Il'ya Mikhailovich Lifshits (1917–1982) On the 80th anniversary

Fiz. Nizk. Temp. **23**, 3–4 (January 1997)

[S1063-777X(97)00101-1]

Academician Il'ya Mikhailovich Lifshits was an outstanding theorist and prominent scientist who left a deep imprint in physics as well as in memories of his colleagues. He was an apprentice and friend of L. D. Landau, who outlined the evolution of a number of new trends in modern physics of the condensed state of matter. The list of branches in which I. M. Lifshits obtained fundamental results includes the dynamics of crystal lattice with defects, the electronic theory of metals, the theory of energy spectra of disordered systems, quantum diffusion and macroscopic quantum effects in solids, the diffusion theory of phase transitions, and physics of macromolecules and biopolymers.

Il'ya Mikhailovich is rightfully regarded as the founder of ''fermiology'' since a new approach to the reconstruction of Fermi branches in the energy spectra of metals is associated with him or with his pupils. The terminology borrowed from publications issued from the Kharkov school guided by I. M. Lifshits is being used for the whole epoch in the works on metal physics.

The research in the dynamic theory of real crystals has determined the development of this field in solid state physics in many respects. The Lifshits theory of regular perturbations, which has become a working tool for calculating the vibrational spectra of crystals with local defects, attracted attention of not only physicists, but also mathematicians due to its elegance.

Il'ya Mikhailovich possessed profuse intuition which helped him to find absolutely new trends in physics. Coalescence in solid solutions, the macroscopic theory of twinning in crystals, quantum diffusion in helium crystals, and phase transitions in macromolecules of biopolymers can serve as examples of complex physical phenomena which were explained theoretically in the works by I. M. Lifshits, thus ensuring advances achieved in many laboratories.

 © 1997 American Institute of Physics

The vast scientific erudition, rigorous thinking and strict principles of Il'ya Mikhailovich were combined with openness and benevolence. These features of Il'mekh's character (Il'mekh was his nickname among colleagues and coworkers) deeply impressed everyone who had the pleasure of meeting him. For this reason, the significance and influence of Il'ya Mikhailovich spread far beyond his scientific school. Il'ya Mikhailovich worked in Kharkov for several decades, so that the Kharkov school in the theory of solids is ''manned'' by his pupils (many of them have become prominent scientists themselves).

The outstanding contribution of I. M. Lifshits to low-temperature physics as well as the role played by him in the organization and development of the journal ''Low Temperature Physics'' are difficult to overestimate.

The editorial board of LTP considers that the publication of the special issue of the journal devoted to the memory of I. M. Lifshits is a proper deed to celebrate his 80th anniversary. Taking into account the vast scientific interests of Il'ya Mikhailovich, we sometimes go beyond the traditional scope of topics of the journal to give the opportunity to the authors to mark by a scientific publication the role of I. M. Lifshits in their research work and life.

Editorial Board

Translated by R. S. Wadhwa

Two-dimensional electron gas in a magnetic field and point potentials

S. Gredeskul and Y. Avishai

*Department of Physics, Ben-Gurion University of the Negev, 84105 Beer-Sheva, Israel**

M. Ya. Azbel'

*Raymond and Beverly Sackler School of Physics and Astronomy, Tel-Aviv University, 69978 Ramat Aviv, Israel***

(Submitted July 15, 1996)
Fiz. Nizk. Temp. **23**, 21–35 (January 1997)

This paper presents a brief review of the electron properties in two-dimensional systems which contain zero-range scatterers which are subjected to a magnetic field. The electron spectrum is described for a periodic arrangement of point scatterers and rational magnetic flux per unit cell. Delocalized states on the Landau levels are constructed for the case of positional disorder. The electron localization in a one-dimensional disordered set of scatterers is studied. Application to the study of electron transmission through quantum dots and ballistic channels is reviewed.
© *1997 American Institute of Physics.* [S1063-777X(97)00301-0]

1. INTRODUCTION

Fifty years ago I. M. Lifshits published a paper, entitled ''On the Theory of Degenerate Regular Perturbations. I. Discrete Spectrum'', in the Journal of Experimental and Theoretical Physics. This paper was a starting point of a series of papers in which I. M. Lifshits developed his famous theory of degenerate perturbations and successfully used it in studying the vibrations of disordered lattices. The theory of degenerate perturbations was found to be a very effective tool for investigation of the local modes that appear outside of an initial spectrum. Lifshits emphasized that an ''important class of perturbations, which can be reduced to the form of a degenerate perturbation, corresponds to local perturbations with a very short range''. For an electron described by the Schrödinger equation, the local perturbation corresponds to a zero range or point potential, which was first introduced by E. Fermi[2,3] for three dimensional ($3D$) systems. Today the point potential model is one of the very popular models in solid state physics and nuclear physics (see, e.g., the monographs in Refs. 4 and 5 and the more recent mathematical review papers[6,7]). This model is especially useful for studying the electronic properties of $2D$ systems in a magnetic field. These systems are ideal objects for application of the local perturbation theory because of a very rarefied initial spectrum. It consists of a zero measure set of the infinitely degenerate, discrete Landau levels[8,9] that leave empty almost all of the energy axis for the new states which occur as a result of degenerate perturbation and which are described by the Lifshits theory (see Eqs. (2) and (3) below).

This paper is actually a mini-review dealing with the electron properties of $2D$ systems with point potentials and magnetic field. In Sec. 2 we describe the main points of the Lifshits theory of degenerate perturbations in the modern formulation, the definition of a point potential, and modification of the main equations of the degenerate perturbation theory for the case of an electron moving in an external field and in a field of a set of point potentials. In Sec. 3 we recall the well-known basic results concerning electron dynamics in the presence of a magnetic field, obtain the scattering amplitude of a point potential in a magnetic field, and describe qualitatively the structure of electron spectrum in a magnetic field and in a field of a single point potential and of a set of point potentials. The Bloch-type electron eigenstates in a field of an ordered set of point potentials and in a so-called rational magnetic field are described in Sec. 4, in which the dispersion laws and energy-flux diagram (Hofstadter-type butterfly[10]) are obtained. Two-dimensional systems containing a disordered set of point potentials are considered in Sec. 5. Here we describe the delocalized states on the Landau levels (Subsec. 5.1), the density of states (DOS) for some of the exactly solvable models (Subsec. 5.2), and the electron localization in the field of a one-dimensionally disordered set of point potentials (Subsec. 5.3). In Sec. 6 we deal with the application of point potentials to the theory of mesoscopic objects such as quantum dots (Subsec. 6.1) and ballistic conducting channels (Subsec. 6.2). In Sec. 7 we give a brief list of some of the unsolved problems and conclude this mini-review.

2. DEGENERATE PERTURBATIONS AND POINT POTENTIALS

Let us consider a quantum system described by a Hamiltonian $H+V$. The degenerate perturbation V is defined as[1]

$$V=\sum_j V_j|v_j\rangle\langle v_j|, \tag{1}$$

where $\{|v_j\rangle\}$ is a set of orthogonal and normalized states. The perturbed eigenstates $|\psi\rangle$ with eigenenergies E lying outside the initial spectrum of the unperturbed Hamiltonian H are the sums of scattered waves

$$|\psi\rangle=\sum_j \eta_j G(E)|v_j\rangle, \quad \eta_j=V_j\langle v_j|\psi\rangle,$$

$$G(E)=(E-H)^{-1}. \tag{2}$$

As was shown in Ref. 1, the eigenenergies E and the corresponding sets of coefficients η_j can be found from the set of equations

$$\sum_j \Lambda_{ij}(E)\,\eta_j = 0. \tag{3}$$

Here

$$\Lambda_{ij}(E) = \frac{\delta_{ij}}{T_j(E)} - (1-\delta_{ij})G_{ij}(E),$$

$$G_{ij}(E) = \langle \upsilon_j | G(E) | \upsilon_j \rangle, \tag{4}$$

where $T_j(E)$ is the scattering amplitude

$$\frac{1}{T_j(E)} = \frac{1}{V_j} - G_{jj}(E). \tag{5}$$

Equations (2) and (3) are the principal ingredients in the Lifshits theory of local perturbations. In the Russian scientific literature they are often called the Lifshits equations (see, e.g., Ref. 11). The main point in this theory is the expression of the perturbation in a form (1). Because of this form, the perturbation (1) is the sum of the projection operators, such that the jth operator projects the state $|\psi\rangle$ onto a number η_j [see Eq. (2)]. Thus the perturbed states (2) lying outside the initial spectrum can be mapped onto the discrete set of coefficients η_j, which map the initial Schrödinger equation $(H+V)|\psi\rangle = E|\psi\rangle$ on the Lifshits equations (3). Note that these equations, in spite of their visible simplicity, are complex equations, because all their coefficients contain the spectral parameter (unknown eigenenergy E) in a complicated, nonlinear way.

The Green's operator $G_V(E) = (E - H_V)^{-1}$ of the perturbed Hamiltonian $H_V = H + V$ can be expressed in terms of the Green's operator $G(E)$ of the unperturbed Hamiltonian and the same matrix $\Lambda(E)$ (4)

$$G_V = G + \sum_{i,j} G|\upsilon_i\rangle (\Lambda^{-1})_{ij} \langle \upsilon_j | G. \tag{6}$$

Using expression (5) for the scattering amplitude, we can rewrite the matrix $\Lambda(E)$ [Eq. (4)] in the form

$$\Lambda(E) = A + Q(E),$$

where

$$A_{ij} = \frac{\delta_{ij}}{V_j},$$

and

$$Q_{ij}(E) = -G_{ij}(E). \tag{7}$$

The diagonal matrix A depends on the perturbation (1) only, while the matrix $Q(E)$ reflects the properties of the unperturbed Hamiltonian H.

Let us consider now a single electron which is described by the Schrödinger equation with an unperturbed Hamiltonian H which includes the interaction with an external (e.g., magnetic) field. In this case an important class of perturbations, which can be reduced to the form of degenerate perturbations, i.e., which can be described by equations such as Eqs. (1)–(6), is formed by zero-range limits of finite-range local potentials, which we call point potentials. A $2D$ point potential can be treated as an attractive zero-range potential with a single bound state with a fixed eigenenergy $-E_b$. There are several ways to introduce such a potential. The first one is based on the idea[12] of regarding a point potential as a delta-function with an infinitely small amplitude. More precisely, the Fourier image of a point potential must be equal to some constant within a circle centered at the origin, and zero otherwise. The corresponding limiting procedure (the constant tends to zero and the radius tends to infinity, keeping the scattering data fixed) can then be used. The second approach[5] is based on the Krein theory of self-adjoint extensions.[13] Finally, a point potential can be described also with the help of a direct zero-range limit in coordinate space.[14] The resulting $2D$ point potential represents a certain generalized function which is less singular than the $2D$ Dirac delta function. The character of this singularity is attributable to the logarithmic singularity of the $2D$ Green's function in coordinate representation with coinciding arguments. A detailed exposition of a general theory and numerous applications of point potentials in all three dimensions can be found in Refs. 4 and 5.

The main equations of degenerate perturbations theory, (2)–(4), and (6), remain valid for the case of an electron moving in a field of a point potential. However, some minor modifications must be made (see details, e.g., in Ref. 15). First, instead of the set of states $|\upsilon_j\rangle$ we must now use the set of states $|\mathbf{r}_j\rangle$ localized at the points $\{r_j\}$, where the point potentials are placed. We can therefore write the scattered wave [Eq. (2)] with the eigenenergy E as follows:

$$\psi(r) = \sum_j \eta_j G(\mathbf{r}, \mathbf{r}_j; E), \tag{8}$$

where $G(\mathbf{r},\mathbf{r}';E) = \langle \mathbf{r}|G(E)|\mathbf{r}'\rangle$ is the Green's function of the unperturbed Hamiltonian in coordinate representation. Second, the scattering amplitude in this case is

$$\frac{1}{T_j(E)} = \lim_{\mathbf{r},\mathbf{r}' \to \mathbf{r}_j} [G_0(\mathbf{r},\mathbf{r}';-E_{bj}) - G(\mathbf{r},\mathbf{r}';E)]$$

$$= G_0^{\mathrm{reg}}(\mathbf{r}_j,\mathbf{r}_j;-E_{bj}) - G^{\mathrm{reg}}(\mathbf{r}_j,\mathbf{r}_j;E). \tag{9}$$

Here $G_0^{\mathrm{reg}}(\mathbf{r},\mathbf{r}';E)$ is the Green's function of the free Hamiltonian $H_0 = -\Delta$ (Δ is the $2D$ Laplace operator), E_{bj} is the binding energy of the point potential placed at the point $\mathbf{r_j}$, and finally $G^{\mathrm{reg}}(\mathbf{r}_j,\mathbf{r}_j;E)$ denotes the regular part of the Green's function. The scattering amplitude (9), together with an obvious equality

$$G_{ij}(E) \equiv G^{\mathrm{reg}}(\mathbf{r}_i,\mathbf{r}_j;E) \tag{10}$$

(in fact, only the diagonal elements are needed in the regularization procedure), defines the matrices $\Lambda(E)$ in Eq. (4), and $Q(E)$ in Eq. (7). The latter matrix Q in this case is called the Krein matrix.[6,7,13] The diagonal matrix A has the elements $\delta_{ij} G_0^{\mathrm{reg}}(\mathbf{r}_j,\mathbf{r}_j;-E_{bj})$. Equation (6) for the perturbed Green's function has the form

$$G_V(\mathbf{r},\mathbf{r}';E)=G(\mathbf{r},\mathbf{r}';E)+\sum_{i,j} G(\mathbf{r},\mathbf{r}_i;E)$$

$$\times(\Lambda^{-1})_{ij}G(\mathbf{r}_j,\mathbf{r}';E). \tag{11}$$

Thus, Eqs. (3), (4), and (8)–(11) give a complete description of the electron spectral properties in a system with point potentials and an external field. The central role in such a description is played by the matrix $\Lambda(E)$: the zeros of its determinant are the eigenenergies E_n of the perturbed Hamiltonian $H+V$, its eigenvectors $\{\eta_j(E_n)\}$, which correspond to the zero eigenvalue (3), determine the wave functions (8) with eigenenergies lying outside the spectrum of the unperturbed Hamiltonian H, and finally the inverse matrix $(\Lambda)^{-1}$ determines the Green's function (11) of the perturbed system $H+V$.

3. COMBINATION OF POINT POTENTIALS AND MAGNETIC FIELD

The motion of electron with charge $-e$ and mass M in a $2D$ system, subject to a perpendicular magnetic field $\mathbf{B}=(0,0,-B)$, is described by the Schrödinger equation

$$-(\nabla+2\pi i\varphi\mathbf{A})^2\psi(\mathbf{r})=E\psi(\mathbf{r}). \tag{12}$$

We use the set of units, where $\hbar=2M=1$ and all the lengths are measured in some scale d. In such units all quantities are dimensionless. In Eq. (12) $\mathbf{A}$ is a vector potential and φ is dimensionless flux

$$\varphi=\frac{\mathbf{\Phi}}{\mathbf{\Phi}_0},\quad \mathbf{\Phi}=Bd^2 \tag{13}$$

($\mathbf{\Phi}_0=hc/e$ are the normal flux quanta). In what follows we choose the Landau gauge of a vector potential

$$\mathbf{A}=(y,0,0).$$

The spectrum of the system consists of a set of the Landau levels[8,9]

$$E_{nL}=\left(n+\frac{1}{2}\right)E_L,\quad n=0,1,\ldots, \tag{14}$$

where $E_L=4\pi\varphi$ is the distance between Landau levels. The normalized eigenfunctions are the Landau functions

$$L_{n,k}(x,y)=(2^{n+1}\pi^{3/2}n!\,l)^{-1/2}\mathrm{e}^{-ikx-1/2(y/l-kl)^2}$$

$$\times H_n(y/l-kl),$$

$$n=0,1,\ldots,-\infty<k<\infty, \tag{15}$$

($l=(2\pi\varphi)^{-1/2}$ is dimensionless magnetic length, and H_n is the Hermite polynomial[16]). Another possible set of eigenfunctions consists of the states $L_{n,m}(x,y)$, which coincide with the gauge transformed states with a fixed angular momentum m in the symmetric gauge. For a fixed number n of the Landau level the angular momentum m is an integer and satisfies the inequality $-\infty<m\leqslant n$. The Green's function in coordinate representation is

$$G(\mathbf{r},\mathbf{r}';E)=-\frac{\Gamma(-a)}{4\pi}\,\mathrm{e}^{-i\pi\varphi(x-x')(y+y')}g(\mathbf{r}-\mathbf{r}';E), \tag{16}$$

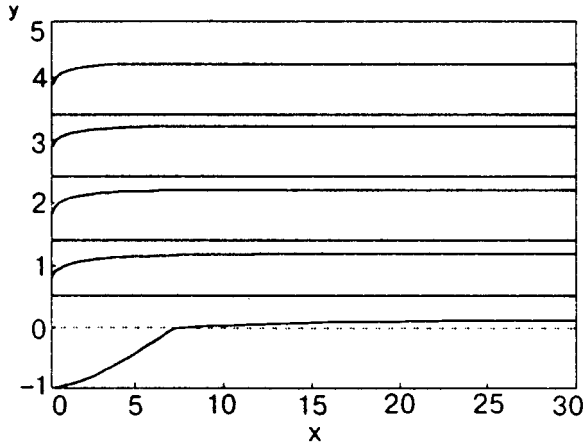

FIG. 1. The s-levels shifted by a single point potential in a magnetic field in the scale $y=E/E_L=a+1/2$ for the positive energies; $y=E/E_b$ for the negative energies; $x=E_L/E_b$. Thus, $y=-1$ corresponds to a bound state without a magnetic field and $y=n+1/2$, $n\geqslant 0$ corresponds to the nth Landau level.

where

$$g(\mathbf{r};E)\equiv F(\xi)=\xi^{-1/2}W_{1/2+a,0}(\xi),\quad \xi=\pi\varphi r^2. \tag{17}$$

Here the gamma Γ and Whittaker W functions[16] are used and a is a new energy parameter

$$a=-\frac{1}{2}+\frac{E}{E_L}.$$

The scattering amplitude of a point potential in a magnetic field in accordance with Eqs. (9) and (16) is

$$\frac{1}{T_j(E)}=-\frac{1}{4\pi}\left(\psi(-a)+\ln\frac{E_L}{E_b}\right., \tag{18}$$

where $\psi(-a)$ is the digamma function.[16] This expression, which was obtained with the help of each one of the three approaches mentioned above,[14,15,17–20] describes now an electron spectrum in a magnetic field and in a field of a single point potential. A point potential scatters only s-states.[21] Therefore, the unperturbed wave functions $L_{n,m}(x,y)$, which correspond in the symmetric gauge to the states with a fixed nonzero angular momentum m, vanish at the origin and therefore are not modified by a point potential. They correspond to the Landau eigenenergies E_{nL} which are zeros of the scattering amplitude (i.e., poles of the digamma function in Eq. (18)). New eigenenergies E_n, which correspond to the scattered s-states, are determined by the poles of the scattering amplitude, Eq. (18). Because of the attractive character of a point potential, the values E_n are shifted down with respect to the Landau levels. Each perturbed s-wave function $\psi_n(\mathbf{r})$ with eigenenergy E_n coincide with the unperturbed Green's function $G(\mathbf{r},0;E_n)$ (16) and therefore has a logarithmic singularity at the origin. The magnetic field dependence of the lowest five shifted eigenenergies is shown[15] in Fig. 1.

Such a picture of the spectrum perturbed by a single point potential, is valid in the ultimate zero-range limit only. The finiteness of the potential radius, $\rho\neq 0$, leads to an additional shift of the eigenenergy E_n and to an additional splitting of the Landau level E_{nL}.[22] For a cylindrically symmetric

potential the states can be classified with respect to the angular momentum m, which they would have in the symmetric gauge. The corresponding shifted eigenenergies $E_{n,m}$ do not depend on the sign of the momentum m,

$$\varepsilon_{n,m} \equiv E_{nL} - E_{n,m} \propto (\rho/l)^{2|m|}/|\ln(\rho^2 E_b)|,$$

and therefore the eigenenergies $E_{n,m}$ with $|m| \leqslant n$, remain twofold degenerated.[23] Note that the finite radius of the potential leads not only to the energy shift, but also modifies the eigenfunctions with $m \neq 0$, so that their intermediate asymptotic relations contain terms proportional to $r^{-|m|}$,

$$\psi_m(\mathbf{r}) \approx r^{|m|} + \frac{r_m^{2|m|}}{r^{|m|}} \quad \rho \ll r,$$

with $r_m \propto \rho/|\ln(\rho^2 E_b)|^{1/2|m|}$ (Ref. 24).

The main qualitative features of the electron spectrum in a magnetic field and in a field of a set of point potentials can be established solely with the help of the fact that the radius of a point potential equals zero. Indeed, consider an electron moving in a field of a set of zero-range potentials which are placed at the points $\{\mathbf{r}_j\}$ and let d be the averaged distance between scatterers (in standard dimensional variables). If the electron wave function vanishes at a point where a scatterer is placed, the electron does not ''feel'' this scatterer. Therefore, if one can construct a linear combination $\psi(\mathbf{r})$ of Landau functions (15) with a fixed Landau level number n, which vanishes at all points $\{\mathbf{r}_j\}$, then this combination will be an exact eigenfunction of an electron in *the presence of this set of point potentials* with the eigenenergy E_{nL} [Eq. (14)], Refs. 22 and 25.

To determine whether such a combination exists, recall that the number of states per scatterer per Landau level for a given density of scatterers equals exactly the dimensionless flux φ [Eq. (13)]. On the other hand, we have exactly one condition $\psi(\mathbf{r})=0$ per scatterer per Landau level which the wave function must satisfy. If the magnetic field B is sufficiently strong, $\varphi>1$, then the condition $\psi(\mathbf{r}_j)=0$ can be satisfied at all points $\{\mathbf{r}_j\}$. This means that only one eigenstate per single scatterer is shifted from each Landau level. The other $\varphi-1$ states remain on the Landau level.[22,25] In a small enough field, $\varphi<1$, all of the states are shifted from the Landau level.

Thus the electron spectrum is naturally divided into two components. The first component consists of the eigenvalues lying outside the Landau levels and is described by the Lifshits equations (3). This component exists in an arbitrary magnetic field. In disordered systems the corresponding eigenstates are localized. (In a recent paper[26] this statement was rigorously proved for a model reduced to the lowest Landau level). The number of these states per point potential per Landau level equals unity for a strong field, $\varphi>1$, and equals φ in the opposite case $\varphi<1$. In the periodic case for the rational flux $\varphi=Q_1/Q$ per unit cell with area d^2, these states are Bloch type. In complete accordance with the group theory predictions,[27,28] they form $\min\{Q_1,Q\}$ dispersive subbands below each Landau level with the number of states per point potential equal Q^{-1} per subband.[17,29,30] Each of these subbands in turn is Q-fold degenerate.

The second component consists of the Landau levels themselves. This component exists only in a sufficiently strong magnetic field $\varphi>1$. The states that realize this component are linear combinations of the Landau functions with a fixed number of levels, which vanish at all points where the point potentials are located. For a spatially uniform (on the average) set of point potentials these states could always be chosen to be delocalized. Indeed, if these states are chosen to be localized, then their localization centers are uniformly distributed. On the other hand, since their eigenenergies are the same, we can rearrange them to be delocalized states.[22] The density of these states is

$$\rho_{\text{sing}}(E) = (\varphi - 1) \sum_n \delta(E - E_{nL}). \tag{19}$$

4. BLOCH STATES OUTSIDE THE LANDAU LEVELS

Let us now consider a set of identical point potentials with binding energy E_b and assume that the positions $\mathbf{r}_j$ form (for simplicity) a square lattice with a unit lattice constant. Thus, $\mathbf{r}_j=(m_j,n_j)$ with integers m_j,n_j. This system represents a particular case of a $2D$ periodic system in a magnetic field. In an irrational magnetic field φ the spectrum of the system is of the devil-staircase type,[31] while for a rational field $\varphi=Q_1/Q$ the spectrum is of a Bloch-type.[27,28,31] In this case two components $T_{(1,0)}$ and $T_{(0,Q)}$ of the magnetic translation operator[27]

$$(T_{\mathbf{R}}\psi)(\mathbf{r}) = \mathrm{e}^{2\pi i \varphi \mathbf{r}\mathbf{A}(\mathbf{R})}\psi(\mathbf{r}+\mathbf{R})$$

($\mathbf{A}$ is the vector potential) commute with one another and with the perturbed Hamiltonian $H+V$. Therefore, the electron wave functions can be choosen as eigenfunctions of all translation operators $T_{\mathbf{R}}$ with $\mathbf{R}=(m,Qn)$ (m and n are integers), with eigenvalues $e^{\mathbf{iqR}}$. The quasimomentum vector $\mathbf{q}=(q_1,q_2)$ is defined in a rectangle $0<q_1$, $Qq_2<2\pi$ (the magnetic Brillouin zone) and the eigenfunctions $\psi_{\mathbf{q}}(\mathbf{r})$ have Bloch-like form

$$\psi_{\mathbf{q}}(\mathbf{r}) = u_{\mathbf{q}}(\mathbf{r})\mathrm{e}^{\mathbf{iqr}}$$

with quasi-Bloch amplitudes

$$u_{\mathbf{q}}(\mathbf{r}+\mathbf{R}) = u_{\mathbf{q}}(\mathbf{r})\mathrm{e}^{-2\pi i \varphi \mathbf{r}\mathbf{a}(\mathbf{R})}.$$

Nevertheless even for rational fluxes the spectrum has a very complicated fractal structure (in spite of simple Bloch-like wave functions as solutions) which is defined by the arithmetical nature of the flux φ.[10,31]

For a field which corresponds to a rational flux, the point character of the periodic potential enables one to study the spectrum of the system in more detail. The first results were obtained in Refs. 7, 17, 29, and 30. Using the symmetry of the problem, the authors expanded the Hamiltonian of the system into a direct integral over possible values of quasimomenta $\mathbf{q}$ and studied the discrete spectra corresponding to each sub-Hamiltonian with a fixed $\mathbf{q}$. An exact expression for the Green's function of the system was obtained and analized for an arbitrary rational flux φ. This enabled us to establish, on the mathematical level of rigor, the main qualitative properties of the spectrum, which were mentioned at the end of the last section.

The next step was carried out in Refs. 19, 20, and 32, where the Lifshits equations (3) in a magnetic field were obtained and studied. Here the symmetry of the equations (3) (after mapping) was used. The matrix $(\Lambda)_{m,n}^{m',n'}$, Eq. (4), depends only on the difference $m-m'$. Therefore, the mapped wave function $\eta_j \equiv \eta_{m,n}(2)$ can be found in the form

$$\eta_{m,n} = e^{iq_1 m}\zeta_n \tag{20}$$

with ζ_n satisfying a one-dimensional difference equation

$$\frac{\ln(E_c/E_b)}{\Gamma(-a)}\,\zeta_{n_0} = \sum_{n=-\infty}^{\infty} F_{n_0,n}\zeta_n\,. \tag{21}$$

Here

$$F_{n_0,n} = \sum_{m=-\infty}^{\infty} F(\xi_{m,n-n_0})e^{im[q_1+\pi\varphi(n+n_0)]},$$

where $F(0) = -\psi(-a)/\Gamma(-a)$, and $F(\xi)$ for $\xi \neq 0$ is defined by Eq. (17) with $\xi_{m,n} = \pi\varphi(m^2+n^2)$. Equation (21) is used below in Subsec. 5.3. Further, the matrix $(\Lambda)_{m,n}^{m',n'}$ is invariant with respect to a simultaneous shift $n \to n+Q$, $n' \to n'+Q$. This results in the final Bloch form of a mapped wave function

$$\eta_{m,n} = e^{i(q_1 m + q_2 n)}c(n)$$

with the Q-periodic discrete Bloch amplitudes $c(n) = c(n+Q)$, which satisfy the equations[19,20,32]

$$\frac{\ln(E_c/E_b)}{\Gamma(-a)}\,c(\lambda) = \sum_{\mu=0}^{Q-1} S_{\lambda\mu}c(\mu), \quad 0 < \lambda, \mu < Q-1. \tag{22}$$

The coefficients in this equation are

$$S_{\lambda\mu} = \sum_{N,m=-\infty}^{\infty} F(\xi_{m,NQ+\mu-\lambda}) \times e^{i[mq_1+(NQ+\mu-\lambda)q_2]+(im\pi Q_1/Q)[NQ+\mu+\lambda]}. \tag{23}$$

In each energy interval between adjacent Landau levels equation (22) has, for a fixed **q**, exactly Q solutions. For a strong field $Q_1 > Q$ these solutions determine Q dispersion laws $E_i(\mathbf{q})$, which correspond to Q sets of coefficients $\eta_{m,n}^{(i)}(\mathbf{q})$, $i = 0,\ldots,Q-1$. These coefficients, together with Eq. (2), determine Q dispersive subbands, which accumulate exactly one state per scatterer per Landau level. Some of the dispersion surfaces are demonstrated in Fig. 2.[35]

For a weak magnetic field $Q_1 < Q$ Eq. (22) can be reduced to a finite set of Q_1 equation,[34] which determine Q_1 dispersive subbands $E_i(\mathbf{q})$, $i = 0,\ldots,Q_1-1$. The other $Q-Q_1$ nontrivial solutions $\eta_{m,n}^{(i)}(\mathbf{q})$, $i = Q_1,\ldots,Q-1$ of Eq. (22), when substituted into Eq. (2), lead to the wave functions $\psi^{(i)}(\mathbf{r})$ which are identically equal to zero.[35] All these results are in complete agreement with the qualitative picture of the spectrum described in the previous section.

Equation (22) is very complicated even for a numerical solution because, as was mentioned above, it is not a standard eigenvalue equation: the energy parameter a appears on both sides of this equation in a nonlinear way. Therefore, let us first discuss different approximate methods of the solution. The dispersive bands are the result of spreading of single scatterer levels discussed in the last section. In the limiting cases of a very strong or very weak scatterer, when the parameter $|\ln(E_L/E_b)|$ is large, these levels are close to the initial Landau levels. In this case one can substitute $E = E_{nL}$ on the right side of Eq. (22) and linearize the left side of this equation with respect to the difference $E - E_{nL}$. Such a procedure significantly simplifies the solution of Eq. (22), reducing it to a standard eigenvalue problem. In Ref. 36, where this approximate scheme was proposed, the linear-

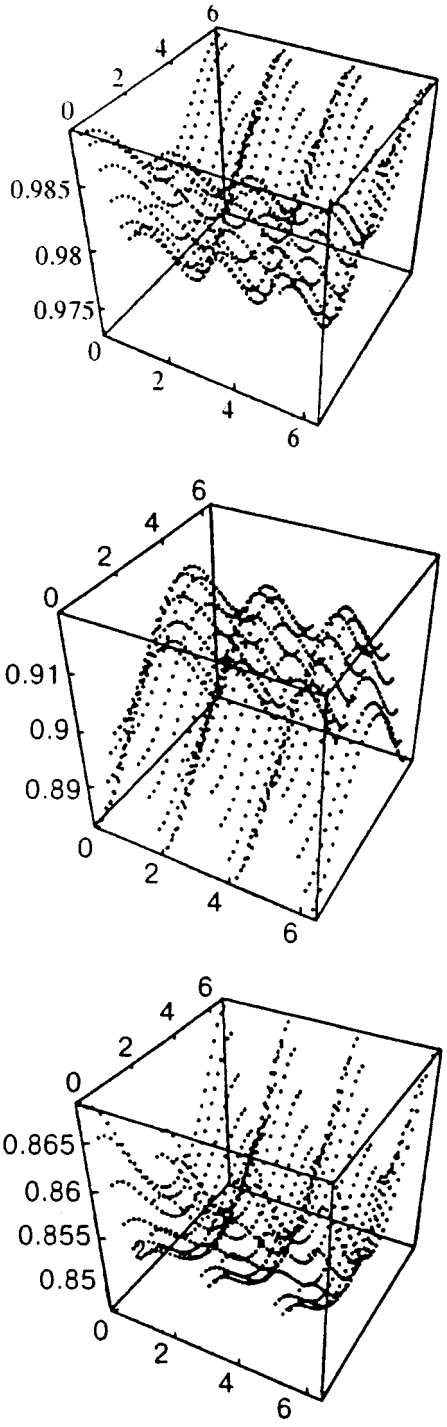

FIG. 2. The dispersion relations $a_i(\mathbf{q}), i-1,2,3$ of the spectrum for flux $\varphi = 5/3$ describing three subbands which are situated between the first two Landau levels.

ized version of Eq. (22) was solved numerically for the energy region lying between the zeroth and the fourth Landau levels and for rational fluxes with denominators $Q \leqslant 9$. All the corresponding dispersion relations were obtained and the energy-flux diagram was constructed. Calculations in a weak magnetic field confirmed the existence of Q_1 dispersive subbands. Numerical results obtained in Ref. 36 indicated that for some rational values of the magnetic field dispersive subbands touch the parent Landau level at some special points $\mathbf{q}$. The condition for touching the nth Landau level is $\varphi \leqslant n+1$. The touching of a given Landau level is evidently related to the symmetry nature of the problem and is not related to the strength of the point potential. Note also that the closer is the energy to the Landau level, the more exact is the approximation applied, so we maintain that numerical results of Ref. 36 concerning this touching are exact. These statements are also used below in subsec. 5.3.

Another approximation can be applied in the strong field limit. Consider Eq. (21) near the nth Landau level. If the Larmor radius is larger than the lattice constant, i.e.,

$$\pi\varphi > n, \tag{24}$$

then, because of an exponential decrease of the Whittaker function [16], one should account only for the nearest neighbors in this equation. Neglecting coupling with non-nearest neighbors, we obtain the well-known Harper equation[37]

$$\zeta_{n+1} + \zeta_{n-1} + 2\zeta_n \cos(2\pi n\varphi + q_1) = \varepsilon\zeta_n, \tag{25}$$

where

$$\varepsilon = \frac{\sqrt{\alpha}}{W_{1/2+a,0}(\alpha)} \frac{\psi(-a) + \ln(E_L/E_b)}{\Gamma(-a)}, \quad \alpha \equiv \pi\varphi, \tag{26}$$

and the single-valued branches of the gamma- and digamma-functions, which correspond to the vicinity of the nth Landau level, are chosen.

It is interesting to note that Harper equation appeared for the first time in the context of electron theory of metals in the opposite limiting case. For a weak magnetic field one can start from the dispersion law $E(\mathbf{p})$ of the system without a magnetic field and use the Peierls substitution, which replaces the quasi-momentum $\mathbf{p}$ by the operator $[(\hbar/i)\nabla - (e/c)\mathbf{A}]$.[38] Then the strong-coupling approximation immediately leads to the Harper equation.[39]

Substituting into Eq. (25) $\zeta_n = \exp(inq_2)c(n,\mathbf{q})$ with the Q-periodic functions $c(n)$, we obtain in the general case $Q \geqslant 3$ (the cases $Q=1,2$ are trivial) the following difference second-order equation:

$$\exp(Qq_2)c(n+1,\mathbf{q}) + \exp(-Qq_2)c(n-1,\mathbf{q}) + 2c(n,\mathbf{q})\cos(2\pi n\varphi + q_1) = \varepsilon c(n,\mathbf{q}).$$

Thus, in the ''ε'' scale the dispersion laws $\varepsilon_i(\mathbf{q})$ and the eigenvectors $c_i(n,\mathbf{q})$ corresponding to the spectral region (24) are *universal* (within the accuracy of the approximation) in the sense that they do not depend either on the Landau level number or on the integer part of the dimensionless flux φ.[40] The integer part of the flux and the concrete position of the eigenenergy between Landau levels enter when one restores the dispersion function $E_i(\mathbf{q})$ from Eq. (26) in the natural energy scale.

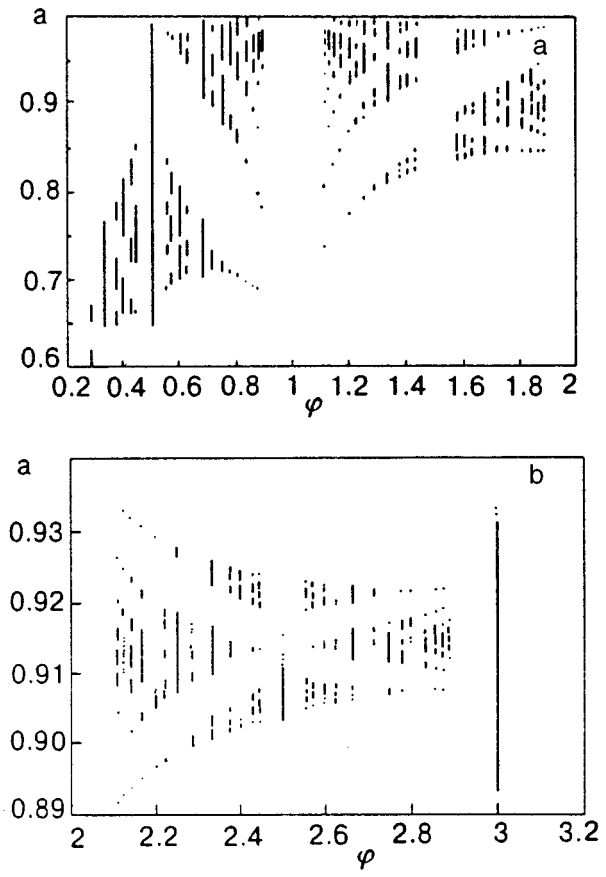

FIG. 3. The energy-flux diagram between the first two Landau levels $0 \leqslant a \leqslant 1$, for the rational fluxes $0 < \varphi < 3$ with denominators up to 9: $0 < \varphi < 2$, where the exact equations are used (a); $2 < \varphi < 3$, where the strong-field approximation (the Harper equation) is used; only diagonal values of the quasi-wave vector $\mathbf{q} = \pi t(1/Q,1)$, $0 < t < 1$, are accounted for (b).

Detailed numerical calculations carried out in Refs. 33 and 40 show that the results of this strong-field approximation are very close to those obtained from the exact equations (22) for all possible values of $\mathbf{q}$ except those corresponding to the touching of the Landau level. This is quite natural since touching is an exact, which cannot be captured by such an approximation. On the contrary, the approximation of very strong or very weak scatterer, discussed in the previous paragraph, becomes exact in the nearest neighborhood of the Landau level. The numerical results obtained with the help of the latter approximation completely coincide in these neighborhoods with the exact ones. In turn, such an approximation fails far away from the Landau levels. Therefore, to obtain correct results one has to use different approximations in different regions of the (E,φ) plane. The part of the energy-flux diagram lying between the zeroth and the first Landau levels is shown in Fig. 3.

5. DISORDERED SET OF POINT POTENTIALS

We now move on and discuss systems containing a disordered set of point potentials. This disorder can be realized

with random positions or strengths of point potentials. Below we will discuss some features of the electron spectrum in such systems: disorder, independent extended states (subsec. 5.1), peculiarities of the DOS (subsec. 5.2), localization properties of a $2D$ system with $1D$ disorder (subsec. 5.3).

5.1. Extended states on the Landau levels

As was mentioned at the end of the previous section, our goal here is to construct a linear combination of Landau functions (15) with a fixed level number n, which vanish on a given set of points (where the point potentials are located). We begin by studying the states on the lowest Landau level E_{0L}. An arbitrary linear combination of Landau functions $L_{0,k}$ has the form

$$\psi_0(\mathbf{r}) = \int_{-\infty}^{\infty} L_{nk}(\mathbf{r}) f_n(k)\, dk. \tag{27}$$

Evidently, this integral equation (27) can be written as[17,41,42]

$$\psi_0(x,y) = \exp(-y^2/2l^2) F_0(z), \tag{28}$$

where $z = x + iy$. The function $F_0(z)$ is defined by

$$F_0(z) = \int_{-\infty}^{\infty} \exp(-ikz - k^2 l^2/2) f_0(k)\, dk.$$

If this integral converges, then it defines an entire function[43] of the complex variable $z = x + iy$. Each entire function is characterized by its order and type. If the modulus of an entire function grows at infinity as $\exp(\tau|z|^{\tau})$, then this function is of order r and of type τ (see exact definition in Ref. 43). We do not, in fact, need the Fourier coefficients $f_0(k)$, since any entire function $F_0(z)$ which, when substituted into Eq. (27), makes $\psi_0(x,y)$ vanish on the sites of the impurities will do. Therefore it is sufficient to construct an entire function $F_0(z)$, which vanishes at all points z_j where the point potentials are located. To this end, let us introduce the Weierstrass product that pertains to the complex sequence $\{z_j\}$.[17,41,42] For a constant density of impurities, the sequence of complex points $\{z_j\}$ is of genus 2 (i.e., the sum of the inverse squares $1/|z_j|^2$ converges), and hence (assuming there is no scatterer at the origin) the Weierstrass product

$$W(z) = \prod_j \left(1 - \frac{z}{z_j}\right) \exp\left(\frac{z}{z_j} + \frac{z^2}{2z_j^2}\right) \tag{29}$$

is well defined and defines an entire function of order $r = 2$. If the distribution of point potentials satisfies some uniformity condition (Lindelöf criterion[43]), then $W(z)$ is also of finite type τ. To control the rate of growth of $\psi_0(x,y)$ on the real axis, we multiply $W(z)$ by an exponent e^{-sz^2} with $s > \tau$. We thus obtain an entire function

$$F_0(z) = e^{-sz^2} W(z) \tag{30}$$

on the order of 2 and of type less equal than $\tau + s$ which falls off with $|x|$ as a Gaussian. Following Eq. (28), it can be easily verified that if the magnetic field is strong enough, i.e.,

$$\frac{1}{2l^2} > s + \tau,$$

the function $\psi_0(x,y)$ also will fall off as a Gaussian along the imaginary axis. This implies the Gaussian falloff of the wave function at infinity. Hence $\psi_0(x,y)$ [Eq. (28)] with $F_0(z)$ [Eq. (30)] is an acceptable wave function.

To proceed further, we shall consider the simplest site disorder case in which the point potentials with random strengths are located at the sites of a square lattice with constant 1. In this case the Weierstrass product $W(z)$ in Eq. (30) must be replaced by the Weierstrass σ-function

$$\sigma(z) = zW(z),$$

where the points z_j in Eq. (29) are all the sites of a square lattice (except the origin). Substituting instead of the Weierstrass σ-function its expression in terms of the Jacobi θ_1-function[16] and using the fact that the type τ of the σ-function equals $\pi/2$,[43] we obtain[17,41,42]

$$\psi_0(\mathbf{r}) = \exp(-y^2/l^2)\, \theta_1(\pi z). \tag{31}$$

Further study shows that if the magnetic field is strong enough $\varphi \geqslant 1$, then this function represents an exact wave function with eigenenergy E_{0L}. In the case $\varphi = 1$, this wave function is delocalized in both directions, but it decreases with $|y|$ if $\varphi > 1$. The wave function (31) is regular (despite the singular nature of the point potential in its limiting form) and is independent of the strength of the disorder. The lattice symmetry and gauge invariance allow one to construct more general wave functions which are delocalized in both directions for an arbitrary field which satisfy the condition $\varphi \geqslant 1$.[44] The simplest example is given by the formula

$$\Psi_{0q}^{(1)}(x,y) = \sum_Y \exp\left(iqY - i\frac{xY}{l^2}\right) \psi_0(x, y - Y). \tag{32}$$

This is a quasi-Bloch wave function with respect to y, and an extended function of x. (By the term quasi-Bloch we mean that it has an x-dependent wave number equal to $q - x/l^2$, in which q is a Bloch-type wave number with $-\pi < q \leqslant \pi$.

This approach can be generalized to an arbitrary index n of the Landau level. In general, the integral equation (27) can be written as follows:[44]

$$\psi_n(x,y) = \exp\left(-\frac{y^2}{l^2}\right) \left\{ H_n\left[\frac{y}{l} - il\frac{d}{dz}\right] F_n(z) \right\} \Bigg|_{z = x+iy},$$

where in the expansion of the differential operator

$$H_n\left[\frac{y}{l} - il\frac{d}{dz}\right],$$

y is assumed to be a constant (independent of z). Unfortunately, here it is not so simple to construct an appropriate analog of the wave function (31) or (32), because of an indirect relation between the point potential positions and the zeros of the entire function $F_n(z)$. For example, it is sufficient to choose $F_n(z)$ in the last equation as

$$F_n(z) = [F_0(z)]^{n+1}.$$

The resulting function $\psi_n(x,y)$ again vanishes at all points z_j, but it represents a true wave function for essentially higher fields. For the site disorder case (the square lattice

considered above), the function $\psi_n(x,y)$ does not grow at infinity when $\varphi \geqslant n+1$ (Ref. 44), while the true wave functions exist in a weaker field $\varphi \geqslant 1$.

5.2. Density of states

The condensation of states on the Landau levels, in a strong field $\varphi > 1$, which was predicted earlier,[25] implies that the most singular part of the DOS has the form (19) (recall that we fixed a finite-dimensional average distance d between point potentials). In the periodic case of a (square) lattice $\mathbf{r}_j = (m_j, n_j)$ of identical point potentials this result was rigorously obtained in Ref. 29. However, the states on the Landau levels are sensitive to the positions of the point potentials, not to their strength. Therefore, Eq. (19) is also valid for any site disorder model.[45] For the position disorder models with uniformly distributed zero-range scatterers, the prediction (19) was confirmed in Ref. 46 for the lowest Landau level and in Ref. 47 for the higher Landau levels. Note that in the two papers[46,47] the projection on the corresponding Landau level was applied. This enables one to use not the point potential, but the conventional $2D$ Dirac delta function as a model of zero-range scatterer.

Let us consider the DOS between Landau levels, which was studied for different types of site-disorder models (square lattice of point potentials). Consider first the so-called Maryland model, which was proposed for a $1D$ case[48] and then generalized to the multidimensional case.[49,50] In the point potential version of this model[51] the random binding energy E_{bj} of the point potential placed at the site $\mathbf{r}_j$ is

$$E_{bj} = E_b \tan(\pi \mathbf{p} \mathbf{r}_j - \omega).$$

Here the constant vector $\mathbf{p}$ is chosen in such a way that $\mathbf{p}\mathbf{r}_j$ is not a rational number for any lattice vector $\mathbf{r}_j$ and a random phase ω is uniformly distributed over the interval $(0, 2\pi)$. For a fixed realization (i.e., for a fixed phase ω) this model is described by an almost periodic Hamiltonian, in which the values of the binding energies could be very large. If $\mathbf{p}\mathbf{r}_j$ is ''sufficiently irrational'', i.e., if it cannot be well approximated by a rational number for any j, then all intervals between the Landau level are densely filled by nondegenerate localized states, which are sums of scattered waves (2) with amplitudes η_j that satisfy some kind of Lifshits equation.[51]

The next two site-disorder models[45] deal with the case in which the binding energy is

$$E_{bj} = E_b \exp(4\pi t_j).$$

Further details depend on the statistics of the random exponents t_j.

The Lloyd model. All t_j are independent random quantities with the same Lorentz distribution

$$p(t) = \frac{\theta}{\pi(t^2+\theta^2)}.$$

Here the expression for the average DOS is obtained in a closed form.[45] Its analysis shows that the DOS is an analytic function of energy and differs from zero for all energies lying between Landau levels. In the limiting cases of very small concentration of point potentials (very strong magnetic field) the DOS between Landau levels is proportional to the concentration of point potentials (the inverse magnetic field). (In the latter case the relative position of the energy with respect to the neighboring Landau levels must be fixed).

Gaussian distribution. All t_j are independent random quantities with the same Gaussian distribution

$$p(t) = (2\pi\omega)^{1/2} \exp(-t^2/2\omega). \tag{33}$$

Here the asymptotic expression for the DOS in the same limiting cases as in the previous paragraph was obtained. The resulting DOS has sharp peaks near the levels E_n which are shifted from the Landau levels E_{nL} in the presence of a single point potential. In the nearest neighborhood of the levels E_n these peaks are Gaussian, but far from E_n they become asymmetric (due to the asymmetric position of E_n with respect to the Landau level E_{nL}). More detailed investigation shows that for higher Landau levels such a form of peaks is valid also in the case of correlated, identically distributed exponents t_j [Eq. (33)].

All these results concern the case of true point potential. But we have already mentioned that a single scatterer with a finite radius splits the Landau levels. Disorder leads to the spreading of these sublevels. If the concentration of scatterers is small and magnetic field is strong enough $\varphi \geqslant 1$, then the sublevels which are most remote from the parent Landau level are spread into the resolved impurity subbands.[23] Each of these subbands accommodates one state (if $-m > n$) or two states (if $|m| < n$) per scatterer (recall that n is the number of the Landau level, and m is an angular momentum). Such an oscillating fine structure of the DOS should manifest itself in the oscillations in the low-temperature specific heat, in the magnetic susceptibility, and perhaps in the transport properties when the flux φ changes by 1 or 2 (Ref. 23).

5.3. Electron localization in a $1D$ disordered system

Let us now consider a site-disorder model where the binding energies are identical along the x direction and are random in the y direction, i.e.,

$$E_{bj} = E_{bn}, \quad \{j\} = (m,n).$$

We assume that the random binding energy E_{bn} takes two possible values: E_1 with the probability $1-c$ and E_2 with the probability c. In what follows we will be interested in the case of weak disorder so that these two binding energies almost coincide. Due to the presence of disorder and with the specific gauge, we can use the representation (20) for the scattering coefficients, which reduced the initial $2D$ system to a purely $1D$ disordered model, which conserves some features of the initial $2D$ problem: it depends explicitly on q_1 and on the magnetic field.

Consider first the unperturbed ordered system with identical binding energies $E_{bj} = E_b$. Fix some rational flux $\varphi = Q_1/Q$ and Landau level number, and choose some dispersive subband (i.e., some concrete segment on the energy-flux diagram). Each value of energy from this subband corresponds to some specific line in the rectangle $0 < q_1$, $Qq_2 < 2\pi$ on the $\mathbf{q}$-plane. In turn, each point of this rectangle corresponds to the eigenstate with quasi-momentum $\mathbf{q}$. Now switch on a weak disorder. The disorder shifts and smears the subband boundaries and, because of the $1D$ nature of the

problem, causes all states to be localized states (in the y direction). The boundaries which do not coincide with some of the Landau levels are the fluctuation boundaries:[52] they depend on the possible strengths of the point potentials. The states in the vicinities of these boundaries are strongly localized. The states lying deeply in the subband are weakly localized. They can be approximately classified by a quasimomentum $\mathbf{q}$ and by the localization length, $\xi(\mathbf{q})$, $\xi(q)\ll 1$. The edges which coincide with the Landau levels, are the stable boundaries: they do not depend on the possible strengths of point potentials. The localization length near such a boundary diverges

$$\xi(E)\propto(E_{nL}-E)^{-\nu}$$

with a critical exponent ν.

These statements are now important within a quantitative description. In the strong field limit, when the Larmor radius is of the order of or less than the lattice constant, each of the Green's functions that enters in the representation (8) essentially differs from zero only when the point $\mathbf{r}$ is close to the corresponding site (m,n). Therefore

1) *the localization properties of the wave function (8) coincide with the localization properties of the set of scattering coefficients* $\{\eta_{mn}\}$ [Eq. (20)],

and

2) the scattering coefficients ζ_n satisfy the *random* Harper equation

$$\zeta_{n+1}+2\zeta_n\cos(2\pi n\varphi+q_1)+\zeta_{n-1}=\varepsilon\zeta_n+v z_n\zeta_n.$$

Here ε is defined by Eq. (26), where $\ln(E_L/E_b)$ is replaced by its average value. The random variable z_n takes values $-c$ with the probability $1-c$, and $1-c$ with the probability c. The intensity v of the effective ''random potential'' $v\zeta_n$ is

$$v=\frac{\sqrt{\pi\varphi}}{W_{1/2+a,0}(\pi\varphi)}\frac{\ln(E_1/E_2)}{\Gamma(-a)}.$$

In the weak disorder limit, $|\ln(E_1/E_2)|\ll 1$, each solution of the ordered Harper equation (which satisfies a fixed initial condition) is slightly modified because of the appearance of a slowly increasing multiplier. The corresponding eigenfunctions are slowly attenuated Bloch solutions. The pertinent inverse localization length $\xi_i^{-1}(\mathbf{q})$ (measured in units of d^{-1}) coincides with the doubled Lyapunov exponent.[52] In the weak scattering approximation it was calculated exactly:[40]

$$\xi_i^{-1}(\mathbf{q})=\frac{1}{Q}\sum_{p=0}^{Q-1}\left\{(1-c)\left(1+\frac{Z_i(p,\mathbf{q})}{v^2c^2}\right)^{-1}+c\left(1+\frac{Z_i(p,\mathbf{q})}{v^2(1-c)^2}\right)^{-1}\right\},$$

$$Z_i(p,\mathbf{q})=4\,\mathrm{Im}^2\left[\frac{c_i(p+1,\mathbf{q})}{c_i(p,\mathbf{q})}\exp(iq_2)\right].$$

The weak scattering regime corresponds to a very large localization length. We define the condition of the weak scattering as

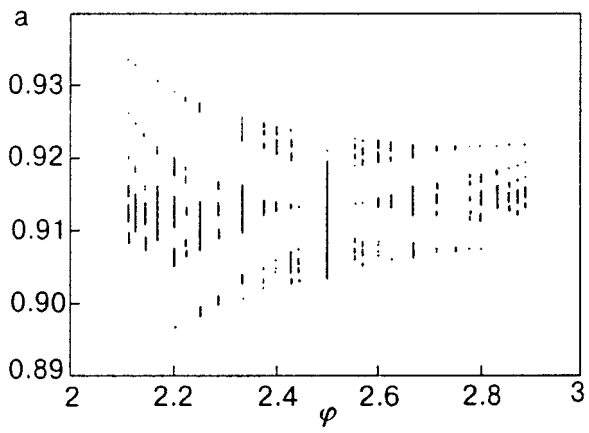

FIG. 4. The energy-flux diagram for the delocalized (with the chosen accuracy) states between the first two Landau levels $0<a\leqslant 1$. The rational fluxes $2<\varphi<3$ with denominators up to 9 are used. Some subbands, which are shown in Fig. 3b, are absent here: the corresponding states are localized.

$$\min_p Z_i(p,\mathbf{q})/v^2>N\gg 1. \tag{34}$$

This inequality determines a domain inside the square $0<q_1$, $Qq_2<2\pi$, such that for each $\mathbf{q}$ in this domain, the localization length (measured in the lattice constant) $\xi_i(\mathbf{q})$ will be larger than N. The dispersion relation $\varepsilon_i(\mathbf{q})$ maps this domain onto an energy interval on the ε axis and, after inversion using Eq. (26), onto a set of intervals on the a axis. Each of these intervals corresponds to a fixed value of the rational flux and a fixed Landau level (i.e., it is located below this level but above the previous one). The set of all these intervals vs flux φ forms a Hofstadter-type butterfly for the localization length. The points of this diagram correspond to those eigenstates which can be classified as slowly attenuated Bloch states. A part of this butterfly is shown in Fig. 4. Note that some subbands, which are present in Fig. 3, are completely absent in Fig. 4. This means that all states of these subbands do not satisfy the weak scattering condition (34), i.e., they are localized.

Let us now consider the localization of states near the stable boundaries. Different types of touching scenarios (possible rational fluxes and possible points in the $\mathbf{q}$ rectangle) were analyzed in Ref. 36. It was shown[19] that in the case of short, correlated $1D$ disorder the localization length diverges with the critical exponent $\nu=1$. Exact expressions for the localization length near the Landau levels, which were obtained for the model considered in this subsection, confirm this result.[40] In particular, for the touching of the first Landau level for $\varphi=2$ the localization length has the form

$$\xi\propto\frac{(\langle\ln(E_L/E_b)\rangle)^2}{(1-a)c(1-c)\ln^2(E_1/E_2)},\quad F=\mathrm{const},$$

and diverges as $a\to 1$ (the energy tends to E_{1L}) and when the disorder vanishes ($c\to 0$, 1, or $E_2-E_1\to 0$).

6. MESOSCOPIC SYSTEMS WITH A POINT POTENTIAL

The electron dynamics in the presence of a single short-range impurity in a magnetic field is related to a number of

physical problems (magnetotransport phenomena in bulk semiconductors, the quantum Hall effect, the conductance of a microconstriction, tunneling through a quantum dot, etc.). Because of the short-range character of the impurity, it is natural to model it by a point potential. Such a simplification often enables one either to obtain an exact solution of the problem, or at least to obtain a solution in a closed form.

6.1. Point potential model of a quantum dot

Let us first discuss the problem of electron transmission through a quantum dot. On the basis of the model proposed and studied in Ref. 53 the dot itself is described by a $2D$ Hamiltonian H_2 which accounts for a magnetic field (12) and a confinement potential $V_c = E_c^2 r^2/4$, $r^2 = x^2 + y^2$, with some characteristic confinement energy E_c. One-dimensional channel is described by a $1D$ free Hamiltonian $H_1 = -d^2/dz^2$. The unperturbed Hamiltonian of the noninteracting system dot-channel is a direct sum $H_2 \oplus H_1$:

$$H = \begin{pmatrix} H_2 & 0 \\ 0 & H_1 \end{pmatrix}.$$

The unperturbed wave function is a two-component vector

$$\Psi_0(\mathbf{r}, z) = \begin{pmatrix} \psi_2(\mathbf{r}) \\ \psi_1(z) \end{pmatrix}.$$

The perturbation V located at the origin describes the zero-range interaction between the dot and the channel

$$V = \begin{pmatrix} T_2^{-1}(E) & \beta \\ \beta & T_1^{-1}(E) \end{pmatrix}.$$

Here $T_2(E)$ is the $2D$ point potential scattering amplitude (18), where E_L is replaced by $(E_L^2 + E_c^2)^{1/2}$, and $T_1(E)$ is the $1D$ point potential scattering amplitude. The $1D$ point potential, however, is the usual repulsive Dirac delta function $k_0 \delta(z)$, $k_0 > 0$ (repulsion models the potential barriers that separate the real quantum dot from the leads, which are modeled here by the channel). Therefore, in accordance with the general rule (9), the scattering amplitude $T_1(E)$ is

$$\frac{1}{T_1(E)} = \frac{i}{2\sqrt{E}} + \frac{1}{k_0}.$$

The constant β describes mixing between the dot and the channel. The case $\beta = 0$ corresponds to two independent subsystems—the dot, which contains the $2D$ point potential, and the channel with the $1D$ point potential. The nonzero value of the parameter β provides a mixing of the two subsystems and thus the dot actually influences the electron transmission along the channel.

The transmission coefficient $T(E)$, which is calculated by a standard method, has the form

$$\frac{1}{T(E)} = 1 + \frac{1}{4E(k_0^{-1} - \beta^2 T_2(E))^2}.$$

At energies $E_n^{\min}$ satisfying the resonant condition $k_0^{-1} = \beta^2 T_2(E)$ the channel is closed: the transmission vanishes. On the other hand, at energies $E_n^{\max}$, which coincide with the poles of the scattering amplitude $T_2(E)$, the transmission is perfect $T(E) = 1$. Note that these energies are simply Landau levels in an effective magnetic field which is shifted by the $2D$ point potential. In this effective field the interlevel distance is not E_L but $(E_L^2 + E_c^2)^{1/2}$. The channel conductance calculated by the Landauer formula has sharp peaks at the energies $E_n^{\max}$.

A similar idea was used for studying the electron spectrum in a periodic array of quantum dots subject to a magnetic field.[54] Consider a lattice of points $\mathbf{R}$. Let $H_{\mathbf{R}}$ be the potential of the dot located at the point $\mathbf{R}$ [i.e., with the confinement potential $V_c(\mathbf{r} - \mathbf{R})$]. The array of dots is described by the Hamiltonian $H = H_0 + U$. Here the unperturbed Hamiltonian $H_0 = \oplus H_{\mathbf{R}}$ is the direct sum of the single-dot Hamiltonians $H_{\mathbf{R}}$ over all sites of the lattice $\{\mathbf{R}\}$. This means that the unperturbed wave function $\Psi_0(\mathbf{r})$ is the vector column

$$\Psi_0(\mathbf{r}) = \begin{pmatrix} \cdots \\ \psi_{\mathbf{R}}(\mathbf{r}) \\ \cdots \end{pmatrix},$$

and each one of its components $\psi_{\mathbf{R}}(\mathbf{r})$ satisfies the Schrödinger equation with the Hamiltonian $H_{\mathbf{R}}$. The mixing of the states in the different dots is realized by adding the perturbation U which in coordinate representation has the form

$$U_{\mathbf{R},\mathbf{R}'}^{\mathbf{r},\mathbf{r}'} = t_{\mathbf{R},\mathbf{R}'} u(\mathbf{r} - \mathbf{R}) u^*(\mathbf{r}' - \mathbf{R}').$$

Here the matrix t has the lattice symmetry

$$t_{\mathbf{R}+\mathbf{R}_0, \mathbf{R}'+\mathbf{R}_0} = t_{\mathbf{R},\mathbf{R}'},$$

which contains only nondiagonal elements, and $u(\mathbf{r})$ is a localized function.

The perturbation U in this model is not a degenerate perturbation. But like the degenerate perturbation (1), which acts in the subspace formed by the set of states $\{|v_j\rangle\}$, the perturbation U acts in the subspace formed by the set of states $\{|u_{\mathbf{R}}\rangle\}$. The spectrum of a single dot is rather rarified $E_{n,m} = (n + 1/2)E_L + (m + 1/2)E_c$ with integers $n, m \geqslant 0$. As a result, the set of the perturbed eigenstates $\Psi(\mathbf{r})$ with eigenenergies lying outside the initial spectrum is sufficiently rich and can be found from the set of Lifshits equations (2)–(4). The index j must be replaced by $\mathbf{R}$,

$$\eta_{\mathbf{R}} = \int u(\mathbf{r} - \mathbf{R}) \psi_{\mathbf{R}}(\mathbf{r}) d\mathbf{R},$$

$t_{\mathbf{R},\mathbf{R}'}$ stands for G_{ij} and

$$\frac{1}{T_{\mathbf{R}}(E)} = \int u^*(\mathbf{r} - \mathbf{R}) G_{\mathbf{R}}(\mathbf{r}, \mathbf{r}') u(\mathbf{r}' - \mathbf{R}) d\mathbf{r} d\mathbf{r}'.$$

These formulas are valid for an arbitrary localized function $u(\mathbf{r})$. A further simplification can be made by using the point potential as $u(\mathbf{r})$. In this case and for a rational flux per plaque an explicit equation for the Bloch-type spectrum can be obtained.[54]

An explicit spectrum of the model can be obtained because of its internal simplicity, which is attributable to the use of degenerate-like perturbation, on the one hand, and to the construction of the unperturbed Hamiltonian as a direct product, on the other. Due to the latter fact, the interaction

between different dots is described independently from the dynamics of a single dot. In a sense, the model considered is an ordered analog of the famous Anderson model,[55] while the case of the point potential lattice, which is considered in Sec. 4, is an ordered analog of the much more complicated Lifshits model.[56,57]

6.2. Point scatterer in a microconstriction

Here we consider electron scattering by a single point potential in a magnetic field and some additional field. The start from the case of the crossed magnetic and electric fields. The modification of the spectrum of the system due to the presence of an impurity is defined by the poles of the scattering amplitude $T(E)$ [Eq. (9)], where $G(\mathbf{r},\mathbf{r}';E)$ now stands for the Green's function in the presence of the two fields. As a result, in contrast with the case of zero electric field, where exactly one bound state exists below each Landau level (see Fig. 1), there are now n novel, nondegenerate, quasi-bound states with energies close to the nth Landau level.[58]

The next group of problems is related to electron transmission through a saddle-point potential

$$V(x,y)=-\frac{x^2}{\kappa^4}+y^2 \tag{35}$$

[here κ is the asymmetry parameter] in the presence of a short-range impurity and magnetic field. Without impurity and a magnetic field the electron wave functions are saddle-point potential waveguide modes,

$$\Psi^{\pm}_{E,n}(x,y)=\Phi_n(y)E\left(-\varepsilon_n,\pm\frac{x\sqrt{2}}{\kappa}\right).$$

Here $\Phi_n(y)$ is a harmonic oscillator wave function, which corresponds to the y-dependent part of the potential [Eq. (35)] and energy $E_n=2n+1$. The x-dependent factor is the Weber-function[16] $E(-\varepsilon,\xi)$, and $\varepsilon_n=\kappa^2(E-2n-1)/2$. The mode $\Psi^{+}_{E,n}(x,y)$ describes the initial wave which comes from the left ($x=-\infty$) and which contains the reflected and transmitted waves with the corresponding transmission coefficient $T_n=(1+\exp(-2\pi\varepsilon_n))^{-1}$ (this result was initially obtained in the WKB approximation in Ref. 59 and reflectivity $R_n=(1+\exp(2\pi\varepsilon_n))^{-1}$.The energy E_n is the threshold for mode n. When $E<E_n$, the mode n is mainly reflected, while in the opposite case this mode is mainly transmitted. The crossover from reflection to transmission turns out to be in the energy band $|E-E_n|\leqslant 2\kappa^{-2}$. The saddle-point potential (35) models a ballistic microjunction, to which the two-terminal conductance in terms of the mode transmission coefficients is given by[60]

$$G_0=\sum_{n=0}^{\infty}T(\varepsilon_n).$$

The presence of a strong point potential, which is located in the central section of the constriction (i.e., near the origin), modifies the spectrum. Exactly one bound state appears below the threshold E_n of each transverse quantization mode. The bound state with the energy far from the threshold decays mostly into the continuous spectra of the above-barrier modes (mode mixing). In this case the point potential correction to the conductance is due to the resonant reflection of the above-barrier modes, which have a downward dip $\Delta G=-1$ on the plot of G vs the Fermi energy E. The bound state close to threshold decays mostly into the continuous spectrum of the threshold mode (tunneling). Here the point potential correction to the conductance is due to resonant transmission of the threshold mode which is seen as a resonant tunneling peak,[61] $\Delta G=+1$.

Let us consider the electron scattering by an isolated point potential in a magnetic field. In the limiting case of a narrow constriction $\alpha\geqslant 1$ an explicit expression for the transmission coefficient as a function of the point potential position and magnetic field could be obtained.[62] Its analysis shows that the transmission coefficient has a sharp peak (the Breit–Wigner resonance) as a function of energy, and that it oscillates as a function of the point potential position.[62] One more aspect of electron transmission through a symmetric [$\alpha=1$] saddle-point potential in a magnetic field is related to the levitation of delocalized states in the quantum Hall effect regime. In the network model of the quantum Hall effect the delocalization of states results from the tunneling of an electron through the saddle points of a smooth random potential, which are connected by equipotential lines.[63] The mixing of Landau levels changes the transmission coefficient of a saddle point, on the average, in such a way that it becomes smaller than 1/2 for the energy at the Landau level.[64] This means that to achieve the 1/2 average transmittivity, the energy should be shifted upwards, which is equivalent to levitation. However, in a smooth potential the Landau level mixing is generally weak since it is associated with a large momentum transfer. Short-range potentials are much more effective in this respect. The average transmission through a saddle point in a strong magnetic field in the presence of a random short-range scatterer was studied in Ref. 65. It was shown that a small portion of the short-range random impurities located in the vicinity of the saddle points reduce the transmission at a given energy even if the mixing of the Landau levels by a smooth potential is disregarded. The resulting upward shift of the energy position of the delocalized state increases with decreasing magnetic field as B^{-4}.[65]

CONCLUSIONS

In summary, we presented a brief review of the results concerning the electron dynamics in a magnetic field and in a field of a point or a short-range potential. We should emphasize that the possibilities contained in the point-potential model for the physics of $2D$ systems in a magnetic field are not exhausted yet. This is a unique model which makes it possible to obtain exact results starting from first principles. In Sec. 4 we presented the results of a calculation of the electron spectrum and wave functions using solely the zero-range property of the scattering potential. It is very intriguing to obtain some transport characteristics, e.g., the Hall conductivity or ac conductivity of the $2D$ periodic system of point potentials. Unexpected application of point potentials is attributable to the superconductivity. The results of Subsec. 5.1 make it possible to construct explicitly an Abrikosov trial function for an arbitrary position of the vortices in

type-II superconductors (this idea was communicated to us by the late Arkadii Aronov). Investigation in this direction is now in progress.

We shall conclude this article with some non-academic words. We believe that the best way to honor the memory of an outstanding scientist is to show that his ideas are alive, that they are at work, and that they are being developed. We attempted to demonstrate this point on a single subject of local perturbations which has been proposed by Ilya Mikhailovich back in 1947. The fact that the topics related to this subject, such as electron localization in a magnetic field, the quantum Hall effect, the physics of quantum dots, and others, are at the forefront of contemporary solid state physics speaks for itself. All the authors, two of whom (M. Ya. Azbel' and S. A. Gredeskul) are privileged to be former pupils of Ilya Mikhailovich Lifshits, would like to express their gratitude and to dedicate this article with love and respect, to the memory of this brilliant scientist and person.

This work was supported, in part, by a grant from Israel Science Foundation, by Raymond and Beverly Sackler Faculty of Exact Sciences, and by J. and R. Meyerhoff Chair for Solid State Physics and Thermodynamics.

