Kitaev, and L. S. Levitov, Pis'ma Zh. Éksp. Teor. Fiz. **41**, 119 (1985) [JETP Lett. **41**, 145 (1985)].
11. P. A. Kalugin, A. Yu. Kitaev, and L. S. Levitov, Zh. Éksp. Teor. Fiz. **91**, 692 (1986) [Sov. Phys. JETP **64**, 410 (1986)].
12. M. Khomoto, B. Sutherland, and Chao Tang, Phys. Rev. B **35**, 1020 (1987).
13. H. Tsunetsugu, T. Fujiwara, K. Ueda, and T. Tokihiro, Phys. Rev. B **43**, 8879 (1991).
14. G. Kasner, H. Schwabe, and H. Bottger, Phys. Rev. B **51**, 10454 (1995).
15. M. A. Marcus, Phys. Rev. B **34**, 5981 (1986).
16. K. Niizeki and T. Akamatsu, J. Phys.: Condens. Matter **2**, 2759 (1990).
17. E. S. Zijlstra and T. Janssen, Mater. Sci. Eng. A **294–296**, 886 (2000).
18. E. S. Zijlstra and T. Janssen, Phys. Rev. B **61**, 3377 (2000).
19. D. V. Olenev, E. I. Isaev, and Yu. Kh. Vekilov, JETP **86**, 550 (1998).
20. Yu. Kh. Vekilov, E. I. Isaev, and S. F. Arslanov, Phys. Rev. B **62**, 14040 (2001).
21. T. Reith and M. Schriber, J. Phys.: Condens. Matter **10**, 783 (1998).
22. M. Janssen, Int. J. Mod. Phys. B **8**, 943 (1994).
23. Yu. Kh. Vekilov and E. I. Isaev, Phys. Lett. A **300**, 500 (2002).

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 1010–1014.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 1127–1132.
Original Russian Text Copyright © 2003 by Zvyagin, Mironov, Ormont.

SOLIDS
Electronic Properties

The Effect of Exchange Interaction on the Energy Spectrum of Electrons in Doped Intentionally Disordered Superlattices

I. P. Zvyagin, A. G. Mironov, and M. A. Ormont*
Moscow State University, Moscow, 119992 Russia
**e-mail: scon281@phys.msu.su*

Received March 25, 2003

Abstract—The electron density distribution is calculated for a doped superlattice with controlled vertical disorder caused by fluctuations of the layer thicknesses (quantum well widths) in the growth direction. At low temperatures, the exchange interaction leads to an increase in the scatter of quantum confinement levels and the formation of a soft gap in the electron density distribution over quantum wells of the superlattice. *© 2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Intentionally disordered superlattices (IDSLs) represent the structures with quantum wells (QWs) in which the distribution of quantum confinement (dimensional quantization) levels can be modified by controlling the thicknesses of layers (QW widths) in the growth stage. The creation of such structures makes it possible to study the influence of the degree and type of disorder on the electron states and their kinetic properties [1–4].

This paper addresses doped IDSLs. The results of recent investigations of GaAs/GaAlAs-based structures homogeneously doped with silicon revealed nontrivial features in the low-temperature vertical (in the SL growth direction) conductivity [5]. In particular, it was found that the vertical conductivity weakly (nonexponentially) depends on temperature ("quasimetallic" behavior) even in strongly disordered systems where the width of a random distribution of the quantum confinement levels significantly exceeds the width of a miniband in the corresponding ordered SL, so that localization of all states in the miniband can be expected. This feature was explained [5] in terms of the Coulomb interaction between electrons.

Recently [6, 7], the equilibrium charge distribution and the electron energy spectrum of an IDSL was calculated in the Hartree approximation (screening of the disorder) with allowance for the Coulomb fields caused by the redistribution of electrons between QWs. It was found that screening of the disorder leads to significant narrowing of the distribution of quantum confinement levels in the QWs. In the case of sufficiently strong doping [6], whereby all QWs in the structure are filled with electrons, the analysis was performed by the method of the local density functional using a standard variational procedure. However, later [7] it was shown that, in the general case (e.g., for moderate doping), the variational problem is nonlinear. The algorithm for solving this problem in the Hartree approximation and the results of numerical calculations can be found in [7].

It is known, however, that the Hartree approximation may be inadequate in the region of low densities, where SLs can feature exchange–correlation effects leading, for example, to the formation of electron superstructures [8]. The aim of this study was to assess the influence of the Coulomb exchange interaction on the electron density distribution in doped composite IDSLs.

2. CALCULATION OF CHARGE DISTRIBUTION IN SUPERLATTICES WITH CONTROLLED DISORDER

We consider an IDSL homogeneously doped with a donor impurity. According to the standard model [6, 7], we assume that QWs in this structure are separated by barriers of constant width and that the positions of quantum confinement levels, calculated in the approximation of isolated wells, fluctuate as a result of controlled random fluctuations of the QW width. The transfer of electrons from donors occurring inside the barriers to the region of QWs leads to the appearance of an internal electric field shifting the quantum confinement levels. This shift is accompanied by a change in the electron population of QWs, which implies the need for self-consistent calculation of the electric fields and the electron density distribution in QWs of the IDSL.

As in [6–8], we assume that QWs are narrow, whereby the consideration can be restricted to the lower subband of the quantum confinement levels in each QW. Indeed, for a GaAs/GaAlAs superlattice with the usual QW width of L = 6 nm (corresponding to the maximum of the distribution of quantum confinement levels), the difference in energy between edges of the first and second subbands is about 150 meV. At the same time, the characteristic energies (the scatter of levels and the miniband width in the corresponding regular SL) usually do not exceed 30–40 meV. In addition, the small width of QWs allows the charge of carriers in a QW to be approximately considered as concentrated in a plane (see [6–8]). This model is analogous to the Visscher–Falikov model of charged layers used for the description of screening related to the charge redistribution in the planes of layers [9]. In this approximation, the action of both an external field applied in the growth direction and the internal field appearing as a result of the charge redistribution between QWs of the SL can be readily described by considering "classical" shifts of the quantum confinement levels V_i in the QWs [8].

The problem of finding the equilibrium charge density distribution with allowance for the Coulomb interaction is conveniently solved by the method of local density functional (see [6–8]). For calculating the electron density distribution over QWs of an IDSL, we performed direct minimization of the free energy considered as a nonlinear electron density functional with allowance for both Hartree and exchange contributions to the system energy.

In the approximation of narrow QWs, the electron density can be expressed as

$$n(z) = \sum_i \nu_i \delta(z - z_i), \tag{1}$$

where z is the coordinate in the growth direction; ν_i is the two-dimensional (2D) electron concentration in the ith layer; $\nu_i = \sum_\sigma \nu_{i\sigma}$; σ is the projection of the electron spin; $i = 1, 2, \ldots, N$; N is the total number of QWs in the given structure; and z_i is the coordinate of the ith QW. Thus, the electron density functional is defined on a discrete manifold $\nu = \{\nu_i\}$, merely as a function of many variables ν_i.

At $T = 0$, the system energy per unit area can be written as [8]

$$\begin{aligned} E(\nu) &= \sum_i E_i^{(0)} \nu_i + \sum_{i,\sigma} \frac{\nu_{i,\sigma}^2}{2\rho_0} \\ &+ \frac{1}{2}\sum_{i,j} V_{ij}(\nu_i - \nu_0)(\nu_j - \nu_0) - \sum_{i,\sigma} C_x \frac{e^2}{\varepsilon} \nu_{i,\sigma}^{3/2}, \end{aligned} \tag{2}$$

where $E_i^{(0)}$ is the energy level in the ith QW in the absence of free electrons (seeding energy); $\rho_0 = m/(\pi\hbar^2)$ is the 2D density of states in the QW; V_{ij} is the electron interaction energy calculated within the framework of the classical Hartree approximation (with allowance for interaction between electrons and a homogeneous positive background); $V_{ij} = -2\pi e^2|z_i - z_j|/\varepsilon$; ε is the permittivity (for simplicity, we assume the permittivities of barriers and QWs to be the same); e is electron charge; $\nu_0 = N_d d$ is the average 2D electron density in the layer; d is the average period of the structure; $z_i = id$; N_d is the volume concentration of the doping impurity; and $C_x = (32/9\pi)^{1/2} \approx 1.06$. The last term in the right-hand side of Eq. (2) describes the exchange interaction to the total energy of the system.

As is known [10], the ground state of the electron gas at low temperatures and not too large electron densities in the QW is spin-polarized (the Stoner ferromagnetism of a 2D electron gas). The spin-ordered state is realized when the concentrations ν_i are below a certain critical value $\nu_p = 8(3 - 2\sqrt{2})(C_x e^2 \rho_0/\varepsilon)^2$ (see, e.g., [8]). In an IDSL with inhomogeneous distribution of electrons over the QWs, it is possible that the electron gas is polarized in some QWs, while being unpolarized in others (with sufficiently large electron densities). We consider the case of sufficiently weak doping, whereby the electron states are ordered with respect to spin in all QWs of the given IDSL. The distribution of electrons over QWs was determined by minimizing the free energy (2) considered as a nonlinear electron density functional.

The algorithm of minimization was as follows. Let $\nu = \{\nu_{i\uparrow}, \nu_{i\downarrow}\}$ be a certain seeding electron density distribution over the QWs. Selecting two QWs, a and b, we can write the following expressions for the corresponding electron densities:

$$\begin{aligned} \nu_{a\uparrow} &= \frac{\bar{\nu}_{ab}}{2}(1 + u)(1 + x_a), \\ \nu_{a\downarrow} &= \frac{\bar{\nu}_{ab}}{2}(1 + u)(1 - x_a), \\ \nu_{b\uparrow} &= \frac{\bar{\nu}_{ab}}{2}(1 - u)(1 + x_a), \\ \nu_{b\downarrow} &= \frac{\bar{\nu}_{ab}}{2}(1 - u)(1 - x_a). \end{aligned} \tag{3}$$

Here, $\bar{\nu}_{ab} = (\nu_a + \nu_b)/2$ is the average QW population in the given pair, $x_a = (\nu_{a\uparrow} - \nu_{a\downarrow})/\nu_a$ is the degree of electron gas polarization in the well a, and $u = \nu_{ab} = (\nu_a - \nu_b)/2\bar{\nu}_{ab}$ is the asymmetry of the QW population in the given pair. The free energy (2) can be rewritten as follows:

$$\begin{aligned} E &= (E_a^{(0)} + E_b^{(0)})\bar{\nu}_{ab} + (E_a^{(0)} - E_b^{(0)})\bar{\nu}_{ab} u \\ &+ \frac{\bar{\nu}_{ab}^2}{4\rho_0}\{(1 + u)^2(1 + x_a^2) + (1 - u)^2(1 + x_b^2)\} \\ &+ V_{ab}(\bar{\nu}_{ab}(1 + u) - \nu_0)(\bar{\nu}_{ab}(1 - u) - \nu_0) \end{aligned}$$

$$+\sum_{j\neq b}V_{aj}(\bar{\nu}_{ab}(1+u)-\nu_0)(\nu_j-\nu_0) \tag{4}$$

$$+\sum_{j\neq a}V_{bj}(\bar{\nu}_{ab}(1-u)-\nu_0)(\nu_j-\nu_0)$$

$$-C_x\frac{e^2}{\varepsilon}\bar{\nu}_{ab}^{3/2}\{(1+u)^{3/2}((1+x_a)^{3/2}+(1-x_a)^{3/2})$$

$$+(1-u)^{3/2}((1+x_b)^{3/2}+(1-x_b)^{3/2})\}+\tilde{E}_{ab},$$

where $\tilde{E}_{ab}$ is the part of the total energy independent of the densities ν_a and ν_b:

$$\tilde{E}_{ab}=\sum_{\substack{i\neq a,b\\ \sigma}}\frac{\nu_{i\sigma}^2}{2\rho_0}+\sum_{i\neq a,b}E_i^{(0)}\nu_i$$

$$+\frac{1}{2}\sum_{\substack{i\neq a,b\\ j\neq a,b}}V_{ij}(\nu_i-\nu_0)(\nu_j-\nu_0) \tag{5}$$

$$-\sum_{i\neq a,b}C_x\frac{e^2}{\varepsilon}(\nu_{i\uparrow}^{3/2}+\nu_{i\downarrow}^{3/2}).$$

Since the electron density is positive, the above quantities obey the condition $-1\le x_a, x_b, u\le 1$. The free energy minimization was performed taking into account that, for $\nu_i<\nu_p$, the electron gas in the ith QW is polarized, $x_i=1$, while for $\nu_i>\nu_p$, the densities of electrons with up and down spins are equal ($\nu_{i\uparrow}=\nu_{i\downarrow}$) and $x_i=0$. A minimum of the free energy as a function of u can be achieved both inside and on the boundaries of the [–1, 1] segment. Inside this segment, the points of extrema of the $E(u)$ function can be determined from the condition

$$\frac{dE(u)}{du}=(E_a^{(0)}-E_b^{(0)})\bar{\nu}_{ab}$$

$$+\frac{\bar{\nu}_{ab}^2}{2\rho_0}\{(1+u)(1+x_a^2)-(1-u)(1+x_b^2)\}$$

$$-2V_{ab}\bar{\nu}_{ab}^2u-\sum_{j\neq a,b}(V_{bj}-V_{aj})\bar{\nu}_{ab}(\nu_j-\nu_0) \tag{6}$$

$$-\frac{3}{2}C_x\frac{e^2}{\varepsilon}\bar{\nu}_{ab}^{3/2}\{(1+u)^{1/2}((1+x_a)^{3/2}+(1-x_a)^{3/2})$$

$$+(1-u)^{1/2}((1+x_b)^{3/2}+(1-x_b)^{3/2})\}=0.$$

In the absence of exchange, the value of $u^{(0)}$ corresponding to a minimum of $E(u)$ within the [–1, 1] segment is

$$u^{(0)}=\frac{2}{\bar{\nu}_{ab}}\frac{\rho_0(E_b^{(0)}-E_a^{(0)})+\rho_0\sum_{j\neq a,b}(V_{bj}-V_{aj})(\nu_j-\nu_0)+(1/2)\bar{\nu}_{ab}(x_a^2-x_b^2)}{1+x_a^2+x_b^2-4\rho_0V_{ab}}. \tag{7}$$

A solution to Eq. (6) with allowance for exchange was found by the method of successive approximations, using (7) as the initial approximation. Accordingly, the system energy was successively minimized for each pair of QWs with allowance for both spin-ordered and nonpolarized states in the wells: for each QW pair (i, j), the total energy was minimized with respect to the concentrations $\nu_{i\uparrow}$, $\nu_{i\downarrow}$, $\nu_{j\uparrow}$, $\nu_{j\downarrow}$. If the iterative procedure converges to a point u inside the [–1, 1] segment, this u value corresponds to an extremum of the function $E(u)$. The energy can also reach a minimum on the boundaries of the segment; the corresponding u value was taken on the boundary where the energy minimum was lower. Then the energy was minimized with respect to populations of the next QW pair and so on (in each stage, the total energy of the system usually decreases). Upon trying all possible pairs of QWs, the procedure was repeated until the system energy ceased to change. The independence of the final charge density distribution of the initial approximation indicated that the obtained distribution corresponds to the ground state of the system studied.

3. DISCUSSION OF RESULTS

The electron density distributions $\nu=\{\nu_i\}$ were calculated, with allowance for the Coulomb fields caused by the redistribution of electrons between QWs, for a given seeding Gaussian distribution of the quantum confinement levels,

$$P_0(E_i^{(0)})=\frac{1}{\sqrt{2\pi}w}\exp\left\{-\frac{(E_i^{(0)}-E_m)^2}{2w^2}\right\}. \tag{8}$$

Figure 1 shows the 2D electron density distributions in QWs calculated without allowance for the exchange energy (Hartree approximation, Fig. 1a) and with allowance for such exchange (Hartree–Fock approximation, Fig. 1b). These calculations were performed for the following parameters: $\rho_0=3\times10^{13}$ eV^{-1} cm^{-2};

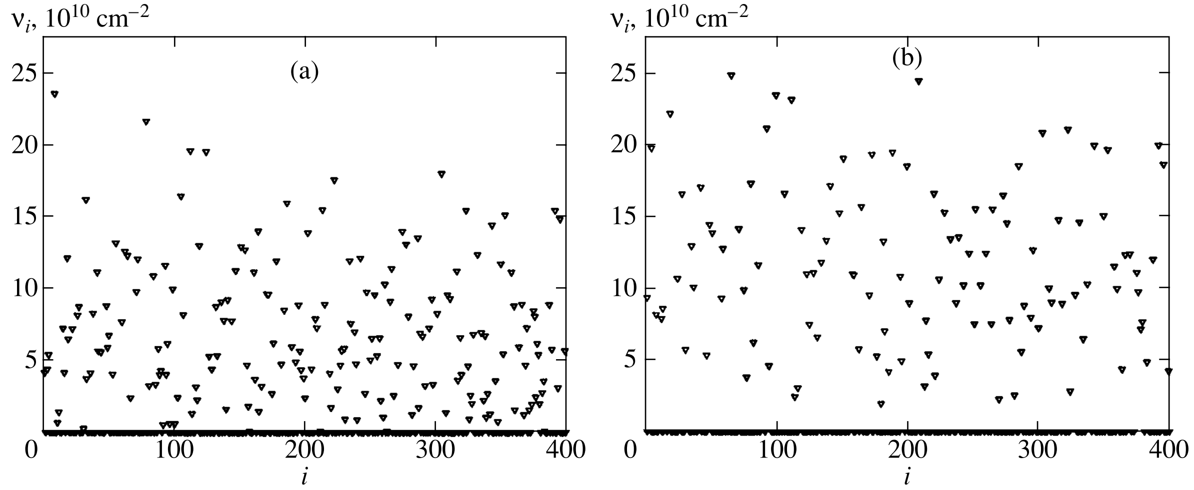


Fig. 1. Plots of the 2D electron density versus QW number calculated (a) without and (b) with allowance for the exchange interaction.

w = 10 meV; ε = 10; d = 33 Å; $N_d = 10^{17}$ cm^{-3}. These values correspond to a spin-ordered state of electrons in all QWs, whereby all densities ν_i are below the critical level $\nu_p \approx 2.9 \times 10^{11}$ cm^{-2}. As can be seen from Fig. 1, allowance for the exchange leads, on the one hand, to a decrease in the number of populated QWs and, on the other hand, to an increase in the 2D densities in the occupied QWs, the levels of which correspond to the low-energy tail of the distribution.

Previously [6, 7], it was demonstrated that allowance for the Coulomb interaction in the Hartree approximation results in the fact that the step in the density of states becomes steeper; that is, the distribution of levels becomes narrower due to the vertical screening caused by the redistribution of electrons between QWs of the IDSL. As is known [1], the 2D density of states $\rho(E)$ (per QW) is related to the distribution function $P(E)$ as

$$\rho(E) = \frac{m}{\pi\hbar^2}\int_{-\infty}^{E} P(E)dE, \qquad (9)$$

so that the density of states determines the number of levels with energies below E.

The results of calculations of the electron density distribution (Fig. 1) can be interpreted in terms of the one-particle energies. An expression for the chemical potential is as follows:

$$\mu = \frac{\partial E}{\partial \nu_i} = \frac{\nu_i}{\rho_0} + E_i^{(0)} + \sum_i V_{ij}(\nu_j - \nu_0) - \frac{3}{2}C_x\left(\frac{e^2}{\varepsilon}\right)\nu_i^{1/2}. \qquad (10)$$

The effective energies of the quantum confinement levels $\tilde{E}_i$ (edges of the lower subbands) can be defined using a relation of the type $\mu = \partial E/\partial \nu_i = \nu_i/\rho_0 + \tilde{E}_i$. Accordingly, the number of effective levels with the energies $\tilde{E}_i$ below E determines the effective density of states $\tilde{\rho}(E)$.

Figure 2 shows the seeding density of states $\rho_0(E)$ calculated without allowance for the Coulomb interaction for the distribution (8) (curve *1*), the density of states $\rho(E)$ calculated in the Hartree approximation (curve *2*), and the effective density of states $\tilde{\rho}(E)$ calculated with allowance for exchange (curve *3*). As can

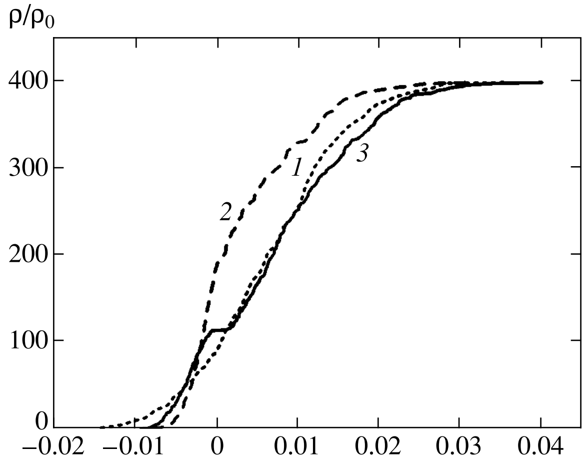


Fig. 2. Plots of the IDSL density of states versus energy (measured relative to the chemical potential): (*1*) the seeding density of states of a doped SL; (*2*) the density of states in the Hartree approximation; (*3*) the effective density of states with allowance for exchange.

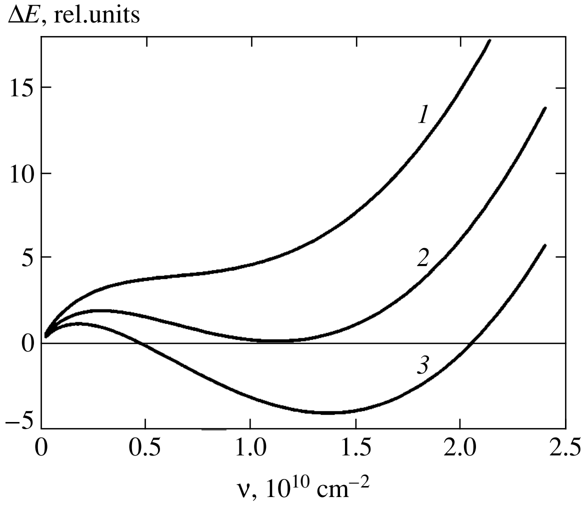


Fig. 3. A change the electron energy ΔE in an asymmetric double QW versus electron density ν in the upper QW for various average densities ν_0: (*1*) small ν_0, the upper QW is empty; (*2*) critical average density, the upper QW starts to fill; (*3*) ν_0 above the critical level.

be seen, the presence of exchange leads to a certain broadening of the distribution function (cf. curves *2* and *3*), that is, to an increase in the degree of disorder.

Figure 1 reveals another important feature in the distribution of electrons over QWs of an IDSL: the exchange significantly decreases the proportion of QWs with small electron densities. This implies that there appears a soft gap in the electron density distribution over QWs of the SL, the width of this gap being about 2×10^{10} cm^{-2}. The presence of such a gap is manifested by a plateau in the density of states (Fig. 2, curve *3*). According to distribution (9), this is accompanied by a dip in the distribution of renormalized energy levels. For the parameters adopted in this calculation, the width Δ of a step in the density of states is about 1 meV.

The appearance of a gap in the electron density is related to the fact that exchange leads to a rigid regime of QW occupation by electrons with increasing average electron density in the structure. This situation resembles the one that takes place in the absence of disorder, during the formation of electron superstructures in SLs [8]. The features of QW occupation in the SLs with disorder can be illustrated in the case of a double QW. Evidently, at a small average density ν_0, only the lower QW is occupied ($\nu = 2\nu_0$), while the upper QW is empty ($\nu = 0$). Accordingly, the system energy as a function of the electron density in the upper QW has a single minimum at $\nu = 0$ (Fig. 3, curve *1*). As the average density ν_0 increases, the upper QW becomes populated as well. The electron density in the upper well increases in a jumplike manner as soon as the energy of the side minimum becomes equal to that at $\nu = 0$ (curve *2*). At large average densities ν_0, the side minimum becomes the main (curve *3*) and then the only one. For the parameters adopted, a minimum value of the density ν corresponding to the onset of filling of the upper QW is about 1.2×10^{10} cm^{-2}. This estimate agrees reasonably well with the gap in the electron density obtained by numerical calculations.

It should be noted that the effects related to the exchange interaction are manifested only at a sufficiently low temperature. Indeed, the exchange energy decreases with increasing temperature [11, 12] and, in addition, the temperature broadening of the step in the electron energy distribution leads to a more homogeneous electron density distribution over QWs. Accordingly, an increase in the temperature makes the effects of exchange interaction significantly less pronounced.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 03-02-17378), the Federal Program “Integration” (project no. I02908/1040) and the Program “Universities of Russia.”

REFERENCES

1. J. D. Dow, Phys. Rev. B **25**, 6218 (1982).
2. A. Chomette, B. Deveaud, A. Regreny, and G. Bastard, Phys. Rev. Lett. **57**, 1464 (1986).
3. A. Chomette, B. Deveaud, J. Y. Emery, and A. Regreny, Solid State Commun. **54**, 75 (1985).
4. T. Yamamoto, M. Kasu, S. Noda, and A. Sasaki, J. Appl. Phys. **68**, 5318 (1990).
5. G. Richter, W. Stolz, P. Thomas, *et al.*, Superlatt. Microstruct. **22**, 475 (1997).
6. I. P. Zvyagin and M. A. Ormont, Fiz. Tekh. Poluprovodn. (St. Petersburg) **33**, 79 (1999) [Semiconductors **33**, 69 (1999)].
7. I. P. Zvyagin, M. A. Ormont, S. D. Baranovskii, and P. Thomas, Phys. Status Solidi B **230**, 193 (2002).
8. I. P. Zvyagin, Zh. Éksp. Teor. Fiz. **114**, 1089 (1998) [JETP **87**, 594 (1998)].
9. P. B. Visscher and L. M. Falikov, Phys. Rev. B **3**, 2541 (1971).
10. E. C. Stoner, Proc. R. Soc. London, Ser. A **165**, 372 (1938).
11. H. Totsuji, H. Tachibana, C. Totsuji, and S. Nara, Phys. Rev. B **51**, 11148 (1995).
12. A. G. Mironov, Vestn. Mosk. Univ., Ser. 3: Fiz. Astron. **5**, 61 (2001).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 1015–1019.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 1133–1137.
Original Russian Text Copyright © 2003 by Kudryashov, Yunovich.

SOLIDS
Electronic Properties

Tunneling Radiative Recombination in *p–n* Heterostructures Based on Gallium Nitride and Other $A^{III}B^{V}$ Semiconductor Compounds

V. E. Kudryashov and A. É. Yunovich*
Moscow State University, Moscow, 119992 Russia
**e-mail: yunovich@scon175.phys.msu.su*

Received April 22, 2003

Abstract—We present summarized data on the tunneling emission in *p–n* heterostructures based on GaN and on a series of cubic $A^{III}B^{V}$ semiconductors, including GaAs, InP, GaSb, and (Ga, In)Sb. The emission in *p–n* heterostructures of the InGaN/AlGaN/GaN type in a spectral interval from 1.9 to 2.7 eV predominates at small currents ($J < 0.2$ mA). The position of maximum $\hbar\omega_{\max}$ in the spectrum approximately corresponds to the applied potential difference U: $\hbar\omega_{\max} = eU$. The tunneling emission is related to a high electric field strength in GaN-based heterostructures. The radiative recombination probability is higher in the structures with piezoelectric fields. The observed spectra are compared to the spectra of tunneling emission from light-emitting diodes based on GaAs, InP, and GaSb. The experimental results for various semiconductors emitting in a broad energy range (0.5–2.7 eV) are described by the equation $\hbar\omega_{\max} = eU = 0.5–2.7$ eV. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Tunneling radiative recombination in the *p–n* junctions in semiconductor structures, accompanied by the emission of light quanta with energies below the bandgap width, is a phenomenon inverse to the interband optical absorption in a strong electric field (Franz–Keldysh effect). This phenomenon was previously studied in cubic semiconductor compounds of the $A^{III}B^{V}$ type [1, 2]. Later [3–5], it was demonstrated that the tunneling emission band predominates in the spectra of light-emitting diodes (LEDs) based on InGaN/AlGaN/GaN heterostructures with a single InGaN quantum well operating at small currents. The model of diagonal tunneling developed in [1, 2] was successfully used for description of the emission spectra of the above GaN based LED structures [3, 4]. The spectra and electrical characteristics of such LEDs were analyzed in detail in [5, 6]. More recently, tunneling radiative recombination at low currents was also studied in some other GaN-based LED structures [7–9]. It was demonstrated that this phenomenon plays an important role in the structures featuring high electric fields in active two-dimensional layers.

This paper continues the previous study [10] and summarizes the experimental results obtained for GaN heterostructures grown in various laboratories. All data were treated within the framework of the diagonal tunneling theory. The role of the piezoelectric fields is discussed. It is established that tunneling emission, both in hexagonal GaN-based structures and in *p–n* junctions based on cubic $A^{III}B^{V}$ semiconductors (GaAs, InP, GaSb), in a broad spectral range from IR to UV can be considered from a common standpoint.

2. EXPERIMENTAL RESULTS

We summarize the data obtained using LEDs based on InGaN/AlGaN/GaN heterostructures grown using metalorganic vapor phase epitaxy (MOVPE) by various research groups, including Nichia Chemical [3–6], Hewlett-Packard [7, 8], CRHEA–CNRS [9], and UniRoyal, LumiLeds [10, 11]. Special attention was devoted to the emission spectra measured at small currents ($J < 0.2$ mA, $j \leq 0.2$ A/cm^2).

The typical emission spectra of LEDs are presented in Fig. 1. Measured at small currents, the spectra display the high-energy fundamental band clearly separated from a longwave band. Figure 2 shows the evolution of the LED spectrum depending on the applied voltage U [9, 10]. At high currents ($J > 10$ mA), the fundamental violet emission band predominates. The maximum of a longwave band significantly shifts when the applied voltage is varied (the black circles in Fig. 2 indicate positions of the energy level $\hbar\omega = eU$). The integral intensity Φ of the longwave band exponentially depends on U as depicted in Fig. 3. Figure 4 shows how the position of the longwave band maximum varies with the applied voltage for various LEDs. As can be seen, the peak energy $\hbar\omega_{\max}$ is approximately equal to the applied voltage: $\hbar\omega_{\max} = eU$. The low-energy emis-

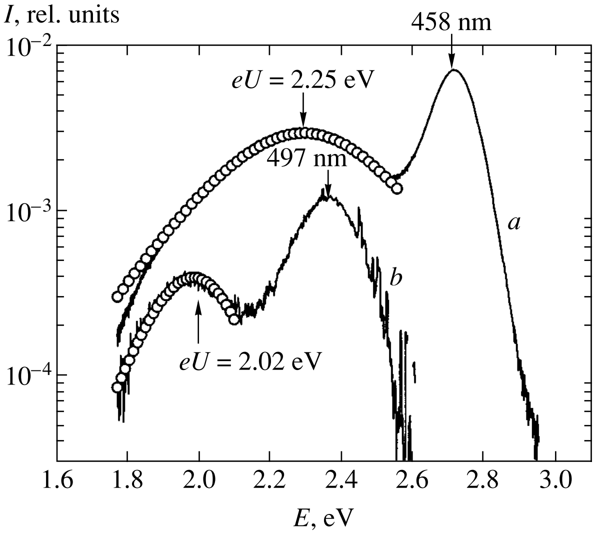


Fig. 1. Electroluminescence spectra of LEDs based on InGaN/AlGaN/GaN heterostructures with quantum wells operating at low currents *J*: (*a*) blue LED of Nichia Chemical [3–5]; (*b*) green LED of Hewlett-Packard [8]. Open circles show the approximation according to formula (1) with the parameters $E_0 = 77$ meV, $\mathbf{E} = 0.5$ MV/cm, $E_g^* = 2.7$ eV, and $mkT = 26$ meV for blue LED; $E_0 = 260$ meV, $\mathbf{E} = 2.8$ MV/cm; $E_g^* = 2.7$ eV, and $mkT = 31$ meV for green LED. Arrows indicate the positions of maxima of the fundamental and tunneling emission bands.

sion band is related to the tunneling radiative recombination.

3. DISCUSSION OF RESULTS

The theory of tunneling radiative recombination [1, 2] was originally developed for homogeneous *p–n* junctions. Assuming that the electric field in the active region is constant, the tunneling emission spectrum according to this model can be described by the equation

$$I(\hbar\omega) \propto \frac{\hbar\omega}{E_g - \hbar\omega}\,\frac{\hbar\omega - eU}{\exp\left(\frac{\hbar\omega - eU}{mkT}\right) - 1} \times \exp\left(-\frac{4}{3}\left(\frac{E_g - \hbar\omega}{E_0}\right)^{3/2}\right), \qquad (1)$$

where E_g is the effective bandgap width and *m* is a parameter depending on the effective mass ratio. The denominator E_0 in the exponent is determined from the theory of the Franz–Keldysh effect,

$$E_0 = [(\hbar/(2m^*_{cv})^{1/2})e\mathbf{E}]^{2/3}, \qquad (2)$$

where m^*_{cv} is the reduced effective mass, **E** is the electric field strength assumed to constant in the region of overlap of the electron and hole wave functions. The integral emission intensity Φ exponentially depends on the applied voltage:

$$\Phi \propto \exp(eU/E_0). \qquad (3)$$

We used Eq. (1) to describe the spectra of heterojunctions in crystals of the wurtzite type. Let us consider changes in the theory that are necessary to describe heterostructures of the InGaN/AlGaN/GaN type with InGaN quantum wells. Figure 5 shows the energy band diagram of such a structure. The electric field **E** in this system depends on the levels of doping in

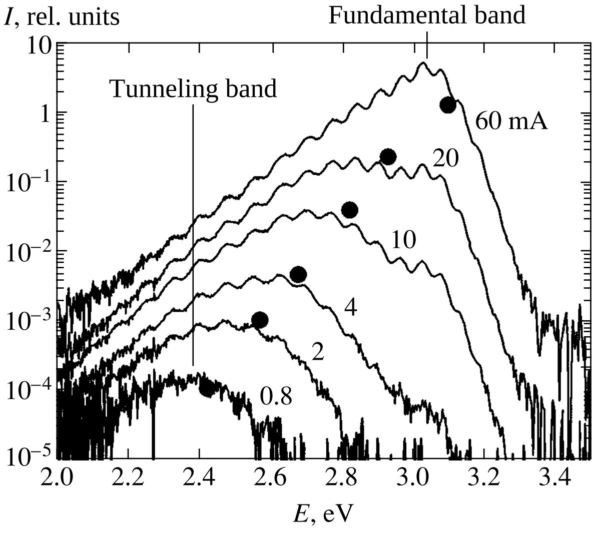


Fig. 2. Evolution of the emission spectrum of LEDs based on InGaN/AlGaN/GaN heterostructures (CRHEA samples [9, 10]) in response to the diode current varied in a broad range. Black circles indicate positions of the energy level $\hbar\omega_{max} = eU$.

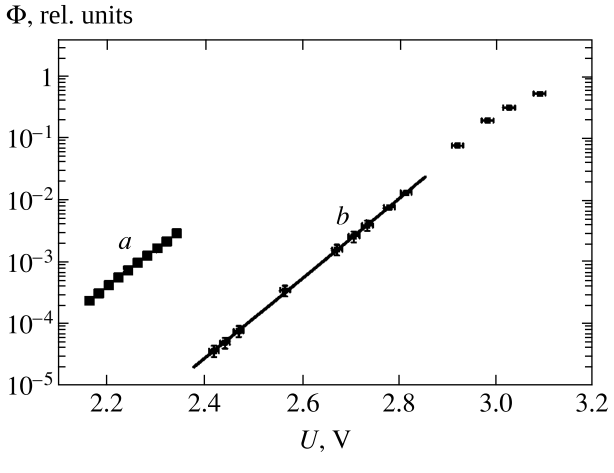


Fig. 3. Plots of the integral intensity $\Phi(U) \propto \exp(eU/E_0)$ versus applied voltage *U*: (*a*) for blue LED (Fig. 1*a*) with $E_0 = 67$ meV; (*b*) for green LED (Fig. 1*b*) with $E_0 = 71$ meV.

the *p* and *n* regions (N_A and N_D, respectively), on the AlGaN layer thickness, and on the widths of GaN barriers and quantum wells. Assuming that the contact potential difference ϕ_c in the structure is on the order of ≈3 V and the total length of the space charge region is $w \approx 80$–100 nm [6–8], the electric field can be estimated as

$$\mathbf{E}_{\phi_c} \approx \phi_c/w \approx (0.4\text{–}0.3) \times 10^6 \text{ V/cm}. \quad (4)$$

The width *w* in relation (4) depends on the distribution of charged impurities (sharp versus smooth) in a complex heterostructure. The values of $\mathbf{E}_{\phi_c}$ in the structures under consideration are significantly higher than typical electric field strengths in the *p–n* junctions in Si or GaAs. Note that **E** enters into the exponent of Eq. (1) and that the tunneling emission probability critically depends on this quantity.

The effective field strength **E** also depends on the piezoelectric fields and the spontaneous polarization fields $\mathbf{E}_p$ in the barriers and quantum wells. In the quantum wells, the polarization fields are antiparallel to the field of a *p–n* junction in the structures grown on Ga-polar (0001) *c*-planes. In the barriers, the fields are parallel to the latter field (due to charging of the heteroboundaries). The tunneling effects depend on the overlap of the electron and hole wave functions under AlGaN barriers. For this reason, the tunneling emission is more probable in the structures featuring strong electric fields in the barriers caused by polarization.

For the InGaN/AlGaN/GaN structures under consideration, the values of electric fields were estimated at $\mathbf{E}_p \approx 0.5 \times 10^6$ V/cm [11–14], which is on the same order of magnitude as the *E* value in Eq. (4). The experimental data presented in Fig. 2 were obtained for samples strongly doped with Si in the wells on the *n* side of the junction. This fact confirms the model of emission related to electron tunneling under barrier of the *p* region. In Fig. 5, this model is illustrated by a thick arrow pointing from the *n* region of the InGaN quantum well to the *p* region of the AlGaN barrier.

In a hexagonal semiconductor, the value of the reduced mass in Eq. (2) depends on the crystallographic direction:

$$m_{cv}^* = (1/m_c^* + 1/m_v^*)^{-1}, \quad (5)$$

where m_c^* and m_v^* are effective masses in the conduction and valence bands, respectively. For the analysis, we adopted the values of m_c^* for the *c* axis in $In_xGa_{1-x}N$, corresponding to the fundamental band position for each sample (by data from [15], $m^*_{c\langle 0001\rangle} = (0.20\text{–}0.15)m_0$ for $x = 0$–0.2). The value of m_v^* was taken equal to the effective mass for the upper valence band of $Al_yGa_{1-y}N$ at the point Γ ($m^*_{v1\langle 0001\rangle} = 0.27m_0$); tunneling from lower valence bands was ignored.

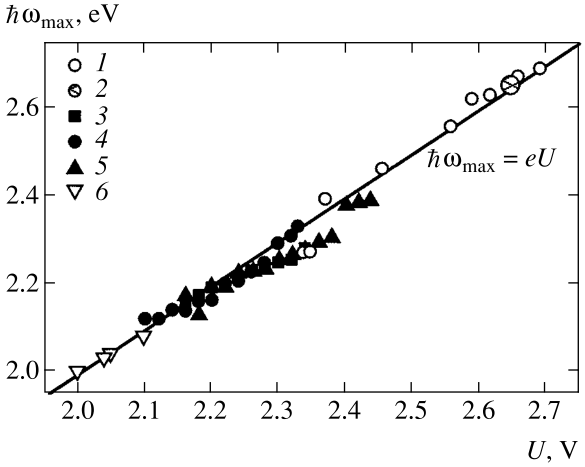


Fig. 4. A plot of the position of maximum of the tunneling emission band versus applied voltage *U* for various LEDs based on InGaN/AlGaN/GaN heterostructures: (*1*) CRHEA, violet (2001); (*2*) UniRoyal, violet (2001); (*3–5*) Nichia Chemical, blue (1996); (*6*) Hewlett-Packard, green (2000).

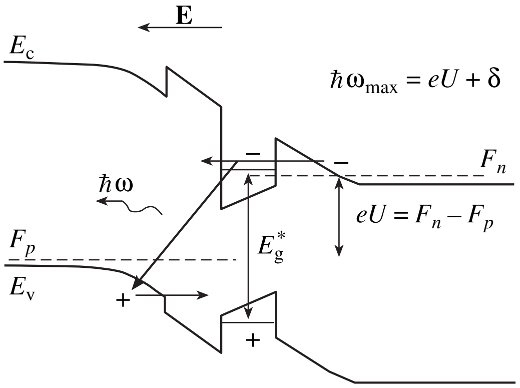


Fig. 5. The energy band diagram of an InGaN/AlGaN/GaN heterostructures with a single InGaN quantum well with allowance for piezoelectric polarization. The arrow indicates a tunneling transition from the quantum well to the *p* region with emission of the $\hbar\omega$ quantum (diagonal tunneling); $F_n - F_p$ is the difference of Fermi quasilevels.

Approximation of the experimental tunneling emission spectra by Eq. (1) with allowance for the changes considered above is illustrated in Fig. 1. The shortwave (blue) band was separated from the spectra in [5, 6]. The voltage *U* in Eq. (1) was taken equal to than measured on the structure, *U* = *V*, which is possible at small currents. In selecting the bandgap for the approximation, this value was varied within the limits of the effective bandgap E_g for the quantum wells.

As for the parameter E_0 in Eq. (1) selected for better approximation of the spectra, this value was varied within $E_0 = 0.15$–0.4 eV. This interval is significantly

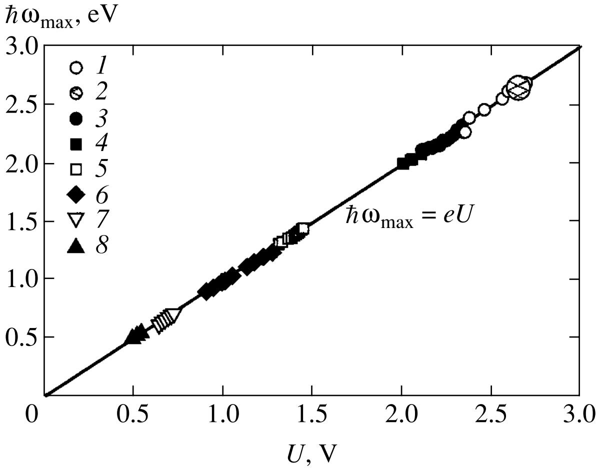


Fig. 6. A plot of the position of maximum of the tunneling emission band versus applied voltage U for various LEDs based on $A^{III}B^{V}$ semiconductors: (*1*) InGaN (CRHEA, violet, 2001); (*2*) InGaN (UniRoyal, violet, 2001); (*3*) InGaN (Nichia Chemical, blue, 1996); (*4*) InGaN (Hewlett-Packard, green, 2000); (*5*) GaAs (1966) [1]; (*6*) InP (1966) [1]; (*7*) GaSb (1969) [1, 16]; (*8*) InGaSb (1975) [17].

higher than the value $E_0 \approx 0.07$ eV determined from the dependence of the integral intensity on the applied voltage (Eq. (3) [5–7], see Fig. 3). The discrepancy can be explained as follows: the longwave slope of the spectrum depends on the tunneling emission at the band tails, which are not taken into account by the theory [1, 2]. The integral intensity determined by the density of states in the bands agrees better with the theory of parabolic bands in the junction plane and is less dependent on the tails of the density of states.

The electric field E was determined from descriptions of the tunneling radiative recombination spectra. This value amounts to $E = (1.2–5) \times 10^6$ V/cm for $E_0 = 0.15–0.4$ eV selected above for approximation of the spectra, and to $\mathbf{E} = (0.3–0.5) \times 10^6$ V/cm for $E_0 = 0.07–0.08$ eV, corresponding to the dependence of the integrated intensity Φ on the applied voltage U. The latter value is more realistic and agrees better with the values obtained from the capacitance measurements and the charge distributions (formula (4)).

The theory [1, 2] predicts the dependence of the tunneling emission peak position on the temperature. The value of $\hbar\omega_{max} - eU$ depends on the statistical factor $\exp((\hbar\omega - eU)/mkT) - 1$ in Eq. (1). According to this, the difference $\hbar\omega_{max} - eU$ is positive at low temperatures ($kT \ll E_0$) and negative at high temperatures ($kT \gg E_0$). The maximum $\hbar\omega_{max}$ coincides with eU for $kT = (1/3)E_0$ [2]. For $E_0 = 0.07–0.08$ eV, we obtain $\hbar\omega_{max} = eU$ at room temperature ($kT = 0.026$ eV), in good agreement with the experimental results (Fig. 4).

The theoretical value of the electric field in the structure takes into account neither inhomogeneities of the heterojunctions nor fluctuations of the potentials related to variations of the InGaN composition. The tunneling is more probable at "hot spots" where the electric field strength is maximum. Previously [7, 8], we observed the tunneling emission bands in LEDs with a low quantum yield. This is apparently related to a nonradiative tunneling caused by the presence of inhomogeneities in the samples with low quantum yield.

In the spectra of high-efficiency LEDs grown at the LumiLeds laboratory [11], no tunneling emission band was observed at small currents. The samples contained four InGaN quantum wells with GaN barriers doped with Si donors. The donated electrons compensate for the polarization charges in such LED structures. These data confirm the model in Fig. 5, in which the diagonal tunneling is more probable due to high piezoelectric fields in the barriers.

It is interesting to compare the results for GaN-based structures to the earlier data reported on the tunneling emission in other (cubic) $A^{III}B^{V}$ semiconductors, including GaAs [1], InP [1], GaSb [1, 16], and (Ga, In)Sb [17]. The plot of $\hbar\omega_{max}$ versus U presents remarkable results: the experimental points fit to the theoretical curve $\hbar\omega_{max} = eU$ in a broad range of wavelengths from IR to UV.

4. CONCLUSIONS

Tunneling radiative recombination (diagonal tunneling), being a phenomenon inverse to the Franz–Keldysh effect, is observed in the *p–n* junctions of GaN-based heterostructures with quantum wells at low currents and high electric field strengths ($E = (0.3–1) \times 10^6$ V/cm). The probability of this phenomenon increases in AlGaN/InGaN heterojunctions with pronounced piezoelectric and spontaneous polarizations.

The experimentally observed emission spectra of GaN-based LEDs at low currents are described well by the theory of tunneling emission. The position of maximum $\hbar\omega_{max}$ of the tunneling emission band at room temperature approximately corresponds to the potential difference U applied to the InGaN/AlGaN/GaN structure: $\hbar\omega_{max} = eU$ in the range from 1.9 to 2.7 eV.

Tunneling radiative recombination is a basic phenomenon for the *p–n* junctions in both homo- and heterostructures based on the direct-band semiconductors of the $A^{III}B^{V}$ type in a broad spectral range from IR to UV ($\hbar\omega = eU = 0.5–2.7$ eV).

ACKNOWLEDGMENTS

The authors are grateful to M. Leroux for fruitful discussion of the results.

The financial support from CNRS (program 2001, project no. 9562) and the Moscow City Committee on Science and Technology (project no. GL-120.02) is gratefully acknowledged.

REFERENCES

1. A. É. Yunovich and A. B. Ormont, Zh. Éksp. Teor. Fiz. **51**, 1292 (1966) [Sov. Phys. JETP **24**, 869 (1966)].
2. T. N. Morgan, Phys. Rev. **148**, 890 (1966).
3. K. G. Zolina, V. E. Kudryashov, A. N. Turkin, *et al.*, MRS Internet J. Nitride Semicond. Res. **1**, 11 (1996).
4. K. G. Zolina, V. E. Kudryashov, A. N. Turkin, and A. É. Yunovich, Fiz. Tekh. Poluprovodn. (St. Petersburg) **31**, 1055 (1997) [Semiconductors **31**, 901 (1997)].
5. V. E. Kudryashov, K. G. Zolina, A. N. Kovalev, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **31**, 1304 (1997) [Semiconductors **31**, 1123 (1997)].
6. A. N. Kovalev, F. I. Manyakhin, V. E. Kudryashov, *et al.*, MRS Internet J. Nitride Semicond. Res. **2**, 11 (1997).
7. A. E. Yunovich, A. N. Kovalev, V. E. Kudryashov, *et al.*, Mater. Res. Soc. Symp. Proc. **449**, 1167 (1997).
8. V. E. Kudryashov, S. S. Mamakin, A. N. Turkin, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **34**, 861 (2001) [Semiconductors **35**, 827 (2001)].
9. P. Vennégues, B. Beaumont, S. Haffouz, *et al.*, J. Cryst. Growth **187**, 167 (1998).
10. A. E. Yunovich, V. E. Kudryashov, A. N. Turkin, *et al.*, in *Abstracts of MRS Symp. Proc.* **723**, L11.4, 647 (2002).
11. A. E. Yunovich, S. S. Mamakin, F. I. Manyakhin, *et al.*, Mater. Res. Soc. Symp. Proc. **711** (K2.4), 71 (2002).
12. P. Lefebre, J. Allegre, B. Gil, *et al.*, Phys. Rev. B **59**, 15363 (1999).
13. S. F. Chichibu, S. P. DenBaars, K. Wada, *et al.*, Mater. Sci. Eng. B **59**, 298 (1999).
14. A. Hangleiter, J. S. Im, H. Kollmer, *et al.*, MRS Internet J. Nitride Semicond. Res. **4S11**, G6.20 (1999).
15. B. Gil, *Group III Nitride Semiconductor Compounds: Physics and Applications* (Clarendon Press, Oxford, 1998).
16. V. M. Stuchebnikov and A. É. Yunovich, Fiz. Tekh. Poluprovodn. (Leningrad) **3**, 1293 (1969) [Sov. Phys. Semicond. **3**, 1078 (1969)].
17. A. I. Lebedev, I. A. Strel'nikova, and A. É. Yunovich, Fiz. Tekh. Poluprovodn. (Leningrad) **10**, 1304 (1976) [Sov. Phys. Semicond. **10**, 773 (1976)].

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 1020–1031.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 1138–1148.
Original Russian Text Copyright © 2003 by Balagurov, Kashin.

SOLIDS
Electronic Properties

Structural Field and Current Fluctuations in the Problem of Conductivity of Inhomogeneous Media: Theory and Numerical Experiment

B. Ya. Balagurov* and V. A. Kashin
Emanuel Institute of Biochemical Physics, Russian Academy of Sciences,
ul. Kosygina 4, Moscow, 119991 Russia
**e-mail: balagurov@deom.chph.ras.ru*

Received April 29, 2003

Abstract—Structural fluctuations Δ_E^2 (in the squares of the electric field strength) and Δ_j^2 (in current density), averaged over the sample volume, are considered in various inhomogeneous conducting systems. Explicit expressions are obtained for Δ_E^2 and Δ_j^2 in the case of a weakly inhomogeneous medium as well as in an approximation linear in one of the components. The critical behavior of quantities Δ_E^2 and Δ_j^2 for a system with the metal–insulator phase transition is considered in the framework of the standard phenomenological similarity hypothesis. Structural fluctuations are determined and tabulated in the course of a numerical experiment on disordered lattices in a wide range of variation of the parameters of the problem. *© 2003 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

The distribution of current and field is obviously nonuniform in an inhomogeneous conducting medium. As a result, the value of the squared electric field E strength or current density j averaged over the sample volume is not equal to the square of the average value $\langle E\rangle$ or $\langle j\rangle$. For this reason, structural fluctuations of the field, Δ_E^2, and of the current, Δ_j^2, calculated by formulas (3) and (4) from Section 2, can serve as the characteristics of the nonuniformity in the distribution of these quantities and also (indirectly) of the system inhomogeneity. In order to avoid confusion, it should be emphasized that quantities Δ_E^2 and Δ_j^2 are determined by the geometry (structure) of the medium and are not associated with temperature fluctuations, which also exist in homogeneous systems and have, as a rule, a small order of magnitude.

Systems with the metal–insulator phase transition are strongly inhomogeneous; it can be expected that field and current fluctuations in such systems are especially strong in the vicinity of the critical concentration (percolation threshold) p_c. For example, it follows from the exact expressions derived by Dykhne [1] for 2D two-component system with critical composition ($p = p_c = 1/2$) that the values of Δ_E^2 and Δ_j^2 increase indefinitely for $h = \sigma_2/\sigma_1 \longrightarrow 0$, where σ_i ($i = 1, 2$) is the conductivity of the ith component. According to [2], an indefinite increase in fluctuations must also be observed for $p \longrightarrow p_c$ and $h = 0$ both in the 2D and in the 3D case. General expressions for the quantities Δ_E^2 and Δ_j^2 obtained in [3] connect these quantities with the dimensionless effective conductivity f. The corresponding relations make it possible to generalize the results obtained in [1, 2] and provide a complete description of the critical behavior of structural fluctuations in field and current.

We will consider the structural fluctuations in field, Δ_E^2, and current, Δ_j^2, in a wide range of variation of the concentration p and the parameter $h = \sigma_2/\sigma_1$. Explicit analytic expressions will be obtained for Δ_E^2 and Δ_j^2 in the case of a weakly inhomogeneous medium as well as in the approximation linear in the impurity concentration. The general relations derived in [3] are used to determine (in the framework of the similarity hypothesis [4]) the critical behavior of structural fluctuations.

We will also consider the results of a numerical experiment for disordered lattices (the problem of associations), in which the effective conductivity σ_e and the quantities Δ_E^2 and Δ_j^2 were determined. Structural fluctuations are represented in graphic form in the entire range of variation of concentration p for a number of fixed values of parameter h. From the behavior of quantities Δ_E^2 and Δ_j^2 in the vicinity of the percolation threshold p_c, their critical indices are determined, which match the corresponding indices of effective conductivity.

It follows from the analytic and numerical results obtained in this study that structural fluctuations Δ_E^2 and Δ_j^2 in a certain range of parameters p and h are connected via a certain relation: $\Delta_j^2 = \Delta_E^2$ in the 2D case and $\Delta_j^2 = 2\Delta_E^2$ in the 3D case. The substitution of the general expressions for Δ_E^2 and Δ_j^2 (see formulas (13) and (14) in the text) into these equalities leads to differential equations for the dimensionless effective conductivity f. The solutions to these equations are expressions coinciding with the corresponding formulas for f, derived in the so-called theory of an effective medium [5].

2. GENERAL RELATIONS

In the problem of conductivity of a heterogeneous medium, it is necessary to solve the equations for direct current,

$$\mathrm{curl}\mathbf{E} = 0, \quad \mathrm{div}\mathbf{j} = 0, \quad \mathbf{j} = \hat{\sigma}(\mathbf{r})\mathbf{E}, \tag{1}$$

where $\mathbf{E} = \mathbf{E}(\mathbf{r})$ is the electric field strength, $\mathbf{j} = \mathbf{j}(\mathbf{r})$ is the current density, and $\hat{\sigma}(\mathbf{r})$ is the conductivity tensor for the medium at point $\mathbf{r}$. The effective conductivity tensor $\hat{\sigma}_e$ is defined in the usual manner:

$$\langle \mathbf{j} \rangle = \hat{\sigma}_e \langle \mathbf{E} \rangle. \tag{2}$$

Here, $\langle \ldots \rangle$ denotes averaging over volume V of the sample for $V \longrightarrow \infty$. Apart from conductivity, it is also important to analyze the quantities

$$\Delta_E^2 = \frac{\langle (\mathbf{E} - \langle \mathbf{E} \rangle)^2 \rangle}{(\langle \mathbf{E} \rangle)^2}, \tag{3}$$

$$\Delta_j^2 = \frac{\langle (\mathbf{j} - \langle \mathbf{j} \rangle)^2 \rangle}{(\langle \mathbf{j} \rangle)^2}, \tag{4}$$

which will be called the structural fluctuations of the electric field strength and of current density.

In view of the familiar identity (see, for example, [1])

$$\langle \mathbf{E} \cdot \mathbf{j} \rangle = \langle \mathbf{E} \rangle \cdot \langle \mathbf{j} \rangle, \tag{5}$$

the effective conductivity of an isotropic medium can be written in the form

$$\sigma_e = \frac{\langle \sigma \mathbf{E}^2 \rangle}{(\langle \mathbf{E} \rangle)^2}. \tag{6}$$

For an N-component system (composite), in accordance with relation (6), we have

$$\sigma_e = \sum_{i=1}^{N} \sigma_i \psi_i, \quad \psi_i = \langle \mathbf{E}^2 \rangle^{(i)} / (\langle \mathbf{E} \rangle)^2. \tag{7}$$

Here, $\langle \ldots \rangle^{(i)}$ indicates an integral over the volume V_i of the ith component, divided by the sample volume V, and σ_i is the conductivity of the ith component. It also follows from relation (5) that

$$\psi_i = \frac{\partial \sigma_e}{\partial \sigma_i}. \tag{8}$$

We can verify the validity of this equality, multiplying it by σ_i and summing over all values of i. This leads to the identity

$$\sum_i \sigma_i \psi_i = \sum_i \sigma_i \frac{\partial \sigma_e}{\sigma_i},$$

in which the left-hand side is equal to σ_e in accordance with relations (7), while the right-hand side is equal to σ_e by virtue of the Euler theorem on homogeneous functions. Expressions (3) and (4) combined with Eqs. (7) and (8) lead to

$$\Delta_E^2 = \sum_i \psi_i - 1, \quad \Delta_j^2 = \frac{1}{\sigma_e^2} \sum_i \sigma_i^2 \psi_i - 1 \tag{9}$$

or

$$\Delta_E^2 = \sum_i \frac{\partial \sigma_e}{\partial \sigma_i} - 1, \quad \Delta_j^2 = \frac{1}{\sigma_e^2} \sum_i \sigma_i^2 \frac{\partial \sigma_e}{\partial \sigma_i} - 1. \tag{10}$$

For an isotropic binary system, we can write the effective conductivity σ_e in the form

$$\sigma_e(p; \sigma_1, \sigma_2) = \sigma_1 f(p, h), \quad h = \sigma_2 / \sigma_1, \tag{11}$$

where p is the concentration (fraction of the occupied volume) of the first component. In this case, expressions (8)–(10) give [3]

$$\psi_1 = f - hf', \quad \psi_1 = f', \quad f' \equiv \partial f(p, h) / \partial h, \tag{12}$$

$$\Delta_E^2 = f + (1 - h) f' - 1, \tag{13}$$

$$\Delta_j^2 = [f(1 - f) - h(1 - h) f'] f^{-2}. \tag{14}$$

Eliminating derivative f' from Eqs. (13) and (14), we obtain the relation [2]

$$h\Delta_E^2 + f^2 \Delta_j^2 = (1 - f)(f - h). \tag{15}$$

For $h = 0$, relation (15) implies (for $p \geq p_c$) that

$$f(p, 0) = (1 + \Delta_j^2)^{-1}. \tag{16}$$

Since $f(p, 0) \longrightarrow 0$ as p tends to the critical concentration p_c, we have $\Delta_j^2 \longrightarrow \infty$ for $h = 0$ and $p \longrightarrow p_c + 0$.

For a two-dimensional randomly inhomogeneous system with the critical composition ($p = p_c = 1/2$), we have, in accordance with [1],

$$f\left(\frac{1}{2}, h\right) = \sqrt{h}. \tag{17}$$

Substituting this relation into Eqs. (13) and (14), we obtain

$$p = \frac{1}{2}: \quad \Delta_E^2 = \Delta_j^2 = \frac{1}{2}(h^{-1/4} - h^{1/4})^2, \tag{18}$$

which coincides with the corresponding result obtained in [1]. It follows from relation (18) that in this case $\Delta_E^2 = \Delta_j^2 \longrightarrow \infty$ for $p = p_c$ and for $h \longrightarrow 0$. It should be noted that expressions (17) and (18) are valid for arbitrary values of h.

For an anisotropic N-component medium, instead of relation (8), we have

$$\langle E_\alpha E_\beta \rangle^{(i)} = \frac{\partial \sigma_{\gamma\delta}^e}{\partial \sigma_{\alpha\beta}^i} \langle E_\gamma \rangle \langle E_\delta \rangle, \tag{19}$$

where $\langle \ldots \rangle^{(i)}$ is the same as in relation (7), $\sigma_{\gamma\delta}^e = (\hat{\sigma}_e)_{\gamma\delta}$, and $\sigma_{\alpha\beta}^i = (\hat{\sigma}_i)_{\alpha\beta}$. In this case, we have

$$\langle \mathbf{E}^2 \rangle = \sum_i \left\{ \frac{\partial \sigma_{\gamma\delta}^e}{\partial \sigma_{xx}^i} + \frac{\partial \sigma_{\gamma\delta}^e}{\partial \sigma_{yy}^i} + \frac{\partial \sigma_{\gamma\delta}^e}{\partial \sigma_{zz}^i} \right\} \langle E_\gamma \rangle \langle E_\delta \rangle, \tag{20}$$

$$\langle \mathbf{j}^2 \rangle = \sum_i \sigma_{\alpha\beta}^i \sigma_{\alpha\gamma}^i \frac{\partial \sigma_{\mu\nu}^e}{\partial \sigma_{\beta\gamma}^i} \langle E_\mu \rangle \langle E_\nu \rangle. \tag{21}$$

It should be noted that the quantities $\langle \mathbf{E}^2 \rangle$ and $\langle \mathbf{j}^2 \rangle$ in the anisotropic case depend on the direction of $\langle \mathbf{E} \rangle$.

3. WEAKLY INHOMOGENEOUS MEDIUM

For an isotropic weakly inhomogeneous medium, we have, in accordance with [6],

$$\sigma_e = \langle \sigma \rangle \left\{ 1 - \frac{1}{D} \frac{\langle (\sigma - \langle \sigma \rangle)^2 \rangle}{(\langle \sigma \rangle)^2} \right\}, \tag{22}$$

where D is the dimension of space ($D = 2, 3$). Using the method developed in [6], we can also find the structural fluctuations

$$\Delta_E^2 = \frac{1}{D} \frac{\langle (\sigma - \langle \sigma \rangle)^2 \rangle}{(\langle \sigma \rangle)^2}, \tag{23}$$

$$\Delta_j^2 = \frac{D-1}{D} \frac{\langle (\sigma - \langle \sigma \rangle)^2 \rangle}{(\langle \sigma \rangle)^2}. \tag{24}$$

Expressions (22)–(24) are valid for an arbitrary coordinate dependence of conductivity σ; $\langle \ldots \rangle$ is the average over the sample volume or, which is the same, over the ensemble. In particular, for an N-component medium, we have

$$\langle (\ldots) \rangle = \sum_{i=1}^{N} p_i (\ldots)_i, \quad \sum_{i=1}^{N} p_i = 1, \tag{25}$$

where p_i is the concentration of the ith component.

For a two-component system, the system of equations (22)–(24) leads to

$$f = 1 - (1-p)(1-h) - \frac{1}{D} p(1-p)(1-h)^2, \tag{26}$$

$$\Delta_E^2 = \frac{1}{D} p(1-p)(1-h)^2, \tag{27}$$

$$\Delta_j^2 = (D-1)\Delta_E^2, \tag{28}$$

where h is the same as in Eq. (11). Formulas (26)–(28) are valid to within the terms on the order of $(1-h)^2$ inclusively in the entire concentration range. In particular, for $D = 2$ and $p = 1/2$, we have, in accordance with Eqs. (26)–(28),

$$f = 1 - \frac{1}{2}(1-h) - \frac{1}{8}(1-h)^2,$$

$$\Delta_E^2 = \Delta_j^2 = \frac{1}{8}(1-h)^2.$$

The same results follow from exact formulas (17) and (18) for $|1-h| \ll 1$. It can easily be verified that the substitution of Eq. (26) into the general relations (13) and (14) leads to expressions (27) and (28).

For a weakly inhomogeneous anisotropic medium, the method developed in [6] (see also [7]) gives

$$\sigma_{\alpha\beta}^e = \langle \sigma_{\alpha\beta} \rangle - \langle \delta\sigma_{\alpha\gamma} \delta\sigma_{\beta\delta} \rangle q_{\gamma\delta}, \quad \delta\hat{\sigma} = \hat{\sigma} - \langle \hat{\sigma} \rangle, \tag{29}$$

$$q_{\alpha\beta} = \overline{\left[\frac{m_\alpha m_\beta}{\mathbf{m} \langle \hat{\sigma} \rangle \mathbf{m}} \right]}. \tag{30}$$

Here, **m** is the unit vector and the bar in Eq. (30) denotes averaging over angles. If the principal axes of tensor $\langle\hat{\sigma}\rangle$ coincide with the Cartesian coordinate axes, matrix $q_{\alpha\beta}$ is diagonal and

$$q_{xx} = \frac{n^{(x)}}{\langle\sigma_{xx}\rangle}, \quad q_{yy} = \frac{n^{(y)}}{\langle\sigma_{yy}\rangle}, \quad q_{zz} = \frac{n^{(z)}}{\langle\sigma_{zz}\rangle}, \qquad (31)$$

where $n^{(\alpha)}$ are the depolarization coefficients of the ellipsoid with the semiaxes

$$a_x = \frac{1}{\langle\sigma_{xx}\rangle}, \quad a_y = \frac{1}{\langle\sigma_{yy}\rangle}, \quad a_z = \frac{1}{\langle\sigma_{zz}\rangle}. \qquad (32)$$

In the same approximation, we find that

$$\langle(\mathbf{E} - \langle\mathbf{E}\rangle)^2\rangle = Q_{\alpha\beta}\langle\delta\sigma_{\alpha\gamma}\delta\sigma_{\beta\delta}\rangle\langle E_\gamma\rangle\langle E_\delta\rangle, \qquad (33)$$

$$\begin{aligned}\langle(\mathbf{j} - \langle\mathbf{j}\rangle)^2\rangle &= \langle\sigma_{\alpha\beta}\rangle\langle\sigma_{\alpha\gamma}\rangle Q_{\beta\gamma\mu\nu}\langle\delta\sigma_{\mu\delta}\delta\sigma_{\nu\rho}\rangle\langle E_\delta\rangle\langle E_\rho\rangle \\ &- 2\langle\sigma_{\alpha\beta}\rangle q_{\beta\gamma}\langle\delta\sigma_{\alpha\mu}\delta\sigma_{\gamma\nu}\rangle\langle E_\mu\rangle\langle E_\nu\rangle \qquad (34) \\ &+ \langle\delta\sigma_{\alpha\beta}\delta\sigma_{\alpha\gamma}\rangle\langle E_\beta\rangle\langle E_\gamma\rangle.\end{aligned}$$

Here, $q_{\alpha\beta}$ is the same as in Eq. (30) and

$$Q_{\alpha\beta\gamma\delta} = \overline{\left[\frac{m_\alpha m_\beta m_\gamma m_\delta}{(\mathbf{m}\langle\hat{\sigma}\rangle\mathbf{m})^2}\right]} = -\frac{\partial q_{\alpha\beta}}{\partial\langle\sigma_{\gamma\delta}\rangle}, \qquad (35)$$

$$Q_{\alpha\beta} = Q_{\alpha\beta\gamma\gamma}. \qquad (36)$$

It can be verified that, in the case of an isotropic medium, expressions (22)–(24) follow from Eqs. (29), (33), and (34). It should be noted that results (22)–(24) are also valid for disordered lattices (the problem of associations), e.g., for a square lattice for $D = 2$ and for a simple cubic lattice for $D = 3$.

4. APPROXIMATION LINEAR IN CONCENTRATION

For an arbitrary relation between the conductivities of the components of a medium, the application of the method described in the preceding section for determining the quantities σ_e, Δ_E^2 and Δ_j^2 involves certain difficulties since in this case we must sum an infinite series in perturbation theory. However, we can calculate the effective conductivity σ_e and the structural fluctuations Δ_E^2 and Δ_j^2 in a quite general form for a binary system with a small concentration of one of the components.

Let us suppose that an isotropic medium with conductivity σ_1 contains an inclusion (body) of an arbitrary shape with conductivity σ_2, which is in an external uniform electric field E_0. At large distances from the body, the expression for potential $\varphi(r)$ has the conventional dipole form

$$r \longrightarrow \infty: \varphi = -\mathbf{E}_0 \cdot \mathbf{r} + \frac{\mathbf{p} \cdot \mathbf{r}}{r^3} + \ldots \qquad (37)$$

in the 3D case and

$$r \longrightarrow \infty: \varphi = -\mathbf{E}_0 \cdot \mathbf{r} + 2\frac{\mathbf{p} \cdot \mathbf{r}}{r^2} + \ldots \qquad (38)$$

in the 2D case. Here,

$$p_\alpha = \Lambda_{\alpha\beta}E_{0\beta}, \quad \Lambda_{\alpha\beta} = v\alpha_{\alpha\beta}, \qquad (39)$$

where v is the volume (area for $D = 2$) of the body (inclusion). In the problem of permittivity, **p** is the electric dipole moment of the body and $\hat{\alpha}$ is the dimensionless tensor of dipole polarizability, which depends on the shape of the inclusion and on the ratio $\varepsilon_2/\varepsilon_1$ of the permittivities of the components. In the problem of conductivity considered here, we must replace the ratio $\varepsilon_2/\varepsilon_1$ by $h = \sigma_2/\sigma_1$.

In the approximation linear in the concentration c of the second component, the effective conductivity of the binary system has the form (see, for example, [8])

$$\sigma_e = \sigma_1\left\{1 + c\frac{4\pi}{D}\mathrm{Tr}\hat{\alpha}\right\}. \qquad (40)$$

Here and below, we assume that the inclusions are identical and oriented at random. In the same approximation, for quantities $\langle\mathbf{E}\rangle$ we have (cf. [8])

$$\langle\mathbf{E}\rangle = \mathbf{E}_0\left\{1 - c\frac{4\pi}{D^2}\mathrm{Tr}\hat{\alpha}\right\}, \qquad (41)$$

where $\mathbf{E}_0$ is the electric field strength in the medium in the absence of inclusions. While deriving formulas (40) and (41), we have used relations (A.5) and (A.6).

The expression for function ψ_1 (see Eq. (7)) can be transformed as

$$\begin{aligned}(\langle\mathbf{E}\rangle)^2\psi_1 &= \frac{1}{V}\int_{V_1}\mathbf{E}^2dV \equiv \frac{1}{V}\int_{V_1}(\mathbf{E} - \langle\mathbf{E}\rangle)^2dV \\ &+ 2\langle\mathbf{E}\rangle\frac{1}{V}\int_{V_1}\mathbf{E}dV - p(\langle\mathbf{E}\rangle)^2,\end{aligned} \qquad (42)$$

where $p = V_1/V$. Taking into account relation (A.17), we obtain

$$\begin{aligned}\frac{1}{V}\int\limits_{V_1}(\mathbf{E}-\langle\mathbf{E}\rangle)^2 dV &= \frac{1}{V}\sum_a\int\limits_{V_e}(\mathbf{E}-\langle\mathbf{E}\rangle)^2 dV \\ &= -c\left\{\frac{4\pi}{D}\left[\frac{1+h}{1-h}\mathrm{Tr}\hat{\alpha}+h\frac{\partial}{\partial h}\mathrm{Tr}\hat{\alpha}\right]+1\right\}(\langle\mathbf{E}\rangle)^2.\end{aligned} \tag{43}$$

Using relation (A.5), we find that

$$\begin{aligned}\frac{1}{V}\int\limits_{V_1}\mathbf{E}dV &= \langle\mathbf{E}\rangle-\frac{1}{V}\sum_a\int\limits_{v}\mathbf{E}dV \\ &= \left\{1+c\frac{4\pi}{D}\frac{1}{1-h}\mathrm{Tr}\hat{\alpha}\right\}\langle\mathbf{E}\rangle.\end{aligned} \tag{44}$$

In relations (43) and (44), summation is carried out over all N_a inclusions so that $c = vN_a/V$. In expression (43), $V_e = V_a - v$, where $V_a = V/N_a$ is the sample volume per inclusion ($V_a \gg v$). Substituting Eqs. (43) and (44) into relation (42), we obtain

$$\psi_1 = 1+c\frac{4\pi}{D}\left\{\mathrm{Tr}\hat{\alpha}-h\frac{\partial}{\partial h}\mathrm{Tr}\hat{\alpha}\right\}. \tag{45}$$

Accordingly, for ψ_2, we obtain, taking into account Eq. (A.12),

$$\psi_2 = c\frac{4\pi}{D}\frac{\partial}{\partial h}\mathrm{Tr}\hat{\alpha}. \tag{46}$$

Substituting relations (45) and (46) into Eq. (9), we finally obtain

$$\Delta_E^2 = c\frac{4\pi}{D}\left\{\mathrm{Tr}\hat{\alpha}+(1-h)\frac{\partial}{\partial h}\mathrm{Tr}\hat{\alpha}\right\}, \tag{47}$$

$$\Delta_j^2 = -c\frac{4\pi}{D}\left\{\mathrm{Tr}\hat{\alpha}+h(1-h)\frac{\partial}{\partial h}\mathrm{Tr}\hat{\alpha}\right\}. \tag{48}$$

It can easily be seen that expressions (40) and (45)–(48) satisfy relations (12)–(14). For a "weakly inhomogeneous" body ($|\sigma_1-\sigma_2|/\sigma_1 \ll 1$), we have, in accordance with [8], the relation

$$\mathrm{Tr}\hat{\alpha} = -\frac{1}{4\pi}[D(1-h)+(1-h)^2], \tag{49}$$

which is valid to within terms on the order of $(1-h)^2$, inclusively. The substitution of this relation into Eqs. (40), (47), and (48) leads to expressions (26)–(28) in the approximation linear in $c = 1 - p$.

For inclusions of a spherical (circular for $D = 2$) shape, $\alpha_{\alpha\beta} = \alpha\delta_{\alpha\beta}$, where

$$\alpha = -\frac{D}{4\pi}\frac{1-h}{D-1+h};$$

consequently, it follows from Eqs. (40), (47), and (48) ($c \ll 1$) that

$$f = 1-cD\frac{1-h}{D-1+h}, \tag{50}$$

$$\Delta_E^2 = cD\left(\frac{1-h}{D-1+h}\right)^2, \quad \Delta_j^2 = (D-1)\Delta_E^2. \tag{51}$$

Expression (50) is in accordance with the result obtained in [6]. In the discrete case (problem of associations), results (50) and (51) are valid for square ($D = 2$) and simple cubic ($D = 3$) lattices.

5. CRITICAL REGION

For a randomly inhomogeneous system with a metal–insulator phase transition, the dimensionless effective conductivity f in the critical region ($h \ll 1$, $|\tau| \ll 1$, where $\tau = (p - p_c)/p_c$) in the framework of the similarity hypothesis has the form [4] (see also [9, 10])

$$\begin{gathered}\tau > 0, \quad \Delta_0 \ll \tau \ll 1: \\ f = \tau^t\left\{A_0 + A_1\frac{h}{\tau^{t/s}}+\dots\right\},\end{gathered} \tag{52}$$

$$|\tau| \ll \Delta_0:\ f = h^s\left\{a_0 + a_1\frac{\tau}{h^{s/t}}+\dots\right\}, \tag{53}$$

$$\begin{gathered}\tau < 0, \quad \Delta_0 \ll |\tau| \ll 1: \\ f = \frac{h}{(-\tau)^q}\left\{B_1 + B_2\frac{h}{(-\tau)^{t/s}}+\dots\right\},\end{gathered} \tag{54}$$

$$\frac{t}{s} = t + q. \tag{55}$$

Here, $\Delta_0 = h^{s/t}$ is the size of the smearing region [4] and p_c is the critical concentration (percolation threshold). Critical indices t, s, and q are positive ($s < 1$) and are connected via relation (55), which is a consequence of the similarity hypothesis [4]. The numerical coefficients in expansions (52)–(54) are on the order of unity; coefficients A_0, A_1, a_0, a, and B_1 are positive and $B_2 < 0$. In accordance with relation (17), $s = 1/2$ and $a_0 = 1$ in the 2D case. It follows from the reciprocity relation [1] that $t = q$, $A_0B_1 = 1$, $A_0B_2 + A_1B_1 = 0$, … for a 2D system [9].

Taking into account relations (52)–(54), we obtain from Eqs. (13) and (14) the critical behavior of the structural fluctuations Δ_E^2 and Δ_j^2 (we write only the principal terms in the corresponding expansions):

$$\tau > 0, \quad \Delta_0 \ll \tau \ll 1:$$

$$\Delta_E^2 \approx \frac{A_1}{\tau^{q'}}, \quad \Delta_j^2 \approx \frac{1}{A_0}\frac{1}{\tau^{t}}, \tag{56}$$

$$|\tau| \ll \Delta_0: \Delta_E^2 \approx s a_0 \frac{1}{h^{1-s}}, \quad \Delta_j^2 \approx \frac{1-s}{a_0}\frac{1}{h^{s}}, \tag{57}$$

$$\tau < 0, \quad \Delta_0 \ll |\tau| \ll 1:$$

$$\Delta_E^2 \approx \frac{B_1}{(-\tau)^{q}}, \quad \Delta_j^2 \approx -\frac{B_2}{B_1}\frac{1}{(-\tau)^{t'}}; \tag{58}$$

here,

$$q' = \frac{t}{s} - t, \quad t' = \frac{t}{s} - q. \tag{59}$$

If relation (55) is satisfied, we have $t' = t$ and $q' = q$ so that the critical indices of the quantity Δ_E^2 (as well as Δ_j^2) above and below the phase-transition point (outside the smearing region) turn out to be identical. For a 2D system, in addition, the ratio $\Delta_E^2/\Delta_j^2 \approx$ const in this case in all three regions where formulas (56)–(58) are valid.

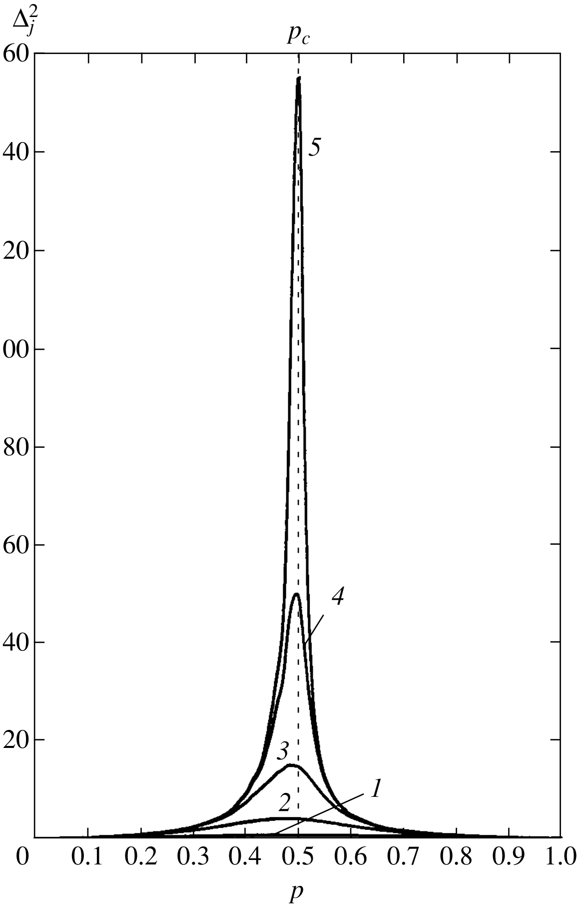


Fig. 1. Quantity Δ_j^2 as a function of concentration p for five values of argument h: $h = 10^{-1}$ (*1*), 10^{-2} (*2*), 10^{-3} (*3*), 10^{-4} (*4*), and 10^{-5} (*5*). Two-dimensional case.

6. RESULTS OF NUMERICAL EXPERIMENT

The numerical experiment on disordered lattices (problem of associations) was carried out in the standard manner [5] (see also [9, 10]). We ascribe potential V_r to each lattice site r. The conductivity of the bond between neighboring sites assumed the value $\sigma_1 = 1$ with probability p (“pure” bonds) and $\sigma_2 = h$ with probability $1 - p$ (“defective” bonds). For a certain fixed concentration p, a “realization” was observed; i.e., the corresponding fraction of $1 - p$ bonds in the initially pure lattice was randomly replaced by defective bonds. Then the system of Kirchhoff equations for preset boundary conditions was solved on a computer. The potentials V_r obtained in this way were used for calculating various effective parameters of the lattice model, e.g., the dimensionless conductivity f and the structural fluctuations Δ_E^2 and Δ_j^2. (The procedure of the numerical experiment is described in greater detail in [9, 10].) In the 3D case, we considered a simple cubic lattice with a size of 51 × 51 × 51 sites, while in the 2D case, we used a square lattice with a size of 401 × 401 sites. While analyzing the critical region for each fixed concentration, we carried out 20 realizations for $D = 3$ and 15 realizations for $D = 2$. The results of the numerical experiment for quantities Δ_E^2 and Δ_j^2 are shown in Figs. 1–4.

In the 2D case, the processing of the results of numerical experiment for Δ_E^2 and Δ_j^2 in the critical region gives the following estimates for indices and coefficients:

$$t = 1.3 \pm 0.1, \quad t' = 1.3 \pm 0.2, \quad s = 0.50 \pm 0.05, \\ q = 1.3 \pm 0.1, \quad q' = 1.3 \pm 0.2, \tag{60}$$

$$A_0 = 1.8 \pm 0.2, \quad A_1 = 1.0 \pm 0.2, \quad a_0 = 1.00 \pm 0.05, \\ B_1 = 0.60 \pm 0.05, \quad B_2 = -0.8 \pm 0.3; \tag{61}$$

these estimates refine the results obtained in [9] to a certain extent. The value of index t corresponds to data available from the literature.

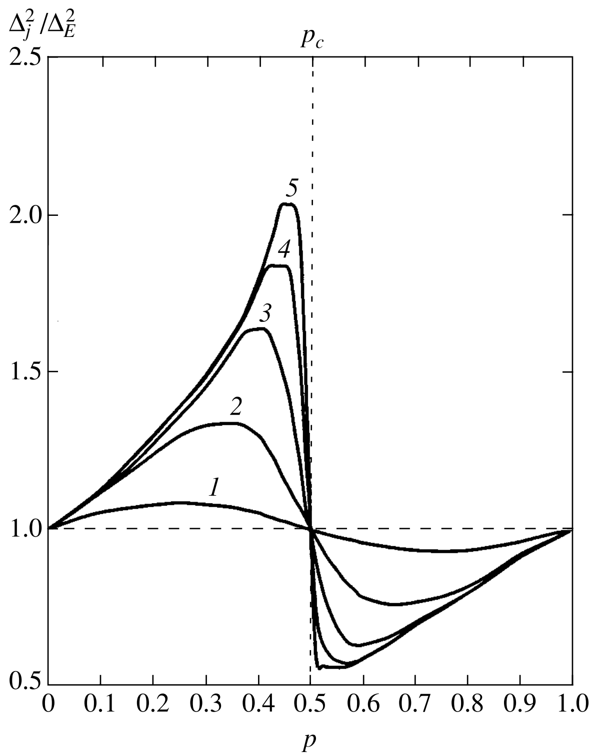


Fig. 2. Ratio Δ_j^2/Δ_E^2 as a function of concentration p for the same values of argument h as in Fig. 1. Two-dimensional case.

Since $q' \approx q$ and $t' \approx t$, we have $\Delta_E^2/\Delta_j^2 \approx \text{const}$ in the concentration ranges indicated in relations (56) and (58). This gives

$$A_0A_1 = 1.80 \pm 0.06, \quad \frac{B_2}{B_1^2} = -2.20 \pm 0.05. \tag{62}$$

Finally, from the ratio of the quantities $\Delta_E^2(\tau > 0)/\Delta_E^2(\tau < 0)$ and $\Delta_j^2(\tau > 0)/\Delta_j^2(\tau < 0)$ taken at points symmetric relative to $\tau = 0$ (outside the smearing region), we obtain

$$\frac{A_1}{B_1} = 1.65 \pm 0.05, \quad A_0\frac{B_2}{B_1} = -2.30 \pm 0.05. \tag{63}$$

It should be noted that the accuracy in determining the combinations of coefficients (62) and (63) is higher than the accuracy in determining the coefficients themselves (see Eq. (61)).

In the 3D case, the processing of the numerical experiment for Δ_E^2 and Δ_j^2 gives

$$\begin{gathered} t = 2.0 \pm 0.1, \quad t' = 2.0 \pm 0.1, \quad s = 0.7 \pm 0.1, \\ q = 0.8 \pm 0.1, \quad q' = 0.8 \pm 0.1, \end{gathered} \tag{64}$$

$$\begin{gathered} A_0 = 0.4 \pm 0.1, \quad A_1 = 3.0 \pm 0.2, \quad a_0 = 1.0 \pm 0.1, \\ B_1 = 0.7 \pm 0.2, \quad B_1 = -1.0 \pm 0.6. \end{gathered} \tag{65}$$

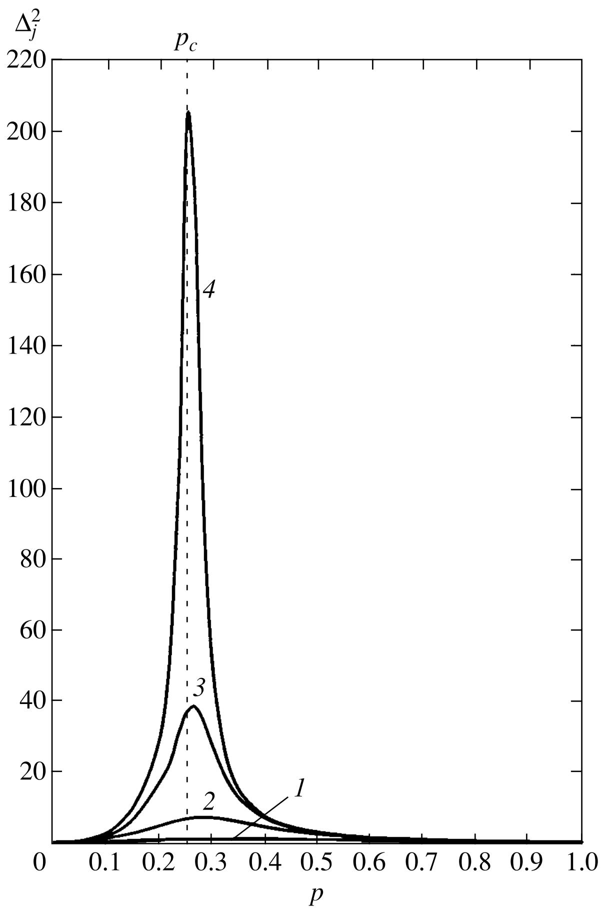


Fig. 3. Quantity Δ_j^2 as a function of concentration p for four values of argument h: $h = 10^{-1}$ (*1*), 10^{-2} (*2*), 10^{-3} (*3*), and 10^{-4} (*4*). Three-dimensional case.

From the ratio of the values of Δ_E^2 (as well as Δ_j^2) above and below the phase-transition point, we obtain, following an analogy with (63),

$$\frac{A_1}{B_1} = 4.2 \pm 0.1, \quad A_0\frac{B_2}{B_1} = -0.6 \pm 0.2. \tag{66}$$

Estimates (64) and (65) are in accordance with the results obtained in [10]. The values of critical indices (64) correspond to the available data from the literature.

Here, we calculate the values of Δ_E^2 and Δ_j^2 only for $h < 1$. However, the system under investigation is randomly inhomogeneous; the macroscopic properties of this system do not change as a result of simultaneous substitutions $p \longrightarrow 1 - p$ and $\sigma_1 \longleftrightarrow \sigma_2$. This means, in particular, that the effective conductivity satisfies the equality $\sigma_e(p; \sigma_1, \sigma_2) = \sigma_e(1 - p; \sigma_2, \sigma_1)$, which leads to the following relation:

$$f(p, h) = hf(1 - p, 1/h). \tag{67}$$

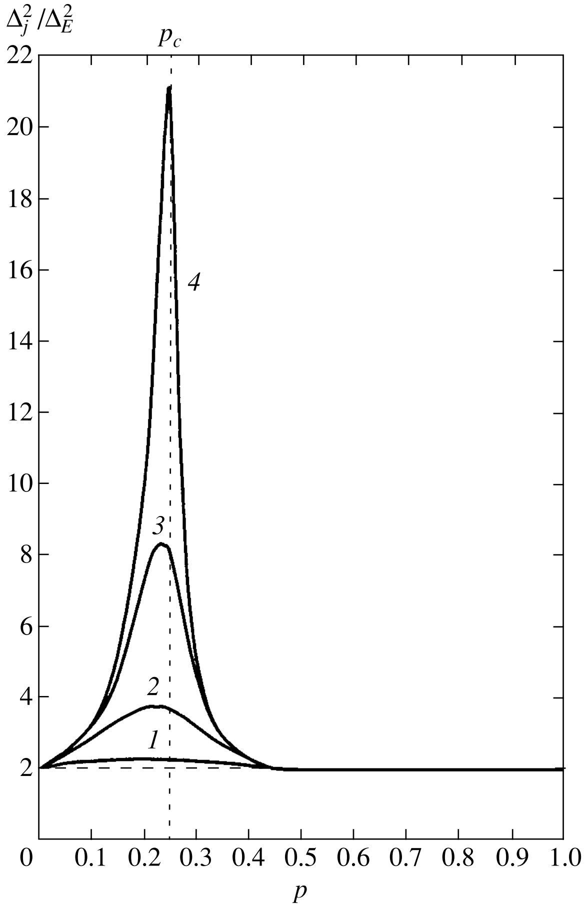


Fig. 4. Ratio Δ_j^2/Δ_E^2 as a function of concentration p for the same values of argument h as in Fig. 3. Three-dimensional case.

For a low concentration p of the first component, we obtain from Eq. (67) taking into account relation (50)

$$p \ll 1:\ f = h\left\{1 + pD\frac{1-h}{(D-1)h+1}\right\}. \tag{68}$$

For structural fluctuations Δ_E^2 and Δ_j^2, it follows from relations (13) and (14) combined with equality (67) that

$$\begin{aligned} \Delta_E^2(p, 1/h) &= \Delta_E^2(1-p, h), \\ \Delta_j^2(p, 1/h) &= \Delta_j^2(1-p, h). \end{aligned} \tag{69}$$

For small values of p, we obtain from Eq. (51) with the help of Eq. (69)

$$p \ll 1:$$
$$\Delta_E^2 = pD\left[\frac{1-h}{(D-1)h+1}\right]^2, \quad \Delta_j^2 = (D-1)\Delta_E^2. \tag{70}$$

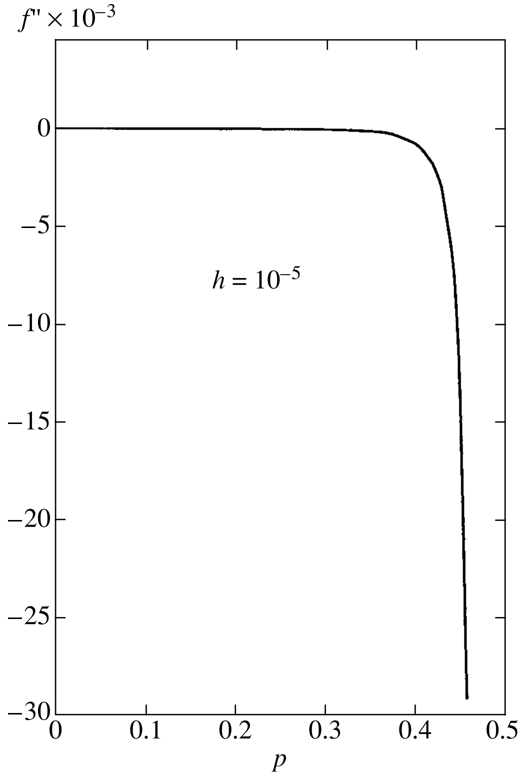


Fig. 5. Second derivative $f'' = \partial^2 f/\partial h^2$ for $h \longrightarrow 0$ as a function of concentration p ($p < p_c$). Two-dimensional case.

Relations (67) and (69) make it possible to determine quantities f as well as Δ_E^2 and Δ_j^2 for $h > 1$ if these quantities are known for $h < 1$ in the entire range of concentration.

For small values of h and for arbitrary p, we have

$$h \longrightarrow 0:\ f(p, h) = f(p, 0) + h\left(\frac{\partial f}{\partial h}\right)_0 + \frac{h^2}{2}\left(\frac{\partial^2 f}{\partial h^2}\right)_0 + \dots.$$

Since $f(p, 0) \equiv 0$ for $p < p_c$, in this case we obtain the following expression for function ψ_1 from relations (12):

$$h \longrightarrow 0:\ \psi_1 \approx -\frac{h^2}{2}\left(\frac{\partial^2 f}{\partial h^2}\right)_0, \quad p < p_c. \tag{71}$$

Calculating ψ_1 in accordance with definition (7), we obtain the second derivative $\partial^2 f/\partial h^2$ for $h = 0$ in the entire interval $0 < p < p_c$ (Figs. 5 and 6).

7. INTERPOLATION FORMULAS

In accordance with formulas (18), (28), (51), and (70), the values of Δ_E^2 and Δ_j^2 in all limiting cases for $D = 2$ are equal:

$$\Delta_j^2 = \Delta_E^2. \tag{72}$$

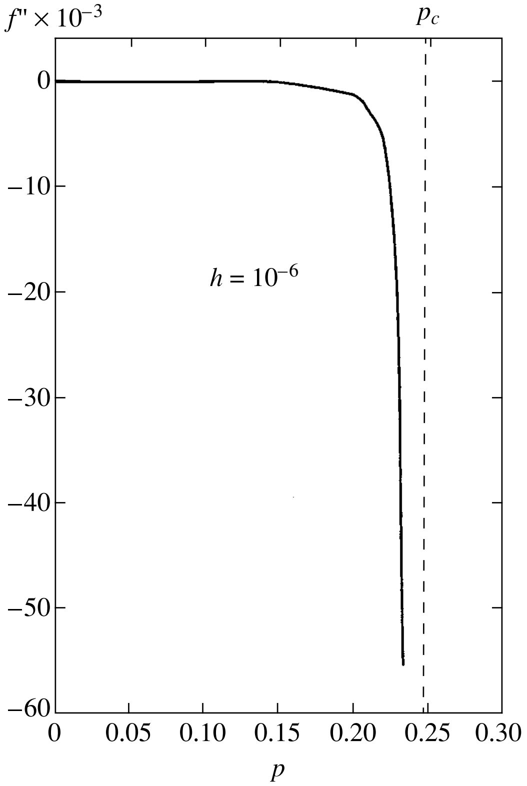


Fig. 6. Quantity f'' for $h \longrightarrow 0$ as a function of p ($p < p_c$). Three-dimensional case.

If we assume that this equality is valid for intermediate values of p and h also, the substitution of expressions (13) and (14) into Eq. (72) gives the following differential equation for function f:

$$(1-h)(f^2+h)\frac{\partial f}{\partial h} + f(f^2-1) = 0. \tag{73}$$

Dividing Eq. (73) by $(1-h)^2f^2$, we note that this equation can be written in the form

$$\frac{\partial}{\partial h}\left[\frac{1}{1-h}\left(f-\frac{h}{f}\right)\right] = 0,$$

whence

$$\frac{1}{1-h}\left(f-\frac{h}{f}\right) = C, \tag{74}$$

where C is the integration constant depending on parameter p.

Relation (74) is a quadratic equation in f; the solution to this equation is the quantity

$$f = \frac{1}{2}\{G(1-h) + [C^2(1-h)^2+4h]^{1/2}\}. \tag{75}$$

Constant C can be determined by comparing Eq. (75) with expression (26) for f in the case of a weakly inhomogeneous medium. The corresponding "boundary condition," which is valid for any D, can be written in the form

$$\left.\frac{\partial f}{\partial h}\right|_{h=1} = 1-p. \tag{76}$$

This gives $C = 2p - 1$ so that we finally obtain

$$f = \left(p-\frac{1}{2}\right)(1-h) + \left[\left(p-\frac{1}{2}\right)^2(1-h)^2 + h\right]^{1/2}, \quad D = 2. \tag{77}$$

Thus, equality (72) for function f leads to an expression coinciding with the formula in the theory of an effective medium (see, for example, [5], where the case $D = 2$ corresponds to $z = 4$).

It follows from the results of the numerical experiment (see Fig. 2) that equality (72) is satisfied approximately for $h \geq 0.1$ so that formula (77) for such values of h is a satisfactory interpolation expression for all functions f in the entire range of concentrations p. Taking into account relation (67), we can write the condition of applicability of Eq. (77) in the form $0.1 \leq h \leq 10$ for all values of p.

For an arbitrary D, in the limiting cases (28), (51), and (70), we have

$$\Delta_j^2 = (D-1)\Delta_E^2. \tag{78}$$

Substituting expressions (13) and (14) into Eq. (78), we obtain the following differential equation for f:

$$(1-h)[(D-1)f^2+h]\frac{\partial f}{\partial h} + [(D-1)f+1]f(f-1) = 0. \tag{79}$$

Dividing Eq. (79) by $(1-h)^2f^2$, we reduce this equation to the form

$$\frac{\partial}{\partial h}\left\{\frac{1}{1-h}\left[(D-1)f-(D-2)-\frac{h}{f}\right]\right\} = 0,$$

whence we obtain the following expression for f:

$$f = \frac{1}{2(D-1)}\{C(1-h)+D-2 + [(C(1-h)+D-2)^2+4(D-1)h]^{1/2}\}. \tag{80}$$

Constant C can be determined from condition (76):

$$C = pD-(D-1). \tag{81}$$

For a 3D system, relations (80) and (81) imply that

$$f = \frac{1}{4}\{3p - 1 + (2 - 3p)h + [(3p - 1 + (2 - 3p)h)^2 + 8h]^{1/2}\}, \quad D = 3. \tag{82}$$

This expression also coincides with the corresponding formula for f from [5], where the case $D = 3$ corresponds to $z = 6$.

The results of numerical experiment (see Fig. 4) show that equality (78) for $D = 3$ is valid to within a few percent for $h \geq 0.1$ so that, in accordance with [5], expression (82) for these values of h satisfactorily describes function f for all values of p in the case of the problem of associations. Taking into account relation (67), we find that the condition of applicability of Eq. (82) is the inequality $0.1 \leq h \leq 10$. If, however, the value of h does not satisfy these inequalities, it follows from Fig. 4 that equality (78) holds in a fairly wide range of concentrations (for $D = 3$) and interpolation formula (82) is applicable.

8. LAYERED MEDIUM

It is also interesting to consider structurally (geometrically) anisotropic media whose anisotropy is determined by the shape, distribution, and orientation of inclusions with an isotropic conductivity of the components. For such a system, we have, instead of relation (8),

$$\psi_{\alpha i} \equiv \frac{\langle(\mathbf{E}^{(\alpha)})^2\rangle^{(i)}}{|\langle\mathbf{E}^{(\alpha)}\rangle|^2} = \frac{\partial\sigma_{\alpha e}}{\partial\sigma_i}. \tag{83}$$

Here, $\sigma_{\alpha e}$ ($\alpha = x, y, z$) are the principal values of the effective conductivity tensor. Index α on $\mathbf{E}^{(\alpha)}$ indicates that field $\langle\mathbf{E}^{(\alpha)}\rangle$ is directed along axis α. Accordingly, instead of relation (10), we obtain

$$\Delta^2_{\alpha E} = \sum_i \frac{\partial\sigma_{\alpha e}}{\partial\sigma_i} - 1, \qquad \Delta^2_{\alpha j} = \frac{1}{\sigma^2_{\alpha e}}\sum_i \sigma_i^2\frac{\partial\sigma_{\alpha e}}{\partial\sigma_i} - 1. \tag{84}$$

It should be noted that structural field and current fluctuations for such media depend on the direction of the average field strength $\langle\mathbf{E}^{(\alpha)}\rangle$.

The limiting case of a structurally anisotropic medium is a layered medium, whose local isotropic conductivity depends on only one coordinate, e.g., z: $\sigma = \sigma(z)$. For such a medium, the quantities $\sigma_{\alpha e}$, as well as $\Delta^2_{\alpha E}$ and $\Delta^2_{\alpha j}$, can be determined in the general form. If field $\langle\mathbf{E}\rangle$ is directed along the x axis, $E(z) = E = \text{const}$ and we have

$$\sigma_{xe} = \langle\sigma\rangle, \quad \Delta^2_{xE} = 0, \qquad \Delta^2_{xj} = \langle(\sigma - \langle\sigma\rangle)^2\rangle/(\langle\sigma\rangle)^2. \tag{85}$$

Accordingly, for $\langle\mathbf{E}\rangle \parallel z$, the condition $j(z) = j = \text{const}$ is preserved so that

$$\sigma_{ze} = \left(\left\langle\frac{1}{\sigma}\right\rangle\right)^{-1}, \qquad \Delta^2_{zE} = \frac{\langle(\sigma^{-1} - \langle\sigma^{-1}\rangle)^2\rangle}{(\langle\sigma^{-1}\rangle)^2}, \quad \Delta^2_{zj} = 0. \tag{86}$$

It can easily be verified that expressions (85) and (86) for an N-component medium satisfy relations (84). For a two-component layered medium, Eqs. (85) and (86) lead to

$$\sigma_{xe} = p\sigma_1 + (1 - p)\sigma_2, \quad \Delta^2_{xE} = 0, \qquad \Delta^2_{xj} = p(1 - p)\frac{(\sigma_1 - \sigma_2)^2}{[p\sigma_1 + (1 - p)\sigma_2]^2}, \tag{87}$$

$$\sigma_{ze} = \frac{\sigma_1\sigma_2}{p\sigma_2 + (1 - p)\sigma_1}, \qquad \Delta^2_{zE} = p(1 - p)\frac{(\sigma_1 - \sigma_2)^2}{[p\sigma_2 + (1 - p)\sigma_1]^2}, \quad \Delta^2_{zj} = 0. \tag{88}$$

APPENDIX

Let us suppose that a body having a conductivity σ_2 is in a medium with conductivity σ_1 and is placed in a uniform electric field of strength $\mathbf{E}_0$. The asymptotic expression for the corresponding potential has the form (37)–(39). We will prove that a series of integrals containing the electric field strength $\mathbf{E} = \mathbf{E}(\mathbf{r})$ can be expressed in terms of the dipole polarizability tensor $\hat{\Lambda}$.

1. Let us consider the vector

$$\mathbf{A} = \int(\mathbf{j} - \sigma_1\mathbf{E})dV, \tag{A.1}$$

where integration is carried out over the entire space. Since the integrand in Eq. (A.1) differs from zero only within the body, we have, on the one hand,

$$\mathbf{A} = -(\sigma_1 - \sigma_2)\int_V \mathbf{E}dV, \tag{A.2}$$

where integration is carried out over the volume V of the body. On the other hand, in view of the equality div$\mathbf{j} = 0$, quantity A can be written in the form

$$A_\alpha = \int \frac{\partial}{\partial x_\beta}(x_\alpha j_\beta + \sigma_1 \varphi \delta_{\alpha\beta}) dV. \tag{A.3}$$

Transforming this integral into a surface integral (over a sphere of radius $R \longrightarrow \infty$) and evaluating this integral with the help of the asymptotic forms (37)–(39), we obtain

$$\mathbf{A} = 4\pi\sigma_1 \mathbf{p} = 4\pi\sigma_1 \hat{\Lambda}\mathbf{E}_0. \tag{A.4}$$

A comparison of Eqs. (A.2) and (A.4) gives

$$\int_V \mathbf{E} dV = -\frac{4\pi}{1-h}\hat{\Lambda}\mathbf{E}_0, \quad h = \frac{\sigma_2}{\sigma_1}. \tag{A.5}$$

Evaluating by the same method the integral

$$\int(\mathbf{E} - \mathbf{E}_0) dV = -\int \boldsymbol{\nabla}(\varphi - \varphi_0) dV,$$

$$\varphi_0(\mathbf{r}) = -\mathbf{E}_0 \cdot \mathbf{r},$$

we obtain

$$\int(\mathbf{E} - \mathbf{E}_0) dV = -\frac{4\pi}{D}\mathbf{p} = -\frac{4\pi}{D}\hat{\Lambda}\mathbf{E}_0, \tag{A.6}$$

where D is the dimension of the space; integration in Eq. (A.6) is carried out over the entire space.

2. Let us now consider the integral

$$J = \int \left\{ \frac{1}{\sigma_1}\mathbf{j} \cdot \tilde{\mathbf{E}} - \frac{1}{\tilde{\sigma}_1}\tilde{\mathbf{j}} \cdot \mathbf{E} \right\} dV, \tag{A.7}$$

where the symbol tilde marks the quantities pertaining to the same problem, but with changed conductivities ($\sigma_1 \longrightarrow \tilde{\sigma}_1$ and $\sigma_2 \longrightarrow \tilde{\sigma}_2$). Since the integrand in Eq. (A.7) differs from zero only within the body, we have

$$J = (h - \tilde{h}) \int_V (\mathbf{E} \cdot \tilde{\mathbf{E}}) dV, \tag{A.8}$$

where $\tilde{h} = \tilde{\sigma}_2/\tilde{\sigma}_1$. On the other hand, the quantity (A.7) can be written in the form

$$J = -\int \nabla \left\{ \frac{1}{\sigma_1}\mathbf{j}\tilde{\varphi} - \frac{1}{\tilde{\sigma}_1}\tilde{\mathbf{j}}\varphi \right\} dV. \tag{A.9}$$

Transforming this integral into a surface integral and evaluating it with the help of the asymptotic forms (37)–(39) and analogous expressions for $\tilde{\varphi}$, we obtain

$$J = 4\pi\mathbf{E}_0(\hat{\Lambda} - \hat{\tilde{\Lambda}})\mathbf{E}_0, \tag{A.10}$$

where $\hat{\Lambda} = \hat{\Lambda}(h)$ and $\hat{\tilde{\Lambda}} = \hat{\Lambda}(\tilde{h})$. A comparison of Eqs. (A.8) and (A.10) gives

$$(h - \tilde{h}) \int_V (\mathbf{E} \cdot \tilde{\mathbf{E}}) dV = 4\pi\mathbf{E}_0(\hat{\Lambda} - \hat{\tilde{\Lambda}})\mathbf{E}_0. \tag{A.11}$$

Using the limiting transition $\tilde{h} \longrightarrow h$, we obtain the relation

$$\int_V \mathbf{E}^2 dV = 4\pi\left(\mathbf{E}_0 \frac{\partial\hat{\Lambda}}{\partial h}\mathbf{E}_0\right). \tag{A.12}$$

3. Finally, let us consider the integral

$$I = \int\left(\frac{1}{\sigma_1}\mathbf{j} - \mathbf{E}_0\right)(\tilde{\mathbf{E}} - \mathbf{E}_0) dV, \tag{A.13}$$

where $\tilde{\mathbf{E}}$ is the same as in Eq. (A.7). The quantity (A.13) can be written in the form

$$I = -\int \nabla \left\{ \left(\frac{1}{\sigma_1}\mathbf{j} - \mathbf{E}_0\right)(\tilde{\varphi} - \varphi_0) \right\} dV. \tag{A.14}$$

Transforming this integral into a surface integral and evaluating it with the help of the asymptotic relations for $\varphi(\mathbf{r})$ and $\tilde{\varphi}(\mathbf{r})$, we find that $I = 0$. Consequently, it follows from Eq. (A.13) that

$$\begin{aligned} &\int_{V_e} (\mathbf{E} - \mathbf{E}_0) \cdot (\tilde{\mathbf{E}} - \mathbf{E}_0) dV \\ &= \int_V (h\mathbf{E} - \mathbf{E}_0) \cdot (\tilde{\mathbf{E}} - \mathbf{E}_0) dV. \end{aligned} \tag{A.15}$$

On the left-hand side of this equality, integration is carried out over the entire space except the volume v of the body. Evaluating the right-hand side of Eq. (A.15) with the help of relations (A.5) and (A.11), we obtain

$$\begin{aligned} &\int_{V_e} (\mathbf{E} - \mathbf{E}_0) \cdot (\tilde{\mathbf{E}} - \mathbf{E}_0) dV = \frac{4\pi}{h - \tilde{h}} \\ &\times \left\{ \tilde{h}\frac{1-h}{1-\tilde{h}}(\mathbf{E}_0\tilde{\Lambda}\hat{\mathbf{E}}_0) - h\frac{1-\tilde{h}}{1-h}(\mathbf{E}_0\hat{\Lambda}\mathbf{E}_0) \right\} - v\mathbf{E}_0^2. \end{aligned} \tag{A.16}$$

Using the limiting transition $\tilde{h} \longrightarrow h$, we obtain from Eq. (A.16)

$$\begin{aligned} &\int_{V_e} (\mathbf{E} - \mathbf{E}_0)^2 dV \\ &= -4\pi\left\{ \frac{1+h}{1-h}(\mathbf{E}_0\hat{\Lambda}\mathbf{E}_0) + h\left(\mathbf{E}_0\frac{\partial\hat{\Lambda}}{\partial h}\mathbf{E}_0\right) \right\} - v\mathbf{E}_0^2. \end{aligned} \tag{A.17}$$

Using relations (A.5) and (A.12), we also find that

$$\int_v (\mathbf{E}-\mathbf{E}_0)^2 dV = 4\pi\left\{\frac{2}{1-h}(\mathbf{E}_0\hat{\Lambda}\mathbf{E}_0) + \left(\mathbf{E}_0\frac{\partial\hat{\Lambda}}{\partial h}\mathbf{E}_0\right)\right\} + v\mathbf{E}_0^2. \tag{A.18}$$

Summing relations (A.17) and (A.18), we obtain

$$\int (\mathbf{E}-\mathbf{E}_0)^2 dV = 4\pi\left\{(\mathbf{E}_0\hat{\Lambda}\mathbf{E}_0) + (1-h)\left(\mathbf{E}_0\frac{\partial\hat{\Lambda}}{\partial h}\mathbf{E}_0\right)\right\}, \tag{A.19}$$

where integration is carried out over the entire space.

REFERENCES

1. A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1971)].
2. B. Ya. Balagurov, Fiz. Tverd. Tela (Leningrad) **20**, 3332 (1978) [Sov. Phys. Solid State **20**, 1922 (1978)].
3. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **93**, 1888 (1987) [Sov. Phys. JETP **66**, 1079 (1987)].
4. A. L. Efros and B. I. Shklovskii, Phys. Status Solidi B **76**, 475 (1976).
5. S. Kirkpatrick, Rev. Mod. Phys. **45**, 574 (1973).
6. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 3rd ed. (Nauka, Moscow, 1992; Pergamon, New York, 1984).
7. C. Herring, J. Appl. Phys. **31**, 1939 (1960).
8. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **89**, 1796 (1985) [Sov. Phys. JETP **62**, 1036 (1985)].
9. B. Ya. Balagurov and V. A. Kashin, Zh. Éksp. Teor. Fiz. **106**, 811 (1994) [JETP **79**, 445 (1994)].
10. B. Ya. Balagurov and V. A. Kashin, Zh. Éksp. Teor. Fiz. **110**, 1001 (1996) [JETP **83**, 553 (1996)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 1032–1052.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 1149–1171.
Original Russian Text Copyright © 2003 by Belyavsky, Kopaev, Sofronov, Shevtsov.

SOLIDS
Electronic Properties

Mirror Nesting of the Fermi Contour and Zero Line of the Superconducting Order Parameter

V. I. Belyavsky*, Yu. V. Kopaev, V. M. Sofronov, and S. V. Shevtsov**
Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 119991 Russia
**e-mail: vib@spu.ac.ru*
***e-mail: kopaev@lebedev.sci.ru*

Received May 22, 2003

Abstract—The superconducting order parameter that emerges owing to pairing of charge carriers with a large total momentum of the pair during screened Coulomb repulsion in a degenerate quasi-two-dimensional electronic system is determined as a function of the momentum of relative motion of the pair. In view of the kinematic constraint associated with Fermi filling, the ordered state exists in a limited domain of the momentum space, the shape and size of this domain being determined by the total momentum of the pair. The order parameter is not a constant-sign function of the momentum and reverses its sign on a certain line in a kinematically allowed domain. Superconducting instability arises for an arbitrarily small value of the repulsive interaction for certain momenta of the pair, for which the mirror nesting condition is satisfied; this results in the formation of a pair Fermi contour, i.e., the line of coincidence of segments of the Fermi contour with the isoline of the kinetic energy of relative motion of the pair. The temperature dependence of the superconducting order parameter is studied. Owing to the proximity effect in the momentum space, superconducting ordering is extended to the kinematically forbidden domain. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Quasi-two-dimensional (2*D*) high-temperature superconducting (HTSC) cuprates in the superconducting (SC) and the normal (N) state exhibit a number of properties which distinguish these compounds from conventional superconductors that can be successfully described in the Bardeen–Cooper–Schrieffer (BCS) theory [1]. In accordance with the BCS theory, the phase transition from the N state (Fermi liquid with a clearly defined Fermi surface) to the SC state is caused by the instability of the ground state of the N phase to the formation of singlet Cooper pairs with zero total momentum for an indefinitely weak attraction between the components of a pair. Since the electron (or hole) dispersion relation is an even function of the momentum, $\varepsilon(\mathbf{k}) = \varepsilon(-\mathbf{k})$, the SC instability evolves on the entire Fermi surface $\varepsilon(\mathbf{k}) = \mu$ (μ is the chemical potential of electrons); this results in the formation of an *s*-wave energy gap in the elementary excitation spectrum. The effective attraction between electrons in the BCS model is due to the electron–phonon interaction and takes place in a narrow (on the order of the characteristic Debye momentum) layer in the momentum space embracing the Fermi surface; this leads to certain considerations concerning the upper limit imposed on the superconducting transition temperature [2].

Undoped (or weakly doped) cuprates are antiferromagnetic (AF) insulators; for this reason, the properties of HTSC compounds are determined to a considerable extent by the competition or coexistence of SC, AF, and, probably, other ordered states [3]. Strong electron correlations in cuprates [4, 5] suggest that the state from which the system passes to the SC state formed in doped compounds upon a decrease in temperature differs from the state of a Fermi liquid [4]. However, angle-resolved photoemission spectroscopy (ARPES) not only unambiguously indicates the presence of the Fermi contour (FC) (which is an analog of the Fermi surface in the electronic 2*D* system) in doped cuprates, but also makes it possible to determine the size and shape of this contour as a function of the doping level x (the deviation of the carrier concentration from the half-filling). For example, in cuprates with the *p*-type doping, a "large" FC is observed, which is determined by the total hole concentration $(1 + x)$, has the shape of a "square with rounded corners" [5], and, hence, exhibits a clearly manifested nesting (as a rule, in the directions along the sides of the 2*D* Brillouin zone). The very existence of the FC is evidence of the fact that, in spite of the important role of electron correlations [4, 5], many properties of HTSC cuprates can be correctly described in terms of the momentum space in the framework of the band theory.

The SC ordering in HTSC cuprates exists in a limited doping interval; at the boundaries of this interval, the SC transition temperature T_c vanishes, attaining its maximum value for an certain optimal doping level x_{opt} within this interval. For $x < x_{\text{opt}}$, the density of states can be considerably suppressed in the temperature range $T_c(x) < T < T^*(x)$, which is interpreted as the emergence of a pseudogap; this can be associated [6] with a low phase stiffness of the SC order parameter. As the doping level increases, the temperature T_p of pair formation, corresponding to the emergence of the modulus of the order parameter, decreases, while the temperature T_ϑ, corresponding to the establishment of phase coherence, increases so that the phase coherence is established for $x \gtrsim x_{\mathrm{opt}}$ directly during the formation of pairs as in conventional superconductors. For $x < x_{\mathrm{opt}}$, incoherent states of pairs exist at $T < T^*$ and the SC phase transition can be treated as the Bose condensation of pairs that already exist.

In the case of singlet pairing, the angular dependence of the SC gap (as well as the pseudogap) observed in HTSC cuprates with the *p*-type doping can correspond to either *d*-wave or anisotropic *s*-wave symmetry. In the latter case, the order parameter $\Delta_{\mathbf{k}}$ as a function of the angular coordinate has a constant sign in contrast to the *d*-wave symmetry, when the sign changes four times during a complete turn of $\Delta_{\mathbf{k}}$ (this is confirmed in a number of experiments sensitive to the phase of the order parameter [7, 8] for some cuprate compounds). In the case of extended *s*-wave symmetry, the order parameter can change its sign depending on the magnitude of the momentum with a constant momentum direction [6].

The momentum dependence of the SC order parameter is associated with the electron–hole asymmetry, which is manifested in the tunnel characteristics of HTSC cuprates [9, 10]; in interpreting these characteristics, it is assumed [11] that the order parameter is a linear function of the hole kinetic energy,

$$\Delta_{\mathbf{k}} = a - b\varepsilon_{\mathbf{k}}, \qquad (1)$$

with constant coefficients *a* and *b*. The presence of coefficient *b* determining the slope of the SC gap leads to a certain shift of the chemical potential μ during SC condensation so that the line corresponding to the pair excitation energy minimum does not coincide with the FC [12].

According to some indications [6], the SC pairing in cuprates results from the repulsion between charge carriers. In contrast to the BCS theory, the gain in the ground-state energy in the SC condensation of pairs is mainly due to the renormalization of kinetic energy [13, 14]. In the case of repulsion, the replacement of the matrix element of interaction $U(|\mathbf{k} - \mathbf{k}'|)$, which is a function of the momentum $\boldsymbol{\kappa} = \mathbf{k} - \mathbf{k}'$ transferred during scattering by its certain constant value $U_0 > 0$ in the BCS equation, leads only to the trivial solution $\Delta = 0$. Consequently, the $U(\kappa)$ dependence in the kinematically allowed scattering domain is significant for the SC pairing with repulsive interaction and must lead to a nontrivial (depending on the momentum) solution for the SC order parameter $\Delta_{\mathbf{k}} \neq 0$.

The superconductivity mechanism [15–19] based on the assumption that the formation of pairs with a large total momentum $\mathbf{K}$ ($K \approx 2k_F$, where k_F is the Fermi momentum in the direction of the momentum of a pair) is the main channel of the SC pairing in the case of screened Coulomb repulsion makes it possible to interpret qualitatively the main experimental results obtained for HTSC cuprates.

In this study, the main conditions required for the SC pairing based on this mechanism are formulated and the SC order parameter depending on the momentum of relative motion of a pair and on temperature is determined and investigated. In Section 2, kinematic constraints imposed on pairing of like-charged particles with a nonzero total momentum of a pair are discussed together with the condition of mirror nesting for segments of the FC that play the role of a pair Fermi contour for the kinetic energy of the relative motion of a pair. Examples of realization of mirror nesting in cuprates are considered. The procedure of reducing the self-consistency equation determining the SC order parameter to an integral equation with a degenerate kernel is described in Section 3. Section 4 is devoted to an analysis of the general structure of the solution to the equation obtained. The parameters determining the general solution are determined and investigated in Section 5 at zero temperature and in Section 6 at a nonzero temperature. In Section 7, the relation between the SC transition temperature and the order parameter at zero temperature is analyzed. In Section 8, a class of discontinuous solutions of the self-consistency equation, which approximately represent an exact solution continuous in the momentum of relative motion of a pair, is considered. The proximity effect in the momentum space, i.e., the extension of SC ordering to a kinematically forbidden domain, is considered in Section 9. The results obtained, which correspond to the internal structure of the SC order parameter (its dependence on the momentum of relative motion of a pair) and its possible symmetry properties in HTSC cuprates, are discussed in the Conclusions (Section 10).

2. PAIRING WITH A LARGE MOMENTUM: MIRROR NESTING

As a result of the kinematic constraints associated with the filling of states inside the FC, momenta $\mathbf{k}_\pm$ of identically charged particles forming a pair with a preset total momentum $\mathbf{K} = \mathbf{k}_+ + \mathbf{k}_-$ ($\mathbf{K}$ pair) must both lie either inside or outside the FC. Thus, these momenta belong to a certain domain Ξ_K of the momentum space, which depends on the momentum of the pair [20, 21] and determines the statistical weight, i.e., the number of quantum states whose linear combination forms the pair. In the general case, this domain is symmetric to the inversion transformation $\mathbf{k} \longleftrightarrow -\mathbf{k}$ of the momentum $\mathbf{k} = (\mathbf{k}_+ - \mathbf{k}_-)/2$ of the relative motion of the pair and consists of two parts $\Xi_K^{(-)}$ and $\Xi_K^{(+)}$, which are inner and outer domains relative to the FC.

The excitation energy of the $\mathbf{K}$ pair, i.e., the kinetic energy of the particles forming this pair,

$$2\xi_{Kk} = \varepsilon_s(\mathbf{K}/2 + \mathbf{k}) + \varepsilon_{s'}(\mathbf{K}/2 - \mathbf{k}) - 2\mu, \qquad (2)$$

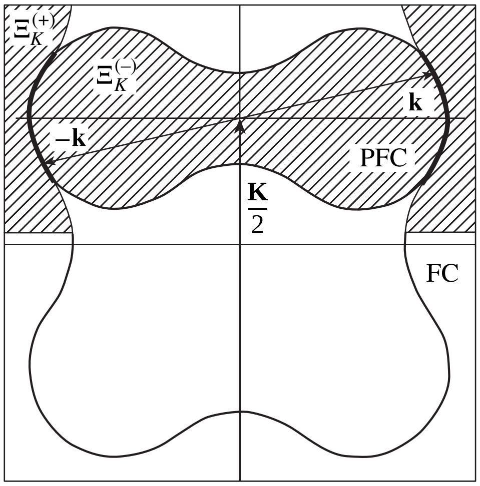


Fig. 1. Intraband mirror nesting. The formation of the PFC (bold lines) when the isoline of the kinetic energy of relative motion of a pair with momentum **K** along a certain direction coincides with a segment of the FC.

measured from the chemical potential 2μ of the pair, vanishes at the boundary separating the filled, $\Xi_K^{(-)}$, and the vacant, $\Xi_K^{(+)}$, parts of the domain of definition of the momentum of relative motion; here, $\varepsilon_s(\mathbf{k}_\pm)$ is the electron (or hole) dispersion relation in the *s*th energy band. The particles forming the **K** pair may belong either to the same ($s = s'$) energy band or to different ($s \neq s'$) energy bands. In the case of excitation of a pair of particles in the same band, the excitation spectrum (2) is symmetric relative to the inversion transformation $\mathbf{k} \longleftrightarrow -\mathbf{k}$ of the momentum of relative motion, while for $s \neq s'$ such symmetry is not observed in the general case.

For $K = 0$, the momenta $\pm\mathbf{k}$ of the relative motion coincide with the momenta of the particles constituting the pair; for this reason, the boundary between the filled part $\Xi_0^{(-)}$ of domain Ξ_0 coinciding with the 2*D* Brillouin zone and the vacant part $\Xi_0^{(+)}$ is the total FC. For $K \neq 0$, the boundary between $\Xi_K^{(-)}$ and $\Xi_K^{(+)}$ generally degenerates into a set of several points. However, such a boundary may be formed by individual segments of the FC for a special form of the dispersion relation and for a definite momentum of the pair. These segments coincide with segments of the isoline of the kinetic energy of relative motion of the **K** pair,

$$\varepsilon_{ss'}^{(r)}(\mathbf{K}, \mathbf{k}) = \varepsilon_s(\mathbf{K}/2 + \mathbf{k}) + \varepsilon_{s'}(\mathbf{K}/2 - \mathbf{k}) - \varepsilon_{ss'}^{(c)}(\mathbf{K}), \quad (3)$$

which corresponds to the difference between the chemical potential 2μ of the pair and the energy of its center of mass:

$$\varepsilon_{ss'}^{(c)}(\mathbf{K}) = \varepsilon_s(\mathbf{K}/2) + \varepsilon_{s'}(\mathbf{K}/2). \quad (4)$$

In this case, the boundary between $\Xi_K^{(-)}$ and $\Xi_K^{(+)}$ plays the role of a sort of FC with respect to the relative motion of the pair and can be referred to as the pair Fermi contour (PFC). The condition determining the PFC has the form

$$\varepsilon_{ss'}^{(r)}(\mathbf{K}, \mathbf{k}) = 2\mu - \varepsilon_{ss'}^{(c)}(\mathbf{K}). \quad (5)$$

Since $\mathbf{k} \longleftrightarrow -\mathbf{k}$, both momenta of the relative motion (**k** and –**k**) belong to the PFC. In this case, the boundary between $\Xi_K^{(-)}$ and $\Xi_K^{(+)}$ possesses mirror symmetry relative to vector **K**; for this reason, relation (5) can be called the condition of mirror nesting. In order to satisfy condition (5), vector **K** must coincide with one of symmetry directions in the 2*D* Brillouin zone. In the case of tetragonal symmetry, these are either the directions along the sides of the Brillouin zone or along its diagonals. It should be noted that the inversion transformation $\mathbf{k} \longleftrightarrow -\mathbf{k}$ under the mirror nesting condition leads to the coincidence of filled states with filled states and vacant states with vacant states.

We can consider several examples of the emergence of a PFC for a certain total momentum of a pair. For instance, suppose that an FC is simply connected and has the shape of a square with rounded corners in accordance with the results of ARPES experiments for cuprates with the *p*-type doping [5]. Then even a small change in the sign of curvature of nearly rectilinear segments of the FC lying in an extended neighborhood of the saddle point of the dispersion relation for holes [5] may ensure the fulfillment of the mirror nesting condition owing to strong anisotropy of the effective masses (this is shown schematically in Fig. 1).

In the case when the Fermi level intersects two energy bands, a doubly connected FC is formed [22] and two segments of the FC belonging to different bands may also ensure the fulfillment of the mirror nesting condition [18] (Fig. 2). In the cases considered above, the PFC is only a part of the total FC on which the excitation energy of a pair with a definite momentum vanishes.

However, we can also consider the case when a PFC in fact coincides with the total FC. For example, in underdoped cuprates, the FC may have the shape of several hole pockets [23, 24] centered at points belonging to symmetric directions as shown in Fig. 3 (as the doping level increases, such pockets are continuously transformed into a large FC [24]). In this case, the mirror nesting condition is perfectly satisfied on the entire line bounding a pocket if one-half of the momentum of the pair corresponds to the center of the pocket. It should be noted that an exactly similar situation takes place in degenerate multivalley semiconductors [25, 26].

The formation of a PFC can also be associated with instability of the ground state of the normal phase, leading to a change in the FC shape as a result of redistribution of charge carriers in the momentum space upon the formation of a strip structure [16, 17]; this is shown schematically in Fig. 4. Due to the hyperbolic metric of the momentum space, the corresponding increase in the kinetic energy, which is compensated either by a partial restoration of the AF order in hole-depleted regions of the real space, or by the emergence of an orbital AF order according to [3], can be minimized.

3. SELF-CONSISTENCY EQUATIONS

In this section we consider the necessary conditions for SC pairing for the repulsive interaction between the particles constituting a pair with a large total momentum. We propose a procedure which reduces the equation for the SC order parameter to an approximate linear singular equation with a degenerate kernel.

The screened Coulomb repulsion between the components of the pair is defined in the entire kinematically allowed domain Ξ_K. The Fermi filling and the singularity in the permittivity associated with it leads to the emergence of the familiar Friedel oscillations of the effective potential; the spatial scale of these oscillations is on the order of the reciprocal Fermi momentum k_F^{-1} [27]. It should be noted in this connection that the interaction leading to pairing in the BCS model is defined only in a narrow strip along the FC with the characteristic width $2\Delta\varepsilon \ll E_F$, where E_F is the Fermi energy and $\Delta\varepsilon$ in the case of the electron–phonon interaction has the meaning of the Debye energy.

The self-consistency equation defining the SC order parameter determined by pairing of particles with a large total momentum $\mathbf{K}$ for a finite temperature T can be written in the form

$$\Delta_{\mathbf{Kk}} = -\frac{1}{2S}\sum_{k' \in \Xi_K} \frac{U(\mathbf{k}-\mathbf{k}')\Delta_{\mathbf{Kk}'}}{\sqrt{\xi^2_{\mathbf{Kk}'} + \Delta^2_{\mathbf{Kk}'}}} h_{\mathbf{Kk}'}(T), \tag{6}$$

where $\mathbf{k}$ and $\mathbf{k}'$ are the momenta of the relative motion of pairs with the same total momentum $\mathbf{K}$, $U(\kappa)$ is the matrix element of the effective interaction potential for the particles constituting a pair, and S is the normalizing area. The momentum $\boldsymbol{\kappa} = \mathbf{k} - \mathbf{k}'$ appearing in the matrix element, which is transferred during scattering and is equal to the difference $\mathbf{k}_+ - \mathbf{k}'_+$ between the final and initial momenta of a particle, is also equal to the change in the momentum of the relative motion as a result of scattering. The temperature factor in the self-consistency equation has the form

$$h_{\mathbf{Kk}'}(T) = \tanh\left(\frac{\sqrt{\xi^2_{\mathbf{Kk}'} + \Delta^2_{\mathbf{Kk}'}}}{2T}\right). \tag{7}$$

Summation in Eq. (6) is carried out over the entire kinematically allowed domain Ξ_K, in which the SC order parameter is defined.

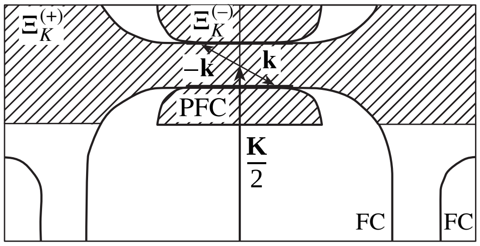


Fig. 2. Interband mirror nesting. The formation of the PFC in the two-band model of the energy spectrum. The Fermi contour consists of two simply connected parts. Only the upper half of the 2*D* Brillouin zone is shown.

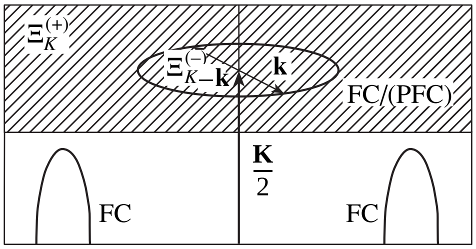


Fig. 3. Fermi contour in the form of hole pockets in strongly underdoped cuprates. Only the upper half of the 2*D* Brillouin zone is shown.

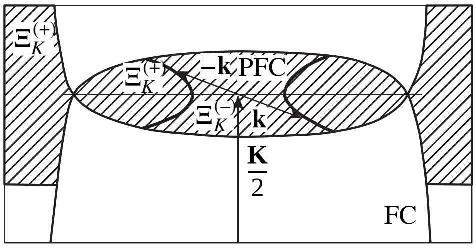


Fig. 4. Formation of the PFC in the case of a strip structure. The PFC (bold lines within Ξ_K) is formed owing to a redistribution of particles in the momentum space. Only the upper half of the 2*D* Brillouin zone is shown.

We will solve Eq. (6) by passing from summation to integration. Omitting index $\mathbf{K}$, which corresponds to the total momentum of a pair, we will treat the quantities $\Delta_{\mathbf{Kk}} = \Delta(\mathbf{k})$ and $\xi_{\mathbf{Kk}} = \xi(\mathbf{k})$ as continuous functions of the momentum of relative motion. Introducing the notation

$$\eta(\mathbf{k}) = \sqrt{\xi^2(\mathbf{k}) + \Delta^2(\mathbf{k})}, \quad f(\mathbf{k}) = h(\mathbf{k})/4\pi\eta(\mathbf{k}), \tag{8}$$

we can write the self-consistency equation (6) in the form

$$\Delta(\mathbf{k}) = -\frac{1}{2\pi}\int_{\Xi} U(|\mathbf{k}-\mathbf{k}'|) f(\mathbf{k}')\Delta(\mathbf{k}')d^2k'. \tag{9}$$

The repulsion between the particles of a pair corre-

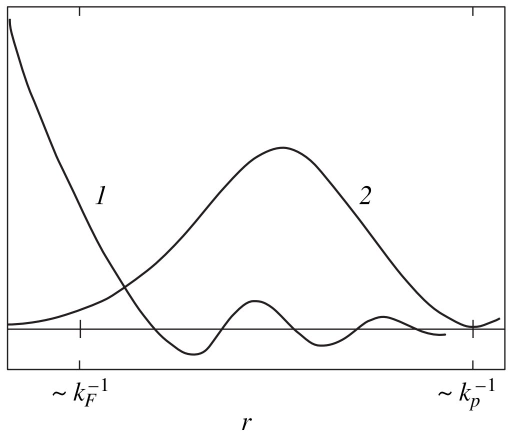


Fig. 5. Curves describing schematically the functions defining the integral on the right-hand side of Eq. (11): function $rU(r)$ (*1*), function $L(r)$ averaged over the angular variable (*2*) and defined in Eq. (12).

sponds to the positive matrix element $U(\mathbf{k}-\mathbf{k}')$ connected to the effective interaction potential $U(r)$ in the r space via the Fourier transformation:

$$U(|\mathbf{k}-\mathbf{k}'|) = \int U(r)\exp[i((\mathbf{k}-\mathbf{k}')\cdot\mathbf{r})]d^2r. \qquad (10)$$

We can establish a simple criterion for the absence of nontrivial solutions to Eq. (9). For this purpose, we multiply this equation by the function $f(\mathbf{k})\Delta(\mathbf{k})$ and integrate the relation obtained with respect to $\mathbf{k}$ within domain Ξ. Expressing $U(\mathbf{k}-\mathbf{k}')$ in accordance with relation (10) and changing the order of integration with respect to $\mathbf{k}$, $\mathbf{k}'$, and $\mathbf{r}$, we obtain

$$\int_\Xi \Delta^2(\mathbf{k}) f(\mathbf{k}) d^2k = -\frac{1}{2\pi}\int U(r)L(\mathbf{r})d^2r, \qquad (11)$$

where the following notation has been introduced for a nonnegative function of argument $\mathbf{r}$:

$$L(\mathbf{r}) = \left|\int_\Xi \Delta(\mathbf{k}) f(\mathbf{k})\exp(i\mathbf{k}\cdot\mathbf{r})d^2k\right|^2. \qquad (12)$$

Integration with respect to $\mathbf{r}$ in Eq. (11) is carried out over the entire $2D$ space. By virtue of definition (8), the inequality $f(\mathbf{k}) \geq 0$ holds and, hence, the left-hand side of Eq. (11) is nonnegative. If potential $U(r)$ is positive for all values of r, the right-hand side of Eq. (11) is, on the contrary, negative; consequently, equality (11) can hold in this case only if $\Delta(\mathbf{k}) \equiv 0$ everywhere in domain Ξ. Thus, if the interaction is purely repulsive in the real space and if $U(r) > 0$ for all values of r, the self-consistency equation has only a trivial solution. Consequently, the SC ordering is possible only in systems for which the effective interaction potential of particles is an alternating function of the distance between these particles. It was mentioned above, in particular, that the screened Coulomb potential in a degenerate Fermi system, which exhibits the Friedel oscillations, is a potential of this kind [27]. Figure 5 shows schematically the dependence $rU(r)$ such that condition (11) can be satisfied for $\Delta(\mathbf{k}) \neq 0$.

