Journal of Experimental and Theoretical Physics, Vol. 97, No. 2, 2003, pp. 217–231.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 2, 2003, pp. 243–258.

ATOMS, SPECTRA, RADIATION

Understanding Atomic Processes in Terms of Coulomb Singularities¶

T. Surić[a,*], E. G. Drukarev[b,], and R. H. Pratt[c]**

[a]*R. Bošković Institute, Zagreb, P.O. Box 180, Zagreb, 10000 Croatia*
**e-mail: tiho@lei2.irb.hr*
[b]*Petersburg Nuclear Physics Institute, Gatchina, Leningrad oblast, 188300 Russia*
***e-mail: drukarev@thd.pnpi.spb.ru*
[c]*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15260, USA*
Submitted December 30, 2002

Abstract—Many aspects of high-energy atomic processes can be described in terms of singularities of a many-body Hamiltonian using the generalized asymptotic Fourier transform (AFT) theory. The study of matrix elements in different kinematic regimes is related to the study of singularities (points of nondifferentiability) of the wave functions and the e–γ interaction. These singularities reflect the singularities of the many-body Hamiltonian. We illustrate the principles of the AFT approach in the simple example of photoabsorption by the electron bound in a potential with a Coulomb singularity. We exhibit two general results that are important for any many-body system: (1) the quality of approximate results in different forms ("gages") depends on the quality of the description of the wave functions in the vicinity of singularities, and (2) due to the character of the Coulomb singularity, photoabsorption cross sections converge slowly to their asymptotic form as the energy increases. However, the slowly converging behavior of these cross sections is due to one common factor (the Stobbe factor), which can be obtained analytically in terms of the characterization of the vicinity of the singularity. The common Stobbe factor explains why ratios of cross sections converge more rapidly than the cross sections themselves. *© 2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

High-energy atomic processes can be described in terms of singularities of the many-body Hamiltonian. The asymptotic Fourier transform (AFT) theory [1, 2] can provide such a description. In this paper, we apply the AFT approach to photoabsorption (at high but nonrelativistic energies) in a simple atomic system, the electron in a potential with a Coulomb singularity. This serves to illustrate general points that are important in a variety of more complex systems. In particular, we illustrate (1) dependence of the required wave function quality on the form of interaction (gauge) utilized and (2) extraction of a common factor (the Stobbe factor) that contains all slow convergence of the matrix elements to their high-energy limit.

The study of single-photon ionization processes resulting in single [3, 4] or multiple [5, 6] ionization of an atom is of fundamental and practical importance. New experimental possibilities, modern synchrotron sources, and experimental methods [7] result in better understanding of the electron correlation effects in complex systems and in processes involving these systems [8–12]. Recently, we proposed a unified description [1, 2] of the processes of high-energy[1] ionization by photoabsorption, based on the mathematically well-founded AFT theory. The idea is based on the close relation between high-energy photoabsorption matrix elements and the AFT of functions with singularities (by a singularity, we mean a point where a function is not differentiable).

According to the AFT theory, the asymptotic Fourier transform of a function with singularities is determined by the behavior of the function in the vicinity of these singularities [13, 14]. Because photoabsorption at high photon energies requires at least one large outgoing electron momentum p, we can generally argue that the analysis is equivalent to the analysis of the asymptotic forms of Fourier transforms (FT). A slow asymptotic decrease for large p, such as $1/p^n$ for example, of the FT of a well-localized function, results only from singularities of that function. By studying singularity structures of the wave functions, which follow from the Schrödinger equation, one is then able to address various, quite general, issues of the matrix element (such as the interaction-form dependence of approximate matrix elements and the nature of convergence with energy to

¶This article was submitted by the authors in English.

[1] By high energy, we mean that the photon energy $\omega \gg E_B$ (where E_B is the binding energy of the state that is ionized), but still $\omega \ll m$ (for a nonrelativistic description of electrons).

high-energy forms) without needing full knowledge of the many-body wave functions. The AFT approach has been applied to double ionization [1] of He-like systems and single ionization of more complex atoms [2].

In this paper, we describe this approach to photoabsorption (and perhaps other related) processes at high but nonrelativistic energies. We use the example of single ionization within a central field independent particle approximation (IPA) model and assume that the IPA potential near the nucleus is point-Coulomb. We use this simple and familiar model in order to illustrate the main points of our general approach in the context of a relatively simple and familiar situation where most features can clearly be seen. We view the idea of the AFT approach as very general, and by presenting it in a relatively simple situation, we would like to motivate its application to other processes. The kinematic situations that occur in high-energy photoabsorption can also occur in other processes, e.g., in charged-particle scattering. In such situations, the AFT approach connects the matrix element of the process to the singularities of the system involved in the process.

Another important motivation in considering the simple system is to illustrate and emphasize two often neglected points, which are general for any photoabsorption process and for which purpose an IPA model is sufficient. The first point is concerned with how the quality of approximate results depends on both the quality of approximate wave functions used and the form of interaction chosen. For example, a plane-wave description of fast electrons is generally inadequate, as discussed in [15–18]. The second point is that due to the Coulomb singularity, any high-energy photoabsorption cross section (for ionization of a system with an interaction having a Coulomb singularity) possesses the Stobbe factor, which must be extracted in order to obtain a fast convergence of the results. This is very important for high-energy studies of photoabsorption. For example, because absolute measurements at higher energies are less accurate than at lower energies, the high-energy results are often obtained from lower energy results assuming some asymptotic behavior. It is sometimes assumed [5, 19] that at some finite energy (not taken sufficiently high), the cross sections for photoabsorption follow the leading-order Born result. This causes errors in cross sections reported.

We consider the adequacy of various forms of matrix elements (length (L), velocity (V), or acceleration (A) forms) in using approximate wave functions of various qualities in the vicinity of a singularity. We demonstrate that nonrelativistic IPA high-energy photoabsorption is determined (up to corrections $O(1/p^2) \approx O(1/\omega)$, where p is the outgoing electron momentum) by the initial state normalization and the point-Coulomb singularity. This result is form-independent, but whether the information about the singularity comes from the interaction (as in the A-form) or from the initial and final state (as in the L- and V-forms), is form-dependent. In such a way, we are able to identify necessary conditions for all the three forms to give the correct high-energy result in the IPA case. We also explicitly obtain the order of magnitude of the error resulting from the error in the description of the wave functions in the vicinity of the singularity. We consider this at two levels of accuracy (depending on the accuracy of the description of the wave functions in the vicinity of the singularity). We first consider the leading-order results in $1/p$ that can be obtained by taking a simple description of the wave functions in the vicinity of the singularity. To illustrate the source of the general Stobbe factor, we then use a description that completely includes the strong *e–N* Coulomb interaction.

We begin in Section 2 with a general discussion of the AFT of singular functions. We discuss the connection to the photoabsorption matrix element and differences (modifications of the asymptotic AFT) required due to the presence of Coulomb functions. In Section 3, we begin the discussion of the behavior of the photoeffect matrix element in an IPA potential. Here, we take the simplest description of the wave functions, which provides an illustration of the main ideas. In Section 4, we consider the simplest case, photoabsorption by an *s*-state in a purely Coulomb potential, neglecting retardation in order to make a comparison with the well-known results obtained in the Born expansion approach. We illustrate our AFT approach there in more detail. In Sections 5–7, we remove the constraints of the simplest case, identifying the resulting additional features. We show how the approach works for non-*s*-states (Section 5) and for a general IPA potential with a point-Coulomb singularity (Section 6). We discuss relativistic and retardation contributions in Section 7. Finally, in Section 8, we show how more accurate results (together with a measure of their error) can be obtained by fully including into the wave functions the interactions that are strong in the vicinity of the singularity (*e–N*). We also assess the importance of the contributions arising from interactions that are weak in the vicinity of the singularity (screening or, more generally, correlations). We discuss the convergence of the results to asymptotic forms with increasing energy. We explicitly obtain a common factor (the Stobbe factor) arising from the *e–N* interaction that contains the entire slowly converging behavior. This explains why ratios of cross sections converge to asymptotic forms much more rapidly than the cross sections themselves. In Section 9, we summarize our conclusions.

2. GENERAL CONSIDERATIONS

In general, the final-state wave function in high-energy photoionization of a many-electron atom is of the form $\exp(i\mathbf{p}\cdot\mathbf{r}_1)\Phi_{\mathbf{p},\lambda}$, where $\mathbf{p}$ denotes the large momentum of one ejected electron (there must be at least one) $\mathbf{r}_1$ is its space coordinate, and λ denotes quantum numbers of other electrons in the final state.

Because the outgoing electron wave function is described by a plane wave and incoming spherical waves at large distances (with appropriate long-range Coulomb logarithmic factors), these oscillations limit the range in $\mathbf{r}_1$ that contributes to the matrix element integral, which can be viewed as an FT in the electron momentum $\mathbf{p}$ of slowly varying functions. (Because we have assumed nonrelativistic energies, there are no oscillations of any retardation factor in the interaction in this range.) We discuss the remaining p-dependence in the Coulomb wave function in this range below. Because large ω necessarily implies large $\mathbf{p}$, the study of the photoabsorption matrix element at large energies is equivalent to the study of the asymptotic form of the FT.

The study of the asymptotic form of the FT arising in our problems is based on the theory of generalized functions [13]. By definition [14], a good function f is an infinitely differentiable function of n variables such that

$$R^l \frac{\partial f}{\partial x^{\alpha_1} \partial x^{\alpha_2} \ldots \partial x^{\alpha_n}} \longrightarrow 0, \quad R \longrightarrow \infty \tag{1}$$

for any l and m and any choice of the indices $\alpha_1, \alpha_2, \ldots, \alpha_n$ (with $\alpha_1 + \alpha_2 + \ldots + \alpha_n = m$), where $R \equiv (x_1^2 + x_2^2 + \ldots + x_n^2)^{1/2}$. (In the terminology of [14], these are called χ functions.) The theorem [13, Theorem 2, p. 15] says that the FT of a good function is a good function [13, 14]. This implies that asymptotically, the Fourier transform $g(p_1, \ldots, p_n)$ of a good function decreases faster than any power of $p \equiv (p_1^2 + p_2^2 + \ldots + p_n^2)^{1/2}$. We call this the AFT theorem. An example of such a function in three dimensions is given $f_1(\mathbf{r}) = \exp(-r^2)$. For large p, the FT $F_1(\mathbf{p}) \propto \exp(-p^2/2)$ of this function decreases exponentially, i.e., faster than any power of $1/p$, in accordance with the AFT theorem.

The functions that appear in our photoionization matrix elements, even for the photoionization of a particle in a potential, are well localized (because the bound state is localized), but are singular [20], i.e., nondifferentiable, at coalescence points. The wave functions, which are eigenstates of a many-body Hamiltonian with Coulomb interactions, have singularities at the singularities of the Hamiltonian, which are located at points where the particles coalesce. We use the term coalescence points for the locations of these singularities. In general, there are double coalescence points where two particles meet[2] or multiple coalescence points where more than two particles coincide. The properties of wave functions in the near vicinity of these singularities, which are well understood for bound states [20, 21], can be extracted from the Schrödinger equation. They are known as coalescence properties, and for s-states, they are often called Kato cusp conditions. We use the term Kato cusp conditions more generally, to denote exact behavior of the wave functions at a two-particle singularity. (There must also be singularities in the e–γ interaction operator, depending on the form taken.)

In the vicinity of a singularity, the functions whose FT is calculated can be written in terms of simpler functions f_s (with s standing for "simple") whose FT is known and a remainder O whose FT is asymptotically negligible. We call this the partitioning $(f_s + O)$ of functions. According to the theory of generalized functions, the FT of a generalized function with singularities is approximated by the FT of these simpler functions f_s, while the size of the FT of the remainder O gives a measure of the accuracy of the approximation. The point is that by taking f_s more accurately in the vicinity of the singularity, we can in principle achieve arbitrary accuracy [14, Theorem 19, p. 52]. A simple example of such a singular function is given as $f_2(\mathbf{r}) = \exp(-r)$. The FT of this function is $F_2(\mathbf{p}) \propto (1 + p^2)^{-2}$, which indeed decreases as a power of p. By partitioning f_2 in terms of polynomials[3] in r and using

$$\lim_{\varepsilon \to 0} \int \exp[-(\varepsilon r + i\mathbf{p} \cdot \mathbf{r})] r^n d^3 r = \frac{4\pi (n+1)!}{(ip)^{n+3}} \begin{cases} 0, & \text{even} \quad n \geq 0, \\ 1, & \text{odd} \quad n > 0 \end{cases} \tag{2}$$

(where we assume that $p \neq 0$ and therefore do not include δ-function terms), we reproduce the expansion of $F_2(\mathbf{p})$ in powers of $1/p$. If we were to use some other partitionings, we would not obtain powers in $1/p$, but rather some other function of p, depending on the nature of f_s used. The point is that the asymptotic FTs of such simpler functions approach the exact FT for large p and the FT of the remainder vanishes faster, in accordance with the theory of the FT of generalized functions.

This is an illustration of the general idea. The functions that appear in our matrix element can be written in terms of simpler functions, which are required by the Kato conditions to have the Coulomb behavior in the vicinity of a two-particle singularity, as will be explained below. In obtaining leading-order results, or in simple cases that we consider for illustration, we use

[2] Finite nuclear size does not affect our conclusions in any way because the distances probed at nonrelativistic energies are much, larger than the size of the nucleus. A finite nuclear size cannot be relevant for photoabsorption (when the total cross section is considered, for example) even at ultrarelativistic energies, see Section 7.

[3] Polynomial partitioning requires a convergence procedure; we multiply each term in f_s with $\exp(-\varepsilon r)$ and let $\varepsilon \longrightarrow 0$ after the integration is performed. This is consistent with the definition of the FT of generalized functions [14, p. 33].

polynomials in r (the interparticle distance) as our simpler functions. These polynomials can be viewed as expansions in r (which can exist in IPA potentials). However, in more general situations (with more electrons involved, e.g., as considered in [1]), we cannot assume that singular functions are expandable in infinite series in r in the vicinity of any singularity. In such cases, we can still partition a singular function in the vicinity of a singularity into a simple function (perhaps a polynomial or the Coulomb function) and a remainder, which may not necessarily be expandable but which vanishes faster than the simple function as r approaches the singularity. In obtaining our full IPA results in Section 8, partitioning in terms of Coulomb functions is required (because all orders in the e–N interaction are required). Such a choice is sufficient for our purposes; it gives accurate results neglecting the order $m\alpha^2/p^2$ (we use the system of units $\hbar = c = 1$) and allows us to collect all Coulomb slowly converging terms in powers of $\pi a/p$ (when full Coulomb functions are used), $a = mZ\alpha$, where Z is the nuclear charge and m is the electron mass. If better accuracy is required, one must go beyond functions with the Coulomb shape in the vicinity of a singularity.

Expansion of wave functions around the origin (which is the position of the e–N singularity) in terms of polynomials has been used previously in both single and double ionization by photoabsorption [16, 17, 22, 23] and in collisions [23]. Here, we illustrate generalizations of these approaches using the AFT theory [1, 2]. We can partition (e.g., use Coulomb functions, which are much better functions than polynomials near the singular point) around singular points that need not be at the origin in general and consider all singularities on the same footing. An important point of this approach is that it clarifies which singularities must be considered for these partitionings (there are more than one singularity in many-body wave functions and interactions in general [1]). As shown in [1], the singularities that must be considered are determined by the kinematics of outgoing electron momenta, identifying situations in which the number of the asymptotic FTs is minimized. Another important point of our approach is that we start from exact matrix elements (with exact wave functions) and extract and collect all contributions in the leading power of $1/p$, which determines the high-energy behavior (and the leading corrections in some cases). With our approach, we identify the dominant terms and avoid losing any of them; we also avoid uncontrolled introduction of spurious contributions. It is illustrated in [1] how both these problems have arisen in the use of approximate wave functions.

There is, however, a point of difference between our asymptotic matrix element and the asymptotic FT. Namely, after isolating the fast oscillating terms of the plane wave, the function left in the integrand still depends on the large momentum variable $\mathbf{p}$ (coming from the final-state wave function) via the pr-dependence (as, for example, in confluent hypergeometric functions in the Coulomb case). It might therefore be more appropriate to talk about a generalization of the FT. We see in what follows that this additional p-dependence is not a problem. The Coulomb modification of the FT results, as we demonstrate in Section 8, is a slowly converging factor, the Stobbe factor.

Accurate evaluation of the matrix element at high energies requires knowledge of both the initial and final-state electron wave functions at the singular point, or at all singular points in general if one goes beyond the IPA. However, how much of this knowledge is actually needed in a given calculation depends on the form of the matrix element used for that calculation. As we demonstrate in considering the leading-order result, only the normalization of the initial wave function at the e–N coalescence is required if the A-form is used. With other forms (V and L), we generally need further information about both the initial and final state. The exception is for the s-state ionization in the V-form, for which we need only the normalization and slope of the initial state at the coalescence. In any case, this knowledge gives us the leading contribution of the high-energy matrix element, which is generally accurate to a relative order of $O(\pi mZ\alpha/p)$.

3. SINGULARITIES OF THE IPA MATRIX ELEMENT INTEGRAND

For a single electron in a potential, in the lowest order of the electron–photon interaction, the matrix element for photoionization by photoabsorption is given by (in units $\hbar = c = 1$)

$$M = \int \Psi_{\mathbf{p}}^{(-)*}(\mathbf{r}) I(\mathbf{r}) \Psi_i(\mathbf{r}) d^3r, \tag{3}$$

where $\Psi_i(\mathbf{r}) = R_{nl}(r) Y_l^m(\hat{\mathbf{r}})$ is the initial bound state normalized to unit integrated probability density (the hat denotes the unit vector), $\Psi_{\mathbf{p}}(\mathbf{r}) = \exp(i\mathbf{p}\cdot\mathbf{r})\Phi_{\mathbf{p}}^{(-)}(\mathbf{r})$ is the final electron continuum state normalized on the momentum scale (to asymptotically approach a distorted plane wave of the momentum $\mathbf{p}$ with an amplitude of $(2\pi)^{-3/2}$), and $I(\mathbf{r})$ is the interaction operator, given in the three commonly used forms (retaining retardation to all orders) as

$$I_V = -ie^{i\mathbf{k}\cdot\mathbf{r}}\boldsymbol{\epsilon}\cdot\nabla, \tag{4}$$

$$I_L = im\left(\omega - \frac{k^2}{2m}\right)\boldsymbol{\epsilon}\cdot\mathbf{r}e^{i\mathbf{k}\cdot\mathbf{r}} - e^{i\mathbf{k}\cdot\mathbf{r}}(\boldsymbol{\epsilon}\cdot\mathbf{r})(\mathbf{k}\cdot\nabla), \tag{5}$$

$$I_A = \left(\omega - \frac{k^2}{2m}\right)^{-1} \times \left[ie^{i\mathbf{k}\cdot\mathbf{r}}(\boldsymbol{\epsilon}\cdot\nabla)V(r) - \frac{1}{m}e^{i\mathbf{k}\cdot\mathbf{r}}(\mathbf{k}\cdot\nabla)(\boldsymbol{\epsilon}\cdot\nabla)\right]. \quad (6)$$

Here, $\boldsymbol{\epsilon}$ is the photon polarization and $V(r)$ is the IPA potential energy of the type $V(r) = -(Z\alpha/r)S(r)$, where $S(r)$ is a screening function, which we assume[4] can be described as a polynomial in r for small $S(r) = (1 + s_1 r + s_2 r^2 + \ldots)$; the potential therefore has only a Coulomb divergence and is differentiable except at $\mathbf{r} = 0$ The singularity of the potential energy results in singularities in the wave functions $\Psi_\mathbf{p}$ and Ψ_i. These functions are not differentiable at the origin. The e–γ interaction operator in the L- and V-forms is regular in this sense, while in the A-form it is singular because it involves the singular potential $V(r)$, Eqs. (4)–(6). The large-p behavior of the Fourier transform of a slowly varying function of r is determined by its behavior near the coalescence point (because $pr \sim 1$, large p corresponds to small r) and only depends on the singular parts of the function. We thus partition the functions $\Psi_i(\mathbf{r})$ and $\Phi_\mathbf{p}^{(-)}(\mathbf{r})$ around the coalescence point $\mathbf{r} = 0$ (the only singular point here). The small-r behavior of these slowly varying portions of the integrand determines the AFT.

The partitioning $f_s + O$ in terms of polynomials of the initial (bound) state with quantum numbers (n, l, m) in an IPA potential with the Coulomb singularity is

$$\Psi_i(\mathbf{r}) = N_i^{IPA} r^l \times \left[1 - \frac{a}{l+1}r + \lambda_2 r^2 + \lambda_3 r^3 + O(r^4)\right] Y_l^m(\hat{\mathbf{r}}). \quad (7)$$

In the simple function f_s (in which the terms are alternately regular and singular, with the regular first term, $r^l Y_l^m(\hat{\mathbf{r}})$), the first two terms are determined solely by the Coulomb singularity of the potential and are therefore known independently of the screening, except for the overall normalization factor N_i^{IPA} (which depends on the choice of the IPA potential). Higher-order terms in f_s in Eq. (7) depend on the screening of the IPA potential, which determines the λ_i coefficients. The fact that the first two terms in the parenthesis in Eq. (7) are determined by the Coulomb singularity is well known; it is a special case of the general behavior of wave functions at coalescence points of many-electron atoms [20, 21]. Namely, in the description of a bound-state many-electron atom wave function around any coalescence (which includes any electron-electron coalescence) in terms of the relative coordinate of the two coalescing particles, the first two terms are determined by the singularity of the corresponding part of the Coulomb potential, up to an overall factor, and there is a remainder that vanishes more rapidly than linearly in the coordinate. These two terms and the normalization constant are all that we need from the initial-state wave function (we also need information from the final state) in order to determine the leading contribution in $1/p$ to high-energy photoabsorption in any form.

In the final-state electron wave function, the situation is very similar, except that the normalization is not affected by screening in the limit of high momenta. According to [24–27], the wave function of a high-energy continuum electron state of momentum $\mathbf{p}$ in the vicinity of the Coulomb singularity of the IPA potential is essentially of a Coulomb form. As shown in [27] using the analytic perturbation theory, the corrections to the Coulomb wave function due to screening in the vicinity of the nucleus ($r \ll 1/a$, where $a = mZ\alpha$ characterizes the unscreened nuclear charge) decrease with the electron momentum as $O(1/p^2)$ relative to the Coulomb functions. This means that in the vicinity of the Coulomb singularity ($r \ll 1/a$), the wave function representing the outgoing electron of momentum $p \gg a$ can be written, following [25, 27], as

$$\Psi_\mathbf{p}^{(-)}(\mathbf{r}) = N_p^C e^{i\mathbf{p}\cdot\mathbf{r}}\left[{}_1F_1\left(-i\frac{a}{p}, 1, -ipr(1+\cos\vartheta)\right) + O\left(\frac{1}{p^2}; pr, \cos\vartheta, s_i\right)\right], \quad (8)$$

where $\cos\vartheta = \hat{\mathbf{p}}\cdot\hat{\mathbf{r}}$, the first term in the right-hand side is the Coulomb term while the second term is the remainder, which vanishes faster than $1/p$ (denoted by $1/p^2$ in O). The functional dependence of O is also shown; the remainder contains all information on screening, symbolized by the coefficients s_i characterizing the small-distance behavior of the screened potential. According to the analytic perturbation theory [25, 27], an even more accurate continuum wave function of the Coulomb shape is obtained in the region $r \ll 1/a$ by shifting the electron momentum in Eq. (8) from p to p_C, by an amount determined by the parameters of the screened potential, and by replacing the normalization N_p^C (if the momentum scale normalization is used) by $\sqrt{p_c/p}\,N_{p_C}^C$. However, although such a Coulomb function is more accurate, its error still decreases as $1/p^2$ with large momentum p. We therefore do not need it here, but we use it in Section 8.

[4] Here, we assume a potential that can be expanded in integral powers of r in the vicinity of a singularity. Using a potential that is expandable in nonintegral powers of r (e.g., the Thomas–Fermi potential $V_{TF} = -Z\alpha/r + C_{TF} + O(\sqrt{r})$, where C_{TF} is a constant) would lead to nonintegral powers of $1/p$.

The result in Eq. (8) is important for our approach because as we show below, it implies that the terms in the partitioning of the final state around the coalescence that contribute to the high-energy matrix element are not affected by screening. We show this to the leading order in $1/p$, further simplifying f_s in the partitioning of Eq. (8). Because distances involved in the process are $r \sim 1/p$ and because we consider high energies for which $p \gg mZ\alpha$, while the wave functions are considered at fixed pr, the terms important for our discussion here involve terms up to linear in the parameter $mZ\alpha/p$, with further terms contributing to higher orders in $1/p$. We write

$$\Psi_{\mathbf{p}}^{(-)}(\mathbf{r}) = N_p^C e^{i\mathbf{p}\cdot\mathbf{r}}\left[1 - i\frac{a}{p}g^{(-)}(i(pr + \mathbf{p}\cdot\mathbf{r})) + O\left(\frac{1}{p^2}; pr, \cos\vartheta, s_i\right)\right], \tag{9}$$

where O includes all contributions of the order a^2/p^2 and higher order contributions from the full Coulomb function in Eq. (8) and

$$g^{(-)}(i\xi) = -\frac{1}{2\pi i}\oint_\Gamma e^{-i\xi t}\ln\left(\frac{t-1}{t}\right)\frac{dt}{t} = \int_0^1 (e^{-i\xi t} - 1)\frac{dt}{t} \tag{10}$$

determines all contributions on the order of a/p to the full Coulomb wave function for $pr \sim 1$. Here, Γ is a counterclockwise-oriented closed contour encircling the cut [0, 1]. By inserting Eqs. (9) and (7) in Eq. (3), we obtain a series of integrals of functions that contain powers of r, the $g^{(-)}$ function, and angular functions. The function $g^{(-)}(i\xi)$ is needed in calculating the leading contribution to the high-energy matrix element in general. It contains the p-dependence through pr, which may appear undesirable at first sight if we want to view this high-energy matrix element as an FT. It follows from explicit calculations, however, that there is no additional p-dependence in an FT integral also involving $g^{(-)}$ function despite the p-dependence of the $g^{(-)}$ function.

The factor $\exp(-\varepsilon r)$ is introduced in order to achieve a convergent integration of each term in the series; after the integration is performed, the limit $\varepsilon \longrightarrow 0$ is taken. As noted in Section 2, this procedure is consistent with the definition of the FT of generalized functions [13]. For the AFT theorem (and, we also assume, for the AFT involving the $g^{(-)}$ function), we must understand the singularities of the integrand. The singularity properties of the wave functions are immediately identifiable in these series, which involve powers of r and angular functions such as powers of $\hat{\mathbf{p}}\cdot\hat{\mathbf{r}} = \cos\vartheta$ and spherical harmonics. For example, r and $\cos\vartheta$ are singular at the origin (as functions of x, y, and z), but their product is not ($r\cos\vartheta = z$), nor are their squares (we also note that $r^l Y_l^m$ is regular).

We see in what follows that the leading contribution to the high-energy matrix element can be obtained in any form using only the first two terms in f_s of the respective partitioning of the initial and final state in Eqs. (7) and (9), while neglecting some of these first two terms may lead to erroneous results in some forms. Higher order terms in the expansion give higher order contributions in $1/p$, as is explained below and is demonstrated using simple examples in Sections 4 and 5. To the leading order in $1/p$, the form-independent high-energy matrix element for photoabsorption in an IPA model is obtained from

$$M = N_i^{IPA} N_p^{C*} \lim_{\varepsilon\to 0}\int \exp(-i\mathbf{p}\cdot\mathbf{r} - \varepsilon r) \times\left[1 + i\frac{a}{p}g^{(-)*}(i(pr + \mathbf{p}\cdot\mathbf{r}))\right] I(\mathbf{r}) r^l\left[1 - \frac{\alpha}{l+1}r\right] Y_l^m d^3r, \tag{11}$$

where $I(\mathbf{r})$ can take forms like Eqs. (4)–(6), which also contain different powers of r and angular functions.

Expression (11) is a form-independent term that gives the leading order in $1/p$ for large p. We see from Eq. (11) that the only difference from the purely Coulomb case is in the initial-state normalization, which depends on the IPA potential. All other terms are determined by the Coulomb singularity. Therefore, in an IPA model with a Coulomb singularity, information about screening persists at high energies only in the initial-state normalization. This behavior for high-energy photoabsorption in an IPA potential is known [26], but it is just one aspect of the persistence of the electron–electron interaction in high-energy photoabsorption, discussed within the AFT approach for two-electron atoms in [2].

In a matrix-element form, in which the interaction operator is regular (such as the V- and L-form, to be denoted by I_V^R and I_L^R) rather than singular (as in the A-form, I_A^S), the contribution from the term in the integrand involving the first terms of the simple functions f_s of both the partitionings of Ψ_i and of $\Phi^{(-)}$ vanishes for any l (while in the A-form, this term makes the leading nonvanishing contribution). The leading nonvanishing contributions in such forms (L or V) involve the product of the first term from $\Phi^{(-)}$ (which is regular, to be denoted as R_f) with the second term in Ψ_i (singular, S_i) and the product of the second term from $\Phi^{(-)}$ (singular, S_f) with the first term in Ψ_i (regular, R_i). These two contributions are of the same order in $1/p$. In summary, in

the leading order in $1/p$, we obtain the nonvanishing contribution from

$$M = \int \exp(-i\mathbf{p}\cdot\mathbf{r} - \varepsilon r)[R_f + S_f]\begin{bmatrix} I_L^R \\ I_V^R \\ I_A^S \end{bmatrix}[R_i + S_i] \longrightarrow \int \exp(-i\mathbf{p}\cdot\mathbf{r} - \varepsilon r)\begin{bmatrix} R_f I_L^R S_i + S_f I_L^R R_i \\ R_f I_V^R S_i + S_f I_V^R R_i \\ R_f I_A^S R_i \end{bmatrix}\cdot \tag{12}$$

We explicitly evaluate Eq. (11) for an H-like potential in the next two sections, and we discuss our leading-order results further, comparing them with the Born-approximation results in different forms. The two approaches must of course agree. We note that the results in Eqs. (11) and (12) reflect the importance of the singularity region. This means that if one wants to improve results, one needs to partition functions in terms of functions that better describe the behavior in the vicinity of the singularity. (This is only one of the points of distinction from a perturbative approach, e.g., the Born expansion. The Born expansion gives the same weight to all regions, while the AFT approach tells us that the singularity region is important for high-energy photoabsorption.) Partitioning of the wave functions in terms of functions that are more accurate in the vicinity of the singularity provides more accurate results. The results in Section 8, for example, are obtained using partitioning in terms of Coulomb functions (which include the *e–N* interaction to all orders).

The integrals involved in evaluating Eq. (11) are elementary and are of two types. The integrals that involve the first term in the square brackets from the final state and powers of r and products of spherical harmonics from the e–γ interaction and the initial state[5] are given by

$$J_1 = \lim_{\varepsilon\to 0}\int \exp(-i\mathbf{p}\cdot\mathbf{r} - \varepsilon r) r^n Y_L^M(\hat{\mathbf{r}}) d^3r = \frac{2\pi(n+2)!}{(ip)^{n+3}} Y_L^M(\hat{\mathbf{p}})\, f_{n+3}^L, \tag{13}$$

where

$$f_k^L = \int_{-1}^{1} \frac{P_L(x)dx}{(x - i\varepsilon)^k} \tag{14}$$

and $P_L(x)$ is the Legendre polynomial. The integrals of the second type, involving $g^{(-)}(i(pr + \mathbf{p}\cdot\mathbf{r}))$ from the final state and powers of r and products of spherical harmonics from the e–γ interaction and the initial state, are given by

$$J_2 = \lim_{\varepsilon\to 0}\int \exp(-i\mathbf{p}\cdot\mathbf{r} - \varepsilon r) g^{(-)*}(i(pr + \mathbf{p}\cdot\mathbf{r})) \times r^n Y_L^M d^3r = \frac{2\pi(n+2)!}{(ip)^{n+3}} Y_L^M(\hat{\mathbf{p}}) h_n^L, \tag{15}$$

where

$$h_n^L = \int_{-1}^{1} dx P_L(x)\left\{\sum_{j=0}^{n+1} \frac{(-1)^{n-j}(x - i\varepsilon)^{n+2-j} - 1}{(n+2-j)(x - i\varepsilon)^{n+3}} + \frac{i\pi + \ln(x - i\varepsilon)}{(x - i\varepsilon)^{n+3}}\right\}. \tag{16}$$

In performing the radial integration in Eq. (15), we used the integral form (the second form in Eq. (10)) of $g^{(-)}(i(pr + \mathbf{p}\cdot\mathbf{r}))$. The integrations over x in Eqs. (14) and (16) are elementary, and we evaluate them for specific L in Sections 4 and 5.

Expressions (13) and (15) show how higher powers in r lead to higher powers in $1/p$. We note that while J_2 gives a nonzero result for any n and L (because g contains both regular and irregular terms), J_1 is zero for n and L for which $r^n Y_L^M$ is regular, in accordance with the AFT theorem. Therefore, depending on the form used, at least one of the two first terms in the f_s functions of electron states makes a contribution to the leading order of the matrix element, while all further terms beyond the first two give higher order contributions.

4. THE SIMPLEST CASE: GROUND STATE IONIZATION OF AN H-LIKE ATOM

We now discuss how the leading contribution to the matrix element is obtained in three forms in Eqs. (4)–(6), in the simple and familiar case of photoionization of the ground state of an H-like atom with retardation neglected.

When we neglect retardation and use the nuclear Coulomb potential, the A-, V-, and L-forms of the matrix element are obtained using

$$I_A = \frac{iZ\alpha}{\omega}\frac{\boldsymbol{\epsilon}\cdot\hat{\mathbf{r}}}{r^2}, \quad I_V = -i\boldsymbol{\epsilon}\cdot\nabla, \quad I_L = im\omega\boldsymbol{\epsilon}\cdot\mathbf{r} \tag{17}$$

[5] These products of spherical harmonics can be combined into one Y_L^M.

for the interaction operators, where ω is the photon energy and $\boldsymbol{\epsilon}$ is the photon polarization. The V- and L-forms of the interaction operator are regular, but the A-form is not, because it was obtained by taking the gradient of the potential. We note that the A-form is irregular at the origin both because it is divergent and because its value near the origin depends on the direction of approach.

As we have seen, the large-p asymptotic behavior of the FT of a slowly varying function of r is determined by its behavior near the coalescence point; it only depends on the singular parts of the function in the small-r limit. We therefore begin by partitioning the functions $\Psi_i(\mathbf{r})$ and $\Phi_{\mathbf{p}}^{(-)}(\mathbf{r})$ around the coalescence point $\mathbf{r} = 0$ (the only singular point here). We write

$$\begin{aligned}\Psi_i(\mathbf{r}) &= N_i(1 - ar + \ldots),\\ \Phi_{\mathbf{p}}^{(-)}(\mathbf{r}) &= N_p^C\left(1 - i\frac{a}{p}g^{(-)}(i(pr + \mathbf{p}\cdot\mathbf{r})) + \ldots\right).\end{aligned} \tag{18}$$

If we proceed as described in the previous section, we obtain the familiar high-energy expression, which is usually obtained in the V-form by assuming that the energetic outgoing electron can be regarded free and can be represented by a plane wave [28–30]. However, in our procedure, we must be more careful and must not make such an assumption, which is incorrect in general (e.g., for non-s-states or even for s-states in the L-form).

Substituting Eqs. (18) and (17) in the matrix element in Eq. (3), we obtain a series of integrals involving powers of r and powers of $\cos\vartheta$. (We choose the $\mathbf{p}$ direction as the z axis in this integration, and only functions of $\cos\vartheta$ therefore appear). Applying $\boldsymbol{\epsilon}\cdot\nabla\Psi_i = \boldsymbol{\epsilon}\cdot\hat{\mathbf{r}}\,(-a + a^2r + \ldots)$, we obtain integrals of the two types in Eqs. (14) and (16). The integral involving the first term from f_s of the partitioning of $\Phi_p^{(-)}$ and terms from the partitioning of Ψ_i is [1]

$$\begin{aligned}&\int \exp[-(\varepsilon + i\mathbf{p}\cdot\mathbf{r})]r^n\boldsymbol{\epsilon}\cdot\hat{\mathbf{r}}\,d^3r\\ &\quad= \frac{2\pi(n+2)!\,\boldsymbol{\epsilon}\cdot\hat{\mathbf{p}}}{(ip)^{n+3}}C_n,\\ &C_n = \begin{cases}-2/(n+1), & \text{even}\quad n \geq -2,\\ 0, & \text{odd}\quad n > -1,\\ i\pi, & n = -1.\end{cases}\end{aligned} \tag{19}$$

We note that the zero result for odd $n > -1$ follows from the AFT theorem because the integrand function is not singular; for even n and for $n = -1$, the integrand function is singular. Integrals involving the term $g^{(-)}(pr, \cos\vartheta)$ in the partitioning of $\Phi_p^{(-)}$ are easily performed using the integral representation in Eq. (10). We obtain [1]

$$\begin{aligned}&\int \exp[-(\varepsilon + i\mathbf{p}\cdot\mathbf{r})]g^{(-)*}(i(pr + \mathbf{p}\cdot\mathbf{r}))r^n\boldsymbol{\epsilon}\cdot\hat{\mathbf{r}}\,d^3r\\ &\quad= \frac{2\pi\boldsymbol{\epsilon}\cdot\hat{\mathbf{p}}}{(ip)^{n+3}}D_n,\end{aligned} \tag{20}$$

where

$$D_n = 2(n+2)!(1 - i\pi)/(n+1)$$

for even $n \geq 0$ and

$$D_n = \sum_{k=0}^{(n-1)/2}\frac{2k+2}{n-2k}$$

for odd $n > 0$, $D_{-2} = i\pi - 2$, and $D_{-1} = -\pi^2/2 - i\pi - 2$. In both Eqs. (19) and (20), the p-dependence of the results is seen by inspection, resulting from the nature of the scaling in pr. Equation (19) shows that higher powers in r lead to higher powers in $1/p$, but nonvanishing contributions come only from singular terms. We use this behavior in identifying the leading contributions in $1/p$ in our calculations. In the partitioning of the final-state wave function in Eq. (18), the first term is of course regular, while the second term ($g^{(-)}$) is singular. The same is true for the initial state, not only for this s-state, but for any state with angular momentum l for which the first terms of f_s in the partitioning around $\mathbf{r} = 0$ are

$$\Psi_i \sim r^l Y_l^m(1 - ar/(l+1) + \ldots).$$

We further note that in the case of an s-state, the contributions from the first term of Ψ (which is then a constant) vanish in the V-form because of the derivative in the interaction operator $\boldsymbol{\epsilon}\cdot\nabla$. For non-$s$-states, there are nonvanishing contributions from this first term (when multiplied with g from $\Phi^{(-)}$), which must be taken into account in order to obtain the correct high-energy matrix element, which would be missing if a continuum plane-wave had been assumed (neglecting the terms in g).

Therefore, in the V-form for the s-state case, the leading contribution involves only the first term in Φ and the second term in the partitioning of the initial state Ψ, justifying the usual calculation involving the plane wave approximation for the final state. Using Eq. (19), we obtain the familiar result for s-state ionization neglecting retardation (dipole approximation),

$$M = -A\frac{\boldsymbol{\epsilon}\cdot\hat{\mathbf{p}}}{p^3}C_0 = 2A\frac{\boldsymbol{\epsilon}\cdot\hat{\mathbf{p}}}{p^3}, \tag{21}$$

where $A = 4\pi a N_i N_p^C$.

We now show that we obtain the same result using the same procedure in the L-form. Here, the singularity of the final state also contributes, however, and an incorrect result is obtained if a plane wave is assumed to provide an adequate description of the energetic electron. In terms of our approach, such an assumption would imply taking the term in the integrand involving the first term from the partitioning of Φ and the second term from the partitioning of Ψ. This contribution is

$$6A\frac{\boldsymbol{\epsilon}\cdot\hat{\mathbf{p}}}{p^3}C_2 = 4A\frac{\boldsymbol{\epsilon}\cdot\hat{\mathbf{p}}}{p^3}$$

(we have put $\omega = p^2/2m$), which is twice the correct result in Eq. (21). However, as we have already explained, we must include all terms contributing to the same power in $1/p$. We must therefore include the term in the integrand that involves the second term $g^{(-)}$ from Φ^- and the first term from Ψ_i. This gives the contribution

$$-\frac{A}{4}\frac{\boldsymbol{\epsilon}\cdot\hat{\mathbf{p}}}{p^3}D_1 = -2A\frac{\boldsymbol{\epsilon}\cdot\hat{\mathbf{p}}}{p^3}.$$

The sum of the two terms gives the correct high-energy limit, Eq. (21), showing that the L-form and the V-form indeed agree.

Finally, we can calculate the photoeffect matrix element in the A-form using the same procedures. As we have already remarked, the electron–photon interaction operator I_A, Eq. (17), is singular at the origin in this form, with a singularity arising from the singularity of the potential. The leading contribution to the matrix element in Eq. (3) in the A-form comes from the first terms in the partitioning of $\Phi^{(-)}$ and Ψ_i only (a term in the integrand that did not contribute in the L- and V-form due to its regularity, not only for the *s*-state case, but for any *l*). All other terms contribute with higher powers an $1/p$. For the *s*-state, the result in the A-form is easily evaluated with the help of Eq. (19) for $n = -2$, again giving the same result, i.e., Eq. (21). (We note that in the A-form, the next-to-leading term in $1/p$ can also be obtained without referring to screening; it involves $g^{(-)}$ or *ar*. We use this fact in Section 8 in discussing convergence toward the high-energy limit.)

5. BEYOND *s*-STATES

We now discuss non-*s*-states, staying within our simple H-like model without retardation and building on the general features already encountered in the *s*-state case. We again utilize the AFT theorem, identifying the singular part of the integrand function and evaluating the dominant contributions via the partitioning of the bound state in Eq. (7) and the partitioning of the final state in Eq. (9) around the electron–nucleus coalescence. The required integrals are given in Eqs. (13) and (14). Now, however, a plane wave does not adequately describe the fast outgoing electron, except in the A-form, in which the electron–photon interaction provides the needed singular behavior for the integrand function.

The leading contribution in $1/p$ to the high-energy photoabsorption matrix element M in Eq. (3), which in the A-form is obtained by taking the leading, regular terms in the partitioning of the initial and final electron states, is

$$M_0^{l,m} = \frac{-2imZ\alpha N_i^C l!}{(2\pi)^{3/2}(ip)^{l+3}}\frac{4\pi}{3} \times \sum_\mu Y_1^{\mu *}(\boldsymbol{\epsilon})\int d\Omega \frac{Y_1^\mu(\hat{\mathbf{r}})Y_l^m(\hat{\mathbf{r}})}{(\cos\vartheta - i\varepsilon)^{l+1}}. \quad (22)$$

The remaining angular integration involves only elementary integrals (13). For $l = 1$, the case that we discuss below in other forms for illustration, Eq. (22) gives

$$M_0^{1,m} = (-1)^{1-m}\frac{8\pi i a}{p^4}N_i N_p^C Y_1^m(\boldsymbol{\epsilon}). \quad (23)$$

(For simplicity, expression (22) is obtained with the *z* axes taken in the direction of $\hat{\mathbf{p}}$. Rotation to fixed coordinates must be made in integrating over electron angles.) We note that the part M_0 of the matrix element M is obtained in the A-form using plane waves. In other forms, the calculation of M_0 requires higher order (singular) terms from both the initial and final states in general. We therefore do not call M_0 the first (plane-wave) Born approximation result, because it is the first Born approximation only in the A-form and is a higher order Born result in other forms in general.

While a plane wave is sufficient for an initial *s*-state in the V-form, this is not true for $l > 0$, as we demonstrate. With the V-form, it is convenient to express $\boldsymbol{\epsilon}\cdot\nabla\Psi_i$ as

$$\boldsymbol{\epsilon}\cdot\nabla[R_l(r)Y_l^m(\hat{\mathbf{r}})] = \sqrt{\frac{4\pi}{3}}\sum_{\nu=-1}^{1} Y_1^{\nu *}(\boldsymbol{\epsilon}) \times\left\{\sqrt{\frac{l+1}{2l+3}}\langle 1,\nu;\, l,m|l+1,m+\nu\rangle Y_{l+1}^{m+\nu}(\hat{\mathbf{r}})\left(\frac{d}{dr}-\frac{l}{r}\right)R_l - \sqrt{\frac{l}{2l-1}}\langle 1,\nu;\, l,m|l-1,m+\nu\rangle Y_{l-1}^{m+\nu}(\hat{\mathbf{r}}) \times\left(\frac{d}{dr}+\frac{l+1}{r}\right)R_l\right\}, \quad (24)$$

where $\langle 1,\nu;\, l,m|L,M\rangle$ are the Clebsch–Gordan coefficients. For $l > 0$, the term with the lowest power in *r* in the partitioning of the function in Eq. (24) around the coalescence ($\mathbf{r} = 0$) is regular and it comes from the sec-

ond term of Eq. (24). Therefore, for the L-form, the term $g^{(-)}$ from the final-state function Φ^- also contributes to the lowest order. The exception, for the V-form, is the *s*-state, as we saw in the previous section, because the second term in Eq. (24) it is zero and the lowest power in r is singular for $l = 0$.

For illustration, we consider the initial $l = 1$ case. Inserting the expression in Eq. (24) for $l = 1$, using a linear polynomial in the partitioning around the coalescence $R_1 = N_i r[1 - (a/2)r + O(r^2)]$, including contributions from the $g^{(-)}$ term, and performing the integration by choosing the direction of **p** as the z axis, we obtain the matrix element in the V-form as

$$\begin{aligned} M_V &= iaN_iN_p^C Y_1^m(\boldsymbol{\epsilon})\int \exp(-i\mathbf{p}\cdot\mathbf{r} - \varepsilon r) \\ &\times\left[\frac{(-1)^m}{2^{|m|}\cdot 3}P_2(\cos\vartheta)r + \frac{2}{3}r\right. \\ &\left. -\frac{i}{p}g^{(-)*}(ipr(1+\cos\vartheta))\right]d^3r \\ &= (-1)^{1-m}\frac{8\pi ia}{p^4}N_iN_p^C Y_1^m(\boldsymbol{\epsilon}), \end{aligned} \tag{25}$$

which coincides with the result obtained in the A-form, Eq. (22). Assuming that a plane wave is an adequate representation of the fast electron wave function and therefore neglecting the contribution from the final state singularity, one would obtain a nonzero result (in the chosen frame, where **p** is directed along the z axes) only for the angular momentum projection $m = 0$, and even that result would be erroneous by the factor 2. The contribution to the matrix element in Eq. (25) coming from the g function part is $-8\pi iaN_iN_p^C Y_1^m(\boldsymbol{\epsilon})/p^4$.

In both the V- and L-forms, the interaction of the outgoing electron with the potential (the final-state interaction) contributes for any $l > 0$. In contrast, in the A-form, the transition operator is singular and only the leading terms (the lowest powers in r) in both the initial and final states are needed for any l to obtain the correct leading contribution.

6. SCREENED POTENTIALS

We now demonstrate that the previous results are sufficient to determine the asymptotic behavior of nonretarded photoabsorption in a general central potential to the leading order in $1/p$. The entire previous discussion, although given for a nuclear point-Coulomb potential model, is in fact valid for a general IPA potential that has a singularity of the Coulomb potential at the nucleus. Our discussion relied on the behavior of wave functions and interactions at this singularity. For the leading terms in the partitionings that we utilized, these behaviors are the same for an IPA potential as long as it is Coulomb at the singularity. To see the effects of the difference between an IPA potential (with the Coulomb singularity) and a pure Coulomb potential, we consider a potential energy of the type $V(r) = -(Z\alpha/r)S(r)$, where $S(r)$ is a screening function that behaves as $S(r) = (1 + s_1 r + s_2 r^2 + \ldots)$ for small r, as assumed in Section 3.

The interaction operators in the L- and V-forms are independent of the potential. In the A-form, we obtain

$$I_A^{IPA} = -\frac{i}{\omega}[V(r), \boldsymbol{\epsilon}\cdot\nabla] = \frac{i}{\omega}\boldsymbol{\epsilon}\cdot\nabla V(r). \tag{26}$$

The partitioning of I_A^{IPA} around the coalescence gives

$$I_A^{IPA} = \frac{iZ\alpha\boldsymbol{\epsilon}\cdot\hat{\mathbf{r}}}{\omega}\left(\frac{1}{r^2} - s_2 - 2s_3 r + \ldots\right). \tag{27}$$

We note that the term involving s_2 in Eq. (27), which is regular, contributes three more powers relative to the first term. In contrast, the first term from wave functions involving screening contribute two more powers in $1/p$.

We thus conclude that in the leading order, the same expression for photoionization at high energies is obtained in the IPA potential and in the Coulomb case (in Section 8, we show that this is in fact true in the first two orders). The normalization factors N_i^{IPA} are different, however, and IPA predictions therefore differ from the purely Coulomb-case prediction by these factors.

7. RETARDATION AND RELATIVITY

We now discuss the inclusion of retardation. This changes the forms of the interaction operators. In the V-form, the change is simple: the retarded interaction is $I_V = -i\exp(i\mathbf{k}\cdot\mathbf{r})\boldsymbol{\epsilon}\cdot\nabla$, where **k** is the photon momentum. In the L- and A-forms, obtained by applying commutator relations to the V-form, the momentum **k** also appears in factors multiplying $\exp(i\mathbf{k}\cdot\mathbf{r})$, as can be seen from Eqs. (4)–(6). In evaluating the integrals, we need to specify how to deal with the photon retardation oscillating term $\exp(i\mathbf{k}\cdot\mathbf{r})$ in the e–γ interaction $I(\mathbf{r})$. One way, particularly if retardation to a certain order in **k** is considered, is to expand $\exp(i\mathbf{k}\cdot\mathbf{r})$ in powers of $\mathbf{k}\cdot\mathbf{r}$. Another way of dealing with $\exp(i\mathbf{k}\cdot\mathbf{r})$ is to attach it to the fast oscillating term $\exp(-i\mathbf{p}\cdot\mathbf{r})$ and consider the FT in the $\Delta = \mathbf{p} - \mathbf{k}$, because $\mathbf{p} - \mathbf{k}$ is large in the nonrelativistic region whenever **p** is large, and our arguments using the AFT theorem in the asymptotic region apply. We note here that in the IPA single ionization from the ground state, retardation effects make a contribution on the relative order of $(v/c)^2$ [31, 32], where v is the velocity of the outgoing electron, which is on the same order as the relativistic contribution. This fact is used in the case of a two-electron atom [1] to argue that retardation effects make the same relative contribution in

single ionization and in double ionization in the region, where the shake-off mechanism is dominant, due to factorization of the matrix element into a (retardation-independent) correlation term and (generally, retardation-dependent) absorption term.

We however note that at relativistic energies, when $\mathbf{p} \sim \mathbf{k}$, neither expansion in k nor the assumption that Δ is large (in comparison to m) is generally valid. It is still true in certain kinematic situations, but these are not dominant for the photoabsorption processes at those energies. Namely, at relativistic energies (and as $\omega \longrightarrow \infty$), a region around the nucleus of the Compton wavelength distances continues to contribute to photoabsorption, for arbitrarily high energies. This implies that although it is a relatively small region, the final state wave function in the whole region, not just at the point of coalescence, is needed, and certain knowledge to all orders in a is required [22]. Under the analyticity assumption, the region is characterized by an expansion around the origin, and expansion of the matrix element as a series in a/Δ is still possible. The plane-wave approximation is no longer valid in any form, but partitioning in terms of Coulomb functions is fully justified.

8. CONVERGENCE OF THE CROSS SECTIONS TO THE HIGH-ENERGY LIMIT: THE STOBBE FACTOR

Here, we discuss the rate of convergence toward the exact IPA high-energy limit for the cross sections that we have obtained using the asymptotic behavior of matrix element (11). The ratio of the first correction to the leading contribution is on the order of $1/p$, and it gives a very slow convergence of the matrix elements and cross sections. In fact, it converges as $\pi a/p \sim \pi\sqrt{E_K/\omega}$, where E_K is the K-shell binding energy. We note that a slowly converging factor (i.e., converging as $\pi a/p$ relative to the asymptotic constant value) exists in the final-state normalization, which is

$$N_p^C = \frac{1}{(2\pi)^{3/2}}\Gamma\left(1 + i\frac{a}{p}\right)e^{\pi a/2p}. \tag{28}$$

However, there are also other Coulomb terms with this slow convergence (e.g., the first correction, which is unaffected by screening). In fact, if partitioning of the initial state was performed in terms of polynomials, a similar slowly converging term would come from each term of the polynomial. Because of this, and also because of the possibility of large Z, we include the e–N interaction completely in both the initial and final states, by partitioning the wave functions in the vicinity of the singularity in terms of Coulomb functions. Formally, this means that we write the initial-state wave function as

$$\Psi_i^{IPA}(\mathbf{r}) = \frac{N_i^{IPA}}{N_i^C}\Psi_i^C(\mathbf{r}) + O[(\lambda_2^C - \lambda_2)r^{l+2}], \tag{29}$$

where λ_2 (λ_2^C) is the coefficient multiplying the third term in a polynomial partitioning of the IPA (Coulomb) wave function (7), $\Psi_i^C(\mathbf{r})$ is a normalized Coulomb wave function with the same quantum numbers as $\Psi_i^{IPA}(\mathbf{r})$, and $O[(\lambda_2^C - \lambda_2)r^{l+2}]$ represents the difference between the Coulomb and screened third term in polynomial partitioning of the wave functions and all higher order differences. The terms represented by O are small, as we discuss below.

For the final state, we take the Coulomb part of Eq. (8), but with a shifted energy and with the corrected normalization [25, 27]. According to [25, 27], as already mentioned, the exact IPA wave function is Coulomb in the vicinity of the e–N singularity. A sufficiently accurate function (containing the dominant terms on the relative order of $1/p^2$) is obtained if the shifted momentum p_C is used instead of the true momentum p. The momentum p characterizes the electron at large distances from the nucleus. If we want to describe the screened wave function in the vicinity of the nucleus by a Coulomb function, we must use, according to [25, 27], the shifted momentum p_C. In addition, if the function is normalized on the momentum scale, the normalization is affected and is given by $N_p^{IPA} = \sqrt{p_C/p}\,N_{p_C}^C$. The final state is therefore given by

$$\Psi_{\mathbf{p}}^{(-)}(\mathbf{r}) = \sqrt{\frac{p_C}{p}}\Psi_{\mathbf{p}_C}^{(-)}(\mathbf{r}) + O\left(\frac{1}{p^2}\right), \tag{30}$$

where the shifted momentum p_C is [25]

$$\frac{p^2}{2m} - \frac{p_C^2}{2m} = \left|E_B^C\right| - \left|E_B^{IPA}\right|, \tag{31}$$

with E_B^C (E_B^{IPA}) denoting the hydrogen-like (IPA) binding energy of the state that is ionized.

We arrive at the following approximation for the IPA matrix element:

$$\begin{aligned} M &= \frac{N_i^{IPA}}{N_i^C}\sqrt{\frac{p_C}{p}}\int \Psi_{\mathbf{p}}^{C(-)*}(\mathbf{r})I_A(\mathbf{r})\Psi_i^C(\mathbf{r})d^3r \\ &+ O\left(\frac{1}{p^2}\right) = \frac{N_i^{IPA}}{N_i^C}\sqrt{\frac{p_C}{p}}M_l^C + O\left(\frac{1}{p^2}\right). \end{aligned} \tag{32}$$

Table 1. The total cross sections σ_K for photoabsorption from the K-shell of He ($Z = 2$) obtained using Eq. (33) in comparison with the full IPA calculations σ_{Sc}^K; $\exp(-a\pi/p)$ is the Stobbe factor

ω, keV	σ_{1s}, b Eq. (33)	σ_{Sc}^K, b Ref. [33]	$\exp(-\frac{a\pi}{p})$
1	396	402	0.542
1.5	107	109	0.596
2	41.7	43.2	0.633
3	10.9	11.2	0.682
4	4.20	4.23	0.715
5	1.99	2.03	0.739
6	1.08	1.10	0.757
8	0.408	0.411	0.784

From Eq. (32), it immediately follows that at high energies,

$$d\sigma^{IPA} = \left(\frac{N_i^{IPA}}{N_i^C}\right)^2 d\sigma^C + O\left(\frac{1}{p^2}\right), \tag{33}$$

where $d\sigma^C$ is the differential cross section obtained from Coulomb H-like wave functions calculated at the shifted momentum p_C and O indicates how rapidly the error decreases. (The momentum p from the phase space cancels p from the factor p_C/p, leaving only the shifted momentum p_C in the right-hand side of Eq. (33).) The error in Eq. (33) is determined by the errors in the wave functions. According to the results in [25], the difference between screened and Coulomb functions is very small when unnormalized functions (with the same first coefficient in the expansion taken) are compared; for potentials with a polynomial expansion, this difference decreases as $1/p^2$ for small r.[6] This implies that dominant terms on the relative order of $1/p^2$ are collected. We illustrate the fast convergence of this procedure for the cases involving low-Z atoms (He) and outer shells of higher Z atoms (the L-state of Ne).

For Coulomb states, the integrals in Eq. (32) can be evaluated analytically. As illustrative examples, we have the cross section per electron for the electron ionization from the 1s, 2s, and 2p shells [15]

$$\sigma_{1s}^C = \frac{32\pi a^5 \alpha p}{3m^3\omega^3(p^2+\alpha^2)} 2\pi \frac{a}{p}\left[\exp\frac{a}{p}\left(\pi - 4\arctan\frac{p}{a}\right)\right] \times \left[\exp\left(\frac{\pi a}{p}\right) - \exp\left(-\frac{\pi a}{p}\right)\right]^{-1}, \tag{34}$$

$$\sigma_{2s}^C = \frac{4\pi a^5 \alpha p(p^2+a^2)}{3m^3\omega^3[p^2+(a/2)^2]^2} 2\pi \frac{a}{p} \times \exp\left[\frac{a}{p}\left(\pi - 4\arctan\frac{2p}{a}\right)\right] \times \left[\exp\left(\frac{\pi a}{p}\right) - \exp\left(-\frac{\pi a}{p}\right)\right]^{-1}, \tag{35}$$

$$\sigma_{2p}^C = \frac{\pi a^7 \alpha p[p^2+(11/12)a^2]}{3m^3\omega^3[p^2+(a/2)^2]^3} 2\pi \frac{a}{p} \times \exp\left[\frac{a}{p}\left(\pi - 4\arctan\frac{2p}{a}\right)\right] \times \left[\exp\left(\frac{\pi a}{p}\right) - \exp\left(-\frac{\pi a}{p}\right)\right]^{-1}. \tag{36}$$

To illustrate the meaning of Eq. (33), we apply it to the calculation of high-energy photoabsorption and compare the result with those obtained within the full Fock-Slater IPA calculations. Our comparison with relativistic calculations is fully justified for low-Z atoms and for s-shells of higher Z atoms, for which retardation and relativistic contributions cancel to a high degree even at higher energies. However, for our illustrative purposes, we also show p-state results for Ne for relatively small photon energies.

In Table 1, we show the total cross section for K-shell ionization obtained from Eq. (33) for $Z = 2$ using $(N_i/N_i^C)^2 = 0.7358$ (which indicates large screening) [33] and compare it with the results of the full IPA calculations from [33]. As we see from Table 1, the agreement between the high-energy result in Eq. (33) and the full IPA calculations is already very good at 1 keV, despite the large screening; in the energy range 1–2 keV, the disagreement is around 1–2%. This is to be compared with the lowest order result, which gives a disagreement of about 50% in the same energy range, as indicated in Table 1 by the value of the factor $\exp(-a\pi/p)$ (see the explanation for this factor below).

For the L-state of Ne, screening is even larger $(N_i/N_i^C)^2 = 0.4386$ for the 2s state and $(N_i/N_i^C)^2 = 0.2277$ for 2p). In Table 2, we show the total cross section for the 2s and 2p states of Ne and compare them

[6] We note that the use of a potential that cannot be expanded in integral powers of r might not give a small correction that vanishes as $1/p^2$, as given in Eq. (33). For example, the Thomas–Fermi potential (see footnote 4) leads to a correction that vanishes slower, i.e., as $1/p^{3/2}$. However, the Thomas–Fermi model fails in the vicinity of the nucleus (which is the region determining high-energy photoabsorption), where it predicts too large an electron density, see, e.g., B.G. Englert and J. Schwinger, Phys. Rev. A **29**, 2331 (1984).

Table 2. The total cross sections σ_{2s} and σ_{2p} for photoabsorption from the respective subshells 2*s* and 2*p* of Ne ($Z = 10$) obtained using Eq. (33) in comparison with the full IPA calculations σ_{Sc}^{2s} and σ_{Sc}^{2p}; $\exp(-a\pi/p)$ is the Stobbe factor

ω, keV	σ_{2s}, b Eq. (33)	σ_{Sc}^{2s}, b Ref. [33]	σ_{2p}, b Eq. (33)	σ_{Sc}^{2p}, b Ref. [33]	$\exp(-\frac{a\pi}{p})$
1	11276	10600	5629	5416	0.039
2	1932	1895	492	495	0.090
4	289	290	37.4	38.9	0.174
8	38.5	39.1	2.51	2.65	0.286
10	19.7	20.0	1.030	1.092	0.325
15	5.67	5.77	0.1987	0.2144	0.398
20	2.309	2.353	0.0607	0.0669	0.449
30	0.637	0.651	0.0112	0.0128	0.519
50	0.122	0.125	0.00129	0.00159	0.600

with the full relativistic IPA calculations. For the 2*s* state, Eq. (33) gives results that converge to the full IPA result very quickly; the disagreement is around 6% at 1 keV and is less than 1% at 4 keV. Similarly, the results in Eq. (33) for the 2*p* state converge rapidly to IPA results in the same energy range. This very good agreement between the results in Eq. (33) and the full IPA results already at relatively small energies, even for Ne, can be explained by the properties of IPA wave functions hi the vicinity of the Coulomb *e*–*N* singularity. Namely, at the photon energy region 1–2 keV, the distances involved (distances around the singularity at which the momentum is transferred between the electron and the nucleus) are within the *K*-shell orbit for Ne, and well within the *K*-shell orbit for He, where the screening is small. Therefore, the shapes of the wave functions at these distances are basically Coulomb. This is a very important point that we use and generalize in our approach. The high-energy photoabsorption is essentially of the Coulomb type. This means that the high-energy behavior of cross sections (we here mean the keV range, as in our examples) is determined by the properties of functions near the singularity, which is of the Coulomb type. The screening effects enter these IPA examples, of course, but in a simple way as a constant factor. By straightforward generalization of these findings in high-energy many-body calculations, we can significantly simplify calculations involving *e*–*e* correlation, as shown in [1].

Another important point that we want to make in this section, relevant for more complex systems [1, 34], is the relatively fast convergence of the ratios of photoabsorption cross sections to the results predicted by lowest order results (the Born approximation results in the A-form). We first note that the slowest converging factor in our examples in Eqs. (34)–(36) is $\exp(-\pi a/p)$. In partitioning wave functions around the coalescence, we obtain this factor by collecting all Coulomb interaction in the final state for each term in the partitioning of the initial state. The factor is therefore present for any state. The existence of a common slowly converging factor provides fast converging ratios of the cross sections. Further, the ratios of the cross sections for ionization from subshells of the same shell converge particularly fast, as we illustrate using our examples for the L-shell, Eqs. (35) and (36). In our examples, the ratio

$$\frac{\sigma_{2s}}{\sigma_{2p}} \sim \omega + \frac{a^2}{12m} + O\left(\frac{a^2}{\omega}\right) \tag{37}$$

is a nearly linear function of the photon energy ω in the keV range. If we had used the lowest-order result in $1/p$, we would obtain $\sigma_{2s}/\sigma_{2p} \sim \omega$, which is very similar to the exact result (in the keV range for Ne, for example), although the first-order results for cross sections differ by an order of magnitude from the exact results in this energy range, as indicated roughly by the factor $\exp(-\pi a/p)$ in Table 2.

9. CONCLUSIONS

We have illustrated the AFT nonrelativistic approach to atomic processes by studying high-energy photoionization (with incident photon energies $\omega \ll m$) of an electron bound in a central potential. We have demonstrated that in this case, high-energy ionization by photoabsorption can be understood in terms of the singularities of the Hamiltonian, which also illustrates more general situations. Our discussion did not depend on the choice of the form [length (L), velocity (V), acceleration (A), etc.] of the photoionization matrix element.

Because photoabsorption at high photon energies requires at least one large outgoing electron momen-

tum, we have argued that the analysis is equivalent to the analysis of the asymptotic form of the FT. Based on the FT theory, we have shown that a slow asymptotic decrease of the photoabsorption matrix element for a large momentum p (such as $1/p^n$) is related to singularities of the e–N potentials. We have demonstrated how this large-momentum behavior can be obtained from the behavior of wave functions and interactions around singularities. With this approach, we can identify the dominant terms and avoid omitting any of them.

We have applied our approach to study the high-energy total cross section for ionization in a central potential with the Coulomb e–N singularity. We have demonstrated that the approach and the final results are form- (gage-) independent. However, the dependence of the final results on the quality of the initial and final state wave functions in the vicinity of a singularity varies with form (gage). We have found that the acceleration form, which places the singularities of the Hamiltonian in the e–γ interaction, has the smallest requirement on the quality of wave functions at the singularity, in the situations considered. We have shown that in the A-form, the leading contribution to the photoabsorption matrix element is the lowest order Born result. In the L- and V-forms, it is generally a higher order Born result, with the exception of the V-form in the ground state ionization, where it is also the lowest order Born result. This means that in general (except in the A-form), the fast electron cannot be represented by a plane wave, even in the high-energy limit. For this leading contribution to the matrix element, the A-form requires only the proper normalization of the initial state at the e–N singularity. In contrast, the L- and V-forms require knowledge of both the normalization and slope of the wave functions at the singularities.

We have discussed slow convergence of the cross sections to the high-energy limit, considering the ionization of an electron in a screened potential. We have demonstrated that by collecting all Coulomb terms in the vicinity of the e–N singularity, we also collect the dominant terms up to the relative order $1/p^2$ and provide fast convergence of the cross sections. Although the neglected terms in the matrix element are still of the relative order $1/p^2$, they are negligible. Thus, we have demonstrated that the high-energy behavior of cross sections (in the keV range, as in our examples) is determined by the properties of functions near the singularity, which is of the Coulomb type. The screening effects enter through normalization factors in the IPA cases. We have also demonstrated that the only slowly converging factor (the Stobbe factor $\exp(-\pi a/p)$, which converges as $1/p$, while all other terms converge faster) is common for ionization from all states. The existence of a common slowly converging factor provides fast converging ratios of the cross sections.

This paper was supported by the Croatian Ministry of Science (grant nos. 00980205 and 0098012) and by the National Science Foundation (grant no. PHY0201595). Two authors (T.S. and E.G.D.) are grateful for the hospitality of the Department of Physics and Astronomy at the University of Pittsburgh during their visits. We also wish to acknowledge the helpful discussions with K. Pisk, D. Kekez, J. Burgdörfer, and M. Dondera.

REFERENCES

1. T. Surić, E. G. Drukarev, and R. H. Pratt, Phys. Rev. A **67**, 022709 (2003).
2. T. Surić, E. G. Drukarev, and R. H. Pratt, Phys. Rev. A **67**, 022710 (2003).
3. E. W. B. Dias, H. S. Chakraborty, P. C. Deshmukh, *et al.*, Phys. Rev. Lett. **78**, 4553 (1997).
4. D. L. Hansen, O. Hemmers, H. Wang, *et al.*, Phys. Rev. A **60**, R2641 (1999).
5. J. A. R. Samson, W. C. Stolte, Z.-X. He, *et al.*, Phys. Rev. A **57**, 1906 (1998).
6. R. Wehlitz, M.-T. Huang, B. D. DePaola, *et al.*, Phys. Rev. Lett. **81**, 1813 (1998).
7. R. Dörner, V. Mergel, O. Jagutzki, *et al.*, Phys. Rep. **330**, 95 (2000).
8. Y. Qiu, Y. Z. Teng, J. Burgdörfer, and Y. Wang, Phys. Rev. A **57**, R1489 (1998).
9. A. S. Kheifets and I. Bray, Phys. Rev. A **58**, 4501 (1998).
10. H. W. van der Hart and C. H. Greene, Phys. Rev. Lett. **81**, 4333 (1998).
11. M. Ya. Amusia, N. B. Avdonina, E. G. Drukarev, *et al.*, Phys. Rev. Lett. **85**, 4703 (2000).
12. E. G. Drukarev, Phys. Rev. A **51**, R2684 (1995).
13. M. J. Lighthill, *Introduction to Fourier Analysis and Generalized Functions*, 2nd ed. (Cambridge Univ. Press, Cambridge, 1970).
14. L. Schwartz, *Théorie des Distributions* (Hermann, Paris, 1950–1951); E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, 2nd ed. (Clarendon Press, Oxford, 1948; Gostekhizdat, Moscow, 1948).
15. H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Springer, Berlin, 1957; Fizmatgiz, Moscow, 1960).
16. T. Åberg, Phys. Rev. A **2**, 1726 (1970).
17. A. Dalgarno and H. R. Sadeghpour, Phys. Rev. A **46**, 3591 (1992).
18. E. G. Drukarev and M. B. Trzhaskovskaya, J. Phys. B **31**, 427 (1998).
19. J. A. R. Samson, Z. X. He, L. Yin, and G. N. Haddad, J. Phys. B **27**, 887 (1994).
20. T. Kato, Commun. Pure Appl. Math. **10**, 151 (1957).
21. R. T. Pack and W. B. Brown, J. Chem. Phys. **45**, 556 (1966).
22. R. H. Pratt, Phys. Rev. **117**, 1017 (1960); Phys. Rev. **119**, 1619 (1960).

23. A. R. P. Rau and U. Fano, Phys. Rev. **162**, 68 (1967).
24. R. H. Pratt and H. K. Tseng, Phys. Rev. A **5**, 1063 (1972).
25. J. McEnnan, L. Kissel, and R. H. Pratt, Phys. Rev. A **13**, 532 (1976).
26. S. D. Oh, J. McEnnan, and R. H. Pratt, Phys. Rev. A **14**, 1428 (1976).
27. A. Bechler and R. H. Pratt, Ann. Phys. (N.Y.) **163**, 28 (1985).
28. L. I. Schiff, *Quantum Mechanics*, 3rd ed. (McGraw-Hill, New York, 1968; Inostrannaya Literatura, Moscow, 1957), p. 420.
29. A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1967; Nauka, Moscow, 1978), p. 1052.
30. J. J. Sakurai, *Advanced Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1967), p. 73.
31. R. H. Pratt, A. Ron, and H. K. Tseng, Rev. Mod. Phys. **45**, 273 (1973).
32. M. A. Kornberg and J. E. Miraglia, Phys. Rev. A **52**, 2915 (1995).
33. J. H. Scofield, UCRL Report No. 51326 (1973).
34. N. B. Avdonina, E. G. Drukarev, and R. H. Pratt, Phys. Rev. A **65**, 052705 (2002).

Journal of Experimental and Theoretical Physics, Vol. 97, No. 2, 2003, pp. 232–258.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 2, 2003, pp. 259–287.

ATOMS, SPECTRA, RADIATION

Instanton versus Traditional WKB Approach to the Landau–Zener Problem¶

V. A. Benderskii[a,b,] *, E. V. Vetoshkin[a], and E. I. Kats[b,c,] **

[a]*Institute for Problems of Chemical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*

[b]*Laue–Langevin Institute, F-38042, Grenoble, France*

**e-mail: bender@icp.ac.ru*

[c]*Landau Institute for Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*

***e-mail: kats@cpd.landau.ac.ru*

Received March 14, 2003

Abstract—Different theoretical approaches to the famous two-state Landau–Zener problem are briefly discussed. Apart from traditional methods of the adiabatic perturbation theory, the Born–Oppenheimer approximation with geometric phase effects, the two-level approach, and the momentum space representation, the problem is treated semiclassically in the coordinate space. In the framework of the instanton approach, we present a full and unified description of the 1D Landau–Zener problem of level crossing. The method enables us to accurately treat all four transition points (appearing at two-level crossing), while the standard WKB approach takes only two of them into account. The latter approximation is adequate for calculating the transition probability or for studying scattering processes, but it does not work in finding the corresponding chemical reaction rates, in which all four transition points can often be relevant in the typical range of parameters. Applications of the method and of the results may concern various systems in physics, chemistry, and biology. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The title of this paper might sound perplexing at first sight. What else can be said about the Landau–Zener (LZ) problem after the numerous descriptions in both research and textbook literature? However, although theoretical (and experimental) investigations of different LZ systems began more than 70 years ago, it still remains an active area of research. Various approaches to the LZ problem that have appeared in the literature (see, e.g., the list of publications [1–67], which is by no means complete) are not fully consistent with each other. We therefore think it important to discuss all these approaches in a single paper. We study the 1D LZ problem [1] of quantum mechanical transitions between the levels of a two-level system at the avoided level crossing. In the LZ theory, a quantum system is placed in a slowly varying external field. Naturally, the system then adiabatically follows the variation of an initially prepared discrete state until its time-dependent energy level crosses another level. Near the crossing point, the adiabaticity condition is evidently violated (because the semiclassical behavior is violated near turning points). The slow variation of the perturbation implies that the duration of the transition process is very long, and therefore the change in the action during this time is great. In this sense, the LZ problem is a semiclassical one (but with respect to time instead of a coordinate in the standard semiclassical problems).

It is well known that the problem presents the most basic model of nonadiabatic transitions that play a very important role in many fields of physics, chemistry, and biology. It is therefore not surprising that numerous monographs and a great number of papers have been devoted to this subject. In the literature, there are, roughly speaking, three approaches to semiclassical modeling of the LZ problem:

(i) the two-level system approach [2–8];

(ii) the adiabatic perturbation theory [9–21] (also see review paper [6]);

(iii) the momentum space representation [22–25].

Because different approaches to the LZ problem have been proposed, one of the immediate motivations of the present paper is to develop a uniform and systematic procedure for handling this problem. We show that the three methods listed above are equivalent for treating tunneling and overbarrier regions of parameters, and none of them can be applied, to the intermediate region of parameters where all four of the states involved in the LZ system are relevant. To study this region is our main objective in this paper. We also address the so-called connection matrices. In the standard textbook treatment of the LZ problem, only transition probabilities are calculated and expressed in terms

¶This article was submitted by the authors in English.

of the genuine two-level LZ formula successively applied at each diabatic level intersection. Evidently, such a procedure is an approximation to the general LZ problem, which includes at least four energy levels even in the simplest case. To solve many important physical or chemical problems, one must find the 4×4 (not only 2×2) connection matrices relating these four states.

While this paper is not intended as a comprehensive review, we detail the key results of the standard WKB and instanton approaches from our own research and the literature within the context of different factors that we feel are important in studying the LZ problem. Specifically, we focus in Section 2 on the Born–Oppenheimer approximation, which is a benchmark in testing semiclassical approximations. In Section 3, we lay the foundation of treating the LZ problem, the adiabatic perturbation theory. Section 4 is devoted to the generalization of the instanton method, which enables us to investigate the LZ problem in the momentum space. We show that for a potential that is linear in a 1D coordinate under consideration, the WKB semiclassical wave functions in the momentum space coincide with the instanton wave functions. For the quadratically approximated (parabolic) potentials, the instanton wave functions are exact and have no singularities (unlike the WKB wave functions; we recall that relations of the same type hold for the WKB and instanton wave functions in the coordinate space [26–29]).

We advocate the instanton approach in this paper, but it is worth noting that many important results have nevertheless been obtained in the framework of the WKB approach [1–8]. For example, one very efficient technique (the so-called propagator method) was proposed and elaborated by Miller and collaborators [34–36] (also see [26]). This approach uses semiclassic propagators (of the Van Vleck–Gutzwiller type), with the contribution coming from the contour around a complex turning point automatically taken into account in terms of the general WKB formalism. The accuracy of the WKB method can be improved considerably [2, 5, 30, 31] (more recent references on the so-called Laplace contour integration can also be found in [32]) by the appropriate choice of the integration path around the turning point. This method appears to be quite accurate for the tunneling and overbarrier regions, but becomes inadequate in the intermediate energy region. This has been overlooked in previous investigations treating this region by a simple interpolation from the tunneling region (with a monotonic decay of the transition probability) to the overbarrier region (with oscillating behavior).

In Section 5, we present all details of the LZ problem for two electronic states using the instanton description of the LZ problem in the coordinate space. The two basic second-order differential (Schrödinger) equations that we consider are written in the so-called diabatic state representation (i.e., in the basis of "crossed" levels). Neglecting higher order spatial derivatives, we find asymptotic solutions, and using the adiabatic–diabatic transformation, we match the solutions in the intermediate region. The complete scattering matrix for the LZ problem is derived in Section 6. In Section 7, we derive the quantization rules for crossing diabatic potentials and briefly discuss the application of the results in some particular models of level crossings that are relevant for the interpretation and description of experimental data on the spectroscopy of nonrigid molecules, on inelastic atomic collisions [33], and nonradiative transitions arising from "intersystem" crossings of potential energy surfaces in the molecular spectroscopy and chemical dynamics (see, e.g., [26] and references therein). In Section 8, we draw our conclusions.

We consider only the 1D case in what follows. The LZ problem for 1D potentials coupled to the thermal bath of harmonic oscillators is shown to reduce to a certain renormalization of the Massey parameter, where the longitudinal velocity entering the expression for this parameter is decreased due to the coupling to transverse oscillations (see [26] and references therein, and also [66, 67] for more recent references). Of course, the energy profile of any real system is characterized by a multidimensional surface. However, it is often possible to identify a reaction coordinate such that the energy barrier between the initial and final states is minimized along this specific direction, and the system can therefore be effectively treated as 1D. In certain systems, the physical interpretation of the reaction coordinate is immediate (e.g., the relative bond length in diatomic molecules), but sometimes finding it is not an easy task (if possible at all) because of the large number of possibilities involved. The latter (multidimensional) case will be studied elsewhere. Unfortunately, the accuracy of the WKB method near the top of the barrier is too poor to make any numbers realistic, and this is one more motivation to use a semiclassical formalism alternative to the WKB, namely, the extreme tunneling trajectory or instanton technique.

2. BORN–OPPENHEIMER APPROXIMATION

It may be useful to illustrate the essential physics of the LZ problem starting with a very well-known picture corresponding to the Born–Oppenheimer approximation [1, 37]. It leads to the separation of nuclear and electronic motions and is valid only because the electrons are much lighter than the nuclei and therefore move much faster. The small parameter of the Born–Oppenheimer approximation is therefore given by

$$\lambda = \left(\frac{m_e}{m}\right)^{1/4} \ll 1, \tag{2.1}$$

where m_e and m are electronic and nuclear masses, respectively. On the other hand, the semiclassical

parameter is

$$\gamma = \frac{m\Omega a^2}{\hbar} \gg 1, \tag{2.2}$$

where a is the characteristic length in the problem and $\Omega \propto m^{-1/2}$ is the characteristic nuclear vibration frequency; therefore, $\gamma \propto \lambda^{-2}$. Important conclusions are drawn from this simple fact. Indeed, the semiclassical condition $\gamma \gg 1$ can be satisfied by formally taking $\hbar \longrightarrow 0$ or, equivalently, $\lambda \longrightarrow 0$. This correspondence allows us to apply either the Born–Oppenheimer or the semiclassical approximation to the separation of scales for nuclear and electronic motions on the same footing.

In the traditional Born–Oppenheimer approach, solution Ψ to the full Schrödinger equation (including the electronic Hamiltonian H_e depending on electronic coordinates r and the nuclear Hamiltonian depending on nuclear coordinates R) is given by an expansion over the electronic Hamiltonian eigenfunctions ϕ_n,

$$\Psi = \sum_n \Phi_n(R)\phi_n(r, R). \tag{2.3}$$

The electronic eigenvalues E_n depend on the nuclear coordinates, and the expansion coefficients $\Phi_n(R)$ are determined by the Born–Oppenheimer equations

$$\left[-\frac{\hbar^2}{2m}\nabla_R^2 + E_n(R) + \frac{\hbar^2}{2m}\sum_{k\neq n} A_{nk}A_{kn} - E\right]\phi_n$$
$$= -\frac{\hbar^2}{2m}\sum_{k,\,m\neq n} (\delta_{nk}\nabla_R - iA_{nk})(\delta_{km}\nabla_R - iA_{km})\phi_m, \tag{2.4}$$

where for $m \neq k$,

$$A_{mk} = i\langle\phi_m|\nabla_R\phi_k\rangle, \tag{2.5}$$

and all the diagonal matrix elements $A_{nn} = 0$.

From (2.4), we can find that in the electronic eigenstate E_n, the nuclei move in the effective potential

$$U_n(R) = E_n(R) + \frac{\hbar^2}{2m}\sum_{k\neq n} A_{nk}A_{kn}, \tag{2.6}$$

and transitions between the electronic states n and m are related to the nonadiabatic operator in the right-hand side of (2.4). This simple observation allows us to rewrite effective potential (2.6) as

$$U_n(R) = E_n(R)$$
$$-\frac{\hbar^2}{2m}\sum_{m\neq n}\frac{\langle\phi_n|\nabla_R H_e|\phi_m\rangle\langle\phi_m|\nabla_R H_e|\phi_n\rangle}{(E_n - E_m)^2}. \tag{2.7}$$

From this seemingly trivial expression, we derive the following important conclusions:

(i) corrections to E_n have the same order $O(\gamma^{-2})$ as the ratio of the nuclear kinetic energy to the potential;

(ii) off-diagonal matrix elements of the nonadiabatic perturbation operator are also small ($\propto O(\gamma^{-2})$); this fact is formulated as the so-called adiabatic theorem that states that no transitions between unperturbed states occur at adiabatic perturbations ($\lambda \longrightarrow 0$).

Because the nonadiabatic effects are characterized by the only small parameter, γ^{-1} (the semiclassical parameter), they can be described in the framework of semiclassical approaches (e.g., WKB or instanton ones). But, we must bear in mind the main problem of the Born–Oppenheimer method: the approximation assumes that the electronic wave functions are real-valued and form a complete basis, but it is impossible to construct such a basis in the entire space, including classically accessible and forbidden regions.

If the requirement of a real-valued basis is relaxed, the diagonal matrix elements $A_{nn} \neq 0$ and the effective adiabatic part of the Born–Oppenheimer Hamiltonian takes the form

$$\hat{H}_n = U_n(R) + \frac{\hbar^2}{2m}(\nabla_R - iA_{nn}(R))^2, \tag{2.8}$$

similarly to the Hamiltonian of a charged particle in the magnetic field $B \propto |\nabla_R \times A_{nn}|$. We can therefore change the phases of the electronic and nuclear wave functions as

$$\phi_n \longrightarrow \phi_n \exp(i\chi_n(R)),$$
$$\Phi_n \longrightarrow \Phi_n \exp(-i\chi_n(R)) \tag{2.9}$$

by changing the "vector potential" appropriately,

$$A_{nn}(R) \longrightarrow A_{nn}(R) + \nabla_R\chi_n(R). \tag{2.10}$$

Thus, we confront an important and, at times, mysterious concept of the geometric (or Berry) phase factor that a quantum mechanical wave function acquires upon a cyclic evolution [38–47]. Most characteristic of the concept of the Berry phase is the existence of a continuous parameter space in which the state of the system can travel along a closed path. In our case, the phase is determined by a nonadiabatic interaction (for more details related to the geometric phase for the Born–Oppenheimer systems, see, e.g., review [48]). This phenomenon (which originally manifested itself as a certain extra phase shift appearing upon some cyclic evolution of an external parameter) has been generalized for the nonadiabatic, noncyclic, and nonunitary cases [49, 50], although most of the Berry phase applications concern systems undergoing adiabatic evolution (see, e.g., review [51]). We also note that in addition to the Berry phase, some higher order corrections to the Born–Oppenheimer approximation also exist (traditionally, and slightly misleadingly, called the geo-

metric magnetism or deterministic friction, see [52]). A practically useful application of the Berry phase concept is the energy level displacements predicted in [53] and observed by NMR [54].

The essential physics of these phenomena can be illustrated as follows. There are two subsystems, fast and slow. The fast subsystem acquires a Berry phase because of the evolution of the slow subsystem. There is a certain feedback effect of the geometric phase on the slow subsystem. As a result, the latter is framed by a gauge field affecting its evolution. The gauge field produces additional (Lorentz-like and electric field–like) forces that must be included into the classical equation of motion. In the case of stochastic external forces (e.g., from surrounding thermal fluctuation media), the Berry phase produces some level broadening for the fast subsystem. In the limit of low temperatures and strong damping, the slow subsystem dynamics can be described by equations of the Langevin type [55]. The general message that we can learn from this fact is that the geometric phases are sources of the dissipative processes for LZ systems.

Thanks to its fundamental origin, this geometric phase has attracted considerable theoretical and experimental attention, but its experimentally observable consequences have been scarce until now. Each opportunity to improve this situation is therefore worth trying. In this respect, the Born–Oppenheimer geometric phase provides a unique opportunity for observation of the geometric phase because it must appear as a nonadiabatic contribution to the standard Bohr–Sommerfeld quantization rule

$$S_n^0 + \chi_n = 2\pi\hbar, \tag{2.11}$$

where S_n^0 is the adiabatic action.

We note that care must be taken when $|E_n(R) - E_m(R)|$ becomes small compared to the characteristic nuclear oscillation energy $\hbar\Omega$. This means that the nonadiabatic interaction energy cannot then be considered as a small perturbation in adiabatic representation (2.4). Fortunately, in the limit

$$|E_n(R) - E_m(R)| < \hbar\Omega,$$

we can start from the other limit with crossing weakly coupled diabatic states and consider the adiabatic coupling as a perturbation. To perform the procedure explicitly, we then need the adiabatic–diabatic transformations

$$\tilde{\Phi}(R) = \exp(i\theta\sigma_y)\Phi(R) \tag{2.12}$$

for the wave functions and

$$\tilde{H} = \exp(i\theta\sigma_y)H\exp(-i\theta\sigma_y) \tag{2.13}$$

for the Hamiltonians, where (H, Φ) and $(\tilde{H}, \tilde{\Phi})$ are the adiabatic and diabatic representations, respectively; σ_y is the corresponding Pauli matrix; and θ is the adiabatic–diabatic transformation parameter (the so-called adiabatic angle).

To illustrate how this works, we consider two coupled crossing effective electronic potentials $U_1(R)$ and $U_2(R)$ (U_{12} is the coupling energy). The corresponding adiabatic and diabatic Hamiltonians are

$$\begin{aligned} H &= -\frac{\hbar^2}{2m}(\nabla_R)^2 + \frac{1}{2}(U_1 + U_2) \\ &+ \left[\frac{1}{2}(U_1 - U_2)\cos(2\theta(R)) + U_{12}\sin(2\theta(R))\right]\sigma_3 \\ &+ \frac{1}{2}\left[-\frac{1}{2}(U_1 - U_2)\sin(2\theta(R)) + U_{12}\cos(2\theta(R))\right]\sigma_1, \end{aligned} \tag{2.14}$$

and

$$\begin{aligned} \tilde{H} &= -\frac{\hbar^2}{2m}(\nabla_R)^2 + \frac{1}{2}(U_1 + U_2) \\ &+ \frac{1}{2}(U_1 - U_2)\sigma_3 U_{12}\sigma_1, \end{aligned} \tag{2.15}$$

where $\sigma_{1,2,3}$ are the Pauli matrices and the adiabatic angle is chosen to eliminate the leading interaction term between the adiabatic states,

$$\cos(2\theta(R)) = \frac{U_1 - U_2}{2U_{12}}. \tag{2.16}$$

The adiabatic–diabatic transformation can also be brought to a more elegant form [16, 56]:

$$(\nabla_R - i\hat{A})\hat{T} = 0, \tag{2.17}$$

where $\hat{T}$ is the sought transformation matrix and the matrix $\hat{A} \equiv A_{nn}$ was introduced above (see (2.5)). The formal solution of Eq. (2.17) can be represented as a contour integral:

$$\hat{T}(s) = \hat{T}(s_0)\exp\left(-\int_{s_0}^{s}\hat{A}(s')ds'\right), \tag{2.18}$$

where s_0 and s are the initial and final points of the contour. Solution (2.18) uniquely determines the transformation matrix $\hat{T}$ for a curl-free field $\hat{A}$,

$$\hat{T}(t_0) = \hat{D}\hat{T}(0), \tag{2.19}$$

where the diagonal matrix $\hat{D}$ can be found from (2.17) and is expressed in terms of the geometric phase factor as

$$D_{kn} = \delta_{kn}\exp(i\chi_k). \tag{2.20}$$

Relations (2.11) and (2.20) completely describe the nonadiabatic transitions, the cornerstone of the LZ

problem. In addition, (2.11) and (2.20) show that the geometric Born–Oppenheimer phases occur from the diabatic-potential crossing points and enter the quantization rules additively with contributions from the turning points. Therefore, our main conclusion in this section is that nonadiabatic phenomena must (and can) be included into the general scheme of the semiclassical approach via the corresponding connection matrices [57] (also see [29]) for the appropriate combinations of crossing and turning points in the problem.

3. ADIABATIC PERTURBATION THEORY

It is almost common student's wisdom nowadays that any solution to the adiabatically time-dependent Schrödinger equation can be represented as an expansion over the complete set of stationary (time-independent) eigenfunctions [1]. In the case under investigation (two-level crossing for the electronic Hamiltonian $H_e(r, t)$), this expansion is given by

$$\Psi(r, t) = c_1(t)\phi_1(r) + c_2(t)\phi_2(r), \tag{3.1}$$

where the wave functions $\phi_{1,2}$ are stationary with respect to nuclear motion. The time-dependent Schrödinger equation can be exactly rewritten as two first-order equations (with respect to time derivatives) for c_1 and c_2,

$$i\hbar\begin{pmatrix} \dot c_1 \\ \dot c_2 \end{pmatrix} = \begin{pmatrix} \tilde H_{11} & \tilde H_{12} \\ \tilde H_{21} & \tilde H_{22} \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \tag{3.2}$$

where

$$\tilde H_{kk'} = \langle \phi_k|\tilde H(t)|\phi_{k'}\rangle, \quad k, k' = 1, 2 \tag{3.3}$$

are the matrix elements for the diabatic Hamiltonian.

The phase transformation

$$c_k(t) = a_k(t)\exp\left(-\frac{i}{\hbar}\int \tilde H_{kk}(t)dt\right) \tag{3.4}$$

(see [6, 8, 10]) reduces (3.2) to the coupled first-order equations

$$\begin{aligned} i\hbar\dot a_1 &= \tilde H_{12}a_2\exp\left(i\int\Omega_{12}(t)dt\right), \\ i\hbar\dot a_2 &= \tilde H_{21}a_1\exp\left(-i\int\Omega_{12}(t)dt\right), \end{aligned} \tag{3.5}$$

where

$$\Omega_{12} = \frac{1}{\hbar}(\tilde H_{22} - \tilde H_{11}). \tag{3.6}$$

A slightly different phase transformation,

$$c_k(t) = \tilde\Phi_k(t)\exp\left(\frac{i}{2\hbar}\int(\tilde H_{11} + \tilde H_{22})dt\right) \tag{3.7}$$

preserves the second-order Schrödinger-like form of the equations for the diabatic functions $\tilde\Phi_{1,2}$,

$$\begin{aligned} \hbar^2\frac{d^2\tilde\Phi_1}{dt^2} - \left[\left(\frac{\tilde H_{11} - \tilde H_{22}}{2}\right)^2 + \tilde H_{12}\tilde H_{21}\right. \\ \left. + \frac{i\hbar}{2}\frac{d}{dt}(\tilde H_{11} - \tilde H_{22})\right]\tilde\Phi_1 = 0. \end{aligned} \tag{3.8}$$

To clarify the mapping of this time-dependent perturbation theory to the two-level crossing problem and the Born–Oppenheimer approach described in Section 2, we consider the two-state Born–Oppenheimer equations in the diabatic representation. From (2.15) for one active space coordinate X, we have

$$-\frac{\hbar^2}{2m}\frac{d^2\tilde\Phi_1}{dX^2} + (\tilde H_{11} - E)\tilde\Phi_1 = \tilde H_{12}\tilde\Phi_2 \tag{3.9}$$

and

$$-\frac{\hbar^2}{2m}\frac{d^2\tilde\Phi_2}{dX^2} + (\tilde H_{22} - E)\tilde\Phi_2 = \tilde H_{21}\tilde\Phi_1. \tag{3.10}$$

If we can neglect the second-order derivatives

$$\frac{\hbar^2}{2m}\frac{d^2\tilde\Phi_{1,2}}{dX^2}$$

and replace the time derivative by $v d/dX$ (where $v = \sqrt{2E/m}$ is the velocity), the change of the variables

$$\tilde\Phi_{1,2} = \exp(ik_0X)c_{1,2}, \quad k_0^2 = \frac{2mE}{\hbar^2} \tag{3.11}$$

transforms the two Born–Oppenheimer equations (3.9) and (3.10) into the two level-crossing equations (3.2) for slow time-dependent perturbations. Obviously, we recognize the standard semiclassical approach in this procedure.

A mapping of the same kind can also be performed for the adiabatic amplitudes $C_{1,2}(t)$ that are related to the diabatic amplitudes $c_{1,2}(t)$ by the adiabatic–diabatic transformation matrix depending on the adiabatic angle θ,

$$\begin{pmatrix} C_1(t) \\ C_2(t) \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} c_1(t) \\ c_2(t) \end{pmatrix}. \tag{3.12}$$

In the adiabatic basis, we have the set of the first-order equations corresponding to (3.2),

$$\begin{pmatrix} \dot C_1 \\ \dot C_2 \end{pmatrix} = \begin{pmatrix} H_{11} & -i\dot\theta \\ i\dot\theta & H_{22} \end{pmatrix}\begin{pmatrix} C_1 \\ C_2 \end{pmatrix}, \tag{3.13}$$

where the nonadiabatic coupling coefficient $\dot{\theta}$ can be related to the off-diagonal operator A_{12} in (2.5) (or to the geometric phase, see Section 2),

$$i\dot{\theta} = A_{12} \equiv i\langle \phi_1 | \dot{\phi}_2 \rangle. \tag{3.14}$$

Transformation (3.11) allows us to reduce the Born–Oppenheimer equations (for the nuclear wave functions $\Phi_{1,2}$ in the adiabatic representation) to (3.13) if and only if the second-order derivatives are neglected (in the spirit of the semiclassical approach) and only $\propto k_0$ terms are kept in the nonadiabatic matrix elements (i.e., higher order contributions with respect to $1/k_0$ are neglected). Expressions (3.12)–(3.14) do allow an entry point into the adiabatic perturbation theory developed by Landau [1] and Dykhne [10, 11] (also see [15, 16]). We follow the same method closely.

We can go one step further and find the combination of the two-level system amplitudes $a_{1,2}$ in (3.4) and (3.5),

$$\begin{aligned} Y(t) &= \Omega_{12}^{-1/2} \exp\left(-\frac{i}{2}\int \Omega_{12} dt\right) a_1 \\ &+ i\Omega_{12}^{-1/2} \exp\left(\frac{i}{2}\int \Omega_{12} dt\right) a_2, \end{aligned} \tag{3.15}$$

satisfying the simple equation

$$\ddot{Y}(t) + \frac{\Omega_{12}^2}{4} Y = 0, \tag{3.16}$$

which is identical to (3.8) and describes oscillations around the crossing point in the adiabatic potential (inverted adiabatic barrier). In the adiabatic perturbation theory, the level-crossing problem is therefore formally reduced to the well-known quantum mechanical phenomenon, the overbarrier reflection. In the latter problem, moreover, the reflection coefficient is equal to 1, in full agreement with the adiabatic theorem.

Evidently, two adiabatic potentials have no real crossing points in the 1D case, and the crossing is therefore possible only at complex values of X or t,

$$\Omega_{12}(\tau_c) = 0; \quad U_1 - U_2 = \pm i U_{12}\big|_{t=\tau_c}. \tag{3.17}$$

In the vicinity of these points, it follows from (3.6) that

$$\Omega_{12} \propto (t - \tau_c)^{1/2}, \tag{3.18}$$

and, therefore,

$$\int \Omega_{12} dt \approx \frac{2}{3}(t - \tau_c)^{3/2}, \tag{3.19}$$

i.e., the crossing points are square root bifurcation points for the function $\Omega_{12}(t)$. Using (3.19), we depicted the Stokes and anti-Stokes lines for Eq. (3.16) in Fig. 1. The diagram shown in this figure is identical to that corresponding to the semiclassical overbarrier reflection problem with the linear turning points under consideration. In the leading approximation, the transition probability P_{12} is determined by integration over the contour $C(\tau_c)$ going around the bifurcation point τ_c,

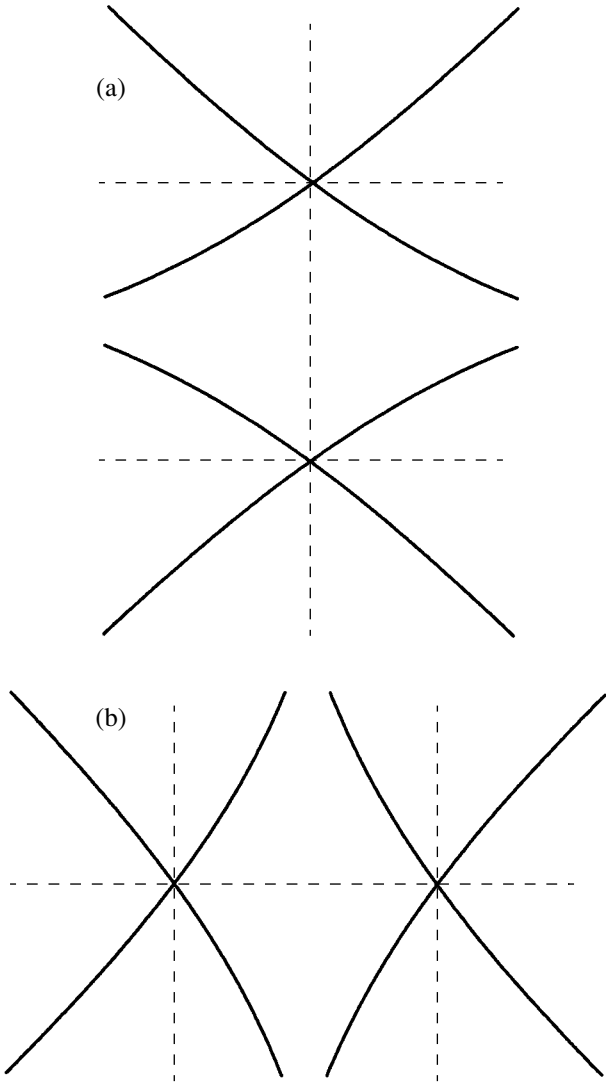

Fig. 1. Stokes (dashed) and anti-Stokes (solid) lines for a pair of close linear turning points replaced by one second-order turning point; (a) classically forbidden region, (b) classically accessible region.

$$P_{12} \approx \exp\left\{\frac{2}{\hbar}\oint\limits_{C(\tau_c)} (H_{11} - H_{22}) dt\right\}. \tag{3.20}$$

In the simplest form of the LZ problem, the diabatic potentials are assumed to be linear functions of t or X (which is the same because $t = X/V$), see Fig. 2 for illustration,

$$U_{1(2)} = U^{\#} \pm FX. \tag{3.21}$$

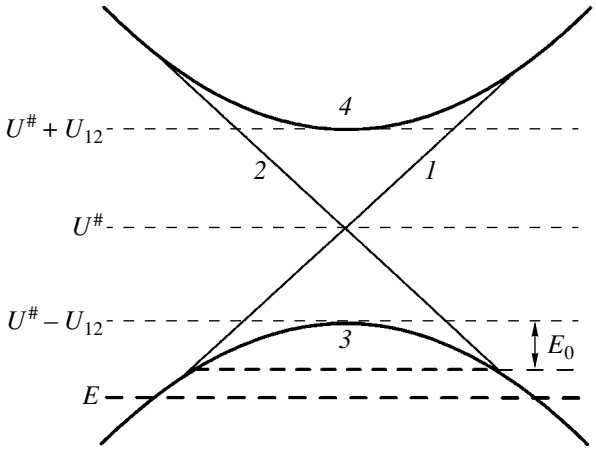

Fig. 2. Adiabatic (*3*, *4*) and diabatic (*1*, *2*) potentials for the LZ problem.

Substituting (3.21) in the general expression for the transition probability (3.20), we then find

$$P_{12} \approx \exp(-2\pi\nu), \tag{3.22}$$

where $\nu = U_{12}^2/2\hbar vF$ is the so-called Massey parameter and

$$v = \sqrt{\frac{2|E - U^{\#}|}{m}}$$

is the velocity.

Some comments about the validity range of the approximation are in order. A question of primary importance for the LZ problem is related to the semiclassical nature of the phenomenon. To illustrate this, we note that, for

$$\Omega_{12}^2 = U_{12}^2 + v^2F^2X^2,$$

Eq. (3.16) is the Weber equation for the real point $X = 0$ (the crossing point of diabatic potentials). Evidently, this correspondence between two complex-conjugate linear crossing points $\pm\tau_c$ and one real crossing point $X = 0$ for the Weber equation is the same as the correspondence between two linear and one second-order turning points in the standard semiclassical treatment of the Schrödinger equation. We can therefore apply the WKB or instanton methods to the LZ problem in the same way as in any semiclassical problem. We now compare the accuracy of the two approaches. If $|E - U^{\#}| \gg \hbar\Omega$ (where Ω is the characteristic frequency of the adiabatic potentials), the WKB method works quite well if two isolated linear turning points in this problem are considered (this is the limit of $k_0a \gg 1$, corresponding to the adiabatic approximation). If this is not the case, the diabatic representation must be used.

4. INSTANTON METHOD IN MOMENTUM SPACE

We do not explain the instanton method in detail here and summarize only the most essential points (see [26–29, 58, 59]). The recipe to find the instanton is based on minimizing the classical action functional in the space of paths connecting the minima in the upside-down potential. It is well known [1] that the expansion of an arbitrary wave function $\Psi(x)$ in terms of the momentum eigenfunctions is simply a Fourier integral,

$$\Psi(x) = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}\exp\left(\frac{ipx}{\hbar}\right)\Phi(p)dp. \tag{4.1}$$

The wave function in the momentum representation $\Phi(p)$ can be written in the semiclassical form

$$\Phi(p) = A(p)\exp\left(-\frac{iW(p)}{\hbar}\right), \tag{4.2}$$

where the action $W(p)$ is determined by the classical trajectory $x_0(p)$ in accordance with the definition

$$\frac{dW}{dp} = x_0(p). \tag{4.3}$$

We use the dimensionless variables $\epsilon = E/\Omega_0$ for the energy, $V = U/\gamma\Omega_0$ for the potential, and $X = x/a_0$ for the coordinate, where E and U are the corresponding dimensional values of the energy and of the potential, a_0 is the characteristic length of the problem (e.g., the tunneling distance), and Ω_0 is the characteristic frequency (e.g., the oscillation frequency around the potential minimum). The dimensionless momentum can be defined as

$$P = \frac{pa_0}{\gamma\hbar}, \tag{4.4}$$

where γ is the semiclassical parameter (we recall that $\gamma \equiv m\Omega_0a_0^2/\hbar$, where m is the mass of the particle, and we believe that $\gamma \gg 1$).

Introducing the semiclassical form (4.2) of the momentum-representation wave function in the standard one-particle 1D Schrödinger equation, we can transform it to the form

$$\left[P^2 + 2\hat{V}\left(X_0 + i\frac{1}{\gamma}\frac{d}{dP}\right) - \frac{2}{\gamma}\epsilon\right]A(P) = 0. \tag{4.5}$$

In the momentum space, $\hat{V}$ is the potential energy operator, which can be expanded in a semiclassical series with respect to $1/\gamma$ (or equivalently, with respect to $\hbar$; we set $\hbar = 1$ in what follows, measuring energies in

units of frequency, except in some intermediate equations where the occurrences of $\hbar$ are necessary for understanding). This expansion allows us to consider $\hat{V}$ as a function V of two independent variables X_0 and d/dP, and we finally obtain

$$V\left(X_0 + \frac{i}{\gamma}\frac{d}{dP}\right) = V(X_0) + \frac{i}{\gamma}\left(\frac{dV}{dX_0}\frac{d}{dP} + \frac{1}{2}\frac{d^2V}{dX_0^2}\frac{dX_0}{dP}\right) + \left(\frac{i}{\gamma}\right)^2\left[\frac{d^2V}{dX_0^2}\frac{d^2}{dP^2}\right. \quad (4.6)$$

$$\left. - \frac{1}{2}\frac{d^3V}{dX_0^3}\left(\frac{dX_0}{dP}\frac{d}{dP} - \frac{1}{3}\frac{d^2X_0}{dP^2}\right) + \frac{1}{24}\frac{d^4X}{dX_0^4}\left(\frac{dX_0}{dP}\right)^2\right] + \ldots,$$

where the dots denote all higher order expansion terms.

In accordance with the general semiclassical rules, we can easily find from (4.5) and (4.6) that the first- and the second-order terms in γ^{-1} become identically zero if the energy-dependent trajectory $X_0(P)$ is determined by the equation

$$P^2 + 2V(X_0) = \frac{2\epsilon}{\gamma} \quad (4.7)$$

and if the so-called transport equation (TE)

$$\frac{dV}{dX_0}\frac{dA}{dP} + \frac{1}{2}\frac{d^2V}{dX_0^2}\frac{d^2W}{dP^2}A \quad (4.8)$$

is also satisfied. The solution of TE (4.8) can be found explicitly as

$$A = \left(\frac{dV}{dX_0}\right)^{-1/2}. \quad (4.9)$$

It follows from (4.9) that semiclassical WKB wave function (4.2) has singularities at all stationary points of the potential V. These points are therefore turning points in the momentum space. This illustrates fundamental difficulties of the WKB procedure, which consist in matching the solutions that become singular on caustic lines separating manifolds with real and imaginary momenta in phase space.

To also illustrate the second drawback of the WKB method, we consider the linear ($V = FX$) and harmonic ($V = X^2/2$) potentials. The trajectories $X_0(P)$ can be trivially determined from (4.7). For the linear potential, $X_0(P)$ is an inverted parabola with a maximum of $X_{0m} = \epsilon F/\gamma$ at $P = 0$. The left and the right branches of the parabola correspond to the opposite motion directions in the classically accessible region $X_0 < X_{0m}$. For the linear potential, the semiclassical WKB wave function in the momentum space,

$$\Phi(P) = \frac{1}{\sqrt{F}}\exp\left\{-\frac{i}{F}\left(\epsilon P - \gamma\frac{P^3}{6}\right)\right\}, \quad (4.10)$$

is the Fourier transform of the coordinate-space Airy function. For the harmonic potential, the corresponding trajectories (4.7) are ellipses and the wave functions have the same functional form in both spaces (momentum and coordinate). It is worthwhile to note that although the WKB functions are not exact, the corresponding eigenvalues coincide with exact quantum mechanical ones.

As we have shown recently [27–29], many important semiclassical problems can be successfully analyzed by the instanton method. Bearing in mind momentum space in this section, we recall the main ideas of the instanton approach. The first step of the approach, derived in [58, 59], is the so-called Wick rotation of the phase space, corresponding to the transformation to imaginary time $t \longrightarrow -it$. During the transformation, both potential and kinetic energies change their signs and the Lagrangian is replaced by the Hamiltonian in the classical equation of motion. In the momentum space, the low-energy instanton wave functions can be constructed using Wick rotation in the momentum space (i.e., the transformation $P \longrightarrow iP$); in addition, the term with energy ϵ in (4.7) must be removed from this equation and taken into account in TE (4.8). In the instanton formalism, the trajectory $X_0(P)$ describes zero-energy motion in the classically forbidden region of the momentum space, where the wave function has the form

$$\Phi(P) = \left(\frac{dV}{dX_0}\right)^{-1/2} Q(P)\exp[-\gamma W(P)], \quad (4.11)$$

and the additional prefactor $Q(P)$ can be represented as

$$\ln Q(P) = \epsilon\int\left(\frac{dV}{dX_0}\right)^{-1} dP. \quad (4.12)$$

In the particular case of a linear potential ($V(X) = FX$), the instanton and WKB functions have the same form. For an arbitrary (nth-order) anharmonic potential, the Schrödinger equation in the momentum space is reduced to the nth-order differential equation, but the nth-order derivatives decrease proportionally to γ^{-n} and the corresponding terms can therefore be taken into account perturbatively. A rigorous mathematical method for performing this procedure (which we use in this paper) has been developed by Fedoryuk [68–70].

To illustrate the instanton approach, we consider the simplest form of the LZ problem illustrated in Fig. 3. For linear potentials with arbitrary line slopes, we have two second-order coupled equations, in the diabatic state representation

$$-\frac{d^2\Theta_1}{dX^2} = \gamma^2(\alpha + f_1X)\Theta_1 = \gamma^2\nu\Theta_2,$$
$$-\frac{d^2\Theta_2}{dX^2} = \gamma^2(\alpha + f_2X)\Theta_2 = \gamma^2\nu\Theta_1, \quad (4.13)$$

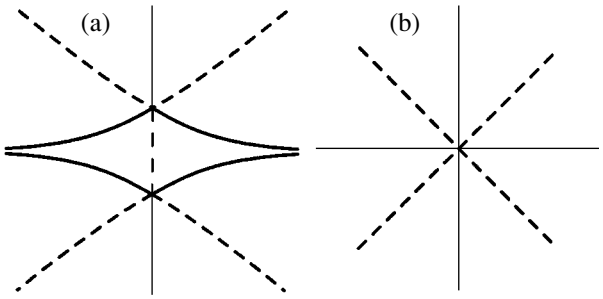

Fig. 3. Stokes (dashed) and anti-Stokes (solid) lines in the vicinity of (a) conjugate bifurcation points $\pm i\tau_c$ and (b) diabatic potentials crossing point $X = 0$.

where $\Theta_{1,2}$ are the eigenfunctions of the corresponding states, and:

$$\Omega^2 = \frac{a^2F^2}{mU_{12}}, \quad F = \sqrt{F_1|F_2|}, \quad \gamma = \frac{a^3Fm^{1/2}}{U_{12}^{1/2}},$$

$$\alpha = 2\frac{U_0 - E}{\gamma\Omega}, \quad f_{1,2} = 2\frac{aF_{1,2}}{\gamma\Omega}, \quad \nu = 2\frac{U_{12}}{\gamma\Omega}.$$

Equations (4.13) can be transformed into the momentum space and can then be rewritten as a single second-order equation:

$$\frac{d^2\Psi_1}{dk^2} + q(k)\Psi_1(k) = 0, \tag{4.14}$$

where we introduced

$$\Psi_1 = \Phi_1 \exp\left[i\frac{\gamma\alpha^{3/2}}{2}\left(\frac{1}{f_1} + \frac{1}{f_2}\right)\left(k + \frac{k^3}{3}\right)\right], \tag{4.15}$$

where Φ_1 is the Fourier transform of Θ_1, $k = P/\gamma\sqrt{\alpha}$, and $q(k)$ is a fourth-order characteristic polynomial

$$q(k) = \lambda^2(1 + k^2)^2 + 2\lambda(ik - 2\nu) \tag{4.16}$$

depending on two parameters

$$\lambda = \frac{1}{2}\gamma\alpha^{3/2}\left(\frac{1}{f_1} - \frac{1}{f_2}\right), \quad \nu = \frac{\gamma v^2}{2(f_1 - f_2)\sqrt{\alpha}}. \tag{4.17}$$

The first parameter λ plays the role of the new semiclassical parameter in the momentum representation, and the second is the known Massey parameter (already defined in (3.22)).

Fortunately, all roots of characteristic polynomial (4.16) can be found analytically quite accurately in the physically most interesting region of parameters. To simplify the expressions (while retaining the complete physical content), we present the results only in the simplest case, where $f_1 = -f_2 \equiv f$ (symmetric slopes of the diabatic potentials). In the classically forbidden region $U^{\#} - E > 0$, $\alpha > 0$, at $\lambda \gg 1$ (equivalently, at $\alpha \gg (f/\gamma)^{2/3}$), all four roots of the polynomial are close to $\pm i$,

$$k_1^{\pm} = i\left(1 \pm \sqrt{\frac{1+\nu}{2\lambda}}\right), \quad k_2^{\pm} = \pm\sqrt{\frac{1-\nu}{2\lambda}} - i. \tag{4.18}$$

In the classically accessible region ($U^{\#} - E < 0$, $\alpha < 0$), the roots are close to ± 1 if $\lambda \gg 1$ (or if $-\alpha \gg (f/\gamma)^{2/3}$),

$$k_1^{\pm} = 1 \pm \left(\frac{\sqrt{1+\tilde{\nu}^2} + \tilde{\nu}}{4\tilde{\lambda}}\right)^{1/2} \pm i\left(\frac{\sqrt{1+\tilde{\nu}^2} - \tilde{\nu}}{4\tilde{\lambda}}\right)^{1/2},$$
$$k_2^{\pm} = -1 \mp \left(\frac{\sqrt{1+\tilde{\nu}^2} + \tilde{\nu}}{4\tilde{\lambda}}\right)^{1/2} \pm i\left(\frac{\sqrt{1+\tilde{\nu}^2} - \tilde{\nu}}{4\tilde{\lambda}}\right)^{1/2} \tag{4.19}$$

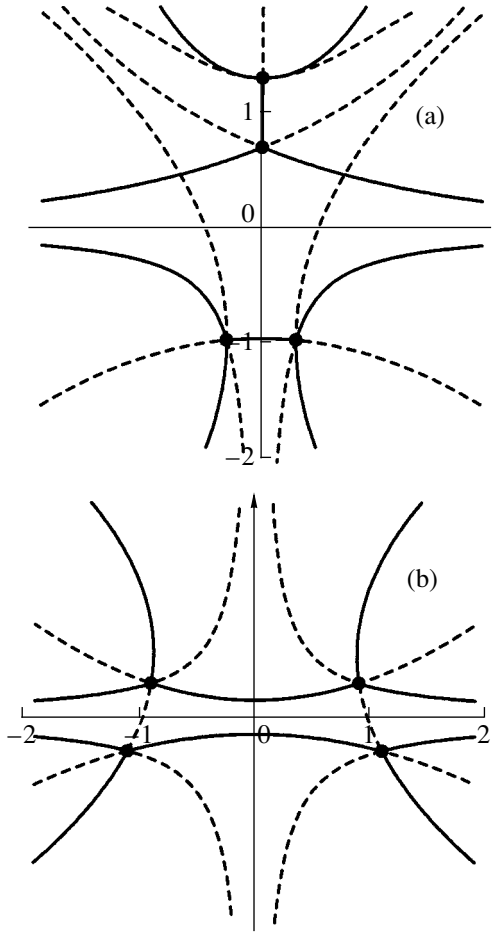

Fig. 4. Stokes (dashed) and anti-Stokes (solid) lines for linear turning points corresponding to classically forbidden (a) and accessible (b) energy regions of the LZ problem.

(the tilde means that in the corresponding quantity, α must be replaced with its modulus).

The roots of characteristic polynomial (4.16) in the classically forbidden region (Eq. (4.18)) and in the classically accessible region (Eq. (4.19)) are formally equivalent to the transition or turning points for the system of two potential barriers or two potential wells, respectively. We can therefore use all the WKB and instanton results known in these cases (see, e.g., our recent paper [29] and references therein for the details). Because only asymptotic solutions and their connections via transition or turning points on the complex plane are usually considered in the semiclassical analysis, the famous Stokes phenomenon [30, 57] of asymptotic solutions plays an essential role, and the distribution of the transition points (which are nothing but the zero points of the characteristic polynomial) and Stokes and anti-Stokes lines determines the phenomenon. We show all the lines emanating from linear turning points in Fig. 2. In the case where the roots form a pair of close linear turning points, each such pair can be replaced with one second-order turning point. The corresponding Stokes and anti-Stokes lines are depicted in Fig. 4.

In the classically forbidden region, the instanton wave functions can be found using roots (4.18),

$$\begin{aligned}\Phi_1^+ &= \frac{(1-ik)^{\nu-1}}{(1+ik)^{\nu+1}}\exp\left[i\lambda\left(k+\frac{k^3}{3}\right)\right],\\ \Phi_1^- &= \frac{(1-ik)^{-\nu}}{(1+ik)^{-\nu}}\exp\left[-i\lambda\left(k+\frac{k^3}{3}\right)\right].\end{aligned} \tag{4.20}$$

As $|k|\longrightarrow\infty$, the function Φ_1^+ decreases proportionally to $|k|^{-2}$ and Φ_1^- is reduced to the Airy function [71, 72]. In the vicinity of the second-order turning points $k=\pm i$, the fourth-order characteristic polynomial is reduced to a second-order one, and Eq. (4.14) is reduced to the Weber equation with the known fundamental solutions [71]

$$D_{-\nu}(\pm 2\sqrt{\lambda}(k+i))$$

as $|k+i|\longrightarrow 0$ and

$$D_{-\nu-1}(\pm 2\sqrt{\lambda}(k-i))$$

as $|k-i|\longrightarrow 0$. The same procedure applied to the classically accessible region leads to the solutions

$$\begin{aligned}\Phi_1^+ &= \frac{(1-k)^{i\tilde{\nu}-1}}{(1+k)^{i\tilde{\nu}+1}}\exp\left[i\tilde{\lambda}\left(k-\frac{k^3}{3}\right)\right],\\ \Phi_1^- &= \frac{(1+k)^{i\tilde{\nu}}}{(1-k)^{i\tilde{\nu}}}\exp\left[-i\tilde{\lambda}\left(k-\frac{k^3}{3}\right)\right],\end{aligned} \tag{4.21}$$

and it is also reduced to the fundamental solutions of the Weber equation

$$D_{i\tilde{\nu}}\left(\pm 2\sqrt{\tilde{\lambda}}(k+1)\exp\frac{i\pi}{4}\right)$$

as $|k+1|\longrightarrow 0$ and

$$D_{i\tilde{\nu}-1}\left(\pm 2\sqrt{\tilde{\lambda}}(k-1)\exp\frac{i\pi}{4}\right)$$

as $|k-1|\longrightarrow 0$.

The same solutions can be obtained for the LZ problem in the two-level approximation using the instanton method in the coordinate space. The reason for this is quite transparent and is based on the fact that for linear diabatic potentials, the limit $k\longrightarrow\pm\infty$ corresponds to the limit $x\longrightarrow\pm\infty$, and the asymptotic behaviors of the solutions are therefore the same in the momentum and in the coordinate space.

The entire analysis can be brought into a more compact form by introducing the so-called connection matrices. In the instanton approach, we consider asymptotic solutions and their connections on the complex coordinate plane. It is therefore important to know the connection matrices. The needed connection matrices can easily be found by matching solutions (4.20) or (4.21) at the second-order turning points via the corresponding fundamental solutions of the Weber equation. This gives the connection matrices

$$\hat{M}_1=\begin{pmatrix}-\cos(\pi\nu) & \dfrac{\sqrt{2\pi}\exp(-2\chi)}{\Gamma(\nu)}\\ \dfrac{\Gamma(\nu)\exp(2\chi)\sin^2(\pi\nu)}{\sqrt{2\pi}} & \cos(\pi\nu)\end{pmatrix}, \tag{4.22}$$

where

$$\chi=\nu-\frac{(\nu-1/2)\ln\nu}{2},$$

and

$$\hat{M}_2=\begin{pmatrix}-\exp(-\pi\tilde{\nu}) & \dfrac{\sqrt{2\pi}\exp(-\pi\tilde{\nu})\exp(-2\tilde{\chi})}{\Gamma(-i\tilde{\nu})}\\ \dfrac{1}{\sqrt{2\pi}}2\Gamma(-i\tilde{\nu})\exp\left(-\dfrac{\pi\tilde{\nu}}{2}\right)\exp(2\tilde{\chi})\sinh(\pi\tilde{\nu}) & \exp(-\pi\tilde{\nu})\end{pmatrix}, \tag{4.23}$$

where

$$\tilde{\chi} = -i\left(\frac{\pi}{4} + \tilde{\nu}(1 - \ln\tilde{\nu})\right) + \frac{1}{4}\ln\tilde{\nu}.$$

As a note of caution at the end of this section, we remind the reader that for the linear diabatic potentials, we initially had two corresponding Schrödinger equations, each of which possesses two fundamental solutions. Therefore, the full LZ problem is characterized by four fundamental solutions that are asymptotic to the left of a given turning point and four fundamental solutions that are asymptotic to the right of the same turning point. Generally speaking, the connection matrices must therefore be 4×4 ones. But because of the symmetry of the potentials, these 4×4 matrices have two 2×2 block structures for the functions Φ_1 and Φ_2, given in (4.22) and in (4.23).

5. LZ PROBLEM FOR TWO ELECTRON STATES (INSTANTON APPROACH IN COORDINATE SPACE)

In Sections 2–4, we investigated the LZ problem in the framework of the adiabatic perturbation theory, the two-level approximation, and the momentum representation. All three methods are equivalent and semiclassical by their nature and are therefore applicable in the tunneling and overbarrier energy regions; they become inadequate within the intermediate region (on the order of $\gamma^{-2/3}$) near the level crossing point. The fact is that the accuracy of these methods depends on the "renormalized" (energy-dependent) semiclassical parameter λ in (4.17), which can be small in the intermediate region ($\lambda \leq 1$ even for $\gamma \gg 1$). To treat this region, we must use the coordinate space presentation, because we need to know the connection matrices for nonadiabatic transitions. In the latter problem, the wave functions outside the level crossing point are more convenient (and have a more compact mathematical form) in the coordinate space.

5.1. Tunneling and Overbarrier Regions

For the smoothness of presentation, we first reproduce the results found in the previous sections for the tunneling and overbarrier energy regions in the coordinate space. In the diabatic representation, we can rewrite two second-order LZ differential equations (4.13) as the fourth-order linear differential equation with constant coefficients at the derivatives

$$\frac{d^4\Phi_1}{dX^4} - 2\gamma^2\alpha\frac{d^2\Phi_1}{dX^2} - 2\gamma^2 f\frac{d\Phi_1}{dX} + \gamma^4(\alpha^2 - \nu^2 - f^2X^2)\Phi_1 = 0 \quad (5.1)$$

(where we consider the case with a symmetric slope $f_1 = -f_2 \equiv f$ for simplicity). In the mathematical formalism elaborated by Fedoryuk [68–70], Eq. (5.1) is reduced by a semiclassical substitution in a set of equations of the order γ^n. The characteristic polynomial for (5.1) is given by

$$F(\lambda) = \lambda^4 - 2\alpha\gamma^2\lambda^2 - 2\gamma^2 f\lambda + \gamma^4(\alpha^2 - \nu^2 - f^2X^2), \quad (5.2)$$

where $\lambda = dW/dX$ by definition.

Solving the equation $F(\lambda) = 0$ perturbatively in $\gamma^{-1} \ll 1$, we find

$$\lambda_j = \lambda_j^0 + u_j, \quad (5.3)$$

where

$$\lambda_j^0 = \pm[\gamma(\alpha \pm \sqrt{\nu^2 + f^2X^2})]^{1/2} \quad (5.4)$$

and

$$u_j = \frac{\gamma f}{2}[(\lambda_j^0)^2 - \alpha\gamma]^{-1}. \quad (5.5)$$

Four asymptotic solutions of (5.1) can then be represented as

$$\{y_j\} \equiv \{\Phi_+^+, \Phi_+^-, \Phi_-^+, \Phi_-^-\} = (\nu^2 + f^2X^2)^{-1/4}\exp\left[\int_0^X \lambda_j(X')dX'\right]. \quad (5.6)$$

They describe the motion with an imaginary momentum in the upper and lower adiabatic potentials

$$\frac{2ma^2}{\hbar^2}(U^{\pm} - E) = \gamma^2(\alpha \pm \sqrt{\nu^2 + f^2X^2}).$$

The subscripts in (5.6) correspond to the upper or lower adiabatic levels, and the superscripts indicate the sign of the action.

Before considering the connection matrices, we use the substitution

$$\Phi_1 = \exp(\kappa X)\phi, \quad (5.7)$$

and choose the κ value such that the first derivative in (5.1) vanishes,

$$\kappa^3 - \gamma^2\alpha\kappa - \frac{1}{2}\gamma^2 f = 0. \quad (5.8)$$

At $\alpha > 3(f/4\gamma)^{2/3}$, we can expand the roots of (5.8) in terms of the parameter

$$\delta = \frac{f}{4\gamma}\alpha^{-3/2} < \frac{1}{3\sqrt{3}}. \quad (5.9)$$

We thus find

$$\kappa_1 = \gamma\sqrt{\alpha}\left(1+\frac{\delta}{2}\right),$$
$$\kappa_2 = \gamma\sqrt{\alpha}\left(-1+\frac{\delta}{2}\right), \quad \kappa_3 = \gamma\sqrt{\alpha}\delta. \tag{5.10}$$

Under condition (5.9), the coefficients at the fourth and at the third-order derivatives in (5.1) are small (proportional to δ and $\sqrt{\delta}$, respectively) and the fourth-order equation (5.1) can be rewritten as two second-order Weber equations with the solutions

$$D_{p^{(1,2)}}(\beta_{(1,2)}X),$$

where

$$p^1 = -1+\frac{\delta}{2}-\nu\left(1-\frac{3\delta}{2}\right),$$
$$p^2 = \frac{\delta}{2}-\nu\left(1+\frac{3\delta}{2}\right), \tag{5.11}$$
$$\beta_{(1,2)} = \left(\frac{\gamma^2 f^2}{\alpha}\right)^{1/4}\left(1\pm\frac{3\delta}{4}\right).$$

The leading terms of these solutions are the same as those found in Section 4. But the Fedoryuk method also gives higher order corrections in δ in tunneling region (5.8).

In the overbarrier energy region where $\alpha < -3(f/4\gamma)^{2/3}$, the roots of Eq. (5.8) are complex conjugate,

$$\frac{\kappa_{(1,2)}}{\gamma\sqrt{\alpha}} = -\frac{\tilde{\delta}}{2}\pm i\left(1+\frac{3\tilde{\delta}^2}{8}\right), \tag{5.12}$$

and

$$\tilde{\delta} = \frac{f}{4\gamma|\alpha|^{3/2}} \tag{5.13}$$

plays the role of a small parameter. Similarly to the case with the tunneling region, the coefficients at higher order derivatives are small and the function ϕ in (5.7) therefore satisfies the Weber equation with the fundamental solutions

$$D_{\tilde{p}^{(1,2)}}(\tilde{\beta}_{(1,2)}X),$$

where

$$\tilde{p}^1 = -1+i\frac{3\tilde{\delta}}{2}+i\nu\left(1+\frac{3\tilde{\delta}}{4}\right),$$
$$\tilde{p}^2 = i\frac{3\tilde{\delta}}{2}+i\nu\left(1-\frac{3\tilde{\delta}}{4}\right),$$
$$\tilde{\beta}_1 = \exp\frac{i\pi}{4}\left(\frac{\gamma^2 f^2}{|\alpha|}\right)^{1/4}, \tag{5.14}$$
$$\tilde{\beta}_2 = \exp\left(-\frac{i3\pi}{4}\right)\left(\frac{\gamma^2 f^2}{|\alpha|}\right)^{1/4}.$$

As was the case with tunneling region (5.11), the leading terms of expansion (5.14) coincide with the results found in the previous sections, but (5.14) also allows us to compute corrections to the leading terms.

We can now find the connection matrices. To do this in the tunneling region, we must establish the correspondence between solutions of fourth-order differential equation (5.1) and solutions for the states localized in the left (L) and in the right (R) wells. In the case where $\alpha \gg f|X|$, the action can be computed for diabatic potentials starting from both wells (R and L),

$$\gamma W^L \approx \gamma W_0^L + k_0 X + \frac{\beta^2}{4}X^2,$$
$$\gamma W^R \approx \gamma W_0^R - k_0 X + \frac{\beta^2}{4}X^2, \tag{5.15}$$

where

$$k_0 = \left(\frac{2m(U^{\#}-E)}{\hbar^2}\right)^{1/2} \equiv \gamma\sqrt{\alpha}$$

is an imaginary momentum and $W_0^{L,R}$ are the actions computed from an arbitrary distant point in the L or R wells, respectively, to the point $X = 0$. On the other hand, in the adiabatic potentials

$$U^{\pm} = U^{\#} \pm \sqrt{U_{12}^2 + f^2X^2},$$

the corresponding actions can be represented as

$$\gamma W^{\pm} - \gamma W_0^{\pm} = k_0 X \pm \frac{\beta^2}{4}X^2\,\mathrm{sgn}\,X. \tag{5.16}$$

Explicitly comparing the semiclassical wave functions in both representations (adiabatic and diabatic ones), it is easy to see that the adiabatic functions in the potential U^- coincide with the diabatic functions for localized L and R states at $X < 0$ and $X > 0$, respectively. The adiabatic functions for the upper potential U^+ correspond to the tails of the diabatic wave functions localized in the opposite wells. In the level crossing region, the L/R diabatic functions are therefore transformed into the R/L functions and the interaction entangles the diabatic states with the same sign of k_0X. Thus, we have only four nonzero amplitudes of the following transitions:

$$\langle\Phi_L^+|\Phi_R^-\rangle,\ \langle\Phi_L^-|\Phi_R^+\rangle,\ \langle\Phi_R^+|\Phi_L^-\rangle,\ \langle\Phi_R^-|\Phi_L^+\rangle. \tag{5.17}$$

Recalling that

$$\gamma W^{\pm} = \gamma\int(\alpha \pm \sqrt{\nu^2 + f^2X^2})^{1/2}dx \approx k_0X \pm \frac{\beta^2}{4}X^2 \pm \frac{\nu}{2}(1 - \ln\nu), \tag{5.18}$$

we conclude that quantum solutions (5.11), asymptotically valid in the vicinity of the level crossing point, match increasing and decreasing solutions (5.6) smoothly, which leads to the Landau description [1] of the level crossing transitions depicted in Fig. 5.

Using expressions (4.22) and (4.23) relating the fundamental solutions of the Weber equation, we can find the 4×4 connection matrix corresponding to (5.17),

$$\begin{pmatrix} \Phi_R^+ \\ \Phi_R^- \\ \Phi_L^+ \\ \Phi_L^- \end{pmatrix} = \begin{bmatrix} \frac{\sqrt{2\pi}\exp(-2\chi)}{\Gamma(\nu)} & 0 & 0 & -\cos(\pi\nu) \\ 0 & \frac{\Gamma(\nu)\exp(2\chi)\sin^2(\pi\nu)}{\sqrt{2\pi}} & -\cos(\pi\nu) & 0 \\ 0 & \cos(\pi\nu) & \frac{\sqrt{2\pi}\exp(-2\chi)}{\Gamma(\nu)} & 0 \\ \cos(\pi\nu) & 0 & 0 & \frac{\Gamma(\nu)\exp(2\chi)\sin^2(\pi\nu)}{\sqrt{2\pi}} \end{bmatrix} \begin{pmatrix} \Phi_L^- \\ \Phi_L^+ \\ \Phi_R^- \\ \Phi_R^+ \end{pmatrix}, \tag{5.19}$$

where

$$\chi = \frac{\nu}{2} - \frac{1}{2}\left(\nu - \frac{1}{2}\right)\ln\nu$$

as above. The matrix in (5.19) has a 2×2 block structure, with each of the identical blocks connecting increasing and decreasing diabatic solutions. However, these diagonal blocks do not correspond to the *L*–*R* transitions for the lower and upper adiabatic potentials separately. Indeed, the 2×2 matrix corresponding to these transitions is

$$\begin{pmatrix} \Phi_R^+ \\ \Phi_R^- \end{pmatrix} = \begin{bmatrix} \frac{\sqrt{2\pi}\exp(-2\chi)}{\Gamma(\nu)} & -\cos(\pi\nu) \\ \cos(\pi\nu) & \frac{\Gamma(\nu)\exp(2\chi)\sin^2(\pi\nu)}{\sqrt{2\pi}} \end{bmatrix} \begin{pmatrix} \Phi_L^- \\ \Phi_R^+ \end{pmatrix}. \tag{5.20}$$

In the diabatic limit (i.e., as $\nu \longrightarrow 0$) the diagonal matrix elements are small ($\propto\nu^{1/2}$ and $\nu^{3/2}$, respectively), and the off-diagonal elements tend to ± 1, as it should be because by definition, there are no transitions between the diabatic potentials.

In the adiabatic limit $\nu \gg 1$, the diagonal matrix elements tend to 1, which implies that the decreasing *L* solution transforms only into the increasing *R* solution, and vice versa. Therefore, the connection matrix in the tunneling region depends only on the Massey parameter ν. We recall that the blocks of the 4×4 connection matrix in (5.19) correspond to the two isolated second-

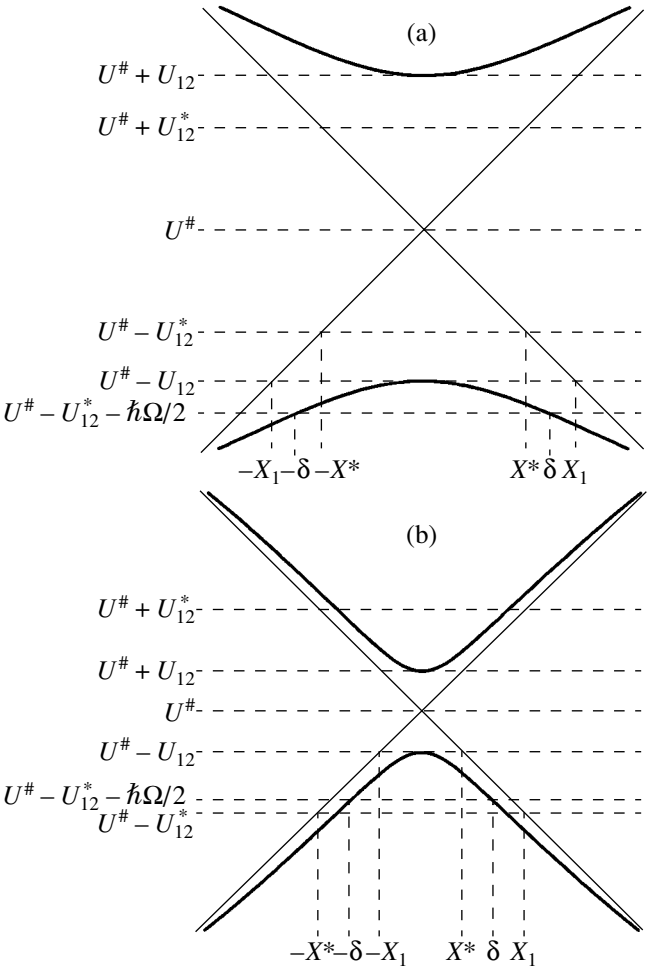

Fig. 5. Relative placement of the adiabatic levels; (a) $U_{12} > U_{12}^*$, (b) $U_{12} < U_{12}^*$ ($U_{12}^* \equiv (3/2)(\hbar^2F^2/4m)^{1/3}$).

order turning points with the Stokes constant (see, e.g., [29])

$$T_2 = \frac{\sqrt{2\pi}}{\Gamma(\nu)}\exp(-2\chi). \quad (5.21)$$

The overbarrier region can be studied similarly. Repeating the procedure described above for the tunneling region (with the evident replacements $k_0 \longrightarrow -ik_0$ and $\beta^2 \longrightarrow i\beta^2$), we obtain the 4×4 connection matrix

$$\hat{U} = \begin{bmatrix} \frac{\sqrt{2\pi}\exp(-2\tilde{\chi})}{\Gamma(-i\nu)} & 0 & 0 & -\exp(-\pi\nu) \\ 0 & \frac{2\Gamma(-i\nu)\exp(-\pi\nu)\exp(2\tilde{\chi})\sinh(\pi\nu)}{\sqrt{2\pi}} & -\exp(-\pi\nu) & 0 \\ 0 & \exp(-\pi\nu) & \frac{\sqrt{2\pi}\exp(-2\tilde{\chi})}{\Gamma(-i\nu)} & 0 \\ \exp(-\pi\nu) & 0 & 0 & \frac{2\Gamma(-i\nu)\exp(2\tilde{\chi})\exp(-\pi\nu)\sinh(\pi\nu)}{\sqrt{2\pi}} \end{bmatrix}, \quad (5.22)$$

where

$$\tilde{\chi} = -\frac{i}{2}\left[\frac{\pi}{4} + \nu(1 - \ln\nu)\right] + \frac{1}{4}(\pi\nu + \ln\nu). \quad (5.23)$$

As already mentioned for the tunneling region, the blocks in (5.22) correspond to the two isolated second-order turning points with the Stokes constant [29]

$$\tilde{T}_2 = \frac{\sqrt{2\pi}}{\Gamma(-i\nu)}\exp(-2\tilde{\chi}). \quad (5.24)$$

Thus, we arrive at the important conclusion that the main peculiarity of the LZ level crossing (in comparison with the standard, e.g., one-potential problems) is that the second-order turning points characterizing the diabatic level crossing for the LZ problem possesses different Stokes constants T_2 (Eq. (5.21)) and $\tilde{T}_2$ (Eq. (5.24)) in the tunneling and in the overbarrier regions.

5.2. Intermediate Energy Region

We can now reap the fruits of our effort in the previous subsection. We first note that Eqs. (5.11) and (5.14) imply that as the energy approaches the top of the barrier, the indices $p^{(i)}$ and $\tilde{p}^{(i)}$ of the parabolic cylinder functions increase and therefore deviate more and more from the value prescribed by the Massey parameter ν. Second, $\beta_{(i)}$ increases as $|\alpha|$ decreases, resulting in a decrease in the values of $|X|$ where the asymptotic smooth matching of the solutions must be performed. As $\delta \longrightarrow 0$, these $|X|$ values are located deeply in the classically forbidden region, where the potentials are close to the diabatic potentials; for $\delta \geq 2\sqrt{3}/3$, these coordinates $|X|$ are on the order of the quantum zero-point oscillation amplitudes and, therefore, the adiabatic representation must be used to find the solution in this region.

These two simple observations give us a conjecture on how to treat the LZ problem in the intermediate energy region. We must first find the energy "window" for the intermediate region. It is convenient to choose the adiabatic potential frequency $\Omega = F/\sqrt{mU_{12}}$ as the energy scale such that the inequality $|\alpha| < 3|f/4\gamma|^{2/3}$ becomes

$$|U^* - E| \leq \frac{3}{2}U_{12}^{1/3}\left(\frac{\Omega}{2}\right)^{2/3} \equiv U_{12}^*. \quad (5.25)$$

In other words, the characteristic interaction energy at the boundaries of the intermediate region is independent of U_{12}. However, the positions of the linear turning points $|X^*|$ corresponding to the energies $U^* \pm U_{12}^*$ depend on the ratio U_{12}/U_{12}^*. These points are located inside or outside the interval $[-a_0\gamma^{-1/2}, a_0\gamma^{-1/2}]$ at $U_{12}/U_{12}^* < 1$ and at $U_{12}/U_{12}^* > 1$, respectively, and the matching conditions in the intermediate energy region are therefore different in the two cases. In the former case, the potentials can be reasonably approximated by a parabola in the asymptotic matching region, and we must therefore work with the Weber equations. In the latter case, the matching is performed in the region where the potentials are linear, and the equations are therefore reduced to Airy equations.

We first investigate the case where $U_{12}/U_{12}^* > 1$. Using the Born–Oppenheimer approach described in

Section 2, we see that the Schrödinger equations for the wave functions $\Psi_\pm$ are decoupled in the adiabatic representation with an accuracy up to γ^{-2},

$$-\frac{d^2\Psi_\pm}{dX^2} + \gamma^2(\alpha \pm \sqrt{\nu^2 + f^2X^2})\Psi_\pm = 0. \qquad (5.26)$$

For $|X| < \nu/f$, Eqs. (5.26) is reduced to Weber equations with the fundamental solutions

$$D_{-1/2-q_1}(\pm\sqrt{2\gamma}X)$$

and

$$D_{-1/2+iq_2}\left(\pm\exp\left(-\frac{i\pi}{4}\right)\sqrt{2\gamma}X\right),$$

where

$$q_1 = \gamma\frac{\nu+\alpha}{2}, \quad q_2 = \gamma\frac{\nu-\alpha}{2} \qquad (5.27)$$

are independent of the Massey parameter ν. Two real solutions of (5.26) correspond to the upper adiabatic potential (classically forbidden region), and two complex solutions correspond to the classically allowed motion under the lower adiabatic potential.

The argument of the Weber functions is $\propto X\sqrt{\gamma}$, and under the condition $X < \nu/f$, their asymptotic expansions determine the interval where the matching is to be performed,

$$\gamma^{-1/2}\left(\frac{U_{12}}{\Omega}\right)^{1/2} > \gamma^{-1/2}. \qquad (5.28)$$

This inequality can be satisfied only at $U_{12}/U_{12}^* > 1$, when the intermediate region is sufficiently broad in comparison with Ω. The exponents q_1 and q_2 in Eq. (5.27) are then large, and our aim is to find the explicit asymptotic expansions of solutions in this case. For this, we closely follow the method in [73] (also see monograph [72]), which is in fact an expansion of the fundamental Weber solutions in the small parameters $1/|q_i|$. This method leads to the asymptotic solution of Eq. (5.26) at $X > 0$, given by

$$\Psi_+^-(X) \approx Y_+^{-1/2}(X+Y_+)^{-q_1}\exp(-\gamma XY_+),$$
$$\Psi_-^-(X) \approx Y_-^{-1/2}(X+Y_-)^{iq_2}\exp(i\gamma XY_-), \qquad (5.29)$$

where $Y_\pm = \sqrt{\nu \pm \alpha + X^2}$. Using the known relation between the fundamental solutions of the Weber equation [71, 72],

$$D_\mu(z) = \exp(-i\pi\mu)D_\mu(z) + \frac{\sqrt{2\pi}}{\Gamma(-\mu)}\exp\left(-i\pi\frac{\mu+1}{2}\right)D_{-\mu-1}(iz),$$

we can find the other two solutions (complementary to (5.29)) as

$$\Psi_+^+(X) = Y_+^{-1/2}\left[-\sin(\pi q_1)(X+Y_+)^{-q_1}\exp(-\gamma XY_+) + \exp(-2\chi_1)\frac{\sqrt{2\pi}}{\Gamma(1/2+q_1)}(X+Y_+)^{q_1}\exp(\gamma XY_+)\right], \qquad (5.30)$$

and

$$\Psi_-^+(X) = Y_+^{-1/2}\left[-i\exp(-\pi q_2)(X+Y_-)^{iq_2}\exp(i\gamma XY_-) + \exp(-2\chi_2)\frac{\sqrt{2\pi}}{\Gamma(1/2-iq_2)}(X+Y_-)^{iq_2}\exp(-i\gamma XY_-)\right], \qquad (5.31)$$

where we introduce the notation

$$\chi_1 = \frac{1}{2}\left(q_1+\frac{1}{2}\right) - \frac{q_1}{2}\ln\left(q_1+\frac{1}{2}\right),$$

$$\chi_2 = -\frac{1}{2}\left(iq_2-\frac{1}{2}\right) + \frac{iq_2}{2}\left[-i\frac{\pi}{2} + \ln\left(q_2+\frac{i}{2}\right)\right].$$

Not surprisingly, solutions (5.29)–(5.31) can be represented as a linear combination of the semiclassical solutions $\Phi_\pm^\pm$ in (5.6) with the coefficients

$$\cos(2\theta_{(1,2)}) = \frac{X}{\sqrt{\nu \pm \alpha + X^2}}. \qquad (5.32)$$

These energy-dependent angles $\theta_{(1,2)}$ coincide with the adiabatic angles introduced above (see (2.12) and (3.12)) at the level crossing point $\alpha = 0$, and $f|X| < \nu$. Both angles take only slightly different values over the entire intermediate region $|\alpha| < \nu$.

