Chapter 2

Semiconductor Heterostructures

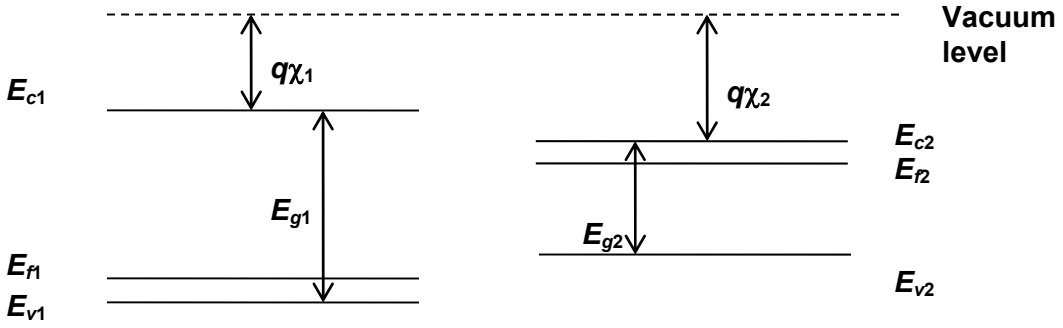
2.1 Introduction

Most interesting semiconductor devices usually have two or more different kinds of semiconductors. In this handout we will consider four different kinds of commonly encountered heterostructures:

a) pn heterojunction diode
b) nn heterojunctions
c) pp heterojunctions
d) Quantum wells, quantum wires, and quantum dots

2.2 A pn Heterojunction Diode

Consider a junction of a p-doped semiconductor (semiconductor 1) with an n-doped semiconductor (semiconductor 2). The two semiconductors are not necessarily the same, e.g. 1 could be AlGaAs and 2 could be GaAs. We assume that 1 has a wider band gap than 2. The band diagrams of 1 and 2 by themselves are shown below.



2.2.1 Electron Affinity Rule and Band Alignment:

How does one figure out the relative alignment of the bands at the junction of two different semiconductors? For example, in the Figure above how do we know whether the conduction band edge of semiconductor 2 should be above or below the conduction band edge of semiconductor 1? The answer can be obtained if one measures all band energies with respect to one value. This value is provided by the vacuum level (shown by the dashed line in the Figure above). The vacuum level is the energy of a free electron (an electron outside the semiconductor) which is at rest with respect to the semiconductor. The electron affinity, denoted by χ (units: eV), of a semiconductor is the energy required to move an electron from the conduction band bottom to the vacuum level and is a material constant. The electron affinity rule for band alignment says that at a heterojunction between different semiconductors the relative alignment of bands is dictated by their electron affinities, as shown in the Figure above. The electron affinity rule implies that the conduction band offset at a heterojunction interface is equal to the difference in the

electron affinities between the two semiconductors. This is shown in the band diagram below. According to the electron affinity rule the conduction band offset ΔE_c is given as,

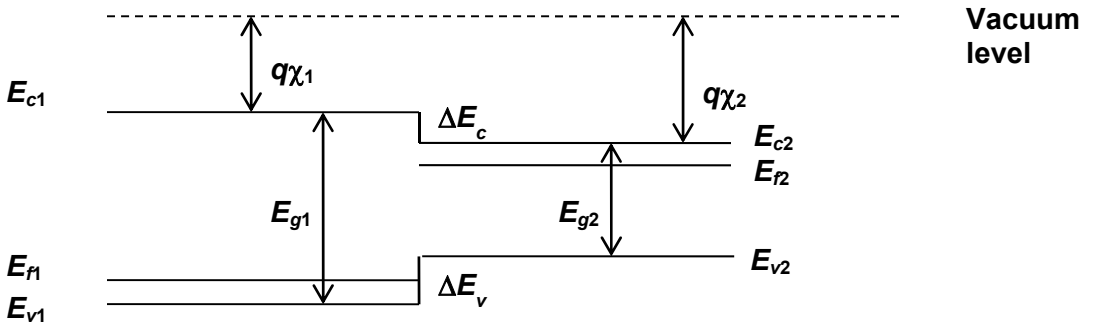
$$\Rightarrow \Delta E_c = q(\chi_2 - \chi_1)$$

The valence band offset is then,

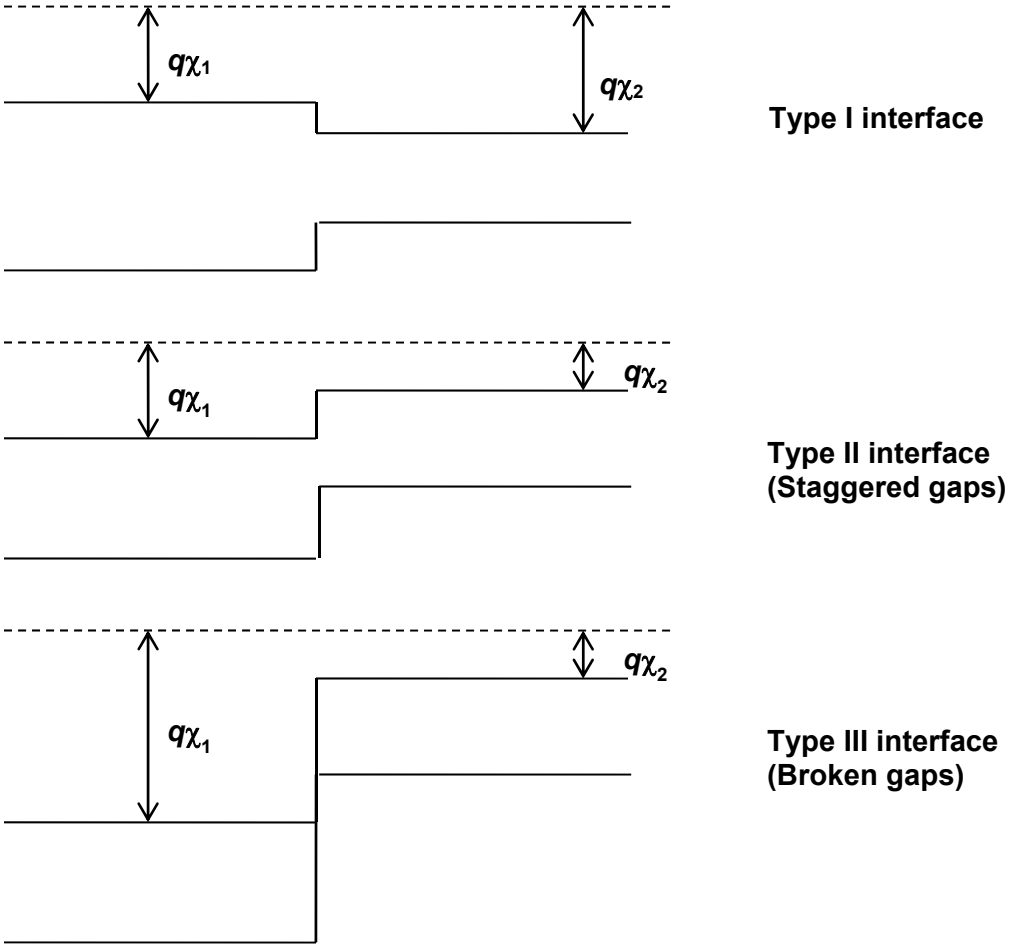
$$\Rightarrow \Delta E_v = (E_{g_1} - E_{g_2}) - \Delta E_c = \Delta E_g - \Delta E_c$$

Note that,

$$\Rightarrow \Delta E_c + \Delta E_v = \Delta E_g$$



Depending upon the difference between χ_1 and χ_2 we could have type I, type II, or type III heterojunction interfaces, as shown below.



The affinity rule does not always work well. The reason is that it attempts to use a bulk property of semiconductors to predict what happens at the interfaces and there is no good reason it should work in the first place. Usually, the conduction band offsets ΔE_c at interfaces are determined experimentally.

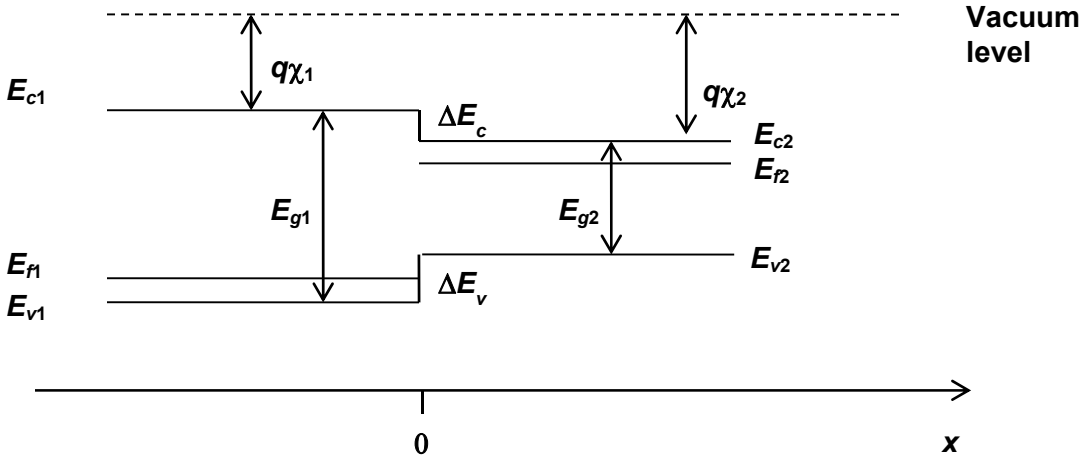
Now we get back to the pn heterojunction and assume that:

i) We have a type I interface

ii) The conduction band offset ΔE_c is known.

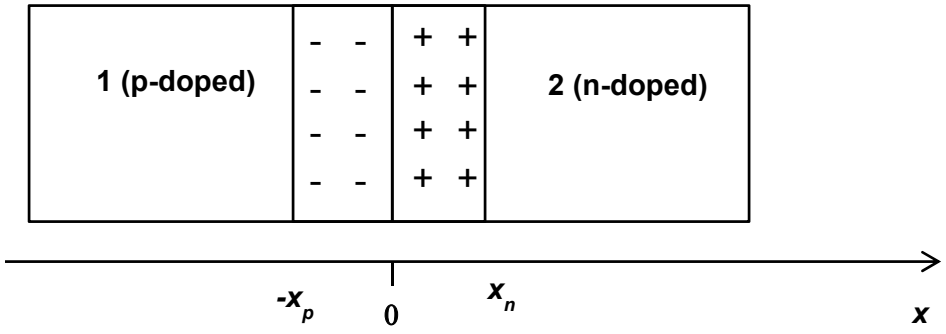
iii) The doping in semiconductor 1 is p-type and equal to N_a and the doping in semiconductor 2 is n-type and equal to N_d and all dopants are ionized.