*E-mail: sergeyg@bgumail.bgu.ac.il
*E-mail: yshai@bgumail.bgu.ac.il
**E-mail: azbel@taunivm.tau.ac.il

[1] I. M. Lifshits, Zh. Éksp. Teor. Fiz. **17**, 1017 (1947).
[2] E. Fermi, Nuovo Cimento **11**, 157 (1934).
[3] E. Fermi, Ric. Sci. **7**, 13 (1936).
[4] Yu. N. Demkov and V. N. Ostrovskii, *Method of the Short-Range Potentials in the Atomic Physics*, Plenum, New York (1988).
[5] S. Albeverio, F. Gestezy, R. Höegh-Krohn, and H. Holden, *Solvable Models in Quantum Mechanics*, Springer, New York (1988).
[6] B. S. Pavlov, Uspekhi Mat. Nauk **42**, 99 (1987) [Russian Math. Surveys **42**, 127 (1987)].
[7] V. A. Geiler, Algebra 1 Analys **3**, 1 (1991) [St. Petersburg Math. J. **3**, 489 (1992)].
[8] L. D. Landau, Zs. Phys. **64**, 629 (1930).
[9] Ya. I. Frenkel' and M. P. Bronshtein, Zh. Russ. Fiz. Khim. O-va, Chast' Fix., **62**, 485 (1930).
[10] D. R. Hofstadter, Phys. Rev. B **14**, 2239 (1976).
[11] A. M. Kosevich, Osnovy Mechaniki Kristalitzeskoj Reshetki, Moscow, Nauka (1972).
[12] F. A. Berezin and L. D. Faddeev, Dokl. Akad. Nauk SSSR **137**, 1011 (1961).
[13] M. G. Krein, Dokl. Akad. Nauk SSSR **52**, 651 (1946).
[14] M. Ya. Azbel', Phys. Rev. Lett. **67**, 1787 (1991).
[15] Y. Avishai, M. Ya. Azbel', and S. A. Gredeskul, Phys. Rev. B **48**, 17280 (1993).
[16] I. S. Gradstein and I. M. Ryzhik, *Tables of Integrals, Sums, Series and Products*, Academic Press, New York (1980).
[17] V. A. Geiler and V. A. Margulis, [Teor. Mat. Fiz. **58**, 461 (1984) [Theoret. Math. Phys. **58**, 302 (1984)].
[18] F. Gesztesy, H. Holden, and P. Seba, in *Schrödinger Operators, Standard and Nonstandard*, P. Exner and P. Seba (Eds.), World Scientific, Singapore (1989), p. 147.
[19] M. Ya. Azbel', Phys. Rev. B **43**, 2435 (1991).
[20] M. Ya. Azbel', Phys. Rev. B **43**, 6717 (1991).
[21] L. D. Landau and E. M. Lifshits, *Quantum Mechanics (Nonrelativistic Theory)*, 3rd ed., Pergamon, Oxford (1977).
[22] E. M. Baskin, L. N. Magarill, M. V. Entin, Zh. Éksp. Teor. Fiz. **75**, 723 (1978) [Sov. Phys. JETP **48**, 365 (1978)].
[23] M. Ya. Azbel' and B. I. Halperin, Phys. Rev. B **52**, 14098 (1995).
[24] S. A. Gredeskul and M. Ya. Azbel', Phys. Rev. B **49**, 2323 (1994).
[25] T. Ando, J. Phys. Soc. Jpn. **36**, 1521 (1974).
[26] T. C. Dorlas, N. Macris, and J. V. Pule, J. Stat. Phys. (in press).
[27] J. Zak, Phys. Rev. A **134**, 1602 (1964).
[28] J. Zak, Phys. Rev. A **134**, 1607 (1964).
[29] V. A. Geiler and V. A. Margulis, Teor. Mat. Fiz. **61**, 140 (1984) [(Theoret. Math. Phys. **61**, 1049 (1984)].
[30] V. A. Geiler, in *Invited Talks of the International Workshop ''Mathematical Aspects of the Scattering Theory and Applications'', St. Petersburg, May 20–24, 1991*, St. Petersburg University (1992), p. 36.
[31] M. Ya. Azbel', Zh. Éksp. Teor. Fiz. **46**, 929 (1964) [Sov. Phys. JETP **19**, 634 (1964)].
[32] M. Ya. Azbel', Phys. Rev. B **45**, 4208 (1992).
[33] S. A. Gredeskul, Y. Avishai, M. Ya. Azbel', and M. Zusman, in *Nonlinearity with Disorder*, L. Vazquez, A. R. Bishop, and S. Petersen (Eds.), World Scientific, Singapore (1995), p. 333.
[34] M. Ya. Azbel', Physica A **200**, 491 (1993).
[35] S. A. Gredeskul, M. Zusman, Y. Avishai, and M. Ya. Azbel', in *Waves in Random and other Complex Media*, Proceedings of the International Workshop ''Waves in Random and other Complex Media'', R. Burridge, L. A. Pastur, and G. Papanicolau (Eds.), Springer, New York (in press).
[36] M. Ya. Azbel', Phys. Rev. B **49**, 5463 (1994).
[37] P. G. Harper, Proceedings of the Phys. Soc. A **68**, 874 (1955): *ibid* 879.
[38] I. M. Lifshits, M. Ya. Azbel', and M. I. Kaganov, *Electron Theory of Metals*, Plenum Press, New York (1973).
[39] G. E. Zilberman, Zh. Éksp. Teor. Fiz. **32**, 296 (1957) [Sov. Phys. JETP **5**, 208 (1957)].
[40] S. A. Gredeskul, M. Zusman, Y. Avishai, and M. Ya. Azbel', in *Hopping and Related Phenomena, Proceedings of VI International Conference HRP-6*, Z. Ovadyahu and O. Millo (Eds.), Hebrew University Publishing, Jerusalem (1995), p. 137.
[41] Y. Avishai, R. M. Redheffer, and Y. B. Band, J. Phys. A **25**, 3883 (1992).
[42] Y. Avishai and R. M. Redheffer, Phys. Rev. B **47**, 2089 (1993).
[43] R. P. Boas, *Entire Functions*, Academic, New York (1954).
[44] S. A. Gredeskul, Y. Avishai, and M. Ya. Azbel', Europhys. Lett. **21**, 489 (1993).
[45] V. A. Geiler and V. A. Margulis, Zh. Éksp. Teor. Fiz. **95**, 1134 (1989) [Sov. Phys. JETP **68**, 654 (1989)].
[46] E. Brezin, D. J. Gross, and C. Itzykson, Nucl. Phys. B **235**, 24 (1984).
[47] K. A. Benedict and J. T. Chalker, J. Phys. **19**, 3587 (1986).
[48] D. R. Grempel, S. Fishman, and R. E. Prange, Phys. Rev. Lett. **49**, 833 (1982).
[49] L. A. Pastur and A. L. Figotin, Pis'ma Zh. Éksp. Teor. Fiz. **37**, 575 (1983) [JETP Lett. (1983)].
[50] A. L Figotin and L. A. Pastur, Commun. Math. Phys. **95**, 401 (1984).
[51] V. A. Geiler and V. A. Margulis, Teor. Mat. Fiz. **70**, 192 (1987) [Theoret. Math. Phys. **70**, 133 (1987)].
[52] I. M. Lifshits, S. A. Gredeskul, and L. A. Pastur, *Introduction to the Theory of Disordered Systems*, Wiley, New York (1988).
[53] V. A. Geiler and I. Yu. Popov, Phys. Lett. A **187**, 410 (1994).
[54] V. A. Geiler and I. Yu. Popov, Z. Phys. B **93**, 437 (1994).
[55] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).
[56] I. M. Lifshits, Zh. Éksp. Teor. Fiz. **44**, 1723 (1963) [Sov. Phys. JETP **17**, 56 (1963)].
[57] I. M. Lifshits, Adv. Phys. **13**, 483 (1964).
[58] C. Kunze, Y. B. Levinson, M. I. Lubin, and E. V. Sukhorukhov, Pis'ma Zh. Éksp. Teor. Fiz. **56**, 55 (1992) [JETP Lett. **56**, 56 (1992)].
[59] H. A. Fertig and B.I. Halperin, Phys. Rev. B **36**, 7969 (1987).
[60] M. Buttiker, Phys. Rev. B **41**, 7906 (1990).
[61] Y. B. Levinson, M. I. Lubin, and E. V. Sukhorukov, Phys. Rev. B **45**, 11936 (1992).
[62] V. A. Geiler, V. A. Margulis, and I. I. Chuchaev, Pis'ma Zh. Éksp. Teor. Fiz. **58**, 668 (1993) [JETP Lett. **58**, 648 (1993)].
[63] J. T. Chalker and P. D. Coddington, J. Phys. C **21**, 2665 (1988).
[64] T. V. Shahbazyan and M. E. Raikh, Phys. Rev. Lett. **75**, 304 (1995).
[65] A. Gramada and M. E. Raikh, Phys Rev. B **54**, 1928 (1996).

This article was published in English in the original Russian journal. It was edited by S. J. Amoretty.

On the theory of skin effect: inclusion of arriving term in collision integral

M. I. Kaganov

7 Agassiz Ave., Belmont, Massachusetts 02178

G. Ya. Lyubarskiĭ

6442 N. Troy, Chicago, Illinois 60645

A. G. Mitina

744 S. Cleveland, Arlington Heights

(Submitted September 23, 1996)
Fiz. Nizk. Temp. **23**, 36–46 (January 1997)

An algorithm for calculating the surface impedance of a normal isotropic metal is constructed by taking into account the arriving term in the collision integral (in the case of specular reflection). Analytic expressions are obtained for scattering probability describing the *s*-, *p*-, and *d*-scattering. © *1997 American Institute of Physics.* [S1063-777X(97)00401-5]

1. At low temperatures, the mean free path l is a macroscopic parameter which is often larger than other parameters of the dimensions of length in good metals (the skin depth δ, the wavelength of sound, the thickness of a plate or wire). Under the condition when $l>\delta$, the skin effect is usually called anomalous. The skin effect anomaly is caused by spatial dispersion of conductivity, i.e., nonlocal relation between the electric field strength $\mathbf{E}=\mathbf{E(r)}$ and the current density $\mathbf{j}=\mathbf{j(r)}$. Formally, spatial dispersion (nonlocality) is a consequence of inclusion of the diffusion term $\mathbf{v}\partial f/\partial \mathbf{r}$ in the Boltzmann kinetic equation for the nonequilibrium component of the distribution function f of conduction electrons ($\mathbf{v}$ is the electron velocity). The inclusion of the diffusion term transforms the integral kinetic equation into an integrodifferential equation. One of the consequences of this transformation is the dependence of the distribution function on the form of interaction of electrons with the sample boundary.[1)] It was shown in Refs. 1 and 2 that the integrodifferential form of the kinetic equation leads to difficulties in the introduction of the concept of mean free path.

The collision integral $I\{f\}$ in the linearized kinetic equation is always the difference of two terms:

$$I\{f\}=\int W(\mathbf{p},\mathbf{p}')f(\mathbf{p}')d^3p'-\int W(\mathbf{p}',\mathbf{p})f(\mathbf{p})d^3p', \qquad (1)$$

where $W(\mathbf{p},\mathbf{p}')$ is the probability density for an electron transition from the state $|\mathbf{p}'\rangle$ to the state $|\mathbf{p}\rangle$, which is normalized in a certain way.

The first integral describes the arrival of electrons in a cell of the phase space (*arriving term*), while the second (with the minus sign) describes the departure of electrons (*departing term*). If we can neglect the arriving term (τ-approximation), the Boltzmann equation is not an integral equation any longer, and the mean free time τ is introduced in the natural way:

$$\frac{1}{\tau}=\int W(\mathbf{p},\mathbf{p}')d^3p'; \quad I\{f\}=-\frac{f(\mathbf{p})}{\tau}. \qquad (2)$$

Naturally, this simplifies the problem significantly: there is no need to solve an integral equation, and τ has the meaning of the ''universal mean free time.''

In the case when the τ-approximation is inapplicable, a universal relaxation time and/or universal mean free path l do not exist. It is necessary to obtain a solution of the kinetic equation which depends on the formulation of the problem. In order to describe the situation more precisely, we shall consider below two problems differing only in the form of the action on conduction electrons. Here we confine ourselves to the case of a spherical Fermi surface, elastic collisions of electrons, and an extremely degenerate electron gas ($T=0$). Scattering changes the direction of electron motion specified by the unit vector $\mathbf{n}$.

The collision integral can be written in the form

$$I\{f\}=\int W(\mathbf{n},\mathbf{n}')f(\mathbf{n}')dO'-\int W(\mathbf{n}',\mathbf{n})f(\mathbf{n})dO', \qquad \mathbf{n}=\mathbf{p}/p, \qquad (3)$$

where dO' is an element of the solid angle of the unit vector $\mathbf{n}'$. In the isotropic case (which we consider here), we have

$$W(\mathbf{n},\mathbf{n}')=W(\theta), \quad \cos\theta=\mathbf{n}\cdot\mathbf{n}', \qquad (4)$$

and

$$I\{f\}=\int W(\theta)f(\mathbf{n}')dO'-\int W(\theta)dO'f(\mathbf{n}). \qquad (5)$$

In order to formulate the problem (to be more precise, its part pertaining to kinetics), we must also know the functions $W(\theta)$, viz., the probability densities of electron scattering through the angle θ. The calculation of this function is beyond the scope of this paper, and we shall assume that the function $W=W(\theta)$ is known and is independent of the specific formulation of the problem. This is quite admissible as long as the reason behind the deviation of the conduction electron state from equilibrium lies in macroscopic (electromagnetic and acoustic) fields whose nonuniformity is smaller than the de Broglie wavelengths of a Fermi electron. In the case of inelastic collisions of electrons, an additional and

rather stringent constraint is imposed on the macroscopic field frequency ω. A violation of this condition makes the scattering probability a nonuniversal function of frequency ω (see Sec. 47 in Ref. 3). Inelastic collisions will be disregarded.

2. The aim of this paper is to demonstrate that the details emerging in the function $W(\theta)$ in the anomalous skin effect do not allow us to confine the analysis of zero momentum [as in the τ-approximation, see (2) and (5)]. But before that, let us determine the change in the ''mean free path'' depending on the nature of action exerted on the metal (in other words, on the kinetic coefficient calculated by us). The resistivity of a metal is determined by the transport mean free path $\tau_{\rm tr}$:

$$\sigma=\frac{ne^2\tau_{\rm tr}}{m^*};\quad \frac{1}{\tau_{\rm tr}}=\int W(\theta)(1-\cos\,\theta)dO;$$
$$m^*=p_F/v_F. \tag{6}$$

While considering the absorption of sound by conduction electrons, we can write the integral equation for determining the ''relaxation time'' in the form of a tensor equation

$$\int W(\theta)[\Psi_{ik}(\mathbf{n})-\Psi_{ik}(\mathbf{n}')]dO'=n_in_k-\frac{1}{3}\delta_{ik}.$$

Hence, $\Psi_{ik}=\tau_{\rm tr}^{\rm tens}[n_in_k-\delta_{ik}/3]$, and

$$\frac{1}{\tau_{\rm tr}^{\rm tens}}=\frac{3}{2}\int W(\theta)(1-\cos^2\,\theta)dO. \tag{7}$$

Naturally, the two times coincide for $W(\theta)\equiv$const. It should be recalled that $\int f dO=0$ so that the case when $W(\theta)=$const corresponds to the τ-approximation.

The physical consequences of expressions (6) and (7) are identical: both transport times are much longer than the reciprocal (total) scattering probability in the case of small-angle scattering due to a decrease in the contribution of scattering through small angles. However, it should be borne in mind while calculating the corresponding kinetic coefficient exactly that $\tau_{\rm tr}^{\rm tens}\neq\tau_{\rm tr}$. For a small-angle scattering, we can write

$$\frac{1}{\tau_{\rm tr}^{\rm tens}}=3\frac{1}{\tau_{\rm tr}}.$$

3. In our opinion, the role of the arriving term in the collision integral in skin effect was considered for the first time in Ref. 1. It is clear a priori[2] that the role of the arriving term is important in the ''intermediate'' case. For $l/\delta\ll 1$, we can use the macroscopic value of electrical conductivity σ in the zeroth approximation in l/δ (it should be borne in mind, however, that the expression for σ contains the time $\tau_{\rm tr}$ depending on the arriving term in the collision integral). For $l/\delta\gg 1$, the impedance in the zeroth approximation in δ/l is completely independent of bulk dissipative characteristics of conduction electrons. The correction terms (to both limiting expressions) depend on the arriving term in the collision integral. It is important that the role of the arriving term is manifested in different ways for different values of l/δ. This means that there is no universal dependence of electrodynamic characteristics of the metal on l/δ. While calculating the impedance ξ and the penetration depth δ, we must fix a special parameter describing the arriving term. In actual practice, there is no need to introduce a special parameter whose presence and magnitude determine the contribution from the arriving term. This parameter emerges in natural way when the function $W=W(\theta)$ is specified. In Ref. 1, the role of the arriving term is demonstrated by the simplest complication of the collision integral [to be more precise, the function $W(\theta)$]. It was assumed that

$$W(\theta)=W_0(1+\alpha\,\cos\,\theta), \tag{8}$$

i.e., we have taken into account not only the s-scattering, when $W(\theta)\equiv W_0$, and the arriving term in the problem of skin effect vanishes, but also the p-scattering whose intensity is described by the coefficient $\alpha(|\alpha|<1)$.

The fact that the kernel of the arriving term of the collision integral is degenerate simplified the solution of the problem on skin effect with the function $W(\theta)$ defined by expression (8). As a result, its action is reduced to the replacement of the electric field strength $E(z)$ by the sum

$$E(z)\to E(z)+\frac{\alpha}{3\sigma}j(z). \tag{9}$$

The role of the arriving term is determined just by the parameter α: $\alpha=0$ corresponds to the τ-approximation.

In Ref. 1, as well as in the well-known publications by Reuter and Sondheimer,[4] the theory of skin effect was constructed for the two limiting cases as regards the type of interaction of an electron with the sample boundary, i.e., the specular reflection of electrons at the boundary, and diffuse scattering.

It is well known that in the case of diffuse scattering, the problem should be solved by the Wiener–Hopf method. As in Ref. 4, we managed to obtain a closed formula for impedance, but the integral appearing in this formula is much more complicated than in Ref. 4. Naturally, the calculation of asymptotic forms is also more complicated. This statement can be refined; we can estimate the difference in the results obtained in Refs. 1 and 4 (see also Ref. 5) and emphasize the difficulties which we managed to overcome.

A function of κ (κ is the wave vector conjugate to the z-coordinate) will be referred to as a function of the first degree of complexity if, by definition, it depends algebraically on the Fourier transform $\overline{K(\kappa)}$ of the kernel of the integral relation between the electric field strength $E(z)$ and the current density $j(z)$. An example of such a function is the function

$$F(\kappa)=\kappa^2-\left(\frac{\alpha}{3}\kappa^2+\frac{2il^2}{\delta^2}\right)\overline{K(\kappa)} \tag{10}$$

introduced in Ref. 1 (see formula (37) from Ref. 1). A function of the second degree of complexity is defined as the integral of the first-degree function, e.g., the function $G_\pm(\kappa)$ in the same publication:

$$G_\pm(\kappa)=\frac{1}{2\pi i}\int_{\mp i\varepsilon-\infty}^{\mp i\varepsilon+\infty}\frac{d\xi}{\xi-\kappa}\ln\frac{F(\xi)}{\xi^2-\kappa_0^2} \tag{11}$$

[see (38) in Ref. 1]. The function

$$H_+(\kappa)=\frac{1}{2\pi i}\int_{-i\varepsilon-\infty}^{-i\varepsilon+\infty}\frac{d\xi}{\xi-\kappa_0}\,\frac{iE(0)\xi-E'(0)}{\xi-\kappa_0}\overline{K(\xi)}\times\exp[G_-(\xi)]$$

[see formula (39) in Ref. 1] is an example of the function of the third degree of complexity.

The result obtained by Reuter and Sonderheimer,[4] i.e.,

$$\zeta_{RS}^{\mathrm{dif}}=\frac{\omega l}{c}\left(\kappa_0-\frac{1}{2\pi i}\int_{-\infty}^{\infty}d\kappa\,\ln\frac{F(\kappa)}{\kappa^2-\kappa_0^2}\right)^{-1} \tag{12}$$

is reduced to the calculation of a function of the second degree of complexity.

The intermediate formulas in Ref. 1 show that, in order to calculate impedance for $\alpha\neq 0$, we must determine the value of a function of the *third* degree of complexity [see formulas (A19) and (A20) in Appendix to Ref. 1]. The mathematical result obtained in Ref. 1 is that the calculation of impedance *could be reduced to the determination of the function G_+ of the second degree of complexity*: for $W(\theta)=W_0(1+\alpha\cos\theta)$ and for a diffuse reflection of electrons at the boundary, the impedance ζ is given by

$$\zeta=\frac{\omega l}{c}i\xi_0\frac{1-S^2}{1+S^2}; \tag{13}$$

Here

$$S=\frac{\xi_0\,\exp[-G_+(\xi_0)]}{\xi_0+\kappa_0},\quad \xi_0=\frac{1}{\delta}(-6i/\alpha)^{1/2},$$

and the function $G_+(k)$ is defined by equalities (10) and (11).

In the present communication, we generalize the results obtained in Ref. 1. It is assumed that

$$W(\theta)=W_0\left(1+\sum_{k=1}^{n}\alpha_k\cos^k\theta\right). \tag{14}$$

This expression is obviously much more general than (8). The natural limitation imposed on the coefficients α_k is that the probability $W(\theta)$ must be positive. As a rule, the value of the coefficient $|\alpha_k|$ decreases with increasing exponent k, although the sum (14) must undoubtedly contain a large number of terms in the case of small-angle scattering.

We consider the normal incidence of an electromagnetic wave on a metallic half-space $z>0$. By definition, the impedance $\zeta=E(0)/H(0)$. According to Ref. 1, it is convenient to introduce the function $\chi=\chi(z,\vartheta)$ defining the current density

$$j(z)=\frac{3\sigma}{4l}\int_0^{\pi}\chi(z,\vartheta)\sin^2\vartheta\,d\vartheta,$$
$$\mathbf{n}=(\sin\vartheta\cos\varphi,\sin\vartheta\sin\varphi,\cos\vartheta), \tag{15}$$

where the mean free path l differs slightly from that defined in Ref. 1:

$$\frac{1}{l}=4\pi W_0\left(1+\sum_{k=1}^{[N/2]}\frac{\alpha_{2k}}{2k+1}\right). \tag{16}$$

The probability $W(\theta)$ (14) differs from $W(\theta)$ appearing in (6) and (7) in the factor $1/v_F$, $[W_0]=[\mathrm{cm}^{-1}]$.

The function $\chi(z,\vartheta)$ satisfies the kinetic equation

$$\cos\vartheta\frac{\partial\chi(z,\vartheta)}{\partial z}+\frac{1}{l}\chi(z,\vartheta)-\int_0^{\pi}\int_0^{2\pi}W(\theta)\chi(z,\vartheta')\sin\vartheta'\cos(\varphi-\varphi')d\varphi'd\vartheta'=E(z)\sin\vartheta. \tag{17}$$

The angle θ can be expressed in terms of the angles ϑ,ϑ' and φ,φ' connected with the vector $\mathbf{n}$ and $\mathbf{n}'$:

$$\cos\theta=\cos\vartheta\cos\vartheta'+\sin\vartheta\sin\vartheta'\cos(\varphi-\varphi'). \tag{18}$$

Although the specific results were obtained by us for specular reflection of electrons at the sample boundary, we will first consider a more general case generalizing the well-known Fuchsconditions:[6]

$$\chi(0,\vartheta)=\int_{\pi/2}^{\pi}\sin\vartheta' Q(\vartheta,\vartheta')\chi(0,\vartheta')d\vartheta',\quad 0\leqslant\vartheta\leqslant\pi/2. \tag{19}$$

The matrix $Q(\vartheta,\vartheta')$ can be defined from an analysis of electron scattering at a rough boundary.[7] A transition to the Fuchs condition is carried out through the substitution of $Q(\vartheta,\vartheta')$ in the form of the δ-function:

$$\chi(0,\vartheta)=Q\chi(0,\pi-\vartheta),\quad 0\leqslant Q\leqslant 1. \tag{20}$$

The substitution of the function (14) for $W(\theta)$ in (8) leads to a system of ordinary integral equations supplemented with a differential equation which must be considered on the semiaxis $z>0$ instead of a single integrodifferential equation. The number of such equations is $N=1+(n/2)^2$ (see Appendix 1).

If we take the Fuchs conditions (20) instead of the boundary conditions (19), the matrix kernel of the system is simplified considerably and assumes a special form of the sum of two kernels, the difference and the ''sum'' kernel.

Finally, if we confine the analysis to the case of specular reflection ($Q=1$), the system under investigation is converted (after transformations) into a system of integral equations on the entire axis with a difference kernel, which can easily be solved explicitly.

In Ref. 1, the following expression was obtained for impedance in the case of specular reflection [it should be recalled that the probability $W(\theta)$ is defined by formula (8)]:

$$\zeta=\frac{\omega l}{i\pi c}\int_{-\infty}^{\infty}\frac{d\kappa}{\kappa^2-2i(l^2/\delta^2)\mu(\kappa,\alpha)}, \tag{21}$$

where

$$\mu(\kappa,\alpha)=\frac{\overline{K(\kappa)}}{1-\frac{1}{3}\alpha\overline{K(\kappa)}};$$

$$\overline{K(\kappa)}\equiv\frac{3}{2}\int_0^1\frac{1-y^2}{1+\kappa^2y^2}dy=-\frac{3}{2\kappa^2}+\frac{3}{4i\kappa}\left(1+\frac{1}{\kappa^2}\right)\ln\frac{1+i\kappa}{1-i\kappa}, \tag{22}$$

$\delta=c/\sqrt{2\pi\sigma\omega}$, and σ depends on the mean free path $l=(4\pi W_0)^{-1}$.

In the general case [see (14) and (16)], if the reflection is specular, expression (21) remains valid. However, the function $\mu(\kappa,\alpha)$ changes significantly. It depends on all the coefficients $\alpha_1,\alpha_2,\ \dots\ ,\alpha_n$, i.e., $\alpha=\{\alpha_1,\ \dots\ ,\alpha_n\}$. Moreover, the mean free path is renormalized [see (16)]. The same function $\mu(\kappa,\alpha)$ characterizes the electric field strength $E=E(z)$.

The computational algorithm for the function $\mu(\kappa,\alpha)$ is derived in Appendices I–III. It has the form

$$\mu(\kappa,\alpha)=\frac{3}{4}\frac{D_{01}(\kappa,\alpha)}{D(\kappa,\alpha)},\quad \alpha=\{\alpha_1,\ \dots\ ,\alpha_n\}, \tag{23}$$

where $D(\kappa,\alpha)$ is the determinant composed of the elements

$$D_{(pq)(p'q')}(l\kappa)=\delta_{p-p'}\delta_{q-q'}-lW_0A_{p'q'}\Pi_{p+p',q+q'}(l\kappa). \tag{24}$$

The parameter α will be omitted. Since the unknowns j_{pq} in the system of equations (A.10) are labeled by two indices, the rows and the columns of the determinant are accordingly labeled by two indices; $D_{01}(l\kappa)$ is the determinant which is obtained from D if we replace the column (0,1) in it by the column $\Pi_{p,q+1}(l\kappa)$.

The functions $\Pi_{p+p',q+q'}$ are defined as follows:

$$\Pi_{p+p',q+q'}(\lambda)\equiv\begin{cases}2\displaystyle\int_0^1(1-t^2)^{(q+q')/2}\frac{t^{p+p'}}{1+\lambda^2t^2}dt, & p+p'\ \text{is even;}\\ 2i\lambda\displaystyle\int_0^1(1-t^2)^{(q+q')/2}\frac{t^{p+p'}+1}{1+\lambda^2t^2}dt, & p+p'\ \text{is odd.}\end{cases} \tag{25}$$

The coefficients have the form

$$A_{p,q}=\alpha_{p+q}C_{p+q}^qB_q;$$

$$B_q=\begin{cases}0, & q\ \text{is even;}\\ \dfrac{\pi}{2q}C_{q+1}^{(q+1)/2}, & q\ \text{is odd;}\end{cases} \tag{26}$$

where C_q^p are binomial coefficients.

In the same approximation in which expression (21) for impedance was obtained, the electric field strength in a metal is given by

$$E(z)=-\frac{i\omega l}{c}H(0)I(z), \tag{27}$$

where

$$I(z)=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\exp(-iz\kappa/l)d\kappa}{\kappa^2-2i(l^2/\delta^2)\mu(\kappa)}. \tag{28}$$

In this notation, we have

$$\zeta=\frac{\omega l}{ic}I(0). \tag{29}$$

Formulas (28) and (29) together with formulas (23)–(26) complete the solution of the problem formulated above. Naturally, for a large number of coefficients $\alpha(n>1)$, these formulas are cumbersome (see below). Besides, it is inconvenient to use integral (28) directly for deriving the asymptotic formulas.

If, for example, $z\gg l$, the integrand function oscillates rapidly so that the integral itself is a sort of ''difference effect.'' If l/δ is very small or very large, it is natural to try to expand the integrand into a power series in l/δ of δ/l respectively. This, however, is not admissible since the ratio of κ^2 to $2i(l^2/\delta^2)\mu(\kappa)$ is not small or large uniformly in κ irrespective of the ratio l/δ. For this reason, it is expedient to deform the integration contour by taking it away from the point $\kappa=0$ and bringing it close to the negative imaginary semiaxis, where the factor $\exp(-i\kappa z/l)$ stops oscillating and decreases rapidly for large z/l.

It should be borne in mind that the function $\mu(\kappa)$ has two branching points $\kappa=\pm i$. This follows from the explicit expressions (25) for the functions $\Pi_{\dots}(\kappa)$ and the function $\mu(\kappa)$ which contain the logarithm $\ln[(1+i\kappa)/(1-i\kappa)]$ (this can be verified easily).

In addition, it was found that for the given deformation, the contour (its left wing) intersects the denominator zero at a single point, which leads to the emergence of a residue.

We denote by $\ln^{(+)}[(1+i\kappa)/(1-i\kappa)]$ and $\ln^{(-)}[(1+i\kappa)/(1-i\kappa)]$ two analytic functions in the ring $|\kappa|>1$, such that the function $\ln[(1+i\kappa)/(1-i\kappa)]$ coincide with the function $\ln^{(+)}[(1+i\kappa)/(1-i\kappa)]$ on the semiaxis $\kappa>1$ and with the function $\ln^{(-)}[(1+i\kappa)/(1-i\kappa)]$ on the negative semiaxis. It can easily be seen that these two functions are defined by the formulas

$$\ln^{(\pm)}\left(\frac{1+i\kappa}{1-i\kappa}\right)=\pm\pi i+\ln\left(\frac{1+1/i\kappa}{1-1/i\kappa}\right),\quad |\kappa|>1.$$

Replacing the function $\ln[(1+i\kappa)/(1-i\kappa)]$ in formulas (25) by the function $\ln^{(+)}[(1+i\kappa)/(1-i\kappa)]$ or by $\ln^{(-)}\times[(1+i\kappa)/(1-i\kappa)]$, we obtain analytic continuations of the functions $\Pi_{\dots}$, and hence of the function $\mu(\kappa)$ [its analytic continuations are $\mu^+(\kappa)$ and $\mu^-(\kappa)$].

We can now bring the integration contour in integral (28) to both banks of the cut $(-i,-i\infty)$. After elementary transformations, this gives

$$I(z)=-i\frac{\exp(-i\kappa_0 z/l)}{\kappa_0-i(l^2/\delta^2)\mu'(\kappa_0)}-3i\frac{l^2}{\delta^2}\Pi(z),$$
$$\Pi(z)=\int_1^\infty\frac{(1/t)(1-1/t^2)\exp(-tz/l)dt}{[t^2D_-(-it)+\frac{3}{2}i(l^2/\delta^2)D_{01}^-(-it)][t^2D_+(-it)+\frac{3}{2}i(l^2/\delta^2)D_{01}^+(-it)]}. \tag{30}$$

These functions are suitable for calculating the asymptotic forms, but as before, these formulas are very cumbersome in view of the structure of the functions $D(\kappa)$ and $D_{01}(\kappa)$ [see (24)].

A natural simplification can be made if only the coefficients α_1 and α_2 differ from zero:

$$W(\theta)=W_0(1+\alpha_1\cos\theta+\alpha_2\cos^2\theta), \tag{31}$$

i.e., if the s-, p- and d-scattering is taken into account.

In this case, we have

$$\mu(\kappa,\alpha)=\frac{3}{4}\frac{\Pi_{02}(\kappa)-\frac{4}{3}\beta_2\Pi_{22}(\kappa)}{1-\beta_1\Pi_{02}(\kappa)-\beta_2(1-\frac{4}{3}\beta_1)\Pi_{22}(\kappa)}, \tag{32}$$

where

$$\beta_1=\frac{\alpha_1}{4(1+\alpha_2/3)};\quad \beta_2=\frac{\alpha_2}{2(1+\alpha_2/3)}, \tag{33}$$

and according to the general formulas (25), the functions $\Pi_{02}(\kappa)$ and $\Pi_{22}(\kappa)$ are given by

$$\Pi_{02}(\kappa)=-\frac{2}{\kappa^2}+\frac{1}{i\kappa}\left(1+\frac{1}{\kappa^2}\right)\ln\frac{1+i\kappa}{1-i\kappa}.$$
$$\pi_{22}(\kappa)=\frac{1}{2\kappa^4}+\frac{4}{3\kappa^2}-\frac{1}{i\kappa^3}\left(1+\frac{1}{\kappa^2}\right)\ln\frac{1+i\kappa}{1-i\kappa}. \tag{34}$$

The determinants [see (23) and (24)] contained the function $\Pi_{12}(\kappa)$. By virtue of the relation

$$\Pi_{12}(\kappa)=2i\kappa\Pi_{02}(\kappa)$$

this function does not appear in the final result. For the same reason, the determinants [the numerator and denominator in formula (32)] are linear functions of $\Pi_{02}(\kappa)$ and $\Pi_{22}(\kappa)$.

Substituting expression (32) into formula (21) and its corollaries, we obtain an expression for the impedance which is valid for the probability $W(\theta)$ taking into account the s-, p-, and d-scattering. Naturally, this expression differs from the formula describing only the s- and p-scattering [see (2) and (21)]. The difference is manifested most clearly if the functions Π_{02} and Π_{22} are expressed in terms of the component of the Fourier operator of conductivity $K(\kappa)$:

$$\Pi_{02}(\kappa)=\frac{4}{3}\overline{K(\kappa)};\quad \Pi_{22}(\kappa)=\frac{4}{3\kappa^2}[1-\overline{K(\kappa)}]. \tag{35}$$

Putting α_2 (or β_2)$=0$, we return to the first formula in (22), and hence to all the corollaries analyzed in Ref. 1.

5. Analyzing expressions (21), (32), and (35) with the function $K(\kappa)$ defined in (22), we can easily see that, at low frequencies (in the limit $l/\delta\to 0$), we arrive at the conventional value of impedance:

$$\zeta_{\rm norm}=\sqrt{\omega/4\pi\sigma_{\rm tr}}, \tag{36}$$

where

$$\sigma_{\rm tr}=\sigma\frac{l_{\rm tr}}{l},\quad l_{\rm tr}^{-1}=\int\int W(\theta)(1-\cos\theta)dO. \tag{37}$$

In the given case, we can write

$$\frac{1}{l_{\rm tr}}=\frac{1}{l}\left(1-\frac{\alpha_1-\alpha_2}{3}\right);\quad \frac{1}{l}=4\pi W_0. \tag{38}$$

Refining the classical formulas (36)–(38), we must take into account the relative correction $\sim l^2/\delta^2$. We proceed from formula (30) for the quantity I(z) which is connected with impedance through expression (29). The first term in (30) for $z=0$ can be easily evaluated with the required accuracy

$$-\frac{i}{k_0-i(l^2/\delta^2)\mu'(k_0)}=\frac{1+i}{2\sqrt{\mu(0)}}\frac{\delta}{l}\left(1+i\frac{l^2}{\delta^2}\frac{\mu''(0)}{2}\right),$$

where the function $\mu(\kappa)$ is defined by (23). The second term in (30) has the order of smallness l^2/δ^2 and hence can be omitted. Thus, we can write [ ]

$$\zeta_{\rm spec}=\frac{1+i}{2}\frac{\omega\delta}{c}\frac{1}{\sqrt{\mu(0)}}\left(1+i\frac{l^2}{\delta^2}\frac{\mu''(0)}{2}\right).\quad l\ll\delta. \tag{39}$$

Let us use this general formula in the case when the probability $W(\theta)$ has the special form (31), and the function $\mu(\kappa)$ is defined by (32). In order to calculate $\mu(0)$ and $\mu''(0)$, it is convenient to use the integral representation of the functions $\Pi_{02}(\kappa)$ and $\Pi_{22}(\kappa)$:

$$\Pi_{02}(\kappa)=2\int_0^1(1-t^2)\frac{dt}{1+\kappa^2t^2};$$
$$\Pi_{22}(\kappa)=2\int_0^1(1-t^2)\frac{t^2dt}{1+\kappa^2t^2} \tag{40}$$

[see (25)]. Elementary calculations lead to

$$\zeta_{\rm spec}=\frac{1-i}{2}\sqrt{\omega/2\pi\sigma_{\rm tr}}\left(1-\frac{1}{5}i\frac{l_{\rm tr}^2}{\delta^2}\frac{1+\alpha_2/3}{1+\alpha_2/5}\right). \tag{41}$$

Pay attention to the fact that the correction term of the parameter α_2 does not enter only into $l_{\rm tr}$. According to formula (32) from Ref. 1, the parameter α_1 appears only in $l_{\rm tr}$.

6. In order to calculate the impedance ζ in the case of extremely anomalous skin effect ($l\gg\delta$), we must know the asymptotic form of the function $\mu=\mu(\kappa)$ for $\kappa\gg 1$. Using formulas (32)–(34) and (41), we obtain

$$\mu(\kappa)\approx\frac{3\pi}{4\kappa}\left[1+\frac{1}{\kappa}\left(\frac{\alpha_1\pi}{4(1+\alpha_2/3)}-\frac{4}{\pi}+\frac{8}{9\pi}\frac{\alpha_2}{1+\alpha_2/3}\right)\right],\quad \kappa>0. \tag{42}$$

Using this formula and (21), we obtain a generalization of formula (33) from Ref. 1:[2)]

$$\zeta_{\text{spec}} \approx \zeta_{\text{an}}^{Q=1}[1+A(\alpha_1,\alpha_2)(\delta^2/2l^2)^{1/3}],$$
$$A(\alpha_1,\alpha_2)=\frac{1}{8}\left(\frac{2}{\pi}\right)^{1/3}(\sqrt{3}+i) \times\left[\frac{4}{\pi}-\frac{\alpha_1\pi}{4(1+\alpha_2/3)}-\frac{8}{9\pi}\frac{\alpha_2}{1+\alpha_2/3}\right]. \tag{43}$$

While deriving this formula, it is convenient to transform the integral (21) to the integral over the semiaxis $\kappa>0$.

7. We can make here the following concluding remarks. First of all, the main result of this publication are not analytic expressions (41) and (43) containing small correction to the well-known limiting formulas, but formulas (21), (23)–(26) which make it possible to calculate impedance in the case of a virtually arbitrary function $W=W(\theta)$ [see (14) and (18)]. Naturally, the larger the number of the terms required for defining the functions $W=W(\theta)$, the more cumbersome the numerical calculations. In our opinion, the formulas derived above can be used in subsequent analysis based on special properties of the scattering function $W=W(\theta)$. We are planning to return to this problem later.

It was mentioned more than once (see Refs. 1 and 2) that the arriving term in the collision integral is responsible for correlation between the scattering at the surface and in the bulk. The formulas derived here (and, naturally, in Ref. 1) confirm this fact. An interesting example can be considered in this connection: for $l\gg\delta$ and for specular reflection of electrons, the main correction depends on the structure of the arriving term [see formula (33) from Ref. 1 and formula (43) of this publication]. According to formula (46) from Ref. 1, which described diffuse scattering for $l/\delta\gg 1$, the term containing the parameter α and appearing due to the introduction of the arriving curve is added to the logarithmically large term $\ln\lambda[\lambda=(3\pi l^2/2\delta^2)^{1/3}]$, i.e., its role is much less significant than in the case of specular reflection of electrons at the boundary.

And the last remark: for $l\gg\delta$, the parameters associated with the arriving term appear only in correction terms. If we confine the analysis to the extremely anomalous skin effect, we can apparently use the perturbation theory, considering that the integral term in Eq. (17) is a perturbation. Such an approach might also be used in subsequent investigations.

APPENDIX I

Let us demonstrate how the kinetic equation (17) with the boundary conditions (19) can be transformed into an equivalent system of integral equations on the semiaxis. We assume that the function $E(z)$ is known. At the first step, we solve formally Eq. (17) which will be first presented in the form

$$\cos\vartheta\frac{\partial\chi}{\partial z}+\frac{1}{l}=S(z,\vartheta),$$

where [see (17) and (18)]

$$S(z,\vartheta)=E(z)\sin\vartheta + \int_0^{\pi}\sin\vartheta' d\vartheta'\int_0^{2\pi}d\varphi' W(\theta)\chi(z,\theta')\cos\varphi'. \tag{A1}$$

This gives

$$\chi(z,\vartheta)=\begin{cases}\chi(0,\vartheta)\exp\left(-\dfrac{z}{l\cos\vartheta}\right) + \displaystyle\int_0^z \exp\left(\frac{\xi-z}{l\cos\vartheta}\right)\frac{S(\xi,\vartheta)}{\cos\vartheta}d\xi; \\ 0\leqslant\vartheta\leqslant\pi/2; \\ -\displaystyle\int_z^{\infty}\exp\left(\frac{\xi-z}{l\cos\vartheta}\right)\frac{S(\xi,\vartheta)}{\cos\vartheta}d\xi; \\ \pi/2\leqslant\vartheta\leqslant\pi.\end{cases} \tag{A2}$$

According to the boundary condition (13), the term $\chi(0,\vartheta)\exp(-z/l\cos\vartheta)$ can be written in the form

$$\chi(0,\vartheta)\exp\left(-\frac{z}{l\cos\vartheta}\right) = -\int_{\pi/2}^{\pi}\sin\vartheta' d\vartheta' Q(\vartheta,\vartheta') \times\int_0^{\infty}\exp\left(\frac{\xi}{l\cos\vartheta'}-\frac{z}{l\cos\vartheta}\right)\frac{S(\xi,\vartheta')}{\cos\vartheta'}d\xi.$$

Let us consider the structure of the function $S(z,\vartheta)$ in greater detail. Using the relation (18) between θ and ϑ,ϑ', and φ, we can write

$$W(\theta)=W_0\left[1+\sum_{k=1}^{n}\alpha_k(\cos\vartheta\cos\vartheta' +\sin\vartheta\sin\vartheta'\cos\varphi)^k\right] =W_0\left[1+\sum_{p+q\leqslant n}\alpha_{p+q}C_{p+q}^q \times\cos^p\vartheta\cos^p\vartheta'\sin^q\vartheta\sin^q\vartheta'\cos^q\varphi\right],$$

where C_p^q are binomial coefficients.

Substituting these relation into (17), we obtain

$$S(z,\vartheta)=E(z)\sin\vartheta + W_0\sum_{p+q\leqslant n}{}' A_{p,q}\cos^p\vartheta\sin^q\vartheta j_{pq}(z), \tag{A3}$$

where new unknown functions $j_{pq}(z)$ have been introduced:

$$j_{pq}(z)\equiv\int_0^{\pi}\chi(z,\vartheta'')\cos^p\vartheta'\sin^{q+1}\vartheta' d\vartheta' \tag{A4}$$

and the coefficients are defined as

$$A_{p,q}=\alpha_{p+q}C_{p+q}^{q}B_q;$$

$$B_q\equiv\int_0^{2\pi}\cos^{q+1}\varphi\,d\varphi=\begin{cases}0 & q \text{ is even}\\ \dfrac{\pi}{2q}C_{q+1}^{(q+1)/2}, & q \text{ is odd.}\end{cases}$$

It should be noted that the index q in the sum (A3) runs through odd values since $B_q=0$ if q is even. The prime on the sum reminds of this circumstance.

We can now obtain the system of integral equations defining all the functions $j_{pq}(z)$ with an odd q [the functions $j_{pq}(z)$ with an even q are not required]. For this purpose, in the definition (A4) of the function $j_{pq}(z)$ we eliminate the function $\chi(z,\vartheta)$ by using (A2); as a result, $j_{pq}(z)$ can be presented by an integral depending only on $S(\xi,\vartheta)$. The second and last step consists in the replacement of the function $S(\xi,\vartheta)$ in the integrand by its expression (A3) in terms of the functions $j_{p'q'}(\xi)$. This leads to the following fundamental system of integral equations:

$$\begin{aligned}j_{pq}(z)-W_0\sum_{p',q'}{}'A_{p'q'}\int_0^{\infty}j_{p'q'}(\xi)[T_{p+p',q+q'}(z-\xi)\\ +T_{(pq)(p'q')}(z,\xi)]d\xi=\int_0^{\infty}E(\xi)[T_{p,q+1}(z-\xi)\\ +T_{(pq)(01)}(z,\xi)]d\xi.\end{aligned}\tag{A5}$$

The following notation has been introduced:

$$\begin{aligned}T_{pq}(z)\equiv\int_0^{\pi/2}\cos^{p-1}\vartheta\,\sin^{q+1}\vartheta\\ \times\exp\left(-\frac{|z|}{l\cos\vartheta}\right)d\vartheta\,\operatorname{sgn}^{p+1}z,\\ T_{(pq)(p'q')}(z,\xi)\equiv-\int_0^{\pi/2}d\vartheta\int_{\pi/2}^{\pi}d\vartheta' Q(\vartheta,\vartheta')\\ \times\cos^{p}\vartheta\,\sin^{q+1}\vartheta\,\cos^{p'-1}\vartheta'\\ \times\sin^{q'+1}\vartheta'\,\exp\left(\frac{\xi}{l\cos\vartheta'}-\frac{z}{l\cos\vartheta}\right).\end{aligned}\tag{A6}$$

This is the system of N ordinary integral equations on the semiaxis mentioned above.

APPENDIX II

As a result of a transition to boundary conditions of the Fuchs type, these equations remain in force, but the expressions of the kernels $T_{(pq)(p'q')}(z,\xi)$ are simplified:

$$T_{(pq)(p'q')}(z,\xi)=(-1)^{p'}QT_{p+p',q+q'}(z,\xi).$$

Even after such a simplification, however, the system of equations can be solved either for $k=1$ and $Q=0$, or for an arbitrary value of k and $Q=1$, i.e., in the case of specular reflection.

In the latter case, if we supplement the definition of the functions $j_{pq}(z)$ and $E(z)$ by putting

$$j_{pq}(z)=(-1)^p j_{pq}(-z),\quad E(z)=E(-z),\quad z<0,\tag{A7}$$

for negative z, we can reduce the system (A5) to the form

$$\begin{aligned}j_{pq}(z)-W_0\sum_{p',q'}{}'A_{p'q'}\int_{-\infty}^{\infty}j_{p'q'}(\xi)T_{p+p',q+q'}(z-\xi)d\xi\\ =\int_{-\infty}^{\infty}E(\xi)T_{p,q+1}(z-\xi)d\xi,\quad -\infty<z<\infty.\end{aligned}\tag{A8}$$

APPENDIX III

Let us suppose that electrons experience specular reflection at the boundary. The system (A8) can be solved in the standard manner through a transition from the functions $E(z)$ and $j_{pq}(z)$ to their Fourier transforms $E(\kappa)$ and $J_{pq}(\kappa)$. The Fourier transforms of the coefficients $T_{p+p',q+q'}(z)$ can be represented in the form

$$\int_{-\infty}^{\infty}T_{p+p',q+q'}(z)\exp(i\kappa z)dz=l\Pi_{p+p',q+q'}(l\kappa),$$

where

$$\begin{aligned}\Pi_{p+p',q+q'}(\lambda)\\ \equiv\begin{cases}2\displaystyle\int_0^1(1-t^2)^{(q+q')/2}\frac{t^{p+p'}}{1+\lambda^2t^2}dt, & p+p' \text{ is even;}\\ 2i\lambda\displaystyle\int_0^1(1-t^2)^{(q+q')/2}\frac{t^{p+p'}+1}{1+\lambda^2t^2}dt, & p+p' \text{ is odd.}\end{cases}\end{aligned}\tag{A9}$$

The system (A8) becomes an algebraic system of equations

$$\begin{aligned}J_{pq}(\kappa)-lW_0\sum_{p',q'}{}'A_{p'q'}J_{p'q'}(\kappa)\Pi_{p+p',q+q'}(l\kappa)\\ =E(k)\Pi_{p,q+1}(l\kappa).\end{aligned}\tag{A10}$$

Solving this system for $J_{01}(\kappa)$, we obtain

$$J_{01}(\kappa)=\frac{D_{01}(l\kappa)}{D(l\kappa)}E(\kappa),\tag{A11}$$

where $D(l\kappa)$ is a determinant whose elements are given by $D_{(pq)(p'q')}(l\kappa)=\delta_{p-p'}\delta_{q-q'}-lW_0A_{p',q'}\Pi_{p+p',q+q'}(l\kappa)$; and $D_{01}(l\kappa)$ is the determinant which is obtained from $D(l\kappa)$ by replacing the column (0,1) in it by the column $D_{(pq)}=\Pi_{p,q+1}(l\kappa)$.

We denote by $J(\kappa)$ the Fourier transform of the current density $j(z)$. According to definition (A4) of $j_{pq}(z)$, formula (15) indicates that the current density $j(z)$ has the form

$$j(z)=\frac{3\sigma}{4l}j_{01}(z).$$

Consequently, the Fourier transform $J(\kappa)$ of current density is given by

$$J(\kappa)=\frac{3\sigma}{4}\frac{D_{01}(l\kappa)}{D(l\kappa)}E(\kappa).\tag{A12}$$

This is a consequence of the kinetic equation and formula (15) expressing the current density in terms of the function $\chi(z,\vartheta)$.

The second independent relation between Fourier transforms of current density and field is the corollary of the corresponding Maxwell's equation:

$$E''(z)+\frac{4\pi i\omega}{c^2}j(z)=0.$$

Carrying out Fourier transformation and considering that $E'(-0)=-E'(+0)$, we obtain

$$-\kappa^2E(\kappa)-2E'(0)+\frac{4\pi i\omega}{c^2}J(\kappa)=0,$$
$$E'(0)=\frac{i\omega}{c}H_y(0). \tag{A13}$$

Eliminating $J(\kappa)$ from relations (A12) and (A13), we obtain

$$E(\kappa)=-\frac{2(i\omega/c)H_y(0)}{\kappa^2-(3i/2\delta^2)(D_0(l\kappa)/D(l\kappa))};$$

$$\delta^2\equiv\frac{c^2}{2\pi\sigma\omega}. \tag{A14}$$

[1)] The Boltzmann integrodifferential equation requires the formulation of boundary conditions for the conduction electron distribution function. The boundary conditions can be obtained from an analysis of electron interaction with the sample boundary.

[2)] We take the opportunity to note that formula (33) in Ref. 1 is incorrect: the factor $(\delta^2/3l^2)$ should be raised to the power 1/3.

[1] M. I. Kaganov, G. Ya. Lubarskiĭ, and E. Chervonko, Zh. Éksp. Teor. Fiz. **102**, 1351 (1992) [Sov. Phys. JETP **75**, 733 (1992)].
[2] M. I. Kaganov, G. Ya. Lubarskiĭ, and E. Chervonko, Zh. Éksp. Teor. Fiz. **102**, 1563 (1992) [Sov. Phys. JETP **75**, 844 (1992)].
[3] I. M. Lifshits, M. Ya. Azbel, and M. I. Kaganov, *Electron Theory of Metals*, Consultants Bureau, NY, 1973.
[4] G. E. Reuter and H. Sondheimer, Proc. Roy. Soc. **A195**, 336 (1948).
[5] G. Ya. Lyubarskiĭ, in *Complex Variables*, vol. 27, 287 (1955).
[6] K. Fuchs, Proc. Camb. Phil. Soc. **34**, 100 (1938).
[7] L. Falkovsky, Adv. Phys. **32**, 753 (1983).

Translated by R. S. Wadhwa

Diffusion in solid helium (A review)

V. N. Grigor'ev

*B. Verkin Institute for Low Temperature Physics and Engineering, National Academy of Sciences of the Ukraine, 310164 Kharkov, Ukraine**

(Submitted July 4, 1996)
Fiz. Nizk. Temp. **23**, 5–20 (January 1997)

The results of experimental investigations of diffusion processes in solid helium are reviewed. It is shown that ^{3}He impurities in hcp ^{4}He are narrow-band quasiparticles whose motion can be adequately described in the existing theory of quantum diffusion. The observed peculiarities of ^{3}He diffusion in the bcc phase of concentrated ^{3}He–^{4}He solutions, i.e., a noticeable diffusive transport at concentrations higher than critical, the independence of the diffusion coefficient D on the concentration for $x<20\%$ ^{3}He, an abrupt increase in D for $x\simeq 20\%$, and a dependence of D on the diffusion length in the latter case, have not received a quantitative interpretation. The relation between diffusion phenomena in concentrated solutions and the percolation problem is emphasized. The results of experiments on vacancy diffusion are analyzed, and it is proved that vacancies in ^{3}He–^{4}He are wide-band quasiparticles. © *1997 American Institute of Physics.* [S1063-777X(97)00201-6]

INTRODUCTION

Diffusion phenomena is solid helium belong to the field of physical studies whose development was determined by the ideas formulated by I. M. Lifshits. The publication in 1969 of the article ''Quantum Theory of Defects in Crystal'' by Andreev and Lifshits[1] changed radically the existing concepts concerning diffusion processes and determined the direction of research work in the field of quantum crystal physics for many years. The term ''quantum crystal'' was introduced by de Boer[2] in 1948 for substances in which the energy of zero-point vibrations of particles is comparable to the total energy of the crystal. Solid helium was always regarded a typical representative of this class of crystals, but prior to the publication by Andreev and Lifshits[1] no manifestations of quantum effects in its macroscopic properties had been predicted. In Ref. 1, several such effects (quantum diffusion, zero-point vacancions, and the possibility of superfluidity of crystals) were predicted simultaneously, which stimulated experimental and theoretical investigations in solid helium.

The main regularities of quantum diffusion have been observed experimentally after a comparatively short time. In accordance with the predictions of Andreev and Lifshits, it was found that ^{3}He impurities in hcp crystals of ^{4}He are narrow-band quasiparticles whose motion is determined by gas laws in many cases. It has been proved theoretically that vacancies in solid helium must be wide-band quasiparticles, and experimental evidences in favor of this hypothesis have been obtained recently.

Subsequent studies have made it possible to generalize the theory to the range of high impurity concentrations, to analyze the properties of other narrow-band particles, to establish the relation between diffusion and percolation phenomena, and to observe quantum diffusion in other systems.

Dozens of publications appearing in this field allow us to make a brief review of the obtained results. Since the basic aspects of theoretical studies are covered in the review by Kagan,[3] the main attention here will be paid to experimental studies which were accounted for (apart of original papers) in a single review[4] of the results obtained at the early stage of quantum diffusion studies.

The present review has the following construction. Basic theoretical concepts and relations are presented in Sec. 1. The main features of NMR method in solid helium are outlined in Sec. 2. Section 3 is devoted to publications concerning the observation of quantum diffusion of impurities in the hcp phase, and Sec. 4 to the effects of phonon-stimulated diffusion and localization of impurities. Section 5 covers the results of investigations in the bcc phase, while Sec. 6 contains the results of direct measurements of vacancy diffusion. Some general aspects of diffusion phenomena in helium are considered in Conclusion, where all the results are summarized.

1. BASIC CONCEPTS AND THEORETICAL RELATIONS

The main idea put forth by Andreev and Lifshits[1] is based on the assumption that the probability of tunnel exchange of neighboring particles in quantum crystals becomes noticeable in view of the large amplitude of zero-point vibrations of atoms, which leads to the conversion of impurities and point defects into peculiar quasiparticles that can move almost freely over the entire volume. In perfect crystals, the motion of quasiparticles is limited only by collisions with one another and with other excitations. Such a quasi-free motion occurs within the energy band whose width $\Delta\cong zJ$ is determined by the tunneling frequency characterized by the exchange integral J (z is the number of nearest neighbors). For example, the motion of impurity atoms in the case of a low concentration is determined by collisions with phonons, leading to a peculiar diffusion flow under these conditions: the diffusion coefficient D must increase upon a decrease in temperature. Andreev and Lifshits[1] derived the following relation:

$$D\sim\frac{a^2\Delta^2}{h\Theta}\left(\frac{\Theta}{T}\right)^9, \tag{1}$$

where a is the separation between nearest neighbors, Θ the Debye temperature, and h the Planck constant. The emergence of such a large exponent for temperature is not surprising. The number of phonons is proportional to T^3, their cross scattering section is proportional to T^4, and the additional factor $(T/\Theta)^2$ associated with low efficiency of collisions appears in view of the difference in the momenta of phonons and impuritons.

Expression (1) was obtained on the basis of a purely qualitative analysis. Subsequent calculations made by Pushkarov[5] showed that the numerical coefficient in this formula is significant ($\sim 10^6$) since the role of Θ is played by an effective temperature approximately equal to $(\Theta/8)$. Later, Slyusarev *et al.*[4] carried out a quantitative analysis for the hcp phase taking into account possible anisotropy and derived the relation

$$D = Aa^2(I^2/\Theta)(\Theta/T)^9, \tag{2}$$

where A is a numerical coefficient equal to $2.3 \cdot 10^5$ for the hcp phase.

An unexpected result was obtained by Kagan and Klinger[7] who proved that the coherent motion of impuritons described by the dependence of type (1) is preserved up to $T \simeq \Theta$ even when the impuriton mean free path is of the order of or smaller than the lattice parameter.

The interaction between impurities inevitably becomes significant upon an increase in the impurity concentration or a decrease in temperature. The situation in which such an interaction plays a decisive role was analyzed simultaneously and independently in Refs. 8–10. It was proved that in the case of gas-like motion, the impurity diffusion coefficient can be written in the form

$$D \simeq a^4 J/h\sigma x, \tag{3}$$

where x is the impurity concentration and σ the impurity scattering cross section.

The problem of determining the cross section of impuritons, which was solved most consistently by Kagan,[11] proved to be very important. The analysis[11] showed that the scattering cross section is quite large ($\sigma \gg a^2$) and depends on the impuriton band width. Such a dependence can be explained qualitatively on the basis of the following considerations. Owing to a large amplitude of zero-point vibrations, the ^{3}He impurity atom occupies a slightly larger volume in the lattice than the ^{4}He atom. This leads to a lattice distortion and to the emergence of elastic interaction between distorted regions. The presence of such an interaction prevents the convergence of two impurity atoms to a distance at which the potential energy exceeds the energy band width. The interaction potential is described by the law $U = U_0(a/r)^3$, and the characteristic distance r_0 at which the interaction energy coincides with the band width is given by

$$r_0 \simeq a(U_0/\Delta)^{1/3}. \tag{4}$$

Obviously, r_0^2 plays the role of σ, and the substitution of (4) into (3) gives

$$D \sim (Ja^2/x)(J/U_0)^{2/3}. \tag{5}$$

According to Slyusarev and Strzhemechny,[12] the inclusion of interaction anisotropy reduces the scattering cross section by a factor of several units, and the final expression for diffusion coefficient in the region of impurity scattering can be written in the form

$$D = \frac{k_0^{-1} z^{5/6} v_0 J a^3}{\Pi^3 \sqrt{2}\Gamma(1/3) a^2 \hbar x} \left(\frac{J}{U_0}\right)^{2/3} \tag{6}$$

(k_0 is the numerical factor taking into account the decrease in the scattering cross section for an anisotropic potential and depending on the crystal lattice type; $k_0 \simeq 0.29$ for the hcp phase).

It is very important to determine the limits of applicability of the expressions obtained above. As regards the temperature, it was noted above[7] that it is virtually unbounded. It should be noted, however, that at high temperatures the quantum diffusion of impurities can be ''screened'' by an exponentially increasing vacancy diffusion which can also be unusual in quantum crystals.

The estimation of the concentration limit is based on a comparison of the average energy of interaction of impurities and the energy band width $U(R) = \Delta$, where $R \sim a/x_0^{1/3}$. A correct estimate of the critical concentration ($x_0 \leqslant 10^{-3}$) was obtained by Guyer *et al.*[13] In view of the large value of $\delta\Delta/\delta\rho$, the concentration x_0 strongly depends on density (see below).

An analysis of the diffusion flow under the conditions when $U(R)$ is slightly larger than Δ was carried out by Andreev[14] who paid attention to the fact that the violation of the condition $U(r) < \Delta$ does not rule out the tunnel exchange which terminates only for $a(\delta U/\delta r) > \Delta$. Andreev[14] proved that diffusion transport for $a(\delta U/\delta r) < \Delta < U$ follows the law

$$D \sim (J^2 a^2/\hbar U_0) x^{-4/3}. \tag{7}$$

This dependence is close to (3), and the two dependences and can hardly be distinguished in experiments, the more so that the range of applicability of (7) is rather small.

A further increase in impurity concentration leads to a considerable expansion of the regions in which quantum diffusion is impossible. As a result, the larger and larger fraction of the crystal is excepted from free motion, clusters of stationary impurities are formed, diffusion becomes of percolation nature, and impurities are localized completely for a certain critical impurity concentration x_c. Such a situation was analyzed by Kagan and Maksimiv[15] who demonstrated the dual role of temperature. On one hand, in the case of coherent motion of impuritons, phonons decelerate the diffusion flow, while on the other hand, in the case of a finite temperature, the localized impuritons can become delocalized due to their interaction with phonons and can subsequently participate in diffusion transport. Taking into account these mechanisms, we can write the diffusion coefficient in the form

$$D = \frac{z a^2 J^2 \Omega_{\rm ph}(T)}{3\hbar[\delta\varepsilon^2 + \Omega_{\rm ph}^2(T)]}, \tag{8}$$

where $\Omega_{\rm ph}=A(\Theta)(T/\Theta)^9$ is the amplitude of impuriton scattering by a phonon, A is a constant of the order of 10^6, and

$$\delta\varepsilon=\alpha U_0 x^{4/3}, \tag{9}$$

where $\delta\varepsilon$ is the mismatching of energy levels at neighboring lattice sites due to interaction between impurities. Kagan and Maksimov[15] also proposed an interpolation formula taking into account various mechanisms and describing the diffusion of impurities in a wide range of concentrations, temperatures, and densities:

$$D\simeq\frac{za^2J^2}{3\hbar}\left[\frac{Q(x)}{\Omega_x+\Omega_{\rm ph}(T)}+\frac{[1-Q(x)]\Omega_{\rm ph}(T)}{\delta\varepsilon^2+\Omega_{\rm ph}(T)}\right]. \tag{10}$$

Here

$$Q(x)=[(x_c-x)/x_c]^t \tag{11}$$

is the percolation function characterizing the fraction of regions where free motion can occur, and

$$\Omega_x=BxU^{2/3}I^{1/3} \tag{12}$$

is the amplitude of scattering of an impuriton by another impuriton ($B=$const). It can be easily seen that equations (10) is transformed into (1), (5), or (8) in limiting cases.

In the region of extremely high temperatures, we must supplement (10) with the term $D_{xv}=D_0\exp(-W/T)$ describing the vacancion diffusion of impurities (see Sec. 6 for details).

2. PECULIARITIES OF NMR MEASURING TECHNIQUE IN ^{3}He–^{4}He SOLID SOLUTIONS

The main experimental results on diffusion of impurities in solid helium were obtained by the NMR method based on spin echo. Spin echo signals appear in spin systems in a nonuniform magnetic field as a result of action of several radio pulses filled with resonant Larmor frequency. The maximum amplitude of echo signal is obtained when two pulses (90 and 180°) are used. This notation characterizes pulses whose amplitudes and durations are selected so that the magnetic moments of the sample are rotated under their action through 90 and 180° respectively. The amplitudes h of echo signals generated in this case are described by the following dependence:

$$h=h_0\exp[-2\tau/T_2-(2/3)\gamma^2G^2D\tau^3]. \tag{13}$$

Here τ is the time interval between 90 and 180° pulses, h_0 is the amplitude of a free induction signal emerging after the 90°-pulse, γ the gyromagnetic ratio, G the magnetic field gradient, and T_2 the spin–spin relaxation time. Relation (13) allows us to determine the value of D directly from the dependence of then echo-signal amplitude on G or τ. The diffusion coefficient measured in this case corresponds to spin diffusion and can differ from the self-diffusion coefficient if the mechanisms of spin transfer without a displacement of atoms are significant. Such processes can occur due to the exchange or dipole–dipole interaction. However, these processes are significant only between nearest neighbors, and hence make a noticeable contribution only when the concentration of ^{3}He in ^{3}He–^{4}He solutions are of the order of several percent. For weak solutions, the measured values of spin diffusion coefficient coincides with the self-diffusion coefficient.

Peculiarities in the application of the spin echo method in the case of weak solutions of ^{3}He in ^{4}He are due to the requirement of combination of the high sensitivity of the setup with the possibility of measuring small values of diffusion coefficient for anomalously long spin–lattice relaxation time T_1 which can be as large as several hours for the most dilute solutions. In this case, the operation under the conditions of equilibrium magnetization of the samples is virtually impossible, and it must be ensured that the value of h_0 in (13) be constant in a given series of measurements. In many cases, this problem was solved by using the method of preliminary saturation.[16] In this method, the sample was subjected to the action of a series of high-power pulses levelling out the population densities of spin sublevels prior to each measurement. If the probing 90 and 180° pulses are applied to the sample with a time interval t_1 which is constant for a given series after switching off the pulses, when the sample magnetization is restores according to the conventional exponential law, the constancy of h_0 whose role is played by the quantity $h_\infty[1-\exp(t_1/T_1)]$ will be ensured.

However, this method is inapplicable to the most dilute solutions in view of a considerable decrease in the amplitude of signals being measured. In this case, for a given t_1 the amplitude decreases in proportion to x^2 (on account the dependence $T_1^{-1}\sim x$). We used an original approach associated with melting of the sample.[17] Intrinsic relaxation times for solid samples are also long, but boundary relaxation processes reducing the value of T_1 to a few seconds play a significant role in the liquid due to rapid diffusion. Before each measurement, the samples of the most dilute solutions were melted by reducing pressure and then crystallized, which allowed us to obtain virtually equilibrium samples during several minutes. (Carrying out melting and crystallization at a lower temperature, we could even obtain the samples with a magnetization higher than the equilibrium value.) The approaches described above made it possible to carry out reliable measurements for low concentrations of ^{3}He down to $6\cdot10^{-5}$.

The conventional method of spin echo using a sequence of 90–180° pulses makes it possible to measure the diffusion coefficient down to 10^{-8} cm^2/s. The method using stimulated echo[18,19] emerging after the action of three 90°-pulses appears to be more promising for measuring lower values of D. In this case, the echo amplitude can be described by the expression

$$h=\frac{h_0}{2}\exp\left[\frac{-2\tau_1}{T_2}-\frac{\tau_2-\tau_1}{T_1}-\gamma^2D\tau_1^2\left(\tau_2-\frac{1}{3}\tau_1\right)\right], \tag{14}$$

where τ_1 is the interval between the first and second pulses and τ_2 between the second and third pulses. The contribution of diffusion to signal damping depends on both signals. This allows us to reduce significantly the role of relaxation damping, thus lowering the limit of measurement of low diffusion coefficients. The gain is especially large in the case of

^{3}He–^{4}He solid solutions due to a very large value of the ratio $T_1/T_2 \geqslant 10^3$. This allows us to measure D down to 10^{-10} cm^2/s.

The limit of the measured values of D can be reduced further only by increasing the applied magnetic field gradient. However, the increase in G under stationary conditions is limited in view of a strong narrowing of echo, necessitating a significant expansion of the transmission band of the receiver, as well as due to an increasing role of spin phase mismatching during the action of pulses. These difficulties can be avoided by using the pulsed gradient applied in the interval between pulses. If in this case the value of the pulsed magnetic field gradient can be made much larger than the constant field gradient, and the interval between the HF field pulses and the gradient can be made short, the expressions for the signal amplitude remain virtually unchanged. The application of pulsed gradient does not impose formal limitations on its magnitude, and hence there are no restrictions on measurement of small diffusion coefficients. In actual practice, the limit of measurements in concentrated solutions could be reduced by two more orders of magnitude to 10^{-12}cm^2/s.

The development and applications of the approaches described above has made it possible to study diffusion in ^{3}He–^{4}He solid solutions in a wide range of concentrations, temperatures, and densities and to discover a large number of interesting and unexpected phenomena.

3. QUANTUM DIFFUSION OF ^{3}He IMPURITY IN hcp PHASE

The first evidence of quantum-mechanical motion of impurities was obtained while studying the concentration dependence of the diffusion coefficient of ^{3}He in the hcp phase of solid helium near the melting curve at $V=21.0$ cm^3/mole. The experiments were carried out by Richards, Pope, and Widom[20] at the Sussex University (UK) at 0.54 K in the concentration range 0.22–3% and by Grigorev, Eselson, Mikheev, and Shulman[21] at B. Verkin Institute for Low Temperature Physics and Engineering, National Academy of Sciences of the Ukraine, who studied the temperature dependence of diffusion coefficient in the temperature interval 0.4–1.4 K for four solutions with ^{3}He concentrations varying from 0.09 to 2.17%. In both publications, it was found that the diffusion coefficient in the region under investigation varies in inverse proportion to concentration according to the predicted dependence (3). In Ref. 21, it was also proved that the diffusion coefficient of the solutions is virtually independent of temperature. The upper temperature limit in these experiments was confined to the hcp–bcc transition.

The next significant step was made in experiments at a higher value of density, in which measurements could be made up to the melting point.[22] The main results of this research were the establishment of a stronger dependence of the diffusion coefficient on density (in comparison with a similar dependence of the spin diffusion coefficient in pure ^{3}He) as well as an exponential increase in the diffusion coefficient observed at higher temperatures (Fig. 1). The first fact could be regarded as a confirmation of a nonlinear dependence of the diffusion coefficient on the exchange integral according to (5), while the second can indicate a significant contribution of vacancion diffusion under these conditions. Similar experiments were carried out later at higher densities up to $V=19.9$ cm^3/mole.[23,24] The concentration dependence constructed according to the results obtained by the Kharkov group are presented in Fig. 2 (this dependence is described by the relation $Dx=10^{-11}$cm^2/s). The results obtained by Richards *et al.* follow in general the same dependence with a slightly differing numerical coefficient. The difference in the absolute values of diffusion coefficient is in all probability due to a strong dependence of the results obtained by the spin echo technique on the magnetic field gradient used, which is difficult to measure with the required accuracy. It should also be noted that considerable deviations from the law $D \sim x^{-1}$ were observed in one of the experiments carried out in Sussex.[24] Possible reasons behind such a behavior will be considered below.