We can now find all the connection matrices for these functions. Although the calculation is straightforward, it must be performed with caution (e.g., because the X-dependent matrices have different functional forms at positive and negative X). For $X > 0$, we obtain

$$\begin{pmatrix} \Psi_-^- \\ \Psi_-^+ \\ \Psi_+^- \\ \Psi_+^+ \end{pmatrix} = \begin{bmatrix} \cos\theta_2 & 0 & 0 & 0 \\ -i\exp(-\pi q_2)\cos\theta_2 & \dfrac{\sqrt{2\pi}\exp(-2\chi_2)\cos\theta_2}{\Gamma(1/2-iq_2)} & 0 & 0 \\ 0 & 0 & \sin\theta_1 & 0 \\ 0 & 0 & -\sin(\pi q_1)\sin\theta_1 & \dfrac{\sqrt{2\pi}\exp(-2\chi_1)\sin\theta_1}{\Gamma(1/2+q_1)} \end{bmatrix} \begin{pmatrix} \Phi_-^+ \\ \Phi_-^- \\ \Phi_+^- \\ \Phi_+^+ \end{pmatrix}, \quad (5.33)$$

and for $X < 0$,

$$\begin{pmatrix} \Psi_-^- \\ \Psi_-^+ \\ \Psi_+^- \\ \Psi_+^+ \end{pmatrix} = \begin{bmatrix} \dfrac{\sqrt{2\pi}\exp(-2\chi_2)\cos\theta_2}{\Gamma(1/2-iq_2)} & -i\exp(-\pi q_2)\cos\theta_2 & 0 & 0 \\ 0 & \cos\theta_2 & 0 & 0 \\ 0 & 0 & \dfrac{\sin\theta_1\sqrt{2\pi}\exp(-2\chi_1)}{\Gamma(1/2+q_1)} & -\sin(\pi q_1)\sin\theta_1 \\ 0 & 0 & 0 & \sin\theta_1 \end{bmatrix} \begin{pmatrix} \Phi_-^- \\ \Phi_-^+ \\ \Phi_+^+ \\ \Phi_+^- \end{pmatrix}. \quad (5.34)$$

The product of the matrix inverse to (5.33) and the matrix in (5.34) determines the sought connection matrix relating the semiclassical solutions in the intermediate energy region (cf. the connection matrices for the tunneling and overbarrier energy regions in (5.20) and (5.22)). Performing this simple algebra, we finally obtain

$$U_{\text{cross}} = \begin{bmatrix} \dfrac{\sqrt{2\pi}\exp(-2\chi_2)}{\Gamma(1/2-iq_2)} & i\exp(-\pi q_2) & 0 & 0 \\ -i\exp(-\pi q_2) & \dfrac{2\exp(2\chi_2)\Gamma(1/2-iq_2)\cosh(\pi q_2)\exp(-\pi q_2)}{\sqrt{2\pi}} & 0 & 0 \\ 0 & 0 & \dfrac{\sqrt{2\pi}\exp(-2\chi_1)}{\Gamma(1/2+q_1)} & \sin(\pi q_1) \\ 0 & 0 & -\sin(\pi q_1) & \dfrac{\cos^2(\pi q_1)\Gamma(1/2+q_1)\exp(2\chi_1)}{\sqrt{2\pi}} \end{bmatrix}. \quad (5.35)$$

This matrix has a two 2×2 block structure, similarly to the connection matrices (5.19) and (5.22) for the tunneling and overbarrier regions. However, unlike matrices (5.19) and (5.22) describing the transitions between the diabatic states, matrix (5.35) corresponds to transitions between adiabatic states. Indeed, at a strong level coupling ($U_{12} > U_{12}^*$), the eigenfunctions are close to the adiabatic functions and only nonadiabatic perturbations induce transitions. Therefore, the off-diagonal matrix elements in (5.35), which have the meaning of the probability that the diabatic state remains unchanged after the transition, are zero. The block with real-valued matrix elements corresponds to the minimum of the upper adiabatic potential, i.e., to an isolated second-order turning point where [29]

$$q_1 = \frac{U^* - E + U_{12}}{\Omega}. \quad (5.36)$$

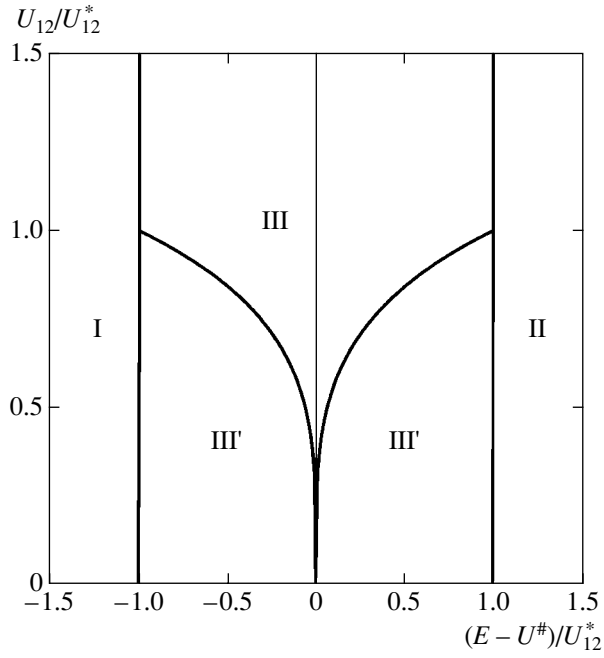

Fig. 6. The E, U_{12} phase diagram: (I) tunneling region, (II) overbarrier region. The two intermediate energy regions III and III' are separated by the line $\nu^* = 0.325$.

The complex-valued block is associated with the maximum of the lower adiabatic potential, and similarly to (5.36), we can find the relation

$$iq_2 = -i\frac{U^* - E + U_{12}}{\Omega} \tag{5.37}$$

for the turning point. For weak level coupling, namely, at $|U^* - E| < U_{12}^*$ and $U_{12} < U_{12}^*$ in the intermediate energy region, the adiabatic potentials can be linearized everywhere except a small neighborhood $|X| < \nu/f \longrightarrow 0$ of the level crossing point; i.e., they can be represented as $\alpha \pm f|X|$. Asymptotic solutions (5.6) are then reduced to a linear combination of the functions

$$\begin{aligned}\Phi_+^{\pm} &\propto (f|X|)^{-1/2}\exp(\pm\xi_+ \operatorname{sgn} X),\\ \Phi_-^{\pm} &\propto (f|X|)^{-1/2}\exp(\pm\xi_- \operatorname{sgn} X),\\ \xi_{\pm} &= \frac{2}{3f}(f|X| \pm \alpha)^{3/2}.\end{aligned} \tag{5.38}$$

All the matrix elements required can now be calculated in the framework of the Landau perturbation theory [1], which can be formulated in terms of the dimensionless variables

$$\tilde{\alpha} = 3\times 2^{-4/3}\frac{U^* - E}{U_{12}}, \quad \tilde{\nu} = 3\times 2^{-4/3}\frac{U_{12}}{U_{12}^*}$$

in order to avoid a divergence of the parameter ν as $\alpha \longrightarrow 0$. The results of our analysis are shown in Fig. 6. The tunneling and overbarrier regions are separated from the intermediate energy region by the lines $|U_{12}^* - E| = U_{12}^*$. The intermediate region is also split into two parts by the line $\nu = \nu^* = 0.325$, where ν^* is the value of the Massey parameter ν at $U_{12}/U_{12}^* = 1$ and $|U^* - E| = U_{12}^*$. In the region $\nu < \nu^*$, perturbation theory is an adequate tool for the problem and the transition matrix elements are proportional to U_{12}/U_{12}^*. At $\nu > \nu^*$, we can use connection matrix (5.35). To illustrate the accuracy of the approximations, we have computed the matrix element M_{11}. The results are shown in Fig. 7. Our computations demonstrate a sufficiently good precision, secured up to two stable digits. The accuracy of the results on the boundaries between the intermediate and overbarrier or tunneling regions is not worse than 3–5% and can easily be improved using interpolation approaches.

6. SCATTERING MATRIX

Phenomena of the LZ type can be considered as (and applied to) scattering processes. The expressions for the 4×4 connection matrices found in Section 5 can be used to calculate the scattering operator (or matrix) $\hat{S}$ that converts an incoming wave into an outgoing one.

We first consider the overbarrier region in the crossing problem with two linear potentials. In this case, in addition to the crossing point chosen as $X = 0$, there are two linear (first-order) turning points $X_0 = \pm|\alpha|/f$ (each turning point for each of the diabatic potentials denoted by L and R). The scattering matrix that relates the asymptotic solutions at $X \ll -X_0$ and $X \gg X_0$ is the product of the 4×4 connection matrix (5.22) and the two known semiclassical connection matrices [57] (also see [29]) describing the wave function evolution from the turning point $-X_0$ to the crossing point 0, and from this point to the turning point $+X_0$, respectively. We thus obtain a 2×2 matrix with the block matrix elements

$$T_{11} = A_{if}\begin{bmatrix}\exp(i(\phi - \phi_0)) & 0\\ 0 & \exp(-i(\phi - \phi_0))\end{bmatrix},$$

$$T_{12} = T_{21}^* = (1 - A_{if}^2)\exp\frac{i\gamma W^*}{2}$$

$$\times\begin{bmatrix} i & -1/2\\ -\exp(-i\gamma W^*) & (i/2)\exp(-i\gamma W^*)\end{bmatrix}, \tag{6.1}$$

$$T_{22} = A_{if}$$

$$\times \begin{bmatrix} 2\cos(\gamma W^* - (\phi - \phi_0)) & -\sin(\gamma W^* - (\phi - \phi_0)) \\ \sin(\gamma W^* - (\phi - \phi_0)) & (1/4)\cos(\gamma W^* - (\phi - \phi_0)) \end{bmatrix},$$

where

$$A_{if} = (1 - \exp(-\pi\nu))^{1/2}$$

is the LZ amplitude of the transition between the diabatic states, $\phi - \phi_0 = \tilde{\chi}$ (see (5.23)), and W^* is the action between the linear turning points.

The diagonal elements in (6.1), proportional to the transition amplitude A_{if}, describe propagating waves (i.e., solutions of the Schrödinger equation in the lower adiabatic potential), and the oscillating blocks correspond to solutions in the upper adiabatic potential. Off-diagonal blocks, proportional to the probability that the initial diabatic states remain unchanged, describe the waves reflected from the linear turning points. The reflection (R) and transmission (T) coefficients, of interest in physical applications, can be found from (6.1) by straightforward calculation,

$$R = -i(1 - A_{if}^2)$$

$$\times [A_{if}^2 \exp((i\gamma W^* - 2i(\phi - \phi_0)) + \exp(-i\gamma W^*))]^{-1},$$

$$T = 2A_{if}\cos(\gamma W^* - (\phi - \phi_0)) \tag{6.2}$$

$$\times [A_{if}^2 \exp(i\gamma W^* - 2i(\phi - \phi_0)) + \exp(-i\gamma W^*)]^{-1}.$$

The poles of the scattering matrix can also be easily found from (6.1), and the corresponding resonance condition is

$$\cos[2(\gamma W^* - (\phi - \phi_0))]$$

$$= -\left(1 - \frac{1}{2}\exp(-2\pi\nu)\right)(1 - \exp(-2\pi\nu))^{-1/2}. \tag{6.3}$$

The action is complex-valued at the resonance points,

$$\mathrm{Re}(\gamma W^* - (\phi - \phi_0)) = \left(n + \frac{1}{2}\right)\pi,$$

$$\mathrm{Im}(\gamma W^* - (\phi - \phi_0)) = -\frac{1}{2}\ln(1 - \exp(-2\pi\nu)). \tag{6.4}$$

The poles of the scattering matrix are in the lower half-plane of complex E on the vertical lines corresponding to the conventional Bohr–Sommerfeld quantization rules ($\gamma W^* = \pi(n + 1/2)$) for the upper adiabatic potential. In the diabatic limit ($\nu \longrightarrow 0$), the imaginary part of the pole positions tends to infinity, and in the adiabatic limit ($\nu \longrightarrow \infty$), the poles move to the real axis. Thus, we see that the eigenstates of the upper adiabatic potential are always quasistationary. The resonance widths are determined by the residues of the scattering matrix elements at the poles and can be shown to be monotonically decreasing functions of ν. In Fig. 8, we show the energy dependence of the transmission coefficient T. In the diabatic limit, $T \longrightarrow 0$, and it increases as U_{12} increases. In the overbarrier region, there appear resonances of widths Γ_n increasing with the energy increase, because the Massey parameter then decreases and $\Gamma_n \propto \exp(-2\pi\nu)$.

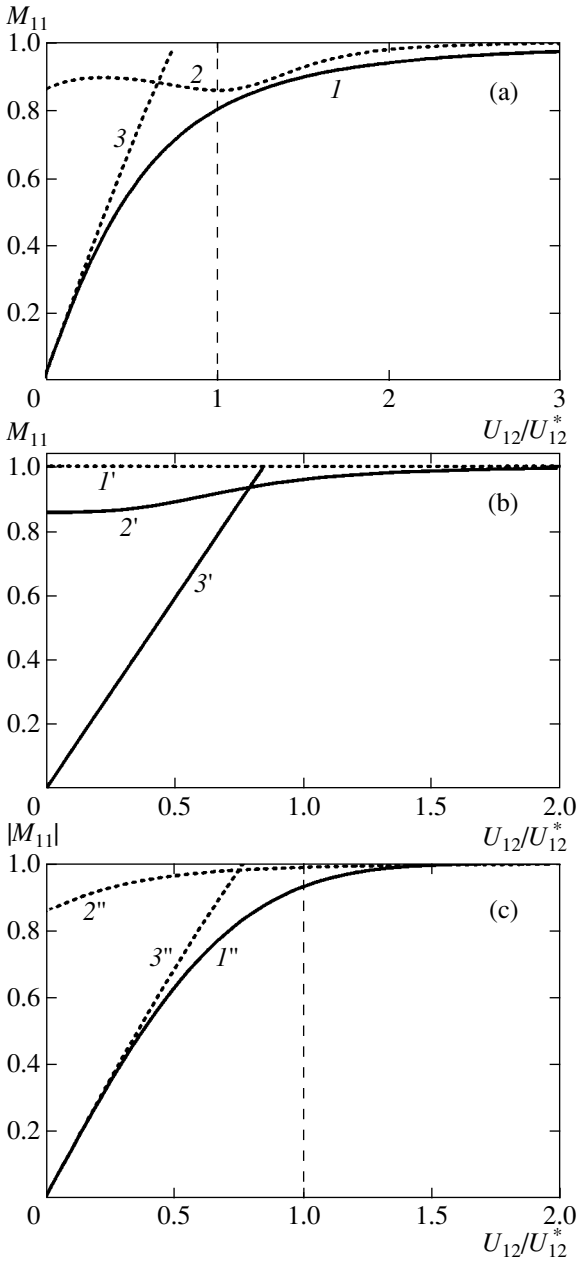

Fig. 7. Transition matrix element M_{11} as a function of U_{12}/U_{12}^*, computed at $\alpha = 0$: (a) on the boundary between tunneling and intermediate energy regions; (b) at $E = U^{\#}$; and (c) on the boundary between the intermediate and overbarrier regions; lines *1–3*, *1'–3'*, *1"–3"* are computed for the corresponding energy regions using (5.19), (5.25), and (5.36), respectively.

We illustrate the energy dependence of the transmitted wave phase in Fig. 9. In accordance with the general scattering theory [1], there are π-jumps of the phase at each quasidiscrete energy level of the upper adiabatic potential. At $U_{12}/U_{12}^* < 1$, the resonance widths are on the order of the interlevel spacings. The amplitudes of the decaying solutions (localized in the well formed by the upper adiabatic potential) increase near the resonances; this behavior is illustrated in Fig. 10. A primarily important point is that the information about decaying solutions contained in the 4×4 connection matrix (e.g., (5.22)) is lost when we use 2×2 scattering matrix (6.1).

The scattering matrix for the tunneling region can be found by minor modifications of the expression already derived. Instead of matrix (6.1), we thus obtain

$$T_{11} = \begin{bmatrix} (1/4)M_{11}\exp(-\gamma W^*) + M_{22}\exp(\gamma W^*) & i((1/4)M_{11}\exp(-\gamma W^*) - M_{22}\exp(\gamma W^*)) \\ -i((1/4)M_{11}\exp(-\gamma W^*) - M_{22}\exp(\gamma W^*)) & (1/4)M_{11}\exp(-\gamma W^*) + M_{22}\exp(\gamma W^*) \end{bmatrix},$$

$$T_{12} = T_{21}^* = \cos(\pi\nu)\exp\frac{i\gamma W^*}{2}\begin{bmatrix} i & -(1/2)\exp(-\gamma W^*) \\ -1 & (i/2)\exp(-\gamma W^*) \end{bmatrix}, \tag{6.5}$$

$$T_{22} = \begin{bmatrix} M_{11} & 0 \\ 0 & M_{22} \end{bmatrix},$$

where M_{11} and M_{22} are the corresponding matrix elements from (5.19).

We also compute the reflection and transmission coefficients

$$R = -i\left[\exp(\gamma W^*) - \frac{1}{4}M_{11}^2\exp(-\gamma W^*)\right]\left[\exp(\gamma W^*) + \frac{1}{4}M_{11}^2\exp(-\gamma W^*)\right]^{-1},$$
$$T = M_{11}\left[\exp(\gamma W^*) + \frac{1}{4}M_{11}^2\exp(-\gamma W^*)\right]^{-1}. \tag{6.6}$$

In the intermediate energy region, the only block matrix element T_{11} requires a special calculation taking into account the contributions from the complex turning points,

$$T_{11} = \begin{bmatrix} \dfrac{\sqrt{2\pi}\exp(-\pi q_2/2)}{\Gamma(1/2 - iq_2)} & i\exp(-\pi q_2) \\ -i\exp(-\pi q_2) & \dfrac{2\Gamma(1/2 - iq_2)\exp(-\pi q_2/2)\cosh(\pi q_2)}{\sqrt{2\pi}} \end{bmatrix}. \tag{6.7}$$

The other matrix elements are the same as in (5.34). Finally, we also find the reflection and the transmission coefficients in the intermediate energy region

$$R = \frac{\exp(-\pi q_2)}{\sqrt{1 + \exp(-2\pi q_2)}}\exp\left[-i\left(\phi - \frac{\pi}{2}\right)\right],$$
$$T = \frac{1}{\sqrt{1 + \exp(-2\pi q_2)}}\exp(-i\phi), \tag{6.8}$$

where $\phi = \arg[\Gamma(1/2 - iq_2)]$.

7. QUANTIZATION RULES FOR CROSSING DIABATIC POTENTIALS

Although instanton trajectories are rather simple objects and can relatively easily be found analytically, calculations of the quantization rules within the instanton approach are rather involved and require the knowledge of the scattering matrix and all the connection matri-

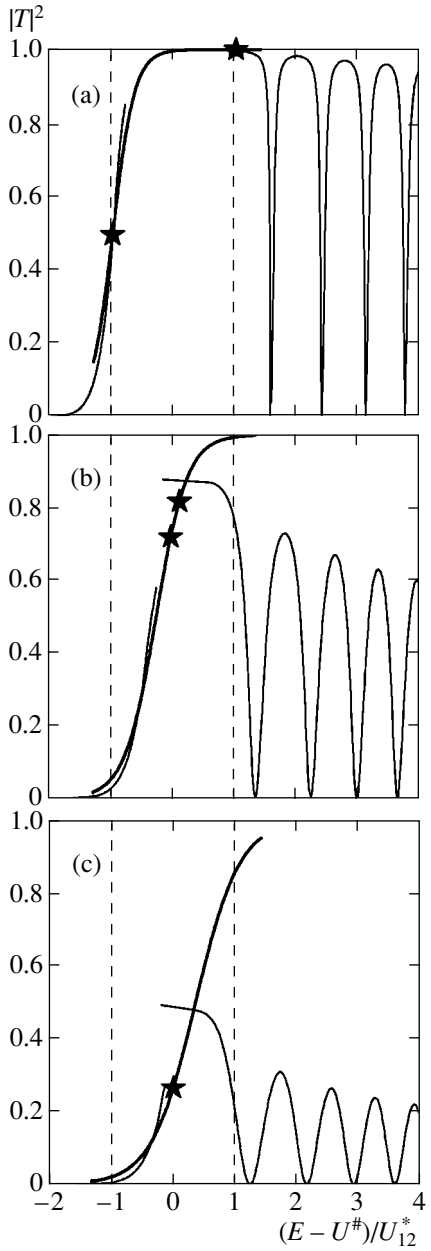

Fig. 8. The T versus E dependence for (a) $U_{12} = U_{12}^*$, (b) $U_{12} = 0.5\,U_{12}^*$, and (c) $U_{12} = 0.25\,U_{12}^*$; stars mark the boundaries of region III', thin lines for the overbarrier and tunneling regions, and bold lines show the results for the intermediate energy region.

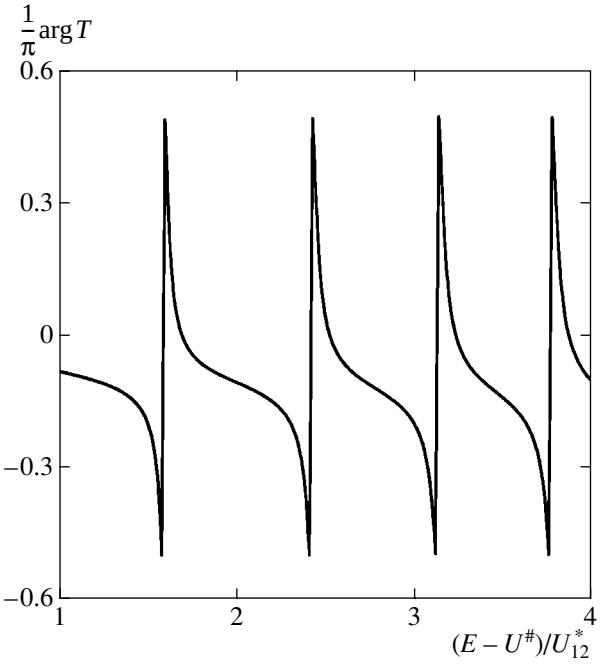

Fig. 9. Transmitted wave phase as a function of E in the overbarrier region at $U_{12} = U_{12}^*$.

ces calculated in the previous sections. In this section, we apply these results to find the quantization rules for the crossing diabatic potentials shown in Fig. 11. Depending on the Massey parameter, the situations shown in the figure exhaust all cases practically relevant for spectroscopy of nonrigid molecules (symmetric or asymmetric double-well and decaying potentials).

Within the instanton approach, the quantization rule can be formulated as the vanishing condition for the amplitudes of the solutions Φ_L^+ and Φ_R^+ that exponentially increase at $X > 0$ and $X < 0$, respectively. Taking into account that $W_L^* = W_R^*$ (the actions in the corresponding wells of the lower adiabatic potential) and using connection matrix (5.19), we obtain the quantization rule

$$\tan(\gamma W_L^*) = \pm\frac{2}{p}\exp(\gamma W_B^*), \tag{7.1}$$

where W_B^* is the action in the barrier formed in the lower adiabatic potential and $p \equiv U_{11}$ is the corresponding matrix element of connection matrix (5.19).

Quantization condition (7.1) differs from the well-known [1] quantization rule for the symmetric double-well potential only by the factor $1/p$ varying from 0 to 1 in the diabatic and adiabatic limits. Therefore, the tunneling splitting at finite values of the Massey parameter ν can be represented as the product

$$\Delta_n = \Delta_n^0 p(\nu) \tag{7.2}$$

of the tunneling splitting Δ_n^0 in the adiabatic potential and the factor

$$p(\nu) = \frac{\sqrt{2\pi}}{\Gamma(\nu)}\gamma^{\nu - 1/2}\exp(-\nu) \tag{7.3}$$

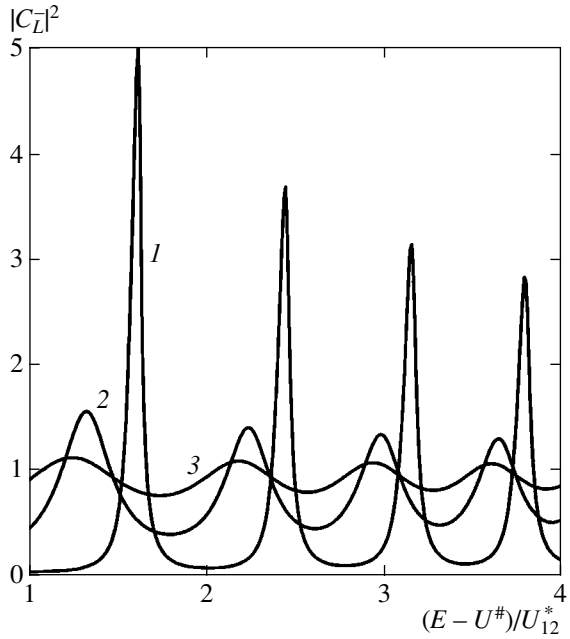

Fig. 10. Amplitudes of the decaying solutions Φ_L^- at $X > 0$ versus E for (*1*) $U_{12} = U_{12}^*$, (*2*) $U_{12} = 0.5U_{12}^*$, and (*3*) $U_{12} = 0.25U_{12}^*$.

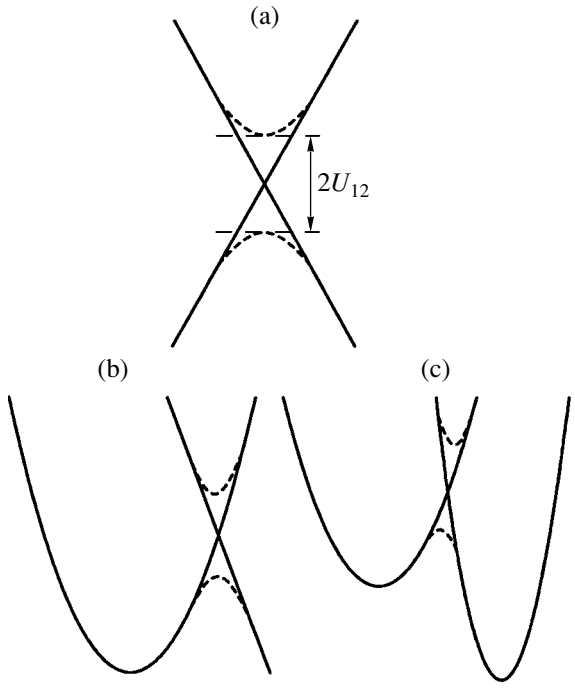

Fig. 11. The diabatic level crossing phenomena: (a) crossing region, (b) bound initial and decay final states, and (c) bound initial and final states.

associated with the transition amplitudes between the diabatic potentials in the crossing region.

It is particularly instructive to consider (7.1) as the standard [1] Bohr–Sommerfeld quantization rule, with both the geometrical φ_n and tunneling χ_n phases included additively in the right-hand side. In the adiabatic limit $p(\nu) \longrightarrow 1$, we find that $\varphi_n \longrightarrow 0$ and (7.1) reduces to the quantization of the symmetric double-well potential. In the diabatic limit, $\varphi_n = -\chi_n$ and the geometric phase compensates the tunneling one. The physical argument leading to this compensation can easily be rationalized as follows. At the reflection at the crossing point $X = 0$, the trajectories in the classically forbidden energy region are the same as those for the tunneling region but with the phase shift π.

We now focus on quantization rules for the overbarrier energy region. Closely following the above analysis for the tunneling region (replacing connection matrix (5.19) by matrix (5.22) and making some other self-evident replacements), after some tedious algebra, we finally obtain the quantization rule

$$\begin{aligned}(1 - \exp(-2\pi\nu))\cos(2\gamma W_L^* + (\phi - \phi_0))\\ \times \cos(\gamma W^* - (\phi - \phi_0))\\ + \exp(-2\pi\nu)\cos^2\left(\gamma W_L^* + \frac{\gamma W^*}{2}\right) = 0,\end{aligned} \tag{7.4}$$

where W^* is the action in the well formed by the upper adiabatic potential and $\phi - \phi_0 = \tilde{\chi}$ is determined from (5.23). Equation (7.4) implies that the eigenstates are determined by the parameter

$$B = \frac{\exp(-2\pi\nu)}{1 - \exp(-2\pi\nu)}. \tag{7.5}$$

In the diabatic limit $\nu \longrightarrow 0$, and hence $B \longrightarrow 1/(2\pi\nu)$, the main contribution to (7.4) is due to the second term which leads to a splitting of degenerate levels in the diabatic potentials. Moreover, because

$$\begin{aligned}\gamma\left(W_L^* + \frac{W^*}{2}\right) = \pi\left\{n + \frac{1}{2}\right.\\ \left. \pm 2\nu\sin\left[\gamma\left(W_L^* + \frac{W^*}{2}\right) - \phi + \phi_0\right]\right\},\end{aligned} \tag{7.6}$$

the splitting increases as the Massey parameter ν increases; the splitting is an oscillating function of the interaction U_{12}.

In the adiabatic limit, as $\nu \longrightarrow \infty$, $\phi - \phi_0 \longrightarrow 0$, and, therefore, $B \approx \exp(-2\pi\nu)$ in accordance with (7.5), the main contribution to (7.4) comes from the first term, which determines the quantization rule for the upper one-well potential and for the lower double-well potential in the overbarrier energy region. In this limit, the parameter B plays the role of the tunneling transition matrix element. For B smaller than the nearest level spacings for the lower and upper potentials, we can find

two sets of quantization rules from (7.4) that lead to two sets of independent energy levels

$$\gamma W^* = \pi\left(n_1 + \frac{1}{2}\right), \quad 2\gamma W_L^* = \pi\left(n_2 + \frac{1}{2}\right). \qquad (7.7)$$

Because the eigenstate energy level displacements depend on U_{12}, resonances can occur at certain values of this parameter, where the independent quantization rules in (7.7) are not correct any more. The widths of these resonances are proportional to $\exp(-2\pi\nu)$ and are therefore strongly diminished as the Massey parameter ν increases. This behavior is easily understood, because the wave functions of the excited states for the lower potential are delocalized in the limit, and their amplitudes in the localization regions for the low-energy states of the upper potential are very small.

A more complicated problem is to derive the quantization rule in the intermediate energy region. We must use connection matrix (5.35) and take the contributions from the imaginary turning points into account. Nevertheless, the quantization rule can finally be written in the simple and compact form

$$\cos(2\gamma W_L^*) = -\exp(-\pi q_2), \qquad (7.8)$$

where $q_2 = \gamma(\nu - \alpha)/2$ is determined by (5.27).

It is useful to illustrate the essence of the general result given above by simple (but nontrivial) examples. We first consider two identical parabolic potentials with their minima at $X = \pm 1$ and with a coupling that does not depend on X. Because of the symmetry, solutions of the Schrödinger equation in this case can be represented as symmetric and antisymmetric combinations of the localized functions

$$\Psi^{\pm} = \frac{1}{\sqrt{2}}(\Phi_L \pm \Phi_R). \qquad (7.9)$$

The functions are orthogonal, and, in addition, the two sets of functions (Ψ_e^+, Ψ_0^-) and (Ψ_0^+, Ψ_e^-) (where the respective subscripts 0 and e denote the ground and the first excited states) correspond to the two possible types of level crossings.

In Fig. 12, we schematically depict the dependence of the positions of levels on the coupling U_{12}. In the energy region $E \le U^* + U_{12}$, where only discrete levels of the lower adiabatic potentials exist, there are pairs of the alternating parity levels (Ψ_e^+, Ψ_0^-) and (Ψ_0^+, Ψ_e^-). The tunneling splittings increase monotonically, because the Massey parameter ν increases, and the barrier decreases with U_{12}. The same level and parity classification remains correct for the energy region above the barrier of the lower adiabatic potential, where the spectrum becomes almost equidistant. However, in the overbarrier region, the resonances occur between levels of the same parity; the sequence of the odd and of the even levels is broken, and level displacements are not monotonic functions of U_{12}. Some of the levels of different parities can pairwise cross. For the upper adiabatic potential, the level sequence is opposite to that for the lower adiabatic potential. We have checked the results of our semiclassical approach and found a remarkably good agreement with the numerical quantum diagonalization.

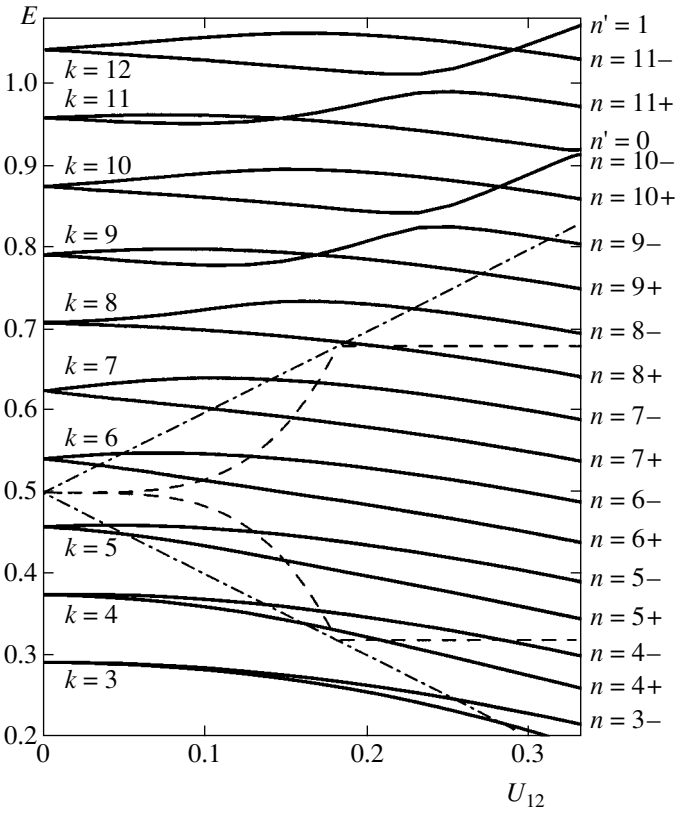

Fig. 12. Level displacements versus U_{12} for two diabatic crossing potentials $(1 \pm X)^2/2$. Dashed lines show the intermediate energy region, dotted–dashed lines show displacements for the top and for the bottom of the adiabatic potentials. k, n, and n' are quantum numbers for the diabatic, lower adiabatic, and upper adiabatic potentials.

The second instructive example involves the crossing of one-well and linear diabatic potentials. This leads to the lower adiabatic decay potential and to the upper one-well adiabatic potential. The quantization rules then correspond to the vanishing amplitudes for the exponentially increasing solutions as $X \longrightarrow -\infty$; in addition, we must require that no waves propagate from the region of infinite motion, i.e., the region $X > 1/2$. Performing the same procedure as above, we find that in the tunneling energy region, the eigenstates are the roots of the equation

$$\tan(\gamma W_L^*) = -i\frac{4}{p^2(\nu)}\exp(2\gamma W_B^*), \qquad (7.10)$$

with the same notation as above.

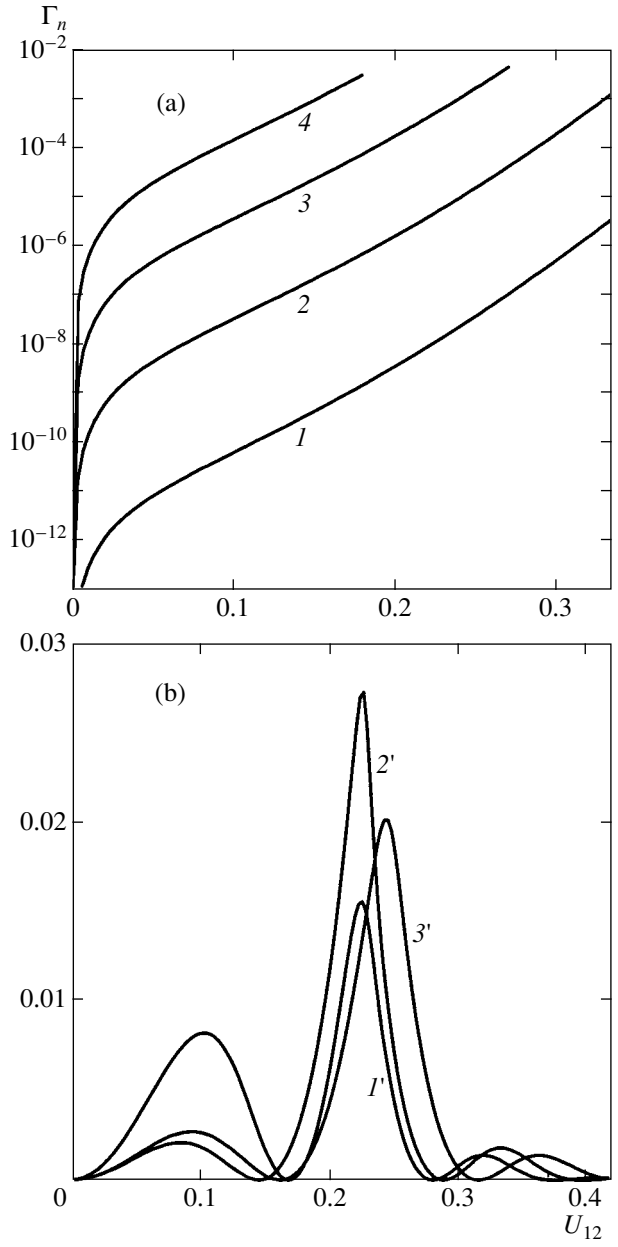

Fig. 13. Γ_n vs. U_{12} for the quasistationary states at the diabatic potentials $(1 \pm X)^2/2$ and $1/2 - X$ crossing; (a) *1–4* are the level energies 0.042, 0.125, 0.208, and 0.292 for the lower adiabatic potential, (b) *1*'–*3*' are the level energies 0.625, 0.708, and 0.792 for the upper adiabatic potential.

To proceed further, it is convenient to introduce a complex action to describe quasistationary states,

$$\gamma W_L^* = \pi\left(\frac{E_n}{\Omega} - i\frac{\Gamma_n}{2\Omega}\right), \tag{7.11}$$

where $\Omega = \partial W_L/\partial E$ is evidently independent of E. The real and imaginary parts of the quantized eigenstates determined from (7.11) are given by

$$E_n = \Omega\left(n + \frac{1}{2}\right),$$
$$\Gamma_n = p^2(\nu)\frac{\Omega}{2\pi}\exp(-2\gamma W_B^*). \tag{7.12}$$

This relation describes the nonadiabatic tunneling decay of quasistationary states of the lower adiabatic potential. Similarly to the case with the crossing of two parabolic potentials (Eq. (7.2)), the tunneling and the adiabatic factors here enter the decay rate multiplicatively. Because the decay rate is proportional to the square of the tunneling matrix element, we have $\Gamma_n \propto p^2(\nu)$, as it should be.

In the overbarrier energy region, the quantization rule is

$$(1 - \exp(-2\pi\nu)\exp[-i(\gamma W_L^* + \phi - \phi_0)]$$
$$\times \cos(\gamma W^* - \phi + \phi_0) + \exp(-2\pi\nu)$$
$$\times \exp\left(-\frac{i\gamma W^*}{2}\right)\cos\left(\gamma W_L^* + \frac{\gamma W^*}{2}\right) = 0, \tag{7.13}$$

and the actions depend on the energy E as

$$\gamma W_L^* = \pi\frac{E}{\Omega}, \quad \gamma W = \pi\left(-\frac{U^* + U_{12}}{\Omega_1} + \frac{E}{\Omega_1}\right), \tag{7.14}$$

where Ω and Ω_1 are E-dependent frequencies of the diabatic and the upper adiabatic potentials.

In the diabatic limit, the decay rate is proportional to the Massey parameter and is given by

$$\Gamma_n \approx \pi\nu\cos^2(\gamma W - \phi + \phi_0). \tag{7.15}$$

In the opposite, adiabatic limit, the decay rate is

$$\Gamma_n \approx \exp(-2\pi\nu)[1 - \sin(2\gamma W_L^* + \phi - \phi_0)]. \tag{7.16}$$

In both limits, the decay rate is an oscillating function of U_{12}. We illustrate the dependence $\Gamma(U_{12})$ for the crossing diabatic potentials $U_1 = (1 + X)^2/2$ and $U_2 = 1/2 - X$ in Fig. 13. We note that while the tunneling decay rate of low-energy states increases monotonically with the Massey parameter ν, the decay rate of highly excited states tends to zero in both (diabatic and adiabatic) limits. There are certain characteristic values of U_{12} at which the right-hand side of (7.15) or (7.16) vanishes, and, therefore, $\Gamma_n = 0$.

The last, more general example that we consider in this section describes two nonsymmetric potentials crossing at $X = 0$,

$$U_1 = \frac{1}{2}(1 + X)^2, \quad U_2 = \frac{1}{2b}(X^2 - 2bX + b). \tag{7.17}$$

In a certain sense, this is the generic case, and as the parameter b entering potential (7.17) varies from 1 to

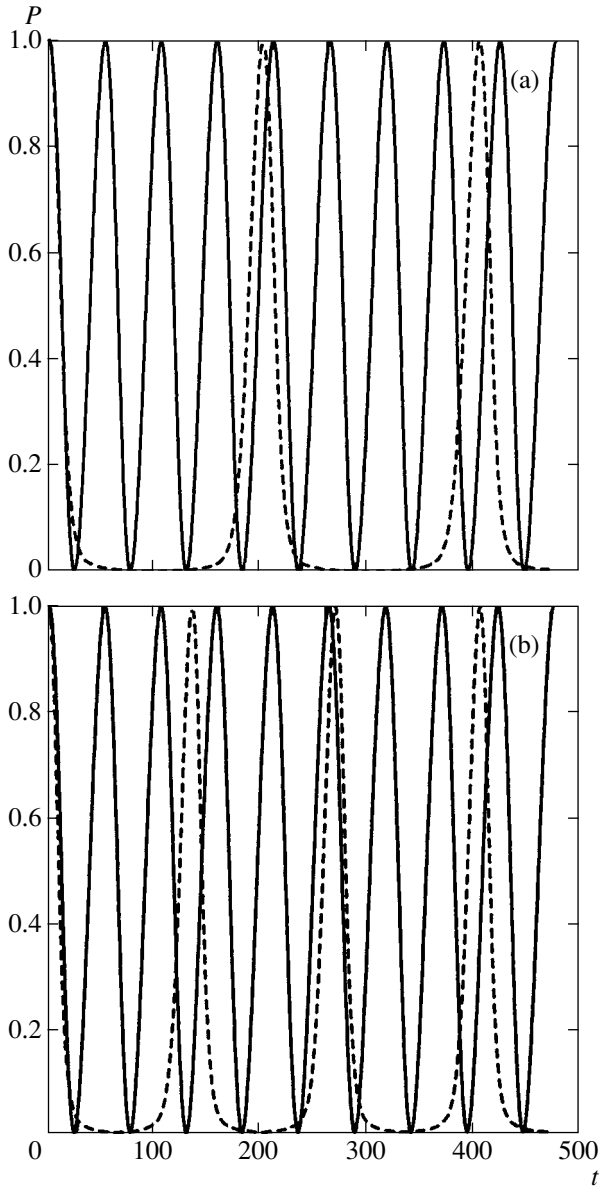

Fig. 14. Survival probability for the localized $n = 0$ state; (a) $b = 1500$, dashed lines $U_{12} = 0.15$, solid lines $U_{12} = 0.21$; (b) $b = 1500$, dashed lines $U_{12} = 0.28$, solid lines $U_{12} = 0.21$.

∞, we recover the two particular examples considered above and pass from two identical parabolic potentials to the crossing of the one-well and linear diabatic potentials. Potentials U_2 of this type were recently investigated by two of the authors (V.B. and E.K.) [64] with the aim of studying the crossover behavior from coherent to incoherent tunneling with the increase of the parameter b; the larger b is, the larger the density of the final states becomes. The criterion for coherent–incoherent crossover behavior found in [64] is based on comparison of the transition matrix elements and the interlevel spacings in the final state. A similar criterion should hold for the LZ level crossing problem, but the tunneling transition matrix elements must then be multiplied by the small adiabatic factor. Therefore, the coherent–incoherent tunneling crossover region moves to the denser density of final states, and the larger U_{12} is, the smaller the region for incoherent tunneling becomes.

A totally different situation occurs for highly excited states. In the diabatic limit, the transition matrix element increases with the Massey parameter ν, and, therefore, at a given b value, the system moves to more incoherent behavior. In the adiabatic limit, the transition matrix element is exponentially small and coherence of the interwell transitions should be restored. However, because the matrix elements are oscillating functions of U_{12} for the intermediate range of this coupling, coherent–incoherent tunneling rates are also nonmonotonically varying functions. These unusual phenomena are illustrated in Fig. 14, where we show time dependence of the survival probability P for the initially prepared state $n = 0$ localized in the left well.

8. CONCLUSION

We have reconsidered a very basic subject, the LZ problem. Currently, about 100 publications per year are related to the LZ problem. Clearly, it is impossible to give a complete analysis of the achievements in this field. Our aim was therefore only to show some recent trends and our new results, to help beginners and experts find cross-references between the many physical phenomena related to the LZ problem. The problem was first addressed long ago, and many results, already classic, are now known from textbooks [1, 37]. Although exact quantum-mechanical calculations are still prohibitively difficult, many important results have been obtained in the framework of the WKB approach [1–65]. The accuracy of the modified WKB methods can be improved considerably; we note, e.g., [30], where the standard WKB was extended by the inclusion of a special type of trajectories in the complex phase plane such that the semiclassical motion along these trajectories is described by the Weber functions. This method, ascending to Landau [1], is equivalent to the appropriate choice of the integration path around the turning point. It appears to be quite accurate for the tunneling and overbarrier regions, where the characteristic fourth-order polynomial (see (4.16)) can be reduced to a second-order polynomial (two pairs of roots are nearly degenerate). However, even in this case, some corrections have been found in [23–25] that cannot be neglected. In the intermediate energy region, where all four roots are noticeably different, the method becomes invalid. In addition, the choice of these additional special trajectories (which must be included to improve the accuracy of the WKB method near the barrier top) depends on a detailed form of the potential far from the top, and, therefore, a nonuniversal procedure is to be performed from the very beginning in each particular case.

We believe we are the first to explicitly address the problem of the behavior in the intermediate energy region. In all previous publications, this region was considered as a very narrow and insignificant one, or at most, the results were obtained by a simple interpolation from the tunneling region (with a monotonic decay of the transition probability) to the overbarrier region (with oscillating behavior). The fact is that classical trajectories can be separated into two classes, "localized" and "delocalized," in the following sense. If the energy is sufficiently close to the minimum or maximum of the potentials, the trajectories can be called confined, because they are determined by the universal features of the potentials in the vicinity of these extremal points. Evidently, this is not the case in the intermediate energy region. In this paper, we have found that contrary to the common belief, the instanton trajectory is a rather simple object and can be explicitly computed even for the intermediate energy region.

Within the framework of the instanton approach, we present a full and unified description of the 1D LZ problem, which can very often be quite a reasonable approximation for real systems. Because different approaches have been proposed to study the LZ problem, we develop a uniform and systematic procedure for handling the problem. We reproduced all the known results for tunneling and overbarrier regions and studied the intermediate energy region. Specifically, we applied our approach to the Born–Oppenheimer scheme, formulated the instanton method in the momentum space, and presented all the details of the LZ problem for two electronic states also using the instanton description of the LZ problem in the coordinate space. Neglecting higher order space derivatives, we found asymptotic solutions; using the adiabatic–diabatic transformation, we then matched the solutions in the intermediate region. Based on these results, we derived the complete scattering matrix for the LZ problem, the quantization rules for crossing diabatic potentials. Our results can be applied to several models of level crossings that are relevant in the interpretation and description of experimental data on spectroscopy of nonrigid molecules and on other systems undergoing crossing and relaxation phenomena.

We also note that in spite of the sufficiently long history of the LZ phenomena, study of it is still in an accelerating stage, and a number of questions remain to be clarified (we mention only several new features of the phenomena that have recently attracted attention, such as the LZ interferometry for qubits [74], LZ theory for Bose–Einstein condensates [75], and multiparticle and multilevel LZ problems [76–79]). Much of the excitement arises from the possibility of discovering novel physics beyond the semiclassical paradigms discussed here. For example, we found in Sections 2 and 3 that the wave functions of nuclei moving along periodic orbits acquire geometric phases (the effect is analogous to the Aharonov–Bohm effect [38], but is related not to external magnetic fields, but to nonadiabatic interactions). The relation between the two phenomena, the geometric phases, and the periodic orbits can be established using the Lagrangian (instead of Hamiltonian) formulation of the problem, which enables one to take into account explicitly the time dependence of the adiabatic process under consideration, using propagator technique [34–36] (also see, e.g., [4, 43]). Proper handling of these aspects is beyond the scope of our work, however. Further experimental and theoretical investigations are required for revealing the detailed microscopic and macroscopic properties of different LZ systems.

In the fundamental problems of chemical dynamics and molecular spectroscopy, transitions from the initial to final states can be treated as a certain motion along the potential energy surfaces of the system under consideration. These surfaces are usually determined within the Born–Oppenheimer approximation (see Section 2). However, the approximation becomes inadequate for the excited vibrational states when their energies are on the order of the electronic interlevel energy spacing or near the dissociation limit. In both cases, nonadiabatic transitions should be taken into account and most of the nonradiative processes occur owing to this nonadiabaticity. Typical examples investigated in [80] are the so-called predissociation, singlet–triplet or singlet–singlet conversion, and vibrational relaxation phenomena.

Slow atomic collisions provide other examples of nonadiabatic transitions between electronic states, where the time dependence of the states is determined by distance and by the relative velocity of the colliding particles [33]. Some examples of nonadiabatic transitions relevant in semiconductor physics can be found in [81], those pertaining to nuclear or elementary particle physics in [82], and those relevant in laser or nonlinear optic physics in [83–86]. The latter topic is interesting not only in its own right, but also as an illustration of novel and fundamental quantum effects related to the LZ model. The off-diagonal electronic state interactions arise from the dipole forces in this case. For relatively short laser pulses, this leads to the time-dependent LZ problem for two electronic states, considered in our paper in detail (also see the laser optic formulation in [83–85]). The probability of finding the system in the upper state after a single resonant passage can be computed in the framework of the LZ model. This is related to one important aspect of the LZ problem, namely, dissipative and noisy environments. When external actions (e.g., fields) driving LZ transitions are reversed from large negative to large positive values, the dissipation reduces tunneling and the system remains in the ground state, or in other words, the thermal excitation from the ground state to the excited state suppresses such adiabatic transitions. However, for the field swept from the resonance point, the tunneling probability becomes larger in the presence of dissipation (see, e.g., [67]). The increasing precision of experimental tests in the femtosecond laser pulse range

enables one to excite well-defined molecular states and to study their time evolution using the second probing laser beam [17].

ACKNOWLEDGMENTS

This paper was supported in part by grants of the Russian Foundation for Basic Research. One of us (E.K.) is indebted to INTAS (grant no. 01-0105) for partial support.

REFERENCES

1. L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed. (Nauka, Moscow, 1989; Pergamon, New York, 1977).
2. E. E. Nikitin, Opt. Spektrosk. **13**, 761 (1962); Discuss. Faraday Soc. **33**, 14 (1962).
3. Yu. N. Demkov, Zh. Éksp. Teor. Fiz. **45**, 195 (1963) [Sov. Phys. JETP **18**, 138 (1964)].
4. G. Hertzberg and H. C. Longuet-Higgins, Discuss. Faraday Soc. **35**, 77 (1963).
5. E. E. Nikitin, Chem. Phys. Lett. **2**, 402 (1968).
6. M. S. Child, Adv. At. Mol. Phys. **14**, 225 (1978).
7. H.-W. Lee and T. F. George, Phys. Rev. A **29**, 2509 (1984).
8. S. Griller and C. Gonera, Phys. Rev. A **63**, 052101 (2001).
9. A. M. Dykhne, Zh. Éksp. Teor. Fiz. **38**, 570 (1960) [Sov. Phys. JETP **11**, 411 (1960)].
10. A. M. Dykhne, Zh. Éksp. Teor. Fiz. **41**, 1324 (1961) [Sov. Phys. JETP **14**, 941 (1961)].
11. A. M. Dykhne and A. V. Chaplik, Zh. Éksp. Teor. Fiz. **43**, 889 (1962) [Sov. Phys. JETP **16**, 631 (1962)].
12. G. V. Dubrovskiĭ, Zh. Éksp. Teor. Fiz. **46**, 863 (1964) [Sov. Phys. JETP **19**, 591 (1964)].
13. L. P. Kotova, Zh. Éksp. Teor. Fiz. **55**, 1375 (1968) [Sov. Phys. JETP **28**, 719 (1968)].
14. P. Pechukas, T. F. George, and K. Morokuma, J. Chem. Phys. **64**, 1099 (1976).
15. J. P. Davis and P. Pechukas, J. Chem. Phys. **64**, 3129 (1976).
16. J.-T. Hwang and P. Pechukas, J. Chem. Phys. **67**, 4640 (1977).
17. B. M. Garraway and S. Stenholm, Phys. Rev. A **45**, 364 (1992).
18. K.-A. Suominen and B. M. Garraway, Phys. Rev. A **45**, 374 (1992).
19. N. V. Vitanov and B. M. Garraway, Phys. Rev. A **53**, 4288 (1996).
20. N. V. Vitanov, Phys. Rev. A **59**, 988 (1999).
21. N. V. Vitanov and K.-A. Suominen, Phys. Rev. A **59**, 4580 (1999).
22. J. B. Delos, W. R. Thorson, and S. K. Knudson, Phys. Rev. A **6**, 709 (1972).
23. C. Zhu, H. Nakamura, N. Re, and V. Aquilanti, J. Chem. Phys. **97**, 1892 (1992); C. Zhu and H. Nakamura, J. Chem. Phys. **97**, 8497 (1992).
24. C. Zhu and H. Nakamura, J. Chem. Phys. **98**, 6208 (1993).
25. C. Zhu and H. Nakamura, J. Chem. Phys. **101**, 4855 (1994); J. Chem. Phys. **101**, 10630 (1994).
26. V. A. Benderskii, D. E. Makarov, and C. A. Wight, *Chemical Dynamics at Low Temperatures* (Wiley, New York, 1994).
27. V. A. Benderskii, E. V. Vetoshkin, and H. P. Trommsdorf, Chem. Phys. **244**, 273 (1999).
28. V. A. Benderskii and E. V. Vetoshkin, Chem. Phys. **257**, 203 (2000).
29. V. A. Benderskii, E. V. Vetoshkin, and E. I. Kats, Zh. Éksp. Teor. Fiz. **122**, 746 (2002) [JETP **95**, 645 (2002)].
30. V. L. Pokrovskii and I. M. Khalatnikov, Zh. Éksp. Teor. Fiz. **40**, 1713 (1961) [Sov. Phys. JETP **13**, 1207 (1961)].
31. B. K. Bykhovskii, E. E. Nikitin, and M. Ya. Ovchinnikova, Zh. Éksp. Teor. Fiz. **47**, 750 (1965) [Sov. Phys. JETP **20**, 500 (1964)].
32. V. M. Akulin and W. P. Schleich, Phys. Rev. A **46**, 4110 (1992).
33. E. E. Nikitin and S. Ya. Umanskii, *Theory of Slow Atomic Collisions* (Atomizdat, Moscow, 1979; Springer, Berlin, 1984), Springer Ser. Chem. Phys., Vol. 30.
34. W. H. Miller and T. F. George, J. Chem. Phys. **56**, 5637 (1972).
35. R. K. Preston, C. Sloane, and W. H. Miller, J. Chem. Phys. **60**, 4961 (1974).
36. F. J. McLafferty and T. G. George, J. Chem. Phys. **63**, 2609 (1975).
37. J. C. Slater, *Quantum Theory of Molecules and Solids*, Vol. 1: *Electronic Structure of Molecules* (McGraw-Hill, New York, 1963).
38. Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).
39. C. A. Mead, J. Chem. Phys. **78**, 807 (1983).
40. M. V. Berry, Proc. R. Soc. London, Ser. A **392**, 451 (1984).
41. M. Wilkinson, J. Phys. A **17**, 3459 (1984).
42. M. V. Berry, J. Phys. A **18**, 15 (1985).
43. H. Kuratsui and S. Ida, Prog. Theor. Phys. **74**, 439 (1985).
44. M. V. Berry, J. Mod. Opt. **34**, 1401 (1987).
45. J. Vidal and J. Wudka, Phys. Rev. A **44**, 5383 (1991).
46. A. Mustafazadeh, Phys. Rev. A **55**, 1653 (1997).
47. R. F. Fox and P. Jung, Phys. Rev. A **57**, 2339 (1998).
48. C. A. Mead, Rev. Mod. Phys. **64**, 51 (1992).
49. Y. Aharonov and J. Anandan, Phys. Rev. Lett. **58**, 1593 (1987).
50. J. Samuel and R. Bhachdari, Phys. Rev. Lett. **60**, 2339 (1988).
51. N. Mukunda and R. Simon, Ann. Phys. (N.Y.) **228**, 20 (1993).
52. M. V. Berry and J. M. Robbins, Proc. R. Soc. London, Ser. A **442**, 659 (1993).
53. J. Moody, A. Shapere, and F. Wilczek, Phys. Rev. Lett. **56**, 893 (1986).
54. D. Suter, G. C. Chingas, R. A. Harris, and A. Pines, Mol. Phys. **61**, 1327 (1987).
55. F. Gaitan, Phys. Rev. A **58**, 1665 (1998).

56. M. Baer, S. H. Lin, A. Alijah, *et al.*, Phys. Rev. A **62**, 032506 (2000).
57. J. Heading, *An Introduction to Phase-Integral Methods* (Wiley–Interscience, London, 1962).
58. A. M. Polyakov, Nucl. Phys. B **129**, 429 (1977).
59. S. Coleman, *Aspects of Symmetry* (Cambridge Univ. Press, Cambridge, 1985).
60. M. S. Child, Mol. Phys. **20**, 171 (1971).
61. M. S. Child, J. Mol. Spectrosc. **53**, 280 (1974).
62. T. Holstein, Philos. Mag. **37**, 49 (1978).
63. G. Stick and M. Thoss, Phys. Rev. Lett. **78**, 578 (1997).
64. V. A. Benderskii and E. I. Kats, Phys. Rev. E **65**, 036217 (2002).
65. N. T. Maintra and E. J. Heller, Phys. Rev. A **54**, 4763 (1996).
66. Y. Kayanuma and H. Nakayama, Phys. Rev. B **57**, 13099 (1998).
67. K. Saito and Y. Kayanuma, Phys. Rev. A **65**, 033407 (2002).
68. M. V. Fedoryuk, Dokl. Akad. Nauk SSSR **158**, 540 (1964).
69. M. V. Fedoryuk, Dokl. Akad. Nauk SSSR **162**, 287 (1965).
70. M. V. Fedoryuk, Usp. Mat. Nauk **21**, 1 (1966).
71. A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1953–1955; Nauka, Moscow, 1965–1967), Vols 1–3.
72. F. W. J. Olver, *Asymptotics and Special Functions* (Academic, New York, 1974; Nauka, Moscow, 1990).
73. F. W. J. Olver, J. Res. Natl. Bur. Stand., Sect. B **63**, 131 (1959).
74. A. V. Shytov, D. A. Ivanov, and M. V. Feigelman, cond-mat/0110490.
75. V. A. Yurovsky, A. Ben-Reuven, and P. S. Julienne, Phys. Rev. A **65**, 043607 (2002).
76. V. L. Pokrovsky and N. A. Sinitsyn, Phys. Rev. B **65**, 153105 (2002).
77. N. A. Sinitsyn, cond-mat/0212017.
78. Y. N. Demkov and V. N. Ostrovsky, Phys. Rev. A **61**, 032705 (2000).
79. D. A. Garanin and E. M. Chudnovsky, Phys. Rev. B **65**, 094423 (2002).
80. H. Eyring, S. H. Lin, and S. M. Lin, *Basic Chemical Kinetics* (Wiley, New York, 1980; Mir, Moscow, 1983).
81. Yu. A. Bychkov and A. M. Dykhne, Zh. Éksp. Teor. Fiz. **48**, 1174 (1965) [Sov. Phys. JETP **21**, 783 (1965)].
82. S. Toshev, Phys. Lett. B **198**, 551 (1987).
83. J. M. Juan and T. F. George, J. Chem. Phys. **68**, 3040 (1978).
84. A. D. Bandrauk and G. Turcott, J. Chem. Phys. **77**, 3867 (1982).
85. H. W. Lee and T. F. George, Phys. Rev. A **35**, 4977 (1987).
86. V. M. Akulin and N. V. Karlov, *Intense Resonant Interactions in Quantum Electronics* (Springer, Berlin, 1992).

Journal of Experimental and Theoretical Physics, Vol. 97, No. 2, 2003, pp. 259–278.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 2, 2003, pp. 288–309.

PLASMA, GASES

Pressure-Produced Ionization of Nonideal Plasma in a Megabar Range of Dynamic Pressures

V. E. Fortov[a], V. Ya. Ternovoĭ[a], M. V. Zhernokletov[b], M. A. Mochalov[b], A. L. Mikhailov[b], A. S. Filimonov[a], A. A. Pyalling[a], V. B. Mintsev[a,*], V. K. Gryaznov[a,], and I. L. Iosilevskiĭ[c]**

[a]*Institute for Chemical Physics Problems, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*

**e-mail: minv@icp.ac.ru*

***e-mail: grvkb@ficp.ac.ru*

[b]*Russian Federal Nuclear Center All-Russian Research Institute of Experimental Physics, Sarov, Nizhni Novgorod oblast, 607200 Russia*

[c]*Moscow Institute for Physics and Technology (State University), Dolgoprudnyĭ, Moscow oblast, 141200 Russia*

Received January 29, 2003

Abstract—The low-frequency electrical conductivity of strongly nonideal hydrogen, helium, and xenon plasmas was measured in the megabar range of pressures. The plasmas in question were generated by the method of multiple shock compression in planar and cylindrical geometries, whereby it was possible to reduce effects of irreversible heating and to implement a quasi-isentropic regime. As a result, plasma states at pressures in the megabar range were realized, where the electron concentration could be as high as $n_e \approx 2 \times 10^{23}$ cm^{-3}, which may correspond to either a degenerate or a Boltzmann plasma characterized by a strong Coulomb ($\Gamma_D = 1$–10) and a strong interatomic ($\Gamma_a = r_a n_a^{1/3} \sim 1$) interaction. A sharp increase (by three to five orders of magnitude) in the electrical conductivity of a strongly nonideal plasma due to pressure-produced ionization was recorded, and theoretical models were invoked to describe this increase. Experimental data available in this region and theoretical models proposed by various authors are analyzed. The possibility of a first-order "phase transition" in a strongly nonideal plasma is indicated. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The behavior of plasma, which is the most widespread state of matter in nature, under the conditions of strong heating and compression is of considerable interest from the general physical point of view; it is also of practical interest for astrophysics, the physics of giant planets, and promising applications in power engineering [1–3]. Particular attention is being given to the ionization composition of a plasma, since this provides a basis for calculating its thermodynamic, transport, and optical properties.

It is well known that plasma can be obtained not only via strong heating up to temperatures commensurate with the ionization potential, $k_B T \sim I$, but also via a strong compression to a state in which the interparticle spacing becomes commensurate with atomic sizes, $r_a \sim n_a^{-1/3}$; the second way is referred to as cold ionization or pressure-produced ionization. While thermal-ionization processes have to date received quite adequate study [1], investigation of pressure-produced ionization is much more complicated since one deals here with a cold ($k_B T \ll I$) compression of a plasma to pressures in the megabar range and densities that considerably exceed solid-state values. Under such conditions, the interaction between particles becomes strong (nonideality), the electron shells of atoms and molecules overlap, and the typical level of electrical conductivity is commensurate with that in metals. Frequently, such a regime is erroneously called metallization, even though Landau and Zeldovich [4], Mott and Davis [5], and Hensel and Frank [6] showed that a metal can be distinguished from a dielectric only by their electron spectra at $T = 0$ but not by the level of the electrical conductivity itself. By way of example, we indicate that dilute tokomak plasmas ($n_e \sim n_i \sim 10^{14}$ cm^{-3}, $T \approx 5$–10 keV) have an electrical conductivity close to that of pure copper [7].

Implementing the isothermal expansion of low-boiling metals at supercritical pressures [6, 8], one can continuously pass from a high-conducting metallic state to low-conducting gaseous dielectric states and establish that a metal–dielectric transition occurs in a narrow range of densities that are close to (Cs, Rb, K) or somewhat greater (Hg) than the matter density at the critical point. For the majority of other metals, which constitute

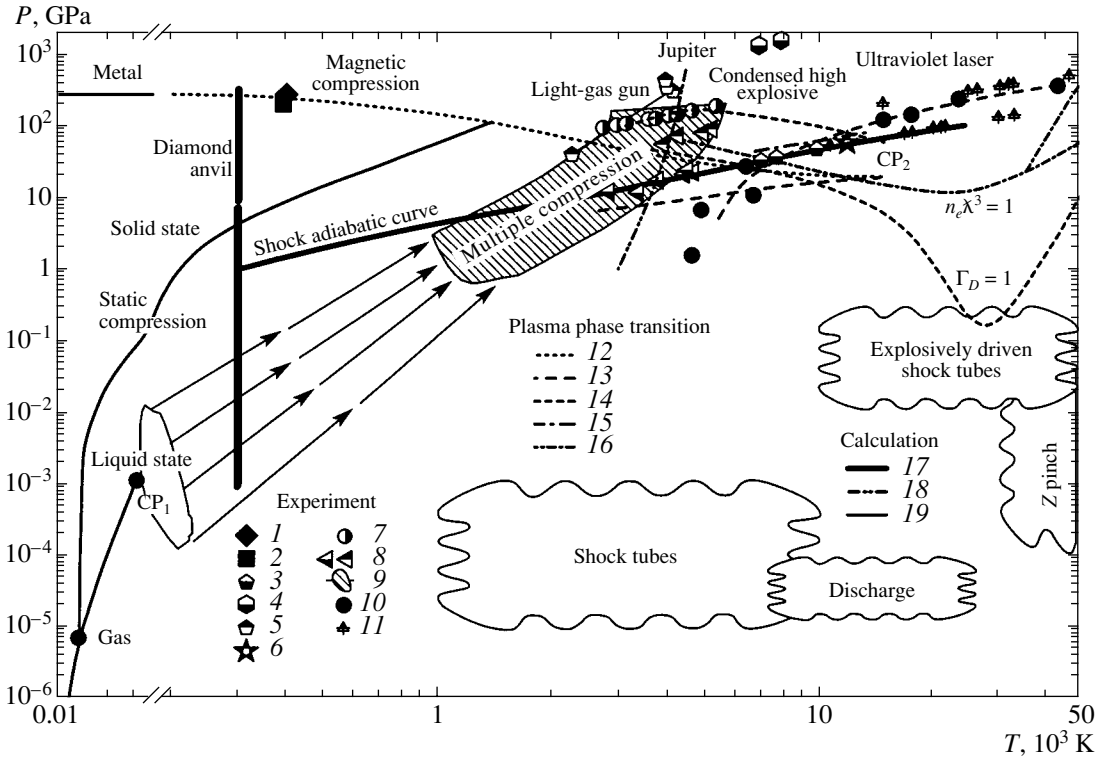

Fig. 1. Phase diagram of hydrogen. The displayed experimental data were obtained in (*1*, *2*) magnetic compression [29, 30], (*3*) *Z* pinch [33], (*4*, *5*) cylindrical compression [26, 27], (*6*) spherical compression [28]; (*7*, *8*) single and multiple compression by means of a light-gas gun [34, 35], (*9*) multiple shock compression [36], and (*10*, *11*) shock compression by a laser [31, 32]. The estimates for the critical point of the plasma phase transition in hydrogen were taken from the articles of (*12*) Beule *et al.* [37], (*13*) Robnic and Kundt [38], (*14*) Saumon and Chabrier [23], (*15*) Haronska *et al.* [39], and (*16*) Mulenko *et al.* [24]. The calculated data correspond to (*17*) compression at a diamond anvil [11], (*18*) the parameters of Jupiter's atmosphere [40], and (*19*) the adiabatic curve for the shock compression of hydrogen [34].

80% of the elements of the periodic table, critical temperatures and pressures are extremely high and are inaccessible to methods of static experiments. On the basis of their recent experiment, DeSilva and Katsouros [9], who were able to reach supercritical pressures via a fast electric explosion of metallic conductors, pointed out that solid-state metals loose their metallic conductivity upon a more significant expansion (by a factor of 5 to 7).

A method that employs the adiabatic expansion of matter preliminarily compressed by strong shock waves to pressures of the megabar range makes it possible to explore a wide region of the phase diagram of metals, including the vicinity of the phase-transition point [3, 10]. Measurements of thermodynamic (Cu, Pb, Bi, Fe, U) and radiative (Bi) properties of some metals according to this procedure permitted determining the equation of state, the coefficient of absorption, and the parameters of the critical points for the metals under study, but they did not confirm the hypothesis [4] that there occur plasma phase transitions caused by the dielectrization of metals in the supercritical region.

A considerable number of studies (see [11] and references therein) motivated by searches for metallic hydrogen [11–13] in connection with its possible high-temperature superconductivity in a metastable medium [14] have been devoted to the metallization ($T = 0$) of dielectrics at high pressures. Estimates of the metallization pressure P^* that were obtained for various substances by methods of the band theory of solids fall within the megabar {$P^*(H_2) \approx 3$ Mbar [11–16], $P^*(Xe) \approx 1.5$ Mbar [17]) and ultramegabar {$P^*(He) \approx 110$ Mbar [18], $P^*(Ne) \approx 1.34$ Gbar [19]} ranges. Although the static experimental technique of diamond anvils presently enables one to obtain pressures as high as about 5 Mbar [11], only in recent years has it become possible to record the metallization of xenon at $P^* = 1.5$ Mbar [20] in such experiments; at the same time, hydrogen seems to remain a dielectric at $P \approx 3$ Mbar [11].

By using the technique of strong shock waves to ensure a compression and an irreversible heating of matter, one can obtain much higher pressures (the world record is about 4 Gbar [21]), the upper limit on them being constrained only by the intensity of the source of their generation and not by the strength of diamond under static conditions. Concurrently, the viscous dissipation of the kinetic energy of the flux in the shock-wave front, along with compression, leads to a considerable heating of matter, and this stimulates the

thermal ($k_BT \sim I$) ionization of a plasma, whose kinetics and thermodynamics have been studied in detail both for an ideal and for a strongly nonideal case (see [1] and [2, 3], respectively). In such experiments, the influence of density effects on the ionization equilibrium is not pronounced against the background of fully developed thermal ionization and is described by various models of ionization-potential reduction [1–3, 22]. It should be noted that a number of theoretical models lose thermodynamic stability upon extrapolation to the region of strong nonideality and that this is attributed in [4, 6, 22–25] to the occurrence of a first-order plasma phase-transition. Thermodynamic states realized to date in static and dynamic experiments [26–36] are displayed in the phase diagram of molecular hydrogen (H_2) in Fig. 1. Also given in the same figure are theoretical estimates for plasma phase transitions, along with relevant critical points [23, 24, 37–39].

In order to separate density and thermal effects of ionization, one must naturally try to suppress the effects of irreversible heating ($k_BT \ll I$) by implementing a quasi-isentropic compression. For this purpose, the compression of the substance in this study was accomplished by means of a sequence of direct and reflected shock waves that emerge upon their reverberation in planar and cylindrical geometries. For the source of generation, we employed explosive devices of end-face and cylindrical throwing. By using processes of multiple shock compression, it proves possible to implement an order of magnitude reduced heating and an approximately tenfold increased compression of a plasma in relation to what we have in a direct wave, as well as to record, in experiments with H_2, He, and Xe, a five orders of magnitude increase in the plasma conductivity over a narrow density range peculiar to the regime of cold ionization of a plasma.

2. GENERATION AND DIAGNOSTICS

A typical layout of experiments to implement multiple shock compression of condensed hydrogen and inert gases in planar geometry is shown in Fig. 2 [36, 41, 42].

Shock waves were generated by the impact of a steel impactor (*2*) 1–3 mm thick and 30–40 mm in diameter accelerated by detonation products of a condensed high explosive (hexogen) (*1*) to velocities of 3–8 km/s with the aid of the gradient-cumulation effect [43]. Explosive throwing devices developed for these experiments ensured that, at the instant of impact against the bottom of the experimental assembly, the diameter of the flat part of the impactor was 15–30 mm. The absence of melting and evaporation of a shock-worker material, as well as the absence of mechanical fracture of the impactor during dynamic acceleration, was tested in a dedicated series of methodological experiments. The transition of a shock wave from a metallic screen (*3*) of thickness 1–1.5 mm to the substance under study (*4*) having an initial thickness of 1 to 5 mm generated, in it, the first shock wave of amplitude pressure P_1 = 0.02–0.8 Mbar; upon being reflected from a transparent sapphire window (*5*) 4–5 mm thick and 20 mm in diameter, this wave excited a repeated-compression shock wave. A further rereflection of shock waves between screen *3* and window *5* led to multiple shock compression of the sample to maximum pressures of $P \approx 1$–2 Mbar, level of which was determined by the velocity of the impinging impactor, its thickness, and the dimensions of the substance being studied.

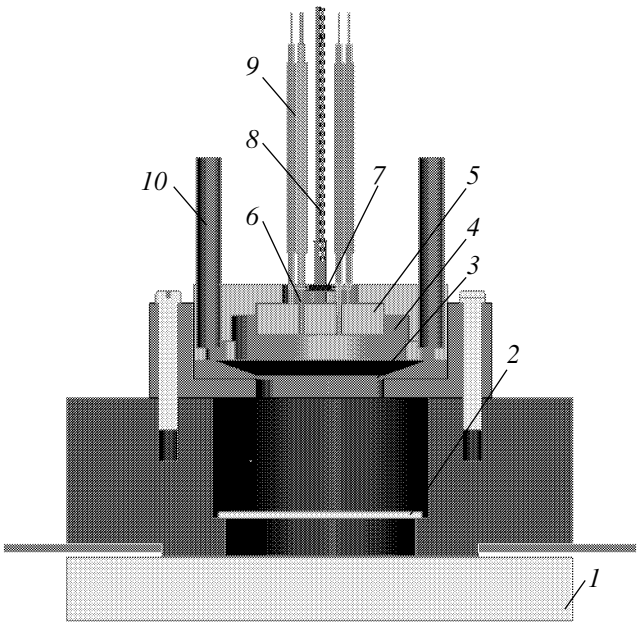

Fig. 2. Layout of experiments aimed at implementing multiple shock compression of condensed hydrogen and inert gases in planar geometry: (*1*) high-explosive charge, (*2*) steel plate, (*3*) bottom of the experimental assembly; (*4*) substance under study, (*5*) leucosapphire window, (*6*) indium electrodes, (*7*) shunting resistance, (*8*) quartz–quartz light guide, (*9*) coaxial electric cables, and (*10*) gas-supplying pipes.

The initial states of the explored substances for a further multiple compression were either in the gas region of the phase diagram at pressure and temperature values of P_0 = 5–35 MPa and T_0 = 77.4–300 K, respectively, or in its liquid region at $P_0 \approx 0.1$–1 MPa and $T \approx 20.4$–160 K. In the latter case, liquefaction was performed from high-purity gases supplied to the assembly through pipes (*10*). In liquefying hydrogen, use was made of a two-contour system of cooling, the external contour being filled with nitrogen, while, in liquefying xenon, the internal contour of the cooling system was filled with ethanol. The temperature in the assembly was monitored by thermocouples and platinum resistance thermometers.

The process of multiple compression was observed by means of fast optic–electronic convertors, as well as by means of a five-channel fiber-optic-coupled pyro-

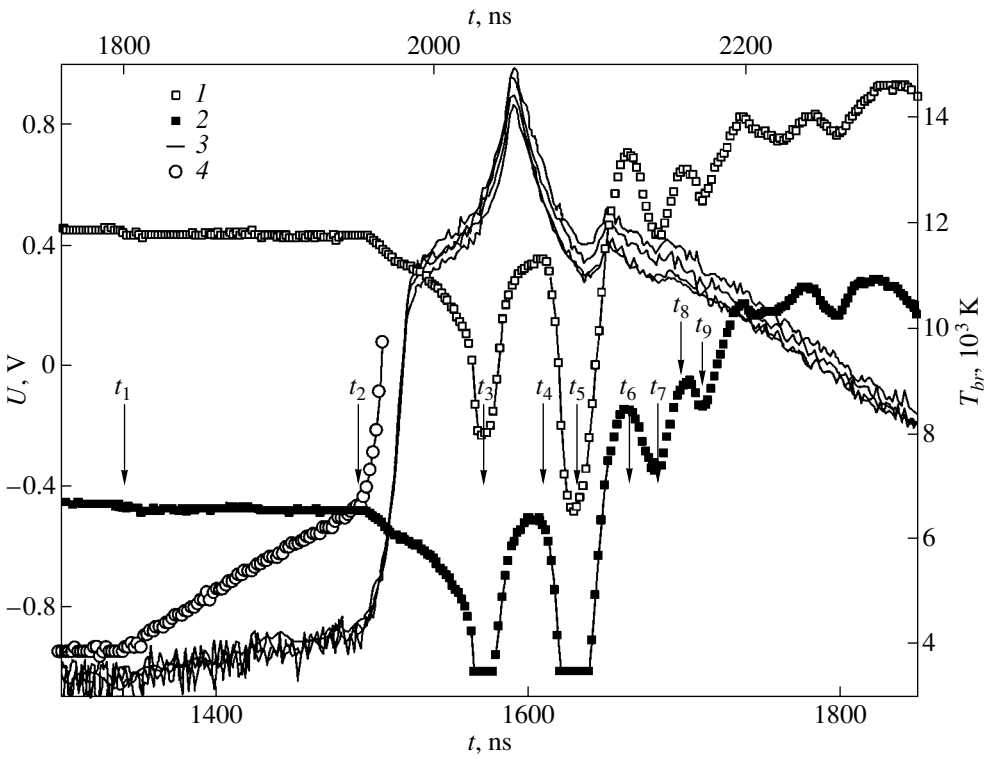

Fig. 3. Typical oscillogram of the experiment: (*1*, *2*) voltage at the "positive" and the "negative" electrode and (*3*, *4*) brightness temperatures of two channels for recording optical radiation (in this representation, the fourth temperature axis is linear in the range 3000–7000 K).

meter with time resolution of 2–5 ns (*8*). Since the shock-compressed sapphire of optic window *5* retained transparency up to $P \approx 20$ GPa and made it possible to record the instants of reflection of shock waves from its surface at still higher pressures and since its electric insulating properties were at an acceptable level under the conditions of compression up to pressures of about 2.2 Mbar [44], five to six reverberations of shock waves could be detected by measuring the conductivity of the compressed layer and optical radiation. Initial stages of the compression process (up to 20 GPa) were recorded in individual experiments with the aid of a VIZAR differential laser interferometer [45]. In order to synchronize the system for measuring the resistance of the compressed layer with that for detecting optical radiation, the light pulse from a laser diode that was connected instead of the measuring cell was recorded before each experimental run through a light guide and through the convertor, along electric cables (*9*) of the system for measuring resistance.

In the scheme chosen for the present experiments (see [38, 41, 42]), the compression and irreversible heating of the substance under study were implemented by series of shock waves arising upon successive reflections from the sapphire window and the steel screen. A hydrodynamic analysis of the process revealed that, following the propagation of the first two waves through the compressed layer, a further compression proceeded in a quasi-isentropic way. This made it possible to advance to the region of higher densities ($\rho/\rho_0 \sim$ 10–100) in relation to the case of single wave compression and to reduce the final temperature, whereby one enhances interparticle-interaction effects, which are of interest for the present investigation. The reverberation of shock waves manifests itself as distinct steps in the oscillograms of radiation and electrical conductivity (Fig. 3).

The measured instants of shock-wave arrival at the plasma-volume boundaries ($t_1 - t_9$) enable one to independently determine, by using the laws of mass, momentum, and energy conservation, the thermodynamic shock-compression parameters P, ρ, and E [3]. Data obtained in this way for the caloric and thermal equations of state of hydrogen, as well as of helium, which was chosen as a reference substance, up to pressures of 30–60 GPa are in accordance with the "chemical" non-ideal-plasma model [2, 3, 22] and with the solutions from the semiempirical equation of state of hydrogen [27, 46]. At pressures in excess of 60 GPa, however, no reliable information about the thermodynamics of the substances being studied could be obtained by means of the procedures used. In that case, the thermodynamic parameters of multiple shock compression at the final stage were calculated on the basis one-dimensional hydrodynamic codes that employ the

semiempirical equations of state from [27, 46] and [47] for hydrogen and structural materials, respectively.

This set of gasdynamic and temperature measurements was used to determine the thermodynamic parameters of shock compression at its initial stages; the results were also used as input data (along with the velocity W of the impinging impactor) in testing one- and two-dimensional gasdynamic codes according to which the parameters in question were determined for the case of higher pressures, densities, and temperatures of multiply compressed matter—the semiempirical equations of state [27, 46], valid over a wide region, were invoked in this determination. The errors in the P, ρ, and T values found in this way are 5, 10, and 20%, respectively.

The electrical conductivity of a shock-compressed plasma was determined by a probe method. An electric current was supplied to the shock-compressed plasma under study by means of electrodes (*6*) that were arranged orthogonally to the plane of the shock-wave front. Further, the current propagated along the shock-compressed sample and then arrived at the surface of steel screen *3*, whereupon it left the compressed region through a grounding electrode. The arising electric signals transferred by high-frequency coaxial cables (*9*) were recorded by multichannel digital oscilloscopes whose transmission bandwidth was 500 MHz. Use was made of two- or three-electrode schemes for recording resistance. In the second case, whose circuit diagram is given in Fig. 4, we were able to get rid of cophasal noises and to record the instants of wave reflection not only from the optic window but also from the screen.

The instants t_1, t_3, t_5, t_7, and t_9 corresponded to wave reflection from the window, and this was recorded with the aid of an electric (lines *1*, *2* in Fig. 3) and an optical (lines *3*, *4*) in Fig. 3) procedure; the instants t_2, t_4, t_6, and t_8 corresponded to wave reflection from the screen. With the aim of eliminating breakdown and arc effects in transmitting the transport current through a plasma, its density was maintained at a level not exceeding 10^4 A/cm^2. By varying this quantity within the range 10^3–10^4 A/cm^2 in a dedicated series of measurements, it was shown that the current–voltage characteristic of the plasma is linear. Determination of the plasma electrical conductivity on the basis of the plasma-gap resistance measured in this way was performed by resorting to numerical and electrostatic simulations of the corresponding electrostatic problem. As a result, the accuracy in measuring the plasma electrical conductivity was estimated at 20 to 50%.

The second series of measurements was performed by employing shock compression under the conditions of cylindrical geometry [48–50] (see Fig. 5).

A cylindrical charge of a high explosive (an alloy formed by trotyl and hexogen in a ratio of 40 : 60), its outer diameter being 30 cm, was initiated over the outer surface at 640 points that generated, at the inner surface of the charge, a highly symmetric detonation wave (the

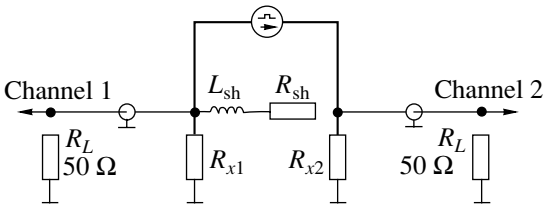

Fig. 4. Circuit diagram of the measurements.

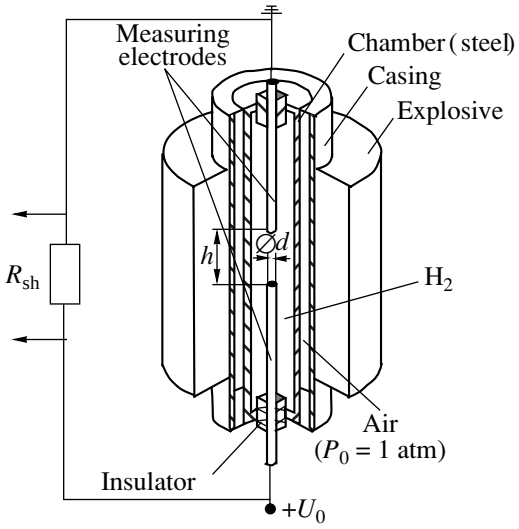

Fig. 5. Layout of cylindrical compression.

difference in time of arrival was not greater than 100 ns). The arrival of this wave at the inner surface caused the centripetal motion of the steel impactor at an initial velocity of $W \approx 5$ km/s. The deceleration of this cylindrical impactor against the metallic surface of the chamber filled with the gas under study at an initial pressure of up to 70 MPa generated a converging shock wave, whose intensity increased as it traveled to the center, this increase being governed by the regularities of geometric cumulation [51]. Thereafter, there occurred successive reflections of the shock wave from the center of symmetry and from the moving inner surface of the chamber, and this gave rise to multiple shock compression, which, as in the case of planar geometry, proved to be close to isentropic compression.

At each instant of time, the profiles of thermodynamic parameters of multiple compression were determined on the basis one- or two-dimensional gasdynamic calculations employing, for the high explosive, the structural materials of the assembly, and target plasmas, semiempirical equations of state that are valid over a wide region. In some special experiments, the process of cylindrical explosive compression was monitored by measuring the velocity of the impactor by electric-contact and fiber-light optical basis methods, as well as by means of examination along the axis with

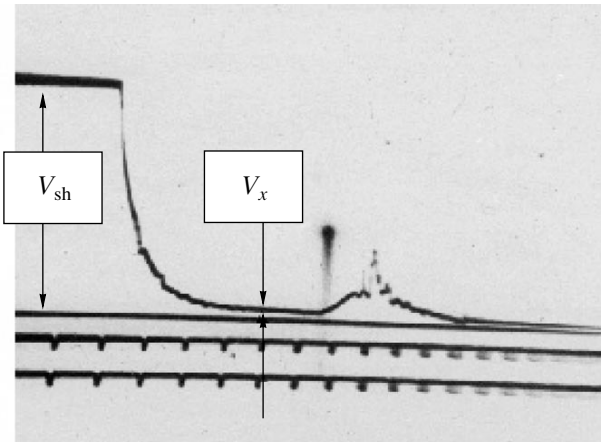

Fig. 6. Oscillogram of the experiment where the initial hydrogen pressure was 70 MPa (the time-scale division here is 1 μs).

two hard-radiation sources [52] emitting beams crossed at an angle of 135°; this made it possible to follow the dynamics of compression and to test the quality of gas-dynamic calculations and yielded additional boundary conditions for the codes used in those calculations. The parameters found in this way for the shock-compressed plasma in the device being considered had the following values: for deuterium, the pressure was 1.25–1.44 TPa at temperatures of 12500–14000 K and densities of 2–2.4 g/cm^3; for xenon, the pressure and density were 200 GPa and 13 g/cm^3, respectively, the latter value being nearly five times as great at the density of solid Al.

The electrical conductivity was measured according to the classic two-point circuit diagram [53] involving a reference resistance connected in parallel with the resistance of the sample under study. The resistance of hydrogen was determined with the aid of two stainless-steel electrodes 2 mm in diameter arranged on the axis of the device with a gap length of $h = 6.5$ mm between their end faces. This procedure of measurements employed the discharge of a large-capacitance capacitor through the resistance R_{sh} shunting the hydrogen sample. The decrease in the compressed-hydrogen resistance R_x led to a decrease in the total resistance, with the result that the voltage across the measuring electrodes changed, which was recorded by an oscilloscope. In these experiments, R_{sh} was 3 Ω. Figure 6 displays an oscillogram that was obtained in one of the experiments at an initial hydrogen pressure of P_0 = 70 MPa. The resistance estimated by using this oscillogram is $R_x \approx 0.2$ Ω.

In determining the electrical conductivity on the basis of the measured resistance values, the actual geometry of the current distribution between the electrodes was taken into account, along with the geometry of the electrodes at each specific instant of compression and with the results obtained by calculating, according to hydrodynamic codes, the profiles of the thermodynamic parameters of hydrogen. The error in the electrical-conductivity values found in this way was estimated at 50%.

The characteristic plasma parameters obtained in some experiments are quoted in Table 1.

3. EXPERIMENTAL RESULTS FOR THE ELECTRICAL CONDUCTIVITY; MODEL OF PRESSURE-PRODUCED IONIZATION

Experiments aimed at implementing multiple shock compression of hydrogen and inert gases make it possible to obtain physical information about an as-yet-unexplored part of the phase diagram, which is depicted in Fig. 1 for hydrogen. One can see that the region of pressures of up to 15 Mbar and temperatures of 3000 to 7000 K was reached by means of dynamic compression. Concurrently, densities are realized that are one order of magnitude higher than those of solid hydrogen and solid inert gases under normal conditions, in which case the mean spacing between protons, $n^{-1/3} \sim 1$ Å, is commensurate to the typical sizes of both molecules (about 0.74 Å) and atoms in the ground state.

From the point of view of physics, this region is of interest since this is the region of strong interaction

Table 1

Substance	Initial state	Final state	P, GPa	ρ, g/cm^3	T, 10^3 K	σ, (Ω cm)$^{-1}$
		Planar geometry				
H_2	P_0 = 25.6 MPa, T_0 = 77.4 K	Maximum compression	227	0.94	5.3	1600
He	P_0 = 28 MPa, T_0 = 77.4 K	Maximum compression	126	1.37	15	1080
Xe	P_0 = 0.1 MPa, T_0 = 160 K	Maximum compression	126	10	25	500
		Cylindrical compression				
H_2	P_0 = 50 MPa, T_0 = 293 K	Maximum compression	1440	2.4	14	550
	P_0 = 70 MPa, T_0 = 293 K	Maximum compression	1250	2	12.5	1100

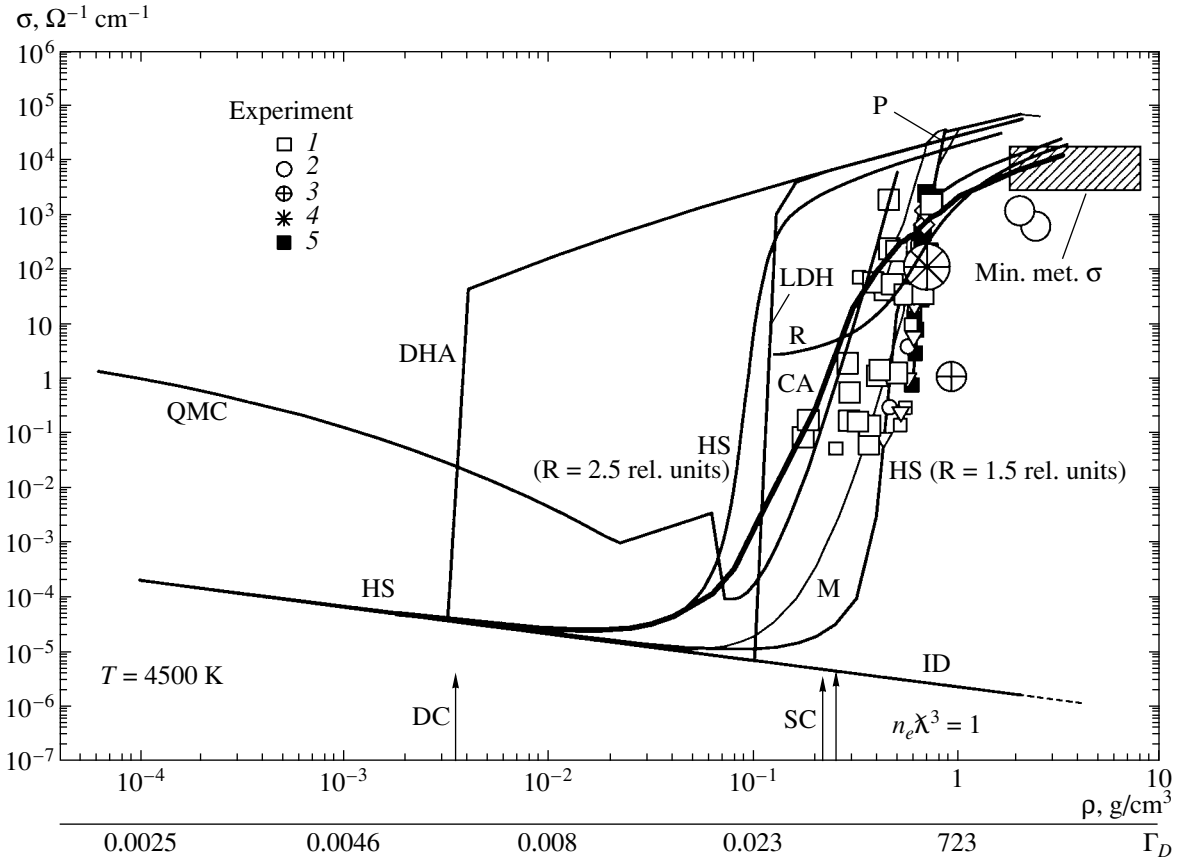

Fig. 7. Electrical conductivity of hydrogen as a function of density. The experimental data of the present study were obtained for (*1*) planar systems and (*2*) cylindrical systems. The remaining data were borrowed from (*3*) [29] and (*4*) [30] (magnetic compression) and from (*5*) [35] (experiments with light-gas guns).

concerning both interatomic (intermolecular) interaction ($\Gamma_a = r_a n_a^{-1/3} \sim 1$—that is, r_a, which is the molecular or the atomic size, is commensurate with the interparticle spacing $n_a^{-1/3}$) and, the latter implying that the mean interaction energy of charged particles,

$$E_C = e^2/r_D, \quad r_D = \left(4\pi e^2 \sum_i Z_i^2 n_i / k_A T\right)^{1/2},$$

is much greater than the mean kinetic energy of thermal motion, E_T ($\Gamma_D = E_C/E_T \sim 10$). The situation is additionally complicated by the fact that the type of statistics changes upon compression—electrons become degenerate, $n_e\lambda_e^3 \approx 200$ [$\lambda_e = (2\pi\hbar^2/m_e k_B T)^{1/2}$ is the thermal de Broglie wavelength], with the result that $E_T \sim k_B T$ as the scale of the kinetic energy of particles gives way to the Fermi energy,

$$E_F = \frac{(3\pi^2 n_e)^{2/3}}{2m_e}, \quad E_T \sim E_F.$$

All these circumstances greatly complicate theoretical description of strongly nonideal states [2, 3], hindering the application of perturbation theory and of parameter-free computer Monte Carlo and molecular-dynamics methods in their classic formulation [54], which were developed for Boltzmann statistics.

Our experimental results for the electrical conductivity of shock-compressed hydrogen and inert gases are displayed in Figs. 7–11, along with the results obtained on the basis of some theoretical models [1–3, 22, 37, 24, 55, 56] and the results of other measurements [29, 30, 35, 36, 41, 42, 57–65].

Let us first point out some general features in the behavior of the electrical conductivity of a strongly nonideal plasma. The most prominent feature is that, at final stages of compression, the electrical conductivity of the plasma increases sharply (by three to five orders of magnitude) in the process of compression in a nar-

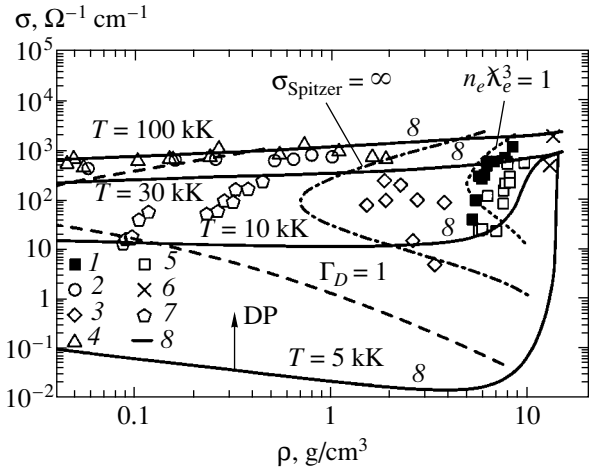

Fig. 8. Electrical conductivity of xenon as a function of density. The displayed experimental data were borrowed from (*1*) [49], (*2*) [57], (*3*) [58], (*4*) [59], (*5*) [41], (*6*) [60], and (*7*) [61]. Also plotted in this figure are the electron-degeneracy line ($n_e\lambda_e^3 = 1$), the line on which the parameter of Coulomb nonideality takes the constant value of $\Gamma_D = 1$, and the line on which the electrical conductivity calculated by Spitzer's formula goes to infinity ($\sigma_{Spitzer} = \infty$). Solid lines *8* represent the electrical conductivity calculated on the basis of the model formulated in the main body of the text.

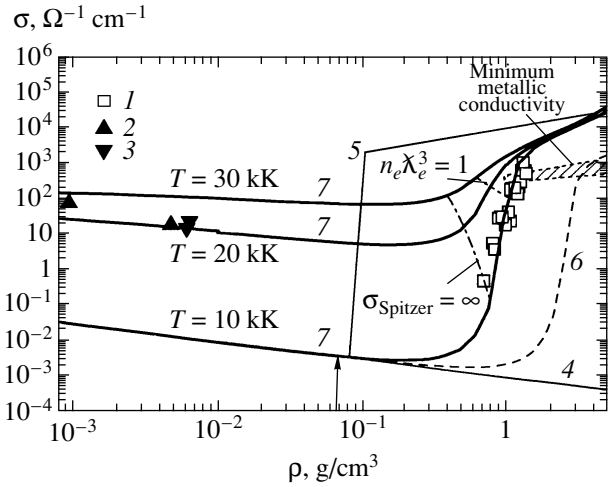

Fig. 9. Electrical conductivity of helium as a function of density: (*1*), (*2*), and (*3*) experimental data from [42], [57], and [61], respectively; (*4*) electrical conductivity calculated with the plasma composition corresponding to the model of an ideal plasma; (*5*) results obtained with the plasma composition calculated on the basis of the Debye–Hückel model [1]; (*6*) results obtained with the plasma composition calculated on the basis of the bounded-atom model [2, 22] featuring a fixed radius of the helium atom ($r_a = 1.3a_0$); and (*7*) results of the present study.

row range of "compressed" densities ($\rho \approx 0.3$–1 g/cm^3 for hydrogen, and $\rho \sim 8$–10 g/cm^3 for xenon) at megabar pressures, reaching values of about 10^2–10^3 Ω^{-1} cm^{-1}, which are characteristic of alkali metals. Our measurements exhibit a pronounced threshold effect in density and are therefore in a qualitative contradiction with models of weakly nonideal plasma [1], which predict a monotonic decrease in the plasma electrical conductivity in response to its isothermal compression [1].

Indeed, it is well known that, at low degrees of the ionization of a plasma,

$$\alpha_i = \frac{n_e}{n_a + n_e} \ll 1,$$

its electrical conductivity is determined by the scattering of electrons on neutral particles and is qualitatively described by the Lorentz formula [3], according to which the electrical conductivity is in direct proportion to the concentration of free electrons; that is,

$$\sigma_{ea} = \frac{2\sqrt{2}}{3\sqrt{\pi}m_e^{1/2}(k_BT)^{1/2}}\frac{n_e}{n_a}\frac{1}{q_{ea}^*(T)}, \qquad (3.1)$$

where q_{ea}^* is the averaged cross section for electron scattering by atoms:

$$\frac{1}{q_{ea}^*(T)} = \frac{1}{(k_BT)^2}\int \exp\left(-\frac{E}{k_BT}\right)\frac{dE}{q_{ea}(E)}.$$

In turn, the composition of a plasma is described by Saha's ionization-equilibrium equation [2]

$$\frac{n_e n_i}{n_a} = \frac{2Q_i}{Q_a}\lambda_e^{-3}\exp\left[-\frac{I - \Delta I(n_e, n_0, T)}{k_BT}\right], \qquad (3.2)$$

where Q_a and Q_i are the partition functions for atoms and ions, respectively, I is the ionization potential, and

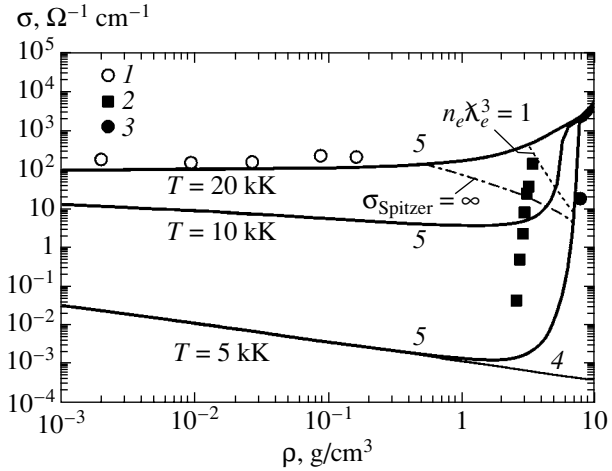

Fig. 10. Electrical conductivity of argon as function of density: (*1*), (*2*), and (*3*) experimental data from [57], [62], and [63], respectively; (*4*) electrical conductivity calculated with the plasma composition corresponding to the model of an ideal plasma; and (*5*) results of the present study.

ΔI is the reduction of the ionization potential due to interparticle interaction (nonideality). Thus, one can see that, at $\alpha_i \ll 1$, it follows from Eqs. (3.1) and (3.2) that, in the absence of interaction ($\Delta I = 0$, and Q_i, Q_a = const), the electrical conductivity under the conditions of isothermal compression is given by

$$\sigma \sim \alpha_i \sim 1/\rho,$$

which corresponds to the curves for an ideal plasma in Figs. 7, 9, and 10.

Nonideality, which must be included under the present conditions for determining the composition of a plasma, is taken into account here by introducing a density-dependent quantity ΔI and less significant dependences for Q_i and Q_a, whereby one arrives at a nonthermal growth of the degree of ionization and at an increase in the plasma electrical conductivity upon isothermal compression in accordance with Eq. (3.1). On the curve representing the electrical conductivity as a function of density at T = constant, there appears a minimum, its depth being greater for lower temperatures. With increasing temperature, this minimum levels out as soon as thermal-ionization effects (at $k_BT \sim I$) become more pronounced than effects associated with pressure-produced ionization, which are significant at $k_BT \ll I$.

As the density increases further at a given temperature, ionization processes described by relation (3.2) are completed. Thereupon, one deals with a strongly ionized ($\alpha_i \sim 1$) weakly nonideal plasma, where, instead of Eq. (3.1), it is necessary to use the Spitzer approximation (which is valid in the case of a nondegenerate plasma) [3]

$$\sigma_{ei} = \gamma(Z)\frac{4\sqrt{2}}{\pi^{3/2}}\frac{(k_BT)^{3/2}}{Ze^2m_e}\frac{1}{\Lambda} \tag{3.3}$$

or, in the case of Fermi statistics, the relation

$$\sigma \sim \frac{n_e}{\Lambda}, \tag{3.4}$$

where Λ is the Coulomb logarithm. This means that, at high temperatures, the exponential dependence in (3.1) and (3.2) gives way to a weaker [logarithmic in the case of (3.3) or linear in the case of (3.4)] dependence on the carrier concentration. In this case, we can take, for an estimate of conductivity, the so-called Regel'–Ioffe "minimal metal" conductivity, which is widely used in the theory of simple metals and semiconductors; that is,

$$\sigma \sim \frac{n_eR_S}{v_T}, \tag{3.5}$$

where R_S is the radius of the Wigner–Seitz cell and v_T is the mean thermal velocity of electrons.

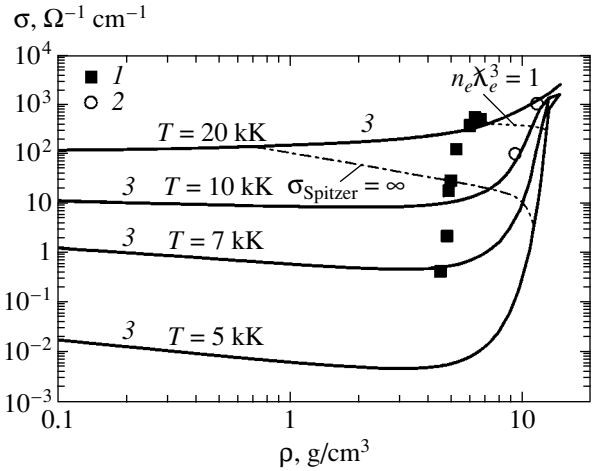

Fig. 11. Electrical conductivity of krypton as function of density: (*1*) and (*2*) experimental data from [64] and [65], respectively, and (*3*) results of the present study.

We can see that the exponential growth of the number of carriers due to the reduction of the ionization potential because of strong interparticle interaction in plasmas of condensed densities is the main reason behind the sharp increase in the measured electrical conductivity.

It should be emphasized that, in exactly the same way as the semiconductor thermal-excitation model (see, for example, [66]) featuring an energy gap $\Delta(\rho)$ that decreases with increasing density, the plasma pressure-produced-ionization model that is based on Eqs. (3.1) and (3.2), which is discussed here, leads to an exponential variation of the electrical conductivity with temperature:

$$\sigma \sim \sigma_0 \exp(-\Delta(\rho)/k_BT).$$

This model was used in [35] to analyze experiments with light-gas guns.

Thus, the data obtained here for the electrical conductivity at $k_BT \ll I$ provide a unique possibility for adequately choosing thermodynamic models that would describe the reduction of the ionization potential. For example, an analysis of the data in Figs. 7–11 reveals that the standard Debye–Hückel model (DHA curve in Fig. 7 and curve *5* in Fig. 9) strongly overestimates effects of Coulomb interaction, leading to pressure-produced ionization at densities that are two orders of magnitude lower than their experimental counterparts.

Having performed the above qualitative analysis, we will now perform quantitative calculations of physical parameters that characterize a dense plasma and compare the results obtained in this way with experimental data.

4. THERMODYNAMICS OF MEGABAR-RANGE PLASMAS

As a model intended for describing experiments devoted to the shock and quasi-isentropic compression of plasmas, we consider the bounded-atom model [67], which explicitly takes into account the finiteness of the phase space for the realization of the bound states of atoms and ions and which was previously used in [2, 3, 67] to describe the thermodynamics of shock-compressed inert gases and cesium. As a matter of fact, this model is a generalization of the Wigner–Seitz solid-state model [1, 3] to the region of condensed-density plasmas via the inclusion of thermal and density-ionization effects. Within this model, atoms and ions are treated as rigid spheres, whose thermodynamic functions are constructed on the basis of molecular-dynamics and Monte Carlo calculations, the contribution of the bound electrons of these atoms and ions being described in the quantum-mechanical Hartree–Fock approximation. Figure 12 shows the energy spectrum of a compressed hydrogen atom.

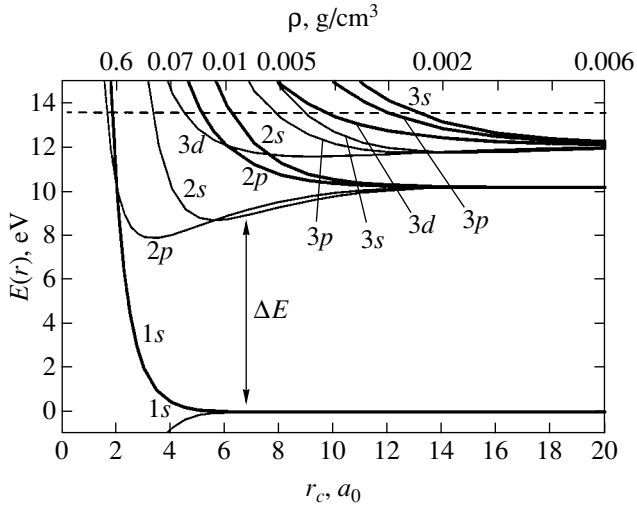

Fig. 12. Energy spectrum of the hydrogen atom for (thick curves) $f_{nl}(r_c) = 0$ and (thin curves) $f'_{nl}(r_c) = 0$.

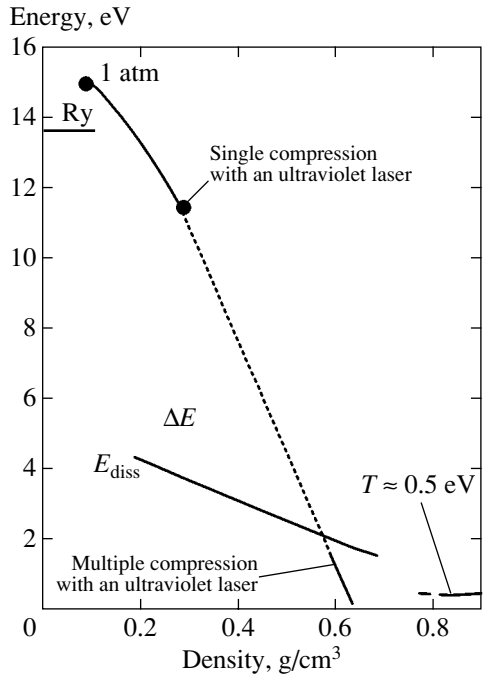

Fig. 13. Bandgap of deuterium.

In the calculation of this spectrum, the radial component of the wave function was required to satisfy the boundary conditions

$$f_{nl}(r)\big|_{r = r_e} = 0, \quad \left.\frac{\partial f_{nl}(r)}{\partial r}\right|_{r = r_c} = 0.$$

Within the solid-state model [68], this corresponds to the upper and lower boundaries of the energy band within which the relevant energy level of an isolated ($r_c \longrightarrow \infty$) atom occurs as the result of compression (decrease in r_c). In this approach, the width of the forbidden band, ΔE (energy gap), can be taken as the energy difference between upper boundary of the ground-state band (curve *1s*) and the lower boundary of the band built on the first excited stated (curve *2p*); as can be seen from Fig. 13, the energy gap, which decreases with increasing density, is in accordance with the experimental data reported in [44, 69] and obtained from a direct treatment of data from experiments devoted to multiple compression of hydrogen and deuterium. Versions of this model were successfully used to describe the thermodynamics of metal plasmas in the region of high and ultrahigh (up to 4 Gbar) pressures [70–72].

The parameter region investigated in the experiments being discussed is characterized by extremely complicated and diversified processes that must be reflected in the corresponding physical models. It should be noted first of all that, in the course of compression, the thermodynamic plasma composition may change sharply, which is accompanied by the emergence of strong interparticle interactions, including Coulomb interaction (between electrons and ions), polarization interaction (between charged and neutral particles), and short-range interaction (between neutral particles). Since the typical interparticle spacing in the plasmas considered here is commensurate with characteristic sizes of atoms and ions, the phase-space part occupied by them becomes inaccessible to other particles; as a result, their kinetic energy grows, and so do the corresponding contributions to the free energy of such strongly compressed disordered structures. Moreover, the energy spectrum of bound states undergoes changes in atomic and molecular systems subjected to

a strong compression. Also, it is necessary to take into account the change in statistics (from Boltzmann to Fermi statistics) for continuous-spectrum electrons, since, under the conditions being considered, the degeneracy parameter $n_e\lambda\!\!\!^{-}_e^3$ changes from 0.001 to 200.

In the present study, the following items of the full thermodynamic approximation were used in calculating the thermodynamic parameters of plasmas in the megabar range of pressures.

The free energy of a quasineutral mixture of electrons, ions, atoms and molecules can be broken down into the contribution of the ideal-gas component and the term that takes into account interparticle interaction; that is,

$$F \equiv F_i^0 + F_e^0 + F_{ii,\,ie,\,ee,\,\ldots}^{\text{int}}. \tag{4.1}$$

It is assumed that heavy particles (atoms, ions, molecules) obey Boltzmann statistics, their contribution having the standard form

$$F_i^0 = \sum_j N_j k_B T\left(\ln\frac{n_j\lambda\!\!\!^{-}_j^3}{Q_j} - 1\right), \tag{4.2}$$

where Q_j stands for the partition functions of atoms and ions.

(A) **Electron degeneracy.** Electrons are treated as a partly degenerate ideal Fermi gas:

$$F_e^0 = 2Vk_BT\pi^{-1/2}\lambda\!\!\!^{-}_e^{-3} \times\left[(\mu_e/k_BT)I_{1/2}(\mu_e/k_BT) - \frac{2}{3}I_{3/2}(\mu_e/k_BT)\right], \tag{4.3}$$

$$\frac{P_e^0}{n_ek_BT} = \frac{2}{3}\frac{I_{3/2}(\mu_e/k_BT)}{I_{1/2}(\mu_e/k_BT)}. \tag{4.4}$$

Here, the electron density n_e and the chemical potential μ_e are related by the equation

$$n_e\lambda\!\!\!^{-}_e^3 = 2\pi^{-1/2}I_{1/2}(\mu_e/k_BT), \qquad I_t(x) = \int_0^\infty \frac{y^t\,dy}{1+\exp(y-t)}. \tag{4.5}$$

The inclusion of electron-degeneracy effects is of paramount importance in the phase-diagram region discussed here, since the degeneracy parameter $n_e\lambda\!\!\!^{-}_e^3$ can be much greater than unity in this region.

(B) **Coulomb interaction.** We applied a version of the pseudopotential-model for multiple ionization [73, 74]. The inclusion of the fact that, at short distances, the interaction of free charges deviates from a Coulomb form is a key point of this model, and there arises, upon explicitly taking this deviation into account, a noticeable positive shift not only in the potential energy of free charges but also in their mean kinetic energy. It should be noted that the depth of the electron–ion pseudopotential, $\Phi_{ie}^*(0)$, is related, in the model, to the boundary separating, in the partition function (4.2), free states of each electron–ion pair from bound states. The electron–ion pseudopotential in the Glauberman–Yukhnovskiĭ form is given by (Fig. 14)

$$\Phi_{ie}^*(r) = -\frac{Z_ie^2}{r}(1 - e^{-r/\sigma_{ie}}), \quad \sigma_{ie} \equiv \sigma_{ie}(n, T), \qquad \Phi_{\alpha\alpha}^*(r) = \frac{Z_\alpha Z_\alpha e^2}{r}, \quad \alpha = i, e. \tag{4.6}$$

For potential (4.6), the parameters of correlation functions were determined from conditions that are formulated immediately below and which are quite general and are valid at arbitrary values of the Coulomb nonideality parameter

$$\Gamma_D = [4\pi(e^2/k_BT)\sum n_\alpha z_\alpha^2]^{-1/2}.$$

Specifically, these are

(i) the condition of local electroneutrality,

$$\int\left\{n_{ei}[F_{ei}(r) - 1] + \sum_j n_{ij}Z_j[F_{ij}(r) - 1]\right\}d\mathbf{r} = -Z_i; \tag{4.7}$$

(ii) the condition of dipole screening,

$$\int\left\{n_{ei}[F_{ei}(r) - 1] + \sum_j n_{ij}Z_j[F_{ij}(r) - 1]\right\} \times\left(\frac{r}{r_D}\right)^2 d\mathbf{r} = -3Z_i; \tag{4.8}$$

(iii) the nonnegativity of correlation functions,

$$F_{ik}(r) \geq 0; \tag{4.9}$$

(iv) the relation between the screening-cloud amplitude and the depth of the electron–ion pseudopotential,

$$F_{ei}(0) \equiv 1 + \Psi_{ei}(0) \approx \beta\Phi_{ei}^*(0), \quad F_{ii}(0) \approx 0. \tag{4.10}$$

At $\Gamma_D \ll 1$, corrections that are associated with the interaction of charges and which were obtained from the conditions in (4.6)–(4.10) are close to Debye corrections, while at $\Gamma_D \gg 1$, they are smaller.

(C) **Short-range repulsion.** The contribution of the short-range repulsion of molecules, atoms, and ions is

described phenomenologically within the soft-sphere approximation [75] generalized to the case of a multicomponent mixture; that is,

$$\frac{\Delta F_{SS}}{Nk_BT} = C_s y^{s/3}(\varepsilon_{SS}/k_BT) + \frac{s+4}{6}Qy^{s/9}(\varepsilon_{SS}/k_BT)^{1/3}, \quad (4.11)$$

$$y = \frac{3Y\sqrt{2}}{\pi}, \quad Y = \frac{4\pi r_c^3}{3} = \frac{\pi\sigma_c^3}{6}, \quad r_c = \left[\frac{\sum n_j r_j^3}{\sum n_j}\right]^{1/3},$$

where

$$C_s = 6 + \frac{6.669}{s-3} = 1.043(s-4)^{0.389}\exp[0.156(4-s)]$$

is the Madelung constant for the potential $V(r) = \varepsilon(r/\sigma)^{-s}$.

We note that the corrections for short-range repulsion to the chemical potential,

$$\frac{\Delta\mu_{jSS}}{k_BT} = \frac{\Delta F_{SS}}{Nk_BT} + \frac{\Delta P_{SS}}{nk_BT}\left[\frac{r_j}{r_c}\right]^3, \quad (4.12)$$

are different for particles having different radii, and this determines the decrease in the ionization (dissociation) energy with increasing matter density.

The above thermodynamic model provides a correct asymptotic behavior at low plasma densities, where it coincides with well-known theories of dilute plasma. In the region of extremely high densities, the applicability of this model was tested by comparing the results that it yields with available experimental data on the thermodynamics of the plasmas of alkali metals [2], inert gases [55–57], and shock-compressed strongly porous metals [72, 74]. Considered individually below are special features of the calculation for each element of interest.

4.1. Hydrogen

In the phase diagram of hydrogen (Fig. 1), the transition to the metallic state at low temperatures is shown in accordance with the estimates given in [22] at a pressure of about 300 GPa. The triple point at which the metal phase coexists with condensed molecular hydrogen and a molecular liquid is predicted in [22] to occur at $P = 100$ GPa and $T = 1500$ K. There are two critical points (CPs) in the molecular-liquid phase. One of these, CP_1, and the curve along which a molecular gaseous hydrogen and a liquid coexist are well known to lie in the low-temperature region. The position of the second critical point (CP_2), which is of greatest interest to us, and the position of the coexistence curve (curve (*12*) that is associated with a sharp change in the degree of dissociation and ionization of hydrogen are not known precisely. According to the estimates presented in [22], $T[CP_2] = 16500$ K, $P[CP_2] = 22.8$ GPa, and $\rho[CP_2] = 0.13$ g/cm^3. Also shown in Fig. 1 are estimates of other authors [23, 24, 37–39] for the coexistence curve and for the critical point at which this plasma phase transition occurs (curves *12–16*).

It can be seen that the parameter region corresponding to multiple shock compression in planar systems [36] (region *9*) and in experiments with light-gas guns (LG) [34, 35] (curves *7, 8*), as well as in cylindrical systems [26, 27] (curves *4, 5*), partly overlap this rather large region of the possible existence of a plasma phase transition. The shock adiabatic curves of a single compression of liquid hydrogen (deuterium) that were obtained in experiments with high-power lasers [31, 32] (curves *10, 11*), in a high-current *Z* pinch [33] (curve *3*), and in explosive spherical systems [28] (star *6*) also occur in the region of our interest, but they lie at higher temperatures. Temperatures of about 700 K were realized in experiments devoted to the isentropic compression of hydrogen by strong magnetic fields in explosive magnetic-compression systems [29, 30] (boxes *1, 2*). Pressures of up to 300 GPa were obtained via the isothermal compression of hydrogen ($T \approx 300$ K) in diamond anvils (DA) [11] (curve *17*). The region where strong Coulomb interaction is operative and the region where the degeneracy of the electron component is significant lie above the curves $\Gamma_D = 1$ and $n_e\lambda_e^3 = 1$, respectively. Curve *19* characterizes the behavior of the shock adiabatic curve for liquid hydrogen, while curve *18* represents parameters that are realized in Jupiter's atmosphere [40]. Figure 1 also displays the regions of typical parameters achievable with the aid of ordinary and explosive shock tubes (see [76] and [77], respectively), in discharges, and in usual low-current pinches [3].

As can be seen, the existence of a large "monomolecular" region ($\rho \le 0.3$ g/cm^3, $-\mu_H \ge D(H_2) \approx 4.5$ eV), where the thermodynamics of hydrogen is almost completely determined by H_2–H_2 interaction, is a feature characteristic of hydrogen. Within the soft-sphere model [75], which is used in the present study, the parameters of the H_2–H_2 interaction were chosen here to be maximally close to those recommended within the rigorous "nonempirical" atom–atom approximation [78], the noncentrality of this interaction being disregarded. The calculations have revealed that the use of the soft-core repulsion $V(r) \propto 1/r^6$ makes it possible to describe the molecular part of the $T = 0$ isotherm ("cold curve") and a considerable part of shock-wave experiments, as well as the results of precise Monte Carlo calculations of $H_2 + H_2$ thermodynamics [78].

The main problem of the chemical model in describing nonideality, including nonideality in the case of dense hydrogen, is that of correctly specifying the entire set of potentials that would simulate the interactions between all members of the mixture being considered. This concerns the interactions involving both charged and neutral particles—first of all, interactions

in H_2–H and H–H pairs. It is of importance that the effective interaction of free atoms that appears in the chemical model differs radically from the singlet (attractive) and triplet (repulsive) branches of the total potential of H–H interaction that are obtained from a rigorous theory, since the contribution of H–H pairs involved in the singlet-branch interaction has already been taken into account in the discharge of intramolecular motion. This is all the more justified for the effective interaction involving (free) charged particles, since, in the chemical model, the contributions of free and bound states must be consistent [see the approximation specified by Eqs. (4.6)–(4.10)]. At present, there is a serious discrepancy between the results obtained within different approaches for the form and parameters of these potentials. Off the monomolecular region, the most glaring contradictions are those in the parameters of the short-range repulsion in H–H and H–H_2 pairs. We note that, according to the present calculations, the parameters of the effective potentials of H_2–$A^{\pm}$ interactions (where the symbol $A^{\pm}$ stands for all charged components) are of no lesser importance. For one of the versions, we would like to indicate the results obtained within the nonempirical atom–atom approximation [78], which lead to relatively large "intrinsic volumes" of the hydrogen atom. In terms of the soft-sphere-model modification introduced in [75] and used here, the results presented in [78] correspond almost exactly to the "additive-volume" approximation,

$$[d(H_2)]^3 \approx 2[d(H)]^3.$$

For $\rho \le \rho^* \approx 0.3$ g/cm^3, this choice leads to results that agree, for $T \le 10$ kK, with the results of precise Monte Carlo calculations [78] and, for $T \ge 10$ kK, with the nonanomalous part of the results obtained by means of a quantum Monte Carlo method (PIMC [79]). At such temperatures, the data are also in satisfactory agreement with the results produced by other versions of *ab initio* approaches, including the method of quantum molecular dynamics (TBMD [80]) and the method of wave packets (WPMD [81]).

Figure 15 displays the entire body of currently available experimental data on single shock compression of liquid hydrogen.

Pressures of up to 25 GPa (point *1* in Fig. 15) were achieved in a direct shock wave generated in experiments with light-gas guns [34]. Investigations aimed at generating shock waves with the aid of high-power lasers [31, 32] (points *2*, *3*) made it possible to obtain pressures of up to 300 GPa and to discover an anomalously high compressibility of deuterium at a pressure of $P > 40$ GPa. However, more recent results that emerged from the *Z*-pinch experiment reported in [33] (points *4*) and from the experiment of Belov *et al.* [28] (points *5*) with explosive spherical systems did not confirm the existence of this anomaly up to $P \approx 70$ GPa.

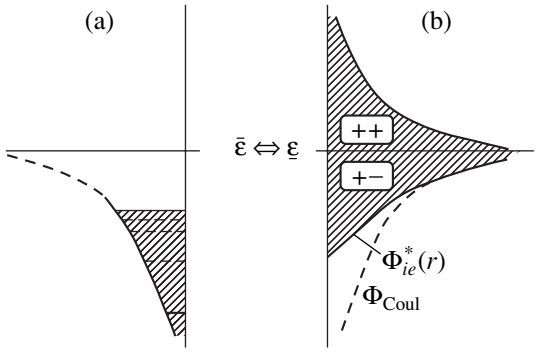

Fig. 14. Glauberman–Yukhnovskiĭ electron–ion pseudopotential: (a) bound states and (b) continuous spectrum.

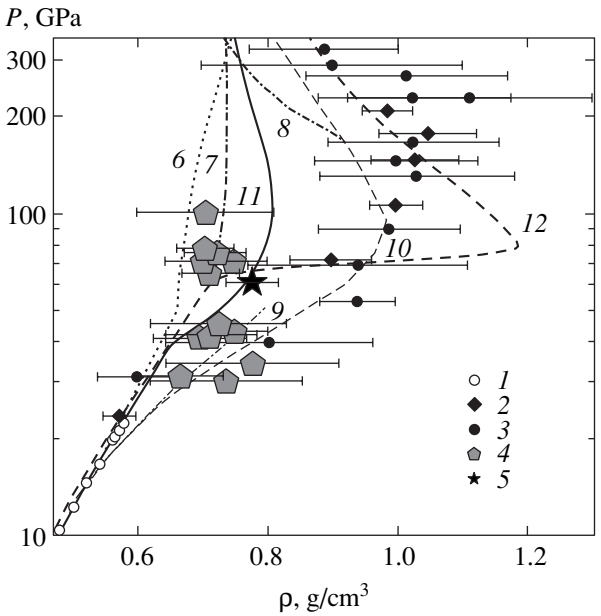

Fig. 15. Shock adiabatic curve for deuterium: (*1*), (*2*), (*3*), (*4*), and (*5*) experimental data from [34], [31], [32], [33], and [28], respectively; (*6*), (*7*), (*8*), (*9*), and (*10*) results of the calculations from [82], [79], [22], [37], and [83], respectively; and (*11*, *12*) results of the present study.

The shock adiabatic curves calculated by using the SESAME equation of state [82] (curve *6*) do not predict this anomaly in the behavior of shock compressibility, nor does it arise in calculations with semiempirical equations of state [27]. The emergence of this anomaly is not expected either if use is made of *ab initio* approaches, such as the quantum Monte Carlo method [79] (curve *7*) and the molecular dynamics method [80]. For deuterium, Ross [83] presented an interpolation equation of state (curve *10*) that qualitatively describes experimental results obtained with the aid of lasers.

The approach considered in the present study also does not reproduce this abrupt change in the behavior of the shock adiabatic curve for deuterium (toward unexpectedly high degrees of compressibility, $\sigma_{\max} \equiv$

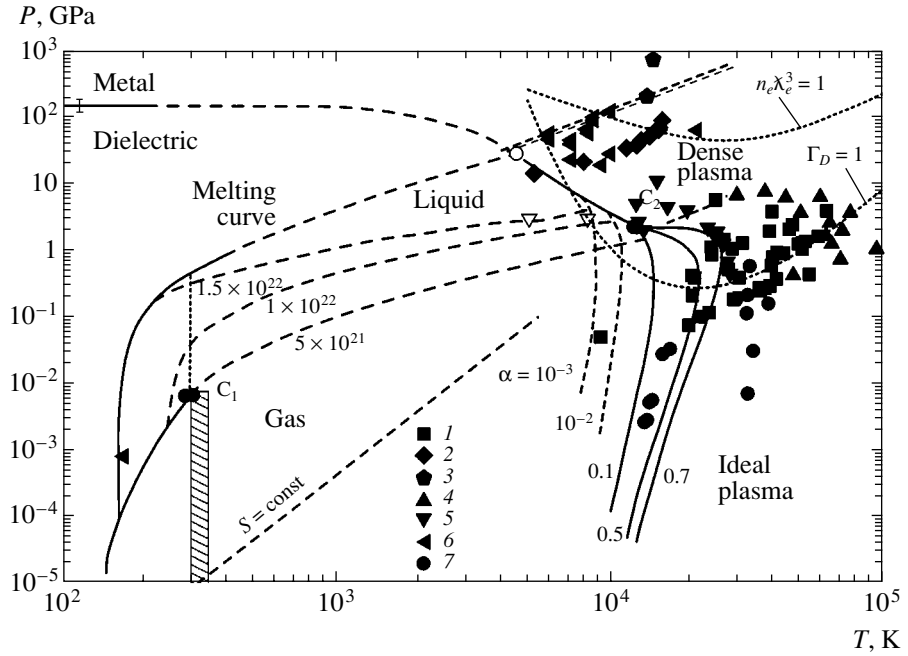

Fig. 16. Phase diagram of xenon. The displayed experimental data were obtained by measuring (i) the equation of state of xenon with the aid of (*1*) the shock compression of gases in [67], (*2*) the shock compression of liquids in [49, 85, 86], and (*3*) quasi-isentropic compression in [49]; (ii) the electrical conductivity in (*4*) [57, 58], (*5*) [59], and (*6*) [41]; and (iii) (*7*) optical properties in [87].

$\rho_{max}/\rho_0 \approx 6.5$ versus the expected value of $\sigma_{max} \approx 4$) in the region $P \approx 0.5$–2 Mbar (curve *11*), nor does it lead to anomalies for $\rho \geq 1$ g/cm^3 that are typical of phase transitions.

The thermodynamics of compressed hydrogen (deuterium) assumes quite a different form if one describes H–H (D–D) interactions by means of the H–H potential introduced in [84] and extensively used in approximate calculations, adopting standard composition rules for H–H_2 interaction. In terms of the soft-sphere-model modification employed in the present study, this corresponds to a much smaller ratio of the intrinsic volumes of H and H_2:

$$\frac{d(\mathrm{H})}{d(\mathrm{H_2})} \approx 0.4 \longrightarrow \frac{2v(\mathrm{H})}{v(\mathrm{H_2})} \approx 0.13.$$

For $\rho \geq 0.3$ mol/cm^3, this choice of the intrinsic size of an atom immediately leads to "pressure-produced dissociation," which is accompanied by a dip in the shock adiabatic curve for deuterium (curve *12*).

4.2. Inert Gases

The phase diagram of xenon is shown in Fig. 16. In experiments devoted to measuring the electrical conductivity of xenon under the conditions of multiple shock-wave loads, its density, pressure, and temperatures took values of up to 9.5 g/cm^3, 120 GPa, and $(5–20) \times 10^3$ K, respectively, the electron concentration ranging up to 3×10^{22} cm^{-3} at a degree of ionization less than or equal to 0.5. In the region of maximum parameter values, the plasma was degenerate ($n_e\lambda_e^3 \sim 50$) and strongly nonideal both with respect to Coulomb ($\Gamma_D \sim 10$) and with respect to interatomic ($\Gamma_a \sim 1$) interaction.

In Fig. 16, the phase boundaries of xenon states are depicted according to [22]. The metallization of xenon under the conditions of static compression at diamond anvils was experimentally observed in [60, 88, 89] at densities of about 12.3 g/cm^3 and pressures in the range 130–150 GPa, this being in agreement with the results of the calculations previously performed in [17]. In accordance with the estimates presented in [22], the phase boundary associated with a metal–dielectric transition intersects the line of xenon melting at the triple point (Tp_2) corresponding to $P \approx 50$ GPa and $T \approx 6000$ K and terminates at critical point C_2 in the plasma region at $P \approx 10$ GPa and $T \approx 10\,000$ K. At high temperatures, the phase transition is accompanied by a sharp change in the concentration of free electrons in a narrow range of plasma densities, and this is shown in Fig. 16 by the curves corresponding to constant values of the degree of xenon-plasma ionization. A vast body of experimental data obtained from shock-wave experiments in measuring the equation of state for a xenon plasma [3, 49, 67, 85, 86, 90, 91], its optical properties [3, 49, 90, 87], and its electrical conductivity [3, 41, 49, 57–59] furnishes no indications of some of the unusual features in the behavior of xenon in this region of its parameters. The experimental observation of a sharp increase in the electrical conductivity in [41] corre-

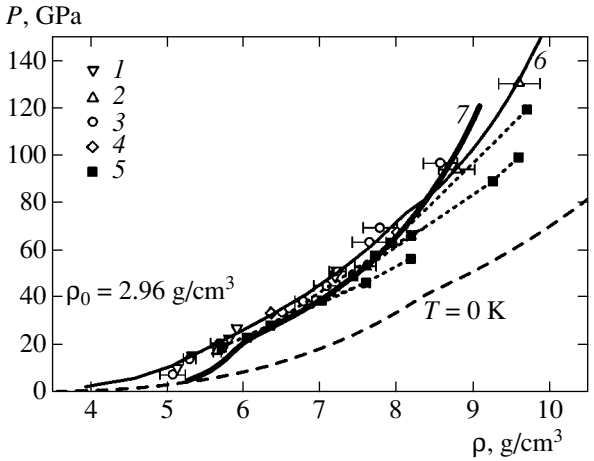

Fig. 17. Shock adiabatic curve for xenon. The experimental data were borrowed from (*1*) [85], (*2*) [86], (*3*) [49], (*4*) [90], and (*5*) [41]. The displayed theoretical results correspond to the calculations performed (*6*) in [49] and (*7*) in the present study (on the shock adiabate for $P = 86.8$ GPa, the characteristic parameters are $T = 29100$ K, $\Gamma_D = 11.9$, $n_e\lambda_e^3 = 2.19$, $n_e = 2.63 \times 10^{22}$ cm^{-3}). The dashed curve represents the "cold" curve from [49].

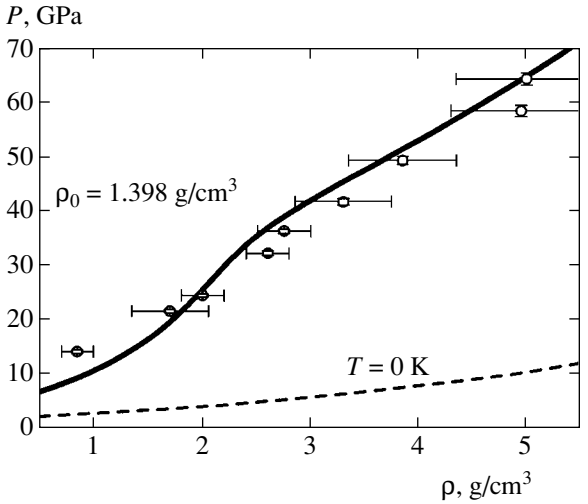

Fig. 18. Shock adiabatic curve for argon. The displayed experimental data (points) were borrowed from [92]. The solid curve was calculated in the present study (on the shock adiabate for $P = 54.6$ GPa, the characteristic parameters are $T = 17900$ K, $\Gamma_D = 8.8$, $n_e\lambda_e^3 = 0.66$, $n_e = 3.78 \times 10^{21}$ cm^{-3}). The dashed curve represents the "cold" curve.

sponds to densities of $\rho \approx 8$–10 g/cm^3 and pressures of about 100 GPa.

For shock-compressed plasmas of liquefied inert gases, the radii of the atoms involved were determined by fitting, to the cold curve ($T = 0$ K) for densities of the experimental range (see Table 2), the results of the calculations within the model specified by Eqs. (4.11) and (4.12).

The relationships between the radii of the atoms involved and of their ions of different ionization multiplicities were determined from a calculation of the relevant electron structure in the bounded-atom approximation by the Hartree–Fock method as implemented within the procedure used previously in [74].

We note that, upon a formal interpolation of the melting curve to the parameter region of our interest, some of the experimental points appear to be in the solid phase.

The applicability of the thermodynamic model considered above was tested by comparing the results derived on its basis with experimental data on the shock compression of liquid xenon that were obtained with light-gas guns [85, 86, 90] and in explosive experiments [49] (see Fig. 17). Figure 17 also displays the results of experiments devoted to multiple shock compression [41]. It can be seen that, by and large, the model reproduces experimental results satisfactorily. The discrepancy manifesting itself at low temperatures and pressures may be attributed to an insufficiently accurate approximation of xenon states in the liquid phase. The use of this model makes it possible to obtain a fairly good description of the shock adiabatic curves for liquid argon and krypton as well (see Figs. 18, 19). Here, the displayed experimental data were borrowed from [64, 92].

It should be noted that good agreement could also be reached for the measured values of the brightness temperature and the speed of sound in these substances.

The situation around a thermodynamic description of helium states proves to be more complicated since available experimental data are much scantier and since the helium phase diagram, which is shown in Fig. 20 [22], is quite unusual. The metallization of helium at low temperatures is expected to occur at extremely high pressures of $P \approx 1.1$ TPa. However, plasma phase transitions associated with a sharp change in the ionization composition of helium must take place at much lower pressures. Indeed, the estimates presented in [22] reveal that, with increasing temperature, the melting of solid helium will be accompanied by its direct transition, first, into a singly ionized plasma state (triple point Tp_3) and then into a doubly ionized state (triple point Tp_4). The plasma-phase-transition curves terminate at the critical points C_1 and C_2, the parameters of the first point ($P \approx 660$ GPa, $T \approx 35000$ K) lying very closely to

Table 2

	s	ε_{SS}, eV	r_{atom}	Q
He	12	0.01354	$2.82a_0$	1
Ar	12	0.0125	$3.2a_0$	1
Kr	12	0.0171	$3.54a_0$	1
Xe	10.5	0.0221	$3.83a_0$	1

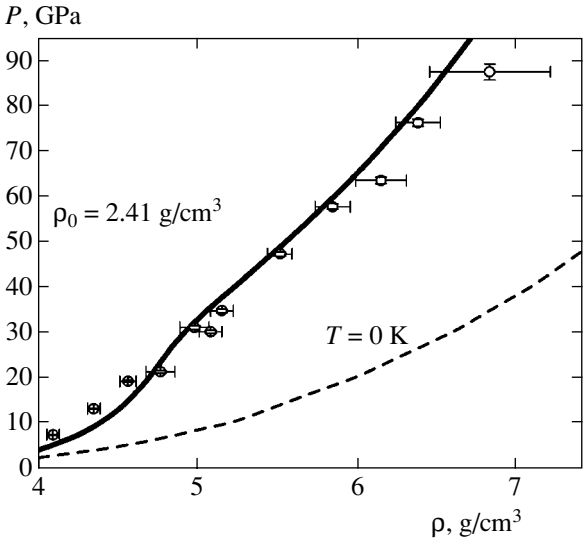

Fig. 19. Shock adiabatic curve for krypton. The displayed experimental data (points) were borrowed from [64]. The solid curve was calculated in the present study (on the shock adiabate for $P = 65.8$ GPa, the characteristic parameters are $T = 20100$ K, $\Gamma_D = 8.06$, $n_e\lambda_e^3 = 0.66$, $n_e = 4.53 \times 10^{21}$ cm^{-3}). The dashed curve represents the "cold" curve.

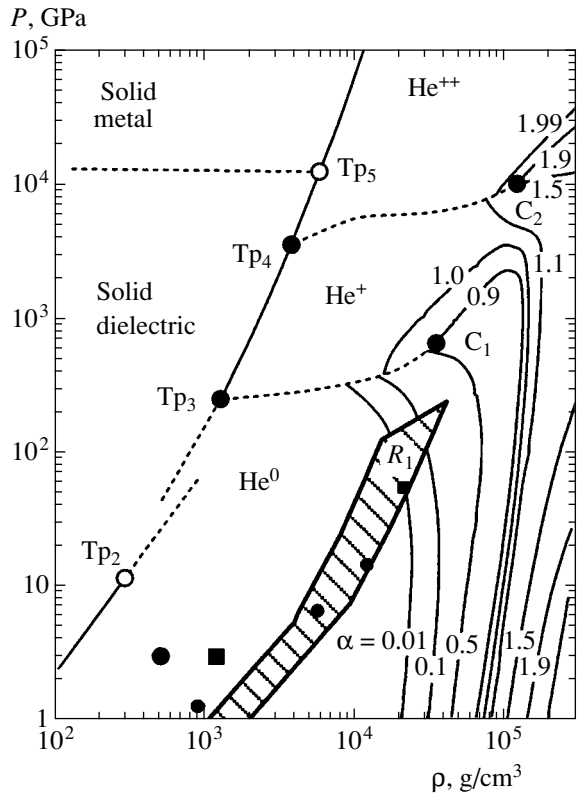

Fig. 20. Phase diagram of helium. The triple points for single and double ionization are denoted by Tp_3 and Tp_4, respectively. The critical points of plasma phase transitions for single and double ionization are C_1 and C_2, respectively. The shaded region represents helium states realized in experiments devoted to multiple shock compression [42].

the experimentally accessible region. In Fig. 20, the closed circles and the closed box represent the parameters of helium that were realized in the experiment of Nellis *et al.* [93] in, respectively, the incident and the reflected shock wave. The shaded region corresponds to helium states realized in experiments devoted to multiple shock compression [42]. It should be emphasized that the thermodynamic model satisfactorily describes the data of Nellis *et al.* [93] on the shock compression of helium in the incident and in the reflected shock wave.

