Energy spectrum of narrow-gap semiconductor heterostructures described by a two-band equation

A. V. Kolesnikov[1)]

L. D. Landau Institute of Theoretical Physics, Russian Academy of Science, 117940 Moscow, Russia

A. P. Silin

P. N. Lebedev Physicotechnical Institute, Russian Academy of Sciences, 117924 Moscow, Russia

(Submitted 21 March 1995)
Pis'ma Zh. Éksp. Teor. Fiz. **61**, No. 9, 733–737 (10 May 1995)

Two effects which arise in narrow-gap heterostructures described by a two-band equation are discussed. The first effect ensures that there are no coincident quantum-well energies in the hole and electron wells in type-II heterostructures which are symmetric under the substitution $E_c(z) = -E_v(-z)$, where $E_c(z)$ $[E_v(z)]$ is the energy position of the bottom (top) of the conduction (valence) band. The second effect is the existence of a localized state whose energy coincides with the inversion point in a semiconductor structure with mutually inverted bands. This state does not depend on the shape of the energy structure or the boundary conditions. © *1995 American Institute of Physics.*

1. The single-band approximation of the effective-mass method is not valid[2] for calculations on type-II heterostructures in which the top of the valence band of one semiconductor (e.g., GaSb) lies higher on the energy scale than the bottom of the conduction band of the other semiconductor (e.g., InAs).[1] The reason is that the depth of the potential well for electrons and holes is greater than the width of the energy gap of the semiconductors making up the lattice. It becomes necessary to use a Hamiltonian which incorporates the interaction of the valence and conduction bands.[3] In this letter we use the two-band Hamiltonian in the form given in Ref. 4:

$$\begin{bmatrix} \varepsilon_g(z)/2+\varphi(z) & Pk_\perp - P\partial_z \\ Pk_\perp + P\partial_z & -\varepsilon_g(z)/2+\varphi(z) \end{bmatrix} \begin{Bmatrix} \psi \\ \chi \end{Bmatrix} = \varepsilon \begin{Bmatrix} \psi \\ \chi \end{Bmatrix}. \tag{1}$$

Here $\varepsilon_g(z)$ is the energy gap of the semiconductor, $\varphi(z)$ determines the position of the gap, $\hbar = 1$, $k_\perp = \sqrt{k_x^2 + k_y^2}$ is the momentum of free motion in the plane of the layers of the heterostructure, the matrix element P is the effective velocity, and ψ and χ are envelope wave functions which correspond to the two interacting bands.

Introducing $E_v(z) = \varphi(z) - \varepsilon_g(z)/2$ and $E_c(z) = \varphi(z) + \varepsilon_g(z)/2$, we easily find equations of the Schrödinger type:

$$\begin{cases} \psi''-\psi'\dfrac{P^2E_v'}{E_v-\varepsilon}+\psi\left[(E_v-\varepsilon)(E_c-\varepsilon)-\dfrac{k_\perp PE_v'}{E_v-\varepsilon}=-P^2k_\perp^2\right]=0, \\ \chi''-\chi'\dfrac{P^2E_c'}{E_c-\varepsilon}+\chi\left[(E_v-\varepsilon)(E_c-\varepsilon)+\dfrac{k_\perp PE_c'}{E_c-\varepsilon}-P^2k_\perp^2\right]=0. \end{cases} \tag{2}$$

2. A semiconductor–semimetal transition occurs in type-II heterostructures as the thicknesses of the layers are varied. This well-known fact is supported by numerous experimental and theoretical studies.[1,5–8] However, numerical calculations[9,10] indicate that, as the thickness of the layers is increased, the lower conduction subband descends below the upper valence subband, without a crossing.

In Ref. 11 we attempted to analytically clarify the situation for a model heterostructure in which the potential can be described by ($u_1>u_2$)

$$\begin{cases} E_c=u_1 \\ \qquad\qquad 0<z,\quad d_1<z<d_1+d_2,\ 2d_1+d_2<z\ \ (\mathrm{GaAs}), \\ E_v=-u_1, \end{cases}$$

$$\begin{cases} E_c=-u_2 \\ \qquad\qquad 0<z<d_1\ \ (\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{As}), \\ E_v=-u_1 \end{cases} \tag{3}$$

$$\begin{cases} E_c=u_1 \\ \qquad\qquad d_1+d_2<z<2d_1+d_2\ \ (\mathrm{GaSb}_y\mathrm{As}_{1-y}). \\ E_v=u_2, \end{cases}$$