Indeed, the alternating function $rU(r)$ in the integral on the right-hand side of Eq. (11) evaluated in polar coordinates, whose maximum (positive) value is attained for $r = 0$ (see Fig. 5), is multiplied by the nonnegative function $L(\mathbf{r})$ whose general dependence on the argument can be easily established. The modulus of the product $\Delta(\mathbf{k})f(\mathbf{k})$ attains its maximal value on the PFC (where $\xi(\mathbf{k}) = 0$), decreasing rapidly with increasing distance from the PFC. For this reason, the main contribution to the integral defining $L(\mathbf{r})$ comes from a small part of domain Ξ in the form of a narrow strip adjoining the PFC. It follows directly from Eq. (9) that, for $U(\mathbf{k}-\mathbf{k}') > 0$, the nontrivial solution $\Delta(\mathbf{k})$, if it exists, must be an alternating function of its argument in domain Ξ, which changes its sign (see below) on a certain line intersecting the PFC. Consequently, the integral in Eq. (12) is a function of r with peaks decreasing upon an increase in the argument and separated by the characteristic distance π/k_P, where k_P is a quantity on the order of the length of the PFC. The first and highest peak corresponds to the fact that the product of the alternating factor $\Delta(\mathbf{k})f(\mathbf{k})$ and the exponent in integral (12) is generally constant in sign. The value of $L(0)$, which is determined by the integral of the product $\Delta(\mathbf{k})f(\mathbf{k})$ over domain Ξ, is generally a small quantity in view of partial mutual compensation of the contributions to this value from the parts of domain Ξ in which function $\Delta(\mathbf{k})$ has opposite signs. Thus, we can assume that function $L(\mathbf{r})$ is close to zero for small values of r. For complete compensation of contributions from $\Delta(\mathbf{k})$ with opposite signs, $L(r) \propto r^4$, which leads to a considerable suppression of the main (positive) extremum of function $rU(r)$, which has a finite limit for $r \longrightarrow 0$. In the opposite limiting case of large r, this function is also small due to the presence of the rapidly oscillating factor $\exp(i\mathbf{k}\cdot\mathbf{r})$ in integral (12). Since the value of k_P is obviously smaller than k_F, the value of the first maximum of function (12) satisfies the condition $r_1 \sim \pi/k_P > \pi/k_F$ so that the first maximum naturally falls in the region of Friedel oscillations (see Fig. 5). It can be seen that in a considerable region of the r space, the integrand on the right-hand side of Eq. (11) is essentially negative, while the contributions from the regions in which it is positive are comparatively small; consequently, the integral itself is negative. In this case, equality (11) can be ensured by the fact that function $f(\mathbf{k})$, appearing to the first power on the left-hand side of Eq. (11) and to the second power on the right-hand side of this equation, has a singularity on the PFC at $T = 0$ and may become indefinitely large upon a decrease in $\Delta(\mathbf{k})$. The amplitude of oscillations of function $L(r)$ depends on $\Delta(\mathbf{k})$ since it is function $\Delta(\mathbf{k})$ that determines the effective bandwidth in the vicinity of the PFC, which makes the main contribution to integral (12). Thus, condition (11) can be satisfied for quite small values of $\Delta(\mathbf{k}) \neq 0$.

Equation (9) is a nonlinear integral Hammerstein equation with the symmetric kernel $U(|\mathbf{k}-\mathbf{k}'|)$. It is well known [28] that all characteristic numbers of such a kernel are real-valued and form a discrete spectrum λ_n having a condensation point $|\lambda_n| \longrightarrow \infty$ for $n \longrightarrow \infty$ in the case of a nondegenerate kernel. If kernel $U(|\mathbf{k}-\mathbf{k}'|)$ is the Fourier transform (10) of the effective potential $U(r)$ which is positive everywhere, it can easily be seen that it is positive definite and, accordingly, all its characteristic numbers are positive. In this case, as was shown earlier, Eq. (9) has only a trivial solution. Consequently, the necessary condition for the existence of a nontrivial solution of the self-consistency equation is the presence of at least one negative characteristic number in the spectrum of kernel $U(|\mathbf{k}-\mathbf{k}'|)$.

The functions determining the nonlinearity of the integral operator in Eq. (9) has a special form $\Delta f(\Delta)$, where $f(\Delta)$ is the nonlinearity factor defined in accordance with Eq. (8). In this sense, Eq. (9) can be classified as a quasi-linear integral equation [28], whose important feature is that, in the case of a degenerate kernel, the solution (if it exists) of the quasilinear equation, as in the case of a linear integral equation, has a form repeating the structure of this kernel [28]. Kernel $U(|\mathbf{k}-\mathbf{k}'|)$ can be approximately reduced to a degenerate kernel if we take into account the fact that both variables $\mathbf{k}$ and $\mathbf{k}'$ are defined in a relatively small domain Ξ of the momentum space. Consequently, in the Taylor expansion of function (10) in argument $\kappa = |\mathbf{k}-\mathbf{k}'|$, we can confine our analysis to only a few first terms. It follows from definition (10) that all coefficients of odd powers of vector $\mathbf{k}-\mathbf{k}'$ in this expansion vanish identically so that the Fourier transform of the interaction potential can be written as

$$U(\kappa) = 2\pi\left[u_0 - \frac{1}{2}u_2\kappa^2 + \frac{3}{8}u_4\kappa^4 - \ldots\right], \tag{13}$$

where

$$u_n = \frac{1}{n!}\int_0^\infty U(r)r^{n+1}dr. \tag{14}$$

If we truncate the series in expansion (13), retaining only a few first terms, we obtain a degenerate kernel which successfully approximates the true nondegenerate kernel in the case of a small domain Ξ.

It should be noted that the Fourier transform $U(\kappa)$ corresponding to the repulsive interaction for small κ is a decreasing function of its argument; consequently, $u_0 > 0$ and $u_2 > 0$. While solving Eq. (9), it is convenient to pass to dimensionless variables, using constants U_0 and r_0 defined by the relations $U_0 r_0^2 = u_0$ and $U_0 r_0^4 = u_2$ as the characteristic scales of energy and length. It should be noted that, in the case of the screened Coulomb potential, parameters r_0 and $U_0 = e^2/r_0$ have the meaning of the screening radius and the characteristic Coulomb energy, respectively. Thus, energy U_0 determines the scale of quantities $\Delta(\mathbf{k})$, $\xi(\mathbf{k})$, and T; the momentum is measured in units of r_0^{-1}; and the dimension of the matrix element is $U_0 r_0^2$. In order to preserve the previous notation for the dimensionless quantities introduced in this way, we must assume in self-consistency equation (9), which does not change its form upon a transition to the dimensionless quantities, that domain Ξ of the momentum space over which the integration in Eq. (9) is carried out is also reduced to dimensionless variables (momentum components k_1 and k_2). If this domain is small enough, we can retain the first two terms in expansion (13) and, hence, arrive at the degenerate kernel

$$U_d(|\mathbf{k}-\mathbf{k}'|) = 2\pi[1-(\mathbf{k}-\mathbf{k}')^2/2]. \tag{15}$$

We can verify that degenerate kernel (15) has four characteristic numbers, from which three are positive, while the fourth number is negative. Two of the positive characteristic numbers are identical, $\lambda_{1,2} \sim u_0^{-1}$, and the third positive number is $\lambda_3 \sim u_2^{-1}$. The negative characteristic number is $\lambda_4 \sim u_0^{-1}/u_2^2$. Obviously, in the case of repulsive interaction, approximation (15) can be regarded as sufficient if $(k + k')^2 < 2$; consequently, in the subsequent analysis, we will assume that domain Ξ is such that any momentum of the relative motion defined in this domain satisfies the condition

$$k^2 < 1/2. \tag{16}$$

Thus, if we assume that the screening radius r_0 amounts to several atomic spacings, condition (16) implies that the characteristic size of domain Ξ must be a fraction of the size of the Brillouin zone. In addition, disregard of the third term in expansion (13) of the next order of smallness is justified only under the condition [29, 30] that the root-mean-square value over domain Ξ is much smaller than the reciprocal of the magnitude of any of the characteristic numbers of degenerate kernel (15):

$$2|u_4|\sqrt{\int_\Xi d^2k\int_\Xi d^2k'(\mathbf{k}-\mathbf{k}')^8} \ll |\lambda|^{-1}. \tag{17}$$

For positive characteristic numbers, this inequality is satisfied under the natural condition that the third term of expansion (13) is small as compared to the second term. For the negative characteristic number of kernel (15), inequality (17) can be reduced to the condition $u_0|u_4| < cu_2^2$, where $c \ll 1$. This relation is obviously violated in the case of the potential $U(r)$ that is positive everywhere, since the quantities u_n defined in (14) must be connected in this case via the opposite inequality following from the Cauchy–Buniakowski inequality for integrals (14). Thus, approximating the actual kernel $U(\mathbf{k} - \mathbf{k}')$ by the degenerate kernel (15), we presume that conditions (16) and (17) are satisfied.

4. GENERAL SOLUTION OF THE SELF-CONSISTENCY EQUATION WITH A DEGENERATE KERNEL

Substituting kernel (15) into Eq. (9) and grouping the terms independent of momentum, linear in the momentum, and quadratic in the momentum on the right-hand side, we can verify that the dependence of the order parameter on the momentum of relative motion of a pair must have the form

$$\Delta(\mathbf{k}) = a + (\boldsymbol{\chi} \cdot \mathbf{k}) - bk^2, \tag{18}$$

where a and b are certain coefficients and $\boldsymbol{\chi}$ is a constant vector. Thus, the solution to Eq. (9) indeed repeats the form of the momentum dependence of degenerate kernel (15) and can be reduced to determining parameters a, b, and $\boldsymbol{\chi}$. In order to find these parameters, we substitute the expression for the matrix element (15) and the explicit form of solution (18) into Eq. (9), after which we equate the coefficients of identical powers of $\mathbf{k}$. As a result, we obtain a system of integral equations in a, b, and $\boldsymbol{\chi}$; one of these equations, obtained from the comparison of the coefficients of $\mathbf{k}$ to the first power, has the form

$$\boldsymbol{\chi} = -\int_{\Xi} f(\mathbf{k})\Delta(\mathbf{k})\mathbf{k}d^2k. \tag{19}$$

This equation implicitly defines vector $\boldsymbol{\chi}$ with components χ_i, $i = 1, 2$, since the right-hand side of this equation contains the dependence on $\boldsymbol{\chi}$ in Δ. It can easily be verified that Eq. (19) has only one solution. In order to prove this, we introduce an auxiliary function $H(\Delta) \equiv \Delta f(\eta(\Delta))$, which is a monotonically increasing function of its argument since $H'(\Delta) > 0$, where the prime indicates differentiation with respect to Δ. Writing Eq. (19) in components, we can prove that its right-hand side is a decreasing function of χ_i. Indeed, differentiation of the integral in Eq. (19) with respect to parameter χ_i, taking into account the fact that $dH(\Delta)/d\chi_i = H'(\Delta)k_i$, shows that this derivative is negative everywhere,

$$-\int_{\Xi} H'(\Delta)k_i^2 d^2k < 0, \tag{20}$$

so that the right-hand side of Eq. (19) is a decreasing function of χ_i, while the left-hand side increases with χ_i. Thus, equality (19) is possible for only one value of $\boldsymbol{\chi}$.

In the case when excitation corresponds to a pair of particles in the same energy band, the obvious solution to Eq. (19) is $\boldsymbol{\chi} = 0$. Indeed, in view of the central symmetry of the integration domain Ξ, the right-hand side of Eq. (19) for $\boldsymbol{\chi} = 0$ also vanishes as an integral of an odd function over a symmetric domain since the energy gap for $\boldsymbol{\chi} = 0$ is an even function of the momentum and the excitation energy satisfies the condition $\xi(-\mathbf{k}) = \xi(\mathbf{k})$. Equation (19) has no other solutions. Thus, in the case of intraband mirror nesting and for the nondegenerate kernel (15), Eq. (9) has a simple solution of the form

$$\Delta(k) = a - bk^2, \tag{21}$$

defining the energy gap in Ξ as a parabolic function of the momentum of the relative motion. Coefficients a and b defining the position of the vertex and the curvature of the parabola are determined by the size and shape of domain Ξ, the form of the one-particle dispersion relation, and the temperature. It should be noted that a and b in relation (21) are dimensionless quantities in units of U_0 and $U_0r_0^2$.

The procedure of determining parameters a, b, and $\boldsymbol{\chi}$ gives, in addition to Eq. (19), two more equations which, on account of the fact that $\boldsymbol{\chi} = 0$, assume the form

$$\begin{gathered}\left(J_0 - \frac{1}{2}J_1 + 1\right)a - \left(J_1 - \frac{1}{2}J_2\right)b = 0,\\ \frac{1}{2}J_0 a - \left(\frac{1}{2}J_1 - 1\right)b = 0,\end{gathered} \tag{22}$$

where the following notation is used for similar integrals in domain Ξ:

$$J_n \equiv \int_{\Xi} f(\mathbf{k})k^{2n}d^2k. \tag{23}$$

All three integrals (23) appearing in system of equations (22) for $n = 0, 1, 2$ are positive and depend on a and b as parameters. Thus, Eqs. (22) form a system of two quasi-linear equations in two unknowns a and b.

Multiplying the condition $k^2 < 1/2$ under which the approximate kernel (15) can be used first by $f(\eta(\mathbf{k}))$ and then by $k^2f(\eta(\mathbf{k}))$, and integrating the obtained inequalities over domain Ξ, we obtain two inequalities $J_1 < J_0/2$ and $J_2 < J_1/2$. From these inequalities, we can conclude that both coefficients of the unknowns in the first equation in system (22) are positive. Consequently, solutions a and b of system (22) have the same (e.g., positive) sign. Consequently, the superconducting order parameter defined by expression (21) changes its sign for a certain $k = k_0$, where $k_0^2 = a/b$, and can be written in the form

$$\Delta(k) = b(k_0^2 - k^2). \tag{24}$$

Parameter b characterizes the energy scale of the superconducting gap, and k_0 determines the radius of the circle in the momentum space, on which the energy gap becomes zero. In the case of the repulsive interaction discussed here, the vanishing of the superconducting order parameter takes place in domain Ξ, since Eq. (9) for $U(|\mathbf{k} - \mathbf{k}'|) > 0$ obviously has no solutions of constant sign. On the circle $k = k_0$, the order parameter changes its sign (the phase of the order parameter changes by π); consequently, this line can be referred to as the line of the π phase shift of the order parameter.

The second equation of system (22) implies that the conditions $a > 0$ and $b > 0$ indicate the fulfillment of the inequality $J_1 > 2$, which imposes certain constraints on possible values of the parameters of the problem. This relation is a consequence of a stronger inequality following from the condition of nontrivial compatibility of system of equations (22). Evaluating the determinant of the system, we can write this condition in the form

$$(J_1/2 - 1)^2 = J_0(J_2/4 - 1), \tag{25}$$

which shows that the inequality $J_2 > 4$ must hold; considering that $J_1 > 2J_2$, we can obtain $J_1 > 8$ from this inequality.

Taking into account the explicit form of integral J_2, as well as the inequality $f(\mathbf{k}) \le (8\pi T)^{-1}$, condition $J_2 > 4$ can be written in the form

$$T < \frac{1}{32\pi}\int_{\Xi} k^4 d^2k. \tag{26}$$

Condition (26) sets the upper limit on the temperature of the emergence of the superconducting order and also indicates that the order associated with the formation of the condensate of pairs with the total momentum **K** at a finite temperature may arise for a fairly large size of the corresponding domain Ξ.

Instead of the pair of the unknown parameters a and b characterizing the superconducting gap, it is convenient to consider another pair of parameters, b and k_0^2, which can be determined using any two equations from (22) and (25). It is convenient to take the second equation from system (22) as one such equation, writing this equation (provided that $k_0^2 = a/b$) in the form

$$J_1 - k_0^2 J_0 = 2. \tag{27}$$

Expressing the left-had side of Eq. (25) with the use of the second equation from system (22), we rewrite relation (25) as

$$J_2 - k_0^4 J_0 = 4. \tag{28}$$

System of equations (27) and (28) is completely equivalent to initial system (22) and can be used for determining parameters b and k_0^2 of the superconducting gap. It should be noted in this connection that integrals J_0, J_1, and J_2, which are functions of Δ, depend on b and k_0^2 themselves.

In the case of interband mirror nesting ($s \neq s'$), the pair excitation energy (2) generally is not an even function of the momentum of relative motion, $\xi(-\mathbf{k}) \neq \xi(\mathbf{k})$. In this case, Eq. (19) defining vector $\boldsymbol{\chi}$ also has a unique solution which might be nonzero. Indeed, writing Eq. (19) in components and analyzing the asymptotic behavior of its right- and left-hand sides for $\chi_i \longrightarrow \pm\infty$, we can verify that the left-hand side increases monotonically from $-\infty$ to $+\infty$ with increasing χ_i, while the right-hand side decreases monotonically from a finite positive value for $\chi_i \longrightarrow -\infty$ to a finite negative value for $\chi_i \longrightarrow +\infty$. Two functions exhibiting such a behavior necessarily have a single point of intersection corresponding to the solution of Eq. (19). It follows from the mirror nesting condition that such a solution may lead to a vector $\boldsymbol{\chi} \neq 0$ collinear to the total momentum **K** of a pair, which reflects the asymmetry of pair excitation energy (2) in different energy bands (when $s \neq s'$).

The equations describing the case when $\boldsymbol{\chi} \neq 0$ can be reduced to the equations considered above for $\boldsymbol{\chi} = 0$. For this purpose, it is sufficient to displace the origin of the momenta of the relative motion from the center of domain Ξ to the point corresponding to the vertex of paraboloid (18). Setting in Eq. (18) $\mathbf{k} = \mathbf{p} + \mathbf{q}_0$, where $\mathbf{q}_0 = \boldsymbol{\chi}/2b$ is the position of the vertex and $\mathbf{p}$ is the momentum measured from the new origin, we can arrive at the previous expression (21) for the order parameter as a function of momentum $\mathbf{p}$, in which a and b are the parameters to be determined. The system of equations for a and b can be derived, using the above method, from fundamental equation (9) by passing in it to new variables $\mathbf{p} = \mathbf{k} - \mathbf{q}_0$ and $\mathbf{p}' = \mathbf{k}' - \mathbf{q}_0$, and it has exactly the same form (22) as in the case when $\boldsymbol{\chi} = 0$. However, the coefficients of the system of equations obtained in this way, which contain integrals J_n over domain Ξ of functions of the form $f(\mathbf{p})p^{2n}$, where $f(\mathbf{p})$ is defined by formula (8), generally depend on vector $\mathbf{q}_0$, which must now appear in the definition (2) of the pair excitation energy. Thus, solution a, b to system of equations (22) turns out to be a function of $\mathbf{q}_0$ as a parameter. In order to determine this parameter, we must use the additional conditions

$$\int_{\Xi} f(\mathbf{p})(a - bp^2)\mathbf{p}d^2p = 0, \tag{29}$$

emerging from Eq. (19) upon a transition to the new variable $\mathbf{p}$. For $\boldsymbol{\chi} = 0$, condition (29) is satisfied identically in view of the system symmetry. For $\boldsymbol{\chi} \neq 0$, this condition implicitly defines the relation between vector $\mathbf{q}_0$ and parameters a and b.

Thus, for $\boldsymbol{\chi} \neq 0$, the SC order parameter is also a simple quadratic function of the form (21), which depends on momentum $\mathbf{p} = \mathbf{k} - \mathbf{q}_0$, with parameters a and b defined by system of equations (22) under the additional condition (29). Consequently, we can confine the subsequent analysis to the case $\boldsymbol{\chi} = 0$.

5. ORDER PARAMETER FOR $T = 0$

In the limiting case of zero temperature, the function $h(T)$ defined in Eq. (7) is equal to unity and Eqs. (27) and (28) assume the form

$$\frac{1}{4\pi}\int\limits_{\Xi} \frac{(k^{2n} - k_0^{2n})d^2k}{\sqrt{\xi^2(\mathbf{k}) + b^2(k^2 - k_0^2)^2}} = 2^n; \qquad (30)$$

here and below, $n = 1, 2$. It is impossible to obtain an analytic solution of Eq. (30) in the entire range of the unknowns ($b > 0$, $0 < k_0^2 < 1/2$), since the integrals in Eq. (30) depend on the shape of the integration domain. However, the form of these integrals can easily be established in two limiting cases of very large ($b \longrightarrow \infty$) and very small ($b \longrightarrow +0$) values of the energy scale of the gap. For example, for large values of b, we can disregard the first term in the radicand in each of the integrals as compared to the second term so that the integrals in Eq. (30) turn out to be proportional to $1/b$ and the solution of system (30) can be easily found.

While analyzing the behavior of the integrals in Eqs. (30) in the physically interesting case of small b, we must take into account the fact that the pair excitation energy (2) vanishes on the line separating the filled, $\Xi^{(-)}$, and the vacant, $\Xi^{(+)}$, parts of domain Ξ, since this line (PFC) is a part of the FC. Consequently, for $b \longrightarrow +0$, the integrals in Eqs. (30) diverge. The solution to system of equations (30) for small values of b is determined by the type of this divergence, which can be established by taking into account the fact that the main contributions to the integrals in Eqs. (30) comes from a small part of domain Ξ in the form of a narrow strip along the line $\xi(\mathbf{k}) = 0$. We denote the width of this strip by $2\Delta k$ and describe the shape of the line $\xi(\mathbf{k}) = 0$ in polar coordinates: $k = k(\varphi)$. Let us consider the integral of the general form,

$$J = \frac{1}{4\pi}\int\limits_{\Xi} \frac{F(k)kdkd\varphi}{\sqrt{\xi^2(k, \varphi) + b^2(k_0^2 - k^2)^2}}, \qquad (31)$$

where $F(k)$ is a continuous function without singularities. We will determine the behavior of this integral for $b \longrightarrow +0$, taking a narrow strip of width $2\Delta k$ along the PFC for the integration domain. Integrating externally with respect to the polar angle between φ_1 and φ_2 corresponding to the end points of the PFC (line $\xi(k, \varphi) = 0$) and internally with respect to argument k between $k(\varphi) - \Delta k$ and $k(\varphi) + \Delta k$, we pass to the new variable ξ in the internal integral. In view of the small width Δk of the strip, variables k and ξ are connected with each other via the linear relation

$$\xi \approx (d\xi(k(\varphi))/dk)(k - k(\varphi)).$$

In addition, in the approximation adopted here, argument k in the integrand of Eq. (31) can be replaced by its constant (for a given φ) value on the PFC, $k = k(\varphi)$. Integrating with respect to ξ, we can obtain the singular (logarithmic in b) contribution to integral (31). In addition to the singular contribution, a regular (without a singularity for $b \longrightarrow +0$) contribution also exists. This contribution is determined by an integral over the entire domain Ξ and, hence, depends on its size and shape as well as on the form of the one-particle dispersion relation and on k_0^2 as a parameter. Thus, integral (31) can be written in the form

$$J = A\ln(1/b) + C, \qquad (32)$$

where the coefficient of the logarithm has the form

$$A = \frac{1}{2\pi}\int\limits_{\varphi_1}^{\varphi_2} \frac{kF(k)}{|\xi'_k|}d\varphi. \qquad (33)$$

Integration in Eq. (33) is carried out along the PFC, i.e., along the line $\xi(k, \varphi) = 0$ on which $k = k(\varphi)$ and $\xi'_k = (\partial\xi/\partial k)_{k = k(\varphi)}$. The part of the integral regular in b and denoted by C can be taken equal to its value for $b = 0$ in the case of small b.

Thus, for small values of b, the system of equations (30) assumes the form

$$(A_n - k_0^{2n}A_0)\ln(1/b) = P_n, \qquad (34)$$

where A_n is defined by expression (33) with $F(k) = k^{2n}$, while the right-hand sides of Eqs. (34) are functions of k_0^2 and are given by

$$P_n = 2^n - C_n + k_0^{2n}C_0. \qquad (35)$$

Here, C_n are the regular parts of integrals (23) at $T = 0$.

Using relation (34), we can show that system of equations (30) obviously has a solution determining the parameters of the SC gap (b and k_0^2). For this purpose, we consider each of the equations (30) separately as the equation defining a certain line on the plane b, k_0^2. The solution of the system corresponds to the point of intersection of these two lines. We set $k_0^2 = 0$ in Eqs. (30); then the left-hand sides of each of these equations define certain functions of argument b, which decrease monotonically from $+\infty$ for $b \longrightarrow +0$ to zero for $b \longrightarrow +\infty$. In view of inequality (16), the second of these functions, corresponding to $n = 2$, is obviously smaller than the first one (with $n = 1$) for any given b. Consequently, the solution $b_1(0)$ to the first equation in system (30), corresponding to $k_0^2 = 0$, turns out to be always larger than solution $b_2(0)$ to the second equation whose right-hand side is greater than the right-hand side of the first equation (Fig. 6). Let us now consider the other limiting case when $b \longrightarrow +0$ and use the asymptotic form of Eqs. (30) in the form (34). Since the right-hand sides of Eq. (34) are finite and $\ln(1/b) \longrightarrow +\infty$ for $b \longrightarrow +0$, the values of $k_0^2(0)$ corresponding to $b \longrightarrow +0$ are defined as $k_{01}^2 = A_1/A_0$ for $n = 1$ and $k_{02}^2 = \sqrt{A_2/A_0}$ for $n = 2$. Taking into account condition (16), we can verify that $k_{0n}^2 < 1/2$. The quantities A_0, A_1, and A_2 defined by integrals of type (33) satisfy the Cauchy–Buniakowski inequality $A_1^2 \le A_0 A_2$; consequently, $k_{01}^2 \le k_{02}^2$, as shown in Fig. 6. It is obvious from this figure that the lines in question intersect at a single point and, hence, system of equations (30) has a solution.

In order to obtain this solution in the case of small values of b, we must divide the equations in (34) termwise, which leads to a closed equation for determining parameter k_0^2 (it was proved above that this equation always has the required solution $k_0^2 < 1/2$). In this case, the energy scale of the superconducting gap at $T = 0$ can be determined from the obtained value of k_0^2 with the help of any equation from system (34), for example,

$$b = \exp\left(-\frac{P_1}{A_1 - A_0 k_0^2}\right); \tag{36}$$

in accordance with Eq. (24), this relation completely defines the dependence of the energy gap on the momentum of relative motion of a pair at $T = 0$. Since $b \ll 1$, the fraction in the exponent is obviously positive.

The following important circumstance is worth noting. In the case of small b considered here, the value of parameter k_0 determined from system of equations (30) is such that the line of the π phase shift of the SC order parameter ($k = k_0$) corresponding to this value necessarily intersects with the PFC. Indeed, let us suppose that b and k_0 are the solutions of system (30). For $b \longrightarrow +0$, the integrals in Eq. (30) are determined by their singular parts emerging from integration over a small neighborhood of the PFC. If the circle $k = k_0$ did not intersect the PFC, the PFC, together with its small neighborhood, would lie completely inside or outside this circle. Consequently, either inequality $k < k_0$ or the opposite inequality $k > k_0$ would be satisfied everywhere within this neighborhood so that the functions in the integrands of system (30) would preserve a constant sign in this neighborhood (obviously, this sign is positive since the sign of the right-hand sides in Eqs. (30) matches the inequality $k > k_0$). In this case, the integral with $n = 2$ in system (30) would be smaller than the integral with $n = 1$ in view of the inequality $k^2 + k_0^2 < 1$ following from condition (16). Thus, under the above assumption that the circle $k = k_0$ and the PFC do not intersect, the equations in system (30) are incompatible. Consequently, for $b \longrightarrow +0$, parameter k_0 is necessarily such that the intersection of the PFC and the circle $k = k_0$ takes place and, hence, the quantity $k^2 - k_0^2$ has different signs on different segments of the PFC (on both sides of the circle $k = k_0$). This conclusion, which is valid for $b \longrightarrow +0$, remains in force for fairly small finite values of parameter b also.

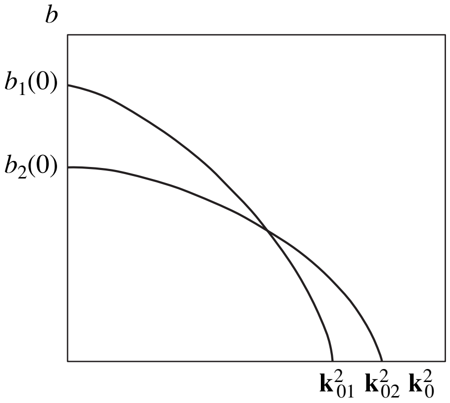


Fig. 6. Schematic illustration of the solution to the system of equations (30).

For values of b and k_0 that are solutions to Eqs. (30), the integral appearing in the second equation (with $n = 2$) must be greater than the integral in the first equation (with $n = 1$), although it contains an additional weighting factor $k^2 + k_0^2 < 1$. Consequently, the positive and negative contributions to the second integral, which stem from regions of domain Ξ where the integrand is positive ($k > k_0$) and negative ($k < k_0$), are compensated to a much lesser extent than in the case of the first integral. Since it is sufficient, while evaluating the integrals for small b, to take into account only the contributions from a small neighborhood of the PFC, the presence of the weighting factor $k^2 + k_0^2$ is manifested the more strongly, the larger the difference between the minimal (k_m) and the maximal (k_M) distances from the center of domain Ξ and the PFC. If the values of k_m and k_M are close, i.e., the PFC is near the line of the π phase shift of the SC order parameter ($k = k_0$) over its entire length, the presence of the weighting factor $k^2 + k_0^2$ may become significant only in the case when the other factor in the integrand of Eqs. (30) varies strongly in the vicinity of the PFC (this is observed for extremely small values of b). Thus, as the PFC approaches the line of the π phase shift, the energy scale of the SC gap must decrease considerably.

Since the circle $k = k_0$ intersects the PFC, we have $k_m < k_0 < k_M$. Therefore, using explicit expressions (33) for quantities A_n (for $F(k) = k^{2n}$), we can easily estimate the combination $A_1 - k_0^2 A_0$ appearing in expression (36) as

$$\left|A_1 - k_0^2 A_0\right| \le \frac{\Delta\varphi}{2\pi}\frac{k_M^3 - k_m^3}{\langle v_F\rangle}, \tag{37}$$

where $\Delta\varphi = \varphi_2 - \varphi_1$ is the angular measure of the PFC and $\langle v_F\rangle$ is a certain dimensionless Fermi velocity averaged over the PFC:

$$\frac{1}{\langle v_F\rangle} = \frac{1}{\Delta\varphi}\int_{\varphi_1}^{\varphi_2}\frac{d\varphi}{|\xi'_k|} \tag{38}$$

(since the PFC is a certain part of the FC, the estimate $|\xi'_k| = k^{-1}|(\mathbf{k}, \nabla_{\mathbf{k}}\xi(\mathbf{k}))| \approx v_F$ holds for this entire part). Thus, expression (36) and estimate (37) show that the width of the SC gap decreases exponentially with the angular measure $\Delta\varphi$ of the PFC, i.e., upon a decrease in the PFC length. In addition, the width of the SC gap decreases exponentially upon a decrease in the difference $k_M - k_m$ characterizing the degree of anisotropy of the PFC.

6. ORDER PARAMETER FOR $T \neq 0$

The SC gap parameters for a nonzero temperature are determined by the same equations (27) and (28), in which the temperature dependence of integrals (23) is contained in the function $f(\eta(\mathbf{k}, T))$ defined by the second relation in (8). It is convenient to write Eqs. (27) and (28) in the form

$$I_n(b, \tau) \equiv \int_\Xi (k^{2n} - k_0^{2n}) f(\mathbf{k}) d^2k = 2^n, \tag{39}$$

where we have introduced a new dimensionless variable $\tau = 2T$. It follows from (26) that system (39) has a solution in a bounded interval of values of parameter τ ($0 \le \tau \le \tau^*$); it is clear from physical considerations that the energy scale b of the SC gap is a monotonically decreasing function of τ, which vanishes at $\tau = \tau^*$. It is impossible to obtain an analytic solution to Eqs. (39), which would determine two functions $b(\tau)$ and $k_0(\tau)$ in the entire range of variation of temperature τ. However, using simple transformations, we can reduce these equations to a form enabling us to analyze the behavior of the solutions to these equations in the two important limiting cases: in the vicinity of absolute zero ($\tau \longrightarrow +0$) and in the vicinity of temperature τ^* corresponding to the disappearance of the SC gap ($\tau \longrightarrow \tau^* - 0$).

Let us differentiate the integrals on the left-hand sides of Eqs. (39) with respect to parameter τ:

$$\frac{\partial I_n}{\partial \tau} = -\frac{1}{4\pi\tau^2}\int_\Xi \frac{k^{2n} - k_0^{2n}}{\cosh^2(\eta(\mathbf{k})/\tau)} d^2k. \tag{40}$$

It is natural to assume that the energy scale of the range of variation of the pair excitation energy $\xi(\mathbf{k})$, corresponding to domain Ξ in the momentum space, is much larger than the energy corresponding to the SC transition temperature. In this case, the main contribution to integrals (40) comes from a small neighborhood of the PFC, while the integrand is vanishingly small in the remaining part of Ξ, where the ratio $\eta(\mathbf{k})/\tau$ is large. For this reason, analogously to the procedure used for evaluating the integrals in Eqs. (30), we pass in Eq. (40) to integration with respect to the angular variable φ and the reduced excitation energy $u = \xi/\tau$, whose range of variation can be extended to infinity. Returning from the partial derivatives (40) to the initial quantities $I_n(b, \tau)$, we can write these quantities in the form

$$I_n(b, \tau) = \int_0^\tau \frac{\partial I_n}{\partial \tau} d\tau + I_n(b, 0), \tag{41}$$

where $I_n(b, 0) = J_n - k_0^2 J_0$ are the vales of the integrals in question for $\tau = 0$. Using the change of variables $\tau = b\left|k^2 - k_0^2\right|/v$, we transform the integral appearing in Eq. (41) to

$$\int_0^\tau \frac{\partial I_n}{\partial \tau} d\tau = -\int_{\varphi_1}^{\varphi_2}\int_{-\infty}^{\infty}\int_{v}^{\infty} \frac{k(k^{2n} - k_0^{2n}) dv\, du\, d\varphi}{4\pi|\xi'_k| v \cosh^2(\sqrt{u^2 + v^2})}, \tag{42}$$

where the external integration with respect to angle φ is carried out, analogously to Eq. (33), over the PFC; $k = k(\varphi)$; and φ_1 and φ_2 correspond to the end points of the PFC. Finally, the double integral over the half-space of variables u and v in Eq. (42) can be reduced, after transformation to polar coordinates ψ and $w = \sqrt{u^2 + v^2}$) and elementary integration with respect to w, to a single integral. After the substitution $\sin\psi = 1/\cosh z$, Eq. (42) assumes the form

$$\int_0^\tau \frac{\partial I_n}{\partial \tau} d\tau = \int_{\varphi_1}^{\varphi_2} G\left(\frac{b|k^2 - k_0^2|}{\tau}\right) \frac{k d\varphi (k_0^{2n} - k^{2n})}{\pi |\xi_k'|}, \quad (43)$$

where symbol $G(v)$ stands for the auxiliary function

$$G(v) = \int_0^\infty \frac{dz}{1 + \exp(2v\cosh z)}, \quad (44)$$

which decreases rapidly upon an increase in its argument. In accordance with Eq. (43), the first term in Eq. (41) is a function of the ratio b/τ. The explicit form of this dependence cannot be established in the general case since it is determined by the specific form of the PFC and, in addition, integral (44) cannot be expressed in terms of the known functions. However, expression (43) makes it possible to analyze the form of the solution to system (39) for $\tau \longrightarrow 0$ and $\tau \longrightarrow \tau^*$ and to trace qualitatively the behavior of this solution in the entire interval $0 \le \tau \le \tau^*$. We assume that the SC gap width for $\tau = 0$ is quite small, so that we can use representation (32) for integrals J_n and write Eqs. (39) combined with Eq. (41) in the form

$$(A_n - k_0^{2n} A_0) \ln\left(\frac{1}{b}\right) + \int_0^\tau \frac{\partial I_n}{\partial \tau} d\tau = P_n, \quad (45)$$

where A_n and P_n have the same meaning as in Eq. (34).

In order to the determine the form of functions $k_0(\tau)$ and $b(\tau)$ in the low-temperature limit, it is sufficient, in evaluating integrals (43), to confine the analysis to the range of large values of b/τ, since parameter b for $\tau \longrightarrow 0$ is close to its value $b(0)$ for $\tau = 0$. In this case, the main contribution to the values of integrals (43) comes from the segments of the PFC for which the value of $|k^2 - k_0^2|$ is minimal; i.e., this contribution comes from the segments adjoining the points of intersection of the PFC with the line $k = k_0$ of the π phase shift of the SC order parameter. In the case of small b, in view of the symmetry of Ξ, such an intersection takes place at least at four points of the momentum space, arranged symmetrically in each quadrant of the coordinate system with the center at $\mathbf{K}/2$.

The contributions to integral (43) from each such point are identical and are determined by their small neighborhood in which the quantity $q = k - k_0$ should be regarded as small and the dependence $q(\varphi)$ should be treated as linear. Integrating in Eq. (43) over this neighborhood, it is convenient to pass from angle φ to variable q, confining the analysis to the first terms of the expansion of difference $k^{2n} - k_0^{2n}$ into a power series in q in view of the smallness of this quantity and assuming that the quantity $|\xi_k'|$ is constant. Since the ratio b/τ is large for $\tau \longrightarrow 0$, function (44) rapidly approaches zero upon an increase in $|q|$; consequently, integration with respect to q can be extended to the entire numerical axis, after which it can be carried out easily with respect to variables q and z and leads to the expression

$$\int_0^\tau \frac{\partial I_n}{\partial \tau} d\tau \approx Q_n \frac{\tau^3}{b^3}, \quad (46)$$

where Q_n are coefficients depending on parameter k_0. Thus, in the low-temperature limit, the system of equations (39) assumes the form

$$(A_n - k_0^{2n} A_0) \ln\left(\frac{1}{b}\right) = P_n - Q_n \frac{\tau^3}{b^3}. \quad (47)$$

The second terms on the right-hand sides of these equations are small corrections; consequently, this system of equations for $\tau \longrightarrow 0$ can be solved by the method of successive approximations. Using the solution to system of equations (34) as the zeroth approximation, we assume that the quantities k_0 and b in the additional terms are equal to their values for $\tau = 0$. The values of $b(\tau)$ and $k_0(\tau)$ can now be determined by the same method as that used for solving system (34). The presence in Eqs. (47) of small terms on the order of τ^3 leads to corrections of the same order of magnitude in the solution,

$$k_0(\tau) = k_0(0) + \gamma_1 \tau^3, \quad b(\tau) = b(0) - \gamma_2 \tau^3, \quad (48)$$

where γ_1 and γ_2 are certain coefficients, the second of which must be a positive number, since the energy scale of the gap obviously decreases upon an increase in temperature. Expressions (48) define the first two terms in the expansions of functions $b(\tau)$ and $k_0(\tau)$ into power series in τ for $\tau \longrightarrow 0$. It follows from Eqs. (48) that the SC order parameter in the vicinity of $\tau = 0$ decreases according to a cubic law upon an increase in temperature.

The decrease in b with increasing temperature must lead to the vanishing of the SC order parameter at a certain temperature $\tau = \tau^*$: $b(\tau^*) = 0$. In order to determine this temperature and establish the way in which the value of $b(\tau)$ approaches zero, we must analyze the behavior of integrals (43) for small values of the ratio b/τ. For small values of its argument, the function $G(v)$ defined by Eq. (44) can be written in the form

$$G(v) \approx \frac{1}{2} \ln\left(\frac{c_1}{v}\right) + c_2 v^2, \quad (49)$$

where $c_1 \approx 1$ and $c_2 > 0$; the singular term (for $v \longrightarrow 0$) in expansion (49) can be derived from Eq. (44) if we

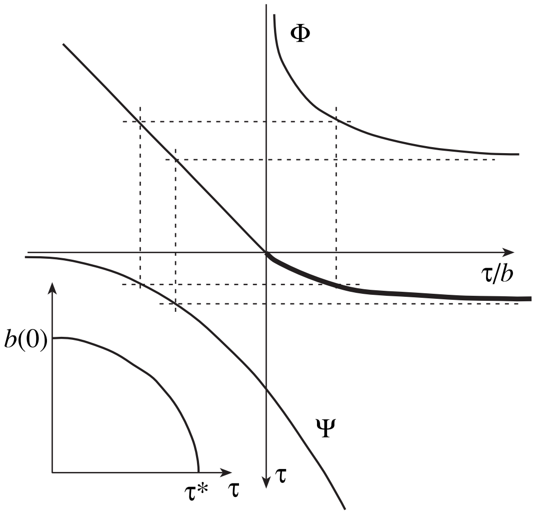


Fig. 7. Graphic solution to Eq. (53), determining the temperature dependence of the order parameter. The inset at the bottom shows the temperature dependence of the SC order parameter.

replace the exponent in Eq. (44) by the two terms of its series expansion in view of the smallness of v. The regular part of expression (49) does not contain a term linear in v, since function $G(v)$ is the primitive of an odd function of argument v, which appears in the integrand in expression (42). Substituting expression (49) into Eq. (43), we can transform the system of equations (45) to

$$(A_n - k_0^{2n}A_0)\ln\left(\frac{1}{\tau}\right) = R_n + S_n\frac{b^2}{\tau^2}, \qquad (50)$$

where S_n and R_n are coefficients depending on k_0.

If we set $\tau = \tau^*$ in Eqs. (50), these equations assume the form

$$(A_n - k_0^{2n}A_0)\ln(1/\tau^*) = R_n \qquad (51)$$

in view of the condition $b(\tau^*) = 0$. System (51) is similar to (34) and can be solved analogously. This enables us to determine the temperature τ^* at which the SC order parameter vanishes, as well as the value of parameter k_0 for $\tau = \tau^*$.

Dependences $b(\tau)$ and $k_0(\tau)$ for $\tau \longrightarrow \tau^* - 0$ can be determined from system of equations (50), in which the terms proportional to b^2/τ^2 are small corrections because $b \longrightarrow 0$. Consequently, Eqs. (50) can be solved by the method of successive approximations using the solution τ^*, $k_0(\tau^*)$ to system (51) as the zeroth approximation. In view of the formal analogy between Eqs. (50) and (47), we can easily find that the behavior of parameters b and k_0 for $\tau \longrightarrow \tau^* - 0$ can be described by the relations

$$b(\tau) \propto \sqrt{1-\tau/\tau^*}, \quad k_0(\tau) = k_0(\tau^*) + \gamma_3(\tau^* - \tau), \qquad (52)$$

where γ_3 is a certain coefficient. It should be noted that the first of these coefficients corresponds to the temperature dependence of the order parameter in the nonsymmetric phase in the vicinity of the transition point, which follows from the general theory of second-order phase transitions. For this reason, the formally introduced temperature τ^* can be treated as the phase-transition temperature below which the system passes to the SC state.

Using Eqs. (45), we can trace qualitatively the $b(\tau)$ dependence in the entire temperature range $0 \le \tau \le \tau^*$ if we preliminarily transform these equations as follows. We subtract from the second equation (with $n = 2$) the first equation multiplied by $2k_0^2$ and then add the quantity $(A_2 - 2k_0^2A_1 + k_0^4A_0)\ln\tau$ to both sides of the equality obtained. Taking into account the explicit form of Eqs. (33) for A_n as integrals along the PFC, as well as expression (43), we can obtain an equation of the form

$$\Phi(\tau/b) = \Psi(\tau), \qquad (53)$$

where

$$\Psi(\tau) = \ln\left(\frac{1}{\tau}\right)\int_{\varphi_1}^{\varphi_2}\frac{k(k^2-k_0^2)^2\,d\varphi}{2\pi|\xi'_k|} - (P_2 - 2k_0^2P_1) \qquad (54)$$

is a simple logarithmic function with a positive coefficient of the logarithm, while the function on the left-hand side of Eq. (53) is defined as

$$\Phi(\tau/b) = \frac{1}{2\pi}\int_{\varphi_1}^{\varphi_2}\frac{k(k^2-k_0^2)^2}{|\xi'_k|} \times\left[2G\left(\frac{b|k^2-k_0^2|}{\tau}\right) + \ln\left(\frac{b}{\tau}\right)\right]d\varphi. \qquad (55)$$

Taking into account definition (44) for $G(v)$, we can prove that the integrand in Eq. (55) is nonnegative and that function (55) decreases monotonically upon an increase in its argument; this function has a finite limit $\Phi_0 > 0$ for $\tau/b \longrightarrow \infty$ and a logarithmic singularity of the form $\ln(b/\tau)$ for $\tau/b \longrightarrow 0$. The definitions of functions Φ and Ψ imply that these functions preserve their behavior for any values of parameter k_0. Taking into account a comparatively small range of this parameter ($0 < k_0^2 < 0.5$), we can qualitatively analyze the dependence of τ/b on τ using relation (53) and assuming that k_0^2 is equal to its certain mean value in the interval $0 < k_0^2 < 0.5$. Plotting the graphs of functions Φ and Ψ of arguments τ/b and τ, respectively, in the first and third quadrants of the coordinate system, and the straight line $\Phi = \Psi$ in the second quadrant, we can obtain the curve describing the dependence of τ/b on τ in the fourth quadrant (Fig. 7). It can be seen that, for small values of τ, the curve has a segment on which $\tau/b \propto \tau$, which corresponds to the fact that function $b(\tau)$ for $\tau \longrightarrow 0$ is close to constant $b(0)$. As $\tau \longrightarrow \tau^* - 0$, the curve describing the dependence of τ/b on τ asymptotically approaches the straight line $\tau = \tau^*$, which corresponds to a decrease in parameter b to zero for $\tau \longrightarrow \tau^*$. The general form of the dependence of the gap parameter b on τ is depicted in Fig. 7.

7. ENERGY GAP AND THE SUPERCONDUCTING TRANSITION TEMPERATURE

The SC transition temperature can be connected to the mean value Δ_0 of the order parameter at zero temperature. Indeed, the gap parameters at zero temperature, $b(0)$ and $k_0(0)$, are defined by system of equations (34), while the transition temperature τ^* and the parameter k_* corresponding to it are defined by system (51); it is convenient to write this system in greater detail, explicitly indicating the argument of all functions:

$$(A_n - k_*^{2n}A_0)\ln(1/\tau^*) = P_n(k_*^2) + T_n(k_*^2). \tag{56}$$

In writing this equation, we have taken into account the fact that, as follows from the derivation of Eqs. (51), $R_n = P_n + T_n$, where T_n denote the integrals along the PFC,

$$T_n(k_*^2) = \int_{\varphi_1}^{\varphi_2} \frac{k(k^{2n} - k_*^{2n})}{2\pi|\xi'_k|}\ln\left(\frac{c_1}{|k^2 - k_*^2|}\right)d\varphi, \tag{57}$$

while $P_n(k_*^2)$ are defined in accordance with Eq. (35); however, in contrast to system (34), these quantities are calculated for $k_0 = k_*$.

Our aim is to establish approximately, to within an order of magnitude, the relation between the mean value $\Delta_0 \equiv \overline{\Delta_0(k)}$ of the SC gap,

$$\Delta_0(k) = b(0)(k_0^2(0) - k^2), \tag{58}$$

at $\tau = 0$ and the transition temperature. First of all, we will prove that the quantities $k_0(0)$ and k_* differ insignificantly from each other and can be treated as identical to a good approximation. Indeed, the value of k_* can be determined from system (56) by termwise division of the equations constituting this system. This leads to a relation defining the quantity k_*, which differs from the equation for $k_0(0)$ following from system (34) in that it contains additional terms of the form

$$(A_1 - k_*^2A_0)T_2(k_*^2) - (A_2 - k_*^4A_0)T_1(k_*^2). \tag{59}$$

It can easily be verified that the last expression is generally quite small. Indeed, in accordance with relation (57), the quantity $T_n(k_*^2)$ differs from the corresponding quantity $(A_n - k_*^{2n}A_0)$ written in the form of an integral along the PFC in the presence of an additional logarithmic term in the integrand. However, owing to the fact that the presence of factor $k^2 - k_*^2$ in integrals (57) vanishes at the points $k = k_*$ where the PFC intersects the line of the π phase shift of the order parameter, the logarithmic singularity is considerably suppressed. In addition, the absolute value of difference $k^2 - k_*^2$ on the entire PFC assumes values which are much smaller than 1/2. Under these conditions, the logarithmic factor can be replaced by a constant value corresponding to a certain value $\tilde{k}$ taken on the PFC. If, accidentally, these values for T_1 and T_2 were identical, expression (59) under investigation would precisely vanish. In view of the small difference between these values, this expression is very small (although it differs from zero) and can be neglected. In this approximation, the equation defining parameter k_* coincides in form with the equation for $k_0(0)$; consequently, we can assume that $k_* = k_0(0)$. Taking this fact into consideration, we subtract the equations forming system (34) from the corresponding equations of system (56) and replace the logarithms in T_n by their average values. Cancelling out $A_n - k^{2n}(0)A_0$ in the equalities obtained and disregarding the difference between the above-mentioned mean values within the approximation used here, we arrive, for $n = 1$ as well as for $n = 2$, at the approximate relation

$$\ln\left(\frac{b(0)}{\tau^*}\right) \approx \ln\left(\frac{c_1}{|\tilde{k}^2 - k_0^2(0)|}\right), \tag{60}$$

where $\tilde{k}$ is the value of k at a certain point of the PFC and, as was noted in connection with Eq. (49), $c_1 \approx 1$. Relation (60) leads to the conclusion that the relation

$$\tau^* \sim \frac{1}{\Delta\varphi}\int_{\varphi_1}^{\varphi_2} |\Delta_0(k)|d\varphi \equiv \Delta_0, \tag{61}$$

holds to within a factor on the order of unity, where integration is carried out along the PFC whose end points correspond to polar angles φ_1 and φ_2, and $\Delta\varphi = \varphi_2 - \varphi_1$. In accordance with relation (61), the SC transition temperature is equal in order of magnitude to the absolute value of the order parameter at zero temperature, averaged over the PFC.

Taking into account the explicit form (58) of quantity $\Delta_0(k)$, we can conclude that the integral in relation (61) is numerically equal to the product of $2b(0)$ by the area of the domain enclosed between the PFC and the line of the π phase shift of the SC order parameter in the interval of polar angles from φ_1 to φ_2. This immediately leads to the following estimate for the SC transition temperature:

$$T^* \lesssim b(0)(k_M^2 - k_m^2), \tag{62}$$

where, as before, k_m and k_M are the minimal and maximal distances from the center of domain Ξ to the PFC. It should be noted that estimate (62) differs considerably from the known result in the BCS theory, according to which the ratio $2\Delta/T_C$ of the doubled width of the SC gap at $T = 0$ to the transition temperature T_C is a universal constant equal to 3.52. In accordance with relation (62), the fraction $b(0)/T^*$ depends on the degree of anisotropy of the PFC; for a low degree of anisotropy (for close values of k_m and k_M), the value of this fraction may in principle be an indefinitely large number.

8. DISCONTINUOUS SOLUTIONS OF THE SELF-CONSISTENCY EQUATION

In addition to the exact continuous solution $\Delta(\mathbf{k})$, self-consistency equation (9) can also have approximate discontinuous solutions. In order to obtain such solutions, we must arbitrarily divide domain Ξ into a certain number of subdomains Ξ_p ($p = 1, 2, \ldots, m$) so that each of these subdomains contains a segment of the PFC and define the mean value of the order parameter,

$$\Delta_p = \frac{1}{\Xi_p}\int\limits_{\Xi_p} \Delta(\mathbf{k}) d^2k, \tag{63}$$

in each such subdomain. Then the mean value of the order parameter in the entire domain Ξ can be written as

$$\Delta = \frac{1}{\Xi}\int\limits_{\Xi} \Delta(\mathbf{k}) d^2k = \sum_{p=1}^{m} \frac{\Xi_p}{\Xi}\Delta_p, \tag{64}$$

where the area of Ξ is equal to the sum of the areas of Ξ_p. Averaging over each subdomain directly in the self-consistency equation (9), we can obtain, using certain approximations, a closed system of m equations in the unknown quantities Δ_p. The set of m values of Δ_p, which is a solution to the obtained system, determines the piecewise-constant function of the momentum of the relative motion, which describes "on average" the behavior of function $\Delta(\mathbf{k})$ in domain Ξ. With increasing number m of the subdomains, the discontinuous solutions obtained in this way represent the exact continuous solution more and more accurately.

It should be noted that, in contrast to the BCS model with attraction between particles, in the case of repulsive interaction, we cannot confine the analysis to only one value of Δ (64) averaged over the entire domain Ξ; i.e., we cannot set $m = 1$ since, as noted above, the self-consistency equation (9) in the case of repulsion has no solutions of constant sign. Consequently, the simplest discontinuous solution for repulsion corresponds to $m = 2$, for which domain Ξ is divided only into two subdomains Ξ_1 and Ξ_2, while the self-consistency equation is reduced to a system of two equations defining the unknown quantities Δ_1 and Δ_2, which, obviously, have opposite signs. In the subsequent analysis, we consider only this simplest case.

We write the self-consistency equation (9) at $T = 0$ in the form

$$\Delta(\mathbf{k}) = -\frac{1}{8\pi^2}\int\limits_{\Xi} \frac{U(\mathbf{k}-\mathbf{k}')\Delta(\mathbf{k}')}{\sqrt{\xi^2(\mathbf{k}') + \Delta^2(\mathbf{k}')}} d^2k' \tag{65}$$

and represent the integral on its right-hand side in the form of the sum of integrals over domains Ξ_1 and Ξ_2. Averaging Eq. (65) written in this way over domain Ξ_1 in accordance with rule (63), we obtain quantity Δ_1 on the left-hand side of this equation and the sum of the integrals with respect to variables **k** and **k**' on its right-hand side. In the first of these integrals, both variables vary within the same domain Ξ_1, while in the second integral, the integration with respect to **k** is carried out over Ξ_1, and with respect to **k**', over domain Ξ_2. Similarly, averaging Eq. (65) over domain Ξ_2, we can obtain the second equation with the left-hand side equal to Δ_2 and the right-hand side in the form of the sum of two integrals with appropriate ranges of variables **k** and **k**'.

Since the kernel $U(\mathbf{k} - \mathbf{k}')$ of the integral equation depends on the difference $\mathbf{k} - \mathbf{k}' = \boldsymbol{\kappa}$ of the arguments, it is convenient to pass to the new pair of variables $\boldsymbol{\kappa}$ and **k**' in each of the integrals obtained. If the integration with respect to the old variables **k** and **k**' is carried out over the same domain (say, Ξ_1), variable $\boldsymbol{\kappa}$ will vary in a certain domain of the momentum space, which will be denoted by Ξ_{11}^*; this domain corresponds to all such vectors of momentum transfer upon scattering, when the initial (**k**) and final (**k**') momenta belong to Ξ_1. If, however, integration with respect to **k** and **k**' is carried out over different domains (e.g., over Ξ_1 and Ξ_2, respectively), variable $\boldsymbol{\kappa}$ is defined in a certain domain Ξ_{12}^* that corresponds to values of $\boldsymbol{\kappa}$ for which the initial momentum **k** is defined in Ξ_1 and the final momentum **k**' is defined in Ξ_2. Generally speaking, domains $\Xi_{pp'}^*$, where $p, p' = 1, 2$, depend on momentum **k**'; however, in the approximate evaluation of the integrals, we can assume that these domains are identical for all values of **k**', having determined the corresponding averaged Jacobians $J_{pp'}$ of transformation of the integration variables from the condition of conservation of the number of states,

$$\int_{\Xi_{p'}}\int_{\Xi_p} d^2k d^2k' = J_{pp'}\int_{\Xi_{p'}}\int_{\Xi^*_{pp'}} d^2\kappa d^2k'. \tag{66}$$

It follows from this relation that $J_{pp'} = \Xi_p/\Xi^*_{pp'}$. Using such averaged Jacobians, integration with respect to variables **k** and **k**' can be carried out independently. It should be noted that $\Xi^*_{pp'} = \Xi^*_{p'p}$.

Let us introduce the values of the matrix element of interaction averaged over domains $\Xi^*_{pp'}$,

$$U_{pp'} = U_{p'p} = \frac{1}{\Xi^*_{pp'}}\int_{\Xi^*_{pp'}} U(\kappa)d^2\kappa, \tag{67}$$

and define functionals $f_p\{\Delta(\mathbf{k})\}$ via the relations

$$\Delta_p f_p\{\Delta(\mathbf{k})\} = \frac{1}{(2\pi)^2\Xi_p}\int_{\Xi_p}\frac{\Delta(\mathbf{k}')d^2k'}{\sqrt{\xi^2(\mathbf{k}') + \Delta^2(\mathbf{k}')}}. \tag{68}$$

Replacing the functionals defined in these relations by their values corresponding to the values Δ_1 and Δ_2 of their argument, averaged over Ξ_1 and Ξ_2, i.e., setting

$$f_p \longrightarrow f_p(\Delta_p) = \frac{1}{(2\pi)^2\Xi_p}\int_{\Xi_p}\frac{d^2k'}{\sqrt{\xi^2(\mathbf{k}') + \Delta_p^2}}, \tag{69}$$

and introducing the notation $\alpha = \Xi_1/\Xi$ and, accordingly, $1 - \alpha = \Xi_2/\Xi$, we can obtain the integral relations connecting two unknown quantities Δ_1 and Δ_2:

$$\begin{aligned} 2\Delta_1 &= -\alpha\Xi U_{11}\Delta_1 f_1 - (1-\alpha)\Xi U_{12}\Delta_2 f_2, \\ 2\Delta_2 &= -\alpha\Xi U_{21}\Delta_1 f_1 - (1-\alpha)\Xi U_{22}\Delta_2 f_2. \end{aligned} \tag{70}$$

After evaluation of functions (69), Eqs. (70) form a closed system of transcendental equations in Δ_1 and Δ_2, which is similar to system (22), which emerges in determining the exact continuous solution of the self-consistency equation (9) with the degenerate kernel (15). The integrals over domains Ξ_p, which appear in relation (69) and contain segments of the PFC, are completely analogous to integrals (23) for $T = 0$ and have logarithmic singularities associated with the PFC as those in function Δ_p. Consequently, for small values of Δ_p, these integrals can be represented in the form (32), where b is replaced by Δ_p. Analysis of the condition of nontrivial compatibility of the system of equations (70) shows that this system may have a nonzero solution only under the condition

$$U_{12}U_{21} - U_{11}U_{22} > 0, \tag{71}$$

which in our case plays the role of the Cauchy–Buniakowski inequality $A_1^2 - A_0A_2 < 0$ for quantities A_0, A_1, and A_2 introduced above in the analysis of the continuous solution to the self-consistency equation. It should be noted that disregard of the dependence of the matrix element $U(\mathbf{k} - \mathbf{k}')$ on its argument, which is usual in the BCS theory, leads to identical mean values of the matrix element ($U_{11} = U_{22} = U_{12}$), which leads the equality $U_{12}U_{21} - U_{11}U_{22} = 0$. This corresponds to the statement that, under the condition $U(\mathbf{k} - \mathbf{k}') = \text{const}$, the self-consistency equation in the case of repulsive interaction may have only a trivial solution. Naturally, system of equations (70) leads only to a trivial solution also for $\alpha = 0$ or $\alpha = 1$, i.e., in the case when one of subdomains Ξ_1 or Ξ_2 coincides with the entire domain Ξ.

We can reduce system of equations (70) to a more symmetric form by multiplying the first of the equations by α and the second by $(1 - \alpha)$, and by assuming that $\alpha\Xi U_{11} \approx (1 - \alpha)\Xi U_{22} \equiv U_0'$ and $\alpha(1 - \alpha)\Xi U_{12} \approx \beta U_0'$, where U_0' and β are certain parameters. Thus, system (70) assumes the form

$$\begin{aligned} 2\alpha\Delta_1 &= -\alpha U_0'\Delta_1 f_1 - \beta U_0'\Delta_2 f_2, \\ 2(1-\alpha)\Delta_2 &= -\beta U_0'\Delta_1 f_1 - (1-\alpha)U_0'\Delta_2 f_2. \end{aligned} \tag{72}$$

System of equations (72) written in an equivalent form,

$$\begin{aligned} 2(1-\alpha)[\alpha\Delta_1 - \beta\Delta_2] &= U_0'\gamma\Delta_1 f_1, \\ 2\alpha[(1-\alpha)\Delta_2 - \beta\Delta_1] &= U_0'\gamma\Delta_2 f_2, \end{aligned} \tag{73}$$

where $\gamma = \beta^2 - \alpha(1 - \alpha)$, was obtained and analyzed in [19], where the filled, $\Xi^{(-)}$, and vacant, $\Xi^{(+)}$, parts of domain Ξ, separated by the PFC, were chosen as domains Ξ_1 and Ξ_2 for averaging the order parameter. In the general case, the quantities $U_{pp'}$ appearing in the system of equations (70) should be treated as phenomenological parameters depending on the method of division of domain Ξ into two parts Ξ_1 and Ξ_2. In this sense, the system of equations (73) leads to a family of discontinuous solutions from which we must choose the one that approximates the exact continuous solution in the best way. If we choose the standard deviation

$$\delta^2 = \frac{1}{\Xi}\int_{\Xi}[\Delta(\mathbf{k}) - \Delta_d(\mathbf{k})]^2 d^2k \tag{74}$$

as the measure of the proximity of the approximate discontinuous solution $\Delta_d(\mathbf{k})$ to the exact solution $\Delta(\mathbf{k})$, the minimum condition for δ^2 taking into account definition (63) for the mean values Δ_1 and Δ_2 leads to the requirement that the quantity $\alpha\Delta_1^2 + (1 - \alpha)\Delta_2^2$ must have a maximum. Since this quantity is nonnegative and vanishes for $\alpha = 0$ (when $\Xi_2 = \Xi$) and $\alpha = 1$ (when $\Xi_1 = \Xi$), this quantity obviously attains its maximal value for a certain choice of division of domain Ξ into two parts and, accordingly, for a certain value of α from the interval $0 < \alpha < 1$. From solutions Δ_1 and Δ_2, the one matching such a method of division of domain Ξ can be chosen as the best approximate solutions.

Since the continuous solution (24) to the self-consistency equation possess circular symmetry relative to the center of domain Ξ, the subdomains Ξ_1 and Ξ_2 in the case when domain Ξ is, for example, a circle of a certain radius k_2 can be naturally chosen so that they possess the same symmetry. If Ξ_1 is a circle $0 < k < k_1$ and Ξ_2 is a ring $k_1 < k < k_2$ ($k_1 < k_2$), the domains $\Xi^*_{pp'}$ introduced above are also circles: Ξ^*_{11} has radius $2k_1$, Ξ^*_{22} has radius $2k_2$, and $\Xi^*_{12} = \Xi^*_{21}$ has radius $k_1 + k_2$. For the sake of simplicity, we can confine our analysis to a PFC in the form of a closed line bounding an anisotropic pocket. In this case, the radius k_1 of the filled domain should be chosen so that the circle $k = k_1$ intersects the PFC in such a way that one part of the PFC belongs to Ξ_1 and the other to Ξ_2.

For the degenerate kernel (15) of the integral equation (9), the mean values of the matrix element (67) assume the form

$$U_{pp'} = 2\pi[1-(k_p+k_{p'})^2/4], \qquad (75)$$

which implies that inequality (71) definitely holds and, hence, the system of equations (73) has a nontrivial solution depending on k_1 as on the parameter to be determined from the minimum condition for quantity (74). After this, parameters b and k_0^2 determining the continuous solution to the self-consistency equation and connected (in the case of the circular domains considered here) with the discontinuous solution Δ_1 and Δ_2 via the relations

$$\Delta_1 = \frac{1}{2}b[2k_0^2 - k_1^2], \quad \Delta_2 = \frac{1}{2}b[2k_0^2 - (k_2^2 + k_1^2)], \qquad (76)$$

following from definition (63) and the form of continuous solution (24).

9. PROXIMITY EFFECT IN THE MOMENTUM SPACE

The emergence of a condensate of pairs with the total momentum $\mathbf{K}$, which is accompanied by transition of particles through the PFC, leads to a change in the chemical potential of the subsystem of particles belonging to domain Ξ of the momentum space. Only such particles are taken into account in the Hamiltonian of the BCS type, which therefore acts in the subspace of pairs with momentum $\mathbf{K}$. It follows from the general condition of thermodynamic equilibrium that a change in the chemical potential in subsystem Ξ changes the chemical potential of the entire electron system, i.e., the redistribution of particles in the momentum space. Thus, the filling of states both outside and inside the FC changes:

$$n_{\mathbf{k}_\pm\sigma} = n^{(0)}_{\mathbf{k}_\pm\sigma} + \delta n_{\mathbf{k}\sigma}. \qquad (77)$$

Here, $n_{\mathbf{k}_\pm\sigma}$ is the occupation number for a state of a particle with momentum $\mathbf{k}_\pm$ and with spin $\sigma = \uparrow, \downarrow$; $n^{(0)}_{\mathbf{k}_\pm\sigma}$ is the corresponding occupation number in the nonsuperconducting phase: $n^{(0)}_{\mathbf{k}_\pm\sigma} = 1$ for $k_\pm < k_F$ and $n^{(0)}_{\mathbf{k}_\pm\sigma} = 0$ for $k_\pm > k_F$, k_F being the Fermi momentum in the direction of the particle momentum $\mathbf{k}_\pm$. The change in the distribution function $\delta n_{\mathbf{k}\sigma}$ associated with a small shift of the chemical potential upon condensation differs substantially from zero only in a small neighborhood of the FC, whose width is determined by the emerging SC order parameter. In order to simplify notation in the subsequent analysis, we will use $\uparrow$ and $\downarrow$ instead of combined indices $\mathbf{k}_{+\uparrow}$ and $\mathbf{k}_{-\downarrow}$. Thus, $n^{(0)}_\uparrow\, n^{(0)}_\downarrow$ can be treated as the characteristic function of domain $\Xi^{(-)}$, equal to unity inside this domain and to zero outside this domain. Similarly, $(1 - n^{(0)}_\uparrow)(1 - n^{(0)}_\downarrow)$ is the characteristic function of domain $\Xi^{(+)}$. Therefore, the characteristic function of the entire kinematically allowed domain Ξ can be written as

$$\Theta_0(\mathbf{k}) = n^{(0)}_\uparrow n^{(0)}_\downarrow + (1 - n^{(0)}_\uparrow)(1 - n^{(0)}_\downarrow), \qquad (78)$$

where $\mathbf{k}$ is the momentum of the relative motion of a pair of particles. The redistribution of particles in the momentum space associated with the condensation of pairs with momentum $\mathbf{K}$ leads to effective "expansion" of domain Ξ: the states which can also participate in the formation of a pair with momentum $\mathbf{K}$ are filled (with a small weight in view of the smallness of the SC order parameter) outside this domain. The superconducting order emerging in the previously inaccessible domain of the momentum space outside Ξ must be described, as a certain secondary effect (the proximity effect in the momentum space), by a self-consistency equation taking into account the partial filling of states participating in the pairing of particles with the total momentum $\mathbf{K}$. Introducing the corresponding function $\Theta(\mathbf{k})$, which characterizes the distribution of particles forming a pair with a given $\mathbf{K}$, in the entire Brillouin zone,

$$\Theta(\mathbf{k}) = n_\uparrow n_\downarrow + (1 - n_\uparrow)(1 - n_\downarrow), \qquad (79)$$

we can write the self-consistency equation for $T = 0$ in the form

$$\Delta_{\mathbf{k}} = -\frac{1}{2}\sum_{k'} \frac{U(\mathbf{k}-\mathbf{k}')}{\sqrt{\xi^2_{\mathbf{k}'} + \Delta^2_{\mathbf{k}'}}}\Theta(\mathbf{k})\Theta(\mathbf{k}'), \qquad (80)$$

where summation is carried out over the entire $2D$ Brillouin zone. Factor $\Theta(\mathbf{k})\Theta(\mathbf{k}')$ on the right-hand side of Eq. (80) takes into account the fact that both states of the relative motion of a pair ($\mathbf{k}'$, from which the scattering takes place during the interaction, and $\mathbf{k}$, to which the pair is scattered) must have a nonzero occupation number. The replacement of this factor by $\Theta_0(\mathbf{k})\Theta_0(\mathbf{k}')$ confines the summation over $\mathbf{k}'$ to domain Ξ and leads to an equation of type (6), which implies, by the way, that factor $\Theta(\mathbf{k})\Theta(\mathbf{k}')$ plays a role similar, to a certain extent, to that of the temperature factor $h_{\mathbf{K}k}(T)$.

The product $\Theta(\mathbf{k})\Theta(\mathbf{k}')$ can be written in the form

$$\Theta(\mathbf{k})\Theta(\mathbf{k}') = [\Theta_0(\mathbf{k}) + \delta\Theta(\mathbf{k})][\Theta_0(\mathbf{k}') + \delta\Theta(\mathbf{k}')], \quad (81)$$

where

$$\begin{aligned}\delta\Theta(\mathbf{k}) &= (2n_\downarrow^{(0)} - 1)\delta n_\uparrow \\ &+ (2n_\uparrow^{(0)} - 1)\delta n_\downarrow + 2\delta n_\uparrow \delta n_\downarrow.\end{aligned} \quad (82)$$

Then, Eq. (80) can be written in the form

$$\begin{aligned}\Delta_\mathbf{k} = &-\frac{1}{2}\sum_{k' \in \Xi} \frac{U(\mathbf{k}-\mathbf{k}')\Delta_{\mathbf{k}'}}{\sqrt{\xi_{\mathbf{k}'}^2 + \Delta_{\mathbf{k}'}^2}}[1 + \delta\Theta(\mathbf{k}')][1 + \delta\Theta(\mathbf{k})] \\ &-\frac{1}{2}\sum_{k' \notin \Xi} \frac{U(\mathbf{k}-\mathbf{k}')\Delta_{\mathbf{k}'}}{\sqrt{\xi_{\mathbf{k}'}^2 + \Delta_{\mathbf{k}'}^2}}\delta\Theta(\mathbf{k}')[1 + \delta\Theta(\mathbf{k})]\end{aligned} \quad (83)$$

for the momenta of the relative motion belonging to the kinematically allowed domain Ξ (i.e., for $\mathbf{k} \in \Xi$), while for the momenta of the relative motion, which do not belong to Ξ (i.e., for $\mathbf{k} \notin \Xi$), Eq. (80) can be written as

$$\begin{aligned}\Delta_\mathbf{k} = &-\frac{1}{2}\sum_{k' \in \Xi} \frac{U(\mathbf{k}-\mathbf{k}')\Delta_{\mathbf{k}'}}{\sqrt{\xi_{\mathbf{k}'}^2 + \Delta_{\mathbf{k}'}^2}}[1 + \delta\Theta(\mathbf{k}')]\delta\Theta(\mathbf{k}) \\ &-\frac{1}{2}\sum_{k' \notin \Xi} \frac{U(\mathbf{k}-\mathbf{k}')\Delta_{\mathbf{k}'}}{\sqrt{\xi_{\mathbf{k}'}^2 + \Delta_{\mathbf{k}'}^2}}\delta\Theta(\mathbf{k}')\delta\Theta(\mathbf{k}).\end{aligned} \quad (84)$$

It can be seen from Eq. (82) that functions $\delta\Theta(\mathbf{k})$ and $\delta\Theta(\mathbf{k}')$ noticeably differ from zero only in a narrow strip of states in the vicinity of the FC. Consequently, we can disregard in Eq. (84) the contribution proportional to the product of these functions and write this equation in the form

$$\Delta_\mathbf{k} = \Delta_\mathbf{k}^{(0)}\delta\Theta(\mathbf{k}), \qquad \mathbf{k} \notin \Xi. \quad (85)$$

The quantity

$$\Delta_\mathbf{k}^{(0)} \equiv -\frac{1}{2}\sum_{k' \in \Xi} \frac{U(\mathbf{k}-\mathbf{k}')\Delta_{\mathbf{k}'}}{\sqrt{\xi_{\mathbf{k}'}^2 + \Delta_{\mathbf{k}'}^2}} \quad (86)$$

is the sum over the kinematically allowed domain Ξ; however, momentum $\mathbf{k}$ belongs to the forbidden domain in which, in accordance with Eq. (85), the order parameter differs noticeably from zero only in a small neighborhood of the FC. Therefore, Eq. (86) can be treated approximately as the value of the order parameter at a certain point on the boundary of the kinematically allowed domain, which corresponds to a momentum close to $\mathbf{k}$.

Functions $\delta\Theta(\mathbf{k})$ and $\delta\Theta(\mathbf{k}')$ in the brackets in Eq. (83) substantially differ from zero only in a neighborhood of the PFC, whose statistical weight is small as compared to the weight of the entire summation domain Ξ. Consequently, these functions can be neglected as compared to unity so that the first sum in Eq. (83) is exactly the solution for the SC order parameter, which was determined taking into account the kinematic constraints imposed on the momentum of the relative motion of a pair. This solution, which has the form (86) provided that $\mathbf{k} \in \Xi$, is the contribution from the kinematically allowed domain to the SC order parameter that depends on the momentum of the relative motion of the pair. In the second sum in Eq. (83), summation with respect to $\mathbf{k}'$ is carried out over the kinematically forbidden domain of the momentum space, which contains the part of the isoline of the kinetic energy of the relative motion, whose segments form the PFC in the kinematically allowed domain (Fig. 8). On this isoline, $\xi_{\mathbf{k}'} = 0$ and, hence, summation with respect to $\mathbf{k}'$ could lead to a singular (logarithmic) contribution to the second sum in Eq. (83), which partly compensates the smallness of the function $\delta\Theta(\mathbf{k}')$ if the isoline passes sufficiently close to the FC, where $\delta\Theta(\mathbf{k}') \neq 0$. However, the role of the parameter truncating the logarithm from above is played in this case by the small energy scale corresponding to the continuation of the order parameter to the kinematically forbidden domain. For this reason, the contributions from the terms equal approximately to $\delta\Theta$ are small as compared to the main contribution and can be neglected.

Thus, the SC order parameter outside the kinematically allowed domain Ξ is defined by Eq. (85), in which the function $\delta\Theta(\mathbf{k})$ itself depends on the order parameter $\Delta(\mathbf{k})$ outside domain Ξ and, hence, has to be determined. However, in order to analyze the proximity effect in the momentum space qualitatively, we can assume that

$$\delta\Theta(\mathbf{k}) = 1 - \frac{\xi_1(\mathbf{k})}{\sqrt{\xi_1^2(\mathbf{k}) + \Delta^2(\mathbf{k})}}, \quad (87)$$

where $\xi_1(k) = \varepsilon(\mathbf{K}/2 + \mathbf{k}) - \mu$ is the excitation energy of a particle beyond the kinematically allowed domain (which contains, however, the state $\mathbf{K}/2 - \mathbf{k}$). Such a representation of function $\delta\Theta(\mathbf{k})$ was chosen analogously to the distribution function v_k^2 emerging in the BCS theory, but differs from this function in factor 2

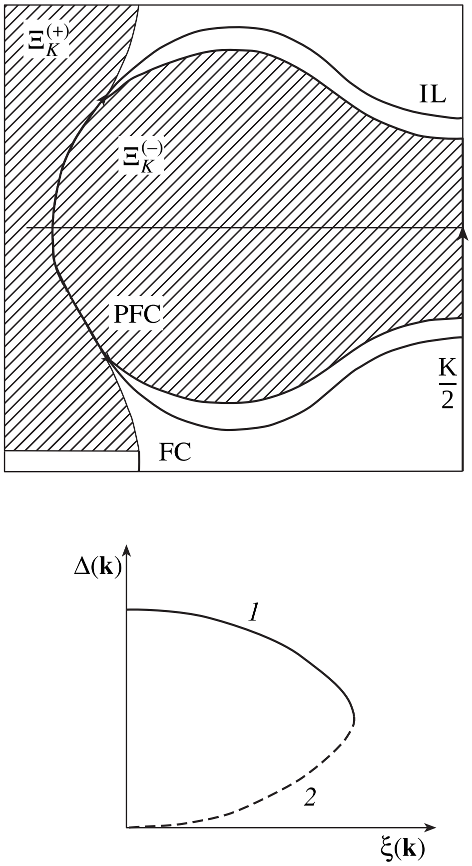


Fig. 8. Kinematically allowed domain Ξ_K (hatched), in which the SC order parameter is defined. A quarter of the Brillouin zone is shown. The isoline (IL) of the excitation energy for a pair with the total momentum **K** (fine curve), coinciding with the FC (bold curve), forms the PFC separating the filled ($\Xi_K^{(-)}$) and vacant ($\Xi_K^{(+)}$) parts of domain Ξ_K. The dependence of the SC order parameter $\Delta(\mathbf{k})$ on the excitation energy $\xi(\mathbf{k})$ in the kinematically forbidden domain of the momentum space is shown below. Curves *1* and *2* are the solutions of the self-consistency equation, which are extended continuously to the forbidden domain of the SC and N solution in the kinematically allowed domain.

taking into account the fact that a particle with any spin component may be located outside the kinematically allowed domain. Omitting argument **k**, we can write Eq. (85) in the form

$$1 - s = \frac{x}{\sqrt{x^2 + s^2}}, \tag{88}$$

where the following nonnegative quantities have been introduced:

$$s \equiv \Delta(\mathbf{k})/\Delta^{(0)}(\mathbf{k}), \quad x \equiv \xi_1(\mathbf{k})/\left|\Delta^{(0)}(\mathbf{k})\right|. \tag{89}$$

Condition $s \geq 0$ stems from the fact that $-\mathbf{k} \in \Xi$ if $\mathbf{k} \notin \Xi$; consequently, if, for example, $n_\uparrow^{(0)} = 0$ and $\delta n_\uparrow > 0$, we have $n_\downarrow^{(0)} = 1$ and $\delta n_\downarrow < 0$ so that definition (82) implies that $\delta\Theta(\mathbf{k}) \geq 0$. Since a particle with momentum $-\mathbf{k}$ of the relative motion corresponds to a state lying deeply in the domain of Fermi filling, we have $\delta n_\uparrow \gg |\delta n_\downarrow|$; in this case, we can approximately assume that $\delta\Theta(\mathbf{k}) \approx \delta n_\uparrow$ (exactly this approximation was used in relation (87)).

The obvious solution $s = 0$ to Eq. (88) for any x corresponds to the N phase. It can also be seen from Eq. (88) that one more solution, $s = 1$, corresponding to a continuous transition of the SC solution from the kinematically allowed domain to the forbidden one, exists on the FC, where $x = 0$, in addition to solution $s = 0$. For $x > 0$, two nontrivial solutions exist, which are transformed to $s \longrightarrow 1$ and $s \longrightarrow 0$ as $x \longrightarrow 0$. Therefore, the SC order parameter in the kinematically forbidden domain must be a continuous extension of the SC solution $s = 1$ at the boundary of this domain. Since Eq. (87) raised to the second power can be reduced to a cubic equation in s, both nontrivial solutions can be written in the explicit form

$$s = \frac{2}{3}\left[1 - \sqrt{1 - 3x^2}\cos\left(\frac{\omega}{3} \mp \frac{\pi}{3}\right)\right], \tag{90}$$

where the upper (lower) sign corresponds to the continuation of solution $s = 1$ ($s = 0$) at the boundary of domain Ξ,

$$\cos\omega = \frac{1 - 18x^2}{(1 - 3x^2)^{3/2}}. \tag{91}$$

Both (real-valued) solutions exist for $0 \leq x \leq x_m$, coinciding for $x = x_m = \sqrt{(5\sqrt{5} - 11)/2}$. The value of the order parameter corresponding to x_m is $s_m = (3 - \sqrt{5})/2$. Thus, the SC order parameter penetrates to a finite depth in the kinematically forbidden domain, differing from zero in a narrow strip in the vicinity of the FC (see Fig. 8). With increasing distance from the FC, the order parameter decreases monotonically and abruptly vanishes at $x = x_m$.

10. CONCLUSIONS

Self-consistency equation (9), defining the SC order parameter in a quasi-two-dimensional electronic system (whose ground state in the N phase is the domain of Fermi filling in the momentum space, which is separated from the domain of vacant one-particle states by the Fermi contour) has a nontrivial solution in the case of the repulsive interaction between electrons, i.e., in the case of a positive scattering length for the relative motion of a pair in the singlet s state (for $u_0 > 0$, where the quantity u_0 defined in Eq. (14) is obviously proportional to the corresponding scattering length).

The necessary condition for the existence of the nontrivial solution is the existence of at least one negative eigenvalue for the kernel $U(\mathbf{k} - \mathbf{k}')$ of the integral equation (9). In the case of a nonnegative kernel, this condition can be satisfied if the interaction potential $U(r)$ for particles in the real space (the Fourier transform of this potential is the kernel of Eq. (9)) assumes negative values within at least one finite segment on the semiaxis $0 < r < \infty$. This property is observed, for example, for the screened Coulomb potential in a Fermi system, which exhibits Friedel oscillations owing to the presence of the FC at large distances as compared to the screening radius: $U(r) \propto r^{-3}\cos 2k_F r$ [31].

However, the existence of a negative eigenvalue for the kernel of Eq. (9) is not a sufficient condition for the emergence of the bound state of a pair, i.e., the emergence of a nontrivial solution to the self-consistency equation. It is necessary that this eigenvalue λ be small in absolute value, as follows from condition (17). It should be noted that an eigenvalue of an integral operator is defined [29] as the factor in front of the integral; therefore, the energy of the bound state emerging in potential $U(r)$ is equal approximately to $|\lambda|^{-1}$. It should also be noted here that the smallness of the amplitude of Friedel oscillations apparently rules out the emergence of bound states in a three-dimensional Fermi system; however, for a quasi-two-dimensional system, it can be expected that at least one local level will appear even in a shallow potential well, owing, among other things, to peculiarities of $2D$ screening [32]. For this reason, the superconducting order in the case of repulsion is inherent, in all probability, precisely in quasi-two-dimensional systems.

For a large total momentum of a pair, the kinematically allowed domain Ξ in the momentum space, which corresponds to the momenta of relative motion, is quite small. Therefore, small momenta of the relative motion correspond to large relative distances between particles in pairs, which correspond exactly to the region of Friedel oscillations. It should be noted in this connection that, in the case of pairing with $K = 0$, when the momenta of the relative motion coincide with the momenta of the particles and are approximately equal to k_F, region $r \lesssim k_F^{-1}$, in which the screened Coulomb potential is definitely positive, is the region of the real space in which the interaction between particles is significant. It should also be noted that the relative smallness of domain Ξ makes it possible to replace the true kernel of integral equation (9) by degenerate kernel (15) which can be used, under condition (17), to obtain a nontrivial solution for the SC order parameter in the explicit form (24).

Mirror nesting, which gives rise to the PFC (i.e., the line on which the segment of the FC coincide with segments of a certain isoline of the kinetic energy of the relative motion of a pair), is the most important condition for the existence of solution (24) in the case of pairing with a large total momentum. It is because of mirror nesting that SC instability under the necessary additional condition (17) can emerge for an arbitrarily small value of repulsive interaction $u_0 > 0$. In a quasi-two-dimensional system such as HTSC cuprates, the mirror nesting condition can be naturally satisfied (at least approximately) for certain total momenta $\mathbf{K}$ of a pair. A deviation from mirror nesting ultimately leads to suppression of SC pairing [18]. It should be noted that this suppression becomes much weaker beyond the framework of the point-particle potential approximation, when the screened Coulomb interaction is taken into account [33].

It follows from Eq. (24) that the SC order parameter changes its sign on a certain line (arc of a circle of radius k_0) intersecting the PFC; in this case, the amplitude of the order parameter substantially depends on the degree of anisotropy of the PFC (if the PFC coincides with the circle $k = k_0$, nontrivial solution (24) does not exist). ARPES experiments show, for example, that strong anisotropy of the electron (hole) dispersion relation in the vicinity of the Fermi level is inherent in HTSC cuprates; the Fermi velocity components for these compounds may differ by more than an order of magnitude [34].

The dependence (24) of the SC order parameter on the momentum of the relative motion corresponds to a certain momentum $\mathbf{K}$ of a pair, which is determined by the mirror nesting condition. The symmetry of the order parameter, corresponding to the zero-current state of a superconductor, is determined by the coefficient in the linear combination of crystallographically equivalent pairs and, hence, depends on the interaction mixing such pairs. A trivial irreducible representation (A_{1g}) of the $2D$ symmetry group of the crystal corresponds to an extended s-wave symmetry of the order parameter, while the other one-dimensional irreducible representation (B_{1g}) corresponds to the d-wave symmetry [19].

ACKNOWLEDGMENTS

This study was partly supported by the Ministry of Education of the Russian Federation (grant no. E02-3.4-147), the Federal Special Program “Integration” (project no. B0049), the Russian Foundation for Basic Research (project no. 02-02-17133), and the Federal Special Program “Studies and Development of Priority Trends in Science and Technology” (state contract no. 40.072.1.1.1173).

REFERENCES

1. J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).
2. V. L. Ginzburg, in *Problems in High-Temperature Superconductivity,* Ed. by V. L. Ginzburg and D. A. Kirzhnits (Nauka, Moscow, 1977), Chap. 1.
3. S. Chakravarty, R. B. Laughlin, D. K. Morr, and C. Nayak, Phys. Rev. B **63**, 094503 (2001).

4. J. Orenstein and A. J. Millis, Science **288**, 468 (2000).
5. Z.-X. Shen, W. E. Spicer, D. M. King, *et al.*, Science **267**, 343 (1995).
6. E. W. Carlson, V. J. Emery, S. A. Kivelson, and D. Orgad, cond-mat/0206217.
7. D. A. Wollman, D. J. V. Harlingen, W. C. Lee, *et al.*, Phys. Rev. Lett. **71**, 2134 (1993).
8. C. C. Tsuei, J. R. Kirtley, C. C. Chi, *et al.*, Phys. Rev. Lett. **73**, 593 (1994).
9. C. Renner, B. Revaz, J.-Y. Genoud, *et al.*, Phys. Rev. Lett. **80**, 149 (1998).
10. Y. de Wilde, N. Miyakawa, P. Guptasarma, *et al.*, Phys. Rev. Lett. **80**, 153 (1998).
11. J. E. Hirsch, Phys. Rev. B **59**, 11962 (1999).
12. J. E. Hirsch, Physica C (Amsterdam) **341–348**, 213 (2000).
13. J. E. Hirsch and F. Marsiglio, Phys. Rev. B **39**, 11515 (1989); F. Marsiglio and J. E. Hirsch, Phys. Rev. B **41**, 6435 (1990).
14. M. R. Norman, M. Randeria, B. Jankó, and J. C. Campuzano, Phys. Rev. B **61**, 14742 (2000).
15. V. I. Belyavskiĭ, V. V. Kapaev, and Yu. V. Kopaev, Zh. Éksp. Teor. Fiz. **118**, 941 (2000) [JETP **91**, 817 (2000)].
16. V. I. Belyavsky and Yu. V. Kopaev, Phys. Lett. A **287**, 152 (2001).
17. V. I. Belyavskiĭ and Yu. V. Kopaev, Zh. Éksp. Teor. Fiz. **121**, 175 (2002) [JETP **94**, 149 (2002)].
18. V. I. Belyavskiĭ, V. V. Kapaev, and Yu. V. Kopaev, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 51 (2002) [JETP Lett. **76**, 44 (2002)].
19. V. I. Belyavsky and Yu. V. Kopaev, Phys. Rev. B **67**, 024513 (2003).
20. P. Fulde and R. A. Ferrel, Phys. Rev., Sect. A **135**, 550 (1964).
21. A. I. Larkin and Yu. N. Ovchinnikov, Zh. Éksp. Teor. Fiz. **47**, 1136 (1964) [Sov. Phys. JETP **20**, 762 (1964)].
22. R. Combescot, Phys. Rev. B **57**, 8632 (1998).
23. R. Hlubina and T. M. Rice, Phys. Rev. B **51**, 9253 (1995).
24. J. Schmalian, D. Pines, and B. Stojković, Phys. Rev. B **60**, 667 (1999).
25. V. L. Gurevich, A. I. Larkin, and Yu. A. Firsov, Fiz. Tverd. Tela (Leningrad) **4**, 185 (1962) [Sov. Phys. Solid State **4**, 131 (1962)].
26. M. L. Cohen, in *Superconductivity*, Ed. by R. D. Parks (Marcel Dekker, New York, 1969; Mir, Moscow, 1972).
27. C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963; Nauka, Moscow, 1967).
28. M. L. Krasnov, *Integral Equations* (Nauka, Moscow, 1975).
29. S. G. Mikhlin, *Lectures on Integral Equations* (Fizmatgiz, Moscow, 1959).
30. V. S. Vladimirov, *Equations of Mathematical Physics* (Nauka, Moscow, 1967; Marcel Dekker, New York, 1971).
31. N. H. March, W. H. Young, and S. Sampanthar, *The Many-Body Problem in Quantum Mechanics* (Cambridge University Press, Cambridge, 1967; Mir, Moscow, 1969).
32. A. Ya. Shik, Fiz. Tekh. Poluprovodn. (St. Petersburg) **29**, 1345 (1995) [Semiconductors **29**, 697 (1995)].
33. Yu. E. Lozovik and V. I. Yudson, Fiz. Tverd. Tela (Leningrad) **17**, 1613 (1975) [Sov. Phys. Solid State **17**, 1054 (1975)].
34. M. Chiao, R. W. Hill, C. Lupien, *et al.*, Phys. Rev. B **62**, 3554 (2000).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 1053–1074.
Translated from Zhurnal Éksperimental’noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 1172–1196.
Original Russian Text Copyright © 2003 by Molotkov.

A New Approach to Unconditional Security in Relativistic Quantum Cryptography

S. N. Molotkov[a,b]
[a] *Institute of Solid State Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
[b] *Department of Computational Mathematics and Cybernetics, Moscow State University, Moscow, 119899 Russia*
e-mail: molotkov@issp.ac.ru

Received February 20, 2003

Abstract—A principally new approach ensuring secure key distribution via an open quantum communication channel is proposed. In contrast to the existing schemes, in which the security is based upon special properties of nonorthogonal states in the Hilbert space, the security of the proposed scheme relies on a spacetime structure of states and on certain constraints imposed by special relativity. Using these factors, it is possible to provide for secure key transmission using practically arbitrary quantum states. *© 2003 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

The main task in cryptography is to provide for the security of information exchange between two or several legitimate parties. Secret information is transmitted via open (insecure) communication channels after being encoded in various systems. Absolutely stable (secure, unintelligible to a third party) are the “one-time pad” systems, frequently called Vernam schemes [1]. The claim of absolute stability (security) of the one-time pad encrypting scheme seems to be the only strictly proven result in cryptography. Originally, the conditions of absolute stability were rationalized by Kotel’nikov in 1941. In 1944, an analogous statement was proved by Shannon (published in 1949 [2]).

In formal terms, a cryptosystem is absolutely secure provided that the information shared by legitimate parties obeys the condition

$$I(M;\, C) = H(C) - H(C|M) = 0, \qquad (1)$$

where M is a bit sequence from an ensemble of messages to be transmitted and C is the bit sequence describing the encoded message from M. Equation (1) directly indicates that a conditional probability of the appearance of a particular encoded text c, provided that message m was selected, is

$$p(c|m) = p(c).$$

This implies that the cryptosystem is absolutely secret provided that the probability of appearance of the encoded message c is independent of the message m.

As noted above, an example of an absolutely secure cryptosystem is offered by the Vernam scheme [1] with one-time pad keys k (k is a bit sequence known only to legitimate parties). Messages from the ensemble M are compressed to m-bit sequences with a length of $H(M)$, after which a random key k is selected with the same length $H(M)$ and the encoded text is obtained in the form of a binary sequence

$$c = m \oplus k.$$

The condition of unambiguous decoding is

$$I(M;\, C|K) = I(M). \qquad (2)$$

The encoded message is decoded by the receiving party via the inverse operation

$$m = c \oplus k = (m \oplus k) \oplus k.$$

This implies that, provided the key k is random, is known only to the legitimate parties, has a length equal to that of the message, and is used only once, the encryption scheme is absolutely secure (unintelligible). The main difficulty encountered in the implementation of this scheme consists in distributing a secrete bit sequence (key) between spatially distant legitimate parties.

The above considerations are purely mathematical and bear no relation to the laws of physics. However, the problem of key distribution between spatially distant legitimate parties cannot be solved without exercising recourse to the physical reality, since the very notions of space and time are physical. Although the information is a mathematical entity, the information carriers are always particular physical objects.

In classical physics, both nonrelativistic and relativistic, the states of physical objects are described in terms of real functions of coordinates and time. It is postulated (and confirmed in experiment) that the state of a physical object can be, in principle, measured to any desired accuracy without perturbing the system state. Therefore, secure key distribution via an insecure

1063-7761/03/9705-1053$24.00 © 2003 MAIK “Nauka/Interperiodica”

communication channel cannot in principle be ensured within the framework of classical physics, since it is impossible to provide for guaranteed detection of an attempt at passive eavesdropping in the course of the information transfer between legitimate parties. For this reason, one-time pad cryptosystems are not widely used in practice.

In view of the impossibility of ensuring secure key distribution by classical signals via an insecure communication channel, cryptosystems with open keys have found rather wide application. This was rendered possible by a remarkable discovery of Diffi and Hellman [3]. Use of the open-key cryptosystems, also referred to as Rivest–Shamir–Adleman (RSA type) systems [4], does not require a common secret key to be preliminarily distributed. The secrecy of open-key cryptography relies on the unproven complexity of calculating the inverse of certain functions with secrets (for example, a discrete logarithm). The known classical algorithms performed using a computational device operating according to the laws of classical physics are characterized by exponentially growing complexity with respect to the volume of input data. At the same time, no polynomial algorithms are known (nor has it been proved that such algorithms cannot exist). Strictly speaking, RSA type systems are not absolutely secure since Shor [5] suggested a quantum algorithm of polynomial complexity for the inversion of a discrete logarithm. This implies that, in principle, any RSA type system is intelligible because no principal prohibitions are known, at least presently, for the creation of a quantum computer. On the other hand, considerable (albeit purely technical) difficulties encountered in realization of the quantum computer can delay the appearance of such devices by several decades. At the same time, another danger still exists for RSA type systems, since there are no strict guarantees that a classical algorithm of polynomial complexity cannot be developed. Thus, RSA type systems are apparently secure in classical physics, but cease to be such in the quantum domain [5].

The task of ensuring a secure key distribution via an insecure communication channel can be solved by means of quantum cryptography. In contrast to the situation with a quantum computer, where only the operation of separate gates for a small amount of qubits have been demonstrated so far, achievements in the practical implementation of quantum cryptography are much more impressive. There are functioning prototypes covering distances on the order of several dozens of kilometers (the present-day record length is 67 km [6]) both via fiber optic lines (e.g., a 23-km communication line under Lake Geneva [7]) and via open space.

The approach to ensuring unconditional security proposed in this paper may prove to be more convenient in practice, since it employs the frequency properties of states that are more stable than polarization degrees of freedom with respect to external perturbations. In addition, most realized cryptosystems represent one or another modification of an interferometer of the Mach–Zehnder type. The proposed scheme does not employ interferometers.

2. NONRELATIVISTIC QUANTUM CRYPTOGRAPHY

The laws of quantum mechanics are more restrictive than those of classical physics in that, generally speaking, any measurement (observation) of a quantum system changes the state of this system. The quantum mechanics offers possibilities of realizing unconditionally secure key distribution between spatially distant legitimate parties, thus providing for the creation of an absolutely secure one-time pad cryptosystem. By unconditional security is implied the security guaranteed by the laws of quantum mechanics representing the basic laws of nature.

The idea of quantum cryptography, originally formulated by Wiesner [8], has become commonly known in the paper of Bennett and Brassard [9]. The unconditional security of nonrelativistic quantum cryptography is based on two closely related prohibitions dictated by the postulates of nonrelativistic quantum mechanics: (i) an unknown quantum state cannot be copied (no-cloning theorem) [10]; (ii) no information can be obtained about nonorthogonal quantum states without perturbing these states [11].

Let us reproduce an elegant proof of the second statement, as suggested by Bennett [11], since this principle will be used in the following analysis. Consider two pure states of a quantum system, described by vectors (rays) in the Hilbert space $|\varphi_0\rangle, |\varphi_1\rangle \in \mathcal{H}$, which are assigned to the classical bits 0 and 1, respectively. One of the states is presented unidentified to an observer. His task is to recognize the state, leaving it, if possible, unperturbed.

The proof is based on reduction ad absurdum. It is assumed that the observer (or eavesdropper)[1] can perform general manipulations over the state according to the laws of quantum mechanics. In the general case, these manipulations reduce to the following. The eavesdropper possesses an auxiliary quantum system occurring in a certain standard quantum state $|a\rangle \in \mathcal{H}_a$. Switching on the interaction between the presented state ($|\varphi_0\rangle$ or $|\varphi_1\rangle$) and the auxiliary quantum system, the eavesdropper allows them to evolve jointly for some time, which is formally described as follows:

$$|\varphi_0\rangle \longrightarrow U(|\varphi_0\rangle \otimes |a\rangle), \tag{3}$$

$$|\varphi_1\rangle \longrightarrow U(|\varphi_1\rangle \otimes |a\rangle), \tag{4}$$

where U is the unitary operator functioning in $\mathcal{H} \otimes \mathcal{H}_a$. Assume that the system upon interaction and joint unitary evolution remains in the initial state, while the aux-

---

[1] In many papers on the subject, the legitimate parties are referred to as A (Alice) and B (Bob) and the eavesdropper is called Eve.

iliary quantum system passes to a new state depending on the unknown presented state as

$$U(|\varphi_0\rangle \otimes |a\rangle) = |\varphi_0\rangle \otimes |a_0\rangle, \quad (5)$$

$$U(|\varphi_1\rangle \otimes |a\rangle) = |\varphi_1\rangle \otimes |a_1\rangle. \quad (6)$$

Complex conjugation of one of these equations, for example, the latter, yields

$$(\langle a| \otimes \langle \varphi_1|)U^{-1} = \langle a_1| \otimes \langle \varphi_1|. \quad (7)$$

Taking a scalar product of Eqs. (5) and (7), we obtain

$$\langle a| \otimes \langle \varphi_1|U^{-1}U|\varphi_0\rangle \otimes |a\rangle = \langle \varphi_1|\varphi_0\rangle\langle a_1|a_0\rangle \quad (8)$$

and by virtue of the unitarity of U,

$$\langle a|a\rangle\langle \varphi_1|\varphi_0\rangle = \langle a_1|a_0\rangle\langle \varphi_1|\varphi_0\rangle. \quad (9)$$

The pure states being normalized vectors, we have

$$\langle a|a\rangle = 1.$$

There are only two possibilities, according to which the states are (i) nonorthogonal, whereby $\langle \varphi_1|\varphi_0\rangle \neq 0$, or (ii) orthogonal, whereby $\langle \varphi_1|\varphi_0\rangle = 0$. In the former case, both parts of Eq. (9) can be divided by the factor $\langle \varphi_1|\varphi_0\rangle \neq 0$ to obtain

$$\langle a_1|a_0\rangle = 1. \quad (10)$$

Since the pure states are edge points of a convex set of states, the latter condition implies that $|a_1\rangle = |a_0\rangle$.

Thus, the initial assumption that the message state remains unperturbed, while the state of the auxiliary quantum system changes depending on the input state, is incorrect. In other words, it is impossible to obtain information about nonorthogonal states without perturbing these states. Any observation (measurement) of such states introduces a perturbation. The perturbation changes transition probabilities in the channel, thus allowing any attempts of eavesdropping to be detected. Various theoretical variants and practical realizations of quantum cryptosystems unavoidably employ the idea of nonorthogonality.

A basically different situation takes place when the states are orthogonal, so that $\langle \varphi_1|\varphi_0\rangle = 0$. In this case, it is impossible to divide both parts of Eq. (9) by a zero factor and there are no formal limitations on the possibility of extracting information from the orthogonal states without perturbing these states. Strictly speaking, the above theorem poses no prohibitions in this case. The fact that information concerning orthogonal states can be obtained in a nonperturbing manner is readily demonstrated by explicitly presenting a measurement that leaves the states unperturbed while gaining information. Indeed, a measurement exists that provides information (reliable with a probability of unity) about orthogonal states without introducing perturbations. Like any measurement, this operation can be described by an orthogonal of nonorthogonal subordinate partition of unity in $\mathcal{H}$.

In the given case, the measurement is described by an orthogonal subordinate partition of unity in a subspace of $\mathcal{H}$ generated by vectors $|\varphi_{0,1}\rangle$, for which

$$I\{\Omega\} = \mathcal{P}_0 + \mathcal{P}_1, \quad \mathcal{P}_{0,1} = |\varphi_{0,1}\rangle\langle\varphi_{0,1}|. \quad (11)$$

The corresponding space of results is the discrete set $\Omega = \{0, 1\}$. The instrument (also called a superoperator) describing the state of the quantum system upon obtaining a particular measurement outcome, has the following form:

$$\mathcal{T}[\ldots] = \mathcal{P}_0[\ldots]\mathcal{P}_0 + \mathcal{P}_1[\ldots]\mathcal{P}_1. \quad (12)$$

The probability of obtaining mutually excluding outcomes for any pair of states is

$$\Pr\{i\} = \mathrm{Tr}\{\mathcal{T}[|\varphi_i\rangle\langle\varphi_i|]\} = 1, \quad i = 0, 1, \quad (13)$$

and a change in the state upon obtaining the ith outcome is

$$|\varphi_i\rangle\langle\varphi_i| = \frac{\mathcal{T}[|\varphi_i\rangle\langle\varphi_i|]}{\Pr\{i\}}. \quad (14)$$

Thus, it is possible to gain reliable information about orthogonal states without perturbing these states. For this reason, using orthogonal states for quantum cryptography purposes in the nonrelativistic case is not worthy of consideration, since an eavesdropper can obtain this information without altering states in the communication channel.

There is one circumstance of principal significance for the following analysis. The orthogonal states can be reliably distinguished without perturbation, provided that such states are entirely accessible for measurements (i.e., as whole objects). This fact is implicitly used in the above proof. Here, the condition of being “accessible as whole objects” means that the entire Hilbert space of states is accessible in which the state carrier is nonzero.

This can be illustrated by an example. Consider the states $|\varphi_{0,1}\rangle \in \mathcal{H}$, and let $\{|e_k\rangle\}$ be an orthonormal basis set in $\mathcal{H}$, so that

$$\begin{aligned}|\varphi_{0,1}\rangle &= \sum_{k<n} a_k^{(0,1)}|e_k\rangle + \sum_{k\geq n} a_k^{(0,1)}|e_k\rangle \\ &= |\tilde{\varphi}_{0,1}\rangle + |\tilde{\varphi}_{0,1}^{\perp}\rangle.\end{aligned} \quad (15)$$

For brevity, we assume that

$$\{|e_k\rangle \in \mathcal{H}_1\}, \quad k < n,$$

$$\{|e_k\rangle \in \mathcal{H}_1^{\perp}\}, \quad k \geq n,$$

$$\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_1^{\perp}.$$

Let the states be inaccessible as a whole, which formally implies that any measurement has to be described by a subordinate partition of unity (orthogonal or nonorthogonal) only in the accessible part of the Hilbert space of states. For example, let $\mathcal{H}_1$ be accessible for the measurements, so that the subordinate partition of unity may contain in the expansion only basis vectors from this subspace. In this case, even orthogonal states cease to be reliably distinguishable. Moreover, reliably indistinguishable without perturbation are the orthogonal states with limitations imposed on the subspace $\mathcal{H}_1$ accessible for measurements ($\langle \tilde{\varphi}_1 | \tilde{\varphi}_0 \rangle = 0$). This statement follows essentially from the normalization of the quantum state.

At first glance, the situation whereby only a part of the space of states is accessible for measurements may seem artificial. In fact, this situation is rather typical. Since quantum objects do not exist beyond spacetime, all the results of measurements (however complicated the space of these results) will unavoidably contain the space and time domains. Therefore, the coordinate representation of an abstract Hilbert space of states, albeit mathematically equivalent to any other representation, is in a certain sense a special one. A situation in which the whole space of states is inaccessible is quite readily realized. To this end, it is sufficient to limit the access to a part of the coordinate space in which the amplitude of the state (wave function) is nonzero. In the coordinate representation, the basis vectors in $\mathcal{H}$ are related to the spatial coordinates and time:

$$|e_k\rangle \longrightarrow |x, t\rangle.$$

In application to problems of quantum cryptography, the situation of limited access to the whole quantum state can be realized by using the states (wave functions) in which the spatial length exceeds the channel length [12, 13]. In this case, the whole propagating state is never present in the channel [12, 13]. However, this requirement is not necessary for ensuring security. As will be shown below, security can be provided using virtually any quantum state.

Unfortunately, nonrelativistic quantum cryptography, posing no restrictions on the maximum propagation velocity, cannot realize a secure protocol of key distribution based on the idea of limited access to the quantum state.

The fact that even orthogonal quantum states are reliably indistinguishable under the conditions of limited access to these states has a general character, being independent of a particular representation of $\mathcal{H}$. Indeed, any measurement can be described using a subordinate partition of unity,

$$I = I_1 + I_1^{\perp}, \tag{16}$$

where I_1 and $I_1^{\perp}$ are units in subspaces $\mathcal{H}_1$ and $\mathcal{H}_1^{\perp}$. With a limited access to the states, the measurements with uncertain outcomes (?) will unavoidably take place. The probabilities of such outcomes are

$$\Pr\{?|i\} = \mathrm{Tr}\{|\varphi_i\rangle\langle\varphi_i| I_1^{\perp}\} \neq 0. \tag{17}$$

For example, with spatially limited access to a part of the amplitude of the state, such outcomes can be interpreted as taking place in the region of the coordinate space inaccessible to the observer. While the total probability of outcomes in the entire space is unity, the observer can extract information about the state only from the part of space accessible for his measurements. If the measuring device did not operate in the accessible part of the space (formally speaking, this implies that the event occurs in the inaccessible part of the space), the observer can assign this uncertain event (?) to any outcome (0 or 1) with a probability of 1/2. In other words, the minimum space of results from which the information about states (0 or 1) can be obtained will unavoidably include uncertain outcomes, $\Omega = \{0, 1, ?\}$, which hinders the obtaining of reliable information even on orthogonal states.

3. RELATIVISTIC QUANTUM CRYPTOGRAPHY

As stated above, secure key distribution via an open channel is impossible in classical physics. The laws of quantum mechanics, being more restrictive than those of classical physics, allow secure key distribution via an open channel to be realized. Here, security is guaranteed by detecting any attempts at eavesdropping. The laws of relativistic quantum field theory are even more restrictive than those of nonrelativistic quantum mechanics. Constraints imposed on the measurability of quantum states in the relativistic domain were originally discussed by Landau and Peierls as long ago as 1931 [14]. Further investigations were undertaken by Bohr and Rosenfeld [15].

The aforementioned theorem [11] concerning the impossibility of obtaining information about nonorthogonal states in a nonperturbing way was proved using only the geometric properties of the Hilbert space of states, without any recourse to the concept of spacetime and the fact that all events (including measurements on the states transmitted via a communication channel) occur in spacetime. Nor was any particular nature of the quantum system specified. However, abstract quantum systems outside spacetime do not exist in nature. Moreover, the very fact of the existence of spacetime results in that only certain elementary quantum systems (particles) do exist, which implies that the basis vectors of unitary irreducible representations of an inhomogeneous Lorentz group in the Hilbert space are interpreted as the vectors of state (wave functions) of elementary particles (electrons, positrons, neutrinos, photons, etc.). Although the states of quantum systems in the relativistic case are described by vectors in the Hilbert space $\mathcal{H}$, the vectors are characterized by an internal “content” in the form of smoothing functions (amplitudes) dependent on the space coordinates and time [16, 17].

It should be noted that, in any real situation, the only acceptable carriers for the transmission of information over large distances in cryptosystems are photons—massless particles (states of a quantum field) propagating with the maximum possible velocity (i.e., with the velocity of light). The existence of this limiting velocity is the basic law of nature. Also, any transmission of information in spacetime with both classical and quantum objects implies a causal link between preparation of an information carrier, distribution via a channel, and measurements on this carrier.

The fact that information carriers are the states of a massless quantum field, together with the relativistic causality principle, allows a new unconditionally secure cryptosystem to be constructed in which security relies on these basic principles rather than on the Hilbert space geometry (i.e., on the special properties of nonorthogonal states in this space). In the relativistic case, secure key distribution is possible using virtually any states of the quantized photon field (even orthogonal states with an effective length below that of the communication channel). Any attempt at eavesdropping is detected by a delay in the results of measurements. In the nonrelativistic case, eavesdropper interference is detected by a change in the vector of state for one of the nonorthogonal states.

The delay found in the relativistic case also represents, in a certain sense, a change in the state, since a shift (translation) in spacetime means a transition to another state. This is essentially a new point that will probably simplify experimental implementation of the proposed approach. For example, using orthogonal states with nonoverlapping frequencies as carriers is much more convenient than encoding into polarization states, because a change in the frequency of a state is more stable with respect to noise than the polarization state. In addition, a change in the state related to a delay can be experimentally measured by usual photodetectors.

Let us consider a quantized electromagnetic field (photons). The electromagnetic field operators can be written (assuming that $c = \hbar = 1$) as [16]

$$A^{\pm}_{\mu}(\hat{x}) = \frac{1}{(2\pi)^{3/2}} \times \int \frac{d\mathbf{k}}{\sqrt{2k_0}} \exp(\pm i\hat{k}\hat{x}) e^{m}_{\mu}(\mathbf{k}) a^{\pm}_{m}(\mathbf{k}), \tag{18}$$

and they obey the commutation relations

$$[A^{-}_{\mu}(\hat{x}), A^{+}_{\nu}(\hat{x}')]_{-} = i g_{\mu\nu} D^{-}_{0}(\hat{x}-\hat{x}'). \tag{19}$$

Here, $D^{-}_{0}(\hat{x}-\hat{x}')$ is the commutation function for the zero-mass field,

$$\begin{aligned} D^{\pm}_{0}(\hat{x}) &= \pm\frac{1}{i(2\pi)^3}\int \frac{d\mathbf{p}}{2p_0}\exp(\pm i\hat{p}\hat{x}) \\ &= \frac{1}{4\pi}\varepsilon(x_0)\delta(\hat{x}^2) \pm \frac{i}{4\pi\hat{x}^2}, \\ \varepsilon(x_0)\delta(\hat{x}^2) &\equiv \frac{\delta(x_0-|\mathbf{x}|)-\delta(x_0+|\mathbf{x}|)}{2|\mathbf{x}|} \end{aligned} \tag{20}$$

and quantities with caps denote four-dimensional vectors,

$$\hat{k} = (k_0, \mathbf{k}), \quad \hat{x} = (x_0, \mathbf{x}).$$

Formally, there are four possible types of photons: two transverse, longitudinal, and temporal, but the latter two are fictitious and can be excluded from consideration via introduction of an indefinite metrics [16, 17]. For our purposes, the simplest way consists in using a particular gauge. In what follows, we consider a subspace of physical states in the Coulomb gauge $A_\mu = (\mathbf{A}, \varphi = 0)$ and deal with two physical transverse states of the electromagnetic field. The generalized operator function represents a vector in the three-dimensional (3D) space,

$$\begin{aligned} \boldsymbol{\varphi}(\hat{x}) &= \frac{1}{(2\pi)^{3/2}}\int \frac{d\mathbf{k}}{\sqrt{2k_0}}\sum_{s=\pm 1}\mathbf{w}(\mathbf{k}, s) \\ &\times \{a(\mathbf{k}, s)\exp(-i\hat{k}\hat{x}) + a^{+}(\mathbf{k}, -s)\exp(i\hat{k}\hat{x})\}, \end{aligned} \tag{21}$$

where $\mathbf{w}(\mathbf{k}, s)$ is a 3D vector describing the photon spirality ($s = \pm 1$),

$$\begin{aligned} \mathbf{w}(\mathbf{k}, \pm) &= \frac{1}{\sqrt{2}}[\mathbf{e}_1(\mathbf{k}) \pm i\mathbf{e}_2(\mathbf{k})], \\ \mathbf{e}_1(\mathbf{k}) &\perp \mathbf{e}_2(\mathbf{k}), \quad |\mathbf{w}(\mathbf{k}, s)|^2 = 1, \end{aligned} \tag{22}$$

and $\mathbf{e}_{1,2}(\mathbf{k})$ are the vectors perpendicular to $\mathbf{k}$. The field operators obey the Maxwell equations:

$$\nabla \times \boldsymbol{\varphi}(\hat{x}) = -i\frac{\partial}{\partial t}\boldsymbol{\varphi}(\hat{x}), \quad \nabla \cdot \boldsymbol{\varphi}(\hat{x}) = 0. \tag{23}$$

Smoothened field operators can be written in the following form:

$$\boldsymbol{\varphi}(f) = \sum_{s=\pm1}\int\boldsymbol{\varphi}(\hat{x},s)f(\hat{x},s)d\hat{x} = \frac{1}{(2\pi)^{3/2}}\int\frac{d\mathbf{k}}{\sqrt{2k_0}} \tag{24}$$
$$\times\sum_{s=\pm1}\mathbf{w}(\mathbf{k},s)\{f(\mathbf{k},s)a^{+}(\mathbf{k},s)+f^{*}(\mathbf{k},s)a(\mathbf{k},s)\},$$

where $f(\mathbf{k}, s)$ are the values of $f(\hat{k}, s)$ on the mass surface, $f(\hat{k}, s)$ being a 4D Fourier image of an arbitrary function $f(\hat{x}, s)$ from the space of base functions $\Omega(\hat{x})$.

Let us consider a one-dimensional model. This approximation is physically justified, since real fiber optic systems are quasi-one-dimensional objects. The information carriers are pure states of the massless quantized electromagnetic field (photons). The generalized basis states are generated by the action of field operators (more precisely, of the generalized functions with operator values) on a cyclic vacuum vector [16, 17]. Taking into account the requirements on the Lorentz invariance, the field operators cannot be simple (even unlimited) operators in $\mathcal{H}$. If the field operators were simple operators, the matrix element $\langle 0|\hat{\varphi}^{-}(\hat{x}')\hat{\varphi}^{+}(\hat{x})|0\rangle$, interpreted as the generation of a particle in $\hat{x}$, propagation, and annihilation in $\hat{x}'$, would be merely a constant independent of $\hat{x}$ and $\hat{x}'$, in contradiction to the causality principle [18].

The generalized functions with operator values are represented as

$$\hat{\varphi}^{+}_{\mu}(\hat{x}) = \int d\hat{k}\exp(i\hat{k}\hat{x})a^{+}_{\mu}(\hat{k})\theta(k_0)\delta(\hat{k}^2), \tag{25}$$
$$\hat{k} = (k_0, k), \quad \hat{x} = (x, t),$$

where $\mu = 0, 1$ is the index of polarization (spirality). The commutation relations are as follows:

$$[a^{-}_{\mu'}(\hat{k}'), a^{+}_{\mu}(\hat{k})] = k_0\delta(k-k')\delta_{\mu,\mu'}, \tag{26}$$

and the generalized basis vectors (linear continuous functionals in $\mathcal{H}$) are

$$a^{+}_{\mu}(\hat{k})|0\rangle = |k\mu\rangle, \quad |\hat{x}\mu\rangle = \varphi_{\mu}(\hat{x})|0\rangle, \tag{27}$$
$$\langle k\mu|k'\mu'\rangle = k_0\delta_{\mu\mu'}\delta(k-k'),$$

where $|0\rangle$ is the vacuum vector and $|k\mu\rangle$, $|\hat{x}\mu\rangle\in\Omega^*$ is the space conjugate to the space of base functions Ω. The physical states (vectors in $\mathcal{H}$) are obtained by smoothening generalized operator functions with the basic functions (amplitudes) from the space $\Omega(\hat{x})$ (the space of infinitely continuously differentiable functions decreasing at infinity faster than any reciprocal polynomial power):

$$|\varphi_{\mu}\rangle = \sum_{\mu}\int d\hat{x}\tilde{\varphi}_{\mu}(\hat{x})\varphi^{+}_{\mu}(\hat{x})|0\rangle \tag{28}$$
$$= \sum_{\mu}\int\frac{dk}{k_0}\tilde{\varphi}_{\mu}(k, k_0 = |k|)|k\mu\rangle.$$

Here, the amplitude $\tilde{\varphi}_{\mu}(\hat{x})$ ($\tilde{\varphi}(k, k_0 = |k|)$) plays the role of the coefficient of expansion over generalized basis states. The construction of $\tilde{\varphi}_{\mu}(\hat{x})\in\Omega(\hat{x})$, $\tilde{\varphi}_{\mu}(\hat{x})\in\Omega^*(\hat{x})$, $|\varphi_{\mu}\rangle\in\mathcal{H}$, and $\Omega\subset\mathcal{H}\subset\Omega^*$ is called the equipped Hilbert space (or the Gelfand triad) [17, 19].

Below we will consider the field states propagating in one direction of the x axis ($k > 0$), representing the very kind of states carrying information between remote legitimate parties:

$$|\varphi_{\mu}\rangle = \int_0^{\infty}\frac{dk}{k}\tilde{\varphi}(k,k)|k\mu\rangle = \int_0^{\infty}dk\varphi(k)|k\mu\rangle \tag{29}$$
$$= \int_{-\infty}^{\infty}d(x-t)\varphi(x-t)|x-t,\mu\rangle,$$

where

$$\varphi(k) = \frac{\tilde{\varphi}(k,k)}{\sqrt{k}},$$
$$\varphi(x-t) = \frac{1}{2\pi}\int_0^{\infty}\exp(-ik(x-t))\varphi(k)dk, \tag{30}$$
$$|x-t,\mu\rangle = \int_0^{\infty}\exp(ik(x-t))|k\mu\rangle dk$$

and

$$\langle\varphi_{\mu}|\varphi_{\mu}\rangle = \int_0^{\infty}\frac{dk}{k}|\tilde{\varphi}(k,k)|^2 = \int_0^{\infty}dk|\varphi(k)|^2 \tag{31}$$
$$= \int_{-\infty}^{\infty}d(x-t)|\varphi(x-t)|^2 = 1.$$

The physical states in $\mathcal{H}$ are determined by the values of amplitude on the mass surface. The amplitude of states propagating in one direction depends only on the difference $\tau = x - t$. This circumstance reflects the fact that, should a given measurement result take place at the time t in the vicinity of the point $(x, x + dx)$, the same result will be obtained at the time t' in the vicinity of the point $(x', x' - x + t + dx)$. For brevity, we will speak of the amplitude $\varphi(\tau)$ set on a branch of the light cone.

The polarization index μ will be omitted as insignificant unless it will be required. The information states will be represented by orthogonal states with nonoverlapping frequency bands. The amplitudes corresponding to the vectors of states 0 and 1 are selected in the form of

$$\operatorname{supp}\varphi_0(k) \in \{\Delta k\}_0$$

and

$$\operatorname{supp}\varphi_1(k) \in \{\Delta k\}_1,$$

respectively. The condition of nonoverlapping frequency bands is

$$\{\Delta k\}_0 \cap \{\Delta k\}_1 = \varnothing,$$

and the vectors of states are automatically orthogonal:

$$\begin{aligned}\langle\varphi_0|\varphi_1\rangle &= \int\limits_{\{\Delta k\}_0 \cap \{\Delta k\}_1} dk\varphi_0^*(k)\varphi_1(k)\\ &= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} d\tau d\tau'\varphi_0^*(\tau)\varphi_1(\tau')\left[\frac{1}{2}\delta(\tau-\tau') + \frac{i}{\pi}\frac{1}{\tau-\tau'}\right] \qquad (32)\\ &= \int\limits_{-\infty}^{\infty} d\tau\varphi_0^*(\tau)\varphi_1(\tau) = 0.\end{aligned}$$

Here, the generalized function is written in the integral representation [20]

$$\begin{aligned}&\int\limits_{-\infty}^{\infty} dx\exp(ik(x-t))\frac{1}{x-t+a} \qquad (33)\\ &= i\pi\operatorname{sgn}(k)\exp(-ika).\end{aligned}$$

As can be seen, the orthogonality of states is a nonlocal property in spacetime. This implies that, once the time t^* in $\tau = x - t^*$ is fixed, the orthogonality "spreads" over a large—formally, an infinitely large (see below)—spatial region. If the coordinate x^* is fixed on $\tau = x^* - t$, the orthogonality is gained due to a large interval of time. In what follows, these considerations will be given operational meaning. The possibility of distinguishing the orthogonal quantum states by means of measurements will unavoidably require either a finite spatial region (at a fixed time section) or a finite time (at a fixed spatial variable).

Strictly speaking, the exact orthogonality and, accordingly, reliable outcome of the measurement would formally require the infinite spacetime region because of the basic nonlocalizability of states in quantum field theory.

In fact, since the amplitude of states is set on the mass surface (whereby the values of $\varphi(k, k_0)$ as the function of two variables are determined only for $k = k_0$, rather than for arbitrary k and k_0), this function is always nonzero in the entire space (i.e., outside any compact, which is an arbitrarily large but finite region). The fact of the nonlocalizability of states in quantum field theory has been known for a long time (see the physical consideration of this question, e.g., in [21]). In the given case, the nonlocalizability can be demonstrated to follow from the Wiener–Paley theorem [22]. Consider a normalized function $\varphi(k)$ such that

$$\int\limits_0^{\infty} dk|\varphi(k)|^2 = 1, \qquad (34)$$

which is equal to zero on the semiaxis $k \le 0$, while being not identical zero. For the Fourier transform $\varphi(\tau)$ of this function, the rate of spatial decay at infinity is determined by the convergence of the integral

$$\int\limits_{-\infty}^{\infty}\frac{\ln|\varphi(\tau)|}{1+\tau^2}d\tau < \infty. \qquad (35)$$

From this it follows that the amplitude $\varphi(\tau)$ cannot decrease even exponentially (not speaking of being zero outside a compact), since otherwise (for $\varphi(\tau)| \propto \exp(-\alpha|\tau|)$) the integral in (35) would diverge. However, the amplitude decay can be arbitrarily close to exponent, with any $\alpha > 0$, so that

$$|\varphi(\tau)| \propto \exp\left(-\frac{\alpha|\tau|}{\ln(\ln\ldots|\tau|)}\right).$$

This degree of localization of the photon field can be also achieved in the 3D case [23], although it was considered for a long time (after the Newton and Wigner paper [24]) that the most rapid possible spatial decay corresponds to a power of 7/2.

The nonlocalizability of the amplitude (differing from zero outside any compact) has deep roots related to causality in the relativistic quantum domain. In particular, Hegerfeld [25] demonstrated that, if the amplitude of a state were localized in a certain finite region of space at the initial time t_0, the free evolution would make the amplitude nonzero in regions arbitrarily far (separated by a spacelike interval) from this finite region at any subsequent time $t > t_0$. This result is inconsistent with relativistic causality, since this behavior implies that information can be transmitted at a rate exceeding the velocity of light (even when the probability of the measurement's outcome is less than unity in the region separated by a spacelike interval from the initial region). Recently [26], it was demonstrated that the principle of causality in decay processes is valid in the relativistic domain for the average values of observables.

Roughly speaking, the state of a free quantized field is different from zero always and everywhere, that is, in the entire spacetime (see, e.g., (35)).

The principle of causality in the relativistic domain was originally formulated in the finite form by Bogoliubov in 1955 (see [16]). In the differential form, it appears as

$$\frac{\delta S^{+}(g(\hat{x}))}{\delta g(\hat{x})} S(g(\hat{y})) = 0, \quad \hat{x} < \hat{y}. \tag{36}$$

This means that, should the state be perturbed in some region where the interaction $g(\hat{x})$ takes place, this cannot influence the results of measurements in regions $\hat{y}$ separated from the former region by a spacetime interval.

In the following considerations, the role of the causality principle is to ensure that information about a quantum state in some spacetime region cannot be obtained until the state has reached this region. No one state can enter a given region or leave it at a velocity faster than that of light. More precisely, no finite information about the state can be obtained differing from zero by more than an exponentially small value. Note also that the smaller the fraction of a state present in the region accessible for the measurement, the lower the probability of obtaining any result in the measurement on this state. This follows directly from the normalization of the quantum state.

The fact that the amplitude of a state differs from zero in the entire space is not a restrictive condition in the problems of quantum cryptography, since it is always possible to select a current spacetime region of dimensions such that the fraction of the amplitude normalization gained outside this region would be exponentially close to zero to within any preset accuracy. Although the very fact of strict nonlocalizability of the amplitude is related to the basic requirements of causality, the manifestations of this nonlocalizability in particular problems are rather of a technical character. Now we will impart more operational meaning to these considerations.

Let the amplitude of a state in the momentum representation be set in a finite but quite arbitrary frequency band $\Delta k \in (0, \infty)$. Let us consider the question as to how the probability of finding a photon inside a spacetime region of size $2T$ (at a fixed frequency band) depends on the shape of the amplitude of this state. These considerations will be required below for the formulation of a secure bit commitment protocol, because the amplitude shape will determine the requirements on the size of a spacetime region necessary for detecting attempts at eavesdropping by monitoring delays in the results of measurements.

Any measurement on a quantum state is described by a subordinate partition of unity in the single-particle subspace of states,

$$I = \int_{0}^{\infty} \frac{dk}{k} |k\rangle\langle k| = \int_{-\infty}^{\infty} \mathcal{M}(d\tau), \tag{37}$$

and the operator value measure

$$\begin{aligned} \mathcal{M}(d\tau) = \frac{d\tau}{2\pi}\left(\int_{0}^{\infty} \frac{dk}{\sqrt{k}} \exp(-ik\tau)|k\rangle\right) \\ \times\left(\int_{0}^{\infty} \frac{dk'}{\sqrt{k'}} \exp(ik'\tau)\langle k'|\right) \end{aligned} \tag{38}$$

describes the probability of detecting a photon within the interval $(\tau, \tau + d\tau)$. Accordingly, the probability of finding the photon in a finite spacetime region T (it should be recalled that the amplitude depends only on the difference $\tau = x - t$) is

$$\Pr(T) = \mathrm{Tr}\{\mathcal{M}(T)|\varphi\rangle\langle\varphi|\} = \int_{-T}^{T} d\tau |\varphi(\tau)|^2, \tag{39}$$

where

$$\begin{aligned} \mathcal{M}(T) = \int_{-T}^{T} \mathcal{M}(d\tau), \quad \mathcal{M}(\bar{T}) = \int_{\bar{T}} \mathcal{M}(d\tau), \\ \bar{T} = (-\infty, \infty)\backslash(-T, T). \end{aligned} \tag{40}$$

Accordingly, the probability of finding the photon outside the $(-T, T)$ region (i.e., in the rest of spacetime) is

$$\Pr(\bar{T}) = \mathrm{Tr}\{\mathcal{M}(\bar{T})|\varphi\rangle\langle\varphi|\} = 1 - \int_{T} d\tau |\varphi(\tau)|^2. \tag{41}$$

Since the amplitude depends only on the difference $\tau = x - t$, these relations can be interpreted as follows. First, as the probability of finding a photon in a coordinate space region of size $\Delta x = 2T$ at a certain fixed moment of time. Second, as the probability of finding a photon in the vicinity of a fixed point x within a time interval of $\Delta t = 2T$ (the meaning of this will be refined below).

The maximum possible probability of finding a single-photon packet within the interval T is given by the expression

$$\Pr_{\max}(T) = \max_{\{\|\varphi\|^2 = 1\}} \mathrm{Tr}\{\mathcal{M}(T)|\varphi\rangle\langle\varphi|\}, \tag{42}$$

where the maximum is taken over all packet shapes $\varphi(k)$ whose carriers fall within the frequency band Δk.

The optimum shape is determined by solving a variational problem for the maximum of the functional

$$\frac{\delta}{\delta\varphi}\left\{\frac{\mathrm{Tr}\{\mathcal{M}(T)|\varphi\rangle\langle\varphi|\}}{\int\limits_{\Delta k}|\varphi(k)|^2 dk}\right\} = 0, \tag{43}$$

which reduces to the eigenvalue problem:

$$\lambda_n\varphi_n(k) = \frac{1}{\pi}\int\limits_{\Delta k}\frac{\sin(k-k')T}{k-k'}\varphi_n(k')dk'. \tag{44}$$

The maximum eigenvalue gives the maximum of the functional, while the eigenfunction corresponding to this eigenvalue describes the optimum shape of the state. This equation has been previously studied in [27, 28], where it was established that the eigenvalues λ_n are positive and form a sequence decreasing with increasing number n ($1 > \lambda_0 > \lambda_1 \ldots > 0$, $n = 0, 1, \ldots \infty$). The eigenvalues depend on the parameter ΔkT, and several first members of this sequence were numerically calculated for various values of ΔkT (at large values of this parameter, the λ_n values rapidly tend to unity; e.g., for $\Delta kT = 4$, $\lambda_0 = 0.99589$) [27]. The asymptotic behavior of eigenvalues with a fixed number n at $\Delta kT \gg 1$ determined by Fuchs [28],

$$\lambda_n(\zeta) \sim 1 - \frac{4\sqrt{\pi}8^n}{n!}\zeta^{n+1/2}e^{-2\zeta}, \quad \zeta = \Delta kT, \tag{45}$$

confirms that the eigenvalues are exponentially close to unity.

Thus, the probability of any measurement in a spatial region (or a time interval) of size $2T$ for a state with the carrier within a frequency band Δk cannot exceed

$$\mathrm{Pr}_{\max}(T) \le 1 - O(\exp(-2\Delta kT)). \tag{46}$$

Therefore, by selecting the size of the spatial region (or the time interval), it is possible to ensure that the probability of detection outside this region (interval) will be arbitrarily small. For example, a probability of 10^{-80} can be considered as zero because this value is below any physically reasonable quantity (the number of atoms in a visible part of the Universe is estimated at 10^{80}). For this reason, the fact of the nonlocalizability of states in quantum field theory is not a restrictive condition: it is always possible to select the size of a region (or the time window) “accommodating” the entire state with exponential accuracy.

In application to problems of quantum cryptography, selection of the necessary region (interval) size will guarantee that an eavesdropper would be capable of gaining only exponentially small information about a transmitted state outside this spatial region (or the time interval). This interval size is on the order of $2T$. Accordingly, any attempts at eavesdropping will be detected by a receiving legitimate party by monitoring the events outside this interval (for details, see below).

Consider a pair of single-photon states (packets), each with a frequency band Δk, and let these frequency bands not overlap, so that the states are orthogonal and, hence, reliably distinguishable. However, the very fact of being reliably distinguishable implies that the whole space of states should be accessible: the measurements distinguishing the given pair of states require access to the space region in which the state is present (otherwise there is nothing to measure).

Now it will be shown that, when one of the orthogonal states ($|\varphi_0\rangle$ or $|\varphi_1\rangle$, with nonoverlapping frequency bands Δk) is transmitted via a communication channel, the obtaining of any information about this state with a probability exceeding an exponentially small value $O(\exp(-2\Delta kT))$ will lead to an unavoidable delay in the results of measurements performed by the legitimate receiving party. By this delay, any attempts at eavesdropping are detected.

Any information transfer by means of both classical and quantum objects implies three stages related by causal links: (i) preparation of an information carrier (quantum state); (ii) transmission of this carrier via a communication channel; and (iii) measurement by the legitimate receiver for gaining the information.

Let the legitimate sender A (Alice) prepare one of the states ($|\varphi_0\rangle$ or $|\varphi_1\rangle$) within a controlled spacetime region of size $2T$ at the initial time $t = 0$ of the bit commitment protocol, this moment being known to everybody including an eavesdropper (Eve). In this spacetime region, the states can be prepared so as to provide for a positive outcome with a probability of $1 - O(\exp(-2\Delta kT))$ (an unsuccessful outcome caused by nonpreparation of the state is of no danger for the key-distribution security). Here, there are two formally equivalent possibilities. In the first case, Alice prepares the state by means of a space-distributed device at $t = 0$ in the entire spatial region of size $\Delta x = 2T$ (Fig. 1a).[2] In the case of using such a distributed device, Alice takes a decision at $t = -T$ on which of the two states, 0 or 1, will be prepared. The state appears at $t = 0$ in the entire spatial region of $x < 0$ ($x > 0$ is the region of the communication channel). In other words, the preparation of a state by means of the distributed device implies that, for causality reasons (Fig. 1a), the spatial region $2T$ has to be covered by the front part of a light cone with the apex at Alice.

Although there are no formal prohibitions for preparing the amplitude of the state at the time $t = 0$ in the entire spatial region of size $2T$, another procedure for

---

[2] An example of the distributed device is given in Section 6. This device comprises a prism and apertures spaced by $2T$ from the prism. The term “distributed” means that the whole device occupies a spatial region of size $2T$.

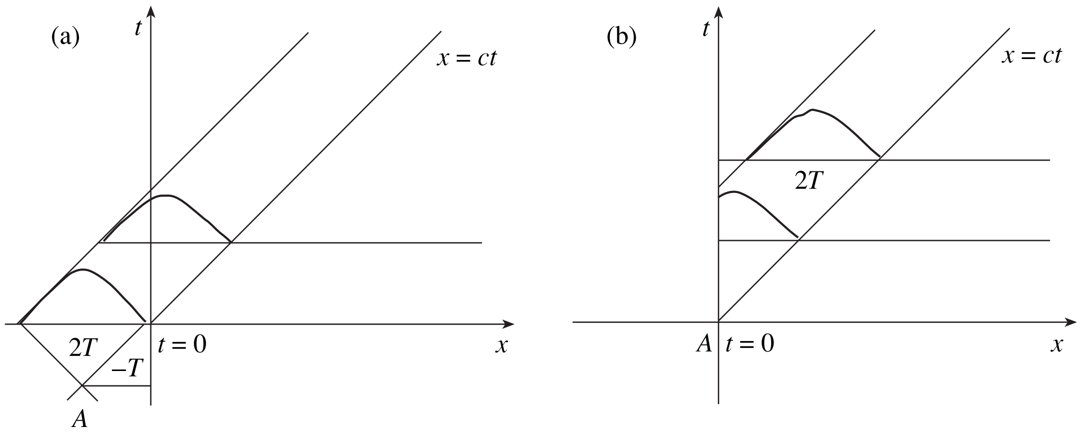


Fig. 1. Schematic diagrams illustrating the preparation of states using (a) distributed and (b) localized devices.

preparation of the state—more convenient in real situations and, naturally, equivalent to the above one—can be performed using a localized source. Here, the term “localized” means that the source size is small, $\delta x_{\text{source}} \ll T$ (Fig. 1b). A state prepared by the localized source immediately enters the communication channel.

