# DEEP LEVELS IN SEMICONDUCTORS

L. V. KELDYSH

P. N. Lebedev Physics Institute, Academy of Sciences, U.S.S.R.

Submitted to JETP editor February 19, 1963

J. Exptl. Theoret. Phys. (U.S.S.R.) **45**, 364-375 (August, 1963)

The problem of local impurity levels having a binding energy comparable with the forbidden band is solved in the two-band approximation, in which the electron and hole bands and their interaction are taken into account. If the two bands are not degenerate, the electron and hole motion can be described by equations of the Dirac type. In the case of triple degeneracy of the valence band, a set of equations is obtained to relate the motion of a scalar particle (electron) with that of a vector particle (hole). Analysis of the equations permits one to explain qualitatively such features of impurity centers with deep levels as attraction to a charged center of carriers with the same charge, capture of both electrons and holes by the same center, etc. The existence of two types of holes (heavy and light) is found to be important for the recombination process: trapping levels correspond mainly to heavy hole states, while large cross sections are due to small light-hole masses.

THE development of the effective-mass method and the detailed clarification of the energy spectra of typical semiconductors have made it possible to solve in exhaustive manner the problem of shallow impurity levels[1-3], i.e., levels with very low binding energies, on the order of $10^{-2}$ eV, formed near impurity atoms having one less valence electron than the atom of the main substance (for example, group III or V elements in Ge). These states are completely determined by the effective charge of the center, i.e., the dielectric constant of the medium, and by the structure of the bottom of that energy band near which they are formed, since their distances from the other bands are on the order of 1 eV, i.e., much larger than the binding energy. They are described quantitatively by an equation (or system of equations) of the Schrödinger type with Coulomb potential, and, generally speaking, by anisotropic effective masses characteristic of the given band. The small effective masses and large dielectric constants typical of all semiconductors give rise to low binding energies and to Bohr orbits with tremendous effective radii (compared with the lattice period).

In contradistinction from the shallow levels, it is customary to classify levels as deep if the binding energies are several electron volts, i.e., comparable with the width of the forbidden band, which is formed as a rule near multiply-charged impurity centers, vacancies, etc. Their nature is by far less clear. It is only known that their role in semiconductor physics is very significant: they are usually recombination centers or traps, and consequently determine the lifetimes of the non-equilibrium carriers.

It is customary to point to two main difficulties in the theoretical description of these levels. First, their radius is much smaller than the radius of the shallow levels, so that there are no grounds for assuming a Coulomb field for the center with correction for the dielectric constant and for using the effective-mass approximation. Second, their distance from both bands—valence and conduction—is of the same order, and they cannot therefore be regarded as belonging to either band, unlike the shallow levels, which are determined by a Shrödinger equation containing the effective mass or other parameters of either the valence or the conduction band.

In the present paper we present, on the basis of a two-band approximation similar to that used by Kane to explain the band structure of InSb[4], an analysis of the deep-level problem, in which at least the second of the foregoing difficulties is resolved. As to the first, from our point of view it has no basic significance at all, since the radii of the bound states deep in the forbidden band, although smaller than the radii of the shallow levels, are nevertheless several times larger than the crystal lattice constant d. Indeed, by virtue of the uncertainty relations the radius a and the energy $\epsilon$ of the bound state satisfy the relation $a\sqrt{m\epsilon} \gtrsim \hbar$, where $\hbar$ is Planck's constant and m is some average effective mass for the electron and hole.

Hence

$$a/d \gtrsim \sqrt{(\hbar/d)^2/m\varepsilon} \gtrsim \sqrt{(\hbar/d)^2/m\Delta} \equiv \sqrt{1/\lambda} \gg 1, \quad (1)$$

where $\Delta$ is the half-width of the forbidden band. The last inequality of (1) follows essentially from the definition of the semiconductor, since it requires that the allowed band be much wider than the forbidden band. If we use values typical of semiconductors such as $d = 3 \times 10^{-8}$, $m = 10^{-28}$ g, and $\Delta = 1$ eV, we get $a/d \gtrsim 3$.

The course of the potential in this region can be calculated quite correctly within the framework of the approximation under consideration. Indeed, the polarization at distances $\sim a$ from the center is produced essentially by the valence electrons, bound in the lattice with energies $\lesssim \hbar^2/ma^2 < \Delta$, i.e., lying at distance of the order of $\Delta$ from the edge of the band. But such states are known to be well described by the Kane model. Deeper electrons contribute to the polarization cloud at smaller distances $\sim d$ and can be taken into account in the region of interest to us by using some fixed value of the dielectric constant.

In this paper, however, we shall not calculate this potential, but confine ourselves to an analysis of the effects connected with the interactions between the valence and conduction bands at a specified potential, particularly a Coulomb potential, with effective charge Ze, taking into account both the charge of the center and the effective dielectric constant. Even this allows us to explain several qualitative peculiarities of the deep levels, which are utterly unintelligible within the framework of the single-band scheme. Of course, no quantitative meaning should be ascribed to the obtained results, both in view of the arbitrariness in the choice of the potential and in view of the extreme idealization of the band structures which will be employed below. The small parameter characterizing the omitted term is the quantity $\lambda$ introduced in (1), equal to the ratio of the width of the forbidden band to the distance to the other bands, the influence of which we neglect in most cases. If the potential is more precisely defined, successive neglect of terms of order $\lambda$ can guarantee an accuracy on the order of 10%.

Before we proceed to a detailed analysis of various specific models, we present now a highly schematized qualitative analysis which permits, however, to obtain in lucid form all the main results and estimate the degree to which they are general. We shall assume here, as in what follows, that the minima of both bands lie at $\mathbf{k} = 0$. Then if the momentum k is small compared with the limiting value $\hbar/d$ but not small compared with $\sqrt{m\Delta}$, the function $\epsilon(\mathbf{k})$, which describes the dependence of the energy on the momentum in both bands, is determined uniquely by the following requirements[5].