The spectrum of the system is found from the equation

$$e^{2k_1d_2}\tan(k_2d_1)\tan(k_3d_1)(1+A^2)(1+B^2)$$
$$+\{(A^2-1)\tan(k_2d_1)-2A\}\{(1-B^2)\tan(k_3d_1)-2B\}=0, \tag{4}$$

where

$$k_1\frac{1}{P}\sqrt{(u_1-\varepsilon)(u_1+\varepsilon)+P^2k_\perp^2},\quad k_2=\frac{1}{P}\sqrt{(u_1+\varepsilon)(u_2+\varepsilon)-P^2k_\perp^2},$$

$$k_3=\frac{1}{P}\sqrt{(u_1-\varepsilon)(u_2-\varepsilon)-P^2k_\perp^2},\quad A^2=\frac{\varepsilon+u_2}{u_1-\varepsilon},\quad B^2=\frac{u_1+\varepsilon}{u_2-\varepsilon}.$$

Below we set $k_\perp=0$, as we can do without altering the physical picture. It is easy to see that Eq. (4) does not change under the substitution $\varepsilon\to-\varepsilon$ (this situation is natural because of the symmetry of the system) and that $\varepsilon=0$ is not a root, no matter what the parameter values are. Figure 1 shows the energy plotted as a function of the well width d_1 for the parameter values $u_1=3/P^2$, $u_2=1/P^2$, and $d_2=1/P$. For $\varepsilon>0$ we see that there are two types of levels: those which "descend" and those which "rise" with increasing d_1. The descending levels are evidently levels which arise in the course of size

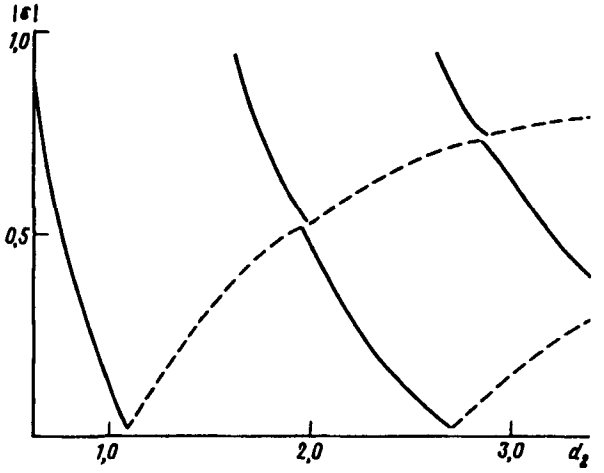

FIG. 1.

quantization of the electrons, while the rising levels are related to the holes in the same way. These levels do not cross; they repel each other. If we trace the wave functions, we see that the wave functions corresponding to the descending levels are localized in the electron well, while those which rise are localized in the hole well. In other words, a semiconductor–semimetal transition does indeed occur, in agreement with Refs. 1 and 8, but it occurs in the way pointed out in Refs. 9 and 10.

As the distance between levels, d_2, tends toward zero, and if the resulting structure repeats periodically (i.e., if a superlattice of the InAs–GaSb type forms[1]), the minimum value of $|\varepsilon|$ periodically vanishes as the layer thickness is varied, in total agreement with Ref. 4.

Below we generalize the results found for a heterostructure of type (3) to a wider class of heterostructures with $E_v(z) = -E_c(-z)$. In other words, the well for holes has the same shape as that for electrons. The absence of a state with $\varepsilon = 0$ corresponds to a splitting of the energy in the two identical wells.

We first note that the substitutions $\varepsilon \to -\varepsilon$, $E_c(-z) \to E_v(z)$, $z \to -z$, $\chi(z) \to i\psi(-z)$, and $\psi(z) \to i\chi(-z)$ do not change system (1). Accordingly, if we know solutions for Eq. (2) for the region $z > 0$,

$$\begin{cases} \Psi(z,\varepsilon) = A\psi_1(z,\varepsilon) + iB\psi_2(z,\varepsilon), \\ \Phi(z,\varepsilon) = C\chi_1(z,\varepsilon) + iD\chi_2(z,\varepsilon), \end{cases} \tag{5}$$

where ψ_1 and ψ_2 are independent solutions of Eqs. (2), and the functions χ_1 (χ_2) are found from (1) with $\psi = \psi_1$ (ψ_2), then we immediately write solutions for the region $z < 0$:

$$\begin{cases} \Psi(z,\varepsilon) = iC\chi_1(-z,-\varepsilon) - D\chi_2(-z,-\varepsilon), \\ \Phi(z,\varepsilon) = -iA\psi_1(-z,-\varepsilon) + B\psi_2(-z,-\varepsilon). \end{cases} \tag{6}$$