Now let us proceed to the question of information extraction by Eve in the communication channel. Upon elapse of the time $2T$, the entire state occurs in the communication channel with a probability exponentially close unity. The evolution of the state transmitted in the channel is described by a unitary operator of translation,

$$\begin{aligned} U(\tau_0) &= \int_0^{\infty} \frac{dk}{k} \exp(ik\tau_0)|k\rangle\langle k| \\ &= \int_{-\infty}^{\infty} |\tau+\tau_0\rangle\langle\tau| d\tau, \end{aligned} \tag{47}$$

where τ_0 is the translation magnitude in spacetime. At the time $t > T$, the state can be presented as

$$\begin{aligned} |\varphi_i\rangle &= \int_{\Delta k} dk \exp(-ik(x-t-T))\varphi_i(k)|k\rangle \\ &= \int_{-\infty}^{\infty} dx \varphi_i(x-t-T)|x-t\rangle. \end{aligned} \tag{48}$$

Now it will be demonstrated that the extraction of any information, not exponentially small, by Eve will lead to a detectable delay at the legitimate receiver end of the channel. In order to gain any information on the state in transit, it is necessary to have access to the region of space in which the amplitude of this state is nonzero. Let us make recourse to the aforementioned Bennett proof [11]. According to Eqs. (3)–(9), distinguishing the orthogonal states without perturbing them is not prohibited. However, this statement is implicitly based upon the assumption that the states are accessible as a whole. The whole state cannot be accessed faster than within a time period of $2T$ (Fig. 2). Obviously, the state cannot enter the region $x \geq x_E$ (Fig. 2) more rapidly than within a propagation time of $2T$. This fact is determined only by the existence of a limiting propagation velocity, whereas the fact that the state is quantum has not yet been employed. It should be recalled that Eve knew the time of the state preparation (i.e., the time when the whole state, to within exponential accuracy, will occur in the communication channel).

The quantum nature of the state is important for the following. The probability of obtaining any result of the measurement in a certain spacetime region cannot be greater than the fraction of the normalization of this state gained in the given region. Obtaining any result of the measurement (irrespective of whether the states are

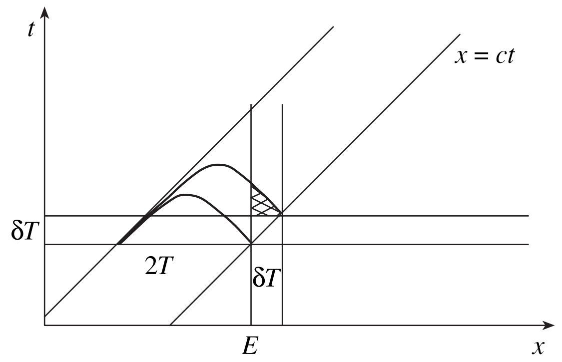


Fig. 2. Schematic diagram illustrating the time delay caused by eavesdropping: the cross-hatched part of a state allows information to be extracted due to a fraction of the state normalization gained in this region. However, this interaction unavoidably introduces a delay by δT, since an unaltered state would propagate forward by δT.

orthogonal or nonorthogonal) requires a finite time (Fig. 2). “Collecting” the state in a certain local region will unavoidably lead to a delay, since otherwise the initial state would propagate further (Fig. 2). The more information gained by Eve about the state, the greater the fraction of the normalization “collected,” and the longer the delay produced. Since the amplitude depends only on the difference $x - t$, the situation can be considered either by fixing the time and treating the coordinate as a variable or vice versa. Let us consider both variants.

Let a state be set with the amplitude $\varphi(x - t)$ (it should be recalled that this is the amplitude of a massless particle propagating at the velocity of light, $c = 1$). The state occupies a region of size $2T$, which implies that

$$\int_T |\varphi(x-t_0)|^2 dx \approx 1,$$

where $\varphi(x-t_0)$ is amplitude in the time section $t = t_0$. In order to obtain the result of any measurement with a probability arbitrarily close to unity, the values of the state amplitude at t_0 for all x are necessary at once in the region where the amplitude is nonzero, since the quantum state is normalized. This information can be obtained by performing a unitary transformation of the entire state, otherwise the transformation will not be unitary. The unitary transformation of the amplitude can be written as

$$U\varphi(x-t_0) = \tilde{\varphi}(x'-t), \quad t > t_0.$$

The amplitude $\tilde{\varphi}(x' - t)$ can be nonzero in a smaller spatial region. The minimal size of the region in x' by the time t is determined by the relativistic causality principle. The matrix elements of the unitary operator are nonzero only when the points (x, t_0) and (x', t) occur in the past portion of the light cone originating at point Γ and covering the region where the state amplitude is nonzero at time t_0. For a time not less than T, the amplitude of the initial state can be transformed in a unitary manner into the amplitude of a state localized arbitrarily strongly in the vicinity of Γ, but this will be a different state. By the time moment of Γ, the state amplitudes are accessible for all x. Now, the measurement outcome and, hence, complete (with a probability of unity) information about the state can be instantaneously obtained. If the initial states are orthogonal, the unitary transformation provides a pair of orthogonal states at Γ, which can be reliably distinguished from each other. It should be emphasized that these are essentially new orthogonal states. “Restoration” or copying (cloning) of the state can also be provided with an inverse unitary transformation “directed” forward in time. A state with the same amplitude shape cannot be obtained earlier than is dictated by the relativistic causality principle, and the amplitude of this state occurs in the front part of the light cone originating at Γ. The new state is also different from the initial one in the sense that the former is delayed in time relative to the latter (which would propagate forward along x by the time $2T$ to cover a distance of $2T$, if it were not an attempt by Eve to obtain information about the state). Here, we are speaking about gaining information about a state in the channel with a probability of unity, whereby Eve will not pass the test for delay with the same probability. Analogous considerations are valid in the case of obtaining information with a probability below unity, but the delay will be smaller than $2T$.

Similar considerations apply to the nonrelativistic case. Ignoring the special relativity constraints, we have to omit the part that pertains to the light cone. Here, the unitary transformations can be formally performed instantaneously and it is possible even to exclude the explicit coordinate, only implicitly assuming that the whole state is accessible for the unitary transformation.

Analogous considerations can be reproduced when the state is transformed in a unitary manner to the state of an auxiliary localized system (usually atomic, as in the case of “light stopping” [40]). This unitary transformation converts the field state into a vacuum state (because of the massless nature and the impossibility of propagating at a zero velocity), and an atomic system state, into a new state. Being unitary, the transformation also requires access to all values of the amplitude of the photon packet at the point of localization of the atomic system. The access is naturally provided in the course of packet propagation at the velocity of light, on reaching the localized atomic system. This process also requires a time of $2T$ for the whole single-photon packet to reach the atomic system (if we are speaking of a result obtained with a probability of unity) and the field occurs in a different (vacuum) state. For the time $2T$, it is possible to identify the state and prepare the same state, but with an unavoidable delay of $2T$ relative to the time of free propagation of the initial packet.

The situation whereby the state penetrates through an auxiliary localized quantum system and locally interacts with this system at each moment of time (by analogy with Eqs. (3)–(9)), unavoidably leads to perturbation of the initial state. This will be illustrated by a simple example. The local interaction implies incomplete access to the Hilbert space of states. Consider the initial state

$$|\varphi\rangle = c_0|e_0\rangle + c_1|e_1\rangle, \quad \langle e_0|e_1\rangle, \tag{49}$$

where $|e_{0,1}\rangle$ are the basis set vectors. Let $|a\rangle$ be an auxiliary quantum system $|a\rangle$ and the joint unitary evolution have the form

$$\begin{aligned} |\varphi\rangle &\longrightarrow U(|\varphi\rangle \otimes |a\rangle) \\ &= c_0 U(|e_0\rangle \otimes |a\rangle) + c_1|e\rangle \otimes |a\rangle. \end{aligned} \tag{50}$$

Here, the locality of the interaction implies that the unitary transformation involves only the accessible basis vectors $|e_0\rangle$. Let the auxiliary system evolve into a new state $|a'\rangle$ such that $\langle a'|a\rangle = 0$ (this simplifying assumption is not of principal importance). The state of the joint system upon evolution is

$$\begin{aligned} |\Psi\rangle &= U(|\varphi\rangle \otimes |a\rangle) = c_0 U(|e_0\rangle \otimes |a\rangle) \\ &+ c_1|e\rangle \otimes |a\rangle = c_0|e_0\rangle \otimes |a'\rangle + c_1|e\rangle \otimes |a\rangle. \end{aligned} \tag{51}$$

The new state of the initial system is calculated as a partial trace over the degrees of freedom of the auxiliary system,

$$\begin{aligned} |\tilde{\varphi}\rangle\langle\tilde{\varphi}| &= \mathrm{Tr}_a\{|\Psi\rangle\langle\Psi|\} \\ &= |c_0|^2|e_0\rangle\langle e_0| + |c_1|^2|e_1\rangle\langle e_1|, \end{aligned} \tag{52}$$

which is by no means identical to the state prior to the local interaction with the auxiliary quantum system,

$$|\varphi\rangle\langle\varphi| = (c_0|e_0\rangle + c_1|e_1\rangle)(c_0^*\langle e_0| + c_1^*\langle e_1|). \tag{53}$$

Now let us consider in more detail the procedure of measurement at the legitimate receiver B (Bob), which allows the delay to be detected. The measurements will be, first, described formally and, then, illustrated by an example of technical realization. It turns out that such measurements are quite readily implemented using filters with a frequency pass band of Δk and a usual photodetector (for details, see below).

The measurement at the receiver end is described by a subordinate partition of unity

$$\begin{aligned} I(\Delta k) &= \int\limits_{(\Delta k)_0} \frac{dk}{k}|k\rangle\langle k| + \int\limits_{(\Delta k)_1} \frac{dk}{k}|k\rangle\langle k| \\ &= \int\limits_{-\infty}^{\infty} \mathcal{M}(d\tau) = \int\limits_{-\infty}^{\infty} \mathcal{M}_0(d\tau) + \int\limits_{-\infty}^{\infty} \mathcal{M}_1(d\tau), \end{aligned} \tag{54}$$

where $\mathcal{M}_i(d\tau)$ $(i = 0, 1)$ is the operator value measure

$$\begin{aligned} \mathcal{M}_i(d\tau) &= \frac{d\tau}{2\pi}\left(\int\limits_{\Delta k_i} \frac{dk}{\sqrt{k}} \exp(-ik\tau)|k\rangle\right) \\ &\times \left(\int\limits_{\Delta k_i} \frac{dk'}{\sqrt{k'}} \exp(ik'\tau)\langle k'|\right). \end{aligned} \tag{55}$$

The measurement is performed in such a manner that it "transmits" only states with the frequency bands falling within $\{\Delta k\}_0$ or $\{\Delta k\}_1$ (the bands are assumed to be equal in width and differ only in position on the frequency axis) or, in other words, only states with effective spacetime lengths not less than T. The probability that the measurement performed by Bob on the state $|\varphi_0\rangle$ will give outcome 0 for the $(L_{\mathrm{ch}}, L_{\mathrm{ch}} + 2T)$ time window (the interval in which the state would occur at a probability exponentially close to unity in the absence of delay) is

$$\begin{aligned} \Pr(i, T) &= \mathrm{Tr}\{\mathcal{M}_i(T)U(L_{\mathrm{ch}})|\varphi_i\rangle\langle\varphi_i|U^{-1}(L_{\mathrm{ch}})\} \\ &= \int\limits_{L_{\mathrm{ch}}}^{L_{\mathrm{ch}}+2T} d\tau|\varphi_i(\tau)|^2 = \lambda_0(2\Delta kT) \\ &= 1 - O(\exp(-2\Delta kT)), \end{aligned} \tag{56}$$

where $U(L_{\mathrm{ch}})$ is the unitary operator describing translation of a state over the channel length L_{ch}. The probability of detecting states outside the time window is exponentially small.

Since all measurements performed by Eve are performed in the communication channel and give information only while the state is present in the channel, this scheme of cryptography eliminates the problem of collective measurements arising in nonrelativistic quantum cryptography where all protocols are formulated only in the Hilbert space of states (i.e., beyond spacetime). Therefore, a key value in every event of a state transfer via the communication channel is the probability of identifying the bit transmitted by Alice and to pass the test during Bob's measurements (54)–(56) on the receiver end. The test is considered successfully passed if measurement (54)–(56) gives an outcome within the $(L_{\mathrm{ch}}, L_{\mathrm{ch}} + 2T)$ time window. In the subsequent key distribution, only messages with an outcome in this window in Bob's measurements are left. Naturally, Eve can preliminary send arbitrary states to yield an outcome within the desired time window in Bob's measurements. These attempts can be detected by means of a protocol whereby a part of the messages are unveiled to determine the flux of errors.

The probability that an eavesdropper possesses some information on the transmitted state is restricted to the extent of her access to a spatial region in which the state is present. If Eve would like to access a region (spatial or temporal) of size δT, the probability of a given outcome of any measurement is determined only by the size of this region and, by virtue of the state normalization condition, cannot exceed

$$\begin{aligned} \Pr_{\mathrm{E}}(\delta T) &= \mathrm{Tr}\{\mathcal{M}(\delta T)|\varphi_i\rangle\langle\varphi_i|\} = \int\limits_{T-\delta T}^{T} d\tau|\varphi_i(\tau)|^2, \\ \delta T &\in -(T, T). \end{aligned} \tag{57}$$

The total probability of correct identification of the given state (without taking into account the temporal test at the receiver end) consists of two parts. If the monitoring took place within a spacetime region of size δT, the state is uniquely identified (due to orthogonality), the probability of such an event being $\Pr_{\mathrm{E}}(\delta T)$. In the absence of monitoring, the probability is $1 - \Pr_{\mathrm{E}}(\delta T)$.

Here, the probability of correct identification is 1/2 (i.e., it equals the probability of simple coin tossing), since no monitoring took place (the eavesdropping device fails to operate).

The total probability of correct identification in the δT time window is

$$\begin{aligned}\mathrm{Pr}_{\mathrm{OK}}(\delta T) &= 1\mathrm{Pr}_{\mathrm{E}}(\delta T) + \frac{1}{2}(1 - \mathrm{Pr}_{\mathrm{E}}(\delta T)) \\ &= \frac{1}{2}(1 + \mathrm{Pr}_{\mathrm{E}}(\delta T)) > \frac{1}{2}.\end{aligned} \tag{58}$$

Due to the finiteness of the velocity of light, access to the δT region will unavoidably lead to a delay in Bob's measurements in the (L_{ch}, $L_{\mathrm{ch}} + 2T$) time window (Fig. 3), since the measurements are performed so that states with effective lengths below $2T$ (or the frequency band width above Δk) are not transmitter. More precisely, the operator measure $\mathcal{M}(d\tau)$ in Eqs. (54), (55) describing the measurement contains a projector onto the subspace of states only with carriers belonging to the frequency band Δk. Among these, the states with shortest lengths are the optimum states corresponding to Eqs. (44)–(46). Any other states, even nondelayed, are more lengthy and cannot be transmitted via the temporal (or spatial) window ($-T$, T) with a probability exceeding that for the optimum states.

Should Eve extract information about the transmitted bits from the optimum states, these states will unavoidably acquire a delay (Fig. 3) dictated by the existence of the limiting velocity, thus yielding outcomes outside the (L_{ch}, $L_{\mathrm{ch}} + 2T$) window in Bob's measurements. The probability for a state delayed by δT to pass the delay test in Bob's measurements does not exceed

$$\begin{aligned}\mathrm{Pr}_{\mathrm{B}}(\delta T) &= \mathrm{Tr}\{\mathcal{M}(\tau \in (L_{\mathrm{ch}}, L_{\mathrm{ch}} + 2T))U(L_{\mathrm{ch}}) \\ &\qquad \times |\varphi_i\rangle\langle\varphi_i| U^{-1}(L_{\mathrm{ch}})\} \\ &= \int\limits_{L_{\mathrm{ch}}}^{L_{\mathrm{ch}}+2T} d\tau |\varphi(L_{\mathrm{ch}} + \tau - \delta T)|^2 = \int\limits_{-T}^{T-\delta T} d\tau |\varphi(\tau)|^2.\end{aligned} \tag{59}$$

The probability of passing the test with a delay of δT under the condition that the eavesdropping device has operated (and that Eve knows the transmitted state) is

$$\mathrm{Pr}_{\mathrm{E}}(\delta T)\mathrm{Pr}_{\mathrm{B}}(\delta T) = \mathrm{Pr}_{\mathrm{E}}(\delta T)(1 - \mathrm{Pr}_{\mathrm{E}}(\delta T)). \tag{60}$$

For a zero delay ($\delta T = 0$), the first multiplier (describing the probability for Eve to obtain the correct result) is zero. The second multiplier is the probability of passing the delay test in Bob's measurements. Should the delay be $\delta T = 2T$, the probability for Eve to obtain the correct result is unity (exponentially close to unity), but the probability of passing the delay test is zero.

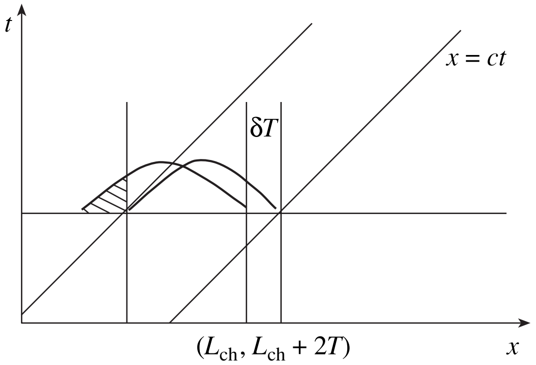


Fig. 3. Schematic diagram illustrating the time delay test: the cross-hatched part of a state delayed in transit does not contribute to the outcomes of measurements in the time window (L_{ch}, $L_{\mathrm{ch}} + 2T$) at the receiver end of the communication channel.

The total probability for Eve to correctly identify the transmitted state and pass the test for delay at the receiver end is

$$\begin{aligned}\delta_{\mathrm{OK}}(\delta T) &= \mathrm{Pr}\{\mathrm{bit}_{\mathrm{E}} \\ &= \mathrm{bit}_{\mathrm{A}} \wedge \mathrm{test}(\tau \in (L_{\mathrm{ch}}, L_{\mathrm{ch}} + 2T)) = O'K\} \\ &= (1 - \mathrm{Pr}_{\mathrm{E}}(\delta T)) \times \frac{1}{2} \times 1 \\ &\quad + \mathrm{Pr}_{\mathrm{E}}(\delta T) \times 1 \times (1 - \mathrm{Pr}_{\mathrm{E}}(\delta T)).\end{aligned} \tag{61}$$

Here, the first multiplier in the first term describes the probability that the eavesdropper device would fail to operate for a given delay δT, the second multiplier gives the probability of correct identification in the absence of monitoring (equal to 1/2), and the third term reflects the probability of passing the time delay test. The latter probability is unity because, if the detector fails to operate, the state freely passes through the eavesdropper without any interactions. The same is valid for the second term, where the first multiplier is the probability for the device to give the outcome 0 or 1. The second term is the probability of identification, which is equal to unity because the states are orthogonal, and the third term is again the probability of passing the time delay test.

The value of $\delta_{\mathrm{OK}}(\delta T)$ reaches maximum for a certain value of δT, this maximum being denoted by $\delta_{\mathrm{OK}}^{\max}$. For the following, let us find the minimum Eve's error, δ_{E}, in determining the transmitted bit while passing the delay test in Bob's measurements. Only such outcomes at the receiver end will be used in the proposed secure bit commitment protocol. For this error, defined as

$$\delta_{\mathrm{E}} = 1 - \delta_{\mathrm{OK}}^{\max}, \tag{62}$$

the maximum probability for Eve to identify the transmitted bit while passing the delay test is attained for

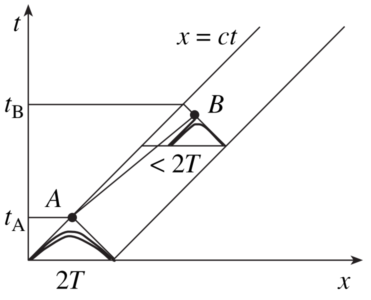


Fig. 4. Schematic diagram illustrating violation of the relativistic causality principle. Hypothetical contraction of the state leads to the possibility of transmitting information at a supraluminal velocity between points A and B separated by a spacelike interval.

$\Pr_E(\delta T) = 1/4$. This yields

$$\delta_E = \frac{7}{16} = \frac{1}{2} - 0.0625 = 43.75\%. \tag{63}$$

Accordingly, the probability that Eve will correctly identify the bit transmitted by Alice and successfully pass the delay test in Bob's measurements is

$$\delta_{OK}^{\max} = \frac{9}{16} = \frac{1}{2} + 0.0625 = 56.25\%. \tag{64}$$

Now it will be demonstrated that the probability of detection in any time window for a noisy channel cannot be higher than the corresponding probability for the ideal channel. This will imply that the probability given by Eq. (61) is the maximum that can be expected by Eve in extracting information about the transmitted state. The presence of noise in a communication channel can increase neither the information obtained by an eavesdropper nor the probability of passing the time delay test.

Since the states in the relativistic case are described by vectors in the Hilbert space of states, a change in the given state under the action of noise can be described by an instrument with allowance of the relativistic constraints. The general form of such an instrument is [29]

$$\mathcal{T}[\ldots] = \sum_k \mathcal{S}_k[\ldots]\mathcal{S}_k^+, \quad \mathcal{S}_k = \sqrt{\lambda_k}|\phi_k\rangle\langle\psi_k|,$$

$$\sum_k \lambda_k \mathcal{S}_k \mathcal{S}_k^+ \le 1, \quad \lambda_k \ge 0, \tag{65}$$

$$\mathrm{Tr}\{\mathcal{T}[|\varphi_{0,1}\rangle\langle\varphi_{0,1}|]\mathcal{M}(T)\}$$

$$= \sum_k \mathrm{Tr}\{|\varphi_{0,1}\rangle\langle\varphi_{0,1}|(\mathcal{S}_k\mathcal{M}(T)\mathcal{S}_k^+)\}$$

$$\le \sum_k \lambda_k \mathrm{Tr}\{|\varphi_{0,1}\rangle\langle\varphi_{0,1}|(|\psi_k\rangle\langle\psi_k|)\} \tag{66}$$

$$\le \sum_k \lambda_k |\langle\psi_k|\varphi_{0,1}\rangle|^2 \le \sum_k \lambda_k \langle\psi_k|\psi_k\rangle\langle\varphi_{0,1}|\varphi_{0,1}\rangle$$

$$\le \langle\varphi_{0,1}|\varphi_{0,1}\rangle = \int_T dx|\varphi(x)|^2 = \int_T dx|\varphi(x - L_{ch})|^2.$$

The last equality in Eq. (66) reflects the fact that the amplitude $\varphi(x)$ of the state

$$|\varphi_{0,1}\rangle = \int_T dx\varphi(x)|x\rangle = \int_T dx\varphi(x - L_{ch})|x\rangle \tag{67}$$

is entirely localized in the region of $x \in (-T, T)$ at the time $t = 0$ and will be entirely localized in the region of $x \in (-T, T) + L_{ch}$ by the time not earlier than L_{ch} and not faster than within $t = \mathrm{dist}(L_{ch})$. Strictly speaking, the problem in field theory is not a single-particle one, since the operator $\hat{S}$ contains processes involving the production of particles and the absorption of other photons falling in the channel as external noise and/or vanishing in the channel. However, detecting these photons apparently does not provide an eavesdropper with additional information about the transmitted bit.

In other words, if some region Ω of the space gains some fraction of the normalization of a given state, the propagation and distortion of the shape of this state (packet) cannot lead to a situation whereby a greater fraction of the normalization of this state will be gained in another region of smaller size ($\mathrm{size}\{\Omega'\} < \mathrm{size}\{\Omega\}$). If this were possible, the classical information could be transmitted by quantum states at a velocity higher than that of light, which contradicts one of the basic laws of physics.

Indeed, consider a pair of identical quantum states prepared in the past (Fig. 4). Let an observer at point A (Alice) perform measurements on one of these states, while the other is transmitted to another observer at point B (Bob). Alice can obtain information about the quantum state not before the time t_A (the state must be covered by the past portion of the light cone originating at point A). Events will always occur (albeit not with a probability of unity because of the principal nonlocalizability of quantum states) when Alice will obtain a certain result. Once this result takes place, Alice can send a classical signal (or arbitrarily strongly localized quantum signal, see (44)–(46)). According to the relativistic causality principle, Bob will receive this information at time t_B (Fig. 4) and will never receive it earlier because these events are separated by a spacelike interval.

Now let the state to be distorted due to evolution and under the action of noise, so that the greater part of the normalization is gained within a smaller spatial region (i.e., the state is contracted). Then, Bob can obtain the information (with the same probability of positive outcome as that for Alice) by measurements on the second quantum state earlier than if he would have received this information from Alice upon accomplishing the same measurements at point A. The fact that such events take place with a probability different from unity is not of much importance, since the probability can be rendered arbitrarily close to unity by increasing the number of quantum states.

Thus, the probability of obtaining any information about a transmitted quantum state and of passing the time delay test at the receiver end is always smaller than unity, irrespective of the quantum state. This fact allows an unconditionally secure cryptographic bit commitment protocol to be formulated. In this cryptosystem, detection of the attempts at eavesdropping is guaranteed to the legitimate parties by the basic laws of nature, including quantum mechanics and special relativity. The fact of orthogonality or nonorthogonality of states, which was of principal importance for certifying security in the nonrelativistic case, is no longer critical in the relativistic domain. This implies that relativistic quantum cryptography can employ any quantum states (naturally, provided that an appropriate procedure of measurements is developed for detecting time delays).

In concluding this section, it should be noted that the fact of alteration of the quantum states (except for orthogonal) as a result of any measurements (observations) on these states can be, in application to quantum cryptography, restated in a different way: the obtaining of any information about a quantum state requires access to a spacetime region in which the state is present, which always unavoidably requires a finite time in view of the finiteness of the velocity of light. The structure of states (orthogonality or nonorthogonality) is not critical, whereas the fact that any quantum state has a carrier (amplitude, smoothening function) in the Minkowski spacetime is of principal importance. The space of the results of measurements on which the operator value measure is defined may possess any structure and unavoidably contains spacetime regions, since no measurements can be performed outside space and time. This remark may seem trivial, but the above circumstance is usually ignored, while being implied implicitly, in solving the problems of quantum cryptography.

It should be also noted that the probabilities of the results of measurements presented above are essentially expressed via the Landau–Peierls wave function φ_{LP} [14]. All these results can be rewritten in terms of the Bialynicki-Birula wave function $\varphi_{\mathrm{B\text{-}B}}$ [30], the square modulus of which in the spacetime representation gives the local energy density. The amplitudes (wave functions) in these representations are related as

$$\sqrt{k}\varphi_{\mathrm{LP}}(k) = \varphi_{\mathrm{B\text{-}B}}(k). \tag{68}$$

The existence of spacetime allows some other quantum cryptographic protocols to be formulated, which would be impossible if only the structure of the Hilbert space of states were taken into account. One of such base cryptographic procedures is the so-called coin-flipping protocol. In the relativistic quantum domain, a secure protocol can be formulated with an arbitrarily small displacement.

4. A SECURE KEY DISTRIBUTION PROTOCOL

Up to the present, several variants of proved unconditional security in quantum cryptography have been reported. Since no proof of the security in general can exist, the considerations have to refer to a particular exchange protocol. Mayers [31] was undoubtedly the first to clearly formulate a criterion of security and prove this in application to the so-called Bennett–Brassard (BB84) protocol [9]; the same protocol was considered in [32]. The proof of Lo and Chau [33] refers to a protocol based on the Einstein–Podolsky–Rosen (EPR) effect, stipulating (in contrast to [31, 32]) that legitimate parties use a (not yet existing) quantum computer. Shor and Preskill [34] simplified the aforementioned proofs by explicitly introducing correcting codes into the cryptographic scheme.

The main difficulty encountered in proving unconditional security arises from the fact that the aforementioned protocols are formulated in terms of “exchange” in the Hilbert space of states and explicitly use neither the notions of causality nor the fact that legitimate users are separated in the coordinate space. However, the transfer of information always implies preparation of an information carrier (quantum state), transmission of this state via a communication channel to a remote receiving party, and subsequent measurement of the state of the information carrier. The formulation of protocols using only the properties of the space of states unavoidably leads to the necessity of proving the stability of these protocols with respect to collective measurements (allowance for which presents the most significant difficulty).

First, let us formulate a security criterion for the key. In contrast to [31], we will adopt a different security criterion that is more convenient in the case under consideration. A secret key must obey two requirements, informally reduced to the following: the key must be the same for both legitimate parties *A* (Alice) and *B* (Bob) and not known to anyone else. More formally, let an *m*-bit sequence obtained by both Alice an Bob according to the protocol be mutually agreed upon to be the key. Then, the key is secret provided that the following conditions are satisfied:

Identity of the key. The probability that every bit in the *m*-bit sequence adopted as the key is different for

Alice ($b_A(i)$) and Bob ($b_B(i)$) is exponentially small:

$$\forall \varepsilon_1 > 0, \quad \exists M,$$
$$\Pr\{b_A(i) \neq b_B(i)\} \le e^{-M} \le \varepsilon_1. \tag{69}$$

In terms of mutual information exchange between Alice and Bob about the resulting bit sequence of length m,

$$I(A;B) \ge m - 2^{-M}. \tag{70}$$

Security of the key. The probability that an eavesdropper (Eve) knows all bits in the key sequence may only exponentially slightly exceed the probability of trivial guessing (coin tossing), 2^{-m} (the probability of error in trivial guessing is 1/2, representing the worst variant), for a resulting bit sequence of length m. Thus, the probability that Eve knows every bit in the sequence is exponentially small as compared to 1/2:

$$\forall \varepsilon_2 > 0, \quad \exists \eta_2, \zeta,$$
$$\Pr\{b_A(i) = b_E(i)\} \le \frac{1}{2} + e^{-\eta_2}, \quad e^{-\eta_2} \le \varepsilon_2. \tag{71}$$

In other words, Eve possesses exponentially small information about bit sequences $b_A(m)$ and $b_B(m)$ adopted as the m-bit key by legitimate parties:

$$I(A;E) \le e^{-\eta_2} \le \varepsilon_2, \quad I(B;E) \le e^{-\eta_2} \le \varepsilon_2. \tag{72}$$

The quantities ε_1, η_1 and ε_2, η_2 can be selected independently of each other. The initial bits transmitted by Alice should not be identified with the bit sequence adopted as the key: every bit in the key sequence is a certain function of the set of initial bits.

Now let us formulate a protocol of secure bit distribution between legitimate parties. Since any receipt of information about transmitted states is accompanied by the appearance of errors, an appropriate protocol guarantees that a sequence adopted as the key is the same for both legitimate parties, Alice and Bob, and is not known to anyone else. The protocol is as follows:

1. Alice and Bob first agree on a large integer $N \gg 1$. The protocol employs $2N$ bits.

2. Alice and Bob synchronize their clocks via an open (insecure) channel, the length of the communication channel being known previously. The same open channel is used to agree on the moments t_i ($i = 1, 2, \ldots, 2n$) of sending messages, such that $t_{i+1} - t_i > 2T$.

3. Alice generates a random sequence b_i of bits (0 and 1) of length $i = 1, \ldots, 2n$.

4. Alice transmits the states $|\varphi_{b_i}\rangle$ at the time moments t_i with carriers in the frequency band Δk (also mutually agreed upon prior to beginning the protocol). The shapes of these states are not necessarily optimum with respect to localization according to (44)–(46).

5. Bob performs measurements on the receiver end, as described by a subordinate partition of unity (54) and (55) and informs Alice of receipt via the open channel.

6. After Alice has transmitted all states, Bob communicates to Alice via the open channel the numbers of transmissions in which the results of measurements had outcomes in the time windows ($t_i + L_{ch}$, $t_i + L_{ch} + 2T$) (messages with outcomes outside these time windows are discarded). Let the number of such outcomes be $2n$.

7. Alice randomly chooses n of $2n$ outcomes and communicates to Bob via an open channel the values of bits transmitted with these messages.

8. Using the open channel, Alice and Bob compare bits in each position of the unveiled sequence and estimate the probability of errors. Let n_{OK} be the number of positions in which the bits coincide, and $n_{\overline{OK}}$, the number of discrepancies. Then, the estimated probability of error is $\delta_{err} = n_{\overline{OK}}/n_s$. If n_s is sufficiently large, the probability of error in the unveiled part of messages is also exponentially close to δ_{err}.

9. Alice and Bob correct errors in the remaining sequence of nonunveiled bits. To this end, a classical error-correcting code $[n, k, d]$ is selected with a minimum distance d according to Hamming and the number of code words 2^k ($d > 2\delta_{err}n + 1$ for a linear code or $d > \delta_{err}n + 1$ for a random linear code, see below). For this purpose, Alice communicates to Bob via the open channel a series of v_i ($i = 1, 2, \ldots, r$; $r = n - k$) test sequences of the given code. In addition, Alice communicates r test parity bits $(\text{parity})_i = v_i n_A$ (n_A and n_B being the sequences of nonunveiled bits of Alice and Bob, respectively). These sequences do not, generally speaking, coincide but differ (with a probability arbitrarily close to unity) in approximately $\delta_{err}n$ positions to be corrected.

10. Knowing the correct parity of subsequences, Bob corrects errors in the n_{sB} sequence. In this step, the sequences of bits for Alice and Bob are identical with a probability arbitrarily close to unity at sufficiently large n (see below).

11. Alice and Bob enhance the key security using the so-called privacy amplification procedure. For this purpose, a certain hash function is selected and declared via the open channel. The privacy amplification consists in reducing a k-bit sequence $\hat{w} = \{w_i\}$, obtained by Alice and Bob upon correcting errors using the correcting code in step 9, to a shorter sequence of length $m < k$ so as to ensure that it is unknown to Eve (or known with an exponentially small probability determined by a selected security parameter). To this end, Alice chooses m sequences $\hat{l} = \{l_i\}$ of length k ($i = 1, 2, \ldots, m$). The final secret m-bit sequence is obtained in the form of parity bits: $(k\hat{e}y) = \hat{w}\hat{l}$.

12. If the length of sequences $\hat{w}$ is insufficient to provide for the required security of the key, the protocol is terminated.

A complete mathematical proof of the unconditional security of this protocol will be published separately. Here we will consider only an outline of this proof so as to clarify the underlying intuitive ideas.

For a sufficiently large number (2*n*) of transmissions in which the results of Bob's measurements had outcomes in the correct time windows (see step 6), estimation of the probability of errors in the unveiled randomly selected part of *n* transmissions chosen from the total number 2*n* of such transmissions guarantees (with a probability arbitrarily close to unity) that the number of errors (noncoinciding bits) in the nonunveiled *n* transmissions is

$$n_{\text{err}} = \delta_{\text{err}} n. \tag{73}$$

This circumstance allows a linear classical error-correcting code $[n, k, d]$ to be selected, which has the code distance $d/n \ge 2\delta_{\text{err}}$ and contains

$$k \ge n(1 - H(2\delta_{\text{err}}))$$

informative symbols. This code ensures error correction in $(d - 1)/2$ positions with a probability arbitrarily close to unity [35–37]. Here,

$$H(x) = -x\log x - (1 - x)\log(1 - x) \tag{74}$$

is the binary entropy function. This statement follows from the achievable Varshamov–Hilbert boundary [37] for linear codes. The estimate can be improved using random linear error-correcting codes, for which $d/n > \delta_{\text{err}}$ and the remaining number of informative bits is

$$k < n(1 - H(\delta_{\text{err}})) \tag{75}$$

(the Shannon limit). However, the estimate following from the Varshamov–Hilbert inequality is constructive. There exist linear regular (nonrandom) codes for which this boundary is achieved (in contrast to the Shannon limit, which can be achieved only on random codes) and is essentially a theorem of existence, rather than a constructive limit: no regular codes are known on which this boundary can be achieved [37].

The proposed protocol works until the error probability becomes $\delta_{\text{err}} < \delta_{\text{E}}$. The code selected by Alice and Bob must correct all fluxes of errors with probabilities below δ_{err}. Eve may know all bits used in the protocol with an error probability of $\delta_{\text{E}} \approx 43.75\%$. Using information concerning the correction of errors (error-correcting code and verification parity sequences) communicated via the open channel, Eve may, in principle, make the probability of her errors arbitrarily close to zero, provided that $\delta_{\text{err}} > \delta_{\text{E}}$.

For $\delta_{\text{err}} < \delta_{\text{E}}$, there exists in principle a random error-correcting code with a code distance of

$$d/n \ge \delta_{\text{err}} \quad (\text{but} \quad d/n < \delta_{\text{E}}), \tag{76}$$

which corrects the errors for Alice and Bob with a probability arbitrarily close to unity (virtually unity for large *n*), but not for Eve. The number of remaining (identical) bits for Alice and Bob upon correction with sufficiently large $n \gg 1$ is approximately

$$nC(\delta_{\text{err}}), \tag{77}$$

where

$$C(\delta_{\text{err}}) = 1 - H(\delta_{\text{err}}) \tag{78}$$

is the transmission capacity of a classical symmetric binary communication channel.

Under the condition (76), Eve occurs essentially in the situation of a binary symmetric channel [35–37], whereby the transmission rate (bit/message) exceeds the transmission capacity $C(\delta_{\text{E}})$ of the channel between Eve and Alice. The error correction by means of codes unfavorable for an eavesdropper (under the condition that $\delta_{\text{err}} < \delta_{\text{E}}$) appears for Eve as the transmission of communications at a rate exceeding the transmission capacity.

For transmission rates exceeding the channel capacity, it is possible to use the Wolfowitz estimate [38] for the probability of error per symbol. For Eve, this probability is not less than

$$\begin{aligned} p_{\text{err}} > 1 &- 4\frac{\text{const}}{n(C(\delta_{\text{err}}) - C(\delta_{\text{E}}))^2} \\ &- \exp\left[-\frac{n(C(\delta_{\text{err}}) - C(\delta_{\text{E}}))}{2}\right]. \end{aligned} \tag{79}$$

According to this, the limiting permissible probability of errors in the communication channel is $\delta_{\text{err}} \approx 21.875\%$, the corresponding Shannon limit being $\delta_{\text{err}} \approx 43.75\%$.

It should be pointed out that the permissible probability of errors (at which a protocol works and allows a secure key to be created) for quantum cryptography on nonorthogonal states (using the so-called BB84 protocol [9]) is 7.5% (with the corresponding Shannon limit $\delta_{\text{err}} \approx 11\%$). This limit appears because the BB84 protocol requires correction of both the phase errors (arising during measurements in different basis sets) and the bit-flip errors (switching from 0 to 1 and vice versa) [34]. The threshold is determined from the equation

$$1 = 2H(2\delta_{\text{err}}), \tag{80}$$

or in the Shannon limit,

$$1 = 2H(\delta_{\text{err}}). \tag{81}$$

After the error-correcting procedure, the bit sequences of legitimate users are identical with a probability arbitrarily close to unity. In principle, after this procedure (with $n \gg 1$), the information obtained by an eavesdropper about the correct bit sequence tends to zero. Nevertheless, the security can be enhanced by using a privacy amplification procedure.

A simple example of the hash function is offered by the function yielding a bit of parity for an m-bit sequence, which is the final secret bit. This secrete bit is the same for both legitimate parties, Alice and Bob, with a probability arbitrarily close to unity, while Eve knows this bit with an arbitrarily small preset probability. If every bit in a correct sequence is known to Eve with a probability of p_e (in contrast to Alice and Bob, for which this sequence is exactly known), the error in determining the secret bit of parity (for certainty, assuming m to be even) is

$$P_{\text{err}}(\text{parity}) = \sum_{i=\text{odd}}^{m-1} C_m^i p_e^i (1-p_e)^{m-i}, \tag{82}$$

where the sum is taken over odd indices i (an error in determining the parity bit appears when Eve makes mistakes in an odd number of positions).

Taking into account that

$$\frac{1}{2}[(x+y)^m - (x-y)^m] = \sum_{i=\text{odd}}^{m-1} C_m^i x^i y^{m-i} \tag{83}$$

and assuming that $x = p_e$ and $y = 1 - p_e$, we obtain

$$P_{\text{err}}(\text{parity}) = \frac{1}{2}[1 - (1-2p_e)^m]. \tag{84}$$

Accordingly, the probability that Eve knows the secret bit is

$$\begin{aligned} P_{\text{OK}}(\text{parity}) &= 1 - P_{\text{err}}(\text{parity}) \\ &= \frac{1}{2}[1-(1-2p_e)^m] = \frac{1}{2} + \varepsilon, \end{aligned} \tag{85}$$

where ε is an exponentially small quantity. According to (77), the probability that Eve would correctly identify the secret bit (if the number m of bits remaining after the correction of errors is large) can be rendered exponentially small as compared to the probability of trivial guessing (always existing and equal to 1/2).

5. SECRET KEY GENERATION IN REAL TIME

Let us turn to the question of secret key generation in real time. The limiting key distribution rate (in bit/s) is determined by the properties of a communication channel, the external noise, and the intensity of eavesdropping. Since the activity of an eavesdropper is not restricted, it is impossible to establish the rate of key distribution in the general case, with arbitrary noise and an unpredictable eavesdropper. In particular, an eavesdropper can perturb the communication channel so strongly that the key distribution will be virtually blocked. All cryptosystems only certify to legitimate parties that the key distribution is secure, provided that all tests stipulated by the protocol have been successfully accomplished. However, the question of the limiting rate of key generation in real time is quite reasonable: this rate depends only on the properties of a particular quantum communication channel (in the case under consideration, on the frequency pass band), essentially representing the transmission capacity of the quantum communication channel. The limiting key generation rate cannot exceed the channel transmission capacity.

Elegant and profound results have been obtained in the theory of transmission capacity of quantum communication channels. These results have been exhaustively reviewed by Kholevo [39]. However, the coding theorems (being formulated only in terms of the properties of the Hilbert space of states) for quantum systems describe the transmission rate as the amount of bits (≤ 1) per transmission, rather than as the transmission capacity in real time. In order to characterize the real-time capacity of a channel, it is necessary to introduce spacetime into the problem. In particular, the real-time transmission capacity can be determined in the case when information is encoded using the polarization states [41]. Below, preliminary results of the author are generalized to the case when information is encoded in the form of a packet, more precisely, when carriers of the informative states $|\varphi_{0,1}\rangle$ are nonoverlapping on the frequency scale.

A quantum communication channel is set by mapping the input operators of density matrices into output operators, also representing the density matrices. In different terms, such a map is described as an instrument or a superoperator. It turns out that measurements performed at the receiver end within a finite time window can be described by introducing a superoperator generating the operator measures $\mathcal{M}(T)$ and $I(\Delta k) - \mathcal{M}(T)$. According to the Krause theorem [29], any mapping of the density matrices into density matrices (superoperator) possessing the property of conservation of the trace, linearity, and complete positive determinacy can be presented in the following form:

$$\mathcal{T}[\ldots] = \sum_k V_k[\ldots]V_k^+, \quad \sum_k V_k^+ V_k = I. \tag{86}$$

Here, index k runs over the space of the results of measurements, which can be either a discrete set or continuum. The probability of outcomes of the measurements is described by an operator value measure related to the operators in representation (86) as

$$\mathcal{M}_k = V_k^+ V_k. \tag{87}$$

The probability of the *k*th outcome for the input state with a density matrix ρ is

$$\Pr(k) = \mathrm{Tr}\{\rho \mathcal{M}_k\}. \tag{88}$$

Note that, generally speaking, a superoperator is not always uniquely reconstructed from its operator value measure. However, no such difficulties are encountered in the problem under consideration.

In the case studied, a superoperator can be represented by a direct sum of operators acting upon the density matrices with carriers in nonoverlapping frequency bands $\{\Delta k\}_0$ and $\{\Delta k\}_1$:

$$\mathcal{T}[\ldots] = \mathcal{T}_0[\ldots] \oplus \mathcal{T}_1[\ldots], \tag{89}$$

where,

$$\begin{aligned}\mathcal{T}_i[\rho_i] &= \sqrt{\mathcal{M}_i(T)}\rho_i\sqrt{\mathcal{M}_i(T)}^+\\ &+ \mathrm{Tr}\{(I_i(\Delta k) - \mathcal{M}_i(T))\rho_i\}|?\rangle\langle ?|.\end{aligned} \tag{90}$$

Here, $|?\rangle$ is a state orthogonal to the first term in (90), which describes the absence of outcomes for operator's measurements in the time window $(-T, T)$. Using a question mark in this notation implies that, in the absence of outcomes in the $(-T, T)$ window, the operator always considers the result of such a measurement as inconclusive. One can readily check that the above operator value measure obeys the formula

$$\begin{aligned}\sqrt{\mathcal{M}_i(T)} &= \sqrt{\int_{-T}^{T}\frac{d\tau}{2\pi}\left(\int_{\{\Delta k\}_i}\frac{dk}{\sqrt{k}}\exp(-ik\tau)|k\rangle\right)\left(\int_{\{\Delta k\}_i}\frac{dk'}{\sqrt{k'}}\exp(ik'\tau)\langle k'|\right)}\\ &= \frac{1}{\pi}\int_0^{\infty}\frac{d\zeta}{\zeta^{1/2}}\frac{\mathcal{M}_i(T)}{(\zeta I_i(\Delta k) + \mathcal{M}_i(T))} = \sum_{n=0}^{\infty}\sqrt{\lambda_n(2\Delta kT)}|\varphi_n(i)\rangle\langle\varphi_n(i)|,\end{aligned} \tag{91}$$

where

$$|\varphi_n(i)\rangle = \int_{\{\Delta k\}_i} dk\varphi_n(k)|k\rangle \tag{92}$$

and $\varphi_n(k)$ is a function satisfying integral equation (46) in the corresponding frequency band $\{\Delta k\}_i$ $(i = 0, 1)$.

The transmission capacity of a channel can be calculated using the Kholevo formula [39],

$$\begin{aligned}C(T) &= \max\{\pi_0, \pi_1\}\\ &\times\left\{H(\mathcal{T}[\rho]) - \sum_{i=0,1}\pi_i H(\mathcal{T}[\rho_i])\right\},\end{aligned} \tag{93}$$

where

$$H(\rho) = -\mathrm{Tr}\{\rho\log\rho\} \tag{94}$$

is the von Neumann quantum entropy. The input density matrix is

$$\begin{aligned}\rho &= \pi_0|\varphi_0\rangle\langle\varphi_0| + \pi_1|\varphi_1\rangle\langle\varphi_1|,\\ \rho_i &= |\varphi_i\rangle\langle\varphi_i|, \quad i = 0, 1,\end{aligned} \tag{95}$$

where π_0 and π_1 are the a priori probabilities. The maximum in formula (93) is attained at $\pi_0 = \pi_1 = 1/2$ (i.e., when the states 0 and 1 are transmitted with equal probabilities).

If the initial states are selected as optimum, possessing the minimum possible effective length for a given frequency band (i.e., the states obeying conditions (42)–(44)), the real-time transmission capacity is

$$C(T) = \frac{1-\varepsilon}{T}\left[\frac{\mathrm{bit}}{\mathrm{s}}\right], \quad 1-\varepsilon = \lambda_0(\Delta kT). \tag{96}$$

Here, the quantity $\lambda_0(\Delta kT)$ is the maximum eigenvalue of the integral equation (44). Formula (96) coincides with an expression for the transmission capacity of a binary classical channel with erasure. The probability of erasure, ε, is related to the real-time transmission rate and the frequency pass band. Thus, even an ideal quantum channel exhibits "noise" related to a finite observation time at the receiver end.

If the measurements at the output of the communication channel are performed in a large time window (formally infinite, $T \longrightarrow \infty$), the reliable distinguishability of orthogonal states (entirely accessible in this limit) allows the number of correctly decoded bit sequences to be $2^{nH(\rho)}$. For orthogonal states, there is no need in collective measurements and it is sufficient to measure states in each individual transmission. Should the measurements be performed in a finite time window $(-T, T)$, there will be outcomes in which the measuring device does not operate within a $2T$ time interval. The probability of such outcomes is ε and, accordingly, the probability of outcomes within the time window is

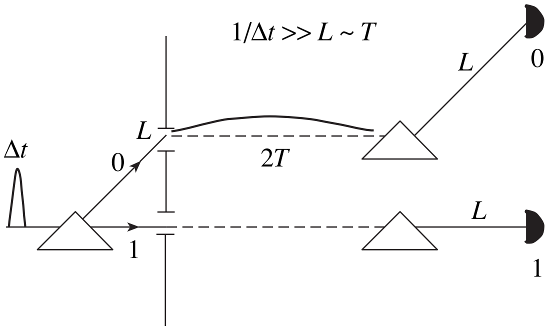


Fig. 5. Schematic diagram illustrating the proposed cryptosystem (see the text for explanations).

$1 - \varepsilon$. The device operation within T allows the states to be identified with confidence. In the absence of outcomes within the given time window, the state can be considered as erased (formally speaking, the state is transformed into a new state (?) at the receiver end; also formally, the state can be considered as a transmitter with a probability of $p_? = 0$). As a result, every typical sequence expands into a Hamming sphere of radius $H(x|y)$, where $H(x|y)$ is the conditional Shannon entropy for the input alphabet ($x = \{0, 1, ?\}$, with the probabilities $\{p_0 = 1/2, p_1 = 1/2, p_? = 0\}$), output alphabet ($y = \{0, 1, ?\}$), and the probabilities of transition in the channel,

$$p(0|0) = p(1|1) = 1-\varepsilon,$$
$$p(0|1) = p(1|0) = 0, \quad p(0|?) = p(1|?) = \varepsilon.$$

Therefore, the number of correctly decoded bit sequences for $n \longrightarrow \infty$ is

$$\frac{2^{nH(\rho)}}{2^{nH(x|y)}} = 2^{nI(x:y)}, \quad I(x:y) = 1-\varepsilon. \tag{97}$$

This formula coincides with the expression for the transmission capacity of a classical binary channel with erasure [36] and determines the maximum rate of key transmission for the given frequency pass band and observation time window.

6. A SIMPLE EXAMPLE OF THE CRYPTOSYSTEM REALIZATION

Let us consider a simple example of implementation of the proposed cryptosystem. This example is intended for illustrative purposes and only demonstrates physical interpretation of the problem. Consider the preparation of states within a finite frequency band. Let a source (e.g., laser) generate a quantum state of sufficiently short duration (and, accordingly, possessing a sufficiently broad frequency spectrum with an approximate width of $\Delta\Omega$). The time of the protocol start is set to within $\Delta t \approx 1/\Delta\Omega$. Then the signal is attenuated down to a single-photon level (e.g., upon passage through an absorber) and transmitted through a filter (frequency-dispersive medium) cutting a band with a width of $\{\Delta k\}_0 \ll \Delta\Omega$ (or a band of the same width centered at some other carrier frequency, $\{\Delta k\}_1 \ll \Delta\Omega$, see Fig. 5) from the broad frequency spectrum $\Delta\Omega$. This procedure can be, in principle, performed similarly to the classical Newton experiment on the dispersion of a "white" (broadband) light beam.

Each transmission is effected by randomly opening one of the apertures (Fig. 5). The farther the aperture from a dispersive element, the higher the accuracy of cutting frequency bands from the broad spectrum. The cut frequency bandwidth is approximately equal to $1/T \propto 1/L$ (L is the distance to the aperture, or the effective length of the state). The proportionality sign implies that the actual frequency band width depends on the prism dispersion (i.e., on the angle of deviation of the corresponding spectral components). Naturally, the probability of cutting a narrow band from a broadband signal is smaller than unity. Therefore, there will be events whereby no signals enter the channel, but such blank preparation outcomes influence only the efficiency of the scheme, while not affecting security of the cryptosystem. Past the aperture, a state enters the communication channel. The leading front enters the channel within a time period of approximately $\Delta t + L$.

The receiver (Fig. 5) employs a procedure whereby only the states carried within a frequency band about $\{\Delta k\}_0$ can pass to a detector. This measurement represents realization of the operator measure cutting (transmitting) only a frequency band of about Δk (see Eqs. (54) and (55)). This circumstance does not allow an eavesdropper to use short-time (broadband) states for compensating a time delay introduced by his activity. Undelayed states carried within a frequency band about $\{\Delta k\}_0$ will make counts in a given time window (L_{ch}, $L_{\text{ch}} + 2T$), naturally, to within a certain accuracy of Δt. The protocol makes use only of the events involving measurements in the given time window. The characteristic intrinsic time constant τ_d of the detector has to obey the condition $\tau_d \ll T$. This requirement arises, in fact, from the need in gaining statistics over a time interval of $2T$ for the time delay test.

To the author's knowledge, detectors with a time constant of $\tau_d \approx 10^{-8}$–10^{-9} s are among quite standard devices. The effective length of a detected state should be on the order of $T \approx 1/\Delta k \gg \tau_d$, which provides for a two-order margin in the monochromaticity, $\Delta k \approx 10^6$–10^7 Hz. Accordingly, the effective length is $cT \approx$ 10–100 m, which is a rather soft condition. The accuracy of clock synchronization has to be $\max\{\tau_d, \Delta t\}$.

Under real conditions, a fiber optic communication line is usually not a straight line between the line ends. This circumstance poses certain additional limitations on the effective length of transmitted states. Evidently, this length cannot be smaller than $L_{\text{curv}} - L_{\text{ch}}$, where

$L_{\rm curv}$ is the true fiber length and $L_{\rm ch}$ is the length of a straight segment between the line ends. Moreover, since the velocity of light in the fiber (c') is somewhat smaller than that in vacuum ($c' < c$), the effective length of the transmitted state cannot be shorter than $c(L_{\rm curv} - L_{\rm ch})/c'$.

It should be noted that a cryptosystem employing frequency states has to be more stable than the system using polarization states, because the carrier frequency breakdowns are related to the Raman processes occurring in the second order of perturbation theory.

7. CONCLUSIONS

The main result of this investigation is that allowance for the constraints posed by special relativity allows the use of practically arbitrary quantum states in quantum cryptography for ensuring unconditionally secure key distribution. In contrast to previously developed cryptosystems, the security of which was based on the geometry of the Hilbert space and the properties of nonorthogonal states, the proposed scheme explicitly takes into account the fact that any quantum states possess carriers in the Minkowski spacetime. In addition, the explicit allowance for the spacetime structure of states allows implementation of other important quantum-cryptographic protocols.

ACKNOWLEDGMENTS

The author is grateful to S.S. Nazin for fruitful discussions and critical remarks.

This study was supported by the Russian Foundation for Basic Research (project no. 02-02-16289); the Ministry of Industry, Science, and Technology of the Russian Federation (project nos. 40.020.1.1.1170 and 37.029.1.1.0031); and the DN-ENN-02 Project.

REFERENCES

1. G. S. Vernam, J. Am. Inst. Electr. Eng. **55**, 109 (1926).
2. C. E. Shannon, Bell Syst. Tech. J. **28**, 658 (1949).
3. W. Diffi and M. E. Hellman, IEEE Trans. Inf. Theory **22**, 644 (1976).
4. R. L. Rivest, A. Shamir, and L. Adleman, Commun. ACM **21**, 120 (1978).
5. P. W. Shor, in *Proceedings of 35th Annual Symposium on the Foundations of Computer Science, Santa Fe, NM, USA, 1994* (IEEE Computer Soc. Press, Los Alamitos, CA, 1994), p. 124; quant-ph/9508027.
6. D. Stucki, N. Gisin, O. Guinnard, G. Ribordy, and H. Zbinden, quant-ph/0203118.
7. N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, quant-ph/0101098; Rev. Mod. Phys. **74**, 145 (2002).
8. S. Wiesner, SIGACT News **15**, 78 (1983).
9. C. H. Bennett and G. Brassard, in *Proceedings of IEEE International Conference on Computer Systems and Signal Processing, Bangalore, India* (1984), p. 175.
10. W. K. Wootters and W. H. Zurek, Nature **299**, 802 (1982).
11. C. H. Bennett, Phys. Rev. Lett. **68**, 3121 (1992); C. H. Bennett, G. Brassard, and N. D. Mermin, Phys. Rev. Lett. **68**, 557 (1992).
12. L. Goldenberg and L. Vaidman, Phys. Rev. Lett. **75**, 1239 (1995).
13. S. N. Molotkov and S. S. Nazin, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 767 (2001) [JETP Lett. **73**, 682 (2001)]; S. N. Molotkov, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 79 (2002) [JETP Lett. **76**, 71 (2002)].
14. L. D. Landau and R. Peierls, Z. Phys. **69**, 56 (1931); L. D. Landau, *Collected Works* (Nauka, Moscow, 1969), Vol. 1, p. 56; L. D. Landau and R. Peierls, Z. Phys. **62**, 188 (1930); L. D. Landau, *Collected Works* (Nauka, Moscow, 1969), Vol. 1, p. 33.
15. N. Bohr and L. Rosenfeld, Mat. Fys. Medd. K. Dan. Vidensk. Selsk. **12**, 3 (1933); N. Bohr, *Collected Scientific Works* (Nauka, Moscow, 1969), Vol. 1, p. 39.
16. N. N. Bogolyubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields*, 2nd ed. (Nauka, Moscow, 1973; Wiley, New York, 1980).
17. N. N. Bogolyubov, A. A. Logunov, A. I. Oksak, and I. T. Todorov, *General Principles of Field Quantum Theory* (Nauka, Moscow, 1987).
18. A. M. Jaffe, Phys. Rev. **158**, 1454 (1967).
19. I. M. Gel'fand and N. Ya. Vilenkin, *Generalized Functions*, Vol. 4: *Some Applications of the Harmonic Analysis. The Equipped Hilbert Spaces* (Fizmatgiz, Moscow, 1961; Academic, New York, 1964).
20. Yu. A. Brychkov and A. P. Prudnikov, *Integral Transforms of Generalized Functions* (Nauka, Moscow, 1977; Gordon and Breach, New York, 1989).
21. D. A. Kirzhnits, Usp. Fiz. Nauk **90**, 129 (1966) [Sov. Phys. Usp. **9**, 692 (1966)].
22. N. Wiener and R. Paley, *Fourier Transform in the Complex Domain* (Am. Math. Soc., New York, 1934; Nauka, Moscow, 1964).
23. I. Bialynicki-Birula, Phys. Rev. Lett. **80**, 5247 (1998).
24. T. D. Newton and E. P. Wigner, Rev. Mod. Phys. **21**, 400 (1949).
25. G. C. Hegerfeldt, Phys. Rev. D **10**, 3320 (1974); G. C. Hegerfeldt and S. N. M. Ruijsenaar, Phys. Rev. D **22**, 377 (1980).
26. S. Wickeramasekara and A. Bohm, quant-ph/0302056.
27. D. Slepian and H. O. Pollak, Bell Syst. Tech. J. **40**, 40 (1961).
28. W. H. Fuchs, J. Math. Anal. Appl. **9**, 317 (1964).
29. K. Kraus, *States, Effects and Operations* (Springer, Berlin, 1983).
30. I. Bialynicki-Birula, Prog. Opt. **36**, 245 (1996).
31. D. Mayers and A. Yao, quant-ph/9802025.

32. E. Biham, M. Boyer, P. O. Boykin, T. Mor, and V. Roychowdhury, quant-ph/9912053.
33. Hoi-Kwong Lo and H. F. Chau, quant-ph/9803006.
34. P. W. Shor and J. Preskill, quant-ph/0003004.
35. C. E. Shannon, Bell Syst. Tech. J. **27**, 397 (1948); Bell Syst. Tech. J. **27**, 623 (1948).
36. R. Gallagher, *Information Theory and Reliable Communication* (Wiley, New York, 1968; Sovetskoe Radio, Moscow, 1974).
37. E. J. Mac Williams and N. J. A. Sloane, *The Theory of Error-Correcting Codes* (North-Holland, Amsterdam, 1977).
38. J. Wolfowitz, Ill. J. Math. **1**, 591 (1957).
39. A. S. Kholevo, Probl. Peredachi Inf. **8**, 63 (1972); Probl. Peredachi Inf. **15**, 3 (1979); Usp. Mat. Nauk **53**, 193 (1998); *Introduction to the Quantum Theory of Information* (MTsNMO, Moscow, 2002), Sovrem. Mat. Fiz., No. 5.
40. M. Fleischauer and M. D. Lukin, Phys. Rev. Lett. **84**, 5094 (2000).
41. S. N. Molotkov, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 683 (2002) [JETP Lett. **76**, 584 (2002)]; Pis'ma Zh. Éksp. Teor. Fiz. **77**, 51 (2003) [JETP Lett. **77**, 47 (2003)].

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 859–867.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 963–972.
Original English Text Copyright © 2003 by Kotikov, Parente.

**NUCLEI, PARTICLES,
AND THEIR INTERACTION**

Small-x Behavior of the Slope $d\ln F_2/d\ln(1/x)$ in the Perturbative QCD Framework[¶]

A. V. Kotikov[a,b] and G. Parente[c]
[a]*Bogoliubov Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, Dubna, 141980 Russia*
[b]*Institut für Theoretische Teilchenphysik, Universität Karlsruhe, D-76128 Karlsruhe, Germany*
e-mail: kotikov@thsun1.jinr.ru
[c]*Departamento de Física de Partículas, Universidade de Santiago de Compostela,*
Santiago de Compostela, 15706 Spain
e-mail: gonzalo@gaes.usc.es

Received April 8, 2003

Abstract—In the leading twist approximation of the Wilson operator product expansion, we show that using an analytic parameterization for the behavior of the x slope of the structure function F_2 at small x in perturbative QCD and applying a flat initial condition in the DGLAP evolution equations leads to very good agreement with the new precise deep inelastic scattering experimental data from HERA. © *2003 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Measurements of the deep inelastic scattering structure function F_2 [1–3] and the derivatives $dF_2/d\ln(Q^2)$ [1, 2, 4] and $d\ln F_2/d\ln(1/x)$ [4, 5] at HERA made it possible to access a very interesting kinematical range for testing theoretical ideas on the behavior of quarks and gluons [6] carrying a very low fraction of the proton momentum, the so-called small-x region. In this limit, one expects that nonperturbative effects can make substantial contributions. However, reasonable agreement between the HERA data and the next-to-leading approximation of perturbative QCD has been observed for $Q^2 \geq 2\ \mathrm{GeV}^2$ (see review [7] and references therein), and therefore, perturbative QCD can describe the evolution of F_2 and its derivatives down to very low Q^2 values, traditionally explained by soft processes. It is fundamentally important to find the kinematical region where the well-established perturbative QCD formalism can be safely applied at small x.

The standard program for studying the x behavior of quarks and gluons is carried out by comparison of data with the numerical solution of the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP)[1] equations [8] by fitting the parameters of the x profile of partons at some initial Q_0^2 and the QCD energy scale Λ [10, 11]. However, for analyzing the small-x region exclusively, there is the alternative of a simpler analysis using some of the existing analytic solutions of the DGLAP equations in the small-x limit [12–15]. This was done in [12], where it was pointed out that the HERA small-x data can be interpreted in terms of the so-called doubled asymptotic scaling phenomenon related to the asymptotic behavior of the DGLAP evolution discovered many years ago [16].

The study in [12] was extended in [13–15] to include the finite parts of anomalous dimensions of Wilson operators and Wilson coefficients.[2] This has led to prediction [14, 15] of the small-x asymptotic form of parton distributions in the DGLAP equation starting at some Q_0^2 with the flat function

$$f_a(Q_0^2) = A_a, \tag{1}$$

where f_a are the parton distributions multiplied by x, A_a are unknown parameters to be determined from data, and $a = q, g$ hereafter.

We refer to the approach in [13–15] as the generalized doubled asymptotic scaling approximation. In the generalized doubled asymptotic scaling approximation, the flat initial conditions in Eq. (1) determine the basic role of the singular parts of anomalous dimensions, as in the standard case [12], while the contribution from finite parts of anomalous dimensions and from Wilson coefficients can be considered corrections, which are, however, important for better agreement with experimental data [14]. In the present paper, similarly

---

[¶] This article was submitted by the authors in English.

[1] At small x, there is a different approach based on the Balitsky–Fadin–Kuraev–Lipatov (BFKL) equation [9], whose application is beyond the scope of this paper. However, we sometimes use the BFKL-based predictions below in discussions and for comparison with our results in generalized DAS.

[2] In the standard doubled asymptotic scaling approximation [16] only the singular parts of the anomalous dimensions are used.

1063-7761/03/9705-0859$24.00

to [12–15], we neglect the contribution from the nonsinglet quark component.

The use of the flat initial condition given in Eq. (1) is supported by the actual experimental situation: low-Q^2 data [1, 4, 17, 18] are well described for $Q^2 \leq 0.4$ GeV^2 by the Regge theory with the Pomeron intercept:

$$\alpha_P(0) \equiv \lambda_P + 1 \;=\; 1.08,$$

which is close to the standard one ($\alpha_P(0) = 1$). The small rise observed in the HERA data [1, 2, 4, 18, 19] at low Q^2 can be naturally explained by including higher twist terms (see [15, 20]). Moreover, HERA data [1, 2, 18, 19] with $Q^2 > 1$ GeV^2 are in good agreement with the predictions from the Gluck–Reya–Vogt (GRV) parton densities [11], which supports our aim of developing an analytic form for parton densities at small x because, at least conceptually, our method is very close to the GRV approach.

The purpose of this paper is to extend the study in [14] to compare the predictions from the generalized doubled asymptotic scaling approach with the new precise H1 data [5] for the F_2 slope. The paper is organized as follows. In Section 2, we address the present situation with experimental data for the slope $d\ln F_2/d\ln(1/x)$ and briefly review some approaches to describe them. For completeness, Sections 3 and 4 contain a compilation of the basic formulas in the generalized doubled asymptotic scaling approximation from [14] needed for the present study. In Section 5, we compare our predictions for the derivative $d\ln F_2/d\ln(1/x)$ with the experimental data and discuss the results.

2. THE SLOPE $d\ln F_2/d\ln(1/x)$: EXPERIMENTAL DATA AND QCD PHENOMENOLOGY

Various teams have been able to fit the available data (mostly separating the low- and high-Q^2 regions) using a steep input at small x, $x^{-\lambda}$, $\lambda > 0$. This is clearly different from the flat input in the doubled asymptotic scaling approach of [12–15], which also describes the experimental results reasonably well. In some sense, this is not very surprising because the modern HERA data (at large Q^2) cannot distinguish between the behavior from a steep input parton parameterization at quite large Q^2 and the steep form acquired after dynamical evolution from a flat initial condition at quite low values of Q^2.

Moreover, for the Q^2 evolution based on the full set of anomalous dimensions obtained at $x \longrightarrow 0$ in [21] within the Balitsky–Fadin–Kuraev–Lipatov (BFKL) formalism [9], the results weakly depend on the form of the initial condition [22], preserving steep ones and changing flat ones. In working with anomalous dimensions at a fixed order in α_s, the initial conditions are important when the data are considered in a wide Q^2 range and it is necessary to adequately choose the form of the parton distribution asymptotic form at some Q_0^2.

As discussed in the Introduction, the use of a flat initial condition leads to the (generalized) doubled asymptotic scaling approximation [13–15]. An alternative to this is the choice of a steep initial condition at some sufficiently large Q_c^2,

$$f_a(x, Q_c^2) \propto x^{-\lambda_c}$$

(subscript c stands for constant), which leads to the Q^2-dependence of $f_a(x, Q^2)$ [23–27] given by (if $x^{-\lambda_c} \gg$ const)

$$\frac{f_a(x, Q^2)}{f_a(x, Q_c^2)} \sim \frac{M_a^+(1+\lambda_c, Q^2)}{M_a^+(1+\lambda_c, Q_c^2)}, \tag{2}$$

where $M_a^+(1+\lambda_c, Q^2)$ is the analytic continuation (from integer n to real $1 + \lambda_c$) of the “+” component of the Mellin moment of $f_a(x, Q^2)$,

$$M_a(n, Q^2) \;=\; \int_0^1 dx x^{n-2} f_a(x, Q^2). \tag{3}$$

For $x^{-\lambda_c} \gg$ const, the slope λ_c must be Q^2-independent [24, 25] and the whole Q^2-dependence in $f_a(x, Q^2)$ comes from the factor $M_a^+(1+\lambda_c, Q^2)$ in front of $x^{-\lambda_c}$ in Eq. (2). Approximations similar to Eq. (2) have been successfully applied in studying the Q^2-dependence of HERA data at large Q^2 (see [28] and references therein).

Considering the low-Q^2 region separately, it is also possible to have good agreement between the F_2 data and its Regge-like behavior [4]. Indeed, at $Q^2 \longrightarrow 0$, F_2 can be determined by the relation

$$F_2 \;=\; \frac{Q^2}{4\pi\alpha_{em}}\sigma_{\rm tot}^{\gamma^* p}, \tag{4}$$

where α_{em} is the electromagnetic coupling constant and $\sigma_{\rm tot}^{\gamma^* p}$ is total (virtual) photoproduction cross section.

A large amount of experimental data on hadronic cross sections for many different processes show a universal rise at large energies, which makes it possible to parameterize all these cross sections as the sum of two different components:

$$\sigma_{\rm tot}^{\gamma^* p} \;=\; A_P s^{\alpha_P(0)-1} + A_R s^{\alpha_R(0)-1}, \tag{5}$$

where s is the center-of-mass energy squared. The constants A_P and A_R are process-dependent magnitudes, and the intercepts $a_P(0) \approx 1.08$ and $a_R(0) \approx 0.5$ (see [29]) are universal process-independent constants. The first and second terms in Eq. (5) correspond to (soft) Pomeron and Reggeon exchange, respectively.

From Eqs. (4) and (5), we immediately find that as $Q^2 \longrightarrow 0$,

$$F_2(x, Q^2) \propto x^{-\varepsilon},$$

and hence

$$f_a(x, Q^2) \propto x^{-\varepsilon}, \quad \varepsilon = \alpha_P(0) - 1 \approx 0.08$$

because $s = Q^2/x$ at small x.

There have been many attempts to study the entire Q^2 region in the Regge-asymptotic framework (see, e.g., the reviews in [7]). The reports in [7] contain a great number of models, but we restrict ourselves to only two of them.

In [30], the fit to F_2 experimental data was sought with

$$f_a(x, Q^2) \propto x^{-\lambda(Q^2)}, \tag{6}$$

and rapidly changing $\lambda(Q^2)$ was found in the transition range $Q^2 \sim 5$–10 GeV^2. Unfortunately, it is rather difficult to reconcile the Regge-like behavior given by Eq. (6) with DGLAP evolution in the entire Q^2 range. Some progress along this line achieved in [27] is also based on the flat initial conditions given by Eq. (1). However, the parton distribution structure in [27] is limited by the Regge-like form of Eq. (6), which allows reconciling it with DGLAP evolution only separately at low Q^2, where $\lambda(Q^2)$ is close to 0 (or to ε), and at large Q^2, where $\lambda(Q^2) \sim \lambda_c$. The structure function F_2 and parton distributions were obtained in [27] for the entire Q^2 range only as a combination of these two representations.

For other types of models (see [32, 33]), the phenomenological Q^2-dependence of $\lambda(Q^2)$ is given by

$$\lambda(Q^2) = \varepsilon\left(1 + \frac{Q^2}{Q^2 + c}\right)$$

with a fitted constant c. This produces soft values of the slope $\lambda(Q^2)$ close to ε at low Q^2 and hard ones, $\lambda(Q^2) \sim \lambda_c \sim 0.2$–$0.3$, at $Q^2 \geq 20$ GeV^2.

New precise experimental data on $\lambda(Q^2)$ became available very recently [5]. The H1 data points are shown in Fig. 1, where one can observe that for a fixed Q^2, λ is independent of x in the range $x < 0.01$ within the experimental uncertainties. Indeed, H1 data are well described by the power behavior [5]

$$F_2(x, Q^2) = Cx^{-\lambda(Q^2)}, \tag{7}$$

where

$$\lambda(Q^2) = \hat{a}\ln(Q^2/\Lambda^2)$$

with $C \approx 0.18$, $\hat{a} \approx 0.048$, and $\Lambda = 292$ MeV. The linear rise of λ with $\ln Q^2$ given by Eq. (7) is plotted in Fig. 2.

As a function of x, $\lambda(Q^2)$ was similarly found by the ZEUS Collaboration. As can be seen in Fig. 8 in [4], the ZEUS data for $\lambda(Q^2)$ are compatible with a constant value on the order 0.1 at $Q^2 < 0.6$ GeV^2, as is expected under the assumption of a single soft Pomeron exchange within the framework of Regge phenomenology. In the case of H1, this behavior can also be inferred from the new preliminary H1 data [34] at quite low values of Q^2.

We point out that even though our results obtained in doubled asymptotic scaling approximation (Eqs. (8)–(11) below) do not have an explicit power-like behavior, they actually mimic a power-law shape over a limited region of x and Q^2 (see Section 4). In addition, we observed earlier [14] that in the generalized doubled asymptotic scaling approximation, the x dependence of the effective slopes is not strong and the F_2 effective slope is in good agreement with old (less precise) H1 data [1]. In Section 5, we repeat the analysis performed in [14], but with the new precise H1 data for the slope [5].

3. Q^2 DEPENDENCE OF F_2 AND PARTON DISTRIBUTIONS IN THE GENERALIZED DOUBLED ASYMPTOTIC SCALING APPROXIMATION

We briefly recall the results of the generalized doubled asymptotic scaling approximation first presented in [14]. The small-x behavior of the parton densities and F_2 in the next-to-leading order approximation is given by[3]

$$f_a(z, Q^2) = f_a^+(z, Q^2) + f_a^-(z, Q^2),$$

$$f_a^-(z, Q^2) \sim \exp(-d_-(1)s - D_-(1)p) + O(z), \tag{8}$$

$$f_g^+(z, Q^2) \sim I_0(\sigma)\exp(-\bar{d}_+(1)s - \bar{D}_+(1)p) + O(\rho), \tag{9}$$

$$\begin{aligned} f_q^+(z, Q^2) &\sim f_g^+(z, Q^2) \\ &\times\left[(1 - \bar{d}^q_{+-}(1)a_s(Q^2))\frac{\rho I_1(\sigma)}{I_0(\sigma)} + 20a_s(Q^2)\right] + O(\rho), \end{aligned} \tag{10}$$

$$F_2(z, Q^2) = e\left(f_q(z, Q^2) + \frac{2}{3}fa_s(Q^2)f_g(z, Q^2)\right), \tag{11}$$

---

[3] Hereafter, $z = x/x_0$, where x_0 is a free parameter that limits the applicability range of Eqs. (8)–(11) and can be fitted from experimental data together with the magnitudes of gluon and sea quark distributions at Q_0^2. As shown in [14], the fits to the F_2 HERA data depend very slightly on the specific x_0 value.

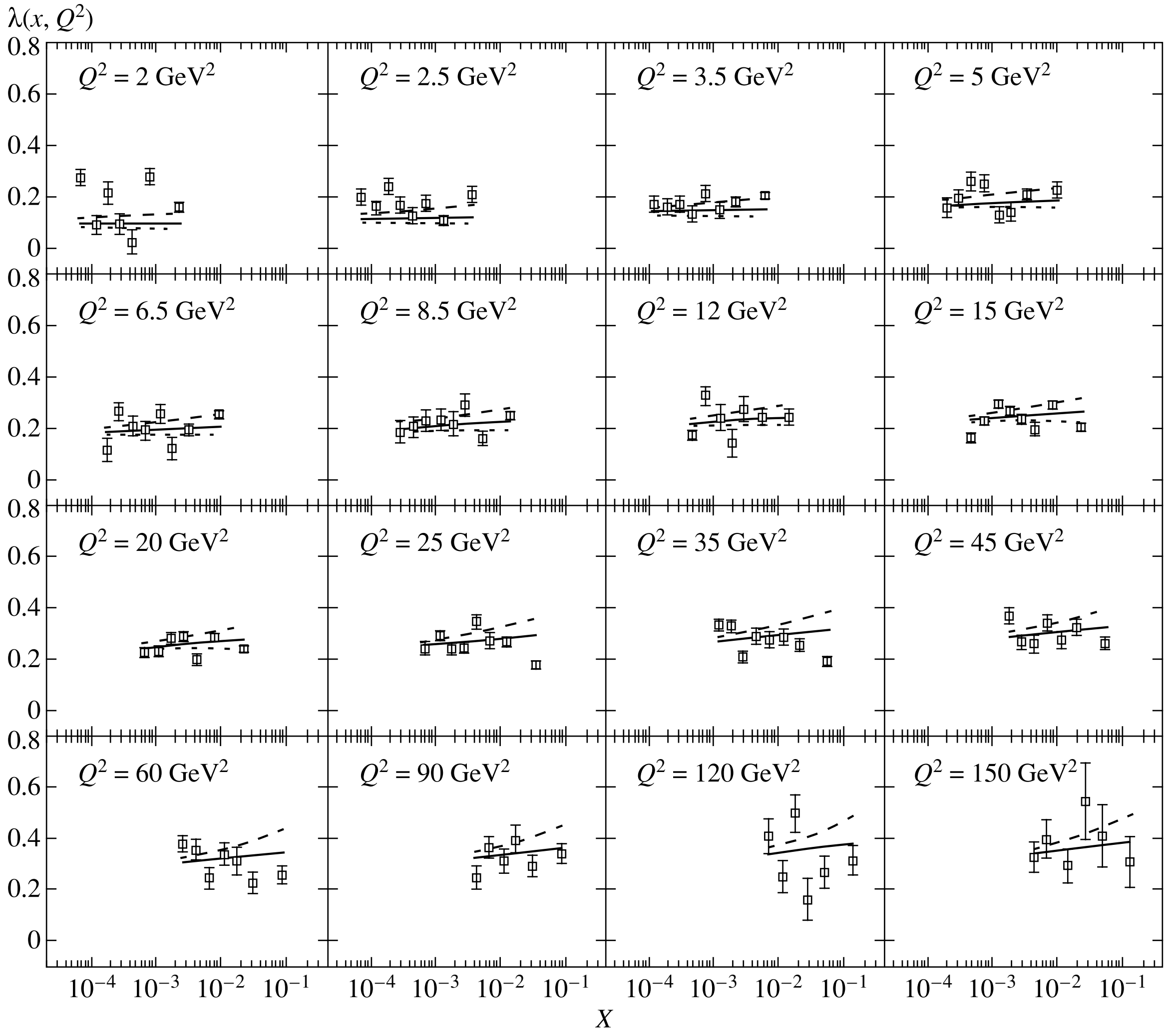


Fig. 1. The derivative $d\ln F_2/d\ln(1/x)$ (the effective slope λ) as a function of x for different values of Q^2. Data points are from H1 [5]. Only statistical uncertainties are shown. The solid line is the result of a fit using $\lambda^{\text{eff}}_{F2}$ in Eq. (15) with fixed Q_0^2 = 1 GeV² and $x_0 = 1$. The dotted line is the same but with the parameters from a fit to the F_2 data in [14]. The dashed line corresponds to the asymptotic expression $\lambda^{\text{eff},\,as}_{F2}$ in Eq. (16).

where

$$e = \frac{\sum_i^f e_i^2}{f}$$

is the average charge square of f effective quarks,

$$a_s = \alpha_s/4\pi,$$

$$s = \ln\left(\frac{a_s(Q_0^2)}{a_s(Q^2)}\right), \quad p = a_s(Q_0^2) - a_s(Q^2),$$

$$D_\pm = d_{\pm\pm} - \frac{\beta_1}{\beta_0}d_\pm, \quad \sigma = 2\sqrt{(\hat{d}_+s + \hat{D}_+p)\ln z}, \qquad (12)$$

$$\rho = \sqrt{\frac{(\hat{d}_+s + \hat{D}_+p)}{\ln z}} = \frac{\sigma}{2\ln(1/z)},$$

and β_0 and β_1 are the first two terms of the QCD β-function.

The components of the leading order anomalous dimension $d_-(n)$ and the singular ($\hat{d}_+$) and regular ($\bar{d}_+(n)$) parts of the leading order anomalous dimension $d_+(n) = \hat{d}_+/(n-1) + \bar{d}_+(n)$ are given by (at $n \longrightarrow 1$)

$$\hat{d}_+ = -\frac{12}{\beta_0}, \quad \bar{d}_+(1) = 1 + \frac{20f}{27\beta_0}, \qquad (13)$$
$$d_-(1) = \frac{16f}{27\beta_0}.$$

The corresponding components in the next-to-leading order can be represented as

$$\hat{d}_{++} = \frac{412}{27\beta_0}f, \quad \hat{d}^q_{+-} = -20, \quad \hat{d}^g_{+-} = 0,$$

$$\bar{d}_{++}(1) = \frac{8}{\beta_0}\left(36\zeta_3 + 33\zeta_2 - \frac{1643}{12}\right.$$
$$\left. + \frac{2}{9}f\left[\frac{68}{9} - 4\zeta_2 - \frac{13}{243}f\right]\right),$$

$$\bar{d}^q_{+-}(1) = 23 - 12\zeta_2 - \frac{13}{81}f, \quad \bar{d}^g_{+-}(1) = \frac{80}{81}f, \tag{14}$$

$$d_{--}(1) = \frac{16}{9\beta_0}\left(2\zeta_3 - 3\zeta_2 + \frac{13}{4}\right.$$
$$\left. + f\left[4\zeta_2 - \frac{23}{18} + \frac{13}{243}f\right]\right),$$

$$d^q_{-+}(1) = 0, \quad d^g_{-+}(1) = -3\left(1 + \frac{f}{81}\right).$$

Some interesting features of the results in Eqs. (8)–(12) are summarized below.

(1) Both the gluon and quark singlet densities given above are presented in terms of two components ("+" and "–") obtained from the analytic Q^2-dependent expressions of the corresponding ("+" and "–") components of parton distribution moments.

(2) The "–" component is constant at small x, whereas the "+" component increases at $Q^2 \geq Q_0^2$ as $\exp\sigma$, where σ contains the positive leading order term $|\hat{d}_+|\, s\ln(1/z)$ and the negative next-to-leading order one $|\hat{D}_+|\, p\ln(1/z)$ (see Eq. (12)). The most important part of the next-to-leading order corrections (i.e., the singular part at $x \longrightarrow 0$) is therefore properly taken into account: it directly enters the argument of the Bessel functions and does not spoil the applicability of perturbation theory at low values of x.

4. Q^2 DEPENDENCE OF THE SLOPE $d\ln F_2/d\ln(1/x)$ IN THE GENERALIZED DOUBLED ASYMPTOTIC SCALING APPROXIMATION

The behavior of the parton densities and the structure function F_2 within the generalized doubled asymptotic scaling approach, given by Eqs. (8)–(11), can be represented by a power-law shape over a limited region of x and Q^2 [14, 15],

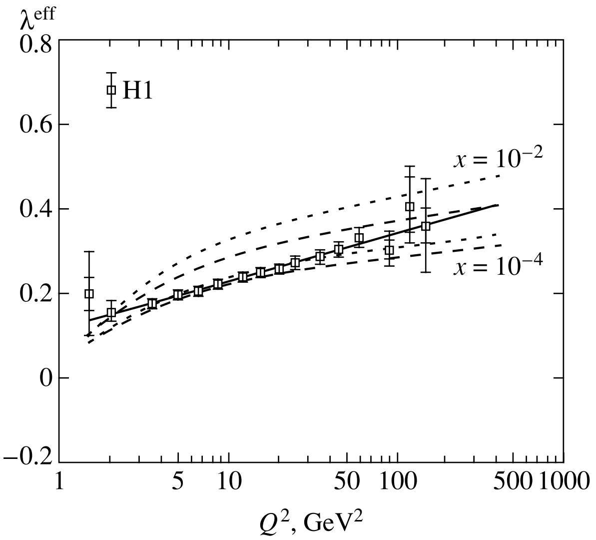


Fig. 2. The derivative $d\ln F_2/d\ln(1/x)$ (the effective slope λ) as a function of Q^2. Data points are from H1 [5]. Outer error bars include statistical and systematical errors added in quadrature. Inner bars correspond to statistical errors. The solid line corresponds to the H1 parameterization [5] given in Eq. (7). Dotted and dashed curves are produced as in Fig. 1. For the lower (upper) curves, the value $x = 10^{-4}$ ($x = 10^{-2}$) was used.

$$f_a(x, Q^2) \propto x^{-\lambda_a^{\rm eff}(x,\,Q^2)}, \quad F_2(x, Q^2) \propto x^{-\lambda_{F2}^{\rm eff}(x,\,Q^2)}.$$

Since

$$\frac{d}{d\ln x} = \frac{d}{d\ln z},$$

the effective slopes can be obtained directly from Eqs. (8)–(11) as

$$\lambda_g^{\rm eff}(z, Q^2) = \frac{f_g^+(z, Q^2)}{f_g(z, Q^2)}\rho\frac{I_1(\sigma)}{I_0(\sigma)},$$

$$\lambda_q^{\rm eff}(z, Q^2) = \frac{f_q^+(z, Q^2)}{f_q(z, Q^2)}\rho\frac{I_2(\sigma)(1 - \bar{d}^q_{+-}(1)a_s(Q^2)) + 20a_s(Q^2)I_1(\sigma)/\rho}{I_1(\sigma)(1 - \bar{d}^q_{+-}(1)a_s(Q^2)) + 20a_s(Q^2)I_0(\sigma)/\rho}, \tag{15}$$

$$\lambda_{F2}^{\rm eff}(z, Q^2) = \frac{\lambda_q^{\rm eff}(z, Q^2)f_q^+(z, Q^2) + (2f)/3a_s(Q^2)\lambda_g^{\rm eff}(z, Q^2)f_g^+(z, Q^2)}{f_q(z, Q^2) + (2f)/3a_s(Q^2)f_g(z, Q^2)}.$$

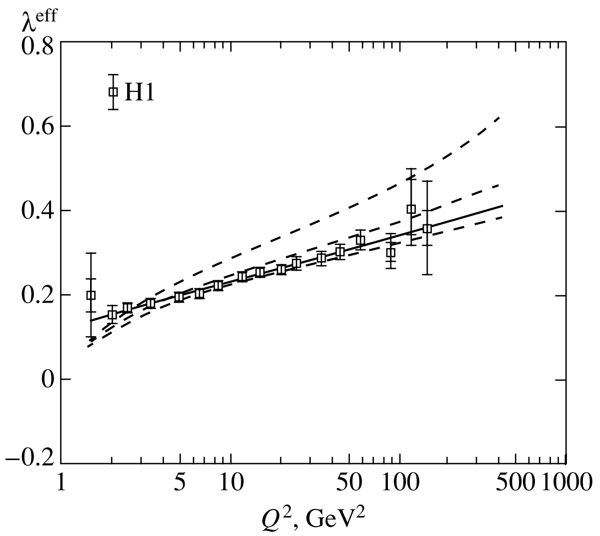


Fig. 3. The derivative $d\ln F_2/d\ln(1/x)$ (the effective slope λ) as a function of Q^2. Data points are from H1 [5]. Error bars and solid line are as in Fig. 2. The dashed lines were calculated with Eq. (16) using $x = a \times 10^{-4}Q^2$ with $a = 0.1$, 1, and 10. Upper curves correspond to larger x.

We emphasize that the gluon effective slope λ_g^{eff} obtained from Eq. (15) is larger than the quark slope λ_q^{eff} [14], which is in excellent agreement with Martin–Stirling–Roberts [35] and Gluck–Reya–Vogt [11] analyses (see also [10]).

On the other hand, the effective slopes λ_a^{eff} and $\lambda_{F2}^{\text{eff}}$ in Eq. (15) depend on the magnitudes A_a of the initial parton distribution and on the chosen input values Q_0^2 and Λ. However, at quite large Q^2, where the "–" component is negligible, the dependence on the initial parton distribution disappears and the asymptotic behavior is then given by[4]

$$\lambda_g^{\text{eff},as}(z, Q^2) = \rho\frac{I_1(\sigma)}{I_0(\sigma)} \approx \rho - \frac{1}{4\ln(1/z)},$$

$$\lambda_q^{\text{eff},as}(z, Q^2) = \rho$$

$$\times\frac{I_2(\sigma)(1-\bar{d}^q_{+-}(1)a_s(Q^2)) + 20a_s(Q^2)I_1(\sigma)/\rho}{I_1(\sigma)(1-\bar{d}^q_{+-}(1)a_s(Q^2)) + 20a_s(Q^2)I_0(\sigma)/\rho}$$

$$\approx \rho - \frac{3}{4\ln(1/z)} + \frac{10a_s(Q^2)}{\rho\ln(1/z)}, \qquad (16)$$

$$\lambda_{F2}^{\text{eff},as}(z, Q^2) = \lambda_q^{\text{eff},as}(z, Q^2)$$

$$\times\frac{1 + 6a_s(Q^2)/\lambda_q^{\text{eff},as}(z, Q^2)}{1 + 6a_s(Q^2)/\lambda_g^{\text{eff},as}(z, Q^2)} + O(a_s^2(Q^2))$$

$$\approx \lambda_q^{\text{eff},as}(z, Q^2) + \frac{3a_s(Q^2)}{\rho\ln(1/z)},$$

where the symbol ≈ denotes that an approximation was made in the expansion of the modified Bessel functions $I_n(\sigma)$ ($n = 0, 1, 2$). These approximations are accurate only at large values of σ (i.e., at large Q^2 and/or small x).

Finally, we note that, to the leading order, the F_2 slope, $\lambda_{F2}^{\text{eff},as}$ is equal to the quark slope $\lambda_q^{\text{eff},as}$ and it coincides with the result in [36] for very large values of σ and a flat input (see also the first paper in [7]). To the next-to-leading order, $\lambda_{F2}^{\text{eff},as}$ lies between the quark and gluon slopes but closer to the former (see Fig. 3 in [14]).

5. COMPARISON WITH EXPERIMENTAL DATA

Using the results of the previous section, we analyzed the HERA data from the H1 Collaboration[5] [5] for the slope $d\ln F_2/d\ln(1/x)$ at small x.

Initially, our results for $\lambda_{F2}^{\text{eff}}$ depend on the five parameters Q_0^2, x_0, A_q, A_g, and $\Lambda_{\overline{MS}}$ ($f = 4$). In our previous paper [14], we fixed $\Lambda_{\overline{MS}}$ ($f = 4$) = 250 MeV, which was a reasonable value extracted from the traditional (higher x) experiments. All other parameters were fitted and good agreement with the F_2 HERA data was achieved for $Q_0^2 \sim 1$ GeV^2 (all results depend on x_0 very slightly).

In this paper, we take $\Lambda_{\overline{MS}}$ ($f = 4$) = 292 MeV, in agreement with the more recent H1 results [5] and other analyses (see [37] and references therein), and we directly fit the data on the slope $d\ln F_2/d\ln(1/x)$ using Eq. (15). The result is shown in Fig. 1. For comparison, we also plot the curves from a fit to the F_2 data in [14], where the value 250 MeV was used. The results are very similar and demonstrate the very important feature

---

[4] The asymptotic formulas in Eq. (16) work quite well at any values of $Q^2 \geq Q_0^2$, because the values of λ_a^{eff} and $\lambda_{F2}^{\text{eff}}$ are equal zero at $Q^2 = Q_0^2$. The use of approximations in Eq. (16) instead of the exact results in Eq. (15) underestimates (overestimates) the gluon (quark) slope at $Q^2 \geq Q_0^2$ only slightly. For F_2, the similarity of the values of $\lambda_{F2}^{\text{eff}}$ and $\lambda_{F2}^{\text{eff},as}$ are as is shown in Fig. 1.

[5] In this paper, we only use the H1 data [5]. The preliminary ZEUS data for the slope $d\ln F_2/d\ln(1/x)$ are only available through points in Figs. 8 and 9 of [4]. They shown quite similar properties in comparison with the H1 data [5]. Unfortunately, the ZEUS numerical values are as yet unavailable and we cannot analyze them in the present paper.

of an approximate x-independence of $\lambda_{F2}^{\text{eff}}$ given by Eq. (15), which is in agreement with the H1 data [5].

Figure 1 also gives the asymptotic values for the slope $\lambda_{F2}^{\text{eff},\,as}$ obtained from Eq. (16). The agreement with the data and with the other curves is also rather good if we take into consideration that no fit is involved in this case because the only parameters entering Eq. (16) are the fixed values $\Lambda_{\overline{MS}}$ $(f = 4) = 292$ MeV and $x_0 = 1$.

Thus, the extremely weak x-dependence of the slope $\lambda(Q^2)$ in the considered region of x and Q^2 supports the possibility of successfully using our generalized doubled asymptotic scaling approximation in the x-independent analysis of the F_2 slope.

Figure 2 shows the experimental data for $\lambda_{F2}^{\text{eff}}$ and the corresponding H1 parameterization [5] written above in Eq. (7). We also plotted the result from Eqs. (16) and (15) using the parameters from our previous paper [14] as in Fig. 1. In both cases, we give it for two representative values of x.

Visual inspection of Fig. 1 shows that the boundaries and mean values of the experimental x ranges [5] increase proportionally with Q^2, which is related to the kinematical restrictions $x \sim 10^{-4}Q^2$ in the HERA experiments (see [1–3, 18] and, e.g., Fig. 1 in [4]).

Figure 3 shows the H1 experimental data [5] for $\lambda_{F2}^{\text{eff}}$ and the H1 parameterization (Eq. (7)) as in Fig. 2, but this time in comparison to the asymptotic values $\lambda_{F2}^{\text{eff},\,as}$ calculated from Eq. (16) using $x = a \times 10^{-4}Q^2$ with a = 0.1, 1, and 10. There is reasonable agreement with the H1 data for $Q^2 > 2$ GeV^2 with a between 0.1 and 1 (the two lower dashed curves in Fig. 3), which approximately corresponds to the middle points of the measured x range.

6. CONCLUSIONS

We have studied the Q^2 dependence of the slope

$$\lambda_{F2}^{\text{eff}} = d\ln F_2/d\ln(1/x)$$

at small x in the framework of perturbative QCD. Our results are in good agreement with the new precise experimental H1 data [5] at $Q^2 \geq 2$ GeV^2, where perturbation theory can be applicable.

Although our approach, which can be called the generalized doubled asymptotic scaling approximation, is based on pure perturbative grounds, flat initial conditions at $Q_0^2 \approx 1$ GeV^2, dynamical evolution to $Q^2 \geq Q_0^2$, and is conceptually very close to the GRV approach but involves the exact analytic Q^2 evolution, it can be reasonably applied for the new precise data of the slope $\lambda_{F2}^{\text{eff}}$.

The agreement between $\lambda(Q^2)$ data and perturbative QCD has already been observed by the H1 [2] and ZEUS [4] collaborations. The obtained linear rise of $\lambda(Q^2)$ with $\ln Q^2$ (see, e.g., Figs. 2 and 3), parameterized by H1 as in Eq. (7), can naively be interpreted in a strongly nonperturbative way, i.e.,

$$\lambda(Q^2) \propto \frac{1}{\alpha_s(Q^2)}.$$

Our analysis, however, demonstrates that the rise can be explained as being proportional to $\ln\ln Q^2$, which is natural in perturbative QCD at low x (see [12–16] and references therein): when the coupling constant is running, the renormalization group leads to the small-x behavior of the parton distribution proportional to $\ln(\alpha_s(Q^2))$ in the leading order of perturbation theory and proportional to $\alpha_s(Q^2)$ in the next-to-leading order (see Eqs. (8)–(12) and discussions after Eq. (14)).

The good agreement between the perturbative QCD and the experiments obtained here and in [14, 15] demonstrates that for $Q^2 > 2$ GeV^2, nonperturbative contributions, such as shadowing effects [38], higher twist effects [39], and others, either are quite small (see also [40] and references therein) or cancel between themselves and/or with $\ln(1/x)$ terms contained in higher orders of perturbation theory. We note, however, that higher twist corrections are important at $Q^2 \leq 1$ GeV^2, as has been demonstrated in [15, 20, 37]. Further efforts in the development of theoretical approaches are needed to isolate the correct contributions from nonperturbative dynamics and higher orders containing strong $\ln(1/x)$ terms.

Moreover, the good agreement between perturbative QCD and experimental data at low Q^2 can be explained by a larger effective scale for the QCD coupling constant [14, 15]. A similar behavior has already been observed in the framework of perturbative QCD [41] and in BFKL-motivated approaches [42–44] (see the recent review in [45] and discussions therein).

We note that large next-to-leading order corrections calculated recently in the BFKL framework [46] (see also [47]) lead to strong suppression of the leading order BFKL results for a high-energy asymptotic behavior of the cross section (see, e.g., [42, 43]). Careful inclusion of next-to-leading order corrections leads to results that are quite close to those obtained in pure perturbative QCD [43]. This can give additional support to the good applicability of perturbation theory in the small-x range, where, as expected before, nonperturbative effects should make a substantial contribution.

As the next step, it could be very useful to apply the generalized doubled asymptotic scaling approach to perform a combined analysis of the HERA data for F_2, $dF_2/d\ln(Q^2)$, $d\ln F_2/d\ln(1/x)$, and F_L. We hope to consider this in a forthcoming paper, including higher twist corrections in the Q^2-evolution approach given by Eqs. (8)–(11). It would also be interesting to consider additional terms in the initial condition given by Eq. (1) proportional to $\ln(1/x)$ and $\ln^2(1/x)$.

We hope that this analysis will be relevant in finding the kinematical region where the well-established perturbative QCD formalism can be safely applied at small x. Moreover, the study should clear up the reason of the good agreement between the small-x relation of F_L, F_2, and $dF_2/d\ln(Q^2)$ obtained in pure perturbative QCD in [48] (based on previous works [26, 49]), the experimental data for these structure functions [2, 50], and the predictions of [51] in the framework of k_t-factorization [21, 22, 52].

ACKNOWLEDGMENTS

This paper was supported in part by the Russian Foundation for Basic Research (project no. 02-02-17513), INTAS (grant no. 366), and by the Heisenberg–Landau program. One of the authors (A.V.K.) was supported in part by an Alexander von Humboldt fellowship during his visit to Karlsruhe University, where the paper was completed. G. P. acknowledges the support of Galician research funds (PGIDT00 PX20615PR) and Spanish CICYT (FPA2002-01161).

REFERENCES

1. C. Adloff *et al.* (H1 Collaboration), Nucl. Phys. B **497**, 3 (1997).
2. C. Adloff *et al.* (H1 Collaboration), Eur. Phys. J. C **21**, 33 (2001).
3. S. Chekanov *et al.* (ZEUS Collaboration), Eur. Phys. J. C **21**, 443 (2001).
4. B. Surrow *et al.* (ZEUS Collaboration), hep-ph/0201025.
5. C. Adloff *et al.* (H1 Collaboration), Phys. Lett. B **520**, 183 (2001).
6. F. J. Yndurain, *The Theory of Quark and Gluon Interactions*, 3rd ed. (Springer, Berlin, 1999).
7. A. M. Cooper-Sarkar, R. C. E. Devenish, and A. De Roeck, Int. J. Mod. Phys. A **13**, 3385 (1998); C. Merino, in *Proceedings of the Madagascar High-Energy Physics International Conference (HEP-MAD'01), Antananarivo, Madagascar,* Ed. by S. Narison (World Sci., 2001); hep-ph/0201088.
8. V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972); Sov. J. Nucl. Phys. **15**, 675 (1972); L. N. Lipatov, Sov. J. Nucl. Phys. **20**, 94 (1975); G. Altarelli and G. Parisi, Nucl. Phys. B **126**, 298 (1977); Yu. L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977).
9. L. N. Lipatov, Sov. J. Nucl. Phys. **23**, 338 (1976); E. A. Kuraev, L. N. Lipatov, and V. S. Fadin, Phys. Lett. B **60**, 50 (1975); Sov. Phys. JETP **44**, 443 (1976); Sov. Phys. JETP **45**, 199 (1977); Ya. Ya. Balitzki and L. N. Lipatov, Sov. J. Nucl. Phys. **28**, 822 (1978); L. N. Lipatov, Sov. Phys. JETP **63**, 904 (1986).
10. A. D. Martin, W. S. Stirling, R. G. Roberts, and R. S. Thorne, Eur. Phys. J. C **23**, 73 (2002); J. Pumplin *et al.* (CTEQ Collaboration), Preprint No. MSU-HEP-011101; hep-ph/0201195.
11. M. Gluck, E. Reya, and A. Vogt, Eur. Phys. J. C **5**, 461 (1998).
12. R. D. Ball and S. Forte, Phys. Lett. B **336**, 77 (1994).
13. L. Mankiewicz, A. Saalfeld, and T. Weigl, Phys. Lett. B **393**, 175 (1997).
14. A. V. Kotikov and G. Parente, Nucl. Phys. B **549**, 242 (1999); Nucl. Phys. A (Proc. Suppl.) **99**, 196 (2001); hep-ph/0010352; in *Proceedings of the International Conference PQFT98* (Dubna, 1998); hep-ph/9810223; in *Proceedings of the 8th International Workshop on Deep Inelastic Scattering, DIS 2000, Liverpool* (2000), p. 198; hep-ph/0006197.
15. A. V. Kotikov and G. Parente, in *Proceedings of 15th International Seminar on Relativistic Nuclear Physics and Quantum Chromodynamics* (Dubna, 2000); hep-ph/0012299; in *Proceedings of the 9th International Workshop on Deep Inelastic Scattering, DIS 2001, Bologna* (2001); hep-ph/0106175; in *Proceedings of 16th Baldin International Seminar on Relativistic Nuclear Physics and Quantum Chromodynamics, Dubna* (2002); hep-ph/0304064.
16. A. De Rújula *et al.*, Phys. Rev. D **10**, 1649 (1974).
17. M. Arneodo *et al.* (NM Collaboration), Phys. Lett. B **364**, 107 (1995); Nucl. Phys. B **483**, 3 (1997); M. R. Adams *et al.* (E665 Collaboration), Phys. Rev. D **54**, 3006 (1996); A. Donnachie and P. V. Landshoff, Nucl. Phys. B **244**, 669 (1984); Nucl. Phys. B **267**, 690 (1986); Z. Phys. C **61**, 134 (1994).
18. J. Breitweg *et al.* (ZEUS Collaboration), Phys. Lett. B **407**, 432 (1997).
19. J. Breitweg *et al.* (ZEUS Collaboration), Phys. Lett. B **487**, 53 (2000); Eur. Phys. J. C **21**, 443 (2001).
20. A. Yu. Illarionov, A. V. Kotikov, and G. Parente, private communication.
21. M. Ciafaloni, Nucl. Phys. B **296**, 249 (1987); S. Catani, F. Fiorani, and G. Marchesini, Phys. Lett. B **234**, 389 (1990); Nucl. Phys. B **336**, 18 (1990); S. Catani, F. Fiorani, G. Marchesini, and G. Oriani, Nucl. Phys. B **361**, 645 (1991).
22. S. Catani, hep-ph/9609237; in *Proceedings of the International Workshop on Deep Inelastic Scattering and Related Phenomena, Rome* (1996), p. 454; Z. Phys. C **75**, 665 (1997).
23. C. Lopez and F. J. Yndurain, Nucl. Phys. B **171**, 231 (1980); Nucl. Phys. B **183**, 157 (1981).
24. V. I. Vovk, A. V. Kotikov, and S. I. Maximov, Theor. Math. Phys. **84**, 744 (1990); L. L. Jenkovszky, A. V. Kotikov, and F. Paccanoni, Sov. J. Nucl. Phys. **55**, 1224 (1992); JETP Lett. **58**, 163 (1993); Phys. Lett. B **314**, 421 (1993); A. V. Kotikov, S. I. Maximov, and I. S. Parobij, Theor. Math. Phys. **111**, 442 (1997).
25. A. V. Kotikov, Phys. At. Nucl. **56**, 1276 (1993).
26. A. V. Kotikov, Phys. At. Nucl. **57**, 133 (1994); Phys. Rev. D **49**, 5746 (1994).

27. A. V. Kotikov, Mod. Phys. Lett. A **11**, 103 (1996); Phys. At. Nucl. **59**, 2137 (1996).
28. C. Lopez, F. Barreiro, and F. J. Yndurain, Z. Phys. C **72**, 561 (1996); K. Adel, F. Barreiro, and F. J. Yndurain, Nucl. Phys. B **495**, 221 (1997).
29. A. Donnachie and P. V. Landshoff, Phys. Lett. B **296**, 227 (1992); Phys. Lett. B **437**, 408 (1998).
30. H. Abramowitz, E. M. Levin, A. Levy, and U. Maor, Phys. Lett. B **269**, 465 (1991); A. Levy, Preprint No. 95-003 (DESY, 1995); hep-ph/9501346.
31. G. M. Frichter, D. W. McKay, and J. P. Ralston, Phys. Rev. Lett. **74**, 1508 (1995).
32. A. Capella, A. B. Kaidalov, C. Merino, and J. Tran Thanh Van, Phys. Lett. B **337**, 358 (1994); A. B. Kaidalov, C. Merino, and D. Pertermann, Eur. Phys. J. C **20**, 301 (2001).
33. P. Desgrolard, L. L. Jenkovszky, and F. Paccanoni, Eur. Phys. J. C **7**, 655 (1999).
34. T. Lastovicka *et al.* (H1 Collaboration), in *Proceedings of the International Workshop on Deep Inelastic Scattering, Cracow* (2002); J. Gayler *et al.* (H1 Collaboration), in *Proceedings of the International Workshop on Deep Inelastic Scattering, Cracow* (2002).
35. A. D. Martin, W. S. Stirling, and R. G. Roberts, Phys. Lett. B **387**, 419 (1996).
36. H. Navelet, R. Peshanski, Ch. Royon, *et al.*, Mod. Phys. Lett. A **12**, 857 (1997).
37. V. G. Krivokhijine and A. V. Kotikov, Preprint No. E2-2001-190 (JINR, 2001); hep-ph/0108224; in *Proceedings of the 16th International Workshop on High Energy Physics and Quantum Field Theory, Moscow* (2001); hep-ph/0206221; Acta Phys. Slov. **52**, 227 (2002); Acta Phys. Pol. B **33**, 2947 (2002); Nucl. Instrum. Methods Phys. Res. A **502**, 624 (2003).
38. E. M. Levin, hep-ph/9706341; hep-ph/9706448; in *Proceedings of the International Workshop on Deep Inelastic Scattering, Chicago* (1997).
39. J. Bartels and C. Bontus, in *Proceedings of the International Workshop on Deep Inelastic Scattering, Chicago* (1997); A. D. Martin and R. S. Thorne, Preprint No. DTP/98/04; hep-ph/9802366; J. Bartels, in *Proceedings of the International Workshop on Deep Inelastic Scattering, Cracow* (2002).
40. J. Bartels, K. Golec-Biernat, and K. Peters, Eur. Phys. J. C **17**, 121 (2000); K. Golec-Biernat, in *Proceedings of the International Workshop on Deep Inelastic Scattering, Cracow* (2002).
41. Yu. L. Dokshitzer and D. V. Shirkov, Z. Phys. C **67**, 449 (1995); A. V. Kotikov, JETP Lett. **59**, 1 (1994); Phys. Lett. B **338**, 349 (1994); W. K. Wong, Phys. Rev. D **54**, 1094 (1996).
42. S. J. Brodsky, V. S. Fadin, V. T. Kim, *et al.*, JETP Lett. **70**, 155 (1999); V. T. Kim, L. N. Lipatov, and G. B. Pivovarov, in *Proceedings of the 8th Blois Workshop at IHEP, Protvino* (1999); Preprint No. IITAP-99-013; hep-ph/9911228; in *Proceedings of the Symposium on Multiparticle Dynamics (ISMD99), Providence, Rhode Island* (1999), Preprint No. IITAP-99-014; hep-ph/9911242.
43. S. J. Brodsky, V. S. Fadin, V. T. Kim, *et al.*, in *Proceedings of the PHOTON2001, Ascona, Switzerland* (2001), Preprint CERN-TH/2001-341, SLAC-PUB-9069; hep-ph/0111390.
44. M. Ciafaloni, D. Colferai, and G. P. Salam, Phys. Rev. D **60**, 114036 (1999); J. High Energy Phys. **07**, 54 (2000); R. S. Thorne, Phys. Lett. B **474**, 372 (2000); Phys. Rev. D **60**, 054031 (1999); Phys. Rev. D **64**, 074005 (2001); G. Altarelli, R. D. Ball, and S. Forte, Nucl. Phys. B **621**, 359 (2002).
45. Bo Andersson *et al.*, Eur. Phys. J. C **25**, 77 (2002).
46. V. S. Fadin and L. N. Lipatov, Phys. Lett. B **429**, 127 (1998); G. Camici and M. Ciafaloni, Phys. Lett. B **430**, 349 (1998).
47. A. V. Kotikov and L. N. Lipatov, Nucl. Phys. B **582**, 19 (2000); in *Proceedings of the 35th Winter School, Repino, St. Petersburg* (2001); hep-ph/0112346; Nucl. Phys. B **661**, 19 (2003).
48. A. V. Kotikov, JETP **80**, 979 (1995); A. V. Kotikov and G. Parente, in *Proceedings of the International Workshop on Deep Inelastic Scattering and Related Phenomena, Rome* (1996), p. 237; hep-ph/9608409; Mod. Phys. Lett. A **12**, 963 (1997); JETP **85**, 17 (1997).
49. A. V. Kotikov, JETP Lett. **59**, 667 (1994); A. V. Kotikov and G. Parente, Phys. Lett. B **379**, 195 (1996).
50. N. Gogitidze *et al.* (H1 Collaboration), J. Phys. G **28**, 751 (2002).
51. A. V. Kotikov, A. V. Lipatov, and N. P. Zotov, Eur. Phys. J. C **27**, 219 (2003).
52. S. Catani, M. Ciafaloni, and F. Hautmann, Phys. Lett. B **307**, 197 (1993); S. Catani and F. Hautmann, Phys. Lett. B **315**, 157 (1993); Nucl. Phys. B **427**, 472 (1994).

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 868–874.
Translated from Zhurnal Éksperimental’noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 973–980.
Original Russian Text Copyright © 2003 by Anisimov, Akhmedzhanov, Zelensky, Kuznetsova.

Influence of Transverse Magnetic Fields and Depletion of Working Levels on the Nonlinear Resonance Faraday Effect

P. M. Anisimov, R. A. Akhmedzhanov, I. V. Zelensky*, and E. A. Kuznetsova

Institute of Applied Physics, Russian Academy of Sciences, Nizhni Novgorod, 603950 Russia

**e-mail: zelensky@appl.sci-nnov.ru*

Received February 18, 2003

Abstract—The nonlinear resonance Faraday effect is studied under the condition of coherent population trapping in ^{87}Rb vapor at the D_1-line $F = 2 \longrightarrow F' = 1$ transition. The influence of transverse magnetic fields on the nonlinear optical Faraday rotation is studied. For the transverse fields perpendicular to the electromagnetic-wave polarization, a simple theoretical model is proposed, which is in good agreement with experimental data. The optimal intensity providing the maximum sensitivity is found based on the results obtained. The influence of working-level depletion on the parameters of Faraday rotation in open systems is studied experimentally and theoretically. The system was closed in the experiment by using an additional laser to increase the sensitivity and extend the dynamic range of measured fields. The importance of compensating for the depletion in the presence of spurious magnetic fields is shown; in particular, the sensitivity was enhanced by a factor of 50 in experiments with a buffer gas. © *2003 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Coherent population trapping (CPT) and electromagnetically induced transparency belong to resonance interference phenomena, which are currently being extensively studied. A strong dispersion under the conditions of electromagnetically induced transparency, along with the suppression of absorption and resonantly enhanced nonlinearity, makes CPT promising for solving various problems of physical optics and opens up numerous possibilities for potential applications [1–3]. In particular, the possibility of using CPT and electromagnetically induced transparency in degenerate systems [4–6] for the development of a new class of optical magnetometers is being actively discussed [7–16].

The problem of precision measurement of a magnetic field is very important from the practical point of view for many studies and applications. Record sensitivities (~10 pG/Hz$^{1/2}$) are now achieved using SQUID technology; however, the necessity of cryogenic cooling substantially restricts the range of its applications. Optical magnetometers often prove to be more convenient for use in various studies. The use of CPT promises a substantial increase in the sensitivity of optical methods for measuring magnetic fields [7–11, 15], achieving the accuracy of SQUID magnetometers in conjunction with convenience typical of optical devices.

One of the effects closely related to CPT in degenerate systems is the nonlinear resonance Faraday effect [11–16], which is characterized by large angles of rotation of the polarization plane. The measurement of the angle of rotation of the polarization plane or, which is the same, of the phase difference between the right and left circular polarizations of an electromagnetic wave proves to be in some cases more convenient than recording CPT resonances in the absorption spectrum. For this reason, it is proposed to use the nonlinear resonance Faraday effect for the development of CPT magnetometers.

Nonlinear resonance polarization rotation is studied, as a rule, in a longitudinal magnetic field, which is parallel to the direction of propagation of the electromagnetic wave. Transverse magnetic fields, which are inevitably present in a laboratory (the Earth’s magnetic field), are eliminated by screening or using compensating schemes. In designing practical magnetometers, one has to take into consideration the effect of a transverse magnetic field and choose optimal parameters. In this paper, we studied experimentally the dependence of the angle of polarization rotation on the magnitude of a longitudinal magnetic field in ^{87}Rb vapor at the D_1-line $F = 2 \longrightarrow F' = 1$ transition for different strengths and directions of a transverse magnetic field. The $F = 2 \longrightarrow F' = 1$ transition was chosen among other transitions between the hyperfine-structure components of the $5s_{1/2}$ and $5p_{1/2}$ levels because it possesses a greater sensitivity to a magnetic field, thereby being most interesting for practical applications.

Another important problem that we considered in this paper is the influence of depletion of working levels in open systems on the nonlinear resonance Faraday

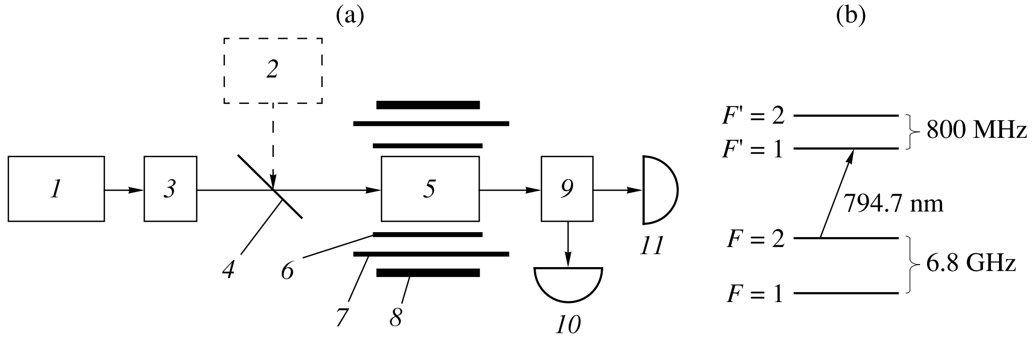


Fig. 1. (a) Scheme of the setup: (*1*) ECDL laser; (*2*) additional laser; (*3*) polarizer; (*4*) semitransparent mirror; (*5*) cell with ^{87}Rb vapor; (*6*) heater; (*7*) solenoid; (*8*) Helmholtz coils; (*9*) polarization beamsplitter; (*10*, *11*) photodiodes. (b) Hyperfine structure if the D_1-line of ^{87}Rb.

effect. Population depletion by optical pumping in an open system results in a decrease in the angle of polarization rotation. The depletion efficiency increases with increasing magnetic field, which destroys CPT. A cardinal solution of this problem would be the use of closed systems. However, the distance between the upper working level and neighboring levels in such systems is, as a rule, small. For example, this distance for the close D_2-line $F = 1 \longrightarrow F' = 0$ transition in ^{87}Rb is smaller than the inhomogeneous Doppler width of the transition, resulting in population depletion by optical pumping via nonresonant upper levels. In other words, the system is no longer closed. In addition, a small distance to neighboring levels leads to a strong nonresonant interaction, which limits the potential sensitivity of a magnetometer [10]. In this paper, we propose using radiation from an auxiliary laser to compensate for optical pumping out of working levels for obtaining a closed system. We studied both the case of pure rubidium and rubidium with a buffer gas, which is widely used to increase the lifetime of the Zeeman coherence. We experimentally demonstrated the increase in the sensitivity and dynamic range of the fields being measured upon compensation of the depletion.

The paper contains five sections. In Section 2, the scheme of the experimental setup is described. In Section 3, the influence of transverse magnetic fields on the nonlinear resonance Faraday effect is considered. Section 4 is devoted to analysis of the influence of the optical depletion in open systems on the effect under study and the possibilities of its compensation. In the Conclusions, the results are summarized and analyzed.

2. EXPERIMENTAL

Figure 1a shows the scheme of the experimental setup. A diode laser *1* with an external cavity (ECDL) tuned to the $F = 2 \longrightarrow F' = 1$ transition of the D_1-line of ^{87}Rb was used as a monochromatic radiation source (Fig. 1b). The laser radiation was linearly polarized. The polarization was controlled with the help of polarizer *3*. The laser beam with a spot size of 2×5 mm passed through a cylindrical cell *5* with a diameter of 56 mm and a length of 55 mm containing vapor of isotopically pure ^{87}Rb. Then the beam was incident on a polarization beamsplitter *9* mounted at an angle of 45° with respect to the polarization plane of laser radiation. The polarization rotation angle and absorption in the cell were calculated from the transmitted intensity in two arms of the beamsplitter measured by photodiodes *10* and *11*. A longitudinal magnetic field was produced by solenoid *7*. A transverse magnetic field was produced with the help of a system of two magnetic coils—Helmholtz coils *8*, 15 cm in diameter. The coils could be rotated around the system axis through an arbitrary angle, thereby varying the orientation of the transverse magnetic field with respect to the direction of polarization of laser radiation. The concentration of rubidium vapor in the cell was varied with the help of heater *6* and was approximately 2×10^{11} cm^{-3}. In this case, the polarization rotation angle was, on the one hand, not too large, so that the mutual orientation of the transverse field and the polarization direction changed weakly over the cell length, and on the other hand, this angle was sufficient to ensure the required accuracy of measurements.

Note that anomalously large polarization rotation angles were observed only in the case of coherent population trapping. When the laser intensity was below a certain threshold, polarization rotation disappeared almost completely. When the laser was tuned to the $F = 1 \longrightarrow F' = 2$ transition (CPT cannot be achieved at this transition [2, 4, 5]), the polarization rotation angle was smaller than the sensitivity of the detecting system (<0.1°) for any intensities of laser radiation.

The measurements were performed using rather high laser intensities on the order of 100 mW/cm^2. Under these conditions, the system was low-sensitive to weak magnetic fields on the order of the Earth's magnetic field (~0.5 G) [16], which allowed us to abandon shielding. The longitudinal component of the

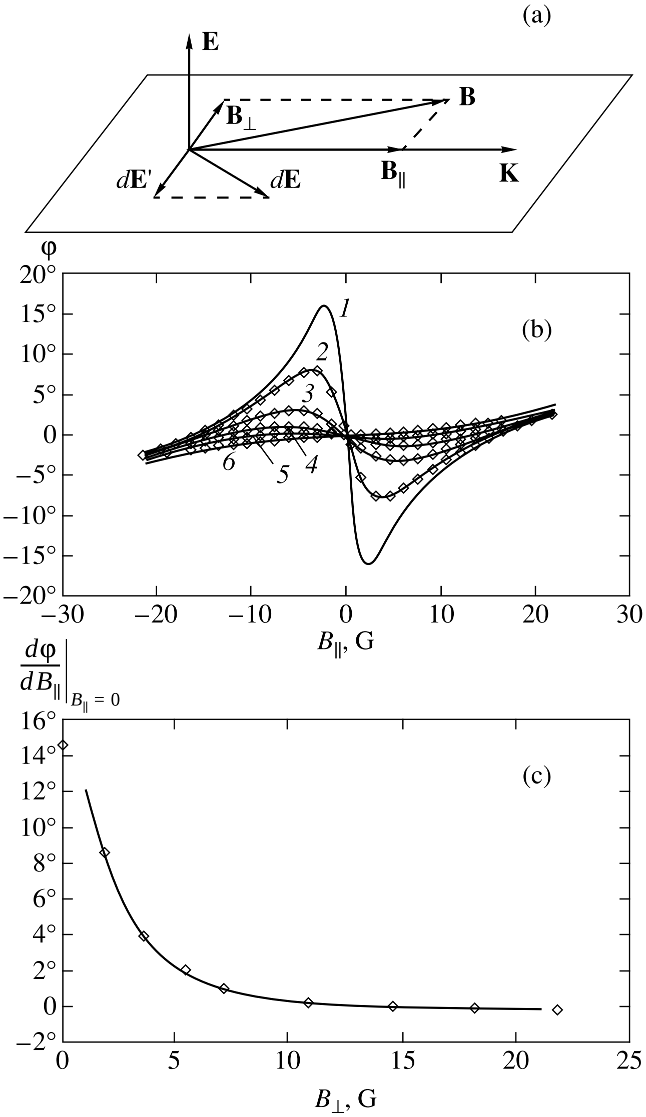


Fig. 2. (a) Scheme of the mutual orientation of the directions of polarization and a magnetic field when a transverse field is orthogonal to polarization. (b) Dependence of the polarization rotation angle on a longitudinal magnetic field for different strengths of a transverse field: $B_\perp$ = 0 (*1*), 3.64 (*2*), 7.28 (*3*), 10.9 (*4*), 14.6 (*5*), 18.2 G (*6*). The solid curves are experimental data; the diamonds are the values calculated from (1). (c) Dependence of the derivative of the polarization rotation angle with respect to the longitudinal magnetic field at field zero (sensitivity) on the transverse field strength: the solid curve is the dependence calculated from (2); the diamonds are experimental data.

Earth's magnetic field was determined by the shift of the zero of the effect and was taken into account in data processing.

We studied the influence of optical depletion of working levels using an additional laser *2*, which was tuned to the $F = 1 \longrightarrow F' = 2$ transition and produced pumping from the $F = 1$ level to the working $F = 2$ level through the upper $F' = 2$ level. The $F = 1 \longrightarrow F' = 2$ transition has no dark states, which provides a high pumping efficiency. In addition, the $F = 1 \longrightarrow F' = 2$ and $F = 2 \longrightarrow F' = 1$ transitions do not have common levels and are separated in frequency by 7.6 GHz, which excluded direct influence of radiation from the additional laser on the working transition. The beams from both lasers overlapped inside the cell. The intensity of the additional laser was 50 mW/cm^2, which, as follows from numerical calculations, is sufficient for the efficient compensation for the working-level depletion. Experiments were performed for two cases with substantially different relaxation constants: pure rubidium vapors and rubidium vapors in the presence of a buffer gas. As a buffer gas, neon at a pressure of 30 Torr was used.

3. INFLUENCE OF A TRANSVERSE MAGNETIC FIELD ON THE NONLINEAR RESONANCE FARADAY EFFECT

Our investigations showed that the influence of a transverse magnetic field on the nonlinear resonance Faraday effect was different for different orientations of polarization of laser radiation with respect to the transverse field. Theoretical analysis of this problem in the general case is rather complicated and can be probably performed only by numerical simulations.

For a transverse field perpendicular to the polarization direction, we propose a simple theoretical model, which is in good agreement with the experiment. In this case, the total magnetic field $\mathbf{B} = \mathbf{B}_\parallel + \mathbf{B}_\perp$ (longitudinal plus transverse) proves to be orthogonal to the polarization direction **E** (Fig. 2a). From the point of view of description of processes occurring in the medium, this situation is similar to the propagation of an electromagnetic wave along a magnetic field. It is known that, in the case of a longitudinal propagation of the wave, polarization rotation is observed, i.e., the response $d\mathbf{E}$ of the medium is orthogonal both to the polarization direction **E** and to the magnetic induction **B**. It is obvious that the latter will be true in our case as well, but the contribution to the polarization rotation angle will be determined not by $d\mathbf{E}$ but by its projection $d\mathbf{E}'$ on a plane orthogonal to the propagation direction. Thus, we obtain the polarization rotation angle

$$\varphi(B_\parallel) = \varphi_0\left(\sqrt{B_\parallel^2 + B_\perp^2}\right)\frac{B_\parallel}{\sqrt{B_\parallel^2 + B_\perp^2}}, \tag{1}$$

where $\varphi_0(B)$ is the dependence of the polarization rotation angle on the longitudinal magnetic field in the absence of a transverse field.

One can easily find from (1) the expression for the derivative of the polarization rotation angle with respect to the longitudinal magnetic field at the zero of the field, which determines the sensitivity of a magnetometer:

$$\left.\frac{d\varphi}{dB_{\parallel}}\right|_{B_{\parallel}=0} = \frac{\varphi_0(B_{\perp})}{B_{\perp}}. \tag{2}$$

The dependence of the polarization rotation angle on the longitudinal magnetic field for different strengths of the transverse field is shown in Fig. 2b. The solid curves are the experimental data, and the diamonds are theoretical values calculated by expression (1) [the experimental curve measured in the absence of a transverse magnetic field was used as $\varphi_0(B)$].

Figure 2c shows the dependence $(d\varphi/dB_{\parallel})|_{B_{\parallel}=0}$ on the transverse field strength. The diamonds are the experimental values. The solid curve is calculated from (2). The 0/0 uncertainty appearing in expression (2) when $B_{\perp} = 0$ and errors in the measurement of $\varphi_0(B)$ (in particular, we neglected the Earth's magnetic field) did not allow us to continue the curve correctly to the zero transverse field. One can see that the model is in excellent agreement with experimental data for all other values of the transverse magnetic field.

Expression (2) has a simple geometrical interpretation. The slope of the dependence of the polarization rotation angle on the longitudinal magnetic field at the zero of the function $(d\varphi/dB_{\parallel})|_{B_{\parallel}=0}$ in the presence of the transverse magnetic field $B_{\perp}$ is equal to the slope of a straight line drawn from the coordinate origin to the point $\varphi_0(B_{\perp})$. Therefore, the sensitivity of the magnetometer will change weakly when a transverse magnetic field is applied, which is perpendicular to polarization and smaller than $B_{\max}$ ($B_{\max}$ is the position of the maximum of the polarization rotation angle in the absence of transverse fields), and it will drastically decrease at larger values of the field.

In the case of another orientation of the perturbing transverse field, the situation was more complicated. When the transverse field was parallel to the light-wave polarization and exceeded $2B_{\max}$, the dependence of the polarization rotation angle on the longitudinal magnetic field exhibited a singularity near zero (Fig. 3), which was similar to that observed in paper [13]. The origin of this singularity is still unclear and requires additional study. However, as in the case of a transverse field orthogonal to laser-radiation polarization considered above, we can state that transverse fields that are smaller than $B_{\max}$ only weakly affect the resonance Faraday polarization rotation.

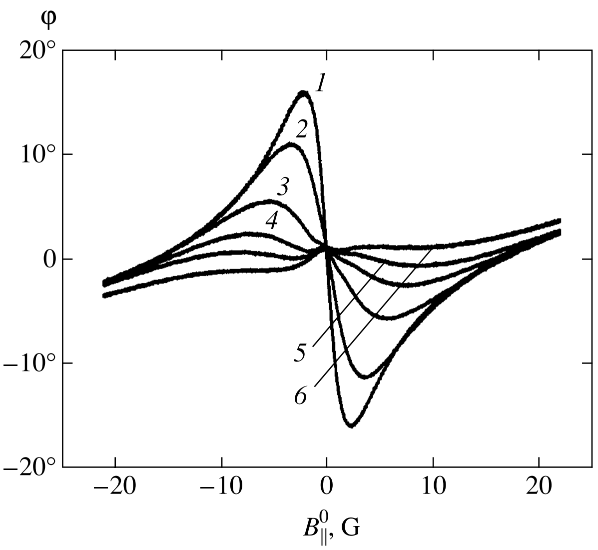


Fig. 3. Dependence of the polarization rotation angle on the longitudinal magnetic field for different strengths of the transverse magnetic field directed parallel to the polarization of an electromagnetic wave: $B_{\perp} = 0$ (*1*), 3.64 (*2*), 7.28 (*3*), 10.9 (*4*), 14.6 (*5*), 18.2 G (*6*).

Therefore, transverse fields of an arbitrary orientation weakly affect the nonlinear resonance Faraday effect until their strength exceeds $B_{\max}$, which should be taken into account, in particular, in the development of a magnetometer. In the presence of transverse fields, which are inevitable in real experimental conditions, it is expedient to use sufficiently high intensities providing the required power broadening of the transition and, hence, the required values of $B_{\max}$. The broadening of the transition reduces the slope of the dependence of the polarization rotation angle on the longitudinal magnetic field and seemingly should reduce the potential sensitivity of magnetometers based on the nonlinear resonance Faraday effect. However, in the case of CPT, the ratio of the slope of dependence of the polarization rotation angle to absorption is independent of the light-wave intensity [15]. Therefore, an increase in the intensity does not result in an increase in the sensitivity limit, which is determined from the signal-to-noise ratio. Moreover, when the sensitivity is restricted by shot noise, the error of measurement of a magnetic field, which can in principle be achieved, decreases as the square root of the light-wave intensity [11]. The real sensitivity of the magnetometer can be increased by increasing the optical thickness of the medium used, as well as by measuring more accurately the polarization rotation angle or, which is the same, the phase shift between the right-hand and left-hand circularly polarized components of laser radiation.

4. INFLUENCE OF OPTICAL DEPLETION OF WORKING LEVELS AND ITS COMPENSATION ON THE NONLINEAR RESONANCE FARADAY EFFECT IN OPEN SYSTEMS

Figure 4a shows the scheme of the (*1*) signal and (*2*) additional laser used to study the compensation for

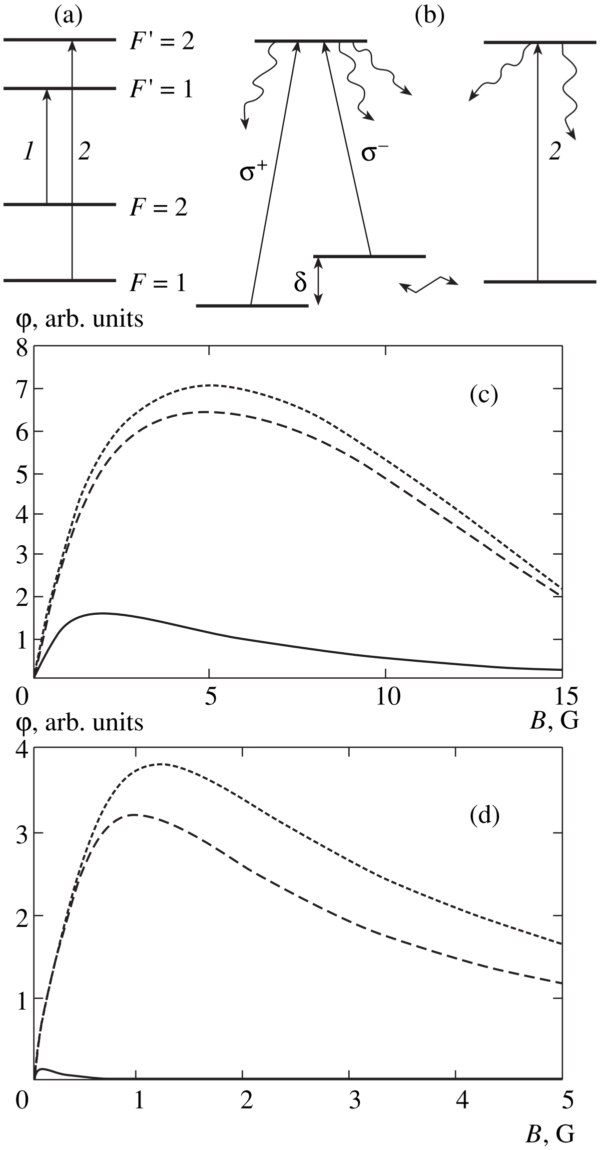


Fig. 4. (a) Scheme of the signal laser (*1*) and additional laser (*2*) used for compensating for the optical decay of the working $F = 2 \longrightarrow F' = 1$ transition. (b) Simplified model used for computer simulation of the optical decay: σ^+ and σ^- are the right-hand and left-hand polarized components of the main laser radiation; δ is the Zeeman splitting. (c, d) Dependences of the polarization rotation angle on the magnetic field strength calculated for rubidium without and with a buffer gas (neon at a pressure of 30 Torr), respectively; the Rabi frequency for the main laser radiation is 20 MHz. The solid curve is the dependence in the absence of compensation; the dotted curve is the dependence calculated for an infinite intensity of the additional laser's radiation (closed system); the dashed curve corresponds to the experiment, when the intensity of the additional laser's radiation was half that of the main laser.

the depletion of the working $F = 2 \longrightarrow F' = 1$ transition. The experiment was performed for two cases with substantially different relaxation constants: in pure rubidium vapor and rubidium vapor in the presence of a buffer gas. We studied the polarization rotation angle for the signal laser radiation resonant with the $F = 2 \longrightarrow F' = 1$ transition as a function of the longitudinal magnetic field. The additional laser, tuned to the $F = 1 \longrightarrow F' = 2$ transition, compensated for depletion of the working levels by populating the lower working $F = 2$ level through the upper $F' = 2$ level, thus making the system closed.

According to the plan of the experiment, we performed numerical simulations using a simplified scheme (Fig. 4b). The $F = 2 \longrightarrow F' = 1$ transition interacting with the main laser radiation was simulated with a polarization Λ scheme with the Zeeman splitting of lower levels $\delta = 2\mu_B gB/\hbar$, where μ_B is the Bohr magneton; g is the Lande g-factor, equal to 1/2 for the F = 2 level of the D_1-line of ^{87}Rb; and B is the magnitude of the applied magnetic field. The $F = 1 \longrightarrow F' = 2$ transition, which has no a dark state, was simulated by a two-level open system excited by an additional laser. The systems were connected by the relaxation of population between levels $F = 1$ and $F = 2$ (low-frequency relaxation) and the relaxation of population of the upper levels $F' = 1$, $F' = 2$ (high-frequency relaxation). The relation between the rates of high-frequency relaxation between the sublevels was chosen in accordance with the optical strengths of the corresponding transitions. The total constant of high-frequency relaxation γ and the constant of low-frequency relaxation Γ were set equal to 6 MHz and 16 kHz, respectively, for pure rubidium vapor and to 150 MHz and 1 kHz, respectively, for rubidium vapor in the presence of a buffer gas. The Rabi frequency for the main laser radiation was 20 MHz, which corresponds to a laser radiation power density equal to 100 mW/cm^2. In the calculations we took into account the Doppler inhomogeneous broadening of the transitions, which was equal to 270 MHz under our experimental conditions.

