Impurity states in substances with narrow energy gaps

L. A. Fal'kovskiĭ

L. D. Landau Institute of Theoretical Physics, USSR Academy of Sciences
(Submitted November 26, 1974)
Zh. Eksp. Teor. Fiz. **68**, 1529–1538 (April 1975)

Impurity states are examined in substances with a narrow energy gap and a large anisotropy of the carrier mass, for which the adiabatic approximation can be used. In the main order of approximation with respect to the mass ratio, the bound states have a Dirac type spectrum with half-integral angular momentum. The corrections are of power-law type, with the exponent depending on the coupling constant. The shift of the critical value of the coupling constant, due to deviation of the potential from Coulomb form at small distances, is found by the WKB method.

PACS numbers: 71.55.

In recent years there is increasing interest in substances with a narrow energy gap. An example of such substances is alloys of bismuth with antimony. At a point L of the reciprocal lattice of such a substance there is an energy gap ϵ_g, whose size in pure bismuth is about 100 K and changes appreciably as antimony is added. The theory of semimetals [1] proposed by Abrikosov and the present writer can be applied to such substances. Abrikosov has shown [2] that the gap can become closed. If the size of the gap is small in comparison with the distance γ to the farther bands the rather complicated spectrum obtained in ref. [1] can be simplified, and takes the following form:

$$\varepsilon=\frac{k_z^2}{4}(M_1^{-1}-M_2^{-1})\pm\left\{\left[\frac{\varepsilon_g}{2}+\frac{k_z^2}{4}(M_1^{-1}+M_2^{-1})\right]^2+k_x^2v_x^2+k_y^2v_y^2\right\}^{1/2}, \quad (1)$$

where v_x and v_y are constants of the order of the usual speeds of electrons in metals, $v \sim 5\times 10^7$ cm/sec, and M_i are positive masses of the order of γ/v^2. The quantity γ is the small parameter of the theory [1]; it is small because the lattice of bismuth differs only slightly from simple cubic form. Experiment gives for γ a value of some tens of electron volts while the masses M_i are close to the order of magnitude of the mass of a free electron.

For positive values of the gap width ϵ_g the two bands described by Eq. (1) do not intersect; this is the so-called direct position of the bands, and the expansion of Eq. (1) near the minimum of the conduction band is of the usual quadratic form

$$\varepsilon=\frac{\varepsilon_g}{2}+\frac{k_x^2}{2m_x}+\frac{k_y^2}{2m_y}+\frac{k_z^2}{2M_1}, \quad (2)$$

where $m_x=\epsilon_g/2v_x^2$, $m_y=\epsilon_g/2v_y^2$. It can be seen from Eq. (2) that as ϵ_g is decreased the small masses m_x and m_y become smaller, but the large mass M_1 does not change. This result agrees with experiment, [3] which gives for the ratio of the longitudinal mass M to the transverse masses m_x, m_y a value $\mu \sim 10^2$–10^3 depending on the size of the gap. The expansion of the spectrum in powers of $\mathbf{k}$ near the maximum of the valence band is obtained from Eq. (2) by changing the sign of ϵ_g and replacing M_1 with $-M_2$.

For negative ϵ_g (inverted position of the bands) $\mathbf{k}=0$ is a saddle point and there is no energy gap in the spectrum, but the density of states is small in the range $\epsilon < |\epsilon_g|$. [4]

In the present paper we consider the influence of the electrostatic field of an impurity introduced into the alloy on the carriers, whose spectrum is described by Eq. (1). The corresponding effective Hamiltonian, the critical-energy operator

$$H_{\alpha\beta}(\mathbf{r})=\begin{vmatrix}\varepsilon_g/2+\hat{k}_z^2/2M_1 & iv_x\hat{k}_x+v_y\hat{k}_y\\ -iv_x\hat{k}_x+v_y\hat{k}_y & -\varepsilon_g/2-\hat{k}_z^2/2M_2\end{vmatrix} \quad (3)$$

was found in a paper by Beneslavskiĭ and the writer [4] by eliminating the farther bands from the Hamiltonian derived in [1]; $\hat{\mathbf{k}}$ is the operator of differentiation, $\hat{\mathbf{k}}=-i\partial/\partial\mathbf{r}$.[1)]

The interaction with the impurity

$$V(r)=-Ze^2/\kappa r, \quad R<r<r_d, \quad (4)$$

is determined by the dielectric constant κ and the difference of valence Z of the impurity. From considerations of symmetry, two principal values of κ are equal; there are three points L in the Brillouin zone, and the third value differs from the two by not more than 15 percent. We shall neglect this difference, and also the difference, of the same order, between the values of v_x and v_y, denoting their common value by v. The range of distances in which the potential can be treated as of the Coulomb form is limited. At large distances $r > r_d$ the Debye screening manifests itself; here r_d is the Debye radius. In pure bismuth at low temperatures $r_d \sim 10^{-6}$ cm; in alloys r_d depends on the carrier concentration. At small distances $r < R \sim v/\gamma$ the potential differs from the Coulomb form owing to interband transitions to levels at energy differences of the order of γ. [5] At distances $r < R$ the potential varies according to the law

$$V(r)=-\frac{Ze^2}{r}e^{-r/r_0}, \quad r_0\ll r<R, \quad (5)$$

which describes the Debye screening in a "good" metal; r_0 is of atomic size. We note that because γ is small one is considering in Eq. (5) distances of macroscopic scale.

