

$\mathbf{k} \cdot \mathbf{p}$  theory of semiconductors  
EE 698D, Advanced Semiconductor Physics  
Debdeep Jena (djena@nd.edu)  
Department of Electrical Engineering  
University of Notre Dame  
(Fall 2004)

## Introduction

### Semiconductor bandstructure

The energy-eigenvalue problem of a semiconductor crystal yields solutions that form what is known as the bandstructure of the semiconductor. It is more commonly known as the  $\varepsilon - \mathbf{k}$  diagram, or the ‘dispersion’-curve of the material.

Most physical phenomena (electronic, optical, magnetic) in semiconductors can be understood by looking at a small portion of the bandstructure. These points are the lowest points in the conduction band and the highest points in the valence band. The highest point of the valence bands are known as the  $\Gamma$ -point, and constitute the  $(k_x = 0, k_y = 0, k_z = 0)$  point in the  $\mathbf{k}$ -space. In most compound semiconductors, the maximum of the valence band and the minimum of the conduction band occur at the same point in the  $\mathbf{k}$ -space - at the  $\Gamma$ -point. Such semiconductors are called direct-gap semiconductors and form the core of most optical devices. If the minimum of the conduction band is reached at some other point in the  $\mathbf{k}$ -space, the semiconductor is indirect-gap. The elemental semiconductors Si and Ge are of this type. To understand the evolution of bandstructure in general and  $\mathbf{k} \cdot \mathbf{p}$  method for calculating bandstructure in particular, it is essential we start with the idea of Bloch waves.

### Bloch Waves

Consider a periodic *lattice*<sup>1</sup> in space of volume  $\Omega$  with period  $\mathbf{R}$ . Bloch theorem states that the solution of Schrodinger equation for the periodic lattice is of the form

$$\psi(\mathbf{k}, \mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r})u(\mathbf{k}, \mathbf{r}), \quad (1)$$

where  $u(\mathbf{k}, \mathbf{r})$  is the Bloch *lattice* function, and  $\psi(\mathbf{k}, \mathbf{r})$  is the Bloch *wavefunction*. Bloch lattice function has the same translational symmetry as the lattice.

If we have an electron in the above lattice with no potential ( $V(\mathbf{r}) = 0$ )<sup>2</sup>, the Bloch wavefunctions are of the form

$$\psi(\mathbf{k}, \mathbf{r}) = \frac{1}{\sqrt{\Omega}} \exp(i\mathbf{K} \cdot \mathbf{r}), \quad (2)$$

---

<sup>1</sup>A lattice is a mathematical set of points.

<sup>2</sup>The nearly-free electron model or the empty-lattice model, as distinct from the free-electron model, which has no potential *and* no lattice.

where the total electron wavevector  $\mathbf{K} = \mathbf{k} + \mathbf{G}$  can be split into a wavevector  $\mathbf{k}$  in the first Brillouin Zone in the  $k$ -space representation of the crystal, and  $\mathbf{G}$  is a reciprocal lattice vector, defined by  $\mathbf{G} \cdot \mathbf{R} = 2\pi m$ ,  $m$  being an integer. Thus, the reciprocal lattice vector  $\mathbf{G}$  brings the total wavevector back to the Brillouin Zone. Then, we can write the Bloch wavefunction as

$$\psi(\mathbf{k}, \mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r}) \underbrace{\frac{1}{\sqrt{\Omega}} \exp(i\mathbf{G} \cdot \mathbf{r})}_{u(\mathbf{k}, \mathbf{r})}, \quad (3)$$

where the Bloch lattice function  $u(\mathbf{k}, \mathbf{r})$  is clearly periodic with the lattice periodicity. Note that for the nearly-free electron model, we actually *know* the Bloch lattice functions, since we know  $\mathbf{G}$ .

Given this, and the lattice symmetries, it is a simple matter to obtain the ‘dispersion’ relation. The dispersion is nothing but the free electron dispersion, but with the symmetries (translational, rotational, compound-symmetries) built into it. We can now look at the ‘reduced zone’ representation of the dispersion, since we are interested in looking at the energy as a function of the ‘reduced’ wavevector  $\mathbf{k}$ , not the total wavevector  $\mathbf{K}$ . The reduced wavevector  $\mathbf{k}$  is known as the Bloch-wavevector, distinct from the *total* wavevector  $\mathbf{K}$  by a reciprocal lattice vector. The respective energies are given by

$$\varepsilon(\mathbf{k}) = \frac{\hbar^2(\mathbf{k} + \mathbf{G})^2}{2m_0}, \quad (4)$$

which becomes  $\hbar^2(\mathbf{k} + \frac{2\pi}{a}n)^2/2m_0$  for the 1-Dimensional case with  $n = 0, \pm 1, \pm 2, \dots$ . This leads to many parabolic energy-dispersions; that is, the free-electron dispersion is repeated periodically in the  $k$ -axis, as shown in Figure 1. The points in energy where the parabolas from different  $G^s$  intersect are the points of degeneracy.

