Tutorial for Semi-Empirical

Band-Structure Calculation

Dragica Vasileska

Arizona State University, Tempe, AZ

Electronic Band Structure Calculation

The basis for discussing transport in semiconductors is the underlying electronic *band* structure of the material arising from the solution of the many body Schrödinger equation in the presence of the periodic potential of the lattice, which is discussed in a host of solid state physics textbooks. The electronic solutions in the presence of the periodic potential of the lattice are in the form of Bloch functions

$$\psi_{n,\mathbf{k}} = u_n(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} \tag{1}$$

where **k** is the wavevector, and *n* labels the band index corresponding to different solutions for a given wavevector. The cell-periodic function, $u_n(\mathbf{k})$, has the periodicity of the lattice and modulates the traveling wave solution associated with free electrons.

A brief look at the symmetry properties of the eigenfunctions would greatly enhance the understanding of the evolution of the bandstructure. First, one starts 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 s- and p-type orbitals. The symmetry (or geometric) properties of these orbitals are made most clear by looking at their angular parts

$$
\begin{aligned}
s &= 1 \\
p_x &= \frac{x}{r} = \sqrt{3}\sin\theta\cos\varphi \\
p_y &= \frac{y}{r} = \sqrt{3}\sin\theta\sin\varphi \\
p_z &= \frac{z}{r} = \sqrt{3}\cos\theta
\end{aligned}
\qquad (2)
$$

Let's denote these states by |S>, |X>, |Y> and |Z>. Once one puts 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 the valence bands only. It turns out that the states near the band-edges behave very much like the |S> and the three p-type states that they had when they were individual atoms.

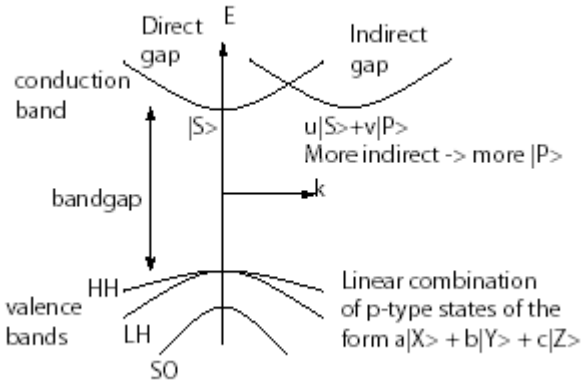


Figure 1: The typical bandstructure of semiconductors. For direct-gap semiconductors, the conduction band state at **k**=0 is s-like. The valence band states are linear combinations of p-like orbitals. For indirect-gap semiconductors on the other hand, even the conduction band minima states have some amount of p-like nature mixed into the s-like state.

Electronic band structure calculation methods can be grouped into two general categories [1]. The first category consists of *ab initio* methods, such as Hartree-Fock or Density Functional Theory (DFT), which calculate the electronic structure from first principles, i.e. without the need for empirical fitting parameters. In general, these methods utilize a variational approach to calculate the ground state energy of a many-body system, where the system is defined at the atomic level. The original calculations were performed on systems containing a few atoms. Today, calculations are performed using approximately 1000 atoms but are computationally expensive, sometimes requiring massively parallel computers.

In contrast to *ab initio* approaches, the second category consists of *empirical* methods, such as the Orthogonalized Plane Wave (OPW) [2], tight-binding [3] (also known as the Linear Combination of Atomic Orbitals (LCAO) method), the $\mathbf{k} \cdot \mathbf{p}$ method [4], and the local [5], or the non-local [6] empirical pseudopotential method (EPM). These methods involve empirical parameters to fit experimental data such as the band-to-band transitions at specific high-symmetry points derived from optical absorption experiments. The appeal of these methods is that the electronic structure can be calculated by solving a one-electron Schrödinger wave equation (SWE). Thus, empirical methods are computationally less expensive than *ab initio* calculations and provide a relatively easy means of generating the electronic band structure. Due to their wide spread usage, in the rest of this section we will review some of the most commonly used ones, namely the empirical pseudopotential method, the tight-binding and the **k·p** method. The empirical pseudopotential method is described in Section 1.1, the tight-binding is discussed in Section 1.2 and the **k·p** method is described in Section 1.3.

Applications of the **k·p** method are given in Section 1.4, which is followed by solutions of the effective mass Schrödinger equation for metal-oxide-semiconductor devices and for heterostructures. We finish this chapter by a brief description of the carrier dynamics that is given in Section 1.5.

Spin-Orbit Coupling

Before proceeding with the description of the various empirical band structure methods, it is useful to introduce the spin-orbit interaction Hamiltonian. The effects of spin-orbit coupling are most easily considered by regarding the spin-orbit interaction energy H_{so} as a perturbation. In its most general form, H_{so} operating on the wavefunctions ψ_k is then given by

Error! Objects cannot be created from editing field codes. (3)

where V is the potential energy term of the Hamiltonian, and σ is the Pauli spin tensor. It can also be written in the following form as an operator on the cell-periodic function

$$H_{SO} = \frac{\hbar}{4m^2c^2}[\nabla V \times \mathbf{p}]\cdot\sigma + \frac{\hbar^2}{4m^2c^2}[\nabla V \times \mathbf{k}]\cdot\sigma \ . \tag{4}$$

The first term is k-independent and is analogous to the atomic spin-orbit splitting term. The second term is proportional to **k** and is the additional spin-orbit energy coming from the crystal momentum. Rough estimates indicate that the effect of the second term on the energy bands is less than 1 percent of the effect of the first term. The relatively greater importance of the first term comes from the fact that the velocity of the electron in its atomic orbit is very much greater than the velocity of a wavepacket made up of wavevectors in the neighborhood of **k**.

The spin-orbit splitting 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 the 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! In fact, the spin-orbit splitting energy Δ 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. In Figure 2, the spin-orbit splitting energy Δ is plotted against an average atomic number and a rough fit using power law is used.

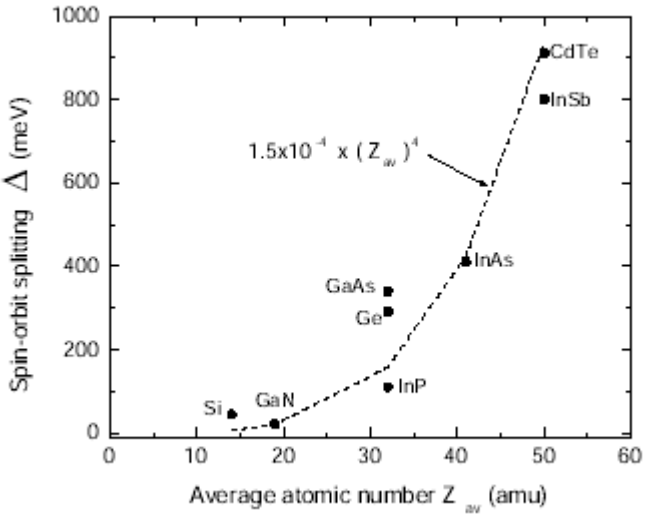


Figure 2. The spin-orbit splitting energy Δ for different semiconductors plotted against the average atomic number Z_{av}.

Rashba and Dresselhaus Spin Splitting

The manipulation of the spin of charge carriers in semiconductors is one of the key problems in the field of spintronics [7]. In the paradigmatic spin transistor proposed by Datta and Das [8], the electron spins, injected from a ferromagnetic contact into a two-dimensional electron system are controllably rotated during their passage from source to drain by means of the Rashba spin-orbit coupling [9]. The coefficient α, which describes the strength of the Rashba spin-orbit coupling, and hence the degree of rotation, can be tuned by gate voltages. This coupling stems from the inversion asymmetry of the confining potential of two-dimensional electron (or hole) systems. In addition to the Rashba coupling, caused by structure inversion asymmetry (SIA), also a Dresselhaus type of coupling contributes to the spin-orbit interaction [10]. The later is due to bulk inversion asymmetry (BIA), and the interface inversion asymmetry (IIA). The BIA and the IIA contributions are phenomenologically inseparable and described below by the generalized Dresselhaus parameter β. Both Rashba and Dresselhaus couplings result in spin splitting of the band and give rise to a variety of spin-dependent phenomena that allow one to evaluate the magnitude of the total spin splitting of electron subbands.

However, usually it is not possible to extract the relative contributions of Rashba and Dresselhaus terms to the spin-orbit coupling. To obtain the Rashba coefficient α, the Dresselhaus contribution is normally neglected. At the same time, Dresselhaus and Rashba terms can interfere in such a way that macroscopic effects vanish though the individual terms are large. For example, both terms can cancel each other, resulting in a vanishing spin splitting in certain k-space directions. This cancellation leads to the disappearance of an antilocalization, the absence of spin relaxation in specific

crystallographic directions, and the lack of SdH beating. In Ref. [11], the importance of both Rashba and Dresselhaus terms was pointed out: turning α such that $\alpha = \beta$ holds, allows one to build a non-ballistic spin-effect transistor.

The consequences of the Rashba and Dresselhaus terms on the electron dispersion and on the spin orientation of the electronic states of the two-dimensional electron gas are summarized below. We consider QWs of the zinc-blende structure grown in the [001] direction. Then, the spin-orbit part of the total Hamiltonian contains the Rashba as well as the Dresselhaus term according to

$$\alpha\left(\sigma_x k_y - \sigma_y k_x\right) + \beta\left(\sigma_x k_x - \sigma_y k_y\right) \tag{5}$$

where **k** is the electron wave-vector, and σ is the vector of the Pauli matrices. Here, the x-axis is aligned along the [100] direction, y-axis is aligned along the [010] direction and z-axis is the growth direction. Note that this Hamiltonian contribution contains only terms linear in **k**. As confirmed experimentally [12], terms cubic in **k** change only the strength of β leaving the Hamiltonian unchanged.

1. The Empirical Pseudopotential Method

The concept of pseudopotentials was introduced by Fermi [13] to study high-lying atomic states. Afterwards, Hellman proposed that pseudopotentials be used for calculating the energy levels of the alkali metals [14]. The wide spread usage of pseudopotentials did not occur until the late 1950s, when activity in the area of condensed matter physics began to accelerate. The main advantage of using pseudopotentials is that only valence electrons have to be considered. The core electrons are treated as if they are frozen in an atomic-like configuration. As a result, the valence electrons are thought to move in a weak one-electron potential.

The pseudopotential method is based on the orthogonalized plane wave (OPW) method due to Herring [2]. In this method, the crystal wavefuntion $\psi_{\mathbf{k}}$ is constructed to be orthogonal to the core states. This is accomplished by expanding $\psi_{\mathbf{k}}$ as a smooth part of symmetrized combinations of Bloch functions $\varphi_{\mathbf{k}}$, augmented with a linear combination of core states. This is expressed as

$$\psi_{\mathbf{k}} = \varphi_{\mathbf{k}} + \sum_t b_{\mathbf{k},t}\Phi_{\mathbf{k},t} \ , \tag{6}$$

where $b_{\mathbf{k},t}$ are orthogonalization coefficients and $\Phi_{\mathbf{k},t}$ are core wave functions. For Si-14, the summation over t in Eq. (6) is a sum over the core states $1s^2\,2s^2\,2p^6$. Since the crystal wave function is constructed to be orthogonal to the core wave functions, the orthogonalization coefficients can be calculated, thus yielding the final expression

$$\psi_{\mathbf{k}} = \varphi_{\mathbf{k}} - \sum_t \langle \Phi_{\mathbf{k},t} | \varphi_{\mathbf{k}} \rangle \Phi_{\mathbf{k},t} \ . \tag{7}$$

To obtain a wave equation for $\varphi_{\mathbf{k}}$, the Hamiltonian operator

$$H = \frac{p^2}{2m} + V_C \ , \tag{8}$$

is applied to Eq. (7), where V_C is the attractive core potential, and the following wave equation results

$$\left(\frac{p^2}{2m} + V_C + V_R \right) \varphi_{\mathbf{k}} = E\varphi_{\mathbf{k}} \, , \tag{9}$$

where V_R represents a short-range, non-Hermitian repulsion potential, of the form

$$V_R = \sum_t \frac{(E - E_t)\langle \Phi_{\mathbf{k},t} | \varphi_{\mathbf{k}} \rangle \Phi_{\mathbf{k},t}}{\varphi_{\mathbf{k}}} \ . \tag{10}$$

E_t in Eq. (10) represents the atomic energy eigenvalue, and the summation over t represents a summation over the core states. The result given in Eq. (9) can be thought of

as wave equation for the pseudo-wave function, $\varphi_{\mathbf{k}}$, but the energy eigenvalue E corresponds to the true energy of the crystal wave function $\psi_{\mathbf{k}}$. Furthermore, as a result of the orthogonalization procedure, the repulsive potential V_R, which serves to cancel the attractive potential V_C, is introduced into the pseudo-wave function Hamiltonian. The result is a smoothly varying pseudopotential $V_P = V_C + V_R$. This result is known as the Phillips-Kleinman cancellation theorem [15] which provides justification why the electronic structure of strongly-bound valence electrons can be described using a nearly-free electron model and weak potentials.

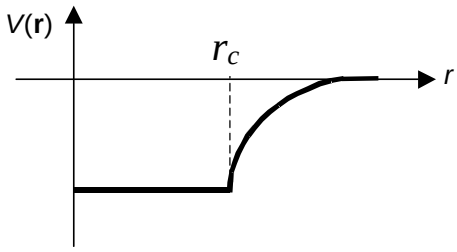
To simplify the problem further, model pseudopotenials are used in place of the actual pseudopotential. Figure 3 summarizes the various models employed. Note that the 3D Fourier transforms (for bulk systems) of each of the above-described model potentials are of the following general form

$$V(q) \sim \frac{Ze^2}{\varepsilon_0 q^2} \cos(qr_c) \ . \tag{11}$$

This q-dependent pseudopotential is then used to calculate the energy band structure along different crystallographic directions, using the procedure outlined in the following section.

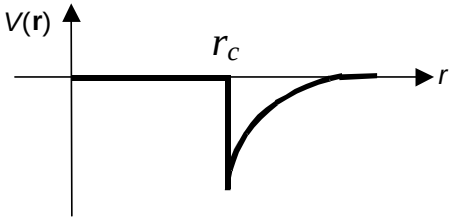
(a) Constant effective potential in the core region:

$$V(r) = \begin{cases} \dfrac{-Ze^2}{4\pi\varepsilon_0 r}; & r > r_C \\ \dfrac{-Ze^2}{4\pi\varepsilon_0 r_C}; & r \le r_C \end{cases}$$



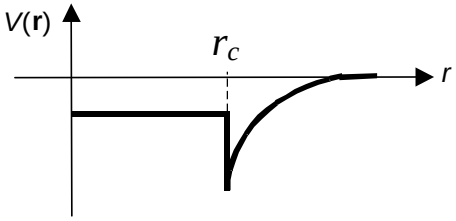
(b) Empty core model:

$$V(r) = \begin{cases} \dfrac{-Ze^2}{4\pi\varepsilon_0 r}; & r > r_C \\ 0; & r \le r_C \end{cases}$$



(c) Model potential due to Heine and Abarenkov:

$$V(r) = \begin{cases} \dfrac{-Ze^2}{4\pi\varepsilon_0 r}; & r > r_C \\ A; & r \le r_C \end{cases}$$



(d) Lin and Kleinman model potentials:

$$V(r) = \begin{cases} 2\dfrac{-Ze^2}{4\pi\varepsilon_0 r}\{1-\exp[-\beta(r-r_C)]\}; & r > r_C \\ 0; & r \le r_C \end{cases}$$

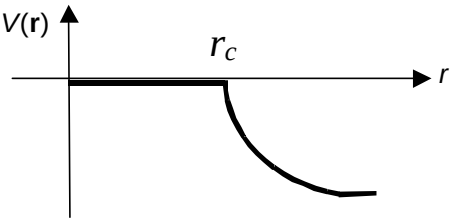


Figure 3. Various model potentials.

1.1 Description of the Empirical Pseudopotential Method

Recall from the previous section that the Phillips-Kleinman cancellation theorem provides a means for the energy band problem to be simplified into a one-electron-like problem. For this purpose, Eq. (9) can be re-written as

$$\left(\frac{p^2}{2m} + V_P \right) \varphi_{\mathbf{k}} = E \varphi_{\mathbf{k}} \ , \tag{12}$$

where V_P is the smoothly varying crystal pseudopotential. In general, V_P is a linear combination of atomic potentials, V_a, which can be expressed as summation over lattice translation vectors **R** and atomic basis vectors $\boldsymbol{\tau}$ to arrive at the following expression

$$V_P(\mathbf{r}) = \sum_{\mathbf{R}} \sum_{\boldsymbol{\tau}} V_a(\mathbf{r} - \mathbf{R} - \boldsymbol{\tau}) \ . \tag{13}$$

To simplify further, the inner summation over $\boldsymbol{\tau}$ can be expressed as the total potential, V_0, in the unit cell located at **R**. Eq. (13) then becomes

$$V_P(\mathbf{r}) = \sum_{\mathbf{R}} V_0(\mathbf{r} - \mathbf{R}) \ . \tag{14}$$

Because the crystal potential is periodic, the pseudopotential is also a periodic function and can be expanded into a Fourier series over the reciprocal lattice to obtain

$$V_P(\mathbf{r}) = \sum_{\mathbf{G}} V_0(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}} \ , \tag{15}$$

where the expansion coefficient is given by

$$V_0(\mathbf{G}) = \frac{1}{\Omega} \int d^3r \ V_0(\mathbf{r}) e^{-i\mathbf{G}\cdot\mathbf{r}} \ , \tag{16}$$

and Ω is the volume of the unit cell.