The resulting band alignment is shown below:

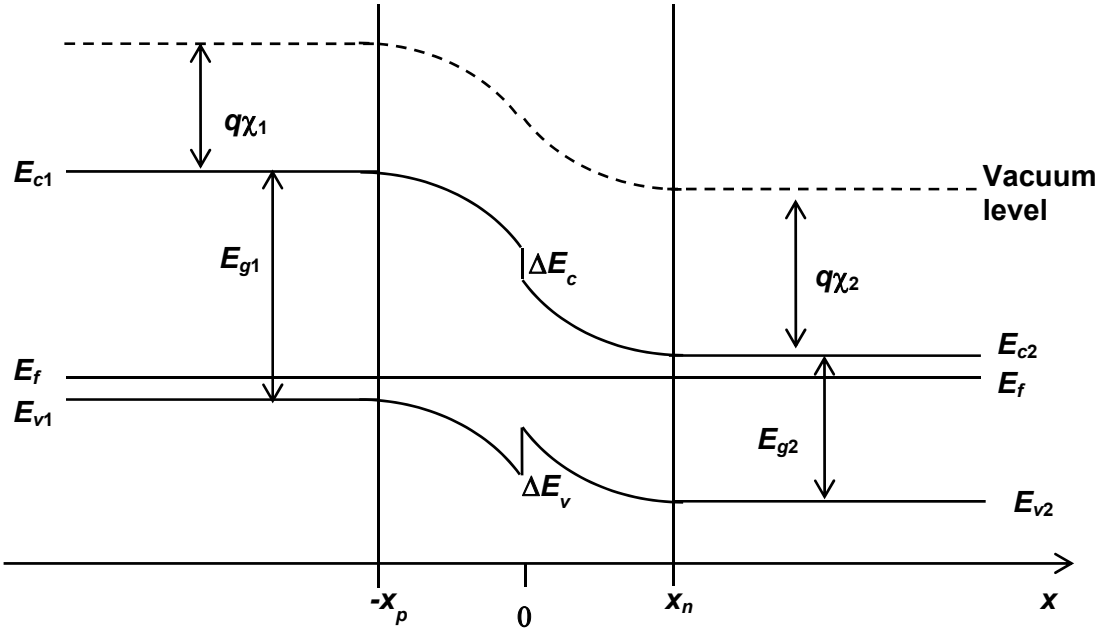


2.2.2 A pn Heterojunction Diode in Thermal Equilibrium:

Clearly the situation shown above in the Figure is not a correct description at equilibrium; the Fermi level is not flat and constant throughout the structure. Let us see how this equilibrium gets established when a junction is formed. There are more electrons in the region $x > 0$ (where $n = n_{no} = N_d$) than in the region $x < 0$ (where $n = n_{po} = \frac{n_{i1}^2}{N_a}$). Similarly, there are more holes in the region $x < 0$ (where $p = p_{po} = N_a$) than in the region $x > 0$ (where $p = p_{no} = \frac{n_{i2}^2}{N_d}$). Consequently, electrons will diffuse from the n-side into the p-side and holes will diffuse from the p-side into the n-side. Electrons on the n-side are contributed by the ionized donor atoms. When some of the electrons near the interface diffuse into the p-side, they leave behind positively changed donor atoms. Similarly, when the holes near the interface diffuse into the n-side, they leave behind negatively changed acceptor atoms. As the electron and holes diffuse, a net positive charge density appears on the right side of the interface ($x > 0$) and a net negative charge density appears on the left side of the interface $(x < 0)$. This situation is shown below. The charged regions are called depletion regions because they are depleted of the majority carriers. As a result of the charge densities on both sides of the interface, an electric field is generated that points in the negative x-direction (from the donor atoms to the acceptor atoms). The drift components of the electron and hole currents due to the electric field are in direction opposite to the electron and hole diffusive currents, respectively. As more electrons and holes diffuse, the strength of the electric field increases until the drift components of the electron and hole currents are exactly equal and opposite to their respective diffusive components. When this has happened the junction is in equilibrium.



Recalling that electrostatic potentials need to be added to the energies in band diagrams, the equilibrium band diagram looks like as shown below. Note that band bending that occurs inside the depletion regions reflecting the presence of an electric field and a corresponding electrostatic potential. Also note that the Fermi level in equilibrium is flat and constant throughout the device. The vacuum level also bends in response to the electric field, as shown.



Band bending implies an electric field and, therefore, a potential difference across the junction. This "built-in" potential V_{bi} can be found as follows. If we look at the "raw" band alignment (i.e. before considering any band bending) qV_{bi} must the difference in the Fermi levels on the two sides of the junction. This is because the bands need to band by this much so that the Fermi level is constant and flat throughout the device in equilibrium.

$$qV_{bi} = E_{f2} - E_{f1}$$

Since,

$$E_{f2} - E_{c2} = KT \ln\left(\frac{N_d}{N_{c2}}\right) \qquad E_{v1} - E_{f1} = KT \ln\left(\frac{N_a}{N_{v1}}\right)$$

Adding these equations we get,

$$qV_{bi} = E_{f2} - E_{f1} = E_{c2} - E_{v1} + KT\ln\left(\frac{N_a N_d}{N_{c2}N_{v1}}\right)$$

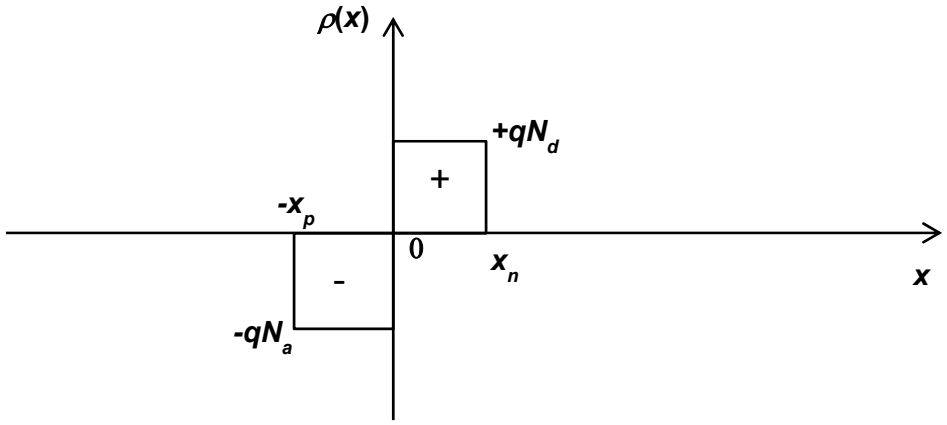
$$qV_{bi} = E_{g1} - \Delta E_c + KT\ln\left(\frac{N_a N_d}{N_{c2}N_{v1}}\right)$$

$$= E_{g2} + \Delta E_v + KT\ln\left(\frac{N_a N_d}{N_{c2}N_{v1}}\right).$$

2.2.3 The Depletion Approximation:

From the band diagram, one can see that the majority carrier concentrations in the depletion regions on both sides of the junction are going to be small because the difference between the Fermi level and the band edges becomes large. The depletion approximation assumes that the majority carrier concentration is exactly zero in the depletion regions of thicknesses x_p and x_n on the p-side and n-side of the junction, respectively. It is a good approximation to calculate the junction electric field and potential. The net charge density in the depletion regions is then due to the ionized donors and acceptors and is as shown below. The net charge on both sides of the junction has to be equal and opposite,

$$qN_d x_n = qN_a x_p$$

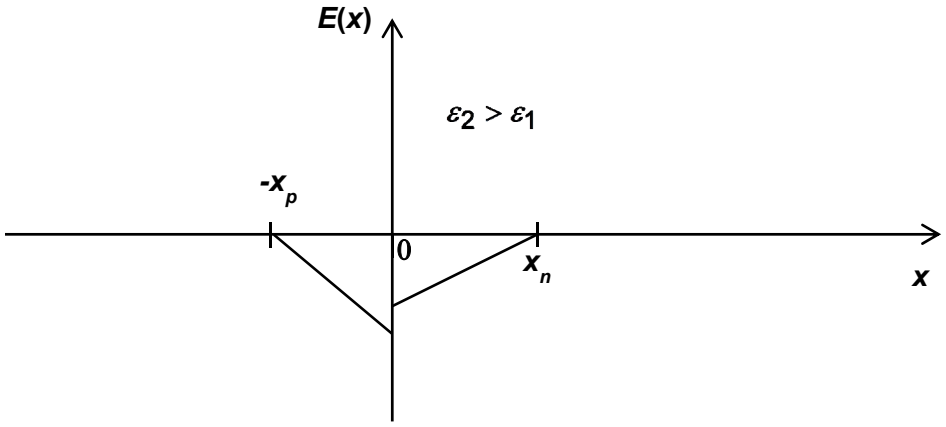


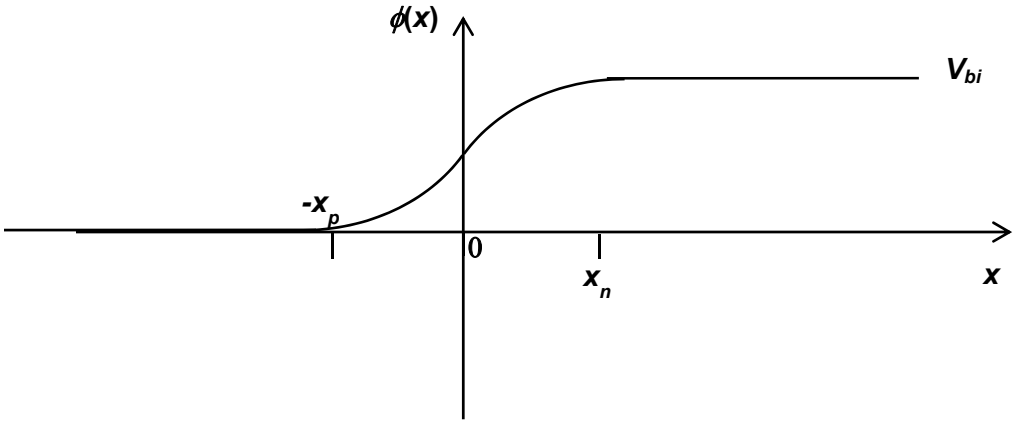
$$\rho(x) = \begin{cases} +qN_d & 0 < x \le x_n \\ -qN_a & -x_p \le x < 0 \\ 0 & \text{elsewhere} \end{cases}$$

Knowing the charge density one can calculate the electric field and the potential drops using Gauss's law,

$$E(x) = \begin{cases} +\dfrac{qN_d(x_n - x)}{\varepsilon_2} & 0 < x \le x_n \\ -\dfrac{qN_a(x + x_p)}{\varepsilon_1} & -x_p \le x < 0 \\ 0 & \text{elsewhere} \end{cases}$$

$$\phi(x) = \begin{cases} \dfrac{qN_a x_p^2}{2\varepsilon_1} + \dfrac{qN_d\left(x_n x - x^2/2\right)}{\varepsilon_2} & 0 < x \le x_n \\ \dfrac{qN_a\left(x + x_p\right)^2}{2\varepsilon_1} & -x_p \le x < 0 \\ \dfrac{qN_a x_p^2}{2\varepsilon_1} + \dfrac{qN_d x_n^2}{2\varepsilon_2} & x > x_n \\ 0 & x < -x_p \end{cases}$$





Potential drop on n-side is,

$$q\frac{N_d x_n^2}{2\varepsilon_2}$$

Potential drop on p-side is,

$$q\frac{N_a x_p^2}{2\varepsilon_1}$$

This implies,

$$V_{bi} = q\frac{N_d x_n^2}{2\varepsilon_2} + q\frac{N_a x_p^2}{2\varepsilon_1}$$

Using the above equation with $qN_d x_n = qN_a x_p$ gives the thicknesses of the depletion regions,

$$x_p = \left[\frac{2}{q}\varepsilon_1\varepsilon_2\frac{N_d}{N_a}\frac{V_{bi}}{(\varepsilon_1 N_a + \varepsilon_2 N_d)}\right]^{\frac{1}{2}}$$

$$x_n = \left[\frac{2}{q}\varepsilon_1\varepsilon_2\frac{N_a}{N_d}\frac{V_{bi}}{(\varepsilon_1 N_a + \varepsilon_2 N_d)}\right]^{\frac{1}{2}}$$

Total thickness of depletion region is,

$$W = x_n + x_p = \left[\frac{2}{q}\frac{\varepsilon_1\varepsilon_2}{(\varepsilon_1 N_a + \varepsilon_2 N_d)}\frac{(N_a + N_d)^2}{N_a N_d}V_{bi}\right]^{\frac{1}{2}}$$

Charge (per unit area) on one side of the junction, given by $qN_d x_n = qN_a x_p = Q$, is,

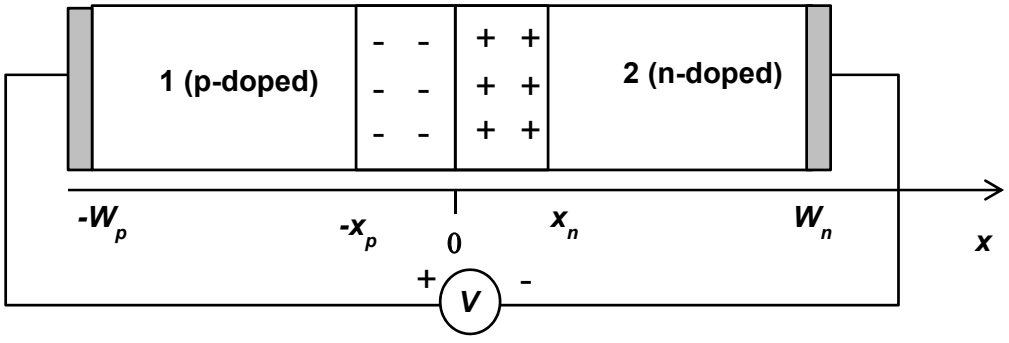
$$Q = \left[2q\frac{\varepsilon_1\varepsilon_2 N_a N_d}{(\varepsilon_1 N_a + \varepsilon_2 N_d)}V_{bi}\right]^{\frac{1}{2}}$$

2.2.4 Quasi-Neutral Regions:

The n and p regions outside the depletion region are called quasi-neutral regions because they always remain, to a very good approximation, charge neutral even when the device is biased. This charge-neutral property is generally true for all good conductors. Significant charge densities cannot be present inside good conductors because if a charge density were present then it would generate electric fields which would in turn generate drift currents that would neutralize the charge density. On length scales longer than the screening length and on time scales longer than the dielectric relaxation time, good conductors are charge neutral.