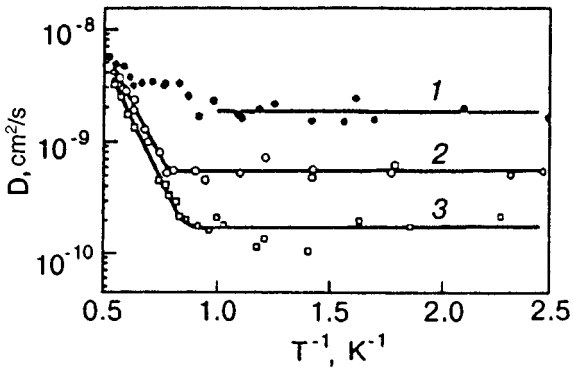

FIG. 1. Temperature dependence of the diffusion coefficient of ^{3}He in the hcp phase of ^{4}He for $V=20.7$ cm^3/mole in solutions with various concentrations of ^{3}He,%: 0.25 (curve *1*), 0.75 (curve *2*), and 2.17 (curve *3*).

The experimental observation of the predicted dependence $D \sim x^{-1}$ suggests that quantum diffusion of ^{3}He impurity takes place in the hcp phase of solid helium. However, Huang *et al.*[25] proved that almost the same dependence can be obtained in the concentration range 10^{-2}–10^{-3} in the framework of the ''model of interaction'' developed by them

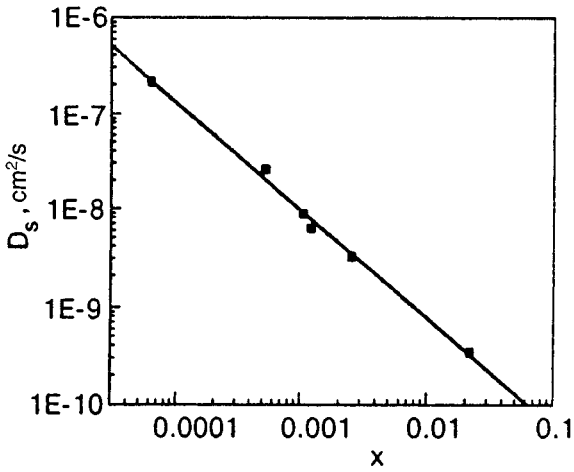

FIG. 2. Concentration dependence of the diffusion coefficient of ^{3}He in the hcp phase of ^{4}He for $V=21$ cm^3/mole in the plateau region.

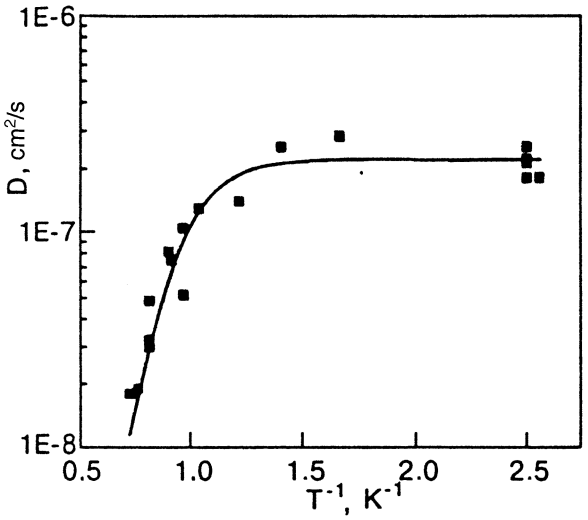

FIG. 3. Temperature dependence of the diffusion coefficient of ^{3}He in the hcp solution with 0.006%^{3}He for $V=21$ cm^3/mole.[26]

without using the concepts of coherent motion of impurity excitations. This circumstance was an additional impetus to carry out investigations in the region of lower concentrations, aimed at observing the dependence $D \sim T^{-9}$ associated with phonon scattering of impuritons and most typical of quantum diffusion.

Such investigations were carried out for the first time at B. Verkin Institute for Low Temperature Physics and Engineering on a solution containing 0.006%^{3}He,[17,26] and later at Sussex[27,28] for concentrations up to 10^{-2}%. A sharp increase (almost by an order of magnitude for dilute solutions) in the diffusion coefficient upon a decrease in temperature was detected. The results of these experiments are presented in Fig. 3. If we disregard the traditional difference in absolute values, the results of both experiments (as well as of subsequent experiments) indicate that the diffusion coefficient in the temperature range 0.8–1.4 K varies in proportion to T^{-9} in accordance with the predictions of the theory. Figure 4 showing the dependence of D^{-1} on T^9 for a 0.06%^{3}He solution illustrates the extent to which the experimental results fit into dependence (1). The results of these experiments unambiguously confirmed the existence of quantum diffusion in weak ^{3}He solutions in the hcp phase of ^{4}He.

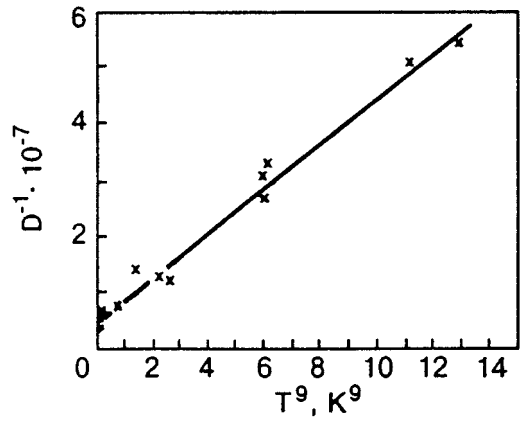

FIG. 4. Dependence of the reciprocal diffusion coefficient of ^{3}He in the hcp solution with 0.006%^{3}He on T^9 (replotted curve in Fig. 3).

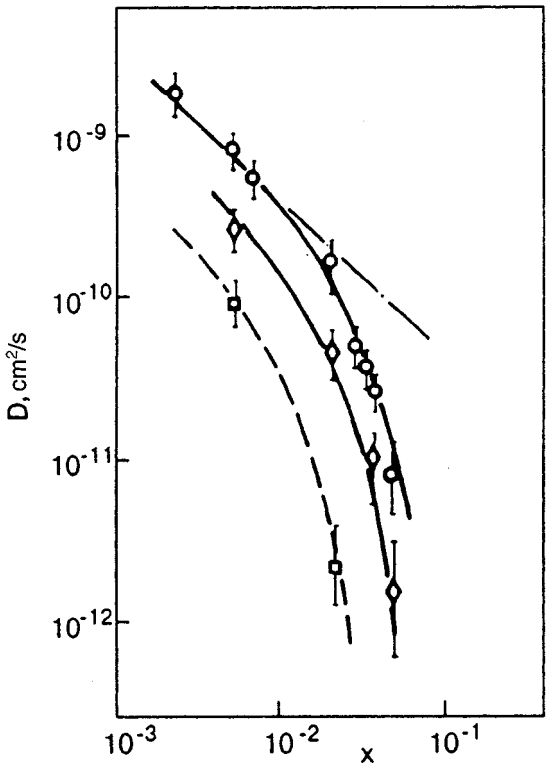

FIG. 5. Concentration dependence of the diffusion coefficient of ^{3}He for various molar volumes V, cm^3/mole: 20.7 (○), 20.5 (◇), and 19.9 (□). The curves correspond to the dependence (11) for $t=1.7$. The dot-and-dash straight line corresponds to the dependence $D \sim x^{-1}$.[23]

4. LOCALIZATION OF IMPURITIES AND PHONON-STIMULATED DIFFUSION

The next important step in experimental investigation was associated with analysis of the behavior of diffusion coefficient with increasing ^{3}He concentration and with observation of impurity localization and phonon-stimulated diffusion predicted by Kagan and Maksimov.[15] Such experiments were made by Mikheev, Maidanov, and Mikhin[23,29] in the concentration range up to 5%^{3}He and for molar volumes 19.9–20.7 cm^3/mole. A tendency to impurity localization was manifested above all in a stronger (as compared to $D \sim x^{-1}$) concentration dependence of diffusion coefficient in the plateau region for $x \geqslant 2$%^{3}He (Fig. 5), which can be naturally explained by a decrease in the factor Q_x in formula (10) with increasing concentration. However, the most reliable data on localization were obtained while studying the dependence of diffusion coefficient on density at a constant concentration. Figure 6 shows the corresponding data for a 2.17%^{3}He solution. It can be clearly seen that the diffusion coefficient decreases abruptly upon a decrease in molar volume associated with a decrease in the value of x_c in (11). The same figure illustrates the formation of a plateau at $D \cong 2 \cdot 10^{-12}$cm^2/s, which can be naturally attributed to the contribution of spin dipole diffusion. The calculation of this

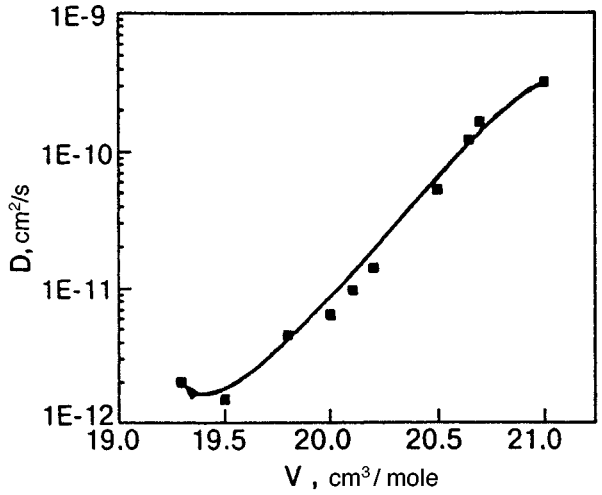

FIG. 6. Dependence of the diffusion coefficient of ^{3}He in a solution with 2.17%^{3}He on molar volume in the plateau region.

contribution was carried out by Slusarev and Strzhemechny[30] and resulted in a value close to that determined experimentally.[1)]

The processing of the obtained results with the help of formulas (9) and (10) made it possible to determine the parameters x_c and t. For $V=20.7$ cm^3/mole, the values of these parameters for the most thoroughly investigated samples were $x_c=(7\pm2)\%^3$He and $t=1.7\pm0.2$. It should be noted that the latter value virtually coincides with the critical index obtained form an analysis of the metal–insulator transition (see, for example, Ref. 31).

Manifestations of phonon-stimulated diffusion were observed while studying the temperature dependence of diffusion coefficient in a sample containing 4%^{3}He (Fig. 7). A sharp increase in diffusion coefficient was observed upon an increase in temperature to $T\sim1$ K, for which the contribution of vacancion diffusion is comparatively small. The observation of all basic mechanisms of quantum diffusion of impurities has made it possible to carry out a complex processing of experimental data on the basis of formula (10) in the entire range of concentrations, temperatures, and molar volumes under investigation.[29] Such a processing gave matching values of basic parameters I, U_0, W, and x_c (Table I) which turn out to be independent of concentration, and their dependence on molar volume was determined by the corresponding ''Grúneisen parameters'': $\gamma_I=\delta\ \ln I/\delta\ \ln V=22$, $\gamma_U=\delta\ln U_0/\delta\ln V=-5$, $\gamma_W=\delta\ln W/\delta\ln V=-7$, $\gamma_x=\delta\ln x_c/\delta\ \ln V=20$. The values of coefficients in formulas (1) and (3), as well as the parameters t and D_0 were assumed to be universal. The curves in Fig. 8 were plotted by using the obtained values of parameters. The possibility of a self-consistent description of all the experimental dependences is the most convincing evidence of the correctness of theoretical ideas concerning quantum diffusion of ^{3}He impurities in the hcp phase of ^{4}He as well as of calculations for various mechanisms of scattering of impurity quasiparticles.

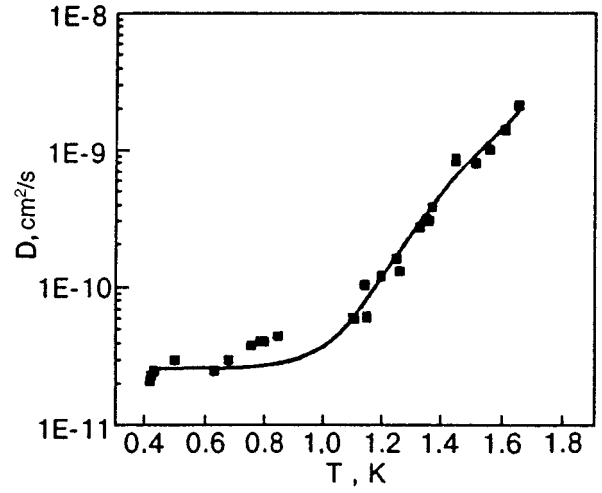

FIG. 7. Temperature dependence of the diffusion coefficient of a solution with 4%^{3}He for $V=20.7$ cm^3/mole.

TABLE I. Parameters of ^{3}He–^{4}He solid solutions.

V, cm^3/mole	$I\cdot10^5$	$U_0\cdot10^2$	Θ, K	W	x_0, %^{3}He
21.0	2.2	2.1_4	26	14	9.3
20.7	1.6	2.3	27	15.4	7.0
20.5	1.3	2.4	28	17.9	5.8
20.2	1.0	2.5_6	29	18.9	4.3
19.9	0.72	2.7	30	20.3	3.2

It is interesting to trace the temperature variation of contribution of various diffusion mechanisms described by corresponding terms in formula (10). This is illustrated in Fig. 9 presenting the data for a 0.25% solution at $V=20.7$ cm^3/mole. It can be seen that in this case each of the three

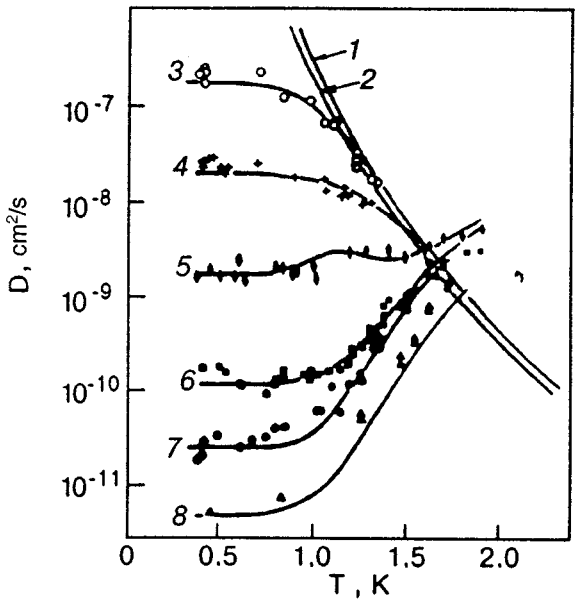

FIG. 8. Temperature dependences of the diffusion coefficient of a solution with various ^{3}He concentrations: theoretical dependences (2) for $V=21.0$ (curve *1*) and 20.5 cm^3/mole (curve *2*), for 21.0 cm^3/mole, $x=0.006$ (curve *3*) and 0.05%^{3}He (curve *4*); for 20.7 cm^3/mole, $x=0.25$ (curve *5*); for 2.17 (curve *6*) and 4.0%^{3}He (curve *7*); for 20.5 cm^3/mole and 4.98%^{3}He (curve *8*). The curves are calculated by formula (10) taking into account the vacancion contribution with the parameters given in Table I.

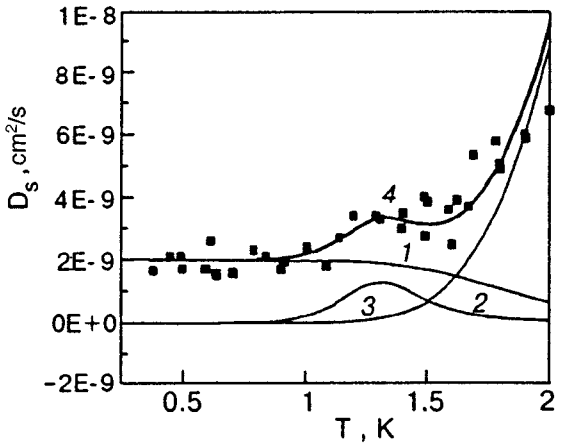

FIG. 9. Temperature dependence of various contributions to diffusion in a solution with 0.25% ^{3}He: contribution from the first term in (10) (curve *1*), from the second term (curve *2*), the vacancion contribution (curve *3*), and the total diffusion coefficient (curve *4*); the symbols ■ correspond to experimental data.

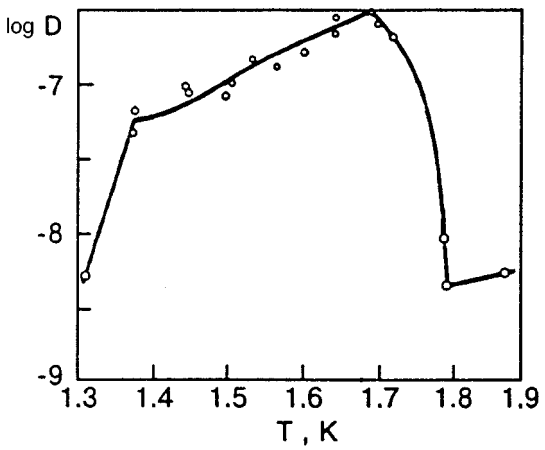

FIG. 10. Temperature dependence of the diffusion coefficient of ^{3}He in a solution with 0.75% ^{3}He on the melting curve in the region of bcc–hcp transitions.

terms becomes significant alternately, while the total values are in good agreement with experimental data.

It is appropriate to mention here the attempt made by Kisvarsanyi and Sullivan[32] to find an alternative explanation of the decrease in the diffusion coefficient with increasing temperature in dilute solutions. Using the fact that the law $D \sim T^{-9}$ can be approximated by an exponential dependence in a bounded interval, these authors attributed the effect observed in Refs. 26 and 27 to the contribution of impuriton scattering by localized vacancies. However, the vacancy concentration required for a quantitative explanation turned out to be so high that their contribution to heat capacity would exceed the total experimental value of heat capacity of helium. The interpretation proposed in Ref. 32 contains some other contradictions (see Ref. 42), but the most important fact is that the vacancy contribution obviously cannot explain the entire body of experimental data.

5. DIFFUSION OF ^{3}He IN THE bcc PHASE OF ^{3}He–^{4}He SOLUTIONS

The phase diagram of solid helium has the following peculiarity: the bcc phase of ^{4}He and dilute ^{3}He–^{4}He solutions exists only at $T>1$ K. This rules out a quantitative analysis of regularities in quantum diffusion of impurities in the bcc phase, and experimental studies of regularities in quantum diffusion of impurities in the bcc phase were carried out in a wide temperature range only for concentrated solutions with $x \geqslant 4\%$. These experiments have not revealed the regularities typical of quantum diffusion, but demonstrated a number of new interesting features of diffusion processes. Above all, it was found that the diffusion coefficients (DC) in coexisting bcc and hcp phases in the vacancion region differ significantly.[33] This is illustrated in Fig. 10 showing the temperature dependence of DC in a 0.75% ^{3}He solution along the melting curve in the region of the bcc–hcp transition. The obtained results indicate that the DC in the bcc phase is almost two orders of magnitude larger than in the hcp phase. This fact was confirmed in later experiments on diffusion[28] as well as during measurements of ion mobility.[34] The reason behind such a discrepancy remains unclear.

Figure 11 shows the temperature dependence of the DC for 3.9 and 6.3% ^{3}He solutions with $V=21.15$ cm^3/mole. The shape of the curve is the same as in the hcp phase, but the dependence has the following two peculiarities: (a) the presence of a large transition region between the exponential dependence at high temperatures and a plateau at low temperatures, and (b) the absence of a concentration dependence for DC. Subsequent measurements[35] revealed that the DC is independent of concentration up to $x \cong 15\%$.

The most unexpected circumstance is connected with a noticeable diffusion flow observed in solutions with ^{3}He concentrations exceeding the critical values typical of the hcp phase. According to estimates obtained by Kalnoj and Strzhemechny,[36,37] the interaction potential for impurities in the bcc phase is more than an order of magnitude larger than in the hcp phase. This must lead to a corresponding decrease in x_c. It is appropriate to note here that other proofs of a

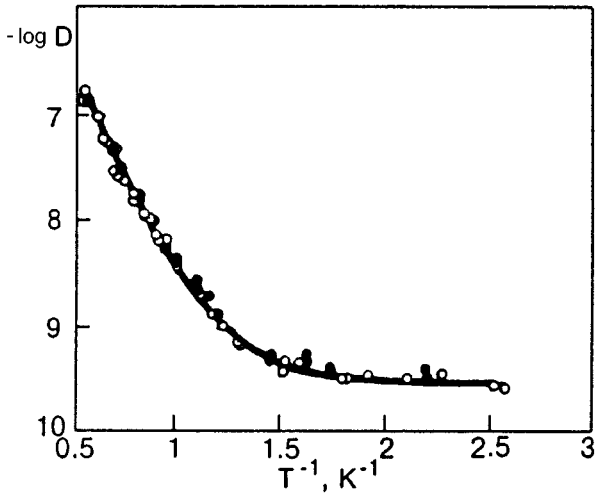

FIG. 11. Temperature dependence of the diffusion coefficient in the bcc phase of solutions with % ^{3}He: 3.9 (○) and 6.3 (●) for 21.15 cm^3/mole.

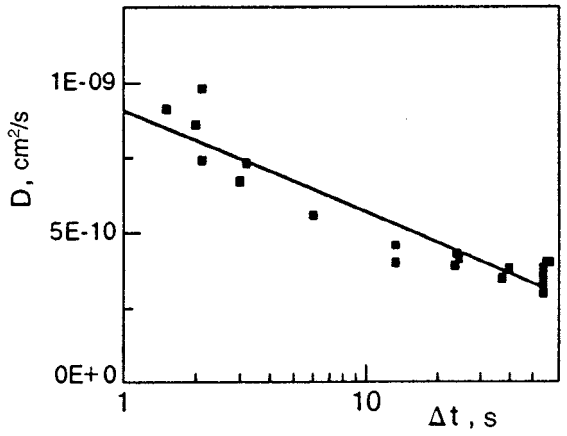

FIG. 12. Dependence of the diffusion coefficient in a solution with 19.7% ^{3}He for 21.15cm^3/mole on the interval between pulses, measured by the spin echo method.[35]

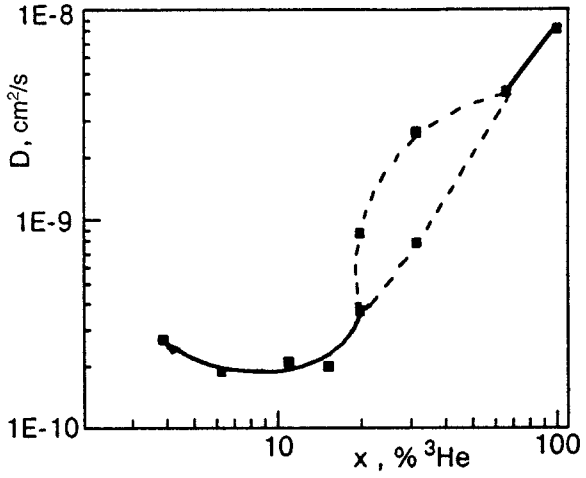

FIG. 13. Concentration dependence of the diffusion coefficient in the plateau region in the bcc phase of solutions ^{3}He–^{4}He.[35]

significant diffusion flow in solid helium at low temperatures for $x > x_c$ also exist. First of all, this concerns the data on the kinetics of decomposition of solid solutions at $T < 0.4$ K (see, for example, Ref. 39). The results of experiments show that the decomposition of solutions occurs at any concentration. The measurements of heat capacity of concentrated solutions[40,41] show that such solutions can be obtained in a nearly equilibrium state. Some results of NMR measurements, e.g., the dependence of diffusion on density,[29] also speak in favor of this hypothesis. In such experiments, an increase in density, and accordingly in the fraction of stationary clusters, must lead to a decrease in the spin echo signal amplitude in view of the absence of relaxation in the case when particles are completely stationary. However, this effect was not observed in these and other experiments.

It was found that the value of diffusion coefficient for a 19.5% ^{3}He solution is higher, and one more peculiarity, i.e., the dependence of the diffusion coefficient on the time interval Δt between probing pulses in the spin echo method, was discovered.[35] The results of a more detailed analysis of this effect presented in Fig. 12 indicate a decrease in the DC with increasing Δt almost by a factor of 3. The same figure shows for comparison the results of similar measurements for a 0.5% solution in the hcp phase, in which this effect does not exist. A $D(\Delta t)$ dependence was also observed in a 31.4% solution. The data on the concentration dependence of D in the plateau region for $V = 21.15$ cm^3/mole for concentrated solutions are presented in Fig. 13. Dashed curves mark the region in which a $D(\Delta t)$ dependence was observed.

In contrast to situation in the hcp phase, where the developed theory provides a quantitative explanation for almost all peculiarities observed in the diffusion of impurities, the results of measurements in the bcc phase can be explained only qualitatively at the best. In Ref. 43, the following empirical formula taking into account the contributions of vacancion transport (D_v) and flip–flop spin exchange processes ^{3}He–^{3}He (D_{33}) was proposed for explaining peculiarities of the temperature dependence of DC in solutions with a concentration $3.9\% \leqslant x \leqslant 15\%$:

$$D = (D_{34} + D_{xv})\left[1 + \frac{D_{33}}{D_{xv} + D_{34} + D_{33}}\right]. \tag{15}$$

This formula also takes into account the fact that the ^{3}He–^{3}He exchange in dilute solutions, for which the probability that ^{3}He atoms are nearest is low, can make a contribution to the attenuation of spin echo signal only in combination with other processes, e.g., the exchange with vacancies or with ^{4}He atoms (D_{34}) ensuring spin transport over distances of the order of diffusion length $l_D \sim \sqrt{D\Delta t}$.

The latter circumstance makes it possible to explain the increase in the diffusion coefficient as well as the emergence of the $D(\Delta t)$ dependence for $x \cong 20\%$. This value of concentration corresponds to the percolation limit for the formation of infinite continuous chains of impurity atoms in bcc structures (see, for example, Ref. 44). When the number of such chains becomes large, DC approaches the value typical of pure ^{3}He. Near the percolation threshold, the dependence of D on Δt appears quite natural and reflects a decrease in the number of chains of length $l \geqslant l_D$ upon an increase in l_D. In this region, diffusion phenomena in helium get mixed with percolation effects, and we can expect that subsequent investigations will be used for a quantitative verification of various versions of the percolation theory.

It should also be noted that the observed $D(\Delta t)$ dependence is similar to that established earlier for samples with a constrained geometry (see, for example, Ref. 45). This analogy, however, is rather formal since in the case of constrained geometry we are actually speaking not of a real change in the diffusion coefficient, but only of the fact that the conventional formulas connecting the attenuation of spin echo signal with DC become inapplicable. On the contrary, in the case under investigation we have a real change in spin transport. Moreover, if the pattern in question is correct, the DC might increase with decreasing sample length d under the conditions when the value of $d \leqslant l_D$ (in the direction of grad **H**). Naturally, the value of DC must be determined by using adequate formulas taking into account the boundedness of the geometry in the case when the spin echo method is employed for measurements.

Thus, the measurements of diffusion coefficient for ^{3}He in the bcc phase of ^{3}He–^{4}He solid solutions led to the discovery of the following additional peculiarities:

(1) considerable difference in DC in coexisting bcc and hcp phases;
(2) the presence of a noticeable spin transport at low temperatures for concentrations $x \geqslant 4\%$ ^{3}He and its independence of concentration up to 15% ^{3}He;
(3) the presence of a noticeable transition region in the temperature dependence of D between the exponential dependence at high temperatures and a plateau at low temperatures;
(4) an abrupt increase in DC in the plateau region for $x \simeq 20\%$ ^{3}He and the dependence of D on diffusion length in this region. The latter circumstance is associated with the percolation behavior of diffusion coefficient.

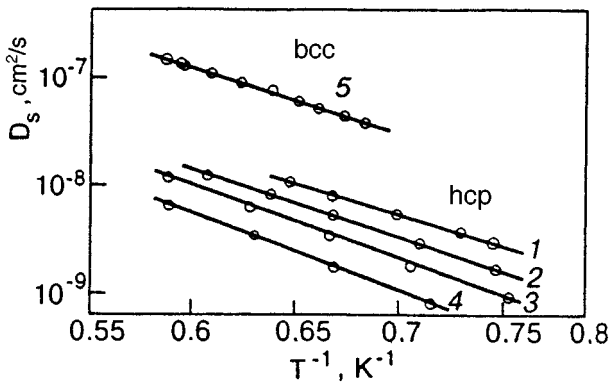

FIG. 14. Temperature dependence of the self-diffusion coefficient in ^{4}He with various molar volumes V, cm^3/mole: 20.82 (curve *1*), 20.67 (curve *2*), 20.49 (curve *3*), 20.30 (curve *4*), and 20.92–21.08 (bcc phase; along the melting curve) (curve *5*).[49,50]

6. DIFFUSION OF VACANCIES IN SOLID HELIUM

The experiments described above proved that ^{3}He impurities in the hcp phase ^{4}He of solutions with $x < 1\%$ behave as quasiparticles with a narrow band $\Delta \leqslant 10^{-4}$ K. On the other hand, according to most theoretical calculations, the band with Δ of vacancies in solid helium must be greater than 1 K. This circumstance stimulates investigations in the kinetics of vacancies since it becomes possible in this case to study a new aspect of quantum diffusion related to wide-band quasiparticles, i.e., quasiparticles for which $\Delta_v > T$.

Although vacancion diffusion was observed by Reich long back in 1963,[46] and a large number of publications appeared later on various vacancion effects in heat capacity, compressibility, velocity of sound, etc. (see, for example, Ref. 47), no convincing evidence of quantum behavior of vacancies has been obtained. One of possible reasons behind such a situation was the lack of direct methods of studying the motion of vacancies. As a rule, information on their behavior was extracted against the background of other more significant contributions.

A method of direct investigation of motion of vacancies in helium crystals was proposed in Ref. 48. The essence of this method lies in the measurement of mobility of a fine-pore membrane frozen into solid helium under the action of small loads which do not exceed the yield stress. The porous membrane was in the form of a movable plate of a parallel-plate capacitor the change in whose capacitance determined the amount of substance flowing though the pores.

In these experiments, the dependence of the velocity of the membrane on the applied load was calculated, and the limiting values for which this dependence was linear (which is typical of vacacion creep) were determined. The measurements were made in the hcp phase of ^{4}He in the interval 20.3–20.9 cm^3/mole at 1.3–1.7 K and in the bcc phase along the melting curve.[49,50] The results of measurements are presented in Fig. 14. All the temperature dependences obtained are exponential.

The further processing of the available data was associated to a considerable extent with the determination of vacancy concentrations. At the first stage, the results obtained by Simmons *et al.*[51] from the x-ray analysis of the temperature dependence of the lattice parameters of helium crystals at constant volume were used. The obtained values of vacancy concentrations and the relation

$$D_s = D_v x \tag{16}$$

were used for obtaining temperature dependences of D_v which turned out to be exponential with the activation energy 4–5 K. This fact could be regarded as an evidence of classical diffusion of vacancies. However, such a treatment involved a number of considerable difficulties.

First of all, the obtained value of the pre-exponential factor could be explained only by introducing a large entropy factor which turned out to be a function of density. Moreover, a comparison of the tunneling probability and the activation motion of vacancies for the obtained values of barriers proved that tunneling must dominate even at $T \leqslant 4.5$ K. Besides, according to estimates, the contribution of vacancies to heat capacity near the melting point for the energy of their formation obtained from x-ray measurements must be larger than the total heat capacity of solid helium. The attempts to overcome these difficulties by treating vacancies as narrow-band quasiparticles or presuming a noticeable contribution from bivacancies also led to various contradictions.

For this reason, subsequent analysis[52] was carried out under the assumption that vacancies are wide-band quasiparticles, which is in accord with a larger part of theoretical estimates and with the approach used by Bernier and Hetherengton[47] for ^{3}He. The concentration of vacancies was estimated on the basis of the data on heat capacity and thermal expansion leading to identical results to within the experimental error and matching to the results obtained from NMR measurements taking into account the contribution from phonon-stimulated diffusion[29] (see Sec. 4). In their analysis, Bernier and Hetherengton[47] assumed a peculiar dependence of the vacancy density of states in ^{3}He:

$$\rho(E) = A(E - \Phi)^2. \tag{17}$$

Heat capacity data processing showed that $A = \Phi^{-3}$. Thus,

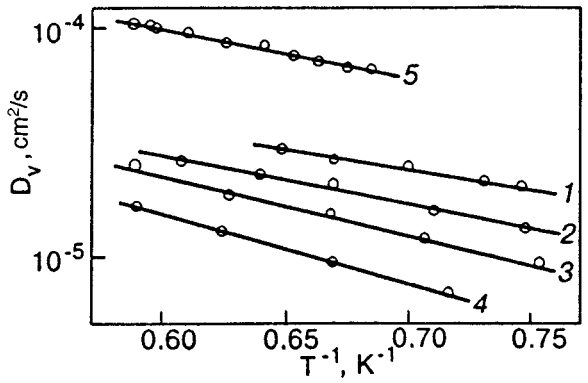

FIG. 15. Temperature dependence of the diffusion coefficient of vacancies in ^{4}He for various molar volumes (notation is the same as in Fig. 14).[50]

TABLE II. Values of parameters in ^{3}He and ^{4}He crystals obtained from comparison of theoretical and experimental results.

Phase	V, cm^3/mole	Φ, K	Δ_v, K	E_a, K
hcp^{4}He	20.30	6.3_5	4.0	7.5
	20.49	6.1_8	3.2	6.3
	20.67	6.0	2.5	5.2
	20.82	5.8_8	2.0	4.3
bcc^{4}He	20.92–21.08	5.1	6.9	5.7
	19.32	12.9	8.2	—
bcc^{3}He	20.12	11.2	6.1	—
	21.10	8.8	4.9	—
	22.48	6.6	3.5	—
	24.30	4.1	3.2	—
	24.40	3.9	3.4	—
	24.50	3.8	3.6	—

the theory contained only one arbitrary parameter, which elevated the reliability of data processing. Under these conditions, it was found that

$$x_v=(T/\Phi)^3\ \exp(-\Phi/T), \tag{18}$$

$$C_v=Rx/T+6T/\Phi+12(T/\Phi)^2. \tag{19}$$

However, dependence (17) has not received a theoretical substantiation, and hence a more traditional situation with an ordinary quadratic energy–momentum relation was analyzed. In this case

$$x_v=4(T/\pi\Delta_v)^{3/2}\ \exp(-\Phi/T), \tag{20}$$

$$C_v=Rx_v[(\Phi/T)^2+3\Phi/T+15/4]. \tag{21}$$

Here we have two fitting parameters determined with different reliability levels. Nevertheless, the ratio D_s/C_v in both approaches does not depend on concentration at all, which makes it possible to determine reliably the DC for vacancies just empirically. The temperature dependence of D_v obtained from such a processing is presented in Fig. 15. The obtained $D_v(T)$ dependence can be interpreted as associated with phonon-stimulated tunneling of vacancies considered long ago by Flynn and Stonecham.[53] These processes lead to the following dependence:

$$D=a^2\Delta^2(\pi/\hbar^2E_aT)\exp(E_a/T), \tag{22}$$

where E_a is the activation energy corresponding to the difference in the lattice energy in the equilibrium state and in an activated state (i.e., the state in which the energy levels at neighboring sites become equal due to the interaction with phonons). The values of E_a and Δ_v obtained from a comparison with experiment are given in Table II. It was found that $\Delta_v>T$, which confirms the hypothesis on vacancions as wide-band quasiparticles.

The approach used above has made it possible to obtain a self-consistent pattern. However, in this case also we encounter a difficulty associated with a significant contribution of band motion of vacancies for the obtained band width if this motion is confined only to two-phonon scattering processes as in the case if impuriton diffusion. In this case, the dependence $D\sim T^{-7}$ must be observed.[36] The absence of such a contribution can be associated, for example, with a significant role of one-phonon processes in the case of a wide band as it was done in the publication by Stamp and Zhang[54] devoted to the interpretation of data on quantum diffusion of μ-mesons in insulators. The estimates obtained on the basis of Stamp–Zhang formulas show that the dependence $D\sim T^{-3}$ typical of such a process must be observed only at $T<1$ K.

Similar experiments and data processing were carried out for a large number of samples of bcc phase of ^{3}He (Fig. 16).[55,56] In this case, it was found that D_v does not depend on temperature, which can be regarded as an evidence of tunnel motion of vacancies. This conclusion is also confirmed by the results of similar processing of the data on spin diffusion obtained by Reich.[46] A comparison of these data with the results of calculations made by Landesman[57] made it possible to obtain much more reliable values of the vacancion band width which also turned out to be larger than the temperature. The obtained results are presented in Table II.

An interesting problem concerning ^{3}He is that of vacancion polarons considered for the first time by Andreev.[58] In this case, the band motion of vacancies is impossible in view of disordering of spins, and the formation of a region with the ferromagnetic spin ordering around a vacancy turns out to be advantageous from the energy point of view in the case of a large band width and low temperature. The presence of

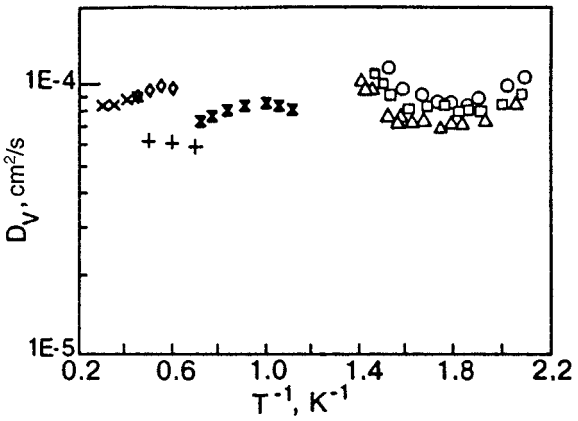

FIG. 16. Temperature dependence of the diffusion coefficient of vacancies in the bcc phase of ^{3}He for various molar volumes V, cm^3/mole: 19.32 (+), 20.12 (∇), 21.10 ($\times$), 22.48 ($\diamond$), 24.30 ($\triangle$), 24.40 ($\square$), and 24.50 ($\bigcirc$).[46,56]

such vacancion polarons affects the static and kinetic properties of solid helium. For example, their effect on spin diffusion was considered by Biushvili and Tugushi[59] and on the mobility of vacancies by Iordanskii.[60] In several publications,[61,62] attempts were made to observe vacancion polarons (see also Ref. 63), but no reliable evidence of such an observation have been obtained.

Concluding the section, we can state that the available data on vacancy diffusion can be described self-consistently in the model of wide-band vacancions whose motion is restricted by disordered spins in ^{3}He and one-phonon scattering processes in ^{4}He. However, the latter statement requires a further experimental and theoretical verification. The most important experiments will be those in ^{4}He at $T<1$ K intended for the observation of band motion of vacancies and in ^{3}He at $T<0.3$ K, in which we can expect a manifestation of the effect of vacancion polarons predicted by Andreev.

CONCLUSIONS

The main result of investigations described here is undoubtedly the discovery and analysis of quantum diffusion in solid helium, which was predicted by Andreev and Lifshits.[1] It was proved that the ^{3}He impurities in the hcp phase of ^{4}He behave like narrow-band quasiparticles whose diffusion flow can be described quantitatively by using the theoretical concepts developed in Ref. 1. Experimental evidence of the fact that vacancies in solid helium at $T\leqslant 1$ K are wide-band quasiparticles were also obtained. However, no clearly manifested effects typical of such particles were discovered. A number of new peculiarities were observed in the bcc phase of concentrated solutions such as a noticeable diffusive transport for $x>x_c$, the absence of a concentration dependence of DC in the interval $4\%\leqslant x\leqslant 15\%$, a sharp increase in the diffusion coefficient in the vicinity of $x\cong 20\%$, and a $D(\Delta t)$ dependence under these conditions). These peculiarities have not yet received a quantitative theoretical explanation. The established relation between diffusion and percolation phenomena in concentrated solution seems to be promising and can open additional opportunities for a more profound analysis of both phenomena.

A few additional remarks will be appropriate here. Naturally, quantum diffusion must be manifested in various properties of quantum crystal. In this connection, we can mention a number of publications in which such effects were considered. Meierovich[64] proved that the formation of delocalized defects must lead to a nonmonotonic temperature dependence of internal friction in the crystal. The attempts to observe this effect[65,66] were not a success probably because it is difficult to ensure a high quality of crystals in such experiments. The effect of quantum-mechanical motion of particles on the shape of the Mössbauer lines as well as the NMR and EPR lines was also studied.[67,68]

The quantum nature of motion of impurities must be manifested directly in the kinetics of decomposition of ^{3}He–^{4}He solid solutions occurring at $T\leqslant 0.4$ K. Under the conditions of decomposition of solutions, diffusion is realized for a finite concentration gradient. The mobility of quasiparticles in this case is mainly limited by U-processes, and relevant experiments make it possible in principle to obtain information on such processes. However, numerous publications in this field failed to provide information on U-processes; moreover, the results of these investigations did not correlate with the available data on diffusion. The obtained values of decomposition time were often too large, depended on the past history of the samples under investigation, and could not be interpreted unambiguously. The only exception was the publication by Shwarts *et al.*,[39] in which reproducible times of decomposition were obtained for the first time, and their decrease with temperature was observed. The authors of Ref. 39 reasonably associated this effect with the dependence $D\sim x^{-1}$ and proved the possibility of quantitative agreement between the obtained results and the diffusion coefficient determined in NMR experiments. Naturally, the diffusion coefficient measured in the presence of a finite concentration gradient can differ from the self-diffusion coefficient. In order to carry out a correct quantitative comparison, the corresponding diffusion problem must be solved under the phase-separation conditions, which has not been done yet. Further experimental and theoretical investigations of this problem are very important.

Another promising trend is associated with the study of extremely dilute solution of ^{3}He in ^{4}He and with the possibility to observe the peculiar effect predicted by Widom[69] under the conditions when the impuriton mean free path λ_i becomes larger then the diffusion length of attenuation of spin-echo signal. Widom[69] proved that in this case an unusual dependence for spin echo amplitude is observed:

$$h=h_0 I_0(2a\gamma r^2 G\Delta/3\hbar), \tag{23}$$

where I_0 is the Bessel function of the first kind. The presence of the band width Δ in the argument of I_0 in formula (23) is worth noting. This circumstance provides a unique opportunity for its direct measurement.

In this connection, it is expedient to estimate the concentration of the solution in which dependence (23) can be observed. The condition of applicability of relation (23) was formulated by Widom,[69] but it actually determines only the possibility of observation of this dependence in NMR experiments and does not contain concentration. Using the condition $\lambda_i>\lambda_D$, expressing λ_i through D, and determining λ_D from the condition $\gamma DG^2\tau^3\sim 1$ [see (13)], we can obtain by using (5)

$$x<(\gamma aGh/U_0)^{1/2}(\Delta/U_0)^{1/6}. \tag{24}$$

Taking into account a weak dependence on Δ in (24), we can write the following expression for the boundary concentration $x_\lambda\cong 3\cdot 10^{-8}(G/U_0)^{1/2}$. Thus, for actual values of G this effect must be observed for $x\sim 10^{-6}$, which is attainable for modern NMR technique (at least in the case when the melting–crystallization method is used for the restoration of sample magnetization).

*E-mail: grigorev@ilt.kharkov.ua

[1)]In Ref. 30, the value of $D_{\rm dip}$ is underestimated. The refined value $D_{\rm dip}=2.6\cdot 10^{-12}$ for $x=4\%$ is given in Ref. 38. The author is grateful to M. A. Strzhemechny who pointed out this circumstance.

[1] A. F. Andreev and I. M. Lifshits, Zh. Éksp. Teor. Fiz. **56**, 2056 (1969) [*sic*].
[2] J. de Boer, Physica **14**, 139 (1948).
[3] Ju. Kagan, J. Low Temp. Phys. **87**, 525 (1992).
[4] V. N. Grigorev, B. N. Eselson, and V. A. Mikheev, Fiz. Nizk. Temp. **1**, 5 (1975) [Sov. J. Low Temp. Phys. **1**, 1 (1975)].
[5] D. I. Pushkarov, Zh. Éksp. Teor. Fiz. **59**, 1755 (1970) [Sov. Phys. JETP **32**, 954 (1970)].
[6] V. A. Slusarev, M. A. Strzhemechny, and I. A. Burakhovich, Fiz. Nizk. Temp. **4**, 698 (1978) [Sov. J. Low Temp. Phys. **4**, 333 (1978)].
[7] Ju. Kagan and M. I. Klinger, J. Phys. **C7**, 2791 (1974).
[8] V. A. Slusarev and M. A. Strzhemechny, in *Low Temperature Physics* [in Russian], vol. 19, Kharkov (1972).
[9] A. Widom and M. G. Richards, Phys. Rev. **6**, 1196 (1972).
[10] A. Landesman and J. M. Winter, in *Proc. LT-13*, London (1974).
[11] Ju. M. Kagan, Fiz. Nizk. Temp. **1**, 573 (1975) [Sov. J. Low Temp. Phys. **1**, 278 (1975)].
[12] V. A. Slusarev and M. A. Strzhemechny, Fiz. Nizk. Temp. **3**, 1229 (1977) [Sov. J. Low Temp. Phys. **3**, 591 (1977)].
[13] R. A. Guyer, R. C. Richardson, and L. J. Zane, Rev. Mod. Phys. **43**, 532 (1971).
[14] A. F. Andreev, Usp. Fiz. Nauk **118**, 251 (1976) [Sov. Phys. Uspekhi **19**, 137 (1976)].
[15] Ju. M. Kagan and L. A. Maksimov, Zh. Éksp. Teor. Fiz. **84**, 792 (1983) [Sov. Phys. JETP **57**, 459 (1983)].
[16] K. L. Garvin and H. A. Reich, Phys. Rev. **115**, 1478 (1959).
[17] V. A. Mikheev, B. N. Eselson, V. N. Grigorev, and N. P. Mikhin, Fiz. Nizk. Temp. **3**, 385 (1977) [Sov. J. Low Temp. Phys. **3**, 186 (1977)].
[18] E. L. Hahn, Phys. Rev. **80**, 580 (1950).
[19] J. E. Tanner, J. Chem. Phys. **56**, 3850 (1972).
[20] M. G. Richards, J. Pope, and A. Widom, Phys. Rev. Lett. **29**, 708 (1972).
[21] V. N. Grigorev, B. N. Eselson, V. A. Mikheev, and Yu. E. Shulman, Pis'ma Zh. Éksp. Teor. Fiz. **17**, 25 (1973) [JETP Lett. **17**, 16 (1973)].
[22] V. N. Grigorev, B. N. Eselson, and V. A. Mikheev, Pis'ma Zh. Éksp. Teor. Fiz. **18**, 289 (1973) [JETP Lett. **18**, 169 (1973)].
[23] V. A. Mikheev, N. P. Mikhin, and V. A. Maidanov, Fiz. Nizk. Temp. **9**, 901 (1983) [Sov. J. Low Temp. Phys. **9**, 465 (1983)].
[24] J. Schratter, A. R. Allen, and M. G. Richards, J. Low Temp. Phys. **57**, 179 (1984).
[25] W. A. Huang, H. A. Holdberg, and H. A. Guyer, Phys. Rev. **B11**, 3374 (1975).
[26] B. N. Eselson, V. A. Mikheev, V. N. Grigorev, and N. P. Mikhin, Zh. Éksp. Teor. Fiz. **74**, 2311 (1978) [Sov. Phys. JETP **47**, 1200 (1978)].
[27] A. R. Allen and M. G. Richards, Phys. Lett. **A65**, 36 (1978).
[28] A. R. Allen, M. G. Richards, and J. Schratter, J. Low Temp. Phys. **47**, 289 (1982).
[29] V. A. Mikheev, V. A. Maidanov, and N. P. Mikhin, Fiz. Nizk. Temp. **8**, 1000 (1982) [Sov. J. Low Temp. Phys. **8**, 505 (1982)].
[30] V. A. Slusarev and M. A. Strzhemechny, Fiz. Nizk. Temp. **9**, 1235 (1983) [Sov. J. Low Temp. Phys. **9**, 635 (1983)].
[31] S. Kirkpatrick, Rev. Mod. Phys. **45**, 574 (1973).
[32] E. G. Kisvarsanyi and N. S. Sullivan, Phys. Rev. **B48**, 16577 (1994).
[33] V. N. Grigorev, B. N. Eselson, and V. A. Mikheev, Zh. Éksp. Teor. Fiz. **64**, 608 (1973) [Sov. Phys. JETP **37**, 309 (1973)].
[34] D. Marti and F. I. B. Williams, J. Phys. (Paris) **34**, 989 (1973).
[35] V. N. Grigorev, B. N. Eselson, and V. A. Mikheev, Fiz. Nizk. Temp. **1**, 1318 (1975) [Sov. J. Low Temp. Phys. **1**, 632 (1975)].
[36] S. E. Kalnoj and M. A. Strzhemechny, Fiz. Nizk. Temp. **8**, 1025 (1982) [Sov. J. Low Temp. Phys. **8**, 515 (1982)].
[37] S. E. Kalnoj, Ph.D. thesis, Kharkov (1987).
[38] M. A. Strzhemechny, D.Sc. thesis, Kharkov (1990).
[39] V. A. Shvarts, N. P. Mikhin, È. Ya. Rudavskii *et al.*, Fiz. Nizk. Temp. **21**, 717 (1995) [Low Temp. Phys. **21**, 556 (1995)].
[40] D. O. Edwards, A. S. McMillan, and J. D. Daunt, Phys. Lett. **1**, 218 (1962).
[41] D. O. Edwards, A. S. McMillan, and J. D. Daunt, Phys. Rev. Lett. **9**, 195 (1962).
[42] V. N. Grigor'ev, Phys. Rev. B (in press).
[43] V. N. Grigorev, B. N. Eselson, and V. A. Mikheev, Zh. Éksp. Teor. Fiz. **66**, 321 (1974) [Sov. Phys. JETP **39**, 153 (1974)].
[44] B. I. Shklovskii and A. L. Èfros, Usp. Fiz. Nauk **117**, 401 (1975) [Sov. Phys. Usp. **18**, 845 (1975)].
[45] D. E. Woessner, J. Phys. Chem. **67**, 1365 (1963).
[46] H. A. Reich, Phys. Rev. **129**, 630 (1963).
[47] M. E. R. Bernier and J. H. Hetherengton, Phys. Rev. **B39**, 11286 (1989).
[48] N. E. Dyumin, S. V. Svatko, and V. N. Grigorev, Fiz. Nizk. Temp. **15**, 524 (1989) [Sov. J. Low Temp. Phys. **15**, 295 (1989)].
[49] N. E. Dyumin, N. V. Zuev, and V. N. Grigorev, Fiz. Nizk. Temp. **18**, 952 (1992) [Sov. J. Low Temp. Phys. **18**, 669 (1992)].
[50] N. E. Dyumin, N. V. Zuev, and V. N. Grigorev, Fiz. Nizk. Temp. **19**, 33 (1993) [Low Temp. Phys. **19**, 23 (1993)].
[51] B. A. Fraas, P. R. Granford, and R. V. Simmons, Phys. Rev. **B39**, 124 (1989).
[52] N. E. Dyumin, N. V. Zuev, V. V. Boiko, and V. N. Grigorev, Fiz. Nizk. Temp. **19**, 980 (1993) [Low Temp. Phys. **19**, 696 (1993)].
[53] C. P. Flynn and A. M. Stonecham, Phys. Rev. **B1**, 3966 (1970).
[54] P. C. E. Stamp and C. Zhang, Phys. Rev. Lett. **66**, 1902 (1991).
[55] N. E. Dyumin, V. V. Boiko, N. V. Zuev, and V. N. Grigorev, Fiz. Nizk. Temp. **20**, 274 (1994) [Low Temp. Phys. **20**, 217 (1994)].
[56] N. E. Dyumin, V. V. Boiko, N. V. Zuev, and V. N. Grigorev, Fiz. Nizk. Temp. **21**, 509 (1995) [Low Temp. Phys. **21**, 395 (1995)].
[57] A. Landesman, J. Low Temp. Phys. **17**, 365 (1974).
[58] A. F. Andreev, Pis'ma Zh. Éksp. Teor. Fiz. **24**, 608 (1976) [JETP Lett. **24**, 564 (1976)].
[59] L. L. Biushvili and A. I. Tugushi, Pis'ma Zh. Éksp. Teor. Fiz. **26**, 286 (1977) [*sic*].
[60] S. V. Iordanskii, Pis'ma Zh. Éksp. Teor. Fiz. **26**, 183 (1977) [JETP Lett. **26**, 171 (1977)].
[61] P. Kumar and N. S. Sullivan, Phys. Rev. Lett. **55**, 963 (1985).
[62] N. S. Sullivan and P. Kumar, Phys. Rev. **B35**, 3162 (1987).
[63] M. E. R. Bernier and E. Suaudeau, Phys. Rev. **B38**, 784 (1988).
[64] A. È. Meierovich, Zh. Éksp. Teor. Fiz. **67**, 744 (1975) [Sov. Phys. JETP **40**, 368 (1975)].
[65] È. L. Andronikashvili, I. A. Gachechiladze, and V. A. Melik-Shakhnazarov, Fiz. Nizk. Temp. **1**, 635 (1975) [Sov. J. Low Temp. Phys. **1**, 305 (1975)].
[66] E. L. Andronikashvili, V. A. Melik-Shakhnasarov, and I. A. Naskidashvili, J. Low Temp. Phys. **23**, 1 (1976).
[67] A. F. Lubchenko and A. Ja. Dzyublik, Phys. Status Solidi **B59**, K99 (1976).
[68] A. F. Lyubchenko and A. Ya. Dzyublik, Fiz. Nizk. Temp. **16**, 652 (1974) [*sic*].
[69] A. Widom, Phys. Rev. **B4**, 1697 (1971).

Translated by R. S. Wadhwa

On magnetoresistance of layered conductors

V. G. Peschansky

*B. Verkin Institute for Low Temperature Physics and Engineering, National Academy of Sciences of the Ukraine, 310164 Kharkov, Ukraine**

(Submitted July 16, 1996)
Fiz. Nizk. Temp. **23**, 47–51 (January 1997)

The dependence of the resistance of a layered conductor with an arbitrary quasi-two-dimensional electron energy spectrum on the magnitude and orientation of a magnetic field relative to the layers is analyzed. It is shown that, when the current flows along the normal to the layers, the resistance of the sample depends significantly on the angle θ between the normal and the strong magnetic field vector; for $\theta=\pi/2$, the resistance increases linearly with the magnetic field in a wide range of magnetic fields. © *1997 American Institute of Physics.* [S1063-777X(97)00501-X]

The electron theory of metals constructed by I. M. Lifshits under the most general assumptions concerning the energy–momentum relation of charge carriers[1] stimulated a large complex of experimental studies of thermodynamic and kinetic parameters of metals in a magnetic field. These parameters, calculated under the assumption that the dependence of the conduction electron energy ε on the momentum $\mathbf{p}$ is known *a priori*, were found to be very sensitive to the form of the electron energy spectrum in the case of a strong magnetic field. This made it possible to solve the inverse problem (formulated by Lifshits) of reconstructing of the Fermi surface $\varepsilon(\mathbf{p})=\varepsilon_F$, which is the main parameter of the electron energy spectrum, from experimental data. It was found that the constant-energy surface in the momentum space, which is equal to the Fermi energy ε_F, has a very complex shape and is open for most of metals. Galvanomagnetic phenomena, which are most sensitive to the topology of the Fermi surface, formed the basis of a reliable spectroscopic method of analysis of the topological structure of the electron energy spectrum.[2]

The well-substantiated concepts concerning conduction electrons in metals are quite suitable for describing the electron properties of a wider class of conductors. For example, these concepts can be rightfully used for studying transport phenomena in low-dimensional conductors with the metal-type conductivity. This property is typical of conductors of organic origin of the type of tetrathiafulvalene salts, tetraselenotetracene halides, and many other layered conductors in which the electrical conductivity along the layers is much larger than the electrical conductivity along the normal $\mathbf{n}$ to the layers. The strong anisotropy of electrical conductivity is apparently associated with a strong anisotropy in the velocity $\mathbf{v}$ of charge carriers on the Fermi surface, i.e., their energy

$$\varepsilon(\mathbf{p})=\sum \varepsilon_n(p_x,p_y)\cos(anp_z/h), \tag{1}$$

depends only slightly on the momentum component $p_z=\mathbf{p}\cdot\mathbf{n}$.

We shall assume that the coefficients of the cosines decrease significantly with increasing n so that

$$A_1=\eta A_0 \ll A_0; \quad A_{n+1}\ll A_n; \tag{2}$$

where A_n is the maximum value of the function $\varepsilon_n(p_x,p_y)$ on the Fermi surface, a the separation between the layers, and h the Planck constant.

The specific form of the quasi-two-dimensional electron energy spectrum leads to a number of peculiar effects which are not observed in conventional metals. These effects include the orientational effect, viz., a strong dependence of kinetic parameters on the angle θ between the normal to the layers and the vector of a strong magnetic field.

We shall consider the dependence of the magnetoresistance of layered conductors on the magnitude and orientation of the magnetic field $\mathbf{H}=(0, H\sin\theta, H\cos\theta)$ relative to the layers.

In order to find the relation between the current density

$$j_i=\int ev_i f(\mathbf{p})2d^3p(2\pi h)^{-3}=\sigma_{ij}(H)E_j \tag{3}$$

and the electric field $\mathbf{E}$, we must solve the kinetic equation for the charge carrier distribution function $f(\mathbf{p})$:

$$[e\mathbf{E}+e(\mathbf{v}\times\mathbf{H})/c]\partial f/\partial\mathbf{p}=W_{\rm col}\{f\}. \tag{4}$$

In an infinitely weak electric field, the deviation of the distribution function $f(\mathbf{p})=f_0(\varepsilon)-\psi(\mathbf{p})\partial f_0/\partial\varepsilon$ from the equilibrium Fermi function $f_0(\varepsilon)$ is small, and the kinetic equation (4) can be linearized in a small perturbation of the system of conduction electrons. In this approximation, the collision integral $W_{\rm col}$ is a linear integral operator acting on the sought function ψ. At low temperatures, when conduction electrons are scattered mainly by impurity atoms or by other crystal defects, the collision integral can be regarded as the operator of multiplication of the nonequilibrium correction to the Fermi function $f_0(\varepsilon)$ and the collision frequency $1/\tau$ to a high degree of accuracy, i.e., the solution of the kinetic equation disregarding the electron–phonon scattering is an eigenfunction of the integral collision operator. In this approximation, the kinetic equation assumes the following simple form:

$$\partial\psi/\partial t_H+\psi/\tau=e\mathbf{E}\cdot\mathbf{v}, \tag{5}$$

and its solution

$$\psi = eE_i\psi_i = eE_i\int_{-\infty}^{t} dt' v_i(t')\exp[(t'-t)/\tau] \tag{6}$$

makes it possible to determine easily the electrical conductivity tensor components:

$$\sigma_{ij} = 2e^3H/c(2\pi h)^3 \int \delta(\varepsilon-\varepsilon_F)d\varepsilon \int dp_H \int_0^T dt v_i\psi_j$$
$$= \langle v_i\psi_j\rangle. \tag{7}$$

Here e and t_H are the charge and the time of motion of a conduction electron in a magnetic field with a period $T=2\pi/\Omega$ according to the equations

$$\partial p_x/\partial t = (v_y\cos\theta - v_z\sin\theta)eH/c;$$
$$\partial p_y/\partial t = -(eH/c)v_x\cos\theta;$$
$$\partial p_z/\partial t = (eH/c)v_x\sin\theta; \tag{8}$$

c is the velocity of light, the angle brackets denote integration over the Fermi surface with the weight factor $2e^3H/(c(2\pi h)^3)$, and the subscript H on t will henceforth be omitted.

It can easily be seen that the velocity of conduction electrons for $\eta \ll 1$ depends weakly on the momentum component $p_H = p_z\cos\theta + p_y\sin\theta$ along the magnetic field, and closed electron orbits in the momentum space are almost indistinguishable for different values of p_H. Hence it follows that the expansion of the electrical conductivity tensor components into a power series in the quasi-two-dimensionality parameter η starts at least with quadratic terms if at least one of the indices of σ_{ij} coincides with z.[3]

The fact that only two frequencies of the Shubnikov–de Haas oscillations of magnetoresistance of organic conductors are observed experimentally (for some magnetic field orientations, their beats are observed when the values of maximum and minimum cross sections of the Fermi surface are close; see, for example, Refs. 4–7 and the literature cited therein) indicates that one group of charge carriers dominates in these conductors. At any rate, there are no grounds for assuming that the compensation of electron and hole volumes are admissible in the presence of several cavities on the Fermi surface. While considering galvanomagnetic phenomena in such conductors, it is sufficient to present the Fermi surface in the form of a weakly corrugated cylinder open in the direction of the p_z-axis.

The resistance of a conductor along the layers is of the same order of magnitude as the resistance of a noncompensated metal, i.e., it differs insignificantly from the resistance in zero magnetic field equal to $1/\sigma_0$ for any orientation of the magnetic field. In contrast to metals, the amplitude of the Shubnikov–de Haas oscillations of magnetoresistance in layered conductors is much larger since these oscillations are formed by almost all the charge carriers on the Fermi surface since its cross sections by the plane $p_H = \text{const}$ are indistinguishable for $\eta = 0$.

On the contrary, the resistivity ρ_{zz} of a layered conductor along the ''difficult'' direction for the current, i.e., along the normal to the layers, is very sensitive to the magnetic field orientation. As a rule, the asymptotic expression for ρ_{zz} is equal to $1/\sigma_{zz}$. It is determined by a considerably larger set of electrical conductivity tensor components and increases abruptly only for some magnetic field orientations, for which the term proportional to $\eta^2\sigma_0$ vanishes in the expansion of σ_{zz} in small parameters η and $\gamma = 1/\Omega\tau$. This forms the basis of the orientation effect.

The asymptotic expression for $\sigma_{zz}(\eta,\gamma)$, i.e.,

$$\sigma_{zz} = \left\langle T^{-1}\int_{-\infty}^{0} dt_1\ \exp(t_1/\tau)\int_0^T dt v_z(t)v_z(t+t_1)\right\rangle \tag{9}$$

can be easily analyzed for small η and γ under the most general assumptions concerning the form of the function $\varepsilon_n(p_x,p_y)$ satisfying condition (2). Confining our analysis only to the terms proportional to η^2, we can easily carry out integration with respect to p_H. This gives

$$\sigma_{zz} = \sum_{n=1}^{\infty}\int_0^T dt\int_{-\infty}^{t} dt'\left(\frac{an}{h}\right)^2\varepsilon_n(t)\varepsilon_n(t')$$
$$\times\exp\frac{t-t'}{\tau}\,\frac{e^3H\cos\theta}{ac(2\pi h)^2}\cos\left\{\frac{an}{h}[p_y(t)-p_y(t')]\tan\theta\right\}, \tag{10}$$

where all the integrands depend only on t and t'. For $\gamma \ll 1$, σ_{zz} assumes the form

$$\sigma_{zz} = ae^2\tau m^*\cos\theta/2\pi h^4\sum n^2I_n^2(\theta)$$
$$+\eta^2\sigma_0(\eta^2\varphi_1+\gamma^2\varphi_2), \tag{11}$$

where m^* is the effective cyclotron mass, and

$$I_n(\theta) = T^{-1}\int_0^T dt\varepsilon_n(t)\cos(p_y(t)an\tan\theta/h). \tag{12}$$

The functions φ_i are of the order of unity and should be taken into account for the values of $\theta = \theta_c$ for which $I_1(\theta)$ vanishes. In this case, the asymptotic expression for magnetoresistance depends considerably on the rate of decrease in the functions $\varepsilon_n(\mathbf{p})$ with increasing n. If, for example, I_n are proportional η^n, the resistance along the normal to the layers in strong fields for $\theta = \theta_c$ increases in proportion to H^2 in the range of fields satisfying the condition $\eta \ll \gamma \ll 1$ instead of attaining saturation. The saturation of resistance is observed in stronger magnetic fields for $\gamma \leqslant \eta$.

The integrand in formula (12) for $\tan\theta \gg 1$ is a rapidly oscillating function, and $I_n(\theta)$ can easily be evaluated by using the stationary phase method. If an electron orbit contains only two stationary points at which v_x vanishes, the asymptotic expression for I_n acquires the form

$$I_n(\theta) = \left(\frac{h}{anD_p\tan\theta}\right)^{1/2}2\cos\left(\frac{anD_p\tan\theta}{2h}-\frac{\pi}{4}\right), \tag{13}$$

where D_p is the diameter of the Fermi surface along the axis p_y.

According to formula (13), the zeros of the function $I_1(\theta)$ are repeated with a period

$$\Delta(\tan\theta) = 2\pi h/aD_p. \tag{14}$$

Additional possibilities of studying anisotropy in the diameters of the Fermi surface for quasi-two-dimensional conductors by measuring galvanomagnetic parameters are due to the presence of electrons orbits strongly elongated along the axis p_z and passing through a large number of unit cells in the momentum space. The period of electron motion in such orbits increases as θ approaches $\pi/2$ and can exceed the mean free time τ of charge carriers for any value of magnetic field. Hence it should be borne in mind that the above formulas are valid for not very large values of $\tan\theta$. In this case, the condition $T \ll \tau$ is satisfied only for electrons which rotate in orbits that do not contain points of self-intersection at which the period of motion diverges logarithmically. For $\tan\theta \geqslant 1/\eta$, self-intersecting electron orbits appear. Let us analyze their role for $\theta = \pi/2$, when a strongly elongated orbit is ''torn'' into two open orbits. In this case, the period of electron motion, which has the form

$$T(p_y) = \int_0^{2\pi h/a} dp_z c/eHv_x \tag{15}$$

in the case of motion in the open section of the Fermi surface by the plane $p_H = p_y = \text{const}$, changes jumpwise.

As an electron approaches the boundary cross section $p_y = p_c$ separating the region of open cross sections from a small fraction of closed cross sections of the Fermi surface, the period of motion $T(p_y)$ increases without limit. This is due to the fact that the cross section $p_y = p_c$ contains points of self-intersection at which open orbits converge as $p_y \to p_c$, and an electron stays for a considerable time near points of self-intersection, since its velocity in the plane perpendicular to the magnetic field is negligibly small.

The velocity of motion of a charge in an open orbit for $\eta \ll 1$ depends on t weakly. For example,

$$\begin{aligned}\partial v_x(t)/\partial t &= (\partial^2\varepsilon/\partial p_x^2)\partial p_x/\partial t + (\partial^2\varepsilon/\partial p_x \partial p_z)\partial p_z/\partial t \\ &= (eH/c)(v_x \partial v_z/\partial p_x - v_z \partial v_x/\partial p_x) \sim \eta. \end{aligned} \tag{16}$$

The period of electron motion in orbits separated by large distances from a self-intersecting orbit is inversely proportional to $v_x(0)$ to within the terms proportional to η (the point $t = 0$ lies on the cross section of the Fermi surface by $p_z = 0$). However, as an electron approaches a self-intersecting orbit $p_y = p_c$, the electron velocity along the x-axis decreases, and small correction in the parameter η should be taken into account. At the point $\mathbf{p}_c = (0, p_c, 0)$ of self-intersection of the electron orbit, the electron velocity $v_x(0)$ vanishes. The small value of the velocity v_x on electron orbits with p_y close to p_c corresponds to a weak dependence of ε on p_x, and we can calculate the period of electron motion in this case by using the power expansion of energy in small p_x. Omitting higher harmonics in formula (1), we obtain

$$\varepsilon = \varepsilon_0(0, p_y) + p_x^2/2m_1 + \varepsilon_1(0, p_y)\cos(ap_z/h). \tag{17}$$

Using relation (17), we can easily calculate the period of electron motion in orbits close to self-intersecting trajectories:

$$T(p_y) = \eta^{-1/2}\Omega_0^{-1}\int_0^{\pi} d\alpha(\xi^2 + \sin^2\alpha)^{-1/2}, \tag{18}$$

where Ω_0 is the frequency of electron rotation in a magnetic field parallel to the normal to the layers, and

$$\xi^2 = \frac{\varepsilon - \varepsilon_0(0, p_y) - \varepsilon_1(0, p_y)}{2\varepsilon_1(0, p_y)}. \tag{19}$$

As we approach a self-intersecting orbit, the value of ξ becomes infinitely small, and the integral in formula (18) diverges logarithmically in proportion to $\ln(1/\xi)$.

In contrast to conventional metals for which the period of motion of charge carriers is greater than or comparable to the mean free time only in an exponentially small (of the order of $\exp(-\Omega_0\tau)$) region of cross sections of the Fermi surface near a self-intersecting orbit in a quasi-two-dimensional conductor the condition $T \geqslant \tau$ is satisfied in a considerably wider region of electron orbits in which ξ can be of the order of unity.

For $\eta^{1/2} < \gamma_0 = 1/\Omega_0\tau$, the main contribution to σ_{zz} comes from a small fraction of charge carriers for which $\eta^{1/2} v_F < v_x \ll v_F$. For such electrons, the velocity can be regarded to a high degree of accuracy as a harmonic function of t of the form

$$v_z = -\varepsilon_1(0, p_y)(a/\hbar)\sin\Omega t, \tag{20}$$

where $\Omega = aeHv_x(0)/\hbar c = \Omega_0 v_x/v_F$, and v_F is the characteristic Fermi velocity of the order of h/am_1.

Simple calculations lead to the following result for the electrical conductivity along the normal to the layers:

$$\sigma_{zz} = \eta^2\sigma_0 \int d\xi \gamma_0^2/(y_0^2 + \xi^2) \cong \eta^2\gamma_0\sigma_0. \tag{21}$$

The contribution to σ_{zz} from electrons belonging to closed cross sections of the Fermi surface is considerably smaller than that determined by the formula (21). Thus, a small region of open orbits near a self-intersecting one makes the major contribution to the electrical conductivity of the sample, and the sample resistance along the normal to the layers in the above-mentioned region of magnetic fields increases linearly with magnetic field.

A further increase in magnetic field leads to narrowing of the region of electron orbits in which the period of motion of an electron is longer than its mean free time. As a result, the contribution to σ_{zz} from electrons near a self-intersecting orbit for $\gamma_0 \ll \eta^{1/2}$ is also proportional to γ_0^2 as well as the contribution from electrons for which $T \ll \tau$. In this case, the resistivity ρ_{zz} is proportional to the square of the magnetic field.

