Size quantization of the holes in a semiconductor with a complicated valence band and of the carriers in a gapless semiconductor

M. I. D'yakonov and A. V. Khaetskiĭ

A. F. Ioffe Physicotechnical Institute, Academy of Sciences of the USSR

(Submitted 2 November 1981)
Zh. Eksp. Teor. Fiz. **82**, 1584–1590 (May 1982)

Size quantization of the holes in a thin film of a semiconductor with a structurally complicated valence band, i.e., a semiconductor of the Ge type, and of the carriers in a gapless semiconductor of the HgTe type is considered. The energy spectrum and the effective masses in the size-quantization subbands are calculated. The characteristics, connected with the presence of surface states, of the size quantization in a gapless semiconductor are discussed.

PACS numbers: 73.20.Cw, 73.60.Fw

1. In a thin semiconductor film the component of the electron or hole quasimomentum along the normal to the film surface is quantized, as a result of which the energy spectrum is a set of two-dimensional subbands. This spectrum is very simple and well known in the case of the simple-conduction-band electrons (when there is only a twofold spin degeneracy). Recently, the phenomenon of size quantization of the holes in semiconductors with a structurally complicated valence band (Ge and GaAs) was also studied experimentally.[1-3] In these semiconductors, as is well known, the valence band consists of light- and heavy-hole subbands with a common vertex at the center of the Brillouin zone. Size quantization of the holes in such semiconductors has not been theoretically investigated.

The present paper is devoted to the computation of the energy spectrum of the holes in semiconductors with the diamond structure in the presence of size quantization. At the same time we consider the size quantization of the carriers in a gapless semiconductor of the HgTe type. Mathematically, these problems are equivalent, since the Bloch wave functions have the same symmetry (the Γ_8 representation), and the volume spectra in the two cases are described by the Luttinger Hamiltonian.

The object of the investigation is the dependence $E(k)$ of the energy on the momentum k parallel to the film surface in the size-quantization subbands. As to the locations of the size-quantization levels for $k=0$, they are given by the standard formula

$$E_n^{(j)} = \hbar^2\pi^2 n^2/2m_j a^2,$$

where a is the film thickness, m_j is the volume effective mass of a particle of the j-th kind, and $n=1,2,3,\ldots$. This is due to the fact that the light and heavy holes (or the electrons and holes in a gapless semiconductor) are quantized independently at $k=0$. Indeed, for a given momentum, there are four states differing in their spin components along the direction of the momentum, so that the components $\pm\frac{3}{2}$ correspond to the heavy, while the $\pm\frac{1}{2}$ components correspond to the light, holes. The spin component along the normal to the film surface is conserved in the purely transverse (i.e., $k=0$) motion; therefore, the reflection of the particle from the film surface is accompanied by a change in sign of its spin component along the direction of the momentum. This change in sign does not lead to the transformation of one type of particles into the other.

If $k \neq 0$, then the reflection of the particle from the surface may be accompanied not only by a change in sign, but also by a change in magnitude of the spin component along the momentum. This implies that a light hole can be transformed into a heavy hole (and vice versa) in the course of its reflection from the surface. Thus, the wave function for $k \neq 0$ should be a linear combination of volume-state wave functions of light and heavy holes with one and the same energy. This mixing of the states is the cause of the complicated character of the function $E(k)$ and, in particular, the deviation of the effective masses in the size-quantization subbands from the volume effective masses.

In the case of a gapless semiconductor, the mixing of the electron and hole wave functions leads to the appearance of specific surface states, which are considered in our previous paper.[4] The corresponding branches of the energy spectrum appear upon the size quantization of the system. Furthermore, a forbidden band appears in the energy spectrum because of the breaking of the cubic symmetry in the thin film. Since, as asserted above, for $k=0$ the electrons and holes become quantized independently, it may be inferred[5] that the width of the forbidden band is equal to the distance between the $k=0$ electron and hole levels with $n=1$, i.e., to the quantity $\hbar^2\pi^2/2m^*a^2$, where m^* is the reduced electron and hole mass. But, as we shall show, this inference is incorrect. In fact, the forbidden-band width is significantly smaller, it being equal to $3\hbar^2\pi^2/2m_h a^2$, where m_h is the volume effective mass of the hole. This circumstance is closely tied with the presence of surface states.