The numerical simulations showed that the compensation of the population depletion becomes efficient when the intensity of the additional laser radiation is two to three times lower than that of the main laser, and there is no point in further increasing the additional laser power. For this reason, we used a comparatively low-power additional laser emitting 50 mW/cm^2. Figures 4c and 4d show the calculated dependences of the polarization rotation angle on the magnetic field for rubidium with and without a buffer gas. The solid curve shows the dependence in the absence of compensation, the dotted curve corresponds to the infinite intensity of the additional laser (closed system), and the dashed curve corresponds to our experimental conditions, when the additional laser radiation intensity was half the intensity of the main laser radiation. One can see that the compensation for the population depletion results in a small (approximately twofold) increase in the slope of the curve at zero, which determines the sensitivity, and in the extension of the range of the fields being measured, i.e., in an increase in the position $B_{\max}$ of the maximum of the polarization rotation angle. Note that $B_{\max}$ also determines the possibility of working in the presence of a transverse magnetic field (see section 3). Our calculations showed that the extension of the dynamic range during the depletion compensation depends on the relation between the constants of low-frequency, Γ, and high-frequency, γ, relaxation, and increases with decreasing Γ/γ.

The experimental dependences of the polarization rotation angle on the magnetic field for rubidium with and without a buffer gas are presented in Figs. 5a and 5b. Curves *1* and *2* were obtained with and without the additional laser, respectively. One can see that the results of numerical simulations (Fig. 4c) are in good qualitative agreement with experiments performed in pure rubidium vapor. The compensation for the population depletion resulted in an increase in the sensitivity by a factor of 1.6 and in a doubling of the dynamic range (i.e., $B_{\max}$ was doubled). The results of measurements with a buffer gas (Fig. 5b) strongly differ from calculations (Fig. 4d). Instead of the expected doubling of the sensitivity and the tenfold extension of the dynamic range, only weak extension of the dynamic range was observed and the sensitivity increased by a factor of 50. This discrepancy is explained by the influence of the Earth's magnetic field, whose strength is approximately equal to $B_E = 500$ mG. The longitudinal component of the Earth's magnetic field was measured from the shift of the zero of the nonlinear resonance Faraday effect and was $B_{E\|} \approx 200$ mG under our experimental conditions, which was taken into account in experimental-data processing. The projection of the Earth's magnetic field in the direction perpendicular to the laser beam can be estimated as

$$B_{E\perp} = \sqrt{B_E^2 - B_{E\|}^2} \approx 450 \text{ mG},$$

which is lower than the width of the dynamic range of $B_{\max}$ when the population depletion is compensated and is substantially greater than the expected width of the dynamic range in the absence of a transverse field without compensation. Therefore, the transverse Earth magnetic field only weakly affected the resonance Faraday effect when the population depletion was compensated; however, its influence was dominant in the absence of compensation.

Therefore, the compensation for the population depletion from the working levels results in an increase in the sensitivity and in a substantial extension of the dynamic range of $B_{\max}$, which is especially important in studies of samples with small constants Γ of low-frequency relaxation in the presence of transverse magnetic fields. Note that the constant of low-frequency relaxation determines the ultimate sensitivity of magnetometers based on the nonlinear resonance Faraday effect [10, 11, 15].

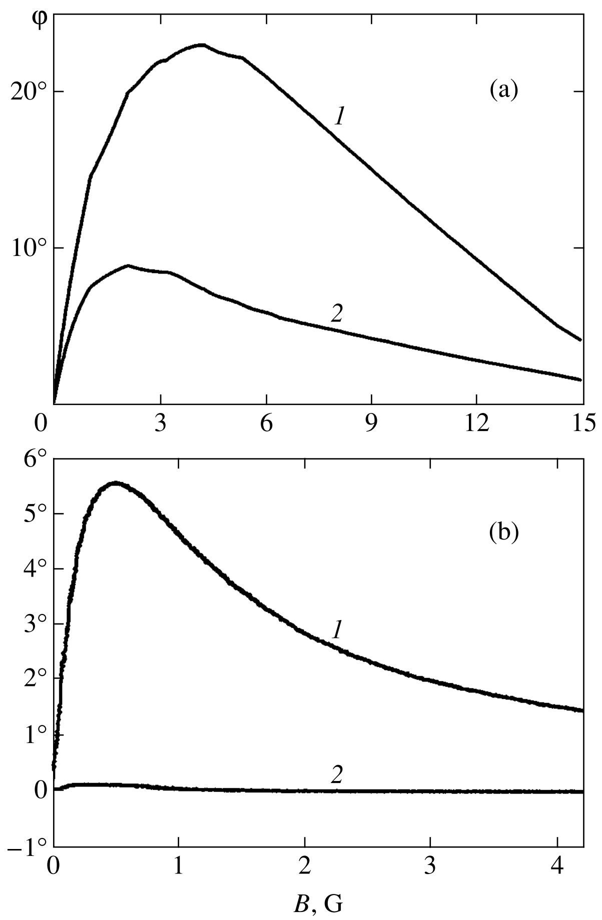


Fig. 5. Experimental dependences of the polarization rotation angle on the magnetic field strength in the presence (*1*) and absence (*2*) of an additional laser. (a) Pure rubidium vapor without a buffer gas; (b) rubidium vapor with a buffer gas (neon at a pressure of 30 Torr).

5. CONCLUSIONS

We have studied the influence of a transverse magnetic field on the nonlinear resonance Faraday effect in the case of coherent population trapping. For the fields orthogonal to the direction of polarization of an electromagnetic wave, a simple theoretical model has been proposed, which is in good agreement with experimental data. It is shown that transverse fields of an arbitrary orientation only weakly affect the nonlinear resonance Faraday effect until their strength does not exceed $B_{\max}$. Our studies have shown that in the development of CPT magnetometers based on the Faraday polarization rotation, it is expedient to use sufficiently high laser radiation intensities, which provide the required power broadening of the transition and, hence, a value of $B_{\max}$ exceeding the strength of transverse fields inevitable under real experimental conditions.

Another important problem that we considered in this paper is the influence of the working-level depletion by optical pumping in open systems on the nonlinear resonance Faraday effect. We have shown that the depletion in the open system reduced the sensitivity and the working range of the magnetometer. The choice of a convenient closed system is often complicated. We have proposed using an additional laser to compensate for the depletion in an open system. We have experimentally demonstrated the increase in the sensitivity and the extension of the dynamic range upon compensation for the depletion. Our studies have shown that the depletion compensation is especially important in the case of samples with small constants of low-frequency relaxation in the presence of transverse magnetic fields. The use of an additional laser to compensate for the depletion in experiments with a buffer gas in the presence of the Earth's magnetic field resulted in an increase in the sensitivity by a factor of 50. Note that reducing the low-frequency relaxation constant is the main method for increasing the ultimate sensitivity of a CPT magnetometer, which emphasizes the importance of our results.

Our studies have shown that the optical compensation for the working-level depletion becomes efficient when the intensity of the additional laser radiation is two to three times lower than that of the main laser. In addition, it is obvious that a laser with a comparatively broad emission line can be used as an additional laser, provided the spectral power is retained. All this reduces the requirements on the quality of the additional laser and makes the method for compensating for the depletion by optical pumping in open systems convenient in practice.

Our results allow one to better understand the phenomenon of the nonlinear resonance Faraday effect and are an important step in the development of real magnetometers based on this effect.

ACKNOWLEDGMENTS

The authors thank R.L. Kolesov, A.G. Litvak, V.A. Mironov, and V.E. Semenov for useful discussions. This work was supported by the Russian Foundation for Basic Research (project nos. 01-02-17779 and 03-02-06664).

REFERENCES

1. B. D. Agap'ev, M. B. Gornyĭ, B. G. Matisov, and Yu. V. Rozhdestvenskiĭ, Usp. Fiz. Nauk **163** (9), 35 (1993) [Phys.–Usp. **36**, 763 (1993)].
2. E. Arimondo, Prog. Opt. **35**, 257 (1996).
3. A. B. Matsko, O. Kocharovskaya, Y. Rostovtsev, *et al.*, Adv. At. Mol. Opt. Phys. **46**, 191 (2001).
4. V. S. Smirnov, A. M. Tumaĭkin, and V. I. Yudin, Zh. Éksp. Teor. Fiz. **96**, 1613 (1989) [Sov. Phys. JETP **69**, 913 (1989)].
5. A. M. Tumaĭkin and V. I. Yudin, Zh. Éksp. Teor. Fiz. **98**, 81 (1990) [Sov. Phys. JETP **71**, 43 (1990)].
6. I. V. Zelenskiĭ and V. A. Mironov, Zh. Éksp. Teor. Fiz. **121**, 1068 (2002) [JETP **94**, 916 (2002)].
7. A. Nagel, L. Graf, A. Naumov, *et al.*, Europhys. Lett. **44**, 31 (1998).
8. R. Wynands and A. Nagel, Appl. Phys. B **68**, 1 (1999).
9. M. O. Scully and M. Fleischhauer, Phys. Rev. Lett. **69**, 1360 (1992).
10. M. Fleischhauer, A. B. Matsko, and M. O. Scully, Phys. Rev. A **62**, 013808 (2000).
11. V. A. Sautenkov, M. D. Lukin, C. J. Bendar, *et al.*, Phys. Rev. A **62**, 023810 (2000).
12. D. Budker, V. V. Yashcuck, and M. Zolotorev, Phys. Rev. Lett. **81**, 5788 (1998).
13. D. Budker, D. F. Kimball, S. M. Rochester, and V. V. Yashcuck, Phys. Rev. Lett. **83**, 1767 (1999).
14. I. Novikova, A. B. Matsko, and G. R. Welch, Opt. Lett. **26**, 1016 (2001).
15. I. Novikova and G. R. Welch, J. Mod. Opt. **49**, 349 (2002).
16. R. A. Akhmedzhanov and I. V. Zelenskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 493 (2002) [JETP Lett. **76**, 419 (2002)].

Translated by M. Sapozhnikov

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 875–889.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 981–995.
Original Russian Text Copyright © 2003 by Bezverbnyĭ.

ATOMS, SPECTRA, RADIATION

Spatial Structure of Invariants in Monochromatic Field Configurations of Dimension $D > 1$

A. V. Bezverbnyĭ
Maritime State University, Vladivostok, 690059 Russia
e-mail: alexb@mail.vntc.ru, alexb@msun.ru

Received February 26, 2003

Abstract—We consider scalar (intensity, ellipticity) and gradient vector invariants for monochromatic field configurations of dimension $D > 1$. We analyze their spatial structure and peculiarities of the vector invariants (divergence and ambiguity, vortex fields) near singular regions (intensity extrema, regions of circular and linear polarizations). We study the convergence and definiteness of physical quantities (the multipole moments of atoms, the light-induced force, the diffusion tensor) with an invariant representation in the basis of these vector invariants. Various spatial structures of singular regions are presented for symmetric two- and three-dimensional configurations of a monochromatic field. *© 2003 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

The field configurations produced by coherent light waves with noncoincident directions and different polarizations are widely used in atomic physics, particularly for laser cooling and trapping of neutral atoms. These configurations are distinguished by the presence of spatial gradients in such field parameters as the ellipticity, intensity, total phase, and the rotation angles of the polarization ellipse. The correlation between the polarization and kinetic parameters of an atomic ensemble in such fields underlies the operation of magnetooptical traps and the action of sub-Doppler cooling mechanisms. The possible new dynamic and polarization effects in atomic structures optically aligned by fields with polarization gradients have been subjects of research in recent years [1]. A common property of these structures is the formation of periodic [2] or quasi-periodic [3] lattices both in the distribution density (dissipative [4] and far-detuned [5] lattices) and in the distribution of multipole atomic moments ρ_κ. For example, dissipative lattices are formed during the simultaneous action of a dipole force $\mathbf{F}_0$, which leads to the spatial localization of atoms, and a radiative friction force $\mathbf{F}_1(\mathbf{r}, \mathbf{v}) = \hat{\mathcal{X}} \cdot \mathbf{v} \sim \mathbf{v}$, which cools the atomic ensemble to sub-Doppler temperatures $k_B T \ll \hbar\gamma$ (γ is the radiative spontaneous excited-state decay constant). Clearly, such lattices are formed every time the technique for sub-Doppler cooling of atoms in light fields with polarization gradients is used.

In the semiclassical approximation, the kinetics of slow atoms in a nonuniformly polarized field is described by the Fokker–Planck equation for the atomic distribution function in phase space [6]. It was shown in [7] that the vector structure of $\mathbf{F}_0$, as well as the tensor structure of $\hat{\mathcal{X}}$ and the diffusion tensor in momentum space $\hat{\mathcal{D}}$ in the zero order in velocity, is determined by the spatial field gradients. Thus, a simple analysis of the gradient structure for an arbitrary field configuration can reveal the characteristic features in the dynamics of individual atoms and the kinetics of an atomic ensemble by determining, for example, the topology and symmetry of the localization regions, the spatial distribution of multipole moments, etc. Our main goal is to investigate the properties of the field gradients in monochromatic configurations of dimension $D > 1$.

The vector field invariants are defined in Section 2. The pattern of divergence or ambiguity of these quantities in singular regions is considered in the next section. We show that, in contrast to one-dimensional configurations, a vortex structure with the vortex centers in regions of circular polarization and with divergence of these gradients on the order of $1/r$ (r is the distance to the vortex center) is typical of the gradients of the total phase and the rotation angle of the polarization ellipse. This fact is connected to the absence of smoothness in the continuous ellipticity parameter $c = \cos(2\varepsilon)$ in these regions (ε is the ellipticity angle). The regions of linear polarization where the gradient of the rotation angle of the minor axis of the polarization ellipse diverges as $1/r$ should also be considered singular. The nodes at which the total field intensity is zero are degenerate regions of configurations. Here, in general, the regions of circular and linear polarizations intersect, the field ellipticity is undefined, and its gradient diverges as $1/r$.

In Sections 4 and 5, we investigate the problems of uniqueness and convergence for the multipole moments ρ_κ^g of ground-state atoms and the kinetic parameters

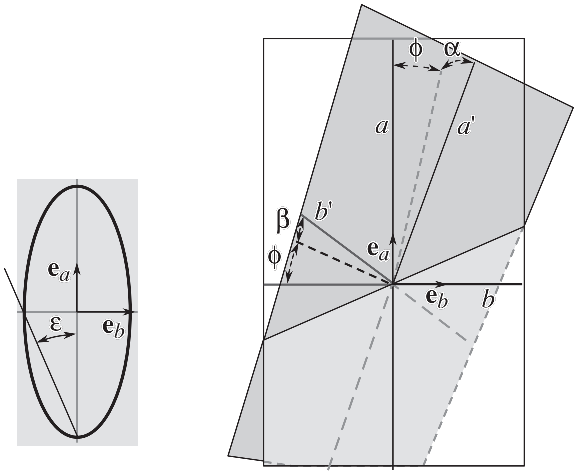


Fig. 1. The local polarization ellipse (left) and the rotation angles of the polarization ellipse under variation of **r**: *a*, *b* and *a*', *b*' are, respectively, the initial and modified directions of the principal axes of the polarization ellipse for field (1).

$\mathbf{F}_0$, $\hat{\mathscr{X}}$, and $\hat{\mathscr{D}}$ defined via vector field invariants. We show that these quantities are finite everywhere, except the nodes, while there is ambiguity in some of the contributions to $\hat{\mathscr{X}}$ and $\hat{\mathscr{D}}$ due to the gradients of ellipticity and total phase (in regions of circular polarization) and the gradient of the rotation angle of the minor axis of the field polarization ellipse (in regions of linear polarization).

In Section 6, we consider the spatial structures of singular regions for well-known and new two- and three-dimensional configurations and analyze the asymptotics of the field invariants in these regions. We show the topological variety of regions with circular and linear polarizations as well as intensity maxima and nodes.

In the Conclusions, we discuss the relationship between the spatial structure of field invariants and the structure of atomic lattices. In particular, we note a significant difference in the structures of (dissipative) lattices, depending on the type of transition. Thus, the predominant localization regions are the intensity minima together with the regions of circular polarization for $J \longrightarrow J$ transitions (J are half-integers) and the intensity maxima together with the regions of circular polarization for $J \longrightarrow J+1$ transitions. The largest effect must be observed at the points of intersection of these regions. In other cases, extended localized structures, for example, in the shape of rings in two-dimensional configurations, can be formed. The peculiarities of the spatial gradients of total phase and the angular parameters of the polarization ellipse have a direct bearing on the vortex component of the dipole force $\mathbf{F}_0$, which affects the dynamics of individual atoms in the localization regions.

2. INVARIANTS OF A MONOCHROMATIC FIELD

Let us consider the field configuration produced by a superposition of s coherent light beams of the same frequency ω with wave vectors $\mathbf{k}_n$:

$$\begin{aligned}\mathbf{E}(\mathbf{r},t) &= \sum_{n=1}^{s}\{\exp(-i\omega t + i\mathbf{k}_n\cdot\mathbf{r})\mathbf{E}_n + \text{c.c.}\}\\ &= \exp(-i\omega t)\mathbf{E}(\mathbf{r}) + \exp(i\omega t)\mathbf{E}^*(\mathbf{r}).\end{aligned} \tag{1}$$

where $(\ldots)^*$ is complex conjugation. Formally, we can separate out the amplitude $E \geq 0$, the total phase Φ, and the unit complex polarization vector ($\mathbf{e}\cdot\mathbf{e}^* = 1$) in the frequency components of the field $\mathbf{E}(\mathbf{r})$, $\mathbf{E}^*(\mathbf{r})$:

$$\begin{aligned}\mathbf{E}(\mathbf{r}) &= E(\mathbf{r})\exp(i\Phi(\mathbf{r}))\mathbf{e}(\mathbf{r}),\\ \mathbf{E}^*(\mathbf{r}) &= E(\mathbf{r})\exp(-i\Phi(\mathbf{r}))\mathbf{e}^*(\mathbf{r}).\end{aligned} \tag{2}$$

We do this by defining them directly in terms of the field invariants $I_0 = \mathbf{E}\cdot\mathbf{E}^*$, $I_1 = \mathbf{E}\cdot\mathbf{E}$, and I_1^*, which are convenient in analyzing specific configurations:

$$\begin{aligned}E = \sqrt{I_0}, \quad \exp(2i\Phi(\mathbf{r})) &= \frac{I_1}{\sqrt{I_1 I_1^*}},\\ \mathbf{e}(\mathbf{r}) &= \frac{\mathbf{E}(\mathbf{r})}{E\exp(i\Phi(\mathbf{r}))}.\end{aligned} \tag{3}$$

Definition (3) implies the choice of $c = \sqrt{I_1 I_1^*}/I_0 \geq 0$ for the ellipticity parameter $c(\mathbf{r}) = \cos(2\varepsilon) = \mathbf{e}\cdot\mathbf{e}$, which is related to the degree of linear field polarization $l = c^2$ [8] and the field ellipticity angle $\varepsilon(\mathbf{r})$ (Fig. 1), because there is the relation $c\exp(2i\Phi) = I_1/I_0$ between $c(\mathbf{r})$ and the phase factor $\exp(2i\Phi(\mathbf{r}))$. Our analysis of non-one-dimensional ($D > 1$) field configurations (see below) shows that for a different choice of relation to the phase factor, the function $c(\mathbf{r})$ will be discontinuous near the typical[1] regions with circular field polarization (isolated points and lines in two- and three-dimensional configurations, respectively). Of fundamental importance is the fact that the phase factor $\exp(i\Phi(\mathbf{r}))$ remains ambiguous in the definition of the unit polarization vector $\mathbf{e}(\mathbf{r})$ in (3) because of the ambiguity in the square root.

As the vector invariants, we choose the gradient functions [7] of amplitude, $\mathbf{g}_1 = \nabla E$; ellipticity parameter, $\mathbf{g}_3 = \nabla c$; and phase,

$$\mathbf{g}_2 = \nabla\Phi = \frac{i}{4}\frac{I_1}{I_1^*}\nabla\frac{I_1^*}{I_1} = \frac{i}{2}\frac{I_1}{\sqrt{I_1 I_1^*}}\nabla\frac{I_1^*}{\sqrt{I_1 I_1^*}}; \tag{4}$$

in contrast to Φ, vector (4) is uniquely defined.

---

[1] The definitions of a continuous ellipticity parameter, $-1 \leq \tilde{c} \leq 1$, are given in Section 6.

The gradients of the following three angular parameters that specify the spatial orientation of the polarization ellipse can serve as additional vector invariants: $\phi(\mathbf{r})$—the angle of rotation relative to the $\mathbf{e}_0$ axis orthogonal to the polarization ellipse as well as $\alpha(\mathbf{r})$ and $\beta(\mathbf{r})$—the angles of rotation of the major, $\mathbf{e}_a$, and minor, $\mathbf{e}_b$, axes of the polarization ellipse relative to the initial polarization plane (Fig. 1), where the axes are

$$\mathbf{e}_0 = \frac{i\mathbf{e}\times\mathbf{e}^*}{\sqrt{1-c^2}}, \quad \mathbf{e}_a = \frac{\mathbf{e}+\mathbf{e}^*}{\sqrt{2(1+c)}}, \quad \mathbf{e}_b = -\frac{i(\mathbf{e}-\mathbf{e}^*)}{\sqrt{2(1-c)}}. \tag{5}$$

In contrast to the angles themselves, invariant representations are possible for their gradients. Thus, in the local cyclic basis

$$\left\{\mathbf{e}_0;\ \mathbf{e}_{\pm1} = \mp\frac{\mathbf{e}_a \pm i\mathbf{e}_b}{\sqrt{2}}\right\},$$

the polarization vector is

$$\mathbf{e} = \cos\left(\varepsilon - \frac{\pi}{4}\right)\exp(-i\phi)\mathbf{e}_{+1} + \sin\left(\varepsilon - \frac{\pi}{4}\right)\exp(i\phi)\mathbf{e}_{-1}.$$

By considering the vector fields $(\mathbf{e}\nabla_i\mathbf{e}^*)$ and $(\mathbf{e}_0\nabla_i\mathbf{e}^*)$, we can obtain the following representations for the gradients of the angular quantities:

$$g_{4,i} = \nabla_i\phi = \frac{\mathrm{Im}(\mathbf{e}\nabla_i\mathbf{e}^*)}{\sqrt{1-c^2}}, \tag{6}$$

$$g_{5,i} = \nabla_i\alpha = \sqrt{\frac{2}{1+c}}\mathrm{Re}(\mathbf{e}_0\nabla_i\mathbf{e}), \quad g_{6,i} = \nabla_i\beta = \sqrt{\frac{2}{1-c}}\mathrm{Im}(\mathbf{e}_0\nabla_i\mathbf{e}). \tag{7}$$

Since, in contrast to $\mathbf{g}_4$, the vectors $\mathbf{g}_5$ and $\mathbf{g}_6$ in (7) implicitly contain phase factors of the form $\exp(\pm i\Phi)$, they are not defined uniquely.

There are relations between the vectors $\mathbf{g}_i$ ($i = 1, \ldots, 6$) that follow from the transversality of the frequency components $\nabla\mathbf{E}(\mathbf{r}) = \nabla\mathbf{E}^*(\mathbf{r}) = 0$ and the corollaries $\nabla^2\mathbf{E}(\mathbf{r}) = -k^2\mathbf{E}(\mathbf{r})$ and $\nabla^2\mathbf{E}^*(\mathbf{r}) = -k^2\mathbf{E}^*(\mathbf{r})$ of the wave equation for the total field of configuration (1). Let us also note an important relation between the gradients of the phase parameters:

$$\sqrt{1-c^2}\mathbf{g}_4 - \mathbf{g}_2 = \frac{\mathbf{G}}{I_0}, \tag{8}$$

where the $G_i = \mathrm{Im}(\mathbf{E}\nabla_i\mathbf{E}^*)$ is a vortex vector ($\nabla\mathbf{G} = 0$).

3. PECULIARITIES OF FIELD INVARIANTS

In one-dimensional field configurations, the wave vectors $\mathbf{k}_n$ of the initial fields are collinear and $\mathbf{g}_5$ and $\mathbf{g}_6$ are always absent. Here, the simplest configurations with only one gradient are realized [9]: $\mathbf{g}_2$ is present in a plane traveling wave; $\mathbf{g}_1$ is present in a standing wave; $\mathbf{g}_3$ is present in two counterpropagating waves of the same intensity with orthogonal linear polarizations, and $\mathbf{g}_4$ is present in counterpropagating waves with orthogonal circular polarizations (the lin ⊥ lin and σ_+–σ_- configurations, respectively [10]). The last three configurations, along with the lin–ϑ–lin model [11], are special cases of the symmetric ε–ϑ–$\bar{\varepsilon}$ model [12, 7] formed by two counterpropagating waves with the same intensities and ellipticities and the angle ϑ between the principal axes of their polarization ellipses. In these models, the invariants (including the gradient ones) can be easily defined in the form of smooth functions for the entire configuration region. If the ellipticity parameter is specified in the range $-1 \le \tilde{c} \le 1$, singularities emerge only in the limiting cases, for example, for $\vartheta \longrightarrow 0$ in the lin–ϑ–lin model [11].

A qualitatively different picture takes place in configurations with polarization gradients of dimension $D > 1$. Irrespective of the method for reconciling the signs of $c(\mathbf{r})$ and $\exp(2i\Phi)$, singularities emerge in the vector invariants in the limiting cases $I_0 = 0$ (field nodes), $I_0 = |I_1|$ (linear polarization), and $I_1 = 0$ (circular polarization).

In a nondegenerate case, $I_0(0) = \mathcal{I} > 0$, the asymptotics near the points with circular polarization is invariant, $I_1 \approx \mathbf{A}(0)\cdot\mathbf{R}$, where, below, $\mathcal{I}$ denotes I_0 at singular points and the radius vector $\mathbf{R}$ is assumed to be given in a local coordinate system with the origin in a singular region (with linear, circular polarization) or at an intensity extremum, depending on the context. Here, $\mathbf{A}(\mathbf{R}) = \nabla I_1$ is a complex vector: $\mathbf{A} = \mathbf{A}_1 + i\mathbf{A}_2$; in general, $\mathbf{A}_1 \times \mathbf{A}_2 \neq 0$. The latter property has a direct bearing on the choice of only a nonnegative ellipticity parameter, $c \approx |\mathbf{A}\cdot\mathbf{R}|/\mathcal{I} \ge 0$, if we want it to be continuous. In exceptional cases when $\mathbf{A} = e^{i\alpha}\mathbf{A}_0$, where $\mathbf{A}_0$ is a real vector and α is a constant phase, we can define a continuous alternating ellipticity parameter in the form

$$-1 \le \tilde{c} \approx \frac{\mathbf{A}_0\cdot\mathbf{R}}{\mathcal{I}} \le 1,$$

while simultaneously specifying the phase factor $e^{2i\Phi} \approx e^{i\alpha}$. In general, however, there is a singularity in the phase factor: $e^{2i\Phi} \approx \mathbf{A}\cdot\mathbf{R}/|\mathbf{A}\cdot\mathbf{R}|$ for $\mathbf{R} \longrightarrow 0$. Accordingly, $\mathbf{g}_2$ diverges when $c \longrightarrow 0$ as

$$\mathbf{g}_2 \approx \frac{[\mathbf{A}_1 \times \mathbf{A}_2] \times \mathbf{R}}{|\mathbf{A}\cdot\mathbf{R}|^2} \tag{9}$$

and its field has a vortex pattern with the vortex axis along $\mathbf{A}_1 \times \mathbf{A}_2$. Obviously, $\mathbf{g}_4$ has the same asymptotics, because the vector $\mathbf{G}$ in (8) is always finite. Being finite, the vector $\mathbf{g}_3$ is undefined at the point $c = 0$ itself, because it has the asymptotics

$$\begin{gathered}\mathbf{g}_3 \approx \frac{\hat{\mathcal{P}}(0)\cdot\mathbf{R}}{\mathcal{J}(\mathbf{R}\cdot\hat{\mathcal{P}}(0)\cdot\mathbf{R})^{1/2}}, \\ \hat{\mathcal{P}} = \mathbf{A}_1 \otimes \mathbf{A}_1 + \mathbf{A}_2 \otimes \mathbf{A}_2,\end{gathered} \tag{10}$$

where $\mathbf{A}_i \otimes \mathbf{A}_i$ denote a direct product of the corresponding vectors.

The gradients $\mathbf{g}_5$ and $\mathbf{g}_6$ are finite in the vicinity of circular polarization:

$$\mathbf{g}_5 \approx \left(\frac{2}{\mathcal{J}^3}\right)^{1/2}(\mathbf{B}_2\cos\Phi - \mathbf{B}_1\sin\Phi),$$

$$\mathbf{g}_6 \approx -\left(\frac{2}{\mathcal{J}^3}\right)^{1/2}(\mathbf{B}_1\cos\Phi + \mathbf{B}_2\sin\Phi),$$

where $\mathbf{B}_1$ and $\mathbf{B}_2$ are, respectively, the real and imaginary components of the everywhere-defined vector

$$B_i = [\mathbf{E}\times\mathbf{E}^*]\cdot\nabla_i\mathbf{E}.$$

However, as we noted above, $\sin\Phi$ and $\cos\Phi$ are not defined uniquely (in contrast to $\sin(2\Phi)$ and $\cos(2\Phi)$).

The asymptotics of the invariants near the points with linear polarization in a nondegenerate case ($\mathcal{J} > 0$) are

$$I_0 \approx \mathcal{J} + (\mathrm{Re}\tilde{\mathbf{A}})\cdot\mathbf{R} + \mathbf{R}\cdot\hat{\mathcal{R}}_1\cdot\mathbf{R},$$

$$I_1 \approx \mathcal{J} + \tilde{\mathbf{A}}\cdot\mathbf{R} + \mathbf{R}\cdot\hat{\mathcal{R}}_2\cdot\mathbf{R},$$

where $\hat{\mathcal{R}}_1$ is a real tensor and $\hat{\mathcal{R}}_2$ is a complex tensor. Clearly, the vectors

$$\mathbf{g}_1 \approx \frac{\mathrm{Re}\tilde{\mathbf{A}}}{2\mathcal{J}}, \quad \mathbf{g}_2 \approx -\frac{\mathrm{Im}\tilde{\mathbf{A}}}{2\mathcal{J}},$$

and $\mathbf{g}_4$ are finite, and the ellipticity gradient $\mathbf{g}_3 \approx -2\hat{\mathcal{R}}_3\mathbf{R}$ is small, because the ellipticity parameter is extremal, $\nabla c = 0$, at the points with linear polarization.

Here, $c \approx 1 - \mathbf{R}\cdot\hat{\mathcal{R}}_3\cdot\mathbf{R}$, with the eigenvalues of the tensor $\hat{\mathcal{R}}_3$ being nonnegative. The quantities $\mathbf{g}_5$ and $\mathbf{g}_6$ for $c \longrightarrow 1$ are

$$\begin{gathered}\mathbf{g}_5 \approx \pm\frac{\mathbf{B}_2}{\sqrt{2\mathcal{J}^3\mathbf{R}\cdot\hat{\mathcal{R}}_3\cdot\mathbf{R}}}, \\ \mathbf{g}_6 \approx \pm\frac{\mathbf{B}_1}{\mathcal{J}^{3/2}\mathbf{R}\cdot\hat{\mathcal{R}}_3\cdot\mathbf{R}},\end{gathered} \tag{11}$$

where the signs of the vectors, being the same, are not defined uniquely. The vectors introduced above have an asymptotic behavior: $\mathbf{B}_1 \approx \mathbf{S}\times\mathbf{R}$, while $\mathbf{B}_2 \approx \hat{\mathcal{S}}\mathbf{R}$, where the real tensor $\hat{\mathcal{S}}$ also contains a symmetric component. The vector $\mathbf{S}$, which defines the plane of the vortex field $\mathbf{B}_1$, can be represented via the parameters of the fields that form configuration (1):

$$\mathbf{S} = 2\sum_{m,n}^{s}(\mathbf{E}(0)\cdot[\mathrm{Re}\mathbf{E}_m\times\mathrm{Re}\mathbf{E}_n])[\mathbf{k}_m\times\mathbf{k}_n].$$

Thus, the vector $\mathbf{g}_5$ is finite and undefined at $c = 1$, while $\mathbf{g}_6$ has a vortex pattern and diverges when $c \longrightarrow 1$.

Near the singular points of the invariant I_0, which is proportional to the intensity of the total field (1), where $\nabla I_0 = 0$, its asymptotics is $I_0 \approx \mathcal{J} + \mathbf{R}\cdot\hat{\mathcal{Q}}\cdot\mathbf{R}$, and the eigenvalues q_m of the tensor $\hat{\mathcal{Q}}$ define the type of extremum or the saddle point.

The nodes ($\mathcal{J} = 0$) where $q_m \geq 0$ are the degenerate cases in field configurations. Here, the gradient

$$\mathbf{g}_1 \approx \frac{\hat{\mathcal{Q}}\cdot\mathbf{R}}{\sqrt{\mathbf{R}\cdot\hat{\mathcal{Q}}\cdot\mathbf{R}}}$$

is undefined. The asymptotics of the invariant $I_1 \approx \mathbf{R}\cdot\hat{\mathcal{Q}}_1\cdot\mathbf{R}$; the condition $|\mathbf{R}\cdot\hat{\mathcal{Q}}_1\cdot\mathbf{R}| \leq \mathbf{R}\cdot\hat{\mathcal{Q}}\cdot\mathbf{R}$ must be satisfied for the complex tensor $\hat{\mathcal{Q}}_1$. In general, the ellipticity parameter $c = |I_1|/I_0$ is undefined and the vector $\mathbf{g}_3$ diverges as $1/R$ near the nodes. The vectors $\mathbf{g}_{2,4}$ have the same pattern of divergence, except for the cases where $(\hat{\mathcal{Q}}_1)^* = e^{i\alpha}\hat{\mathcal{Q}}_1$. The vector $\mathbf{g}_6$ also has a $1/R$ divergence at the nodes.

The dimension of singular regions, obviously, depends on the dimension of the field configuration. The regions of circular polarization for $I_0 > 0$ are specified by two conditions: $\mathrm{Re}I_1 = 0$ and $\mathrm{Im}I_1 = 0$. Usually, these are isolated points in two-dimensional configurations and one-dimensional lines in three-dimensional configurations. However, if $I_1 \sim |I_1|$, to within the constant phase factor, which corresponds to the case $\mathbf{A} = e^{i\alpha}\mathbf{A}_0$ considered above, then only one condition remains and the dimension of the region increases. An example of a two-dimensional model with regions of circular polarization in the form of closed lines is given in Section 6.1. The regions of linear polarization that correspond to the maximum values of $c = 1$, as well as the regions of extremal intensities of field (1), satisfy the conditions $\nabla F = 0$ ($F = \{c; I_0\}$), which, in general, leads to D independent equations that define these regions in the form of isolated points. However, the amount of independent equations may turn out to be less. The various intensity extrema and regions of linear polarization in the form of isolated points, lines, and two-dimensional surfaces (in three-dimensional configurations) are given in Section 6.

4. THE MULTIPOLE MOMENTS OF ATOMS

Let us consider the problem of uniqueness and finiteness at the singular points of the physical quantities with gradient invariants present in their representations.

It was shown in [13, 14] that the stationary[2] multipole moments ρ_κ of atoms at rest with a total ground-state angular momentum $J_g > 0$ aligned by field (1) in a cycle of resonant optical pumping can be represented in an invariant way as expansions into a basis of tensor products of the vectors **e** and **e***. The characteristic relaxation times for such stationary distributions (in internal degrees of freedom) are $t_\kappa \sim \max\{\gamma^{-1}, (\gamma S)^{-1}\}$, where S is the saturation parameter defined via the Rabi frequency Ω and the field frequency detuning $\delta = \omega - \omega_0$ from the atomic transition frequency as

$$S = \frac{|\Omega^2|}{\gamma^2/4 + \delta^2}. \tag{12}$$

For example, at small saturations, $S \ll 1$, the expansion of the multipole moments ρ_κ^g of ranks $0 < \kappa \le 2J_g$ for ground-state atoms at rest is [14]

$$\rho_\kappa^g = \sum_{l=\bar{\kappa}-\kappa}^{\kappa} a^{\kappa,l}\{\{\mathbf{e}\}_l \otimes \{\mathbf{e}^*\}_{\bar{\kappa}-l}\}_\kappa. \tag{13}$$

Here, $\bar{\kappa} = 2[(\kappa+1)/2]$, where $[X]$ denotes taking an integer part of the number X. The following notation is used in (13): $\{\cdots\otimes\cdots\}_\kappa$ for an irreducible tensor product of rank κ [15], and $\{\mathbf{e}\}_l = \{\cdots\{\{\mathbf{e}\otimes\mathbf{e}\}_2\otimes\mathbf{e}\}_3\cdots\otimes\mathbf{e}\}_l$ are the abbreviations suggested in [16, 17].

For slow atoms under the condition

$$\min(\gamma, \gamma S) \gg \max\left(kv, \frac{v}{\Delta r}\right), \tag{14}$$

where Δ_r is the spatial scale for the polarization gradients, a similar expansion holds for the linear (in velocity) corrections $\delta\rho_\kappa^g$ [14]:

$$\begin{aligned}\delta\rho_\kappa^g = \frac{1}{\gamma S}\Bigg[&\sum_{l=\bar{\kappa}-\kappa}^{\kappa}((\delta a)_3^{\kappa,l}\mathbf{v}\cdot\mathbf{g}_3 + (\delta a)_4^{\kappa,l}\mathbf{v}\cdot\mathbf{g}_4)\\ &\times\{\{\mathbf{e}\}_l\otimes\{\mathbf{e}^*\}_{\bar{\kappa}-l}\}_\kappa\\ &+\sum_{l=\underline{\kappa}-\kappa}^{\kappa}((\delta a)_5^{\kappa,l}\mathbf{v}\cdot\mathbf{g}_5 + (\delta a)_6^{\kappa,l}\mathbf{v}\cdot\mathbf{g}_6)\\ &\times\{\{\mathbf{e}\}_l\otimes\{\mathbf{e}^*\}_{\underline{\kappa}-l}\}_\kappa\Bigg].\end{aligned} \tag{15}$$

Here, $\underline{\kappa} = \overline{\kappa + 1} - 1$.

There is no ambiguity in expressions (13) and (15) because the phase factors $\exp(\pm i\Phi)$ in the polarization vectors **e** and **e*** (3) are not uniquely defined. The evenness of $\bar{\kappa}$ ensures an even number of vectors **e** and **e*** for the tensor products with the coefficients $a^{\kappa,l}$, $(\delta a)_3^{\kappa,l}$, and $(\delta a)_4^{\kappa,l}$. The resulting phase factors are $[\exp(\pm 2i\Phi)]^n$ and are uniquely defined in accordance with (3). The phase factors for the contributions with the coefficients $(\delta a)_5^{\kappa,l}$ and $(\delta a)_6^{\kappa,l}$ have a similar structure: for an odd number $\underline{\kappa}$ of vectors **e** and **e*** in each tensor product, the arising ambiguity is offset by the corresponding phase factors $\exp(\pm i\Phi)$ in the vectors $\mathbf{g}_{5,6}$.

Analyzing the finiteness of expressions (13) and (15), we note that the expansion coefficients $a^{\kappa,l}$, $(\delta a)_3^{\kappa,l}$, $(\delta a)_4^{\kappa,l}$, $(\delta a)_5^{\kappa,l}$, and $(\delta a)_6^{\kappa,l}$ are fractional rational functions of the ellipticity parameter $c(\mathbf{r})$ without singularities for $c \longrightarrow 0$, while for $c \longrightarrow 1$, the coefficients $(\delta a)_{4,5,6}^{\kappa,l}$ and $(\delta a)_3^{\kappa,l}$ have $1/\sqrt{1-c^2}$ and $1/(1-c^2)$ singularities, respectively. Nevertheless, the resulting contributions to $\delta\rho_\kappa^g$ are everywhere finite, except for the nodes. For the first-rank moment $\delta\rho_1^g$, this was shown in [14] for the $1/2 \longrightarrow 1/2$ transition. It turns out that the gradients that diverge ($\mathbf{g}_{2,4}$) and that are undefined ($\mathbf{g}_3$) in the vicinity of $c \longrightarrow 0$ enter into the multipoles of a higher rank at $J_g > 1/2$ in the finite combinations $c^n\mathbf{g}_m$, where $n \ge 1$ and $m = 3, \ldots, 6$. The

---

[2] An expansion of form (13) also holds in a nonstationary case for an equilibrium initial distribution in Zeeman atomic sublevels.

finiteness in the other limit $c \longrightarrow 1$ can be shown only after the addition of all contributions by taking into account the singularities of the tensors $\{\{\mathbf{e}\}_l \otimes \{\mathbf{e}^*\}_L\}_\kappa$ in the region of linear polarization where $\mathbf{e}^* \approx \mathbf{e}$. Therefore, the divergences of the individual contributions are cancelled out and the asymptotics proves to be finite. For example, the diverging gradient $\mathbf{g}_6$ enters in the finite combination $\sqrt{1-c}\,\mathbf{g}_6$. Since the asymptotics of $\delta\rho^g_\kappa$ for specific transitions with $J > 1/2$ are cumbersome, they are not given here.

Clearly, expansions (13) and (15) are inapplicable near the nodes ($S = 0$): here, $t_\kappa \longrightarrow \infty$ and conditions (14) for the ellipticity parameter are violated because $\Delta r \longrightarrow 0$. In this case, the structure of the multipole moments near the nodes is not determined by the local values of I_0, $\mathbf{e}(\mathbf{r})$, $\mathbf{e}^*(\mathbf{r})$, and their gradients but depends on the averaged (integrated) field parameters.

5. KINETIC COEFFICIENTS

Let us consider the application of vector invariants to the problem of invariant representation of kinetic coefficients, which is of relevant interest in analyzing the localization and cooling of atoms in field configurations with $D > 1$. It was shown in [7] that in the semiclassical approximation equivalent to the condition

$$\min(\gamma, \gamma S) \gg \frac{\hbar k^2}{2m}, \tag{16}$$

under the approximations given in the preceding section, the kinetic evolutionary stage ($t \gg \max(t_\kappa)$) of an atomic ensemble in a light field is described by the Fokker–Planck equation for the atomic distribution function $f(\mathbf{r}, \mathbf{p})$ in phase space,

$$\begin{aligned}(\partial_t + \mathbf{v}\cdot\nabla) f &= -\nabla_{\mathbf{p}} \cdot [(\mathbf{F}_0 + \hat{\mathcal{X}} \cdot \mathbf{v}) f] \\ &+ \frac{1}{2}(\nabla_{\mathbf{p}} \cdot (\hat{\mathcal{D}}^{\text{ind}} + \hat{\mathcal{D}}^{\text{spont}}) \cdot \nabla_{\mathbf{p}}) f,\end{aligned} \tag{17}$$

($\mathbf{v} = \mathbf{p}/M$ is the atomic velocity) with the following invariant expansions for the coefficients of (17):

$$\begin{aligned}\mathbf{F}_0 &= \hbar\gamma S \sum_{i=1}^{4} F_i \mathbf{g}_i, \\ \hat{\mathcal{X}} &= \hbar\left(\sum_{i;\,j=1}^{4} X_{ij}\mathbf{g}_i \otimes \mathbf{g}_j + \sum_{l;\,m=5}^{6} X_{lm}\mathbf{g}_l \otimes \mathbf{g}_m\right),\end{aligned} \tag{18}$$

$$\begin{aligned}\hat{\mathcal{D}}^{\text{ind}} &= \hbar^2\gamma S \\ &\times\left(\sum_{i;\,j=1}^{4} D_{ij}\mathbf{g}_i \otimes \mathbf{g}_j + \sum_{l;\,m=5}^{6} D_{lm}\mathbf{g}_l \otimes \mathbf{g}_m\right).\end{aligned} \tag{19}$$

Here, the expansion coefficients F, X, and D are functions of only E, c, and detuning δ of the light field frequency from resonance.[3] As an example, we give these coefficients in explicit form for the 1/2 ⟶ 1/2 transition in the low-saturation approximation, $S \ll 1$ (these coefficients are given in [7] for the 1/2 ⟶ 3/2 and 1 ⟶ 2 transitions):

$$F_1 = -\frac{2\tilde{\delta}c^2}{3}, \quad F_2 = \frac{c^2}{3}, \quad F_3 = -\frac{\tilde{\delta}c}{3}, \tag{20}$$

$$\begin{aligned}&X_{13} = 3\tilde{\delta}c, \quad X_{23} = -\frac{3c}{2}, \quad X_{33} = -\frac{3\tilde{\delta}c^2}{2A^2}, \\ &X_{43} = \frac{3c}{2A}, \quad X_{\left(\begin{smallmatrix}55\\66\end{smallmatrix}\right)} = \pm 24\tilde{\delta}A^2 c\mathbb{N}, \\ &X_{\left(\begin{smallmatrix}56\\65\end{smallmatrix}\right)} = -3A(1 \pm c)(\mathbb{A} \mp \mathbb{B})\mathbb{N},\end{aligned} \tag{21}$$

$$\begin{aligned}&D_{11} = \frac{c^2}{4}(1 + 12\tilde{\delta}^2 A^2), \quad D_{12} = -\frac{3\tilde{\delta}A^2c^2}{2}, \\ &D_{13} = -\frac{c}{8}(12\tilde{\delta}^2c^2 - 1), \quad D_{14} = \tilde{\delta}Ac^2, \\ &D_{22} = \frac{c^2}{4}(1 + 3A^2), \quad D_{23} = \frac{3\tilde{\delta}c^3}{4}, \\ &D_{24} = -\frac{Ac^2}{2}, \quad D_{33} = \frac{2 - c^2 + 12\tilde{\delta}^2c^4}{16A^2}, \\ &D_{34} = -\frac{\tilde{\delta}c^3}{2A}, \quad D_{44} = \frac{c^2}{2}, \\ &D_{\left(\begin{smallmatrix}55\\66\end{smallmatrix}\right)} = \frac{1 \pm c}{8}[1 + (\mathbb{C} \pm \mathbb{D})\mathbb{N}], \\ &D_{56} = 8\tilde{\delta}Ac^3\mathbb{N}, \\ &D_{i>j} = D_{ji},\end{aligned} \tag{22}$$

the remaining coefficients are equal to zero. Here, $A = \sqrt{1-c^2}$ is the coefficient related to the degree of linear

---

[3] The spontaneous diffusion tensor $\hat{\mathcal{D}}^{\text{spont}}$ has an expansion similar to (13) and is digonalized in the basis $\{\mathbf{e}_0, \mathbf{e}_a, \mathbf{e}_b\}$ in accordance with [7].

polarization A^2 in (1), $\tilde{\delta} = \delta/\gamma$ is the relative detuning,

$$\mathbb{N} = [2(9 - c^2 + 36\tilde{\delta}^2 A^2)]^{-1}, \quad \mathbb{A} = (12\tilde{\delta}^2 - 1)c,$$
$$\mathbb{B} = 12\tilde{\delta}^2 + 3, \quad \mathbb{C} = 2c^2(13 + 84\tilde{\delta}^2) - 18(1 + 4\tilde{\delta}^2),$$
$$\mathbb{D} = 8c^3(12\tilde{\delta}^2 - 1).$$

Since the vectors $\mathbf{g}_{5,6}$ are present in expressions (18) and (19) only in the form of quadratic combinations, no ambiguity due to the phase factors arises. Analysis of closed atomic dipole $J \longrightarrow J + 1$ (integer and half-integer J) and $J \longrightarrow J$ (half-integer J) transitions shows that, at least for $J \le 2$, the physical quantities $\mathbf{F}_0$, $\hat{\mathcal{X}}$, and $\hat{\mathcal{D}}^{\text{ind}}$ are everywhere finite but nonsmooth at the singular points if the approximations (14) and (16) hold. Let us consider this issue for coefficients (20)–(22) for the 1/2 $\longrightarrow$ 1/2 transition.

The contribution with $\mathbf{g}_1$ in the expression for the force $\mathbf{F}_0$, along with the analogous contribution in the expression for the tensor $\hat{\mathcal{D}}^{\text{ind}}$, is everywhere defined, because it enters in the combination $I_0\mathbf{g}_1$ (given $S \propto I_0$). For the tensor $\hat{\mathcal{X}}$, coefficients (21) are inapplicable near the nodes for the reasons given in the preceding section. Here, allowance for the nonlinear (in velocity) contributions similar to that made in [11] for a one-dimensional lin–ϑ–lin model is required to refine $\hat{\mathcal{X}}$.

Let us consider the vicinities of the points with circular polarization. Analysis of the force $\mathbf{F}_0$ and the tensor $\hat{\mathcal{X}}$ shows that the contributions with the ellipticity gradient $\mathbf{g}_3$ enters in the finite form $c\mathbf{g}_3$. The components containing $\mathbf{g}_{2,4}$ also enter in the finite form $c^2\mathbf{g}_2$ or in the combination $\mathbf{g}_2 - \mathbf{g}_4 \approx -\mathbf{G}/I_0$ without singularities. The contributions with $\mathbf{g}_{5,6}$ in the expression for the tensor $\hat{\mathcal{X}}$ are also finite, but their asymptotics is fundamentally different for $c \longrightarrow 0$ for $J \longrightarrow J + 1$ and $J \longrightarrow J$ transitions. Thus, for the 1/2 $\longrightarrow$ 1/2 transition, according to (21), this is the antisymmetric tensor

$$\hat{\mathcal{X}} \approx \frac{\hbar}{2}(\mathbf{g}_5 \otimes \mathbf{g}_6 - \mathbf{g}_6 \otimes \mathbf{g}_5),$$

which corresponds to the contribution of the effective Lorentz force $\hat{\mathcal{X}}\mathbf{v} \approx \mathbf{B}_{\text{eff}} \times \mathbf{v}$ with the effective field $\mathbf{B}_{\text{eff}} = \hbar\mathbf{g}_5 \times \mathbf{g}_6$. For the 1/2 $\longrightarrow$ 3/2 transition, the asymptotics is

$$\hat{\mathcal{X}} \approx \frac{2\hbar}{4\tilde{\delta}^2 + 1}[(4\tilde{\delta}^2 - 1)$$
$$\times (\mathbf{g}_5 \otimes \mathbf{g}_5 + \mathbf{g}_6 \otimes \mathbf{g}_6) + 4\tilde{\delta}(\mathbf{g}_5 \otimes \mathbf{g}_6 - \mathbf{g}_6 \otimes \mathbf{g}_5)],$$

where the first term, being a symmetric tensor, describes the radiative cooling processes due to the contributions from $\mathbf{g}_{5,6}$.

Near the points with linear polarization, the singularities of the tensor $\hat{\mathcal{X}}$ are related only to the terms with $\mathbf{g}_{5,6}$. We see from expressions (21) that these contributions are finite in accordance with the asymptotics of the vectors $\mathbf{B}_i$ earlier introduced but are undefined at the point of linear polarization:

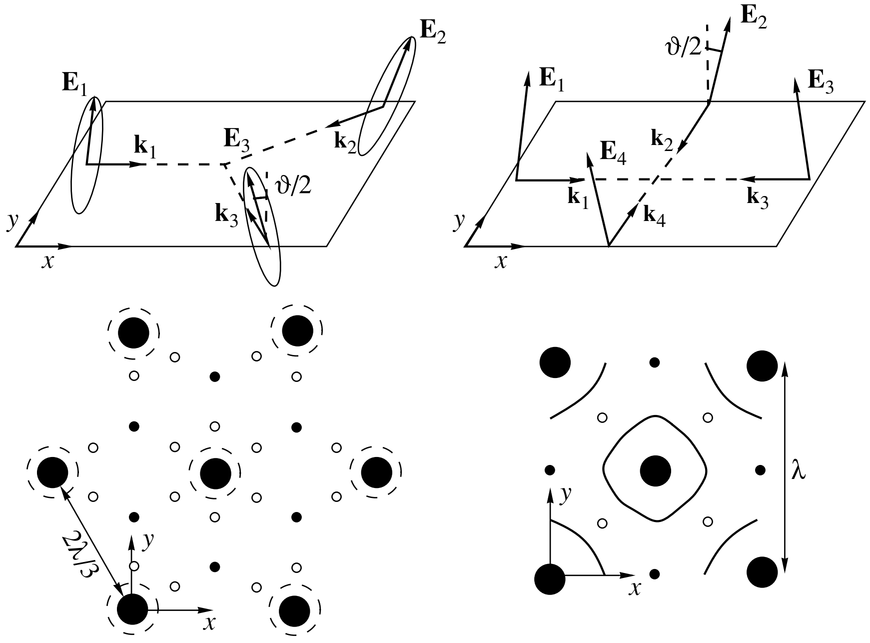
$$\hat{\mathcal{X}} \approx \frac{3\hbar}{2}\frac{(\mathbf{B}_2 - 2\tilde{\delta}\mathbf{B}_1) \otimes \mathbf{B}_1}{I_0^3\mathbf{R} \cdot \tilde{\mathcal{Q}}_3 \cdot \mathbf{R}}. \tag{23}$$

For other transitions, the asymptotics is similar but has different coefficients of the terms.

The proof of the finiteness of the diffusion tensor $\hat{\mathcal{D}}^{\text{ind}}$ near the points with linear and circular polarizations is more cumbersome. Let us consider the contributions from the most critical gradients only briefly. In the region of circular polarization, the terms with the coefficients D_{12}, D_{13}, D_{14}, D_{23}, and D_{34} are finite, because the combinations $c^2\mathbf{g}_{2,4}$ and $c\mathbf{g}_3$ are continuous and finite. The contributions with D_{22}, D_{24}, and D_{44} are finite both because the combinations $\mathbf{G} \otimes \mathbf{G}$ arise when the contributions are added and because the vectors $(c\mathbf{g}_{2,4})$ are finite. The finiteness of the contributions with $\mathbf{g}_{5,6}$ can be shown in a similar way. It should be noted that ambiguities similar to (23) also take place in the expression for the diffusion tensor both at the points of circular polarization (due to the contributions $c^2\mathbf{g}_{2,4} \otimes \mathbf{g}_{2,4}$ and $\mathbf{g}_3 \otimes \mathbf{g}_3$) and at the points of linear polarization (due to the contributions with $\mathbf{g}_{5,6}$).

6. LIGHT FIELD CONFIGURATIONS

We supplement and refine our analysis of the field invariants in singular regions for configuration dimensions $D > 1$ with characteristic examples that have the following common features: the initial light beams with the same amplitude $\tilde{E}$ are arranged in such a way that there is no drift of atoms that rapidly expels them from the region of interaction with the field. This implies that there should be no straightening of the dipole force, which is expressed by the condition $\langle \mathbf{F}_0(\mathbf{r}) \rangle_d = 0$ for the average force in any direction on the scale $d \gg \lambda$ corresponding to the beam diameter. These configurations are suitable and actively used in experiments (models A2, C, and D) on the formation of atomic lattices [2, 4] and their investigation. All of the configurations, except for case B, are formed by $D + 1$ light beams, so the structure of the invariants has a spatial periodicity character-



(a) (b)

Fig. 2. (Top) two-dimensional field configurations (a) A and (b) B; (bottom) the arrangement of the points corresponding to extremal invariants I_0 and I_1.

ized by reciprocal lattice vectors $\mathbf{s}_n = \mathbf{k}_1 - \mathbf{k}_{n+1}$, $n = 1, \ldots, D$ [18].

6.1. Two-Dimensional Field Configurations

(A) Let us consider the most general configuration composed of three beams with the same ellipticity $\tilde{\varepsilon}$ that satisfies these conditions. This configuration is a generalization of the one-dimensional ε–ϑ–$\bar{\varepsilon}$ configuration that was first suggested in [12]. Let the wave vectors $\mathbf{k}_n$ of the three traveling waves lie in the same $\{\mathbf{k}\}$ plane at an angle of 120° relative to one another (Fig. 2a) and the inclinations between the major polarization axis of each wave and the normal $\mathbf{e}_z$ to the plane be identical and equal to $\vartheta/2$. For this configuration, the invariants I_0 and I_1 are

$$I_0 = \tilde{E}^2[3 + C(Z^*Z - 3)], \quad Z = \sum_{n=1}^{3} \exp(i\mathbf{k}_n \cdot \mathbf{r}),$$

$$C = (\mathbf{e}_i \cdot \mathbf{e}_j^*)_{i \neq j},$$

$$I_1 = \tilde{E}^2[(1 + 2C)(Z^2 - 2Z^*) - 2H(Z^2 - 3Z^*)],$$

$$H = (\mathbf{e}_i \cdot \mathbf{e}_j)_{i \neq j},$$

where

$$\mathbf{e}_i = \frac{\mathbf{e}_z + \epsilon \cdot \mathbf{k}_i \times \mathbf{e}_z}{\sqrt{1 + |\epsilon|^2}}$$

is the polarization vectors of the ith beam and

$$\epsilon = \frac{t + ie}{1 - ite}$$

is specified by the parameters $\tilde{\varepsilon}$ and ϑ of the beam polarization ellipse. Below, we use the notation $e = \tan\tilde{\varepsilon}$ and $t = \tan\vartheta/2$.

Let us separate out the following values of ϑ and $\tilde{\varepsilon}$:

A1: for the ellipticity

$$\tan^2\tilde{\varepsilon}_0 = \frac{3\cos\vartheta + 1}{3\cos\vartheta - 1},$$

where $C = 0$ and the amplitude of the total field $E(\mathbf{r}) = \sqrt{3}\,\tilde{E}$ is constant ($\mathbf{g}_1 = 0$), while field (1) has isolated points with circular polarization defined by the condition $Z = 0$ (small filled circles in Fig. 2a) that are supplemented with the points (small open circles) for $\tilde{\varepsilon} = 0$ defined by the condition $Z^2 = 2Z^*$ at $Z \neq 0$;

A2: for $\vartheta \longrightarrow \pi$ and $\tilde{\varepsilon} \longrightarrow 0$ (the linear beam polarizations lie in the $\{\mathbf{k}\}$ plane), at the points denoted by small filled circles in Fig. 2a, the maxima of $E(\mathbf{r})$ coincide with the points of circular polarization, while the field nodes with linear polarization that are formed only in this and other limiting ($\vartheta = 0$) cases are located at the points denoted by large filled circles;

A3: for $\vartheta \sim \tilde{\varepsilon} \longrightarrow 0$ (the beam polarizations are nearly linear and almost perpendicular to the $\{\mathbf{k}\}$ plane), and the minima of E and the points of circular polarization are very close in the regions denoted by small filled circles in Fig. 2a, while the maxima of E are located at the points denoted by large filled circles.

The fixed (independent of $\tilde{\varepsilon}$ and ϑ) isolated points with circular polarization defined by the condition $Z = 0$ and denoted by small filled circles are the first peculiarity of the configuration. For linearly polarized beams, $\tilde{\varepsilon} = 0$, and at inclinations $\mathfrak{t} \gtrsim 4$, they are supplemented with the closed lines with circular (or nearly circular) polarization (marked by the dotted lines in Fig. 2a) that are formed around the fixed points with linear polarization $\mathbf{r}_{\text{lin}}$ (see below). Here, the curvature of the lines is

$$kr_0 \approx \frac{2}{\sqrt{3}} \arccos \frac{2\mathfrak{t}^2 - 1}{2(1 + \mathfrak{t}^2)}.$$

At smaller inclinations ϑ, these lines transform into figures of six isolated points in the shape of regular hexagons [19] (e.g., the open circles in case A1). As the ellipticity $\tilde{\varepsilon}$ of the initial beams increases, this structure is destroyed.

Near the fixed isolated points with circular polarization $\mathbf{r}_{\text{circ}}$ (below, we chose the point $\mathbf{r}_1 = (x_1 = \lambda/3, y_1 = \lambda/3\sqrt{3})$ in Fig. 2), the asymptotics is invariant,

$$I_0 \approx 9\tilde{E}^2[4(\mathfrak{e}^2 + \mathfrak{t}^2) + k^2(x^2 + y^2)(2 - \mathfrak{t}^2 - \mathfrak{e}^2 + 2(\mathfrak{e}\mathfrak{t})^2][8(1 + \mathfrak{e}^2)(1 + \mathfrak{t}^2)]^{-1},$$

and is indicative of extremal intensities at these points: intensity minima at $\tilde{\varepsilon} > \tilde{\varepsilon}_0$ and local intensity maxima at $\tilde{\varepsilon} < \tilde{\varepsilon}_0$. It follows from the asymptotics of the other invariant $I_1 \approx \mathbf{A} \cdot \mathbf{R}$ that

$$\mathbf{A} = \frac{9(K\tilde{E})^2}{4} \exp(i\varphi_0)(1 - \cos 2\tilde{\varepsilon} - \cos\vartheta)(\mathbf{e}_x + i\mathbf{e}_y),$$

where $\mathbf{e}_x$ and $\mathbf{e}_y$ are the Cartesian unit vectors and

$$\varphi_0 = -\frac{\pi}{6} - 2\arctan\frac{\mathfrak{t}}{\mathfrak{e}} + 2\arctan(\mathfrak{t}\mathfrak{e})$$

is a constant phase. Thus, the phase factor near $\mathbf{r}_1$ is $\exp(i2\Phi(\mathbf{r})) \approx \exp(i(\tilde{\varphi} + \varphi_0))$, where $\tilde{\varphi} = \arctan(y/x)$ is the polar angle. It can be shown that the signs of $\tilde{\varphi}$ alternate at the adjacent points $\mathbf{r}_{\text{circ}}$. This implies that the vector fields $\mathbf{g}_{2,4} \propto \pm\nabla\tilde{\varphi}$ near the lattice points $\mathbf{r}_{\text{circ}}$ are directed azimuthally and that the circulation direction of these fields is related to the polarization of the field $\mathbf{E}(\mathbf{r}_{\text{circ}})$.

The fixed isolated points with linear polarization $\mathbf{r}_{\text{lin}}$ marked by large filled circles, whose positions are defined by the condition $Z = 3$ or $\cos(\mathbf{k}_n \cdot \mathbf{r}_{\text{lin}}) = 1$, $n = 1, \ldots, 3$, are the second peculiarity of the configuration. Near these points (below, we chose $\mathbf{r}_0 = (x_0 = 0; y_0 = 0)$ in Fig. 2),

$$I_0 \approx 9\tilde{E}^2[4(1 + \mathfrak{e}^2\mathfrak{t}^2) - 9k^2(x^2 + y^2)(2 - \mathfrak{t}^2 - \mathfrak{e}^2 + 2(\mathfrak{e}\mathfrak{t})^2)][4(1 + \mathfrak{e}^2)(1 + \mathfrak{t}^2)]^{-1}.$$

These points are also the intensity extrema, but, in contrast to $\mathbf{r}_{\text{circ}}$, there will be local maxima for $\tilde{\varepsilon} > \tilde{\varepsilon}_0$ and minima for $\tilde{\varepsilon} < \tilde{\varepsilon}_0$ at them. Analysis of the fields $\mathbf{g}_{5,6}$ near these points shows their vortex structure: the vector $\mathbf{g}^{(5)} \propto [\mathbf{c}_1 \times \mathbf{r}]/r$ is finite but undefined at the point of linear polarization; the vector $\mathbf{g}^{(6)} \propto [\mathbf{c}_2 \times \mathbf{r}]/r^2$ asymptotically diverges, in accordance with the general results (11). Here,

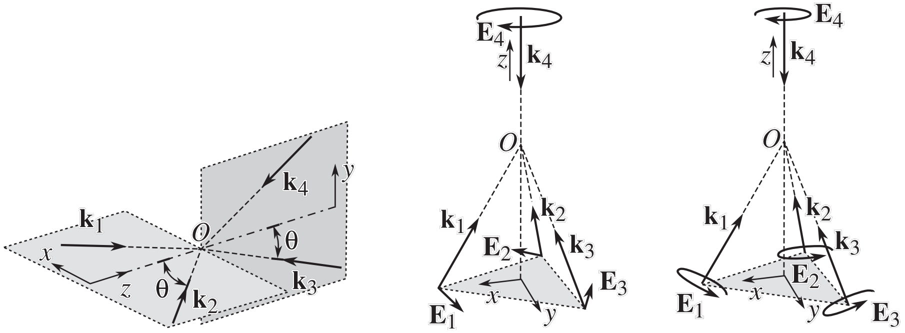
$$\mathbf{c}_1 = \frac{27\mathfrak{e}\mathfrak{t}(\mathfrak{e}^2 - 1)}{2\sqrt{(1 + \mathfrak{e}^2)^3(1 + \mathfrak{t}^2)(1 + (\mathfrak{e}\mathfrak{t})^2)}}\mathbf{e}_z,$$

$$\mathbf{c}_2 = \frac{2(\mathfrak{e}^2 - 1)\mathfrak{t}}{(1 + \mathfrak{t}^2)\mathfrak{e}}\mathbf{c}_1,$$

and $\mathbf{e}_z$ is the normal to the $\mathbf{k}$ plane.

It should also be noted that the field $\mathbf{g}_5$ exhibits more significant vortex structures at the saddle points $\mathbf{r}_{\text{sad}}$ of the $c(\mathbf{r})$ surface.

(B) In two-dimensional configurations formed by four or more beams, the structure of I_0 and I_1 significantly depends on the ratio of the beam phases $\tilde{\Phi}_i$ [18], which causes difficulties in the experimental implementation of time-stable configurations. In addition, the spatial structure of the invariant fields generally loses its periodicity [4]. On the other hand, for a certain choice of parameters, unique topological structures atypical of three-beam configurations become possible in principle. Let us consider the four-beam configuration that is a combination of two perpendicular lin–ϑ–lin configurations [11]: the beams have the same intensity, are linearly polarized, and are inclined at an angle of $\vartheta/2$ to the normal $\mathbf{e}_z$ to the $\{\mathbf{k}\}$ plane with mutually



(a) (b) (c)

Fig. 3. Three-dimensional field configurations (a) C, (b) D, and (c) E.

perpendicular adjacent wave vectors $\mathbf{k}_n$ (Fig. 2b). For this model,

$$I_0 = 4\tilde{E}^2\left\{\cos^2\frac{\vartheta}{2}[\cos^2(kx) + 2\cos(\Delta\Phi)\cos(kx)\cos(ky) + \cos^2(ky)] + \sin^2\frac{\vartheta}{2}[\sin^2(kx) + \sin^2(ky)]\right\},$$

$$I_1 = 2\tilde{E}^2(\cos\vartheta + \cos(2kx) + 4\exp(i\Delta\Phi) \times \cos^2\frac{\vartheta}{2}\cos(kx)\cos(ky) + \exp(2i\Delta\Phi) \times [\cos\vartheta + \cos(2ky)]),$$

where $\Delta\Phi$ is the relative phase shift between the lin–ϑ–lin configurations. Whereas the relative change of $\tilde{\Phi}_i$ in a three-beam configuration leads to a simple shift of the entire spatial structure of the invariants in the $\{\mathbf{k}\}$ plane, now the topological field structure of the invariants changes with $\Delta\Phi$.

At $\Delta\Phi = \pi/2$ and $\vartheta = \pi/2$, the field shape is similar to case *A*: circular polarization takes place at the isolated points $\mathbf{r}_{\text{circ}}$ (Fig. 2b, small open circles), the vector fields $\mathbf{g}_{2,4}$ have the shape of vortices with the centers at $\mathbf{r}_{\text{circ}}$, and the intensity distribution is uniform ($E = 2\tilde{E}$).

At $\Delta\Phi = 0$, the vector fields $\mathbf{g}_2$ and $\mathbf{g}_4$ disappear, which, for example, implies the disappearance of the corresponding contributions to the force $\mathbf{F}_0$ (18), called the light pressure force; this cannot be achieved in a three-beam configuration. The intensity distribution is nonuniform: the nodes and the maxima in Fig. 2b are indicated by small and large filled circles, respectively. The regions of circular polarization have the shape of closed $\tilde{I}_1(x, y) = 0$ lines defined by the function

$$\tilde{I}_1(x, y) = \cos(2kx) + 2\cos(kx)\cos(ky) + \cos(2ky).$$

This corresponds to the case where an ellipticity parameter $\tilde{c}$ in the form of a smooth alternating function

$$\tilde{c} = \frac{\tilde{I}_1(x, y)}{2[1 + \cos(kx)\cos(ky)]}$$

for the entire region of the field configuration can be introduced.

6.2. Three-Dimensional Field Configurations

Configurations C, D, and E given below (Fig. 3) show a wider topological variety of possible spatial structures of the field invariants even in relatively simple four-beam three-dimensional models. Cases C and D are well known [2, 4] and have been commonly used to study the various processes in light-induced spatial atomic lattices.

(C) This configuration is formed by four linearly polarized beams of the same intensity: one pair of beams with polarizations along the *y* axis have wave vectors $\mathbf{k}_{1,2}$ in the *xz* plane with the same angle ϑ between $\mathbf{k}_i$ ($i = 1, 2$) and the *z* axis, and the other pair of beams with polarizations along the *x* axis have wave vectors $\mathbf{k}_{3,4}$ located in the *yz* plane at the same angle to the *z* axis and are directed toward the first pair (Fig. 3a). The gradients $\mathbf{g}_{5,6}$ are absent in the configuration, because there is the symmetry *z* axis along which the vector $\mathbf{e}_0$ defined in (5) is always directed.

In this model, the invariants are

$$I_0 = 2\tilde{E}^2[2 - \cos(2X) - \cos(2Y)],$$

$$I_1 = 2\tilde{E}^2\{\cos(2Z)[\cos(2X) + \cos(2Y) - 2] + i\sin(2Z)[\cos(2X) - \cos(2Y)]\},$$

where $X = kx\sin\vartheta$, $Y = ky\sin\vartheta$, and $Z = kz\cos\vartheta$.

The singular regions also have a simple shape: the field nodes lie on the straight lines specified by the conditions $\cos(2X) = \cos(2Y) = 1$, while the maxima lie at the straight lines $\cos(2X) = \cos(2Y) = -1$. The regions of circular polarization are specified by the straight lines $\cos(2Z) = \cos(2X) - \cos(2Y) = 0$. The regions of linear polarization are surfaces. The relative positions of the singular regions are shown in Fig. 4, where the part of the periodic structure corresponding to the halves of the spatial periods in each direction is considered. The figure shows the structure (3) of the $c(\mathbf{r}) = 0.1$ surface around the line of circular polarization, which demonstrates the significant difference between the ellipticity gradients near the nodes (1) and the intensity maxima (2).

Of considerable interest are the asymptotics of the invariants near the nodes. For example, near the z axis containing nodes,

$$I_0 \approx 4\tilde{E}^2(X^2 + Y^2),$$

$$I_1 \approx 4\tilde{E}^2[i\sin(2Z)(Y^2 - X^2) - \cos(2Z)(Y^2 + X^2)].$$

For the ellipticity parameter and the total phase, we have

$$c = [1 - \sin^2(2\tilde{\varphi})\sin^2(2Z)]^{1/2},$$

$$\tan(2\Phi) = \cos(2\tilde{\varphi})\tan(2Z),$$

where $\tan\tilde{\varphi} = Y/X$. Near the points of intersection of the node straight lines and the straight lines corresponding to circular polarization, the phase factor is

$$\exp(2i\Phi) = i\,\mathrm{sgn}(\tilde{Y}^2 - \tilde{X}^2)$$

and the ellipticity parameter is $c \approx |\cos(2\tilde{\varphi})|$. It follows from their form that, when the signs of $\exp(2i\Phi)$ and $c(\mathbf{r})$ are reconciled and if the relation $c\exp(2i\Phi) = I_1/I_0$ holds, it is more convenient to define an alternating continuous ellipticity parameter $-1 \le \tilde{c} \le 1$ in this case.[4] For example, we may choose $\tilde{c} \approx \cos(2\tilde{\varphi})$ for $\exp(2i\Phi) = -i$. As a result, vector $\mathbf{g}_3$ diverges as $1/r$ at the degenerate point, while the magnitudes of $\mathbf{g}_{2,4}$ are negligible.

(D) The four-beam configuration with the wave vectors $\mathbf{k}_i$ that form a tetrahedron [2] is more complex: the circularly polarized wave $\mathbf{k}_4$ is aligned with the z axis; the remaining waves are linearly polarized, with their polarizations lying in the xy plane (Fig. 3b). Since the vector $\mathbf{e}_0$ (5) is along the symmetry z axis, the gradients $\mathbf{g}_{5,6}$ are absent here, as in the previous case. The invariants are

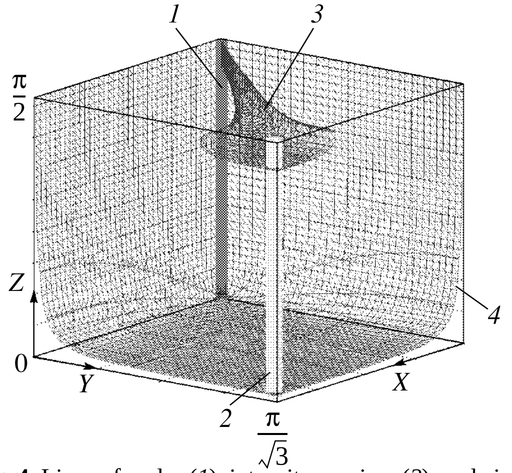


Fig. 4. Lines of nodes (*1*), intensity maxima (*2*), and circular polarization (*3*), and the surface of linear polarization (*4*) in configuration C.

$$I_0 = \tilde{E}^2\left[4 - \sum_{m<n}\cos\left((\mathbf{s}_m - \mathbf{s}_n)\cdot\mathbf{r} + (-1)^{m+n}\frac{2\pi}{3}\right) + \sqrt{2}\sum_m \cos(\mathbf{s}_m\cdot\mathbf{r})\right],$$

$$I_1 = \tilde{E}^2\exp\left(-\frac{3i\tilde{\Phi}_D}{2}\right) \times\left\{\sum_{m<n}\exp\left[i\left((\mathbf{s}_m + \mathbf{s}_n)\cdot\mathbf{r} - \frac{2\pi(m+n)}{3}\right)\right]\right\} \times\left\{\sum_{m<n}\exp[i(\mathbf{s}_m + \mathbf{s}_n)\cdot\mathbf{r}] + \sqrt{2}\exp(i\tilde{\Phi}_D)\right\},$$

where $\mathbf{s}_m = \mathbf{k}_4 - \mathbf{k}_m$ ($m = 1, \ldots, 3$) are the reciprocal lattice vectors [18],

$$\tilde{\Phi}_D = \sum_m \mathbf{s}_m\cdot\mathbf{r}.$$

There are no nodes in this configuration, while the intensity minima, $I_0 \approx 0.35\tilde{E}^2$, and maxima, $I_0 = (11 + 6\sqrt{2})\tilde{E}^2/2$, lie at isolated points. The regions of circular polarization represent lines of two types. Regions I defined by the equation

$$\sum_{m<n}\cos\left((\mathbf{s}_m - \mathbf{s}_n)\cdot\mathbf{r} + (-1)^{m+n}\frac{2\pi}{3}\right) = \frac{3}{2}$$

---

[4] The above configuration B at $\Delta\Phi = 0$ is also this case.

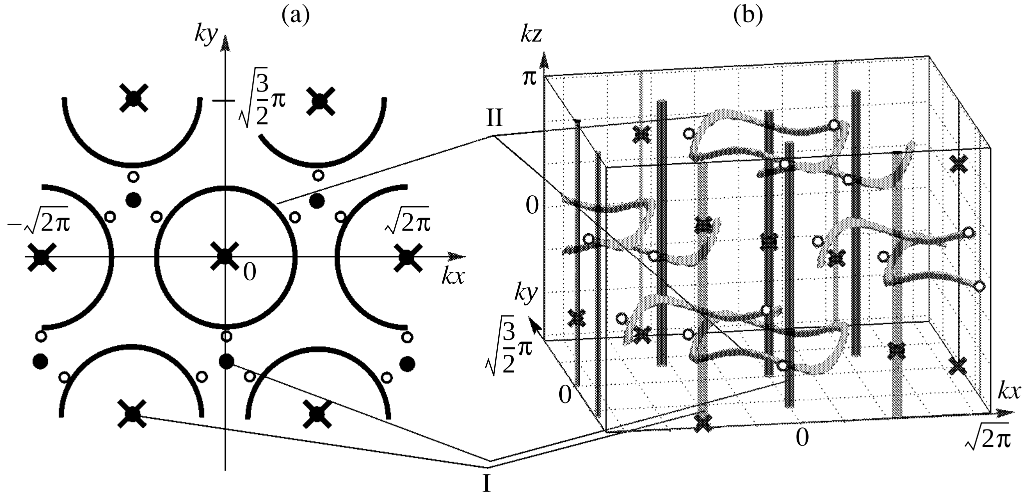


Fig. 5. Points of intensity minima (open circles) and maxima (crosses) and lines of circular polarization in configuration D: (a) top view and (b) side view.

are straight lines parallel to the z axis. Regions II defined by the equation

$$\sqrt{2}\sum_{m}\cos(\mathbf{s}_m\cdot\mathbf{r}) + \sum_{m<n}\cos((\mathbf{s}_m-\mathbf{s}_n)\cdot\mathbf{r}) = -\frac{5}{2}$$

are nonplane closed lines whose projections onto the xy plane are circumferences of radius $kr_0 \approx 1.75$. The relative positions of the singular regions are shown in Fig. 5. Note that the intensity minima are close to regions II, while the intensity maxima lie only on lines I surrounded by regions II. The regions of linear polarization, which are surfaces with a complex structure, are not shown.

(E) Finally, let us consider a symmetrized four-beam configuration with the wave vectors $\mathbf{k}_i$ directed in the same way as in the previous case and with the same circular polarizations for all waves (Fig. 3c). All types of gradients $\mathbf{g}_i$ ($i = 1, \ldots, 6$) are available here. Let the origin O be chosen so that $\mathbf{E}_m\cdot\mathbf{E}_n = -\tilde{E}^2/3$ at $m \neq n$ for the components $\mathbf{E}_m$ ($m = 1, \ldots, 4$) of the configuration beams. The invariants

$$I_0 = \tilde{E}^2\left[4-\frac{2}{3}\sum_{m<n}\cos((\mathbf{k}_m-\mathbf{k}_n)\cdot\mathbf{r})\right], \tag{24}$$

$$\begin{aligned} I_1 &= \frac{2\tilde{E}^2}{3}\sum_{m\neq n}\exp[i(\mathbf{k}_m+\mathbf{k}_n)\cdot\mathbf{r}+i\tilde{\Phi}_{mn}], \\ &= \frac{8\tilde{E}^2}{3}\left[\exp\left(-\frac{2\pi i}{3}\right)\cos((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{r})\right. \\ &\left.+\cos((\mathbf{k}_1+\mathbf{k}_3)\cdot\mathbf{r})+\exp\frac{2\pi i}{3}\cos((\mathbf{k}_1+\mathbf{k}_4)\cdot\mathbf{r})\right] \end{aligned} \tag{25}$$

by virtue of the relations

$$\sum_{m=1}^{4}\mathbf{k}_m = 0, \quad \mathbf{E}_m\cdot\mathbf{E}_m = 0.$$

Here, the phase factors

$$\tilde{\Phi}_{mn} = \frac{\pi}{3}((-1)^{[(m+n)/2]}-(-1)^{[(m+n+1)/2]})$$

are defined via the operation $[a]$ of taking an integer part of the number a.

Analysis of invariants (24) and (25) allows the distribution of singular regions to be determined. The field nodes lie at the isolated points and form a body-centered lattice $\{\mathbf{r}_{\min}\}$ specified, according to (24), by relations of

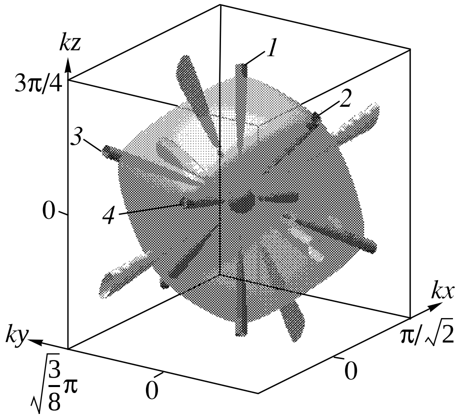


Fig. 6. The field node (at the center), the ray directions with circular (*1*–*4*) and linear (remaining) polarizations, and the surface of maximum intensity around the node in configuration E.

the form $\cos(\mathbf{s}_n \cdot \mathbf{r}_{\min}) = 1$, where $\mathbf{s}_n$ at $n = 1, \ldots, 3$ are the reciprocal lattice vectors ($\mathbf{s}_n$ coincide in configurations D and E). The regions of circular polarization are straight lines specified, according to (25), by the equations

$$\cos((\mathbf{k}_1 + \mathbf{k}_2) \cdot \mathbf{r}) = \cos((\mathbf{k}_1 + \mathbf{k}_3) \cdot \mathbf{r}) = \cos((\mathbf{k}_1 + \mathbf{k}_4) \cdot \mathbf{r}).$$

Four such straight lines pass through each node that is an isolated point (the $c(\mathbf{r}) = 0.1$ surfaces are shown in Fig. 6). The regions of linear polarization are also straight ray lines L. The three straight lines corresponding to linear polarization are mutually perpendicular and specify the rays perpendicular to the faces of an imaginary cube (the $c(\mathbf{r}) = 0.995$ surfaces are shown). Here, we can also easily see a cubic symmetry of the surfaces corresponding to the maximum field intensity. The asymptotics of the invariants in the region of nodes are (we consider the point $\mathbf{r}_{\min} = 0$ and chose the reciprocal lattice vectors $\mathbf{s}_m = \mathbf{k}_4 - \mathbf{k}_m$, $m = 1, \ldots, 3$)

$$I_0 \approx \tilde{E}^2 \left[\sum_n (\mathbf{s}_n \cdot \mathbf{r})^2 - \frac{2}{3} \sum_{m<n} (\mathbf{s}_m \cdot \mathbf{r})(\mathbf{s}_n \cdot \mathbf{r}) \right], \tag{26}$$

$$I_1 \approx -\frac{4\tilde{E}^2}{3} \left[\exp\frac{2\pi i}{3} (\mathbf{s}_1 \cdot \mathbf{r})(\mathbf{s}_2 \cdot \mathbf{r}) + (\mathbf{s}_1 \cdot \mathbf{r})(\mathbf{s}_3 \cdot \mathbf{r}) + \exp\left(-\frac{2\pi i}{3}\right) (\mathbf{s}_2 \cdot \mathbf{r})(\mathbf{s}_3 \cdot \mathbf{r}) \right]. \tag{27}$$

Their representations in Cartesian coordinates are

$$I_0 \approx \frac{16(k\tilde{E})^2 (x^2 + y^2 + z^2)}{9},$$

$$I_1 \approx \frac{8\exp\left(\frac{i\pi}{3}\right) (k\tilde{E})^2 ((x - iy)^2 - 2\sqrt{2} z (x + iy))}{9},$$

while the representation of the ellipticity parameter c in the corresponding spherical coordinates,

$$c = \frac{|\sin\theta|}{2\sqrt{2}} [9 + 7\cos(2\theta) + 2\sqrt{2}\sin(3\varphi - 2\theta) - 2\sqrt{2}\sin(3\varphi + 2\theta)]^{1/2}, \tag{28}$$

shows its complex structure at a singular point. At the node $\mathbf{r}_{\min} = 0$, the four directions of circular polarization are specified by the spherical angles $1(\theta = 0; \varphi)$, $2(\alpha_0; 0)$, $3(\alpha_0; 2\pi/3)$, and $4(\alpha_0; -2\pi/3)$ (Fig. 6), where $\cos\alpha_0 = 1/3$, and correspond to the directions of the diagonals in a body-centered lattice. The explicit form of (27) and (28) shows that the gradient invariants $\mathbf{g}_2$, $\mathbf{g}_3$, and $\mathbf{g}_4$ diverge near the nodes at $1/r$.

In the above three-dimensional configurations, the common features in the spatial structure of the vector invariants $\mathbf{g}_n$ ($n = 1, \ldots, 6$) are: the vortex structure of the phase gradients $\mathbf{g}_{2,4}$ around the lines of circular polarization and the vortex structure of the field $\mathbf{g}_6$ around the lines of linear polarization (case E).

7. CONCLUSIONS

Let us summarize the main points of our analysis. There are singular regions in field configurations of dimension $D > 1$ in which the vector field invariants are either undefined or diverge: for the gradients of the total phase and the rotation angle of the polarization ellipse around its axis, these are the regions of circular polarization where their fields generally have a vortex structure and diverge as $1/r$; for the gradient of the rotation angle of the minor axis of the polarization ellipse, these are the regions of linear polarization where its field also has a vortex structure and diverges as $1/r$. The field nodes are degenerate regions. The field ellipticity in them is undefined, and the ellipticity gradient diverges as $1/r$; the regions of circular and linear polarizations can intersect here. Near the nodes, the semiclassical approximation is inapplicable in analyzing the kinetics of atoms.

By analyzing the structure of the invariants in configurations of dimension $D > 1$, we can qualitatively distinguish the main features of the kinetics of atomic ensembles, which follows from the structure of the friction force $\mathbf{F}_0$ and tensor $\mathcal{X}$ (18). Aspects related to the cooling and trapping mechanisms and the physical interpretation of the various contributions in these expansions were discussed in [7, 20]. For detunings $|\delta| \gtrsim 10\gamma$, the gradients of amplitude $\mathbf{g}_1$ and ellipticity $\mathbf{g}_3$ give the largest contribution to $\mathbf{F}_0$ and $\mathcal{X}$. It follows from the explicit form (21) for the simplest $1/2 \longrightarrow 1/2$ transition that a steady-state kinetic regime is possible only for detunings $\delta > 0$, when radiative cooling takes place.[5] This is a common property of $J \longrightarrow J$ transitions with half-integer J. It then follows from the explicit form of $F_{1,3}$ from (20) and the properties of $\mathbf{g}_{1,3}$ that atoms are attracted either to the intensity minima or to the regions of circular polarization, while they are expelled from the regions of maxima and linear polar-

---

[5] The coefficient $\mathcal{X}_{33} < 0$, which is attributable to a manifestation of the Sisyphus cooling mechanism [10, 11].

ization. For $J \longrightarrow J + 1$ transitions, the situation is different [7]: cooling takes place for $\delta < 0$, the regions of circular polarizations and the intensity maxima are the localization centers, and the intensity minima and the regions of linear polarization are the expulsion centers.

Our numerical simulations of the dynamics of atoms with the $1/2 \longrightarrow 1/2$ and $1/2 \longrightarrow 3/2$ transitions described by the Langevin equation corresponding to Eq. (17) for the above two-dimensional configurations A and B showed [21] that the degree of atomic localization near the intensity extrema is appreciably higher than that in the region of circular polarization if these singular regions are spatially separated. The remaining components of $\mathbf{F}_0$, called the light pressure force, may have an effect here. In the regions of circular polarization, being proportional to the vectors $\mathbf{g}_{2,4}$, they have a vortex structure and reduce the localization effects. The superposition of the localization and expulsion produced by various gradients can lead to a multiply connected topological structure of the localization regions. For example, if the center of attraction due to the gradient $\mathbf{g}_1$ ($\mathbf{g}_3$) coincides with the center of repulsion due to the gradient $\mathbf{g}_3$ ($\mathbf{g}_1$), then the localization regions in two-dimensional configurations have the shape of a ring at this point [21]. Obviously, the largest effect in point localization is achieved at the points of intersection of circular polarization with the corresponding intensity extrema. In this sense, the $J \longrightarrow J + 1$ transitions have a significant advantage, because the intersection of the intensity maxima with the regions of circular polarizations is easily realizable in both two-dimensional (A2) and three-dimensional configurations where there are such points of intersection[6] in models C, D, and E. These regions, which are quite stable with respect to the fluctuations that form the field configuration, determine the main structure of the dissipative lattices currently implemented in experiments [2, 4]. In contrast, at the intensity nodes, the ellipticity is undefined and has a large gradient in their vicinity, while their positions are sensitive to fluctuations of the initial beams.

The force and diffusion contributions that originate from the other vector invariants $\mathbf{g}_2$, $\mathbf{g}_4$, and $\mathbf{g}_{5,6}$ affect the dynamics of atoms. These effects are noticeable even for detunings $|\delta| \sim 10\gamma$, which is attributable to the large values of these gradients in certain configuration regions. As a result, additional structures are formed in the density of atomic lattices with a topology that does not reduce to localization points and localized vortex particle flows emerge. Their influence significantly decreases the mean time of atomic confinement in the localization region. The dynamic aspects can be highly varied here.

The method of invariants allows us to approach the problem of a probe light field in a different way. This problem is of considerable importance in spectroscopy of atomic lattices and Bragg scattering [4], photorefraction by atomic lattices [23], etc. Clearly, even at low probe field intensities, the structure of the invariants for the general field can change qualitatively. First, the addition of an extra field disrupts their spatial periodicity. Second, the degenerate points (nodes) can disappear (or appear at new positions). Third, there is the problem of estimating the influence of fluctuations in parameters of the beams that form a configuration on the configuration properties and, eventually, on the kinetic and polarization parameters of atomic ensembles.

ACKNOWLEDGMENTS

This work was supported in part by the Program of the Ministry of Education of Russia “Basic Research in the Higher School in the Field of Natural and Humanitarian Sciences: Universities of Russia” (project no. UR.01.01.062).

REFERENCES

1. G. Grynberg, C. Triché, L. Guidoni, *et al.*, Europhys. Lett. **51**, 506 (2000); E. Demler and F. Zhou, Phys. Rev. Lett. **88**, 163001 (2002).
2. G. Grynberg, B. Lounis, P. Verkerk, *et al.*, Phys. Rev. Lett. **70**, 2249 (1993).
3. L. Guidoni, C. Triché, P. Verkerk, *et al.*, Phys. Rev. Lett. **79**, 3363 (1997).
4. G. Grynberg and C. Robilliard, Phys. Rep. **355**, 335 (2001).
5. D. L. Haycock, S. E. Hamman, G. Klose, *et al.*, Phys. Rev. A **55**, R3991 (1997); D. Boiron, A. Michaud, J. M. Fournier, *et al.*, Phys. Rev. A **57**, R4106 (1998); S. Friebel, C. D’Andrea, J. Walz, *et al.*, Phys. Rev. A **57**, R20 (1998).
6. J. Dalibard and C. Cohen-Tannoudji, J. Phys. B **18**, 1661 (1985); J. Javanainen, Phys. Rev. A **44**, 5857 (1991).
7. A. V. Bezverbnyĭ, O. N. Prudnikov, A. V. Taĭchenachev, *et al.*, Zh. Éksp. Teor. Fiz. **123**, 437 (2003) [JETP **96**, 383 (2003)].
8. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 7th ed. (Nauka, Moscow, 1988; Pergamon Press, Oxford, 1975).
9. G. Nienhuis, P. van der Straten, and S.-Q. Shang, Phys. Rev. A **44**, 462 (1991).
10. J. Dalibard and C. Cohen-Tannoudji, J. Opt. Soc. Am. B **6**, 2023 (1989).
11. V. Finkelstein, P. R. Berman, and J. Guo, Phys. Rev. A **45**, 1829 (1992).
12. O. N. Prudnikov, A. V. Taĭchenachev, A. M. Tumaĭkin, and V. I. Yudin, Pis’ma Zh. Éksp. Teor. Fiz. **70**, 439 (1999) [JETP Lett. **70**, 443 (1999)].

---

[6] An example of a configuration where there are no such points of intersection, but the localization and sub-Doppler cooling effects also take place, is considered in [22].

13. A. V. Taĭchenachev, A. M. Tumaĭkin, V. I. Yudin, and G. Nienkhaus, Zh. Éksp. Teor. Fiz. **114**, 125 (1998) [JETP **87**, 70 (1998)].
14. A. V. Bezverbnyĭ, Zh. Éksp. Teor. Fiz. **118**, 1066 (2000) [JETP **91**, 921 (2000)].
15. D. A. Varshalovich, A. I. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Sci., New York, 1988).
16. N. L. Manakov, S. I. Marmo, and A. V. Meremianin, J. Phys. B: At. Mol. Opt. Phys. **29**, 2711 (1996).
17. N. L. Manakov, A. V. Meremianin, and A. Starace, Phys. Rev. A **61**, 022103 (2000).
18. K. I. Petsas, A. B. Coates, and G. Grynberg, Phys. Rev. A **50**, 5173 (1994).
19. A. V. Bezverbny, G. Nienhuis, and A. M. Tumaikin, Opt. Commun. **148**, 151 (1998).
20. A. V. Bezverbnyĭ, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 162 (2001) [JETP Lett. **74**, 144 (2001)].
21. A. V. Bezverbnyĭ, Izv. Vyssh. Uchebn. Zaved., Fiz. **46** (5), 7 (2003).
22. C. Mennerat-Robilliard, L. Guidoni, K. I. Petsas, *et al.*, Eur. Phys. J. D **1**, 33 (1998).
23. S. Guibal, C. Mennerat-Robilliard, D. Larousserie, *et al.*, Phys. Rev. Lett. **78**, 4709 (1997).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 890–914.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 996–1022.
Original Russian Text Copyright © 2003 by Karnakov, Popov.

ATOMS, SPECTRA, RADIATION

A Hydrogen Atom in a Superstrong Magnetic Field and the Zeldovich Effect

B. M. Karnakov[a] **and V. S. Popov**[b]
[a]Moscow Institute of Engineering Physics, Kashirskoe sh. 31, Moscow, 115409 Russia
e-mail: karnak@theor.mephi.ru
[b]Institute for Theoretical and Experimental Physics, ul. Bol'shaya Cheremushkinskaya 25, Moscow, 117218 Russia
e-mail: markina@heron.itep.ru

Received April 23, 2003

Abstract—We consider the problem of a hydrogen atom in a superstrong magnetic field, $B \gg B_a = 2.35 \times 10^9$ G. The analytical formulas that describe the energy spectrum of this atom are derived for states with various quantum numbers n_ρ and m. A comparison with available calculations shows their high accuracy for $B \gg B_a$. We note that the derived formulas point to a manifestation of the Zeldovich effect, i.e., a rearrangement of the atomic spectrum under the influence of strong short-range Coulomb potential distortion. We discuss the relativistic corrections to level energies, which increase in importance with magnetic field and become significant for $B \gtrsim 10^{14}$ G. We suggest the parameters in terms of which the Zeldovich effect has the simplest form. Analysis of our precision numerical calculations of the energy spectrum for a hydrogen atom in a constant magnetic field indicates that the Zeldovich effect is observed in the spectrum of atomic levels for superstrong fields, $B \gtrsim 5 \times 10^{11}$ G. Magnetic fields of such strength exist in neutron stars and, possibly, in magnetic white dwarfs. We set lower limits for the fields $B_{\min}$ required for the manifestation of this effect. We discuss some of the properties of atomic states in a superstrong magnetic field, including their mean radii and quadrupole moments. We calculated the probabilities of electric dipole transitions between odd atomic levels and a deep ground level.

1. INTRODUCTION

A peculiar effect of rearrangement of the energy spectrum, which was first pointed out in [1] and was subsequently called the Zeldovich effect, arises under certain conditions in systems for which the interaction potential breaks up into two parts with highly incommensurable radii. In this paper, we show that this effect manifests itself in the spectrum of hydrogen atomic levels in an extremely strong magnetic field, $B \gg B_a = 2.35 \times 10^9$ G. This problem is of great interest in astrophysics, solid-state physics, and atomic physics and has been considered by many authors, starting from the pioneering paper by Schiff and Snyder [2]. Magnetic fields of such strength are characteristic of neutron stars [3, 4] and magnetic white dwarfs. Dozens of studies in which the energy spectrum of this atom was calculated by various numerical methods have been published (see, e.g., [5–14] and references therein).

Nevertheless, as far as we know, there are no correct analytical expressions for the energy spectrum of a hydrogen atom in a strong magnetic field in the literature. The standard formula for the ground-state binding energy

$$\varepsilon_0(B) \equiv \frac{\hbar^2\lambda}{2m_e} = \frac{m_e e^4}{2\hbar^2}\ln^2\frac{B}{B_a} \qquad (1.1)$$

[15, §112], even for $B \sim 10^{13}$ G, i.e., at the validity boundary of the nonrelativistic approximation, gives only order-of-magnitude values of ε_0, which are a factor of 3 larger than the result of an accurate numerical calculation. Moreover, the dependence that is the inverse of (1.1),

$$B(\lambda) = B_a \exp(\lambda a_B), \qquad (1.2)$$

which is of particular interest for astrophysical applications, is completely unsatisfactory, because it gives $B(\lambda)$ underestimated by two orders of magnitude. The paper by Hasegawa and Howard [16], who derived a formula for the ground state of a hydrogen atom that is asymptotically exact in the limit $B \longrightarrow \infty$, constitutes an exception. However, this formula also has a low accuracy for $B \lesssim 10^{15}$ G (see Table 1) and is inapplicable to excited states.

This discrepancy stems from the fact that both the derivation of formula (1.1) and some of the remarks on the properties of the spectrum of excited hydrogen atomic states given in [15] require significant refinements. In particular, even and odd (with respect to the reflection of the electron coordinate along the magnetic field) levels are described by distinctly different expressions and are greatly shifted relative to each other, which is clearly seen from Fig. 1. In this case, the

Table 1. The ground level of a hydrogen atom in a strong magnetic field

λ^2	$\mathcal{H}$				$\frac{a_H}{a_B}$
	Number	Eq. (2.13)	[16]	Eq. (2.16)	
7.5796	100	80	>	128	0.10
9.4543	200	174	>	250	0.071
11.703	400	368	>	488	0.050
15.325	1000	962	>	1181	0.032
18.610	2000	1962	>	2.31(3)	0.022
22.408	4000	3966	>	4.53(3)	0.016
23.747	5000	4970	>	5.63(3)	0.014
28.282	1.0(4)	9991	>	1.11(4)	0.010
32.92	1.878(4)	1.880(4)	3.64(4)	2.05(4)	7.3(–3)
41.159	5.0(4)	5.014(4)	8.96(4)	5.37(4)	4.5(–3)
47.783	1.0(5)	1.003(5)	1.71(5)	1.06(5)	3.2(–3)
55.062	–	2.0(5)	3.27(5)	2.10(5)	2.2(–3)
65.805	–	5.0(5)	7.82(5)	5.21(5)	1.4(–3)
74.772	–	1.0(6)	1.52(6)	1.047(6)	1.0(–3)
98.740	–	5.0(6)	7.16(6)	5.146(6)	4.5(–4)
109.945	–	1.0(7)	1.40(7)	1.024(7)	3.2(–4)

Note: λ^2 is the binding energy of the ground $1s(1^+)$ state in rydbergs; $\mathcal{H}$ is the reduced magnetic field, $\mathcal{H} = 18777$ corresponds to $B = B_{\text{cr}}$; in all tables, $a(b) \equiv a \times 10^b$; the second column gives the results of our numerical calculations. The ">" sign means that the corresponding value has an error larger than 100%.

hydrogen atomic spectrum, $E_n^{(0)} = -m_e e^4/2\hbar^2 n^2$, is rearranged in even states under the influence of short-range ($r \lesssim r_0 \ll a_B$) Coulomb potential distortion. The possible manifestation of this effect imposes stringent requirements on the pattern of Coulomb potential distortion: for slow particles, it must be resonant in pattern [1].

Let us briefly describe the content of this paper. In Section 2, we derive the formulas that describe the energy spectrum of a nonrelativistic hydrogen atom in a superstrong magnetic field. Their comparison with available numerical calculations for states with various magnetic quantum numbers in Section 3 shows that these asymptotic formulas are highly accurate for fields $B \gg B_a$. The relativistic corrections to the level spectrum are discussed in Section 4.

In Section 5, we consider the general properties of the Zeldovich effect and point out the variables in terms of which it has the most natural form. Our numerical calculations of the energy spectrum for a hydrogen atom in a magnetic field are analyzed in Section 6. The correction of the lower-level binding energies determined in [9–11] for the quantum defects δ_{nl} strongly suggests that the Zeldovich effect (which is actually contained in the results of numerical calculations of these works, but was overlooked by their authors) is present in the spectrum of atomic levels in magnetic fields $B \gtrsim 5 \times 10^{11}$ G. We set a lower limit for the magnetic fields B_{min} required for this effect to emerge. In Section 7, we calculate the mean radii of *ns*-states, the quadrupole moment of the ground level, and the probabilities of electric dipole transitions to the 1*s* ground level when $B \gg B_{\text{min}}$. In Section 8, we discuss the peculiarities of the Zeldovich effect in a superstrong magnetic field and in a three-dimensional potential, the relationship between Eqs. (2.13) and (5.1), etc. In the Appendices, we give details of our calculations, discuss the asymptotic formulas for the ground-state energy for $\mathcal{H} \longrightarrow \infty$ and the peculiar features of a one-dimensional hydrogen atom, and give formulas related to the Whittaker function.

In this paper, we use the following notation: $\varepsilon = \hbar^2\lambda^2/2m_e$ is the level binding energy, $a_B = \hbar^2/m_e e^2$ is the Bohr radius, $B_a = e^3 m_e^2 c/\hbar^3$ is the atomic unit of magnetic field, $\mathcal{H} = B/B_a$ is the reduced magnetic field, $B_{\text{cr}} = m_e^2 c^3/e\hbar = \alpha^{-3}B_a = 4.414 \times 10^{13}$ G is the critical or Schwinger magnetic field [17], $m = 0, \pm1, \ldots$ is the magnetic quantum number, $\omega_L = eB/m_e c$ is the cyclotron or Larmor frequency, $\alpha_H = \sqrt{\hbar/m_e\omega_L} = \sqrt{\hbar c/eB}$ is the magnetic length or Landau radius, and $\hbar = m_e =$

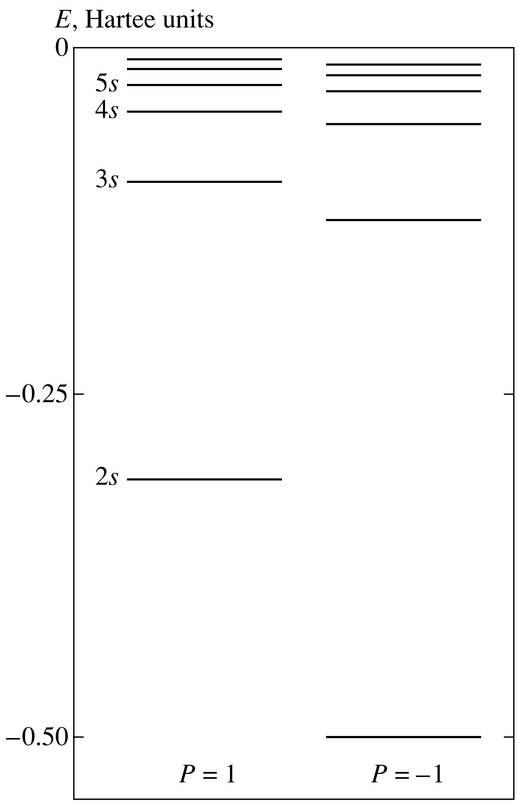


Fig. 1. The energy spectrum of a hydrogen atom for $\mathcal{H} = 3120$, or $B \approx 0.73 \times 10^{13}$ G. The series of even and odd levels ($P = \pm 1$) are shown. The 2*s* level is halfway between the odd levels with $n = 1$ and 2; the binding energy of the ground 1*s* level is 6.98 at. units ≈190 eV.

$e = 1$ (the Hartree atomic units, in which $c = \alpha^{-1} = 137$ is the speed of light).

The results presented below were partly announced in [18, 19].

2. ASYMPTOTIC FORMULAS FOR THE ENERGY SPECTRUM OF A HYDROGEN ATOM IN A STRONG MAGNETIC FIELD

Choosing the vector potential of the external magnetic field to be $\mathbf{A} = \frac{1}{2}[\mathbf{B} \cdot \mathbf{r}]$ and directing the *z* axis along **B**, we have the following nonrelativistic Hamiltonian of the system under consideration:

$$\hat{H} = -\frac{1}{2}\Delta_\perp + \frac{1}{8}\mathcal{H}^2\rho^2 + \frac{1}{2}\mathcal{H}(\hat{l}_z + \hat{\sigma}_z) - \frac{1}{2}\frac{\partial^2}{\partial z^2} - \frac{1}{\sqrt{\rho^2 + z^2}}. \tag{2.1}$$

For a superstrong magnetic field, where $a_H \ll a_B$, the adiabatic approximation can be used to solve the Schrödinger equation [2, 15]. In this case, the solution can be written as

$$\Psi_{nn_\rho m\sigma_z}(\mathbf{r}, \sigma) = R_{n_\rho m}(\boldsymbol{\rho})\chi_n(z)\varphi_{\sigma_z}(\sigma). \tag{2.2}$$

Here, $R_{n_\rho m}(\boldsymbol{\rho})$ are the standard functions of the transverse electron motion in a purely magnetic field [15], $\chi_n(z)$ is the wave function of the longitudinal electron motion, and $\varphi_{\sigma_z}(\sigma)$ is the spin part of the wave function. The bound-state spectrum for the Hamiltonian can be represented as

$$E_{nn_\rho m\sigma_z} = \left(n_\rho + \frac{1}{2}(|m| + m + \sigma_z + 1)\right)\mathcal{H} - \frac{1}{2}\lambda^2_{n|m|n_\rho}, \tag{2.3}$$

where $\sigma_z = \pm 1$ is the electron spin component along the magnetic field, *n* is the quantum number for the longitudinal motion, and the last term defines the shift of the corresponding Landau level produced by the attractive Coulomb potential; for the ground state, $n_r = m = 0$, $\sigma_z = -1$, and $E_0 = -\lambda^2/2$. This shift can be determined from the Schrödinger equation for the longitudinal part of the wave function:

$$\left\{-\frac{1}{2}\frac{\partial^2}{\partial z^2} + U_{\text{eff}}(|z|) + \frac{\lambda^2}{2}\right\}\chi_n(z) = 0, \tag{2.4}$$

in which the effective (averaged over the rapid transverse electron motion) potential energy is given by the expression [2, 15]

$$U_{\text{eff}}(|z|) = -\iint \frac{1}{\sqrt{\rho^2 + z^2}}|R_{n_\rho m}(\boldsymbol{\rho})|^2 d^2\rho. \tag{2.5}$$

Note the following properties of the effective potential (specific expressions for it depend on the quantum numbers n_ρ and $|m|$ for the transverse motion).

(1) At $|z| \lesssim a_H \ll a_B$, $U_{\text{eff}}(|z|) \sim 1/a_H$. In this case, $U_{\text{eff}} a_H^2 \sim 1/a_H \ll 1$, so the effective potential at such distances is a shallow one-dimensional potential well.

(2) At $|z| \gg a_H$, the effective potential is a Coulomb one.

(3) Note also that $U_{\text{eff}}(|z|) + 1/|z| > 0$; i.e., the effective potential energy curve lies above the Coulomb one.

The effective potential for the $m = n_\rho = 0$ states is

$$U_{\text{eff}}(|z|) = -\frac{\sqrt{2}}{a_H}\int_0^\infty \exp\left(-\frac{\sqrt{2}|z|x}{a_H} - x^2\right)dx = -\sqrt{\frac{\pi}{2}}a_H^{-1}\operatorname{erfc}\left(\frac{|z|}{\sqrt{2}a_H}\right)\exp\left(\frac{z^2}{2a_H^2}\right), \tag{2.6}$$

and its limiting expressions are

$$U_{\mathrm{eff}}(|z|) = \begin{cases} -\sqrt{\pi/2}\,a_H^{-1} + |z|/a_H^2, & z \ll a_H, \\ -1/|z| + a_H^2/|z|^3, & z \gg a_H. \end{cases} \tag{2.6'}$$

The solutions of Eq. (2.4) have a certain parity, $P = \pm 1$, relative to the reflection $z \longrightarrow -z$. We begin with even states for which $\chi'(0) = 0$.

2.1. The Spectrum of Even Levels

Since the effective potential in Eq. (2.4) at $|z| \lesssim a_H \ll 1$ is a shallow one-dimensional potential well, the wave functions of even states at such distances change only slightly, and, in the zeroth approximation, $\chi(z) = 1$ and $\chi'(z) = 0$ if the terms with U_{eff} and λ^2 are considered as a perturbation. A more accurate value of $\chi'(z)$ can be obtained from (2.4) if we discard the term with the binding energy, substitute $\chi(z)$ with unity in the term with the effective potential, and integrate the resulting equation over z. As a result, we obtain

$$\chi'(z) = 2\int_0^z U_{\mathrm{eff}}(|z|)dz. \tag{2.7}$$

For the subsequent analysis, we will need the values of $\chi'(z)$ at distances $a_H \ll z \ll a_B$, where the effective potential is a Coulomb one. Taking into account the dependence of the functions $R_{n_\rho m}(\boldsymbol{\rho})$ on ρ (the polynomials in variable ρ^2 multiplied by $\rho^{|m|}\exp(-\rho^2/4a_H^2)$) and using the integral

$$\int_0^\infty x^{s-1}e^{-x}\ln x\,dx = \Gamma'(s) = \Gamma(s)\psi(s), \tag{2.8}$$

where $\psi(z)$ is the logarithmic derivative of the gamma function, we see that the derivative $\chi'(z)$ at the above distances is[1]

$$\chi'(z) \approx -2\ln\frac{z}{a_H} + A_{|m|n_\rho}, \tag{2.9}$$

the term $A_{|m|n_\rho}$ does not depend on B. In particular, using the expression for the wave functions of transversally nodeless states (with $n_\rho = 0$)

$$R_{0m}(\boldsymbol{\rho}) = \{\pi(2a_H^2)^{(1+|m|)}(|m|!)\}^{-1/2} \times \rho^{|m|}\exp\left(im\varphi - \frac{\rho^2}{4a_H^2}\right), \tag{2.10}$$

we obtain

$$A_{|m|0} = -\ln 2 + \psi(1 + |m|). \tag{2.11}$$

Recall that $\psi(1) = -\gamma$, where $\gamma = 0.5772\ldots$ is the Euler constant, and $\psi(n+1) = -\gamma + \sum_{k=1}^{n} 1/k$ for $n = 1, 2, 3, \ldots$ [20].

On the other hand, the solution of Eq. (2.4) with a purely Coulomb potential, which exponentially decreases at large distances, $z \longrightarrow \infty$, is described by the Whittaker function

$$\chi(z) = \mathrm{const}\,W_{\nu,1/2}(2|z|\nu), \quad \nu = 1/\lambda \tag{2.12}$$

(see Appendix A). At small distances, $|z| \ll 1$, this solution is (A.3) with $x = 2\lambda|z|$. Joining formulas (2.9) and (2.12) and using (A.4), we obtain an equation for the spectrum of even states:

$$\ln\mathcal{H} = \lambda + 2\left[\ln\lambda + \psi\left(1 - \frac{1}{\lambda}\right) + 2\gamma + \ln 2\right] + A_{|m|n_\rho}. \tag{2.13}$$

We emphasize that this equation directly defines the λ dependence of the magnetic field for even levels; the expression $\exp\{-A_{|m|n_\rho}\}\mathcal{H}(\lambda_{n|m|n_\rho})$ is a universal function (for a given n) that is the same for states with different quantum numbers $|m|$ and n_ρ.

Let us explain the derivation of Eq. (2.13). If $U_{\mathrm{eff}}(z)$ is a shallow short-range potential well, then we may set $z = \infty$ in formula (2.7) at distances $r \gg a_H$, with

$$\chi'(z) \approx \mathrm{const} = 2\int_0^\infty U_{\mathrm{eff}}(z)dz.$$

In this case, solution (2.12) takes the form $\chi(z) = \exp(-\lambda|z|)$ and joining the solutions yields the standard equation

$$\lambda = -2\int_0^\infty U_{\mathrm{eff}}(z)dz,$$

which defines the location of the only discrete level that exists in a shallow one-dimensional short-range poten-

---

[1] We should first integrate $1/\sqrt{z^2+\rho^2}$ in the effective potential (2.5) over z and pass to the limit $|z| \longrightarrow \infty$ by discarding the decreasing expansion terms. The integral with $\ln z$ can then be easily calculated, and the z-independent integral with $\ln\rho$ can be expressed in terms of integrals of type (2.8). Note also that, although $\chi'(z) \gg 1$ in (2.9), $\chi(z) \approx 1$ and $|a_H\chi'(z)| \ll 1$ as previously.

tial well (see, e.g., [15, §45]). The same technique was used in [15, §112] for a distorted Coulomb potential. In this case, however, the substitutions of the Whittaker function with $\exp(-\lambda|z|)$ and z with a_B in (2.7) and (2.9) are unjustified. This is the reason why the error of formula (1.1) for the ground state is large and the resulting equation is inapplicable to all excited states.

The basic properties of the even-level spectrum can be easily understood if we notice an analogy between Eq. (2.13) and the equation that defines the s-state spectrum in a three-dimensional attractive Coulomb potential $U(r) = -e^2/r$ distorted at small distances, $r \ll a_B$, by a short-range potential $V_s(r)$, which can itself bind the particle (electron). This equation is (see Section 5)

$$\lambda + 2\ln\lambda + 2\psi\left(1 - \frac{1}{\lambda}\right) = \frac{1}{a_0}, \qquad (2.14)$$

where a_0 is the renormalized (Coulomb-modified) scattering length in the potential V_s. A comparison of Eqs. (2.13) and (2.14) indicates that an analog of the inverse scattering length is

$$\frac{1}{a_0(B)} \equiv \ln\mathcal{H} - 4\gamma - 2\ln 2 - A_{|m|n_\rho}. \qquad (2.15)$$

In this case, under the validity conditions for the approach in question, the inequality $a_H \ll a_0(B)$ must be satisfied. Thus, for $\mathcal{H} = 10^4$ at $m = n_\rho = 0$, we have $a_0(B) \approx 0.15$, while $a_H = 0.01$.

Bearing in mind this analogy, we note the following properties of the even-level spectrum:

(1) For each pair of quantum numbers $|m|$ and n_ρ, Eq. (2.13) has an infinite number of roots $\lambda^{(+)}_{n|m|n_\rho} > 0$. The lowest (with $n = 1$) root with $\lambda^{(+)}_{1|m|n_\rho} \gg 1$ corresponds to a deep (on the atomic scale) level.

(2) The remaining roots ($n \geq 2$) correspond to excited states. The corresponding energy levels are located between the adjacent unshifted Coulomb $n's$ levels with the principal quantum numbers n' equal to $(n-1)$ and n.

Note that an equation similar to (2.13), in which $\psi(1 - 1/\lambda)$ was substituted for $\psi(1) = -\gamma$, was derived in [16] for the lower (deep) $n = 1^+$ level. This equation is definitely inapplicable to excited even states, but it is asymptotically exact for the ground level. This substitution results in an appreciable loss of accuracy at large but finite $\mathcal{H}$. However, even the inclusion of the next expansion term $\psi(1 - 1/\lambda)$ in λ^{-1} gives the equation

$$\ln\mathcal{H} = \lambda + 2\ln\lambda + 2\gamma + 2\ln 2 + A_{|m|n_\rho} - \frac{\pi^2}{3\lambda}, \qquad (2.16)$$

which has a accuracy of several percent for $\mathcal{H} \geq 10^5$ (see Table 1).

Equations (2.13) and (2.16) explicitly specify the λ dependence of the magnetic field $\mathcal{H}$ and implicitly specify the inverse function $\lambda = \lambda(\mathcal{H})$. The asymptotics of the ground-level binding energy $\varepsilon_0(\mathcal{H})$ in the limit $\mathcal{H} \longrightarrow \infty$ can also be easily obtained. However, since this asymptotics is established very slowly, we give the corresponding formulas in Appendix B.

2.2. The Spectrum of Odd Levels

Let us now discuss the properties of odd states that follow from Eq. (2.4). The energy spectrum E_{ns} for s levels in a three-dimensional potential $U(r)$ is known [21] to coincide with the spectrum for even levels in a symmetric one-dimensional potential $U(|z|)$ of the same form. Therefore, taking into account the above properties of $U_{\text{eff}}(|z|)$, we may assert that, if the potential in the corresponding three-dimensional problem is written as

$$U_{\text{eff}}(r) \equiv -\frac{1}{r} + \left(U_{\text{eff}}(r) + \frac{1}{r}\right),$$

then the last term may be considered as a small distortion of the Coulomb potential and can be taken into account by using perturbation theory. As a result, we can write the following expression for the energies of odd levels:

$$E^{(-)}_{n|m|n_\rho} \equiv -\frac{1}{2}\left(\lambda^{(-)}_{n|m|n_\rho}\right)^2 = -\frac{1}{2n^2} + \int\left(U_{\text{eff}}(r) + \frac{1}{r}\right)\psi^2_{ns}(r)d^3r, \qquad (2.17)$$

where $\psi_{ns}(r)$ are the unperturbed wave functions of ns states in the Coulomb potential. In this case, $\overline{1/r} = 1/n^2$, and the levels are slightly shifted upward relative to the Coulomb levels, because the integrand in (2.17) is positive.

Substituting, at $m = n_\rho = 0$, expression (2.5) for the effective potential into (2.17), writing

$$\psi^2_{ns}(r) = \psi^2_{ns}(0)e^{-2\kappa r}\sum_{k=0}^{2(n-1)} c_n^{(k)} r^k,$$

$$\psi^2_{ns}(0) = \frac{1}{\pi n^3}, \quad \kappa = \frac{1}{n},$$

$$c_n^{(0)} = 1, \quad c_n^{(1)} = -\frac{2(n-1)}{n}, \ldots$$

and performing the integration yields

$$\left(\lambda^{(-)}_{n00}\right)^2 = -2\overline{U}_{\text{eff},\,n} - \frac{1}{n^2}, \qquad (2.18)$$

where

$$\overline{U}_{\mathrm{eff},\,n} = -\frac{4\sqrt{2}}{a_H n^3}\sum_{k=0}^{2(n-1)} c_n^{(k)}\Gamma(k+3)$$
$$\times\int_0^{\infty} e^{-x^2}\left(2\kappa+\frac{\sqrt{2x}}{a_H}\right)^{-k-3}dx.$$

Hence, integrating by parts, we can derive the following asymptotic expansion for the binding energy when $\mathcal{H} \longrightarrow \infty$:

$$(\lambda_{n00}^{(-)})^2 = \frac{1}{n^2}$$
$$-\frac{4}{n^3}\left\{\frac{\ln\mathcal{H}}{\mathcal{H}}+\frac{a_n^{(0)}}{\mathcal{H}}+\frac{a_n^{(1)}}{\mathcal{H}^{3/2}}+\dots\right\}, \qquad (2.19)$$

where the expansion coefficients for the lower odd levels are[2]

$$a_1^{(0)} = -(2+\ln 2+\gamma) \approx -3.27036,$$
$$a_2^{(0)} = -(3.5-\ln 2+\gamma) \approx -3.38407,$$
$$a_3^{(0)} = -(13/3-\ln 4.5+\gamma) \approx -3.40647,$$
$$a_1^{(1)} = 4\sqrt{2\pi} \approx 10.0265, \qquad (2.20)$$
$$a_2^{(1)} = 4\sqrt{2\pi} \approx 10.0265,$$
$$a_3^{(1)} = \frac{16\sqrt{2\pi}}{3} \approx 13.3678.$$

3. THE ACCURACY OF THE ASYMPTOTIC FORMULAS

Let us now discuss our results and compare them with available numerical calculations. The hydrogen atomic spectrum was calculated with a precision accuracy for several $m = 0$ states below the ground Landau level over a wide B range in [9–11, 14]. Table 1 illustrates the dependence $\mathcal{H}(\lambda)$ for the ground atomic state. For the given λ^2, this table compares the corresponding magnetic field strengths taken from [9, 14] and calculated using both formulas (2.13) and (2.16) and the asymptotic formula by Hasegawa and Howard [16]. In addition, it gives the a_H/a_B ratio, whose smallness is required for the adiabatic approximation to be applicable. As expected, the binding energy for the ground state in a strong magnetic field is large on the scale of ordinary atomic energies.

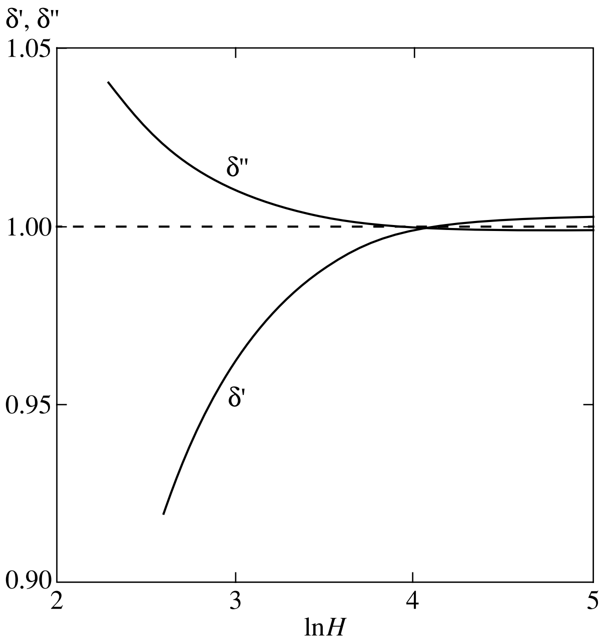


Fig. 2. The accuracy of Eq. (2.13) for the ground level. The values of δ' and δ'' give the errors of the corresponding approximations; see (3.1).

Note that the dependence $\lambda^2(\mathcal{H})$ is given by Eq. (2.13) with a much higher accuracy than the inverse dependence $\mathcal{H}(\lambda)$. Figure 2 shows the errors

$$\delta'(\lambda) = \frac{\tilde{\mathcal{H}}-\mathcal{H}}{\mathcal{H}}, \quad \delta''(\mathcal{H}) = \frac{\tilde{\lambda}^2-\lambda^2}{\lambda^2}, \qquad (3.1)$$

where λ^2 and $\mathcal{H}$ are the exact values, and the tilde denotes the values calculated using Eq. (2.13). This difference stems from the fact that $\mathcal{H}(\lambda)$ has a sharp exponential pattern, which is one of the reasons why formula (1.2) is untenable.

Above, in connection with formula (2.13) for the spectrum of even levels, we pointed out a peculiar scaling relation for $\mathcal{H}(\lambda_{n|m|n_\rho})$ at various magnetic quantum numbers m:

$$\xi_m \equiv \ln\mathcal{H}(\lambda_{n|m|n_\rho})-4\gamma-2\ln 2-A_{|m|n_\rho}$$
$$= F(\lambda_{n|m|n_\rho}), \qquad (3.2)$$

in which

$$F(\lambda) = \lambda+2[\ln\lambda+\psi(1-1/\lambda)] \qquad (3.3)$$

is a universal function. As we see from Fig. 3, scaling (3.2) holds good for states with various m at $\lambda^2 > 12$. A deviation from relation (3.2) is observed at $\lambda^2 < 10$, but in this range, the points ξ_m cluster along a smooth curve close to (3.3).

---

[2] Here, we corrected the misprint made in formula (17) from [18] in the numerical coefficient $a_1^{(1)}$.

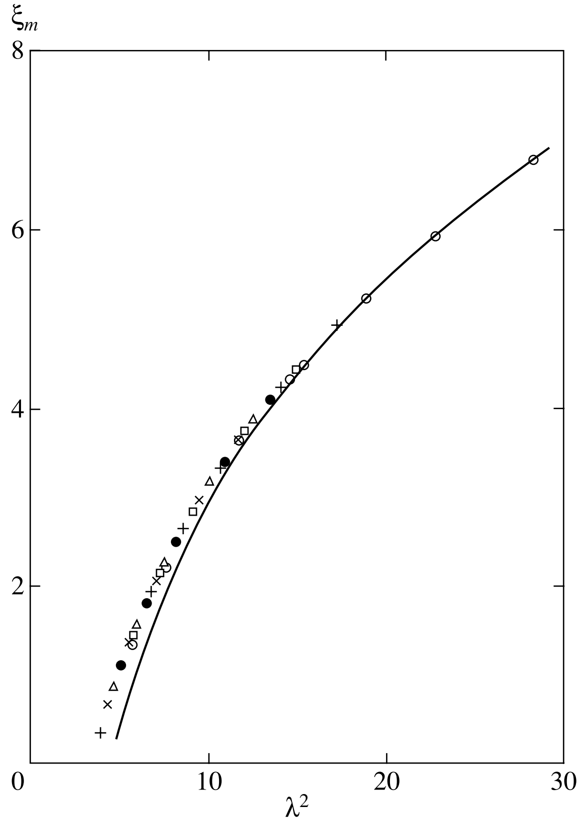


Fig. 3. Checking the scaling relation (3.2): the solid line represents the function $F(\lambda)$; ○, +, □, ●, △, and × represent the results of numerical calculations [9, 10] for $m = 0, -1, -2, -3, -4$, and -5, respectively.

Table 2 presents the binding energies for the lower excited atomic states with $m = 0$ from [9] and calculated using the asymptotic formulas (2.13), (2.18), and (2.19) for the states with quantum numbers $n = 2^+$, $m = n_\rho = 0$ and $n = 1^-$, $m = n_\rho = 0$ (in [9], they are classified as the 2*s* and 2*p* states with $m = 0$, respectively). Below, we make the following remarks regarding the results presented in Table 2.

(1) A comparison with the data in Table 1 shows that the approach under consideration provides a higher accuracy for these (excited in the longitudinal direction of motion) states. This circumstance has a simple explanation, because the longitudinal size of the localization region for the wave function increases for excited states, which results in an extension of the validity range for the adiabatic approximation to progressively lower magnetic field strengths.

(2) For the odd $n = 1^-$ (2*p*) state, the level shift in a magnetic field $\mathcal{H} \gtrsim 60$ is small and completely confirms the remarks made in Section 2.2.

(3) However, the even $n = 2^+$ (2*s*) state is much more interesting. As we see from Table 2, the corresponding level is greatly shifted relative to the unperturbed Coulomb levels, $E_n = -1/2n^2$, and is almost halfway between the levels with the principal quantum numbers $n = 1$ and $n = 2$ over a wide range of magnetic fields, $\mathcal{H} = 2 \times 10^2$–$2 \times 10^4$ ($\tilde{\lambda}_{12}^2 = 0.625$ corresponds to this level location). Note that a similar situation takes place for more strongly excited even levels. Thus, for the $n = 3^+$ level at $\mathcal{H}$ equal to 10^2 and 10^3, formula (12) gives $\lambda^2 = 0.17471$ and 0.18876, respectively, while $\tilde{\lambda}_{23}^2 = 0.18056$. These properties of the even-state spectrum are related to the manifestation of the Zeldovich effect in them (see Section 5).

4. THE ENERGY SPECTRUM OF A RELATIVISTIC ELECTRON

(In this section, we use standard units.) The results obtained above for the spectrum of Hamiltonian (2.4) are asymptotically exact for $B \longrightarrow \infty$. In this case, however, the transverse electron velocity indefinitely increases, the Schrödinger equation becomes inapplicable, and the Dirac equation should be used. Thus, at $B = 10^{13}$ G (or $\mathcal{H} \approx 4000$), the kinetic energy of the transverse electron motion is $0.1 m_e c^2$ in order of magnitude.

Let us discuss the generalization of our results to this case and first note the following property of bound states in the problem under consideration. There is a wide range of strong magnetic fields in which the transverse electron motion is relativistic (and can even be ultrarelativistic), while the longitudinal electron motion is nonrelativistic and can be studied in terms of the approach outlined in Section 2. This is particularly clearly seen for the bound states that originate from the ground Landau level of the transverse electron motion under the action of a Coulomb potential. From the Dirac equation for the bispinor $\Psi_E = \begin{pmatrix}\varphi(\mathbf{r})\\ \chi(\mathbf{r})\end{pmatrix}$, we have the equations

$$c\boldsymbol{\sigma}_\perp\cdot\left(\mathbf{p}_\perp + \frac{e}{2c}[\mathbf{B}\times\boldsymbol{\rho}]\right)\chi + c\sigma_z\hat{p}_z\chi = \left(E - m_e c^2 + \frac{e^2}{r}\right)\varphi, \tag{4.1'}$$

$$c\boldsymbol{\sigma}_\perp\cdot\left(\hat{\mathbf{p}}_\perp + \frac{e}{2c}[\mathbf{B}\times\boldsymbol{\rho}]\right)\varphi + c\sigma_z\hat{p}_z\varphi = \left(E + m_e c^2 + \frac{e^2}{r}\right)\chi. \tag{4.1''}$$

Considering initially only the transverse electron motion (neglecting the Coulomb potential), we note that the (degenerate in energy) electronic states with the

Table 2. Binding energies $\lambda^2(\mathcal{H})$ for the $2s$ and $2p$ states

$\mathcal{H}$	$n = 2^+$ ($2s$)		$n = 1^-$ ($2p$)		
	[9]	Eq. (2.13)	[9]	Eq. (2.18)	Eq. (2.19)
10	0.41790	0.44291	0.76530	0.67081	–
20	0.44768	0.46440	0.82676	0.77299	0.6065
40	0.47640	0.48696	0.87748	0.84901	0.7996
60	0.49261	0.50046	0.90186	0.88287	0.8588
100	0.51236	0.51760	0.92723	0.91619	0.9065
140	0.52496	0.52888	0.94092	0.93330	0.9280
200	0.53794	0.54079	0.95306	0.94804	0.9453
400	0.56206	0.56342	0.97073	0.96856	0.9678
1000	0.59171	0.59207	0.98499	0.98433	0.98418
2000	0.61248	0.61250	0.99119	0.99093	0.99089
3000	–	0.62390	–	0.99348	0.99344
5000	–	0.63766	–	0.99570	0.99569
1.0(4)	–	0.65525	–	0.99759	0.99758
2.0(4)	–	0.67163	–	0.99866	0.99866
5.0(4)	–	0.69150	–	0.99939	0.99939
1.0(5)	–	0.70528	–	0.99967	0.99967

quantum numbers $n_\rho = 0$, $m = -|m| = 0, -1, -2, \ldots$, and $\sigma_z = -1$ correspond to the lower Landau level. For them,

$$E^{(0)}_{n_\rho m \sigma_z} = m_e c^2, \quad \varphi^{(0)}(\boldsymbol{\rho}) = R_{n_\rho m}(\boldsymbol{\rho})\begin{pmatrix}0\\1\end{pmatrix}, \qquad \chi^{(0)} = \begin{pmatrix}0\\0\end{pmatrix}, \tag{4.2}$$

with the radial functions $R_{n_\rho m}(\boldsymbol{\rho})$ having the same form as those in the nonrelativistic case. Using the adiabatic approximation and assuming[3] that $e^2/a_H \ll m_e c^2$ (or $B \ll 10^{18}$ G), we obtain from Eqs. (4.1) and (4.2)

$$\varphi = \varphi^{(0)}(\boldsymbol{\rho}) \cdot \psi_l(z), \qquad \chi \approx \frac{i\hbar}{2m_e c}\frac{\partial}{\partial z}\varphi = \frac{i\hbar}{2m_e c} R_{n_\rho m}(\boldsymbol{\rho})\psi_l'(z)\begin{pmatrix}0\\1\end{pmatrix} \tag{4.3}$$

(in this case, $|\chi| \ll |\varphi|$) and an equation for the longitudinal part of the wave function $\psi_l(z)$ that matches the nonrelativistic equation (2.4). Accordingly, the energy spectrum now takes the form

$$E_{n_\rho = 0,\, m,\, \sigma_z = -1} = m_e c^2 - \frac{\hbar^2\lambda^2}{2m_e}, \tag{4.4}$$

where the binding energy is given by Eq. (2.13).

The relativistic corrections to formula (4.4) were considered in many papers, in particular, in [7, 8], based on an accurate numerical solution of the Dirac equation and in [5] using the adiabatic approximation. The relativistic corrections were shown to be very small. Thus, according to [5, 7], they are about 3×10^{-5} of the nonrelativistic $\lambda^2/2$ value for $\mathcal{H} = 5 \times 10^3$.

However, for the excited states associated with higher Landau levels, the situation is slightly different. At $B \sim B_{\text{cr}}$, the energy of the transverse motion is no longer equal to $m_e c^2$, which, in turn, affects the longitudinal motion. We illustrate this using the Klein–Gordon equation for a scalar particle as an example:

$$c^2\left(\hat{\mathbf{p}} + \frac{e}{2c}[\mathbf{B} \times \mathbf{r}]\right)^2 \Psi_E + m_e^2 c^4 \Psi_E = \left(E + \frac{e^2}{r}\right)^2 \Psi_E. \tag{4.5}$$

In the adiabatic approximation, we now have

$$\Psi_E \equiv \Psi_{nmn_\rho}(\mathbf{r}) \approx R_{n_\rho m}(\boldsymbol{\rho})\psi_l(z)$$

and the energy of the transverse motion

$$E^{(0)}_{n_\rho m} = \sqrt{m_e^2 c^4 + (2n_\rho + m + |m| + 1)\hbar\omega_L m_e c^2}.$$

The equation for the longitudinal part of the wave func-

---

[3] The satisfaction of this inequality ensures that the nonrelativistic approximation can be used to describe the longitudinal electron motion.

tion takes the form

$$-\frac{\hbar^2}{2m_e}\psi_l''(z) - \epsilon\left\langle \frac{e^2}{\sqrt{\rho^2+z^2}}\right\rangle_{n_\rho m}\psi_l(z) = \epsilon\Delta E\psi_l(z), \quad (4.6)$$

where $\epsilon = E^{(0)}_{n_\rho m}/m_e c^2$ and $\Delta E = E - E^{(0)}_{n_\rho m}$, and the values averaged over the transverse motion, $-\langle e^2/\sqrt{\rho^2+z^2}\rangle_{n_\rho m}$, are equal to the corresponding effective potential in (2.5). Thus, we see that the relativistic transverse particle motion results in charge renormalization: $e^2 \longrightarrow \epsilon e^2$, which significantly affects the spectra of both even and odd states. For example, when the distortion of the seed Coulomb potential $U_C = -e^2/|z|$ at small distances is neglected, the odd levels are

$$\Delta E_n^{(-)} = -\epsilon\frac{m_e e^4}{2\hbar^2 n^2}. \quad (4.7)$$

The generalization of the above approach to the Dirac equation should be considered separately.

5. THE ZELDOVICH EFFECT IN ATOMIC SPECTRA

The peculiar properties of the discrete spectrum in a Coulomb field distorted at small ($0 < r < r_0$) distances were first considered by Zeldovich [1] in connection with the energy levels of an electron in an extrinsic semiconductor with a dielectric constant $\varepsilon \gg 1$. This author showed that at the time ($g = g_0$) when a bound s level emerges in a short-range potential $V_s(r) = -g\mathcal{V}(r)$ or when there is a resonance in the scattering of low-energy particles (i.e., a real or virtual level with an energy close to zero), the atomic spectrum is rearranged: the Coulomb level E_{ns} rapidly sinks to $E_{n-1,\,s}$, $n = 2, 3, \ldots$, while the ground level E_{1s} plunges steeply downward. In this case, the relative width of the rearrangement region (in coupling constant g of the strong potential) is $\Delta g/g_0 \sim r_0/a_B \ll 1$.

A similar behavior of the s and p levels of the electron spectrum was found in the relativistic Coulomb problem with a nuclear charge $Z > 137$ [22], where the $1s_{1/2}$ level at the critical value of $Z = Z_{\rm cr}$ disappears from the discrete spectrum,[4] going into the lower continuum $E < -m_e c^2$ (the Dirac sea). It was pointed out in [29] that this effect could show up in the nuclear level shifts of the lightest hadronic atoms ($\bar{p}p$, K^-p, etc.). In particular, it was noted that the level shift could be positive ($\Delta E_{ns} > 0$, i.e., the level is pushed upward), although the short-range potential V_s that produces these shifts is attractive.