Thus, for a layered conductor with a quasi-two-dimensional electron energy spectrum, the region of magnetic fields for which the resistance along the normal to the layers in a single crystal increases linearly with magnetic field is quite large.

This research was partly financed by the International Soros Program supporting education in the field of science of the International Foundation ''Regeneration,'' Grant SPU 042051.

*E-mail: peschansky@ilt.kharkov.ua

[1] I. M. Lifshits, *Selected Works. Electron Theory of Metals. Physics of Polymers and Biopolymers*, Nauka, Moscow (1994).
[2] I. M. Lifshits and V. G. Peschansky, Zh. Éksp. Teor. Fiz. **35**, 1251 (1958); *ibid.* **38**, 188 (1960) [Sov. Phys. JETP **8**, 875 (1958); *ibid.* **11**, 137 (1960)].
[3] V. G. Peschansky, J. A. Roldan Lopez, and T. G. Yao, J. Phys. (Paris) **1**, 1469 (1991).
[4] M. V. Kartsovnik, P. A. Kononovich, V. N. Laukhin *et al.*, Zh. Éksp. Teor. Fiz. **97**, 1305 (1990) [Sov. Phys. JETP **70**, 735 (1990)].
[5] I. F. Shcheglov, P. A. Kononovich, V. M. Kartsovnic *et al.*, Synth. Met. **39**, 357 (1990).
[6] M. Tokumoto, A. G. Swanson, J. S. Brooks *et al.*, J. Phys. Soc. Jpn. **59**, 2324 (1990).
[7] R. Yagi, Y. Iye, T. Osada, and S. Kagoshima, J. Phys. Soc. Jpn. **59**, 3069, (1990).

Translated by R. S. Wadhwa

Surface magnetization of metals

S. S. Nedorezov and E. F. Rofe-Beketova

*Ukrainian Academy of Engineering and Education, 310003 Kharkov, Ukraine**
(Submitted July 10, 1996)
Fiz. Nizk. Temp. **23**, 52–57 (January 1997)

The surface magnetization of metals associated with electrons whose orbits are tangential to the metal boundary is investigated, taking into account anisotropy in the energy spectrum of conduction electrons. It is shown that the angular dependence of magnetization has peculiarities determined by the curvature of the Fermi surface at the reference points. Explicit expressions have been obtained for the surface magnetic susceptibility of electrons in bismuth, and appropriate numerical estimates of its value have been made. © *1997 American Institute of Physics.* [S1063-777X(97)00601-4]

INTRODUCTION

The problem concerning the effect of the metal boundary on the diamagnetic susceptibility of electrons was encountered even in the first publications appearing after the creation of the theory of diamagnetism by Landau.[1] Subsequent investigations are closely related to the Lifshits representation[2] of electrons in metal as a gas of quasiparticles (conduction electrons) with an arbitrary energy– momentum relation. The publications by Lifshits and Kosevich[3–5] on the theory of the de Haas–van Alphen effect and quantum oscillations in metals and metallic layers (see also the review in Ref. 6) determined the theoretical and experimental studies of quantum magnetic and size effects in metals.

Special attention was paid to the effect of metal boundary on the magnetization component of conduction electrons, which does not oscillate with a variation of the magnetic field H (see Ref. 7 for the literature concerning this problem). The semiclassical approximation can be used for calculating quantum oscillations of magnetization for quantum energy levels of conduction electrons in a magnetic field,[4] but the calculation of the smooth magnetization component requires the knowledge of exact values of energy levels. However, the thermodynamic theory of perturbations in the powers of H is inapplicable for an analysis of contribution of special electron orbits to the surface magnetization of metals. The results obtained by Lifshits and Kosevich[8] proved to be very important in this respect (see below).

Taking into account the boundary surface of a metal, we can represent the thermodynamic quantities in the form of the sum of two terms one of which is proportional to volume V and the other to the area S of the boundary surface. For example, for the magnetic moment $\mathbf{M}$ of conduction electrons, we can write

$$\mathbf{M}=V\mathbf{M}^{(v)}+S\mathbf{M}^{(s)}. \tag{1}$$

The first term in (1) is associated with the Landau magnetic levels and is determined by electron orbits away from the metal (Fig. 1a), while the second term is determined by electron orbits near the metal boundary (Figs. 1b and c).

The orbits of electrons hopping over the metal surface (see Fig. 1b) are responsible for magnetic surface levels (MSL). The interest to such levels has increased considerably after the experimental discovery of these levels by Khaikin[9] (see also Ref. 10). The contribution of MSL to the magnetization of metals was studied by many authors.[8,11,12] On the basis of the semiclassical approximation for MSL, the following analytic dependence of $M^{(s)}$ on H was obtained:[11,12]

$$M^{(s)}_{\text{quasi}}\propto H^{-1/3}, \tag{2}$$

which leads to a strong surface magnetization of metals in weak magnetic fields.

Another approach to this problem was indicated by Lifshits and Kosevich.[8] They proved that, while calculating $M^{(s)}$, we must take into account exact values of MSL since the value of $M^{(s)}$ differs considerably from $M^{(s)}_{\text{quasi}}$. According to the estimate obtained by Kosevich,[13] $M^{(s)}/M^{(s)}_{\text{quasi}}<0.1$. According to the statement made in Ref. 8 (which, however, was not proved), the magnetic field dependence of magnetic moment does not contain the singularity described by formula (2). The calculations of $M^{(s)}$ in the semiclassical approximation were continued until this problem was solved by one of the authors of the present publication.[14] An explicit expression obtained for $M^{(s)}$ taking into account the exact values of MSL has made it possible to calculate the contribution of MSL to the magnetization of metals. This contribution indeed proved to be equal to zero.

It was shown in Ref. 15 that the surface magnetization of metals is determined by electrons with the orbits tangential to the metal surface (see Fig. 1c). It has a root dependence[1)] on the magnetic field H:

$$M^{(s)}\propto H^{1/2} \tag{3}$$

and is much larger than the surface corrections of the order of λ_F/L, where λ_F is the Fermi wavelength for electrons and L the sample length.

In the present research, the surface magnetization of metals is studied by taking into account the anisotropy in the energy spectrum of conduction electrons. The angular dependence of magnetization on the magnetic field orientation is analyzed for an ellipsoidal Fermi surface (FS). Explicit expressions for the surface magnetic susceptibility of electrons in bismuth are obtained, and the numerical values of its magnitude are estimated. The predicted singularities in the angular dependence of surface magnetization are determined by

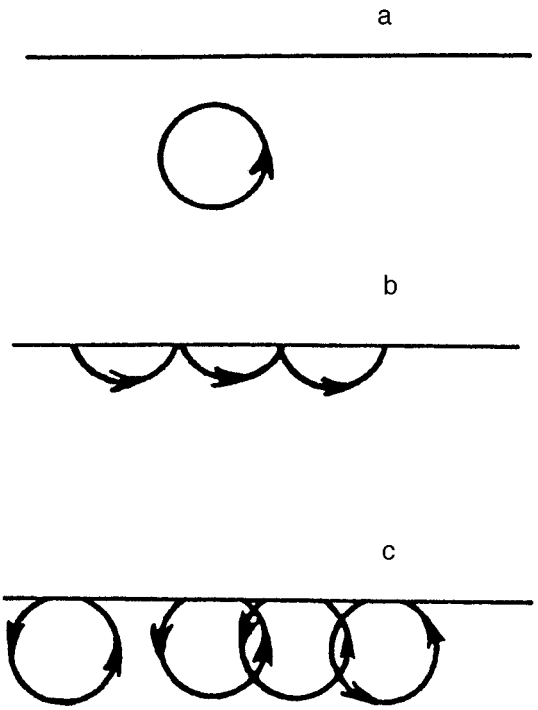

FIG. 1. Types of conduction electron trajectories in a plate placed in a parallel magnetic field.

the curvature of the FS at the reference points and are due to conduction electrons with the orbits tangential to the metal boundary.

SURFACE MAGNETIZATION OF CONDUCTION ELECTRONS

According to Ref. 7, the surface magnetization of electrons with the orbits tangential to the metal boundary (see Fig. 1c) is defined by the formula

$$M^{(s)}=\frac{\beta}{2}\left(\frac{e}{c}\right)^{3/2}\left(\frac{H}{\hbar}\right)^{1/2}\sum_j \int_0^{\zeta} K_j^{3/4}(\varepsilon)k_j^{-1/2}(\varepsilon,\mathbf{N})d\varepsilon. \quad (4)$$

Here $K_j(\varepsilon)$ is the Gaussian curvature of the constant-energy surface $E(\mathbf{p})=\varepsilon$ at the jth reference point with the normal parallel to $\mathbf{H}$ (see Fig. 2), $k_j(\varepsilon,\mathbf{N})$ is the curvature of the normal cross section perpendicular to $\mathbf{N}$ at the jth point, and the vector $\mathbf{N}$ is perpendicular to the boundary surface of the metal. The magnetic field $\mathbf{H}$ is parallel to the metal boundary. The energy ε and the chemical potential ζ are measured from the bottom of the relevant energy band.

For the number β in (4), we have[17]

$$\beta=\frac{3}{4\pi^4}\left\{\frac{2\sqrt{2}-1}{8}\zeta\left(\frac{5}{2}\right)+\frac{\pi^{1/6}}{2^{1/3}}\Gamma\left(\frac{5}{6}\right)\sum_{k=1}^{\infty}(-1)^k\frac{1}{k^{11/6}}\right.$$
$$\times\int_0^{\infty}\left[\cos\left(2\pi ky(x)+\frac{\pi}{12}\right)\right.$$
$$\left.\left.+\cos\left(2\pi ky(-x)+\frac{\pi}{12}\right)-\cos\left(\frac{\pi}{12}\right)\right]dx\right\}$$
$$=0.78\cdot 10^{-2}, \quad (5)$$

where the function $\gamma(x)$ is defined as

$$\gamma(x)=\pi^{-1}\arctan[Ai(x)/Bi(x)], \quad (6)$$

and $Ai(x)$ and $Bi(x)$ are the first and second order Airy functions, respectively. In the case of the quadratic isotropic energy–momentum relation $E=p^2/2m$, relation (4) leads to

$$M^{(s)}2\beta\left(\frac{e}{c}\right)^{3/2}\left(\frac{\zeta}{2m\hbar}\right)^{1/2}H^{1/2}. \quad (7)$$

For an anisotropic energy–momentum relation

$$E(p)=\frac{p_x^2}{2m_1}+\frac{p_y^2}{2m_2}+\frac{p_z^2}{2m_3}, \quad (8)$$

the calculation of the Fermi surface curvature at the reference points gives

$$M^{(s)}=2\beta\left(\frac{e}{c}\right)^{3/2}\left(\frac{\zeta}{2\hbar\sqrt{m_1m_2m_3}}\right)^{1/2}\frac{|\mathbf{h}_m|^{5/2}}{|\mathbf{n}_m\times\mathbf{h}_m|}H^{1/2}, \quad (9)$$

where

$$\mathbf{h}_m=\frac{1}{H}(H_x\sqrt{m_1}\mathbf{i}+H_y\sqrt{m_2}\,\mathbf{j}+H_z\sqrt{m_3}\mathbf{k}),$$
$$\mathbf{n}_m=N_x\sqrt{m_1}\mathbf{i}+N_y\sqrt{m_2}\,\mathbf{j}+N_z\sqrt{m_3}\mathbf{k}. \quad (10)$$

Here the normal $\mathbf{N}$ to the boundary surface of the metal and the magnetic field strength $\mathbf{H}$ are defined in the intrinsic reference frame of expansion (8), and $\mathbf{H}\perp\mathbf{N}$, $|\mathbf{N}|=1$.

It follows from (4) that the surface magnetization is determined by the curvature of constant-energy surfaces at the reference points (see Fig. 2). The largest contribution to magnetization comes from conduction electrons near the reference points with the largest curvature of the Fermi surface, which leads to a considerable anisotropy in surface magnetization. Let us prove this for the case of an ellipsoidal FS elongated along the x-axis ($m_1\gg m_2,m_3$). Denoting $m=m_1$, $\lambda=m_2/m_1$, $\mu=m_3/m_2$ and introducing the unit vectors $\mathbf{h}=\mathbf{H}/H$, $\mathbf{s}=\mathbf{h}\times\mathbf{N}$, we obtain from (9)

$$M^{(s)}=2\beta\left(\frac{e}{c}\right)^{3/2}\left(\frac{\zeta H}{2m\hbar}\right)^{1/2}\Phi(\varphi), \quad (11)$$

where the function

$$\Phi(\varphi)=\frac{1}{\lambda\mu^{1/4}}[h_x^2+\lambda(h_y^2+\mu h_z^2)]^{5/4}(s_z^2+\mu s_y^2+\lambda\mu s_x^2)^{-1/2} \quad (12)$$

defines the magnetization $M^{(s)}$ as a function of the angle φ of rotation of the magnetic field $\mathbf{H}$. In the isotropic case ($\lambda=\mu=1$), we have $\Phi=1$.

An analysis of the angular dependence of $M^{(s)}$ in the case of strong anisotropy ($\lambda\ll 1$) leads to the following results. In the direction $\mathbf{h_0}$ ($\mathbf{h_0}=\mathbf{i}\times\mathbf{N}$) perpendicular to the axis along which the FS is elongated (x-axis), the function $\Phi(\varphi)$ assumes the minimum value

$$\Phi(0)=\left(\frac{\lambda}{\mu}\right)^{1/4}(N_z^2+\mu N_y^2)^{5/4}$$
$$\times[N_x^2(N_z^2+\mu N_y^2)+\lambda\mu(1-N_x^2)^2]^{-1/2}. \quad (13)$$

The value of $\Phi(0)$ is very sensitive to the angle γ at which the X-axis is tilted to the metal boundary

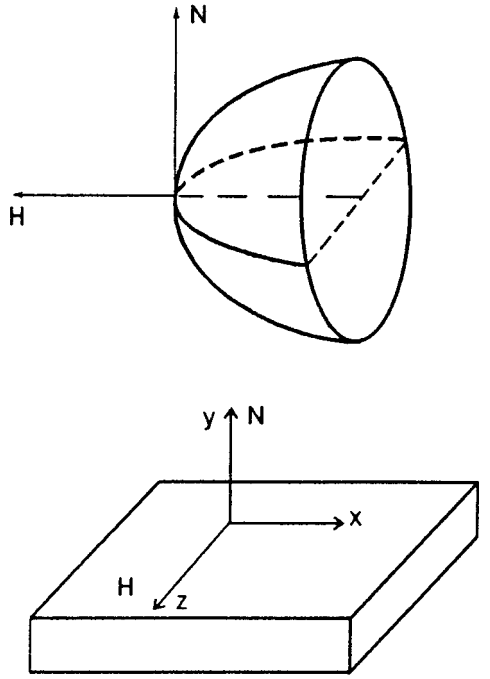

FIG. 2. Reference points of the Fermi surface, determining the surface magnetization of metals.

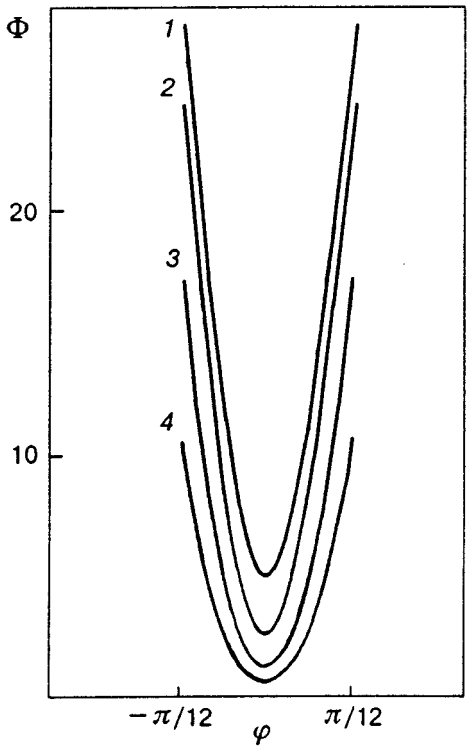

FIG. 3. Angular dependence $\Phi(\varphi)$ of surface magnetization of metals in the case of strong anisotropy of the Fermi surface ($\lambda \approx 4.503 \cdot 10^{-3}$, $\cos\varphi = \mathbf{h}_0 \cdot \mathbf{h}$, $N_z = 0$, $N_x = \sin\gamma$) for various values of $\gamma = 0$ (curve *1*), 6°23′ (curve *2*), 15° (curve *3*), and 30° (curve *4*).

($N_x = \sin\gamma$), varying from values of the order of $\lambda^{1/4}$ to the value $\sim\lambda^{-1/4}$ for $\gamma = 0$. The abrupt increase in the value of $\Phi(0)$ in the angular interval $|\gamma| \sim \sqrt{\lambda}$ is due to a decrease in the curvature $k_j(\varepsilon, \mathbf{N})$ of the normal cross section of the FS at the reference points.

The reference points of the FS with the minimum curvature correspond to the direction $\mathbf{h}_0$. When **H** is tilted relative to $\mathbf{h}_0$, the value of $\Phi(\varphi)$ increases abruptly to $\sim\lambda^{-1}$. For small angles of inclination of the X-axis ($\gamma < 1$), the $\Phi(\varphi)$ dependence is defined by the formula

$$\Phi(\varphi) \approx \frac{|\sin\varphi|^{3/2}}{\lambda\mu^{1/4}}(N_y^2 + \mu N_z^2)^{-1/2}, \tag{14}$$

where φ is the angle between **H** and $\mathbf{h_0}$. The function $\Phi(\varphi)$ assumes the maximum value $\Phi(\pi/2)$ in the direction corresponding to the reference points of the FS with the maximum curvature. Formula (14) is applicable for $\varphi \neq 0$. In a narrow angular interval $|\varphi| \sim \sqrt{\lambda} \ll 1$, the values of $\Phi(\varphi)$ decrease to $\Phi(0)$.

In the vicinity of the direction $\mathbf{h_0}$, ($\varphi < 1$) and for small angles of inclination of the X-axis to the metal boundary ($\gamma < 1$), the angular dependence of the surface magnetization is very sensitive to the value of γ and is defined as

$$\Phi(\varphi) \sim \frac{1}{\lambda\mu^{1/4}}[\sin^2\varphi + \lambda(N_z^2 + \mu N_y^2)]^{5/4} \times [(N_y \sin\varphi + N_z \sin\gamma)^2 + \mu(N_z \sin\varphi - N_y \sin\gamma)^2 + \lambda\mu]^{-1/2}. \tag{15}$$

For $|\sin\varphi| > \lambda$, formula (15) coincides with (14) and described a sharp decrease in the surface magnetization in the angular interval $|\varphi| \sim \sqrt{\lambda}$ (see Fig. 3).

The above analysis shows that anisotropy in the electron energy spectrum is manifested clearly in the angular dependence of surface magnetization. The typical dependence of $M^{(s)}$ on the angle φ along with its field dependence ($\sim\sqrt{H}$) can be useful for an analysis of experimental dependence of magnetization of conduction electrons with orbits tangential to the metal boundary. The FS model considered above can be used directly for calculating the surface magnetization due to conduction electrons in small groups, e.g., electrons in a semimetal.

SURFACE MAGNETIC SUSCEPTIBILITY OF ELECTRONS IN BISMUTH

Proceeding from the known electron energy spectrum of bismuth[18] and calculating $M^{(s)}$ by formula (9), we obtain the following expression for magnetic susceptibility $\chi^{(s)} = \partial M^{(s)}/\partial H$:

$$\chi^{(s)} \beta \left(\frac{e}{c}\right)^{3/2} \left(\frac{\zeta}{2m\hbar}\right)^{1/2} H^{-1/2} F(\varphi), \tag{16}$$

where the function $F(\varphi)$ is defined by the sample shape and depends significantly on the magnetic field orientation. We describe the results of calculations od $\chi^{(s)}$ for a plate with the normal **N** parallel to the trigonal (C_3), bisector (C_1), and binary (C_2) crystallographic axes of bismuth. The angle φ determines the direction of **H** in the plane of the plate.

1. $\mathbf{N}\|C_3$.

$$F(\varphi) = \frac{1}{\lambda\mu^{1/4}}\sum_{k=0}^{2}\frac{[\alpha_c\cos^2(\varphi+2\pi k/3)+\lambda\sin^2(\varphi+2\pi k/3)]^{5/4}}{[\alpha_s\sin^2(\varphi+2\pi k/3)+\mu\cos^2(\varphi+2\pi k/3)]^{1/2}}, \tag{17}$$

where the angle φ is the angle between the C_1-axis and **H**,

$$\alpha_c=\cos^2\alpha+\lambda\mu\sin^2\alpha, \quad \alpha_s=\sin^2\alpha+\lambda\mu\cos^2\alpha, \tag{18}$$

α is the angle of inclination of the major axis of the ellipsoid to the C_1-axis, $\lambda=m_2/m_1, \mu=m_3/m_2$, and $m=m_1$. In the isotropic case ($\lambda=\mu=1$), $F(\varphi)=3$, i.e., to the number of electron ellipsoids in bismuth.

2. $\mathbf{N}\|C_1$.

$$F(\varphi) = \frac{1}{\lambda\mu^{1/4}}\sum_{k=0}^{2}\frac{(\alpha_s\cos^2\varphi+\beta_k(\lambda,\alpha_c)\sin^2\varphi+\delta\gamma_k\sin 2\varphi)^{5/4}}{(\alpha_c\sin^2\varphi+\beta_k(\mu,\alpha_s)\cos^2\varphi+\delta\gamma_k\sin 2\varphi)^{1/2}}, \tag{19}$$

where $\delta=1/4(1-\lambda\mu)\sin 2\alpha$, and

$$\beta_k(x,y)=\begin{cases} x, & k=0 \\ \dfrac{1}{4}(x+3y), & k=1,2; \end{cases}$$

$$\gamma_k=\begin{cases} 0, & k=0 \\ (-1)^k\sqrt{3}, & k=1,2. \end{cases} \tag{20}$$

The angle of rotation of the magnetic field **H** in the plane of the plate is measured from the axis C_3.

3. $\mathbf{N}\|C_2$.

The angular dependence $F(\varphi)$ in this case is defined by formula (19) in which $\beta_k(\lambda,\alpha_c)$ is replaced by $\beta_k(\alpha_c,\lambda)$, $\beta_k(\mu,\alpha_s)$ by $\beta_k(\alpha_s,\mu)$, and γ_k by

$$\gamma_k=\begin{cases} 2, & k=0 \\ -1, & k=1,2. \end{cases} \tag{21}$$

In this case, the last two terms in (19) coincide due to the equivalence of the two ellipsoids relative to the plate boundary.

The results of calculations of $F(\varphi)$ determining the angular dependence of the surface magnetic susceptibility $\chi^{(s)}$ are presented in Fig. 4. In accordance with the experimental data on the Fermi surface in bismuth,[18] we choose the following values of the parameters λ, μ, and α: $\lambda=4.503\cdot10^{-3}$, $\mu=1.7355$, $\alpha=6°23'$.

The contributions from individual ellipsoids (see Fig. 4b) have a clearly manifested anisotropic nature with a dip in the direction of minimum values of the FS curvature at the reference points. In the case when $\mathbf{N}\|C_3$, the summation over ellipsoids leads to smoothing of these singularities (see Fig. 4a). A different angular dependence is observed for $\mathbf{N}\|C_1, C_2$. In such cases, the singularities are not smoothed, and the function $F(\varphi)$ has a sharp dip in the direction of the C_3 axis. This is associated with different roles of ellipsoids in the formation of $F(\varphi)$, especially for $\mathbf{N}\|C_1$, when the contribution of the first ellipsoid can be neglected (see Fig. 4b).

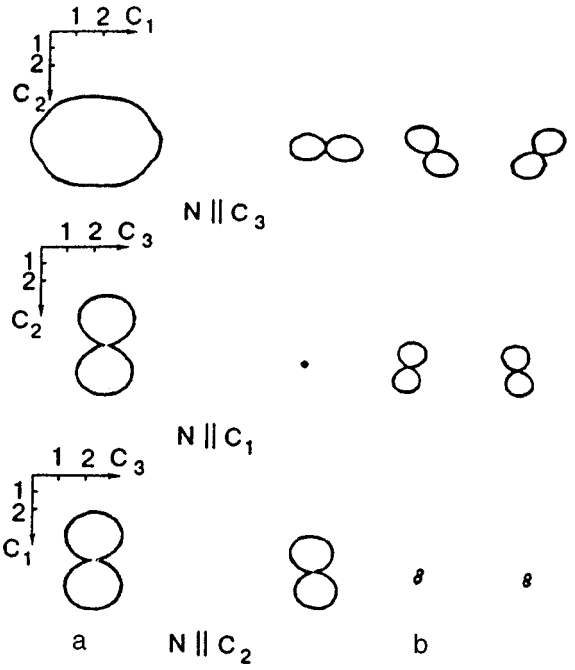

FIG. 4. Angular dependence $F(\varphi)$ of surface magnetic susceptibility of electrons in bismuth: the total contribution (a) and contributions from individual electron ellipsoids (b).

In weak magnetic fields, the field dependence of the magnetic susceptibility $\chi(H)$ of the plate is determined by its surface component $\chi^{(s)}\propto H^{-1/2}$ [see (16)]. For a plate with the normal $\mathbf{N}\|C_3$, the relative variation of the angular function $F(\varphi)$ is small, and we can use any orientation of the magnetic field **H** in the plane of the plate. For $\mathbf{N}\|C_1$ and $\mathbf{N}\|C_2$, the function $F(\varphi)$ has the maximum value in the directions C_2 and C_1, respectively, which must be taken into account while choosing the orientation of **H**.

The obtained angular dependence $\chi^{(s)}(\varphi)$ (see Fig. 4) is due to conduction electrons with orbits tangential to the meatal boundary. Experimental observation of such a dependence of $\chi^{(s)}$ on the angle φ along with its field dependence can serve as a confirmation of the existence of the electron states under investigation. Also, an analysis of the angular dependence of magnetic susceptibility can give information on the electron energy spectrum of metals.

In conclusion, we express our gratitude to the editorial board of LTP for publishing the jubilee issue dedicated to the memory of Il'ya Mikhailovich Lifshits and for the opportunity of participating in this issue. This is especially important for one of the authors (S. N.) whose research work was carried out under the direct support and guidance of I. M. Lifshits starting from post graduate work and finishing with D.Sc. dissertation.

*E-mail: docents@uenpa.kharkov.ua

[1)]The field dependence of the surface magnetic susceptibility $\chi^{(s)}\propto H^{-1/2}$ obtained in Ref. 15, and the corresponding root dependence of $M^{(s)}$ on $H(\chi^{(s)}=\partial M^{(s)}/\partial H)$ are in accord with the results of subsequent calculations of magnetic moment carried out in Ref. 16 by using another method. Shishido[16] analyzed the contribution of MSL to the magnetization of metals once again. The approach used by this author did not allow him to derive an exact expression for the coefficient $C_{-1/6}$ of $H^{-1/3}$. The upper

estimate $|C_{-1/6}| < 0.002$ obtained by Shishido does not contradict the zero result.[14]

[1] L. D. Landau, Z. Phys. **64**, 629 (1930).
[2] I. M. Lifshits, M. Ya. Azbel, and M. I. Kaganov, *Electron Theory of Metals*, Nauka, Moscow (1971).
[3] I. M. Lifshits and A. M. Kosevich, Izv. Akad. Nauk SSSR, Ser. Fiz. **19**, 395 (1955).
[4] I. M. Lifshits and A. M. Kosevich, Zh. Éksp. Teor. Fiz. **29**, 730 (1955) [Sov. Phys. JETP **2**, 636 (1955)].
[5] I. M. Lifshits and A. M. Kosevich, Zh. Éksp. Teor. Fiz. **29**, 743 (1955) [Sov. Phys. JETP **2**, 646 (1955)].
[6] A. S. Kondrat'ev, S. S. Nedorezov, and V. G. Peschanskii, Fiz. Metal. Metalloved. **5**, 5 (1991).
[7] S. S. Nedorezov, Zh. Éksp. Teor. Fiz. **64**, 624 (1973) [Sov. Phys. JETP **37**, 317 (1973)].
[8] I. M. Lifshits and A. M. Kosevich, Dokl. Akad. Nauk SSSR, Ser. Fiz. **91**, 795 (1953) [*sic*].
[9] M. S. Khaikin, Zh. Éksp. Teor. Fiz. **39**, 212 (1960) [Sov. Phys. JETP **12**, 152 (1960)].
[10] T. W. Nee and R. E. Prange, Phys. Rev. Lett. **25A**, 582 (1967).
[11] M. C. Steel, Phys. Rev. **88**, 451 (1952).
[12] R. B. Dingle, Proc. Roy. Soc. **A219**, 463 (1953).
[13] A. M. Kosevich, Ph.D. thesis, Kharkov (1953).
[14] S. S. Nedorezov, Zh. Éksp. Teor. Fiz. **60**, 1938 (1971) [Sov. Phys. JETP **33**, 1045 (1971)].
[15] S. S. Nedorezov, Pis'ma Zh. Éksp. Teor. Fiz. **14**, 597 (1971) [JETP Lett. **14**, 415 (1971)].
[16] F. Shishido, Phys. Rev. Lett. **152A**, 443 (1991).
[17] S. S. Nedorezov, Fiz. Nizk. Temp. **6**, 924 (1980) [Sov. J. Low Temp. Phys. **6**, 449 (1980)].
[18] V. S. Édelman, Usp. Fiz. Nauk **123**, 257 (1977) [Sov. Phys. Uspekhi **20**, 819 (1977)].

Translated by R. S. Wadhwa

The theory of kinetic effects in two-dimensional degenerate gas of colliding electrons

R. N. Gurzhi, A. N. Kalinenko, and A. I. Kopeliovich

*B. Verkin Institute for Low Temperature Physics and Engineering, National Academy of Sciences of the Ukraine, 310164 Kharkov, Ukraine**

(Submitted July 9, 1996)
Fiz. Nizk. Temp. **23**, 58–72 (January 1997)

A mathematical method based on a reduced representation of the electron–electron collision operator acting in the space of quasi-equilibrium functions is constructed. A number of kinetic phenomena such as the evolution of highly anisotropic and high-energy electron distributions, the quasi-hydrodynamic effect in electrical conduction, and a new nonlinear transport mode are described from a unified point of view. Kinetic effects which can be observed in experiments on electron beam propagation and electrical conduction of (GaAs)Al wires with a high mobility of charge carriers are predicted. © *1997 American Institute of Physics.* [S1063-777X(97)00701-9]

INTRODUCTION

A reduction of the dimensionality of an electron system with a weak repulsion from $3D$ to $2D$ does not lead to a rearrangement of its ground state, but modifies significantly the processes of electron–electron relaxation. Energy relaxation in a degenerate two-dimensional ($2D$) electron system is accelerated by a factor of $\ln(\varepsilon_F/T)$ as compared to the $3D$ case (ε_F is the Fermi energy and T the temperature).[1,2] Relaxation in the direction of electron velocity (angular, or momentum relaxation) undergoes more radical changes.[3–5] Collisions of two arbitrary electrons generally leads to scattering through a small angle $\varphi \simeq T/\varepsilon_F \ll 1$. The only exception is the collision of electrons with nearly opposite momenta: the angle of deviation from antiparallelism is $\eta \leqslant T/\varepsilon_F \ll 1$. In this case, scattering can be of any type ($\varphi \simeq 1$). It should be noted that, in contrast to the $3D$ case, the probabilities of both types of processes in a $2D$ system have the same order of magnitude. Processes of the second type are effective in the case of relaxation of an electron distribution even in momentum and corresponding to the relaxation time of the same order of magnitude as in the $3D$ case: $\tau_s \propto (\varepsilon_F/T)^2$. Both types of collisions are ineffective in the case of relaxation of a distribution which is odd in momentum: $\tau_a \simeq \tau_s(\varepsilon_F/T)^2 \propto T^{-4}$. The ineffectiveness of collisions of the second type is due to the fact that the rotation of a pair of electrons with opposite momenta through an arbitrary angle φ does not affect the odd component of the distribution.

These results refer to weakly anisotropic and slightly nonuniform electron distributions. As the characteristic angular scale φ_0 characterizing the deviation from equilibrium decreases, the relaxation time for odd distribution decreases and becomes of the order of τ_s for $\varphi_0 \simeq \sqrt{T/\varepsilon_F}$.[6]

The effects in which the above peculiarities in angular relaxation in $2D$ systems play a significant role were discovered recently in heterostructures. These are hydrodynamic effects in electrical conductivity of wires with a two-dimensional electron gas ($2D$EG), leading to a reverse temperature dependence of resistivity,[7] as well as the experiments with electron beams injected into $2D$EG (see, for example, Refs. 8 and 9; these experiments might provide the most detailed information on angular relaxation). Finally, according to Ref. 10, the effect of $2D$EG entrainment by a ballistic phonon beam[11] can also be used for studying the angular electron relaxation.

This research aims at a consistent theoretical description of momentum relaxation in a $2D$ degenerate electron system and an analysis of some new effects associated with strongly anisotropic electron distributions and nonlinear phenomena. The general scheme of description of angular relaxation at various stages is given in Sec. 1. In Sec. 2, the developed mathematical apparatus is applied for analyzing effects of the hydrodynamic type. The evolution of strongly anisotropic distributions for which electron excitation energy exceeds the temperature significantly (high-energy beams) is considered in Sec. 3. In conclusion, relaxation of strongly anisotropic distributions are considered on qualitative level, and the results pertaining to experimentally observed effects in (GaAs)Al-based structures are formulated.

1. TRANSFORMATION OF COLLISION OPERATOR

1.1. Energy distributions close to equilibrium

The linearized integral of electron–electron collisions can be written in the form

$$J\{\chi_{\mathbf{p}}\} = \frac{\pi^2}{2Th^5} \int W_{\mathbf{p}\mathbf{p}_1\mathbf{p}_2\mathbf{p}_3} n_{\mathbf{p}} n_{\mathbf{p}_1} (1-n_{\mathbf{p}_2})(1-n_{\mathbf{p}_3}) \times (\chi_{\mathbf{p}} + \chi_{\mathbf{p}_1} - \chi_{\mathbf{p}_2} - \chi_{\mathbf{p}_3}) \delta(\varepsilon_{\mathbf{p}} + \varepsilon_{\mathbf{p}_1} - \varepsilon_{\mathbf{p}_2} - \varepsilon_{\mathbf{p}_3}) \times \delta(\mathbf{p} + \mathbf{p}_1 - \mathbf{p}_2 - \mathbf{p}_3) d^2p_1 d^2p_2 d^2p_3,$$

$$W_{\mathbf{p}\mathbf{p}_1\mathbf{p}_2\mathbf{p}_3} = (\Phi_{\mathbf{p}-\mathbf{p}_3} - \Phi_{\mathbf{p}-\mathbf{p}_2})^2 + \Phi^2_{\mathbf{p}-\mathbf{p}_3} + \Phi^2_{\mathbf{p}-\mathbf{p}_2}. \qquad (1)$$

Here $\mathbf{p}$ is the two-dimensional electron momentum, $\varepsilon_{\mathbf{p}}$ the electron energy measured from the Fermi level, $n_{\mathbf{p}}$ the Fermi distribution function, the nonequilibrium correction is written in the form $f_{\mathbf{p}} = -\chi_{\mathbf{p}} \partial n/\partial \varepsilon$ (in should be noted that such a form is convenient only for distributions in which the characteristic energy ε of electron excitations does not exceed significantly the temperature T), and $\Phi_{\mathbf{p}}$ the Fourier transform of the function $\Phi(\vec{\rho})$ connected with the screened electron interaction potential $U(\vec{\rho}, y, y')$ through the following relation:

$$\Phi(\vec{\rho})=\int |\Psi(y')|^2|\Psi(y+y')|^2U(\vec{\rho},y;y')dydy',$$

where $\vec{\rho}$ and y are the components of the vector difference in the coordinates of two electrons in the conduction plane and in the perpendicular direction respectively. In view of screening, U also depends on the transverse coordinate y' of one of the electrons, and $\Psi(y)$ is the wave function of an electron in the layer.

The frequency $\tau_{\rm en}^{-1}$ of electron–electron energy relaxation can be obtained from (1) in the relaxation time approximation, i.e., in the same way as the coefficient of $-\chi_{\mathbf{p}}\partial n/\partial\varepsilon$ in (1):

$$t_{\rm en}^{-1}\simeq\tau_0^{-1}\,\frac{\varepsilon^2+2\pi^2T^2}{\varepsilon_F^2}\,\ln\left(\frac{\varepsilon_F}{T+|\varepsilon|}\right),$$

$$\tau_0^{-1}=\frac{\pi^2Wmp_F^2}{4h^5}. \tag{2}$$

In order to simplify calculations, we assume that the value of W is constant.

It is convenient for subsequent analysis to introduce the following functional which is quadratic in the arbitrary functions $\chi_{\mathbf{p}}$ and $\psi_{\mathbf{p}}$:

$$F\{\psi,\chi\}=\int \psi_{\mathbf{p}}\,J\{\chi_{\mathbf{p}}\}d^2p=\frac{1}{4}\int R_{\mathbf{pp}_1\mathbf{p}_2\mathbf{p}_3}(\psi_{\mathbf{p}}+\psi_{\mathbf{p}_1}-\psi_{\mathbf{p}_2}-\psi_{\mathbf{p}_3})(\chi_{\mathbf{p}}+\chi_{\mathbf{p}_1}-\chi_{\mathbf{p}_2}-\chi_{\mathbf{p}_3})\times d^2pd^2p_1d^2p_2d^2p_3;$$

$$R_{\mathbf{pp}_1\mathbf{p}_2\mathbf{p}_3}=\frac{\pi^2}{2Th^5}\,W_{\mathbf{pp}_1\mathbf{p}_2\mathbf{p}_3}n_{\mathbf{p}}n_{\mathbf{p}_1}(1-n_{\mathbf{p}_2})(1-n_{\mathbf{p}_3})\times\delta(\varepsilon_{\mathbf{p}}+\varepsilon_{\mathbf{p}_1}-\varepsilon_{\mathbf{p}_2}-\varepsilon_{\mathbf{p}_3})\,\delta(\mathbf{p}+\mathbf{p}_1-\mathbf{p}_2-\mathbf{p}_3). \tag{3}$$

We have used the following symmetry properties of the quantity R: $R_{1234}=R_{2134}=R_{1243}=R_{3412}$. The functional $F\{\chi,\chi\}$ is often used in the formulation of the variational principle in kinetics problems.

If the functions $\chi_{\mathbf{p}}$ and $\psi_{\mathbf{p}}$ have a certain parity relative to the substitution $\mathbf{p}\to-\mathbf{p}$, the functional (3) can be written in the form

$$F=\int \Phi_{\mathbf{pkq}}(\hat{D}_{\mathbf{k}}^{\mp}\hat{D}_{\mathbf{q}}^{-}\psi_{\mathbf{p}})(\hat{D}_{\mathbf{k}}^{\mp}\hat{D}_{\mathbf{q}}^{-}\chi_{\mathbf{p}})d^2pd^2kd^2q,$$

$$R_{\mathbf{pp}_1\mathbf{p}_2\mathbf{p}_3}=4\Phi_{\mathbf{p}-\mathbf{k}/2-\mathbf{q}/2,\mathbf{k},\mathbf{q}}\delta(\mathbf{p}+\mathbf{p}_1-\mathbf{p}_2-\mathbf{p}_3),$$

$$\mathbf{k}=\mathbf{p}+\mathbf{p}_1,\quad \mathbf{q}=\mathbf{p}-\mathbf{p}_2=\mathbf{p}_3-\mathbf{p}_1, \tag{4}$$

where the finite difference (sum) operator

$$\hat{D}_a^{\mp}\chi_{\mathbf{p}}=\chi_{\mathbf{p}+\mathbf{a}/2}\mp\chi_{\mathbf{p}-\mathbf{a}/2}.$$

has been introduced. The operators $\hat{D}_{\mathbf{k}}^{-}$ and $\hat{D}_{\mathbf{k}}^{+}$ in (4) correspond respectively to odd and even functions $\chi_{\mathbf{p}}$ (the functions $\chi_{\mathbf{p}}$ and $\psi_{\mathbf{p}}$ are even or odd simultaneously since otherwise $F\equiv0$). Using the arbitrariness of the function ψ, we obtain the following expression from (4) for the collision operator:

$$J\{\chi_{\mathbf{p}}\}=\mp\int \hat{D}_{\mathbf{k}}^{\mp}\hat{D}_{\mathbf{q}}^{-}\Phi_{\mathbf{pkq}}\hat{D}_{\mathbf{k}}^{\mp}\hat{D}_{\mathbf{q}}^{-}\chi_{\mathbf{p}}d^2kd^2q, \tag{5}$$

which is naturally identical to expression (1).

The exact expression (5) itself is not convenient for estimating the angular relaxation frequencies since the quantity Φ is a sharp function of energy. If we assume that the function χ can be regarded as independent of ε in view of the closeness to energy equilibrium (the basis of this assumption will be considered in the next section), the expression for the angular relaxation operator can be integrated with respect to the energies of the states $\mathbf{p},\mathbf{k}-\mathbf{p}$, and $\mathbf{p}-\mathbf{q}$:

$$\int J\{\chi(\alpha)\}d\varepsilon_{\mathbf{p}}\simeq\mp\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}V(\eta,\varphi)\times(\hat{D}_{\eta}^{\mp}\hat{D}_{\varphi}^{-})^2\chi(\alpha)d\eta d\varphi,$$

$$V(\eta,\varphi)=\int \Phi_{\mathbf{pkq}}\,d\varepsilon_{\mathbf{p}}\,d\varepsilon_{\mathbf{k}-\mathbf{p}}\,d\varepsilon_{\mathbf{p}-\mathbf{q}};$$

$$\hat{D}_{\beta}^{\pm}\chi(\alpha)=\chi\left(\alpha+\frac{\beta}{2}\right)\pm\chi\left(\alpha-\frac{\beta}{2}\right); \tag{6}$$

where α is the angular variable corresponding to the state $\mathbf{p}$. It is more convenient to derive this expression not from (5), but from (4) by assuming that the function ψ is also independent of energy. While deriving (6), we assumed that fixed values of the quantities q and k correspond to fixed values of angular displacements φ and η: $\sin(\varphi/2)=q/2p_F$, $\sin(\eta/2)=k/2p_F$. Deviations from this correspondence of the order of T/ε_F are significant only in the angular intervals $\eta\leqslant T/\varepsilon_F$ and $\pi-\eta\leqslant T/\varepsilon_F$. However, it can be verified that the inclusion of the contribution from these regions leads only to insignificant changes in the numerical coefficients. If we assume that the value of W changes insignificantly upon a change in the energy of electron states in the region of thermal blurring of the Fermi surface, i.e., $W\simeq W(\eta,\varphi)$, we can write

$$V(\eta,\varphi)=\frac{\pi^2m^3}{8Th^5}\,\frac{W(\eta,\varphi)}{1+\cos\,\eta}\left[\frac{\Delta}{2\sinh(\Delta/2T)}\right]^2;$$

$$\Delta=2\varepsilon_F\,\frac{\sin\,\varphi\,\sin\,\eta}{1+\cos\,\eta};\quad \eta,\pi-\eta\gg T/\varepsilon_F; \tag{7}$$

where Δ is the energy transfer during a collision.

Let us also consider angular relaxation, taking into account weak dependence of χ on ε. In the stationary and spatially homogeneous case, the energy dependence has the form of a ''local drift'' [3]:

$$\chi(\alpha,\varepsilon)\simeq\overline{\chi(\alpha)}-\frac{\varepsilon}{2\varepsilon_F}\,\frac{d^2\overline{\chi(\alpha)}}{d\alpha^2}, \tag{8}$$

where $\overline{\chi(\alpha)}$ is an arbitrary odd function. Such a relation between the energy and angular dependences, which is also observed for the true drift $\chi=\mathbf{u}\cdot\mathbf{p}$, is imposed by ''horizontal transitions,'' i.e., collisions with a small energy transfer: $\Delta\ll qv_F$ (but at the same time, $q\ll p_F$). It is convenient to use in calculations the functional (4) with the function ψ which also has the structure (8). The result has the form

$$F\{\psi,\chi_a\}\simeq \int\int_{-\pi}^{\pi}\int \overline{\psi(\alpha)}\Big\{V(\eta,\varphi)\Big(\hat{D}_\eta^-\hat{D}_\varphi^- - \frac{\Delta}{\varepsilon_F}\hat{D}_\eta^+\hat{D}_\varphi^+\frac{d^2}{d\alpha^2}\Big)^2 - V_1(\eta,\varphi) \times(\hat{D}_\varphi^-)^2\frac{d^4}{d\alpha^4}\Big\}\overline{\chi(\alpha)}d\eta\, d\varphi\, d\alpha,$$

$$V_1(\eta,\varphi)=\frac{1}{3}V(\eta,\varphi)\frac{\Delta^2+4\pi^2T^2}{\varepsilon_F^2}. \tag{9}$$

In view of the arbitrariness of $\overline{\psi(\alpha)}$ and the fact that

$$F\{\psi,\chi_a\}\simeq F\{\overline{\psi},\chi_a\}, \tag{10}$$

(this relation is proved in Appendix A), relations (9) immediately lead to the expression for the operator $\int J\{\chi_a\}d\varepsilon$ of odd angular relaxation.

For small η and φ, the operators $\hat{D}_\eta^\pm$ and $\hat{D}_\varphi^\pm$ can be expanded into series in this parameters. It is interesting to note that the principal terms of the expansion, which are proportional to $(\eta\varphi)^2$, are present in (6), but are cancelled out in expression (9) (in the term with V). Thus, the inclusion of even a weak energy dependence of the function χ is in general significant for estimating the rate of odd angular relaxation.

1.2. High-energy nonequilibrium distributions ($T\ll|\varepsilon|\ll\varepsilon_F$)

If the excitation energy for nonequilibrium electrons is much higher than temperature, the value of $\chi=-(dn/d\varepsilon)^{-1}f$ increases exponentially with $|\varepsilon|$ within the excitation region (in which f is not vanishingly small). Consequently, in the parentheses in (1) containing χ we can neglect three terms in comparison with the one corresponding to the maximum value of $|\varepsilon|$. This means that, for $T\to 0$, the collision integral can be written in the form

$$J\{f_{\mathbf{p}}\}=-\int(\nu_{\mathbf{p}'\mathbf{p}}f_{\mathbf{p}}-\nu_{\mathbf{p}\mathbf{p}'}f_{\mathbf{p}'})d^2p',$$

$$\nu_{\mathbf{p}\mathbf{p}'}=\frac{\pi^2}{2h^5}\begin{cases}2\int U_{\mathbf{p}\mathbf{p}_1\mathbf{p}_2\mathbf{p}'}d^2{}_{p_1}d^2p_2, \\ \quad \varepsilon\varepsilon'>0,\varepsilon\varepsilon_1>0,\varepsilon\varepsilon_2<0,\\ -\int U_{\mathbf{p}\mathbf{p}'\mathbf{p}_2\mathbf{p}_3}d^2p_2d^2p_3, \\ \quad \varepsilon\varepsilon'<0,\varepsilon\varepsilon_2<0,\varepsilon\varepsilon_3<0,\end{cases}$$

$$U_{\mathbf{p}\mathbf{p}_1\mathbf{p}_2\mathbf{p}_3}=W_{\mathbf{p}\mathbf{p}_1\mathbf{p}_2\mathbf{p}_3}\delta(\varepsilon+\varepsilon_1-\varepsilon_2-\varepsilon_3)\times\delta(\mathbf{p}+\mathbf{p}_1-\mathbf{p}_2-\mathbf{p}_3),$$

$$\varepsilon\equiv\varepsilon_{\mathbf{p}},\quad \varepsilon'\equiv\varepsilon_{\mathbf{p}'},\quad \varepsilon_1\equiv\varepsilon_{\mathbf{p}_1},\quad \varepsilon_2\equiv\varepsilon_{\mathbf{p}_2}. \tag{11}$$

It can be seen from these formulas that $\nu_{\mathbf{p},\mathbf{p}'}=0$ for $|\varepsilon'|<|\varepsilon|$. This is not surprising since at the given ''pretemperature'' stage of relaxation, the departure from the state $\mathbf{p}$ is realized only downwards on the scale of energy values, while the arrival in this state is possible only from above. In this approximation, instead of (3) we have

$$F\{\psi,f\}=-\int U_{\mathbf{p}\mathbf{p}_1\mathbf{p}_2\mathbf{p}_3}(\psi_{\mathbf{p}}+\psi_{\mathbf{p}_1}-\psi_{\mathbf{p}_2}-\psi_{\mathbf{p}_3})\times f_{\mathbf{p}}d^2pd^2p_1d^2p_2d^2p_3$$
$$=-\int\nu_{\mathbf{p}\mathbf{p}'}(\psi_{\mathbf{p}}-\psi_{\mathbf{p}'})\times f_{\mathbf{p}'}d^2pd^2p',$$
$$|\varepsilon_2|,|\varepsilon_3|,|\varepsilon_4|<|\varepsilon_1|. \tag{12}$$

Assuming that the electron spectrum is quadratic and isotropic, and the quantity W is constant, we obtain the following expression (11), accurate to within insignificant quantities of the order of $\varepsilon/\varepsilon_F$:

$$\nu_{\mathbf{p}\mathbf{p}'}\simeq 2C\begin{cases}\dfrac{|K(\varphi,\epsilon)-K(\varphi,\epsilon')|}{\zeta\,\sin(\varphi/2)}, & \epsilon\epsilon'>0,\\ -\arcsin\dfrac{|\epsilon+\epsilon'|}{\sqrt{(\delta\epsilon)^2+4\sin^2\varphi}}, & \epsilon\epsilon'<0,\end{cases}$$

$$K(\varphi,\epsilon)=(\cos^2(\varphi/2)-\epsilon\zeta)^{1/2}\Theta(\cos^2(\varphi/2)-\epsilon\zeta);\zeta=1+\left(\frac{\delta\epsilon}{4\sin\varphi/2}\right)^2;$$

$$\epsilon\equiv\varepsilon_{\mathbf{p}}/\varepsilon_F,\quad \epsilon'\equiv\varepsilon_{\mathbf{p}'}/\varepsilon_F;\quad \delta\epsilon=\epsilon'-\epsilon;\quad |\epsilon|,|\epsilon'|\ll 1;$$
$$C\equiv 2\tau_0^{-1}p_F^{-2}; \tag{13}$$

where φ is the angle between $\mathbf{p}$ and $\mathbf{p}'$, and $\Theta(x)$ the Heaviside function. For angles φ for which expression (13) can be expanded into a power series in ϵ,ϵ', we have

$$\nu_{\mathbf{p}\mathbf{p}'}\simeq\frac{C}{\sin\varphi}\begin{cases}2|\delta\epsilon|, \\ \quad \epsilon\epsilon'>0;\ \ \varphi\gg|\delta\epsilon|\sqrt{|\epsilon'|},\pi-\varphi\gg\sqrt{|\epsilon'|}\\ -|\epsilon+\epsilon'|, \\ \quad \epsilon\epsilon'<0;\ \ \varphi,\pi-\varphi\gg|\delta\epsilon|.\end{cases} \tag{14}$$

The kernel $\nu_{\mathbf{p}\mathbf{p}'}$ in (14) was found to be even in momentum, i.e., invariant relative to the substitution $\varphi\to\pi-\varphi$ (by definition, the angle φ varies in the limits $0\leqslant\varphi\leqslant\pi$). The odd component appears in the next approximation:

$$\nu^a_{\mathbf{p}\mathbf{p}'}=-C\frac{\cos\varphi}{\sin^3\varphi}(\epsilon'^2-\epsilon^2)\Theta(\epsilon\epsilon')\mathrm{sgn}\,\epsilon';$$
$$\varphi,\pi-\varphi\gg\sqrt{|\epsilon'|}. \tag{15}$$

A rapid decrease in the odd component of the kernel $\nu^2_{\mathbf{p}\mathbf{p}'}\sim\varphi^{-3}$ is in accord with the general statement made in Refs. 3 and 4 concerning the low rate of odd angular relaxation. It should be noted, however, that the evenness of the kernel in approximation (14) is a consequence of the above assumption about the constancy of W. The expression obtained earlier in Ref. 12 and actually analogous to (14) does not possess a definite parity. In the general case, for an arbitrary dependence of W on its variables, the slow rate of odd relaxation follows from (12). Indeed, is we take for ψ a slowly varying odd function of the angle independent of energy, the functional $F\{\psi,f\}\sim\epsilon^3$; the additional degree of

smallness in comparison with the collision frequency $\tau_{\rm en}^{-1}\sim\epsilon^2$ appears in the expansion of the parentheses with four ψ into a power series in the small momentum transfer $\mathbf{p}_3-\mathbf{p}_1$, or in the small total momentum $\mathbf{p}+\mathbf{p}_1$ of the colliding pair since $|\mathbf{p}+\mathbf{p}_1|\cdot|\mathbf{p}_3-\mathbf{p}_1|/p_F^2\leqslant\epsilon$. (It should be noted that the functional $F\{\psi,\chi\}$ is the total change in the value of $\psi_{\mathbf{p}}$ per unit time due to collisions, and hence characterizes the rate of odd angular relaxation associated with transport of particles through large angles for a give choice of ψ.) If, however, we assume that W is a smooth function of its variables, we can use (12) to obtain more detailed estimates associated with odd angular relaxation. We choose the function f in (12) in the form of $f_{\mathbf{p}}=\delta(\mathbf{p}-\mathbf{p}_0)$ and assume that the odd function $\psi(\varphi)$ is equal to 1/2 for $0<\varphi<\tilde{\varphi}$ and $\psi=0$ for $\tilde{\varphi}<\varphi<\pi/2$, where φ is the angle between $\mathbf{p}$ and $\mathbf{p}_0$ and $\sqrt{\epsilon}_0\ll\tilde{\varphi}<\pi/2,\varepsilon_0$ being the energy of the state $\mathbf{p}_0$. In this case, the expression in the parentheses with ψ in (12) differs from zero only if three of the states contained in it lie in regions of nonzero values of ψ, while the fourth state (say, $\mathbf{p}_3$) lies outside this region. This means that the contribution to the integral comes only from narrow neighborhoods of the angle $\tilde{\varphi}$ (of the order of $\varepsilon_0/\tilde{\varphi}_0$ for the states $\mathbf{p}_1$ and $\mathbf{p}_3$; see the above inequality for the values of $|\mathbf{p}_3-\mathbf{p}_1|$ and $|\mathbf{p}+\mathbf{p}_1|$).

Consequently, it follows from (12) and (11) that

$$m\int\int_{-\tilde{\varphi}}^{\tilde{\varphi}}J\{\delta^a(\mathbf{p}-\mathbf{p}_0)\}d\varepsilon\ d\varphi=m\int\int_{-\tilde{\varphi}}^{\tilde{\varphi}}\nu^a_{\mathbf{pp}_0}d\varepsilon\ d\varphi$$
$$-\frac{1}{2}\int\nu_{\mathbf{pp}_0}d^2p\sim\epsilon_0^3\tilde{\varphi}^{-2};$$

$$2\,\delta^a(\mathbf{p}-\mathbf{p}_0)=\delta(\mathbf{p}-\mathbf{p}_0)-\delta(\mathbf{p}+\mathbf{p}_0). \tag{16}$$

The additional power of ϵ_0 appears due to integration with respect to ε_1 in the thermal layer. Differentiating (16) with respect to $\tilde{\varphi}$, we obtain the following relation which is in accord with (15):

$$\int\nu^a_{\mathbf{pp}'}d\varepsilon\sim(\epsilon')^3\varphi^{-3};\quad\varphi,\pi-\varphi\gg\sqrt{|\epsilon'|}. \tag{17}$$

2. EFFECTS OF THE HYDRODYNAMIC TYPE IN ELECTRICAL CONDUCTION OF TWO-DIMENSIONAL CONDUCTORS

The temperature minimum in the resistivity of 2*D*EG wires discovered by Molekamp and de Jong[7] indicates a manifestation of the hydrodynamic mechanism of electrical conduction. It was shown in Ref. 13 and 14 that electron–electron collisions in 2*D* conductors can lead not only to a conventional Poiseuille flow of electron liquid over a bounded 2*D*EG sample,[15] but also to new effects of the hydrodynamic type. The latter were obviously observed in the experiments.[7] In this section, we develop the theory of inhomogeneous current states in 2*D*EG, which provides a description of these hydrodynamic effects.

The linearized kinetic equation for the nonequilibrium correction to the distribution function has the form

$$v_z\frac{\partial\chi}{\partial z}+I\{\chi\}=e\mathbf{E}\cdot\mathbf{v},\quad I=\left(\frac{\partial n}{\partial\varepsilon}\right)^{-1}J. \tag{18}$$

This equation is written for a strip $-d/2\leqslant z\leqslant+d/2$; the electric field $\mathbf{E}$ is directed along the strip, and the collision integral J is defined by (1). We separate the components in Eq. (18) with different parities in momentum:

$$v_z\frac{\partial\chi_s}{\partial z}+I\{\chi_a\}=e\mathbf{E}\cdot\mathbf{v}, \tag{19}$$

$$v_z\frac{\partial\chi_a}{\partial z}+I\{\chi_s\}=0. \tag{20}$$

If we can neglect low-rate odd relaxation altogether, in the case of diffuse scattering of electrons at the boundary we have

$$\chi_s=\frac{z}{v_z}\,e\mathbf{E}\cdot\mathbf{v},$$
$$\chi_a=\left(\frac{d^2}{8}-\frac{z^2}{2}\right)\frac{1}{v_z}\,I\left\{\frac{e\mathbf{E}\cdot\mathbf{v}}{v_z}\right\}-\frac{de\mathbf{E}\cdot\mathbf{v}}{2|v_z|}. \tag{21}$$

In this approximation, the current density

$$j(z)=\frac{4\varepsilon_F m^2}{h^2}\int\chi_a(z,\alpha)\cos\,\alpha\ d\alpha \tag{22}$$

diverges for small angles $\alpha\simeq v_z/v$ since the first term for χ_a in (21) behaves approximately as $\alpha^{-2}(I\{\alpha^{-1}\}\sim\alpha^{-1}$, a more exact estimate will be given below). Thus, ''grazing'' electrons flying at a small angle α to the boundary make the main contribution to conduction, which can be determined only by taking into account the low-rate odd relaxation. It should be emphasized that the main singularity in (21) is associated with the first term for χ_a describing effects of the hydrodynamic type: this term is principal for all values of α in the hydrodynamic limit $d\gg l_s=v_F\tau_s$.

It is impossible to solve the system of equations (19), (20) exactly. However, a semiquantitative description of hydrodynamic-type phenomena can be obtained by using the following self-consistent procedure. We introduce the mean free paths for the even and odd angular relaxation, which are determined by the characteristic angular scale φ_0 of variation of the function χ:

$$I\{\chi_s\}\simeq\frac{\chi_s v_F}{l_s(\varphi_0)},\quad I\{\chi_a\}\simeq\frac{\chi_a v_F}{l_a(\varphi_0)}.$$

In this approximation, we can easily obtain a solution of Eqs. (19) and (20), which, however, is rather cumbersome. To the same accuracy, we can use the following compact expression:

$$\chi_a\simeq\frac{d^2 l_a(\varphi_0)e\mathbf{E}\cdot\mathbf{v}}{d^2+\alpha^2 l_s(\varphi_0)l_a(\varphi_0)},$$
$$\varphi_0\simeq d/\sqrt{l_s(\varphi_0)l_a(\varphi_0)}. \tag{23}$$

For the transport mean free path $l_{\rm tr}$ determined by the Lifshits relation

$$\frac{1}{d}\int_{-d/2}^{d/2}j(z)dz=\frac{e_N^2}{p_F}\,l_{\rm tr}\,E,$$

where N is the two-dimensional density of charge carriers, we obtain

$$l_{\rm tr} \simeq \frac{d^2}{l_s(\varphi_0)\varphi_0} \simeq d\sqrt{\frac{l_a(\varphi_0)}{l_s(\varphi_0)}}. \tag{24}$$

If the dependences $l_s(\varphi_0)$ and $l_a(\varphi_0)$ are known, relation (23) is an equation for φ_0 that can be used to express φ_0 in terms of the quantities $d, l_s, l_a (l_s \propto T^{-2}, l_a \simeq (\varepsilon_F/T)^2 l_2 \propto T^{-4}$ correspond to $\varphi_0 \simeq 1$). Let us find the dependences $l_s(\varphi_0), l_a(\varphi_0)$ with the help of expressions (6) and (9). It follows from (7) that the inequality $|\varphi\eta| \leqslant \gamma \equiv T/\varepsilon_F$, determining the ranges of the angles φ and η which make the contribution in formulas (6) and (9) that cannot be regarded as exponentially small.

Let us first consider not too anisotropic distributions: $\varphi_0 \gg \sqrt{\gamma}$. In this case, the total interval of variation of η can be divided into three asymptotic regions: region I with $\eta \ll \gamma/\varphi_0$, in which $\eta \ll \varphi_0$, and $\varphi_{\max} \gg \varphi_0$; region II with $\gamma/\varphi_0 \ll \eta \ll \varphi_0$, in which $\eta, \varphi \ll \varphi_0$, and region III with $\eta \gg \varphi_0$, in which $\varphi \ll \varphi_0$. The operator $\hat{D}_\eta^{\pm}$ can be expanded in the small parameter η/φ_0: $\hat{D}_\eta^{+}\chi(\alpha) \simeq 2\chi(\alpha)$ and $\hat{D}_\eta^{-}\chi(\alpha) \simeq \eta\chi'(\alpha)$ in region I, in the parameter φ/φ_0 in region III, and in both these parameters in region II. While evaluating $l_s(\varphi_0)$, we can obviously neglect the contribution from regions II and III, which is proportional to the small parameter $(\varphi/\varphi_0)^2$. In region I, we can neglect the integral terms $(\hat{D}_\varphi^{-})^2\chi(\alpha) \simeq 2\chi(\alpha)$. As a result, an approximate estimate of the quantity $l_s(\varphi_0)$ has the form

$$l_s^{-1}(\varphi_0) - \gamma \int_{\gamma}^{\gamma/\varphi_0} d\eta \int_0^{\gamma/\eta} d\varphi \sim 2 l_0^{-1} \gamma^2 \ln(\varphi_0^{-1}),$$

$$\varphi_0 \gg \sqrt{\gamma}, \tag{25}$$

where $l_0 = v_F \tau_0$, and τ_0 has the form (2). While estimating the value of $l_a(\varphi_0)$ by formula (6), we can write the contributions from the above regions in the form

$$l_a^{-1}(\varphi_0) \sim \gamma \left\{ \int_0^{\gamma/\varphi_0} d\eta \left(\frac{\eta}{\varphi_0}\right)^2 \int_0^{\gamma/\eta} d\varphi + \int_{\gamma/\varphi_0}^{\varphi_0} d\eta \left(\frac{\eta}{\varphi_0}\right)^2 \int_0^{\gamma/\eta} d\varphi \left(\frac{\varphi}{\varphi_0}\right)^2 + \int_{\varphi_0}^{\pi} d\eta \int_0^{\gamma/\eta} d\varphi \left(\frac{\varphi}{\varphi_0}\right)^2 \right\} \sim l_0^{-1} \frac{\gamma^4}{\varphi_0^4} \ln \frac{\varphi_0^2}{\gamma},$$

$$\varphi_0 \gg \sqrt{\gamma}. \tag{26}$$

Region III makes the dominating contribution in the parameter $\ln(\varphi_0^2/\gamma)$.

It should be recalled that expression (6) corresponds to the assumption that the function χ_a is independent of energy. It will be shown in Appendix A that for relatively large φ_0, the energy dependence acquires the form of a ''local drift'' (8), and the value of $l_a(\varphi_0)$ should be estimated by using expression (9). In this case, the expansion in region II leads to terms of the order of

$$(\varphi\eta/\varphi_0^2)^2[(\varphi/\varphi_0^2) + (\eta/\varphi_0^2)],$$

which reduces its contribution as compared to (26); on the contrary, the contribution from regions I and III increases and becomes predominant:

$$l_a^{-1}(\varphi_0) \sim l_0^{-1} \frac{\gamma^4}{\varphi_0^4} \ln(\varphi_0^{-1}), \quad \varphi_0 \gg \sqrt{\gamma}. \tag{27}$$

Thus, the inclusion of the energy dependence of χ_a leads to a redistribution of the contribution to $l_a(\varphi_0)^{-1}$ between the region, which does not change the result significantly. It will be shown in Appendix A that the true value of $l_a(\varphi_0)^{-1}$ for $\varphi_0 \gg \sqrt{\gamma}$ is determined by the larger of the quantities (26) and (27).

In the case of more anisotropic distributions ($\varphi_0 \ll \sqrt{\gamma}$), the division of the total interval of variation of η into three asymptotic region is carried out as follows: region I with $\eta \ll \varphi_0$, in which the expansion in the parameter η/φ_0 is possible; region III with $\eta \gg \gamma/\varphi_0$, in which the expansion can be carried out in the parameter φ/φ_0, and region II with $\varphi_0 \ll \eta \ll \gamma/\varphi_0$, in which the expansion is not possible. In the case of an even distribution, region III is ineffective, while regions I and II give

$$l_s^{-1}(\varphi_0) \sim l_0^{-1} \gamma^2 \ln(\gamma^{-1}), \quad \varphi_0 \ll \sqrt{\gamma}. \tag{28}$$

For odd distributions, regions I and III are ineffective as compared to region II, and its contribution, according to (6), has the form

$$l_a^{-1}(\varphi_0) \sim l_0^{-1} \gamma^2 \ln \frac{\gamma}{(\gamma + \varphi_0)^2}, \quad \varphi_0 \ll \sqrt{\gamma}. \tag{29}$$

Expressions (25)–(29) refine the dependences $l_s(\varphi_0)$ and $l_a(\varphi_0)$ obtained in Refs. 6 and 14. It should be noted the a more accurate analysis carried out here establishes a slow dependence $l_s(\varphi_0)$ which was not observed in Refs. 6 and 14. It can be seen from relations (25)–(27) that the ratio of lengths $l_a(\varphi_0)/l_s(\varphi_0) \gg 1$ and decreases rapidly with φ_0 for $\varphi_0 \gg \sqrt{\gamma}$. This leads to a decreases in the transport length with φ_0 (see (24)), and hence to the inverse temperature dependence of resistivity $l_{\rm tr}^{-1} \propto T^{-1}$ for $d^2 \ll \gamma l_s^2$ (the latter inequality corresponds to $\varphi_0 \gg \sqrt{\gamma}$). However, it can be seen from (28) and (29) that $l_a(\varphi_0) > l_s(\varphi_0)$, and for $\varphi_0 \ll \sqrt{\gamma}$ the inverse temperature dependence defined by (24) is weaker than the linear dependence determined by the conventional Fuchs term $d \ln(l_s/d)$ [disregarded in (24)]. Thus, for $d^2 \ll \gamma l_s^2$ we are dealing with the Fuchs situation in electrical conduction with a weak increase in resistivity with temperature.

The above analysis carried out in the framework of two relaxation times has the upper limit on the plane thickness d defined by the inequality $d^2 \ll l_s l_a$. In this limit, $\varphi_0 \ll 1$, and hence we can disregard the conservation of the total electron momentum in collisions: the departure of a nonequilibrium electron from the grazing region of width φ_0 during the time $v_F^{-1} l_a(\varphi_0)$ leads to a rapid relaxation at the boundary, i.e., is equivalent to its disappearance. In the opposite limit $d^2 \gg l_s l_a$, this approximation is inapplicable, and it can be proved that the conventional hydrodynamic situation is realized.

Till now, the analysis was carried out in the linear approximation. However, inverse temperature dependence of resistivity was observed by Molekamp and de Jong[7] under nonlinear conditions, when nonlinearity is ''trivial,'' i.e., is reduced to the necessity of taking into account the depen-

dence of temperature on the current if the temperature exceeds the correction to the chemical potential $\chi(\alpha)$. If the electron system loses its energy due to emission of phonons (the mean free path for these collisions is l_{ep}), and the condition $l_s \ll l_{ep}$ ensuring the closeness of the electron system to equilibrium in energy is satisfied, it can easily be verified that the linearity condition has the form $l_{ep}\varphi_0^2 \gg l_{\rm tr}$. This condition is obviously violated for $l_a(\varphi_0) \gg l_{ep}$, when the odd relaxation associated with electron–phonon collisions dominates. In the opposite limiting case $l_{ep}\varphi_0^2 \ll l_{\rm tr}$, we cannot use the linearized collision integral (1). (For the sake of simplicity, we assume that the electron energy and momentum change significantly during electron–phonon collisions.)

Essentially nonlinear electron–electron relaxation in two-dimensional systems requires special consideration; here we only indicate the following important circumstance. The nonequilibrium distribution corresponding to a shift in chemical potential which is odd in momentum (i.e., the distribution of the form $n[\varepsilon_{\mathbf{p}} - \chi(\alpha)]$ with $\chi(-\alpha) = -\chi(\alpha)$, $\chi \ll \varepsilon_F$) for $T \to 0$ relaxes very slowly to the equilibrium state (at any rate, relaxation is absent to within quantities of the order of $(\chi/\varepsilon_F)^3$, inclusively). In order to verify this, note that, by virtue of conservation laws, the two initial states $\mathbf{p}$ and $\mathbf{p}_1$ and two final states $\mathbf{p}_2$ and $\mathbf{p}_3$ lie on the same circle in the $\mathbf{p}$-space. The points $\mathbf{p}$ and $\mathbf{p}_1$ as well as the points $\mathbf{p}_2$ and $\mathbf{p}_3$ are diametrically opposite relative to the center of this circle. Let us first consider a circle of radius p_F displaced relative to the Fermi surface (FS) through a small distance δp. This circle will contain the initial and final states of collision if it includes pairs of diametrically opposite points corresponding to occupied states as well as pairs of opposite points corresponding to free states. However, the displacement of points of distorted FS relative to the circle under investigation as well as their displacement relative to the unperturbed FS is odd in the first order in χ/ε_F. This means that free regions of the circle are opposite to occupied regions, and hence collisions are ruled out. An increase in the radius of the circle does not change the situation: final states but not the initial state become possible (the converse situation takes place upon a decrease in the radius). A considerable increase in displacement δp does not change the situation either. In the second approximation in χ/ε_F, the situation changes. Consequently, the role of effective temperature in expression (2) for $\tau_{\rm en}$ is played by the quantity χ^2/ε_F. Preliminary analysis shows that the relaxation time for the FS deformed in the odd way can be even higher than $\tau_0(\varepsilon_F/\chi)^4$. In the case of the evenly deformed FS, the role of effective temperature is played by the quantity χ, and the relaxation time is of the order of $\tau_0(\varepsilon_F/\chi)^2$.

