Impurity States in Zero Gap Semiconductors

B. L. GEL'MONT AND M. I. D'YAKONOV

A. F. Ioffe Physico-technical Institute, USSR Academy of Sciences

Submitted June 3, 1971

Zh. Eksp. Teor. Fiz. **62**, 713–724 (February, 1972)

It is shown that quasi-discrete acceptor states exist in a semiconductor with zero forbidden gap, provided the hole mass m_p is much larger than the electron mass m_n. The relative width of the ground level is of the order of $(m_n/m_p)^{3/2}$. In this connection the correction to the density of states due to the donors is a smooth function of the energy. The presence of quasi-discrete acceptor levels against the background of the continuous spectrum of an electron in the conduction band leads to a number of interesting features in the thermodynamic and kinetic properties of such semiconductors. In particular it is shown that the addition of acceptors to an n-type semiconductor does not lead to compensation at low temperatures.

1. INTRODUCTION

THE energy spectrum of electrons in semiconductors having the structure of diamond or of zincblende can be four-fold degenerate at the point $\mathbf{p} = 0$ ($\mathbf{p}$ denotes the quasimomentum). This point corresponds to the top of the valence band in Ge, Si, InSb, GaAs, etc. In the crystals HgTe and α-Sn the point of degeneracy occurs on the boundary of the valence band and of the conduction band, so that these semiconductors have zero forbidden gaps. The role of impurities in such semiconductors has not been investigated before.

In semiconductors with zero forbidden gaps, the acceptor levels occur in the continuous spectrum corresponding to an electron in the conduction band; therefore, strictly speaking, bound acceptor states do not exist (the same applies to donor states whose energies fall into the continuous spectrum of the valence band). One can only expect a certain change of the density of states for energies of the order of the Bohr energy of the impurity state. We shall demonstrate that if the masses of the hole and electron are markedly different, then the broadening of the level corresponding to the heavy mass is small, that is, the level turns out to be quasi-discrete. The relative width of the level corresponding to the ground state is proportional to the ratio of the electron mass to the hole mass raised to $^3/_2$ power. The presence of a quasi-discrete level in the background of the continuous spectrum should manifest itself in the dependence of the kinetic properties of the system on temperature, electric field, etc.

The problem of determining the position of the quasi-discrete acceptor level and its width in a semiconductor with zero forbidden gap is solved in the present article. For definiteness we assume the hole to be heavier by far than the electron. Therefore we talk about a quasi-discrete acceptor level, since for such a ratio of the masses the donor levels are broadened by an amount which is comparable with the distance between them, so that it is impossible to talk about any kind of bound states of an electron on a donor. Therefore donors should be ionized at arbitrary temperatures whereas acceptors are neutral at low temperatures.

In the zero-gap semiconductors known to us, the mass of the hole is larger than the electron mass. In the case of the opposite mass ratio, the donor levels will be quasi-discrete and all of the conclusions of the present work will be applicable to them.

2. BASIC EQUATIONS

In the presence of the point of degeneracy $\mathbf{p} = 0$, in the effective mass approximation the Hamiltonian has the form (see[1])

$$H=\frac{1}{2m_0}\left[\left(\gamma_1+\frac{5}{2}\gamma\right)\hat{p}^2-2\gamma(\mathbf{p}\hat{\mathbf{J}})^2\right]+V(\mathbf{r}), \tag{1}$$

where m_0 is the mass of the free electron; $\hat{\mathbf{p}}$ is the momentum operator; J_x, J_y, and J_z are the numerical 4×4 matrices for the components of the angular momentum corresponding to the value $J = ^3/_2$; $\gamma = (2\gamma_2 + 3\gamma_3)/5$; γ_1, γ_2, and γ_3 are the band parameters introduced by Luttinger,[1] and $V(\mathbf{r})$ is the impurity potential.

The Hamiltonian (1) is written down neglecting the undulating nature of the constant energy surfaces. This approximation works very well, for example, in the interaction between the p-band (from which the four bands considered by us, which are split-off as a consequence of the spin-orbit interaction, stem) and the nearest s-band.

The dependence of the energy on the momentum has the form

$$E_\pm(\mathbf{p})=\frac{\gamma_1\pm 2\gamma}{2m_0}p^2. \tag{2}$$

From this it is clear how the band parameters γ_1 and γ are related to the effective masses. If $\gamma_1 + 2\gamma$ and $\gamma_1 - 2\gamma$ have the same sign, then $E_\pm(p)$ corresponds to bands of light and heavy holes. In this case the masses of the light and heavy holes are given by $m_0(\gamma_1 + 2|\gamma|)^{-1}$ and $m_0(\gamma_1 - 2|\gamma|)^{-1}$ respectively. However, if $\gamma_1 + 2\gamma$ and $\gamma_1 - 2\gamma$ have opposite signs, then $E_\pm$ describe the conduction band and the valence band, which are degenerate for $\mathbf{p} = 0$. There is no forbidden gap then. For the sake of definiteness, we shall assume in what follows that $2\gamma > \gamma_1 > 0$. Then $m_n = m_0(\gamma_1 + 2\gamma)^{-1}$ is the electron mass and $m_p = m_0(2\gamma - \gamma_1)^{-1}$ is the hole mass.