5. ELECTRICAL CONDUCTIVITY OF NONIDEAL PLASMAS

In order to describe the electrical conductivity over a broad range of parameters where electrons may obey either Boltzmann or Fermi statistics, expressions (3.1)–(3.4) were combined into an interpolation expression within the τ approximation [94]; that is,

$$\sigma = \frac{4e^2(k_BT)^{-3/2}}{3\sqrt{\pi}m_e}\frac{2}{\lambda_e^2}\int_0^\infty \varepsilon^{3/2}\tau(\varepsilon)\left(-\frac{\partial f_0}{\partial\varepsilon}\right)d\varepsilon, \qquad (5.1)$$

where f_0 is the electron distribution; τ is the relaxation time,

$$\tau^{-1}(\varepsilon) = \sqrt{\frac{2\varepsilon}{m_e}}\left[\sum_j \gamma_j n_j Q_{ej}(\varepsilon) + n_a Q_{ea}(\varepsilon)\right],$$

Q_{ea} and Q_{ei} are the transport cross sections for, respectively, electron–atom and electron–ion scattering; and γ_j is a correction for electron–electron scattering. For the case where the change in statistics occurs, this correction was interpolated as [22]

$$\gamma_j = \gamma_j^B - (\gamma_j^B - 1)\frac{T_F}{\sqrt{T_F^2 + T^2}}$$

with T_F being the Fermi temperature and γ_j^B is a correction for the Boltzmann plasma.

In the Born approximation, the cross section for electron scattering on a Coulomb potential is given by

$$Q_{ej} = \frac{Z_j^2\pi e^4}{\varepsilon^2}\Lambda_j,$$

$$\Lambda_j = \frac{1}{2}\left[\ln(1+\chi_j) - \frac{\chi_j}{1+\chi_j} - \frac{Z_j b\chi_j^2}{(1+\chi_j)^2}\right], \qquad (5.2)$$

$$b = \frac{\sum_k Z_k n_k}{\sum_k Z_k^2 n_k + n_e}, \quad \chi_j = \left(\frac{r_{\max j}}{r_{\min j}}\right)^2,$$

where

$$r_{\max j} = \left[4\pi\frac{e^2}{k_B T}\sum_{i=2}^{L} Z_i^2 n_i + e^2\left(\frac{\partial n_e}{\partial \mu_e}\right)_T^{id}\right]^{-1/2} \quad (5.3)$$

is the Debye screening radius with allowance for electron degeneracy and

$$r_{\min j} = \max\{Z_j e^2/4\varepsilon, \lambda\!\!\!^{-}_e\}$$

is the minimum impact parameter.

In the case of Boltzmann statistics, the expression for Λ_j reduces to the ordinary Coulomb logarithm $\Lambda_j = \ln\chi_j$ in the limit of a weakly nonideal plasma ($\chi_j \gg 1$). For a strongly nonideal plasma ($\chi_j \sim 1$), it proves to be finite, not leading to an nonphysical divergence in the expression for the electrical conductivity, and this gives sufficient grounds to use it in tentative calculations.

Expression (5.1) for the electrical conductivity takes into account the fact that, in the region of high compressions, free electrons are partly or fully degenerate. In this case, the Boltzmann distribution of electrons is replaced by the Fermi–Dirac distribution, the radius of screening being calculated with allowance for electron degeneracy [formula (5.3)]. As a result, the approximation specified by Eq. (5.1) yields Spitzer's asymptotic behavior for a fully ionized Boltzmann plasma. For a full degeneracy of electrons, $\gamma = 1$ and the electrical conductivity tends to expression (3.4), while for a partly ionized plasma, the estimate in (3.1) is valid.

The effect of pressure-produced ionization is the most pronounced in the case of hydrogen (see Fig. 7). In this figure, our data on the quasi-isentropic compression of liquid and gaseous hydrogen in (points *1*) planar and (points *2)* cylindrical geometries are contrasted against the results of its compression by means of a light-gas gun (points *5*) [35], as well as against the results of explosive cylindrical compression by means of an axial magnetic field (points *3* and *4* from [29] and [30], respectively). In view of a light molecular weight, the multiple shock compression of hydrogen is accompanied by a relatively weak heating of this substance—even at maximum pressures of 1 to 10 Mbar, the characteristic values of its temperature do not exceed $T = 10^4$ K and this favors the "cold"-ionization" regime ($k_B T \ll I$). For hydrogen compressed to densities of $\rho \approx 0.01$–1.2 g/cm^3 and heated to $T \sim 10^4$ K at pressures below 15 Mbar, a wide spectrum of plasma states characterized by a fully developed ionization, $\alpha = 1$, and a high electron concentration of $n_e \approx 2 \times 10^{23}$ cm^{-3} were realized in experiments. At maximum compressions, the plasma in question is degenerate, $n_e \lambda\!\!\!^{-3}_e \approx 200$, and is strongly nonideal both with respect to Coulomb ($\Gamma_D \approx 10$) and to interatomic ($\Gamma_a \approx 1$) interaction.

It is interesting to note that an extrapolation of the simplest plasma models to this region of strong nonideality leads to the thermodynamic instability of Debye–Hückel models (Coulomb collapse)—arrow DC in Fig. 7—and to the divergence of Spitzer's formula—arrow SC. The first of these approximations is depicted by the DHA curve in Fig. 7 and predicts pressure-produced ionization at a density value that is approximately two orders of magnitude lower than its experimental counterpart.

The shock compression of matter leads to the overlap of the wave functions for neighboring atoms and, hence, to the percolation conductivity mechanism [55], which is described in terms of the density-dependent reduction of the ionization potential (curve *P*),

$$\frac{\Delta I}{k_B T} = e^2\left(\frac{4\pi n}{\zeta}\right)^{1/3}.$$

A decrease in the ionization potential with increasing density is also predicted by the Mott model [5], which was used in [22] to construct a semiempirical wide-range model of ionization equilibrium and transport properties (curve *M* in Fig. 7) of compressed and hot matter; that is,

$$\frac{\Delta I}{k_B T} = -I\ln\left(1 + \exp\left(-2\frac{R - a(I)}{\Delta(I)}\right)\right).$$

Here, the parameters *a*, *R*, and Δ were chosen is such a way as to reproduce experimental data on pressure-produced ionization of alkali metals. It can be seen that the proposed approximations provide a good qualitative description of experimental results.

By using the ring (Debye) approximation in a grand canonical ensemble of statistical mechanics (LDH) to describe Coulomb nonideality, one can reduce the discrepancy between the theoretical and experimental results down to one order of magnitude. The remaining discrepancy can be removed by introducing the hard-sphere model to describe the short-range repulsion of atoms and ions (curve HS) and by taking into account the compression-induced change in the energy spectrum of atoms and ions within a simplified version of the model considered at the beginning of Section 4 (curve CA). An attempt at taking into account the jump-like character of the electrical conductivity in nonideal plasmas was made Redmer *et al.* [56]. The results of their calculations are represented by curve *R* in Fig. 7. The QMC curve corresponds to the calculation of the electrical conductivity by the quantum Monte Carlo method in [24].

Figures 8–11 display the results obtained by studying the electrical conductivity of shock-compressed Xe, Ar, Kr, and He. In just the same way as in the case of hydrogen, one observes here, at "low" temperatures ($k_B T \ll I$), the pressure-produced-ionization effect

occurring at higher plasma densities of $\rho \approx 1$–10 g/cm^3. For many-electron atoms, it is also natural to expect that, as compression is increased further, the first ionization induced by pressure will be followed by the next steps of multiple ionization with the emergence of subsequent additional boundaries of phase transitions corresponding to the second and still higher steps of ionization (see Fig. 20). Unfortunately, experimental investigation of these regimes of multiple ionization is presently beyond the capabilities of the explosive experimental equipment that we have at our disposal.

Along with the results of multiple ("cold") compression, Figs. 8–11 exhibit data obtained previously in [57–59] by measuring the electrical conductivity of singly and doubly compressed plasmas. Almost one order of magnitude higher temperatures are realized in experiments with inert gases than in those with hydrogen, the effects of thermal ionization becoming dominant in the former. The role of these effects becomes more pronounced with the increasing molecular weight of the substances being studied, and this is especially spectacular for xenon (see Fig. 8). It can be seen that, upon thermal ionization [$T \approx (2$–$10) \times 10^4$ K], a high level of the electrical conductivity [about 10^3 $(\Omega\ \mathrm{cm})^{-1}$] is achieved even at low densities of $\rho < 1$ g/cm^3; on the other hand, only at extremely high compressions up to densities of $\rho \sim 10$ g/cm^3 is the same level of electrical conductivity ensured in cold ($T \sim 10^4$ K) matter under conditions of pressure-produced ionization. It can also be seen that, with increasing molecular weight of substances, the jump in the electrical conductivity due to pressure-produced ionization decreases, falling down to only two orders of magnitude in xenon. It is worthy of note that the values of the electrical conductivity of xenon plasma that were measured in multiple-shock-compression experiments are close to those obtained under static conditions of diamond anvils (the cross in Fig. 8).

It is important to note that some of the models discussed here lose thermodynamic stability in the region of the experiments in question. With some qualifications, this may be considered as an indication of a plasma first-order phase transition leading to the stratification of a strongly nonideal plasma into phases characterized by different degrees of ionization and compressibility [1–4]. A sharp increase in the electrical conductivity of a dense plasma in our experiments suggests the occurrence of such a phase transition.

6. CONCLUSIONS

Thus, we believe that it is necessary to conduct experiments aimed at directly measuring the concentration of free electrons in the region of the possible phase transition. The first studies devoted to such measurements in plasmas have already been performed by the authors.

The chemical model used to describe the equation of state for plasmas in the region of ultrahigh densities yields satisfactory results in this region, which is not traditional for the model in question; nonetheless, the model calls for further refinements associated with taking into account the rearrangement of the internal structure of plasma particles under the conditions of ultradense plasma states.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 00-15-96738 and 00-02-17550), by the Ministry for Industry and Science of the Russian Federation (state contract no. 40.009.1.1.1192), and by the Presidium of the Russian Academy of Sciences (Interdisciplinary Program for Scientific Research into the Physics and Chemistry of Extreme States of Matter).

REFERENCES

1. *Encyclopedia of Low-Temperature Plasma*, Ed. by V. E. Fortov (Nauka, Moscow, 2000).
2. V. K. Gryaznov, I. L. Iosilevskiĭ, Yu. G. Krasnikov, *et al.*, *Thermal Properties of Working Media of Gas-Phase Nuclear Reactor*, Ed. by V. M. Ievlev (Atomizdat, Moscow, 1980).
3. V. E. Fortov and I. T. Yakubov, *Nonideal Plasma* (Énergoatomizdat, Moscow, 1994).
4. L. D. Landau and Yu. B. Zel'dovich, Zh. Éksp. Teor. Fiz. **14**, 32 (1944).
5. N. F. Mott and E. A. Davis, *Electron Processes in Non-Crystalline Materials*, 2nd ed. (Clarendon Press, Oxford, 1979; Mir, Moscow, 1982).
6. F. Hensel and E. U. Frank, Rev. Mod. Phys. **40**, 697 (1968).
7. A. G. Barsukov, P. E. Kovrov, V. M. Kuligin, *et al.*, in *Proceedings of 8th IAEA Conference* (1989), Vol. 1, p. 83.
8. V. A. Alekseev and A. A. Vedenov, Usp. Fiz. Nauk **102**, 665 (1970) [Sov. Phys.–Usp. **13**, 522 (1970)].
9. A. W. DeSilva and J. D. Katsouros, Phys. Rev. E **57**, 5945 (1998).
10. A. V. Bushman, B. L. Glushak, V. K. Gryaznov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **44**, 375 (1986) [JETP Lett. **44**, 480 (1986)].
11. E. G. Maksimov, Usp. Fiz. Nauk **169**, 1223 (1999) [Phys.–Usp. **42**, 1121 (1999)].
12. E. Wigner and H. B. Huntington, J. Chem. Phys. **3**, 764 (1935).
13. A. A. Abrikosov, Astron. Zh. **31**, 112 (1954).
14. N. W. Ashkroft, Phys. Rev. Lett. **21**, 1748 (1968).
15. E. G. Brovman, Yu. Kogan, and A. Kholac, Zh. Éksp. Teor. Fiz. **61**, 2429 (1971) [Sov. Phys. JETP **34**, 1300 (1971)].
16. V. P. Trubitsyn, Fiz. Tverd. Tela (Leningrad) **8**, 862 (1966) [Sov. Phys. Solid State **8**, 688 (1966)].

17. M. Ross and A. K. McMahan, Phys. Rev. B **21**, 1658 (1980).
18. D. A. Young, A. K. McMahan, and M. Ross, Phys. Rev. B **24**, 5119 (1981).
19. J. C. Boettger, Phys. Rev. B **33**, 6788 (1986).
20. K. A. Goettel, J. H. Eggert, I. F. Silvera, and W. C. Moss, Phys. Rev. Lett. **62**, 665 (1989).
21. E. N. Avrorin, B. K. Vodolaga, B. A. Simonenko, and V. E. Fortov, Usp. Fiz. Nauk **163** (5), 1 (1993) [Phys.–Usp. **36**, 337 (1993)].
22. W. Ebeling, A. Förster, V. Fortov, V. Gryaznov, and A. Polishchuk, *Thermophysical Properties of Hot Dense Plasmas* (Teubner, Stuttgart, 1991).
23. D. Saumon and G. Chabrier, Phys. Rev. Lett. **62**, 2397 (1989); D. Saumon and G. Chabrier, Phys. Rev. A **46**, 2084 (1992).
24. I. A. Mulenko, E. N. Olejnikova, A. L. Khomkin, *et al.*, Phys. Lett. A **289**, 141 (2001).
25. G. É. Norman and A. N. Starostin, Teplofiz. Vys. Temp. **8**, 413 (1970).
26. M. A. Mochalov and O. N. Kuznetsov, in *Abstracts of III Khariton Topical Readings* (Sarov, 2001), p. 108.
27. F. V. Grigor'ev, S. B. Kormer, O. L. Mikhaĭlova, *et al.*, Zh. Éksp. Teor. Fiz. **75**, 1683 (1978) [Sov. Phys. JETP **48**, 847 (1978)].
28. S. I. Belov, G. V. Boriskov, A. I. Bykov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 508 (2002) [JETP Lett. **76**, 433 (2002)].
29. R. S. Hawke, T. J. Burgers, D. E. Duerre, *et al.*, Phys. Rev. Lett. **41**, 994 (1978).
30. A. I. Pavlovskiĭ, G. V. Boriskov, *et al.*, in *Megagauss Technology and Pulsed Power Applications*, Ed. by C. M. Fowler (Plenum, New York and London, 1987).
31. L. B. Da Silva, P. Celliers, G. W. Collins, *et al.*, Phys. Rev. Lett. **78**, 483 (1997).
32. A. N. Mostovych, Y. Chan, T. Lehecha, *et al.*, Phys. Rev. Lett. **85**, 3870 (2000).
33. M. D. Knudsen, D. L. Hanson, J. B. Bailey, *et al.*, Phys. Rev. Lett. **90**, 035505 (2003).
34. N. S. Holmes, M. Ross, and W. J. Nellis, Phys. Rev. B **52**, 15835 (1995).
35. W. J. Nellis, S. T. Weir, and A. C. Mitchell, Rev. High Pressure Sci. Technol. **7**, 870 (1998); S. T. Weir, A. C. Mitchell, and W. J. Nellis, Phys. Rev. Lett. **76**, 1860 (1996).
36. V. E. Fortov, V. Ya. Ternovoĭ, S. V. Kvitov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 874 (1999) [JETP Lett. **69**, 926 (1999)]; V. Ya. Ternovoĭ, A. S. Filimonov, V. E. Fortov, *et al.*, Physica B (Amsterdam) **265**, 6 (1999).
37. D. Beule, W. Ebeling, A. Förster, *et al.*, Phys. Rev. B **59**, 14177 (1999).
38. M. Robnic and W. Kundt, Astron. Astrophys. **120**, 227 (1983).
39. P. Haronska, D. Krempt, *et al.*, Wiss. Z. Wilhelm-Pieck-Univ. Rostock, Naturwiss. Reihe **36**, 98 (1987).
40. W. J. Nellis, Planet. Space Sci. **48**, 671 (2000).
41. V. B. Mintsev, V. Ya. Ternovoĭ, V. K. Gryaznov, *et al.*, in *Shock Compression of Condensed Matter-1999*, Ed. by S. C. Schmidt, D. P. Dandekar, and J. W. Forbes (Woodbury, New York, 2000), p. 987.
42. V. Ya. Ternovoĭ, A. S. Filimonov, A. A. Pyalling, *et al.*, in *Shock Compression of Condensed Matter-2001*, Ed. by M. D. Furnish, N. N. Thadhani, and Y. Horie (AIP Press, New York, 2002), p. 107.
43. V. Ya. Ternovoĭ, in *Nonstationary Problems in Hydrodynamics* (Inst. Gidrodin., Sib. Otd. Akad. Nauk SSSR, 1980), No. 48, p. 141.
44. S. T. Weir, A. C. Mitchell, and W. J. Nellis, J. Appl. Phys. **80**, 1522 (1996).
45. L. M. Barker, T. G. Trucano, J. L. Wize, and J. R. Asay, in *Proceedings of Conference on Shock Waves in Condensed Matter-85* (Plenum, New York, 1986), p. 455.
46. H. Juranek, R. Redmer, G. Roepke, *et al.*, Plasma Phys. **39**, 251 (1999).
47. A. V. Bushman, I. V. Lomonosov, and V. E. Fortov, *Equations of State of Metals at High Energy Densities* (Inst. Probl. Khim. Fiz., Chernogolovka, 1992).
48. I. A. Adamskaya, F. V. Grigor'ev, O. L. Mikhaĭlova, *et al.*, Zh. Éksp. Teor. Fiz. **93**, 647 (1987) [Sov. Phys. JETP **66**, 366 (1987)].
49. V. D. Urlin, M. A. Mochalov, and O. L. Mikhailova, High Press. Res. **8**, 595 (1992).
50. V. D. Urlin, M. A. Mochalov, and O. L. Mikhaĭlova, Zh. Éksp. Teor. Fiz. **111**, 2099 (1997) [JETP **84**, 1145 (1997)].
51. E. I. Zababakhin, *Mechanics in USSR for 50 Years* (Nauka, Moscow, 1979).
52. A. I. Pavlovskiĭ, G. D. Kuleshov, G. V. Sklizkov, *et al.*, Dokl. Akad. Nauk SSSR **160**, 68 (1965) [Sov. Phys. Dokl. **10**, 30 (1965)].
53. A. A. Brish, M. S. Tarasov, and V. A. Tsukerman, Zh. Éksp. Teor. Fiz. **38**, 22 (1960) [Sov. Phys. JETP **11**, 15 (1960)].
54. V. M. Zamalin, G. É. Norman, and V. S. Filinov, *Monte Carlo Method in Statistical Thermodynamics* (Nauka, Moscow, 1977; Mir, Moscow, 1977).
55. A. A. Likal'ter, Zh. Éksp. Teor. Fiz. **113**, 1094 (1998) [JETP **86**, 598 (1998)].
56. R. Redmer, G. Roepke, S. Kuhlbrodt, and H. Reinholz, Contrib. Plasma Phys. **41**, 163 (2001).
57. 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)].
58. V. B. Mintsev and V. E. Fortov, Pis'ma Zh. Éksp. Teor. Fiz. **30**, 401 (1979) [JETP Lett. **30**, 375 (1979)].
59. V. B. Mintsev, V. E. Fortov, and V. K. Gryaznov, Zh. Éksp. Teor. Fiz. **79**, 116 (1980) [Sov. Phys. JETP **52**, 59 (1980)].
60. M. I. Eremets, E. A. Gregoryanz, V. V. Struzhkin, *et al.*, Phys. Rev. Lett. **85**, 2797 (2000).
61. S. V. Dudin, V. E. Fortov, *et al.*, in *Shock Compression of Condensed Matter-1997*, Ed. by S. C. Schmidt, D. P. Dandekar, and J. W. Forbes (AIP Press, New York, 1998), p. 793.

62. L. A. Gatilov, V. D. Glukhodedov, F. V. Grigor'ev, *et al.*, Prikl. Mekh. Tekh. Fiz., No. 1, 99 (1985).
63. M. I. Dolotenko, A. I. Bykov, *et al.*, in *Megagauss and Megaamper Pulsed Power and Related Topics*, Ed. by V. Chernishev, V. Selimir, and L. Plyashkevitch (VHI-IEF, Sarov, 1997), p. 805.
64. V. D. Glukhodedov, S. I. Kirshanov, T. S. Lebedeva, and M. A. Mochalov, Zh. Éksp. Teor. Fiz. **116**, 551 (1999) [JETP **89**, 292 (1999)].
65. L. I. Veeser, C. A. Ekdah, H. Oona, *et al.*, in *Abstracts of VIII International Conference on Megagauss Magnetic Field Generation and Related Topics, Tallahassee* (1998), p. 239.
66. K. Seeger, *Semiconductor Physics* (Springer, Berlin, 1974; Mir, Moscow, 1977).
67. V. K. Gryaznov, M. V. Zhernokletov, V. N. Zubarev, *et al.*, Zh. Éksp. Teor. Fiz. **78**, 573 (1980) [Sov. Phys. JETP **51**, 288 (1980)].
68. J. M. Zaiman, *Principles of the Theory of Solids*, 2nd ed. (Cambridge Univ. Press, London, 1972; Mir, Moscow, 1966).
69. W. J. Nellis, S. T. Weir, N. C. Holmes, *et al.*, UCRL-125039, LLNL (1996).
70. V. K. Gryaznov, I. L. Iosilevskiĭ, and V. E. Fortov, Pis'ma Zh. Tekh. Fiz. **8**, 1378 (1982) [Sov. Tech. Phys. Lett. **8**, 592 (1982)].
71. V. K. Gryaznov and V. E. Fortov, Teplofiz. Vys. Temp. **25**, 1208 (1987).
72. V. K. Gryaznov, M. V. Zhernokletov, I. L. Iosilevskiĭ, *et al.*, Zh. Éksp. Teor. Fiz. **114**, 1242 (1998) [JETP **87**, 678 (1998)].
73. I. L. Iosilevskiĭ, Teplofiz. Vys. Temp. **18**, 355 (1980).
74. V. K. Gryaznov, I. L. Iosilevskiĭ, and V. E. Fortov, in *Shock Waves and Extremal States of Substance*, Ed. by V. E. Fortov, L. V. Al'tshuler, R. F. Trunin, and A. I. Funtikov (Nauka, Moscow, 2000).
75. D. A. Young, UCRL-52352, LLNL (1977).
76. A. G. Gaydon and I. R. Hurle, *The Shock Tube in High-Temperature Chemical Physics* (Chapman and Hall, London, 1963; Mir, Moscow, 1966).
77. V. B. Mintsev and V. E. Fortov, Teplofiz. Vys. Temp. **20**, 745 (1982).
78. E. S. Yakub, Teplofiz. Vys. Temp. **28**, 664 (1990); Physica B (Amsterdam) **265**, 31 (1999).
79. C. Pierleoni, D. M. Ceperley, B. Bernu, and W. R. Magro, Phys. Rev. Lett. **73**, 2145 (1994).
80. L. Collins, I. Kwon, J. Kress, *et al.*, Phys. Rev. E **52**, 6202 (1995); S. Günter and A. Kötnies, Phys. Rev. E **55**, 907 (1997).
81. M. Knaup, P. Reinhard, and C. Topffer, Contrib. Plasma Phys. **39**, 57 (1999).
82. G. I. Kerley, in *Theoretical Equation of State for Deuterium* (National Technical Information Service, Springfield, VA, 1972), NTIS Document No. LA-47766.
83. M. Ross, Phys. Rev. B **58**, 669 (1998).
84. F. Ree, M. Ross, and D. Young, J. Chem. Phys. **79**, 1487 (1983).
85. R. Keeler, M. Van Thiel, and B. Alder, Physica (Amsterdam) **31**, 1437 (1965).
86. W. Nellis, M. Van Thiel, and A. Mitchel, Phys. Rev. Lett. **48**, 816 (1982).
87. M. I. Kulish, V. K. Gryaznov, V. B. Mintsev, *et al.*, Teplofiz. Vys. Temp. **33**, 967 (1995).
88. K. A. Goettel, J. H. Eggert, I. F. Silvera, and W. C. Moss, Phys. Rev. Lett. **62**, 665 (1989).
89. R. Reichlin, K. E. Brister, A. K. McMahan, *et al.*, Phys. Rev. Lett. **62**, 669 (1989).
90. H. B. Radousky and M. Ross, Phys. Lett. A **129**, 43 (1988).
91. V. E. Fortov, V. K. Gryaznov, V. B. Mintsev, *et al.*, Contrib. Plasma Phys. **41**, 215 (2001).
92. F. V. Grigor'ev, S. B. Kormer, O. L. Mikhaĭlova, *et al.*, Zh. Éksp. Teor. Fiz. **88**, 1271 (1985) [Sov. Phys. JETP **61**, 751 (1985)].
93. W. J. Nellis, N. C. Holmes, A. C. Mitchell, *et al.*, Phys. Rev. Lett. **53**, 1248 (1984).
94. V. K. Gryaznov, Yu. V. Ivanov, A. N. Starostin, and V. E. Fortov, Teplofiz. Vys. Temp. **14**, 643 (1976).

Translated by A. Isaakyan

Journal of Experimental and Theoretical Physics, Vol. 97, No. 2, 2003, pp. 279–289.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 2, 2003, pp. 310–321.

SOLIDS
Structure

The Magnetoelastic Contribution to Thermal Expansion of Rare-Earth Metal Scheelites $RLiF_4$ (R = Tb–Yb)

Z. A. Kazei*, O. A. Shishkina, R. I. Chanieva, R. Yu. Abdulsabirov, and S. L. Korableva

Moscow State University, Vorob'evy gory, Moscow, 119899 Russia
**e-mail: kazei@plms.phys.msu.su*

Received January 29, 2003

Abstract—The paper presents systematic experimental and theoretical studies of thermal expansion for rare-earth metal scheelites $RLiF_4$ (R = Tb–Ho, Tm, and Lu). Pronounced thermal expansion anomalies were observed. The magnetoelastic contributions were determined taking into account corrections for changes in the phonon contribution in the $RLiF_4$ series according to the Debye thermal expansion model. The calculated multipole moments of various orders for various rare-earth metal ions were compared to analyze the applicability of the quadrupole approximation to totally symmetric modes in the scheelite structure. For some ions (Ho and Tm), the magnetoelastic contributions to thermal expansion could not be described by the temperature dependences of their quadrupole moments, that is, multipole moments made considerable contributions. The totally symmetric magnetoelastic coefficients for the scheelite structure were determined from the experimental data on magnetoelastic contributions. These coefficients were compared with those for the zircon structure. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Magnetoelastic interaction in rare-earth metal compounds, which is strong compared with transition metal compounds with open *d* shells, plays an important role in the formation of their magnetic characteristics and is responsible for various magnetoelastic effects, for instance, for structural phase transitions of a Jahn–Teller nature [1]. One of the simplest spontaneous effects is the magnetoelastic contribution to thermal expansion, which is not accompanied by a descent in crystal symmetry and is described by totally symmetric magnetoelastic modes. Magnetoelastic effects are related to additional lattice deformations caused by changes in the magnetic state of ions and can also manifest themselves in the absence of a magnetic field, for instance, in the form of spontaneous magnetostriction accompanying crystal magnetic ordering or as a deformation caused by magnetoelastic coupling between the anisotropic 4*f* shell electron cloud and the lattice. Changes in the anisotropy of electron cloud distributions during cooling caused by a decrease in the population of excited levels are responsible for magnetoelastic anomalies of lattice parameters.

In the quadrupole approximation, the magnetoelastic effects depend on magnetoelastic coefficients, which are a characteristic of the whole series of rare-earth metal compounds, on the one hand, and on the response function of the rare-earth metal ion determined by its electronic structure, on the other. The main problem in studying magnetoelastic phenomena is to determine magnetoelastic coefficients. The currently available theoretical calculations of magnetoelastic coefficients performed, for instance, according to the model of exchange charges [2], contain many unknown parameters, which vary in fairly broad ranges within the framework of the model. The calculations of first principles require knowledge of many structural and electronic parameters of the compound under study. In addition, such calculations cannot be consistently compared with experiment to substantiate the correctness of the selected model and calculation results. For this reason, calculated magnetoelastic coefficients can only be considered estimates. Currently, the approach based on treatment of magnetoelastic coefficients as phenomenological parameters and their determination from comparison with experiment has gained broader acceptance.

In this work, we studied the magnetoelastic contribution to thermal expansion of rare-earth metal scheelites. Our goal was to explore totally symmetric magnetoelastic modes in the structure of scheelite and analyze the applicability of the quadrupole approximation to their description. In [3, 4], similar studies were performed for rare-earth metal vanadates and phosphates with zircon structures. The crystal field parameters for rare-earth metal zircons and scheelites are substantially different. The same rare-earth metal ions therefore have different spectra and, accordingly, behave quite differently in these tetragonal structures. For this reason, systematic studies of magnetoelastic interactions for a series of rare-earth metal scheelites and their comparison with those in rare-earth metal zir-

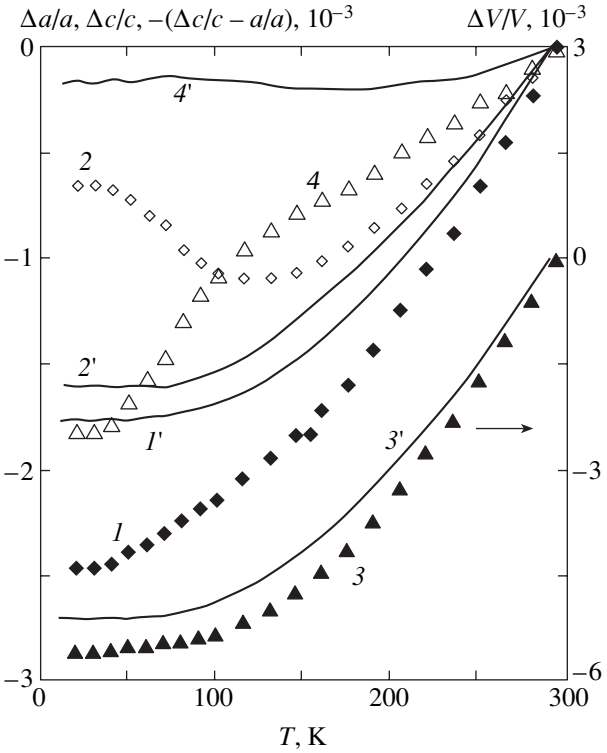

Fig. 1. Relative temperature-induced changes in the (*1*) $\Delta a/a$ and (*2*) $\Delta c/c$ parameters, (*3*) volume $\Delta V/V$, and (*4*) degree of tetragonality ($\Delta c/c - \Delta a/a$) for the $TbLiF_4$ unit cell; for comparison, similar dependences of (*1'*) $\Delta a/a$, (*2'*) $\Delta c/c$, (*3'*) $\Delta V/V$, and (*4'*) ($\Delta c/c - \Delta a/a$) are shown for $LuLiF_4$.

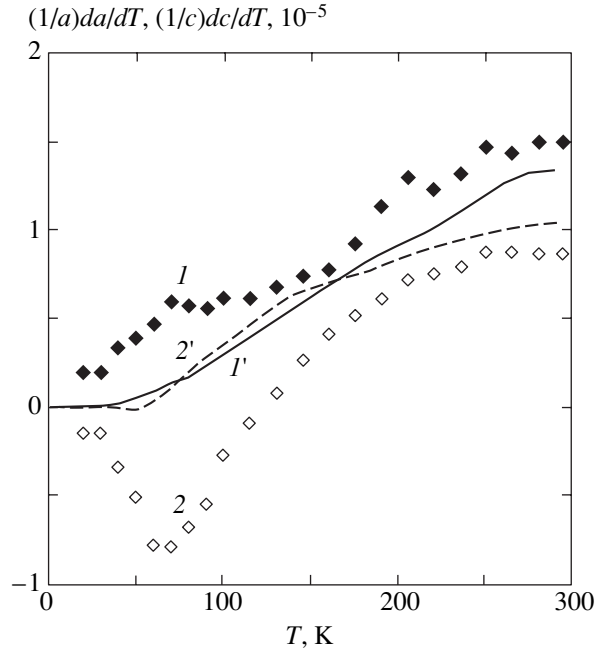

Fig. 2. Thermal expansion coefficients (*1*) normally $(1/a)da/dT$ and (*2*) along $(1/c)dc/dT$ to the tetragonal axis of $TbLiF_4$; for comparison, similar dependences of (*1'*) $(1/a)da/dT$ and (*2'*) $(1/c)dc/dT$ are shown for $LuLiF_4$.

cons taking into account the differences in the crystal fields and structure parameters is of obvious interest.

2. SAMPLES AND PROCEDURE FOR MEASUREMENTS

Thermal expansion of $RLiF_4$ scheelites (R = Tb–Ho, Tm, and Lu) was studied on a Gegerflex diffractometer with a CF-107 (Oxford Instruments) flow helium cryostat. Compound single crystals were grown by the Bridgman technique. Plates with an area of 1×4 mm^2 and thickness of 0.5–1 mm were cut from these crystals normally to the *a* and *c* axes. Temperature measurements of the *a* and *c* parameters were performed based on the (600) (CoK_β radiation, $2\theta \sim 130°$–$140°$) and (0012) (FeK_β radiation, $2\theta \sim 150°$–$160°$) reflections. The relative accuracy of measurements was $\delta a/a \approx \delta c/c \approx \cot\theta\Delta\theta \approx 10^{-5}$ ($\Delta\theta \approx 0.003°$).

3. RESULTS

3.1. Thermal Expansion of $RLiF_4$ (R = Tb–Ho, Tm, Lu)

We performed measurements for $RLiF_4$ crystals with R = Tb, Dy, Ho, and Tm. According to the calculations, the magnetoelastic contribution to thermal expansion of these crystals should be largest in magnitude. Consider the special features of thermal expansion of rare-earth metal scheelites for the example of $TbLiF_4$. Figure 1 shows the experimental temperature dependences of the *a* and *c* parameters [$\Delta a/a = a(T)/a_0 - 1$, $\Delta c/c = c(T)/c_0 - 1$, $a_0 = a(290\text{ K})$, $c_0 = c(290\text{ K})$, curves *1* and *2*] and volume $\Delta V/V = V(T)/V_0 - 1$, $V_0 = a_0^2 c_0$, curve *3*] for the tetragonal unit cell of $TbLiF_4$. The normalization that we use rules out the systematic measurement error related to crystal alignment. Similar dependences for $LuLiF_4$ are shown by solid lines in the same figure (curves *1'*, *2'*, *3'*). These dependences are used to determine the phonon contribution to the thermal expansion of $RLiF_4$. The isomorphous $LuLiF_4$ compound does not contain magnetic ions, and its temperature dependences of $\Delta c/c$ and $\Delta a/a$ have the usual Debye form; these values equal 16×10^{-3} and 18×10^{-3}, respectively, at $T = 20$ K.

The $\Delta a/a$ curve for the Tb scheelite goes below the corresponding curve for the Lu compound at all temperatures, and the temperature dependence of $\Delta a/a$ differs from the Debye dependence, which is characterized by saturation below 100–80 K. The $\Delta a/a$ curve for $TbLiF_4$ has a singularity at $T \sim 80$ K corresponding to a local minimum of the thermal expansion coefficient $\alpha_a = (1/a)da/dT$ (Fig. 2).

The temperature dependence of $\Delta c/c$ for the Tb scheelite has a nonmonotonic character and contains a minimum in the region of 120 K. The thermal expansion coefficient along the tetragonal axis, α_c =

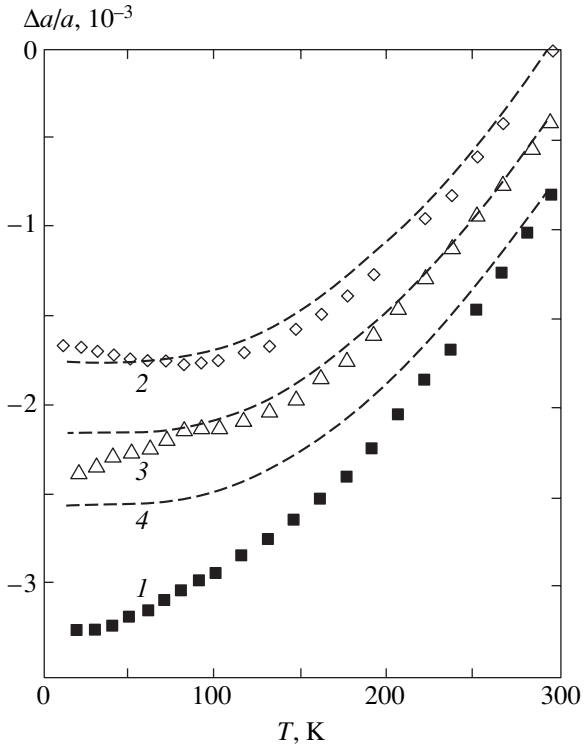

Fig. 3. Relative temperature-induced changes in the $\Delta a/a = a(T)/a_0 - 1$ parameter for the tetragonal unit cells of rare-earth metal scheelites [$a_0 = a(290\ \text{K})$]: (*1*) $TbLiF_4$ (a_0 = 5.197 Å), (*2*) $DyLiF_4$ (a_0 = 5.186 Å), (*3*) $TmLiF_4$ (a_0 = 5.148 Å), and (*4*) $LuLiF_4$ (a_0 = 5.126 Å); each subsequent curve is shifted along the *y* axis by -4×10^{-4} from the previous one.

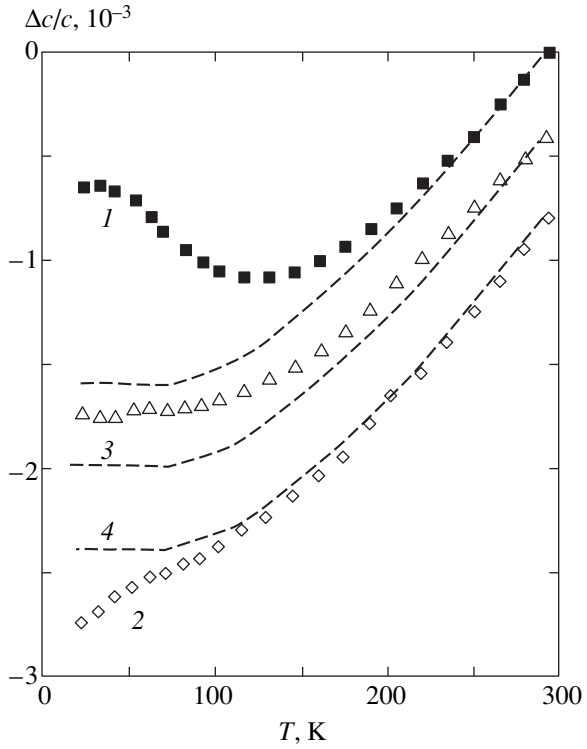

Fig. 4. Relative temperature-induced changes in the $\Delta c/c = c(T)/c_0 - 1$ parameter for the tetragonal unit cells of rare-earth metal scheelites [$c_0 = c(290\ \text{K})$]: (*1*) $TbLiF_4$ (c_0 = 10.87 Å), (*2*) $DyLiF_4$ (c_0 = 10.81 Å), (*3*) $TmLiF_4$ (c_0 = 10.65 Å), and (*4*) $LuLiF_4$ (c_0 = 10.55 Å); each subsequent curve is shifted along the *y* axis by -4×10^{-4} from the previous one.

$(1/c)dc/dT$, changes sign at $T \sim 120$ K and has an extremum at about 80 K, which coincides with the α_a coefficient minimum (Fig. 2). The magnetoelastic contribution along both $TbLiF_4$ axes is insignificant above 200 K but sharply increases at lower temperatures. In the $TbLiF_4$ crystal, the magnetoelastic contribution is positive for the *c* parameter and negative for *a*. This increases the thermal expansion anisotropy along and normally to the tetragonal axis. The anomalies of the $\Delta a/a$ and $\Delta c/c$ curves for $TbLiF_4$ cause the volume anomaly $\Delta V/V = \Delta c/c + 2\Delta a/a \sim 5 \times 10^{-3}$. The magnetoelastic contribution to the volume deformation determined with respect to the Lu compound is, however, insignificant, about 5×10^{-4} (cf. curves *3* and *3*' in Fig. 1).

The structure of $TbLiF_4$ remains tetragonal in the whole temperature range of our experiments. Only the degree of tetragonality determined by the difference $\Delta c/c - \Delta a/a$ (curve *4*) changes as temperature decreases. Note that, for the Lu compound, thermal expansions along and normally to the tetragonal axis are comparable in magnitude and the degree of tetragonality changes insignificantly as temperature varies (curve *4*'). Temperature-induced changes in the degree of tetragonality of the Tb scheelite are more substantial. The representation of the thermal expansion in terms of the totally symmetric deformations $\Delta\varepsilon^{\alpha 1}/\varepsilon^{\alpha 1} \sim \Delta V/V$ and $\Delta\varepsilon^{\alpha 2}/\varepsilon^{\alpha 2} \sim (\Delta c/c - \Delta a/a)$ is useful because the magnetoelastic contributions of these deformations give direct information about the value and sign of the corresponding magnetoelastic coefficients (see below). In the Tb scheelite as distinguished from rare-earth metal vanadates [3], both magnetoelastic contributions are negative, and the contribution to the volume deformation is substantially smaller in magnitude. This is explained by the ratio between the magnetoelastic contributions along the *c* and *a* axes in zircon and scheelite structures.

Similar lattice parameter anomalies are observed for the $DyLiF_4$, $HoLiF_4$, and $TmLiF_4$ scheelites (Figs. 3, 4). These anomalies can be quantitatively estimated from the difference between the $\Delta a/a$ and $\Delta c/c$ values for a given scheelite and for $LuLiF_4$. The magnetoelastic contributions in the scheelites are negative at all temperatures for the *a* parameter but positive for the *c* parameter, at least at not very low temperatures. Figures 3 and 4 show that the magnetoelastic contribution is maximum for the Tb scheelite and decreases in the

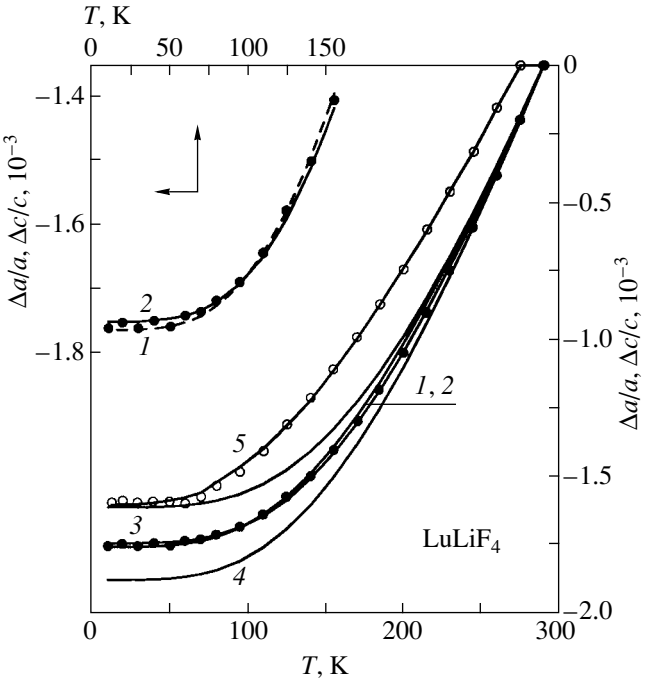

Fig. 5. Relative temperature-induced changes in the (*1*–*4*) $\Delta a/a$ and (*5*) $\Delta c/c$ parameters, experimental ($LuLiF_4$, symbols) and calculated for the tetragonal scheelite unit cell; calculations by (1) with the parameters (*1*) $T_D = 790$ K and $a_0 = 18.8 \times 10^{-6}$, (*2*) $T_D = 710$ K and $a_0 = 16.8 \times 10^{-6}$, (*3*, *4*) $T_D = 740$, 840 K and $a_0 = 18.8 \times 10^{-6}$, and (*5*) $T_D = 600$ K and $a_0 = 13.0 \times 10^{-6}$.

Dy and Tm compounds. The thermal expansion of $DyLiF_4$ along both axes is larger than that of $LuLiF_4$ at $T > 100$ K, which is, in our view, evidence of different phonon contributions in the Dy and Lu scheelites. In addition, $|\Delta a/a| > |\Delta c/c|$ for $DyLiF_4$ at high temperatures, but this ratio changes at $T < 50$ K because the magnetoelastic contributions to the *a* and *c* parameters change sign.

The special features of the thermal expansion of $RLiF_4$ are similar to those observed for rare-earth metal phosphates [4] and vanadates [3]. They are caused by magnetoelastic interaction between the aspherical 4*f* shell cloud of the rare-earth metal ion and the lattice. At high temperatures, at which all multiplet levels are equally populated, the electron shell is spherically symmetric. Electron cloud asphericity, which is characterized by quadrupole and multipole moments, changes as temperature lowers because excited level populations then gradually decrease (become frozen). As a result, magnetoelastic coupling causes the appearance of additional magnetoelastic anomalies of lattice parameters.

Systematic experimental data on the magnetoelastic contribution to the thermal expansion of rare-earth metal scheelites or an analysis of the phenomenon are lacking in the literature. We can only cite [5], where the temperature dependences of the *a* and *c* parameters were obtained for $HoLiF_4$ and $TmLiF_4$; these dependences were analyzed in the multipole approximation using the magnetoelastic coefficients calculated by the model of exchange charges. This approach cannot be used to correctly compare magnetoelastic effects in various structures, and, as mentioned, the calculated magnetoelastic coefficients can only be considered estimates. For this reason, we will analyze totally symmetric magnetoelastic modes using the approach applied in [3] to rare-earth metal zircons.

3.2. Thermal Expansion of the Scheelite Lattice ($LuLiF_4$)

To separate the purely magnetoelastic contribution to thermal expansion, we must correctly estimate the phonon contribution. X-ray measurements of $LuLiF_4$ allow the phonon contributions $\Delta a_{ph}/a$ and $\Delta c_{ph}/c$ to be determined for the other $RLiF_4$. It is only necessary to introduce corrections for their variation along the series of rare-earth metal ions. The correction for phonon contribution changes, which amounts to 10–15% of the contribution value, can, in our view, be introduced using fairly rough models, for instance, the Debye model of solids.

According to the Debye model, thermal expansion, for instance, along axis *a*, is described by the universal function of heat capacity at a constant volume $C_v(t)$,

$$\Delta a_{ph}/a = \alpha_0 T_D \int_0^t [C_v(t)/3R]dt, \qquad (1)$$

where $t = T/T_D$ is the reduced temperature and T_D and α_0 are two independent coefficients.

The experimental data obtained for $LuLiF_4$ and the $\Delta a_{ph}/a$ and $\Delta c_{ph}/c$ curves calculated by (1) with various T_D and α_0 parameter values are shown in Fig. 5. The thermal expansion of $LuLiF_4$ along axis *a* is seen to be fairly well described by the Debye formula with the coefficients $T_D = 790$ K and $\alpha_0 = 18.8 \times 10^{-6}$ (curve *1*). A comparison of curves *1* and *2* ($T_D = 710$ K and $\alpha_0 = 16.8 \times 10^{-6}$) shows that, generally, a coupled pair of the coefficients can only be determined within the accuracy of our measurements; that is, we can simultaneously somewhat decrease or increase both T_D and α_0, the agreement remaining satisfactory to within experimental errors. To determine the Debye temperature more accurately, we must perform precision measurements at temperatures below 150 K. Calculations show that changes in the Debye temperature within ±50 K cause very substantial changes in the photon contribution compared with the magnetoelastic contribution (compare curve *1* with *3* and *4*; $\alpha_0 = 18.8 \times 10^{-6}$ and $T_D = 740$ and 840 K). The thermal expansion along tetragonal axis *c* is approximated by the Debye dependence with $T_D = 600$ K and $\alpha_0 = 13.0 \times 10^{-6}$ (curve *5*). Note that we

further use the Debye model only as a simple and convenient form for describing the thermal expansion of the $RLiF_4$ lattice and introducing corrections for changes in the phonon contribution; the calculated T_D and α_0 values are neither used nor discussed in what follows.

4. THEORETICAL CONSIDERATION

4.1. Magnetoelastic Contribution to Thermal Expansion of Tetragonal Compounds

The rare-earth ion contribution to thermal expansion in the absence or presence of a magnetic field is calculated using a Hamiltonian including the crystal field Hamiltonian H_{CF}, the Zeeman term H_Z, and the one-particle magnetoelastic Hamiltonian H_{ME}. It is sufficient to calculate thermodynamic properties for the ground multiplet, and the crystal field Hamiltonian will therefore be written in terms of equivalent operators O_n^m in the form

$$\begin{aligned} H_{CF} &= \alpha_J B_2^0 O_2^0 + \beta_J (B_4^0 O_4^0 + B_4^4 O_4^4 + B_4^{-4}\Omega_4^4) \\ &\quad + \gamma_J (B_6^0 O_6^0 + B_6^4 O_6^4 + B_6^{-4} O_6^4), \end{aligned} \tag{2}$$

where α_J, β_J, and γ_J are the Stevens parameters and the number of crystal field parameters B_n^m for the scheelite tetragonal structure (local symmetry S_4) is seven. As distinguished from the structure of zircon (local symmetry D_{2d}), the H_{CF} Hamiltonian contains two additional terms with off-diagonal imaginary operators Ω_4^4 and Ω_6^4. Accurate to coefficient i, the matrix elements of these operators coincide with the elements of the corresponding Stevens operators O_4^4 and O_6^4 [6].

The Zeeman interaction between rare-earth metal ions and magnetic field **H** is described by the Hamiltonian

$$H_Z = -g_J \mu_B \mathbf{JH}, \tag{3}$$

where g_J is the Lande factor, **J** is the rare-earth ion angular momentum operator, and μ_B is the Bohr magneton.

Magnetoelastic effects are most often calculated using the quadrupole approximation. In the magnetoelastic Hamiltonian, we will only explicitly write the invariants that transform under the totally symmetric representation, which are essential to the totally symmetric modes that we wish to calculate,

$$H_{ME} = -\alpha_J O_2^0 (B_{20}^{\alpha 1}\varepsilon^{\alpha 1} + B_{20}^{\alpha 2}\varepsilon^{\alpha 2}), \tag{4}$$

where $B_{20}^{\alpha 1}$ and $B_{20}^{\alpha 2}$ are the quadrupole magnetoelastic coefficients. The symmetrized deformations, isotropic $\varepsilon^{\alpha 1}$ and tetragonal $\varepsilon^{\alpha 2}$, are expressed via the Cartesian components of the deformation tensor as follows:

$$\begin{aligned} \varepsilon^{\alpha 1} &= \sqrt{1/3}(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}), \\ \varepsilon^{\alpha 2} &= \sqrt{2/3}[\varepsilon_{zz} - (\varepsilon_{xx} + \varepsilon_{yy})/2]. \end{aligned} \tag{5}$$

The totally symmetric magnetoelastic modes are found as usual, from the condition of free energy $F = E_E - k_B T \ln Z$ minimum. In the partition function

$$Z = \sum_i \exp(-E_i/k_B T),$$

it is sufficient to calculate the E_i energy levels for the rare-earth metal ion based on the crystal field Hamiltonian H_{CF} and the Zeeman term H_Z in the presence of a magnetic field. The elastic energy of a tetragonal crystal is calculated only taking into account totally symmetric deformations determined via the symmetrized components of the deformation tensor [7],

$$E_E = \frac{1}{2} C_0^{\alpha 1} (\varepsilon^{\alpha 1})^2 + C_0^{\alpha 12} \varepsilon^{\alpha 1} \varepsilon^{\alpha 2} + \frac{1}{2} C_0^{\alpha 2} (\varepsilon^{\alpha 2})^2, \tag{6}$$

where C_0^{μ} are the symmetrized elastic constants of the lattice that do not take into account magnetic interactions. Their relation to the C_{ij} Cartesian components is given, e.g., in [8].

The totally symmetric magnetoelastic anomalies of a tetragonal lattice are described in the quadrupole approximation as [8, 9]

$$\left(\frac{\Delta\varepsilon^{\alpha 1}}{\varepsilon^{\alpha 1}}\right)_{me} = \frac{B_{20}^{\alpha 1} C_0^{\alpha 2} - B_{20}^{\alpha 2} C_0^{\alpha 12}}{C_0^{\alpha 1} C_0^{\alpha 2} - (C_0^{\alpha 12})^2} Q_{20}(T, H), \tag{7}$$

$$\left(\frac{\Delta\varepsilon^{\alpha 2}}{\varepsilon^{\alpha 2}}\right)_{me} = \frac{B_{20}^{\alpha 2} C_0^{\alpha 1} - B_{20}^{\alpha 1} C_0^{\alpha 12}}{C_0^{\alpha 1} C_0^{\alpha 2} - (C_0^{\alpha 12})^2} Q_{20}(T, H), \tag{8}$$

where

$$Q_{20}(T, H) = \alpha_J \langle O_2^0 \rangle = \alpha_J \frac{1}{Z} \langle i | O_2^0 | i \rangle \exp\left(-\frac{E_i}{k_B T}\right).$$

The quadrupole moment $\langle O_2^0 \rangle = \langle 3J_z^2 - J(J+1) \rangle$ (the observed O_2^0 operator value) is zero for cubic crystals, whose x, y, and z axes are equivalent ($\langle J_x^2 \rangle = \langle J_y^2 \rangle = \langle J_z^2 \rangle = \langle J(J+1)/3 \rangle$), and becomes nonzero when the z axis is a distinguished direction, for instance, in tetragonal crystals. It follows that the quadrupole moment

Table 1. Crystal field parameters of rare-earth metal scheelites determined in the formalism of tensor (C_k^g) or equivalent (O_n^m) operators; the B_4^{-4} and B_6^{-4} parameters are nonzero in the crystallographic system of coordinates; the crystal field parameter sets used in the calculations are labeled by asterisks

$RLiF_4$	Method	B_2^0, cm^{-1}	B_4^0, cm^{-1}	B_6^0, cm^{-1}	B_4^4, cm^{-1}	B_6^6, cm^{-1}	B_4^{-4}, cm^{-1}	B_6^{-4}, cm^{-1}	Refs.
$PrLiF_4$	(C_k^g)*	489	−1043	−42	1242	1213	–	23	[10]
$NdLiF_4$	C_k^g	441	−906	−26	1114	1072	–	21	[10]
$NdLiF_4$	(C_k^g)*	532	−934	−76	−1161	−916	244	534	[11]
$TbLiF_4$	C_k^g	474	−433	64	1080	744	–	286	[12]
$TbLiF_4$	(O_n^m)*	237	−54	4	−854	−477	−739	−291	[13]
$DyLiF_4$	O_n^m	207	−69.5	−0.6	−776.5	−536	−581	−285	ext.
$DyLiF_4$	(O_n^m)*	202	−66	1.5	−756	−340	−654	−272	opt.
$HoLiF_4$	C_k^g	379	−626	−52	831	608	–	–	[14]
$HoLiF_4$	(O_n^m)*	189.3	−78.3	−3.3	−657	−322	−568	−253	[15]
$ErLiF_4$	C_k^g	377	−642	−71	861	625	–	–	[14]
$ErLiF_4$	(O_n^m)*	190	−80	−2.3	−771	−363	−667	−222	[16]
$TmLiF_4$	C_k^g	368	−717	−65	919	619	–	–	[17]
$TmLiF_4$	(O_n^m)*	184	−90	−4.06	−727	−328	−628	−284	[18]
$YbLiF_4$	(O_n^m)*	185	−76	0	−618	−288	−534	−177	[13]

describes the degree of electron cloud asphericity and its changes induced by temperature variations or by applying an external magnetic field determine magnetoelastic anomalies of thermal expansion.

Analysis and description of magnetoelastic effects in the multipole approximation are considerably more complex. For a rare-earth ion in a site with tetragonal environment symmetry, the one-particle magnetoelastic Hamiltonian linear in deformation tensor components ε^μ (harmonic approximation) contains 26 invariants in the multipole approximation [3]. Usually, magnetoelastic effects are described at the level of the quadrupole approximation, which reduces the number of terms in the Hamiltonian to five. However, it is not obvious that the simplest quadrupole approximation is sufficient for correctly describing the totality of magnetoelastic effects in systems with strong magnetoelastic interactions.

To summarize, totally symmetric magnetoelastic modes are determined in the multipole approximation by temperature-induced variations in all multipole moments of rare-earth ions Q_{nm}, which can be calculated numerically from known crystal field parameters. The contributions of multipole moments to totally symmetric magnetoelastic modes are described by equations similar to (7) and (8), which contain multipole moments Q_{nm} and multipole magnetoelastic coefficients $B_{nm}^{\alpha 1}$ and $B_{nm}^{\alpha 2}$.

4.2. The Multipole Moments of Rare-Earth Metal Ions in Scheelites and Criteria of the Applicability of the Quadrupole Approximation

For most of the $RLiF_4$ scheelites formed by rare-earth metals R, crystal field parameters can be considered as reliably established from detailed optical stud-

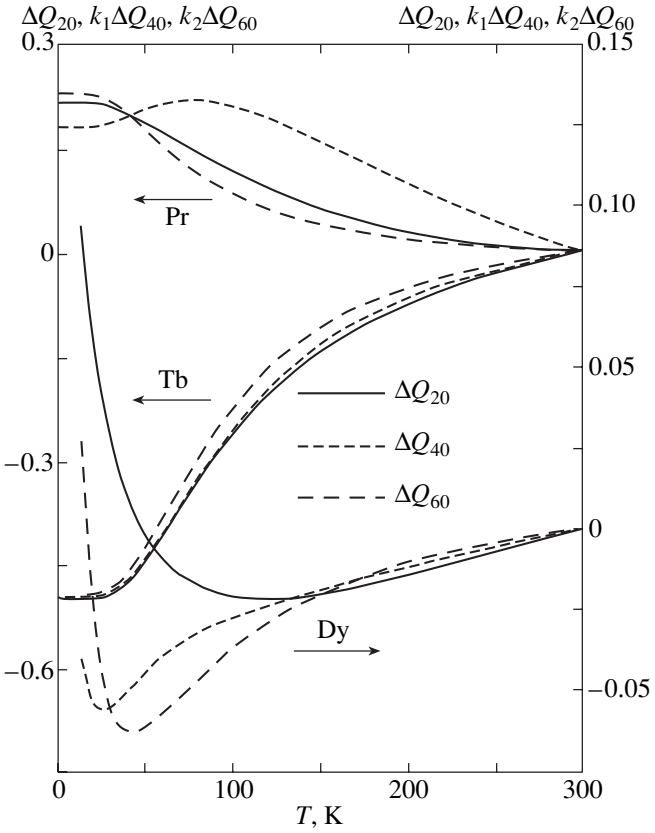

Fig. 6. Calculated temperature dependences of the ΔQ_{20}, $k_1\Delta Q_{40}$, and $k_2\Delta Q_{60}$ multipole moments of rate-earth metal scheelites $PrLiF_4$ ($k_1 = 1$ and $k_2 = 0.15$), $TbLiF_4$ ($k_1 = -0.82$ and $k_2 = 2.9$), and $DyLiF_4$ ($k_1 = -0.7$ and $k_2 = 0.4$).

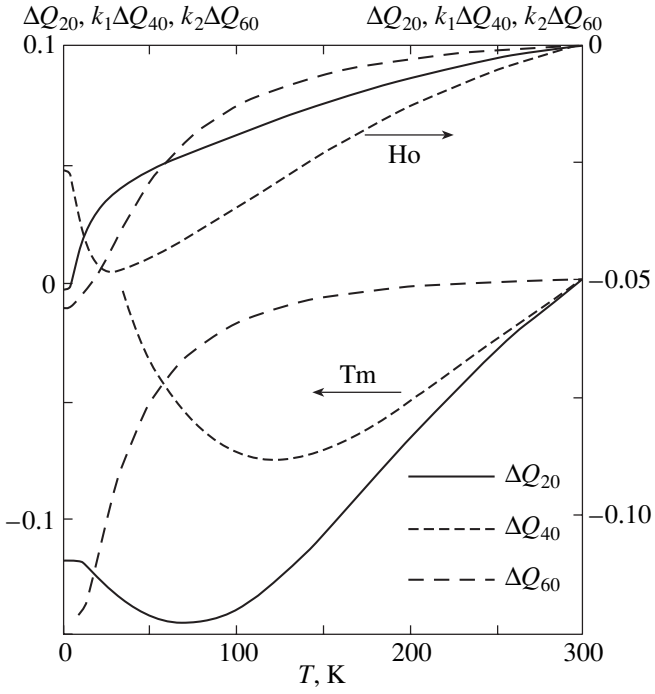

Fig. 7. Calculated temperature dependences of the ΔQ_{20}, $k_1\Delta Q_{40}$, and $k_2\Delta Q_{60}$ multipole moments of rate-earth metal scheelites $HoLiF_4$ ($k_1 = -0.45$ and $k_2 = -0.045$) and $TmLiF_4$ ($k_1 = -1$ and $k_2 = 0.15$).

ies [10–18]. The available data on the crystal fields of these compounds are summarized in Table 1, in which the method of measurements and the basis for which the corresponding parameters were determined are also indicated. The crystal field parameters used in our calculations are labeled by asterisks. For some compounds of the family ($DyLiF_4$ and $YbLiF_4$), only crystal field parameters calculated by the model of exchange charges are available; these parameters were not given reliable experimental substantiation. In certain instances, we were able to select between alternative sets of crystal field parameters for these compounds based on our experimental data. Note that optical experiments are often described in rotated coordinate systems for which exact rotation angles are not known. This impedes the use of the obtained crystal field parameters for describing magnetoelastic effects.

Numerical calculations show that fourth- or sixth-order multipole moments of $RLiF_4$ are comparable to ($TbLiF_4$ and $DyLiF_4$) or even substantially larger than ($HoLiF_4$ and $TmLiF_4$) the quadrupole moments. The temperature dependences of three diagonal multipole moments ΔQ_{20}, ΔQ_{40}, and ΔQ_{60} for several ions (Pr, Tb–Ho, and Tm) are shown in Figs. 6 and 7. The Q_{44} and Q_{64} off-diagonal moments (and the Ω_{44} and Ω_{64} moments proportional to them) and their temperature variations are, as a rule, small for rare-earth metal ions in scheelites; we will not discuss them in this work. For convenience of comparison, scaled dependences of multipole moments $k_1\Delta Q_{40}$ and $k_2\Delta Q_{60}$ are given. The magnitude and sign of the k_1 and k_2 scale factors were selected such that changes in Q_{20}, Q_{40}, and Q_{60} in the temperature range 10–300 K were comparable in magnitude. The smaller the k_1 and k_2 factors in magnitude, the more substantial the expected contribution of the fourth- and sixth-order multipole moments, respectively, to thermal expansion. Note that the Q_{20} and Q_{60} moments are, as a rule, negative, and Q_{40}, positive, because second- and sixth-order crystal field parameters are positive, and fourth-order parameters, negative. The criteria of quadrupole approximation applicability can be formulated differently for two different groups of ions.

If the temperature dependences of all moments are similar, as, for instance, for $TbLiF_4$, we can use the "effective" quadrupole approximation, in which the effective magnetoelastic coefficient is some combination of quadrupole and multipole magnetoelastic coefficients. It is then difficult to use experimental data for drawing conclusions about the contributions of multipole moments of different orders to totally symmetric magnetoelastic modes.

Table 2. Coefficients $A_1 = 3(\Delta c_{me}/c)/\Delta Q_{20}$ and $A_2 = 3(\Delta a_{me}/a)/\Delta Q_{20}$ and quadrupole magnetoelastic coefficients $B^{\alpha 1}$ and $B^{\alpha 2}$ for rare-earth metal scheelites; $(\Delta c_{me}/c)$, $(\Delta a_{me}/a)$, and ΔQ_{20} are the magnetoelastic contributions to thermal expansion along axes *c* and *a* and the quadrupole moment change in the temperature range 10–290 K, respectively

$RLiF_4$	A_1, 10^{-2}	A_2, 10^{-2}	$B^{\alpha 1}$, 10^3 K/(formula unit)	$B^{\alpha 2}$, 10^3 K/(formula unit)
$TbLiF_4$ [a)]	−0.645	0.366	0.03	−1.37
$TbLiF_4$ [b)]	−0.642	0.353	−0.03	−1.35
DyLiF [b)]	−0.75	0.375	−0.2	−1.54
$TbVO_4$ [a)]	0.75	−1.02	−5.40	5.13

[a)] For the interval 10–290 K;
[b)] for the interval 10–160 K.

However, if the temperature dependences of Q_{i0} multipole moments are essentially different, as for Tm and Ho, a comparison with experimental data allows us to identify the multipole moment that makes the major contribution to thermal expansion and to estimate the actual magnetoelastic coefficients. If the contribution of fourth- and sixth-order multipole moments to the magnetoelastic anomalies of thermal expansion is significant, the temperature dependences of $\Delta a_{me}/a$ and $\Delta c_{me}/c$ should contain singularities corresponding to these moments. Calculations of the temperature dependences of the multipole moments of rare-earth metal ions and a comparison of these dependences with those determined experimentally make it possible to estimate the contributions to thermal expansion of different multipole moments regardless of the deficiency of information about magnetoelastic coefficients of various orders.

An analysis of the experimental data on $TmLiF_4$ and $HoLiF_4$ shows that a noticeable contribution is made by the sixth-order multipole moments of the corresponding ions. Indeed, for $TmLiF_4$, changes in the Q_{60} sixth-order multipole moments are almost an order of magnitude ($k_2 = 0.15$) larger than changes in the quadrupole moment, and the temperature dependences of these two moments are different. The temperature dependence of the $|\Delta Q_{20}|$ quadrupole moment has a maximum at 80 K, and the moment begins to decrease at lower temperatures, which is at variance with the thermal expansion anomalies of $TmLiF_4$. The ΔQ_{60} moment changes monotonically, and the rate of its growth is maximum at about 50 K, where a thermal expansion anomaly is observed. The $\Delta c_{me}/c$ and $\Delta a_{me}/a$ magnetoelastic contributions are proportional to the ΔQ_{60} multipole moment in $TmLiF_4$, which means that the $B_{60}^{\alpha 1}$ and $B_{60}^{\alpha 2}$ sixth-order magnetoelastic coefficients are not too small compared with the corresponding quadrupole magnetoelastic coefficients. Calculations according to the model of exchange charges give similar ratios between the magnetoelastic coefficients of various orders [19]. A similar behavior of multipole moments is observed for the Ho ion in the scheelite structure ($k_1 = -0.45$ and $k_2 = -0.045$, Fig. 7).

As the crystal field was reliably determined for $TmLiF_4$, and corrections to the phonon contribution compared with $LuLiF_4$ have no noticeable influence on the character of the temperature dependences of $\Delta c_{me}/c$ and $\Delta a_{me}/a$, the reasoning given above and the conclusions drawn from it should be considered valid. Magnetoelastic coefficients change not very substantially along the series of rare-earth metal ions within the structure type under consideration. We can therefore expect that other rare-earth metal ions should be characterized by similar ratios between the magnetoelastic coefficients of various orders. To summarize, our analysis of the experimental data on the thermal expansion of $RLiF_4$ leads us to conclude that the contributions of the fourth- and sixth-order multipole moments to totally symmetric magnetoelastic modes in the series of scheelites formed by rare-earth metals can exceed the contributions of their quadrupole moments; that is, the quadrupole approximation is not quite valid for these compounds.

5. THE MAGNETOELASTIC CONTRIBUTION TO THE THERMAL EXPANSION OF $RLiF_4$ (R = Tb–Ho, Tm)

We used the experimental data on $TbLiF_4$ compound, for which the effective quadrupole approximation is valid, to determine its quadrupole magnetoelastic coefficients. The magnetoelastic contribution in $TbLiF_4$ was found taking into account corrections for phonon contribution variations. For this purpose, we used an optimization procedure in which the T_D and α_0 Debye model parameters were varied within ±20% from their values for $LuLiF_4$. The optimization was performed to make the resulting magnetoelastic contribution proportional to the quadrupole moment and reduce the largest difference (observed at low temperatures) between the phonon thermal expansion values along axes *c* and *a* in $TbLiF_4$ and $LuLiF_4$ below some set value, namely, $\delta(\Delta c/c)_0 \sim \delta(\Delta a/a)_0 \sim 1 \times 10^{-4}$.

The optimization gave the proportionality factors between the magnetoelastic contributions and quadrupole moments for the Tb and Dy scheelites, $\Delta c_{me}/c = A_1\Delta Q_{20}/3$ and $\Delta a_{me}/a = A_2\Delta Q_{20}/3$ [see (7), (8)]. The A_1 and A_2 values are listed in Table 2. The magnetoelastic contributions along axes *c* and *a* and the calculated quadrupole moments for these scheelites are shown in Fig. 8. To make comparison convenient, the magnetoelastic contribution along *a* and the quadrupole moment are given with the corresponding factors,

$A\Delta a_{me}/a$ and $A_1\Delta Q_{20}$ ($A = A_1/A_2 \sim -1$ and $A_1 \sim 10^{-2}$, see Table 2). As has been mentioned, there is no crystal field data on $DyLiF_4$. The extrapolated (intermediate between the crystal fields of the neighboring Tb and Ho ions) crystal field did not give an acceptable description of the magnetoelastic contribution. We therefore performed additional optimization of the extrapolated crystal field of $DyLiF_4$ using data on the *g*-factors and splittings of the lower Dy^{3+} levels. This improved agreement with experiment. Of course, in the absence of spectroscopic information about $DyLiF_4$, complete crystal field optimization cannot be performed and a unique reliable set of parameters cannot be obtained.

6. DISCUSSION

When the effective quadrupole approximation is valid, the contribution of rare-earth metal ions to thermal expansion is determined by the $B_{20}^{\alpha 1} = B^{\alpha 1}$ and $B_{20}^{\alpha 2} = B^{\alpha 2}$ quadrupole magnetoelastic coefficients and temperature-induced quadrupole moment changes $\Delta Q_{20} = Q_{20}(T) - Q_{20}(290\ \text{K})$ [see (7), (8)]. Consider some special features of the temperature dependences of ΔQ_{20} quadrupole moments in the series of rare-earth metal scheelites (Fig. 9). The largest ΔQ_{20} change in the temperature range 10–290 K was observed for $TbLiF_4$. Calculations show (see Fig. 9) that, below 50 K, the quadrupole moment of the Dy scheelite and the derivative of the quadrupole moment of the Tm scheelite change their signs.

For the other scheelites, quadrupole moment changes are insignificant, as is shown in the inset in Fig. 9. Although $\Delta Q_{20}(T)$ is small for the Ho and Er scheelites, it has sharp anomalies at about 30–50 K, where the phonon contribution is almost constant. These low-temperature anomalies can be studied by more sensitive methods, for instance, with the use of strain gauges. In $HoLiF_4$ and $TmLiF_4$ (curves *3*, *5*), quadrupole moment changes in the temperature range 10–290 K are small and should not cause observable magnetoelastic effects. Other multipole moments of these ions, however, experience substantial changes as temperature varies. It is likely that the behavior of these moments determines the magnetoelastic contributions to thermal expansion.

The anomalies of the $\Delta Q_{20}(T)$ dependences are interesting to relate to the special features of the spectra and wave functions of rare-earth metal ions in the scheelite structure. An analysis shows that the sharp change in $\Delta Q_{20}(T)$ at $T \sim 90$ K observed for the Tb scheelite is related to a decrease in the population of two lower singlets, which form a quasi-doublet and make the major contribution. This contribution decreases as the population of the next excited level situated at a distance of $E_3 \sim 180$ K grows. A different situation is observed for the Dy scheelite. The contribution of the lower doublet is very small, whereas the contributions of the three higher doublets situated at $T = 5$, 62, and 88 K are comparable in magnitude and gradually decrease as the populations of these levels decrease

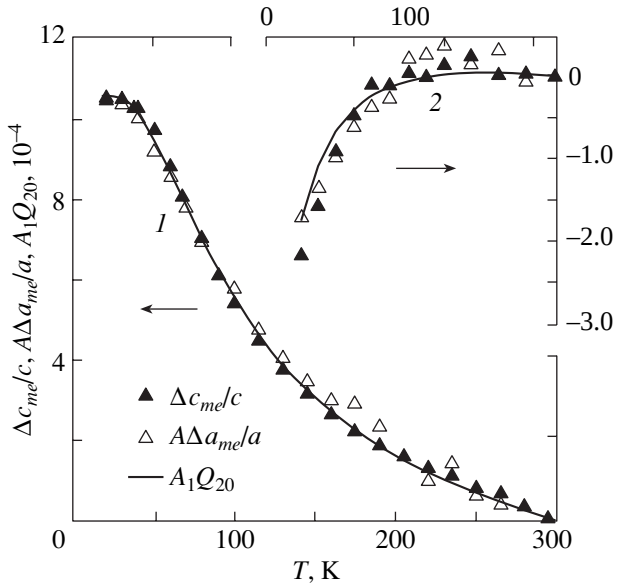

Fig. 8. Magnetoelastic contributions $\Delta c_{me}/c$ and $A\Delta a_{me}/a$ and quadrupole moments $A_1\Delta Q_{20}$ for (*1*) $TbLiF_4$ and (*2*) $DyLiF_4$, experimental and calculated in the quadrupole approximation; the $A = A_1/A_2$ and A_1 parameter values for the specified rare-earth metal ions are listed in Table 2 ($A \sim -1$ and $A_1 \sim 10^{-2}$).

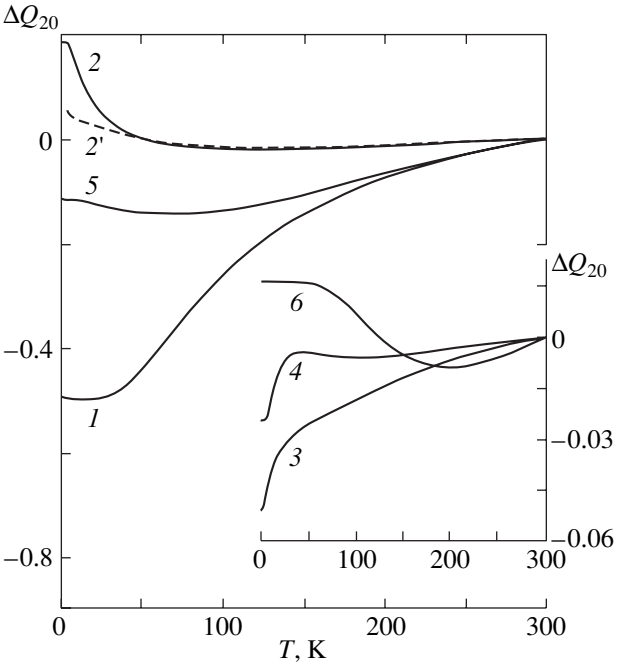

Fig. 9. Calculated quadrupole moments Q_{20} for $RLiF_4$: (*1*) $TbLiF_4$, (*2*, *2'*) $DyLiF_4$, and (*5*) $TmLiF_4$; given in the inset are calculated quadrupole moments Q_{20} drawn on an enlarged scale for (*3*) $HoLiF_4$, (*4*) $ErLiF_4$, and (*6*) $YbLiF_4$.

with a decrease in temperature. This explains the change in sign of $\Delta Q_{20}(T)$ at low temperatures. Also note that the quadrupole and multipole moments of $DyLiF_4$ are very sensitive to crystal field parameters, as can be seen from a comparison of curves *2* and *2'* for the extrapolated and optimized crystal fields (Fig. 9).

The experimental magnetoelastic contribution to the thermal expansion of the Tb and Dy scheelites correlates with changes in the calculated quadrupole moments ΔQ_{20}. This leads us to conclude that, in a first approximation, their quadrupole magnetoelastic coefficients $B^{\alpha 1}$ and $B^{\alpha 2}$ are not too different. We used the C_{11}, C_{12}, C_{13}, and C_{33} elasticity constants of $YLiF_4$ [20] to calculate the symmetrized elasticity constants in (7) and (8). This gave [in 10^5 K/(formula unit) units] $C_0^{\alpha 1} = 13.0$, $C_0^{\alpha 2} = 5.03$, and $C_0^{\alpha 12} = 0.7$. The hierarchy of the elasticity constants for the scheelite structure is the same as for zircons, but their values are on average two times smaller. Next, we determined the $B^{\alpha 1}$ and $B^{\alpha 2}$ magnetoelastic quadrupole coefficients (Table 2) from the A_1 and A_2 proportionality coefficients and the symmetrized elasticity constants. For comparison, Table 2 contains similar data on $TbVO_4$. At the specified relative elasticity constants $C_0^{\alpha 1} \sim C_0^{\alpha 2} \ll C_0^{\alpha 12}$, the sign and magnitude of the $B^{\alpha 2}$ coefficient are determined by the magnetoelastic mode $(\Delta\varepsilon^{\alpha 2}/\varepsilon^{\alpha 2})_{me}/\Delta Q_{20} \approx B^{\alpha 2}/C_0^{\alpha 2}$ [the change in the degree of tetragonality, Eq. (8)], and the $B^{\alpha 1}$ coefficient, which is comparable in magnitude, by the magnetoelastic mode $(\Delta\varepsilon^{\alpha 1}/\varepsilon^{\alpha 1})_{me}/\Delta Q_{20} \approx B^{\alpha 1}/C_0^{\alpha 1}$ [the change in volume, Eq. (7)].

Note that the signs of the magnetoelastic coefficients for the scheelites are opposite to those obtained for the vanadates and phosphates with zircon structures, and their magnitudes are different [3, 4]. For the scheelites, the $B^{\alpha 1}$ coefficient is substantially smaller than $B^{\alpha 2}$; that is, the magnetoelastic contribution to volume deformation is insignificant compared with the magnetoelastic contribution to changes in the degree of tetragonality. The $B^{\alpha 1}$ magnetoelastic coefficient is determined as the difference of two large values, and small changes in A_1 and A_2 or the $C_0^{\alpha 1}$, $C_0^{\alpha 2}$, and $C_0^{\alpha 12}$ elasticity constants can therefore change not only the magnitude but also the sign of this coefficient (see Table 2). This causes large errors in its determination, and we only obtain estimates of $B^{\alpha 1}$ compared with $B^{\alpha 2}$. The largest magnetoelastic coefficient $B^{\alpha 2}$ for the scheelite structure is approximately four times smaller than that for zircons. It follows that the magnetoelastic anomalies of thermal expansion of rare-earth metal scheelites should be two times smaller than those of rare-earth metal zircons, changes in their quadrupole moments being comparable.

The experimental data on the scheelite structure are evidence of the existence of substantial multipole contributions to spontaneous magnetoelastic effects such as magnetoelastic thermal expansion anomalies. Systematic studies of rare-earth metal ion compounds with other structures show that the quadrupole approximation is quite acceptable and sufficient for describing various magnetoelastic effects in them. A wide range of magnetoelastic effects (thermal expansion anomalies, Young's modulus, the ΔE effect, and magnetostriction) in tetragonal phosphates RPO_4 [9] and vanadates RVO_4 [3] with zircon structures and paramagnetic aluminate–garnets and gallate–garnets containing rare-earth metal ions from Tb to Yb with orthorhombic local symmetry [21] can be described by the quadrupole approximation. The necessity of using the multipole approximation in similar situations for the scheelites formed by rare-earth metal ions is a new and interesting result for the theory of magnetoelasticity.

7. CONCLUSIONS

In this work, we performed theoretical and experimental studies of the thermal expansion of $RLiF_4$ scheelites formed by rare-earth metal ions. We observed substantial thermal expansion anomalies for scheelites with Tb–Ho and Tm ions and determined the $\Delta c_{me}/c$ and $\Delta a_{me}/a$ magnetoelastic contributions to thermal expansion taking into account corrections for changes in the phonon contribution along the series of rare-earth metal ions according the Debye model of thermal expansion. The experimental $\Delta c_{me}/c$ and $\Delta a_{me}/a$ values were used to calculate the $B^{\alpha 1}$ and $B^{\alpha 2}$ totally symmetric quadrupole coefficients, which were substantially different in magnitude. A comparison with rare-earth metal vanadates RVO_4 showed that the $B^{\alpha 1}$ and $B^{\alpha 2}$ magnetoelastic coefficients and ΔQ_{20} quadrupole moment changes had opposite signs for the two series of tetragonal structures, whereas the $\Delta c_{me}/c$ and $\Delta a_{me}/a$ magnetoelastic contributions to thermal expansion determined by the products of these factors were of the same sign. Different ΔQ_{20} quadrupole moment signs are caused by different signs of the B_2^0 second-order parameter for the scheelites and vanadates formed by rare-earth metal ions.

Independent experimental data on magnetoelastic coefficients and their hierarchy in different structures are necessary for testing various models in the theory of magnetoelasticity. In particular, calculations of the magnetoelastic coefficients for the scheelite structure by the model of exchange charges [2] can be compared with experiment to reveal their interrelation with the structural parameters of the crystals, ion charges, dipole moments, etc.

The conditions of the applicability of the quadrupole approximation to describe the totally symmetric magnetoelastic modes were considered based on a

comparison of the calculated quadrupole and dipole moments of various rare-earth metal ions in the scheelite structure and the experimental magnetoelastic contributions to thermal expansion. It was shown that multipole moments could make a considerable contribution for some ions and the magnetoelastic thermal expansion anomalies of scheelites with rare-earth metal ions could not be described solely by the temperature dependences of their quadrupole moments. The conclusion was drawn that the fourth- and sixth-order multipole magnetoelastic coefficients were comparable with the quadrupole magnetoelastic coefficients. Substantial multipole contributions and the necessity of going beyond the quadrupole approximation with the scheelites containing rare-earth metal ions are of certain interest for the theory of magnetoelasticity. For this reason, other spontaneous and induced magnetoelastic effects in these scheelites for which multipole contributions may be of significance appear worthy of theoretical and experimental investigation.

ACKNOWLEDGMENTS

This study was partly supported by the Russian Foundation for Basic Research (project no. 03-02-16809) and by the International Scientific-Technological Center (grant no. 2029).

REFERENCES

1. G. A. Gehring and K. A. Gehring, Rep. Prog. Phys. **38**, 1 (1975).
2. B. Z. Malkin, in *Spectroscopy of Solids Containing Rare-Earth Ions*, Ed. by A. A. Kaplyanskii and R. M. Macfarlane (North-Holland, Amsterdam, 1987), p. 13.
3. Z. A. Kazeĭ and N. P. Kolmakova, Zh. Éksp. Teor. Fiz. **109**, 1687 (1996) [JETP **82**, 909 (1996)].
4. Z. A. Kazeĭ, N. P. Kolmakova, and O. A. Shishkina, Fiz. Tverd. Tela (St. Petersburg) **39**, 106 (1997) [Phys. Solid State **39**, 91 (1997)].
5. R. Yu. Abdulsabirov, Z. A. Kazeĭ, S. L. Korableva, and D. N. Terpilovskiĭ, Izv. Akad. Nauk, Ser. Fiz. **57**, 138 (1993).
6. S. A. Al'tshuler and B. M. Kozyrev, *Electron Paramagnetic Resonance in Compounds of Transition Elements* (Nauka, Moscow, 1972; Halsted, New York, 1975).
7. E. de Lacheisserie, Ann. Phys. (Paris) **5**, 267 (1970).
8. P. Morin, J. Rouchy, and D. Schmitt, Phys. Rev. B **37**, 5401 (1988).
9. V. I. Sokolov, Z. A. Kazeĭ, N. P. Kolmakova, and T. V. Solov'yanova, Zh. Éksp. Teor. Fiz. **99**, 945 (1991) [Sov. Phys. JETP **72**, 524 (1991)].
10. L. Esteromitz, F. J. Bartoli, R. E. Allen, *et al.*, Phys. Rev. B **19**, 6442 (1979).
11. H. De Leebeeck and C. Gorller-Walrand, J. Alloys Comp. **225**, 75 (1995).
12. H. P. Christensen, Phys. Rev. B **17**, 4060 (1978).
13. A. K. Kupchikov, B. Z. Malkin, A. L. Natadze, and A. I. Ryskin, Fiz. Tverd. Tela (Leningrad) **29**, 3335 (1987) [Sov. Phys. Solid State **29**, 1913 (1987)].
14. H. P. Christensen, Phys. Rev. B **19**, 6564 (1979).
15. N. I. Apladze, M. A. Koreĭba, and B. Z. Malkin, Zh. Éksp. Teor. Fiz. **104**, 4171 (1993) [JETP **77**, 1021 (1993)].
16. L. A. Bumagina, V. I. Krotov, B. Z. Malkin, and A. Kh. Khasanov, Zh. Éksp. Teor. Fiz. **80**, 1543 (1981) [Sov. Phys. JETP **53**, 792 (1981)].
17. H. P. Christensen, Phys. Rev. B **19**, 6573 (1979).
18. A. V. Vinokurov, B. Z. Malkin, A. I. Pominov, and A. L. Stolov, Fiz. Tverd. Tela (Leningrad) **28**, 381 (1986) [Sov. Phys. Solid State **28**, 211 (1986)].
19. A. V. Vinokurov, S. L. Korableva, B. Z. Malkin, *et al.*, Fiz. Tverd. Tela (Leningrad) **30**, 801 (1988) [Sov. Phys. Solid State **30**, 459 (1988)].
20. P. Blanchfield and G. A. Saunders, J. Phys. C: Solid State Phys. **12**, 4673 (1979).
21. N. P. Kolmakova, R. Z. Levitin, V. N. Orlov, and N. F. Vedernikov, J. Magn. Magn. Mater. **87**, 218 (1990).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 97, No. 2, 2003, pp. 290–304.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 2, 2003, pp. 322–337.

SOLIDS
Electronic Properties

Energy Spectrum of *p* Electrons in a Plane Crystal Lattice and in a Strong Magnetic Field

R. O. Zaitsev[a, *] and Yu. V. Mikhailova[b]

[a]*Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia*
e-mail: agydel@veernet.iol.ru
[b]*State Research Center NII Teplopribor, Moscow, 129085 Russia*

Received December 4, 2002

Abstract—Harper equations are derived for a p_x, p_y electronic system. Analysis is carried out for extreme points of the quasi-continuous spectrum in the cases when the number of magnetic flux quanta through a unit cell is a rational number and calculations are made for square and triangular lattices as well as for a honeycomb lattice with two nonequivalent atoms. The possibility of application of the results for explaining the fractional Hall effect is considered. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Equations determining the energy spectrum of the system under the action of a magnetic field form a system of difference equations and are known as Harper equations [1]. In the problem considered here, there are two limits, viz., the energy spectrum consisting of discrete strongly degenerate Landau levels in the absence of a periodic potential and the energy bands associated with Bloch waves emerging due to translational lattice symmetry in zero magnetic field.

The problem of electrons in a lattice with periods $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$, which move in a magnetic field $\mathbf{H} \parallel \mathbf{a}_3$, is characterized by the natural parameter $\zeta = (HS_{\mathbf{a}_1\mathbf{a}_2})/\Phi_0$ equal to the ratio of the magnetic flux through an area element constructed on vectors $\mathbf{a}_1$, $\mathbf{a}_2$ to the flux quantum $\Phi_0 = \hbar c/|e|$. If this parameter is rational, i.e., $\zeta = p/q$, we have a magnetic superlattice constructed on vectors $q\mathbf{a}_1$, $q\mathbf{a}_2$. This simplifies calculations significantly. In the simplest case of a square lattice with a single atom in the unit cell with rational number $\zeta = eHa^2/\hbar c$, the energy spectrum was calculated by Hofstadter [2] and is known as the Hofstadter butterfly. In the case of a weak magnetic potential, the Landau levels split into subbands [3], while degeneracy is partially removed in a strong field [4].

In the Hofstadter case mentioned above, the Schrödinger equation in the strong coupling approximation leads to Harper equations

$$\Psi_{n+1} + 2\Psi_n \cos(2\pi n\zeta - \varphi) + \Psi_{n-1} = E\Psi_n, \qquad (1)$$

where Ψ_n is the wave function at the nth site, $\varphi = ap_y/\hbar$ is the phase associated with the electron quasimomentum, a is the lattice constant, and E is the energy in units of the jump integral to the nearest neighbors.

All possible self-energies of the Harper equations lie in the range $|E| \le 4$ and the parameter can be confined to the region $0 < |\zeta| < 1$. In fact, it is sufficient to consider the range $0 < |\zeta| < 1/2$ since the Harper equation is invariant to the substitution $\zeta \longrightarrow 1 - \zeta$. It should also be noted that Eqs. (1) permit the substitution $E \longrightarrow -E$.

The numerical method proposed by Hofstadter was based on the fact that Harper equation (1) can be written in the form

$$\begin{pmatrix} \Psi_{n+1} \\ \Psi_n \end{pmatrix} = \hat{T}_n \begin{pmatrix} \Psi_n \\ \Psi_{n-1} \end{pmatrix} \qquad (2)$$

with the transition matrix

$$\hat{T}_n = \begin{pmatrix} E - 2\cos(2\pi n\zeta - \varphi) & -1 \\ 1 & 0 \end{pmatrix}. \qquad (3)$$

Transition matrices T_n have determinants equal to unity. In the case when $\zeta = p/q$, Eqs. (2) are periodic in n with period q. Consequently, matrix $\hat{A} = \prod_{n=1}^{q} T_n$ has eigenvalues whose absolute value is smaller than unity. Since $\det\hat{A} = 1$, this requirement is equivalent to the condition $\mathrm{Tr}(\hat{A}) = 2\cos\theta$, which leads to the entire excitation spectrum.

It is interesting to note that the Hofstadter solution could be modeled experimentally [5] using the analogy between the equations describing the propagation of microwaves through a scattering matrix and the Harper equations for an electron. The passage of a microwave through a (100) matrix of scattering elements inserted

in a waveguide was studied in [5], where the bands reproducing the Hofstadter butterfly were determined.

The experimental realization of a butterfly with typical lattice parameters of 0.1 nm requires magnetic fields on the order of 10^5 T, which is beyond the technically accessible limit. The only way to solve this problem is to use artificial superstructures. The first indications of magnetically split subbands are in fact contained in [6–9].

The quantum Hall effect is studied on structures consisting of atoms with an unfilled p shell. For this reason, we will study a two-dimensional system of p_x and p_y electrons in a square lattice or in a honeycomb lattice. The s electrons in a triangular lattice or in a honeycomb lattice are equally interesting since the properties of such electrons in the 2D case are equivalent to the properties of p_z electrons placed in the same lattices.

Theoretical investigations of the quantum Hall effect have been confined so far to analysis of a 2D system of Landau levels taking into account a weak Coulomb interaction in perturbation theory. The only exceptions were paper [9] and a recent publication [10]. In [9] a group-theoretical analysis is carried out for the electronic structure of a 2D electron gas in an external magnetic field, while the conditions for the emergence of the fractional Hall effect are studied in [10] in the framework of self-consistent perturbation theory. Several series of fractional relations are derived in the concluding part of this paper.

2. HARPER EQUATIONS FOR THE p-ELECTRON SYSTEM

Let us consider equations for the creation operators in the Heisenberg representation:

$$i\hbar\frac{\partial\hat{a}_{k,\mathbf{r}}(t)}{\partial t} = -[\hat{H},\hat{a}_{k,\mathbf{r}}(t)] = \sum_{n,\mathbf{s}}\hat{H}^{k,\mathbf{r}}_{n,\mathbf{s}}\hat{a}_{n,\mathbf{s}}(t). \quad (4)$$

Here, $\hat{a}_{k,\mathbf{r}}(t)$ is the annihilation operator for an electron belonging to a cell with coordinate $\mathbf{r}$ and located in a degenerate state with number k.

In the semiclassical approximation, we have

$$\hat{H}^{k,\mathbf{r}}_{n,\mathbf{s}} = \int\Psi^*_k(\mathbf{R}-\mathbf{r})\hat{h}(\mathbf{R})\Psi_n(\mathbf{R}-\mathbf{s})d\mathbf{R},$$
$$\Psi_\pm(\mathbf{R}) = \exp\left(-\frac{ie}{c\hbar}\int^{\mathbf{R}}\mathbf{A}(\mathbf{r})d\mathbf{r}\right)(x\pm iy)\rho\left(\frac{R}{r_B}\right). \quad (5)$$

The atomic wave function written in this way diagonalizes the atomic Hamiltonian for $\mathbf{r} = \mathbf{s}$:

$$H^{k,\mathbf{r}}_{k,\mathbf{r}} = \mathrm{diag}(\mu_B H, -\mu_B H). \quad (6)$$

If $\mathbf{r} \neq \mathbf{s}$, it is convenient to replace $\mathbf{R} \longrightarrow \mathbf{R} + (\mathbf{r} + \mathbf{s})/2$ and then integrate over the product $\rho(\mathbf{R} + (\mathbf{r} - \mathbf{s})/2)\rho(\mathbf{R} - (\mathbf{r} + \mathbf{s})/2)$ of radial functions.

In the remaining factors, we can disregard $\mathbf{R}$ as compared to $(\mathbf{r} - \mathbf{s})$. This gives

$$H^{k,\mathbf{r}}_{n,\mathbf{s}} = \exp\left(\frac{ie}{c\hbar}\int_s^{\mathbf{r}}\mathbf{A}(\mathbf{r})d\mathbf{r}\right)\int\rho^*_k\left(\left|\mathbf{R}-\frac{\mathbf{r}-\mathbf{s}}{2}\right|\right)\hat{h}(\mathbf{R})$$
$$\times\rho_n\left(\mathbf{R}+\frac{\mathbf{r}-\mathbf{s}}{2}d\mathbf{R}\right)\begin{pmatrix}-1/4 & -\exp(2i\psi)/4\\ -\exp(-2i\psi)/4 & -1/4\end{pmatrix}. \quad (7)$$

Here, ψ is the angle of rotation of the unit cell vectors. In our case, $\psi = \pi/2$.

We write Eqs. (4) for a square lattice in explicit form

$$i\frac{\partial\hat{a}_{\mathbf{p}}(t)}{\partial t} = 2[\cos(p_x a)+\cos(p_y a)]\hat{a}_{\mathbf{p}}(t)$$
$$+2[\cos(p_x a)-\cos(p_y a)]\hat{b}_{\mathbf{p}}(t)+\mu_B H\hat{a}_{\mathbf{p}}(t),$$
$$i\frac{\partial\hat{b}_{\mathbf{p}}(t)}{\partial t} = 2[\cos(p_x a)-\cos(p_y a)]\hat{a}_{\mathbf{p}}(t) \quad (8)$$
$$+2[\cos(p_x a)+\cos(p_y a)]\hat{b}_{\mathbf{p}}(t)-\mu_B\mathbf{H}\hat{a}_{\mathbf{p}}(t).$$

Here, $\hat{a}_{\mathbf{p}}(t)$ and $\hat{b}_{\mathbf{p}}(t)$ are the annihilation operators for the x and y electrons with momentum $\mathbf{p}$ and a is the unit cell size. The scalar jump integral is assumed to be equal to unity, which corresponds to the hole representation. To pass to the electron representation, it is sufficient to carry out the substitution $E \longrightarrow -E$; p_x, p_y are generalized quasimomenta determined by the vector potential. In the Landau gauge $\mathbf{A} = (0, Hx, 0)$, we have the general formula for a square lattice,

$$p_x = \hat{k}_x, \quad p_y = k_y - \frac{eH\hat{x}}{c}, \quad \hat{x} = i\frac{\partial}{\partial p_x}, \quad (9)$$

where k_y is the integral of motion determined from the cyclic condition in y. Since the representation of equations of motion in terms of the differential operator $\hat{x}$ is unsuitable for numerical computations, it is expedient to return to the equations of motion in the form of difference equations (in variable x) for a given $ak_y = \varphi$, $0 < \varphi < 2\pi$.

It is convenient to introduce the cell number along the x axis, $x = an$, and to eliminate the time dependence with the help of transitions to states with a preset energy E. As a result, we obtain the following system of difference equations:

$$E\hat{a}_{n,\varphi} = [\hat{a}_{n-1,\varphi}+\hat{a}_{n+1,\varphi}]+2\hat{a}_{n,\varphi}\cos(\varphi-\zeta n)$$
$$+[\hat{b}_{n-1,\varphi}+\hat{b}_{n+1,\varphi}]-2\hat{b}_{n,\varphi}\cos(\varphi-\zeta n)+\mu_B H\hat{a}_{n,\varphi},$$
$$E\hat{b}_{n,\varphi} = [\hat{b}_{n-1,\varphi}+\hat{b}_{n+1,\varphi}]-2\hat{a}_{n,\varphi}\cos(\varphi-\zeta n) \quad (10)$$
$$+[\hat{a}_{n-1,\varphi}+\hat{a}_{n+1,\varphi}]-2\hat{b}_{n,\varphi}\cos(\varphi-\zeta n)-\mu_B H\hat{b}_{n,\varphi}.$$

These equations generalize the Harper equations for a plane system of p electrons in a square lattice. It is important to note that there are two dimensionless parameters in the problem, i.e.,

$$u = \mu_B H, \quad \zeta = \frac{eHa^2}{c\hbar}, \quad \frac{u}{\zeta} = \frac{\hbar^2}{2ma^2} = \frac{m^*}{m}. \quad (11)$$

In experiment, each of these parameters is small, but their ratio depends to a considerable extent on the jump integral. For a small value of the jump integral, effective mass $m^* \gg m$, i.e., $\zeta \ll u$, which means that the Zeeman splitting is larger than the spacing between the Landau levels.

Equations (10) are difference equations in variable n, while variable φ is a parameter on which the unknown functions depend without any shift. For this reason, we can omit the dependence on φ. In view of the linearity of the equations under study, we will henceforth omit the operator symbols.

In order to transform Eqs. (10) to the same form (2) as for s electrons, we pass to new variables

$$x_n^+ = a_n + b_n; \quad x_n^- = a_n - b_n,$$

which gives

$$\begin{aligned} Ex_n^+ &= 2[x_{n-1}^+ + x_{n+1}^+] + \mu_B H x_n^-, \\ Ex_n^- &= 4x_n^- \cos(\varphi - \zeta n) + \mu_B H x_n^+. \end{aligned} \quad (12)$$

As a result, we obtain equations suitable for computations:

$$Ex_n^+ = 2[x_{n-1}^+ + x_{n+1}^+] + \frac{\mu_B^2 H^2 x_n^+}{E - 4\cos(\varphi - \zeta n)}. \quad (13)$$

We write these equations in the matrix form proposed by Hofstadter [2],

$$\hat{R}_{n+1} = \hat{\Gamma}_n \hat{R}_n,$$

where

$$\hat{\Gamma}_n = \begin{pmatrix} E - W(\epsilon, \varphi - \zeta n) & -1 \\ 1 & 0 \end{pmatrix}, \quad (14)$$

in which $\epsilon = E/2$, $r_n = \varphi - \zeta n$,

$$\Delta^2 = \frac{\mu_B^2 H^2}{4}, \quad W(\epsilon, r_n) = \frac{\Delta^2}{\epsilon - 2\cos r_n},$$
$$\hat{R}_n = \begin{pmatrix} x_n^+ \\ x_{n-1}^+ \end{pmatrix}, \quad \hat{R}_{n+1} = \begin{pmatrix} x_{n+1}^+ \\ x_n^+ \end{pmatrix}. \quad (15)$$

If $H = 0$, we have two independent solutions:

$$\epsilon = 2\cos\varphi, \quad \epsilon = 2\cos\theta. \quad (16)$$

For $\zeta = 2\pi k$, we also obtain two spectral branches:

$$\epsilon_\pm = (\cos\varphi + \cos\theta) \pm \sqrt{(\cos\theta - \cos\varphi)^2 + \Delta^2}. \quad (17)$$

We denote by $\hat{Q}_m$ the product of an arbitrary number m of matrices $\hat{\Gamma}_n$. It can be noted that the determinant of any of these matrices is equal to unity so that either both eigenvalues of matrix Q are modulo equal to unity, or one eigenvalue is modulo greater than unity, while the other eigenvalue is modulo smaller than unity. The condition of boundedness of the wave function for all n imposes the following condition on the eigenvalues of matrix Q: both eigenvalues must be modulo equal to unity. In other words, we have

$$\mathrm{Tr}\hat{Q}_m(\epsilon) = 2\cos\theta, \quad (18)$$

where θ is the second dimensionless quasimomentum. Proceeding from this relation, we find four spectral branches for $\zeta = \pi$:

$$\epsilon(\theta, \varphi) = \pm\sqrt{2 + \Delta^2 + \cos\theta + \cos 2\varphi \pm \sqrt{2\Delta^2[2 + \cos 2\varphi + \cos\theta] + [\cos\theta - \cos 2\varphi]^2}}.$$

In order to estimate the form of the energy spectrum qualitatively, we calculated the eigenvalues for rational values of parameter

$$\lambda = \frac{\zeta}{2\pi}, \quad \zeta = 2\pi\frac{n_1}{n_2}, \quad (19)$$

where n_1 and n_2 are integers such that $1 \le n_2 \le 36$ and $1 \le n_1 < n_2$ (see Fig. 1).

3. ENERGY SPECTRUM IN A TRIANGULAR LATTICE

For a triangular lattice, the energy spectrum for $H = 0$ has the form

$$\epsilon(\mathbf{k}) = 2[\cos(k_+ a) + \cos(k_- a) + \cos(-k_+ a + k_- a)], \quad (20)$$

where the jump integral is assumed to be equal to unity as before, and quasimomenta have the form

$$k_+ = \frac{k_x + \sqrt{3}k_y}{2}, \quad k_- = \frac{-k_x + \sqrt{3}k_y}{2}.$$

We direct the magnetic field along the z axis and use a symmetric gauge; i.e., we write the magnetic potential in the form

$$A_x = -\frac{yH}{2}, \quad A_y = \frac{xH}{2}, \quad A_z = 0.$$

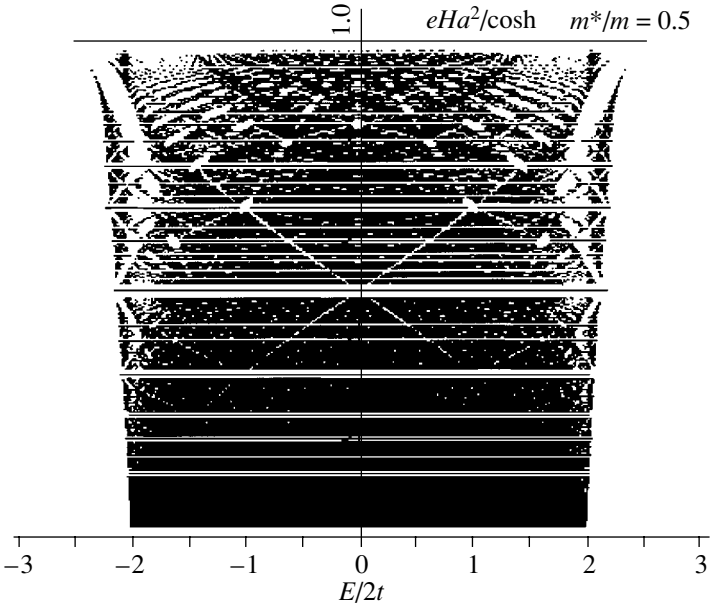

Fig. 1. Electronic spectra of p electrons in a square lattice.

As a result, we obtain the Hamiltonian

$$\begin{aligned}\hat{H}/2 = {} & \exp\left[\frac{a}{2}\frac{\partial}{\partial x} + \frac{\sqrt{3}a}{2}\frac{\partial}{\partial y} + i\frac{eaH}{4\hbar c}(-y+\sqrt{3}x)\right] \\ & + \exp\left[\frac{a}{2}\frac{\partial}{\partial x} - \frac{\sqrt{3}a}{2}\frac{\partial}{\partial y} - i\frac{eaH}{4\hbar c}(+y+\sqrt{3}x)\right] \\ & + \exp\left[-\frac{a}{2}\frac{\partial}{\partial x} + \frac{\sqrt{3}a}{2}\frac{\partial}{\partial y} + i\frac{eaH}{4\hbar c}(y+\sqrt{3}x)\right] \\ & + \exp\left[-\frac{a}{2}\frac{\partial}{\partial x} - \frac{\sqrt{3}a}{2}\frac{\partial}{\partial y} - i\frac{eaH}{4\hbar c}(-y+\sqrt{3}x)\right] \\ & + \exp\left[a\frac{\partial}{\partial x} - i\frac{eaH}{2c\hbar}y\right] + \exp\left[-a\frac{\partial}{\partial x} + i\frac{eaH}{2c\hbar}y\right].\end{aligned} \tag{21}$$

Operator $\exp\{\mathbf{a}\partial/\partial\mathbf{r}\}$ is the operator for the function argument shift by $\mathbf{a}$.

In order to reduce the equations to the Harper form, we pass to a nonorthogonal system of coordinates, assuming that e_x and $e_+ = e_x/2 + \sqrt{3}\,e_y/2$ are unit vectors; in this case, lattice sites have integral coordinates (n_x, n_+), and the transition to the new system of coordinate is described by the formulas

$$x = x_0 + \frac{1}{2}y_+, \quad y = \frac{\sqrt{3}}{2}y_+ \tag{22a}$$

or

$$x_0 = x - \frac{1}{\sqrt{3}}y, \quad y_+ = \frac{2}{\sqrt{3}}y. \tag{22b}$$

In this case, operators

$$\pm\frac{1}{2}\frac{\partial}{\partial x} + \frac{\sqrt{3}}{2}\frac{\partial}{\partial y}$$

are written in the form

$$\begin{aligned}\frac{1}{2}\frac{\partial}{\partial x} + \frac{\sqrt{3}}{2}\frac{\partial}{\partial y} &= \frac{\partial}{\partial y_+}, \\ -\frac{1}{2}\frac{\partial}{\partial x} + \frac{\sqrt{3}}{2}\frac{\partial}{\partial y} &= -\frac{\partial}{\partial x_0} + \frac{\partial}{\partial y_+}.\end{aligned} \tag{23}$$

We denote by $\zeta = eH\sqrt{3}\,a^2/4c\hbar$; then the Hamiltonian in the new variables can be written in the following symmetric form:

$$\begin{aligned}\frac{\hat{H}}{2} = {} & \exp\left[a\frac{\partial}{\partial y_+} + i\zeta\frac{x_0}{a}\right] + \exp\left[-a\frac{\partial}{\partial y_+} - i\zeta\frac{x_0}{a}\right] \\ & + \exp\left[-a\frac{\partial}{\partial x_0} + a\frac{\partial}{\partial y_+} + i\zeta\frac{(y_+ + x_0)}{a}\right] \\ & + \exp\left[a\frac{\partial}{\partial x_0} - a\frac{\partial}{\partial y_+} - i\zeta\frac{(y_+ + x_0)}{a}\right] \\ & + \exp\left[-a\frac{\partial}{\partial x_0} + i\zeta\frac{y_+}{a}\right] + \exp\left[a\frac{\partial}{\partial x_0} - i\zeta\frac{y_+}{a}\right].\end{aligned} \tag{24}$$

It should be noted that operator $\exp\{\mathbf{a}\partial/\partial\mathbf{r}\}$ is the operator of the function argument shift by $\mathbf{a}$; consequently,

the result of action of Hamiltonian (24) corresponds to the Harper equations:

$$\begin{aligned} E\Psi &= \Psi(x_0, y_+ + a)\exp\frac{i\zeta x_0}{a} \\ &+ \Psi(x_0, y_+ - a)\exp\left(-\frac{i\zeta x_0}{a}\right) \\ &+ \exp\left(-\frac{i\zeta(x_0 + y_+)}{a}\right)\Psi(x_0 + a, y_+ - a) \\ &+ \exp\left(\frac{i\zeta(x_0 + y_+)}{a}\right)\Psi(x_0 - a, y_+ + a) \\ &+ \exp(-i\zeta y_+)\Psi(x_0 + a, y_+) + \exp(i\zeta y_+)\Psi(x_0 - a, y_+). \end{aligned} \tag{25}$$

The solution for their solvability is the finiteness of the wave function for $\mathbf{r} \longrightarrow \infty$. It should be noted that the parameter $\zeta = |e|Ha^2\sqrt{3}/\hbar c$ of the problem is equal to the ratio of the magnetic flux through a unit cell to the flux quantum. Lattice sites have integral arguments. Consequently, Eq. (25) can be written in the form

$$\begin{aligned} E\Psi(n_x, n_y) &= \Psi(n_x, n_y + 1)\exp(i\zeta n_x) \\ &+ \Psi(n_x, n_y - 1)\exp(-i\zeta n_x) \\ &+ \exp(-i\zeta(n_x + n_y))\Psi(n_x + 1, n_y - 1) \\ &+ \exp[i\zeta(n_x + n_y)]\Psi(n_x - 1, n_y + 1) \\ &+ \exp(-i\zeta n_y)\Psi(n_x + 1, n_y) \\ &+ \exp(i\zeta n_y)\Psi(n_x - 1, n_y). \end{aligned} \tag{26}$$

We will seek the solution to this equation in the form

$$\Psi(n_x, n_y) = \exp\left(-\frac{\zeta i(n_x^2 - n_y^2)}{2}\right)W(n_x, n_y). \tag{27}$$

Function $W(n_x, n_y)$ satisfies the equation

$$\begin{aligned} W(n_x, n_y)E &= \exp(i\zeta(n_x + n_y))W(n_x, n_y + 1)\exp\frac{i\zeta}{2} \\ &+ \exp(-i\zeta(n_x + n_y))W(n_x, n_y - 1)\exp\frac{i\zeta}{2} \\ &+ \exp(-2i\zeta(n_x + n_y))W(n_x + 1, n_y - 1) \\ &+ \exp(2i\zeta(n_x + n_y))W(n_x - 1, n_y + 1) \\ &+ \exp(-i\zeta(n_x + n_y))\exp\left(-\frac{i\zeta}{2}\right)W(n_x + 1, n_y) \\ &+ \exp(i\zeta(n_x + n_y))\exp\left(-\frac{i\zeta}{2}\right)W(n_x - 1, n_y). \end{aligned} \tag{28}$$

It should be noted that the coefficients in Eq. (28) are functions of only the sum $n_x + n_y$.

This allows us to pass to the variables $N = n_x + n_y$, $M = n_x - n_y$ and seek the solution in the form

$$W(N, M) = e^{i\varphi M}S_N. \tag{29}$$

Then Eqs. (28) lead to a chain of linear difference equations

$$\begin{aligned} &S_N[E - 2\cos(2r_N)] \\ &= 2S_{N+1}\cos\left(r_N + \frac{\zeta}{2}\right) + 2S_{N-1}\cos\left(r_N - \frac{\zeta}{2}\right), \end{aligned} \tag{30}$$

where $r_N = N\zeta - \varphi$. To carry out numerical computations, it is convenient to write Eq. (30) in form (14),

$$\hat{G}_{N+1} = \hat{\Upsilon}_N\hat{G}_N,$$

where

$$\begin{aligned} \hat{\Upsilon}_N &= \begin{pmatrix} U(\epsilon, r_N, \zeta) & V(r_N, \zeta) \\ 1 & 0 \end{pmatrix}, \\ \hat{G}_N &= \begin{pmatrix} S_N \\ S_{N-1} \end{pmatrix}, \quad \hat{G}_{N+1} = \begin{pmatrix} S_{N+1} \\ S_N \end{pmatrix}, \end{aligned} \tag{31}$$

$$\begin{aligned} \epsilon &= \frac{E}{2}, \quad U(\epsilon, r_N, \zeta) = \frac{\epsilon - \cos 2r_N}{\cos(r_N + \zeta/2)}, \\ V(\beta, \zeta) &= -\frac{\cos(r_N - \zeta/2)}{\cos(r_N + \zeta/2)}, \quad r_N = N\zeta - \varphi. \end{aligned} \tag{32}$$

The physical condition imposed on the wave function is its boundedness for all values of N.

We assume that matrix $\hat{A}(N)$ is periodic in N with period m; i.e.,

$$2\pi(N + m)\kappa = 2\pi N\kappa + 2\pi i. \tag{33}$$

This means that parameter $\kappa = \zeta/2\pi$ is a rational number: $\kappa = i/m$ (see Fig. 2).

We denote by $\hat{Q}_k$ the product of an arbitrary number k of matrices $\hat{\Upsilon}_N$. It can be noted that the determinant of any of these matrices is equal to unity so that either both eigenvalues of matrix Q are modulo equal to unity, or one eigenvalue is modulo greater than unity, while the other eigenvalue is modulo smaller than unity. The condition of boundedness of the wave function for all n imposes the following condition on the magnitude of eigenvalues of matrix Q: both eigenvalues must be

modulo equal to unity. Thus, we again arrive at Eq. (18).

Let us write the solution to this equation in the simplest cases of integral values of parameter $2\pi/\zeta$. For $\zeta = \pi$, we have

$$\epsilon(\theta, \varphi)_{\pm} = \cos 2\varphi \pm 2\sin\frac{\theta}{2}. \quad (34a)$$

For $\zeta = 2\pi/3$, we have the following cubic equation for determining the spectrum:

$$\epsilon^3 + \frac{9}{4}\epsilon + \frac{3}{4} + \frac{1}{2}\cos 3\varphi\cos\theta - \frac{1}{4}\cos 6\varphi = 0. \quad (34b)$$

For $\zeta = \pi/2$, we have four spectral branches:

$$\epsilon(\theta, \varphi) = \pm\sqrt{1 + \cos^2 2\varphi \pm \cos\frac{\theta}{2}\cos 2\varphi}. \quad (34c)$$

4. ENERGY SPECTRUM FOR A HONEYCOMB LATTICE

We study phenomena associated with the quantum Hall effect dealing with a 2D gas formed at the interface between two binary compounds of the GaAs and $Al_xGa_{1-x}As$ type, each of which has an unfilled p shell. In the ideal case, a 2D honeycomb lattice of alternating cations and anions is formed at the interface (see Fig. 3). The average occupancies n_a and n_b for p electrons, e.g., for Ga (A) and As (B), satisfy the electroneutrality condition $n_a + n_b = 4$, which determines the value of the chemical potential $\mu = -(\epsilon_a + \epsilon_b)/2$. Here and below, we denote by ϵ_a and ϵ_b the one-particle energies of corresponding electrons. The difference $r = \epsilon_a - \epsilon_b$ is determined by electrostatic properties of this heterojunction and is regarded as a given energy parameter.

4.1. p_x, p_y Electrons

We assume that jumps occur between neighboring atoms each of which has a wave function proportional to x or y, while the p_z shell remains unfilled.

In order to study the magnetic properties associated with orbital splitting, we must choose two complex conjugate wave functions $x \pm iy$, characterized by different energies in a given magnetic field $\pm H$ (the magnetic field is measured in energy units).

The matrix of the integrals of transition between the nearest ions is proportional to the overlap integral, which is set equal to unity.

In the nearest neighbor approximation, the inverse one-particle Green function can be written in the form

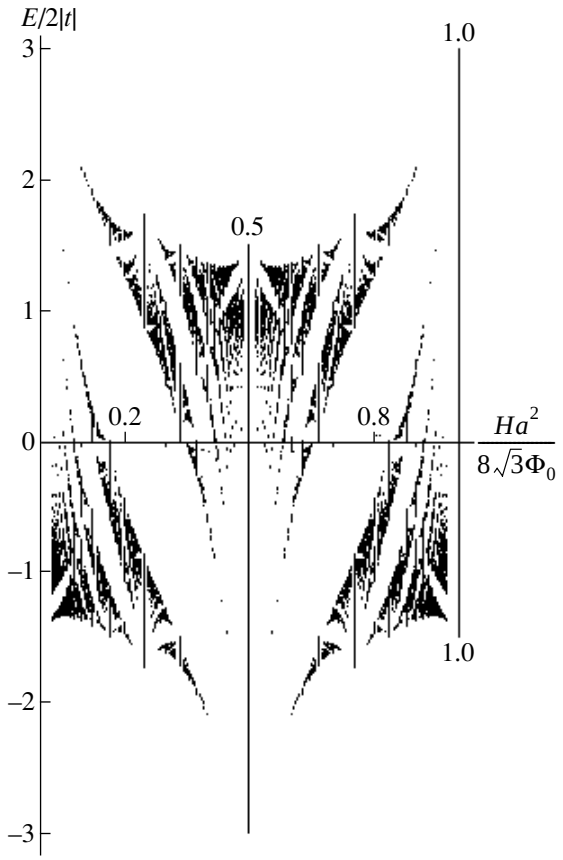

Fig. 2. Energy spectrum of s electrons in a triangular lattice.

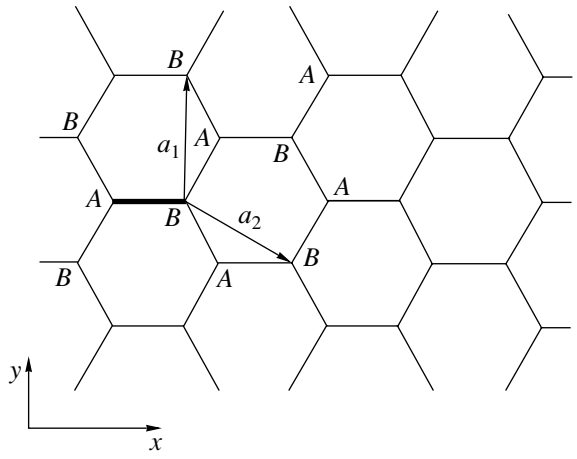

Fig. 3. Unit cell of a honeycomb lattice.

of the matrix

$$G^{-1}(E) = \begin{pmatrix} E - \epsilon_a^+ & 0 & \alpha & \beta \\ 0 & E - \epsilon_a^- & \gamma & \delta \\ \alpha^* & \gamma^* & E - \epsilon_b^+ & 0 \\ \beta^* & \delta^* & 0 & E - \epsilon_b^- \end{pmatrix}. \quad (35)$$

Here, the following notation is introduced:

$$\epsilon_a^{\pm} = \frac{r}{2} - \mu \pm H, \quad \epsilon_b^{\pm} = -\frac{r}{2} - \mu \pm H,$$

$$\alpha = \delta = 1 + \nu_1^* + \nu_2^*,$$

$$\beta = 1 + \nu_1^* \exp(-2i\psi) + \nu_2^* \exp(-2i\psi),$$
$$\gamma = 1 + \nu_1^* \exp(-2i\psi) + \nu_2^* \exp(2i\psi), \tag{36}$$

$$\nu_k = \exp(i\mathbf{q} \cdot \mathbf{r}_k),$$

$$\psi = \frac{\pi}{3}, \quad \mathbf{r}_1 = \left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right), \quad \mathbf{r}_2 = \left(\frac{\sqrt{3}}{2}, -\frac{1}{2}\right).$$

In this case, the matrix elements are calculated from atomic wave functions corresponding to the unit orbital angular momentum and differing in the sign of its component.

Thus, the equation for determining the electron energy spectrum in zero magnetic field has the form

$$\begin{aligned} E_+^a x_1 + \alpha x_3 + \beta x_4 &= 0, \\ E_-^a x_2 + \gamma x_3 + \delta x_4 &= 0, \\ E_+^b x_3 + \alpha^* x_1 + \gamma^* x_2 &= 0, \\ E_-^b x_4 + \beta^* x_1 + \delta^* x_2 &= 0, \end{aligned} \tag{37}$$

where $E_{\pm}^a = E - \epsilon_a \mp H$, $E_{\pm}^b = E - \epsilon_b \mp H$. The effect of magnetic field on the orbital motion of electrons between unit cells is determined by the gradient transformation of the phase:

$$\mathbf{q} \longrightarrow \hat{p} = -i\hbar\frac{\partial}{\partial \mathbf{r}} + \frac{e}{c}\mathbf{A}, \quad \mathbf{A} = (0, Hx, 0). \tag{38}$$

Subsequent transformations are reduced to the expression of the exponents of the differentiation operators in formulas (36) for quantities ν_k in terms of finite shifts along the unit vectors of a unit cell. All these transformations are given in the Appendix.

As a result, we obtain a system of difference equations with real coefficients (see formulas (A.7)):

$$\begin{aligned} &E_+^a x_1(n+1) + S_{n+1}(34) + 2x_3(n)\cos\theta_n \\ &\quad + 2x_4(n)\cos\theta_n^+ = 0, \\ &E_-^a x_2(n+1) + S_{n+1}(34) + 2x_4(n)\cos\theta_n \\ &\quad + 2x_3(n)\cos\theta_n^- = 0, \\ &E_+^b x_3(n) + S_n(12) + 2x_1(n+1)\cos\theta_n \\ &\quad + 2x_2(n+1)\cos\theta_n^- = 0, \\ &E_-^b x_4(n) + S_n(12) + 2x_2(n+1)\cos\theta_n \\ &\quad + 2x_1(n+1)\cos\theta_n^+ = 0. \end{aligned} \tag{39}$$

Here, the notation

$$\theta_n = \frac{eH}{4c}\left(n + \frac{1}{2}\right) + \frac{\varphi}{2}, \quad \theta_n^{\pm} = \theta_n \pm \frac{2\pi}{3},$$
$$S_n(12) = x_1(n) + x_2(n),$$
$$S_n(34) = x_3(n) + x_4(n) \tag{40}$$

has been introduced, where φ is an arbitrary phase having the meaning of conserved quasimomentum.

In order to transform system (39) to the form used by Hofstadter, we solve it in $x_{1,2}(1+n)$ and $x_{3,4}(n)$ for the given $S_n(12)$ and $S_{1+n}(34)$:

$$\begin{aligned} x_1(1+n) &= -\frac{2}{3}E_+^b x_3(n)\cos\theta_n + \frac{2}{3}E_-^b x_4(n)\cos\theta_n^- \\ &\quad + \frac{2}{3}S_n(12)[\cos\theta_n^- - \cos\theta_n], \\ x_2(1+n) &= -\frac{2}{3}E_-^b x_4(n)\cos\theta_n + \frac{2}{3}E_+^b x_3(n)\cos\theta_n^+ \\ &\quad + \frac{2}{3}S_n(12)[\cos\theta_n^+ - \cos\theta_n], \\ x_3(n) &= -\frac{2}{3}E_+^a x_1(n+1)\cos\theta_n + \frac{2}{3}E_-^a x_2(n+1)\cos\theta_n^+ \\ &\quad + \frac{2}{3}S_{n+1}(34)[\cos\theta_n^+ - \cos\theta_n], \\ x_4(n) &= \frac{2}{3}E_+^a x_1(n+1)\cos\theta_n^- - \frac{2}{3}E_-^a x_2(n+1)\cos\theta_n \\ &\quad + \frac{2}{3}S_{n+1}(34)[\cos\theta_n^- - \cos\theta_n]. \end{aligned} \tag{41}$$

We solve the obtained system of equations for $x_{1,2}(1+n)$ and $x_{3,4}(n)$, after which we find $x_3(n) + x_4(n) = S_n(34)$

and, hence, the relation between the given $S_{n+1}(34)$ and $S_n(34)$, $S_n(12)$:

$$S_{(1+n)}(12) = S_{(n)}(12)\frac{U_n(\theta_n)}{\Delta_n(\theta_n)} + S_{(1+n)}(34)\frac{W_{12}(\theta_n)}{\Delta_n(\theta_n)},$$
$$S_{(1+n)}(34) = -S_{(n)}(12)\frac{W_{34}(\theta_n)}{U_n(\theta_n)} + S_{(n)}(34)\frac{\Delta_n(\theta_n)}{U_n(\theta_n)}. \quad (42)$$

Here, we have introduced the following notation:

$$U_n(\theta_n) = -2\cos\theta_n + \frac{2}{9}[(E_+^aE_+^b + E_-^aE_-^b)\cos\theta_n + E_+^aE_-^b\cos\theta_n^- + E_-^aE_+^b\cos\theta_n^+], \quad (43)$$

$$W_{12}(\theta_n) = \frac{4}{9}[E_+^b(\cos\theta_n - \cos\theta_n^+)^2 - E_-^b(\cos\theta_n - \cos\theta_n^-)^2] - \frac{1}{9}E_+^bE_-^b(E_+^a + E_-^a),$$
$$W_{34}(\theta_n) = \frac{4}{9}[E_+^a(\cos\theta_n - \cos\theta_n^-)^2 + E_-^a(\cos\theta_n - \cos\theta_n^+)^2] - \frac{1}{9}E_+^aE_-^a(E_+^b + E_-^b). \quad (44)$$

The determinant of system (41) has the form

$$\Delta_n(\theta_n) = 1 - \frac{4}{9}[(E_+^aE_+^b + E_-^aE_-^b)\cos^2\theta_n + E_+^aE_-^b\cos^2\theta_n^- + E_-^aE_+^b\cos^2\theta_n^+] + \frac{1}{9}E_+^aE_-^aE_+^bE_-^b. \quad (45)$$

Thus, we obtain a system of two recurrence equations,

$$\begin{pmatrix} S_{n+1}(12) \\ S_{n+1}(34) \end{pmatrix} = \hat{R}\begin{pmatrix} S_n(12) \\ S_n(34) \end{pmatrix}, \quad (46)$$

where

$$\hat{R} = \begin{pmatrix} \dfrac{U_n^2(\theta_n) - W_{12}(\theta_n)W_{34}(\theta_n)}{\Delta_n(\theta_n)U_n(\theta_n)} & \dfrac{W_{12}(\theta_n)}{U_n(\theta_n)} \\ -\dfrac{W_{34}(\theta_n)}{U_n(\theta_n)} & \dfrac{\Delta_n(\theta_n)}{U_n(\theta_n)} \end{pmatrix}.$$

It can easily be verified that the determinant of 2D matrix $\hat{R}$ is equal to unity.

Direct computations show that the following relation holds:

$$U_n^2(\theta_n) - W_{12}(\theta_n)W_{34}(\theta_n) = \Delta_n(\theta_n)V_n(\theta_n); \quad (47)$$

here, we have

$$V_n(\theta_n) = 4\cos^2\theta_n - \frac{1}{9}(E_a^+ + E_a^-)(E_b^+ + E_b^-). \quad (48)$$

It can be concluded from here that matrix $\hat{R}$ has the form

$$\hat{R} = \begin{pmatrix} V_n(\theta_n)/U_n(\theta_n) & W_{12}(\theta_n)/U_n(\theta_n) \\ -W_{34}(\theta_n)/U_n(\theta_n) & \Delta_n(\theta_n)/U_n(\theta_n) \end{pmatrix}. \quad (49)$$

The obtained system has coefficients periodically depending on the number of flux quanta passing through a unit cell. The equation for determining the excitation spectrum has the canonical form (18).

In the limiting case when the intrinsic orbital angular momentum can be disregarded, we have $E_\pm^{a,b} = E^{a,b}$ and

$$U_n(\theta_n) \longrightarrow u_n(\theta) = 2\cos\theta_n\left[-1 + \frac{1}{9}E_aE_b\right],$$
$$V_n(\theta_n) \longrightarrow v_n(\theta) = 4\cos^2\theta_n - \frac{4}{9}E_aE_b,$$
$$W_{12}(\theta_n) \longrightarrow w_{12}(\theta_n) = E_bw_n(\theta_n),$$
$$W_{34}(\theta_n) \longrightarrow w_{34}(\theta_n) = E_aw_n(\theta_n), \quad (50)$$

where

$$w_n(\theta_n) = \frac{2}{3}\left\{[1 + 2\cos^2\theta_n] - \frac{1}{3}E^bE_a\right\}. \quad (51)$$

We can now carry out the scale transformation $S_{(n)}(12) = s_{(n)}(12)\sqrt{E_b}$, $S_{(n)}(34) = s_{(n)}(34)\sqrt{E_b}$. As a result, the elements of matrix $\hat{R}$ become functions of the product E_aE_b only:

$$\hat{R} = \begin{pmatrix} v_n(\theta_n)/u_n(\theta_n) & \dfrac{\sqrt{E_aE_b}w_n(\theta_n)}{u_n(\theta_n)} \\ -\dfrac{\sqrt{E_aE_b}w_n(\theta_n)}{u_n(\theta_n)} & \dfrac{\Upsilon_n(\theta_n)}{u_n(\theta_n)} \end{pmatrix}, \quad (52)$$

where

$$\Upsilon_n(\theta_n) = 1 - \frac{4}{9}E_aE_b\left[\frac{3}{2} + \cos^2\theta_n\right] + \frac{1}{9}(E_aE_b)^2. \quad (53)$$

Thus, in this limit, it is sufficient to analyze the dispersion equations in this limit in two variables $F = E_aE_b$ and θ with the help of the general relation (18). In this

case, the energy of the system is expressed in terms of F and energy shift through the formulas $E = \pm\sqrt{r^2 + F} + \mu$.

4.2. p_z and s Electrons in a Honeycomb Lattice

We assume that jumps occur between neighboring atoms, each of which has a wave function independent of coordinates x and y.

For this reason, the equations for real amplitudes have the form (see Eqs. 39)

$$\begin{aligned} E_a x_1(n+1) + x_3(n+1) + 2x_3(n)\cos\theta_n &= 0, \\ E_b x_3(n) + x_1(n) + 2x_1(n+1)\cos\theta_n &= 0. \end{aligned} \tag{54}$$

This immediately leads to the system of recurrence relations of type (49),

$$\begin{pmatrix} x_1(n+1) \\ x_3(n+1) \end{pmatrix} = \hat{P}\begin{pmatrix} x_1(n) \\ x_3(n) \end{pmatrix}, \tag{55}$$

where

$$\hat{P} = \begin{pmatrix} -\dfrac{1}{2\cos\theta_n} & -\dfrac{E_b}{2\cos\theta_n} \\ \dfrac{E_a}{2\cos\theta_n} & \dfrac{E_aE_b - 4\cos^2\theta_n}{2\cos\theta_n} \end{pmatrix}.$$

It can easily be verified that the 2D transformation matrix corresponding to system (55) has a determinant equal to unity.

Our task is simplified significantly if we introduce the scale transformation $x_1 = y_1\sqrt{E_b}$, $x_3 = y_3\sqrt{E_a}$. As a result, matrix $\hat{P}$ depends only on products E_aE_b:

$$\hat{P} = \begin{pmatrix} -\dfrac{1}{2\cos\theta_n} & -\dfrac{\sqrt{E_aE_b}}{2\cos\theta_n} \\ \dfrac{\sqrt{E_aE_b}}{2\cos\theta_n} & \dfrac{E_aE_b - 4\cos^2\theta_n}{2\cos\theta_n} \end{pmatrix}. \tag{56}$$

Therefore, we can analyze the dispersion equations in two variables $F = E_aE_b$ and θ using the general relation (18). In this case, the energy of the system can be expressed in terms of F, and the shift, with the help of the formula $E = \pm\sqrt{r^2 + F} + \mu$.

By way of an example, we can write the excitation spectrum for $eH/c\hbar = \pi$:

$$\epsilon(\varphi, \omega) = \pm\sqrt{r^2 + 3 - 2\cos\varphi \pm 4\sin(\omega/2)\sin(\varphi/2)} - \mu.$$

The density of states of s and p electrons in the honeycomb lattice for $m^* \ll m$ are shown in Fig. 4; it can be seen that the graphs are identical.

5. SUSCEPTIBILITY IN THE LIMIT OF WEAK FIELDS

It can be seen from Figs. 1, 2, and 4 that the excitation spectrum in the limit $H \longrightarrow 0$ becomes continuous, but it cannot be calculated. For this reason, we consider the limiting case using conventional perturbation theory.

In order to calculate the orbital susceptibility, we will use the Fourier expansion of the vector potential satisfying the transverse gauge [11]

$$\mathrm{div}\mathbf{A} = 0, \quad \mathbf{A}(\mathbf{q}) = (A_x(q_y), 0, 0). \tag{57}$$

Then we determine the first current correction for a given shape $\xi_\mathbf{p}$ of the Fermi surface calculated for zero field:

$$\langle j_x\rangle = -2Te^2 \times \sum_{\omega,\,\mathbf{p}}\left[v_s^2 G_\omega\left(\mathbf{p} + \frac{\mathbf{q}}{2}\right)G_\omega\left(\mathbf{p} - \frac{\mathbf{q}}{2}\right) + \xi''_{xx}G_\omega(\mathbf{p})\right]A_x(\mathbf{q}). \tag{58}$$

Here, $G_\omega(\mathbf{p}) = (i\omega - \xi_\mathbf{p})^{-1}$ is the thermodynamic Green function and v_s is the symmetrized velocity operator in the direction of the x axis:

$$\begin{aligned} v_s &= \frac{1}{2}\left[\frac{\partial}{\partial p_x}\xi\left(\mathbf{p} + \frac{\mathbf{q}}{2}\right) + \frac{\partial}{\partial p_x}\xi\left(\mathbf{p} - \frac{\mathbf{q}}{2}\right)\right], \\ \mathbf{q} &= (0, q_y, 0), \quad \xi''_{xx} = \frac{\partial^2}{\partial p_x^2}\xi(\mathbf{p}). \end{aligned} \tag{59}$$

In order to find the orbital susceptibility in a uniform magnetic field, we must expand the integrand into a power series in wave vector $\mathbf{q}$. Integrating the second term by parts, we note that the first sum is cancelled out for $\mathbf{q} = 0$. The first nonvanishing term turns out to be proportional to q_y^2 and the corresponding coefficient of $A_xq_y^2$ is precisely the orbital magnetic susceptibility χ [11]:

$$\chi(T) = \frac{e^2}{6}\sum\left\{\left[\frac{\partial v_x}{\partial p_x}\frac{\partial v_y}{\partial p_y} - \frac{\partial v_y}{\partial p_x}\frac{\partial v_x}{\partial p_y}\right] + \frac{3}{2}\left[v_x\frac{\partial^2 v_x}{\partial p_y^2} + v_y\frac{\partial^2 v_y}{\partial p_x^2}\right]\right\}n'_F(\xi_\mathbf{p}). \tag{60}$$

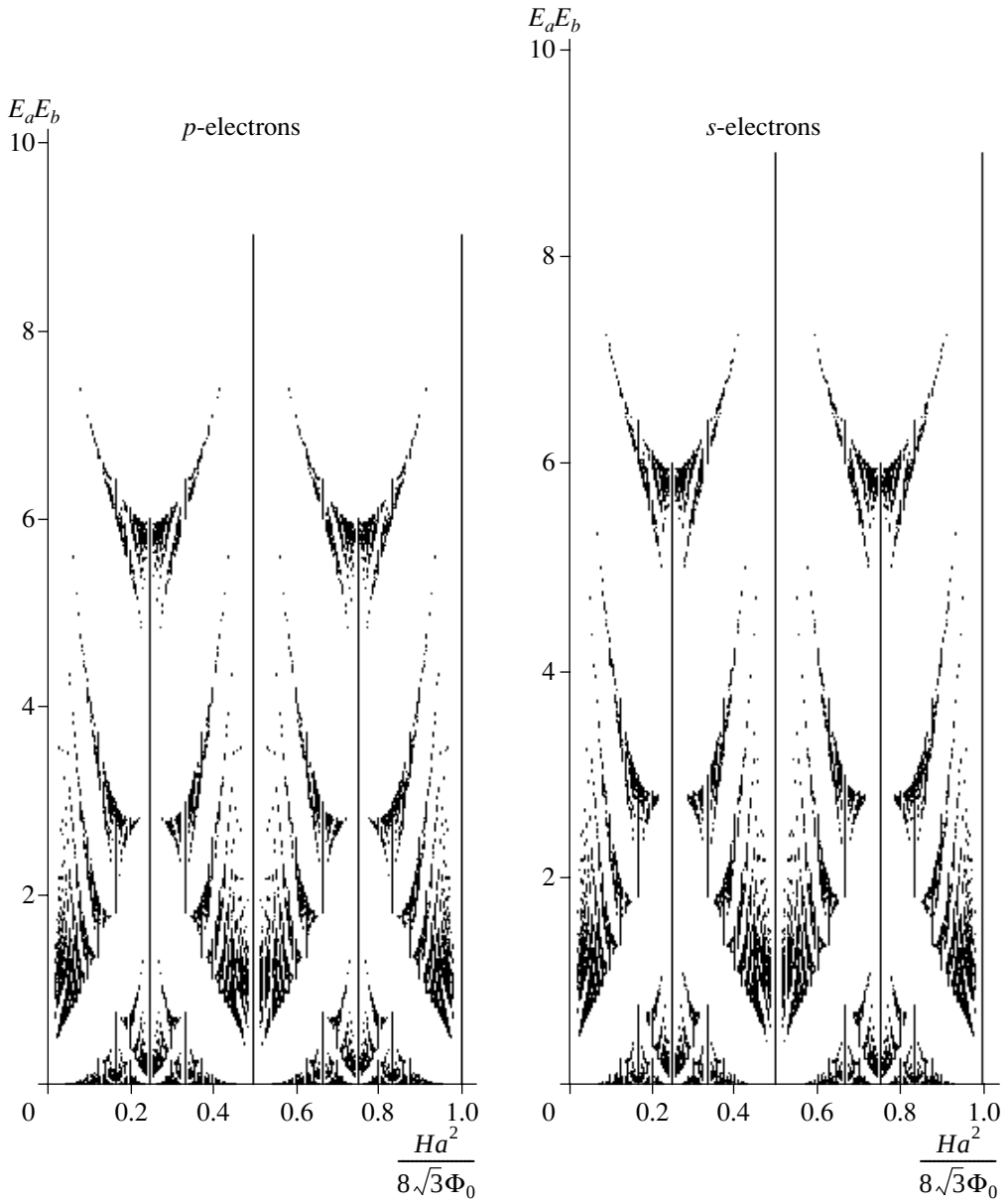

Fig. 4. Energy spectra of p and s electrons in a honeycomb lattice for $m^* \ll m$.

Here, $v_\alpha = \partial\xi/\partial p_\alpha$ are the velocity components and summation over momenta is carried out within the first Brillouin zone.

For a low concentration of particles, integration is carried out in the vicinity of the minimal or maximal value of the energy surface; the value of the curvature tensor

$$\frac{\partial v_x}{\partial p_x}\frac{\partial v_y}{\partial p_y} - \frac{\partial v_y}{\partial p_x}\frac{\partial v_x}{\partial p_y} = \frac{\partial^2\xi}{\partial p_x^2}\frac{\partial^2\xi}{\partial p_y^2} - \left(\frac{\partial^2\xi}{\partial p_x \partial p_y}\right)^2 \quad (61)$$

is positive and can be reduced to the product of reciprocal masses. Consequently, in all cases of cubic, triangular, and square lattices, relation (60) is reduced to the Landau diamagnetism formula.

At the points of Van Hove singularity, the velocity components vanish and the curvature tensor (61) is found to be negative. For this reason, for the Fermi energy intersecting the Van Hove singular points, the susceptibility is positive and turns to infinity at zero temperature.

For a finite temperature, we have $\chi \approx \ln(1/T)$ for 2D lattices and $\chi \approx \ln^2(1/T)$ for bcc lattices.

6. DISCUSSION: INTERMEDIATE FIELDS

For a finite value of magnetic field, when the number of flux quanta through a unit cell is a partial fraction N/K, the size of the unit cell increases by a factor of K,

which corresponds to a proportional K-fold decrease in the size of the first Brillouin zone and to the transformation of Brillouin zone into the Bravais–Zak subband (see detailed article by Zak [9]).

A. Let us first consider the case when $N = 1$.

For square and triangular lattices of s electrons, Eqs. (18) give exactly K solutions for each given K, each solution being dependent on two quasimomenta θ and φ.

In the case of a square lattice of p_x, p_y electrons, for each K we obtain $2K$ solutions from Eqs. (15). As the magnetic field increases, the division into bands pertaining to the upper and lower orbital doublets becomes ever clearer. However, for a small effective mass ($m^* \ll m$), a considerable overlapping of subbands belonging to different components of the intrinsic orbital angular momentum takes place.

In the case of a honeycomb lattice with two atoms in a cell and for p_z electrons, the division into the upper and lower subgroups of states is well pronounced; each subgroup contains K subbands arising as solutions to the corresponding system of equations (55).

In the general case of p_x, p_y electrons in a honeycomb lattice, the arrangement of levels is again determined by Harper-type equations (49). In this case, there are two independent parameters, viz., orbital splitting proportional to the magnetic field and the splitting associated with the electrostatic potential difference at the site where different atoms are located. It follows from Eqs. (49) that, for a given number m, there exist $4K$ states each of which depends on two quasimomenta.

B. $N \neq 1$.

It was shown by Dzyaloshinski (10) that number N determines the degeneracy of each quantum state calculated for a given K. This circumstance makes it possible to explain qualitatively the reasons for the emergence of the fractional Hall effect. The values of Coulomb matrix elements calculated from Wannier-type wave functions turn out to be much larger than the width of an individual magnetic subband. For this reason, it is natural to assume that the corresponding Hubbard energy is infinitely large.

In this approximation, the filling of the lower magnetic subband occurs through the so-called end factors equal to the sum of the mean occupation numbers of the initial and final states, $f = n_0 + n_1$.

Let us consider for the sake of simplicity the filling of the lower subband of s electrons, which corresponds to transitions from an empty state to an N-fold degenerate one-particle state. In this case, the equation of state can be expressed in terms of the product of the degeneracy by the end factor [12]:

$$n = Nf\sum_{\mathbf{q}} n_F(f\epsilon_1(\mathbf{q}) - \mu), \quad f = n_0 + n_1. \tag{62}$$

Here, n is the average number of particles per unit cell, n_0 and n_1 are average occupation numbers of empty and one-particle states, and $\epsilon_1(\mathbf{q})$ are solutions of the Harper equations for an empty square or triangular lattice, which correspond to the lower level depending on the 2D quasimomentum $\mathbf{q}$. In the limiting case of an infinitely large Hubbard energy, quantities n_0 and n_1 can be expressed in terms of degeneracy N and the average number n of electrons with the help of the normalization condition and the definition of n_1 in terms of degeneracy,

$$n_0 + Nn_1 = 1, \quad n_1 = \frac{n}{N}. \tag{63}$$

It is also convenient to replace the sum over quasimomenta by integrals with respect to the angles corresponding to the first magnetic subband. As a result, we obtain the explicit form of the equation of state:

$$n = Nf(n)\int_0^{2\pi}\int_0^{2\pi} n_F[f(n)\epsilon_1(\alpha,\beta) - \mu]\frac{d\alpha d\beta}{(2\pi)^2},$$
$$f(h) = 1 - n\frac{N-1}{N}. \tag{64}$$

It is convenient to write this equation in terms of density of states. For $T = 0$, we have

$$n = \frac{NS}{1 + S(N-1)}, \quad S = \int^{\mu/f} \rho_1(E)dE,$$
$$\rho_1(E) = \int_0^{2\pi}\int_0^{2\pi} \delta[E - \epsilon_1(\alpha,\beta)]\frac{d\alpha d\beta}{(2\pi)^2}. \tag{65}$$

The values of the emerging energy gaps are determined by the corresponding Coulomb matrix elements. However, the arrangement of these gaps on the energy scale coincides with the energy positions of Van Hove singularities within each allowed magnetic subbands. For a given finite field, the position of each correlation gap is determined by the condition of passage of the Fermi surface through one or several equivalent Van Hove points.