2.2.5 A Reverse Biased pn Heterojunction:

Now we attach metal (ohmic) contacts to the n and p sides, and apply a voltage V on the p contact w.r.t. the n contact, as shown below. We assume that, $V < 0$ (reverse bias).

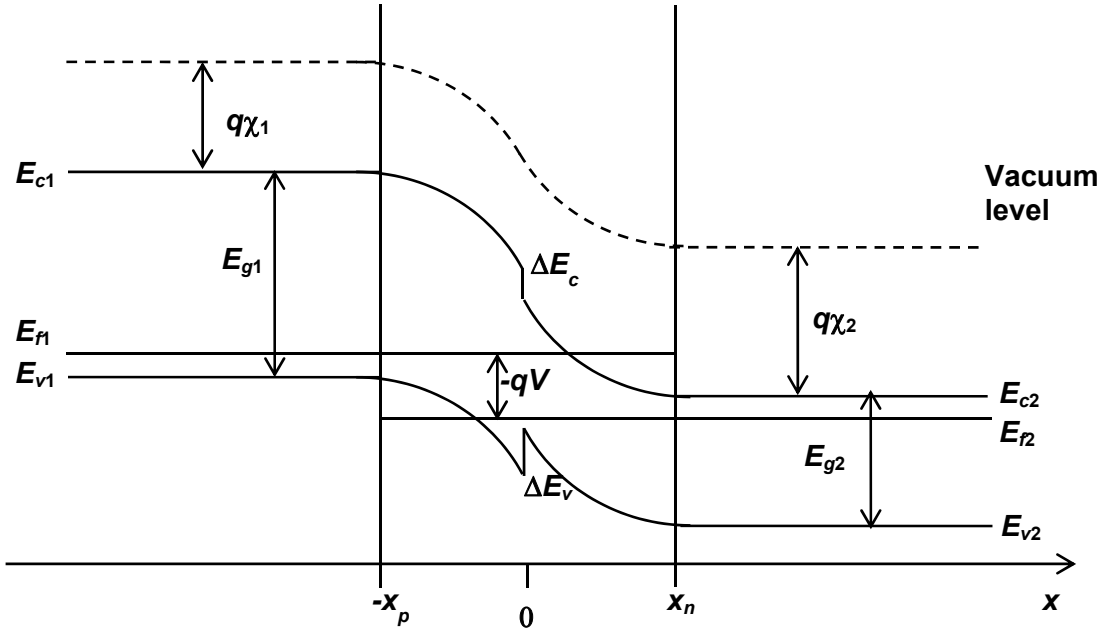


We will also assume that the applied potential falls completely across the depletion region (i.e. across the junction) and not across the conductive n or p regions (quasi-neutral regions) or across the metal contacts. The effect of the applied voltage is taken into account by changing the electrostatic potential across the depletion region from V_{bi} to $V_{bi} - V$. Therefore, the depletion region width will change (and increase because $V < 0$) to accommodate the added potential,

$$x_n(V) = \left[\frac{2}{q}\varepsilon_1\varepsilon_2\frac{N_d}{N_a}\frac{(V_{bi}-V)}{(\varepsilon_1 N_a + \varepsilon_2 N_d)}\right]^{\frac{1}{2}}$$

$$x_p(V) = \left[\frac{2}{q}\varepsilon_1\varepsilon_2\frac{N_a}{N_d}\frac{(V_{bi}-V)}{(\varepsilon_1 N_a + \varepsilon_2 N_d)}\right]^{\frac{1}{2}}$$

The peak electric field in the junction will also increase and the resulting band diagram will look as shown in the Figure below.



Note that now since an external voltage has been applied the device is no longer in equilibrium and the Fermi levels on the n-side and the p-side are not the same. The splitting of the Fermi levels is exactly equal to the applied voltage,

$$qV = E_{f2} - E_{f1}$$

A current will also flow through the device but we will postpone the discussion of the current till we discuss the forward biased case below.

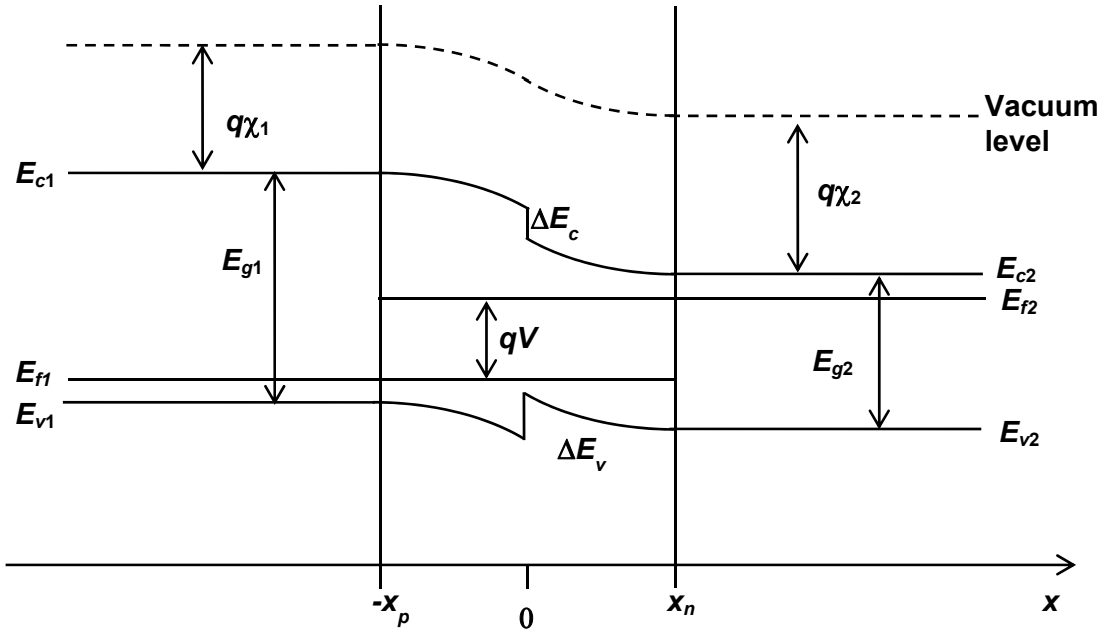
2.2.6 A Forward Biased pn Heterojunction:

Now we consider the case $V > 0$ (forward bias). We will as before assume that the applied potential falls completely across the depletion region (i.e. across the junction) and not across the conductive n or p regions (quasi-neutral regions) or across the metal contacts. The effect of the applied voltage is taken into account by changing the electrostatic potential across the depletion region from V_{bi} to $V_{bi} - V$. Therefore, the depletion region width will change (and decrease because $V > 0$) to accommodate the added potential,

$$x_n(V) = \left[\frac{2}{q}\varepsilon_1\varepsilon_2\frac{N_d}{N_a}\frac{(V_{bi}-V)}{(\varepsilon_1 N_a + \varepsilon_2 N_d)}\right]^{\frac{1}{2}}$$

$$x_p(V) = \left[\frac{2}{q}\varepsilon_1\varepsilon_2\frac{N_a}{N_d}\frac{(V_{bi}-V)}{(\varepsilon_1 N_a + \varepsilon_2 N_d)}\right]^{\frac{1}{2}}$$

The peak electric field in the junction will also decrease and the resulting band diagram will look as shown in the Figure below.



Note again that since an external voltage has been applied the device is no longer in equilibrium and the Fermi levels on the n-side and the p-side are not the same. The splitting of the Fermi levels is exactly equal to the applied voltage,

$$qV = E_{f2} - E_{f1}$$

2.2.7 Minority Carrier Injection and Current Flow:

Calculating current flow in pn hetero junction diodes is complicated. The model presented here, called the drift-diffusion model, is valid only if the band offsets ΔE_C and ΔE_V are small (compared to KT). The essential assumption in the drift-diffusion is that carrier drift and diffusion in the n-doped and p-doped regions (not including the depletion regions) are the main bottleneck for electron transport and transport

through the depletion regions or at the heterointerface are not the main bottlenecks. These assumptions do not always hold. Nevertheless, drift-diffusion model provides some useful insights.

The current flow in a forward biased pn device can be understood as follows. In equilibrium, the drift components of both the electron and hole currents balanced the respective diffusive components. When a forward bias is applied, the electric field in the junction is reduced and therefore the drift components of the electron and hole currents also decrease. The diffusion components of the electron and hole currents are then larger than the drift components and, therefore, a net current will flow through the device and this current will be diffusive in nature. Below we calculate this current.

The assumption here is that the main bottleneck to electron and hole transport in the device comes from the quasi-neutral regions and electron and hole transport through the depletion region (whether due to drift or diffusion) is not the bottleneck. Consequently, one may assume that the electrons in the depletion region are in equilibrium with the electrons in the n-doped side (and share the same Fermi level) and holes in the depletion region are in equilibrium with the holes on the p-doped side (and share the same Fermi level). This is the reason why the Fermi levels E_{f1} and E_{f2}, as drawn in the band diagram under forward bias, are extended into the depletion regions. We can therefore write,

$$n(-x_p) = N_{c1}\, e^{\left(E_{f2}(-x_p)-E_c(-x_p)\right)/KT} = N_{c1}\, e^{\left(E_{f1}(-x_p)-E_c(-x_p)\right)/KT}\, e^{\left(E_{f2}(-x_p)-E_{f1}(-x_p)\right)/KT}$$

$$= \frac{n_{i1}^2}{N_a}\, e^{qV/KT} = n_{po}\, e^{qV/KT}$$

$$p(-x_p) = N_{v1}\, e^{\left(E_v(-x_p)-E_{f1}(-x_p)\right)/KT} = N_a$$

And,

$$n(x_n) = N_{c2}\, e^{\left(E_{f2}(x_n)-E_c(x_n)\right)/KT} = N_d$$

$$p(x_n) = N_{v2}\, e^{\left(E_v(x_n)-E_{f1}(x_n)\right)/KT} = N_{v2}\, e^{\left(E_v(x_n)-E_{f2}(x_n)\right)/KT}\, e^{\left(E_{f2}(x_n)-E_{f1}(x_n)\right)/KT}$$

$$= \frac{n_{i2}^2}{N_d}\, e^{qV/KT} = p_{no}\, e^{qV/KT}$$

When a voltage is applied across the junction, the electric field in the depletion region is reduced, and the diffusion currents exceed the drift currents. Electrons diffuse from the n-side through the depletion region into the p-side and holes diffuse from the p-side through the depletion region into the n-side. What happens to the electrons once they make it to the p-side? They keep diffusing but they suddenly find a great number of holes with which to recombine. The generation-recombination rate of these "injected" electrons (which are minority carriers) on the p-side is given by,

$$R_e(x) - G_e(x) = \frac{n(x) - n_{po}}{\tau_{e1}}$$

By assumption, there is no potential drop across the quasi-neutral p-doped region, therefore the electric field in this region is (almost) zero and the electron current is entirely due to diffusion,

$$J_e(x) = q\, D_{e1} \frac{\partial n(x)}{\partial x}.$$

Since,

$$\frac{\partial n}{\partial t} - \frac{1}{q}\frac{\partial}{\partial x} J_e(x) = G_e(x) - R_e(x)$$