2. The energy spectrum in the vicinity of the point of degeneracy of the bands is described by the Luttinger Hamiltonian[6]

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

where m_0 is the free-electron mass; $\gamma = (2\gamma_2 + 3\gamma_3)/5$; γ_1, γ_2, and γ_3 are the Luttinger parameters; p is the quasi-momentum operator; and J_x, J_y, and J_z are 4×4

0038-5646/82/050917-04$04.00

matrices corresponding to the spin $\frac{3}{2}$. The Hamiltonian (1) has been written in the spherical approximation, to which we limit ourselves here.

In an unbounded crystal, the solutions to the Schrödinger equation with the Hamiltonian (1) give, as is well known, two energy-spectrum branches corresponding to the free motion of particles with effective masses

$$m_1=m_0/(\gamma_1+2\gamma),\quad m_2=m_0/(\gamma_1-2\gamma).$$

These two branches correspond, depending on the values of the Luttinger parameters, to either light and heavy holes in a normal semiconductor of the Ge type (when $\gamma_1+2\gamma>0$ and $\gamma_1-2\gamma>0$) or electrons and holes in a gapless semiconductor of the HgTe type (when $\gamma_1+2\gamma>0$ and $\gamma_1-2\gamma<0$).

The Schrödinger equation with the Hamiltonian (1) is a system of four second-order equations for the components of the spinor ψ. The most general conditions at the film boundaries have the form

$$\psi+l\hat{S}\mathbf{n}\nabla\psi=0,$$

where n is the unit vector along the normal to the surface and $\hat{S}$ is a dimensionless matrix whose general form can be established from symmetry considerations. The characteristic length l is determined by the dimension of the region near the crystal boundary where the effective-mass approximation is inapplicable. For narrow-band semiconductors $l\sim(\hbar^2/mE_g)^{1/2}$ (Ref. 7), where E_g in our case is the distance to the nearest s band. In the region where the spectrum is parabolic (i.e., where $E\ll E_g$), the characteristic wave vector q is such that $ql\ll 1$. In this case we can neglect those terms in the boundary conditions which contain $\nabla\psi$, thereby reducing these conditions to the normal requirement that $\psi=0$ at the boundaries.

Let us orient the z axis along the direction perpendicular to the film surface. Let the film boundaries be located at $z=\pm a/2$. The energy spectrum is determined from the Schrodinger equation $H\psi=E\psi$ with the boundary conditions $\psi(-a/2)=\psi(a/2)=0$. Let the wave vector k in the film-surface plane be oriented along the x axis. Let us denote the wave-vector component along the z axis by q. For a given values of k and q the Schrödinger equation possesses four independent solutions of the form

$$\varphi_\alpha^{(j)}(k,q)\exp(ikx+iqz),\tag{2}$$

where the superscript $j=1,2$ numbers of the two branches of the energy spectrum ($E_j=\hbar^2(k^2+q^2)/2m_j$), while the subscript α numbers the two degenerate states corresponding to a given j. In the representation in which the matrix J_z is diagonal, the spinors $\varphi_\alpha^{(j)}(k,q)$ can be chosen in the form

$$\varphi_1^{(1)}(k,q)=\begin{pmatrix} i\lambda \\ 1 \\ -i \\ -\lambda \end{pmatrix},\quad \varphi_1^{(2)}(k,q)=\begin{pmatrix} 1 \\ -i\lambda \\ \lambda \\ -i \end{pmatrix},\quad \lambda(k,q)=\frac{k\sqrt{3}}{k-2iq}.\tag{3}$$

The spinors $\varphi_2^{(1)}$ and $\varphi_2^{(2)}$ are obtained from (3) through complex conjugation: $\varphi_2^{(j)}=(\varphi_1^{(j)})^*$.

