

NOTES 12: The Kane Model and k dot p Perturbation Theory

As we have seen, band-gaps are very common in the band structure of crystalline materials, independent of their electrical or optical properties. In normal-gap and narrow-gap semiconductors having a direct band gap (i.e., the minimax points occur at the same point in k space), there is a very useful model of bandstructure first developed by E. Kane in the late 1950s (the “golden years” of solid-state science). The Kane model is particularly useful for the studies of narrow band semiconductor where the conduction and valence bands are close to each other. The mutual interaction of the conduction band and valence bands strongly distorts the band edges near $\vec{k} = 0$ so that the band nature is non-parabolic¹. For the description of the electronic and optical properties of a narrow-band semiconductor, a multiple-bands model is a must. In the case of carrier transports in a heterostructure where the conduction and valence bands are close to each other in dissimilar materials, the Kane model may be applied too.

The k-p method

The Schrödinger equation for a perfect, uniform crystal is

$$\frac{p^2}{2m_0} \psi_{n\vec{k}} + V(\vec{r}) \psi_{n\vec{k}} = U_n(\vec{k}) \psi_{n\vec{k}}, \quad (1)$$

where $V(\vec{r})$ is the periodic potential. m_0 is the free electron mass. When the Bloch theorem is applied, $\psi_{n\vec{k}}$ is in the explicit form as,

$$\psi_{n\vec{k}} = \exp(i\vec{k} \cdot \vec{r}) u_{n\vec{k}}(\vec{r}), \quad (2)$$

where $u_{n\vec{k}}(\vec{r})$ has the periodicity of crystal lattice. The equation for $u_{n\vec{k}}(\vec{r})$ is,

$$\left\{ \frac{p^2}{2m_0} + V(\vec{r}) + \frac{\hbar}{m_0} (\vec{k} \cdot \hat{p}) \right\} u_{n\vec{k}}(\vec{r}) = \left[U_n(\vec{k}) - \frac{\hbar^2 k^2}{2m_0} \right] u_{n\vec{k}}(\vec{r}) \quad (3)$$

where the index n is to label the energy band and $U_n(\vec{k})$ is the electron energy with momentum \vec{k} at the n^{th} band.

In an atom, there exists the interaction between the electron spin moment and the orbital magnetic momentum. Hence the energy of states in a semiconductor may be

modified by the spin-orbit coupling. The effect becomes more significant as the semiconductor element becomes heavier (i.e. the atomic number is larger). The Hamiltonian for the spin-orbit coupling is written as¹,

$$H_{so} = -\frac{i\hbar^2}{4m_0^2 c^2} \hat{\sigma} \cdot (\nabla V \times \hat{p}) \quad (4)$$

As the spin-orbit coupling is taken into account, Eq.(3) need be modified to ,

$$\left\{ \frac{p^2}{2m_0} + V(\vec{r}) + \frac{\hbar}{m_0} (\vec{k} \cdot \vec{p}) + \frac{\hbar}{4m_0^2 c^2} \hat{\sigma} \cdot (\nabla V \times \hat{p}) + \frac{\hbar^2}{4m_0^2 c^2} \hat{\sigma} \cdot (\nabla V \times \vec{k}) \right\} u_{n\vec{k}}(\vec{r}) = U'_n(\vec{k}) u_{n\vec{k}}(\vec{r}) \quad (5)$$

$$U'_n(\vec{k}) = U_n(\vec{k}) - \frac{\hbar^2 k^2}{2m_0}$$

The fifth term is very small and thus can be neglected. Here the crystal momentum $\hbar\vec{k}$ is far less than the atomic momentum \vec{p} . At the point $\vec{k} = 0$, the **k.p** equation becomes,

$$\left\{ \frac{p^2}{2m} + V(\vec{r}) \right\} u_{n0}(\vec{r}) = U'_n(0) u_{n0}(\vec{r}) \quad (6)$$