And,

$$\frac{\partial n}{\partial t} = 0 \text{ (no time dependence in steady state)}$$

We get,

$$\Rightarrow D_{e1}\frac{\partial^2 n(x)}{\partial x^2} = \frac{n(x)-n_{po}}{\tau_{e1}}$$

$$\Rightarrow \frac{\partial^2 n(x)}{\partial x^2} = \frac{n(x)-n_{po}}{L_{e1}^2}$$

$$L_{e1} = \sqrt{D_{e1}\,\tau_{e1}} = \text{Minority carrier diffusion lenght}$$

Similarly, for the holes injected into the n-side the corresponding equation is,

$$\frac{\partial^2 p(x)}{\partial x^2} = \frac{p(x)-p_{no}}{L_{h2}^2}$$

$$L_{h2} = \sqrt{D_{h2}\,\tau_{h2}}$$

To solve the above two equations we need appropriate boundary conditions. Consider first the electrons on the p-side. One boundary condition is,

$$n(-x_p) = n_{po}\; e^{qV/KT}$$

The second boundary condition is usually obtained by assuming that at the left most metal contact the recombination time is extremely fast so that the electron density must equal the equilibrium density,

$$n(-W_p) = n_{po}$$

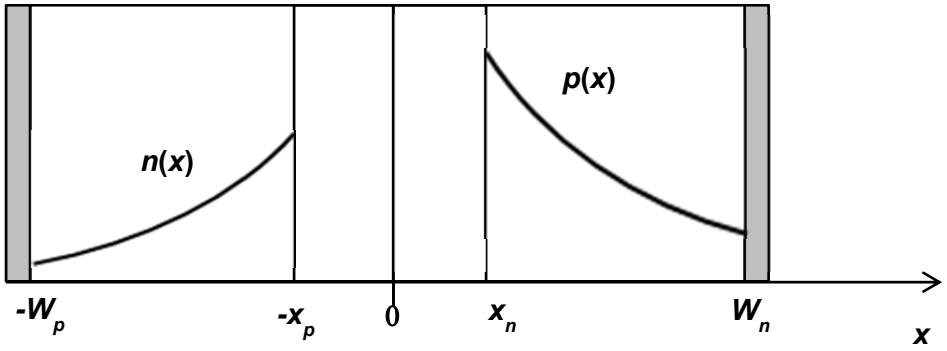
The solution for the electron density is then,

$$n(x) = n_{po} + n_{po}\frac{\sinh\left(\frac{W_p + x}{L_{e1}}\right)}{\sinh\left(\frac{W_p - x_p}{L_{e1}}\right)}\left(e^{\frac{qV}{KT}} - 1\right) \qquad -W_p \le x \le -x_p$$

Similarly, for the holes on the n-side one obtains,

$$p(x) = p_{no} + p_{no}\frac{\sinh\left(\frac{W_n - x}{L_{h2}}\right)}{\sinh\left(\frac{W_n - x_n}{L_{h2}}\right)}\left(e^{\frac{qV}{KT}} - 1\right) \qquad x_n \le x \le W_n$$

These solutions are sketched in the Figure below.



The electron current density $J_e(x)$ (due to diffusion) can now be found on the p-side,

$$J_e(x) = q D_{e1} \frac{\partial n(x)}{\partial x}$$

$$= q \frac{n_{i1}^2}{N_a} \frac{D_{e1}}{L_{e1}} \frac{\cosh\left(\frac{W_p + x}{L_{e1}}\right)}{\sinh\left(\frac{W_p - x_p}{L_{e1}}\right)} \left(e^{\frac{qv}{KT}} - 1 \right) \qquad -W_p \le x \le -x_p$$

The hole current density on the n-side is,

$$J_h(x) = -q D_{nh} \frac{\partial p(x)}{\partial x}$$

$$= q \frac{n_{i2}^2}{N_d} \frac{D_{h2}}{L_{h2}} \frac{\cosh\left(\frac{W_n - x}{L_{n2}}\right)}{\sinh\left(\frac{W_n - x_n}{L_{n2}}\right)} \left(e^{\frac{qv}{KT}} - 1 \right) \qquad x_n \le x \le W_n$$

The total current density is the sum of the electron and the hole density,

$$J_T = J_e(x) + J_h(x)$$

The total current density must be independent of position in the steady state. We can find J_T if we know both $J_e(x)$ and $J_h(x)$ at any one location. If one ignores recombination and generation processes inside the depletion region then the electron and hole current density must be constant throughout the depletion region,

$$J_e(-x_p) = J_e(x_n) = J_e(x) \quad \{-x_p \le x \le x_n\}$$
$$J_h(-x_p) = J_h(x_n) = J_h(x) \quad \{-x_p \le x \le x_n\}$$

But,

$$J_e(-x_p) = q \frac{n_{i1}^2}{N_a} \frac{D_{e1}}{L_{e1}} \coth\left(\frac{W_p - x_p}{L_{e1}}\right) \left(e^{\frac{qv}{KT}} - 1 \right)$$

$$J_h(x_n) = q \frac{n_{i2}^2}{N_d} \frac{D_{h2}}{L_{h2}} \coth\left(\frac{W_n - x_n}{L_{h2}}\right) \left(e^{\frac{qv}{KT}} - 1 \right)$$

$$\Rightarrow J_T = q \left\{ \frac{n_{i1}^2}{N_a} \frac{D_{e1}}{L_{e1}} \coth\left(\frac{W_p - x_p}{L_{e1}}\right) + \frac{n_{i2}^2}{N_d} \frac{D_{h2}}{L_{h2}} \coth\left(\frac{W_n - x_n}{L_{h2}}\right) \right\} \left(e^{\frac{qv}{KT}} - 1 \right)$$

The current in the external circuit is,

$$I = A J_T = I_o \left(e^{\frac{qv}{KT}} - 1 \right)$$

where,

$$I_o = qA \left\{ \frac{n_{i1}^2}{N_a} \frac{D_{e1}}{L_{e1}} \coth\left(\frac{W_p - x_p}{L_{e1}}\right) + \frac{n_{i2}^2}{N_d} \frac{D_{h2}}{L_{h2}} \coth\left(\frac{W_n - x_n}{L_{h2}}\right) \right\}$$

2.2.8 Majority Carrier Dynamics and Quasi-Neutrality:

We know $n(x)$ for $-W_p \le x \le -x_p$. Since $n(x)$ in this range is greater than the equilibrium electron density (n_{po}) the presence of these injected electrons can create charge imbalances, resulting in net negative charge density on the p-side and cause large electric fields. But this never actually happens because the majority carriers (i.e. the holes) move to quickly screen these injected carriers and maintain charge neutrality ("quasi neutrality" as it is called). Therefore, in steady state,

$$\Delta p(x) = p(x) - N_a = \Delta n(x) = n(x) - n_{po} \quad -W_p \le x \le -x_p$$

and similarly on the n-side,

$$\Delta n(x) = n(x) - N_d = \Delta p(x) = p(x) - p_{no} \quad x_n \le x \le W_n$$

The excess majority carrier concentration therefore equals the excess minority carrier concentration to maintain charge neutrality. On the p-side the hole currents is,

$$J_h(x) = qp(x)\mu_{h1}E(x) - q\,D_{h1}\frac{\partial p(x)}{\partial x}$$

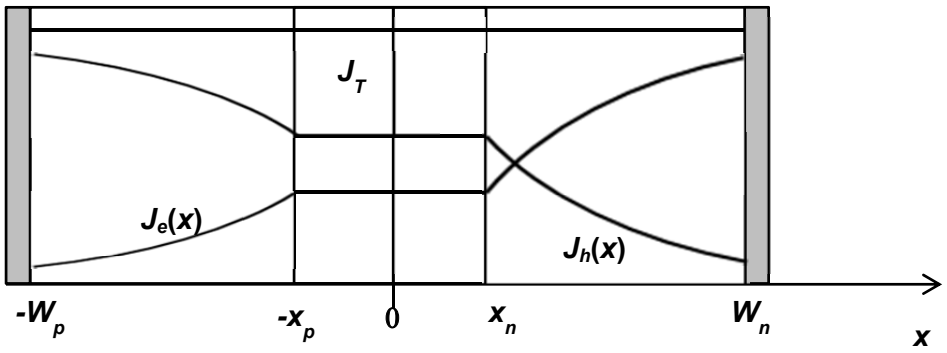
$$= J_T - J_e(x)$$

We know $J_e(x)$ on the p-side and we know J_T and therefore we can calculate $J_h(x)$. On the n-side, the electron current is,

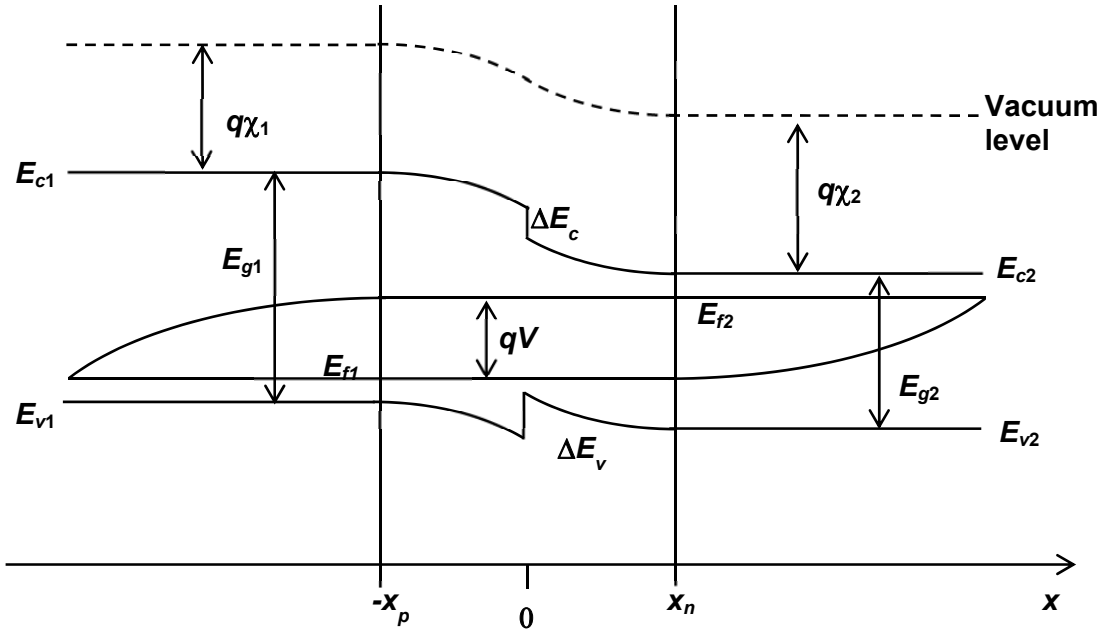
$$J_e(x) = qn(x)\mu_{e2}E(x) + q\,D_{e2}\frac{\partial n(x)}{\partial x}$$

$$= J_T - J_h(x)$$

Again, we know $J_h(x)$ on the n-side and we can calculate $J_h(x)$. All the current so obtained densities are plotted in the Figure below.