1. Its value in the electron and hole bands is given by two branches of one analytic function, since both are obtained by solving the same Schrödinger problem.

2. As $k^2$ approaches zero, $\epsilon_e(\mathbf{k}) \approx \Delta + k^2/2m_e$ in the electron band and $\epsilon_h(\mathbf{k}) \approx -\Delta - k^2/2m_h$ in the hole band.

This function is of the form

$$\varepsilon(\mathbf{k}) = k^2/2M + \sqrt{\Delta^2 + k^2\Delta/m},$$
$$1/M = 1/2\, m_e - 1/2\, m_h, \quad 1/m = 1/2\, m_e + 1/2\, m_h. \quad (2)$$

Higher orders of $k^2$ are omitted both under and outside the radical sign, in view of the assumed smallness of $(kd/\hbar)^2$. In the energy region under consideration they are of the order of $\lambda$. An exact calculation for the two-band model[4] yields the same result. It shows also that the mass M, which is outside the radical sign, is brought about by the interaction with other bands, while m is connected with the branching under consideration, i.e., the interaction between the electron and hole bands, by virtue of which $m/M \sim \lambda$. The term in front of the radical can therefore be omitted henceforth. Experimental data on semiconductors with narrow forbidden bands confirm qualitatively that the masses of the electrons and light holes are of the same order (in the case of Ge we have in mind, of course, the electrons not in the main lateral minima on the (111) axis, but in the additional minimum at $k = 0$).

In accordance with the general principles of quantum mechanics, the Schrödinger equation (at least the symbolic one) for a system described by (2) should be in the form

$$\{\Delta\sqrt{1 - \hbar^2\nabla^2/m\Delta} + V(\mathbf{r})\}\psi(\mathbf{r}) = \varepsilon\psi(\mathbf{r}), \quad (3)$$

where $V(\mathbf{r})$ is the external potential and $\nabla$ the differentiation operator. Let us apply to this equation the operator $\Delta\sqrt{1 - \hbar^2\nabla^2/m\Delta} - V(\mathbf{r})$. Then

$$\left\{-\frac{\hbar^2}{2m}\nabla^2 + \frac{\varepsilon}{\Delta}V(\mathbf{r}) - \frac{1}{2\Delta}V^2(\mathbf{r}) + \frac{1}{2}\left[\sqrt{1 - \frac{\hbar^2}{m\Delta}\nabla^2},\ V(\mathbf{r})\right] - \frac{\varepsilon^2 - \Delta^2}{2\Delta}\right\}\psi(\mathbf{r}) = 0. \quad (4)$$

Expression (4) has the form of the Schrödinger equation for an ordinary particle with mass m, but has several characteristic features. The role of the effective energy is played in it by the quantity $(\epsilon^2 - \Delta^2)/2\Delta$, which goes over in the vicinity of each band into the energy reckoned from the bottom of this band. The potential $V(\mathbf{r})$ has $\epsilon/\Delta$

as a coefficient, i.e., it reverses sign with $\epsilon$: if it is attractive for electrons, it is repulsive for holes, and vice versa. Finally, the most interesting circumstance is the appearance of an additional potential $V^2(\mathbf{r})/2\Delta$, which is attractive for both electrons and holes, regardless of the sign of the initial potential. At small distances it always predominates, for $V(r)$ tends to infinity as $r$ tends to zero independently of the character of the screening. This fact explains why the same center can capture both electrons and holes. Capture of the type of carrier which is repelled at large distances calls in this case for activation energy.

In addition, an increase in the degree of singularity of the attractive potential can cause the particle to fall into the center, i.e., recombination, if $V^2(\mathbf{r})$ increases more rapidly than $1/r^2$ with increasing $r$, or exactly as $1/r^2$ but with a coefficient larger than $\hbar^2\Delta/8m$[6]. Thus, it is precisely the Coulomb potential $V \sim 1/r$ which serves as the dividing line between the potentials for which bound levels can exist in the forbidden band and the potentials for which instantaneous recombination occurs. From this point of view, the impurities existing in the neutral state only are those for which the potential increases more rapidly than $(\hbar/2r)\sqrt{\Delta/m}$ in the given crystal even for single ionization; singly-charged centers are those for which this situation arises in the case of double ionization, etc. Finally, the usual recombination center should have in the initial state a potential that increases slowly in the sense indicated above, so that it can capture at a stationary level an electron which falls into the center after the capture of a hole, i.e., after the charge increases by unity.

It is easy to understand why the Coulomb potential is the recombination limit. To demonstrate this we repeat the usual arguments showing that for a free particle this role is played by the potential $1/r^2$. Allowing for the uncertainty relation $\Delta k \Delta r \gtrsim h$, the total energy

$$\varepsilon = k^2/2m + V(r) \gtrless \hbar^2/2mr^2 - |V(r)| \qquad (5)$$

is bounded from below if $V(\mathbf{r})$ increases more slowly than $1/r^2$. In the opposite case the particle falls into the center, i.e., $\epsilon$ can tend to $-\infty$. This argument is essentially based on the fact that the kinetic energy increases like $k^2 \gtrsim \hbar^2/r^2$. But it is seen from (2) that if we neglect the term $k^2/2M$, as we have done, the kinetic energy increases in our case in proportion to $|\mathbf{k}|$ when $\mathbf{k}$ is large, so that in a relation of the type (5) even a Coulomb increase is sufficient to cause recombination. An account of the $k^2/2M$ term limits this recombination at very large momenta, i.e., very short distances, which are of no interest to us.

We note finally that Eq. (4) contains one more term—the commutator of kinetic into potential energy. It is equivalent to some effective spin-orbit coupling and behaves at small distances like $\nabla V(\mathbf{r})\nabla$. We shall not analyze it in detail at present, since it has a somewhat different form in real models, which we now proceed to consider.