The position of each point in this case can be determined from an analysis of the corresponding equation of motion (12), (31), (46), or (49).

We can prove that the Harper equations for a square lattice determine the only Van Hove surface, for which we obtain fractional values of $S = S_1 = 1/2$ for $T = 0$.

For a triangular lattice, the Harper equations give only two Van Hove surfaces for which we have $S = S_1 = 1/4$ or $S = S_2 = 3/4$ for $T = 0$.

At the points of Van Hove singularities, the density of states becomes infinitely large. Accordingly, the sus-

ceptibility and conductivity at these points also turn to infinity at these points at $T = 0$. However, these singularities disappear even for a slight increase in temperature.

Thus, for $T = 0$, we obtain series of critical singularities. For a square lattice, we have

$$n_{c1} = \frac{N}{1+N} \text{ or } \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \ldots. \tag{66}$$

For a triangular lattice, we must distinguish between the cases of even and odd number of levels, i.e., between the cases with even and odd number K.

For an even K, pair overlapping takes place; i.e., we can see $M = K/2$ narrow subbands in the limit of large $K = 2M$. It follows hence that the degeneracy of each level is doubled in this case; i.e., the number N in the equation of state (65) assumes only even values.

In the case of an odd K, the number of subbands coincides with the number K itself. Therefore, the degeneracy of each state coincides with the number N, which runs through the entire series of integral values.

Thus, in the case of a triangular lattice, we have two series of critical values:

$$n_{c1} = \frac{N}{3+N} \text{ or } n_{c2} = \frac{3N}{1+3N}. \tag{67}$$

In the case of even values of K, the numbers N are even: $N = 2, 4, 6, \ldots$. In the case of odd values of K, N can be any integer: $N = 1, 2, 3, 4, 5, 6, \ldots$.

In the case of square and honeycomb lattices, we can also obtain various series of critical values for p electrons, which correspond to the Van Hove-type Fermi surfaces; this case should be considered separately.

APPENDIX

Transition to Harper Equations

In order to single out the effect of the displacement operators

$$\exp\left(\pm i\frac{\sqrt{3}}{2\hbar}\hat{p}_x\right)\Psi(x,y) = \Psi\left(x \mp \frac{\sqrt{3}}{2}, y\right),$$
$$\exp\left(\pm i\frac{1}{2\hbar}\hat{p}_x\right)\Psi(x,y) = \Psi\left(x, y \mp \frac{1}{2}\right), \tag{A.1}$$

we can use the general formula, which is valid under the condition that $[\hat{A}, \hat{B}]$ is not an operator:

$$\exp(\hat{A}+\hat{B}) = \exp\hat{A}\exp\hat{B}\exp\left(-\frac{1}{2}[\hat{A},\hat{B}]\right). \tag{A.2}$$

This leads to the following four relations:

$$\nu_1^*\Psi(x,y) = \exp\left\{-\frac{i\sqrt{3}eH}{8c} + \frac{ieH\sqrt{3}n_1}{4c}\right\}$$
$$\times\Psi\left(\frac{\sqrt{3}}{2}(n_1-1), \frac{1}{2}(n_2+1)\right)$$
$$\times\exp\left\{-\frac{i\sqrt{3}eH}{8c} + \frac{ieH\sqrt{3}n_1}{4c}\right\}\Psi(\mathbf{r}-\mathbf{a}_2)$$
$$= \exp\left\{\frac{ieH\sqrt{3}(n_1-1/2)}{4c}\right\}\psi(n_1-1, n_2). \tag{A.3a}$$

Here and below, we use the relations written in accordance with Fig. 3 ($x = n_1\sqrt{3}/2$, $y = n_2 - n_1/2$):

$$\nu_2^*\Psi(x,y)$$
$$= \exp\left(-\frac{i}{\hbar}\left(-\frac{\sqrt{3}}{2}i\hbar\frac{\partial}{\partial x} - \frac{1}{2}i\hbar\frac{\partial}{\partial y} + \frac{e}{2c}Hx\right)\right)\Psi(x,y)$$
$$= \exp\left(-\frac{i}{\hbar}\left(-\frac{\sqrt{3}}{2}i\hbar\frac{\partial}{\partial x} + \frac{e}{2c}Hx\right)\right)\Psi\left(x, y-\frac{1}{2}\right)$$
$$= \exp\left(-\frac{ieHx^2}{2c\sqrt{3}}\right)\exp\left(\frac{ieH(x-\sqrt{3}/2)^2}{2c\sqrt{3}}\right)\Psi\left(x-\frac{\sqrt{3}}{2}, y-\frac{1}{2}\right)$$
$$= \exp\left(\frac{i\sqrt{3}eH}{8c}\right)\exp\left(-\frac{ieHx}{2c}\right)\Psi\left(x-\frac{\sqrt{3}}{2}, y-\frac{1}{2}\right)$$
$$= \exp\left(\frac{i\sqrt{3}eH}{8c}\right)\exp\left(-\frac{ieHx}{2c}\right)\Psi(\mathbf{r}-\mathbf{a}_1)$$
$$= \exp\left(-\frac{ieH\sqrt{3}(n_1-1/2)}{4c}\right)\psi(n_1-1, n_2-1), \tag{A.3b}$$

$$\nu_1\Psi(x,y)$$
$$= \exp\left(\frac{i}{\hbar}\left(-\frac{\sqrt{3}}{2}i\hbar\frac{\partial}{\partial x} + \frac{1}{2}i\hbar\frac{\partial}{\partial y} - \frac{e}{2c}Hx\right)\right)\Psi(x,y)$$
$$= \exp\left(\frac{i}{\hbar}\left(-\frac{\sqrt{3}}{2}i\hbar\frac{\partial}{\partial x} - \frac{e}{2c}Hx\right)\right)\Psi\left(x, y-\frac{1}{2}\right)$$
$$= \exp\left(\frac{ieHx^2}{2c\sqrt{3}}\right)\exp\left(-\frac{ieH(x+\sqrt{3}/2)^2}{2c\sqrt{3}}\right)\Psi\left(x+\frac{\sqrt{3}}{2}, y-\frac{1}{2}\right)$$
$$= \exp\left(-\frac{i\sqrt{3}eH}{8c}\right)\exp\left(-\frac{ieHx}{2c}\right)\Psi\left(x+\frac{\sqrt{3}}{2}, y-\frac{1}{2}\right) \tag{A.3c}$$

$$= \exp\left(-\frac{ieH\sqrt{3}(n_1 + 1/2)}{4c}\right)\psi(n_1 + 1, n_2),$$

$$\nu_2\Psi(x, y)$$

$$= \exp\left(\frac{i}{\hbar}\left(-\frac{\sqrt{3}}{2}i\hbar\frac{\partial}{\partial x} - \frac{1}{2}i\hbar\frac{\partial}{\partial y} + \frac{e}{2c}Hx\right)\right)\Psi(x, y)$$

$$= \exp\left(\frac{i}{\hbar}\left(-\frac{\sqrt{3}}{2}i\hbar\frac{\partial}{\partial x} + \frac{e}{2c}Hx\right)\right)\Psi\left(x, y + \frac{1}{2}\right)$$

$$= \exp\left(-\frac{ieHx^2}{2c\sqrt{3}}\right)\exp\left(\frac{ieH(x + \sqrt{3}/2)^2}{2c\sqrt{3}}\right)\Psi\left(x + \frac{\sqrt{3}}{2}, y + \frac{1}{2}\right)$$

$$= \exp\left(\frac{i\sqrt{3}eH}{8c}\right)\exp\left(\frac{ieHx}{2c}\right)\Psi\left(x + \frac{\sqrt{3}}{2}, y + \frac{1}{2}\right) \quad \text{(A.3d)}$$

$$= \exp\left(\frac{i\sqrt{3}eH}{8c}\right)\exp\left(\frac{ieHx}{2c}\right)\Psi(\mathbf{r} - \mathbf{a}_1)$$

$$= \exp\left(\frac{ieH\sqrt{3}(n_1 + 1/2)}{4c}\right)\psi(n_1 + 1, n_2 + 1).$$

In this way, we can transform Eqs. (37) to a system of difference equations:

$$E_+^a x_1(n_1, n_2) + x_3(n_1, n_2)$$

$$+ \exp\left(-\frac{i\sqrt{3}eH}{8c}\right)\exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right)x_3(n_1 - 1, n_2)$$

$$+ \exp\left(\frac{i\sqrt{3}eH}{8c}\right)\exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right)x_3(n_1 - 1, n_2 - 1)$$

$$+ x_4(n_1, n_2) + \exp(2i\varphi) \quad \text{(A.4a)}$$

$$\times \exp\left(-\frac{i\sqrt{3}eH}{8c}\right)\exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right)x_4(n_1 - 1, n_2)$$

$$+ \exp(-2i\varphi)\exp\left(\frac{i\sqrt{3}eH}{8c}\right)\exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right)$$

$$\times x_4(n_1 - 1, n_2 - 1) = 0,$$

$$E_-^a x_2(n_1, n_2) + x_3(n_1 n_2) + \exp(-2i\varphi)$$

$$\times \exp\left(-\frac{i\sqrt{3}eH}{8c}\right)\exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right)x_3(n_1 - 1, n_2)$$

$$+ \exp(2i\varphi)\exp\left(\frac{i\sqrt{3}eH}{8c}\right)\exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right)$$

$$\times x_3(n_1 - 1, n_2 - 1) + x_4(n_1, n_2) \quad \text{(A.4b)}$$

$$+ \exp\left(-\frac{i\sqrt{3}eH}{8c}\right)\exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right)x_4(n_1 - 1, n_2)$$

$$+ \exp\left(\frac{i\sqrt{3}eH}{8c}\right)\exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right)$$

$$\times x_4(n_1 - 1, n_2 - 1) = 0,$$

$$E_+^b x_3(n_1, n_2) + x_2(n_1, n_2) + \exp(2i\varphi)$$

$$\times \exp\left(\frac{i\sqrt{3}eH}{8c}\right)\exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right)x_2(n_1 + 1, n_2)$$

$$+ \exp(-2i\varphi)\exp\left(-\frac{i\sqrt{3}eH}{8c}\right)\exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right)$$

$$\times x_2(n_1 + 1, n_2 + 1) + x_1(n_1, n_2) \quad \text{(A.4c)}$$

$$+ \exp\left(\frac{i\sqrt{3}eH}{8c}\right)\exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right)x_1(n_1 + 1, n_2)$$

$$+ \exp\left(-\frac{i\sqrt{3}eH}{8c}\right)\exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right)$$

$$\times x_1(n_1 + 1, n_2 + 1) = 0,$$

$$E_-^b x_4(n_1, n_2) + x_2(n_1, n_2)$$

$$+ \exp\left(\frac{i\sqrt{3}eH}{8c}\right)\exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right)x_2(n_1 + 1, n_2)$$

$$+ \exp\left(-\frac{i\sqrt{3}eH}{8c}\right)\exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right)x_2(n_1 + 1, n_2 + 1)$$

$$+ x_1(n_1, n_2) + \exp(-2i\varphi) \quad \text{(A.4d)}$$

$$\times \exp\left(\frac{i\sqrt{3}eH}{8c}\right)\exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right)x_1(n_1 + 1, n_2)$$

$$+ \exp(2i\varphi)\exp\left(-\frac{i\sqrt{3}eH}{8c}\right)\exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right)$$

$$\times x_1(n_1 + 1, n_2 + 1) = 0.$$

In all these equations, we carry out the substitution

$$x_k(n_1, n_2) \longrightarrow \exp(i\alpha n_2 - i\alpha n_1/2)x_k(n_1, n_2). \quad \text{(A.5)}$$

As a result, we obtain a system of difference equations that do not contain a dependence on the second argument:

$$E_+^a x_1(n_1) + x_3(n_1) + \exp\left(-\frac{i\sqrt{3}eH}{8c}\right) \times \exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right)\exp\left(\frac{i\alpha}{2}\right)x_3(n_1-1) + \exp\left(\frac{i\sqrt{3}eH}{8c}\right) \times \exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right)\exp\left(-\frac{i\alpha}{2}\right)x_3(n_1-1) + x_4(n_1) + \exp(2i\varphi)\exp\left(-\frac{i\sqrt{3}eH}{8c}\right) \times \exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right)\exp\left(\frac{i\alpha}{2}\right)x_4(n_1-1) + \exp(-2i\varphi)\exp\left(\frac{i\sqrt{3}eH}{8c}\right)\exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right) \times \exp\left(-\frac{i\alpha}{2}\right)x_4(n_1-1) = 0, \quad \text{(A.6a)}$$

$$E_-^a x_2(n_1) + x_4(n_1) + \exp\left(-\frac{i\sqrt{3}eH}{8c}\right) \times \exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right)\exp\left(\frac{i\alpha}{2}\right)x_4(n_1-1) + \exp\left(\frac{i\sqrt{3}eH}{8c}\right) \times \exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right)\exp\left(-\frac{i\alpha}{2}\right)x_4(n_1-1) + x_3(n_1) + \exp(-2i\varphi)\exp\left(-\frac{i\sqrt{3}eH}{8c}\right) \times \exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right)\exp\left(\frac{i\alpha}{2}\right)x_3(n_1-1) + \exp(2i\varphi)\exp\left(\frac{i\sqrt{3}eH}{8c}\right)\exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right) \times \exp\left(-\frac{i\alpha}{2}\right)x_3(n_1-1) = 0, \quad \text{(A.6b)}$$

$$E_+^b x_3(n_1) + x_1(n_1) + \exp\left(\frac{i\sqrt{3}eH}{8c}\right) \times \exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right)\exp\left(\frac{i\alpha}{2}\right)x_1(n_1+1) + \exp\left(-\frac{i\sqrt{3}eH}{8c}\right)\exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right)\exp\left(-\frac{i\alpha}{2}\right)x_1(n_1+1) + x_2(n_1) + \exp(2i\varphi)\exp\left(\frac{i\sqrt{3}eH}{8c}\right) \times \exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right)\exp\left(\frac{i\alpha}{2}\right)x_2(n_1+1) + \exp(-2i\varphi)\exp\left(-\frac{i\sqrt{3}eH}{8c}\right)\exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right) \times \exp\left(-\frac{i\alpha}{2}\right)x_2(n_1+1) = 0, \quad \text{(A.6c)}$$

$$E_-^b x_4(n_1) + x_2(n_1) + \exp\left(\frac{i\sqrt{3}eH}{8c}\right) \times \exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right)\exp\left(-\frac{i\alpha}{2}\right)x_2(n_1+1) + \exp\left(-\frac{i\sqrt{3}eH}{8c}\right)\exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right)\exp\left(\frac{i\alpha}{2}\right)x_2(n_1+1) + x_1(n_1) + \exp(-2i\varphi)\exp\left(\frac{i\sqrt{3}eH}{8c}\right) \times \exp\left(-\frac{ieH\sqrt{3}n_1}{4c}\right)\exp\left(-\frac{i\alpha}{2}\right)x_1(n_1+1) + \exp(2i\varphi)\exp\left(-\frac{i\sqrt{3}eH}{8c}\right)\exp\left(\frac{ieH\sqrt{3}n_1}{4c}\right) \times \exp\left(\frac{i\alpha}{2}\right)x_1(n_1+1) = 0. \quad \text{(A.6d)}$$

We can write the equations derived above in the form used by Hofstadter. For this purpose, we carry out substitution $n_1 \longrightarrow n$ in Eqs. (A.6) and pass to real coefficients:

$$\begin{aligned}
&E_+^a x_1(n+1) + x_3(n+1) + x_4(n+1) \\
&+ 2x_3(n)\cos\theta_n + 2x_4(n)\cos\theta_n^+ = 0, \\
&E_-^b x_2(n+1) + x_3(n+1) + x_4(n+1) \\
&+ 2x_4(n)\cos\theta_n + 2x_3(n)\cos\theta_n^- = 0, \\
&E_+^a x_3(n) + x_1(n) + x_2(n) \\
&+ 2x_1(n+1)\cos\theta_n + 2x_2(n+1)\cos\theta_n^- = 0, \\
&E_-^b x_4(n) + x_1(n) + x_2(n) \\
&+ 2x_2(n+1)\cos\theta_n + 2x_1(n+1)\cos\theta_n^+ = 0.
\end{aligned} \quad \text{(A.7)}$$

REFERENCES

1. P. G. Harper, Proc. Phys. Soc. London, Sect. A **68**, 874 (1955).
2. D. R. Hofstadter, Phys. Rev. B **14**, 2239 (1976).
3. V. Ya. Demikhovskiĭ and A. A. Perov, Zh. Éksp. Teor. Fiz. **114**, 1795 (1998) [JETP **87**, 973 (1998)].
4. V. Ya. Demikhovskiĭ and D. V. Khomitskiĭ, Zh. Éksp. Teor. Fiz. **120**, 191 (2001) [JETP **93**, 168 (2001)].
5. U. Kuhl and H.-J. U. Stockman, Phys. Rev. Lett. **80**, 3232 (1998).
6. T. Schlosser, K. Ensslin, J. P. Kotthaus, and M. Holland, Europhys. Lett. **33**, 683 (1996).
7. R. R. Gerhardts, D. Weiss, and U. Wulf, Phys. Rev. B **43**, 5192 (1991).
8. D. Weiss, M. L. Roukes, A. Menshing, *et al.*, Phys. Rev. Lett. **66**, 2790 (1991).
9. J. Zak, Phys. Rev. A **136**, A776 (1964).
10. I. E. Dzaloschinski, Phys. Rev. B **65**, 205325 (2002).
11. G. Wentzel, Phys. Rev. **108**, 1593 (1957).
12. R. O. Zaĭtsev, Zh. Éksp. Teor. Fiz. **112**, 2223 (1997) [JETP **85**, 1218 (1997)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 97, No. 2, 2003, pp. 305–316.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 2, 2003, pp. 338–350.

SOLIDS
Electronic Properties

Dynamic Phenomena Connected with Magnetic and Antiferroelectric Interactions in Trirutiles

I. F. Mirsaev* and E. A. Turov**

Institute of Metal Physics, Ural Division, Russian Academy of Sciences, ul. S. Kovalevskoĭ 18, Yekaterinburg, 620219 Russia

** e-mail: mirsaev@imp.uran.ru*

*** e-mail: turov@imp.uran.ru*

Received December 23, 2002

Abstract—Dynamic effects caused by the magnetoelectric and antiferroelectric interactions in tetragonal antiferromagnets are studied. The analysis is based on the example of trirutiles that are a series of antiferromagnets with different exchange structures and orientation states. We are mainly dealing with the excitation by an alternating electric field $\mathbf{E}(t)$ of spin waves typical of these magnets (antiferroelectric resonance) and the nuclear magnetoelectric resonance connected with these interactions. In the first case, special emphasis is placed on specific magnons (antimagnons), where only the antiferromagnetism vectors **L** take part in oscillations, whereas the total ferromagnetism vector **M** remains unchanged. The nuclear magnetoelectric resonance can be generated by oscillations of both **L** and **M** caused by field $\mathbf{E}(t)$. In this way, the field contributes to the hyperfine field, which acts on the nuclear spins. It is shown that the magnetic and antiferroelectric interactions in the dynamics can manifest themselves both at high (usually, exchange) frequencies ω_E (antiferroelectric resonance) and at rather low nuclear frequencies of $\omega_n \ll \omega_E$. Particular cases of magnetic structures (phases) are considered where field $\mathbf{E}(t)$ can excite not only antimagnons, but also quasiantiferromagnons that have lower eigenfrequencies than those of quasimagnons (relativistic and semirelativistic). © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In certain magnetically ordered materials (magnets), magnetoelectric and/or antiferroelectric interactions of magnetic moments (sublattice magnetizations) with electric field **E** occur. These interactions are connected with the presence of terms of the form

$$M_i L_j E_k \tag{1}$$

in thermodynamic potential Φ (its density in the homogeneous case) in the case of the magnetoelectric interaction and

$$L_{1i} L_{2j} E_k \tag{2}$$

in the case of the antiferromagnetic interaction. In formulas (1) and (2), **M** is the total magnetization vector and **L** and $\mathbf{L}_{1,2}$ are the antiferromagnetism vectors. In centrosymmetric crystals whose space group contains an inversion $\bar{1}$, invariants of form (1) exist in the case where vector **L** is centroantisymmetric. Hence, $\bar{1}\mathbf{L} = -\mathbf{L}$, since **M** is a centrosymmetric vector ($\bar{1}\mathbf{M} = \mathbf{M}$) and **E** is a centroantisymmetric vector. Invariants of form (2) exist if one of the antiferromagnetic vectors (say, $\mathbf{L}_1$) is centrosymmetric and the other one ($\mathbf{L}_2$) is centroantisymmetric.

In statics, a term of form (1) generates a known magnetoelectric effect, namely, the magnetization of the antiferromagnet by an external electric field $\mathbf{E} = \mathbf{E}_0 = \text{const}$ [1] and, due to contribution (2), field $\mathbf{E}_0$ can generate magnetic orientational (or even structural) phase transitions. The influence of constant electric field $\mathbf{E}_0$ on the antiferromagnetic resonance frequency for trirutiles and Cr_2O_3 in a two-sublattice approximation is considered in [1]. This allows us to apply a model where $\mathbf{L}^2 = \text{const}$ [1], and it is assumed that the magnetic susceptibility parallel to vector **L** is nonzero ($\chi_\parallel \neq 0$). However, in this case, the antiferroelectric interaction caused by the presence of four sublattices is not taken into account.

In dynamics, the magnetoelectric and antiferroelectric interactions exhibit themselves in a significantly greater variety of phenomena, in particular, for alternating fields $\mathbf{E}(t)$ and multisublattice ($n > 2$) magnets. First, these interactions are associated with the possibility of excitation by an alternating electric field $\mathbf{E}(t)$ of magnons of different types (including such that cannot be excited by magnetic field $\mathbf{H}(t)$), since only an antiferromagnetism vector (or vectors) takes part in magnon oscillations and oscillations of the total local ferromagnetism vector **M** are absent. The existence of such magnons was predicted for the first time in [2–6]. In these papers, antiferromagnetic crystals with at least

four sublattices (rhombohedral α-Fe_2O_3 and Cr_2O_3, rhombic orthoferrites, and hexagonal $CsMnF_3$) are considered. The authors of [2–6] called such magnons electrically active. It is typical of all these antiferromagnets that magnetic ions present in these materials are in noncentrosymmertical positions; hence, the symmetry center element $\bar{1}$ existing in these antiferromagnets permutes such ions with each other. Unfortunately, the society of magnetologists paid very little attention to these fundamental studies that opened a new interesting field in spin dynamics. Only in 2001 did one of the authors of [7] draw attention to the existence of purely antiferromagnetic magnons excited by field $\mathbf{E}(t)$ in a two-sublattice ferromagnet. He emphasized the possibility of the existence of electrically active magnons and named them antimagnons. A ferromagnet (not an antiferromagnet as in [2–6]) is considered in [7], and it is the ferromagnetic structure with two magnetic ions in a noncentrosymmetric position, but linked through a symmetry center, that is the simplest system in which antimagnons exist. The authors of papers [8–11] developed this area of research. They considered the totality of phenomena determined by the dynamic manifestation of the magnetoelectric and antiferroelectric interactions in multilattice ferro-, antiferro-, and ferrimagnets. As well as in all papers cited above, eigenfrequency ω_L of antimagnons possesses an exchange nature, and the antimagnons can be excited not only by field $\mathbf{E}(t)$, but also by magnetic field $\mathbf{H}(t)$ (at the same frequency) if a constant electric field $\mathbf{E} = \mathbf{E}_0 = \text{const}$ is applied to the sample. Moreover, some magnetoelectric and antiferroelectric phenomena in the low frequency region, where, in particular, $\omega \approx \omega_n \ll \omega_E$ (ω_n is the NMR frequency), were predicted. In other words, the nuclear magnetoelectric resonance generated by field $\mathbf{E}(t)$ at frequency ω_n was predicted [12]. Leskovets *et al.* [13] investigated the high-frequency properties of an antiferromagnetic $KNiPO_4$ crystal, which has no symmetry center, in constant and alternating fields $\mathbf{H}$ and $\mathbf{E}$.

In this paper, the phenomena mentioned above are considered in tetragonal antiferromagnets with the trirutile structure (they have not been investigated in [2–6]) from the points of view of both antimagnons and nuclear magnetoelectric resonance. Much attention has also been also given to the study of magnetic structures (phases) existing in trirutiles, in which field $\mathbf{E}(t)$ is excited by other (different from antimagnon) spin waves (quasiantiferromagnons). Being relativistic and semirelativistic, these structures have rather low eigenfrequencies.

We chose trirutiles as a concrete object for the investigation, because these materials represent a rather large class of centrosymmetric antiferromagnetic compounds with various exchange structures and orientation states [1]. The results allow us to discuss the known ones and to recommend new ones. Note that some results for tetragonal crystals are presented in [12, 14].

2. CRYSTAL AND MAGNETIC STRUCTURES: THERMODYNAMIC POTENTIAL

The following antiferromagnetic compounds are tetragonal crystals with the trirutile structure: Fe_2TeO_6 (T_N = 219 K), Cr_2TeO_6 (T_N = 105 K), Cr_2WO_6 (T_N = 69 K), V_2WO_6 (T_N = 370 K), etc. [1]. The magnetic atoms in these crystals are in the fourfold position $4e\{mm\}$ of the group $P4_2/mnm$ (D_{4h}^{14}): 1(0, 0, z), 2(0, 0, $1 - z$), 3(1/2, 1/2, 1/2 + z), 4(1/2, 1/2, 1/2 – z). In the braces, the island (local) symmetry of the position is indicated. The four magnetic sublattices with magnetizations M_ν (ν = 1, 2, 3, 4) are associated with four (one ferromagnetic and three antiferromagnetic) base vectors:

$$\begin{aligned}
\mathbf{M} &= \mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3 + \mathbf{M}_4,\\
\mathbf{L}_a &= \mathbf{M}_1 + \mathbf{M}_2 - \mathbf{M}_3 - \mathbf{M}_4,\\
\mathbf{L}_b &= \mathbf{M}_1 - \mathbf{M}_2 + \mathbf{M}_3 - \mathbf{M}_4,\\
\mathbf{L}_c &= \mathbf{M}_1 - \mathbf{M}_2 - \mathbf{M}_3 + \mathbf{M}_4.
\end{aligned} \tag{3}$$

The first two vectors are centrosymmetric, and the other vectors are centroantisymmetric.

We take the elements $\bar{1}$, 2_{1x}, and 4_{2z} as generators of the group. Then, we write the transformation table for the components of the base vectors. We take into account not only rotations (and reflections), but also permutations of magnetic ions under the action of these elements [1, 14]. (Since $2_{1x} = 4_{2z} \times 2_d$, we can use for this purpose a table similar to Table 4.4 from [1], which determines the transformation rules for $\mathbf{M}$ and $\mathbf{L}_\xi$ (3) ($\xi = a, b, c$) with respect to elements $\bar{1}$, 4_{2z}, and 2_d of the group $P4_2/mnm$.)

First, we compose such a table for the rhombic group $Pnmm$, which is a subgroup of the true tetragonal group $P4_2/mnm$ since $4_z^2 = 2_z$. The first column of the table enumerates phases Γ_n of group $Pnmm$ in terms of components of vectors $\mathbf{M}$ and $\mathbf{L}$. The numbers +1 or –1 in the third column show whether the corresponding functions in the second column change sign under the action of elements $\bar{1}$, 2_{1x}, and 2_z, which are generators of group $Pnmm$. The magnetic point groups, with respect to which the functions (in the second column) determining the corresponding phases for group $Pnmm$ are invariant, are written in the fourth column. Here, $g' = g \times 1'$ (where g is an element of the point group and 1' is the time reversion operation). In the last column, the results of the transformation under the action of element 4_{2z}, which complements group $Pnmm$ with respect to $P4_2/mnm$, are presented.

From the viewpoint of the rhombic group, each row of the table corresponds to a certain magnetic structure (phase), in which only the components presented in this row are nonzero in the ground state. After adding the 4_{2z}

axis, these phases can be united into one phase, which includes the components associated with this axis. (For example, phases Γ_4 and Γ_5 can compose the phase $\Gamma_4 + \Gamma_5$, in which $\mathbf{L}_b^0 \parallel [110]$, $\mathbf{L}_c = 0$ or $\mathbf{L}_c^0 \parallel [110]$, $\mathbf{L}_b^0 = 0$ [1]. Here and below, the superscripted zero denotes the ground state.) However, these phases can exist separately. This is determined by the concrete type of the magnetic anisotropy energy and by the magnitude and direction of the magnetic field. Considering only the phases typical of the rhombic symmetry, we come, due to the presence of the 4_{2z} axis, to certain relations between the constants of the thermodynamic potential and the equality of the corresponding magnon frequencies for the two phases connected by this axis.

Using the table, we write an invariant expression for the thermodynamic potential density Φ including the magnetoelectric and antiferroelectric interactions, as well as the Zeeman energy in a magnetic field. In the bilinear approximation in vectors (3), we have

$$\Phi = \Phi_{mag} + \Phi_{ME} + \Phi_{AFE}. \tag{4}$$

Here,

$$\begin{aligned}\Phi_{mag} &= \frac{1}{2}A_M\mathbf{M}^2 + \frac{1}{2}\sum_\xi A_\xi \mathbf{L}_\xi^2 \\ &+ \frac{1}{2}K_M(M_x^2 + M_y^2) + \frac{1}{2}\sum_\xi K_\xi(L_{\xi x}^2 + L_{\xi y}^2) \\ &+ r(L_{bx}L_{cy} + L_{by}L_{cx}) + p(M_xL_{ay} + M_yL_{ax}) - \mathbf{M}\cdot\mathbf{H}\end{aligned} \tag{5}$$

is the magnetic energy, A_M and A_ξ ($\xi = a, b, c$) are exchange constants; the other constants are relativistic. Then,

$$\begin{aligned}4M_0\Phi_{ME} &= -s_1(M_xE_x + M_yE_y)L_{bz} \\ &- s_2(M_xL_{bx} + M_yL_{by})E_z - s_3M_z(E_xL_{bx} + E_yL_{by}) \\ &- s_4M_zE_zL_{bz} - d_1(M_xE_y + M_yE_x)L_{cz} \\ &- d_2(M_xL_{cy} + M_yL_{cx})E_z - d_3M_z(E_xL_{cy} + E_yL_{cx})\end{aligned} \tag{6}$$

defines the magnetoelectric interaction,

$$\begin{aligned}4M_0\Phi_{AFE} &= -f_1L_{az}(E_xL_{by} + E_yL_{bx}) \\ &- f_2L_{az}(E_xL_{cx} + E_yL_{cy}) - f_3(L_{ax}E_y + L_{ay}E_x)L_{bz} \\ &- f_4(L_{ax}E_x + L_{ay}E_y)L_{cz} - f_5L_{az}L_{cz}E_z\end{aligned} \tag{7}$$

defines the antiferroelectric interaction, and M_0 is the magnitude of the sublattice magnetization vector. Note that, for particular magnetic structures, only some terms remain in general expressions (6) and (7).

Transformation of vectors $\mathbf{M}$, $\mathbf{L}_\xi$ ($\xi = a, b, c$), and $\mathbf{E}$ under the action of the symmetry elements of the *Pnmm* and $P4_2/mnm$ groups

Γ_i	$\mathbf{M}, \mathbf{L}, \mathbf{E}$	$\bar{1}\,2_{1x}2_z$	Magnetic group	4_{2z}
Γ_1	M_x, L_{ay}	+1 + 1 –1	$\bar{1}\,2_x2'_z$	M_y, L_{ax}
Γ_2	M_y, L_{ax}	+1 – 1 –1	$\bar{1}\,2'_x\,2'_z$	$-M_x, -L_{ay}$
Γ_3	M_z	+1 – 1 +1	$\bar{1}2'_x4_z$	M_z
Γ_4	L_{bx}, L_{cy}, E_y	–1 – 1 –1	$\bar{1}'2'_x\,2'_z$	$L_{by}, L_{cx}, -E_x$
Γ_5	L_{by}, L_{cx}, E_x	–1 + 1 –1	$\bar{1}'2_x2'_z$	$-L_{bx}, -L_{cy}, E_y$
Γ_6	L_{az}	+1 + 1 +1	$\bar{1}\,2_x4'_z$	$-L_{az}$
Γ_7	L_{bz}	–1 + 1 +1	$\bar{1}'2_x4'_z$	L_{bz}
Γ_8	L_{cz}, E_z	–1 – 1 +1	$\bar{1}'2'_x\,4'_z$	$-L_{cz}, E_z$

We consider the equimodule model $\mathbf{M}_v^2 = M_0^2$ that corresponds to the conditions

$$\begin{aligned}\mathbf{M}^2 + \sum_\xi \mathbf{L}_\xi^2 &= (4M_0)^2, \\ \mathbf{M}\cdot\mathbf{L}_a + \mathbf{L}_b\cdot\mathbf{L}_c &= 0, \\ \mathbf{M}\cdot\mathbf{L}_b + \mathbf{L}_a\cdot\mathbf{L}_c &= 0, \\ \mathbf{M}\cdot\mathbf{L}_c + \mathbf{L}_a\cdot\mathbf{L}_b &= 0,\end{aligned} \tag{8}$$

as well as to the Landau–Lifshitz equations. For four sublattices, these equations have the form [14]

$$\begin{aligned}\dot{\mathbf{M}} &= \gamma\left(\mathbf{M}\times\frac{\partial\Phi}{\partial\mathbf{M}} + \mathbf{L}_a\times\frac{\partial\Phi}{\partial\mathbf{L}_a} + \mathbf{L}_b\times\frac{\partial\Phi}{\partial\mathbf{L}_b} + \mathbf{L}_c\times\frac{\partial\Phi}{\partial\mathbf{L}_c}\right), \\ \dot{\mathbf{L}}_a &= \gamma\left(\mathbf{M}\times\frac{\partial\Phi}{\partial\mathbf{L}_a} + \mathbf{L}_a\times\frac{\partial\Phi}{\partial\mathbf{M}} + \mathbf{L}_b\times\frac{\partial\Phi}{\partial\mathbf{L}_c} + \mathbf{L}_c\times\frac{\partial\Phi}{\partial\mathbf{L}_b}\right), \\ \dot{\mathbf{L}}_b &= \gamma\left(\mathbf{M}\times\frac{\partial\Phi}{\partial\mathbf{L}_b} + \mathbf{L}_a\times\frac{\partial\Phi}{\partial\mathbf{L}_c} + \mathbf{L}_b\times\frac{\partial\Phi}{\partial\mathbf{M}} + \mathbf{L}_c\times\frac{\partial\Phi}{\partial\mathbf{L}_a}\right), \\ \dot{\mathbf{L}}_c &= \gamma\left(\mathbf{M}\times\frac{\partial\Phi}{\partial\mathbf{L}_c} + \mathbf{L}_a\times\frac{\partial\Phi}{\partial\mathbf{L}_b} + \mathbf{L}_b\times\frac{\partial\Phi}{\partial\mathbf{L}_a} + \mathbf{L}_c\times\frac{\partial\Phi}{\partial\mathbf{M}}\right).\end{aligned} \tag{9}$$

Here, γ is the magnitude of the magnetomechanical ratio.

Note that, in this approach, the oscillations of the spin system are considered independently of the lattice vibrations excited by field $\mathbf{E}(t)$, as well as oscillations of polarization $\mathbf{P}$ connected with these vibrations. Actually, we make the assumption (which is justified in

most cases) that the eigenfrequencies of the latter oscillations are high as compared with spin frequencies interesting for us. This allows us to immediately write thermodynamic potential (4) in terms of electric field $\mathbf{E}(t)$ assuming that $\mathbf{P}$ follows $\mathbf{E}$ in an equilibrium way. The situation can be more complicated for the case of nonuniform spin waves if we take into account the interaction of these waves with acoustic vibrations of the lattice.

3. SPIN-WAVE REPRESENTATIONS (VIBRATIONAL MODES)

Passing to vibrational modes and their eigenfrequencies, we note that we use here the concept of spin-wave representations [1, 15–17] that allows us to divide the vibrational variables into independent groups before writing the equations of motion. According to the algorithm formulated in [1, 15] for a rhombic crystal, the variables from two rows Γ_m and $\Gamma_{m'}$ of the table for which the products of the numbers (+1 or –1) give the numbers of row Γ_n are related to the same spin-wave representation of phase Γ_n under consideration. Schematically, we write $\Gamma_m \cdot \Gamma_{m'} = \Gamma_n$. It can occur that $m = m'$.

Unfortunately, we cannot hitherto formulate an equally simple general rule for determining the spin-wave representations for tetragonal crystals as has been done for rhombic crystals. Here, the general techniques of the corepresentation theory of space groups developed in [16, 17] could be applied. However, the spin-wave representations obtained for an appropriate rhombic group give us sufficient information for obtaining the representations for the tetragonal case using some additional considerations connected with the symmetry and the particular form of the thermodynamic potential. This should be done separately for each magnetic structure considered. This way is simpler and more popular than the use of the general theory [16, 17]. Below, we demonstrate this assertion using concrete examples. The magnetic groups presented in the table allow us to justify the choice of the spin-wave representations by verifying the invariance of the corresponding equations of motion. (We should keep in mind that the elements of the magnetic group act only on the dynamic (oscillation) variables.)

4. EASY-AXIS STRUCTURE. PHASE $\Gamma_7(L^0_{bz})$ (MAGNETIC GROUP $\bar{1}'2_x4_z$)

We studied earlier this antiferromagnetic phase with centroantisymmetric primitive vector $\mathbf{L}^0_b \parallel z$ ($L^0_{bz} = 4M_0$) characteristic of the Fe_2TeO_6 compound [14]. Now, we present the results in more detail.

According to the table, there are two transverse quasiantiferromagnetic representations for this phase, namely,

$$\Gamma_{15}(M_x, L_{ay}, L_{by}, L_{cx}) \tag{10}$$

and

$$\Gamma_{24}(M_y, L_{ax}, L_{bx}, L_{cy}), \tag{11}$$

which are the same as in the rhombic case, but remain independent when the 4_{2z} axis is taken into account. However, phases Γ_{15} and Γ_{24} are equivalent from the viewpoint of symmetry: namely, one of them can be obtained from the other via rotation of the 4_{2z} axis. Therefore, it is sufficient to solve the formulated problem for one of these phases. Let it be Γ_{15} (10).

From potential Φ specified by expressions (5)–(7), we separate the quadratic form Φ_2 in appropriate variables (10) (taking into account equimodule model (8)):

$$\begin{aligned}\Phi_2 = \frac{1}{2}(\tilde{A}_M M_x^2 + \tilde{A}_a L_{ay}^2 + \tilde{A}_c L_{cx}^2 + K_b L_{by}^2) \\ + pM_x L_{ay} + rL_{by}L_{cx} \\ - (s_1 M_x + f_3 L_{ay})E_x(t) - M_x H_x(t).\end{aligned} \tag{12}$$

Here, in addition to magnetoelectric (s_1) and antiferroelectric (f_3) interactions, we take into account the Zeeman interaction in the alternating field $\mathbf{H} \parallel x$. According to (10), this field (just as field $\mathbf{E} \parallel x$) must excite mode Γ_{15}.[1] In formula (12), we use the notation $\tilde{A}_\kappa = A_\kappa - A_b + K_\kappa$, where $A_\kappa > 0$, $A_b < 0$, $K_b > 0$, $K_\kappa < 0$, and $\kappa = M, a, c$.

Taking into account formula (12), from the corresponding Landau–Lifshitz equations (9), we obtain the following system (for $E_x, H_x \propto \exp(-i\omega t)$):

$$\begin{aligned}i\omega M_x &= \omega_0(K_b L_{by} + rL_{cx}), \\ i\omega L_{by} &= -\omega_0(\tilde{A}_M M_x + pL_{ay} - s_1 E_x - H_x), \\ i\omega L_{ay} &= -\omega_0(\tilde{A}_c L_{cx} + rL_{by}), \\ i\omega L_{cx} &= \omega_0(\tilde{A}_a L_{ay} + pM_x - f_3 E_x),\end{aligned} \tag{13}$$

where $\omega_0 = 4\gamma M_0$.

For $p = r = 0$, the first and the second pairs of Eqs. (13) split into two independent pairs. Moreover, the second pair with variables L_{ay} and L_{cx} excited by

[1] The constant magnetic field $\mathbf{H}_0$ should be taken into account (as it has been done in the two-sublattice approximation in [1]). However, in our case, this results in a significant complication of the problem and is hardly justified. The matter is that the considered phase $\Gamma_7(L^0_{bz})$ does not contain vector $\mathbf{M}$ that is generated by field $\mathbf{H}_0$. However, this is another phase and, in general, one should seek other modes instead of (10) and (11).

field E_x produces an exchange, purely antiferromagnetic mode (by definition, antimagnonic). However, in this case, we are more interested in the other pair of equations (13) with variables M_x and L_{by}. For this pair, we can also take into account the relativistic terms with $p \neq 0$ and $r \neq 0$. However, we should bear in mind that the corresponding variables L_{ay} and L_{cx} follow M_x and L_{by} in a quasiequilibrium way. Therefore, the first variables can be expressed via the second variables by minimizing Φ_2 (12) in L_{ay} and L_{cx} ($\partial\Phi_2/\partial L_{ay} = \partial\Phi_2/\partial L_{cx} = 0$). This means that we should set $\omega = 0$ in the second pair of equations (13). We obtain

$$L_{ay} = -\frac{p}{\tilde{A}_a}M_x + \frac{f_3}{\tilde{A}_a}E_x, \quad L_{cx} = -\frac{r}{\tilde{A}_c}L_{by}. \tag{14}$$

Substituting Eqs. (14) in the first pair of equations (13), we find a solution for the quasiantiferromagnetic mode

$$M_x = (\alpha_0 E_x + \chi_0 H_x)\frac{\omega_{AF}^2}{\omega_{AF}^2 - \omega^2},$$
$$L_{by} = \frac{i\omega}{\omega_0 K_b^*}M_x, \tag{15}$$

where $\chi_0 = 1/A_M^*$ and $\alpha_0 = s_1^*/A_M^*$ are the static magnetic and magnetoelectric susceptibilities, respectively, with the resonance frequency

$$\omega_{AF} = \omega_0\sqrt{A_M^* K_b^*}, \tag{16}$$

and

$$A_M^* = \tilde{A}_M - \frac{p^2}{\tilde{A}_a}, \quad K_b^* = K_b - \frac{r^2}{\tilde{A}_c},$$
$$s_1^* = s_1 - \frac{f_3 p}{\tilde{A}_a}$$

is the relativistic renormalization of constants.

As for the exchange mode arising due to the second pair of equations (13), the influence of terms with p and r on the mode frequency ω_E is insignificant, because for

$$\omega_E = \omega_0\sqrt{\tilde{A}_a\tilde{A}_c} \gg \omega_{AF}, \tag{17}$$

quantities M_x and L_{by} remain almost in equilibrium, i.e., close to zero. The optical mode is actually antimagnonic (the antimagnon is electically active):

$$L_{ay} = \frac{i\omega_0\tilde{A}_c}{\omega}L_{cx} = \frac{f_3}{\tilde{A}_a}\frac{\omega_E^2}{\omega_E^2 - \omega^2}E_x. \tag{18}$$

This mode is excited by field $\mathbf{E} \parallel x$. Moreover, in contrast to quasiantiferromagnetic mode (15) caused by the magnetoelectric and Zeeman interactions, this mode is caused by the antiferroelectric interaction.

Thus, spin-wave representation Γ_{15} (10) produces two modes, namely, an electrically active antimagnonic mode with variables (18) and frequency ω_E (17) and a quasiantiferromagnetic mode, which is simultaneously magnetically and electrically active with resonance frequency ω_{AF} (16).

The corresponding results for spin representation Γ_{24} (11) can be obtained from (12)–(18) via the substitution $x \longleftrightarrow y$ and $\omega \longrightarrow -\omega$.

The antiferromagnetic resonance should be observable at the antinode of field $H_x(t)$, and the electrically active (antiferroelectric) resonance should be observable at the antinode of field $E_x(t)$; we can assume that $|H_x| \approx |E_x|$ (in the CGS system). Therefore, according to relations (15), the magnetization ratio for these two resonances is determined via the ratio of the static magnetoelectric and magnetic susceptibilities:

$$\left|\frac{M_x^E}{M_x^H}\right| \approx \frac{\alpha_0}{\chi_0}. \tag{19}$$

These susceptibilities in some antiferromagnets can be comparable in magnitude. Then, one can hope that, along with the antiferromagnetic resonance, the antiferroelectric resonance can be observed in such antiferromagnets. It is important that, unlike the exchange antimagnonic resonance, frequency ω_{AF} can lie in the microwave region. Nevertheless, for tetragonal antiferromagnets, the antiferroelectric resonance should be experimentally sought in Fe_2TeO_6 and in rare-earth phosphates and vanadates (compounds of the $TbPO_4$ and $GdVO_4$ types) [1].

5. EASY-PLANE STATE

5.1. Centroantisymmetric Phases $\Gamma_4(L_{bx}^0, L_{cy}^0)$ *and* $\Gamma_5(L_{by}^0, L_{cx}^0)$

The easy-plane trirutiles Cr_2TeO_6, Cr_2WO_6, and V_2WO_6 belong to centroantisymmetric structures with vectors $\mathbf{L}_b^0$ and $\mathbf{L}_c^0$ in the ground state. In the rhombic case, they include antiferromagnetic phases $\Gamma_4(L_{bx}^0, L_{cy}^0)$ and $\Gamma_5(L_{by}^0, L_{cx}^0)$. In a tetragonal crystal, these phase are energetically equivalent (one of them can be obtained from the other by the action of element 4_{2z}). Hence, they can be united into one phase $\Gamma_4 + \Gamma_5$ with magnetic group $\bar{1}'2_z'$.

Hitherto, the experiment gives no information on the orientation of $\mathbf{L}_b^0$ and $\mathbf{L}_c^0$ in the basal plane (see the discussion in [1], Section 7.2). However, the study of thermodynamic potential Φ_{mag} (5) for the tetragonal phase $\Gamma_4 + \Gamma_5$, to which, generally speaking, biquadratic

anisotropic terms of type $L_{bx}^2 L_{by}^2$, etc., should be added, shows (at least for the equimodule model) that the minimum of Φ_{mag} corresponds to the same phases $\Gamma_4(L_{bx}^0, L_{cy}^0)$ and $\Gamma_5(L_{by}^0, L_{cx}^0)$ as in the rhombic case and the diagonal phases with $\mathbf{L}_b^0 \parallel [110]$, $\mathbf{L}_c^0 = 0$ or $\mathbf{L}_c^0 \parallel [110]$, $\mathbf{L}_b^0 = 0$ [1]. The table gives four modes for each of the first two phases, namely, these are antimagnonic mode $\Gamma_{46}(L_{az}, \Delta L_{bx}, \Delta L_{cy})$ and quasiantiferromagnetic modes $\Gamma_{35}(M_z, L_{by}, L_{cx})$, $\Gamma_{18}(M_x, L_{ay}, L_{cz})$, and $\Gamma_{27}(M_y, L_{ax}, L_{bz})$ in phase $\Gamma_4(L_{bx}^0, L_{cy}^0)$; and analogous modes $\Gamma_{56}(L_{az}, \Delta L_{by}, \Delta L_{cx})$, $\Gamma_{34}(M_z, L_{bx}, L_{cy})$, $\Gamma_{28}(M_y, L_{ax}, L_{cz})$, and $\Gamma_{17}(M_x, L_{ay}, L_{bz})$ in phase $\Gamma_5(L_{by}^0, L_{cx}^0)$. Here, Δ denotes vibrational variables.

The aforesaid assertions can be verified on the basis of the invariance condition for the equations of motion with respect to magnetic groups $\bar{1}'2_x'2_z'$ and $\bar{1}'2_x2_z'$ for Γ_4 and Γ_5, respectively.

Here, two cases can occur, where the primitive base vectors are (1) $\mathbf{L}_b^0$ ($L_b^0 \gg L_c^0$) and (2) $\mathbf{L}_c^0$ ($L_c^0 \gg L_b^0$). Case (1) corresponds to Cr_2TeO_6; case (2), to Cr_2WO_6 and V_2WO_6 (see Table 7.1 in [1]). In the ground state, the noncollinearity is caused by the relativistic interaction (the term with r in (5)). Due to this interaction, the magnetic structure with vectors $\mathbf{L}_b^0$ and $\mathbf{L}_c^0$ should have the form of a plane cross extended in the direction of the primitive base vector (see Fig. 6.21 in [1]).

5.2. Magnetic Structure with Vector $\mathbf{L}_b^0$ (Cr_2TeO_6)

1. Weakly collinear phase $\Gamma_4(L_{bx}^0, L_{cy}^0)$. Consider a magnetic structure with primitive base vector $\mathbf{L}_b^0 \parallel x$ ($L_{bx}^0 \approx 4M_0$) corresponding to compound Cr_2TeO_6. This orientation state is associated with weakly noncollinear phase $\Gamma_4(L_{bx}^0, L_{cy}^0)$, where $L_{bx}^0 \gg L_{cy}^0$. The requirement of the energy minimum for ground state Φ_{mag}^0 of phase Γ_4 implies that

$$\lambda \equiv \frac{L_{cy}^0}{L_{bx}^0} \approx -\frac{r}{\tilde{A}_c} \ll 1, \tag{20}$$

where $\tilde{A}_c = A_c - A_b + K_c - K_b$; $A_b, K_b < 0$, and $A_c, K_c > 0$.

Using relations (8) and (20), as well as the condition of energy minimum $\partial\Phi_{mag}^0/\partial L_{bx}^0 = 0$, in thermodynamic potential Φ (4), we separate quadratic form Φ_2 in the spin-wave variables (transversal with respect to $\mathbf{L}_b^0 \parallel x$) corresponding to the above-listed modes of phase Γ_4. As a result, we have

$$\begin{aligned}\Phi_2 = {} & \frac{1}{2}\tilde{A}_M M_y^2 + \frac{1}{2}(\tilde{A}_M - K_M)M_z^2 + \frac{1}{2}\tilde{A}_a L_{ay}^2 \\ & + \frac{1}{2}(\tilde{A}_a - K_a)L_{az}^2 + 2\frac{r^2}{\tilde{A}_c}L_{by}^2 - \frac{1}{2}K_b L_{bz}^2 \\ & + \frac{1}{2}\tilde{A}_c \Delta L_{cy}^2 + \frac{1}{2}(\tilde{A}_c - K_c)L_{cz}^2 \\ & + \lambda s_2 L_{ay}E_z - s_3 M_z E_x - f_1 L_{az}E_y,\end{aligned} \tag{21}$$

where $\tilde{A}_\kappa = A_\kappa - A_b + K_\kappa - K_b$ and $A_\kappa, K_\kappa > 0$ ($\kappa = M, a, c$). In writing relation (21), we did not take into account an insignificant renormalization (of order λr and λp) of exchange constants A_κ and anisotropy K_κ due to the relativistic interactions with r and p in formula (5).

First, consider the antimagnonic mode $\Gamma_{46}(L_{az}, \Delta L_{bx}, \Delta L_{cy})$. By virtue of relation (21), it can be excited by electric field $\mathbf{E}(t) \parallel \mathbf{L}_c^0 \parallel y$ due to the antiferroelectric interaction (the term containing f_1 in formula (21)). Taking into account formula (21) and the relation $\mathbf{E}(t) \propto \exp(-i\omega t)$, we can obtain for this mode from the Landau–Lifshitz equation (9)

$$\begin{aligned}L_{az} = \beta_{zy}E_y = \frac{f_1}{\tilde{A}_a - K_a}\frac{\omega_{Lb}^2}{\omega_{Lb}^2 - \omega^2}E_y, \\ \Delta L_{cy} = -\frac{i\omega}{\omega_0 \tilde{A}_c}L_{az}, \quad \Delta L_{bx} = \frac{r}{\tilde{A}_c}\Delta L_{cy}.\end{aligned} \tag{22}$$

Here, β_{zy} is the component of the antiferroelectric susceptibility tensor and ω_{Lb} is the frequency of the antimagnonic resonance, which is defined as

$$\omega_{Lb}^2 = \omega_0^2 \tilde{A}_c(\tilde{A}_c - K_a), \tag{23}$$

where $\omega_0 = \gamma L_{bx}^0 \approx 4\gamma M_0$. Here, in contrast with [14], we take into account the relativistic interactions. One can see from formulas (22) and (23) that, taking these interactions into account, essentially we change nothing; we only add some minor corrections to exchange frequency ω_{Lb} (23). However, we shall show below that these interactions play an important role for quasiantiferromagnetic modes $\Gamma_{35}(M_z, L_{by}, L_{cx})$ and $\Gamma_{27}(M_y, L_{ax}, L_{bz})$. Now, we determine the heat loss Q connected with the excitation of antimagnons by electric field $\mathbf{E}(t) =$

$\mathbf{E}_0 \exp(-i\omega t)$. Representing the antiferroelectric interaction for mode Γ_{46} in the form

$$\Phi_{AFE} = -P_y E_y, \quad P_y = f_1 L_{az}, \tag{24}$$

we see that [12]

$$Q = -\overline{P_y \dot{E}_y}, \tag{25}$$

where the bar above the symbols denotes the averaging over time $t \gg 2\pi/\omega$. Here, P_y is the component of the effective polarization vector connected with oscillations of L_{az} (22) (the last term in formula (21)). Introducing the dissipation (for instance, according to Bloch, by using the substitution $\omega \longrightarrow \omega + i\Gamma$) and taking into account relations (22), we obtain from formula (25) that

$$Q = \frac{1}{2}\omega f_1 \beta''_{zy} |E_{0y}|^2, \tag{26}$$

where β''_{zy} is the imaginary part of the susceptibility and $\beta_{zy} = \beta'_{zy} + i\beta''_{zy}$ (we do not write the obvious formula for β''_{zy}).

Passing to quasiantiferromagnetic mode $\Gamma_{35}(M_z, L_{by}, L_{cx})$, we find from Eqs. (9) and (21) that

$$M_z = \alpha_{zx} E_x = \frac{s_3}{\tilde{A}_M - K_M} \frac{\omega_{rb}^2}{\omega_{rb}^2 - \omega^2} E_x,$$
$$L_{by} = -\frac{i\omega \tilde{A}_c}{4\omega_0 r^2} M_z, \quad L_{cx} = \frac{r}{\tilde{A}_c} L_{by}, \tag{27}$$

where α_{zx} is the component of the magnetoelectric susceptibility tensor and

$$\omega_{rb} = 2\omega_0 r \sqrt{(\tilde{A}_M - K_M)/\tilde{A}_c} \tag{28}$$

is the eigenfrequency of oscillations determined by the relativistic interaction (the term containing r in formula (5)) and dependent on the ratio of exchange constants. The presence of the low-frequency mode Γ_{35} is connected with the noncollinearity of the ground state (an admixture of vector $\mathbf{L}_c^0$). Note, however, that the biquadratic anisotropy described by invariant $I = q_b L_{bx}^2 L_{by}^2 + q_c L_{cx}^2 L_{cy}^2$ in the expression for Φ_{mag} can also contribute to the magnitude of frequency ω_{rb} (28). Taking into account this anisotropy, we should renormalize constant r in formula (28):

$$r \longrightarrow r^* = \sqrt{r^2 + 8 q_b M_0^2 \tilde{A}_c}. \tag{29}$$

For mode $\Gamma_{18}(M_x, L_{ay}, L_{cz})$, the solutions to the Landau–Lifshitz equations (9) have the form

$$L_{ay} = \beta_{yz} E_z = -\lambda \frac{s_2}{\tilde{A}_a} \frac{\omega_E^2}{\omega_E^2 - \omega^2} E_z,$$
$$L_{cz} = \frac{i\omega}{\omega_0(\tilde{A}_c - K_c)} L_{ay}, \quad M_x = \frac{r}{\tilde{A}_c} L_{ay}. \tag{30}$$

The excitation of this mode is caused by the magnetoelectric interaction connected in relations (21) with longitudinal oscillations $M_x = -\lambda L_{ay}$; therefore, susceptibility β_{yz} containing parameter $\lambda \ll 1$ (14) is small. In this case, the eigenfrequency

$$\omega_E = \omega_0 \sqrt{\tilde{A}_a(\tilde{A}_c - K_c)} \tag{31}$$

of oscillations is, in contrast to eigenfrequency (28), an exchange frequency. This is connected with the fact that the transverse oscillations of mode $\Gamma_{18}(M_x, L_{ay}, L_{cz})$ are associated in relations (21) with the exchange energy with coefficients $\tilde{A}_a$ and $\tilde{A}_c - K_c$, and the energy of transverse oscillations for mode $\Gamma_{35}(M_z, L_{by}, L_{cx})$ contains both the exchange part (with coefficient $\tilde{A}_M - K_M$) and the relativistic part (with coefficient $2r^2/\tilde{A}_c$); therefore, its eigenfrequency is relativistic.

One can see from relations (21) that quasiantiferromagnetic mode $\Gamma_{27}(M_y, L_{ax}, L_{bz})$ with eigenfrequency

$$\omega_{AF} = \omega_0 \sqrt{\tilde{A}_M |K_b|} \tag{32}$$

(semiexchange, semirelativistic) is not excited by electric field $\mathbf{E}(t)$, because the magnetoelectric and antiferroelectric interactions are absent for this mode.

Note that the results obtained here for phase $\Gamma_4(L_{bx}^0, L_{cy}^0)$ (where $L_{bx}^0 \gg L_{cy}^0$) can be extended to vibrational modes of phase $\Gamma_5(L_{by}^0, L_{cx}^0)$ (where $L_{by}^0 \gg L_{cx}^0$) by using the substitution $x \longleftrightarrow y$, $\omega \longrightarrow -\omega$.

2. Collinear phase with vector $\mathbf{L}_b^0 \parallel$ [110]. Consider united phase $\Gamma_4 + \Gamma_5$, in which the primitive vector $\mathbf{L}_b^0$ (or $\mathbf{L}_c^0$) is directed along the diagonal of the base square [110]. It is possible that it is this phase that is realized in the easy-plane trirutiles mentioned above. It has been noted that the experiment does not allow one to make an unambiguous choice [1]. Moreover, this state is collinear, since, by virtue of conditions (8), we

have only two possibilities in the equimodule model, namely,

$$\mathbf{L}_b^0 \parallel [110], \quad \mathbf{L}_c^0 = 0 \tag{33a}$$

or

$$\mathbf{L}_c^0 \parallel [110], \quad \mathbf{L}_b^0 = 0. \tag{33b}$$

In diagonal phases (33), the tetragonal axis 4_{2z} mixes up representations $\Gamma_{18}(M_x, L_{ay}, L_{cz})$ and $\Gamma_{27}(M_y, L_{ax}, L_{bz})$ (or Γ_{28} and Γ_{17}), as well as $\Gamma_{35}(M_z, \Delta L_{by}, \Delta L_{cx})$ and $\Gamma_{46}(L_{az}, \Delta L_{bx}, \Delta L_{cy})$ (or Γ_{34} and Γ_{56}). As a result, the united quasiferromagnetic and quasiantiferromagnetic modes are formed with the representations

$$\Gamma_{18} + \Gamma_{27} \equiv \Gamma_{18+27}(M_x, M_y, L_{ax}, L_{ay}, L_{bz}, L_{cz}) \tag{34}$$

and

$$\begin{aligned}&\Gamma_{35} + \Gamma_{46}\\ &\equiv \Gamma_{35+46}(M_z, L_{az}, \Delta L_{bx}, \Delta L_{by}, \Delta L_{cx}, \Delta L_{cy}),\end{aligned} \tag{35}$$

respectively. The first of these modes is not excited by an electric field (therefore, we do not consider it). The second mode is excited by field $\mathbf{E}(t) \perp z$.

Suppose that $\mathbf{L}_b^0 \parallel [110]$ and $\mathbf{L}_c^0 = 0$. Then, taking into account conditions (8) corresponding to the equimodule model, we obtain from relation (4) the following quadratic form Φ_2 for quasiantiferromagnetic mode $\Gamma_{35} + \Gamma_{46}$:

$$\begin{aligned}\Phi_2 = {} & \frac{1}{2}(\tilde{A}_M - K_M)M_z^2 + \frac{1}{2}(\tilde{A}_a - K_a)L_{az}^2\\ & + \tilde{A}_c L_{cx}^2 - 4r\Delta L_{bx}L_{cx} - rM_zL_{az}\\ & - \frac{\sqrt{2}}{2}(s_3M_z + f_1L_{az})(E_x + E_y),\end{aligned} \tag{36}$$

where

$$\tilde{A}_\kappa = A_\kappa - A_b + K_\kappa - K_b, \quad \kappa = M, a, c,$$

$$A_\kappa, K_\kappa > 0, \quad A_b, K_b < 0.$$

It is convenient to represent the solutions to the Landau–Lifshitz equations (9) obtained taking into account relation (36) in a rotated coordinate system with the axes $x' \parallel \mathbf{L}_b^0 \parallel [110]$, $y' \parallel [\bar{1}10]$, and $z' \parallel z \parallel 4_{2z}$. Considering that, in this coordinate system, we have $L'_{by} = -L_{bx}\sqrt{2}$, $L'_{cy} = -L_{cx}\sqrt{2}$, and $E'_x = (E_x + E_y)/\sqrt{2}$, we find that quasiantiferromagnetic mode Γ_{35+46} has an exchange frequency ω_{Lb} (23). Near the resonance, $\omega = \omega_{Lb}$, we approximately obtain

$$\begin{aligned}L_{az} &= \frac{f_1}{\tilde{A}_a - K_a}\frac{\omega_{Lb}^2}{\omega_{Lb}^2 - \omega^2}E'_x, \quad L'_{cy} = -\frac{i\omega}{\omega_0\tilde{A}_c}L_{az},\\ L'_{by} &= \frac{i\omega_0 s_3}{\omega}\left(1 + \frac{\Omega_1^2}{\omega_{Lb}^2 - \omega^2}\right)E'_x,\\ M_z &= -2\frac{r}{\tilde{A}_c}L_{az},\end{aligned} \tag{37}$$

and $\Delta L'_{bx} = L'_{cx} = 0$. Here, we assume that

$$\omega_{Lb}^2 \gg \Omega_2^2 = 2\omega_0^2 r\frac{s_3}{f_1}(\tilde{A}_a - K_a)$$

and use the notation

$$\Omega_1^2 = \omega_0^2 r\frac{f_1}{s_3}[2(\tilde{A}_M - K_M) + \tilde{A}_c].$$

Formulas (37) imply that the quasiantiferromagnetic mode under consideration is transverse (relative to $\mathbf{L}_b^0 \parallel [110]$) and is excited by electric field $\mathbf{E}(t) \parallel \mathbf{L}_b^0 \parallel [110]$. The magnetoelectric and antiferroelectric susceptibilities corresponding to this mode have a pole at antimagnonic frequency ω_{Lb} (23), but they are not antimagnons.

5.3. Magnetic Structure with Vector $\mathbf{L}_c^0$ (Cr_2WO_6 and V_2WO_6)

1. Weakly collinear phase $\Gamma_5(L_{by}^0, L_{cx}^0)$. Now, let vector $\mathbf{L}_c^0 \parallel x$ ($L_c^0 \approx 4M_0$) be the primitive base vector. This orientation state corresponds to phase $\Gamma_5(L_{by}^0, L_{cx}^0)$. In this case we have

$$\varepsilon = \frac{L_{by}^0}{L_{cx}^0} \approx -\frac{r}{\tilde{A}_b} \ll 1 \tag{38}$$

and quadratic form Φ_2 written for transverse oscillations (with respect to $\mathbf{L}_c^0 \parallel x$) has the form

$$\begin{aligned}\Phi_2 = {} & \frac{1}{2}\tilde{A}_M M_y^2 + \frac{1}{2}(\tilde{A}_M - K_M)M_z^2 + \frac{1}{2}\tilde{A}_a L_{ay}^2\\ & + \frac{1}{2}(\tilde{A}_a - K_a)L_{az}^2 + \frac{1}{2}\tilde{A}_b L_{by}^2\\ & + \frac{1}{2}(\tilde{A}_b - K_b)L_{bz}^2 + \frac{2r^2}{\tilde{A}_b}L_{cy}^2 - \frac{1}{2}K_c L_{cz}^2\\ & - d_2M_yE_z - d_3M_zE_y - f_2L_{az}E_x.\end{aligned} \tag{39}$$

Now, we have $\tilde{A}_\kappa = A_\kappa - A_c + K_\kappa - K_c$, $\kappa = M, a, b$ ($A_c, K_c < 0, A_\kappa, K_\kappa > 0$).

In this phase, mode $\Gamma_{17}(M_x, L_{ay}, L_{bz})$, as well as mode $\Gamma_{27}(M_y, L_{ax}, L_{bz})$ of phase $\Gamma_4(L^0_{bx}, L^0_{cy})$ considered above, is not excited by field $\mathbf{E}(t)$. Antimagnonic mode $\Gamma_{56}(L_{az}, \Delta L_{by}, \Delta L_{cx})$ with exchange frequency

$$\omega_{Lc} = \omega_0\sqrt{\tilde{A}_b(\tilde{A}_a - K_a)} \tag{40}$$

is excited by electric field $\mathbf{E}(t) \parallel x$, and quasiantiferromagnetic mode $\Gamma_{34}(M_z, L_{bx}, L_{cy})$ with relativistic frequency

$$\omega_{rc} = 2\omega_0 r\sqrt{(\tilde{A}_M - K_M)/\tilde{A}_b} \tag{41}$$

is excited by field $\mathbf{E}(t) \parallel y$. Mode variables L_{az}, ΔL_{by}, ΔL_{cx} and M_z, L_{bx}, L_{cy} can be defined by formulas (22) and (27), respectively, applying to them the substitutions

$$\beta_{zy} \longrightarrow \beta_{zx}, \quad \alpha_{zx} \longrightarrow \alpha_{zy},$$

$$E_x \longleftrightarrow E_y, \quad f_1 \longrightarrow f_2, \quad s_3 \longrightarrow d_3,$$

as well as the change of indices $b \longleftrightarrow c$. The heat loss due to the excitation of antimagnonic mode $\Gamma_{56}(L_{az}, \Delta L_{by}, \Delta L_{cx})$ can be determined by formula (26) by using the substitutions $f_1 \longrightarrow f_2$ and $y \longrightarrow x$. The renormalization of constant $r \longrightarrow r^*$ arising when taking into account biquadratic term $q_c L^2_{cx} L^2_{cy}$ in expression (5) for Φ_{mag} can be determined from formula (29) with the help of the change of indices $b \longleftrightarrow c$.

For mode $\Gamma_{28}(M_y, L_{ax}, L_{cz})$, using relation (39) in the Landau–Lifshitz equations (9), we obtain

$$\begin{aligned} M_y &= \alpha_{yz}E_z = \frac{d_2}{\tilde{A}_M}\frac{\omega^2_{AF}}{\omega^2_{AF} - \omega^2}E_z, \\ L_{cz} &= \frac{i\omega}{\omega_0|K_c|}M_y, \quad L_{ax} = \frac{r}{\tilde{A}_b}M_y. \end{aligned} \tag{42}$$

Here, eigenfrequency

$$\omega_{AF} = \omega_0\sqrt{\tilde{A}_M|K_c|}, \tag{43}$$

unlike frequency ω_E (31) of mode $\Gamma_{18}(M_x, L_{ay}, L_{cz})$, is a semiexchange, semirelativistic frequency. This is connected with the fact that modes $\Gamma_{18}(M_x, L_{ay}, L_{cz})$ and $\Gamma_{28}(M_y, L_{ax}, L_{cz})$ are nonequivalent here because they relate to different magnetic structures with base vectors $\mathbf{L}^0_b$ and $\mathbf{L}^0_c$, respectively. Hence, the energies and eigenfrequencies of these modes are also different.

The results presented here relate to phase $\Gamma_5(L^0_{by}, L^0_{cx})$ with $L^0_{cx} \gg L^0_{by}$. These results also remain valid for phase $\Gamma_4(L^0_{bx}, L^0_{cy})$ with $L^0_{cy} \gg L^0_{bx}$ after substituting $x \longrightarrow y$ and $\omega \longrightarrow -\omega$.

2. Collinear phase with $\mathbf{L}^0_c \parallel [110]$. If the primitive base vector in united phase $\Gamma_4 + \Gamma_5$ is $\mathbf{L}^0_c \parallel [110]$ and $\mathbf{L}^0_b = 0$ by virtue of conditions (8), then electric field $\mathbf{E}(t) \parallel [110]$ again excites quasiantiferromagnetic mode $\Gamma_{35} + \Gamma_{46}$ (35). It can be described by formulas (37) after the change of indices $b \longrightarrow c$ and constants $s_3 \longrightarrow d_3$, $f_1 \longrightarrow f_2$ in these formulas.

6. EXCITATION OF NUCLEAR MAGNETIC RESONANCE BY ELECTRIC FIELD

The effects of excitation of antimagnons and quasiantiferromagnons by electric field $\mathbf{E}(t)$ considered above are determined by magnetoelectric and antiferroelectric interactions. The antimagnon frequency is in the optical (exchange) region, and the quasiantiferromagnon frequency can be rather low (a relativistic frequency in the microwave region).

In the region of lower (radio) frequencies, these interactions can be manifested in the effect of excitation of the NMR by electric field $\mathbf{E}(t)$, where a quasiequilibrium relation between $\mathbf{E}(t)$ and $\mathbf{L}(t)$ can exist [12].

In the absence of an external magnetic field, the NMR frequency

$$\omega_n = \gamma_n H_{n\nu} \tag{44}$$

is determined by the constant part of the hyperfine field

$$H_{n\nu} = FM_{\nu 0} \tag{45}$$

of the electron subsystem. Here, γ_n is the nuclear gyromagnetic ratio, F is the hyperfine interaction constant, and $M_{\nu 0}$ is the magnetization of the νth sublattice in the ground state.

First, consider the NMR excitation by electric field $\mathbf{E}(t) \parallel y$ in magnetic structures with primitive base vector $\mathbf{L}^0_b \parallel x$ (Cr_2TeO_6) corresponding to phase $\Gamma_4(L^0_{bx}, L^0_{cy})$. We shall consider only the exchange approximation and frequencies

$$\omega \approx \omega_n \ll \omega_{Lb} \tag{46}$$

that are lower than antimagnon frequency ω_{Lb} (23). In this case, oscillations of L_{az} generated, in accordance

with Eqs. (22), by field $E_y(t)$ follow this field in the quasiequilibrium way:

$$L_{az} = \frac{f_1}{A_a - A_b} E_y, \tag{47}$$

where $A_a > 0$, $A_b < 0$.

The effective magnetic fields on nuclei, which are generated by magnetization M_ν of sublattices, consist of the constant component $\mathbf{H}_{n\nu}$ (45) and the alternating component $\mathbf{h}_\nu$ determined by oscillations ΔM_ν, which are connected with L_{az}. Since $\mathbf{L}_b^0 \neq 0$ and $\mathbf{M}^0 = \mathbf{L}_a^0 = 0$ in the ground state and $\mathbf{L}_c^0 = 0$ in the exchange approximation accepted here, by virtue of formulas (3), we have $M_{10} = M_{30} = M_0$ and $M_{20} = M_{40} = -M_0$. Hence, constant fields $\mathbf{H}_{n\nu}$ in direction $\mathbf{L}_b^0 \parallel x$ are equal to

$$H_{n1} = H_{n3} = FM_0, \quad H_{n2} = H_{n4} = -FM_0. \tag{48}$$

Definition (3) implies that L_{az} is associated with oscillations of magnetization $\Delta M_{1z} = \Delta M_{2z} = L_{az}/4$, $\Delta M_{3z} = \Delta M_{4z} = -L_{az}/4$, which correspond to fields

$$\begin{aligned} h_{1z} &= h_{2z} = -h_{3z} = -h_{4z} \\ &= \frac{1}{4} F L_{az} = \frac{1}{4} \frac{F f_1}{A_a - A_b} E_y \equiv h_z(t). \end{aligned} \tag{49}$$

Now, we determine the nuclear magnetization $\mathbf{m}_\nu = \mathbf{m}_{\nu 0} + \Delta\mathbf{m}_\nu$ for every sublattice from the equation of motion [12]

$$\frac{d\mathbf{m}_\nu}{dt} = \gamma_n(\mathbf{m}_{\nu 0} \times \mathbf{h}_\nu + \Delta\mathbf{m}_\nu \times \mathbf{H}_{n\nu}), \tag{50}$$

where $\mathbf{m}_{n0} = \chi_{n0}\mathbf{H}_{n\nu}$ and $\Delta\mathbf{m}_\nu$ are the constant and dynamic components of $\mathbf{m}_\nu$ and χ_{n0} is the static nuclear susceptibility. Taking into account relations (48) and (49), we find from formula (50) that

$$m_{1y} = m_{4y} = -m_{2y} = -m_{3y} = \frac{i\omega\omega_n\chi_{n0}h_z}{\omega_n^2 - \omega^2}, \tag{51}$$

$$\begin{aligned} m_{1z} &= m_{2z} = -m_{3z} = -m_{4z} \\ &= \frac{\omega_n^2\chi_{n0}h_z}{\omega_n^2 - \omega^2} = \chi_n(\omega)h_z. \end{aligned} \tag{52}$$

Here, $m_{\nu y} = \Delta m_{\nu y}$, $m_{\nu z} = \Delta m_{\nu z}$, and

$$\chi_n(\omega) = \chi_{n0}\frac{\omega_n^2}{\omega_n^2 - \omega^2} \tag{53}$$

is the susceptibility determining the magnetization component which is in phase with field $\mathbf{h}_\nu$.

Relations (51) and (52) imply that the total nuclear magnetization generated by electric field $E_y(t)$ is equal to zero,

$$\mathbf{m} = \sum_\nu \mathbf{m}_\nu = 0. \tag{54}$$

Nevertheless, the resonance in field $E_y(t)$ at the NMR frequency (44) exists. This can be found if we introduce the dissipation by the substitution $\omega \longrightarrow \omega + i\Gamma_n$ and calculate the heat loss (absorbed energy) [12].

The total energy absorption for all sublattices is defined by the expression

$$Q = \sum_\nu Q_\nu = -\sum_\nu \overline{m_{\nu z}\dot{h}_z}. \tag{55}$$

Taking into account relations (52), this equality implies that

$$Q = 2\omega\chi_n''(\omega)|h_z|^2, \tag{56}$$

where $\chi_n''(\omega)$ is the imaginary part of $\chi_n(\omega)$, which is equal [12] to

$$\chi_n''(\omega) = \chi_{n0}\frac{2\omega\Gamma_n\omega_n^2}{(\omega_n^2 - \omega^2)^2 + (2\omega\Gamma_n)^2}. \tag{57}$$

If the primitive base vector is $\mathbf{L}_c^0 \parallel y$ ($\mathbf{L}_b^0 = 0$), then quantity h_z defined by expression (49) should be replaced in formula (56) by

$$h_z = \frac{1}{4}\frac{F f_2}{A_a - A_c} E_y, \tag{58}$$

where $A_a > 0$ and $A_c < 0$.

The above discussion relates to phase $\Gamma_4(L_{bx}^0, L_{cy}^0)$. The nuclear magnetoelectric resonance in the equivalent phase $\Gamma_5(L_{by}^0, L_{cx}^0)$ can be excited by electric field $\mathbf{E}(t) \parallel x$, which is also connected with L_{az} via a quasiequilibrium relation of form (47). In this case, alternating magnetic field h_z determining the energy absorption Q (56) can be calculated by formulas (49) for $\mathbf{L}_b^0 \parallel y$, $\mathbf{L}_c^0 = 0$ or (58) for $\mathbf{L}_c^0 \parallel x$, $\mathbf{L}_b^0 = 0$ by replacing E_y with E_x in these formulas.

Now, consider the NMR excitation in the diagonal phase with $\mathbf{L}_b^0 \parallel [110]$ ($\mathbf{L}_c^0 = 0$) for $\mathbf{E}(t) \perp z$. We again

consider only the exchange approximation and low frequencies

$$\omega \approx \omega_n \ll \omega_{rb} \ll \omega_{Lb}. \tag{59}$$

In this case, oscillations of quantities M_z and L_{az} connected with field $\mathbf{E}(t)$ in a quasiequilibrium way arise. Minimizing Φ_2 (36) in M_z and L_{az} and passing to the primed coordinate system ($x' \parallel [110]$, $y' \parallel [\bar{1}10]$, $z' \parallel z \parallel 4_{2z}$), we obtain

$$M_z = \frac{s_3}{A_M - A_b}E'_x, \quad L_{az} = \frac{f_1}{A_a - A_b}E'_x. \tag{60}$$

These oscillations are associated with effective magnetic fields

$$\begin{aligned} h_{1z} &= h_{2z} = \frac{F}{4}(M_z + L_{az}), \\ h_{3z} &= h_{4z} = \frac{F}{4}(M_z - L_{az}), \end{aligned} \tag{61}$$

and constant fields

$$H'_{n1} = H'_{n3} = -H'_{n2} = -H'_{n4} = FM_0 \tag{62}$$

are directed along $\mathbf{L}_b^0 \parallel [110] \parallel x'$. Taking into account relations (61) and (62), from the equations of motion (50), we obtain

$$\begin{aligned} \Delta m'_{1y} &= -\Delta m'_{2y} = \frac{i\omega}{\omega_n}\chi_n(\omega)h_{1z}, \\ \Delta m'_{3y} &= -\Delta m'_{4y} = \frac{i\omega}{\omega_n}\chi_n(\omega)h_{3z}, \\ \Delta m_{1z} &= \Delta m_{2z} = \chi_n(\omega)h_{1z}, \\ \Delta m_{3z} &= \Delta m_{4z} = \chi_n(\omega)h_{3z}, \end{aligned} \tag{63}$$

which gives

$$\begin{aligned} m'_y &= \sum_\nu \Delta m'_{\nu y} = 0, \\ m_z &= \sum_\nu \Delta m_{\nu z} = \chi_n(\omega)FM_z = \frac{\chi_n(\omega)Fs_3}{A_M - A_b}E'_x. \end{aligned} \tag{64}$$

Thus, in this case, in contrast to relation (54), the total nuclear magnetization m_z arises, which is caused by oscillations of M_z due to the magnetoelectric interaction.

The heat loss for all sublattices is again determined by formula (56), in which we should set

$$|h_z|^2 = |h_z^M|^2 + |h_z^L|^2, \tag{65}$$

where

$$\begin{aligned} h_z^M &= \frac{1}{4}FM_z = \frac{1}{4}\frac{Fs_3}{A_M - A_b}E'_x, \\ h_z^L &= \frac{1}{4}FL_{az} = \frac{1}{4}\frac{Ff_1}{A_a - A_b}E'_x \end{aligned} \tag{66}$$

are effective magnetic fields generated by oscillations of M_z and L_{az}.

The above formulas (60)–(66) are also valid for the antiferromagnetic structure with $\mathbf{L}_c^0 \parallel [110]$ ($\mathbf{L}_b^0 = 0$) after substituting $A_b \longrightarrow A_c$, $s_3 \longrightarrow d_3$, $f_1 \longrightarrow f_2$ in these formulas. Thus, there exists a resonance energy absorption at the NMR frequency ω_n (44). The intensity of this resonance is determined by alternating field $h_z(t)$, the order of magnitude of which can be estimated from the relation [12]

$$h_z(t) \approx (\alpha F)E_i(t), \quad i = x, y,$$

where the magnetoelectric susceptibility is $\alpha \approx 10^{-5}$–10^{-2} (in CGS units) and the hyperfine interaction constant is $F \approx 10^2$–10^3. In magnets with $\alpha F \approx 1$, quantity $h_z(t)$ can be sufficient for observing the NMR in electric field $E_i(t)$, $i = x, y$.

7. CONCLUSIONS

In the group of tetragonal trirutiles, i.e., four-sublattice antiferromagnets, magnetic ions occupy a noncentrosymmertical position due to the presence of both magnetoelectric and antiferroelectric interactions. Moreover, they have two types of exchange magnetic structures with centroantisymmetric base vectors $\mathbf{L}_b^0$ and $\mathbf{L}_c^0$ (see Table 7.1 in [1]) with different orientation states, namely, easy-axis and easy-plane, collinear and weakly collinear. As a whole, this makes them rather rich objects for the development of spin dynamics considered in this paper and is connected with the very wide manifestation of these interactions. It turns out that not only purely antiferromagnetic magnons (electroactive antimagnons) with exchange eigenfrequencies (in the pure form, they are described by formulas (23) and (40)) can exist, but also semirelativistic quasiantiferromagnetic magnons excited by both fields $\mathbf{E}(t)$ and $\mathbf{H}(t)$ (formulas (16) and (43)) or, as in ordinary antiferromagnets, only by field $\mathbf{H}(t)$ by virtue of the absence of the corresponding magnetoelectric and antiferroelectric interactions for them. These are modes Γ_{17} and Γ_{27} (see, e.g., frequency (32) for the latter mode). There are also electrically and magnetically active purely relativistic magnons (28) and (41). It seems that they have the lowest frequencies among all magnons excited by field $\mathbf{E}(t)$. Lastly, the exchange magnons corresponding to formulas (18) are electrically active antimagnons only in a certain approximation.

Physical phenomena connected with a large number of magnon types excited by both magnetic and electric fields can play a practical role. It seems possible that, by analyzing the results presented, we can clarify or solve the disputable question about the magnetic structure and orientation state of the trirutiles under investigation in addition to the data obtained from neutron diffractometry and the static magnetoelectric effect [1].

If the primitive base vector ($\mathbf{L}_b^0$ or $\mathbf{L}_c^0$) is directed along the $\langle 100 \rangle$ edges of the base square, the problem can be solved for $\mathbf{L}_b^0$ and $\mathbf{L}_c^0$ by comparing modes $\Gamma_{18}(M_x, L_{ay}, L_{cz})$ and $\Gamma_{28}(M_y, L_{ax}, L_{cz})$ for $\Gamma_{28}(M_y, L_{ax}, L_{cz})$ with $\mathbf{E} \parallel z$ in both the cases. The observation of the first mode with exchange frequency (31), but with a very low intensity (or nonobservation due to such an intensity) would correspond to the easy-plane structure with vector $\mathbf{L}_b^0$, and the existence of quasiantiferromagnetic mode Γ_{28} with semirelativistic frequency (43) would support the structure with principal base vector $\mathbf{L}_c^0$.

At the same time, the question whether base vectors $\mathbf{L}_b^0$ and $\mathbf{L}_c^0$ are directed along an edge or a diagonal of the base square (in the latter case, one of them is equal to zero, since the structure is collinear [1]) can also be answered in the case of field $\mathbf{E} \parallel \langle 100 \rangle$. In this case, if the quasiantiferromagnetic mode with relativistic frequency ω_{rb} (28) (or ω_{rc} (41)) is excited, vector $\mathbf{L}_b^0$ (or $\mathbf{L}_c^0$), as well as $\mathbf{E}$, is directed along an edge. If the quasiantiferromagnetic mode with exchange frequency ω_{Lb} (23) (or ω_{Lc} (40)) is excited, then vector $\mathbf{L}_b^0$ (or $\mathbf{L}_c^0$) is directed along a diagonal.

Similarly, an experiment can help to determine the NMR magnetic structure: if the principal base vector $\mathbf{L}_b^0$ (or $\mathbf{L}_c^0$) is directed along the edge $\langle 100 \rangle$, the total nuclear magnetization generated by field $\mathbf{E}(t) \parallel y$ (or $\mathbf{E}(t) \parallel x$) is equal to zero (i.e., $\mathbf{m} = \sum_\nu \Delta m_\nu = 0$), whereas, for $\mathbf{L}_b^0$ or $\mathbf{L}_c^0$ directed along a diagonal in the coordinate system with field $\mathbf{E}_x' \parallel [110]$, the total component is

$$m_y' = \sum_\nu \Delta m_{\nu y}' = 0$$

and the total component $m_z = \sum_\nu \Delta m_{\nu z} = 4\chi_n(\omega) h_z^M$ is nonzero.

Thus, the results presented in this paper indicate that experimental investigations of antiferromagnetic, antiferroelectric, and nuclear magnetoelectric resonances in trirutiles are necessary.

ACKNOWLEDGMENTS

The work was supported financially by the Russian Foundation for Basic Research (project no. 02-02-16440).

REFERENCES

1. E. A. Turov, A. V. Kolchanov, V. V. Men'shenin, *et al.*, *Symmetry and Physical Properties of Antiferromagnetics* (Fizmatlit, Moscow, 2001).
2. D. A. Yablonskiĭ and V. N. Krivoruchko, Fiz. Nizk. Temp. **14**, 656 (1988) [Sov. J. Low Temp. Phys. **14**, 363 (1988)].
3. V. N. Krivoruchko and D. A. Yablonskiĭ, Zh. Éksp. Teor. Fiz. **94** (9), 268 (1988) [Sov. Phys. JETP **67**, 1886 (1988)].
4. D. A. Yablonskiĭ and V. N. Krivoruchko, Fiz. Tverd. Tela (Leningrad) **30**, 2069 (1988) [Sov. Phys. Solid State **30**, 1765 (1988)].
5. V. V. Eremenko, V. N. Krivoruchko, N. M. Lavrinenko, *et al.*, Fiz. Tverd. Tela (Leningrad) **30**, 3605 (1988) [Sov. Phys. Solid State **30**, 2070 (1988)].
6. D. A. Yablonskiĭ and V. N. Krivoruchko, in *Problems in Physical Kinetics and Physics of Solid State* (Naukova Dumka, Kiev, 1990), p. 444.
7. E. A. Turov, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 92 (2001) [JETP Lett. **73**, 87 (2001)].
8. E. A. Turov and A. V. Kolchanov, in *Magnetism of Transition Metals and Alloys* (Ural. Otd. Ross. Akad. Nauk, Ekaterinburg, 2000), p. 85.
9. E. A. Turov and A. V. Kolchanov, Phys. Met. Metallogr. **92** (Suppl.), S67 (2001).
10. A. V. Kolchanov and E. A. Turov, Phys. Met. Metallogr. **92** (Suppl.), S82 (2001).
11. I. F. Mirsaev and E. A. Turov, Zh. Éksp. Teor. Fiz. **121**, 419 (2002) [JETP **94**, 356 (2002)].
12. E. A. Turov, Preprint of Inst. of Metal Physics, Ural Division, Russian Academy of Sciences, Ekaterinburg (2001).
13. V. V. Leskovets, M. I. Kurkin, V. V. Nikolaev, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **44**, 1272 (2002) [Phys. Solid State **44**, 1330 (2002)].
14. E. A. Turov and I. F. Mirsaev, Fiz. Nizk. Temp. **28**, 822 (2002) [Low Temp. Phys. **28**, 592 (2002)].
15. E. A. Turov, A. V. Kolchanov, V. V. Men'shenin, *et al.*, Usp. Fiz. Nauk **168**, 1303 (1998) [Phys.–Usp. **41**, 1191 (1998)].
16. Yu. A. Izyumov and N. A. Chernoplekov, in *Neutron Spectroscopy*, Ed. by R. P. Ozerov (Énergoatomizdat, Moscow, 1983; Consultants Bureau, New York, 1994).
17. V. G. Bar'yakhtar, I. M. Vitebskiĭ, and D. A. Yablonskiĭ, Zh. Éksp. Teor. Fiz. **76**, 1381 (1979) [Sov. Phys. JETP **49**, 703 (1979)].

Translated by E. Pankratiev

Journal of Experimental and Theoretical Physics, Vol. 97, No. 2, 2003, pp. 317–330.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 2, 2003, pp. 351–366.

SOLIDS
Electronic Properties

Single-Electron Charging of Triangular Quantum Dots in a Ring Interferometer

V. A. Tkachenko[a,*], A. A. Bykov[a], D. G. Baksheev[a], O. A. Tkachenko[a], L. V. Litvin[a], A. V. Latyshev[a], T. A. Gavrilova[a], A. L. Aseev[a], O. Estibals[b], and J. C. Portal[b]

[a]*Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia*

**e-mail: vtkach@isp.nsc.ru*

[b]*Grenoble High Magnetic Fields Laboratory, MPI-FKF and CNRS, B.P. 166, F-38042 Grenoble, France*

Received January 10, 2003

Abstract—Small-size semiconductor ring interferometers operating in the Coulomb blockade regime have been experimentally and theoretically studied. The conductance as a function of the gate voltage exhibits narrow quasiperiodic peaks, which are further split into doublets. Based on the experimental structural data, a three-dimensional electrostatic potential, the energy spectrum, and the single-electron transport in the interferometer were modeled. The electron system can be divided into two triangular quantum dots connected by single-mode microcontacts to each other and to the reservoirs. A model of quantum dot charging in this system is proposed that explains the appearance of doublets in the conductance–gate voltage characteristics. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Many of the effects manifested in mesoscopic electron transport [1–3] are explained in terms of the quantum interference [4–6] or the Coulomb blockade [7–10]. These phenomena, observed under conditions of coherent transport or under the opposite conditions of sequential tunneling, are usually studied in special submicron devices. In particular, the Coulomb blockade was studied in much detail for a single closed quantum dot [11–14] representing an electron pool separated by tunneling barriers from the reservoirs containing a two-dimensional electron gas (2DEG). In contrast, the interference regime is observed in open systems featuring a strong coupling between a nanodevice and the conducting channels. An example is offered by a ballistic semiconductor ring interferometer [15–20] representing a quantum ring connected to the reservoirs by short quantum wires.

Recently, devices appeared featuring both the interference and charging phenomena [21–25]. The conductance of combined devices representing quantum dots incorporated into the arms of a ballistic interferometer [21–23] exhibited both the Aharonov–Bohm (AB) oscillations [4, 6], related to the interference via two paths, and the peaks caused by the gate voltage variations, related to the Coulomb blockade [10, 11] of the tunneling via the quantum dot. Thus, it was experimentally established that processes of the sequential tunneling and relaxation, leading to quantum dot charging, may coexist with the coherent ring transport.

New experiments [24, 25] showed that small-size megaohm quantum rings may exhibit both periodic Coulomb oscillations and AB oscillations even without quantum dots incorporated. Despite the close dimensions of the device used in the experiments cited, the electron transport regimes in the two cases were different. In the experiments of Holleitner *et al.* [25], the ring was separated from the 2DEG reservoirs by potential barriers controlled by the finger gates occurring in the same plane. When the barriers were sufficiently high, the ring exhibited single-electron charging as a whole, whereas the motion of electrons in the ring remained free. In this case, the source of the Coulomb oscillations is obvious and there is no problem of coexistence between the ring interference and the Coulomb blockade.

Our experiments [24] were performed with an interferometer containing a wide in-plane split gate controlling the electron transmission in both finger and ring channels. The results were interpreted assuming that only two regions of high electron density at the ring entrance and exit are charged rather than the whole ring. The possibility of the existence of such regions, called triangular quantum dots, follows from the electrostatic considerations [20]. It was suggested [24] that the triangular quantum dots are spaced by tunneling barriers from each other and from the reservoirs and are charged by the Coulomb blockade mechanism [10]. However, no clear evidence was found in favor of that hypothesis.

This study continues investigations initiated previously [20, 24]. In addition to detecting periodic peaks in the conductance as a function of the gate voltage, we observed a doublet splitting of these peaks evidencing a Coulomb interaction between the triangular quantum dots. Earlier, it was demonstrated theoretically [26, 27] and established experimentally [28] that the Coulomb blockade in a sequence of two quantum dots is essentially different from that in a single quantum dot. In the former case, the peak distribution is less regular and the existence of a mutual capacitance leads to the appearance of doublets instead of the periodic peaks observed in the latter case. Recently, such doublets were observed in parallel-connected quantum dots incorporated into the arms of an AB interferometer [23]. In this study, we have modeled a three-dimensional electrostatic potential, a single- electron transport, and the quantum transfer in a ring interferometer. Based on the results, it is possible to conclude that a small-size large-resistance interferometer exhibits an analogous effect of the charge interaction between triangular quantum dots. Thus, we have confirmed for the first time that two triangular quantum dots influence the conductance of the interferometer.

The aforementioned hypothesis concerning the tunneling insulation of triangular quantum dots [24] is at variance with both experiment and model calculations. The results of modeling showed that the presence of tunneling barriers in the ring arms suppresses the AB effect. The hypothesis also contradicts the strong influence of the mesoscopic state of the ring on the critical temperature for observation of the Coulomb oscillations. A more realistic pattern corresponds to the quantum dots connected by single-mode quantum wires to each other and to the reservoirs. By modeling the two-dimensional electron transport, it is possible to understand how this can system trap electrons and possess a large resistance. Following the investigations [29–35], we have elucidated the reasons of single-electron charging of the quantum dots beyond the tunneling regime. Based on a modified variant of the orthodox theory of the Coulomb blockade, we explain the mesoscopic character of renormalization of the Coulomb energy.

An initial information for the modeling is based on the structural data, including the images of the surface of device structures obtained by scanning electron microscopy (SEM) or by atomic force microscopy (AFM). The AFM images substantially supplement the SEM data by explaining "narrowness" of the channels connecting reservoirs and triangular quantum dots. Reliability of the modeling results and conclusions is checked by comparison of the calculated and measured values such as the initial 2DEG density, the characteristic resistance of the interferometer, the Coulomb oscillation period, the peak splitting, and the critical temperature for observing Coulomb blockade effects.

2. EXPERIMENT

2.1. Structural Data and the Aharonov–Bohm Effect

The sample devices were based on the GaAs/AlGaAs heterojunctions fabricated by molecular beam epitaxy techniques. The main special feature of the heterostructures studied was the small thickness (3 nm) of the spacer between the 2DEG and a bulk doped AlGaAs layer. For this reason, the 2DEG had a large density of $n_s = 1.45 \times 10^{12}$ cm^{-2} (upon illumination) and a relatively small mobility of $\mu = 3.4 \times 10^4$ cm^2/(V s) (at $T = 4.2$ K).

The interferometer geometry was formed by reactive ion sputtering, with a lateral profile determined by the electron-beam lithography at a 100 nm resolution. The device surface patterns (studied on the control samples only) exhibited significant variations. An example is offered by Fig. 1a, schematically representing an SEM image of one of the control samples, where shaded regions represent the etched paths *1* and *2*. The bright regions on the left and right correspond to the reservoirs, those at the top and bottom represent the 2DEG regions of the in-plane split gate, and a ring is formed around the etched site *3* (antidot) at the center.

The SEM images reproduce well the lateral dimensions, but hide the in-depth variations, which also significantly influence the shape of the electrostatic potential. Figures 1b and 1c show an AFM image and the corresponding profile of the surface of a different sample. As can be seen, there are etched bands above the quantum wires that appeared as a result of unintentional exposure of the resist between close lithographic paths *1* and *2*. These bands can either increase the device resistance or even completely insulate the ring from the reservoirs. These bands essentially increase the electrostatic barriers separating the ring from the reservoirs in the same manner as do the in-plane finger gates in a device studied in [25]. In addition, Figs. 1b and 1c reveal a rise of 10–15 nm in the surface relief near the etched channels, which is evidence of a damaged layer formed as a result of the reactive ion sputtering. This layer brings the charged semiconductor surface closer to the interferometer channels, thus effectively narrowing the channels.

For investigating the transport properties, we have selected a series of samples with a resistance of $R \sim 10^5$–10^6 Ω in a range of the bias voltage variation within several tenths of volt on the in-plane split gate. The same values were characteristic of quantum wires with a length of about 500 nm. The resistance was measured in a two-point-probe scheme (with potential contacts at a large distance from the sample) using a lock-in amplifier, for an ac signal in the frequency range from 7 to 800 Hz. The current was selected so that the drain–source voltage would not exceed $k_B T/e$ (to provide for the linear regime), where k_B is the Boltzmann constant, T is the absolute temperature, and e is the electron charge.

As noted above, the structural variations of the obtained samples were large, so that only a part of the devices were suited for the investigation. Indeed, if the ring was broken, the main interference effect—the AB oscillations—could not be studied. Nevertheless, we observed the AB effect in two samples possessing a high resistance (see Fig. 2 in this paper and [2, Fig. 24]). The AB effect was much more seldom in weak magnetic fields than in strong ones. However, the AB oscillation period was constant over a broad range of variation of the resistance (2×10^4–2×10^6 Ω), magnetic field strength (0–4 T), and temperature (0.1–4.2 K), which is evidence of the same origin of the observed phenomena. The oscillations in the conductance G with respect to the magnetic inductance B were identified using the relation $\pi r^2 \Delta B = \phi_0$, where ΔB is the oscillation period, r = 130 and 210 nm are the effective radii of the rings in sample nos. 1 and 2, and $\phi_0 = hc/e$ is the magnetic flux quantum. The relation between ΔB and ϕ_0 is indicated by the correlated phase rotations of the magnetic-field-induced oscillations, whereby minima on one curve correspond to maxima on the other curve (Fig. 2). A twofold decrease in the AB oscillation period observed for $B > 10$ T is explained by the appearance of $hc/2e$ oscillations [24].

2.2. Coulomb Oscillations

The curves of the conductance G versus gate voltage V_g for $B = 0$ exhibit small-scale oscillations on the background of large-scale peaks (Figs. 3 and 4). The gate-voltage-controlled oscillations are of interest, since no such oscillations were observed for the control quantum wires. Oscillations with a period of 6 mV were observed in a broad range of temperatures and resistances (Figs. 3a–3c). The periodic character of these oscillations agrees with predictions of the Coulomb blockage theory for the structures with a single Coulomb island [10, 11]. The Coulomb character of the small-scale oscillations is indicated by the fact that the hill-to-valley conductance ratio near the gate voltage threshold grows when the temperature is decreased to 1.3 K (Fig. 3b and 3c), and by the appearance of clearly pronounced Coulomb blockade plateaus between narrow Coulomb peaks at a still lower temperature of 0.1 K (Figs. 4a and 4b). In sample no. 2, the oscillations are less periodic than in sample no. 1, but the characteristic period of 5–6 mV is still observed. Coincidence of the oscillations periods is evidence of the same characteristic size of the Coulomb island in both samples, despite different radii of the rings.

For sample no. 2, the main effect consists in the splitting or pairing of the Coulomb peaks (Fig. 4). In this respect, the period of 6 mV refers to the average distance between neighboring pairs. In fact, the pattern of oscillations is more complicated than simple repetition of the equidistant doublets. Both the splitting magnitude and the peak shape change with the gate voltage.

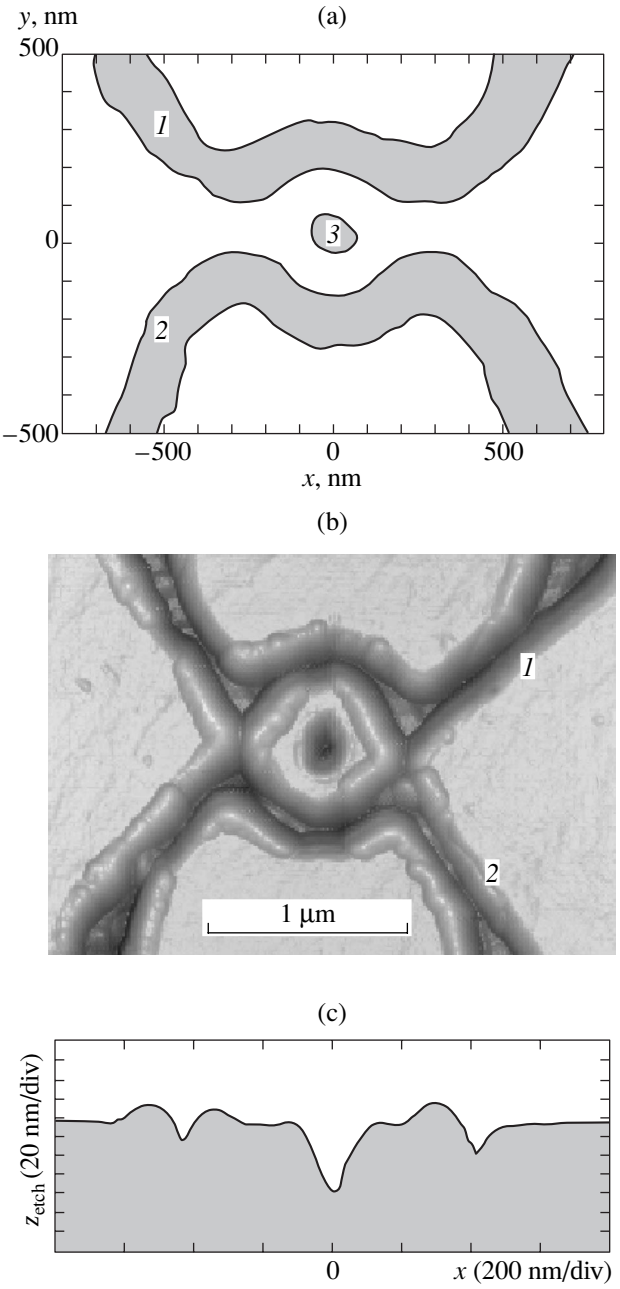

Fig. 1. Ring interferometer structures: (a) a diagram of etched regions *1–3* determining the device pattern (SEM data); (b) AFM image of a different control sample; (c) etching depth z_{etch} versus coordinate x profile corresponding to the central cross section of the structure imaged in (b), revealing a deep minimum at the center (antidot) and side minima corresponding to the additional etching bands at the entrance and exit of the ring.

When the splitting is maximum, we may speak of a doubling of the Coulomb oscillation frequency (Fig. 4c). The conductance of sample no. 2 was measured only at T = 0.1 K, but the doubling and splitting of the $G(B)$ peaks at T = 1.3 K was also observed for sample no. 1 (see Fig. 3d and the inset in Fig. 3c). The

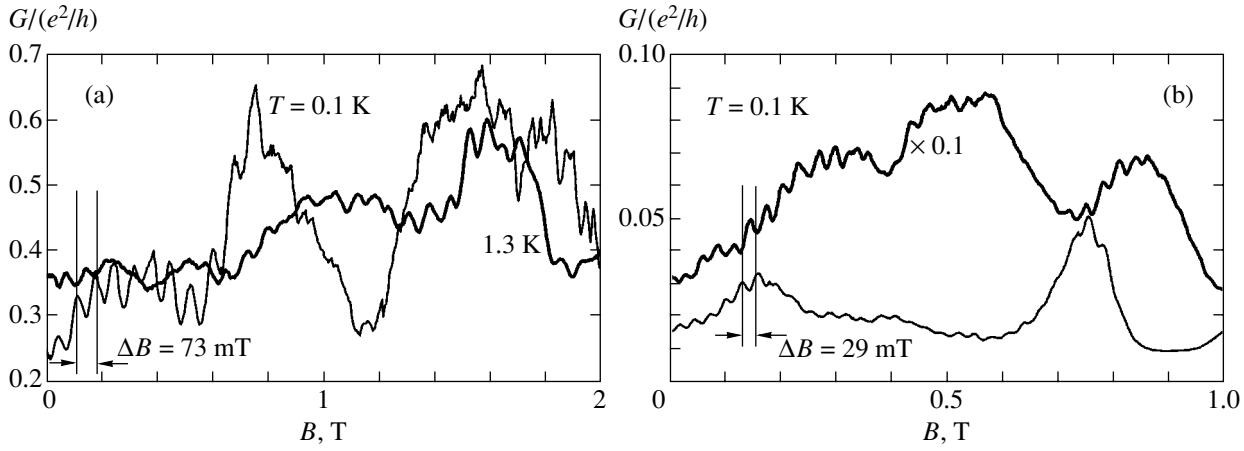

Fig. 2. Low-temperature magnetic field dependences of the conductance $G(B)$ for samples nos. 1 (a) and 2 (b) in the region of small fields. The two curves correspond to different mesoscopic states of the samples. Oscillations with the period ΔB are observed on the background of mesoscopic fluctuations of the conductance.

results of theoretical and experimental investigations of the Coulomb blockade in double quantum dots [23, 26–28], suggest that the observed doublet splitting of the Coulomb peaks is related to the single-electron charging of the two quantum dots at the entrance and exit of the ring.

The Coulomb nature of the small-scale gate-voltage-induced oscillations is confirmed by some additional measurements. In contrast to oscillations of the hopping conductivity in closed systems [36] and the gate-voltage-controlled interference oscillations in the open interferometer [19], the positions of the conductance peaks observed in our experiments were almost independent of the magnetic field (Fig. 5). In a moderate magnetic field of $B \approx 1$ T, the Coulomb peaks are subject to small shifts and their splitting is more pronounced as compared to that in a zero field (Fig. 5b). In addition, the system exhibits random switching of the phase of oscillations in response to sharp changes in the constant polarization charge q_0 of the Coulomb island (to within several tenths of the elementary charge) [24]. This behavior is typical of single-electron transistors.

In concluding the presentation of the experimental part of this study, it should be emphasized that the measurements were performed in the linear regime, which does not provide for an exhausting verification of the Coulomb nature of the observed effects (requiring nonlinear current–voltage characteristics to be measured at various gate voltages in the low-temperature limit). In this sense, we have directly measured neither the Coulomb gap, nor the Coulomb energy E_C, nor the full capacitance C_Σ of the Coulomb island. Such measurements have been recently performed for an interferometer (with dimensions close to those of the device studied in our experiments) characterized by periodic Coulomb peaks without doublets. The energy of the single-electron charging of a ring Coulomb island was 0.2 meV [25]. In another experiment, each of the two quantum dots incorporated into the arms of an interferometer had an energy of $E_C \sim 1.7$ meV [23].

Maintaining the linear regime, we measured a critical temperature ($T_c \approx 8$ K) at which the small-scale periodic oscillations disappear, which yields an estimate of $E_C \approx k_B T_c \approx 0.7$ meV. Actually, the small-scale oscillations are not always registered even at 4.2 K and are observed in a narrower interval of gate voltages than that for the lower temperatures. For example, oscillations in the left-hand part of Fig. 3a are smoothened and retained as the temperature increases from 4.2 to 6 K ($T_c \approx 8$ K). In contrast, the right-hand part of this figure exhibits almost no such oscillations even at 4.2 K (so that $T_c \approx 4$ K). An analogous difference of T_c in the other mesoscopic state of sample no. 1 leads to a strong decrease in amplitude of the Coulomb oscillations at $V_g > -310$ mV at $T = 1.3$ K (Figs. 3b and 3c). These observations indicate that the critical temperature depends on the mesoscopic state of samples and the electron density (both these values being controlled by the gate voltage) and they vary in the course of thermal cycling and under illumination. At the same time, according to the Coulomb blockade theory, the critical temperature in the tunneling regime must depend only on the electric capacitance of the island ($C_\Sigma \approx$ const), that is, $k_B T_c \approx E_C = e^2/2C_\Sigma$ [10]. Therefore, the observed variation of the critical temperature contradicts the assumption on the tunneling insulation of the Coulomb island. Moreover, the background conductance is weakly sensitive to the temperature (Fig. 3a). Thus, the background current does not require thermal activation, which is evidence of the absence of low wide tunneling barriers in the system. The agreement between these estimates and conclusions, on the one hand, and the electrostatic behavior of the device under consideration, on the other hand, was verified as described below.

3. MODEL CALCULATIONS AND DISCUSSION OF RESULTS

3.1. Electrostatic Description

Figure 6 shows the results of modeling of the electrostatic potential and charge density in a ring interferometer. Note that the exact etching pattern of the experimentally studied devices was unknown and the calculations were performed using microscopic data obtained for the control structures. The base pattern was analogous to that depicted in Fig. 1a. Taking into account peculiarities of this nanostructure, we have solved the complete electrostatic problem as described in [20, 35, 37, 38]. The numerical solution of the three-dimensional (3D) Poisson equation was obtained with allowance of the following features: (i) base heterostructure parameters (layer thicknesses, doping levels, and permittivities); (ii) random impurity distribution in the doped layer; (iii) two-level etching depth and the influence of a damaged layer covering the etched regions; and (iv) electron density distribution within the framework of the Thomas–Fermi model with a 2D density of states. The boundary conditions in equilibrium and under illumination assumed the existence of a common Fermi level E_F fixed at the surface, deep centers, and, the substrate. Under the nonequilibrium conditions, we used the frozen charge condition, whereby the charge density on the surface and in the doped layer was independent of the voltages applied to the electron system. The latter condition allowed the device capacity relative to the gate electrode to be determined.

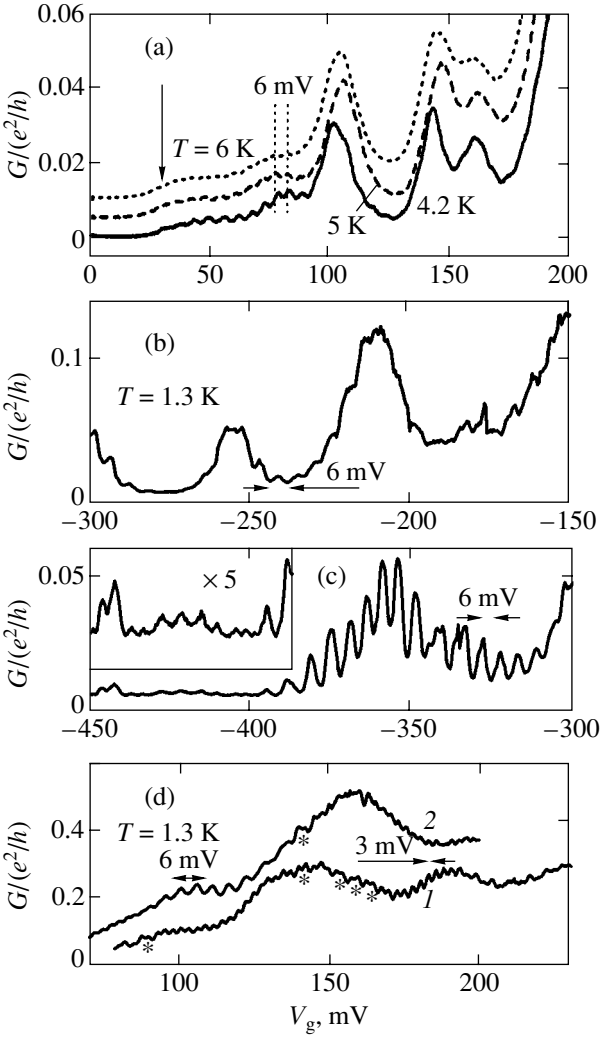

Fig. 3. Plots of the conductance G versus gate voltage V_g for sample no. 1 showing small-scale oscillations with a period of 6 meV on the background of large-scale peaks: (a) the effect of temperature; note a decrease in the critical temperature for $V_g > 90$ mV (the arrow indicates the threshold gate voltage; for the sake of clarity, the curves are shifted upward along the ordinate axis); (b–d) the examples of oscillations in various mesoscopic states of the sample at $T = 1.3$ K; the inset in (c) shows oscillations near the threshold; (d) doubling of the oscillation period and the appearance of doublets (indicated by the asterisks).

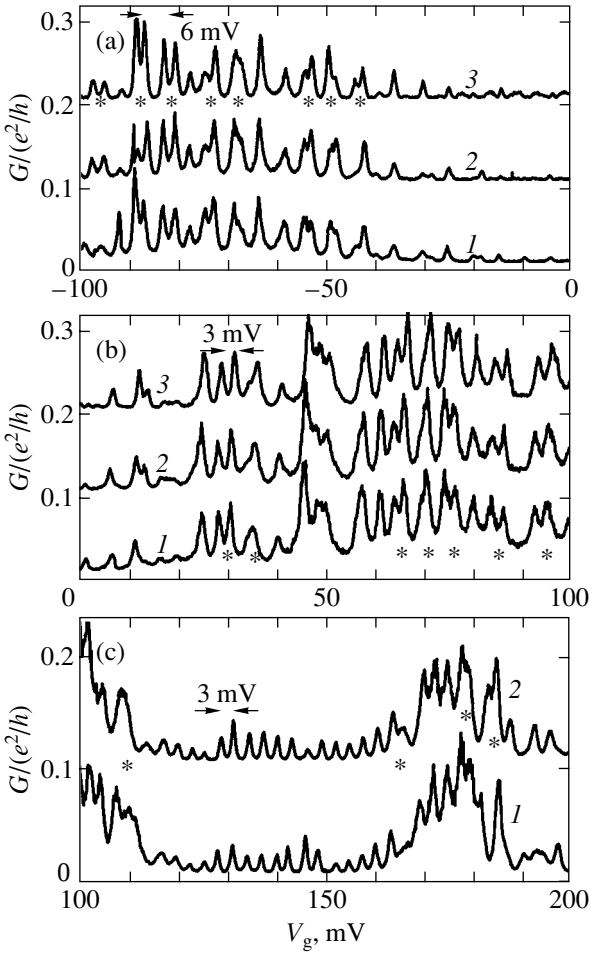

Fig. 4. Plots of the conductance G versus gate voltage V_g for sample no. 2 at low temperatures ($T = 0.1$ K): (a, b) splitting of the Coulomb peaks in sequential measurements (*1–3*) for the same mesoscopic state; (c) doubling of the frequency of oscillations in sequential measurements (*1, 2*) for another mesoscopic state (doublets are indicated by the asterisks).

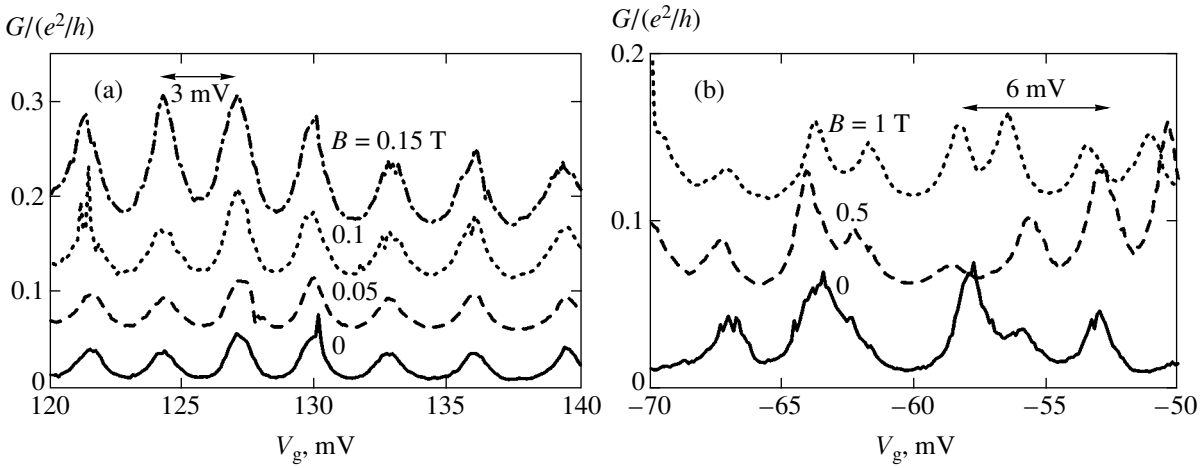

Fig. 5. The effect of (a) weak and (b) moderate magnetic field on the conductance–gate voltage characteristics of sample no. 2 at $T = 0.1$ K (for the sake of clarity, the curves are shifted upward along the ordinate axis).

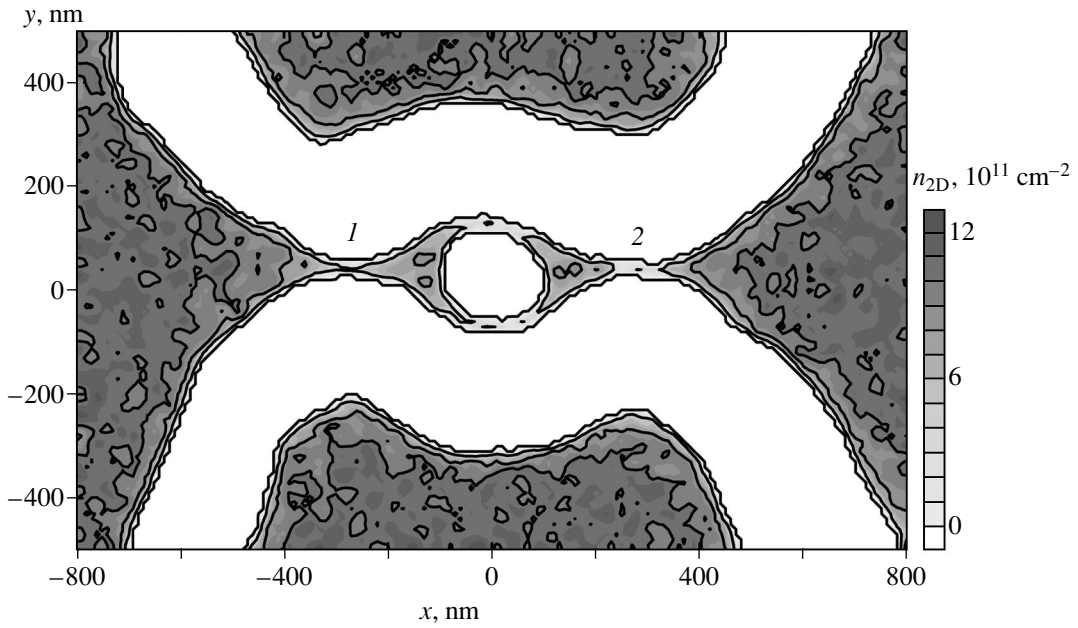

Fig. 6. A calculated map of the electron density n_{2D} versus coordinates x, y in the 2DEG plane ($n_{2Dmax} = 12 \times 10^{11}$ cm^{-2}). Triangular quantum dots *1* and *2* connected by narrow channels form a ring interferometer. The 2DEG regions above and below the ring serve as the in-plane split gate.

As can be seen from Fig. 6, the shape of the etched bands is such that almost the entire area of the electron system of the ring interferometer, except the central antidot, is occupied by triangular lakes. These lakes are connected to each other and to the reservoirs via narrow channels possessing a much smaller area. The distance between ring arms is close to the effective diameter (~260 nm) found for sample no. 1 from the AB oscillation period. The dimensions of triangular quantum dots are close to the ring radius, which makes this device configuration different from that of large ballistic interferometers studied previously [19, 20].

The electron lake area is low-sensitive with respect to details of the internal geometry of the interferometer. For this reason, the capacitance characteristics of the control samples, despite the aforementioned technological variations, must be close to those of real devices. The solution to the 3D electrostatic problem yields the total gate capacitance

$$C_{g1} + C_{g2} = (4.4 \pm 0.2) \times 10^{-17} \text{ F}.$$

Thus, a change in the total interferometer charge by one electron requires a gate voltage increment of 3.65 mV (or 7.3 mV for one-half of the interferometer).

These values coincide with small and large periods of the small-scale conductance oscillations to within 20%, which coincides with the scale of fluctuations in the oscillation period observed in experiment. This coincidence is retained upon variation in the shape of narrowings and the electron density within broad limits. These results confirm the single-electron character of the small-scale oscillations and indicate that a possible reason for the oscillations with a period of 5–7 mV observed up to a temperature of 8 K is charging of one of the triangular quantum dots, rather than of the whole ring. Note that the agreement between theory and experiment in determining the gate capacitance is remarkably good, taking into account that the electrostatic calculations were based only on the restricted geometric data.

For determining the total capacitance of the interferometer and quantum dots, we calculated the capacitance of the electron system of the device with respect to the reservoirs. This was done by introducing a small difference between the Fermi level of the device and that of the reservoirs. An analogous approach was used for determining the capacitance between quantum dots. Note that this procedure is less precise than that of determining the gate capacitance, since the quantum dots are less insulated from each other and the reservoirs. However, in the case of narrow channels, the error is relatively small because the field lines between wide parts of the electron system pass predominantly outside the channels and the 2DEG plane. We have checked that a change in the channel width slightly influences the resulting capacitance. From the standpoint of electrostatics, there is little difference even between the cases of closed and open channels.

The value of the interferometer capacitance with respect to the reservoirs is virtually the same as that relative to the gates. The capacitances of one triangular quantum dot with respect to another (C_{dd}), the adjacent reservoir (C_{dr1}), and each half of the in-plane split gate (C_{dg1}, C_{dg2}) were also found to be close:

$$C_{d\Sigma} = C_{dr1} + C_{dr2} + C_{dd} + C_{dg1} + C_{dg2}$$
$$= (6.3 \pm 1.0) \times 10^{-17}\ \text{F}.$$

It is interesting to note that both capacitance and the size (100 nm) of our triangular quantum dot coincided with the analogous values determined from the data of transport spectroscopy for small quantum dots incorporated into the ring arms [23]. According to the results of calculations, the number of electrons per quantum dot in our system was 80–100, while that found in [23] was as small as 14–16. This difference is consistent with the ratio of densities of the initial 2DEG in the two cases.

3.2. The Coulomb Energy and Subband Spectrum

In the closed regime, a triangular quantum dot of the interferometer would possess a Coulomb energy of $E_C \approx 1.3$ meV. However, judging from the measured maximum temperature at which the single-electron oscillations with a 6-mV period were observed, $k_B T_c \approx 0.7$ meV. The critical temperature T_c further decreases by half outside a certain interval with respect to the gate voltage in the threshold region (Figs. 3a and 3b). The difference between the calculated values of E_C and $k_B T_c$ cannot be fully attributed to the technological variations and calculation uncertainties. The agreement of theory and experiment with respect to capacitances was observed for various semiconductor nanostructures [35, 38]. For example, the results obtained for a quantum dot with controlled transitions from the open to closed regime [33] showed that the capacitance calculated for the closed regime agrees with the measured Coulomb energy, whereas the critical temperature in the open regime is much lower than that estimated from the electrostatic description [35]. The above discrepancy is of a basic nature: the absence of the tunneling insulation of a quantum dot leads to a strong decrease (renormalization) of the single-electron charging energy E_C^* as compared to $e^2/2C_\Sigma$ [29–32, 34, 35], while the charging effect can be retained. An example is offered by a "semiopen" quantum dot in which one input is a tunneling junction and the other, a single-mode microcontact with a transmission coefficient close to unity [30–32]. Before publication of the theoretical results [30], the nonlocal character of the resistance was not taken into account in the Coulomb blockade theory and it was commonly accepted that the microcontact resistance $h/2e^2$ excludes charging of the quantum dot [10–13, 29]. However, the effects predicted in [30] concerning the charging of a semiopen quantum dot and a decrease in the Coulomb energy were confirmed in experiment [31, 32].

As was noted above, an increase in the electron density in our interferometer resulted in that the electron oscillations disappeared at lower temperatures. At the same time, the background conductance and the total electrostatic capacitance did not change significantly. This observation is most probably explained by the fact that the interferometer is not a closed device. This assumption arises by analogy with that adopted in [30–32] and does not contradict the results obtained by modeling the potential and the energy spectrum of our device.

Figure 6 gives a qualitative notion about continuous electron system of the interferometer, in which the channel width is sensitive to variations in the technological parameters. The main reason for a small channel width are the depleted regions (about 40 nm wide), but this contribution is insufficient (at a 2DEG density of 10^{12} cm^{-2}) for driving the model interferometer out of the ballistic regime. For this reason, we have calculated the potential taking into account the information provided by AFM (Figs. 1b and 1c). First, allowance for the 25-nm-deep etching bands between the reservoirs and the ring made the width of the entrance channel

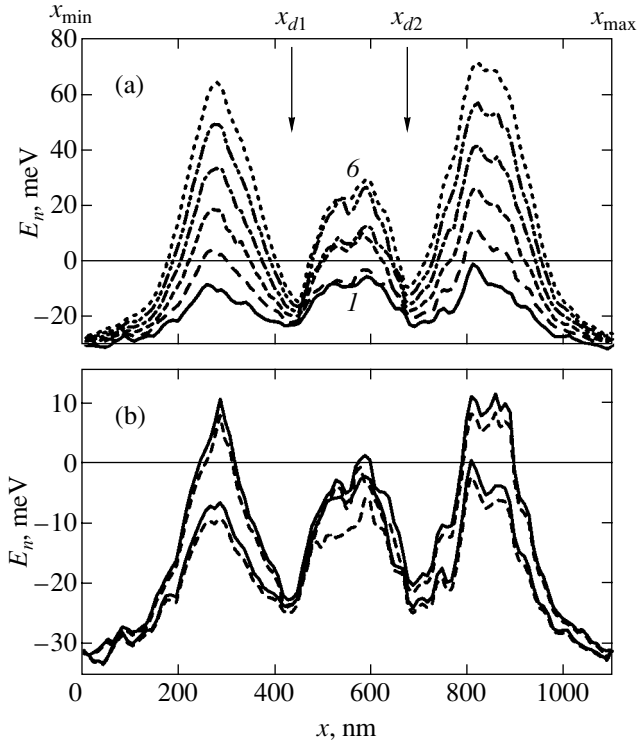

Fig. 7. The energy spectrum of one-dimensional subbands of the interferometer studied: (a) the quantization levels $E_n(x)$ (n = 1 to 6 bottom to top) of the electron motion with respect to the y axis indicated in Fig. 6; the local minima E_n x_{d1} and x_{d2} correspond to triangular quantum dots and the local maxima E_n, to the channels; (b) effect of the gate voltage on the positions of subbands; solid curves $E_{1,2}(x)$ correspond to V_g = 0 and dashed curves, to V_g = 0.2 V applied to one half of the in-plane split gate.

closer to that of the ring. Second, the presence of damaged layers produced by the reactive ion sputtering was taken into account by shifting the vertical insulator–semiconductor junctions 20 nm away from the etching surface (as compared to Fig. 1a). We have checked that further narrowing (by 10 nm) of the gaps between the etching regions breaks the quantum wires and creates nontransparent barriers.

As can be seen from Fig. 6, a shift of the antidot from the center and deviation from the circular shape lead to a certain asymmetry of the ring. In addition, the pattern clearly shows the result of a random distribution of charged impurities in the doped layer above a thin (3 nm) AlGaAs spacer (this is a fluctuational potential, the role of which increases in the narrow channels as a result of reduced screening [37]). The electron density in narrow parts of the ring is much smaller than that in the triangular quantum dots and depends on the charged state of impurities, which also makes our device different from the system studied in [25].

Figure 7 shows the transverse quantization levels E_n (with respect to y) as functions of the x coordinate. The positions of levels were determined by solving a one-dimensional Schrödinger equation with the potential

$$V_i(y) = U(x_i, y, z_0) + E_{z0}(x_i, y),$$

where U is the electrostatic potential determined for the system depicted in Fig. 6, z_0 is the coordinate of the heterojunction of the GaAlAs/GaAs spacer, and E_{z0} is the lower level in the one-dimensional potential well with respect to z. The variable y was restricted to an interval that did not include the split gate regions. Since the device cross section contains two potential wells, the energy levels E_n in the ring are combined in pairs, but diverge significantly (on the order of 10 meV) in the entrance channels. As can be seen, electrons fill only one lowest subband in the narrowest sites of the input quantum wires (the other subbands occur above the Fermi level E_F = 0). In the ring, electrons fill only two lowest subbands corresponding to localization of the wave functions $\psi_{1,2}(y)$ in different arms. At the same time, about ten subbands fall below the Fermi level in the triangular quantum dots. These high subbands form almost nontransparent barriers in the channels. Thus, the quantum dots contain many levels in the vicinity of E_F, which belong to the "closed" subbands and are characterized by large lifetimes [34, 35, 39].

Note that a change in the gate voltage within a typical experimental range of 0.2 V leads to a small shift of the subbands as compared to the intersubband distances in the quantum wires (Fig. 7b). A shift on the same order of magnitude may be caused by a change in the random distribution of charged impurities. Because of technological scatter, the barrier height variation within several millielectronvolts seems to be unavoidable. However, the distances between subbands in the barriers are sufficiently large (10 meV) and stable to provide that the quantum wires would remain single-mode. Thus, the pattern of subbands shows that the electron system of the interferometer separates into two quantum dots connected via single-mode channels to each other and to the reservoirs. These features distinguish the device under consideration from more homogeneous rings of the closed [25] and ballistic [20] interferometers.

3.3. Electron Transmission

In this section, it will be shown that, even with the single-mode microcontacts of the interferometer open, the total conductance of the device remains low because of multiple reflections of an electron in a four-barrier potential of complicated geometry. For the same reason, the electron lifetime on the levels of closed subbands can be quite large, which is important for quantum dot charging. In calculating the device transmission function, we use a model of the coherent multimode transmission. This model, not taking into account the charging effects, cannot be formally used for

describing the conductance oscillations. However, it is still possible, based on the pattern of resonances in the transmission coefficient, to determine the average distance between quasilevels in the quantum dots and in the ring, to assess the influence of the device geometry on the conductance, and to evaluate the role of the nonlocal character of the resistance.

The multimode transmission coefficient T was determined by solving a two-dimensional Schrödinger equation [20, 38] with the effective potential

$$V_{\text{eff}}(x, y) = U(x, y, z_0) + E_{z0}(x, y),$$

corresponding to the electron density distribution depicted in Fig. 6 and the spectrum of one-dimensional subbands ion Fig. 7a.

Figure 8 presents the electron transmission coefficient as a function of the incident electron energy for the whole device and separate parts, including the channels connecting the ring to the reservoirs and the quantum dots. The minima of the subbands corresponding to the left and right quantum dots are denoted by x_{d1} and x_{d2}, respectively (Fig. 7a). The values of $x_{\min}$ and $x_{\max}$ correspond to the left and right boundaries of the interval for which the electrostatic potential and the transverse subbands have been calculated. In this case, the interval $x_{\min} < x < x_{d1}$ corresponds to the left microcontact; $x_{\min} < x < x_{d2}$, to the left quantum dot (Fig. 8, curve *1*); $x_{d2} < x < x_{\max}$, to the right microcontact; and $x_{d1} < x < x_{\max}$, to the right quantum dot (curve *2*). In calculating the transmission of the interferometer parts, the potential outside the given x interval was continued by the boundary values.

The calculation was performed for 20 subbands (adding additional higher subbands does not influence the results). The calculated transmission curves for the right and left entrance channels are presented in Fig. 8a. As can be seen, the transmission function exhibits quantization and the levels for opening of the left and right microcontacts correspond to different energies. For $E \approx E_F = 0$, the transmission of the left microcontact (L) is close to unity, while that of the right channel (R), to 1/2. The difference between transmissions of the quantum dots is determined by microcontacts at the entrance and exit of the microinterferometer. For this reason, the right quantum dot is less transparent with respect to the energy than the left quantum dot (Fig. 8b and 8c). The situation remains qualitatively the same when a voltage of +0.2 V is applied to one-half of the in-plane split gate, but the thresholds for the opening of the microcontacts shift by approximately 2 meV toward lower energies, which corresponds to a decrease in the potential barrier height in Fig. 7b.

The results of modeling of the transmission of the whole interferometer and the individual quantum dots (Fig. 8b) give the pattern of resonance scattering on the quasilevels, that is, information on the energy spectrum. For the interferometer under consideration, the number of resonances (30) and the characteristic energy range (from –6 meV to 0) agrees to a good accuracy with the number of quasiclassical states in this interval (34.5) determined from the electrostatic description. The difference can be explained by the nonrigorous character of the quasiclassical approach, nonuniform level distribution, and the conditional character of the interface between electron systems of the interferometer and the reservoirs. The density of single-particle states in the interferometer and in the quantum dots can be estimated as 10 and 5 per meV (that is, the average distance between levels is smaller than the

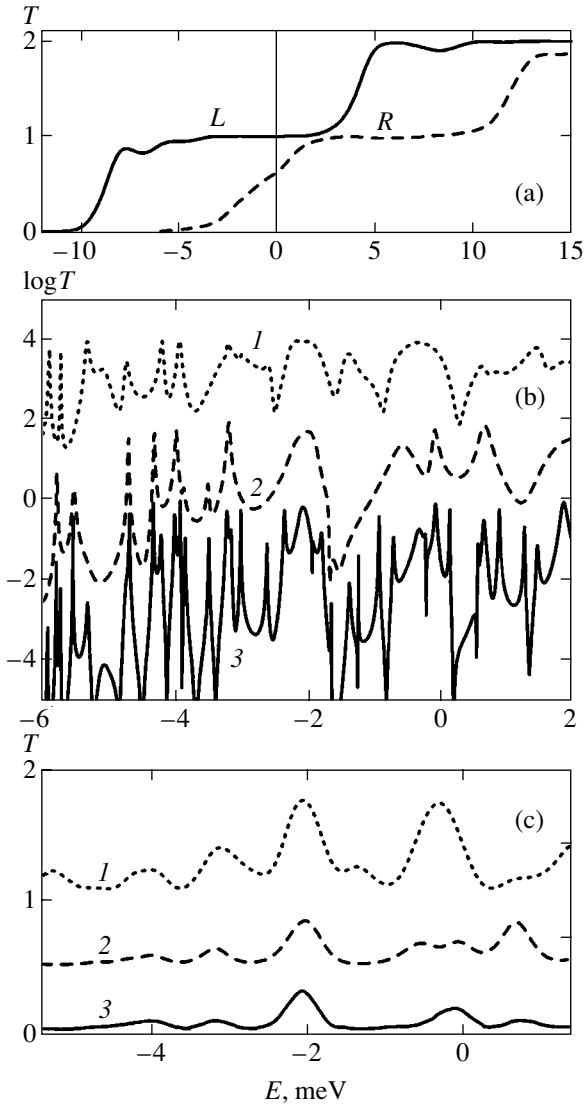

Fig. 8. The electron transmission coefficient of the interferometer versus the incident particle energy: (a) left (L) and right (R) entrance channels; (b, c) triangle quantum dots (*1*, *2*) and the whole device (*3*); (c) the result of averaging upon thermal smearing of the Fermi distribution for T = 1.3 K; for the sake of clarity, the curves in (b, c) are shifted upward along the ordinate axis.

Coulomb energy. Since the critical temperature for observation of the small-scale gate-voltage-controlled Coulomb oscillations exceeds this distance, the oscillations cannot be explained in terms of the resonance transmission through these levels. On the other hand, a small distance between the single-particle levels is a necessary condition for treating the quantum dots as structureless Coulomb islands.

The interference in the quantum dots gives rise to both narrow (0.01–0.1 meV) and wide (0.5 meV) resonances. The pattern of small-width resonances in Fig. 8b is evidence of a large electron lifetime 10^{-1}–10^{-10} s on the quasilevels, that is, of a weak mixing between the localized states of the closed subbands and the decay states. Averaging of the transmission coefficient with respect to the incident particle energy with allowance of the thermal smearing of the Fermi distribution leads to complete leveling of the narrow resonances already at $T = 1.3$ K. The characteristic distance between wide resonances is on the order of 1–3 meV. Such resonances are retained even upon averaging (Fig. 8c) and contribute to the observed large-scale mesoscopic oscillations in the conductance (see Figs. 2 and 3).

As can be seen from Fig. 8c, the less transparent right quantum dot determines the low conductance of the whole device, including the main features of the averaged transmission. The averaged transmission of the whole device is smaller than that of the individual quantum dots: $\langle T(E)\rangle \ll 1$. In this situation, the Landauer formula for the ballistic conductance

$$G = \frac{2e^2}{h}\langle T(E_F)\rangle$$

shows that $G \ll e^2/h$ and the quantization of the microcontact transmission (Fig. 8a) is not manifested in the conductance of the interferometer (Fig. 8b, curve *R*). Therefore, a low conductance of the device as such is not indicative of the tunneling regime. Note that the character of the quantum scattering on the quantum dot quasilevels (Fig. 8b) changes only slightly when the microcontact is open, whereby the transmission coefficient becomes higher than 1/2 (Fig. 8a). Therefore, the rate of electron exchange between localized states of the quantum dot and the environment is determined by interference in the whole system rather than by the transmission of individual microcontacts. In our opinion, this is the main reason for electron localization at the quantum dot, related to the nonlocal character of the resistance and leading to the Coulomb blockade regime. Additional insulation of the quantum dots can be provided by any factors leading to the breakage of quantization of the microcontacts transmission. Such factors were apparently present in our experiments, since the conductance of the control quantum wires remained low in a broad range of the gate voltage variation, while the calculated microcontact transmission shows evidence of quantization (Fig. 8a).

Analysis of the system transmission reveals an effect that is significantly dependent on the state of the quantum wires connecting the quantum dots. A comparison of Figs. 7a and 8b shows that, as long as the motion in one of the arms is forbidden in the classical sense ($E < E_2(x = 580\text{ nm}) = -2.6$ meV), each resonance of the interferometers corresponds to a transmission peak for one of the quantum dots and vise versa. However, when the motion in both arms is allowed in the classical sense, the frequency of transmission resonances for the interferometer is greater than that for the quantum dots. A difference in the number of resonances refers to the states of the ring motion not manifested in the transmission of individual quantum dots. The ring motion arises as a result of the interference in a system of four single-mode microcontacts. Analogous narrow resonances with the levels of a ring motion are known for a large open interferometer, where these resonance are also more frequent than wide resonances with the levels of small-size triangular quantum dots [20]. A shift of the ring levels in a magnetic field causes the AB effect. In our case, the ring is small ($r \approx 130$ nm) and the characteristic distance between resonances of the ring motion is 0.6 meV. This distance allows the AB oscillations to be observed at temperatures of up to 4.2 K.

One may expect that a transition from the classically allowed ring motion to tunneling will suppress the AB effect. This assumption can be verified for a simple model of the interferometer as a point junction of one-dimensional wires [6, 40], using the calculated subband position ($E_1(x)$ in the quantum wires and $E_{1,2}(x)$ in the ring) as the one-dimensional potential. Then, the wave function of the electron motion along the wires is determined by numerically solving the Schrödinger equation with the well-known matching conditions at the junction [40]. For $B = 0$, this model gives the same average distance between resonances of the ring motion as that obtained by the multimode calculation. The sensitivity of the interferometer transmission with respect to the magnetic field virtually vanishes when the particle energy decreases 1–2 meV below the maximum of the potential E_2 in the ring. The tunneling barrier width is small (below 100 nm) and the arm with the potential E_1 remains open (as in Fig. 7). This conclusion retains validity when the quantum wire length and the shape of the one-dimensional potential are varied within realistic limits. For this reason, we do not accept the notion of wide tunneling barriers in the samples exhibiting AB oscillations.

To summarize the results of experiments and model calculations, we can ascertain that observation of the AB effect in rings with a resistance of 10^5–10^6 Ω implies that electrons fill all channels without large ($\gtrsim$100 nm) breaks. In addition, a weak temperature sensitivity of the background conductance points to the absence of low wide tunneling barriers in the interferometer. Finally, variations in the critical temperature and a discrepancy between the electrostatic description

and the Coulomb energy exclude the tunneling insulation of triangular quantum dots. At the same time, observation of the Coulomb blockade excludes a multimode character of the connecting quantum wires. The interferometer most probably operates in a semiopen regime, whereby one microcontact at the entrance or exit determines a large resistance of the device, while the other single-mode channels kept electron at least in one of the triangular quantum dots (the right one in Fig. 7). Opening of the microcontacts leads to additional renormalization of the Coulomb energy, but the Coulomb oscillations at low temperatures are still retained.

3.4. Single-Electron Transport

Strictly speaking, adequate description of the charging process in semiopen systems requires the development of new approaches [30]. However, such theories are now available only for devices of the simplest geometry. For this reason, we have numerically modeled the single-electron transport in a ring interferometer using the orthodox theory of the Coulomb blockade [10]. The question as to whether the orthodox theory is applicable to a system with open microcontacts is a subject of discussion, since this model ignores the interference and replaces real barriers by tunneling junctions characterized by electrostatic capacitances and large localized resistances. It is assumed that the charge inside the tunneling junctions is frozen so as to determine the constant polarization charge q_0 of the Coulomb islands. In our case, these conditions are not obeyed, but the main factor—the Coulomb blockade of the system of localized states of the quantum dots—remains valid.

Below we extend the notion of the charging of such systems, developed recently [34, 35] for a quasi-one-dimensional quantum dot, to a ring interferometer. In the case of the interferometer, there are no obvious reasons (such as the absence of intermode mixing in a quasi-one-dimensional quantum dot) for electron localization in closed subbands. However, the necessary localization is provided by the multiple reflections of an electron in a system of four single-mode microcontacts and by the unidentified factors responsible for the large resistance of the control samples of quantum wires.

The localized states of the triangular quantum dots form two islands charged by the mechanism of the Coulomb blockade. The isolation of these Coulomb islands can be described by introducing effective tunneling junctions with the resistances $R > h/e^2$, which account for the weak interaction between the localized states and the moving electron. The characteristic charging time (RC) in this case is determined by an effective capacitance which will be evaluated below. The central role in this estimation is played by a simple modification of the Coulomb blockade theory, which is used to describe occupation of the lower subband in the microcontacts. The charge of ballistic electrons, concentrated in the single-mode channels, forms the variable polarization charges q_{01} and q_{02}. These charges are determined by the self-consistent potential of the quantum wires and depend on the gate voltage and the withdrawal voltage V_w on the quantum wire [34, 35].

A linear expansion of the function $q_0(V_w, V_g)$ in the vicinity of the characteristic values $V_w = 0$, $V_g = V_{g0}$ yields a simple formula,

$$q_0 = \text{const} - C_w V_w + C_{wg}\Delta V_g.$$

This phenomenological relation closes the system of equations of the orthodox Coulomb blockade theory, in which q_0 is considered as a free parameter. Since the relations between charges and voltages in the electrostatics of conductors are linear, substituting the variable quantity q_0 leads to the usual case of $q_0 = \text{const}$, but with modified values of capacitances. For example, with a zero voltage on the island, the number of electrons n in this island can be expressed through the total and gate capacitances as

$$\begin{aligned} ne &= C_\Sigma V_w - C_g V_g - q_0 \\ &= (C_\Sigma + C_w)V_w - (C_g + C_{wg})V_g\text{—const}. \end{aligned}$$

Then, formulas of the orthodox theory yield a renormalized Coulomb energy,

$$E_C^* = e^2/2(C_\Sigma + C_w),$$

where the additive C_w to the total capacitance summarizes the corresponding corrections to all contributions due to the effective tunneling junctions.