The general solution to the Schrödinger equation with given values of the wave vector k and the energy E is a superposition of solutions of the type (2), where the values of q for $j=1,2$ should be determined from the equation

$$E=\hbar^2(k^2+q_1^2)/2m_1=\hbar^2(k^2+q_2^2)/2m_2.\tag{4}$$

Notice that the quantities q_1 and q_2 given by Eq. (4) may turn out to be purely imaginary. Each of the quantities q_1 and q_2 can assume two values differing in sign. Thus, there are in all eight independent solutions for given values of k and E. The general solution can be written in the form

$$\psi=e^{ikx}\sum_{j,\alpha}[A_\alpha^{(j)}\varphi_\alpha^{(j)}(k,q_j)\exp(iq_jz)+B_\alpha^{(j)}\varphi_\alpha^{(j)}(k,-q_j)\exp(-iq_jz)].\tag{5}$$

The use of the explicit form (3) of the spinors $\varphi_\alpha^{(j)}$ in the boundary conditions $\psi(-a/2)=\psi(a/2)=0$ gives rise to a system of eight homogeneous equations for the coefficients $A_\alpha^{(j)}$ and $B_\alpha^{(j)}$. This system reduces to two independent systems of four equations each for the coefficients $A_1^{(1)}$, $A_2^{(2)}$, $B_1^{(1)}$, and $B_2^{(2)}$ and the coefficients $A_2^{(1)}$, $A_1^{(2)}$, $B_2^{(1)}$, and $B_1^{(2)}$. The determinants of the two systems are identical (which corresponds to a twofold Kramers degeneracy of the levels), and have the form

$$\begin{vmatrix} \lambda_1^*\cos(q_1a/2) & \lambda_1\cos(q_1a/2) & \cos(q_2a/2) & \cos(q_2a/2) \\ \cos(q_1a/2) & \cos(q_1a/2) & -\lambda_2\cos(q_2a/2) & -\lambda_2^*\cos(q_2a/2) \\ \lambda_1^*\sin(q_1a/2) & -\lambda_1\sin(q_1a/2) & \sin(q_2a/2) & -\sin(q_2a/2) \\ \sin(q_1a/2) & -\sin(q_1a/2) & -\lambda_2\sin(q_2a/2) & \lambda_2^*\sin(q_2a/2) \end{vmatrix}=0,\tag{6}$$

where $\lambda_j=\lambda(k,q)$. Expanding the determinant (6), we obtain the equation

$$1-\cos q_1a\cos q_2a=-f(q_1,\ q_2,\ k)\sin q_1a\sin q_2a,\tag{7}$$

where

$$f(q_1,\ q_2,\ k)=[4k^4+k^2(q_1^2+q_2^2)+4q_1^2q_2^2]/6q_1q_2k^2.\tag{8}$$

Equation (7), together with Eq. (4), determines the sought energy spectrum. For $m_1, m_2>0$, it describes the size quantization of the holes in a complicated valence band; for $m_1>0$, $m_2<0$, the size quantization of the carriers in a gapless semiconductor.

For $k=0$ we have $f(q_1,q_2,0)=\infty$; then it follows from Eq. (7) that $q_1a=\pi n$ or $q_2a=\pi n$. Thus, as has already been noted in Sec. 1, for $k=0$ we have two independent systems of levels with energies

$$E_n^{(j)}(0)=\hbar^2\pi^2n^2/2m_ja^2.$$

This is valid both for the holes in a complicated valence band and for the carriers in a gapless semiconductor. Below we shall consider these two cases separately. The energy-spectrum branches $E_n^{(j)}(k)$ will be numbered by the indices j and n indicating which size-quantization level the branch in question goes over into at $k=0$.