The problem of finding the impurity states in a semiconductor with zero forbidden gap is related to the problem of determining the acceptor level in an ordinary semiconductor with a valence band which is four-fold degenerate for $\mathbf{p} = 0$. Such a problem in general presents considerable difficulty, since it is

necessary to solve a system of four equations in partial derivatives. The acceptor ground state energy was calculated in[2-7] by a variational method for specific semiconductors.

Neglecting the undulating nature and for a spherically symmetric impurity potential $V(r)$, one can substantially simplify the problem by using the fact that the Hamiltonian (1) commutes with the operator $\mathbf{F} = \mathbf{L} + \mathbf{J}$,[4] where $\mathbf{L} = \hbar^{-1}(\mathbf{r} \times \mathbf{p})$ is the orbital angular momentum operator. The energy levels are degenerate with respect to the projection M of the vector $\mathbf{F}$.

We shall seek the wave function in the form

$$\psi_{FM}(r, \theta, \varphi) = \sqrt{2F+1} \sum_l (-1)^{l-3/2+M} \cdot \times R_{Fl}(r) \sum_{m\mu} \begin{pmatrix} l & 3/2 & F \\ m & \mu & -M \end{pmatrix} Y_{lm}(\theta, \varphi)\chi_\mu, \quad (3)$$

where Y_{lm} are spherical functions and χ_μ denotes the eigenvector of the matrix J_z, $J_z\chi_\mu = \mu\chi_\mu$. The index μ takes the values $\pm 3/2, \pm 1/2$. The Wigner 3-j symbol is used in the expansion (3). For a given value of F ($F = 1/2, 3/2, 5/2 \ldots$) the quantity l may take the four values $l = F \pm 3/2$ and $l = F \pm 1/2$, with the exception of the case $F = 1/2$ when only the values $l = 1$ and 2 are possible.

One can derive two independent systems of equations for the radial functions R_{Fl} associated with a definite value of F. Let us cite one of these systems, relating the functions $R_{F,F+1/2}$ and $R_{F,F-3/2}$:

$$(\gamma_1 - 2\gamma\cos\alpha_F)P_F^+P_FR_{F+1/2} + 2\gamma\sin\alpha_F P_F^+P_{-F}R_{F-3/2} + \frac{2m_0}{\hbar^2}(E-V)R_{F+1/2} = 0,$$
$$(\gamma_1 + 2\gamma\cos\alpha_F)P_{-F}^+P_{-F}R_{F-3/2} + 2\gamma\sin\alpha_F P_{-F}^+P_FR_{F+1/2} + \frac{2m_0}{\hbar^2}(E-V)R_{F-3/2} = 0, \quad (4)$$

where

$$\cos\alpha_F = \frac{2F-3}{4F}, \quad \sin\alpha_F \geqslant 0, \quad P_F = \frac{d}{dr} + \frac{F+3/2}{r},$$
$$P_F^+ = \frac{d}{dr} - \frac{F-1/2}{r}. \quad (5)$$

We have omitted the first subscript F from the functions R. A second pair of equations, relating the functions $R_{F,F+3/2}$ and $R_{F,F-1/2}$, can be derived from Eqs. (4) and (5) by the substitution $F \to -(F+1)$ in the expression for $\cos\alpha_F$ and in the operators P_F and P_F^+.

For $F = 1/2$ one will have $R_{F-3/2} = R_{F-1/2} = 0$, $\sin\alpha_{1/2} = 0$, and in this case we obtain the following equation for the functions R_l

$$(\gamma_1 + 2\gamma)\left(\frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} - \frac{l(l+1)}{r^2}\right)R_l + \frac{2m_0}{\hbar^2}(E-V)R_l = 0, \quad (6)$$

that is, the usual radial Schrödinger equation for an electron in p- and d-states ($l = 1, 2$).

Equations of the form (6) were used in[6,8,9] to determine the energies of the acceptor states in a semiconductor with a forbidden gap of finite width in the case when the valence band is four-fold degenerate at the center of the Brillouin zone.

3. THE POSITION AND WIDTH OF THE ACCEPTOR LEVELS

Equations (4) and (6) are valid (neglecting the undulating nature of the constant energy surfaces) both for degenerate valence bands and for a semiconductor with zero forbidden gap. Let us show that in the latter case the system (4) and the analogous system of equations for the functions $R_{F+3/2}$ and $R_{F-1/2}$ do not give bound states.

In fact, Eqs. (4) have four independent solutions which behave like $\exp(\pm ik_+r)$ and $\exp(\pm ik_-r)$ as $r \to \infty$, where

$$k_+^2 = \frac{2m_0E}{\hbar^2(\gamma_1+2\gamma)}, \quad k_-^2 = \frac{2m_0E}{\hbar^2(\gamma_1-2\gamma)}. \quad (7)$$

In the case of a degenerate valence band, both of the quantities k_+ and k_- are pure imaginary for levels located in the forbidden gap (impurity—acceptor), so that two damped and two increasing solutions exist as $r \to \infty$. On the other hand, as one can easily show, two independent solutions exist which are finite as $r \to 0$. From these for arbitrary E one can construct a linear combination, containing only one increasing exponential as $r \to \infty$. The coefficient associated with this increasing exponential vanishes for certain discrete values of E, which therefore correspond to the bound states.