where $u_{n0}(\vec{r})$ is the complete and orthogonal set of basis functions. It is necessary to find the matrix representation for the Hamiltonian H in Eq.(3) or (5) as follows,

$$H u_{n\vec{k}}(\vec{r}) = (H_0 + H' + H_{so}) u_{n\vec{k}}(\vec{r}) = U'_n(\vec{k}) u_{n\vec{k}}(\vec{r}) \quad (7)$$

$$H_0 = \frac{p^2}{2m_0} + V(\vec{r}), \quad H' = \frac{\hbar}{m_0} (\vec{k} \cdot \vec{p}) \quad (8)$$

in a properly chosen set of basis functions $u_{n0}(\vec{r})$ in the Hilbert space. $u_{n\vec{k}}(\vec{r})$ is the linear superposition of

$$u_{n\vec{k}}(\vec{r}) = \sum_n a_{n'n}(\vec{k}) u_{n0}(\vec{r}) \quad (9)$$

The basis functions and the symmetry of semiconductor crystal

A perfect semiconductor crystal allows symmetry transformations such as rotation, reflection, inversion². Under these symmetry operations, the crystal lattice space repeats itself or remains invariant. The collections of symmetry operations consist of a group which may be defined by axioms of multiplication. When a proper set of basis functions

is chosen, every symmetry operation can be represented by a matrix. The matrices obey the same rules of multiplication that designated for the group's elements, thereby define a representation. Generally a group could have as many representations as possible depending on the choices of basis. It could be n-dimensional (n=1,2,3, etc). However, among them, the fundamental ones are the irreducible representation. Suppose the representation matrices are labeled as $\Gamma(R)$. When there is a similarity transformation, by which all the matrices can be transformed into a quasi-diagonal form like

$$\Gamma(R) = \begin{pmatrix} \Gamma_1(R) & 0 \\ 0 & \Gamma_2(R) \end{pmatrix} \quad (10)$$

, then the representation is called reducible. If there is no kind of similarity transformations existing, then the representation is irreducible. The trace of a representation matrix $\Gamma(R)$ is defined as a character χ . The rank of the matrix is l . The group elements are categorized into classes. Those matrices belong to the same class if they can be transformed into each other by similarity transformations. The characters of matrices in one class are the same. The number of irreducible representation is equal to the number of classes. The fundamental theorems about irreducible representations and their characters in a detailed discussion can be found in ².

The Hamiltonian should be invariant under all symmetry transformations. These transformations construct a symmetry group of Hamiltonian or a Schrödinger equation group. It is proven that (a) the eigenfunctions with the same energy level form a basis for an irreducible representation of the Schrödinger equation group, (b) without the accidental degeneracy (the degeneracy not caused by symmetry), an irreducible representation of Hamiltonian is to characterize a definite energy level. The dimension of irreducible representation is equal to the degree of degeneracy. Hence without the exact knowledge of energy bands, the irreducible representations of group provide insights into the degeneracy of bands and the transformation properties of the basis functions. The theoretical analysis from group theory and experimental data indicates the bottom of conduction band at the $\vec{k}=0$ point has the symmetry properties of atomic s functions under the operation of T_d group; the top of valence bands at the $\vec{k}=0$ point has the

symmetry properties of atomic p functions under the operation of tetrahedral T_d group (for the diamond structure Ge, or the zinc blende structure InSb).

The Löwdin perturbation method

For a Hamiltonian of system, assume that the eigenstates have been figured out and these states are zeroth-order or "unperturbed" from their mutual interactions. These eigenfunctions can be divided into two classes (A) and (B). Those arbitrary energy states that are of interest are put in class (A). Those less important bands are put in the class (B). Their influences on bands in class (A) are treated as perturbations. Löwdin's theorem³ claims that the matrix elements between the j and j' states in class (A) denoted as $U_{jj'}^A$ are,

$$U_{jj'}^A = H_{jj'} + \sum_r^B \frac{H'_{jr} H'_{rj'}}{U - H_{rr}} \quad (11)$$

where the first term $H_{jj'}$ is the exact matrix element between basis functions in class (A), while the second term describes the perturbations from the partitioned bands in class (B). The indices r run through the energy states in class (B).