To apply this formalism to the zincblende lattice, it is convenient to choose a two-atom basis centered at the origin (**R** = 0). If the atomic basis vectors are given by $\boldsymbol{\tau}_1 = \boldsymbol{\tau} = -\boldsymbol{\tau}_2$, where $\boldsymbol{\tau}$, the atomic basis vector, is defined in terms of the lattice constant a_0 as $\boldsymbol{\tau} = a_0(1/8,1/8,1/8)$, $V_0(\mathbf{r})$ can be expressed as

$$V_0(\mathrm{r}) = V_1(\mathrm{r}-\tau) + V_2(\mathrm{r}+\tau) \ , \tag{17}$$

where V_1 and V_2 are the atomic potentials of the cation and anion. Substituting Eq. (17) into Eq. (16), and using the displacement property of Fourier transforms, $V_0(\mathbf{r})$ can be recast as

$$V_0(\mathrm{G}) = e^{i\mathrm{G}\cdot\tau}V_1(\mathrm{G}) + e^{-i\mathrm{G}\cdot\tau}V_2(\mathrm{G}) \ . \tag{18}$$

Writing the Fourier coefficients of the atomic potentials in terms of symmetric ($V_S(\mathbf{G})=V_1+V_2$)) and antisymmetric ($V_A(\mathbf{G})=V_1-V_2$)) *form factors*, $V_0(\mathbf{G})$ is given by

$$V_0(\mathrm{G}) = \cos(\mathrm{G}\cdot\tau)V_S(\mathrm{G}) + i\sin(\mathrm{G}\cdot\tau)V_A(\mathrm{G}) \ , \tag{19}$$

where the prefactors are referred to as the symmetric and antisymmetric structure factors. The form factors above are treated as adjustable parameters that can be fit to experimental data, hence the name empirical pseudopotential method. For diamond-lattice materials, with two identical atoms per unit cell, the V_A=0 and the structure factor is simply $\cos(\mathbf{G}\cdot\boldsymbol{\tau})$. For zinc-blende lattice, like the one in GaAs material system, $V_A\neq 0$ and the structure factor is more complicated.

Now with the potential energy term specified, the next task is to recast the Schrödinger equation in a matrix form. Recall that the solution to the Schrödinger wave equation in a periodic lattice is a Bloch function, which is composed of a plane wave component and a cell periodic part that has the periodicity of the lattice, i.e.

$$\varphi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\sum_{\mathbf{G}'}U(\mathbf{G}')e^{i\mathbf{G}'\cdot\mathbf{r}} \ . \tag{20}$$

By expanding the cell periodic part $u_{\mathbf{k}}(\mathbf{r})$ of the Bloch function appearing in Eq. (20) into Fourier components, and substituting the pseudo-wave function φ_k and potential V_0 into the Schrödinger wave equation, the following matrix equation results

$$\sum_{\mathrm{G}}\left\{\left[\frac{\hbar^2(\mathbf{k}+\mathbf{G})^2}{2m}-E\right]U(\mathbf{G})+\sum_{\mathrm{G'}}V_0(|\mathbf{G}-\mathbf{G}'|)U(\mathbf{G}')\right\}=0\,. \tag{21}$$

The expression given in Eq. (21) is zero when each term in the sum is identically zero, which implies the following condition

$$\left[\frac{\hbar^2(\mathbf{k}+\mathbf{G})^2}{2m}-E\right]U(\mathbf{G})+\sum_{\mathbf{G'}}V_0(|\mathbf{G}-\mathbf{G}'|)U(\mathbf{G}')=0\ . \tag{22}$$

In this way, the band structure calculation is reduced to solving the eigenvalue problem specified by Eq. (22) for the energy E. As obvious from Eq. (20), $U(\mathbf{G}')$ is the Fourier component of the cell periodic part of the Bloch function. The number of reciprocal lattice vectors used determines both the matrix size and calculation accuracy.

The eigenvalue problem of Eq. (22) can be written in the more familiar form $\mathbf{HU} = E\mathbf{U}$, where **H** is a matrix, **U** is a column vector representing the eigenvectors, and E is the energy eigenvalue corresponding to its respective eigenvector. For the diamond lattice, the diagonal matrix elements of **H** are then given by

$$H_{i,j}=\frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{G}_i|^2\ , \tag{23}$$

for $i = j$, and the off-diagonal matrix elements of **H** are given by

$$H_{i,j}=V_S(|\mathbf{G}_i-\mathbf{G}_j|)\cos[(\mathbf{G}_i-\mathbf{G}_j)\cdot\boldsymbol{\tau}], \tag{24}$$

for $i \neq j$. Note that the term $V_s(0)$ is neglected in arriving at Eq. (23), because it will only give a rigid shift in energy to the bands. The solution to the energy eigenvalues and corresponding eigenvectors can then be found by diagonalizing matrix **H**.

1.2 Implementation of the Empirical Pseudopotential Method for Si and Ge

For a typical semiconductor system, 137 plane waves are sufficient, each corresponding to vectors in the reciprocal lattice, to expand the pseudopotential. The reciprocal lattice of a face-centered cubic (FCC), i.e. diamond or zinc-blende structure, is a body-centered cubic (BCC) structure. Reciprocal lattice vectors up to and including the 10^{th}-nearest neighbor from the origin are usually considered which results in 137 plane waves for the zinc-blende structure. The square of the distance from the origin to each equivalent set of reciprocal lattice sites is an integer in the set $|\mathbf{G}^2| = 0, 3, 4, 8, 11, 12, \ldots$ where $|\mathbf{G}^2|$ is expressed in units of $(2\pi/a_o)^2$. Note that the argument of the pseudopotential term V_S in Eq. (24) is the difference between reciprocal lattice vectors. It can be shown that the square of the difference between reciprocal lattice vectors will also form the set of integers previously described. This means that V_S is only needed at discrete points corresponding to nearest-neighbor sites. The pseudopotential, on the other hand, is a continuous quantity. Therefore, its Fourier transform $V_S(q)$ is also a continuous function that is shown in Figure 4. The points corresponding to the first three nearest neighbors are also indicated on this figure.

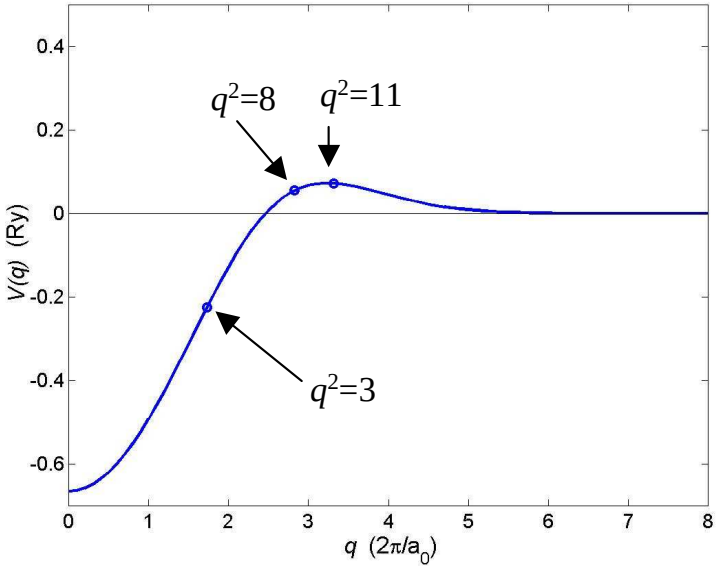


Figure 4. Fourier transform of the pseudopotential. (Note that $q = |\mathbf{G} - \mathbf{G}'|$)

Recall that the pseudopotential is only needed at a few discrete points along the $V(q)$ curve. The discrete points correspond to the q^2-values that match the integer set described previously. There is some controversy, however, regarding the value of V_S as q vanishes. There are two common values seen in the literature: $V_1(0)$= -3/2E_F and $V_1(0)$ = 0. In most cases, the term $V_S(0)$ is ignored because it only gives a rigid shift in energy to the bands. The remaining form factors needed to compute the band structure for non-polar materials correspond to q^2 = 3, 8, and 11. For q^2 = 4, the cosine term in Eq. (24) will always vanish. Furthermore, for values of q^2 greater than 11, $V(q)$ quickly approaches zero. This comes from the fact that the pseudopotential is a smoothly varying function, and only few plane waves are needed to represent it. If a function is rapidly varying in space, then many more plane waves would be required. Another advantage of the empirical pseudopotential method is that only three parameters are needed to describe the band structure of non-polar materials.

Table 1: Local Pseudopotential Form Factors.

Form Factor (Ry)	Si	Ge
V_3	-0.2241	-0.2768
V_8	0.0551	0.0582
V_{11}	0.0724	0.0152

Using the form factors listed in Table 1, where the Si form factors are taken from [16] and the Ge form factors are taken from [17], the band structures for Si and Ge are plotted in Figure 5 [18]. Note that spin-orbit interaction is not included in these simulations. The lattice constants specified for Si and Ge are 5.43Å and 5.65Å, respectively. Si is an indirect band gap semiconductor. Its primary gap, i.e. minimum gap, is calculated from the valence band maximum at the Γ-point to the conduction band minimum along the Δ direction, 85% of the distance from Γ to X. The band gap of Si, is calculated to be E_g^{Si} = 1.08 eV, in agreement with experimental findings. Ge is also an indirect band gap semiconductor. Its band gap is defined from the top of the valence band at Γ to the conduction band minimum at L. The band gap of Ge is calculated to be E_g^{Ge} = 0.73 eV. The direct gap, which is defined from the valence band maximum at Γ to the conduction band minimum at Γ, is calculated to be 3.27 eV and 0.82 eV for Si and Ge, respectively. Note that the curvature of the top valence band of Ge is larger than that of Si. This corresponds to the fact that the effective hole mass of Si is larger than that of Ge. Note that the inclusion of the spin-orbit interaction will lift the triple degeneracy of the bands at the Γ point, leaving doubly-degenerate heavy and light-hole bands and a split-off band

moved downward in energy by few 10's of meV (depending upon the material under consideration).

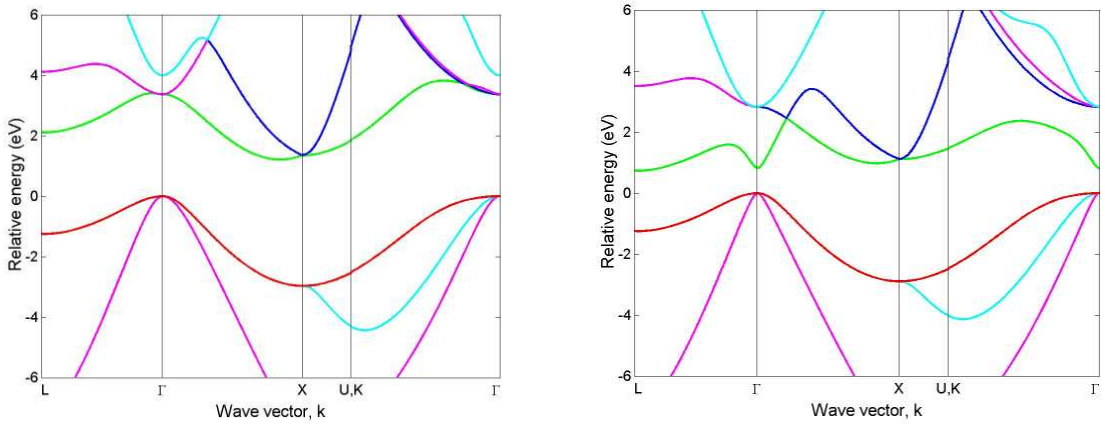


Figure 5. Left panel: band structures of silicon. Right panel: band structure of germanium.

In summary, the local empirical pseudopotential method described in this section is rather good for an accurate description of the optical gaps. However, as noted by Chelikowsky and Cohen [19], when these local calculations are extended to yield the valence-band electronic density of states, the results obtained are far from satisfactory. The reason for this discrepancy arises from the omission of the low cores in the derivation of the pseudopotential in the previous section. This, as previously noted, allowed the usage of a simple plane wave basis. To correct for the errors introduced, an energy-dependent non-local correction term is added to the local atomic potential. This increases the number of parameters needed but leads to better convergence and more exact band-structure results [20,21].

1.3 Empirical Pseudopotential Method for GaN

In the previous section, the implementation of the empirical pseudopotential method for diamond and zinc-blende material systems was explained in details and representative bandstructure simulation results for the Si and Ge material system were presented. In this section we want to point out (without giving much details) that the empirical pseudopotential method can be successfully used for describing the optical gaps in a variety of state-of-the-art materials, such as GaN, AlN and InN, that exhibit a wurtzite structure. The great interest in the group-III nitride material system can be attributed to the promissing electrical and optical properties of the binary compounds AlN, GaN, InN and their alloys. These binary compounds usually crystallize in the wurtzite structure (α-nitrides). In has been demonstrated that, by using modified structure parameters, representative of a wurtzite material system, and discussed in more details in Ref. [22], one can obtain with the Empirical Pseudopotential Method the complete band dispersion of valence and conduction bands and reliably provide necessary band-structure parameters such as the optical gaps and the effective masses. Necessary ingredients in these calculations are the use of the continuous ionic model potentials, which are screened by the model dielectric function derived for semiconductors by Levine and Louie [23]. Such approach allows for a continuous description in the reciprocal space, the explicit inclusion of bond charges, and the exploitation of the ionic model potential transferability to other crystal structures, namely, the wurtzite crystal. In Ref. [24], it was shown by way of an example of wurtzite phase nitrides that crystal-specific anisotropies can be taken into account via proper choice of the screening function. A band-structure of

a GaN material system obtained by using the approach given in Ref. [24] is shown in Figure 6.

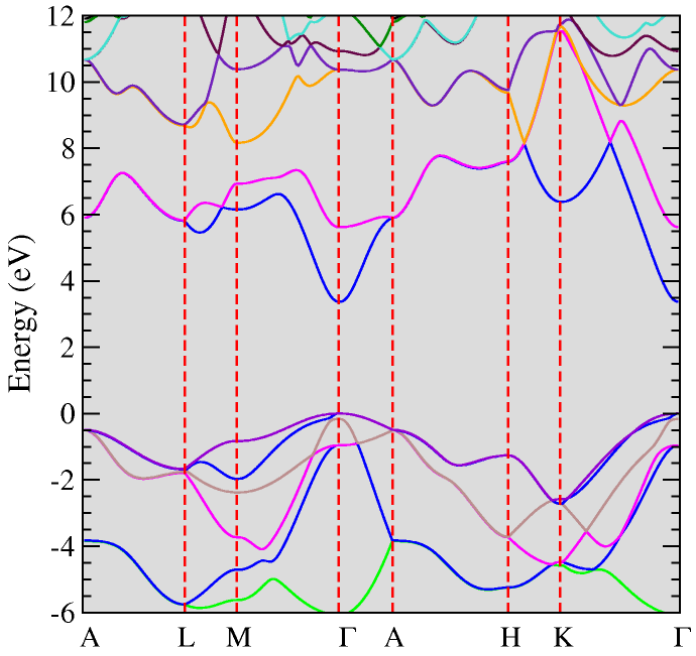


Figure 6. GaN bandstructure as calculated by Yamakawa *et al*. [22].

2. The Tight-Binding Method

Tight binding (TB) is a semi-empirical method for electronic structure calculations. While it retains the underlying quantum mechanics of the electrons, the Hamiltonian is parametrized and simplified before the calculation, rather than constructing it from first principles. The method is detailed by Slater and Koster [25], who laid the initial ground work. Conceptually, tight binding works by postulating a basis set which consists of atomic-like orbitals (i.e. they share the angular momentum components of the atomic orbitals, and are easily split into radial and angular parts) for each atom in the system, and the Hamiltonian is then parametrized in terms of various high symmetry interactions between these orbitals. For tetrahedral semiconductors, as already noted, a conceptual basis set of 1 *s*-like orbital and 3 *p*-like orbitals has been

used. In the most common form of tight binding (nearest neighbor, orthogonal TB), the orbitals are assumed to be orthogonal and interactions between different orbitals are only allowed to be non-zero within a certain distance, which is placed somewhere between the first and second nearest neighbors in the crystal structure. A further simplification which is made, is to neglect three-center integrals (i.e. an interaction between orbitals on atoms A and B which is mediated by the potential on atom C), meaning that each interaction is a function of the distance between the atoms only.