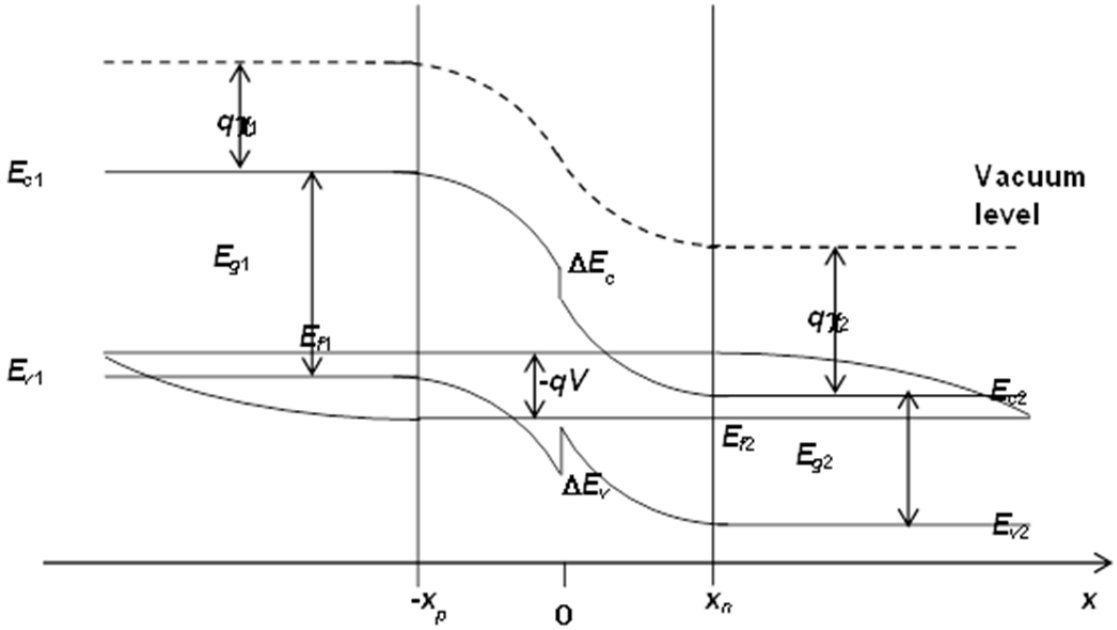


We can now sketch the electron and hole Fermi levels through the entire device in forward bias. These are shown in the Figure below. As the injected minority carries reach equilibrate with the majority carriers, the minority carrier Fermi level approaches the majority carrier Fermi level.



2.2.9 Cuurent Flow in Reverse Biased pn Heterojunction:

We now go back to the case of the reverse biased pn heterojunction and study the current flow. Note that all formulas derived for the forward bias case are also applicable to the reverse bias case. The band diagram in reverse bias is then as shown below.



The expression for the total current is therefore also valid for the reverse bias case,

$$I = I_o\left(e^{\frac{qv}{KT}} - 1\right)$$

When the reverse bias voltage is large the current approaches $-I_o$. Why is there a small negative current $-I_o$? This is due to electron and hole generation in the n and p quasi-neutral regions. When $V > 0$ in forward bias, the electron (hole) density on the p-side (n-side) increases above its equilibrium value as a results of minority carrier injection via increased diffusion. When $V < 0$ in reverse bias, the electron (hole) density on the p-side (n-side) is pulled below its equilibrium value. This happens because in reverse bias the electric field in the depletion region increases and the drift components of the electron and hole currents exceed the corresponding diffusion components. The large electric field in the depletion region in reverse bias sucks the minority carriers and transfers them across the depletion region. Consequently, minority carrier generation rate exceeds the recombination rate in the quasi-neutral regions, and this generation of minority carriers on both sides of the depletion region and the subsequent transfer of these generated carriers across the depletion region is responsible for the current in a reverse biased diode. An important rule of thumb to remember is that when the conduction electron Fermi level is larger than the valence electron Fermi level (or the hole Fermi level) then the recombination rate exceeds the generation rate (as in a forward biased device) and when the conduction electron Fermi level is smaller than the valence electron Fermi level (or the hole Fermi level) then the generation rate exceeds the recombination rate (as in a reverse biased device).

2.2.10 Generation and Recombination in the Depletion Region:

Generation and recombination in the depletion region also contributes to the total diode current in both forward and reverse bias. Consider the electron current inside the depletion region. The continuity equation in steady state gives,

$$-\frac{1}{q}\frac{\partial}{\partial x}J_e(x) = G_e(x) - R_e(x)$$

Integrating over the depletion region gives,

$$J_e(x_n) - J_e(-x_p) = q\int_{-x_p}^{x_n}[R_e(x) - G_e(x)]dx$$

Similarly, we get for the hole current,

$$J_h(-x_p) - J_h(x_n) = q\int_{-x_p}^{x_n}[R_h(x) - G_h(x)]dx$$

In the derivation for the total diode current, we added the electron and hole currents at one edge of the depletion region. The same procedure now yields,

$$J_T = q\left\{\frac{n_{i1}^2}{N_a}\frac{D_{e1}}{L_{e1}}\coth\left(\frac{W_p - x_p}{L_{e1}}\right) + \frac{n_{i2}^2}{N_d}\frac{D_{h2}}{L_{h2}}\coth\left(\frac{W_n - x_n}{L_{h2}}\right)\right\}\left(e^{\frac{qv}{KT}} - 1\right) + q\int_{-x_p}^{x_n}[R_e(x) - G_e(x)]dx$$

The current in the external circuit is,

$$I = AJ_T = I_o\left(e^{\frac{qv}{KT}} - 1\right) + qA\int_{-x_p}^{x_n}[R_e(x) - G_e(x)]dx$$

The second term on the right hand side also generally has an exponential dependence on the applied voltage. For example, in the case of trap-assisted recombination-generation (the dominant mechanism in silicon pn diodes),

$$G_e(x) - R_e(x) \propto \left(np - n_i^2\right)$$

Inside the depletion region,

$$np = n_i^2 e^{(E_{fc} - E_{fh})/KT}$$

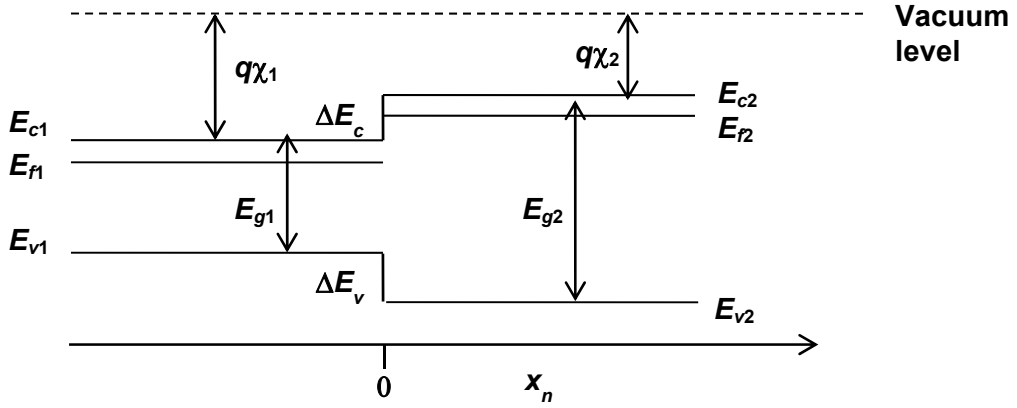
and therefore,

$$G_e(x) - R_e(x) \propto \left(e^{\frac{qv}{KT}} - 1 \right)$$

2.3 A nn Heterojunction

2.3.1 A nn Heterjunction in Thermal Equilibrium:

Consider the following nn heterojunction.



The band diagram in equilibrium looks like as shown below. There is a small depletion region formed on the right side of the junction and a small accumulation region formed on the left side of the junction.

The built-in voltage of the junction is,

$$qV_{bi} = E_{f2} - E_{f1}$$

which can also be written as,

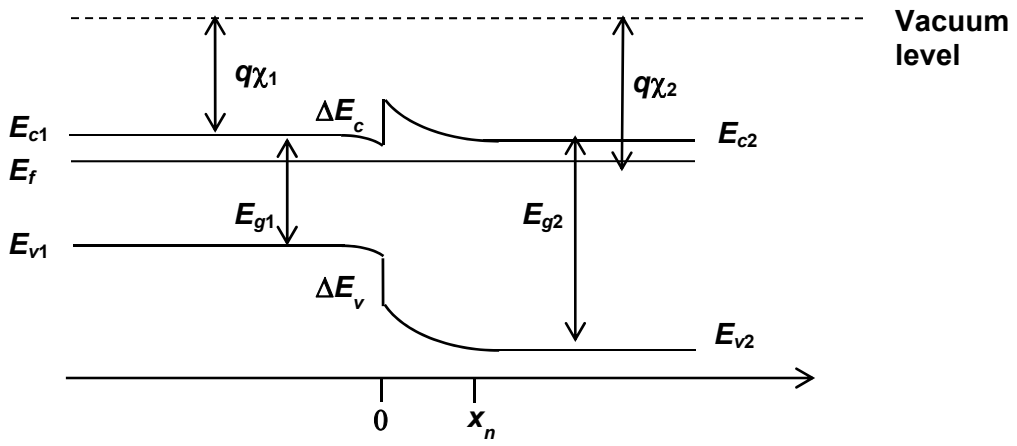
$$qV_{bi} = E_{f2} - E_{f1} = \Delta E_c + KT \ln\left(\frac{N_{d2}}{N_{d1}} \frac{N_{c1}}{N_{c2}} \right)$$

One can go further and calculate the electric field and potential on both sides of the junction. Let the potential drop in semiconductor 1 be ϕ_1 and let it be ϕ_2 in semiconductor 2.

$$\Rightarrow \phi_1 + \phi_2 = V_{bi}$$

The potential drop in the depletion region can be related to the thickness of the depletion region,

$$\phi_2 = q \frac{N_{d2} x_n^2}{2\varepsilon_2}$$



In region $-\infty \le x \le 0$, assume that the potential is zero at $-\infty$. Then we can write,

$$n(x) = N_{c1}\, e^{\frac{E_f - E_{c1}(x)}{KT}}$$

$$= N_{c1}\, e^{\frac{E_f - E_{c1}(-\infty) + q\phi(x)}{KT}}$$

$$n(x) = N_{c1}\, e^{\frac{E_f - E_c(-\infty)}{KT}}\, e^{\frac{q\phi(x)}{KT}}$$

$$n(x) = N_{d1}\, e^{\frac{q\phi(x)}{KT}}$$

Poisson equation gives,

$$-\varepsilon_1 \frac{\partial^2 \phi(x)}{\partial x^2} = q[N_{d1} - n(x)]$$

$$-\varepsilon_1 \frac{\partial^2 \phi(x)}{\partial x^2} = q N_{d1}\left[1 - e^{\frac{q\phi(x)}{KT}}\right]$$

$$\Rightarrow \frac{\partial}{\partial x}\left(\frac{\partial \phi}{\partial x}\right)^2 = 2q\frac{N1}{\varepsilon_1}\left[e^{\frac{q\phi(x)}{KT}} - 1\right]\frac{\partial \phi}{\partial x}$$

Integrate from $-\infty$ to 0 using the boundary conditions,

$$\left\{\begin{matrix}\phi(-\infty) = 0 \\ \phi(0) = \phi_1\end{matrix}\right\}$$

to get,

$$\left(\frac{\partial \phi}{\partial x}\right)^2\Bigg|_{x=0} = \frac{2N_{d1}KT}{\varepsilon_1}\left[\left(e^{\frac{q\phi_1}{KT}} - 1\right) - \frac{q\phi_1}{KT}\right]$$

$$\Rightarrow E(x)\big|_{x=0^-} = -\sqrt{\frac{2N_{d1}KT}{\varepsilon_1}}\sqrt{\left(e^{\frac{q\phi_1}{KT}} - 1\right) - \frac{q\phi_1}{KT}}$$