However, in the case of a zero gap semiconductor, the quantities k_+^2 and k_-^2 have different signs for arbitrary E; therefore there is one increasing solution, one decreasing solution, and two oscillating solutions as $r \to \infty$. Just as previously, for arbitrary E one can form a solution which is finite as $r \to 0$ and does not increase as $r \to \infty$. This solution, however, contains two oscillating exponentials as $r \to \infty$; it is impossible to make the coefficients associated with these two terms simultaneously vanish by the choice of the single parameter E. Thus, bound states are not present for $F \geq 3/2$.

In the unusual case $F = 1/2$, Eq. (6) may give discrete levels (against the background of the continuous spectrum). If $V(r)$ is the Coulomb potential, then a hydrogen-like spectrum is obtained from Eq. (6) with the lowest level corresponding to the value of the principal quantum number $n = 2$ (since $l = 1, 2$). We note the asymmetry of Eq. (6) with respect to electrons and holes. For $\gamma_1 + 2\gamma > 0$ Eq. (6) gives bound states for the electron, if the impurity is a donor ($V(r) < 0$).

Let us emphasize that the possibility of classifying the states according to the quantum number F and the separation of Eq. (6) for $F = 1/2$ are due to the neglect of the anisotropy of the energy surfaces. Upon taking the undulations into account, in general we would not obtain any bound states.

Thus, in a zero-gap semiconductor the role of the impurities only reduces to a change in the density of states of the continuous spectrum. However, as was stated in the Introduction, in the case $m_p \gg m_n$ the change in the density of states caused by acceptors must have the form of narrow peaks, whose widths are small in comparison with the distance between them. In what follows we shall assume that $\beta = (2\gamma - \gamma_1)/(2\gamma + \gamma_1) = m_n/m_p \ll 1$. We shall show that in this connection the position of the quasi-discrete levels with respect to the top of the valence band coincides with the position of the acceptor levels in an ordinary semiconductor having a four-fold degenerate valence band, provided the ratio of the masses of the light and heavy holes is very small. The

width of the levels will be calculated to within a numerical coefficient.

Let us change in Eqs. (4) to dimensionless quantities, having chosen the Bohr radius a_p as the unit of length and the Bohr energy of a hole as the unit of energy

$$r = a_p\rho, \quad E = m_p e^4 \varepsilon / 2\hbar^2\varkappa^2, \quad a_p = \hbar^2\varkappa / m_p e^2,$$

where $\varkappa$ denotes the dielectric constant. Having taken $V(r)$ to be the Coulomb potential $V(r) = e^2(\varkappa r)^{-1}$, we rewrite Eq. (4) in the form

$$(P_F^+ P_F + 2/\rho - \varepsilon) R_{F+1/2} + (1+\beta)\sin(\alpha_F/2) P_F^+ Y_F = 0, \quad (8)$$

$$(P_{-F}^+ P_{-F} + 2/\rho - \varepsilon) R_{F-3/2} + (1+\beta)\cos(\alpha_F/2) P_{-F}^+ Y_F = 0, \quad (9)$$

where the new function Y_F is introduced in the following way:

$$-\beta Y_F = \sin(\alpha_F/2) P_F R_{F+1/2} + \cos(\alpha_F/2) P_{-F} R_{F-3/2}. \quad (10)$$

In Eqs. (8) and (9) it is impossible to pass directly to the limit $\beta = 0$. In order to realize this transition, side by side with the function Y_F we also introduce the function

$$X_F = \cos(\alpha_F/2) P_F R_{F+1/2} - \sin(\alpha_F/2) P_{-F} R_{F-3/2}. \quad (11)$$

In order to obtain a system of equations for the functions X_F and Y_F, let us divide Eqs. (8) and (9) by $2/\rho - \epsilon$ and then we operate on (8) from the left with the operator P_F, and on (9) with the operator P_{-F}. Here we use the relation

$$P_F P_F^+ = P_{-F} P_{-F}^+ = \Delta_{F-1/2} = \frac{d^2}{d\rho^2} + \frac{2}{\rho}\frac{d}{d\rho} - \frac{F^2 - 1/4}{\rho^2}.$$

Then we obtain

$$\left(\Delta_{F-1/2} + \frac{2}{\rho} - \varepsilon\right) X_F + \frac{2}{\rho(2-\varepsilon\rho)}\left[\left(\frac{d}{d\rho} + \frac{1}{2\rho}\right) X_F - \frac{F}{\rho}(\cos\alpha_F X_F + \sin\alpha_F Y_F)\right] = 0, \quad (12)$$

$$\left[\Delta_{F-1/2} + \beta\left(\varepsilon - \frac{2}{\rho}\right)\right] Y_F + \frac{2}{\rho(2-\varepsilon\rho)}\left[\left(\frac{d}{d\rho} + \frac{1}{2\rho}\right) Y_F + \frac{F}{\rho}(\cos\alpha_F Y_F - \sin\alpha_F X_F)\right] = 0. \quad (13)$$

Putting $\beta = 0$ here, we arrive at a system of equations which is identical to that which describes the acceptor states in an ordinary semiconductor in the limit of an infinitesimal ratio of the masses of the light and heavy holes. The statement made above concerning the position of the quasi-discrete levels also follows from here. The dependence of the energy of the ground acceptor level ($F = 3/2$) on the ratio of the masses of the light and heavy holes was calculated numerically by the authors in[9], and it was shown that a small mass ratio corresponds to the limiting value $\epsilon_0 = 4/9$. According to what has been said, this same value also determines the position of the highest quasi-discrete level in a zero gap semiconductor, provided that $\beta \ll 1$.