The quantitative description of the method presented below is due to Chadi and Cohen [3]. Let us denote the position of the atom in the primitive cell as

$$\mathbf{r}_{jl} = \mathbf{R}_j + \mathbf{r}_l \,, \tag{25}$$

where R_j is the position of the *j*-th primitive cell and r_l is the position of the atom within the primitive cell. Let $h_l(r)$ be the Hamiltonian of the isolated atom, such that

$$h_l \phi_{ml}\left(\mathbf{r}-\mathbf{r}_{jl}\right) = E_{ml}\phi_{ml}\left(\mathbf{r}-\mathbf{r}_{jl}\right) \,, \tag{26}$$

where E_{ml} and ϕ_{ml} are the eigenvalues and the eigenfunctions of the state indexed by *m*. The atomic orbitals ϕ_{ml} are called Löwdin orbitals [26], and they are different from the usual atomic wavefunctions in that they have been constructed in such a way that wavefunctions centered at different atomic sites are orthogonal to each other. The total Hamiltonian of the system is then

$$H_0 = \sum_{j,l} h_l\left(\mathbf{r}-\mathbf{r}_{jl}\right) \,. \tag{27}$$

Note that the sum over *l* refers to a sum within the different atoms in the basis, therefore, *l* = 1, 2 for diamond and zinc-blende crystals. The unperturbed Bloch functions, that have the proper translational symmetry, are constructed to be of the following form

$$\Phi_{mlk} = \frac{1}{\sqrt{N}} \sum_j e^{i\mathbf{r}_{jl}\cdot\mathbf{k}} \phi_{ml}\left(\mathbf{r} - \mathbf{r}_{jl}\right) . \tag{28}$$

The eigenvalues of the total Hamiltonian $H = H_0 + H_{int}$ (where H_{int} is the interaction part of the Hamiltonian) are then represented as a linear combination of the Bloch functions

$$\Psi_k = \sum_{ml} c_{ml} \Phi_{mlk} \ . \tag{29}$$

Operating with the total Hamiltonian of the system H on Ψ_k, and using the orthogonality of the atomic wavefunctions, one arrives at the following matrix equation

$$\sum_{ml} \left[H_{m'l',ml} - E_k \delta_{mm'} \delta_{ll'} \right] c_{ml} = 0 \ , \tag{30}$$

where the matrix element appearing in the above expression is given by

$$H_{m'l',ml}(\mathbf{k}) = \sum_j e^{i(\mathbf{R}_j + \mathbf{r}_l - \mathbf{r}_{l'})\cdot\mathbf{k}} \left\langle \phi_{mlk}\left(\mathbf{r} - \mathbf{r}_{jl}\right) \middle| H \middle| \phi_{mlk}\left(\mathbf{r} - \mathbf{r}_{jl'}\right) \right\rangle . \tag{31}$$

Note that in the simplest implementation of this method, instead of summing over all the atoms, one sums over the nearest-neighbor atoms only. Also note that the index m represents the *s*- and *p*-states of the outermost electrons ($|s\rangle, |X\rangle, |Y\rangle$ and $|Z\rangle$), and l is the number of distinct electrons in the basis. For the case of tetrahedrally coordinated semiconductors, the number of nearest-neighbors is 4, and are located at

$$\begin{cases} d_1 = (1,1,1)a_0/4 \\ d_2 = (1,-1,-1)a_0/4 \\ d_3 = (-1,1,-1)a_0/4 \\ d_4 = (-1,-1,1)a_0/4 \end{cases} . \tag{32}$$

For a diamond lattice, one also defines the following matrix elements

$$\begin{cases} V_{ss} = 4V_{ss\sigma} \\ V_{sp} = -4V_{sp\sigma}/\sqrt{3} \\ V_{xx} = 4\left[V_{pp\sigma}/3 + 2V_{pp\pi}/3\right] \\ V_{xy} = 4\left[V_{pp\sigma}/3 - V_{pp\pi}/3\right] \end{cases} . \tag{33}$$

As an example, consider the matrix element between two *s*-states

$$H_{s_1,s_2} = \left[e^{i\mathbf{k}\cdot\mathbf{d}_1} + e^{i\mathbf{k}\cdot\mathbf{d}_2} + e^{i\mathbf{k}\cdot\mathbf{d}_3} + e^{i\mathbf{k}\cdot\mathbf{d}_4}\right]\left\langle s_1\middle|H_{\text{int}}\middle|s_2\right\rangle = g_1(\mathbf{k})V_{ss\sigma} \ . \tag{34}$$

Notice the appearance of the Bloch sum $g_1(\mathbf{k})$ in Eq. (34). This observation suggests that for different basis states, there will be four different Bloch sums g_1 through g_4, of the form

$$\begin{cases} g_1(\mathbf{k}) = \left[e^{i\mathbf{k}\cdot\mathbf{d}_1} + e^{i\mathbf{k}\cdot\mathbf{d}_2} + e^{i\mathbf{k}\cdot\mathbf{d}_3} + e^{i\mathbf{k}\cdot\mathbf{d}_4}\right] \\ g_2(\mathbf{k}) = \left[e^{i\mathbf{k}\cdot\mathbf{d}_1} + e^{i\mathbf{k}\cdot\mathbf{d}_2} - e^{i\mathbf{k}\cdot\mathbf{d}_3} - e^{i\mathbf{k}\cdot\mathbf{d}_4}\right] \\ g_3(\mathbf{k}) = \left[e^{i\mathbf{k}\cdot\mathbf{d}_1} - e^{i\mathbf{k}\cdot\mathbf{d}_2} + e^{i\mathbf{k}\cdot\mathbf{d}_3} - e^{i\mathbf{k}\cdot\mathbf{d}_4}\right] \\ g_4(\mathbf{k}) = \left[e^{i\mathbf{k}\cdot\mathbf{d}_1} - e^{i\mathbf{k}\cdot\mathbf{d}_2} - e^{i\mathbf{k}\cdot\mathbf{d}_3} + e^{i\mathbf{k}\cdot\mathbf{d}_4}\right] \end{cases} . \tag{35}$$

It is also important to note that the Hamiltonian matrix elements between a *s*- and a *p*-state on the same atom, or two different *p*-states on the same atom, are zero because of symmetry in diamond and zincblende crystals. The 8×8 secular determinant representing all possible nearest-neighbor interactions between the tight-binding *s*- and *p*-orbitals centered on each atom in the crystal is

$$\begin{vmatrix} & S1 & X1 & Y1 & Z1 & S2 & X2 & Y2 & Z2 \\ S1 & E_s - E_k & 0 & 0 & 0 & V_{ss}g_1 & V_{sp}g_2 & V_{sp}g_3 & V_{sp}g_4 \\ X1 & 0 & E_p - E_k & 0 & 0 & -V_{sp}g_2 & V_{xx}g_1 & V_{xy}g_4 & V_{xy}g_3 \\ Y1 & 0 & 0 & E_p - E_k & 0 & -V_{sp}g_3 & V_{xy}g_4 & V_{xx}g_1 & V_{xy}g_2 \\ Z1 & 0 & 0 & 0 & E_p - E_k & -V_{sp}g_4 & V_{xy}g_3 & V_{xy}g_2 & V_{xx}g_1 \\ S2 & V_{ss}g_1^* & -V_{sp}g_2^* & -V_{sp}g_3^* & -V_{sp}g_4^* & E_s - E_k & 0 & 0 & 0 \\ X2 & V_{sp}g_2^* & V_{xx}g_1^* & V_{xy}g_4^* & V_{xy}g_3^* & 0 & E_p - E_k & 0 & 0 \\ Y2 & V_{sp}g_3^* & V_{xy}g_4^* & V_{xx}g_1^* & V_{xy}g_2^* & 0 & 0 & E_p - E_k & 0 \\ Z2 & V_{sp}g_4^* & V_{xy}g_3^* & V_{xy}g_2^* & V_{xx}g_1^* & 0 & 0 & 0 & E_p - E_k \end{vmatrix} . \quad (36)$$

The tight-binding parameters appearing in Eqs. (33) and (36) are obtained by comparison with empirical pseudopotential calculations, which are presented in Ref. [3].

Table 2. Chadi and Cohen tight-binding parameters [3].

	E_p - E_s	V_{ss}	V_{sp}	V_{xx}	V_{xy}
C	7.40	-15.2	10.25	3.0	8.3
Si	7.20	-8.13	5.88	1.71	7.51
Ge	8.41	-6.78	5.31	1.62	6.82

Using the above-described method one can quite accurately describe the valence bands, whereas the conduction bands are not reproduced that well due to the omission of the interaction with the higher-lying bands. The accuracy of the conduction bands can be improved with the addition of more overlap parameters. However, there are only four conduction bands and the addition of more orbitals destroys the simplicity of the method.

3. The k·p method

In contrast to the previously described empirical pseudopotentials and the tight-binding methods, the **k·p** method is based upon perturbation theory [27,28]. In this method, the energy is calculated near a band maximum or minimum by considering the wavenumber (measured from the extremum) as a perturbation.

3.1 k•p General Description

To get a better understanding of the method, let us assume that the Schrödinger equation studied is mono-dimensional and stationary. To further elaborate the problem, also assume that the particle is surrounded by a potential $V = V_{\sim} + V_{\cup}$, where $V_{\sim}$ is the periodic potential that has the periodicity of the lattice, and $V_{\cup}$ is the confinement potential. For this particle, the mono-dimensional Schrödinger wave equation is

$$H_0\psi(x) = \left[\frac{p^2}{2m} + V(x)\right]\psi(x) = \lambda\psi(x) \ , \tag{37}$$

and $V_{\cup} = 0$ if $x \notin [-x_0, x_0]$; and $V_{\cup} = -V_0$ otherwise. Here, V_0 and x_0 are some arbitrary positive constants. If $V_{\cup}$ is small, then the solutions to the mono-dimensional Schrödinger equation are of the Bloch form (as discussed in the introduction part of this section) and are repeated here for completeness for a 1D case. Namely,

$$\psi_k(x) = e^{i\mathbf{k}\cdot x}u_k(x) \ , \tag{38}$$

where $u_k(x)$ is cell periodic part of the Bloch function. The Schrödinger equation can then be written as

$$H_k u_k(x) = \left[\frac{p^2}{2m} + V(x) + \frac{\hbar}{m} \mathrm{k} \cdot \mathrm{p} \right] u_k(x) = \left[E_k - \frac{\hbar^2 k^2}{2m} \right] u_k(x) \,. \tag{39}$$

The term $(\hbar/m)\mathrm{k}\cdot\mathrm{p}$ is treated as a perturbation to H_0 for determining $u_k(x)$ and E_k in the vicinity of k = 0 in terms of the complete set of cell-periodic wavefunctions and energy eigenvalues at k = 0, which are assumed known. To simplify the form of Eq. (39), it is convenient to define

$$E_k^{'} = E_k - \frac{\hbar^2 k^2}{2m} \,. \tag{40}$$

To deal with this problem, we now assume that we have an orthonormal basis $\{\zeta_i\}_{i=1}^n$ of eigenvectors (associated to their eigenvalues λ_i) of the operator $p^2/2m + V_{\sim}$ that are of a fixed parity (the orbitals may be of *s*- or *p*-type). We then project operator H_k on the finite dimensional space generated by the ζ_i's, to get

$$\begin{aligned} \left\langle \zeta_i \middle| H_k \middle| \zeta_j \right\rangle &= \lambda_j \delta_{ij} + \frac{\hbar}{m} k \left\langle \zeta_i \middle| p \middle| \zeta_j \right\rangle + \left\langle \zeta_i \middle| V_{\cup} \middle| \zeta_j \right\rangle \\ &= \lambda_j \delta_{ij} + k P_{ij} + Q_{ij} \quad , \end{aligned} \tag{41}$$

i.e. we arrive at the symmetric eigenvalue matrix

$$H(k) = Q + \begin{bmatrix} \lambda_1 & \cdots & \left(kP_{ij}\right) \\ \vdots & \ddots & \vdots \\ \left(k\bar{P}_{ij}\right) & \cdots & \lambda_n \end{bmatrix} , \tag{42}$$

the solutions of which provide us the eigenvalues and the corresponding eigenvectors.

3.2 k•p Theory Near the Γ Point and for Bulk Materials

In general practical situation, one either has a bulk-like system or lower-dimensional systems such as 2D and 1D electron gases in which there is a confinement in

one and two directions, respectively. We want to point out that these lower dimensional systems are frequently encountered in state-of-the-art devices which makes this general discussion of the **k•p** method very useful. For a general system, with spin-orbit interaction included in the model, and using the result for the 1D case given in the previous section, the Schrödinger equation is of the following general form:

$$\left[\frac{p^2}{2m_0}+\frac{\hbar}{4m_0^2c^2}(\sigma\times\nabla V)\cdot\mathbf{p}+\frac{\hbar^2k^2}{2m_0}+\frac{\hbar\mathbf{k}}{m_0}\cdot\left(\mathbf{p}+\frac{\hbar}{4m_0c^2}(\sigma\times\nabla V)\right)\right]u_{n\mathbf{k}}(r)=E_{n\mathbf{k}}u_{n\mathbf{k}}(r) \quad (43)$$

The Hamiltonian in Eq. (43) can be divided into two terms

$$\left[H(\mathbf{k}=0)+W(\mathbf{k})\right]u_{n\mathbf{k}}=E_{n\mathbf{k}}u_{n\mathbf{k}}\ , \quad (44)$$

where the only **k**-dependence is preserved in $W(\mathbf{k})$. Next, as in the 1D case, we assume that the local single particle of Hamiltonian $H(\mathbf{k}=0)$ has a complete set of eigenfunctions u_{n0}, i.e.

$$H(\mathbf{k}=0)u_{n0}=E_{n0}u_{n0}\ . \quad (45)$$

An arbitrary ("well behaving") lattice periodic function can be written as a series expansion using the eigenfunctions u_{n0}. We then insert an expansion

$$u_{n\mathbf{k}}=\sum_m c_m^n(\mathbf{k})u_{m0} \quad (46)$$

in Eq. (45) and find matrix equation for determining the unknown coefficients $c_m^n(\mathbf{k})$. We multiply from left by u_{n0}^*, integrate and use the orthogonality of the basis functions, to get

$$\sum_m \left[\left(E_{n0} - E_{n\mathbf{k}} + \frac{\hbar^2 k^2}{2m_0} \right) \delta_{nm} + \frac{\hbar \mathbf{k}}{m_0} \cdot \left\langle u_{n0} \right| \left(\mathbf{p} + \frac{\hbar}{4m_0 c^2} (\sigma \times \nabla V) \right) \left| u_{m0} \right\rangle \right] c_m^n(\mathbf{k}) = 0 \ . \qquad (47)$$

Solving the above matrix equation then gives the exact eigenstates of the Eq. (43). However, this looks good only in principle, the reason being that the calculation becomes increasingly complicated as **k** increases. One has to increase the number of states in the expansion given in Eq. (47) and the calculations become numerically unfeasible. Therefore this approach is practical only for small wavevector values.

When **k** is small, the non-diagonal terms are small and the lowest order solution for eigenstate $u_{n\mathbf{k}} = u_{n0}$ is $c_m^n(\mathbf{k}) = \delta_{nm}$, and the corresponding eigenvalue is given by

$$E_{nk} = E_{n0} + \frac{\hbar^2 k^2}{2m_0} \ . \qquad (48)$$

If the nondiagonal terms are small, one can improve the above result by using the second order perturbation theory

$$E_{n\mathbf{k}} = E_{n0} + \left\langle u_{n0} \left| \frac{\hbar^2 k^2}{2m_0} \right| u_{n0} \right\rangle + \sum_{m \neq n} \frac{\left\langle u_{n0} \left| H_I \right| u_{m0} \right\rangle \left\langle u_{m0} \left| H_I \right| u_{n0} \right\rangle}{E_{n0} - E_{m0}} \ , \qquad (49)$$

where

$$H_I = \frac{\hbar \mathbf{k}}{m_0} \cdot \left(\mathbf{p} + \frac{\hbar}{4m_0 c^2} (\sigma \times \nabla V) \right) \ . \qquad (50)$$

Since the kinetic energy operator is a scalar, the second order eigen-energies can be written as

$$E_{n\mathbf{k}} = E_{n0} + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar^2}{m_0^2} \sum_{m \neq n} \frac{\left| \pi_{nm} \cdot \mathbf{k} \right|^2}{E_{n0} - E_{m0}} , \qquad (51)$$

where

$$\begin{cases} \pi = \mathbf{p} + \dfrac{\hbar}{4m_0c^2}(\sigma \times \nabla V) \ . \\ \pi_{nm} = \langle u_{n0} | \pi | u_{m0} \rangle \end{cases} \tag{52}$$

The vector **k** can be taken outside the integral in Eq. (51). It is seen that the eigenvalue depends in the vicinity of the Γ point quadratically on the wave vector components. Then, Eq. (51) is often written as

$$E_{n\mathbf{k}} = E_{n0} + \frac{\hbar^2}{2} \sum_{\alpha\beta} k_\alpha \frac{1}{\mu_{\alpha\beta}} k_\beta, \quad \alpha, \beta : x, y, z \tag{53}$$

where

$$\frac{1}{\mu_n^{\alpha\beta}} = \frac{1}{m_0}\delta_{\alpha\beta} + \frac{2}{m_0^2} \sum_{m \neq n} \frac{\pi_{mn}^\alpha \pi_{nm}^\beta}{E_{n0} - E_{m0}} \tag{54}$$

is an effective mass tensor.

3.3 Kane's Theory

k•p theory, as discussed in Section ?? is essentially based on perturbation theory. A more exact approach, capable of including strong band to band interactions, is provided by Eq. (47). Note that the inclusion of a complete set of basis states in Eq. (47) is not feasible numerically. However, one can improve the **k•p** theory drastically if it includes in Eq. (47) those bands that are strongly coupled and correct this approximation by treating the influence of distant (energetically) bands perturbatively. This procedure can be made consistently if the electron bands can be divided into two groups. In the first group of bands there is a strong interband coupling – the number of bands in this group is very limited (up to 8, say). The second group of bands is only weakly interacting with the first set. This interaction is treated by perturbation theory. This approach is called Kane's

[27,28] model and it has shown to be very predictive for the III-V compound semiconductors.