The electric fields on both sides of the junction are related by Gauss's law,

$$\varepsilon_2 E(x)\big|_{x=0^+} = \varepsilon_1 E(x)\big|_{x=0^-} = -qN_{d2}x_n$$

$$\Rightarrow \sqrt{2\varepsilon_1 N_{d1} KT}\sqrt{\left(e^{\frac{q\phi_1}{KT}} - 1\right) - \frac{q\phi_1}{KT}} = qN_{d2}x_n \qquad (1)$$

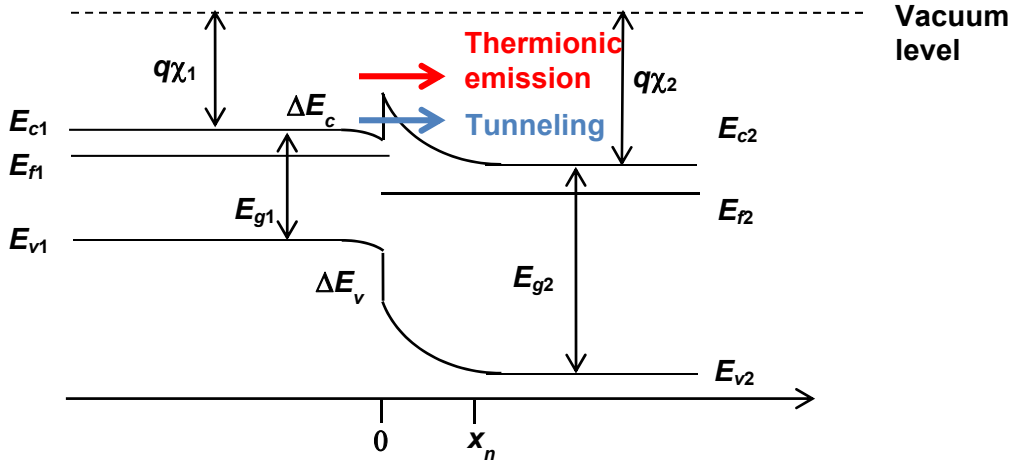
also,

$$\phi_2 + \frac{qN_{d1}x_n^2}{2\varepsilon_1} = V_{bi} \qquad (2)$$

Equations (1) and (2) can be solved for x_n, ϕ_1 and ϕ_2

2.3.2 Current Flow by Thermionic Emission and Quantum Mechanical Tunneling:

Suppose now a voltage is applied from an external source such that the potential on the right side is increased by V. The band diagram looks as shown in the Figure.



Computing current flow across the junction depends on prior knowledge or assumption of the main bottleneck to current flow in the device. The bottleneck can be either transport across the heterojunction interface via thermionic emission or quantum mechanical tunneling or the bottleneck can be the subsequent transport of the electrons via drift-diffusion. Current flow across the heterojunction interface is by two mechanisms:

i) Thermionic emission over the heterojunction barrier

ii) Quantum mechanical tunneling through the heterojunction barrier

Both these mechanisms are depicted in the band diagram above. Generally transport across the interface rather than drift-diffusion in the quasi-neutral regions is the main bottleneck to current flow. Suppose the electron energy band dispersion in both semiconductor 1 is,

$$E(k) = E_{c1} + \frac{\hbar}{2m_{e1}}\left(k_x^2 + k_y^2 + k_z^2\right)$$

The electron velocity in the x-direction is given by,

$$v_x(k_x) = \frac{1}{\hbar}\frac{\partial E(k)}{\partial k_x} = \frac{\hbar k_x}{m_{e1}}$$

Let the quantum mechanical transmission probability of electrons for getting through the barrier (either through it or over it) be $t(\vec{k})$. Electron flux going from left to right is:

$$F_{L\to R} = 2\int_{\infty}^{\infty}\frac{dk_y}{2\pi}\int_{-\infty}^{\infty}\frac{dk_z}{2\pi}\int_{0}^{\infty}\frac{dk_x}{2\pi}\cdot v_x(k_x)\,t(\vec{k})\,f(E(k)-E_{f1})[1-f(E(k)-E_{f2})]$$

Similarly, the flux from right to left is,

$$F_{R\to L} = 2\int_{\infty}^{\infty}\frac{dk_y}{2\pi}\int_{-\infty}^{\infty}\frac{dk_z}{2\pi}\int_{-\infty}^{0}\frac{dk_x}{2\pi}\cdot v_x(k_x)\,t(\vec{k})\,f(E(k)-E_{f2})[1-f(E(k)-E_{f1})]$$

The total electron flux is the difference between the right moving and left moving fluxes,

$$F = F_{L\to R} + F_{R\to L}$$

$$= 2\int_{\infty}^{\infty}\frac{dk_y}{2\pi}\int_{-\infty}^{\infty}\frac{dk_z}{2\pi}\int_{0}^{\infty}\frac{dk_x}{2\pi}\cdot v_x(k_x)\,t(\vec{k})\,[f(E(k)-E_{f1})-f(E(k)-E_{f2})]$$

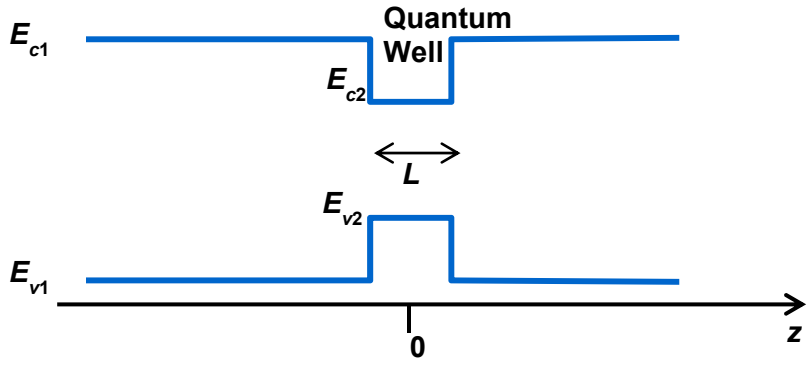
And the current is then,

$$I = -qAF = -2qA\int_{\infty}^{\infty}\frac{dk_y}{2\pi}\int_{-\infty}^{\infty}\frac{dk_z}{2\pi}\int_{0}^{\infty}\frac{dk_x}{2\pi}\cdot v_x(k_x)\,t(\vec{k})\,[f(E(k)-E_{f1})-f(E(k)-E_{f2})]$$

The negative sign comes in because current is in direction opposite to the electron flux.

2.4 Quantum Wells, Energy Subbands, and Reduced Density of States in Two Dimensions

Consider a thin layer of a semiconductor with a smaller bandgap sandwiched between two wider bandgap semiconductors, as shown below.



If the length L of the smaller bandgap material is smaller than the electron decoherence length then the electron energies in the smaller band gap material are quantized. The electrons in the smaller bandgap material are confined in the potential well formed by the conduction band discontinuities but are free in the plane of the quantum well (the x-y plane). Semiconductor 2 is called the quantum well and semiconductor 1 is called the barrier. We will solve this problem using the effective mass theorem.

2.4.1 Effective Mass Theorem:

The effective mass theorem is very useful in the context of semiconductor heterostructures. Consider a semiconductor with conduction band energy dispersion given by,

$$E_c(\vec{k}) = E_c + \frac{\hbar^2}{2}(\vec{k}-\vec{K}_c).M_e^{-1}.(\vec{k}-\vec{K}_c)$$

According to the effective mass theorem the wavefunction $\psi(\vec{r})$ of an electron in the semiconductor in the vicinity of wavevector $\vec{K}_c$ is most generally written (even in the presence of an added potential $V(\vec{r})$) as a product of the Bloch function $\psi_{c,\vec{K}_c}(\vec{r})$ and a slowly varying envelope function $\phi(\vec{r})$,

$$\psi(\vec{r}) = \phi(\vec{r})\,\psi_{c,\vec{K}_c}(\vec{r})$$

where $\phi(\vec{r})$ satisfies the effective mass Schrodinger equation,

$$\left[E_c(-i\nabla) + V(\vec{r})\right]\phi(\vec{r}) = E\phi(\vec{r})$$

The above equation can be used to find the energy of the electron.

For the quantum well problem, we assume parabolic-isotropic conduction and valence bands for both the semiconductors. Consider first the conduction band states. The conduction band energy dispersions are,

$$E_{c1}(k) = E_{c1} + \frac{\hbar^2}{2m_{e1}}\left(k_x^2 + k_y^2 + k_z^2\right)$$

$$E_{c2}(k) = E_{c2} + \frac{\hbar^2}{2m_{e2}}\left(k_x^2 + k_y^2 + k_z^2\right)$$

In semiconductor 1, the barrier, the effective mass equation is,

$$\left(\frac{-\hbar^2\nabla^2}{2m_{e1}} + E_{c1}\right)\phi(\vec{r}) = E\phi(\vec{r})$$

In semiconductor 2, the quantum well, we have,

$$\left(\frac{-\hbar^2\nabla^2}{2m_{e2}} + E_{c2}\right)\phi(\vec{r}) = E\phi(\vec{r})$$

The solutions are of the form,

$$\left.\begin{aligned}&\phi(\vec{r}) = A\,\sin(k_z z)\frac{e^{ik_x x}}{\sqrt{L_x}}\frac{e^{ik_y y}}{\sqrt{L_y}}\\ &\text{or}\\ &\phi(\vec{r}) = B\,\cos(k_z z)\frac{e^{ik_x x}}{\sqrt{L_x}}\frac{e^{ik_y y}}{\sqrt{L_y}}\end{aligned}\right\} -\frac{L}{2} \le z \le \frac{L}{2}$$

$$\left.\phi(\vec{r}) = C e^{-\alpha(z-L/2)}\frac{e^{ik_x x}}{\sqrt{L_x}}\frac{e^{ik_y y}}{\sqrt{L_y}}\right\} \frac{L}{2} < z \le \infty$$

$$\left.\phi(\vec{r}) = D e^{\alpha(z+L/2)}\frac{e^{ik_x x}}{\sqrt{L_x}}\frac{e^{ik_y y}}{\sqrt{L_y}}\right\} -\infty \le z < -\frac{L}{2}$$

The solution form corresponds to the electron confined in the potential well formed by the conduction band discontinuities but free in the plane of the quantum well (the x-y plane). We have also assumed that the area of the quantum well in the x-y plane is $L_x L_y$. The effective mass equation in the well gives,

$$E = E_{c2} + \frac{\hbar^2}{2m_{e2}}\left(k_z^2 + k_x^2 + k_y^2\right)$$

and in the barrier the effective mass equation gives,

$$E = E_{c1} + \frac{\hbar^2}{2m_{e1}}(-\alpha^2 + k_x^2 + k_y^2)$$

$$\Rightarrow \frac{\hbar^2 k_z^2}{2m_{e2}} = \Delta E_c + \frac{\hbar^2}{2}\left(\frac{1}{m_{e1}} - \frac{1}{m_{e2}}\right)\left(k_x^2 + k_y^2\right) - \frac{\hbar^2\alpha^2}{2m_{e1}} \qquad (1)$$

2.4.2 Envelope Function Boundary Conditions:

Solving Schrodinger equation requires boundary conditions. In textbook quantum mechanics these boundary conditions are the continuity of the wavefunction and that of its derivative at all boundaries. The second boundary condition comes from the continuity of the probability current at a boundary. The effective mass equation satisfies more complicated boundary conditions. For the simple case where the energy band disperions are parabolic and isotropic the boundary conditions are simple and are given below:

i) The envelope function must be continuous at a boundary
ii) The derivative of the envelope function weighted by the inverse effective mass is continuous at a boundary