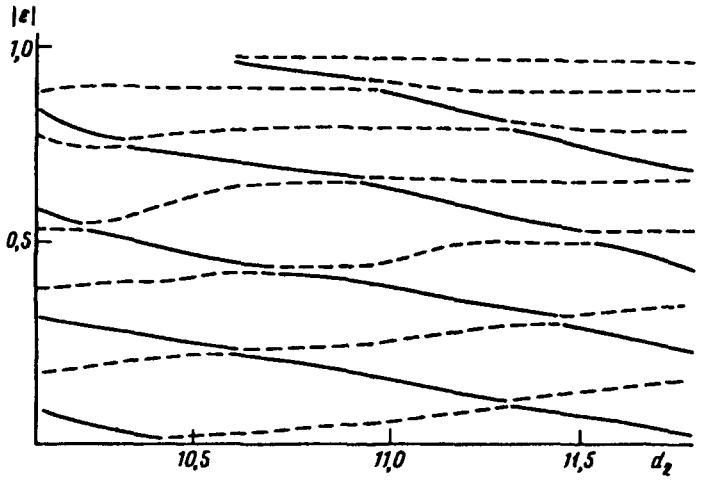

FIG. 2.

Since we are interested in only the presence or absence of a state with $\varepsilon=0$, we set $\varepsilon=0$ directly in (5) and (6). This operation is not totally rigorous mathematically (we should have constructed an equation for finding the spectrum and used this equation to determine whether a state with $\varepsilon=0$ exists for any parameter values). Nevertheless, it can be justified by the continuity of the dependence of the ε quantization equation found from (5) and (6), provided that the point $\varepsilon=0$ is not a singular point of differential equations (2). We postpone a discussion of this case to Sec. 3.

If we now require that Ψ and Φ vanish at $\pm\infty$ as boundary conditions on the solutions, in accordance with the proposition that the particles are localized, then functions which decay at $\pm\infty$ are solutions ψ_1 and χ_1 (for definiteness, we need to set the constant B and D equal to zero in the case of increasing functions). Equation (2) is real, however, and we cannot join solutions which are real at $z>0$ and imaginary at $z<0$ at the point $z=0$, provided that ψ_1 and χ_1 do not vanish simultaneously at this point.

We are thus left with only the problem of determining a form of $E_v(z)$ and $E_c(z)$ which does not cause ψ_1 and χ_1 to vanish at $z=0$. Substituting $\psi=\sigma\sqrt{E_v}$ and $\phi=\rho\sqrt{E_c}$ into (2), we find the following equations for σ and ρ:

$$\begin{cases} \sigma''+\sigma(E_vE_c-P^2(k_\perp+E_v'/2E_v)^2-P^2E_v''/2E_v-P^2E_v'^2/E_v^2)=0, \\ \rho''+\rho(E_vE_c-P^2(k_\perp-E_c'/2E_c)^2-P^2E_c''/2E_c-P^2E_c'^2/E_c^2)=0. \end{cases} \tag{7}$$

Since near the point $z=0$, we now have $E_cE_v<0$. For this general situation we have wave functions of the same form as in case (3), discussed above, if we assume $E_c\sim z^\alpha+\text{const}$ in the region $z\sim 0$, $0\leqslant\alpha\leqslant 1$.

It is because of the conditions $\Psi(\pm\infty)\to 0$ and $\Phi(\pm\infty)\to 0$ that we can choose $\psi_{1,2}$ and $\chi_{1,2}$ to be real. For the superlattice discussed in Ref. 4, solutions are Bloch functions which do not decay at $\pm\infty$, and it is not possible to choose $\psi_{1,2}$ and $\chi_{1,2}$ to be real.

3. We now consider a spatially nonuniform contact between two semiconductors with mutually inverted bands. A structure of this type, of infinite size, was discussed in Ref. 12. It was shown for such systems in Refs. 12 and 13 that in the case $E_v(z) = -\alpha^2 E_c(z)$ there exists a state, localized near the inversion point, with an energy which coincides with inversion point. This result was found by transforming the two-band equations into supersymmetric equations. The method used in Ref. 13 does not work for arbitrary E_c and E_v forming an inverted contact, however, since it is not a natural candidate for the role of superpotential $W^2 = E_c E_v$. Furthermore, one can show that the case $E_v(z) = -\alpha^2 E_c(z)$ is the only one in which two-band equations can be transformed into supersymmetric equations.

However, if the energy gap of a semiconductor (whether it is finite or infinite is unimportant; the only point of importance is that there be one inversion point) is a continuous function of the coordinate z, which changes sign at the inversion point, the result of Refs. 12 and 13 remains valid. We can show that the existence of a state with $\varepsilon = 0$ is a consequence of only the very fact that the bands intersect. (The origin of coordinate coincides with the inversion point.)