Within Kane's theory one constructs a new basis of p-symmetric atomic Bloch states as a linear combination of the "directed orbital" atomic Bloch states $\left|u_{n0}\right\rangle$ discussed in the previous chapter. The new basis set will consist of eigenfunctions of operators J and its component in the z-direction J_z. The new basis set is then denoted by $\left|jm_j\right\rangle$, where j = ½, 3/2 and m_j = j, j-1, …, -j. This gives six subbands. These can be considered together with the s-symmetric conduction band. The resulting 8-band model gives a good description of the electronic structure of III-V semiconductors near the Γ point. The new basis set is given in terms of the directed orbitals shown in Table 3. In the literature, there are another sets of basis functions that differ from the set of functions given here by a unitary transformation $\left|jm_j\right\rangle' = U\left|jm_j\right\rangle, \quad UU^{\dagger} = I$.

Table 3. The atomic basis states at Γ point. The eigenvalues in the fourth column correspond to Eq. (). The zero point of energy has been set to the bottom of the conduction band.

u_i	$\left\|j,m_j\right\rangle$	ψ_{J,m_J}	$E_i(k=0)$	
u_1	$\left\|\frac{1}{2},\frac{1}{2}\right\rangle$	$i\left\|S\uparrow\right\rangle$	0	Γ_6
u_3	$\left\|\frac{3}{2},\frac{1}{2}\right\rangle$	$-\sqrt{\frac{2}{3}}\left\|Z\uparrow\right\rangle+\frac{1}{\sqrt{6}}\left\|X+iY)\uparrow\right\rangle$	$-E_0$	Γ_8
u_5	$\left\|\frac{3}{2},\frac{3}{2}\right\rangle$	$\frac{1}{\sqrt{2}}\left\|X+iY)\uparrow\right\rangle$	$-E_0$	Γ_8

u_7	$\left\|\frac{1}{2},\frac{1}{2}\right\rangle$	$\frac{1}{\sqrt{3}}\left\|X+iY)\downarrow\right\rangle+\frac{1}{\sqrt{3}}\left\|Z\uparrow\right\rangle$	$-E_0-\Delta$	Γ_7
u_2	$\left\|\frac{1}{2},-\frac{1}{2}\right\rangle$	$i\left\|S\downarrow\right\rangle$	0	Γ_6
u_4	$\left\|\frac{3}{2},-\frac{1}{2}\right\rangle$	$-\sqrt{\frac{2}{3}}\left\|Z\downarrow\right\rangle-\frac{1}{\sqrt{6}}\left\|X-iY)\uparrow\right\rangle$	$-E_0$	Γ_8
u_6	$\left\|\frac{3}{2},-\frac{3}{2}\right\rangle$	$\frac{1}{\sqrt{2}}\left\|X-iY)\downarrow\right\rangle$	$-E_0$	Γ_8
u_8	$\left\|\frac{1}{2},-\frac{1}{2}\right\rangle$	$-\frac{1}{\sqrt{3}}\left\|X-iY)\uparrow\right\rangle+\frac{1}{\sqrt{3}}\left\|Z\downarrow\right\rangle$	$-E_0-\Delta$	Γ_7

The atomic Bloch states in Table 3 are eigenstates of the Hamiltonian $H(\mathbf{k}=0)$ and include spin-orbit interaction. Γ_6 corresponds to the conduction band, Γ_8 denotes the heavy-hole ($m_j = \pm 3/2$) and Γ_8 ($m_j = \pm 1/2$) the light-hole band. Γ_7 is known as split-off band. If we neglect in Eq. () the spin-orbit term that depends on the wavevector, i.e. the term

$$\frac{\hbar\mathbf{k}}{m_0}\cdot(\pi-\mathbf{p})=\frac{\hbar\mathbf{k}}{m_0}\cdot\left[\frac{\hbar}{4m_0c^2}(\sigma\times\nabla V)\right], \tag{55}$$

then the matrix presentation of the Hamiltonian

$$H(\mathbf{k})=H(\mathbf{k}=0)+\frac{\hbar^2k^2}{2m_0}+\frac{\hbar\mathbf{k}\cdot\mathbf{p}}{m_0} \tag{56}$$

is given in Table 4 using the basis set of Table 3. Note that this Hamiltonian does not yet include the influence of distant bands which makes the effective mass of the valence band to differ from the electron rest mass. Some notations used in Table 4 are given below

$$
\begin{aligned}
k_{\pm} &= \frac{1}{\sqrt{2}}\left(k_x \pm i k_y\right) \\
E_0 &= E_{\Gamma_6} - E_{\Gamma_8} \\
\Delta &= E_{\Gamma_8} - E_{\Gamma_7} \\
P &= \frac{-i}{m_0}\left\langle S\left|p_x\right|X\right\rangle = \frac{-i}{m_0}\left\langle S\left|p_y\right|Y\right\rangle = \frac{-i}{m_0}\left\langle S\left|p_z\right|Z\right\rangle
\end{aligned}
\tag{57}
$$

We now calculate the dispersion as a function of **k** for the 8-band **k•p** theory by diagonalizing the Hamiltonian in Table 4. For bulk systems, or heterostructures in which we have confinement only in one direction, the Hamiltonian in Table 4 is easy to diagonalize if the z-axis of the coordinate system is taken in the direction of the wavevector. In this case $k_z = k$ and accordingly $k_{\pm} = 0$. This choice is possible since it can be shown that the Hamiltonian is isotropic and, therefore, the eigenvalues and eigenvectors depend on the magnitude of **k**-only. In this coordinate system the Hamiltonian is brought into a block form

Table 4. Eight band Hamiltonian $H(\mathbf{k})$.

	$\left\|iS\uparrow\right\rangle$	$\left\|\frac{3}{2},\frac{1}{2}\right\rangle$	$\left\|\frac{3}{2},\frac{3}{2}\right\rangle$	$\left\|\frac{1}{2},\frac{1}{2}\right\rangle$	$\left\|iS\downarrow\right\rangle$	$\left\|\frac{3}{2},-\frac{1}{2}\right\rangle$	$\left\|\frac{3}{2},-\frac{3}{2}\right\rangle$	$\left\|\frac{1}{2},-\frac{1}{2}\right\rangle$
$\left\|iS\uparrow\right\rangle$	$\frac{\hbar^2k^2}{2m_0}$	$-\sqrt{\frac{2}{3}}P\hbar k_z$	$P\hbar k_+$	$\sqrt{\frac{1}{3}}P\hbar k_z$	0	$-\sqrt{\frac{1}{3}}P\hbar k_-$	0	$-\sqrt{\frac{2}{3}}P\hbar k_-$
$\left\|\frac{3}{2},\frac{1}{2}\right\rangle$	$-\sqrt{\frac{2}{3}}P\hbar k_z$	$-E_0+\frac{\hbar^2k^2}{2m_0}$	0	0	$\sqrt{\frac{1}{3}}P\hbar k_-$	0	0	0
$\left\|\frac{3}{2},\frac{3}{2}\right\rangle$	$P\hbar k_-$	0	$-E_0+\frac{\hbar^2k^2}{2m_0}$	0	0	0	0	0
$\left\|\frac{1}{2},\frac{1}{2}\right\rangle$	$\sqrt{\frac{1}{3}}P\hbar k_z$	0	0	$-E_0+\Delta+\frac{\hbar^2k^2}{2m_0}$	$\sqrt{\frac{2}{3}}P\hbar k_-$	0	0	0
$\left\|iS\downarrow\right\rangle$	0	$\sqrt{\frac{1}{3}}P\hbar k_+$	0	$\sqrt{\frac{2}{3}}P\hbar k_+$	$\frac{\hbar^2k^2}{2m_0}$	$-\sqrt{\frac{2}{3}}P\hbar k_z$	$P\hbar k_-$	$\sqrt{\frac{1}{3}}P\hbar k_z$
$\left\|\frac{3}{2},-\frac{1}{2}\right\rangle$	$-\sqrt{\frac{1}{3}}P\hbar k_+$	0	0	0	$-\sqrt{\frac{2}{3}}P\hbar k_z$	$-E_0+\frac{\hbar^2k^2}{2m_0}$	0	0
$\left\|\frac{3}{2},-\frac{3}{2}\right\rangle$	0	0	0	0	$P\hbar k_+$	0	$-E_0+\frac{\hbar^2k^2}{2m_0}$	0
$\left\|\frac{1}{2},-\frac{1}{2}\right\rangle$	$-\sqrt{\frac{2}{3}}P\hbar k_+$	0	0	0	$\sqrt{\frac{1}{3}}P\hbar k_z$	0	0	$-E_0+\Delta+\frac{\hbar^2k^2}{2m_0}$

$$H=\begin{bmatrix} H_{4\times4} & 0 \\ 0 & H_{4\times4} \end{bmatrix}, \tag{58}$$

where the 4×4 matrix is given by

$$H_{4\times4}=\begin{bmatrix} \frac{\hbar^2k^2}{2m_0} & -\sqrt{\frac{2}{3}}P\hbar k_z & 0 & \sqrt{\frac{1}{3}}P\hbar k_z \\ -\sqrt{\frac{2}{3}}P\hbar k_z & -E_0+\frac{\hbar^2k^2}{2m_0} & 0 & 0 \\ 0 & 0 & -E_0+\frac{\hbar^2k^2}{2m_0} & 0 \\ \sqrt{\frac{1}{3}}P\hbar k_z & 0 & 0 & -E_0-\Delta+\frac{\hbar^2k^2}{2m_0} \end{bmatrix}, \tag{59}$$

and where for bulk case one can assume $k_z = k$. The eigenvalues (doubly degenerate) are obtained by finding the roots of the determinant equation

$$\left|H - E(\mathbf{k})\mathbf{I}\right| = 0 . \tag{60}$$

Denoting $\lambda(\mathbf{k}) = E(\mathbf{k}) - \dfrac{\hbar^2 k^2}{2m_0}$, the eigenvalues, i.e. the roots of Eq. () are

$$\begin{aligned} &\lambda(\mathbf{k}) = -E_0 \\ &\lambda(\mathbf{k})\left[\lambda(\mathbf{k}) + E_0\right]\left[\lambda(\mathbf{k}) + E_0 + \Delta\right] = \hbar^2 k^2 P^2 \left[\lambda(\mathbf{k}) + E_0 + \frac{2\Delta}{3}\right] . \end{aligned} \tag{61}$$

This last equation corresponds to the original formulation of Kane [27,28]. In his derivation he uses a different basis set corresponding to eigenfunctions of operators L^2, L_z, S^2, S_z, but the eigenvalues and dispersion relations are equal since the basis sets are related by a unitary transformation. From the first equation, one obtains the dispersion for the *HH*-band:

$$E_{hh} = -E_0 - \frac{\hbar^2 k^2}{2m_{hh}}; \quad \frac{1}{m_{hh}} = \frac{1}{m_0} \tag{62}$$

Note that the effective hole mass is still equal to the bare electron mass. From the second equation we obtain the other three dispersion relations as follows. We assume that the coefficient $\hbar^2 k^2 P^2$ is small. We then obtain the lowest order solution by setting this term equal to zero and obtain (of course) the original band edge positions

$$\lambda^0_{\Gamma_6} = 0, \;\; \lambda^0_{\Gamma_8 lh} = -E_0, \quad \lambda^0_{\Gamma_7} = -E_0 - \Delta \; . \tag{63}$$

The zero of the energy scale is taken to be at the conduction band edge. Now, the first order solution is obtained by each band by inserting the zero order solution on the RHS of Eq. () and also in the left hand side in all other terms except for one becoming zero if the

substitution is made. The first order eigenvalue is then obtained analytically. For example, for the conduction band one obtains

$$\begin{aligned}
&\lambda^1_{\Gamma_6}(\mathbf{k})\left[\lambda^0_{\Gamma_6}(\mathbf{k})+E_0\right]\left[\lambda^0_{\Gamma_6}(\mathbf{k})+E_0+\Delta\right]=\hbar^2k^2P^2\left[\lambda^0_{\Gamma_6}(\mathbf{k})+E_0+\frac{2\Delta}{3}\right]\Leftrightarrow\\
&\lambda^1_{\Gamma_6}(\mathbf{k})\left[E_0\right]\left[E_0+\Delta\right]=\hbar^2k^2P^2\left[E_0+\frac{2\Delta}{3}\right]\Rightarrow\\
&\lambda^1_{\Gamma_6}(\mathbf{k})=\hbar^2k^2P^2\frac{\left[E_0+\frac{2\Delta}{3}\right]}{\left[E_0\right]\left[E_0+\Delta\right]}
\end{aligned} \tag{64}$$

i.e.

$$E_{\Gamma_6}(\mathbf{k})=\hbar^2k^2P^2\frac{\left[E_0+\frac{2\Delta}{3}\right]}{\left[E_0\right]\left[E_0+\Delta\right]}+\frac{\hbar^2k^2}{2m_0}=\frac{\hbar^2k^2}{2}\left(\frac{1}{m_0}+\frac{4P^2}{3E_0}+\frac{2P^2}{3\left(E_0+\Delta\right)}\right), \tag{65}$$

which means that the effective mass of the electrons in the vicinity of the conduction band edge is

$$\frac{1}{m_c}=\frac{1}{m_0}+\frac{4P^2}{3E_0}+\frac{2P^2}{3\left(E_0+\Delta\right)}\ . \tag{66}$$

For the light hole and the split-off bands, one obtains by similar procedure

$$\begin{aligned}
&E_{lh}=-E-\frac{\hbar^2k^2}{2m_{lh}};\quad \frac{1}{m_{lh}}=\frac{1}{m_0}-\frac{4P^2}{3E_0}\\
&E_{so}=-E-\Delta-\frac{\hbar^2k^2}{2m_{lh}};\quad \frac{1}{m_{so}}=\frac{1}{m_0}-\frac{2P^2}{3\left(E_0+\Delta\right)}
\end{aligned} \tag{67}$$

Note that because of the relative magnitudes of the matrix elements of the dipole operator, band edge energies and spin-orbit splitting conduction band obtains a positive effective mass, whereas the light-hole and split-off bands have negative effective electron mass. Kane [27,28] used this method to describe the energy band-structure in a *p*-type germanium and silicon, and indium antimonide.

3.4. Coupling With Distant Bands

To describe the coupling with distant bands, we consider the wave equation

$$(H_0 + W)\psi = E\psi \,, \tag{68}$$

where

$$W = \frac{\hbar}{m_0}\mathbf{k}\cdot\mathbf{p} + \frac{\hbar^2 k^2}{2m_0} \,. \tag{69}$$

We assume that the eigenvalues E_l corresponding to eigenstates $|l\rangle$ $l = 1, 2, ..., 8$ of the Hamiltonian H_0 are close to each others on the energy scale. These are the 8 bands considered in Kane's model. These eigenstates are strongly coupled by the operator *W*. We assume that there is another set of eigenstates $|\nu\rangle$ of H_0 only weakly coupled by *W*. We now calculate the correction to the 8 lowest eigenvalues of H_0 caused by the distant bands. Let ψ be solution to Eq. () including this correction, i.e.

$$\psi = \sum_l c_l |l\rangle + \sum_\nu c_\nu |\nu\rangle \,. \tag{70}$$

Inserting Eq. (70) into Eq. (69), we obtain

$$\begin{aligned} &\sum_l c_l \left[(E_l - E)\delta_{lm} + \langle m|W|l\rangle \right] + \sum_\nu c_\nu \langle m|W|\nu\rangle = 0 \\ &\sum_\nu c_\nu \left[(E_\nu - E)\delta_{\mu\nu} + \langle \mu|W|\nu\rangle \right] + \sum_l c_l \langle \mu|W|l\rangle = 0 \end{aligned} \tag{71}$$

Since the coupling to the distant bands $|\nu\rangle$ is weak, one can conclude that if ψ is one of the lowest 8 eigenvalues, the relative magnitudes of the expansion coefficients are: $|c_l| \cong 1$ and $|c_\nu| << 1$. The second of the above equations then gives

$$c_\nu \cong \frac{1}{E - E_\nu} \sum_l c_l \left\langle \nu \middle| W \middle| l \right\rangle \ . \tag{72}$$

Inserting this into the first of Eqs. (71), we obtain

$$\sum_l c_l \left[\left(E_l - E \right) \delta_{lm} + \left\langle m \middle| W \middle| l \right\rangle + \left\langle m \middle| W \sum_\nu \frac{\left| \nu \right\rangle \left\langle \nu \right|}{E - E_\nu} W \middle| l \right\rangle \right] = 0 \ . \tag{73}$$

It is obvious that the influence of the distant bands can be taken into account by replacement

$$W \to \tilde{W} = W + W \sum_\nu \frac{\left| \nu \right\rangle \left\langle \nu \right|}{E - E_\nu} W \ . \tag{74}$$