As will be seen from what follows, the interaction with the impurity is characterized by the dimensionless constant $\alpha = Ze^2/\kappa v$. For the parameters of pure bismuth $\alpha \sim 0.1$ for Z = 1. For small α the impurity levels are shallow; they lie in the energy gap near the extrema of the bands. For sizable α—and we shall not exclude this possibility—there occur so-called deep levels, for which the ionization energy is comparable with the gap. The radius of a bound state in the xy plane is $v/|\alpha\epsilon_g|$; it is assumed that this quantity lies in the interval from R to r_d described by the Coulomb law (4). For $|\alpha| > \frac{1}{2}$ there is "collapse to the center" like the instability for nuclei with $Z > 137$ in the Dirac equation. In this case, as is well known (cf., e.g., [6]), the decisive region is

that of small distances, in which the potential differs from the Coulomb law because of the finite size of the nucleus. In a potential cut off at small distances bound states also exist for somewhat larger values of Z. But for some value α_{cr} a bound state level drops out into a continuous spectrum, which leads to instability of nuclei with $Z > Z_{cr}$.

Near r = R the potential (5) changes much more rapidly than the Coulomb potential (4). For this reason, as will be shown later, a potential barrier appears at $r \lesssim R$, and the quantity R plays the role of a cut-off parameter for the Coulomb potential (4). The critical value of α at which an impurity level drops into the continuous spectrum is somewhat different from $\frac{1}{2}$. Impurities with $|\alpha| > \alpha_{cr}$ are not very effective; they behave like neutral structures of rather small radius.

The main qualitative features of deep levels in semiconductors have been considered by Keldysh, [7] who studied the example of the Kane model. In particular, the ground state of this complex spectrum was found by means of a variational method. The small parameter which we have at our disposal, the mass ratio $1/\mu$, allows us to use the adiabatic approximation. It turns out that in zeroth order in $1/\mu$ the problem of impurity levels in a Coulomb field has an exact solution. The first-order correction is determined by the equation of one-dimensional motion of a heavy mass M, depending on the coordinates and on the interaction with the impurity. An analysis of this equation enables us to ascertain in what sorts of cases the zeroth-order spectrum is discrete, continuous, or quasidiscrete. We shall construct the adiabatic approximation for a system of two equations describing two close-together bands, and then consider two separated motions: a fast two-dimensional motion and a slow one-dimensional motion. The last section is devoted to "collapse to the center."

THE ADIABATIC APPROXIMATION IN THE TWO-BAND MODEL

The equations describing impurity states near the point L are of the form

$$\{H_{\alpha\beta}(r)+[V(r)-\varepsilon]\delta_{\alpha\beta}\}\Phi^{(\beta)}(r)=0, \tag{6}$$

where H(r) and V(r) are given by the expressions (3)–(5).

Let us separate out the fast motion in the variables ρ = (x, y). We write the free Hamiltonian (3) of Eq. (6) in the form

$$H(r)=H^{(0)}(\rho)+T(z), \tag{7}$$

where

$$H^{(0)}(\rho)=\lim_{M_i\to\infty} H(r), \quad T_{\alpha\beta}(z)=(2M)_{\alpha\beta}^{-1}\hat{k}_z^2,$$

and the matrix $M^{-1}_{\alpha\beta}$ is diagonal: $M_{11}^{-1} = M_1^{-1}$, $M_{22}^{-1} = M_2^{-1}$.