The Luttinger-Kohn Hamiltonian

Here is only focus on the valence bands which are triply degenerate. The basis functions at $\vec{k} = 0, |X\rangle, |Y\rangle$ and $|Z\rangle$, are chosen to be in class (A), the rest of bands (include the conduction band) are put in class (B). Here the momentum operators \hat{p}_x, \hat{p}_y and \hat{p}_z are transformed the same as the coordinates x, y and z . Under the operations of $3C_2$ of T_d group², the coordinates transform like $\{x \rightarrow -x, y \rightarrow -y, z = z\}, \{x \rightarrow -x, y \rightarrow y, z = -z\}$ and $\{x \rightarrow x, y \rightarrow -y, z = -z\}$. Hence ,

$$\langle X | \hat{p}_x | X \rangle = -\langle X | \hat{p}_x | X \rangle = 0, \quad \langle X | \hat{p}_y | X \rangle = -\langle X | \hat{p}_y | X \rangle = 0, \quad \langle X | \hat{p}_z | X \rangle = -\langle X | \hat{p}_z | X \rangle = 0, \quad (12)$$

$$\langle X | H' | X \rangle = \frac{\hbar}{m_0} \{ \langle X | \hat{p}_x | X \rangle k_x + \langle X | \hat{p}_y | X \rangle k_y + \langle X | \hat{p}_z | X \rangle k_z \} = 0, \quad (13)$$

and so on. Hence the Luttinger-Kohn Hamiltonian H^{LK} has the matrix elements,

$$H_{11}^{LK} = \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar^2}{m_0^2} \sum_{\alpha,\beta} k_\alpha k_\beta \sum_r^{(B)} \frac{\langle X | \hat{p}_\alpha | r \rangle \langle r | \hat{p}_\beta | X \rangle}{U_v - U_r} \quad (14)$$

$$H_{12}^{LK} = \frac{\hbar^2}{m_0^2} \sum_{\alpha,\beta} k_\alpha k_\beta \sum_r^{(B)} \frac{\langle X | \hat{p}_\alpha | r \rangle \langle r | \hat{p}_\beta | Y \rangle}{U_v - U_r} \quad (15)$$

according to the Lödwin's theorem. The Hamiltonian of Schrödinger equation is invariant under all the symmetry transformations. It implies the matrix element H_{11}^{LK} and H_{12}^{LK} should coincide itself as the result of symmetry transformation. Under the operations of $3C_4$, the product obeys $\langle X | \hat{p}_x | r \rangle \langle r | \hat{p}_x | X \rangle = -\langle X | \hat{p}_x | r \rangle \langle r | \hat{p}_x | X \rangle$, and $\langle X | \hat{p}_x | r \rangle \langle r | \hat{p}_x | X \rangle = 0^2$. It is concluded that H_{11}^{LK} doesn't contain any $k_x k_y$ term. By examining the rest operations in the crystal space group T_d on the interactions terms of Eq.(14)-(15), the Luttinger-Kohn Hamiltonian must be written into ⁴,

$$(H^{LK}) = \begin{pmatrix} Lk_x^2 + M(k_y^2 + k_z^2) & Nk_x k_y & Nk_x k_z \\ Nk_x k_y & Ly^2 + M(k_z^2 + k_x^2) & Nk_y k_z \\ Nk_x k_z & Nk_y k_z & Lk_z^2 + M(k_x^2 + k_y^2) \end{pmatrix}, \quad (16)$$