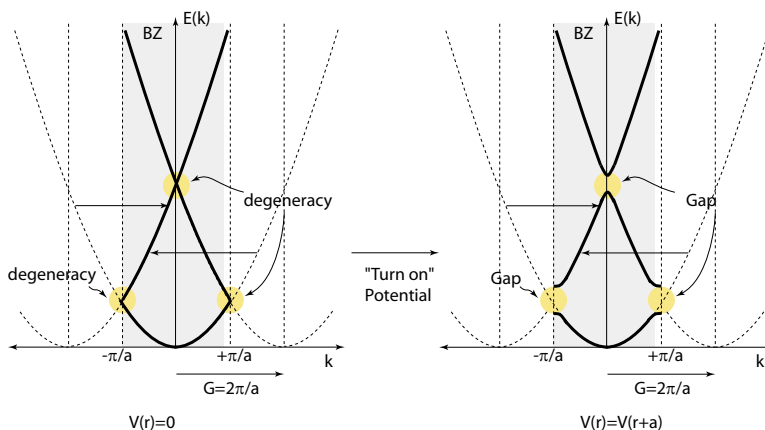


Figure 1: Bandstructure in the nearly-free electron model. Note that for 1-D, there can be degeneracies *only* at the zone center ( $\mathbf{k} = \mathbf{0}$  or at the BZ edge ( $\mathbf{k} = \pm\pi/a$ )). For 2-D and 3-D cases, there can be *accidental* degeneracies at other points in the BZ as well. Can you prove it?

For example, considering a 1-D lattice of spacing  $a$  we have a dispersion as shown in Figure 1. Note specially the degeneracies at the zone center ( $\mathbf{k} = \mathbf{0}$  and at the BZ edge ( $\mathbf{k} = \pm\pi/a$ ). The reciprocal lattice vector  $G = 2\pi/a$  has been used to translate all wavevectors to the 1st BZ, leading to *multiple* energies for the same  $\mathbf{k}$ . However, we have to keep mind that the different  $\varepsilon(\mathbf{k})$  values are actually energies with different  $\mathbf{G}$  values, and hence different  $\mathbf{K}$ . The different sets of  $\varepsilon(\mathbf{k})$  are the different *bands*. One should take special note of the degeneracies, highlighted by orange circles.

The degeneracies can be<sup>3</sup> split in the presence of the crystal potential. If it turns out that all three degeneracies are split by the crystal potential, the bands are now separated by gaps; the magnitudes of the gaps are determined by the magnitude of the crystal potential<sup>4</sup>.

Thus, the periodic crystal potential splits the allowed energies into bands, separated by forbidden gaps. One can show easily that the number of states (in  $-\pi/a \leq k \leq +\pi/a$ ) in each band in the 1st Brillouin Zone is equal to the number of atoms  $N$  in the whole crystal. Allowing for spin degeneracy of 2, the total number of electrons one can fit into a band is  $2N$ .

## Perturbation theory

For  $\mathbf{k} \cdot \mathbf{p}$  theory, we will need some elementary results of time-independent perturbation theory. We list the results here. They can be found in any textbook of quantum mechanics. There are two levels of the solution - the first is degenerate perturbation theory, and the next non-degenerate theory. For  $\mathbf{k} \cdot \mathbf{p}$  theory, the non-degenerate case is important.

Assume that we have solved the Schrodinger equation for a particular potential with Hamiltonian  $H^{(0)}$

$$H^{(0)}|n\rangle = \varepsilon_n^{(0)}|n\rangle, \quad (5)$$

and obtained the eigenfunctions  $|n\rangle$  and eigenvalues  $\varepsilon_n^{(0)}$ . Now let us see how the solution set (eigenfunction, eigenvalue) changes for a potential that differs from the one we have solved for by a small amount. Denote the new Hamiltonian by  $H = H^{(0)} + W$ , where  $W$  is the perturbation.

If the eigenvalues are non-degenerate, the *first order* energy correction is given by

$$\Delta\varepsilon_n^{(1)} \approx \langle n|W|n\rangle, \quad (6)$$

and there is no correction (to first order, in the absence of non-diagonal matrix elements) in the eigenfunction. This is just the diagonal matrix element of the perturbing potential. The second order correction arises from from non-diagonal terms; the energy correction is given by

$$\Delta\varepsilon_n^{(2)} \approx \sum_{m \neq n} \frac{|\langle n|W|m\rangle|^2}{\varepsilon_n^{(0)} - \varepsilon_m^{(0)}}, \quad (7)$$

---

<sup>3</sup>The degeneracy will be split if the perturbing potential (here the crystal potential  $V(\mathbf{r})$ ) has non-zero matrix elements with the states that are degenerate, i.e.,  $\langle 1|V(\mathbf{r})|2\rangle \neq 0$ . This in turn will be determined by the symmetry properties of the states participating in the degeneracy.