Specific model potentials V_s were used in these calculations: a square potential well $\mathcal{V}(r) = \theta(r_0 - r)$ [1, 29], a parabolic potential corresponding to a constant volume charge density inside the nucleus [22], and separable finite-rank potentials [29–33]. The general pattern of this phenomenon, which can take place in all systems for which the interaction potential breaks up into two parts (short-range and long-range ones) with highly incommensurable radii and which weakly depends on the specific form of $V_s(r)$, was pointed out in [1] and, in more detail, in [32–34].

In [33], the Zeldovich effect was considered based on the equation

$$\lambda + 2[\ln\lambda r_0 + \psi(1-\lambda^{-1}) + c_0] = a_B/a_s, \quad (5.1)$$

which defines the ns-level locations when the long-range part of the potential ($r > r_0$) has a Coulomb form.[5] Here, $c_0 = \ln 2 + 2\gamma = 1.848$, r_0 is the joining point, a_s is the s-scattering length for the inner potential (at $r < r_0$), and the small terms on the order of r_0 and $r_0\ln(\lambda r_0)$ were discarded in (5.1). The validity condition for Eq. (5.1) is $r_0 \ll a_B$. In the problem considered by Zeldovich [1], this condition is ensured by the fact that $a_B = \varepsilon m_e/m_{\rm eff} \gg r_0 \sim 1$ (here, m_e is the electron mass, $m_{\rm eff}$ is the effective electron mass in the lattice, and r_0 is the ion radius). For hadronic atoms, we have $a_B = 57.6$ Fm for a $\bar{p}p$ atom, $a_B = 83.5$ Fm for K^-p, $a_B = 51.4$ Fm for Σ^-p, etc., while the strong-interaction radius is $r_0 \approx$ 2–3 Fm. Therefore, there is also a small parameter $r_0/a_B \approx 1/30$–$1/40$ here.

Equation (5.1) shows that the ns-level energies are rigidly related and can be expressed in terms of the energy of one of these levels without requiring the solution of the Schrödinger equation (see [35, Fig. 1]). The properties of Eq. (5.1) were analyzed in detailed in [41] in connection with the then available evidence [42] for a large shift of the $1s$ level in a proton–antiproton atom. According to (5.1), a shallow deuteron-type bound state could exist in a $\bar{p}p$ system, which would be of great interest in nuclear physics. Subsequently, however, it emerged that the experimental results [42] are incorrect, the level shifts for a $\bar{p}p$ atom are small, and interest in the Zeldovich effect (at least in the field of nuclear physics) fell sharply. An overview of the main results obtained in [35, 38–41] and other papers of this period can be found in [43].

---

[4] In this case, we should include finite sizes of the nucleus in the analysis, i.e., cutoff the Coulomb potential at small distances [23], without which the problem is ill-posed because of the "fall to the center" [15, 24–26]. Taking this into account, the critical charge is $Z_{\rm cr} \approx 170$ for a spherical superheavy nucleus [27, 28].

[5] See also [36, 37]. Note that the name "Zeldovich effect" was offered in [33]. The Zeldovich effect for the states with an angular momentum $l \neq 0$ is peculiar [38–40] compared to the s states. The peculiarities of this effect in the presence of absorption in the system (which is the case, for example, in hadronic atoms) were considered in [40, 41].

Turning to a hydrogen atom in a superstrong magnetic field, $B \gg B_a$, we note that such a field compresses the atom whose characteristic transverse size is equal to the Landau radius. In this case, a_H acts as the cutoff radius of the Coulomb potential, which is small compared to the Bohr radius for $\mathcal{H} \gg 1$: $a_H/a_B = \sqrt{B_a/B} = 1/\sqrt{\mathcal{H}}$. Therefore, it is not surprising that Eq. (2.13), similar to (5.1), in which the specific relationship between r_0 and a_B depends on the quantum numbers n_ρ and m for the electron motion in a magnetic field, is obtained for the energy spectrum of a hydrogen atom in this limit. In particular, for the nodeless states with $n_\rho = m = 0$,

$$\frac{r_0}{a_B} = \exp\left(-\frac{1}{2}(\ln 2 + \gamma)\right)\frac{a_H}{a_B} = \frac{0.5298}{\sqrt{\mathcal{H}}}, \qquad (5.2)$$

and the condition $r_0 \ll a_B$ is satisfied if $\mathcal{H} \gg 1$.

Equation (2.13) defines the level energy in a given field $\mathcal{H}$ or the magnetic field strength for the measured level shift. It is identical to the corresponding equation in the theory of a $\bar{p}p$ atom, in which the parameter ξ is

$$\xi \equiv \ln\mathcal{H} - (3\gamma + \ln 2) = a_B/a_{cs}, \qquad (5.3)$$

where a_{cs} is the Coulomb-modified low-energy $\bar{p}p$-scattering length.[6] This shows that the Zeldovich effect must be observed in the spectrum of hydrogen atomic levels for $\mathcal{H} \gg 1$. The physical causes of this is clear: the 1*s* ground level of the atomic spectrum that sank steeply downward (its binding energy for $\mathcal{H} \gtrsim 500$ is several tens of times larger than the ionization potential of the hydrogen atom) acts as the quasi-nuclear state *Qs* that perturbs the Coulomb spectrum (see [35, Fig. 1]).

Let us discuss the cause of the different patterns of even- and odd-level shifts in a one-dimensional potential $U_{\text{eff}}(|z|)$ and the relationship to the Zeldovich effect. As was shown in [1], in the three-dimensional short-range Coulomb problem, large *ns*-level shifts are possible only for a distorting short-range potential $V_s(r)$ in which an intrinsic shallow *s* level, real or virtual, is available. In the remaining cases, the Coulomb level shifts are small even if intrinsic levels of the discrete spectrum, deep on the atomic scale, are available in the potential V_s. In the one-dimensional case, the wave functions of even and odd states satisfy the same Schrödinger equation (2.4) on the semiaxis $r \equiv |z| > 0$ but different boundary conditions at zero: $\chi^{(-)}(0) = 0$ for odd levels and $\chi^{(+)\prime}(0) = 0$ for even levels with $\chi^{(+)}(0) \neq 0$. As we noted above, the spectrum of odd levels in a symmetric one-dimensional potential coincides with the spectrum of *ns* levels in the corresponding centrally symmetric three-dimensional potential. The wave functions in these cases are related by $\psi_{ns}(r) = \chi_n^{(-)}(r)/r$. Therefore, $\psi_{ns}(0) < \infty$ and the odd-level shifts relative to the Coulomb levels are small, because the Coulomb potential distortion is weak.[7]

There is no such an analogy for even states, because now $\psi_{ns}(r) = \chi_n^{(+)}(r)/r \propto 1/r \longrightarrow \infty$ when $r \longrightarrow 0$ and the singular solutions of the Schrödinger equation are usually excluded from analysis, except for the case of a zero-range three-dimensional potential (δ potential), which is specified by the boundary condition (see [44, p. 27])

$$\left.\frac{d\ln\chi(r)}{dr}\right|_{r=0} = -\kappa_0, \qquad (5.4)$$

or

$$\psi(r) \approx \text{const}\left(\frac{1}{r} - \kappa_0 + O(r)\right), \quad r \longrightarrow 0. \qquad (5.4')$$

Here, the parameter κ_0 defines the energy $E_0 = -\hbar^2\kappa_0^2/2m$ of the real (for $\kappa_0 > 0$) or virtual ($\kappa_0 < 0$) *s* level. Note that the even solutions $\chi_n^{(+)}(z)$ of the Schrödinger equation for a symmetric one-dimensional potential $U(|z|)$ are uniquely related to the solutions $\psi_{ns}(r) = \chi^{(+)}(r)/r$ of the three-dimensional Schrödinger equation for the *s* states with the same potential $U(r)$, with

$$\int\psi_{ns}^2(r)d^3r = 2\pi\int_{-\infty}^{\infty}[\chi_n^{(+)}(|z|)]^2dz \qquad (5.5)$$

and boundary condition (5.4) with $\chi^{(+)\prime}(0) = -\kappa_0 = 0$ being satisfied. Thus, the spherically symmetric potential is the superposition of a potential $U(r) \equiv U_{\text{eff}}(r)$ and a zero-range potential localized at $r = 0$, for which $\kappa_0 = 0$.

It remains to note that $\kappa_0 = 0$ implies that a zero-range potential models the short-range potential $V_s(r)$ when a bound state emerges in it; in this case, the (nonrenormalized by the effective potential) scattering length is $a_0 = \kappa_0^{-1} = \infty$. However, precisely this property of the short-range potential is required [1] for the clear-

---

[6] Here, we use the approximation of a zero range of action of the forces, $r_s = r_{cs} = 0$. The influence of the effective range on the location of the quasi-nuclear level in a $\bar{p}p$ system was analyzed in detail in [41]. For the scattering lengths a_s and a_{cs}, see remark (5) in Section 8.

[7] As was noted above, $U_{\text{eff}}(|z|) + 1/|z| > 0$, so the distorting potential is repulsive and, hence, cannot be resonant.

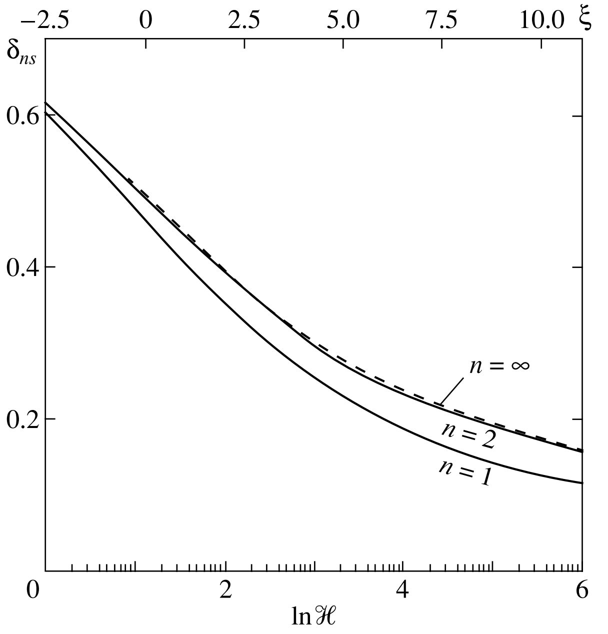


Fig. 4. Quantum defects for *ns* levels versus $\xi = a_B/a_{cs}$ or reduced magnetic field. The solid and dashed curves refer to the lower states with $n = 1$ and 2 and to states with $n \gg 1$, respectively.

est manifestation of the Zeldovich effect, which corresponds to $\xi = 0$ in Table 3.

6. THE ZELDOVICH EFFECT IN A SUPERSTRONG MAGNETIC FIELD: ANALYSIS OF THE NUMERICAL CALCULATIONS

The question arises as to what can be said about the Zeldovich effect based on available experimental data. Magnetic fields $B \gg B_a$ are encountered in astrophysics (magnetic white dwarfs, neutron stars), but here an additional analysis is required. However, there are numerous works in which the Schrödinger equation for a hydrogen atom in a uniform magnetic field was solved on a computer with a high accuracy (up to 10–12 significant figures) by various numerical methods for $\mathcal{H} \lesssim 10^5$ or $B \lesssim 2 \times 10^{14}$ G. The results of these calculations are in complete agreement [6–14]. We use these data by considering them as the result of a numerical experiment. First, however, some words should be said about the distinctive features of the Zeldovich effect.

Previously, it was noted [35] that the most suitable variable for describing this effect is not the energy E or $\lambda = \sqrt{-2E}$ but the dimensionless parameter

$$\nu = \frac{Z}{\lambda} = \sqrt{\frac{Z^2 m_e e^4}{2\hbar^2 |E|}} \qquad (6.1)$$

(if the potential $V(r) = -Z/r$ at $r > r_0$ for a hydrogen atom $Z = 1$). In atomic physics, this variable is called the effective quantum number and is usually denoted by n^* [45]. A characteristic property of the Zeldovich effect is that the ν_n values for the entire series of shifted atomic *ns* levels are (with a high accuracy) periodic in *n*. This periodicity is seen from Table 3, which gives the quantum defects $\delta_n = \nu_n - (n - 1)$ for *ns* states. The change of δ_n from $n = 2$ to $n = \infty$ (at a given $\mathcal{H}$) does not exceed 1%, which characterizes the degree of periodicity of ν_n in the region of the atomic spectrum (i.e., at $\nu > 1$), and is clearly seen from Fig. 4. On the other hand, for the lower-lying 1*s* level, which sinks arbitrarily deeply with increasing $\mathcal{H}$ [21, 46], this change is more significant (see, e.g., the first column in Table 3, which refers to an exact resonance: $a_{cs} = \infty$, $\xi = 0$ (the level formation time with an allowance made for the Coulomb interaction)). When $n \longrightarrow \infty$, $\nu_n = n - 1/2$ and $\delta_n = 1/2$ in this

Table 3. The δ_n and ρ_n values for a hydrogen atom in a magnetic field

n	$\mathcal{H} = 11.3$ ($\xi = 0$)			$\mathcal{H} = 1000$		$\mathcal{H} = 3120$		$\mathcal{H} = 10^5$	
	δ_n	δ_n/δ_∞	ρ_n	δ_n	ρ_n	δ_n	ρ_n	δ_n	ρ_n
1	0.4695	0.9391	–	0.2541	–	0.2179	–	0.1447	–
2	0.4964	0.9927	0.738	0.2996	0.544	0.2649	0.500	0.1907	0.393
3	0.4987	0.9973	0.641	0.3017	0.441	0.2668	0.399	0.1920	0.302
4	0.4993	0.9986	0.606	0.3022	0.399	0.2672	0.359	0.1923	0.267
5	0.4996	0.9992	0.583	0.3024	0.377	0.2674	0.337	0.1924	0.249
10	0.4999	0.9998	0.539	0.3026	0.337	0.2676	0.300	0.1925	0.218
$n \longrightarrow \infty$	0.5000	1.0000	0.500	0.3027	0.303	0.2677	0.271	0.1926	0.193
δ_1/δ_∞	0.939			0.839		0.814		0.751	

Note: The quantum defects δ_n for the *ns* levels of a hydrogen atom (at $n_\rho = m = 0$) and the relative level shifts (6.4) are given.

case. Table 3 also gives the values of δ_n and ρ_n for the case where $\rho_2 = 1/2$; i.e., the $2s$ level has the binding energy

$$\varepsilon_{2s} = \frac{1}{2}(\varepsilon_{1s}^{(0)} + \varepsilon_{2s}^{(0)}) = \frac{5}{12} \text{ at. units } = 8.5 \text{ eV}.$$

The ratio δ_1/δ_∞ changes more significantly than δ_2/δ_∞, more specifically, in the range from 0.95 to 0.75 when $\mathcal{H}$ changes from 100 to 10^5.

The periodicity property of ν_n can be easily explained if, using the identity [20] $\psi(1 - z) = \pi\cot(\pi z) + \psi(z)$, we write (2.13) as ($n_\rho = m = 0$)

$$\cot(\pi\nu) - d(\nu) = \xi = \frac{1}{2\pi}(\ln\mathcal{H} - 3\gamma - \ln 2). \quad (6.2)$$

The function $d(\nu)$ is numerically small in the region of the atomic spectrum:

$$d(\nu) = \frac{1}{\pi}\left[\ln\nu - \psi(\nu) - \frac{1}{2\nu}\right]$$

$$= \begin{cases} 0.0246, & \nu = 1, \\ 0.0065, & \nu = 2, \\ 1/12\pi\nu^2, & \nu \longrightarrow \infty. \end{cases}$$

Hence, for $n \gg 1$,

$$\delta_n = \delta_\infty - \frac{\sin^2\pi\delta_\infty}{12\pi n^2},$$
$$\delta_\infty = \frac{1}{\pi}\operatorname{arccot}\xi, \quad 0 < \delta_\infty < 1, \quad (6.3)$$

which is indicative of the ν_n periodicity at large n; as we see from Table 3, the latter extends to $n = 2$.

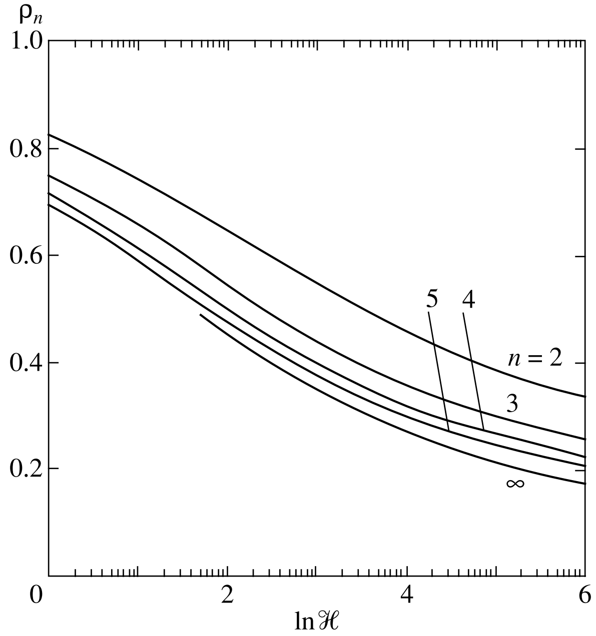


Fig. 5. Relative *ns*-level shifts (6.4) versus magnetic field.

Occasionally, the opinion that the rearrangement of the atomic spectrum shows up most clearly in level shifts is expressed. This is qualitatively true, but quantitatively the atomic level shifts do not obey such a simple law as do ν_n and δ_n. Indeed, let us consider the relative *ns*-level shift:

$$\rho_n = \frac{E_{ns} - E_{n-1}^{(0)}}{E_n^{(0)} - E_{n-1}^{(0)}}$$
$$= \frac{n^2(2n - 2 + \delta_n)}{(2n-1)(n-1+\delta_n)^2}\delta_n, \quad n \geq 2 \quad (6.4)$$

($E_n^{(0)} = -Z^2/2n^2$ is the Coulomb spectrum). In contrast to δ_n, the dependence of ρ_n on the level number is significant, as we see from Fig. 5. Similarly, if $\rho_n = 1/2$, i.e., if the *ns* level is exactly halfway between the adjacent Coulomb levels, then the quantum defect δ_n for it is by no means equal to 0.5:

$$\delta_n = (n-1)\left[n\sqrt{\frac{2}{2n^2 - 2n + 1}} - 1\right]$$
$$= \begin{cases} 0.265, & n = 2, \\ 0.353, & n = 3, \end{cases}$$

and $\delta_n = 1/2 - 3/8n + \ldots$ for $n \gg 1$. The values of ρ_n and δ_n are close only for highly excited ($n \gg 1$) states,

$$\rho_n = \delta_n + \frac{3\delta_n(1-\delta_n)}{2n} + O\left(\frac{1}{n^2}\right),$$
$$\delta_n < \rho_n < 1, \quad (6.5)$$

and for a superstrong magnetic field, where the quantum defects themselves are small,

$$\rho_n = \frac{2n^2}{(2n-1)(n-1)}\delta_n + O(\delta_n^2). \quad (6.6)$$

Thus, the constancy of the quantum defects δ_n for $2 \leq n < \infty$ is a reliable indicator of the Zeldovich effect in the atomic level spectrum (in a short-range Coulomb field).

Note also that our definition of the quantum defect slightly differs from the definition adopted in atomic physics [15, 46], where it is commonly assumed that $\nu_n = n - \Delta_{nl}$ (the Rydberg correction Δ_{nl} is virtually independent of the principal quantum number). In our case, it is convenient to measure δ_{nl} not from the initial

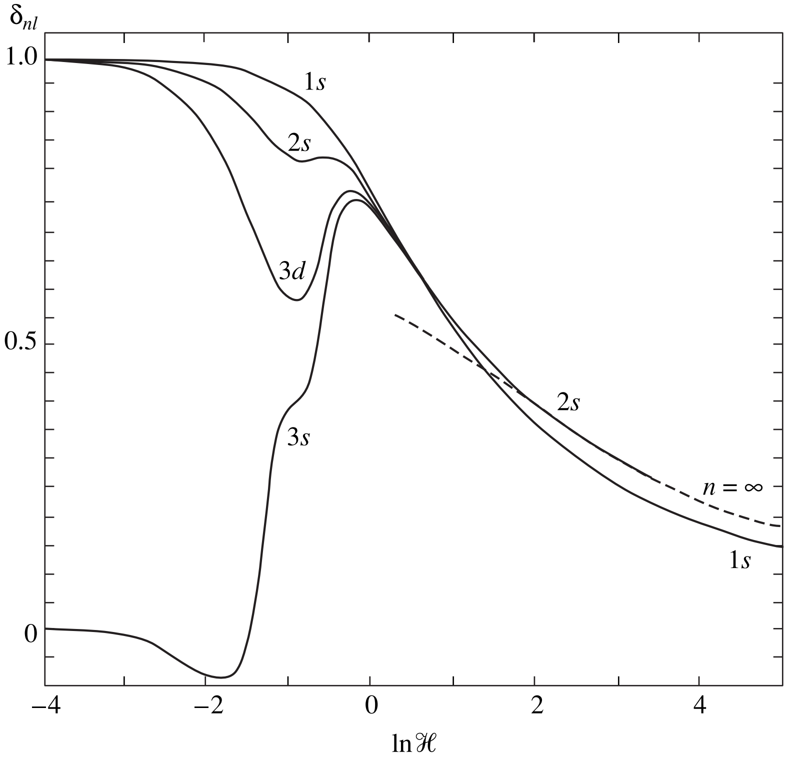


Fig. 6. Numerically calculated quantum defects $\delta_{nl}(\mathcal{H})$ for even levels [9]. The hydrogen atomic states according to their classification [9] with a switched-off magnetic field are given near the curves. The dashed curve refers to *ns* states with $n \gg 1$.

level n in a free ($\mathcal{H} = 0$) atom but from the limiting value of n_f to which the parameter ν tends when $\mathcal{H} \longrightarrow \infty$ in the Schrödinger equation (since $a_H \longrightarrow 0$, this corresponds to the motion in a one-dimensional Coulomb potential).

The quantum defects δ_{nl} can be easily calculated by using the binding energies $\varepsilon_{nl} = \lambda_{nl}^2$ (in rydbergs) from [9–11]. In classifying the states, we will number them by the quantum numbers n and l for a free hydrogen atom following [9, 11]. At $\mathcal{H} \gg 1$, n_ρ, m, n_f, and P, where n_ρ and m describe the rapid motion about the magnetic field [15], n_f is the level number in a one-dimensional Coulomb potential (the slow electron motion along $\mathcal{H}$),[8] and $P = \pm 1$ is the parity of the wave function relative to the reflection $z \longrightarrow -z$, become "good" quantum numbers. Below, we give the correspondence between the quantum numbers (n, l) and n_f^P for $n_\rho = m = 0$, because the energies of only these states were calculated in [9]. As the magnetic field increases, we have

$$1s \longrightarrow 1^+, \quad 2s \longrightarrow 2^+, \quad 3d \longrightarrow 3^+, \quad 3s \longrightarrow 4^+ \ldots \tag{6.7}$$

for the even $P = +1$ states and

$$2p \longrightarrow 1^-, \quad 3p \longrightarrow 2^-, \quad 4f \longrightarrow 3^- \ldots \tag{6.7'}$$

for the odd $P = -1$ states.

The δ_{nl} values for them are shown in Figs. 6 and 7. The Zeldovich effect is established at $\mathcal{H} > \mathcal{H}_{\min} \sim 100$ for even states and starting from $\mathcal{H}_{\min} \approx 1$ for odd states. The dashed line in Fig. 6 indicates the limiting curve ($n \longrightarrow \infty$) constructed from Eq. (6.3). The difference between this curve and the curve for the 2*s* level closely corresponds to Fig. 4. At the same time, for $\mathcal{H} < \mathcal{H}_{\min}$, even the order in the which the $\delta_{nl}(\mathcal{H})$ curves are arranged sharply differs from the order characteristic of the Zeldovich effect. For *ns* levels, $\delta_n \propto 1/\ln(\mathcal{H}/\ln^2\mathcal{H})$ when $\mathcal{H} \longrightarrow \infty$; i.e., $\delta_{ns} \longrightarrow 0$ slowly. As a result, the range of magnetic fields in which the Zeldovich effect shows up for even states is broad and covers $B = 10^{12}$–10^{14} G characteristic of neutron stars.

A comment is required on the relative positions of the 3*s* and 3*d* states. In the absence of a magnetic field, three degenerate states,[9] 3*s*, 3*p*, and 3*d*, of which 3*p* has a negative parity and does not interact with the other two states, while 3*s* and 3*d* are states with the same symmetry, correspond to the principal quantum number

---

[8] Note that n_f is identical to the quantum number n in Eq. (2.4) (but not to the principal quantum number n in the Coulomb problem).

[9] This is the so-called random degeneracy attributable to the hidden symmetry group of the Coulomb field [47, 48].

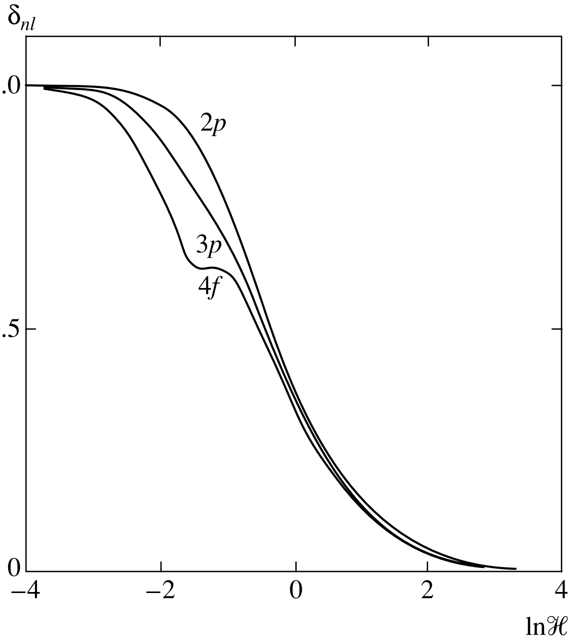


Fig. 7. The same as the previous figure for the odd $2p$, $3p$, and $4f$ levels corresponding to the 1^-, 2^-, and 3^- states of a one-dimensional hydrogen atom. The results from [11] were also used to construct the curve for the $2p$ level.

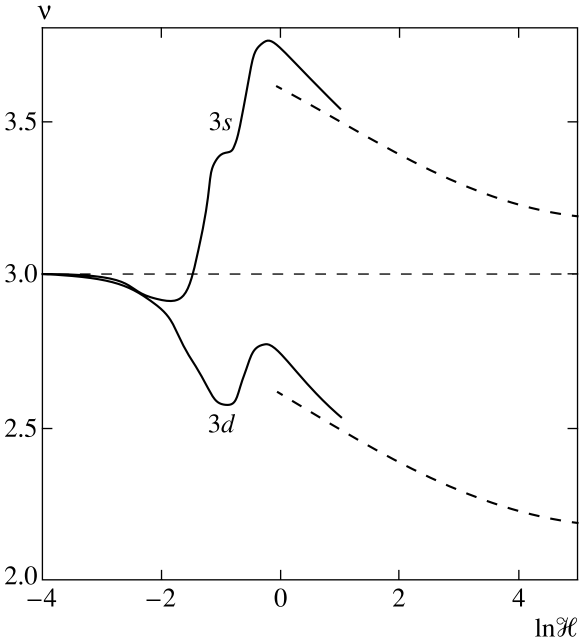


Fig. 8. $\nu = Z/\lambda$ versus magnetic field for the $3s$ and $3d$ states: the solid curves were recalculated from the data of [9], and the dashed curves were calculated from the asymptotic equation (2.13).

$n = 3$. These terms are mutually pushed apart with increasing $\mathcal{H}$ (Fig. 8), because their intersection would be in conflict with the Wigner–Neumann theorem.[10] As we see from Fig. 8, numerical calculations brought in this case only to $\mathcal{H} = 10$ [9] satisfactorily agree with the asymptotic equation (2.13). The $3d$ level lies below the $3s$ level, which can be qualitatively explained by the fact that its radial wave function has no nodes.[11] Of course, for complete clarity, we would like to continue the numerical calculations of the energies ε_{3s} and ε_{sd} at least to $\mathcal{H} \sim 50$.

The quantum defect δ_{1s} for the ground level differs markedly from δ_{ns}, $n \geq 2$. This quantity can also be used as a kind of a test for the Zeldovich effect by considering the ratio

$$R_{12} = \delta_1/\delta_2 = \left(\sqrt{\varepsilon_1/\varepsilon_2} - \sqrt{2\varepsilon_1}\right)^{-1}, \qquad (6.8)$$

where $\varepsilon_{1,2}$ are the binding energies of the two lower states, $1s$ and $2s$, expressed in atomic units $m_e e^4/\hbar^2 = 27.21$ eV. In Fig. 4 from [19], one curve was recalculated from the numerical data [9] and the other curve (curve *2*) corresponds to the solution of Eq. (2.13). These curves approach each other for $\mathcal{H} \gtrsim 200$, when a small parameter $r_0/a_B < 1/20$ appears in the problem. In this region, the shift of the $2s$ atomic level is related to the location of the lower $1s$ level by the relation peculiar to the Zeldovich effect.

The $\delta_{nl}(\mathcal{H})$ curves for three odd states shown in Fig. 7 indicate that an analog of the Zeldovich effect can exist for them in the range $1 \lesssim \mathcal{H} \lesssim 10^3$, although in this case there is no deep quasi-nuclear (or $1s$, as for even states) level in the system. In the above $\mathcal{H}$ range, the quantum defects for the $2p$, $3d$, and $4f$ states are close, which results in the characteristic ν_{nl} periodicity. However, at $\mathcal{H} \gtrsim 1000$, where the Zeldovich effect is observed for even states, the odd-level shifts are very small. Thus, for the $2p$ level at $\mathcal{H} = 2000$, $\varepsilon_{2p} = 0.991189\ldots$ (in atomic units; see [9, Table II]) and $\delta_{2p} = \varepsilon_{2p}^{-1/2} - 1 = 0.00443$. According to (6.6), the $2p$-level shift is only 1.2% of the spacing between the unperturbed levels $E_n^{(0)}$ with $n = 1$ and 2, and it is even smaller for $\mathcal{H} \gtrsim> 10^4$.

Turning to states with a magnetic quantum number $m \neq 0$, we will first obtain a simple estimate for the cutoff radius r_0 of the Coulomb field. In the adiabatic approximation, the effective potential for the electron

---

[10] See [15, §79]. Similarly, one may expect the $4s$, $4d$ (even) and $4p$, $4f$ (odd) levels to interact between themselves. Indeed, Fig. 7 shows a clear irregularity in the behavior of the $\delta_{4f}(\mathcal{H})$ curve.

[11] Note also that it follows from the formula for the diamagnetic level shift in a weak magnetic field [15] that $\nu_{nl} = n + [A_n - B_n l(l+1)]\mathcal{H}^2 + \ldots$, with $B_n > 0$. Therefore, the larger the orbital angular momentum l, the lower the values of ν_{nl} (for a particular n); this is also the case for finite field strengths $\mathcal{H}$ in view of the Wigner–Neumann theorem.

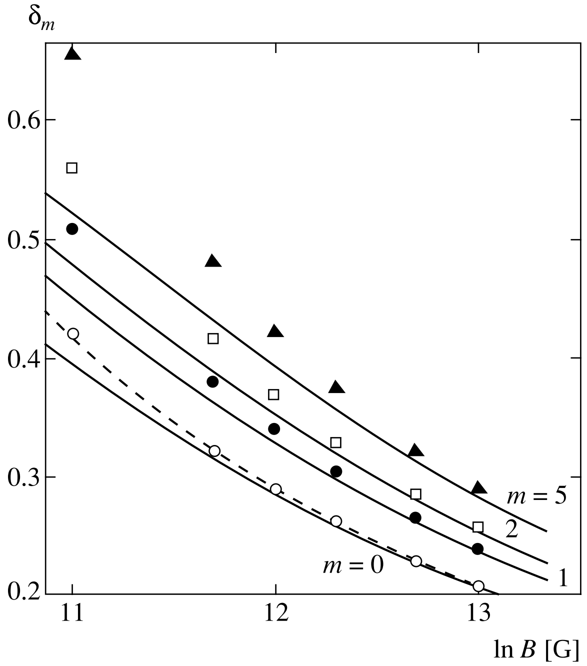


Fig. 9. Quantum defects δ_m for $n_\rho = 0$ states in a superstrong magnetic field. The solid curves were constructed from Eq. (2.13); the δ_m values extracted from the numerical calculations [9] are denoted by ○ ($m = 0$), ● ($m = 1$), □ ($m = 2$), and △ ($m = 5$).

motion along the field is given by expression (2.5). For transversally nodeless states ($n_\rho = 0$, $m < 0$), we have

$$U_{\mathrm{eff}}(z) = -\frac{1}{|m|!a_H}e^{\xi}\sum_{j=0}^{|m|}\frac{|m|!}{j!(|m|-j)!}(-\xi)^j \times\Gamma\left(|m|-j+\frac{1}{2},\xi\right), \qquad (6.9)$$

where $\xi = z^2/2a_H^2$ and $\Gamma(\alpha, \xi)$ is an incomplete gamma function. Hence, at $|z| \ll a_H$, we obtain (2.6') for $m = 0$ and

$$U_{\mathrm{eff}}(z) = -\frac{1}{\sqrt{2}a_H}\left\{\frac{\Gamma(|m|+1/2)}{|m|!} - \frac{\Gamma(|m|-1/2)}{4(|m|!)}\left(\frac{z}{a_H}\right)^2 + \dots\right\}, \quad |m| \geq 1, \qquad (6.10)$$

at $|z| \gg a_H$, the effective potential is a Coulomb one:

$$U_{\mathrm{eff}}(z) = -\frac{1}{|z|}\left\{1 - (|m|+1)\left(\frac{a_H}{z}\right)^2 + \dots\right\}. \qquad (6.11)$$

Joining these expressions, we see that the cutoff radius is equal in order of magnitude to

$$r_0 \approx [|m|!/\Gamma(|m|+1/2)]a_H. \qquad (6.12)$$

At $m = 0$, this estimate gives $r_0 \approx 0.564a_B$, which is close to the exact value of (5.2). At $|m| \gg 1$, the ratio r_0/a_B increases as $\sqrt{|m|}$. As a result, the accuracy of Eq. (2.13) decreases (at a fixed $\mathcal{H}$), which is also confirmed by our numerical calculation (Fig. 9).

Note that Eq. (2.13) for the level spectrum at $m \neq 0$ is identical to Eq. (5.1) if we set

$$r_0 = \exp\left\{\frac{1}{2}\left[-(\ln 2 + \gamma) + \sum_{k=1}^{|m|}\frac{1}{k}\right]\right\}a_H, \qquad (6.13)$$

i.e., $r_0/a_H = 0.530, 0.874, 1.12, 1.32, \dots$ for $m = 0, 1, 2, 3$.

The energies of the states with $n_\rho = 0$ and $m = 0, 1, \dots, 5$ in the range of magnetic fields $B = 10^{11}$–10^{13} G were calculated in [10] and used in Fig. 9. We see from this figure that, as the magnetic quantum number $|m|$ increases, the validity range for Eq. (2.13) is displaced toward increasingly large field strengths, and it begins only from $B \gtrsim 3 \times 10^{12}$ G for $m = 5$.

7. THE MEAN ATOMIC RADIUS, QUADRUPOLE MOMENT, AND RADIATIVE TRANSITION PROBABILITIES

At $\mathcal{H} \gg 1$, the size of the hydrogen atom transverse to the field is equal to the Landau radius a_H, while the longitudinal wave function is expressed in terms of the Whittaker function (see Appendix A):

$$\xi_\lambda(z) = \mathrm{const}\,\sigma W_{\nu,1/2}(2\lambda|z|), \quad -\infty < z < \infty, \qquad (7.1)$$

where $\nu = 1/\lambda$, $\sigma = 1$ for even states, and $\sigma = \mathrm{sgn}z$ for odd states. This function is the solution of the Schrödinger equation for a one-dimensional hydrogen atom to which the electron motion along the magnetic field (the z axis) reduces if $r_0 \sim a_H \ll a_B$. The integrals in the normalization and in the expression for the rms radius

$$a_\parallel = \langle z^2\rangle^{1/2} = \frac{1}{2\lambda} \times\left\{\int_0^\infty W_{\nu,1/2}^2(x)x^2dx\Big/\int_0^\infty W_{\nu,1/2}^2(x)dx\right\}^{1/2} \qquad (7.2)$$

were calculated in [41], which allows us to derive an analytical formula for $a_\parallel$ (see Appendix C). Let us intro-

Table 4. Characteristic sizes of a hydrogen atom in a superstrong magnetic field

$\mathcal{H}$	$\frac{a_H}{a_B}$	$a_\parallel/a_B$		β	
		1*s*	2*s*	1*s*	2*s*
0	–	1.000	3.74	–	–
100	0.100	0.353	3.25	3.53	32.5
213	6.85(–2)	0.303	3.10	4.42	45.2
425	4.85(–2)	0.264	2.97	5.44	61.2
851	3.43(–2)	0.232	2.86	6.76	83.4
1000	3.16(–2)	0.225	2.83	7.12	89.6
2000	2.24(–2)	0.200	2.74	8.93	122
2130	2.17(–2)	0.198	2.73	9.12	126
4255	1.53(–2)	0.178	2.66	11.6	174
5000	1.41(–2)	0.173	2.64	12.3	187
1.0(4)	1.00(–2)	0.157	2.57	15.7	257
5.0(4)	4.47(–3)	0.126	2.44	28.2	546
1.0(5)	3.16(–3)	0.116	2.40	36.7	759

Note: The parameter $\beta = a_\parallel/a_H$ characterizes the degree of elongation of the electron cloud along the field.

duce the ratio $\beta = a_\parallel/a_H$ that characterizes the shape of the atom (i.e., the electron localization region). The values of a_H, $a_\parallel$, and β are given in Table 4. We see from this table that the atom is cigar-shaped (for 1*s* at $\mathcal{H} \lesssim 1000$) or even needle-shaped (for *ns* states, $n \geq 2$) [49]. Its longitudinal size also decreases with increasing $\mathcal{H}$, especially for the 1*s* level, asymptotically when $\mathcal{H} \longrightarrow \infty$

$$\beta = \frac{a_\parallel}{a_H} \propto \frac{\sqrt{\mathcal{H}}}{\ln(\mathcal{H}/\ln^2\mathcal{H})}, \quad n = 1. \tag{7.3}$$

For excited ($n \geq 2$) states, $a_\parallel \propto n^2 a_B$ is on the order of the Bohr radius, although it also decreases with increasing $\mathcal{H}$.

The deformation of the electron cloud by the magnetic field gives rise to a quadrupole moment of the atom, which affects the interatomic interaction. In a strongly magnetized hydrogen plasma, apart from the ordinary Van der Waals attraction, the quadrupole–quadrupole interaction between atoms can also become significant [50]. Recently, the quadrupole moment has been numerically calculated for the ground state of a hydrogen atom [14] by using (i) the variational method and (ii) the solution of the differential equations that follow from the expansion of the wave function in terms of Landau orbitals. Below, we give a simple asymptotic formula for Q that refers to $\mathcal{H} > \mathcal{H}_{\min}$.

In view of axial symmetry, the quadrupole moment tensor is

$$\begin{gathered} Q_{xx} = Q_{yy} = \frac{1}{2}Q, \quad Q_{zz} = -Q, \\ Q_{\alpha\beta} = 0, \quad \alpha \neq \beta, \\ Q = 2\langle z^2 \rangle - \langle \rho^2 \rangle \end{gathered} \tag{7.4}$$

($Q_{zz} < 0$, because the electron charge $e < 0$). Using formula (C.2), we find that

$$Q = 2\left\{a_\parallel^2(\mathcal{H}) - \frac{1+|m|}{\mathcal{H}}\right\}, \quad n_\rho = 0, \tag{7.5}$$

$$\begin{gathered} a_\parallel^2 = \frac{18\nu^2}{(2-\nu)^2(3-\nu)^2} \\ \times \frac{{}_3F_2(-\nu, -\nu, 3; 4-\nu, 4-\nu; 1)}{{}_3F_2(-\nu, -\nu, 1; 2-\nu, 2-\nu; 1)} \end{gathered} \tag{7.6}$$

(here, $0 < \nu = \lambda^{-1} < 1$ and $\lambda = \lambda(\mathcal{H})$ can be determined from Eq. (2.13)). As a result, we obtain the solid curve for $Q(\mathcal{H})$ in Fig. 10. At $\mathcal{H} \geq 50$, this curve agrees well with the numerical calculation, especially for method (ii), which, as was noted in [14], is more accurate. Q decreases with increasing $\mathcal{H}$ (at $\mathcal{H} \geq 10$), because a superstrong magnetic field compresses the atom not only in the direction transverse to B but also in the longitudinal direction. The asymptotic formula for $Q(\mathcal{H})$ when $\mathcal{H} \longrightarrow \infty$ (see Appendix B) also clearly follows from (7.5) and (7.6).

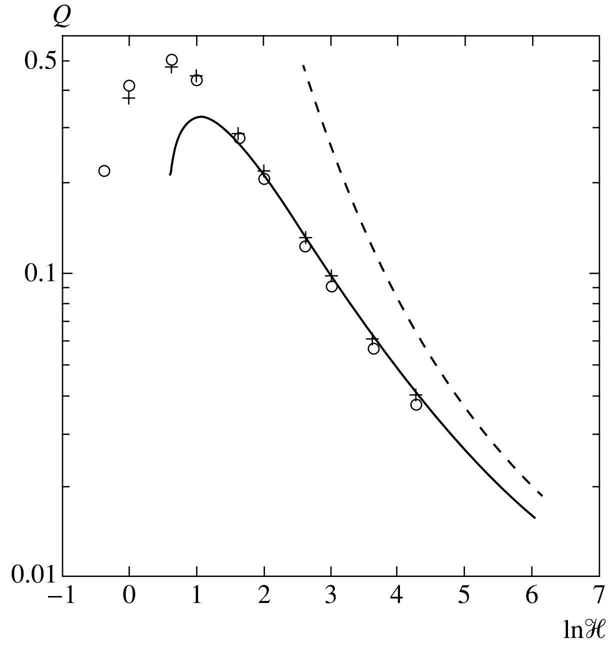


Fig. 10. The ground-state quadrupole moment for a hydrogen atom. The solid curve was constructed from formula (7.5); the numerical calculations [14] are denoted by ○ and + (for methods (i) and (ii), respectively; see the text). The dashed curve corresponds to the asymptotic formula (B.5).

Measurements of the spectrum of the lines emitted during transitions between atomic levels can also give useful information. Let us consider the $E1$ transitions from the excited odd electron states with $m = 0$ in the lower Landau zone, $n_\rho = 0$, to the ground 1s level at $\mathcal{H} > \mathcal{H}_{\min}$, when $a_H \ll a_\parallel \ll a_B$ and the Zeldovich effect takes place. In the decay probability

$$w(n, P = -1 \longrightarrow 1s) = \frac{4\omega^3}{3c^3}|\mathbf{d}_{fi}|^2,$$

only the matrix element of the dipole moment component along the field, i.e., in a one-dimensional hydrogen atom, is nonzero. Assuming that the binding energy $\varepsilon_{1s} = 1/2\nu^2 \gg 1$ (i.e., $\nu = \delta_{1s} \ll 1$), we substitute the simpler function (A.7) for the Whittaker function (7.1). For the initial state

$$E_n^- = -\frac{1}{2n^2},$$

$$\psi_n^-(z) = \sqrt{\frac{2}{n^3}}\exp\left(-\frac{|z|}{n}\right){}_1F_1\left(1-n, 2; \frac{2|z|}{n}\right)$$

(as we noted above, the spectrum of these levels coincides with the spectrum of s levels in a central Coulomb field). Since function (7.1) cuts off the matrix element at distances $|z| \sim \nu \ll 1$, it will suffice to take the asymptotics at zero of $\psi_n^-(z)$: $\psi_n^-(z) = \sqrt{2/n^3}(z - z^2 + \ldots)$. As a result, we obtain

$$\langle n^-|\hat{d}_z|1s\rangle = -\int_{-\infty}^{\infty}\psi_\lambda(z)z\psi_n^-(z)dz = -(2\nu)^{5/2}n^{-3/2} + \ldots, \qquad (7.7)$$

$$w(n, P = -1 \longrightarrow 1s) = \frac{16}{3}\frac{\alpha^3}{n^3\nu}(1 - c_1\nu + \ldots),$$

where $\alpha = 1/137$, $c_1 = 3 - 2\ln 2 \approx 1.61$ (see (A.7) and (A.13)), the transition probability is measured here in units of $m_e e^4/\hbar^3 = 4.13 \times 10^{16}$ s^{-1}, and $n \equiv n_f^- = 1, 2, 3, \ldots$ is the quantum number of the initial state with a parity of $P = -1$. As the ground level deepens, probability (7.7) increases, while for a $\bar{p}p$ atom, it decreases:[12]

$$w(np \longrightarrow \nu s) = \frac{16\alpha^3}{9n^3}\nu\left[1 - \frac{2}{3}\nu + O(\nu^2)\right]. \qquad (7.7')$$

This difference stems from the fact that $\psi_n^-(z) \propto z$ in a one-dimensional hydrogen atom at small distances, while $\chi_{np}(r) \propto r^2$ in a three-dimensional hydrogen atom. Therefore, the probability of the electron being found in the localization region of the wave function for the final 1s state in the latter case is much lower.

Let us compare (7.7) with the probability of a spontaneous $E1$ 2p ⟶ 1s transition with a switched-off magnetic field:

$$\frac{w(1^- \longrightarrow 1s)}{w_H(2p \longrightarrow 1s)} = \frac{2187}{16\nu} \approx 140\nu^{-1}, \qquad (7.8)$$

where [51] $w_H(2p \longrightarrow 1s) = (2/3)^8\alpha^3 = 6.27 \times 10^8$ s^{-1} (in both cases, we consider the transition from the level closest to the ground level). Thus, the probability of an $E1$-transition to a deep s level in a superstrong magnetic field is much higher than that in a free ($\mathcal{H} = 0$) hydrogen atom. At $B \sim 10^{13}$ G, we have the following estimates: $\varepsilon_{1s} \sim 300$ eV and $\nu \sim 0.2$, whence $w \propto 10^{12}$ s^{-1}. Accurate radiative transition probabilities can be numerically calculated by using analytical expressions for the integrals with the Whittaker function and the degenerate hypergeometric function calculated in connection with the theory of a $\bar{p}p$ atom [41].

To conclude this section, we note that the dipole approximation is applicable if $kR \ll 1$, where $k = w/c$ is

---

[12] A comparison with the numerical calculation [41] shows that the first two terms of this expansion can be retain for $\nu \leq 0.3$.

the photon momentum, and R is the characteristic size of the radiator. In our case,

$$a_H \ll a_\parallel, \quad \omega = E_i - E_f = \frac{1}{2\nu^2}\left(1 - \frac{\nu^2}{n^2}\right) \approx \frac{1}{2\delta^2}$$

and the dimensionless parameter $kR \sim a_\parallel/2c\delta^2$ changes within the range from 0.013 at $\mathcal{H} = 10^3$ to 0.02 at $\mathcal{H} = 10^5$ (here, $\delta \equiv \delta_{1s}$ is the quantum defect of the 1*s* level; see Table 3). Therefore, the dipole approximation has a good accuracy, while the angular photon distribution is

$$w(\theta) \propto [\mathbf{n} \times \mathbf{d}_{fi}]^2 \propto \sin^2\theta, \tag{7.9}$$

where $\mathbf{n} = c\mathbf{k}/\omega$ and θ is the angle between $\mathbf{k}$ and $\mathbf{B}$; in particular, no photons emerge along field $\mathbf{B}$.

Note also[13] that these photons have a 100% linear polarization in the $(\mathbf{B}, \mathbf{k})$ plane. The matrix element that corresponds to the photon emission is $M \propto \mathbf{B} \cdot \mathbf{e} \propto z_i e_i$, where $\mathbf{e}$ is the photon polarization vector, and z is a unit vector along field $\mathbf{B}$. The three-dimensional photon density matrix is (see [51])

$$\begin{aligned}\rho_{ik} &= \frac{1}{2}(e_i^{(1)}e_k^{(1)} + e_i^{(2)}e_k^{(2)}) \\ &+ \frac{1}{2}\xi_1(e_i^{(1)}e_k^{(2)} + e_i^{(2)}e_k^{(1)}) - \frac{1}{2}\xi_2(e_i^{(1)}e_k^{(2)} - e_i^{(2)}e_k^{(1)}) \\ &+ \frac{1}{2}\xi_3(e_i^{(1)}e_k^{(1)} - e_i^{(2)}e_k^{(2)}),\end{aligned} \tag{7.10}$$

where ξ_i are the Stokes parameters and the unit vectors $\mathbf{e}^{(1)}$ and $\mathbf{e}^{(2)}$ were chosen as follows:

$$\mathbf{e}^{(1)} = \frac{1}{\sin\theta}[\mathbf{z} \times \mathbf{k}], \quad \mathbf{e}^{(2)} = [\mathbf{e}^{(1)} \times \mathbf{n}].$$

Substituting this into the expression for $|M|^2 \propto z_i z_k \rho_{ik}$ yields

$$|M|^2 = \frac{1}{2}(\mathbf{e}^{(2)} \cdot \mathbf{z})^2[1 - \xi^{(3)}], \tag{7.11}$$

where the first and second factors define the angular distribution (7.9) and the photon polarization, respectively. On the other hand, $|M|^2 \propto 1 + \boldsymbol{\xi}_f\boldsymbol{\xi}_d$, where $\boldsymbol{\xi}_d$ and $\boldsymbol{\xi}_f$ are the Stokes parameters that characterize the detector and the polarization of the emitted photon [51]. Equation (7.11) yields $\boldsymbol{\xi}_f = (0, 0, -1)$, and we obtain hence

$$\rho_{ik} = \frac{1}{2}\begin{pmatrix} 1 + \xi_3 & \xi_1 - i\xi_2 \\ \xi_1 + i\xi_2 & 1 - \xi_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{7.12}$$

This implies that the photon is linearly polarized along the unit vector $\mathbf{e}^{(2)}$, i.e., in the $(\mathbf{B}, \mathbf{k})$ plane.

8. CONCLUSIONS

(1) In their work entitled “Interaction and Annihilation of Antinucleons with Nucleons,” which was written in 1957 but not published, L.L. Landau, I.Ya. Pomeranchuk, and K.A. Ter-Martirosyan investigated the spectrum of atomic levels in a Coulomb field distorted at small distances by strong interaction. As in [35], these authors joined the inner and outer wave functions at point $r = r_0 \ll a_B$, where the strong and Coulomb potentials are equal in magnitude. They derived an equation for level energies that is essentially identical to Eq. (5.1) and qualitatively analyzed the influence of strong interaction between p and $\bar{p}$ on the level shifts in a $p\bar{p}$ atom.

One of us (V.P.) is very grateful to K.A. Ter-Martirosyan for information about this work and for the possibility of familiarizing ourselves with it in the manuscript. Note, however, that it was unknown to us when studies [35, 41] were performed.

(2) Table 5 gives the binding energies for the first two *s* levels, their quantum defects, and the limiting values of δ_∞ for highly excited levels. We took ε_1 from [9] and calculated ε_2 from Eq. (2.13)—a procedure that is inverse to the procedure used previously in the theory of a $p\bar{p}$ atom [35]. Note that the relative 2*s*-level shift (6.4) changes from $\rho_2 = 0.584$ at $\mathcal{H} = 400$ to 0.393 at $\mathcal{H} = 10^5$. For $\varepsilon_2 = 8.5$ eV ($\mathcal{H} \approx 3100$), this level is exactly halfway between the unshifted Coulomb levels $E_n^{(0)}$ with $n = 1$ and 2.

We see from Table 5 that the inequality $\delta_\infty < \delta_2$ that is inverse to the inequality characteristic of the Zeldovich effect is satisfied at $\mathcal{H} \lesssim 200$. Therefore, the lower limit for the magnetic fields required for this effect to show up in the spectrum of atomic levels is $\mathcal{H}_{\min} \sim 200$ (for even states), or $B \sim B_{\min} \sim 5 \times 10^{11}$ G, which also follows from Fig. 5. On the other hand, $\mathcal{H}_{\min} \sim 1$ for odd states (1^-, 2^-, etc.) (see Fig. 6).

(3) There is a significant difference between the Zeldovich effects for three-dimensional local potentials and in the problem with a magnetic field. In the former case, the rearrangement can be repeated several times as the coupling constant g increases: thus, for a square potential well, this occurs [1] at $g \approx g_N = (2N + 1)^2\pi^2\hbar^2/8r_0^2$, $N = 0, 1, 2\ldots$ In the latter case, only the 1*s* ground level can become deep with increasing $\mathcal{H}$, while the remaining levels lie at $\nu > 1$ (which is a characteristic property of the one-dimensional Coulomb potential; see Appendix D). Therefore, only one cycle of spectrum rearrangement takes place in a magnetic field (however, each pair of quantum numbers n_ρ and $|m|$ has its own cycle).

---

[13] This remark, as well as the considerations given below, belong to M.I. Vysotskiĭ.

Table 5

B, G	$\mathcal{H}$	ε_1	ε_2	δ_1	δ_2	δ_∞
0	0	13.6	3.40	1.0	1.0	1
2.35(11)	100	103.1	6.90	0.3632	0.397	0.3937
4.7(11)	200	128.6	7.31	0.3252	0.363	0.3635
9.4(11)	400	159.2	7.64	0.2923	0.334	0.3357
1.0(12)	425	161.5	7.65	0.2902	0.332	0.3334
2.0(12)	851	198.5	7.96	0.2618	0.307	0.3082
2.35(12)	1000	208.4	8.05	0.2554	0.300	0.3027
4.70(12)	2000	253.2	8.33	0.2318	0.278	0.2807
5.0(12)	2130	257.1	8.35	0.2300	0.276	0.2788
1.0(13)	4255	309.6	8.61	0.2096	0.257	0.2592
1.18(13)	5000	323.1	8.67	0.2052	0.252	0.2549
2.53(13)	1.0(4)	384.8	8.91	0.1880	0.235	0.2378
B_{cr}	1.878(4)	447.9	9.11	0.1743	0.222	0.2237
1.18(14)	5.0(4)	560.0	9.41	0.1559	0.203	0.2045
2.35(14)	1.0(5)	650.1	9.59	0.1447	0.191	0.1926
5.0(14)	2.13(5)	759	9.81	0.1339	0.177	0.1782
1.0(15)	4.3(5)	868	9.97	0.1252	0.168	0.1688
5.0(15)	2.13(6)	1.16(3)	10.3	0.1082	0.149	0.1501
1.0(16)	4.3(6)	1.31(3)	10.4	0.1021	0.144	0.1431

Note: ε_1 and ε_2 are the binding energies (in eV) of the 1*s* and 2*s* levels in a magnetic field, $\delta_{1(2)}$ are their quantum defects, δ_∞ is the quantum defect for highly excited levels, and $B_{\mathrm{cr}} = 4.414 \times 10^{13}$ G is the Schwinger field in quantum electrodynamics [17].

(4) The characteristic magnetic field strengths are B_a and $B_{\mathrm{cr}} = \alpha^{-2}B_a$, at which the spacing between the adjacent Landau levels is

$$\hbar\omega_H = \frac{\hbar eB}{m_e c} = \begin{cases} m_e e^4/\hbar^2, & B = B_a, \\ m_e c^2, & B = B_{\mathrm{cr}}. \end{cases} \tag{8.1}$$

The ratio $\eta = \varepsilon_0(\mathcal{H})/\hbar\omega_H$ decreases with increasing field (Fig. 11); in particular, $\eta = 0.831$ at $B = B_a$ and $\eta = 8.77 \times 10^{-3}$ at $B = B_{\mathrm{cr}}$. Thus, in a superstrong magnetic field, the energy range of the atomic spectrum, including the deep 1*s* level, accounts for only a small fraction of $\hbar\omega_H$.

Note that $\hbar\omega_H/m_e c^2 = B/B_{\mathrm{cr}}$. At $B \geq B_{\mathrm{cr}}$, the transverse electron motion is completely relativistic. However, for the ground state, this has almost no effect on the energy spectrum (see [5, 7] and Section 4 of this paper). The validity range for the approach developed above, including Eq. (2.13), is

$$\alpha B_{\mathrm{cr}} \leq B \leq \alpha^{-1}B_{\mathrm{cr}}, \quad \alpha = e^2/\hbar c = 1/137. \tag{8.2}$$

For even more intense magnetic fields, the influence of an external field on the vacuum properties and the anomalous electron magnetic moment should be taken into account.

(5) Let us explain the relationship between Eqs. (5.1) and (2.13). They include the "strong" scattering length a_s and the Coulomb-modified scattering length $a_0 = a_{cs}$ in the $l = 0$ state. When we pass from (5.1) to (2.13), the "large" logarithm $\ln(r_0/a_B)$ goes into the Coulomb renormalization of the scattering length: $1/a_s \longrightarrow 1/a_{cs}$, which can be significant if the system has a (real or virtual) level with a low energy. At the same time, the effective radius changes only slightly when the Coulomb interaction is included. Let us illustrate this behavior for *NN* scattering in the ${}^1S_0$ state, where [52]

$$a_s = -23.748 \pm 0.010 \text{ Fm}, \quad r_s = 2.75 \pm 0.05 \text{ Fm},$$

$$a_{cs} = -7.8098 \pm 0.0023 \text{ Fm}, \quad r_{cs} = 2.767 \pm 0.010 \text{ Fm},$$

and the subscripts *a* and *cs* refer to the *np* and *pp* systems, respectively. The virtual level $|\epsilon| = 0.067$ MeV (a singlet deuteron) corresponds to a large value of a_s.

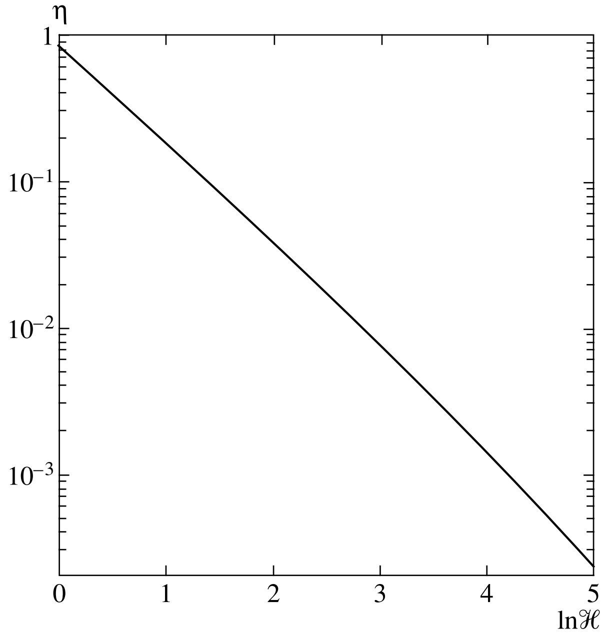


Fig. 11. The ratio of ground-level binding energy ϵ_0 to Landau energy $\hbar\omega_H$.

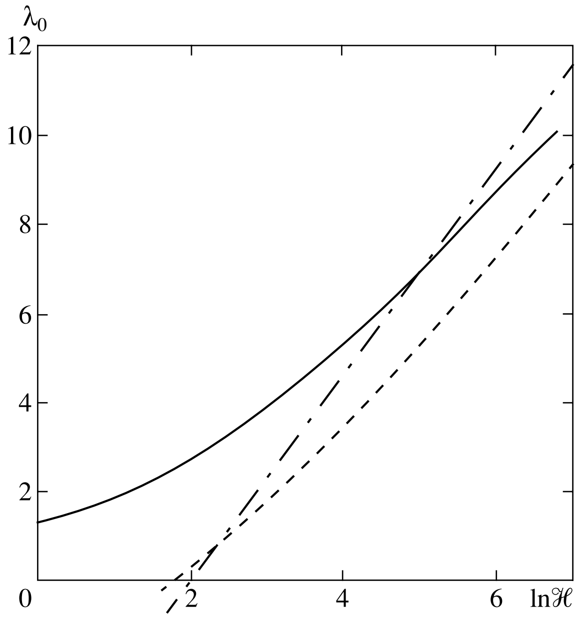


Fig. 12. The dependence $\lambda(\mathcal{H})$ for the ground level. The solid curve was constructed from the results of numerical calculations [9] for $\mathcal{H} < 10^5$ and from Eq. (2.13) for $\mathcal{H} > 10^5$. The dashed and dash-dotted curves correspond to the asymptotic formula (B.3) and the paper [13].

The determination of a_{cs} and r_{cs} for an arbitrary potential $V_s(r)$ can be found, for example, in [53–55].

(6) The discussion of hydrogen atomic energy levels for $\mathcal{H} \gg 1$ given in [15, §112] requires some clarifications.

First, formula (1.1) derived in [15] is not identical to the asymptotic formula for the ground-level binding energy $\varepsilon_0(\mathcal{H})$ when $\mathcal{H} \longrightarrow \infty$. A comparison of (1.1) and (B.3) shows that the correct asymptotic formula for $\varepsilon_0(\mathcal{H})$ is not $\varepsilon_0 \approx \frac{1}{2}\ln^2\tilde{c}\,\mathcal{H}$ with some constant $\tilde{c}$, but additionally contains $1/\ln^2\mathcal{H}$ under the logarithm sign. In this case, as was noted above, formula (1.1) has virtually no validity range.[14]

Further, the assertion [15, 56] that all the levels of the discrete spectrum, except the ground level, in a superstrong magnetic field are doubly degenerate (in parity $P = \pm 1$) and are described by the Balmer formula is unlikely to be true.[15] The latter is true for odd levels (see Fig. 7), but even levels in fields $B \sim 10^{11}$–10^{14} G typical of neutron stars are greatly shifted relative to $E_n^{(0)}$, which is seen from Fig. 5 and, particularly clearly, from Fig. 1. The corresponding quantum defects $\delta_n^+ \sim [\ln(\mathcal{H}/\ln^2\mathcal{H})]^{-1} < 0.05$ only for $\mathcal{H} > 10^{11}$, or $B > 10^{20}$ G; even in this case, the shifts of even levels relative to odd levels are not very small ($\rho_2 \approx 1/8$; see (6.6)). If such magnetic fields actually existed in the early universe [57], obviously there were still no hydrogen atoms at the time.

(7) Thus, the data on the quantum defects of *ns* levels extracted from previous numerical calculations [6–11] indicate that the Zeldovich effect manifests itself in the energy spectrum of a hydrogen atom in extremely strong magnetic fields, $B \gtrsim 10^{12}$ G. Such fields are typical of neutron stars and, possibly, magnetic white dwarfs [3, 4]. The establishment of this effect with increasing $\mathcal{H}$ is clearly seen from Figs. 5–8. These figures show that, essentially, it has already been observed (in a numerical experiment). A confirmation of this conclusion by direct astrophysical observations would be of great interest.

ACKNOWLEDGMENTS

We would like to thank M.B. Voloshin, M.I. Vysotskii, V.D. Mur, A.I. Nikishov, I.D. Novikov, L.B. Okun', Yu.A. Simonov, and K.A. Ter-Martirosyan for discussions of the results obtained and useful remarks, which helped to remove several inaccuracies and to improve the text; we thank S.G. Pozdnyakov for help with the numerical calculations, V.M. Weinberg for reference

---

[14] In [15], formula (1.1) is ascribed to Elliott and Loudon, but it is not contained in their paper [56] at all (these authors calculated the energy $\varepsilon_0(\mathcal{H})$ numerically). Note also that, according to [13], $\lambda \approx \ln(0.01\mathcal{H})$, but this formula is not identical to the correct asymptotic formula (B.3); see also Fig. 12.

[15] This mistake probably results from the use of a one-dimensional hydrogen model atom [46, 56]. This approximation is valid in the limit $B \longrightarrow \infty$. However, for magnetic fields $B \sim 10^{11}$–10^{14} G, it is not yet accurate enough (when calculating the energies of even levels), which is clearly seen from Figs. 1 and 4–6.

to [12, 13], and M.N. Markina for help in preparing the manuscript. This study was supported in part by the Russian Foundation for Basic Research (project no. 01-02-16850).

APPENDIX A

The Schrödinger equation in a Coulomb field $U(r) = -Z/r$ has a solution that decreases at infinity:

$$\begin{aligned}\chi_l(r) &= W_{\nu,\, l+1/2}(x)\\ &= e^{-x/2}x^{\nu}\left[1+\frac{(l+\nu)(l+1-\nu)}{x}+O(x^{-2})\right],\end{aligned}\tag{A.1}$$

where $\nu = Z/\lambda$, $\lambda = \sqrt{-2E}$, $x = 2\lambda r$, l is the orbital angular momentum, and W is the Whittaker function. In general, this function has a singularity at zero:

$$W_{\nu,\, l+1/2}(x) = \frac{\Gamma(2l+1)}{\Gamma(l+1-\nu)}x^{-l}+\ldots, \quad x \longrightarrow 0. \tag{A.2}$$

In particular, for the s states,

$$\begin{aligned}W_{\nu,\, 1/2}(x) &= \frac{1}{\Gamma(1-\nu)}\\ &\times[1-c_1x\ln x-c_2x+c_3x^2\ln x+O(x^2)],\end{aligned}\tag{A.3}$$

where $c_1 = \nu$, $c_2 = 1/2 + \nu[\psi(1-\nu) - 1 + 2\gamma]$, $c_3 = \frac{1}{2}\nu^2$, etc.

Hence, we have for $r \longrightarrow 0$

$$\begin{aligned}&\frac{d}{dr}\ln W_{\nu,\, 1/2}(2\lambda r)\\ &= -\{\lambda+2[\ln\lambda r+\psi(1-\lambda^{-1})+c_0]+O(r\ln\lambda r)\},\end{aligned}\tag{A.4}$$

where $\nu = 1/\lambda$, $c_0 = 2\gamma + \ln 2$, and $Z = 1$. Using this expansion when joining with the wave function from the inner region $r < r_0$ immediately leads to Eqs. (5.1) and (2.13).

A singularity of type (A.2) disappears for an integer $\nu = n \geq l + 1$ when solution (A.1) differs from the wave function for the discrete hydrogen atomic spectrum only by the numerical factor:

$$\begin{aligned}\chi_{nl}(r) &= \frac{(-1)^{n-l-1}}{n\sqrt{(n+l)!(n-l-1)!}}\\ &\times W_{n,\, l+1/2}\left(\frac{2r}{n}\right) = \frac{2^{l+1}\xi_{nl}}{(2l+1)!n^{3/2}}r^{l+1}\\ &\times\left[1-\frac{r}{l+1}+\frac{2n^2+l+1}{(2l+2)(2l+3)n^2}r^2+\ldots\right], \quad r \longrightarrow 0,\end{aligned}\tag{A.5}$$

where

$$\xi_{nl} = \sqrt{\left(1-\frac{1^2}{n^2}\right)\ldots\left(l-\frac{l^2}{n^2}\right)}, \quad \xi_{n0} \equiv 1,$$

$$\int_0^{\infty}\chi_{nl}^2 dr = 1.$$

When $\nu \longrightarrow n$, given the identity

$$\lim_{\nu \to n}\frac{\psi(1-\nu)}{\Gamma(1-\nu)} = (-1)^n(n-1)!, \quad n = 1, 2, 3, \ldots,$$

expansion (A.3) transforms into (A.5) for $l = 0$.

Let us now determine the form of function (A.1) for a deep s level. We find from the integral representation [20]

$$W_{\nu,\, 1/2}(x) = \frac{e^{-x/2}}{\Gamma(1-\nu)}\int_0^{\infty}e^{-t}\left(1+\frac{x}{t}\right)^{\nu}dt \tag{A.6}$$

for $\nu \longrightarrow 0$ that

$$W_{\nu,\, 1/2}(x) = e^{-x/2}[1+\nu u(x)+O(\nu^2)], \tag{A.7}$$

where

$$u(x) = \int_0^{\infty}e^{-t}\ln(t+x)dt = \ln x - e^x\mathrm{Ei}(-x), \tag{A.8}$$

and $\mathrm{Ei}(z)$ is the integral exponent (see 4.837.1 and 6.223 in [58]); $u(0) = -\gamma$, $u(x) = \ln x + O(x^{-1})$ for $x \longrightarrow \infty$ and $u(x) = 0$ for $x = x_0 = 0.28501\ldots$. Taking into account the integrals

$$\int_0^{\infty}e^{-\lambda x}x^{b-1}\ln x dx = \Gamma(b)\lambda^{-b}[\psi(b)-\ln\lambda], \tag{A.9}$$

$$\int_0^{\infty}e^{-x}x^{b-1}u(x)dx = \Gamma(b)\psi(b+1), \quad b > 0, \tag{A.10}$$

we can easily see that the normalized (to within ν^2) wave function is

$$\begin{aligned}\chi_{\nu}(z) &= c_{\nu}e^{-\lambda|z|}[1+\nu u(2\lambda|z|)+O(\nu^2)],\\ c_{\nu} &= \nu^{-1/2}[1-(1-\gamma)\nu+\ldots],\end{aligned}\tag{A.11}$$

whence follows formula (B.4). In the limit $\nu \longrightarrow 0$ (a deep level), we obtain the function

$$\chi_{\lambda}(z) = \sqrt{\lambda}\exp(-\lambda|z|), \quad \lambda = 1/\nu, \tag{A.12}$$

which corresponds to a one-dimensional δ potential.

In Section 7, when calculating the dipole transition matrix element, we encountered the integral

$$J_k = \frac{1}{k!}\int_0^\infty e^{-x}u(2x)x^k dx = \psi(k+1) + \ln 2 + F_k(1/2) \qquad \text{(A.13)}$$

for integer $k = 2, 3, \ldots$. Introducing the parameter μ and denoting

$$F_k(\mu) = -\frac{1}{2^{k+1}k!}\int_0^\infty e^{\mu x}x^k \mathrm{Ei}(-x)dx, \quad \mu < 1, \qquad \text{(A.14)}$$

we obtain

$$F_0(\mu) = -\frac{1}{2\mu}\ln(1-\mu), \qquad F_1(\mu) = \frac{1}{4\mu^2}\left[\frac{\mu}{1-\mu} + \ln(1-\mu)\right], \qquad \text{(A.15)}$$

$$F_k(\mu) = \frac{1}{2k}\frac{d}{d\mu}F_{k-1}(\mu) \qquad \text{(A.16)}$$

(see 6.224.1 in [58]). Hence, for $\mu = 1/2$,

$$J_0 = 2\ln 2 - \gamma, \quad J_1 = 2 - \gamma,$$

$$J_2 = 2\ln 2 + 1 - \gamma = 1.809, \quad J_3 = 2.089,$$

etc., which gives the coefficient c_1 in (7.7)

APPENDIX B

Here, we consider the asymptotic formula for the lower-level energy in the limit $\mathcal{H} \longrightarrow \infty$.

In a superstrong magnetic field, this level can, in principle, sink arbitrarily deeply [46, 56]. Assuming that $\lambda \gg 1$ and taking into account the fact that [20] $\psi^{(n)}(1) = (-1)^{n+1}n!\zeta(n+1)$, we write Eq. (2.13) as

$$\mathcal{H} = \exp\left\{\lambda + 2\ln\lambda + k_0 - \sum_{n=1}^{\infty} k_n\lambda^{-n}\right\}, \qquad \text{(B.1)}$$

where $k_0 = \gamma + \ln 2$, $k_n = 2\zeta(n+1)$, and $\zeta(s)$ is the Riemann zeta function: $k_1 = \pi^2/3$, $k_2 = 2.404$, $k_3 = \pi^4/45$, etc. The first term of this expansion is identical to formula (1.2), while the first three terms correspond to Eq. (4.25) in the paper by Hasegawa and Howard [16]. The relative error of this approximation is

$$\delta_3 \equiv \frac{|\Delta\mathcal{H}|}{\mathcal{H}} \approx \frac{\pi^2}{3\lambda} = \frac{2.33}{\sqrt{\epsilon}} \qquad \text{(B.2)}$$

and slowly decreases with increasing level binding energy ε (here, ε is in atomic units). If we retain the four expansion terms in (B.1), including $-\pi^2/3\lambda$, then the error is $\delta_4 \approx 1.2\varepsilon^{-1}$. For example, $\delta_3 = 0.71$ and $\delta_4 = 0.06$ at $\mathcal{H} = 10^5$, $\delta_3 = 0.52$ and $\delta_4 = 0.037$ at $\mathcal{H} = 10^6$, and $\delta_3 = 0.33$ and $\delta_4 = 0.017$ at $\mathcal{H} = 10^8$. Thus, Eq. (2.16), which contains only elementary functions, defines the dependence $\mathcal{H}(\lambda)$ at $\mathcal{H} > 10^5$ with an accuracy of several percent, which is probably high enough for astrophysical applications. At the same time, in the range $100 < \mathcal{H} \lesssim 10^4$, only Eq. (2.13) defines the magnetic field for the 1*s*-level energy with an acceptable accuracy.

We derive the sought-for asymptotic formula for $\varepsilon_0(\mathcal{H}) = \lambda^2/2$ from (B.1) by an iterative method:

$$\lambda = \ln\frac{\mathcal{H}}{\ln^2\mathcal{H}} - (\gamma + \ln 2) + \ldots = \ln\frac{c_\infty\mathcal{H}}{\ln^2\mathcal{H}} + O\left(\frac{\ln\ln\mathcal{H}}{\ln\mathcal{H}}\right), \quad \mathcal{H} \longrightarrow \infty, \qquad \text{(B.3)}$$

$$\varepsilon_0(\mathcal{H}) \approx \frac{1}{2}\ln^2\left(c_\infty\frac{\mathcal{H}}{\ln^2\mathcal{H}}\right), \qquad \text{(B.4)}$$

where $c_\infty = 0.2807$. In the series of even levels with an arbitrary magnetic quantum number m, the asymptotic formula for the deep level is also given by (B.3) but with a different constant:

$$c_\infty \longrightarrow c_\infty^{(m)} = \exp\left\{-\left[\gamma + \ln 2 + \sum_{k=1}^{|m|}\frac{1}{k}\right]\right\}$$

(the binding energy of these levels decreases with increasing $|m|$).

For the quadrupole moment of the ground state, we find using formulas (B.3) and (C.4) that

$$Q = \frac{1}{\lambda^2} + \frac{5}{3\lambda^3} + O(\lambda^{-4}) = \Lambda^{-2} + \left[2(\gamma + \ln 2) + \frac{5}{3}\right]\Lambda^{-3} + \ldots, \qquad \text{(B.5)}$$

$$\Lambda = \ln\frac{\mathcal{H}}{\ln^2\mathcal{H}}$$

(cf. expansion (4) from [14]). However, the asymptotics (B.3)–(B.5) are established very slowly: even at $\mathcal{H} = 10^5$, the ground-level binding energy calculated from (B.4) differs from its numerically calculated value [9] by a factor of 1.7; see Fig. 12.

A similar situation also takes place in several other quantum-mechanical problems, including the Stark effect in a strong field [59, 60]. If the asymptotics contains not only the degrees of the expansion parameter but also the logarithms, then its onset is delayed and the calculation of even a very large number of the coefficients of the perturbation series does not allow the function represented by this series to be restored at infinity (see [60]). This requires going outside the scope of the perturbation theory and studying an exact solution.

APPENDIX C

The longitudinal (along the magnetic field) size of a hydrogen atom is characterized by the moments of the electron density distribution R_α. For $\mathcal{H} \gg 1$,

$$R_\alpha = \langle |z|^\alpha \rangle = (2\lambda)^{-\alpha} \int_0^\infty W^2_{\nu,\,1/2}(x) x^\alpha dx \Big/ \int_0^\infty W^2_{\nu,\,1/2}(x) dx, \tag{C.1}$$

where $\nu = 1/\lambda$ and we used (7.1). The integrals appears in this formula were calculated in [41], whence

$$R_\alpha = A(\alpha, \nu) \times \frac{{}_3F_2(-\nu, -\nu, \alpha+1; \alpha+2-\nu, \alpha+2-\nu; 1)}{{}_3F_2(-\nu, -\nu, 1; 2-\nu, 2-\nu; 1)} a_B^\alpha, \tag{C.2}$$

$$A(\alpha, \nu) = \Gamma(\alpha+1)\left[\frac{\Gamma(\alpha+2)\Gamma(2-\nu)}{\Gamma(\alpha+2-\nu)}\right]^2 \left(\frac{\nu}{2}\right)^\alpha, \tag{C.3}$$

where ${}_3F_2(\ldots)$ is the generalized hypergeometric series, with

$$A(0, \nu) = 1, \quad A(1, \nu) = 2\nu/(2-\nu)^2,$$

$$A(2, \nu) = 18\nu^2[(2-\nu)(3-\nu)]^{-2}, \ldots .$$

For a deep level,

$$R_\alpha = \Gamma(\alpha+1)\left(\frac{\nu}{2}\right)^\alpha \times \{1 + 2[\psi(\alpha+2) - \psi(2)]\nu + O(\nu^2)\}, \tag{C.4}$$

$$a_\parallel = \frac{1}{\sqrt{2}\lambda}\left[1 + \frac{5}{6\lambda} + O(\lambda^{-2})\right], \quad \lambda \gg 1, \tag{C.5}$$

where $\psi(x) = \Gamma'(x)/\Gamma(x)$. The first term of this expansion corresponds to the approximation of a δ function (A.12).

The calculation using formula (C.2) involves no difficulty, because the power series for ${}_3F_2$ converge as $\sum n^{-(\alpha+4)}$, irrespective of ν. As a result, we obtain the rms radii $a_\parallel = \sqrt{R_2} = \langle z^2 \rangle^{1/2}$ given in Table 4. Since $a_\parallel \gg a_H$ in a superstrong magnetic field, the size of the hydrogen atom is determined by $a_\parallel$. It should be noted that $a_\parallel$ is smaller by an order of magnitude than the Bohr radius; i.e., the size of the hydrogen atom along the field also significantly decreases.

When the magnetic field is switched off,

$$a_\parallel = n\sqrt{(5n^2+1)/6}\,a_B, \tag{C.6}$$

for ns levels [15]; in particular, $a_\parallel = a_B$ for the 1*s* level and $a_\parallel = 3.742 a_B$ for the 2*s* level.

APPENDIX D

A hydrogen atom in D dimensions is described by the Schrödinger equation

$$\Delta_D \psi + 2\left(E + \frac{1}{r}\right)\psi = 0, \quad r = \left(\sum_{i=1}^D x_i^2\right)^{1/2} \tag{D.1}$$

and has a discrete energy spectrum:

$$E_n = -\frac{1}{2[n + (D-3)/2]^2}, \quad n = 1, 2, 3, \ldots . \tag{D.2}$$

For $n > 1$, there is a random level degeneracy attributable to the hidden symmetry group $SO(D+1)$ of the Coulomb potential [47, 48, 61, 62]. However, the ground state is nondegenerate for any dimension of space D. Its energy, mean radius, and the wave function are

$$E_{1s} = -\frac{2}{(D-1)^2}, \quad \langle r \rangle = \frac{1}{4}(D^2 - D),$$

$$R_{1s}(r) = \frac{2^D}{[(D-1)^D \Gamma(D)]^{1/2}} \exp\left(-\frac{2r}{D-1}\right), \tag{D.3}$$

$$\int_0^\infty R_{1s}^2 r^{D-1} dr = 1,$$

as can be easily verified by the direct substitution into (D.1). For an arbitrary $\alpha > -D$, we have

$$\langle r^\alpha \rangle^{1/\alpha} = \frac{1}{4}(D-1)[\Gamma(D+\alpha)/\Gamma(D)]^{1/\alpha}, \tag{D.4}$$

for example, $\langle r^2 \rangle^{1/2} / \langle r \rangle = \sqrt{1 + D^{-1}}$, and we obtain $\Delta r / \langle r \rangle = D^{-1/2}$ for the rms deviation.

The energy $E_{1s} \longrightarrow -\infty$ and the mean radius $\langle r \rangle \longrightarrow 0$ when $D \longrightarrow 1$, indicative of the "fall to the center" known from quantum mechanics [15]. This case, described by the Hamiltonian

$$\hat{H} = -\frac{1}{2}\frac{d^2}{dz^2} - \frac{1}{|z|}, \quad -\infty < z < \infty, \tag{D.5}$$

was considered in detail in [21]. It was shown in [21] that Hamiltonian (D.5) is a Hermitian but self-adjoint operator (according to the general theory of singular potentials in quantum mechanics [24–26]). However, it admits a self-adjoint extension. From the physical point of view, introducing an additional condition[16] to Hamil-

---

[16] Necessary for the unambiguous determination of the energy spectrum.

tonian (D.5) corresponds to choosing a particular method for cutting off the potential $U(z) = -1/|z|$ at small distances (in our case, at $|z| \lesssim a_H$). The location of the lower level $E_{1s}(\mathcal{H})$ may be specified as such an additional condition; cf. the potential [25] $U(r) = -g^2/2r^2$ for $g > l + 1/2$.

As we see from (D.2), for a dimension $D > 1$, the energy spectrum is limited from below and there is no fall to the center.

REFERENCES

1. Ya. B. Zel'dovich, Fiz. Tverd. Tela (Leningrad) **1**, 1637 (1959) [Sov. Phys. Solid State **1**, 1497 (1959)]; Ya. B. Zel'dovich, *Selected Works. Particles, Nuclei, Universe* (Nauka, Moscow, 1985).
2. L. I. Schiff and H. Snyder, Phys. Rev. **55**, 59 (1939).
3. R. H. Garstang, Rep. Prog. Phys. **40**, 105 (1977).
4. M. A. Liberman and B. Johansson, Usp. Fiz. Nauk **165**, 121 (1995) [Phys.–Usp. **38**, 117 (1995)].
5. K. A. U. Lindgren and J. T. Virtamo, J. Phys. B **12**, 3465 (1979); J. T. Virtamo and K. A. U. Lindgren, Phys. Lett. A **71**, 329 (1979).
6. G. Fonte, P. Falsaperla, G. Schiffrer, and D. Stanzial, Phys. Rev. A **41**, 5807 (1990).
7. S. P. Goldman and Z. Chen, Phys. Rev. Lett. **67**, 1403 (1991).
8. Z. Chen and S. P. Goldman, Phys. Rev. A **45**, 1722 (1992); Phys. Rev. A **48**, 1107 (1993).
9. Jang-Haur Wang and Chen-Shiung Hsue, Phys. Rev. A **52**, 4508 (1995).
10. D. Lai and E. E. Salpeter, Phys. Rev. A **53**, 152 (1996).
11. Yu. P. Kravchenko, M. A. Liberman, and B. Johansson, Phys. Rev. A **54**, 287 (1996).
12. V. M. Weinberg, V. A. Gani, and A. E. Kudryavtsev, Zh. Éksp. Teor. Fiz. **113**, 550 (1998) [JETP **86**, 305 (1998)].
13. V. A. Gani, A. E. Kudryavtsev, V. A. Lensky, and V. M. Weinberg, Zh. Éksp. Teor. Fiz. **123**, 457 (2003) [JETP **96**, 402 (2003)].
14. A. Y. Potekhin and A. V. Turbiner, Phys. Rev. A **63**, 065402 (2001).
15. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory,* 5th ed. (Fizmatlit, Moscow, 2002; Pergamon, New York, 1977).
16. H. Hasegawa and R. E. Howard, J. Phys. Chem. Solids **21**, 179 (1961).
17. J. Schwinger, Phys. Rev. **82**, 664 (1951).
18. B. M. Karnakov, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 73 (2003) [JETP Lett. **77**, 68 (2003)].
19. V. S. Popov, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 79 (2003) [JETP Lett. **77**, 74 (2003)].
20. *Higher Transcendental Functions (Bateman Manuscript Project)*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1953; Nauka, Moscow, 1965), Vol. 1.
21. V. M. Galitskiĭ, B. M. Karnakov, and V. I. Kogan, *Problems on Quantum Mechanics* (Nauka, Moscow, 1992).
22. V. S. Popov, Zh. Éksp. Teor. Fiz. **60**, 1228 (1971) [Sov. Phys. JETP **33**, 665 (1971)].
23. I. Pomeranchuk and Ya. Smorodinsky, J. Phys. (Moscow) **9**, 97 (1945); I. Ya. Pomeranchuk, *Collected Scientific Works* (Nauka, Moscow, 1972), Vol. 2, p. 21.
24. K. M. Case, Phys. Rev. **80**, 797 (1950).
25. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953; Inostrannaya Literatura, Moscow, 1960), Vol. 2.
26. A. M. Perelomov and V. S. Popov, Teor. Mat. Fiz. **4**, 48 (1970).
27. V. S. Popov, Pis'ma Zh. Éksp. Teor. Fiz. **11**, 254 (1970) [JETP Lett. **11**, 162 (1970)]; Yad. Fiz. **12**, 429 (1970) [Sov. J. Nucl. Phys. **12**, 235 (1970)].
28. Ya. B. Zel'dovich and V. S. Popov, Usp. Fiz. Nauk **105**, 403 (1971) [Sov. Phys. Usp. **14**, 673 (1971)].
29. A. E. Kudryavtsev, V. E. Markushin, and I. S. Shapiro, Zh. Éksp. Teor. Fiz. **74**, 432 (1978) [Sov. Phys. JETP **47**, 225 (1978)].
30. H. van Haeringen, C. V. M. van der Mee, and R. van Wageningen, J. Math. Phys. **18**, 941 (1977).
31. H. van Haeringen and L. P. Kok, Phys. Rev. C **24**, 1827 (1981).
32. A. M. Badalyan, L. P. Kok, M. I. Polikarpov, and Yu. A. Simonov, Phys. Rep. **82**, 31 (1982).
33. L. P. Kok, J. W. de Maag, H. H. Brouwer, and H. van Haeringen, Phys. Rev. C **26**, 2381 (1982).
34. D. A. Kirzhnits and F. M. Pen'kov, Zh. Éksp. Teor. Fiz. **82**, 657 (1982) [Sov. Phys. JETP **55**, 393 (1982)].
35. A. E. Kudryavtsev and V. S. Popov, Pis'ma Zh. Éksp. Teor. Fiz. **29**, 311 (1979) [JETP Lett. **29**, 280 (1979)].
36. T. L. Trueman, Nucl. Phys. **26**, 57 (1961).
37. A. I. Nikishov and V. I. Ritus, Zh. Éksp. Teor. Fiz. **52**, 223 (1967) [Sov. Phys. JETP **25**, 145 (1967)].
38. B. M. Karnakov, V. D. Mur, and V. S. Popov, Dokl. Akad. Nauk SSSR **279**, 345 (1984) [Sov. Phys. Dokl. **29**, 938 (1984)].
39. B. M. Karnakov, A. E. Kudravtsev, V. D. Mur, and V. S. Popov, Nuovo Cimento B **98**, 63 (1987).
40. B. M. Karnakov, A. E. Kudryavtsev, V. D. Mur, and V. S. Popov, Zh. Éksp. Teor. Fiz. **94**, 65 (1988) [Sov. Phys. JETP **67**, 1333 (1988)].
41. V. S. Popov, A. E. Kudryavtsev, and V. D. Mur, Zh. Éksp. Teor. Fiz. **77**, 1727 (1979) [Sov. Phys. JETP **50**, 865 (1979)]; Zh. Éksp. Teor. Fiz. **80**, 1271 (1981) [Sov. Phys. JETP **53**, 650 (1981)].
42. M. Izycki *et al.*, Z. Phys. A **297**, 1 (1980).
43. V. S. Popov, in *Proceedings of International Conference on High Energy Nuclear Physics, Hungary* (1987), p. 257; Preprint No. 136, ITEP (Inst. of Theoretical and Experimental Physics, Moscow, 1987).
44. A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions, and Decays in Nonrelativistic Quantum Mechanics*, 2nd ed. (Nauka, Moscow, 1971; Israel Program for Scientific Translations, Jerusalem, 1966).

45. D. R. Hartree, Proc. Cambridge Philos. Soc. **24**, 89 (1927); *The Calculation of Atomic Structures* (Wiley, New York, 1957; Inostrannaya Literatura, Moscow, 1960).
46. R. Loudon, Am. J. Phys. **27**, 649 (1959).
47. V. A. Fock, Z. Phys. **98**, 145 (1935).
48. V. Bargmann, Z. Phys. **99**, 576 (1935).
49. B. B. Kadomtsev, Zh. Éksp. Teor. Fiz. **58**, 1765 (1970) [Sov. Phys. JETP **31**, 945 (1970)].
50. Yu. E. Lozovik and A. V. Klyuchnik, Phys. Lett. A **66**, 282 (1978).
51. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Relativistic Quantum Theory* (Nauka, Moscow, 1968; Pergamon Press, Oxford, 1971), Vol. 1.
52. O. Dumbarais, R. Kock, J. J. de Swart, and P. Kroll, Nucl. Phys. B **216**, 277 (1983).
53. H. A. Bethe, Phys. Rev. **76**, 38 (1949).
54. M. L. Goldberger and K. M. Watson, *Collision Theory* (Wiley, New York, 1964; Mir, Moscow, 1967).
55. V. D. Mur, A. E. Kudryavtsev, and V. S. Popov, Yad. Fiz. **37**, 1417 (1983) [Sov. J. Nucl. Phys. **37**, 844 (1983)].
56. R. J. Elliott and R. Loudon, J. Phys. Chem. Solids **15**, 196 (1960).
57. D. Grasso and H. Rubinstein, Phys. Rep. **348**, 163 (2001).
58. I. S. Gradshteĭn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 4th ed. (Fizmatgiz, Moscow, 1962; Academic, New York, 1980).
59. L. Benassi, V. Grecchi, E. Harrell, and B. Simon, Phys. Rev. Lett. **42**, 704 (1979); Phys. Rev. Lett. **42**, 1430 (1979).
60. D. I. Kazakov and V. S. Popov, Zh. Éksp. Teor. Fiz. **122**, 675 (2002) [JETP **95**, 581 (2002)]; Pis'ma Zh. Éksp. Teor. Fiz. **77**, 547 (2003) [JETP Lett. **77**, 453 (2003)].
61. S. P. Alliluev, Zh. Éksp. Teor. Fiz. **33**, 200 (1957) [Sov. Phys. JETP **6**, 156 (1957)].
62. G. Dyerdi and J. Revai, Zh. Éksp. Teor. Fiz. **48**, 1445 (1965) [Sov. Phys. JETP **21**, 967 (1965)].

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 915–921.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 1023–1029.
Original Russian Text Copyright © 2003 by Matveev.

ATOMS, SPECTRA, RADIATION

Emission and Electron Transitions in an Atom Interacting with an Ultrashort Electromagnetic Pulse

V. I. Matveev
Lomonosov Pomorskiĭ State University, Arkhangel'sk State Technical University, Arkhangel'sk, 163006 Russia
e-mail: matveev.victor@pomorsu.ru

Received May 27, 2002; in final form, May 5, 2003

Abstract—Electron transitions and emission of an atom interacting with a spatially inhomogeneous ultrashort electromagnetic pulse are considered. The excitation and ionization probabilities are obtained as well as the spectra and cross sections of the reemission of such a pulse by atoms. By way of an example, one- and two-electron inelastic processes accompanying the interaction of ultrashort pulses with hydrogen- and helium-like atoms are considered. The developed technique makes it possible to take into account exactly the spatial nonuniformity of the ultrashort pulse field and photon momenta in the course of reemission. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The interaction of atoms with electromagnetic pulses (e.g., laser radiation) is usually studied for pulses with a duration much longer than the characteristic times of the target atom. The interaction of long pulses with atoms gives rise to an exceptionally rich pattern for theoretical and experimental investigations. Such investigations are described in a large number of reviews and monographs. Contemporary theoretical investigations correspond in many respects to the experimental tendency of designing more powerful lasers and generation of ultrashort pulses. Advances in laser technology provided sources of laser pulses of 20–30 fs in duration with a peak intensity up to 10^{21} W/cm^2 [1, 2]. The generation of 4-fs pulses was reported in [3]. Electromagnetic pulses of 0.25 fs in duration have been observed in recent experiments [4]. Experimental requirements and the complication of the physical situation stimulated the further development [5, 6] of the Keldysh theory; the modification and evolution of the Coulomb–Born approximation [7–11], which made it possible to simultaneously take into account a strong ac field as well as the Coulomb field of the atomic core; the development and construction of relativistic theories of tunnel and multiphoton ionization [12, 13]; and the development of numerical approaches for calculating the probabilities of transitions of atomic electrons in strong fields (see, for example, [14, 15] and the references cited therein). At the same time, the methods for generating ultrashort subfemtosecond pulses and even shorter pulses 10^{-21}–10^{-22} s [19] in duration are being actively discussed in the literature (see, for example, [16–19]). This may open new prospects for studying the interaction of ultrashort electromagnetic pulses with matter. In particular, it becomes possible to analyze the processes accompanying the interaction of atoms with strong electromagnetic pulses of a duration smaller than the characteristic atomic periods. Let us estimate the value of the Keldysh parameter,

$$\gamma = \frac{(2\mathcal{G}m)^{1/2}\omega}{eE},$$

where $\mathcal{G}$ is the atomic ionization potential, m and e are the electron mass and charge, and E is the strength of the external field of frequency ω. The atomic time unit is equal approximately to 2.42×10^{-17} s. For a pulse of $(1/3) \times 10^{-17}$ s, the photon energy is approximately equal to 1.24 keV. Let us suppose that the intensity of incident radiation is on the order of 10^{21} W/cm^2; in this case, $\gamma \approx 0.4$ for the hydrogen atom. Such a value of the Keldysh parameter indicates the inapplicability of perturbation theory for describing the interaction of atoms with ultrashort pulses of a strong electromagnetic field.

An additional possibility [20] (see also [21]) of studying the interaction of atoms with ultrashort electromagnetic pulses and a direct experimental confirmation of processes in question can be achieved using collision experiments. For example, in experiments described in [20], double and single ionization of helium atoms by an impact of the uranium ion U^{92+} with an energy of 1 GeV/nucleon was studied and a superintense pulse ($I > 10^{19}$ W/cm^2) of duration 10^{-18} s was simulated. The obtaining of such parameters of the electromagnetic pulse by other methods is extremely complicated. It is well known [22] that the fields produced by relativistic and ultrarelativistic charged particles are close to the field of a light wave. This circumstance makes it possible to use the so-called equivalent photon method [22, 23] based on the replacement of virtual photons by real light field quanta. For fields created by fairly large charges ($Z > 72$), perturbation theory is inapplicable [24] even for infinitely large collision energies; this necessitates the application of nonperturbative methods for describing the processes occurring in such fields. One of the most advanced approaches is as follows. It is well known [25] that the field of a charge moving uniformly along a straight line with an ultrarelativistic velocity is concentrated in the plane perpendicular to the direction of its motion and passes through the point at which the charge is located at a given instant. However, this circumstance is manifested explicitly [26] only after the implementation of the singular gauge transformation proposed earlier [27] (see also [28]), when the field potentials can be written [26] in the form of a function proportional to the Dirac delta function concentrated in the above-mentioned plane. The latter circumstance suggests an instantaneous action of such a field on an atom and makes it possible to solve exactly the Dirac equation for atomic electrons in the ultrarelativistic limit [29]. In addition, the effective strengths of the fields produced by high-charge ions may attain values of 10^{11} V/cm (cf. the characteristic atomic electric field strength, which is on the order of 5×10^9 V/cm). An additional considerable enhancement of the field [26] also occurs due to relativistic compression.

It should be noted that it is often difficult to take into account the interaction of atoms with strong electromagnetic pulses with a duration exceeding the characteristic atomic time periods by using nonperturbative methods; in this case, numerical methods are required. By way of example, we mention publication [30], in which the excitation and ionization of helium atoms by short pulses of a strong electromagnetic field of 3.8–15.2 fs are considered (see also [14, 15, 31, 32] and the references cited therein). The necessity of taking into account the spatial inhomogeneity of the electromagnetic field pulse (over the size of the target atom), which extends beyond the dipole approximation limits, involves additional difficulties. The number of publications in this direction (see, for example, [33–36] and the literature cited therein), in which only the first correction to the dipole approximation was included, is scarce.

In many practically important cases, perturbations are not small enough for the application of perturbation theory. However, the time of action of perturbations in many cases is much shorter than the characteristic atomic times; this makes it possible to solve the problem without limiting the intensity of perturbations and to carry out calculations analytically [29, 37–44]. In the cases considered here, the characteristic atomic time τ_a is assumed to be much larger than the duration τ of ultrashort pulses. For this reason, the approximation of sudden perturbations [37], which does not limit the perturbation intensity and requires only the fulfillment of the inequality $\tau/\tau_a \ll 1$ for its application, may serve as the basis for the solution.

In this study, we consider the excitation and ionization of light (nonrelativistic) atoms during the interaction with a spatially inhomogeneous ultrashort pulse of the electromagnetic field and calculate the probabilities of excitation and ionization as well as the spectra and cross sections of reemission of such a pulse by an atom. The developed method makes it possible to take into account exactly the spatial inhomogeneity of the field generated by an ultrashort pulse as well as the momenta of photons during reemission.

2. GENERAL

The potential of electromagnetic waves (the vector potential $\mathbf{A}$ and the scalar potential φ) are often chosen in such a way that the scalar potential is equal to zero. In such a calibration, the potential of interaction of an electron with the external electromagnetic field has the form (here and below, atomic units are used)

$$V(\mathbf{r}, t) = -\frac{1}{c}(\hat{\mathbf{p}} \cdot \mathbf{A} + \mathbf{A} \cdot \hat{\mathbf{p}}) + \frac{1}{c^2}\mathbf{A}^2, \qquad (1)$$

where $\hat{\mathbf{p}}$ is the electron momentum operator and $c = 137$ atomic units is the velocity of light. We assume that the vector potential of the wave field is a function of coordinate $\mathbf{r}$ and time t, $\mathbf{A}(\mathbf{r}, t) = \mathbf{A}(\eta)$, where the wave phase $\eta = \omega_0 t - \mathbf{k}_0 \cdot \mathbf{r}$. Here, the wave vector $\mathbf{k}_0$ is such that $|\mathbf{k}_0| = \omega_0/c$, ω_0 being the cyclic frequency. We carry out the gauge transformation

$$\mathbf{A}' = \mathbf{A} + \nabla f, \quad \varphi' = \varphi - \frac{1}{c}\frac{\partial f}{\partial t},$$

where $f = \mathbf{A} \cdot \mathbf{r}$. This gives

$$\mathbf{A}' = -\mathbf{k}_0\left(\mathbf{r} \cdot \frac{d\mathbf{A}}{d\eta}\right), \quad \varphi' = -\mathbf{E} \cdot \mathbf{r},$$

where

$$\mathbf{E} = \mathbf{E}(\mathbf{r}, t) = -|\mathbf{k}_0|\frac{d\mathbf{A}}{d\eta}.$$

In such a calibration, the potential $V(\mathbf{r}, t)$ of interaction of an electron with the electromagnetic field assumes the form

$$V(\mathbf{r}, t) = -\frac{1}{c}(\hat{\mathbf{p}} \cdot \mathbf{A}' + \mathbf{A}' \cdot \hat{\mathbf{p}}) + \frac{1}{c^2}(\mathbf{A}')^2 + \mathbf{E} \cdot \mathbf{r}. \qquad (2)$$

We assume that the estimates $p \sim 1$ and $r \sim 1$ (which obviously hold for an electron in the hydrogen atom or in atoms with small nuclear charges on the order of unity) are valid for a nonrelativistic electron. In this case, in expression (2) we can disregard the first two terms as compared to the third term, and the potential of interaction of the electron with the electromagnetic field assumes the simple form

$$V(\mathbf{r}, t) = \mathbf{r} \cdot \mathbf{E}(\mathbf{r}, t). \tag{3}$$

The potential of interaction between atomic electrons and an electromagnetic pulse of the Gaussian form,

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 \exp\left[-\alpha^2\left(t - \frac{\mathbf{k}_0 \cdot \mathbf{r}}{\omega_0}\right)^2\right]\cos(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r}), \tag{4}$$

can be written in the form

$$V(t) \equiv V(\{\mathbf{r}_a\}, t) = \sum_{a=1}^{a=N} \mathbf{E}(\mathbf{r}_a, t) \cdot \mathbf{r}_a, \tag{5}$$

where $\{\mathbf{r}_a\}$ is the set of coordinates of atomic electrons ($a = 1, \ldots, N$), N being the number of atomic electrons. Let α in formula (4) assume such values that $V(t)$ effectively differs from zero only during a time interval $\tau \sim \alpha^{-1}$, which is much smaller than the characteristic periods of an unperturbed atom, described by Hamiltonian H_0. Then the amplitude of the transition of the atom from the initial state φ_0 to a final state φ_n as a result of action of a sudden perturbation $V(t)$ has the form [38]

$$a_{0n} = \langle\varphi_n|\exp\left(-i\int_{-\infty}^{\infty} V(t)dt\right)|\varphi_0\rangle, \tag{6}$$

where φ_0 and φ_n belong to the complete orthonormal system of eigenfunctions of the unperturbed Hamiltonian H_0. In accordance with relation (6). the choice of the perturbation in the form (5) makes it possible to express probabilities $w_{0n} = |a_{0n}|^2$ in terms of the well-known inelastic atomic form factors [45, 46],

$$w_{0n} = \left|\langle\varphi_n|\exp\left(-i\mathbf{q} \cdot \sum_a \mathbf{r}_a\right)|\varphi_0\rangle\right|^2, \tag{7}$$

where

$$\mathbf{q} = \int_{-\infty}^{\infty} dt\mathbf{E}(\mathbf{r}, t) = \frac{\sqrt{\pi}}{\alpha}\mathbf{E}_0 \exp\left(-\frac{\omega_0^2}{4\alpha^2}\right). \tag{8}$$

These formulas allow us to easily calculate the probabilities w_{0n} of excitation or ionization of an atom. Thus, the probabilities of atomic excitation or ionization by a spatially inhomogeneous pulse are formally the same as the probabilities of excitation and ionization by a spatially homogeneous electromagnetic pulse (formula (4) with the formal equality $\mathbf{k}_0 \cdot \mathbf{r} = 0$). The spatially inhomogeneous and homogeneous cases lead to different results for the probabilities of reemission of a pulse incident on an atom.

3. REEMISSION OF AN ULTRASHORT PULSE BY AN ATOM

In the approximation of sudden perturbations, the evolution of the initial state φ_0 has the form

$$\Psi_0(t) = \exp\left[-i\int_{-\infty}^{t} V(t')dt'\right]\varphi_0, \tag{9}$$

where $\Psi_0(t) \longrightarrow \varphi_0$ as $t \longrightarrow -\infty$. We introduce the complete and orthonormal system of functions

$$\Phi_n(t) = \exp\left[i\int_{t}^{\infty} V(t')dt'\right]\varphi_n, \tag{10}$$

where $\Phi_n(t) \longrightarrow \varphi_n$ as $t \longrightarrow \infty$. Obviously, amplitude (6) can be written in the form

$$a_{0n} = \langle\Phi_n(t)|\Psi_0(t)\rangle. \tag{11}$$

For this reason, we will calculate the amplitude of the emission of a photon in the first order of perturbation theory as the corrections to states (9) and (10) in the interaction of atomic electrons with the electromagnetic field [22]:[1]

$$\begin{gathered} U = -\sum_{a, \mathbf{k}, \sigma}\left(\frac{2\pi}{\omega}\right)^{1/2} \\ \times \mathbf{u}_{\mathbf{k}\sigma}[\overset{+}{a}_{\mathbf{k}\sigma}\exp(-i\mathbf{k} \cdot \mathbf{r}_a) + a_{\mathbf{k}\sigma}\exp(i\mathbf{k} \cdot \mathbf{r}_a)]\hat{\mathbf{p}}_a. \end{gathered} \tag{12}$$

Here, $\overset{+}{a}_{\mathbf{k}\sigma}$ and $a_{\mathbf{k}\sigma}$ are the creation and annihilation operators for a photon with frequency ω, momentum $\mathbf{k}$, and polarization σ ($\sigma = 1, 2$); $\mathbf{u}_{\mathbf{k}\sigma}$ are unit vectors of polarization; $\mathbf{r}_a$ are the coordinates of atomic electrons ($a = 1, \ldots, N$); and $\hat{\mathbf{p}}_a$ are the momentum operators for atomic electrons. In this case, the amplitude of photon emission associated with a transition of the atom from state φ_0 to state φ_n has the form

$$\begin{gathered} b_{0n}(\omega) = i\left(\frac{2\pi}{\omega}\right)^{1/2}\mathbf{u}_{\mathbf{k}\sigma}\int_{-\infty}^{\infty} dt\exp(i\omega t)\langle\Phi_n(t)| \\ \times \sum_a \exp(-i\mathbf{k} \cdot \mathbf{r}_a)\hat{\mathbf{p}}_a|\Psi_0(t)\rangle. \end{gathered} \tag{13}$$

Integrating by parts with respect to time and omitting the terms vanishing upon the elimination of interaction

---

[1] Sudden perturbation $V(t)$ is taken into account in functions $\Phi_n(t)$ and $\Psi_0(t)$ without imposing limitations on the value of $V(t)$.

with the electromagnetic field (for $t \longrightarrow \pm\infty$), we obtain

$$b_{0n}(\omega) = -\left(\frac{2\pi}{\omega}\right)^{1/2}\mathbf{u}_{\mathbf{k}\sigma}\int_{-\infty}^{\infty}dt\frac{\exp(i\omega t)}{i\omega}\langle\varphi_n|$$
$$\times\sum_a \exp(-i\mathbf{k}\cdot\mathbf{r}_a)\frac{\partial V(t)}{\partial\mathbf{r}_a}\exp\left[-i\int_{-\infty}^{\infty}V(t')dt'\right]|\varphi_0\rangle. \tag{14}$$

Representing the element of integration with respect to the photon momentum in the form

$$(2\pi)^{-3}d\mathbf{k} = (c2\pi)^{-3}d\Omega_{\mathbf{k}}\omega^2 d\omega$$

and carrying out the summation of $|b_{0n}(\omega)|^2$ over polarizations, we obtain the corresponding spectrum for the photon emission into the unit solid angle $d\Omega_{\mathbf{k}}$, accompanied by the atomic transition from state φ_0 into state φ_n:

$$\frac{d^2W_{0n}}{d\Omega_{\mathbf{k}}d\omega} = \frac{1}{(2\pi)^2c^3\omega}\left|\langle\varphi_n|\sum_a\exp(-i\mathbf{k}\cdot\mathbf{r}_a)\right.$$
$$\left.\times\left[\frac{\partial\tilde{V}(\omega)}{\partial\mathbf{r}_a}\times\mathbf{n}\right]\exp\left[-i\int_{-\infty}^{\infty}V(t')dt'\right]|\varphi_0\rangle\right|^2. \tag{15}$$

Here, $\tilde{V}(\omega)$ is the Fourier transform of function $V(t)$, defined in accordance with relation (5),

$$\tilde{V}(\omega) = \int_{-\infty}^{\infty}V(t)\exp(i\omega t)dt$$
$$= \sum_{a=1}^{N}\mathbf{E}_0\cdot\mathbf{r}_a\exp\left(i\frac{\omega}{\omega_0}\mathbf{k}_0\cdot\mathbf{r}_a\right)f_0(\omega), \tag{16}$$

$$f_0(\omega) = \frac{\sqrt{\pi}}{2\alpha}\left\{\exp\left[-\frac{(\omega-\omega_0)^2}{4\alpha^2}\right]+\exp\left[-\frac{(\omega+\omega_0)^2}{4\alpha^2}\right]\right\}, \tag{17}$$

and the vector product is given by

$$\frac{\partial\tilde{V}(\omega)}{\partial\mathbf{r}_a}\times\mathbf{n} = f_0(\omega)\exp\left(-i\frac{\omega}{\omega_0}\mathbf{k}_0\cdot\mathbf{r}_a\right)$$
$$\times\left(\mathbf{E}_0\times\mathbf{n}+i\frac{\omega}{\omega_0}(\mathbf{E}_0\cdot\mathbf{r}_a)[\mathbf{k}_0\times\mathbf{n}]\right). \tag{18}$$

Formula (15) describes the spectrum of photon emission accompanied by the atomic transition from state φ_0 to state φ_n, i.e., the partial spectrum. Carrying out summation over all final states φ_n of the atom in formula (15), we obtain the total emission spectrum

$$\frac{d^2W}{d\Omega_{\mathbf{k}}d\omega} = \frac{1}{(2\pi)^2}\frac{1}{c^3\omega}\langle\varphi_0|\sum_{a,a'}\exp[-i\mathbf{k}\cdot(\mathbf{r}_a-\mathbf{r}_{a'})]$$
$$\times\left[\frac{\partial\tilde{V}(\omega)}{\partial\mathbf{r}_a}\times\mathbf{n}\right]\cdot\left[\frac{\partial\tilde{V}^*(\omega)}{\partial\mathbf{r}_{a'}}\times\mathbf{n}\right]|\varphi_0\rangle. \tag{19}$$

Thus, we have obtained the total atomic emission spectrum over the time of action of the sudden perturbation $V(t)$.

In the case of a one-electron hydrogen-like atom, formula (19) is simplified and assumes the form

$$\frac{d^2W}{d\Omega_{\mathbf{k}}d\omega}$$
$$= \frac{1}{(2\pi)^2}\frac{1}{c^3\omega}\langle\varphi_0|\left[\frac{\partial\tilde{V}(\omega)}{\partial\mathbf{r}}\times\mathbf{n}\right]\cdot\left[\frac{\partial\tilde{V}^*(\omega)}{\partial\mathbf{r}}\times\mathbf{n}\right]|\varphi_0\rangle. \tag{20}$$

Integrating this formula with respect to the photon emission angles $d\Omega_{\mathbf{k}}$, we obtain

$$\frac{dW}{d\omega} = \frac{2}{3\pi}\frac{1}{c^3\omega}\langle\varphi_0|\frac{\partial\tilde{V}(\omega)}{\partial\mathbf{r}}\cdot\frac{\partial\tilde{V}^*(\omega)}{\partial\mathbf{r}}|\varphi_0\rangle. \tag{21}$$

The average over the ground state of the hydrogen-like atom can be evaluated easily. As a result, the total emission spectrum of the hydrogen-like atom with nuclear charge Z has the form

$$\frac{dW}{d\omega} = \frac{2}{3\pi}\frac{1}{c^3\omega}|f_0(\omega)|^2\mathbf{E}_0^2\left(1+\frac{\omega^2}{Z^2c^2}\right). \tag{22}$$

Since this spectrum is proportional to $|f_0(\omega)|^2$, the atom, in accordance with formula (17), mainly emits photons belonging to the continuous spectrum with characteristic frequencies of $|\omega-\omega_0| \le 1/\tau$.

In order to calculate the total emission spectrum of a two-electron helium-like atom, we consider separately the terms with $a = a'$ and with $a \ne a'$ in formula (19) and, accordingly, write the spectrum in the form

$$\frac{d^2W}{d\Omega_{\mathbf{k}}d\omega} = \frac{d^2W_1}{d\Omega_{\mathbf{k}}d\omega}+\frac{d^2W_2}{d\Omega_{\mathbf{k}}d\omega}, \tag{23}$$

where

$$\frac{d^2W_1}{d\Omega_{\mathbf{k}}d\omega} = \frac{1}{(2\pi)^2}\frac{1}{c^3\omega} \times \langle\varphi_0|\sum_a\left[\frac{\partial\tilde{V}(\omega)}{\partial\mathbf{r}_a}\times\mathbf{n}\right]\cdot\left[\frac{\partial\tilde{V}^*(\omega)}{\partial\mathbf{r}_a}\times\mathbf{n}\right]|\varphi_0\rangle, \tag{24}$$

$$\frac{d^2W_2}{d\Omega_{\mathbf{k}}d\omega} = \frac{1}{(2\pi)^2}\frac{1}{c^3\omega} \times \langle\varphi_0|\sum_{a,\,a'(a\neq a')}\exp[-i\mathbf{k}\cdot(\mathbf{r}_a-\mathbf{r}_{a'})] \times\left[\frac{\partial\tilde{V}(\omega)}{\partial\mathbf{r}_a}\times\mathbf{n}\right]\cdot\left[\frac{\partial\tilde{V}^*(\omega)}{\partial\mathbf{r}_{a'}}\times\mathbf{n}\right]|\varphi_0\rangle. \tag{25}$$

In formula (24), we can integrate with respect to the photon emission angles in the general form; this gives

$$\frac{dW_1}{d\omega} = \frac{2}{3\pi}\frac{1}{c^3\omega}\langle\varphi_0|\sum_a\frac{\partial\tilde{V}(\omega)}{\partial\mathbf{r}_a}\cdot\frac{\partial\tilde{V}^*(\omega)}{\partial\mathbf{r}_a}|\varphi_0\rangle. \tag{26}$$

We can calculate the average over the ground state of the helium-like atom appearing in this formula by describing the wave function of the ground state in the form of the product of one-electron hydrogen-like wave functions with the effective charge Z. This gives

$$\frac{dW_1}{d\omega} = 2\frac{2}{3\pi}\frac{1}{c^3\omega}|f_0(\omega)|^2\mathbf{E}_0^2\left(1+\frac{\omega^2}{Z^2c^2}\right). \tag{27}$$

Comparing this expression with formula (22), we arrive at the conclusion that this part of the emission spectrum of the two-electron atom corresponds to incoherent emission by two electrons, since formula (27) can be derived from Eq. (22) by multiplying it by the number of terms in formula (26), which in our case is equal to the number of emitting electrons, $N = 2$. The part of the spectrum represented by formula (25) contains $N(N-1) = N^2 - N$ terms and corresponds to the mixed (coherent + incoherent) form of emission.

It is difficult to integrate formula (25) with respect to photon emission angles in the general form; for this reason, using formula (18), we can write this expression in the form

$$\frac{d^2W_2}{d\Omega_{\mathbf{k}}d\omega} = \frac{1}{(2\pi)^2}\frac{1}{c^3\omega}|f_0(\omega)|^2 \times\langle\varphi_0|\sum_{a,\,a'(a\neq a')}\exp[-i\mathbf{k}\cdot(\mathbf{r}_a-\mathbf{r}_{a'})]\Big\{[\mathbf{E}_0\times\mathbf{n}]^2 + i\frac{\omega}{\omega_0}([\mathbf{E}_0\times\mathbf{n}]\cdot[\mathbf{k}_0\times\mathbf{n}])\mathbf{E}_0\cdot(\mathbf{r}_a-\mathbf{r}_{a'}) + (\mathbf{E}_0\cdot\mathbf{r}_a)(\mathbf{E}_0\cdot\mathbf{r}_{a'})\frac{\omega^2}{\omega_0^2}[\mathbf{k}_0\times\mathbf{n}]^2\Big\}|\varphi_0\rangle. \tag{28}$$

The average over the ground state of the helium atom appearing in this formula can be evaluated easily, which gives

$$\frac{d^2W_2}{d\Omega_{\mathbf{k}}d\omega} = \frac{2}{(2\pi)^2}\frac{1}{c^3\omega}|f_0(\omega)|^2 \times\left[\frac{16Z^4}{4Z^2+(\mathbf{k}-\mathbf{k}_0\omega/\omega_0)^2}\right]^2\Bigg\{[\mathbf{E}_0\times\mathbf{n}]^2 -\frac{\omega}{\omega_0}[\mathbf{E}_0\times\mathbf{n}]\cdot[\mathbf{k}_0\times\mathbf{n}]\frac{8\mathbf{E}_0\cdot(\mathbf{k}-\mathbf{k}_0\omega/\omega_0)}{4Z^2+(\mathbf{k}-\mathbf{k}_0\omega/\omega_0)^2} +\left[\frac{4\mathbf{E}_0\cdot(\mathbf{k}-\mathbf{k}_0\omega/\omega_0)}{4Z^2+(\mathbf{k}-\mathbf{k}_0\omega/\omega_0)^2}\frac{\omega}{\omega_0}\mathbf{k}_0\times\mathbf{n}\right]^2\Bigg\}. \tag{29}$$

We can now integrate with respect to the photon emission angle, directing the z axis along vector $\mathbf{k}_0$; this gives

$$\frac{dW_2}{d\omega} = \frac{1}{2\pi}\frac{1}{c^3\omega}|f_0(\omega)|^2 \times\mathbf{E}_0^2\int_{-1}^{1}dx\left[\frac{16Z^4}{(4Z^2+2\omega^2c^{-2}(1+x))^2}\right]^2 \times\left[1+x^2+\frac{8\omega^2c^{-2}(1-x^2)x}{4Z^2+2\omega^2c^{-2}(1+x)} +\frac{16\omega^4c^{-4}(1-x^2)^2}{(4Z^2+2\omega^2c^{-2}(1+x))^2}\right]. \tag{30}$$

We denote by $I(\alpha)$ the integral with respect to x appearing on the right-hand side of this formula, where $\alpha = \omega^2/2c^2Z^2$. The evaluation of this integral is elementary, but cumbersome; as a result, we obtain

$$I(\alpha) = \frac{2-2\alpha+8\alpha^2}{15\alpha^3} -\frac{2+18\alpha+68\alpha^2+120\alpha^3+80\alpha^4}{15\alpha^3(1+2\alpha)^5}. \tag{31}$$

In accordance with relation (23), the total spectrum can be represented in the form

$$\frac{dW}{d\omega} = \frac{dW_1}{d\omega}+\frac{dW_2}{d\omega}, \tag{32}$$

where $dW_1/d\omega$ can be expressed with the help of formula (27), and

$$\frac{dW_2}{d\omega} = \frac{1}{2\pi}\frac{1}{c^3\omega}|f_0(\omega)|^2\mathbf{E}_0^2 I(\alpha). \tag{33}$$

4. CONCLUSIONS

Thus, we have calculated the total spectra of reemission of an ultrashort electromagnetic pulse by hydrogen-like and helium-like atoms. We managed to take into account exactly both the spatial inhomogeneity of the pulse field over the atomic size and the momentum of the photons being emitted. Using expressions (27) and (33) and introducing the number N of electrons in the atom, we can write formula (32) for the total spectrum $dW/d\omega$ in the form

$$\begin{aligned}\frac{dW}{d\omega} &= \frac{2}{3\pi}\frac{1}{c^3\omega}|f_0(\omega)|^2 \\ &\times \mathbf{E}_0^2\left[N\left(1+\frac{\omega^2}{c^2Z^2}\right)+N(N-1)\frac{3}{8}I(\alpha)\right].\end{aligned} \tag{34}$$

Such a notation for the spectrum appears as convenient since formula (34) coincides with formula (22) for $N = 1$ and, hence, describes the spectrum of a hydrogen-like atom with the effective nuclear charge Z and the emission spectrum of a helium-like atom for $N = 2$. For atoms with an arbitrary number of electrons ($N \geq 2$), formula (34), in accordance with the line of reasoning following formula (27), may form the basis for a qualitative description and for estimating the dependence of the spectrum of reemission of a spatially inhomogeneous pulse on the number of atomic electrons. The spectrum (34) of reemission of a spatially inhomogeneous pulse consists of the incoherent (proportional to N) and coherent (proportional to N^2) parts. Since $I(\alpha) \longrightarrow 8/3$ for $\omega \longrightarrow 0$, the reemission spectrum is proportional to N^2 in the low-frequency range (when $\omega^2/Z^2c^2 \ll 1$) and is coherent by nature. Obviously, the case of low frequencies corresponds to the coherent nature of the reemission of a spatially homogeneous pulse by many-electron atoms, while in the high-frequency range (when $\omega^2/Z^2c^2 \gg 1$ and $I(\alpha) \longrightarrow 0$), the spectrum is proportional to N and is incoherent by nature.

In accordance with [25], in order to obtain the cross section of pulse reemission, the spectra defined by formulas (19) and (34) must be multiplied by ω and divided by the energy flux I expressed in terms of the integral of the absolute value of the Poynting vector $S(t) = c(4\pi)^{-1}\mathbf{E}^2$ with respect to time,

$$I = \int_{-\infty}^{\infty} dt S(t) = \frac{c}{4\pi}\mathbf{E}_0^2\frac{\sqrt{\pi}}{2\sqrt{2}\alpha}\left[\exp\left(-\frac{\omega_0^2}{2\alpha^2}\right)+1\right]. \tag{35}$$

ACKNOWLEDGMENTS

This study was supported financially by the Ministry of Education of the Russian Federation (grant no. E02-3.2-512) and by the Russian Foundation for Basic Research (project no. 01-02-17047).

REFERENCES

1. G. A. Mourou, Ch. P. J. Barty, and M. D. Perry, Phys. Today **51**, 22 (1998).
2. T. Brabec and F. Krausz, Rev. Mod. Phys. **72**, 545 (2000).
3. N. Zhavoronkov and G. Korn, Phys. Rev. Lett. **88**, 203901 (2002).
4. E. Hertz, N. A. Papadogiannis, G. Nersisyan, *et al.*, Phys. Rev. A **64**, 051801 (2001).
5. V. I. Usachenko and V. A. Pazdzersky, J. Phys. B **35**, 761 (2002).
6. K. Mishima, M. Hayashi, J. Yi, *et al.*, Phys. Rev. A **66**, 033401 (2002).
7. S. Basile, F. Trombetta, G. Ferrante, *et al.*, Phys. Rev. A **37**, 1050 (1988).
8. C. Leone, S. Bivona, R. Burlon, *et al.*, Phys. Rev. A **40**, 1828 (1989).
9. L. Rosenberg and F. Zhou, Phys. Rev. A **46**, 7093 (1992).
10. H. S. Reiss and V. P. Krainov, Phys. Rev. A **50**, R910 (1994).
11. G. Duchateau, E. Cormier, and R. Gayet, Phys. Rev. A **66**, 023412 (2002).
12. N. Milosevic, V. P. Krainov, and T. Brabec, J. Phys. B **35**, 3515 (2002).
13. H. K. Avetissian, A. G. Markossian, and G. F. Mkrtchian, Phys. Rev. A **64**, 053404 (2001).
14. A. D. Kondorskiy and L. P. Presnyakov, J. Phys. B **34**, L663 (2001).
15. J. B. West, J. Phys. B **34**, R45 (2001).
16. S. E. Harris and A. V. Sokolov, Phys. Rev. Lett. **81**, 2894 (1998).
17. I. P. Christov, M. M. Murnane, and H. C. Kapteyn, Opt. Commun. **148**, 75 (1998).
18. A. V. Sokolov, D. D. Yavuz, and S. E. Harris, Opt. Lett. **24**, 557 (1999).
19. A. E. Kaplan and P. L. Shkolnikov, Phys. Rev. Lett. **88**, 074801 (2002).
20. R. Moshammer, W. Schmitt, J. Ullrich, *et al.*, Phys. Rev. Lett. **79**, 3621 (1997).
21. A. V. Selin, A. M. Ermolaev, and C. J. Joachain, Phys. Rev. A **67**, 012709 (2003).
22. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 3rd ed. (Nauka, Moscow, 1989; Pergamon Press, Oxford, 1982).
23. C. A. Bertulani and G. Baur, Phys. Rep. **163**, 209 (1998).
24. J. Eichler and W. E. Meyrhof, *Relativistic Atomic Collisions* (Academic, New York, 1995).
25. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 7th ed. (Nauka, Moscow, 1988; Pergamon Press, Oxford, 1975).
26. A. J. Baltz, Phys. Rev. A **52**, 4970 (1995).
27. N. Toshima and J. Eichler, Phys. Rev. A **42**, 3896 (1990).

28. V. I. Matveev and M. M. Musakhanov, Zh. Éksp. Teor. Fiz. **105**, 280 (1994) [JETP **78**, 149 (1994)].
29. A. J. Baltz, Phys. Rev. Lett. **78**, 1231 (1997).
30. A. Scrinzi and B. Piraux, Phys. Rev. A **56**, R13 (1997).
31. J. Bauer, J. Plucinski, B. Piraux, *et al.*, J. Phys. B **34**, 2245 (2001).
32. G. Lagmago Kamta, T. Grosges, B. Piraux, *et al.*, J. Phys. B **34**, 857 (2001).
33. C. C. Chirilǎ, N. J. Kylstra, R. M. Potvliege, *et al.*, Phys. Rev. A **66**, 063411 (2002).
34. M. W. Walser, C. H. Keitel, A. Scrinzi, *et al.*, Phys. Rev. Lett. **85**, 5082 (2000).
35. D. B. Miloǒevič, S. Hu, and W. Becker, Phys. Rev. A **63**, 011403 (2001).
36. N. J. Kylstra, R. M. Potvliege, and C. J. Joachain, J. Phys. B **34**, L55 (2001).
37. A. M. Dykhne and G. L. Yudin, Usp. Fiz. Nauk **125**, 377 (1978) [Sov. Phys. Usp. **21**, 549 (1978)].
38. V. I. Matveev and É. S. Parilis, Usp. Fiz. Nauk **138**, 573 (1982) [Sov. Phys. Usp. **25**, 881 (1982)].
39. J. Eichler, Phys. Rev. A **15**, 1856 (1997).
40. G. L. Yudin, Zh. Éksp. Teor. Fiz. **80**, 1026 (1981) [Sov. Phys. JETP **53**, 523 (1981)].
41. V. I. Matveev, Fiz. Élem. Chastits At. Yadra **26**, 780 (1995) [Phys. Part. Nucl. **26**, 329 (1995)].
42. I. C. Percival, in *Atoms in Astrophysics*, Ed. by P. G. Burke, W. Eissner, D. G. Hummer, and I. C. Percival (Plenum, New York, 1983; Mir, Moscow, 1998).
43. V. I. Matveev, Zh. Éksp. Teor. Fiz. **121**, 260 (2002) [JETP **94**, 217 (2002)].
44. A. B. Migdal, *Qualitative Methods in Quantum Theory* (Nauka, Moscow, 1975; Benjamin, Reading, Mass., 1977).
45. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 4th ed. (Nauka, Moscow, 1989; Pergamon, New York, 1977).
46. A. R. Holt, J. Phys. B **2**, 1209 (1969).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 922–931.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 1030–1040.

PLASMA,
GASES

Measurements of the Electron Concentration and Conductivity of a Partially Ionized Inert Gas Plasma

N. S. Shilkin*, S. V. Dudin, V. K. Gryaznov, V. B. Mintsev, and V. E. Fortov

Institute of Problems of Chemical Physics, Russian Academy of Sciences,
Chernogolovka, Moscow oblast, 142432 Russia
**e-mail: shilkin@ficp.ac.ru*

Received March 3, 2003

Abstract—The results are presented of experiments performed to measure the electron concentration and conductivity of a partially ionized inert gas plasma in a magnetic field. The plasma was generated behind the front of incident and reflected shock waves excited by explosively driven linear generators. A magnetic field of about 5 T was formed inside a solenoid wound on the generator channel. Measurements were taken at $P = 30$–650 MPa, $T = 6000$–17000 K, and a Coulomb nonideality parameter of 0.01–2.8. Electron concentrations calculated from measured Hall voltages reached 1.6×10^{21} cm^{-3}. The recorded conductivities were in the range 0.1–200 Ω^{-1} cm^{-1}. The experimental results were compared with various models of the thermodynamic and transport properties of a nonideal plasma. *© 2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The determination of the electron concentration of a low-density plasma does not create difficulties for either measurements or theory, which calculates ionization equilibria and transport properties using well-grounded models [1, 2]. The situation becomes much more complex as the plasma density increases, when the mean potential energy of the interaction between plasma particles ($E_p = e^2/r_D$, where r_D is the Debye shielding radius) is comparable to their mean kinetic energy ($E_k = kT$) and the plasma becomes nonideal (the nonideality parameter is $\Gamma = E_p/E_k \sim 1$) [3–5].

Interest in a nonideal plasma stems from both the necessity of describing the fundamental properties of media under extreme conditions and the feasibility of its practical use [4].

A rigorous theoretical description of a dense plasma is a complex problem [3, 4, 6], primarily because taking into account interparticle interactions in a plasma requires a rigorous solution to the quantum-mechanical many-body problem, which has only been solved in the simplest cases. For this reason, additional approximations and simplifications are introduced for selecting models and methods for solving equations. The calculation results therefore strongly depend on the initial assumptions and significantly change when these assumptions are refined.

This brings to the fore experimental studies of plasmas with strong interparticle interactions [4]. However, a dense plasma is not a simple object for experimental studies either. Among the difficulties of its generation and parameter determination, it should be stressed that measurements are always performed under high temperature and pressure conditions, which entails performing experiments in a pulsed mode. This enhances the requirements on measuring devices, which must record the results in submicrosecond times with an acceptable accuracy. In addition, the unit for measurements should have an energy source of a fairly large capacitance and provide high-speed energy transfer to the object under study.

Currently, electric and dynamic methods for generating nonideal plasmas are widely used [4]. Fairly extensive experimental studies of their thermodynamic, transport, and optical properties have been performed [3, 4]. There remains, however, much uncertainty in the data on the electron concentration of a nonideal plasma in the region of its partial ionization. In this parameter range, a "chemical" model is usually applied to study ionization equilibria. This model enables reliable data to be obtained for a weakly nonideal plasma. With a nonideal plasma, an uncertainty arises in calculations of plasma parameters, including the concentration of electrons, because of the absence of a rigorous criterion for dividing them into free and bound electrons. Direct experimental measurements of this important parameter are lacking.

In this work, the concentration of electrons in a nonideal plasma was determined using the approach based on Hall voltage measurements, which is extensively used in solid-state physics [7].

2. PLASMA DIAGNOSTICS AND GENERATION

Because a high energy concentration is necessary to produce a nonideal plasma, we used the dynamic method based on gas compression and irreversible heating in the front of high-power shock waves. This technique is a reliable means of generating spatially uniform plasma bunches with a characteristic size (about 1 cm) sufficient for probe diagnostics. For self-similar flows, the conservation laws at the shock-wave discontinuity are written in a simple algebraic form [8], which allows the thermodynamic parameters of a shock-compressed gas to be calculated from the measured flow hydrodynamic parameters.

We used explosively driven linear generators [9, 10], one of which is schematically shown in Fig. 1. The working channel of the generator was a polyvinyl plastic tube approximately 30 cm long with an inner diameter of about 5 cm. The active explosive charge (ammonite or RDX) was loaded into the channel. The total charge length was 12–15 cm. It was initiated by an electric detonator. A shock wave was formed when detonation products expanded into the gas to be studied. The thickness of plasma bunches was determined by the difference between the velocity D of the shock wave front and the mass velocity U of plasma motion. Under stationary conditions, D and U were constant and the thickness of the plasma bunch linearly increased with time. Plasma diagnostics were performed in both incident and obstacle-reflected shock waves. An obstacle made of organic glass was placed at a distance of 70–100 mm from the end face of the charge. The interaction between an incident plasma bunch and the obstacle created a reflected shock wave, which caused further plasma heating and ionization.

The characteristics of the plasma formed behind the shock wave front are determined by the initial gas state and shock wave velocity D. Changing them enables a plasma with the required parameters to be obtained. In this work, the initial gas pressure was $P_0 = 0.4$ MPa, the initial temperature was $T_0 = 300$ K, and the range of shock wave velocities was $D = 2$–10 km/s. The velocity D depended on the selected condensed explosive. Bulk ammonite gave velocities D of 2–3 km/s in Ar and Xe, and pressed desensitized RDX gave D of 9–11 km/s in He. The D, P_0, and T_0 parameters were used to calculate the thermodynamic [11] and electrophysical plasma parameters. The calculations were based on the assumption that the flow was one-dimensional and stationary. The validity of this assumption was checked in a special series of optical (high-speed filming) and electric contact probe experiments. It was shown that, at a distance of 50–100 mm from the end face of the charge, the flow was one-dimensional and stationary, which allowed a uniform plasma bunch with a characteristic size of several centimeters to be obtained. One of the problems was the presence of lateral unloading waves, which noticeably curved the plane shock wave front as the wave moved from the end charge face. To solve this problem, the measuring cell was mounted where the plane-front region of the plasma bunch had a size (20–25 mm) sufficient for measuring plasma properties. Of no less importance is the selection of the generator material when a shock-compressed plasma is formed in explosively driven linear generators. On the one hand, this material must resist the necessary initial pressures; on the other, the velocity of shock-wave perturbation propagation over generator walls must be lower than the velocity of the shock wave in the gas under study to prevent the “channel” effect. We studied argon and xenon plasmas in the region of low shock wave velocities using generators made of polyvinyl plastic; the velocity of sound in polyvinyl plastic was $c_s \approx 2.3$ km/s.

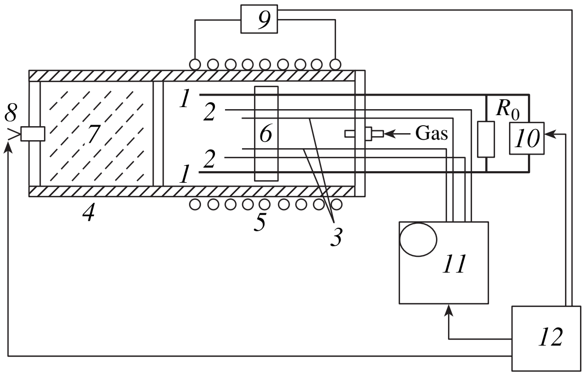


Fig. 1. Scheme of the experimental unit: *1*, power probes; *2*, probes for conductivity measurements; *3*, probes for electron concentration measurements; *4*, generator channel; *5*, solenoid; *6*, obstacle; *7*, explosive; *8*, electric detonator; *9*, solenoid current source; *10*, source of current through a plasma; *11*, oscilloscopes; and *12*, delayed pulse generator.