3. EVOLUTION OF STRONGLY ANISOTROPIC HIGH-ENERGY DISTRIBUTIONS

Although spatial dispersion effects play a significant role in $2D$ electron beams, we shall confine ourselves in this section to the analysis of pretemperature relaxation in momentum space, which is described by the equation

$$\partial f_{\mathbf{p}}/\partial t = J\{f_{\mathbf{p}}\}, \tag{30}$$

where the collision integral $J\{f_{\mathbf{p}}\}$ has the form (11). It will be indicated in the Conclusion how the results of this research can be used for analyzing experiments with electron beams, i.e., taking spatial dispersion into account.

Let us suppose that the distribution of nonequilibrium electrons at the initial instant is described by the function $f_0 = A p_F^{-2}\delta(\mathbf{p} - \mathbf{p}_0)$, where $\varepsilon_0 \gg T$. The coefficient of the δ-function obviously decreases with time in proportion to $\exp(-t/\tau_{\rm en})$ with $\tau_{\rm en}(\varepsilon_{\mathbf{p}})$ from (2) $(T=0)$.

For times $t \ll \tau_{\rm en}(\varepsilon_0)$, the distribution of electrons experiencing a collision can be determined in the first approximation in the collision integral:

$$f_{\mathbf{p}} \simeq A p_F^{-2} \nu_{\mathbf{p}\mathbf{p}_0} t. \tag{31}$$

We shall be interested in the evolution of the distribution for times exceeding $\tau_{en}(\varepsilon_0)$. Obviously, the energy of nonequilibrium electrons decreases after each collision (to approximately one third of the previous value) since a nonequilibrium electron shares energy with equilibrium electrons. Thus, nonequilibrium electrons ''fall'' on the FS (holes rise above it), each next collision occurring at a much lower rate than the previous collision ($\tau_{\rm en} \sim \varepsilon^{-2}$). In the limit $\varepsilon \to 0$, the process terminates: $\nu_{\mathbf{p}\mathbf{p}'} = 0$ for $\varepsilon_{\mathbf{p}'} = 0$. Consequently, an arbitrary angular distribution of electrons over the Fermi surface is a solution of the equation $J\{f\} = 0$ describing the stationary situation which is final for the pretemperature stage. Since the total energy of electrons is conserved during collisions, we must distinguish between two types of solutions of the steady-state equation

$$f_\infty = \lambda(\alpha)\delta(\varepsilon) + \kappa(\alpha)\delta'(\varepsilon). \tag{32}$$

The second term ensures a nonzero energy of the final distribution.

Let us formulate the following question: what is the final distribution of particles over the FS for the pretemperature relaxation stage? In other words, what is the limit of the solution of Eq. (30) with an arbitrary initial condition $f(t=0) = f_0$ for $t \to \infty$? In order to answer this question, we multiply both sides of Eq. (30) by the function $\tilde{f}$, which is a solution of the equation

$$\tilde{J}\{\tilde{f}\} = 0,$$

$$\tilde{J}\{\tilde{f}\} = -\int \nu_{\mathbf{p}'\mathbf{p}}(\tilde{f}_{\mathbf{p}} - \tilde{f}_{\mathbf{p}'})\, d^2p' \tag{33}$$

($\tilde{J}$ is the transposed operator for J) and integrate with respect to $\mathbf{p}$ and time between $t=0$ and ∞ for $f = f_\infty$. This gives the following relation between the final and initial distributions:

$$m\int_0^{2\pi}\left[\lambda(\alpha)\tilde{f}(\alpha,\varepsilon) + \kappa(\alpha)\,\frac{\partial}{\partial\varepsilon}\,\tilde{f}(\alpha,\varepsilon)\right]\Big|_{\varepsilon=0} d\alpha$$

$$= \int f_0 \tilde{f} d^2p. \tag{34}$$

The solutions of Eq. (33) are quantities conserved during collisions. The functions $\tilde{f} = \text{const}$, $\mathbf{p}$, and ε are obviously such solutions in view of the conservation of the number of particles, momentum, and energy. Do other solutions of Eq. (33) exist? In view of isotropy of the problem, such solutions

can be sought in the form $\tilde{f}_n(\varepsilon)\exp(i\alpha n)$. Assuming that solutions with $\tilde{f}_n(0) \neq 0$ exist and omitting the arbitrary coefficient, we can write $\tilde{f}_n(\varepsilon)=1-g_n(\varepsilon), g_n(\varepsilon)\to 0$ for $\varepsilon\to 0$. As a result, we obtain the following equation for g_n from (33):

$$g_n(\varepsilon)-\tau_{\rm en}(\varepsilon)\int \nu_{\mathbf{p}'\mathbf{p}} g_n(\varepsilon')\mathrm{e}^{i\alpha' n}d^2p'=\tau_{\rm en}(\varepsilon)\tilde{J}\{\mathrm{e}^{i\alpha n}\}. \tag{35}$$

In order to analyze this equation, it is important to estimate its right-hand side.

For odd n, we have the following order-of-magnitude estimate:

$$\tau_{\rm en}(\varepsilon)\tilde{J}\{\mathrm{e}^{i\alpha n}\}\simeq n^2\epsilon\,\frac{\ln(n^2|\epsilon|)}{\ln|\epsilon|}, \quad |\epsilon|n^2\ll 1. \tag{36}$$

The smallness of this expression for $\epsilon\to 0$ is a direct consequence of the low rate of odd angular relaxation considered in Sec. 1. It is convenient to obtain estimate (36) from (12) by putting $f_{\mathbf{p}}=\delta(\mathbf{p}-\mathbf{p}_0), \psi_{\mathbf{p}}=\mathrm{e}^{i\alpha n}$ and dividing the integration domain into three regions in analogy with the computational algorithm used for deriving expression (26). A rapid decrease in the right-hand side of (35) with ϵ makes it possible to solve this equation by iterations in the second term on the left-hand side of (35). Indeed, $\tilde{J}\{\tilde{g}_n\}\approx\tau_{\rm en}^{-1}g_n$ since in this case we are speaking of energy relaxation, and there are no grounds for the compensation of the nonintegral and integral terms on the left-hand side of (35). Moreover, according to estimates, the ratio of these terms is small:

$$\tau_{\rm en}\, g_n^{-1}\tilde{J}\{g_n\mathrm{e}^{i\alpha n}\}-1\simeq\frac{1}{6}\left[1-\frac{3}{2}\,\frac{\ln n^2}{\ln(|\epsilon|^{-1})}\right]\leqslant\frac{1}{6}.$$

Thus, Eq. (35) is solvable for odd n, and hence additional odd solutions of Eq. (33) with $\tilde{f}_n(0)\neq 0$ do exist. It should be noted that for reasons similar to those considered in the previous section, the functions g_n with small n correspond to the ''local drift'' (8):

$$g_n(\varepsilon)\simeq-\frac{1}{2}\,\epsilon n^2\mathrm{e}^{i\alpha n}, \quad |\epsilon|\ll 1, \tag{37}$$

while for $n=\pm 1$ this expression corresponds to actual drift, i.e., exact solution of Eq. (33), to a high degree of accuracy.

For even n, we have

$$\tau_{\rm en}(\varepsilon)\tilde{J}\{\mathrm{e}^{i\alpha n}\}\simeq\frac{\ln(|n|+1)}{\ln(|\epsilon|^{-1})}, \quad |\epsilon|n^2\ll 1. \tag{38}$$

Let us prove that Eq. (35) cannot be solved for even n in view of a weak energy dependence of its right-hand side. We first assume that g tends to zero according to the same law as for the right-hand side of (35) (i.e., $g\propto|\ln|\epsilon||^{-1}$). In this case, the left-hand side is considerably smaller than the right-hand side: in the main approximation in $|\ln|\epsilon||^{-1}$, the operator $\tilde{J}$ acts on the logarithmic function of energy as on the constant ($\tilde{J}\{g\mathrm{e}^{i\alpha n}\}\simeq g\tilde{J}\{\mathrm{e}^{i\alpha n}\}$). If, however, we assume that g decreases with energy at a faster rate, the first term dominates on the left-hand side of (35), and we arrive at a contradiction again.

It can easily be seen that additional solutions of Eq. (33) with $\tilde{f}_n(0)=0$ do not exist. This is obvious for odd n since in this case we actually apply the τ-approximation for the operator $\tilde{J}$ acting on the function depending significantly on energy. For even n, we can write $\tilde{f}$ in the form $\tilde{f}_n=\varepsilon[1-q_n(\varepsilon)]$ and arrive at the right-hand side of the equation for the function q_n which has a weak dependence on ε similar to that in Eq. (35) for g_n.

In order to make relation (34) informative to the maximum possible extent, we construct from additional odd solutions of Eq. (33) their following linear combination, depending on the parameter α_0:

$$G_{\mathbf{p}}(\alpha_0)\equiv G(\varepsilon,\alpha-\alpha_0),$$

$$G(\varepsilon,\alpha)=\frac{1}{2\pi}\sum_{n=\pm 1,\pm 3,\ldots}^{\infty}[1-g_n(\varepsilon)]\mathrm{e}^{i\alpha n}. \tag{39}$$

This function is the odd component of $\delta(\alpha)$ for $\varepsilon=0$. The conserved quantity G corresponds to a local conservation of the number of particles in the angle α. It is obvious form the physical point of view that the absence of a similar even conserved quantity and the quantity $\varepsilon[1-q_n(\varepsilon)]\mathrm{e}^{i\alpha n}$ of the type of local energy density indicates complete isotropization of the even component in the distribution of the number of particles and their energies at the pretemperature relaxation stage, in other words, the independence of the quantities λ_s and κ in (34) of α. Substituting $\tilde{f}=G,1,\varepsilon$ into (34), we obtain

$$\lambda_a(\alpha)=m^{-1}\int f_{0\mathbf{p}'}G_{\mathbf{p}'}(\alpha)d^2p', \tag{40}$$

$$\lambda_s=(2\pi m)^{-1}\int f_{0\mathbf{p}}\,d^2p, \tag{41}$$

$$\kappa=(2\pi m)^{-1}\int \varepsilon f_{0\mathbf{p}}\,d^2p. \tag{42}$$

These formulas complete the solution of the problem on the final state of pretemperature relaxation. More rigorous mathematical substantiation of these formulas will be given in Appendix B. It can be seen from (40) that the function $G(\varepsilon_0,\alpha)$ is the final angular distribution over the FS for a nonequilibrium state which initially has the form[1)] $m\delta^a(\mathbf{p}-\mathbf{p}_0), \alpha_{\mathbf{p}_0}=0$. Substituting the solution of Eq. (35) in the zeroth order of iteration into (39), we obtain the explicit form of the function G:

$$G(\varepsilon,\alpha)\simeq\tau_{\rm en}(\varepsilon)\int \nu^{\alpha}_{\mathbf{p}'\mathbf{p}}\delta(\alpha')d^2p'$$

$$\simeq-\frac{2}{3}\,\frac{\epsilon}{|\ln|\epsilon||}\,\frac{\cos\alpha}{\sin^3\alpha},$$

$$|\pi-\alpha|,\alpha\gg|\epsilon|^{1/2}. \tag{43}$$

Obviously, $|G|\simeq|\epsilon|^{-1/2}$ in the region $\alpha,|\pi-\alpha|\leqslant\sqrt{|\epsilon|}$. However, G cannot be a monotonic function of the angle for $\varepsilon>0$. This statement follows from the fact that G in (43) is negative for $\varepsilon>0$ and from the ''normalization condition''

$$\int_{-\pi/2}^{\pi/2} G(\varepsilon,\alpha)d\alpha \cong \frac{1}{2}, \tag{44}$$

which indicates the conservation of the number of particles. The proof of relation (44) is given in Appendix B.

The odd distribution $\lambda_a(\alpha)$ sets in even after a few first collisions, i.e., approximately over the time $\tau_{\rm en}(\varepsilon_0)$. (The zeroth approximation of the iterative method for Eq. (35) indicates that the first collision is taken into account, while the next iterations take into account a small contribution to g_n from the next collisions.) It can be see from estimate (38) that isotropization of the even component (number density of particles and energy density) requires approximately $\ln(\varepsilon_F/\varepsilon_0)$ collisions.

It should be borne in mind, however, that the pretemperature stage of relaxation considered above under the assumption $T\to 0$ terminates when the average excitation energy becomes equal to the electron temperature T, i.e., after the thermalization period $\tau_{\rm en}(T)$ [see (2)] (approximately after $\ln(\varepsilon_0/T)$ collisions). Thus, isotropization of the even distribution is not completed at the pretemperature stage if $\varepsilon_0^2 \leqslant T\varepsilon_F$.

The electron temperature T stabilized as a result of thermolization is determined by the initial temperature T_0 of equilibrium electrons (the temperature of the thermostat) as well as the energy of the electron beam:

$$T=\sqrt{T_0^2+6\pi^{-2}\kappa}. \tag{45}$$

This relation follows from the equality of the energy of the electron system at the pretemperature and temperature stages of relaxation.

Let us consider the applicability of the linear approximation for collision integral in the given case. It can easily be seen that the most stringent constraint boils down to the requirement that any region of the **p**-space with a radius of the order of ε/v_F must contain a much smaller number of excitations than the maximum possible number $h^{-2}(\varepsilon/v_F)^2$. In this case, the linear approximation is applicable for small-angle collisions with the momentum transfer $\sim\varepsilon/v_F$ also. This requirement is satisfied for

$$\kappa \ll \varepsilon_0^3\varepsilon_F^{-1}. \tag{46}$$

In the opposite limit, the pretemperature stage is transformed into the linear mode of relaxation of the FS deformed in the odd way (for $T_0 \leqslant T$), which was mentioned at the end of the previous section. It should be noted, however, that even when the inequality (46) is observed, the linear approximation is inevitably violated upon a transition from the pretemperature stage to the temperature stage (for $T_0 \leqslant T$), but is restored again for the time $t \gg \tau_{\rm en}(T)$.

CONCLUSIONS

Qualitative ideas concerning the relaxation of strongly anisotropic distributions are outlined in Refs. 4, 5, and 14 and can be formulated as follows. After several electron–electron collisions, i.e., after a time $t > \tau_{\rm en}(\varepsilon_0)$, the electron distribution component which is even in the momentum becomes isotropic, while the odd component, which acquires an angular size of the order of $\sqrt{(\varepsilon_0+T)/\varepsilon_F}$, is broadened slowly with time (ε_0 is the excitation energy for electrons in the beam). This long-lived nonequilibrium state, which is odd in the **p**-space, has the form of electron and hole beams moving in opposite directions in the coordinate space. Since an electron is converted into a hole (and vice versa) over a time of the order of $\tau_{\rm en}$, the evolution of this electron–hole formation is realized though one-dimensional diffusion over a time much longer than $\tau_{\rm en}$.[5,14] (It is interesting to note that as a result the current state is blurred symmetrically both in the direction of injection of the primary beam, and in the opposite direction.) Under certain conditions, the secondary electron and hole beams can be observed directly in experiments with narrow primary electron beams injected into a $2D$EG. In the electrical conduction of $2D$EG wires (see Sec. 2) whose width $d \ll \sqrt{l_a l_s}$, $d \gg l_s\sqrt{T/\varepsilon_F}$, such electron–hole formations make the main contribution to current. The transport mean free path of a charge carrier moving at an angle $\alpha \ll 1$ to the boundary can be written as the one-dimensional diffusion length of the charge carrier:

$$[l_{\rm tr}]^{-1}=\frac{I_s(\alpha)\alpha^2}{d^2}+\sum_l\, l_{ai}^{-1}(\alpha).$$

The second term takes into account all processes of odd relaxation (due to electron–electron collisions, impurities, and phonons), which take a carrier out of the angular interval of the order of α.

The consistent analysis of the problem carried out in this paper allowed us to substantiate the qualitative conclusion drawn in Refs. 13 and 14. In addition, we refined the dependences of the mean free paths $l_a(\varphi_0)$ and $l_s(\varphi_0)$ on the parameters of the problem [see formulas (25)–(29)]. Among other things, we have established a slow dependence of the even relaxation rate on the anisotropy scale of the distribution, which allowed us to determine the lower boundary of the layer width d for the region of inverse temperature dependence of resistivity: $d \geqslant l_s\sqrt{T/\varepsilon_F}$. Although the length $l_a(\varphi_0)$ is noticeably larger than $l_s(\varphi_0)$ in the region $T/\varepsilon_F \leqslant d/l_s \ll \sqrt{T/\varepsilon_F}$, the slow increase in resistivity with temperature typical of the Fuchs situation in the electrical conduction of plates dominates.

The analysis of the evolution of high-energy strongly anisotropic distribution disregarding spatial anisotropy effects (Sec. 3) can nevertheless be used for studying the results of experiments with narrow electron beams in $2D$EG. It should be noted that, in the case of a strong dependence of the distribution function on the coordinate **r**, the momentum distribution $f_{\mathbf{p}}=\int f_{\mathbf{p}}(\mathbf{r})d^2r$ of particles in the linear approximation satisfies Eq. (30). In other words, all the results concerning the evolution of the initial distribution $f_{0\mathbf{p}}$ are valid for a beam injected in pulses into the momentum space at the initial instant (naturally, until the electron beam collides with the boundaries). The application of a magnetic field leads only to a rotation of the distribution in the momentum space: $f(\alpha,\varepsilon,t)\to f(\alpha-\Omega t,\varepsilon,t),\Omega$ being the cyclotron frequency (although even a very weak magnetic field leads to a strong beam broadening in the coordinate space).[4,14] It was mentioned at the end of the previous section that the process of thermalization is nonlinear for $T_0 \leqslant T$, and hence the mo-

mentum distribution in the given case cannot be determined without taking spatial dispersion effects into consideration for $t \geqslant \tau_{en}(T)$. It can be stated, however, that these effects do not hamper the stabilization of local temperature in the coordinate space since the thermalization length $l_{en}(T) = v_F \tau_{en}(T)$ for such a time has the minimum value among all characteristic lengths. According to (45), the condition $T - T_0 \ll T_0$, which allows us to avoid the solution of the problem on the coordinate dependence of temperature, is reduced to the inequality $\sqrt{\kappa_T} \ll T_0$. Here κ_T is the value of the quantity $\varkappa$ (which is proportional to the energy density of the beam) at the instant $t = \tau_{en}(T)$ [the substitution $\kappa \to \kappa_T$ must also be made in condition (46)]. The realization of these linearity conditions is facilitated by the fact that the initial energy density of the injected beam decreases significantly during its propagation.

The relation between the momentum distribution of particles $f_{\mathbf{p}} \equiv f(\alpha, \varepsilon)$ and their distribution in the coordinate space in the problem on electron beams is determined from the following simple physical considerations. The displacement of a particle in the direction perpendicular to the velocity of the primary beam is

$$r_\perp \simeq v_F \int_0^t \alpha(t')dt' \simeq \alpha(t) t v_F,$$

where $\alpha(t) \ll 1$ is the angle of deviation of the particle from the initial direction of the beam. For this reason, the distribution of particles over $r_\perp$ repeats qualitatively the function $\int f(\alpha, \varepsilon, t)d\varepsilon$, where $\alpha \simeq r_\perp / v_F t$, and t stands for the time of motion of the particle between the emitter and the detector. If the separation between the emitter and the detector is $L \leqslant l_{en}(\varepsilon_0)$, a binary collision is hardly probable, and the momentum distribution is defined by (31) with $t \simeq L/v_F$ (both for electrons deviating through small angles from the primary beam and for particles moving in the opposite direction). If $L > l_{en}(\varepsilon_0)$ [but $L < l_{en}(t)$], a particle participates in several collisions before hitting the detector, and the angular distribution is described by the function $G(\varepsilon_0, \alpha)$ (see Sec. 3). This function is positive and has a peak in the angular range $\alpha \leqslant \sqrt{\epsilon_0}$. Then it changes sign, and $G \propto -\alpha^{-3}$ for $\alpha \gg \sqrt{\epsilon_0}$. Negative values of G correspond to holes, i.e., the detection of a positive charge. The estimate $t \simeq L/v_F$ is preserved in this case also since each next mean free path is approximately six times longer than the previous one due to a decrease in energy during a collision, i.e., the total mean free path is approximately equal to the last mean free path. If, however, $L \gg l_{en}(T)$, the main broadening of the electron and hole beams occurs at the temperature stage of relaxation. It follows from the results of Sec. 2 that the angular width of beams $\varphi_0 \simeq \sqrt{T/\varepsilon_F}(t/\tau_s)^{1/4}$, and the quantity t is the time of diffuse traversing the path $L(t \simeq L^2/v_F l_s)$ by the electron–hole formation. It should be borne in mind, however, that after its first collision, a hole returns to the emitter region with a probability approximately equal to 1/2, and $t \simeq l_{en}/v_F$ for such holes for any $L \gg l_{en}(\varepsilon)$.

In order to observe these effects associated with the beams, the emitter must create a narrow primary beam, while the detector must transmit particles in a much wider angular range as compared to the emitter. Indeed, electrons reach the detector separated by the distance $r_\perp$ from the primary beam with the mean angle of deviation from the primary beam $\sim r_\perp / t v_F$, and hence a detector having the same angular width as in the emitter would not detect most of electrons.

It was mentioned above that a new nonlinear mode of electron–electron relaxation in $2D$ degenerate systems exists, in which odd deformations of the Fermi surface with respect to the angle relax at a much slower rate than even deformations. The ranges of existence of the nonlinear mode have been determined in problems on electrical conduction of $2D$EG wires and on the evolution of electron beams injected into a two-dimensional gas. It should be noted that under the nonlinear conditions, the inclusion of spatial dispersion forms an independent problem which must be solved separately.

This research was carried out under partial financial support of the International Science Foundation and the Ukrainian Government (Grant No. U2D200).

APPENDIX

A. Approximation of frequent collisions

Under the conditions when the characteristic length and time intervals for a system exceed significantly the quantities $l_{en}(T) \simeq l_s$ and τ_s, the electron system is close to energy equilibrium. In the problem on static electrical conductivity of a $2D$EG wire, these conditions are reduced to the inequality $x = l_s(\alpha)\alpha/d \ll 1$. The presence of new effects associated with the two- dimensional nature of the system is associated with the inequality

$$y = l_s(\alpha)/l_\alpha(\alpha) \ll 1.$$

Successive expansion of the kinetic equation (19) in these small parameters leads to the following chain of equations:

$$I_0\{\chi^{00}\} = 0, \tag{A1}$$

$$v_z \frac{\partial \chi^{00}}{\partial z} + I_0\{\chi^{10}\} = 0, \tag{A2}$$

$$I_1\{\chi^{00}\} + I_0\{\chi^{01}\} = 0, \tag{A3}$$

$$v_z \frac{\partial \chi^{10}}{\partial z} + I_0\{\chi^{02} + \chi^{20}\} + I_1\{\chi^{01}\} + I_2\{\chi^{00}\} = e\mathbf{E}\mathbf{v}.$$

$$I = I_0 + I_1 + I_2,$$

$$\chi = \chi^{00} + \chi^{01} + \chi^{10} + \chi^{02} + \chi^{20} + \dots . \tag{A4}$$

The operator I_0 describing energy relaxation differs from I in that both operators $\hat{D}_{\mathbf{q}}^-$ in (5) are replaced by the operator $\hat{D}_{\mathbf{q}\varepsilon}$:

$$\hat{D}_{\mathbf{q}\varepsilon}\chi(\varepsilon_{\mathbf{p}}, \alpha_{\mathbf{p}}) = \frac{1}{2}[\chi(\varepsilon_{\mathbf{p}+\mathbf{q}/2}, \alpha_{\mathbf{p}+\mathbf{q}/2}) - \chi(\varepsilon_{\mathbf{p}-\mathbf{q}/2}, \alpha_{\mathbf{p}+\mathbf{q}/2}) + \chi(\varepsilon_{\mathbf{p}+\mathbf{q}/2}, \alpha_{\mathbf{p}-\mathbf{q}/2}) - \chi(\varepsilon_{\mathbf{p}-\mathbf{q}/2}, \alpha_{\mathbf{p}-\mathbf{q}/2})].$$

Here ε and α are the energy and angular variables defining the position of the point $\mathbf{p}$. The operator I_1 is equal to the

sum of two operators each of which was obtained by the replacements of one of the operators $\hat{D}_{\mathbf{q}}^{-}$ in (5) by $\hat{D}_{\mathbf{q}\varepsilon}$, and the other by $\hat{D}_{\mathbf{q}\alpha}$:

$$\hat{D}_{\mathbf{q}\alpha}\chi(\varepsilon_{\mathbf{p}},\alpha_{\mathbf{p}})=\frac{1}{2}\,[\chi(\varepsilon_{\mathbf{p}+\mathbf{q}/2},\alpha_{\mathbf{p}+\mathbf{q}/2}) - \chi(\varepsilon_{\mathbf{p}+\mathbf{q}/2},\alpha_{\mathbf{p}-\mathbf{q}/2}) + \chi(\varepsilon_{\mathbf{p}-\mathbf{q}/2}),\alpha_{\mathbf{p}+\mathbf{q}/2}) - \chi(\varepsilon_{\mathbf{p}-\mathbf{q}/2},\alpha_{\mathbf{p}-\mathbf{q}/2})].$$

Both operators $\hat{D}_{\mathbf{q}}^{-}$ in I_2 are replaced by $\hat{D}_{\mathbf{q}\alpha}$. The first superscript on χ indicates the order of the expansion in the parameter x, while the second shows the order of the expansion in the parameter $\sqrt{y}$. Equation (A4) is the sum of equations of the orders (02) and (20).

Let us first prove that an arbitrary odd function of α, i.e., $\chi^{00}=\overline{\chi(\alpha)}$ independent of ε, is the only solution of Eq. (A1) (apart from the solution $\chi^{00}\sim\varepsilon$ corresponding to a small variation of temperature). Equation (A1) is equivalent to the detailed balancing condition

$$\hat{D}_{\mathbf{k}}^{\pm}\hat{D}_{\mathbf{q}\varepsilon}\chi_{s,\alpha}^{00}=0.$$

This is obvious if we note that relation (A1) leads to $F_{\varepsilon}\{\chi^{00},\chi^{00}\}=0$, where the functional F_s can be obtained from F by replacing the operators $\hat{D}_{\mathbf{q}}^{-}$ in (4) by $\hat{D}_{\mathbf{q}\varepsilon}$ and taking into account the fact that the kernel of Φ is essentially positive. Naturally, the function $\overline{\chi(\alpha)}$ is a solution of this equation since $\hat{D}_{\mathbf{q}\varepsilon}\overline{\chi}=0$. In order to prove the uniqueness of this solution, it is sufficient to consider infinitely small $\mathbf{q}$:

$$\frac{\partial}{\partial\varepsilon}\,\chi_{s,a}^{00}(\mathbf{p}+\mathbf{k}/2)=\pm\frac{\partial}{\partial\varepsilon}\,\chi_{s,a}^{00}\,(\mathbf{p}-\mathbf{k}/2),$$

which gives $\chi_a^{00}=\overline{\chi(\alpha)}$ and $\chi_s^{00}\sim\varepsilon$.

The field term $e\mathbf{E}\cdot\mathbf{v}$ is not included in Eq. (A1): the identity $\int\partial n/\partial\varepsilon I_0\{\chi_a\}d\varepsilon=0$ shows that otherwise the equation would be unsolvable. It cannot be included either in Eqs. (A2) (since its left-hand side is even in $\mathbf{p}$) and (A3) (in view of the identity $\int\partial n/\partial\varepsilon I_1\{\widetilde{\chi}\}d\varepsilon=0$). Thus, Eqs. (A2) and (A3) allow us in principle to express the corrections χ^{10} and χ^{01} in terms of $\overline{\chi}$, while the function $\overline{\chi(\alpha)}$ can be determined from the condition of solvability of Eq. (A4):

$$\int d\varepsilon\left[v_z\frac{\partial\chi^{10}}{\partial z}+I_1\{\chi^{01}\}+I_2\{\overline{\chi}\}-e\mathbf{E}\cdot\mathbf{v}\right]\frac{\partial n}{\partial\varepsilon}=0. \quad \text{(A5)}$$

This algorithm supplements the approach developed in Secs. 1.1 and 2: Eq. (A3) defines the energy-dependent correction χ^{01} which is essential for estimating the rate of odd angular relaxation.

While analyzing Eq. (A3), it is convenient to separate three domains of integration with respect to the angle η as in the derivation of expressions (25) and (26). For $\varphi_0\gg\sqrt{T/\varepsilon_F}$, region II makes the main contribution in $\ln(\varphi_0^2\varepsilon_F/T)$ to the quantity $I_1\{\chi^{00}\}$. However, the contribution form region II containing $\ln(\varphi_0^2\varepsilon_F/T)$ to $I_0\{\chi^{01}\}$ competes with the contributions from regions I and III containing $\ln\varphi_0^{-1}$. If we retain in (A3) only the contribution from region II and take into account the possibility of expansion of the function $\overline{\chi}$ in $\mathbf{q}$ and $\mathbf{k}$ in region II, we obtain from (A3)

$$\int\hat{D}_{\mathbf{k}}^{-}\hat{D}_{\mathbf{q}\varepsilon}\Phi_{\mathbf{pkq}}\left(p_F^2\varphi\eta\frac{d^2\chi}{d\alpha^2}+2\hat{D}_{\mathbf{q}\varepsilon}\chi^{01}\right)d^2kd^2q=0,$$

where the integration with respect to $\mathbf{k}$ is carried out within region II. This equation can be solved by equating to zero the expression in the parenthesis in the integrand. In this case, χ^{01} has the form of the ''local drift'' (8). Thus, for not too anisotropic distributions $[\ln(\varphi_0^2\varepsilon_F/T)>2\ln(\varphi_0^{-1})]$, the ''local drift'' describes successfully the energy dependence of the nonequilibrium correction; for more anisotropic contributions, the energy dependence is weaker than (8).

It can be seen from Eqs. (A1)–(A3) that a significant energy dependence of the quantity $I\{\chi^{00}+\chi^{10}+\chi^{01}\}$ can appear only in second-order terms in the small parameters x and y. This proves relation (10): the correction ψ^{01} can be neglected to within the required accuracy.

B. The limit $t\to\infty$ for pretemperature relaxation stage

In order to control the limiting transition $t\to\infty$, we introduce the term $-\nu f$ describing infinitely slow damping of states for $\nu\to0$ into the right-hand side of Eq. (30). Let the nonequilibrium correction at the time instant $t=0$ have the form $f(0)=\delta(\mathbf{p}-\mathbf{p}_0)$. Integrating the kinetic equation with respect to t from 0 to ∞, we obtain

$$(-\hat{J}_{\mathbf{p}}+\nu)\Gamma_{\nu}(\mathbf{p}_0,\mathbf{p})=\nu\delta(\mathbf{p}-\mathbf{p}_0),$$

$$\Gamma_{\nu}(\mathbf{p}_0,\mathbf{p})=\nu\int_0^{\infty}f_{\mathbf{p}}(t)dt. \quad \text{(A6)}$$

The quantity $\Gamma_{\nu}(\mathbf{p}_0,\mathbf{p})$ is the value of nonequilibrium correction averaged over the damping time ν^{-1}. This quantity obviously tends to the required final distribution to which the initial nonequilibrium state evolves with time for $\nu\to0$. Let us consider the transposed equation

$$(-\tilde{\hat{J}}_{\mathbf{p}}+\nu)\Gamma_{\nu}(\mathbf{p}_1,\mathbf{p})=\nu\delta(\mathbf{p}-\mathbf{p}_1). \quad \text{(A7)}$$

Multiplying both sides of (A6) by $\Gamma_{\nu}(\mathbf{p}_1,\mathbf{p})$ and integrating with respect to $\mathbf{p}$, we obtain, according to (A7),

$$\Gamma_{\nu}(\mathbf{p}_1,\mathbf{p}_0)=\Gamma_{\nu}(\mathbf{p}_0,\mathbf{p}_1). \quad \text{(A8)}$$

Being the final state of pretemperature relaxation, the function $\Gamma_{\nu}(\mathbf{p}_0,\mathbf{p})$ for $\nu\to0$ is obviously not small only near the Fermi surface for $\varepsilon_{\mathbf{p}}\to0$. (It can easily be seen from (A7) and (A8) that Γ_{ν} is not small in the range of momenta where $\nu_{\mathrm{en}}(\varepsilon_{\mathbf{p}})\simeq\nu$.) It is convenient to introduce the angular density of particles on the FS, i.e.,

$$G_{\nu}(\mathbf{p}_0,\alpha)=m\int\Gamma_{\nu}(\mathbf{p}_0,\mathbf{p})d\varepsilon$$

and, accordingly, the angular energy density

$$\varepsilon_{\nu}(\mathbf{p}_0,\alpha)=m\int\varepsilon\Gamma_{\nu}(\mathbf{p}_0,\mathbf{p})d\varepsilon.$$

It follows from (A7) and (A8) that the function $G_{\nu}(\mathbf{p},\alpha_0)$ satisfies the equation

$$(-\tilde{\hat{J}}+\nu)G_\nu(\mathbf{p},\alpha_0)=\nu\delta(\alpha-\alpha_0).$$

If we write the solution of this equation in the form

$$G_\nu(\mathbf{p},\alpha_0)=\frac{1}{2\pi}\sum_n\,[1-g_{\nu n}(\varepsilon)]\mathrm{e}^{in(\alpha-\alpha_0)},$$

it becomes obvious that for $\nu\to 0$, the odd component of the function $G_\nu(\mathbf{p},\alpha_0)$ tends to the function $G_{\mathbf{p}}(\alpha_0)$ which is defined by expression (39). Indeed, equations for the quantities $g_{\nu n}(\varepsilon)$ differ from (35) only in that $\nu_{\mathrm{en}}(\varepsilon)$ is supplemented with an infinitely small term ν. For odd n, these equations can be solved by the iterative method (see Sec. 3). The absence of even solutions of Eq. (33) (apart from a constant) indicates that the even component of $G_\nu(\mathbf{p},\alpha_0)$ attains the constant value for $\nu_{\mathrm{en}}(\varepsilon)\gg\nu$. Thus, the assumption made in Sec. 3 concerning the relation between the absence of additional even solutions of (33) and the isotropy of the even component of the final distribution is proved. Similar arguments applied to the energy density $\mathscr{E}_\nu(\mathbf{p},\alpha_0)$ prove that the quantity κ in (32) is constant. The value of the functions $G_{\nu s}(\mathbf{p},\alpha_0)$ and $\mathscr{E}_\nu(\mathbf{p},\alpha)$ for $\nu_{\mathrm{en}}(\varepsilon)\gg\nu$ can be easily obtained from the conservation laws: multiplying (A6) by a constant and ε and integrating with respect to $\mathbf{p}$, we obtain $G_{\nu s}(\mathbf{p},\alpha_0)=(2\pi)^{-1}$ and $\mathscr{E}_\nu(\mathbf{p},\alpha_0)=(2\pi)^{-1}\varepsilon$. In order to obtain the ''normalization condition'' (44), we must multiply both sides of (A6) by $\mathrm{sgn}(\mathbf{p}\cdot\mathbf{p}_0)$, integrate with respect to $\mathbf{p}$, and use expression (16).

*E-mail: gurzhi@ilt.kharkov.ua

[1)]Negative values of the initial nonequilibrium state for $\varepsilon>0$ and positive values for $\varepsilon<0$ are naturally deprived of any physical meaning for $T\to 0$. However, it is convenient to analyze the evolution of the initial electron ($\varepsilon>0$) or hole ($\varepsilon<0$) nonequilibrium state by dividing them into components having different parities.

[1] A. V. Chaplik, Zh. Éksp. Teor. Fiz. **60**, 1845 (1971) [Sov. Phys. JETP **33**, 997 (1971)].
[2] C. Hodges, H. Smith, and J. W. Wilkins, Phys. Rev. **B4**, 302 (1971).
[3] R. N. Gurzhi, A. I. Kopeliovich, and S. B. Rutkevich, Zh. Éksp. Teor. Fiz. **83**, 290 (1982) [Sov. Phys. JETP **56**, 159 (1982)]; Adv. Phys. **36**, 221 (1987).
[4] R. N. Gurzhi, A. N. Kalinenko, and A. I. Kopeliovich, Fiz. Nizk. Temp. **19**, 1046 (1993) [Low Temp. Phys. **19**, 44 (1993)].
[5] R. N. Gurzhi, A. N. Kalinenko, and A. I. Kopeliovich, Phys. Rev. **B52**, 4744 (1995).
[6] R. N. Gurzhi, A. N. Kalinenko, and A. I. Kopeliovich, Phys. Low Dim. Struct. **2**, 75 (1994).
[7] L. W. Molenkamp and M. J. M. de Jong, Phys. Rev. **B49**, 5038 (1994); Phys. Rev. **B51**, 13389 (1995).
[8] L. W. Molenkamp, M. J. P. Brugmans, H. Van Houten, and C. T. Foxon, Semicond. Sci. Technol. **7**, B228 (1992).
[9] J. Heremans, B. K. Fuller, C. M. Thrush, and D. L. Partin, Phys. Rev. **B52**, 5767 (1995).
[10] R. N. Gurzhi, A. I. Kopeliovich, and T. Paszkiewicz, in *Die Kunst of Phonons* (ed. by T. Pazkievich), Plenum Press, New York (1994).
[11] L. G. Challis, in *Low-Dimensional Semiconductor Structures* (ed. by J. Butcher), Plenum Press, New York (1992).
[12] B. Laikhtman, Phys. Rev. **B45**, 1259 (1992).
[13] R. N. Gurzhi, A. N. Kalinenko, and A. I. Kopeliovich, Fiz. Nizk. Temp. **21**, 114 (1995) [Low Temp. Phys. **21**, 87 (1995)].
[14] R. N. Gurzhi, A. N. Kalinenko, and A. I. Kopeliovich, Phys. Rev. Lett. **72**, 3872 (1995).
[15] R. N. Gurzhi, Usp. Fiz. Nauk **94**, 689 (1968) [Sov. Phys. Uspekhi **11**, 255 (1968)].

Translated by R. S. Wadhwa

The high-frequency conductivity tensor of a two-dimensional electron gas with electron impurity states in a magnetic field

N. V. Gleizer and A. M. Ermolaev

Kharkov State University, 310077 Kharkov, Ukraine
(Submitted May 13, 1996; revised June 18, 1996)
Fiz. Nizk. Temp. **23**, 73–78 (January 1997)

The Lifshits method of local perturbations is used for studying the properties of a two-dimensional electron-impurity system in a quantizing magnetic field perpendicular to the plane of electron motion. The high-frequency conductivity tensor of the system is calculated in the model of independent point impurity atoms, taking into account the impurity states of electrons and spatial dispersion. The dissipative component of the conductivity has narrow resonant peaks at frequencies of electron transitions between the Landau levels and local levels, which are induced by the magnetic field. In zero magnetic field, these peaks merge into one broad peak lying above the threshold frequency for electron transitions from a local level to the energy band. Numerical values of the peak heights are obtained for semiconducting structures with a two-dimensional electron gas. © *1997 American Institute of Physics.*
[S1063-777X(97)00801-3]

The problem of impurity states of quasiparticles in solids was formulated for the first time in classical works by I. M. Lifshits on the theory of vibrations of nonideal crystal lattices, which appeared more than fifty years ago.[1] Lifshits formulated and solved the problem on the effect of impurity atoms on the phonon spectrum, predicted local vibrations, and developed a computational method for physical parameters of systems perturbed by impurity atoms. This theory was used later for studying impurity states of other quasiparticles (electrons and magnons).[2,3] A natural continuation of studies in this direction was the analysis of electron impurity states in a magnetic field.

The problem on electron impurity states in conductors in the presence of a magnetic field has specific features. As a matter of fact, the bound state of an electron in the field of an impurity in a three-dimensional conductor appears only when the potential well in which the electron falls is deep enough for the uncertainty in the particle energy in the well to be much smaller than the well depth.[4] If this condition is not observed, electrons experience only potential scattering by the impurity center accompanied by an insignificant phase shift, and the bound state is not formed. The motion of an electron in a magnetic field is bounded in two directions, and the electron drifts along the field. In a strong magnetic field, the situation resembles the one-dimensional case, when the bounds state emerges in a well of any depth. Thus, the magnetic field localizes electrons at attracting impurities even when localization in zero field is impossible. In a magnetic field, specific local and quasi-local states of electrons appear due to joint action of attracting impurities and magnetic field. For this reason, such states are referred to as magnetoimpurity states.

The idea of magnetic localization of electrons at isolated impurity atoms was put forth by Skobov[5] and Bychkov.[6] Skobov derived an exact expression for the amplitude of electron scattering by a point center in a magnetic field and observed its resonant nature. Bychkov predicted a bound state of an electron split by an attracting impurity from the first Landau level. He observed that Landau quantization leads to the ''reproduction'' of bound state. A small-radius attracting impurity removes degeneracy in the position of the center of the Larmor ''orbit'' and splits a level from each Landau level. The levels split from the second, third, etc. Landau level get in the continuous spectral region and become quasi-local. Being in resonance with the Landau states, these levels acquire a finite width which is inversely proportional to the lifetime of an electron near the impurity. The influence of magnetoimpurity states on the de Haas–val Alphen effect in metals was considered in Refs. 7 and 8.

In a two-dimensional electron–impurity system, an infinitely weak attractive potential of the impurity atom leads to the formation of a bound state of the electron. The binding energy in such a state is exponentially small as compared to the depth of the impurity potential well.[4] Bound states correspond to the poles of the amplitude of electron scattering by an impurity atom. These states are located on the physical sheet of the Riemann surface of scattering amplitude as a function of the electron energy and affect significantly the low-temperature properties of two-dimensional systems. The presence of such poles indicates that it is impossible to calculate the kinetic parameters of a two-dimensional electron–impurity system on the basis of the perturbation theory in the scattering potential. The exact expression of scattering amplitude is required.

The amplitude of scattering of quasiparticles by impurity centers in a solid is usually calculated by using the method of local perturbations developed by Lifshits.[1,2] or the method of zero-radius potentials.[9,10] The exact expression obtained by these methods for the amplitude of electron scattering by short-range impurity atoms was used recently for calculating the static conductivity of a two-dimensional electron gas.[11] In this publication, peculiarities of static conductivity of two-dimensional metals, heterojunctions, and inversion layers at the semiconductor boundary (which cannot be obtained on the basis of the perturbation theory) were predicted.

A system of local layers alternating with the Landau

levels is formed in a quantizing magnetic field perpendicular to the plane of electron motion. In the case of short-range impurity potentials of various types, their positions were determined in Refs. 12, 13 by the method of local perturbations. Such a spectral structure of the electron–impurity system is manifested in optical experiments with a two-dimensional electron gas in a magnetic field.[14] In spite of the large number of publications devoted to study of high-frequency properties of such systems,[14–16] the calculations of high-frequency conductivity of two-dimensional electron gas taking into account the local states of electrons have not been carried out yet.

In this paper, we describe the results of calculations of the high-frequency conductivity tensor for a two-dimensional electron gas taking into consideration the impurity states of electrons in a magnetic field perpendicular to the plane of electron motion. We used the effective mass approximation. Point impurity atoms are assumed to be scarce and distributed at random. The pole structure of the amplitude of electron scattering by isolated impurity atoms in a quantizing magnetic field is taken into account. The frequency of the electromagnetic field is assumed to be higher than the electron collision frequency. The results presented here can be used in an analysis of the inversion layer at the boundary between a semiconductor and an insulator, as well as thin metallic films under the conditions when electrons fill only the lower energy level associated with size quantization. It should be noted that the application of the convenient model of point impurity centers in Ref. 11 and in this publication indicates that the impurity potential is regarded as strongly screened. Otherwise, the obtained results can be used only for a qualitative comparison with experimental data.

In order to calculate the dynamic conductivity tensor of a two-dimensional electron gas in a magnetic field, we shall use the Kubo formula which makes it possible to express the conductivity tensor in the form of the product of two one-electron Green's functions averaged over the configurations of impurity atoms.[17] Neglecting apical corrections and using the spectral representation of the average one-electron Green's function, we obtain a relation between the conductivity tensor and the spectral density ρ of this function. The latter, being a function of the electron energy ε, has δ-shaped peaks at local levels. In the approximation of isolated impurity atoms, the contribution of local levels to the spectral density of Green's function has the form

$$\delta\rho_s(\kappa,\varepsilon)=|v_0|n_i(\varepsilon-\varepsilon_{\kappa s})^{-2}\delta[1-v_0F_s(\varepsilon)]. \tag{1}$$

Here κ and s are the orbital and spin quantum numbers of an electron in a magnetic field, $\varepsilon_{\kappa s}$ the electron energy in the state $|\kappa s\rangle$, v_0 the constant characterizing the scattering potential intensity, n_i the number density of impurity atoms, and $F_s(\varepsilon)$ the function appearing in the Lifshits equation[1,2]

$$1-v_0F_s(\varepsilon)=0 \tag{2}$$

for local levels.

The contribution of local levels to the high-frequency conductivity tensor is connected with function (1) through the relation

$$\begin{aligned}\delta\sigma_{\alpha\beta}(\mathbf{q},\omega)=\frac{ie^2}{\omega S}\sum_{\kappa\kappa' s}\int_{-\infty}^{\infty}d\varepsilon V^{\alpha}_{\kappa\kappa'}(-\mathbf{q})\\ \times V^{\beta}_{\kappa'\kappa}(\mathbf{q})\delta\rho_s(\kappa',\varepsilon)[f(\varepsilon)-f(\varepsilon_{\kappa s})][(\varepsilon\\ -\varepsilon_{\kappa s}+\hbar\omega+io)^{-1}+(\varepsilon-\varepsilon_{\kappa s}-\hbar\omega\\ -io)^{-1}],\end{aligned} \tag{3}$$

where $\mathbf{q}$ and ω are the wave vector and frequency, e is the electron charge, f the Fermi function, $V_{\kappa\kappa'}(\mathbf{q})$ are the matrix elements of the operator

$$\mathbf{V}(\mathbf{q})=\frac{1}{2}\,(\mathbf{v}e^{i\mathbf{q}\cdot\mathbf{r}}+e^{i\mathbf{q}\cdot\mathbf{r}}\mathbf{v})$$

($\mathbf{r}$ is the radius vector and $\mathbf{v}$ the operator of electron velocity in a magnetic field) in the Landau basis, S is the sample area, $\alpha,\beta=x,y$, and electrons move in the plane $z=0$. Relation (3) is obtained by the method used in Ref. 18 in the three-dimensional case. In the approximation used by us, it must be supplemented with the well-known contribution[14–16,19] calculated without taking into account the impurity states of electrons.

Using formula (1) for the δ-function of a complex argument,[4] we obtain from (3)

$$\begin{aligned}\delta\sigma_{xx}(q,\omega)=\frac{ie^2\omega_c^2n_i}{2\pi l^2\omega}\sum_{knn' s}r_{ks}(\varepsilon_{n' s}-\varepsilon^l_{ks})^{-2}\\ \times[f(\varepsilon^l_{ks})-f(\varepsilon_{ns})][(\varepsilon^l_{ks}-\varepsilon_{ns}+\hbar\omega\\ +io)^{-1}\\ +(\varepsilon^l_{ks}-\varepsilon_{ns}-\hbar\omega-io)^{-1}]Q_{n'n}(q).\end{aligned} \tag{4}$$

Here ω_c is the cyclotron frequency, l the magnetic length, ε_{ns} and ε^l_{ks} are the positions of the nth Landau level and the kth local energy level of an electron with the spin component s,

$$r_{ks}=-\left[\left.\frac{dF_s(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=\varepsilon^l_{ks}}\right]^{-1}$$

is the residue of the amplitude of electron scattering by an impurity center relative to the pole ε^l_{ks};

$$Q_{n'n}(q)=\left[\frac{d}{dq}\,\varphi_{n'n}(q)\right]^2,$$

$$\varphi_{n'n}=(n!/n'!)^{1/2}\xi^{1/2(n'-n)}\,\exp\left(-\frac{\xi}{2}\right)L_n^{n'-n}(\xi);$$

$L_n^{n'-n}$ are generalized Legendre polynomials, $\xi=\hbar q^2/(2m\omega_c)$, m is the effective mass of the electron, and the wave vector $\mathbf{q}$ is parallel to the y-axis. The component $\delta\sigma_{yy}$ can be obtained from (4) by replacing $Q_{n'n}$ by $(n'-n)^2\varphi^2_{n'n}(q)/q^2$.

The high-frequency Hall conductivity $\delta\sigma_{yx}$ differs from (4) in that it contains, instead of $Q_{n'n}$, the factor

$$-\frac{i}{2q}\,(n'-n)\,\frac{d}{dq}\,[\varphi_{n'n}(q)]^2,$$

and the term $(\varepsilon_{ks}^{l}-\varepsilon_{ns}-\hbar\omega-io)^{-1}$ in the brackets appears with the minus sign. While deriving formula (4), we used only the fact of existence of local levels. Their characteristics ε_{ks}^{l} and r_{ks} are arbitrary. These parameters can be calculated on the basis of a definite model of scattering potential or from a comparison of the theory with the experiment.

If we disregard spatial dispersion ($q=0$) in formula (4), the latter assumes the form

$$\delta\sigma_{xx}(\omega)=\delta\sigma_{yy}(\omega)=\frac{ie^2\hbar^2 n_l}{4\pi m^2 l^4\omega}\sum_{kns} r_{ks} \times[f(\varepsilon_{ks}^{l})-f(\varepsilon_{ns})][(\varepsilon_{k}^{l}-\varepsilon_{n}+\hbar\omega+io)^{-1} +(\varepsilon_{k}^{l}-\varepsilon_{n}-\hbar\omega-io)^{-1}][n(\varepsilon_{n-1}-\varepsilon_{k}^{l})^{-2} +(n+1)(\varepsilon_{n+1}-\varepsilon_{k}^{l})^{-2}], \quad (5)$$

where ε_n and ε_k^l are the positions of the nth Landau level and the kth local level without spin splitting. The component $\delta\sigma_{xy}(\omega)$ can be obtained from (5) by multiplication by $-i$ and by the sign reversal in the second term in the brackets. Expressions (4) and (5) have resonant singularities at frequencies corresponding to electron transitions between the Landau levels and local levels, which are induced by an electromagnetic field. The resonant frequencies are given by

$$\omega_{kn}=|\varepsilon_k^l-\varepsilon_n|/\hbar. \quad (6)$$

The real component of (5) determining the absorption of electromagnetic field energy by electrons has peaks (which have the δ-shape in the given approximation) at resonant frequencies. The peaks are blurred when the finite width of the Landau levels and local levels is taken into account. Figure 1 shows the real (curve *1*) and imaginary (curve *2*) parts of the dimensionless circular component of conductivity

$$\delta C=\delta\sigma_+\frac{\pi\hbar^3\omega_1^2(1+\omega_1/\omega_c)^2}{2e^2 r n_i}. \quad (7)$$

as a function of $x=\omega/\omega_1-1$ in the vicinity of the frequency $\omega_1=|v_0|/(2\pi\hbar l^2)$ corresponding to electron transitions between the Landau level $n=1$ and a local level split from it. Here $\delta\sigma_+=\delta\sigma_{xx}+i\delta\sigma_{yx}$. The calculations were made for the case of short-range attractive impurity potential ($v_0<0$) and for $\gamma/\omega_1=0.1$, where $\hbar\gamma$ is the total width of the energy levels participating in resonant transitions. The ratio of the maximum value of the real component $\delta\sigma_+$ and the static conductivity $e^2 n_e/(m\nu)$ (n_e is the number density of electrons and ν the collision frequency associated with scattering by impurities) is given by

$$A=4\frac{n_i}{n_e}\frac{\omega_1}{\omega_c}\frac{\nu}{\gamma}\left(1+\frac{\omega_1}{\omega_c}\right)^{-2}. \quad (8)$$

Using the values of the parameters $n_i/n_e=0.5$, $\omega_1/\omega_c=0.1$, and $\nu=\gamma$ typical of semimetals and semiconducting structures with a degenerate two-dimensional electron gas,[14] we obtain $A=17\%$ from (8). It should be noted that in the two-dimensional case, the resonant frequency ω_1 is proportional to the magnetic field strength H. In the three-dimensional conductor, the frequency is proportional to H^2.[7,8]

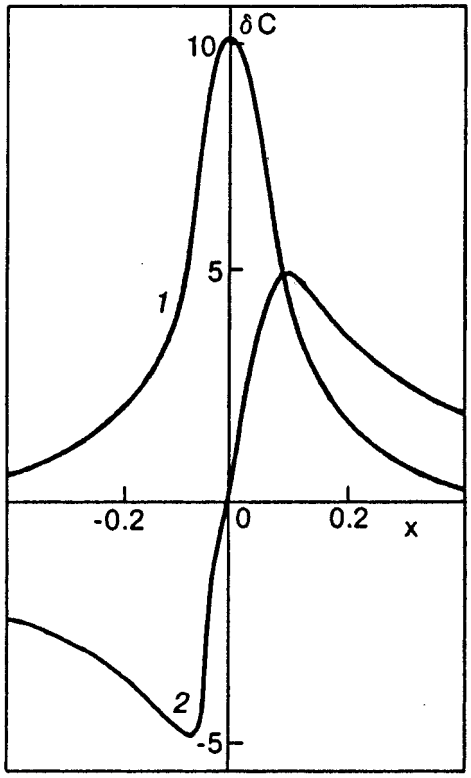

FIG. 1. Frequency dependence of the real (curve *1*) and imaginary (curve *2*) components of conductivity (7) near the resonance ($x=\omega/\omega_1-1$).

In zero magnetic field and in the case when short-range impurity atoms attract electrons, one local level ε_l splits from the bottom of the band. In this case, the sum over n in (5) can be replaced by an integral. As a result, $\delta\sigma_{xy}=0$, and formula (5) gives

$$\mathrm{Re}\ \delta\sigma(\omega)=\frac{e^2 r n_i}{\hbar^4\omega^3}\Theta(\varepsilon_l+\hbar\omega)(\varepsilon_l+\hbar\omega) \times[f(\varepsilon_l) -f(\varepsilon_l+\hbar\omega)]+(\omega\to-\omega), \quad (9)$$

$$\mathrm{Im}\ \delta\sigma(\omega)=\frac{e^2 r n_i}{\pi\hbar^4\omega^3}(\varepsilon_l+\hbar\omega)\ln\left|\frac{\varepsilon_F-\varepsilon_l-\hbar\omega}{\varepsilon_F-\varepsilon_l}\right| -(\omega\to-\omega), \quad (10)$$

where ε_F is the Fermi energy, Θ the Heaviside function, and $(\omega\to-\omega)$ indicates the term obtained from the previous term by reversing the sign of frequency. Formula (10) was derived at zero temperature ($T=0$). As expected, the real conductivity component (9) is an even function of frequency, while the imaginary component (10) is odd. It can be seen from (9) that the real conductivity component has a threshold at the frequency $\omega_g=|\varepsilon_l|/\hbar$. In the vicinity of the threshold, $\mathrm{Re}\ \delta\sigma\sim\omega-\omega_g$. It increases with frequency, attains its peak value, and then decreases according to the law ω^{-2}. For $T\to0$, the threshold frequency is shifted to the point $\omega_g+\varepsilon_F/\hbar$ in accordance with the Pauli exclusion principle. At this point, the imaginary conductivity component (10) has a logarithmic singularity.

In order to estimate the contribution (9), we approximate the density of states of two-dimensional electrons unperturbed by impurities by a rectangular step of width w and height $m/(2\pi\hbar^2)$. In this case, the function F appearing in Eq. (2) is given by

$$F(\varepsilon) = -\frac{m}{2\pi\hbar^2}\ln\left|\frac{w-\varepsilon}{\varepsilon}\right|.$$

In the case of attractive impurity potential and for $|v_0| \ll \hbar^2/m$, the solution of Eq. (2) has the form

$$\varepsilon_l = -w\ \exp\left(-\frac{2\pi\hbar^2}{m|v_0|}\right)$$

in accordance with the well-known result.[4] In this case, we have

$$r = 2\pi\hbar^2|\varepsilon_l|/m.$$

We use the following values of parameters: $m = 10^{-31}$ kg, $|\varepsilon_l| = 0.01$ eV, $\varepsilon_F/\hbar\omega_g = 0.5$, and $\hbar\omega_g/T = 10$, which are typical of semiconducting structures with a two-dimensional electron gas.[14] In this case, the ratio of the maximum value of (9) to the static conductivity is equal to $4.8(n_i/n_e) \times (\nu/\omega_g)$. If $n_i/n_e = 0.5$ and $\nu/\omega_g = 0.1$, this ratio is equal to 24%.

It was mentioned above that the impurity levels of a two-dimensional electron gas in a magnetic field are local. They split from degenerate Landau levels in the downward direction ($v_0 < 0$) or upwards ($v_0 > 0$) depending on the type of impurity potential. The electron energy spectrum can be studied taking into account their impurity states by analyzing high-frequency properties. The contribution of these states to the high-frequency conductivity calculated here leads to narrow resonant peaks of high-frequency parameters determined by the frequencies (6) of electron transitions between the Landau levels and local levels. In zero magnetic field, when the electron energy spectrum contains only one local level at the bottom of the band, these peaks merge into a single broad peak. It lies above the threshold frequency of activation of the local level.

The results obtained here can be used for studying the absorption of high-frequency radiation incident at right angle to the inversion layer at the boundary between a semiconductor and an insulator. The absorption of radiation polarized in the plane of the layer is proportional to the real component of conductivity (5). The inversion n-layer at the boundary between Si and SiO_2 is studied most thoroughly.[14] The simplest model of such a system is a plane occupied by electrons and immersed in a three-dimensional insulator. The conductivity tensor of such a system in a magnetic field in the absence of impurity states of electrons was calculated in Ref. 19. It follows from the equations presented in this publication and formula (5) that the absorption coefficient as a function of frequency has peaks at the cyclotron frequency and at frequencies (6). The peculiarities of conductivity considerably affect the properties of plasmons and magnetoplasmons in inversion layers also. This will apparently lead to the deformation of the magnetoplasmon spectrum in the vicinity of resonant frequencies (6).

The authors are grateful to V. G. Peschanskii for helpful advice and to N. A. Ermolaeva for her assistance in numerical calculations.

[1] I. M. Lifshits, *Selected Works. Physics of Real Crystals and Disordered Systems* [in Russian], Nauka, Moscow (1987).
[2] I. M. Lifshits, S. A. Gredeskul, and L. A. Pastur, *Introduction to the Theory of Disordered Systems*, Wiley, New York, 1988.
[3] Yu. A. Izyumov and M. V. Medvedev, *The Theory of Magnetically Ordered Crystals with Impurities*, Consultants Bureau, NY, 1973.
[4] L. D. Landau and E. M. Lifshits, *Quantum Mechanics* [in Russian], Nauka, Moscow (1989).
[5] V. G. Skobov, Zh. Éksp. Teor. Fiz. **37**, 1467 (1959) [Sov. Phys. JETP **10**, 1039 (1959)].
[6] Yu. A. Bychkov, Zh. Éksp. Teor. Fiz. **39**, 689 (1960) [Sov. Phys. JETP **12**, 483 (1960)].
[7] A. M. Ermolaev and M. I. Kaganov, Pis'ma Zh. Éksp. Teor. Fiz. **6**, 984 (1967) [JETP Lett. **6**, 395 (1967)].
[8] A. M. Ermolaev, Zh. Eksp. Teor. Fiz. **54**, 1259 (1968) [Sov. Phys. JETP **27**, 673 (1968)].
[9] Yu. N. Demkov and V. N. Ostrovskii, *Zero-Range Potentials and Their Applications in Atomic Physics*, Plenum, NY, 1988.
[10] S. Albeverio *et al.*, *Solvable Models in Quantum Mechanics*, Springer-Verlag, Germany (1988).
[11] V. A. Geiler, V. A. Margulis, and I. I. Chuchaev, Fiz. Tverd. Tela (St. Petersburg) **37**, 837 (1995) [Sov. Phys. Solid State **37**, 455 (1995)].
[12] A. M. Kosevich and L. V. Tanatarov, Fiz. Tverd. Tela (Leningrad) **6**, 3423 (1964) [Sov. Phys. Solid State **6**, 2738 (1964)].
[13] E. P. Bataka and A. M. Ermolaev, Izv. Vuzov, Ser. Fizika No. 1, 111 (1983).
[14] T. Ando, A. Fowler, and F. Stern, *Electronic Properties of Two-Dimensional Systems*, American Physical Society, New York (1985).
[15] V. M. Gokhfeld, M. I. Kaganov, and V. G. Peschanskii, Fiz. Nizk. Temp. **12**, 1173 (1986) [Sov. J. Low Temp. Phys. **12**, 661 (1986)].
[16] V. G. Peschanskii, H. Kheir Bek, and S. N. Savel'eva, Fiz. Nizk. Temp. **18**, 1012 (1992) [Sov. J. Low Temp. Phys. **18**, 711 (1992)].
[17] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskii, *Methods of Quantum Field Theory in Statistical Physics*, Prentice Hall, Englewood Cliffs, NJ, 1963.
[18] E. A. Kaner and A. M. Ermolaev, Zh. Éksp. Teor. Fiz. **92**, 2245 (1987) [Sov. Phys. JETP **65**, 1266 (1987)].
[19] K. W. Chiu and J. J. Quinn, Phys. Rev. **B9**, 4724 (1974).

Translated by R. S. Wadhwa

Principles of finite-dimensional perturbation theory

I. V. Krasovskiĭ

*B. Verkin Institute for Low Temperature Physics and Engineering, National Academy of Sciences of the Ukraine, 310164 Kharkov, Ukraine**

V. I. Peresada

Kharkov State University, 310077 Kharkov, Ukraine
(Submitted June 17, 1996; revised July 26, 1996)
Fiz. Nizk. Temp. **23**, 79–91 (January 1997)

The theory of finite-dimensional perturbations of self-adjoint operators, which is aimed at the solution of physical problems, is reviewed. Special attention is paid to the special kind of operators, which permits efficient application of the **J**-matrix technique. The spectral density of a periodic **J**-matrix is calculated. © *1997 American Institute of Physics.* [S1063-777X(97)00901-8]

1. INTRODUCTION

Many problems in the theory of elementary excitations in solids can be reduced to the following: self-adjoint operators H and H_0 such that $H=H_0+V$, where V is a finite-dimensional (local, degenerate) perturbation, are defined in a separable Hilbert space G. Finite dimensionality of the operator V indicates that there exists an orthonormal basis $\{e_i\}_{i=0}^{\infty}$ of the space G, such that for a certain number r the condition $Ve_i=0$ is satisfied for all $i\geqslant r$. The spectral parameters (the spectrum, eigenvectors, and spectral functions) of the operator H_0 are known. The theory of finite-dimensional perturbations aims at determining the spectral characteristics of the operator H as well as the quantities $\mathrm{Tr}\{f(H)-f(H_0)\}$ which are variations of thermodynamic functions associated with the perturbation V. Since the value of V is not necessarily small, the small-perturbation theory based on the power expansion in V is inapplicable here.

Under certain assumptions concerning the models, the above-mentioned physical problems include the determination of the influence of a point defect on the spectrum, eigenfunctions, and thermodynamics of the crystal[1–20 1)] as well as the problems on linear and planar defects in the crystal.[3] The fact that this class of problems includes the determination of thermodynamic functions of a solid solution[1,21,22] and the calculation of the distribution function for squares of vibrational frequency of a perfect lattice[24] is less obvious. Naturally, this list does not exhaust all possible applications of the finite-dimensional perturbation theory. The fundamentals of this theory were developed by I. M. Lifshits in his publications[25–29] where no assumptions concerning some special properties of the operators H and H_0 were made. It will be proved below that the further evolution of the theory has become possible on the basis of the method of Jacobi matrices (J-matrix technique)[11–13,19,20,24,30–35] taking into account the special form of the matrix of the operator H.

In this paper, we shall review the theory of finite-dimensional perturbations and acquaint the reader with some new results obtained in this field.

2. THE THEORY IN THE GENERAL CASE

This chapter is based on publications by I. M. Lifshits[25–29] and M. G. Krein.[36]

2.1. Operators with discrete spectra

At the beginning, we assume that the operator H_0 acting in the N-dimensional space (N is finite or $N=\infty$) has only a discrete spectrum $\{\lambda_i\}_{i=0}^{N-1}$ and the corresponding normalized eigenvectors $\{\psi_i\}_{i=0}^{N-1}$. Let us consider the equation for eigenvectors of the operator $H=H_0+V$:

$$(H_0+V-zI)\phi=0$$

or

$$\phi=-R(z)V\phi, \quad R(z)=(H_0-zI)^{-1}. \tag{1}$$

Since $Ve_i=0$, $i\geqslant r$, the condition $\det(I+RV)=0$ of solvability of system (1) has an especially simple form when the matrix $I+RV$ and the vector ϕ are presented in the basis $\{e_i\}_{i=0}^{N-1}$, namely,

$$\Delta(z)\equiv\det[I+R(z)V]=|\delta_{lj}+(R(z)Ve_j,e_l)|_0^{r-1}=0, \tag{2}$$

$$[R(z)Ve_j,e_l]=\sum_i\sum_{k=0}^{r-1}\frac{(e_k,\psi_i)(\psi_i,e_l)}{\lambda_i-z}V_{kj},$$

$$V_{kj}=(Ve_j,e_k). \tag{3}$$

The solutions z_q of Eq. (2) are eigenvalues of the operator H. It follows from Eq. (1), presented in the basis $\{e_i\}_{i=0}^{N-1}$, that the corresponding eigenvectors have the form

$$\phi_q=-\sum_i\sum_{j=0}^{r-1}a_j\sum_{k=0}^{r-1}\frac{(e_k,\psi_i)}{\lambda_i-z_q}V_{kj}\psi_i, \tag{4}$$

where a_j are the solutions of the system of equations

$$\sum_{j=0}^{r-1}[I+R(z_q)V]_{lj}a_j=0, \quad l=0,\ldots,r-1.$$

It should be noted that Eqs. (2) and (4) can be simplified if nonzero eigenvalues and the corresponding normalized eigenvectors $\{\eta_i\}_{i=0}^{r_0-1}$ of the operator V are known. Then we

can put $e_i=\eta_i (i=0,\ldots,r_0-1)$, and $V_{kj}=V_{kk}\delta_{kj}$ in this case. Further, the number r has the minimum value in this case (it is equal to the rank r_0 of the operator V).

If all the eigenvalues λ_i are simple, the meromorphic function $\Delta(z)$ can be decomposed in partial fractions, and Eq. (2) assumes the form

$$1+\sum_i \frac{c_i}{\lambda_i - z}=0. \tag{5}$$

It was shown by Lifshits[26] that the coefficients c_i are defined as follows:

$$c_i = V(i)+\sum_{i_0\neq i}\frac{V(i,i_0)}{\lambda_{i_0}-\lambda_i}+\ldots$$

$$+\frac{1}{(r-1)!}\sum_{i_0\neq i,\ldots,i_{r-2}\neq i}\frac{V(i,i_0,\ldots,i_{r-2})}{(\lambda_{i_0}-\lambda_i)\ldots(\lambda_{i_{r-2}}-\lambda_i)},$$

$$V(i_1,\ldots,i_n)=\begin{vmatrix} V_{i_1 i_1} & \cdots & V_{i_1 i_n}\\ \vdots & & \vdots \\ V_{i_n i_1} & \cdots & V_{i_n i_n}\end{vmatrix},$$

$$\text{where}\quad V_{ik}=(V\psi_k,\psi_i).$$

Equation (5) can be used for a qualitative analysis of the shifts in the energy levels under the action of a perturbation.[26] However, it is easier to use the Courant–Fischer theorem. The following statement is a simple corollary of this theorem[37,38]:

Lemma 1. Let

(a) the eigenvalues of the operator H_0 be ordered as follows: $\lambda_0\geqslant\lambda_1\geqslant\ldots\geqslant\lambda_{N-1}$;

(b) the rank of the matrix V is equal to r_0, the number of negative eigenvalues is n, and the number of positive eigenvalues is $p=r_0-n$;

(c) $z_0\geqslant z_1\geqslant\ldots\geqslant z_{N-1}$ are eigenvalues of the operator H.

Then the interval $(\lambda_{s+1},\lambda_s)$ contains not more than r_0 eigenvalues z_i, the interval $(-\infty,\lambda_{N-1})$ not more than n, and the interval (λ_0,∞) not more than p such eigenvalues. Also, $\lambda_{i-p}\geqslant z_i\geqslant\lambda_{i+n}$ (we formally put here $\lambda_j=\infty$ for $j<0$ and $\lambda_j=-\infty$ for $j>N-1$). For example, if a certain level λ_m is x-fold degenerate and $x>r_0$, λ_m is also an eigenvalue of the operator $H=H_0+V$ of multiplicity not lower than $x-r_0$.