<sup>4</sup>Specifically, gap magnitudes are determined by the Fourier components of the crystal potential at the particular reciprocal lattice vector determining the particular band. Can you give a simple physical reason for this?

where  $|m\rangle$  are all other eigenfunctions. The correction to the eigenfunction is

$$|p\rangle = |n\rangle + \sum_{m \neq n} \frac{\langle m|W|n\rangle}{\varepsilon_n - \varepsilon_m} |m\rangle. \quad (8)$$

Thus, the total perturbed energy is given by

$$\varepsilon_n \approx \varepsilon_n^{(0)} + \Delta\varepsilon_n^{(1)} + \Delta\varepsilon_n^{(2)} = \varepsilon_n^{(0)} + \langle n|W|n\rangle + \sum_{m \neq n} \frac{|\langle n|W|m\rangle|^2}{\varepsilon_n^{(0)} - \varepsilon_m^{(0)}}, \quad (9)$$

and the perturbed eigenfunction is given by the the equation before last.

Some more facts will have a direct impact on bandstructure calculation by  $\mathbf{k} \cdot \mathbf{p}$  method. The total second-order perturbation  $\Delta\varepsilon_n^{(2)}$  arises due to the *interaction* between different eigenvalues. Whether interaction between states occurs or not is determined by the matrix elements  $\langle n|W|m\rangle$ ; if it vanishes, there is no interaction. Whether the states vanish or not can typically be quickly inferred by invoking the *symmetry* properties of the eigenfunctions and the perturbing potential  $W$ .

Let us look at the effect of interaction of a state with energy  $\varepsilon_n^{(0)}$  with all other eignestates. Interacting states with energies  $\varepsilon_m^{(0)}$  higher than  $\varepsilon_n^{(0)}$  will lower the energy  $\varepsilon_n$  by contributing a negative term; i.e., they push the energy down. Similarly, states with energies lower than  $\varepsilon_n^{(0)}$  will push it up. The magnitude of interaction scales inversely with the difference in energies; so, the strongest interaction is with the nearest energy state.

This is all the basic results that we need for  $\mathbf{k} \cdot \mathbf{p}$  theory. The last homework that needs to be done is a familiarity with the consequences of symmetry, which we briefly cover now.

## Symmetry

A brief look at the symmetry properties of the eigenfunctions would greatly simplify solving the final problem, and greatly enhance our understanding of the evolution of bandstructure. First, we start by looking at the energy eigenvalues of the individual *atoms* that constitute the semiconductor crystal. All semiconductors have tetrahedral bonds that have  $sp^3$  hybridization. However, the individual atoms have the outermost (valence) electrons in in  $s$ - and  $p$ -type orbitals. The symmetry (or geometric) properties of these orbitals are made most clear by looking at their angular parts -

$$s = 1 \quad (10)$$

$$p_x = \frac{x}{r} = \sqrt{3} \sin \theta \cos \phi \quad (11)$$

$$p_y = \frac{y}{r} = \sqrt{3} \sin \theta \sin \phi \quad (12)$$

$$p_z = \frac{z}{r} = \sqrt{3} \cos \theta. \quad (13)$$

The spherical s-state and the p-type lobes are depicted in Figure 2. Let us denote the states by  $|S\rangle, |X\rangle, |Y\rangle, |Z\rangle$ .

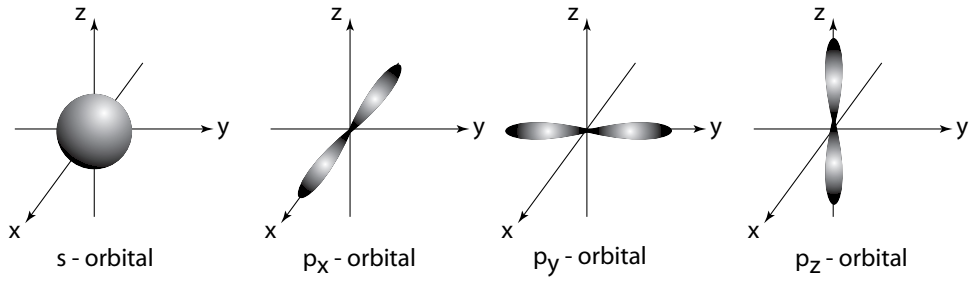


Figure 2: s- and p-orbitals of atomic systems. The s-orbital is spherical, and hence symmetric along all axes; the p-orbitals are antisymmetric or odd along the direction they are oriented - i.e., the  $p_x$  orbital has two lobes - one positive, and the other negative.

Once we put the atoms in a crystal, the valence electrons hybridize into  $sp^3$  orbitals that lead to tetrahedral bonding. The crystal develops its own bandstructure with gaps and allowed bands. For semiconductors, one is typically worried about the bandstructure of the conduction and valence bands only. It turns out that the states near the band-edges behave very much like the the  $|S\rangle$  and the three p-type states that they had when they were individual atoms.

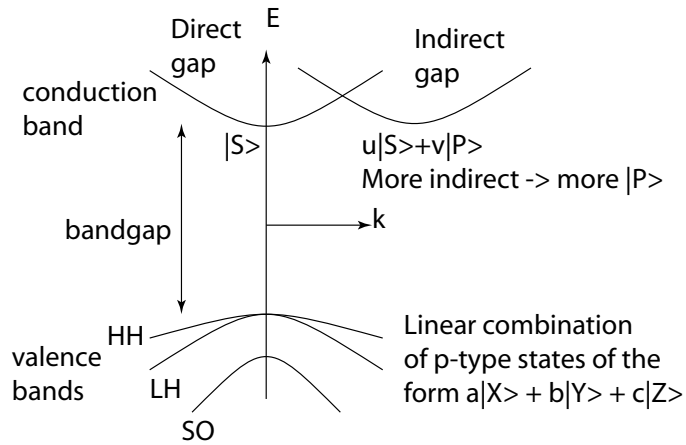


Figure 3: The typical bandstructure of semiconductors. For direct-gap semiconductors, the conduction band state at  $\mathbf{k} = 0$  is s-like. The valence band states are linear combinations of p-like orbitals. For indirect-gap semiconductors on the other hand, even the conduction band minima states have some amount of p-like nature mixed into the s-like state.