The main difference between these models and the schematic equation (3) lies in the fact that in these models we deal with a multi-component $\psi$-function (number of components equal to the number of bands considered), and Eq. (3) is then replaced by a system of first-order equations. Following the method of Luttinger and Kohn[1], we represent the solution of the complete Schrödinger equation

$$\left\{-\frac{\hbar^2}{2m_0}\nabla^2 + W(\mathbf{r}) + V(\mathbf{r})\right\}\psi(\mathbf{r}) = \varepsilon\psi(\mathbf{r}) \qquad (6)$$

in the form

$$\psi(\mathbf{r}) = \sum_{n\mathbf{k}} c_n(\mathbf{k})\, e^{i\mathbf{k}\mathbf{r}/\hbar} u_{n0}(\mathbf{r}). \qquad (7)$$

Here $m_0$ is the mass of the free electron $W(\mathbf{r})$ the periodic potential of the crystal lattice, $V(\mathbf{r})$ the potential of the impurity center, and $u_{n0}$ the purely periodic (i.e., corresponding to $\mathbf{k} = 0$) solutions of (6) without the external potential $V(\mathbf{r})$. We assume for simplicity that the minima of both the electron and hole bands lie at $\mathbf{k} = 0$. The eigenvalues corresponding to $u_{n0}$ are denoted $\epsilon_{n0}$. We can then readily obtain for the functions

$$\varphi_n(\mathbf{r}) = \sum_{\mathbf{k}} c_n(\mathbf{k})\, e^{i\mathbf{k}\mathbf{r}/\hbar} \qquad (8)$$

the following system of equations:

$$\left\{-\frac{\hbar^2}{2m_0}\nabla^2 + \varepsilon_{n0} - \varepsilon\right\}\varphi_n(\mathbf{r}) - \sum_{n'}\left\{\frac{i\hbar\, \mathbf{p}_{nn'}}{m_0}\nabla + \tilde{V}_{nn'}(\mathbf{r})\right\}\varphi_{n'}(\mathbf{r}) = 0, \qquad (9)$$

where

$$\tilde{V}_{nn'}(\mathbf{r}) = \frac{1}{(2\pi)^3}\int e^{i\mathbf{k}(\mathbf{r}-\mathbf{r}')/\hbar}\, u^*_{n0}(\mathbf{r}')\, V(\mathbf{r}')\, u_{n'0}(\mathbf{r}')\, d\mathbf{k}\, d\mathbf{r}', \qquad (10)$$

$$\mathbf{p}_{nn'} = -i\hbar\int u^*_{n0}(\mathbf{r})\,\nabla u_{n'0}(\mathbf{r})\, d\mathbf{r}, \quad \mathbf{p}_{nn} = 0, \quad \mathbf{p}_{nn'} = \mathbf{p}^*_{n'n}. \qquad (11)$$

In the last formula the integral is taken over the volume of the unit cell, for which the functions $u_{n0}$ are normalized. By virtue of the orthogonality of this system of functions, the potential $\tilde{V}_{nn'}(\mathbf{r})$ reduces to

$$\tilde{V}_{nn'}(\mathbf{r}) = V(\mathbf{r})\,\delta_{nn'} \qquad (12)$$

accurate to terms of order $\lambda$, if $V(\mathbf{r})$ changes noticeably over distances of order $\hbar/\sqrt{m\Delta}$, as we assume.

Let us now break down the entire infinite set of functions $\varphi_n$ into two groups. The first, designated by indices c and v, will include those numbers n for which $u_{n0}$ correspond to the bottom of the conduction or valence band. The second group includes all the remaining ("far") bands. They will be designated by the index j ($j \neq c, v$). The possible degeneracy of the c and v bands will be taken into account by means of a supplementary index, which will not be written out for the time being, for it is best to introduce it in different fashion for different specific cases. Using standard perturbation theory with respect to $\mathbf{kp}_{nn'}$,[1] we can eliminate the $\varphi_j$ from the equations for $\varphi_{c,v}$. The system (9) then reduces to

$$\left\{\frac{\hbar^2}{2m_0} D^{cc}_{\alpha\beta}\nabla_\alpha\nabla_\beta + V(\mathbf{r}) + \Delta - \varepsilon\right\}\varphi_c(\mathbf{r}) - \left\{\frac{i\hbar}{m_0}\mathbf{p}_{cv}\nabla - \frac{\hbar^2}{2m_0} D^{cv}_{\alpha\beta}\nabla_\alpha\nabla_\beta\right\}\varphi_v(\mathbf{r}) = 0,$$

$$\left\{\frac{\hbar^2}{2m_0} D^{vv}_{\alpha\beta}\nabla_\alpha\nabla_\beta + V(\mathbf{r}) - \Delta - \varepsilon\right\}\varphi_v(\mathbf{r}) - \left\{\frac{i\hbar}{m_0}\mathbf{p}_{vc}\nabla - \frac{\hbar^2}{2m_0} D^{vc}_{\alpha\beta}\nabla_\alpha\nabla_\beta\right\}\varphi_c(\mathbf{r}) = 0, \tag{13}$$

$$D^{cc}_{\alpha\beta} = -\delta_{\alpha\beta} + 2\sum_j \frac{(p_{cj})_\alpha (p_{jc})_\beta}{m_0(\varepsilon_{j0} - \Delta)},$$

$$D^{cv}_{\alpha\beta} = 2\sum_j \frac{(p_{cj})_\alpha (p_{jv})_\beta}{m_0(\varepsilon_{j0} + \Delta)} \quad \text{etc.} \tag{14}$$

The indices $\alpha$ and $\beta$ run through the three values x, y, and z; the energy is reckoned from the center of the forbidden band the width of which is $2\Delta$. We emphasize once more that generally speaking $\varphi_c(\mathbf{r})$ and $\varphi_v(\mathbf{r})$ have in the case of band degeneracy several components. All the quantities $D^{cc}_{\alpha\beta}$, $D^{cv}_{\alpha\beta} = (D^{vc}_{\alpha\beta})^*$, and $D^{vv}_{\alpha\beta}$ are square matrices.