It can be shown that the Hamiltonian of distant band interaction $\tilde{W} - W$ is given by Table 5, where the following notation has been used

$$\begin{aligned} F(\mathbf{k}) &= Ak^2 + \frac{B}{2}\left(k^2 - 3k_z^2 \right) \\ G(\mathbf{k}) &= Ak^2 - \frac{B}{2}\left(k^2 - 3k_z^2 \right) \\ H(\mathbf{k}) &= -iDk_z \left(k_x - ik_y \right) \\ I(\mathbf{k}) &= \frac{\sqrt{3}}{2} B(k_x^2 - k_y^2) - iDk_x k_y \\ A &= \frac{L + 2M}{3}, \ \ B = \frac{L - M}{3}, \ \ C^2 = D^2 - 3B^2, \ \ D = \frac{N}{\sqrt{3}} \end{aligned} \quad , \tag{75}$$

where

$$
\begin{aligned}
L &= \frac{\hbar^2}{2m_0} + \frac{\hbar^2}{m_0^2}\sum_{\nu}\frac{\left|\left\langle X\left|p_x\right|\nu\right\rangle\right|^2}{E-E_\nu} \\
M &= \frac{\hbar^2}{2m_0} + \frac{\hbar^2}{m_0^2}\sum_{\nu}\frac{\left|\left\langle X\left|p_y\right|\nu\right\rangle\right|^2}{E-E_\nu} \\
N &= \frac{\hbar^2}{m_0^2}\sum_{\nu}\frac{\left\langle X\left|p_x\right|\nu\right\rangle\left\langle\nu\left|p_y\right|Y\right\rangle + \left\langle X\left|p_y\right|\nu\right\rangle\left\langle\nu\left|p_x\right|Y\right\rangle}{E-E_\nu} \\
\frac{1}{m_c^{'}} &= \frac{2}{m_0^2}\sum_{\nu}\frac{\left|\left\langle X\left|p_x\right|\nu\right\rangle\right|^2}{E-E_\nu} + \frac{1}{m_0}
\end{aligned}
\qquad (76)
$$

Table 5. Hamiltonian of the distant band interaction.

$$
\tilde{W}-W = \begin{bmatrix}
 & \left|iS\uparrow\right\rangle & \left|\frac{3}{2},\frac{1}{2}\right\rangle & \left|\frac{3}{2},\frac{3}{2}\right\rangle & \left|\frac{1}{2},\frac{1}{2}\right\rangle & \left|iS\downarrow\right\rangle & \left|\frac{3}{2},-\frac{1}{2}\right\rangle & \left|\frac{3}{2},-\frac{3}{2}\right\rangle & \left|\frac{1}{2},-\frac{1}{2}\right\rangle \\
\left|iS\uparrow\right\rangle & \frac{\hbar^2k^2}{2}\left(\frac{1}{m_c}-\frac{1}{m_0}\right) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\left|\frac{3}{2},\frac{1}{2}\right\rangle & 0 & G-\frac{\hbar^2k^2}{2m_0} & iH^* & -\frac{1}{\sqrt{2}}(G-F) & 0 & 0 & I & i\sqrt{\frac{3}{2}}H \\
\left|\frac{3}{2},\frac{3}{2}\right\rangle & 0 & -iH & F-\frac{\hbar^2k^2}{2m_0} & \frac{i}{\sqrt{2}}H & 0 & -I & 0 & -\sqrt{2}I \\
\left|\frac{1}{2},\frac{1}{2}\right\rangle & 0 & -\frac{1}{\sqrt{2}}(G-F) & -\frac{i}{\sqrt{2}}H & \frac{1}{2}(F+G)-\frac{\hbar^2k^2}{2m_0} & 0 & -i\sqrt{\frac{3}{2}}H & \sqrt{2}I & 0 \\
\left|iS\downarrow\right\rangle & 0 & 0 & 0 & 0 & \frac{\hbar^2k^2}{2}\left(\frac{1}{m_c}-\frac{1}{m_0}\right) & 0 & 0 & 0 \\
\left|\frac{3}{2},-\frac{1}{2}\right\rangle & 0 & 0 & -I^* & i\sqrt{\frac{3}{2}}H^* & 0 & G-\frac{\hbar^2k^2}{2m_0} & -iH & -\frac{1}{\sqrt{2}}(G-F) \\
\left|\frac{3}{2},-\frac{3}{2}\right\rangle & 0 & I^* & 0 & \sqrt{2}I^* & 0 & iH^* & F-\frac{\hbar^2k^2}{2m_0} & -\frac{i}{\sqrt{2}}H^* \\
\left|\frac{1}{2},-\frac{1}{2}\right\rangle & 0 & -i\sqrt{\frac{3}{2}H^*} & -\sqrt{2}I^* & 0 & 0 & -\frac{1}{\sqrt{2}}(G-F) & \frac{i}{\sqrt{2}}H & \frac{1}{2}(F+G)-\frac{\hbar^2k^2}{2m_0}
\end{bmatrix}
$$

3.5 The Luttinger-Kohn Hamiltonian

In the Luttinger-Kohn approximation [29] it is assumed that E_0 and Δ are large enough, so that coupling of the Γ_8 bands with the split-off band is weak. This allows one to derive 4×4 Hamiltonian submatrix. The derivation consists of unitary transformation from the basis set $\left|X\right\rangle,\left|Y\right\rangle,\left|Z\right\rangle$ multiplied by the spin functions $\left|\uparrow\downarrow\right\rangle$ parts to the *jj* –

coupled subspace $\left|3/2,3/2\right\rangle,\left|3/2,-3/2\right\rangle$ (*HH*-states) and $\left|3/2,1/2\right\rangle,\left|3/2,-1/2\right\rangle$ (*LH*-states). The Hamiltonian is

$$H=\begin{array}{c} \\ \left|\frac{3}{2},\frac{3}{2}\right\rangle \\ \left|\frac{3}{2},\frac{1}{2}\right\rangle \\ \left|\frac{3}{2},-\frac{1}{2}\right\rangle \\ \left|\frac{3}{2},-\frac{3}{2}\right\rangle \end{array}\begin{array}{cccc} \left|\frac{3}{2},\frac{3}{2}\right\rangle & \left|\frac{3}{2},\frac{1}{2}\right\rangle & \left|\frac{3}{2},-\frac{1}{2}\right\rangle & \left|\frac{3}{2},-\frac{3}{2}\right\rangle \\ F-E_0 & H & -I & 0 \\ H^* & G-E_0 & 0 & I \\ -I^* & 0 & G-E_0 & H \\ 0 & I^* & H^* & F-E_0 \end{array} . \tag{77}$$

The eigenvalues of the above Hamiltonian are obtained from determinant equation $\left|\mathbf{H}-E\mathbf{I}\right|=0$, which gives

$$\begin{aligned} E_{\Gamma_8}(k) &= -E_0+\frac{1}{2}\left(F+G\right)\pm\sqrt{\left(\frac{F-G}{2}\right)^2+\left|I\right|^2+\left|H\right|^2} \\ E_{\Gamma_8}(k) &= -E_0+Ak^2\pm\sqrt{B^2k^4+C^2\left(k_x^2k_y^2+k_y^2k_z^2+k_z^2k_x^2\right)} \end{aligned} \tag{78}$$

If one assumes that $\mathbf{k}\parallel z$-axis, it arrives at

$$E_{\Gamma_8}(k)=-E_0+\left(A\pm B\right)k^2 \ . \tag{79}$$

Inserting matrix elements A and B the *HH-LH* dispersion is given by

$$E=-E_0-\frac{\hbar^2k^2}{2m^*};\quad \frac{1}{m^*}=-\frac{1}{m_0}+\frac{2}{m_0^2}\sum_{m\neq\Gamma_8}\frac{\left\langle\frac{3}{2},\pm\frac{3}{2}\middle|p_z\middle|u_{m0}\right\rangle}{E_m+E_0} \ . \tag{80}$$

The last term is a correction coming from the coupling with the distant bands. This gives the *HH*-band a negative electron effective mass (and a positive hole effective mass).

The Hamiltonian given in Eq. (77) can be written in a more familiar form in terms of Luttinger parameters. First, the Luttinger parameters $\gamma_1, \gamma_2, \gamma_3$ are written in terms of L, M, N, R, S, and T. The matrix elements of the *HH-LH* Hamiltonian are then given by

$$\begin{aligned} F(\mathbf{k}) &= -\frac{\hbar^2}{2m_0}\left[(\gamma_1+\gamma_2)\left(k_x^2+k_y^2\right)+(\gamma_1-2\gamma_2)k_z^2\right] \\ G(\mathbf{k}) &= -\frac{\hbar^2}{2m_0}\left[(\gamma_1-\gamma_2)\left(k_x^2+k_y^2\right)+(\gamma_1+2\gamma_2)k_z^2\right] \\ H(\mathbf{k}) &= \frac{\hbar^2}{2m_0}\sqrt{3}\gamma_3 k_z\left(k_x-ik_y\right) \\ I(\mathbf{k}) &= -\frac{\hbar^2}{2m_0}\sqrt{3}\left[\gamma_2\left(k_x^2-k_y^2\right)-2i\gamma_3 k_x k_y\right] \end{aligned} \tag{81}$$

4. Applications of k•p to Quasi-2D Electron and Hole Systems

4.1 Heterostructure Devices

Development of Molecular Beam Epitaxy [30] has been pushed by device technology to achieve structures with atomic layer dimensions and this has led to an entirely new area of condensed matter physics and investigation of structures exhibiting strong quantum size effects. MBE has played a key role in the discovery of phenomena like the two dimensional electron and hole gases [31], quantum Hall effect [32], and new structures like quantum wires [33], quantum dots [34], etc. The continued miniaturization of solid state devices is leading to the point where quantization-induced phenomena become more and more important. These phenomena have shown that the role of material purity, native defects and interface quality are very critical to the device performance. Modulation doping is employed to achieve adequate carrier densities in one region of the

device which is physically separated from the source of the carriers, the ionized impurities.

The low temperature mobility of modulation doped GaAs/AlGaAs structures is a good measure of the GaAs/AlGaAs material quality. This depends very strongly on the epitaxial structure, particularly the placement and quantity of dopant impurities. The two dimensional electron gas (2DEG) that exists at the interface between GaAs and the wider band gap AlGaAs exhibits a very high mobility at low temperatures. Even at room temperatures, the mobility is larger than that of bulk GaAs. Two factors contribute to this higher mobility, both arising from the selective doping of AlGaAs buffer layers rather than the GaAs layers in which the carriers reside. The first is the natural separation between the donor atoms in the AlGaAs and the electrons in the GaAs. The second is the inclusion of an undoped AlGaAs spacer layer in the structure. Such structures are quite complicated but can be easily fabricated using MBE techniques. A typical heterostructure begins with the bulk GaAs wafer upon which a GaAs buffer layer or superlattice is grown. The latter is used to act as a barrier to the out-diffusion of impurities and defects from the substrate. It also consists of a GaAs cap layer and alternating layers of AlGaAs and GaAs. The common practice is to use a doping for the AlGaAs layers in the active region but nowadays undoped AlGaAs layers are used and a delta doped layer is included. This delta doped layer along with the growth of superlattices restricts the formation of defects, known as D–X centers, to a minimum. There are two important AlGaAs layers on either side of the δ–doped layer and they are called buffer and spacer layer, respectively. The spacer layer is closer to the GaAs quantum well and is of high purity to prevent scattering of the channel carriers by the ionized impurities. A usual

practice is to use undoped AlGaAs layers to have very good confinement of the charge carriers in the well. Other device parameters that have to be considered are the composition of the Aluminum in AlGaAs. There is a compromise in the value chosen for *x*: if *x* is smaller than 0.2 then the band discontinuity will be too small to properly confine carriers in the well; if *x* is too large then defects, termed as D-X centers, tend to appear in AlGaAs. To overcome this problem Aluminum content is limited to about 20% and other variations like δ–doping layer and growth of superlattices are introduced into the MBE techniques.

A prototypical GaAs/AlGaAs heterostructure used, for example, for quantum wires and dots formation that utilize the split-gate technique, is shown in Figure 7. The zero-temperature conduction band profile along the growth direction is shown on the right panel of Figure 7. One can see that the electrons are localized in the quantum-well region, and the presence of this quantization can, in principle, lead to the modification of the electron and hole effective masses.

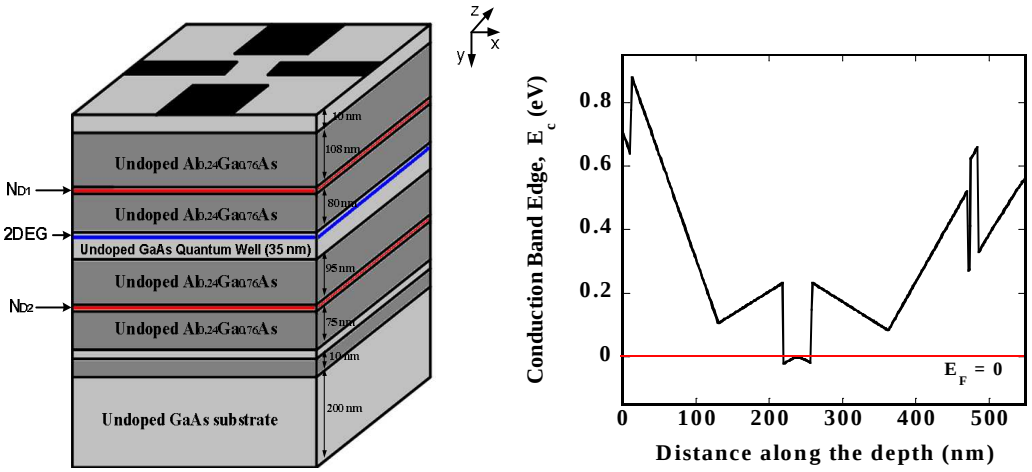


Figure 7. Heterostructure used in the fabrication of the spin filter realized with quantum point contacts at Arizona State University [35,36].

4.1 A Study of Spin-Resolved Band Structure in GaAs Quantum Wells as Applied to Spintronics

Today, the research in *spintronics* – the field of semiconductor electronics based upon exploiting the quantum mechanical property of electron spin as an information carrying entity – owes its existence, in general, to two primary factors. The first is theoretical: the spin component of the electron's wave function can retain its form (i.e., it's orientation or, if referring to an ensemble of electrons, the coherence) in semiconductor transport for much longer times than can the spatial components. In fact, the spin relaxation time τ_s can be on the order of nanoseconds, as opposed to femtoseconds for the spatial momentum relaxation time [37,38]. The second factor is technological: Improvements in lithography and the continuing progress in developing efficient spin filter injection/detection mechanisms that polarize (or detect polarized) electrons in the 2DEG gas of semiconductor heterostructures for spintronic devices by various means, other than by applying a cumbersome external magnetic field to lift the spin degeneracy [39,40], have prompted the race to develop novel devices such as the spin transistor (SPINFET). The starting point for understanding spin injection/detection, the spin-flip scattering mechanisms and their relative impact on transport is, of course, the determination of an accurate spin-resolved band structure.

This section is organized as follows. We first present an application of the multi-band **k·p** to calculations of a band-structure in quantum well, in the absence of stress, strain, or magnetic fields. The effect of bulk inversion asymmetry, a source of spin-splitting in the bands, is then introduced in the model. Furthermore, the **k·p** model, Ddiscussed in section ?? and developed within our group is enhanced to account for spin

splitting in the conduction band due to any structural inversion asymmetry (SIA) that may exist. The model, incorporating SIA effects, is then compared to a similar model, recently published in the literature. Current work includes investigation of the energy level structure in the spin filter that will be realized at Arizona State University using the heterostructure shown in Figure 7.

General Characteristics of the Solver

The Fermi energy of the electrons involved in the transport through our QPC is on the order of 10 meV. This energy is low enough such that a band structure calculation accurate only around a point of high symmetry in the Brillouin zone (Γ point) is sufficient. For this situation, the **k·p** method of calculating band structure, explained in details in Section ??, is applicable.

Eppenga *et al*. [41], using the Kane basis set and including spin-orbit coupling and remote band effects via Löwdin perturbation theory [42], arrive at the following Hamiltonian

$$\mathrm{H} = \begin{bmatrix} E_{EL} & -\sqrt{2}P_z & P_z & \sqrt{3}P_+ & 0 & -P_- & -\sqrt{2}P_- & 0 \\ & E_{LH} & G_1 & \sqrt{2}G_+ & -P^*_+ & 0 & -\sqrt{3}G_- & G_2 \\ & & E_{SO} & -G_+ & -\sqrt{2}P^*_+ & \sqrt{3}G_- & 0 & \sqrt{2}G_2 \\ & & & E_{HH} & 0 & -G_2 & -\sqrt{2}G_2 & 0 \\ & & & & E_{EL} & \sqrt{2}P_z & -P_z & -\sqrt{3}P_- \\ & & & & & E_{LH} & G_1 & \sqrt{2}G_- \\ & & & & & & E_{SO} & -G_- \\ & & & & & & & E_{HH} \end{bmatrix}. \qquad (82)$$

The “empty spaces” in Eq. (82) denote Hermitian conjugates. The diagonal matrix elements are

$$\begin{aligned} E_{EL} &= E_g + s\tilde{e} \\ E_{LH} &= -\gamma_1\tilde{e} + \gamma_2\tilde{e}_1 \\ E_{SO} &= -\Delta - \gamma_1\tilde{e} \\ E_{HH} &= -\gamma_1\tilde{e} + \gamma_2\tilde{e}_1 \end{aligned}, \tag{83}$$

where E_g is the bandgap and Δ is the spin-orbit split-off energy. The γ_1, γ_2 and s parameters are effective mass parameters that modify the free-electron term. Also,

$$\tilde{e} = \frac{\hbar^2}{2m}\left(k_x^2 + k_y^2 + \tilde{k}_z^2\right), \quad \tilde{e}_1 = \frac{\hbar^2}{2m}\left(2\tilde{k}_z^2 - k_x^2 - k_y^2\right), \quad e_2 = \frac{\hbar^2}{2m}\left(k_x^2 - k_y^2\right). \tag{84}$$

The off-diagonal terms are given as,

$$\begin{aligned} P_z &= \sqrt{1/3}\left(i\rho\tilde{k}_z + \beta k_x k_y\right) \\ P_{\pm} &= \sqrt{1/6}\left[i\rho\left(k_x \pm ik_y\right) + \beta\tilde{k}_z\left(k_y \pm ik_x\right)\right] \\ G_1 &= \sqrt{2}\gamma_2\tilde{e}_1 \\ G_2 &= -\sqrt{3}\gamma_2 e_2 + i2\sqrt{3}\gamma_3 k_x k_y \\ G_{\pm} &= \sqrt{6}\gamma_3\tilde{k}_z\left(k_x \pm ik_y\right) \end{aligned}. \tag{85}$$

In these terms, $\rho = -\frac{\hbar^2}{m}\int_{unit\ cell} dr\ \phi_s^* \frac{\partial}{\partial z}\phi_z$ and accounts for the coupling of the conduction band s-states of the Kane basis with the valence band z-state. Terms containing γ_3 result in an anisotropic band structure near the Γ point if $\gamma_2 \neq \gamma_3$. The parameter β is due to the bulk inversion asymmetry (BIA) and causes spin-splitting of the bands.