A comparison of theory and experiment shows that the correction to the gate capacitance is small, whereas the additives to other capacitances are large. Indeed, when V_g changes by 200 mV, the barrier heights U_b in the quantum wires vary only by $\Delta U_b \approx 2$ meV. Denoting the distance between the delocalized states by ΔE_d, we can estimate the change in the charge of the system of delocalized states as $(2e/\Delta E_d)(U_b/e\Delta V_g)e\Delta V_g$, where $2e$ is the maximum charge on a delocalized state. This yields

$$C_{wg}\,[\text{aF}] \approx 3.2/\Delta E_d\,[\text{meV}].$$

As can be seen from Fig. 8c, $\Delta E_d > 1$ meV and, hence, C_{wg} is about ten times smaller than the gate capacitance C_g of the quantum dot. In turn, $C_w V_w$ can be estimated as $(2e/\Delta E_d)(eV_w/2)$, where $eV_w/2$ is a change in the barrier height in the quantum wire upon application of the

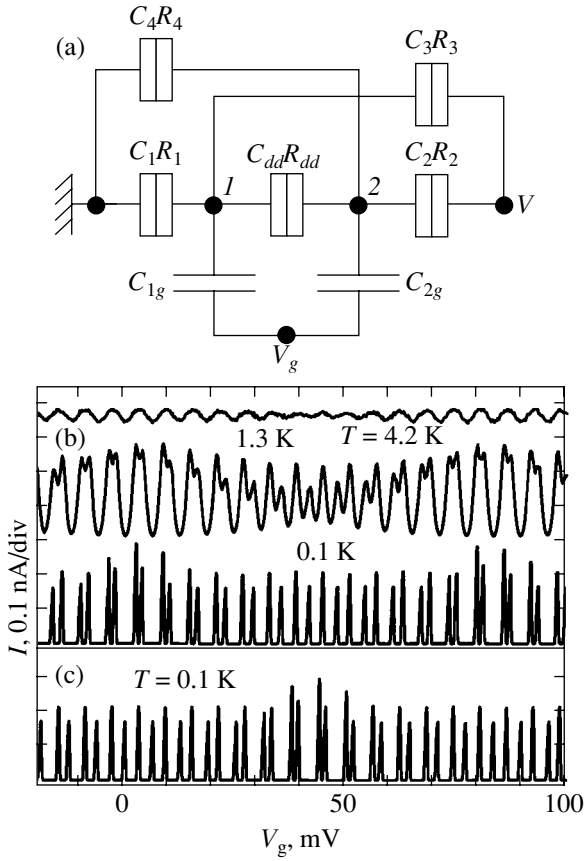

Fig. 9. The results of modeling of the single-electron charging of quantum dots in a ring interferometer: (a) equivalent scheme (C_3R_3 and C_4R_4 junctions describe the coupling of quantum dots *1* and *2* to the far reservoirs); (b, c) plots of the current $I(V_g)$ calculated for the parameters C_{1g} = 25 aF, C_{2g} = 27 aF, C_{dd} = 30 aF, $C_1 = C_2$ = 50 aF, $C_3 = C_4$ = 10 aF, R_{dd} = 3 MΩ, R_1 = 0.1 MΩ, R_2 = 0.4 MΩ, R_3 = 0.2 MΩ, R_4 = 0.8 MΩ, V = 0.25mV; (b) the pattern of doublet splitting of the Coulomb peaks and variation of the splitting magnitude for $q_{01} = q_{02} = 0$ (for the sake of clarity, the curve for T = 0.1 K is shifted downward); (c) A change in the phase of the variation of the Coulomb peak splitting in response to a change in q_{01} by $e/2$.

voltage V_w relative to the environment. From this we obtain

$$C_w\,[\text{aF}] \approx 160/\Delta E_d\,[\text{meV}],$$

which is a large correction to C_Σ. For realistic values of $\Delta E_d \sim$ 1–3 meV, the Coulomb energy decreases by a factor of 4–2 as compared to the case of the tunneling insulation of a triangular quantum dot. In practice, the density of the localized states is determined is determined by the shape of the electrostatic potential and the value of ΔE_d sharply changes upon overloading, illumination, or a change in the gate voltage. This leads to different renormalization of the Coulomb energy and to variation of the critical temperature for observing the Coulomb oscillations.

Below we will take into account the renormalization of the Coulomb energy by increasing the effective capacitance between the quantum dots and between each dot and the reservoirs by a factor of 1.5–3 as compared to the electrostatic value. This increase corresponds to the experimental threshold temperature of 8 K and allows the charging time to be estimated as $RC > 10^{-12}$ s, which agrees with the lifetime (10^{-11} s) of the localized states. The resistances of effective tunneling transitions are assumed to be constant, although real values exhibit mesoscopic fluctuations with respect to the gate voltage. Here, we model only the current charging the capacitances and ignore the ballistic current that increases the background conductance and levels the single-electron features [34, 35]. In the case of low temperatures and the Coulomb blockade, the system may also feature an inelastic cotunneling current that does not influence the charge of the quantum dots [41]. Therefore, our model overstates the sharpness and reproducibility of the shape of the Coulomb blockade features as compared to those observed in experiment. Within the framework of this simple Coulomb blockade model, we attempt to reconstruct the most important features of the gate-voltage-controlled oscillations observed in two samples with different characteristic temperatures. Of special importance is the case of sample no. 2, in which splitting of the Coulomb peaks is clearly observed at T = 0.1 K. It would be of interest to trace the transition from this case to the smoothened periodic oscillations observed for sample no. 1 at T = 4.2 K without changing the parameters of the equivalent scheme.

For this purpose, we have modeled a system of two interacting Coulomb islands, representing triangular quantum dots (Fig. 9a). A pair of parallel interdot microcontacts in our model was replaced by a single tunneling junction. Figures 9b and 9c present the results of calculations for T = 0.1 K. The current was calculated by the Monte Carlo method, which is a standard approach to a system with several Coulomb islands [42, 43]. Since an electron from a localized state in the quantum dot can pass to the far reservoir without influencing the charge of the intermediate quantum dot, we introduced the additional shunting junctions (C_3R_3, C_4R_4 in Fig. 9a) into the effective scheme. This possibility can be provided by three mechanisms, including the inelastic and elastic cotunneling [41, 43] and the passage without loss of the phase coherence. A symmetric system describes splitting of the conductance peaks even with neglect of the effective tunneling junctions shunting the intermediate quantum dots [26, 27]. However, it was established that, for the real system parameters and low temperatures (T = 0.1 K), the result is unstable with respect to violation of the equality of gate capacitances and residual polarization charges, q_{01} and q_{02}, of the two quantum dots. For example, a 10% difference in the gate capacitances leaves the system practically blocked: the charging current becomes

small to provide for the observation of most Coulomb peaks because of a large-period amplitude modulation [26, 27]. This prediction of the theory contradicts our experiment, but introduction of the shunting junctions improves the situation. The resistances of all junctions, selected within 10^5–10^6 Ω, could be varied without significantly influencing the pattern of peak splitting. The capacitances of the quantum dots (except for the gate and total values) could be also taken rather arbitrarily.

Let the gate capacitances of the quantum dots differ by 10%. As can be seen from Fig. 9b, the main effects predicted by the calculation (the presence of doublets and the variable peak spacing) are qualitatively similar to the experimental features (see Fig. 4). At the same time, the calculated peaks are more pronounced, homogeneous, and regular in repetition. It is interesting to note that the magnitude of splitting in the calculation continuously varies within 1–3 mV. Previously, the changes in this value were explained by variations of the interdot capacitance or coupling [26–28]. In our case, the reason is different because neither capacitance nor coupling depends on the gate voltage. Nevertheless, as the gate voltage increases, an initially small splitting exhibits an increase and then decreases again. The analogous, although not as regular, pattern was observed in our experiments (see Fig. 4c). The explanation of the behavior observed in Fig. 9b consists in taking into account the coupling of quantum dots to the far reservoirs. Excluding the interaction between the two quantum dots, we obtain two system of peaks with different periods, which exhibit several coincidences. If the quantum dots interact, the peaks cannot approach each other and a minimum splitting (1.5 mV) can be considered as analogous to the anti-cross-section of levels. In the case of identical gate capacitances, the splitting is constant, but its magnitude strongly depends on the polarization charges. The splitting is minimum in the case of equal residual polarization charges of the two quantum dots, $q_{01} = q_{02}$, and it is maximum (3 mV) in case when $q_{01} = q_{02} \pm e/2$; that is, the positions of the two systems of equidistant peaks are arbitrary to within the difference of the free parameters $q_{01} - q_{02}$. The maximum difference of the gate capacitances of the two quantum dots provides for a coupling of the systems of equidistant peaks to within the phase of oscillations in the splitting magnitude (which depends on $q_{01} - q_{02}$, see Figs. 9b and 9c). In the calculation, the residual polarization charges of the quantum dots (more precisely, corrections to the gate capacitances C_{wg} related to the filling of quantum wires) were independent of the gate voltage, which is an idealization as compared to the situation depicted in Fig. 4.

The results of modeling show that, when the temperature increases to 1.3 K, the minimum splitting is difficult to determine because the peaks exhibit merging. However, it is still possible to observe alternation of the regions featuring oscillations with the main and double frequency and the transition doublets, as is shown in Fig. 3d for sample no. 1. The calculation performed using this model for $T = 4.2$ K explains the oscillations with a period of 6 mV observed in Fig. 3a and demonstrates the absence of oscillations at a double frequency. We have established that the oscillations are manifested to the same extent as in the model of one island with a total capacitance equal to that of the triangular quantum dot. At $T = 8$ K, the oscillations completely disappear. If the capacitance of the effective tunneling junctions is doubled, the oscillations disappear at $T = 4$ K, while the low-temperature splitting of the peaks is retained. This numerical verification confirms the hypothesis of the Coulomb (single-electron charging) interaction between the quantum dots. Thus, the main properties of the Coulomb oscillations are explained by the charging of either one or two triangular quantum dots in the ring interferometer.

4. CONCLUSIONS

We have studied a semiconductor ring interferometer with a small ring radius ($r \approx 0.1$ μm) and a large resistance ($R \gg h/e^2$). The conductance of such devices shows manifestations of both the quantum interference and the Coulomb blockade. It was found that the Coulomb peaks exhibit splitting into doublets in the region of very low temperatures.

The results of structural investigations using scanning electron microscopy and atomic force microscopy provided initial data for the numerical modeling of electrostatics, energy spectrum, and single-electron transport in the device under consideration. It was established that the interferometers represent a system of two triangular quantum dots connected via single-mode microcontacts. The main properties of the experimentally observed gate-voltage-controlled oscillations of the conductance are explained by charging of the quantum dots: the main measured oscillation period (about 6 mV) corresponds to adding a single electron to one quantum dot, while the doublet splitting of the Coulomb peaks (about 1.5 mV) reflects a Coulomb interaction between two charged quantum dots.

We have demonstrated that systems of localized states of the quantum dots are retained on the background of the continuous spectrum of the open lower subband of the microcontacts. A modified orthodox Coulomb blockade model describes the discrete charging in such systems and renormalization of the Coulomb energy with allowance of the continuous filling of the delocalized electron states. Thus, we have considered for the first time a ring interferometer in which the overbarrier transmission coexists with the single-electron charging of triangular quantum dots.

ACKNOWLEDGMENTS

The authors are grateful to V.P. Migal' and L.V. Nenasheva for kindly providing the sample structures and to

Z.D. Kvon and A.G. Pogosov, and M.V. Entin for stimulating discussions.

This study was supported by the Russian Foundation for Basic Research (project no. 01-02-16892), the Federal Targeted Program "Physics of Solid State Nanostructures" (project no. 01.40.01.09.04), and the "Low-Dimensional Quantum Structures" Program of the Russian Academy of Sciences. One of the authors (V.A.T.) also gratefully acknowledges the support from INTAS, grant 01-0014.

REFERENCES

1. *Mesoscopic Phenomena in Solids*, Ed. by B. L. Altshuler, P. A. Lee, and R. A. Webb (North-Holland, Amsterdam, 1991).
2. *Quantum Coherence in Mesoscopic Systems*, Ed. by B. Kramer (Plenum, New York, 1991).
3. *Mesoscopic Electron Transport*, Ed. by L. Sohn, L. P. Kouwenhoven, and G. Schön (Kluwer, Dordrecht, 1997).
4. Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).
5. D. Yu. Sharvin and Yu. V. Sharvin, Pis'ma Zh. Éksp. Teor. Fiz. **34**, 285 (1981) [JETP Lett. **34**, 272 (1981)].
6. M. Buttiker, Y. Imry, and R. Landauer, Phys. Lett. A **96A**, 365 (1983); M. Buttiker, Y. Imry, and M. Ya. Azbel, Phys. Rev. A **30**, 1982 (1984).
7. I. O. Kulik and R. I. Shekhter, Zh. Éksp. Teor. Fiz. **62**, 623 (1975) [Sov. Phys. JETP **41**, 308 (1975)].
8. D. V. Averin and K. K. Likharev, Zh. Éksp. Teor. Fiz. **90**, 733 (1986) [Sov. Phys. JETP **63**, 427 (1986)].
9. L. S. Kuz'min and K. K. Likharev, Pis'ma Zh. Éksp. Teor. Fiz. **45**, 389 (1987) [JETP Lett. **45**, 495 (1987)].
10. K. K. Likharev, IBM J. Res. Dev. **32**, 144 (1988).
11. L. I. Glazman and R. I. Shekhter, J. Phys.: Condens. Matter **1**, 5811 (1989).
12. U. Meirav, M. A. Kastner, and S. J. Wind, Phys. Rev. Lett. **65**, 771 (1990).
13. P. L. McEuen, E. B. Foxman, U. Meirav, *et al.*, Phys. Rev. Lett. **66**, 1926 (1991).
14. L. P. Kouwenhoven, N. C. van der Vaart, A. T. Johnson, *et al.*, Z. Phys. B **85**, 367 (1991).
15. G. Timp, A. M. Chang, J. E. Cunningham, *et al.*, Phys. Rev. Lett. **58**, 2814 (1987).
16. C. J. B. Ford, A. B. Fowler, J. M. Hong, *et al.*, Surf. Sci. **229**, 307 (1990).
17. K. Ismail, S. Washburn, and K. Y. Lee, Appl. Phys. Lett. **59**, 1998 (1991).
18. A. A. Bykov, Z. D. Kvon, E. B. Ol'shanetskiĭ, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **58**, 538 (1993) [JETP Lett. **58**, 543 (1993)].
19. Z. D. Kvon, L. V. Litvin, V. A. Tkachenko, and A. L. Aseev, Usp. Fiz. Nauk **169**, 471 (1999) [Phys.–Usp. **42**, 402 (1999)].
20. O. A. Tkachenko, V. A. Tkachenko, D. G. Baksheev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 366 (2000) [JETP Lett. **71**, 255 (2000)].
21. A. Yakoby, M. Heiblum, H. Shtrikman, *et al.*, Phys. Rev. Lett. **74**, 4047 (1995).
22. R. Shuster, E. Buks, M. Heiblum, *et al.*, Nature **385**, 417 (1997).
23. A. W. Holleitner, C. R. Decker, H. Qin, *et al.*, Phys. Rev. Lett. **87**, 256802 (2001).
24. A. A. Bykov, D. G. Baksheev, L. V. Litvin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 631 (2000) [JETP Lett. **71**, 434 (2000)].
25. A. Fuhrer, S. Lüsher, T. Ihn, *et al.*, Nature **413**, 822 (2001).
26. I. M. Ruzin, V. Chandrasekhar, E. I. Levin, and L. I. Glazman, Phys. Rev. B **45**, 13469 (1992).
27. K. A. Matveev, L. I. Glazman, and H. U. Baranger, Phys. Rev. B **54**, 5637 (1996).
28. F. R. Waugh, M. J. Berry, D. J. Mar, *et al.*, Phys. Rev. Lett. **75**, 705 (1995); F. R. Waugh, M. J. Berry, and C. H. Crouch, Phys. Rev. B **53**, 1413 (1996).
29. K. A. Matveev, Phys. Rev. B **51**, 1743 (1995).
30. I. L. Aleiner and L. I. Glazman, Phys. Rev. B **57**, 9608 (1998).
31. C. Möller, H. Buhmann, S. F. Godijn, and L. M. Molenkamp, Phys. Rev. Lett. **81**, 5197 (1998).
32. S. M. Cronenwett, S. M. Maurer, S. R. Patel, *et al.*, Phys. Rev. Lett. **81**, 5904 (1998).
33. C.-T. Liang, M. Y. Simons, C. G. Smith, *et al.*, Phys. Rev. Lett. **81**, 3507 (1998).
34. V. A. Tkachenko, D. G. Baksheev, O. A. Tkachenko, and C.-T. Liang, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 229 (2001) [JETP Lett. **74**, 209 (2001)].
35. O. A. Tkachenko, V. A. Tkachenko, D. G. Baksheyev, *et al.*, J. Phys.: Condens. Matter **13**, 9515 (2001).
36. A. B. Fowler, in *Physics of Nanostructures*, Ed. by J. H. Davies and A. R. Long (IOP, London, 1991), p. 163.
37. J. A. Nixon, J. H. Davies, and H. U. Baranger, Phys. Rev. B **43**, 12638 (1991).
38. O. A. Tkachenko, V. A. Tkachenko, D. G. Baksheyev, *et al.*, J. Appl. Phys. **89**, 4993 (2001).
39. P. G. Silvestrov and Y. Imry, Phys. Rev. Lett. **85**, 2565 (2000); cond-mat/9903299.
40. Jian-Bai Xia, Phys. Rev. B **45**, 3593 (1992).
41. D. V. Averin and A. A. Odintsov, Phys. Lett. A **140**, 251 (1989).
42. K. K. Likharev, N. S. Bakhvalov, G. S. Kazacha, and S. I. Serdyukova, IEEE Trans. Magn. **25**, 1436 (1989).
43. L. R. C. Fonseca, A. N. Korotkov, K. K. Likharev, and A. A. Odintsov, J. Appl. Phys. **78**, 3238 (1995).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 97, No. 2, 2003, pp. 331–342.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 2, 2003, pp. 367–378.

SOLIDS
Electronic Properties

Exchange and Spin-Fluctuation Mechanisms of Superconductivity in Cuprates

N. M. Plakida[a,*], L. Anton[a,b], S. Adam[a,c], and Gh. Adam[a,c]
[a] *Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia*
**e-mail: plakida@thsun1.jinr.ru*
[b] *Institute of Atomic Physics INFLPR, R-76900, Bucharest, Romania*
[c] *Institute of Physics and Nuclear Engineering, R-76900, Bucharest-Măgurele, Romania*
Received January 20, 2003

Abstract—A microscopic theory of superconductivity is considered in the framework of the Hubbard *p–d* model for the CuO_2 plane. The Dyson equation is derived in the nonintersecting diagram approximation using the projection technique for the matrix Green function of the Hubbard operator. The solution of the equation for the superconducting gap shows that interband transitions for Hubbard subbands lead to antiferromagnetic exchange pairing as in the *t–J* model, while intraband transitions additionally lead to spin-fluctuation pairing of the *d*-wave type. The calculated dependences of the superconducting transition temperature on the hole concentration and of the gap on the wave vector are in qualitative agreement with experiments. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The spin-fluctuation mechanism of high-temperature superconductivity in copper-oxide compounds (cuprates), which is determined by the high intensity of the antiferromagnetic exchange interaction, is attracting considerable attention of researchers at present (see reviews [1, 2]). Anderson [3] was the first to point out a special role of strong electronic correlations in cuprates and the antiferromagnetic exchange interaction associated with these correlations; he also proposed the exchange pairing mechanism in the framework of the *t–J* model. Subsequent investigations of superconductivity based on the *t–J* model were carried out mainly in the mean field approximation (MFA) (see [4, 5] and the literature cited therein).[1] In the low-electron-density limit, superconducting pairing in the **T** matrix approximation was analyzed in [7]. Corrections to the MFA were considered with the help of the diagrammatic technique for the *t–J* model [8] and using the equations of motion method for Green functions [9]. In the latter publication, where a self-consistently numerical solution of the Dyson equation was obtained in the approximation of nonintersecting diagrams for the mass operator, a non-Fermi-liquid behavior was revealed in the normal phase, while in the superconducting phase the spin-fluctuation *d* wave pairing was obtained in addition to the exchange interaction in the MFA.

Numerical methods were also used by many authors for investigating the *t–J* model and the Hubbard model for finite-size clusters (see reviews [2, 10]); however, the results of these studies are contradictory. For example, a stable *d*-wave pairing in the *t–J* model was detected in [11], while the authors of [12] did not observe any long-range superconducting correlations in the initial Hubbard model.

In this connection, it is expedient to investigate the Hubbard model taking into account self-energy corrections without reducing the Hubbard model to the *t–J* model. In the latter model, time-dependent interband transitions are eliminated with the help of the Schrieffer–Wolf transformation, leading to instantaneous exchange interaction in Hubbard subbands. This transformation is essentially analogous to the reduction of the electron–phonon model with retarded interaction to the reduced Bardeen–Cooper–Schrieffer (BCS) model with instantaneous interaction in a limited region of electron energies. Consequently, in order to confirm the results obtained in the framework of the *t–J* model, it is important to estimate the retardation effect for interband transitions in the initial Hubbard model.

In this study, we develop a microscopic theory of superconductivity in the framework of the *p–d* model [13] for the CuO_2 plane in the limit of strong correlations. We use the Hubbard operator technique in the method of thermodynamic Green functions, which leads to the Dyson equation [14] that is solved in the approximation of nonintersecting diagrams for the mass operator. In this way, we can for the first time leave the limits of the MFA, which was formerly used for analyzing equations for the superconducting gap (see [15–19]).

It should be noted that the dynamic mean field theory, which has been widely used for calculating the

[1] We will not discuss the spinon–holon theories, which are not substantiated sufficiently in our opinion [6].

electronic structure in systems with strong correlation, cannot be applied directly for an analysis of the *d*-wave type superconductivity. In this case, a generalization of the theory is required; such a generalization was proposed in [20, 21], where the mass operator was calculated in the cluster approximation at four representative **k** points of the Brillouin zone. Using our method, we obtain an equation of the Migdal–Eliashberg type taking into account the total **k** dependence of the mass operator. An analysis of the obtained equations shows that retardation effects are insignificant for interband transitions, which leads to antiferromagnetic exchange pairing of mobile charge carriers in the entire subband both in the MFA and in the *t–J* model, while intraband transitions determine spin-fluctuation pairing in an energy range of the order of exchange energy *J*.

The derivation of the Dyson equation for the effective Hubbard model is carried out in Section 2. The mean field approximation is considered in Section 3. The equation for the gap, taking into account the spin-fluctuation pairing, is derived in Section 4; the solution of this equation is discussed in Section 5.

2. DYSON EQUATION

Let us consider the *p–d* model for the CuO_2 plane in the limit of strong Coulomb correlations at copper sites [13]:

$$H = \sum_{i,\sigma}\{\epsilon_d \tilde{d}^+_{i\sigma}\tilde{d}_{i\sigma} + \epsilon_p c^+_{i\sigma}c_{i\sigma}\} + \sum_{i,j,\sigma} V_{ij}\{\tilde{d}^+_{i\sigma}c_{j\sigma} + \text{H.c.}\}, \quad (1)$$

where operators $\tilde{d}^+_{i\sigma}$ and $c^+_{i\sigma}$ describe the production of single-hole states of the *d* or *p* type in the *i*th cell of a square lattice for the CuO_2 plane with energies ϵ_d and $\epsilon_p = \epsilon_d + \Delta_{pd}$, respectively. In view of the large value of Coulomb energy at copper sites ($U_d \approx 8$ eV) as compared to $\Delta_{pd} \approx 3$ eV, we take into account only singly occupied 3*d* states: $\tilde{d}^+_{i\sigma} = d^+_{i\sigma}(1 - n^d_{i,-\sigma})$. Hybridization of the *p* states of oxygen in the Wannier representation and the 3*d* states of copper is described by parameters $V_{ij} = 2t_{pd}\nu_{ij}$, where coefficients $\nu_0 = \nu_{jj} \approx 0.96$, $\nu_1 = \nu_{j,\, j\pm a_{x/y}} \approx -0.14$, and $\nu_2 = \nu_{j,\, j\pm a_x \pm a_y} \approx -0.02$ [22]. Since hybridization $V_0 \approx 2t_{pd} \approx \Delta_{pd}$ in a cell is much stronger than hybridization between the nearest (proportional to ν_1), second (proportional to ν_2), and next neighbors, $V_0 \gg |V_{i\neq j}|$, we must first diagonolize one- and two-particle states in a cell, introduce the Hubbard operators of transition between these states, and then calculate the matrix elements of the transition between different cells. As a result of application of such cluster perturbation theory, we obtain the following effective Hubbard model, in which the lower Hubbard subband describes one-hole states of the Cu of the *d* type, while the upper Hubbard subband describes two-hole singlet states of the *p–d* type [22]:

$$H = E_1\sum_{i,\sigma} X^{\sigma\sigma}_i + E_2\sum_i X^{22}_i + \sum_{i\neq j,\sigma}\{t^{11}_{ij}X^{\sigma 0}_i X^{0\sigma}_j + t^{22}_{ij}X^{2\sigma}_i X^{\sigma 2}_j + 2\sigma t^{12}_{ij}(X^{2\bar{\sigma}}_i X^{0\sigma}_j + \text{H.c.})\}, \quad (2)$$

where we have introduced the Hubbard operators $X^{nm}_i = |in\rangle\langle im|$ for the four states $n, m = |0\rangle, \langle\sigma|; |2\rangle = |\uparrow\downarrow\rangle$, $\sigma = \pm 1/2$; and $\bar{\sigma} = -\sigma$. The Hubbard operators obey the multiplication rules ($X^{nm}_i X^{kl}_i = \delta_{mk}X^{nl}_i$) and satisfy the completeness condition ($X^{00}_i + X^{\sigma\sigma}_i + X^{\bar{\sigma}\bar{\sigma}}_i + X^{22}_i = 1$), which shows that only one quantum state $|in\rangle$ can be occupied at any site *i*. In Hamiltonian (2), we have introduced energies $E_1 = \tilde{\epsilon}_d - \mu$ and $E_2 = 2E_1 + \Delta$, where $\tilde{\epsilon}_d$ is the renormalized energy of a *d* hole, μ is the chemical potential, and $\Delta \approx \Delta_{pd}$ is the renormalized energy difference for the Hubbard subband, which plays the role of the Coulomb energy *U* in the standard Hubbard model. The parameters of interband transitions for subbands $\alpha, \beta = 1, 2$ are defined by the relation $t^{\alpha\beta}_{i\neq j} = K_{\alpha\beta}V_{ij}$, where $K_{\alpha\beta} \leq 1$ (see [22]) so that the effective subband width $W = 8t_{\text{eff}} \approx t_{pd} \approx \Delta/2$ and Hamiltonian (2) correspond to the Hubbard model in the strong correlation limit. It should be noted that we carry out here the transformation for a simple version of the *p–d* model (1) containing the minimal number of parameters (t_{pd} and Δ_{pd}). The application of the cluster perturbation theory to a more general *p–d* model [13] including the Coulomb interaction U_d and U_p at copper and oxygen sites, the *p–d* interaction U_{pd}, as well as the *p–p* hybridization t_{pp}, leads to an equally effective Hubbard model (2), but with renormalized parameters (see [23–25]).

In order to analyze the quasiparticle spectrum and superconductivity in model (2), we compose the equation for a one-particle matrix Green function [26],

$$\tilde{G}_{ij\sigma}(t-t') = \langle\langle \hat{X}_{i\sigma}(t)\big|\hat{X}^\dagger_{j\sigma}(t')\rangle\rangle = \int_{-\infty}^{+\infty}\frac{d\omega}{2\pi}e^{-i\omega(t-t')}\frac{1}{N}\sum_{\mathbf{q}}e^{i\mathbf{q}\cdot(\mathbf{i}-\mathbf{j})}\tilde{G}_\sigma(\mathbf{q},\omega), \quad (3)$$

where we have introduced the anticommutator Green function of four-component Nambu operators $\hat{X}_{i\sigma}$ and $\hat{X}^\dagger_{i\sigma} = (X^{2\sigma}_i X^{\bar{\sigma}0}_i X^{\bar{\sigma}2}_i X^{0\sigma}_i)$. Green function (3) can be conveniently written in the form of a 2×2 supermatrix consisting of 2×2 matrices for two subbands for the

normal, $\hat{G}_{ij\sigma}(\omega)$, and anomalous, $\hat{F}_{ij\sigma}(\omega)$, components:

$$\hat{G}_{ij\sigma}(\omega) = \begin{pmatrix} \hat{G}_{ij\sigma}(\omega) & \hat{F}_{ij\sigma}(\omega) \\ \hat{F}^{\dagger}_{ij\sigma}(\omega) & -\hat{G}_{ij\sigma}(-\omega) \end{pmatrix}. \quad (4)$$

Let us construct the equation for Green function (3), applying a projection method of the Mori type, which makes it possible to avoid uncontrollable uncoupling for higher order Green functions. This method is described in detail in our previous publications [9, 14]; for this reason, we will describe here only the results of computations. Successive differentiation of Green function (3) with respect to time t and t' followed by separation of linear parts in the obtained equations of motion using a projection method allows us to write the equation for the Fourier component of this function in the form of the Dyson equation:

$$(\tilde{G}_{\sigma}(\mathbf{q}, \omega))^{-1} = (\tilde{G}^{0}_{\sigma}(\mathbf{q}, \omega))^{-1} - \tilde{\Sigma}_{\sigma}(\mathbf{q}, \omega). \quad (5)$$

In the generalized mean field approximation, the zero-order Green function has the form

$$\tilde{G}^{0}_{\sigma}(\mathbf{q}, \omega) = (\omega\tilde{I} - \tilde{E}_{\sigma}(\mathbf{q}))^{-1}\tilde{\chi}, \quad (6)$$

where $\tilde{I}$ is a 4×4 unit matrix. Assuming that the system is in the paramagnetic state ($\langle X_i^{\sigma\sigma}\rangle = \langle X_i^{\bar{\sigma}\bar{\sigma}}\rangle$) and anomalous means in the $\tilde{\chi} = \langle\{\hat{X}_{i\sigma}, \hat{X}^{\dagger}_{i\sigma}\}\rangle$ matrix are equal to zero in the case of d-wave pairing,

$$\langle X_i^{02}\rangle = \langle X_i^{0\downarrow} X_i^{\downarrow 2}\rangle = \langle c_{i\downarrow} c_{i\uparrow}\rangle = 0,$$

we obtain the following representation for this matrix:

$$\tilde{\chi} = \tau_0 \times \begin{pmatrix} \chi_2 & 0 \\ 0 & \chi_1 \end{pmatrix}.$$

Here, τ_0 is a unit matrix, and correlation functions $\chi_2 = \langle X_i^{22} + X_i^{\sigma\sigma}\rangle = n/2$ and $\chi_1 = \langle X_i^{00} + X_i^{\bar{\sigma}\bar{\sigma}}\rangle = 1 - \chi_2$ depend only on the average number of holes n:

$$n = \langle N_i\rangle = \sum_{\sigma}\langle X_i^{\sigma\sigma}\rangle + 2\langle X_i^{22}\rangle. \quad (7)$$

The one-particle excitation spectrum in Green function (6) is defined by the matrix

$$\tilde{E}_{ij\sigma} = \langle\{[\hat{X}_{i\sigma}, H], \hat{X}^{\dagger}_{j\sigma}\}\rangle\tilde{\chi}^{-1}. \quad (8)$$

The mass operator in Dyson equation (5) is determined by the proper part of the many-particle Green function of "irreducible operators" $\hat{Z}^{(ir)}_{i\sigma} = [\hat{X}_{i\sigma}, H] - \sum_l \tilde{E}_{il\sigma}\hat{X}_{l\sigma}$:

$$\tilde{\Sigma}_{\sigma}(\mathbf{q}, \omega) = \tilde{\chi}^{-1}\langle\langle \hat{Z}^{(ir)}_{\mathbf{q}\sigma} \big| \hat{Z}^{(ir)\dagger}_{\mathbf{q}\sigma}\rangle\rangle^{(prop)}_{\omega}\tilde{\chi}^{-1}. \quad (9)$$

Equations (6), (5), and (9) give exact representation for one-particle Green function (3). However, the evaluation of this functions requires the use of some approximations for the many-particle Green function in mass operator (9), describing the contribution from inelastic processes of scattering of one-particle Fermi-like excitations from spin and charge fluctuations in the system.

3. MEAN FIELD APPROXIMATION

Let us first consider the electronic spectrum in the mean field approximation, which can be described by Green function (6) with matrix (8) of one-particle excitations:

$$\tilde{E}_{ij\sigma} = \begin{pmatrix} \hat{\omega}_{ij} & \hat{\Delta}_{ij\sigma} \\ \hat{\Delta}^{\dagger}_{ij\sigma} & -\hat{\omega}_{ji} \end{pmatrix}. \quad (10)$$

Here, matrices $\hat{\omega}_{ij}$ and $\hat{\Delta}_{ij\sigma}$ determine the normal and anomalous components of the complete matrix, respectively. The one-particle excitation spectrum in the normal state was investigated in detail in [22]. Here, we will consider only the results of calculations. Diagonalization of matrix $\hat{\omega}_{ij}$ in the $\mathbf{q}$ representation leads to the following expression for the one-particle excitation spectrum in the normal phase:

$$\Omega_{2,1}(\mathbf{q}) = \frac{1}{2}[\omega_2(\mathbf{q}) + \omega_1(\mathbf{q})] \pm \frac{1}{2}\{[\omega_2(\mathbf{q}) - \omega_1(\mathbf{q})]^2 + 4|W^{21}|^2\}^{1/2}. \quad (11)$$

Here, the excitation spectra for the subbands of singlet and one-hole states in the absence of hybridization are defined by the equations

$$\omega_2(\mathbf{q}) = E_1 + \Delta + a^{22} + \tilde{t}^{22}(\mathbf{q}),$$

$$\omega_1(\mathbf{q}) = E_1 + a^{11} + \tilde{t}^{11}(\mathbf{q}),$$

and their interaction is described by the function

$$W^{21}(\mathbf{q}) = a^{21} + \tilde{t}^{21}(\mathbf{q}) = \frac{\chi_1}{\chi_2}W^{12}(\mathbf{q}).$$

In these equations, coefficients $a^{\alpha\beta}$ determine renormalization of the chemical potential and the effective

parameters of jump $\tilde{t}^{\alpha\beta}(\mathbf{q})$ can be written in the form

$$\tilde{t}^{\alpha\beta}(\mathbf{q}) = \frac{t_{pd}}{N}\sum_{\mathbf{k}}\nu(\mathbf{k})K^{\alpha\beta}(\mathbf{k}-\mathbf{q}), \quad (12)$$

where the renormalization of parameters $t_{ij}^{\alpha\beta}$ of jumps between lattice sites is determined by coefficients $K^{\alpha\beta}(\mathbf{q})$. Explicit expressions for coefficients $a^{\alpha\beta}$ and $K_{ij}^{\alpha\beta}$ are given in Appendix 1. If we take into account only the jumps between the nearest and second neighbors, coefficient $\nu(\mathbf{q})$ has the form

$$\nu(\mathbf{q}) = 2\sum_{j\neq 0}\nu_{0j}e^{-i\mathbf{q}\cdot\mathbf{j}} = 8\nu_1\gamma(\mathbf{q}) + 8\nu_2\gamma'(\mathbf{q}), \quad (13)$$

where $\gamma(\mathbf{q}) = (\cos q_x + \cos q_y)/2$ and $\gamma'(\mathbf{q}) = \cos q_x \cos q_y$.

It was proved in [22] that, in the case of half-filling ($n = 1$), the system is in the dielectric state with a forbidden gap width on the order of Δ; in the case of *p*-type doping ($n = 1 + \delta > 1$), the Fermi level falls into the upper Hubbard subband of singlet states. In this case, the dispersion of one-particle excitations and the weight of Hubbard subbands are determined by doping to a considerable extent. For a low hole concentration, jumps between nearest neighbors are suppressed due to strong antiferromagnetic correlations and the spectrum is determined by jumps between second neighbors with dispersion $\gamma'(\mathbf{q})$ in formula (13). For a high concentration of holes ($n \geq 1$, 2), antiferromagnetic correlations become insignificant and the spectrum is mainly determined by jumps between nearest neighbors with dispersion $\gamma(\mathbf{q})$ in formula (13).

Let us consider the anomalous component $\hat{\Delta}_{ij\sigma}$ of matrix (10), which determines the superconducting gap. For two subbands, this 2×2 matrix has the form

$$\hat{\Delta}_{ij\sigma} = \delta_{ij}\begin{pmatrix} b_\sigma^{22} & b_\sigma^{21} \\ b_\sigma^{12} & b_\sigma^{11}\end{pmatrix} + (1-\delta_{ij})\begin{pmatrix} \Delta_{ij\sigma}^{22} & \Delta_{ij\sigma}^{21} \\ \Delta_{ij\sigma}^{12} & \Delta_{ij\sigma}^{11}\end{pmatrix}. \quad (14)$$

One-site correlation functions $b_\sigma^{\alpha\beta}$ for *d*-wave pairing are equal to zero (see Appendix 1), while components for sites $i \neq j$ are determined by the correlation functions

$$\begin{aligned}\chi_2\Delta_{ij\sigma}^{22} &= -2\sigma t_{ij}^{21}\langle X_i^{02}N_j\rangle = -\chi_1\Delta_{ij\sigma}^{11}, \\ \chi_2\Delta_{ij\sigma}^{21} &= \frac{1}{2}(t_{ij}^{22}\langle X_i^{02}N_j\rangle + t_{ij}^{11}\langle N_jX_i^{02}\rangle) = -\chi_1\Delta_{ij\sigma}^{12}.\end{aligned} \quad (15)$$

Assuming the rule for multiplying Hubbard operators, we can write anomalous means in these equations in the form

$$\langle X_i^{02}N_j\rangle = \langle X_i^{0\downarrow}X_i^{\downarrow 2}N_j\rangle = \langle c_{i\downarrow}c_{i\uparrow}N_j\rangle,$$

as in [17–19]. Thus, pairing in the MFA is described by the anomalous correlation function for a pair of particles at the same site, but in different Hubbard subbands, $X_i^{02} = X_i^{0\downarrow}X_i^{\downarrow 2}$, and the number of particle operator N_j.

In order to derive an equation for the superconducting gap, we must calculate the correlation function $\langle X_i^{02}N_j\rangle$. In [17, 19], this was done using the method of equations of motion for Green functions (Roth method), in which one-particle operators at the same side corresponded to different times as, for example, in function $\langle c_{i\downarrow}(t)|c_{i\uparrow}(t')N_j(t')\rangle$. This uncoupling procedure for operators at the same site is ambiguous and leads to several solutions determined by the method of uncoupling (this was discovered in [17, 19]). In [18], this method was not used; however, the solution of a self-consistent system of equations also led to ambiguous results.

Here, we calculate the anomalous correlation function $\langle X_i^{02}N_j\rangle$ without resorting to any uncoupling. For this purpose, we consider the equation for the corresponding Green function of the initial operators

$$L_{ij}(t-t') = \langle\langle X_i^{02}(t)|N_j(t')\rangle\rangle.$$

Differentiating the Green function with respect to tine *t*, we obtain for its Fourier component the equation

$$(\omega - E_2)L_{ij}(\omega) \approx 2\delta_{ij}\langle X_i^{02}\rangle$$

$$+\sum_{m\neq i,\sigma}2\sigma t_{im}^{12}\{\langle\langle X_i^{0\bar{\sigma}}X_m^{0\sigma}\big|N_j\rangle\rangle_\omega - \langle\langle X_i^{\sigma 2}X_m^{\bar{\sigma}2}\big|N_j\rangle\rangle_\omega\},$$

where we have omitted on the right-hand side the contributions corresponding to intraband transitions since these transitions lead to only a small correction to the pair production energy at a site:

$$|E_2| \approx \Delta \gg |t_{ij}^{\alpha\alpha}|.$$

Using the spectral representation, we obtain the following expression for the correlation function with $i \neq j$:

$$\langle X_i^{02}N_j\rangle = \int_{-\infty}^{+\infty}\frac{d\omega}{1-\exp(-\omega/T)}\sum_{m\neq i,\sigma}2\sigma t_{im}^{12}$$

$$\times\left\{-\frac{1}{\pi}\mathrm{Im}\left[\frac{1}{\omega_2 - E_2 + i\varepsilon}(\langle\langle X_i^{0\bar{\sigma}}X_m^{0\sigma}\big|N_j\rangle\rangle_{\omega+i\varepsilon}\right.\right.$$

$$\left.\left. - \langle\langle X_i^{\sigma 2}X_m^{\bar{\sigma}2}\big|N_j\rangle\rangle_{\omega+i\varepsilon})\right]\right\}.$$

Depending on the position of the chemical potential in the upper (singlet) or lower (single-hole) subband, the main contribution to the integral on the right-hand side

of this equation comes from the corresponding many-particle Green function. In both cases, the contribution from the pole at $\omega = E_2$ is exponentially small (on the order of $\exp(-\Delta/T) \ll 1$).

Let us now consider the case of *p*-type doping ($n > 1$), when the Fermi level lies in the upper subband, $\mu \approx \Delta$, and one-site energies $E_2 \approx E_1 \approx -\Delta$. In this case, the contribution from the lower subband, which is proportional to

$$-\frac{1}{\pi}\mathrm{Im}\langle\langle X_i^{0\bar{\sigma}} X_m^{0\sigma} | N_j\rangle\rangle_{\omega + i\varepsilon} \approx \delta_{mj}\langle X_i^{0\bar{\sigma}} X_j^{0\sigma}\rangle\delta(\omega - 2E_1),$$

makes an exponentially small contribution on the order of $\exp(-2\Delta/T) \ll 1$. As a result, we obtain the following estimate for the anomalous correlation function:

$$\langle X_i^{02} N_j\rangle \approx -\frac{1}{\Delta}\sum_{m \neq i,\,\sigma} 2\sigma t_{im}^{12}\langle X_i^{\sigma 2} X_m^{\bar{\sigma}2} N_j\rangle. \qquad (16)$$

Here, we evaluated the integral disregarding the retardation effects, i.e., omitting the frequency dependence in the denominator $1/(\omega - E_2)$ since the excitation energy $|E_2| \approx \Delta$ is much higher than the characteristic excitation energies in the singlet subband on the order of $|t_{ij}^{22}|$. Thus, direct evaluation of the anomalous correlation function shows that retardation effects can be neglected in the case of the exchange interaction associated with an interband transition.

Using the two-site approximation in formula (16), $m = j$, which is normally employed in the derivation of the *t–J* model, we obtain the representation

$$\langle X_i^{02} N_j\rangle = \langle c_{i\downarrow} c_{i\uparrow} N_j\rangle = -\frac{4t_{ij}^{12}}{\Delta} 2\sigma\langle X_i^{\sigma 2} X_j^{\bar{\sigma}2}\rangle,$$

where the relation $X_j^{\bar{\sigma}2} N_j = 2X_j^{\bar{\sigma}2}$ is taken into account. In this approximation, for a gap in the singlet subband, we obtain, in accordance with Eq. (14), the expression

$$\Delta_{ij\sigma}^{22} = J_{ij}\langle X_i^{\sigma 2} X_j^{\bar{\sigma}2}\rangle\chi_2^{-1}, \qquad (17)$$

corresponding to the conventional expression for the gap in the *t–J* model with exchange interaction $J_{ij} = 4(t_{ij}^{12})^2/\Delta$.

Thus, sequentially applying the Hubbard operator technique, we have proved that the anomalous correlation function of the type $\langle c_{i\downarrow} c_{i\uparrow} N_j\rangle$ in the MFA for the Hubbard model are connected unambiguously with anomalous means for a pair of electrons (holes) at neighboring sites and that pairing is ensured by the standard exchange interaction as in the *t–J* model (see, for example, [4, 9]). Consequently, the statement made in [19] concerning a special role of interstitial excitations in the compound operator method, which is based on application of uncoupling for Hubbard operators at the same site,

$$\langle X_i^{02} N_j\rangle = \langle X_i^{0\downarrow} X_i^{\downarrow 2} N_j\rangle$$
$$= \langle c_{i\downarrow} c_{i\uparrow} N_j\rangle \to \langle c_{i\downarrow}(t) | c_{i\uparrow}(t') N_j(t')\rangle,$$

proved to be erroneous. In addition, direct computation of the anomalous correlation function in the Hubbard model shows that retardation effects in exchange pairing associated with interband transitions can be disregarded; this makes it possible to substantiate the results obtained in the *t–J* model for instantaneous exchange interaction. A more general formula (16) derived by us can also be used for a more consistent analysis of the exchange-type pairing in the Hubbard model if three-site terms are taken into account.

4. EQUATION FOR THE SUPERCONDUCTING GAP

Let us consider the mass operator (9), which can be written in the form of a supermatrix like Green function (4):

$$\tilde{\Sigma}_{ij\sigma}(\omega) = \tilde{\chi}^{-1}\begin{pmatrix} \hat{M}_{ij\sigma}(\omega) & \hat{\Phi}_{ij\sigma}(\omega) \\ \hat{\Phi}_{ij\sigma}^{\dagger}(\omega) & -\hat{M}_{ij\bar{\sigma}}(-\omega) \end{pmatrix}\tilde{\chi}^{-1}, \qquad (18)$$

where matrices $\hat{M}$ and $\hat{\Phi}$ denote the normal and anomalous component of the total matrix (see formula (38) in Appendix 2). In accordance with relation (9), these matrices are defined in terms of many-particle Green functions describing the inelastic scattering of one-particle Fermi-like excitations $X_1(t)$ at spin and charge fluctuations; in the general form, these fluctuations are represented by Bose-type operators $B_{1'}(t)$ (see formulas (35)–(37) in Appendix 1). Here, we will calculate the mass operator in the approximation of nonintersecting diagrams, corresponding to the self-consistent Born approximation. This approximation can be described by noncorrelated propagation of Fermi-like and Bose-like excitations and corresponds to the skeleton loop diagram in the diagrammatic language. Consequently, in this approximation, the time-dependent pair correlation function of operators $X_1(t)$ and $B_{1'}(t)$ for sites $(1 \neq 1', 2 \neq 2')$ can be written in the form of the product of correlation functions:

$$\langle B_{1'}(t) X_1(t) B_{2'}(t') X_2(t')\rangle \approx \langle X_1(t) X_2(t')\rangle\langle B_{1'}(t) B_{2'}(t')\rangle.$$

Using the spectral theorems for Green functions, we can write the many-particle Green function in the form

$$\langle\langle B_{1'} X_1 | B_{2'} X_2\rangle\rangle_{\omega} \approx \frac{1}{\pi^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{d\omega_1 d\omega_2}{\omega - \omega_1 - \omega_2} \qquad (19)$$
$$\times N(\omega_1, \omega_2)\mathrm{Im}\langle\langle X_1 | X_2\rangle\rangle_{\omega_1}\mathrm{Im}\langle\langle B_{1'} | B_{2'}\rangle\rangle_{\omega_2},$$

where $N(\omega_1, \omega_2) = (1/2)[\tanh(\omega_1/2T) + \coth(\omega_2/2T)]$. In this approximation, we obtain a self-consistent system of equations for mass operator (18) and for one-particle Green function (4). In order to solve this system, we must also specify the spectral density of boson excitations, which is determined by the retarded commutator Green function $\langle\langle B_{1'}|B_{2'}\rangle\rangle_{\omega_2}$.

In the lowest order in interband hybridization, we can introduce independent Green functions for the subband of singlet and one-hole states. For definiteness, we will consider the case of *p*-type doping ($n > 1$), when the Fermi level lies in the upper subband ($\mu \approx \Delta$). Taking into account the relations obtained in the MFA for the one-particle excitation spectrum (11) in the normal phase and (17) for the gap, we obtain the Green functions for the singlet band in the form of a 2×2 matrix:

$$\hat{G}_\sigma^{22}(\mathbf{q}, \omega) = \chi_2\left\{\omega\hat{\tau}_0 - \left[\Omega_2(\mathbf{q}) + \frac{M_\sigma^{22}(\mathbf{q}, \omega)}{\chi_2}\right]\hat{\tau}_3 - \left[\Delta_\sigma^{22}(\mathbf{q}) + \frac{\Phi_\sigma^{22}(\mathbf{q}, \omega)}{\chi_2}\right]\hat{\tau}_1\right\}^{-1}, \tag{20}$$

where τ_0, τ_3, and τ_1 are the Pauli matrices. In accordance with relation (19), the normal ($M_\sigma^{22}(\mathbf{q}, \omega)$) and anomalous ($\Phi_\sigma^{22}(\mathbf{q}, \omega)$) components of the mass operator for the singlet subband have the form

$$M_\sigma^{22}(\mathbf{q}, \omega) = \frac{1}{N}\sum_{\mathbf{k}}\int_{-\infty}^{+\infty} d\omega_1 K^{(+)}(\omega, \omega_1|\mathbf{k}, \mathbf{q}-\mathbf{k}) \times\left\{-\frac{1}{\pi}\mathrm{Im}[K_{22}^2 G_\sigma^{22}(\mathbf{k}, \omega_1) + K_{12}^2 G_\sigma^{11}(\mathbf{k}, \omega_1)]\right\}, \tag{21}$$

$$\Phi_\sigma^{22}(\mathbf{q}, \omega) = \frac{1}{N}\sum_{\mathbf{k}}\int_{-\infty}^{+\infty} d\omega_1 K^{(-)}(\omega, \omega_1|\mathbf{k}, \mathbf{q}-\mathbf{k}) \times\left\{-\frac{1}{\pi}\mathrm{Im}[K_{22}^2 F_\sigma^{22}(\mathbf{k}, \omega_1) - K_{12}^2 F_\sigma^{11}(\mathbf{k}, \omega_1)]\right\}. \tag{22}$$

The kernel of the integral equation has the same form as in the Eliashberg theory:

$$K^{(\pm)}(\omega, \omega_1|\mathbf{k}, \mathbf{q}-\mathbf{k}) = t_{pd}^2|\nu(\mathbf{k})|^2 \times\int_{-\infty}^{+\infty} d\omega_2 \frac{N(\omega_1, \omega_2)}{\omega - \omega_1 - \omega_2}\left[\frac{1}{\pi}\mathrm{Im}\,\chi_{sc}^{(\pm)}(\mathbf{q}-\mathbf{k}, \omega_2)\right], \tag{23}$$

where the spectral density of boson excitations is determined by the corresponding dynamic susceptibilities for charge and spin fluctuations:

$$\chi_{sc}^{(\pm)}(\mathbf{q}, \omega) = \chi_c(\mathbf{q}, \omega) \pm \chi_c(\mathbf{q}, \omega) = -\left[\langle\langle \mathbf{S}_\mathbf{q}|\mathbf{S}_{-\mathbf{q}}\rangle\rangle_\omega \pm \frac{1}{4}\langle\langle \delta N_\mathbf{q}|\delta N_{-\mathbf{q}}\rangle\rangle_\omega\right]. \tag{24}$$

Analogous expressions can also be obtained for the Green function for the one-hole Hubbard subband $G_\sigma^{11}(\mathbf{k}, \omega)$ (see Appendix 2).

Similar computations carried out in [9] in the *t–J* model for a single subband show that the complete solution of the self-consistent system of equations for mass operators (21) and (22) and Green functions (20) for two subbands involves considerable computational difficulties. For this reason, we will estimate the type of the superconducting transition, confining ourselves to the weak coupling approximation for the mass operator. In this approximation, the kernel of integral equation (23) for excitation energies (ω, ω_1) close to the Fermi energy can be approximated by the function

$$K^{(\pm)}(\omega, \omega_1|\mathbf{k}, \mathbf{q}-\mathbf{k}) \approx -\frac{1}{2}\tanh\left(\frac{\omega_1}{2T}\right)\lambda^{(\pm)}(\mathbf{k}, \mathbf{q}-\mathbf{k}), \tag{25}$$

where $|\omega, \omega_1| \le \omega_s \ll W$ and ω_s is the characteristic energy of bosons responsible for pairing. In this approximation, the effective interaction is determined by the static susceptibility:

$$\lambda^{(\pm)}(\mathbf{k}, \mathbf{q}-\mathbf{k}) = t_{pd}^2|\nu(\mathbf{k})|^2\mathrm{Re}\,\chi_{sc}^{(\pm)}(\mathbf{q}-\mathbf{k}, \omega_2 = 0). \tag{26}$$

The weak coupling approximation (25) can be used for estimating the contributions to mass operator (21), (22) due to the Green functions $G_\sigma^{22}(\mathbf{k}, \omega_1)$ and $F_\sigma^{22}(\mathbf{k}, \omega_1)$ of the singlet subband, in which the excitation energy ω_1 is small. The contributions to the mass operator from the Green functions $G_\sigma^{11}(\mathbf{k}, \omega_1)$ and $F_\sigma^{11}(\mathbf{k}, \omega_1)$ of the one-hole subband for excitation energies $\omega_1 \approx \Delta$ much larger than the singlet subband width turn out to be negligibly small (on the order of $(t_{\mathrm{eff}}/\Delta)^2$) and can be neglected.

Using the approximations adopted for the mass operator of the Green function, we arrive at the equa-

tion for the superconducting gap in a singlet subband of the BCS type:

$$\tilde{\Delta}_\sigma^{22}(\mathbf{q}) = \frac{1}{N}\sum_{\mathbf{k}}[J(\mathbf{k}-\mathbf{q}) - K_{22}^2\lambda^{(-)}(\mathbf{k}, \mathbf{q}-\mathbf{k})] \times \frac{\tilde{\Delta}_\sigma^{22}(\mathbf{k})}{2\mathscr{E}_2(\mathbf{k})}\tanh\frac{\mathscr{E}_2(\mathbf{k})}{2T}, \quad (27)$$

where the superconducting gap is given by $\tilde{\Delta}_\sigma^{22}(\mathbf{q}) = \Delta_\sigma^{22}(\mathbf{q}) + \Phi_\sigma^{22}(\mathbf{q}, \omega \approx 0)/\chi_2$. The energy of quasiparticle excitations in the superconducting phase has the standard form $\mathscr{E}_2(\mathbf{q}) = \varepsilon_2^2(\mathbf{q}) + |\tilde{\Delta}_\sigma^{22}(\mathbf{q})|^2]^{1/2}$, where we have introduced the renormalized energy $\varepsilon_2(\mathbf{q}) \approx \Omega_2(\mathbf{q}) + M_\sigma^{22}(\mathbf{q}, \omega = \varepsilon_2(\mathbf{q}))/\chi_2$ of one-particle excitations, measured from the Fermi energy: $\varepsilon_2(\mathbf{q}_F) = 0$. In Eq. (27), integration with respect to $\mathbf{k}$ for the exchange interaction is carried out over the entire singlet subband, while integration with respect to energy in the second term proportional to $\lambda^{(-)}(\mathbf{k}, \mathbf{q} - \mathbf{k})$ is carried out near the Fermi energy in a layer having a thickness on the order of $\pm\omega_s$.

A similar equation can also be obtained in the case of *n*-type doping ($n = 1 + \delta \le 1$), when the chemical potential lies in the one-hole band ($\mu \approx 0$). In this case, using the weak coupling approximation for the mass operators and the Green function for the one-hole subband, we arrive at the equation for gap $\tilde{\Delta}_\sigma^{11}(\mathbf{q})$ similar to Eq. (27).

5. RESULTS AND DISCUSSION

In order to solve Eq. (27) for the gap, we must use model (26) for static susceptibility. It is sufficient to consider the contribution from spin fluctuations only, the excitation energy of which is much smaller than the energy of charge fluctuations: $\omega_s \approx J \ll \omega_c \approx W$. In this case, interaction (27) can be written in the form $\lambda_s(\mathbf{k}, \mathbf{q} - \mathbf{k}) = t_{pd}^2|\nu(\mathbf{k})|^2\chi_s(\mathbf{q} - \mathbf{k})$, where we can use for static spin susceptibility $\chi_s(\mathbf{q} - \mathbf{k})$ the model proposed in [27] on the basis of numerical calculations:

$$\chi_s(\mathbf{q}) \approx \frac{1}{\omega_s}\langle \mathbf{S}_\mathbf{q}\mathbf{S}_{-\mathbf{q}}\rangle = \frac{\chi_0(\xi)}{1 + \xi^2[1 + \gamma(\mathbf{q})]}. \quad (28)$$

This model is determined by two parameters, viz., the correlation length ξ and the characteristic energy $\omega_s \le J$ of antiferromagnetic spin fluctuations. Susceptibility (28) has a peak for $\mathbf{q} = \mathbf{Q} = (\pi, \pi)$, i.e., for the antiferromagnetic wave vector for which $1 + \gamma(\mathbf{Q}) = 0$. In this case, the value of susceptibility is determined by coefficient $\chi_0(\xi)$, whose value can be found from the normalization condition for spin $S = 1/2$ at a site:

$$\frac{1}{N}\sum_i\langle \mathbf{S}_i\mathbf{S}_{-i}\rangle = \frac{1}{N}\sum_\mathbf{q}\langle \mathbf{S}_\mathbf{q}\mathbf{S}_{-\mathbf{q}}\rangle = \frac{3}{4}(1-\delta).$$

This equations leads to

$$\chi_0(\xi) = \frac{3(1-\delta)}{4\omega_s C(\xi)},$$

where

$$C(\xi) = \frac{1}{N}\sum_\mathbf{q}\{1 + \xi^2[1 + \gamma(\mathbf{q})]\}^{-1}.$$

For large values of χ, we have $\chi_0(\xi) \propto \xi^2/\ln\xi$.

Let us first consider an analytic estimate for the superconducting transition temperature T_c, solving Eq. (27) for the *d*-wave pairing with a gap in the form

$$\tilde{\Delta}_\sigma^{22}(\mathbf{q}) = \Delta_\sigma^d(\cos q_x - \cos q_y) = \Delta_\sigma^d\eta(\mathbf{q}).$$

Multiplying both sides of this equation by $\eta(\mathbf{q})$ and integrating with respect to $\mathbf{q}$, we obtain the following equation for T_c:

$$1 = \frac{1}{N}\sum_\mathbf{k}\frac{1}{2\varepsilon_2(\mathbf{k})}\tanh\frac{\varepsilon_2(\mathbf{k})}{2T_c} \times [J\eta^2(\mathbf{k}) + \lambda_s(4\gamma(\mathbf{k}))^2\eta^2(\mathbf{k})]. \quad (29)$$

Here, while integrating the spin-fluctuation contribution, we have taken into account the fact that the spin susceptibility $\chi_s(\mathbf{q} - \mathbf{k})$ in Eq. (27) has a peak at $\mathbf{q} - \mathbf{k} = \mathbf{Q} = (\pi, \pi)$; this enables us to obtain the estimate

$$\frac{1}{N}\sum_\mathbf{q}\eta(\mathbf{q})\chi_s(\mathbf{q} - \mathbf{k}) \approx -\frac{\eta(\mathbf{k})}{N}\sum_{\mathbf{q}'}\chi_s(\mathbf{q}') = -\frac{\eta(\mathbf{k})3(1-\delta)}{4\omega_s}$$

and to introduce the effective spin-fluctuation interaction constant

$$\lambda_s = \frac{3}{4\omega_s}(2\nu_1 K_{22}t_{pd})^2(1-\delta) \approx \frac{t_{\text{eff}}^2}{\omega_s}.$$

Passing to integration with respect to energy, we can write Eq. (29) in the form

$$1 \approx \int_{-\mu}^{\tilde{W}-\mu}\frac{d\epsilon}{2\epsilon}\tanh\frac{\epsilon}{2T_c} \times [JN_d(\epsilon) + \theta(\omega_s - |\epsilon|)\lambda_s N_{sf}(\epsilon)], \quad (30)$$

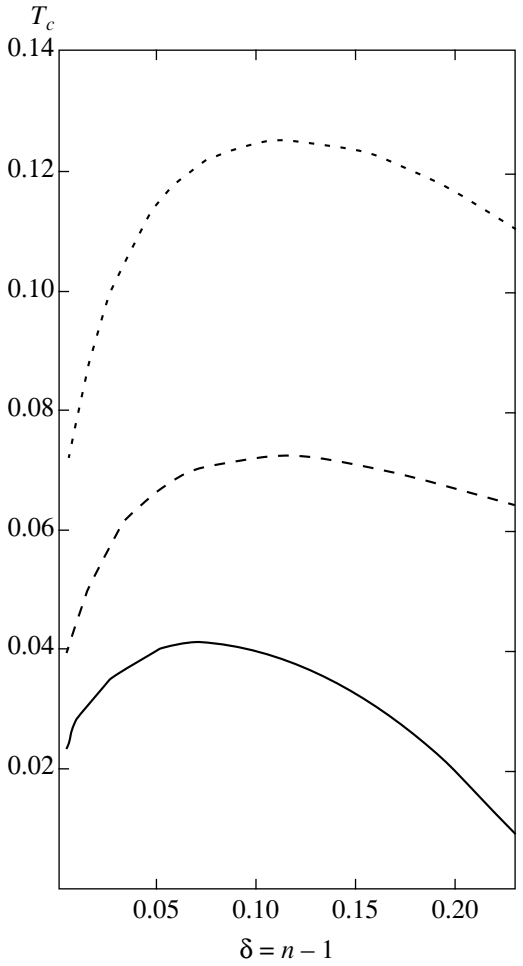

Fig. 1. $T_c(\delta)$ value (in units of t_{eff}) taking into account the spin-fluctuation contribution (solid curve), exchange interaction (dashed curve), and both contributions (dotted curve).

where $\theta(x) = 1$ for $x < 1$ and $\theta(x) = 0$ for $x > 1$. Integration with respect to energy is carried out over the renormalized width $\tilde{W}$ of the singlet subband with the weighted density of states

$$N_d(\epsilon) = \frac{1}{N}\sum_{\mathbf{k}} \eta^2(\mathbf{k})\delta(\epsilon - \varepsilon_2(\mathbf{k}))$$

for the exchange interaction and

$$N_{sf}(\epsilon) = \frac{1}{N}\sum_{\mathbf{k}} \eta^2(\mathbf{k})(4\gamma(\mathbf{k}))^2\delta(\epsilon - \varepsilon_2(\mathbf{k}))$$

for the spin-fluctuation interaction. Both densities of state are normalized to unity:

$$\int_{-\infty}^{+\infty} d\epsilon N_{d,\,sf}(\epsilon) = 1.$$

It should be noted that the main contribution to the density of states for a two-dimensional lattice comes from regions in the vicinity of the Van Hove singularities, $|\mathbf{k}| = (0, \pi), (\pi, 0)$. While the exchange interaction makes a large contribution ($\eta^2(\mathbf{k}) = 4$) in these regions, the contribution of the spin-fluctuation interaction is suppressed ($\gamma^2(\mathbf{k}) = 0$). Consequently, the effective coupling constant $V_{ex} = JN_d(0)$ for the exchange interaction can be larger than the spin-fluctuation constant $V_{sf} = \lambda_s N_{sf}(0)$ even for $\lambda_s \approx t_{\text{eff}}^2/\omega_s > J$. Considering, in addition, that the range of spin-fluctuation pairing ω_s is much smaller than the region in which the exchange interaction $\tilde{W}$ operates, we conclude that the exchange interaction ensures higher superconducting transition temperatures.

Standard estimates in the logarithmic approximation for $T_c \ll \omega_s \ll \mu$ give $T_c^{ex} \approx \sqrt{\mu(\tilde{W} - \mu)}\exp(-1/V_{ex})$ for exchange pairing and $T_c^{sf} \approx \omega_s \exp(-1/V_{sf})$ for spin-fluctuation pairing. It should be noted that both T_c^{ex} and T_c^{sf} vanish for $n \longrightarrow 0$ due to the tendency to zero of effective interactions V_{ex} and V_{sf} proportional to the density of states $N_{d,\,sf}(\epsilon)$; this is in accordance with the results obtained in [7]. When both contributions are taken into account, it is convenient to write the expression for T_c in the form

$$T_c \approx \omega_s \exp\left(-\frac{1}{\tilde{V}_s}\right), \qquad \tilde{V}_s = V_{sf} + \frac{V_{ex}}{1 - V_{ex}\ln(\mu/\omega_s)}, \tag{31}$$

introducing the effective coupling constant $\tilde{V}_s$ for the spin-fluctuation interaction; this constant increases considerably when the exchange contribution is taken into account. Indeed, assuming here for estimates that $\mu = \tilde{W}/2 \approx 0.35$ eV, $\omega_s \approx J \approx 0.13$ eV, and $V_{sf} \approx V_{ex} = 0.2$, we obtain $\tilde{V}_s \approx 0.2 + 0.25 = 0.45$ and $T_c \approx 160$ K, while the spin-fluctuation contribution alone gives $T_c^0 \approx \omega_s \exp(-1/V_{sf}) \approx 10$ K. Similar estimates can be obtained for the case of *n*-type doping $n \leq 1$. In the case of the standard Hubbard model with equal parameters of jumps between sites, we obtain identical $T_c(\delta)$ dependences on the carrier concentration δ in view of electron–hole symmetry. In the effective *p–d* model (2), this symmetry is absent and the $T_c(\delta)$ dependence for the subband of singlet states considered above will differ from that obtained for the subband of one-hole states in the case of *n*-type doping.

These estimates were confirmed by the numerical solution of Eq. (27) for the gap using direct summation

in the **k** space. We sought solutions of the *d*-wave type only, for which the superconducting gap satisfies the condition $\tilde{\Delta}_\sigma^{22}(q_x, q_y) = -\tilde{\Delta}_\sigma^{22}(q_y, q_x)$. Under this condition, the identity $\langle X_i^{\sigma 2} X_i^{\bar{\sigma}2}\rangle = (1/N)\sum_{\mathbf{k}} \langle X_{\mathbf{k}}^{\sigma 2} X_{-\mathbf{k}}^{\bar{\sigma}2}\rangle = 0$ following from the algebra of Hubbard operators is satisfied automatically. The determination of the superconducting transition temperature $T_c(\delta)$ in linearized equation (27) was reduced to the search for maximal eigenvalues of the discrete Fredholm integral equation, and the **k** dependence of the gap was determined by the corresponding eigenfunction of this equation. For the parameters of the *p–d* model, we chose the following values: $\Delta_{pd} = 2t_{pd} = 3$ eV. $t_{\text{eff}} \approx K_{22} 2\nu_1 t_{pd} \approx 0.14 t_{pd} \approx 0.2$ eV. For the exchange interaction, we assumed the standard value $J = 0.4 t_{\text{eff}}$ from the *t–J* model. The parameters in the model of spin susceptibility (28) were such that the antiferromagnetic correlation length for its characteristic value $\xi = 3$ and the energy $\omega_s = 0.15$ eV of spin correlations were assumed to be independent of the hole concentration.

Figure 1 shows the results of numerical solution for the superconducting transition temperature T_c (in units of $t_{\text{eff}} \approx 0.2$ eV) as a function of the hole concentration $\delta = n - 1$ in the singlet subband. The spin-fluctuation interaction gives a value of the superconducting transition temperature equal approximately to one-third of the value corresponding to the inclusion of both contributions. A decrease in correlation length ξ leads to a considerable reduction of the spin-fluctuation contribution to T_c. The maximal temperature $T_c^{\max} \approx 0.12 t_{\text{eff}} \approx 270$ K is attained for optimal doping $\delta_{\text{opt}} \approx 0.13$. It is well known that the weak coupling approximation (25) leads to elevated values of T_c: the complete inclusion of contributions due to the mass operator lowers T_c significantly (see, for example, the calculations made in [9] in the *t–J* model). As $\delta \longrightarrow 0$, T_c also tends to zero according to our calculations, although the possibility of the emergence of antiferromagnetic instability and a pseudogap in the quasiparticle spectrum accompanying this instability [28] must be taken into consideration in this case. The pseudogap also suppresses T_c. Taking into account additionally a decrease in the correlation length ξ with increasing hole concentration, we obtain much lower values of T_c for large δ. On the whole, the $T_c(\delta)$ dependence is in qualitative agreement with experiment and with the results of simulation for the *t–J* model [11]. Figure 2 shows the dependence $\tilde{\Delta}^{22}(\mathbf{k})$ of the superconducting gap on the wave vector **k** in the first quarter of the Brillouin zone ($0 \le k_x, k_y \le 1$ in units of π/a) for $\delta = 0.13$ for three temperatures: $T = 0$ (a), $T = 0.5T_c$ (b), and $T = 0.9T_c$ (c). The numbers and symbols on isolines indicate the gap width and sign (+ or –). The Fermi surface is shown by circles. Calculations reproduce the *d*-wave symmetry of the gap, but with a much more complicated dependence than in the simple

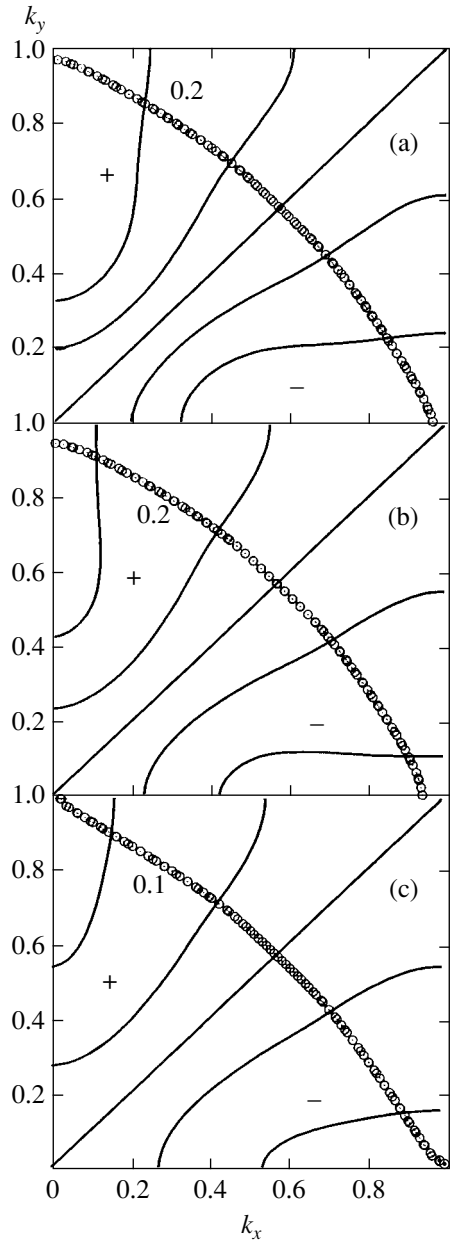

Fig. 2. Dependence of gap width $\tilde{\Delta}^{22}(\mathbf{k})$ on wave vector **k** in the first quarter of the Brillouin zone ($0 \le k_x, k_y \le 1$ in units of π/a) for $\delta = 0.13$ at three temperatures: $T = 0$ (a), $T = 0.5T_c$ (b), and $T = 0.9T_c$ (c). The numbers and symbols indicate the gap width and sign (+ or –). The Fermi surface is shown by circles.

model in the form $\tilde{\Delta}_\sigma^{22}(\mathbf{q}) = \Delta_\sigma^d(\cos q_x - \cos q_y)$. The inclusion of the spin-fluctuation contribution alone gives maximal values of the gap outside the Fermi surface, which lies in the vicinity of the antiferromagnetic Brillouin zone in the case of optimal doping (see Fig. 2): $|k_x| + |k_y| = \pi$. This is due to the weakness of the spin-fluctuation interaction in the vicinity of the antiferromagnetic Brillouin zone, where $\gamma^2(\mathbf{k}) = 0$ in accor-

dance with the remark made by Schrieffer [29] during the discussion of phenomenological models of spin-fluctuation pairing. Consequently, the inclusion of the exchange contribution is significant for attaining high values of T_c in the case of a "large" Fermi surface.

In conclusion, we describe the main results of our calculations in the framework of the *p–d* Hubbard model (2). The exchange interaction associated with interband jumps is the most important for superconducting pairing of the *d*-wave type. Retardation effects for this mechanism are insignificant; consequently, it can be approximated by the instantaneous exchange interaction as in the *t–J* model. Pairing of electrons (holes) in this mechanism is precisely caused by the decrease in the electron kinetic energy in the presence of strong correlations (exclusion of double filling of quantum states) due to interband jumps in a lattice with a short-range antiferromagnetic order. It was noted by Anderson [30] that there is no such mechanism in phenomenological spin-fermion models disregarding strong correlations. The spin-fluctuation pairing associated with scattering from spin fluctuations makes a decisive contribution to the increase in the superconducting transition temperature, but is manifested effectively only for a high intensity of spin fluctuations. The obtained results are confirmed by calculations made in the framework of the *t–J* model [9]. However, the more complicated nature of superconducting correlations in the Hubbard model in the MFA as compared to the *t–J* model (cf. formulas (16) and (17)) requires the inclusion of dynamic processes associated with high-energy interband jump. This apparently explains the discrepancy in the results obtained in numerical calculations for the *t–J* model [11] and for the Hubbard model [12], which take into account only a limited number of excited states.

The calculations made here involve certain approximations. In order to refine the obtained results, a self-consistent computation should be made for Green function (4) and for mass operator (18), as was done in [9] for the *t–J* model. It is also necessary to derive expression (24) for the dynamic spin susceptibility without resorting to the weak coupling approximation (25). Such computations will be carried out in future.

ACKNOWLEDGMENTS

The authors are grateful to P. Horsch, F. Mancini, and V.S. Udovenko for fruitful discussions. One of the authors (N.P.) thanks Prof. P. Fulde for hospitality offered during his visit to the Max-Planck Institute of Complex Systems. S. Adam and Gh. Adam would like to acknowledge partial financial support from the Romanian MER (grant no. 7038GR).

APPENDIX 1

In order to calculate the one-particle excitation matrix (8), we must consider the equations of motion for the Hubbard operators:

$$Z_i^{\sigma 2} = [X_i^{\sigma 2}, H] = (E_1 + \Delta)X_i^{\sigma 2} + \sum_{l \neq i,\, \sigma'} (t_{il}^{22} B_{i\sigma\sigma'}^{22} X_l^{\sigma' 2} - 2\sigma t_{il}^{21} B_{i\sigma\sigma'}^{21} X_l^{0\bar{\sigma}'}) - \sum_{l \neq i} X_i^{02} (t_{il}^{11} X_l^{\sigma 0} + 2\sigma t_{il}^{21} X_l^{2\bar{\sigma}}), \tag{32}$$

$$Z_i^{0\bar{\sigma}} = [X_i^{0\bar{\sigma}}, H] = E_1 X_i^{0\bar{\sigma}} + \sum_{l \neq i,\, \sigma'} (t_{il}^{11} B_{i\sigma\sigma'}^{11} X_l^{0\bar{\sigma}'} - 2\sigma t_{il}^{12} B_{i\sigma\sigma'}^{12} X_l^{\sigma' 2}) - \sum_{l \neq i} X_i^{02} (t_{il}^{22} X_l^{2\bar{\sigma}} + 2\sigma t_{il}^{12} X_l^{\sigma 0}), \tag{33}$$

$$Z_i^{2\bar{\sigma}} = -(Z_i^{\bar{\sigma} 2})^\dagger, \quad Z_i^{\sigma 0} = -(Z_i^{0\sigma})^\dagger. \tag{34}$$

Here, operators $B_{i\sigma\sigma'}^{\alpha\beta}$ describe spin and charge fluctuations:

$$B_{i\sigma\sigma'}^{22} = (X_i^{22} + X_i^{\sigma\sigma})\delta_{\sigma'\sigma} + X_i^{\sigma\bar{\sigma}}\delta_{\sigma'\bar{\sigma}} = \left(\frac{1}{2}N_i + S_i^z\right)\delta_{\sigma'\sigma} + S_i^\sigma \delta_{\sigma'\bar{\sigma}}, \tag{35}$$

$$B_{i\sigma\sigma'}^{21} = \left(\frac{1}{2}N_i + S_i^z\right)\delta_{\sigma'\sigma} - S_i^\sigma \delta_{\sigma'\bar{\sigma}}, \tag{36}$$

$$B_{i\sigma\sigma'}^{11} = \delta_{\sigma'\sigma} - B_{i\sigma\sigma'}^{21}, \quad B_{i\sigma\sigma'}^{12} = \delta_{\sigma'\sigma} - B_{i\sigma\sigma'}^{22}. \tag{37}$$

After carrying out required commutations, we obtain the one-particle excitation matrix in form (10) with the components

$$\chi_2 a^{22} = \sum_{m \neq i} V_{im}(K_{22}\langle X_i^{2\bar{\sigma}} X_m^{\bar{\sigma} 2}\rangle - K_{11}\langle X_m^{\sigma 0} X_i^{0\sigma}\rangle),$$

$$\chi_2 a^{21} = -\sum_{m \neq i} V_{im}(K_{22}\langle X_i^{\sigma 0} X_m^{\bar{\sigma} 2}\rangle + K_{11}\langle X_i^{\bar{\sigma} 2} X_m^{\sigma 0}\rangle) - 2\sigma \sum_{m \neq i} V_{im} K_{12}(\langle X_i^{\sigma 0} X_m^{0\sigma}\rangle - \langle X_m^{2\bar{\sigma}} X_i^{\bar{\sigma} 2}\rangle).$$

Renormalization of the jump parameters is determined

by the correlation functions

$$\chi_2 K_{ij}^{22} = K_{22}\chi_{ij}^{cs} - K_{11}\langle X_i^{02} X_j^{20}\rangle,$$

$$\chi_1 K_{ij}^{11} = K_{11}(\chi_{ij}^{cs} + 1 - n) - K_{22}\langle X_i^{02} X_j^{20}\rangle,$$

$$\chi_2 K_{ij}^{21} = 2\sigma K_{12}\left(\chi_{ij}^{cs} - \frac{1}{2}n - \langle X_i^{02} X_j^{20}\rangle\right),$$

where the static correlation function for particles and spins has the form $\chi_{ij}^{cs} = (1/4)\langle N_i N_j\rangle + \langle \mathbf{S}_i \mathbf{S}_j\rangle$. One-site anomalous correlation functions in Eq. (14) are defined as

$$\chi_2 b_\sigma^{22} = \sum_{m\neq i} V_{im}\{K_{22}(\langle X_i^{\bar\sigma 2} X_m^{\sigma 2}\rangle - \langle X_i^{\sigma 2} X_m^{\bar\sigma 2}\rangle) - 2\sigma K_{12}(\langle X_i^{\sigma 2} X_m^{0\sigma}\rangle + \langle X_i^{\bar\sigma 2} X_m^{0\bar\sigma}\rangle)\},$$

$$\chi_1 b_\sigma^{11} = -\sum_{m\neq i} V_{im}\{K_{11}(\langle X_i^{0\sigma} X_m^{0\bar\sigma}\rangle - \langle X_i^{0\bar\sigma} X_m^{0\sigma}\rangle) - 2\sigma K_{12}(\langle X_i^{0\sigma} X_m^{\sigma 2}\rangle + \langle X_i^{0\bar\sigma} X_m^{\bar\sigma 2}\rangle)\},$$

$$\chi_2 b_\sigma^{21} = \sum_{m\neq i} V_{im}\{K_{22}(\langle X_i^{0\sigma} X_m^{\sigma 2}\rangle + \langle X_i^{0\bar\sigma} X_m^{\bar\sigma 2}\rangle) - 2\sigma K_{12}(\langle X_i^{0\sigma} X_m^{0\bar\sigma}\rangle - \langle X_i^{0\bar\sigma} X_m^{0\sigma}\rangle)\}.$$

APPENDIX 2

While calculating mass operator (9), we used the following functions for irreducible operators describing scattering of one-particle excitations from spin and charge fluctuations (taking into account equations of motion (32)–(34)): $Z_{i,\sigma,2}^{(ir)} = \sum_{l\neq i,\sigma'}(t_{il}^{22}\delta B_{i\sigma\sigma'}^{22} X_l^{\sigma' 2} - 2\sigma t_{il}^{21}\delta B_{i\sigma\sigma'}^{21} X_l^{0\bar\sigma'})$, $Z_{i,0,\bar\sigma}^{(ir)} = \sum_{l\neq i,\sigma'}(t_{il}^{11}\delta B_{i\sigma\sigma'}^{11} X_l^{0\bar\sigma'} - 2\sigma t_{il}^{12}\delta B_{i\sigma\sigma'}^{12} X_l^{\sigma' 2})$. Here, $\delta B_{i\sigma\sigma'}^{\alpha\beta} = B_{i\sigma\sigma'}^{\alpha\beta} - \langle B_{i\sigma\sigma'}^{\alpha\beta}\rangle$.

As a result, the mass operator in representation (18) assumes the form

$$\hat M_{ij\sigma}(\omega) = \left\langle\!\!\left\langle \begin{pmatrix} (Z_i^{\sigma 2})^{(ir)} \\ (Z_i^{0\bar\sigma})^{(ir)} \end{pmatrix} \middle| ((Z_j^{2\sigma})^{(ir)} (Z_j^{\bar\sigma 0})^{(ir)}) \right\rangle\!\!\right\rangle_\omega,$$

$$\hat\Phi_{ij\sigma}(\omega) = \left\langle\!\!\left\langle \begin{pmatrix} (Z_i^{\sigma 2})^{(ir)} \\ (Z_i^{0\bar\sigma})^{(ir)} \end{pmatrix} \middle| ((Z_j^{\bar\sigma 2})^{(ir)} (Z_j^{0\sigma})^{(ir)}) \right\rangle\!\!\right\rangle_\omega. \quad (38)$$

Computation of these functions in the approximation of nonintersecting diagrams leads to the following expression for the mass operator matrix:

$$\hat M_\sigma(\mathbf{q},\omega) = \frac{1}{N}\sum_{\mathbf{k}}\int_{-\infty}^{+\infty} d\omega_1 K^{(+)}(\omega,\omega_1|\mathbf{k},\mathbf{q}-\mathbf{k}) \times\left\{-\frac{1}{\pi}\mathrm{Im}[\hat P_2^{(+)} G_\sigma^{22}(\mathbf{k},\omega_1) + \hat P_1^{(+)} G_\sigma^{11}(\mathbf{k},\omega_1)]\right\},$$

$$\hat\Phi_\sigma(\mathbf{q},\omega) = \frac{1}{N}\sum_{\mathbf{k}}\int_{-\infty}^{+\infty} d\omega_1 K^{(-)}(\omega,\omega_1|\mathbf{k},\mathbf{q}-\mathbf{k}) \times\left\{-\frac{1}{\pi}\mathrm{Im}[\hat P_2^{(-)} F_\sigma^{22}(\mathbf{k},\omega_1) - \hat P_1^{(-)} F_\sigma^{11}(\mathbf{k},\omega_1)]\right\},$$

where the coefficients for the Green functions are defined by the matrices

$$\hat P_2^{(\pm)} = \begin{pmatrix} K_{22}^2 & \pm 2\sigma K_{21}K_{22} \\ 2\sigma K_{21}K_{22} & \pm K_{21}^2 \end{pmatrix},$$

$$\hat P_1^{(\pm)} = \begin{pmatrix} K_{21}^2 & \pm 2\sigma K_{21}K_{11} \\ 2\sigma K_{21}K_{11} & \pm K_{11}^2 \end{pmatrix}.$$

The explicit form of the mass operator for the singlet band in the diagonal approximation is given in Eqs. (21) and (22).

REFERENCES

1. Yu. A. Izyumov, Usp. Fiz. Nauk **169**, 225 (1999) [Phys.–Usp. **42**, 215 (1999)].
2. D. J. Scalapino, Phys. Rep. **250**, 329 (1995); cond-mat/9908287.
3. P. W. Anderson, Science **235**, 1196 (1987); P. W. Anderson, *The Theory of Superconductivity in the High-T_c Cuprates* (Princeton Univ. Press, Princeton, N.J., 1997).
4. N. M. Plakida, V. Yu. Yushankhai, and I. V. Stasyuk, Physica C (Amsterdam) **160**, 80 (1989); V. Yu. Yushankhai, N. M. Plakida, and P. Kalinay, Physica C (Amsterdam) **174**, 401 (1991).
5. N. M. Plakida, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 38 (2001) [JETP Lett. **74**, 36 (2001)].
6. N. M. Plakida, Condens. Matter Phys. (Ukraine) **5**, 707 (2002); cond-mat/0210385.
7. M. Yu. Kagan and T. M. Rice, J. Phys.: Condens. Matter **6**, 3771 (1994).
8. Yu. A. Izyumov and B. M. Letfulov, Int. J. Mod. Phys. B **6**, 3771 (1992).
9. N. M. Plakida and V. S. Oudovenko, Phys. Rev. B **59**, 11949 (1999).
10. E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994).

11. S. Sorella, G. B. Martins, F. Becca, *et al.*, Phys. Rev. Lett. **88**, 117002 (2002).
12. Z. B. Huang, H. Q. Lin, and J. E. Gubernatis, Phys. Rev. B **64**, 205101 (2001).
13. V. J. Emery, Phys. Rev. Lett. **58**, 2794 (1987); C. M. Varma, S. Schmitt-Rink, and E. Abrahams, Solid State Commun. **62**, 681 (1987).
14. N. M. Plakida, Physica C (Amsterdam) **282**–**287**, 1737 (1997).
15. R. O. Zaĭtsev and V. F. Ivanov, Fiz. Tverd. Tela (Leningrad) **29**, 2554 (1987) [Sov. Phys. Solid State **29**, 1475 (1987)]; Fiz. Tverd. Tela (Leningrad) **29**, 3111 (1987) [Sov. Phys. Solid State **29**, 1784 (1987)]; R. O. Zaitsev and V. F. Ivanov, Int. J. Mod. Phys. B **5**, 153 (1988); Physica C (Amsterdam) **153**–**155**, 1295 (1988).
16. N. M. Plakida and I. V. Stasyuk, Mod. Phys. Lett. **2**, 969 (1988).
17. J. Beenen and D. M. Edwards, Phys. Rev. B **52**, 13636 (1995).
18. A. Avella, F. Mancini, D. Villani, and H. Matsumoto, Physica C (Amsterdam) **282**–**287**, 1757 (1997); T. Di Matteo, F. Mancini, H. Matsumoto, and V. S. Oudovenko, Physica B (Amsterdam) **230**–**232**, 915 (1997).
19. T. D. Stanescu, I. Martin, and Ph. Phillips, Phys. Rev. B **62**, 4300 (2000).
20. Th. Maier, M. Jarrell, Th. Pruschke, and J. Keller, Phys. Rev. Lett. **85**, 1524 (2000).
21. A. I. Lichtenstein and M. I. Katsnelson, Phys. Rev. B **62**, R9283 (2000).
22. N. M. Plakida, R. Hayn, and J.-L. Richard, Phys. Rev. B **51**, 16599 (1995).
23. L. F. Feiner, J. H. Jefferson, and R. Raimondi, Phys. Rev. B **53**, 8751 (1996); R. Raimondi, J. H. Jefferson, and L. F. Feiner, Phys. Rev. B **53**, 8774 (1996).
24. V. Yu. Yushankhai, V. S. Oudovenko, and R. Hayn, Phys. Rev. B **55**, 15562 (1997).
25. V. V. Val'kov and S. G. Ovchinnikov, *Quasi-particles in Strongly Correlated Systems* (Sib. Otd. Ross. Akad. Nauk, Novosibirsk, 2001), Chap. 8.
26. D. N. Zubarev, Usp. Fiz. Nauk **71**, 71 (1960) [Sov. Phys.–Usp. **3**, 20 (1960)].
27. J. Jaklić and P. Prelovsék, Phys. Rev. Lett. **74**, 3411 (1995); Phys. Rev. Lett. **75**, 1340 (1995).
28. P. Prelovsék and A. Ramšak, Phys. Rev. B **63**, 180506(R) (2001).
29. J. R. Schrieffer, J. Low Temp. Phys. **99**, 397 (1995).
30. P. W. Anderson, Adv. Phys. **46**, 3 (1997).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 97, No. 2, 2003, pp. 343–357.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 2, 2003, pp. 379–393.

SOLIDS
Electronic Properties

Resonant Tunneling and a Nonlinear Response in RF Fields

V. F. Elesin
Moscow Institute of Engineering Physics (Technical University), Moscow, 115409 Russia
e-mail: VEF@supercon.mephi.ru
Received February 6, 2003

Abstract—The problem of resonant tunneling through a double-barrier nanostructure in a strong alternating electric field is solved completely. To this end, a perturbation method is proposed. Electron wavefunctions and a nonlinear response are obtained in analytical form over wide ranges of field frequencies and amplitudes, using the perturbation method and the semiclassical approximation. The semiclassical expression for the current allows for contributions of all orders with respect to the field, i.e., electron transitions with the emission and absorption of any number of photons. This enables one to find the limits of resonant current and output power. The case of $\hbar\omega \gg \Gamma$ is considered, where $\hbar$ is the rationalized Planck constant, ω is the field frequency, and Γ is the resonance level width. It is established that the maximum resonant current is approximately as high as half the resonant constant current. For the quantum regime of oscillation, the output power can be 10^6–10^7 W/cm^2 at $\omega = 10^{13}$ s^{-1} and the output power rises with ω, in contrast to the well-known classical regime, where the power decreases rapidly. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Quantum interference of electrons forms the basis of many fundamental phenomena. One of these phenomena is resonant tunneling, i.e., the zero-reflection transport of electrons with the energy ε through a double-barrier structure when ε is equal to the resonance energy ε_R; if ε differs from ε_R by an amount larger than the width Γ of the resonance energy level, then the reflection grows and the current decreases sharply.

The current was first calculated by Esaki and Tsu [1]. Later, Chang *et al.* [2] did this for a more realistic structure, viz., the resonant-tunneling diode (RTD). Experimentally, they observed a resonant dependence of current I_0 on voltage V and demonstrated the existence of negative differential conductance. These results highlighted the potential of the RTD as an amplifier or oscillator [3, 4]. From a theoretical viewpoint, the findings raised the problem on resonant tunneling in an applied alternating field and an ac response to the field [5–14].

When an alternating electric field $E\cos(\omega t)$ is applied, the tunneling is accompanied by the absorption and emission of photons $\hbar\omega$. Radiative transitions occur between the states with quasienergies $\varepsilon + n\hbar\omega$, where $n = 0, \pm1, \pm2, \ldots$. It is reasonable to expect that the intensity of transitions will greatly increase if one of the quasienergy levels coincides with the resonance energy ε_R: $\varepsilon - \varepsilon_R = n\hbar\omega$. Spatiotemporal resonance will thus arise, especially when $\omega = \delta \gg \Gamma$; from here on, we set $\hbar = 1$. If the resonance condition $\varepsilon - \varepsilon_R \equiv \delta = \omega$ is observed, the in-phase current $I_c(x, t) = I_c(x)\cos(\omega t)$ must be strongest. The resonant increase in the alternating current for $\omega = \delta \gg \Gamma$ was revealed theoretically in [9].

In early theoretical investigations, the nonresonant regime was considered, i.e., the case $\delta < \Gamma$, with the structure biased into the region of negative differential conductance in the current–voltage characteristic (δ is set by adjusting the bias voltage); examples of this are numerical and analytical studies [5–8] and references therein. In addition, the focus of research was mostly on the collector current $I_c(a)$, where a is the width of the quantum well (QW) [5, 6, 14]. It has been found that with $\omega > \Gamma$ the ac amplitude rapidly decreases with increasing ω and the current can even change its sign. Accordingly, it has been concluded that the cutoff frequency of the RTD is approximately equal to Γ (see, e.g., [8]).

However, the last conclusion is inconsistent with the above physical model. In fact, this conclusion goes against experiments such as those of Sollner *et al.* [3] and Brown *et al.* [4], who detected resonance in an RTD in a 2.5-THz laser field and observed oscillations at frequencies of up to 712 GHz. Also note that analytical studies tend to follow semiphenomenological approaches that do not yield any explicit solution to the Schrödinger equation [5] or treat the problem by tunneling Hamiltonian [6] or similar methods.

In this connection, it should be pointed out that the system of electrons which perform resonant tunneling and interact with an alternating field is extremely sensitive to the energy of electrons injected from the collector and to the boundary conditions. Therefore, an attempt was made to solve the problem in a rigorous

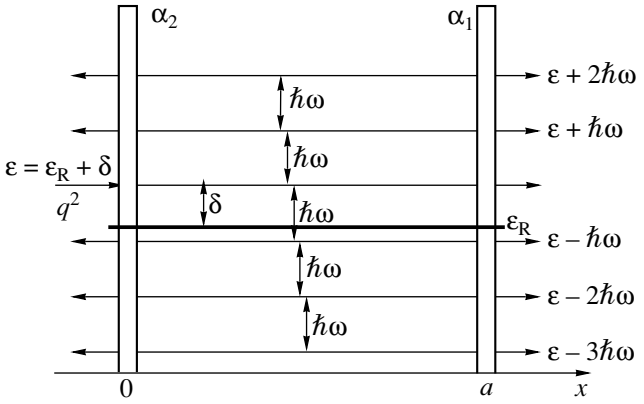

Figure.

quantum-mechanical formulation for a very simple model of the RTD [9, 10]. Explicit formulas for the current in the presence of a weak alternating field were derived by finding exact analytical solutions to the Schrödinger equation subject to open-system boundary conditions. It has been shown that (i) the current amplitude $I_c(x)$ inside the QW ($0 \le x < a$ and the so-called reduced current I_c (see Eq. (31)) exhibit a resonance peak at $\omega = \delta$ and (ii) the two currents retain their signs over the whole frequency range of interest. The resonance arises if the electron energy (and hence the bias voltage) satisfies the condition $\delta > \Gamma$. This regime was termed the quantum regime [9]. Thus, the model of resonant tunneling in an alternating field appears to be physically consistent. It essentially implies that the RTD can oscillate at frequencies much higher than Γ if operated in the quantum regime, since I_c determines field amplification.

By contrast, the collector current $I_c(a)$ (the current outside the QW) has been found to exhibit no resonance and to change its sign at $\omega \ge \Gamma$ [9, 10]. It was this current that was calculated in previous studies, such as [5, 6, 8]. Most experimental and theoretical investigations dealt with the classical regime with a maximum absolute value of differential conductance, in which case $I_c(x)$ is maximum at $\omega = 0$ and decreases as rapidly as $\sim 1/\omega^4$ with increasing ω when $\omega \gg \Gamma$ [9, 10].

In the above-mentioned works, wave functions and currents were calculated in a linear approximation. The aim of the present study is to derive analytical formulas for wave functions and currents in the RTD in a strong field over a wide frequency range. The inclusion of current nonlinearity is required to calculate the output intensity of an oscillating RTD. An approach will be proposed that makes it possible to find nonlinear corrections of any order to the wave function and current. This approach will be used to find a current proportional to the third power of the ac field.

The central result of this study is a semiclassical formula of the ac response in the quantum regime for fields in a wide frequency range and with magnitudes eEa up to a level much higher than $\omega \gg \Gamma$, where e is the electron charge and E is the field amplitude. The expressions include terms of any order in the field amplitude, thus taking into account electron transitions with the emission and absorption of any number of photons. The formalism enables one to find the limits of resonant current and output intensity. In particular, it will be shown that the current at the resonant frequency $\omega = \delta \gg \Gamma$ attains a very high value of about half the constant resonance current I_0 at $\delta = 0$ if the field satisfies the condition $eEa = 2.8\omega$.

The maximum output intensity in the limit $\omega \ll \Gamma$ for the classical and the quantum regime will be estimated. This will also be done in the limit $\omega \gg \Gamma$. In particular, it will be demonstrated that for $\omega = 10^{13}$ s^{-1}, the output power can reach values as high as 10^6–10^7 W/cm^2 if the RTD is operated in the quantum regime. For the well-known classical regime, the output intensity will be shown to sharply decrease with increasing frequency, in agreement with experimental results [4].

This paper is organized as follows. Section 2 gives basic equations, boundary conditions, and expressions for currents in terms of wave functions. Section 3 deals with the linear approximation. Sections 4 and 5 present the perturbative calculations of nonlinear corrections to wave functions and current amplitudes, respectively. In Section 6, the behavior of a nonlinear response is analyzed. In Section 7, the wave function is determined. In Section 8, the nonlinear ac response to a strong field is calculated. The general formulas are used in Section 9 to show that resonant tunneling exists over a wide range of fields. The limiting values of field and output powers are estimated in Section 10.

2. FORMULATION OF THE PROBLEM

Let us explore coherent tunneling within the model used in the earlier study [9]. To simplify the presentation, we consider a one-dimensional QW (or quantum dot) with potential barriers in the form of the delta function, located at $x = 0$ and $x = a$ (see figure). A stationary electron flux proportional to q^2 comes to the QW from $x = -\infty$, with $\varepsilon \approx \varepsilon_R$. Inside the QW, there is an applied electric field $E(t)$ of angular frequency ω with the potential

$$U(x, t) = 2U(x)\cos(\omega t),$$
$$U(x) = \begin{cases} xU, & 0 < x < a, \\ Ua, & x > a, \quad U = -eE/2. \end{cases} \tag{1}$$

The electron wave function obeys the Schrödinger equation

$$i\frac{\partial\Psi}{\partial t} = -\frac{\partial^2\Psi}{\partial x^2} + \alpha[\delta(x) + \delta(x - a)]\Psi + U(x, t)\Psi, \tag{2}$$

where $\hbar = 2m = 1$. We seek a steady-state solution of Eq. (2) in the form [9]

$$\Psi(x, t) = \sum_n \psi_n(x)\exp[-it(\varepsilon + n\omega)], \quad n = 0, \pm1, \pm2, \ldots. \tag{3}$$

Partial wave functions $\psi_n(x)$ refer to electrons of quasienergies $\varepsilon + n\omega$ and satisfy the infinite system of equations

$$\psi_n'' + p_n^2\psi_n = U(x)(\psi_{n+1} + \psi_{n-1}), \quad \psi_n'' \equiv \frac{d^2\psi_n}{dx^2}, \tag{4}$$

$$p_n^2 = p^2 + n\omega, \quad \varepsilon = p^2.$$

If the field is weak in the sense that $Ua/\Gamma \ll 1$, series (3) can be truncated when solving system (4). In a linear approximation with respect to the field Ua, three wave functions will suffice: $\psi_0^{(0)}$ and $\psi_{\pm1}^{(1)}$. In this context, one can calculate the constant current I_0, the polarization in-phase current $I_c(x, t)$, and the reactive current $I_s(x, t)$,

$$I_c(x, t) = I_c(x)\cos(\omega t), \quad I_s(x, t) = I_s(x)\sin(\omega t), \tag{5}$$

in the linear approximation.

In calculating the nonlinear ac response, we will focus on third-order corrections with respect to the field. Accordingly, we will be interested in the wave functions $\psi_{\pm2} \equiv \tilde{\psi}_{\pm2}$ and $\psi_0^{(2)} \equiv \tilde{\psi}_0$, which vary as $(Ua)^2$, and the wave functions $\psi_{\pm1}^3 \equiv \tilde{\psi}_{\pm1}$, which vary as $(Ua)^3$. The partial wave functions obey the equations

$$\psi_0''^{(0)} + p^2\psi_0^{(0)} = 0, \quad \psi_{\pm1}''^{(1)} + p_{\pm1}^2\psi_{(1)}^{(1)} = U(x)\psi_0^{(0)}, \tag{6}$$

$$\tilde{\psi}_0'' + p^2\tilde{\psi}_0 = U(x)[\psi_{+1}^{(1)} + \psi_{-1}^{(1)}], \tag{7}$$

$$\tilde{\psi}_{\pm2}'' + p_{\pm2}^2\tilde{\psi}_{\pm2} = U(x)\psi_{\pm1}^{(1)}, \tag{8}$$

$$\tilde{\psi}_{\pm1}'' + p_{\pm1}^2\tilde{\psi}_{\pm1} = U(x)[\tilde{\psi}_0 + \tilde{\psi}_{\pm2}]. \tag{9}$$

As in [9], we set the boundary conditions

$$\psi_0^{(0)}(0)(1-\beta) + \frac{\psi_0'^{(0)}(0)}{ip} = q, \quad \psi_0^{(0)}(a)(1-\beta) - \frac{\psi_0'^{(0)}(a)}{ip} = 0, \quad \beta = \frac{\alpha}{ip},$$

$$\psi_n^{(1)}(0)(1-\beta_n) + \frac{\psi_n'^{(1)}(0)}{ip_n} = 0, \quad \psi_n^{(1)}(a)(1-\beta_n) - \frac{\psi_n'^{(1)}(a)}{ip_n} = 0, \tag{10}$$

$$\tilde{\psi}_n(0)(1-\beta_n) + \frac{\tilde{\psi}_n'(0)}{ip_n} = 0, \quad \tilde{\psi}_n(a)(1-\beta_n) - \frac{\tilde{\psi}_n'(a)}{ip_n} = 0, \quad \beta_n = \frac{\alpha}{ip_n}. \tag{11}$$

These conditions describe the incoming and the outgoing flux of electrons, their reflection from the barriers, and departure to region $x > a$.

Currents I_0 and $I_c(x)$ can be expressed in terms of the partial wave functions as

$$I_0 = -ie[\psi_0^*\psi_0' - \psi_0(\psi_0^*)'], \tag{12}$$

$$I_c(x) = (I_1^{(1)}(x) + I_{-1}^{(1)}(x)) + (I_{101}(x) + I_{-101}(x)) + (I_{102}(x) + I_{-102}(x)) + (I_2(x) + I_{-2}(x)), \tag{13}$$

where

$$I_n^{(1)} = -ie[(\psi_0^*\psi_n'^{(1)} + \psi_0'\psi_n^{(1)*}) - \text{c.c.}], \tag{14}$$

$$I_{\pm101} = -ie[(\psi_0^{(0)*}\tilde{\psi}_{\pm1}' + \tilde{\psi}_{\pm1}^*\psi_0'^{(0)}) - \text{c.c.}], \tag{15}$$

$$I_{\pm102} = -ie[(\tilde{\psi}_0^*\psi_{\pm1}'^{(1)} + \psi_{\pm1}^{(1)}\tilde{\psi}_0') - \text{c.c.}], \tag{16}$$

$$I_{\pm2} = -ie[(\psi_{\pm1}^{(1)*}\tilde{\psi}_{\pm2}' + \tilde{\psi}_{\pm2}^*\psi_{\pm1}'^{(1)}) - \text{c.c.}]. \tag{17}$$

Here, I_0 is the direct current, $I_n^{(1)}$ is the linear current, I_{101} and I_{102} are the nonlinear currents that correspond to transitions between energy levels ε_0 and $\varepsilon_0 \pm \omega$, and I_2 is the nonlinear current that corresponds to transitions between energy levels $\varepsilon_0 \pm \omega$ and $\varepsilon_0 \pm 2\omega$.

3. FIRST APPROXIMATION

Wave functions $\psi_n^{(1)}$ and current amplitudes $I_c(x)$ and $I_s(x)$ linear in field were found in [9]. The expressions for the currents were presented in simple and clear form by introducing the small parameters ω/ε_R and Γ/ε_R, which fit naturally into the context of the RTD.

In this section we will derive the main results of study [9] in a simpler way, which eliminate terms diverging for $\omega \longrightarrow 0$ The new approach also enables one to find nonlinear corrections of any order.

Equations (6)–(9) subject to boundary conditions (10) and (11) have the solution

$$\begin{aligned}\psi_0^{(0)}(x) &= A\exp(-px) + B\exp(-ipx)\\ &\equiv \gamma_0\cos(px) + i\delta_0\sin(px),\\ \Delta_0 A &= q(2-\beta)\exp(-2ipa), \quad \Delta_0 B = q\beta, \qquad (18)\\ \Delta_0 &\approx \frac{4}{\Gamma}(i\delta-\Gamma), \quad \delta = \varepsilon-\varepsilon_R, \quad \Gamma = \frac{4p^3}{\alpha^2 a},\end{aligned}$$

$$\begin{aligned}\psi_n^{(1)} &= A_n\exp(ip_n x) + B_n\exp(-ip_n x)\\ &\quad -\frac{Ux}{\omega_n}\psi_0 - \frac{2U}{\omega^2}\psi_0',\\ \Delta_n A_n &= q_n(2-\beta_n)\exp(-2ip_n a) + \beta_n\tilde{q}_n, \qquad (19)\\ \Delta_n B_n &= \beta_n q_n + (2-\beta_n)\tilde{q}_n, \quad \omega_n = -n\omega,\\ \Delta_n &= \frac{4}{\Gamma}[i(\delta+n\omega)-\Gamma],\end{aligned}$$

$$q_n = \frac{2Uip}{\omega_n^2}[A(2-\beta_n)+\beta_n B],$$

$$\begin{aligned}\tilde{q}_n &= -\frac{2Uip}{\omega_n^2}\exp[i(p-p_n)a]\\ &\times[\beta_n A + (2-\beta_n)B\exp(-2ip_n a)].\end{aligned} \qquad (20)$$

It has been shown [9] that the quantities

$$\gamma_n = A_n + B_n, \quad \delta_n = A_n - B_n \qquad (21)$$

can be decomposed into two terms,

$$\gamma_n = \gamma_n^{(1)} + \gamma_n^{(2)}, \quad \delta_n = \delta_n^{(1)} + \delta_n^{(2)}, \qquad (22)$$

where the terms $\gamma_n^{(1)}$ and $\delta_n^{(1)}$ are responsible for divergence at $\omega \longrightarrow 0$. These quantities can be written in the form

$$\gamma_n^{(1)} = \frac{2Uip}{\omega_n^2}\delta_0, \quad \delta_n^{(1)} = \frac{2Uip}{\omega_n^2}\gamma_0, \qquad (23)$$

$$\begin{aligned}\gamma_n^{(2)} &= -\frac{U\alpha A a^2}{p^2\Delta n} = -\frac{4iUa}{\Gamma\Delta_0\Delta_n},\\ \delta_n^2 &= (\beta_n - 1)\gamma_n^{(2)}.\end{aligned} \qquad (24)$$

Using the expressions for $\psi_0^{(0)}$ and $\psi_n^{(1)}$, we obtained in [9] the following expressions:

$$I_0 = \frac{\Gamma^2 Q}{\Gamma^2+\delta^2}, \qquad (25)$$

$$\begin{aligned}I_n^{(1)}(x) &= ep\{(K_n^{(1)} + \text{c.c.})\cos[(p-p_n)x]\\ &\quad - i(F_n^{(1)} - \text{c.c})\sin[(p-p_n)x]\},\\ K_n^{(1)} &= \delta_0^*\gamma_n + \gamma_0^*\delta_n, \quad F_n = \delta_0^*\delta_n + \gamma_0^*\gamma_n,\end{aligned} \qquad (26)$$

$$\begin{aligned}I_c^{(1)}(x) &\approx \frac{Ua\Gamma^2\delta Q}{(\delta^2+\Gamma^2)[\Gamma^2+(\delta+\omega)^2][\Gamma^2+(\delta-\omega)^2]}\\ &\times\left[(\delta^2+\omega^2+\Gamma^2) - \omega^2\frac{2x}{a}\right], \quad Q = q^2 p.\end{aligned} \qquad (27)$$

From here on, except for Section 10, we assume that $Q = 1$.

It is difficult to compensate for divergent terms in the current, especially for nonlinear corrections. Nevertheless, it was shown [10] that $\psi_n^{(1)}$ can be represented in a much simpler form by compensating for terms that diverge for $\omega \longrightarrow 0$. Specifically, we use decomposition (22) and expand functions $\cos(p_n x)$ and $\sin(p_n x)$ appearing with $\gamma_n^{(1)}$ and $\delta_n^{(1)}$ into a series in $\omega/p^2 \ll 1$; as a result, we change from Eqs. (19) to

$$\psi_n^{(1)} = \gamma_n^{(2)}\cos(p_n x) + i\delta_n^{(2)}\sin(p_n x). \qquad (28)$$

Substituting Eqs. (28) and (18) into relation (14), we arrive at (27) in a much simpler way.

Equation (27) indicates that in the high-frequency limit, the current amplitude varies linearly with x since ω/ε_R is small. It should be emphasized that currents $I_c(0)$, $I_c(a)$, and the reduced current I_c that determines amplification in the RTD exhibit essentially different frequency dependences:

$$I_c(0) = I_c\frac{\delta^2+\Gamma^2+\omega^2}{\delta^2+\Gamma^2}, \qquad (29)$$

$$I_c(a) = I_c\frac{\delta^2+\Gamma^2-\omega^2}{\delta^2+\Gamma^2}, \qquad (30)$$

$$\begin{aligned}I_c &= \frac{1}{a}\int_0^a I_c(x)dx\\ &= \frac{Ua\Gamma^2\delta}{[\Gamma^2+(\delta+\omega)^2][\Gamma^2+(\delta-\omega)^2]}.\end{aligned} \qquad (31)$$

Expressions (29)–(31) contain the same resonant denominators; distinct numerators reflect the competition among the contributions of radiative transitions between electron states (see [10] for details). In particular, the numerator for collector current $I_c(a)$ decreases sharply at resonance, $\omega = \delta$; more importantly, $I_c(a)$ changes its sign at $\omega > \sqrt{\delta^2 + \Gamma^2}$ and exhibits no peak (resonance) at $\omega \approx \delta$. This behavior can be attributed to the absence of constructive interference outside the QW. Recall that only collector current $I_c(a)$ was calculated in [5, 6, 8]. Those results suggested that the cutoff frequency is approximately equal to Γ [8].

It follows from Eqs. (27) and (31) that current $I_c(x)$ inside the QW and, in particular, reduced current $I_c \equiv I_c(a/2)$ (determining the amplification and generation of the ac field in the RTD) retain their signs over the whole frequency range and have two peaks, which are located at $\omega = 0$ and $\omega^2 = \delta^2 + \Gamma^2$. The former is related to the classical regime and the latter to the quantum regime. It is the quantum regime that corresponds to resonant tunneling in an alternating field.

Thus, it is essential to properly take into account the coordinate dependence reflecting the interference in the alternating field. The existence of the dependence follows from the continuity equation

$$I_c(0) - I_c(a) = \frac{\partial}{\partial t}\int_0^a |\psi|^2 dx. \tag{32}$$

Using relations (28) and (18), it is easy to show that the model exactly obeys Eq. (32):

$$I_c^{(1)}(0) - I_c^{(1)}(a) = \frac{2\omega^2}{\delta^2 + \Gamma^2} I_c^{(1)}. \tag{33}$$

4. NONLINEAR CORRECTIONS TO THE WAVE FUNCTION

A convenient method for obtaining nonlinear corrections can be constructed by decomposing $\psi_n^{(1)}$ into a sum of two terms with renormalized coefficients. First, let us determine $\tilde{\psi}_2$. Equation (8) and boundary conditions (11) have the same form as their counterparts for $\psi_n^{(1)}$ with $\psi_0^{(0)}$ replaced by $\psi_1^{(1)}$ on the right-hand side. The solution to Eq. (8) is

$$\begin{aligned}\tilde{\psi}_2 &= A_2 \exp(ip_2 x) + B_2 \exp(-ip_2 x) \\ &\quad + \frac{Ux}{\omega_{21}}\psi_1^{(1)} - \frac{2U}{\omega_{21}^2}\psi_1'^{(1)}, \\ \omega_{21} &= p_1^2 - p_2^2 = -\omega,\end{aligned} \tag{34}$$

where coefficients A_2 and B_2 satisfy the system of equations

$$\begin{aligned}A_2(2-\beta_2) - \beta_2 B_2 &= q_2 \equiv \frac{2Uip_1}{\omega_{21}^2}[A_1(2-\beta_2) + \beta_2 B_1], \\ -A_2\beta_2 + (2-\beta_2)B_2 \exp(-2ip_2 a) &= \tilde{q}_2 \\ \equiv -\frac{2Uip_1}{\omega_{21}^2}[A_1\beta_2 + (2-\beta_2)B_1 \exp(-2ip_1 a)] &\\ \times \exp[i(p_1 - p_2)a].&\end{aligned} \tag{35}$$

These equations could be obtained from those for $A_1^{(1)}$ and $B_1^{(1)}$ by the substitutions $p \longrightarrow p_1$, $A \longrightarrow A_1$, and $B \longrightarrow B_1$ on the right-hand side. Accordingly, we can carry out the separation and compensation of the diverging expression for γ_2 and δ_2 as in Section 3, which yields

$$\tilde{\psi}_2 = \tilde{\gamma}_2 \cos(p_2 x) + i\tilde{\delta}_2 \sin(p_2 x), \tag{36}$$

$$\begin{aligned}\tilde{\delta}_2 &= \tilde{\gamma}_2(\beta_2 - 1), \quad \tilde{\gamma}_2 = -\frac{U\alpha A_1 a^2}{p_1^2 \Delta_2}, \\ A_1 &= \frac{\beta_1}{2}\gamma_1^{(2)}.\end{aligned} \tag{37}$$

Finally, we substitute the expressions for A_1 and $\gamma_1^{(2)}$ into the formula for $\tilde{\gamma}_2$ to obtain

$$\tilde{\gamma}_2 = 8\left(\frac{Ua}{\Gamma}\right)^2 \frac{1}{\Delta_0 \Delta_1 \Delta_2}. \tag{38}$$

In a similar way, we derive

$$\tilde{\psi}_{-2} = \gamma_{-2} \cos(p_{-2} x) + i\delta_{-2} \sin(p_{-2} x), \tag{39}$$

$$\begin{aligned}\tilde{\delta}_{-2} &= \tilde{\gamma}_{-2}(\beta_{-2} - 1), \\ \tilde{\gamma}_{-2} &= 8\left(\frac{Ua}{\Gamma}\right)^2 \frac{1}{\Delta_0 \Delta_{-1} \Delta_{-2}}.\end{aligned} \tag{40}$$

For $\tilde{\psi}_0$ the problem can be formally reduced to the above if we seek a solution in the form of the sum

$$\tilde{\psi}_0 = \tilde{\psi}_{01} + \tilde{\psi}_{0-1},$$

where $\tilde{\psi}_{01}$ and $\tilde{\psi}_{0-1}$ correspond to $\psi_{+1}^{(1)}$ and $\psi_{-1}^{(1)}$,

respectively. We thus arrive at

$$\tilde{\psi}_0 = \gamma_0 \cos(px) + i\delta \sin(px), \quad (41)$$

$$\tilde{\delta}_0 = \tilde{\gamma}_0(\beta - 1),$$

$$\tilde{\gamma}_0 = 8\left(\frac{Ua}{\Gamma}\right)^2 \frac{1}{\Delta_0^2}\left(\frac{1}{\Delta_1} + \frac{1}{\Delta_{-1}}\right). \quad (42)$$

For $\tilde{\psi}_1$ and $\tilde{\psi}_{-1}$, we follow the same approach to obtain

$$\tilde{\psi}_1 = \tilde{\gamma}_1 \cos(p_1 x) + i\tilde{\delta}_1 \sin(p_1 x),$$

$$\tilde{\delta}_1 = \tilde{\gamma}_1(\beta - 1), \quad (43)$$

$$\tilde{\gamma}_1 = 16i\left(\frac{Ua}{\Gamma}\right)^3 \frac{1}{\Delta_0\Delta_1}\left(\frac{1}{\Delta_0\Delta_1} + \frac{1}{\Delta_0\Delta_{-1}} + \frac{1}{\Delta_1\Delta_2}\right), \quad (44)$$

$$\tilde{\psi}_{-1} = \tilde{\gamma}_{-1} \cos(p_{-1} x) + i\tilde{\delta}_{-1} \sin(p_{-1} x),$$

$$\tilde{\delta}_{-1} = \tilde{\gamma}_{-1}(\beta_{-1} - 1), \quad (45)$$

$$\tilde{\gamma}_{-1} = 16i\left(\frac{Ua}{\Gamma}\right)^3 \frac{1}{\Delta_0\Delta_{-1}} \times \left(\frac{1}{\Delta_0\Delta_1} + \frac{1}{\Delta_0\Delta_{-1}} + \frac{1}{\Delta_{-1}\Delta_{-2}}\right). \quad (46)$$

Conceptually, it is not difficult to find corrections of any order. It is important to note that any such correction is finite for $\omega \longrightarrow 0$.

5. THIRD-ORDER NONLINEAR RESPONSE

We will seek contributions to $I_c^{(3)}(x)$ separately so as to clarify the physical meaning of each contribution (another reason is that the expressions will be rather awkward). We start with current $I_2(x)$. Substituting the expression for $\tilde{\psi}_2$ into formula (17), we represent $I_2(x)$ as

$$I_2(x) = ep\{(\tilde{K} + \text{c.c})\cos[(p_1 - p_2)x] - i(\tilde{F}_2 - \text{c.c.})\sin[(p_1 - p_2)x]\}, \quad (47)$$

which is similar to linear current Eq. (26). From here on we use the notation

$$\tilde{K}_2 = \delta_1^*\tilde{\gamma}_2 + \gamma_1^*\tilde{\delta}_2, \quad \tilde{F}_2 = \delta_1^*\tilde{\delta}_2 + \gamma_1^*\tilde{\gamma}_2.$$

Using Eqs. (37) and (38), we obtain the current

$$I_2(x) = \Phi_2(\omega)\left[(\delta + 2\omega) - \omega\frac{x}{a}\right], \quad I_{-2}(x) = I_2(x, -\omega),$$

$$\Phi_2(\omega) = -\frac{(Ua)^3\Gamma^2}{8(\delta^2 + \Gamma^2)[\Gamma^2 + (\delta + \omega)^2][\Gamma^2 + (\delta + 2\omega)^2]}, \quad (48)$$

so that reduced current I_2 and collector current $I_2(a)$ are expressed as

$$I_2(\omega) = \Phi_2(\omega)\left(\delta + \frac{3}{2}\omega\right), \quad (49)$$

$$I_2(a, \omega) = \Phi_2(\omega)(\delta + \omega). \quad (50)$$

In a similar manner, we represent the other contributions:

$$I_{101}(x) = \frac{(Ua)^3\Gamma^2}{8(\delta^2 + \Gamma^2)}\Pi(\omega)\left(\delta + \omega\frac{x}{a}\right), \quad (51)$$

$$I_{-101}(\omega) = I_{101}(-\omega),$$

$$\Pi(\omega) = \Pi_1(\omega) + \Pi_2(\omega) + \Pi_3(\omega), \quad (52)$$

$$\Pi_1(\omega) = \frac{3\delta^2 - \Gamma^2 + \omega^2 + 4\delta\omega}{(\delta^2 + \Gamma^2)[\Gamma^2 + (\delta + \omega)^2]^2}, \quad (53)$$

$$\Pi_2(\omega) = \Pi_2(-\omega) = \frac{3\delta^2 - \Gamma^2 - \omega^2}{(\delta^2 + \Gamma^2)[\Gamma^2 + (\delta + \omega)^2][\Gamma^2 + (\delta - \omega)^2]}, \quad (54)$$

$$\Pi_3(\omega) = \frac{3\delta^2 - \Gamma^2 + 5\omega^2 + 8\delta\omega}{[\Gamma^2 + (\delta + \omega)^2]^2[\Gamma^2 + (\delta + 2\omega)^2]}, \quad (55)$$

$$I_{102}(\omega) + I_{-102}(\omega) = -\frac{(Ua)^3\Gamma^2\delta}{2(\delta^2 + \Gamma^2)[\Gamma^2 + (\delta + \omega)^2][\Gamma^2 + (\delta - \omega)^2]}. \quad (56)$$

Accordingly, the reduced and collector currents I_{101} and $I_{101}(a)$ are given by

$$I_{101} = \frac{(Ua)^3\Gamma^2\Pi(\omega)(\delta + \omega/2)}{8(\delta^2 + \Gamma^2)}, \quad (57)$$

$$I_{101}(a) = \frac{(Ua)^3\Gamma^2\Pi(\omega)(\delta + \omega)}{8(\delta^2 + \Gamma^2)}. \quad (58)$$

We see from Eq. (56) that the sum $I_{102} + I_{-102}$ is independent of x. The formulas presented in this section describe a third-order nonlinear ac response of the RTD in the entire frequency range of interest and for any coordinate $0 \le x \le a$.

6. ANALYSIS OF THE BEHAVIOR OF THE NONLINEAR RESPONSE

Let us analyze the results obtained in the entire frequency interval. Initially, we examine the response in the limit $\omega \ll \Gamma$, in which case the current is indepen-

dent of x. Adding up the current components given by (51), (56), and (48), we find

$$I_c^{(3)} \approx -\frac{3\delta(Ua)^3\Gamma^2(\Gamma^2-\delta^2)}{2(\delta^2+\Gamma^2)^4} \tag{59}$$

or, due to Eq. (27),

$$I_c = \frac{\delta Ua\Gamma^2}{(\delta^2+\Gamma^2)^2}\left[1-\frac{3(Ua)^2(\Gamma^2-\delta^2)}{2(\delta^2+\Gamma^2)^2}\right]. \tag{60}$$

Notice that the nonlinear correction essentially changes its sign at $\delta = \Gamma$. If $\delta < \Gamma$, the second right-hand summand is positive ($Ua < 0$) and amplification of I_c decreases with increasing field magnitude. If $\delta > \Gamma$, the amplification of I_c grows with the field. Equation (60) was first derived in [11] by another method (see below). It agrees with the numerical results presented in [12].

Let us now proceed to the other limiting case, $\omega \gg \Gamma$. We start with the classical regime, when δ corresponds to the maximal value of the negative differential conductance. A fairly lengthy calculation yields the following expression for reduced current:

$$I_c = \frac{Ua\Gamma^2\delta}{\omega^4}\left[1-15\left(\frac{U}{4\omega}\right)^2\right], \quad \delta \sim \Gamma \ll \omega. \tag{61}$$

More interesting is the behavior of the reduced current in the high-frequency quantum regime with the quasi-resonance condition $\omega \approx \delta \gg \Gamma$. It follows from (48), (51), and (56) that

$$I_c \approx \frac{Ua}{4\omega}\left[1-\left(\frac{U}{4\omega}\right)^2\right], \quad \delta \approx \omega. \tag{62}$$

Two features of Eq. (62) are worth noting. First, unlike in formula (61) for the classical regime, the linear term varies as $1/\omega$ and not as $1/\omega^4$; this indicates considerable amplification even for $\omega \gg \Gamma$. Second, with increasing ω, the nonlinear term decreases (as $\sim 1/\omega^2$).

As regards the collector current amplitude (see Eq. (30)), it changes its sign near the resonant frequency, so it is sensitive to the choice of δ. If $\delta = \omega$, then Eqs. (48), (51), and (56) imply that

$$I_c(a) \approx \frac{Ua\Gamma^2}{4\omega^3}\left[1-\frac{(Ua)^2}{6\omega^2}\right]. \tag{63}$$

For $\delta = \sqrt{\omega^2+\Gamma^2}$, we have

$$I_c(a) \approx \frac{Ua\Gamma^2}{2\omega^3}\left[3-\frac{1}{3}\left(\frac{Ua}{\omega}\right)^2\right]. \tag{64}$$

It is seen that the collector current, as well as the reduced current, decreases as rapidly as $\sim 1/\omega^3$ with increasing ω (cf. Eq. (61)).

7. WAVE FUNCTIONS IN A STRONG FIELD

The method proposed in Section 4 enables one to derive the wave-function of any order for a wide frequency range and to a desired accuracy. On the other hand, the method requires the summation of a series if the RTD operates in a strong field, $Ua \gg \Gamma$. In the case when $Ua \gg \omega$ and $Ua \gg \Gamma$, the wave functions can be found in the semiclassical approximation. Indeed, the applicability conditions in this case are given by the inequalities

$$\omega \ll \varepsilon_R, \quad Ua \ll \varepsilon_R, \tag{65}$$

which are always satisfied for RTDs. The semiclassical approximation was employed in [13, 14], where the wave function was calculated only in the collector region $x \geq a$. However, the wave function and current $I_c(x)$ were not determined in the most interesting region $0 \leq x \leq a$. Moreover, the authors of [13, 14] used the QW model with semiclassical barriers, in which numerical computations are extremely cumbersome only for large numbers n of resonance levels.

With our model the potential barriers are in the form of the delta function (see Section 2), the tunneling problem is exactly solvable for a zero applied field, and the semiclassical approximation is used for an alternating field only. The strategy enables one to set boundary conditions in a simpler and rigorous manner for any n. It provides an exactly solvable equation of tunneling. The problem on resonant tunneling in an electromagnetic field can thus be solved completely in the case $\omega < \varepsilon_R$ and $Ua < \varepsilon_R$.

For a strong applied field, we seek steady-state solutions $\Psi_1(x, t)$, $\Psi_2(x, t)$, and $\Psi_3(x, t)$ of the Schrödinger equation (2) in three respective regions: (i) $x \leq 0$, (ii) $0 \leq x \leq a$, and (iii) $a \leq x$. The boundary conditions are

$$\begin{aligned} \Psi_1(0, t) &= \Psi_2(0, t), \\ \frac{\partial\Psi_2(0, t)}{\partial x} - \frac{\partial\Psi_1(0, t)}{\partial x} &= \alpha\Psi_1(0, t), \end{aligned} \tag{66}$$

$$\begin{aligned} \Psi_3(a, t) &= \Psi_2(a, t), \\ \frac{\partial\Psi_3(a, t)}{\partial x} - \frac{\partial\Psi_2(a, t)}{\partial x} &= \alpha\Psi_3(a, t). \end{aligned} \tag{67}$$

These conditions represent the continuity of wave functions and the discontinuity of their derivatives at $x = 0$ and $x = a$. Expand the wave function for region (ii) as

$$\begin{aligned} \Psi_2(x, t) = &\exp\{-i\varepsilon t + ipx + iS_p(x, t)\} \\ &+ \exp\{-i\varepsilon t - ipx + iS_{-p}(x, t)\}. \end{aligned} \tag{68}$$

The right-hand terms refer to electrons moving in the forward and the reverse direction, respectively, inside the QW. Semiclassically, we neglect the second-order derivatives of $S_{\pm p}(x, t)$ to obtain the well-known equations [15]

$$-\frac{\partial S_p(x,t)}{\partial x} = 2p\frac{\partial S_p}{\partial t} + W(x)\cos(\omega t), \quad W(x) = 2U(x), \tag{69}$$

$$-\frac{\partial S_{-p}(x,t)}{\partial x} = -2p\frac{\partial S_{-p}}{\partial t} + W(x)\cos(\omega t). \tag{70}$$

The solution to Eq. (69) can be written as

$$S_p(x,t) = S_p^U(x,t) + S_p\left(\frac{x}{2p} - t\right), \tag{71}$$

$$S_p^U(x,t) = \exp\left\{i\omega\left(t - \frac{x}{2p}\right)\right\} \times\left[-\int^x dx'\frac{W(x')}{4p}\exp\left(\frac{i\omega x'}{xp}\right)\right] + \text{c.c.}, \tag{72}$$

where S_p^U is a solution of the inhomogeneous equation and $S_p(x/2p - t)$ is the solution of the homogeneous equation subjected to the boundary conditions (67) and (66). To find S_{-p}, one should replace p with $-p$ in (71) and (72).

Choosing the origin at point $x = a$ and introducing the tunneling functions $f_p(z)$ and $f_{-p}(\tilde{z})$, we can write wave function (68) in the form

$$\Psi_2(x,t) = \exp\{-i\varepsilon t + ipx + iS_p^U(x,t)\} f_p(z) + \exp\{-i\varepsilon t - ipx + iS_p^U(x,t)\} f_{-p}(\tilde{z}), \tag{73}$$

where

$$S_p^U(x,t) = -\frac{1}{4p} \times\left\{\exp(i\omega z)\int_a^x W(x')dx'\exp\left[-\frac{i\omega(x'-a)}{2p}\right] + \text{c.c.}\right\}, \tag{74}$$

$$S_{-p}^U(x,t) = \frac{1}{4p} \times\left\{\exp(i\omega\tilde{z})\int_a^x W(x')dx'\exp\left[\frac{i\omega(x'-a)}{2p}\right] + \text{c.c.}\right\}, \tag{75}$$

$$z = \frac{x-a}{2p} - t, \quad \tilde{z} = -\frac{x-a}{2p} - t. \tag{76}$$

We seek the wave functions in regions (i) and (iii) in the form

$$\Psi_1(x,t) = q\exp(-i\varepsilon t + ipx) + \exp(-i\varepsilon t - ipx) f_1(\tilde{z}), \tag{77}$$

$$\Psi_3(x,t) = \exp(-i\varepsilon t + ipx) f(z). \tag{78}$$

In Eq. (77), the first right-hand summand represents the incoming electron flow from $x = -\infty$ and the second one describes the reflection characterized by function $f_1(\tilde{z})$. Function Ψ_3 characterizes the electron flow in direction $x = +\infty$ with tunneling function $f(z)$. Boundary conditions (67) imply that, for $x = a$,

$$f_p(-t) = \frac{1}{2}\left(1 - \frac{\alpha}{ip}\right) f(-t), \tag{79}$$

$$f_{-p}(-t) = \frac{1}{2}\frac{\alpha}{ip}\exp(ipa) f(-t). \tag{80}$$

Eliminating function $f_1(\tilde{z})$ by means of boundary conditions (66) at $x = 0$ and using expressions (79) and (80), we arrive at the equation for $f(z)$ with displaced arguments:

$$\left(2 - \frac{\alpha}{ip}\right)^2\exp\left[iS_p^U\left(-\frac{a}{2p} - t\right)\right] \times f\left(-t - \frac{a}{2p}\right) - \left(\frac{\alpha}{ip}\right)^2 \times\exp\left[2ipa + iS_{-p}\left(\frac{a}{2p} - t\right)\right] f\left(\frac{a}{2p} - t\right) = 1. \tag{81}$$

Introducing the new function

$$F(z) = f(z - T), \quad T = \frac{a}{2p}, \tag{82}$$

we obtain the equation for $F(z)$,

$$\left(2 - \frac{\alpha}{ip}\right)^2 F(z) - \left(\frac{\alpha}{ip}\right)^2 F(z + 2T)\exp(2ipa)A(z) = Y(z), \tag{83}$$

where

$$A(z) = \exp\{i[S_{-p}(z + T) - S_p(z - T)]\}, \tag{84}$$

$$Y(z) = \exp\{-iS_p(z - T)\}, \tag{85}$$

with

$$S_p = \frac{1}{4p} \times \left\{ \exp(i\omega z) \int_0^a W(x')dx' \exp\left[-\frac{i\omega(x'-a)}{2p} \right] + \text{c.c.} \right\}. \quad (86)$$

Equation (83) can be solved exactly. Following the approach used in [13, 14], we seek a solution to Eq. (83) in the form

$$F(z) = \sum_{k=-\infty}^{\infty} c_k \Phi_k(z), \qquad \Phi_k = \exp(-ikz\omega)\Phi_0(z), \quad (87)$$

where $\Phi_0(z)$ satisfies the equation

$$\Phi_0(z) = \Phi_0(z + 2T)A(z). \quad (88)$$

To calculate expansion coefficients c_k, we substitute $F(z)$ from Eq. (87) into (83) and perform the weighted integration with respect to z using the weight function $\exp(ikz\omega)$:

$$c_k = \frac{Y_k}{\Delta_k}, \quad Y_k = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} \frac{Y(z)\exp(ik\omega z)dz}{\Phi_0(z)}, \quad (89)$$

$$\Delta_k = \left(2 - \frac{\alpha}{ip}\right)^2 - \left(\frac{\alpha}{ip}\right)^2 \exp(2ipa - 2ik\omega T). \quad (90)$$

If $\alpha/p \gg 1$, then resonance determinants Δ_k have the form (19). The function $\Phi_0(z)$ is sought in the form

$$\Phi_0(z) = \exp\left[i \sum_{m=-\infty}^{\infty} b_m \exp(i\omega z m) \right]. \quad (91)$$

Condition (88) allows us to determine coefficients b_m, after which (91) becomes

$$\Phi_0(z) = \exp\left\{ \sum_{n=-\infty}^{\infty} \exp(in\omega z) \frac{\omega}{2\pi} \times \int_0^{2\pi/\omega} \frac{\exp(im\omega z')\ln A(z')dz'}{1 - \exp(2in\omega T)} \right\}. \quad (92)$$

In view of definition (84) for $A(z)$, we obtain the exact formula

$$\Phi_0(z) = \exp\left\{ -\frac{i}{2p} \int_0^a W(x)dx \times \cos\frac{\omega x}{2p} \left[\frac{\exp(i\omega z)}{1 - \exp(2i\omega T)} + \text{c.c.} \right] \right\}. \quad (93)$$

It is important to note that $|\Phi_0(z)|^2 = 1$. The frequencies of interest satisfy the condition $\omega T \approx \omega/\varepsilon_R \ll 1$, so that Eq. (93) can be reduced to

$$\Phi_0(z) \approx \exp\left[i\frac{\overline{W}}{\omega} \sin(\omega z) \right], \qquad \overline{W} = \frac{1}{a} \int_0^a W(x)dx = Ua. \quad (94)$$

Finally, we substitute expression (93) into Eq. (89) to obtain the Fourier component Y_k,

$$Y_k \approx J_k\left(\frac{\overline{W}}{\omega}\right), \quad (95)$$

where J_k are the Bessel functions; formula (95) is accurate to the second order in ωT. We thus obtain $F(z)$ in the form of the series

$$F(z) = \Phi_0(z) \sum_{k=-\infty}^{\infty} \frac{\exp(-ik\omega z) J_k\left(\frac{\overline{W}}{\omega}\right)}{\Delta_k(\omega)} \quad (96)$$

or, in the integral representation,

$$F(z) = \Phi_0(z)\exp[(\Gamma + i\delta)z] \times \int_{-\infty}^{-z} dz' \exp\left[(\Gamma + i\delta)z' + \frac{i\overline{W}}{\omega} \sin(\omega z) \right]. \quad (97)$$

Notice that Eq. (97) is similar to the expression for $f(z)$ derived by Sokolovski [14] in the case $x \geq a$, although the two formulas use different values of Γ, δ, and $\overline{W}$ because they are based on different models.

8. NONLINEAR RESPONSE OF AN RESONANT-TUNNELING DIODE IN A STRONG ALTERNATING FIELD

For $0 \leq x \leq a$, we take wave function (73) to calculate the current in the semiclassical approximation:

$$I(x, t) = 4[|f_p(z)|^2 - |f_{-p}(\tilde{z})|^2], \quad (98)$$

small terms ω/ε_R and Ua/ε_R being omitted. Substituting f_p and f_{-p} from Eqs. (79) and (80) into (98), we express the current in terms of tunneling functions $f(z)$ and $f(\tilde{z})$:

$$I(x, t) = |f(z)|^2 + \frac{\alpha^2}{4p^2}[|f(z)|^2 - |f(\tilde{z})|^2]. \quad (99)$$

The first right-hand term varies with x insignificantly, since the coordinate correction in $f(z)$ is proportional to the small quantity

$$\omega\frac{x-a}{2p} \approx \frac{\omega}{\varepsilon_R}.$$

The second term, though proportional to $\omega(x-a)/2p$ as well, includes the large factor $(\alpha/p)^2$ due to constructive interference of electrons in the QW. It is this term that ensures effective resonant tunneling (quantum regime) and is responsible for the coordinate dependence of the current.

Since $|f(x)|^2 = 1$, it is convenient (bearing in mind that $|\Phi_0(z)|^2 = 1$) to use the expression

$$\begin{aligned}|f(z)|^2 &= \Gamma^2\int_0^\infty d\tau_1 \exp[-(\Gamma - i\delta)\tau_1] \\ &\times \int_0^\infty d\tau_2 \exp[-(\Gamma + i\delta)\tau_2] \\ &\times \exp[iA\sin(\omega z) + iB\cos(\omega z)],\end{aligned} \quad (100)$$

where

$$\begin{aligned}A &= \frac{\overline{W}}{\omega}[\cos(\omega\tau_2) - \cos(\omega\tau_1)], \\ B &= \frac{\overline{W}}{\omega}[\sin(\omega\tau_2) - \sin(\omega\tau_1)].\end{aligned} \quad (101)$$

In Eq. (100), we omitted the correction to z proportional to T on the grounds that T is a small correction to the first right-hand term of Eq. (99) and is cancelled out in the difference appearing in the second term.

Since we are interested in the current proportional to $\cos(\omega t)$, it is convenient to separate it from the very outset using the Fourier transform

$$I_c(x) = \frac{\omega}{\pi}\int_0^{2\pi/\omega} \cos(\omega t) I(xt) dt. \quad (102)$$

Since

$$\begin{aligned}&\frac{\omega}{2\pi}\int_0^{2\pi/\omega} dt \exp(i\omega k t) \\ &\times \exp[\exp(i\omega t)a^+ + \exp(-i\omega t)a^-] \\ &= (i)^k\left(\frac{a^+}{a^-}\right)^{k/2} J_k(2(a^+a^-)^{1/2}),\end{aligned} \quad (103)$$

we arrive at the following expression for current:

$$\begin{aligned}I_c(x) &= 8i\Gamma^2\int_0^\infty d\tau_1 \exp[-(\Gamma - i\delta)\tau_1] \\ &\times \int_0^\infty d\tau_2 \exp[-(\Gamma + i\delta)\tau_2] J_1\left(\frac{2\overline{W}}{\omega}\sin\left[\omega\frac{\tau_2 - \tau_1}{2}\right]\right) \\ &\times \left\{\cos\left[\omega\frac{\tau_2 + \tau_1}{2}\right] + \frac{a-x}{a}\frac{\omega}{\Gamma}\sin\left[\omega\frac{\tau_2 + \tau_1}{2}\right]\right\}.\end{aligned} \quad (104)$$

The second summand in this expression plays a key role, originating from the bracketed difference of Eq. (99). Introducing the new variable $t = \tau_2 - \tau_1$ and integrating with respect to τ_1, we obtain the final expression for current within the QW ($0 \le x \le a$):

$$\begin{aligned}I_c(x) &= \frac{4\Gamma^2}{\omega^2 + 4\Gamma^2}\int_0^\infty dt \exp(-\Gamma t) J_1\left(\frac{2\overline{W}}{\omega}\sin\frac{\omega t}{2}\right)\sin(\delta t) \\ &\times \left[\left(2\Gamma\cos\frac{\omega t}{2} - \omega\sin\frac{\omega t}{2}\right)\right. \\ &\left. + \frac{a-x}{a}\frac{\omega}{T}\left(\omega\cos\frac{\omega t}{2} + 2\Gamma\sin\frac{\omega t}{2}\right)\right].\end{aligned} \quad (105)$$

The expression for reduced current assumes a simpler form:

$$\begin{aligned}I_c &= 2\Gamma\int_0^\infty dt \exp(-\Gamma t) J_1\left(\frac{2\overline{W}}{\omega}\sin\frac{\omega t}{2}\right) \\ &\times \sin(\delta t)\cos\frac{\omega t}{2}.\end{aligned} \quad (106)$$

Expressions (105) and (106) are valid in a wide range of frequencies $\omega < \varepsilon_R$ and field amplitudes $Ua < \varepsilon_R$ and give a complete solution to the problem of resonant tunneling in an alternating field. A framework is thus provided for modeling oscillations in the RTD, within which one could calculate the limits of current and output intensity. Also note that Eqs. (105) and (106) take into account the contributions of all orders with respect

to field, i.e., electron transitions involving absorption and emission of any number of photons. Moreover, Eqs. (105) and (106) imply all the results of earlier analytical studies [9–11, 14], the formulas of Sections 2 and 6, and the numerical results obtained in [12].

Equation (105) can be generalized to the case of potential barriers differing in α [10]. Let α_1 and α_2 characterize the barriers at $x = a$ and $x = 0$, respectively. Then we must modify the formulas by replacing Γ with Γ_j if the former appears at the $(1 - x/a)$ term; otherwise, Γ is to be replaced with Γ_{12}. The new quantities are defined as

$$\Gamma_j = \frac{4p^3}{a\alpha_j^2}, \quad \Gamma_{12} = \frac{1}{2}(\Gamma_1 + \Gamma_2).$$

Note that the coordinate dependence remains linear for any field due to the fact that x is multiplied by the small parameter ω/ε_R.

Let us now calculate the constant current flowing through the RTD,

$$I_0(\beta) = \frac{\omega}{\pi}\int_0^{2\pi/\omega} I_c(x, t)dt,$$

for a high field. Due to Eq. (103), integration with respect to τ_2 yields

$$I_0(\beta) = \Gamma\int_0^\infty dte^{-\Gamma t}J_0\left(\frac{2\overline{W}}{\omega}\sin\frac{\omega t}{2}\right)\cos(\delta t). \quad (107)$$

Notice that Eq. (107) becomes Eq. (25) if $U = 0$.

Finally, we find the reactive current:

$$\begin{aligned} I_s(x) &= \frac{\omega}{\pi}\int_0^{2\pi/\omega} dt\sin(\omega t)I(xt) \\ &= \frac{4\Gamma^2}{\omega^2 + 4\Gamma^2}\int_0^\infty dte^{-\Gamma t}J_1\left(\frac{2\overline{W}}{\omega}\sin\frac{\omega t}{2}\right)\sin(\delta t) \\ &\times\left[\left(\omega\cos\frac{\omega t}{2} + 2\Gamma\sin\frac{\omega t}{2}\right)\right. \\ &\left. + \frac{a-x}{a}\frac{\omega}{\Gamma}\left(2\Gamma\cos\frac{\omega t}{2} - \omega\sin\frac{\omega t}{2}\right)\right]. \end{aligned} \quad (108)$$

9. RESONANT TUNNELING IN A STRONG ALTERNATING FIELD

We start with limiting cases and compare the results with what was derived in preceding sections. Consider a linear approximation with respect to field. Expanding the Bessel function $J_1(x) \approx x/2$ and calculating the integrals appearing in Eqs. (105) and (106), we arrive at Eqs. (27) and (31), respectively. Clearly, it is the second summand in Eq. (105) that is responsible for resonant tunneling in the field and for the quantum regime. This term vanishes at $x = a$.

Let us find the current in the low-frequency limit $\omega \ll \Gamma$. It follows from (105) or (106) that

$$I_c(\overline{W}) = 2\Gamma\int_0^\infty dte^{-\Gamma t}J_1(\overline{W}t)\sin(\delta t),$$

as derived in our earlier study [11]. Following [11], we represent I_c as

$$\begin{aligned} I_c(\overline{W}) &= -\frac{\Gamma}{\sqrt{2\overline{W}}}\frac{\sqrt{\sqrt{x^2+y^2}-x}}{\sqrt{x^2+y^2}}, \\ x = 1 + \frac{\overline{W}^2(\Gamma^2-\delta^2)}{(\Gamma^2+\delta^2)^2}, &\quad y = \frac{2\overline{W}^2\delta\Gamma}{(\Gamma^2+\delta^2)^2}. \end{aligned} \quad (109)$$

As $\overline{W} \longrightarrow 0$, the current tends to zero, the first two expansion terms constituting the right-hand side of Eq. (60). In the opposite limiting case $Ua \gg \Gamma$, the current decreases in magnitude but remains negative:

$$I_c \approx -\frac{\delta\Gamma}{(\overline{W}a)^2}. \quad (110)$$

Thus, there should be a peak. It is easy to show that I_c attains its peak value for $\overline{W}a \approx 1.3\Gamma$ and is equal to

$$I_c(\overline{W}_m, \omega \longrightarrow 0) \approx -0.33, \quad (111)$$

so that the peak value is one-third of the constant resonance current I_0 (see (25)).