At the energies $\epsilon \sim \Delta$ of interest to us, all the momenta are of order $\sqrt{m\Delta}$, and all terms of the type $D_{\alpha\beta}\nabla_\alpha\nabla_\beta$ can be omitted, with accuracy of order $\lambda$. In other words, the effective mass is brought about in the case of small $\Delta$ only by the $\mathbf{kp}$ interaction of the neighboring bands[4]. An exception are the bands for which all the matrix elements, of momentum $\mathbf{p}$, coupling them to the other c and v bands, vanish, Singular cases of this type, as is well known, are heavy holes. We therefore begin the analysis with an artificial model, in which heavy holes play no role whatever.

This situation can be attained in the Kane model[4] if it is assumed that the spin-orbit splitting energy is $\Lambda \gg \Delta$, but has a sign opposite that of the real semiconductors InSb, Ge, etc. The upper valence band is then the band with total momentum $j = 1/2$, which is doubly degenerate in the spin. Exactly the same result is obtained if the usual sign of $\Lambda$ is retained, but it is assumed that the triple representation corresponds to the conduction band, while the valence band is not degenerate. To be specific, we shall consider in what follows the first of these variants. Then there are two functions $\varphi^{\pm}_c$ corresponding to spin projections $+1/2$ and $-1/2$, along the momentum axis, and six functions $\varphi^{\pm}_{v\alpha}$, which transform like x, y, and z with respect to the index $\alpha$. Accordingly

$$(p_{cv\alpha})_\beta = p\delta_{\alpha\beta}. \tag{15}$$

It is more convenient, however, to go over to a different representation of the valence functions, in which the total momentum j is diagonal, and with it the spin-orbit interaction[4,7]:

$$j = \tfrac{1}{2}, \quad \chi^{\pm}_1 = \pm[\varphi^{\pm}_{vz} - (\varphi^{\mp}_{vx} \pm i\varphi^{\mp}_{vy})]/\sqrt{3}, \tag{16}$$

$$j = \tfrac{3}{2}, \quad \chi^{\pm}_{3h} = [\varphi^{\pm}_{vx} \pm i\varphi^{\pm}_{vy}]/\sqrt{2},$$

$$\chi^{\pm}_{3l} = [2\varphi^{\pm}_{vz} + (\varphi^{\mp}_{vx} \pm i\varphi^{\mp}_{vy})]/\sqrt{6}. \tag{17}$$

We now turn again to (13), eliminating from the second equation the spin-orbit term $(1/2 - j)\Lambda$, which was not considered so far. The analysis is best continued in the k-representation, introducing a special system of coordinates in which the z axis is directed along the vector $\mathbf{k}$. It is then seen directly that the wave function $\chi_{3h}$ of the heavy holes is not connected at all with the others, and the equation for it can be left out. The remaining equations assume the form

$$[\varepsilon - \Delta - V(-i\hbar\nabla_{\mathbf{k}})]\varphi_c(\mathbf{k}) - (p/\sqrt{3}m_0)\, k\,[\sqrt{2}\chi_{3l}(\mathbf{k}) + \sigma_z\chi_1(\mathbf{k})] = 0,$$

$$[\varepsilon + \Delta - V(-i\hbar\nabla_{\mathbf{k}})]\chi_1(\mathbf{k}) - (p^*/\sqrt{3}m_0)\, k\sigma_z\varphi_c(\mathbf{k}) = 0,$$

$$[\varepsilon + \Delta + \Lambda - V(-i\hbar\nabla_{\mathbf{k}})]\chi_{3l}(\mathbf{k}) - \sqrt{2/3}\, p^* k\varphi_c(\mathbf{k})/m_0 = 0, \tag{18}$$

where $\sigma_z$ is the third Pauli matrix.

In the energy region $\epsilon \sim \Delta$ we get from the third equation

$$\chi_{3l}(\mathbf{k}) \approx \sqrt{\frac{2}{3}}\frac{p^* k}{m_0\Lambda}\varphi_c(k) \sim \frac{\Delta}{\Lambda}\varphi_c(\mathbf{k}), \tag{19}$$

and if $\Lambda \gg \Delta$ the contribution of $\chi_{3l}$ in the first equation can be neglected. But then the remaining

two equations constitute the Dirac equation for the bispinor $\begin{pmatrix}\varphi_c\\ \chi_1\end{pmatrix}$ in the chosen special coordinate frame. In the coordinate representation and in an arbitrary coordinate system they assume the usual form

$$[\varepsilon - \Delta - V(\mathbf{r})]\,\varphi_c(\mathbf{r}) + i\hbar s\,\sigma\nabla\chi_1(\mathbf{r}) = 0,$$
$$[\varepsilon + \Delta - V(\mathbf{r})]\,\chi_1(\mathbf{r}) + i\hbar s\,\sigma\nabla\varphi_c(\mathbf{r}) = 0, \qquad (20)$$

in which the constant $s = p/\sqrt{3}\,m_0$ plays the role of the velocity of light, and the effective mass is given by

$$m = \frac{\Delta}{s^2} = m_0\frac{3m_0\Delta}{p^2}. \qquad (21)$$

Elimination of one of the functions from the system (20) leads directly to Eq. (4), except that the commutator of the potential into kinetic energy is replaced by the term

$$(\sigma\ \mathrm{grad}\ \ln[\varepsilon \pm \Delta - V(\mathbf{r})])\,(\sigma\ \mathrm{grad})$$
$$= \frac{\partial \ln[\varepsilon \pm \Delta - V(\mathbf{r})]}{\partial r}\left(\frac{\mathbf{r}}{r}\nabla + \frac{\sigma\mathbf{L}}{r}\right), \qquad (22)$$

where $\mathbf{L}$ is the momentum operator, and the signs + and − appear in the equations for $\chi_1$ and $\varphi_c$, respectively. The first part of (22) is written already for a spherically symmetrical potential.