Since the levels are strictly discrete for $\beta = 0$, then it is clear that their width should be small for small values of β. One can determine this width if the dependence of the phase of the wave functions on the energy is known at infinity. As is evident from Eqs. (12) and (13) the functions X_F and Y_F behave in the following manner as $\rho \to \infty$:

$$Y_F \sim \rho^{-1}\sin(\rho\sqrt{\beta\varepsilon} + \eta_F), \quad X_F = \frac{2F\sin\alpha_F}{\varepsilon^2\rho^3} Y_F. \quad (14)$$

From formulas (10) and (11) we find the asymptotic behavior of the functions $R_{F+1/2}$ and $R_{F-3/2}$:

$$R_{F+1/2} \sim \sin(\alpha_F/2)(\beta/\varepsilon)^{1/2}\cos(\rho\sqrt{\beta\varepsilon} + \eta_F), \quad R_{F-3/2} = \operatorname{ctg}(\alpha_F/2) R_{F+1/2}. \quad (15)$$

One can express the change $\Delta\rho_F(\epsilon)$ of the density of states with quantum numbers F and M, per acceptor, in terms of the phase η_F (see, for example, Kittel's book[10]):

$$\Delta\rho_F(\varepsilon) = \frac{1}{\pi}\frac{d\eta_F}{d\varepsilon}. \quad (16)$$

Thus, it is necessary to calculate the phase η_F for energies which are close to the energies ϵ_{Fn} of the discrete levels, which are determined by the system of equations (12) and (13) for $\beta = 0$. One can do this by utilizing the following property. There are two characteristic distances governing the variation of the wave function: The Bohr radius a_p of a hole (the corresponding region is $\rho \lesssim 1$) and the wavelength $\sim\sqrt{a_p a_n}$ of an electron whose energy is near that of a quasi-discrete level ($a_n = \hbar^2\kappa/m_n e^2$ denotes the electron's Bohr radius, the corresponding region is $\rho \sim \beta^{-1/2}$). Owing to the strong difference between these lengths, an intermediate region $1 \ll \rho \ll \beta^{-1/2}$ exists where one can "join together" the solution oscillating at infinity with the asymptotic solution of the system (12) and (13) for $\beta = 0$ and $\rho \gg 1$.

The total system of Eqs. (12) and (13) has a continuous spectrum of eigenvalues in contrast to the limiting system with $\beta = 0$, which has discrete eigenvalues ϵ_{Fn}. Out of the four linearly independent solutions of the limiting system, two solutions are finite at the origin. Out of these solutions, for any arbitrary value of ϵ one can put together a superposition such that the obtained solution does not have any increasing exponentials at infinity. Then for $\rho \gg 1$ one can neglect the decreasing exponential, and the function Y_F is determined from the equation $\Delta_{F-1/2} Y_F = 0$. It will contain terms which decrease like $\rho^{-(F+1/2)}$ and increase like $\rho^{F-1/2}$, and the increasing term is absent only for $\epsilon = \epsilon_{Fn}$. For a small deviation of ϵ from ϵ_{Fn} the coefficient associated with $\rho^{F-1/2}$ is proportional to the relative deviation $\delta\epsilon/\epsilon = (\epsilon - \epsilon_{Fn})/\epsilon$. Thus, in the interval $1 \ll \rho \ll \beta^{-1/2}$ the function Y_F has the form

$$Y_F = C_1\rho^{-(F+1/2)} + C_2(\delta\varepsilon/\varepsilon)\rho^{F-1/2}, \quad (17)$$

where C_1 and C_2 are constants of the same order of magnitude, which are finite for $\beta = 0$. The coefficient associated with $\rho^{F-1/2}$ contains an additional term proportional to the small parameter β and determining the shift in the position of the level. We shall assume that the energy ϵ_{Fn} is determined with this shift taken into consideration.

For $\rho \sim \beta^{-1/2}$, in Eq. (13) side by side with the term $\Delta_{F-1/2} Y_F$ the term $\beta\epsilon Y_F$ also becomes important. Therefore one can obtain a solution, valid over the entire region $\rho \gg 1$ (also including the region $\rho \gg \beta^{-1/2}$), from the equation

$$(\Delta_{F-1/2} + \beta\varepsilon) Y_F = 0. \quad (18)$$

The solution of this equation, which goes over into

expression (17) for $\rho \ll \beta^{-1/2}$, is given by

$$Y_F = C_1 \cdot 2^{-F}(\beta\varepsilon)^{F/2}\Gamma(1-F)\rho^{-1/2}J_{-F}(\rho\sqrt{\beta\varepsilon}) + C_2(\delta\varepsilon/\varepsilon)\cdot 2^F(\beta\varepsilon)^{-F/2}\Gamma(1+F)J_F(\rho\sqrt{\beta\varepsilon}), \quad (19)$$

where J_F denotes a Bessel function of order F and Γ denotes the Gamma function. By using the asymptotic form of the Bessel functions for large values of the argument and comparing expression (19) with (14) we find

$$\operatorname{tg}(\eta_F + \pi F/2 + \pi/4) = 2\delta\varepsilon/\varepsilon_{Fn}\nu_{Fn}, \quad (20)$$

where

$$\nu_{Fn} = (-1)^{F+1/2}\frac{\Gamma(1-F)}{\Gamma(1+F)\cdot 2^{2F-1}}\frac{C_1}{C_2}(\beta\varepsilon_{Fn})^F \quad (21)$$

denotes the relative half-width of the quasi-discrete level ϵ_{Fn}. In fact, from formulas (16) and (20) we obtain the following expression for the change $\Delta\rho_F$ of the density of states near the energy ϵ_{Fn}:

$$\Delta\rho_F = \frac{1}{\pi\varepsilon_{Fn}}\frac{\nu_{Fn}/2}{(\nu_{Fn}/2)^2 + (\delta\varepsilon/\varepsilon_{Fn})^2}. \quad (22)$$

Thus, the level widths are small in comparison with the distance between them. For the ground state ($F = 3/2$) the relative width is of the order of $(m_n/m_p)^{3/2} \ll 1$.