Taking into account the continuity of the dependence of eigenvalues and eigenvectors of the operator H on eigenvalues of the operator H_0 and introducing arbitrarily small corrections to the eigenvalues λ_i, we can also use Eqs. (2) and (4) in the case when the eigenvalue z of the operator H coincides with λ_m for a certain m. However, it is expedient to derive an explicit formula for this case. We write the equation $(H_0-zI)\phi=-V\phi$ in the form

$$(H_0-zI)\phi'=-V\phi, \tag{6}$$

where $\phi=\phi'+\eta_m$, and η_m is the orthogonal projection of the vector ϕ onto the eigenspace G_m of the operator H_0 corresponding to the eigenvalue $z=\lambda_m$. Equation (6) is valid since, by definition, $(H_0-zI)\eta_m=0$. Thus, (H_0-zI) in Eq. (6) can be considered as the orthogonal adjunct to G_m, where the resolvent $R(z)=(H_0-zI)^{-1}$ is defined correctly. Using the same arguments for (6) as in the derivation of formulas (2) and (4) from (1), we obtain

$$\phi=\eta_m-\sum_{i\neq m}\sum_{j=0}^{r-1}a_j\sum_{k=0}^{r-1}\frac{(e_k,\psi_i)}{\lambda_i-\lambda_m}V_{kj}\psi_i, \tag{7}$$

where a_j are solutions of the system of equations

$$a_l+\sum_{j=0}^{r-1}a_j\sum_{i\neq m}\sum_{k=0}^{r-1}\frac{(e_k,\psi_i)(\psi_i,e_l)}{\lambda_i-\lambda_m}V_{kj}=(\eta_m,e_l),$$

$$l=0,\ldots,r-1. \tag{8}$$

Among other things, Lemma 1 states that, if the spectrum of the operator H is contained in a certain interval, this operator can have not more than r_0 eigenvalues outside this interval. The upper and lower boundaries of these eigenvalues can be obtained conveniently by using the Gerschgorin theorem (see, for example, Ref. 37) formulated for the case of a Hermitian matrix: each eigenvalue of the $N\times N$ matrix H lies at least in one of the intervals

$$|H_{ii}-z|\leqslant\sum_{k(k\neq i)}|H_{ik}|,\quad i=0,\ldots,N-1,$$

where H_{ik} are the matrix elements of H in an arbitrary basis, and summation is carried out over all the elements of the ith row, excluding H_{ii}.

2.2. Operators with discrete and continuous spectra

2.2.1. Limiting transition to continuous spectrum. Shift function

In many physical models of systems with a large number M of particles, a fraction of eigenvalues of the Hamiltonian are ''condensed'' in one or several intervals of continuous spectrum for $M\to\infty$. (The remaining ''isolated'' eigenvalues form a discrete spectrum.) Hence it is important to generalize the above formulas to the case when a fraction of the eigenvalues λ_i is ''condensed'' into a continuous spectrum. Following Lifshits,[27] we introduce a set of self-adjoint operators H_α possessing only a discrete spectrum and depending on the parameter α. Let $\lambda_p^{(\alpha)}$ be their eigenvalues $\lambda_0^{(\alpha)}\geqslant\lambda_1^{(\alpha)}\geqslant\ldots\geqslant\lambda_{N-1}^{(\alpha)}$, and $\psi_p^{(\alpha)}(x)$ the normalized eigenfunctions depending on the argument x. By definition, we assume that the following conditions are satisfied for a part of the spectrum of the operator H_α (this part is called the quasicontinuous spectrum).

(1) There exists a piecewise continuous function $\lambda(u)$ independent of α and having a piecewise continuous derivative, such that

$$\lambda_p^{(\alpha)}\equiv\lambda^{(\alpha)}(u_p)=\lambda(u_p)+O(\alpha),\quad u_p=p\alpha,$$

$$\lambda_p^{(\alpha)}-\lambda_{p-1}^{(\alpha)}=\alpha\left[\left.\frac{d\lambda}{du}\right|_{u=u_p}+O(\alpha)\right].$$

(2) There exists a set of functions $\chi_u(x)$ such that the following relations are observed in a certain range of x:

$$\chi_p^{(\alpha)}(x)=\chi_{u_p}(x)+O(\alpha), \quad \chi_p^{(\alpha)}=\psi_p^{(\alpha)}/\sqrt{\alpha},$$

$$\chi_p^{(\alpha)}-\chi_{p-1}^{(\alpha)}=\alpha\left[\left.\frac{\partial\chi_u}{\partial u}\right|_{u=u_p}+O(\alpha)\right].$$

In the limit $\alpha\to 0$, the quasi-continuous spectrum is transformed into a continuous spectrum. For systems of many particles, $\alpha\sim 1/M$.

In order to generalize the theory presented in Sec. 2.1 to the case of operators with a continuous spectrum, we apply it to the operator $H_\alpha=H_{0\alpha}+V$ and then proceed to the limit $\alpha\to 0$.

Since the perturbation V is finite-dimensional, it follows from Lemma 1 that the continuous spectrum of the limiting operator $H=H_0+V=H_{\alpha\to 0}$ coincides with the continuous spectrum of the unperturbed limiting operator $H_0=H_{0,\alpha\to 0}$. In this case, the ''action'' of the operator V can ''split'' not more than r_0 discrete eigenvalues from an interval of the continuous spectrum.

We can now write formulas (2)–(4) for H_α and proceed to the limit $\alpha\to 0$. This transition affects only the sums over i. For formulas (7) and (8), the limit is obvious; for example,

$$\sum_{i\neq m}\frac{(e_k,\psi_i)}{\lambda_i-\lambda_m}\psi_i$$

is replaced by

$$\int\frac{(e_k,\chi_u)}{\lambda(u)-\lambda_m}\chi_u du+\sum_{l\neq m}\frac{(e_k,\psi_i)}{\lambda_i-\lambda_m}\psi_i,$$

where the sum is taken over the discrete spectrum of the operator $H_{\alpha\to 0}$, while the integral takes over the continuous spectrum (in the sense of the Cauchy principal value if λ_m belongs to the continuous spectrum of $H_{\alpha\to 0}$). Similarly, if z belongs to the discrete spectrum of the limiting operator $H_{\alpha\to 0}$, formulas (3) and (4) are generalized: the sums over i are replaced by integrals over the continuous spectrum plus the sums over the discrete spectrum. In order to proceed to the limit in (3) and (4) in the case when z belongs to the continuous spectrum, we introduce the shift function $\xi(\lambda)$ on the continuous spectrum of the operator $H_{\alpha\to 0}$ according to the formulas[2)]

$$z_p^{(\alpha)}\equiv z^{(\alpha)}(u_p)=\lambda^{(\alpha)}(u_p)+\alpha\left.\frac{d\lambda}{du}\right|_{u=u_p}\xi^{(\alpha)}[\lambda(u_p)],$$

$$\xi^{(\alpha)}[\lambda(u_p)]=\xi[\lambda(u_p)]+O(\alpha), \tag{9}$$

under the condition that $d\lambda/du\neq 0$. Here $z_p^{(\alpha)}$ are eigenvalues of the operator $H_\alpha(z_0^{(\alpha)}\geqslant z_1^{(\alpha)}\geqslant\ldots\geqslant z_{N-1}^{(\alpha)})$.

According to Lifshits, the following relation holds for the sum over the quasi-continuous spectrum:

$$\sum_p\frac{\alpha f(u_p)}{\lambda_p-z_q}=\int'\frac{f(u)du}{\lambda(u)-\lambda(u_q)}-\left.\frac{\pi f(u_q)}{\frac{d\lambda}{du}}\right|_{u=u_q}\cot\,\pi\xi[\lambda(u_q)]+O(\alpha), \tag{10}$$

where the prime indicates that the integral is taken in the sense of the principal Cauchy value.

Since formula (5) is valid in the case of a simple spectrum, we have

$$\tan\,\pi\xi[\lambda(u)]=\frac{\pi c(u)}{\frac{d\lambda}{du}\left(1+\int'\frac{c(v)dv}{\lambda(v)-\lambda(u)}\right)},$$

$$c(u)=V(u)+\int'\frac{V(u,u_0)du_0}{\lambda(u_0)-\lambda(u)}+\ldots+\frac{1}{(r-1)!}\int'\cdots\int'\times\frac{V(u,u_0,\ldots,u_{r-2})du_0\ldots du_{r-2}}{[\lambda(u_0)-\lambda(u)]\ldots[\lambda(u_{r-2})-\lambda(u)]},$$

$$V(u_1,\ldots,u_n)=|V_{u_iu_k}|_1^n, \quad V_{u_iu_k}=(V\chi_{u_k},\chi_{u_i})$$

(for simplicity, we assume that H has only a continuous spectrum).

Substituting (10) into (4), we obtain the formulas for the eigenvectors of the simple quasi-continuous spectrum, which are expressed in terms of the shift function. Some of the relations for eigenvectors used in the subsequent analysis are presented in Refs. 27 and 28.

The role of the shift function is that the changes in thermodynamic functions (the traces $\mathrm{Tr}\{f(H_0+V)-f(H_0)\}$) associated with the perturbation V can be expressed in terms of this function.[29] Indeed,

$$\begin{aligned}\mathrm{Tr}\{f(H_{0\alpha}+V)-f(H_{0\alpha})\}&=\sum_k\{f(z_k^{(\alpha)})-f(\lambda^{(\alpha)})\}=\sum_k\{f[\lambda_k^{(\alpha)}+\xi(\lambda(u_k)]\\&\times[\lambda_k^{(\alpha)}-\lambda_{k-1}^{(\alpha)})+o(\alpha)]-f(\lambda_k^{(\alpha)})\}+S\\&=\sum_k\left\{\left.\frac{df}{d\lambda}\right|_{\lambda=\lambda_k^{(\alpha)}}\xi[\lambda(u_k)](\lambda_k^{(\alpha)}-\lambda_{k-1}^{(\alpha)})+o(\alpha)\right\}+S,\end{aligned}$$

where S is the sum over the indices k for which $z_k^{(\alpha)}$ or $\lambda_k^{(\alpha)}$ do not belong to the quasi-continuous spectrum.

In the limit $\alpha\to 0$, we have

$$\mathrm{Tr}\{f(H_0+V)-f(H_0)\}=\int\frac{df}{d\lambda}\xi(\lambda)d\lambda+\sum_l[f(z_i)-f(\lambda_i)], \tag{11}$$

where the integral is taken over the continuous spectrum, z_i is the discrete eigenvalue of H, while λ_i is the discrete eigenvalue of H_0 or the boundary of the continuous spectrum. The numbers z_i and λ_i are determined from the condition that $z_i^{(\alpha)}$, $\lambda_i^{(\alpha)}$, $i=0,\ldots,N-1$ are arranged in a decreasing order. However, the sum in (11) should not necessarily be calculated in this way (see the remark following formula (19)).

2.2.2. Calculation of the density of states and spectral functions of Hamiltonian

The density of states $\eta(\lambda)$ in the continuous spectrum of the operator H can be defined as follows:

$$\eta(\lambda)=\lim_{N\to\infty}\frac{1}{N}\sum_{i=0}^{N-1}\tilde{\rho}_{ii}(\lambda),\quad \tilde{\rho}_{ij}(\lambda)=\frac{(dE_\lambda e_j e_i)}{d\lambda},\quad (12)$$

where E_λ is the unit expansion of the operator

$$H=\int \lambda\, dE_\lambda;$$

and $\tilde{\rho}_{ij}$ is the spectral function corresponding to the vectors e_i and e_j.

Let us consider a crystal lattice with a low concentration of identical local defects, such that the defects can be regarded as independent. Let us suppose that the Hamiltonian H_0 describes a perfect lattice, and the Hamiltonian $H=H_0+V$ describes the lattice with a *single* defect.

We calculate the density of states $\eta(\lambda)$ in the lattice with defects. Substituting into formula (11) the function

$$\theta_x(\lambda)=\begin{cases}0, & \lambda>x,\\ 1, & \lambda\leqslant x,\end{cases}$$

where x is the real number from the continuous spectrum, Lifshits[29] found that

$$\mathrm{Tr}(E_x-E_x^0)=-\xi(x),\quad (13)$$

where E_x^0 and E_x are unit expansions of the operators H_0 and H respectively. Assuming that the concentration of defects is c, we obtain from (13)

$$\eta(x)=\eta_0(x)-c\xi'(x),\quad (14)$$

where $\eta(x)$ and $\eta_0(x)$ are the densities of states in a lattice with defects and in a perfect lattice, respectively.

In problems such as the determination of the Raman spectrum (see, for example, Ref. 39), it is sometimes important to know the spectral functions $\tilde{\rho}_{ij}(\lambda)$ of the operator H. Since

$$\tilde{R}_{ml}(z)=\int_{-\infty}^{\infty}\frac{\tilde{\rho}_{ml}(\lambda)d\lambda}{\lambda-z},$$

where $\tilde{R}(z)=(H-zI)^{-1}$, it follows from Ref. 40 that

$$\tilde{\rho}_{ml}(x)=\frac{1}{\pi}\,\mathrm{Im}\lim_{y\downarrow 0}\tilde{R}_{ml}(x+iy).\quad (15)$$

almost everywhere on the real axis. In order to determine the matrix elements $\tilde{R}_{ml}(z)$, we use the Dyson equation $\tilde{R}=R-RV\tilde{R}$ which can easily be verified by post-multiplying it by $H-zI$. The Dyson equation leads to the relation

$$\tilde{R}=(I+RV)^{-1}R=R-RV(I+RV)^{-1}R,\quad (16)$$

which can be used for determining $\tilde{R}_{ml}(z)$.

2.2.3. Calculation of shift functions

According to Krein,[36] the value of $\xi(\lambda)$ is determined unambiguously by formula (11) from H_0 and V to within values on a set of measure zero. In this case, the spectra of the operators H_0 and H should not necessarily be simple. Consequently, formulas (11) can be regarded as a definition of the function $\xi(\lambda)$, which is more general than (9). (Formulas (13) and (14) obviously remain valid.) By substituting the function $f(\lambda)=1/(\lambda-z)$ and calculating $\mathrm{Tr}\{\tilde{R}(z)-R(z)\}$, Krein found that

$$\xi(x)=\frac{1}{\pi}\lim_{y\downarrow 0}\arg\,\Delta(x+iy),\quad (17)$$

where x and y are real numbers and $\Delta(z)$ is defined in analogy with (2):

$$\Delta(z)=|\,\delta_{lj}+[R(z)Ve_j,e_l]|_0^{r-1}.\quad (18)$$

If, in addition, $\tau_i(i=1,\ldots,r_0)$ are eigenvalues of the operator V, from which the number of positive values is p and the number of negative values is $n=r_0-p$, we have

$$-n\leqslant\xi(x)\leqslant p.\quad (19)$$

(It should be noted that formulas (17) and (19) remain valid if we expand the range of the shift function to the entire set of real numbers by putting

$$\mathrm{Tr}\{f(H_0+V)-f(H_0)\}=\int_{-\infty}^{\infty}\frac{df}{d\lambda}\,\xi(\lambda)d\lambda.$$

In this case, $\int_{-\infty}^{\infty}|\xi(\lambda)|d\lambda\leqslant\sum_{i=1}^{r_0}|\tau_i|$.)

The branch of the argument in (17) can be easily established by substituting $\xi(\lambda)$ into (11), where we chose for $f(\lambda)$ a certain trial function. For example, for $f(\lambda)=\lambda$ we have

$$\mathrm{Tr}\;V=\int\xi(\lambda)d\lambda+\sum_i(z_i-\lambda_i).\quad (20)$$

3. THE THEORY IN THE CASE OF BLOCK-TRIDIAGONAL AND TRIDIAGONAL MATRICES

3.1. Method of Jacobi matrices

The finite-dimensional perturbation theory was developed and found new applications on the basis of the method of Jacobi matrices[30] (see also Refs. 11–13, 19, 20, 24, 31–35). In this method, special type of matrices is considered, which allows us to obtain simpler formulas for $\Delta(z)$ and for matrix elements of the resolvent $R(z)$ than those presented in the previous section.

The method reduces the problem to an analysis of a set of tridiagonal (Jacobi, or J-) matrices of the form

$$J=\begin{pmatrix}a_0 & b_0 & & & 0\\ b_0 & a_1 & b_1 & & \\ & b_1 & a_2 & b_2 & \\ 0 & & \ddots & \ddots & \ddots\end{pmatrix}$$

$$b_1\neq 0,\quad i=0,1,\ldots\quad (21)$$

with real matrix elements, which allows us to use the well-developed theory of J-matrices and the theory of orthogonal polynomial which is closely related to it.

For the further analysis, we will require some basic concepts from the theory of J-matrices (see, for example, Ref. 41).

The first- and second-degree polynomials $[p_i^J(z)$ and $q_i^J(z)$ respectively] associated with the J-matrix of the type (21) are defined as follows:

$$p_{-1}^J=0; \quad p_0^J=1;$$

$$p_{i+1}^J=-b_i^{-1}((a_i-z)p_i^J+b_{i-1}p_{i-1}^J), \quad i=0,1,\ldots$$

$$q_0^J=0; \quad q_1^J=b_0^{-1};$$

$$q_{i+1}^J=-b_i^{-1}((a_i-z)q_i^J+b_{i-1}q_{i-1}^J), \quad i=1,2,\ldots . \quad (22)$$

The spectrum of the truncated $n\times n$ Jacobi matrix $J^{(n)}$ (i.e., $J_{ik}^{(n)}=\delta_{ik}a_i+\delta_{i+1,k}b_i+\delta_{i-1,k}b_k$, $i,k=0,1,\ldots,n-1$) coincides with the roots $\{z_k^{(n)}\}_{k=1}^n$ of the polynomial $p_n^J(z)$. The evaluation of these roots is simplified by the fact that the roots of the polynomial $p_n^J(z)$ are simple, while the roots of two adjacent polynomial $p_n^J(z)$ and $p_{i+1}^J(z)(i=1,2,\ldots)$ alternate. The (non-normalized) eigenvector $\mathbf{x}$ of the matrix $J^{(n)}$ corresponding to the eigenvalue $z_k^{(n)}$ has the components $x_j=p_j(z_k^{(n)}), j=0,\ldots,n-1$.

If the matrix of the operator in the orthonormal basis $\{e_i\}_{i=0}^\infty$ has the form (21), the equality $e_i=p_i^J(J)e_0$, $i=0,1,\ldots$, is valid. Consequently, we can write

$$\delta_{ij}=(e_i,e_j)=[p_i^J(J)p_j^J(J)e_0,e_0]$$
$$=\int_{-\infty}^{\infty}p_i^J(x)p_j^J(x)\rho^J(x)dx, \quad (23)$$

where $\rho^J(x)=(dE_xe_0,e_0)/dx$ is the spectral density of the matrix J, and E_x is the unit expansion of the matrix J.

Equality (23) indicates that the polynomials $\{p_i^J(x)\}_{i=0}^\infty$ form an orthonormal system relative to the weight function (or weight) $\rho^J(x)$. (It should be noted that the properties of relevant polynomials are studied in detail for some weight functions. This refers to Laguerre, Jacobi, Hermite, and other polynomials; see, for example Ref. 42).

The range of applicability of the J-matrix techniques can be divided into *two classes* depending on the possibility of obtaining exact and approximate solutions of the problems:

(1) The methods gives exact solutions for a number of problems for a linear chain ($1D$ system) with interaction between nearest neighbors. We are speaking of problems in the theory of finite-dimensional perturbations, in which the matrix of the Hamiltonian H has the well-known tridiagonal form $H=L+V$, where L is the J-matrix corresponding to polynomials with a known weight and V the finite-dimensional matrix. In the case of a homogeneous chain, the matrices L are identical (i.e., $a_i=a$, $b_i=b$, $i=0,1,\ldots$), and hence the representation $L=p^{J_{\mathrm{Ch}}}+qI$, $p=4b$, $q=a-2b$, where $J_{\mathrm{Ch}\,ii}=1/2$, $J_{\mathrm{Ch}\,ii+1}=1/4$, $i=0,1,\ldots$, is valid. The polynomials $p_i^{J_{\mathrm{Ch}}}(x)$ are the Chebyshev polynomials of the second kind, which are orthogonal on the interval $[0,1]$. Their weight function is well known: $\rho_{\mathrm{Ch}}(\lambda)=(8/\pi)\sqrt{\lambda(1-\lambda)}$ if $\lambda\in[0,1]$ and $\rho_{\mathrm{Ch}}(\lambda)=0$ in the opposite case. Consequently, the spectral density of the matrix L is given by

$$\rho(\lambda)=\frac{1}{|\rho|}\rho_{\mathrm{Ch}}\left(\frac{\lambda-q}{p}\right). \quad (24)$$

(2) The method provides only approximate solutions for a much wider class of problems for $1D$, $2D$, and $3D$ systems since the matrix of any Hamiltonian can be reduced to a set of tridiagonal matrices by using the Lanczos algorithm (see, for example, Refs. 34 and 43). The theoretical basis of this algorithm will be given below.

We choose a certain vector φ (which will be henceforth referred to as a generating vector) and consider the subspace K_φ invariant to the operator H and stretched on the vectors $\varphi,H\varphi,H^2\varphi,\ldots$. Orthogonalization of the sequence of these vectors gives a basis in which the matrix H is tridiagonal.

The Lanczos algorithm involves the simultaneous construction of this basis and the matrix of the operator H in this basis. Namely, we put

$$\varphi_0=\varphi/\sqrt{(\varphi,\varphi)};$$

$$a_0=(H\varphi_0,\varphi_0); \quad \varphi_1'=H\varphi_0-a_0\varphi_0;$$

$$b_0=\sqrt{(\varphi_1',\varphi_1')}; \quad \varphi_1=\varphi_1'/b_0;$$

$$a_1=(H\varphi_1,\varphi_1); \quad \varphi_2'=H\varphi_1-a_1\varphi_1-b_0\varphi_0;$$

$$b_1=\sqrt{(\varphi_2',\varphi_2')}; \quad \varphi_2-\varphi_2'/b_1, \text{ etc.}$$

At the ith step, we have

$$a_i=(H\varphi_i,\varphi_i); \quad \varphi_{i+1}'=H\varphi_i-a_i\varphi_i-b_{i-1}\varphi_{i-1};$$

$$b_i=\sqrt{(\varphi_{i+1}',\varphi_{i+1}')}; \quad \varphi_{i+1}=\varphi_{i+1}'/b_i.$$

Thus, we obtain the matrix (21) in the basis $\{\varphi_i\}_{i=0}^\infty$ of the space K_φ.

In actual practice, we calculate only a finite number of elements of the J-matrix. Using various properties of J-matrices, we can obtain approximate expressions for discrete energy levels and the quantities $(f(H)\varphi,\varphi)$, where $f(x)$ is a certain function, even at this stage (we shall not consider this question here; see Refs. 30, 34, and 41). If, however, the asymptotic behavior of the elements H_{ik} for $i,k\to\infty$ of the matrix H can be determined, it becomes possible to apply the finite-dimensional perturbation theory to operators with a continuous spectrum. This theory coincides with the theory applied in the class 1 of problems described above. Indeed, if the asymptotic behavior of the matrix elements H_{ik} for $i,k\to\infty$ coincides with the behavior of the matrix elements of the J-matrix L corresponding to the well-known system of orthogonal polynomials, we have the representation $H\approx L+V$, where V is the finite-dimensional tridiagonal matrix. (It should be noted that we can extend the range of applicability of the method by considering J-matrices in the Hilbert space of operators and using the Liouville equation.[44–47]) It should be emphasized that the spectrum of J-matrix is *simple*. This means that, if the initial system has a degenerate spectrum, a complete description of its properties can be obtained by choosing several generating vectors (whose number must be not smaller than the spectrum degeneracy) and consider a set of J-matrices.

3.2. Generalization of the method of *J*-matrices

The method of Jacobi matrices can be generalized[19,20] to the case of a special class of Hermite matrices of the form

$$H=\begin{pmatrix} A_0 & B_0 & & & 0 \\ B_0^* & A_1 & B_1 & & \\ & B_1^* & A_2 & B_2 & \\ 0 & & \ddots & \ddots & \ddots \end{pmatrix}, \tag{25}$$

where A_i, B_i are $n\times n$ matrices with real matrix elements for $i=0,1,\ldots$, and inverse matrices B_i^{-1} exist.

Such a generalized method has the following potentialities.

(1) It provides an exact solution for problems with a linear chain in which atoms interact with *several* (to be more precise, n) ''spheres'' of its nearest neighbors, which are similar to problems of class 1 of the method of Jacobi matrices. We are speaking of the problems in which the matrix H of the Hamiltonian is $(2n+1)$-diagonal ($H=(H_{jk})$, where $H_{jk}=0$ if $|j-k|>n$) and can be presented in the form $H=T_n(L)+V$, where $T_n(x)$ is the n-degree polynomials and L is the J-matrix corresponding to polynomials with a known weight, and the matrix V is finite-dimensional. For example, in the case of a homogeneous chain with a point defect, which is interesting from the physical point of view, the rows of the matrix H are identical (i.e., $H_{jj}=a_0$; $H_{j,j+1}=H_{j,j-1}=a_1$; $H_{j,j+n}=H_{j,j-n}=a_n$, where the numbers $\{a_i\}_{i=0}^{n}$ do not depend on the index j) starting from the rth row. It can easily be seen that such a Hamiltonian can be represented in the form $H=T_n(J_{\mathrm{Ch}})+V$.

Obviously, the $(2n+1)$-diagonal matrix has a block-tridiagonal structure (25) with blocks of dimensionality n.[3)]

(2) It also provides an approximate solution for a wider class of problems for $1D$, $2D$, and $3D$ systems since for $n>1$ we can use an algorithm, similar to the Lanczos algorithm and involving n generating orthonormal vectors, and reduce the matrix to the $(2n+1)$-diagonal form. Such a reduction can be useful when the initial system has a degenerate spectrum (since a block-tridiagonal matrix with blocks of dimensionality n *can* have an n-fold degenerate spectrum). A further analogy between the cases with $n=1$ and $n>1$ is clear, but its technical realization is not trivial: the main problem is to estimate the asymptotic behavior of matrix elements. To our knowledge, no publications in this field exist.

3.3. Calculation of spectrum, spectral functions of the Hamiltonian, and shift functions

In accordance with the above arguments, let us write the formulas for calculating the spectrum, spectral functions of the operator $H=H_0+V$, and the shift functions for the case when H and H_0 are block-tridiagonal matrices $[H_0=T_n(L)+V]$. Then the formulas for tridiagonal matrices are obtained as a special case.

Let L be a J-matrix with the spectral density $\rho(\mu)$, $H=T_n(L)+V$ ($H_{ij}=0$ if $|i-j|>n$) the matrix of a self-conjugate operator in the orthonormal basis $\{e_i\}_{i=0}^{\infty}$ of the Hilbert space, $T_n(x)$ the n-degree polynomial, and V a $(2n+1)$-diagonal operator such that $Ve_i=0$ if $i\geqslant r$. We assume that the matrix elements H_{ij} and the coefficients of the polynomial $T_n(x)$ are real-valued.

For subsequent analysis, we shall need the matrix polynomials $P_s^M(z)$ and $Q_s^M(z)$ of the first and second kind, respectively, associated with the block-tridiagonal matrix M of the form (25), which are known to be defined by the formulas

$$P_0^M=I, \quad P_1^M=B_0^{-1}(zI-A_0),$$

$$P_{i+1}^M=-B_i^{-1}[(A_i-zI)P_i^M+B_{i-1}^*P_{i-1}^M],$$

$$i=1,2,\ldots,$$

$$Q_0^M=0, \quad Q_1^M=B_0^{-1},$$

$$Q_{i+1}^M=-B_i^{-1}[(A_i-zI)Q_i^M+B_{i-1}^*Q_{i-1}^M],$$

$$i=1,2,\ldots,$$

where I is a unit $n\times n$ matrix. In particular, for $n=1$ we obtain formulas (22).

We also introduce the vector matrices

$$P^M(z)=\begin{pmatrix} P_0^M(z) \\ P_1^M(z) \\ \vdots \end{pmatrix}, \quad Q^M(z)=\begin{pmatrix} Q_0^M(z) \\ Q_1^M(z) \\ \vdots \end{pmatrix}.$$

It was shown in Ref. 19 that determinant (18) for the special case of the block-tridiagonal matrix $H=H_0+V, H_0=T_n(L)$ under investigation has the form

$$\Delta(z)=c\widetilde{\Delta}(z), \quad \widetilde{\Delta}(z)=|F_{ij}(z)|_0^{n-1},$$

$$F_{ij}(z)=\delta_{ij}+\sum_{s=0}^{[r/n]}\sum_{k=0}^{n-1}[VP^H(z)]_{sn+k,j}$$

$$\times\left\{\sum_{m=0}^{n-1}P_{s,km}^{T_n(L)}(z)R_{im}(z)+Q_{s,ki}^{T_n(L)}(z)\right\},$$

$$R_{im}(z)=\int_{-\infty}^{\infty}\frac{p_m^L(\mu)p_i^L(\mu)}{T_n(\mu)-z}\,\rho(\mu)d\mu. \tag{26}$$

(V is multiplied by the vector composed of the matrix polynomials $P_i^H(z)$ according to conventional rules for the calculation of matrix products), where the factor c is a real number depending on H_0 and V.

It should be noted that in order to find the matrix elements $R_{im}(z)$ in (26), we must know the spectral density $\rho(\mu)$, while in formulas (3) and (18) a large set of spectral parameters (the spectrum and eigenfunctions of the unperturbed Hamiltonian) is used. In addition, in contrast to (18), the dimensionality of the determinant in (26) does not depend on the dimensionality of the $r\times r$ perturbation matrix. For example, in the case when H and H_0 are J-matrices, we have

$$\widetilde{\Delta}(z)=1+\sum_{s=0}^{r-1}[Vp^H(z)]_s\left\{p_s^L(z)\int_{-\infty}^{\infty}\frac{\rho(\mu)d\mu}{\mu-z}\right.$$

$$\left.+q_s^L(z)\right\}. \tag{27}$$

It follows from formulas (2) and (17) that the discrete spectrum of the operator $H=T_n(L)+V$ is determined from the condition $\widetilde{\Delta}(z)=0$, and the shift function from the formula

$$\xi(x)=\frac{1}{\pi}\lim_{y\downarrow 0}\arg\widetilde{\Delta}(x+iy)+\delta, \tag{28}$$

where δ is an integer that can easily be determined together with the branch $\arg(z)$ by the formulas (19) and (20).

For example, in the case when H and $H_0=L$ are Jacobi matrices, we can obtain from (27)

$$\xi(x)=\frac{1}{\pi}\arctan \times\frac{\pi\Sigma_k(Vp^H(x))_k p_k^L(x)\rho(x)}{1+\Sigma_k(Vp^H(x))_k\left(p_k^L(x)\int_{-\infty}^{\infty\prime}\rho(\mu)(\mu-x)^{-1}d\mu+q_k^L(x)\right)}. \tag{29}$$

The eigenvector of the operator H corresponding to the point z_k of the discrete spectrum is $\mathbf{x}=P^H(z_k)\mathbf{c}$. Here the n-component vector $\mathbf{c}$ is determined from the equation $F(z_k)\mathbf{c}=0$.

For the spectral functions of the operator H, we can derive the following expression:

$$\widetilde{\rho}_{ns+i,nt+j}(x)=\sum_{m,l=0}^{n-1}P^H_{s,im}(x)P^H_{t,jl}(x)\widetilde{\rho}_{ml}(x), \tag{30}$$

where

$$\widetilde{\rho}_{ml}(x)=\frac{1}{\pi}\operatorname{Im}\lim_{y\downarrow 0}\widetilde{R}_{ml}(x+iy),\quad 0\leqslant m,l\leqslant n-1$$

and the matrix elements of the resolvent $\widetilde{R}(z)=(H-zI)^{-1}$ are defined as follows:

$$\widetilde{R}_{ml}(z)=\frac{1}{\widetilde{\Delta}(z)}\sum_j\widetilde{\Delta}_{jm}(z)\Bigg[R_{jl}(z)-\sum_{s,i}[VQ^H(z)]_{ns+i,l}\Bigg\{\sum_k P^{T_n(L)}_{s,ik}(z)R_{jk}(z)+Q^{T_n(L)}_{s,ij}(z)\Bigg\}\Bigg], \tag{31}$$

where $\widetilde{\Delta}_{jm}(z)$ is the cofactor of the element $F_{jm}(z)$. The summation over the indices i,j,k is carried out between 0 and $n-1$.

In the case when H and $H_0=L$ are J-matrices, we can obtain the following expression for the spectral density $\widetilde{\rho}_{00}(x)\equiv\widetilde{\rho}(x)$ from (30):

$$\widetilde{\rho}(x)=\frac{\rho(x)}{G(x)},\quad G(x)=\left\{1+\sum_k(Vp^H)_k(p_k^LS+q_k^L)\right\}^2+\left\{\sum_k(Vp^H)_kp_k^L\right\}^2\rho^2\pi^2=|\lim_{y\downarrow 0}\widetilde{\Delta}(x+iy)|^2, \tag{32}$$

where

$$S(x)=\int_{-\infty}^{\infty\prime}\frac{\rho(\mu)}{\mu-x}\,d\mu.$$

If, in addition, $H_0=J_{\mathrm{Ch}}$, the following expression is valid:

$$\widetilde{\rho}(x)=\frac{\rho_{\mathrm{Ch}}(x)}{\Sigma_{k=0}^{2r-1}c_kp_k^H(x)},\quad c_k=\int_{-\infty}^{\infty}\rho_{\mathrm{Ch}}(x)p_k^H(x)dx. \tag{33}$$

This expression was initially used in Ref. 24 for calculating an approximate expression for the distribution functions of the squares of vibrational frequencies of a perfect crystal lattice. It should be noted that there is no need to evaluate the integral in (33): as a matter of fact, according to the orthogonality relations (23) for the polynomials $p_i^{J_{\mathrm{Ch}}}(x)$, $c_k=d_0$ in the expansion

$$p_k^H(x)=\sum_{i=0}^k d_ip_i^{J_{\mathrm{Ch}}}(x),$$

which can be constructed by comparing the coefficients of the powers of the variable x on both sides of the equality.

3.4. Calculations of spectral density of a periodic Jacobi matrix

It was mentioned above more than once that matrices L with identical rows, which are associated with the Chebyshev polynomials, play a significant role in physical problems. In a certain sense, these matrices correspond to systems of identical particles with identical interactions. In this respect, some complex crystal lattices correspond to the so-called periodic Jacobi matrices. (The Jacobi matrix L_t of type (21) is called a periodic matrix with a period t if $a_i=a_{i+tm}, b_i=b_{i+tm}, i,m=0,1,\ldots$.) A simple example of such a lattice is a linear chain of two different alternating spins s and σ, which is described by the Heisenberg Hamiltonian with the interaction between nearest neighbors, if it contains a point spin defect χ. Let us suppose that the spin of the defect is at the site with index 0, and the spins s and σ are at the odd and even sites of the chain respectively. We denote by $|0\rangle$ the normalized state of the complete spin ordering, which is defined by the formulas $\chi_0^2|0\rangle=\chi|0\rangle$, $s_i^z|0\rangle=s|0\rangle$ and $\sigma_{i+1}^z|0\rangle=\sigma|0\rangle$, $i=\pm1,\pm3,\ldots$. In the basis $e_0=\chi_0^-|0\rangle/\sqrt{2\chi}$, $e_i=(s_i^-|0\rangle+s_{-i}^-|0\rangle)/2\sqrt{s}$, $e_{i+1}=(\sigma_{i+1}^-|0\rangle+\sigma_{-i-1}^-|0\rangle)/2\sqrt{\sigma}$, $i=1,3,\ldots$, the matrix of the Hamiltonian has the form $H=L_2+V$, where V is a 2×2 matrix and L_2 is a periodic J-matrix with a period 2. A similar representation of the operator H can be obtained in the basis $g_i=(s_i^-|0\rangle-s_{-i}^-|0\rangle)/2\sqrt{s}$, $g_{i+1}=(\sigma_{i+1}^-|0\rangle-\sigma_{-i-1}^-|0\rangle)/2\sqrt{\sigma}$, $i=1,3,\ldots$ supplementing $\{e_i\}_{i=0}^{\infty}$ to the basis of a one-magnon space. Thus, in order to calculate the shift function and the spectrum of the Hamiltonian on the basis of the formulas from Sec. 3.3, it remains for us to determine the spectral density of the matrix L_2.

The problem on a Bloch electron in an external magnetic field also leads to the structure of a periodic tridiagonal matrix (see, for example, Refs. 50–55).

Let us determine the spectral density $\rho(x)$ of a t-periodic J-matrix L_t, after which the entire theory developed in Sec. 3 can be applied to the matrix $L=L_t$.

For this purpose, we introduce a block-tridiagonal matrix E_t of type (25) with the following blocks of dimensionality $t\times t$: $A_i=0$, $B_i=I, i=0,1,\ldots$.

The following inequality holds:

$$S_t(L_t)=E_t+V, \tag{34}$$

where $S_t(x) = p_t^{L_t}(x) - b_{t-1}q_{t-1}^{L_t}(x)$, and $V_{ik}=0$ for $i,k\geqslant t-1$ (the indices of matrix elements are $i,k,=0,1,\ldots$).

The equality (34) can be derived from the results obtained in Refs. 56 and 57, but it is much more convenient to carry out its direct proof.

Let L be an infinite (in both directions) t-periodic tridiagonal matrix of a symmetric operator in the basis $\{e_i\}_{i=-\infty}^{\infty}$ of a certain space with the matrix elements

$$(Le_i,e_{i+k})=\delta_{-1k}b_{i-1}+\delta_{0k}a_i+\delta_{1k}b_i,$$

$$a_i=a_{i+tm}, \quad b_i=b_{i+tm}, \quad i=0,1,\ldots,t-1,$$

$$m=0,\pm 1,\pm 2,\ldots .$$

We associate the matrix L with two systems of polynomials $p_i(x)$ and $q_i(x)$ defined by the initial conditions $p_{-1}=0$, $p_0=1$, $q_{-1}=1$, $q_0=0$ and by the same recurrence relation $y_{i+1}=\{(x-a_i)y_i-b_{i-1}y_{i-1}\}/b_i$, $i=0,1,\ldots$ (here $y_i=p_i, i=0,1,\ldots$ or $y_i=q_i, i=0,1,\ldots$).

We define the matrix E through the relations $E_{i,i+j}=E_{i,i-j}=\delta_{jt}, i,j=0,\pm 1,\pm 2,\ldots$.

Lemma 2.

$$p_t(L)+q_{t-1}(L)=E. \tag{35}$$

Proof of the lemma. Using the recurrence relations for the polynomials $p_i(x), q_i(x)$ as well as the obvious equality $Le_i=b_{i-1}e_{i-1}+a_ie_i+b_ie_{i+1}$, we find by induction that

$$p_k(L)e_0=e_k-q_k(L)e_{-1}, \quad k=0,1,\ldots . \tag{36}$$

Consequently, we have

$$p_t(L)e_0+q_{t-1}(L)e_0=e_t+\varphi, \tag{37}$$

$$\varphi=-q_t(L)e_{-1}+q_{t-1}(L)e_0. \tag{38}$$

Substituting into (38) the expression for $q_t(L)$ in terms of L, i.e., $q_{t-1}(L), q_{i-2}(L)$, obtained from the recurrence relation and then applying the operator L to the vector e_{-1}, we have

$$\varphi=\frac{b_{t-2}}{b_{t-1}}(-q_{t-1}(L)e_{-2}+q_{t-2}(L)e_{-1}). \tag{39}$$

This is the first step in the transformation of the formula for φ. The second step involves a similar operation for $q_{t-1}(L)$ in (39). Continuing the process, at the $(t-1)$th step we obtain

$$\varphi=\frac{b_0}{b_{t-1}}(-q_1(L)e_{-t}+q_0(L)e_{-t+1})=e_{-t}, \tag{40}$$

i.e.,

$$p_t(L)e_0+q_{t-1}(L)e_0=e_t+e_{-t}. \tag{41}$$

Using (36), (41), and the equality $p_t(L)e_{-1}-p_{t-1}(L)e_0=e_{-t-1}$ which can be proved in analogy with the derivation of the formula (40), we obtain

$$\begin{aligned}(p_t+q_{t-1})e_k&=(p_t+q_{t-1})(p_ke_0+q_ke_{-1})=p_k(e_t+e_{-t})\\&+q_k(p_te_{-1}+e_{t-1}-p_{t-1}e_0)=p_ke_t\\&+q_ke_{t-1}+p_ke_{-t}+q_k(p_te_{-1}-p_{t-1}e_0)\\&=e_{t+k}+p_ke_{-t}+q_ke_{-t-1}=e_{k+1}+e_{k-t},\end{aligned}$$

(for brevity, we have omitted the argument L of the polynomials), which completes the proof of the lemma.

Equality (34) follows directly from Lemma 2 if we premultiply and postmultiply relation (35) by the operator W of the orthoprojection onto the subspace stretched on the vectors $\{e_i\}_{i=0}^{\infty}$ and note that the equality $WL^kW=(WLW)^k+V_k$, where $V_ke_i=0$ for $i\geqslant k-1$, holds for any natural k.

Since the matrix E has a $2t$-fold degenerate spectrum filling the interval $[-2,2]$, and the spectrum of the matrix L is not more than doubly degenerate, relation (35) leads to the well-known fact the spectrum of the matrix L is formed by a set of t intervals, which is the inverse image of the segment $[-2,2]$ in the mapping $S_t(\lambda)=\nu$. Consequently, the continuous spectrum of the matrix $L_t=WLW$ consists of a set of the same intervals.

It follows from formula (34) that in addition to the continuous spectrum filling the interval $[-2,2]$, the operator $S_t(L_t)$ can have up to $t-1$ discrete eigenvalues (the multiplicity of each of these values corresponds to its degeneracy) which can be sought by using the method described in Sec. 3.3. If λ_0 is one of such eigenvalues of multiplicity m, there exists a basis $\{\psi_i\}_{i=1}^{m}$ of the corresponding eigen subspace, such that $L_t\psi_i=\varepsilon_i\psi_i$, where ε_i satisfies the equation $S_t(\varepsilon_i)=\lambda_0$. Thus, the matrix L_t can have up to $t-1$ discrete eigenvalues. (It should be recalled that the spectrum of a J-matrix is simple.)

Since the spectral functions of the operator E_t corresponding to the block A_0 are defined by the formula

$$\rho_{(0)jk}(x)=\delta_{jk}\frac{1}{4}\rho_{\mathrm{Ch}}\left(\frac{x+2}{4}\right), \quad j,k=0,1,\ldots,t-1, \tag{42}$$

we can write

$$\begin{aligned}\lim_{y\downarrow 0}R_{jk}(x+iy)&=\lim_{y\downarrow 0}\int_{-\infty}^{\infty}\frac{\rho_{(0)jk}(\lambda)d\lambda}{\lambda-(x+iy)}\\&=\delta_{jk}\left\{-\frac{x}{2}+i\pi\rho_{(0)jj}(x)\right\}\end{aligned} \tag{43}$$

and the spectral functions $\widetilde{\rho}_{ij}(x)$ of the operator E_t+V are calculated by the formulas (30) and (31) in which the block dimensionality is $n=t$.

In order to determine the spectral density of the matrix L_t, we write Eq. (34) in the form

$$\int_{-\infty}^{\infty}\lambda\left\{\sum_{i=0}^{t-1}\frac{dE_{\varepsilon_i}}{d\varepsilon_i}\left|\frac{d\varepsilon_i}{d\lambda}\right|\right\}d\lambda=\int_{-\infty}^{\infty}\lambda\frac{d\widetilde{E}_{\lambda}}{d\lambda}d\lambda, \tag{44}$$

where $\lambda = S_t(\varepsilon_i)$, $i = 0,1,\ldots,t-1$, and E_x and $\widetilde{E}_x$ are unit expansions of the operators L_t and $E_t + V$, respectively.

Relation (44) leads to

$$\left(\int_{-\infty}^{\infty} \lambda \left\{ \sum_{i=0}^{t-1} \frac{dE_{\varepsilon_i}}{d\varepsilon_i} \left| \frac{d\varepsilon_i}{d\lambda} \right| \right\} d\lambda e_0, e_j \right)$$

$$= \left(\int_{-\infty}^{\infty} \lambda \frac{d\widetilde{E}_\lambda}{d\lambda} d\lambda e_0, e_j \right), \quad j = 0,1,\ldots,t-1. \tag{45}$$

Introducing the notation

$$\rho'(\varepsilon_i) = \rho(\varepsilon_i) \left| \frac{d\varepsilon}{d\lambda} \right|_{\varepsilon=\varepsilon_i} \Bigg|,$$

where $\rho(\varepsilon)$ is the spectral density of the matrix L_t, we obtain from system (45)

$$\sum_{i=0}^{t-1} p_j^{L_t}(\varepsilon_i) \rho'(\varepsilon_i) = \widetilde{\rho}_{0j}(\lambda), \quad j = 0,1,\ldots,t-1. \tag{46}$$

Hence $\rho'(\varepsilon_i) = D_i / D$, where $D = |p_m^{L_t}(\varepsilon_n)|_0^{t-1}$, and D_i is a $t \times t$ determinant which can be obtained from D by replacing the ith column by the column composed by the quantities $\widetilde{\rho}_{01}(\lambda)$. The expression for D can be simplified by expanding the polynomials $p_m^{L_t}(\varepsilon_n)$ in the recurrence formulas successively in each row of the determinant starting from the last row. Repeating this operation $t-1$ times, we obtain

$$D = \frac{|\varepsilon_n^m|_0^{t-1}}{b_{t-2} b_{t-3}^2 \ldots b_0^{t-1}} = \prod_{n=1}^{t-1} \prod_{k=0}^{n-1} \frac{\varepsilon_n - \varepsilon_k}{b_k}.$$

Thus, the spectral density of the matrix L_t on the ith interval of the continuous spectrum has the form

$$\rho(\varepsilon_i) = \frac{D_i}{D} \left| \frac{dS_t(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=\varepsilon_i} \Bigg|, \quad S_t(\varepsilon_i) \in [-2,2],$$

$$i = 0,1,\ldots,t-1, \tag{47}$$

and the correction to spectrum density from each point ε_k of the discrete spectrum has the form

$$\rho_k(\varepsilon) = \delta(\varepsilon - \varepsilon_k) |(\psi_k, e_0)|^2, \tag{48}$$

where the eigenvector ψ_k is assumed to be normalized.

Relations (47) and (48) complete the solution of the problem formulated above. It should be noted that it is simpler to determine the coefficients of the polynomial $S_t(x)$ and the matrix elements of the operator V directly by formula (34) proceeding from the equality of the matrix elements on the right and left sides of the relation [in analogy with the calculation of $T_n(x)$ and V by the formula $H = T_n(L) + V$].

CONCLUSIONS

It follows from the above discussion that practical calculations of spectral and thermodynamic characteristics in problems pertaining to the finite-dimensional perturbation theory can be reduced to the determination of the matrix elements of the resolvent $R(z)$ of the unperturbed operator and to the calculation of the determinant $\Delta(z) = |I + R(z)V|_0^{r-1}$. Special properties of the operators H and H_0 lead to the following simplifications of calculations.

(1) If the operators H and H_0 have a block-tridiagonal structure with blocks of dimensionality n, for an arbitrary rank of the perturbation operator the determinant $\Delta(z)$ coincides with the determinant $\widetilde{\Delta}(z)$ of dimensionality n to within an insignificant factor.

(2) In the case of the representation $H_0 = f(L)$, where $f(x)$ is a certain function and L the Jacobi matrix, the determination of matrix elements of the resolvent R_{ml} requires the knowledge of the spectral density of the J-matrix L, while a larger set of spectral characteristics, viz., the spectrum and the eigenfunctions of the unperturbed Hamiltonian, are required in the general case for calculating the values of $R_{ml}(z)$.

It was mentioned in Sec. 3 that these two conditions are satisfied by the Hamiltonians for a large class of physical problems.

*E-mail: ivk@igorvk.kharkov.ua

[1)]The publications[1,3,7,8,21,22,25–28] are contained in Ref. 23.

[2)]We confine the analysis to the case when the spectrum of the operator H_α is simple.

[3)]The method can be easily extended to the case when the polynomial $T_n(x)$ is replaced by an integrable function $g(x_1,\ldots,x_d)$ of several variables.[20] Among other things, this allows us to analyze *exactly* some of $2D$ and $3D$ problems by using the given method. This refers to the problems in which the relevant Hamiltonians can be represented in the form $H = H_0 + V$, where V is a finite-dimensional perturbation and H_0 a function of several J-matrices with known spectral densities. For example, the space of one-magnon excitations of a simple cubic spin lattice with the interaction between nearest neighbors and with a point spin defect can be expanded into an orthogonal sum of subspaces, in which the matrices of the Heisenberg Hamiltonian have the form $H = L_1 \otimes I \otimes I + I \otimes L_2 \otimes I + I \otimes I \otimes L_3 + V$, where L_i are J-matrices with known spectral densities and I is the unit matrix. The corresponding functions $g(x_1,x_2,x_3)$ in this case are polynomials in their variables. On the other hand, according to Ref. 20, the function $g(x) = e^x$, for example, is connected with the model of a linear chain with an infinitely large range of interaction. In the case when $g(x_1,\ldots,x_d)$ is a polynomial, we can use the fact that the matrix of the Hamiltonian can be regarded as a block-tridiagonal matrix, but, in contrast to the $1D$ case, the block has an infinitely large dimensionality (i.e., $n = \infty$).

[1] I. M. Lifshits, Nuovo Cimento, Suppl. **3**, 716 (1956).
[2] Yu. A. Izyumov, Adv. Phys. **14**, 569 (1965).
[3] I. M. Lifshits and A. M. Kosevich, Rep. Progr. Phys. **29**, 217 (1966).
[4] *Point Defects in Metals, Springer Tracts in Modern Physics* **81**, Berlin (1978).
[5] R. F. Woods and T. M.Wilson, in *Proc. of the Int. Conf. Defects in Insulating Crystals*, Riga (1981).
[6] A. A. Maradudin, in *Proc. of the Int. Conf. Defects in Insulating Crystals*, Riga (1981).
[7] I. M. Lifshits, Zh. Éksp. Teor. Fiz. **18**, 293 (1948).
[8] I. M. Lifshits and V. I. Peresada, Uchen. Zap. Kharkov Univ. **64**, Trudy Fiz. Otd. Fiz. Mat. Fak. **6**, 37 (1955).
[9] G. F. Koster and J. C. Slater, Phys. Rev. **96**, 1208 (1954).
[10] V. I. Peresada, Zh. Éksp. Teor. Fiz. **38**, 1140 (1970) [sic].
[11] V. I. Peresada and V. P. Tolstoluzhskii, Fiz. Nizk. Temp. **3**, 788 (1977) [Sov. J. Low Temp. Phys. **3**, 383 (1977)].
[12] V. I. Peresada and E. S. Syrkin, Fiz. Tverd. Tela (Leningrad) **18**, 336 (1976) [Sov. Phys. Solid State **18**, 197 (1976)].
[13] I. A. Gospodarev, E. S. Syrkin, and V. I. Peresada, Poverkhnost' **7**, 95 (1986).
[14] Yu. A. Izyumov and M. V. Medvedev, *Theory of Magnetically Ordered Crystals with Impurities* [in Russian], Nauka, Moscow (1970).

[15] Yu. A. Izyumov and M. V. Medvedev, Zh. Éksp. Teor. Fiz. **48**, 574 (1965) [Sov. Phys. JETP **21**, 381 (1965)].
[16] T. Wolfram and J. Callaway, Phys. Rev. **130**, 2207 (1963).
[17] T. Oguchi and I. Ono, J. Phys. Soc. Jpn. **26**, 32 (1969).
[18] E. S. Syrkin and S. B. Feodos'ev, Fiz. Nizk. Temp. **20**, 586 (1994) [Low Temp. Phys. **20**, 463 (1994)].
[19] I. V. Krasovsky and V. I. Peresada, J. Phys. **A28**, 1493 (1995).
[20] I. V. Krasovsky and V. I. Peresada, J. Phys. **A29**, 133 (1996).
[21] I. M. Lifshits and G. I. Stepanova, Zh. Éksp. Teor. Fiz. **31**, 156 (1956) [Sov. Phys. JETP **4**, 151 (1956)].
[22] I. M. Lifshits and G. I. Stepanova, Zh. Éksp. Teor. Fiz. **30**, 938 (1956) [Sov. Phys. JETP **3**, 656 (1956)].
[23] I. M. Lifshits, *Selected Works. Physics of Real Crystals and Disordered Systems* [in Russian], Nauka, Moscow (1987).
[24] V. I. Peresada and V. N. Afanas'ev, Zh. Éksp. Teor. Fiz. **58**, 135 (1970) [Sov. Phys. JETP **31**, 78 (1970)].
[25] I. M. Lifshits, Dokl. Akad. Nauk SSSR **48**, 83 (1945).
[26] I. M. Lifshits, Zh. Éksp. Teor. Fiz. **17**, 1017 (1947).
[27] I. M. Lifshits, Zh. Éksp. Teor. Fiz. **17**, 1076 (1947).
[28] I. M. Lifshits, Uchen. Zap. Kharkov Univ. **28**, Zap. NII Mat. Mekh. Kharkov State Univ., Kharkov. Mat. Obshch. **20**, 77 (1950).
[29] I. M. Lifshits, Usp. Mat. Nauk **7**, 171 (1952).
[30] V. I. Peresada, *Physics of Condensed State* [in Russian], vyp. 2, 172 (1968).
[31] V. I. Peresada, Zh. Éksp. Teor. Fiz. **53**, 605 (1967) [Sov. Phys. JETP **26**, 389 (1967)].
[32] V. I. Peresada, V. N. Afanas'ev, and V. S. Borovikov, Fiz. Nizk. Temp. **1**, 461 (1975) [Sov. J. Low Temp. Phys. **1**, 227 (1975)].
[33] R. Haydock, V. Heine, and M. J. Kelly, J. Phys. **C5**, 2845 (1972).
[34] H. Ehrenreich, F. Seitz, and D. Turnbull (eds.), *Solid State Physics: Advances in Research and Applications* **35**, Academic Press, New York (1980).
[35] D. G. Pettifor and D. L. Weaire (eds.), *The Recursion Method and Its Applications*, Springer, Berlin (1985).
[36] M. G. Krein, Matem. Sbornik **33**, 597 (1953).
[37] P. Lancaster, *Theory of Matrices*, Academic Press, New York (1969).
[38] I. V. Krasovsky and V. I. Peresada, Fiz. Nizk. Temp. **20**, 433 (1994) [Low Temp. Phys. **20**, 343 (1994)].
[39] P. D. Loly, Can. J. Phys. **65**, 1272 (1987).
[40] I. I. Privalov, *Boundary Properties of Analytical Functions* [in Russian], Gostekhteoretizdat, Moscow (1950).
[41] N. I. Akhiezer, *Classical Moment Problem*, Oliver and Boyd, Edenburgh, 1965.
[42] G. Szego, *Orthogonal Polynomials*, American Mathematical Society (1939).
[43] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford (1965).
[44] M. H. Lee, Phys. Rev. **B26**, 2547 (1982).
[45] M. H. Lee, J. Math. Phys. **24**, 2512 (1983).
[46] J. Florencio, Jr. and M. H. Lee, Phys. Rev. **B35**, 1835 (1987).
[47] J. F. Annet, W. Matthew, C. Foulkes, and R. Haydock, J. Phys.: Condens. Matter **6**, 6455 (1994).
[48] N. I. Akhiezer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space*, 3rd edition Pitman, Boston, Mass., 1981.
[49] F. V. Atkinson, *Discrete and Continuous Boundary Problems*, Academic Press, New York–London (1964).
[50] M. Ya. Azbel, Zh. Éksp. Teor. Fiz. **46**, 929 (1964) [Sov. Phys. JETP **19**, 634 (1964)].
[51] D. Hofstadter, Phys. Rev. **B14**, 2239 (1976).
[52] G. H. Wannier, G. M. Obermair, and R. Ray, Phys. Status Solidi **B93**, 337 (1979).
[53] P. B. Wiegmann and A. V. Zabrodin, Phys. Rev. Lett. **72**, 1890 (1994).
[54] V. M. Gvozdikov, J. Phys.: Condens. Matter **6**, 6245 (1994).
[55] M. Y. Choi and J. Yi, Phys. Rev. **B52**, 13769 (1995).
[56] Ya. L. Geronimus, Izv. Akad. Nauk SSSR, Ser. Matem. **5**, 203 (1941).
[57] P. B. Naiman, Dokl. Akad. Nauk SSSR **143**, 277 (1962).

Translated by R. S. Wadhwa

Rayleigh-type vibrations localized at the free surface of a fcc crystal

A. M. Kosevich, D. V. Matsokin, and S. E. Savotchenko

*B. Verkin Institute for Low Temperature Physics and Engineering, National Academy of Sciences of the Ukraine, 310164 Kharkov, Ukraine**

(Submitted August 13, 1996)
Fiz. Nizk. Temp. **23**, 92–96 (January 1997)

The waves localized near the free surface (001) of a fcc crystal and propagating along the [110] direction are analyzed in the model of central interaction of nearest neighbors. The frequencies of these waves fall in the gaps of the frequency spectrum of bulk harmonic vibrations for a fixed value of the wave vector k along the surface. The long-wave limit and the case of wave vectors close to the Brillouin zone boundary are studied analytically. These limiting dependences are in accord with the results obtained earlier by other authors by numerical methods. The analytical calculations in the limiting intervals of vector k are supplemented with numerical calculations for arbitrary values of wave vectors. It is significant that the waves under investigation have a displacement component perpendicular to the crystal surface and hence can be studied by standard methods of inelastic scattering of helium atoms. © *1997 American Institute of Physics.* [S1063-777X(97)01001-3]

INTRODUCTION

I. M. Lifshits initiated theoretical studies of localized vibrations in a crystal with defects (see, for example, the reviews in Refs. 1, 2). He considered local vibrations near point defects and introduced the classification (which is generally accepted at present) of local vibrations associated with the dimensions of defects responsible for the localization of crystal lattice vibrations, such as the vibrations near point defects (impurities), near linear defects (dislocations), and near two-dimensional, or planar defects (including the free surface of the crystal). Although the study of surface waves in the theory of elasticity (Rayleigh waves) has a century-long history, investigation of surface waves in the dynamic theory of a discrete crystal lattice remains essential. The application of multilayered crystalline systems in electronics and acoustoelectronics, which is often based on resonant properties of such systems, involves the analysis of peculiarities of the vibrational spectrum, which are associated with surfaces separating individual monocrystalline layers. This research aims at the solution of a similar problem in the simple model of a fcc crystal with the interaction between nearest neighbors.

In Sec. 1, the model of central interaction between nearest neighbors, characterized by only one free parameter, is formulated. However, even this primitive model used earlier more than once[3–5,7] makes it possible to observe peculiarities for vibrations localized near a planar defect. If the defect plane coincides with the crystallographic plane (001), the dynamics of a wave with the displacement parallel to the surface (shear wave) and the dynamics of a Rayleigh-type wave (with the displacement vector having a vertical component and lying in the sagittal plane) in a localized wave propagating in the direction [110] can be separated. The surface wave of the first type, which has only one component, was analyzed in Refs. 3, 5. Here we study vibrations of the second type, which are bipartite. The latter circumstance complicates analytical calculations significantly and makes them cumbersome, but still permits a detailed description of the limiting cases (of small wave vectors, which is the long-wave limit, and of wave vectors close to the Brillouin zone boundary).

In Sec. 2, the dependence of the surface wave frequency on the wave vector is studied. According to the Lifshits general theory, localized vibrations are characterized by frequencies lying outside the continuous frequency spectrum of harmonic vibrations of the crystal lattice. In the case under investigation, propagating harmonic vibrations with a fixed wave vector in the direction indicated above have a spectrum with a gap under the lowermost frequency which adjoins the long-wave frequency of one of transverse waves, and a gap in the frequency spectrum of harmonic vibrations[4,6] for wave vectors near the Brillouin zone boundary (Fig. 1). The frequencies of vibrations localized near the free surface must fall in these gaps. In order to find possible frequencies of surface vibrations, the localized solutions of the equations describing volume vibrations of the crystal, which satisfy the boundary conditions at the crystal surface, are used.

The obtained results are analyzed in Sec. 3. Two types of localized waves are observed, viz., the low-frequency wave existing for all wave vectors in the Brillouin zone and the high-frequency wave existing only in the gap of the spectrum of volume vibrations for wave vectors near the Brillouin zone. The velocity of the low-frequency wave in the long-wave limit coincides with the velocity of a quasi-transverse volume wave. Consequently, it has been proved in conformity to the results obtained earlier that a Rayleigh wave propagating in the [110] direction in the fcc crystal does not exist. The coincidence of the limiting cases of the energy–momentum relation for surface waves and the results of numerical calculations[4,6] is confirmed as well as the qualitative coincidence of the general form of the dependence of frequency of surface vibrations with the pattern for surface waves localized near the free surface (110) in the fcc crystal.[8]

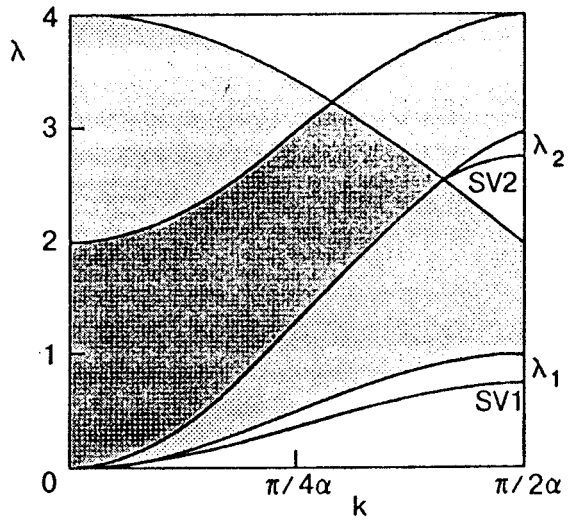

FIG. 1. Dependence of the squared frequency $\lambda = m\omega^2/(4\alpha)$ on the component of the wave vector k along the crystal surface for surface waves SV1 and SV2. Hatched regions correspond to bulk vibrational states.

1. THEORETICAL MODEL

Let us consider the dynamics of a simple fcc crystal consisting of identical atoms of mass m, confining the analysis to the central interaction of nearest neighbors. We choose the coordinate axes in the standard way by determining the coordinates of the nearest neighbors of the selected atom in the form $a(\pm 1,\pm 1,0)$, $a(\pm 1,0,\pm 1)$, $a(0,\pm 1,\pm 1)$. The equations of motion for atoms of the crystal in the harmonic approximation have the form

$$mu_i(\mathbf{n}) = -\sum_{\mathbf{n}'} A_{ik}(\mathbf{n}-\mathbf{n}')u_k(\mathbf{n}'), \tag{1}$$

where $\mathbf{n}$ is the vector number of an atom in the chosen coordinate system and $A_{ik}(\mathbf{n})$ the dynamic matrix of force constants of the crystal, which takes into account the symmetry of the problem and ensures translational and rotational invariance of the energy of the crystal. The force matrix, taking into account the interaction between the nearest and the next neighbors was derived in Ref. 7. We shall use only the approximation of nearest neighbors. In this case, nonzero force matrices $A_{ik}(\mathbf{n})$ in the bulk of the crystal have the form

$$\begin{aligned}
&A_{ik}(0,0,0) = 8\alpha\delta_{ik},\\
&A_{ik}(1,1,0) = -\alpha(\delta_{i1}+\delta_{i2})(\delta_{k1}+gd_{k2}),\\
&A_{ik}(1,0,1) = -\alpha(\delta_{i1}+\delta_{i3})(\delta_{k1}+\delta_{k3}),\\
&A_{ik}(0,1,1) = -\alpha(\delta_{i2}+\delta_{i3})(\delta_{k2}+\delta_{k3}),
\end{aligned} \tag{2}$$

where α is the single force constant in the adopted model ($\alpha>0$). The force matrix for the remaining nearest neighbors is written in a similar way.

Let the crystal in question occupy the half-space $z>0$. The presence of the free surface ($z=0$) is taken into account in the equations of motion of atoms of the boundary layer by changing the corresponding matrices of force constants: the interaction with atoms in the region $z>0$ is zero, and the ''self- action'' element can be written in the form

$$A_{ik}(0,0,0) = \alpha\begin{pmatrix} 6 & 0 & 0\\ 0 & 6 & 0\\ 0 & 0 & 2\end{pmatrix}.$$

We seek the solution of Eq. (1) in the form of a wave localized near the free surface. We assume that the wave propagates along the symmetry line ΓX in the direction of the nearest neighbor: $\mathbf{k}=(k,k,0)$. In this case, we can write

$$u_i(\mathbf{n},t) \equiv u_i(n_x,n_y,-s,t) = u_i^0 q^s \times \exp[iak(n_x+n_y)-i\omega t], \tag{3}$$

where q is the quantity responsible for the decrease in the amplitude towards the bulk of the crystal ($|q|<1$), and s labels planes in the bulk of the crystal ($s=0$ for the boundary plane). Substituting (2) and (3) into (1), we obtain the following equations in the bulk of the crystal:

$$\begin{aligned}
&\left[3-\cos 2ka-\left(q+\frac{1}{q}\right)\cos ka-\frac{m\omega^2}{2\alpha}\right]u_x^0\\
&\quad+[1-\cos 2ka]u_y^0+\left[i\left(q-\frac{1}{q}\right)\sin ka\right]u_z^0=0,\\
&[1-\cos 2ka]u_x^0+\left[3-\cos 2ka-\left(q+\frac{1}{q}\right)\cos ka\right.\\
&\quad\left.-\frac{m\omega^2}{2\alpha}\right]u_y^0+\left[i\left(q-\frac{1}{q}\right)\sin ka\right]u_z^0=0,\\
&\left[i\left(q-\frac{1}{q}\right)\sin ka\right]u_x^0+\left[i\left(q-\frac{1}{q}\right)\sin ka\right]u_y^0\\
&\quad+\left[4-2\left(q+\frac{1}{q}\right)\cos ka-\frac{m\omega^2}{2\alpha}\right]u_z^0=0.
\end{aligned} \tag{4}$$

It can easily be seen that the system (4) splits into the equation for the horizontal shear wave $u_x^0=-u_y^0$, $u_z^0=0$ with the energy–momentum relation

$$\omega^2 = \frac{2\alpha}{m}\left[2-\left(q+\frac{1}{q}\right)\cos ka\right],$$

and the system of equations for a wave with the Rayleigh polarization ($u_x^0=u_y^0, u_z^0\neq 0$):

$$\begin{cases}
\left[4-2\cos 2ka-\left(q+\frac{1}{q}\right)\cos ka-2\lambda\right]u_x^0\\
\quad+\left[i\left(q-\frac{1}{q}\right)\sin ka\right]u_z^0=0,\\
\left[i\left(q-\frac{1}{q}\right)\sin ka\right]u_x^0+\left[2-\left(q+\frac{1}{q}\right)\cos ka-\lambda\right]u_z^0=0,
\end{cases} \tag{5}$$

where $\lambda = m\omega^2/(4\alpha)$.

The solvability of system (5) leads to the obvious relation

$$\left[2-\left(q+\frac{1}{q}\right)\cos ka-\lambda\right]\left[4-2\cos 2ka-\left(q+\frac{1}{q}\right)\cos ka-2\lambda\right]+\left(q-\frac{1}{q}\right)^2\sin^2 ka=0. \tag{6}$$

It defines two branches of the dependence $q=q_\beta(\lambda,k)$, $\beta=1,\ 2$. This means that the surface wave with the vertical polarization (SV) is a bipartite wave:

$$\mathbf{u}=(\mathbf{u}_1^0 q_1^s+\mathbf{u}_2^0 q_2^s)\exp[iak(n_x+n_y)-i\omega t]. \tag{7}$$

It should be noted that, if we put $q=e^{i\varphi}(|q|=1)$ into (6), the latter relation will define the spectrum of possible bulk vibrations with the polarization in question for a fixed k. The regions of these states in Fig. 1 are hatched.

2. ENERGY–MOMENTUM RELATION FOR A WAVE WITH RAYLEIGH POLARIZATION

In order to find the energy–momentum relation, i.e., the dependence $\omega=\omega(k)$ for the wave (7), we substitute (7) into the equation of motion of surface atoms ($z=0$). This gives

$$\begin{aligned}&(3-2\cos 2ka-2\lambda)(u_{1x}+u_{2x})-\cos ka(q_1u_{1x}\\&\quad+q_2u_{2x})+i\sin ka(q_1u_{1z}+q_2u_{2z})=0,\\&i\sin ka(q_1u_{1x}+q_2u_{2x})+(1-\lambda)(u_{1z}+u_{2z})\\&\quad-\cos ka(q_1u_{1z}+q_2u_{2z})=0,\end{aligned} \tag{8}$$

where the superscript ''0'' on the wave amplitude is omitted for the sake of simplicity.

Since the components $u_{\alpha x}$ and $u_{\alpha z}(\alpha=1,2)$ are connected through the relation

$$u_{\alpha x}=i\,\frac{2-(q_\alpha+1/q_\alpha)\cos ka-\lambda}{(q_\alpha-1/q_\alpha)\sin ka}\,u_{\alpha z}, \tag{9}$$

following from (5), the homogeneous linear system (8) defines two amplitudes (e.g., u_{1z} and u_{2z}). By equating to zero the determinant of this system, we obtain an equation for the energy–momentum relation of a surface wave. Unfortunately, an explicit analytic expression for $\omega(\mathbf{k})$ cannot be found since the obtained determinant leads to a high-degree algebraic equation. For this reason, we have to confine the analysis to obvious limiting dependences, supplementing them with numerical calculations on a computer.

3. ANALYSIS OF RESULTS

The long-wave limit ($k\to 0$) of surface vibrations can be analyzed most easily. Only one localized wave (SV1) exists near the free surface of a fcc crystal. However, the velocity of this wave in the long-wave limit is equal to the velocity of the bulk transverse acoustic wave:

$$\omega_1(k)=s_1k,\quad s_1^2=\frac{4\alpha}{m}\,a^2. \tag{10}$$

For this reason, from the point of view of the theory of elasticity, a wave localized near the free surface does not exist in the [110] direction. This conclusion is naturally in accord with the known results in the theory of elasticity (see, for example, Ref. 9).

The splitting of the energy–momentum relation of the surface wave from the acoustic spectrum for small k has the order of magnitude of

$$\frac{s_1k-\omega_1(k)}{s_1k}\sim(ak)^2$$

and increases with k according to an obvious law.

In the limit $ak=\pi/2$, the squared frequency of the SV1 wave is $\omega_1^2=4\alpha\lambda_1/m$, where

$$\lambda_1=\frac{1}{4}\,(7-\sqrt{17}). \tag{11}$$

In the main approximation in small $\xi=(\pi/2)-ak$, we can obtain the following expansion:

$$\omega_1(k)=\omega_1-0{,}6\omega_V\left(\frac{\pi}{2}-ak\right)^2, \tag{12}$$

where $\omega_V^2=4\alpha/m$.