Let us consider by way of an example the known[8] solutions of the Dirac equation in a Coulomb field $V(r) = Ze^2/r$ ($Ze$ is the charge of the center, divided by the effective dielectric constant, which generally speaking differs from its value at infinity). We note only the results of greatest interest to us. The energy levels are determined by the two quantum numbers n and $\gamma$:

$$\varepsilon_{n\gamma} = \Delta/\sqrt{1 + \alpha^2/(n+\gamma)^2}, \qquad \alpha = Ze^2/\hbar s,$$
$$n = 0, 1, 2, 3, \ldots, \qquad \gamma^2 = (j+\tfrac{1}{2})^2 - \alpha^2, \qquad (23)$$

and j is the total momentum. The quantity $\gamma^2$ determines the effective centrifugal potential. At sufficiently large Z it becomes negative, and the corresponding stationary levels, as can be seen from (23), vanish, i.e., there is falling into the center and a resultant decrease in Z. All the electronic levels lie in the upper half of the forbidden band ($\epsilon > 0$ for $Z > 0$). This rule is violated, however, if the deviations of the field from a Coulomb one are taken into account. Finally, we note also that the wave function of the continuous spectrum behaves like $r^{\gamma-1}$ as $r \to 0$, for both attractive and repulsive centers. Thus, for $j = \frac{1}{2}$ the probability of finding an electron near the center increases like $\exp(-2\alpha^2 \ln r)$, which can noticeably increase the cross section for the capture of slow electrons at large Z.

We now turn to a case of greater practical interest, when the spin-orbit splitting is small compared with the width of the forbidden band. Such a model is closer to the real situation occurring in Ge, and particularly in crystals of the GeAs type, where the minima of both bands lie at $\mathbf{k} = 0$. Neglecting the spin-orbit coupling, we can disregard in general the presence of electron spin. Then we have one function $\varphi_c$ and three functions that transform like the components of the vector $\varphi_v$. The vector $\mathbf{p}_{cv}$ is then parallel to $\varphi_v$. But then, as can be readily verified, only the band of light holes, defined by the condition $\varphi_{\Lambda v}(\mathbf{k}) \parallel \mathbf{k}$, is directly coupled with the conduction band, and this explains the smallness of the effective mass in this band. The effective masses of the two other hole bands are entirely connected with the values of $D^{vv}_{\alpha\beta}$, which must therefore be left in the equation. We make, however, one more simplification, assuming that the matrices $D_{\alpha\beta}$ have not cubic but spherical symmetry. We then get from (13) the following equations:

$$[\varepsilon - \Delta - V(\mathbf{r})]\,\varphi_c(\mathbf{r}) + i\hbar s\,\mathrm{div}\,\varphi_v(\mathbf{r}) = 0,$$
$$\left[-\frac{\hbar^2}{2M}\,\mathrm{rot\,rot} - \frac{\hbar^2}{2M'}\,\mathrm{grad\,div} + \varepsilon + \Delta - V(\mathbf{r})\right]\varphi_v(\mathbf{r})$$
$$+ i\hbar s\,\mathrm{grad}\,\varphi_c(\mathbf{r}) = 0, \qquad (24)^*$$

where M is the mass of the heavy hole, M′ is connected with the mass of the light holes $M_l$ by the relation $M_l^{-1} = M'^{-1} + m^{-1}$ (here $M_l \approx m$, since $M' \gg m$), and the parameters s and m are determined, as in the preceding case, by (21). The second term $M'^{-1}$ in the second equation can be omitted, since it yields only a small correction to the mass of the light holes. From the same consideration, the curl curl operator in the first term can be replaced by $-\nabla^2$.

Let us investigate now the possible types of solutions of the system (24) for a spherically symmetrical potential $V(\mathbf{r})$.

First, there are solutions satisfying the supplementary condition $\varphi_v = 0$, meaning that there is no contribution of the light holes to the wave function. Then $\varphi_c$ also vanishes, and the remaining equation

$$[(\hbar^2/2M)\nabla^2 + \varepsilon + \Delta - V(r)]\,\varphi_v(r) = 0 \qquad (25)$$

describes the motion of the heavy hole in the poten-

*rot = curl.

tial V(r). In order to satisfy automatically the supplementary condition, we must seek $\varphi_V$ in the form

$$\varphi_v(r) = \varphi(r)\, Y^{(0)}_{lm}(\mathbf{n}), \quad Y^{(0)}_{lm}(\mathbf{n}) = -i/[\sqrt{l(l+1)}]\,[\mathbf{r}\nabla]\,Y_{lm}(\mathbf{n}), \quad \mathbf{n} = \mathbf{r}/r. \tag{26}$$

Then the radial function $\varphi(\mathbf{r})$ satisfies the relation

$$\frac{\hbar^2}{2M}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\varphi}{dr}\right) + \left[\varepsilon + \Delta - V(r) - \frac{\hbar^2}{2M}\frac{l(l+1)}{r^2}\right]\varphi = 0. \tag{27}$$

Among these states, however, there are no spherically symmetrical ones ($l = 0$), as can be verified directly from (26).