where the notations are,

$$L = \frac{\hbar^2}{2m_0} + \frac{\hbar^2}{m_0} \sum_{r,c}^{(B)} \frac{|\langle X | \hat{p}_x | r \rangle|^2}{U_v - U_r}, \quad M = \frac{\hbar^2}{2m_0} + \frac{\hbar^2}{m_0} \sum_{r,c}^{(B)} \frac{|\langle X | \hat{p}_y | r \rangle|^2}{U_v - U_r} \quad (17)$$

$$N = \frac{\hbar^2}{m_0^2} \sum_{r,c}^{(B)} \frac{\langle X | \hat{p}_x | r \rangle \langle r | \hat{p}_y | Y \rangle + \langle X | \hat{p}_y | r \rangle \langle r | \hat{p}_x | Y \rangle}{U_v - U_r} \quad (18)$$

The commonly used Luttinger constants are defined from the L , M and N parameters,

$$\frac{\hbar^2}{2m_0} \gamma_1 = -\frac{1}{3}(L+2M) \quad \frac{\hbar^2}{2m_0} \gamma_2 = -\frac{1}{6}(L-M), \quad \frac{\hbar^2}{2m_0} \gamma_3 = -\frac{1}{6}N \quad (19)$$

The basis functions for spin-orbit coupling

For the description of valence bands, it is more convenient to describe the atomic p state using the two good quantum numbers, orbital $l=1$ and magnetic $m=1,0,-1$. l and m are

the eigenvalues for the two commuting operators: the squared angular momentum \hat{L}^2 and the projected angular momentum \hat{L}_z as,

$$\hat{L}^2|l, m\rangle = \hbar^2 l(l+1)|l, m\rangle, \hat{L}_z|l, m\rangle = m\hbar|l, m\rangle \quad (20)$$

where the eigenfunctions are the spherical harmonics which are linear superpositions of the three spatial function X , Y and Z ,

$$|1,0\rangle = |z\rangle, \quad |1,\pm\rangle = \mp \frac{1}{\sqrt{2}}(|X\rangle \pm i|Y\rangle) \quad (21)$$

For the description of conduction band, there is orbital $l=0$, $m=0$ and the eigenfunction is $|0,0\rangle = |s\rangle$. When the spin-orbit coupling is taken into account, the conduction band is two-fold degenerate and the valence band is six-fold degenerate. The basis function for the conduction band is $|s \uparrow\rangle$, $|s \downarrow\rangle$ and the basis functions for valence band are $|X \uparrow\rangle$, $|Y \uparrow\rangle$, $|Z \uparrow\rangle$, $|X \downarrow\rangle$, $|Y \downarrow\rangle$ and $|Z \downarrow\rangle$, respectively. Here the spinors \uparrow for spin up and \downarrow for spin down are multiplied. The total momentum operator is defined as $\hat{J} = \hat{L} + \hat{s}$ and the projected total momentum operator $\hat{J}_z = \hat{L}_z + \hat{s}_z$. \hat{s} is the spin operator and \hat{s}_z is the spin's projection. The eigenfunctions for spins,

$$\hat{s}^2|s, m_s\rangle = \hbar^2 s(s+1)|s, m_s\rangle, \quad \hat{s}_z|s, m_s\rangle = \hbar m_s|s, m_s\rangle \quad (22)$$

where s assumes the value $1/2$ and m_s assume the value $1/2$ and $-1/2$. Neither \hat{L} or \hat{s} commute with the Hamiltonian. In fact, the squared total momentum \hat{J}^2 and projected total momentum \hat{J}_z are commute with the Hamiltonian. The eigenfunctions are derived from,

$$\hat{J}^2|j, M\rangle = j(j+1)\hbar^2|j, M\rangle, \quad \hat{J}_z|j, M\rangle = m\hbar|j, M\rangle \quad (23)$$