We shall look for the solution of Eq. (6) in the form of an expansion

$$\Phi^{(\alpha)}(\rho, z)=\sum_n u_n(z)\psi_{nz}^{(\alpha)}(\rho) \tag{8}$$

in terms of the eigenfunctions $\psi_{nz}^{(\alpha)}(\rho)$ of the eigenvalue problem of the operator $H^{(0)}(\rho) + V(\rho, z)$:

$$\{H_{\alpha\beta}^{(0)}(\rho)+[V(\rho, z)-\lambda_n(z)]\delta_{\alpha\beta}\}\psi_{nz}^{(\beta)}(\rho)=0. \tag{9}$$

whose eigenvalues are $\lambda_n(z)$. The variable z on which the potential $V(\rho, z)$ depends occurs in Eq. (9) as a parameter. Substituting Eq. (8) in Eq. (6) and using the orthogonality of the $\psi_{nz}^{(\alpha)}(\rho)$ for any fixed z,

$$\langle\alpha n'|\alpha n\rangle=\sum_\alpha\int d^2\rho\,\bar{\psi}_{n'z}^{(\alpha)}(\rho)\psi_{nz}^{(\alpha)}(\rho)=\delta_{nn'},$$

we get the equation for the determination of $u_n(z)$ and ϵ:

$$\begin{aligned}[\langle\alpha n|\beta n\rangle T_{\alpha\beta}(z)+\lambda_n(z)-\varepsilon]u_n(z)&=-2^{-1}u_n(z)\langle\alpha n|M_{\alpha\beta}^{-1}\hat{k}_z^2|\beta n\rangle\\ -(\hat{k}_z u_n(z))\langle\alpha n|M_{\alpha\beta}^{-1}\hat{k}_z|\beta n\rangle&-\sum_{n'}{}'\langle\alpha n|T_{\alpha\beta}(z)|\beta n'\rangle u_{n'}(z).\end{aligned} \tag{10}$$

There is summation over the repeated indices α, β, but not over n.

As in the ordinary adiabatic approximation, the potential energy $\lambda_n(z)$ of the slow motion (10) is the average energy of the fast motion for a fixed position of the slow particle, which is here represented by the projection of the motion on the z axis. The first term in the square brackets on the left-hand side of Eq. (10), the kinetic energy of the slow motion, can be written in the form

$$\langle\alpha n|\beta n\rangle T_{\alpha\beta}(z)=[2M_n(z)]^{-1}\hat{k}_z^2, \tag{11}$$

where the coordinate-dependent reduced mass $M_n(z)$ of the slow motion is given by the expression

$$\frac{1}{M_n(z)}=\int d^2\rho\left[\frac{|\psi_{nz}^{(1)}(\rho)|^2}{M_1}-\frac{|\psi_{nz}^{(2)}(\rho)|^2}{M_2}\right]. \tag{12}$$

In the ordinary adiabatic approximation there is no α, β matrix structure, the coefficient of T given as an angle brackets in Eq. (11) is equal to unity owing to normalization, and the mass of the heavy particle depends neither on the coordinates nor on the interaction. We get this result in the case of shallow levels, when only one term survives in the brackets in Eq. (12), the first for a donor impurity and the second for an acceptor; the omitted term is proportional to the square of the coupling constant α. The reduced mass $M_n(z)$ is not definite in sign, as it should be, because Eq. (6) simultaneously describes an electron in the conduction band and a hole in the valence band.

The first two terms in the right member of Eq. (10) are of the usual form for the adiabatic approximation, and are small to the extent that the characteristic distances z of the slow motion are small relative to the characteristic distances q^{-1} of the fast motion. The relative order of magnitude of the last term in the right member is $(\lambda_n - \epsilon)/(\lambda_n - \lambda_{n+1})$. The condition

$$zq\ll 1, \quad \left|\frac{\lambda_n-\varepsilon}{\lambda_n-\lambda_{n+1}}\right|\ll 1 \tag{13}$$

for the right-hand side of Eq. (10) to be small can be satisfied, as we shall see when we find the behavior of $\lambda_n(z)$, if the mass ratio is sufficiently large and the state of the slow motion is not a high excited state. Equation (10) finally takes the form

$$\{[2M_n(z)]^{-1}\hat{k}_z^2+\lambda_n(z)-\varepsilon\}u_n(z)=0. \tag{14}$$

We are interested in the case of large mass $M_n(z)$. Then the z motion, if it is finite, occurs near the equilibrium point $z = z_0$ at which $\lambda_n(z)$ has an extremum. The amplitude of the deviation $z - z_0$, and also the deviation of ϵ from $\lambda_n(z)$, are small for large $M_n(z)$. Accordingly, to determine these quantities we need to know the position of the extremum and the behavior of $\lambda_n(z)$ near it. Starting from the form (4) of the potential $V(\rho, z)$, which for given ρ has its extremum at z = 0, we make the assumption, which is confirmed by the subsequent calculation, that the extremum of $\lambda_n(z)$ is at z = 0. Solving Eq. (9) for z = 0, we find $\lambda_n(0)$, i.e., $\epsilon = \lambda_n(0)$ in zeroth approximation in the mass ratio μ^{-1}. When we then calculate by perturbation theory the expansions of $\lambda_n(z)$ and

$M_n(z)$ near z = 0 and solve Eq. (14), we find the corrections in powers of μ^{-1} to $\lambda_n(0)$ in the expansion of ϵ. It may turn out that already at small z the z motion is infinite and Eq. (14) has no discrete levels. An analysis of Eq. (14) is necessary for this reason and also to find out in what cases the spectrum of Eq. (9) at z = 0 corresponds to discrete levels of the original Eq. (6). We now proceed with this program.