3. *The spectrum of the holes in a complicated valence band*. In this case $m_1=m_l$ and $m_2=m_h$, where m_l and m_h are the effective masses of the light and heavy holes respectively. Equation (7) can be rewritten in the form

$$\operatorname{tg}(q_1a/2)\operatorname{ctg}(q_2a/2)=-f\pm(f^2-1)^{1/2}.\tag{9}$$

The two signs in Eq. (9) correspond to the two different

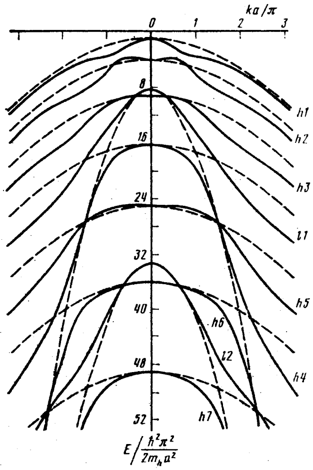

FIG. 1. Size-quantization energy spectrum of the holes in the valence band of germanium. The dashed curves depict the parabolas (10) corresponding to the independent quantization of the light and heavy holes; $\beta = 0.12$.

sets of $E(k)$ curves. It is easy to see that, on choosing the upper sign from Eq. (9), we obtain the branches $E_{n_1}^{(l)}(k)$ and $E_{n_2}^{(h)}(k)$ with even n_1 and odd n_2; on choosing the lower sign, the branches with odd n_1 and even n_2. The indicated sets of curves can intersect, since they are obtained from different equations. Specifically, the branches with the same j intersect when their n numbers are of different parities, while the branches with different j intersect when the n's are of the same parity. The points of intersection are determined by the conditions $q_1 a = \pi l_1$ and $q_2 a = \pi l_2$, where l_1 and l_2 are nonzero whole numbers of the same parity. In this case Eq. (7) is satisfied irrespective of the value of k, and Eq. (4) gives the E and k values corresponding to the points of intersection.

Figure 1 shows the energy spectrum, which was obtained through a numerical solution of Eqs. (4) and (7) with $\beta = m_l/m_h = 0.12$ (Ge). Schematically, this spectrum can be described as follows. Let us draw the set of parabolas

$$E=\hbar^2(k^2+\pi^2 l_1^2/a^2)/2m_l, \qquad E=\hbar^2(k^2+\pi^2 l_2^2/a^2)/2m_h, \tag{10}$$

which correspond to the independent quantization of the light and heavy holes for all k (the dashed curves in Fig. 1). The parabolas corresponding to the light and heavy holes intersect. The qualitative form of the true spectrum can be obtained by replacing the intersections of the parabolas (10) with l_1 and l_2 values of different parities by anti-intersections. The intersections corresponding to l_1 and l_2 values of the same parity remain in the true spectrum.

This scheme provides a rough qualitative description of the size-quantization spectrum. In fact, the effective masses in the size-quantization subbands differ significantly from the volume masses of the light and heavy holes. It follows from Eqs. (4) and (7) that, for $ka \ll (\pi n)^{1/2}$,

$$E_n^{(j)}(k) = E_n^{(j)}(0) + \hbar^2 k^2/2m_n^{(j)},$$

where the effective masses $m_n^{(l)}$ are given by the expression

$$\frac{1}{m_n^{(l)}} = \frac{1}{m_l}\left[1 - \frac{3\beta^{1/2}}{\pi n}\,\frac{(-1)^n - \cos(\pi n \beta^{-1/2})}{\sin(\pi n \beta^{-1/2})}\right]. \tag{11}$$

The expression for $m_n^{(h)}$ is obtained from (11) by substituting m_h for m_l and β^{-1} for β. Several first values of the ratio $m_n^{(j)}/m_j$ for Ge are given in Table I.

TABLE I. Values, computed from the formulas (11) and (15), of the effective masses in the size-quantization subbands for Ge and HgTe.

n	Ge		HgTe	
	$m_n^{(l)}/m_l$	$m_n^{(h)}/m_h$	$m_n^{(e)}/m_e$	$m_n^{(h)}/m_h$
1	0.94	0.18	1.31	−0.11
2	0.94	−0.61	1.14	−3.96
3	0.93	1.07	1.09	−1.85
4	0.92	0.51	1.06	8.05
5	0.91	−3.78	1.05	4.87
6	0.88	1.07	1.04	2.67
7	0.81	0.68	1.03	2.22

Notice that, according to Eq. (4), two regions can be distinguished in the (E, k) plane: the region $E > \hbar^2 k^2/2m_l$, where the wave vectors q_1 and q_2 are both real, and the region $E < \hbar^2 k^2/m_l$, where q_2 is real and q_1 is purely imaginary. In the second of these regions the transformation of a heavy hole into a light one in the course of a reflection from the surface is impossible, in view of which the level of admixture of the states of the light hole in the wave function (5) attenuates exponentially with distance from the surface into the interior of the film.