The numerical coefficient C_1/C_2, determined by the asymptotic behavior of the function Y_F for $\beta = 0$ and $\rho \gg 1$, enters into formula (21). Let us show that one can express the coefficient C_2 in terms of the coefficient C_1, which determines the asymptotic form of the bound state wave function, existing for $\beta = 0$. In order to show this, we utilize a relation which one can derive from the system of Eqs. (8) and (9) by a method analogous to that used in the book[11]:

$$-X_F\frac{\partial x_F}{\partial\varepsilon} + x_F\frac{\partial X_F}{\partial\varepsilon} - Y_F\frac{\partial y_F}{\partial\varepsilon} + y_F\frac{\partial Y_F}{\partial\varepsilon} = \frac{1}{\rho^2}\int_0^\rho (R^2_{F+1/2} + R^2_{F-3/2})\rho^2 d\rho, \quad (23)$$

where

$$x_F = \cos(\alpha_F/2)R_{F+1/2} - \sin(\alpha_F/2)R_{F-3/2},$$
$$y_F = \sin(\alpha_F/2)R_{F+1/2} + \cos(\alpha_F/2)R_{F-3/2}.$$

If we pass to the limit $\rho \to \infty$ in relation (23), then we are able to determine the desired relation between C_2 and C_1. For $\beta = 0$ the asymptotic behavior of the functions R has the following form:

$$R_{F+1/2} = -\frac{2F\sin(\alpha_F/2)}{\varepsilon}\frac{C}{\rho^{F+3/2}} + \frac{2F\sin\alpha_F\cos(\alpha_F/2)}{(2F-1)\varepsilon^2}\frac{\delta\varepsilon}{\varepsilon}C_2\rho^{F-5/2},$$
$$R_{F-3/2} = \frac{2F\sin\alpha_F\sin(\alpha_F/2)}{(2F+1)\varepsilon^2}\frac{C_1}{\rho^{F+5/2}} + \frac{2F\cos(\alpha_F/2)}{\varepsilon}\frac{\delta\varepsilon}{\varepsilon}C_2\rho^{F-3/2}. \quad (24)$$

Using formulas (11), (17), and (23) we find

$$C_2 = -\varepsilon^2(2FC_1)^{-1}, \quad (25)$$

here we regard the wave functions for $\beta = 0$ and $\epsilon = \epsilon_{Fn}$ to be normalized according to the usual condition

$$\int_0^\infty (R^2_{F+1/2} + R^2_{F-3/2})\rho^2 d\rho = 1.$$

Finally, the relative half-width of the quasi-discrete levels is expressed in terms of the single constant C_1:

$$\nu_{Fn} = 2^{2(1-F)}\pi^{-1}\Gamma^2(1-F)C_1^2\varepsilon_{Fn}^{F-2}\beta^F. \quad (26)$$

The half-width of the ground acceptor level amounts to

$$\nu_0 = 2C_1^2\varepsilon_0^{-1/2}\beta^{3/2}. \quad (27)$$

Numerical calculation on an electronic computer gives $\epsilon_0 = 4/9$ and $C_1 \approx 2$. For HgTe $m_n = 0.026\ m_0$, $m_p = 0.4\ m_0$, $\kappa = 20$,[12] $\beta \approx 0.065$, so that in this case $E_0 \approx 6$ MeV and $\nu_0 \approx 0.2$.

Let us compare the asymptotic behavior of the wave function of the acceptor level in the case of a zero gap semiconductor and in the case of a semiconductor having a finite forbidden gap, when the ratio of the masses of the light and heavy holes is small. In the latter case expression (24) determines the behavior of the wave function in the region $1 \ll \rho \ll (m_h/m_l)^{1/2}$ (m_h and m_l denote the masses of the heavy and light holes). Upon taking the finite value of the ratio of these masses into account, instead of expression (19) we obtain the following asymptotic form of the function Y_F for $\rho \gg 1$:

$$Y_F = C_1(-1)^{F+3/2}\frac{2^{1-F}}{\pi}\left(\frac{m_l\varepsilon}{m_h}\right)^{F/2}\frac{\Gamma(1-F)}{\sqrt{\rho}}K_F\left(\rho\sqrt{\frac{\varepsilon m_l}{m_h}}\right), \quad (28)$$

where K_F denotes a Macdonald function of order F. For $r \gg \sqrt{a_l a_h}$, where a_l and a_h denote the Bohr radii of the light and heavy holes, we obtain the following result from Eqs. (10), (11), and (28):

$$R_{F+1/2}\cos\left(\frac{\alpha_F}{2}\right) = R_{F-3/2}\sin\left(\frac{\alpha_F}{2}\right) = C_1\frac{2^{1/2-F}\sqrt{\pi}}{\Gamma(F)}\left(\frac{\varepsilon m_l}{m_h}\right)^{F/2+1/2}\frac{a_h}{r}\exp\left(-\frac{r\sqrt{\varepsilon}}{\sqrt{a_h a_l}}\right). \quad (29)$$

Thus, in an ordinary semiconductor the wave function of the acceptor state has a long exponential tail with a characteristic scale of the order of $\sqrt{a_l a_h}$. The presence of such a tail may manifest itself in kinetic phenomena, in which the overlap of the wave functions of neighboring acceptors is essential.