THE SPECTRUM OF THE FAST MOTION

Let us find at z = 0 the solution of Eq. (9), for which the free Hamiltonian is

$$H^{(0)}(\rho)=\begin{vmatrix} \varepsilon_g/2 & v(i\hat{k}_x+\hat{k}_y) \\ v(-i\hat{k}_x+\hat{k}_y) & -\varepsilon_g/2 \end{vmatrix}$$

and the interaction is taken in the Coulomb form

$$V(\rho,z)=-\frac{Ze^2}{\varkappa(\rho^2+z^2)^{1/2}}.$$

We change to cylindrical coordinates ρ and φ = arc tan (y/x). Calculating

$$\hat{k}_y\pm i\hat{k}_x=e^{\mp i\varphi}\left(\pm\frac{\partial}{\partial\rho}-\frac{i}{\rho}\frac{\partial}{\partial\varphi}\right),$$

we see that the solution of Eq. (9) is to be looked for in the form

$$\psi_0^{(\alpha)}(\rho,\varphi)=\frac{e^{il\varphi}}{(2\pi)^{1/2}}\begin{vmatrix} \psi^{(1)}(\rho) \\ \psi^{(2)}(\rho)\,e^{i\varphi} \end{vmatrix} \tag{15}$$

with integer values of the angular momentum component l; here the subscript zero means z = 0, and we omit the index n denoting a set of quantum numbers.

Substituting Eq. (15) in Eq. (9), we get

$$\begin{aligned} &\left(\frac{\varepsilon_g}{2}+V(\rho,0)-\lambda\right)\psi^{(1)}(\rho)+v\left(\frac{\partial}{\partial\rho}+\frac{l+1}{\rho}\right)\psi^{(2)}(\rho)=0,\\ &-v\left(\frac{\partial}{\partial\rho}-\frac{l}{\rho}\right)\psi^{(1)}(\rho)+\left(-\frac{\varepsilon_g}{2}+V(\rho,0)-\lambda\right)\psi^{(2)}(\rho)=0, \end{aligned} \tag{16}$$

where $V(\rho, 0) = -Ze^2/\kappa\rho$. With the substitution

$$\psi^{(\alpha)}(\rho)=\rho^{-1/2}\chi^{(\alpha)}(\rho) \tag{17}$$

the system (16) reduces to the system of equations for the radial functions of the Dirac equation in a Coulomb field.[8] The difference between the spectrum of our problem,

$$\lambda_n(0)=\frac{|\varepsilon_g|}{2}\left\{1+\frac{\alpha^2}{[n_r+((l+1/2)^2-\alpha^2)^{1/2}]^2}\right\}^{-1/2}\operatorname{sign}\alpha,\quad \alpha=\frac{Ze^2}{\varkappa v}, \tag{18}$$

and the corresponding Dirac spectrum is that here $l + 1/2$ takes half-integral values. For $\alpha \ll 1$ there are hydrogenlike levels:

$$\lambda_n(0)=\frac{|\varepsilon_g|}{2}\left[1-\frac{\alpha^2}{2(n_r+|l+1/2|)^2}\right]\operatorname{sign}\alpha, \tag{19}$$

whose ground state is deeper than that of the hydrogen atom by a factor 4.

The sign of the levels (18), (19) depends on the sign of the impurity charge. For example, for an acceptor impurity $\alpha < 0$ and the discrete levels are located in the lower half of the energy gap.

We need the expression for the wave function

$$\begin{aligned} \chi^{(1,2)}&=\frac{\rho^\beta e^{-\rho/2}}{\Gamma(2\beta+1)}\left[\frac{\Gamma(2\beta+n_r+1)\,(\varepsilon_g/2\pm\lambda)\,qv}{n_r!\alpha\varepsilon_g^2(l+1/2+\alpha\varepsilon_g/2qv)}\right]^{1/2}\\ &\times[\mp n,F(1-n_r,2\beta+1,\rho)+(l+1/2+\alpha\varepsilon_g/2qv)F(-n_r,2\beta+1,\rho)]; \end{aligned} \tag{20}$$

$$\beta=[(l+1/2)^2-\alpha^2]^{1/2},\quad q=(\varepsilon_g^2/4-\lambda^2)^{1/2}/v,\quad \varepsilon_g>0;$$