The low-frequency ($\omega \ll \omega_{pe} = \sqrt{4\pi n_e e^2/m_e}$) plasma conductivity was measured by the four-probe method extensively used in solid-state physics. This method allows the influence of near-electrode phenomena to be avoided in shock-compressed plasma measurements [12, 13]. The transport electric current I was supplied through two external electrodes (Fig. 1, probes *1*), and the voltage U_c caused by this current was taken from two internal electrodes (Fig. 1, probes *3*). The conductivity was calculated with an accuracy of about 30% by the equation

$$\sigma = \frac{IK}{U_c h}, \tag{1}$$

where h is the plasma thickness and K is the geometrical factor of the measuring cell, which takes into account the size of the probes and the sample, and their mutual arrangement. This coefficient was calculated theoretically and refined by electrolytic modeling. Prior to an explosion, the obstacle with the probes was placed into an electrolyte having known parameters; the geometry of measurements was preserved. The conductivity of the electrolyte was calculated from its concentration using handbook data [14]. The errors in K, h, I, and U_c were approximately 10% without taking into account

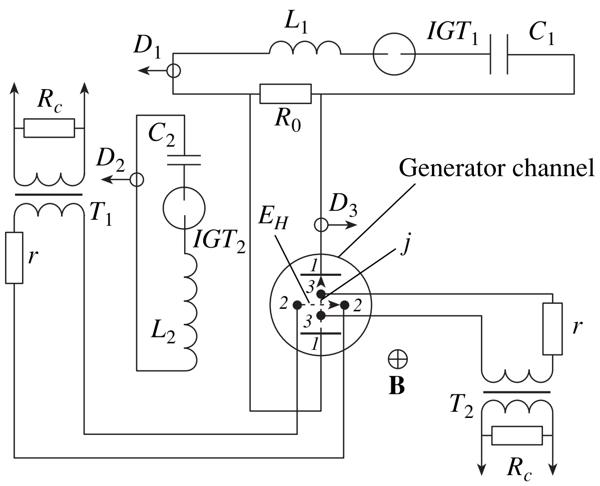


Fig. 2. Electrical circuit for measurements: $L_1C_1R_0$, circuit that forms current through a plasma; L_2C_2, solenoid circuit; D_1–D_3, Rogovski loops; T_1–T_2, pulse transformers; IGT_1 and IGT_2, ignitrons; *1*, power probes; *2*, probes for electron concentration measurements; and *3*, probes for conductivity measurements.

magnetic field effects. This gave a 20% accuracy of conductivity determinations. With the magnetic field included, the error in I and U_c increased by a factor of 1.5–2.

The procedure for determining the electron concentration was based on measuring Hall effect parameters. The Hall effect is the appearance of a Hall electric field in a medium in magnetic field B when current I noncollinear with field B flows through the medium. The Hall electric field is

$$E_H = R_H[\mathbf{j} \times \mathbf{B}], \tag{2}$$

where j is the current density in the sample and R_H is the Hall coefficient. For charge carriers of one type, $R_H = \pm r_H/n_{\pm}e$, where n is the concentration of the carriers, the minus sign corresponds to electronic conductivity, the plus sign corresponds to the positively charged carriers (ions and holes), and r_H is a theoretical factor on the order of one that takes into account the velocity distribution of the ions and the mechanism of their scattering; for scattering by a Coulomb potential, $r_H = 1.93$ [7, 15, 16].

In electron concentration measurements, a plasma was placed into a magnetic field. The transport current I was passed through one pair of probes (Fig. 1, probes *1*) immersed into the plasma, and the Hall voltage U_H was taken from the other pair of probes (Fig. 1, probes *2*). The Hall coefficient was calculated with an accuracy of about 50% by the equation

$$R_H = \frac{U_H h}{BIQ}, \tag{3}$$

where Q is the geometric factor of the cell, which was determined in a special series of measurements with a weakly nonideal helium plasma.

In an incident wave, the magnetic field penetrates into the plasma volume during characteristic experiment times on the order of 1 μs (the plasma skin depth is $\delta \approx 5$ cm $> r = 2.5$ cm, where r is the radius of the plasma bunch). In a reflected wave, we have frozen magnetic field conditions ($\delta < r$).

The main difficulty that limits probe measurements in a shock-compressed plasma in a pulsed magnetic field is a high level of extraneous signals generated by changes in the magnetic flux through the measuring circuits. With Hall effect measurements, the problem is complicated by a decrease in the recorded signal intensity as the concentration of electrons increases. To solve this problem, the control circuits were minimized and placed in a plane parallel to the magnetic field. The probes were arranged symmetrically with respect to the axis of the solenoid. All measuring leads were connected to the explosive assembly coaxially with the magnetic field direction. To identify extraneous signals generated by magnetic field changes against the background of the total signal, the measuring probes were made longer than the power probes in several experiments. The moving plasma first connected to the measuring probes. When plasma connected to the power probes, signal was added to the noise on the measuring probes.

A pulsed magnetic field was generated by a discharge of a capacitor bank of capacitance $C_2 = 1$ mF (Fig. 2) through an inductance solenoid $L_2 = 20$ μH wound on the explosively driven generator. A 20–30 kA current flowed through it in experiments; the derivative of the current was measured by a Rogovski loop D_2 accurate to about 10%. The magnetic induction in the center of the solenoid was determined by the equation

$$B = \frac{\mu_0 NI}{\sqrt{l^2 + d^2}},$$

where μ_0 is the magnetic constant, $N = 25$ is the number of solenoid turns, $I = 25$–30 kA is the electric current flowing through the solenoid, $l = 0.15$ m is the length of the solenoid, and $d \approx 0.05$ m is the diameter of the solenoid.

The source of the current supplied to the plasma was capacitor bank C_1. Current pulses were formed by an RLC circuit ($C_1 = 100$ μF, $L_1 \approx 300$ μH, and $R_0 \approx 1$ Ω). Current switching was effected by an IRT-6 ignitron. Before power probes were closed by a plasma, the current flowed through shunt resistance R_0. The limiting voltage of capacitor bank C_1 charging was 2–2.5 kV; it was determined by the absence of breakdown between power electrodes spaced 12–20 mm apart on the experimental assembly. Inductive sensor D_3 measured the current flowing through the plasma. It was placed directly on the measuring cell. The voltage for conductivity and electron concentration determinations was measured by pulsed transformers T_1 and T_2 made of ferrite high-frequency cores. Resistors r limited currents and thereby prevented core saturation. The transformers were tested to find if they reproduced microsecond pulse fronts with a time resolution no lower than 0.1–0.3 μs. Resistances $R_c = 50\ \Omega$ were placed at oscilloscope inputs to correlate measuring circuits. The signals from the transformers and Rogovski loops were transferred to a complex of S9-8 dual trace storage oscilloscopes. To remove currents induced by high-power electric and magnetic fields, the measuring oscilloscopes were placed into a screened room. In each experiment, two or three independent oscillograms of each parameter were digitized and sent to a PC through a general-use channel. The recording apparatus, devices for switching kiloampere currents, and devices for blasting explosives were triggered by a delay pulse generator at the required time instants. The delay times were selected to ensure maximum currents through the solenoid and the plasma bunch at the instant when it flew up to the obstacle.

We also measured the velocity D of the shock wave front with the use of three pairs of probes of different lengths. If the times of the arrival of a plasma bunch to the probes and the distance between the probes are known, it is easy to calculate D. The accuracy of D determinations was 1–3%. The thickness of the plasma bunch was calculated from the Hugoniot curves for inert gases [17] taking into account the distance from the charge end face to the obstacle with the probes.

Spatially uniform samples are needed to simplify interpreting Hall effect data. This requirement predetermined our selection of the method for plasma generation. The use of shock waves allows uniform dense plasma bunches several centimeters in size to be obtained. The procedure is characterized by high reproducibility of measurement results.

Correct Hall voltage measurements also require the use of a spatially uniform magnetic field freely penetrating into a plasma. A plasma is a moving object in our experiments, and its flow should be such that the time of magnetic field diffusion into its volume be much shorter than the characteristic hydrodynamic time; that is, the magnetic Reynolds number determined as the ratio between these times should be smaller than one.

The estimates obtained in [18, 19] for $B = 5$ T and $r = 2.5$ cm show that such flows arise in an argon plasma at $D < 4$ km/s and in a xenon plasma at $D < 3$ km/s. In addition, we may expect the formation of a plasma with nonideality parameters up to one in argon and up to three in xenon behind the reflected shock wave front at the specified velocities of the incident shock wave front.

To simplify the interpretation of measurement results, the conductivity of the plasma must be higher than that of detonation products. According to [20], the characteristic conductivity of detonation products behind the detonation front is fairly high (0.1–1 Ω^{-1} cm^{-1}), but the expansion of detonation products substantially (10–100 times) decreases it. It follows from our experimental oscillograms that the arrival of detonation products to the power and measuring probes did not cause sharp signal changes, which was evidence that detonation products did not shunt the plasma.

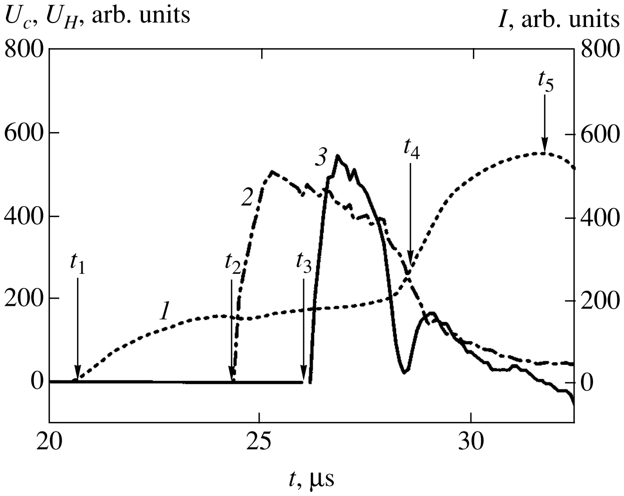


Fig. 3. Typical oscillogram of experiments with an argon plasma: (*1*) current through the plasma, (*2*) voltage U_C on probes for conductivity measurements, and (*3*) voltage U_H on probes for electron concentration measurements; t_1–t_3 are the instants of plasma arrival at probes *1–3*, t_4 is the beginning of reflection, and t_5 is the end of reflection.

A typical experimental oscillogram taken from [21] is reproduced in Fig. 3, where the incident and reflected shock wave regions are shown. The plasma conductivity increases in reflected waves, its resistance decreases and becomes smaller than the shunt resistance, and the current through the plasma therefore increases. The Hall electromotive force decreases in the reflected wave because of an increase in the electron concentration caused by further plasma heating and compression. Voltage taken from the probes for determining conductivity in the reflected wave decreases because the resistance of the plasma decreases more rapidly than the current increases.

3. RESULTS AND DISCUSSION

Measurements were performed in shock-compressed, partially ionized (the degree of ionization $\alpha = 10^{-6}$–10^{-1}) helium, argon, and xenon plasmas. The electronic component was not degenerate ($n_e \lambda_e^3 \ll 1$). The range of parameters that we studied was $0.01 < \Gamma < 2.8$, $P = 20$–650 MPa, and $T = 6000$–22000 K.

The thermodynamic parameters and the composition of shock-compressed plasmas were calculated

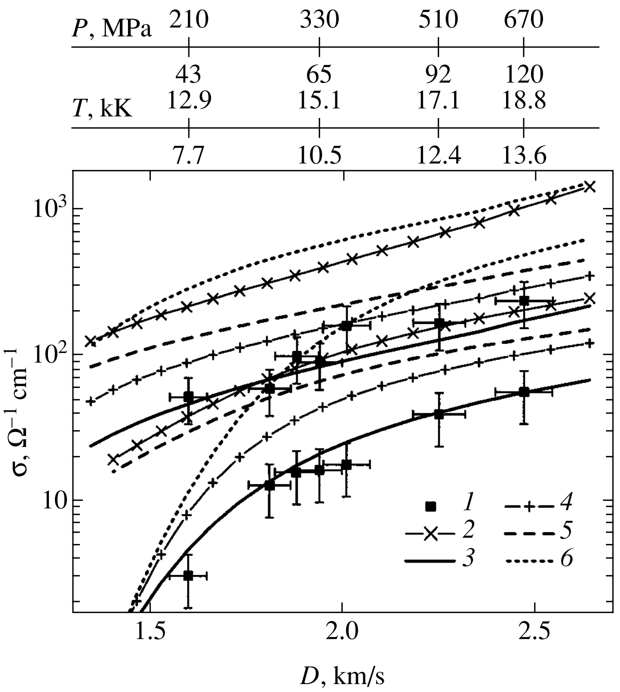


Fig. 4. Calculated and experimental conductivity of a xenon plasma: (*1*) our experimental data, (*2*) calculations by the Spitzer equation, (*3*) calculations by (6), (*4*) additive approximation, (*5*) Coulomb conductivity component, and (*6*) conductivity resulting from electron scattering by atoms.

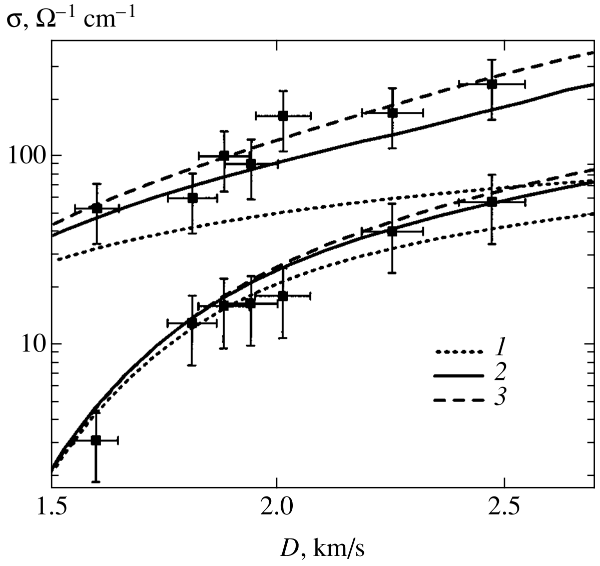


Fig. 5. Calculated and experimental xenon plasma conductivity: (*1*) ideal plasma approximation with the number of energy levels $n = 100$ for calculating the partition functions of atoms and ions; (*2*) the Debye approximation in a grand canonical ensemble, the partition functions of atoms and ions are calculated for energy levels E_n not exceeding $I - \Delta I$, where ΔI is the decrease in the ionization potential; and (*3*) Debye approximation in a grand canonical ensemble, the partition functions of atoms and ions and calculated for the ground state only; the experimental values are shown by symbols.

using the SAHA-IV code based on the chemical plasma model [1, 4, 5, 22]. In this model, a plasma is treated as a quasi-neutral mixture of particles of different kinds (electrons, atoms, and variously charged ions) that interact with each other. The Coulomb interaction of the particles was described at the level of the Debye approximation in a grand canonical ensemble [22], which was earlier applied in [17, 23] to describe Coulomb effects in a shock-compressed plasma. At high degrees of plasma compression, the repulsion of atoms and ions at short distances [24] caused by overlap of their outer electron shells becomes important. This effect was described in the approximation of "soft" spheres [25]. The ratio between the number densities of the particles of different kinds n_j was determined from the condition of minimum free energy [2]. No corrections to the thermodynamic functions for magnetic field effects were introduced.

The experimental conductivities are plotted in Figs. 4–6 in σ–*D* coordinates, where *D* is the velocity of the incident shock wave front and σ is the plasma conductivity behind the incident and obstacle-reflected shock wave fronts. All data correspond to an initial gas pressure of 0.4 MPa. The calculated values are shown by curves. The lower groups of curves correspond to the incident wave, and the upper curves, to the reflected wave. The calculated plasma pressures and temperatures are plotted on additional abscissa axes in Figs. 4 and 5. The lower values correspond to the plasma parameters in the incident wave, and the upper values, to the plasma parameters in the reflected wave.

In the presence of an external magnetic field and in the case of a spherically symmetrical conduction band, the equation for the transverse conductivity component obtained from the kinetic equation in the τ approximation has the form [26]

$$\sigma = \frac{e^2}{3\pi^2}\int_0^\infty -\frac{\partial f_0}{\partial \varepsilon}\frac{\tau(\varepsilon)}{1+(\Omega\tau)^2}\frac{k^3(\varepsilon)}{m(\varepsilon)}d\varepsilon, \qquad (4)$$

where $\Omega = eH/m_ec$ is the electron cyclotron frequency, ε is the kinetic energy of the electrons, *k* is the wave vector modulus, τ is the momentum relaxation time, and $f_0(\varepsilon)$ is the equilibrium distribution function of the electrons, which is related to the electron concentration by the equation

$$n_e = \frac{1}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2}\int_0^\infty f_0(\varepsilon)\varepsilon^{1/2}d\varepsilon\Bigg/\int_0^\infty f_0(\varepsilon)d\varepsilon. \qquad (5)$$

Assuming that the conduction band is parabolic ($\varepsilon = \hbar^2k^2/2m_e$) and taking into account that the electrons obey classical statistics, we obtain the final equation for conductivity calculations,

$$\sigma = \frac{4}{3\sqrt{\pi}}\frac{e^2 n_e (kT)^{-5/2}}{m_e}\int_0^\infty \frac{\varepsilon^{3/2}\tau(\varepsilon)\exp(-\varepsilon/kT)}{1+(\Omega\tau)^2}d\varepsilon, \quad (6)$$

where

$$\tau(\varepsilon) = \left[\sum_j \frac{(\nu_{ei}(\varepsilon))_j}{\gamma_j} + \nu_{ea}(\varepsilon)\right]^{-1}, \quad (7)$$

the summation is over ions with different charges; γ_j is a factor taking into account electron collisions and depending on the ionization multiplicity [27]; ν_{ei} and ν_{ea} are the energy-dependent frequencies of electron collisions with ions and atoms,

$$\nu_{ea} = \sqrt{\frac{2\varepsilon}{m_e}}n_a Q_{ea}, \quad \nu_{ei} = \sqrt{\frac{2\varepsilon}{m_e}}n_i Q_{ei}; \quad (8)$$

n_a and n_i are the concentrations of atoms and ions, respectively; and Q_{ea} and Q_{ei} are the transport cross sections of electron scattering by atoms and ions, respectively. The cross sections of electron scattering by atoms were taken from [28]. The cross sections of electron scattering by ions were calculated as

$$Q_{ei} = \sum_j \frac{Z_j^2 \pi e^4}{\varepsilon^2} I_i, \quad (9)$$

where

$$I_i = \frac{1}{2}\ln\left[1+\left(\frac{2\varepsilon r_d}{e^2}\right)^2\right], \quad r_d = \sqrt{\frac{kT}{8\pi e^2 n_e}}.$$

In the absence of a magnetic field, (6) transforms into the Frost interpolation equation [29]. The Frost equation is constructed to give the Lorentz and Spitzer asymptotic behaviors for weakly and completely ionized plasmas, respectively, in the limiting cases.

The main problems in calculating the conductivity of a partially ionized plasma are the selection of the thermodynamic model for determining the component composition of the plasma and the model for describing its transport properties. Let us estimate the influence of each of these factors separately. First, we will perform calculations using different transport property models with a set thermodynamic model. Next, we will use different thermodynamic models with a set model of transport properties.

Different methods for calculating the conductivity of a partially ionized xenon plasma are compared in Fig. 4 for a thermodynamic model that takes into account Coulomb interactions at the level of the Debye approximation in a grand canonical ensemble and a model of soft spheres to describe the repulsion of heavy particles [25]. Curve *2* corresponds to conductivity calculations by the Spitzer equation [27]. The calculations according to the model described above (which is similar to the Frost equation) are shown by curve *3*. Curve *4* corresponds to independent electron scattering by atoms and ions [30]. We then have

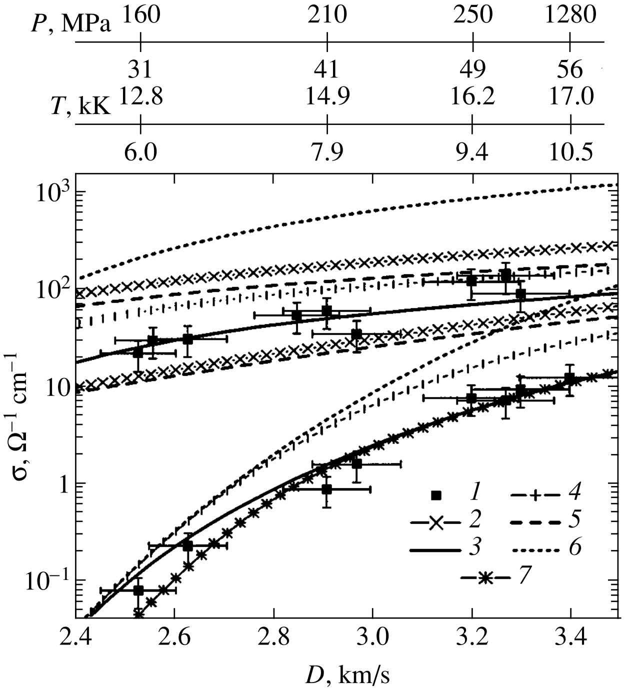


Fig. 6. Conductivity of a nonideal argon plasma (thermodynamic properties are described in the Debye approximation for a grand canonical ensemble): (*1*) our experimental data, (*2*) calculations by the Spitzer equation, (*3*) calculations by (6) with $B = 0$, (*4*) additive approximation; (*5*) Coulomb conductivity component, (*6*) conductivity resulting from electron scattering by atoms, and (*7*) calculations by (6) with $B = 5$ T.

$$\frac{1}{\sigma} = \frac{1}{\sigma_{ei}} + \frac{1}{\sigma_{ea}}.$$

This equation is based on the assumption that the total resistance of the conducting region is the sum of the neutral and charged component resistances to the motion of electrons; σ_{ei} is the Coulomb conductivity component, which corresponds [29] to including ion–electron collisions alone; and σ_{ea} is the conductivity [29] only taking into account electron–neutral collisions. The calculated σ_{ei} and σ_{ea} values are shown by curves *5* and *6*, respectively. According to Fig. 4, calculations with different models of the transport properties of a dense plasma give a large spread of conductivity values. The experimental data are reasonably described by (6) only, the other methods exaggerate conductivities compared with experiment. The dependences for

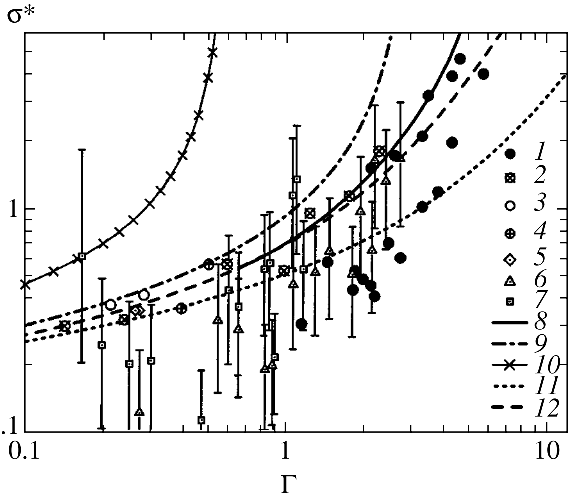


Fig. 7. Dependences of the dimensionless conductivities of nonideal media on the nonideality parameter: $\sigma^* = \sigma_C/\sigma_S$ for experimental values and $\sigma^* = \sigma/\sigma_S$ for theoretical models, $\sigma_S = 2^{5/2}(kT)^{3/2}/e^2\sqrt{m_e}$; (*1–4*) are data from [12, 13]; (*5*), (*6*), and (*7*) are our data for helium, xenon, and argon, respectively; (*8*) T-matrix, (*9*), (*10*), and (*11*) are data from [33, 27, 32, 34], respectively; and (*12*) is dependence (10) at $K = 1$ and $\Omega\tau = 0$.

the other thermodynamic models of a dense xenon plasma have the same shape as those shown in Fig. 4.

Our data on the conductivity of a partially ionized xenon plasma are compared in Fig. 5 with the results of calculations by (6) for different thermodynamic models. The calculated conductivity curves for the incident wave differ insignificantly and are not at variance with the measured conductivities. For the reflected wave, the difference between the experimental and calculated conductivities becomes noticeable. The conclusion can be drawn that the “ideal plasma” model with the number of energy levels $n = 100$ in the partition functions for the atoms and ions gives the worst agreement with the experimental data.

Different methods for calculating the conductivity of a partially ionized argon plasma at the level of the Debye approximation in a grand canonical ensemble are compared in Fig. 6. At velocities $D < 2.9$ km/s, the electronic component is magnetized behind the incident wave front. The magnetic field contribution is seen when we compare curves *7* (obtained taking the magnetic field into account) and *3* (obtained ignoring the magnetic field). The behavior of the curves in Fig. 6 qualitatively coincides with that of the curves in Fig. 4. Most models predict exaggerated conductivity values compared with experiment. Equation (6) satisfactorily describes our measurement results.

When the contribution of atom–electron collisions was substantial, the Coulomb conductivity component (Fig. 7) was separately estimated by a method similar to that described in [31]. Namely, the frequency of ion–electron collisions $\left(\sum_j (\nu_{ei}(\varepsilon))_j/\gamma_j\right)$ in the equation for conductivity was multiplied by a factor K that brought the calculated conductivities in coincidence with the experimental values. The Coulomb conductivity component σ_C was then calculated as

$$\sigma_C = \frac{1}{K}\frac{4e^2 n_e (kT)^{-5/2}}{3\sqrt{\pi}m_e} \times \int_0^\infty \frac{d\varepsilon\,\varepsilon^{3/2}\exp(-\varepsilon/kT)}{1+(\Omega\tau)^2}\left[\sum_j \frac{(\nu_{ei}(\varepsilon))_j}{\gamma_j}\right]^{-1}, \qquad (10)$$

where τ in the equation for the magnetization parameter is determined by electron collisions with both ions and atoms.

The Coulomb contributions to conductivity obtained as described above are shown in Fig. 7, which also presents several theoretical approximations and experimental data from other works. Curve *9* corresponds to the Spitzer theory, which describes well the conductivity of a completely ionized gas or a gas with a low degree of ionization. At $\Gamma \geq 3$, Spitzer conductivities are indeterminate. The experimental data on shock compression of inert gases and air [12, 13] are shown by curves *1–4*. Curve *12* corresponds to theoretical dependence (6) at $K = 1$ and $\Omega\tau = 0$. This curve has a correct asymptotic behavior in the limit of a weakly nonideal plasma. Curve *10* corresponds to a solution to the converging kinetic equation from [32], which is valid to the first order in Γ. Interparticle interactions are taken into account in this equation through a screened Coulomb potential.

For charge carriers of two types (electrons and ions in our problem), calculations give the Hall coefficient in the form [15]

$$R_H = \frac{r_i n_i \mu_i^2 - r_e n_e \mu_e^2}{e(n_i\mu_i + n_e\mu_e)^2}. \qquad (11)$$

Here, index “*e*” corresponds to electrons; “*i*,” to ions; $r_{e,\,i}$ and $\mu_{e,\,i}$ are the Hall factors and mobilities of the charged components, respectively; and n_e and n_i are their concentrations.

We studied a singly ionized plasma; that is, $n_e = n_i$ in our experiments. Negative Hall electromotive force values were obtained. This leads us to conclude that the mobility of the electrons was higher than that of the ions; that is, $\mu_e > \mu_i$.

The drift mobility μ_d of electrons differs from the Hall mobility μ_H, $r_H = \mu_H/\mu_d$ [16]. This requires estimating the Hall factor r_H to refine the electron concentration. The Hall factor was calculated taking into account the magnetic field value by the equation taken from [16],

$$r_H = \frac{\left\langle \dfrac{\mu_e^2}{1+\mu_e^2 B^2} \right\rangle}{\left\langle \dfrac{\mu_e}{1+\mu_e^2 B^2} \right\rangle^2 + \dfrac{e^2 B^2}{m_e^2}\left\langle \dfrac{\mu_e^2}{1+\mu_e^2 B^2} \right\rangle^2}, \qquad (12)$$

where angle brackets denote averaging over all energies according to the equation

$$\langle Y \rangle = -\frac{2}{3}\int_0^\infty Y \varepsilon^{3/2} \frac{\partial f_0}{\partial \varepsilon} d\varepsilon \Big/ \int_0^\infty \varepsilon^{1/2} f_0 d\varepsilon. \qquad (13)$$

Here, f_0 is the distribution function of the electrons and $\mu_e = \mu_e(\varepsilon) = e\tau(\varepsilon)/m_e$ is the energy dependence of the mobility of the electrons. The energy dependence of the momentum relaxation time $\tau(\varepsilon)$ was taken the same as in conductivity calculations.

The Hall factor shown in Fig. 8 as a function of the nonideality parameter was obtained only taking into account electron–ion and electron–electron collisions. In the limit of a weakly ionized plasma when $\Gamma \longrightarrow 0$, we obtain the result known from solid-state physics, namely, $r_H \longrightarrow 1.93$, which corresponds to electron scattering by ionized impurities. The Hall factor value is determined by the type of the functional dependence of the momentum relaxation time on energy. This dependence changes at $\Gamma \approx 1$, which explains the kink present in the curve.

Apart from the Debye approximation in a grand canonical ensemble, the ideal plasma [2] and Debye–Hückel models, which are chemical model variants [22], were used to describe Coulomb interaction effects for comparison with the experimental electron concentrations. The ideal plasma model completely ignores interparticle interactions and is therefore only applicable to a strongly rarefied or high-temperature plasma. Along with various models of Coulomb nonideality, the influence of bound states on the total thermodynamic values and the calculated electron concentrations n_e was analyzed using several variants of imposing limitations on partition functions [22].

The experimental dependences of the electron concentrations obtained taking into account the Hall factor (12) values on the shock wave front velocity at a $P_0 = 0.4$ MPa initial gas pressure are compared in Fig. 9 with the data calculated using various theoretical models. The calculated degrees of plasma ionization are plotted on the additional abscissa axis in Fig. 9. The lower and upper values correspond to the plasma parameters in the incident and reflected waves, respectively.

Two different limitations were imposed on partition functions when calculations were performed in the ideal plasma approximation. In one variant, up to a hundred energy levels of the atom and ion were used; in the other, the only ground state was included. The Coulomb interaction of free charges was taken into account not only in the Debye approximation in a grand canonical ensemble but also including terms quadratic in the nonideality parameter. Calculations in the Debye–Hückel approximation are only possible in a narrow range of low shock wave front velocities for both argon and xenon. The velocities at which we were able to use this model are denoted by arrows in Fig. 9. The higher shock compression rates correspond to larger nonideality parameters, which result in the well-known [22] model instability sometimes interpreted as a phase transition.

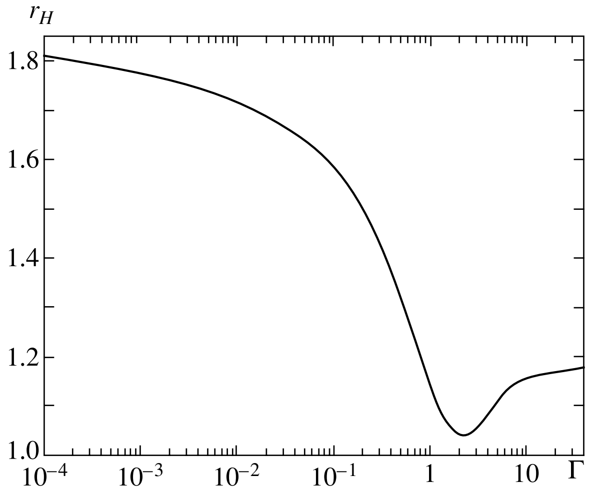


Fig. 8. Electronic Hall factor taking into account electron collisions with charged particles.

Moderately nonideal states ($\Gamma < 0.3$) are formed in shock-compressed argon behind the incident wave front in the velocity range of our measurements (2.3–3.5 km/s) at an initial pressure of 0.400 ± 0.001 MPa. At such nonideality parameter values, the theoretical models diverge insignificantly and do not contradict our experimental results. The Γ values in argon behind the reflected wave front reached 1.2. Figure 9a shows that, on the whole, the experimental n_e values behind the reflected wave front in argon lie below the values obtained with the theoretical models described above. This result can, for instance, be explained by the undervaluation of the Hall factor, which may be caused by the use of inaccurate cross sections for electron scattering by atoms and an incorrect functional dependence of the momentum relaxation time on energy.

The experimental n_e values obtained for the xenon plasma are shown in Fig. 9b. The calculated Coulomb nonideality parameter Γ for the xenon plasma reached 1.3 behind the incident wave front and 2.8 behind the reflected wave front. The calculated and experimental n_e values for the incident wave do not contradict each

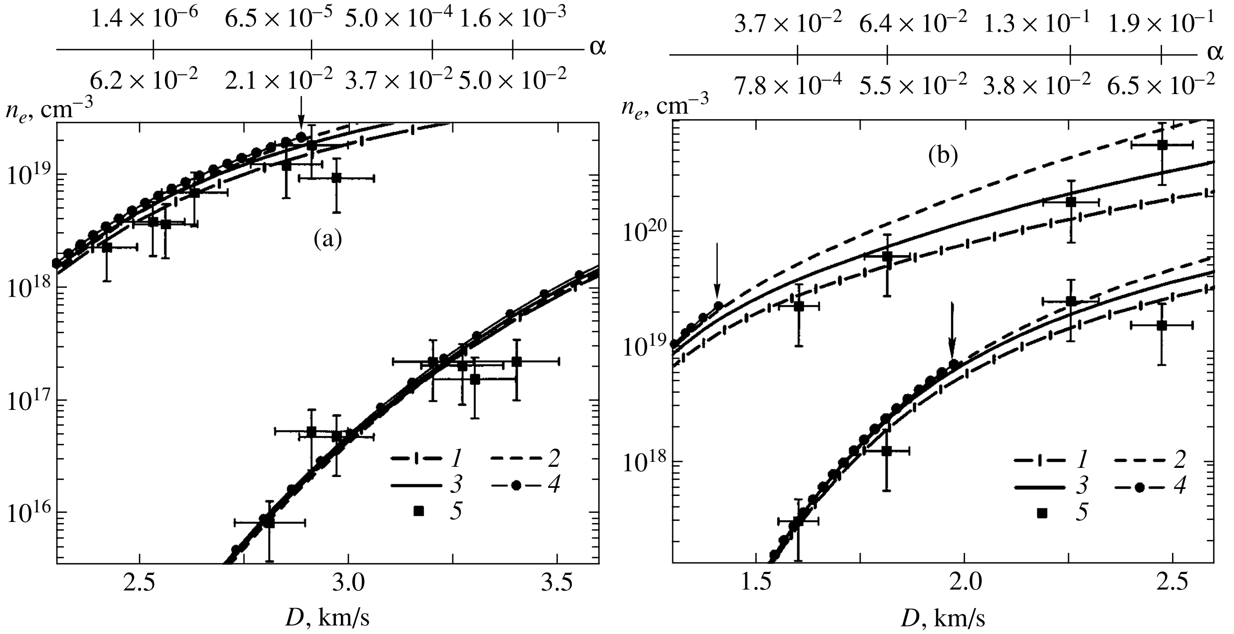


Fig. 9. Experimental and calculated electron concentrations in nonideal (a) argon and (b) xenon plasmas: (*1*) ideal plasma approximation with the number of energy levels $n = 100$ in the partition functions of atoms and ions; (*2*) ring approximation in a grand canonical ensemble, expansion to the second order in Γ, the partition functions of atoms and ions are limited to the ground state; (*3*) Debye approximation in a grand canonical ensemble, the partition functions of atoms and ions were calculated including E_n levels that did not exceed $I - \Delta I$; (*4*) Debye–Hückel model; and (*5*) experiment.

other. Noticeable discrepancies are observed for the reflected wave. It can be suggested that curves *3* and *4* in Fig. 9b predict exaggerated n_e values at $\Gamma > 1.5$, whereas the Debye approximation in a grand canonical ensemble is valid up to $\Gamma = 3$.

If the experimental electron concentrations are analyzed taking into account the experimental plasma conductivity values, the conclusion can be drawn that, at the current level of the accuracy of measurements, a noncontradictory description of the electron concentration and conductivity is attained using the Debye approximation in a grand canonical ensemble or the ideal plasma approximation with partition functions of atoms and ions calculated for the ground state.

It appears that drawing more definite conclusions on the applicability of various models requires more precise measurements.

REFERENCES

1. W. Ebeling, Physica (Amsterdam) **43**, 293 (1969).
2. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 2nd ed. (Nauka, Moscow, 1964; Pergamon Press, Oxford, 1980).
3. *Encyclopedia of Low-Temperature Plasma*, Ed. by V. E. Fortov (Nauka/Interperiodika, Moscow, 2000), Vol. 1.
4. V. E. Fortov and I. T. Yakubov, *Nonideal Plasma* (Énergoatomizdat, Moscow, 1994).
5. W. Ebeling, A. Förster, V. Fortov, V. Gryaznov, and A. Polishchuk, *Thermophysical Properties of Hot Dense Plasmas* (Teubner, Stuttgart, 1991).
6. G. É. Norman and A. N. Starostin, Teplofiz. Vys. Temp. **8**, 413 (1970).
7. K. Zeeger, *Semiconductor Physics* (Springer, Berlin, 1974; Mir, Moscow, 1977).
8. Ya. B. Zel'dovich and Yu. P. Raĭzer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Fizmatgiz, Moscow, 1963; Academic, New York, 1966 and 1967), Vols. 1 and 2.
9. V. E. Fortov, Yu. V. Ivanov, A. N. Dremin, *et al.*, Dokl. Akad. Nauk SSSR **221**, 1307 (1975) [Sov. Phys. Dokl. **20**, 295 (1975)].
10. V. B. Mintsev and V. E. Fortov, Teplofiz. Vys. Temp. **20**, 745 (1982).
11. V. K. Gryaznov, I. L. Iosilevskiĭ, and V. E. Fortov, Prikl. Mekh. Tekh. Fiz. **3**, 70 (1973).
12. Yu. V. Ivanov, V. E. Fortov, V. B. Mintsev, and A. N. Dremin, Zh. Éksp. Teor. Fiz. **71**, 216 (1976) [Sov. Phys. JETP **44**, 112 (1976)].
13. V. B. Mintsev, V. E. Fortov, and V. K. Gryaznov, Zh. Éksp. Teor. Fiz. **79**, 116 (1980) [Sov. Phys. JETP **52**, 59 (1980)].
14. *A Brief Handbook of Physicochemical Values*, Ed. by K. P. Mishchenko and A. A. Ravdel' (Khimiya, Leningrad, 1972).
15. E. V. Kuchis, *Methods for Studying the Hall Effect* (Sovetskoe Radio, Moscow, 1974).
16. P. S. Kireev, *Physics of Semiconductors* (Vysshaya Shkola, Moscow, 1969).

17. V. K. Gryaznov, M. V. Zhernokletov, V. N. Zubarev, *et al.*, Zh. Éksp. Teor. Fiz. **78**, 573 (1980) [Sov. Phys. JETP **51**, 288 (1980)].
18. S. V. Dudin, V. E. Fortov, V. K. Gryaznov, *et al.*, in *Shock Compression of Condensed Matter*, Ed. by S. C. Schmidt, D. P. Dankedar, and J. W. Forbes (AIP Press, New York, 1997), p. 793.
19. S. V. Dudin, V. B. Mintsev, A. E. Ushnurtsev, *et al.*, in *Megagauss and Megaamper Pulsed Technology and Applications* (VNIIÉF, Sarov, 1993), p. 733.
20. A. P. Ershov, P. I. Zubkov, Yu. N. Il'yanovich, *et al.*, in *Proceedings of Third International Conference on Megagauss Magnetic Field Generation and Related Topics* (Nauka, Moscow, 1984), p. 397.
21. N. S. Shilkin, S. V. Dudin, V. K. Gryaznov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 582 (2003) [JETP Lett. **77**, 486 (2003)].
22. V. K. Gryaznov, I. L. Iosilevskiĭ, Yu. G. Krasikov, *et al.*, in *Thermal Properties of Working Media of Gas-Phase Nuclear Reactor*, Ed. by V. M. Ievlev (Atomizdat, Moscow, 1980), p. 301.
23. V. K. Gryaznov, M. V. Zhernokletov, I. L. Iosilevskiĭ, *et al.*, Zh. Éksp. Teor. Fiz. **114**, 1242 (1998) [JETP **87**, 678 (1998)].
24. A. A. Likalter, Zh. Éksp. Teor. Fiz. **56**, 240 (1969) [Sov. Phys. JETP **29**, 133 (1969)].
25. V. E. Fortov, V. K. Gryaznov, V. B. Mintsev, *et al.*, Contrib. Plasma Phys. **41**, 215 (2001).
26. B. M. Askerov, *Kinetic Effects in Semiconductors* (Nauka, Leningrad, 1970).
27. L. Spitzer and R. Härm, Phys. Rev. **89**, 977 (1953).
28. L. S. Frost and A. V. Phelps, Phys. Rev., Sect. A **136**, 1538 (1964).
29. L. S. Frost, J. Appl. Phys. **32**, 2029 (1961).
30. S. C. Lin, E. L. Resler, and A. Kantrowitz, J. Appl. Phys. **26** (1) (1955); Vopr. Raket. Tekh. **1** (31), 13 (1956).
31. V. B. Mintsev, Tr. Mosk. Fiz.–Tekh. Inst., Ser.: Obshch. Mol. Fiz. **10**, 192 (1978).
32. H. A. Gould and H. E. deWitt, Phys. Rev. **155**, 68 (1967).
33. R. Redmer, Phys. Rep. **282**, 35 (1997).
34. R. Redmer, Phys. Chem. **204**, 135 (1998).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 932–947.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 1041–1057.
Original Russian Text Copyright © 2003 by Gus'kov, Rozanov, Cherkasov.

Formation of the Charge and Energy Distributions of Heavy Ions in a Material in the Diffusion Model

S. Yu. Gus'kov, V. B. Rozanov, and A. N. Cherkasov*

Department of Quantum Radiophysics, Lebedev Physical Institute, Russian Academy of Sciences,
Leninskii pr. 53, Moscow, 119991 Russia
**e-mail: al_cherkasov@rambler.ru*

Received October 31, 2002

Abstract—We investigate the formation of the charge and energy distributions of ions that slow down and randomly change their charges in collisions with particles of the medium. We study the influence that the spread of ions in charge has on the shape of the Bragg curve. The suggested diffusion approximation for the kinetic equation of heavy ions allows the parameters of the ion charge and energy distributions to be easily determined in the entire ion path. The parameters of the ion charge distribution are shown to be related to the ionization–recombination cross sections. The ion distributions calculated in the suggested analytical model are compared with the results of numerical calculations. We have obtained good agreement between the analytical, numerical, and experimental results. *© 2003 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Interest in the slowing down of ions stems from the possible accomplishment of the thermonuclear fusion chain reaction using beams of heavy ions. Being oriented to this problem, we will consider the slowing-down of high-energy ions of heavy elements in a material that can be both weakly ionized and almost completely ionized. Below, we will not be concerned with the behavior of ions as they slow down to velocities comparable to the velocities of the material particles, because the ions almost completely lose their energy in this case.

The subject of charged particle slowing down in a material has a long history, beginning with the studies by Bethe and Bohr. Subsequently, the slowing-down theory was significantly improved, but attention to how the charge distribution of ions as they slow down influences the energy losses has arisen only recently. To some extent, the influence of a change in charge on the slowing-down of ions was taken into account in [1–5]. The authors of these papers considered the change in the charge of an ion during its collision with plasma particles as a random transition between a few charge states. Their exact solutions are valid only for several special cases and are inapplicable to the slowing-down of highly charged ions in a material.

The suggested approximate method, which is based on the equation that describes the kinetics of changes in the charges and energies of heavy ions, is applicable to ions of heavy elements with a large number of charge states. This paper has the following structure. In Section 2, we consider the dynamic equations for an ion beam in the charge moment and diffusion approximations. In Section 3, we analyze the analytical ion charge and energy dependences of the cross sections for the processes and compare them with the numerical calculations and experimental data. Approximate analytical solutions, equations for the mean charge and the variance of the charge distribution, and the variance of the fast ion energy distribution are considered in Sections 4 and 5. In Section 6, we compare our results with the numerical calculations and published experimental data.

2. DYNAMIC EQUATIONS FOR THE SLOWING-DOWN AND CHANGE IN THE CHARGES OF HEAVY IONS

The problem of heavy ions slowing-down leads to a system of coupled kinetic equations for the flux densities of ions in different charge states Z:

$$\begin{gathered}\frac{\partial N_Z}{\partial l}-\frac{\partial(S_Z N_Z)}{\partial E} = -(\sigma_i(Z)+\sigma_r(Z))N_Z \\ +\sigma_i(Z-1)N_{Z-1}+\sigma_r(Z+1)N_{Z+1}, \\ Z = 0, 1, 2, \ldots, Z_1, \quad \sigma_i(Z_1) = \sigma_r(0) = 0.\end{gathered} \tag{1}$$

Here,

$$l = \int_0^x n_e(x')dx', \tag{2}$$

$$S_Z = -\frac{\partial E}{\partial l} \tag{3}$$

is the stopping power of the material that depends on the ion charge, n_e is the total electron density of the medium (including bound electrons), and σ_i and σ_r are the ionization and recombination cross sections that were reduced to the electron density of the medium by using the formula

$$\sigma = \frac{\alpha}{n_e v}, \tag{4}$$

where α is the ionization or recombination rate, and v is the ion velocity relative to the particles of the medium. The condition for a zero ionization cross section of a completely stripped atom is obvious; a zero recombination cross section for an ion with zero charge, i.e., a neutral atom, is required to exclude from consideration the negative ions whose formation probability in the present problem is extremely low.

Below, to analyze the equations, we will use a special system of units in which the unit of energy is the maximum ion energy and the unit of mean free path l is a value close to the total mean free path of the ion with a maximum energy and charge equal to the nuclear charge:

$$\xi = \frac{l}{l_0}, \quad \varepsilon = \frac{E}{E_0}, \tag{5}$$

where E_0 is the maximum energy (the initial energy for a monochromatic ion beam),

$$l_0 = \frac{m_e}{M}\frac{E_0^2}{2\pi Z_1^2 e^4 L(Z_1, E_0)} \tag{6}$$

(Z_1 is the ion nuclear charge, M is the ion mass, and L is the Coulomb logarithm of the slowing-down). In this system of units, the stopping power of the material has a very simple form (we restrict our analysis to the nonrelativistic case, which is justified up to energies of ~100 MeV/nucleon):

$$s_Z(\varepsilon) = \frac{l_0 S_Z}{E_0} = \frac{Z^2}{Z_1^2}\frac{1}{\varepsilon}\frac{L(Z, \varepsilon)}{L(Z_1, \varepsilon = 1)}. \tag{7}$$

Formulas for the stopping power were derived long ago by Bethe [6] and Bohr [7] from the linearized electrodynamic equations for a moving particle. A simple approximate formula for the energy losses can be easily derived by combining the formulas by Bethe and Bohr into a single expression:

$$\begin{gathered}-\frac{dE}{dl} = \frac{4\pi Z_2 Z^2 e^4}{m v^2} \\ \times\left((1-\kappa)\ln\frac{2mv^2}{\bar{I}} + \kappa\ln\frac{mv^3}{Ze^2\omega_p}\right).\end{gathered} \tag{8}$$

Here, $\bar{I}$ is the mean excitation energy of the atoms and ions of the medium, ω_p is the plasma frequency, and κ is the ionization coefficient of the medium. Formula (8) has many relativistic and nonrelativistic corrections (see, e.g., [8–10]). For the subsequent analysis, the exact form of the Coulomb logarithm in formula (8) is unimportant. We only note that the stopping power does not increase infinitely as the ion velocity tends to zero, $v \longrightarrow 0$ (see [10, 11]), but dependence (7) remains valid down to low energies for the case under consideration.

Let us rewrite Eqs. (1) in the new notation:

$$\begin{gathered}\frac{\partial N_Z}{\partial \xi} - \frac{\partial(s_Z N_Z)}{\partial \varepsilon} = -(\tilde{\sigma}_i(Z) + \tilde{\sigma}_r(Z))N_Z \\ + \tilde{\sigma}_i(Z-1)N_{Z-1} + \tilde{\sigma}_r(Z+1)N_{Z+1}, \\ Z = 0, 1, \ldots, Z_1,\end{gathered} \tag{9}$$

where

$$\tilde{\sigma}_i = l_0\sigma_i, \quad \tilde{\sigma}_r = l_0\sigma_r.$$

Below, we will omit the tilde over the dimensionless cross sections $\tilde{\sigma}_i$ and $\tilde{\sigma}_r$. The dimensionless ionization and recombination cross sections show the relative rate of change in charge compared to the ion slowing-down.

In general, the slowing-down of an ion may be considered to be a slow process against the background of which fast random changes in charge occur; these changes force the ion to acquire an equilibrium (at a given velocity relative to the medium) charge state. Some of the empirical formulas that describe the dependence of the equilibrium charge of an ion in a cold material on its velocity will be considered at the end of the next section.

This picture, which is always valid for cold targets, can change for a strongly ionized material [12]. The most effective recombination process, the recombination via electron–ion exchange in the medium, is absent when ions of heavy elements with a high charge state and a high energy slow down in strongly ionized plasma. A sharp reduction in recombination causes the equilibrium charge to shift to the range of high values, and the ionization–recombination cross sections near the equilibrium charge prove to be much smaller than those for the motion of an ion in a cold material. As a result, all terms of the kinetic equation become of the same order of magnitude.

Let us obtain the first approximation to the solution of system (9). We add Eqs. (9) with different Z:

$$\frac{\partial n}{\partial \xi} - \sum_Z \frac{\partial s_Z N_Z}{\partial \varepsilon} = 0, \quad n = \sum_Z N_Z. \tag{10}$$

We assume that the ion spread in charge is negligible, so only ions with a mean charge $\bar{z}$ (below, z denotes a continuous approximation of the discrete variable Z) may be considered. In this case, the equation reduces to the well-known approximation of continuous slowing-down. In this approximation, we obtain the following system together with the equation for the mean charge:

$$\begin{gathered}\frac{\partial n}{\partial \xi} - \frac{\partial s n}{\partial \varepsilon} = 0, \\ \frac{\partial \bar{z}}{\partial \xi} - s\frac{\partial \bar{z}}{\partial \varepsilon} = \sigma_i(\bar{z}, \varepsilon) - \sigma_r(\bar{z}, \varepsilon).\end{gathered} \tag{11}$$

Solving Eqs. (11) yields the first approximation to the solution of problem (9) in which the ion energy is uniquely related to the coordinate ξ:

$$\begin{gathered}n = \frac{1}{s(\bar{z}, \varepsilon)}\delta\left(\xi - \int_{\varepsilon}^{1}\frac{1}{s(\bar{z}, \varepsilon')}d\varepsilon'\right), \\ \bar{z} = z_0 + \int_{\varepsilon}^{1}\frac{1}{s(\bar{z}, \varepsilon')}(\sigma_i(\bar{z}, \varepsilon') - \sigma_r(\bar{z}, \varepsilon'))d\varepsilon'.\end{gathered} \tag{12}$$

For ions of fixed charge, for example, for the completely stripped nuclei of light elements, more accurate kinetic models and their solutions can be found in [13]. For the ions of heavy elements of interest to us, the most important mechanism responsible for the deviation of the solution from expressions (12) is the spread of ions in the charge state and the related spread in energy.

Let us consider the possibility of using the method of charge moments to obtain the next (in order of accuracy) solutions of the problem. In addition to the mean charge, we introduce the second central moment of the charge distribution or the charge variance D:

$$D = \frac{1}{n}\sum_Z Z^2 N_Z - \bar{Z}^2. \tag{13}$$

We also take into account the explicit dependence of s_Z on Z. The variables in the function $s(Z, \varepsilon)$ are assumed to be separable. This is a realistic assumption (see formula (7)). According to this assumption, we write

$$s(Z, \varepsilon) = Z^2 s_1(\varepsilon) \tag{14}$$

and change the variable

$$\eta = \int_0^{\varepsilon}\frac{d\varepsilon'}{s_1(\varepsilon')}; \tag{15}$$

using (7), we obtain

$$\eta \approx Z_1^2\varepsilon^2/2.$$

We also introduce an ion flux density that also depends on η:

$$\bar{N}(\xi, \eta) = N(\xi, \varepsilon)s_1(\varepsilon), \tag{16}$$

where $\bar{N}$, just as N, is assumed to be normalized to unity:

$$\int \bar{N}(\xi = 0, \eta)d\eta = 1.$$

Given these transformations, system (9) takes the form

$$\begin{gathered}\frac{\partial N_Z}{\partial \xi} - Z^2\frac{\partial N_Z}{\partial \eta} = -(\sigma_i(Z) + \sigma_r(Z))N_Z \\ + \sigma_i(Z-1)N_{Z-1} + \sigma_r(Z+1)N_{Z+1}.\end{gathered} \tag{17}$$

Summing Eqs. (17) over Z, multiplying them by Z and Z^2, and adding them yields the following equations for the moments:

$$\begin{gathered}\frac{\partial n}{\partial \xi} - \frac{\partial}{\partial \eta}((\bar{Z}^2 + D)n) = 0, \\ \frac{\partial \bar{Z}n}{\partial \xi} - \frac{\partial}{\partial \eta}(\bar{Z}(\bar{Z}^2 + 3D)n) = \sum_Z(\sigma_i - \sigma_r)N_Z, \\ \frac{\partial((\bar{Z}^2 + D)n)}{\partial \xi} - \frac{\partial}{\partial \eta}(\bar{Z}^2(\bar{Z}^2 + 6D)n) \\ = 2\sum_Z Z(\sigma_i - \sigma_r)N_Z + \sum_Z(\sigma_i + \sigma_r)N_Z,\end{gathered} \tag{18}$$

where

$$\bar{Z} = \frac{1}{n}\sum_Z Z N_Z$$

and we use the expressions that follow from statistics and that are valid at least when $D \ll \bar{Z}^2$:

$$\begin{gathered}\sum_Z Z^3 N_Z = \bar{Z}(\bar{Z}^2 + 3D)n, \\ \sum_Z Z^4 N_Z = \bar{Z}^2(\bar{Z}^2 + 6D)n.\end{gathered} \tag{19}$$

To close the moment equations, some assumptions regarding the charge dependences of the ionization and recombination cross sections should be made to close the moment equations. We assume that the range of change in variables important for the solution of the problem, more specifically, near the mean charge at a given ion energy, the cross sections can be fitted by a linear function (see Fig. 1), so

$$\begin{gathered}\sigma_i(z,\varepsilon)+\sigma_r(z,\varepsilon) = a(\varepsilon),\\ \sigma_i(z,\varepsilon)-\sigma_r(z,\varepsilon) = b(\varepsilon)(q(\varepsilon)-z),\\ a>0,\quad b>0.\end{gathered}\tag{20}$$

In this case, we managed to close Eqs. (18) by writing

$$\begin{gathered}\frac{\partial n}{\partial\xi}-\frac{\partial}{\partial\eta}((\bar{Z}^2+D)n) = 0,\\ \frac{\partial\bar{Z}n}{\partial\xi}-\frac{\partial}{\partial\eta}(\bar{Z}(\bar{Z}^2+3D)n) = b(\varepsilon)(q(\varepsilon)-\bar{Z})n,\\ \frac{\partial((\bar{Z}^2+D)n)}{\partial\xi}-\frac{\partial}{\partial\eta}(\bar{Z}^2(\bar{Z}^2+6D)n)\\ = 2b(\varepsilon)(q(\varepsilon)\bar{Z}-\bar{Z}^2-D)n+a(\varepsilon)n.\end{gathered}\tag{21}$$

Equations (21) are a closed system of partial differential equations to determine the charge moments. For this system to be unambiguously solved, we must specify initial conditions, for example, the condition for a monochromatic, singly charged ion beam:

$$\begin{gathered}n(\xi=0,\eta) = \delta(\eta),\quad \bar{Z}(\xi=0,\eta) = Z_0,\\ D(\xi=0,\eta) = 0,\end{gathered}\tag{22}$$

where $\delta(\eta)$ is the delta function. Since the equations are linear in variables ξ and η, the solutions of the problem with an arbitrary initial condition can be obtained from the solutions of the problems with initial conditions of the form (22).

Let us consider the influence of particular terms in the kinetic equation on the solution of the problem. To this end, we pass to a continuous charge variable in Eq. (9). Assuming the ionization and recombination cross sections to be smooth in the charge variable, we can use a Taylor expansion on the right-hand side of the equation. Retaining the terms to the second order inclusive, we obtain the equation

$$\frac{\partial N}{\partial\xi}-\frac{\partial(sN)}{\partial\varepsilon} = -\frac{\partial}{\partial z}(\beta_a N)+\frac{\partial^2}{\partial z^2}(\beta_t N),\tag{23}$$

where

$$\beta_a = \sigma_i-\sigma_r,\quad \beta_t = \frac{1}{2}(\sigma_i+\sigma_r).\tag{24}$$

The function $\beta_t(\varepsilon, z)$ is positive and has a charge minimum that lies near the equilibrium charge. Recall that the equilibrium charge is determined by the equality between the ionization and recombination cross sections at a given ion charge. In contrast, the function $\beta_a(\varepsilon, z)$ changes sign as the charge passes through its equilibrium value. The form of the solution to the kinetic equation is directly related to these singularities in the charge dependence of its coefficients. An example of the charge dependences of coefficients β_a and β_t is shown in Fig. 1.

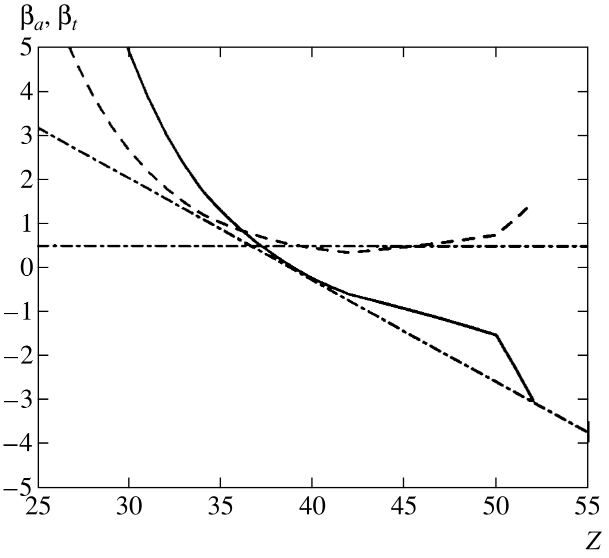


Fig. 1. An example of the dependence of coefficients β_a (solid curve) and β_t (dashed curve) on ion charge (1.5 MeV/u iodine ions in a hydrogen plasma). The dash-dotted straight lines are the linear fits to the cross sections near the equilibrium charge given by (20). In the range $q-2<z<q+4$, the error of these first does not exceed 30%.

The initial and boundary conditions for Eq. (23) are

$$\begin{gathered}N(\xi=0,\varepsilon,z) = \phi(\varepsilon,z),\\ \beta_a N-\frac{\partial}{\partial z}(\beta_t N)\Big|_{Z=Z_1,0} = 0,\end{gathered}\tag{25}$$

where $\phi(\varepsilon, z)$ is the function that describes the energy and charge distribution of ions in the beam before their interaction with the medium.

Equation (23) is the charge diffusion equation with a gathering force. Its solution is qualitatively shown in Fig. 2. We will return to the solutions of Eqs. (21) and (23) after considering the singularities of their coefficients.

3. IONIZATION AND RECOMBINATION CROSS SECTIONS

The main difficulty in studying Eq. (9) is that its coefficients are functions of both ion charge and energy. Below, we attempt to find suitable fits for the coefficients of this equation, which, in particular, will help justify the transformations made in the previous section.

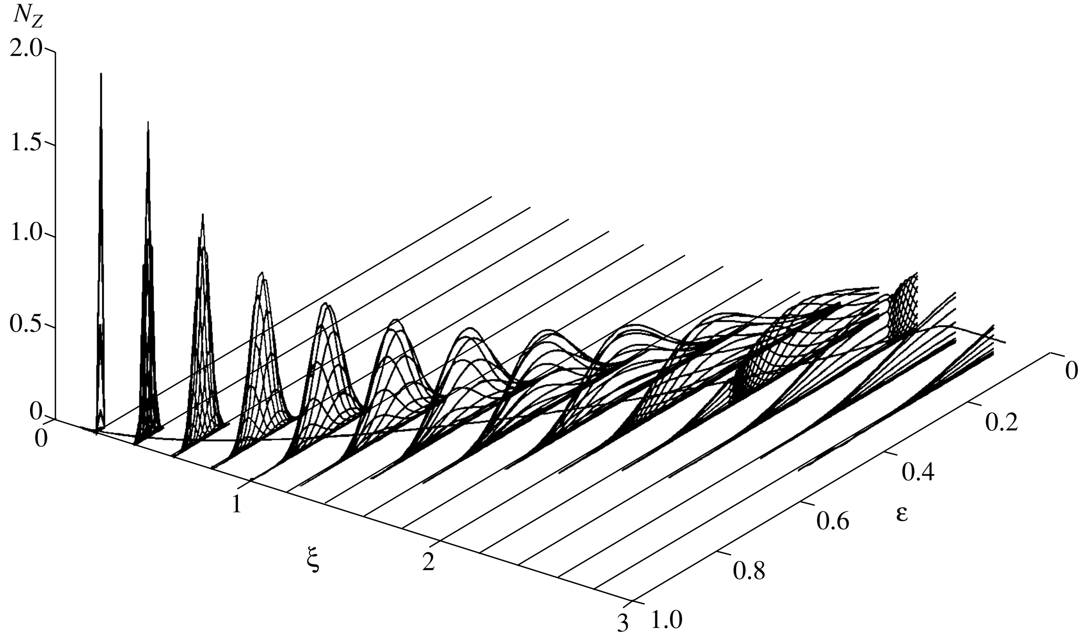


Fig. 2. The qualitative energy and depth distribution of ions with different charges. A monochromatic and singly charged ion beam is at the entrance.

Note that in our formation of the problem, exact (taken from a good theory or from experimental data) cross sections for the processes can be used to solve Eq. (9). However, "smooth" formulas, including interpolation ones, that can satisfactorily fit the behavior of the real cross sections are required to solve Eq. (23).

Below, we give simple analytical expressions for the ionization and recombination cross sections. For the electron impact ionization cross section, the Thomson ionization cross section may serve as the first approximation. For the ionization of the target ion by an electron accompanied by the electron detachment from the *n*th level, the cross section is

$$\sigma_{ne} = \pi\frac{\varepsilon - U_n}{U_n\varepsilon^2},$$

where $\varepsilon = v^2/2$, and v is the relative particle velocity; according to our assumptions, the latter is equal to the ion velocity. This and other formulas of this section were written by using the atomic system of units: the unit of cross sections is $\hbar^4/m^2e^4 = 2.8 \times 10^{-17}$ cm^2.

For a collision with a plasma ion, the Thomson ionization cross section is

$$\sigma_{ni} = \pi Z_t^2\frac{4\varepsilon - U_n}{4U_n\varepsilon^2},$$

where Z_t is the effective charge of the target ion. It follows from a comparison of these formulas that the main ionization mechanism is a collision with ions or atoms of the material. Even if the material is not ionized, but the energies are high, the effective target ion charge Z_t is generally larger than unity (except the case where the target is hydrogen).

The semiempirical formula by Lotz [20] more accurately describes the ionization cross sections during collisions with target electrons:

$$\sigma_{ne} = 2.17\frac{\ln(\varepsilon/U_n)}{U_n\varepsilon}. \tag{26}$$

The Lotz cross section differs from the Thomson cross section by no more than a factor of 2, as illustrated by Fig. 3.

According to the classical theory, the ionization cross section during the collision of a fast ion with a target ion is described by the formula

$$\sigma_{ni} = \frac{\pi Z_t^2}{U_n^2}G\left(\frac{\varepsilon}{U_n}\right), \tag{27}$$

where $G(x)$ was derived in [14, 15] from classical models for collisions of an external electron with atomic electrons. The ionization cross sections for an iodine ion in a hydrogen plasma calculated using these formulas are plotted in Fig. 3 (curves I). Even more accurate methods for calculating the ionization cross sections based on quantum-mechanical models can be found in [16].

Consider the recombination cross section. Let us write the cross sections for the main processes that lead to the capture of an electron by a moving ion: the cross sections for radiative capture, electron–ion exchange, and three-particle recombination.

Radiative capture is the main recombination process for completely ionized plasma of moderately high density. For the radiative capture cross section, we use the Kramers classical formula

$$\sigma_{nR} = A\frac{U_n^2}{\varepsilon\varepsilon_\gamma}\frac{1}{n}, \tag{28}$$

where

$$A = 7.5\times10^{-6}a_B^2 = 2.1\times10^{-22}\ \text{cm}^2,$$
$$\varepsilon_\gamma = \varepsilon + U_n,$$

and U_n in our case corresponds to unoccupied levels. Assuming that $v^2 > z^2/n^2$, integrating the cross section over n, and assuming that the last occupied level is

$$n_g = (Z_1 - Z)^{1/3},$$

we obtain the following simple expression for the radiative capture cross section:

$$\sigma_R = \frac{A}{8\varepsilon^2}\frac{Z^4}{(Z_1 - Z)^{2/3}}. \tag{29}$$

If we restrict our analysis to the cross section for recombination to the ground level, then we will derive an expression that differs only slightly from the above expression:

$$\sigma_R = \frac{A}{8\varepsilon^2}f(Z), \quad f(Z) = \frac{Z^4}{(Z_1 - Z)}.$$

The three-particle recombination rate can become significant at high target densities. Let us write a formula for the three-particle recombination cross section following Zeldovich and Raizer [17]:

$$\sigma_3 = \frac{\pi^2 Z_1^3 n_e}{\varepsilon^5}. \tag{30}$$

Several empirical formulas for the cross sections can also be found in [18–20].

In a nonionized material or for a incompletely ionized target plasma, the recombination via electron transfer from a target ion to a penetrating ion plays a leading role. We will estimate the cross section for recombination via electron exchange by using the result obtained in the Oppenheimer–Brinkmann–Kramers (OBK) theory [21]. According to this theory, the cross section for the transition of an electron from a state with energy E_i to a state with energy E_f is

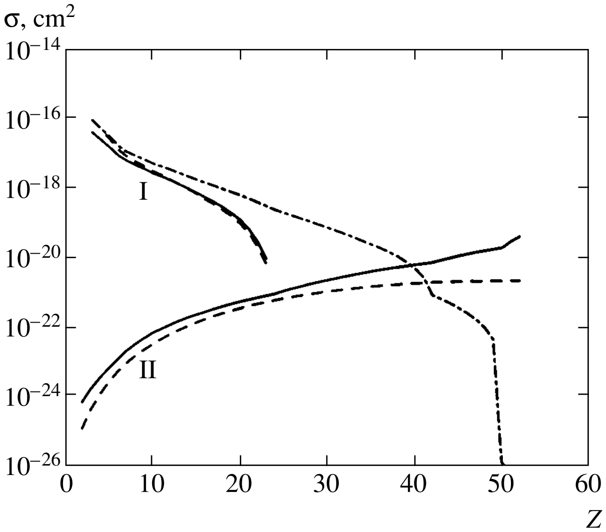


Fig. 3. Ionization (I) and recombination (II) cross sections for iodine ions with energy 1.5 MeV/*u* in a hydrogen plasma with a temperature of 10 eV and a density of 10^{17} cm^{-3} versus charge *Z*. Curves I represent the Thomson (solid curve), Lotz [20] (dashed curve), and Gryzinski [14] (dash-and-dotted curve) cross sections; curves II represent the cross sections for photorecombination (29) (solid curve) and charge exchange (32) (dashed curve). The conditions were taken from [11]. The electron binding energies were calculated in terms of the Thomas–Fermi model for the ground state.

$$\sigma_{OBK}^{i\to f} = 4\times10^4\frac{Z^2\varepsilon^4E_i^{5/2}E_f^{3/2}}{(\varepsilon^2 + 2\varepsilon(E_i + E_f) + (E_i - E_f)^2)^5}. \tag{31}$$

To calculate the charge exchange cross section, we use the correction of the second Born approximation from [22],

$$\sigma_{\text{ct}} = \left(0.3 + \frac{5\pi v}{2^{12}}\right)\sigma_{OBK}. \tag{32}$$

Approximating the binding energy of the captured electron by

$$E_f = \frac{Z^2}{2n^2}$$

and assuming that $n = (Z_1 - Z + 1)^{1/3}$ for the last energy level occupied in the ion, we obtain an approximate expression for the charge exchange cross section at high ion velocities:

$$\sigma_{\text{ct}} = 2.1\times10^3\frac{\sum_n N_nE_{\text{in}}^{5/2}}{\varepsilon^6}\frac{Z^5}{(Z_1 - Z + 1)^{2/3}}. \tag{33}$$

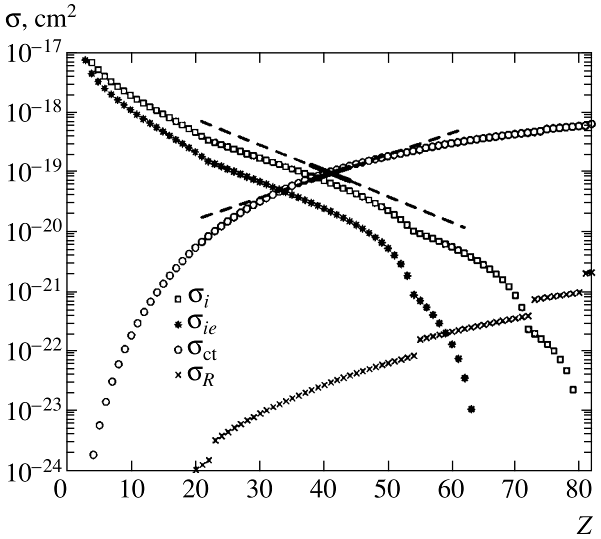


Fig. 4. Calculated ionization and recombination cross sections for a lead ion with energy 11.6 MeV/*u* in the plasma of a capillary discharge in hydrocarbon with a temperature of 3.3 eV. The dotted lines represent the exponential fits to the ionization and recombination cross sections near the equilibrium charge.

This formula is suitable for fitting the charge exchange cross section only for high energies of the moving ion. The cross sections calculated using Brinkmann–Kramers formula (31) yield correct dependences on the ion charge and energy, but they unsatisfactorily describe the experimental data. We note once again that we do not seek to derive exact formulas for the cross sections. Thus, a more accurate theory [23, 24] or the following semiempirical formula suggested in [25] for nonionized targets can be used for the charge exchange cross section:

$$\sigma_{\rm ct} = 1.1\times10^{-8}\ \text{cm}^2\ Z^{0.5}/Z_2^{1.8}u^{-4.8} \times[1-\exp(-0.037u^{2.2})][1-\exp(-2.44\times10^{-5}u^{2.6})], \quad (34)$$

$$u = E[\text{keV/amu}]/(Z^{0.7}Z_2^{1.25}), \quad u\ge 10, \quad Z\ge 3.$$

Formulas for the empirical charge exchange cross sections can also be found in [26].

To write the interpolation formulas of this section, we used hydrogen-like energy levels. The binding energies calculated in this way are lower than their true values. This leads primarily to an underestimation of the recombination cross sections because of their sharp dependence on bound electron energy, which, in turn, is responsible for the low equilibrium ion charge calculated in this way and can significantly affect the accuracy of calculating the parameters of the plasma ion distribution function. To avoid this, we must use more accurate binding energies of the electrons in a moving ion, for example, those obtained in the Thomas–Fermi model.

The cross sections with the energy levels calculated in the Thomas–Fermi model are plotted in Fig. 3.

Note also that the cross sections σ_i and σ_r are conveniently fitted by exponentials of the form

$$\sigma_i = \sigma_{i0}\exp(-a_i z), \quad \sigma_r = \sigma_{r0}\exp(a_r z). \quad (35)$$

The exponential fits to the ionization and recombination cross sections are shown in Fig. 4. We used the exponential dependences in our test calculations of the slowing-down of a beam of heavy ions in a material.

To conclude this section, let us consider an important question concerning the equilibrium charge *q*. The equilibrium charge is determined by the equality between the ionization and recombination cross sections, and the formulas given in this section can be used to determine it. For example, for an almost completely ionized target, the ionization cross section (27) should be equated to the radiative recombination cross section (28), while for a cold material, the recombination will be determined by charge exchange and can be calculated using formula (32). The formulas of this section allow an ion energy dependence of the equilibrium charge *q* to be established. In addition, they can be compared with published empirical formulas, which are used mainly for a cold target material. Thus, by comparing the analytical expressions for the stopping power of a material derived in [27] with experimental data, Basko [9] suggested the following formula for the equilibrium charge:

$$q = Z_1\left(1+\left(0.62Z_1^{2/3}\frac{1}{v}\right)^{1.7}\right)^{-1/1.7}, \quad (36)$$

while Brown and Moak [28] suggested the expression

$$q = Z_1\left(1-a\exp\left(\frac{-v}{Z_1^{2/3}}\right)\right), \quad (37)$$

where $a \approx 1$. Based on formula (37) and the Bohr rule for the effective ionization energy, Armel and Funkhouser [29] derived an expression for the successive ionization potentials that matched the formulas following from the Thomson–Fermi model on the asymptotics.

Figure 5 shows the equilibrium charge calculated using semiempirical formulas and, for comparison, using the formulas for the ionization and recombination cross sections from this section. As an example, we again consider an iodine ion in hydrogen, in completely ionized (solid curve in Fig. 5) and cold (dashed curve) materials. In addition, this figure shows dependences (36) (dash-dotted curve) and (37) (dotted curve). We see from the figure that there is a range of conditions where the differences between the formulas are large, although the empirical curves are within the boundaries defined by the formulas for the cross sections in plasma and a cold material. Formula (37) is preferred for a cold target, while formula (36) is suitable for calculating the equilibrium charge for a strongly ionized target.

However, in the entire range of states for the target material, explicit formulas similar to (36) or (37) most likely cannot be constructed and the only method of determining the equilibrium charge is to equate the corresponding ionization and recombination cross sections.

4. APPROXIMATE ANALYTICAL SOLUTIONS

Let us consider Eq. (23) by taking into account the singularities of its coefficients (see Section 2).

Equation (23) is an analog of the diffusion equation and actually contains the coordinate and energy transfer and charge diffusion terms. It reduces to the standard diffusion equation if the charge remains near its equilibrium value and if the variance of the charge distribution is so small that the term with β_a may be disregarded. In this case, the dispersion of the charge distribution must satisfy the condition

$$\beta_a(q \pm \sqrt{D}) \ll \beta_t. \tag{38}$$

Consider the simplest approximation that allows the parameters of the ion charge distribution to be determined. We change variables:

$$\varepsilon = \varepsilon,$$

$$u = \xi - \int_{\varepsilon}^{1} \frac{d\varepsilon'}{s(\varepsilon', \bar{z}(\varepsilon'))},$$

where $\bar{z}$ is the solution of system (11), i.e., the mean charge. The equation in these variables can be written as

$$\begin{aligned} &\frac{\partial N}{\partial u}\left(1 - \frac{s(\varepsilon, z)}{s(\varepsilon, \bar{z}(\varepsilon))}\right) - \frac{\partial (sN)}{\partial \varepsilon} \\ &= -\frac{\partial}{\partial z}(\beta_a N) + \frac{\partial^2}{\partial z^2}(\beta_t N). \end{aligned} \tag{39}$$

If we disregard the charge dependence, i.e., assume that $s(\varepsilon, z) = s(\varepsilon, \bar{z}(\varepsilon))$, then we obtain the equation

$$-\frac{\partial (sN)}{\partial \varepsilon} = -\frac{\partial}{\partial z}(\beta_a N) + \frac{\partial^2}{\partial z^2}(\beta_t N). \tag{40}$$

Under certain conditions, the term related to the slowing-down of ions may be ignored in Eq. (23). This is appropriate for large values of the sum of the ionization and recombination cross sections, $\beta_t \ll 1$, and corresponds to a situation in which many collisions take place in a segment of the path where the slowing-down is small. In particular, it can be shown that at the onset of slowing-down, if the ion charge differs significantly from the equilibrium charge,

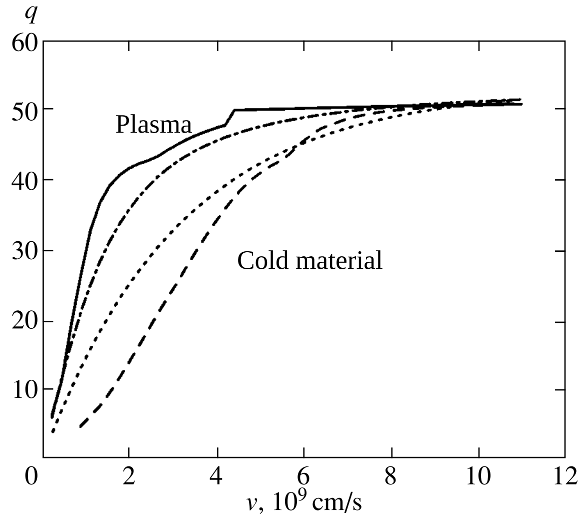


Fig. 5. Equilibrium charge of an iodine ion in hydrogen versus velocity. The equilibrium charge was calculated from the equality between the ionization and recombination cross sections using the formulas in the text for a completely ionized target (solid curve) and a cold material (dashed curve) and calculated using formulas (36) (dash-dotted curve) and (37) (dotted curve).

$$\beta_a \approx \frac{M}{4\pi z_1^2 m} \approx 10^2.$$

An analytical solution for the resulting equation can be found by using suitable fits for the cross sections. Thus, for example, let us assume that the following linear fits (see Fig. 1) can be used for the coefficients β_a and β_t near the equilibrium charge:

$$\begin{aligned} \beta_a &= b(q - z), \\ \beta_t &= a, \end{aligned} \tag{41}$$

where q is the equilibrium charge at a given ion velocity. In this case, Eq. (23) takes the form

$$\frac{\partial N}{\partial \xi} + b\frac{\partial}{\partial z}((z - q)N) - a\frac{\partial^2 N}{\partial z^2} = 0 \tag{42}$$

and its solution can be written as

$$N(\xi, z) = \frac{1}{\sqrt{2\pi D}}\exp\left(-\frac{(z - \bar{z})^2}{2D}\right), \tag{43}$$

where

$$\frac{dz}{d\xi} = \beta_a(\bar{z}) = b(q - z),$$
$$\frac{dD}{d\xi} + 2bD = 2a. \tag{44}$$

Thus, we see that the charge dependence of the ion flux density is roughly Gaussian with a variance whose increase, according to (44), is limited by the derivative of the coefficient β_a with respect to charge. Note that this term always prevents the charge variance from increasing, because the derivative $\partial\beta_a/\partial z$ is always negative. The maximum charge dispersion can easily be estimated from the second equation in (44) (we assume that the initial variance does not exceed its limiting value),

$$D_{\max} = \frac{a}{b}. \tag{45}$$

Occasionally, when the charge changes by one from its equilibrium value, the ionization or recombination processes begin to sharply dominate. In this case, an estimate from formula (45) gives a variance that is approximately equal to unity. This is in agreement with the following qualitative considerations. An ion with an equilibrium charge can both be ionized and capture an electron with an equal probability. Once the ion has changed its charge, the return of the ion charge to its equilibrium value will be most probable. As a result, no more than three ion charge states will be involved and the variance of the charge distribution does not exceed unity.

Without writing out the general solution of system (44), we will give the solution in the approximation of constant coefficients:

$$z = z_0 e^{-b\xi} + q(1 - e^{-b\xi}), \quad D = \frac{a}{b}(1 - e^{-2b\xi}). \tag{46}$$

Let us compare the mean charge and variance of the charge distribution calculated using these formulas for chlorine ions in cold hydrogen with the experimental data presented in Fig. 9. As we see, these simple formulas describe well the experimental data if the ion charges are close to their equilibrium value, as in our case.

Let us return to the system of equations (21) that we derived by the method of moments. Its solution is significantly simplified by the fact that the ion flux density far from some characteristic curve in the coordinate–energy plane is very low. We introduce the following new coordinates related to this characteristic:

$$dx = d\xi - \frac{1}{\bar{z}^2 + D}d\eta,$$
$$dy = d\xi + \frac{1}{\bar{z}^2 + D}d\eta.$$

In the new variables, the first equation in (21) can be written as

$$2\frac{\partial n}{\partial x} = \frac{n}{\bar{z}^2 + D}\left(-\frac{\partial(\bar{z}^2 + D)}{\partial x} + \frac{\partial(\bar{z}^2 + D)}{\partial y}\right). \tag{47}$$

As we see, it does not contain the derivative with respect to the variable *y*. Since the function $n(\xi, \eta)$ rapidly decreases sideways from the characteristic, we require that

$$\frac{\partial n}{\partial y} = 0. \tag{48}$$

For the variance and the mean charge, we derive the following equations from Eqs. (21) using these expressions, which are valid on the $dy = 0$ characteristic:

$$\begin{gathered}\frac{\partial\bar{z}}{\partial\xi} - (\bar{z}^2 + D)\frac{\partial\bar{z}}{\partial\eta} - \bar{z}\frac{2\bar{z}^2 + D}{\bar{z}^2 + D}\frac{\partial D}{\partial\eta} \\ - b(q - \bar{z}) = 0, \\ \frac{\partial\overline{D}}{\partial\xi} - \frac{\bar{z}^4}{(\bar{z}^2 + D)}\frac{\partial\overline{D}}{\partial\eta} - \bar{z}D\frac{4\bar{z}^2 + 9D}{\bar{z}^2 + D}\frac{\partial\bar{z}}{\partial\eta} \\ + 2bD - 2a = 0.\end{gathered} \tag{49}$$

Since the variables ξ and η are equivalent in the kinetic equation, all parameters of the ion charge distribution must change only along the characteristic. Assuming, based on these considerations, that

$$\frac{\partial\bar{z}}{\partial y} = 0, \quad \frac{\partial D}{\partial y} = 0, \tag{50}$$

we obtain the following system of ordinary differential equations for the mean charge and the variance of the charge distribution along the characteristic:

$$\begin{gathered}\frac{\partial\bar{z}}{\partial x} - \frac{\bar{z}(\bar{z}^2 + (1/2)D)}{(\bar{z}^2 + D)^2}\frac{\partial D}{\partial x} - b(q - \bar{z}) = 0, \\ \frac{\partial D}{\partial x} + \frac{D(\bar{z}^2 + (1/2)D)}{(\bar{z}^2 + D)^2}\frac{\partial\overline{D}}{\partial x} \\ - \bar{z}D\frac{2\bar{z}^2 + (9/2)D}{(\bar{z}^2 + D)^2}\frac{\partial\bar{z}}{\partial x} + 2bD - 2a = 0.\end{gathered} \tag{51}$$

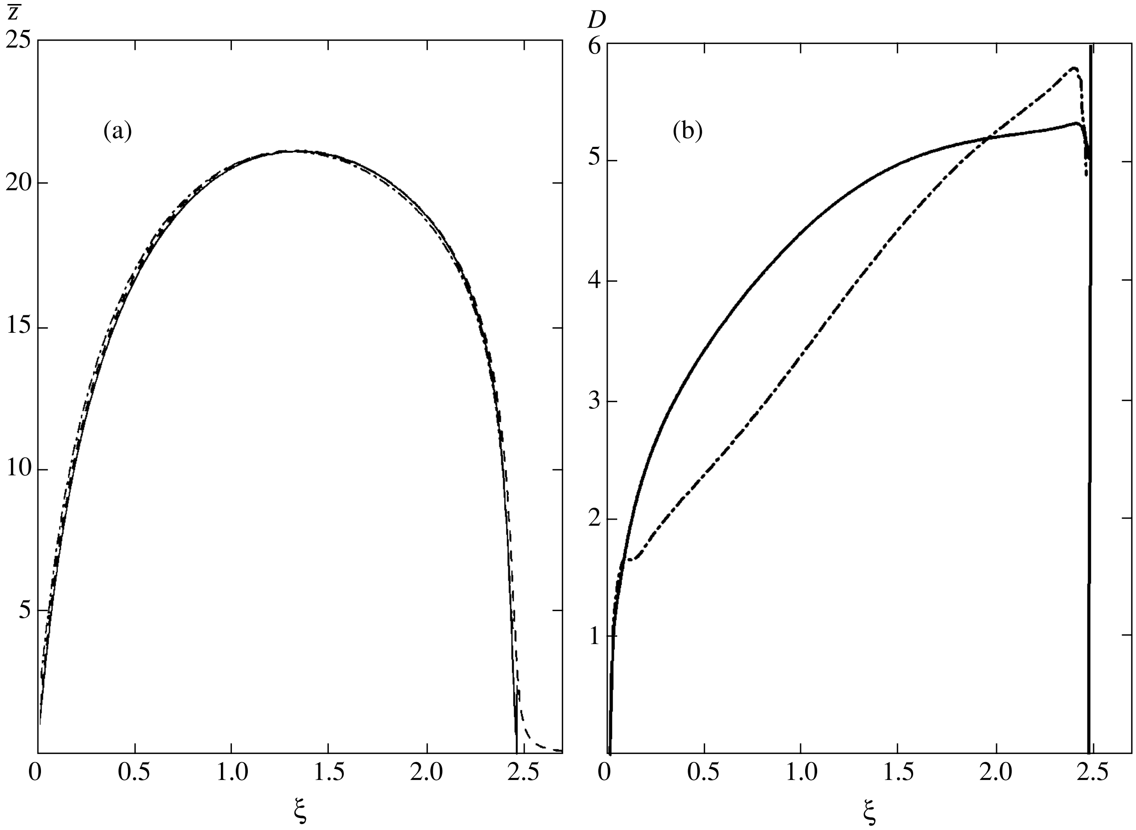


Fig. 6. The test problem. The mean charge (a) and the variance of the charge distribution (b) calculated using formulas (51) (solid line) and obtained from the numerical solution (dash–dotted line). The dashed line in panel (a) is the solution of the problem in the continuous ionization approximation.

As we see from the first equation in (51), the spread of ions in charge affects the mean charge only at the very beginning of the ion path, when the charge variance changes greatly. However, at the beginning of the ion path, this term is still small ($\sim D/\bar{z}$), because the variance is small. Thus, the mean charge along the ion slowing-down characteristic can be calculated in the zero variance approximation (see (44)). The charge variance can then be determined from the second equation in (51).

The charge variance and the mean charge on the characteristic calculated by the method of moments are shown in Fig. 6 in comparison with their values obtained by numerically calculating the test problem. As we see from the figure, the charge variances slightly differ (by no more than 30%). This difference stems from the fact that condition (50) is poorly satisfied at the beginning of the ion path, where the charge differs greatly from its equilibrium value.

5. THE VARIANCE OF THE ENERGY DISTRIBUTION OF FAST IONS

In the model considered above, the distribution of ions with mean charge z and variance D_z is formed at each depth ξ. They can be calculated by using Eqs. (21).

An important physical process that affects the energy variance D_ε is the change in ion charge via collisions with particles of the medium for the energy distribution formed in the previous segment of the path. Because of the change in the charge of ions, ions with charge $\Delta z = \pm 1$ (or even with a larger charge difference) may prove to have the same energy. Recall that the ionization rate exceeds the slowing-down rate by many times (by a factor of $M/4\pi Z_1^2 m$), particularly at the beginning of the ion motion if its charge differs from the equilibrium value.

If there were no change in ion charge, then the energy variance for different charges would be a result of only the initial spread of ions in energy. Because of the change in charge, the accumulated energy variance is approximately the same for different charges, since a large charge that lost much of its energy can become small, causing the energy variance of small charges to increase and vice versa.

Let us consider the evolution of two extreme charges in the ion charge distribution, $z' = \bar{z} - \sqrt{D}$ and

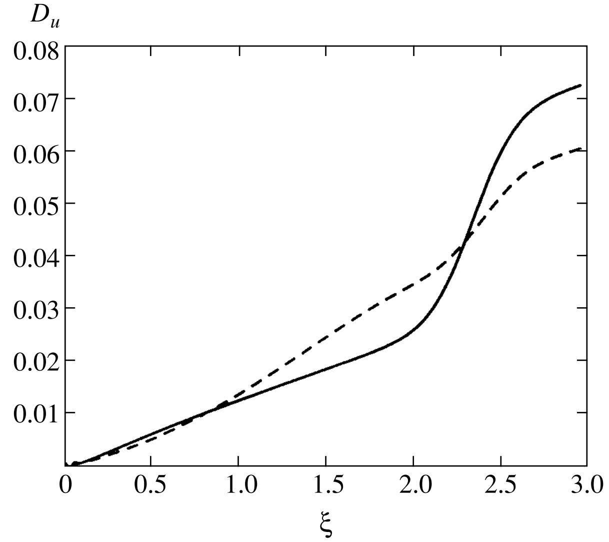


Fig. 7. Comparison of the variances for the variable u determined by numerically solving the kinetic equation (solid line) and calculated using formula (52) (dashed line).

$z'' = \bar{z} + \sqrt{D}$, and the corresponding values of the variance η:

$$\eta' = \eta_0 - \int_0^{\xi} z'^2(\xi)d\xi,$$

$$\eta'' = \eta_0 - z''^2(\xi)d\xi.$$

The difference between these values characterizes the rms deviation of the variable η:

$$\begin{aligned}\eta'' - \eta' &= 2\sqrt{D_\eta} \\ &= \int_0^{\xi}((\bar{z} + \sqrt{D})^2 - (\bar{z} - \sqrt{D})^2)d\xi' = 4\int_0^{\xi}\bar{z}\sqrt{D}d\xi'.\end{aligned} \tag{52}$$

The derived expression relates the rms deviation of the variable η to the rms deviation of the ion charge $\sqrt{D}$. It is valid if the charge variance is small compared to the mean charge. The mean square of the deviation of the ion energy can be related to the variance of the variable D_η as follows (if the higher order moments are disregarded):

$$\begin{aligned}D_\eta &= \overline{\eta^2} - \bar{\eta}^2 = \frac{Z_1^2}{4}(\overline{\varepsilon^4} - (\overline{\varepsilon^2})^2) \\ &= \frac{Z_1^2}{4}(\bar{\varepsilon}^2(\bar{\varepsilon}^2 + 6D_\varepsilon) - (\bar{\varepsilon}^2 + D_\varepsilon)^2) = Z_1^2\bar{\varepsilon}^2 D_\varepsilon.\end{aligned} \tag{53}$$

On short paths, we can derive an approximate formula for the energy variance from this expression and formula (52):

$$\sqrt{D_\varepsilon} = \frac{2}{Z_1\varepsilon}\int_0^{\xi}\bar{z}\sqrt{D}d\xi'. \tag{54}$$

Note the difference between the behaviors of the charge and energy variances. The charge variance rapidly increases at the beginning of the path, because the ionization and recombination rates generally exceed the ion slowing-down rate, but the increase in charge dispersion is limited by (45). In contrast, the energy variance increases over the entire path and becomes significant at its end. Figure 7 shows the energy variance calculated using these formulas in comparison with its numerically calculated value. We see that the derived formulas describe well the dependence of the variance on the path depth.

6. COMPARISON OF THE ANALYTICAL AND NUMERICALLY CALCULATED DISTRIBUTIONS

Let us compare our analytical solutions with the numerical solutions of Eq. (9). To improve the stability of the solution, we pass to the variables ξ and u, where

$$u = \xi - \int_{\varepsilon}^{1}\frac{d\varepsilon'}{s(\bar{z}, \varepsilon')}, \tag{55}$$

and $\bar{z}(\varepsilon)$ can be determined by solving the system of equations

$$\begin{aligned}\frac{d\varepsilon}{d\xi} &= -s(\bar{z}, \varepsilon), \\ \frac{d\bar{z}}{d\xi} &= \sigma_i(\bar{z}, \varepsilon) - \sigma_r(\bar{z}, \varepsilon).\end{aligned} \tag{56}$$

We use this method, because the solution of the ion slowing-down problem in the continuous ionization approximation, i.e., when the charge and energy spreads of ions are disregarded, has the simplest form in these variables. Indeed, in this approximation, the coordinate is uniquely related to the ion energy and the solution is trivial:

$$u = 0.$$

In the variables ξ and u, Eq. (9) takes the form

$$\begin{aligned}\frac{\partial\tilde{N}_Z}{\partial\xi} + \frac{\partial\tilde{N}_Z}{\partial u} - \frac{\partial}{\partial u}\left(\frac{s_Z}{\bar{s}}\tilde{N}_Z\right) &= -(\sigma_i(Z) + \sigma_r(Z))\tilde{N}_z \\ &+ \sigma_i(Z-1)\tilde{N}_{Z-1} + \sigma_i(Z+1)\tilde{N}_{Z+1}, \\ Z &= 0, 1, \ldots, Z_1,\end{aligned} \tag{57}$$

where

$$\tilde{N}_Z(\xi, u) = N_Z(\xi, \varepsilon)s(\bar{z}, \varepsilon).$$

To solve this equation, we used the finite-element representation of the function $\tilde{N}$ in variable u,

$$\tilde{N}(\xi, u) = \sum_m \tilde{N}_m(\xi)\varphi_m(u),$$

with the piecewise-linear basis functions $\varphi_m(u)$. An explicit Euler scheme was used to obtain the discrete approximation of the derivative with respect to ξ.

We checked out the numerical scheme by using a test problem in which the following model dependences of the coefficients in the equation were specified:

$$Z_0 = 1,$$
$$Z_1 = 40,$$
$$s(Z, \varepsilon) = \frac{Z^2}{Z_1^2}\frac{1}{\varepsilon}, \tag{58}$$
$$\sigma_i(Z, \varepsilon) = \exp(-0.1Z)\frac{100}{\varepsilon},$$
$$\sigma_r(Z, \varepsilon) = \exp(0.1Z)\frac{1}{\varepsilon^2}.$$

The results of our solution of the kinetic equation with the model dependences of the coefficients are shown in Fig. 8. As we see from this figure, the Bragg curves in the exact calculation and in the continuous ionization approximation, which disregards the fluctuations in ion charge, differ significantly. The total ion flux that is a step in the continuous ionization approximation, becoming zero at a depth equal to the total ion path, proves to be smoothed in the exact calculation. In this case, the point at which the flux is equal to half its initial value lies near the total path calculated in the continuous ionization approximation.

Figure 6 shows the numerically calculated mean charge and mean square of the deviation of the charge from its mean and the dependences obtained by using our analytical calculations with formulas (44). As we see from the figure, the coincidence is quite satisfactory. Figure 7 shows the $D_u(\xi)$ curves for the variable u obtained in our test calculation and using formula (52). These curves also closely coincide.

7. COMPARISON WITH EXPERIMENTAL DATA

Experiments on the slowing-down of ions in a cold material have long been carried out. In recent years, it has also become possible to experimentally study the behavior of high-energy ions in plasma. The previously predicted (see, e.g., [12]) peculiarities of the slowing-down of heavy charged particles in an ionized material have been confirmed experimentally. Thus, the stopping power for plasma is anomalously large compared to that for a cold material. This phenomenon is attributable to a decrease in the recombination cross section of ions in plasma due to a weakening of the charge exchange with particles of the medium. The decrease in recombination causes an increase in effective ion charge and a corresponding increase in energy losses.

The slowing-down of high-energy ions in a plasma is rather difficult to investigate, because it is necessary to prepare the plasma with known parameters and to synchronize the plasma and ion beam production times with the time of beam entry. Below, we give references to the experimental studies of the slowing-down of an ion beam in plasma known to us. The synchronization and ion access are most easily achieved in laser plasma experiments. The plasma produced by CO_2-laser illumination of an aluminum surface was used, for example, by the authors of [30]. The mean charge states of copper ions were investigated in the experiment described in this paper. Thus, the measured ion charge changed from 10 to 16 for the copper nuclear charge of 29. The authors compared their distributions with a Gaussian and pointed out a close coincidence in many cases. It should be noted, however, that the region of ion charge states lies within one electron shell (except charge 10).

Plasma with well-known parameters can be produced in capillary discharge experiments. This plasma production method was used by the authors of [31].

Note the interesting possibility of determining the ionization and electron capture cross sections using the model presented in this paper from experimental data on the charge distribution of ions slowing down in a material. To this end, we use data from [30]. This paper presents the measured charge distributions for Cl^{13+} and Cl^{15+} ions at various depths of penetration into cold hydrogen and strongly ionized hydrogen plasma.

The plots for the mean charge and the variance of the chlorine ion charge distribution in hydrogen at various initial charges shown in Fig. 9 were constructed by using experimental data from [31]. The solid lines in this figure indicate the analytical solutions obtained here using formulas (46). We see from the experimental data that, irrespective of how the mean charge behaves (whether it tends to its equilibrium value from high or low values), the variance of the charge distribution tends to the same steady-state value. According to the approximate solutions of the kinetic equation, the steady-state variance is equal to the ratio of the coefficients a and b of the linear fit (20) to the ionization and recombination cross sections. In our case, this ratio is approximately equal to

$$D = \frac{a}{b} = 0.9.$$

Since the ion charge in our case is close to its equilibrium value (approximately equal to 13.2), formula (46) may be used to calculate the variance. As we showed

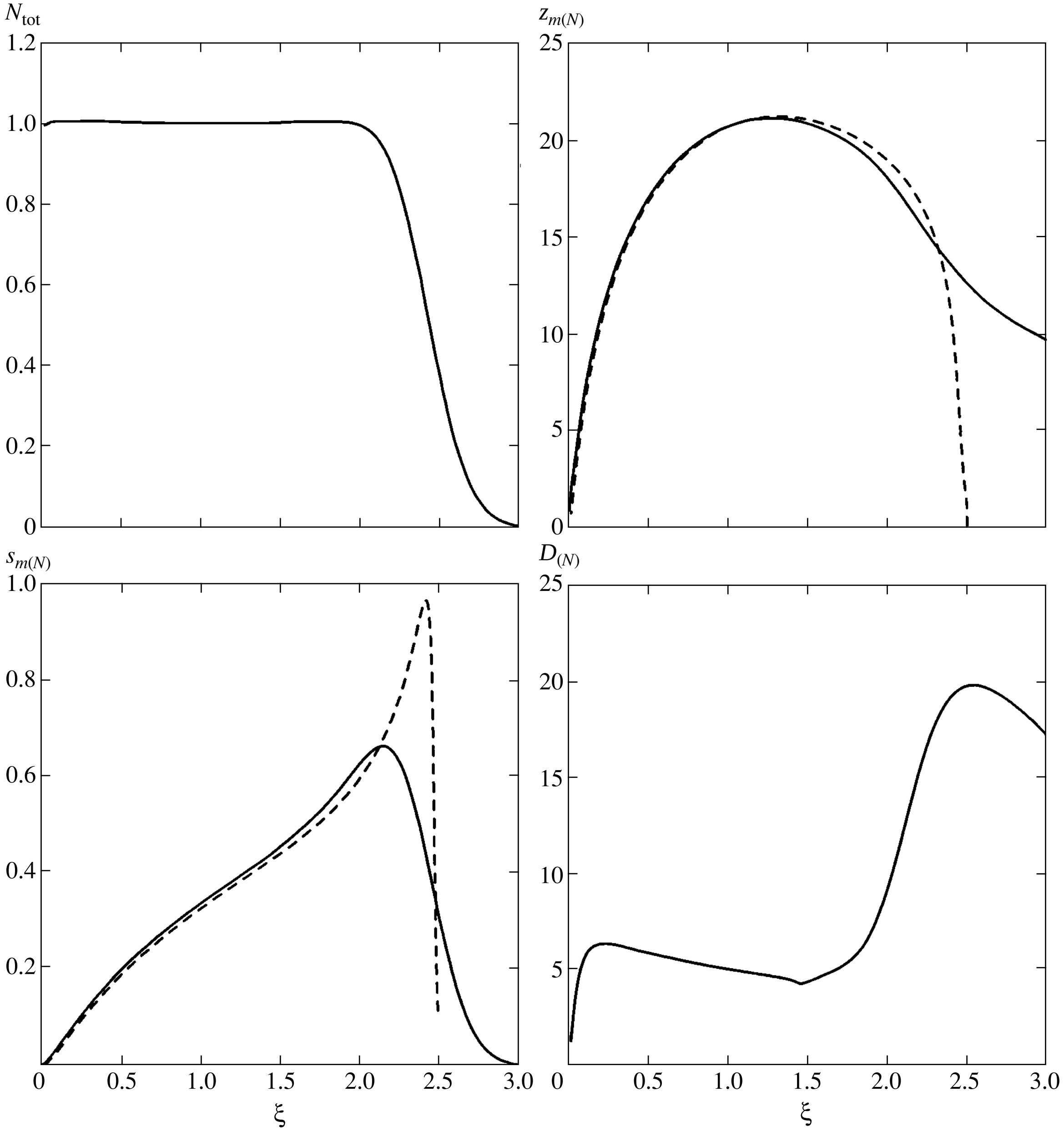


Fig. 8. The results of numerical calculations: N_{tot} is the total ion flux, $z_{m(N)}$ is the flux-averaged charge, $s_{m(N)}$ are the mean energy losses per unit path (Bragg curve), and $D_{(N)}$ is the flux-averaged variance of the charge distribution as a function of the dimensionless path ξ. The dotted lines indicate the solutions in the continuous ionization approximation. The test problem.

above, both the mean charge and the variance tend to their equilibrium values exponentially, with the coefficient b being the exponent. Using these dependences, we can estimate these coefficients:

$$a = 1.35\times10^{-19}\ \text{cm}^2, \quad b = 1.5\times10^{-19}\ \text{cm}^2.$$

These values are close to those that were used by the authors of [31] for their computer simulations of the slowing-down problem. Thus, for the ionization cross section of an ion with charge 13, they used 1.5 × 10^{-19} cm^2, which is close to our equilibrium cross section $\sigma_i = a = 1.35 \times 10^{-19}$ cm^2. According to the first equation of system (44), analysis of the behavior of the mean charge yields a coefficient b close to its value obtained from the path dependence of the variance.

The asterisks in Fig. 9 separately indicate the points that correspond to the mean charge and the variance of the ion charge distribution in a hydrogen plasma. As we see from the figure, the mean charge in this case exceeds the mean ion charge in cold hydrogen at an appropriate density, while the variance of the charge distribution in plasma is appreciably smaller than that

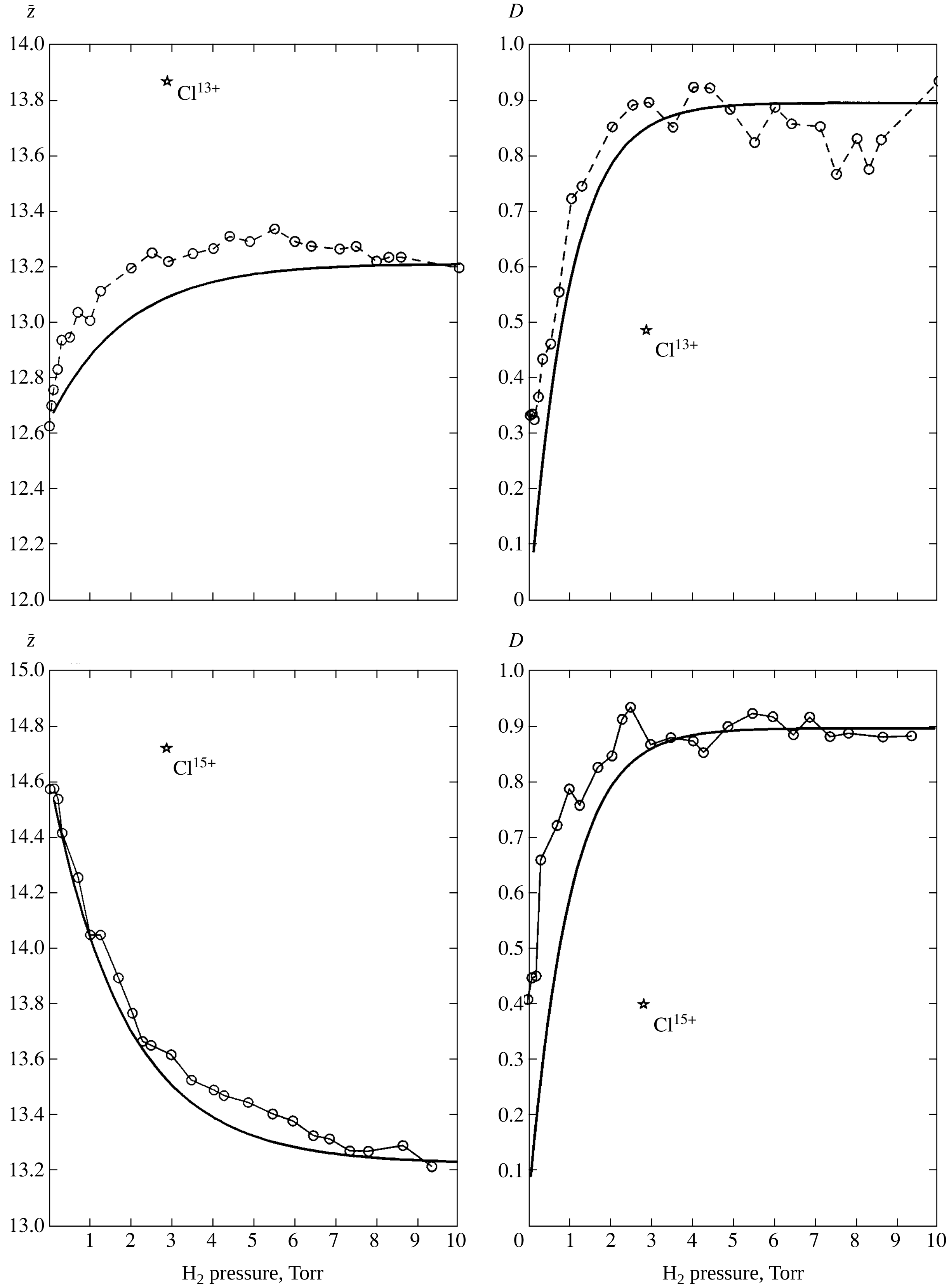


Fig. 9. Dynamics of the changes in the mean charge and charge variance of Cl ions with initial energy 1.5 MeV/*u* and initial charges 13 and 15 in cold hydrogen. The circles represent the experimental data from [31], and the solid curves represent the theoretical dependences (46). The asterisks indicate isolated measurements in hydrogen plasma.

in a cold material. This result agrees with the previously derived formulas for the variance. Indeed, for the increased ion charge, the sum of the ionization and recombination cross sections (coefficient a) becomes smaller, while the slope of the charge dependence of the cross section difference (coefficient b) becomes larger. As a result, the variance decreases.

Thus, the above analysis shows that the dependences of the ionization and recombination cross sections on ion charge can be estimated by using our analytical solutions of the ion slowing-down problem in the diffusion model. Since we can ensure that an ion during its slowing-down in plasma will pass through different charge states by varying the initial ion charge and energy, this method can yield an interesting means of calculating the ionization and recombination cross sections from experimental data. On the other hand, the charge variance is rather easy to estimate. The formula for the steady-state charge variance can be written in such a way that the derivatives of the logarithms of the ionization and recombination cross sections, which are smoother functions than the cross sections themselves, will be used in it:

$$D = \frac{\partial}{\partial z}\ln\left(\frac{\sigma_i}{\sigma_r}\right). \tag{59}$$

At the same time, the variance itself rapidly reaches its steady-state value.

8. CONCLUSIONS

We have suggested a diffusion method for the kinetic equation that describes the motion of high-energy ions in plasma or a cold material. Based on our model, we obtained approximate solutions that describe the mean ion charge and the variances of the ion charge distributions as a function of the ion path in the target. Our calculation of the variance of the charge distribution is based on the fact that the mean square of the deviation from the mean charge for heavy ions is much smaller than the square of the charge, and the ion charge distribution (at a given energy and the same initial charge) is a narrow peak localized near the mean charge. This is attributable to the peculiarities of the charge behavior of the coefficients of the kinetic equation that describes the slowing-down of ions in a material.

We also described the method for numerically solving the ion transport equation and performed calculations whose results were used to test the analytical solutions of the diffusion method. Our numerical calculations confirm the conclusion that we reached when analyzing the problem that the ion charge and charge variance change along the characteristic in the path–energy plane. The mean charge along this characteristic can be calculated by using the formulas of the continuous ionization approximation.

We considered the problem of determining the variance of the ion energy distribution. Allowance for the changes in ion charges leads to a significant increase in the energy spread of ions at the final stage of their slowing-down, which, in turn, leads to a smearing of the peak in the Bragg curve for the mean energy losses of ions in a material.

We compared the analytical solutions of the problem with available experimental data. In general, the experimental results agree well with the analytical formulas. In our opinion, the existing differences between the experimental data and the analytical model calculations for the slowing-down of high-energy ions in weakly ionized plasma are attributable to the insufficiently accurate models of the cross sections. In this case, the experimental data lead us to conclude that a particular model is valid, because the analytical method developed here allows information about the cross sections to be extracted from the mean charge and the charge variance.

ACKNOWLEDGMENTS

We thank the staff of B.Yu. Sharkov's laboratory at the Institute for Experimental and Theoretical Physics for a valuable discussion on the interactions of heavy ions with materials.

REFERENCES

1. O. Volmer, Nucl. Instrum. Methods **121**, 373 (1974).
2. K. B. Winterbon, Nucl. Instrum. Methods **144**, 311 (1977).
3. F. Besenbacher, J. U. Andersen, and E. Bonderup, Nucl. Instrum. Methods **168**, 1 (1980).
4. P. Sigmund, Nucl. Instrum. Methods Phys. Res. B **69**, 113 (1992).
5. A. Narmänn and P. Sigmund, Phys. Rev. A **49**, 4709 (1994).
6. H. Bethe, Ann. Phys. (Leipzig) **5**, 324 (1930).
7. N. Bohr, Philos. Mag. **25**, 10 (1913).
8. J. Lindhard and A. H. Sorensen, Phys. Rev. A **53**, 2443 (1996).
9. M. M. Basko, Fiz. Plazmy (Moscow) **10**, 1195 (1984) [Sov. J. Plasma Phys. **10**, 689 (1984)].
10. T. Peter and Ju. Meyer-ter-Vehn, Phys. Rev. A **43**, 1998 (1991).
11. T. Peter and Ju. Meyer-ter-Vehn, Phys. Rev. A **43**, 2015 (1991).
12. E. Nardi and Z. Zinamon, Phys. Rev. Lett. **49**, 1251 (1982).
13. V. S. Remizovich, D. B. Rogozkin, and M. I. Ryazanov, *Fluctuations of Charged Particle Randes* (Énergoatomizdat, Moscow, 1988).
14. M. Gryzinski, Phys. Rev. A **138**, 305 (1965).
15. L. Vriens, Proc. R. Soc. London **90**, 935 (1966).

16. V. P. Shevelko, *Single and Multiple Ionization of Atoms and Ions by Electron Impact* (Harwood Academic, New York, 1999).
17. Ya. B. Zel'dovich and Yu. P. Raĭzer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, 2nd ed. (Fizmatgiz, Moscow, 1966; Academic, New York, 1966 and 1967), Vols. 1 and 2.
18. V. P. Kovalev, *Ion Effective Charge* (Énergoatomizdat, Moscow, 1991).
19. B. M. Smirnov, *Ions and Excited Atoms in Plasma* (Atomizdat, Moscow, 1974).
20. W. Lotz, Z. Phys. **220**, 466 (1969).
21. H. C. Brinkmann and H. A. Kramers, Proc. K. Ned. Akad. Wet. **33**, 973 (1930).
22. R. A. Mapleton, *Theory of Charge Exchange* (Springer, New York, 1972).
23. J. Eichler and F. T. Chen, Phys. Rev. A **20**, 104 (1979).
24. J. Eichler, Phys. Rev. A **32**, 112 (1985).
25. A. S. Schlachter, J. W. Stearns, W. G. Graham, *et al.*, Phys. Rev. A **27**, 3372 (1983).
26. R. K. Janev, L. P. Presnyakov, and V. P. Shevelko, *Physics of Highly Charged Ions* (Springer, New York, 1985).
27. V. S. Nikolaev and I. S. Dmitriev, Phys. Lett. A **28**, 277 (1968).
28. M. D. Brown and C. D. Moak, Phys. Rev. B **6**, 90 (1972).
29. M. S. Armel and M. S. Funkhouser, Phys. Rev. A **64**, 64502 (2001).
30. C. Couillaud, R. Deicas, Ph. Nardin, *et al.*, Phys. Rev. E **49**, 1545 (1994).
31. M. Chabot, D. Gardes, P. Box, *et al.*, Phys. Rev. E **51**, 3504 (1995).
32. S. Andriamonje, B. Blank, R. Delmoral, *et al.*, Nucl. Instrum. Methods Phys. Res. B **87**, 116 (1994).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 948–957.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 1058–1067.
Original English Text Copyright © 2003 by Caruso, Strangio.

**PLASMA,
GASES**

Ignition Thresholds for Deuterium–Tritium Mixtures Contaminated by High-*Z* Material in Cone-Focused Fast Ignition[¶]

A. Caruso and C. Strangio
CR ENEA Frascati, Frascati, RM, Italy
e-mail: caruso@frascati.enea.it

Received June 6, 2003

Abstract—Evaluations of the energy for thermonuclear ignition of a compressed deuterium–tritium mixture contaminated by a high-*Z* material are presented. Mixing at the atomic level is considered and the results are given as a function of the contaminant fraction. The reference situation is that of cone-focused fast ignition (CFFI). The numerical 2*D* simulations for this study were performed by a Lagrangian 2*D* hydrocode that includes real matter EoS, real matter opacity coefficients, and packages for finite-range energy deposition by reaction products and the relative in-flight reactions. A simple estimate is presented for the effects of high-*Z* material blobs on the ignition energy (macroscopic mixing). Possible sources for fuel contamination in CFFI are discussed. *© 2003 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Thermonuclear ignition thresholds for a compressed deuterium–tritium mixture contaminated by a high-*Z* material at the atomic level were evaluated as functions of the contaminant fraction. A short pulse of protons was used to start the ignition of a cylindrical assembly of compressed fuel uniformly contaminated by gold at the atomic level. As a reference, a study of the ignition of a clean target at different proton energies was performed first and, after this, the ignition conditions for contaminated targets were found for the selected proton energy. Protons with the proper energy can be used to mock up deposition by fast electrons such that a part of the study can also be considered useful in predicting the performances for this energy vector.

The 2*D* code COBRAN was used to perform the study. COBRAN includes a “real matter” equation of state and opacity coefficients. The package for driving energy deposition includes light/heavy charged particles, ray-traced laser light, and ray-traced X-rays. The thermonuclear reaction treatment includes finite-range charged particle diffusion and nonthermal nuclear reactions. The diffusion and energy deposition of neutrons is treated by a Monte Carlo code. However, in the cases considered here, the effect of neutrons was marginal and this part of the code was normally not activated [1].

A simple estimate for the effects of high-*Z* material blobs on the ignition energy (macroscopic mixing) are presented below. The estimate is given in terms of the “cool” surface of the high-*Z* blobs.

This study can be relevant for the so-called cone-focused fast ignition (CFFI) [2, 3], because in experiments performed for this scheme, the imploded material was found to be contaminated by the cone high-*Z* material [3, 4]. In the CFFI, a shell containing a layer of DT fuel is actually imploded by soft X-rays or laser-light-induced ablation, sliding along the external surface of a cone composed by the high-*Z* material (see Fig. 1a). As a result of the implosion, a blob of compressed fuel is formed near the cone tip (see Fig. 1b). At this time, a short laser pulse is focused inside the cone

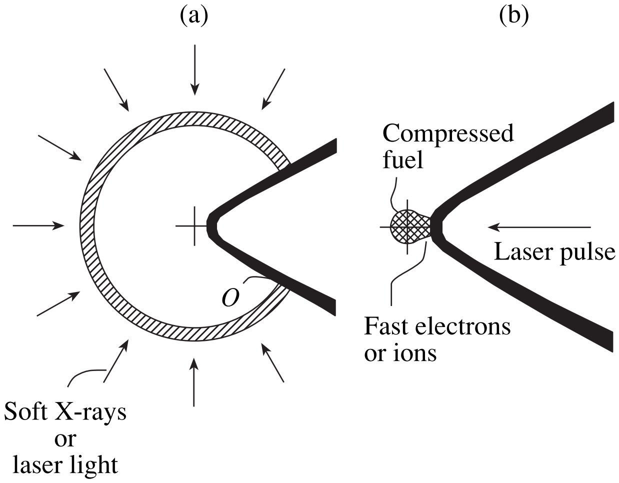


Fig. 1. CFFI scheme. (a) Ablation pressure induced by X-rays or laser imploding a fuel shell to form a compressed fuel assembly near the tip of the cone. (b) A short laser pulse is used to ignite the fuel by fast electrons or energetic light ions.

[¶] This article was submitted by the authors in English.

to produce a forward jet of fast electrons. Passing through the cone material, the electrons can create an ignition spark on the compressed fuel assembly located nearby. Alternative energy vectors can be light ions (e.g., protons) produced either by using the fast electrons as a virtual cathode with respect to a properly composed cone tip [3] or by causing the explosion of a low-*Z* foil set in place of the cone terminal portion [1, 5].

High-*Z* contamination was found in experiments performed in the indirect drive mode. It was indicated that vaporization of the cone material (gold) by X-rays (*m*-lines) passing through the imploding shell was responsible for this effect [4]. Such a source of contamination should be absent in the direct-drive mode. However, another source, potentially also active in this case, can be excited by the onset of the Kelvin–Helmholtz instability on the sliding interface between the imploding shell and the cone surface. We have studied this transient process and found the shell transit near the cone tip to be the most critical stage.

During the implosion, the pressure in the acceleration stage can range from 1 to 100 Mb on the shell and from 1000 to 2000 Mb within the shell material at the transit near the cone tip. Pressures of this order are also exerted on the interface between the shell and the cone material. At such values of pressure, the cohesion forces in the cone material can be neglected and the shell containment by the cone is inertial. This statement is valid for gold as the cone material because the maximum tensile strength of Au estimated from the vaporization specific energy is on the order of 0.14 Mb (see [6]). A very sketchy representation can be used to infer some of the basic flow features just near the leading shell points sliding on the cone (represented by *O* in Fig. 1a). Consider a reference frame attached to *O* and assume a locally planar 2*D* flow pattern. In this frame, the cone material impinges from the left at a velocity of $\mathbf{V}_0$ ($|\mathbf{V}_0|$ is the implosion velocity, see Fig. 2a). The region behind the leading edge of the imploding fuel shell is roughly mocked up by a stagnating uniform pressure *p*-wave starting after a rise-front (*f*). At *f*, the pressure passes from 0 to *p* over a distance assumed negligible. The main features of the flow are an oblique shock wave (angle ϕ) deflecting the cone material velocity by angle χ, from $\mathbf{V}_0$ to $\mathbf{V}$ [7], and a slip surface (*s*) between the cone and the shell materials, represented by the pressure wave.

Assuming the initial pressure on the impinging cone material negligible and assigning p, $\mathbf{V}_0$, and the initial material (solid-state) density ρ_s, the shock parameters follow from the Hugoniot curve

$$\varepsilon(\rho, p) = p\left(\frac{1}{\rho_s} - \frac{1}{\rho}\right)/2,$$

where ε is the internal energy per unit mass [7] ($\varepsilon(\rho_s, 0) = 0$) and ρ is the density behind the shock.

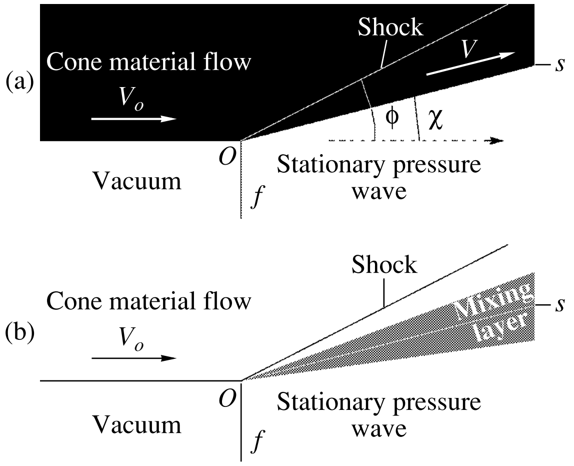


Fig. 2. (a) The hydrodynamical flow induced in the cone material (black) by the pressure of the imploding shell (gray). (b) Instabilities at the slip surface *s* mix cone and shell materials.

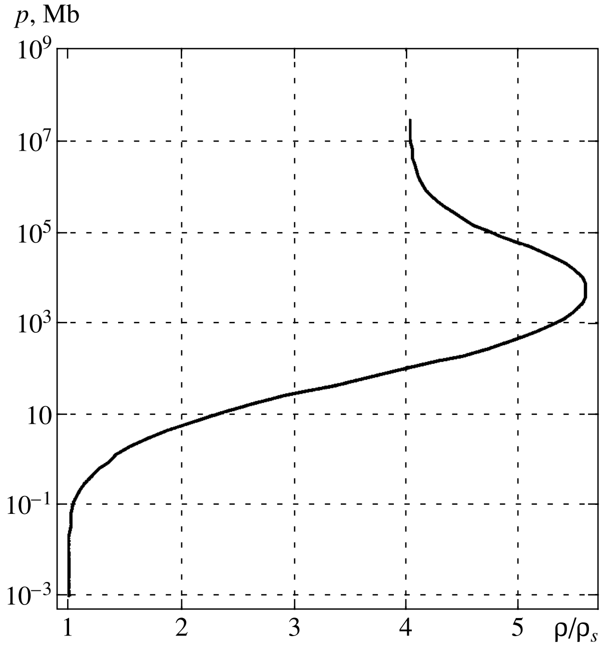


Fig. 3. Hugoniot adiabat for a shock in gold at the initial solid-state density ρ_S and at the initial pressure set to 0. The adopted equation of state for gold was that on line in the COBRAN code.

With p given, ρ follows. To evaluate $\varepsilon(\rho, p)$, we took the equation of state for “real materials” included in the previously mentioned 2*D* code COBRAN. In Fig. 3, the resulting Hugoniot adiabat is shown for Au.

Using mass, momentum, and tangential velocity conservation at the shock, we easily determine the quantities V, $\sin\chi$, and $\sin\phi$ (see Fig. 2) as simple functions of the assigned quantities p, ρ_s, and $\mathbf{V}_0$ and of the density ρ corresponding to p in the Hugoniot adiabat.

Table 1

Reference case	p	V_0	ρ_{DT}	$c_{s\,DT}$	V	ρ_{Au}	$c_{s\,Au}$	χ	ϕ
First shock wave	1	6	0.78	1.7	≈6	26.2	0.42	1	7.5
End of accelerating pulse	100	15	10	4.3	14.8	76	1.6	7.6	10
Near cone tip transit	1200	13	40	7.2	9.6	103	4.4	33	42

We define the quantities

$$c_0 = \left(\frac{p}{\rho_s}\right)^{1/2}, \quad \xi = \frac{\rho}{\rho_s}, \quad M = \frac{V_0}{c_0}.$$

The oblique shock is formed if

$$V_0 > c_0\left(\frac{\xi}{\xi - 1}\right)^{1/2}.$$

The relevant quantities for this study are V, χ, and ϕ given by

$$V = V_0\left(1 - \frac{1 + 1/\xi}{M^2}\right)^{1/2},$$

$$\sin\chi = \frac{1}{M}\left[\frac{\xi - M^2(\xi - 1)}{1 - \xi(M^2 - 1)}\right]^{1/2},$$

$$\sin\phi = \frac{1}{M}\left(\frac{\xi}{\xi - 1}\right)^{1/2}.$$

In Table 1, the above quantities are reported for regimes that can occur in the implosion of a thin spherical shell (e.g., with the in-flight aspect ratio equal to 14). The shell is first set at a low adiabat of $\alpha = 0.3$ by a pressure pulse of $p_0 = 1$ Mb (the first shock wave, α is the ratio of the thermal electronic pressure to the Fermi pressure). A gradual (adiabatic) rise of the pressure up to $p = 100p_0$ then accelerates the shell to the maximum velocity. The shell material is finally left to freely implode towards the cone tip, where the pressure is expected to be 1000–2000 Mb. The densities (ρ_{DT}, ρ_{Au}) and the sound velocities (c_{sDT}, c_{sAu}) were consistently estimated at the “real matter” level in the flowing cone material (gold) and for $\alpha = 0.3$ in the “DT shell” set at rest. The units are Mb (pressure), g/cm^3 (densities), cm/μs (velocities), and degrees (angles). The effects of the shock wave on the gold flow appear relatively modest in the first two stages of the implosion. Severe flow distortion is produced near the cone tip.

In what follows, we report stability calculations for the slip surface and the corresponding quasilinear evaluation for the mixing layer aperture. Numerical estimates are presented for the reference equilibrium cases listed in Table 1.

For any mechanism of fuel poisoning, it is not easy to evaluate the spectrum of sizes or masses of Au finally mixed into the compressed fuel as they undergo a complex evolution before being trapped in the ignition spark area. Two types of contamination can be expected, one in the form of mixing at the atomic level, and the other as a distribution of Au blobs in the fuel. For small blobs, a two-step evolution can be expected. In the first, a gold blob immersed in a DT plasma at a temperature of 10^8 K or higher and a density of $\rho_{DT} = 200$ g/cm^3 is ablated by electronic thermal conduction and brought to the temperature and pressure equilibrium with the DT fuel. The already mentioned equation of state predicts a density of $\rho_{Au} \approx 435$ g/cm^3 for gold and an average charge of $Z_{Au} \approx 70$. The second step is the diffusion of DT ions through the gold plasma. The diffusion coefficient of the average DT ion through gold plasma is

$$k_{DT} = \frac{7 \times 10^{-14} T_i^{5/2}}{Z_{Au}^2 \rho_{Au}} \mathrm{CGS} \cdot \mathrm{K},$$

and the diffusion length in a disassembling time is therefore given by

$$R_{crit} = \sqrt{k_{DT} R_s / c_s},$$

where R_{crit} represents the largest blob that can be diffused through during the spark lifetime. The previous estimate can be set as

$$R_{crit} = \frac{0.003}{\sqrt{\rho_{DT}\rho_{Au}}}\ \mathrm{cm}$$

by setting the ion temperature to 10^8 K in the previous formula, $R_s = 0.5/\rho_{DT}$ cm (as typically required for ignition), and $Z_{Au} = 70$. The mass corresponding to these values is

$$M_{crit} \approx \rho_{Au} R_{crit}^3 = \frac{2.7 \times 10^{-8}}{\rho_{Au}^{1/2} \rho_{DT}^{3/2}}.$$

For $\rho_{Au} \approx 435$ g/cm^3 and $\rho_{DT} \approx 200$ g/cm^3, $R_{crit} \approx 0.1$ μm and $M_{crit} \approx 5 \times 10^{-13}$ g.

Below, we report evaluations for the amount of Au (atomic mixing) or the sizes of blobs (macroscopic mixing) necessary to obtain the relevant effect on the ignition threshold in the two contamination modes.

2. IGNITION STUDIES

The ignition studies for contamination at the atomic level were performed for cylinders of a compressed DT + Au uniform mixture at initial densities of approximately 200 g/cm^3, irradiated by a proton beam along the axis. Advantages in using light ions for fast ignition were considered in [8]. At any rate, proton beams with the proper pulse and particle energy can be used to roughly mock up deposition by fast electrons as both these energy vectors give energy mostly to the electrons in the target.

The length of the cylinder (S) was taken equal to the diameter ($2R$) and set to 300 μm. The intensity distribution of the beam as a function of the radius r was Gaussian, proportional to $\exp(-r^2/R_0^2)$ for $r \le 2R_0$ and 0 outside. The scale R_0 was set to 25 μm. The pulse duration was taken as 15 ps, short enough to make the expansion negligible during irradiation. The time dependence was a linear ramp for the first 5 ps and constant for the remaining 10 ps.

The initial fuel composition was assigned by giving the numeric fractions

$$f_{\mathrm{Au}} = \frac{n_{\mathrm{Au}}}{n_{\mathrm{Au}} + n_{\mathrm{D}} + n_{\mathrm{T}}},$$

$$f_{\mathrm{D}} = f_{\mathrm{T}}, \quad f_{\mathrm{Au}} + f_{\mathrm{D}} + f_{\mathrm{T}} = 1,$$

where n_{Au}, n_{D}, and n_{T} are the numeric densities of gold, deuterium, and tritium. In the simulations, $n_{\mathrm{D}} = n_{\mathrm{T}}$ was initially assigned.

The initial target density was either set equal to 200 g/cm^3 or adjusted for each value of f_{Au} such that the total pressure (p) was equal to that of a clean reference target at 200 g/cm^3 and $\alpha = 0.3$, assuming uniform temperature. The second case was considered to simulate equilibrium between the contaminated region where the ignition spark is formed and a contiguous clean fuel at the specified reference conditions. For our “realistic” equation of state, this implied that ρ increases with f_{Au}.

In Fig. 4, we show the electron temperature distribution in a clean reference target ($f_{\mathrm{Au}} = 0$, $\rho = 200$ g/cm^3) at the end of an igniting proton pulse of 40 kJ and 1 MeV.

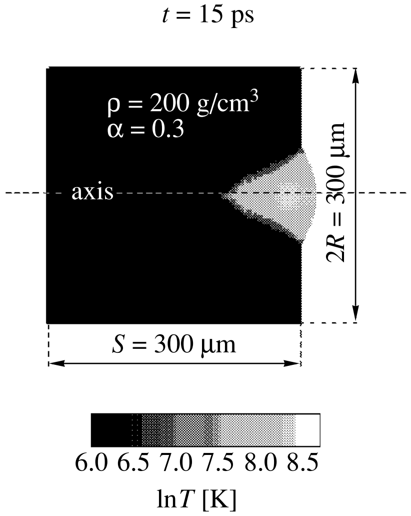


Fig. 4. Electron temperature distribution in a pure DT cylindrical target at the end of the ignition proton pulse (pulse energy 40 kJ, proton energy 1 MeV).

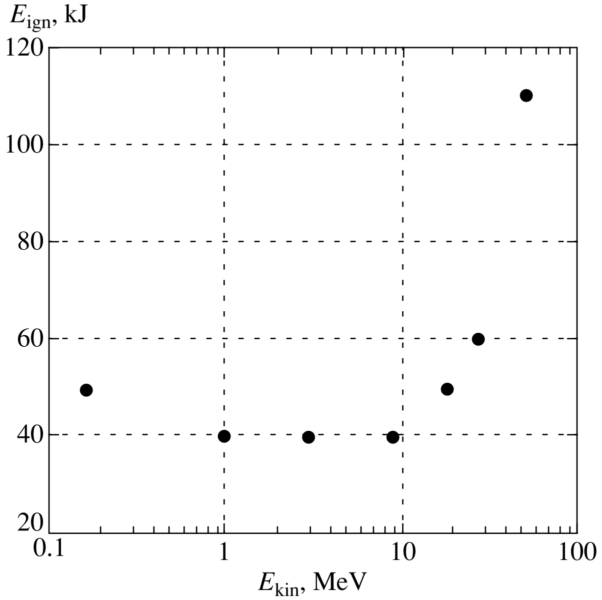


Fig. 5. Ignition energy for a 200 g/cm^3 DT fuel by a proton beam as a function of the proton kinetic energy.

2.1. Ignition of a Clean Fuel

The ignition energy (E_{ign}) for a clean fuel at $\rho = 200$ g/cm^3 is represented in Fig. 5 as a function of the ion (proton) kinetic energy (E_{kin}). No ignition occurs at less than 40 kJ. The total fusion energy released by the target after complete disassembling is about 540 MJ, the radiated energy is about 5 MJ, and the burned fraction is $f_b = 0.38$.

The ignition of a compressed clean fuel by light ions demonstrates several typical features. One, clearly seen in Fig. 4, is the rather spiked structure of the spark despite the smooth Gaussian space distribution of the irradiation. This feature is simply explained by noting that the range of the ions increases with temperature, and this depends on the amount of energy deposited. The ion range is therefore a self-consistent feature.

During irradiation, a bleaching wave is formed that allows a deeper ion penetration into the target in the regions where the temperature is higher, namely, near the ion beam axis. Neglecting thermal diffusion, nuclear heating, and the “cold” target range, this effect can be estimated, for the range (L) equation, as

$$\rho L \propto E_{\mathrm{kin}}^{1/2} T_e^{3/2} \tag{1a}$$

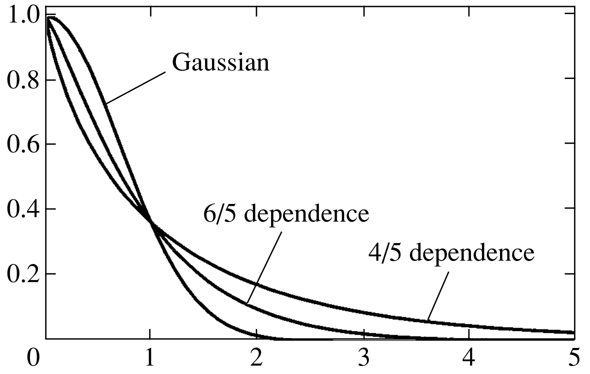


Fig. 6. Space distributions for ρL, T_e, and E_s according to Eqs. (2a) and (2b).

and, for the energy equation,

$$\rho L T_e \propto E_s(r, t), \qquad (1b)$$

where $E_s(r, t)$ is the energy deposited per unit surface.

These equations imply that

$$T_e \propto E_{\rm kin}^{-1/5} E_s^{2/5}(r, t) \propto E_{\rm kin}^{-1/5} \exp\left[-\left(\frac{r}{R_0}\right)^{4/5}\right], \qquad (2a)$$

$$\rho L \propto E_{\rm kin}^{1/5} E_s^{3/5}(r, t) \propto E_{\rm kin}^{1/5} \exp\left[-\left(\frac{r}{R_0}\right)^{6/5}\right]. \qquad (2b)$$

The different dependences are presented in Fig. 6. The spiked structure is clearly shown for both ρL and T_e.

Peculiar structures and behaviors can be detected in burn waves. In Fig. 7, we show the development and propagation of the burn wave for a target irradiated by a beam of protons of 40 kJ and 1 MeV. Up to 80 ps, burn propagation occurs via a mode featured by in-cavity burning and alpha-particle ablation-driven shock front. The shock generated by the ablation pressure increases the density up to approximately 570 g/cm^3. In the ablation front, the density drops to about 36 g/cm^3 over a distance of about 40 μm. Near t = 80 ps, a new propagation mode starts. A quasispherical shock wave detaches from the tip of the ablation front and the reaction propagates as a detonation wave starting from a point near the initial fuel assembly center. The produced thermonuclear power W_{Tn} and energies are also shown in Fig. 7; we note the bell-like time dependence of the power. Density profiles along the z axis of the cylinder are shown in Fig. 8 for two propagation modes.

The propagation pattern can change substantially if the igniting pulse energy is well above the threshold. This can be seen in Fig. 9, where the density maps are presented in the case where the energy of the igniting pulse rises from 40 kJ to 70 kJ, with the other parameters remaining unchanged. After a fast radial propagation, the detonation propagates along the target axis. The power release of thermonuclear energy presents a plateau corresponding to the propagation along the axis of the cylinder.