States of this type form a second group of solutions, in which there is no contribution of the heavy holes (curl $\varphi_V = 0$). For these we have $\varphi_C(\mathbf{r}) = \varphi_C(\mathbf{r})$, $\varphi_V(\mathbf{r}) = \mathbf{n}\varphi_V(\mathbf{r})$, and the equation for the radial functions

$$[\varepsilon - \Delta - V(r)]\,\varphi_c(r) + i\hbar s\frac{1}{r^2}\frac{d(r^2\varphi_v)}{dr} = 0,$$

$$[\varepsilon + \Delta - V(r)]\,\varphi_v(r) + i\hbar s\frac{d\varphi_c}{dr} = 0 \tag{28}$$

coincides with the Dirac equation for $j = 1/2$.

The greatest interest attaches to the third group of solutions, in which the potential V(r) tangles together the light and heavy holes, so that none of the supplementary conditions used above is satisfied, i.e., div $\varphi_V \not\equiv 0$, curl $\varphi_V \not\equiv 0$, and $l \neq 0$. These solutions are analyzed in detail in the Appendix. Their main feature is that the behavior of $\varphi(\mathbf{r})$ at large distances from the center is determined by the mass of the electron or of the light hole m, while near the center there is a narrow region with width of order $\hbar/\sqrt{M\Delta}$, where $|\varphi|^2$ increases sharply because of the presence of heavy holes (we have in mind a center that attracts holes, i.e., $V(r) > 0$). At sufficiently large Z, when $\alpha^2 \gtrsim m/M$, such levels appear above the bottom of the conduction band. These are essentially additional electronic levels, but they can be arbitrarily called virtual heavy-hole levels, since they fall into the region of the continuous spectrum but have a wave function of symmetry different than the neighboring continuous-spectrum functions. The capture of a hole by such a center, i.e., the departure of a randomly captured electron from such a center, can therefore occur sufficiently slowly, i.e., these will be quasistationary levels of relatively low width. The cross section for the capture on them should have a resonant character of the Breit-Wigner type.

If the lifetime of such a level is sufficiently long, then there is a noticeable probability of transition from this level to levels of the same center in the forbidden band, which are therefore stationary in the full meaning of the word. This is how an electron is captured by a negatively charged center. In real multiply charged centers, an important role can be played in such capture by the presence of other electrons in the center. Indeed, for example, the ground-state level of the three-electron system $Cu^{---}$ in Ge lies at 0.26 eV, under the bottom of the conduction band. It is quite probable, however, that this system also has excited levels, some of which can fall in the conduction band. This makes possible the already described resonant capture of electrons at these excited levels with subsequent cascade[9] transition into the ground state.

A detailed quantitative analysis of the recombination process calls for calculation of the polarization corrections to the potential and for a more accurate account of the real band structure. All these are beyond the scope of the present paper, which is aimed at showing the possibility in principle of describing deep levels in semiconductors within the framework of an approximation of the effective-mass type, and explain at the same time such qualitative features of these levels as the ability of capturing carriers of either polarity, for example.

In conclusion I wish to express my deep gratitude to G. E. Pikus, whose remarks played an important role in the performance of this work, and particularly in its formulation.

## APPENDIX

The system of equations (24) leads, after elimination of the component $\varphi_C(\mathbf{r})$, to the following equation for the vector $\varphi_V(\mathbf{r})$:

$$\left\{\frac{\hbar^2}{2M}\,\mathrm{rot}\,\mathrm{rot}\,\varphi_v(\mathbf{r}) - \frac{\hbar^2\Delta}{m}\,\mathrm{grad}\,\frac{\mathrm{div}\,\varphi_v(\mathbf{r})}{\varepsilon - \Delta - V(\mathbf{r})} - [\varepsilon + \Delta - V(\mathbf{r})]\,\varphi_v(\mathbf{r})\right\} = 0. \tag{A.1}$$

For the case of interest to us, that of an acceptor impurity, $V(\mathbf{r}) > 0$ and if $\epsilon < \Delta$ the determination of the lowest level reduces to the variational problem

$$\min\Phi\{\varphi_v\} = \min\int\left\{\frac{\hbar^2}{2M}|\,\mathrm{rot}\,\varphi_v|^2 + \frac{\hbar^2}{m}\frac{\Delta}{\Delta - \varepsilon + V(\mathbf{r})}|\,\mathrm{div}\,\varphi_v|^2 - V(\mathbf{r})\,|\varphi_v|^2\right\}d\mathbf{r} = -\varepsilon - \Delta \tag{A.2}$$

with the additional normalization condition

$$\int\left\{\frac{\hbar^2}{m\Delta}\frac{\Delta^2}{[\Delta - \varepsilon + V(\mathbf{r})]^2}|\,\mathrm{div}\,\varphi_v|^2 + |\varphi_v|^2\right\}d\mathbf{r} = 1. \tag{A.3}$$

Equation (A.2) calls, however, for additional explanations, since the integrand in it depends on $\epsilon$ as a parameter. Therefore, strictly speaking, we have in mind the following procedure: We introduce under the integral sign in lieu of $\epsilon$ a new parameter $\eta(\epsilon \to \eta)$, carry out the variation for a specified $\eta < \Delta$, determine the extremal $\Phi_{\min}(\eta)$ and $\varphi_V(\eta)$, and then solve the transcendental equation

$$\Phi_{min}(\varepsilon) = -\varepsilon - \Delta, \qquad (A.4)$$

which determines the sought-for level. We shall see, however, that the real problem is much simpler.