Miller *et al.*[3] report an experimental study of the optical transitions between the size-quantization subbands for the conduction and valence bands of GaAs in heterostructures consisting of thin GaAs layers separated by relatively thick $Al_xGa_{1-x}As$ layers. Besides transitions between subbands with the same n, "forbidden" transitions between subbands with different n were observed. Such transitions would be forbidden if the light and heavy holes were quantized independently. The authors of Ref. 3 relate the occurrence of these forbidden transitions with the distortion of the bottom of the potential well formed by the GaAs layer. As follows from the results of the present paper, such forbidden transitions are possible even without allowance for this effect. Indeed, the dependence of the wave function (5) on the coordinate z does not have the simple form $\cos(\pi n z/a)$ at all nonzero values of k, and, consequently, the function (5) is no longer orthogonal to the wave function of the n'-th ($n' \neq n$) subband of the conduction band. The transitions with $n' \neq n$ are forbidden only when $k = 0$.

4. *The spectrum of the carriers in a gapless semiconductor.* The only difference between this and the preceding case lies in the fact that now $m_1 > 0$ and $m_2 < 0$ in Eq. (4). Let us set $m_1 = m_e$ and $m_2 = -m_h$, where m_e and m_h are the electron and hole masses respectively. Three essentially different regions can now be distinguished in the (E, k) plane (see Fig. 2): I) the region $E > \hbar^2 k^2/2m_e$, where q_1 is real and q_2 is purely imaginary II) the region $E < -\hbar^2 k^2/2m_h$, where q_1 is purely imaginary and q_2 is real; III) the region $-\hbar^2 k^2/2m_h < E < \hbar^2 k^2/$

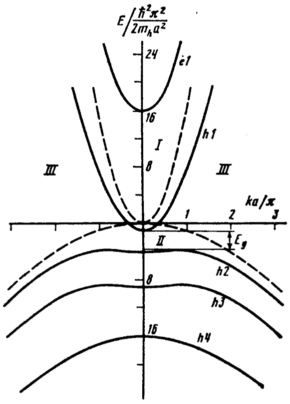

FIG. 2. Size-quantization energy spectrum of the carriers in the gapless semiconductor HgTe. The dashed curves indicate the boundaries of the regions I, II, and III; for $ka \gg 1$ the curve $h1$ gives the energy spectrum of the surface states; E_g is the forbidden-band width; $\beta = 0.063$.

$2m_e$, where q_1 and q_2 are both purely imaginary.

Let us set $q_1 = i\varkappa_1$ and $q_2 = i\varkappa_2$ in the region III, and consider the limiting case $\varkappa_1 a \gg 1, \varkappa_2 a \gg 1$. Then Eq. (7) assumes the form

$$4k^4 - k^2(\varkappa_1^2 + \varkappa_2^2) + 4\varkappa_1^2\varkappa_2^2 - 6\varkappa_1\varkappa_2 k^2 = 0, \tag{12}$$

and, according to Eq. (4), $k^2 - \varkappa_1^2 = \beta(\varkappa_2^2 - k^2)$, where $\beta = m_e/m_h$. Equation (12) coincides with the equation that describes the surface states in a gapless semiconductor.[4] The dependence on the mangitude of the mass ratio β contains either one or two energy-spectrum branches corresponding to these states. For $\beta < \frac{1}{3}$, there exists one branch of the electronic type with effective mass m_{s1} given by the formula[4]

$$m_e/m_{s1} = 1 - {}^1/_4[1 + (3\beta)^{1/2}]^2. \tag{13}$$

It follows from Eq. (12) that $\varkappa_1 \sim \varkappa_2 \sim k$; therefore, the above-considered limiting case corresponds to the condition $ka \gg 1$. Thus, in the $\beta < \frac{1}{3}$ case one of the branches of the size-quantization spectrum should go over at high k values into the branch of surface states with effective mass given by (13). All the remaining branches of the spectrum are contained in the regions I and II.