Highly structured detonation waves can also be found. In Fig. 10, a 27-MeV beam of protons starts to ignite at 60 kJ. Both the radial wave and tip starting

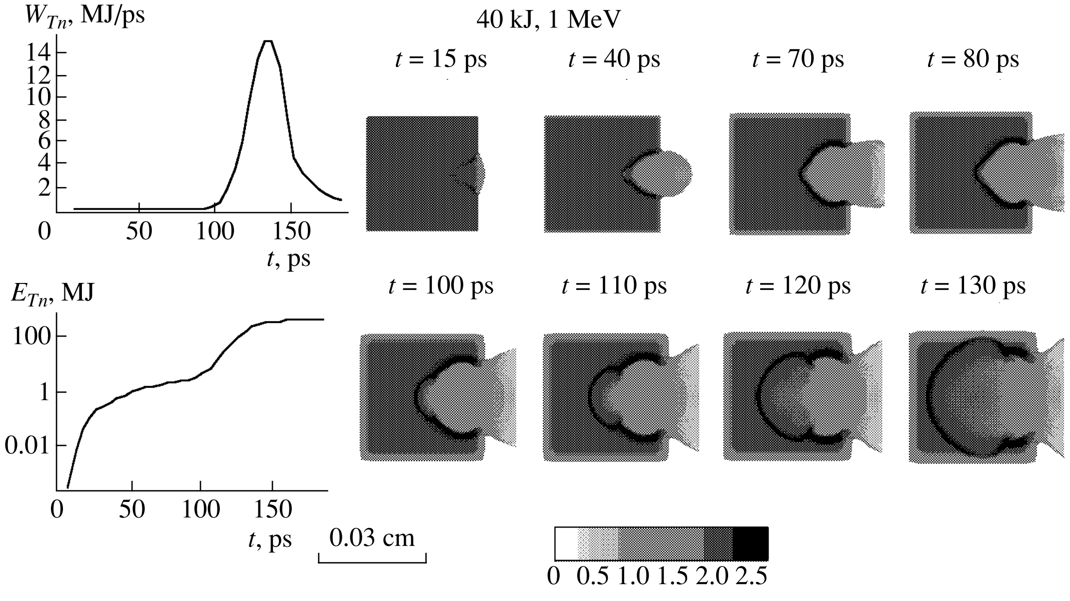


Fig. 7. Burn propagation for the ignition spark shown in Fig. 1. Density maps. The propagation for the lowest energy needed to ignite the considered fuel assembly is represented. After propagation in the α-particle ablation-driven mode, a detonation wave starts near the initial fuel assembly center, at the tip of the ablation front. The produced thermonuclear (Tn) power and energies are also shown; we note the bell-like time dependence of the power.

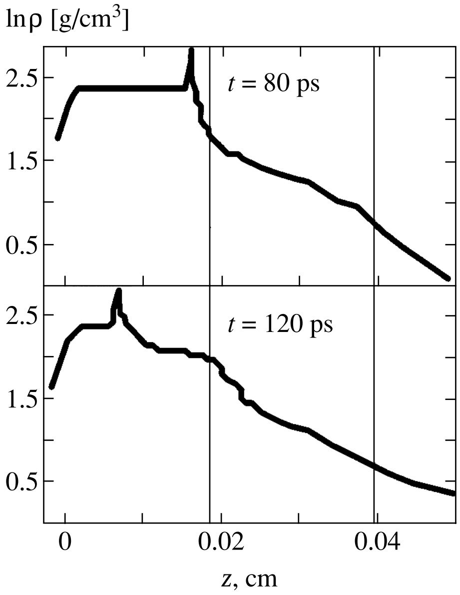


Fig. 8. Density profile along the cylinder axis at different times showing features of the two burn propagation modes.

waves propagate and ultimately coalesce in a single wave propagating along the axis.

2.2. Ignition of a Fuel Contaminated by High-Z Au Materials

In this section, we present the results of simulations performed to study fast ignition thresholds of a fuel uniformly doped by gold. Studies of the effects of light impurities (C, Si, and O) on the performances of directly driven targets have been published in the past [9].

In the case of $f_{\rm Au} \neq 0$, the energy $E_{\rm kin}$ was set to 1 MeV (corresponding to the ignition energy $E_{\rm ign}$ = 40 kJ in the clean case, see Fig. 5) and $f_{\rm Au}$ was increased from 0.001 to 1.5%. The introduction of high-Z contamination affects the ignition thresholds via several mechanisms. Some of these are detectable by inspection of Fig. 11. The electron temperature maps are given at the end of the igniting pulse for different degrees of doping. It can be seen that by increasing $f_{\rm Au}$, the needed electron temperature increases. At the largest doping, the spark is imbedded in a thick radiation wave. The presence of gold affects the heat capacity and, via this, the igniting-beam penetration depth, which is a self-consistent feature. The spiked structure disappears. The difference between T_e and T_i increases with $f_{\rm Au}$. The situation evolves very rapidly, however. Actually, 10 ps later, for $f_{\rm Au}$ = 1%, for instance, due to diffusive losses and relaxation effects, the electron temperature decreases to 26 keV and the maximum ion temperature achieves a value of approximately 39 keV. In this same case, the simulations show that due to the reduced DT content (about 57%) and to a smaller burned fraction (about 25% instead of 38% of the clean fuel), the yield is decreased to about 40% of that of the clean fuel case. It is important to remark, however, that the calculations presented here are relevant to assess whether ignition occurs for a given contamination of the spark region. The previous considerations relative

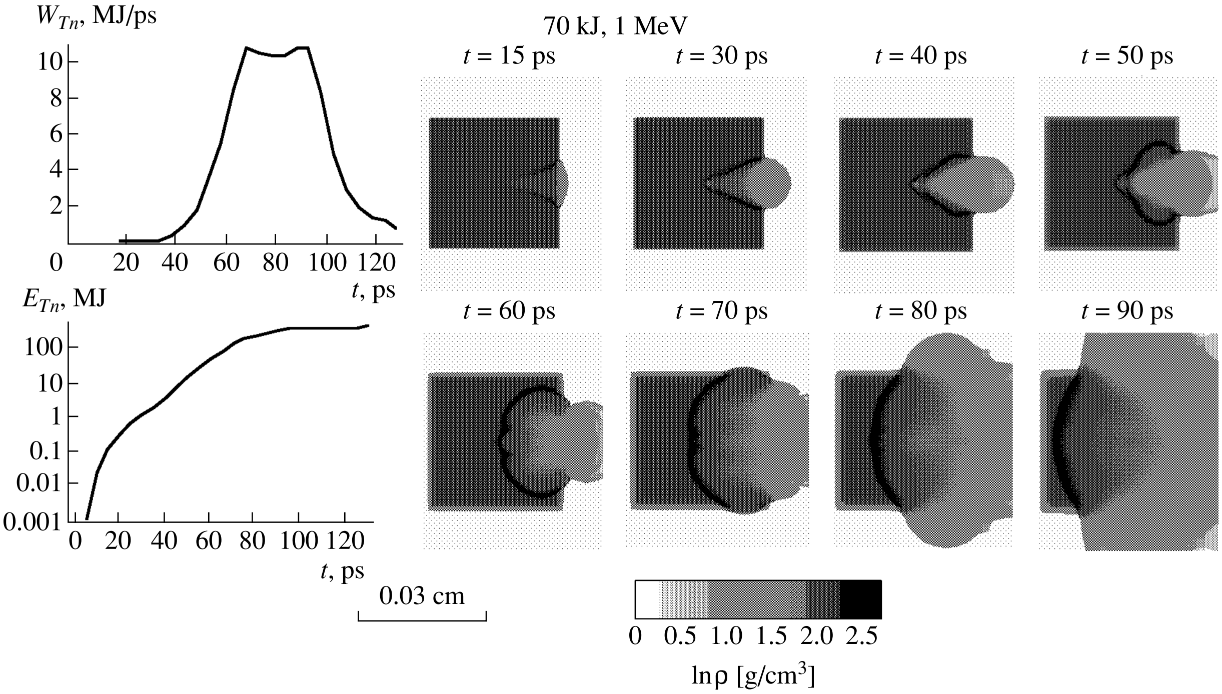


Fig. 9. Burn propagation through a target irradiated by an energy pulse well above the threshold. The detonation propagates first in the radial direction and then along the axis. The plateau on the burn-rate curve is connected to the axial propagation (to be compared with the bell-like shape shown in Fig. 7).

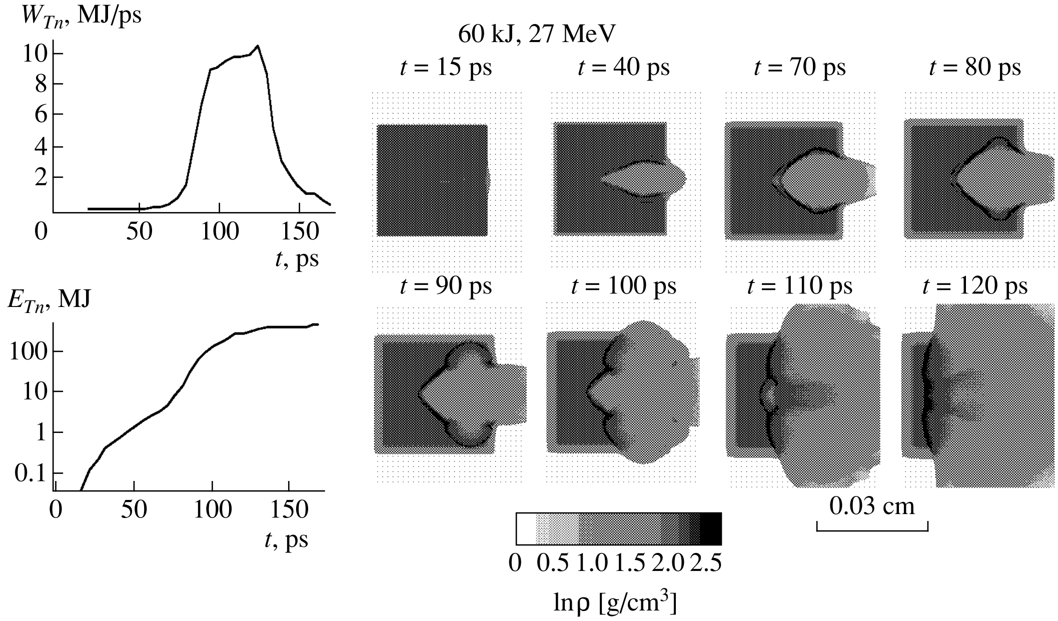


Fig. 10. Burn propagation for a target ignited by a 27-MeV proton beam. Density maps. The burn wave is highly structured. This results from simultaneous propagation of waves in the radial direction and from the heated cone tip. The situation is a hybrid of the cases shown in Figs. 5 and 7.

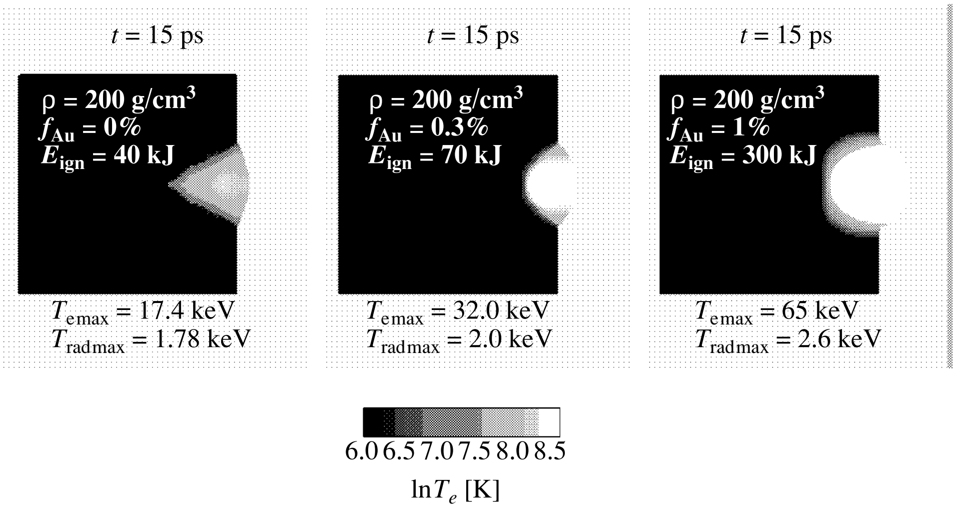


Fig. 11. Comparison between the electron temperature distribution maps for increasing f_{Au} at the end of proton-beam irradiation. We note the change in the spark shape. The thick gray feature for $f_{Au} = 1\%$ is due to radiation preheating, and the maximum radiation temperature increases with f_{Au}. The same trend occurs for the difference in temperature between electrons and ions. This, however, is a highly transient situation: 10 ps later, for $f_{Au} = 1\%$, for instance, due to diffusive losses and relaxation effects, the electron temperature decreases to 26 keV and the maximum ion temperature achieves the value about 39 keV.

to the yield can be taken as meaningful in the case where contamination is truly uniform.

The results of the investigation for $f_{Au} \neq 0$ are shown in Fig. 12, where the data for the assigned density or pressure are shown. The two fitting curves are similar (exponential functions), with the one for a constant pressure (circles) being somewhat shifted to the right. In both cases, E_{ign} represents a steep increase as f_{Au} approaches 0.2% (about 14% by mass). For greater values of f_{Au}, the increase in E_{ign} is very fast.

3. EFFECTS OF MACROSCOPIC CONTAMINATION

With regard to the effects of macroscopic mixing, we estimate the typical size of the blobs that have a substantial effect on the ignition threshold. Assuming the surface of a contaminant blob to be a sink for the fuel energy through the electron thermal flow, we can estimate the order of magnitude of the power W_a absorbed in the ignition spark by the contaminant as

$$W_a \approx \kappa_e T_e^{7/2} 4\pi \sum R_b.$$

Here, κ_e is the Spitzer thermal conductivity coefficient and $\sum R_b$ is the sum of the blob radii contained in the spark. The total energy absorbed before spark disassembling is

$$E_a \approx W_a(R_s/c_{s\mathrm{DT}});$$

this value has to be compared with the spark thermal energy E_{th}. For a standard spark ($T \approx 10$ keV, $\rho_{\mathrm{DT}}R_s \approx 0.5$ g/cm^2), we find that

$$\frac{E_a}{E_{\mathrm{th}}} \approx \frac{3.8\sum R_b}{R_s}.$$

Setting this ratio to 0.5 implies that

$$\sum R_b \approx 0.13 R_s.$$

4. KELVIN–HELMHOLTZ INSTABILITY AT THE SHELL–CONE INTERFACE

In the reference cases in Table 1, a treatment of the Kelvin–Helmholtz instability for compressible fluids with a discontinuous density and velocity of sound was required. The adopted geometry is shown in Fig. 13. The xy plane is taken on the slip surface s and the wavenumber is taken in the xz plane. The flow is assumed at rest for $z < 0$, in the region occupied by the imploding shell material. In the $z > 0$ region, where the cone material flows, the velocity is set to velocity V resulting from the shock analysis. Assuming that the perturbed quantities behave as $\exp[-i\omega t + i(k_x x + k_z z)]$, we write the dispersion relation

$$c^2 r^2(\nu^2\omega^2 - 4)(\omega - 2)^4 - [\nu^2(\omega - 2)^2 - 4c^2] = 0, \quad (3)$$

where

$$r = \frac{\rho_{\mathrm{Au}}}{\rho_{\mathrm{DT}}}, \quad c = \frac{c_{s\mathrm{Au}}}{c_{s\mathrm{DT}}}, \quad \nu = V\frac{\cos\psi}{c_{s\mathrm{DT}}},$$

and the complex frequency ω is normalized by $k_x V\cos\psi/2$. The study was performed numerically and the results are as follows. The interface is always unstable for waves with $90° \geq \psi \geq \psi_2$, where $\psi_2 \geq 0$. In this interval, the growth rate γ (the imaginary part of ω) starts from 0 at $\psi = 90°$, achieves a maximum γ_m at some ψ_m, and vanishes again for $\psi = \psi_2$ if $\psi_2 > 0$. Furthermore, unstable modes can be found when $\gamma \neq 0$ for $\psi_2 = 0$. All these unstable modes correspond either to a reduced component of velocity V along x ($\psi_2 > 0$) or to a sufficiently small value of V itself (for $\psi_2 = 0$ and $\gamma \neq 0$). It is therefore possible to say that for these modes, the fluids behave as incompressible. The results obtained in the reference cases are reported in Table 2, where ψ_m and ψ_2 are expressed in degrees,

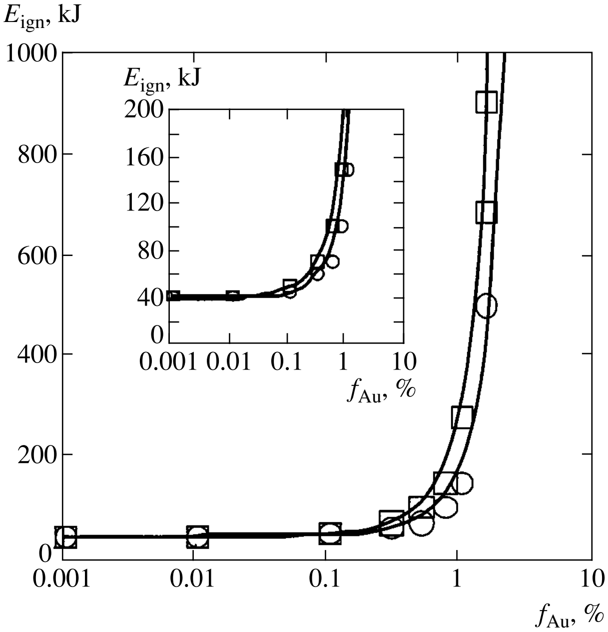


Fig. 12. The ignition energy for 1 MeV protons represented as a function of the Au doping. Squares correspond to the fuel at the assigned density (ρ = const = 200 g/cm^3); circles, to fuel at the assigned pressure (p = const). In both cases, the ignition threshold represents a steep increase as f_{Au} approaches 0.2% (see magnification in the inset). The curves are exponential fits.

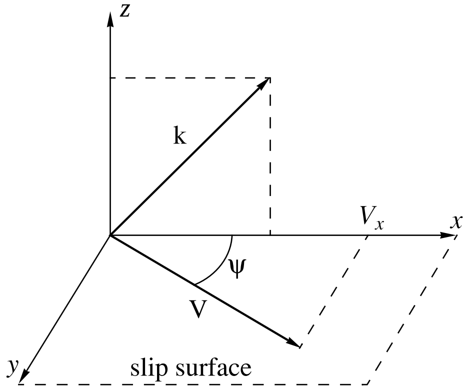


Fig. 13. Geometry for the Kelvin–Helmholtz instability calculations.

$$\gamma_0 = \frac{\gamma_m}{k_x c_{s\mathrm{DT}}}, \quad \tau = \frac{\Delta}{\gamma_m t_{\mathrm{transit}}\lambda_x},$$

Δ is the shell thickness, $t_{\mathrm{transit}} = \Delta/V_0$, and $\lambda_x = 2\pi/k_x$. When $\tau(\lambda_x/\Delta) < 1$, the instability arises.

Table 2

Reference case	ψ_m	ψ_2	γ_0	τ
First shock wave	86	70	0.046	10
End of accelerating pulse	82	73	0.11	5
Near cone tip transit	48	0	0.24	1

In conclusion, the slip surface is always unstable. The maximum growth rate is found in the regime near the tip cone, where all perturbations with $0° \le \psi < 90°$ are unstable. The formation of mixing layers like that represented in Fig. 2b can be expected over most of the implosion process.

Semiempirical theory [10] and numerical simulations [11, 12] have been used in the description of mixing layers. Quoted in [10] are experiments and theories for turbulent mixing in the case of large differences in density or supersonic flows. Difference in density seems to slightly affect the width of the mixing layer (halfwidth h), whereas this becomes somewhat narrower for supersonic flows.

In what follows, we adopt a tentative dimensional model to estimate h in the reference cases. We assume that

$$h^2 \approx D(x/V), \quad D = a\gamma_m h^2,$$

a is a suitable number, x is the distance from O along $\mathbf{V}$ (see Fig. 2b), and γ_m is evaluated for a typical wavenumber $k_x \approx h^{-1}$. Because

$$\gamma_m \propto h^{-1},$$

we find that

$$h \propto x.$$

For $a \approx 0.2$, the previous model applied to incompressible flows with a uniform density agrees with experiment [10]. In the cases considered here, the angle $\theta_{\text{mix}} = 2h/x$ is

$$\theta_{\text{mix}} = 2a\gamma_0 c_{s\text{DT}}/V.$$

The maximum thickness of the mixing layer is

$$\delta_{\text{mix}} \approx \theta_{\text{mix}}\Delta;$$

in the previously assumed reference situations, the results, ordered for increasing pressure, are $\theta_{\text{mix}} \approx 5$, 10, and 70 mrad, and hence, for $\Delta = 100$ μm, we have $\delta_{\text{mix}} = 0.5$, 1, and 7 μm.

It seems safe to conclude that at the sliding surface, the Kelvin–Helmholtz instability, which is active during the shell implosion, is especially effective in the "near tip" region, where a substantial amount of matter can be collected and mixed into the fuel. An indication for this can be found by the following considerations.

Typically, the fuel masses of interest in CFFI (M_f) are on the order of 3 mg at a density of about 200 g/cm^3. The typical spark masses (M_{spark}) are on the order of 3 μg. The quantities M_f and M_{spark} are to be compared with that of gold (M_{Au}) involved in the mixing. Estimating the thickness of the mixing zone as $\delta_{\text{mix}} \approx \theta_{\text{mix}}\Delta$, the volume involved in mixing by a shell sweeping the cone surface by a length L along the cone up to the cone apex is

$$V_{\text{mix}} = \pi \sin\theta L^2 \delta_{\text{mix}},$$

where θ is the cone half aperture. The amount of gold involved in mixing can be estimated as $0.5\rho_{\text{Au}}V_{\text{mix}}$, and assuming that a fraction f is entrained to the fuel, it is found that

$$M_{\text{Au}} = 0.5\rho_{\text{Au}}V_{\text{mix}}f.$$

If this estimate is applied to the case of a near tip for a cone with $\theta = 30°$, at $L = 400$ μm and $f = 0.3$, it follows that $M_{\text{Au}} \approx 30$ μg or $M_{\text{Au}} \approx 10M_{\text{spark}}$. In the ignition study, it was found that a 14% mass contamination was sufficient to make ignition practically impossible. Under the previous conditions, this critical situation can be created at a mass of about 30 times that of the spark. Clearly, the relevance of Kelvin–Helmholtz instability can be completely assessed only after using a more complete description that would include the final space distribution of the contaminant. However, the previous estimate indicates that substantial amounts of contaminant can be potentially involved in the process.

5. CONCLUSIONS

The process of fast ignition by light ions has been studied, and the ignition thresholds for different ion kinetic energies have been found for a fuel composed of 50% deuterium and 50% tritium. The self-consistent structure of the igniting spark has also been determined by analytic methods. The physics of the burn propagation has been studied in some detail as a function of the irradiation parameters. Because the cone material can contaminate the fuel in CFFI, a parametric study has been performed to find how the ignition thresholds depend on the level of contamination. This study, performed for an atomically mixed contaminant, has shown that a level of atomic contamination of about 0.2% is sufficient to render ignition practically impossible.

Evaluations have also been performed to find the possible effects on ignition of contaminant blobs mixed into a pure DT fuel.

Finally, it has been shown that in addition to cone vaporization due to X-rays passing through the imploding shell in the indirect drive mode, another source of contamination can be a turbulent mixing at the interface cone-imploding shell. This mechanism could also be active for directly driven implosions.

ACKNOWLEDGMENTS

We are grateful to N.V. Zmitrenko (IMM, Russian Academy of Sciences) for very useful discussions on the Kelvin–Helmholtz instability.

REFERENCES

1. A. Caruso and C. Strangio, Laser Part. Beams **19**, 295 (2001).
2. R. Kodama, P. A. Norreys, K. Mima, *et al.*, Nature **412**, 798 (2001).
3. S. Hatchett *et al.*, in *Proceedings of 43rd Meeting on Division of Plasma Physics, Long Beach, California* (Am. Phys. Soc., 2001).
4. R. B. Stephens *et al.*, Preprint GA-A24140 (2002).
5. A. Caruso and R. Gratton, Phys. Lett. A **36**, 275 (1971).
6. Ya. B. Zel'dovich and Yu. P. Raĭzer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, 2nd ed. (Nauka, Moscow, 1966; Academic, New York, 1966 and 1967), Vols. 1 and 2.
7. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics*, 4th ed. (Nauka, Moscow, 1988; Pergamon, New York, 1987).
8. M. Roth, T. E. Cowan, M. H. Key, *et al.*, Phys. Rev. Lett. **86**, 436 (2001).
9. R. A. Sacks and D. H. Darling, Nucl. Fusion **27**, 447 (1987).
10. H. Schlichting, *Boundary Layer Theory*, 6th ed. (McGraw-Hill, New York, 1968; Nauka, Moscow, 1974).
11. V. A. Andronov *et al.*, Preprint LA-12896 (1995).
12. M. Lesieur, *Turbulence in Fluids*, 2nd ed. (Kluwer Academic, Dordrecht, 1990).

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 958–968.
Translated from Zhurnal Éksperimental’noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 1068–1079.
Original Russian Text Copyright © 2003 by Rosenfeld.

SOLIDS
Structure

Physical Reasons for the Emergence of Jumps during Rotation of the Magnetization of a Two-Sublattice Ferrimagnet at Low Temperatures

E. V. Rosenfeld
Institute of Metal Physics, Ural Division, Russian Academy of Sciences,
ul. S. Kovalevskoĭ 18, Yekaterinburg, 620219 Russia
e-mail: rosenfeld@imp.uran.ru

Received January 27, 2003

Abstract—It is shown that the contribution of the intersublattice exchange interaction to the magnetic anisotropy energy of a two-sublattice ferrimagnet can come only from higher order constants satisfying the condition $2K_2 + 3K_3 + \ldots > 0$. For this reason, for different signs of the first anisotropy constants of the sublattices, this contribution may cause a spontaneous spin-reorientation second-order transition, but not first-order transitions and jumps during magnetization rotation, which are associated with such a transition. Such jumps can appear only when the opposite inequality is satisfied, and the corresponding contribution to anisotropy can be ensured only by a fairly strong magnetoelastic interaction. *© 2003 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

The field dependences of magnetization of crystals of the hexagonal system are usually described on the basis of the standard phenomenological expansion of their energy,

$$\mathscr{E}(\mathbf{H}, \mathbf{M}) = K_1(M)\cos^2\theta + K_2(M)\cos^4\theta + \ldots - \mathbf{H}\cdot\mathbf{M}. \quad (1)$$

The anisotropy constants K_1, K_2, etc., appearing in this expansion, can be determined from the condition of the best coincidence of the dependences of magnetization **M** on the external magnetic field **H**, obtained as a result of minimization of expression (1), with the experimental magnetization curves. Usually, it is sufficient to confine analysis to the first two or three terms of the expansion, where the anisotropy constants K_2 and K_3, which describe the same contributions to energy that determine the anisotropy of cubic crystals, are small as compared to K_1. However, experimental investigations of the behavior of compounds of rare-earth (R) and transition (T) 3*d* metals often reveal anomalies, which have to be described using values of K_2 comparable to or even exceeding the value of K_1.

Such an increase in the modulus of the ratio $|K_2/K_1|$ always indicates the possibility of a spin-reorientation transition. Depending on the sign of K_2, this can be either a second-order transition ($K_2 > 0$) or a first-order transition ($K_2 < 0$). In the present study, we are mainly interested in the first-order magnetization processes (FOMP), i.e., in the magnetization jumps during magnetization rotation. In this case, the angular dependence of energy (1) acquires two minima (Fig. 1a). As a result, in an external field directed approximately along the “hard” axis, the smooth rotation of magnetization experiences a jump (Fig. 1b). The list of compounds exhibiting such jumps can be found in [1].

In zero magnetic field, two minima appear on the angular dependence of energy only if the following condition is satisfied [2]:

$$0 < K_1 < -2K_2. \quad (2)$$

The meaning of this condition is very simple. The term in expansion (1) proportional to $\cos^4\theta$ (or $\sin^4\theta$ if we use a power expansion in $\sin\theta$) describes a curve which is a second-order parabola near one of its extrema and a fourth-order parabola near the other extremum. The latter extremum is found to be “flat” since the curvature of the curve $K_2\cos^4\theta$ at this point is equal to zero, and the main of the three condition combined in (2) ($K_2 < 0$) requires that this extremum is a maximum. If the first term has a minimum in this region ($K_1 > 0$), this minimum will be preserved after the summation with the flat maximum. If, in addition, the first term is not very large, $K_1 < 2|K_2|$, the minimum located at the foot of the flat peak is also preserved (see Fig. 1a). In this simple model, two minima always correspond to the directions of magnetization **M** along ($\theta = 0$) and across ($\theta = \pi/2$) the hexagonal axis *c* (*z*). Supplementing expansion (1) with the third term, $K_3\cos^6\theta$, we can shift the minima to the region $0 < \theta < \pi/2$ [3].

Condition (2) is sufficient but not necessary for the emergence of magnetization jumps. The second minimum on the $\mathscr{E}(\theta)$ curve may also appear after the appli-

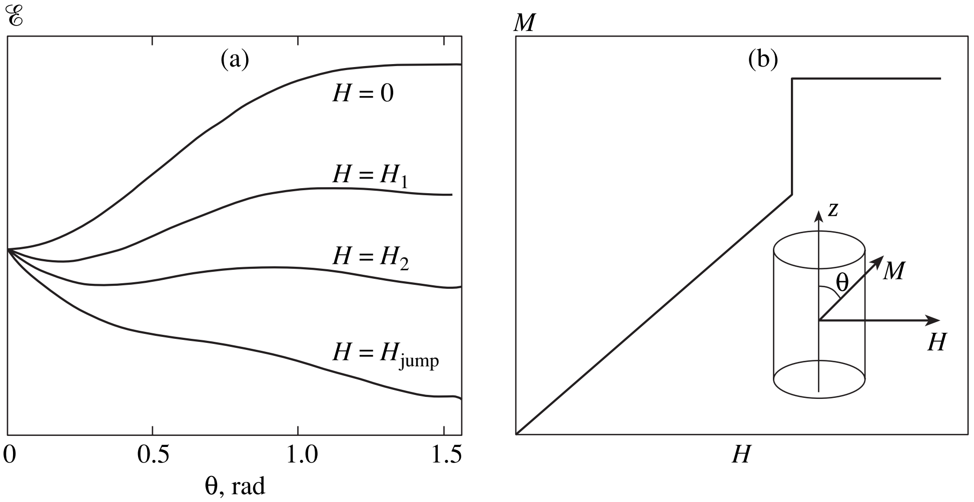


Fig. 1. (a) Angular dependence of energy and (b) projection of magnetization on the external field applied along the hard axis as a function of the field strength $0 < H_1 < H_2 < H_{\text{jump}}$. A jump appears when the metastable state vanishes in the field $H = H_{\text{jump}}$.

cation of an external field in the hard direction. In this case, in zero field, the curvature of the $\mathcal{E}(\theta)$ curve near the maximum must be smaller than near the minimum (the curve with a flatter peak). If the Zeeman energy in the region of the flatter peak is taken into account, the second minimum appears if

$$K_2 < 0, \quad 4K_2 < K_1 < -6K_2. \tag{3}$$

Thus, only the condition $K_2 < 0$ is necessary for the emergence of magnetization jumps. If the opposite inequality $K_2 > 0$ holds, it is not the maximum but the minimum on the $\mathcal{E}(\theta)$ dependence that turns out to be flatter. In this case, there is only one minimum; however, as the value of K_2 increases, this minimum can move to the region $0 < \theta < \pi/2$, indicating the emergence of a spontaneous (zero-field) spin-reorientation second-order transition (Fig. 2).

Such an approach, which is ideal from the viewpoint of combining simplicity and quality of fitting in the phenomenological description, does not explain the physical reasons for the emergence of magnetization jumps. As a matter of fact, there are no grounds to expect the emergence of large second-order constants in the R or T sublattice. In the former case, this is hardly probable in accordance with the crystal field theory [4, 5], while in the latter case, in accordance with experimental evidence for similar compounds of T metals with nonmagnetic yttrium and lanthanum, only one first-order constant is sufficient for describing anisotropy. Consequently, neither the R nor the T sublattice alone may have two minima on the angular dependence of energy. Hence, a meticulous analysis of all contributions to the magnetic anisotropy energy is required to find out which interactions are responsible for the emergence of a large second-order constant in general and for the formation of metastable states in particular.

The two main contributions to the magnetic anisotropy energy come from the exchange interactions between magnetic moments and from their interaction with the lattice. These contributions will be considered in the next two sections of this article. Confining the analysis to the low-temperature region, we will disregard the change in the exchange energy associated with the change in the modulus of spin during its rotation [6, 7]. For this reason, we will henceforth assume that the exchange contribution comes only from the intersublat-

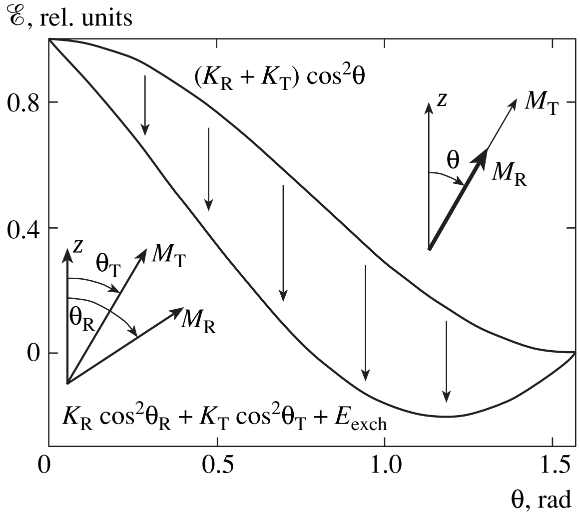


Fig. 2. In the case of an infinitely strong intersublattice exchange interaction, the magnetic moments of the sublattices are parallel and the anisotropy constant of the crystal is equal to the sum of the sublattice constants (upper curve). As the exchange energy decreases, the moments of the sublattices start deviating towards their easy axes in the region $0 < \theta < \pi/2$. This gives rise to a new degree of freedom, which lowers the energy (lower curve). As a result, the maximum becomes steeper and the minimum becomes more gently sloping or is even shifted to the region $0 < \theta < \pi/2$.

tice exchange interaction whose energy changes with the angle between the spins of the sublattices. This contribution to the second anisotropy constant of the crystals is always positive and is responsible for the second-order spin-reorientation transition. As regards magnetization jumps, the only possible negative contribution to K_2 comes from the magnetoelastic interaction (if we disregard, in contrast to [1], the possibility of formation of three or more sublattices with specific exchange interactions between these sublattices). The conclusions drawn will be used for describing the magnetization curves for compounds $\mathrm{Nd}_x\mathrm{Y}_{1-x}\mathrm{Co}_5$.

2. CONTRIBUTION OF INTERSUBLATTICE EXCHANGE TO ANISOTROPY

We will first consider a classical two-sublattice magnet, i.e., a magnet with a very strong exchange interaction in the sublattices. In this case, their magnetization $\mathbf{M}_\mathrm{R}$ and $\mathbf{M}_\mathrm{T}$ can be treated as classical vectors whose magnitudes depend only on the temperature. It is clear that, if $\mathbf{M}_\mathrm{R}$ and $\mathbf{M}_\mathrm{T}$ are not parallel to the crystal axes, these vectors cannot be parallel to each other and the angle between these vectors changes upon their rotation. It is shown in [8] that in such a situation we must explicitly take into account the contribution $-I_\mathrm{RT}(\mathbf{M}_\mathrm{R}\cdot\mathbf{M}_\mathrm{T})/M_\mathrm{R}M_\mathrm{T}$ from the intersublattice exchange interaction to anisotropy, where I_RT is the constant of this interaction. This contribution is most significant when the easy axes of the sublattices are mutually perpendicular (precisely this case will be considered below).

Confining our analysis to the first anisotropy constants K_1R and K_1T of the sublattices and denoting by θ_R, θ_T, and θ_H the angles formed by vectors $\mathbf{M}_\mathrm{R}$, $\mathbf{M}_\mathrm{T}$, and $\mathbf{H}$ with the c axis, we write the energy of the system in the form

$$\begin{aligned}\mathscr{E}(\theta_\mathrm{R},\theta_\mathrm{T}) &= -I_\mathrm{RT}\cos(\theta_\mathrm{R}-\theta_\mathrm{T})\\ &-K_\mathrm{1T}\cos^2\theta_\mathrm{T}+K_\mathrm{1R}\cos^2\theta_\mathrm{R}\\ &-H\{M_\mathrm{R}\cos(\theta_H-\theta_\mathrm{R})+M_\mathrm{T}\cos(\theta_H-\theta_\mathrm{T})\}.\end{aligned}\tag{4}$$

This function of two variables can easily be reduced to a function of one variable, similar to (1):

$$\begin{aligned}\mathscr{E}(\theta_\mathrm{R},\theta_\mathrm{T}) \longrightarrow \mathscr{E}(\theta_\mathrm{T}) &= \min\{E_\mathrm{R}(\theta_\mathrm{R},\theta_\mathrm{T})\}\\ &-K_\mathrm{1T}\cos^2\theta_\mathrm{T}-HM_\mathrm{T}\cos(\theta_H-\theta_\mathrm{T}),\\ E_\mathrm{R}(\theta_\mathrm{R},\theta_\mathrm{T}) &= -I_\mathrm{RT}\cos(\theta_\mathrm{R}-\theta_\mathrm{T})\\ &+K_\mathrm{1R}\cos^2\theta_\mathrm{R}-HM_\mathrm{R}\cos(\theta_H-\theta_\mathrm{R}).\end{aligned}\tag{5}$$

Minimization is carried out here over θ_R; i.e., we assume that any rotation of vector $\mathbf{M}_\mathrm{T}$ is accompanied by the rotation of vector $\mathbf{M}_\mathrm{R}$ to the position corresponding to the minimal value of energy E_R. If we now assume that the Zeeman energy is small as compared to the anisotropy constants of the sublattices, and the latter constants are smaller than the intersublattice exchange energy, we can carry out this minimization in analytic form in the lowest order in $K_\mathrm{1R}/I_\mathrm{RT}$:

$$\begin{aligned}\sin(\theta_\mathrm{R}-\theta_\mathrm{T}) &\approx \frac{K_\mathrm{1R}}{I_\mathrm{RT}}\sin(2\theta_\mathrm{T}),\\ \mathscr{E}(\theta_\mathrm{T}) &\approx (K_\mathrm{1R}-K_\mathrm{1T})\cos^2\theta_\mathrm{T}-\frac{(K_\mathrm{1R})}{2|I_\mathrm{RT}|}\\ &\times\sin^2(2\theta_\mathrm{T})-H|M_\mathrm{R}\pm M_\mathrm{T}|\cos(\theta_H-\theta_\mathrm{T}).\end{aligned}\tag{6}$$

Since the angle between vectors $\mathbf{M}_\mathrm{R}$ and $\mathbf{M}_\mathrm{T}$ is small in this case, the crystal magnetization is equal simply to the sum or difference of the magnetizations of the sublattices depending on the sign of I_RT. Comparing relations (1) and (6), we obtain the total anisotropy constants for a crystal with energy (1) as functions of the anisotropy constants of the sublattices and the exchange parameter:

$$K_1 \approx K_\mathrm{1R}-K_\mathrm{1T}-2\frac{K_\mathrm{1R}^2}{|I_\mathrm{RT}|},\quad K_2 \approx 2\frac{K_\mathrm{1R}^2}{|I_\mathrm{RT}|}.\tag{7}$$

Thus, the contribution to K_2 emerging due to the intersublattice exchange interaction is always positive in this approximation.

It can easily be verified that this conclusion remains valid for any value of I_RT. Indeed, for an infinitely strong intersublattice exchange interaction, the moments of the sublattices are always exactly parallel to each other. The crystal energy in this case is equal to the sum of the anisotropy energies of the sublattices (see the upper curve in Fig. 2). If the intersublattice exchange becomes finite, a new degree of freedom appears since the angle between the magnetizations of the sublattices may change. However, the emergence of new degrees of freedom for any system leads only to a decrease in its energy. Consequently, when the exchange becomes finite, the energy must become lower, which is indicated by arrows in Fig. 2. However, in the high-symmetry directions (along the c axis or in the basal plane), the moments of the sublattices always remain parallel and the difference in the energies corresponding to these states remains unchanged. In any other positions, the moments start deviating toward their easy axes, the angle between these moments increases, and the energy decreases. Obviously, the energy peak in this case becomes sharper and the energy minimum becomes more gently sloping, which indicates the emergence of a positive anisotropy constant K_2.

We did not depict in Fig. 2 the curve with a more gently sloping minimum, confining our analysis to the case when the value of $|I_{\mathrm{RT}}|$ decreases to such an extent that the following condition holds [8]:

$$|I_{\mathrm{RT}}(K_{1\mathrm{R}} - K_{1\mathrm{T}})| < 2K_{1\mathrm{R}}K_{1\mathrm{T}}. \tag{8}$$

In this case, the energy minimum, remaining the single minimum, is displaced to the interior of the region $0 < \theta < \pi/2$; i.e., a noncollinear magnetic structure is formed as a result of the second-order phase transition. Minimizing Eq. (4) for $H = 0$, we can easily determine the position of the minimum,

$$\begin{gathered}\theta_{\mathrm{R}} = \arcsin\sqrt{\frac{1}{2}\left[1 + \frac{1 - 2\alpha_{\mathrm{R}}(\alpha_{\mathrm{R}} - \alpha_{\mathrm{T}})}{\sqrt{1 + 4\alpha_{\mathrm{R}}\alpha_{\mathrm{T}}}}\right]},\\ \theta_{\mathrm{T}} = \operatorname{sgn} I_{\mathrm{RT}} \arccos(\alpha_{\mathrm{R}} - \alpha_{\mathrm{T}}) - \theta_{\mathrm{R}},\end{gathered} \tag{9}$$

where

$$\alpha_{\mathrm{R}} = \frac{I_{\mathrm{RT}}}{2K_{1\mathrm{R}}}, \quad \alpha_{\mathrm{T}} = \frac{I_{\mathrm{RT}}}{2K_{1\mathrm{T}}}.$$

In this case, the application of expression (5) for energy is less effective in spite of its dependence on only one variable and strongly complicates the derivation of Eqs. (9). However, approximate formulas (6) and (7) are very helpful. These formulas make it possible to estimate the minimal value of the second anisotropy constant, which must be inherent in at least one of the sublattices for the emergence of a metastable state. In the lowest order in $K_{1\mathrm{R}}/I_{\mathrm{RT}}$, condition (2) now assumes the form

$$\begin{gathered}2\frac{K_{1\mathrm{R}}^2}{|I_{\mathrm{RT}}|} < \frac{K_{1\mathrm{R}} - K_{1\mathrm{T}}}{2} + \frac{K_{1\mathrm{R}}^2}{|I_{\mathrm{RT}}|} < -K_{2\nu},\\ \nu = \mathrm{R}, \mathrm{T}.\end{gathered} \tag{10}$$

Thus, the conditions under which even small second anisotropy constants of the sublattices may cause magnetization jumps in a classical two-sublattice magnet still exist. It is necessary that the first anisotropy constants of the sublattices be close in value and have opposite signs, while the exchange interaction between the sublattices must be quite strong. It is only in this case that an essentially positive exchange contribution to the second anisotropy constant of the crystal may not play a decisive role. However, as the value of $|I_{\mathrm{RT}}|$ decreases, this contribution becomes predominant and magnetization jumps cannot appear.

The results obtained for a classical ferrimagnet cannot be used directly for analyzing the magnetic properties of real RT compounds. As a matter of fact, the exchange in the R sublattice is practically absent for the latter compounds. For this reason, the magnetization of this sublattice cannot be treated as a classical vector, which results in a radical change in energy E_{R} in expression (5).

Assuming that only the magnetization of the T sublattice is a classical vector, while R ions are “paramagnetic” in the exchange field created by this lattice, we obtain, instead of Eqs. (4) and (5),

$$\begin{gathered}\mathscr{E}(\theta_{\mathrm{T}}) = F_{\mathrm{R}}(\mathbf{m}) - K_{1\mathrm{T}}\cos^2\theta_{\mathrm{T}} - HM_{\mathrm{T}}\cos(\theta_H - \theta_{\mathrm{T}}),\\ \mathbf{m} = \mathbf{M}_{\mathrm{T}}/M_{\mathrm{T}},\\ F_{\mathrm{R}}(\mathbf{m}) = -T\ln(Q(\mathbf{m})),\\ Q(\mathbf{m}) = \operatorname{Tr}\left\{-\frac{H_{\mathrm{R}}(\mathbf{m})}{T}\right),\\ H_{\mathrm{R}}(\mathbf{m}) = I_J(\mathbf{m}\cdot\mathbf{J}) + k_1J_z^2 + k_2J_z^4\\ + \ldots + k_{66}(J_+^6 + J_-^6) - g_J\mu_B(\mathbf{H}\cdot\mathbf{J}),\end{gathered} \tag{11}$$

where T is the absolute temperature in energy units and k_i are anisotropy parameters. The magnetic moment $\mathbf{M}_{\mathrm{R}}$ of an R ion is now transformed from a classical vector into an ordinary quantum-mechanical mean:

$$\begin{gathered}\mathbf{M}_R = g_J\operatorname{Tr}\{\mathbf{J}\rho(\mathbf{m})\},\\ \rho(\mathbf{m}) = \frac{1}{Q(\mathbf{m})}\exp\left(-\frac{H_{\mathrm{R}}(\mathbf{m})}{T}\right).\end{gathered} \tag{12}$$

Here, $\mathbf{J}$ and $g_J\mathbf{J}$ are the operators of the total angular momentum and the magnetic moment of the R ion,[1] g_J is the Lande factor, and $\mathbf{m}$ is the unit vector directed along the magnetic moment $\mathbf{M}_{\mathrm{T}}$ of T sites.

If the intersublattice exchange I_J is very strong, the free energy F_{R}, which coincides with the ground-state energy of the R ion at $T = 0$, is given in the first order of perturbation theory by

$$\begin{gathered}F_{\mathrm{R}}(\theta) = K_{1\mathrm{R}}\cos^2\theta + K_{2\mathrm{R}}\cos^4\theta + \ldots\\ \approx [k_1 + k_2(3J - 2)]J\left(J - \frac{1}{2}\right)\cos^2\theta\\ + k_2J\left(J - \frac{1}{2}\right)(J - 1)\left(J - \frac{3}{2}\right)\cos^4\theta + \ldots.\end{gathered} \tag{13}$$

In this case, condition (2) for the emergence of two energy minima can be written in explicit form. If the energy of the R sublattice must have two minima, we obtain the following condition from relations (2) and (13):

$$0 < k_1 < -k_2(J - 1)(2J - 3).$$

---

[1] We assume that these vectors are parallel.

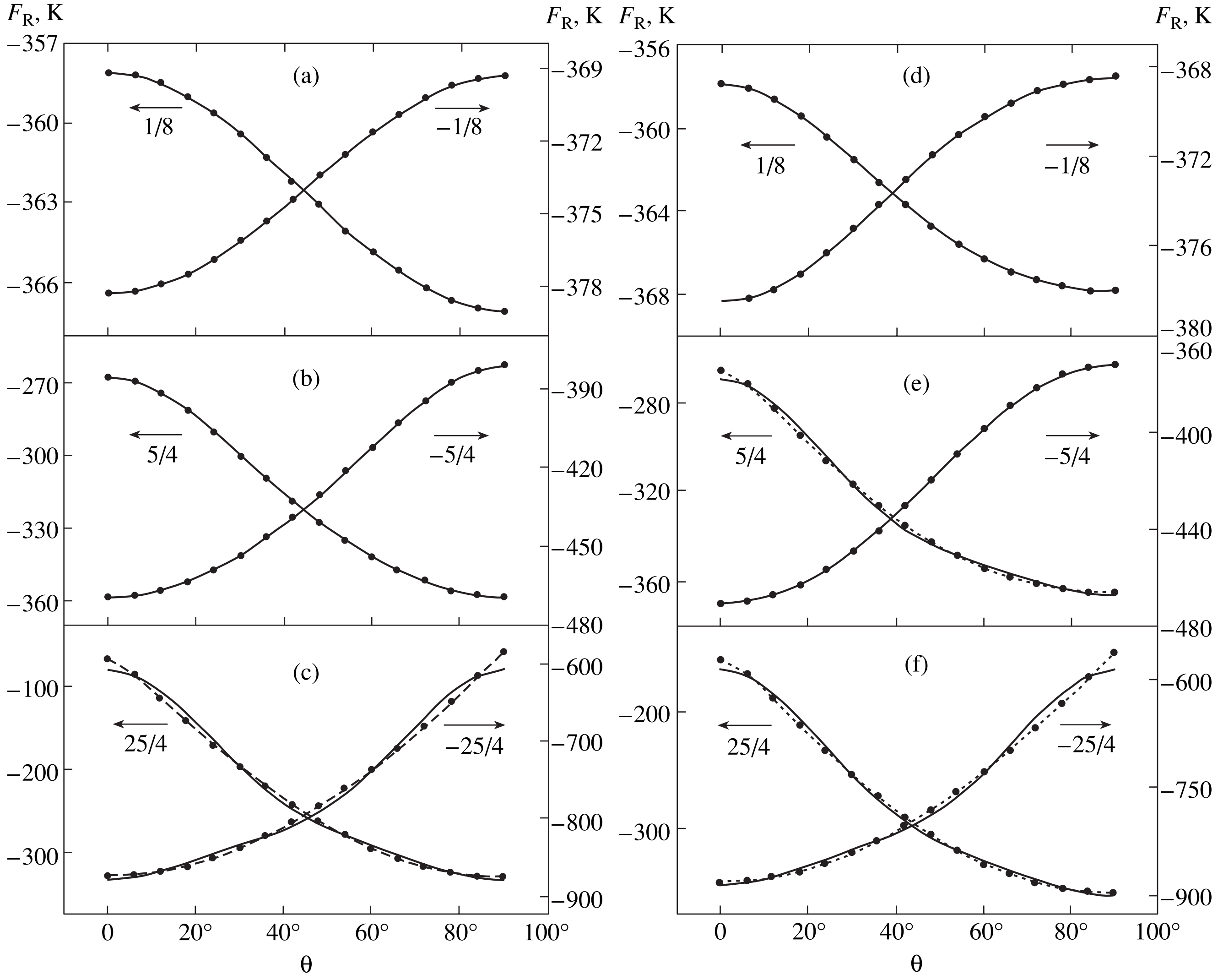


Fig. 3. Dependences $F_R(\theta)$ at $T = 4$ K, $J = 9/2$. $g_J = 8/11$. And $I = 300$ K in the cases when only k_1 (a–c) and only k_2 (d–f) differ from zero in Hamiltonian (11). The dashed curves with circles correspond to the results of calculation, while solid curves describe approximation by functions (16)–(18) and (19)–(21), respectively, for cases (a–c) and (d–f). The values of k_1J^2/I_J (a–c) and k_2J^4/I_J (d–f) are shown near the arrows.

In the general case, $|k_1| \gg |k_2|J^2$ for hexagonal lattices [4, 5], so that the R sublattice in this model cannot be a trigger putting in operation the mechanism of magnetization jump formation. However, the total energy $\mathscr{E}(\theta_T)$ in Eqs. (11) in which the contribution $F_R(\theta_T)$ plays the same role as $K_2\cos^4\theta$ in Eq. (1) may have two minima also. For $H = 0$, function $F_R(\theta_T)$ must have a flatter peak at $\theta_T = 0$ and a sharp minimum at $\theta_T = \pi/2$:

$$0 < -\frac{\partial^2 F_R}{\partial \theta_T^2}\bigg|_{\theta_T = 0} < 2K_{1T} < \frac{\partial^2 F_R}{\partial \theta_T^2}\bigg|_{\theta_T = \pi/2}. \tag{14}$$

In this case, when the anisotropy energy $-K_{1T}\cos^2\theta_T$ of the T sublattice is added to $F_R(\theta_T)$, the minimum at $\theta_T = \pi/2$ is preserved and another minimum is formed at $\theta_T = 0$. Substituting relation (13) into (14), we obtain

$$K_{1R} + 2K_{2R} < K_{1T} < K_{1R}. \tag{15}$$

Consequently, as in the case of a classical ferrimagnet, we again come to the conclusion that a magnetization jump could in principle emerge if the intersublattice exchange were strong and the first anisotropy constants compensated each other almost exactly. However, we are not aware of two-sublattice systems in which the existence of magnetization jumps would be indeed associated with such a compensation of the first sublattice anisotropy constants. The anisotropy energy in such a compound would be much lower that in a similar compound of a T metal with yttrium or lanthanum; the experimentally observed picture is quite different.

In real RT compounds, the anisotropy of the R sublattice is often on the same order of magnitude as the intersublattice exchange; we numerically calculated the angular dependence of F_R just in this range of the parameters. For the R element, we took Nd ($J = 9/2$ and $g_J = 8/11$) and carried out calculations for various values of the ratio k_ν/I_J, $\nu = 1, 2$.

If we do not confine our analysis to the lowest order in perturbation theory, any operator from $k_m J_z^{2m}$ in the Hamiltonian H_R (11) is sufficient for the simultaneous emergence of all the terms proportional to even powers of $\cos\theta$ in expansion (13). However, the relation between constants K_{jR} is obviously determined to a considerable extent by a parameter k_m that differs from zero and by the ratio of this parameter to I_J.

Figures 3a–3c show the results of calculation of the $F_R(\theta)$ dependence for the case when only k_1 in Eqs. (11) differs from zero. It can be seen that, in accordance with relation (13), for small values of $|k_1|/I_J$, we obtain a symmetric function proportional to $\cos^2\theta$. However, as the ratio $|k_1|/I_J$ increases, the peak of the function $F_R(\theta)$ becomes steeper and steeper, while the minimum becomes more and more gently sloping. This tendency is preserved even after the ground state changes, i.e., for $k_1 > I_J/(2J - 1)$. Naturally, the role of higher order constants in expansion (13) increases simultaneously. The results of fitting of the numerically calculated functions $F_R(\theta)$ to the first three terms in expansion (13) are given below:

$$\begin{aligned} F_R(\theta)|_{k_1 = -0.5\text{ K}} &\approx -370 - 9.43\cos^2\theta + 0.48\cos^4\theta - 0.04\cos^6\theta, \\ F_R(\theta)|_{k_1 = 0.5\text{ K}} &\approx -367 + 8.61\cos^2\theta + 0.37\cos^4\theta + 0.04\cos^6\theta \end{aligned} \tag{16}$$

in Fig. 3a,

$$\begin{aligned} F_R(\theta)|_{k_1 = -5\text{ K}} &\approx -382 - 142\cos^2\theta + 87\cos^4\theta - 32.6\cos^6\theta, \\ F_R(\theta)|_{k_1 = 5\text{ K}} &\approx -358 + 68.5\cos^2\theta - 16.3\cos^4\theta + 38.7\cos^6\theta \end{aligned} \tag{17}$$

in Fig. 3b, and

$$\begin{aligned} F_R(\theta)|_{k_1 = -25\text{ K}} &\approx -607 - 721\cos^2\theta + 878\cos^4\theta - 430\cos^6\theta, \\ F_R(\theta)|_{k_1 = 25\text{ K}} &\approx -333 + 220\cos^2\theta - 305\cos^4\theta + 338\cos^6\theta \end{aligned} \tag{18}$$

in Fig. 3c. It should be noted here that the conditions for the emergence of more gently sloping maxima and minima at $\theta = 0, \pi/2$ change in expansions with a number of terms exceeding two (instead of $K_2 < 0$ ($K_2 > 0$), we have $2K_2 + 3K_3 + \ldots < 0$ ($2K_2 + 3K_3 + \ldots > 0$)).

Figures 3d–3f show the results of such computations for the case when only the value of k_2 in expression (11) for H_R differs from zero. The approximation was carried out using the following functions:

$$\begin{aligned} F_R(\theta)|_{k_2 = -0.025\text{ K}} &\approx -368.5 - 5.33\cos^2\theta + 5.44\cos^4\theta + 0.88\cos^6\theta, \\ F_R(\theta)|_{k_2 = 0.025\text{ K}} &\approx -367.8 + 5.11\cos^2\theta + 3.78\cos^4\theta + 1.02\cos^6\theta \end{aligned} \tag{19}$$

in Fig. 3d,

$$\begin{aligned} F_R(\theta)|_{k_2 = -0.25\text{ K}} &\approx -372 - 104\cos^2\theta - 12.6\cos^4\theta + 18.2\cos^6\theta, \\ F_R(\theta)|_{k_2 = 0.25\text{ K}} &\approx -366 + 70.7\cos^2\theta - 103\cos^4\theta + 129\cos^6\theta \end{aligned} \tag{20}$$

in Fig. 3e, and

$$\begin{aligned} F_R(\theta)|_{k_2 = -1.25\text{ K}} &\approx -587 - 778\cos^2\theta + 935\cos^4\theta - 456\cos^6\theta, \\ F_R(\theta)|_{k_2 = 1.25\text{ K}} &\approx -3589 + 156\cos^2\theta - 210\cos^4\theta + 249\cos^6\theta \end{aligned} \tag{21}$$

in Fig. 3f. In this case, for a small ratio $|k_2J^2/I_J|$, the term proportional to $\cos^4\theta$ plays the leading role and the $F_R(\theta)$ dependence may acquire a flatter peak (see Fig. 3d). However, as the value of $|k_2J^2/I_J|$ increases, the contribution from the corrections of the next orders of magnitude becomes larger, leading to a deviation of $\mathbf{M}_R$ from $\mathbf{M}_T$. As a result, the expression for F_R acquires a considerable exchange contribution and the curvature of the $F_R(\theta)$ curve in the vicinity of the maximum turns out to be larger than in the vicinity of the minimum for any sign of k_2 (see Figs. 3e and 3f). It should be emphasized that this conclusion is valid even for large negative values of k_2 (or, which is the same, K_{2R}; see relation (13)) for which the function $\min\{E_R\}$ in relations (5) would have a more gently sloping peak in the classical case.

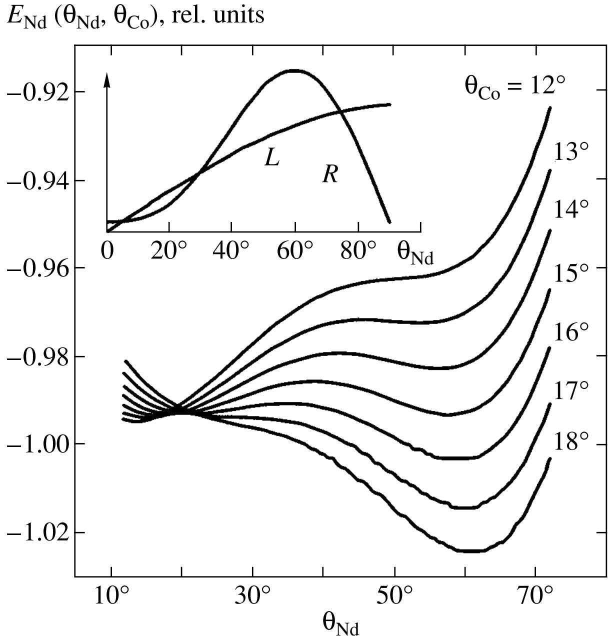


Fig. 4. Dependence of energy $E_{Nd}(\theta_{Nd}, \theta_{Co})$ of the $NdCo_5$ alloy on θ_{Nd} (first equation in (22)) for the values of the parameters chosen in [10]. The inset shows the graphic solution of the second equation in (22); letters *L* and *R* mark the curves corresponding to the left- and right-hand sides.

Thus, a decrease in the intersublattice exchange in a ferrimagnet with a paramagnetic R sublattice, as well as in a classical two-sublattice ferrimagnet, leads to an increase in higher order anisotropy constants of the crystal. The magnitudes of these constants may become equal to or even exceed the first constant, but almost in all cases these constants satisfy the condition $2K_2 + 3K_3 + \ldots > 0$. Consequently, the intersublattice exchange interaction energy can only lead to the emergence of a second-order spin-reorientation transition. A first-order transition may take place only for $2K_2 + 3K_3 + \ldots < 0$; however, in order to explain this transition, we must consider some other interactions that have been disregarded so far.

3. MAGNETIZATION JUMPS IN THE SYSTEM OF $Nd_xY_{1-x}Co_5$ ALLOYS

At the end of the 1970s, Ermolenko [9] discovered strong magnetization jumps in alloys of the system $Nd_xY_{1-x}Co_5$ with $x \le 0.4$, which were subsequently interpreted [10] on the basis of the model of a classical two-sublattice magnet. In such alloys, Co atoms exhibit an easy axis anisotropy, Nd atoms exhibit an easy plane anisotropy, while Y atoms are a nonmagnetic analog of neodymium atoms. For this reason, the magnetization and the anisotropy constant of the R sublattice, as well as the energy of the intersublattice exchange, are proportional to the concentration *x* of Nd atoms in this system of alloys.

The dependences of the total energy on θ_{Co} calculated in [10] for all alloys with $x > 0.1$ have two minima, one of which lies at $\theta_{Co} = 0$ and the other is shifted towards $\theta_{Co} = \pi/2$ with increasing *x*. This made it possible not only to explain the jumps on the magnetization curves for alloys with $x \le 0.4$, but also to predict the existence of a magnetization jump in the alloy $NdCo_5$ in a field having a strength on the order of 350 kOe and parallel to the *c* axis. This prediction was subsequently confirmed experimentally [11], which suggests that the set of parameters and the mechanism for the emergence of magnetization jumps proposed in [10] should be considered again more attentively.

According to Ermolenko [10], the key point is that the first anisotropy constant of the Nd sublattice is assumed to be zero. This value was chosen since the magnetization curves for alloys with different values of *x* had approximately the same slope for small angles of deviation of the magnetization from the *c* axis; in other words, the contribution from the Nd sublattice to anisotropy was equal to zero for small values of θ_{Nd}. From the viewpoint of the above analysis, this means that function $F_{Nd}(\theta)$ has a flatter peak in this case. Ermolenko [10] obtained the value of this maximum using model (4) and introducing a term proportional to $\sin^4\theta_{Nd}$. As a result, instead of relations (5), the following relations are obtained using the sine expansion adopted in [10]:

$$\begin{aligned} E_{Nd}(\theta_{Nd}, \theta_{Co}) &= -I_{NdCo}\cos(\theta_{Nd} - \theta_{Co}) \\ &\quad - K_{2Nd}\sin^4\theta_{Nd}, \\ \sin(\theta_{Nd} - \theta_{Co}) &= \frac{4K_{2Nd}}{I_{NdCo}}\sin^3\theta_{Nd}\cos\theta_{Nd}. \end{aligned} \tag{22}$$

It can easily be verified that if the intersublattice exchange I_{NdCo} is not very strong, the last equation in θ_{Nd} has three solutions in a certain range of θ_{Co}. It is this situation that takes place for the values of parameters chosen in [10] (see the inset to Fig. 4). For $\theta_{Co} < 12°$, vector $\mathbf{M}_{Nd}$ may lie only in the vicinity of the *c* axis, while for $\theta_{Co} > 18°$, this vector lies only in the vicinity of the basal plane. However, for $12° < \theta_{Co} < 18°$, vector $\mathbf{M}_{Nd}$ has two stable states simultaneously (Fig. 4).

Thus, the magnetization $\mathbf{M}_{Nd}$ of the Nd sublattice in the model used in [10] behaves as a trigger. It performs jumplike transitions from the position near the *c* axis to the position near the basal plane and back when vector $\mathbf{M}_{Co}$ rotates only through 6° from $\theta_{Co} = 12°$ to $\theta_{Co} = 18°$. Accordingly, the surface describing the total energy

$\mathscr{E}(\theta_{\mathrm{Nd}}, \theta_{\mathrm{Co}})$ has two local minima separated by a narrow "ridge" in the band $12° < \theta_{\mathrm{Co}} < 18°$.

It was mentioned above that the assumption used in [10] not only helped to explain the experimental results available at that time, but also led to correct predictions. However, the analysis carried out by us here indicates that the magnetization of the R sublattice cannot behave in such a manner. Indeed, on the one hand, the existence of a crystal field in which $|K_{2\mathrm{R}}| \gg |K_{1\mathrm{R}}|$ is highly improbable. On the other hand, it was emphasized at the end of the previous section that, even in this case, the flat energy peak for "paramagnetic" R ions emerges if anisotropy is small as compared to the intersublattice exchange. Since R ions in real compounds of the $\mathrm{RCo_5}$ type are indeed virtually paramagnetic, while the intersublattice exchange is relatively weak, the R sublattice itself cannot be a trigger and cannot ensure the existence of a metastable state for the Co sublattice.

Thus, it becomes important to construct a noncontradictory model in which the contribution of the R sublattice to the energy of the $\mathrm{Nd}_x\mathrm{Y}_{1-x}\mathrm{Co}_5$ system would ensure the angular dependence of the total energy close to that obtained in [10]. The only contribution to the anisotropy that has not been included so far is the magnetoelastic interaction; therefore, it is natural to attribute the magnetization jumps precisely to this interaction. Indeed, a rotation of the magnetization vector deforms the crystal lattice so that the total energy of the crystal decreases. In particular, in the vicinity of the hard axis, magnetostriction always reduces the anisotropy constant and flattens the energy peak. This effect, if it is strong enough, automatically ensures the emergence of magnetization jumps, imparting stability to the direction of magnetization, which was formerly hard. Consequently, it only remains to verify whether the value of the magnetoelastic energy can be sufficiently high.

In order to solve this problem, let us obtain the simplest estimates proceeding in the framework of the general theory [12]. In order to estimate the main magnetoelastic contribution, we must introduce the dependence of the anisotropy parameters k_i in Hamiltonian (11) on the lattice parameters a, b, c and on strain ε,

$$c = (1+\varepsilon)c_0, \quad a = b = \left(1-\frac{1}{2}\varepsilon\right)a_0,$$

$$k_1\left(\frac{c}{a}\right) = k_1\left(\frac{c_0}{a_0}\right) \tag{23}$$

$$+\frac{3}{2}\varepsilon\frac{c_0}{a_0}\frac{\partial}{\partial x}k_1(x)\bigg|_{x=c_0/a_0} + \ldots \equiv k_{cf} + \varepsilon q_{me} + \ldots,$$

and to take the elastic energy into account. Confining ourselves to only one elastic constant C, writing the exchange parameters taking into account the de Gennes rule in the form $I_J = (g_J - 1)I$, and assuming that I is independent of ε, we obtain the Hamiltonian for the R ion in the form

$$H_{\mathrm{R}} = (g_J-1)I(\mathbf{m}\cdot\mathbf{J}) + k_{\mathrm{cf}}\left[J_z^2 - \frac{J(J+1)}{3}\right]$$

$$+\delta H_{\mathrm{me}}(\varepsilon) - g_J\mu_B(\mathbf{H}\cdot\mathbf{J}), \tag{24}$$

$$\delta H_{\mathrm{me}}(\varepsilon) = \varepsilon q_{\mathrm{me}}\left[J_z^2 - \frac{J(J+1)}{3}\right] + \frac{1}{2}C\varepsilon^2.$$

In contrast to relation (11), we are using here, instead of J_z^2, the equivalent operator $\hat{O}_{20}$ as is usually done [3, 4]. In addition to the conventional anisotropy constant k_{cf} associated with the crystal field of the undeformed lattice, the coefficient of this operator now contains the magnetoelastic contribution $\varepsilon q_{\mathrm{me}}$. The magnitude of this contribution depends on strain ε, which is determined from the minimum condition for the mean value of energy (24). The sign of the quantity q_{me} is determined only by the sign of strain ε, and the magnetoelastic correction to energy is always negative,

$$\langle\delta H_{\mathrm{me}}\rangle = -\frac{q_{\mathrm{me}}^2}{C}\left[\langle J_z^2\rangle - \frac{J(J+1)}{3}\right]^2, \tag{25}$$

where $\langle X\rangle \equiv \mathrm{Tr}(X\rho)$.

Formulas (11) and (12), which are associated with computation of the density matrix ρ, mean values, etc., remain valid for Hamiltonian (24) also; this Hamiltonian must only be self-consistent in $\langle J_z^2\rangle$. However, the first formula in Eqs. (11), which describes the angular dependence of the total energy $\mathscr{E}(\theta_{\mathrm{Co}})$, must be modified.

In binary RT compounds, the crystal is deformed uniformly, the mean values are identical for all R sites, and formula (11) for $\mathscr{E}(\theta_{\mathrm{Co}})$ remains valid. However, it was mentioned above that, in the solid solution $\mathrm{R}_x\mathrm{Y}_{1-x}\mathrm{Co}_5$, the cells in which R sites are occupied by yttrium atoms make zero contribution to the anisotropy energy and the energy of exchange with the Co sublattice. Consequently, when calculated per unit cell, these energies must be simply proportional to concentration x. The dependence of the magnetoelastic energy on x turns out to be more complicated. As the concentration deviates from $x = 1$, strain ε changes and becomes nonuniform itself, together with the density of the magnetoelastic energy. For this reason, the magnetoelastic correction to energy per unit cell is a complex nonlinear function of x. However, we will not try to calculate this function, since our goal is not to describe the experiment quantitatively, but only to analyze the possibility of the emergence of magnetization jumps due to

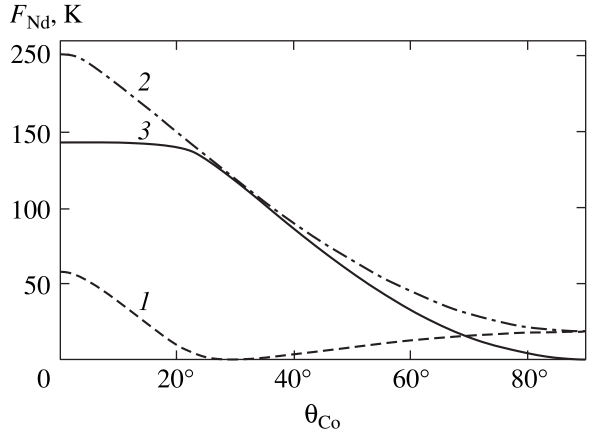


Fig. 5. Angular dependences of the magnetoelastic contribution (curve *1*) and the sum of the contributions from the magnetocrystal anisotropy and the intersublattice exchange interaction (curve *2*) to the total energy F_{Nd} of the Nd sublattice (curve *3*) for the $NdCo_5$ alloy. The values of the parameters are given in the text.

magnetostriction.[2] Since we are interested in the range of low Nd concentrations, we will assume that the deformed regions in the vicinity of each Nd atom do not overlap. In this approximation, the magnetoelastic energy is simply proportional to concentration x and the total energy of the system $R_xY_{1-x}Co_5$ assumes the form

$$\mathscr{E}(x, \theta_{Co}) = xF_R(\theta_{Co}) - F_{1Co}\cos^2(\theta_{Co}) - \mathbf{H}\cdot\mathbf{M}_{Co}. \quad (26)$$

This expression differs from the corresponding expression in Eqs. (11) only in the emergence of the dependence on concentration and in the requirement of self-consistency.

Figures 5 and 6 show the results of numerical calculations based on formula (26) for the system $Nd_xY_{1-x}Co_5$ with the following values of parameters: $I = 350$ K, $J = 9/2$, $g_J = 8/11$, $k_{cf} = 10$ K, $k_{me} \equiv q_{me}^2 J(J + 1)/C = 20$ K, $K_{1Co} = 42$ K, and $M_{Co} = 8\mu_B$/structural unit. These values of the parameters of the Co sublattice are borrowed from [9, 10], while the values of the parameters of the Nd sublattice show that they were not specially fitted.

Figure 5 shows the dependence of the energy F_{Nd} of the neodymium atom on angle θ_{Co}, obtained from self-consistent calculations for the above values of the parameters. It can be seen that the magnetoelastic contribution (curve *1*) to the anisotropy energy is relatively small and amounts to about a quarter of the sum of the contributions from the magnetocrystal anisotropy and exchange (curve *2*). However, the emergence of this contribution transforms the sharp peak at $\theta_{Co} = 0$ discussed in the previous section into a broad and gently sloping peak similar to that obtained in [10]. Taking into account the anisotropy energy of the Co sublattice, we now obtain two minima on each angular dependence of the total energy for all $Nd_xY_{1-x}Co_5$ alloys with $x > 0.1$ (see the insets to Figs. 6). Consequently, the removal of a magnetic field must lead to the emergence of jumps on the magnetization curves for these alloys and to hysteresis phenomena associated with these jumps. The calculated magnetization curves for alloys with $x = 0.1$ and $x = 0.2$ in a field perpendicular to the c axis are shown in Fig. 6a. The experimental magnetization curves [9] for the alloys with $x = 0.1$ and $x = 0.2$ are marked by bullets and dark squares in this figure. Figure 6b shows the calculated magnetization curves for alloys with $x = 0.3$, 0.4, and 1 in a field parallel to the c axis and the corresponding experimental values (dark circles [9], squares [9], and triangles [11]). The insets to Figs. 6a and 6b show the calculated dependences of the energies of various alloys on the direction of their magnetization.

In spite of the absence of special fitting, the agreement between these results obtained for $T = 0$ and experiment can be regarded as satisfactory in our opinion. However, in addition to low-temperature magnetization jumps, compounds $Nd_xY_{1-x}Co_5$ and analogous compounds $Tb_xY_{1-x}Co_5$ exhibit spin-reorientation transitions emerging at higher temperatures. Consequently, to determine the true values of parameters of the R sublattice, it is necessary to match the experimental and calculated positions of both magnetization jumps and the spin-reorientation transitions, which is beyond the framework of this paper.

It remains for us to find out whether or not so large magnetoelastic contributions to the anisotropy of $Nd_xY_{1-x}Co_5$ alloys are possible. At low temperatures, the values of elastic moduli of these compounds were investigated in [13], where the value of $C_{33} \approx 2.5 \times 10^{11}$ J/m^3 was obtained. The magnetostrictive strains $\varepsilon \approx \delta l/l \approx \lambda_2^{\alpha,2}$ in this system of alloys attain values of about 3×10^{-3} [14]. Consequently, the magnetoelastic contribution $-C(\delta l/l)^2/2 \approx -10^6$ J/m^3 can indeed approach a value equal to 0.1 of the total magnetic anisotropy energy, which is equal approximately to 2×10^7 J/m^3 [15]. However, in this case another interesting question arises: how can the magnetoelastic contribution to the first anisotropy constant, i.e., the value of q_{me} in Eq. (17), be so large for $\varepsilon \approx 10^{-3}$?

Such a ratio of the contributions would be quite natural for pure rare-earth (RE) metals, for which the coefficient for Y_{20} in the expansion of the crystal field potential at a site into a series in spherical functions is given by [5]

$$k_1 \approx Q\left[\left(\frac{c}{a}\right)_{id} - \frac{c}{a}\right]. \quad (27)$$

---

[2] If we assume that the strain is the same in all unit cells and its magnitude is proportional to x, we will obviously obtain a function of x^2.

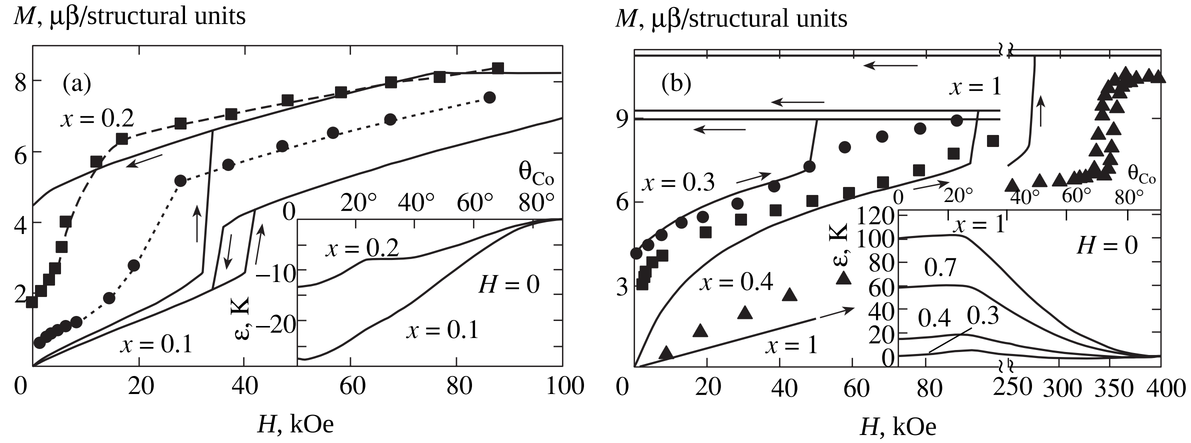


Fig. 6. Magnetization curves in fields parallel to the *a* (a) and *c* (b) axes for the alloys $Nd_xY_{1-x}Co_5$ for (a) $x = 0.1$ (circles correspond to experiment [9]), $x = 0.2$ (squares correspond to experiment [9]) and (b) $x = 0.3$ (circles correspond to experiment [9]), $x = 0.4$ (squares correspond to experiment [9]), $x = 1$ (triangles correspond to experiment [11]). The insets show the angular dependences of the energy of these alloys calculated by formula (26).

Here, $(c/a)_{\mathrm{id}} = \sqrt{8/3} \approx 1.633$ is the ideal value of the ratio c/a, for which the hcp lattice becomes cubic, and Q is the coefficient depending on the charges of neighboring ions and on the distance to these ions. Since $c_0/a_0 \approx 1.58$–1.6 for pure RE metals, the value of $k_{\mathrm{cf}} = Q(1.633 - c_0/a_0)$ turns out to be approximately two orders of magnitude lower than Q, and $q_{\mathrm{me}} = 1.5(c_0/a_0)Q \approx 2.5Q$. Consequently, the relation $\varepsilon q_{\mathrm{me}}/k_{\mathrm{cf}} \approx 10^{-1}$ indeed holds for $\varepsilon \approx 10^{-3}$.

When the crystal field at a site of an RE metal is calculated, the contributions of different signs come to relation (27) from ligands lying in “their own” and “alien” hexagonal planes. It can apparently be expected that the anisotropy constant must be equal to the sum of two large and opposite contributions, which compensate each other almost exactly but vary in different ways upon lattice deformation in other cases also, when small variations of the lattice parameters lead to large relative changes in the anisotropy constant.

It should be emphasized that this assumption is not associated with the desire to explain the emergence of magnetization jumps. This assumption can be verified in independent calculations and experiments. In particular, it is sufficient, knowing the experimental values of elastic moduli and magnetostriction, to compare the magnetoelastic energy and the magnetic anisotropy energy for this purpose.

As regards RCo_5 compounds, the crystal field created by cobalt ions at an R site was also calculated [16]. In this case, in contrast to pure RE metals, the lattice geometry is such that the possibility of compensation of the contributions from different hexagonal planes to the crystal field is ruled out. For this reason, the values of the first anisotropy constant obtained in [16] for different RE metals are two orders of magnitude higher and have signs opposite to those in experiment. This contradiction could in principle be eliminated by introducing certain effects of strong screening, which reverse the sign of the effective charge of the ligand [17]. However, on account of experimental values of magnetostriction constants and elastic constants [13, 14], another reason has to be sought in our opinion. It is more natural to assume the existence of one more contribution to the crystal field of as yet unknown origin, which has a close value and the opposite sign.

4. CONCLUSIONS

(1) The contributions of the intersublattice exchange and magnetoelastic interactions to the magnetic anisotropy of a two-sublattice ferrimagnet lead to an increase in the values of experimentally determined values of higher-order anisotropy constants, which may turn out to be comparable to the first constant.

(2) Higher order constants associated with the intersublattice exchange satisfy the inequality $2K_2 + 3K_3 + \ldots > 0$. The same constants emerging when the magnetoelastic contribution is taken into account satisfy the opposite inequality $2K_2 + 3K_3 + \ldots < 0$.

(3) The emergence of jumps during the rotation of the magnetization vector is possible only if the latter inequality holds. Consequently, it is precisely magnetoelastic interactions (provided that their intensity exceeds a certain threshold value) that can ensure the emergence of magnetization jumps in two-sublattice ferrimagnets.

(4) The possibility of estimating this threshold value in analytic form for a magnet with a paramagnetic sublattice requires special analysis. For a classical magnet, the inclusion of magnetostriction only indicates the addition of the term $K_{\mathrm{me}}[\cos^2\theta_{\mathrm{R}} - 1/3]^2$ to energy (4). Consequently, the quantity K_{me} in this case plays the role of constant $K_{2\mathrm{R}}$ and can be estimated using formula (10).

(5) For an appreciable value of the magnetoelastic correction to the anisotropy constant for small strains, the magnetic anisotropy energy must be the sum of two large and approximately equal contributions of opposite signs, which exhibit different dependences on the lattice parameters. In particular, in the case of RCo_5, a contribution to the crystal field at an R site must exist, which compensates the potential of cobalt ions.

(6) If the emergence of magnetization jumps is indeed associated with magnetoelastic interactions, the absence of a hysteresis loop on experimental magnetization curves can be explained quite easily. In a real crystal, strong internal compressive and tensile stresses are always present; these stresses stabilize the seeds of the phase with magnetization lying in the vicinity of the *c* axis in some regions and with magnetization lying near the basal plane in other regions. For this reason, a transition may occur due to shifts in phase boundaries, as was noted in [18] without specifying the nature of the phases.

(7) The calculations of energy (20) of $Nd_xY_{1-x}Co_5$ alloys with the above values of parameters at nonzero temperatures give the region of spontaneous spin reorientation, which are in satisfactory agreement with experiment. However, in contrast to the traditional theory [8], these calculations predict the transformation of spin-reorientation transitions in temperature into first-order transitions. The parameter determining whether the spin reorientation will occur continuously or will experience a jump is the magnetoelastic interaction energy. If this energy is small in the transition region, only one minimum exists and a second-order transition takes place; if, however, this energy is large, there are two minima and a first-order transition takes place. This could be confirmed or rejected experimentally by analyzing Barkhausen magnetic and/or acoustic noise upon a change in the temperature in the transition region.

ACKNOWLEDGMENTS

The author is grateful to Yu.P. Irkhin for numerous valuable discussions and to A.S. Ermolenko and N.V. Mushnikov for fruitful discussions of experimental results.

This study was supported by the Russian Foundation for Basic Research (project no. 02-02-16440).

REFERENCES

1. V. Yu. Irkhin, J. Phys.: Condens. Matter **14**, 6865 (2002).
2. L. G. Onoprienko, Fiz. Met. Metalloved. **17**, 350 (1964).
3. T. Asti and F. Bolzoni, J. Magn. Magn. Mater. **20**, 29 (1980).
4. M. T. Hutchings, Solid State Phys. **16**, 227 (1964).
5. T. Kasuya, *Magnetism* (Academic, New York, 1966), Vol. 2B, p. 212.
6. E. R. Callen and H. B. Callen, J. Phys. Chem. Solids **16**, 310 (1960).
7. A. S. Ermolenko, E. V. Rozenfeld, Yu. P. Irkhin, *et al.*, Zh. Éksp. Teor. Fiz. **69**, 1743 (1975) [Sov. Phys. JETP **42**, 885 (1975)].
8. Yu. P. Irkhin and E. V. Rozenfeld, Fiz. Tverd. Tela (Leningrad) **16**, 485 (1974) [Sov. Phys. Solid State **16**, 310 (1974)].
9. A. S. Ermolenko, Fiz. Met. Metalloved. **50**, 952 (1980).
10. A. S. Ermolenko, Fiz. Met. Metalloved. **53**, 706 (1982).
11. M. I. Bartashevich, T. Goto, M. Yamaguchi, *et al.*, Solid State Commun. **87**, 1093 (1993).
12. E. R. Callen and H. B. Callen, Phys. Rev., Sect. A **139**, 455 (1965).
13. A. V. Deryagin, G. M. Kvashnin, and A. M. Kapitonov, Fiz. Tverd. Tela (Leningrad) **26**, 3106 (1984) [Sov. Phys. Solid State **26**, 1871 (1984)].
14. A. V. Andreev, A. V. Deryagin, and S. M. Zadvorkin, Zh. Éksp. Teor. Fiz. **85**, 974 (1983) [Sov. Phys. JETP **58**, 566 (1983)].
15. E. Tatsumoto, T. Ocamoto, H. Fujii, and G. Inoue, J. Phys. Suppl. C1 **32**, 550 (1971).
16. Yu. P. Irkhin, E. I. Zabolotskiĭ, E. V. Rozenfeld, and V. P. Karpenko, Fiz. Tverd. Tela (Leningrad) **15**, 2963 (1973) [Sov. Phys. Solid State **15**, 1976 (1973)].
17. V. Yu. Irkhin and Yu. P. Irkhin, Phys. Rev. B **57**, 2697 (1998).
18. Yu. P. Irkhin, Fiz. Tverd. Tela (St. Petersburg) **45**, 676 (2003) [Phys. Solid State **45**, 709 (2003)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 969–977.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 1080–1089.
Original Russian Text Copyright © 2003 by Aplesnin.

SOLIDS
Structure

Nonadiabatic Interaction of Acoustic Phonons with Spins S = 1/2 in the Two-Dimensional Heisenberg Model

S. S. Aplesnin

Kirenskiĭ Institute of Physics, Siberian Division, Russian Academy of Sciences,
Akademgorodok, Krasnoyarsk, 660036 Russia
e-mail: apl@iph.krasn.ru

Received February 27, 2003

Abstract—The ground state of a two-dimensional antiferromagnet having spins S = 1/2 and interacting with acoustic phonons is investigated in the nonadiabatic approximation using the quantum-mechanical Monte Carlo method. The critical parameters of the spin–phonon coupling, corresponding to the formation of bound spin–phonon excitations, crystal symmetry lowering, and the emergence of a gap in the spin excitation spectrum, as well as the antiferromagnet–quantum spin liquid transition, are determined. The orthorhombicity parameter, the sublattice magnetization, the violation of the spherical symmetry of spin–spin correlation functions, and the magnetic moment in Gd_2CuO_4 and Eu_2CuO_4 are calculated. *© 2003 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Intense studies of electronic, elastic, and magnetic properties of high-temperature superconductors and manganites with a colossal magnetoresistance have lead to the conclusion that the electronic structure is closely related to magnetic and lattice fluctuations. For weakly doped superconducting cuprates, one of the hypotheses is associated with the formation of a quasigap due to the generation of coupled spin–phonon excitations. This hypothesis is confirmed by the results of optical measurements [1]; for example, Raman spectra are explained on the basis of coupled excitations in a system consisting of two magnons and a phonon. In the low-temperature range, a number of observed structural distortions are due to lattice modulation, and superstructural reflections exhibit tetragonal symmetry as well as symmetry lower than the orthorhombic symmetry. For $La_{1.6-x}Nd_{0.4}Sr_xCuO_4$, these transitions exist in the normal phase at T < 80 K [2]. In $La_{2-x}Sr_xCuO_4$ (x < 0.05), the isotopic effect is observed with an intensity comparable to that for traditional superconductors [3]. Another feature of weakly doped semiconducting cuprates is associated with their thermal conductivity, which cannot be described using the theory of a Fermi liquid and presumes the existence of certain delocalized quasiparticles [4]. These experimental facts indicate the existence of two characteristic energy scales: the electron–phonon and the spin–phonon interactions.

Some peculiarities in magnetic properties, which are also observed in allied compounds with a tetragonal T' structure and with CuO planes in R_2CuO_4 (R = Eu, Gd, Nd), can be due to the interaction between lattice and spin fluctuations. Such compounds are characterized by low values of the magnetic moment of copper ions ($\sigma \approx 0.4$) and a relatively high Néel temperature ($T_N \approx$ 230–280 K). The following has been observed: strong anharmonicity in local displacements in the CuO plane in the temperature range 145 K < T < 175 K in the absence of structural transitions up to 393 K and a minimum in the temperature dependence of the square of average displacements of copper ions along the [100] direction at T = 175 K [5] due to antiferromagnetic spin fluctuations. The antiferromagnet Gd_2CuO_4 with a tetragonal symmetry exhibits the electron spin resonance at ω_0 = 18.2 cm^{-1} [6], which can be explained by the orthorhombic distortion of lattice planes with a stochastic arrangement of the orthorhombicity vectors along the c axis. At T = 20 K, the resonance disappears and the susceptibility increases strongly [7], which can be attributed to the coherent orthorhombicity state (although the elastic scattering of neutrons and X-ray studies do not confirm this effect). These effects are probably associated with the formation of bound spin–phonon quasiparticles, i.e., lattice and spin fluctuations coupled dynamically with each other.

Antiferromagnets with the spin–phonon interaction were considered in the adiabatic approximation, in which the interaction between spin and acoustic phonons can be reduced to the four-spin exchange interaction and to the effective interaction between the spins of next-to-nearest neighbors. For certain parameters of this model, a spin nematic state with violation of the spherical symmetry of the spin–spin correlation functions is formed [8] and the long-range magnetic order disappears [9]. The interaction between the spin and the elastic subsystems leads to nonlinear interactions not only between spins, but also between phonons. For this reason, a correct solution should be carried out taking into account the nonadiabatic interaction between spins and phonons; this can be done using the quantum-mechanical Monte Carlo method based on the continuous-time algorithm.

2. COMPUTATIONAL MODEL AND METHOD

For quasi-two-dimensional magnets, the interplanar exchange is several orders of magnitude weaker than the intraplanar exchange; consequently, we can confine our analysis to the interaction between the spins of the nearest neighbors and with acoustic modes of vibrations in the plane of the lattice. In the harmonic approximation, the Hamiltonian for a coupled spin–phonon system has the form

$$\begin{aligned} H = \sum_{i,j} & [J + \alpha(u_{i,j} - u_{i+1,j})] \\ & \times \left[S^z_{i,j}S^z_{i+1,j} + \frac{1}{2}(S^+_{i,j}S^-_{i+1,j} + S^-_{i,j}S^+_{i+1,j})\right] \\ & + [J + \alpha(u_{i,j} - u_{i,j+1})] \\ & \times \left[S^z_{i,j}S^z_{i,j+1} + \frac{1}{2}(S^+_{i,j}S^-_{i,j+1} + S^-_{i,j}S^+_{i,j+1})\right] \\ & + \frac{1}{2}M\dot{u}^2_{i,j} + \frac{1}{2}K(u_{i,j} - u_{i+1,j})^2 + \frac{1}{2}K(u_{i,j} - u_{i,j+1})^2, \end{aligned} \tag{1}$$

where $S^{z,\pm}$ are the components of the spin operator $S = 1/2$ at a lattice site, $u_{i,j}$ is the displacement of an ion over the translation vectors of the lattice, M is the ion mass, and K is the elastic rigidity constant of the lattice ($J > 0$). Using the canonical transformation

$$\hat{u}_{\mathbf{r}} = \frac{1}{\sqrt{N}}\sum_{\mathbf{q}}\sqrt{\frac{\hbar}{2M\Omega(\mathbf{q})}}(b_{\mathbf{q}} + b^\dagger_{-\mathbf{q}})e^{i\mathbf{q}\cdot\mathbf{r}}, \tag{2}$$

we pass from variables $u_{i,j}$ to the creation ($b^\dagger$) and annihilation (b) operators of phonon with momenta $q_\beta = 2\pi n/L$, $n = 1, 2, \ldots, L$, where $\beta = x, y$ and the lattice constant $a = 1$. The transformed Hamiltonian has the form

$$\begin{aligned} H = \sum_{i,j} & J_{i,j}\mathbf{S}_i \cdot \mathbf{S}_j + \sum_{q_x, q_y}\sum_{n,m}\alpha\sqrt{\frac{\hbar}{2M\Omega(\mathbf{q})}}e^{i\mathbf{q}\cdot\mathbf{r}} \\ & \times (b_{\mathbf{q}} + b^+_{-\mathbf{q}})[(1 - \cos q_x - i\sin q_x)\mathbf{S}_{n,m}\cdot\mathbf{S}_{n+1,m} \\ & + (1 - \cos q_y - i\sin q_y)\mathbf{S}_{n,m}\cdot\mathbf{S}_{n,m+1}] \\ & + \sum_{\mathbf{q}}\hbar\Omega(\mathbf{q})b^\dagger_{\mathbf{q}}b_{\mathbf{q}}, \end{aligned} \tag{3}$$

where $\Omega(\mathbf{q}) = \omega_0\sqrt{2 - \cos q_x - \cos q_y}$ and $\omega_0 = \sqrt{2K/M}$.

In computations, the spin–phonon interaction constant α and the excitation energy ω normalized to exchange are used. As the computational method, we choose the quantum-mechanical Monte Carlo method combining the algorithms of world lines and continuous time [10] on a plane of dimensions $N = 32 \times 32$ with periodic boundary conditions at temperature $\beta = J/T = 50$. In accordance with this method, the Hamiltonian is divided into three parts: the diagonal part,

$$H_0 \propto JS^z_iS^z_j + \hbar\Omega(\mathbf{q})n_{\mathbf{q}},$$

where $n_{\mathbf{q}}$ is the occupation number of phonons with the same momentum; and two nondiagonal parts,

$$V_J \propto \frac{J}{2}(S^+_iS^-_j + S^-_iS^+_j),$$

$$V_\alpha \propto \alpha\sqrt{\frac{\hbar}{M\Omega(\mathbf{q})}}(b_{\mathbf{q}} + b^\dagger_{-\mathbf{q}})\hat{S}_i\hat{S}_j.$$

Applying the Trotter formula [11], we can disregard the commutation of operators V_j and V_α to within $\tau_0^2\alpha J/(2\sqrt{\omega_0})$. This leads to a systematic error whose maximal value is 15% for $\alpha = 4$, $\omega_0 = 8$, and $\tau_0 = 0.5$. Following [10] we express operators $\exp[-\tau_0(H_0/2 + V_J)]$ and $\exp[-\tau_0(H_0/2 + V_\alpha)]$ on the imaginary time segment τ_0 in terms of the evolution operator σ_{ev} in the interaction representation $\exp(-\tau_0 H) = \exp(-\tau_0 H_0)\sigma_{ev}$, where

$$\begin{aligned} \sigma_{ev} = 1 - & \int_0^{\tau_0} d\tau V_{J,\alpha}(\tau) + \ldots + (-1)^m \\ & \times \int_0^{\tau_0} d\tau_m \ldots \int_0^{\tau_2} d\tau_1 V_{J,\alpha}(\tau_m)\ldots V_{J,\alpha}(\tau_1) + \ldots, \end{aligned} \tag{4}$$

and

$$V_{J,\alpha}(\tau) = e^{\tau H_0}V_{J,\alpha}e^{-\tau H_0}, \quad V_{J,\alpha}|\beta\rangle = -q_{\gamma\beta}(J,\alpha)|\gamma\rangle.$$

Summation and integration of two operators V_J and V_α in Eq. (4) are carried out using a stochastic procedure of sampling various kink–antikink configurations in accordance with their weights. The probability of the formation of a kink–antikink pair is given by

$$W = |q_{\gamma\beta}(J,\alpha)|^2\exp[(\tau_2 - \tau_1)E_{\gamma\beta}], \quad \tau_2 - \tau_1 < \tau_0.$$

A subprocess of kink shift along the time axis with probability $W = \exp(\Delta\tau E_{\gamma\beta})$ is possible. The use of global spin flips at a site and a change in the occupation number of phonons with momentum q lead to a finite transition probability $W \sim q_{\gamma\beta}$ on interval τ_0. As a result, the total projection of the spin changes and discontinuities are observed on world lines with even numbers. Since computations lead to only an even number of nondiagonal changes of trajectories $q^{2n}_{\gamma\beta}(J)$ and $q^{2n}_{\gamma\beta}(\alpha)$, we can avoid obtaining the minus sign due to an increase in the systematic error. As an eigenfunction of Hamiltonian H_0, we choose the S^z representation of ↑ and ↓ spins; the occupation numbers of phonons with momentum $\mathbf{q}$ are $n_\mathbf{q} = 0, 1, 2, \ldots$ (the maximum number is not limited).

The spectral density of magnetic and spin–phonon excitations can be determined from the corresponding time correlation functions calculated in imaginary time for $\tau > 0$. We define the spin correlator in the form

$$\begin{aligned}&\langle S^-(\tau)S^+(0)\rangle\\ &= \sum_\nu |\langle\nu|S^+|\mathrm{vac}\rangle|^2 \exp[-(E_\nu - E_0)\tau],\end{aligned} \tag{5}$$

where $|\nu\rangle$ is the complete set of eigenstates of Hamiltonian H_0, $H_0|\nu\rangle = E_\nu|\nu\rangle$, $H_0|\mathrm{vac}\rangle = E_0|\mathrm{vac}\rangle$. For the vacuum state, we choose the Néel arrangement of spins with energy $E_0/NJ = 1/4$. Let us redefine the spin correlator (5) as

$$\begin{aligned}\langle S^-(\tau)S^+(0)\rangle &= \int_0^\infty d\omega \rho_s(\omega) e^{-\omega\tau},\\ \rho_s(\omega) &= \sum_\nu \delta(\omega - \Omega_\nu)|\langle\nu|S^+|\mathrm{vac}\rangle|^2,\\ \Omega_\nu &= E_\nu - E_0,\end{aligned} \tag{6}$$

where $\rho_s(\omega)$ defines the spectral density of magnetic excitations. We treat spin–phonon excitations as coupled excitations of spins, appearing as a result of action of operators $\hat{S}^z$ and $\hat{S}^+$, and phonons induced by the creation operators $b^\dagger$ on the wave function of vacuum $n_\mathbf{q} = 0$. We represent the time correlators in the form

$$\begin{aligned}\langle b(\tau)S^z(\tau)b^\dagger(0)S^z(0)\rangle &= \sum_\nu |\langle\nu|b^\dagger S^z|\mathrm{vac}\rangle|^2\\ &\times \exp[-(E_\nu - E_0)\tau], \quad \tau \ge 0,\\ \langle b(\tau)S^-(\tau)b^\dagger(0)S^+(0)\rangle &= \sum_\gamma |\langle\nu|b^\dagger S^+|\mathrm{vac}\rangle|^2\\ &\times \exp[-(E_\gamma - E_0)\tau], \quad \tau \ge 0.\end{aligned} \tag{7}$$

Analogously to (6), we define the spectral density of coupled spin–phonon excitations in the form

$$\rho_{sp}(\omega) = \rho_{bS^z}(\omega) + \rho_{bS^+}(\omega).$$

In fact, the Monte Carlo method is used for calculating the time correlator on a finite interval $0 < \tau < \tau_0$. In order to reproduce the spectral density in a wide range of energies, we must solve the integral equation (6). For this purpose, we use the stochastic procedure optimizing the deviation [12]

$$D = \int_0^{\tau_0} |G(\tau) - G_t(\tau)| G^{-1}(\tau) d\tau \tag{8}$$

of the computed correlator $G(\tau)$ from the true correlator $G_t(\tau)$ with the spectral density $\rho_t(\omega)$.

In order to calculate the nondiagonal operators, we use a symmetrized representation of the wave function in the imaginary time interval τ_0. For example, we seek the eigenvalue of operators $(b^\dagger_\mathbf{q}, b_\mathbf{q})$ for $\tau \longrightarrow 0$ on the basis of the functions

$$\begin{aligned}\nu(\tau_i) &= c_1|n_{\mathbf{q}1}, n_{\mathbf{q}2}, n_{\mathbf{q}3}, \ldots\rangle\\ &+ c_2|n_{\mathbf{q}1} + 1, n_{\mathbf{q}2}, n_{\mathbf{q}3}, \ldots\rangle\\ &+ c_3|n_{\mathbf{q}1}, n_{\mathbf{q}2} + 1, n_{\mathbf{q}3}, \ldots\rangle + \ldots\end{aligned}$$

The displacement of an ion at a site is defined as

$$\begin{aligned}\langle u(\mathbf{r})\rangle &= \sqrt{\frac{\hbar}{2MN}}\\ &\times \sum_\nu \langle\nu_j|(b_\mathbf{q} + b^\dagger_{-\mathbf{q}})e^{i\mathbf{q}\cdot\mathbf{r}}\Omega^{-1/2}(\mathbf{q})|\nu_j\rangle.\end{aligned} \tag{9}$$

The mean square displacement of the ion is defined as

$$\langle u^2\rangle = \frac{\hbar}{2MN}\sum_\mathbf{q}\frac{2n_\mathbf{q} + 1}{\Omega(\mathbf{q})}.$$

In the ground state, the number of phonons for a harmonic oscillator with $\alpha \longrightarrow 0$ is equal to zero. Therefore, it is important to calculate the change in zero-point vibrations as a result of action of the magnetic system on the elastic one; i.e.,

$$\langle U_n^2\rangle = \langle u^2(\alpha)\rangle - \langle u^2(\alpha = 0)\rangle.$$

In the subsequent analysis, we will use the quantities

$$\langle U\rangle = \frac{1}{N}\sum_r \frac{\langle u(r)\rangle}{\sqrt{\hbar/2MN\omega_0}}, \quad \langle U^2\rangle = \sum_r \frac{\langle U_n^2(r)\rangle}{\hbar/2NM\omega_0}.$$

Correlated vibrations of ions and their momentum dependence can be determined from the correlation

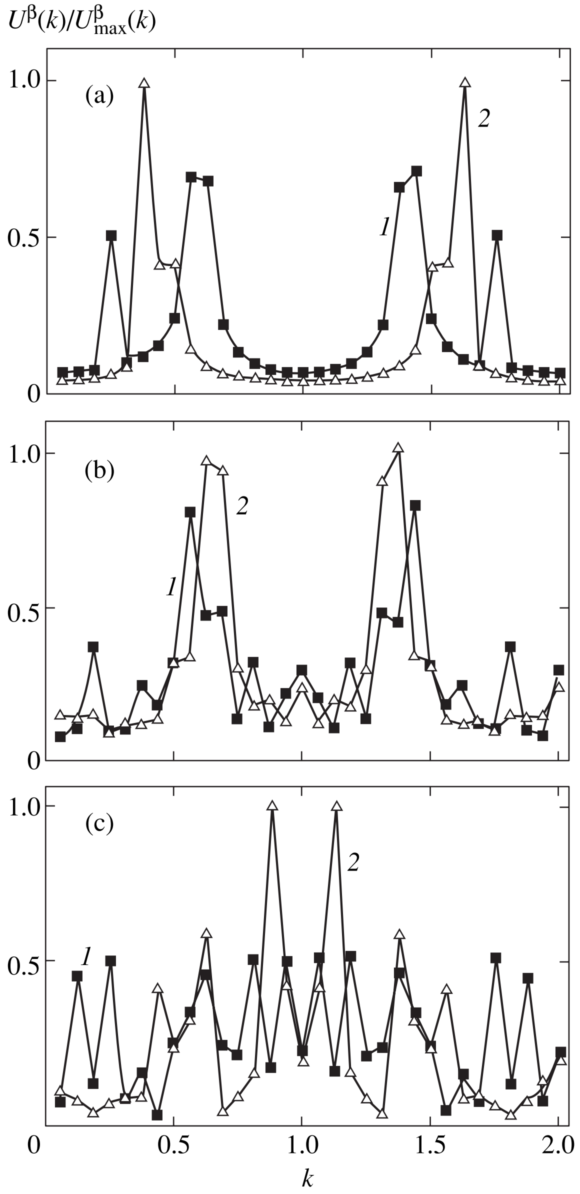


Fig. 1. The lattice structural factor $U^\beta(k)/U^\beta_{\max}(k)$ normalized to the maximum value and calculated in two directions β = [10] (*1*) and [01] (*2*) for $\omega_0/J = 6$, $\alpha/\alpha_{c3} = 0.3$ (a), 0.7 (b), and 1 (c).

phonon density function $\langle n(\mathbf{q})n(\mathbf{q}+\mathbf{p})\rangle$. The wave vectors of incommensurability of lattice and magnetic fluctuations were determined from the ion displacement structural factor

$$U(\mathbf{q}) = \frac{1}{N}\sum_{\mathbf{q}} \langle u_0 u_{\mathbf{r}}\rangle e^{i\mathbf{q}\cdot\mathbf{r}}$$

in two directions, [10] and [01], and the magnetic structural factor

$$S^z(\mathbf{k}) = \frac{1}{N}\sum_{\mathbf{k}} \langle S_0^z S_{\mathbf{r}}^z\rangle e^{i\mathbf{k}\cdot\mathbf{r}}.$$

In the computational procedure, the first 20000 Monte Carlo steps per spin were omitted and the averaging was carried out over 8000 Monte Carlo steps per spin. This value is much larger than the time of attainment of thermodynamic equilibrium calculated from the sublattice magnetization,

$$\langle\sigma(0)\sigma(t)\rangle - \langle\sigma(0)\rangle\langle\sigma(\tau_{\max})\rangle = A\exp(-t/t_0),$$

where t is the number of Monte Carlo steps; t_0 = 3000 and 7000 Monte Carlo steps per spin for $\alpha/\alpha_c = 0.3$ and 0.75, respectively; and α_c is the critical parameter of the spin–phonon coupling, for which the long-range magnetic order disappears. The mean square error amounts approximately to 3% for the sublattice magnetization, 1% for the energy, 2% for the spin–spin correlation functions, and 4% for the average phonon occupation number.

3. DISCUSSION

The processes of inelastic scattering and formation of magnon and phonon bound states are determined by the density of states of the initial quasiparticles. In the two-dimensional Heisenberg model, the density of magnon states diverges logarithmically at the middle of the band and the interaction between quasiparticles is symmetric for points Γ and X of the band. In the case when the dispersion curves for magnons and phonons intersect, which is observed for $v_{ph} < v_m$ at $\omega_0/J < 2$ (v_{ph} and v_m are the velocities of phonons and magnons, respectively), additional singularities are formed in the density of states of these quasiparticles. The calculations were made for ω_0/J = 1, 2, 4, 6, 8, and 10; the figures illustrate the typical cases when $\omega_0/J = 1$ and $\omega_0/J = 6$. Under the action of the magnetic system, the structural factor of lattice fluctuations shown in Fig. 1 becomes spatially anisotropic. Ladder-type fluctuations containing two nearest chains in the [01] direction and quasi-one-dimensional chain fluctuations in the [10] direction, which are separated by distance $r \approx 7$–10, are formed in the magnetic subsystem. The energy per bond in an antiferromagnet with a square lattice is 1.3 times smaller than the energy in an antiferromagnetic chain. Consequently, lattice fluctuations facilitate local extension of the lattice along one of the symmetry directions of the initial square lattice. Ladder-type fluctuations accompanied by dynamic local lattice dimerization also lower the magnetic energy; the approximated dependence of this energy has the form

$$|E_m(\alpha) - E_m(0)| \approx A(\alpha/\alpha_{c3})^{1.80(6)},$$

$$A = \begin{cases} 0.11(1), & v_m > v_{ph}, \\ 0.18(2), & v_m < v_{ph}. \end{cases}$$

The gain in the magnetic energy is almost an order of magnitude higher than the loss in the elastic energy.

The average displacement

$$U^{\beta}_{av} = \frac{1}{N}\sum_{i,j} u^{\beta}_{i,j}, \quad \beta = x([10]), y([01])$$

is depicted in Fig. 2. Anisotropy of lattice fluctuations leads to anisotropy of displacement $U^x_{av} - U^y_{av}$ (Fig. 2c) and lowers the crystal symmetry from tetragonal to orthorhombic. With increasing interaction between the magnetic and elastic subsystems, zero-point vibrations at a certain wave vector Q are enhanced, as well as their correlation $\langle U_i^2 U_j^2\rangle \propto \langle N(0)N(Q)\rangle$, depicted in Fig. 3. The maximal value of the correlator is attained at the wave vector $Q_{\max} = (0.75–0.9)\pi$, $\alpha_{c2} < \alpha < \alpha_{c3}$ and reflects a coherent vibration of ions with localized spin excitations in the [10] direction. For $v_m > v_{ph}$, in the interval $\alpha_{c2} < \alpha < \alpha_{c3}$ of the parameters, the local orthorhombicity parameter shown in Fig. 2c decreases sharply, its value being within the computational error. The change in the symmetry of structural distortions is in qualitative agreement with the replacement of the condensed mode (π, 0) for $\delta < 0.5$ ($\delta = J_{i,\,i+1} - J_{i,\,i-1}$) by the optical mode ($\pi$, π) for $\delta > 0.5$, calculated on a square lattice by the method of exact diagonalization in the adiabatic approximation [13].

A qualitatively different behavior of elastic and magnetic properties is observed in the case when $v_m < v_{ph}$. The lattice volume and the orthorhombicity parameter increase monotonically for $\alpha > \alpha_{c2}$ and the change in zero-point vibrations is an order of magnitude smaller as compared to the case when $v_m > v_{ph}$ (Fig. 3a). Anisotropy of correlated vibrations also increases and the results of calculations can be correctly described by the power dependence

$$\langle U^2_{\beta}\rangle - \langle U^2_{\gamma}\rangle \approx 0.24(3)[(\alpha - \alpha_{c2})/\alpha_{c3}]^{0.41(3)}.$$

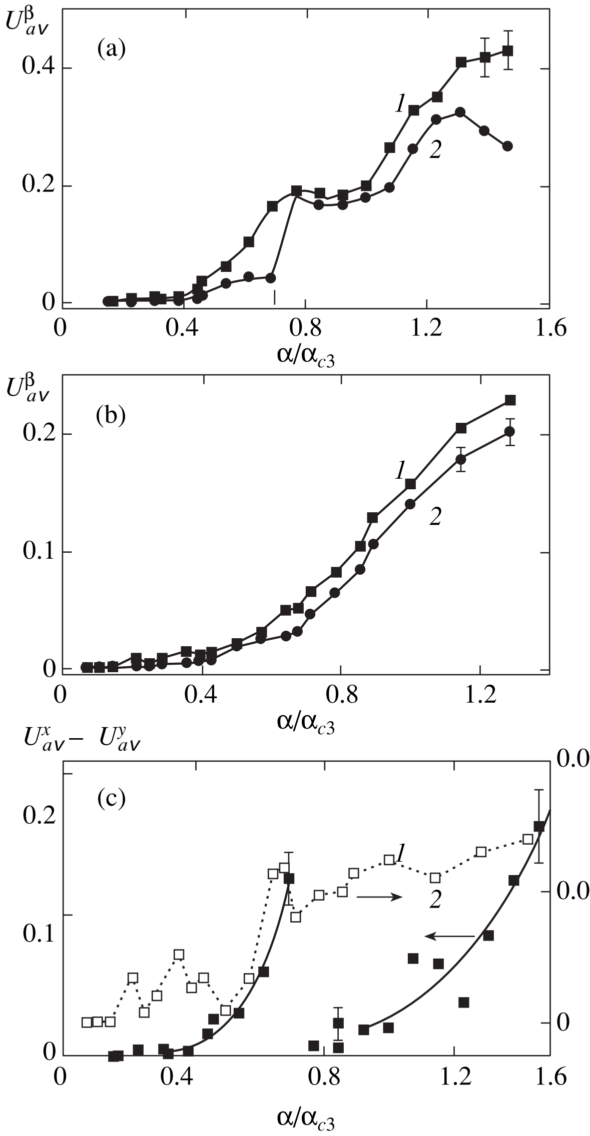


Fig. 2. Dependence of lattice-averaged displacements U^{β} of ions, normalized to $\sqrt{\hbar/M\omega_0}$, for ω_0/J = 1 (a) and 6 (b) in directions β = [10] (*1*) and [01] (*2*) on the normalized spin–phonon interaction constant and the dependence of the orthorhombicity parameter $U^x_{av} - U^y_{av}$ for ω_0/J = 1 (*1*), 6 (*2*) on α/α_{c3} (c).

The changes in the lattice parameters in the region of the critical values $\alpha_{c2,\,3}$, in which the typical values of the upper boundary of the region of acoustic vibrations in quasi-two-dimensional antiferromagnets R_2CuO_4 (R = La, Gd, Eu, Nd) are $M \approx 4 \times 10^{-22} g$, and $\omega_0 = J$, are equal to

$$U_{c2} = 0.005(1)\ \text{Å}, \quad U_{c3} = 0.04(02)\ \text{Å}$$

for $\omega_0 \approx 4 \times 10^{12}$ Hz [14] and

$$U_{c2} = 0.002(1)\ \text{Å}, \quad U_{c3} = 0.007(2)\ \text{Å}$$

for $\omega_0 \approx 10^{14}$ Hz. The lattice-averaged change in the exchange interaction in the region of the antiferromagnet–quantum spin liquid phase transition constitutes approximately 1% ($dJ/J \approx 0.01$), which is an order of magnitude smaller than local exchange fluctuations.

The linear decrease in the magnetic moment at a site upon an increase in the spin–phonon coupling constant can be approximated by the dependence

$$\frac{\sigma}{\sigma(0)} = \begin{cases} 1.14 - 1.3\alpha/\alpha_{c3}, & v_m > v_{ph}, \\ 1.12 - 0.96\alpha/\alpha_{c3}, & v_m > v_{ph}. \end{cases}$$

In the range of parameters $0.15 < \alpha/\alpha_{c3} < 0.7$. For α =

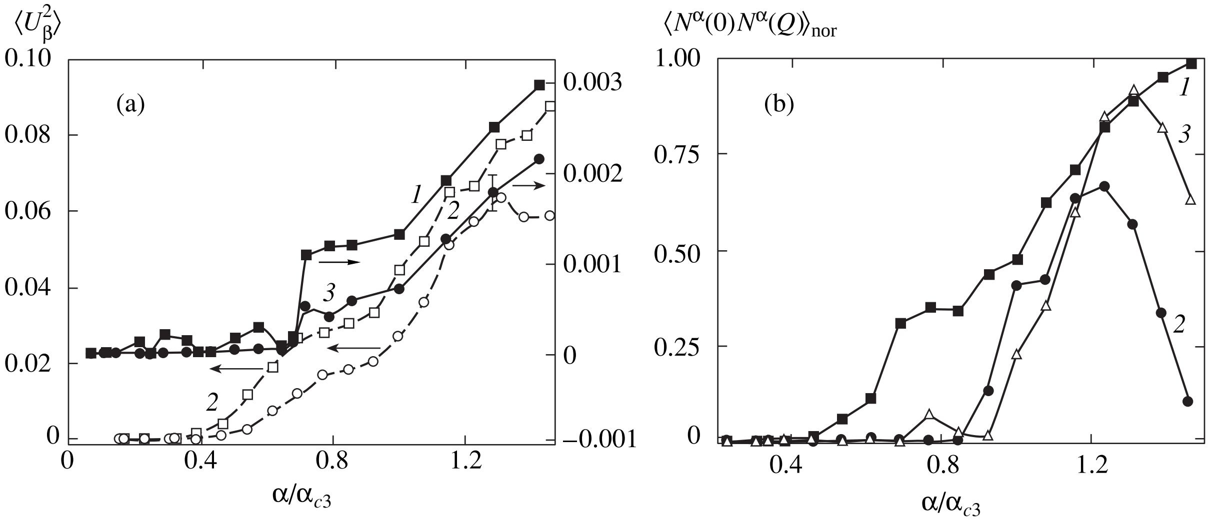


Fig. 3. Dependences (a) of the mean square displacement $\langle U_\beta^2 \rangle$ of ions normalized to $\hbar/M\omega_0$ for β = [10] (*1*, *2*) and [01] (*3*, *4*), $\omega_0/J = 1$ (*2*, *4*) and 6 (*1*, *3*) on the normalized spin–phonon interaction constant and (b) of the maximum value of the phonon density correlator on the wave vector Q for $\omega_0/J = 1$, β = [10] (*1*), [01] (*2*), and [11] (*3*) on α/α_{c3}.

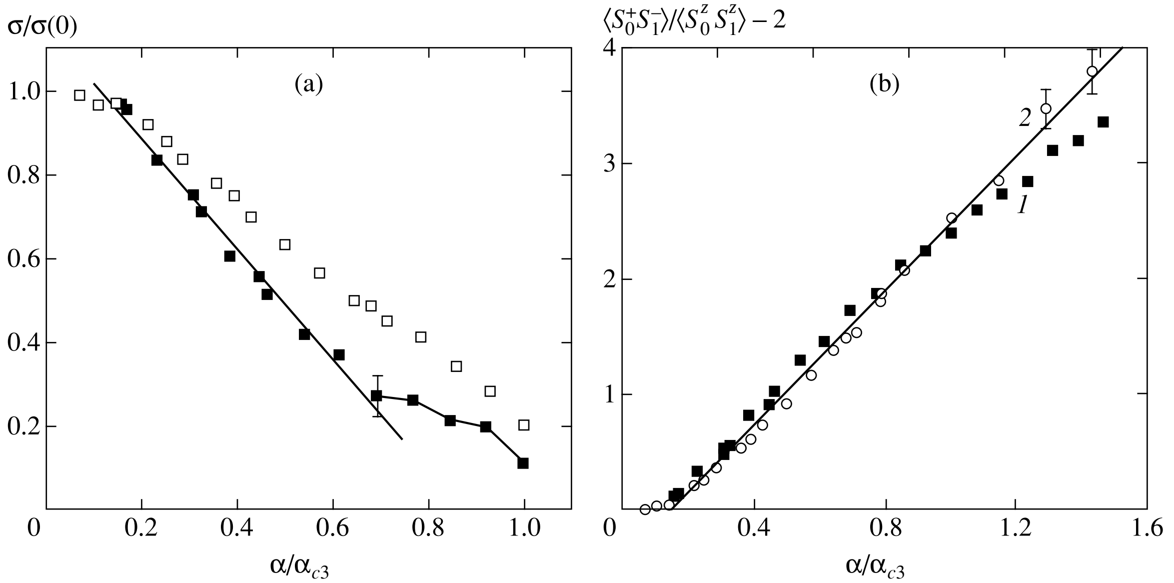


Fig. 4. Normalized magnetic moment $\sigma/\sigma(0)$ at a site (a) and the ratio of the correlation functions between the transverse components of spins and longitudinal components for $r = 1$ (b) for $\omega_0/J = 1$ (*1*) and 6 (*2*) as functions of the spin–phonon interaction parameter.

α_{c3}, the magnetic moment abruptly vanishes. The typical dependences are shown in Fig. 4a. In the region of the critical parameters of the spin–phonon coupling, the spin–spin correlation functions and the correlation radius are spatially anisotropic,

$$\left|1-\frac{\sum_r \langle S^z(i,j)S^z(i,j+r)\rangle}{\sum_r \langle S^z(i,j)S^z(i+r,j)\rangle}\right| \approx 0.02\text{–}0.04,$$

and anisotropy of the spin correlation functions between directions [11] and [10] is on the order of 0.1. For $\alpha > \alpha_{c1}$, the spherical symmetry of spin–spin correlation functions depicted in Fig. 4b is violated. This fact serves as a criterion for determining the value of the spin–phonon interaction parameter α_{c1} and is in qualitative agreement with the results obtained by Andreev and Grishchuk [8], who obtained a spin-nematic state in the Heisenberg model with competing antiferromagnetic interactions and the four-spin exchange. In the vicinity of the wave vector $Q = (\pi, \pi)$ corresponding to

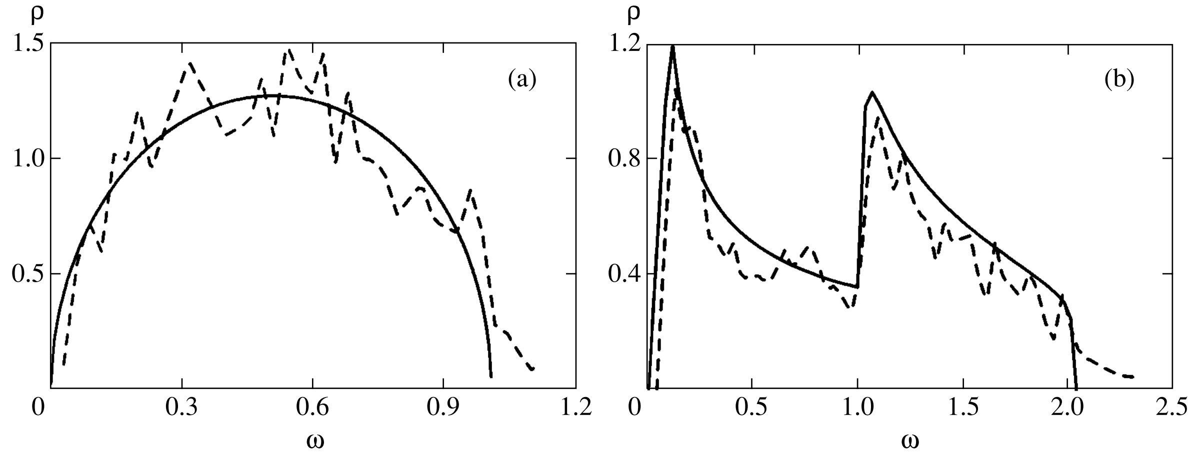


Fig. 5. Model densities of states defined analytically, $\rho(\omega) = (8/\pi)\sqrt{\omega(1-\omega)}$ (a) and numerically (b) (solid curves). The reconstructed density of states are depicted by the dashed curves.

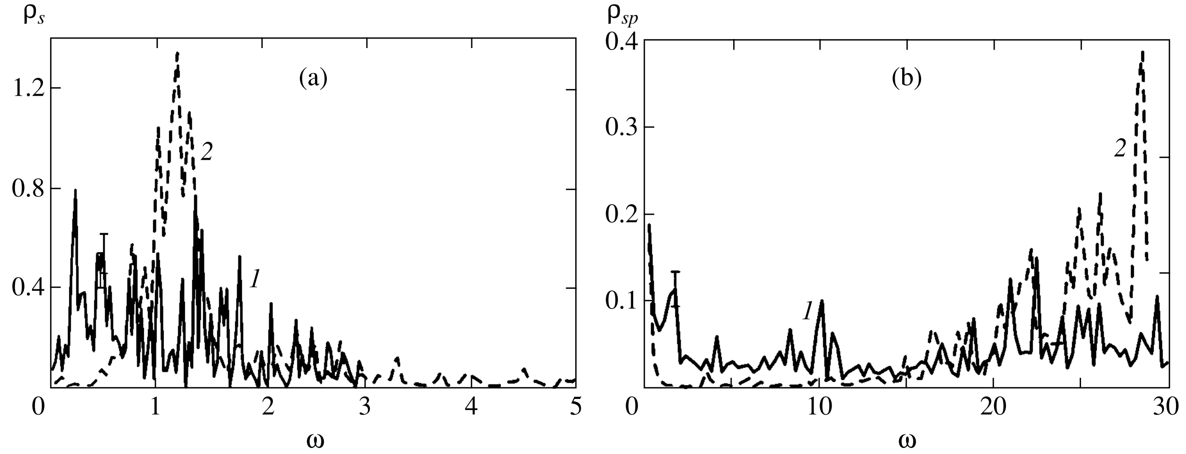


Fig. 6. Density of states for (a) spin and (b) coupled spin–phonon excitations for $\omega_0/J = 6$, $\alpha/\alpha_{c3} = 0.8$ (*1*) and 1.14 (*2*).

the antiferromagnetic structural factor, satellites with the incommensurability vector of the spin density are observed in the interval $q_{in} = (0.7–0.95)\pi$. The intensity of the satellites varies in the limits

$$\frac{S^z(q_{in})}{S^z(\pi,\pi)} \approx \begin{cases} 0.05, & \alpha = \alpha_{c1}, \\ 0.15, & \alpha = \alpha_{c2}, \\ 0.3, & \alpha = \alpha_{c3}. \end{cases}$$

The procedure of reconstructing the spectral density of states [12] for given models can be successfully used for determining the band boundaries and the positions of the peaks of the function $\rho(\omega)$ on the energy scale to within 5%. The intensity has a saw-tooth shape and fluctuates in the limits of 10–20%. Figure 5 shows the reconstructed and model densities of states defined analytically ($\rho(\omega) = (8/\pi)\sqrt{\omega(1-\omega)}$) and numerically. The time correlator

$$G(\tau) = \int_0^{\omega_{\max}} e^{-\omega\tau}\rho(\omega)d\omega$$

was calculated over 100 points τ_i, $i = 1, 2, \ldots, 100$. The typical densities of spin excitations and coupled spin–phonon excitations are shown in Fig. 6. For $\alpha > \alpha_{c2}$, a gap is observed in the spectral density of spin excitations. The dependence of the gap energy on the magnitude of the spin–phonon coupling, together with the approximating power function

$$\Delta_s/J \approx [(\alpha-\alpha_{c2})/\alpha_{c3}]^{0.50(8)}, \quad \omega_0 = J,$$

is depicted in Fig. 7 (solid curve). In the density of coupled spin–phonon excitations, one can single out a quasigap. The maximal density $\rho(\omega)$ corresponds to quasiparticles with zero energy and with the quasiparticle

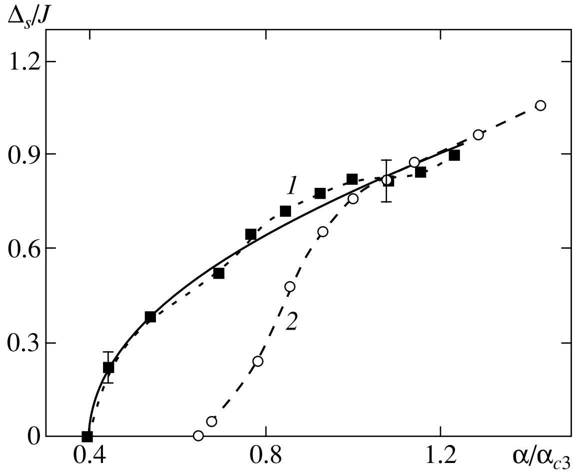


Fig. 7. The energy Δ_s of the gap in the spin excitation spectrum as a function of the spin–phonon interaction parameter for $\omega_0/J = 1$ (*1*) and 6 (*2*).

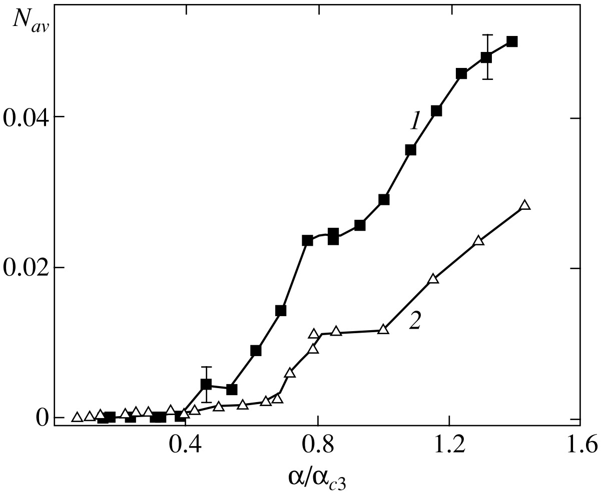


Fig. 8. Average occupation number for phonons as a function of the normalized spin–phonon interaction constant for $\omega_0/J = 1$ (*1*) and 6 (*2*).

excitation energy $\omega \approx 2(\omega_0 + 2J)$, $2(\omega_0 + J)$, and $3(\omega_0 + J)$ for $v_{ph} > v_m$ and $\omega \approx 4(\omega_0 + J)$ and $8(\omega_0 + J)$ for $v_{ph} < v_m$. For $\omega_0 \approx 4 \times 10^{12}$ Hz [14] and $J \approx 0.1$ eV, these estimates are in satisfactory agreement with the optical data on the absorption spectra of $Sr_2CuCl_2O_2$ [1], which show a broad peak in the vicinity of 4000 cm^{-1}, as well as with the values of excitation energy $E^{MC} \approx$ 4400 cm^{-1} calculated by the Monte Carlo method. The observed excitations reveal a close relation between the spin and lattice degrees of freedom in the CuO_2 plane.

On the plane including the upper boundary ω_0 of the region of acoustic vibrations and the spin–phonon coupling parameter α, three critical lines can be singled out. As the spin–phonon coupling constant attains the critical value with the approximation dependence $\alpha_{c1} = 0.16(2)\omega_0/J$, coupled spin–phonon excitations are formed analogously to the formation of polarons in systems with electron–phonon coupling. As the value of α increases, the quasiparticle density becomes higher and the spectral density $\rho(\omega = 0)$ of bound spin–phonon excitations has a finite value at $\omega = 0$ for $\alpha_{c2} = 0.39(6)(\omega_0/J)^{0.85(4)}$. A gap Δ_s appears in the spin excitation spectrum and the crystal symmetry is lowered. If we treat the gap width Δ_s as an order parameter of singlet pairs of spins, an inhomogeneous state consisting of a long-range magnetic order and a singlet state is realized in the range of parameters $\alpha_{c2} < \alpha < \alpha_{c3}$. This resembles the coexistence of the normal and anomalous phases in liquid helium and in a type II superconductor in a magnetic field. For constant α_{c3}, which can be approximated by the power dependence $\alpha_{c3} = 0.62(4)(\omega_0/J)^{0.85(6)}$, the long-range magnetic order disappears and a quantum spin liquid is formed.

The correlated state of lattice fluctuations can be destroyed by thermal phonons. The critical temperature can be estimated from the condition of equality of the thermal energy E_{heat} of phonons and the energy $E_{\text{bin}} = (N_{ph} - N_{ph,\,c2})\omega_0/2$ of the bound state of phonons. In the Debye approximation, $E_{\text{heat}} = 3\pi^4 k_B T^4/(5\Theta^3)$, where Θ is the Debye temperature, the corresponding critical temperature is

$$T^* \approx 0.74\Theta^{3/4}[0.02(\alpha - \alpha_{c3})/\alpha_{c3}]^{1/4}.$$

The dependence of the average number $N_{av} = (1/N)\sum_k N_k$ of phonons on the normalized value of the spin–phonon coupling constant is shown in Fig. 8. Lattice fluctuations are connected with magnetic fluctuations which change under the action of the magnetic field and temperature at $T \sim \Delta_s$. For $\alpha > \alpha_{c3}$, the lowest temperature at which the soliton lattice can be broken is determined by thermal phonons; for $\theta = 400$ K, we have $T^*_{c3} \approx 22$ K.

The low values of the magnetic moment $\sigma = 0.4(1)$ for Gd_2CuO_4 and Eu_2CuO_4 [15], which were obtained from elastic scattering of neutrons, as well as the values $\sigma = 0.35(4)$ determined from the electron spin resonances at Gd^{3+} ions in Eu_2CuO_4 [16], are probably due to the spin–phonon interaction with parameters $\alpha/\alpha_{c3} \approx$ 0.3 and 0.35 leading to the formation of coupled spin–phonon excitations. This changes the acoustic excitation spectrum. For example, an anomaly is observed in the lower branch of the acoustic phonon excitation spectrum along the ΓX direction in the isostructural compound Nd_2CuO_4 [14]. The corresponding changes in the lattice constant are on the order of 2×10^{-3} Å and are manifested in the X-ray spectra in the form of an ellipsoidal displacement of oxygen ions in the *ab* plane at right angles to the Cu–O bond [17].

The lifetime of coupled spin–phonon quasiparticles and the average relaxation time are proportional to the matrix element of the spin–phonon interaction operator for a phonon transition from the ground state to an excited state with a simultaneous change in the spin configurations of the two spins. In accordance with the "golden Fermi rule,"

$$\frac{1}{\tau_0} = \frac{2\pi}{\hbar}\left|\langle \mathrm{exc}|\hat{V}_{sph}|0\rangle\right|^2 N_{ph},$$

the quasiparticle lifetime $\tau_0 = 0.6 \times 10^{-7}$ s ($\langle \mathrm{exc}|$ denotes the excited state). The relaxation time distribution is described by the power law $P(\tau) \propto (\tau/\tau_0)^{5/4}$ for $\tau < \tau_0$.

4. CONCLUSIONS

Let us summarize the main results. The interaction between the elastic and magnetic subsystems leads to anisotropy of the elastic vibrations of the lattice as well as in magnetic properties; the change in the latter properties occurs at the three characteristic parameters of the spin–phonon interaction. For $\alpha = \alpha_{c1}$, coupled lattice and spin fluctuations are formed and the spherical symmetry of the spin–spin correlation functions is broken. For $\alpha = \alpha_{c2}$, a gap opens in the spin excitation spectrum and the crystal symmetry is lowered. The singlet state and the long-range antiferromagnetic order may coexist. For $\alpha = \alpha_{c3}$, the magnetic moment at a site vanishes and the antiferromagnet–quantum spin liquid phase transition takes place. The constant of the spin–phonon coupling corresponding to a decrease in the magnetic moments of quasi-two-dimensional antiferromagnets Gd_2CuO_4 and Eu_2CuO_4 are determined.

REFERENCES

1. V. V. Struzhkin, A. F. Goncharov, H. K. Mao, *et al.*, Phys. Rev. B **62**, 3895 (2000).
2. N. Ichikawa, S. Uchida, J. M. Tranquada, *et al.*, Phys. Rev. Lett. **85**, 1738 (2000).
3. T. Dahm, Phys. Rev. B **61**, 6381 (2000).
4. J. Takeya, Y. Ando, S. Komiya, *et al.*, Phys. Rev. Lett. **88**, 077001 (2002).
5. E. I. Golovenchits, V. A. Sanina, A. A. Levin, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **39**, 1600 (1997) [Phys. Solid State **39**, 1425 (1997)].
6. V. A. Pashchenko, S. Huant, A. A. Stepanov, and P. Wyder, Phys. Rev. B **61**, 6889 (2000).
7. A. A. Stepanov, P. Wyder, T. Chattopadhyay, *et al.*, Phys. Rev. B **48**, 12979 (1993).
8. A. F. Andreev and I. A. Grishchuk, Zh. Éksp. Teor. Fiz. **87**, 467 (1984) [Sov. Phys. JETP **60**, 267 (1984)].
9. S. S. Aplesnin, Fiz. Tverd. Tela (St. Petersburg) **39**, 1404 (1997) [Phys. Solid State **39**, 1246 (1997)].
10. N. V. Prokof'ev and B. V. Svistunov, Phys. Rev. Lett. **81**, 2514 (1998); N. V. Prokof'ev, B. V. Svistunov, and I. S. Tupitsyn, Zh. Éksp. Teor. Fiz. **114**, 570 (1998) [JETP **87**, 310 (1998)].
11. M. Suzuki, J. Stat. Phys. **43**, 883 (1986).
12. A. S. Mishchenko, N. V. Prokof'ev, A. Sakamoto, and B. V. Svistunov, Phys. Rev. B **62**, 6317 (2000).
13. S. Tang and J. E. Hirsch, Phys. Rev. B **37**, 9546 (1988).
14. E. Rampf, U. Schröder, F. W. de Wette, *et al.*, Phys. Rev. B **48**, 10143 (1993).
15. T. Chattopadhyay, J. W. Lynn, N. Rosov, *et al.*, Phys. Rev. B **49**, 9944 (1994).
16. C. Rettori, S. B. Oseroff, D. Rao, *et al.*, Phys. Rev. B **54**, 1123 (1996).
17. P. Adelman, R. Ahrens, G. Czjzek, *et al.*, Phys. Rev. B **46**, 3619 (1992).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 978–988.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 1090–1102.
Original Russian Text Copyright © 2003 by Voronkova, Grigor'ev, A. Zuev, V. Zuev, Kiryukhin, Manykin, Chmyrev, Shcherbakov.