Indeed, by virtue of the condition $m \ll M$ the second term in the integrand of (A.2) is very large and positive in all cases when div $\varphi_V$ differs noticeably from zero. Therefore the requirement that the functional be a minimum leads to the condition div $\varphi_V \approx 0$, which is satisfied, as can be readily verified, with such an accuracy that the second term becomes much smaller than the first in (A.2). Then the terms containing div $\varphi_V$ in (A.2) and (A.3) can be omitted, and the problem is reduced to the usual form, but with the limitation that the variation is carried out on the class of functions satisfying the condition div $\varphi_V = 0$. One possibility of satisfying this condition was indicated above, namely formula (26). Expansion of $\varphi_V$ in spherical vectors shows that in addition to (26) there are solutions of the type

$$\varphi_v(\mathbf{r}) = \varphi_{\parallel}(r)\, \mathbf{Y}_{lm}^{(-1)}(\mathbf{n}) + \varphi_{\perp}(r)\, \mathbf{Y}_{lm}^{(+1)}(\mathbf{n}), \qquad (A.5)$$

which form with (26) a complete system. Here

$$\mathbf{Y}_{lm}^{(-1)}(\mathbf{n}) = \mathbf{n} Y_{lm}(\mathbf{n}), \quad \mathbf{Y}_{lm}^{(+1)}(\mathbf{n}) = \frac{r}{\sqrt{l(l+1)}} \nabla Y_{lm}(\mathbf{n}). \qquad (A.6)$$

A particular case of (A.6) are the solutions (28) considered above. For the functions (A.5) we have

$$\operatorname{div} \varphi_v(r) = \left\{ \frac{1}{r^2} \frac{d}{dr}(r^2 \varphi_{\parallel}) - \frac{\sqrt{l(l+1)}}{r} \varphi_{\perp} \right\} Y_{lm}(\mathbf{n}),$$

so that when $l \neq 0$ the condition div $\varphi_V$ can be satisfied by putting

$$\varphi_{\perp}(r) = \frac{1}{r\sqrt{l(l+1)}} \frac{d}{dr}(r^2 \varphi_{\parallel}), \qquad (A.7)$$

after which one independent radial function $\varphi_{\parallel}(\mathbf{r})$ is left for variation, and Eqs. (A.2) and (A.3) reduce to

$$\Phi_l\{\varphi_{\parallel}\} = \int \left\{ \frac{\hbar^2}{2Mr^2} \left[ \frac{d^2}{dr^2} \frac{r^2 \varphi_{\parallel}}{\sqrt{l(l+1)}} + \sqrt{l(l+1)}\, \varphi_{\parallel} \right]^2 - V(r) \left[ \varphi_{\parallel}^2 + \frac{1}{r^2} \left( \frac{d}{dr} \frac{r^2 \varphi_{\parallel}}{\sqrt{l(l+1)}} \right)^2 \right] \right\} r^2 dr, \qquad (A.8)$$

$$\int \left\{ \frac{1}{r^2} \left( \frac{d}{dr} \frac{r^2 \varphi_{\parallel}}{\sqrt{l(l+1)}} \right)^2 + \varphi_{\parallel}^2(r) \right\} r^2 dr = 1. \qquad (A.9)$$

In a potential of the Coulomb type the lowest level is six-fold degenerate (with account of the spin) with $l = 1$, as expected for a vector particle. Account of the spin-orbit interaction in the band structure would split it into four-fold and doubly degenerate levels with $j = 3/2$ and $j = 1/2$, as is observed at the shallow levels in Ge[3].

Using the two-parameter family of functions $\varphi_{\parallel}(\mathbf{r}) = A(\alpha, \gamma)(1 + 2\gamma\alpha r) \exp(-\alpha r)$, we can obtain for the energy of the ground state in a Coulomb field

$$\varepsilon_0 \approx +1.4 \frac{Z^2 e^4 M}{2\hbar^2} - \Delta, \quad \alpha = 1.7 \frac{Me^2}{\hbar^2}, \quad \gamma = \frac{1}{4}. \qquad (A.10)$$

The solution obtained from (A.8), as in (26), describes essentially the states of heavy holes, as demonstrated at least by the absence of the mass m from these equations. In other words, $\varphi_V$, like any other vector, can be resolved into solenoidal (heavy holes) and potential (light holes) components:

$$\varphi_v(\mathbf{r}) = \operatorname{rot} \mathbf{B}(\mathbf{r}) + \operatorname{grad} P(\mathbf{r}), \qquad (A.11)$$

and the condition div $\varphi_V = 0$ excludes the possible presence of heavy holes, while the contribution of the states of the conduction band also vanishes by virtue of the first equation of (24). However, unlike (26), the relation (A.7) is approximate, so that generally speaking it is not compatible with the exact equations for $\varphi_{\parallel}(\mathbf{r})$ and $\varphi_{\perp}(\mathbf{r})$, i.e., div $\varphi_V$ is not rigorously equal to zero in this case and is a small quantity of order higher than $\sqrt{m/M}$.

Using the solution obtained as a zeroth approximation, we obtain the component $\varphi_c(\mathbf{r})$. To this end, multiplying the second equation of (24) by $-i\hbar\sqrt{\Delta/m}$ div, we get

$$\frac{\hbar^2}{2m} \nabla^2 \varphi_c(\mathbf{r}) + \frac{[\varepsilon - V(\mathbf{r})]^2 - \Delta^2}{2\Delta} \varphi_c(\mathbf{r}) = -\frac{i\hbar}{2\sqrt{m\Delta}} \varphi_v(\mathbf{r}) \operatorname{grad} V(\mathbf{r}). \qquad (A.12)$$

For a spherically symmetrical potential, the right half reduces by virtue of (A.5) and (A.6) to $\varphi_{\parallel}(\mathbf{r})[dV(\mathbf{r})/d\mathbf{r}]\, Y_{lm}(\mathbf{n})$, and therefore $\varphi_c(\mathbf{r}) = \varphi_c(\mathbf{r})\, Y_{lm}(\mathbf{n})$. The radial function $\varphi_c(\mathbf{r})$ can be expressed in the usual fashion in terms of the solutions of the homogeneous radial equation

$$\frac{\hbar^2}{2m} \left\{ \frac{1}{r^2} \frac{d}{dr} \left( r^2 \frac{df}{dr} \right) - \frac{l(l+1)}{r^2} f \right\} + \frac{[\varepsilon - V(r)]^2 - \Delta^2}{2\Delta} f = 0, \qquad (A.13)$$

one of which, $f^{(0)}(\mathbf{r})$, is regular at zero, and the other, $f^{(\infty)}(\mathbf{r})$, is regular at infinity:

$$\varphi_c(r) = \frac{f^{(0)}(r)}{\hbar s} \int_r^{\infty} \varphi_{\parallel}(r') \frac{dV(r')}{dr'} f^{(\infty)}(r') r'^2 dr' + \frac{f^{(\infty)}(r)}{\hbar s} \int_0^r \varphi_{\parallel}(r') \frac{dV(r')}{dr'} f^{(0)}(r') r'^2 dr'. \qquad (A.14)$$