Also known as Dresselhaus splitting [10], BIA induced spin-splitting occurs in zinc-blende semiconductors due to the fact that two different kinds of atoms (e.g., Ga and As, Ga and Sb, etc...) exist, resulting in asymmetrical wave functions about an axis of symmetry (e.g., (100) axis). This means that while Kramer's theorem, $E(k\uparrow) = E(-k\downarrow)$

is satisfied for all values of k , the situation away from $k = 0$ is that $E(k\uparrow) \neq E(k\downarrow)$. In fact, spin-splitting of the conduction band (Γ_6 band) due to BIA is proportional to k^3 for small values of k in bulk zinc-blende semiconductors. However, in 2DEG systems such as heterostructures, a linear dependence on k occurs too. The valence bands exhibit linear BIA splitting in both the bulk and 2DEG cases. Thorough discussions of the k-dependence of BIA spin-splitting can be found in Zawadzki and Pfeffer [43] and Silsbee [44].

Finally, note that the matrix operator of Eq. (82) does not include the effects of stress and strain or the influence of a magnetic field. Of course, these effects undoubtedly will modify the k-dependence on spin-splitting. However, it must be mentioned that the Kane model inclusive of these effects has been derived by several researchers, notably Trebin et al. [45].

Quantum Well Structure

To test the applicability of the Kane approach, Eq. (82) has been applied to the *symmetrical* quantum well shown in Figure 8.

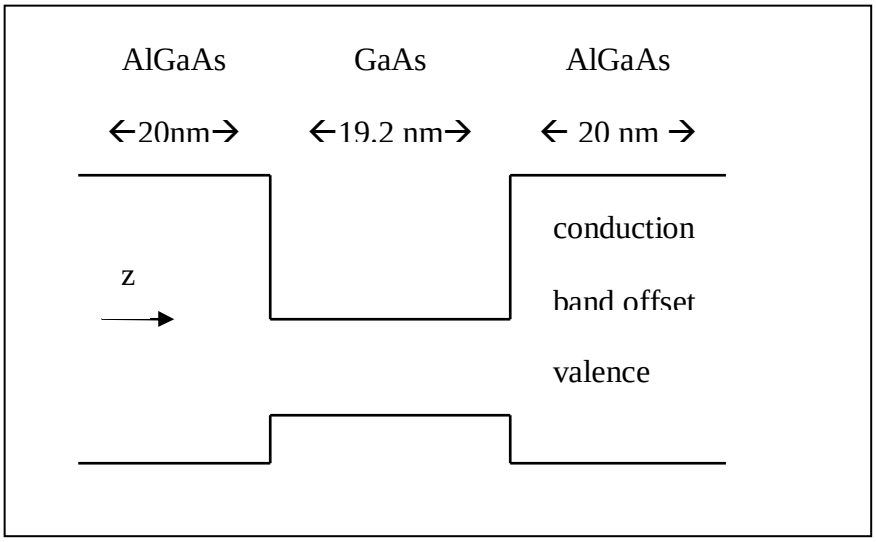


Figure 8. Symmetrical Quantum Well. Though not shown here, the valence split off band also exhibits a band offset. The dimensions of the well are chosen so as to be able to compare to the tight-binding model results of Chang and Schulman [46].

To start, one applies the operator $\hat{p} \to -i\hbar\nabla$ to all k_z terms (denoted by the ~) in the eigenvalue equation with matrix operator given in Eq. (82). This operation is done to the k_z terms only, since quantization is assumed to be along the z-direction. In doing this, it is convenient to note that the resultant matrix with the matrix operator given explicitly included then takes the form,

$$\begin{pmatrix} F_{11}\nabla_z^2 + B_{11}\nabla_z + C_{11} + D_{11} & \cdots & F_{18}\nabla_z^2 + B_{18}\nabla_z + C_{18} + D_{18} \\ \vdots & \ddots & \vdots \\ F_{81}\nabla_z^2 + B_{81}\nabla_z + C_{81} + D_{81} & \cdots & F_{88}\nabla_z^2 + B_{88}\nabla_z + C_{88} + D_{88} \end{pmatrix} \begin{pmatrix} \chi_{1,z} \\ \chi_{2,z} \\ \vdots \\ \chi_{7,z} \\ \chi_{8,z} \end{pmatrix} = \varepsilon \begin{pmatrix} \chi_{1,z} \\ \chi_{2,z} \\ \vdots \\ \chi_{7,z} \\ \chi_{8,z} \end{pmatrix}, \quad (86)$$

where the *F* and *B* terms denote coefficients to 2nd and 1st order partial derivatives w.r.t. z, respectively. The *C* terms denote potential offsets, and the *D* terms indicate all other terms not operated on by the momentum operator. Keep in mind that all terms *B*, *C*, *D*, and *F* are actually *z*–dependent functions, though not denoted as such to ease the notation.

To ensure Hermiticity of the resulting matrix at the heterostructure interfaces, I use following discretization scheme as noted in Eppenga et al. [17]:

$$B(z)\frac{\partial}{\partial z} \rightarrow \frac{1}{2}\left(B(z)\frac{\partial}{\partial z} \; + \; \frac{\partial}{\partial z}B(z) \right) , \tag{13}$$

$$F(z)\frac{\partial^2}{\partial z^2} \rightarrow \frac{\partial}{\partial z}F(z)\frac{\partial}{\partial z} \; . \tag{14}$$

Equations (13) and (14) yield

$$\begin{aligned} B(z)\frac{\partial \chi}{\partial z} \; &= \; \frac{1}{4\Delta z}\left[2B(z)\chi_z + 1 - 2B(z)\chi_z - 1 + \left(B(z+1) - B(z-1) \right)\chi_z \right] \\ &\approx \; B(z)\frac{\chi_{z+1} - \chi_{z-1}}{2\Delta z} \end{aligned} \tag{15}$$

and

$$\frac{\partial \chi}{\partial z} F(z) \frac{\partial \chi}{\partial z} = \left(\frac{F(z)}{\Delta z^2} + \frac{F(z+1) - F(z-1)}{4\Delta z^2} \right) \chi_{z+1} + \left(\frac{F(z)}{\Delta z^2} - \frac{F(z+1) - F(z-1)}{4\Delta z^2} \right) \chi_{z-1}$$
$$- \left(\frac{2F(z)}{\Delta z^2} \right) \chi_z \qquad (16)$$
$$\approx \frac{F(z)\chi_{z+1} + F(z)\chi_{z-1} - 2F(z)\chi_z}{\Delta z^2}$$

It was later noticed the approximations of equation (15) or (16) did not degrade results, so I used them to transform the matrix from the bulk to the z-quantized situation. Thus,

$$F(z) \frac{\partial^2 \chi}{\partial z^2} \rightarrow \frac{F(z)\chi_{z+1} + F(z)\chi_{z-1} - 2F(z)\chi_z}{\Delta z^2} \qquad (17)$$

$$B(z) \frac{\partial \chi}{\partial z} \rightarrow B(z) \frac{\chi_{z+1} - \chi_{z-1}}{2\Delta z} \qquad (18)$$

At this point, a word of caution is in order. Applying equations (13) and (14) is known as using “symmetrizing” boundary conditions. Burt [13] has argued that this approach is technically incorrect for abrupt interfaces. However, as observed in the simulations of Meney et al. [23], for the energy ranges of interest (less than 50meV) the error is miniscule, especially when using an 8-band model.

Applying (17) and (18) to (10) results in a matrix of the form shown in Figure 3.

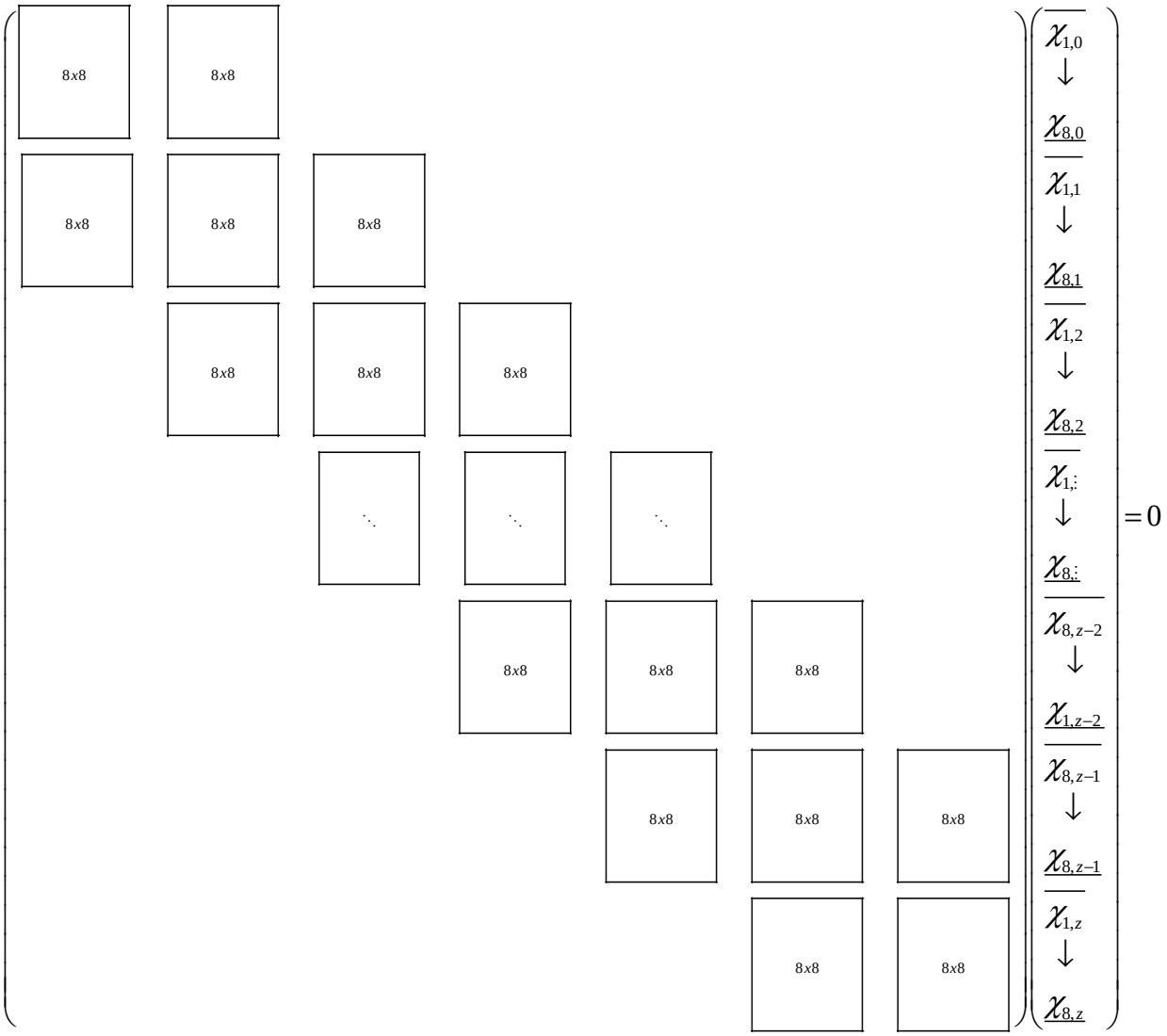


Figure 3: General form of the expanded 8-band Kane Matrix. Each point in real-space, along the quantized z-axis corresponds to an 8-row block in this matrix.

To further clarify the discretization scheme, figure 4 shows what the scheme would be for a simple 2-band matrix evaluated at 3 points in real-space [(z - 1), z, (z + 1)] along the z axis. Of course, all the coefficients (B, C, D, F) are z-dependent.

Also, note that the C terms only appear in the diagonal, as they are the potential offsets inclusive in the $E_n(k_0)$ term of (6).

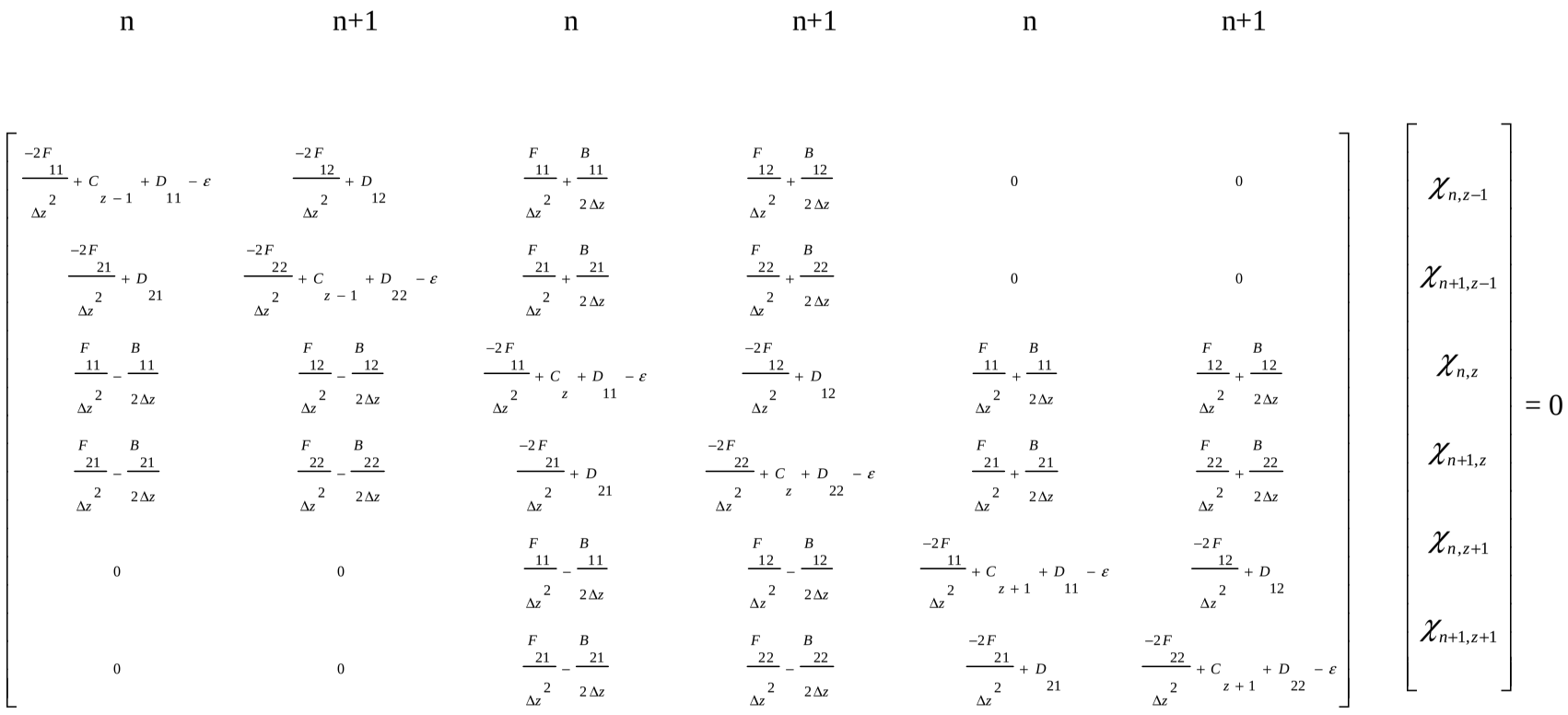


Figure 4. Example of the discretization scheme for a simple 2 x 2 matrix,

showing the distribution of the B, C, D, F coefficients given in equations (17) and (18).

The matrix denoted in Figure 3 was coded in FORTRAN and evaluated using the LAPACK ZHEEVX eigenvalue/eigenvector solver subroutine. The parameters γ_1, γ_2, γ_3, s and the parameterized value of ρ of (11a,c) were chosen in accordance with Table II of Eppenga et al. [17], as shown below in Figure 5. The parameter β was chosen to be ~17 eV-A° in the GaAs well and ~12 eV-A° in the AlGaAs cladding layers, in accordance with another Eppenga and Schuurmans paper [24]. Additionally, a mesh spacing Δ_z of 1nm was used.

	$(2m/\hbar^2)P^2$ (eV)	s	γ_1	γ_2	γ_3
GaAs	28.196	−3.519	1.673	−0.266	0.250
$Ga_{0.75}Al_{0.25}As$	25.959	−2.903	1.789	−0.086	0.385

Figure 5. k.p parameters used in this work (reproduced from Table II of [17]).