For direct-gap semiconductors, for states near the conduction-band minimum ( $\mathbf{k} = 0$ ), the Bloch lattice-function  $u_c(\mathbf{k}, \mathbf{r}) = u_c(0, \mathbf{r})$  possesses the same symmetry properties as a  $|S\rangle$  state<sup>5</sup>. In other words, it has spherical symmetry. The states at the valence band maxima

<sup>5</sup>If the semiconductor has *indirect* bandgap, the conduction-band minimum state is no longer  $|S\rangle$ -like; it

for all bands, on the other hand, have the symmetry of p-orbitals. In general, the valence band states may be written as *linear combinations* of p-like orbitals. Figure 3 denotes these properties. So, we see that the Bloch lattice-functions retain much of the symmetries that the atomic orbitals possess. To put it in more mathematical form, let us say that we have the following Bloch lattice-functions that possess the symmetry of the s- and  $p_x, p_y, p_z$ -type states -  $u_s, u_x, u_y, \&u_z$ . Then, we make the direct connection that  $u_c$  is the same as  $u_s$ , whereas the Bloch lattice-functions of the valence bands  $u_v^s$  are linear combinations of  $u_x, u_y, \&u_z$ .

Without even knowing the exact nature of the Bloch lattice-functions, we can immediately say that the matrix element between the conduction band state and *any* valence band state is

$$\langle u_c | u_v \rangle = 0, \quad (14)$$

i.e., it vanishes. This is easily seen by looking at the orbitals in Figure 2; the p-states are odd along one axis and even along two others; however, the s-states are even. So, the product, integrated over all unit cell is zero. Note that it does not matter which valence band we are talking about, since all of them are linear combinations of p-orbitals.

Next, we look at the momentum-matrix element,  $\langle u_c | \mathbf{p} | u_v \rangle$  between the conduction and valence bands. Since we do not know the linear combinations of  $u_x, u_y, \&u_z$  that form the valence bands yet, let us look at the momentum-matrix elements  $\langle u_s | \mathbf{p} | u_i \rangle$ , with  $i = x, y, z$ . The momentum operator is written out as  $\mathbf{p} = -i\hbar(\mathbf{x}\partial/\partial x + \mathbf{y}\partial/\partial y + \mathbf{z}\partial/\partial z)$ , and it is immediately clear that

$$\langle u_s | \mathbf{p} | u_i \rangle = \langle u_s | p_i | u_i \rangle \equiv P, \quad (15)$$

i.e., it does not vanish. Again, from Figure 2, we can see that the momentum operator along any axis makes the odd-function even, since it is the derivative of that function. The matrix-element is *defined* to be the constant  $P$ . We also note that

$$\langle u_s | p_i | u_j \rangle = 0, (i \neq j). \quad (16)$$

To go into a little bit of detail, it can be shown<sup>6</sup> that the valence band states may be written as the following extremely simple linear combinations

$$u_{HH,\uparrow} = -\frac{1}{\sqrt{2}}(u_x + iu_y), \quad (17)$$

$$u_{HH,\downarrow} = \frac{1}{\sqrt{2}}(\bar{u}_x - i\bar{u}_y), \quad (18)$$

$$u_{LH,\uparrow} = -\frac{1}{\sqrt{6}}(\bar{u}_x + i\bar{u}_y - 2u_z), \quad (19)$$

$$u_{LH,\downarrow} = \frac{1}{\sqrt{6}}(u_x - iu_y + 2\bar{u}_z), \quad (20)$$

$$u_{SO,\uparrow} = -\frac{1}{\sqrt{3}}(\bar{u}_x + i\bar{u}_y + u_z), \quad (21)$$

$$u_{SO,\downarrow} = \frac{1}{\sqrt{3}}(u_x - iu_y - \bar{u}_z) \quad (22)$$

---

has mixed  $|S\rangle$  and p-characteristics.

<sup>6</sup>Broido and Sham, Phys. Rev B, **31** 888 (1986)

and note that

$$\langle u_s | \mathbf{p} | \bar{u}_i \rangle = 0, \quad (23)$$

which in words means that Bloch lattice-functions of opposite spins do not interact. With a detailed look at perturbation theory and symmetry properties, we are in the (enviable!) position of understanding  $\mathbf{k} \cdot \mathbf{p}$  theory with ease.