$F(\alpha, \beta, \rho)$ is the confluent hypergeometric function, $\Gamma(\alpha)$ the gamma function; with a change of sign of ϵ_g there is a change of the sign of the function $\chi^{(1)}$. In the case when the radial quantum number $n_r = 0$ there is a restriction on the possible values of l: for $\alpha\epsilon_g > 0$, l is nonnegative, and for $\alpha\epsilon_g < 0$, l is negative. The ground state has quantum numbers $n_r = 0$, $|l + 1/2| = 1/2$, and the following energy and characteristic momentum:

$$\lambda_0(0)=|\varepsilon_g|\,(1/4-\alpha^2)^{1/2}\operatorname{sign}\alpha,\quad q_0=|\alpha\varepsilon_g|/v. \tag{21}$$

The characteristic distance q^{-1} is usually large compared with R, and the actual form of the potential (4) at small distances has little effect on the positions of the levels for $\alpha < 1/2$. For $\alpha > 1/2$, however, the quantity β becomes imaginary for some values of l, and so-called "collapse to the center" occurs. In particular this is expressed in the fact that the function (20) for the ground state oscillates, in contradiction with its definition. In this case it is necessary to take into account the departure of the potential from the Coulomb form at small distances. We shall return to this question after examining the slow motion for $\alpha < 1/2$.

THE SLOW MOTION

To calculate the $1/\mu$ correction to the spectrum (18) we go back to Eq. (14). The potential energy $\lambda_n(z)$ and the reduced mass $M_n(z)$ are calculated by means of the solution of Eq. (9) with the use of Eq. (12). It is easy to find $\lambda_n(z)$ for small z by treating the deviation of $V(\rho, z)$ from the previously used value $V(\rho, 0)$ by perturbation theory:

$$\lambda_n(z)-\chi_n(0)=\langle\alpha n|V(\rho,z)-V(\rho,0)|\alpha n\rangle.$$

Using Eqs. (15) and (17), we find from this for the ground state

$$\lambda_0(z)-\lambda_0(0)=-2\alpha|\alpha\varepsilon_g|\int d\rho|\chi^{(1)}|^2[(\rho^2+z^2)^{-1/2}-\rho^{-1}], \tag{22}$$

where ρ and z are in dimensionless units $(2q_0)^{-1}$. Substituting the expression (20) in Eq. (22), we can verify that for $z \ll 1$ the important region in the integral is $\rho \sim z$, and $e^{-\rho}$ is to be replaced with unity. We get

$$\lambda_0(z)-\lambda_0(0)=-\frac{2\alpha|\alpha\varepsilon_g|}{\Gamma(2\beta+1)}\int_0^\infty\rho^{2\beta-1}\left(\frac{\rho}{(\rho^2+z^2)^{1/2}}-1\right)d\rho=\frac{\alpha|\alpha\varepsilon_g|\Gamma(1-\beta)}{2^{2\beta}\beta\Gamma(1+\beta)}|z|^{2\beta}. \tag{23}$$

Since the integral (22) is determined by the behavior of χ at small ρ, the power-law exponent 2β does not change when we go to excited states.

Consider $M_n(z)$, given by Eq. (12). For the ground state and at z = 0 we find from Eq. (20)

$$M_0(0)=\bar{M}\varepsilon_g/[\lambda_0(0)-\bar{\lambda}], \tag{24}$$

where we have introduced the notations

$$\bar{M}=\frac{M_1M_2}{M_1+M_2},\quad \bar{\lambda}=\frac{\varepsilon_g}{2}\frac{M_1-M_2}{M_1+M_2}.$$

If the coupling constant α is such that the position of the level is close to the point of mass reversal $\bar{\lambda}$, it is necessary to calculate the dependence of $M_n(z)$ on the coordinate. For small z this can be done with perturbation theory, like the way the dependence of $\lambda_n(z)$ was found:

$$M_0^{-1}(z)-M_0^{-1}(0)=b|z|^{2\beta}/\beta\bar{M}. \tag{25}$$

The expression for b, which depends on α only weakly, is a sum over excited states n, which has not been calculated in explicit form.

The solution of Eq. (14) with λ and M in the forms (23)–(25) is not known. The spectrum and the wave function u(z) can be written out explicitly for $|\alpha| \ll 1$, when $\beta = 1/2$. In the case of inverted position of the bands,