These boundary conditions at the $z = L/2$ boundary give,

$$\phi(\vec{r})\big|_{z=(L/2)^-} = \phi(\vec{r})\big|_{z=(L/2)^+}$$

$$\frac{1}{m_{e2}}\frac{\partial\phi(\vec{r})}{\partial z}\bigg|_{z=(L/2)^-} = \frac{1}{m_{e1}}\frac{\partial\phi(\vec{r})}{\partial z}\bigg|_{z=(L/2)^+}$$

Similar boundary conditions apply at $z = -L/2$. These boundary conditions give,

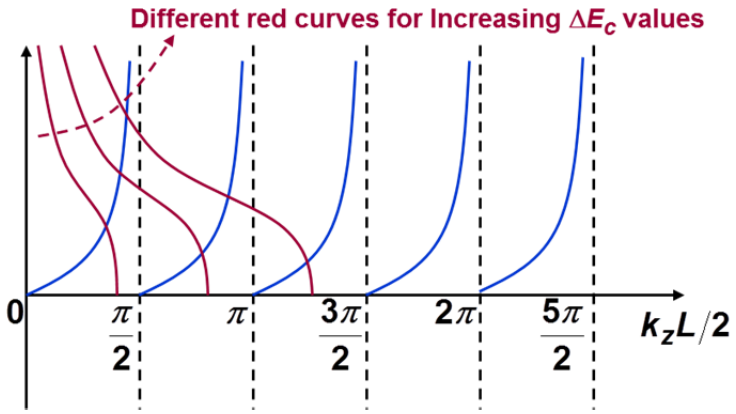
$$\tan\left(\frac{k_z L}{2}\right) = \frac{\alpha}{k_z}\frac{m_{e2}}{m_{e1}} \quad \{\text{for the cosine solutions} \qquad (2)$$

or

$$\cot\left(\frac{k_z L}{2}\right) = \frac{-\alpha}{k_z}\frac{m_{e2}}{m_{e1}} \quad \{\text{for the sine solutions} \qquad (3)$$

Equations (2) and (3) together with Equation (1) yield discrete values of k_z (denote them by $k_z(n)$) for every value of $k_{||}^2 = k_x^2 + k_y^2$ (and $n = 1,2,3,\ldots$). An easy way to graphically solve Equations (2) and (3) is to substitute for α using Equation (1) and then plot the right hand and left hand sides as a function of k_z and observe where they intersect. This is demonstrated in the Figure below where the right hand sides are plotted for different values of the conduction band discontinuity ΔE_c or the depth of the potential well. Note that the number of solution depend on the value of ΔE_c. Suppose these discrete values $k_z(n)$ have been found, Let,

$$E_n = \frac{\hbar^2 k_z^2(n)}{2m_{e2}} \qquad \text{for } n = 1,2,3,\ldots.$$



Note that these discrete $k_z(n)$ values depend on $k_{||}^2 = k_x^2 + k_y^2$ and so we should write $E_n(k_{||})$ to be more explicit. However, if $m_{e2} \approx m_{e1}$ then this dependence is weak (see Equation (1)) and may be ignored. The total energy of an electron confined in the quantum well can finally be written as,

$$E(n, k_{||}) = E_{c2} + E_n + \frac{\hbar^2 k_{||}^2}{2m_{e2}}$$

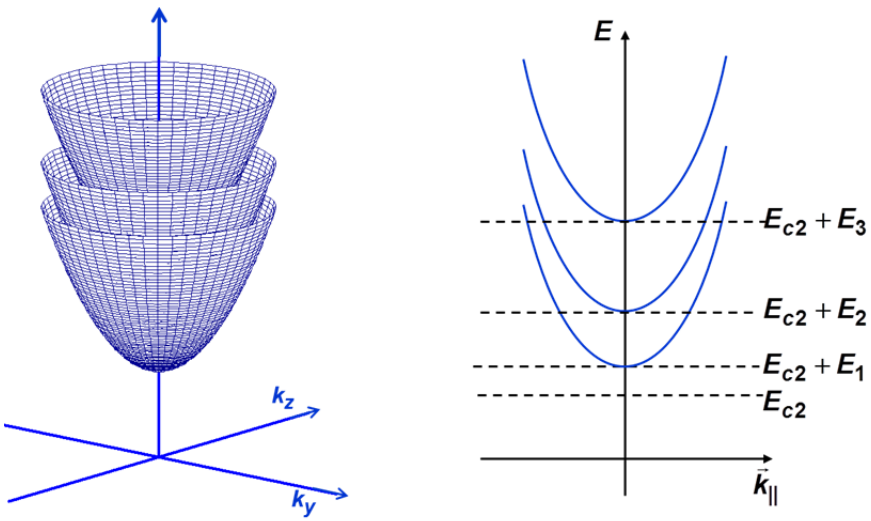
In the limit $\Delta E_c \to \infty$ the values of E_n are given by simple textbook infinite potential well result,

$$k_z(n) = \frac{n\pi}{L} \qquad \Rightarrow \qquad E_n = \frac{\hbar^2}{2m_{e2}}\left(\frac{n\pi}{L}\right) \qquad n = 1,2,3...$$

The electron energies in the quantum well given by,

$$E(n, k_{||}) = E_{c2} + E_n + \frac{\hbar^2 k_{||}^2}{2m_{e2}}$$

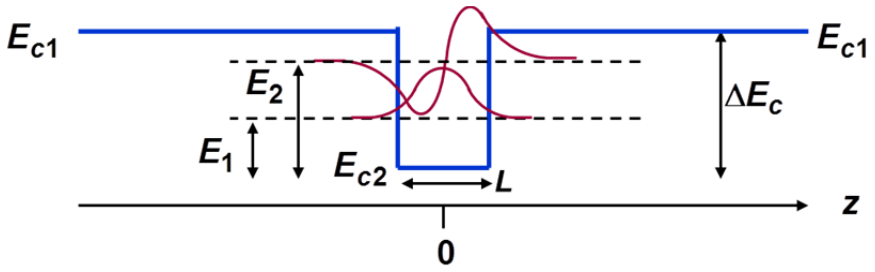
are plotted as a function of the in-plane wavevector $k_{||}$ in the Figure below.



2.4.3 Discussion – 2-Dimensional Electron Gas and Energy Subbands:

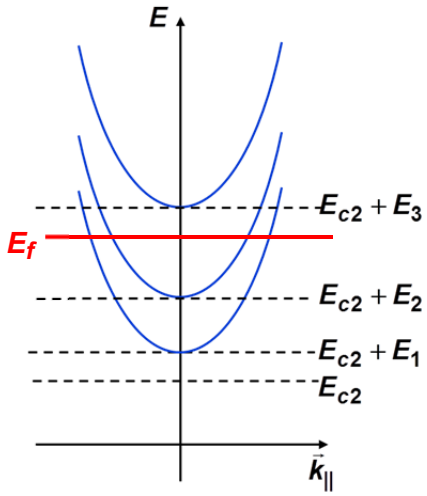
The electron motion is confined in the z-direction. Therefore its energy due to motion in the z-direction can only take discrete values. However, in the x- and y-directions (in the plane of the quantum well) the electron is free and its energy due to motion in these directions has the standard free electron form. Electrons so confined in a quantum well are said to constitute a 2 dimensional electron gas since the

effective degrees of freedom for motion are only 2 (as opposed to 3). As a result of this quantum confinement the conduction band energy dispersion gets modified. The 3D conduction band splits into several 2D conduction subbands, as shown in the Figure above. The electron wavefunctions (the envelope functions) are depicted for the first two confined energy levels ($n = 1,2$) in the Figure below.



2.4.4 2-Dimensional Density of States:

Suppose we know the Fermi level E_f, the question then is what is the electron density (per cm^{-2}) in the quantum well. The question can be answered if one first recalls from basic solid state physics that the number of allowed states per unit area in k-space in 2 dimensions equal $2A/(2\pi)^2$. So in an area $d^2\vec{k}_{||}$ in k-space the number of allowed states is $2A\, d^2\vec{k}_{||}/(2\pi)^2$.



The total number of electrons in all the confined levels in the quantum well is,

$$N = \sum_j 2 \times A \int \frac{d^2\vec{k}_{||}}{(2\pi)^2} f\left(E_c\left(j, k_{||}\right) - E_f\right)$$

$$\Rightarrow n = \sum_j 2 \times \int \frac{d^2\vec{k}_{||}}{(2\pi)^2} f\left(E_c\left(j, k_{||}\right) - E_f\right)$$

As a standard practice, we want to be able to write the above expression as a one dimensional integral over energy in the form,

$$n = \int_{-\infty}^{\infty} dE\; g_{QW}(E)\, f(E - E_f)$$

where $g_{QW}(E)$ is the conduction band density of states (number of states per unit area of the quantum well per unit energy) for the quantum well. Using the energy relation,

$$E(n,k_{||}) = E_{c2} + E_n + \frac{\hbar^2 k_{||}^2}{2m_{e2}}$$

one can proceed as follows,

$$n = \sum_j 2 \times \int \frac{d^2\vec{k}_{||}}{(2\pi)^2} f(E_c(j,k_{||}) - E_f)$$

$$= \sum_j \int_{E_{c2}+E_j}^{\infty} dE \left(\frac{m_{e2}}{\pi\hbar^2}\right) f(E - E_f)$$

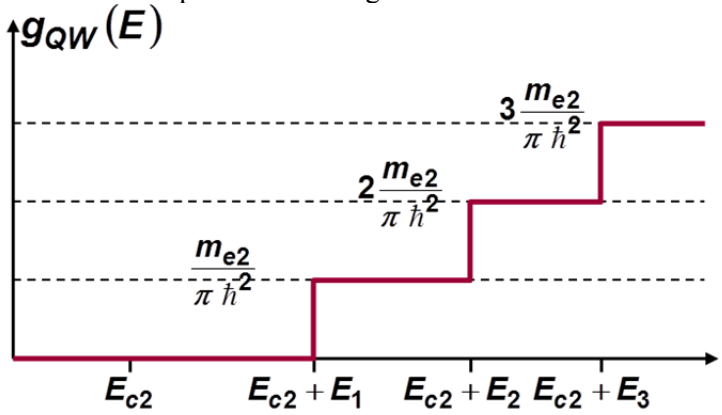
$$= \int_{-\infty}^{\infty} dE \sum_j \left(\frac{m_{e2}}{\pi\hbar^2}\right) \theta(E - E_{c2} - E_j) f(E - E_f)$$

$$= \int_{-\infty}^{\infty} dE\, g_{QW}(E) f(E - E_f)$$

where,

$$g_{QW}(E) = \sum_j \left(\frac{m_{e2}}{\pi\hbar^2}\right) \theta(E - E_{c2} - E_j)$$

The density of states function is plotted in the Figure below.



The 3D conduction band density of states in a bulk semiconductor with energy dependence proportional to $\sqrt{E - E_{c2}}$ gets modified into the staircase structure shown above as a result of quantum confinement. Each confined energy level contributes to the density of states an amount equal to $m_{e2}/\pi\hbar^2$.