Let us compare formulas (105) and (106) with the results of Section 6 over the whole frequency range. Taking the expansion of the Bessel function up to third-order terms,

$$J_1(y) \approx \frac{y}{2} - \frac{y^3}{16},$$

we again arrive at Eqs. (61) and (62) for classical and quantum regimes. In the next approximation, the

collector current has the form

$$I_c^{(3)}(a) = \frac{(Ua)^3\Gamma^2}{64\omega^2}\left\{\left[\frac{\delta-2\omega}{[\Gamma^2+(\delta-\omega)^2][\Gamma^2+(\delta-2\omega)^2]} + \frac{\delta+2\omega}{[\Gamma^2+(\delta+\omega)^2][\Gamma^2+(\delta+2\omega)^2]}\right] - \frac{3}{\delta^2+\Gamma^2}\left[\frac{\delta-\omega}{\Gamma^2+(\delta-\omega)^2}+\frac{\delta+\omega}{\Gamma^2+(\delta+\omega)^2}\right] + \frac{4\delta}{[\Gamma^2+(\delta-\omega)^2][\Gamma^2+(\delta+\omega)^2]}\right\}. \quad (112)$$

A comparison of Eq. (112) with Eqs. (50), (56), and (58) obtained by perturbation theory shows that there is agreement between them in all the limiting cases. Specifically, Eq. (112) is identical to Eq. (60), (63), or (64) for $\omega \ll \Gamma$, $\delta = \omega$, or $\delta = \sqrt{\omega^2+\Gamma^2}$, respectively. Thus, the two approaches yield identical results.

Let us now analyze currents in the case of strong fields: $\overline{W} \gg \Gamma$ and $\overline{W} \gg \omega$. Resonant tunneling in a strong alternating field is one of the most interesting. Consider the quantum regime with $\omega = \delta \gg \Gamma$. In Eq. (106), we change to integration over the interval $0 \le t \le \pi/\omega$ to eventually obtain an approximation for reduced current accurate to the second order in Γ/ω:

$$I_c^R(\beta) = \frac{8}{\pi}\int_0^1 z\,dz\sqrt{1-z^2}J_1(\beta z) = \frac{4}{\beta}J_1^2\left(\frac{\beta}{2}\right), \quad \beta = \frac{2\overline{W}}{\omega} = \frac{Ua}{\omega}. \quad (113)$$

This is one of the major results of the study. It is seen that with $\beta \longrightarrow 0$ the current decreases and the first two terms of the expansion in β are the same as the corresponding right-hand terms of Eq. (62), which was derived by perturbation theory. As β increases ($Ua \gg \omega$), the current exhibits a succession of peaks decreasing in height with zero-level valleys in between. It is important to note that dependence $I_c^R(\beta)$ is the same as the Fraunhofer diffraction pattern from a circular aperture [16]. This fact indicates the coherent character of resonant tunneling. Also note that the peaks and valleys are reached when the momentum eE/ω gained by an electron in the field is a multiple of $\hbar/p$.

Let us determine the alternating field amplitude at which the current amplitude attains its first maximum. Solving the equation

$$\frac{dI_c^R(z_0)}{dz_0} = 0, \quad J_1(z_0) = 2z_0J_2(z_0), \quad z_0 = \frac{\beta_0}{2}, \quad (114)$$

we find that $z_0 \approx 1.4$ and $U_0a/\omega = 2.8$, so that the current maximum is

$$I_c^R(z_0) = \frac{2J_1^2(z_0)}{z_0} \approx -0.41. \quad (115)$$

Thus, the reduced alternating current attains a value comparable to the constant current at resonance, $I_0(\delta = 0) = 1$ (see (25)). The reduced current amplitude is much higher than that for the classical regime (see relation (120)) and the collector current; in fact, it is somewhat larger than the maximum for $\omega \gg \Gamma$ as given by (111). Thus, constructive interference makes it possible to achieve large power gains and output intensities at superhigh frequencies $\omega \gg \Gamma$.

We now examine the quantum regime ($\omega = \delta \gg \Gamma$) in terms of constant current. Let us compare constant current $I_0(\beta)$ in a strong field with I_0 for a zero field. Formula (107) provides this approximation accurate to the second order in Γ/ω:

$$I_0(\beta) \approx J_1^2\left(\frac{\beta}{2}\right). \quad (116)$$

For $\beta = \beta_0$, $I_0(\beta_0) \approx 0.29$. Notice that this value is much larger than $I_0 = \Gamma^2/\omega^2$ for $\beta = 0$ and $\delta = \omega \gg \Gamma$ (see Eq. (25)). Physically, the growth in current $I_0(\beta)$ is due to the resonant radiative transitions of electrons due to the alternating field. To see this, note that Eqs. (113) and (116) imply

$$I_0(\beta) = \frac{\beta}{4}I_c^\beta.$$

Here, the right-hand term is proportional to the total number of electrons that have performed radiative transitions. The above relation can be used for the experimental verification of the oscillatory behavior predicted for I_c^R, since constant currents are easier to measure.

Let us proceed to the classical (nonresonant) regime with $\omega \gg \Gamma$, $\omega \gg \delta$, and $\delta \sim \Gamma$. In this case, Eq. (106) can be brought to the form

$$I_c = \frac{2\Gamma}{i\omega}\int_0^1 dz\,J_1(\beta z)\left[\frac{e^{\Delta\varphi}-e^{-\Delta\varphi}}{e^{\Delta}-e^{-\Delta}} - \text{c.c.}\right], \quad \Delta = \frac{\pi(\Gamma - i\delta)}{\omega}, \quad \varphi(z) = 1 - \frac{2}{\pi}\arcsin z. \quad (117)$$

For high frequencies $\Delta \ll 1$, we have

$$I_c = \frac{4\pi^2\delta\Gamma^2}{3\omega^3}\int_0^1 dz\,J_1(\beta z)\varphi(1-\varphi^2). \quad (118)$$

With $\beta \longrightarrow 0$ (weak fields), it is easy to see that Eq. (118) exactly coincides with Eq. (61). With an arbi-

trary β, the integral in (118) cannot be evaluated easily, but can be approximated as follows. Comparing the integrands in formulas (113) and (118), we notice that they include analogous terms: $z\sqrt{1-z^2}$ and $\varphi(1-\varphi^2)$, respectively. The point is that both of them vanish for $z = 0$ and $z = 1$ and have approximately equal respective maxima. Accordingly, we can change from Eq. (118) to

$$I_c^{K\Lambda} \approx \frac{4\pi^2\delta\Gamma^2}{3\omega^3}\frac{\pi}{2\beta}J_1^2\left(\frac{\beta}{2}\right). \tag{119}$$

Approximation (119) was verified to within 10% by numerical integration of Eq. (118). Thus, the reduced current in the classical regime differs from that in the quantum regime (see Eq. (113)) by a factor of $O((\Gamma/\omega)^3)$. The maximum of the former current is

$$I_c^{K\Lambda}(\beta_0) \approx 0.7\frac{\delta\Gamma^2}{\omega^3}. \tag{120}$$

Let us explore the dependence of collector current $I_c(a)$ on β and ω for $\delta = \omega \gg \Gamma$. We change to integration over the interval $0 < t < \pi/\omega$ to eventually obtain

$$\begin{aligned} I_c(a) &= \frac{16\Gamma^2}{\omega^2}\left[\frac{1}{\beta}J_1^2\left(\frac{\beta}{2}\right) - \kappa(\beta)\right], \\ \kappa(\beta) &= \int_0^1 dz z^2\left(1 - \frac{2}{\pi}\arcsin z\right)J_1(\beta z). \end{aligned} \tag{121}$$

It follows that $I_c(a)$ tends to zero as $\beta \longrightarrow 0$; the first two terms of the expansion in β are identical to the respective right-hand terms of Eq. (63). In the opposite limit $\beta \gg 1$, the current decreases as $\overline{W}^{-2}$:

$$I_c(a) \approx -\frac{2}{\pi}\left(\frac{4\Gamma}{\overline{W}}\right)^2. \tag{122}$$

Compared with Eqs. (113), the maximum of $I_c(a)$ differs by an $O((\Gamma/\omega)^2)$ factor.

Thus, the reduced current, as well as the current $I_c(x)$ inside the QW, shows resonant behavior not only in weak fields (see Eq. (31)) but also in strong ones. The resonance nature of collector current at $\omega = \delta$ is suppressed in fields of any magnitude.

10. OUTPUT POWER LIMITS FOR RESONANT-TUNNELING DIODE

The above expressions for current enable one to construct a theory of oscillation of the RTD. A steady-state field amplitude can be calculated by substituting an expression for the reduced current amplitude into the field equation

$$\frac{E}{2\tau_0} = -\frac{2\pi}{\chi}I_c, \tag{123}$$

where τ_0 is the time constant that characterizes resonator loss and χ is the dielectric constant [9]. We will focus on the limits of fields and output intensities. If $\omega \ll \Gamma$, we substitute expression (110) to obtain for $Ua \gg \Gamma$

$$U_0 a = \frac{\delta\Gamma^2}{(U_0 a)^2}\frac{Q}{Q_{th}}, \tag{124}$$

where Q is the pumping current and

$$Q_{th} = \frac{\chi}{4\pi\tau_0\Gamma a}$$

is the threshold current. Equation (124) gives the limiting field

$$U_0 a \approx \Gamma\left(\frac{Q}{Q_{th}}\right)^{1/3} > \Gamma, \tag{125}$$

which is bounded only by the maximum allowable pumping current and the resonance width Γ.

For the quantum regime with $\omega \gg \Gamma$, it follows from (123) and (113) that the field obeys the equation

$$\beta^2 = \tilde{q}J_1^2\left(\frac{\beta}{2}\right), \quad \tilde{q} = \frac{4\pi\tau_0}{\chi}\frac{a}{\omega}. \tag{126}$$

It is seen that the field magnitude is limited only by the maximum admissible pumping current. For the first minimum of function $J_1(\beta/2)$, where $\beta/2 = 4$, we have

$$U_p(\omega)a = 8\omega, \quad \omega \gg \Gamma. \tag{127}$$

Thus, if the RTD oscillates at $\omega \gg \Gamma$ in the quantum regime, it produces a maximum field magnitude much higher than that in the low-frequency regime $\omega \ll \Gamma$ and $\delta \approx \Gamma$:

$$\frac{U_p(\omega)}{U_0(0)} \approx \frac{8\omega}{\Gamma}.$$

It turns out that output power of RTD in the quantum regime varies as ω^2 and can be as high as 10^6–10^7 W/cm^2 at $\omega = 10^{13}$ s^{-1}.

For comparison, let us estimate the output power used in the classical regime with $\delta \approx \Gamma$ and $\omega \gg \Gamma$. Using approximation (119), we see that, for a given Q, the output power is about $(\Gamma/\omega)^3$ times lower than in the quantum regime. Accordingly, the output power should decrease as ω^{-2} with increasing ω; presumably, this behavior was observed in the experiment of Brown

et al. [4]. Thus, the RTD operated in the quantum regime appears to have great potential as an oscillator with superhigh frequencies $\omega \gg \Gamma$.

11. CONCLUSIONS

Electrons experience no reflection in passing through a symmetric double-barrier structure if $\varepsilon = \varepsilon_R$, the current reaching its maximum possible value: $I_0(\varepsilon_R) = 1$. Reflection from the structure is suppressed by destructive interference. The large electron density inside the structure is due to constructive interference, the QW width being a multiple of the de Broglie wavelength.

Two questions have long been pursued (see, for example, [8] and the literature cited therein): (i) Is resonant tunneling possible in an alternating field? (ii) What is the maximum ac response that can be achieved by varying the frequency and amplitude of the field and the parameters of the structure? This study provides answers to these questions in sufficient detail. It has been established that resonant tunneling of electrons with energy $\varepsilon = \varepsilon_R + \omega$ also occurs at $\omega \gg \Gamma$. In this case, the reduced current amplitude I_c^R is as large as about $0.41I_0(\varepsilon_R)$ when the optimal field amplitude is $Ua = 2.8\omega$. The field frequency is bounded from above only by energy ε_R of resonance level and may considerably exceed the resonance level width Γ.

If the field is strong enough, $Ua \gg \omega$, then the current oscillates with the field, the maxima being reached when the momentum eE/ω gained by an electron in the field is a multiple of the number of the de Broglie half-waves $\hbar/p$ contained in the QW width. Moreover, the curve current describing current I_c^R as a function of Ua/ω exactly follows the Fraunhofer diffraction pattern from a circular aperture. This exact similarity indicates that electrons absorbing and emitting photons experience interference. The possibility of attaining high alternating currents suggests that the RTD could produce strong oscillations if operated in the resonant tunneling regime. Calculations show that in the so-called quantum regime, when the energy (voltage) is chosen outside the region of negative differential conductance, the output power varies as ω^2; the former was found to be 10^6–10^7 W/cm^2 at $\omega = 10^{13}$ s^{-1}.

In contrast, the classical regime ($\delta < \Gamma$, negative differential conductance region) is characterized by output intensity rapidly decreasing with increasing ω and by the current amplitude differing from that in the resonant regime by a factor of $O((\Gamma/\omega)^3)$. It should be emphasized that as far as we know the focus of research, whether theoretical or experimental, has been on the classical regime.

Written in explicit and simple form, expressions (105) and (106) for currents contain all the results of earlier analytical studies [9–11, 14], the perturbation formulas of Sections 2 and 6, and the numerical results obtained in [12]. Moreover, the expressions (105) are applicable for very strong fields $Ua \gg \omega$, which appears to be unexplored numerically by independent researchers. We are also unaware of independent studies in which explicit analytical formulas are derived for alternating currents in the QW for the case of a strong field; the results obtained in [14] are probably the most relevant in this respect, but alternating currents were not determined. Liou and Roblin [16] numerically calculated the currents through an RTD in a strong field in the classical (nonresonant) regime; their findings are in qualitative agreement with our results.

Concerning the linear theory for weak fields $Ua \ll \Gamma$ (see Section 1), a detailed comparison with other theoretical investigations has been made in our previous studies [9, 10]. The main point is formulated below.

In our opinion, the predictions that resonant tunneling is impossible at $\omega \gg \Gamma$, that the cutoff frequency is about Γ, etc., which were made, for example, in [5, 6], stem from the fact that the problem was solved by using methods that were not quite adequate. The authors followed approximate approaches that do not yield an exact solution to the Schrödinger equation [5] or treated the problem by tunneling Hamiltonian [6] or similar semiphenomenological methods. However, it is essential to correctly include interference and the competition of radiative transitions between different spatial states [10]. The approximate approaches cannot allow for these processes.

A final remark is in order. Although our study was made within a very simple model, we hope that the results are also valid, at least qualitatively, in more realistic contexts. The reason for this is the fundamental character of phenomena determining resonant tunneling. Indeed, a generalization of the model on the basis of numerical methods to the cases of realistic barriers, an electrostatic field, etc. demonstrates coincidence of results for the same level width Γ.

ACKNOWLEDGMENTS

The author is grateful to Yu.V. Kopaev for stimulating discussions and to I.Yu. Kateev and A.I. Podlivaev for numerical calculations and assistance in preparing the manuscript.

This study was conducted within the Federal Program “Integration” (project no. AO133). It was also supported by the Ministry of Industry, Science, and Technologies of the Russian Federation within the program “Physics of Solid-State Nanostructures” (project no. 99-1140) and the project “Theory of Interaction of Strong Electromagnetic Fields with the Electron System of RTDs and Lasers.”

REFERENCES

1. L. Esaki and R. Tsu, Appl. Phys. Lett. **22**, 562 (1973).

2. L. L. Chang, L. Esaki, and R. Tsu, Appl. Phys. Lett. **24**, 593 (1974).
3. T. C. L. G. Sollner, P. E. Tannenwald, D. D. Peck, and W. D. Goodhue, Appl. Phys. Lett. **45**, 1319 (1984).
4. E. R. Brown, J. R. Södestrom, C. D. Parker, *et al.*, Appl. Phys. Lett. **58**, 2291 (1991).
5. H. C. Liu, Phys. Rev. B **43**, 12538 (1991); Erratum: Phys. Rev. B **48**, 4977 (1993).
6. M. P. Anatram and S. Datta, Phys. Rev. B **51**, 7632 (1995).
7. R. K. Mains and G. I. Haddad, J. Appl. Phys. **64**, 3564 (1988); J. Appl. Phys. **64**, 504 (1988).
8. H. C. Liu and T. C. L. G. Sollner, Semicond. Semimet. **41**, 359 (1994).
9. V. F. Elesin, Zh. Éksp. Teor. Fiz. **116**, 704 (1999) [JETP **89**, 377 (1999)].
10. V. F. Elesin, Zh. Éksp. Teor. Fiz. **121**, 925 (2002) [JETP **94**, 794 (2002)].
11. V. F. Elesin, Phys. Low-Dimens. Semicond. Struct. **1**/**2**, 55 (2000).
12. V. F. Elesin, I. Yu. Kateev, and A. I. Podlivaev, Fiz. Tekh. Poluprovodn. (St. Petersburg) **34**, 1373 (2000) [Semiconductors **34**, 1321 (2000)]; Usp. Fiz. Nauk **170**, 333 (2000) [Phys.–Usp. **43**, 291 (2000)].
13. D. Sokolovski and M. Yu. Sumetskij, Teor. Mat. Fiz. **64**, 233 (1985).
14. D. Sokolovski, Phys. Rev. B **37**, 4201 (1988).
15. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 3rd ed. (Fizmatgiz, Moscow, 1960; Pergamon Press, Oxford, 1975).
16. W.-R. Liu and P. Roblin, IEEE Trans. Electron. Dev. **41**, 1098 (1994).

Translated by A. Sharshakov

Journal of Experimental and Theoretical Physics, Vol. 97, No. 2, 2003, pp. 358–401.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 2, 2003, pp. 394–444.

REVIEWS

Superconductivity in Doped Nondegenerate Insulators

A. I. Agafonov* and É. A. Manykin**

Institute of Superconductivity and Solid-State Physics, Russian Research Center Kurchatov Institute, Moscow, 123182 Russia

**e-mail: aai@isssph.kiae.ru*

***e-mail: edmany@isssph.kiae.ru*

Received December 5, 2002

Abstract—An analysis of experimental data on the change in the electronic structure and the position of the chemical potential depending on the doping level in superconducting tungsten bronze Na_xWO_3 and cuprates $La_{2-x}Sr_xCuO_4$ and $Nd_{2-x}Ce_xCuO_4$ demonstrates the importance of the general problem of superconductivity in the impurity bands of doped nondegenerate insulators. The theory of superconductivity in the impurity bands of doped nondegenerate insulators is formulated. The approach to a description of a doped insulator is substantiated on the basis of Holstein–Anderson (Fröhlich–Anderson) model with electron correlations at impurity sites distributed at random in the initial lattice. An insulator–metal transition observed in the normal phase is caused by attenuation of spin fluctuations in the doped system upon an increase in the doping level and/or temperature. This transition is characterized by the presence of narrow allowed bands in the initial insulator. In contrast to the BCS theory, the equation for the superconducting gap does not arise in the description of such a peculiar superconductivity. Instead, a nontrivial solution to the equation for singlet bosons localized at lattice sites must exist for a superconducting transition. The formation of such bosons is a precursor to the emergence of charged extended bosons in the doped system. In the general case, both singlet and triplet pairing channels are possible in the superconducting state. The spin-triplet channel exists only in the case of a finite spectral density of spin fluctuations in a doped compound. In this case, single-particle Green functions are nondiagonal in the spin index. The results of analysis of the phonon mechanism of superconductivity with a spin-singlet pairing channel are considered. The superconductor–metal transition in a doped compound, which is induced by changes in the temperature and/or the doping level, as well as the isotopic effect, is studied using numerical methods. The results are compared with the available data for HTSC materials. © *2003 MAIK "Nauka/Interperiodica".*

CONTENTS

1. Introduction 358
1.1. Superconducting Tungsten Bronzes 359
1.2. Superconducting Doped Cuprates 360
1.3. The Problem of Superconductivity in Doped Nondegenerate Insulators 363
2. Model of a Doped Insulator 364
3. System of Equations for the Superconducting State 365
4. Insulator–Metal Transition 369
4.1. Normal Green Functions 370
4.2. Numerical Analysis of the Insulator–Metal Transition 371
4.3. Insulator State 372
4.4. Temperature-Induced Insulator–Metal Transition 375
4.5. Doping-Induced Insulator–Metal Transition 377
4.6. Effect of the Allowed Bandwidth of an Insulator on the Insulator–Metal Transition 378
5. Analysis of the Triplet Pairing Channel in the Superconducting State 380
6. Phonon-Mediated Mechanism of Superconductivity 380
6.1. Numerical Analysis of the Superconductor–Metal Transition 382
6.2. The State of a Poor Paramagnetic Metal 384
6.3. Superconducting State 385
6.4. Temperature-Induced Superconductor–Metal Transition 387
6.5. Doping-Induced Superconductor–Metal Transition 391
6.6. Doping Dependence of the Superconducting Transition Temperature at Large Doping Levels 393
6.7. Electron Photoemission Spectra 394
6.8. Isotopic Effect 395
7. Conclusions 398
References 399

1. INTRODUCTION

The development of the theory of superconductivity in metals has stimulated to a considerable extent the search for and construction of the theory of supercon-

ductivity in doped semiconductors (insulators) [1, 2]. These studies were based on the possibility of degeneracy for an insulator with a high doping level [1]. However, the superconducting transition temperature T_c predicted by the BCS theory for degenerate insulators is relatively low. This is due to the fact that such materials are characterized by low electron concentrations and densities of electron states at the Fermi level as compared to metals. In addition, the attenuation of quasiparticles in the vicinity of the Fermi surface can be quite strong for high doping levels.

It should be noted that the term "degenerate semiconductors" is applied to superconductors for which the Fermi level either lies in the conduction band or the valence band, or it is in the immediate vicinity of the edges of these bands at distances on the order of k_BT [2–4]. If these conditions are not observed for the position of the Fermi level, a doped semiconductor (insulator) is called nondegenerate. However, the Fermi gas can be degenerate in doped nondegenerate semiconductors for which the Fermi level lies in impurity bands. In a degenerate electron system, the mean distance between particles, which decreases upon an increase in the doping level, must be on the order of (or smaller than) the mean de Broglie wavelength, which can increase with decreasing temperature *T*.

The first object of investigation of superconductivity in doped superconductors was apparently compound $SrTiO_3$ [5]. It was found that the value of T_c in strontium titanate depends on the charge carrier concentration. The highest value of $T_c = 0.43$ K was attained by doping this compound with niobium [6]. After this, superconductivity in doped SnTe and GeTe was discovered [7]. At present, at least two more families of doped insulators (in addition to metals) are known to exhibit superconducting properties. A transition metal element and oxygen are parts of these compounds. The first family of oxide systems, which is known as tungsten bronze, includes doped compounds M_xWO_3, where M stands for elements of groups I, II, and III in the Mendeleev periodic table and $0 < x < 1$. The second family includes doped layered cuprates with CuO_2 planes as a common structural element.

Many of the known models of superconductivity in doped tungsten bronzes and cuprates have the same foundation as in the theory of superconductivity in metals. It is based on the assumption that doping leads to the emergence of charge carriers, which are not connected with impurity ions and can migrate freely in the material. This leads to metallization of the material as a result of its doping. For example, the valence electron of sodium in Na_xWO_3 is not connected to the Na^+ ion and can freely migrate to the conduction band of WO_3. Accordingly, the most used approach for describing superconductivity in tungsten bronzes was based on the model of superconductivity in degenerate semiconductors (insulators) with a degenerate Fermi system. In the case of cuprates, it is usually assumed that doping leads to the emergence of charge carriers which migrate freely to cuprate layers; as a result, the number of charge carriers per unit cell of the CuO_2 lattice changes. In this case, theoretical results on the change in the electron structure upon variable filling of the CuO_2 lattice immediately lead to the conclusion that doping must lead to metallization of either the upper Hubbard band in the case of *n*-type doping, or of the first filled band of the initial insulator in the case of *p*-type doping.

However, the superconducting transition scenario must match the experimental data on the change in the electron structure and the position of the chemical potential depending on the doping level in superconducting tungsten bronzes and cuprates. These results are of fundamental importance for choosing an appropriate model for describing superconductivity in doped insulators since it becomes possible to determine directly the part of the band structure of a doped insulator in which a superconducting band appears. We will briefly review in the Introduction the available experimental results for tungsten bronzes and cuprates. These results indicate that at least Na_xWO_3, $La_{2-x}Sr_xCuO_4$, and $Nd_{2-x}Ce_xCuO_4$ are not degenerate insulators in the doping regions corresponding to the emergence of superconductivity in these compounds. It should be emphasized that, in the case of cuprates, this conclusion is equivalent to the statement that, upon doping, the chemical potential does not leave the region of the insulator gap in the initial ($x = 0$) electron-correlated material.

1.1. Superconducting Tungsten Bronzes

A WO_3 crystal is a insulator with a gap of 2.77 eV [8]. Doped M_xWO_3 compounds, where M = Na, K, Rb, Cs, Ca, Sr, Ba, In, and Tl, are crystallized in several different symmetries depending on the doping level, the radius, and the valence of the interstitial ion [9, 10]. The structural element in these materials is the octahedron WO_6 and impurity atoms occupy interstitial positions in the lattice. The interest in these doped oxide systems [11–16] is due to the facts that these systems (i) pass from the insulator to the metallic state for certain values of the doping index *x* and (ii) are superconducting in the tetragonal and hexagonal phases.

Superconductivity was first observed in the tetragonal phase I of $Na_{0.30}WO_3$ with the superconducting transition temperature $T_c \approx 0.5$ K [11]. Subsequent studies showed that the value of T_c in this phase (existing for $0.2 < x < 0.5$) increases almost exponentially with decreasing doping level, attaining the value $T_c \approx 3$ K at the boundary of the structural transition at $x = 0.2$ [17]. It is important to note that the insulator–metal transition is also realized in this phase for $x_{DM} = 0.25$ [18]. For $x < 0.2$, there is a structural transition to the tetragonal phase II in which superconductivity in Na_xWO_3 was not observed. This structural transition and the difference in tetragonal phases was discussed in [17].

Later, a superconducting transition were detected in the hexagonal phase formed as a result of certain ordering of WO_6 octahedra, in other tungsten bronzes M_xWO_3 doped with elements M = K, Rb, Cs, Ca, Sr, Ba, In, and Tl with a superconducting transition temperature in the range from 1 to 7 K [11–13, 15].

Numerous experiments were made to study the properties in the normal (insulator and metallic) and superconducting states of these compounds. However, the theoretical basis for the observed phenomena was quite contradictory. Mott [19] proposed a model according to which sodium in Na_xWO_3 is ionized and gives away a 3*s* electron to the conduction band of WO_3 formed by hybridized $5d(t_{2g})$ orbitals of W and 2*p* orbitals of O. The insulator–metal transition was explained by a strong scattering of quasiparticles in the conduction band and their Anderson localization in the field of disordered charge distribution, associated with Na^+ ions. The emergence of the superconducting transition in doped materials was essentially based on the theory of superconductivity in degenerate superconductors in accordance with the phonon-mediated mechanism [2].

Among all elements denoted by M and exhibiting the superconducting state in the M_xWO_3 compound, the most detailed experimental studies of the variation of the electron structure and the position of the chemical potential depending on the doping level were carried out for Na_xWO_3 and $Na_xTa_yW_{1-y}O_3$ [18, 20]. Photoelectron spectroscopic data revealed [18] that the impurity band in the insulator gap of the initial material appears as a result of doping in the insulator state of these compounds. These impurity states exhibit the pinning of the Fermi level in the insulator as well as in the metallic state in the vicinity of the insulator–metal transition. As the doping level rises, this impurity band starts overlaping the conduction band. The disagreement between the photoelectron spectroscopic data for Na_xWO_3 and the model of a doped degenerate semiconductor was also discussed in [20].

The possibility of the formation of a superconducting phase in a surface layer of $Na_{0.05}WO_3$ at a temperature approximately equal to 90 K was demonstrated in [21]. It was found [22] that at low temperatures ($T <$ 90 K), the temperature dependence of spin–lattice relaxation in $Na_{0.05}WO_3$ is qualitatively the same as the dependence predicted by the BCS theory for a gap of $\Delta = 160$ K. The photoelectron spectrum of filled states in a doped material is also given in [21]. A broad peak of the spectral density at an energy exceeding the upper edge of the valence band by approximately 1.5 eV was reported. Apparently, it is these states in the insulator gap that must determine the physics of the observed phenomena.

Thus, the above results indicate that Na_xWO_3 is not a degenerate insulator in the doping region corresponding to the superconducting state on the phase diagram for this compound. The emergence of superconductivity in doped Na_xWO_3 is associated with the formation of impurity bands in the insulator gap in the initial WO_3 crystal. Consequently, a correct model of the superconducting state in M_xWO_3 must lead to the emergence of a superconducting gap in these impurity bands. Since an insulator–metal transition also takes place in the case of doping, these impurity bands must be formed by both extended and localized states.

We are not aware of detailed studies on the rearrangement of the electron structure induced by doping for other superconducting M_xWO_3 tungsten bronzes, where M = K, Rb, Cs, Ca, Sr, Ba, In, and Tl. Probably, valence electrons for some of these elements are weakly coupled with the corresponding impurity ions and can freely migrate to the allowed bands of the initial insulator, leading to its degeneracy. However, this mechanism apparently cannot be universal for all the elements. At any rate, this doping mechanism is absent in the case of sodium.

1.2. Superconducting Doped Cuprates

The discovery of high-temperature superconductivity in doped cuprates immediately gave rise to a theory [23] two of three concepts of which are used most frequently even in contemporary theoretical models. These two concepts can be formulated as follows [24, 25]: (i) high-temperature superconductivity is determined by processes occurring in CuO_2 layers; (ii) doping supplies charge carriers which migrate freely in these layers, changing the number of charge carriers per unit cell of the CuO_2 lattice. These processes can be described with the help of Hubbard-model Hamiltonians for a strongly correlated electron system in the lattice [24, 25]. The second concept, which essentially determines the doping mechanism in HTSC materials, leads to the scenario of phase transitions controlled by variable filling of cuprate planes.

Initial cuprates are in the antiferromagnetic insulator state. In the case of doping, the antiferromagnetic ordering disappears soon (e.g., for $x \approx 0.02$ for $La_{2-x}Sr_xCuO_4$). As the value of x increases further, a smooth transition to the state of a poor metal with a high value of T_c takes place. In the $d_{x^2-y^2}$ Hubbard model used for describing CuO_2 planes, the Mott–Hubbard insulator state is formed in the case of half-filling (in the absence of doping) and the chemical potential lies in the insulator gap whose emergence is associated with electron correlations on $d_{x^2-y^2}$ orbitals of copper. In the *d–p* model, a comprehensive idea is given for $3d_{x^2-y^2}$ and $2p_\sigma$ orbitals of cuprate planes and it is possible to describe the initial insulator state in which the chemical potential also lies in the insulator gap. Since this model takes into account the hybridization of

$3d_{x^2-y^2}$ and $2p_\sigma$ orbitals in addition to their electron correlations, this insulator state is characterized by a charge transfer gap.

If we assume that charge carriers introduced in the material by doping are free, migrate to CuO_2 layers, and change the filling per unit cell in these planes, then we see there are common features among the predictions of the models of strongly correlated electrons concerning the change in the density of single-particle states and the chemical potential depending on the doping level. The predicted features can be formulated as follows [24–30]. In the case of half-filling, a insulator gap appears in the density of single-particle states. Even for low levels of *p*-type doping, the chemical potential is abruptly displaced from the initial position in the gap to the edge of the lower Hubbard band (or of the upper filled band in the *p–d* model; see Fig. 53 in [24] and Fig. 22 in [25]). A further increase in the doping level x leads to a further displacement of the chemical potential to the region of the lower Hubbard band so that $\partial\mu/\partial x < 0$. This jump of potential $\mu(x)$ through the initial insulator gap is accompanied by a considerable reconstruction of the density of states in the lower and upper Hubbard bands: the gap in single-particle states disappears, and the density of states forms a single broad band. However, the states arising in the region of the initial insulator gap are unfilled in the case of *p*-type doping and chemical potential μ is not pinned by these states. In the case of *n*-type doping, potential μ jumps through the initial insulator gap to the edge of the upper Hubbard band. The subsequent dynamics of $\mu(x)$ in the region of this band is obvious ($\partial\mu/\partial x > 0$ and the states in the region of the initial gap are filled completely).

The dynamics of the variation of the single-particle density of states in cuprate planes and of the position of μ upon a change in x has determined the most frequently used trend in the construction of the theory of high-T_c superconductivity. It is based on the assumption that in the case of *p*-type doping, chemical potential μ lies in the region of the initially filled band of the initial insulator state; in other words, *p*-type doping leads to metallization of this band of single-particle states in cuprate planes in the case of their variable filling. A review of the main models developed in this field for explaining the superconducting properties of cuprates is given in [31]. These models differ substantially in the type of interactions in cuprate planes, which are introduced for the electron subsystem taking into account the singularities of the Fermi surface and which may lead to Cooper pairing.

However, the approach to the HTSC problem described above does not agree with the known results on the variation of the single-particle density of states and the position of the chemical potential upon an increase in the doping level in cuprates. For example, soon after the discovery of high-temperature superconductivity, the electron structure of the superconductor $La_{1.8}Sr_{0.2}CuO_4$ obtained from the electron photoemission spectra (PES) and inverse photoemission spectra (IPES) clearly demonstrated the pinning of chemical potential by electron states in the insulator gap (see Fig. 1 in [32]). An analysis of experimental papers on photoelectron spectroscopy of doped cuprates published before 1990 leads to the conclusion [33] that (i) a density of states emerges in the insulator gap upon a transition from the insulator to conducting phase, and the Fermi level is in the bulk of the gap, its position depending only slightly on the doping level; (ii) the level filling function in the vicinity of potential μ is a clearly manifested Fermi distribution; and (iii) the results obtained from the energy–momentum relation show that the states in the gap are inherent in the host material. It was noted later in review [24] that the experimentally observed behavior of the chemical potential in cuprates as a function of the doping level cannot be explained from the point of view of the theoretical models based on the two concepts of the Anderson theory for a strongly correlated electron system in CuO_2 planes with variable filling [see p. 815 in [24]). Analysis of the existing photoelectron spectroscopic data led to the conclusion [24] that potential μ changes weakly with the doping level x, remaining in the energy range corresponding to the position of the insulator gap of the initial cuprates in the normal and superconducting states. Doping leads to the emergence of single-particle states in the region of this gap where the chemical potential pinning takes place [34–38]. It is these states that determine the physics of phase transitions in doped cuprates.

Since the phase diagram depending on the doping level of HTSC materials includes insulator–superconductor–metal transitions, the spectral density arising in the initial insulator gap as a result of doping must be determined by the contribution from the bands of localized as well as extended states; i.e., the Fermi system can be a two-component system. In the initial insulator gap, these bands can be referred to as impurity bands. It should be noted that a variable-range hopping-type conduction typical of doped insulators is observed in the impurity bands of localized states in the insulator phase of $La_{2-x}Sr_xCuO_4$ ($x \le 0.05$) [39, 40]. In addition, the existence of superconductivity and localization also follows from the results of analysis of the Hall effect and conductivity [41] as well as thermal conductivity [42] in doped cuprates for various doping levels x and various temperatures.

It is important to establish the relation between these two types of charge carriers in the gap. This was done in [43] using the results on optical conductivity of $La_{2-x}Sr_xCuO_4$. The total number of charge carriers in the gap states contributing to photoconduction was determined from the sum rule, while the contribution from free charge carriers was estimated from the Drude peak. As a result, approximate dependences of the total number of charge carriers and free charge carriers on x were obtained (see Fig. 11 in [43]). It was found that the

number of free charge carriers in the gap states emerging as a result of doping is by almost an order of magnitude smaller than the total number of charge carriers. It becomes clear that the localization effects play the basic role (see [44]) and the approximation of a one-component Fermi system is insufficient in this case.

In spite of reliable data on the coexistence of localization and superconductivity, the experimentally obtained dependences are analyzed for doped cuprates in a large number of publications only on the basis of models for a one-component Fermi system including only extended electron states. For example, the singularities of Fermi contours as functions of the doping levels are constructed and analyzed on the basis of the one-component model from the results of ARPES [34, 44–48]. If we take into account the coexistence of the bands of localized and extended states in the vicinity of the Fermi surface, the pattern of the Fermi contours and their modification as a result of doping will change considerably. To our knowledge, contributions from localized and extended states to the photoelectron spectra have not been separated, although the emergence of localized states in the spectra in the vicinity of the Fermi surface was noted in [35].

One of the problems widely discussed in recent years concerns the condensation energy of the superconducting state of cuprates and the possibility of "violation" of the optical sum rule [49–53], which reflects the integral relation between real parts of dynamic frequency-dependent conductivities in the normal and superconducting states [54]. Naturally, this sum rule and its variation upon a transition to the superconducting state should be analyzed after a correct model for doped cuprates has been established and the conductivity as a function of the doping level, as well as the frequency and temperature dependence of conductivity, has been calculated. However, this problem, as well as the problem of the origin of high-T_c superconductivity, has not been solved as yet. It was shown for the model of a one-component Fermi system with a strong electron–phonon coupling that the violation of the optical sum rule can be attributed to the temperature dependence of the relaxation rate determining the conductivity in the normal state [55]. For a Fermi system with overlapping bands of localized and extended states, which will be described below, the conductivity has not been calculated. In this model, the characteristic density of localized states is considerably higher than the density of extended states in the vicinity of the Fermi surface and the conductance of such a system cannot be presented as the sum of the contributions from charge carriers of two types. It is known, however, that the frequency-dependent conductivity due to localized states may be proportional to temperature [56] (precisely this type of conductivity dependence was used in [55]).

The theoretical approaches to HTSC systems can also be divided into two groups differing in the spatial charge distribution in cuprate planes. Stripe models (see [57] and the literature cited therein) based on the assumption on spatial nonuniformity of the charge carrier distribution in CuO_2 planes with variable filling were initially proposed for describing the fine structure of the neutron magnetic scattering spectra for doped cuprates [58, 59]. This made it possible to easily explain the observed splitting of the initial Bragg peak for a wave vector $\mathbf{Q} = \{\pm\pi/a, \pm\pi/a\}$ (a is the distance between the nearest copper atoms in the plane) into four peaks for $|\mathbf{Q}| \pm \delta$, where $\delta = x$ (see [57]). Another possibility of describing the observed fine structure of the neutron peak was demonstrated in the model of a one-component Fermi system (only extended states) with a uniform spatial distribution of charge carriers in cuprate planes, which is essentially anisotropic in the space of wave vectors [60, 61]. The basic concept of this model consists in the existence of generalized Van Hove singularities located in the vicinity of wave vectors $\{\pm\pi/a, 0\}$ and $\{0, \pm\pi/a\}$ in the spectra of HTSC systems. An analysis of the spin susceptibility whose imaginary part is proportional to the neutron magnetic scattering cross section has made it possible to explain the fine structure of the neutron peak without using the assumption on stripes. The possibility of the existence of Van Hove singularities in the electron density follows from an analysis of the ARPES photoelectron spectra for HTSC materials (see [60, 61]).

Without refuting the approach developed in [60, 61], we would only like to single out the region of the electron structure in which possible singularities in the density may be significant for the HTSC problem. In accordance with the above analysis of experimental data, this region lies in the initial insulator gap of cuprates.

The dependence of the chemical potential shift in $La_{2-x}Sr_xCuO_4$ (p-type doping) and $Nd_{2-x}Ce_xCuO_4$ (n-type doping) was studied experimentally in [62, 63] using X-ray electron spectroscopy (XPS). It was found that the chemical potential of $La_{2-x}Sr_xCuO_4$ is independent of x up to the optimal doping level $x \le 0.15$. A further increase in the doping level x shifts the chemical potential μ towards the first filled band ($\partial\mu/\partial x < 0$). The total shift of potential μ from $x = 0$ to $x = 0.3$ amounted to only 230 meV, which is much smaller than the insulator gap width (see Fig. 3 in [63]). The same figure shows the result for the shift of μ in $Nd_{2-x}Ce_xCuO_4$ (n-type doping). The chemical potential was shifted towards the first unfilled band ($\partial\mu/\partial x > 0$), and the total shift of μ upon a change from $x = 0$ to $x = 0.3$ amounted to only 220 meV. The obtained values of the chemical potential shift in these materials, which are smaller than the initial insulator gap width ($\approx$2.0 eV for La_2CuO_4 and $\approx$1.5 eV for Nd_2CuO_4) show that the chemical potential in these materials lies in the initial insulator gap in the normal (insulator and metallic) as well as in the superconducting state.

It should be noted, however, that various methods of photoelectron spectroscopy lead to a result of fundamental importance, according to which the shift of μ is

small as compared to the initial gap, although the $\mu(x)$ dependences may differ considerably. For example, the variations in the electron structure as a result of doping of $La_{2-x}Sr_xCuO_4$ were studied in [64] using angular-resolution photoelectron spectroscopy (ARPES). The shift of $\mu(x)$ was determined from the position of the step on the low-temperature spectra for different values of x. It was found that the total shift of potential μ upon a change from $x = 0$ to $x = 0.3$ in the case of p-type doping amounted to approximately 100 meV, but potential μ was shifted towards the first unfilled band ($\partial\mu/\partial x > 0$). The behavior of $\mu(x)$ contradicts the X-ray electron spectroscopic (XPS) data [63]. This may be due to the fact that the correct inclusion of the shift of the inner levels is important in X-ray electron spectroscopy, and the energy resolution in this method is much worse than in the ARPES method in which, however, the surface states of the material may give a significant contribution to the photoelectron spectra. Nevertheless, it is important that both methods give the shifts of $\mu(x)$ which are small as compared to the insulator gap in the entire doping region.

Thus, the results presented above indicate that the chemical potential remains in the region of the initial insulator gap at least for $La_{2-x}Sr_xCuO_4$ (p-type doping) and $Nd_{2-x}Ce_xCuO_4$ (n-type doping). The chemical potential shift due to doping is small as compared to the gap. The superconducting gap appears in the impurity bands, which must be formed by extended as well as localized states. In other words, the Fermi system is a two-component system and the concentration of free charge carriers in the vicinity of the chemical potential is an order of magnitude lower than the concentration of localized charge carriers.

1.3. The Problem of Superconductivity in Doped Nondegenerate Insulators

We described the contradictions between the scenario of phase transitions controlled by variable filling of the unit cell of a insulator and experimental data on the change in the band structure and the position of the chemical potential depending on the doping level in superconducting tungsten bronze Na_xWO_3 and cuprates $La_{2-x}Sr_xCuO_4$ and $Nd_{2-x}Ce_xCuO_4$. The main reason for these contradictions is the mechanism of doping in the theory of phase transitions controlled by variable filling. This doping mechanism can be reduced to the assumption that the charge carriers introduced by an impurity are not bound to impurity ions and can migrate freely in the material.

The rejection of this assumption necessitates an analysis of the general problem of superconductivity of nondegenerate doped insulators. Apparently, HTSC materials are characterized by high densities of single-particle states at the Fermi energy and by relatively low concentrations of charge carriers. Both these positions hold for rather narrow impurity bands with a high density of states. The bands for localized and extended states must coexist in the vicinity of the Fermi surface. Such a situation can be realized in the impurity bands of doped insulators.

In this review, we consider the theory of insulator–metal and superconductor–metal phase transitions in doped nondegenerate insulators. The model of a doped compound used by us is substantiated in Section 2. This model includes the Holstein–Anderson (Fröhlich–Anderson) Hamiltonian with electron correlations at impurity sites distributed at random in the initial lattice. We consider a certain ensemble of initially localized impurity levels, which interact with the phonon field and are hybridized with the initial band states of the insulator. A closed system of equations for the normal and superconducting states in a doped insulator is derived in Section 3 using the method of temperature Green functions. The results of numerical analysis of the insulator–metal transition in the impurity bands of a doped compound, which is induced by changes in temperature and/or the doping level, are given in Section 4. The effect of the insulator band width on the insulator–metal transition is analyzed. It is shown that the allowed energy bands of the initial insulator must be quite narrow for the insulator–metal transition in question. The key moment for the superconducting transition is the formation of local zero-spin bosons at impurity sites with zero spin, which are precursors of the emergence of extended charged bosons in the doped system. In the general case, both the singlet and the triplet pairing channel are possible in the superconducting state. An analysis of the triplet pairing channel is given in Section 5. The spin-triplet channel exists only if the parameter representing spin fluctuations in a doped insulator is finite. The phonon-mediated mechanism of superconductivity with a spin-singlet pairing channel is considered in Section 6. The results of numerical investigation of the superconductor–metal transition in a doped compound, which is induced by variations of temperature and/or the doping level, and the isotopic effect are described.

It is well known that so far final conclusions concerning the nature of high-T_c superconductivity have not been drawn [31]. For this reason, along with a detailed analysis of superconductivity mechanisms in CuO_2 planes with variable filling, various possibilities of the formation of the superconducting state should be investigated. We primarily mean here the well-known studies aimed at the search for superconductivity in semiconductors [2]. However, it is appropriate to analyze the state of the art of the general problem of superconductivity in the impurity bands of doped semiconductors taking into account the role of hybridization in a doped system and Coulomb correlations at impurity orbitals. Such an analysis can be useful in the search for new superconducting compounds (not necessarily on the basis of 2D conductivity in planes). Many of questions touched upon in this review require more detailed

treatment and refining, which is typical of all theoretical approaches that have been developed at present. In this respect, the theory of peculiar superconductivity in doped nondegenerate insulators presented in this review is one of possible theories and is proposed as such for future discussion.

2. MODEL OF A DOPED INSULATOR

For the superconducting doped nondegenerate insulators considered above, the fermion system must be a two-component system. Since the chemical potential lies in the region of the initial insulator gap of the undoped material both in the normal (insulator and metallic) and in the superconducting state, the bands of extended and localized states must be formed in this gap upon doping. The origin of the latter states in the case of doped cuprates was considered in the literature [24, 25, 43] and was associated with apical and/or interstitial O atoms as, for example, in $La_2CuO_{4+\delta}$ [43, 65–69] or with doping impurity atoms, e.g., Sr^{2+}, in the case of substitution of La^{3+} in $La_{2-x}Sr_xCuO_4$ [39, 68, 70–73]. It should be noted that the initial multicomponent cuprates are apparently self-doped compounds with a deviation from stoichiometry on the order of 10% [74].

Since it can expected that the electron–phonon interaction is strong in cuprates, the dynamic properties of small-radius polarons were studied in polaron models with a large number of lattice sites using various analytical and numerical approaches [75–81]. In the most used polaron model, tunnel bonds between sites in a certain periodic lattice are considered as well as the electron–phonon interaction of the Holstein (local mode) or Fröhlich type, leading to "shaking" of electrons at a lattice site [80]. In this case, the initial band states of the insulator are not taken into account. One possible objection to this model is that, for values of model parameters admissible for cuprates, the bands of extended states are so narrow (large effective masses of fermions on the order of $100m_e$) that dynamic fluctuations destroy these bands [79].

The necessary condition for the emergence of the superconducting state is the formation of extended charged bosons. The formation of on-site boson–fermion mixed states could be a sort of a precursor to such a superconducting state [82, 83]. If we consider a certain ensemble of such impurity sites distributed at random in the lattice of a insulator and take into account the jumps of particles over the ensemble of lattice sites, single-particle transitions over the sites will lead to the formation of extended fermion states [84] (insulator–metal transition occurring upon the attainment of a certain threshold doping level; see Section 4), while two-particle transitions of initially localized charged bosons over lattice-site states may lead to the formation of extended charged bosons [85]. We believe that in such an approach to the study of superconductivity in doped nondegenerate insulators, it is important to take into account the hybridization between initially localized single-particle impurity states and the band states of the insulator. This interaction leads to jumps of particles over the ensemble of impurity sites even in the second order in the matrix element of hybridization. First, in contrast to the tunnel (or hopping) mechanism, the exponential smallness associated with overlapping of the wave functions at neighboring sites does not arise. Second, the available ARPES data show that the symmetry of the wave functions of extended single-particle states in the vicinity of the Fermi surface in doped HTSC materials is close to the symmetry of the wave functions of the initial cuprates in the vicinity of the insulator gap [33, 34]. Hybridization between the initial localized impurity-site states and the band states of the insulator makes it possible to easily explain the symmetry of the wave functions of extended states in narrow impurity bands.

The chemical composition of cuprate planes does not change upon doping as, for example, in $La_{2-x}Sr_xCuO_4$. The hybridization of 3*d*-Cu and 2*p*-O mainly determines the initial band structure in the vicinity of the insulator gap. For $x = 0.02$, the magnetic ordering in this compound disappears and a further increase in x leads to insulator–superconductor–metal transitions in the paramagnetic phase of a doped insulator.

Upon the introduction of a substitutional impurity, for example, in $La_{2-x}Sr_xCuO_4$, La^{3+} atoms in LaO layers are randomly replaced by Sr^{2+} atoms. Both valence electrons of Sr are spent for the formation of valence bonds; consequently, the formation of singly filled impurity orbital of the acceptor type can be expected. The cuprate layer CuO_2 in this compound is between two intermediate layers with composition $La_{1-x/2}Sr_{x/2}O$. The centers of impurity orbitals lie outside the CuO_2 plane. The interaction of impurity orbitals with the band states of the CuO_2 plane leads to hybridization between these states. For the sake of simplicity, we confine our analysis to the one-band approximation for the density of states in the CuO_2 plane. Taking into account the strong anisotropy of the physical parameters of cuprates, we will disregard the charge transfer along the c (z) axis, which, however, leads to the formation of a bond between different structural cells in $La_{2-x}Sr_xCuO_4$.

The structure of Nd_2CuO_4 is similar to that of La_2CuO_4 except for the positions of O atoms outside CuO_2 planes. In the case of *n*-type doping, the substitution of Ce^{4+} atoms for Nd^{3+} atoms must lead to the formation of a singly charged donor level in $Nd_{2-x}Ce_xCuO_4$. In the case of La_2CuO_{4+y}, excess oxygen atoms can occupy interstitial positions in the vicinity of CuO_2 planes. In this case, all the valence electrons of O atoms can participate in the formation of impurity bands.

In tungsten bronzes, 5*d* and 2*p* orbitals determine the initial band structure near the insulator gap and an impurity of alkali metals which gives *s* states occupies interstitial states in the tetragonal or hexagonal phase. In the case of an alkaline impurity in M_xWO_3, doping may lead to the formation of an ensemble of singly filled impurity orbitals.

Using the Anderson–Haldane model for describing a doped semiconductor [75, 84–86], we can write the Hamiltonian of the model in question in the form

$$H = \sum_{\mathbf{k},\sigma}\varepsilon_{\mathbf{k}}n_{\mathbf{k}\sigma} + \sum_{j,\sigma}\varepsilon_0 n_{j\sigma} + \sum_j U n_{j\sigma}n_{j,-\sigma} + \sum_{j,\mathbf{k},\sigma}\{V_{\mathbf{k}j}a^+_{\mathbf{k}\sigma}a_{j\sigma} + \text{H.c.}\} + \sum_{j,\sigma,\lambda}\hbar\Omega_\lambda\xi_{j\lambda}n_{j\sigma}\phi_\lambda + \sum_\lambda\hbar\Omega_\lambda\left(b^+_\lambda b_\lambda + \frac{1}{2}\right). \quad (2.1)$$

Here, the number-of-particles operator $n_{\mathbf{k}\sigma} = a^+_{\mathbf{k}\sigma}a_{\mathbf{k}\sigma}$ corresponds to the initial valence band of a insulator with the dispersion relation $\varepsilon_{\mathbf{k}}$ (we are using the one-band approximation for the insulator); $\mathbf{k}$ is the wave vector; $\sigma = \uparrow, \downarrow$ is the spin index; the number-of-particles operator $n_{j\sigma} = a^+_{j\sigma}a_{j\sigma}$ corresponds to the ensemble of sites distributed at random in the initial lattice of the insulator; ε_0 is the initial energy of a localized electron state at a site; j is the site number; U is the electron correlation energy in the site; $V_{\mathbf{k}j}$ is the hybridization matrix element; $\xi_{j\lambda}$ represents the electron–phonon interaction; b^+_λ is the phonon production operator; and $\phi_\lambda = b_\lambda + b^+_\lambda$. In the Fröhlich model, λ should be replaced by the phonon wave vector $\mathbf{q}$, while in the Holstein model of the local mode, we have $\lambda = j$.

Solutions for Hamiltonian (2.1) depend on the initial filling of a site $2A = \langle j|n_{j,\uparrow} + n_{j,\downarrow}|j\rangle$. We consider the case of one electron per impurity orbital: $2A = 1$. The total number of impurity sites $N_s = \sum_j$ can be presented in the form $N_s = xN_t$, where N_t is the total number of states in the initial band of the insulator and x will be treated as the doping level. While studying superconductivity, we will use the Holstein model for the electron–phonon interaction. The result will be qualitatively the same as in the case of the Fröhlich model.

3. SYSTEM OF EQUATIONS FOR THE SUPERCONDUCTING STATE

In the following analysis, we will use the thermodynamic variables with a given total number of fermions in the doped system. This condition makes it possible to determine the chemical potential μ. In the standard representation, the Matsubara Green functions are defined as

$$g_{\sigma\sigma_1}(\nu, \nu_1; \tau - \tau_1) = -\langle T_\tau\{\tilde{a}_{\nu\sigma}(\tau)\tilde{\bar{a}}_{\nu_1\sigma}(\tau_1)\}\rangle,$$

$$f^{(+)}_{\sigma\sigma_1}(\nu, \nu_1; \tau - \tau_1) = \langle T_\tau\{\tilde{\bar{a}}_{\nu\sigma}(\tau)\tilde{\bar{a}}_{\nu_1\sigma}(\tau_1)\}\rangle,$$

where $\nu = j, \mathbf{k}$ and $\langle\ldots\rangle$ indicates the Gibbs averaging.

In the Anderson–Haldane model (2.1), using the Fourier transformation, we obtain the first pair of equations in the superconducting state:

$$g_{\sigma\sigma_1}(\mathbf{k}, \nu; \omega_n) = g^{(0)}_{\sigma\sigma}(\mathbf{k}, \mathbf{k}; \omega_n)\times\left(\delta_{\mathbf{k}\nu}\delta_{\sigma\sigma_1} + \sum_j V_{\mathbf{k}j}g_{\sigma\sigma_1}(j, \nu; \omega_n)\right), \quad (3.1)$$

$$f^{(+)}_{\sigma\sigma_1}(\mathbf{k}, \nu; \omega_n) = -f^{(0)}_{\sigma\sigma}(\mathbf{k}, \mathbf{k}; \omega_n)\sum_j V_{j\mathbf{k}}f^{(+)}_{\sigma\sigma_1}(j, \nu; \omega_n). \quad (3.2)$$

Here, $\nu = j_1, \mathbf{k}_1, \omega_n = (2n + 1)\pi T$, and

$$g^{(0)}_{\sigma\sigma}(\mathbf{k}, \mathbf{k}; \omega_n) = (i\omega_n - \varepsilon_{\mathbf{k}})^{-1}, \quad (3.3)$$

$$f^{(0)}_{\sigma\sigma}(\mathbf{k}, \mathbf{k}; \omega_n) = (i\omega_n + \varepsilon_{\mathbf{k}} - 2\mu)^{-1}. \quad (3.4)$$

In the superconducting state, we have obtained the second pair of equations in the two approximations which were also used in the Eliashberg theory [87]. For example, formula (2.1) leads to the following equation for the normal Green function:

$$\left(\frac{\partial}{\partial\tau} + \varepsilon_0\right)g_{\sigma\sigma_1}(j, \nu; \tau) = -\delta_{j\nu}\delta_{\sigma\sigma_1} - \sum_{\mathbf{k}}V_{j\mathbf{k}}g_{\sigma\sigma_1}(\mathbf{k}, \nu; \tau) + U\langle T_\tau\{\tilde{a}_{j\sigma}(\tau)\tilde{\bar{a}}_{j,-\sigma}(\tau)\tilde{a}_{j,-\sigma}(\tau)\tilde{\bar{a}}_{\nu\sigma_1}(0)\}\rangle + \hbar\Omega\xi\langle T_\tau\{\tilde{a}_{j\sigma}(\tau)\tilde{\phi}(\tau)\tilde{\bar{a}}_{\nu\sigma_1}(0)\}\rangle. \quad (3.5)$$

For two-particle Green functions, the mean-field approximation and the initial electron–phonon vortex in irreducible self-energy parts are used in Eq. (3.5) (see, for example, [87]). Proceeding in the same manner for the anomalous Matsubara Green function and carrying out the Fourier transformation, we obtain the second pair of equations for the superconducting state:

$$g_{\sigma\sigma_1}(j, \nu; \omega_n) = g^{(0)}_{\sigma\sigma}(j, j; \omega_n)\times\left[\delta_{j\nu}\delta_{\sigma\sigma_1} + \sum_{\mathbf{k}}V_{j\mathbf{k}}g_{\sigma\sigma_1}(\mathbf{k}, \nu; \omega_n)\right.$$

$$-\Bigg(Ug_{\sigma,-\sigma}(j,j;+0)+T(\hbar\Omega\xi)^2 \times\sum_{n_1}d^0(\omega_n-\omega_{n_1})g_{\sigma,-\sigma}(j,j;\omega_{n_1})\Bigg)g_{-\sigma,\sigma_1}(j,\nu;\omega_n) +\Bigg(Uf_{\sigma,-\sigma}(j,j;+0)+T(\hbar\Omega\xi)^2 \times\sum_{n_1}d^0(\omega_n-\omega_{n_1})f_{\sigma,-\sigma}(j,j;\omega_{n_1})\Bigg)f^{(+)}_{-\sigma\sigma_1}(j,\nu;\omega_n) +\Bigg(T(\hbar\Omega\xi)^2\sum_{n_1}d^0(\omega_n-\omega_{n_1})f_{\sigma,\sigma}(j,j;\omega_{n_1})\Bigg) \times f^{(+)}_{\sigma,\sigma_1}(j,\nu;\omega_n)\Bigg] \tag{3.6}$$

and

$$f^{(+)}_{\sigma\sigma_1}(j,\nu;\omega_n)=f^{(0)}_{\sigma\sigma}(j,j;\omega_n)\Bigg[-\sum_{\mathbf{k}}V_{\mathbf{k}j}f^{(+)}_{\sigma\sigma_1}(\mathbf{k},\nu;\omega_n) +\Bigg(Uf^{(+)}_{\sigma,-\sigma}(j,j;+0)+T(\hbar\Omega\xi)^2 \times\sum_{n_1}d^0(\omega_n-\omega_{n_1})f^{(+)}_{\sigma,-\sigma}(j,j;\omega_{n_1})\Bigg)g_{-\sigma,\sigma_1}(j,\nu;\omega_n) +\Bigg(Ug_{-\sigma,\sigma}(j,j;+0)+T(\hbar\Omega\xi)^2 \times\sum_{n_1}d^0(\omega_n+\omega_{n_1})g_{-\sigma,\sigma}(j,j;\omega_{n_1})\Bigg)f^{(+)}_{-\sigma,\sigma_1}(j,\nu;\omega_n) +\Bigg(T(\hbar\Omega\xi)^2\sum_{n_1}d^0(\omega_n-\omega_{n_1})f^{(+)}_{\sigma\sigma}(j,j;\omega_{n_1})\Bigg) \times g_{\sigma\sigma_1}(j,\nu;\omega_n)\Bigg]. \tag{3.7}$$

Here, the following notation has been used:

$$d^0(\omega_m=2m\pi T)=(i\omega_m-\Omega)^{-1}-(i\omega_m+\Omega)^{-1},$$

$$g^{(0)}_{\sigma\sigma}(j,j;\omega_n)=\Bigg(i\omega_n-\varepsilon_0-Ug_{-\sigma,-\sigma}(j,j;-0) +T(\hbar\Omega\xi)^2\sum_{n_1}d^0(\omega_n-\omega_{n_1})g_{\sigma\sigma}(j,j;\omega_{n_1})\Bigg)^{-1} \tag{3.8}$$

and

$$f^{(0)}_{\sigma\sigma}(j,j;\omega_n)=\Bigg(i\omega_n+\varepsilon_0+Ug_{-\sigma,-\sigma}(j,j;-0) -T(\hbar\Omega\xi)^2\sum_{n_1}d^0(\omega_n+\omega_{n_1})g_{\sigma\sigma}(j,j;\omega_{n_1})-2\mu\Bigg)^{-1}. \tag{3.9}$$

The system of equations (3.1), (3.2), (3.6), and (3.7) taking into account definitions (3.3), (3.4), (3.8), and (3.9) forms a closed system of equations for the superconducting state of a doped compound. While solving this system of equations, averaging should be carried out over a random distribution $\{\mathbf{R}_j\}$ of the ensemble of impurity sites in the initial lattice [84, 88]. For extended fermion and boson states (as well as for localized states), this system can be reduced to two subsystems of eight equations with eight unknown Green functions. While determining the chemical potential μ, one should match the solution to these subsystems. In addition, the solution to the total system depends self-consistently on three types of parameters determined by Green functions for localized fermion and boson states. The first of these parameters is the site filling

$$A_\sigma=g_{-\sigma,-\sigma}(j,j;-0). \tag{3.10}$$

In the general case, $A_\sigma\neq A_{-\sigma}$ and the solution to this system corresponds to the magnetic phase of the doped compound. This phase will acquire a magnetic moment in view of the difference in the density distribution both for localized and for extended states, as well as in their filling in the two spin subspaces [89].

We will confine our analysis to the paramagnetic state alone, for which $A_\sigma=A_{-\sigma}=A$. The spectral expansion of the Green function has the form

$$g_{\sigma\sigma}(j,j;\omega_n)=\int_{-\infty}^{\infty}\frac{\rho^f_{\mathrm{loc}}(\varepsilon)d\varepsilon}{i\omega_n-\varepsilon+\mu}, \tag{3.11}$$

where $\rho^f_{\mathrm{loc}}(\varepsilon)\geq 0$ is the real-valued density of single-particle localized states per impurity atom, which has the normalization

$$\int_{-\infty}^{\infty}\rho^f_{\mathrm{loc}}(\varepsilon)d\varepsilon=1. \tag{3.12}$$

Using Eqs. (3.10) and (3.11) and applying the sum rule in ω_n [87], we obtain

$$A=\lim_{\tau\to-0}g_{\sigma\sigma}(j,j;\tau)=\int_{-\infty}^{\infty}d\omega\frac{\rho^f_{\mathrm{loc}}(\omega)}{\exp\left(\dfrac{\omega-\mu}{T}\right)+1}. \tag{3.13}$$

The next parameter is the spin-fluctuation parameter λ:

$$\lambda = g_{\downarrow\uparrow}(j, j; +0) = [g_{\uparrow\downarrow}(j, j; +0)]^*. \quad (3.14)$$

This parameter appears when we leave the (self-consistent) Hartree–Fock approximation [86] for the electron-correlated system. This parameter is connected with the nondiagonal form of the Green functions in the spin variable. Using the representation

$$g_{\downarrow\uparrow}(j, j; \omega_n) = \int_{-\infty}^{\infty} \frac{\rho_{\mathrm{loc}}^{sf}(\varepsilon)d\varepsilon}{i\omega_n - \varepsilon + \mu}, \quad (3.15)$$

where the real-valued density of spin fluctuations has the property

$$\int_{-\infty}^{\infty} \rho_{\mathrm{loc}}^{sf}(\varepsilon)d\varepsilon = 0, \quad (3.16)$$

we obtain

$$\lambda = \int_{-\infty}^{\infty} d\omega \frac{\rho_{\mathrm{loc}}^{sf}(\omega)}{\exp\left(\frac{\omega - \mu}{T}\right) + 1}. \quad (3.17)$$

The last important parameter is determined by the quantity β whose squared modulus gives the probability of a localized boson with zero spin located at the jth site:

$$\beta = f_{\downarrow\uparrow}^{(+)}(j, j; +O) = [f_{\uparrow\downarrow}(j, j; +0)]^*. \quad (3.18)$$

In contrast to parameters A and λ, which exist both for the normal and for the superconducting state, parameter β determines the possibility for the emergence of superconductivity in a doped compound. Using the representation

$$f_{\downarrow\uparrow}^{(+)}(j, j; \omega_n) = \int_{-\infty}^{\infty} \frac{\rho_{\mathrm{loc}}^{b0}(\varepsilon)d\varepsilon}{i\omega_n - \varepsilon + \mu}, \quad (3.19)$$

where the real-valued boson spectral density has the property

$$\int_{-\infty}^{\infty} \rho_{\mathrm{loc}}^{b0}(\varepsilon)d\varepsilon = 0, \quad (3.20)$$

we obtain

$$\beta = \int_{-\infty}^{\infty} d\omega \frac{\rho_{\mathrm{loc}}^{b0}(\omega)}{\exp\left(\frac{\omega - \mu}{T}\right) + 1}. \quad (3.21)$$

For the paramagnetic superconducting state, we introduce the self-energy parts

$$\Sigma_{\pm}^{n}(\omega_n) = -T(\xi\hbar\Omega)^2 \times \sum_{n_1} d^0(\omega_n \pm \omega_{n_1}) g_{\sigma\sigma}(j, j; \omega_{n_1}), \quad (3.22)$$

$$\Sigma_{\pm}^{f}(\omega_n) = -T(\xi\hbar\Omega)^2 \times \sum_{n_1} d^0(\omega_n \pm \omega_{n_1}) g_{-\sigma\sigma}(j, j; \omega_{n_1}), \quad (3.23)$$

$$\Delta_{\pm}^{s}(\omega_n) = -T(\xi\hbar\Omega)^2 \times \sum_{n_1} d^0(\omega_n \pm \omega_{n_1}) f_{\downarrow\uparrow}^{(+)}(j, j; \omega_{n_1}), \quad (3.24)$$

$$\Delta^{t}(\omega_n) = -T(\xi\hbar\Omega)^2 \times \sum_{n_1} d^0(\omega_n - \omega_{n_1}) f_{\sigma\sigma}^{(+)}(j, j; \omega_{n_1}). \quad (3.25)$$

Taking into account relations (3.10) and (3.22), we can reduce definitions (3.8) and (3.9) to the form

$$g^{(0)}(l, l; \omega_n) = (i\omega_n - \varepsilon_0 - UA - \Sigma_{-}^{n}(\omega_n))^{-1} \quad (3.26)$$

and

$$f^{(0)}(l, l; \omega_n) = (i\omega_n + \varepsilon_0 + UA + \Sigma_{+}^{n}(\omega_n) - 2\mu)^{-1}. \quad (3.27)$$

The system of equations for extended states in the paramagnetic phase is reduced to the form

$$g_{\uparrow\uparrow}(\mathbf{k}, \mathbf{k}_1) = g^{(0)}(\mathbf{k}, \mathbf{k})\left(\delta_{\mathbf{k}\mathbf{k}_1} + \sum_j V_{\mathbf{k}j} g_{\uparrow\uparrow}(j, \mathbf{k}_1)\right), \quad (3.28)$$

$$g_{\downarrow\uparrow}(\mathbf{k}, \mathbf{k}_1) = g^{(0)}(\mathbf{k}, \mathbf{k}) \sum_j V_{\mathbf{k}j} g_{\downarrow\uparrow}(j, \mathbf{k}_1), \quad (3.29)$$

$$g_{\uparrow\uparrow}(j, \mathbf{k}) = g^{(0)}(l, l) \times \left(\sum_{\mathbf{k}_1} V_{j\mathbf{k}_1} g_{\uparrow\uparrow}(\mathbf{k}_1, \mathbf{k}) - (\lambda U - \Sigma_{-}^{f}) g_{\downarrow\uparrow}(j, \mathbf{k}) + (\beta U - \Delta_{-}^{s}) f_{\downarrow\uparrow}^{(+)}(j, \mathbf{k}) - \Delta^{t} f_{\uparrow\uparrow}^{(+)}(j, \mathbf{k})\right), \quad (3.30)$$

$$g_{\downarrow\uparrow}(j, \mathbf{k}) = g^{(0)}(l, l) \times \left(\sum_{\mathbf{k}_1} V_{j\mathbf{k}_1} g_{\downarrow\uparrow}(\mathbf{k}_1, \mathbf{k}) - (\lambda U - \Sigma_-^f) g_{\uparrow\uparrow}(j, \mathbf{k}) - (\beta U - \Delta_+^s) f_{\uparrow\uparrow}^{(+)}(j, \mathbf{k}) - \Delta^t f_{\downarrow\uparrow}^{(+)}(j, \mathbf{k}) \right), \quad (3.31)$$

$$f_{\downarrow\uparrow}^{(+)}(j, \mathbf{k}) = f^{(0)}(l, l) \times \left(-\sum_{\mathbf{k}_1} V_{\mathbf{k}_1 j} f_{\downarrow\uparrow}^{(+)}(\mathbf{k}_1, \mathbf{k}) + (\beta U - \Delta_-^s) g_{\uparrow\uparrow}(j, \mathbf{k}) + (\lambda U - \Sigma_+^f) f_{\uparrow\uparrow}^{(+)}(j, \mathbf{k}) - \Delta^t g_{\downarrow\uparrow}(j, \mathbf{k}) \right), \quad (3.32)$$

$$f_{\uparrow\uparrow}^{(+)}(j, \mathbf{k}) = f^{(0)}(l, l) \times \left(-\sum_{\mathbf{k}_1} V_{\mathbf{k}_1 j} f_{\uparrow\uparrow}^{(+)}(\mathbf{k}_1, \mathbf{k}) - (\beta U - \Delta_+^s) g_{\downarrow\uparrow}(j, \mathbf{k}) + (\lambda U - \Sigma_+^f) f_{\downarrow\uparrow}^{(+)}(j, \mathbf{k}) - \Delta^t g_{\uparrow\uparrow}(j, \mathbf{k}) \right), \quad (3.33)$$

$$f_{\downarrow\uparrow}^{(+)}(\mathbf{k}, \mathbf{k}_1; \omega_n) = -f^{(0)}(\mathbf{k}, \mathbf{k}) \sum_j V_{j\mathbf{k}} f_{\downarrow\uparrow}^{(+)}(j, \mathbf{k}_1; \omega_n), \quad (3.34)$$

$$f_{\uparrow\uparrow}^{(+)}(\mathbf{k}, \mathbf{k}_1; \omega_n) = -f^{(0)}(\mathbf{k}, \mathbf{k}) \sum_j V_{j\mathbf{k}} f_{\uparrow\uparrow}^{(+)}(j, \mathbf{k}_1; \omega_n). \quad (3.35)$$

The system of equations for localized states in the paramagnetic phase assumes the form

$$g_{\uparrow\uparrow}(j, j_1) = g^{(0)}(l, l) \times \left(\delta_{jj_1} + \sum_{\mathbf{k}} V_{j\mathbf{k}} g_{\uparrow\uparrow}(\mathbf{k}, j_1) - (\lambda U - \Sigma_-^f) g_{\downarrow\uparrow}(j, j_1) + (\beta U - \Delta_-^s) f_{\downarrow\uparrow}^{(+)}(j, j_1) - \Delta^t f_{\uparrow\uparrow}^{(+)}(j, j_1) \right), \quad (3.36)$$

$$g_{\uparrow\uparrow}(\mathbf{k}, j) = g^{(0)}(\mathbf{k}, \mathbf{k}) \sum_{j_1} V_{\mathbf{k}j_1} g_{\uparrow\uparrow}(j_1, j), \quad (3.37)$$

$$g_{\downarrow\uparrow}(j, j_1) = g^{(0)}(l, l) \times \left(\sum_{\mathbf{k}} V_{j\mathbf{k}} g_{\downarrow\uparrow}(\mathbf{k}, j_1) - (\lambda U - \Sigma_-^f) g_{\uparrow\uparrow}(j, j_1) - (\beta U - \Delta_+^s) f_{\uparrow\uparrow}^{(+)}(j, j_1) - \Delta^t f_{\downarrow\uparrow}^{(+)}(j, j_1) \right), \quad (3.38)$$

$$g_{\downarrow\uparrow}(\mathbf{k}, j) = g^{(0)}(\mathbf{k}, \mathbf{k}) \sum_{j_1} V_{\mathbf{k}j_1} g_{\downarrow\uparrow}(j_1, j), \quad (3.39)$$

$$f_{\downarrow\uparrow}^{(+)}(j, j_1) = f^{(0)}(l, l) \times \left(-\sum_{\mathbf{k}} V_{\mathbf{k}j} f_{\downarrow\uparrow}^{(+)}(\mathbf{k}, j_1) + (\beta U - \Delta_-^s) g_{\uparrow\uparrow}(j, j_1) + (\lambda U - \Sigma_+^f) f_{\uparrow\uparrow}^{(+)}(j, j_1) - \Delta^t g_{\downarrow\uparrow}(j, j_1) \right), \quad (3.40)$$

$$f_{\downarrow\uparrow}^{(+)}(\mathbf{k}, j; \omega_n) = -f^{(0)}(\mathbf{k}, \mathbf{k}) \sum_{j_1} V_{j_1\mathbf{k}} f_{\downarrow\uparrow}^{(+)}(j_1, j; \omega_n), \quad (3.41)$$

$$f_{\uparrow\uparrow}^{(+)}(j, j_1) = f^{(0)}(l, l) \times \left(-\sum_{\mathbf{k}} V_{\mathbf{k}j} f_{\uparrow\uparrow}^{(+)}(\mathbf{k}, j_1) - (\beta U - \Delta_+^s) g_{\downarrow\uparrow}(j, j_1) + (\lambda U - \Sigma_+^f) f_{\downarrow\uparrow}^{(+)}(j, j_1) - \Delta^t g_{\uparrow\uparrow}(j, j_1) \right), \quad (3.42)$$

$$f_{\uparrow\uparrow}^{(+)}(\mathbf{k}, j; \omega_n) = -f^{(0)}(\mathbf{k}, \mathbf{k}) \sum_{j_1} V_{j_1\mathbf{k}} f_{\uparrow\uparrow}^{(+)}(j_1, j; \omega_n). \quad (3.43)$$

Thus, eight equations (3.28)–(3.35) for extended fermion and boson states and eight equations (3.36)–(3.43) for localized states taking into account the self-energy parts (3.22)–(3.25) introduced above; definitions (3.3), (3.4), (3.26), and (3.27), as well as conditions (3.13), (3.17), and (3.21) for matched parameters A, λ, and β, respectively; and chemical potential μ determined from the condition of conservation of the total number of fermions $(1 + x)N_t$ completely determine the superconducting state in model (2.1) of the nondegenerate doped insulator analyzed here. In contrast to the theory of superconductivity in metals, these equations do not include the equation for the superconducting gap. Nevertheless, it will be shown below that

this gap appears in the density of single-particle states in the superconducting state.

If Eq. (3.21) has only the trivial solution $\beta = 0$, it can easily be seen that the state of the doped material is normal. In this state, which is defined by parameters μ, A, and λ, the electron structure is determined only by single-particle normal Green functions for localized and extended electron states. These normal Green functions have poles in the region of the initial band as well as in the region of the initial insulator gap, where impurity bands of localized as well as extended states are formed. It was noted above that the formation of impurity bands of extended single-particle states is associated with hybridization leading to one-electron transitions initial impurity site ⟶ band state ⟶ another site ⟶ band state, and so on.

The emergence of the superconducting state is determined by the existence of only a nontrivial solution $\beta \neq 0$ to Eq. (3.21) for the anomalous singlet function $f^{(+)}_{\downarrow\uparrow}(j, j)$. Such a solution indicates the formation of localized bosons with zero spin. In addition to these bosons, localized bosons with spin 1 must also be present in accordance with Eqs. (3.34)–(3.41). It is this parameter β which specifies Green functions $f^{(+)}_{\downarrow\uparrow}(\mathbf{k}, \mathbf{k}_1)$ for extended charged bosons with zero spin and $f^{(+)}_{\uparrow\uparrow}(\mathbf{k}, \mathbf{k}_1)$ for extended charged bosons with spin 1 (see Eqs. (3.26)–(3.33)). As a result, the superconducting state in the general case is characterized by pairing of quasiparticles both in the singlet and in triplet channels.

Thus, this theory of superconductivity in doped insulators does not contain a BCS-type equation for the energy gap. Instead, the key equation for the superconducting state is Eq. (3.21) describing the formation of localized bosons with zero spin in the doped system. Such bosons can be regarded as precursors for the emergence of superconductivity in the system since hybridization-induced two-quasiparticle transitions over the impurity ensemble (a localized boson at an impurity site ⟶ pair of coupled quasiparticles $(-\mathbf{k}, \mathbf{k})$ in the impurity band of extended states ⟶ localized boson at another impurity site, etc.) may lead to the formation of a superconducting condensate.

The symmetry of a doped system does not include spatial inversion. For this reason, the superconducting state is mixed in the general case. This means that both singlet and triplet pairing channels for quasiparticles may be opened. In Section 5, we will demonstrate the relation between the spin-triplet channel and spin fluctuations in a doped system; i.e., the formation of a triplet channel is always associated with a purely electronic mechanism of the emergence of the superconducting state.

Spin fluctuations represented by parameter λ attenuate upon an increase in the doping level x and/or temperature (Section 4). If $\lambda = 0$, the superconducting state can arise due to a purely phonon mechanism (Section 6). It will be shown below that in this model superconducting bosons appear with a center of mass at rest only. This is a result of averaging over the ensemble of impurity atoms distributed at random over the initial lattice. However, we will first analyze the phase diagram of the normal state depending on the doping level and temperature.

4. INSULATOR–METAL TRANSITION

In addition to the doped tungsten bronzes and cuprates considered in the Introduction, two families of transition metal oxides $R_{1-x}A_xTiO_3$ and $R_{1-x}A_xVO_3$ (R = La, Nd, Pr, Y; and A is a bivalent alkali-earth element Ca, Sr, or Ba) are known, in which no superconducting transition has been detected, but the insulator–metal transition takes place upon doping. At both limits of doping levels ($x = 0$ and $x = 1$), these stoichiometric compounds are insulators [90]. The initial insulator state of compounds $RTiO_3$ and RVO_3 is determined by strong electron correlations (Mott–Hubbard insulators). The substitution of a bivalent alkali-earth element for a trivalent rare-earth element (*p*-type doping) leads to an insulator–metal transition [90–94]. This transition occurs for $x \approx 0.05$ in $La_{1-x}Sr_xTiO_3$, $x \approx 0.35$ in $Y_{1-x}Ca_xTiO_3$, and $x \approx 0.2$ in $La_{1-x}Sr_xVO_3$.

In the other limit, $x = 1$, these two families of transition metal oxides $ATiO_3$ and AVO_3 also belong to insulators and the substitution of rare-earth element R^{3+} for alkali-earth element A^{2+} (*n*-type doping) also leads to an insulator–metal transition [95, 96]. This transition for compounds $R_xBa_{1-x}TiO_3$ with R = Y, La, Nd occurs for $x \approx 0.2$–0.3, while this transition in $La_xSr_{1-x}TiO_3$ occurs in $x \approx 0.08$.

In a recent review [25], compounds $R_{1-x}A_xTiO_3$ and $R_{1-x}A_xVO_3$ were classified as electron-correlated systems with a phase transition controlled by variable filling of the band structure. The band structure was determined on the basis of the Hubbard model taking into account $3d^1$ orbitals of Ti^{3+} and $2p$ orbitals of oxygen. It was assumed that doping changes orbital fillings; i.e., the theoretical analysis in [25] was based on the two concepts of the Anderson theory discussed in the Introduction.

It should be noted that photoelectron spectroscopic methods were used for obtaining data on the variation of the band structure and the position of the chemical potential in compounds $R_{1-x}A_xTiO_3$ and $R_{1-x}A_xVO_3$ upon doping [90–94]. The results presented in these publications clearly indicate the formation of electron states in the region of the initial insulator gap. The formation of gap states was observed upon doping with rare-earth elements R of the initial band insulator $ATiO_3$ [95] as well as upon doping of the Mott insulator $RTiO_3$ with alkali-earth elements A [96]. It was concluded [96] that the change in the spectral density observed in the

vicinity of the chemical potential in the region of the initial insulator gap contradicts the results of the Hubbard model. The scenario of the reconstruction of the density of states in the Hubbard model was discussed in the Introduction.

Naturally, various types of insulator–metal transitions may take place in the impurity states (in particular, the Mott transition associated with overlapping of atomic orbitals upon an increase in the doping level). The possibility of the Mott transition in doped oxides of transition metals was discussed in [97]. For this mechanism, the effective Bohr radius a_{eff} of the impurity orbital is essential. The effective radius of doped cuprates in the insulator state was estimated in [39, 71, 97–101] as $a_{\text{eff}} = 4$–8 Å. At the lower boundary of this estimate corresponding to small-radius polarons or deep impurity levels with a_{eff} on the order of the atomic spacing, the Mott criterion may be violated.

We will consider below the results for the insulator–metal phase transition predicted in model (2.1). This transition is due to the formation of a pseudogap in the spectrum of single-particle states in a doped compound and is induced by a strong rearrangement of the impurity bands of localized and extended states upon a change in temperature and/or doping level.

4.1. Normal Green Functions

For the paramagnetic normal state, the system of equations (3.28)–(3.35) for extended states can be reduced to the form

$$g_{\sigma\sigma_1}(\mathbf{k}, \mathbf{k}_1) = g^{(0)}(\mathbf{k}, \mathbf{k}) \times\left(\delta_{\mathbf{k}\mathbf{k}_1}\delta_{\sigma\sigma_1} + \sum_j V_{\mathbf{k}j} g_{\sigma\sigma_1}(j, \mathbf{k}_1)\right), \tag{4.1}$$

$$g_{\sigma\sigma_1}(j, \mathbf{k}) = g^{(0)}(l, l) \times\left(\sum_{\mathbf{k}_1} V_{j\mathbf{k}_1} g_{\sigma\sigma_1}(\mathbf{k}_1, \mathbf{k}) - (\lambda U - \Sigma_-^f) g_{-\sigma,\sigma_1}(j, \mathbf{k})\right). \tag{4.2}$$

Similarly, for localized states in the normal phase, we obtain from the system of equations (3.36)–(3.43)

$$g_{\sigma\sigma_1}(j, j_1) = g^{(0)}(l, l) \times\left(\delta_{jj_1}\delta_{\sigma\sigma_1} + \sum_{\mathbf{k}} V_{j\mathbf{k}} g_{\sigma\sigma_1}(\mathbf{k}, j_1) - (\lambda U - \Sigma_-^f) g_{-\sigma,\sigma_1}(j, j_1)\right), \tag{4.3}$$

$$g_{\sigma\sigma_1}(\mathbf{k}, j) = g^{(0)}(\mathbf{k}, \mathbf{k}) \sum_{j_1} V_{\mathbf{k}j_1} g_{\sigma\sigma_1}(j_1, j). \tag{4.4}$$

The matrix element of hybridization has the form

$$V_{\mathbf{k}j} = V_{\mathbf{k}l} \exp(i\mathbf{k}\cdot\mathbf{R}_j), \tag{4.5}$$

where $\mathbf{R}_j$ is the radius vector of the jth impurity center.

We will solve Eqs. (4.1), (4.2) and (4.3), (4.4) by using the following approach. We write Eq. (4.4) for localized states in the form

$$g^{(0)^{-1}}(\mathbf{k}, \mathbf{k}) g_{\sigma\sigma_1}(\mathbf{k}, j) = V_{\mathbf{k}j} g_{\sigma\sigma_1}(j, j) + \sum_{j_1 \neq j} V_{\mathbf{k}j_1} g_{\sigma\sigma_1}(j_1, j). \tag{4.6}$$

Substituting into this equation the matrix element $g_{\sigma\sigma_1}(j_1, j)$, which is nondiagonal in j and is defined by Eq. (4.3), and using Eq. (4.6) for $-\sigma\sigma_1$ spin matrix elements of Green functions, we obtain

$$\begin{aligned}&(g^{(0)^{-1}}(\mathbf{k}, \mathbf{k}) - xN_t V_{\mathbf{k}l}^2 g^{(0)}(l, l)) g_{\sigma\sigma_1}(\mathbf{k}, j) \\ &+ (\lambda U - \Sigma_-^f) g^{(0)}(l, l) g^{(0)^{-1}}(\mathbf{k}, \mathbf{k}) g_{-\sigma,\sigma_1}(\mathbf{k}, j) \\ &= V_{\mathbf{k}j}(g_{\sigma\sigma_1}(j, j) + (\lambda U - \Sigma_-^f) g^{(0)}(l, l) g_{-\sigma,\sigma_1}(j, j)) \\ &+ g^{(0)}(l, l) \sum_{j_1 \neq j} \sum_{\mathbf{k}_1 \neq \mathbf{k}} V_{\mathbf{k}j_1} V_{j_1\mathbf{k}_1} g_{\sigma\sigma_1}(\mathbf{k}_1, j).\end{aligned} \tag{4.7}$$

Carrying out summation over j_1 in the last term on the right-hand side of this equation, we must perform averaging over a random distribution of impurity centers. Using the configuration averaging technique [84, 88], we obtain

$$\left\langle \sum_{j_1 \neq j, \mathbf{k}_1 \neq \mathbf{k}} V_{\mathbf{k}j_1} V_{j_1\mathbf{k}_1} \right\rangle_{av} \propto x \sum_{\mathbf{k}_1 \neq \mathbf{k}} \delta(\mathbf{k}_1 - \mathbf{k}) = 0$$

and

$$\left\langle \sum_{j_1 \neq j, \mathbf{k}_1 \neq \mathbf{k}} V_{\mathbf{k}j_1} V_{j_1\mathbf{k}_1} g(\mathbf{k}_1, j; \omega_n) \right\rangle_{av} \propto x^2.$$

The evaluation of the last term on the right-hand side of Eq. (4.7) gives a power series in x starting with x^2. In some cases, we have taken into account the first term of this series and found that the densities of states change insignificantly on the whole in the case of low doping levels, but in the energy ranges where the values of imaginary parts of the self-energy components (3.22) and (3.23) (or (4.11); see below) are small, the inclusion of this contribution led to negative values of the density of states. Assuming that the relative doping levels are such that $x^2 \ll 1$, we will disregard the terms containing double sums.

Equation (4.7) forms the a system of two equations for the spin matrix elements of the Green functions $g_{\uparrow\uparrow}(\mathbf{k}, j)$ and $g_{\downarrow\uparrow}(\mathbf{k}, j)$. This system can be solved easily. As a result, the Green functions can be expressed in terms of j-diagonal functions $g_{\uparrow\uparrow}(j, j)$ and $g_{\downarrow\uparrow}(j, j)$. Substituting these expressions into Eq. (4.3), we obtain a

system of two equations for the Green functions $g_{\uparrow\uparrow}(j, j)$ and $g_{\downarrow\uparrow}(j, j)$. The solution to this system determines the density of localized states and the spectral density of their spin fluctuations. We used a similar approach for solving the system of equations (3.28)–(3.35) for extended states.

Omitting simple but cumbersome computations, we write the final expressions for the sought Green functions determining the normal state of a doped insulator. For localized states, we obtain

$$g_{\sigma\sigma_1}(j, j; \omega_n) = \frac{1}{2} \times (g_{\rm loc}(\omega_n; A-\lambda) + (-1)^{\sigma-\sigma_1} g_{\rm loc}(\omega_n; A+\lambda)), \tag{4.8}$$

where

$$g^{-1}_{\rm loc}(\omega_n; A\pm\lambda) = h(\omega_n; A\pm\lambda) - \sum_{\mathbf{k}} \frac{V^2_{\mathbf{k}l}}{g^{(0)^{-1}}(\mathbf{k}, \mathbf{k}) - xN_t V^2_{\mathbf{k}l} h^{-1}(\omega_n; A\pm\lambda)}, \tag{4.9}$$

$$h(\omega_n; A\pm\lambda) = i\omega_n - \varepsilon_0 - U(A\pm\lambda) - \Sigma_-(\omega_n; A\pm\lambda), \tag{4.10}$$

$$\Sigma_-(\omega_n; A\pm\lambda) = -T(\xi\hbar\Omega)^2 \times \sum_{n_1} d^0(\omega_n - \omega_{n_1}) g_{\rm loc}(\omega_{n_1}; A\pm\lambda). \tag{4.11}$$

For extended states, we obtain

$$g_{\sigma\sigma_1}(\mathbf{k}, \mathbf{k}; \omega_n) = \frac{1}{2}(g_{\rm ext}(\mathbf{k}; \omega_n; A-\lambda) + (-1)^{\sigma-\sigma_1} g_{\rm ext}(\mathbf{k}; \omega_n; A+\lambda)), \tag{4.12}$$

where

$$g^{-1}_{\rm ext}(\mathbf{k}; \omega_n; A\pm\lambda) = g^{(0)^{-1}}(\mathbf{k}, \mathbf{k}; \omega_n) - \frac{xN_t V^2_{\mathbf{k}l}}{h(\omega_n; A\pm\lambda) - \sum_{\mathbf{k}} V^2_{\mathbf{k}l} g^{(0)}(\mathbf{k}, \mathbf{k}; \omega_n)}. \tag{4.13}$$

4.2. Numerical Analysis of the Insulator–Metal Transition

We choose for our computations the model of the symmetric valence band with a density $\mathbf{k}$ states per spin

$$\rho^{(0)}(\varepsilon) = \begin{cases} \dfrac{N_t}{\pi D_b^2}[D_b^2 - \varepsilon^2]^{1/2}, & |\varepsilon| \le D_b, \\ 0, & |\varepsilon| > D_b. \end{cases} \tag{4.14}$$

Here, $2D_b$ is the valence band width and N_t is the total number of states in the band.

The analytic continuation of the Matsubara Green functions to the upper part of the complex plane ω makes it possible to obtain the Fourier components of the retarded time Green functions at finite temperatures. The latter make it possible to determine the spectral densities

$$\rho_{\rm loc}(\omega; A\pm\lambda) = -\frac{1}{\pi}{\rm Im} g_{\rm loc}(\omega_n; A\pm\lambda)\big|_{i\omega_n \to \omega + i0}, \tag{4.15}$$

$$\rho_{\rm ext}(\mathbf{k}, \omega; A\pm\lambda) = -\frac{1}{\pi}{\rm Im} g_{\rm ext}(\mathbf{k}; \omega_n; A\pm\lambda)\big|_{i\omega_n \to \omega + i0}. \tag{4.16}$$

Functions (4.15) determine the density of localized states per spin per impurity atom,

$$\rho^f_{\rm loc}(\omega) = \frac{1}{2}(\rho_{\rm loc}(\omega; A-\lambda) + \rho_{\rm loc}(\omega; A+\lambda)), \tag{4.17}$$

and the corresponding spectral density of spin fluctuations,

$$\rho^{sf}_{\rm loc}(\omega) = \frac{1}{2}(\rho_{\rm loc}(\omega; A-\lambda) - \rho_{\rm loc}(\omega; A+\lambda)). \tag{4.18}$$

In addition, the self-energy parts (4.11) continued analytically from a discrete set of points to the entire upper half-plane ω are also determined by functions (4.15):

$$\Sigma_-(\omega_n; A\pm\lambda) = (\xi\Omega)^2 \int_{-\infty}^{\infty} \rho_{\rm loc}(x, A\pm\lambda) dx \times \left[\frac{n(\mu - x) + N}{\omega - x - \Omega + i\gamma} + \frac{n(x-\mu) + N}{\omega - x + \Omega + i\gamma}\right]. \tag{4.19}$$

Here, $N = 1/(\exp(\Omega/T) - 1)$, and the positive constant γ was regarded as finite in the calculations so that $\gamma \ll \Omega$. The finite value of γ can describe the attenuation of phonons.

Substituting Eq. (4.17) into (3.13), we obtain the following integral equation for A:

$$A = \frac{1}{2}\int_{-\infty}^{\infty} d\omega \frac{\rho_{\rm loc}(\omega; A-\lambda) + \rho_{\rm loc}(\omega; A+\lambda)}{\exp\left(\dfrac{\omega-\mu}{T}\right) + 1}. \tag{4.20}$$

Equations (3.17) and (4.18) lead to the integral equation for λ:

$$\lambda = \frac{1}{2}\int_{-\infty}^{\infty} d\omega \frac{\rho_{\rm loc}(\omega; A-\lambda) - \rho_{\rm loc}(\omega; A+\lambda)}{\exp\left(\dfrac{\omega-\mu}{T}\right) + 1}. \tag{4.21}$$

Functions (4.16) determine the density of extended states,

$$\rho^{f}_{\mathrm{ext}}(\omega) = \frac{1}{2}\sum_{\mathbf{k}}(\rho_{\mathrm{ext}}(\mathbf{k};\,\omega;\,A-\lambda)+\rho_{\mathrm{ext}}(\mathbf{k};\,\omega;\,A+\lambda)), \tag{4.22}$$

and the corresponding spectral density of spin fluctuations,

$$\rho^{sf}_{\mathrm{ext}}(\omega) = \frac{1}{2}\sum_{\mathbf{k}}(\rho_{\mathrm{ext}}(\mathbf{k};\,\omega;\,A-\lambda)-\rho_{\mathrm{ext}}(\mathbf{k};\,\omega;\,A+\lambda)). \tag{4.23}$$

From the condition of conservation of the total number of fermions, we obtain the following equation for chemical potential μ in a doped compound:

$$\frac{1+x}{2}N_t = xN_tA + \int_{-\infty}^{\infty}\frac{\rho^{f}_{\mathrm{ext}}(\omega)d\omega}{\exp\left(\frac{\omega-\mu}{T}\right)+1}. \tag{4.24}$$

Since spectral density (4.16) possesses the property

$$\int d\omega\rho_{\mathrm{ext}}(\mathbf{k};\,\omega;\,A\pm\lambda) = 1,$$

we can obtain $N_t = 2\int\rho^{f}_{\mathrm{ext}}(\omega)d\omega$. Taking into account this property, we can reduce Eq. (4.24) for the chemical potential to the form

$$\frac{1}{2}+\frac{1}{xN_t}\int_{-\infty}^{\infty}\frac{\rho^{f}_{\mathrm{ext}}(\omega)d\omega}{\exp\left(\frac{\mu-\omega}{T}\right)+1} = \int_{-\infty}^{\infty}\frac{\rho^{f}_{\mathrm{loc}}(\omega)d\omega}{\exp\left(\frac{\omega-\mu}{T}\right)+1}. \tag{4.25}$$

The expression on the right-hand side of this equation gives A. For this reason, we always have $A > 1/2$.

The poles of the retarded Green function may lie in the region of the initial band or in the gap both for localized and for extended states. We will describe below the densities of states normalized to the concentration of impurity atoms. For example, extended states will be characterized by the following density per spin per impurity atom:

$$\frac{1}{xN_t}\rho^{f}_{\mathrm{ext}}(\omega). \tag{4.26}$$

In order to demonstrate clearly the effect of doping on the density of extended states in the valence band, we introduce the change in the density of states in this spectral region:

$$\Delta\rho^{f}_{\mathrm{ext}}(\omega) = \frac{1}{xN_t}(\rho^{f}_{\mathrm{ext}}(\omega)-\rho^{(0)}(\omega)). \tag{4.27}$$

In the expressions obtained for Green functions, we use the substitution

$$V_{\mathbf{k}l} = V_kN_t^{-1/2},$$

where V_k has the dimensions of energy. In our calculations, we assume that V_k is independent of $\mathbf{k}$.

The expression for the electron energy per impurity atom has the form

$$E_{\mathrm{el}}(T,x) = 2\int_{-\infty}^{\infty}\frac{\omega(\rho^{f}_{\mathrm{ext}}(\omega)+\rho^{f}_{\mathrm{loc}}(\omega))d\omega}{\exp\left(\frac{\omega-\mu}{T}\right)+1} - U(A^2-\lambda^2).$$

Thus, six constants D_b, ε_0, V_k, U, Ω, and $\xi\Omega$ with dimensions of energy determine model (2.1) completely. The solution to this equation for the normal paramagnetic state depends on the three self-consistent parameters A, λ, and μ. These parameters can be determined from three integral equations (4.20), (4.21), and (4.25). Two of these equations, (4.20) and (4.21), contain two spectral densities $\rho_{\mathrm{loc}}(\omega; A\pm\lambda)$, and Eq. (4.25) also contains $\rho_{\mathrm{ext}}(\mathbf{k};\,\omega;\,A\pm\lambda)$. Functions $\rho_{\mathrm{loc}}(\omega; A\pm\lambda)$ can be determined from the integral equation (4.15) taking into account Eqs. (4.9), (4.10), and the definition of self-energy parts (4.11) or (4.19). Functions $\rho_{\mathrm{ext}}(\mathbf{k};\,\omega;\,A\pm\lambda)$ can be determined from the solution to Eq. (4.16) taking into account definitions (4.13), (4.10), and (4.11). The latter quantities, (4.11), are determined in turn by the values of $\rho_{\mathrm{loc}}(\omega; A\pm\lambda)$.

We solved the obtained system of equations at finite temperatures using the iteration procedure. The error in the total number of states did not exceed 10^{-3} per impurity atom. Iterations were carried out until the absolute variation of self-consistent parameters for the last two iterations became smaller than 10^{-5}. The accuracy of calculations became lower upon an increase in the doping level x. For a given value of x, an optimal accuracy existed for a certain temperature. The computation error increased upon a deviation from this temperature.

4.3. Insulator State

The reason for the emergence of a insulator state in a doped compound at low doping levels and temperatures is associated with spin fluctuations represented by parameter λ in the obtained solution (4.8)–(4.13). Solid curves in Fig. 1 represent the electronic structure of the normal state of a doped compound with $x = 0.03$ at temperature $T = 40$ K. In our calculations, we used the initial band of the insulator in form (4.14) with a half-

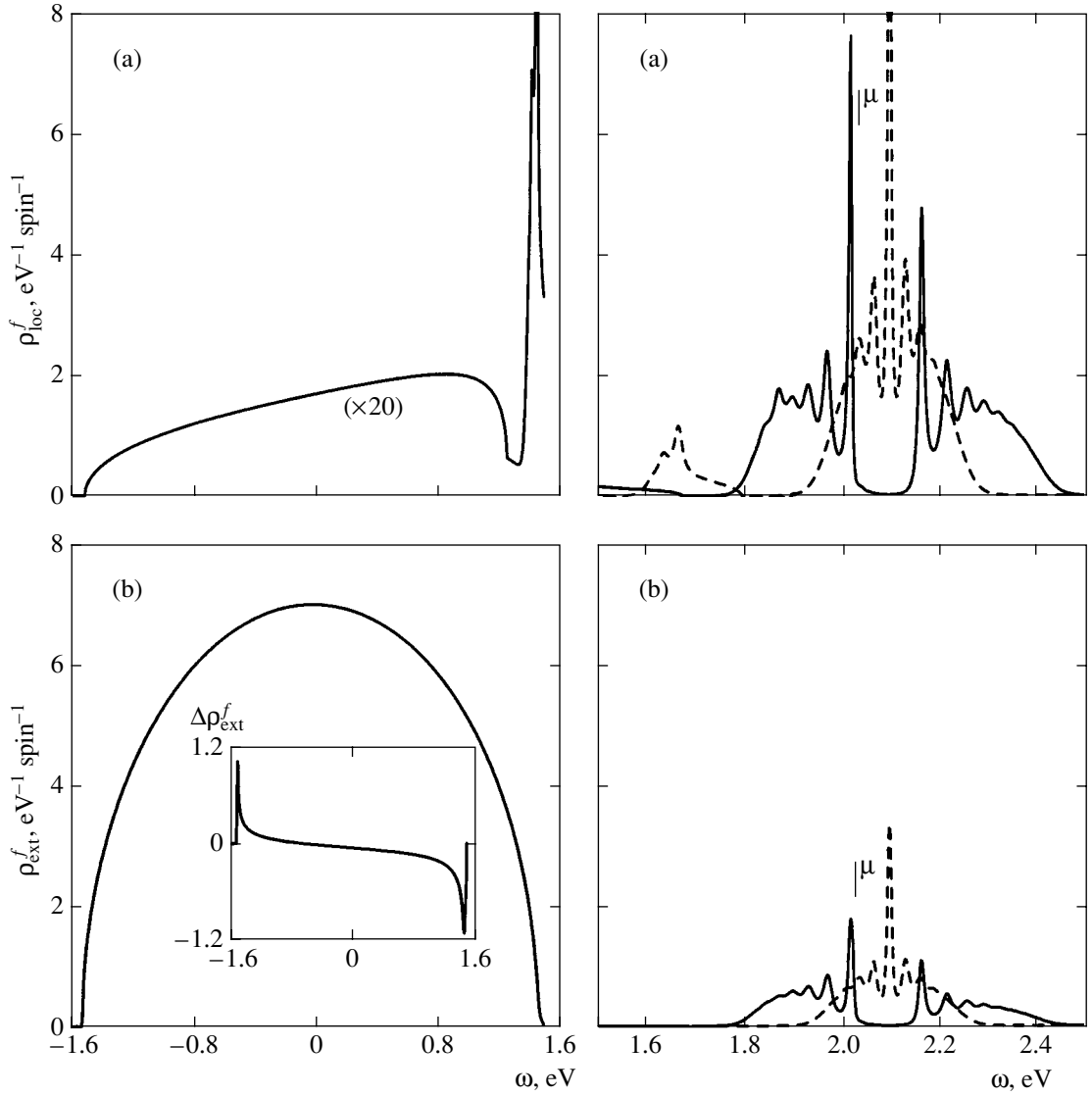

Fig. 1. Electronic structure of the spin-fluctuation insulator state for $x = 0.03$ and $T = 40$ K (solid curves): the spectral densities of (a) localized and (b) extended states. The dashed curve shows the electronic structure in the region of the initial insulator gap in the state of a poor metal, obtained for $\lambda = 0$. Parameters: $D_b = 1.5$ eV, $\varepsilon_0 = D_b - 0.2$ eV, $V_k = 1.2$ eV, $U = 0.6$ eV, $\xi = 5$, $\hbar\Omega = 20$ meV, and $\gamma = 0.2\hbar\Omega$.

width $D_b = 1.5$ eV, while the initial impurity level in this band was 0.2 eV lower than its upper boundary at 1.5 eV. It was found that $\lambda = 0.352$ and $A = 0.598$.

The $\rho^f_{loc}(\omega)$ dependence (solid curve in Fig. 1a) shows the density of localized states; accordingly, $\rho^f_{ext}(\omega)$ (solid curve in Fig. 1b) shows the single-particle density of extended states per spin per impurity atom both in the region of the initial valence band and in the region of the initial insulator gap. It can be seen that the structure has many peaks in the density of states $\rho^f_{loc}(\omega)$ and $\rho^f_{ext}(\omega)$ in the vicinity of chemical potential μ, which are due to multiphonon processes. The position of μ shown by the arrow in Fig. 1 is higher than the upper edge of the initial valence band of the insulator by approximately 0.5 eV.

It can be seen in Fig. 1 that a insulator gap in which the chemical potential is located appears in the impurity bands in localized and extended states. We will call this gap a pseudogap Δ_p. In accordance with numerical calculations, the gap width for localized and extended states may be different. For the parameters used in the calculations and for $x = 0.03$ at $T = 40$ K, the width of the pseudogap is approximately the same for both types of charge carriers. Since the density of localized states is considerably higher than the density of extended states in the vicinity of the chemical potential, we will determine the pseudogap from the position of the peaks

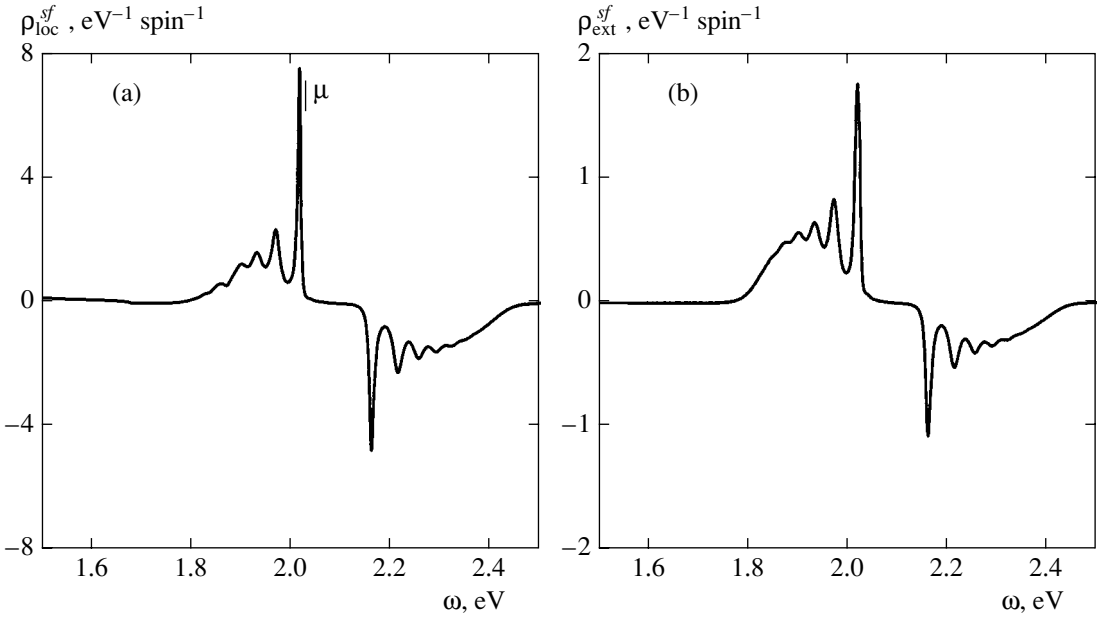

Fig. 2. Densities of spin fluctuations for (a) localized and (b) extended states in the vicinity of the chemical potential in the insulator state for $x = 0.03$ and $T = 40$ K (Fig. 1). The parameters are the same as in Fig. 1.

in the densities of localized states. It is equal to Δ_p = 146 meV (see Fig. 1).

The density of extended states in the initial band of the insulator varies significantly. The inset to Fig. 1b shows the variation $\Delta\rho^f_{ext}(\omega)$ of this density. In the vicinity of the upper edge of the band, $\Delta\rho^f_{ext} < 0$; consequently, the density of extended states is much lower than the initial density $\rho^{(0)}(\omega)$. In the vicinity of the lower edge of the band, $\Delta\rho^f_{ext} > 0$, and the density of extended states in this region is higher than the initial density. The change in the total number of extended states in the region of the initial band turns out to be negative. For the result represented in Fig. 1, this change is equal to –0.215 states. A band of localized states, which is shown in Fig. 1a with a 20 magnification, also appears in this region. The total number of these states to the initial boundary of the valence band D_b is 0.243.

The decrease in the total number of extended states in the region of the initial valence band is due to the fact that a part of these states splits from the initial band and forms two impurity bands of extended states in the region of the initial insulator gap (see Fig. 1b). The total number of these states in the entire gap is 0.215 states per spin per impurity atom. These two bands are determined by the spectral densities $\rho_{ext}(\mathbf{k}; \omega; A + \lambda)$ (4.16) taking into account formulas (4.13) and (4.10), (4.11). For $x = 0.03$ and $T = 40$ K in the vicinity of the chemical potential, these two bands are separated by the pseudogap. The gap in the spectrum of the impurity bands of localized states appears for the same reason. This spectral density is determined by the values of $\rho_{loc}(\omega; A \pm \lambda)$ (4.15) taking into account formulas (4.9) and (4.10). The total number of these states in the entire gap is equal to 0.757 states per spin per impurity atom. As a result, a pseudogap appears in the densities of single-particle states and the normal state is insulator.

In the region of the initial valence band, the corresponding pair of spectral densities always overlap considerably and will not be considered separately here.

In order to demonstrate clearly the role of spin fluctuations in the formation of the insulator state, we also show in Fig. 1 (dashed curves) the densities of states calculated for $\lambda = 0$. In this case, there is no pseudogap in the spectral densities of states, and the compound is in the state of a poor metal, for which the band for localized and extended states coexist on the Fermi surface, the total number of localized fermions being several times larger than the number of extended fermions. The chemical potential in this state lies at the highest peaks on the densities of localized and extended states; it can clearly be seen that the position of μ for the insulator state is lower.

In the state of a poor metal, the electron energy E_{el} = 0.953 eV per impurity atom, while the insulator state is characterized by E_{el} = 0,932 eV per impurity atom. Consequently, spin fluctuations responsible for the emergence of a insulator gap lower the electron energy. For this reason, from the two states, the insulator state induced by spin fluctuations is realized. Figure 2 shows the densities ρ^{sf}_{loc} and ρ^{sf}_{ext} for localized and extended states in the vicinity of the chemical potential (the

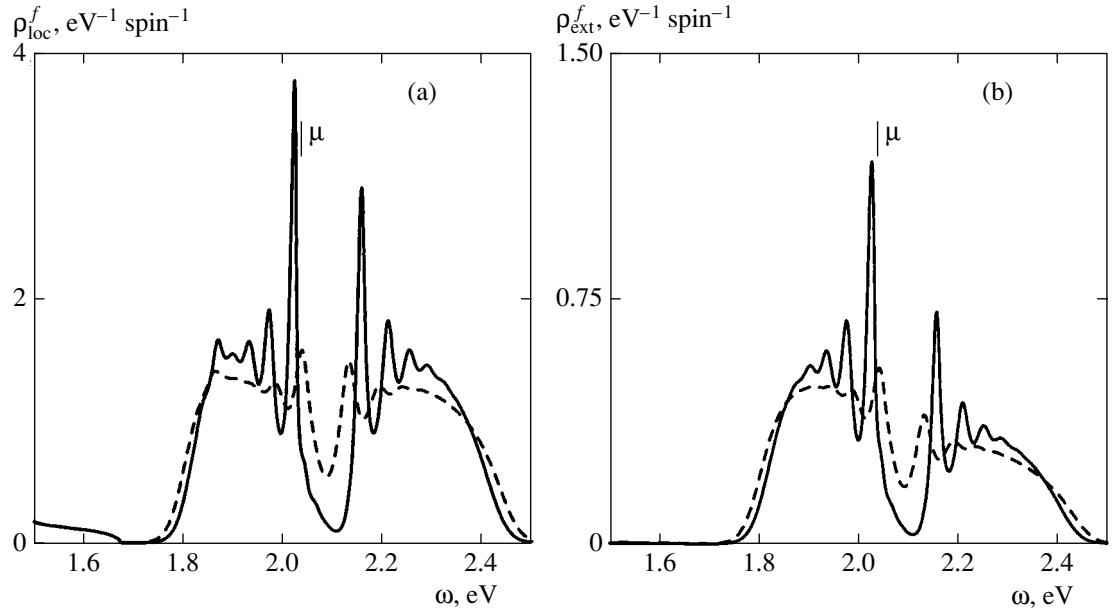

Fig. 3. Temperature variation of the electronic structure in the vicinity of the chemical potential near the insulator–metal transition. Solid curves show the densities of (a) localized and (b) extended states in the vicinity of the pseudogap at $T = 80$ K. Dashed curves represent the electronic structure of the metallic state at $T = 150$ K. The parameters and doping level x are the same as in Fig. 1.

region of the initial valence band also has a finite spectral density of these states, but it is not shown in the figure). The values of spectral densities of spin fluctuations averaged over the entire energy interval where their values are specified are equal to zero (see Eq. (3.16)). The set of values of λ in the integral equation (4.21), taking into account (4.18), is mainly determined by the positive part of ρ^{sf}_{loc} represented in Fig. 2, while the contribution to λ from the negative part of this spectral density is suppressed by the Fermi distribution function.

4.4. Temperature-Induced Insulator–Metal Transition

For a fixed doping level, an increase in temperature leads to a decrease in parameter λ and the pseudogap, and a smooth increase in the density of extended states in the pseudogap region. Such a behavior of the spectral density leads to a temperature-induced insulator–metal transition.

For $x = 0.03$ and for the parameters considered above, the gap in the density of extended states is preserved up to $T \approx 70$ K. The charge transfer phenomena are characterized by a certain activation energy. At high temperatures, this density is finite in the entire pseudogap region in which the chemical potential is located and we can expect a transition to metal-type conductivity. However, transport phenomena still differ considerably from those observed in metals since in this case the bands of localized and extended states coexist on the Fermi surface.

Solid curves in Fig. 3 represent the densities of localized (Fig. 3a) and extended (Fig. 3b) states in the vicinity of the pseudogap at $T = 80$ K and at the same doping level $x = 0.03$. Here, $\lambda = 0.350$ and $A = 0.597$. In the pseudogap region, the density of extended states becomes finite. However, in the vicinity of μ (see Fig. 3), this density is approximately one-third of the density of localized states. The value of Δ_p is 134 meV, which is 12 meV lower than the gap width at $T = 40$ K.

With increasing temperature, parameter λ decreases, and two bands of localized states become closer and overlap. A similar behavior is also observed for the two bands of extended states. The dashed curves in Fig. 3 show the electronic structure of the metallic state at $T = 150$ K. It was found that $\lambda = 0.345$ and $A = 0.596$. Chemical potential μ in this state is 27 meV higher than that indicated in Fig. 3 for $T = 80$ K. The densities of states become smoother, but we can still speak of pseudogap $\Delta_p = 94$ meV, which is manifested in singularities of photoelectron spectra.

The change in the electronic structure of the metallic state of a doped compound upon a further increase in temperature is depicted in Fig. 4. Here, the densities of localized states (curves *1* in Fig. 4a) and extended states (curves *2* in Fig. 4a) are shown by solid curves for $T = 180$ K and by dashed curves for $T = 250$ K. It was found that at $T = 180$ K, the spin-fluctuation parameter is $\lambda = 0.329$ and the corresponding densities of spin fluctuations are shown by solid curves *1* and *2* in Fig. 4b. It is important to note that although the pseudogap vanishes at temperatures above 180 K (see Fig. 4), spin fluctuations are preserved in the metallic phase up to much higher temperatures. This is clearly demonstrated by dashed curves *1* and *2* in Fig. 4b, which represent the

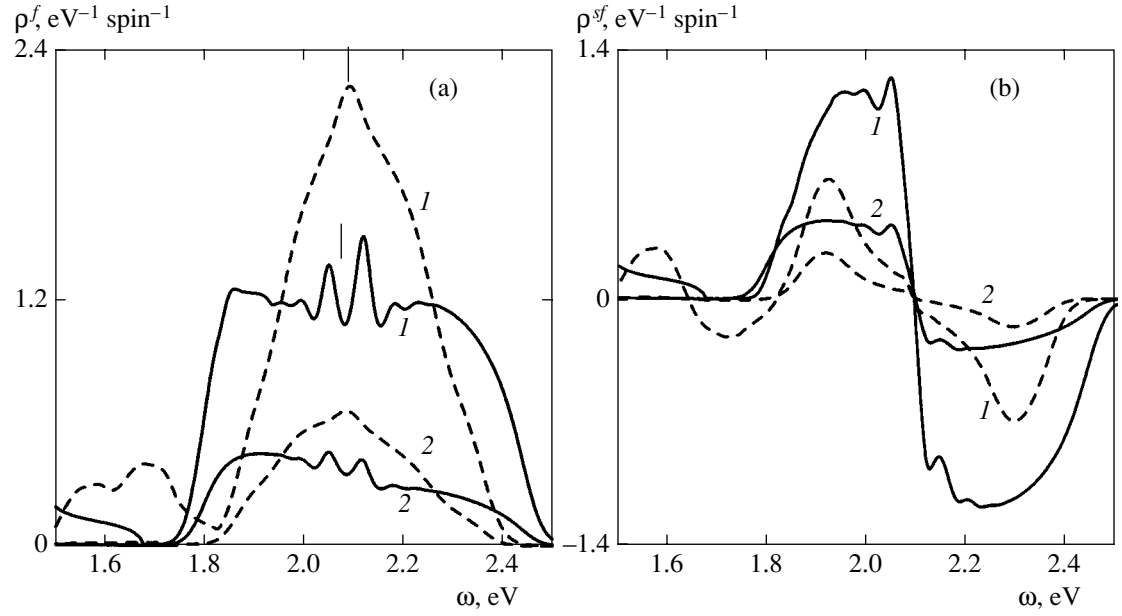

Fig. 4. (a) Electronic structure and (b) spectral density of spin fluctuations in the state of a poor metal at $T = 180$ K (solid curves) and 250 K (dashed curves). Localized states are represented by curves *1*, and extended states, by curves *2*. The parameters and doping level x are the same as in Fig. 1.

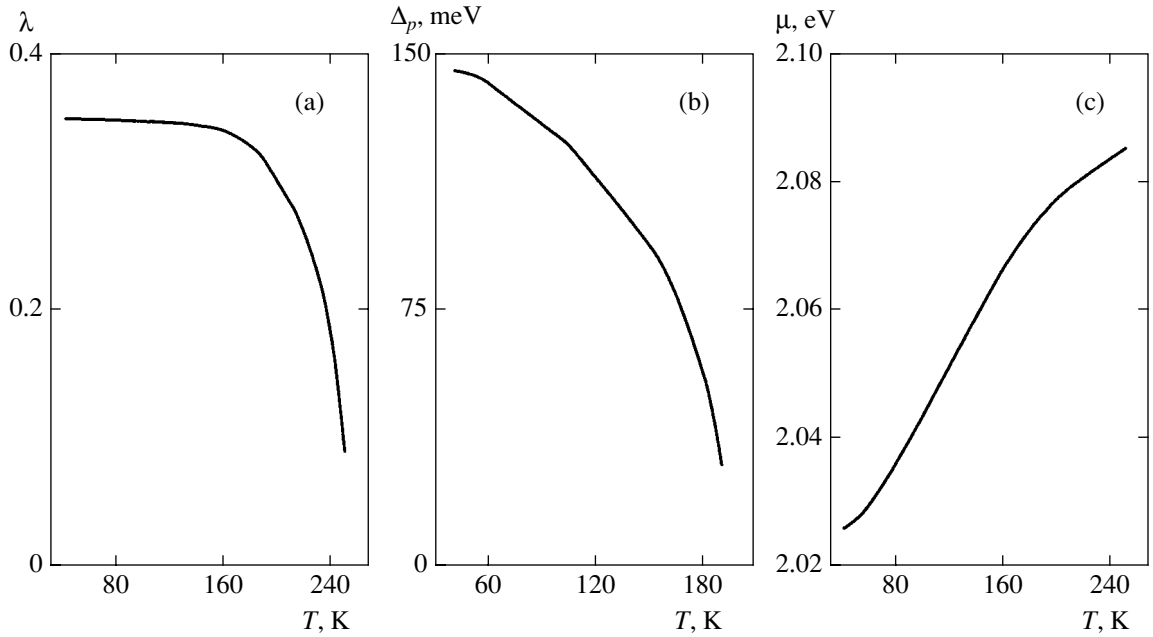

Fig. 5. Temperature dependences of (a) the spin-fluctuation parameter, (b) the pseudogap, and (c) the chemical potential. The parameters and doping level x are the same as in Fig. 1.

spectral densities of spin fluctuations at $T = 250$ K. For this state, $\lambda = 0.091$.

Thus, the temperature-induced insulator–metal transition has the following features. At low doping levels and temperatures, the doped compound is in the insulator state induced by spin fluctuations. This state is characterized by a pseudogap in the densities of localized and extended states. The chemical potential lies in the pseudogap. The charge transfer in this compound exhibit an activation dependence, the activation energy being a function of T.

As the temperature increases, spin fluctuations attenuate (Fig. 5a), the pseudogap decreases (Fig. 5b) and the density of both localized and extended states increases in the pseudogap region (see Figs. 1–4). This determines a smooth continuous transition to the state

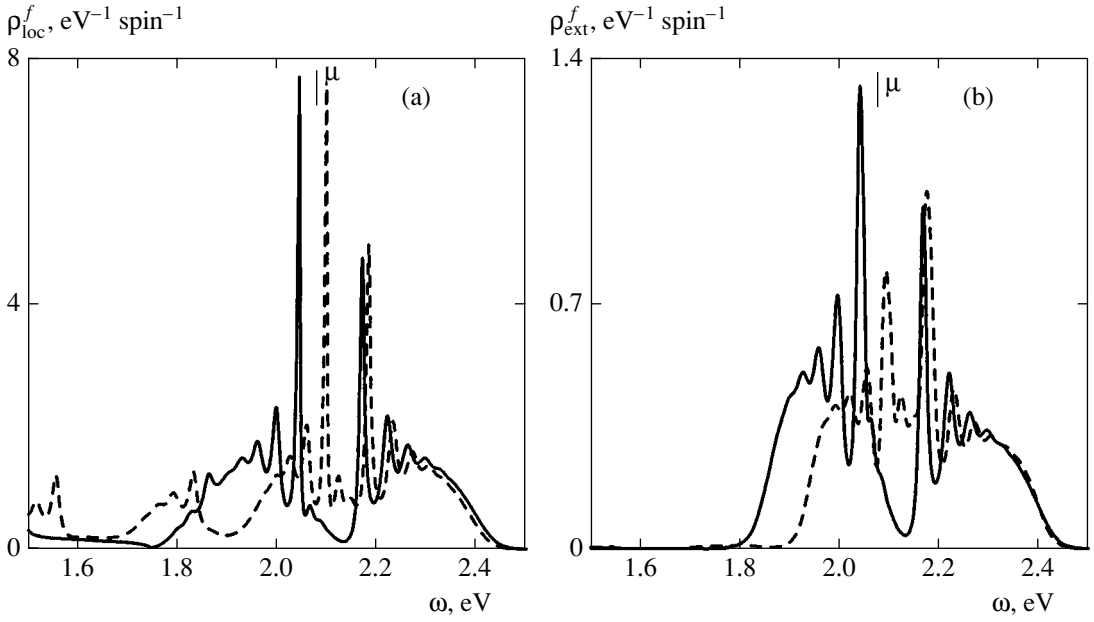

Fig. 6. Variation of electronic structure in the vicinity of the chemical potential for the doping-induced insulator–metal transition. Solid curves indicate the densities of (a) localized and (b) extended states near the pseudogap for $x = 0.05$. Dashed curves represent the electronic structure of the metallic state for $x = 0.1$. The parameters and the temperature are the same as in Fig. 1.

with a metal-type conductivity. First, even in the metallic phase, the dip in the vicinity of μ in the densities of localized and extended states disappears. The photoelectron spectroscopic data in this case can be interpreted as pseudogap collapse (see Fig. 5b). Then spin fluctuations in the doped compound attenuate (see Fig. 5a). The chemical potential increases with *T*, its shift being 60 meV upon a change in temperature from 40 to 250 K (Fig. 5c).

However, this metallic state is peculiar for at least two reasons. First, the bands of localized and extended states coexist on the Fermi surface, the density of localized states being much higher than the density of extended states. For this reason, such a state of a poor metal may exhibit anomalous transport properties as compared to those of ordinary metals. Second, single-particle Green functions in this state are nondiagonal in the spin subspace. The presence of spin fluctuations may lead to singularities in the behavior of the spin susceptibility of the doped compound.

4.5. Doping-Induced Insulator–Metal Transition

Here, we consider the results of calculations on the change in the electronic structure in the vicinity of the chemical potential depending on the doping level *x* at a fixed temperature $T = 40$ K. For $x = 0.03$, the doped compound is in the insulator state (see Figs. 1 and 2). As the doping level increases, parameter λ and the pseudogap decrease, while the density of extended states increases smoothly in the pseudogap region. This leads to a doping-induced insulator–metal transition.

Solid curves in Figs. 6a and 6b illustrate the electronic structure of the normal state of a doped compound for $x = 0.05$. Parameter $\lambda = 0.317$ decreases significantly, while the value of $A = 0.605$ is higher than the value corresponding to $x = 0.03$. In the pseudogap region, the density of extended states is finite, but it is smaller than the density of localized states at the chemical potential level. The pseudogap Δ_p determined from the position of the peaks in the densities of states is found to be equal to 129 meV.

As the value of *x* increases, two bands of localized states and two bands of extended states are shifted to the right and converge pairwise. The overlapping of the two pairs of bands increases in this case. As a result, the values of λ and Δ_p decrease and the density of states in the pseudogap increases. The dashed curves in Fig. 6 show the electronic structure of the metallic state for $x = 0.1$. It was found that $\lambda = 0.231$ and $A = 0.616$. Chemical potential μ in the same state is 41 meV higher than that corresponding to $x = 0.05$ in Fig. 6. Apparently, the dip in the density of states with its characteristic width of $\Delta_p = 90$ meV shown in Fig. 6 for $x = 0.1$ can still be manifested in the photoelectron spectra.

At high doping levels, the pseudogap in the densities of states disappears soon. Spin fluctuations are preserved in the metallic phase up to much higher values of *x*. For $x = 0.15$, a pseudogap $\Delta_p = 32$ meV is observed for localized states, but does not appear in the spectral function for extended states. However, spin fluctuations are present for both types of states and $\lambda = 0.148$.

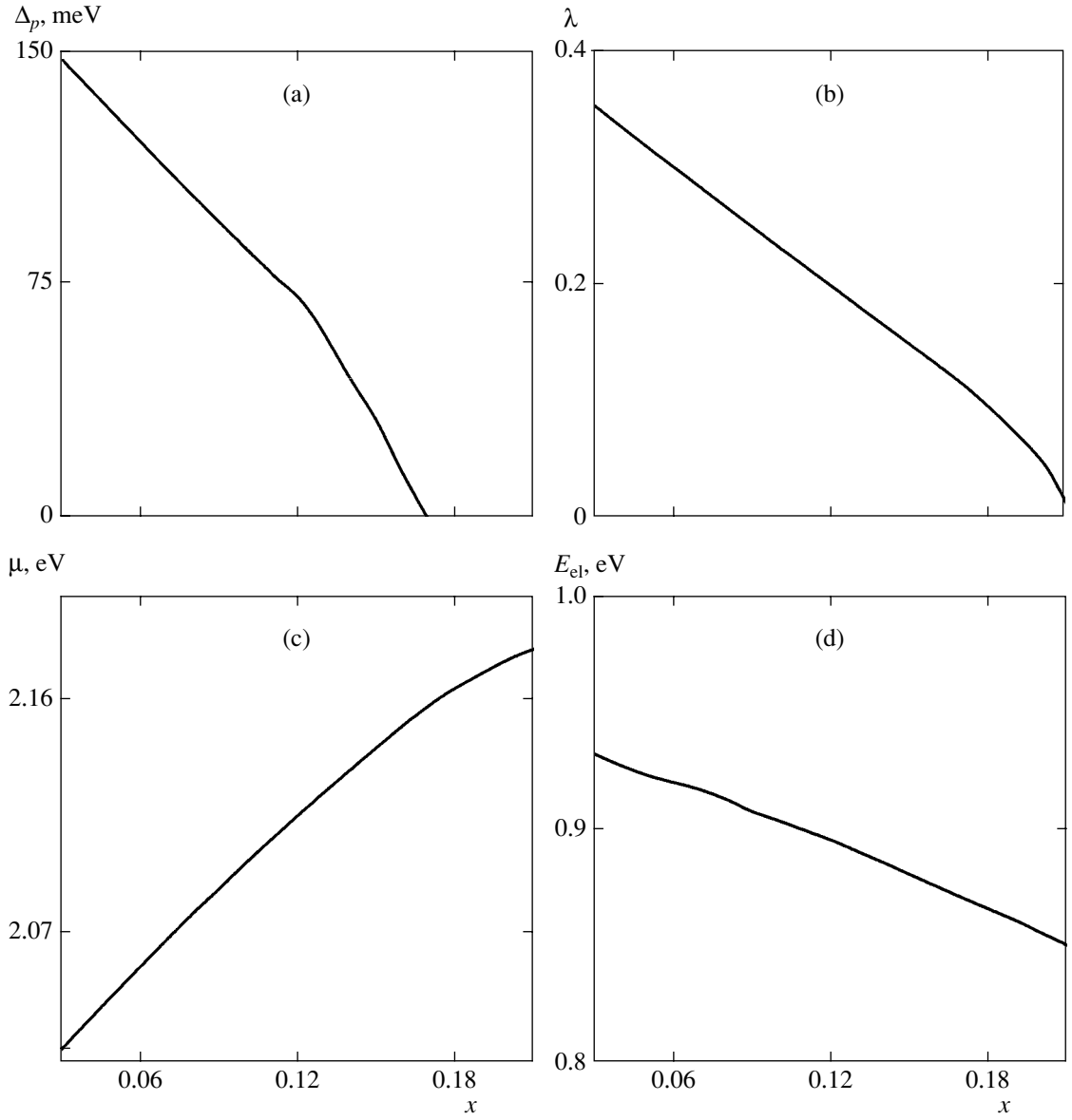

Fig. 7. Dependence of (a) the pseudogap width, (b) the spin-fluctuation parameter, (c) the chemical potential, and (d) the electron energy per impurity atom on the doping level at a constant temperature of $T = 40$ K. The parameters are the same as in Fig. 1.

Thus, an increase in the doping level at a fixed temperature leads to a smooth insulator–metal transition. In this case, the pseudogap in the densities of states in the vicinity of the chemical potential is the first to disappear (Fig. 7a). Then spin fluctuations attenuate for large x (Fig. 7b). Accordingly, $\lambda \longrightarrow 0$ and the doped compound passes to the state of a poor paramagnetic metal, in which the bands of localized and extended states coexist on the Fermi surface, the density of localized states considerably exceeding the density of extended states (see Fig. 6). The chemical potential increases with T, its shift upon a change in the doping level from 0.03 to 0.21 amounting to 155 meV (Fig. 7c). We can also plot the dependence of the electron energy E_{el} per impurity atom on x (Fig. 7d). Although the value of E_{el} decreases with increasing doping level, the product E_{el} increases with x.

4.6. Effect of the Allowed Bandwidth of an Insulator on the Insulator–Metal Transition

It is important to note that a peculiarity of the insulator–metal transition is that the allowed bands of the initial insulator must be quite narrow. Figure 8 shows the results of calculations of the electronic structures for two insulator states of doped compounds in which only the width $2D_b$ of the initial allowed band varies, the remaining model parameters being the same as in Fig. 1. As the bandwidth increases the structure of the impurity bands in the vicinity of the chemical potential

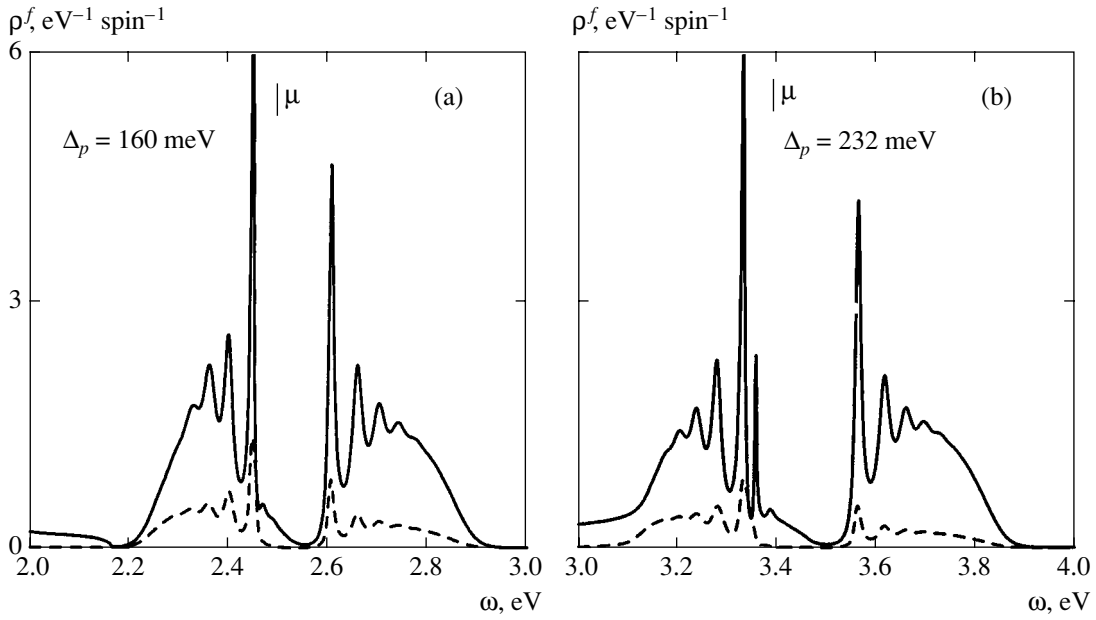

Fig. 8. Electronic structure of the spin-fluctuation insulator state depending on the half-bandwidth of the initial insulator: (a) $D_b = 2$ eV; (b) $D_b = 3$ eV. Solid curves correspond to localized states and dashed curves to the extended states. The parameters (except D_b) are the same as in Fig. 1.

shifts to the left, towards the initial band; it is important to note that in this case the pseudogap increases. The pseudogap is present in the spectra of both localized and extended states.

With increasing D_b, parameter A decreases, approaching its lower boundary of 1/2, while parameter λ increases, tending to 1/2, which is its upper boundary. For example, for the valence band halfwidth $D_b = 4$ eV of the valence band, we have $A = 0.542$, $\lambda = 0.464$, and $\Delta_p \approx 270$ meV, while the valence band halfwidth for $D_b = 1.5$ eV is equal to $\Delta_p = 146$ meV (see Fig. 1). It is worth noting that the spectral density of extended states is small as compared to the density of localized states (Fig. 8).

For $D_b = 1.5$ eV, the insulator–metal transition occurs in the vicinity of the doping level $x = 0.05$ (see Fig. 6). It was found that, as the allowed bandwidth increases, this transition occurs for larger values of x. For example, for $D_b = 4$ eV, $T = 40$ K, and $x = 0.12$, a clearly manifested pseudogap $\Delta_p^{\text{ext}} \approx 100$ meV appears in the density of extended states; however, the density of localized states within the pseudogap is finite. Accordingly, the insulator–metal transition will occur for $x > 0.12$.

Thus, for the insulator–metal transition studied here, the allowed bands of the initial insulator must be quite narrow. In insulators with *sp* bands, the predicted transition can hardly be observed since these bands are broad as a rule. Narrower bands can be expected in insulators with the *pd* hybridization. This is due to the fact that the wave functions of *d* electrons are coupled with the atomic potential more rigidly. For this reason, the initial insulator must contain an element from transition metals as a component. This element must take part in the formation of the band structure of the insulator in the vicinity of the insulator gap. It should be noted that, in accordance with photoelectron spectroscopic data, the typical width of the main peak of the valence band of cuprate is $2D_b \approx 3$ eV [34].

In model (2.1), the electron–phonon interaction affects the width of impurity bands. The smaller the parameter of this interaction, the narrower the bands. Accordingly (other conditions being equal), the pseudogap is wider and the spin fluctuations leading to the dielectrization of the compound are stronger. It can be stated that the electron–phonon interaction tends to suppress spin fluctuations in a doped compound and transform it to the state of a poor metal.

It can be predicted intuitively that a decrease in temperature may lead to a transition to the superconducting state from the normal state in which the pseudogap can be finite. However, the density of extended states in the pseudogap region must be quite high. It was shown above that the pseudogap or a deep valley in the density of states in the vicinity of the chemical potential appears when spin fluctuations in the doped compound are strong. These spin fluctuations determine parameter λ, which can vary within the limits $0 \le \lambda \le 1/2$. When the value of λ is closer to the right limit of this interval, it can be stated that spin fluctuations in the doped compound are strong.

Proceeding analogously to with the approach used for obtaining the Green functions for the normal state,

we can easily obtain solutions to the eight equations (3.28)–(3.35) for extended fermion and boson states and the eight equations (3.36)–(3.43) for localized states. On account of self-energy components (3.22)–(3.25) introduced above, this solution depends on four matched parameters A (3.13), λ (3.17), β (3.21), and chemical potential μ. This solution will not be given here since, first, it is very cumbersome and, second, we failed to obtain a numerical solution to this system with these mutually consistent parameters.

In model (2.1) with the electron–phonon interaction, the possibility of a superconducting transition in the vicinity of the insulator–metal transition was not investigated. Nevertheless, in the general case we can analyze the role of spin fluctuations in the superconducting state.

5. ANALYSIS OF THE TRIPLET PAIRING CHANNEL IN THE SUPERCONDUCTING STATE

Let us consider the case when $\lambda = 0$ and single-particle Green functions are diagonal in the spin index so that $g_{-\sigma,\sigma}(\nu, \nu_1) = 0$. In accordance with Eq. (3.23), $\Sigma^f_\pm = 0$. Then we obtain from Eq. (3.31)

$$f^{(+)}_{\uparrow\uparrow}(j, \mathbf{k}) = -\frac{\Delta^t}{\beta U - \Delta^s_+} f^{(+)}_{\downarrow\uparrow}(j, \mathbf{k}). \tag{5.1}$$

Taking this expression into account, we obtain from Eqs. (3.34) and (3.35)

$$f^{(+)}_{\uparrow\uparrow}(\mathbf{k}, \mathbf{k}) = -\frac{\Delta^t}{\beta U - \Delta^s_+} f^{(+)}_{\downarrow\uparrow}(\mathbf{k}, \mathbf{k}_1). \tag{5.2}$$

In the case considered here, Eqs. (3.32) and (3.33) can be reduced to the form

$$f^{(0)^{-1}}(l, l) f^{(+)}_{\downarrow\uparrow}(j, \mathbf{k}) + \sum_{\mathbf{k}_1} V_{\mathbf{k}_1 j} f^{(+)}_{\downarrow\uparrow}(\mathbf{k}_1, \mathbf{k}) = (\beta U - \Delta^s_-) g_{\uparrow\uparrow}(j, \mathbf{k}) \tag{5.3}$$

and

$$f^{(0)^{-1}}(l, l) f^{(+)}_{\uparrow\uparrow}(j, \mathbf{k}) + \sum_{\mathbf{k}_1} V_{\mathbf{k}_1} f^{(+)}_{\uparrow\uparrow}(\mathbf{k}_1, \mathbf{k}) = -\Delta^t g_{\uparrow\uparrow}(j, \mathbf{k}). \tag{5.4}$$

Using relations (5.1) and (5.2) for anomalous Green functions, we obtain from Eqs. (5.3) and (5.4)

$$g_{\uparrow\uparrow}(j, \mathbf{k})\left(1 - \frac{\beta U - \Delta^s_-}{\beta U - \Delta^s_+}\right)\Delta^t = 0. \tag{5.5}$$

Proceeding in a similar way, we can easily obtain from Eqs. (3.38), (3.40)–(3.43) the analogous expression

$$g_{\uparrow\uparrow}(j, j_1)\left(1 - \frac{\beta U - \Delta^s_-}{\beta U - \Delta^s_+}\right)\Delta^t = 0. \tag{5.6}$$

Since the first two cofactors on the left-hand sides of Eqs. (5.5) and (5.6) differ from zero, the only solution to these equations has the form

$$\Delta^t = 0. \tag{5.7}$$

Taking into account the definition of the self-energy part (3.25) of Δ^t, we arrive at the conclusion that the anomalous Green function $f^{(+)}_{\sigma\sigma}$ for triplet bosons is equal to zero.

Thus, in the general case, both singlet and triplet pairing channels are possible in the superconducting state. A spin-triplet channel is realized only if the parameter representing spin fluctuations in a doped insulator is finite, $\lambda \neq 0$ (3.17). It is only in this case that single-particle Green functions are nondiagonal in the spin index.

A purely electronic mechanism that takes into account both the singlet and the triplet pairing channels in the superconducting state of a doped nondegenerate insulator was studied in [85] for model (2.1), but without taking into account the electron–phonon interaction. In the remaining part of this review, we will analyze the phonon mechanism of superconductivity in model (2.1) for a doped insulator. The analysis will be carried out for the paramagnetic phase under the assumptions that spin fluctuations can be neglected (see Section 4).