$\epsilon_g < 0$, there are no discrete levels, since the z motion is infinite. If $\epsilon_g > 0$ and $0 < \alpha \ll 1$, we have $M_n(z) = M_1$, and Eq. (14), written in the dimensionless units

$$\mu = M_1 v^2/\varepsilon_g, \quad E = [\varepsilon - \lambda_0(0)]/2\alpha^2\varepsilon_g, \tag{26}$$

takes the form

$$\left(-\frac{1}{\mu}\frac{d^2}{dz^2} + |z| - E\right) u(z) = 0. \tag{27}$$

Its solution can be expressed in terms of the Airy function:

$$u(z) = c\,\mathrm{Ai}\,[\mu^{1/3}(z-E)], \quad z \geqslant 0.$$

The energies of states even in z are found from the condition that the first derivation is continuous at z = 0:

$$\mathrm{Ai}'(-\mu^{1/3}E) = 0,$$

which has as its first root $\mu^{1/3}E = 1.02$. From this and Eqs. (19) and (22) we find the following result for the ground state:

$$\varepsilon = \frac{\varepsilon_g}{2}(1 - 2\alpha^2 + 4.08\alpha^2\mu^{-1/3}), \tag{28}$$

which agrees with the result of Kohn and Luttinger for the shallow levels in silicon.[9] The ground state of an acceptor impurity is obtained from (28) by changing the sign of the parentheses and, in accordance with Eq. (24), replacing M_1 with M_2 in the definition of μ, Eq. (26).

If $\alpha \lesssim 1/2$ but is not small, the energies and radii of deep states can be estimated by means of Eqs. (23)–(25). We first discuss the possibilities that exist here. It is clear from physical considerations that discrete levels can exist only for the direct position of the bands, $\epsilon_g > 0$; their energies $|\epsilon| < \epsilon_g/2$. This can be verified by examining the asymptotic behavior at large distances of the wave function $\Phi^{(\alpha)}(r)$ of Eq. (6). For $r \to \infty$, where $V(r) \to 0$, the function for a bound state must be decreasing. This is possible only for the indicated range of eigenvalues. Furthermore, for $\alpha > 0$ the potential $\lambda(z)$ in Eq. (14) increases with increasing z, from the value $\lambda(0)$, Eq. (18), lying in the upper half of the energy gap, to $|\epsilon_g|/2$. When the sign of α is changed, the potential $\lambda(z)$ changes sign. Because the masses M_1 and M_2 are different the reduced mass M(0) of Eq. (14) changes sign at an energy which is not at the center of the energy gap. For this reason, in the region of small z, the slow motion can be finite also for $\epsilon_g < 0$, which leads to the appearance of quasidiscrete levels.

What has been said will become clearer if we consider the quasiclassical integral from which we can estimate the energy of the ground state of Eq. (14):

$$\int \left\{ \mu[\varepsilon - \lambda(z)] \left[\frac{\lambda(0) - \bar{\lambda}}{\varepsilon_g} + \frac{b}{\beta}|z|^{\beta} \right]^{-1} \right\}^{1/2} dz \sim 1, \tag{29}$$

where $\mu = 2M_1M_2v^2/|\epsilon_g|(M_1 + M_2)$, z is in units $v/|\alpha\epsilon_g|$, and $\epsilon - \lambda$ is in units $\alpha^2|\epsilon_g|$.

If the level is not too close to the point of reversal of the mass

$$\left| \frac{\lambda(0) - \bar{\lambda}}{\varepsilon_g} \right| > \beta^{-1}\mu^{-\beta}, \tag{30}$$

then the coordinate dependence of the reduced mass, given by the second term in the denominator in Eq. (29) need not be taken into account, and we find for the energy and the radius:

$$\varepsilon - \lambda_0(0) \sim \alpha|\alpha\varepsilon_g|\beta^{-1/(1+\beta)} \left| \frac{\lambda_0(0) - \bar{\lambda}}{\mu\varepsilon_g} \right|^{\beta/(1+\beta)}, \quad z \sim \left| \beta \frac{\lambda_0(0) - \bar{\lambda}}{\mu\varepsilon_g} \right|^{1/(2+2\beta)}. \tag{31}$$

For $\alpha \ll 1$ one gets from this the last term in Eq. (28). It can be seen that the longitudinal size of a bound state in the z direction is small compared with the longitudinal size, the smallness being that indicated by the second expression in (31).

We point out that for normal position of the bands there cannot be any levels in the interval from the center of the gap to about $\bar{\lambda}$. For inverted position of the bands there can be quasidiscrete levels in this interval, of donor or acceptor type according to the sign of $\bar{\lambda}$.

For small $\lambda_0(0) - \bar{\lambda}$, when the inequality opposite to (30) is satisfied, levels can exist for a definite relation between the signs of $(\lambda_0(0) - \bar{\lambda})/\epsilon_g$ and b. If $(\lambda_0(0) - \bar{\lambda})/\epsilon_g > 0$ and $b < 0$, then in the region of small z a donor impurity creates a narrow potential well, and in the case of an acceptor impurity there is a potential barrier. We get the following estimate for the energy and radius of donor states:

$$\varepsilon - \lambda_0(0) \sim \alpha^2\varepsilon_g\mu^{-1}\beta^{-1/\beta} \left(\frac{\varepsilon_g}{\lambda_0(0) - \bar{\lambda}} \right)^{(1-\beta)/\beta}, \quad z \sim \left| \frac{\beta}{\varepsilon_g}(\lambda_0(0) - \bar{\lambda}) \right|^{1/2\beta} \tag{32}$$

and for acceptor states:

$$\varepsilon - \lambda_0(0) \sim \alpha^2\varepsilon_g\mu^{-\beta}\beta^{-1}, \quad z \sim \mu^{-1/2}. \tag{33}$$