## $\mathbf{k} \cdot \mathbf{p}$ theory

Substituting the Bloch wavefunction into Schrodinger equation, we obtain a equation similar to the Schrodinger equation, but with two extra terms -

$$\underbrace{\left[ H^0 + \frac{\hbar}{m_0} \mathbf{k} \cdot \mathbf{p} + \frac{\hbar^2 k^2}{2m_0} \right]}_W u(\mathbf{k}, \mathbf{r}) = \varepsilon(\mathbf{k}) u(\mathbf{k}, \mathbf{r}), \quad (24)$$

where  $u(\mathbf{k}, \mathbf{r})$  is the Bloch lattice function.

## No spin-orbit interaction

Let us first look at  $\mathbf{k} \cdot \mathbf{p}$  theory *without* spin-orbit interaction. We will return to spin-orbit interaction later. In the absence of spin-orbit interaction, the three valence bands are degenerate at  $\mathbf{k} = 0$ . Let us denote the bandgap of the (direct-gap) semiconductor by  $E_g$ .

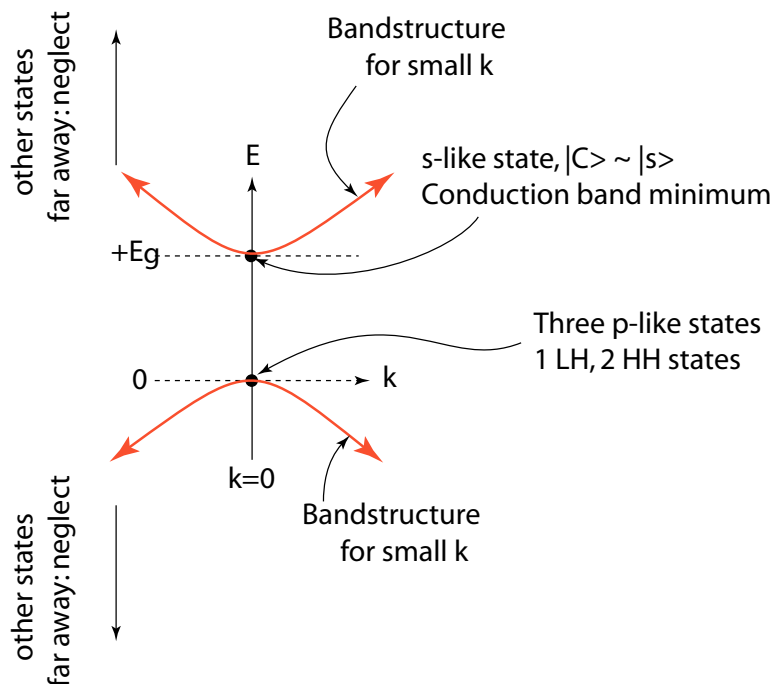


Figure 4:  $\mathbf{k} \cdot \mathbf{p}$  bandstructure in the absence of spin-orbit coupling.

Let us look at the eigenvalues at  $\mathbf{k} = 0$ , i.e., at the  $\Gamma$  point for a direct-gap semiconductor. So the Bloch lattice functions are  $u(0, \mathbf{r})$ . We assume that we have solved the eigenvalue problem for  $\mathbf{k} = 0$ , and obtained the various eigenvalues (call them  $\varepsilon_n(0)$ ) for the corresponding eigenstates (call them  $|n\rangle$ ). We look at only four eigenvalues - that of the conduction band ( $|c\rangle$ ) at  $\mathbf{k} = 0$ , and of the three valence bands - heavy hole ( $|HH\rangle$ ), light hole ( $|LH\rangle$ ) and the split-off band ( $|SO\rangle$ ). In the absence of spin-orbit interaction, they are all degenerate. The corresponding eigenvalues for a cubic crystal are given by ( $\varepsilon_c(0) = +E_g, \varepsilon_{HH}(0) = 0, \varepsilon_{LH}(0) = 0, \varepsilon_{SO}(0) = 0$ ) respectively, where  $E_g$  is the (direct) bandgap.

Using the two results summarized in the last section, we directly obtain that the  $n^{\text{th}}$  eigenvalue is perturbed to

$$\varepsilon_n(\mathbf{k}) \approx \varepsilon_n(\mathbf{0}) + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar^2}{m_0^2} \sum_{m \neq n} \frac{|\langle u_n(0, \mathbf{r}) | \mathbf{k} \cdot \mathbf{p} | u_m(0, \mathbf{r}) \rangle|^2}{\varepsilon_n(0) - \varepsilon_m(0)}, \quad (25)$$

which can be written in a more instructive form as

$$\varepsilon_n(\mathbf{k}) = \varepsilon_n(\mathbf{0}) + \frac{\hbar^2 k^2}{2m^*}, \quad (26)$$

where

$$\frac{1}{m_n^*} = \frac{1}{m_0} \left[ 1 + \frac{2}{m_0 k^2} \sum_{m \neq n} \frac{|\langle u_n(0, \mathbf{r}) | \mathbf{k} \cdot \mathbf{p} | u_m(0, \mathbf{r}) \rangle|^2}{\varepsilon_n(0) - \varepsilon_m(0)} \right] \quad (27)$$

is the reciprocal *effective mass* of the  $n^{\text{th}}$  band.