Compared with the rapidly decreasing $\varphi_{\parallel}(\mathbf{r})$ and $\varphi_{\perp}(\mathbf{r})$, the functions $f^{(0)}(\mathbf{r})$ and $f^{(\infty)}(\mathbf{r})$ are slowly varying. The corresponding characteristic lengths are of the order $\hbar\sqrt{\Delta/m(\Delta^2-\epsilon^2)}$, i.e., $\sim\sqrt{M/m}\,\alpha^{-1}$. Therefore at large distances from the center ($\gg\alpha^{-1}$) the main contribution to the total wave function is made by the second term of (A.14) which in this region is of the form

$$\varphi_c(r)\approx Cf^{(\infty)}(r),\quad C=\frac{1}{\hbar s}\int_0^\infty \varphi_{\parallel}(r)\frac{dV(r)}{dr}f^{(0)}(r)\,r^2\,dr. \tag{A.15}$$

The smallness of C is ensured by the fact that in the integral of (A.15) the function $f^{(0)}(\mathbf{r})$, which is normalized to a volume on the order of $(M/m\alpha^2)^{3/2}$, is integrated over a much smaller volume $\sim\alpha^{-3}$, determined by the decrease of $\varphi_{\parallel}(\mathbf{r})$. For the Coulomb case under consideration we have for $l=1$

$$\varphi_c(r)\approx\frac{Ze^2}{\hbar s}\frac{\gamma+3}{2}\frac{\Gamma(\gamma+1)\,\Gamma(1/2-\mu+\gamma)}{\Gamma(2\gamma+1)}\left(\frac{2\varkappa\hbar^2}{1.7\,Ze^2M}\right)^{\gamma+1}W_{\mu,\gamma}(2\varkappa r)$$

$$=\frac{Ze^2}{\hbar s}\frac{\gamma+3}{2}\frac{\Gamma(\gamma+1)\,\Gamma(1/2-\mu+\gamma)}{\Gamma(2\gamma+1)}$$

$$\times\left[2\frac{m}{M}\frac{\Delta^2-\varepsilon^2}{\Delta(\varepsilon_0+\Delta)}\right]^{(\gamma+1)/2}W_{\mu,\gamma}(2\varkappa r);$$

$$\mu=-\alpha\sqrt{\frac{\varepsilon^2}{\Delta^2-\varepsilon^2}},\quad \gamma^2=\left(l+\frac{1}{2}\right)^2-\alpha^2,$$

$$\varkappa=\sqrt{\frac{m(\Delta^2-\varepsilon^2)}{\hbar^2\Delta}}. \tag{A.16}$$

Here $W_{\mu,\gamma}(2\varkappa r)$ is the Whittaker function. The relative amplitude of this component is of order m/M when $\epsilon_0\sim\Delta$, i.e., it is sufficiently small. However, it decreases much more slowly than $\varphi_{\parallel}(\mathbf{r})$ at large distances.

Let us consider, finally, the case when $V(\mathbf{r})$ is so large that the energy of the hole ground state $\epsilon_0$ is larger than $\Delta$, i.e., it falls in the conduction band. At first glance the variational procedure employed is not applicable at all in this case, since the coefficient preceding $|\operatorname{div}\varphi_V|^2$ in (A.3) is not positive definite. This is obviously due to the fact that there are arbitrarily high conduction band states (in the scheme under consideration), so that the sought hole state can certainly not correspond to the absolute minimum of the functional. But we are seeking the lowest hole state. Consequently, it should be orthogonal in the zeroth approximation to all the conduction-band states, which is equivalent to requiring that $\varphi_c(\mathbf{r})=0$, i.e., $\operatorname{div}\varphi_V=0$. Thus, the problem reduces again to variation of the approximate functional (A.8) and its solution is given by (A.10). The small contribution of the conduction-band states is also expressed by (A.14)—(A.16), in which, however, $f^{(0)}(\mathbf{r})$, $f^{(\infty)}(\mathbf{r})$, and $W_{\mu,\gamma}$ are already functions of the continuous spectrum. The virtual levels thus obtained should play an important role in the capture of carriers by the center or in photoconductivity, and also give the resonance scattering of the conduction electrons.

[1] W. Kohn and J. M. Luttinger, Phys. Rev. **97**, 869 (1955) and **98**, 915 (1955).

[2] W. Kohn and D. Schechter, Phys. Rev. **99**, 1903 (1955).

[3] W. Kohn, Sol. State Phys. **5**, 257 (1957).

[4] E. O. Kane, J. Phys. Chem. Sol. **1**, 249 (1957).

[5] L. V. Keldysh, JETP **33**, 994 (1957), Soviet Phys. JETP **6**, 763 (1958).

[6] L. D. Landau and E. M. Lifshitz, Quantum Mechanics, Pergamon, 1958.

[7] Roth, Lax, and Zwerdling, Phys. Rev. **114**, 90 (1959).

[8] A. I. Akhiezer and V. B. Berestetskiĭ, Kvantovaya elektrodinamika (Quantum Electrodynamics), 1953, Sec. 12.

[9] M. Lax, Phys. Rev. **119**, 1502 (1960).

Translated by J. G. Adashko
62