SOLIDS
Structure

Changes in the State of Impurities and the Corresponding Force Fields in Silicon after High-Temperature Cyclic Annealings

G. M. Voronkova[a], V. V. Grigor'ev[a], A. V. Zuev[a], V. V. Zuev[a,*], A. D. Kiryukhin[a], É. A. Manykin[a], V. I. Chmyrev[b], and S. A. Shcherbakov[c]

[a]*Moscow State Institute of Engineering Physics (Technical University), Kashirskoe sh. 31, Moscow, 115409 Russia*
[b]*Kurnakov Institute of General and Inorganic Chemistry, Russian Academy of Sciences, Leninskiĭ pr. 31, Moscow, 117907 Russia*
[c]*Dorodnitsyn Computer Center, Russian Academy of Sciences, Moscow, 119991 Russia*
**e-mail: solid@mephi.ru*

Received April 10, 2003

Abstract—The influence of high-temperature annealings followed by cooling under different conditions on changes in the specific resistance and photoconductivity relaxation time in silicon and on the formation of the acceptor state of gold, when gold diffused from a layer vacuum-deposited on the surface, was studied in experiments including sequences of processes conducted using various temperatures, cooling conditions, and surface states. The experimental data are analyzed based on the concept of changes in the state of impurities present in silicon and in the corresponding dynamic and static force fields. *© 2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The displacements of impurity atoms under various thermal treatments and during in-diffusion should noticeably depend on the spatial characteristics and values of force ("elastic") fields created by defects of various origins. This dependence is, for instance, a basis of inside gettering, or collecting impurities on the corresponding defect formations such as oxygen precipitates under specially selected thermal treatment conditions [1]. Special fields localized near the surface are also formed under outside gettering conditions, when the surface of a material contains some imperfections such as micropores [2]. Clearly, force fields may be both static and dynamic with respect to the diffusing impurities. Dynamic fields change as the impurity diffuses because of the evolution of the spatial distribution of defects (in particular, vacancies and interstitial atoms), their concentration, and fluxes. When impurity atom nanoformations interact in real silicon, changes in their structure and composition, which influence the spectrum of electron levels in the forbidden band, can manifest themselves as various macrocharacteristic peculiarities, for instance, as ultra-accurate compensation [3], bistability of defects (in particular, impurity atoms) with a negative electron correlation energy (U^- defects) [4], and changes in photoconductivity spectra, carrier lifetimes [1], the type of conductivity, and specific resistance values [5].

The diversity of combinations of various external conditions of actions on a real crystal, such as temperature, annealing duration and conditions, the rate of cooling, etc., results in the production of formations that differ in shape, structure, and composition. These formations may have different systems of electronic levels in the forbidden band of the semiconductor, which results in the diversity of electrophysical properties. Our goal is to obtain properties of the greatest interest to science and technology, such as a high specific resistance and a comparatively long lifetime of minority carriers for designing detectors of particles. This requires purposeful studies of interaction phenomena in impurity systems and the influence of their force fields on the formation of electronic levels by impurities during their diffusion into a real crystal.

In this work, we performed a system of experiments to determine to what extent the internal state of a real silicon crystal changes under cyclic thermal treatments that precede the in-diffusion of gold and influence the formation of the acceptor state of gold atoms. These changes were monitored based on obvious indications of gold penetration into silicon, such as binding part of free electrons by impurity atoms in the deep acceptor state and the formation of a recombination channel. As a consequence, the specific resistance ρ increased and the photoconductivity relaxation time τ changed.

2. EXPERIMENTAL

We used KEF-20 electronic silicon grown by the Czochralski method in the form of plates about 0.15 cm thick with the specific resistance (determined by the four-probe method) $\rho \approx 25$–$31\ \Omega$ cm, which corresponded to a concentration of free electrons of

$n \approx (1.72–1.39) \times 10^{14}$ cm^{-3}, their mobility $\mu_n \approx$ 1450 cm^2/(V s) [6], and $\tau \approx$ 51–60 μs. Time τ was determined by a noncontact method from time changes in the reflected power of a microwave field, which followed time changes in nonequilibrium photoconductivity excited in the material by pulsed laser radiation with a wavelength of λ = 1.06 μm [7]. The content of oxygen was 10^{18} cm^{-3}. Prior to and after thermal treatment, the material was subjected to mechanical and chemical treatments according to the well-known scheme [8]. The samples were plates of various shapes. High-temperature treatments were performed under continuous evacuation at 0.1 Pa in a quartz tube, which was preliminarily etched and calcined at 1150°C for several hours. The conductivity type was controlled using a thermoprobe.

Preliminary thermal treatment of the samples was performed in cycles (the duration of each cycle was 5 h) during t = 5, 10, 25, 35, and 50 h at annealing temperatures T = 850, 900, 1000, and 1100°C. The samples for each annealing temperature and each series of measurements were cut in the form of plates from one plate to achieve uniformity of initial material parameters.

At a high temperature (1100°C), a noticeable deposit was formed on the walls of the cold quartz tube section and the surface of the samples turned dull. The samples were then weighed on a VLR-200 balance before and after thermal treatment.

Gold was thermally deposited on initial and thermally treated (operation no. 1) silicon samples until a visible film was formed [9]. Gold diffusion was performed at T_{Au} = 900°C for t_{Au} = 1 h with subsequent slow cooling (operation no. 2). To reach a region with a constant specific resistance, the samples were subjected to layer-by-layer etching. The thickness of the removed film was 100–150 μm. The next operation (no. 3) was thermal treatment of the samples at T = 900°C for t = 1 h followed by quenching in air. After the ρ and τ parameters were measured at T = 300 K, the samples were again thermally treated at T = 900°C for t = 1 h, which was followed by slow cooling (operation no. 4). This procedure was used to reveal various electronic states of gold atoms present in the material after the initial diffusion of gold followed by slow cooling and the transition of the gold atoms from an electrically inactive to the active state and vice versa. Clearly, experiments organized as described above also gave information about the thermal stability of the material at 900°C in a vacuum.

Because the gold diffusion coefficient at 850°C is lower than at higher temperatures, the diffusion was conducted for 5, 25, and 50 h with slow cooling after termination of diffusion. Property variations were controlled by simultaneously performing the diffusion of gold into the initial material and the thermally treated samples.

Another series of experiments were conducted to reveal the influence of preliminary long-term thermal treatment of silicon on lateral gold atom displacements during high-temperature diffusion. These experiments were performed using long plates ($L \approx$ 1–2 cm). Gold was thermally deposited (in a vacuum) along half the plate length on one or both plate sides. Prior to deposition, some plates were subjected to high-temperature thermal treatment (T = 850°C, t = 50 h); this was done in 5 h long cycles followed by slow cooling. The diffusion was performed at the temperature of thermal treatment (850°C) in 5-h cycles over 15–50 h with subsequent slow cooling (in a furnace) to room temperature. After each cycle, part of samples were removed from the process. The surface of the thermally treated samples was not processed before four-probe and relaxation time measurements. This was possible because no noticeable oxide film was formed at 850°C in vacuum, and the surface remained visually the same as after chemical etching. Specific resistance measurements were performed both on the side opposite to that onto which gold was deposited (one-sided deposition) and on the side with deposited gold. Because of the small thickness of the deposited gold layer, time τ measurements could be performed along the whole plate length; laser radiation (λ = 1.065 μm) penetrated deeply into the sample; that is, they excited nonequilibrium carriers, whose presence influenced the microwave field reflection value. The lateral properties of the plates were measured at a distance of approximately 0.5 cm from the edge of the deposited gold film.

We intentionally selected silicon with a majority carrier concentration lower than that used in [3, 8, 10] to observe noticeable thermal treatment effects on the specific resistance of samples at lower temperatures and reasonable thermal treatment durations [5].

3. RESULTS

3.1. Preliminary Thermal Treatment

The ρ and τ parameters measured after preliminary thermal treatments are shown in Figs. 1–4. The duration of thermal treatment t is indicated for each pair of parameters (Figs. 2–4) or, if possible, by special symbols (Fig. 1). The ρ and τ values decreased 2.5–3 and 14–50 times, respectively, compared with those for the initial sample. The smallest change in τ was observed after preliminary thermal treatment at 1100°C. A decrease in the specific resistance was evidence of the formation of thermal donors with a comparatively low embedding energy, and a decrease in the photoconductivity relaxation time τ showed that the embedding energy of the defects produced in the forbidden band was sufficient for increasing the probability of recombination.

This can be explained under the assumption that the produced thermal donors are characterized by a spread of electron binding energies in the forbidden band, as suggested in [1, 11], and/or not only shallow thermal donors but also deep acceptor states, whose concentra-

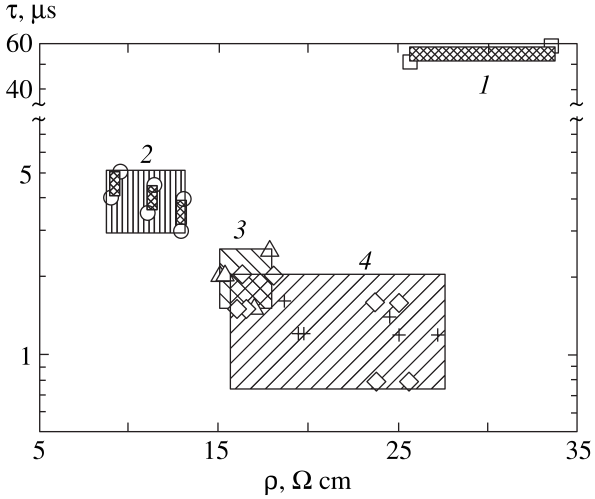


Fig. 1. The influence of duration *t* of preliminary thermal treatment at 850°C on specific resistance ρ and photoconductivity relaxation time τ at room temperature: (*1*) initial sample, (*2*) $t = 5$ h, (*3*) $t = 25$ h, and (*4*) $t = 50$ h (two series). The regions of the spread of experimental parameter values are hatched.

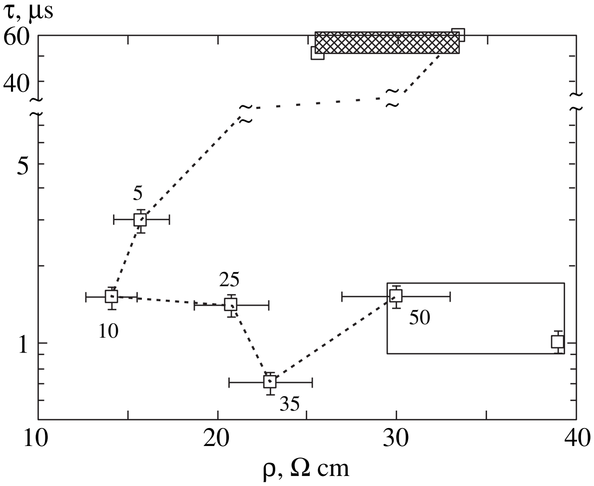


Fig. 2. Distribution of τ and ρ parameters depending on the duration of thermal treatment (given in hours near symbols) at 900°C. Parameter values for two different samples after thermal treatment for 50 h are conventionally shown by a rectangle. The region of the spread of initial sample parameters is hatched.

tion is initially lower than that of thermal donors, are formed. The ρ value increases as the duration of preliminary thermal treatment increases and reaches the initial ρ_0 value. At high temperatures (1000 and 1100°C), it begins to exceed ρ_0 at treatment durations of 20–50 h. At 1100°C, we invariably observed a peculiar behavior of the resistance. While remaining higher than the initial ρ_0 value, it decreased as the duration of preliminary thermal treatment increased. Experiments were repeated with three series of samples; the same trend was observed for all of them (Fig. 4). As at equal annealing temperatures, defects whose size exceeds the critical value can only grow and smaller defects can only continue to diminish with time; the behavior of the specific resistance described above can reasonably be related to a size, composition, and structure dependence of the electronic characteristics of the growing defects [1, 12], namely, to the disappearance of their donor properties caused by an increase in size and changes in compositions and structures. Simultaneously, acceptor-type formations initially grown from nuclei of the corresponding critical size continue to grow and retain their acceptor properties. Their action, namely, free electron capture from the conduction band, begins to prevail over that of donor forms, which donate electrons to the conduction band. At a compara-

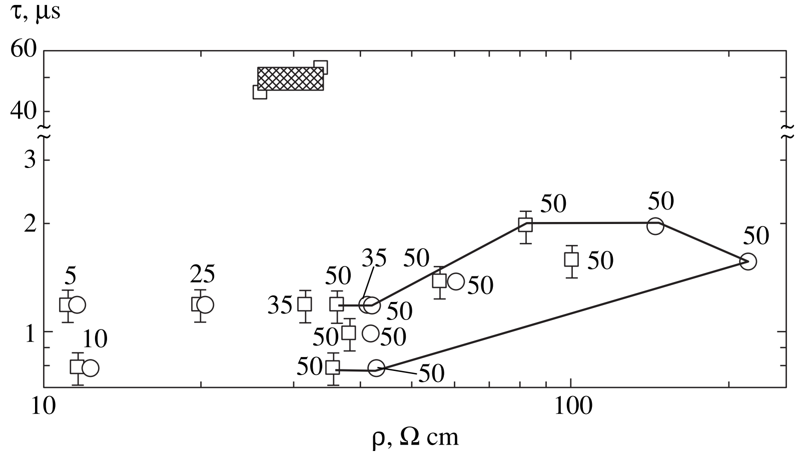


Fig. 3. Distribution of τ and ρ parameters depending on the duration of thermal treatment (given in hours near symbols) at 1000°C; □ and ○ correspond to specific resistance measurements on the opposite sides of each plate subjected to thermal treatment. The region of the spread of parameters for six samples subjected to thermal treatment for 50 h is outlined. The region of the spread of initial sample parameters is hatched.

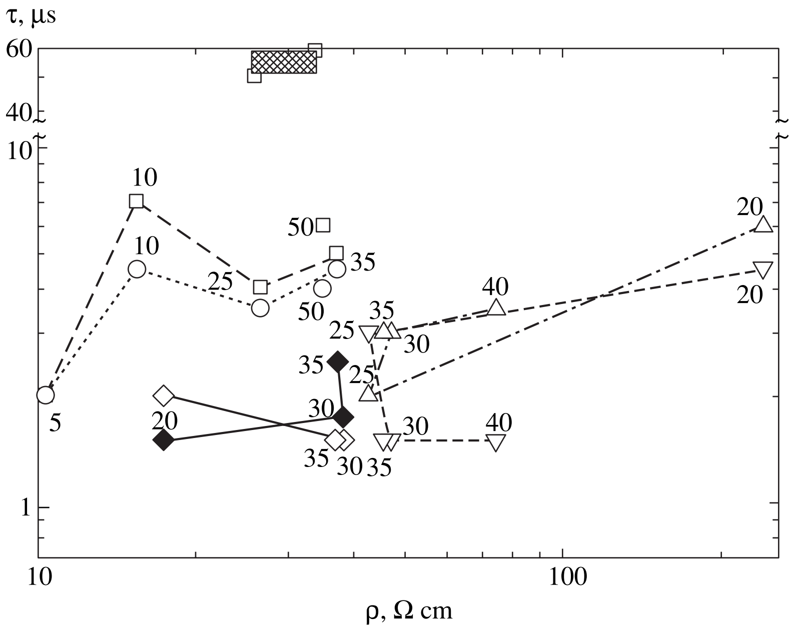


Fig. 4. Distribution of τ and ρ parameters depending on the duration of thermal treatment (given in hours near symbols) at 1100°C; ○ and □ correspond to photoconductivity relaxation time measurements using the first and last pulsed laser action peaks in the first thermal treatment series; ▽ and △, the same for the second series, and ◇ and ◆, the same for the third series.

tively low temperature of preliminary thermal treatment (850°C), the ρ value increased but did not exceed that of the initial sample (Fig. 1). This observation and the data obtained for high temperatures are evidence of the influence of the diffusion coefficient of atoms (for instance, oxygen) that create donor forms on the growth of these forms. Time τ, which decreased after the first cycle, remained small during subsequent cycles; that is, the concentration of defects that determined the rate of recombination did not change substantially. We can only note that the τ value after preliminary thermal treatment at 1100°C was several times larger than at other temperatures. At this temperature, substantial outflow of a substance (supposedly, silicon or silicon combined with oxygen) from the bulk was noticeable for samples with a free surface, which caused weight loss in a sample (by 1–2% of 100–150 mg). Under these conditions, intense generation of vacancies from the surface to bulk material can occur. Saturation with vacancies can favor displacements of impurities inside crystals. These displacements are different from those that occur at lower temperatures and leave the surface of the material bright as at the beginning of the process. Accordingly, the parameters of impurity formations can be different, which manifests itself by larger τ values and the peculiarities of ρ variations depending on the duration of preliminary thermal treatment.

3.2. The Influence of Preliminary Thermal Treatment on Manifestations of the Acceptor State of Gold

The data given in Fig. 5 show how preliminary thermal treatment influences manifestations of the acceptor state of gold atoms that are capable of binding free electrons from the conduction band. After the diffusion of gold into the initial sample during 5, 25, and 50 h at 850°C followed by slow cooling, the resistance did not increase above the initial ρ_0 value. It even slightly decreased, as is characteristic of the formation of thermal donors accompanied by a decrease in τ. A similar behavior of ρ and τ after the diffusion of gold was observed for the sample subjected to preliminary thermal treatment for 5 h. However, the samples that were preliminarily thermally treated for 25 and 50 h gave radically different results, namely, the specific resistance ρ increased to $(1–8) \times 10^4$ Ω cm after the diffusion of gold during 25 and 50 h followed by slow cooling. This is a consequence of binding the majority of free electrons at deep energy levels capable of localizing electrons even at room temperature. The conductivity type remained unchanged, which was evidence of the predominance of electron exchange between these levels and the conduction band states.

Note that some of the data given in Fig. 5 refer to the samples that were heated at 850°C for 50 h, either during gold diffusion or during preliminary thermal treatment followed by gold diffusion. Their ρ values are, however, substantially different. It follows that thermal treatment causes the appearance of differences in the static and dynamic internal fields. These fields influence binding diffusing gold in the electrically active and inactive states.

The differences in the τ values are less noticeable. We can only note that the τ value after preliminary thermal treatment for 25 and 50 h at 850°C followed by

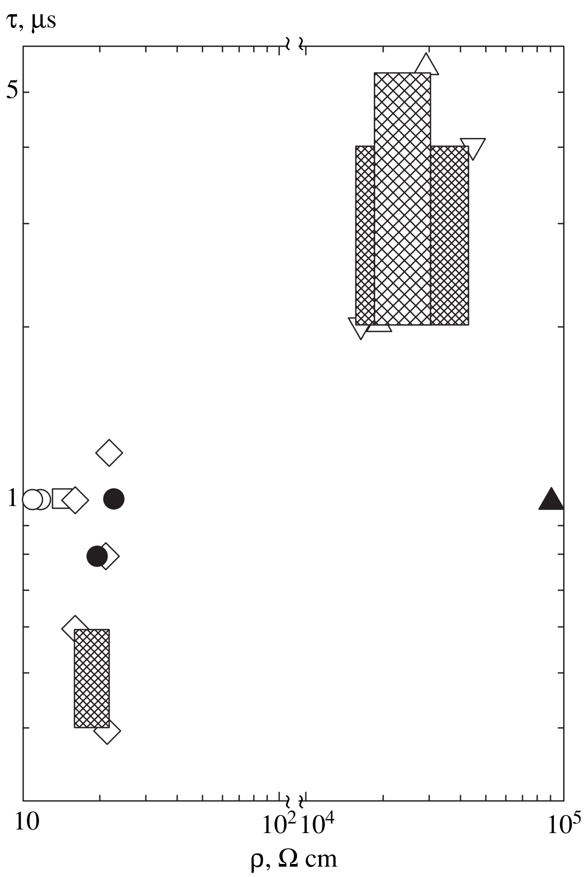


Fig. 5. The influence of duration t of preliminary thermal treatment at 850°C followed by gold diffusion at the same temperature ($T_{\rm Au}$ = 850°C) on specific resistance ρ and photoconductivity relaxation time τ at room temperature: □, $t = t_{\rm Au}$ = 5 h; ○, initial sample, $t_{\rm Au}$ = 5 h; △, $t = t_{\rm Au}$ = 25 h; ▽, t = 50 h, $t_{\rm Au}$ = 25 h; ◇, initial sample, $t_{\rm Au}$ = 25 h; ▲, $t = t_{\rm Au}$ = 50 h; and ●, initial sample, $t_{\rm Au}$ = 50 h. Hatched regions give the spread of experimental values.

gold diffusion for 25 h was larger than the τ value after gold diffusion into the initial material and the material subjected to preliminary thermal treatment for t = 5 h. At t = 50 h, an increase in the duration of gold diffusion increases ρ and decreases τ. It is physically clear that these changes are related to an increase in the concentration of gold acceptor states.

After the diffusion of gold ($T_{\rm Au}$ = 900°C, $t_{\rm Au}$ = 1 h, slow cooling) in samples subjected to thermal treatment at T = 900, 1000, and 1100°C, the specific resistance increased (Fig. 6, operation no. 2). The longer the time of preliminary thermal treatment, the larger this increase. An exception was the sample annealed at T = 1100°C, whose specific resistance increased to a greater extent in t = 35 h than in t = 50 h (Fig. 6c). This behavior has something in common with that of the ρ value after preliminary thermal treatment at 1100°C (Fig. 4). Quenching the same samples after the removal of the gold layer that remained on the surface and heating them at 900°C for an hour increased their resistance (Fig. 6, operation no. 3). Subsequent annealing (T = 900°C, t = 1 h) of the samples subjected to thermal treatment at T = 900°C (t = 5–50 h), 1000°C (t = 5–35 h), and 1100°C (t = 5–25 h) followed by slow cooling returned ρ to the initial values. After preliminary thermal treatment at T = 1000°C (t = 50 h) and 1100°C (t = 35–50 h), annealing followed by slow cooling did not decrease ρ, which remained high (Figs. 6b, 6c). After diffusion ($T_{\rm Au}$ = 900°C, $t_{\rm Au}$ = 1 h) and slow cooling, the range of times τ broadened depending on the duration of thermal treatment (Fig. 7, operation no. 2). Subsequent quenching in the absence of gold on the surface at 900°C for 1 h noticeably decreased the spread of τ values (Fig. 7, operation no. 3). Further annealing at 900°C for 1 h again increased the spread of τ depending on the duration of thermal treatment (Fig. 7, operation no. 4).

3.3. Diffusion of Nonuniformly Deposited Gold

The dependences of the specific resistance on the duration of gold diffusion at 850°C are shown in Fig. 8 for various crystal regions; the crystals were subjected to preliminary thermal treatment at 850°C for 50 h. The same figure contains data on the initial material obtained under similar conditions (gold diffusion followed by slow cooling).

The behavior of the specific resistance was as follows:

(1) after gold diffusion, the specific resistance of the samples subjected to preliminary thermal treatment was always an order of magnitude higher than that of the initial samples;

(2) if a gold film was deposited on one side of a silicon plate, an increase in the specific resistance was larger for the opposite plate side at a distance of plate thickness from the gold film;

(3) lateral gold diffusion caused a systematically larger increase in the specific resistance at the side opposite to that where the gold was deposited;

(4) if gold was deposited on both sides, lateral diffusion equally increased specific resistance on both sides; this increase was, however, lower than with one-sided deposition, which could be explained by the influence of the state of the surface on the penetration of gold atoms and preserved their electrically active state.

Gold diffusion into the initial material not subjected to preliminary thermal treatment followed by slow cooling always decreased the specific resistance in comparison with its initial value, as is characteristic of the formation of thermal donors, which occurs in parallel with gold diffusion. A small increase in the specific resistance with time can be caused by a change in the structure of thermal donors. Fluxes of atoms (supposedly oxygen) that arise in the formation of thermal donors can shift gold atoms to regions and into formations where they are electrically inactive.

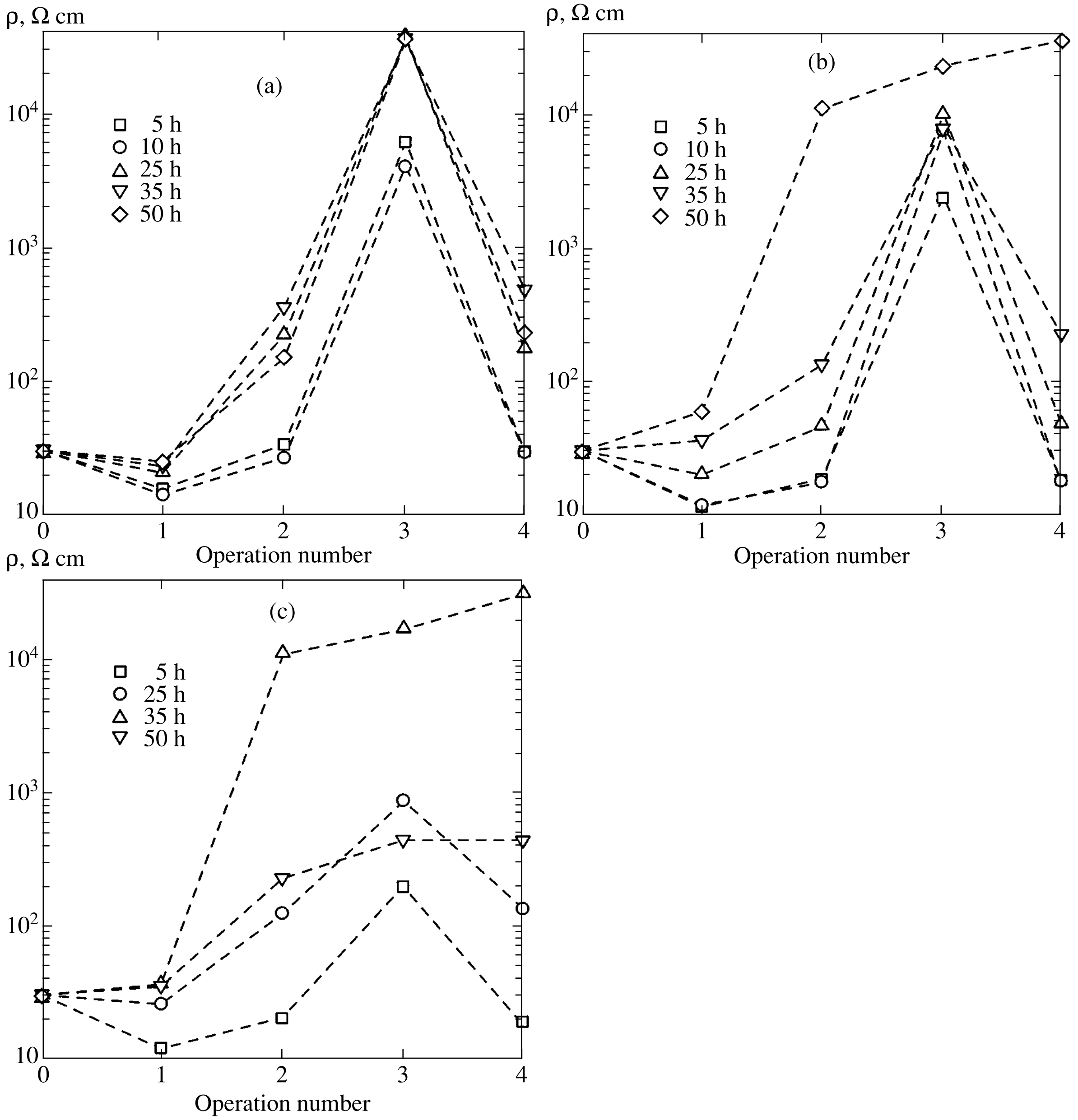


Fig. 6. Dependences of specific resistance on sample processing. Operation numbers: 0, initial sample; 1, thermal treatment at T = 900 (a), 1000 (b), and 1100°C (c) for various time intervals indicated in the figures; 2, T_{Au} = 900°C, t_{Au} = 1 h (slow cooling), etching; 3, T = 900°C, t = 1 h (quenching); and 4, T = 900°C, t = 1 h (slow cooling). For convenience, points corresponding to equal preliminary thermal treatment durations are connected by dashed lines.

As concerns the photoconductivity relaxation time, it always and everywhere decreased compared with the initial sample. However, as previously, it was systematically higher after gold diffusion in the samples that were subjected to preliminary thermal treatment and equaled 2–5 μs against 0.8–1.4 μs in the initial samples.

4. PRINCIPAL EXPERIMENTAL RESULTS AND DISCUSSION

1. Under all preliminary thermal treatment conditions, the specific resistance decreased after the first annealing cycle and increased during subsequent cycles, even to values higher than the specific resistance of the initial samples (ρ_0) at high temperatures. Simultaneously, the photoconductivity relaxation time τ decreased approximately 50 times after the first cycle compared with the initial samples and remained small, on the order of 1 μs (Fig. 7). Such parameter behaviors can be explained using the literature data on thermal donors originating from oxygen [1, 5, 11–13]. At high annealing temperatures (550–850°C), thermal donors different from those formed at comparatively low tem-

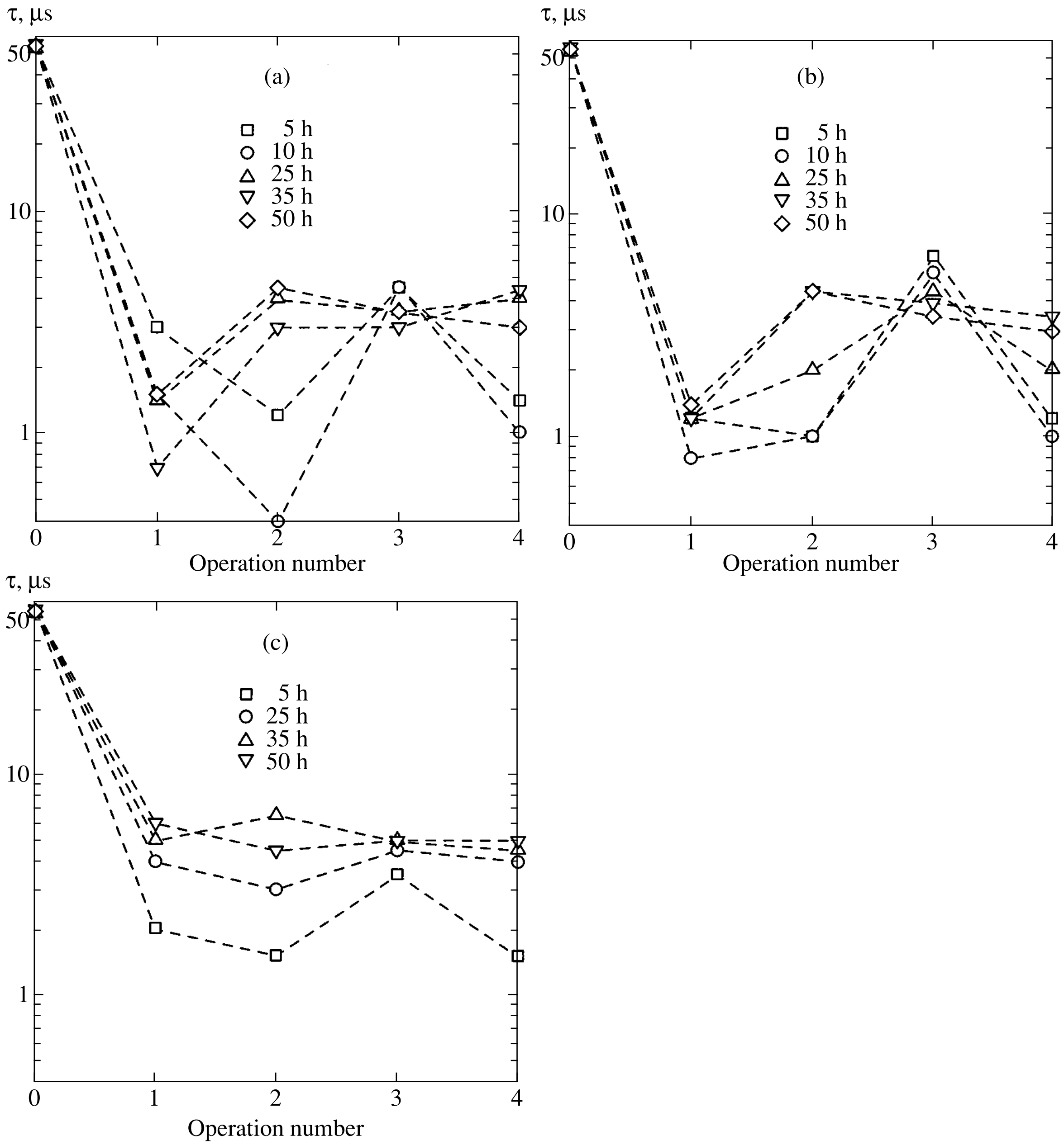


Fig. 7. Dependences of photoconductivity relaxation times on sample processing. Operation numbers are the same as in Fig. 6.

peratures (450°C) can appear. Small donors that correspond to small oxygen clusters are known to be formed at 450°C [5, 13, 14]. At higher temperatures (550–800°C), these thermal donors decay and are replaced by new ones with the embedding energies of electronic states near the middle of the forbidden band, which should manifest itself by a short lifetime of minority carriers [1]. These thermal donors were called new oxygen donors in [11]; they represented new SiO_x phase formations. At the interface between this phase and the silicon matrix, surface states providing intense recombination are formed. The electrons transferred from the high-energy state in the wideband SiO_x phase cause an increase in the conductivity of the silicon matrix, that is, a decrease in the resistance of electronic silicon after thermal treatment [11, 14]. Above 900°C, annihilation of thermal donors remaining after growth, however, occurs in hole materials [5]. Transformations in the system of oxygen precipitates are determined by the presence of nuclei of the corresponding structure with a critical growth size. The structure and size, in turn, depend on the temperature and, of course, the concentration of dispersedly distributed interstitial oxygen. According to recent experimental data, these processes can also be influenced by the presence of carbon, which can be kicked out of the nodal state by interstitial silicon formed as a result of the removal of stresses as a result of the growth of oxygen precipitates. This interstitial carbon can easily diffuse to oxygen precipitates, attach to them, and change their electric activity, that is, change the system of electron energy levels in the forbidden band [13]. In our experiments, we used slow cooling after every cycle; that is, the system could fairly longly exist at a low temperature. Under these conditions, a noticeable concentration of defects of various structures and compositions can be formed after the first cycle. These defects are responsible for both shallow and deep donor states. During subsequent annealing cycles, when the samples are held at high temperatures, the precipitates responsible for shallow donor states dissolve because their critical size does not correspond to the critical growth size at these temperatures. The concentrations of larger SiO_x phase precipitates with deeper electronic states can remain substantial also during subsequent slow coolings and sufficient for providing an effective channel of carrier recombination. This manifests itself by a short photoconductivity relaxation time, which does not return to the value characteristic of the initial material. The concentration of disperse oxygen decreases after each cycle because of the formation of precipitates of various structures. This follows from a decrease in infrared absorption in the corresponding wave number range (1100–1300 cm^{-1}), more exactly, at the absorption maximum of disperse oxygen (1107 cm^{-1}) [12]. Note that, judging from the retention of the short photoconductivity relaxation time observed in all our experiments, the concentration of deep electronic levels related to thermal donors remained noticeable in spite of high annealing temperatures, at which no noticeable concentration of thermal donors was detected in [5] in specific resistance measurements, most likely because cooling was fast in [5], whereas it was slow in our experiments. In addition, the relaxation time of photoconductivity observed in our experiments was proportional to the lifetime of carriers and could serve as a parameter more sensitive to the formation of a low concentration of deep levels than specific resistance changes. It is easy to imagine a situation when the concentration of newly formed defects adds little to the equilibrium concentration of electrons or holes, whereas their lifetime decreases rapidly. For instance, at a 10^{13} cm^{-3} concentration of defects and a recombination coefficient of $\gamma \approx 5 \times 10^{-8}$ cm^3/s, the lifetime of holes is on the order of 2 μs; that is, it decreases 25 times compared with the initial value of 50 μs, whereas an increase in the conductivity against the background of the initial concentration 5×10^{13} cm^{-3} amounts only to 2%. This increase cannot be reliably measured, considering measurement error limits. It follows that preliminary annealings followed by slow cooling substantially change the internal state of the material compared with the initial sample; that is, they change internal force fields. This is substantiated by changes in the ρ and τ parameters depending on the temperature and duration of processes. Experiments with the gold diffusion lend additional support to this conclusion. The formation of gold acceptor states capable of binding conduction electrons and decreasing conductivity substantially depends on force fields in the crystal.

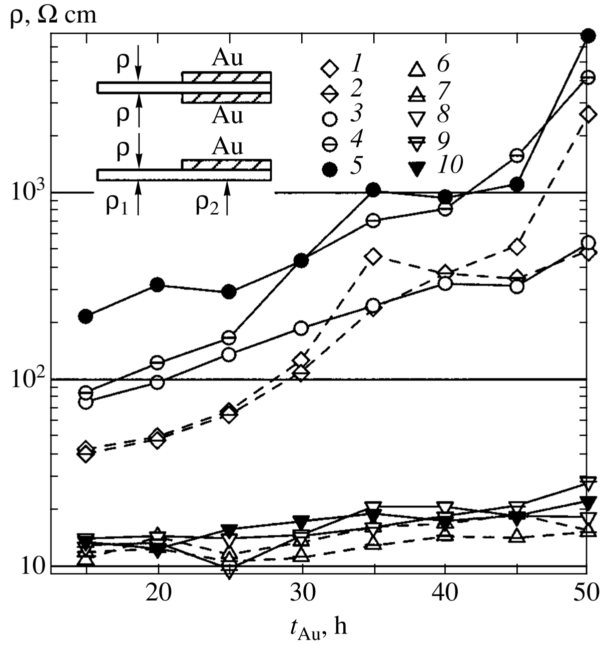


Fig. 8. Changes in ρ depending on the duration of gold diffusion at T_{Au} = 850°C (*1–5*) into the sample subjected to preliminary thermal treatment for t = 50 h at T = 850°C followed by slow cooling and (*6–10*) into the initial sample. Shown at the top by arrows and corresponding denotations (ρ, ρ_1, ρ_2) are the points of specific resistance measurements on a plate by the four-probe method; (*1, 2, 6, 7*) diffusion of gold from both sides, (*3–5, 8–10*) diffusion of gold from one side, (*3, 8*) ρ, (*4, 9*) ρ_1, and (*5, 10*) ρ_2.

2. Under correlated conditions, that is, when the temperature of gold diffusion equals that of preliminary thermal treatment, it is reasonable to assume that the formation of defects and the corresponding elastic fields more weakly influences the penetration of gold. If the duration of gold diffusion is shorter than that of preliminary annealings, the formation of precipitates and, accordingly, changes in internal fields differently affect the material that was subjected to preliminary thermal treatment and the initial material, where gold diffusion occurs in parallel with the formation of impurity inclusions accompanied by dynamic changes in elastic fields because of impurity atom fluxes toward precipitation sites. This is substantiated by the experimental data (Fig. 5), which show that the diffusion of gold at 850°C for 25 and 50 h after preliminary thermal treatments increases the specific resistance by orders of magnitude in comparison with the specific resistance after the diffusion of gold into the initial sample not subjected to preliminary thermal treatment. It should be emphasized that these results were obtained under identical conditions for samples that were slowly cooled after the diffusion of gold. Note that, if the samples were quenched, we would obtain a high resistance for both of them [10]. Gold atoms remain in the acceptor state after slow cooling; this state binds part of the conduction electrons and increases the specific resistance of the material compared with the initial sample. This is evidence of a higher uniformity of force fields (static and dynamic) in crystals subjected to preliminary thermal treatment. When gold diffuses into the initial material, which is subsequently slowly cooled (at a rate of 1°C/min or in a furnace), gold atoms are kicked out of nodal positions, where they have acceptor properties, by interstitial silicon atoms. Simultaneously, gold atoms travel over interstices toward regions where they are electrically inactive, that is, where they do not possess systems of electronic levels within the forbidden band of silicon, under the action of elastic force fields created by oxygen and, possibly, carbon atoms that move to regions of their accumulation. Such entrainment of gold atoms during slow cooling gives a physically reasonable explanation of the difference in the properties of slowly cooled and quenched samples.

Quite significant are the results of experiments with diffusion from a nonuniformly deposited source of gold into the samples subjected to preliminary thermal treatment at 850°C for 50 h and into the initial material (Fig. 8) with subsequent slow cooling. These results lend support to the abovementioned peculiarities of an increase in the resistance both under the deposited layer and at the side of it at a distance of about 0.5 mm in the samples that were subjected to thermal treatment.

Also informative are experiments with changes in the internal crystal fields that influence gold diffusion. Consider the results presented in Figs. 6 and 7 for different durations of preliminary thermal treatments at 900, 1000, and 1100°C and subsequent two-sided gold diffusion for 1 h at 900°C followed by slow cooling. Studies of these samples were continued by annealing at 900°C for 1 h with quenching (rapid cooling in air) after the removal of the gold film and subsequent annealing at 900°C for 1 h followed by slow cooling. Under correlated conditions (thermal treatment and gold diffusion at 900°C), the specific resistance noticeably increased after gold diffusion (Fig. 6a, operation no. 2). The longer the duration of thermal treatment, the larger the increase in the specific resistance, which is in close agreement with the model described above. However, the quenching of these samples with gold removed from their surfaces after annealing at 900°C for 1 h showed that a noticeable fraction of gold atoms could leave the regions where they were electrically inactive, pass to electrically active states, additionally bind electrons from the conduction band, and increase the resistance (Fig. 6a, operation no. 3). Subsequent annealing at 900°C for 1 h followed by slow cooling allowed gold atoms to return to the regions of their electrical inactivity; the specific resistance then returned to the value it had before quenching (Fig. 6a, operation no. 4).

A similar specific resistance behavior depending on the operation performed was observed for thermal treatments at 1000°C for 5, 10, 25, and 35 h (Fig. 6b, operation nos. 1–4) and at 1100°C for 5 and 25 h (Fig. 6c, operation nos. 1–4).

An increase in the resistance after gold diffusion followed by slow cooling was, however, maximum in materials subjected to preliminary thermal treatments at 1000°C (t = 50 h) and 1100°C (t = 35 and 50 h) (Figs. 6b, 6c, operation no. 2). Subsequent annealing at 900°C with quenching (operation no. 3) and slow cooling (operation no. 4) did not cause a decrease in the resistance, which remained unchanged or even increased (Figs. 6b, 6c, operation nos. 3, 4).

Once more note changes in internal force (elastic) fields in materials depending on the duration and temperature of preliminary thermal treatment, which manifest themselves in different specific resistance values after gold diffusion at 900°C for 1 h followed by slow cooling (Fig. 6, operation no. 2), that is, in different concentrations of electrically active gold atoms capable of binding part of the free electrons present in the conduction band before diffusion.

As concerns the photoconductivity relaxation time τ, its changes are also determined by structural changes in the system of electrically active states after different preliminary thermal treatments of different durations (Fig. 7, operation no. 1), after gold diffusion followed by slow cooling (Fig. 7, operation no. 2), after annealing followed by quenching (Fig. 7, operation no. 3), and after annealing at 900°C for 1 h (Fig. 7, operation no. 4). The influence of the structure of defects on the rate of recombination after preliminary thermal treatment determines the complex dependence of τ on the duration of processes (Fig. 7, operation no. 1), namely, the longer the duration of thermal treatment at 1100°C, the larger the τ value, whereas a nonmonotonic dependence is observed at 1000 and 900°C. After the diffusion of gold at 900°C for 1 h followed by slow cooling, the dependence of τ on the duration of thermal treatment remains complex. Sometimes, the sequence of τ values corresponding to different thermal treatment durations changes and the spread of τ values increases (Fig. 7, operation no. 2). Annealing at 900°C for 1 h after the removal of the layer of gold and subsequent quenching, when the resistance of all samples increases (Fig. 6, operation no. 3), causes pronounced accumulation of τ values in the range 3–6 μs (Fig. 7, operation no. 3). If thermal treatment is performed at 1000°C, we observe a well-defined dependence of τ on treatment duration, namely, the longer the duration of thermal treatment, the smaller the τ value (Fig. 7b, operation no. 3). Subsequent annealing at 900°C for 1 h followed by slow cooling, when part of the gold atoms resume the electrically inactive state, again distributes τ values over thermal treatment durations in approximately the same order as was observed after operation no. 2; that is, after the diffusion of gold followed by slow cooling. The convergence of τ times to one value (operation no. 3) is evidence of more uniform recombination parameters formed after quenching compared with the state after slow cooling (operation nos. 2, 4). Physically, this is plausible. Indeed, during slow cooling, gold atoms have enough time to “track” the difference in internal fields formed in different thermal treatment cycles. After quenching, there remain a large number of gold atoms under identical conditions (compare the specific resistance data at T = 900°C, t = 1 h after quenching samples annealed without gold on the surface). Importantly, the photoconductivity relaxation time and, accordingly, the lifetime of carriers after the introduction of gold atoms can be longer than those observed after preliminary thermal treatment alone (compare operation nos. 1, 3 in Figs. 7a–7c). The introduction of additional recombination levels such as gold atoms in the acceptor state should seemingly intensify recombination processes (decrease τ) that occur in parallel with those in which formations produced after thermal treatment participate. However, a different behavior of τ may be evidence of the exclusion of part of defect levels from recombination because of the formation of electrically active gold atom states on their base. Such suggestions were made, for instance, in [15]. Naturally, we should then assume that the states excluded from recombination had recombination coefficients larger than those of gold. What is more, times τ after prolonged thermal treatments followed by the diffusion of gold (t = 25, 35, and 50 h) remain unchanged both after annealing with quenching (operation no. 3) and after subsequent annealing with slow cooling (operation no. 4); that is, we observe stable relaxation times at 900°C for 1 h; similarly, we observe stable specific resistances of the samples subjected to preliminary thermal treatment at T = 1000°C (t = 50 h) and 1100°C (t = 35 and 50 h).

Note that a decrease in τ after thermal treatments prevents us from identifying the time caused by electron recombination on acceptor gold in the pure form. Accordingly, we cannot observe comparatively longer times τ that are possible when gold exactly compensates the phosphor, that is, under the conditions when the concentration of electrons in the conduction band remains equal to the number of free gold atom levels [16].

If we assume that the cross section of electron capture by the neutral acceptor gold state is $\sigma_n = 8.5 \times 10^{-17}$ cm^2 [17], the recombination coefficient at T = 300 K is $\gamma_n = 1.5 \times 10^{-9}$ cm^3/s, and the concentration of capture centers is 1.4×10^{14} cm^3 (which is comparable with the concentration of shallow donors), then the lifetime with respect to capture into the neutral acceptor state is 4.8 µs. This is close to the τ values observed after quenching (Fig. 7, operation no. 3). Hole capture into the acceptor level of gold occurs at a higher rate because of the action of the attractive potential. The capture cross section is then $\sigma_p \approx 9 \times 10^{-15}$ cm^2; accordingly, $\gamma_p = 1.7 \times 10^{-7}$ cm^3/s, and the lifetime of holes at the same concentration of attracting centers is approximately 4×10^{-8} s, which is much smaller than the measured photoconductivity relaxation times. Because of the experimental limitations of the method for measuring times τ that we use, we can only determine the larger component of photoconductivity relaxation times that corresponds to the lifetime of electrons.

5. CONCLUSIONS

Our experiments were aimed at revealing the influence of high-temperature thermal treatments on the internal state of semiconducting silicon caused by changes in the structure and concentration of defects (presumably, oxygen precipitates). Changes in this internal state were studied by measuring changes in the specific resistance and photoconductivity relaxation time determined by a noncontact method without additional temperature actions on the system of defects in silicon. As a result, we obtained experimental data on the character of changes in the specific resistance and photoconductivity relaxation time depending on cyclic thermal treatments at 850, 900, 1000, and 1100°C for 5 to 50 h; five-hour heating cycles were followed by slow cooling. We also used the diffusion of gold as a probing method. Gold atoms in the electrically active state are capable of binding free electrons. This can cause a noticeable increase in the specific resistance at room temperature and intensification of recombination processes with changes in the photoconductivity relaxation time compared with the initial material. The preservation of gold atoms in the electrically active state depends on the degree of uniformity of internal crystal fields, which can be studied by slowly cooling and/or quenching samples after gold diffusion.

We found that, after gold diffusion followed by slow cooling, the samples subjected to preliminary thermal treatment contained electrically active gold in a higher concentration than the initial samples, and the longer the time (the higher the temperature) of thermal treatment, the larger this concentration (specific resistance). This allows us to suggest that the force microfields that influence gold atom displacements become more uniform after thermal treatments. Understanding changes in the internal state of crystals allowed us to prepare samples with a 2×10^4 Ω cm resistance from KEF-20 electronic silicon, whose relaxation time was 5 µs. These characteristics did not change after annealing at 900°C for 1 h followed by either slow cooling or quenching; that is, thermal stability of these parameters was attained under these conditions.

Internal electronic state transformations after preliminary thermal treatment manifest themselves in the convergence of photoconductivity relaxation times into the range 3–5 µs after quenching the samples annealed at 900°C for 1 h preceded by preliminary thermal diffusion of gold.

Changes in the internal fields of the samples subjected to preliminary thermal treatment were also observed for lateral gold diffusion both into the initial material and into the materials subjected to thermal treatments. A noticeable increase in the specific resistance at a distance of 0.5 cm from the edge of the gold layer was observed only for the material subjected to thermal treatment during 50 h and was not observed for the initial sample.

As far as we know, such systematic experiments have been conducted for the first time. They show that both preliminary thermal treatment and changes in the internal force fields of defects in crystals influence manifestations of the acceptor state of gold atoms in silicon.

ACKNOWLEDGMENTS

The authors thank Yu.A. Voronov and L.K. Orlova for providing the facilities for the chemical treatment of the samples and O.S. Sal’nik for providing the initial silicon.

This work was financially supported by the Russian Foundation for Basic Research (project no. 01-01-00507).

REFERENCES

1. J. M. Hwang and D. K. Schroder, J. Appl. Phys. **59**, 2476 (1986).
2. V. Raineri, P. G. Fallica, G. Percolla, *et al.*, J. Appl. Phys. **78**, 3727 (1995).
3. Yu. A. Bykovskiĭ, V. V. Zuev, A. D. Kiryukhin, and É. A. Manykin, Pis’ma Zh. Tekh. Fiz. **9**, 983 (1983) [Sov. Tech. Phys. Lett. **9**, 422 (1983)].
4. A. G. Nikitina and V. V. Zuev, Fiz. Tekh. Poluprovodn. (St. Petersburg) **27**, 1211 (1993) [Semiconductors **27**, 668 (1993)].
5. V. Cazcarra and P. Zunino, J. Appl. Phys. **51**, 4206 (1980).
6. R. A. Smith, *Semiconductors* (Cambridge Univ. Press, Cambridge, 1959; Inostrannaya Literatura, Moscow, 1962).
7. Yu. A. Bykovskiĭ, V. V. Zuev, A. D. Kiryukhin, *et al.*, Zh. Tekh. Fiz. **69** (4), 54 (1999) [Tech. Phys. **44**, 399 (1999)].
8. Yu. A. Bykovskiĭ, G. M. Voronkova, V. V. Grigor’ev, *et al.*, Neorg. Mater. **38**, 647 (2002).
9. W. Lerch and N. A. Stolwijk, J. Appl. Phys. **83**, 1312 (1998).
10. Yu. A. Bykovskiĭ, G. M. Voronkova, V. V. Grigor’ev, *et al.*, Fiz. Tekh. Poluprovodn. (St. Petersburg) **36**, 794 (2002) [Semiconductors **36**, 740 (2002)].
11. K. Hölzlein, G. Pensl, and M. Schulz, Appl. Phys. A **34**, 155 (1984).
12. Chyng Yan Kung, J. Appl. Phys. **65**, 4654 (1989).
13. Y. Kamiura, Y. Takeuchi, and Y. Yamashita, J. Appl. Phys. **87**, 1681 (2000).
14. A. Bourret, J. Thibault-Desseaux, and D. N. Seidman, J. Appl. Phys. **55**, 825 (1984).
15. A. Mesli, E. Courcelle, T. Zundel, *et al.*, Phys. Rev. B **36**, 8049 (1987).
16. Yu. A. Bykovskiĭ, G. M. Voronkova, V. V. Grigor’ev, *et al.*, in *Proceedings of Scientific Session of Moscow Inst. of Engineering Physics, 2001* (Moscow, 2001), Vol. 4, p. 140.
17. R. H. Wu and A. K. Peaker, Solid-State Electron. **25**, 643 (1982).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 989–995.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 1103–1111.
Original Russian Text Copyright © 2003 by Balaev, Bayukov, Vasil'ev, Velikanov, Ivanova, Kazak, Ovchinnikov, Abd-Elmeguid, Rudenko.

Magnetic and Electrical Properties of $Fe_{1.91}V_{0.09}BO_4$ Warwickite

A. D. Balaev[a], O. A. Bayukov[a], A. D. Vasil'ev[a], D. A. Velikanov[a], N. B. Ivanova[b], N. V. Kazak[a,]*, S. G. Ovchinnikov[a], M. Abd-Elmeguid[c], and V. V. Rudenko[a]

[a]*Kirensky Institute of Physics, Siberian Division, Russian Academy of Sciences, Krasnoyarsk, 660036 Russia*
[b]*Krasnoyarsk State Technical University, Krasnoyarsk, 660074 Russia*
[c]*Physikalisches Institut II, Universität zu Köln, 50937 Köln, Germany*
**e-mail: nat@iph.krasn.ru; sgo@iph.krasn.ru*

Received April 16, 2003

Abstract—We have performed a complex investigation of the structure and the magnetic and electrical properties of a warwickite single crystal with the composition $Fe_{1.91}V_{0.09}BO_4$. The results of Mössbauer measurements at $T = 300$ K indicate that there exist “localized” (Fe^{2+}, Fe^{3+}) and “delocalized” ($Fe^{2.5+}$) states distributed over two crystallographically nonequivalent positions. The results of magnetic measurements show that warwickite is a P-type ferrimagnet below $T = 130$ K. The material exhibits hopping conductivity involving strongly interacting electrons. The experimental data are analyzed in comparison to the properties of the initial (unsubstituted) Fe_2BO_4 warwickite. The entire body of data on the electric conductivity and magnetization are interpreted on a qualitative basis. © *2003 MAIK “Nauka/Interperiodica”.*

1. INTRODUCTION

Most transition metal oxyborates with the general chemical formula $M^{2+}M^{3+}BO_4$ crystallize in an orthorhombic structure of warwickite ($Mg_{1.5}Ti_{0.5}BO_4$) representing a system of linear, weakly interacting ribbons comprising two internal and one external chains in which octahedrally coordinated divalent and trivalent transition metal atoms are randomly distributed over nonequivalent crystallographic positions of two types. In recent years, these compounds have drawn the attention of researchers due to an unusual combination of properties inherent in significantly disordered, strongly correlated electron systems [1–6].

From the theoretical standpoint, by studying the Heisenberg one-dimensional chains with integer or half-integer spins, it is possible to assess the influence of disorder on the spectrum of excitations, calculate the thermodynamic parameters, and describe the properties of the ground state. Experimentally, it is possible to observe a number of interesting phenomena such as temperature-induced magnetic transitions, unusual magnetization and heat capacitance curves, and the Mott transition from the delocalized to localized state in the conductivity. In addition, the possibility of synthesizing warwickites with most transition elements opens wide possibilities for a systematic investigation into their physical properties.

To the present, some monometallic (Fe_2BO_4 [4], Mn_2BO_4) [6]) and a series of bimetallic oxoborates of the $(M,M')BO_4$ type have been synthesized, where either both M and M' are transition elements ($FeCoBO_4$ [7], $FeMnBO_4$ [8], $ScMnBO_4$ [2]) or M can be a nonmagnetic alkali earth metal such as Mg or Ca [1].

From the standpoint of magnetic order, bimetallic oxoborates with a nonmagnetic metal M represent Heisenberg one-dimensional antiferromagnetic chains with integer or half-integer spins. These compounds have been theoretically and experimentally studied to a certain extent by Continentino *et al.* [9–11]. Systems with half-integer spin were represented by $MgTiBO_4$, and those with integer spins, by $MgVBO_4$. The choice of vanadium as a transition metal is related to large dimensions of its 3*d* orbitals, which accounts for the Heisenberg nature of exchange interactions. The results of investigations showed that behavior of the magnetic susceptibility in a broad temperature interval and the magnetization curves of these systems obey a power law characteristic of strongly disordered spin chains, both with $S = 1/2$ and with $S = 1$. It was demonstrated that $MgVBO_4$ exhibits a phase transition to the spin glass state at $T = 6$ K.

In this paper, we present $Fe_{1.91}V_{0.09}BO_4$ warwickite, a new compound with two magnetic 3*d* metal ions. The magnetic and electrical properties of this compound will be described in comparison to a rather thoroughly studied Fe_2BO_4 warwickite (see, e.g., [4, 12–14]). The most interesting features of this initial (unsubstituted) compound are the phase transitions of three types: a structural transition from monoclinic to orthorhombic structure at $T = 317$ K is accompanied by an electron transition from the semiconductor–semiconductor state (related to delocalization of the charge carriers) and is

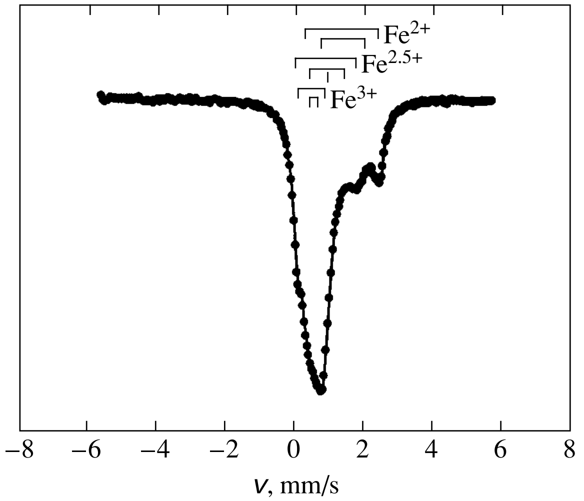


Fig. 1. The room-temperature Mössbauer spectrum of $Fe_{1.91}V_{0.09}BO_4$ warwickite.

followed by the magnetic phase transition from the paramagnetic to *P*-type ferrimagnetic state at $T = 155$ K.

2. SAMPLE PREPARATION AND EXPERIMENTAL METHODS

$FeVBO_4$ single crystals were grown using a solution melt technology in the Fe_2O_3–V_2O_3–B_2O_3–(70PbO + $30PbF_2$, wt %) system with intermediate cooling of the solution-melt from 900 to 780°C. Attfield *et al.* [4] met serious difficulties in their attempts to grow Fe_2BO_4 single crystals from a solution-melt, for which reason they used solid-phase synthesis and obtained only polycrystalline samples incorporating some other phases (removed by magnetic methods). We have succeeded in obtaining $FeVBO_4$ crystals possessing a regular shape and smooth surface. The samples had the form of needles (whiskers) with a length of up to 1 cm and a thickness of 0.10–0.15 mm.

Elemental composition of the grown samples was determined by energy-dispersive X-ray spectroscopy (EDAX ZAF quantification procedure), which showed that the relative content of iron and vanadium is 95.42 and 4.58 at. %, so that a formula unit of the substituted warwickite is $Fe_{1.91}V_{0.09}BO_4$.

X-ray diffraction was studied on a setup of the D8 ADVANCE type using CuK_α radiation (λ = 1.5406 Å). Room-temperature scans over 2θ = 13.4°–89.7° showed evidence for an orthorhombic structure (*Pnam*). The lattice parameters of our samples are presented in Table 1 in comparison to the data for Fe_2BO_4 reported in [15]. The unit cell volume of $Fe_{1.91}V_{0.09}BO_4$ is 274.82 Å^3, which virtually coincides with the value for Fe_2BO_4 (275.02 Å^3). However, in contrast to the room-temperature orthorhombic structure of $Fe_{1.91}V_{0.09}BO_4$, Fe_2BO_4 exhibits pronounced monoclinic distortions.

Table 1. Crystal lattice parameters of substituted and unsubstituted warwickites

	a, Å	*b*, Å	*c*, Å	β	*V*, Å^3
$Fe_{1.91}V_{0.09}BO_4$	3.1727	9.3831	9.2317	89.993	274.84
Fe_2BO_4	3.1688	9.3835	9.2503	90.22	275.02

The temperature and field dependences of magnetization were measured on a vibrating-sample magnetometer with a superconducting coil. The zero-field magnetization measurements were performed using a SQUID magnetometer.

The Mössbauer spectra were obtained using a $^{57}Co(Cr)$ source. The measurements were performed on single crystal powders of $Fe_{1.91}V_{0.09}BO_4$ with a linear density of 5–10 mg/cm and a natural room-temperature iron content.

The dc resistivity measurements in a range of temperatures T = 90–430 K were performed in a two-contact scheme. High resistances were measured using a teraohmmeter E6-13A capable of measuring resistances up to 10^{13} Ohm. The contacts were made of an indium-based paste and their ohmic behavior was checked by measuring current–voltage characteristics. The temperature was measured by a copper–constantan thermocouple placed immediately at a sample. The samples were heated and cooled at a rate of 1 K/min, so as to avoid parasitic temperature gradients.

3. MÖSSBAUER MEASUREMENTS

The room-temperature Mössbauer spectrum of a $Fe_{1.91}V_{0.09}BO_4$ single crystal (Fig. 1) represents a superposition of several quadrupole doublets. The resolution of spectral lines is lower than that in the spectrum of unsubstituted warwickite [4, 14], which can be related both to a decrease in the electron delocalization temperature for Fe^{2+}–Fe^{3+} and to the vanadium additive.

Figure 2 shows the probability distribution functions of the quadrupole splitting, $P(E_Q)$, for three valence states of iron (Fe^{3+}, $Fe^{2.5+}$, Fe^{2+}) in the spectrum of substituted warwickite. These functions possess a qualitative character, since the fitting parameters (isomer shifts) were the same for all distributions. As can be seen, Fe^{3+} and Fe^{2+} cations occupy nonequivalent (with respect to local environment) crystallographic positions I and II, respectively, while $Fe^{2.5+}$ cations possess local environments of three types differing by the values of quadrupole splitting. Thus, the observed spectrum of $Fe_{1.91}V_{0.09}BO_4$ should be approximated by seven quadrupole doublets. An analogous model was used [14] for interpreting the spectrum of unsubstituted Fe_2BO_4 warwickite.

The hyperfine structure parameters determined upon fitting a model curve to the experimental spectrum by the least square technique assuming the Lorentzian line shape are presented in Table 2. The isomer shifts of Fe^{2+}(I, II) and Fe^{3+}(I, II) are typical of the localized states of these cations occurring in a high-spin state with octahedral oxygen coordination. The isomer shifts of $Fe^{2.5+}$(I, IIa, IIb) are characteristic of a mixed valence of Fe^{2+} and Fe^{3+} cations, which appears as a result of the fast electron exchange between these ions. The ratio of the isomer shifts indicates that the electron density on an iron nucleus is higher in position II than in position I. At the same time, the values of quadrupole splitting show that the coordination octahedron in position II is more distorted than that in position I. A relatively high symmetry in position I allows us to assign the $Fe^{2.5+}$(I) singlet to this position, and the $Fe^{2.5+}$(IIa) and $Fe^{2.5+}$(IIb) doublets, to nonequivalent states in position II.

The distribution of Fe^{2+}, $Fe^{2.5+}$, and Fe^{3+} ions over crystallographic positions I and II, as determined from the Mössbauer measurements, allows us to write the formula of vanadium-substituted warwickite as follows:

$$(\mathrm{Fe}^{3+}_{0.383}\mathrm{Fe}^{2+}_{0.313}\mathrm{Fe}^{2.5+}_{0.299})_{\mathrm{I}}(\mathrm{Fe}^{3+}_{0.278}\mathrm{Fe}^{2+}_{0.279}\mathrm{Fe}^{2.5+}_{0.359}\mathrm{V}^{2+}_{0.09})_{\mathrm{II}}\mathrm{BO}_4.$$

The deficit of cations in position I (~0.005) according to this formula is small compared to the content of vanadium in the crystal and may be related to experimental uncertainty. Thus, the experimental data show that vanadium occupies only positions II in the given warwickite structure. Taking into account the condition of electroneutrality, the formal valence of vanadium in this compound is 2+.

A comparison of the cation distribution established in the crystal studied to that in the unsubstituted warwickite [13, 14],

$$(\mathrm{Fe}^{3+}_{0.36}\mathrm{Fe}^{2+}_{0.38}\mathrm{Fe}^{2.5+}_{0.26})_{\mathrm{I}}(\mathrm{Fe}^{3+}_{0.36}\mathrm{Fe}^{2+}_{0.38}\mathrm{Fe}^{2.5+}_{0.26})_{\mathrm{II}}\mathrm{BO}_4,$$

suggests that vanadium replaces Fe^{2+} ions in positions II, after which the total amount of iron ions subjected to fast electron exchange in the crystal increases from 0.52 (in Fe_2BO_4) to 0.66 (in $Fe_{1.91}V_{0.09}BO_4$) per formula unit. Therefore, the introduction of vanadium into the warwickite crystal changes the ratio of delocalized atoms in positions I and II. In particular, the presence of vanadium in sublattice II increases the number of delocalized atoms in this sublattice from 0.26 (in Fe_2BO_4) to 0.359 (in $Fe_{1.91}V_{0.09}BO_4$) per formula unit. This increase is significantly greater than that in sublattice I (from 0.26 to 0.299 per formula unit).

Apparently, vanadium entering into a sublattice may induce some ordering of the cations with different valences in this sublattice (similar to the ordering effect observed in manganese-containing warwickite [6]), for example, of the type V–Fe^{2+}–Fe^{3+}–V–Fe^{2+}–Fe^{3+}. Probably, the two nonequivalent states, $Fe^{2.5+}$(IIa) and $Fe^{2.5+}$(IIb), appear as a result of this process.

4. MAGNETIZATION MEASUREMENTS

Figure 3 shows the temperature dependence of magnetization, $M(T)$, in $Fe_{1.91}V_{0.09}BO_4$ warwickite measured with a SQUID magnetometer in zero magnetic field. Figure 4 presents an $M(T)$ curve measured with a vibrating-sample magnetometer for an external magnetic field of $H = 1$ kOe applied parallel to the a axis of a needle crystal. A broad maximum observed in the region of $T = 60$ K is characteristic of a P-type ferrimagnetic order, while a transition to the paramagnetic state takes place at $T = 130$ K. The $M(T)$ curve in Fig. 3 displays a feature (step) at $T = 120$ K, which was previously reported [12–14] for Fe_2BO_4. Continentino *et al.* [12] assigned this feature to the Verwey transition in an impurity phase of magnetite (Fe_3O_4), while Douvalis *et al.* [13, 14] pointed to the existence of a series of transitions in the 45–130 K range and attributed these features to different temperature dependences of mag-

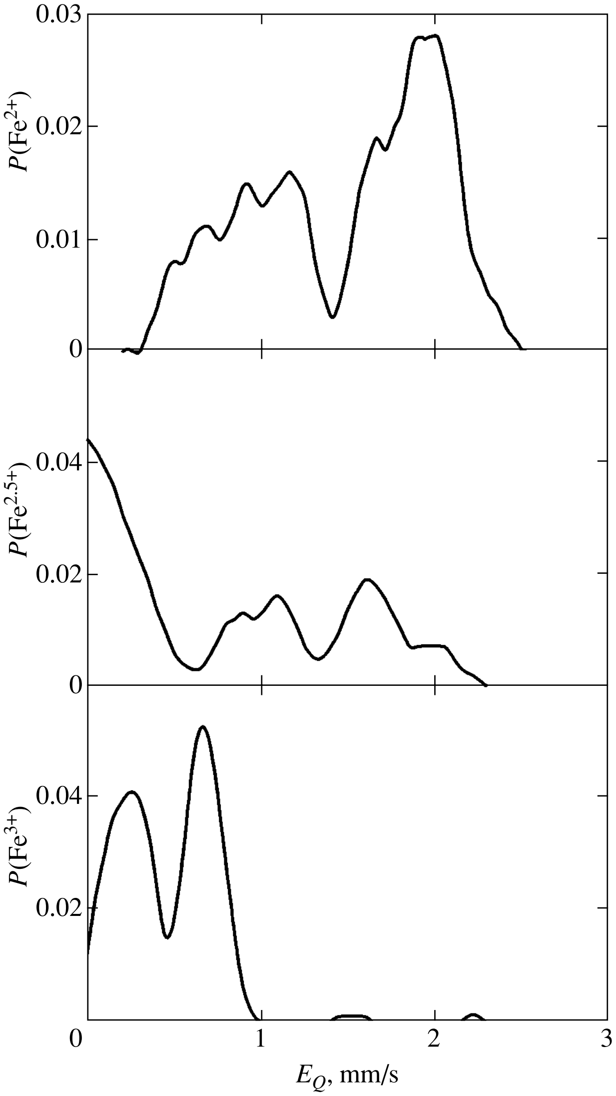


Fig. 2. The probability distribution functions of quadrupole splitting in $Fe_{1.91}V_{0.09}BO_4$ warwickite.

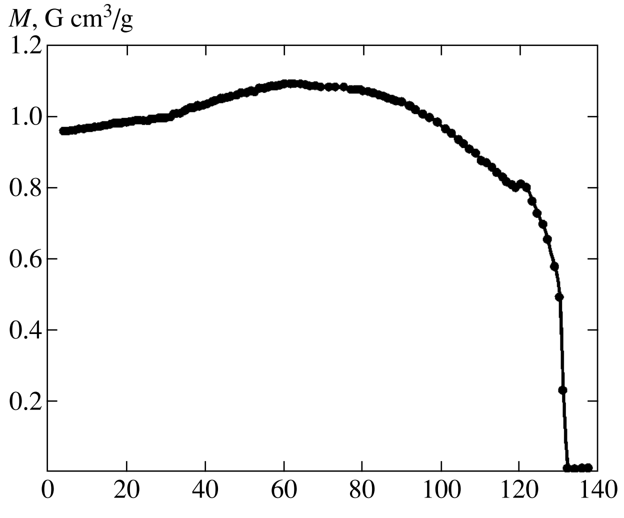


Fig. 3. The temperature dependence of magnetization for $Fe_{1.91}V_{0.09}BO_4$ in a zero field.

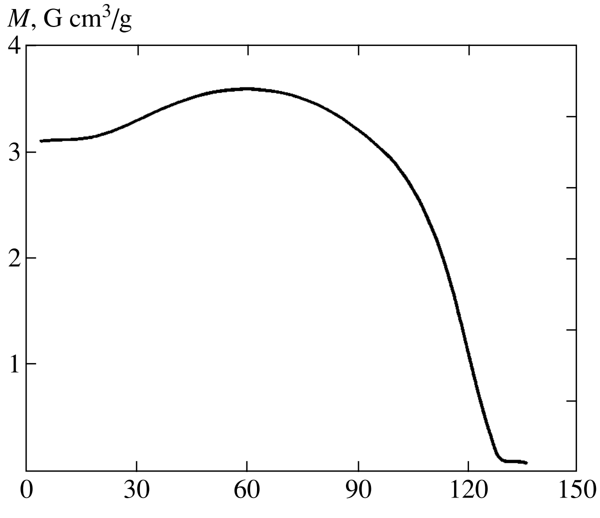


Fig. 4. The temperature dependence of magnetization for $Fe_{1.91}V_{0.09}BO_4$ in a field of $H = 1$ kOe.

netization for the two sublattices in unsubstituted Fe_2BO_4. In our experiments, neither magnetic measurements nor X-ray diffraction on high-quality single crystals showed evidence of the presence of magnetite impurity. Apparently, the feature at $T = 120$ K is inherent in $Fe_{1.91}V_{0.09}BO_4$ warwickite and probably in Fe_2BO_4 as well.

The $M(T)$ curve in Fig. 4 exhibits, generally, the same character as that reported fir Fe_2BO_4 [12]. However, the presence of vanadium ions in the structure even in small amounts decreases the magnetic transition temperature. A long-range magnetic order in $Fe_{1.91}V_{0.09}BO_4$ is established at $T = 130$ K, whereas the same transition in Fe_2BO_4 takes place at 155 K. The same trend was observed previously in a series of solid solutions of the $Fe_{1-x}V_xBO_3$ system [16].

Figure 5 presents the magnetization curves of a $Fe_{1.91}V_{0.09}BO_4$ single crystal at 4.2 and 100 K, showing the presence of an uncompensated magnetic moment of 0.1 μ_B per formula unit. This value is greater than that reported for Fe_2BO_4 (0.06 μ_B) [4]. At $H > 1$ T, the $M(H)$ plots represent linear dependences without any features, which corresponds to rotation of the magnetization vectors of sublattices toward the external magnetic field direction. As can be seen from Fig. 5, there is a nonzero magnetization in the region of small fields, which is probably related to the presence of a magnetic crystallographic anisotropy.

Table 2. Hyperfine structure parameters determined from the Mössbauer spectrum of $Fe_{1.91}V_{0.09}BO_4$ warwickite

	$\delta(\alpha_{Fe})$, mm/s	ΔE_Q, mm/s	Γ, mm/s	A, %
Fe^{3+}(I)	0.378	0.192	0.350	20
Fe^{3+}(II)	0.287	0.743	0.286	14.5
Fe^{2+}(I)	1.196	1.236	0.576	16.4
Fe^{2+}(II)	1.192	2.147	0.344	14.6
$Fe^{2.5+}$(I)	0.76	–	0.504	15.6
$Fe^{2.5+}$(IIa)	0.75	1.006	0.583	13.3
$Fe^{2.5+}$(IIb)	0.702	1.736	0.304	5.5

5. ELECTRIC RESISTANCE

On the whole, the temperature dependence of the electric resistance observed for $Fe_{1.91}V_{0.09}BO_4$ single crystals (Fig. 6) is typical of 3*d* metal borates such as Fe_2BO_4, VBO_3, and $Fe_{1-x}V_xBO_3$ [16], exhibiting a sharp increase in the resistivity and “dielectrization” of the sample at low temperatures. Features of the electron systems behavior in these materials can be revealed only in logarithmic plots of resistivity versus reciprocal temperature.

Various laws describing the behavior of $\rho(T)$ in most cases can be considered as particular cases of the general relation

$$\rho(T) = A_0 \exp(\Delta_n/T)^{1/n}, \quad n = 1, 2, 3, 4.$$

For $n = 1$, this formula describes conductivity of a simple activation type with the activation energy Δ_1. The hopping conductivity of noninteracting electrons in disordered insulators and semiconductors at low temperatures obeys the Mott law ($n = 4$). This type of conductivity was observed in solid solutions $Fe_{1-x}V_xBO_3$ and in VBO_3 single crystals below room temperature. The value of $n = 2$ is indicative of a Coulomb interaction between localized electrons in a three-dimensional

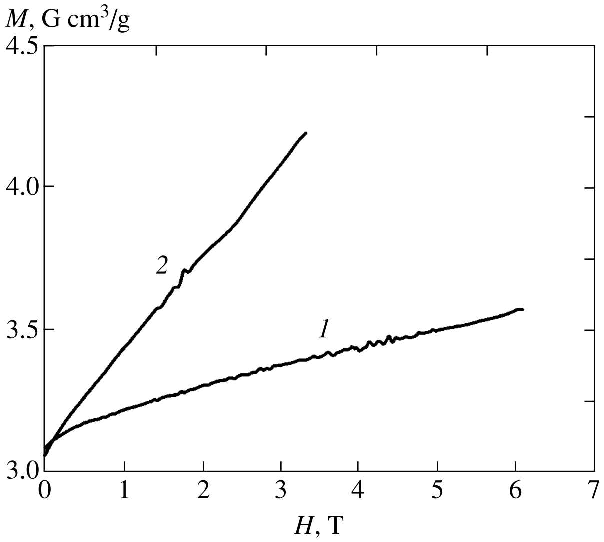


Fig. 5. The magnetization curves for $Fe_{1.91}V_{0.09}BO_4$ at T = 4.2 (*1*) and 100 K (*2*).

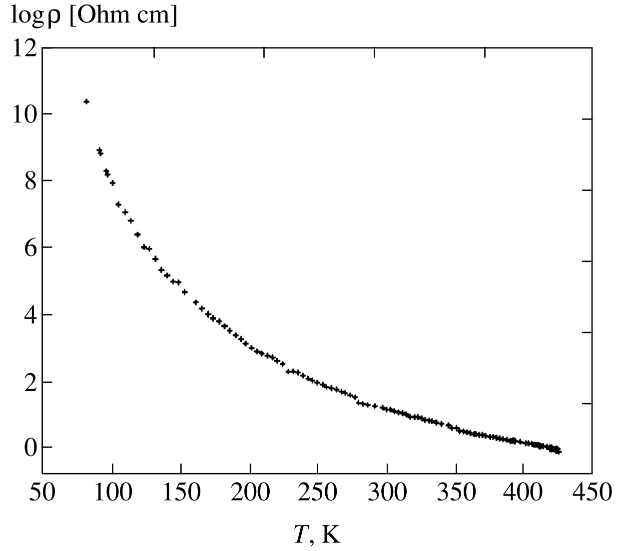


Fig. 6. The plot of $\log\rho$ versus T for $Fe_{1.91}V_{0.09}BO_4$.

(3D) system. Efros and Shklovskii [17] showed that this interaction is manifested by the density of one-electron states tending to zero near the Fermi level. In this case, the resistance varies with the temperature as $\ln R \propto (\Delta_2/T)^{1/2}$, where

$$\Delta_2 \sim e^2/\varepsilon\zeta,$$

e is the electron charge, ε is the dielectric constant, and ζ is a linear size of localization.

According to [13], the conductivity of unsubstituted Fe_2BO_4 warwickite obeys a simple activation law with an activation energy of $\Delta_n \approx 0.33$ eV. Our investigation of the electric resistance of a $Fe_{1.91}V_{0.09}BO_4$ single crystal revealed deviation from the linear temperature dependence, $\ln\rho(T) \propto T^{-1}$. This behavior of the resistivity can be explained by the existence of various competitive mechanisms of electric conductivity. Least squares processing of the experimental data leads to the following empirical relation:

$$R(T) = A_1\exp(\Delta_1/T) + A_2\exp(\Delta_2/T)^{1/2},$$

where the coefficients A_1 and A_2 weakly depend on the temperature. The first term corresponds to a conductivity component with the simple activation character, while the second term describes the hopping of strongly correlated electrons (Fig. 7). The former activation energy is $\Delta_1 = 0.15 \pm 0.01$ eV. The Coulomb interaction energy amounts to $\Delta_2 = 4.92 \pm 0.01$ eV, which is characteristic of oxides.

Theoretical investigation of the electron structure of a natural warwickite of the $MgTiBO_4$ type [18, 19] showed that a correct description of the electron properties of this system requires taking into account electron correlations at the 3d metal sites. According to the results of these calculations, the Fermi level E_F is situated inside a d band of the transition metal, so that doping of such compounds can shift the E_F value toward the bandgap.

Having studied the electrical properties of unsubstituted compound Fe_2BO_4, Attfield *et al.* [20] suggested that the charge ordering in this crystal is also related to the Coulomb repulsion. A broad, weakly pronounced semiconductor–semiconductor transition (without significant change in the activation energy) observed at T = 317 K in Fe_2BO_4 [4] is not manifested in substituted compound $Fe_{1.91}V_{0.09}BO_4$ warwickite. However, it should be noted that the measurements reported in [4] were performed on polycrystalline samples and, hence,

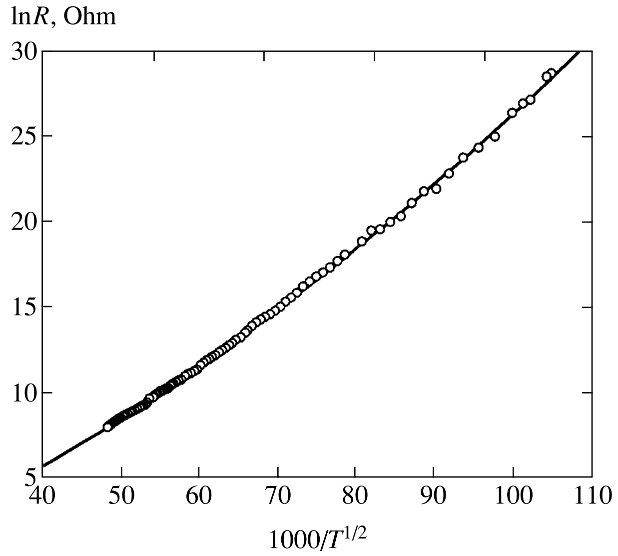


Fig. 7. The plot of $\ln R$ versus $T^{-1/2}$ for $Fe_{1.91}V_{0.09}BO_4$.

the possible influence of grain boundaries should be borne in mind.

6. DISCUSSION OF RESULTS

Apparently, the magnetic and electrical properties of the new compound $Fe_{1.91}V_{0.09}BO_4$ differ from those of Fe_2BO_4, although one cannot speak of radical changes. The available experimental data reveal important features in common for Fe_2BO_4 and the new $Fe_{1.91}V_{0.09}BO_4$ warwickite:

(i) the orthorhombic structure at high temperatures; (ii) the existence of two nonequivalent crystallographic positions; (iii) the *P*-type ferrimagnetic ordering.

According to the neutron diffraction data for unsubstituted Fe_2BO_4, the structure of this warwickite represents a system of linear, weakly interacting ribbons, each comprising two internal and one external chains (Fig. 8). The (M,M')O_6 octahedra (edge-sharing) form infinite chains parallel to the short crystallographic *a* axis. The ribbons are linked by common vertices and trigonal BO_3 groups. Cations belonging to octahedra of the same ribbon interact via two oxygen atoms occupying vertices of the common edge. In this case, the M–O–M' angle approximately equals 90°, so that an indirect 90° exchange takes place inside each ribbon. The interaction between cations belonging to the adjacent ribbons is mediated by a single oxygen atom occupying the vertex shared by octahedra.

It can be suggested that, when a part of the iron ions are replaced by vanadium in the $Fe_{1.91}V_{0.09}BO_4$ crystal, the existing 3*D* magnetic structure is retained and also represents a system of substructures having the form of ribbons. The introduction of vanadium leads to an increase in the saturation magnetic moment, which can be related to a decrease in the absolute magnetization of sublattice II.

The results of our investigation showed that the introduction of vanadium leads to modification of the electrical properties of the system. To reveal the origin of these changes, let us compare the electron structures of Fe_2BO_4 and $Fe_{1.91}V_{0.09}BO_4$. These compounds are characterized by strong electron correlations in the narrow *d* bands, forming both local magnetic moments and the dielectric ground state. The presence of strong electron correlations hinders reliable calculations of the energy band structure by traditional one-electron methods of the band theory in the local electron density functional approximation.

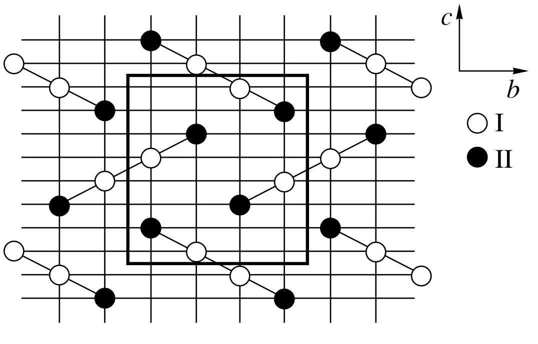


Fig. 8. A schematic diagram of the structure of $Fe_{1.91}V_{0.09}BO_4$ warwickite in a (100) plane: I and II are crystallographically nonequivalent cation positions.

For the related borate $FeBO_3$, the band structure was calculated using the local spin density functional [21] and in a generalized gradient approximation taking into account nonlocal corrections to the local density functional [22]. Unfortunately, the latter paper, devoted to pressure-induced changes in the lattice parameters and the magnetic state, presented neither the band structure proper nor the density $N(E)$ of one-electron states. For this reason, our considerations are based primarily on the results reported in [21] for the partial contributions to the density of one-electron states from various orbitals (B2*s*, B2*p*, O2*s*, O2*p*, and Fe3*d*). According to these calculations, the bottom of the empty conduction band (C) and the top of the valence band (V) are formed by the *s*-, *p*-hybrid states of B and O. A narrow *d* band of Fe is situated near the valence band top.

In the one-electron calculation, the Fermi level falls inside the *d* band and the crystal acquires metal properties. Allowance of the strong electron correlations leads to modification of the calculated local spin density functional [21], whereby the main effect consists in splitting of the *d* band into the filled lower Hubbard band (LHB) and the empty upper Hubbard band (UHB) separated by a large gap on the order of *U*, the intraatomic Coulomb matrix element. The typical value of *U* in 3*d* metals is about 5 eV.

The proposed model of the electron band structure of the unsubstituted Fe_2BO_4 crystal is depicted in Fig. 9a. Here, a dashed line above the LHB shows an acceptor impurity level accounting for the activation conductivity with the activation energy E_a. The charge carriers are represented by the conduction electrons from LHB.

Substituting vanadium for a part of iron leads to an increase in the degree of disorder in the crystal, which results in the appearance of a pseudogap with a mobility threshold ε_{c1} at the conduction band bottom and ε_{c2} at the valence band top. The Fermi level occurs in the region of localized states (Fig. 9b). As a result of the Coulomb interactions, the density of states at the Fermi level $N(\varepsilon_F)$ is zero [17]. In the case of one-electron states, the conductivity would obey the usual Mott law, $\ln\sigma(T) \propto (\Delta_4/T)^{1/4}$, characterizing thermally activated jumps of variable length. Since charge carriers at the valence band top are strongly correlated electrons, it is by no means surprising that the conductivity of $Fe_{1.91}V_{0.09}BO_4$ is closer to the Efros–Shklovskii law, $\ln\sigma(T) \propto (\Delta_2/T)^{1/2}$, where $\Delta_2 \sim e^2/\varepsilon\zeta$ [17] and ζ is the electron localization length. In our case, the localiza-

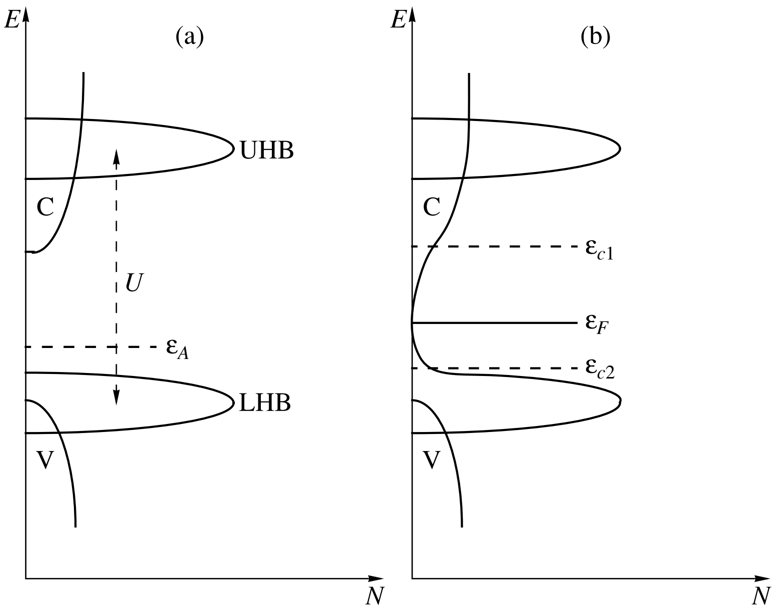


Fig. 9. A schematic energy diagram of the density of states for (a) Fe_2BO_4 and (b) $Fe_{1.91}V_{0.09}BO_4$ (see the text for explanations).

tion of 3*d* electrons is caused by the intraatomic strong electron correlations and it would be natural to assume that $\zeta \sim a_B$, where a_B is the Bohr radius, so that $\Delta_2 \sim U$.

The experimental results showed evidence for the existence of two competing mechanisms of conductivity, hopping and activation (with the activation energy Δ_1). This fact indicates that both localized electrons and band carriers are present in the system.

In conclusion, it should be noted that the $Fe_{1.91}V_{0.09}BO_4$ single crystals, in contrast to unsubstituted Fe_2BO_4, exhibit no structural transition. Adding vanadium significantly decreases the temperature of magnetic ordering, changes the type of conductivity, and alters the distribution of Fe^{2+}, $Fe^{2.5+}$, and Fe^{3+} cations over crystallographically nonequivalent positions.

ACKNOWLEDGMENTS

This study was supported in part by the Russian Foundation for Basic Research (project no. 03-02-16286) and by the Federal Program "Integration" (project no. B0017).

REFERENCES

1. J. C. Fernandes, R. B. Guimãraes, M. A. Continentino, *et al.*, Phys. Rev. B **50**, 16754 (1994).
2. R. B. Guimãraes, J. C. Fernandes, M. A. Continentino, *et al.*, Phys. Rev. B **56**, 292 (1997).
3. R. Norrestam, Z. Kristallogr. **189**, 1 (1989).
4. J. P. Attfield, A. M. T. Bell, L. M. Rodriguez-Martinez, *et al.*, J. Mater. Chem. **9**, 205 (1999).
5. M. A. Continentino, B. Boechat, R. B. Guimaraes, *et al.*, J. Magn. Magn. Mater. **226–230**, 427 (2001).
6. R. Norrestam, M. Kritikos, and A. Sjoerdin, J. Solid State Chem. **114**, 311 (1995).
7. M. J. Buerger and V. Venkatakrishnan, Mater. Res. Bull. **7**, 1201 (1972).
8. K. Bluhm and A. Utzolino, Z. Naturforsch. B **50**, 1450 (1995).
9. M. A. Continentino, J. C. Fernandes, R. B. Guimaraes, *et al.*, Philos. Mag. B **73**, 601 (1996).
10. B. Boechat, A. Saguia, and M. A. Continentino, Solid State Commun. **98**, 411 (1996).
11. A. Saguia, B. Boechat, and M. A. Continentino, J. Magn. Magn. Mater. **230**, 1300 (2001).
12. M. A. Continentino, A. M. Pedreira, R. B. Guimaraes, *et al.*, Phys. Rev. B **64**, 014406-1 (2001).
13. A. P. Douvalis, V. Papaefthymiou, A. Mookarika, *et al.*, J. Phys.: Condens. Matter **12**, 177 (2000).
14. A. P. Douvalis, V. Papaefthymiou, A. Mookarika, *et al.*, Hyperfine Interact. **126**, 319 (2000).
15. J. P. Attfield, J. F. Clarke, and D. A. Perkins, Physica B (Amsterdam) **180**, 581 (1992).
16. N. B. Ivanova, V. V. Rudenko, A. D. Balaev, *et al.*, Zh. Éksp. Teor. Fiz. **121**, 354 (2002) [JETP **94**, 299 (2002)].
17. A. L. Efros and B. I. Shklovskii, J. Phys. C **8**, L49 (1975).
18. M. Matos, R. Hoffmann, A. Latge, *et al.*, Chem. Mater. **8**, 2324 (1996).
19. D. C. Marcucci, A. Latge, E. V. Anda, *et al.*, Phys. Rev. B **56**, 3672 (1997).
20. J. P. Attfield, A. M. Bell, L. M. Rodriguez-Martinez, *et al.*, Nature **396**, 655 (1998).
21. A. V. Postnikov, St. Bartkowski, M. Neumann, *et al.*, Phys. Rev. B **50**, 14849 (1994).
22. K. Parlinski, Eur. Phys. J. B **27**, 283 (2002).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 97, No. 5, 2003, pp. 996–1004.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 5, 2003, pp. 1112–1120.
Original Russian Text Copyright © 2003 by Zavitaev, Yushkanov, Yalamov.

SOLIDS
Electronic Properties

Interaction of Electromagnetic Radiation with a Cylindrical Particle of Finite Length

É. V. Zavitaev, A. A. Yushkanov*, and Yu. I. Yalamov
Moscow State Regional University, Moscow, 107005 Russia
**e-mail: yushkanov@mtu-net.ru*

Received February 21, 2003

Abstract—The magnetic absorption cross section of a cylindrical metal particle of finite length is calculated. A general case is considered when the ratio of the transversal dimension of a particle to its length may take arbitrary values. Diffuse reflection of electrons from the internal surface of a particle is chosen as the boundary condition for the problem. Limiting cases are considered, and the results are discussed. *© 2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

When interacting with electromagnetic radiation, small metal particles exhibit properties different from those of bulky metal samples [1]. This phenomenon is attributed to the fact that the linear dimension R of a particle is comparable to or less than the mean free path Λ of electrons, $R < \Lambda$. Hence, the reflection of electrons from the boundary of a particle has a significant effect on their response to an external electromagnetic field. Therefore, when the condition $R < \Lambda$ is fulfilled, one of the basic optical characteristics—the absorption cross section—exhibits a nontrivial behavior as a function of the ratio R/Λ. At room temperature, the value of Λ for pure metals with good conductivity (aluminum, copper, silver, etc.) ranges from 10 to 100 nm. The size of experimentally investigated particles amounts to several nm; therefore, condition $R < \Lambda$ is fulfilled sufficiently often.

Equations of macroscopic electrodynamics are valid only for large particles, $R \gg \Lambda$. Therefore, the Mie theory, which describes the interaction between electromagnetic waves and metal bodies within macroscopic electrodynamics, is inapplicable to investigating the properties of small metal particles.

One can apply the standard kinetic theory of conductivity electrons in metal [2] as an apparatus capable of describing the response of electrons to an external electromagnetic field. In this case, no restrictions are imposed on the relation between the mean free path of electrons and the sample size.

In [3, 4], the authors constructed a theory for the interaction between electromagnetic radiation and a spherical particle. In the limiting case $R \ll \Lambda$ at low frequencies (in the far infrared region), this result coincides with that obtained in [5, 6]. In these studies, an approach based on the Boltzmann equation for conductivity electrons in metal is applied. An alternative approach to the problem is proposed and developed in [7, 8] and other works. Recently, there has been increasing interest in the problem of interaction between electromagnetic waves and nonspherical particles [9]. We also mention the works in which the authors attempted to take into consideration quantum-mechanical phenomena in this problem, which is especially important at low temperatures [10–12].

A number of studies [13, 14] were devoted to the description of the interaction between electromagnetic radiation and a cylindrical particle. However, in these works, only strongly oblong particles (the length of a particle is many times greater than its radius) were considered. The question of the absorption of a cylinder of arbitrary length was not raised.

The aim of the present paper is to generalize the analysis to the case when the ratio of the radius of a cylindrical particle to its length is arbitrary. We apply a kinetic method to derive a distribution function that describes a linear response of conductivity electrons to the ac magnetic field of a plane electromagnetic field. Using the distribution function obtained, we calculate the absorption cross section of the magnetic field energy of the wave.

2. STATEMENT OF THE PROBLEM

Consider a cylindrical particle of a nonmagnetic metal of radius R and length L in the field of a plane electromagnetic wave of frequency ω bounded from above by the near infrared range ($\omega < 2 \times 10^{15}$ s^{-1}). Assume that the direction of the magnetic field in the electromagnetic wave coincides with the cylinder axis. The nonhomogeneity of the external field and the skin effect are neglected (we assume that $R < \delta$, where δ is the skin depth). In the frequency range considered, the contribution of the current of dipole electric polarization to the absorption cross section is small compared with the contribution of eddy currents induced by the

1063-7761/03/9705-0996$24.00 © 2003 MAIK "Nauka/Interperiodica"

external magnetic field of the wave [3]. Therefore, we do not take into account the effect of the external electric field of the wave.

We also make the following common physical assumptions: the conductivity electrons are considered as a degenerate Fermi gas, and the response of electrons to the external ac magnetic field is taken into account by the Boltzmann equation within the relaxation-time approximation. In the boundary conditions, we assume diffuse reflection of electrons from the internal surface of the cylinder (i.e., an electron may be reflected with equal probability at an arbitrary angle ranging from 0° to 90°).

Based on the assumptions made, we can describe the absorption of electromagnetic-wave energy as follows: a uniform time-periodic magnetic field $\mathbf{H} = \mathbf{H}_0\exp(-i\omega t)$ of the wave induces an eddy electric field in the particle. This field is determined from the Maxwell equation

$$\operatorname{rot}\mathbf{E} = -\left(\frac{1}{c}\right)\frac{\partial\mathbf{H}}{\partial t} \tag{1}$$

and can be represented as

$$\mathbf{E} = \frac{1}{2c}\left[\mathbf{r}\times\frac{\partial\mathbf{H}}{\partial t}\right] = \frac{\omega}{2ic}[\mathbf{r}\times\mathbf{H}_0]\exp(-i\omega t), \tag{2}$$

where $\mathbf{r}$ is a radius vector (the origin of the coordinates is chosen on the axis of the particle). The eddy electric field acts on conductivity electrons in the particle and leads to a deviation f_1 of their distribution function f from the equilibrium Fermi distribution f_0:

$$f(\mathbf{r},\mathbf{v}) = f_0(\varepsilon) + f_1(\mathbf{r},\mathbf{v}), \quad \varepsilon = \frac{m\mathbf{v}^2}{2}, \tag{3}$$

where $\mathbf{v}$ is the electron velocity.

This gives rise to the eddy electric current

$$\mathbf{j} = e\int\mathbf{v}f\frac{2d^3(m\mathbf{v})}{h^3} = 2e\left(\frac{m}{h}\right)^3\int\mathbf{v}f_1d^3\mathbf{v} \tag{4}$$

(where h is the Planck constant) and to energy dissipation in the bulk of the particle. The energy $\bar{Q}$ dissipated per unit time is given by [15]

$$\bar{Q} = \int\overline{(\operatorname{Re}\mathbf{E})(\operatorname{Re}\mathbf{j})}d^3r = \frac{1}{2}\operatorname{Re}\int\mathbf{j}\cdot\mathbf{E}^*d^3r; \tag{5}$$

here, the bar denotes time averaging and the asterisk denotes complex conjugation.

In formula (4), we used a standard normalization for the distribution function f under which the density of electron states is equal to $2/h^3$. Below, we will use a stepwise approximation for the equilibrium function $f_0(\varepsilon)$ [16]:

$$f_0(\varepsilon) = \theta(\varepsilon_F - \varepsilon) = \begin{cases}1, & 0 \le \varepsilon \le \varepsilon_F,\\ 0, & \varepsilon_F < \varepsilon,\end{cases} \tag{6}$$

where $\varepsilon_F = mv_F^2/2$ is the Fermi energy.

The problem is reduced to finding a deviation f_1, caused by the eddy field (2), of the electron distribution function from the equilibrium distribution f_0. In a linear approximation with respect to the external field, f_1 satisfies the kinetic equation [2, 16]

$$-i\omega f_1 + \mathbf{v}\frac{\partial f_1}{\partial\mathbf{r}} + e(\mathbf{v}\cdot\mathbf{E})\frac{\partial f_0}{\partial\varepsilon} = -\frac{f_1}{\tau}, \tag{7}$$

where the collision integral is taken in the approximation of the relaxation time τ:

$$\left(\frac{df_1}{dt}\right)_s = -\frac{f_1}{\tau}. \tag{8}$$

To determine function f_1 unambiguously, we have to impose a boundary condition on the cylindrical surface of the particle. As such a condition, we take the condition if diffuse reflection of electrons from the surface [2]

$$f(\mathbf{r},\mathbf{v}) = 0 \text{ for } \begin{cases}|r_\perp| = R,\\ \mathbf{r}_\perp\cdot\mathbf{v}_\perp < 0,\end{cases} \tag{9}$$

where $\mathbf{r}_\perp$ and $\mathbf{v}_\perp$ are the projections of the electron radius vector $\mathbf{r}$ and velocity $\mathbf{v}$ onto a plane perpendicular to the cylinder axis. Solving Eq. (7) by the method of characteristics [14], we obtain

$$f_1 = A(\exp(-\nu t) - 1)/\nu, \quad t \ge 0, \tag{10}$$

where

$$\begin{aligned}\nu &= \frac{1}{\tau} - i\omega, \quad A = e(\mathbf{v}\cdot\mathbf{E})\frac{\partial f_0}{\partial\varepsilon}\\ &= \frac{e\omega}{2ic}\left(\frac{\partial f_0}{\partial\varepsilon}\right)[\mathbf{v}\times\mathbf{r}]\cdot\mathbf{H}_0\exp(-i\omega t).\end{aligned} \tag{11}$$

Here, ν and A are constant along the trajectory (characteristic). The parameter t is the time of flight of an electron along the trajectory from the boundary to point $\mathbf{r}$ at velocity $\mathbf{v}$ and is determined differently depending on the place in the particle surface where the electron is reflected. If an electron is reflected from the lateral surface, then

$$t = \frac{\mathbf{r}_\perp \cdot \mathbf{v}_\perp + [(\mathbf{r}_\perp \cdot \mathbf{v}_\perp)^2 + (R^2 - r_\perp^2)\mathbf{v}_\perp^2]^{1/2}}{\mathbf{v}_\perp^2}. \tag{12}$$

This is clear from the following geometrical considerations. Using the obvious vector equality $\mathbf{r} = \mathbf{r}_0 + \mathbf{v}t$, where $\mathbf{r}_0$ is the radius vector of an electron at the moment of reflection from the particle boundary, and projecting it onto a plane perpendicular to the cylinder axis, we obtain $\mathbf{r}_\perp = \mathbf{r}_{0\perp} + \mathbf{v}_\perp t$, where the vectors $\mathbf{r}_\perp$, $\mathbf{r}_{0\perp}$, and $\mathbf{v}_\perp$ are components of the original vectors in the projection plane. Squaring both sides of the last equality and solving the equation obtained for t, we arrive at formula (12).

Formulas (10)–(12) completely determine the solution f_1 to Eq. (7) with boundary condition (9). This fact enables us to calculate current (4) and dissipated power (5).

3. REFLECTION CONDITIONS OF ELECTRONS FROM THE LATERAL SURFACE AND END FACES

One of the most important moments in the solution of the problem posed is taking into consideration the reflection of electrons not only from the lateral surface, as is done in [13, 14], but also from the end faces of the cylinder. Let us derive these conditions.

Introduce a vector $\mathbf{b}$ by rewriting the equality $\mathbf{r}_\perp = \mathbf{r}_{0\perp} + \mathbf{v}_\perp t$ as $\mathbf{r}_{0\perp} = \mathbf{r}_\perp - \mathbf{b}$. Squaring both sides of the latter equality, we obtain a quadratic equation $b^2 - 2r_\perp b\cos\varphi - (R^2 - r_\perp^2) = 0$. Solving this equation for b, we obtain $b = r_\perp\cos\varphi + \sqrt{r_\perp^2\cos^2\varphi + R^2 - r_\perp^2}$. Here, we took into account that $|r_{0\perp}| = R$ and $\mathbf{r}_\perp \cdot \mathbf{b} = r_\perp b\cos\varphi$, where the angle φ is a cylindrical coordinate in the space of coordinates (see below).

An electron situated at a distance of z from one of the end faces is reflected from this face if its flight time $t_z = z/v_z$ along the cylinder axis is less than the flight time $t_\perp = b/v_\perp$ perpendicular to the cylinder axis, i.e.,

$$\frac{z}{v_z} < \frac{b}{v_\perp}. \tag{13}$$

Formula (13) represents one of the reflection conditions of electrons from the end face. Another condition is given by the inequality

$$\frac{-(L-z)}{v_z} < \frac{b}{v_\perp}. \tag{14}$$

Physically, this means that electrons situated at a distance of z from the chosen end but fly in the opposite direction strike the opposite end face. The “–” sign stresses the fact that the projection v_z of the electron velocity is negative in this case.

To obtain conditions for the reflection of electrons from the lateral surface, it suffices to change the signs of relations in inequalities (13) and (14).

4. ABSORPTION CROSS SECTION

To determine the absorption cross section of the energy of external magnetic field, we divide the mean dissipated power $\overline{Q}$ (see formula (5)) by the mean power flux $cH_0^2/8\pi$ in the wave:

$$\begin{aligned}\sigma &= \frac{8\pi\overline{Q}}{cH_0^2} = \frac{8\pi}{cH_0^2}\frac{1}{2}\int \mathrm{Re}(\mathbf{j}\times\mathbf{E}^*)d^3r \\ &= \frac{8\pi}{cH_0^2}\frac{1}{2}\mathrm{Re}\int\left[e\int \mathbf{v} f_1 \frac{2d^3(mv)}{h^3}\right]\mathbf{E}^* d^3r.\end{aligned} \tag{15}$$

When calculating the integral in (15), it is convenient to pass to cylindrical coordinates both in the space of coordinates ($r_\perp$, α, z; the polar axis is axis z; vector $\mathbf{H}_0$ is parallel to axis z) and in the velocity space (($v_\perp$, φ, v_z). The cylinder axis coincides with axis z. In cylindrical coordinates, field (2) has only the φ component:

$$\mathbf{E} = E_\varphi \mathbf{e}_\varphi, \quad E_\varphi = \frac{i\omega}{2c} r_\perp H_0 \exp(-i\omega t). \tag{16}$$

Accordingly, current (4) also has only the φ component (the current lines represent closed circles with centers on the axis Z in planes perpendicular to this axis).

Then, the absorption cross section is given by

$$\begin{aligned}\sigma &= \frac{1}{2}\frac{8\pi}{cH_0^2}\mathrm{Re}\int\left[e\int v_\varphi \frac{e(\mathbf{E}_\varphi\cdot\mathbf{v}_\varphi)}{\nu}\frac{\partial f_0}{\partial\varepsilon}\right. \\ &\quad \left.\times(\exp(-\nu t) - 1)\frac{2d^3(mv)}{h^3}\right]E_\varphi^* d^3r \\ &= \frac{1}{2}\frac{8\pi}{cH_0^2}2e^2\frac{m^3}{h^3}\mathrm{Re}\int[E_\varphi\int v_\varphi^2\delta(\varepsilon-\varepsilon_F) \\ &\quad \times(1-\exp(-\nu t))d^3v]E_\varphi^* d^3r \\ &= \frac{2\pi e^2 m^3\omega^2}{c^3\nu h^3}\mathrm{Re}\int[\int v_\varphi^2\delta(\varepsilon-\varepsilon_F) \\ &\quad \times(1-\exp(-\nu t))d^3v]r_\perp^2 dr_\perp.\end{aligned} \tag{17}$$

Depending on the reflection place of an electron (end face or lateral surface), the parameter t has a different form and the limits of integration with respect to z and $v_\perp$ are also different. Therefore, it is convenient to decompose the cumbersome expression (17) into two parts, σ_1 and σ_2; the first part represents the contribution of reflection from the lateral surface to the absorption, whereas the second part represents the contribution of reflection from the end faces of the cylinder.

To simplify the integration and the analysis of the results, we introduce the following notations:

$$\Gamma = R/L, \quad \xi = r_\perp/R, \qquad (18)$$
$$\eta = \xi\cos\varphi + \sqrt{1-\xi^2\sin^2\varphi},$$

$$z(x, y) = \frac{\nu R}{v_F} = \frac{R}{\tau v_F} - i\frac{R\omega}{v_F} = \frac{R}{\Lambda} - i\frac{R\omega}{v_F} = x - iy. \quad (19)$$

It is obvious that

$$b = R\eta. \quad (20)$$

Below, we give a detailed account of the calculation of each part of (17).

4.1. Contribution of the Lateral Surface

Let σ_1 be the contribution of the reflection of electrons from the lateral surface to the absorption cross section.

Using the properties of the Dirac δ function, we have

$$\delta(\varepsilon - \varepsilon_F) = \frac{2}{m}\delta(v_z^2 + v_\perp^2 - v_F^2) = \frac{2}{m}\delta[v_z^2 - (v_F^2 - v_\perp^2)]$$
$$= \frac{2}{m}\delta[(v_z - \sqrt{v_F^2 - v_\perp^2})(v_z + \sqrt{v_F^2 - v_\perp^2})]$$
$$= \frac{1}{m\sqrt{v_f^2 - v_\perp^2}}[\delta(v_z - \sqrt{v_f^2 - v_\perp^2})$$
$$+ \delta(v_z + \sqrt{v_f^2 - v_\perp^2})].$$

Due to the symmetry of the problem, we replace the integration over the entire range of velocities v_z by the integration over the positive interval and double the result obtained.

Let us rewrite the reflection condition from the lateral surface ($v_z > 0$) as $v_z < (z/b)v_\perp$ or $\sqrt{v_F^2 - v_\perp^2} < z/b$ (according to the definition of the Dirac δ function, $\delta(v_z - \sqrt{v_F^2 - v_\perp^2}) = 0$ when $v_z \neq \sqrt{v_F^2 - v_\perp^2}$). Thus,

$$\frac{b}{v_\perp}\sqrt{v_F^2 - v_\perp^2} < z. \quad (21)$$

Let us introduce the dimensionless velocity

$$\rho = v_\perp/v_F. \quad (22)$$

From (21), we obtain $b\sqrt{1-\rho^2}/\rho < z$; i.e., the reflection condition of electrons from the lateral surface is given by

$$b\sqrt{1-\rho^2}/\rho < z < L. \quad (23)$$

For the integration limits to satisfy the requirements of the problem posed, we write out an additional inequality for the lower limit $l_{\rm N}$ of integration with respect to the coordinate z:

$$l_{\rm N} = \begin{cases} b\sqrt{1-\rho^2}/\rho, & b\sqrt{1-\rho^2}/\rho < L, \\ L, & b\sqrt{1-\rho^2}/\rho > L. \end{cases} \quad (24)$$

Hence, one should impose an additional condition when integrating with respect to ρ as well:

$$\begin{cases} \rho > \kappa, & b\sqrt{1-\rho^2}/\rho < L, \\ \rho < \kappa, & b\sqrt{1-\rho^2}/\rho > L, \end{cases} \quad (25)$$

where

$$\kappa = \frac{b}{\sqrt{b^2 + L^2}} = \frac{\eta\Gamma}{\sqrt{\eta^2\Gamma^2 + 1}}. \quad (26)$$

Taking into account notations (18), (19), and (22), we obtain

$$\sigma_1 = \frac{3\pi n e^2 R^3 v_f}{mc^3}\mathrm{Re}\frac{y^2}{z(x, y)}\int_0^1 \xi^3 d\xi \int_0^{2\pi}\sin^2\varphi d\varphi$$
$$\times\int_\kappa^1 \int_{b\sqrt{1-\rho^2}/\rho}^{L}\left(1 - \exp\left(-\frac{\eta z(x, y)}{\rho}\right)\right)\frac{\rho^3}{\sqrt{1-\rho^2}}d\rho dz, \quad (27)$$

where $n = 2(m/h)^3 4\pi v_F^3/3$ is the concentration of conductivity electrons in metals. Here, we immediately integrated with respect to the angle α, using the fact that the integrand is independent of this angle, and with respect to the variable v_z, taking into account the integration rule for expressions containing the Dirac δ function.

Note that, strictly speaking, formula (27) must contain two terms; however, due to (24), the second term of the sum vanishes (since the integral with respect to z is taken over a symmetric interval).

Integrating with respect to the coordinate z, we have $\sigma_1 = \sigma_0 F_1(x, y, \Gamma)$, where

$$\sigma_0 = \frac{3\pi n e^2 R^3 v_F L}{mc^3},$$

$$F_1(x, y, \Gamma) = \mathrm{Re}\frac{y^2}{z(x, y)}\int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi d\varphi$$

$$\times \int_\kappa^1 \frac{\rho^3}{\sqrt{1-\rho^2}}\left(1 - \exp\left(-\frac{\eta z(x, y)}{\rho}\right)\right) d\rho$$

$$+ \mathrm{Re}\frac{y^2\Gamma}{z(x, y)}\int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi d\varphi \tag{28}$$

$$\times \int_\kappa^1 \eta\rho^2\left(\exp\left(-\frac{\eta z(x, y)}{\rho}\right) - 1\right) d\rho.$$

4.2. Contribution of End Faces

Denote the contribution of reflections from end faces to the absorption cross section by σ_2.

The condition of reflection of electrons from one of the end faces ($v_z > 0$) is given by

$$0 < z < b\sqrt{1-\rho^2}/\rho. \tag{29}$$

For the integration limits to satisfy the requirements of the problem posed, we write out an additional inequality for the upper limit l_V of integration with respect to z:

$$l_V = \begin{cases} b\sqrt{1-\rho^2}/\rho, & b\sqrt{1-\rho^2}/\rho < L, \\ L, & b\sqrt{1-\rho^2}/\rho > L. \end{cases} \tag{30}$$

Hence, we arrive again at the additional condition for integrating with respect to the variable ρ:

$$\begin{cases} \rho > \kappa, & b\sqrt{1-\rho^2}/\rho < L, \\ \rho < \kappa, & b\sqrt{1-\rho^2}/\rho > L. \end{cases} \tag{31}$$

Taking into account notations (18), (19), and (22), we obtain

$$\sigma_2 = \frac{3\pi n e^2 R^3 v_F}{mc^3}\left[\mathrm{Re}\frac{y^2}{z(x, y)}\int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi d\varphi\right.$$

$$\times \int_\kappa^1 \int_0^{b\sqrt{1-\rho^2}/\rho} \frac{\rho^3}{\sqrt{1-\rho^2}}\left(1 - \exp\left(-\frac{\nu z}{v_F\sqrt{1-\rho^2}}\right)\right) d\rho dz$$

$$+ \mathrm{Re}\frac{y^2}{z(x, y)}\int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi d\varphi \tag{32}$$

$$\left.\times \int_0^\kappa \int_0^L \frac{\rho^3}{\sqrt{1-\rho^2}}\left(1 - \exp\left(-\frac{\nu z}{v_F\sqrt{1-\rho^2}}\right)\right) d\rho dz\right].$$

Here, we have again integrated immediately with respect to the angle α, using the fact that the integrand is independent of this angle, and with respect to the variable v_z, taking into account the integration rule for expressions containing the Dirac δ function (recall that the flight time of an electron to one of the end faces is expressed as $t = z/v_z$).

Now, integrating with respect to the coordinate z, we have $\sigma_2 = \sigma_0 F_2(x, y, \Gamma)$, where

$$F_2(x, y, \Gamma) = \mathrm{Re}\frac{y^2}{z(x, y)}\left[\frac{\Gamma}{z(x, y)}\int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi d\varphi\right.$$

$$\left.\times \int_\kappa^1 \rho^3 \exp\left(-\frac{\eta z(x, y)}{\rho} - 1\right) d\rho\right]$$

$$+ \mathrm{Re}\frac{y^2}{z(x, y)}\left[\frac{\Gamma}{z(x, y)}\int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi d\varphi\right.$$

$$\left.\times \int_0^\kappa \rho^3\left(\exp\left(-\frac{z(x, y)}{\Gamma\sqrt{1-\rho^2}}\right) - 1\right) d\rho\right] \tag{33}$$

$$+ \mathrm{Re}\frac{y^2}{z(x, y)}\left[\Gamma\int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi d\varphi \int_\kappa^1 \eta\rho^2 d\rho\right.$$

$$\left.+ \int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi d\varphi \int_0^\kappa \frac{\rho^3}{\sqrt{1-\rho^2}} d\rho\right].$$

The dimensionless absorption cross section is given by

$$F(x, y, \Gamma) = F_1(x, y, \Gamma) + F_2(x, y, \Gamma). \tag{34}$$

The results of numerical calculations are represented in Figs. 1–5.

5. ABSORPTION IN THE LOW-FREQUENCY REGIME

Now we dwell on the case when the frequency ω of the external field and the collision rate of electrons $1/\tau$ in the bulk of metal are low compared with the collision rate of electrons with the surface of a particle. In other words, we consider the situation when $|z| \ll 1$.

In this approximation, the exponential functions entering the expressions $F_1(x, y, \Gamma)$ and $F_2(x, y, \Gamma)$ can be expanded by the well-known Taylor formula. After

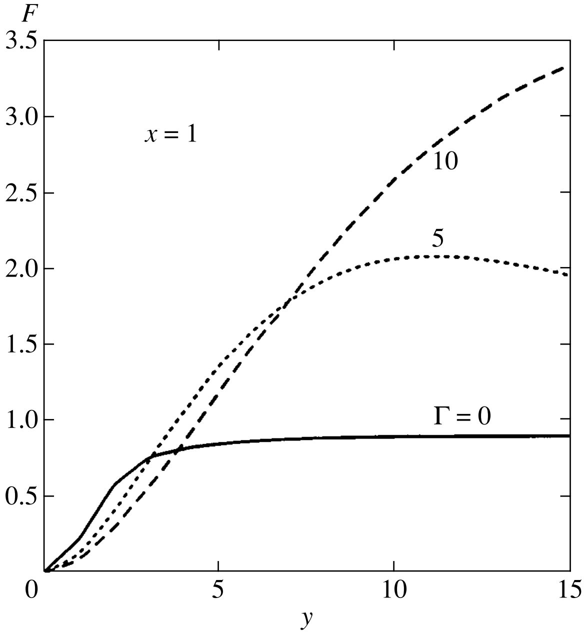


Fig. 1. Dimensionless absorption cross section F versus the dimensionless frequency y for a fixed dimensionless inverse mean free path $x = 1$ and various values of the radius-to-length ratio Γ of a particle.

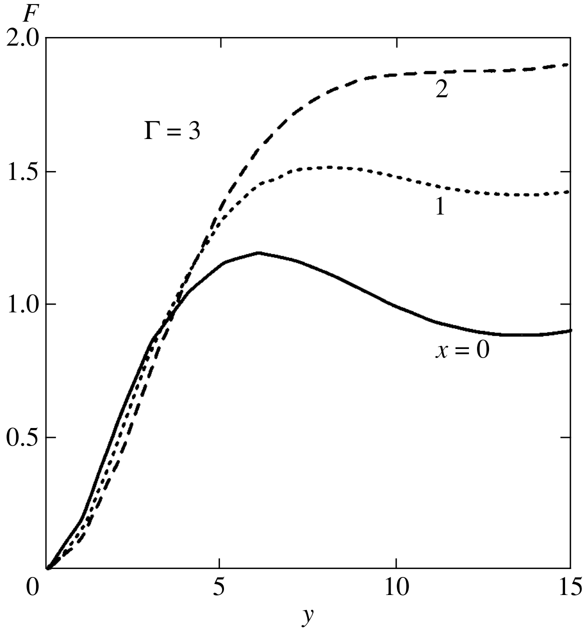


Fig. 2. Dimensionless absorption cross section F versus the dimensionless frequency y for a fixed radius-to-length ratio $\Gamma = 3$ of a particle and various values of the inverse mean free path x.

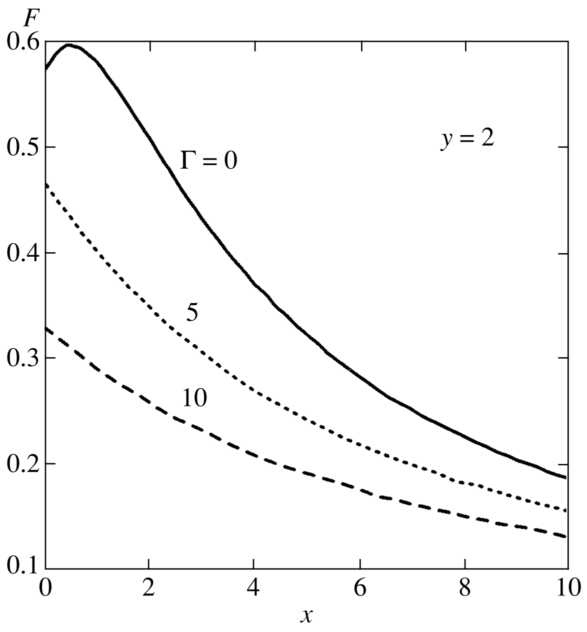


Fig. 3. Dimensionless absorption cross section F versus the dimensionless inverse mean free path x for a fixed dimensionless frequency $y = 2$ and various values of the radius-to-length ratio of a particle.

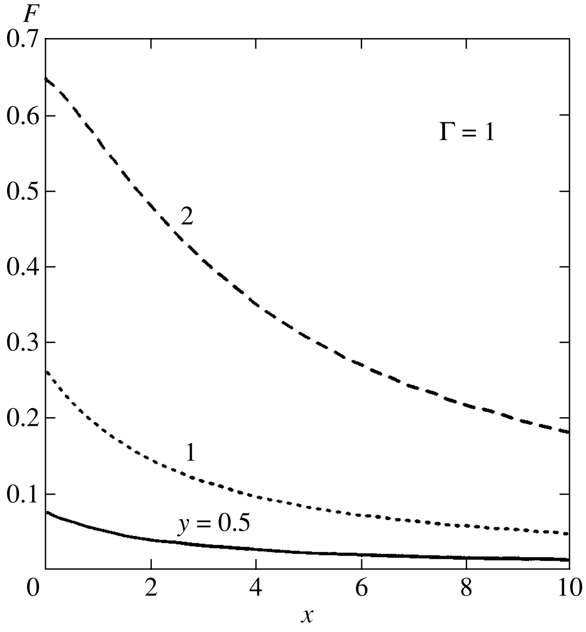


Fig. 4. Dimensionless absorption cross section F versus the dimensionless inverse mean free path x for a fixed radius-to-length ratio $\Gamma = 1$ of a particle and various values of dimensionless frequency.

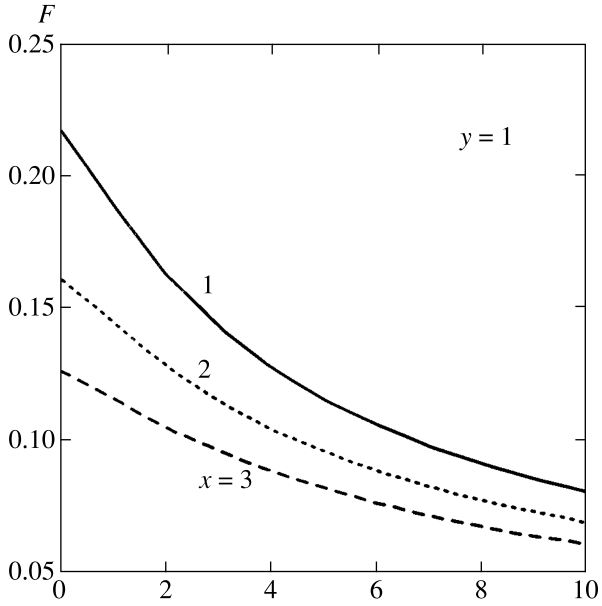


Fig. 5. Dimensionless absorption cross section F versus the radius-to-length ratio Γ of a particle for a fixed dimensionless frequency $y = 1$ and various values of the dimensionless inverse mean free path x.

substituting the expansions of the exponential functions into the integrals and performing algebraic transformations, we can see that expression $F_2(x, y, \Gamma)$ has a single term, while the integrals entering expression $F_1(x, y, \Gamma)$ are substantially simplified. Then,

$$F_1(x, y, \Gamma) = y^2\int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi d\varphi \int_\kappa^1 \frac{\eta\rho^2}{\sqrt{1-\rho^2}} d\rho - y^2\Gamma\int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi d\varphi \int_\kappa^1 \eta^2\rho d\rho, \tag{35}$$

$$F_2(x, y, \Gamma) = \frac{y^2}{2\Gamma}\int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi d\varphi \int_0^\kappa \frac{\rho^3}{1-\rho^2} d\rho. \tag{36}$$

After integrating with respect to ρ, expressions (35) and (36) are reduced to

$$F_1(x, y, \Gamma) = \frac{y^2}{2}\int_0^1 \xi^3 d\xi \int_0^{2\pi} \eta \sin^2\varphi \times \left[\frac{\pi}{2} - \arcsin\left(\frac{\eta\Gamma}{\sqrt{\eta^2\Gamma^2+1}}\right)\right] d\varphi, \tag{37}$$

$$F_2(x, y, \Gamma) = -\frac{y^2}{2\Gamma}\int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi \times \left[\ln(\sqrt{1-\kappa^2}) + \frac{\kappa^2}{2}\right] d\varphi. \tag{38}$$

The absorption cross section σ is expressed as

$$\sigma = \sigma_0 F(x, y, \Gamma) = \sigma_0 \frac{y^2}{2}\int_0^1 \xi^3 d\xi \int_0^{2\pi} \eta \sin^2\varphi d\varphi \times \left[\frac{\pi}{2} - \arcsin\left(\frac{\eta\Gamma}{\sqrt{\eta^2\Gamma^2+1}}\right)\right] d\varphi - \frac{y^2}{2\Gamma} \times \int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi \left[\ln(\sqrt{1-\kappa^2}) + \frac{\kappa^2}{2}\right] d\varphi. \tag{39}$$

This formula shows that a specific feature of the absorption cross section in the low-frequency limit is its quadratic dependence on the dimensionless frequency y of the external field (see [13]). A new result, compared with the earlier ones, is that the expression obtained takes into account the dependence of the absorption cross section on the ratio of the cylinder radius to its length (the absorption cross section depends on Γ).

Now, we consider possible limiting cases.

(a) In the case of an infinite cylinder ($\Gamma \longrightarrow 0$),

$$\sigma = \frac{2\pi^2 n e^2 R^3 v_F L}{5mc^3} y^2. \tag{40}$$

Since real particles have a finite radius R, we can determine a correction to the absorption. Indeed, when $\Gamma \ll 1$, $\kappa \approx \eta\Gamma$, and $\arcsin\kappa \approx \eta\Gamma$ (the contribution of the second integral entering in (39) is negligible due to the smallness of κ), we have

$$\sigma = \sigma_0 \frac{y^2}{2}\int_0^1 \xi^3 d\xi \int_0^{2\pi} \eta \sin^2\varphi d\varphi \left[\frac{\pi}{2} - \eta\Gamma\right] d\varphi = \frac{2}{15}\pi\sigma_0 y^2\left(1 - \frac{5}{8}\Gamma\right). \tag{41}$$

(b) Let us determine the correction to the absorption in the case of a disc, when $\Gamma \gg 1$. Here,

$$\frac{\pi}{2} - \arcsin\left(\frac{\eta\Gamma}{\sqrt{\eta^2\Gamma^2+1}}\right) = \arcsin\left(\frac{1}{\sqrt{\eta^2\Gamma^2+1}}\right) \approx \arcsin\left(\frac{1}{\eta\Gamma}\right) \approx \frac{1}{\eta\Gamma},$$

$$\sqrt{1-\kappa^2} = \frac{\kappa}{\eta\Gamma} \approx \frac{1}{\eta\Gamma}.$$

Hence, after integrating, we obtain

$$\sigma = \sigma_0 \frac{y^2}{2\Gamma}\left[\frac{\pi}{8} + \int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi \ln(\eta\Gamma) d\varphi\right]. \tag{42}$$

The dominant term in this expression is the one containing $\ln\Gamma$ ($\Gamma \gg 1$); therefore,

$$\sigma \approx \sigma_0 \frac{y^2}{8}\pi \frac{\ln\Gamma}{\Gamma}. \tag{43}$$

The logarithmic singularity in the formula for the absorption cross section (43) is directly related to the well-know logarithmic singularity of the conductivity of a thin metal layer [2].

6. ABSORPTION IN THE HIGH-FREQUENCY REGIME

When $|z| \gg 1$ and $|z|/\Gamma \gg 1$, there exists an asymptotics expression for (34). Neglecting the terms containing exponential functions because they rapidly decay and carrying out algebraic transformations, we arrive at the following expression for the dimensionless absorption cross section $F(z)$:

$$\begin{aligned} F(z) = {} & \mathrm{Re}\frac{y^2}{z(x, y)}\int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi d\varphi \\ & \times \int_0^1 \frac{\rho^3}{\sqrt{1-\rho^2}} d\rho - \mathrm{Re}\frac{y^2}{z(x, y)}\frac{\Gamma}{z(x, y)} \\ & \times \int_0^1 \xi^3 d\xi \int_0^{2\pi} \sin^2\varphi d\varphi \int_0^1 \rho^3 d\rho. \end{aligned} \tag{44}$$

This expression is easily integrated to give

$$F(z) = \mathrm{Re}\frac{y^2}{z(x, y)}\frac{\pi}{6}\left(1 - \frac{3}{8}\frac{\Gamma}{z(x, y)}\right). \tag{45}$$

As a result, we obtain the following expression for the absorption cross section:

$$\begin{aligned} \sigma(z) &= \sigma_0 \mathrm{Re}\left(\frac{y^2}{z(x, y)}\frac{\pi}{6}\left(1 - \frac{3}{8}\frac{\Gamma}{z(x, y)}\right)\right) \\ &= \sigma_0 \frac{\pi}{6} y^2 \left[\frac{x}{x^2 + y^2} - \frac{3}{8}\frac{\Gamma(x^2 - y^2)}{(x^2 - y^2)^2 + 4x^2y^2}\right]. \end{aligned} \tag{46}$$

The first term in this expression corresponds to the classical result (the Drude formula) [16], which is valid for any Γ provided that the condition $|z|/\Gamma \gg 1$ holds.

Consider the case $y \gg x$. In this case, analysis of formula (46) implies that, if $x \gg \Gamma$, then the dominant term in the absorption is the first term. Otherwise, when $\Gamma \gg x$, the second term is dominant. A similar situation arises when $x \gg y$.

In [9], the authors obtained, after a number of simplifications, an expression that describes the absorption cross section of a highly prolate ellipsoid of revolution (actually an infinite cylinder) in the high-frequency limit (for $x \ll 1$):

$$\sigma_{\mathrm{el}} = \frac{27}{128}\frac{\pi^2 e^2 n v_F R V}{2mc^3}, \tag{47}$$

where V is the volume of the ellipsoid. The authors of [9] did not take into account electron collisions in the bulk of a particle, restricting the analysis to surface collisions.

Let us determine the ratio of specific (per unit volume) absorption cross sections (46) and (47) in the high-frequency limit (the second term in (46) can be neglected). We have

$$\frac{\sigma(z)}{\sigma_{\mathrm{el}}} = 1.13. \tag{48}$$

Thus, an exact kinetic calculation yields a correction equal to 13% to the results of [9].

7. NUMERICAL ANALYSIS OF THE RESULTS

Formula (34) allows one to calculate the dimensionless absorption cross section $F(x, y, \Gamma)$ of a cylindrical particle and the absorption cross section of electromagnetic radiation,

$$\sigma = \sigma_0 F(x, y). \tag{49}$$

By varying Γ, we can obtain the absorption cross section for cylinders of various degrees of oblongness.

When $\Gamma \longrightarrow 0$ ($\kappa \longrightarrow 0$), from (34) we obtain

$$\begin{aligned} F(x, y) = {} & y^2 \mathrm{Re}\int_0^1 \xi^3 d\xi \int_0^1 \int_0^{2\pi} \frac{\rho^3}{\sqrt{1-\rho^2}} \\ & \times \frac{1}{z}\left(1 - \exp\left(-\frac{z\eta}{\rho}\right)\right)\sin^2\varphi d\rho d\varphi. \end{aligned} \tag{50}$$

This expression coincides with the result obtained in [14] for an infinite cylinder.

The dimensionless absorption cross section F is a complicated function of a combination of three dimensionless quantities x, y, and Γ.

Figure 1 and 2 represent the dimensionless absorption cross section F versus the dimensionless frequency y of the external field for a given radius and a constant radius-to-length ratio of a particle. Taking into account the fact that the particle length is finite leads to results different from those obtained in [14]. The amplitude and the period of the frequency dependence of the dimensionless absorption cross section F considerably increase. This is associated with the fact that, in the present paper, in addition to the reflection of electrons from the lateral surface of a particle, we have considered the reflection of particles from the end faces. This makes an additional contribution to the absorption of the energy of external electromagnetic field by a particle. As the particle radius increases, the oscillations of the absorption cross section are smoothed out. An increase in the radius of a particle (Fig. 1), as well as an increase in the ratio of the radius to the length of a particle (Fig. 2), leads to an increase in the absorption cross section.

Figures 3 and 4 show the dimensionless absorption cross section F versus the dimensionless inverse mean free path x of electrons. For a given frequency, the absorption is the greater, the greater the oblongness of a particle (Fig. 3). In turn, an increase in the frequency leads to an increase in the absorption cross section. As the particle radius increases further, the absorption decreases throughout the frequency range and reaches its asymptotic value. In this case, all three curves merge into a single curve. If the radius-to-length ratio of a particle is constant (Fig. 4), then the absorption decreases as the frequency of the external field decreases. As Γ increases, the absorption is somewhat less at the same frequencies.

In the present paper, the dimensionless absorption cross section F depends, in addition to x and y, on the radius-to-length ratio Γ of a particle (Fig. 5). As the particle radius increases (for constant y), the absorption cross section decreases because the dominant contribution to the absorption is made by the interaction of electrons with the end faces of a cylinder, while the contribution of the lateral surface is small. At higher frequencies y (the particle radius remaining constant), the absorption cross section is greater because the electric-field intensity increases with frequency.

REFERENCES

1. Yu. I. Petrov, *Physics of Small Particles* (Nauka, Moscow, 1984), Chap. 7.
2. J. M. Ziman, *Electrons and Phonons* (Clarendon Press, Oxford, 1960; Inostrannaya Literatura, Moscow, 1962), Chap. 11.
3. A. G. Lesskis, V. E. Pasternak, and A. A. Yushkanov, Zh. Éksp. Teor. Fiz. **83**, 310 (1982) [Sov. Phys. JETP **56**, 170 (1982)].
4. A. G. Lesskis, A. A. Yushkanov, and Yu. I. Yalamov, Poverkhnost, No. 11, 115 (1987).
5. H. J. Trodahl, Phys. Rev. B **19**, 1316 (1979).
6. H. J. Trodahl, J. Phys. C: Solid State Phys. **15**, 7245 (1982).
7. E. A. Bondar’, Opt. Spektrosk. **75**, 837 (1993) [Opt. Spectrosc. **75**, 496 (1993)].
8. E. A. Bondar’, Opt. Spektrosk. **80**, 89 (1996) [Opt. Spectrosc. **80**, 78 (1996)].
9. P. M. Tomchuk and B. P. Tomchuk, Zh. Éksp. Teor. Fiz. **112**, 661 (1997) [JETP **85**, 360 (1997)].
10. L. P. Gor’kov and G. M. Éliashberg, Zh. Éksp. Teor. Fiz. **48**, 1407 (1965) [Sov. Phys. JETP **21**, 940 (1965)].
11. É. A. Manykin, P. P. Poluéktov, and Yu. G. Rubezhnyĭ, Zh. Éksp. Teor. Fiz. **70**, 2117 (1976) [Sov. Phys. JETP **43**, 1105 (1976)].
12. A. A. Lushnikov, V. V. Maksimenko, and A. Ya. Simonov, *Dispersed Metal Films* (Akad. Nauk USSR, Kiev, 1976), p. 72.
13. É. V. Zavitaev, A. A. Yushkanov, and Yu. I. Yalamov, Zh. Tekh. Fiz. **71** (11), 114 (2001) [Tech. Phys. **46**, 1460 (2001)].
14. É. V. Zavitaev, A. A. Yushkanov, and Yu. I. Yalamov, Opt. Spektrosk. **92**, 851 (2002) [Opt. Spectrosc. **92**, 784 (2002)].
15. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 3rd ed. (Nauka, Moscow, 1992; Pergamon, New York, 1984).
16. W. A. Harrison, *Solid State Theory* (McGraw-Hill, New York, 1970; Mir, Moscow, 1972).

Translated by I. Nikitin