Let us look at the conduction band effective mass. It is given by

$$\frac{1}{m_c^*} = \frac{1}{m_0} \left[ 1 + \frac{2}{m_0 k^2} \left[ \frac{1}{2} \left( \frac{k^2 P^2}{E_g} \right) + \frac{1}{6} \left( \frac{k^2 P^2}{E_g} \right) + \frac{1}{3} \left( \frac{k^2 P^2}{E_g} \right) \right] \right]. \quad (28)$$

Here we have used the form of Bloch lattice functions given in Equations (17)-(22). Cancelling  $k^2$ , and recasting the equation, we get

$$m_c^* \approx \frac{m_0}{1 + \frac{2P^2}{m_0 E_g}}. \quad (29)$$

To get an estimate of the magnitude of the momentum matrix element  $P$ , we do the following. Looking at the momentum matrix element, it is in the form  $\langle u_c | \mathbf{p} | u_h \rangle$ . The momentum operator will extract the  $k$ -value of the state it acts on. Since the valence (and conduction) band edge states actually occur outside the first Brillouin Zone at  $|\mathbf{k}| = G = 2\pi/a$  and are folded back in to the  $\Gamma$ -point in the reduced zone scheme, it will extract a value  $|P| \approx \hbar \cdot 2\pi/a$ , where  $a$  is the lattice constant of the crystal. Using this fact, and a typical lattice constant of  $a \approx 0.5 \text{ nm}$  we find that

$$\frac{2P^2}{m_0} = \frac{8\pi^2 \hbar^2}{m_0 a^2} \approx 24 \text{ eV}. \quad (30)$$

In reality, the momentum matrix element of most semiconductors is remarkably constant! In fact, it is a very good approximation to assume that  $2P^2/m_0 = 20 \text{ eV}$ , which leads to the relation

$$m_c^* \approx \frac{m_0}{1 + \frac{20 \text{ eV}}{E_g}}, \quad (31)$$

which in the limit of narrow-gap semiconductors becomes  $m_c^* \approx (E_g/20)m_0$ , bandgap in eV. This is a remarkably simple and powerful result!

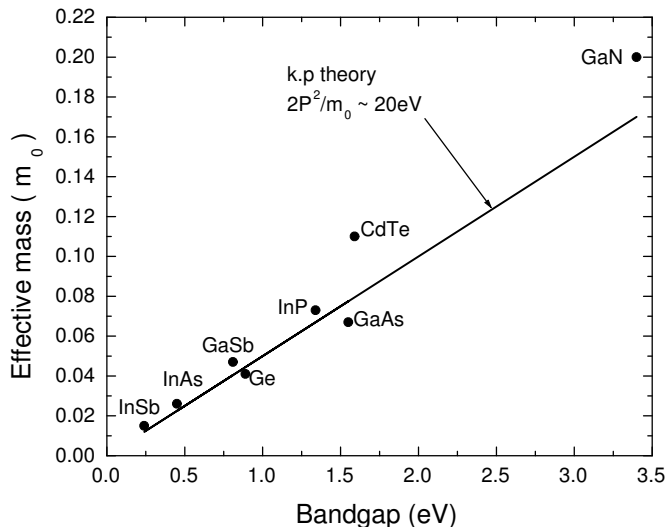


Figure 5: Conduction band effective masses predicted from  $\mathbf{k} \cdot \mathbf{p}$  theory. Note that the straight line is an approximate version of the result of  $\mathbf{k} \cdot \mathbf{p}$  theory, and it does a rather good job for all semiconductors.

It tells us that the effective mass of electrons in a semiconductor increases as the bandgap increases. We also know exactly *why* this should happen as well: the conduction band energies have the strongest interactions with the valence bands. Since valence band states are lower in energy than the conduction band, they ‘push’ the energies in the conduction band upwards, increasing the curvature of the band. This directly leads to a lower effective mass. The linear increase of effective mass with bandgap found from the  $\mathbf{k} \cdot \mathbf{p}$  theory is plotted in Figure 5 with the experimentally measured conduction band effective masses. One has to concede that theory is rather accurate, and does give a very physical meaning to why the effective mass should scale with the bandgap.

Finally, in the absence of spin-orbit interaction, the bandstructure for the conduction band is

$$\varepsilon_c(\mathbf{k}) \approx E_g + \frac{\hbar^2 k^2}{2m_c^*}, \quad (32)$$

where the conduction band effective mass is used. Note that this result is derived from perturbation theory, and is limited to small regions around the  $\mathbf{k} = 0$  point only. One rule of thumb is that the results from this analysis hold only for  $|\mathbf{k}| \ll 2\pi/a$ , i.e., far from the BZ edges.

## With spin-orbit interaction

What is spin-orbit interaction? First, we have to understand that it is a *purely* relativistic effect (which immediately implies there will be a speed of light  $c$  somewhere!). Putting it in words, when electrons move around the positively charged nucleus at relativistic speeds, the electric field of the nucleus Lorentz-transforms to a magnetic field seen by the electrons. The transformation is given by

$$\mathbf{B} = -\frac{1}{2} \frac{(\mathbf{v} \times \mathbf{E})/c^2}{\sqrt{1 - \frac{v^2}{c^2}}} \approx -\frac{1}{2} \frac{\mathbf{v} \times \mathbf{E}}{c^2}, \quad (33)$$

where the approximation is for  $v \ll c$ . To give you an idea, consider a Hydrogen atom - the velocity of electron in the ground state is  $v \approx \alpha c$  where  $\alpha = \frac{1}{137}$  is the fine structure constant, and the consequent magnetic field seen by such an electron (rotating at a radius  $r_0 = 0.53\text{\AA}$ ) from the nucleus is - hold your breath - 12 Tesla! That is a *very* large field, and should have perceivable effects.