The results, depicted in Figure 6a for the valence bands, are in good agreement with the Chang and Schulman tight binding calculation of Figure 6b [22], which requires several more adjustable fitting parameters other than $\gamma_1, \gamma_2, \gamma_3$, s and ρ. The labels of the bands (HH1, LH1, etc...) of Fig. 6b also apply to Fig. 6a. and define the predominant character of the band at $k = 0$. This was proven by comparing the relative amplitudes of the χ_j envelope functions (where j signifies the HH, LH, etc… band index) in equation (9) for the various sub-bands at $k = 0$. In producing Fig. 6a the band offset depicted in figure 2 was chosen to be split 85/15 % over the conduction/ valence bands, to be consistent with [17] and [22]. However, experimental studies have since indicated that a 65/35 % split is more realistic.

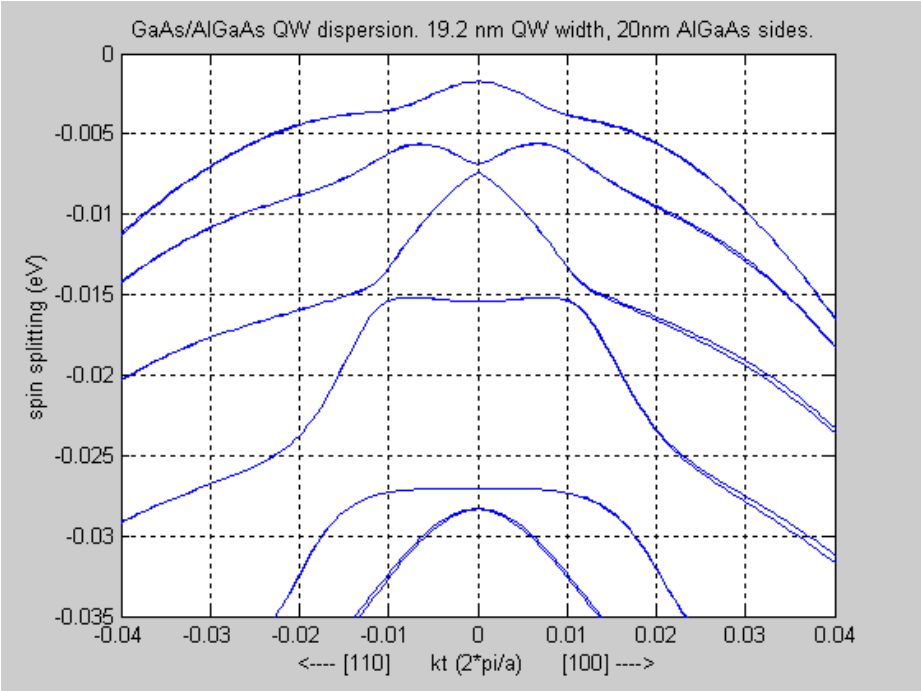


Figure 6a (top): valence band dispersion in 19.2 nm quantum well, as calculated with 8-band k.p method. Figure 6b (bottom): Chang & Schulman tight-binding results. The labels (HH1, HH2, etc…) apply to Figure 6a also, as explained in the text.

Although Figure 6 indicates that the 8-band k.p model is useful for accurately describing band structure in the vicinity of a point of symmetry in the Brillouin zone, the model does of course have severe limitations. Indeed, it was found that in using (11) to calculate the bulk band structure of GaAs, the conduction band suddenly curves down and goes negative at about 10% of the path length along 111 ($\Gamma \rightarrow$ L). Thus, at large real values of *k*, the bandgap can "disappear" and "spurious" solutions to the eigenvalue problem exist [23]. Therefore, at large values of *k* the 8-band k.p approach is not appropriate. Other methods such as tight-binding methods must be used. Higher order k.p models, such as the 24-band k.p model recently reported by Radhia et al. [25], could also possibly be used, though they would be computationally inefficient and computer-memory intensive.

III. Present Research - Inclusion of SIA Effects into the 8-band Model

A. Theory

Since the results of Figure 6 indicate that the 8-band solver developed in section II is accurate for low values of *k* (i.e., the energy range of interest spintronic devices operated at cryogenic temperatures), the solver is currently being used to investigate SIA effects in quantum GaAs/AlGaAs quantum wells, by modifying the Hamiltonian matrix operator of (11) to include SIA terms.

As previously mentioned, SIA is of great interest in spintronics, since the promise of many proposed spintronic devices (refs.[1] - [5]) relies on the principle of being able to modulate the SIA via application of an external electric field. The "Rashba Effect" [26] predicts that an electric field applied perpendicular to the plane of a 2DEG will cause SIA. In turn, this

electric field will then relativistically [15, 26] induce an effective magnetic field in the plane of the 2DEG, effectively lifting the spin-degeneracy of the charge carriers.

Aside from external electric fields, SIA can in principle be caused by anything that results in an asymmetrical quantum well. Effectively, this means that the penetration of the wave function in the cladding layers is not identical to both sides of the well. As reported by Silsbee [20], this can occur during MBE growth. For example, if one attempts to grow a heterostructure formed by a cation and anion of one kind of atom (denoted by C1 and A1 respectively) and a cation and anion of another type of atom (C2 and A2), one would desire the following growth pattern:

…A1-C1-A1-C1-A1-C1-A1=C2-A2-C2-A2-C2-A2-C2=A1-C1-A1-C1-A1-C1… ,

where the '=' sign indicates the heterostructure interface. Typically though MBE growth yields

…A1-C1-A1-C1-A1-C1-A1=C2-A2-C2-A2-C2-A2-C2-A2=A1-C1-A1-C1-A1-C1… .

Note that the 2nd structure is asymmetrical.

The general form of the SIA term to be added to (11) to account for SIA splitting of the conduction band is

$$Hso = \alpha_1 \sigma \cdot (k \times \hat{z}) \tag{19}$$

where σ Pauli spin matrix, as discussed in Section II and in ref. [15].

The term α_1 is a coefficient, typically called the "Rashba Coefficient". In fact, the precise nature of this term has been controversial. Until recently many people believed that this term is proportional to the electric field of the conduction band (i.e., $-\nabla V$), in which case (19) has an obvious similarity and/or connection to eqn. (8) which describes spin-orbit coupling. However, Zawadzki and Pfeffer [19] point out that the average electric field in a bound state of a quantum well is zero. Interestingly, they report that α_1 has a dependence on the *valence band offsets* at the

interfaces. I am presently studying this matter. Nevertheless, the value of α_1 has been empirically estimated for various structures, one of which I am investigating, as will be discussed later in this section.

Furthermore, Winkler [27] has stated that the SIA splitting of the HH and LH valence bands (Γ_8 bands) should take a form similar to (8),

$$H_{8v} = \alpha_2 J \cdot (k \times \hat{z}) + \alpha_3 J' \cdot (k \times \hat{z}) . \qquad (20)$$

Since the basis set considered is the same as the Kane basis of earlier, in (20) J is defined as the angular momentum matrix operator for particles of momentum $j = 3/2$ (ref. [15,28]), and J' is defined as J^3 (according to [28], since J and J^3 are linearly independent any 4 x 4 matrix such as equation (20) can be represented as an expansion of J and J^3) .

Most models of SIA induced spin-splitting include only a 2 x 2 term (i.e., equation (19). However, in this study, both equations (19) and (20) are added to the Hamiltonian matrix operator (11). Also, I have decided to add a matrix operator for the split-off hole Γ_7 bands. In doing this, I assume that it has the same form as equation (19) (using a new coefficient α_4)since the split-off band is also characterized by $J = 1/2$.

All the above SIA terms are represented in the matrix of Figure 7, which are then added to the Hamiltonian matrix operator (11). (The .PDF version of this report color-codes the terms for each coefficient in Figure 7.) . Currently, I am trying to determine if the blank spaces in the matrix should actually correspond to some type of coupling terms.

$\lvert J, J_z\rangle$		$\frac{1}{2},\frac{1}{2}$	$\frac{3}{2},\frac{1}{2}$	$\frac{1}{2},\frac{1}{2}$	$\frac{3}{2},\frac{3}{2}$	$\frac{1}{2},-\frac{1}{2}$	$\frac{3}{2},-\frac{1}{2}$	$\frac{1}{2},-\frac{1}{2}$	$\frac{3}{2},-\frac{3}{2}$
$\lvert J, J_z\rangle$		$CB\uparrow$	$LH\uparrow$	$SO\uparrow$	$HH\uparrow$	$CB\downarrow$	$LH\downarrow$	$SO\downarrow$	$HH\downarrow$
$\frac{1}{2},\frac{1}{2}$	$CB\uparrow$	0				$\alpha_1 k_-$			
$\frac{3}{2},\frac{1}{2}$	$LH\uparrow$		0		$\sqrt{\frac{3}{2}}\alpha_2 k_+ + \frac{7}{8}\sqrt{3}\alpha_3 k_+$		$\alpha_2 k_- + \sqrt{\frac{5}{2}}\alpha_3 k_-$		0
$\frac{1}{2},\frac{1}{2}$	$SO\uparrow$			0				$\alpha_4 k_-$	
$\frac{3}{2},\frac{3}{2}$	$HH\uparrow$		$\sqrt{\frac{3}{2}}\alpha_2 k_- + \frac{7}{8}\sqrt{3}\alpha_3 k_-$		$0+0$		$0+0$		$0+\frac{3}{4}k_+\alpha_3$
$\frac{1}{2},-\frac{1}{2}$	$CB\downarrow$	$\alpha_1 k_+$				0			
$\frac{3}{2},-\frac{1}{2}$	$LH\downarrow$		$\alpha_2 k_+ + \frac{5}{2}k_+\alpha_3$		$0+0$		$0+0$		$\sqrt{\frac{3}{2}}\alpha_2 k_- + \frac{7}{8}\sqrt{3}\alpha_3 k_-$
$\frac{1}{2},-\frac{1}{2}$	$SO\downarrow$			$\alpha_4 k_+$				0	
$\frac{3}{2},-\frac{1}{2}$	$HH\downarrow$		0		$0+\frac{3}{4}k_-\alpha_3$		$\sqrt{\frac{3}{2}}\alpha_2 k_+ + \frac{7}{8}\sqrt{3}\alpha_3 k_+$		$0+0$

Figure 7. Matrix of SIA spin-splitting terms to be added to the Hamiltonian matrix operator (eqn. 11). As discussed in the text, the $\alpha_{2,3}$ α_4 coefficients refer to conduction band, HH/LH band, and split-off band SIA terms, respectively.

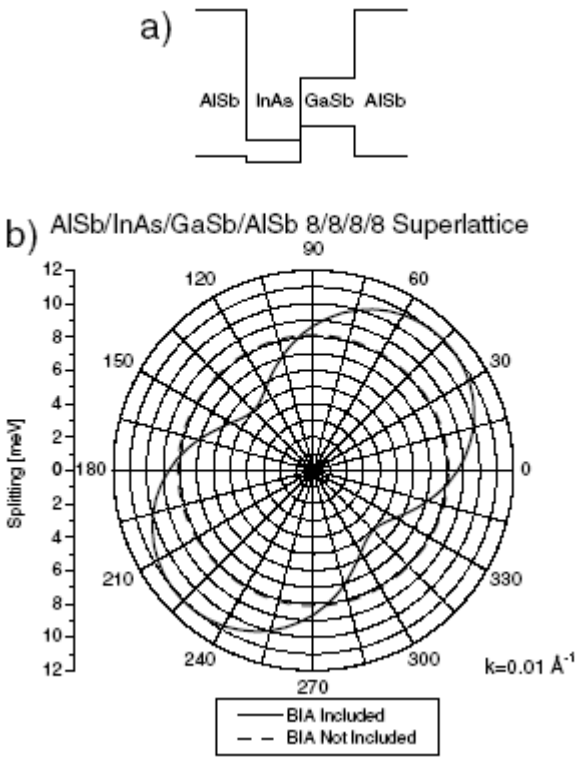
B. Application to an Asymmetric Quantum Well

Using an 8-band solver apparently similar to the one derived in this report, Cartioxa et al. [29] have recently investigated the relative contributions of BIA and SIA terms to conduction band splitting in an asymmetrical quantum well, as depicted in Figure 8a. In this structure each layer is ~ 3.0 nm thick. In Figure 8b they show the BIA and SIA splitting for a particular magnitude of $k_{\square}$ ($k_{\square} = 0.01(2\pi/a)$, a is the lattice spacing) as this vector is rotated through the k_x k_y plane.

To test my SIA modeling, I have attempted to reproduce their results. As indicated in Figure 8C, my results show good agreement with Cartioxa et al when $\alpha_{2,3} = \alpha_4 = 0$ (I have not studied the case when these other coefficients are non-zero.) However, in order to match their results, I had to calculate the modified Luttinger parameters (as in Figure 5), which they did not provide, for AlSb, InAs, and GaSb using data from [9] and [17]. However, the calculated values may not be that accurate. Consequently, the values chosen for the BIA and SIA coefficients (37.5 eV- A°, 100.5 eV- A°) were 2.5 times the values Carioxa et al. quoted.

As expected, Figure 8(b) and 8(c) indicate that the SIA splitting of the conduction band is isotropic (independent of the component of the vector, k_x or k_y). The BIA splitting is greater along 110 ($k_x = k_y$) than 100 ($k_y = 0$) due to the differences in the HH effective mass for these two directions in *k*-space.

In the near future, the *k*-dependence of SIA on the valence bands will be investigated. As noted in Zawadzki and Pfeffer [19], SIA is expected to be linear in *k* for LH bands and proportional to k^3 for HH bands.



C)

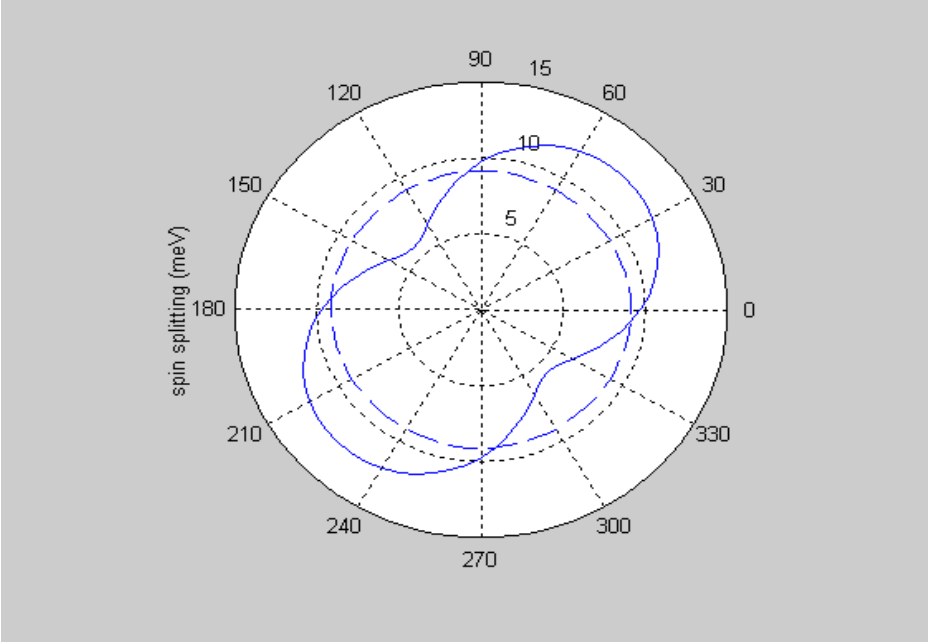


Figure 8: (a) Asymmetrical QW of Cartioxa et al [29]. (b) Cartioxa et al. spin splitting for well of (a). (c) My results, using estimated values of the modified Luttinger parameters.

4.2 k·p for Nano-Scale p-Channel MOSFET Devices

Electron transport in Si inversion layers has been the primary subject of research for many years now, but hole transport has been relegated to the background mainly due to the complex valence band-structure in Si. Hole transport is affected by the warping and anisotropy of the valence bands and the band-structure cannot be approximated with an effective mass picture or with an analytical band model. The advent of alternate device structures [47,48,49] aimed at boosting the speed and density of VLSI circuits however, seems to have revived interest. The important alternate device technologies are buried channel strained SiGe p-channel MOSFETs and surface channel strained Si.

In this section we describe a new way of incorporating band-structure and quantum effects on hole transport in conventional Si p-channel MOSFETs. This is achieved by coupling a 2D Poisson–1D discretized 6×6 **k·p** Hamiltonian solver (discussed in Section ??, for the case when the conduction band contribution is neglected) self-consistently to the Monte Carlo transport kernel (explained in more details in Section ??). At present our method only includes phonon scattering in the full band model. This method is generic and can easily be extended to model strained layer MOSFETs by incorporating an additional strain Hamiltonian into the band-structure kernel.

The band-structure is calculated using the **k·p** method, the Hamiltonian for which is given in Eq. (), and is repeated here for completeness:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{\mathbf{k.p}} & 0 \\ 0 & \mathbf{H}_{\mathbf{k.p}} \end{bmatrix} + \mathbf{H}_{\mathbf{so}} + \mathbf{I}V(z), \tag{87}$$

where $\mathbf{H_{k.p}}$ and $\mathbf{H_{so}}$ are the 6×6 $\mathbf{k}\cdot\mathbf{p}$ and the spin-orbit Hamiltonians respectively, $\mathbf{I}$ is a 6×6 identity matrix, $V(z)$ is the confining potential along the device depth. Replacing the vector k_z with its operator notation as k_z=-i∂/∂z, and using a finite difference discretization, Eq. (82) can be recast into an eigenvalue equation for the eigenenergies in the *xy*-plane, for different values of the in-plane **K**-vector, $\mathbf{K}_{\|}(k_x,k_y)$.The solution of the eigenvalue problem involves the diagonalization of a tridiagonal block matrix whose rank is given by $6\times N_Z$, where N_Z is the number of mesh points along the depth direction. For the 3D (bulk) carriers in the source and drain, we only have the first two terms of Eq. (82). This 6×6 Hamiltonian can easily be diagonalized to give the eigenvalues of 3D carriers at (k_x, k_y, k_z).