In addition to SV1, a high-frequency localized vibration (SV2) corresponding to the ''window'' in the vibrational spectrum exists near the Brillouin zone boundary (see Fig. 1). The limiting square of the frequency of SV2 is $\omega_2^2=4\alpha\lambda_2/m$, where

$$\lambda_2=\frac{1}{4}\,(7+\sqrt{17}). \tag{13}$$

For small ξ, the energy–momentum relation for SV2 can be written in the form

$$\omega_2(k)=\omega_2-0{,}7\omega_V\left(\frac{\pi}{2}-ak\right)^2. \tag{14}$$

The limiting values of squared frequencies of the waves SV1 and SV2, i.e., the quantities (11) and (13), coincide with the values obtained earlier as a result of numerical calculations.[4,6]

The curve describing the $\omega=\omega(k)$ dependence away from the Brillouin zone boundary can be plotted on the basis of numerical calculations. In complete agreement with the results obtained by Allen *et al.*,[6] this curve terminates at the point corresponding to the ''window'' in the spectrum of bulk vibrations at which the lower boundary of pseudo-longitudinal bulk waves intersects the upper boundary of pseudo-transverse bulk waves (see Fig. 1). A similar result was described by Franchini *et al.*[8] who studied surface waves at the free surface of a fcc crystal of another orientation [namely, (110)]. The authors of Refs. 6, 8 concluded that the dependence $\omega=\omega(k)$ for the so-called resonant states serves as a continuation of the curve describing the energy–momentum relation of the wave SV2 in the continuous spectrum. However, the experience of studying quasi-localized surface waves in an isotropic medium[10,11] shows that this problem must be investigated separately.

It should be noted in conclusion that the surface waves studied here have a displacement component perpendicular to the crystal surface, and hence can be observed, for example, in experiments on scattering of electromagnetic waves at the crystal surface. The main contribution to such a scattering comes just from the component perpendicular to the crystal surface.[12]

The authors are grateful to E. S. Syrkin and A. V. Tutov for fruitful discussion of the obtained results.

This research was carried out under partial support of the ISF, Grant No. U2I200 and the ISSEP, grant No. 042032 as well as project ''Twin'' of the Ukrainian State Committee on Science, Technology, and Industrial Policy.

*E-mail: kosevich@ilt.kharkov.ua

[1] I. M. Lifshits, Nuovo Cimento Suppl. **3**, 716 (1956).
[2] I. M. Lifshits and A. M. Kosevich, Rept. Progr. Phys. **29**, 217 (1966).
[3] I. M. Gelfgat, Fiz. Tverd. Tela **19**, 1711 (1977) [Sov. Phys. Solid State **19**, 998 (1977)].
[4] V. I. Peresada and E. S. Syrkin, Fiz. Nizk. Temp. **3**, 229 (1977) [Sov. J. Low Temp. Phys. **3**, 110 (1977)].
[5] A. M. Kosevich, E. S. Syrkin, and A. V. Tutov, Fiz. Nizk. Temp. **22**, 804 (1996) [Low Temp. Phys. **22**, 617 (1996)].
[6] R. E. Allen, G. P. Alldredge, and F. W. de Wette, Phys. Rev. **B4**, 1661 (1971).
[7] G. Leibfried, *Macroscopic Theory of Mechanical and Thermal Properties of Crystals*, Handbuch der Physik, Vol. 7, PtI (S. Flügge, Ed.) Springer, Berlin, 1955.
[8] A. Franchini, G. Santoro, V. Bortolani, and R. F. Wallis, Phys. Rev. **B38**, 12139 (1988).
[9] D. C. Gazis, R. Herman, and R. F. Wallis, Phys. Rev. **119**, 533 (1960).
[10] A. M. Kosevich and A. V. Tutov, Fiz. Nizk. Temp. **19**, 1273 (1993) [Low Temp. Phys. **19**, 905 (1993)].
[11] A. M. Kosevich and A. V. Tutov, Phys. Lett. **A213**, 265 (1996).
[12] G. Braco, T. Tatarek, F. Tommasini *et al.*, Phys. Rev. **B36**, 2928 (1987).

Translated by R. S. Wadhwa

The smoothing of the surface structure in solids under irradiation

V. V. Slezov and V. M. Apalkov

*National Research Center ''Kharkov Physicotechnical Institute,'' 310108 Kharkov, Ukraine**

Yu. I. Boĭko

Kharkov State University, 310077 Kharkov, Ukraine

H. D. Carstanjen

Max Planck Institute of Metal Science, Heisenbergstrasse 1, Stuttgard, 70569 Germany

(Submitted May 15, 1996)
Fiz. Nizk. Temp. **23**, 97–109 (January 1997)

The general system of equations describing the smoothing of surface structure in solids under irradiation is formulated. It is shown that under real conditions the system can be reduced to a simpler form which has an exact solution. A general formula describing the smoothing of surface structure with time is derived. As a special case, this formula gives the evolution (smoothing) of the relief in the absence of radiation, which is associated with the difference in curvatures for different roughnesses. The physical reason behind the intensification of smoothing under irradiation is the difference between the positions of the centers of gravities for concentration profiles of vacancies and interstitials. The radiation gives rise to a new type of dependence of the rate of smoothing on the roughness parameters of the surface, which permits the experimental separation of the contribution from radiation to smoothing and simultaneously makes it possible to determine some parameters which are difficult to measure.
© *1997 American Institute of Physics.* [S1063-777X(97)01101-8]

1. BASIC SYSTEM OF EQUATIONS

Diffusive processes in solids are important from the point of view of the theory and applications since they determine many important properties of solids. The works by I. M. Lifshits[1,2] play the main role in the study of this processes. The corresponding problem formulated in these publications in the general form stimulated the development of a new trend in investigations. This paper also continues the research work associated with I. M. Lifshits.

The evolution (smoothing) of roughnesses on the surface of a solid is analyzed for determining important constants of the material such as the surface and volume diffusion coefficients.[3,4] These constants are required for describing many processes associated with the transport of mass in solids (such as fritting of powders, creep, and formation of thin surface layers). In this problem, it is very important to take into account the role of sources of point defects created by various types of radiation. This paper is devoted to an analysis of this problem.

The physical reason behind smoothing of the surface structure in solids is the difference in the equilibrium concentrations of vacancies in convex and concave regions of the surface. The equilibrium concentration of vacancies in the convex and concave regions is $K>0$, $u_K>u_0$ and $K<0$, $u_K<u_0$, respectively, where u_K is the equilibrium surface concentration of vacancies on the surface with the curvature K, while u_0 is the equilibrium concentration of vacancies corresponding to a plane surface ($K=0$). Thus, the vacancies at convexities and concavities should be redistributed to attain a thermodynamically equilibrium state corresponding to a plane surface. The concentration of surface vacancies is determined by their redistribution over the surface due to nonuniformity in the surface curvature as well as by the flow of vacancies from the bulk since the concentration of vacancies in the bulk of the sample is uniform and corresponds to an equilibrium value c_0, while near the surface it is also determined by the curvature K.

The problem of smoothing the surface structure under the influence of the above factors was considered theoretically by Mullins[3] and tested experimentally by Hoehne and Sizmann.[4] Here we consider the problem in which the smoothing of the surface is initiated not only by the curvature of the relief, but also by sources of vacancies and interstitial atoms. Such a situation emerges for various types of radiation acting on the surface of a solid. Point defects formed as a result of irradiation make an additional contribution to the process of smoothing.

In order to consider the smoothing of roughnesses, we introduce the volume (c) and the surface (u) concentrations of vacancies. The concentrations are normalized to a lattice site.[1)] The system of equations describing the change of the sample surface as a result of diffusion flows and taking into account the presence of a source of vacancies has the form[5]

$$\begin{cases} \dfrac{\partial u}{\partial t} = D_S \Delta_S u - \nu(u-u_K) + \dfrac{D_V}{a} \left.\dfrac{\partial c}{\partial \mathbf{n}}\right|_{z=f(x,y,t)} \\ \dfrac{\partial c}{\partial t} = D_V \Delta c + I(x,y,z,t) \\ \left.\dfrac{\partial z}{\partial t}\right|_{z=f(x,y,t)} = \alpha \nu (u-u_K)\cos(\mathbf{n},z), \end{cases} \tag{1}$$

where D_V, D_S are the volume and surface diffusion coefficients for vacancies, ν the frequency of vacancy absorption

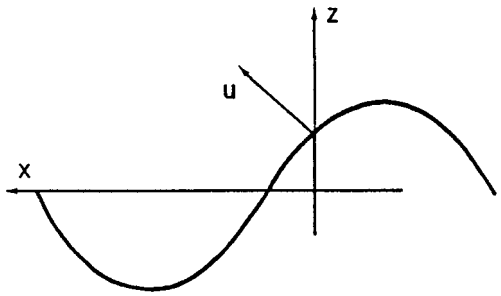

FIG. 1. Orientation of coordinate axes. The z-axis is directed to the bulk of the sample, and the normal $\mathbf{n}$ to the sample surface forms an acute angle with the z-axis.

by the surface, u_K the equilibrium surface concentration of vacancies, which corresponds to the curvature K: $(u_K = u_0 + u_0 \gamma_S \Omega K/(kT)$, γ_S is the surface tension, $\Omega = a^3$, T is the sample temperature, $I(x,y,z,t)$ the intensity of a source of vacancies, i.e., the number of vacancies generated per unit time per lattice site, the z-axis is directed into the bulk of the sample, $z = f(x,y,t)$ is the equation of the surface at the instant t, and $\mathbf{n}$ the inward normal to the sample surface, $\cos(\mathbf{n},z) > 0$ (Fig. 1).

The boundary conditions to system (1) have the form

$$D_V \left.\frac{\partial c}{\partial \mathbf{n}}\right|_{z=f(x,y,t)} = \left.\left(\frac{\beta D_V}{a}\, c - \frac{a}{\tau_S}\, u\right)\right|_{z=f(x,y,t)}, \qquad (2)$$

where $\beta D_V = D'$ (D' determines the last jump for a vacancy emerging from the bulk to the free surface, β is the dimensionless coefficients that takes into account the presence of a potential barrier for a vacancy emerging at the sample surface ($0 < \beta \leqslant 1$), and τ_S is the time of departure of the vacancy to the bulk, $\tau_S = 1/\nu$.

The coefficients in (2) are connected through the following condition: in equilibrium, the vacancy concentration in the bulk is constant, i.e., $\partial c/\partial \mathbf{n} = 0$; in this case, relation (2) leads to $u_K/c_K = \beta D_V \tau_S / a^2$, where c_K is the equilibrium volume concentration of vacancies corresponding to the curvature K: $c_K = c_0 + c_0 \gamma_S \Omega K/kT$, and c_0 is the equilibrium volume concentration of vacancies, corresponding to a plane surface. For dilute solutions of vacancies in the surface region, the equality of their chemical potentials leads to

$$\mu_S = \psi_S + kT \ln u_K = \mu_V = \psi_V + kT \ln c_K,$$

$$\frac{u_K}{c_K} = \exp\left(\frac{\psi_V - \psi_S}{kT}\right) = \beta \frac{D_V \tau_S}{a^2}. \qquad (3)$$

The initial condition for the system (1), (2) have the form

$$\begin{cases} u[x,y,f(x,y,0),t=0] = u_K[x,y,f(x,y,0)] \\ f(x,y,t=0) = f^{(0)}(x,y) \\ D_V \left.\dfrac{\partial c}{\partial \mathbf{n}}\right|_{z=f(x,y,0)} = \left.\left(\beta \dfrac{D_V}{a}\, c - \dfrac{a}{\tau_S}\, u\right)\right|_{z=f(x,y,0)} \\ c(x,y,z,t)\big|_{t=0} = c_0, \quad z > f(x,y,0). \end{cases}$$

In the given problem, we have two characteristic time scales: the short time (of absorption) $\tau_S = 1/\nu \sim a^2/(\beta D_V)$ (this expression for τ_S follows from (3)), and the long (relaxation) time $\tau_{\text{rel}} \sim \lambda^2/D_S$, where λ is the separation between identical roughnesses on the surface. For example, for $D_V \sim 10^{-12}$ cm^2/s, $D_S \sim 10^{-8}$ cm^2/s, $\lambda \sim 10^{-3}$ cm, and $a \sim 10^{-8}$ cm, we obtain $\tau_S \sim 10^{-4}$ s and $\tau_{\text{rel}} \sim 10^2$ s. Since $a \ll \lambda$, $\tau_{\text{rel}} \gg \tau_S$ (for $\beta \sim 1$). In view of these inequalities and for a time $t \gg \tau_S$, the first equation of the system (1) and the boundary condition (2) taking into account (3) in the form

$$D_V \left.\frac{\partial c}{\partial \mathbf{n}}\right|_{z=f(x,y,t)} = \left.\left(\frac{\beta D_V}{a}\, c - \frac{a}{\tau_S}\, u\right)\right|_{z=f(x,y,t)}$$

$$= \left.\left(\frac{\beta D_V}{a}\,(c - c_K) - \frac{a}{\tau_S}\,(u - u_K)\right)\right|_{z=f(x,y,t)},$$

indicate that the difference between the vacancy concentrations u and u_K is small as well as the difference between $c|_{z=f(x,y,t)}$ and c_K:

$$|u - u_K| = \frac{1}{\nu}\left| D_S \Delta_S u + \frac{D_V}{a} \left.\frac{\partial c}{\partial \mathbf{n}}\right|_{z=f(x,y,t)} - \frac{\partial u}{\partial t}\right| \sim \frac{1}{\nu}\left| D_S \frac{u_K}{\lambda^2} + \frac{D_V}{a}\frac{c_K}{\lambda} + I_0 a^2 + \frac{u_K}{t}\right| \sim \frac{u_K}{\nu \tau_{\text{rel}}} + \left(\frac{a}{\lambda}\right)\frac{u_K}{\beta \tau_S \nu} + \frac{u_K}{t\nu} + \frac{I_0 a^2}{\nu} \ll u_K;$$

$$|c|_{z=f(x,y,t)} - c_K| = \left|\frac{a^2(u - u_K)}{\beta D_V \tau_S} + \frac{a}{\beta}\left.\frac{\partial c}{\partial \mathbf{n}}\right|_{z=f(x,y,t)}\right| \sim \left|\frac{u - u_K}{u_K}\, c_K + \frac{a}{\lambda}\frac{c_K}{\beta} + I_0 a^2 \tau_S\right| \ll c_K,$$

where we used the estimate $\partial c/\partial \mathbf{n} \sim c_K/\lambda + I_0 \Omega / D_V$ (I_0 is the number of vacancies per unit surface area of the sample per unit time) and the condition $I_0 a^2 \tau_S \ll u_K$ (this means that the number of vacancies emerging in a unit area of the surface during the time $\sim \tau_S$ is much smaller than the equilibrium number density). In other words, the tuning of concentration u and u_K and of $c|_{z=f(x,y,t)}$ to c_K occurs during the short time τ_S. This statement can be formally obtained from (1) and (2) for $\nu \to \infty$ in all the terms except $\nu(u - u_K)$. Using the equations from the system (1) and (2), we can obtain the following correction in $1/\nu$ to the quantities of interest. For example, the first equation of the system (1) and the boundary condition (2) immediately lead to the correction to $c|_{z=f(x,y,t)}$:

$$c|_{z=f(x,y,t)} = c_K + \frac{a^2 D_S}{\beta D_V} \Delta_S u_K = c_K + \frac{D_S}{\nu} \Delta_S c_K.$$

It follows hence that the expansion is actually carried out in powers of $D_S/(\lambda^2 \nu)$. Taking into account what has been said above, we can write the initial system of equations (1) and (2) to within the first nonzero approximation in $D_S/(\lambda^2 \nu)$ in the form

$$\begin{cases} D_V \Delta c + I(x,y,z,t) = 0 \\ \left.\dfrac{\partial z}{\partial t}\right|_{z=f(x,y,t)} = \left(D_S a \Delta_S u_K + D_V \left.\dfrac{\partial c}{\partial \mathbf{n}}\right|_{z=f(x,y,t)}\right) \cos(n,z) \end{cases} \qquad (4)$$

with the initial and boundary conditions

$$\begin{cases} c(x,y,z,t)|_{z=f(x,y,t)}=c_K(x,y,t) \\ f(x,y,t=0)=f^{(0)}(x,y) \end{cases}.$$

In the system of equations (4) we take into account the fact that for a real experimental time $[t \gg \tau_{\max}=\max(\lambda^2/D_S, l^2/D_V] \sim 10^2$ s, where l is the characteristic depth at which a source of vacancies is located) and for a long characteristic time of variation of roughness of the sample surface ($(f/\dot{f}) \gg \tau_{\max}$, which is observed under real conditions since $(f/\dot{f}) \sim 10^5$ s),[4] we can neglect the time derivatives $(\partial c/\partial t \approx \partial \mu/\partial t \approx 0)$ and obtain a quasi-stationary state to within the terms $\sim \tau_{\max}/t$, where the depth up to which the solution can be regarded as stationary is defined by the condition $z \ll \sqrt{D_V t}$. Thus, the time t is a parameter at this quasi-stationary stage. In other words, we must first determine quasi-steady-state fluxes for a given shape of the surface relief, and then find a closed equation determining the time variation of the surface on the basis of these fluxes. It follows from the second equation of system (4) that we must determine only the vacancy flux $\partial c/\partial \mathbf{n}|_{z=f(x,y,t)}$ from the bulk to the surface. For this purpose, we must solve the first equation from system (4) to the above accuracy and with the boundary condition $c|_{z=f(x,y,t)}=c_K$.

2. SOLUTION OF THE BASIC SYSTEM OF EQUATIONS

Henceforth, we shall consider the shape of the surface periodic in x, i.e.,

$$z=f(x,t)=\sum_n z_n(t)\mathrm{e}^{-in\omega x}, \quad \omega=\frac{2\pi}{\lambda}, \tag{5}$$

where λ is the characteristic distance between equivalent surfaces (period in x). For a given shape of the surface, the dependence on y is absent, and the variable y can be omitted. Using (5), we obtain the following expression for c_K:

$$\begin{aligned} c_K(x,t) &= c_0 + c_0 \frac{\gamma_S \Omega K}{kT} = c_0 \\ &\quad + \frac{c_0}{kT}\gamma_S \Omega \sum_n z_n(t) n^2 \omega^2 \mathrm{e}^{-in\omega x} \\ &= c_0 + B\sum_n z_n(t) n^2 \omega^2 \mathrm{e}^{-in\omega x}, \end{aligned} \tag{6}$$

where $B=c_0\gamma_S\Omega/(kT)$, and we have taken into account the fact that

$$K \approx -\partial^2 z/\partial x^2 = \sum_n z_n n^2 \omega^2 \mathrm{e}^{-in\omega x}.$$

for small z_n/λ.[6]

Thus, we must solve the following equation in c:

$$\begin{cases} D_V \Delta c(x,z,t) + I(x,z,t) = 0 \\ c[x, z=f(x,t), t] = c_K(x,t) \end{cases}, \tag{7}$$

where t appears only as a parameter through $z_n(t)$.

In order to solve Eq. (7), we introduce a new function $V(\hat{x},\hat{z})$ and carry out the substitution of variables

$$c(x,z,t)=V(\hat{x},\hat{z},t)+c_K(\hat{x},t), \tag{8}$$

where $\hat{x}=x$ and $\hat{z}=z-f(x,t)$. Substituting (8) into (7), we obtain a nonhomogeneous equation for V with zero boundary conditions on the plane surface $\hat{z}=0$:

$$\begin{cases} \Delta V - f_{\hat{x}\hat{x}} V_{\hat{z}} - 2 f_{\hat{x}} V_{\hat{x}\hat{z}} = -c_{K,\hat{x}\hat{x}} - \dfrac{I(\hat{x},\hat{z})}{D_V}, \\ V(\hat{x},\hat{z}=0)=0, \end{cases} \tag{9}$$

where φ_x denoted the derivative of φ with respect to x: $\varphi_x=\partial\varphi/\partial x$, and $\varphi_{xx}=\partial^2\varphi/\partial x^2$.

The differential equation (9) contains two inhomogeneous terms: $c_{K,\hat{x}\hat{x}}$ and I/D_V. In view of the linearity of the homogeneous part of Eq. (9), we can consider the following two separate problems:

$$\begin{cases} \Delta V_C - f_{\hat{x}\hat{x}} V_{C,\hat{z}} - 2 f_{\hat{x}} V_{C,\hat{x}\hat{z}} = -c_{K,\hat{x}\hat{x}} \\ V_C(\hat{x},\hat{z}=0)=0 \end{cases} \tag{10}$$

and

$$\begin{cases} \Delta V_I - f_{\hat{x}\hat{x}} V_{I,\hat{z}} - 2 f_{\hat{x}} V_{I,\hat{x}\hat{z}} = -\dfrac{I(\hat{x},\hat{z})}{D_V} \\ V_I(\hat{x},\hat{z}=0)=0. \end{cases} \tag{11}$$

The solution of Eq. (9) in this case is equal to the sum of solutions of Eqs. (10) and (11): $V=(V_C+V_I)$ (superposition principle). In this case, in accordance with (8), the concentration $c(x,z,t)$ is given by

$$c(x,z,t)=c_C(x,z,t)+c_I(x,z,t),$$

where we have introduced the following notation:

$$\begin{aligned} c_C(x,z,t) &= V_C(\hat{x},\hat{z},t)-c_K(\hat{x},t) \\ c_I(x,z,t) &= V_I(\hat{x},\hat{z},t). \end{aligned} \tag{12}$$

2.1. The role of surface and volume diffusion in smoothing of the surface structure of a solid

Let us first consider Eq. (10). We shall seek the solution in the form of a power series in z_n/λ to within the first-order terms in z_n/λ: $V_C=V_C^{(0)}+V_C^{(1)}$, where $V_C^{(0)}$ and $V_C^{(1)}$ denote the zeroth order in z_n/λ. It follows from (7) that

$$f_{\hat{x}} = -i\sum 2\pi(z_n/\lambda)n \exp(-in\omega\hat{x})$$

is of first order in z_n/λ; in this case, Eq. (10) can be written to within z_n/λ in the form

$$\begin{cases} \Delta V_C^{(0)} + \Delta V_C^{(1)} - f_{\hat{x}\hat{x}} V_{C,\hat{z}}^{(0)} - 2 f_{\hat{x}} V_{C,\hat{x}\hat{z}}^{(0)} = -c_{K,\hat{x}\hat{x}} \\ V_C^{(0)}(\hat{x},\hat{z}=0) + V_C^{(1)}(\hat{x},\hat{z}=0)=0. \end{cases} \tag{13}$$

Since $c_K \propto z_n/\lambda$ (according to (6)), Eq. (13) splits into the following system of equations each of which corresponds to its own order of smallness in z_n/λ:

$$\begin{cases} \begin{cases} \Delta V_C^{(0)}=0 \\ V_C^{(0)}(\hat{x},\hat{z}=0)=0 \end{cases} \\ \begin{cases} \Delta V_C^{(1)} - f_{\hat{x}\hat{x}} V_{C,\hat{z}}^{(0)} - 2 f_{\hat{x}} V_{C,\hat{x}\hat{z}}^{(0)} = -c_{K,\hat{x}\hat{x}} \\ V_C^{(1)}(\hat{x},\hat{z}=0)=0. \end{cases} \end{cases} \tag{14}$$

It follows from the first equation of system (14) that $V_C^{(0)}=0$. Substituting this result into the second equation of system (14), we obtain the two-dimensional Poisson equation for $V_C^{(1)}$:

$$\begin{cases} \Delta V_C^{(1)} = -c_{K,\hat{x}\hat{x}} \\ V_C^{(1)}(\hat{x},\hat{z}=0)=0. \end{cases} \tag{15}$$

It is well known[7] that the solution of Eq. (15) can be conveniently expressed in terms of the Green's function of the two-dimensional Laplace operator: $G_2(\hat{x},\hat{z})=1/(4\pi) \times \ln[(\hat{x}^2+\hat{z}^2)/L^2]$, where L is the sample size ($L\to\infty$) and $G_2(\hat{x},\hat{z})$ satisfies the condition $\Delta G_2(\hat{x},\hat{z}) = \delta(\hat{x})\delta(\hat{z})$, where $\delta(\hat{x})$ is the Dirac delta-function. It should be recalled that, in view of the absence of the dependence on y, the two-dimensional Laplace operator $\Delta=\partial_{\hat{x}}^2+\partial_{\hat{x}}^2$ actually appears in Eqs. (7)–(15). The boundary condition $V_C^{(1)}|_{\hat{z}=0}=0$ is taken into account by specular continuation of the source to the entire space (this is a standard approach in electrostatics for the problem with a charge near the metal surface). After the substitution of the explicit expression for $c_{K,\hat{x}\hat{x}}(\hat{x},t)$ from (6), $V_C^{(1)}$ can be expressed in terms of tabulated integrals:

$$\begin{aligned} V_C^{(1)}(\hat{x},\hat{z},t) &= \int_{-\infty}^{\infty} dx' \int_0^{\infty} dz' \, \frac{1}{4\pi} \, c_{K,x'x'}(x',t)\{\ln[(\hat{x}-x')^2 \\ &\quad +(\hat{z}-z')^2]-\ln[(\hat{x}-x')^2+(\hat{z}+z')^2]\} \\ &= \int_{-\infty}^{\infty} dx' \int_0^{\infty} dz' \, \frac{B}{4\pi} \sum_n z_n(t) n^4 \omega^4 e^{-in\omega x'} \\ &\quad \times\{\ln[(\hat{x}-x')^2+(\hat{z}-z')^2]-\ln[(\hat{x}-x')^2 \\ &\quad +(\hat{z}+z')^2]\} \\ &= B\sum_n z_n(t) n^2 \omega^2 e^{-in\omega\hat{x}}(e^{-n\omega\hat{z}}-1). \end{aligned}$$

Using relation (12), we obtain

$$\begin{aligned} c_C(x,z,t) &= V_C(\hat{z},\hat{x},t)+c_K(\hat{x},t) \\ &= c_0 + B\sum z_n(t) n^2 \omega^2 e^{-in\omega\hat{x}} e^{-n\omega\hat{z}}, \end{aligned} \tag{16}$$

where $\hat{x},\hat{z}$ and x,z are connected through relation (8). The latter relation immediately leads to the following expression for the derivative with respect to the normal $\mathbf{n}$ in the new coordinates to within $f_{\hat{x}}\sim z_n/\lambda$:

$$\frac{\partial}{\partial \mathbf{n}} = \frac{\partial}{\partial \hat{z}} - f_{\hat{x}} \frac{\partial}{\partial \hat{x}}. \tag{17}$$

Using the relation (3) between u_K and c_K as well as the fact that $\cos(\mathbf{n},z)\cong 1$ to within $(z_n/\lambda)^2$, we substitute (6), (16), and (17) into the second equation of the system (4) to obtain the equation describing the smoothing of the surface of a solid in the absence of irradiation:

$$\begin{aligned} \frac{\partial f(x,t)}{\partial t} &= -\Bigg\{ D_V \frac{c_0\gamma_S\Omega}{kT} \sum_n z_n(t)\omega^3 n^3 e^{-in\omega x} \\ &\quad + D_S a\left(\frac{\beta D_V \tau_S}{a^2}\right) \frac{c_0\gamma_S\Omega}{kT} \sum_n z_n(t)\omega^4 n^4 e^{-in\omega x}\Bigg\}, \end{aligned}$$

$$\begin{aligned} \frac{dz_n(t)}{dt} &= -z_n(t)\Bigg\{ D_V \frac{c_0\gamma_S\Omega}{kT} \omega^3 n^3 \\ &\quad + D_S a\left(\frac{\beta D_V \tau_S}{a^2}\right) \frac{c_0\gamma_S\Omega}{kT} \omega^4 n^4\Bigg\}, \end{aligned} \tag{18}$$

where we have used the explicit expression (5) for $f(x,t)$.

The obtained equation (18) coincides (to within the factor 2) with the sum of equations (9) and (14) from Ref. 4 under the condition $(\beta D_V\tau_S)/a^2=1$ (this condition indicates that $c|_{z=f(x,t)}=u$) and after the substitution of the density for the vacancy concentration ($c\to c/a^3$ and $u\to u/a^2$) and corresponds exactly to the solution obtained in Ref. 3. It should be noted that Hoehne and Sizmann[4] took into account twice the same reason behind vacancy redistribution in Eq. (6).

2.2. Source of vacancies and increase in the roughness on the surface of a solid

Let us now consider Eq. (11) describing the correction to (18) associated with an additional source of vacancies. We take the source in the form of a δ-function: $I(\hat{x},\hat{z})=I_0\Omega\,\delta(\hat{z}-l)$, where I_0 is the number of vacancies emerging at the surface per unit time per unit area. The absence of a dependence of I on $\hat{x}$ indicates that the source repeats completely the shape of the sample surface. Taking this type of the source of vacancies into account, we can write Eq. (11) in the form

$$\begin{cases} \Delta V_I - f_{\hat{x}\hat{x}} V_{I,\hat{z}} - 2 f_{\hat{x}} V_{I,\hat{x}\hat{z}} = -\dfrac{I_0\Omega}{D_V}\,\delta(\hat{z}-l) \\ V_I(\hat{x},\hat{z}=0,\,t)=0. \end{cases} \tag{19}$$

We seek the solution of this equation in the form of a power series in z_n/λ. Substituting $V_I=V_I^{(0)}+V_I^{(1)}$ into Eq. (19) (where $V_I^{(0)}$ and $V_I^{(1)}$ have the zeroth and the first order in z_n/λ, respectively), we obtain the following expressions accurate to within z_n/λ:

$$\begin{cases} \Delta V_I^{(0)} + \Delta V_I^{(1)} - f_{\hat{x}\hat{x}} V_{I,\hat{z}}^{(0)} - 2 f_{\hat{x}} V_{I,\hat{x}\hat{z}}^{(0)} = -\dfrac{I_0\Omega}{D_V}\,\delta(\hat{z}-l) \\ V_I^{(0)}(\hat{x},\hat{z}=0,\,t)+V_I^{(1)}(\hat{x},\hat{z}=0,\,t)=0. \end{cases} \tag{20}$$

Using the fact that $f_x\sim z_n/\lambda$, we obtain the following equation for the zeroth order $V_I^{(0)}$:

$$\begin{cases} \Delta V_I^{(0)} = -\dfrac{I_0\Omega}{D_V}\,\delta(\hat{z}-I) \\ V_I^{(0)}(\hat{x},\hat{z}=0,\,t)=0. \end{cases}$$

As in the case considered above (Eq. (15)), the solution can be expressed in terms of Green's function taking into account specular continuation. The integrals can easily be evaluated, which gives

$$\begin{aligned} V_I^{(0)}(\hat{x},\hat{z}) &= -\int_{-\infty}^{\infty} dx' \int_0^{\infty} dz' \, \frac{I_0\Omega}{4\pi D_V}\,\delta(z'-l) \\ &\quad \times\{\ln[(\hat{x}-x')^2+(\hat{z}-z')^2]-\ln[(\hat{x}-x')^2 \\ &\quad +(\hat{z}+z')^2]\} = \frac{I_0\Omega}{D_V}\begin{cases} \hat{z}, & l>\hat{z}>0 \\ l, & \hat{z}>l \end{cases}. \end{aligned}$$

It can be seen that the results corresponds exactly to the electrostatic analogy (a positively charged plane near a metal surface). It should be noted that the above result for $V_I^{(0)}$ is inapplicable for large $\hat{z} > \sqrt{D_V t}$, where the solution of the initial system (1) is naturally nonstationary. However, in order to find the fluxes from the bulk to the surface of the sample, we must know the behavior of the vacancy concentration only near the surface $\hat{z} \approx 0$, where the problem is stationary, and the above result f $V_I^{(0)}$ is applicable. Considering that the solution $V_I^{(0)}$ is a function of $\hat{z}$ alone, we obtain from (20) the following first order equation for $V_I^{(1)}$ in z_n/λ:

$$\begin{cases} \Delta V_I^{(1)} = f_{\hat{x}\hat{x}} V_{I,\hat{z}}^{(0)} \\ V_I^{(1)}(\hat{x}, \hat{z}=0,\ t) = 0. \end{cases}$$

As before, the solution of this equation can be expressed in terms of tabulated integral:

$$\begin{aligned} V^{(1)}(\hat{x},\hat{z},t) = &- \int_{-\infty}^{\infty} dx' \int_0^l dz' \, \frac{I_0 \Omega}{4\pi D_V} \\ &\times \sum_n z_n(t) n^2 \omega^2 \mathrm{e}^{-in\omega x'} \\ &\times \{\ln[(\hat{x}-x')^2 + (\hat{z}-z')^2] - \ln[(\hat{x}-x')^2 \\ &+ (\hat{z}+z')^2]\} \\ = -\frac{I_0 \Omega}{D_V} &\begin{cases} \sum_n z_{(t)} \mathrm{e}^{-in\omega\hat{x}} \\ \quad \times [\mathrm{e}^{-n\omega\hat{z}} - 1 + \mathrm{e}^{-n\omega l} \sinh(n\omega\hat{z})], \\ l > \hat{z} > 0 \\ \sum_n z_n(t) \mathrm{e}^{-in\omega\hat{x}} [1 - \cosh(n\omega l)] \mathrm{e}^{-n\omega\hat{z}}, \\ \hat{z} > l. \end{cases} \end{aligned}$$

Then the total value for $V_I(\hat{x},\hat{z},t)$ is given by

$$\begin{aligned} V_I(\hat{x},\hat{z},t) &= V_I^{(0)}(\hat{x},\hat{z}) + V_I^{(1)}(\hat{x},\hat{z},t) \\ &= \frac{I_0 \Omega}{D_V} \begin{cases} \hat{z} - \sum_n z_n(t) \mathrm{e}^{-i\omega n\hat{x}} \\ \quad \times [\mathrm{e}^{-n\omega\hat{z}} - 1 + \mathrm{e}^{-n\omega l} \sinh(n\omega\hat{z})], \\ ll > \hat{z} > 0 \\ l - \sum_n z_n(t) \mathrm{e}^{-in\omega\hat{x}} \\ \quad \times [1 - \cosh(n\omega l)] \mathrm{e}^{-n\omega\hat{z}}, \\ \hat{z} > l. \end{cases} \end{aligned} \tag{21}$$

Using (4), (12), (17), and (21), we can find the correction to the rate of variation of the surface associated with the presence of a source of vacancies:

$$\begin{aligned} \left.\frac{\partial z}{\partial t}\right|_{z=f(x,t)} &= D_V \left.\frac{\partial c_I(x,z,t)}{\partial \mathbf{n}}\right|_{z=f(x,t)} = D_V \left.\frac{\partial V_I(\hat{x},\hat{z},t)}{\partial \mathbf{n}}\right|_{\hat{z}=0} \\ &= D_V \left(\left.\frac{\partial V_I(\hat{x},\hat{z},t)}{\partial \hat{z}}\right|_{\hat{z}=0} - f_x \left.\frac{\partial V_I(\hat{x},\hat{z},t)}{\partial \hat{x}}\right|_{\hat{z}=0} \right); \end{aligned}$$

$$\left.\frac{\partial z}{\partial t}\right|_{z=f(x,t)} = I_0\Omega + I_0\Omega \sum_n z_n n\omega \ \mathrm{e}^{-in\omega x}[1 - \mathrm{e}^{-n\omega l}].$$

Using (5), we obtain

$$\frac{dz_0(t)}{dt} = I_0 \Omega, \tag{22}$$

$$\frac{dz_n(t)}{dt} = I_0 \Omega z_n(t) n\omega [1 - \mathrm{e}^{-n\omega l}], \quad n \neq 0. \tag{23}$$

Equation (22) describes the shear of the surface without a change in its shape, while Eq. (23) describes the change in the shape of the surface. It can be seen from (23) that the presence of a source of vacancies is responsible for an enhancement of roughness ($dz_n/dt > 0$).

Analyzing the equation for V_I in the initial system of coordinates x,z, we can explain qualitatively why an external source of vacancies increases the roughness of the surface. Indeed, Eq. (11) for V_I in the initial system of coordinates has the form

$$\begin{cases} \Delta V_I(x,z,t) = -\dfrac{I_0 \Omega}{D_V} \, \delta(z - l - f(x,t)) \\ V_I(x, z = f(x), t) = 0. \end{cases}$$

Since the concentration V_I at the surface is identically equal to zero, the vacancy flux proportional to the gradient of V_I is directed along the normal to the sample surface: $\partial V_I / \partial \vec{\tau}|_{z=f(x,t)} = 0$, where $\vec{\tau}$ is the vector tangential to the surface (in analogy with the behavior of the electric field near the metal surface). It can be seen from the simple geometrical construction (Fig. 2) that vacancy flows near concavities are denser than near convexities. This means that the number of vacancies emerging at a concavity is larger than the number of vacancies at a convexity, which leads to an increase in the surface roughness.

3. A SOURCE OF INTERSTITIALS AND SMOOTHING OF ROUGHNESSES ON THE SURFACE OF A SOLID

Let us now take into account the fact that irradiation of a sample creates, in addition to vacancies, interstitial atoms whose fluxes can also lead to a change in the sample surface. In the absence of radiation, the equilibrium concentration of interstitials c_0^{at} is several orders of magnitude smaller than

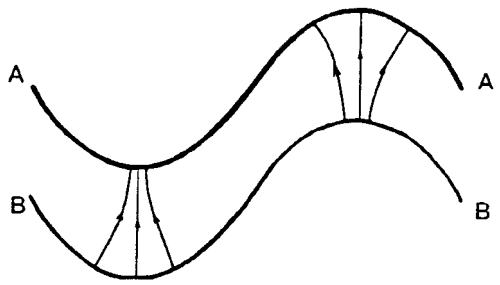

FIG. 2. Flow lines of point defects created by irradiation. The curve *AA* is the contour of the sample surface. The *BB* curve indicates the position of a delta-shaped source of point defects.

the corresponding concentration of vacancies. For this reason, we shall disregard the terms in Eq. (4) associated with the curvature, i.e., put $u_K^{\rm at}=c_K^{\rm at}=0$.

In this case, the system of equations (4) for interstitial atoms assumes the form

$$\begin{cases} D_V^{\rm at}\Delta c_{\rm at}+I_{\rm at}(x,y,z,t)=0 \\ \left.\dfrac{\partial z}{\partial t}\right|_{z=f(x,y,t)} = -\left(D_V^{\rm at}\left.\dfrac{\partial c_{\rm at}}{\partial \mathbf{n}}\right|_{z=f(x,y,t)}\right)\cos(\mathbf{n},z) \end{cases} \tag{24}$$

with the boundary condition

$$c_{\rm at}(x,y,z,t)|_{z=f(x,y,t)}=0.$$

The index ''at'' indicates that the quantities correspond to interstitial atoms. It should be noted that the main difference between the systems (24) and (4) is that the second equation in system (24) has the minus sign in front of the volume flux (while in the case of vacancies we have the plus sign).

The solution of the first equation in system (24) with a delta-shaped source leads to the result (21), and we can write for dz_n/dt in this case

$$\left.\frac{\partial z}{\partial t}\right|_{z=f(x,t)} = -I_0^{\rm at}\Omega - I_0\Omega\sum_n z_n(t)n\omega\ {\rm e}^{-in\omega x} \times[1-{\rm e}^{-n\omega l_{\rm at}}],$$

$$\frac{dz_0(t)}{dt}=-I_0^{\rm at}\Omega; \tag{25}$$

$$\frac{dz_n(t)}{dt}=-I_0^{\rm at}\Omega z_n(t)n\omega[1-{\rm e}^{-n\omega l_{\rm at}}], \quad n\neq 0. \tag{26}$$

The above equations show that, in contrast to a source of vacancies, an additional source of interstitial atoms smoothes the roughness of the surface. This is illustrated in Fig. 2. The number of interstitials emerging at a concave surface is larger than at a convex surface, which leads to the smoothing of the sample surface.

4. INTENSIFICATION OF SURFACE SMOOTHING UNDER IRRADIATION

Summing up (22) and (25), (23) and (26), we obtain the total contribution to $\dot z_n$ due to irradiation, i.e., the presence of a source of vacancies and interstitial atoms:

$$\frac{dz_0(t)}{dt}=(I_0^{\rm vac}-I_0^{\rm at})\Omega,$$

$$\frac{dz_n(t)}{dt}=-z_n(t)\Omega n\omega[I_0^{\rm at}(1-{\rm e}^{-n\omega l_{\rm at}})-I_0^{\rm vac}(1-{\rm e}^{-n\omega l_{\rm vac}})], \quad n\neq 0,$$

where the quantities with the indices ''at'' and ''vac'' correspond to interstitials and vacancies, respectively. If $I_0^{\rm at}=I_0^{\rm vac}=I_0$ (i.e., if the number of vacancies emerging at the surface is equal to the number of emerging interstitial atoms), the above equations assume the form

$$\frac{dz_0(t)}{dt}=0, \tag{27}$$

$$\frac{dz_n(t)}{dt}=-z_n(t)\Omega n\omega I_0[{\rm e}^{-n\omega l_{\rm vac}}-{\rm e}^{-n\omega l_{\rm at}}], \quad n\neq 0. \tag{28}$$

It follows from Eq. (27) that when, the sources of vacancies and interstitials are taken into account, the surface as a whole does not move. For $n\omega l_{\rm vac,at}\ll 1$, we obtain from (28)

$$\frac{dz_n(t)}{dt}=-z_n(t)\Omega n^2\omega^2 I_0[l_{\rm at}-l_{\rm vac}]. \tag{29}$$

Since an interstitial atom acquires, as a result of irradiation, an additional momentum (as compared to vacancies) directed to the bulk of the sample, $l_{\rm at}>l_{\rm vac}$, and then $\dot z_n<0$ (the surface roughness is smoothed). In order to obtain an estimate, we take $I_0\approx 10^{15}\ {\rm s}^{-1}\cdot{\rm cm}^{-2}$, $\Omega\approx 10^{-23}\ {\rm cm}^3$, $l_{\rm at}-l_{\rm vac}\approx 10^{-5}$ cm, and $\lambda\approx 10^{-3}$ cm; it follows from (29) that

$$\frac{dz_1}{dt}\approx -z_1\cdot 10^{-5}(s^{-1}),$$

which coincides in order of magnitude with the temperature contribution for $T\sim 1000$ K.[4]

When the sample is bombarded with ions, the profile of generated point defects has a Gaussian shape, and the delta-shaped source is a good approximation for a narrow distribution of defects. If sources have an arbitrary shape $I^{\rm at}(\hat z), I^{\rm vac}(\hat z)$ (e.g., in the case of bombardment with neutrons), in view of the linearity of system (11) we can easily obtain (by using the superposition principle) the following expression for smoothing rate:

$$\frac{dz_n(t)}{dt}=-z_n(t)n\omega\int_0^\infty d\xi[I^{\rm at}(\xi)-I^{\rm vac}(\xi)](1-{\rm e}^{-n\omega\xi}).$$

If $I_0^{\rm at}=\int d\xi I^{\rm at}(\xi)=I_0^{\rm vac}=\int d\xi I^{\rm vac}(\xi)$, we have

$$\frac{dz_n(t)}{dt}=-z_n(t)n\omega\int_0^\infty d\xi[I^{\rm vac}(\xi)-I^{\rm at}(\xi)]{\rm e}^{-n\omega\xi}. \tag{30}$$

In the case of bombardment with neutrons, the profile of generated point defects has a shape of a step of length d: $I(\hat z)=I_0(\Omega/d)\theta(d-\hat z)$. Using (30), we then obtain the following expression for the rate of smoothing of the surface structure (for $dn\omega\ll 1$):

$$\frac{dz_n(t)}{dt}=-z_n(t)\Omega n^2\omega^2\frac{I_0}{2}[d_{\rm at}-d_{\rm vac}]. \tag{31}$$

Taking (18), (28), and (31) into account, we obtain the general equation describing the rate of smoothing of the sample surface having already taken into account the temperature factor and irradiation, i.e.,

$$\frac{dz_n(t)}{dt}=-z_n(t)\left(D_V\frac{c_0\gamma_S\Omega}{kT}\omega^3n^3 + D_Sa\left[\frac{\beta D_V\tau_S}{a^2}\right]\frac{c_0\gamma_S\Omega}{kT}\omega^4n^4+I_0\Omega(l_{\rm at}-l_{\rm vac})\omega^2n^2\right) \tag{32}$$

for a delta-shaped profile of point defects generated by radiation, which corresponds to the bombardment with ions, or

$$\frac{dz_n(t)}{dt} = -z_n(t)\left(D_V \frac{c_0\gamma_S\Omega}{kT}\,\omega^3 n^3 + D_S a\left[\frac{\beta D_V\tau_S}{a^2}\right]\frac{c_0\gamma_S\Omega}{kT}\,\omega^4 n^4 + \frac{I_0}{2}\,\Omega(d_{\rm at} - d_{\rm vac})\omega^2 n^2\right) \quad (33)$$

in the case of a step-shaped profile of generated defects, which corresponds to neutron-type of irradiation. After integration, we have

$$\left(\frac{\lambda}{2\pi}\right)^4 \ln\frac{z_n(t)}{z_n(0)} = -\left[B_n + C_n\left(\frac{\lambda}{2\pi}\right) + P_n\left(\frac{\lambda}{2\pi}\right)^2\right]t, \quad (34)$$

where

$$B_n = D_S a\left(\frac{\beta D_V\tau_S}{a^2}\right)\frac{c_0\gamma_S\Omega}{kT}\,n^4;$$

$$C_n = D_V\frac{c_0\gamma_S\Omega}{kT}\,n^3; \quad P_n = I_0\Omega(l_{\rm at}-l_{\rm vac})n^2$$

(in the case of a delta-shaped profile of point defects generated by irradiation with ions) or $P_n = (I_0/2)\Omega(d_{\rm at} - d_{\rm vac})n^2$ (in the case of a step-shaped profile of point defects generated by neutron irradiation). In can be seen from (34) that the first two terms on the right-hand side describe the role of surface and volume diffusion in surface smoothing, while the third term describes the role of radiation. All the three terms have different dependences on λ. Thus, analyzing Eq. (34), we can determine the role of each factor in the process under investigation from the experimentally determined dependences $z_n(t)/z_n(0)$ (for different values of λ). For example, we can obtain information on the values of D_V and D_S as well as physical parameters characterizing radiation (e.g., $l_{\rm at}-l_{\rm vac}$).

Let us analyze expression (34) in greater detail for a delta-shaped profile of the sources:

$$\ln\frac{z_n(t)}{z_n(0)} = -\left(\frac{(2\pi n)^4}{\tau_S}\,\beta\,\frac{c_0\gamma_S\Omega}{kTa}\,\frac{D_S\tau_S D_V\tau_S}{\lambda^4} + \frac{(2\pi n)^3}{\tau_S}\,\frac{c_0\gamma_S\Omega}{kTa}\,\frac{D_V\tau_S a}{\lambda^3} + \frac{(2\pi n)^2}{\tau_S}\,I_0 a^2\tau_S\,\frac{(l_{\rm at}-l_{\rm vac})a}{\lambda^2}\right)t. \quad (35)$$

According to this equation, there exist two characteristic values of λ:

$$\lambda_1 \sim a\,\frac{D_S}{D_V}\,\frac{\beta D_V\tau_S}{a^2},$$

for which the first and second terms on the right-hand side of Eq. (35) become equal, and

$$\lambda_2 \sim a\,\frac{c_0\gamma_S\Omega}{kT(l_{\rm at}-l_{\rm vac})}\,\frac{1}{I_0 a^2(a^2/D_V)},$$

for which the second and third terms become equal, where

$$\frac{c_0\gamma_S\Omega}{kT(l_{\rm at}-l_{\rm vac})}$$

is the difference between the equilibrium concentration of vacancies near the surface with the curvature $K\sim(l_{\rm at}-l_{\rm vac})^{-1}$ from the equilibrium concentration near the plane surface, where $l_0 a^2(a^2/D_V)$ is the concentration of point defects emerging at the surface due to irradiation over the time $\sim a^2/D_V$. Depending on the value of λ and the relation between λ_1 and λ_2, different terms in Eq. (35) play the leading role. For example, considering that the right-hand side in Eq. (35) is a function of λ, we can observe a segment with the dependence of $1/\lambda^3$ only for a relatively low radiation intensity such that $\lambda_1<\lambda_2$, i.e.,

$$\frac{I_0 a^2\tau_S}{c_0\gamma_S\Omega[kT(l_{\rm at}-l_{\rm vac})]^{-1}} < \frac{1}{\beta}\,\frac{D_V}{D_S},$$

where $I_0 a^2\tau_S$ is the concentration of point defects emerging at the surface as a result of irradiation during the time $\sim\tau_S$. Under this condition, the transition from the dependence $1/\lambda^4$ to $1/\lambda^3$ on the right-hand side of (35) occurs at $\lambda\sim\lambda_1$, while the transition to the dependence $1/\lambda^2$ takes place for $\lambda\sim\lambda_2$. If, however, the radiation intensity is high, the second term on the right-hand side of (35) is manifested weakly, and the transition from the dependence $1/\lambda^4$ to $1/\lambda^2$ occurs almost immediately.

5. THE RATE OF SURFACE STRUCTURE SMOOTHING FOR A FINITE TIME OF ABSORPTION OF POINT DEFECTS BY THE SURFACE

In previous sections, we considered a fast absorption of point defects by the surface, when the surface concentration of the defects can be regarded as equal to its equilibrium value. In some cases, however, this approximation is insufficient, and a more exact analysis of the basis system of equations (1), (2) is required. Under the quasi-steady-state condition $l \gg \tau_{\max} = \max(\lambda^2/D_S, l^2/d_V)$ the system of equations (1), (2) can be written in the form

$$\begin{cases} D_S\Delta_S u - \nu(u-u_K) + \dfrac{D_V}{a}\left.\dfrac{\partial c}{\partial \mathbf{n}}\right|_{z=f(x,y,t)} = 0 \\ D_V\Delta c + I(x,y,z,t) = 0 \\ \left.\dfrac{\partial z}{\partial t}\right|_{z=f(x,y,t)} = a\,\nu(u-u_K)\cos(\mathbf{n},z), \end{cases} \quad (36)$$

$$D_V\left.\frac{\partial c}{\partial \mathbf{n}}\right|_{z=f(x,y,t)} = \left.\left(\frac{\beta D_V}{a}\,c - \frac{a}{\tau_S}\,u\right)\right|_{z=f(x,y,t)}. \quad (37)$$

As before, the shape of the surface will be taken in the form (5); in this case,the solution of the system of equations (36), (37) is independent of y, and the solution at the sample surface can be presented in the form of Fourier series in x:

$$\begin{cases} c|_{z=f(x,t)} = c_0 + \sum\limits_{n\neq 0} c_n(t)\mathrm{e}^{-in\omega x} \\ u(x,t) = u_0 + \sum\limits_{n\neq 0} u_n(t)\mathrm{e}^{-in\omega x} \\ \left.\dfrac{\partial c}{\partial \mathbf{n}}\right|_{z=f(x,t)} = \sum\limits_{n\neq 0} \eta_n(t)\mathrm{e}^{-in\omega x}, \end{cases} \tag{38}$$

where c_0, u_0 are equilibrium concentrations of vacancies corresponding to a plane surface, the time dependence appears only through $z_n(t)$, and the term with $n=0$ is omitted in the summation over n. Substituting (38) into (36) and (37) and considering that $\cos(\mathbf{n},z)\approx 1$ to within $(z_n/\lambda)^2$, we obtain the following system of equations:

$$\begin{cases} D_V \Delta c(x,z,t) + I(x,z,t) = 0, \\ c|_{z=f(x,t)} = c_0 + \sum\limits_{n\neq 0} c_n(t)\mathrm{e}^{-in\omega x} \\ D_S u_n(t)\omega^2 n^2 + \nu\left(u_n(t) - \dfrac{\gamma_S \Omega}{kT} u_0 z_n(t)\omega^2 n^2\right) \\ \quad - \dfrac{D_V}{a}\eta_n(t) = 0 \\ D_V \eta_n(t) = \dfrac{\beta D_V}{a} c_n(t) - \dfrac{a}{\tau_S} u_n(t) \\ \dfrac{dz_n(t)}{dt} = -aD_S u_n(t)\omega^2 n^2 + D_V \eta_n(t), \end{cases} \tag{39}$$

where $n \neq 0$, and the expression for u_K in terms of curvature $u_K = u_0 + u_0\gamma_S\Omega K/kT$, $K \approx -d^2 f(x,t)/dx^2$ have been used. In order to find the solution of the first equation of system (39), we shall use the method of substitution of variables described in Sec. 2. As a result, we obtain the following expression for the delta-shaped source of vacancies to within $(z_n/\lambda)^2$:

$$\begin{aligned} c(x,z,t) = c_0 + \sum_n c_n(t)\mathrm{e}^{-in\omega\hat{x}}\mathrm{e}^{-n\omega\hat{z}} + \frac{I_0\Omega}{D_V}\hat{z} \\ - \frac{I_0\Omega}{D_V}\sum_n z_n(t)\mathrm{e}^{-in\omega\hat{x}}[\mathrm{e}^{-n\omega\hat{z}} - 1 \\ + \mathrm{e}^{-n\omega l}\mathrm{sh}(n\omega\hat{z})], \end{aligned}$$

$$\hat{z} < l, \tag{40}$$

where $\hat{x} = x$ and $\hat{z} = z - f(x,t)$. Substituting (40) into the last equation of system (38), we obtain the equation connecting $\eta_n(t)$ and $c_n(t)$:

$$\eta_n(t) = -c_n(t)n\omega + \frac{I_0\Omega}{D_V} n\omega z_n(t)[1-\mathrm{e}^{-n\omega l}], \quad n\neq 0. \tag{41}$$

Taking into account (39), we obtain the following system of equations:

$$\begin{cases} D_S u_n(t)\omega^2 n^2 + \nu\left(u_n(t) - \dfrac{\gamma_S\Omega}{kT} u_0 z_n(t)\omega^2 n^2\right) \\ - \dfrac{D_V}{a}\eta_n(t) = 0 \\ D_V \eta_n(t) = \dfrac{\beta D_V}{a} c_n(t) - \dfrac{a}{\tau_S} u_n(t) \qquad n\neq 0. \\ \eta_n(t) = -c_n(t)n\omega + \dfrac{I_0\Omega}{D_V} n\omega z_n(t)[1-\mathrm{e}^{-n\omega l}] \\ \dfrac{dz_n(t)}{dt} = -aD_S u_n(t)\omega^2 n^2 + D_V \eta_n(t) \end{cases} \tag{42}$$

Using the first three equations of system (42), we can express $u_n(t)$, $c_n(t)$, and $\eta_n(t)$ in terms of $z_n(t)$:

$$u_n(t) = \frac{(1+n\omega a\beta^{-1})\gamma_S\Omega(kT)^{-1}u_0 z_n(t)\omega^2 n^2 + I_0\Omega(a\nu)^{-1}n\omega z_n(t)[1-\mathrm{e}^{-n\omega l}]}{(1+D_S\omega^2 n^2\nu^{-1})(1+n\omega a\beta^{-1}) + n\omega a\beta^{-1}};$$

$$c_n(t) = \frac{a^2}{\beta D_V \tau_S}\frac{\gamma_S\Omega(kT)^{-1}u_0 z_n(t)\omega^2 n^2 + I_0\Omega(a\nu)^{-1}n\omega z_n(t)[1-\mathrm{e}^{-n\omega l}]\{1+\tau_S(\nu+D_S\omega^2 n^2)\}}{(1+D_S\omega^2 n^2\nu^{-1})(1+n\omega a\beta^{-1}) + n\omega a\beta^{-1}};$$

$$\eta_n(t) = -c_n(t)n\omega + \frac{I_0\Omega}{D_V} n\omega z_n(t)[1-\mathrm{e}^{-n\omega l}], \tag{43}$$

where the relation $\tau_S = 1/\nu$ has been used. Substituting (42) into the last equation of system (41), we obtain the rate of smoothing of the surface of a solid:

$$\frac{dz_n(t)}{dt} = -z_n(t)P_1 - z_n(t)P_2 + z_n(t)P_3,$$

$$\ln\left(\frac{z_n(t)}{z_{n(0)}}\right) = -(P_1+P_2+P_3)t. \tag{44}$$

$$P_1 = \frac{\gamma_S\Omega(kT)^{-1}\omega^3 n^3 c_0 D_V}{[1+n^2(\tau_\lambda\nu)^{-1}](1+n\omega a\beta^{-1}) + n\omega a\beta^{-1}};$$

$$P_2 = \frac{\gamma_S\Omega(kT)^{-1}\beta D_V\tau_S a^{-2}\omega^4 n^4 c_0 a D_S}{1+n^2(\tau_\lambda\nu)^{-1} + [1+\beta/(n\omega a)]^{-1}};$$

$$P_3 = -\frac{I_0\Omega\omega^2 n^2 l}{[1+n^2(\tau_\lambda\nu)^{-1}](1+n\omega a\beta^{-1}) + n\omega a\beta^{-1}}, \tag{45}$$

where $\tau_\lambda = (\omega^2 D_S)^{-1} = \lambda^2/4\pi D_S$ as well as the relation $u_0 = c_0(\beta D_V \tau_S/a^2)$ as well as the condition $l\omega n \ll 1$ have been used.

Equation (44) can be easily generalized to the case when the source of interstitials is taken into account in addition to the source of vacancies: the terms P_1 and P_2 do not change in view of the smallness of the equilibrium concentration of interstitial atoms as compared to the equilibrium concentration of vacancies (see Sec. 3), while the expression for the term P_3 associated with sources of point defects assumes the form

$$P_3 = \frac{I_0 \Omega \omega^2 n^2 I_{\rm at}}{[1+n^2/(\tau_\lambda^{\rm at}\nu^{\rm at})](1+n\omega a/\beta^{\rm at})+n\omega a/\beta^{\rm at}} - \frac{I_0 \Omega \omega^2 n^2 l_{\rm vac}}{[1+n^2/(\tau_\lambda^{\rm vac}\nu^{\rm vac})](1+n\omega a/\beta^{\rm vac})+n\omega a/\beta^{\rm vac}}, \tag{46}$$

where the condition $I_0^{\rm at} = I_0^{\rm vac} = I_0$ has been used. Equations (44) and (46) show that for $\omega n a/\beta \ll 1$ and $(\tau_\lambda^{\rm vac}\nu^{\rm vac})^{-1} \ll 1$, $(\tau_\lambda^{\rm at}\nu^{\rm at})^{-1} \ll I$ the result (32) is reproduced as the zeroth approximation in these small parameters.

If sources have the form of a step (as in the case of bombardment with neutrons), in analogy with (33), expression (46) assumes the form

$$P_3 = \frac{1}{2}\frac{I_0 \Omega \omega^2 n^2 d_{\rm at}}{[1+n^2/(\tau_\lambda^{\rm at}\nu^{\rm at})](1+n\omega a/\beta^{\rm at})+n\omega a/\beta^{\rm at}} - \frac{1}{2}\frac{I_0 \Omega \omega^2 n^2 d_{\rm vac}}{[1+n^2/(\tau_\lambda^{\rm vac}\nu^{\rm vac})](1+n\omega a/\beta^{\rm vac})+n\omega a/\beta^{\rm vac}}. \tag{47}$$

For a slow absorption of point defects by the surface ($\tau_\lambda \nu \ll 1$) and for $n\omega a/\beta \ll 1$, equations (44) and (46) lead to the following expression (for a delta-shaped profile of the source of point defects):

$$\ln\left(\frac{z_n(t)}{z_n(0)}\right) = -\left[\frac{2\pi}{\lambda}\, n\, \frac{c_0 \gamma_S \Omega}{kT}\, \frac{D_V^{\rm vac}}{D_S^{\rm vac}}\, \nu^{\rm vac} + \left(\frac{2\pi}{\lambda}\right)^2 n^2\, \frac{c_0 \gamma_S \Omega}{kT}\left(\frac{\beta D_V \tau_S}{a^2}\right)\nu^{\rm vac} + I_0 \Omega\left(\frac{l_{\rm at}\nu^{\rm at}}{D_S^{\rm at}} - \frac{l_{\rm vac}\nu^{\rm vac}}{D_S^{\rm vac}}\right)\right] t. \tag{48}$$

It can be seen from (48) that for $\tau_\lambda^{\rm vac,at}\nu^{\rm vac,at} \to 0$, the surface is not smoothed ($\dot{z}_n \to 0$), which is natural since point defects emerging at the sample surface under these conditions have time to be redistributed uniformly over the entire surface before absorption.

Knowing the diffusion coefficients and other constants and analyzing Eqs.(44)–(47) as functions of λ experimentally, we can easily determine the quantities $\nu^{\rm vac,at}$ and $\beta^{\rm vac,at}$ which are difficult to measure.

Thus, we have proved that the presence of point defects created by irradiation leads to a faster smoothing of the surface of a solid. The observed effect is important for many applied problems. It should be noted that we did not take into account the effects associated with evaporation of a substance upon irradiation, which can lead to a dependence on the depth of location of a source of point defects and on the vapor pressure, or to a deterioration of the surface (increase in its roughness), or to acceleration of smoothing of the surface structure.

This research was carried out under partial financial support from the International Soros Program on Promotion of Education in Science (Grant ISSEPSPU No. 042062).

*E-mail: kfti@kfti.kharkov.ua

[1)] $c/a^3 = n_V$, $u/a^2 = n_S$, n_V and n_S are the volume and surface densities of vacancies and a is the lattice parameter.

[1] E. M. Lifshitz and V. V. Slezov, Zh. Éksp. Teor. Fiz. **35**, 479 (1958) [Sov. Phys. JETP **8**, 331 (1958)].
[2] E. M. Lifshitz, Zh. Éksp. Teor. Fiz. **44**, 1349 (1963) [Sov. Phys. JETP **17**, 909 (1963)].
[3] W. W. Mullins, J. Appl. Phys. **30**, 77 (1959).
[4] K. Hoehne and R. Sizmann, Phys. Status Solidi **5**, 577 (1971).
[5] V. V. Slezov and L. V. Tanatarov, Metallofizika **10**, 90 (1988); V. V. Slezov, *Soviet Scientific Reviews*, Sec. A (ed. by I. M. Khalatnikov), vol. 17, part 3 (1995).
[6] A. V. Pogorelov, *Differential Geometry* [in Russian], Nauka, Moscow (1969).
[7] V. S. Vladimirov, *Equations of Mathematical Physics*, Dekker, NY, 1971.

Translated by R. S. Wadhwa

New exact solutions of the Schrödinger equation with potentials of spin and soliton origin

V. V. Ulyanov, O. B. Zaslavskiĭ, and Yu. V. Vasilevskaya

Kharkov State University, 310077 Kharkov, Ukraine
(Submitted June 24, 1996)
Fiz. Nizk. Temp. **23**, 110–119 (January 1997)

New classes of exact solutions of the Schrödinger equation with simple explicit analytic expressions for potential fields, energy levels, and wave functions of stationary states are considered. The solutions are discovered with the help of new original methods elaborated in the quantum theory of spin systems. The corresponding effective potentials are compared to similar models of soliton origin. The main attention is paid to peculiar phenomena such as quasi-exact solvability, potentials with multiple and flexible profiles, fourth-order extrema, finite-band spectra and structural transformations in energy bands, and the spin–soliton analogy. © *1997 American Institute of Physics.* [S1063-777X(97)01201-2]

1. INTRODUCTION

Exact solutions of these equations have been obtained. In order to clarify the following considerations, we consider this question in greater detail . . . [1)]

I. M. Lifshits

The quantum-mechanical problem of exact solutions of the Schrödinger equation for stationary states of a particle in a potential field remains the object of interest for specialists. The number of potentials permitting a simple exact solution was scarce until recently. The situation has changed after the development of the method of inverse scattering problem (ISP) in soliton theory.[3] Finally, a noticeable breakthrough in the problem of exact solutions of the Schrödinger equation has been made[4] of when original methods in the theory of spin systems were elaborated, and several new classes of exact solutions were indicated, together with the directions of a search for such solutions. Moreover, exact solutions of the Schrödinger equation with simple explicit expressions for potentials, energy levels, and wave functions of stationary states form a solid ''golden'' foundation of the quantum theory.

The problem of exact solution naturally attracts persistent attention. We can single out three periods in which new models with exact solutions were created.

The first models with exact solutions with simple expressions for potentials, energy levels and wave functions of stationary states (harmonic oscillator, Morse, Eckart, and Peschel–Teller potentials, infinitely deep rectangular well, and delta-well) appeared at the early stage of evolution of quantum mechanics at the end of the twenties.

This stage was followed by a period of various generalizations, complications, and compositions of different models.[5,6] The search for new models with exact solutions continued, but the results were either too complicated and cumbersome, or too abstract.

The first breakthrough into the region of new simple models with exact solutions is associated with the evolution of soliton theory, since in the case of ISP, soliton formations play the role of potential fields in the time-independent Schrödinger equation for a certain quantum particle (pseudoparticle). Such soliton-type exact solutions are characterized by fixed energy levels (which are integrals of motion of the Korteweg–de Vries nonlinear evolution equation)[3] for potentials with flexible profiles, and the states of the continuous spectrum correspond to complete transparency for scattered particles.

Another trend is associated with the discovery of a basically new object in quantum mechanics, viz., quasi-exactly solvable models. The insight moment was the discovery of a whole class of new models with exact solutions in connection with the development of new methods in the theory of spin systems.[7] Spin systems in relevant effective potential fields with exact solutions of the Schrödinger equation are characterized by various potential profiles depending on the choice of the spin parameter and the values of magnetic fields and by simple formulas for the energy spectrum and corresponding wave functions.[4]

These two new trends have made a sound contribution to the depository of problems with exact solutions.

The subsequent evolution of quasi-exactly solvable models resulted in the development of a new branch of mathematics and mathematical physics (see Refs. 8 and 9 and the literature cited therein), but here we shall confine the analysis to only a few physical aspects of the problem.

Why are exact solutions of the type under investigation so important?

It should be emphasized above all that they form the basis of stationary states in quantum mechanics (see, for example, Refs. 10–14). It should also be noted that exact solutions are foundations of new problems. The significance of exact solutions as the basis of approximate methods is also worth noting. The models with exact solutions are often used in the analysis of complex phenomena for which exact regularities are unknown or in the case when basic properties are independent of the form of the potential and can be studied by disregarding complications introduced by the potential details. For example, I. M. Lifshits[1] often used the so-called separable model of perturbation. It should also be noted that new theories are usually verified on the basis of well-known special cases with exact solutions. In addition, the problems

with exact solutions serve as test examples in numerical methods.

In all probability, the problem of exact solutions will remain important in the future.

2. EFFECTIVE SPIN POTENTIALS

As a rule, it is easier to study differential equations than their discrete analogs . . .

I. M. Lifshits

Spin systems form a special class of quantum systems whose Hamiltonian contains spin operators (effective spin, pseudospin, etc.). Such systems are encountered in many fields of physics (magnetism and superconductivity, nuclear physics, and interaction between light and a substance). The description of such systems involves special methods of theoretical physics since the commutation relations for spin components differ from the Bose- and Fermi-type relations. The methods for many-particle systems have received wide application, while systems with one-spin Hamiltonian remained unnoticed. Anisotropic paramagnets can be mentioned in this connection by way of an example. This also applies to many-particle systems since collective degrees of freedom of the one-particle type can describe in some cases the motion of the system as a whole.[4]

Spin is a quantum-mechanical concept of essentially discrete origin. For this reason, the equation required for an analysis of the energy spectrum of spin systems has a matrix form. For large values of spin, this complicates the analysis of the properties of the system with the help of standard quantum-mechanical methods. It turned out, however, that we can introduce a rigorous potential description for a wide class of spin systems such that the energy spectrum of the spin system coincides with certain energy levels of a pseudoparticle moving in a potential field of a simple shape. Such an exact spin–coordinate correspondence also forms the basis for the development of various approximate methods for describing spin systems, e.g., the perturbation theory and the semiclassical approximation. It is especially important that this leads to new exact solutions of the Schrödinger equation in an appropriate coordinate system.[4]

In the approach proposed by us, the Hamiltonians constructed from generators of a certain Lie group (in particular, spin operators) are considered, and the eigenvalue and eigenvector problem is solved by using the concept of generalized[15] coherent states (including spin states). In the obtained coordinate representation, such a Hamiltonian becomes a differential operator, e.g., the Schrödinger operator with a certain effective potential energy. The potentials obtained for simple spin systems either have the form of nonlocalized wells, or are periodic. In all cases, various multiparametric potential models (both symmetric and asymmetric) exist.

Going over to specific examples, we first consider one of the simplest spin-Hamiltonians (here and below, we use dimensionless quantities without loss of generality), corresponding to the so-called easy-axis paramagnet in a transverse magnetic field:[7]

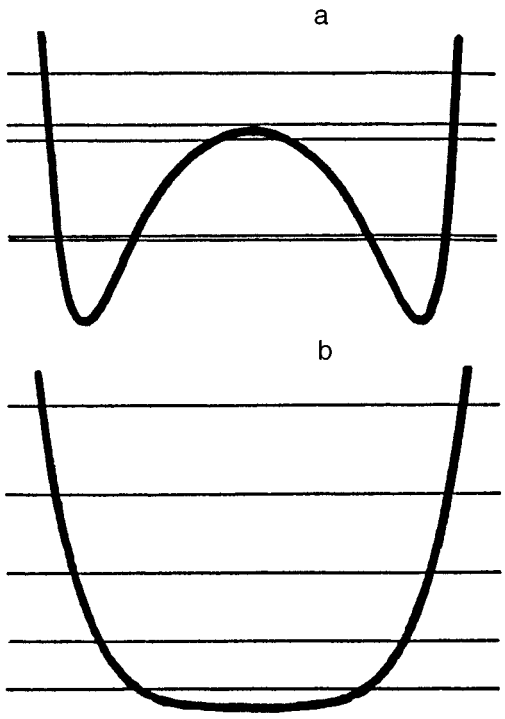

FIG. 1. Typical effective potentials (3) for $S=2$: a well with two minima for $B \ll B_0$ (a) and a well with a fourth-order minimum for $B=B_0$ (b).

$$H=-S_y^2-BS_x. \tag{1}$$

Here S_j are operators of spin components and B is proportional to the applied magnetic field. If we are dealing with the problem of determination of stationary states of such a system with the help of the above method, we arrive at the standard one-dimensional Schrödinger equation

$$\frac{d^2\psi}{d\xi^2}+[E-U_{\text{eff}}(\xi)]\psi=0 \tag{2}$$

for a pseudoparticle with a quadratic energy–momentum relation (spinon) moving in the effective potential field constructed from hyperbolic functions:

$$U_{\text{eff}}(\xi)=\frac{B^2}{4}\sinh^2\xi-B(S+1/2)\cosh\xi, \tag{3}$$

where S is the magnitude of spin and ξ can be regarded as a certain dimensionless coordinate. It turns out[4] that the energy eigenvalues E of the spin system (1) coincide with the lower $2S+1$ energy levels of a spinon in the potential field (3).