Spin-orbit splitting occurs in the bandstructure of crystal precisely due to this effect. Specifically, it occurs in semiconductors in the valence band, because the valence electrons are very close to the nucleus, just like electrons around the proton in the hydrogen atom. Furthermore, we can make some predictions about the magnitude of splitting - in general, the splitting should be more for crystals whose constituent atoms have higher atomic number - since the nuclei have more protons, hence more field!

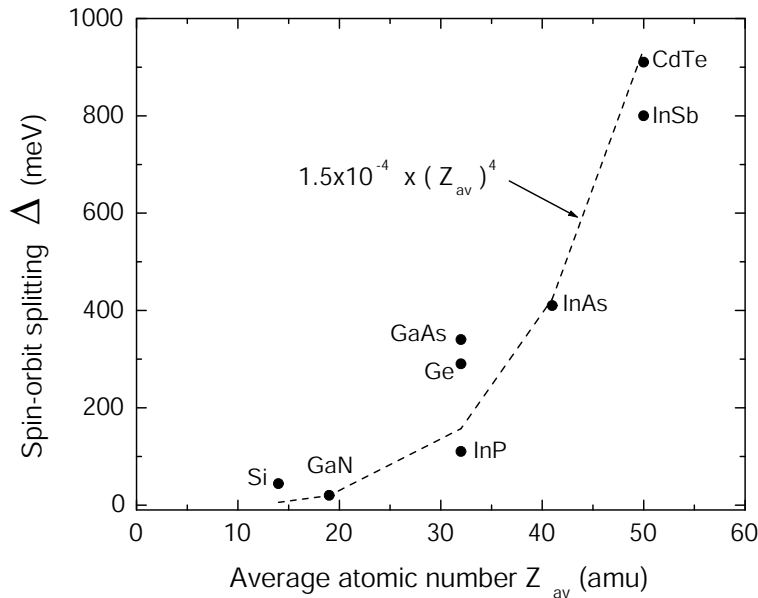


Figure 6: The spin-orbit splitting energy  $\Delta$  for different semiconductors plotted against the average atomic number  $Z_{av}$ . It is a well-known result that the spin-orbit splitting for atomic systems goes as  $Z^4$ ; the situation is not very different for semiconductors.

In fact, the spin-orbit splitting energy  $\Delta$  of semiconductors increases as the *fourth* power of the atomic number of the constituent elements. That is because the atomic number

is equal to the number of protons, which determines the electric field seen by the valence electrons. I have plotted  $\Delta$  against average atomic number in Figure 6, and shown a rough fit to a  $Z_{av}^4$  polynomial. For a detailed account on the spin-orbit splitting effects, refer to the textbooks (Yu and Cardona) mentioned in the end of this chapter.

Let us now get back to the business of building in the spin-orbit interaction to the  $\mathbf{k} \cdot \mathbf{p}$  theory. Spin-orbit coupling splits the 3 degenerate valence bands at  $\mathbf{k} = 0$  into a degenerate HH and LH states, and a split-off state separated by the spin-orbit splitting energy  $\Delta$ . The eigenvalues at  $\mathbf{k} = 0$  are thus given by ( $\varepsilon_c(0) = +E_g, \varepsilon_{HH}(0) = 0, \varepsilon_{LH}(0) = 0, \varepsilon_{SO}(0) = -\Delta$ ) respectively.

These bandgap  $E_g$ , the spin-orbit splitting  $\Delta$ , and the momentum matrix element  $P$  (or, equivalently, the conduction-band effective mass  $m_c^*$ ) evaluated in the last section are the *inputs* to the  $\mathbf{k} \cdot \mathbf{p}$  theory to calculate bandstructure - that is, they are *known*.

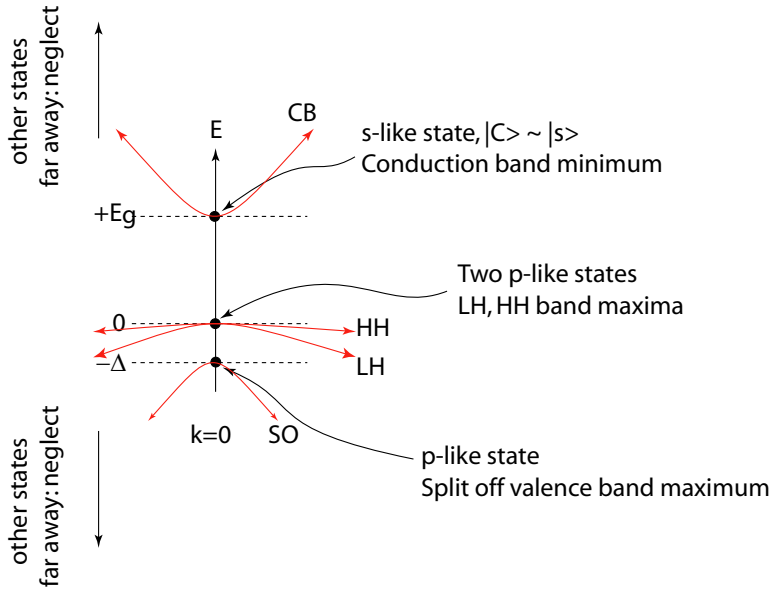


Figure 7:  $\mathbf{k} \cdot \mathbf{p}$  bandstructure with spin-orbit splitting.