If we assume that the energy bands have the behavior $E_c \sim z^\alpha$ and $E_v \sim z^\beta$ near the origin ($z \sim 0$), vanishing at the point $z = 0$, with exponents α and β which may generally be different for values of z greater than and less than zero, then the solutions of Eq. (2) near the origin can be written

$$\begin{cases} \Psi \sim z^{1+\alpha} A_1 + A_2, \\ \Phi \sim z^{1+\beta} A_3 + A_4, \end{cases} \tag{8}$$

where A_i are constants. General solutions of Eqs. (2) for the regions $z > 0$ and $z < 0$ are

$$\begin{cases} \Psi = C y_1(z) + C_1 y_2(z), \\ \Phi = D g_1(z) + D_1 g_2(z); \end{cases} \tag{9}$$

$$\begin{cases} \Psi = C y_1'(z) + C_2 y_2'(z), \\ \Phi = D g_1'(z) + D_2 g_2'(z). \end{cases} \tag{10}$$

Here C, D, C_i, and D_i are constants; the functions $y_2(g_2)$ and $y_2'(g_2')$ have the behavior $z^{1+\alpha}$ and $z^{1+\beta}$ at $z \sim 0$; and the functions $y_1(g_1)$ and $y'^1(g'^1)$ are asymptotically independent of z at $z \sim 0$ and are equal. Again, $y_1^{(\prime)}$ ($y_2^{(\prime)}$) is a particular solution of Eq. (2), and $g_1^{(\prime)}$ ($g_2^{(\prime)}$) is found from the function y_1 (y_2), replacing ψ in (1).

Through the choice of solutions and constants described above, we have achieved continuous solutions at the inversion point ($z = 0$), and we still have enough constants left to satisfy arbitrary boundary conditions.

Because of the presence of the singular point in (2), which arises as a result of the band intersection, we were able to choose different constants C_i and D_i in solutions (9) and (10). We might also point out that there should be an odd number of inversion points in a sample of infinite size [i.e., an odd number of singular points in Eqs. (2), at which the coordinate z takes on values on the entire number line]. Otherwise, we would not be able

to achieve a decay of solutions (2) at $\pm\infty$ simultaneously. A situation of this sort with two inversion points was discussed in Refs. 14 and 15, where there was of course no state with $\varepsilon = 0$.

We are indebted to M. Altarelli and B. A. Volkov for a discussion of these results. This study was supported in part by a grant from the International Science Foundation (N9Z000), a grant from the Russian Fund for Fundamental Research (Grant 93-02-14700), and an L. D. Landau Stipend from the KFA Korschungszentrum Jülig GmbH.

[1)]E-mail: alkol@itp.ac.ru

---

[1]L. Esaki, Lect. Notes Phys. **133**, 302 (1980).
[2]A. P. Silin, Kratk. Soobshch. Fiz. **12**, 13 (1985).
[3]L. V. Keldysh, Zh. Éksp. Teor. Fiz. **45**, 364 (1963) [Sov. Phys. JETP **18**, 253 (1963)].
[4]M. de Dios Leyva, R. P. Alvarez, and J. L. Gondar, Phys. Status Solidi (b) **125**, 221 (1984).
[5]G. A. Sai-Halasz, R. Tsu, and L. Esaki, Appl. Phys. Lett. **30**, 651 (1977).
[6]G. A. Sai-Halasz, L. Esaki, and W. A. Harrison, Phys. Rev. B **18**, 2812 (1978).
[7]L. L. Chang, N. Kawai, G. A. Sai-Halasz *et al.*, Appl. Phys. Lett. **35**, 939 (1979).
[8]G. Bastard, E. E. Mendez, L. L. Chang, and L. Esaki, J. Sci. Technol. **21**, 531 (1982).
[9]M. Altarelli, Lect. Notes Phys. **177**, 174 (1983).
[10]M. Altarelli, Phys. Rev. B **28**, 842 (1983).
[11]A. V. Kolesnikov and A. P. Silin, Preprint No. 21, P. N. Lebedev Physics Institute, Moscow, 1993.
[12]O. A. Pankratov and B. A. Volkov, JETP Lett. **42**, 178 (1985).
[13]O. A. Pankratov, S. V. Pakhomov, and B. A. Volkov, Solid State Commun. **61**, 93 (1987).
[14]B. G. Idlis and M. Sh. Usmanov, Fiz. Tekh. Poluprovodn. **26**, 329 (1992) [Sov. Phys. Semicond. **26**, 186 (1992)].
[15]V. Korenann and H. D. Drew, Phys. Rev. B **25**, 6446 (1986).

Translated by D. Parsons