4. INFLUENCE OF DONORS ON THE DENSITY OF STATES

In contrast to acceptors, donors in a zero gap semiconductor with $m_p > m_n$ do not give any quasi-discrete levels with $F \geq 3/2$. The correction to the density of states caused by the donors is a smooth function of the energy. One can show this by calculating, just as in the preceding section, the phase of the wave functions for the attractive Coulomb potential. We shall not present the corresponding rather cumbersome calculations here. One can easily verify this conclusion if the potential V(r) is taken to be a spherically symmetric potential well. In this case the problem can be solved exactly for any arbitrary ratio of the masses, since the equations for the functions X_F and Y_F can be separated and reduce to the usual Schrödinger equations for electron and hole. The phase is determined by the conditions for continuity of the functions $R_{F+1/2}$, $R_{F-3/2}$, X_F, and Y_F at the boundary of the well.

We note that one can estimate the order of magnitude of the level width $\Gamma = \nu E_0$ by using the ordinary formula, without taking the spinor nature of the wave functions into consideration:

$$\Gamma \sim \hbar^{-1}2\pi\int d^3\mathbf{p}|V_{0\mathbf{p}}|^2\delta(E_p - E_0),$$

where $E_p = p^2/2m_p$, $E_0 \sim m_n e^4/\kappa^2\hbar^2$ denotes the energy level, V_{0p} denotes the matrix element of the potential

between ordinary Coulomb functions corresponding to the bound state of an electron on a donor and the state of an electron in the valence band with momentum p far away from the center. Since $pa_p \sim (m_n/m_p)^{1/2} \ll 1$, then it is necessary to take into consideration the distortion of the wave functions of the continuous spectrum by the Coulomb potential. Calculation according to this formula gives $\nu \sim 1$ in accordance with what was stated above.

One can estimate the width of the acceptor levels in precisely the same way. In order to do this one should set $E_p = p^2/2m_n$ and $E_0 \sim m_p e^4/\kappa^2\hbar^2$. Then we obtain $\nu \sim (m_n/m_p)^{3/2}$, which agrees with the results of the preceding section.

5. INFLUENCE OF THE ELECTRON-ELECTRON INTERACTION

We have solved the problem of the position and width of impurity levels within the framework of the one-electron problem. One can briefly formulate the results in the following manner. The wave function of an electron, located in the field of an impurity atom in a semiconductor with zero forbidden gap, consists of two parts, a part which is attenuated and an oscillating part at large distances from the impurity. If the masses of the electron and hole are comparable, then the amplitude of the oscillating part is comparable with the value of the wave function near the impurity, and therefore the electron has a large probability to become free. In this connection the levels are markedly broadened. The presence of an impurity atom leads to a small (nonresonance) change in the density of states.

However, if the hole mass (for the sake of definiteness) is much larger than the electron mass, then as usual donor bound states do not exist (the broadening of the levels is of the order of the distance between them), whereas the acceptor levels are quasi-discrete with a relative width proportional to $(m_n/m_p)^{3/2}$ for the ground level.

Near the maxima of the density of states, one can describe the wave function in the following manner. At first it decreases over the Bohr radius a_p of a hole; for $r > a_p$ the exponential decrease is replaced by a decrease according to a power law (see formula (24)). For $r > \sqrt{a_p a_n}$ the wave function oscillates with a period $\sim \sqrt{a_p a_n}$; however, if $\beta \ll 1$ then the amplitude of the oscillating part is small in comparison with the value of the function at the origin.

Now let us qualitatively discuss the role of the electron-electron interaction. At absolute zero temperature and in the absence of donors, when there are no free carriers, the interelectron interaction is realized by means of virtual interband transitions. In this connection the screening of the impurity charge occurs over distances which are larger than the electron's Bohr radius a_n. Such a result gives, for example, an estimate of the dielectric constant within the framework of the random phase approximation.[13] In general the expression for the dielectric constant in[13] is valid only for wave vectors which are larger than the reciprocal of the electron Bohr radius. However, the results of this work can be used in order to estimate the characteristic radius of screening. Since the Bohr radius of the hole, inside of which the major part of the wave function is concentrated, is small in comparison with the electron Bohr radius, then the interelectron interaction has very little influence on the position of the acceptor level. The period of the oscillations of the tail of the wave function is also small in comparison with the distance over which screening is realized.

One can reach a similar conclusion based on the results of[14], in which it was shown that the electron-electron interaction leads to a strong renormalization of the dependence of the energy on the momentum in the range of energies of the order of an electron Bohr energy. However, in the case of a large ratio of the masses of hole and electron, the Bohr energy of the hole is considerably larger than the electron Bohr energy, and therefore the problem of the position and width of the acceptor levels can be solved within the framework of the one-electron approximation. The electron-electron interaction has an appreciable influence only on the position and width of the highly excited levels, for which the radius of the state is comparable with the Bohr radius of the electron.