Using the same results as for the case without spin-orbit splitting, it is rather easy now to show the following. The bandstructure around the  $\Gamma$  point for the four bands and the corresponding effective masses can be written down. For the conduction band, we have

$$\varepsilon_c(k) \approx E_g + \frac{\hbar^2 k^2}{2m_c^*}, \quad (34)$$

where the effective mass is now given by

$$\frac{1}{m_c^*} = \frac{1}{m_0} \left[ 1 + \frac{2}{m_0 k^2} \left[ \frac{1}{2} \left( \frac{k^2 P^2}{E_g} \right) + \frac{1}{6} \left( \frac{k^2 P^2}{E_g} \right) + \frac{1}{3} \left( \frac{k^2 P^2}{E_g + \Delta} \right) \right] \right], \quad (35)$$

which can be re-written as

$$m_c^* \approx \frac{m_0}{1 + \frac{2P^2}{3m_0} \left( \frac{2}{E_g} + \frac{1}{E_g + \Delta} \right)}, \quad (36)$$

which is the same as the case without the SO-splitting if one puts  $\Delta = 0$ .  $2P^2/m_0 \approx 20eV$  is still valid.

Spin-orbit splitting causes changes in the valence bandstructure. We chose not to talk about valence bands in the last section, since the degeneracy prevents us from evaluating the perturbed eigenvalues. However, with spin-orbit splitting, it is easy to show the following.

The HH valence bandstructure is that of a free-electron, i.e., the effective mass is the same as free-electron mass; so,

$$\varepsilon_{HH}(k) = -\frac{\hbar^2 k^2}{2m_0}, \quad (37)$$

and the light-hole bandstructure is given by

$$\varepsilon_{LH}(k) = -\frac{\hbar^2 k^2}{2m_{LH}^*}, \quad (38)$$

where the light-hole effective mass is given by

$$m_{LH}^* = \frac{m_0}{1 + \frac{4P^2}{3m_0 E_g}}. \quad (39)$$

Finally, the split-off valence bandstructure is given by

$$\varepsilon_{SO}(k) = -\Delta - \frac{\hbar^2 k^2}{2m_{SO}^*}, \quad (40)$$

where the split-off hole effective mass is given by

$$m_{SO}^* = \frac{m_0}{1 + \frac{2P^2}{3m_0(E_g + \Delta)}}. \quad (41)$$

This model is known as the Kane-model of  $\mathbf{k} \cdot \mathbf{p}$  bandstructure, after Kane's celebrated paper<sup>7</sup> of 1956. There is a very good section on the uses of this form of bandstructure calculation in the text by S. L. Chuang (Physics of Optoelectronic Devices, 1995).  $\mathbf{k} \cdot \mathbf{p}$  is very useful in calculating optical transition probabilities and oscillator strengths.

The effects of strain can be incorporated into the  $\mathbf{k} \cdot \mathbf{p}$  theory rather easily, and the shifts of bands can be calculated to a great degree of accuracy. The theory is easily extendable to heterostructures, in particular, to quantum wells for calculating density of states, gain in lasers, and so on. The most popular  $\mathbf{k} \cdot \mathbf{p}$  calculations employ what is called a 8-band  $\mathbf{k} \cdot \mathbf{p}$  formalism. Where do the eight bands come from? We have already seen all 8 - it is the four bands we have been talking about all along, with a spin degeneracy of 2 for each band.

To make the calculations more accurate, one can include bands higher than the conduction band and lower than the valence band. However, the effects of these distant bands are weak, and scale inversely as the energy separation, as we have seen. Thus, they are rarely used.

---

<sup>7</sup>E. O. Kane, J. Phys. Chem. Solids, **1**, 82 (1956)

## Further reading

As Kittel states in his text on Solid State Physics, learning how to calculate bandstructure is an *art*, not learnt from book only, but by experience. My personal favorites for bandstructure theory and applications are two books -

1) *Fundamentals of Semiconductors* (Yu and Cardona, Springer, 1999).

Chapter 2 in this comprehensive text has one of the best modern treatments of semiconductor bandstructure. It makes heavy usage of group theory, which can be intimidating for beginners, but nevertheless very rewarding. The authors do not assume that you come all prepared with results from group theory - they actually have 'crystallized' the results that are needed from group theory in the chapter.

2) *Energy Bands in Semiconductors* (Donald Long, Interscience Publishers, 1968).

An old and classic monograph, it still remains one of the few books entirely devoted to the topic. The theory is covered in 80 pages, and the rest of the book analyzes bandstructures of specific materials.