where the possible values of quantum number of j are $3/2$ and $1/2$. Thus one has the four-fold degenerate valence band labeled by $|3/2, 3/2\rangle$, $|3/2, 1/2\rangle$, $|3/2, -1/2\rangle$, $|3/2, -3/2\rangle$ and a two-fold valence band defined by $|1/2, -1/2\rangle$, $|1/2, 1/2\rangle$. Hence the six-fold degenerate valence bands are split into four-fold degenerate bands Γ_8 and two-fold degenerate Γ_7 under the influences from the spin-orbit coupling. Here Γ_8 and Γ_7 are notations for the irreducible representations of double group T_d at Brillouin zone center Γ . The energy

splitting between Γ_8 valence band and Γ_7 spin split-off band is Δ . This is a result from the theory of double group T_d and is confirmed by experimental measurements. Similarly, the two-fold conduction band Γ_6 are denoted as $|1/2, -1/2\rangle$, $|1/2, 1/2\rangle$. The total angular momentum states can be written into the expansion of uncoupled and individual basis,

$$|j, M\rangle = \sum_{m=-1}^1 \sum_{m_s=-1/2}^{1/2} C(l, m, s, m_s, j, M) |l, m\rangle |s, m_s\rangle, \quad (24)$$

where $C(l, m, s, m_s, j, M)$ are named Clebsch-Gordan (CG) coefficients in atomic physics. With the help of the tabled CG coefficients and the proper introduction of phase, the Bloch basis functions (the zero is dropped in the subscript for u) are chosen to be,

$$\begin{aligned} u_1 &= \left| c, \frac{1}{2}, \frac{1}{2} \right\rangle = |S \uparrow\rangle \\ u_2 &= \left| c, \frac{1}{2}, -\frac{1}{2} \right\rangle = |S \downarrow\rangle \\ u_3 &= \left| v, \frac{3}{2}, \frac{3}{2} \right\rangle = \frac{1}{\sqrt{2}} |(X + iY) \uparrow\rangle \\ u_4 &= \left| v, \frac{3}{2}, \frac{1}{2} \right\rangle = -i\sqrt{\frac{2}{3}} |Z \uparrow\rangle + i\frac{1}{\sqrt{6}} |(X + iY) \downarrow\rangle, \\ u_5 &= \left| v, \frac{3}{2}, -\frac{1}{2} \right\rangle = \sqrt{\frac{2}{3}} |Z \downarrow\rangle + \frac{1}{\sqrt{6}} |(X - iY) \uparrow\rangle \\ u_6 &= \left| v, \frac{3}{2}, -\frac{3}{2} \right\rangle = i\frac{1}{\sqrt{2}} |(X - iY) \downarrow\rangle, \\ u_7 &= \left| v, \frac{1}{2}, \frac{1}{2} \right\rangle = \frac{1}{\sqrt{3}} |(X + iY) \downarrow\rangle + \sqrt{\frac{1}{3}} |Z \uparrow\rangle \\ u_8 &= \left| v, \frac{1}{2}, -\frac{1}{2} \right\rangle = \frac{1}{\sqrt{3}} |(X - iY) \uparrow\rangle - \sqrt{\frac{1}{3}} |Z \downarrow\rangle \end{aligned}$$

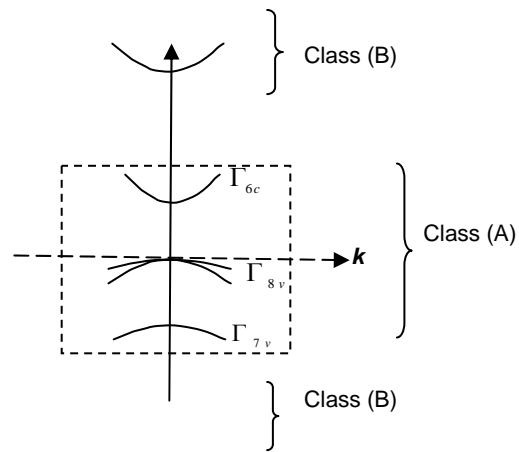


Figure The energy bands and the Löwdin perturbation method.