To include carrier scattering within the transport kernel, the density of states of the system (2D and 3D) are required. For the 2D case, we tabulate the in-plane **K-** vector, $\mathbf{K}_{\|}$, as a function of carrier energy (ε_{2d}), band (ν) and subband (n) indices, and the in-plane azimuth angle(ϕ). For the 3D case, the K-vector, $\mathbf{K}_{3D}$ is tabulated as a function of carrier energy (ε_{3d}), band index (ν), and the azimuth (ϕ) and elevation (θ) angles. In order to set up the inverse problem, the discretized eigenvalue Eq. (82) for the 2D system can be recast into a eigenvalue equation for $|\mathbf{K}_{\|}|$ [50] as shown by the following equation, where $\mathbf{D_n}$ operates on $|\mathbf{K}_{\|}|^{\mathbf{n}}$.

$$\begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{D}_2^{-1}.[\mathbf{D}_0 - \mathbf{I}E] & -\mathbf{D}_2^{-1}.\mathbf{D}_1 \end{bmatrix} \begin{bmatrix} \psi_K \\ \psi_K^{(1)} \end{bmatrix} = K \begin{bmatrix} \psi_K \\ \psi_K^{(1)} \end{bmatrix} \tag{88}$$

Since $\varepsilon_\nu^n(k_x,k_y)$ is quadratic in $|\mathbf{K}_{\|}|$, the problem involves diagonalizing a matrix whose rank is twice as large as that of the discretized $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian, i.e. $12\times N_Z$. In the 3D

case, using a similar technique, one can show that the problem involves diagonalizing a matrix whose rank is twice that of the **k·p** Hamiltonian, i.e. 12. Thus, for the 3D case, one can tabulate the values initially and these can be used throughout the simulation. The computational complexity for the 2D case led us to make the following simplifying assumptions.

Using a sufficiently high vertical electric field ~5MV/cm, a triangular test-potential was generated and used to tabulate the dispersions and density of states (DOS) of the ground state subbands in each band (heavy-hole HH, light-hole LH and split off-SO). It was then assumed that for the case of a real confining potential in the device, the dispersions in each subband for a particular band would be given by the tabulated (triangular-well) dispersion of the ground state subband of the corresponding band, thus allowing us to capture the basic features of subband anisotropy, warping and nonparabolicity. The only effect of the 'real' confining potential in the device would be the translations of the dispersions on the energy axis by the subband energies at the Γ point.

$$\varepsilon_\nu^n(k_x,k_y) \approx [\varepsilon_\nu^0(k_x,k_y) - \varepsilon_\nu^0(0,0)] + \varepsilon_\nu^n(0,0) \ . \tag{89}$$

For the inverse problem, a similar approach is used. The triangular test potential is used in the inverse solver, in order to tabulate the in-plane **K**-vectors $\mathbf{K}_\parallel^{n,\nu}(\varepsilon_{2d},\phi)$ for a set of chosen (ε_{2D},ϕ). Having so tabulated in-plane **K**-vectors for the lowest subband in each band, we assume that the same dispersion holds also when employing the actual device potential for all the subbands of the given band, i.e.

$$K_{\parallel}^{n,\nu}(\varepsilon_{2D},\phi) \approx K_{\parallel}^{0,\nu}(\varepsilon_{2D},\phi) \ . \qquad (90)$$

The Monte Carlo particle based simulator handles the transport of holes through the device and is described in much more details in Section ??. Having calculated the hole band-structure in the contacts and the active device region *i.e.* under the gate, the quantum mechanical hole density in the channel, is calculated self consistently with the Poisson equation and the 2D band-structure code. Holes are then initialized in real space based on the local carrier density and their energy is initialized by assuming a thermal distribution. The present version of our simulator accounts only for phonon scattering within the isotropic approximation, but uses overlap factors calculated from the actual eigenfunctions in each subband. As the carriers drift under the influence of the electric field due to the applied bias, the confining potential changes and this in turn changes the eigenenergies and the eigenfunctions. As a result, the scattering rates must be updated frequently during the simulation. Within the scope of the current model we have assumed the holes to be quasi 3D particles in the source and drain regions and have used appropriate models to treat these boundary conditions effectively**.** When converting a bulk (3D) Monte Carlo particle into a low-dimensionality (2D) particle occupying a subband in the inversion layer, the difference between the carrier energy ε_{3D} and the in-plane kinetic energy ε_{2D} gives the subband energy ε_v^n. The carrier subband is then determined by choosing a subband with the minimum error in subband energy and ε_v^n, the calculated energy. In the opposite case of converting a 2D-particle into a bulk carrier, the 3D carrier energy is given by $\varepsilon_{3D}=\varepsilon_{2D}+\varepsilon_v^n$. By scanning the elevation angle θ from the tabulated values of the 3D **K**-vector and preserving the in plane azimuth ϕ, the $\mathbf{K_{3D}}$ vector

which minimizes the error in the magnitude of the in-plane $\mathbf{K_{2D}}$ vector is chosen as the 3D carrier momentum of the bulk particle.

The isosurfaces of the lowest Heavy, Light and Split off subbands, for the case of the triangular test potential are shown in Figure 9. Note the strong warping of the heavy hole band when compared with the fairly regular shapes for the light hole and the split off bands which makes it extremely difficult for analytical band models to describe the valence band-structure accurately. The hole-density of states is determined by performing a surface integral over these isosurfaces and these are then used to determine the carrier scattering rates in the channel.

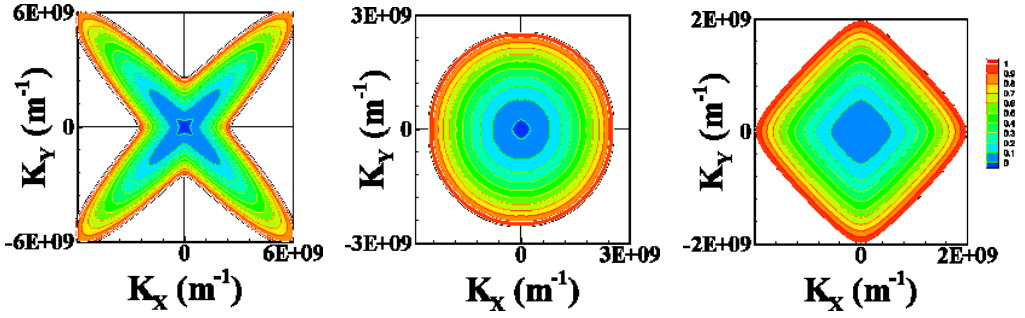


Figure 9. Isosurfaces of the lowest lying HH, LH and SO subbands on a (001) oriented substrate.

The density of states for the confined carriers is shown in Figure 10. The deviation of the density of states obtained by a full band calculation from a regular step-like profile expected out of an effective-mass type approximation is clearly seen in the case of the light hole and split off bands, whereas the heavy hole density of states looks more like a step function.

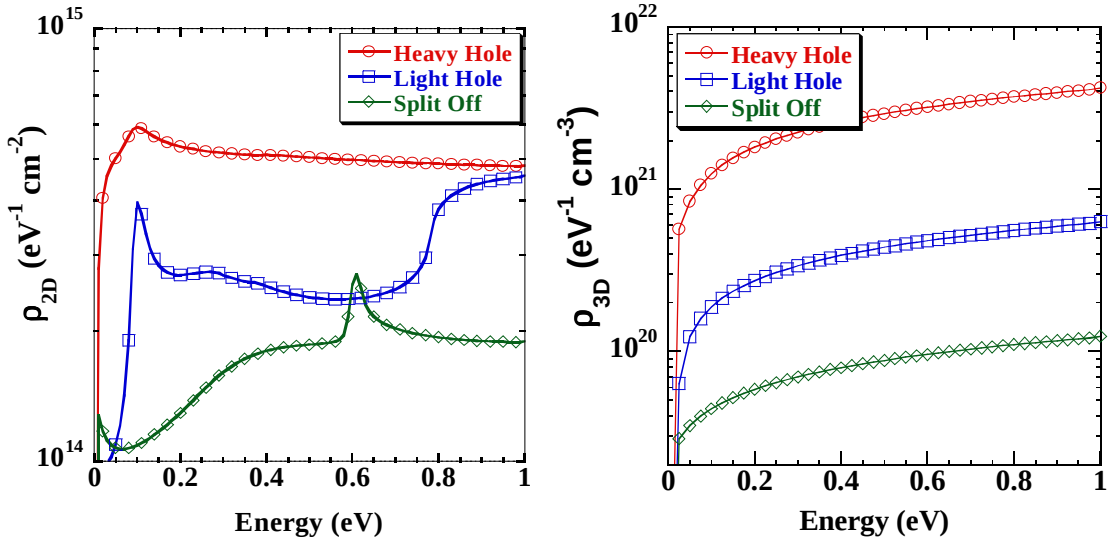


Figure 10. Density of States for channel (triangular test potential) and bulk (3D) carriers respectively from **k·p** calculations.

The output characteristics of a 25 nm p-channel conventional Si MOSFET are shown in Figure 11. Significant DIBL is seen in the output characteristics in this case. Only phonon scattering (acoustic and optical phonons) are included in this calculation. An equivalent effective mass two band (Heavy and Light Hole bands) model with similar scattering mechanisms included, underestimates the current by about 14%. Thus, it is clear that the effective mass approximations is not reliable and, therefore, band-structure calculations are required to accurately predict the output current under high field transport conditions in nano-scale MOSFETs, in particular the *p*-channel ones.

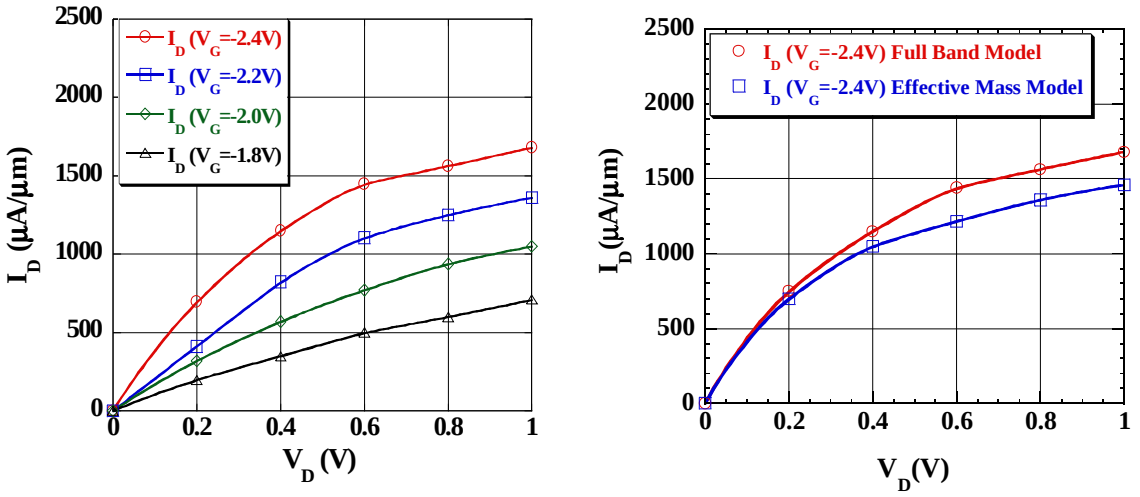


Figure 11. Output characteristics of a 25 nm p-channel Si MOSFET calculated using the full band and the effective mass model.

References

[1] P. Y. Yu and M. Cardona, Fundamentals of Semiconductors, Springer-Verlag, Berlin, 1999.

[2] C. Herring, Phys. Rev., 57 (1940) 1169.

[3] D. J. Chadi and M. L. Cohen, Phys. Stat. Sol. (b), 68 (1975) 405.

[4] J. Luttinger and W. Kohn, Phys. Rev., 97 (1955) 869.

[5] M. L. Cohen and T. K. Bergstresser, Phys. Rev., 141 (1966) 789.

[6] J. R. Chelikowsky and M. L. Cohen, Phys Rev. B, 14 (1976) 556.

[7] D. D. Awshalom, D. Loss and N. Samarth, Semiconductor Spintronics and Quantum Computation, Springer, Berlin, 2002.

[8] S. Datta and B. Das, Appl. Phys. Lett., Vol. 56, 665 (1990).

[9] Y. A. Bychkov, E. I. Rashba,, J. Phys. C: Solid State Phys., 17, No. 9, pp. 6039-6045, 30 Nov. 1984.

[10] G. Dresselhaus, *Physica Review*, Vol. 100, No. 2, pp. 580-586, 15 Oct. 1955

[11] J. Schliemann et al, Phys. Rev. Lett., Vol. 90, 146801(2003).

[12] S. D. Ganichev et al., Phys. Rev. Lett., Vol. 92 (2004).

[13] E. Fermi, Nuovo Cimento, 11 (1934) 157.

[14] H. J. Hellman, J. Chem. Phys., 3 (1935) 61.

[15] J. C. Phillips and L. Kleinman, Phys. Rev., 116 (1959) 287.

[16] J. R. Chelikowsky and M. L. Cohen, Phys Rev. B, 10 (1974) 12.

[17] L. R. Saravia and D. Brust, Phys. Rev., 176 (1968) 915.

[18] S. Gonzalez, Masters Thesis, Arizona State University, 2001.

[19] J. R. Chelikowsky and M. L. Cohen, Phys. Rev. B, 10 (1974) 5059.

[20] K. C. Padney and J. C. Phillips, Phys. Rev. B, 9 (1974) 1552.

[21] D. Brust, Phys. Rev. B, 4 (1971) 3497.

[22] Shynia Yamakawa, Computational Electronics, 481-485, December 2003.

[23] Z. H. Levine and S. G. Louie, Phys. Rev. B 25, 6310 - 6316 (1982).

[24] S. Schumacher, N. H. Kwong, and R. Binder, Phys. Rev. B 76, 245324 (2007)

[25] J. C. Slater and G. F. Coster, Phys. Rev., 94 (1954) 1498.

[26] P.-O. Löwdin, Journal of Chemical Physics, 19 (1951) 1396.

[27] E. O. Kane, J. Phys. Chem. Solids, 1 (1956) 82.

[28] E. O. Kane, J. Phys. Chem. Solids, 1 (1957) 249.

[29] J. M. Luttinger and W. Kohn, Phys. Rev. 97, 869 (1955).

[30] H. Kroemer, Proc. IEEE 63, 988 (1975).

[31] . Bastard, Wave Mechanics Applied to Semiconductor Heterostructures (Halsted Press, New York, 1989), ch. 3.

[32] von Klitzing, K.; Dorda, G.; Pepper, M, Physical Review Letters, vol.45, no.6, p. 494-7 (1980).

[33] A. T. Tilke et al., Physical Rev. B, vol. 68, 075311 (2003).

[34] Reed MA, Randall JN, Aggarwal RJ, Matyi RJ, Moore TM, Wetsel AE (1988). "Observation of discrete electronic states in a zero-dimensional semiconductor nanostructure". Phys Rev Lett 60 (6): 535–537. doi:10.1103/PhysRevLett.60.535. PMID 10038575. (1988).

[35] M.J. Gilbert and J.P. Bird, Applied Physics Letters, Vol. 77, pp. 1050-1052, 2000

[36] A. Ashwin, Masters Thesis, January 2005.

[37] S. Das Sarma et al., *IEEE Transactions on Magnetics*, Vol. 36, No. 5, Sept. 2000.

[38] I. Zutic, J. Fabian, S. Das , *Reviews of Modern Physics*, Vol. 76, pp. 323-410, April 2004

[39] S.A. Wolf et al., Science, Vol. 294, No. 10, pp. 1468-1493, Nov. 2001.

[40] M.J. Gilbert and J.P. Bird, Applied Physics Letters, Vol. 77, pp. 1050-1052, 2000

[41] R. Eppenga, M.F. Schuurmans, S. Colak, Physical Review B, Vol. 36, No. 3, pp. 1554-1564, 15 July 1987.

[42] P. Löwdin, The Journal of Chemical Physics, Vol. 19, No. 11, pp. 1396-1401, Nov. 1951.

[43] W. Zawadzki, P. Pfeffer, Semicond. Sci. Technol. 19, (2004) R1-R17.

[44] R. Silsbee, Journal of Physics: Condensed Matter 16, (2004) R179-R207.

[45] H.R. Trebin, U. Rossler, R. Ranvaud, Physical Review B, Vol. 20, No. 2, pp. 686-700, 15 July 1979

[46] Y. Chang, J.N. Schulman, Physical Review B, Vol. 31, No. 4, pp. 2069-2079, 15 Feb. 1985

[47] P. M. Garone, V. Venkataraman, and J. C. Sturm, IEEE Electron Device Lett., 13(1), 56 (1992).

[48] N. Collaert, P. Verheyen, K. De Meyer, R. Loo, M. Caymax, IEEE Trans. Nanotechnology, 1, 190 (2002).

[49] R. Oberhuber, G. Zandler, and P. Vogl, Phys. Rev. B, 58 (15), 9941 (1998).

[50] M.V. Fischetti, Z. Ren, P. M. Solomon, M. Yang, and K. Rim, Journal of Applied Physics, 94(2), 1079 (2003).