If $(\lambda_0(0) - \bar{\lambda})/\epsilon_g < 0$ and $b > 0$, the roles of donor and acceptor impurities are interchanged. Finally, if $(\lambda_0(0) - \bar{\lambda})/\epsilon_g$ and b are positive, there are donor levels of the type (33); for negative $(\lambda_0(0) - \bar{\lambda})/\epsilon_g$ and b there are acceptor states (33).

In the case of infinite z motion it is interesting to calculate the singularities of the density of states near $\epsilon = \lambda_n(0)$. To do so one can again use the adiabatic approximation, since the character of the singularity is determined by the behavior of $u_n(z)$ in the region of small z. The calculation, which we shall not present, shows that the phase of the wave function, which determines the variation of the number of states, has no singularity at the point $\epsilon = \lambda_n(0)$ itself. However, in the region above a barrier, for example for an acceptor impurity for $\epsilon > \lambda_n(0)$ and inverted position of the bands, a behavior of the phase is found which is characteristic of the scattering of slow particles by a virtual level.

THE NONCOULOMB POTENTIAL AT SMALL DISTANCES

Let us examine what are the consequences of the departure of the potential (5) from the Coulomb form at small distances. If $\alpha < 1/2$, the shift of the levels $\lambda_n(0)$, Eq. (18), owing to this difference is small, since the Bohr characteristic distance $q^{-1} \sim v/|\alpha\epsilon_g|$ is large in comparison with R. The correction to the ground-state energy is calculated by perturbation theory:

$$\delta\lambda_0(0) = \frac{\alpha|\alpha\varepsilon_g|}{\beta\Gamma(2\beta+1)} \left(\frac{2R\alpha\varepsilon_g}{v} \right)^{2\beta} \left[1 - c\frac{\varkappa}{(\ln\varkappa)^{2\beta}} \right]. \tag{34}$$

We have expressed the r_0 of Eq. (5) in terms of $\kappa = \exp(R/r_0)$, requiring that the expressions (4) and (5) match at r = R; the constant $c \sim 1$ is determined by the lower limit $r \sim r_0$ of the region of Eq. (5).

We note that the correction (34) depends on the coupling constant α, the dielectric constant κ, and the ratio $R\epsilon_g/v$, which is of the order of the mass ratio μ^{-1}. Therefore the quantity (34) can be of the same order of magnitude as the adiabatic corrections given by the first expressions in Eqs. (31)–(33). Here, however, it is

necessary to recall that along with the adiabaticity conditions (13), which can be written out explicitly by means of Eqs. (31)—(33), we also had satisfied the condition that R be small compared with the characteristic values of z indicated in Eqs. (31)—(33):

$$R|\alpha\varepsilon_g|/v \ll z.$$

Otherwise in calculations of the potential $\lambda(z)$, Eq. (22), one must take into account the departure of the field from the Coulomb form at small distances.

This effect is important for $\alpha \to 1/2$. The main question is what then happens to the fast-motion levels of Eq. (18). To answer this we make the usual substitution for the Dirac equation[6]: $\chi^{(1)} = (\epsilon_g/2 - V + \lambda)^{1/2} f$. After eliminating $\chi^{(2)}$ from Eq. (16) the equation for f reduces to self-adjoint form:

$$f''(\rho) + k^2(\rho) f(\rho) = 0, \tag{35}$$

where

$$k^2(\rho) = \frac{(V-\lambda)^2 - \varepsilon_g^2/4}{v^2} - \frac{l^2-1/4}{\rho^2} - \frac{3}{4}\left(\frac{V'}{\varepsilon_g/2 - V + \lambda}\right)^2 - \frac{V''/2 + (l+1/2)V'/\rho}{\varepsilon_g/2 - V + \lambda} \tag{36}$$

the potential V is given by Eqs. (4) and (5) with $r = \rho$; primes denote differentiation with respect to ρ; for definiteness we are considering the case $\alpha > 0$. Using Eqs. (4) and (5), we find

$$k^2(\rho) = \frac{\lambda^2 - \varepsilon_g^2/4}{v^2} + \frac{2\alpha\lambda}{v\rho} + \frac{\alpha^2 - l^2}{\rho^2} - \frac{3 + (2l-1)(2 + (2\lambda + \varepsilon_g)\rho/\alpha v)}{4\rho^2(1 + (2\lambda + \varepsilon_g)\rho/2\alpha v)^2}, \quad \rho > R,$$

$$k^2(\rho) = \frac{\alpha^2\varkappa^2}{R^2}\varkappa^{-2\rho/R} - \frac{l^2}{\rho^2} - \frac{(l+1/2)\ln\varkappa}{\rho R} - \frac{5\ln^2\varkappa}{4R^2}, \quad r_0 \ll \rho < R, \tag{37}$$

where we have added the well-known Langer term $-1/4\rho^2$, since we intend to use the WKB method, and in the second equation have neglected terms of order $R\epsilon_g/v \ll 1$, and also used the fact that in the region in question the variation of the exponential in Eq. (5) is most important.