Such an approach to the description of spin systems was used for studying the low-temperature physical properties (such as energy spectrum, magnetization, susceptibility, and spin tunneling) of anisotropic paramagnets.[4,7]

For a magnetic field $B>B_0=2S+1$, the obtained nonlocalized potential (3) has the form of a solitary well, while for $B<B_0$ it is transformed into a well with two minima (Fig. 1a). It is important that for $B=B_0$ this model has the form of a well with a fourth-order minimum (Fig. 1b).

For large values of spin ($S \gg 1$), the spin–coordinate correspondence established above is convenient for studying the properties of a spin system[4] by using the standard

quantum-mechanical methods for the Schrödinger equation (2); however, for small values of spin, the eigenvalues of the spin-Hamiltonian (1) can be determined directly in the discrete spin representation as the roots of the characteristic equation. Taking into account the symmetry properties of the problem, we can obtain simple analytic expressions for energy levels and wave functions of stationary states. Various specific problems were considered in our earlier publications.[4,17,18] Here we shall concentrate attention on a special case of exact solutions for model (3) with the critical value of the parameter $B=B_0=2S+1$, when the potential has the form of a well with a fourth-order minimum:

$$U_{\text{eff}}(\xi)=-\frac{B_0^2}{2}+B_0^2\,\sinh^4\,\xi/2. \tag{4}$$

It is convenient for subsequent analysis to introduce the new coordinate $x=\xi/2$ and to measure energy from the minimum of potential (4). In the new variables, Eq. (1) assumes the form

$$\frac{d^2\psi}{dx^2}+[\varepsilon-u(x)]\psi=0, \tag{5}$$

where

$$\varepsilon=4E+2B_0^2;\;\; u(x)=Q\,\sinh^4\,x,\;\; Q=4B_0^2. \tag{6}$$

Thus, exact solutions for a potential field with a fourth-order minimum (6) exist for the values of $Q=4(2S+1)^2$, where $S=0,1/2,1,3/2,2,\ldots$. It was noted above that in this case simple explicit expressions for energy levels and wave functions of stationary states are obtained for small values of parameter S. Omitting computational details, we shall write some of these expressions (the first indices on energy correspond to the upper signs in the formulas). If $S=0$, for the ground state we have

$$\varepsilon_0=2;\;\; \psi_0(x)=A_0\,\exp\left(-\frac{1}{2}\cosh 2x\right).$$

If $S=1/2$, the ground state is supplemented with the first excited state:

$$\varepsilon_0=3;\;\; \psi_0(x)=A_0\,\exp(-\cosh 2x)\cosh x;$$

$$\varepsilon_1=11;\;\; \psi_1(x)=A_1\,\exp(-\cosh 2x)\sinh x.$$

For $S=1$, the formulas acquire radicals:

$$\varepsilon_{0,2}=16\mp 2\sqrt{37};$$

$$\psi_{0,2}(x)=A_{0,2}\,\exp\left(-\frac{3}{2}\cosh 2x\right)\left(\cosh 2x\pm\frac{\sqrt{37}\mp 1}{6}\right);$$

$$\varepsilon_1=14;\;\; \psi_1(x)=A_1\,\exp\left(-\frac{3}{2}\cosh 2x\right)\sinh 2x.$$

If $S=3/2$, we have the following expressions for four states with low-lying energy levels:

$$\varepsilon_{0,2}=19\mp 4\sqrt{13};\;\; \psi_{0,2}(x)=A_{0,2}\,\exp(-2\cosh 2x)$$
$$\times\left(\cosh 3x\pm\frac{\sqrt{13}\pm 1}{2}\cosh x\right);$$

$$\varepsilon_{1,3}=35\mp 4\sqrt{21};\;\; \psi_{1,3}(x)=A_{1,3}\,\exp(-2\cosh 2x)$$
$$\times\left(\sinh 3x\pm\frac{\sqrt{21}\mp 3}{2}\sinh x\right).$$

For $S=2$, we confine ourselves to only two energy levels. These expressions contain not only radicals:

$$\varepsilon_{2n}=\frac{130}{3}+\frac{8}{3}\sqrt{313}\,\cos\left[\frac{1}{3}\arccos\left(\frac{1765}{313^{3/2}}\right)+\varphi_n\right],$$

$$n=0,1,2;$$

$$\varphi_{0,1}=\pm\frac{2}{3}\pi,\;\; \varphi_2=0;\;\; \varepsilon_{1,2}=40\mp 2\sqrt{109}.$$

Normalization quantities are connected with modified Bessel's functions (Macdonald functions); for example, $A_0=[K_0(1)]^{-1/2}$ for $S=0$.

It should be emphasized that the number of exact solutions in the model under investigation is finite. This property is known as ''quasi-exact solvability.'' [8] In this case, stationary states with exact solutions (the multiplet $2S+1$ of energy levels) lie in the lower part of the energy spectrum (starting from the ground state), while the remaining stationary (''superspin'') states have no exact solutions.

It is well known that the power potential $u(x)=Qx^4$ (quadruple oscillator) has no exact solutions, being an important element of many essentially anharmonic systems.[19] Thus, a paradoxical situation takes place: a simpler model of the fourth-degree potential has no exact solutions, while a more complex 4-hyperbolic model (6), which is close to the former model both qualitatively and quantitatively, has such solutions. A comparison of the energy levels of these two models shows that they become closer upon an increase in the parameter S in (6), and the relative error is of the order of $S^{-2/3}$.

It is important to note that we have thus obtained another method of determining the energy levels of a quadruple oscillator with the help of model (6) with exact solutions. In a more general case, this applies to a mixed quadratic–quadruple oscillator in the exactly solvable model (3).

For the system with the spin-Hamiltonian (1) under investigation, we can also introduce a description on the basis of a certain periodic model of potential. However, we shall illustrate such a possibility for a spin system with the Hamiltonian of a more general type:

$$H=\alpha S_z^2-\beta S_y^2-BS_x, \tag{7}$$

describing, for example, a biaxial paramagnet in a magnetic field B perpendicular to anisotropy axes with the constants $\alpha,\beta\geqslant 0$. In this case, we can also arrive at the standard Schrödinger equation of type (2) with the periodic effective potential

$$U_{\text{eff}}(\xi)=\frac{W_1\,\text{sn}^2\xi-W_2\,\text{cn}\xi}{\alpha+\beta\,\text{cn}^2\xi},$$

$$W_1=\frac{B^2}{4}-a\beta S(S+1),\;\; W_2=(\alpha+\beta)B(S+1/2), \tag{8}$$

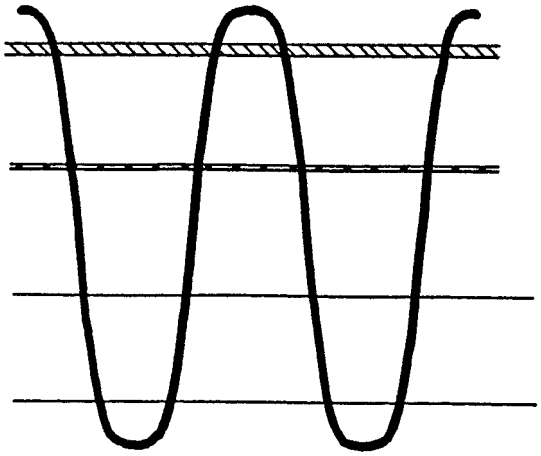

FIG. 2. Energy bands of the effective potential (8) with fourth-order extrema and exact solutions ($S=3/2$).

constructed from Jacobi elliptic functions with the modulus $k=\sqrt{\beta/(\alpha+\beta)}$.

Since the spin system (7) has an exact solution in the form of simple explicit formulas for energy levels and vectors of stationary states for small S, the Schrödinger equation also has corresponding exact solutions. The eigenvalues of the spin-Hamiltonian (7) coincide in this case with $2S+1$ extreme energy levels for the lower energy bands (alternating bottom and top of the energy band) in the potential field (8).

With increasing magnetic field B, the minima of the wells in the cells of the periodic potential are transformed from binary to solitary minima through quadruple minima, while the shape of barrier peaks changes in the reverse order. It is interesting to note that for $\alpha=\beta$ there exists a critical value of magnetic field $B_0=2\alpha\sqrt{S(S+1)}$ for which the potential has a quadruple minimum and a quadruple maximum simultaneously (Fig. 2). We shall confine our analysis to specific examples for this case putting $\alpha=1$.

We will not consider here wave functions but describe the results for some energy levels E. Energy levels are labeled according to the spin system (the first index corresponds to the upper sign in the formulas). Their positions in the energy bands of effective potentials will be indicated simultaneously.

If $S=0$, $E_0=0$ (the bottom of the ground-state band). For $S=1/2$, $E_{0,1}=\mp\sqrt{3}/2$ (extreme levels of the first gap). For $S=1$, $E_{0,2}=\mp 3$ (the bottom of the ground-state band and the bottom of the second band), while $E_1=0$ (the top of the first band). If $S=3/2$, $E_{0,1}=-3\sqrt{2}\mp\sqrt{15}/2$ (extreme energy levels of the first gap) and $E_{2,3}=3\sqrt{2}\mp\sqrt{15}/2$ (extreme energy levels of the third gap). For $S=2$, $E_{0,4}=\mp 6\sqrt{3}$ (the bottom of the ground-state band and the bottom of the first band), $E_{1,3}=\mp\sqrt{33}$ (the tops of the first and third bands), and $E_2=0$ (the bottom of the second band).

It was mentioned above that one of extreme energy levels in each of $2S+1$ lower bands is a spin level, which allows us to refer to these bands as spin bands. However, higher (''superspin'') bands deserve special attention. It was found that their arrangement has interesting peculiarities. If we assume that the value of S varies continuously, all the superspin bands match pairwise so that half the superspin gaps are closed. In addition, one of the superspin bands is converted into a spin band, and exact solutions for spin bands appear. These transformations of energy bands can be called a periodic structural spin transition.

Putting $\alpha=0$ and $\beta=1$ in formulas (8) and taking into account the limiting properties of elliptic functions, we obtain the effective potential of a uniaxial spin system (3).

If $B=0$, potential (8) for integral S can be reduced to the Lamé–Eines finite-band conoidal potential for which all superspin bands are interlocked to form a single infinitely large band.

The above examples of spin systems illustrate a remarkable feature of effective potentials, viz., the existence of exact solutions only for a part of stationary states.

In addition to stationary states having exact solutions, there exists an infinite set of stationary states lying above them (on the energy scale), for which this property is absent in general. Also, this property is manifested in the above models of potentials not for all possible values of parameters appearing in them, but only for certain values associated with integers (of multiplicity $2S+1$). The remaining values of parameters correspond to ''extraspin'' potentials with ''extraspin'' stationary states (energy levels, wave functions, and energy bands).

Thus, we have discovered a new peculiar property of quantum-mechanical systems known as quasi-exact solvability. The origin of such models of potentials usually lies in the group-theoretical properties of systems. Among other things, the basis of spin systems is the spin–coordinate correspondence.

As regards physical objects, quasi-exact solvability was apparently manifested for the first time for spin systems.

Until now, we considered only one-dimensional problems. However, a remarkable property of models with quasi-exact solutions is that exact solutions can be also obtained in two-dimensional cases even without separating the variables. This was demonstrated for the first time in Ref. 21. It is appropriate to mention here the publications[22,23] in which models with quasi-exact solutions were constructed in the presence of a magnetic field.

3. LOCALIZED SOLITON POTENTIALS

. . . We consider here a special case permitting an exact solution and leading to the simplest expressions . . .
. . . By way of an example, we depict the pattern of deformations emerging in this model . . .

I. M. Lifshits

The close relation between the nonlinear evolution Korteweg–de Vries (KdV) equation and the one-dimensional time-independent Schrödinger equation established in soliton theory[3] makes it possible to indicate new types of exact solutions of the latter equation. The authors studying solitons were interested only in the method of solution of the Cauchy problem for the KdV equation, and the Schrödinger equation played only an auxiliary role. If, however, we consider this dependence only from a purely quantum-mechanical point of view, we find that the Schrödinger equation has a multipara-

metric family of potentials for which simple exact explicit expressions exist for potential models as well as for energy levels and wave functions of stationary states. The discrete part of energy spectrum (for localized potentials) is most interesting, while the continuous part is characterized, in accordance with the steady-state theory of scattering, by zero reflection (supertransparency), i.e., its reflection coefficient is equal to zero.

It is important that, in contrast to other models with exact solutions in quantum mechanics, for which the potential is specified, and the Schrödinger equation (the eigenvalue problem) is solved (as a result, energy levels are obtained), in models of soliton origin the energy levels are defined straightaway, and each set of the levels is characterized by potentials with exact solutions. In other words, we are dealing with an inverse problem: energy levels (arranged arbitrarily) are given, and potentials are constructed for them directly. In this case, each set of energy levels is characterized not by a single fixed potential profile, but by an infinite multiparametric family of potential models.

The relation between energy levels and potentials becomes of primary importance, while wave functions of bound states play an insignificant role. This was inherited from the ISP method in which the wave functions of bound states do not appear in a complete form, but are represented only through their asymptotic parameters (they appear as so-called scattering data). However, explicit exact solutions can be obtained for wave functions also.

Choosing N successive bound stationary states arbitrarily with parameters $0<k_1<k_2< \ldots <k_N$ (soliton numbering), we can always use the formulas

$$u(x)=-2\frac{\partial^2}{\partial x^2}\ln F,$$

$$F=\det\left[\delta_{lm}+\frac{2\sqrt{k_l k_m}}{k_l+k_m}\exp(2\gamma_l)\right],$$

$$\gamma_l=k_l x-4k_l^3 t+\Delta_l+\delta_l, \tag{9}$$

$l,m=1,2,\ldots,N$ (δ_{lm} is the Kronecker delta) corresponding to the N-soliton solution of the KdV equation, to construct a family of spatially localized potentials with energy levels $E_n=-k_{N-n}^2$, $n=0,1,\ldots,N-1$ and corresponding wave functions which satisfy an equation of the type (5) and at the same time have a simpler form (we do not write them here). The quantity t having the physical meaning of time for solitons, appears here as a constant parameter which will be referred to as ''quasi-time,'' the quantities Δ_l are determined by energy parameters k_m, while phase corrections δ_l for $t=0$ affect the symmetry properties of models (9).

Potentials (9) are negative, and their deformations upon the variation of quasi-time t occur with area conservation.

In the region of continuous energy spectrum, models (9) are reflectionless in the case of particle scattering.[20]

For $t\to\pm\infty$, potential (9) consists of N different solitary reflectionless Eckart wells which are combined for $t=0$ into a composite well acquiring various symmetric properties (if all $\delta_l=0$) as well as asymmetric forms.

In the case when $N=1$, the shape of model (9) is fixed; it is a one-level reflectionless Eckart well for which we can always choose the origin of the coordinate x so that $u(x)=-2k_1^2/\cosh^2 k_1 x$. A similar profile is also obtained for $N>1$ if $\delta_l=0$, $k_l=lk_1$, $l=1,\ldots,N$ and $t=0$: $u(x)=-N(N+1)k_1^2/\cosh^2(k_1x)$.

For $N=2$, models (9) have various forms determined by two parameters: quasi-time t and the ratio of energy parameters k_2/k_1 (the phase corrections δ_1 and δ_2 can be made equal to zero by an appropriate choice of the parameters x and t). In view of the equality $u(-x)=u(x)$, for $t=0$ we have symmetric models (9) whose specific forms can be determined from the behavior of $u(x)$ near $x=0$, where a peak, i.e., double well, exists for $k_2/k_1<\sqrt{3}$ and a minimum, i.e., a solitary well (in particular, an Eckart well for $k_2/k_1=2$) exists for $k_2/k_1>\sqrt{3}$. For $k_2/k_3=\sqrt{3}$, we have a well with a flat bottom, i.e., a fourth-order minimum:

$$u(x)=-4k_1^2+\frac{8}{3}k_1^6x^4-\frac{32}{15}k_1^8x^6+O(x^8).$$

For $t\neq 0$, we have asymmetric models.

While one- and two-level potentials (9) are well known from the corresponding soliton profiles, it is not possible to speak of three-level models with four independent parameters $k_3/k_1, k_2/k_1, t$ and one of the phase corrections (say, δ_1; the other two corrections affect only the origin of x and t and can be disregarded). If $t=0$ and $\delta_1=0$, we have symmetric forms of potentials (9) which are determined above all by the behavior near $x=0$. Simple calculations show that

$$u(x)=-2(k_3^2-k_2^2+k_1^2)+2[k_3^4-(k_2^2-k_1^2)(4k_3^2-3k_2^2+k_1^2)]x^2-\frac{4}{3}\{k_3^6-(k_2^2-k_1^2)[6k_3^4-2(5k_2^2-3k_1^2)k_3^2+5k_2^4-5k_2^2k_1^2+k_1^4]\}x^4+O(x^6).$$

Thus, under the condition

$$k_3^4-4(k_2^2-k_1^2)k_3^2+(k_2^2-k_1^2)(3k_2^2-k_1^2)=0$$

we have fourth-order extrema. On the plane $(k_2^2/k_1^2, k_3^2/k_1^2)$, this corresponds to a branch of a hyperbola whose one part corresponds to fourth-order minima (Fig. 3e), and the other part to fourth-order maxima (Fig. 3f). A sixth-order minimum appears at the junction of minima and maxima (Fig. 3g) for the values $k_2/k_1=\sqrt{2+\sqrt{2}}$ and $k_3/k_1=\sqrt{3+2\sqrt{2}}$ for which the expansion of the potential has the form

$$u(x)=-2(2+\sqrt{2})k_1^2+\frac{4}{45}(17+12\sqrt{2})k_1^8x^6-\frac{8}{315}(58+41\sqrt{2})k_1^{10}x^8+O(x^{10}).$$

Figure 3 shows some typical symmetric profiles corresponding to various relations between energy parameters. For $t\neq 0$ and/or $\delta_1\neq 0$, we obtain asymmetric profiles.

Thus, a set of potentials with simple explicit exact solutions in quantum mechanics is considerably enriched with localized multiparametric models of various types. Apart

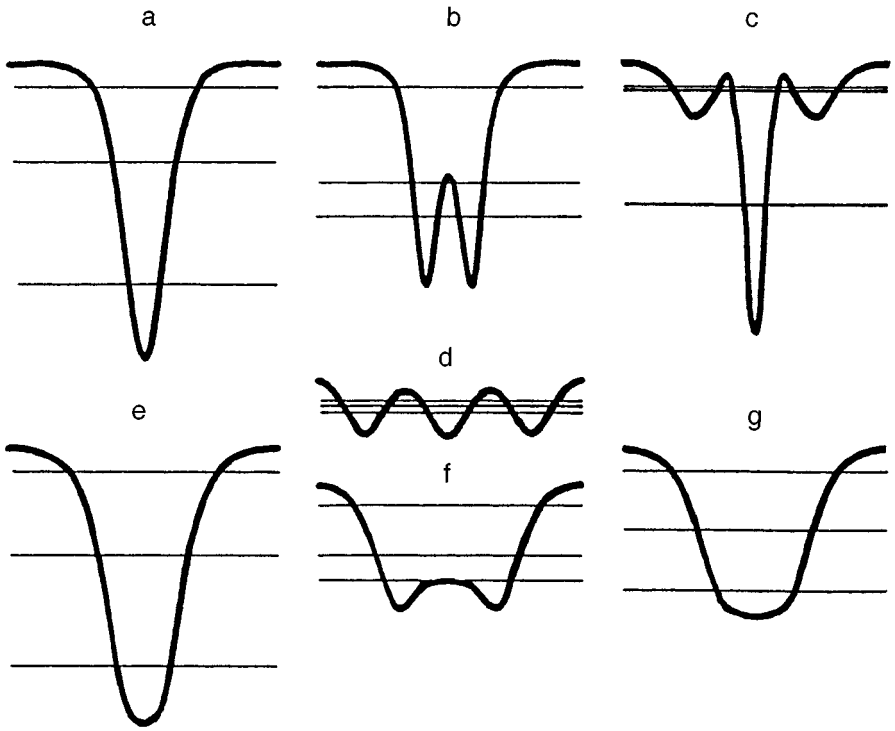

FIG. 3. Characteristic profiles of exactly solvable three-level symmetric potential models of soliton origin.

from localized potentials, periodic (or nearly periodic) finite-band potential models of soliton origin also exist,[3] but we shall not consider them here.

A transition to the quantum-mechanical situation and the formulation of the problem on exact solutions of the Schrödinger equation of the soliton origin have not been discussed in monographs on quantum mechanics from this point of view (to our knowledge, the only exception is Ref. 16).

4. SPIN–SOLITON ANALOGY

In spite of the apparent physical difference between these problems, a deep-rooted similarity can be observed.

I. M. Lifshits

The spin and soliton models of potential fields have many features in common.

For example, the spin–coordinate correspondence in the theory of spin systems is an analog of so-called Lax pair in the soliton theory, i.e., the soliton–coordinate correspondence. In both cases, there exists a certain associated Schrödinger equation which can be used for solution of the main problems, viz., the problem of energy spectrum (or stationary states in general) in spin systems and the Cauchy problem in soliton theory for nonlinear evolutionary Korteweg–de Vries equation. The role of a link in the spin–coordinate correspondence is played by the representation by coherent spin states, while a similar role in the soliton–coordinate correspondence is played by the Gelfand–Levitan–Marchenko integral transformation in the ISP method.

In both cases, the initial elements are discrete (spin and soliton), while the coordinate representations with one-dimensional potential fields and a certain energy matching which are put in correspondence to them are continuous. In the spin case, we have the spin spectrum and superspin energy levels for nonlocalized potentials, while in the soliton case we have a discrete soliton spectrum for localized potentials and a continuous ''supersoliton'' spectrum with complete transparency. Another object in common is the integral analogy: the basic parameter of potential models in the spin case is the integer $2S+1$, while an equivalent element in the soliton case is the number of solitons N.

Finite-band potentials typical of soliton periodic (and nearly periodic) models are encountered in spin models also.

The clear-cut splitting of the energy spectrum into two parts is a common feature for spin and soliton systems: in spin systems, this is manifested in the quasi-exact solvability of the models, while in soliton systems, in reflectionless (and finite-band) potentials.

We can assume that the corresponding Schrödinger equation describes the states of some pseudo-particles, viz., ''spinon'' in the case of spin systems and ''vrieson'' in the case of soliton systems. A peculiar feature of the spinon is that a spin system can be characterized by nonlocalized as well as periodic potentials. The vrieson is distinguished by the fact that both localized and periodic potentials can correspond to a soliton system.

Furthermore, potential models in both cases are characterized by a variety of forms, several parameters, and the emergence (in addition to discrete profiles mentioned above) of analogous special potential profiles: with a fourth-order minimum, with two wells and possible location of energy levels in the critical region near the peak of the barrier separating the wells, etc. In both cases, various asymmetric models as well as periodic potentials exist along with symmetric models.

The analogy can be traced in many aspects including the introduction of new concepts (quasi-exact solvability, associated Schrödinger equation, spin–coordinate correspondence, reflectionless and finite-band potentials, pseudoparticles, etc.) which play a significant role.

In both cases, the discovery of exact solutions was an auxiliary result in basic investigations aimed at the development of new methods in the theory of spin and soliton systems. The soliton boom exploited the quantum theory to a considerable extent, but advances in soliton theory have not been reflected yet in the monographs on quantum mechanics and its applications (except Ref. 16). The same situation is observed for the problem of quasi-exact solutions and new approaches to the theory of spin systems, although we are mainly dealing with a relation between two branches of quantum mechanics. Mathematical physics and pure mathematics play an increasingly large role in quasi-exactly solvable problems,[8,9] as was also the case with the soliton theory.

Thus, the two completely different systems supplying models with exact solutions to the quantum theory have a number of analogous properties. It would be interesting to find out whether these are purely formal analogies or a deep-rooted structural similarity.

5. CONCLUSIONS

The goal of this article will be reached if we have managed to clarify . . . new and unusual ideas at least partially . . .

I. M. Lifshits

Thus, we have traced briefly the emergence of new quantum-mechanical problems with exact solutions emerging in connection with systems of different origins (spin and soliton systems). The problems have enriched the family of potential models with simple explicit analytical expressions for energy levels and wave functions of stationary states in quantum mechanics. The new class of exact solutions is distinguished above all by a variety of potential profiles.

It is necessary to grasp many aspects of the new state-of-the-art in the problem of exact solutions (such as quasi-exact solvability, spin–soliton analogy, and structural transformations in the energy bands of spin models) and to generalize the obtained results, for example, on the basis of more complex spin and soliton systems.

The models under consideration have a clear physical meaning even at the stage of their creation since these models appeared not as a result of mathematical transformations, but as a result of analysis of specific systems (such as spin and soliton systems).

It should be noted in conclusion that quantum-mechanical problems with exact solutions form a new actively developing trend.

The authors are extremely grateful for the opportunity to publish this material in the journal issue dedicated to I. M. Lifshits, who was both an outstanding scientist and an excellent person. This research was made by the representatives of three generations of theorists belonging to his physics school.

We are deeply indebted to V. M. Tsukernik with whom we made the first steps in the development of new methods in the theory of spin systems.

[1)]The epigraphs to the sections of this paper are borrowed from the works by I. M. Lifshits.[1,2]

[1] I. M. Lifshits, *Selected Works. Physics of Real Crystals and Disordered Systems* [in Russian], Nauka, Moscow (1987).
[2] I. M. Lifshits, *Selected Works. Electron Theory of Metals. Physics of Polymers and Biopolymers* [in Russian], Nauka, Moscow (1994).
[3] V. E. Zakharov, S. V. Manakov, S. P. Novikov, and L. P. Pitaevskii, *Theory of Solitons. Method of Inverse Problem* [in Russian], Nauka, Moscow (1992).
[4] V. V. Ulyanov and O. B. Zaslavskii, Phys. Rept. **216**, 179 (1992).
[5] V. G. Bagrov, D. M. Gitman, I. M. Ternov *et al.*, *Exact Solutions of Relativistic Wave Equations*, Kluwer, Dordrecht (1990).
[6] I. V. Komarov, L. I. Ponomarev, and S. F. Slavyanov, *Spheroidal and Coulomb Spheroidal Functions* [in Russian], Nauka, Moscow (1976).
[7] O. B. Zaslavskii, V. V. Ulyanov, and V. M. Tsukernik, Fiz. Nizk. Temp. **9**, 511 (1983) [Sov. J. Low Temp. Phys. **9**, 259 (1983)].
[8] A. V. Turbiner, Commun. Math. Phys. **118**, 467 (1988); Contept. Math. **160**, 263 (1994).
[9] A. G. Ushveridze, Fiz. Elem. Chastits At. Yadra **20**, 1185 (1989) [Sov. J. Particles and Nuclei **20**, 504 (1989)].
[10] L. D. Landau and E. M. Lifshits, *Quantum Mechanics*, Pergamon, Oxford (1972).
[11] S. Flügge, *Practical Quantum Mechanics*, Springer, Heidelberg (1971).
[12] V. M. Galitskii, B. M. Karnakov, and V. I. Kogan, *Problems in Quantum Mechanics* [in Russian], Nauka, Moscow (1981).
[13] V. V. Ulyanov, *Problems in Quantum Mechanics and Quantum Statistics* [in Russian], Vysshaya Shkola, Kharkov (1980).
[14] V. V. Ulyanov, *Integral Methods in Quantum Mechanics* [in Russian], Vysshaya Shkola, Kharkov (1982).
[15] A. M. Perelomov, *Generalized Coherent States and Their Applications* [in Russian], Nauka, Moscow (1987).
[16] V. V. Ulyanov, *Methods of Quantum Kinetics* [in Russian], Vysshaya Shkola, Kharkov (1987).
[17] O. B. Zaslavskii and V. V. Ulyanov, Zh. Éksp. Teor. Fiz. **87**, 1724 (1984) [Sov. Phys. JETP **60**, 991 (1984)].
[18] O. B. Zaslavskii and V. V. Ulyanov, Tekh. Mekh. Phys. **71**, 260 (1987).
[19] S. I. Chan, D. Stelman, and L. E. Thompson, J. Chem. Phys. **41**, 2828 (1964).
[20] I. Kay and H. E. Moses, J. Appl. Phys. **27**, 1503 (1956).
[21] M. A. Shifman and A. V. Turbiner, Commun. Math. Phys. **126**, 347 (1989).
[22] O. B. Zaslavskii, Phys. Lett. **A190**, 373 (1994).
[23] O. B. Zaslavskii, J. Phys. **A27**, L447 (1994).

Translated by R. S. Wadhwa

Hamiltonian description of the motion of discontinuity surfaces

A. V. Kats

*Kharkov Military University, 310077 Kharkov, Ukraine**

V. M. Kontorovich

*Institute of Radioastronomy, National Academy of Sciences of the Ukraine, 310002 Kharkov,Ukraine***

(Submitted July 12, 1996)
Fiz. Nizk. Temp. **23**, 120–128 (January 1997)

The Hamiltonian description of the motion of arbitrary discontinuity surfaces is proposed on the basis of the variational principle, taking into account the conservation laws in terms of consecutively introduced volume potentials (of the Clebsch type) as constraints. Such a method of introduction of Hamiltonian variables makes it possible to generalize the known canonical variables for the interface between two media to the cases of shock waves and slip surfaces. The results are compared with the introduction of surface Hamiltonian variables through the canonical transformations of bulk Hamiltonian variables. The results permit a direct generalization to the case of magnetohydrodynamics, plasmas, superfluid liquids, and other media for which the bulk Hamiltonian equations are known. © *1997 American Institute of Physics.* [S1063-777X(97)01301-7]

INTRODUCTION

The surface canonical variables were introduced by Zakharov[1] for describing the potential motion of the free surface of a liquid. The superior convenience of this approach to nonlinear and stochastic problems renders the transition to canonical variables also suitable for describing the motion of an arbitrary discontinuity surface in hydrodynamics, i.e., for a shock wave and slip surface; this forms the subject matter of this publication. It should be noted that in Ref. 1 as well as the later publication by Sinitsyn and one of the authors,[2,3] in which the results obtained by Zakharov were generalized for the interface between two media, the surface Hamiltonian variables were virtually guessed. Later, a regular method of derivation of surface Hamiltonian variables from the bulk variables with the help of canonical transformations applicable at the interface between two incompressible liquid was proposed.[4–6] Here we describe a method of introduction of surface Hamiltonian variables as natural boundary conditions of the variational problem (see, for example, Ref. 7). Both these approaches can be used in other cases also, when bulk variables are known:[8–11] magnetohydrodynamics, plasma, magnets, superfluid liquids, etc.

At the same time, the variational principle is formulated so that it allows us to introduce Hamiltonian potentials in the bulk as Lagrangian multipliers at constraints by using the conservation laws for this purpose. It should be noted that such an approach was used fruitfully by Dirac[12] in quantum electrodynamics; for continuous media, this approach was applied in Refs. 13 and 14.[1)] Although bulk Hamiltonian variables in conventional hydrodynamics are known for a long time,[18–21] the method of their introduction is often purely directive (especially in the case of a nonpotential flow). In this connections, it remains unclear, for example, how many pairs of Clebsch variables are required[10,22,23] and whether the entropy is included in the set of such variables.[4,22] As regards the representation of the velocity for a nonpotential flow, the above-mentioned approach corresponds to Lin's variational principle[13] and leads unambiguously to Clebsch vector potentials, excluding arbitrariness in their number.

VARIATIONAL PRINCIPLE WITH CONSTRAINTS

The bulk equations must be equivalent to the equations of hydrodynamics of an ideal liquid (the notation is the same as in the monograph by Landau and Lifshitz[24]):

$$\frac{d\mathbf{v}}{dt} = -\frac{\nabla p}{\rho} + \mathbf{g}, \tag{1}$$

$$\frac{\partial \rho}{\partial t} + \operatorname{div}\, \rho\mathbf{v} = 0, \tag{2}$$

$$\frac{ds}{dt} = 0; \quad \frac{d}{dt} = \frac{\partial}{\partial t} + (\mathbf{v}\cdot\nabla). \tag{3}$$

We will derive these equations by using the variational principle for the action

$$S = \int dt L = \min,$$

$$L = \int dr \mathfrak{I}, \quad \mathfrak{I} = \frac{\rho v^2}{2} - \epsilon, \quad \epsilon \equiv \rho\varepsilon. \tag{4}$$

The internal energy can be supplemented with the terms describing potential forces such as the force of gravity written symbolically in (1) (this force will be omitted below). We introduce canonical variables by the Lagrangian method, taking into account (nonintegrable) constraints (2) and (3). The constraints introduce into the Lagrangian the required temporal and spatial derivatives

$$\mathfrak{I} \to \mathfrak{I}' = \mathfrak{I} + \mathfrak{I}_c, \quad \mathfrak{I}_c = \varphi(\dot{\rho} + \operatorname{div}(\rho\mathbf{v})) + \sigma \frac{ds}{dt}. \tag{5}$$

By varying the action (4) with the Lagrangian density (5) independently in $\mathbf{v}$, ρ and s, we obtain a representation of velocity and equations which will be written below for a more general case [Eqs. (27), (29)]. The variation in φ and σ reproduces the constraints (2) and (3). For the simples isoentropic case, we arrive at a purely potential flow

$$\frac{\delta S}{\delta \mathbf{v}} = 0 \rightarrow \mathbf{v} = \nabla \varphi.$$

The Lagrangian multiplier of the constraint expressing the law of mass conservation turned out to be just the velocity potential, which in turn indicates that the density ρ and the potential φ are canonically conjugate variables.[20] We shall use this circumstance below in a more complex case of a nonpotential flow by introducing the bulk Clebsch potentials by Lin's method[13,15] as well as the surface Hamiltonian variables.

CANONICAL VARIABLES

We can now write the equations (for the case under investigation) in the Hamiltonian form:[20]

$$\frac{\delta S}{\delta \varphi} = 0 \rightarrow \dot{\rho} = \frac{\delta H}{\delta \varphi}; \quad \frac{\delta S}{\delta \rho} = 0 \rightarrow \dot{\varphi} = -\frac{\delta H}{\delta \rho}. \tag{6}$$

In this case, generalized coordinates and momenta have the form

$$q = (\rho, \mathbf{v}), \quad \rho = (\varphi, 0).$$

The density of the Hamiltonian is given by

$$\hat{H} \equiv p\dot{q} - \mathfrak{I} = \varphi\dot{\rho} - \mathfrak{I}. \tag{7}$$

(Pay attention to the fact that generalized momentum and pressure have the same notation.) Zakharov[1] was the first to show that, in the presence of a free boundary

$$z = \zeta(\mathbf{r}_\perp, t), \quad \mathbf{r}_\perp = (x, y), \tag{8}$$

it is possible to introduce ''two-dimensional canonical variables:''

$$\zeta(\mathbf{r}_\perp, t), \quad \psi(\mathbf{r}_\perp, t) = \varphi|_\zeta, \quad (\rho = 1). \tag{9}$$

These variables were introduced by using the bulk equation for velocity potential (the Laplace equation) and the fact the coordinate conjugate to it in the given case of an incompressible liquid is constant. In view of the latter fact, the potential can be in principle excluded from the Hamiltonian. In this case, the essential variables are the elevation and the boundary value of potential, which corresponds to the existence of one (continual) physical degree of freedom.

These Hamiltonian variables were used in the analysis of disturbance on the liquid surface (see, for example, Refs. 10, 25 and 26). The investigation of internal waves in Ref. 2 (see also Ref. 3) resulted in the following generalization of the Zakharov variable to the case of the boundary between two media:

$$\zeta; \quad \psi = (\rho_1\varphi_1 - \rho_2\varphi_2) \equiv [\rho\varphi], \quad z = \zeta. \tag{10}$$

(This generalization is obvious, but it is very important for subsequent analysis.)

CANONICAL TRANSFORMATIONS

Later, a method of derivation of surface canonical variables from the bulk variables through a canonical transformations was proposed[4–6] and generalized to a nonpotential flow (see below) as well as to the case of magnetohydrodynamics. We write the bulk Hamiltonian equations for a potential flow:[20]

$$H = \int dV \left(\frac{\rho v^2}{2} + \epsilon(\rho) \right). \tag{11}$$

Following Kontorovich *et al.*,[5] we consider a transition to two incompressible liquids with densities ρ_1 and ρ_2 which are separated by the boundary $z = \zeta(\mathbf{r}_\perp, t)$. For this purpose, we introduce the representation

$$\rho = f(\zeta - z) \equiv \rho_2 + (\rho_1 - \rho_2)\Theta(\zeta - z), \tag{12}$$

where the continuous function Θ describing a real or fictitious transition layer tends to the function of discontinuity

$$\Theta(\zeta - z) \rightarrow \theta(\zeta - z) = 1(\zeta > z); \quad 0(\zeta < z). \tag{13}$$

Naturally, the change in density at a given point as a result of a limiting transition is determined only by the motion of the interface and is described by the function $\zeta(\mathbf{r}_\perp, t)$. On the other hand, the derivative is transformed to the δ function:

$$f' \rightarrow \delta(\zeta - z).$$

Let us now go over from the old coordinate ρ and momentum φ to new canonical variables through a canonical transformation,[27] assuming that the elevation ζ is a new coordinate. The generating functional

$$F(\varphi, \zeta) = \int dV \varphi f(\zeta - z)$$

is transformed as a result of a limiting transition into

$$F(\varphi, \zeta) = \int d\mathbf{r}_\perp \left\{ \int^\zeta \rho_1\varphi_1 dz + \int_\zeta \rho_2\varphi_2 dz \right\},$$

and the variation in ζ gives the new momentum

$$\psi \equiv \frac{\delta F}{\delta \zeta} = (\rho_1\varphi_1 - \rho_2\varphi_2)|_{z=\zeta} \equiv [\rho\varphi], \tag{14}$$

so that ζ and $\psi \equiv [\rho\varphi]$ form a canonically conjugate pair coinciding with (10). It will also be obtained below from the variational principle.

CLEBSCH REPRESENTATION AND VECTOR POTENTIAL

In order to describe a nonpotential flow, we shall use the Clebsch representation[18,19,21]

$$\mathbf{v} = \nabla\varphi - \frac{\lambda}{\rho}\nabla\mu - \frac{\sigma}{\rho}\nabla s, \tag{15}$$

where the Clebsch parameters satisfy the relations

$$\dot{\mu} = \frac{\delta H}{\delta \lambda} = -(\mathbf{v}\cdot\nabla)\mu, \ \frac{d\mu}{dt} = 0; \quad \dot{\lambda} = -\frac{\delta H}{\delta \mu} = -\operatorname{div} \lambda\mathbf{v}.$$

A generalization of the previous approach with the help of a canonical transformation makes it possible to introduce surface variables in this case also.[6] For example, omitting the entropy terms, we can introduce the generating function in the form

$$F(\varphi,\mu;\zeta,\tilde{\lambda})=\int dV\{\varphi f(\zeta-z)+\mu\tilde{\lambda}f(\zeta-z)\},$$

where

$$\rho=f(\zeta-z),\quad \lambda=\tilde{\lambda}f(\zeta-z).$$

As a result, we arrive at the canonical equations

$$\frac{\partial\psi}{\partial t}=-\frac{\delta H}{\delta\zeta},\quad \frac{\partial\zeta}{\partial t}=\frac{\delta H}{\delta\psi},$$

in which the role of surface Hamiltonian variables is played by ζ, $\psi\equiv[\rho(\varphi+\mu\lambda)]$.

Although the three scalar fields φ, λ, and μ may appear as sufficient for describing an arbitrary vector velocity field,[23] it is not so. The incompleteness of the description becomes obvious if we calculate the helicity (Hopf invariant)[28,29]

$$I=\int dV(\mathbf{v}\cdot\mathrm{curl}\ \mathbf{v}).$$

In the representation under investigation (for isoentropic flows), it vanishes identically, while for topologically complex flows it must differ from zero, characterizing the degree of knoffing.[9,30,31]

Apparently, the intuitive conception that the three scalar fields are sufficient becomes false in the case of emergence of special points and lines as in the above example, since in this case the conditions of expansibility into a series (forming the basis of the Darboux theorem[32] that expresses this possibility exactly) are violated.

Let us introduce vector canonical variables in analogy with the scalar potential, but using the momentum conservation as an additional constraint:

$$\mathfrak{I}'=\mathfrak{I}+\varphi(\dot{\rho}+\mathrm{div}\ \rho\mathbf{v})+A_l(\partial_t\rho v_l+\partial_m\Pi_{lm}),$$

$$\Pi_{lm}=\rho v_l v_m+p\,\delta_{lm}. \tag{16}$$

Variation in $\mathbf{v}$ leads to

$$\delta\mathbf{v}:\ v_l=\dot{A}_l+(v\,V)A_l+v_m\partial_l A_m+\partial_l\varphi.$$

Alternately, solving this equation for velocity, we obtain

$$K_{lm}v_m=\partial_l\varphi+\dot{A}_l; \tag{17}$$

$$K_{lm}=\delta_{lm}-\partial_m A_l-\partial_l A_m. \tag{18}$$

Thus, the Lagrangian multiplier of the constraint expressing the conservation of momentum proved to be velocity vector potential. In view of the nonlinearity of equations, however, its role in hydrodynamics is much less significant than in electrodynamics. The generalized coordinates and momenta have the form

$$(\rho,\rho\mathbf{v})(\text{coordinate}),\quad (\varphi,\mathbf{A})(\text{momentum}). \tag{19}$$

It can be seen that this method leads to the emergence of a vector potential as a canonical variable, which is used for describing a nonpotential flow. It should be emphasized that this representation makes it possible to describe a liquid flow in canonical variables by using a number of functions smaller than the number of vector Clebsch variables. However, the emergence of a tensor nonlinear in $\mathbf{A}$ and inverse to K_{ij} in the Hamiltonian after elimination of velocity complicates the analysis. For this reason, we shall not consider this interesting case here.

CLEBSCH VECTOR VARIABLES

In order to reduce the Lagrangian to a more conventional form, we represent the velocity as the total derivative of the displacement $\vec{\xi}$:

$$\mathbf{v}=\frac{d\vec{\xi}}{dt}=\dot{\vec{\xi}}+(\mathbf{v}\cdot\nabla)\vec{\xi}. \tag{20}$$

Using (2) as a relation with the Lagrangian multiplier $\vec{\lambda}$, we can easily derive canonical equations. We write these equations by introducing preliminarily the displacement by a radius vector

$$\vec{\xi}(\mathbf{r},\,t)=\mathbf{r}-\vec{\mu}(\mathbf{r},\,t), \tag{21}$$

for a comparison with the Clebsch variables and the variable $\vec{\mu}(\mathbf{r},t)$ instead of the field variable $\vec{\xi}(\mathbf{r},t)$. Assuming, without loss of generality, that the displacement $\vec{\xi}=0$ for $t=0$, we obtain $\vec{\mu}(\mathbf{r},t)=0$ so that $\vec{\mu}(\mathbf{r},t)$ determines Lagrangian coordinates of a liquid particle located at a point with Euler coordinates $\mathbf{r}$ at the instant t. By fixing $\vec{\mu}$, we can invert this relation and determine the trajectory $\mathbf{r}(\vec{\mu},t)$ of the particle; using (21), we can also determine the displacement:

$$\vec{\xi}=\vec{\xi}(\mathbf{r}(\vec{\mu},t),t),\quad \mathbf{r}(\vec{\mu},0)=\vec{\mu}. \tag{22}$$

This leads to the following expression for the velocity:

$$\mathbf{v}=\frac{d\vec{\xi}}{dt}\Big|_{\vec{\mu}}=\partial_t\mathbf{r}\big|_{\vec{\mu}}=\dot{\vec{\xi}}\big|_{\mathbf{r}}+(\mathbf{v}\cdot\nabla)\vec{\xi}, \tag{23}$$

so that the equation of motion for $\vec{\mu}(\mathbf{r},t)$ coincides with the transport equation for a conserved quantity: $d\vec{\mu}/dt=0$. Obviously, the variables $\vec{\lambda}$ and $\vec{\mu}$ correspond to the conserved Lagrangian coordinates ($\vec{\mu}$) and momenta ($\vec{\lambda}/\rho$) of the particle. This corresponds to the introduction of the vector set of Clebsch parameters by Lin.[4,15,22] A transition to scalar Clebsch variables is carried out in a special case when $\vec{\mu}$ and $\vec{\lambda}$ are proportional to the same constant vector.

For the bulk density of the Lagrangian function, we obtain

$$\mathfrak{I}'=\mathfrak{I}+\mathfrak{I}'_c,\quad \mathfrak{I}'_c=\varphi(\dot{\rho}+\mathrm{div}(\rho\mathbf{v}))+\vec{\lambda}\frac{d\vec{\mu}}{dt}+\sigma\frac{ds}{dt}, \tag{24}$$

which leads to the following representation of velocity:

$$\frac{\delta}{\delta\mathbf{v}}:\mathbf{v}=\nabla\varphi-\frac{\lambda_m}{\rho}\nabla\mu_m-\frac{\sigma}{\rho}\nabla s \tag{25}$$

and the bulk equations

$$\frac{\delta}{\delta\vec{\mu}}:\dot{\vec{\lambda}}+\partial_m(\vec{\lambda}v_m)=0, \tag{26}$$

$$\frac{\delta}{\delta\vec{\lambda}}:\ \frac{d\vec{\mu}}{dt}=0, \tag{27}$$

$$\frac{\delta}{\delta\rho}:\ \frac{d\varphi}{dt}+w-\frac{v^2}{2}=0, \tag{28}$$

$$\frac{\delta}{\delta\varphi}:\ \dot{\rho}+\mathrm{div}(\rho\mathbf{v})=0, \tag{29}$$

$$\frac{\delta}{\delta s}:\ \dot{\sigma}+\mathrm{div}(\sigma\mathbf{v})+\rho T=0, \tag{30}$$

$$\partial/\partial\sigma:\ ds/dt=0. \tag{31}$$

The canonical coordinates and the momenta conjugate to them are given by

$$q=(\rho,\vec{\mu},s),p=(\varphi,\vec{\lambda},\sigma). \tag{32}$$

Differentiating (25) with respect to time and using Eqs. (26)–(31), we can verify that velocity satisfies the Euler equation (1).

It should be noted that in this case the Hamiltonian has the form

$$\hat{H}=\sum\ p\dot{q}-\mathfrak{I}'\equiv\hat{\hat{H}}-\mathrm{div}(\rho\mathbf{v}\varphi), \tag{33}$$

where $\hat{\hat{H}}$ is the energy density for a liquid, and the divergence term is significant.

HAMILTONIAN VARIABLES ON A SURFACE OF DISCONTINUITIES

Let the equation of a discontinuity surface (of any type) have the form

$$R(\mathbf{r},t)=0, \tag{34}$$

where the function R is either given (defining the shape of a fixed rigid boundary), or is to be determined (e.g., of a free boundary) from the solution of the hydrodynamic problem. Under obvious constraints, it can be represented locally in the form (8): $R\to z-\zeta(\mathbf{r}_\perp,t)$. Differentiating (34) with respect to time, we obtain the kinematic condition

$$\frac{dR}{dt}=\frac{\partial R}{\partial t}+\mathbf{u}\nabla R=0, \tag{35}$$

where

$$\left.\frac{d\mathbf{r}}{dt}\right|_{R=0}=\mathbf{u},\quad \mathbf{u}=\mathbf{n}u_n$$

is the velocity of the boundary which, without loss of generality, will be assumed to be directed along the normal $\mathbf{n}=\nabla R/|\nabla R|$ (the normal is directed from medium 1 to medium 2). It should be noted that, in the case of a contact break, u_n coincides with the velocity of liquid v_n. We use the kinematic condition as a constraint in the variational principle with the Lagrangian multiplier ψ.

Since we had to integrate by parts in order to obtain the bulk equations of motion, only the surface term containing the boundary values of bulk variations is left after the vanishing of the coefficients of variations of bulk variables. Integrating by parts, taking into account the identity

$$\int dt\int d\mathbf{r}\,\frac{\partial f}{\partial t}=\int dt\left(\frac{\partial}{\partial t}\int d\mathbf{r}f-\int d\Sigma[u_nf]\right)$$

$$=-\int dt\int d\mathbf{r}_\perp[f\mathbf{u}\cdot\nabla R]$$

and adding the term corresponding to the variation of the boundary, we obtain the following expression for action:

$$(\delta S)_{\mathrm{bound}}=\int dt\int d\mathbf{r}_\perp[(\varphi\delta\rho+\vec{\lambda}\delta\vec{\mu}+\sigma\delta s)\mathbf{v}'\nabla R$$

$$+\rho\varphi(\delta\mathbf{v})\nabla R+\mathfrak{I}\delta\zeta], \tag{36}$$

where the brackets indicate the jump in the corresponding quantity at the boundary ($[X]\equiv X_1-X_2$), and we consider that $|\mathfrak{I}_c'|=0$ according to the bulk equations. It should be noted that the surface term in (36) contains only the boundary variations of generalized coordinates (including velocity) $\delta q=(\delta q)|_\zeta$ and not momenta.

The Hamiltonian potentials on both sides of a discontinuity are not independent. We modify the Lagrangian through the surface terms by introducing additional constraints on the canonical variables at the boundary so that the obtained variational equations are equivalent to hydrodynamic boundary conditions. An analysis shows that it is sufficient to impose two additional (scalar and vector) constraints at the discontinuity, i.e., to require the continuity of the mass fluxes

$$\gamma:[\rho\mathbf{v}']=0\leftrightarrow[\rho\mathbf{v}'\nabla R]=0, \tag{37}$$

and of the vector Lagrangian coordinate

$$\vec{\eta}:[\rho\vec{\mu}v_n']=0\leftrightarrow[\rho\vec{\mu}(\mathbf{v}'\nabla R)]=0, \tag{38}$$

where γ and $\vec{\eta}$ are the Lagrangian multipliers, and

$$\mathbf{v}'=\mathbf{v}-u \tag{39}$$

is the velocity of the liquid relative to the boundary. For convenience of introduction of surface variables, we shall also use the kinematic condition (35) with the multiplier ψ as a constraint.

The requirement of continuity of the flux of the quantity $\vec{\mu}$ is in accord with its interpretation as a set of Lagrangian coordinates for a liquid particle: these coordinates do not change upon an intersection of a shock wave, while in the case of a contact break or slip surface, the flux vanishes due to the equality to zero of the normal velocity component of the liquid. It should also be noted that such an interpretation of $\vec{\mu}$ is not necessary from the formal point of view, but indicates a certain choice of calibration.[2)] With such a choice, the canonical equations are written in the Euler representation, but include Lagrangian coordinates and momenta as field variables.

As a result, we obtain the surface density of the Lagrangian in the form

$$\mathfrak{I}_\Sigma=-\psi(\dot{R}+u\nabla R)+[\rho\mathbf{v}'\Gamma]\nabla R, \tag{40}$$

$$\Gamma=\gamma+\vec{\eta}\vec{\mu}. \tag{41}$$

(we have omitted the term with the surface energy; see also the next section). This leads to the following expression for the Lagrangian:

$$L \to L' + L_\Sigma, \quad L_\Sigma = \int d\mathbf{r}_\perp \mathfrak{I}_\Sigma, \tag{42}$$

so that the momentum conjugate to the elevation of the boundary is equal to ψ (10), (14): $\psi = \delta L / \delta \dot{\zeta}$. The surface density of the Hamiltonian and the Hamiltonian itself are given by

$$H_\Sigma = \psi\dot{\zeta} - \mathfrak{I}_\Sigma \equiv \psi \mathbf{u}\cdot\nabla R - [\rho \mathbf{v}'\Gamma]\nabla R, \tag{43}$$

$$H = \widetilde{H} + \int d\mathbf{r}_\perp \psi(\mathbf{u}\cdot\nabla R) - \int d\mathbf{r}[\mathrm{div}(\rho \mathbf{v}\varphi) + \mathrm{div}(\rho \mathbf{v}'\Gamma)].$$

Assuming now that the variations of the boundary values of all bulk variables on both sides of the discontinuity and the introduced surface variables (ψ, ζ, γ, and $\vec{\eta}$) are independent, we obtain the system of boundary conditions (it should be emphasized that conditions (44), (45), (47), and (48) are satisfied on both sides of the boundary):

$$\delta\rho: \quad (\varphi + \Gamma)(\mathbf{v}'\nabla R) = 0, \tag{44}$$

$$\delta\mathbf{v}: \quad \rho(\varphi + \Gamma) = 0 \to \Gamma = -\varphi, \tag{45}$$

$$\delta\mathbf{u}: \quad \psi + [\rho\Gamma] = 0 \to \psi = [\rho\varphi], \tag{46}$$

$$\delta\vec{\mu}: \quad (\vec{\lambda} + \rho\vec{\eta})(\mathbf{v}'\nabla R) = 0, \tag{47}$$

$$\delta s: \quad \sigma(\mathbf{v}'\nabla R) = 0, \tag{48}$$

$$\delta\gamma: \quad [\rho \mathbf{v}'\nabla R] = 0, \tag{49}$$

$$\delta\vec{\eta}: \quad [\rho\vec{\mu}(\mathbf{v}'\nabla R)] = 0, \tag{50}$$

$$\delta\psi: \quad \dot{\zeta} - \mathbf{u}\nabla(z - \zeta) = 0, \tag{51}$$

$$\delta\zeta: \quad [\Sigma p\dot{q} - \hat{H} + \mathrm{div}(\rho \mathbf{v}\varphi + \rho \mathbf{v}'\Gamma)] - (\partial_t\psi + \nabla_\perp(\mathbf{u}_\perp\psi)) + \alpha\nabla_\perp\left(\frac{\nabla_\perp\zeta}{(1 + (\nabla_\perp\zeta)^2)^{1/2}}\right) = 0. \tag{52}$$

Here α is the surface tension (the corresponding standard term

$$H_\alpha = \int d\Sigma\,\alpha = \alpha\int d\mathbf{r}_\perp\left[\frac{1}{(1 + (\nabla_\perp\zeta)^2)^{1/2}} - 1\right] + \mathrm{const}$$

in the Hamiltonian was omitted).

In the variation of the surface term $[\rho \mathbf{v}'\Gamma]\nabla R$ in (43), it is convenient to transform it first to the bulk term with the help of the identity

$$\int d\mathbf{r}_\perp[\mathbf{X}\nabla R] = \int d\mathbf{r}\ \mathrm{div}\mathbf{X},$$

carry out variation, and then integrate by parts.

Equations (44)–(52) form a complete system of boundary conditions in canonical variables for the bulk equations of motion (25)–(31). Before going over to an analysis of various types of discontinuities, let us analyze the obtained equations. Obviously, Eq. (44) is a consequence of (45), i.e., without loss of generality, we can assume that the boundary variation of density $\delta\rho|_{z=\zeta}$ is zero. Taking into account the remaining boundary conditions and bulk equations, we can transform Eq. (52) which has the most cumbersome form to

$$[p + \rho v_n'^2] + \alpha\nabla_\perp\left(\frac{\nabla_\perp\zeta}{(1 + (\nabla_\perp\zeta)^2)^{1/2}}\right) = 0, \tag{53}$$

which corresponds to the continuity of the normal momentum flux component. Instead of Eqs. (44)–(52), we shall henceforth consider a system of independent equations (45)–(51) and (53). It can be seen that this system can be simplified further since Eqs. (47), (48), and (50) contain the same factor $\mathbf{v}' \cdot \nabla R$, i.e., the normal velocity component of the liquid relative to the boundary. Consequently, the obtained system of equations as well as the standard hydrodynamic system[24] contains a ''fork'' corresponding to two classes of discontinuities: shock waves through which mass is transferred and $j = \rho v_n' \neq 0$, so that $v_n' \neq 0$, and the slip surfaces and contact breaks for which $v_n' = 0$.

SURFACE TENSION

In the above analysis, the surface tension was taken into account, as usual, by introducing a special surface term H_α into the energy, and accordingly into the Lagrangian function. However, using the limiting transition (13) and (15), we can obtain this term from the bulk internal energy density. For this purpose, we must take into account the dependence of the latter quantity on local density gradients (weak spatial dispersion), which reflects the finiteness of the range of intermolecular forces responsible for surface tension:

$$\rho\varepsilon(\rho) \equiv \epsilon(\rho) \to \epsilon(\rho, \nabla\rho). \tag{54}$$

Such a dependence on local gradients is naturally included in the conventional formulation of the variational principle, where the Lagrangian function is a function of not only generalized coordinates, but also of their gradients (which are essential for the formulation of field equations). On the other hand, for the equation of state this corresponds to the well-known Ornstein–Zernike approximation which is used for describing, for example, critical opalescence (we should not only come too close to the critical point at which dispersion is not weak any longer in view of a nontrivial contribution from fluctuations) and for deriving the Van der Waals equation. The isotropy of interactions indicates that the argument of energy is the square of density gradient, and hence we must retain in the limiting transition the derivative of the Θ-function describing the actial shape of the transition layer (and the form of interpolation) as a multiplier of the δ-function:

$$\nabla\rho \to (\rho_2 - \rho_1)\mathbf{n}\,\frac{d\Theta}{dR}\,\sqrt{1 + (\nabla\zeta)^2}.$$

Accordingly, we can write

$$\int dV\epsilon[\rho, (\nabla\rho)^2] \to \int dV\epsilon(\rho) + \int d\Sigma\,\alpha,$$

where

$$\alpha=\frac{\partial\epsilon}{\partial(\nabla\rho)^2}(\rho_2-\rho_1)^2\frac{d\Theta}{dR}\bigg|_{R=0}. \tag{55}$$

This leads to the estimate $\alpha\sim\epsilon a^2/l$, where a is the atomic scale and l the transition layer thickness.

HAMILTONIAN VARIABLES ON THE SHOCK WAVE SURFACE

If $v_n'\neq 0$, the mass flux through the discontinuity differs from zero and is continuous according to (49):

$$\rho v_n'=j\neq 0,\quad [j]=0. \tag{56}$$

Bearing this in mind, we obtain the following expressions on the discontinuity surface from (47), (48), and (50), respectively:

$$\vec{\eta}=-\vec{\lambda}/\rho\rightarrow[\vec{\lambda}/\rho]=0, \tag{57}$$

$$\sigma|_\zeta=0, \tag{58}$$

$$[\vec{\mu}]=0. \tag{59}$$

Consequently, the Lagrangian invariants $\vec{\mu}$ and $\vec{\lambda}/\rho$ are continuous, and the momentum σ conjugate to entropy vanishes at the boundary. The continuity of $\vec{\mu}$ leads to the continuity of Γ which, according to (45), results in the continuity of the potential φ at the boundary:

$$\Gamma\equiv\gamma+\vec{\lambda}\vec{\mu}/\rho=-\varphi\rightarrow[\varphi]=0,\quad \psi=[\rho]\varphi|_\zeta. \tag{60}$$

For the boundary values of velocity, we obtain

$$\mathbf{v}|_\zeta=\nabla\varphi-(\lambda_m\nabla\mu_m)/\rho. \tag{61}$$

Taking into account what has been said above, we arrive at the continuity of the tangential velocity component:

$$[\mathbf{v}_\tau]=[\nabla_\tau\varphi-(\lambda_m\nabla_\tau\mu_m)/\rho]=0,\quad n\nabla_\tau\equiv 0, \tag{62}$$

where the index τ denotes the tangential component. Using the obtained conditions, we can verify that the energy flux through the boundary is continuous.

Supplementing these equations with (53), we see that they describe a shock wave (in accordance with the above estimates, we must put $\alpha=0$ in the case of a gas).

It is interesting to note that the fluxes of the quantities $\vec{\mu}$, $\vec{\lambda}/\rho$, and φ are continuous:

$$[j_{\vec{\mu}}]=[j_{\vec{\lambda}/\rho}]=[j_\varphi]=0;$$

$$j_{\vec{\mu}}\equiv\rho\vec{\mu}v_n',\quad j_{\lambda/\rho}\equiv\vec{\lambda}v_n',\quad j_\varphi\equiv\rho\varphi v_n'. \tag{63}$$

HAMILTONIAN VARIABLES ON THE SLIP SURFACES AND CONTACT BREAK

Let us now suppose that $v_n'=0$ on one side of a discontinuity. Then it follows from (49) that this quantity is equal to zero on the other side of the discontinuity also, i.e., the normal velocity component of the liquid is equal to the velocity of the boundary, and the liquid does not intersect the discontinuity. Equations (44), (47), (48), and (50) in this case do not impose any additional constraints. Using relation (51), the quantity u_n can be eliminated:

$$\mathbf{v}'\nabla R=0,\quad j=0,\quad v_n=u_n, \tag{64}$$

$$\dot{\zeta}=\mathbf{v}\cdot\nabla R=v_n\sqrt{1+(\nabla_\perp\zeta)^2}, \tag{65}$$

expression (45) can be written in explicit form

$$\rho(\Gamma+\varphi)=0\rightarrow\gamma+(\varphi+\vec{\eta}\vec{\mu})|_\zeta=0, \tag{66}$$

and expression (53) can be reduced to

$$[p]+\alpha\nabla_\perp\left(\frac{\nabla_\perp\zeta}{(1+(\nabla_\perp\zeta)^2)^{1/2}}\right)=0. \tag{67}$$

Supplementing these equations with the result of (45) and (46), i.e.,

$$\psi=[\rho\varphi], \tag{68}$$

we obtain the set of the boundary conditions (64)–(68) which must be satisfied by Hamiltonian variables for $[v_n]=0$. For velocity components, we obtain

$$\mathbf{v}_\tau=\nabla_\tau\varphi-\frac{\lambda_m}{\rho}\nabla_\tau\mu_m-\frac{\sigma}{\rho}\nabla_\tau s, \tag{69}$$

$$v_n=\frac{\partial\varphi}{\partial n}-\frac{\lambda_m}{\rho}\frac{\partial\mu_m}{\partial n}-\frac{\sigma}{\rho}\frac{\partial s}{\partial n},\quad [v_n]=0. \tag{70}$$

It follows hence that the boundary conditions correspond to a slip surface for $[\mathbf{v}_\tau]\neq 0$ and to a contact break for $[\mathbf{v}_\tau]=0$.

It should be noted that the obtained equations indicate a considerable freedom in the choice of the boundary values of Hamiltonian variables, which corresponds to a gauge freedom in the choice of potentials. For all types of discontinuities, additional surface Hamiltonian variables, viz., the coordinate $\zeta(\mathbf{r}_\perp,t)$ and the momentum $\psi=[\rho\varphi]|_{z=\zeta}$ are significant. The expression for momentum derived by us differs by a canonical transformation from the symmetric expression

$$\psi=[\rho\varphi+\vec{\lambda}\vec{\mu}+\sigma s]_{z=\zeta}, \tag{71}$$

generalizing the expression given in Refs. 5 and 6 to the vector Clebsch variables.

CONCLUSIONS

The variational principle with additional constraints of various types has been used for introducing the bulk canonical variables (including Clebsch-type potentials) as well as surface canonical variables for all types of discontinuities, including shock waves which can also be regarded as a Hamiltonian system. The latter statement is not trivial on account of the dissipative nature of shock waves.

The obtained system of conditions at the surfaces of discontinuities in canonical variables is minimal and equivalent to the ordinary system of hydrodynamic conditions at discontinuities (i.e., discontinuities of mass, energy, and momentum fluxes). Owing to minimality and gauge freedom, additional constraints corresponding to a specific problem can be imposed.

A canonical transformation from the bulk to surface Hamiltonian variables has been considered; this transformation is used for the introduction of terms with surface tension.

Vector Clebsch potentials make it possible to describe the most general type of flow, and the continuity conditions

for fluxes of Hamiltonian potentials at a discontinuity indicate that the surface of discontinuity should not necessarily coincide with the equipotential surface; this makes it possible to avoid a number of difficulties in applications.

The authors are grateful to the editorial board of the jubilee issue devoted to I. M. Lifshits for the opportunity to take part in its publication. Our attitude to I. M. Lifshits whom we owe a lot can be expressed in one word: love. One of the authors (V.K.) has preserved perpetual interest in hydrodynamics since writing his degree thesis on hydrodynamics of superfluid helium under the guidance of I. M. Lifshits. The authors are grateful to participants and organizers of the Working Group ''Singular Limit for Waves with Dispersion II'' (DW2-singlimit) and the Council on Nonlinear Dynamics at the Russian Academy of Sciences for fruitful discussions and for support. V. M. Kontorovich thanks the International Soros Program for Supporting Education in Science of the International Foundation ''Revival,'' Grant SPU042029.

*E-mail: bondarew@metrog.kharkov.ua
**E-mail: rai@ira.kharkov.ua

[1)]Of course, the boundary conditions on the bulk Hamiltonian variables (and, correspondingly, the choice of the surface canonical variables) should lead to a continuity of flow of mass, energy, and momentum; however, they are not uniquely determined by these conditions. Evidently, the use of the variational principle, which is connected with the introduction of the same canonical variables would be the simplest method of obtaining such a system which, on the one hand, would be complete, and on the other, would allow the greatest gauge freedom (see notes in Refs. 15–17).

[2)]One can easily be convinced that there exists a gauge (and simultaneously canonical) transformation which transforms $\vec{\mu}$ into $\vec{\mu}+s\cdot\mathbf{b}$, where $\mathbf{b}$ is a constant vector which disrupts the natural continuity of $\vec{\mu}$. Attempts in the literature at using the entropy s in place of the scalar Clebsch variable are connected with this circumstance. From what has been noted above, the limited nature and unsuitability of such a representation are obvious, in particular, because of the differences in the boundary conditions for $\vec{\mu}$ and s.

[1] V. E. Zakharov, Prilk. Mekh. Tekhn. Fiz. **2**, 89 (1968).
[2] Yu. A. Sinitsyn and V. M. Kontorovich, in *Studies of Turbulent Structure of Ocean* [in Russian], Izd. MGI, Sevastopol' (1975).
[3] V. M. Kontorovich, Izv. Vuzov, Radiofizika **19**, 872 (1976).
[4] V. P. Goncharov and V. I. Pavlov, *Hydrodynamic Problems in Hamiltonian Description* [in Russian], Izd. MGU, Moscow (1993).
[5] V. M. Kontorovich, H. Kravchik, and V. Time, Preprint No. 158, Institute of Radiophysics and Electronics, Kharkov (1980).
[6] V. M. Kontorovich, H. Kravchik, and V. Time, in *Interaction and Self-Interaction of Waves in Nonlinear Media* [in Russian], part II, Donish, Dushanbe (1988).
[7] L. D. Landau and E. M. Lifshitz, *The Theory of Elasticity* [3rd English edition, Pergamon, 1986].
[8] V. E. Zakharov, Izv. Vuzov, Radiofizika **17**, 431 (1974).
[9] V. E. Zakharov and E. A. Kuznetsov, Soc. Sci. Review, Sect. S **4**, 167 (1984).
[10] V. E. Zakharov, V. S. L'vov, and G. Falkovich, *Kolmogorov Spectra of Turbulence. Wave Turbulence*, Springer Verlag, New York (1992).
[11] V. L. Pokrovskii and I. M. Khaloatnikov, Zh. Éksp. Teor. Fiz. **71**, 1974 (1976) [Sov. Phys. JETP **44**, 1036 (1976)].
[12] P. Dirac, *Lectures on Quantum Mechanics*, New York (1964).
[13] C. C. Lin, in Proc. Int. School of Physics, Course XXI, Acad. Press, New York (1963).
[14] L. I. Sedov, Usp. Mat. Nauk **20**, 121 (1965).
[15] V. L. Berdichevskii, *Variational Principles in the Mechanics of Continuous Medium* [in Russian], Nauka, Moscow (1983).
[16] A. N. Kraiko, Prikl. Mat. Mekh. **45**, 256 (1981).
[17] M. V. Lurie, Prikl. Mat. Mekh. **30**, 747 (1966); **33**, 602 (1969).
[18] H. Lamb, *Hydrodynamics*, Hydrodynamics, 6th ed., Cambridge, 1932.
[19] F. Stuart, Thesises, Dublin (1900) (citation according to Lamb, Ref. 18).
[20] H. Bateman, *Partial Differential Equations of Mathematical Physics*, CUP, Cambridge (1932).
[21] A. Clebsch, J. Angew. Math., Crelle **56**, No. 1 (1859).
[22] R. I. Seliger and G. B. Whitham, Proc. Roy. Soc. **A305**, 1 (1968).
[23] B. I. Davydov, Dokl. Akad. Nauk SSSR **69**, 165 (1949).
[24] L. D. Landau and E. M. Lifshits, *Hydrodynamics*, Fluid Mechanics, 2nd ed., Pergamon, Oxford, 1987.
[25] B. B. Kadomtsev and V. M. Kontorovich, Izv. Vuzov, Radiofizika **17**, 511 (1974).
[26] A. V. Kats and V. M. Kontorovich, Zh. Prikl. Mekh. Tekh. Fiz. No. 6, 97 (1974).
[27] L. D. Landau and E. M. Lifshits, *Mechanics*, Gos. Izd. Fiz. Mat. Lit., Moscow (1958) [3rd ed., Pergamon, Oxford, 1976].
[28] H. Hopf, Math. Annalen **B104**, 637 (1931).
[29] M. Steenbeck and F. Krause, Z. Naturforsch. **B21a**, 1285 (1966).
[30] H. Moffatt, J. Fluid Mech. **106**, 49 (1981).
[31] E. A. Kuznetsov and A. V. Mikhailov, Phys. Lett. **77A**, 37 (1980).
[32] P. K. Rashevskii, *Geometrical Theory of Partial Differential Equations* [in Russian], Gostekhizdat, Moscow (1947).

Translated by R. S. Wadhwa