Screening of the impurity charge appears upon increasing the temperature or the concentration of donors. The screening radius decreases with an increase of the electron concentration; however, it is comparable with the Bohr radius of a hole only when the Fermi level $\mu \sim E_0 \cdot (m_p/m_n)^3$, that is, it lies considerably above the ground acceptor level.

The question about discrete donor levels with $F = 1/2$, described by Eq. (6), remains open. These levels are discrete only in the approximation of neglecting the undulating nature of the constant-energy surfaces. Taking account of the undulating nature leads to the result that they become quasi-discrete. In addition, for these states the electron-electron interaction turns out to be essential (the radius of screening is of the order of the radius of these states). One can therefore assume that in reality these levels will be markedly broadened.

6. DEPENDENCE OF THE ELECTRON CONCENTRATION ON THE TEMPERATURE

The existence of sharp acceptor levels in the background of the continuous spectrum of the conduction band leads to a number of interesting features in the thermodynamics and kinetics of the free carriers in a semiconductor with zero forbidden gap.

For a large difference in the masses of the electrons and holes, the donors and acceptors in such a semiconductor play substantially different roles. Since, as is shown above, bound states of the electrons on donors do not exist, then the concentration of free electrons is equal to the concentration N_D of donors even for $T = 0$ (even for a low concentration of donors). On the other hand, the acceptor levels are sharp; therefore, for any noticeable ionization of the acceptors it is necessary that the temperature should be of the order of the ground state energy E_0 of the acceptor. (We consider only the case of a sufficiently low concentration of acceptors, $N_A a_p^3 \ll 1$, so that the wave functions of neighboring acceptors overlap very little.

For HgTe this condition means $N_A \ll 10^{20}$ cm^{-3}). Since the forbidden energy gap is absent, free carriers appear at arbitrarily low temperature even in the absence of donors. Thus, for $N_D = 0$ the intrinsic conductivity appears at lower temperatures than the impurity conductivity associated with ionization of the acceptors.

If there are both donors and acceptors, then at low temperature a semiconductor with zero forbidden gap will be n-type, independently of the relation between their concentrations. Compensation does not occur at low temperatures, provided the Fermi level lies below the acceptor level. With an increase of the temperature an unusual compensation begins, and at high temperature the semiconductor becomes p-type, if the concentration of acceptors exceeds the concentration of donors. During the course of this compensation, in a certain temperature interval the concentration of free electrons may decrease with increasing temperature.

Let us find the number of free carriers as a function of the temperature T for an acceptor concentration N_A and a donor concentration N_D. We shall assume the acceptor levels to be sharp. One can determine the fermi level μ from the equation

$$p + N_D = n + N_A \left[1 + \frac{1}{4}\exp\left(\frac{E_0 - \mu}{T}\right)\right]^{-1}, \qquad (30)$$

where E_0 denotes the position of the ground acceptor level, and p and n denote the concentrations of holes and electrons. The factor $\frac{1}{4}$ takes account of the fourfold degeneracy of the ground level.

Let us consider the most interesting case, when $\mu > T$, so that the electrons are degenerate but the holes are nondegenerate. Then from Eq. (30) we obtain

$$\frac{3\sqrt{\pi}}{4}\left(\frac{m_p}{m_n}\right)^{3/2} e^{-x} - x^{3/2} = \left(\frac{E_0}{T}\right)^{5/2}\left\{n_A\left[1 + \frac{1}{4}\exp\left(\frac{E_0}{T} - x\right)\right]^{-1} - n_D\right\}, \qquad (31)$$

$$p = \frac{3\sqrt{\pi}}{4}\left(\frac{m_p}{m_n}\right)^{3/2}\left(\frac{T}{E_0}\right)^{3/2} e^{-x} N_0, \quad n = \left(\frac{xT}{E_0}\right)^{3/2} N_0, \qquad (32)$$

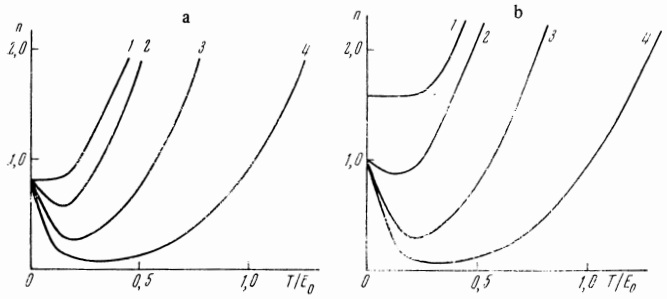
where

$$n_A = N_A / N_0,\ n_D = N_D / N_0,\ N_0 = (2m_n E_0)^{3/2} / 3\pi^2\hbar^3,\quad x = \mu / T.$$

The characteristic concentration N_0 is chosen in such a manner that for $n = N_0$ the Fermi level would coincide with the energy of the acceptor level. For HgTe $N_0 \approx 10^{16}$ cm^{-3}.

From Eqs. (31) and (32) one can determine the concentration of electrons and holes as a function of the temperature. As an example, we present curves calculated on an electronic computer for the case when $(m_p/m_n)^{3/2}(3\sqrt{\pi}/4) = 100$, which corresponds to the mass ratio $m_p/m_n \approx 18$, which is close to the value which exists in HgTe. The results of calculations for various values of the dimensionless concentrations n_A and n_D are shown in the Figures.