6. PHONON-MEDIATED MECHANISM OF SUPERCONDUCTIVITY

In the case considered here, normal single-particle Green functions are diagonal in the spin index and only a spin-singlet pairing channel is realized. Henceforth, spin indices will be omitted. The system of equations for extended states has the form

$$g(\mathbf{k}, \mathbf{k}_1; \omega_n) = g^{(0)}(\mathbf{k}, \mathbf{k})\left(\delta_{\mathbf{k}\mathbf{k}_1} + \sum_j V_{\mathbf{k}j} g(j, \mathbf{k}_1; \omega_n)\right), \tag{6.1}$$

$$g(j, \mathbf{k}; \omega_n) = g^{(0)}(l, l) \times \left(\sum_{\mathbf{k}_1} V_{j\mathbf{k}_1} g(\mathbf{k}_1, \mathbf{k}; \omega_n) + (\beta U - \Delta^s_-) f^{(+)}(j, \mathbf{k}; \omega_n)\right), \tag{6.2}$$

$$f^{(+)}(j, \mathbf{k}; \omega_n) = f^{(0)}(l, l) \times \left(-\sum_{\mathbf{k}_1} V_{\mathbf{k}_1 j} f^{(+)}(\mathbf{k}_1, \mathbf{k}; \omega_n) + (\beta U - \Delta_-^s) g(j, \mathbf{k}; \omega_n) \right), \tag{6.3}$$

$$f^{(+)}(\mathbf{k}, \mathbf{k}_1; \omega_n) = -f^{(0)}(\mathbf{k}, \mathbf{k}) \sum_j V_{j\mathbf{k}} f^{(+)}(j, \mathbf{k}_1; \omega_n). \tag{6.4}$$

Similarly, for localized states we have

$$g(j, j_1; \omega_n) = g^{(0)}(l, l) \left(\delta_{jj_1} + \sum_{\mathbf{k}} V_{j\mathbf{k}} g(\mathbf{k}, j_1; \omega_n) + (\beta U - \Delta_-^s) f^{(+)}(j, j_1; \omega_n) \right), \tag{6.5}$$

$$g(\mathbf{k}, j; \omega_n) = g^{(0)}(\mathbf{k}, \mathbf{k}) \sum_{j_1} V_{\mathbf{k}j_1} g(j_1, j; \omega_n), \tag{6.6}$$

$$f^{(+)}(j, j_1; \omega_n) = f^{(0)}(l, l) \times \left(-\sum_{\mathbf{k}} V_{\mathbf{k}j} f^{(+)}(\mathbf{k}, j_1; \omega_n) + (\beta U - \Delta_-^s) g(j, j_1; \omega_n) \right), \tag{6.7}$$

$$f^{(+)}(\mathbf{k}, j; \omega_n) = -f^{(0)}(\mathbf{k}, \mathbf{k}) \sum_{j_1} V_{j_1\mathbf{k}} f^{(+)}(j_1, j; \omega_n). \tag{6.8}$$

For solving Eqs. (6.1)–(6.4) and (6.5)–(6.8), we will use the following approach. Substituting the nondiagonal Green function $g(j_1, j \neq j_1)$ defined by expression (6.5) into the right-hand side of Eq. (6.6) and using Eqs. (6.7) and (6.8), we obtain

$$a_g g(\mathbf{k}, j; \omega_n) + b_f f^{(+)}(-\mathbf{k}, j; \omega_n) = V_{\mathbf{k}, j}[g(j, j) - g^{(0)}(l, l)(\beta U - \Delta_-^s) f^{(+)}(j, j)] + g^{(0)}(l, l) \sum_{j_1 \neq j, \mathbf{k}_1 \neq \mathbf{k}} V_{\mathbf{k}j_1} V_{j_1\mathbf{k}_1} g(\mathbf{k}_1, j). \tag{6.9}$$

Here, we have

$$a_g(\mathbf{k}; \omega_n) = g^{(0)^{-1}}(\mathbf{k}, \mathbf{k}; \omega_n) - xN_t V_{\mathbf{k}l}^2 g^{(0)}(l, l; \omega_n), \tag{6.10}$$

$$b_f = (\beta U - \Delta_-^s) g^{(0)}(l, l) f^{(0)^{-1}}(-\mathbf{k}, -\mathbf{k}; \omega_n). \tag{6.11}$$

The second equation connecting the nondiagonal functions $g(\mathbf{k}, j; \omega_n)$ and $f^{(+)}(-\mathbf{k}, j; \omega_n)$ with diagonal functions $g(j, j)$ and $f^{(+)}(j, j)$ can easily be obtained from Eq. (6.8) by substituting $f^{(+)}(j_1, j \neq j_1)$ (6.7) into it and using Eqs. (6.5) and (6.6):

$$b_g g(\mathbf{k}, j; \omega_n) + a_f f^{(+)}(-\mathbf{k}, j; \omega_n) = V_{\mathbf{k}, j}[(\beta U - \Delta_-^s) f^{(0)}(l, l) g(l, l) - f^{(+)}(j, j)] + f^{(0)}(l, l) \sum_{j_1 \neq j, \mathbf{k}_1 \neq \mathbf{k}} V_{\mathbf{k}j_1} V_{j_1\mathbf{k}_1} f^{(+)}(-\mathbf{k}, j). \tag{6.12}$$

Here, we have

$$a_f(-\mathbf{k}; \omega_n) = f^{(0)^{-1}}(-\mathbf{k}, -\mathbf{k}; \omega_n) - xN_t V_{\mathbf{k}l}^2 f^{(0)}(l, l; \omega_n), \tag{6.13}$$

$$b_g = (\beta U - \Delta_-^s) f^{(0)}(l, l) g^{(0)^{-1}}(\mathbf{k}, \mathbf{k}; \omega_n). \tag{6.14}$$

Using Eqs. (6.1)–(6.4), we can obtain two more similar equations connecting the nondiagonal Green functions $g(j, \mathbf{k}; \omega_n)$ and $f^{(+)}(j, \mathbf{k}; \omega_n)$ with the required functions $g(\mathbf{k}, \mathbf{k})$ and $f^{(+)}(-\mathbf{k}, \mathbf{k})$. These equations contain similar terms with double sums, which are analogous to the last terms on the right-hand sides of Eqs. (6.9) and (6.12).

For low doping levels ($x^2 \ll 1$), such terms with double sums can be disregarded (see Section 4.1).

Solving Eqs. (6.9) and (6.12), we obtain functions $g(\mathbf{k}, j; \omega_n)$ and $f^{(+)}(-\mathbf{k}, j; \omega_n)$ expressed in terms of diagonal functions $g(j, j)$ and $f^{(+)}(j, j)$. Substituting these expressions into Eqs. (6.5) and (6.7), we obtain a system of two equations for $g(j, j)$ and $f^{(+)}(j, j)$. Omitting simple but cumbersome calculations, we can present the solution to this system for localized fermions and bosons in the form

$$g(j, j; \omega_n) = \frac{f^{(0)^{-1}}(l, l) - S_2(\omega_n)}{(g^{(0)^{-1}}(l, l) - S_1)(f^{(0)^{-1}}(l, l) - S_2) - (\beta U - \Delta_-^s)^2 S_3^2}, \tag{6.15}$$

$$f^{(+)}(j, j; \omega_n) = \frac{(\beta U - \Delta_-^s(\omega_n)) S_3(\omega_n)}{(g^{(0)^{-1}}(l, l) - S_1)(f^{(0)^{-1}}(l, l) - S_2) - (\beta U - \Delta_-^s)^2 S_3^2}. \tag{6.16}$$

Here, we have used the following notation:

$$S_1(\omega_n; x) = \sum_{\mathbf{k}} \frac{V_{\mathbf{k}l}^2}{a_g a_f - b_g b_f}[a_f - b_f f^{(0)}(l, l)(\beta U - \Delta_-^s)], \tag{6.17}$$

$$S_2(\omega_n; x) = \sum_{\mathbf{k}} \frac{V_{\mathbf{k}l}^2}{a_g a_f - b_g b_f}[a_g - b_g g^{(0)}(l, l)(\beta U - \Delta_-^s)], \tag{6.18}$$

$$S_3(\omega_n; x) = 1 + xN_t g^{(0)}(l, l; \omega_n) f^{(0)}(l, l; \omega_n) \sum_{\mathbf{k}} \frac{V_{\mathbf{k}l}^4}{a_g a_f - b_g b_f}. \tag{6.19}$$

Proceeding in a similar way, we obtain from Eqs. (6.1)–(6.4) the following expressions for extended fermions and bosons:

$$g(\mathbf{k}, \mathbf{k}; \omega_n) = \frac{f^{(0)^{-1}}(-\mathbf{k}, -\mathbf{k}; \omega_n) - D_2(\mathbf{k}; x)}{(g^{(0)^{-1}}(\mathbf{k}, \mathbf{k}) - D_1)(f^{(0)^{-1}}(-\mathbf{k}, -\mathbf{k}) - D_2) - (\beta U - \Delta_-^s)^2 D_3^2}, \tag{6.20}$$

$$f^{(+)}(-\mathbf{k}, \mathbf{k}; \omega_n) = -\frac{(\beta U - \Delta_-^s) D_3(\mathbf{k}; x)}{(g^{(0)^{-1}}(\mathbf{k}, \mathbf{k}) - D_1)(f^{(0)^{-1}}(-\mathbf{k}, -\mathbf{k}) - D_2) - (\beta U - \Delta_-^s)^2 D_3^2}. \tag{6.21}$$

Here, we have used the notation

$$D_1(\mathbf{k}, \omega_n; x) = \frac{xN_t V_{\mathbf{k}l}^2 c_f}{c_g c_f - (\beta U - \Delta_-^s)^2}, \tag{6.22}$$

$$D_2(\mathbf{k}, \omega_n; x) = -\frac{xN_t V_{\mathbf{k}l}^2 c_g}{c_g c_f - (\beta U - \Delta_-^s)^2}, \tag{6.23}$$

$$D_3(\mathbf{k}, \omega_n; x) = \frac{xN_t(\beta U - \Delta^s) V_{\mathbf{k}l}^2}{c_g c_f - (\beta U - \Delta_-^s)^2}, \tag{6.24}$$

$$c_g(\omega_n) = g^{(0)^{-1}}(l, l) - \sum_{\mathbf{k}} V_{\mathbf{k}l}^2 g^{(0)}(\mathbf{k}, \mathbf{k}), \tag{6.25}$$

$$c_f(\omega_n) = f^{(0)^{-1}}(l, l) - \sum_{\mathbf{k}} V_{\mathbf{k}l}^2 f^{(0)}(\mathbf{k}, \mathbf{k}). \tag{6.26}$$

6.1. Numerical Analysis of the Superconductor–Metal Transition

Here, we also use expression (4.14) for the initial density of states in the valence band of a insulator.

The analytic continuation of the Matsubara Green functions to the upper part of the complex plane ω makes it possible to obtain the Fourier components of the retarded temporal Green function at finite temperatures. This allows us to determine the spectral densities

$$\rho_{\text{loc}}^f(\omega) = -\frac{1}{\pi}\text{Im}g(j, j; \omega_n)\Big|_{i\omega_n \to \omega + i\delta}, \tag{6.27}$$

$$\rho_{\text{loc}}^{b0}(\omega) = -\frac{1}{\pi}\text{Im}\, f^{(+)}(j, j; \omega_n)\Big|_{i\omega_n \to \omega + i\delta}, \tag{6.28}$$

$$\rho_{\text{ext}}^f(\mathbf{k}, \mathbf{k}; \omega) = -\frac{1}{\pi}\text{Im}g(\mathbf{k}, \mathbf{k}; \omega_n)\Big|_{i\omega_n \to \omega + i\delta}, \tag{6.29}$$

$$\rho_{\text{ext}}^{b0}(-\mathbf{k}, \mathbf{k}; \omega) = -\frac{1}{\pi}\text{Im}\, f^{(+)}(-\mathbf{k}, \mathbf{k}; \omega_n)\Big|_{i\omega_n \to \omega + i\delta} \tag{6.30}$$

(δ is a positive infinitesimal).

Expression (6.29) enables us to determine the density of single-particle extended states:

$$\rho_{\text{ext}}^f(\omega) = \sum_{\mathbf{k}} \rho_{\text{ext}}^f(\mathbf{k}, \mathbf{k}; \omega). \tag{6.31}$$

The equation for μ has the same form as Eq. (4.25), but $\rho_{\text{ext}}^f(\omega)$ is now defined by Eq. (6.31), while $\rho_{\text{loc}}^f(\omega)$ is defined by formula (6.27).

We are also interested in the energy distribution of superconducting bosons. The momentum distribution of extended charged bosons has the form

$$\rho_{\text{ext}}^{b0}(\mathbf{k}) = \left|\lim_{\tau \to +0} f^{(+)}(-\mathbf{k}, \mathbf{k}; \tau)\right|^2. \tag{6.32}$$

Taking into account relation (6.30), we can reduce expression (6.32) to the form

$$\rho_{\text{ext}}^{b0}(\mathbf{k}) = \left| \int_{-\infty}^{\infty} \frac{\rho_{\text{ext}}^{b0}(-\mathbf{k}, \mathbf{k}; \omega) d\omega}{\exp\left(\frac{\omega - \mu}{T}\right) + 1} \right|^2. \quad (6.33)$$

This expression leads to the total number of bosons per unit volume:

$$N_{\text{ext}}^{b0} = 2\sum_{\mathbf{k}} \rho_{\text{ext}}^{b0}(\mathbf{k}) = \int_{-D_b}^{D_b} \rho_{\text{ext}}^{b0}(\varepsilon_{\mathbf{k}}) d\varepsilon_{\mathbf{k}}, \quad (6.34)$$

where the boson distribution over energy eigenvalues $\varepsilon_{\mathbf{k}}$ of the initial insulator has the form

$$\rho_{\text{ext}}^{b0}(\varepsilon_{\mathbf{k}}) = 2\rho^{(0)}(\varepsilon_{\mathbf{k}})\rho_{\text{ext}}^{b0}(\mathbf{k}(\varepsilon_{\mathbf{k}})).$$

If we use the above property of spectral density (6.31),

$$\int \rho_{\text{ext}}^{f}(\mathbf{k}, \mathbf{k}; \omega) d\omega = 1,$$

the energy distribution for bosons can be written in the form

$$\rho_{\text{ext}}^{b0}(\omega) = 2\int_{-D_b}^{D_b} \rho^{(0)}(\varepsilon_{\mathbf{k}}) d\varepsilon_{\mathbf{k}} \rho_{\text{ext}}^{f}(\mathbf{k}, \mathbf{k}; \omega) \times \left| \int_{-\infty}^{\infty} \frac{\rho_{\text{ext}}^{b0}(-\mathbf{k}, \mathbf{k}; \omega) d\omega}{\exp\left(\frac{\omega - \mu}{T}\right) + 1} \right|^2. \quad (6.35)$$

This energy distribution for extended bosons can be determined both in the valence band region and in the insulator gap of the initial insulator. Consequently, the total concentration of bosons can be conveniently written in the form

$$N_{\text{ext}}^{b0} = N_{\text{ext},\,b}^{b0} + N_{\text{ext},\,g}^{b0}.$$

Parameter β determined by the spectral density of the boson state at an impurity site makes it possible to determine the concentration of local charged bosons, which has the form

$$N_{\text{loc}}^{b0} = 2x\beta^2 N_t.$$

Analytic continuation of the self-energy parts (3.22) and (3.24) from a discrete set of points to the entire upper half-plane ω has the form

$$\Sigma_{-}^{n}(\omega) = (\xi\Omega)^2 \int_{-\infty}^{\infty} \rho_{\text{loc}}^{f}(x) dx \times \left[\frac{n(\mu - x) + N}{\omega - x - \Omega + i\gamma} + \frac{n(x - \mu) + N}{\omega - x + \Omega + i\gamma} \right], \quad (6.36)$$

$$\Sigma_{+}^{n}(\omega) = -(\xi\Omega)^2 \int_{-\infty}^{\infty} \rho_{\text{loc}}^{f}(x) dx \times \left[\frac{n(\mu - x) + N}{\omega + x + \Omega - 2\mu + i\gamma} + \frac{n(x - \mu) + N}{\omega + x - \Omega - 2\mu + i\gamma} \right], \quad (6.37)$$

$$\Delta_{-}^{s}(\omega) = (\xi\Omega)^2 \int_{-\infty}^{\infty} \rho_{\text{loc}}^{b0}(x) dx \times \left[\frac{n(\mu - x) + N}{\omega - x - \Omega + i\gamma} + \frac{n(x - \mu) + N}{\omega - x + \Omega + i\gamma} \right]. \quad (6.38)$$

In the results described below, the density of both localized and extended states is normalized to the concentration (4.26) of substitutional atoms. In the region of the initial valence band, we consider for the sake of visualization the change in the density of extended states per substitutional atom (4.27). In our calculations, the value of $V_k = V_{\mathbf{k}l} N_t^{1/2}$ was regarded as independent of $\mathbf{k}$.

The expression for the electron energy per impurity atom has the form

$$E_{\text{el}}(T, x) = 2\int_{-\infty}^{\infty} \frac{\omega(\rho_{\text{ext}}^{f}(\omega) + \rho_{\text{loc}}^{f}(\omega)) d\omega}{\exp\left(\frac{\omega - \mu}{T}\right) + 1} - U(A^2 + \beta^2). \quad (6.39)$$

The temperature dependence of the electron energy $E_{\text{el}}(T, x)$ allowed us to determine the electronic heat capacity per impurity atom:

$$C_{\text{el}}(T, x) = \frac{\partial E_{\text{el}}(T, x)}{\partial T}. \quad (6.40)$$

It should be noted that in the atomic limit, when $V_{\mathbf{k}l} \longrightarrow 0$, it follows from Eqs. (6.10) and (6.21) that $f^{(+)}(-\mathbf{k}, \mathbf{k}; \omega_n) \longrightarrow 0$ and $g(\mathbf{k}, \mathbf{k}; \omega_n) \longrightarrow g^{(0)}(\mathbf{k}, \mathbf{k}; \omega_n)$. In addition, it follows from Eqs. (6.15) and (6.16) that $g(j, j; \omega_n)$ and $f^{(+)}(j, j; \omega_n)$ are transformed, respectively, into expressions (39) and (41) from [83], which describe one-site boson–polaron mixed states.

The solution to Hamiltonian (2.1) for the superconducting state depends on three self-consistent parameters A, μ, and β and, accordingly, there are only two self-consistent parameters for the normal state since $\beta = 0$. We solved the obtained system of equations for

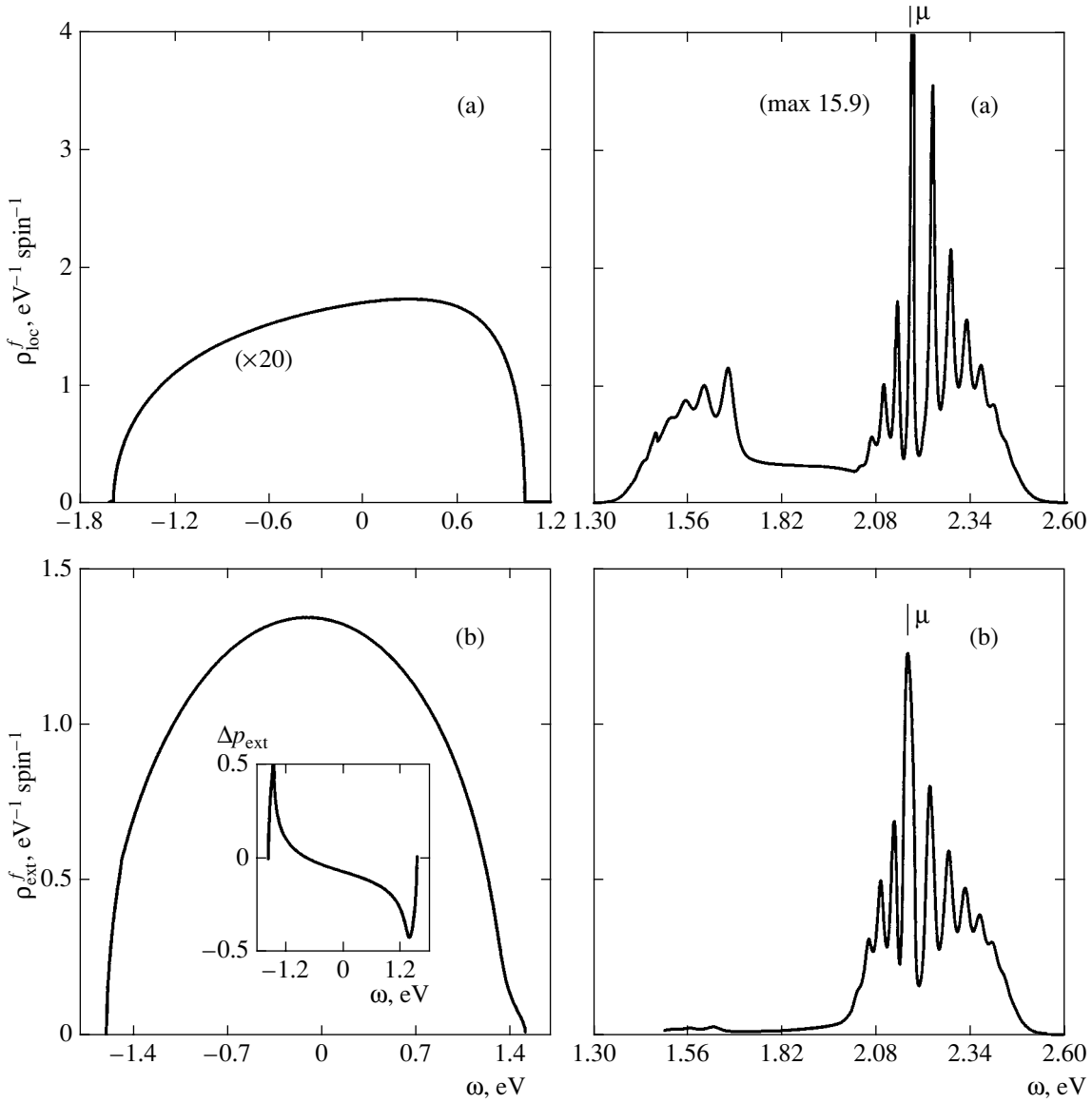

Fig. 9. The state of a poor paramagnetic metal at $x = 0.15$ and $T = 45$ K: (a) localized states and (b) extended states. Parameters: $D_b = 1.5$ eV, $\varepsilon_0 = D_b - 0.2$ eV, $V_k = 1.3$ eV, $U = 0.5$ eV, $\xi = 6$, $\hbar\Omega = 30$ meV, and $\gamma = 0.1\hbar\Omega$.

finite temperatures by using an iterative procedure. For doping levels $x \approx 0.15$, the error in the total number of states did not exceed 5×10^{-4} per impurity atom. Iterations were carried out until the absolute variation of self-consistent parameters for the last two iterations became smaller than 10^{-5}. For a given value of x, there existed an optimal accuracy at a certain temperature. A deviation from this temperature deteriorated the accuracy of computations.

6.2. The State of a Poor Paramagnetic Metal

Figure 9 shows the electron structure of the normal state of a doped compound at $x = 0.15$, calculated under the assumption that $\beta = 0$ at $T = 45$ K. The $\rho^f_{\text{loc}}(\omega)$ dependence (Fig. 9a) shows the density of localized states, while $\rho^f_{\text{ext}}(\omega)$ (Fig. 9b) shows the single-particle density of extended states both in the region of the initial valence band and in the region of the initial insulator gap. It is worth noting that single-particle densities of states $\rho^f_{\text{loc}}(\omega)$ and $\rho^f_{\text{ext}}(\omega)$ have a multipeak structure due to many-phonon processes in the vicinity of chemical potential μ lying deeply in the initial insulator gap.

The density of extended states in the initial insulator band varies significantly. The inset to Fig. 9b shows the density variation $\Delta\rho^f_{\text{ext}}(\omega)$. In the vicinity of the upper edge of the band, $\Delta\rho^f_{\text{ext}} < 0$ and, accordingly, the density of extended states is much lower than the initial

density $\rho^{(0)}(\omega)$. In the vicinity of the lower edge of the band, $\Delta\rho^{f}_{\text{ext}} > 0$, and the density of extended states is higher than the initial density in this region. The change in the total number of extended states in the region of the initial band is negative. For the result represented in Fig. 9, this variation is equal to –0.1874 states per spin per impurity atom. The band of localized states shown in Fig. 9a also appears in this region.

The decrease in the total number of extended states in the region of the initial band is associated with the fact that a part of these states split from the initial band and form impurity bands of extended states in the region of the insulator gap (see Fig. 9b). The total number of these states in the entire gap is 0.1874.

The parameters calculated for the normal state represented here are as follows: $A = 0.6202$, $\mu = 2.1620$ eV, and the electron energy of the state is $E_{\text{el}} = 0.7472$ eV per impurity atom. It can easily be seen from Fig. 9 that this electronic structure corresponds to the state of a poor metal. At the energy of the chemical potential, bands of both localized and extended states coexist. For the energy equal to μ, the density of localized states is almost an order of magnitude higher than the density of extended states. It is important to note that this band of extended states is quite narrow, its effective width being on the order of 0.5 eV. It is characterized by a relatively low total number of states, equal to 0.1874. At the same time, this band, which is filled approximately by half, has a high density of states. It can be seen from Fig. 9b that the maximal density of extended states in this band is approximately equal to the maximal density of states in the initial valence band of the insulator.

6.3. Superconducting State

It was noted above that the normal state is characterized by an electron energy $E_{\text{el}} = 0.7472$ eV per impurity atom. For the same parameters of calculations, the superconducting state possesses a lower energy; it was found that $E_{\text{el}} = 0.7461$ eV per impurity atom in this state. We have determined the probability amplitude $\beta = 2.635 \times 10^{-2}$ for the boson state filling at a site, the decrease in the chemical potential $\Delta\mu = -0.36$ meV in the superconducting state, and the increase in the single-particle site filling $\Delta A = 0.52 \times 10^{-3}$ as compared to the normal state.

Figure 10 shows the densities of single-particle localized (a) and extended (b) states in the region of the initial insulator gap ($\omega > 1.5$ eV) in the superconducting state of a doped compound at $x = 0.15$ and $T = 45$ K (it will be shown below that the superconducting transition temperature $T_c > 45$ K). Here, the changes in the densities of single-particle states in the region of the initial valence band are not indicated. In the superconducting state, energy gaps in the densities of single-particle extended (solid curve in Fig. 10d) and localized states (solid curve in Fig. 10c) appear in the vicinity of the chemical potential. In order to visualize the rearrangement of single-particle spectra upon a superconducting transition more clearly, the spectral densities in the state of a poor metal are also shown in Figs. 10d and 10c (dashed curves). The latter correspond to the results represented in Fig. 9. It can be seen that the position and width of the gap are the same for both types of charge carriers. The obtained value of the gap width is $\Delta_s = 24.36$ meV.

It should be noted that the density of single-particle states in the vicinity of the superconducting gap decreases abruptly, but remains finite both for localized and for extended states. Among other things, this leads to a singularity on the temperature dependence of the electron heat capacity, which will be discussed below. It was found that the density of single-particle states in the superconducting gap decreases with temperature.

The spectral density $\rho^{b0}_{\text{loc}}(\omega)$ of the localized boson state at a site for this superconducting state is shown in Fig. 11. This spectral density is finite both in the region of the initial valence band ($|\omega| \le 1.5$ eV) and in the region of the initial insulator gap. The substitution of $\rho^{b0}_{\text{loc}}(\omega)$ into Eq. (3.21) gives the probability amplitude $\beta = 2.635 \times 10^{-2}$ of filling of the boson state at a lattice site. It should be noted that the substitution $\rho^{b0}_{\text{loc}}(\omega) \longrightarrow -\rho^{b0}_{\text{loc}}(\omega)$ reverses the sign of β, but does not affect the density of single-particle states shown in Fig. 10.

Figure 12 shows the energy distribution of extended charged bosons (solid curves) per impurity atom. It can be seen that the energy distribution of bosons is very broad and is defined both in the region of the initial valence band ($|\omega| \le 1.5$ eV) and in the region of the initial insulator gap. The position of the chemical potential is also shown in Fig. 12. This distribution differs qualitatively from the distribution of the superconducting condensate in the BCS theory. The total concentration of extended bosons defined in the region of the initial band is $N^{b0}_{\text{ext},\,b} = 8.843 \times 10^{-5}$ per impurity atom, while the corresponding total concentration of extended bosons in the region of the initial gap is $N^{b0}_{\text{ext},\,g} = 7.525 \times 10^{-6}$ per impurity atom. Accordingly, the total concentration of extended bosons is $N^{b0}_{\text{ext}} = 0.96 \times 10^{-4} x N_t$. For the results presented in Figs. 10–12, $x = 0.15$. The concentration of localized charged bosons is $N^{b0}_{\text{loc}} = 2\beta^2 x N_t$, where $\beta = 2.635 \times 10^{-2}$ (see above). It can hence be concluded that the concentration of impurity-site (localized) charge bosons in the superconducting state is an order of magnitude higher than the concentration of extended charged bosons.

Such a broad energy distribution of extended charged bosons (see Fig. 12) can be explained as follows. The electron–phonon interaction plays a significant role in the emergence of localized bosons with zero spin at impurity sites. As soon as these localized

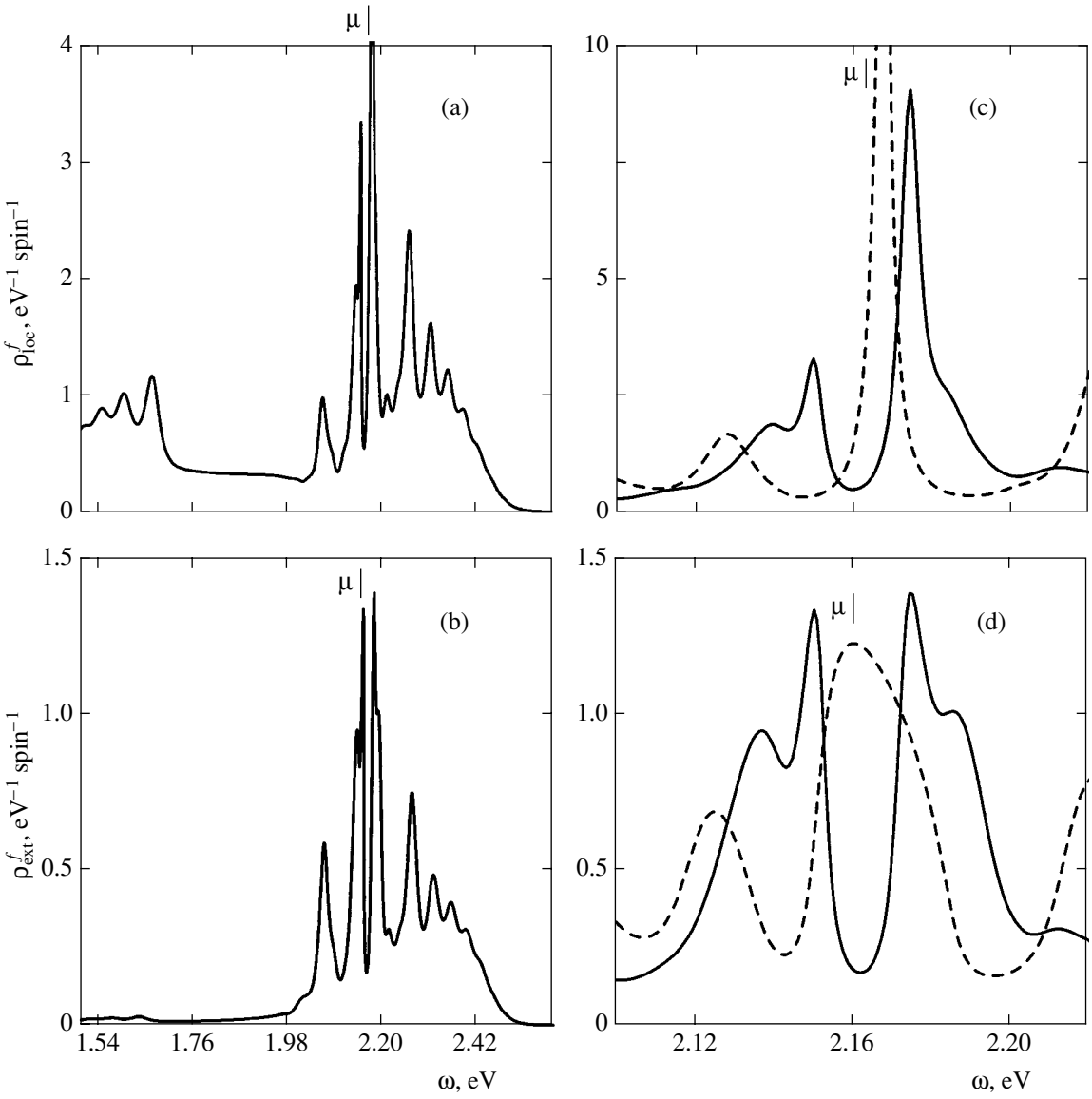

Fig. 10. Densities of single-particle (a) localized and (b) extended states in the vicinity of the initial insulator gap (ω > 1.5 eV) in the superconducting state of a doped compound at $x = 0.15$ and temperature $T = 45$ K. Figures 10c and 10d show the rearrangements of single-particle spectra in the vicinity of the superconducting gap upon a transition from the poor metal state (dashed curves) to the superconducting state (solid curves). Parameters are the same as in Fig. 9.

bosons appear in the system, a channel of two-particle transitions over the lattice-site ensemble is formed: a localized boson at the jth site ⟶ $-\mathbf{k}$, $\mathbf{k}$ pair in the bands of extended states ⟶ a boson at the j_1th site, and so on. The existence of this channel is due to hybridization in model (2.1) and leads to the movement of charged bosons over the impurity sites or, in other words, to the delocalization of bosons and the formation of extended charged bosons. Single-particle transitions induced by hybridization mainly determine the widths of impurity bands for fermion states. These widths correlate with the width of the energy distribution of extended bosons. This becomes clear if we take into account the fact that hybridization also plays a significant role in the formation of such bosons.

Thus, we have obtained a superconducting state in a nondegenerate doped insulator. The conventional equation for the superconducting gap is not required for a superconducting transition as in the BCS theory. Instead of this equation, the leading role is played by the equation determining the formation of localized bosons with zero spin, while the main parameter is the value of β determined by the spectral density of the localized boson state (see Fig. 11). Although the theory does not contain equations for the superconducting gap, this gap appears in the spectrum of single-particle states of extended states. However, the density of single-particle states in the gap can be finite (see Fig. 10). Below, we will consider the thermodynamics of the superconducting state.

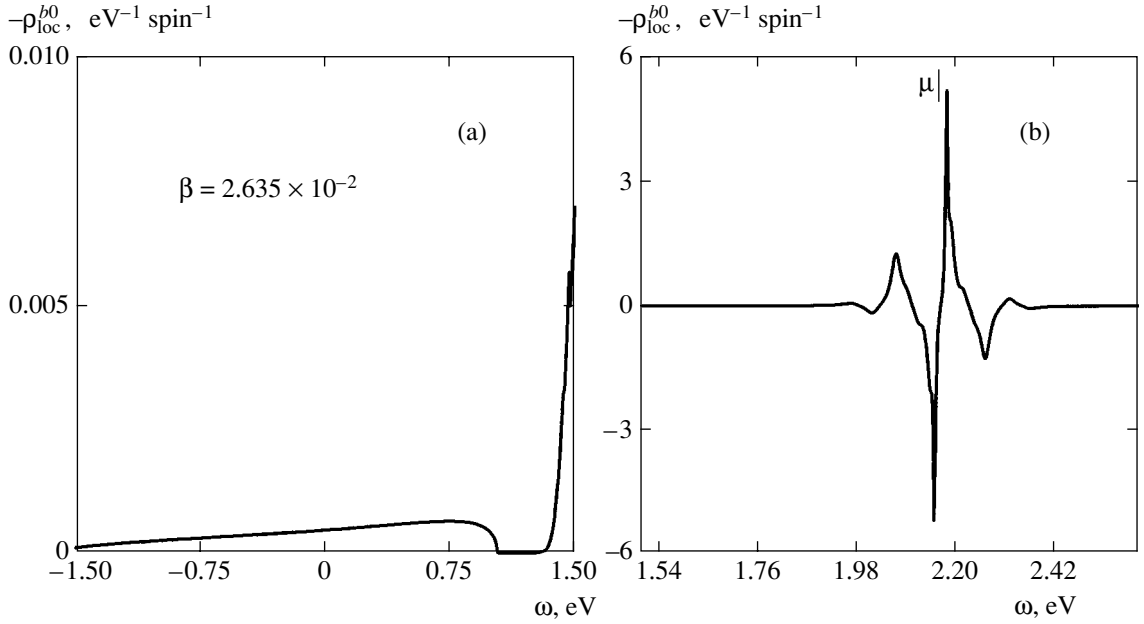

Fig. 11. Spectral density of a localized boson for a mixed boson–polaron state in the superconducting state shown in Fig. 10. The boson filling of a lattice site is equal to the squared coefficient of the localized boson state, which amounts to $\beta = 2.635 \times 10^{-2}$. The single-particle site filling $A = 0.6207$.

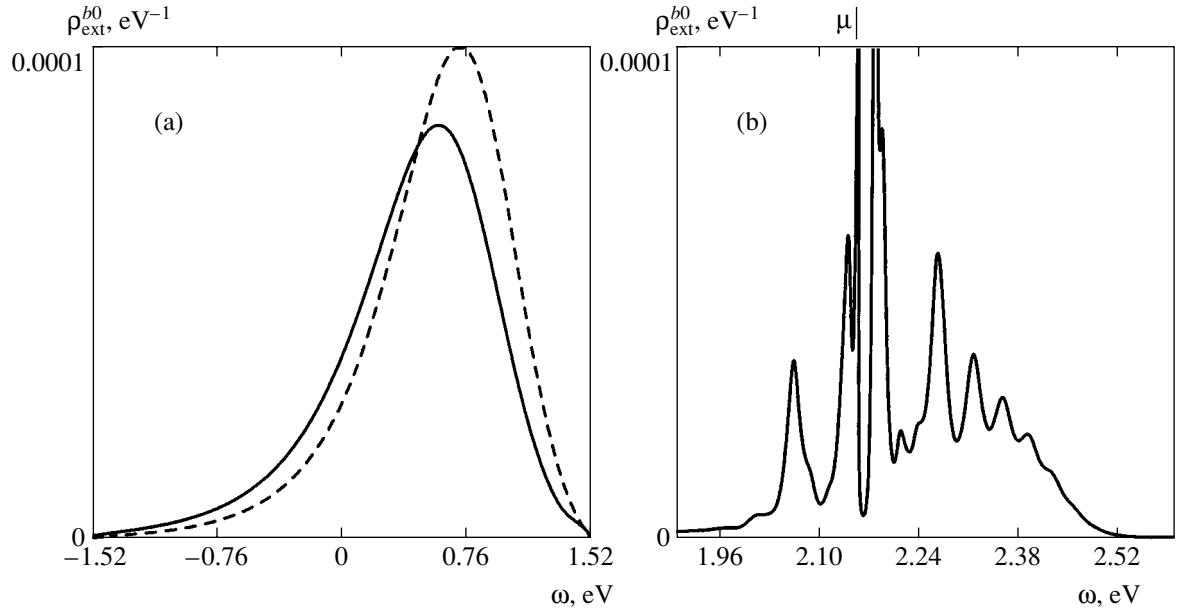

Fig. 12. Energy distribution of extended charged bosons (solid curves) per impurity atom in the superconducting state shown in Fig. 10. The dashed curve indicates the distribution of extended charged bosons over the energy eigenvalues $\varepsilon_{\boldsymbol{k}}$ of the initial insulator.

6.4. Temperature-Induced Superconductor–Metal Transition

For a fixed doping level, the spectral density of a localized boson attenuates upon an increase in temperature. Accordingly, parameter β and the superconducting gap decrease, while the density of single-particle states increases smoothly in the gap region. Such a behavior of spectral densities predetermines the temperature-induced superconductor–metal transition.

The spectral densities of the localized boson state for three different temperatures are shown in Fig. 13. The attenuation of the spectral density upon an increase

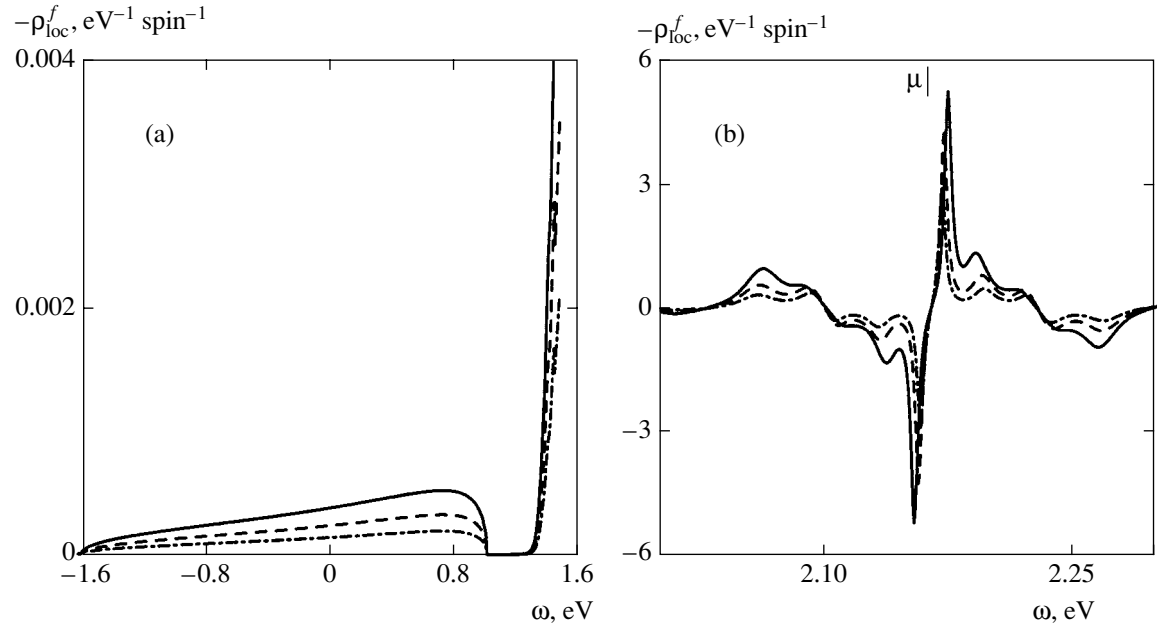

Fig. 13. Spectral densities of the localized boson state for three different temperatures: the solid, dashed, and dot-and-dash curves correspond to T = 50, 54, and 55 K, respectively. The parameters are the same as in Fig. 9.

in T reduces the value of β. For example, $\beta(T = 50\ \mathrm{K}) = 2.146 \times 10^{-2}$, $\beta(T = 54\ \mathrm{K}) = 1.328 \times 10^{-2}$, and $\beta(T = 55\ \mathrm{K}) = 7.78 \times 10^{-3}$. It will be shown below that the latter temperature is close to the superconducting transition temperature $T_c \approx 55.4$ K. It should be noted that the chemical potential practically does not change upon heating and the value of μ or all the three spectral densities is 2.1616 eV.

Figure 14 shows the single-particle densities of localized and extended states in the superconducting states at the same temperatures as in Fig. 13. The spectrum of localized states (Fig. 14a) at T = 50 K displays two clearly manifested peaks in the vicinity of the chemical potential with a dip between the peaks in the density of states. The position of these two peaks correlates with the position of the superconducting gap in the density of extended states at the same temperature (Fig. 14b). The gap width is Δ_s = 19.8 meV. In the region of this gap, the densities of both extended and localized states are finite. As the temperature increases, the densities of single-particle increase and the superconducting gap decreases in the vicinity of the chemical potential. At T = 54 K, we have Δ_s = 12.93 meV. At a higher temperature of T = 55 K, the gap in the density of localized states disappears, but the gap in the density of extended states is still preserved. The gap width determined from the position of two peaks in the density of states in the vicinity of μ; i.e., Δ_s^{ext} = 9.14 meV. Figure 14 also shows the density of states in the normal metallic state at T = 55.5 K immediately after the superconductor–metal transition. The electronic structure of the normal state shown in the figure clearly demonstrates the tendency of how the densities of single-particle states in the superconducting state vary near the transition to the normal state.

Figure 15 shows the energy distributions for extended charged bosons in the superconducting states at the same three temperature values as in Figs. 13 and 14. These distributions are defined both in the initial band of the insulator and in the initial insulator gap. The figure also shows the distributions for extended charged bosons over energy eigenvalues $\varepsilon_{\mathbf{k}}$ of the initial insulator (dashed curves). As the temperature increases, the density of superconducting bosons decreases, but the distributions remain broad. The concentration of extended bosons decreases accordingly. Similarly, a decrease in the probability of a localized boson state being filled leads to a decrease in the concentration of localized bosons.

Figure 16 shows the temperature dependence of parameter β and the gap width Δ_s in the spectra of single-particle states, which are determined from the positions of two peaks in the densities of states in the vicinity of μ. In the vicinity of the superconducting transition temperature $T_c \approx 55.4$ K, both these values are successfully approximated by the characteristic dependences. For example, for $\beta(T)$, we obtained

$$\beta = \beta_0\sqrt{1 - T/T_c}, \tag{6.41}$$

where $\beta_0 \approx 0.0915$. For the superconducting gap width, we have

$$\Delta_s = \Delta_0\sqrt{1 - T/T_c}, \tag{6.42}$$

where $\Delta_0 \approx 107.56$ meV.

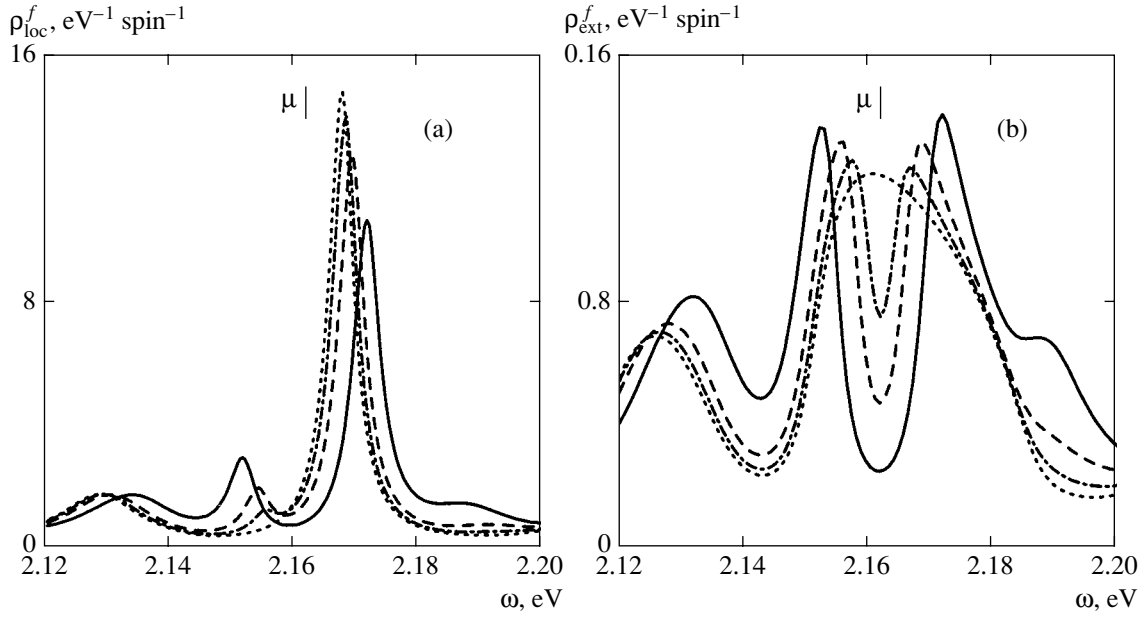

Fig. 14. Densities of single-particle (a) localized and (b) extended states in superconducting states at the same three temperatures as in Fig. 13: the solid, dashed, and dot-and-dash curves correspond to $T = 50$, 54, and 55 K, respectively. The dashed curve represents the densities of states in the normal metallic state at $T = 55.5$ K immediately after the superconductor–metal transition ($T_c \approx 55.4$ K).

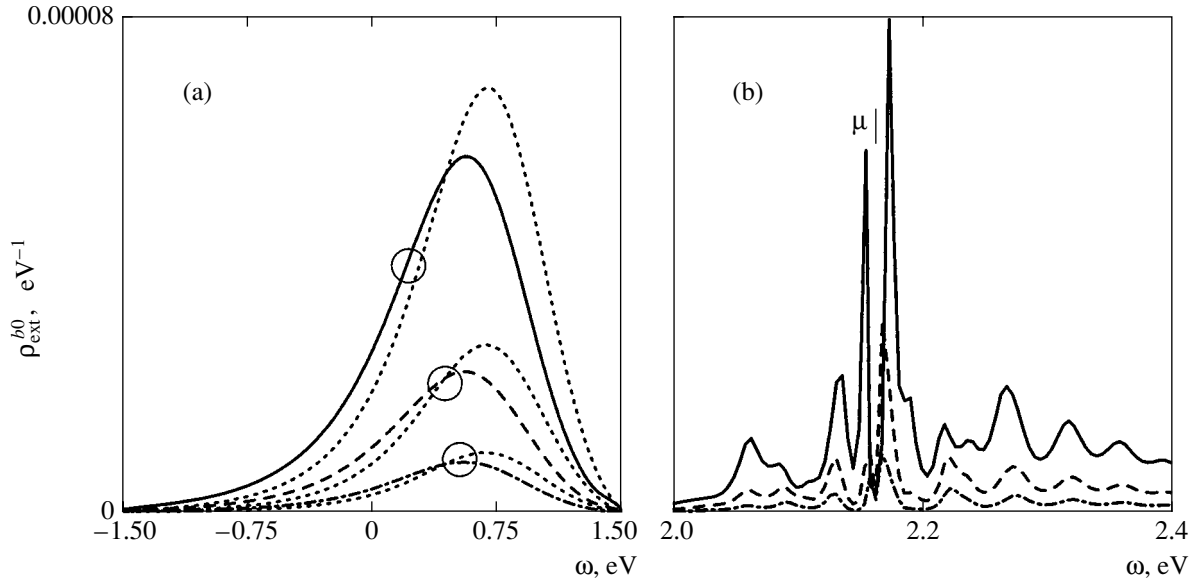

Fig. 15. Energy distribution of extended charged bosons in the superconducting states corresponding to the same three temperatures as in Figs. 13 and 14: the solid, dashed, and dot-and-dash curves correspond to $T = 50$, 54, and 55 K, respectively. The dashed curves represent the distribution of extended charged bosons over energy eigenvalues $\varepsilon_{\mathbf{k}}$ of the initial insulator.

It should be noted that although the $\Delta_s(T)$ dependence shown in Fig. 16 is in qualitative agreement with the result predicted by the BCS theory, there are two differences between them. First, at zero temperature, the ratio $\Delta_s(T = 0)/T_c$ considerably exceeds the value of 3.53 predicted by the BCS theory. In accordance with Fig. 16, this ratio is larger than $\Delta_s(T = 42\ \text{K})/T_c \approx 5.75$. Second, the ratio of Δ_0 (see Eq. (6.42)) to T_c in the vicinity of T_c considerably exceeds the value of 6.12 predicted by the BCS theory. For the result presented here, this ratio is 22.53.

Extended bosons have an energy distribution both in the region of the initial insulator gap and in the region of the initial valence band (see Figs. 12 and 15). The

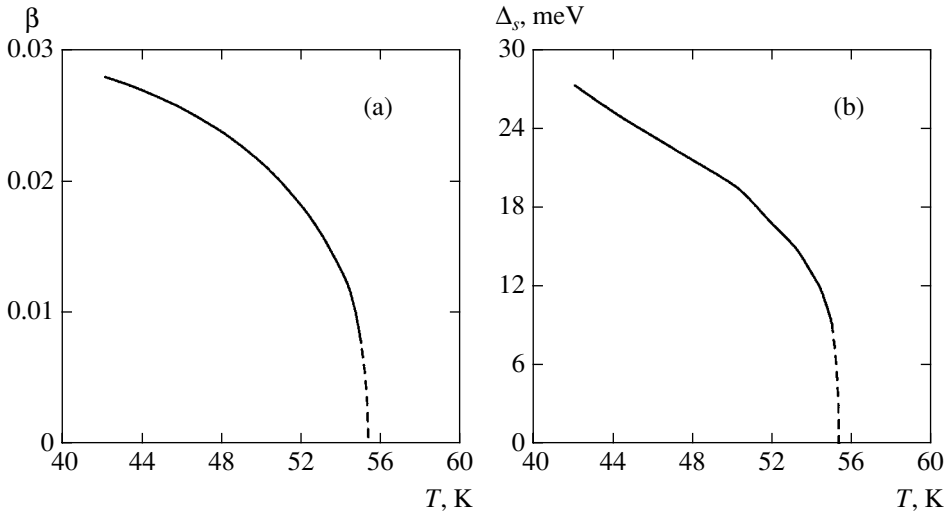

Fig. 16 Temperature dependences $\beta(T)$ of the probability amplitude of filling of a localized boson state and of the gap Δ_s in the spectra of single-particle states, determined from the position of two peaks in the densities in the vicinity of μ. Dashed curves approximate these dependences according to Eqs. (8.41) and (8.42), in which the constants were determined from the data corresponding to the last three points for each dependence. The parameters are the same as in Fig. 9.

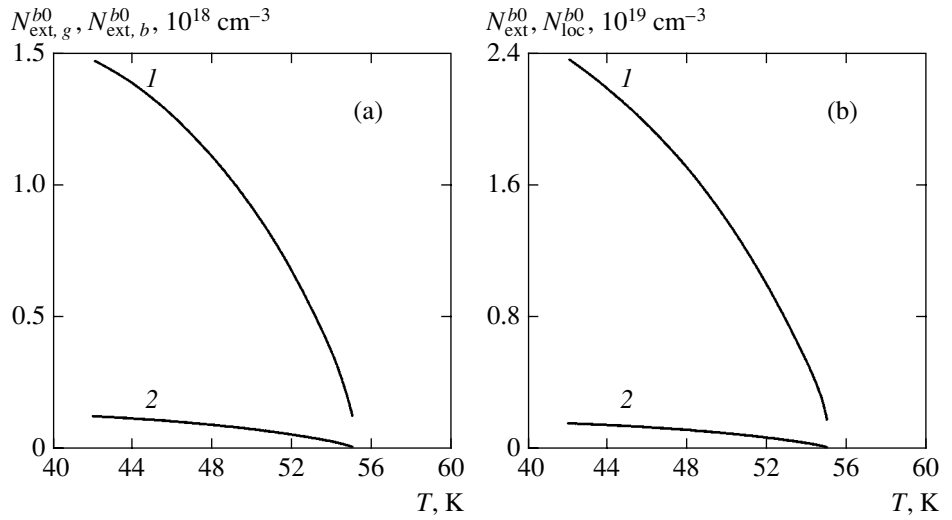

Fig. 17. Temperature dependences of boson concentrations in the superconducting state: (a) concentration $N^{b0}_{\mathrm{ext},\,b}(T)$ of extended bosons in the region of the initial valence band (curve *1*) and $N^{b0}_{\mathrm{ext},\,g}(T)$ in the region of the initial insulator gap (curve *2*); (b) concentration $N^{b0}_{\mathrm{loc}}(T)$ of localized bosons (curve *1*) and the total concentration $N^{b0}_{\mathrm{ext}}(T)$ of extended bosons (curve *2*). The parameters are the same as in Fig. 9 and $N_t = 0.1$ Å^{-3}.

temperature dependences of their concentrations are shown in Fig. 17a. The number of bosons $N^{b0}_{\mathrm{ext},\,g}(T)$ in the initial insulator gap (curve *2* in Fig. 17a) is many times smaller than their number $N^{b0}_{\mathrm{ext},\,b}(T)$ in the initial band (curve *1* in Fig. 17a). The ratio of these concentrations for different values of temperature is found to be constant to a high degree of accuracy and equal to $N^{b0}_{\mathrm{ext},\,g}(T)/N^{b0}_{\mathrm{ext},\,b}(T) \approx 0.0847$.

The temperature dependences of the total concentration $N^{b0}_{\mathrm{ext}}(T)$ of extended bosons (curve *2*) and the con-

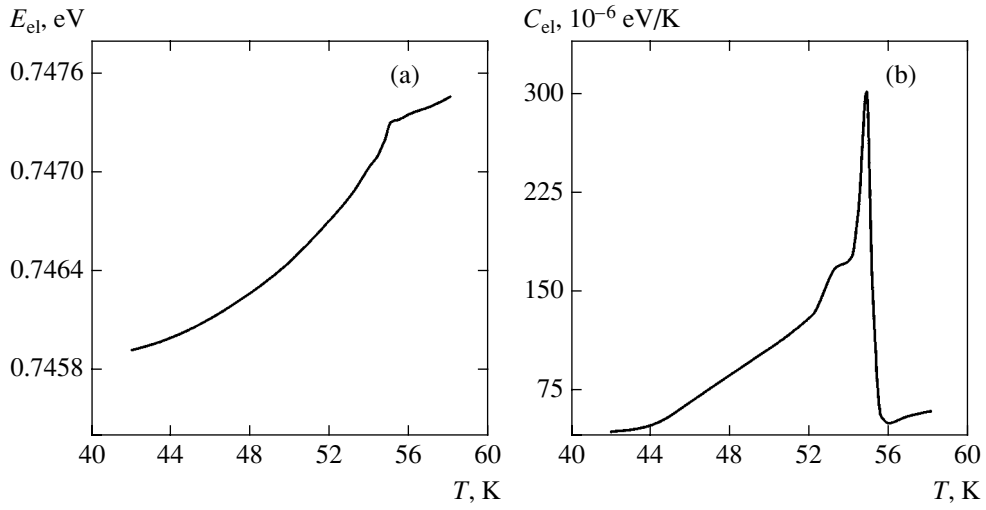

Fig. 18. Temperature dependence (a) of the electron energy for the superconductor–metal transition and (b) of the electron heat capacity. The parameters are the same as in Fig. 9.

centration $N_{loc}^{b0}(T)$ of localized bosons (curve *1*) are presented in Fig. 17b. Both values vanish at the superconducting transition temperature $T_c \approx 55.4$ K. The number of extended bosons is many times smaller than the number of localized bosons. The ratio of these concentrations is found to be constant to a high degree of accuracy and equal to $N_{ext}^{b0}(T)/N_{loc}^{b0}(T) = 0.072$.

The accuracy of calculations of the electron energy up to a temperature of $T \approx 54.5$ K was quite high ($\approx 10^{-5}$ eV), while the typical variation of energy was $\approx 10^{-4}$ eV upon an increase in temperature by 1 K. As T_c was approached, a prolonged relaxation of calculation parameters was observed in the use of the iterative procedure. This deteriorated the accuracy of calculations, although the rate of variation of energy with temperature was considerably higher than the above-mentioned value. The typical variation of energy beyond the transition to the metallic state was approximately equal to 3×10^{-5} eV upon an increase in temperature by 1 K, and the accuracy of its determination was slightly worse than the above value. The temperature dependence of the electron energy upon the superconductor–metal transition is shown in Fig. 18a. The kink in this dependence in the vicinity of the transition is worth noting. The temperature dependence of the electron heat capacity is shown in Fig. 18b. The clearly manifested heat capacity jump is due to the superconducting transition occurring when the temperature decreases below T_c. The transition width amounts approximately to 1 K.

In accordance with Fig. 18b, the heat capacity jump ΔC_{el} is approximately equal to 0.25×10^{-3} eV K^{-1} per impurity atom. In units of the Boltzmann constant k_B, this jump is equal to

$$\Delta C_{el} = 2.9 k_B \tag{6.43}$$

per impurity atom.

In the BCS theory, the heat capacity jump per charge carrier is given by

$$\Delta C_{el} = 7 k_B T_c/\mu. \tag{6.44}$$

Relation (6.43) shows that the heat capacity jump per free charge carrier in the theory of unusual superconductivity in doped insulators is very large as compared to the typical value (6.44) in the BCS theory. It is well known that the heat capacity jump for doped HSC materials is large (on the order of several J mol^{-1} K^{-1}) [102–105]. When recalculated per charge carrier introduced in the material by doping, the observed value of the heat capacity jump is in accordance with the estimate obtained on the basis of the above relation (6.43).

The following remarks can be made in conclusion. Physical properties in the normal metallic state are determined by the Fermi surface and by the density of states. Consequently, for the results described above both for the normal and the superconducting phases, we can find the concentration N_{loc}^{f} of localized states and N_{ext}^{f} of extended states of fermions. As a result, we obtain the following relations between concentrations of various types of charge carriers in the superconducting state of a nondegenerate doped insulator:

$$N_{loc}^{f} \gg N_{ext}^{f} \gg N_{loc}^{b0} \gg N_{ext}^{b0}.$$

These relations clearly demonstrate the peculiar nature of the superconducting state in a doped compound.

6.5. Doping-Induced Superconductor–Metal Transition

The Green function for localized fermions (6.15) and bosons (6.16) contain complex functions S_i, $i = 1$,

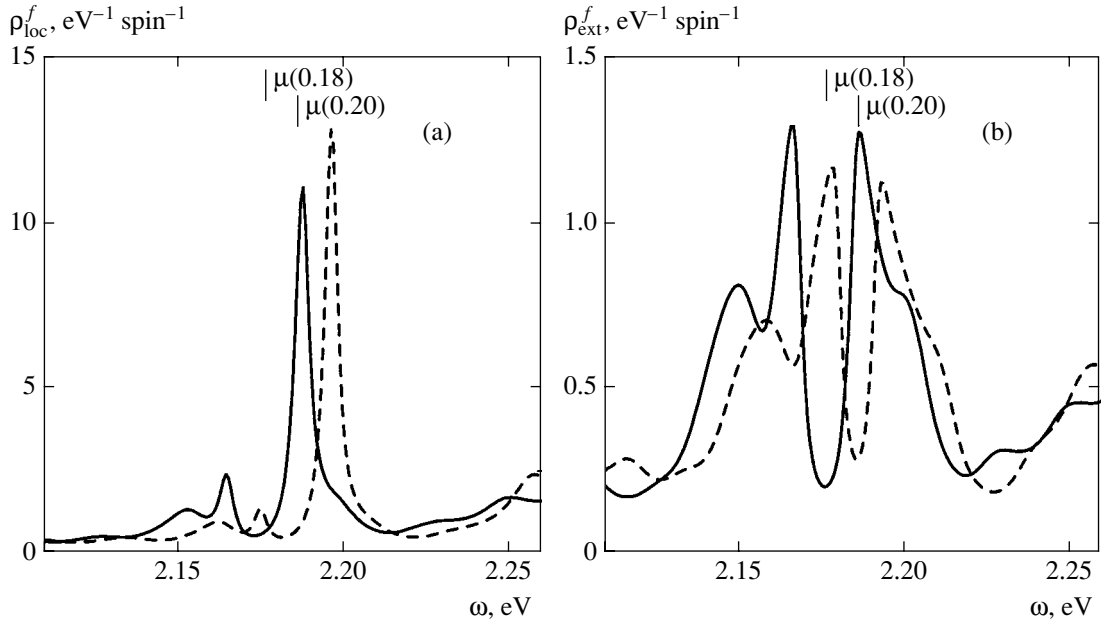

Fig. 19. Variation of single-particle spectra of impurity bands in the vicinity of the chemical potential upon an increase in x at T = 45 K: (a) localized states and (b) extended states. Solid curves correspond to x = 0.18 and dashed curves, to x = 0.2. The parameters (except x) are the same as in Fig. 9.

3 (6.17)–(6.19). The imaginary parts of functions S_i determining the attenuation of these localized particles depend on the doping level x. Similarly, in the Green function for extended fermions (6.20) and bosons (6.21) contain complex functions D_i (i = 1, 3) (6.22)–(6.24). These functions also depend on the doping level x and make a contribution to the attenuation of these extended particles. This leads to a qualitative change in the ground state of the doped compound upon a change in the doping level. We will consider the results of calculating the change in the superconducting state as a function of x for a fixed temperature and discuss the transition of the doped compound to the normal state upon an increase in the value of x.

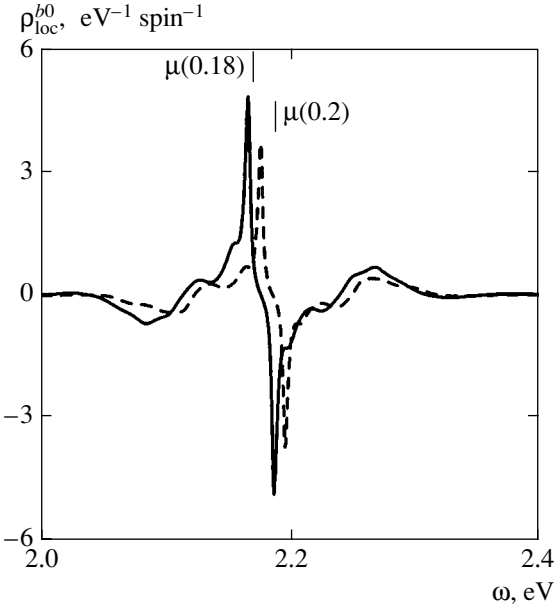

Fig. 20. Spectral densities of the localized boson state in the vicinity of the chemical potential for x = 0.18 (solid curves) and x = 0.2 (dashed curves). The parameters (except x) are the same as in Fig. 9.

Figure 10–12 show the spectral densities in superconducting states at T = 45 K for the doping level x = 0.15. As the doping level increases, parameter β characterizing localized bosons decreases. Figure 19 illustrates the change in the single-particle spectra of impurity bands in the vicinity of the chemical potential upon an increase in x at T = 45 K. For x = 0.18, the gaps in the densities of localized (solid curve in Fig. 19a) and extended states (solid curve in Fig. 19b) are slightly different, their width being Δ_s^{loc} = 22.8 meV and Δ_s^{ext} = 19.7 meV, respectively. The density of states in the gap is finite, and the value of β = 0.0201. An increase in x shifts the distribution to the right; the chemical potential increases, while the value of β decreases. For x = 0.2, we have β = 0.0134. The gap in the density of localized states disappears rapidly, but for x = 0.2 it can still be determined (see the dashed curve in Fig. 19a): Δ_s^{loc} = 20.5 meV. The density of extended states in the gap region (Δ_s^{ext} = 15.1 meV) has increased (see the dashed curve in Fig. 19b) as compared to the density for x = 0.18.

Figure 20 shows the attenuation of the spectral density of a localized boson state with increasing x. This

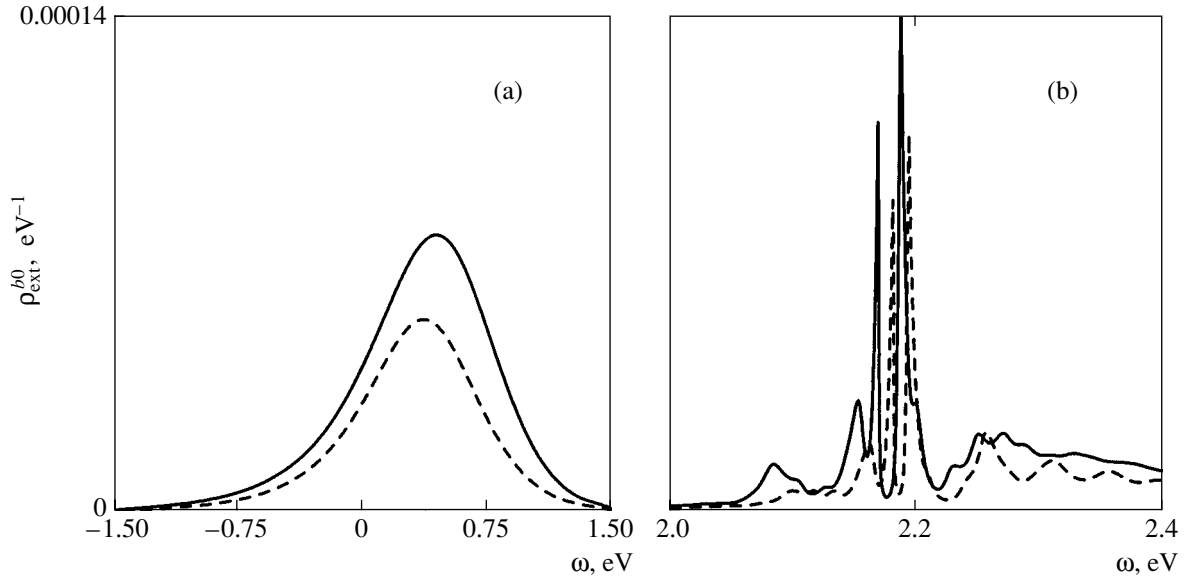

Fig. 21. Energy distributions of extended charged bosons in superconducting states for $x = 0.18$ (solid curves) and $x = 0.2$ (dashed curves). The parameters are the same as in Fig. 20.

leads to a decrease in β. It was found that parameter β and, accordingly, the concentration of localized bosons vanish for $x = 0.228$. Figure 21 shows the energy distributions of extended charged bosons in superconducting states for $x = 0.18$ (solid curves) and $x = 0.2$ (dashed curves). The temperature is the same ($T = 45$ K). These distributions remain broad but decrease with increasing x. It was found that the concentration of extended bosons also decreases, vanishing for $x = 0.228$. Consequently, at this doping level, a superconductor–metal transitions takes place at $T = 45$ K.

Thus, we have established that there exists an upper limit for the doping levels at which a superconducting state may arise in a doped compound at a fixed temperature. In addition, the lower limit for x must also exist. This is associated with the role of spin fluctuations leading to dielectrization of the ground state at low doping levels (see Section 4). The results obtained by us earlier [85] also lead to the conclusion that an insulator–superconductor transition occurs for low values of x. However, we analyzed a purely electronic mechanism of phase transitions in a doped compound, disregarding the electron–phonon interaction.

Unfortunately, we failed to numerically analyze the solution to the general system of equations (3.28)–(3.35) and (3.36)–(3.43) (although this solution can be obtained quite easily) taking into account both the spin fluctuations leading to the formation of a triplet pairing channel (see Section 5 and [85]) and the electron–phonon interaction. The latter may lead to singularities in phase transitions for low values of x. In particular, the insulator–superconductor transition can apparently be realized via an intermediate phase of a poor metal state with strong spin fluctuations (see Section 4.5).

6.6. Doping Dependence of the Superconducting Transition Temperature for Large Doping Levels

The effect of doping level x on the superconducting state at a fixed temperature established above leads to the conclusion that the superconducting transition temperature must also be a function of x. In this section, we will study this dependence for large doping levels.

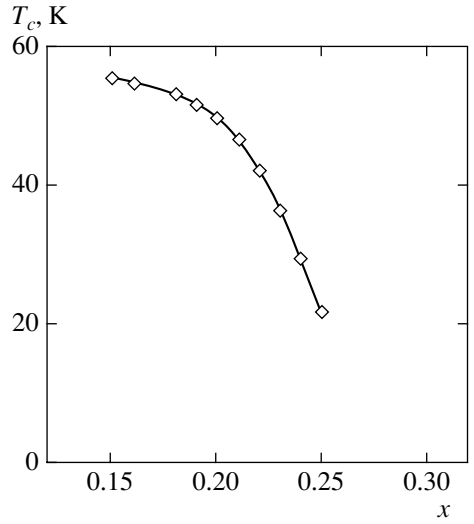

Fig. 22. Dependence of the superconducting transition temperature on the doping level for $x \geq 0.15$. The parameters (except x) are the same as in Fig. 9.

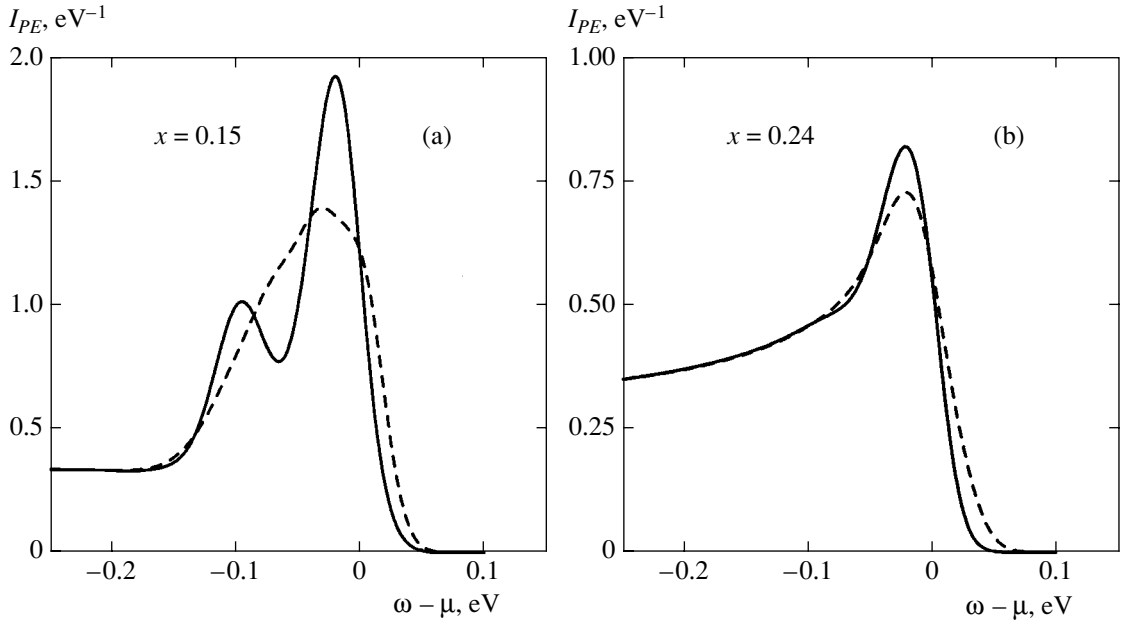

Fig. 23. Photoelectron spectra for superconducting (solid curves) and normal (dashed curves) states for two doping levels $x = 0.15$ and 0.24. For $x = 0.15$, temperature $T = 45$ K for the solid curve and $T = 80$ K for the dashed curve. For $x = 0.24$, temperature $T = 25$ K for the solid curve and $T = 90$ K for the dashed curve. The parameters are the same as in Fig. 9.

Figure 22 shows the $T_c(x)$ dependence for $x \geq 0.15$. It can be seen that the value of T_c decreases with increasing x. However, in the normal phase, the densities of localized and extended states normalized to unit volume increase with the doping level and, hence, we can expect that the conductivity of a doped compound will increase. In spite of this, however, superconductivity is not observed in a more conducting metallic state.

It should be noted that, for values of $x > 0.25$, the accuracy of calculations and the convergence of the iterative procedure become worse. For low doping levels (as compared to those represented in Fig. 22), it is meaningless to carry out calculations since spin fluctuations in the system in the ground state may be significant. However, the effect of spin fluctuations is dual. On the one hand, these fluctuations lead to the emergence of a channel for the formation of triplet bosons in the superconducting state. On the other hand, these fluctuations suppress superconductivity, causing the emergence of a insulator pseudogap in the densities of single-particle states.

6.7. Electron Photoemission Spectra

One of the most distinct feature of the electron angular-resolution photoemission spectra (ARPES) is the so-called peak–dip–hump structure inherent in superconducting cuprates [45–47, 106–109]. The dependence of the photoemission line profile on the wave vector, temperature, and doping level was studied comprehensively. The theoretical interpretations of the observed line profile are contradictory. We will consider the photoemission spectra integrated over the wave vector as function of the temperature and doping level within the theory of superconductivity in doped nondegenerate insulators described above.

The observed photocurrent is determined by the single-particle density of localized and extended states. We assume that the process cross section for both types of charge carriers is the same. The typical total instrumental resolution including the analyzer and the monochromator is on the order of 30–45 meV [46, 106–109]. Consequently, the photocurrent can be written in the form

$$I_{PE}(\omega) = \frac{1}{\delta\sqrt{\pi}}\int_{-\infty}^{\infty} d\omega_1 \frac{(\rho^f_{\text{ext}}(\omega_1) + \rho^f_{\text{loc}}(\omega_1))}{\exp\left(\dfrac{\omega_1 - \mu}{T}\right) + 1} \times \exp\left[-\left(\frac{\omega - \omega_1}{\delta}\right)^2\right], \tag{6.45}$$

where δ is determined by the instrumental resolution. In our calculations, we assumed that $\delta = 23$ meV. The energy was measured from the chemical potential level in the superconducting state.

Figure 23 shows the photoelectron spectra for superconducting (solid curves)) and normal (dashed curves) states for two doping levels $x = 0.15$ and 0.24. It should be noted above all that the convolution of single-particle densities of states with the Gaussian distribution function of the instrumental resolution leads to a considerably smoothing of photoelectron spectra. This becomes clear if we compare the spectra recorded for

$x = 0.15$ (Fig. 23) with the densities of states shown in Figs. 9 and 10 for the same doping level. For $x = 0.15$, the transition from the normal ($T = 80$ K) to the superconducting state ($T = 45$ K) is accompanied by a change in the photoelectron spectrum so that the line profile acquires a peak–dip–hump structure. For a worse instrumental resolution (higher value of δ), the dip in the spectrum decreases. A different situation is observed for the higher doping level $x = 0.24$. No structure of this type appears in the line profile upon a transition from the normal ($T = 90$ K) to the superconducting state ($T = 25$ K), but a rearrangement of the single-particle density of states can clearly be seen.

Thus, the singularity in the electron photoemission spectra associated with the peak–dip–hump profile of the line can be attributed to the rearrangement of the bands of localized and extended states coexisting in the vicinity of the Fermi surface without using additional assumptions such as, the existence of local Cooper pairs or the emergence of new branches of collective excitations in HTSC materials.

6.8. Isotopic Effect

The possibility of the phonon mechanism of superconductivity is often substantiated by the observation of isotopic effect. When an atom of mass M_1 is replaced by its isotope of mass M_2, the expected change in the frequency of modes associated mainly with the motion of the atom is given by the approximate relation $\Omega_1/\Omega_2 = \sqrt{M_2/M_1}$. In the BCS theory, the dependence of T_c on the ion mass appears in view of the relationship between the Debye temperature and M. The predicted dependence has the form $T_c M^\alpha = \text{const}$, where $\alpha \leq 1/2$ depends on the band structure of the metal, including regions far from the Fermi surface (both above and below it) and on the electron–electron interaction. Naturally, we are speaking of complete substitution of an isotope.

At present, a large body of experimental data have been obtained that indicate that the electron–phonon interaction in HTSC materials is strong (see the Introduction). However, the data on the isotopic shift of the superconducting transition temperature in cuprates are contradictory. For example, it was reported in [110–112] that the substitution of ^{18}O for ^{16}O in cuprates leads to a small isotopic shift in T_c, which is characterized by the value of $\alpha \approx 0.03$. On the other hand, an isotopic shift in T_c, which is comparable to the prediction of the BCS theory or even a considerably larger shift with $\alpha \approx 1$, was observed in [113–118] as a result of such a substitution. The concept of isotopic shift in T_c of HTSC cuprates has also been formulated in [113, 119, 120]. In the underdoped region, i.e., for a doping level lower than the optimal value, for which the value of T_c is maximal for a given compound, the substitution of ^{18}O for ^{16}O leads to an isotopic shift in T_c, which is larger than or comparable to the result predicted by the BCS theory. In the overdoped region with large values of x, this shift is smaller.

However, some doping regions for various cuprates may be unattainable in experiments. This may lead to singularities in the observed behavior of $\alpha(x)$ and to some discrepancies in the statements concerning the magnitude of the isotopic shift. On the phase diagram of $La_{2-x}Sr_xCuO_4$, there exists a critical doping level $x \approx 0.05$ for which a superconducting transition takes place. As the value of x increases, the value of T_c becomes higher, attaining its maximum for the optimal value of $x \approx 0.15$. Then the value of T_c decreases in the overdoped region and becomes equal to zero for $x \approx 0.3$. This $T_c(x)$ curve has the same form typical of doped cuprates. A similar phase diagram was obtained for compounds $HgBa_2CuO_{4+x}$ [121, 122] and $Bi_2Sr_2CuO_{6+x}$ [123], for which both doping regions and, hence, both limiting points of x for which superconductivity exists are attainable in experiment. In other cuprates $Bi_2Sr_2CaCu_2O_{8+x}$ [123] and $YBa_2Cu_3O_{7-x}$ [124], only the regions up to the optimal doping level are experimentally attainable.

In the solution (6.15)–(6.26) obtained for the superconducting state, a dependence of T_c on the phonon energy $\hbar\Omega$ and, hence, on the ion mass appears due to the dependence of self-energy parts (6.36)–(6.38) on Ω. The definitions of these self-energy parts contain only the spectral density $\rho^f_{\text{loc}}(\omega)$ of a localized fermion for Eqs. (6.36)–(6.37) and the spectral density $\rho^{b0}_{\text{loc}}(\omega)$ of a localized singlet boson for Eq. (6.38).

In $La_{2-x}Sr_xCuO_4$, the cuprate layer is between two intermediate layers having the composition $La_{1-x/2}Sr_{x/2}O$. The charge carriers introduced in the system during doping are initially localized. Proceeding from the value of the effective Bohr radius, the region of localization of a charge carrier may include positions of the Sr ion and four O ions closest to it. Analysis of the effect of isotope ^{18}O being substituted for oxygen ^{16}O on the effective frequency of modes associated mainly with the motion of O and Sr atoms is, strictly speaking, a separate problem. Here, we describe the results on the expected maximal isotopic shift of T_c for the phonon mechanism of formation of the superconducting state in doped nondegenerate insulators. To analyze this shift in the solution obtained in Section 6, we carry out the substitution $\Omega \longrightarrow \Omega\sqrt{16/18}$ for the phonon frequency without changing the remaining parameters.

For the parameters represented in Fig. 9 and for $x = 0.15$, the superconducting transition temperature is approximately equal to 55.4 K. This value of T_c was obtained for the phonon energy $\hbar\Omega = 30$ meV. The corresponding results of investigation of the superconducting state and the superconductor–metal transition upon a change in x and T are shown in Figs. 10–18. The value

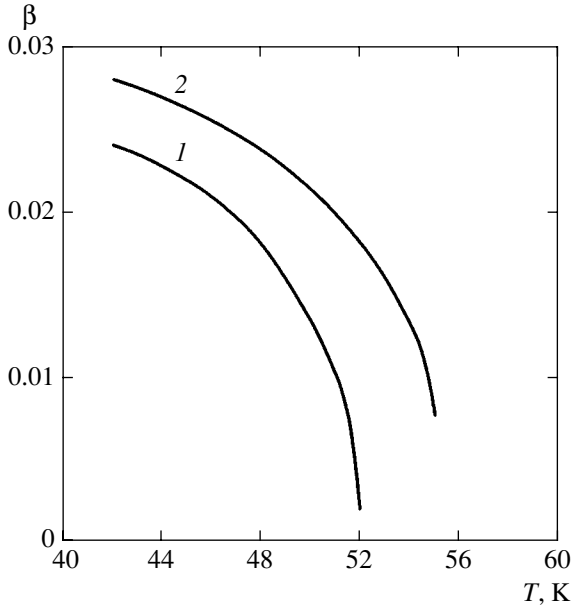

Fig. 24. Temperature dependences $\beta(T)$ of the probability amplitude of filling of a localized boson state for $x = 0.15$; curve *1* corresponds to $\Omega = 30\sqrt{16/18}$ meV, curve *2* corresponds to the dependence depicted in Fig. 16.

of T_c decreases with decreasing phonon energy. Curve *1* in Fig. 24 describes the temperature dependence $\beta(T)$ of the probability amplitude of filling of a localized boson state for the same parameters, including x, except $\hbar\Omega = 30\sqrt{16/18}$ meV ≈ 28.28 meV. In order to describe the isotopic effect more clearly, we also show the $\beta(T)$ dependence for $\Omega = 30$ meV (curve *2* in Fig. 24, which corresponds to the dependence shown in Fig. 16). It was found that the change in the phonon frequency for the same remaining parameters represented in Fig. 9 lowers the superconducting transition temperature to a value equal to $T_c \approx 52.07$ K. Consequently, $\alpha = 0.526$ for $x = 0.15$, which is close to the results obtained in the BCS theory.

It should be noted that, in accordance with Fig. 24, the quantity of $\beta(T)$ tends to a constant value at low temperatures. This value also depends on the phonon frequency. For example, for $T = 42$ K, we have $\beta = 0.02808$ for $\hbar\Omega = 30$ meV. The substitution $\Omega \longrightarrow \Omega\sqrt{16/18}$ reduces the value of β to 0.02412 at the same temperature. In this case, for parameter α_1 appearing in the dependence $\beta M^{\alpha_1} = \text{const}$, we obtain $\alpha_1 = 1.29$.

A change in the phonon frequency for the same remaining parameters considerably changes the electronic structure of the superconducting state. Figure 25a shows the densities of single-particle localized (curve *1*) and extended (curve *2*) states in the vicinity of the chemical potential in the superconducting state at a temperature of $T = 52$ K, which is very close to the superconducting transition temperature $T_c \approx 52.07$ K. Here, we used the changed phonon frequency. It should be noted that an insignificant dip is still present in the density of extended states (see the inset to Fig. 25a). The spectral density of localized bosons in the vicinity of μ is also shown in Fig. 25b. This density gives a small value of parameter $\beta = 0.0021$.

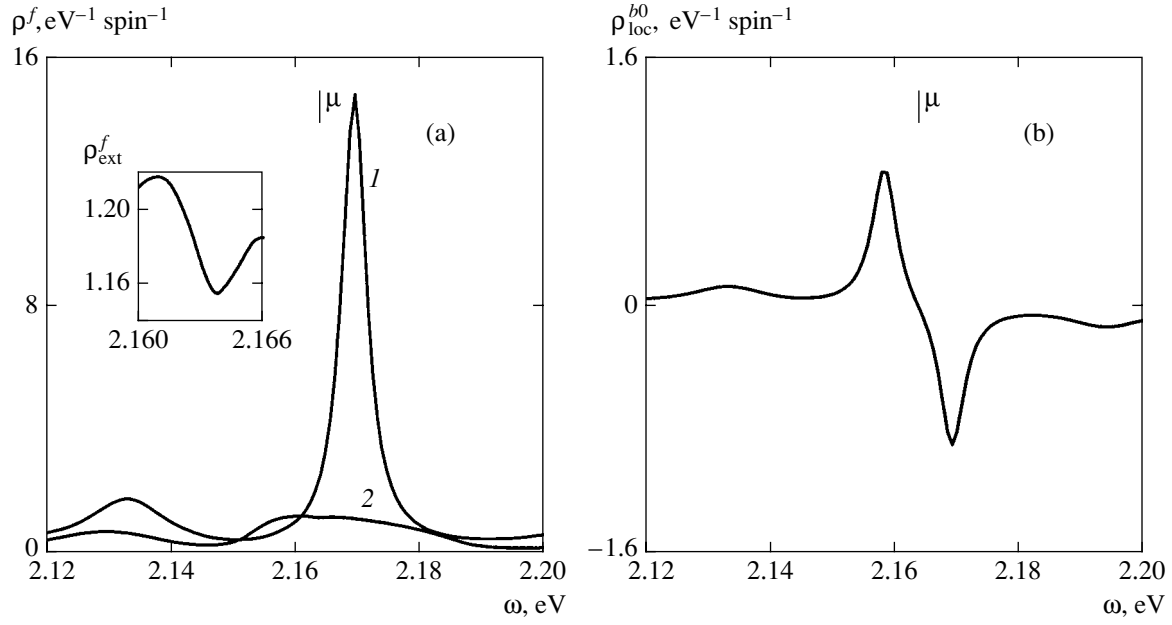

Fig. 25. (a) Densities of single-particle localized (curve *1*) and extended (curve *2*) states in the vicinity of the chemical potential in the superconducting state at temperature $T = 52$ K; the inset shows the singularity in the density of extended states. (b) Spectral density of localized bosons near μ. The parameters (except the phonon energy $\hbar\Omega$) are the same as in Fig. 9.

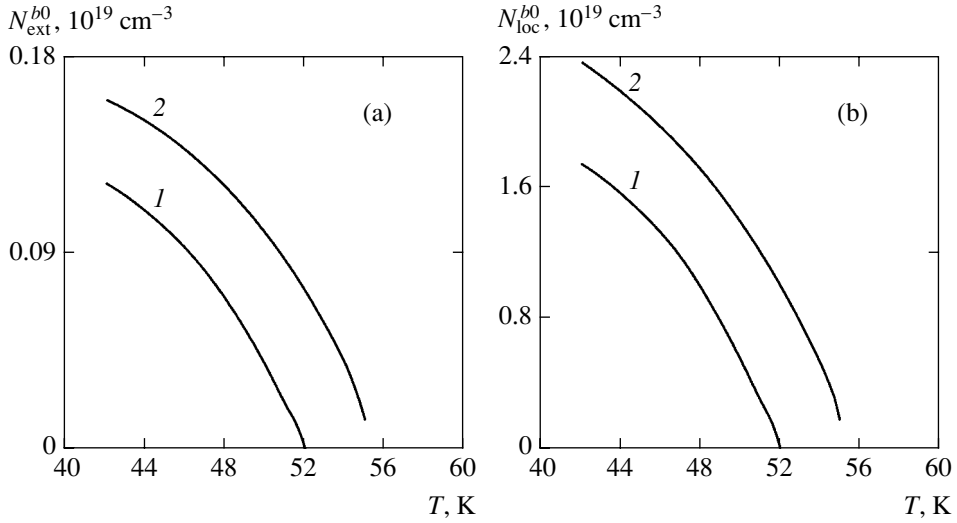

Fig. 26. Temperature dependences of the concentration of extended bosons (curve *1* in (a)) and localized bosons (curve *1* in (b)) in the superconducting state obtained for $x = 0.15$ and $\hbar\Omega = 30\sqrt{16/18}$ meV. Curves *2* show the corresponding dependences for $\hbar\Omega = 30$ meV. The remaining parameters are the same as in Fig. 9.

The isotopic effect is also manifested in the temperature dependence of the concentration of extended bosons (curve *1* in Fig. 26a) and localized bosons (curve *1* in Fig. 26b). These curves were obtained for a changed boson frequency. The critical temperature at which these boson concentrations vanish is equal to $T_c \approx 52.07$ K. In order to demonstrate the isotopic effect clearly, we also show in Fig. 26 the dependences of boson concentrations for $\hbar\Omega = 30$ meV with $T_c \approx 55.4$ K (curves *2* in Figs. 26a and 26b correspond to the dependences shown in Fig. 17). It can be seen that a decrease in the phonon energy decreases the boson concentration. It should be noted that curve *1* in Fig. 26a represents the total concentration of extended bosons, while the ratio of the number of bosons $N^{b0}_{ext,\,g}(T)$ in the initial insulator gap to the number of boson $N^{b0}_{ext,\,b}(T)$ in the region of the initial valence band is constant and equal to $N^{b0}_{ext,\,g}(T)/N^{b0}_{ext,\,b}(T) = 0.0823$. The number of extended bosons is many times smaller than the number of localized bosons. The ratio of these concentration is constant to a high degree of accuracy and is equal to $N^{b0}_{ext}(T)/N^{b0}_{loc}(T) = 6.95 \times 10^{-2}$. Both these ratios are slightly smaller than those obtained for $\hbar\Omega = 30$ meV.

It should be noted that, in accordance with Fig. 26, the boson concentrations display a tendency to saturation at low temperatures. These limiting values of concentrations also depend on the phonon frequency. We can estimate this influence of the isotopic effect from the relation $N^{b0}M^{\alpha_2} = \text{const}$, using the results obtained for the lowest temperature $T = 42$ K. For example, $\alpha_2 = 2.58$ for localized bosons and $\alpha_2 = 2.32$ for extended bosons.

Let us now study the isotopic effect for a high doping level. For $x = 0.24$ and phonon energy $\hbar\Omega = 30$ meV, temperature $T_c = 29.24$ K (see Fig. 22), and the temperature dependence $\beta(T)$ of the probability amplitude of filling of a localized boson state is represented by curve *1* in Fig. 27. The replacement of the phonon frequency by $\hbar\Omega = 30\sqrt{16/18} \approx 28.28$ meV without changing the remaining parameters considerably reduces the temperature region in which $\beta(T) \neq 0$ and reduces the superconducting transition temperature to $T_c \approx 25.72$ K (curve *2* in Fig. 27). The data obtained for the isotopic shift of T_c show that $\alpha = 1.08$ for $x = 0.24$, which is much higher than the limiting value predicted by the BCS theory.

The influence of the isotopic effect on the temperature dependences of concentrations of localized bosons for $x = 0.24$ is illustrated in Fig. 28, where curve *1* was obtained for the phonon energy $\hbar\Omega = 30$ meV, while curve *2* was obtained for $\hbar\Omega = 30\sqrt{16/18}$ meV $\approx$ 28.28 meV. In the vicinity of T_c, these dependences are linear. It can be seen that a decrease in the phonon energy leads to a sharp decrease in the concentration of localized bosons. The ratio of these concentrations for the two dependences presented here is approximately equal to 16.59 for $T = 25$ K. An equally large difference in the concentrations is also observed for extended bosons since it was found that the ratio of the total concentration $N^{b0}_{ext}(T)$ of extended bosons to the concentration $N^{b0}_{loc}(T)$ of localized bosons is almost constant and equal approximately to 0.377 for $x = 0.24$ for both values of the phonon energy.

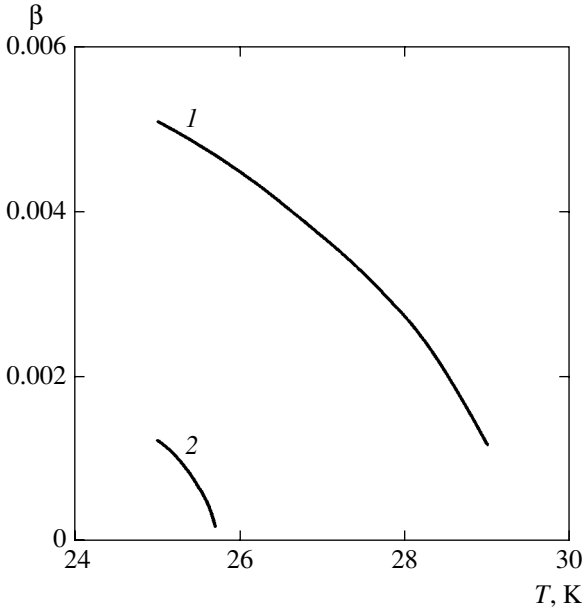

Fig. 27. Temperature dependences $\beta(T)$ of the probability amplitude of filling of the localized boson state for $x = 0.24$; curve *1* was obtained for the phonon energy $\hbar\Omega = 30$ meV; curve *2* was obtained for $\hbar\Omega = 30\sqrt{16/18}$ meV. The superconducting transition temperature for curve *1* corresponds to T_c in Fig. 22 for $x = 0.24$.

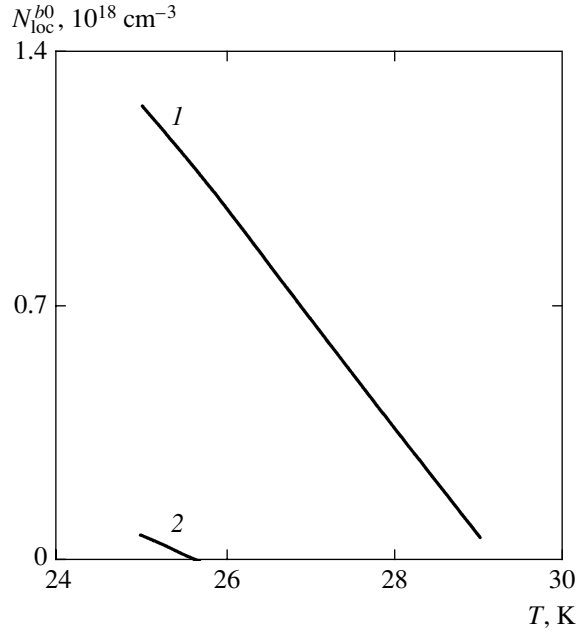

Fig. 28. The influence of the isotopic effect on the temperature dependences of concentrations of localized bosons for $x = 0.24$. Curve *1* was obtained for the phonon energy $\hbar\Omega = 30$ meV and curve *2*, for $\hbar\Omega = 30\sqrt{16/18}$ meV ≈ 28.28 meV. The remaining parameters are the same as in Fig. 27.

Thus, the possibility of the isotopic effect having a strong influence on the boson concentration in the superconducting state at temperatures differing considerably from T_c can be predicted.

7. CONCLUSIONS

Analysis of available experimental data on the change in the electronic structure and the position of the chemical potential upon a change in the doping level in superconducting tungsten bronze Na_xWO_3 and cuprates $La_{2-x}Sr_xCuO_4$ and $Nd_{2-x}Ce_xCuO_4$ stimulated the study of the general problem of superconductivity in impurity bands of doped insulators. The emergence of superconductivity in the above-mentioned three doped compounds is associated with the formation of impurity bands in the insulator gap of initial crystals, while the Fermi system has two components. The concentration of free charge carriers in the vicinity of the chemical potential for these cuprates is almost an order of magnitude lower than the concentration of localized charge carriers.

A theory of insulator–metal and superconductor–metal phase transition in doped nondegenerate insulator has been constructed. In the normal phase, a insulator state can be induced by spin fluctuations in the doped system at low doping levels and temperatures. Insulator pseudogaps Δ_p are formed in the spectral densities of impurity bands for localized and extended states. The gap width is determined both by the doping level and by the temperature. Spin fluctuations attenuate in the doped system, and, accordingly, $\Delta_p \longrightarrow 0$ upon an increase in the doping level and/or temperature. However, the density of extended states in the pseudogap, which is smaller than the density of localized states, increases smoothly even for final values of the pseudogap width. This leads to the scenario of a smooth insulator–metal transition upon an increase in the temperature or the doping level. For example, the characteristic temperature at which the semiconductor-type dependence of conductivity changes to the metallic type decreases smoothly with an increase in the doping level. This metallic state is peculiar for at least two reasons. First, regions of localized and extended states coexist on the Fermi surface, the density of localized states being much higher than the density of extended states. For this reason we can expect that such a poor metal will display anomalous transport properties as compared to conventional metals. Second, single-particle Green functions in this state are nondiagonal in the spin subspace. The presence of spin fluctuations may lead to singularities in the behavior of the spin susceptibility of a doped compound. Analysis of the effect of the insulator band width on the insulator–metal transition proved that the necessary condition for this transition is the existence of narrow allowed bands for the initial insulator.

The superconducting phase in a doped nondegenerate compound exhibits rather peculiar properties as compared to the predictions of the BCS theory. The mechanism of a superconducting transition in a doped nondegenerate insulator is associated with the formation of mixed boson–fermion states localized at impurity sites. In contrast to the BCS theory, the theory of

this unusual superconductivity does not contain an equation for the superconducting gap. Instead of this equation, a nontrivial solution to the equation for local singlet bosons at impurity sites plays the key role for the emergence of superconductivity. The formation of such bosons is a precursor for the emergence of extended charged bosons in the doped system. Although the theory does not contain an equation for the superconducting gap, this gap appears in the spectrum of single-particle states of extended states. However, the density of single-particle states in the gap may be finite. The obtained relations between the concentrations N_{loc}^{f} of localized fermions, N_{ext}^{f} of extended fermions, N_{loc}^{b0} of localized bosons, and N_{ext}^{b0} of extended bosons clearly indicate the peculiar nature of the superconducting state $N_{\mathrm{loc}}^{f} \gg N_{\mathrm{ext}}^{f} \gg N_{\mathrm{loc}}^{b0} \gg N_{\mathrm{ext}}^{b0}$.

The energy distribution for extended bosons differs qualitatively from the energy distribution of the superconducting condensate in the BCS theory. As a matter of fact, the electron–phonon interaction plays a significant role in the formation of zero-spin bosons localized at impurity sites. As soon as such localized bosons appear in the system, a channel of two-particle transition over the impurity-site ensemble is formed: a boson localized at the jth site $\longrightarrow$ **–kk** pair in the regions of extended states $\longrightarrow$ a boson at the j_1th site, and so on. This channel is due to hybridization in model (2.1) and leads to the motion of charged bosons over the impurity sites or, in other words, to delocalization of bosons and the formation of extended charged bosons. Single-particle transitions induced by hybridization mainly determine the widths of the impurity bands for fermion states. These bandwidths correlate with the width of the energy distribution of extended bosons since hybridization also plays a significant role in the formation of such bosons. Typically, the number of extended bosons in the region of the initial band, away from the chemical potential, is more than an order of magnitude larger than the number of bosons in the gap region in the vicinity of the chemical potential. In the general case, both the singlet and the triplet pairing channels are possible in the superconducting state. The spin-triplet channel is realized only if the spectral density of spin fluctuations in the doped compound is finite. In this case, single-particle Green functions are nondiagonal in the spin index, which may lead to singularities in the spin susceptibility of the compound. For the phonon mechanism of superconductivity with a spin-singlet pairing channel, a superconductor–metal transition induced by variations of temperature and/or doping level in a doped compound is predicted as well as a strong isotopic effect as compared to the limiting value predicted by the BCS theory for weak and strong electron–phonon interactions. Such transitions are observed in doped HTSC materials. The obtained temperature width of the superconducting transition (≈1 K) and the heat capacity jump per charge carrier ($\approx k_B$) are also typical of cuprates. As regards the isotopic shift of T_c, which is large according to predictions ($\alpha \approx 1$), the available information on the value of this shift in cuprates is contradictory.

The observation of superconductivity (with any value of T_c) in other doped nondegenerate insulators would be a good verification of the theory. The optimal composition of these compounds must contain three components for interstitial impurities and four components for substitutional impurities (an additional component is required for an element replaced by an impurity). A insulator must possess a large capacity for the impurity being introduced (≈10 at.% without precipitation of impurity phases). The remaining two components must satisfy the following requirements. On account of the required small width of allowed bands, the initial insulator must contain an element from transition or rare-earth metals. The atomic (*d* or *f*) orbitals of this element must participate in the formation of the band structure of the insulator in the vicinity of the insulator gap. An important parameter of the theory is the electron–phonon interaction. For this reason, the last component must be a light electronegative element enhancing the ionic bonds in the initial material.

ACKNOWLEDGMENTS

This study was partly supported finantially by the Russian Foundation for Basic Research (project no. 02-02-16320).

REFERENCES

1. D. Pines, Phys. Rev. **109**, 280 (1958).
2. M. Cohen, in *Superconductivity*, Ed. by R. D. Parks (Marcel Dekker, New York, 1969; Mir, Moscow, 1972), Vols. 1, 2.
3. A. I. Ansel'm, *Introduction to the Theory of Semiconductors* (Nauka, Moscow, 1978).
4. J. S. Blakemore, *Semiconductor Statistics* (Pergamon Press, Oxford, 1962; Mir, Moscow, 1964).
5. J. F. Schooley, W. R. Hosler, and M. L. Cohen, Phys. Rev. Lett. **12**, 474 (1964).
6. E. Ambler, J. H. Colwell, W. R. Hosler, and J. F. Schooley, Phys. Rev. **148**, 280 (1966).
7. R. A. Hein, J. W. Gibson, R. S. Allgaier, *et al.*, in *Low Temperature Physics, LT9*, Ed. by J. G. Daunt *et al.* (Plenum, New York, 1965), p. 604.
8. G. Hoppmann and E. Salje, Opt. Commun. **30**, 199 (1979); F. P. Koffyberg, K. Dwight, and A. Wold, Solid State Commun. **30**, 433 (1979).
9. P. Hagenmuller, Progr. Solid State Chem. **5**, 71 (1971).
10. A. Magneli, Nature **169**, 791 (1952).
11. Ch. J. Raub, A. R. Sweedler, M. A. Jensen, *et al.*, Phys. Rev. Lett. **13**, 746 (1964).
12. A. J. Bevolo, H. R. Shanks, P. H. Sidles, and G. C. Danielson, Phys. Rev. B **9**, 3220 (1974).
13. M. R. Skokan, W. G. Moulton, and R. C. Morris, Phys. Rev. B **20**, 3670 (1979).

14. W. A. Kamitakahara, K. Schrnberg, and H. R. Shanks, Phys. Rev. Lett. **43**, 1607 (1979).
15. L. H. Cadwell, R. C. Morris, and W. G. Moulton, Phys. Rev. B **23**, 2219 (1981).
16. H. Srikanth and A. K. Raychaudhuri, J. Phys.: Condens. Matter **5**, L551 (1993).
17. H. R. Shanks, Solid State Commun. **15**, 753 (1974).
18. G. Hollinger, P. Pertosa, J. P. Doumerc, *et al.*, Phys. Rev. B **32**, 1987 (1985).
19. N. F. Mott, Philos. Mag. **35**, 111 (1977).
20. H. Hochst, R. D. Bringans, and H. R. Shanks, Phys. Rev. B **26**, 1702 (1982).
21. S. Reich and Y. Tsabba, Eur. Phys. J. B **9**, 1 (1999).
22. A. Shengelaya, S. Reich, Y. Tsabba, and K. A. Muller, Eur. Phys. J. B **12**, 15 (1999).
23. P. W. Anderson, Science **235**, 1196 (1987).
24. E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994).
25. M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1998).
26. T. Tohyama and S. Maekawa, Physica C (Amsterdam) **191**, 193 (1992).
27. N. Bulut, D. J. Scalapino, and S. R. White, Phys. Rev. B **50**, 7215 (1994).
28. S. Haas, A. Moreo, and E. D. Dagotto, Phys. Rev. Lett. **74**, 4281 (1995).
29. A. Moreo, S. Haas, A. W. Sandvik, and E. Dagotto, Phys. Rev. B **51**, 12045 (1995).
30. R. Preuss, W. Hanke, C. Grober, and H. G. Evertz, Phys. Rev. Lett. **79**, 1122 (1997).
31. Yu. V. Kopaev, Usp. Fiz. Nauk **172**, 712 (2002).
32. B. Reichl, T. Riesterer, J. G. Bednorz, and K. A. Muller, Phys. Rev. B **35**, 8804 (1987).
33. G. M. Éliashberg, in *Physical Properties of High Temperature Superconductors*, Ed. by D. M. Grinberg (World Sci., Singapore, 1990; Mir, Moscow, 1990).
34. Z.-X. Shen and D. S. Dessau, Phys. Rep. **253**, 1 (1995).
35. C. Quitmann, J. Ma, R. J. Kelly, *et al.*, Physica C (Amsterdam) **235**–**240**, 1019 (1994).
36. R. O. Anderson, R. Claessen, J. W. Allen, *et al.*, Phys. Rev. Lett. **70**, 3163 (1990).
37. J. D. Perkins, R. J. Birgeneau, J. M. Graybeal, *et al.*, Phys. Rev. B **58**, 9390 (1998).
38. K. Tamasaku and S. Uchida, Physica C (Amsterdam) **235**–**240**, 1321 (1994).
39. E. Lai and R. J. Gooding, Phys. Rev. B **57**, 1498 (1998).
40. O. E. Parfenov, A. A. Nikonov, and S. N. Barilo, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 719 (2002) [JETP Lett. **76**, 616 (2002)].
41. B. Beschoten, S. Sadewasser, G. Guntherodt, and C. Quitmann, Phys. Rev. Lett. **77**, 1837 (1996).
42. J. Takeya, Y. Ando, S. Komiya, and X. P. Sun, Phys. Rev. Lett. **88**, 077001 (2002).
43. S. Uchida, T. Ido, H. Takagi, *et al.*, Phys. Rev. B **43**, 7942 (1991).
44. M. V. Sadovskiĭ, Sverkhprovodimost: Fiz. Khim. Tekh. **3**, 337 (1995).
45. Z.-X. Shen and J. R. Schrieffer, Phys. Rev. Lett. **78**, 1771 (1997).
46. A. Ino, C. Kim, M. Nakamura, *et al.*, Phys. Rev. B **65**, 94504 (2002).
47. H. Ding, J. R. Engelbrecht, Z. Wang, *et al.*, Phys. Rev. Lett. **87**, 227001 (2001).
48. D. Manske, I. Eremin, and K. H. Bennemann, Phys. Rev. Lett. **87**, 177005 (2001).
49. D. N. Basov, S. I. Woods, A. S. Katz, *et al.*, Science **283**, 49 (1999).
50. H. J. A. Molegraaf, C. Presura, D. van der Marel, *et al.*, Science **295**, 2239 (2002).
51. A. F. Santander-Syro, R. P. S. M. Lobo, N. Bontemps, *et al.*, cond-mat/0111539.
52. J. E. Hirsh and F. Marsiglio, Phys. Rev. B **62**, 15131 (2000).
53. A. S. Katz, S. I. Woods, E. J. Singley, *et al.*, Phys. Rev. B **61**, 5930 (2000).
54. M. Tinkham, *Introduction to Superconductivity*, 2nd ed. (McGraw-Hill, New York, 1996; Atomizdat, Moscow, 1980).
55. A. E. Karakozov, E. G. Maksimov, and O. V. Dolgov, cond-mat/0208170.
56. N. F. Mott and E. A. Davis, *Electronic Processes in Non-Crystalline Materials*, 2nd ed. (Clarendon Press, Oxford, 1979; Mir, Moscow, 1982).
57. J. Orenstein and A. J. Millis, Science **288**, 468 (2000).
58. J. Tranquada, B. J. Sternlieb, J. D. Axe, *et al.*, Nature **375**, 561 (2000).
59. J. Zaanen, Science **286**, 251 (1999).
60. A. A. Abrikosov, Phys. Rev. B **57**, 8656 (1998).
61. A. A. Abrikosov, Phys. Rev. B **62**, 15156 (2000).
62. A. Ino, T. Mizokawa, A. Fujimori, *et al.*, Phys. Rev. Lett. **79**, 2101 (1997).
63. N. Harima, J. Matsuno, A. Fujimori, *et al.*, Phys. Rev. B **64**, 220507(R) (2001).
64. A. Ino, T. Mizokawa, K. Kobayashi, *et al.*, Phys. Rev. Lett. **81**, 2124 (1998).
65. G. A. Thomas, D. H. Rapkine, S. L. Cooper, *et al.*, Phys. Rev. B **45**, 2474 (1992).
66. C. Chaillout, I. Chenavas, S.-W. Cheong, *et al.*, Physica C **170**, 87 (1990).
67. P. C. Hammel and D. I. Scalapino, Philos. Mag. B **74**, 523 (1996).
68. A. I. Agafonov and É. A. Manykin, Zh. Éksp. Teor. Fiz. **114**, 1765 (1998) [JETP **87**, 956 (1998)].
69. C. R. A. Catlow, M. S. Islam, and X. Zhang, J. Phys.: Condens. Matter **10**, L49 (1998).
70. A. I. Agafonov and É. A. Manykin, Pis'ma Zh. Éksp. Teor. Fiz. **65**, 419 (1997) [JETP Lett. **65**, 439 (1997)].
71. P. P. Edwards, N. F. Mott, and A. S. Alexandrov, J. Supercond. **11**, 151 (1998).
72. P. C. Hammel, B. W. Statt, R. L. Martin, *et al.*, Phys. Rev. B **57**, R712 (1998).
73. J. K. Perry, J. Tahir-Kheli, and W. A. Goddard, Phys. Rev. B **65**, 144501 (2002).
74. P. W. Anderson, Science **288**, 480 (2000).
75. S. Ciuchi, F. de Pasquale, S. Fratini, and D. Feinberg, Phys. Rev. B **56**, 4494 (1997).
76. E. V. L. de Mello and J. Ranninger, Phys. Rev. B **55**, 14872 (1997); Phys. Rev. B **58**, 9098 (1998).

77. K. A. Muller, J. Supercond. **12**, 3 (1999).
78. P. B. Allen and V. Perebeinos, Phys. Rev. B **60**, 10747 (1999).
79. B. K. Chakraverty, J. Ranninger, and D. Feinberg, Phys. Rev. Lett**. 82**, 2621 (1999).
80. A. S. Alexandrov, Phys. Rev. B **61**, 12315 (2000).
81. M. Zoli, Phys. Rev. B **61**, 14523 (2000).
82. A. I. Agafonov and É. A. Manykin, Physica B (Amsterdam) **259**–**261**, 458 (1999).
83. A. I. Agafonov and É. A. Manykin, Zh. Éksp. Teor. Fiz. **119**, 763 (2001) [JETP **92**, 665 (2001)].
84. A. I. Agafonov and É. A. Manykin, Phys. Rev. B **52**, 14571 (1995).
85. A. I. Agafonov and É. A. Manykin, Zh. Éksp. Teor. Fiz. **117**, 182 (2000) [JETP **90**, 160 (2000)].
86. F. D. M. Haldane and P. W. Anderson, Phys. Rev. B **13**, 2553 (1976).
87. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics* (Fizmatgiz, Moscow, 1962; Prentice Hall, Englewood Cliffs, N.J., 1963).
88. F. Yonezawa and T. Matsubara, Progr. Theor. Phys. **35**, 357 (1966).
89. A. I. Agafonov and É. A. Manykin, Zh. Éksp. Teor. Fiz. **109**, 1405 (1996) [JETP **82**, 758 (1996)].
90. K. Morikawa, T. Mizokawa, A. Fujimori, *et al.*, Phys. Rev. B **54**, 8446 (1996).
91. Y. Okada, T. Amira, Y. Tokura, *et al.*, Phys. Rev. B **48**, 9677 (1993).
92. Y. Taguchi, Y. Tokura, T. Arima, and F. Inaba, Phys. Rev. B **48**, 511 (1993).
93. T. Katsufuji, Y. Okimoto, and Y. Tokura, Phys. Rev. Lett. **75**, 3497 (1995).
94. F. Inaba, T. Arima, T. Ishihawa, *et al.*, Phys. Rev. B **52**, R2221 (1995).
95. S. W. Robey, L. T. Hudson, C. Eylem, and B. Eichorn, Phys. Rev. B **48**, 562 (1993).
96. A. Fujimori, I. Hase, N. Nakamura, *et al.*, Phys. Rev. B **46**, 9841 (1992).
97. C. Y. Chen, E. C. Branlund, Chin Sung Bae, *et al*., Phys. Rev. B **51**, 3671 (1995).
98. N. F. Mott, *Metal–Insulator Transitions*, 2nd ed. (Taylor and Francis, London, 1990; Nauka, Moscow, 1979).
99. P. P. Edwards, T. V. Ramakrishnan, and C. N. R. Rao, J. Phys. Chem. **99**, 5228 (1995).
100. C. Y. Chen, R. J. Birgeneau, M. A. Kastner, *et al.*, Phys. Rev. B **43**, 392 (1991).
101. G. A. Thomas, in *High Temperature Superconductivity*, Ed. by D. P. Tunstall and W. Barford (Adam Hilger, Bristol, 1991), p. 169.
102. M. V. Nevitt, G. W. Grabtree, and T. E. Klippert, Phys. Rev. B **36**, 2398 (1987).
103. A. Junod, A. Bezinse, and J. Muller, Physica C (Amsterdam) **152**, 50 (1988).
104. S. E. Inderhees, M. B. Salamon, T. A. Friedman, and D. M. Ginsberg, Phys. Rev. B **36**, 2401 (1987).
105. M. B. Salamon, S. E. Inderhees, J. P. Rice, *et al.*, Phys. Rev. B **38**, 885 (1988).
106. J. D. Lee and A. Fujimori, Phys. Rev. Lett. **87**, 167008 (2001).
107. A. Lanzara, P. V. Bogdanov, X. J. Zhou, *et al.*, Nature **412**, 510 (2001).
108. M. R. Norman, H. Ding, J. C. Campuzano, *et al.*, Phys. Rev. Lett. **79**, 3506 (1997).
109. J. C. Campuzano, H. Ding, M. R. Norman, *et al.*, Phys. Rev. Lett. **83**, 3709 (1999).
110. B. Batlogg, R. J. Cava, A. Jayaraman, *et al.*, Phys. Rev. Lett. **58**, 2333 (1987).
111. D. E. Morris, R. M. Kuroda, A. G. Markelz, *et al.*, Phys. Rev. B **37**, 5936 (1988).
112. J. P. Franck, in *Physical Properties of High Temperature Superconductors III*, Ed. by D. M. Ginsberg (World Sci., Singapore, 1994), p. 189.
113. M. K. Crawford, M. N. Kunchur, W. E. Farneth, *et al.*, Phys. Rev. B **41**, 282 (1990).
114. M. K. Crawford, W. E. Farneth, E. M. McCarron III, *et al.*, Science **250**, 1390 (1990).
115. H. J. Bornemann, D. E. Morris, H. B. Liu, and P. K. Narwankar, Physica C (Amsterdam) **191**, 211 (1992).
116. J. P. Franck, S. Harker, and J. H. Brewer, Phys. Rev. Lett. **71**, 283 (1993).
117. G. M. Zhao, K. Conder, H. Keller, and K. A. Muller, J. Phys.: Condens. Matter **10**, 9055 (1998).
118. D. Zech, H. Keller, K. Conder, *et al.*, Nature **371**, 681 (1994).
119. J. P. Franck, J. Jung, and M. A.-K. Mohamed, Phys. Rev. B **44**, 5318 (1991).
120. G. M. Zhao, K. Conder, H. Keller, *et al.*, J. Phys.: Condens. Matter **10**, 9055 (1998).
121. A. Fukuoka, A. Tokiwa-Yamamoto, M. Iton, *et al.*, Physica C (Amsterdam) **265**, 13 (1996).
122. J. Hofer, R. Molinski, H. Keller, *et al.*, Physica C (Amsterdam) **297**, 103 (1998).
123. W. A. Groen, D. M. de Leeuw, L. F. Feiner, *et al.*, Physica C (Amsterdam) **165**, 55 (1990).
124. T. R. Chien, W. R. Datars, M. D. Lan, *et al.*, Phys. Rev. B **49**, 1342 (1994).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 97, No. 2, 2003, pp. 402–431.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 124, No. 2, 2003, pp. 445–476.

REVIEWS

X-ray Optical Activity: Applications of Sum Rules¶

J. Goulon[a], A. Rogalev[a, *], F. Wilhelm[a], C. Goulon-Ginet[a], P. Carra[a], I. Marri[a], and Ch. Brouder[b]

[a]*European Synchrotron Radiation Facility 38043, Grenoble Cedex, France*
**e-mail: rogalev@esrf.fr*
[b]*Laboratoire de Minéralogie-Cristallographie UMR-CNRS 7590, Universite Paris-VI, 1PGP F-75252, Paris, Cedex 05, France*

Submitted December 20, 2002

Abstract—Edge-selective sum rules are proposed for a variety of X-ray dichroisms related to natural or nonreciprocal optical activity. Four spherical operators are identified that mix orbitals of different parities in what is assumed to be the ground state. The orbital anapole moment $\mathbf{\Omega}^{(1)}$ is primarily responsible for the magnetochiral dichroism; the time-even rank-2 tensor $\mathbf{N}^{(2)} = [\mathbf{L}, \mathbf{\Omega}]^{(2)}$ for natural circular dichroism; the time-odd rank-2 tensor $\mathbf{W}^{(2)} = [\mathbf{L}, \mathbf{n}]^{(2)}$ for nonreciprocal magnetic linear dichroisms. At higher orders, the time-odd rank-3 tensor $\mathbf{\Gamma}^{(3)} = [\mathbf{L}, \mathbf{L}, \mathbf{\Omega}]^{(3)}$ can also contribute to all nonreciprocal dichroisms. The physical content of these operators is analyzed. For every magnetoelectric group, one can predict which dichroic effect can be measured with either a single crystal or a powdered sample. Experimental spectra are produced to illustrate the value of the sum rules and the practical conditions of their application. Regarding nonreciprocal activity, one should be cautious about discussing magnetic symmetry because the deep core hole can couple the true ground state with low-lying excited states. *© 2003 MAIK "Nauka/Interperiodica".*

CONTENTS

1. Introduction 402
2. Edge-Selective *E*1*E*2 Sum Rules 403
2.1. Parity-Mixing Operators 403
2.2. Spherical Polarization Tensors 404
2.3. Symmetry Groups in XOA 405
2.3.1. Transformation Properties 405
2.3.2. Symmetry Groups and *E*1*E*2 Absorption 406
2.4. Effective Operators of XOA 407
2.4.1 Spherical Basis 407
2.4.2. Cartesian Basis 408
3. Physical Implications 409
3.1. Orbital Magnetoelectric Operators 409
3.1.1. Spin and Orbital Anapoles 409
3.1.2. Operators Conserved by $I\Theta$ 410
3.1.3. Operators Not Conserved by $I\Theta$ 412
3.2. Natural X-ray Optical Activity 413
3.3. Rotational Isotropy 414
3.3.1. XNCD Spectra 414
3.3.2. Nonreciprocal XMχD and XMLD Spectra 416
4. Applications of X-ray Optical Activity 419
4.1. Magnetoelectric Symmetry 419
4.1.1. Chromium Sesquioxide Cr_2O_3 419
4.1.2. Vanadium Sesquioxides $(V_{1-x}Cr_x)_2O_3$ 422
4.2. Effective Operators and Cross Densities of States 425
4.2.1. Applications of the XOA Sum Rules 425
4.2.2. Cross Densities of States 426
5. Conclusions 428
Appendix 428
References 429

1. INTRODUCTION

Systems with broken inversion symmetry play a fascinating role not only in physics but also in chemistry and in life sciences, where molecular recognition processes are very often controlled by chirality. In 1958, Zel'dovich [1] introduced the concept of the anapole to describe parity-violating interactions. For nearly 40 years, atomic and nuclear physicists were angling for nuclear anapoles [2, 3] until the anapole moment of ^{133}Cs was finally measured in 1997 [4]. In solid-state physics, the concept of the anapole has attracted much less attention even though there is a long established literature dealing with toroidal multiple moments [5–8]. It was realized recently that X-ray optical activity (XOA) can offer unique experimental access to orbital anapole moments and to a whole family of related operators. It is the aim of the present paper to analyze the physical content of these operators, especially for magnetoelectric solids in which parity (I) and time-reversal (Θ) symmetries are broken, while the structure remains invariant under the product $I\Theta$ [9].

¶This article was submitted by the authors in English.

Unlike magnetooptical effects such as Faraday rotation or magnetic circular dichroism, which refer primarily to electric dipole ($E1E1$) transitions, optical activity is associated with transition probabilities that mix multipole moments of opposite parities (e.g., $E1M1$ or $E1E2$). The Curie principle thus states that optical activity can be observed only in parity nonconserving systems. We recall that properties related to optical activity can be either even ("natural") or odd ("nonreciprocal") with respect to the time-reversal operator Θ. We have discussed elsewhere [10] how to transpose the theories of optical activity currently used at optical wavelengths into the X-ray spectral range. Following Buckingham [11] and Barron [12], we found it most convenient to describe XOA by introducing a complex gyration tensor

$$\zeta_{\alpha\beta\gamma} = \zeta'_{\alpha\beta\gamma} - i\zeta''_{\alpha\beta\gamma}.$$

In core level spectroscopes, magnetic dipole transitions ($M1$) are very weak [13], and it seems perfectly legitimate to neglect the $E1M1$ terms. Under such conditions, this Cartesian gyration tensor is dominated by the electric dipole ($E1_\alpha$)–electric quadrupole ($E2_{\beta\gamma}$) interference terms,

$$\begin{aligned}\zeta'_{\alpha\beta\gamma} &= \mathrm{Im}\{E1_\alpha E2_{\beta\gamma}\} + \mathrm{Im}\{E1_\beta E2_{\alpha\gamma}\},\\ \zeta''_{\alpha\beta\gamma} &= \mathrm{Re}\{E1_\alpha E2_{\beta\gamma}\} - \mathrm{Re}\{E1_\beta E2_{\alpha\gamma}\}.\end{aligned} \quad (1)$$

The imaginary part (ζ'') is antisymmetric with respect to interchange of the α, β subscripts and is responsible for natural XOA; the real part (ζ') is symmetric and contributes to nonreciprocal effects [12].

Every Stokes component S_j is associated with a well-identified dichroism related to XOA [14–19]:

1. The X-ray magnetochiral dichroism (XMχD),

$$\mathrm{XM}\chi\mathrm{D}(S_0) \propto [\zeta'_{\beta\beta\gamma} + \zeta'_{\alpha\alpha\gamma}].$$

2. The nonreciprocal X-ray magnetic linear dichroism (XMLD),

$$\mathrm{XMLD}(S_1) \propto [\zeta'_{\beta\beta\gamma} - \zeta'_{\alpha\alpha\gamma}].$$

3. The nonreciprocal, Jones X-ray magnetic linear dichroism (XMLD),

$$\mathrm{XMLD}(S_2) \propto 2\zeta'_{\alpha\beta\gamma}.$$

4. The X-ray natural circular dichroism (XNCD),

$$\mathrm{XNCD}(S_3) \propto 2\zeta''_{\alpha\beta\gamma}.$$

In several cases, XNCD spectra were successfully reproduced using *ab initio* calculations in the general framework of the multiple scattered wave theory [16, 20]. To the best of our knowledge, however, no suitable code for simulation of the nonreciprocal XOA is presently available. This is why in this paper we focus on the exploitation of edge-selective sum rules, which may give access to the expectation values of a series of effective operators that mix orbitals of opposite parities in what is assumed to be the true multielectronic ground state $|\psi_g\rangle$. In the next section, we recast the XOA sum rules in their general framework. In Section 3, we analyze the physical content of the four effective operators that were identified as responsible for XOA. An important result in this section is the possibility of predicting which dichroism can be observed experimentally for a given magnetic class. In Section 4, referring to several specific examples, we develop some considerations on what can be learned regarding the magnetoelectric symmetry and how the XOA operators can be accessed in practice.

Throughout this paper, we retain the same terminology (i.e., scalar, vector, deviator, septor, …) for the decomposition of Cartesian or spherical tensors into their irreducible representations [21, 22]. We systematically use normal fonts for Cartesian tensors and boldface fonts for spherical tensors. Irreducible tensors of an even rank that have odd parity or irreducible tensors of an odd rank with even parity are commonly called pseudotensors. For clarity, we prefer to call irreducible tensors of an odd rank and odd parity polar tensors rather than true tensors. Polar vectors that are odd under time reversal are called toroidal for reasons explained in Section 3; pseudovectors that are even under time reversal are called antitoroidal by analogy.

2. EDGE-SELECTIVE $E1E2$ SUM RULES

2.1. Parity-Mixing Operators

Optical sum rules are commonly used in atomic physics [23]. In 1992, Thole *et al.* [24] established a useful sum rule for X-ray magnetic circular dichroism (XMCD): it states that the integrated dichroic signal is proportional to $\langle L_z\rangle$, i.e., to the ground state expectation value of the angular momentum operator acting on the electronic shell that accepts the excited photoelectron. For XMCD spectra, one is mostly concerned with electric dipole (El) transitions satisfying the selection $l = l_c \pm 1$, where l_c characterizes the angular momentum in the initial core state and l is the angular momentum in the final excited state. In 1998, using a similar approach, Natoli *et al.* [20] already established the sum rule

$$\Sigma_{E1E2} = \int_{\Delta E} \frac{\sigma_{E1E2}(E)}{E^2} dE \propto \left\langle \psi_g \middle| \mathbf{N}^{(2)}(l, l') \middle| \psi_g \right\rangle \quad (2)$$

for X-ray natural circular dichroism (XNCD), where σ_{E1E2} denotes the X-ray absorption cross section due to the $E1E2$ interference terms in a finite energy range ΔE that must include, whenever this is relevant, the two partners (j_+, j_-) of the spin-orbit split edge. In the right-

hand side, l still refers to the final angular momentum of the electric dipole ($E1$) transition and l' refers to the electric quadrupole transition ($E2$) satisfying the selection rule $l' = l_c \pm 0.2$, excluding $l' = l_c = 0$. Obviously, l and l' have opposite parities and the operator $\mathbf{N}^{(2)}(l, l')$ probes the mixing of atomic orbitals of the corresponding parities. The problem with Eq. (2) was that the rank-2 spherical tensor $\mathbf{N}^{(2)}$ was given no clear physical meaning in [20]. We also found it desirable to extend this sum rule to all $E1E2$ dichroisms.

Regarding Eq. (2), there is still a serious limitation that was underlined by Di Matteo and Natoli in the comprehensive review article [25]. Due to the core hole perturbation, $|\psi_g\rangle$ is merely a virtual or pseudo ground state of the system. Intuitively, one may fear that the core hole does affect orbitals of opposite parities differently. If we expand $|\psi_g\rangle$ in terms of stationary states ($|\Psi_n\rangle$), the quantity obtained is actually given by

$$\langle \psi_g | \mathbf{O}^{(q)} | \psi_g \rangle = \sum_{n, n'} \alpha_n^* \alpha_{n'} \langle \Psi_n | \mathbf{O}^{(q)} | \Psi_{n'} \rangle, \qquad (3)$$

where $\mathbf{O}^{(q)}$ is the pertinent parity-mixing effective operator. At present, there is no proof that the sum over all configurations cancels the effects of the core hole and restores the property of a true ground state, as this is implicitly assumed for the popular XMCD sum rules. Contrary to [25], we are not even certain that the cross terms ($n \neq n'$) can be neglected a priori: typically, in the case of XMCD where $\mathbf{O}^{(1)} = L_z$, the matrix elements in the right-hand side of Eq. (3) are precisely those that contribute to the Van Vleck paramagnetism and are usually nonzero. We consider this difficulty again in Section 4.

2.2. Spherical Polarization Tensors

The electric dipole ($E1$) and electric quadrupole ($E2$) transition operators are $\hat{\boldsymbol{\epsilon}} \cdot \mathbf{r}$ and $\hat{\boldsymbol{\epsilon}} \cdot \mathbf{r}\mathbf{k} \cdot \mathbf{r}$, respectively. We recall that for the left circularly polarized light with a wave vector $\mathbf{k}$, $\hat{\boldsymbol{\epsilon}} = (\mathbf{i} - i\mathbf{j})/\sqrt{2}$, where $\mathbf{i}$ and $\mathbf{j}$ are unit vectors such that

$$\mathbf{i} \times \mathbf{j} = \mathbf{k}/k = \hat{k}.$$

It is therefore natural to describe the angular dependence of the interference between the $E1$ and $E2$ transitions by coupling $\hat{\boldsymbol{\epsilon}}$ first with $\hat{k}$ (as spherical tensors) and then with $\hat{\boldsymbol{\epsilon}}^*$ to obtain

$$\mathbf{T}_\beta^{(b)} = [\hat{\boldsymbol{\epsilon}}^*, [\hat{\boldsymbol{\epsilon}}, \hat{k}]^{(2)}]_\beta^{(b)}.$$

The coupling of spherical tensors is described in standard textbooks (e.g., [26]). However, the tensors $\mathbf{T}_\beta^{(b)}$ do not have a well-defined behavior under time reversal and are destined to decompose into their time-reversal even ($\theta = 1$) and time-reversal odd ($\theta = -1$) parts $\mathbf{T}_\beta^{(b, \theta)}$. The $E1E2$ absorption cross section (σ_{E1E2}) and sum rules (Σ_{E1E2}) are therefore written as

$$\sigma_{E1E2} = \sum_{b=1}^{3} \sum_{\beta=-b}^{b} \sum_{\theta=\pm 1} (-1)^\beta \mathbf{T}_\beta^{(b, \theta)} \boldsymbol{\sigma}_{-\beta}^{(b, \theta)},$$

$$\Sigma_{E1E2} = \sum_{b=1}^{3} \sum_{\beta=-b}^{b} \sum_{\theta=\pm 1} (-1)^\beta \mathbf{T}_\beta^{(b, \theta)} \boldsymbol{\Sigma}_{-\beta}^{(b, \theta)},$$

where $\boldsymbol{\sigma}_{-\beta}^{(b, \theta)}$ and $\boldsymbol{\Sigma}_{-\beta}^{(b, \theta)}$ are rank-b spherical tensors.

To investigate the time-reversal symmetry of $\mathbf{T}_\beta^{(b, \theta)}$, we write it in terms of

$$\mathbf{X}_\beta^{(a, b)} = [[\hat{\boldsymbol{\epsilon}}^*, \hat{\boldsymbol{\epsilon}}]^{(a)}, \hat{k}]_\beta^{(b)}$$

(see [27] and Table 1). Here, $[\hat{\boldsymbol{\epsilon}}^*, \hat{\boldsymbol{\epsilon}}]^{(a)}$ is a rank-a spherical tensor. As proved in the next section, the time-reversal properties of $\mathbf{X}_\beta^{(a, b)}$ can readily be deduced from the fact that the action of the time-reversal operator Θ on $\hat{\boldsymbol{\epsilon}}$ and $\mathbf{k}$ is $\Theta\hat{\boldsymbol{\epsilon}} = \hat{\boldsymbol{\epsilon}}^*$ and $\Theta\mathbf{k} = -\mathbf{k}$. The action of Θ on $\mathbf{X}_\beta^{(a, b)}$ is therefore given by

$$\Theta \mathbf{X}_\beta^{(a, b)} = (-1)^{a+1} \mathbf{X}_\beta^{(a, b)}.$$

We note that complex conjugation has a different action,

$$\mathbf{X}_\beta^{(a, b)*} = (-1)^{a+1+b-\beta} \mathbf{X}_{-\beta}^{(a, b)}.$$

We now consider the possible values of a and b satisfying the triangle conditions $0 \leq a \leq 2$ and $|a - 1| \leq b \leq a + 1$.

1. For $a = 0$ i.e., $\hat{\boldsymbol{\epsilon}}^* \cdot \hat{\boldsymbol{\epsilon}} = 1$, it immediately follows that $b = 1$ and $\sigma_{E1E2} \propto \hat{k}\,\sigma^{(1, -1)}$ or $\Sigma_{E1E2} \propto \hat{k}\,\Sigma^{(1, -1)}$. This is obviously the case of XMχD.

2. For $a = 1$, i.e., $[\hat{\boldsymbol{\epsilon}}^*, \hat{\boldsymbol{\epsilon}}]^{(1)} \propto i\hat{k}$, the result is still rather simple if we assume that the electromagnetic wave remains transverse inside the sample, i.e., if the condition $\hat{\boldsymbol{\epsilon}} \cdot \mathbf{k} = 0$ is satisfied; then, the only choice for b is 2. This is typically the case of XNCD, and it was previously established that the spherical tensors $\boldsymbol{\sigma}^{(2, +1)}$ and $\boldsymbol{\Sigma}^{(2, +1)}$ are rank-2 pseudodeviators [20].

3. Finally, if $a = 2$, the problem becomes more complicated because values of $b = 1, 2, 3$ are possible, which implies that the tensor property $\sigma^{(b, -1)}$ can be a

Table 1. Polarization tensors of XOA

$\mathbf{T}^{(1,+1)} \equiv 0$ (transversality condition)		
$a = 0$	$\mathbf{T}_0^{(1,-1)}(\hat{k}) = -\frac{1}{2}\sqrt{\frac{3}{5}}\hat{k}$	XMχD (S_0)
$a = 2$	$\mathbf{T}_0^{(2,+1)}(\hat{\epsilon}, \hat{k}) = \frac{\sqrt{3}}{2}[[\hat{\epsilon}^*, \hat{\epsilon}]^{(1)}, \hat{k}]_0^{(2)}$	XNCD (S_3)
$a = 2$	$\mathbf{T}_{\pm 2}^{(2,-1)}(\hat{\epsilon}, \hat{k}) = \frac{1}{2}[[\hat{\epsilon}^*, \hat{\epsilon}]^{(2)}, \hat{k}]_{\pm 2}^{(2)}$	XMLD (S_1, S_2)
$\mathbf{T}^{(3,+1)} \equiv 0$		
$a = 2$	$\mathbf{T}_0^{(3,-1)}(\hat{\epsilon}, \hat{k}) = [[\hat{\epsilon}^*, \hat{\epsilon}]^{(2)}, \hat{k}]_0^{(3)}$	XMχD (S_0)
$a = 2$	$\mathbf{T}_{\pm 2}^{(3,-1)}(\hat{\epsilon}, \hat{k}) = [[\hat{\epsilon}^*, \hat{\epsilon}]^{(2)}, \hat{k}]_{\pm 2}^{(3)}$	XMLD (S_1, S_2)

vector, a deviator, or a septor. The option $\{a = 2; b = 1\}$ again yields the same vector contribution to XMχD; the options $\{a = 2; b = 2, 3\}$ can be shown to contribute to nonreciprocal XMLD.

This discussion and the relation between $\mathbf{T}_\beta^{(b,\theta)}$ and $\mathbf{X}_\beta^{(a,b)}$ show that the nonzero tensors are $\mathbf{T}^{(1,-1)}$, $\mathbf{T}^{(2,+1)}$, $\mathbf{T}^{(2,-1)}$, and $\mathbf{T}^{(3,-1)}$. These tensors transform as

$$\Theta\mathbf{T}_\beta^{(b,\theta)} = \theta\mathbf{T}_\beta^{(b,\theta)}$$

under time-reversal symmetry and as

$$\mathbf{T}_\beta^{(b,\theta)*} = \theta(-1)^{b-\beta}\mathbf{T}_{-\beta}^{(b,\theta)}$$

under complex conjugation. We note that all tensors $\boldsymbol{\sigma}^{(b,\theta)}$ and $\boldsymbol{\Sigma}^{(b,\theta)}$ are time-reversal odd with the unique exception of XNCD $\{a = 1; b = 2\}$.

At this stage, within the limits of validity of Eqs. (2), several important results already follow without heavy calculations: because the effective vector operator of XMχD is odd under I and Θ, it can only be a toroidal vector; the effective operator of XNCD must be a time-even pseudodeviator; the effective operators of XMLD (S_1, S_2) must combine a pseudodeviator and a polar septor, which must again be odd with respect to both I and Θ. These results are summarized in Table 1.

2.3. Symmetry Groups in XOA

In magnetic samples, the time-reversal operator 0 plays a key role and the point and space groups have to be replaced by magnetic point and space groups [29]. The representation theory of magnetic groups is difficult because Θ is antilinear and representations are replaced by corepresentations [30]. Theorems involving characters are no longer valid for corepresentations because the equivalence between corepresentations D and D' is not determined by the existence of a matrix A such that $D' = ADA^{-1}$ [30]. Nevertheless, we show that these complications can be circumvented for the representations of symmetries involved in XOA.

2.3.1. Transformation properties. The X-ray absorption cross section σ including electric dipole and quadrupole transitions is proportional to

$$\sigma(\hat{\epsilon}, \mathbf{k}) \propto \sum_f \langle\psi_g|\hat{\epsilon}^* \cdot \mathbf{r} - \frac{i}{2}\hat{\epsilon}^* \cdot \mathbf{r}\mathbf{k} \cdot \mathbf{r}|\psi_f\rangle$$
$$\times \langle\psi_f|\hat{\epsilon} \cdot \mathbf{r} + \frac{i}{2}\hat{\epsilon} \cdot \mathbf{r}\mathbf{k} \cdot \mathbf{r}|\psi_g\rangle\delta(E_f - E_g - \hbar\omega).$$

We now successively transform a physical state with the parity I, time-reversal Θ, rotation R, and translation $T_\mathbf{R}$ operations and consider how $\sigma(\hat{\epsilon}, \mathbf{k})$ is modified.

To investigate the transformation of the absorption cross section under parity, we first consider the one-electron spineless case. The action of the parity operator (I) on the system transforms the wavefunctions as $(I\psi_f)(\mathbf{r}) = \psi_f(-\mathbf{r})$ and $(I\psi_g)(\mathbf{r}) = \psi_g(-\mathbf{r})$, and the matrix elements become

$$\langle I\psi_f|\hat{\epsilon} \cdot \mathbf{r}|I\psi_g\rangle = \int d\mathbf{r}\psi_f^*(-\mathbf{r})\hat{\epsilon} \cdot \mathbf{r}\psi_g(-\mathbf{r})$$
$$= \int d\mathbf{r}'\psi_f^*(\mathbf{r}')\hat{\epsilon} \cdot (-\mathbf{r}')\psi_g(\mathbf{r}') = \langle\psi_f|(-\hat{\epsilon}) \cdot \mathbf{r}|\psi_g\rangle.$$

The same result holds generally for a many-body system with spin. Moreover,

$$\langle I\psi_f|\hat{\epsilon} \cdot \mathbf{r} + \frac{i}{2}\hat{\epsilon} \cdot \mathbf{r}\mathbf{k} \cdot \mathbf{r}|I\psi_g\rangle$$
$$= \langle\psi_f|I\left(\hat{\epsilon} \cdot \mathbf{r} + \frac{i}{2}\hat{\epsilon} \cdot \mathbf{r}\mathbf{k} \cdot \mathbf{r}\right)I|\psi_g\rangle$$
$$= \langle\psi_f|(-\hat{\epsilon}) \cdot \mathbf{r} + \frac{i}{2}(-\hat{\epsilon}) \cdot \mathbf{r}(-\mathbf{k}) \cdot \mathbf{r}|\psi_g\rangle.$$

Therefore, if $\sigma(\hat{\epsilon}, \mathbf{k}; I)$ denotes the absorption cross section of the system transformed by parity, and if parity is a symmetry of the system (such that the energies of $I\psi_f$ and $I\psi_g$ coincide with the respective energies of ψ_f and ψ_g), we obtain $\sigma(\hat{\epsilon}, \mathbf{k}; I) = \sigma(-\hat{\epsilon}, -\mathbf{k})$.

For time-reversal symmetry (Θ), we start from the basic equation

$$\langle \Theta\phi | \Theta\psi \rangle = \langle \phi | \psi \rangle^* = \langle \psi | \phi \rangle$$

(see [30, 31]). Hence,

$$\left\langle \Theta\psi_g \middle| \Theta\left(\hat{\epsilon}\cdot\mathbf{r} + \frac{i}{2}\hat{\epsilon}\cdot\mathbf{r}\mathbf{k}\cdot\mathbf{r}\right)\middle|\psi_f\right\rangle$$
$$= \langle \psi_g | \hat{\epsilon}\cdot\mathbf{r} + \frac{i}{2}\hat{\epsilon}\cdot\mathbf{r}\mathbf{k}\cdot\mathbf{r} | \psi_f \rangle^*$$
$$= \langle \psi_f | \hat{\epsilon}^*\cdot\mathbf{r} - \frac{i}{2}\hat{\epsilon}^*\cdot\mathbf{r}\mathbf{k}\cdot\mathbf{r} | \psi_g \rangle.$$

On the other hand, the antilinearity of the time-reversal operator yields

$$\left| \Theta\left(\hat{\epsilon}\cdot\mathbf{r} + \frac{i}{2}\hat{\epsilon}\cdot\mathbf{r}\mathbf{k}\cdot\mathbf{r}\right)\psi_f \right\rangle$$
$$= \left(\hat{\epsilon}^*\cdot\mathbf{r} - \frac{i}{2}\hat{\epsilon}^*\cdot\mathbf{r}\mathbf{k}\cdot\mathbf{r}\right)|\Theta\psi_f\rangle,$$

and therefore,

$$\langle \Theta\psi_g | \hat{\epsilon}^*\cdot\mathbf{r} - \frac{i}{2}\hat{\epsilon}^*\cdot\mathbf{r}\mathbf{k}\cdot\mathbf{r} | \Theta\psi_f \rangle$$
$$= \langle \psi_f | \hat{\epsilon}^*\cdot\mathbf{r} - \frac{i}{2}\hat{\epsilon}^*\cdot\mathbf{r}\mathbf{k}\cdot\mathbf{r} | \psi_g \rangle.$$

Similarly,

$$\langle \Theta\psi_f | \hat{\epsilon}\cdot\mathbf{r} + \frac{i}{2}\hat{\epsilon}\cdot\mathbf{r}\mathbf{k}\cdot\mathbf{r} | \Theta\psi_g \rangle$$
$$= \langle \psi_g | \hat{\epsilon}\cdot\mathbf{r} + \frac{i}{2}\hat{\epsilon}\cdot\mathbf{r}\mathbf{k}\cdot\mathbf{r} | \psi_f \rangle.$$

Finally, if $\sigma(\hat{\epsilon}, \mathbf{k}; \Theta)$ denotes the absorption cross section of the time-reversed system and if the system is invariant under Θ, we obtain $\sigma(\hat{\epsilon}, \mathbf{k}; \Theta) = \sigma(\hat{\epsilon}^*, -\mathbf{k})$.