Figure 2 shows the size-quantization energy spectrum in a gapless semiconductor, obtained through a numerical solution of Eqs. (4) and (7) with $\beta = 0.063$ (HgTe). The most interesting characteristic of this spectrum is the fact that the branch corresponding at $k = 0$ to the hole-quantization level with $n = 1$ has an electronic, and not a hole, character. For $ka \gg 1$ this branch represents the energy spectrum of the surface states. The wave function changes its character when this branch crosses the boundary between the regions II and III: the electron density has its maximum in the middle of the film (i.e., at $z = 0$) when $ka \ll 1$ and in the vicinity of the film surfaces when $ka \gg 1$. As can be seen from Fig. 2 the forbidden-band width is determined by the distance between the $n = 1$ and $n = 2$ size-quantization levels of the hole:

$$E_g \approx 3\hbar^2\pi^2/2m_h a^2.$$

For $\frac{1}{3} < \beta < 3$ there exist two branches of surface states—of the electronic and hole types.[4] In this case, as can be shown, both of these branches exist: in the size-quantization spectrum they go over to the $n = 1$ size-quantization levels for the electron and hole, respectively, when $ka \gg 1$, but $k \to 0$. The forbidden-band width is then determined by the distance between these levels. But this case is of no practical interest, since in the known gapless semiconductors $\beta < \frac{1}{3}$.

From Eqs. (4) and (7) we find that, for $ka \ll (\pi n)^{1/2}$, the expressions for the energy in the size-quantization subbands have the form

$$E_n^{(e)}(k) = \frac{\hbar^2\pi^2 n^2}{2m_e a^2} + \frac{\hbar^2 k^2}{2m_n^{(e)}}, \quad E_n^{(h)}(k) = -\frac{\hbar^2\pi^2 n^2}{2m_h a^2} - \frac{\hbar^2 k^2}{2m_n^{(h)}}, \tag{14}$$

where the effective masses $m_n^{(e)}$ are given by the expression

$$\frac{1}{m_n^{(e)}} = \frac{1}{m_c}\left(1 + \frac{3\beta^{1/2}}{\pi n}\frac{(-1)^n - \mathrm{ch}(\pi n\beta^{-1/2})}{\mathrm{sh}(\pi n\beta^{-1/2})}\right). \tag{15}$$

The expression for $m_n^{(h)}$ is obtained from (15) by substituting m_h for m_e and β^{-1} for β. Values of the ratios $m_n^{(e)}/m_e$ and $m_n^{(h)}/m_h$ for HgTe are given in Table I. It can be seen that the effective masses $m_n^{(e)}$ differ slightly from the volume mass m_e of the electron, whereas the masses $m_n^{(h)}$ significantly differ from the hole mass m_h and, for the first values of n even have the opposite sign.

[1] R. M. Kolbas *et al.*, Solid State Commun. **31**, 1033 (1979).

[2] N. N. Ovsyuk and M. P. Sinyukov, Pis'ma Zh. Eksp. Teor. Fiz. **32**, 366 (1980) [JETP Lett. **32**, 342 (1980)].

[3] P. C. Miller, D. A. Kleinman, W. A. Nordland, and A. C. Gossard, Phys. Rev. B **22**, 863 (1980).

[4] M. I. D'yakonov and A. V. Khaetskiĭ, Pis'ma Zh. Eksp. Teor. Fiz. **33**, 115 (1981) [JETP Lett. **33**, 110 (1981)].

[5] J. G. Broerman, Phys. Rev. Lett. **45**, 747 (1980).

[6] J. M. Luttinger, Phys. Rev. **102**, 1030 (1956).

[7] V. A. Volkov and T. N. Pinster, Surf. Sci. **81**, 181 (1979).

Translated by A. K. Agyei