Figure a corresponds to the case $n_D = 0.8$, so that for T = 0 the Fermi energy is somewhat smaller than the ground state energy E_0 of the acceptor. For a small concentration of acceptors (curve 1) at low temperatures, the concentration of electrons remains constant, and subsequently begins to monotonically increase with increasing temperature owing to the formation of electron-hole pairs. With an increase of the concentration of acceptors, a minimum appears on the curve showing the dependence of n on T. In the region of low temperatures, the decrease in the concentration of free electrons with increasing temperature is due to the capture of electrons by acceptors. Upon further increase of the temperature, the role of thermal generation of pairs increases in proportion to the population of acceptors, which leads to an increase in the concentration of electrons.



Results of a numerical calculation of the dependence of the free electron concentration n on the temperature T for various acceptor concentrations: 1) $n_A = 0$, 2) $n_A = 1$, 3) $n_A = 10$, 4) $n_A = 50$ and donor concentrations a) $n_D = 0.8$ and b) $n_D = 1.6$. All concentrations are expressed in units of N_0.

Figure b corresponds to the case when for T = 0 the Fermi energy is higher than the acceptor level ($n_D = 1.6$). In contrast to the preceding case now the addition of acceptors leads to a partial compensation even at T = 0, when with an increase of the concentration of acceptors the concentration of free electrons decreases to the value $n = N_0$ for $n_A = n_D - 1$. A further increase of n_A does not lead to any decrease in the concentration of free electrons since their capture by acceptors becomes energetically unfavorable. If $n_A > n_D - 1$, then two competing processes are going on when the temperature is different from zero: The thermal hopping of free electrons to unoccupied acceptor levels and the thermal generation of electron-hole pairs. At low temperatures the first process dominates, and at higher temperatures the second process dominates. As a result, just as in Fig. a, a minimum appears in the curves n(T). Under these conditions a minimum also appears in the curve showing the temperature dependence of the conductivity; this minimum will be more pronounced due to the fact that the role of the impurity scattering increases upon ionization of the acceptors.

We note that such an unusual dependence of the free electron concentration and of the conductivity on the temperature permits us to anticipate that the current-voltage characteristics of a zero gap semiconductor will have a decreasing region. In fact, heating up the carriers by a field may, for the appropriate concentrations of donors and acceptors, lead to a drop in the concentration of electrons, and together with this a drop in the conductivity.

It is necessary to bear in mind that the electrons, captured by the acceptors, can begin to participate in the conductivity since the acceptor level is quasi-discrete and the lifetime of an electron in this level is finite. It is obvious that the mobility of these electrons is small.

The nonmonotonic nature of the temperature depend-

ence of the Hall constant and of the conductivity in certain samples of HgTe was observed in[15]. One can conjecture that this nonmonotonicity is related to acceptor levels in the conduction band.

[1]J. M. Luttinger, Phys. Rev. **102,** 1030 (1956).
[2]Walter Kohn and Daniel Schechter, Phys. Rev. **99,** 1903 (1955).
[3]W. Kohn, in Solid State Physics, edited by F. Seitz and D. Turnbull, Academic Press, 1957, Vol. 5, pp. 257–320.
[4]D. Schechter, J. Phys. Chem. Solids **23,** 237 (1962).
[5]K. S. Mendelsohn and H. M. James, J. Phys. Chem. Solids **25,** 729 (1964).
[6]V. I. Sheka and D. I. Sheka, Zh. Eksp. Teor. Fiz. **51,** 1445 (1966) [Sov. Phys. JETP **24,** 975 (1967)].
[7]K. S. Mendelsohn and D. R. Schultz, Phys. Status Solidi **31,** 59 (1969).
[8]Nunzio O. Lipari and A. Baldereschi, Phys. Rev. Lett. **25,** 1660 (1970).
[9]B. L. Gel'mont and M. I. D'yakonov, Fiz. Tekh. Poluprov. **5,** 2191 (1971) [Sov. Phys. Semicond. **5,** 1905 (1972)].
[10]Charles Kittel, Quantum Theory of Solids, Wiley, 1963.
[11]L. D. Landau and E. M. Lifshitz, Kvantovaya mekhanika (Quantum Mechanics) Fizmatgiz, 1963, p. 593. [English Transl., Pergamon Press, 2nd ed., 1965].
[12]S. H. Groves, R. N. Brown, and C. R. Pidgeon, Phys. Rev. **161,** 779 (1967).
[13]L. Liu and David Brust, Phys. Rev. Lett. **20,** 651 (1968).
[14]A. A. Abrikosov and S. D. Beneslavskii, Zh. Eksp. Teor. Fiz. **59,** 1280 (1970) [Sov. Phys. JETP **32,** 699 (1971)].
[15]V. I. Ivanov-Omskiĭ, B. T. Kolomiets, V. K. Ogorodnikov, and K. P. Smekalova, Fiz. Tekh. Poluprov. **1,** 289 (1967) [Sov. Phys. Semicond. **1,** 232 (1967)]; Fiz. Tekh. Poluprov. **4,** 264 (1970) [Sov. Phys. Semicond. **4,** 214 (1970)].

Translated by H. H. Nickle
86