For the analysis of Eq. (35) we use the WKB method, which, as Kraĭnov[10] has shown for the Dirac equation, gives practically the exact value of the critical charge of the nucleus. Picking out the coordinate-dependent terms in Eq. (37) as an effective potential energy, we see that the difference from ref.[10] is in the specific character of the cut-off of the Coulomb potential. Owing to the last term in the second equation of (37), which appears because of the fast variation of the potential, there is a potential barrier for the ground state over a rather wide region:

$$\ln\ln\varkappa/\ln\varkappa > 1 - \rho/R > 0.$$

Nevertheless, as α increases the ground level drops, and for some coupling constant α_{cr} it reaches the value $\lambda = -\epsilon_g/2$, i.e., merges with the continuous spectrum of holes. The corresponding quasiclassical integral determining the position of the level is easily calculated, and we get the relation

$$R = \frac{4\alpha_{kp}^2 - 1}{\varepsilon_g \alpha_{kp}} v \exp\left(-2 - \frac{3\pi}{2(4\alpha_{kp}^2 - 1)^{1/2}}\right),$$

connecting α_{cr} with the cut-off radius R. Effects of the slow motion change α_{cr}. This change is small, of the order μ^{-1}; we shall not present the calculation of it here.

CONCLUSION

Let us briefly summarize the results of this paper. The impurity levels existing in a substance with a narrow energy gap and an anisotropic spectrum of free carriers are located inside the energy gap when the position of the bands is the direct one. For the inverted position of the bands quasidiscrete levels can exist if the coupling constant is not too small. The states are characterized by a Bohr radius $v/|\alpha\epsilon_g|$ which is large on the atomic scale. For $|\alpha| > \alpha_{cr}$ such states are unstable; neutral structures of smaller radius are stable. This is apparently the explanation of the small effectiveness of impurities in bismuth. The effect of a large anisotropy of the carrier spectrum is to make the bound states have the shape of ellipsoids of revolution with small longitudinal axis. The anisotropy of the mass allows us to use the adiabatic approximation. The condition for its applicability is that the longitudinal size of a bound state be small compared with the transverse radius [expressions for z in Eqs. (31)—(33)], and also that the fractional corrections to the energy levels (18), as given by Eqs. (31)—(33), be small.

I am happy to express my gratitude to A. A. Abrikosov, who called my attention to the fact that impurity states is gapless substances had not been studied previously.

[1)] We take $\hbar = 1$.

[1] A. A. Abrikosov and L. A. Fal'kovskiĭ, Zh. Eksp. Teor. Fiz. **43**, 1089 (1962) [Sov. Phys.-JETP **16**, 769 (1963)].
[2] A. A. Abrikosov, J. Low Temp. Phys. **8**, 315 (1972).
[3] G. Oelgart and R. Herrmann, Phys. Status Solidi B **58**, 181 (1973).
[4] S. D. Beneslavskiĭ and L. A. Fal'kovskiĭ, Fiz. Tverd. Tela **16**, 1360 (1974) [Sov. Phys.-Solid State **16**, 876 (1974)].
[5] A. A. Abrikosov, Zh. Eksp. Teor. Fiz. **44**, 2039 (1963) [Sov. Phys.-JETP **17**, 1372 (1963)].
[6] Ya. B. Zel'dovich and V. S. Popov, Usp. Fiz. Nauk **105**, 403 (1971) [Sov. Phys.-Uspekhi **14**, 673 (1972)].
[7] L. V. Keldysh, Zh. Eksp. Teor. Fiz. **45**, 364 (1963) [Sov. Phys.-JETP **18**, 253 (1964)].
[8] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, Relativisticheskaya kvantovaya teoriya (Relativistic Quantum Theory), Part 1, "Nauka", 1968. English translation: Addison-Wesley, 1971.
[9] W. Kohn and J. M. Luttinger, Phys. Rev. **98**, 915 (1955).
[10] V. P. Kraĭnov, ZhETF Pis. Red. **13**, 359 (1971) [JETP Lett. **13**, 255 (1971)].

Translated by W. H. Furry
167