Next we consider a transformation by the rotation R. Starting again with the one-electron spinless case, we have

$$\langle R\psi_f | \hat{\epsilon}\cdot\mathbf{r} | R\psi_g \rangle = \int d\mathbf{r}\psi_f^*(R\mathbf{r})\hat{\epsilon}\cdot\mathbf{r}\psi_g(R\mathbf{r})$$
$$= \int d\mathbf{r}'\psi_f^*(\mathbf{r}')\hat{\epsilon}\cdot(R^{-1}\mathbf{r}')\psi_g(\mathbf{r}') = \langle \psi_f | (R\hat{\epsilon})\cdot\mathbf{r} | \psi_g \rangle.$$

More generally, for a many-body system with spin, if $\sigma(\hat{\epsilon}, \mathbf{k}; R)$ denotes the absorption cross section of the system transformed by the rotation R, we find that

$$\sigma(\hat{\epsilon}, \mathbf{k}; R) = \sigma(R\hat{\epsilon}, R\mathbf{k}).$$

The last transformation that we need is translation. In X-ray absorption spectroscopy, the dipole and quadrupole approximations are valid because the core states are localized and the origin of coordinates can be taken at the absorbing atom. If the system is translated, the origin is no longer the absorbing atom, the dipole and quadrupole approximations are not valid, and we must use the full absorption cross section [32]

$$\sigma(\hat{\epsilon}, \mathbf{k}) = \frac{4\pi^2\hbar\alpha}{m^2\omega}\sum_f \left|\langle \psi_f | e^{i\mathbf{k}\cdot\mathbf{r}} X | \psi_g \rangle\right|^2 \delta(E_f - E_g - \hbar\omega),$$

where

$$X = \hbar\hat{\epsilon}\cdot\nabla - (g/2)\mathbf{s}\cdot(\mathbf{k}\times\hat{\epsilon})$$

and $\mathbf{s}$ is the spin operator. The operator X is not modified by translation. Thus, the translation $T_{\mathbf{R}}$ acting on the system by $T_{\mathbf{R}}\psi_f(\mathbf{r}) = \psi_f(\mathbf{r} + \mathbf{R})$ and $T_{\mathbf{R}}\psi_g(\mathbf{r}) = \psi_g(\mathbf{r} + \mathbf{R})$ transforms $\langle \psi_f | e^{i\mathbf{k}\cdot\mathbf{r}} X | \psi_g \rangle$ into

$$\langle T_{\mathbf{R}}\psi_f | e^{i\mathbf{k}\cdot\mathbf{r}} X | T_{\mathbf{R}}\psi_g \rangle = e^{-i\mathbf{k}\cdot\mathbf{R}} \langle \psi_f | e^{i\mathbf{k}\cdot\mathbf{r}} X | \psi_g \rangle.$$

Therefore, if $T_{\mathbf{R}}$ is a symmetry of the system, we obtain that $\sigma(\hat{\epsilon}, \mathbf{k}; T_{\mathbf{R}}) = \sigma(\hat{\epsilon}, \mathbf{k})$ and the absorption cross section is independent of translations of the system.

At this stage, we have shown that a transformation of the physical system can be replaced by a simultaneous transformation of the polarization and wave vectors. We analyze next the consequences of this result for the angular and polarization dependence of σ_{E1E2}.

2.3.2. Symmetry groups and *E*1*E*2 absorption. As discussed in Section 2.2, the *E*1*E*2 absorption cross section can be written as

$$\sigma_{E1E2} = \sum_{b=1}^{3}\sum_{\beta=-b}^{b}(-1)^{\beta}\sum_{\theta=\pm1}\mathbf{T}_{\beta}^{(b,\theta)}\sigma_{-\beta}^{(b,\theta)}. \qquad (4)$$

In a reference frame where the wave vector is directed along z axis, the polarization vector is

$$\hat{\epsilon} = \begin{pmatrix} \cos\psi\cos\chi + i\sin\psi\sin\chi \\ \sin\psi\cos\chi - i\cos\psi\sin\chi \\ 0 \end{pmatrix},$$

which represents an elliptically polarized wave for which the ellipse axes are at an angle of ψ with the reference frame axes, and the circular polarization rate is $\sin 2\chi$. We recall that

$$S_1/S_0 = \cos 2\chi \cos 2\psi, \quad S_2/S_0 = \cos 2\chi \sin 2\psi,$$

$$S_3/S_0 = \sin 2\chi.$$

In this frame, the nonzero tensor components are

$$\mathbf{T}_0^{(1,-1)} = -\frac{1}{2}\sqrt{\frac{3}{5}},$$

$$\mathbf{T}_0^{(2,+1)} = \frac{1}{2}\sin 2\chi,$$

$$T_{\pm 2}^{(2,-1)} = \pm\frac{1}{2\sqrt{6}} e^{\pm 2i\psi}\cos 2\chi,$$

$$\mathbf{T}_0^{(3,-1)} = -\frac{1}{\sqrt{10}},$$

$$\mathbf{T}_{\pm 2}^{(3,-1)} = \frac{1}{2\sqrt{3}} e^{\pm 2i\psi}\cos 2\chi.$$

In particular, $\mathbf{T}_{\pm 3}^{(3,-1)} = 0$.

We proved that the action of Θ on the system can be replaced by its action on $T_\beta^{(b,\theta)}$, which was found to be

$$\Theta \mathbf{T}_\beta^{(b,\theta)} = \theta \mathbf{T}_\beta^{(b,\theta)}.$$

This result is nontrivial because the action of the time-reversal operator Θ on a spherical tensor is usually described by

$$\mathbf{T}_m^{(j)} \longrightarrow (-1)^{j-m} T_{-m}^{(j)}$$

(see [30, 33]) or

$$\Theta \mathbf{T}_m^{(j)} = (-1)^m \mathbf{T}_{-m}^{(j)}$$

(see [31]). Here, the result is different because the time-reversal operator does not act directly on the spherical tensor. Its action on the system is translated into a simpler action on the polarization and wave vectors.

More generally, any symmetry operation S acting on the system can be written as

$$S = I^p \Theta^t R T_{\mathbf{R}},$$

where $p = 1$ or $p = 0$ if S contains or does not contain the inversion, $t = 1$ or $t = 0$ if S contains or does not contain the time-reversal symmetry, R denotes a rotation, and $T_{\mathbf{R}}$ denotes a translation. From the identity $\sigma(\epsilon, \mathbf{k}; I) = \sigma(-\epsilon, -\mathbf{k})$, we see that the action of the parity operator on the system reverses the $E1E2$ absorption cross section (i.e., $\sigma_{E1E2}(I) = -\sigma_{E1E2}$). Therefore, the action of a general symmetry operation S on the system transforms σ_{E1E2} into

$$\sigma_{E1E2}(S) = \sum_{b=1}^{3}\sum_{\beta=-b}^{b}(-1)^\beta \times \sum_{\theta=\pm 1}(-1)^p \theta^t \mathbf{D}^{(b)}_{\beta'\beta}(R)\mathbf{T}^{(b,\theta)}_{\beta'}\sigma^{(b,\theta)}_{-\beta}, \tag{5}$$

where $\mathbf{D}^{(b)}_{\beta'\beta}(R)$ is the Wigner rotation matrix. This result justifies the use of the character method, which was employed by Tenenbaum in [34] and which we use in Section 3.

For a magnetic group G_M containing g_m elements, the form of the absorption cross section is obtained by taking the average over the elements of the group,

$$\langle \sigma_{E1E2}\rangle = \sum_{b=1}^{3}\sum_{\beta=-b}^{b}(-1)^\beta \sum_{\theta=\pm 1}\langle \mathbf{T}^{(b,\theta)}_\beta\rangle \sigma^{(b,\theta)}_{-\beta}, \tag{6}$$

where

$$\langle \mathbf{T}^{(b,\theta)}_\beta\rangle = \frac{1}{g_m}\sum_{S\in G_M}(-1)^p \theta^t \mathbf{D}^{(b)}_{\beta'\beta}(R)\mathbf{T}^{(b,\theta)}_\beta.$$

2.4. Effective Operators of XOA

2.4.1. Spherical basis. The $E1E2$ sum rules were calculated by Carra and collaborators using the powerful method of group generators [27, 35–37]. A key achievement was to show that all operators $\mathbf{\Sigma}^{(b,\theta)}$ can be built from the triad of mutually orthogonal vector operators:

1. $\mathbf{n} = \mathbf{r}/r$, which is a time-reversal even polar vector typically associated with the electric dipole moment;

2. the orbital angular momentum $\mathbf{L}$, which is a time-reversal odd axial vector;

3. the toroidal vector $\mathbf{\Omega} = [(\mathbf{n}\times\mathbf{L}) - (\mathbf{L}\times\mathbf{n})]/2$, which is odd with respect to both I and Θ.

Because $\mathbf{\Omega}$ can be rewritten as the commutator, $\mathbf{\Omega} = i[\mathbf{n}, \mathbf{L}^2]/2$, we show in Section 3 that it is proportional to the orbital anapole moment defined in [2].

Important results have been established.

1. The XMχD sum rule involves the ground state expectation value of the toroidal vector $\mathbf{\Omega}^{(1,-1)}$ projected along the direction of the wave vector $\mathbf{k}$.

2. The XNCD sum rule must yield the expectation value of the Θ-even pseudodeviator $\mathbf{N}^{(2,+1)} = [\mathbf{L}, \mathbf{\Omega}]^{(2)}$, which is obtained for $a = 1$; $b = 2$.

3. For $a = b = 2$, the effective operator must be a Θ-odd pseudodeviator, which was identified with $\mathbf{W}^{(2,-1)} = [\mathbf{L}, \mathbf{n}]^{(2)}$. Its ground state expectation value appears in the nonreciprocal XMLD sum rule.

4. For $a = 2$ and $b = 3$, the effective operator is the Θ-odd septor $\mathbf{\Gamma}^{(3,-1)} = [[\mathbf{L}, \mathbf{L}]^{(2)}, \mathbf{\Omega}]^{(3)}$. Its ground state expectation value is involved in the XMχD sum rule and in the nonreciprocal XMLD sum rule.

As long as the definition of the polarization tensors $\mathbf{T}_\beta^{(b,\theta)}$ given in the previous section is applicable, we can use the following generic formulations of the XOA sum rules, to be called the Carra–Jerez–Marri equations hereafter [37]:

for XNCD (S_3),

$$\Sigma_{E1E2} = \frac{-8\pi^2\alpha}{3\hbar c}\sin 2\chi(2l_c+1)\sum_{l,l'} R_l^{(1)}R_{l'}^{(2)} \times a^{(2,+1)}(l_c, l, l')\sqrt{\frac{3}{2}}\langle \mathbf{N}_0^{(2,+1)}(l, l')\rangle, \quad (7)$$

for XMχD (S_0),

$$\Sigma_{E1E2} = \frac{-2\pi^2\alpha}{\hbar c}(2l_c+1)\sum_{l,l'} R_l^{(1)}R_{l'}^{(2)} \times\left\{\frac{2}{5}a^{(1,-1)}(l_c, l, l')\langle \mathbf{\Omega}_0^{(1,-1)}(l, l')\rangle - \frac{16}{\sqrt{10}}b^{(3,-1)}(l_c, l, l')\langle \Gamma_0^{(3,-1)}(l, l')\rangle\right\}, \quad (8)$$

and for XMLD (S_1, S_2),

$$\Sigma_{E1E2} = \frac{16\pi^2\alpha\cos 2\chi}{\hbar c}(2l_c+1)\sum_{l,l'} R_l^{(1)}R_{l'}^{(2)} \times\sum_{\beta=\pm 2}\left\{a^{(2,-1)}(l_c, l, l')\frac{ie^{i\beta\psi}}{3\beta}\langle \mathbf{W}_\beta^{(2,-1)}(l, l')\rangle + \frac{e^{i\beta\psi}}{\sqrt{3}}b^{(3,-1)}(l_c, l, l')\langle \Gamma_\beta^{(3,-1)}(l, l')\rangle\right\}. \quad (9)$$

In these equations, $R_l^{(1)}$ and $R_{l'}^{(2)}$ denote the radial dipole and quadrupole integrals that are classically defined as

$$R_l^{(1)} = \int_0^{\rho_{MT}} r^3 dr\,\phi_c(r)\varphi_l(r),$$

$$R_{l'}^{(2)} = \int_0^{\rho_{MT}} r^4 dr\,\phi_c(r)\varphi_{l'}(r),$$

where the core state and photoelectron radial wave functions are $\phi_c(r)$ and $\varphi_{l,l'}$, respectively; $\phi_c(r)$ is typically localized in a muffin-tin sphere of radius ρ_{MT}. The expressions for the numerical factors $a^{(2,+1)}$, $a^{(2,-1)}$, $a^{(1,-1)}$, $b^{(3,-1)}$ are given in Table 2.

2.4.2. Cartesian basis. For linear dichroism experiments, it is more appropriate to express Eq. (9) in terms of Hermitian Cartesian effective operators. This can easily be done using the relations (see, e.g., [22, 38])

$$\mathbf{W}_{\pm 2}^{(2,-1)} = \frac{1}{2}[W_{XX}^{(2,-1)} - W_{YY}^{(2,-1)}] \pm \frac{1}{2}[W_{XY}^{(2,-1)} + W_{YX}^{(2,-1)}],$$

$$\mathbf{\Gamma}_{\pm 2}^{(2,-1)} = \frac{\sqrt{3}}{2}[\Gamma_{XXZ}^{(3,-1)} - \Gamma_{YYZ}^{(3,-1)}] \pm \frac{\sqrt{3}}{2}[\Gamma_{XYZ}^{(3,-1)} + \Gamma_{YXZ}^{(3,-1)}],$$

where $\{X, Y, Z\}$ are Cartesian coordinates in the reference frame used to define the polarization tensors in Section 2.3.2. Hence, the two effective operators defined in Eq. (9) can now be rewritten as

$$\begin{aligned} &i[e^{2i\psi}\langle \mathbf{W}_{+2}^{(2,-1)}\rangle - e^{-2i\psi}\langle \mathbf{W}_{-2}^{(2,-1)}\rangle] \\ &= \sin 2\psi[\langle W_{YY}^{(2,-1)}\rangle - \langle W_{XX}^{(2,-1)}\rangle] \\ &- \cos 2\psi[\langle W_{XY}^{(2,-1)}\rangle + \langle W_{YX}^{(2,-1)}\rangle], \end{aligned} \quad (10)$$

$$\begin{aligned} &[e^{2i\psi}\langle \mathbf{\Gamma}_{+2}^{(3,-1)}\rangle - e^{-2i\psi}\langle \mathbf{\Gamma}_{-2}^{(3,-1)}\rangle] \\ &= \sin 2\psi[\langle \Gamma_{XYZ}^{(3,-1)}\rangle + \langle \Gamma_{YXZ}^{(2,-1)}\rangle] \\ &- \cos 2\psi[\langle \Gamma_{YYZ}^{(3,-1)}\rangle - \langle \Gamma_{XXZ}^{(3,-1)}\rangle]. \end{aligned} \quad (11)$$

Since

$$S_1/S_0 = \cos 2\chi\cos 2\psi, \quad S_2/S_0 = \cos 2\chi\sin 2\psi,$$

it becomes obvious that within the defined reference frame, $[W_{YY}^{(2,-1)} - W_{XX}^{(2,-1)}]$ is the effective operator responsible for the Jones dichroism XMLD (S_2) and $[\langle W_{XY}^{(2,-1)}\rangle + \langle W_{YX}^{(2,-1)}\rangle]$ is the effective operator of XMLD (S_1). It can be seen that the septor terms $[\langle \Gamma_{XYZ}^{(3,-1)}\rangle + \langle \Gamma_{YXZ}^{(3,-1)}\rangle]$ and $[\langle \Gamma_{YYZ}^{(3,-1)}\rangle - \langle \Gamma_{XXZ}^{(3,-1)}\rangle]$ also contribute to XMLD ($S2$) and XMLD (S_1), respectively. Typically, the contributions of $[W_{YY}^{(2,-1)} - W_{XX}^{(2,-1)}]$ and $[\langle \Gamma_{YYZ}^{(3,-1)}\rangle - \langle \Gamma_{XXZ}^{(3,-1)}\rangle]$ are in quadrature with respect to the angular dependence 2ψ.

Identical conclusions can be reached by directly decomposing the rank-3 gyration tensor $\zeta_{\alpha\beta\gamma}$ into rotational invariants following procedures reviewed in [22]. Such a decomposition yields one scalar $\zeta^{(0)}$, three vectors $\zeta_\gamma^{(1)}$, two deviators $\zeta_{\alpha\beta}^{(2)}$, and one septor $\zeta_{\alpha\beta\gamma}^{(3)}$. Because the $E1E2$ interference terms have no scalar part, it follows that $(\zeta^{(0)}) = 0$. Regarding the vector com-

Table 2. Numerical factors

$$a^{(1,-1)}(l_c, l, l') = \frac{(l_c + l + 1)(l_c + l - 2l')(l_c + 2l' - l + 1)}{(l_c + l)(l_c + l + 2)(l + l' + 1)^2}$$

$$a^{(2,+1)}(l_c, l, l') = \frac{2(2l + 1)(2l' + 1)[6 + 3l_c(l_c + 1) - 2l(l + 1) - l'(l' + 1)]}{(l + l' + 1)(l_c - 3l' + 2l)(l_c + 3l' - 2l + 1)(l_c + l)^2(l_c + l + 2)^2}$$

$$a^{(2,-1)}(l_c, l, l') = \frac{(l' - l)(2l + 1)(2l' + 1)[6 + 3l_c(l_c + 1) - 2l(l + 1) - l'(l' + 1)]}{2(l_c - 3l' + 2l)(l_c + 3l' - 2l + 1)(l_c + l)^2(l_c + l + 2)^2}$$

$$b^{(3,-1)}(l_c, l, l') = \frac{2(2l + 1)(2l' + 1)}{(l + l' + 1)(l_c - 3l' + 2l)(l_c + 3l' - 2l + 1)(l_c + l)^2(l_c + l + 2)^2}$$

ponents, it follows from Section 2.2 that only the vector part collinear with k_γ is involved, i.e., $\zeta_\gamma^{(1)\prime} = \delta_{\alpha\beta}\zeta'_{\alpha\beta\gamma}$. Therefore, $\zeta_\gamma^{(1)}$ must be identified with the expectation value of the anapole component $\langle \Omega_\gamma^{(1,-1)} \rangle = \langle \mathbf{\Omega}_0^{(1,-1)} \rangle$.

Two pseudodeviators can be generated by a symmetric contraction of $\zeta'_{\alpha\beta\gamma}$ [22]:

$$[\zeta'^{(2)}_{\alpha\beta}]_1 = -\frac{1}{2}[\epsilon_{\alpha ij}\zeta'_{ji\beta} + \epsilon_{\beta ij}\zeta'_{ji\alpha}],$$

$$[\zeta'^{(2)}_{\alpha\beta}]_2 = -\frac{1}{2}[\epsilon_{ji\beta}\zeta'_{\alpha ij} + \epsilon_{ji\alpha}\zeta'_{\beta ij}].$$

Given the symmetry properties of the gyration tensor, we can check that $[\zeta'^{(2)}_{\alpha\beta}]_1 = 0$; therefore

$$[\zeta'^{(2)}_{\alpha\beta}]_2 \propto \langle W_{\alpha\beta}^{(2,-1)} \rangle.$$

In Section 3, we address the inverse problem: assuming that we know some physical realization of a rank-2 irreducible tensor $\langle W_{\alpha\beta}^{(2,-1)} \rangle$, we can generate its embedded form in the rank-3 tensor space [22],

$$\zeta'^{(2)}_{\alpha\beta\gamma} \propto 2\epsilon_{\alpha\beta\delta}\langle W_{\delta\gamma}^{(2,-1)} \rangle + \epsilon_{\delta\beta\gamma}\langle W_{\alpha\delta}^{(2,-1)} \rangle,$$

$$\zeta''^{(2)}_{\alpha\beta\gamma} \propto 2\epsilon_{\alpha\beta\delta}\langle N_{\delta\gamma}^{(2,+1)} \rangle + \epsilon_{\delta\beta\gamma}\langle N_{\alpha\delta}^{(2,+1)} \rangle.$$

Keeping in mind that $\zeta'^{(2)}_{\alpha\beta\gamma}$ must be symmetric and $\zeta''^{(2)}_{\alpha\beta\gamma}$ must be antisymmetric under the transposition of the α, β subscripts, we finally find after proper symmetrization that

for XMLD (S_1),

$$[\zeta'^{(2)}_{\beta\beta\gamma} - \zeta'^{(2)}_{\alpha\alpha\gamma}] \propto [\langle W_{\alpha\beta}^{(2,-1)} \rangle + \langle W_{\beta\alpha}^{(2,-1)} \rangle];$$

for XMLD (S_2),

$$[\zeta'^{(2)}_{\alpha\beta\gamma} + \zeta'^{(2)}_{\beta\alpha\gamma}] \propto [\langle W_{\alpha\alpha}^{(2,-1)} \rangle - \langle W_{\beta\beta}^{(2,-1)} \rangle],$$

and for XNCD (S_3),

$$[\zeta''^{(2)}_{\alpha\beta\gamma} - \zeta''^{(2)}_{\beta\alpha\gamma}] \propto \langle N_{\gamma\gamma}^{(2,+1)} \rangle.$$

On the other hand, the septor $\Gamma_{\alpha\beta\gamma}^{(3,-1)}$ must be the natural irreducible representation of the rank-3 tensor $\zeta'_{\alpha\beta\gamma}$. It expected to contribute to both XMLD (S_1) and XMLD (S_2) because

$$[\zeta'^{(3)}_{\beta\beta\gamma} - \zeta'^{(3)}_{\alpha\alpha\gamma}] \propto [\Gamma_{\beta\beta\gamma}^{(3,-1)} - \Gamma_{\alpha\alpha\gamma}^{(3,-1)}],$$

$$[\zeta'^{(3)}_{\alpha\beta\gamma} + \zeta'^{(3)}_{\beta\alpha\gamma}] \propto [\Gamma_{\alpha\beta\gamma}^{(3,-1)} + \Gamma_{\beta\alpha\gamma}^{(3,-1)}].$$

The two approaches are indeed equivalent.

3. PHYSICAL IMPLICATIONS

3.1. Orbital Magnetoelectric Operators

3.1.1. Spin and orbital anapoles. As was first pointed out by Zel'dovich [1], a toroidal solenoid generates not only an annular magnetic field $\mathbf{H}_a(\mathbf{r})$, but also the so-called toroidal current $\mathbf{j}_a(\mathbf{r})$ along the torus axis z [1–3, 39]. The anapole moment $\mathbf{A}$ is defined as the root-mean-square (rms) radius of $\mathbf{j}_a(\mathbf{r})$. As emphasized long ago in [5], one should not confuse the anapole moment with the toroidal dipole moment $\mathbf{M}_{td}$ in the theory of classical electrodynamics [7]. Khriplovich [2]

and others [40] have nevertheless proved that in a stationary state, where $\mathbf{j}_a(\mathbf{r})$ is time-independent, the two moments become equivalent up to a factor of 4π, i.e.,

$$\mathbf{A} = 4\pi\mathbf{M}_{td}.$$

We use this equivalence in Section 4 because the current literature on magnetoelectric solids mostly refers to toroidal dipole moment.

In solid-state physics, annular magnetic fields can be associated with either spin or orbital currents. The magnetoelectric character of a spin anapole [42] is schematically illustrated in Fig. 1. In the presence of a magnetic field **H**, the energy of each spin carrier (electron) depends on its location on the annular orbit to which the electrons are constrained: their distribution is no longer uniform; consequently, an electric polarization **P** is generated in the direction that is mutually orthogonal to **H** or $\mathbf{j}_a(\mathbf{r})$. The case of an orbital anapole was also envisaged by Ginzburg, Gorbatsevich, Kopaev, and their collaborators [8] many years ago, but in a different theoretical perspective.

Following [2] or [43], we can decompose the total anapole moment into its spin and orbital components,

$$\langle \mathbf{A}_{\text{spin}}^{(1,-1)} \rangle = 2\pi\mu_B \sum_i \langle \mathbf{S}_i \times \mathbf{r}_i \rangle, \tag{12}$$

$$\begin{aligned} \langle \mathbf{A}_{\text{orb}}^{(1,-1)} \rangle &= -i\frac{2\pi\mu_B}{3} \sum_i \langle [\mathbf{L}_i^2, \mathbf{r}_i] \rangle \\ &= \frac{2\pi\mu_B}{3} \langle \mathbf{\Omega}^{(1,-1)} \rangle, \end{aligned} \tag{13}$$

where it can be checked that the operator $-i[\mathbf{L}^2, \mathbf{r}]$ is both Hermitian and Θ-odd [43]. We immediately obtain

$$\mathbf{M}_{td}^{\text{orb}} = \frac{\mu_B \langle \mathbf{\Omega} \rangle}{6}.$$

At this stage, we must recall Lloyd's theorem, which states that for (diamagnetic) systems that have an even number of electrons and integral spin, the expectation value of Hermitian Θ-odd operators vanishes [43].

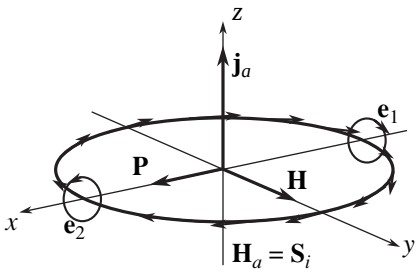

Fig. 1. Magnetoelectric character of a spin anapole: in an external magnetic field **H**, the electron distribution is no longer uniform and induces an electric polarization orthogonal to **H**.

3.1.2. Operators conserved by $I\Theta$. We now consider the perturbation of a system that is in a remanent magnetoelectric state, when one electron is annihilated in a core state and one electron $\{\mathbf{n}, \mathbf{L}, \mathbf{S}\}$ is created in a virtual ground state. By analogy with [44], we expand the energy $U(n, L, S)$ into a MacLaurin series [27]. Using a Cartesian basis, we obtain

$$\begin{aligned} U(n, L, S) &= U(0, 0, 0) \\ &+ \sum_m \left\{ \frac{1}{m!} [n_\alpha \partial_{n_\alpha} + L_\beta \partial_{L_\beta} + S_\beta \partial_{S_\beta}]^m U(0, 0, 0) \right\}. \end{aligned} \tag{14}$$

We need to retain only the magnetoelectric interference terms that are invariant in the product $I\Theta$ and contribute to the nonreciprocal XOA. In the case of *K*-shell ionization, the spin does not play any role and can be neglected. Starting with $m = 2$, we obtain a rank-2 Cartesian tensor $[a_{\alpha\beta}]_{\text{orb}}$ that is the ground state expectation value of the orbital part of the one-electron magnetoelectric tensor associated with the dyad $[L \otimes n]$. Indeed, we can decompose $[a_{\alpha\beta}]_{\text{orb}}$ into its irreducible representations, which include a pseudoscalar (i.e., the trace), the dual vector of the antisymmetric part, and the traceless pseudodeviator of the totally symmetric part. It is then straightforward to show that the integrated XMχD signal, via the expectation value of the orbital anapole moment, is proportional to the dual vector of $[a_{\alpha\beta}]_{\text{orb}}$,

$$\Sigma_{\text{XM}\chi\text{D}(S_0)} \propto \langle \Omega_\gamma^{(1,-1)} \rangle \propto \frac{1}{2}\epsilon_{\alpha\beta\gamma}[a_{\alpha\beta}]_{\text{orb}}.$$

We note that this is a direct transposition of the result established a long time ago by Ascher [39] and by Gorbatsevich *et al.* [41, 42], who pointed out that for a toroidal magnetoelectric solid, the total anapole moment ($\langle A_{\text{total}} \rangle$) is proportional to the dual vector of the rank-2 magnetoelectric tensor $[a_{\alpha\beta}]$, i.e.,

$$\langle A_\gamma \rangle \propto \frac{1}{2}\epsilon_{\alpha\beta\gamma}[a_{\alpha\beta}].$$

We also show that at the $m = 2$ order, $\Sigma_{XMLD}(S_1)$ and $\Sigma_{XMLD}(S_2)$ can similarly be related to the irreducible zero-trace symmetric pseudodeviator $[a_{\alpha\beta}]_{\text{orb}}^{(2)}$ because we have seen that

$$[\langle W_{\alpha\beta}^{(2,-1)} \rangle + \langle W_{\beta\alpha}^{(2,-1)} \rangle] \propto [a_{\alpha\beta} + a_{\beta\alpha}]_{\text{orb}},$$

$$[\langle W_{\alpha\alpha}^{(2,-1)} \rangle - \langle W_{\beta\beta}^{(2,-1)} \rangle] \propto [a_{\alpha\alpha} - a_{\beta\beta}]_{\text{orb}}.$$

This implies that XMLD (S_1) and the Jones dichroism XMLD (S_2) are also first-order magnetoelectric effects, but of orbital nature. Typically, XMLD (S_2) can be

detected in magnetoelectric solids characterized by a magnetoelectric tensor that has nonzero diagonal terms. More precisely, this dichroism should be observed whenever the diagonal terms are not equal in the plane (0, 0, *k*). The dichroism XMLD (S_1) is expected to be detectable only when the magnetoelectric tensor has symmetric off-diagonal terms, a situation which is less frequent.

At the *K*-edge, there is no hope of accessing the spin part of the one-electron magnetoelectric tensor $[S \otimes n]$, which can also be decomposed into the spin anapole $A_{\rm spin}^{(1,-1)}$ and the pseudodeviator $A_{\rm spin}^{(2,-1)}$. *A priori*, nothing can be said about the relative sign and magnitude of the spin and orbital parts of the total magnetoelectric tensor. At most, it may be supposed that for transition metal oxides, the spin part $[a_{\alpha\beta}]_{\rm spin}$ should be much larger than the orbital part $[a_{\alpha\beta}]_{\rm orb}$. To date, magnetoelectric susceptibility measurements were largely dominated by the spin contribution, and to the best of our knowledge, there is not a single example where the orbital part has been extracted. What makes X-ray absorption spectroscopy attractive is indeed its capability to selectively probe the orbital contributions to the magnetoelectric tensor.

At the $m = 4$ order, additional magnetoelectric interference terms can be identified that are odd with respect to parity *I* and time-reversal Θ but remain invariant under the product *I*Θ. Such terms can only be obtained from the two rank-4 Cartesian tensors

$$[b_{\beta\gamma\delta\epsilon}]_{\rm orb} = [L \otimes L \otimes L \otimes n],$$

$$[c_{\beta\gamma\delta\epsilon}]_{\rm orb} = [L \otimes n \otimes n \otimes n].$$

We recall that the rank-4 tensor $[L \otimes n \otimes L \otimes n]$, which contributes to the so-called biquadratic susceptibility [45], is obviously parity-even and therefore cannot contribute to XOA. As far as XOA is concerned, we must only retain irreducible tensors of rank ≤3 that are linear with respect to *n*. We are then left with (at most) three independent Cartesian septors obtained by decomposing $[b_{\beta\gamma\delta\epsilon}]_{\rm orb}$ into irreducible representations. The latter are related to the six dual rank-3 tensors generated by antisymmetric contraction [22], i.e.,

$${}_\tau t_{\alpha\beta\gamma} \propto \epsilon_{\alpha\delta\epsilon}[b_{\beta\gamma\delta\epsilon}]_{\rm orb},$$

with τ varying from 1 to 6. Three independent tensors are easily identified,

$${}_1\Gamma_{\alpha\beta\gamma} = [L \otimes L \otimes \Omega]_{\alpha\beta\gamma} = [Q_{LL}^{(2,+1)} \otimes \Omega^{(1,-1)}]_{\alpha\beta\gamma},$$

$${}_2\Gamma_{\alpha\beta\gamma} = [\Pi_{LL} \otimes L \otimes n]_{\alpha\beta\gamma} = [\Pi_{LL}^{(1,+1)} \otimes W^{(2,-1)}]_{\alpha\beta\gamma},$$

$${}_3\Gamma_{\alpha\beta\gamma} = [L \otimes \Pi_{LL} \otimes n]_{\alpha\beta\gamma},$$

where

$$[\Pi_{LL}^{(1,+1)}]_\gamma = [L \times L]_\gamma = iL_\gamma$$

can be viewed as an example of an antitoroidal vector operator that is not Hermitian. As a consequence, only the expectation value of the first septor operator is real and can contribute to XOA. At this stage, it becomes more convenient to return to the representation in the spherical basis. We first observe that for the septor

$$\langle \mathbf{\Gamma}_0^{(3,-1)} \rangle = \langle \mathbf{Q}_{LL}^{(2,+1)}, \mathbf{\Omega}^{(1,-1)} \rangle$$

to exist, it is sufficient but not necessary that $\langle \mathbf{\Omega} \rangle$ and $\langle \mathbf{Q}_{LL} \rangle \neq 0$ individually. We note that the tensor $\mathbf{Q}_{LL}^{(2,+1)}$, which is Θ- and *I*-even, has the same symmetry and angular dependence as the charge quadrupole operator, although the matrix elements are different. Interestingly, $\mathbf{Q}_{LL}^{(2,+1)}$ was recently shown [46, 47] to be also the effective operator responsible for the reciprocal XMLD of magnetooptical origin [48].

At this stage, it can be anticipated that in analyzing XMχD spectra, we could experience serious difficulties in disentangling the contributions of $\mathbf{\Gamma}_0^{(3,-1)}$ and $\mathbf{\Omega}_0^{(1,-1)}$, especially if these two operators appearing in Eq. (8) are both allowed by symmetry. In principle, the higher order septor term can be expected to be smaller. As discussed in Section 3.3.2 below, comparison of the XMχD spectra recorded with a single crystal or a powder can be very helpful in verifying the validity of this assumption. We assume that we can perform XMLD (S_2) experiments with a single crystal; Eqs. (9)–(11) show that the effective operators $[W_{YY}^{(2,-1)} - W_{XX}^{(2,-1)}]$ and have $[\Gamma_{YYZ}^{(2,-1)} - \Gamma_{XXZ}^{(2,-1)}]$ the same angular dependence (2ψ) when the crystal is rotated around the direction of the incident X-ray beam, but we already pointed out that the two contributions are in quadrature. This implies that the higher order septor must induce only a small phase shift with respect to the dominant XMLD (S_2) signal. The same conclusion must obviously be true for nonreciprocal XMLD (S_1) experiments. Again, the comparison of nonreciprocal XMLD spectra recorded with a single crystal or a powdered sample could be most helpful in evaluating the importance of the septor term. This option is also considered in Section 3.3.2.

In Table 3, we have summarized the effective XOA operators that are irreducible representations of a given magnetoelectric point group. Table 3 is a spinoff of the work by Tenenbaum [34] who listed the number of independent components of the spherical tensors up to rank 4 for 90 magnetic point groups. We recall, however, that this application was justified in Section 2.3.2. For each magnetoelectric class, we indicated the number of independent, nonzero components of the anapole

Table 3. X-ray optically active magnetoelectric group $G_{mag}[\chi\Omega, \chi W, \chi\Gamma]$

Anapole		$\mathbf{\Omega}^{(1)} \neq 0$		$\mathbf{\Omega}^{(1)} = 0$	
Deviator		$\mathbf{W}^{(2)} = 0$	$\mathbf{W}^{(2)} \neq 0$	$\mathbf{W}^{(2)} \neq 0$	
Septor		$\mathbf{\Gamma}^{(3)} \neq 0$		$\mathbf{\Gamma}^{(3)} = 0$	$\mathbf{\Gamma}^{(3)} \neq 0$
AFM	AFE				222[0, 2, 1]
				4/*m*'*m*'*m*'[0, 1, 0]	*m*'*m*'*m*'[0, 2, 1]
		$\bar{3}$'*m* [1, 0, 2]	$\bar{1}$'[3, 5, 7]; *mmm*'[1, 1, 2]	422[0, 1, 0]	4'/m'[0, 2, 2]
		4/*m*'*mm*[1, 0, 1]	2/*m*'[1, 3, 3]; 2'/*m*[2, 2, 4]	$\bar{4}$'2*m*' [0, 1, 0]	4'/*m*'*mm*'[0, 1, 1]
		$\bar{4}$'2'*m* [1, 0, 1]	$\bar{3}$' [1, 1, 3]	$\bar{4}$'*m*'2 [0, 1, 0]	4'*mm*'[0, 1, 1]
		6/*m*'*mm*[1, 0, 1]	$\bar{4}$' [1, 1, 1]; 4/*m*'[1, 1, 1]	622[0, 1, 0]	$\bar{4}$2*m* [0, 1, 1]
		$\bar{6}$'2'*m* [1, 0, 1]	$\bar{6}$' [1, 1, 1]; 6/*m*'[1, 1, 1]	$\bar{6}$'2*m*' [0, 1, 0]	$\bar{4}$2'*m*' [0, 1, 1]
				$\bar{4}$'*m*'2 [0, 1, 0]	32[0, 1, 1]
					$\bar{3}$'*m*' [0, 1, 1]
AFM	FE	3*m*[1, 0, 2]	*mm*2[1, 1, 2]		
		4*mm*[1, 0, 1]; 6*mm*[1, 0, 1]	(2*mm*)*[1, 1, 2]		4'[0, 2, 2]
			(*m*2*m*)*[1, 1, 2]		
FM	AFE	32'[1, 0, 2]			$\bar{4}$ [0, 2, 2]
		42'2'[1, 0, 1]; 62'2'[1, 0, 1]			
FM	FE		3[1, 1, 3]	4*m*'*m*'[0, 1, 0]	*m*'*m*'2[0, 2, 1]
			4[1, 1, 1]; 6[1, 1, 1]	6*m*'*m*'[0, 1, 0]	3*m*'[0, 1, 1]
Weak FM	AFE		22'2'[1, 1, 2]		
			(2'22')*[1, 1, 2]; (2'2'2)*[1, 1, 2]		
Weak FM	FE		1[3, 5, 7]; 2[1, 3, 3]; 2'[2, 2, 4]		
			m[2, 2, 4]; *m*'[1, 3, 3]		
			m'*m*2'[1, 1, 2]; *mm*'2'[1, 1, 2]		

* Nonstandard groups.

($\chi_\Omega \leq 3$), of the pseudodeviator $\mathbf{W}^{(2,-1)}(\chi_W \leq 5)$, and of the pseudoseptor $\Gamma^{(3,-1)}(\chi_\Gamma \leq 7)$. We have identified 34 "toroidal point groups" (but only 31 classes) [6, 39] that admit the anapole as an irreducible representation and we found that all of them also admit $\mathbf{\Gamma}^{(3)}$ as an irreducible representation. We have also found 22 "nontoroidal groups" that admit the pseudodeviator $\mathbf{W}^{(2,-1)}$ as an irreducible representation and may exhibit nonreciprocal XMLD; interestingly, 13 of them still admit the pseudoseptor $\mathbf{\Gamma}^{(3,-1)}$ as an irreducible representation. Not listed in Table 3 are the magnetic classes that are not magnetoelectric but still admit $\mathbf{\Gamma}^{(3)}$ as irreducible representations,

$$6';\ \bar{6}\,;\ 6'/m;\ 6'22';\ 6'mm';\ \bar{6}\,m2;\ 6'/mmm'$$

$$23;\ m'3;\ m'3m;\ 4'32;\ \bar{4}\,3m.$$

For nonreciprocal XOA to be detectable, it is imperative that the orbital magnetoelectric group must belong to the groups listed in Table 3. This is not sufficient, unfortunately, because Table 3 does not tell us whether the specific representations $\mathbf{W}^{(2,-1)}_{(\pm 2)}$ and $\mathbf{\Gamma}^{(3,-1)}_{(0,\pm 2)}$ are allowed. This is where Eq. (6) has to be used. In the specific case of the Jones dichroism, one can alternatively exploit the fact that the Cartesian tensors $W^{(2,-1)}_{\alpha\beta}$ must have the same form as the magnetoelectric tensors in [49] or [50]: using Eq. (10), it is then a trivial exercise to identify which magnetoelectric groups give a nonreciprocal dichroism XMLD (S_2).

3.1.3. Operators not conserved by $I\Theta$. The so-called higher order magnetoelectric effects, or the induced magnetoelectric effects in paramagnetic systems [51], are commonly associated with rank-3 susceptibility tensors referring to $H_\alpha H_\beta E_\gamma$ or $E_\alpha E_\beta H_\gamma$ [52].

The corresponding tensors are therefore odd with respect to $I\Theta$ and can be identified with cross terms in the MacLaurin expansion of the energy U at the intermediate order ($m = 3$). As pointed out in [52], these additional terms must be taken into consideration for magnetic groups that are compatible with either piezomagnetism or piezoelectricity. Neither group $\bar{3}'m'$ of Cr_2O_3 nor groups $2/m'$ and $2'/m$ to be considered in Section 4 for $(V_{1-x}Cr_x)_2O_3$ belong to these classes, but we nevertheless feel it useful to look at the relevant effective operators listed in Table 4.

Because we are primarily interested in the σ_{E1E2} X-ray absorption cross section, we first consider the case of the odd-parity *HHE* susceptibilities. The Hermitian operators listed in the first column of Table 4 can be seen as describing magnetic field-induced magnetoelectric properties. It immediately appears, however, that the first three operators in the first column are the effective operators for natural optical activity, as discussed in more detail in Section 3.2. Of particular importance is the scalar term, which is a parity-violating energy of orbital origin but is fully consistent with the formulation in [I]. For example, in a population of resolved chiral species in a disordered fluid phase, each enantiomer must bear an orbital anapole moment with a well-defined sign in the molecular coordinate system, but because the orientation of molecules is random in a disordered fluid phase, there is obviously no magnetochiral dichroism that can be detected. In the presence of a strong external field **H**, the magnetoelectric energy of the system becomes

$$[\mathbf{L}\cdot\boldsymbol{\Omega}]_{\mathbf{H}} = [\mathbf{L}_0\cdot\boldsymbol{\Omega}_0] + [\chi_{\text{orb}}\mathbf{H}\cdot\boldsymbol{\Omega}_0] + [\mathbf{L}_0\cdot\Delta\boldsymbol{\Omega}_{\text{H}}] + \ldots,$$

where

$$\mathbf{L} = \mathbf{L}_0 + \chi_{\text{orb}}\mathbf{H}.$$

Neglecting the field-induced anapole moment $\Delta\Omega_H$ in the first approximation, we expect the system to minimize its magnetoelectric energy with an anisotropic angular distribution of the anapole preferably oriented along the direction of the external magnetic field **H**. As a consequence, one may anticipate that a (weak) paramagnetochiral dichroism (XMχD) might be found. Baranova and Zel'dovich [53] and others [54, 55] predicted a long time ago that such a dichroism should be detectable at optical wavelengths, where the contribution of the *E*1*M*1 interference terms is dominant; but the theory of optical magnetochiral dichroism (OMχD) is more complicated because the Zeeman effect and the contribution of the spin anapole must also be taken into account. The first OMχD spectra were reported rather recently in solutions of paramagnetic chiral compounds [56, 57] and even in diamagnetic systems [58, 59]. Nevertheless, no XMχD could unfortunately be detected as yet on chiral paramagnetic solutions. On the other hand, the problem of detecting XMχD spectra using powdered samples of magnetic chiral complexes is different because the orientations of the crystallites are frozen and another way to define the quantization axis must be found. In this case, it is desirable to combine the electric and magnetic fields in a geometry depending on the magnetic group of the sample. Further work is in progress at the ESRF in order to explore this possibility, which is reminiscent of the induced magnetoelectric effect detected in the paramagnetic phase of $NiSO_4\cdot 4H_2O$ [60].

Table 4. Operators for induced magnetoelectric susceptibilities

$I = -1, \Theta = +1$(*HHE*) *E*1*E*2 compatible	$I = +1, \Theta = -1$(*EEH*) *E*2*E*2 + *E*1*E*3 compatible
Piezo-electric	Piezo-magnetic
$[\mathbf{L}\cdot\boldsymbol{\Omega}]^{(0)}$	$[\mathbf{n}\cdot\boldsymbol{\Omega}]^{(0)}$
$[\mathbf{L}\times\boldsymbol{\Omega} - \boldsymbol{\Omega}\times\mathbf{L}]^{(1)}$	$[\mathbf{n}\times\boldsymbol{\Omega} - \boldsymbol{\Omega}\times\mathbf{n}]^{(1)}$
$[\mathbf{L}, \boldsymbol{\Omega}]^{(2)}$	$[\mathbf{n}, \boldsymbol{\Omega}]^{(2)}$
$[\mathbf{L}, \mathbf{W}^{(2)}]^{(3)}$	$[\mathbf{n}, \mathbf{W}^{(2)}]^{(3)}$

For completeness, we have also listed the effective operators related to the parity-even *EEH* susceptibilities in column 2 of Table 4. These operators can be seen as describing the magnetoelectric properties induced by the electric field. The first term is the free energy violating the time-reversal symmetry. It should be kept in mind that the corresponding systems having even parity are, strictly speaking, no longer relevant to optical activity. In the X-ray range, they could nevertheless contribute to the σ_{E2E2} or σ_{E1E3} absorption cross sections, which are unfortunately significantly smaller than σ_{E1E2}.

3.2. Natural X-ray Optical Activity

According to [27], the effective operator

$$\mathbf{N}^{(2,+1)} = [\mathbf{L}, \boldsymbol{\Omega}]^{(2)}$$

associated with natural XOA is the Θ-even direct product of two Θ-odd operators that are both related to orbital magnetism. This suggests viewing the natural XOA either as a "degenerate" case of orbital magnetism or as a particular case of the "induced orbital magnetoelectric effect." We also note that $\langle\mathbf{N}^{(2,+1)}\rangle$ may well be nonzero even when either ($\langle\mathbf{L}\rangle$) or $\langle\boldsymbol{\Omega}^{(1,-1)}\rangle$ is zero. This can be easily illustrated with the case of diamagnetic chiral compounds: Lloyd's theorem implies that $\langle\boldsymbol{\Omega}^{(1,-1)}\rangle$ must vanish, whereas $\langle[\mathbf{L}, \boldsymbol{\Omega}]^{(2,+1)}\rangle$, which is Hermitian but Θ-even, can remain perfectly finite.

There is another case that deserves special attention: if the expectation value of the orbital anapole moment

is nonzero along the direction of the wave vector **k**, then the system must exhibit a magnetochiral dichroism (XMχD) in addition to the natural circular dichroism (XNCD). Moreover, one would expect the external magnetic field not only to create a magnetization vector **M**, but also to stabilize one isomer with respect to its enantiomer as a consequence of the parity-violating free energy [$\mathbf{\Omega} \cdot \mathbf{M}$]; this effect has been proved experimentally using OMχD [57]. This experiment may shed new light on a long-lasting debate regarding the existence of chirality in prebiotic chemistry [61–63], because it suggests that the action of a strong magnetic field can suffice to resolve optical enantiomers. This would revitalize the old view of Pasteur [64] that an intrinsic dissymmetric force is inherent to the physical world. We recall that Pasteur in his remarkable intuition tried hard for many years to show that chirality and magnetism are connected [64], but he could not prove this within the knowledge of his time.

As already illustrated with the first column of Table 4, the dyad [**L**, **Ω**] can be decomposed into three irreducible representations: the pseudoscalar

$$\mathbf{N}^{(0)} = [\mathbf{L} \cdot \mathbf{\Omega}]^0,$$

the dual polar vector

$$\mathbf{N}^{(1)} = [\mathbf{L} \times \mathbf{\Omega} - \mathbf{\Omega} \times \mathbf{L}]^{(1)},$$

and the pseudodeviator $\mathbf{N}^{(2)}$, which was shown to play a key role in XOA. A priori, the scalar part $\mathbf{N}^{(0)}$ could only be associated with the *E*1*M*1 interference terms that dominate optical activity at optical wavelengths but can be neglected in the X-ray range as proved in the next subsection. One may wonder, however, whether any specific XOA effect may be related to the vector term $\mathbf{N}^{(1)}$. An interesting indication can be found in the early works [65] and [66], where it was suggested that a new type of optical activity can be measured in the reflectivity mode for several crystal classes. More recently [67], it was pointed out that these mysterious classes are precisely associated with the irreducible vector part of the optical activity tensor. In Table 5, following [67], we have listed the crystal classes that can contribute to a scalar-, vector-, or tensor-type XOA according to symmetry. However, we must identify where the theory developed in Section 2.2 is to be modified in order to become compatible with the eventual detection of the vector part of natural optical activity in the X-ray regime. The solution to this puzzling problem was more or less given out in [68], where it was pointed out that in all crystal classes exhibiting vector optical activity, the electromagnetic wave propagating inside the crystal is not transversally polarized but has a so-called skew polarization with an axial component. It should be realized, however, that the absorption cross section σ_{E1E2} of such a very weak axial component is a second-order dichroism, whose detection would be a considerable challenge for experimentalists. Recently, we nevertheless succeeded in detecting the vector type of natural optical activity of a diamagnetic zinc oxide (ZnO) single crystal, in a geometry optimized for X-ray resonant scattering [69].

Induced natural optical activity can also be predicted to occur as a consequence of the $m = 4$ terms in Eq. (14). The two rank-3 operators

$$\mathbf{\Gamma}^{(3,-1)} = [[\mathbf{L}, \mathbf{L}], \mathbf{\Omega}]^{(3,-1)},$$

$$\mathbf{\Delta}^{(3,+1)} = [[\mathbf{n}, \mathbf{L}], \mathbf{\Omega}]^{(3,+1)}$$

can induce natural optical activity. We have already emphasized that $\mathbf{\Gamma}^{(3,-1)}$ is odd with respect to I and Θ, whereas $\mathbf{\Delta}^{(3,+1)}$ is even with respect to both I and Θ. Typically, $\Delta^{(3,+1)}$ originates in the $m = 4$ biquadratic magnetoelectric susceptibility and appears as the operator responsible for electrogyration in centrosymmetric solids; under high magnetic fields, $\mathbf{\Gamma}^{(3,-1)}$ can induce XNCD signals in noncentrosymmetric magnetic systems, possibly in powdered samples. It seems that the latter operator could be responsible for the so-called quadratic Faraday effect of optically active systems. We recall that there are magnetic groups that are not magnetoelectric but nevertheless admit $\mathbf{\Gamma}^{(3,-1)}$ as irreducible representations. Nevertheless, the $m = 4$ terms in Eq. (14) are expected to be rather small; as of yet, we have failed to prove that electrogyration can be measured in the X-ray spectral range.

Table 5. Rotational invariants of natural optical activity

Natural OA	Irreducible parts of $\mathbf{L} \otimes \mathbf{\Omega}$		
Crystal classes	Pseudoscalar	Polar vector	Pseudo-deviator
1; 2; 3; 4; 6	1	1	1
622; 32; 422; 222	1	0	1
m; *mm*2	0	1	1
$\bar{4}$; $\bar{4}2m$	0	0	1
6*mm*; 3*m*; 4*mm*	0	1	0
432; 23	1	0	0

3.3. Rotational Isotropy

3.3.1. XNCD spectra. It was obvious from the beginning that the X-ray natural circular dichroism (XNCD) can hardly be detectable in powders or solutions because the rank-3 tensor *E*1*E*2 has no scalar part. This is not surprising because the spherical harmonics

associated with the electric dipole ($l = 1$) and electric quadrupole ($l = 2$) are orthogonal in a sample that is orientationally isotropic. For the sake of illustration, we have reproduced in Fig. 2a the cobalt *K*-edge X-ray absorption near edge (XANES) and the XNCD spectra of two resolved enantiomers of the chiral "propeller-like" complex

$$\mathbf{1}^{\pm} = 2[Co(en)_3Cl_3] \cdot NaCl \cdot 6H_2O.$$

In these compounds, the ligand field has the D_3 point group symmetry. As already reported elsewhere [18], the XNCD spectra of the two enantiomers have opposite signs. In Fig. 2b, we have compared the XNCD spectra recorded with either a single crystal or a pellet of a powdered sample of the same enantiomer. With the powdered sample, the strong XNCD signature assigned to the *E*1*E*2 interference terms totally vanishes. However, a very weak signal that has the opposite sign is left in the preedge range (the normalized amplitude is approximately 2.5×10^{-4}). It is impossible to transform the XNCD spectrum of a given enantiomer into the spectrum of its mirror image by a simple rotation. This implies that the very weak signal observed in the powdered sample cannot be explained by any residual orientational order in the powder. It is therefore our interpretation that this weak signal should be of a different nature and can be associated with small *E*1*M*1 pseudoscalar interference terms.

Regarding the photoexcitation of deep core states, a monoelectronic *M*1*M*1 transition is forbidden for two independent reasons: (i) the angular momentum operator **L** has a zero eigenvalue for a spherically symmetric 1*s* core state (e.g., in the case of *K*-edge photoionization); (ii) in a central-field atomic model, one-electron radial wave functions with the same l and different energies are orthogonal, and the magnetic dipole transition matrix element therefore vanishes. In a many-body picture, the second argument no longer applies because different potentials must be used to describe initial and final one-electron states [13], but argument (i) is still a problem. It is our interpretation that *E*1*M*1 transitions can nevertheless be allowed in the case of a multielectron excitation process. This interpretation is supported by the derivation [37] of a two-particle *E*1*M*1 sum rule via the same procedure as that described in Section 2.3. The calculated effective operator was identified as a two-particle orbital pseudoscalar $\mathbf{N}^{(0)} = \mathbf{L} \cdot \mathbf{\Omega}$. Typically, one X-ray photon would cause the simultaneous photoexcitation of two electrons, one in the deep *K*-shell and the other in the valence band. That the effective operator vanishes ($\mathbf{N}^{(0)} \equiv 0$) for a single particle follows from the definition,

$$\mathbf{N}^{(0)} = \mathbf{L} \cdot [(\mathbf{L} \times \mathbf{n}) - (\mathbf{n} \times \mathbf{L})]/2.$$

The experimental and theoretical results thus suggest that (difficult) XNCD experiments on powdered samples could possibly give access to the effective operators of parity-mixing many-body processes, of which very little is presently known.

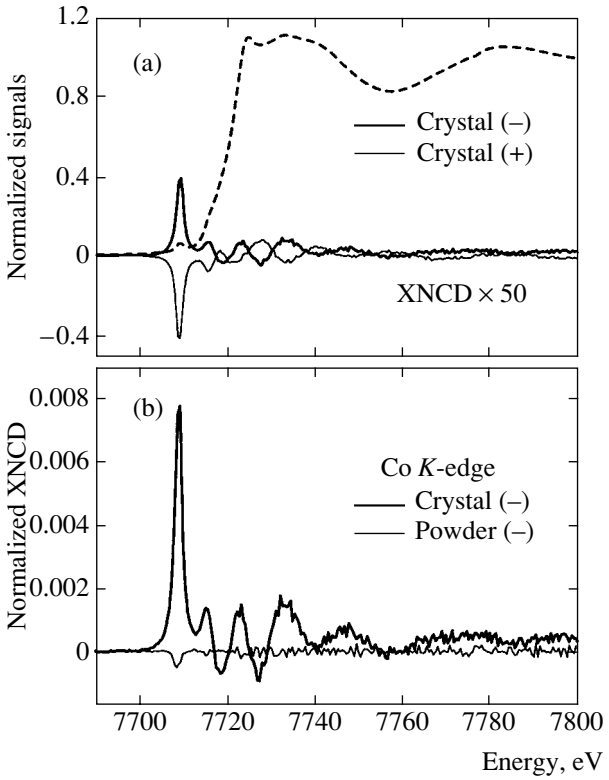

Fig. 2. Co *K*-edge XNCD spectra of the resolved enantiomers of the chiral complex $\mathbf{1}^{\pm} = 2[Co^*(en)_3Cl_3] \cdot NaCl \cdot 6H_2O$. a—XNCD spectra recorded with single crystals of (+) and (–) enantiomers. A polarization-averaged XANES spectrum was added for comparison. b—XNCD spectra of the (–) enantiomer as a single crystal or as a powdered pellet. Note the very weak inverted signal obtained with the powdered pellet.

When no single crystal is available, there is still the possibility of recovering a well-detectable $XNCD_{E1E2}$ signal: the idea is to artificially break the orientational isotropy of space, e.g., by investigating liquid crystal phases aligned in a high magnetic field or chiral ferromagnets below their Curie temperature [70]. As an example, we report the XNCD spectra of another stereogenic organometallic complex,

$$\mathbf{2}^{\pm} = [(L_1)(L_2)Fe^{II}(L_3)(L_4)],$$

dissolved in an aligned liquid crystal. In this tetra-coordinated iron complex, the absorbing atom (Fe) is clearly in a chiral ligand field because all the four ligands are different (the point group C_1): L_1 is the cyclopentadienyl ligand ($\eta^5 - C_5H_5$), L_2 is an iodine atom (–I), L_3 is a carbonyl group (–CO), and L_4 is a chiral tertiary phosphine ($-PPh_2R$) with R = (–NMe–C*HMePh). The stereoselective synthesis of the corre-

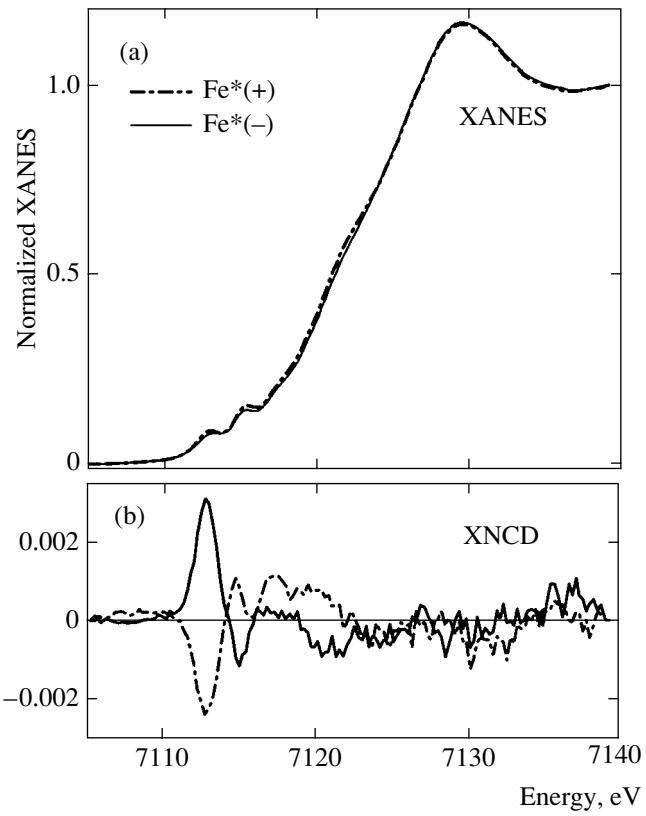

Fig. 3. Fe *K*-edge XNCD spectra of the two resolved enantiomers of the chiral complex $\mathbf{2}^{\pm}$ = (η^5 – C_5H_5)Fe*(–I)(–CO)[–PPh_2(–NMe–C*HMePh)] dissolved in a liquid crystal phase (Merck ZLI 4814). All spectra were recorded in the fluorescence detection mode using a high magnetic field (5 T) to align the liquid crystal and the solute. a—Polarization-averaged XANES spectra of each (±) enantiomer. b—XNCD spectra of the two (±) enantiomers.

sponding diastereoisomers was first described in [71] and was reproduced for us at the University of Dijon (France). Because no large-size single crystals could be grown, the enantiomers were dissolved in a liquid crystal that was known to exhibit a strong diamagnetic anisotropy (Merck: MLC-6204; T_c = 66°C) and each chiral sample was aligned in a 5 T magnetic field directed along the wave vector **k** of the incident X-ray beam. We recall that the exploitation of mesophase-oriented solutes has become a very popular technique in NMR and ESR since the pioneering work of Saupe in 1963 [72, 73]. In Fig. 3, we have reproduced the Fe *K*-edge XANES and XNCD spectra of the two enantiomeric solutions. The two XNCD spectra have clearly opposite signs, as expected. The price that we had to pay was clearly a dramatic loss of sensitivity, not only because the solubility of the chiral complexes was very poor, but also because the (unknown) order parameter of the solute itself inside the liquid crystal phase was probably rather low. We note that *ab initio* simulations of the experimental XNCD spectra turned out to be impossible unfortunately due to the lack of information regarding the preferential orientational order of the solute in the oriented liquid crystal.

3.3.2. Nonreciprocal XMχD and XMLD spectra. Magnetochiral dichroism (XMχD) spectra of Cr_2O_3 were successively recorded using either a single crystal or a powdered sample [14]. As illustrated by Fig. 4, the most significant difference between the two spectra is a reduction of the signal, by approximately 6 : 1 in the experiment carried out with the powdered sample. It also appears that the normalized intensity of the magnetochiral dichroism spectrum measured with the single crystal exceeds the intensity of the XNCD spectra reproduced in Fig. 2; this might well be consistent with our remark that the magnetoelectric susceptibilities $[a_{\alpha\beta}]_{\text{orb}}$ appear in the lowest order term $m = 2$ of the series expansion of the energy $U(\mathbf{L}, \mathbf{n})$, whereas the effective operators for XNCD contribute to the $m = 3$ susceptibilities.

The primary aim of this section is to show that the proven capability to record XMχD spectra using powdered samples is fully consistent with the proposed sum rule analysis and also consistent with our interpretation that the leading term in Eq. (3) should be the contribution of the projection of the orbital anapole moment $\mathbf{\Omega}_0^{(1)}$ along the direction of the wave vector **k**. Since the pioneering works of Astrov [74, 75], it has been well documented that the key step in measurement of magnetoelectric susceptibility is the creation of a remanent state characterized by a strong polarization of the magnetoelectric domains. This is rather well understood for Cr_2O_3, which has only two magnetoelectric domains (±) that can be exchanged by reversing the time and are illustrated with Fig. 5. If $n_{(+)}$ and $n_{(-)}$ denote the number densities of the two types of domains, we are directly concerned in our experiment with the magnetoelectric polarization ratio

$$\rho_{ME} = \frac{n_{(+)} - n_{(-)}}{n_{(+)} + n_{(-)}}.$$

We found it most convenient to adapt the model proposed in [76] to describe the nucleation of magnetoelectric domains by annealing.

We start from a crystal that is described by the tensors $\langle \mathbf{T}_\beta^{(b,\theta)} \rangle_X$ in the crystalline axes. In a powder, the crystalline axes of a given crystallite *i* are rotated with respect to the reference frame of the experiment, with the rotation described by the Euler angles ϕ_i, θ_i, and ψ_i. We assume that the electric and magnetic fields are parallel to the *z* axis of the reference frame of the experiment and that the magnetoelectric tensor of the crystal is diagonal (with $\alpha_{xx} = \alpha_{yy}$) in the reference frame of the

crystal. The magnetocrystalline energy of the crystallite is therefore proportional to (see [77])

$$U_{ME}(\theta_i) = -\mathbf{E}\cdot\alpha\cdot\mathbf{H}$$
$$= -EH(\alpha_{zz}\cos^2\theta_i + \alpha_{xx}\sin^2\theta_i)$$

for a domain of the magnetoelectric type and to $-U_{ME}(\theta_i)$ for domains of other types. At a temperature of T_N, the polarization ratio is given by

$$\frac{n_{(+)} - n_{(-)}}{n_{(+)} + n_{(-)}} = \rho_{ME}(\theta_i) = \tanh\left(\frac{U_{ME}(\theta_i)}{kT_N}\right). \qquad (15)$$

The XOA experiments refer to tensors $\langle \mathbf{T}_\beta^{(b,\theta)}\rangle_X$ that are not only parity-odd but also change sign for domains of different types. For a crystallite *i*, the tensors in the reference frame of the experiment become

$$\sum_{\beta'} \mathbf{D}^{(b)}_{\beta'\beta}(\phi_i, \theta_i, \psi_i)\langle \mathbf{T}_{\beta'}^{(b,\theta)}\rangle_X \tanh(U_{ME}(\theta_i)/kT_N).$$

To obtain the tensor components $\langle \mathbf{T}_\beta^{(b,\theta)}\rangle_p$ of the powder, we calculate the average of the last expression over ϕ_i, θ_i, and ψ_i. The average over ϕ_i and ψ_i gives $\beta' = 0$ and $\beta = 0$. From

$$\mathbf{D}^{(b)}_{00}(\phi_i, \theta_i, \psi_i) = P_b(\cos\theta_i)$$

(where P_b is a Legendre polynomial), we find that

$$\langle \mathbf{T}_\beta^{(b,\theta)}\rangle_p = \delta_{\beta,0}\langle \mathbf{T}_0^{(b,\theta)}\rangle_X I_b,$$

where

$$I_b = \int_0^1 P_b(x)\tanh\{\alpha[(1+\beta^2)x^2 - \beta^2]\}dx,$$

with

$$\beta^2 = -\frac{a_{xx}}{a_{zz}} \ll 1, \quad \alpha = -\frac{EH\alpha_{zz}}{kT_N}.$$

We note that the integral is restricted by symmetry to the range 0 to 1. In this context, the macroscopic magnetoelectric susceptibility $J_2 = (2I_2 + I_0)/3$ measured with a powdered sample must be given by [76]

$$J_2 = \int_0^1 x^2\tanh\{\alpha[(1+\beta^2)x^2 - \beta^2]\}dx.$$

In Table 6, we have regrouped the calculated values of the integrals I_b (≤3) for typical values of the parameters α and β^2. Because the efficiency of the magnetoelectric annealing procedure carried out with the single

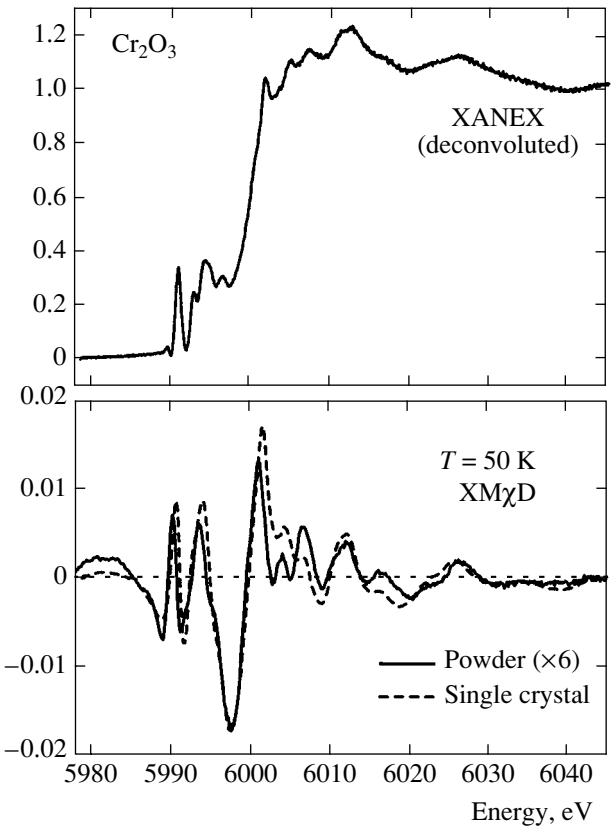

Fig. 4. Cr *K*-edge XMχD spectra of Cr_2O_3 recorded with either a single crystal (**c** || **k**) or a powdered sample. The differential absorption spectra refer to the 180° domains grown under the condition of time-reversality after magnetoelectric annealing. The upper trace reproduces a high-energy resolution (deconvoluted) XANES spectrum.

crystal is unknown, the values listed in Table 6 are systematically normalized with a constant scaling factor slightly less than unity, the $\alpha_0 \approx 0.995$ if $\alpha_0 = 3$. Following [76], we have assumed in the first two examples that, at least near the Néel temperature T_N, $\beta^2 \approx 0$, and we compared the integrals obtained with $\alpha_0 = 3$ and $\alpha = 1$. In the last simulation, which seems to be a reasonable approximation of our experimental conditions, we selected $\alpha = 1$ and $\beta^2 = 0.15$. In all cases, I_3 is quite small (≤0.05); it even reverses its sign when the magnetoelectric polarization of a well-oriented crystallite is assumed to be as effective as in the case of a single crystal ($\alpha_0 = 3$). We also note that the experimental value $J_2 \approx 0.3$ reported in [75] is very close to the asymptotic value that should be measured when the annealing process is as effective in the powder as in the single crystal. In practice, unless very careful annealing procedures are used (e.g., heating the powder at 1300°C in an inert atmosphere), the local electric field in the powder can be dramatically reduced due to the hygroscopic character of the powder, while the conductivity increases, as pointed out in [77]. As a result, we expect a certainly

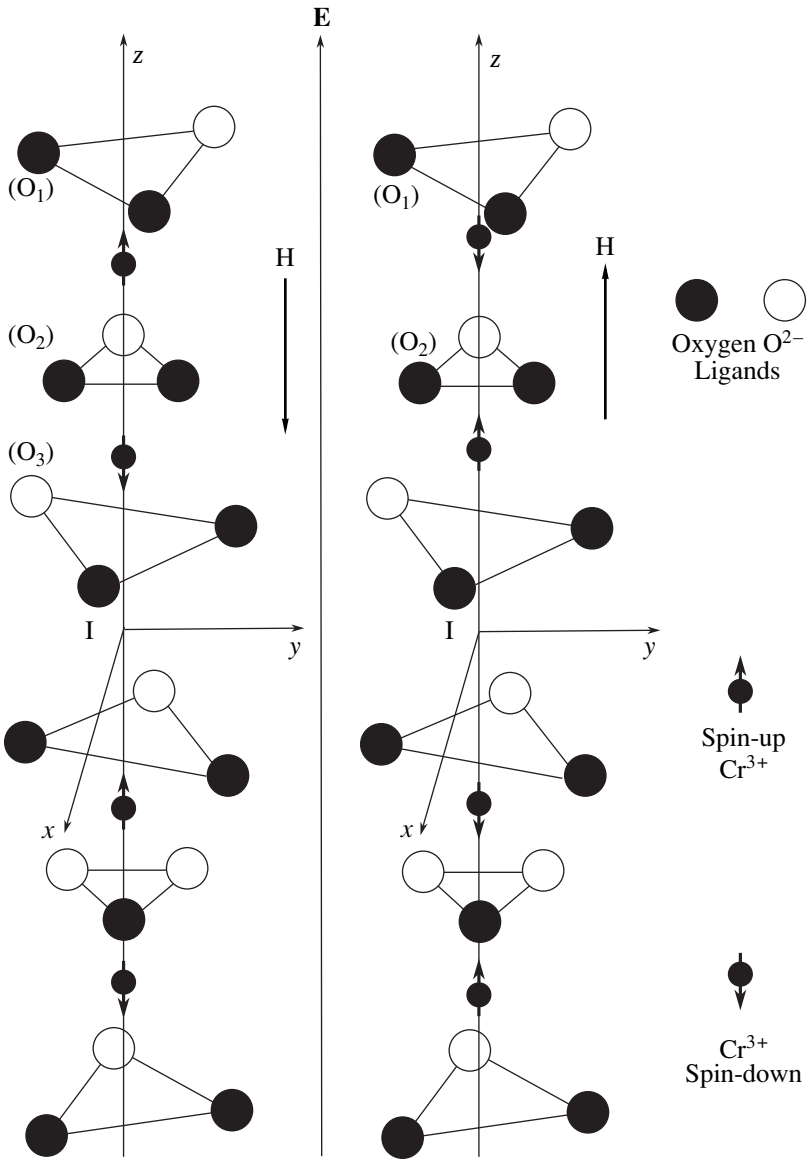

Fig. 5. Schematic representation of the two 180° antiferromagnetic domains grown by magnetoelectric annealing with antiparallel (left panel) and parallel (right panel) electric (**E**) and magnetic (**H**) fields.

lower efficiency of our annealing, as reflected by the $\alpha \approx 0.76$ for $\alpha = 1$.

An important result of this calculation is that $\langle \mathbf{\Gamma}_0^{(3,-1)} \rangle$ should make only a very small contribution in the powdered sample; because the two XMχD spectra displayed in Fig. 4 exhibit only minor differences, we are therefore led to the important conclusion that the application of the sum rule to the XMχD spectrum recorded with the single crystal should yield a reasonable estimate of the orbital anapole moment.

A further question is whether nonreciprocal XMLD spectra can also be recorded using powdered samples. The answer strongly depends on how the magnetoelectric annealing procedure is conducted. We can assume, for instance, that the same type of annealing is again performed with a powdered sample of Cr_2O_3, but in a different geometry, $\mathbf{E} \parallel \mathbf{H} \perp \mathbf{k}$. In other terms, we have set a different direction of quantification for the magnetoelectric domains in the laboratory coordinates $\{X, Y, Z\}$, whereas the free energy in the crystallite coordinates $\{x, y, z\}$ remains unchanged. Hence,

$$\langle W_{XX}^{(2,-1)} \rangle - \langle W_{YY}^{(2,-1)} \rangle \propto \int_0^1 [(a_{zz} - a_{xx})x^2 + a_{xx}]_{\text{orb}} \times \tanh\{\alpha[(1+\beta^2)x^2 - \beta^2]\} dx.$$

Table 6. Rotational average integrals for powdered samples

Parameters	I_0	I_1	I_2	I_3	J_2
$\alpha_0 = 3.0$, $\alpha = 3.0$, $\beta^2 = 0.0$	0.565	0.387	0.150	−0.008	0.289
$\alpha_0 = 3.0$, $\alpha = 1.0$, $\beta^2 = 0.0$	0.296	0.218	0.109	0.026	0.172
$\alpha_0 = 3.0$, $\alpha = 1.0$, $\beta^2 = 0.15$	0.202	0.184	0.131	0.051	0.155

We thus expect a nonreciprocal XMLD (S_2) signal to be measurable in the powdered sample: it should now be proportional to J_2, under the assumption that

$$(a_{zz} - a_{xx})_{\text{orb}} \neq 0.$$

It would be interesting to compare such a nonreciprocal XMLD spectrum with a test experiment carried out with a single crystal in a geometry satisfying the condition $\mathbf{c} \perp \mathbf{k}$, with the magnetoelectric annealing still performed with $\mathbf{E} \parallel \mathbf{H} \parallel \mathbf{c}$. A comparison of this type would yield valuable information regarding the importance of the septor term in XMLD experiments.

In powdered samples, as suggested in [78], annealing could be carried out in electric and magnetic fields arbitrarily oriented with respect to each other. In the crystallite coordinates $\{x, y, z\}$, the relevant magnetoelectric free energy must be replaced by

$$U_{ME} \propto -EH[(a_{zz}\cos^2\theta_i + a_{xx}\sin^2\theta_i)\cos\beta_0 + (a_{zz} - a_{xx})\sin\theta_i\cos\theta_i\sin\phi_i\sin\beta_0].$$

Because the term proportional to $\sin\phi_i$ has zero average in the calculation of the modified integral J_2, we can anticipate that the cost of this is a further reduction of the annealing efficiency proportional to $\cos\beta_0$, where β_0 denotes the angle between the electric and magnetic fields. This result was not really unexpected. It is, however, restricted to magnetoelectric solids that have a diagonal magnetoelectric tensor with $a_\parallel \gg -a_\perp$. We will consider the general case of the magnetoelectric annealing of powdered samples depending on the magnetoelectric symmetry of the crystallites elsewhere [79].

4. APPLICATIONS OF X-RAY OPTICAL ACTIVITY

4.1. Magnetoelectric Symmetry

4.1.1. Chromium sesquioxide Cr_2O_3. The detection of rather intense XMχD spectra in the magnetoelectric phase of Cr_2O_3 (eskolaite) is somewhat puzzling because the universally cited magnetic group of Cr_2O_3, i.e., $\bar{3}'m'$ does not admit the anapole as an irreducible representation, $\mathbf{\Omega}_0^{(1)} = 0$. According to Table 3, the septor $\mathbf{\Gamma}_\beta^{(3,-1)}$ must have only a single nonzero component, but the table itself does not tell us whether this component is for $\beta = 0$. This is precisely where the method developed in Section 2.3 can help us. For the magnetic group $\bar{3}'m'$, we obtain

$$\langle \mathbf{T}_\beta^{(1,-1)} \rangle = 0,$$

$$\langle \mathbf{T}_\beta^{(2,+1)} \rangle = 0,$$

$$\langle \mathbf{T}_\beta^{(2,-1)} \rangle = \delta_{\beta,0}\mathbf{T}_0^{(2,-1)},$$

$$\langle \mathbf{T}_3^{(3,-1)} \rangle = \langle \mathbf{T}_{-3}^{(3,-1)} \rangle = \frac{\mathbf{T}_3^{(3,-1)} + \mathbf{T}_{-3}^{(3,-1)}}{2}.$$

The other components of $\mathbf{T}^{(3,-1)}$ are zero, including the one for $\beta = 0$. Moreover, we note that when the wave vector is directed along the z axis of the reference frame (which is also the **c** axis of the crystal), then $\mathbf{T}_{\pm 3}^{(3,-1)} = 0$. In other terms, $\mathbf{T}^{(3,-1)}$ cannot be detected in this geometry. In conclusion, there is no optical activity of any type compatible with the group $\bar{3}'m'$ in the geometry of the experiment.

It must also be kept in mind that whatever the true magnetic group of Cr_2O_3 may ultimately be, a component $\Gamma_0^{(3,-1)}$, if any, should make only a very weak contribution to the spectrum recorded in the powdered sample. It was argued in the previous section that the spectrum recorded with the single crystal should yield a reasonable estimate of some orbital anapole moment. In all cases, this would imply a reduction of the magnetic symmetry in what we have previously called a pseudoground state. At this stage, we are left with interpretations of two types, which we now consider successively.

1. The observed reduction of the ground state magnetic symmetry of Cr_2O_3 is related to experimental conditions favoring some metamagnetic phase.

Some ambiguity may possibly stem from the fact that the XMχD spectra were recorded in the presence of a rather modest magnetic field (0.5 T) directed along the **c** axis. We recall that this magnetic field was required only to grow single antiferromagnetic domains, no magnetic field being needed *a priori* to record the XMχD spectra. It has been argued that the

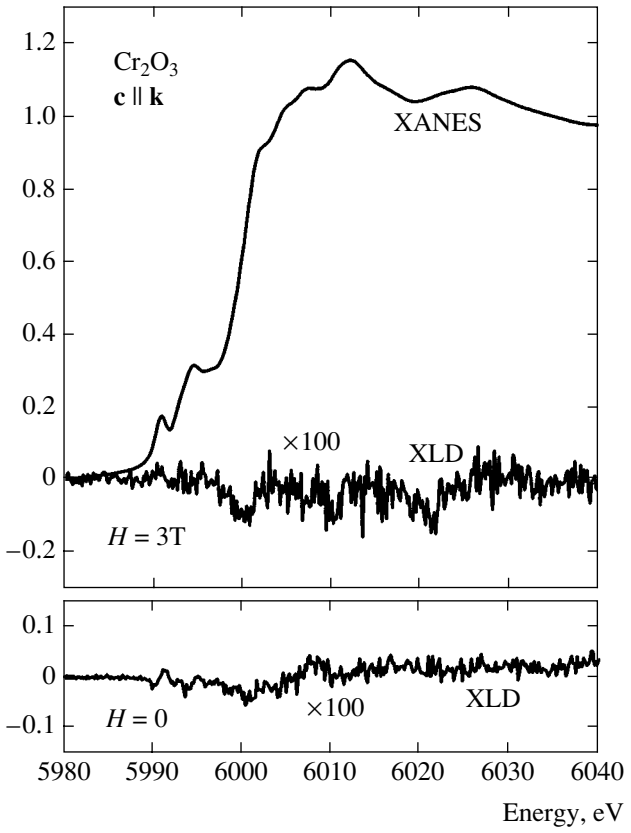

Fig. 6. Cr *K*-edge XLD spectra of Cr_2O_3 with and without magnetoelectric annealing using a 3 T magnetic field. The crystal and the geometry of the experiment were the same as for recording XMχD spectra (**c** || **k**). The upper trace reproduces a XANES spectrum (raw data) for comparison.

axial magnetic field can modify magnetic symmetry of the sample and that metamagnetic domains of *3m*' symmetry can (eventually) contaminate the measurements. This is, however, contradicted by the fact that the sample did not exhibit any measurable XMCD spectrum at the Cr *K*-edge. Moreover, a quick inspection of Table 3 immediately shows that the group 3*m*' again admits only the septor, but not the anapole, as an irreducible representation.

It has been known for decades that the magnetic group of Cr_2O_3 changes beyond the critical spin-flop transition [80, 81]. Recent investigations initiated in [82–84] have confirmed that when a strong magnetic field (up to 20 T) is applied along the **c** axis, a toroidal order can be detected that is associated with the spin-flop magnetic group 2'/*m*. Because the critical spin-flop field at 100 K is 5.8 T [85], it is very unlikely that spin-flop domains could develop in a field of 0.5 T. Recent crystal topography experiments carried out with the powerful method of polarized second-harmonic generation have proved that no spin-flop domain can be detected below the critical spin-flop field [85].

There are further experimental data that also concur in ruling out any contribution of spin-flop domains. Unpublished X-ray linear dichroism spectra, e.g., XLD (S_1) spectra, were recorded in the presence of a higher magnetic field (3 T) following a magnetoelectric annealing procedure carried out with a strong electric field (1 kV/cm) in the geometry **E** || **H** || **c** || **k**. Under such experimental conditions, one would expect the hypothetical spin-flop domains to have a stronger weight, with the practical consequence that reciprocal or nonreciprocal XLD signals should become detectable. As illustrated in Fig. 6, we found no conclusive evidence of such a dichroism. For comparison, we have included in Fig. 6 a natural XLD spectrum recorded in the absence of any magnetic field: the goal was to check carefully whether the (possibly) imperfect alignment of the **c** axis with the wave vector **k** can generate any artifactual dichroism. This is clearly not the case. These negative experiments support our view that it is very unlikely that spin-flop domains can contribute to the XMχD experiment performed with a much weaker magnetic field (0.5 T). It is also noteworthy that all diagonal terms of the magnetoelectric tensor are expected to vanish in the 2'/*m* spin-flop phase [50]: no effective magnetoelectric annealing can then occur in the geometry of our XMχD experiment and the separation of domains of opposite time-reversality becomes impossible.

2. There is a reduction of the magnetic symmetry due to some partially unquenched angular momentum that has a different quantization axis than the spins.

There is nothing sacrilegious in envisaging that the orbital part of the magnetoelectric tensor $[a_{\alpha\beta}]_{\text{orb}}$ (which has so far never been measured) can reveal a symmetry reduction with respect to the point group $\bar{3}\,m \otimes \Theta$, which admits $\bar{3}'\,m'$ as a subgroup. We recall that this subgroup corresponds to the highest magnetic symmetry compatible with the chemical cell (measured above T_N) and with the antiferromagnetic spin configuration. As pointed out in [86], the angular momentum unquenched by covalent bonding, with a different quantization axis than the spins, must be a widespread phenomenon in antiferromagnetic solids.

In our case, there can be no anapole component $\mathbf{\Omega}_0$ along the **c** axis without a (small) orbital moment **L** and an orthogonal electric dipole in directions perpendicular to **c** at the Cr sites. We can therefore expect a (small) canting of the total magnetic moment. We here reopen a fairly old debate that started when Foner [87] reported that the parallel magnetic susceptibility of Cr_2O_3 does not drop to zero below 4 K. This led to the active search for a canted structure of the magnetic moments until Silverstein and Jacobs found that Van Vleck susceptibility calculations can explain the residual contribution of $\chi_{\|}$ [88]. We recall that the Van Vleck susceptibility takes into account localized orbital moments consistent

with a Zeeman perturbation restricted to the first and second order,

$$\chi_{\text{orb}}^{ii} = \frac{N\mu_B^2}{3kT}\{\langle\Psi_g|L_i|\Psi_g\rangle\langle\Psi_g|L_i|\Psi_g\rangle\} + \frac{2N\mu_B^2}{3}\sum_{n\neq g}\frac{\langle\Psi_g|L_i|\Psi_n\rangle\langle\Psi_n|L_i|\Psi_g\rangle}{E_n - E_g}, \quad (16)$$

where N is the Avogadro number. In the particular case where the ground state is orbitally nondegenerate, the first term vanishes. This was assumed by Silverstein and Jacobs, who considered a Cr ion in a cubic crystal field with a weak trigonal field. The Van Vleck susceptibility can then only result from the temperature-independent second term, which couples the ground state to higher crystal field levels. Parallel susceptibility measurements refer to the component L_z, but the weak trigonal field splitting of the corundum structure yields an even higher coupling for $L_{x,y}$, resulting in stronger Van Vleck contributions to $\chi_\perp$ [89]. The angular momentum alone cannot yield the orbital anapole moment; we also need electric dipoles. Recently, Muto *et al.* [90] tried to simulate magnetoelectric spectra at optical wavelengths and pointed out that an antisymmetric twist field with trigonal symmetry must be introduced in the microscopic model in order to mix odd-parity orbitals in the stationary states of the system. This antisymmetric twist is clearly essential to produce a nonzero local orbital anapole moment. At this stage, it is tempting to conclude that the symmetry reduction is caused by a substantial admixture of low-lying crystal field levels in the virtual ground state. We do not even require an external field to induce the Zeeman second-order perturbation; the strong exchange field responsible for the antiferromagnetic order and the spin-orbit coupling could play the same role. We recall that the strong local perturbation caused by the deep core hole can obviously also cause such a substantial coupling as predicted in Section 2.1 (see Eq. (3)).

It remains to be proved experimentally, however, that there is no unquenched angular momentum in the ground state of Cr_2O_3; one should also reinvestigate whether some small ordered canting of the magnetic moments associated with angular momentum can occur. Careful neutron diffraction studies [91] failed to detect any large, ordered canting of the magnetic moments, but the authors admitted openly that neutron diffraction cannot disprove models with canting angles less than three degrees. This implies that with a measured spin moment of $2.48\mu_B$ at each Cr site, orthogonal orbital moments as large as $0.13\mu_B$ may not be seen. This leaves ample space for some orbital magnetism involving only the ground state wave functions. The authors of [92] suspected that the covalent character in the Cr–O bonds might involve "a small spin transfer from the Cr ($3d$) orbitals to the O ($2p$) shell," but they noted that the transferred moment is too small to be detected by neutron diffraction. This problem was recently reconsidered in [93] via spherical neutron polarimetry, and it was confirmed that a reduction of $\langle S_z\rangle$ from $2.98\mu_B$ to $2.48\mu_B$ is definitely too large to be explained solely by the Heisenberg "zero-point deviation" (8%) deduced from neutron inelastic scattering measurements in [92]. Using a simple model based on a covalent overlap of the metal $3d(t_{2g})$ orbitals with the oxygen $2p$ orbitals, Brown *et al.* [93] pointed out that the symmetry constraints preclude a net magnetization of the oxygen atoms, and the only effect of covalent mixing is therefore the lowering of the measured moment $\langle S_z\rangle$ on the Cr sites; no change of the accepted $\bar{3}'m'$ magnetic group is required. This is only true if the spin-orbit coupling can be neglected as discussed below.

More sophisticated unrestricted Hartree–Fock calculations [94, 95] revealed that covalency effects are particularly important in chromium sesquioxide and can explain the well-known differences in the magnetic structures of Cr_2O_3 and Fe_2O_3. Contrary to the model in [93], Dovesi *et al.* [94] found a large splitting between the spin-up (t_{2g}^{α}) and spin-down (t_{2g}^{β}) states and observed that the covalent electron transfer involves the substantial contribution of the Cr $3d(e_g)$ orbitals in the ground state. This result suggests that there should be a significant contribution of the first term of the orbital susceptibility χ_{orb} expressed by Eq. (16). In the general framework of band structure calculations, the orbital part of susceptibility must also include terms, such as those predicted in [96], that have the same source as the temperature-independent Van Vleck susceptibility in localized ions.

A quick inspection of our XMχD spectra convinced us that the $E1E2$ dichroic signal is the most intense for mixed-parity excited states that can be identified as $\{p(\text{O}) + e_g^{\alpha}\}$ and $\{p(\text{O}) + e_g^{\beta}\}$ above the Fermi level in the unrestricted Hartree–Fock calculations in [94]. As a consequence of the crystal field symmetry, there cannot be any net spin moment delocalized on the oxygen atoms. However, the calculations produce clear evidence of a local polarization of each oxygen atom: the part of the electron cloud facing Cr_1 (α) is β polarized, while that facing Cr_2 (β) is α polarized, the maximum polarization occurring along the directions of the chemical bonds. Regarding orbital moments possibly associated with the covalent bonding, one should keep in mind that the spin-orbit coupling is expected to lower the crystal field symmetry, especially in the plane perpendicular to the **c** axis. Thus, the calculation in [94] strongly suggests that a small orbital magnetic moment perpendicular to **c** can occur at every chromium site. This is also fully consistent with the observation in [86] that highly aspherical spin densities with zero spatial average are most often associated with nonzero angular momentum distributions.

Dovesi *et al.* [94, 95] reiterated the claim that the magnetic symmetry of the antiferromagnetic phase of Cr_2O_3 is reduced to R3*c* (class 3*m*), which is a subgroup of $R\bar{3}c$. It is not clear from their paper how this claim was justified. It seems that the only magnetic constraint imposed on the calculation was that the difference between the numbers of majority-spin and minority-spin electrons per unit cell $n_\alpha - n_\beta$ was set to zero, while the program was expected to retain only solutions for which two consecutive Cr atoms have α- and β-type net atomic spin densities. Indeed, the group 3*m* would ideally explain our XOA experiments:

1. the magnetic crystal class 3*m* (space group *R*3*c*) admits the anapole as an irreducible representation, which is consistent with the observation of the XMχD spectra;

2. the crystal class 3*m* does not admit the pseudodeviator $\mathbf{W}^{(2,-1)}$ as an irreducible representation, which is consistent with the absence of detectable XMLD (S_1, S_2) spectra [14];

3. the crystal class 3*m* admits the septor $\Gamma^{(3,-1)}$ as an irreducible representation—using the procedure described in Section 2.3, we were able to check that $\langle \mathbf{\Gamma}_0^{(3,-1)} \rangle \neq 0$. From the experiment carried out with the powdered sample, we expect only a small contribution of this septor term to the XMχD spectra. On the other hand, it is easy to check that $\langle \mathbf{\Gamma}_{\pm 2}^{(3,-1)} \rangle = 0$.

Unfortunately, the magnetoelectric group 3*m* is definitely incompatible with all published magnetoelectric susceptibility measurements, including the magnetoelectric annealing procedure that we used, because it is easy to verify that the generic magnetoelectric tensor of this group has no diagonal term [50].

The point raised by Dovesi *et al.* that a structural change could occur below T_N would be consistent with the observation reported by several authors long ago that the lattice parameters change quite significantly below T_N [97]. Unfortunately, very high quality crystal structure data are required to refine the true magnetic space group. If we trust the interpretation that our XMχD spectra imply a reduction of magnetic symmetry below T_N, then we must seek a magnetoelectric group consistent with both XOA and the well-established magnetoelectric susceptibility measurements. The only magnetic groups that can reconcile these two experiments are $\bar{3}'$ and 3: this is because their generic magnetoelectric tensors simultaneously have the same diagonal terms as the group $\bar{3}'m'$ and the same off-diagonal terms as the group 3*m* [50].

We note that only group $\bar{3}'$ is suitable for an antiferromagnetic solid, whereas group 3 would imply that the system is ferromagnetic, which is not the case. Similarly, a very important observation [39] is that the existence of a magnetoelectric toroidal group requires that in the high-temperature paramagnetic phase, the compound must belong to one of the eight ordinary groups:

$$m\!mm,\ 4/m\!mm,\ \bar{3}\,m,\ \bar{6}\,m2,\ 6/m\!mm,\ m3,\ \bar{4}\,3m,\ m3m.$$

Therefore, as far as the corundum point group $\bar{3}\,m$ is concerned, the only antiferromagnetic toroidal subgroups that deserve attention are $\bar{3}$, 3*m*, and $\bar{3}'\,m$, which are all subgroups of $\bar{3}'\,m'$. We note that only $\bar{3}'$ has a magnetoelectric tensor with diagonal elements. This was the basic argument that led us to propose this group as the true magnetic group describing the spin and orbital magnetoelectric effects in Cr_2O_3 [14]. According to Table 3, magnetic group $\bar{3}'$ must admit irreducible representations of the type $\mathbf{W}_\beta^{(2,-1)}$. Using Eq. (6), we find that $\mathbf{W}_{\pm 2}^{(2,-1)} = 0$ in our experimental configuration (**H** ∥ **E** ∥ **c** ∥ **k**). This is fully consistent with the fact that we failed to detect any nonreciprocal XMLD signal in this geometry. As discussed in Section 3.3.2, a nonreciprocal dichroism XMLD (S_2) might, however, be detected if the wave vector **k** is set perpendicular to the **c** axis, the annealing being still performed with **H** ∥ **E** ∥ **c**. Unfortunately, no experiment has yet been performed in this geometry.

4.1.2. Vanadium sesquioxides $(V_{1-x}Cr_x)_2O_3$. The magnetic structure of chromium-doped vanadium sesquioxides $(V_{1-x}Cr_x)_2O_3$ in the so-called antiferromagnetic "insulating" low-temperature phase is another controversial subject. It dates back to 1980, when Word *et al.* [98] reported a careful neutron diffraction study on pure vanadium sesquioxide (karelianite). They confirmed that in the monoclinic antiferromagnetic insulating phase, the crystal has a distorted *I*2/*a* symmetry and that the vanadium atoms carry a magnetic moment approximately given by $1.2\mu_B$, tilted away from the trigonal **c** axis by 71° and perpendicular to the **a** axis. However, the observation of a forbidden reflection for $l = 6h + 3$ [99] led them to envisage that the magnetic group might not be $2/m \otimes \Theta$ as is usually accepted, but rather a low-symmetry group 2 [98]. They tentatively explained this symmetry reduction by a small magnetic contribution of the oxygen lattice [99]. This puzzling observation was nevertheless considered a "minor issue" even though it was admitted by Moon himself [100] and by von Laar and Yethiraj [101] that a reduction of the magnetic symmetry could be perfectly envisaged. Moon explicitly mentioned in his paper that orbital moments could result in a reduction of symmetry.

The neutron diffraction study in [98] is pertinent here because class 2 is magnetoelectric; according to Table 3, it simultaneously admits $\mathbf{\Omega}^{(1,-1)}$, $\mathbf{W}^{(2,-1)}$, and $\mathbf{\Gamma}^{(3,-1)}$ as irreducible representations, and the measurement of nonreciprocal XMLD (S_1, S_2) spectra must then be allowed by symmetry. Using a crystal of chro-

mium-doped vanadium sesquioxide, i.e., $(V_{1-x}Cr_x)_2O_3$ with $x = 0.028$, we observed in the monoclinic low-temperature phase what is still believed to be the first example of a nonreciprocal X-ray magnetic linear dichroism [15]. The crystal borrowed from Paolasini was initially assumed to have been cleaved perpendicularly to the hexagonal **c** axis, but it was realized recently that it was slightly miscut. Thus, the nonreciprocal XMLD (S_1) spectra reproduced in Fig. 7 were recorded after a magnetoelectric annealing process conducted in the geometry **E** || **H** || **k** with **c** tilted away from **k** by approximately 10°. Because the signal was found to change its sign when the annealing was performed with parallel or antiparallel electric or magnetic fields and to vanish above the Néel temperature $T_N = 181$ K, we feel that there is very little doubt left regarding the nonreciprocal character of this signal. We emphasize that the orientations of the crystallographic axes **a** and **b** were unfortunately unknown in this experiment: this makes it impossible to clarify whether the nonreciprocal dichroism that was measured is to be interpreted as the Jones dichroism XMLD (S_2) associated with the effective operator $[W_{bb}^{(2,-1)} - W_{aa}^{(2,-1)}]$ or as a true dichroism of the type XMLD (S_1) associated with symmetric off-diagonal terms $[W_{ab}^{(2,-1)} + W_{ba}^{(2,-1)}]$. We note that we refer here to the crystal axes and not to the laboratory frame. Clearly, future experiments of this type would call for a detailed (systematic) analysis of the angular dependence of the signal with respect to 2ψ, even though one can anticipate that such experiments should be very demanding in terms of beam time allocation. Moreover, because no experiment has yet been performed with a powdered sample, no indications are available as to whether the septor terms make any significant contribution.

We note that the nonreciprocal XMLD signal measured in $(V_{1-x}Cr_x)_2O_3$ and the nonreciprocal XMχD signal of Cr_2O_3 are of approximately the same order of magnitude, the nonreciprocal XMLD signal being perhaps slightly less intense. We insist that several reasons make it impossible to interpret the spectra reproduced in Fig. 7 as classical magnetooptical (reciprocal) XMLD spectra [48]: (i) a nonreciprocal dichroism changes its sign when the magnetic field is reversed, while this is not the case for the magnetooptical XMLD spectra; (ii) in the experiment illustrated with Fig. 7, the magnetic field was oriented along the direction of the wavevector **k**, whereas the magnetic field is typically set perpendicular to **k** in magneto-optical experiments; (iii) the intensity of our nonreciprocal XMLD signal exceeds (by one order of magnitude at least) the highest intensity that one would expect for a reciprocal, Θ-even XMLD signal. Everyone who has tried to measure a reciprocal XMLD signal at a *K*-edge would agree with us that this is always a very challenging experiment.

As pointed out in [15], a careful examination of the spectra reproduced in Fig. 7 reveals that there is unambiguously a weak dichroism contribution that does not change its sign when the magnetic field is reversed. Our interpretation is that this residual reciprocal signal can result either from the (small) monoclinic distortion and/or from the fact that the crystal was slightly miscut.

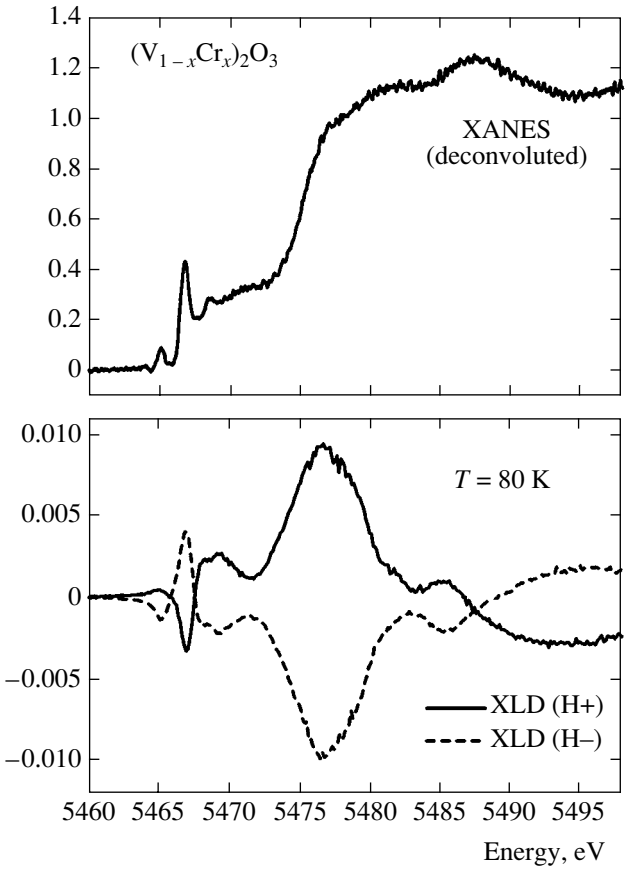

Fig. 7. V *K*-edge nonreciprocal XMLD spectra of $(V_{1-x}Cr_x)_2O_3$ recorded in the monoclinic antiferromagnetic insulating phase below T_N after magnetoelectric annealing performed with either *parallel* (+) or *antiparallel* (–) electric and magnetic fields in the geometry (**c** || **k** || **E** || **H**). The differential absorption spectra refer to the domains grown under the condition of time-reversality after magnetoelectric annealing. The upper trace reproduces a high-energy resolution (deconvoluted) XANES spectrum.

As in the case of Cr_2O_3, the dichroism intensity seems to be most intense for the final states of mixed parity $3d(e_g) + O(p)$. Dovesi *et al.* [95] also performed unrestricted Hartree–Fock calculations on V_2O_3, but their calculations were unfortunately conducted with the high-temperature corundum structure of V_2O_3 and still with the previous magnetic group R3*c*. It would be desirable to reproduce such calculations with the distorted monoclinic structure $I2/a$ that is widely accepted for the low-temperature anti-ferromagnetic phase.

In a recent theoretical study of V_2O_3, Di Matteo *et al.* [102] have identified two magnetoelectric subgroups of $2/m \otimes \Theta$ that can be compatible with the

X-ray diffraction data in [103]: 2/*m*' and 2'/*m*. From the tensor tables in [50], it immediately follows that the generic magnetoelectric tensor of the group 2'/*m* has no diagonal terms and cannot give any dichroism of the type XMLD (*S*2). Moreover, because $[a_{cc}] = 0$, no magnetoelectric annealing is possible in our experimental configuration. In contrast, $[a_{cc}] \neq 0$ for the group 2/*m*', which looks like the ideal choice for nonreciprocal XOA experiments in our experimental configuration because this magnetic group admits the anapole as an irreducible representation along the **c** axis, while

$$[W_{bb}^{(2,-1)} - W_{aa}^{(2,-1)}] \neq 0$$

and

$$[W_{ab}^{(2,-1)} - W_{ba}^{(2,-1)}] \neq 0;$$

it is also easy to verify that

$$[\Gamma_{bbc}^{(3,-1)} - \Gamma_{aac}^{(3,-1)}] \neq 0$$

and

$$[\Gamma_{abc}^{(3,-1)} - \Gamma_{bac}^{(3,-1)}] \neq 0.$$

As in the case of Cr_2O_3, there are several indications suggesting that orbital magnetism should also exist in V_2O_3. We would like to draw attention to the experimental fact that the parallel magnetic susceptibility does not drop to zero at low temperatures for V_2O_3 and Cr_2O_3 [104, 105]. This was again interpreted as the signature of a temperature-independent Van Vleck orbital magnetism. Very recently, Tanaka developed an interesting model [106] according to which each vanadium ion with $S = 1$ also has an orbital magnetic moment approximately given by $0.7\mu_B$; it was even suggested in [106] that these orbital moments can be slightly tilted away from the plane of the antiferromagnetic spin lattice, with the practical consequence that the $2/m \otimes \Theta$ symmetry is broken, thus making the low-temperature phase magnetoelectric. This would be consistent with the observation of a nonreciprocal XMLD spectrum if we additionally admit that there is locally some ordered electric dipole. Precisely this was recently considered in [27], where it was suggested that some cooperative Jahn–Teller distortion occurring at the monoclinic phase transition would also tilt the electric moments. We note that the development of an antiferroelectric order is compatible with the magnetoelectric group 2/*m*' and can possibly explain the highly destructive character of the phase transition for single crystals of any size. Indeed, as discussed in the previous subsection, there is still the risk that due to the core hole perturbation, the cross terms in Eq. (3) allow probing some pseudoground state of artificially reduced symmetry because mixes the true ground state with low crystal field levels.

Recently, Di Matteo and Jansen [107] reported that they failed to measure any magnetoelectric susceptibility using the same single crystal as the one used in our nonreciprocal XMLD experiment, and they immediately questioned our interpretation. They also doubted the efficiency of the annealing process in our experiment by alleging that the conductivity of the $(V_{1-x}Cr_x)_2O_3$ crystal would exceed the conductivity losses of Cr_2O_3 by 15 orders of magnitude. According to our own tests, this figure is erroneously excessive. At the Néel temperature T_N, the conductivity of $(V_{1-x}Cr_x)_2O_3$ crystal (approximately 3×10^3 Ω cm [108]) was estimated to be five orders of magnitude higher than the measured conductivity (0.3 GΩ cm) of our Cr_2O_3 crystal. Under such conditions, the calculated dielectric relaxation time $\tau_R \approx 0.36$ ms (to be compared with τ_R & 36 s for Cr_2O_3) still seems compatible with the fast microscopic dynamics of the magnetoelectric annealing process, as long as one accepts a low leakage current (<10 μA)) at the polarizing electrodes in order to evacuate the accumulated charges.[1] It seems to us that the static magnetic field method apparently used in [107] to measure the magnetoelectric susceptibility of this chromium-doped vanadium sesquioxide crystal is totally inappropriate for systems that have rather large conductivity losses as explained in classical textbooks on magnetoelectric media [51]; this is precisely why pulse methods or methods exploiting magnetic fields modulated at a very high frequency were developed by several groups in the late 1960s, in particular by Al'shin and Astrov, who used an alternating magnetic field at a frequency of 4 MHz. Thus, due to the conductivity losses of the crystal, the failure of the experiments reported in [107] is not unexpected, but the inadequacy of the experimental method does not allow them to draw any conclusion regarding the questioned magnetoelectric nature of this $(V_{1-x}Cr_x)_2O_3$ crystal in the low-temperature monoclinic phase.

Anyhow, comparison of XOA experiments with magnetoelectric susceptibility measurements is not straightforward, as is illustrated by the following differences.

1. Nonreciprocal XOA probes only the orbital part of some average, spineless, one-electron magnetoelectric tensor. In contrast, macroscopic magnetoelectric susceptibility measurements have been discussed up to now essentially by considering, in the first place, how the spins are supposed to be ordered in a given low-temperature phase. Nothing is really known, however, regarding the relative contributions of the spin and orbital currents in such a magnetoelectric solid, and it is not even clear whether magnetoelectric susceptibility measurements would be sensitive enough to detect the contribution of orbital currents. One can easily imagine

[1] F. de Bergevin drew our attention to this important point. This led us to check the reality of the low leakage current which we had neglected in our reports.

a situation where some terms of the magnetoelectric tensor have a purely orbital origin or a vanishingly small spin contribution: in this case, the standard magnetoelectric susceptibility measurement can possibly fail and lead to erroneous conclusions. One may also envisage the converse case of magnetoelectric solids, where the orbital part of the magnetoelectric tensor is partially quenched: there might exist geometries under which no XOA can be detected, even though the standard magnetoelectric susceptibility measurements allow one to expect a signal.

2. Nonreciprocal XOA yields local, element-selective information that cannot be obtained by conventional magnetoelectric susceptibility measurements. This could be turned to formidable advantage if several absorption edges can be probed selectively. This advantage has a counterpart, however: the perturbation induced by the deep core hole might jeopardize the possibility of drawing firm conclusions regarding the magnetic symmetry of the true ground state as a consequence of Eq. (3).

3. Macroscopic magnetoelectric susceptibility measurements require the use of intense electric or magnetic fields. In contrast, nonreciprocal XOA experiments per se do not require any electric or magnetic field and are inherently insensible to the conductivity losses of the sample. In the experiment discussed in this section, a magnetoelectric annealing process was used only to create remanent magnetoelectric states of opposite time-reversality. Other types of annealing could possibly produce the same result; for example, galvanomagnetic annealing or simply the magnetic annealing could suffice under proper symmetry conditions. Figure 1 is a typical example where the action of the magnetic field on a spin anapole induces a local electric polarization and can induce the nucleation of an antiferroelectric order, without applying any electric field. The existence of an orbital anapole could possibly play the same role.

4.2. Effective Operators and Cross Densities of States

4.2.1. Applications of the XOA sum rules. In this subsection, we wish to report on our first attempts to use Carra–Jerez–Marri equations (7)–(9) in order to derive the expectation values of the relevant *E*1*E*2 effective operators. It is instructive to first compare some practical details concerning the XOA sum rules and the XMCD sum rules in the soft X-ray range [24, 110].

1. Renormalizing the XOA dichroism spectra against the XANES spectra cannot exempt us from calculating the two radial integrals $R_l^{(1,2)}$ numerically. We found that this can be most conveniently done with the so-called FDMNES code [111], because we could easily check that these integrals are nearly constant over the energy range selected for the integration.

2. The XOA sum rules do not introduce any renormalization with respect to the number of holes in the band that accepts the photoelectron, as this is typically the case with the XMCD sum rules.

3. In establishing the sum rules, we implicitly considered transitions between atomic multiplets of pure configurations, with l_c, l, and l' being well-identified quantum numbers. This may restrict Eqs. (7)–(9) to *E*1*E*2 transitions towards partially filled, localized bands of finite width,

$$\Delta E = E_{\text{cutoff}} - E_{\text{Fermi}}.$$

There is some ambiguity regarding the definition of E_{cutoff}, however. It is rather unclear whether one should set the cutoff energy at the inflection point of the edge spectrum or beyond the most intense signatures of the dichroism spectra, i.e., slightly above the absorption edge. In order to warrant the numerical stability of the calculations, we were led to systematically set E_{cutoff} above the edge, but this is rather questionable when strong shape resonances of chiral-EXAFS signatures contribute to the experimental spectra.

4. In our opinion, the most serious difficulty is still of an experimental nature and concerns the extreme sensitivity of the sum rules to baseline distortions that may be caused by instabilities of the X-ray beam or by radiation damage to the sample. These problems can hardly be avoided over long data-acquisition times. Error bars therefore strongly depend on the amplitude of the measured dichroism.

In Table 7, we have regrouped the expectation values of the effective operators that were extracted from our nonreciprocal XOA experiments using Eqs. (8) and (9). Two calculations were carried out systematically. In the first calculation, we assumed that the contribution of the septor $\Gamma_{0,\pm2}^{(3,-1)}$ can be neglected; in the second, we assumed that the whole dichroism is entirely due to this septor term. As discussed in the previous sections, the second assumption is highly improbable regarding the measured XMχD spectra of Cr_2O_3; this is why the corresponding result is only quoted in parentheses. For the nonreciprocal XMLD experiments, the situation is more ambiguous due to the lack of information regarding the exact orientation of the crystal. Under the present conditions, the only option is to refer to the laboratory frame, and therefore the relevant effective operator is to be written as $[W_{XY}^{(2,-1)} + W_{YX}^{(2,-1)}]$. By analogy with the previous case, one may suppose that the contribution of the septor terms must be negligible.

It appears clearly from Table 7 that the expectation value of the anapole moment is rather small for Cr_2O_3. Since we missed any pertinent reference for comparison, we tried to convert the calculated orbital anapole

Table 7. Expectation values of the nonreciprocal XOA operators

Compound	Cr *K*-edge Cr_2O_3		V *K*-edge V_2O_3 [H^+]		V *K*-edge V_2O_3 [H^-]	
Effective operator	Ω_z	$\Gamma_z^{(3)}$	$W_{XY}^{(2)} + W_{YX}^{(2)}$	$\Gamma_{YYZ}^{(3)} - \Gamma_{XXZ}^{(3)}$	$W_{XY}^{(2)} + W_{YX}^{(2)}$	$\Gamma_{YYZ}^{(3)} - \Gamma_{XXZ}^{(3)}$
atom. units	0.03	(−0.03)	−0.84	(−0.48)	+0.90	(+0.52)
$R^{(1)}$ rad. integral	-8.21×10^{-5}	-8.21×10^{-5}	-9.47×10^{-5}	-9.47×10^{-5}	-9.47×10^{-5}	-9.47×10^{-5}
$R^{(2)}$ rad. integral	-7.62×10^{-6}	-7.62×10^{-6}	-1.03×10^{-5}	-1.03×10^{-5}	-1.03×10^{-5}	-1.03×10^{-5}

moment into an average toroidal dipole moment per unit cell using the relation

$$\mathbf{M}_{td}^{\text{orb}} = \frac{N\langle \mathbf{\Omega}_0 \rangle}{6} = 0.02\mu_B a_0,$$

where N is here the number of Cr atoms per unit cell and a_0 is the Bohr radius. It then becomes immediately obvious that $\mathbf{M}_{td}^{\text{orb}}$ is several orders of magnitude smaller than the spin toroidal dipole moment

$$\mathbf{M}_{td}^{\text{spin}} = 45\mu_B a_0$$

that was reported recently for the magnetoelectric crystal $Ga_{2-x}FeO_3$ [109]. If this comparison makes sense, it would leave virtually no hope for extracting the orbital part of the magnetoelectric tensor $[a_{\alpha\beta}]_{\text{orb}}$ from magnetoelectric susceptibility measurements, because such measurements are not sufficiently accurate at present. Interestingly, the values quoted for $[\langle W_{XY}^{(2,-1)} \rangle + \langle W_{YX}^{(2,-1)} \rangle]$ in the antiferromagnetic insulating phase of the $(V_{1-x}Cr_x)_2O_3$ crystal are one order of magnitude larger. As expected, the sign is reversed for magnetoelectric domains of opposite time-reversality. We recall that in our nonreciprocal XMLD experiments, we essentially measure a linear combination of $[W_{bb}^{(2,-1)} - W_{aa}^{(2,-1)}]$ and $[W_{ab}^{(2,-1)} + W_{ba}^{(2,-1)}]$, whereas in XM$\chi$D experiments, one would measure $[W_{ab}^{(2,-1)} - W_{ba}^{(2,-1)}]$. Thus, the preliminary result quoted in Table 7 for

$$[W_{XY}^{(2,-1)} + W_{YX}^{(2,-1)}] \approx \pm 0.8 \text{ a.u.}$$

looks rather consistent with the model proposed in [106], where a rather large ground state orbital moment is predicted for V_2O_3, while this is certainly not true for Cr_2O_3.

To illustrate the case of the XNCD sum rules, we selected the cobalt *K*-edge XNCD spectra of the two enantiomeric complexes

$$\mathbf{1}^{(\pm)} = 2[Co(en)_3Cl_3] \cdot NCl \cdot 6H_2O$$

(see [18]), which were introduced in Section 3.2.1. In Table 8, we also included additional results taken from our XNCD spectra database:

1. the titanium *K*-edge XNCD spectra of two nonenantiomorphous crystals of potassium titanic phosphate, i.e., **3** = $KTiOPO_4$ (space group $Pna2_1$, class $mm2$)—these crystals were cut normal to the conjugated [120] and $[1\bar{2}0]$ directions;

2. the iodine L_1-edge XNCD spectrum of lithium iodate, i.e., **4** = $LiIO_3$ (space group $P6_3$, class 6);

3. the tellurium L_1-edge XNCD spectrum of paratellurite, i.e., **5** = TeO_2 (space group $P4_12_12$, class 422).

As confirmed by Table 8, the pseudodeviators

$$\langle \mathbf{N}^{(2,+1)} \rangle = \langle [\mathbf{L}, \mathbf{\Omega}]^{(2)} \rangle$$

of the enantiomers $\mathbf{1}^{(+)}$ and $\mathbf{1}^{(-)}$ have nearly the same absolute value but opposite signs as anticipated from symmetry. In fact, the complex $\mathbf{1}^{(\pm)}$ turned out to be the most favorable example due to its very strong preedge XNCD signal. For potassium titanic phosphate crystals, one would expect $\langle \mathbf{N}^{(2,+1)} \rangle$ to exhibit inverted signs in the case of XNCD spectra recorded with the wave vector parallel to the [120] and $[1\bar{2}0]$ directions. In practice, the situation is much less favorable because there are two inequivalent Ti sites in the unit cell and we found them to contribute to dichroisms of opposite signs [112]. The XNCD signal measured at the Ti *K*-edge is therefore very weak and the poor signal-to-noise ratio makes it more difficult to exploit the sum rule quantitatively. Nevertheless, the calculated values of $\langle \mathbf{N}^{(2,+1)} \rangle$ have the expected opposite signs and their low magnitudes are consistent with the average of the effective operator over the two inequivalent Ti sites.

4.2.2. Cross densities of states. The requirement that the final states are localized is a severe restriction, especially in the so-called "chiral-EXAFS" regime, which we have explored in the case of TeO_2 [19]. One may thus question whether Eq. (7) is suitable for analyzing the iodine L_1-edge XNCD spectrum of α-$LiIO_3$, because in this particular example, the most intense signatures are clearly located in the continuum [16], i.e., well beyond the intense $2s \longrightarrow 5p$ white line. In the continuum of states, we are convinced that it may be a

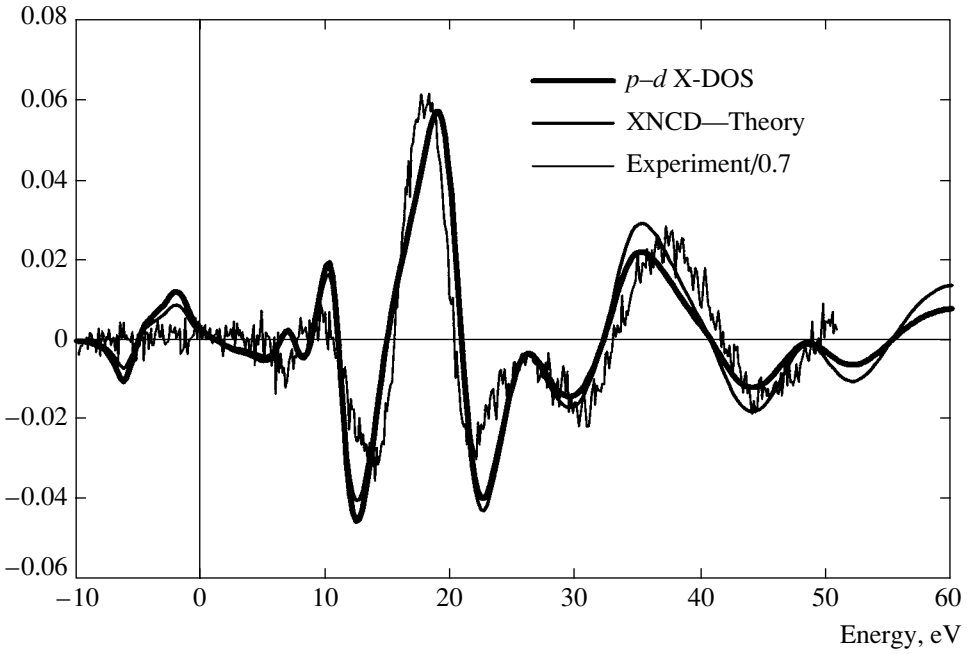

Fig. 8. Comparison of the experimental iodine L_1-XNCD spectrum of $LiIO_3$ with the calculated p–d cross density of states and the simulated XNCD spectrum.

better strategy in the context of XOA to transpose the so-called "differential" formulation of the sum rule, which is now commonly used to analyze the K-edge X-ray magnetic circular dichroism (XMCD) spectra in the so-called "Magnetic-EXAFS" regime [113–115]. Such a differential reformulation of Eq. (2) is given by

$$\frac{\Delta\sigma(E)}{E^2} \approx \frac{4\pi^2\alpha}{\hbar c}\sqrt{\frac{2\pi}{45}}\frac{S_3}{S_0}\frac{d}{dE}\left\{\sum_{\beta} Y_2^{\beta *} R_l^{(1)} R_{l'}^{(2)} \gamma(l, l') \times \left\langle \psi_f \middle| \mathbf{N}_{\beta}^{(2,+1)}(l, l') \middle| \psi_f \right\rangle \right\}, \tag{17}$$

where $\mathbf{N}^{(2,+1)}$ can be identified with $[\mathbf{L}, \Omega]^{(2)}$ and $\gamma(l, l')$ is a numerical factor. In this monoelectronic approach, we can define cross densities of states (X-DOS) $\langle \mathbf{N}^{(2,+1)}(E) \rangle$ that are related to the retarded one-electron Green function $G^+(E)$ by

$$\langle \mathbf{N}^{(2,+1)}(E) \rangle = -\frac{1}{\pi} \times \mathrm{Tr}\{\mathbf{N}^{(2,+1)}(l, l')\mathrm{Im}G^+(E)\}. \tag{18}$$

It follows from this definition that these cross densities of states refer to the effective operator of XNCD, i.e.,

$$\mathbf{N}^{(2)} = [\mathbf{L}, \boldsymbol{\Omega}]^{(2)}.$$

According to Eq. (17), experimentally measured XNCD spectra must be directly proportional to the X-DOSs. This is confirmed by Fig. 8, where experimental and simulated iodine L_1-edge XNCD spectra of α-$LiIO_3$ are compared with the (p–d) X-DOS calculated with an LMTO code [116]. The agreement looks

Table 8. Expectation values of the XNCD operators

Compound Absorption edge	$Coen_3$[+] Co K-edge	$Coen_3$[–] Co K-edge	$KTiOPO_4$ [120] Ti K-edge	$KTiOPO_4$ [1$\bar{2}$0] Ti K-edge	$LiIO_3$ I L_1-edge	TeO_2 Te L_1-edge
Effective operator	$\langle[\mathbf{L}, \boldsymbol{\Omega}]^2\rangle$	$\langle[\mathbf{L}, \boldsymbol{\Omega}]^2\rangle$	$\langle[\mathbf{L}, \boldsymbol{\Omega}]^2\rangle$	$\langle[\mathbf{L}, \boldsymbol{\Omega}]^2\rangle$	$\langle[\mathbf{L}, \boldsymbol{\Omega}]^2\rangle$	$\langle[\mathbf{L}, \boldsymbol{\Omega}]^2\rangle$
atom. units	+0.424	–0.409	+0.016	–0.011	+0.50	+0.77
$R^{(1)}$ rad. integral	-5.31×10^{-5}	-5.31×10^{-5}	-1.15×10^{-4}	-1.15×10^{-4}	2.33×10^{-5}	2.53×10^{-5}
$R^{(2)}$ rad. integral	-3.96×10^{-6}	-3.96×10^{-6}	-1.27×10^{-5}	-1.27×10^{-5}	-2.16×10^{-6}	-2.57×10^{-6}

very encouraging and clearly stimulates our attempt to extend Eqs. (17), (18) to nonreciprocal optical activity.

We finally note that although the definition of cross densities of states makes no reference to the ground state properties (Ψ_g), this does not mean that the deep core hole has no influence on their calculation.

5. CONCLUSIONS

In conclusion, X-ray optical activity appears as a new, element-specific spectroscopy to study orbital magnetism in parity nonconserving solids. As far as the proposed *E*1*E*2 sum rules may give us access to the true ground state expectation values of magnetoelectric orbital operators, nonreciprocal XOA might reveal hidden space–time symmetry properties in magnetoelectric crystals, because XOA probes only the weak orbital part of a monoelectronic magnetoelectric tensor, whereas it is extremely difficult to disentangle the orbital part from the spin part in classical magnetoelectric susceptibility measurements. For instance, XOA can reveal a reduction of the magnetic symmetry whenever the partially unquenched angular momentum has a quantization axis different from the one of the spins. We note, however, that the *E*1*E*2 sum rules can yield the expectation values of the pertinent parity-mixing operators only for a pseudo ground state. Due to the strong perturbation caused by the deep core hole, we cannot exclude the contribution of cross terms involving the ground state and low-lying excited states, as is the case with the temperature-independent Van Vleck paramagnetism. In this context, we note that the contribution of Van Vleck paramagnetism to XMCD has been observed very recently at the same ESRF beam line in a paramagnetic insulator (EuF_3) and a paramagnetic metal (Pd) [117].

In this paper, we have clarified which time-reversal-odd effective operator should be responsible for the magnetochiral dichroism XMχD (S_0) and the nonreciprocal linear dichroisms XMLD (S_1, S_2). The comparison of the nonreciprocal dichroism spectra recorded with single crystals or powdered samples has been shown to be particularly helpful in evaluating the relative importance of the higher-order septor $\langle \Gamma^{(3,-1)} \rangle$ terms with respect to the contributions of the orbital anapole $\langle \mathbf{\Omega}^{(1,-1)} \rangle$ or the pseudodeviator $\langle \mathbf{W}^{(2,-1)} \rangle$. In the specific case of Cr_2O_3, there is very little doubt left that the observed magnetochiral dichroism is related to the orbital anapole operator. On the other hand, the orbital toroidal moment ($\mathbf{M}_{td}$) derived from the sum rule was found to be several orders of magnitude smaller than the spin $\mathbf{M}_{td}$ determined independently for a typical magnetoelectric crystal from diffraction data; this result seems to confirm that it would be very difficult to access the orbital part of the magnetoelectric tensor using conventional magnetoelectric susceptibility measurements.

Potential applications of natural XOA in inorganic or bioinorganic chemistry are still heavily impeded by the prerequisite that one should first obtain large-size single crystals of resolved enantiomers in order to be able to record accurate XNCD spectra. We have shown that this difficulty can be circumvented if, for instance, the chiral species is soluble in a liquid crystal phase that can be aligned in a magnetic field. We are still seeking further alternative approaches. The time-even pseudodeviator

$$\langle \mathbf{N}^{(2,+1)} \rangle = \langle [\mathbf{L}, \mathbf{\Omega}]^{(2)} \rangle$$

could be used to study and quantify ligand-induced asymmetry effects that are suspected to play an important role in asymmetric synthesis. More work is underway in order to extend the calculation of cross density of states and make systematic numerical simulations of them possible.

The authors are particularly indebted to E. Katz for drawing their attention to the formulation of the orbital anapole given in [2] and for many stimulating discussions. They wish to thank C. Moise and D. Perly (Laboratoire de Synthèse & Electrosynthèse Organométalliques, Université de Bourgogne, Dijon), who carried out the synthesis and purification of the chiral compound **2**. One of us (Ch. B.) is grateful to S. Di Matteo and C.R. Natoli for communicating a preprint before publication. This work was supported in part by INTAS (grant no. 01-822) and IPGP (grant no. 1921).

APPENDIX

The *E*1*E*2 absorption cross section was given in Eq. (4) as a product of spherical tensors $\mathbf{T}_\beta^{(b,\theta)}$ describing the X-rays and $\sigma_\beta^{(b,\theta)}$ describing the sample. Here, we give the relation between the sample spherical tensors $\sigma_\beta^{(b,\theta)}$ and the sample Cartesian tensor

$$A_{lmn} + iA'_{lmn} = 4\pi^2 \alpha \hbar \omega k i \sum_f \langle \psi_g | r_l | \psi_f \rangle \langle \psi_f | r_m r_n | \psi_g \rangle \times \delta(E_f - E_g - \hbar\omega), \quad \text{(A.1)}$$

where A_{lmn} and A'_{lmn} are real. With this relation, the tables given in [49] can be used to determine the form of the sample spherical tensors as a function of the magnetic point group. From the definition of the sample Cartesian tensor, it is clear that $A_{lmn} = A_{lnm}$. This property must therefore be added when using the tables in [49]. The sample Cartesian tensor is parity-odd. To investigate its transformation under time-reversal sym-

metry, we replace $|\psi_f\rangle$ and $|\psi_g\rangle$ with $|\Theta\psi_f\rangle$ and $|\Theta\psi_g\rangle$ in Eq. (A.1), which gives

$$\Theta(A_{lmn} + iA'_{lmn}) = -(A_{lmn} + iA'_{lmn})^*.$$

Therefore, A_{lmn} is time-reversal odd and A'_{lmn} is time-reversal even. The relation between spherical and Cartesian tensors is given by the following formulas:

$$\sigma_{\pm1}^{(1,-1)} = \pm\frac{1}{\sqrt{30}}(2A_{xxx} \pm 3iA_{xxy} - A_{xyy} - A_{xzz} \mp iA_{yxx} + 3A_{yxy} \pm 2iA_{yyy} \mp iA_{yzz} + 3A_{zxz} \pm 3iA_{zyz}), \quad (A.2)$$

$$\sigma_0^{(1,-1)} = -\frac{1}{\sqrt{15}}(3A_{xxz} + 3A_{yyz} - A_{zxx} - A_{zyy} + 2A_{zzz});$$

$$\sigma_{\pm2}^{(2,-1)} = \mp\frac{1}{\sqrt{6}}(A_{xxz} \pm iA_{xyz} \pm iA_{yxz} - A_{yyz} - A_{zxx} \mp 2iA_{zxy} + A_{zyy}),$$

$$\sigma_{\pm1}^{(2,-1)} = \frac{1}{\sqrt{6}}(\pm iA_{xxy} - A_{xyy} + A_{xzz} \mp iA_{yxx} + A_{yxy} \pm iA_{yzz} - A_{zxz} \mp iA_{zyz}), \quad (A.3)$$

$$\sigma_0^{(2,-1)} = i(A_{yxz} - A_{xyz});$$

$$\sigma_{\pm3}^{(3,-1)} = \mp\frac{1}{2\sqrt{2}}(A_{xxx} \pm 2iA_{xxy} - A_{xyy} \pm iA_{yxx} - 2A_{yxy} \mp iA_{yyy}),$$

$$\sigma_{\pm2}^{(3,-1)} = \frac{1}{2\sqrt{3}}(2A_{xxz} \pm 2iA_{xyz} \pm 2iA_{yxz} - 2A_{yyz} + A_{zxx} \pm 2iA_{zxy} - A_{zyy}), \quad (A.4)$$

$$\sigma_{\pm1}^{(3,-1)} = \pm\frac{1}{2\sqrt{30}}(3A_{xxx} \pm 2iA_{xxy} + A_{xyy} - 4A_{xzz} \pm iA_{yxx} + 2A_{yxy} \pm 3iA_{yyy} \mp 4iA_{yzz} - 8A_{zxz} \mp 8iZ_{zyz}),$$

$$\sigma_0^{(3,-1)} = -\frac{1}{\sqrt{10}}(2A_{xxz} + 2A_{yyz} + A_{zxx} + A_{zyy} - 2A_{zzz}).$$

Finally,

$$\sigma_{\pm2}^{(2,+1)} = \frac{1}{\sqrt{6}}(\mp iA'_{xxz} + A'_{xyz} + A'_{yxz} \pm iA'_{yyz} \pm iA'_{zxx} - 2A'_{zxy} \mp iA'_{zyy}),$$

$$\sigma_{\pm1}^{(2,+1)} = \mp\frac{1}{\sqrt{6}}(A'_{xxy} \pm iA'_{xyy} \mp iA'_{xzz} - A'_{yxx} \mp iA'_{yxy} + A'_{yzz} \pm iA'_{zxz} - A'_{zyz}), \quad (A.5)$$

$$\sigma_0^{(2,+1)} = A'_{xyz} - A'_{yxz}.$$

REFERENCES

1. Ya. B. Zel'dovich, Zh. Éksp. Teor. Fiz. **33**, 1531 (1958) [Sov. Phys. JETP **6**, 1184 (1958)].
2. I. B. Khriplovich, *Parity Non-Conservation in Atomic Phenomena* (Gordon and Breach, New York, 1991).
3. M.-A. Bouchiat and C. Bouchiat, Rep. Prog. Phys. **60**, 1351 (1997).
4. C. S. Wood, S. C. Bennett, D. Cho, *et al.*, Science **275**, 1759 (1997).
5. V. M. Dubovik and A. A. Cheskov, Fiz. Elem. Chastits At. Yadra **5**, 791 (1974) [Sov. J. Part. Nucl. **5**, 318 (1974)].
6. V. M. Dubovik, S. S. Krotov, and V. V. Tugushev, Kristallografiya **32**, 1987 (1987) [Sov. Phys. Crystallogr. **32**, 314 (1987)].
7. V. M. Dubovik and V. V. Tugushev, Phys. Rep. **187**, 145 (1990).
8. V. L. Ginsburg, A. A. Gorbatsevitch, Yu. V. Kopaev, and B. A. Volkov, Solid State Commun. **50**, 339 (1984).
9. I. E. Dzyaloshinskiĭ, Zh. Éksp. Teor. Fiz. **37**, 881 (1960) [Sov. Phys. JETP **10**, 628 (1960)].
10. J. Goulon, C. Goulon-Ginet, A. Rogalev, *et al.*, Eur. Phys. J. B **12**, 373 (1999).
11. A. D. Buckingham, Adv. Chem. Phys. **12**, 107 (1968).
12. L. D. Barron, *Molecular Light Scattering and Optical Activity* (Cambridge Univ. Press, Cambridge, 1982).
13. J. W. Cooper, Phys. Rev. A **47**, 1841 (1993).
14. J. Goulon, A. Rogalev, F. Wilhelm, *et al.*, Phys. Rev. Lett. **88**, 237401 (2002).
15. J. Goulon, C. Goulon-Ginet, A. Rogalev, *et al.*, Phys. Rev. Lett. **85**, 4385 (2000).
16. J. Goulon, C. Goulon-Ginet, A. Rogalev, *et al.*, J. Chem. Phys. **108**, 6394 (1998).
17. L. Alagna, T. Prosperi, S. Turchini, *et al.*, Phys. Rev. Lett. **80**, 4799 (1998).
18. B. Stewart, R. D. Peacock, L. Alagna, *et al.*, J. Am. Chem. Soc. **121**, 10233 (1999).
19. J. Goulon, C. Goulon-Ginet, A. Rogalev, *et al.*, J. Synchrotron Radiat. **7**, 182 (2000).
20. C. R. Natoli, Ch. Brouder, Ph. Sainctavit, *et al.*, Eur. Phys. J. B **4**, 1 (1998).
21. J. A. Schouten, *Tensor Analysis for Physicists* (Clarendon Press, Oxford, 1964).
22. J. Jerphagnon, D. Chemla, and R. Bonneville, Adv. Phys. **27**, 609 (1978).
23. A. J. Starace, Phys. Rev. B **5**, 1773 (1972).
24. B. T. Thole, P. Carra, F. Sette, and G. van der Laan, Phys. Rev. Lett. **68**, 1943 (1992).
25. S. Di Matteo and C. R. Natoli, J. Synchrotron Radiat. **9**, 9 (2002).

26. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum* (World Sci., New York, 1988).
27. P. Carra, A. Jerez, and I. Marri, cond-mat/0104582.
28. J. Luo, G. T. Trammell, and J. P. Hannon, Phys. Rev. Lett. **71**, 287 (1993).
29. W. Ludwig and C. Falter, *Symmetries in Physics* (Springer, Berlin, 1988).
30. E. P. Wigner, *Group Theory* (Academic, New York, 1960).
31. R. G. Sachs, *The Physics of Time Reversal* (Univ. of Chicago Press, Chicago, 1987).
32. Ch. Brouder, J. Phys.: Condens. Matter **2**, 701 (1990).
33. L. C. Biedenharn and J. D. Louck, *The Racah–Wigner Algebra in Quantum Theory* (Addison-Wesley, Reading, 1981).
34. J. Tenenbaum, Proc.–Indian Acad. Sci., Sect. A **64**, 74 (1967).
35. P. Carra and R. Benoist, Phys. Rev. B **62**, R7703 (2000).
36. P. Carra, J. Magn. Magn. Mat. **233**, 8 (2001).
37. P. Carra, A. Jerez, and I. Marri, Phys. Rev. B **67**, 045111 (2003).
38. S. Brasselet and J. Zyss, J. Opt. Soc. Am. B **15**, 257 (1998).
39. E. Ascher, Helv. Phys. Acta **39**, 40 (1966).
40. E. E. Radescu and D. H. Vlad, Phys. Rev. E **57**, 6030 (1998).
41. A. A. Gorbatsevich, Yu. V. Kopaev, and V. V. Tugushev, Zh. Éksp. Teor. Fiz. **85**, 1107 (1983) [Sov. Phys. JETP **58**, 643 (1983)].
42. A. A. Gorbatsevich and Y. Kopaev, Ferroelectrics **161**, 321 (1994).
43. R. R. Lewis, Phys. Rev. A **49**, 3376 (1994).
44. L. D. Landau and E. M. Lifshitz, *Statistical Physics* (MIR, Moscow, 1951).
45. K. Siratori, K. Kohn, and E. Kita, Acta Phys. Pol. A **81**, 431 (1992).
46. P. Carra, H. König, B. T. Thole, and M. Altarelli, Physica B (Amsterdam) **192**, 182 (1993).
47. G. van der Laan, Phys. Rev. B **57**, 112 (1998).
48. G. van der Laan, B. T. Thole, G. A. Sawatzky, *et al.*, Phys. Rev. B **34**, 6529 (1986).
49. R. R. Birss, *Symmetry and Magnetism*, Vol. 3 of *Series on Selected Topics in Solid State Physics*, Ed. by E. P. Wohlfarth (North-Holland, Amsterdam, 1964).
50. J.-P. Rivera, Ferroelectrics **161**, 165 (1994).
51. T. H. O'Dell, *The Electrodynamics of Magneto-Electric Media*, Vol. 11 of *Series on Selected Topics in Solid State Physics*, Ed. by E. P. Wohlfarth (North-Holland, Amsterdam, 1970).
52. E. Ascher, Philos. Mag. **17**, 149 (1968).
53. N. B. Baranova and B. Ya. Zel'dovich, Mol. Phys. **38**, 1085 (1979).
54. G. Wagnière and A. Meier, Chem. Phys. Lett. **93**, 78 (1982).
55. L. D. Barron and J. Vrbancich, Mol. Phys. **51**, 715 (1984).
56. G. L. J. A. Rikken and E. Raupach, Nature **390**, 493 (1997).
57. G. L. J. A. Rikken, E. Raupach, V. Krstic, and S. Roth, Mol. Phys. **100**, 1155 (2002).
58. P. Kleindienst and G. H. Wagnière, Chem. Phys. Lett. **288**, 89 (1998).
59. N. G. Kalugin, P. Kleindienst, and G. H. Wagnière, Chem. Phys. **248**, 105 (1999).
60. S. L. Hou and N. Bloembergen, Phys. Rev. **138A**, 1218 (1965).
61. V. A. Avetisov and V. I. Gol'danskii, Usp. Fiz. Nauk **166**, 873 (1996) [Phys.–Usp. **39**, 819 (1996)].
62. J. L. Bada, Nature **374**, 594 (1995).
63. S. F. Mason and G. E. Tranter, Proc. R. Soc. London, Ser. A **397**, 45 (1985).
64. L. Pasteur, Rev. Sci. **7**, 2 (1884).
65. W. Voigt, Ann. Phys. (Leipzig) **18**, 651 (1905).
66. F. I. Fedorov, Opt. Spektrosk. **6**, 237 (1959).
67. J. Jerphagnon and D. Chemla, J. Chem. Phys. **65**, 1522 (1976).
68. E. B. Graham and R. E. Raab, Philos. Mag. **66**, 269 (1992).
69. J. Goulon, A. Rogalev, *et al.*, private communication.
70. J. Goulon, A. Rogalev, C. Train, *et al.*, private communication.
71. H. Brünner and J. Doppelberger, Bull. Soc. Chim. Belg. **84**, 923 (1975).
72. A. Saupe and G. Englert, Phys. Rev. Lett. **11**, 462 (1963).
73. H. Kelker and R. Hatz, *Handbook of Liquid Crystals* (Chemie, Weinheim, 1980).
74. D. N. Astrov, Zh. Éksp. Teor. Fiz. **40**, 1035 (1961) [Sov. Phys. JETP **13**, 729 (1961)].
75. B. I. Al'shin and D. N. Astrov, Zh. Éksp. Teor. Fiz. **44**, 1195 (1963) [Sov. Phys. JETP **17**, 809 (1963)].
76. T. H. O'Dell, Philos. Mag. **13**, 921 (1966).
77. S. Shtrikman and D. Treves, Phys. Rev. **130**, 986 (1963).
78. R. M. Hornreich, in *Magnetoelectric Interaction Phenomena in Crystals*, Ed. by A. J. Freeman and H. Schmid (Gordon and Breach, London, 1975), p. 211.
79. J. Goulon *et al.*, private communication.
80. S. Foner and M. Hanabusa, J. Appl. Phys. **34**, 1246 (1963).
81. J. Ohtani and S. Kohn, J. Phys. Soc. Jpn. **53**, 3744 (1984).
82. D. V. Belov, G. P. Vorob'ev, A. M. Kadomtseva, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **58**, 603 (1993) [JETP Lett. **58**, 579 (1993)].
83. Yu. Popov, A. M. Kadomtseva, D. V. Belov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 302 (1999) [JETP Lett. **69**, 330 (1999)].
84. S. S. Krotov, A. M. Kadomtseva, Yu. F. Popov, *et al.*, J. Magn. Magn. Mat. **226–230**, 963 (2001).
85. M. Fiebig, D. Fröhlich, and H.-J. Thiele, Phys. Rev. B **54**, R12681 (1996).
86. V. P. Plakthy, Solid State Commun. **79**, 313 (1991).
87. S. Foner, Phys. Rev. **130**, 183 (1963).
88. S. D. Silverstein and I. S. Jacobs, Phys. Rev. Lett. **12**, 670 (1964).

89. W. H. Brumage, C. R. Quade, and C. C. Lin, Phys. Rev. **131**, 949 (1963).
90. M. Muto, Y. Tanabe, T. Izuka-Sakano, and E. Hanamura, Phys. Rev. B **57**, 9586 (1998).
91. L. M. Corliss, J. M. Hastings, R. Nathans, and G. Shirane, J. Appl. Phys. **36**, 1099 (1965).
92. E. J. Samuelsen, M. T. Hutchings, and G. Shirane, Physica (Amsterdam) **48**, 13 (1970).
93. P. J. Brown, J. B. Forsyth, E. Lelièvre-Berna, and F. Tasset, J. Phys.: Condens. Matter **14**, 1957 (2002).
94. M. Catti, G. Sandrone, G. Valerio, and R. Dovesi, J. Phys. Chem. Solids **57**, 1735 (1996).
95. M. Catti, G. Sandrone, and R. Dovesi, Phys. Rev. B **55**, 16122 (1997).
96. R. Kubo and Y. Obata, J. Phys. Soc. Jpn. **11**, 547 (1956).
97. S. Greenwald, Nature **177**, 286 (1956).
98. R. E. Word, S. A. Werner, W. B. Yelon, *et al.*, Phys. Rev. B **23**, 3533 (1981).
99. W. B. Yelon, S. E. Werner, and R. E. Word, J. Appl. Phys. **52**, 2237 (1981).
100. R. M. Moon, Phys. Rev. Lett. **25**, 527 (1970).
101. M. Yethiraj, J. Solid State Chem. **88**, 53 (1990).
102. S. Di Matteo, N. B. Perkins, and C. R. Natoli, Phys. Rev. B **65**, 054413 (2002).
103. L. Paolasini, C. Vettier, F. de Bergevin, *et al.*, Phys. Rev. Lett. **82**, 4719 (1999).
104. P. H. Carr and S. Foner, J. Appl. Phys. **31**, 344S (1960).
105. M. Greenwood, R. W. Mires, and A. R. Smith, J. Chem. Phys. Lett. **54**, 1417 (1971).
106. A. Tanaka, J. Phys. Soc. Jpn. **71**, 1091 (2002).
107. S. Di Matteo and A. G. M. Jansen, Phys. Rev. B **66**, 100402(R) (2002).
108. H. Kuwamoto, J. M. Honig, and J. Appel, Phys. Rev. B **22**, 2626 (1980).
109. Yu. Popov, A. M. Kadomtseva, G. P. Vorob'ev, *et al.*, Zh. Éksp. Teor. Fiz. **114**, 263 (1999) [JETP **87**, 146 (1998)].
110. R. Wu, D. Wang, and A. J. Freeman, Phys. Rev. Lett. **71**, 3581 (1993).
111. Y. Joly, Phys. Rev. B **63**, 125120 (2001).
112. G. Benayoun, PhD Thesis (Univ. Joseph Fourier, Grenoble, 2001).
113. P. Strange, H. Ebert, J. B. Staunton, and B. L. Gyorffy, J. Phys.: Condens. Matter **1**, 2959 (1989).
114. H. Ebert, Rep. Prog. Phys. **59**, 1665 (1996).
115. G. Y. Guo, J. Phys.: Condens. Matter **8**, L747 (1996).
116. O. K. Andersen and O. Jepsen, Phys. Rev. Lett. **53**, 2571 (1984).
117. A. Rogalev, F. Wilhelm, *et al.*, submitted to J. Magn. Magn. Mat.