2.4.5 Valence Band Energy Subbands:

Suppose the valence band energy dispersions are,

$$E_{v1}(k) = E_{v1} - \frac{\hbar^2}{2m_{h1}}\left(k_x^2 + k_y^2 + k_z^2\right)$$

$$E_{v2}(k) = E_{v2} - \frac{\hbar^2}{2m_{h2}}\left(k_x^2 + k_y^2 + k_z^2\right)$$

In semiconductor 1, the barrier, the effective mass equation for the valence band electrons is,

$$\left(\frac{-\hbar^2\nabla^2}{2m_{h1}} - E_{v1}\right)\phi(\vec{r}) = -E\phi(\vec{r})$$

In semiconductor 2, the quantum well, we have,

$$\left(\frac{-\hbar^2\nabla^2}{2m_{h2}} - E_{c2}\right)\phi(\vec{r}) = -E\phi(\vec{r})$$

The solutions are again of the form,

$$\left.\begin{aligned}&\phi(\vec{r}) = A\,\sin(k_z z)\frac{e^{ik_x x}}{\sqrt{L_x}}\frac{e^{ik_y y}}{\sqrt{L_y}}\\ &\text{or}\\ &\phi(\vec{r}) = B\,\cos(k_z z)\frac{e^{ik_x x}}{\sqrt{L_x}}\frac{e^{ik_y y}}{\sqrt{L_y}}\end{aligned}\right\} -\frac{L}{2} \le z \le \frac{L}{2}$$

$$\left.\phi(\vec{r}) = C e^{-\alpha(z-L/2)}\frac{e^{ik_x x}}{\sqrt{L_x}}\frac{e^{ik_y y}}{\sqrt{L_y}}\right\} \frac{L}{2} < z \le \infty$$

$$\left.\phi(\vec{r}) = D e^{\alpha(z+L/2)}\frac{e^{ik_x x}}{\sqrt{L_x}}\frac{e^{ik_y y}}{\sqrt{L_y}}\right\} -\infty \le z < -\frac{L}{2}$$

The solution form corresponds to the electron in the valence band confined in the potential well formed by the valence band discontinuities but free in the plane of the quantum well (the x-y plane). We have also assumed that the area of the quantum well in the x-y plane is $L_x L_y$. The effective mass equation in the well gives,

$$E = E_{v2} - \frac{\hbar^2}{2m_{h2}}\left(k_z^2 + k_x^2 + k_y^2\right)$$

and in the barrier the effective mass equation gives,

$$E = E_{v1} - \frac{\hbar^2}{2m_{h1}}(-\alpha^2 + k_x^2 + k_y^2)$$

$$\Rightarrow \frac{\hbar^2 k_z^2}{2m_{e2}} = \Delta E_v + \frac{\hbar^2}{2}\left(\frac{1}{m_{h1}} - \frac{1}{m_{h2}}\right)\left(k_x^2 + k_y^2\right) - \frac{\hbar^2\alpha^2}{2m_{h1}} \qquad (1)$$

Using the effective mass boundary conditions give,

$$\tan\left(\frac{k_z L}{2}\right) = \frac{\alpha}{k_z}\frac{m_{h2}}{m_{h1}} \quad \{\text{for the cosine solutions} \qquad (2)$$

or

$$\cot\left(\frac{k_z L}{2}\right) = \frac{-\alpha}{k_z}\frac{m_{h2}}{m_{h1}} \quad \{\text{for the sine solutions} \qquad (3)$$

The rest of the discussion proceeds exactly as in the case of the conduction band energy levels. The total energy of a valence band electron confined in the quantum well can finally be written as,

$$E\left(n, k_{||}\right) = E_{v2} - E_n - \frac{\hbar^2 k_{||}^2}{2m_{h2}}$$

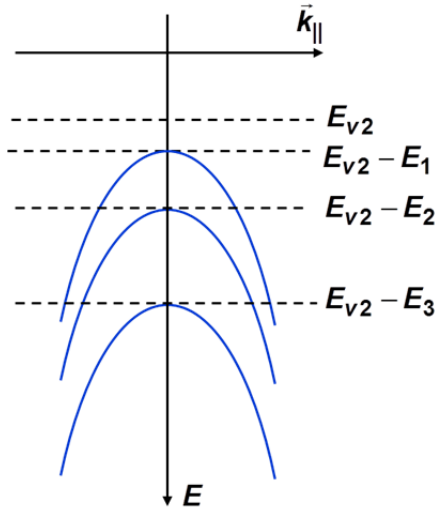
In the limit $\Delta E_v \to \infty$ the values of E_n are given by simple textbook infinite potential well result,

$$k_z(n) = \frac{n\pi}{L} \qquad \Rightarrow \qquad E_n = \frac{\hbar^2}{2m_{h2}}\left(\frac{n\pi}{L}\right) \qquad n = 1,2,3\ldots$$

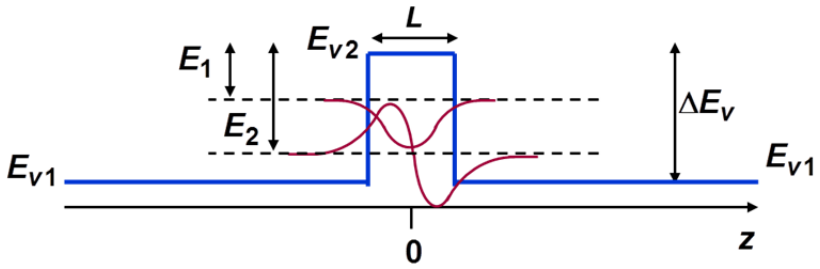
The electron energies in the quantum well given by,

$$E(n,k_{||}) = E_{v2} - E_n - \frac{\hbar^2 k_{||}^2}{2m_{h2}}$$

are plotted as a function of the in-plane wavevector $k_{||}$ in the Figure below.



The wavefunctions (the envelope functions) for the first two confined energy levels are shown in the Figure below.



2.4.6 2-Dimensional Density of States:

Suppose we know the Fermi level E_f, the question then is what is the hole density (per cm^{-2}) in the quantum well. The total number of holes in all the confined levels in the quantum well is,

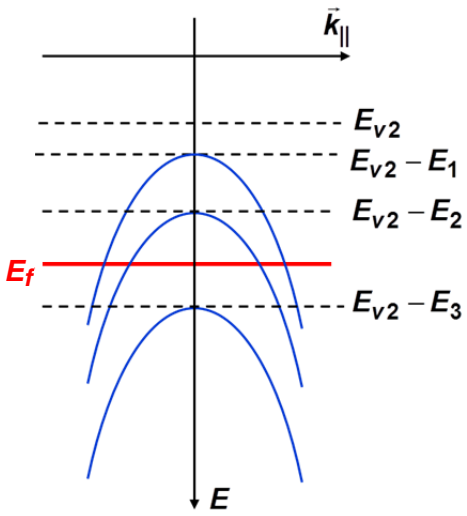
$$P = \sum_j 2 \times A \int \frac{d^2\vec{k}_{||}}{(2\pi)^2} \left[1 - f\left(E_v\left(j,k_{||}\right) - E_f\right)\right]$$

$$\Rightarrow p = \sum_j 2 \times \int \frac{d^2\vec{k}_{||}}{(2\pi)^2} \left[1 - f\left(E_v\left(j,k_{||}\right) - E_f\right)\right]$$

As a standard practice, we want to be able to write the above expression as a one dimensional integral over energy in the form,

$$p = \int_{-\infty}^{\infty} dE \; g_{QW}(E) \left[1 - f(E - E_f)\right]$$

where $g_{QW}(E)$ is the valence band density of states (number of states per unit area of the quantum well per unit energy) for the quantum well.



Using the energy relation,

$$E(n,k_{||}) = E_{v2} - E_n - \frac{\hbar^2 k_{||}^2}{2m_{h2}}$$

one can proceed as follows,

$$p = \sum_j 2 \times \int \frac{d^2\vec{k}_{||}}{(2\pi)^2} \left[1 - f\left(E_v(j,k_{||}) - E_f\right)\right]$$

$$= \sum_j \int_{E_{v2}-E_j}^{\infty} dE \left(\frac{m_{h2}}{\pi\hbar^2}\right) \left[1 - f(E - E_f)\right]$$

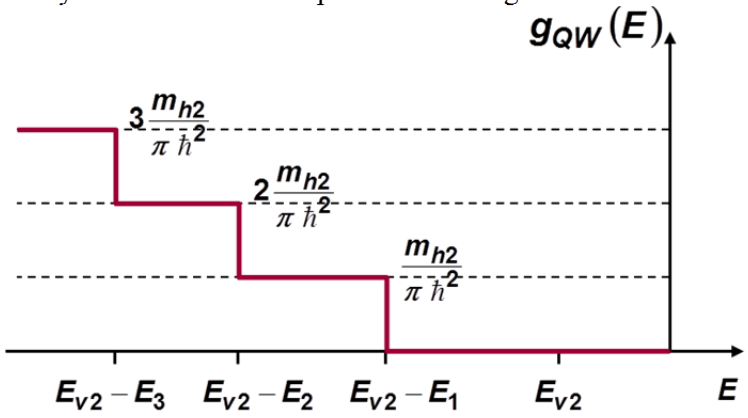
$$= \int_{-\infty}^{\infty} dE \sum_j \left(\frac{m_{h2}}{\pi\hbar^2}\right) \theta\left(E - E_{v2} + E_j\right) \left[1 - f(E - E_f)\right]$$

$$= \int_{-\infty}^{\infty} dE\, g_{QW}(E) \left[1 - f(E - E_f)\right]$$

where,

$$g_{QW}(E) = \sum_j \left(\frac{m_{h2}}{\pi\hbar^2}\right) \theta\left(E - E_{v2} + E_j\right)$$

The valence band density of states function is plotted in the Figure below.



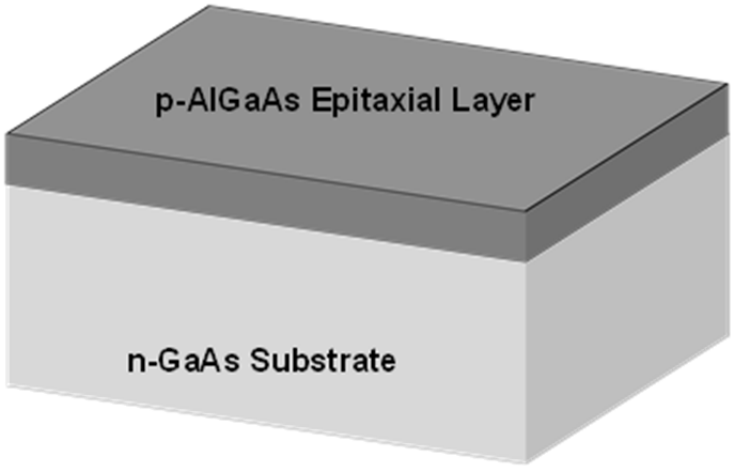
The 3D valence band density of states in a bulk semiconductor with energy dependence proportional to $\sqrt{E_{v2} - E}$ gets modified into the staircase structure shown above as a result of quantum confinement. Each confined energy level contributes to the density of states an amount equal to $m_{h2} / \pi \hbar^2$.

2.5 Heterostructures: Issues Related to Growth, Lattice Mismatch, and Strain

Compound semiconductor layers are usually grown by two methods:

i) MBE (Molecular Beam Epitaxy)
ii) MOCVD or MOVPE: (Metal-Organic Chemical Vapor Deposition) or (Metal-Organic Vapor Phase Epitaxy)

In both case are starts with a single crystal substrate (e.g. $GaAs$ or InP) and grows epitaxial Layers one by one on the substrate. The epitaxial layers are also doped (n-type or p-type) as they are grown. For example, to make a $n-GaAs/p-AlGaAs$ hetero junction, one can start with a n-doped $GaAs$ substrate and grow epitaxially a layer of $p-AlGaAs$ on top.



2.5.1 Lattice Constant Matching:

The interface is usually very sharp of abrupt (on atomic scale). Very good quality crystal material can be obtained by epitaxy provided the lattice constant a of the grown layer is identical to (or close to) the lattice constant a_{sub} of the substrate. This means that on a given substrate (of lattice constant a_{sub}) one may only grow additional layers that all have lattice constants close to a_{sub}. If the lattice constant of the grown layer is not exactly identical to the lattice constant a_{sub} of the substrate, then the grown layer stretches (if $a < a_{sub}$) or compresses ($a > a_{sub}$) so that its lattice constant is identical to the substrate lattice constant. Such coherently strained layers are called pseudomorphic. However, if the thickness h of the coherently strained layer exceeds a certain critical thickness h_c the coherent strain relaxes and this process generates crystal dislocations (crystal defects).

One way to understand the generation of dislocations is as follows. An elastically strained layer contains elastic energy (just like a stretched or compressed spring). A crystal dislocation line also contains energy. As the thickness of the coherently strained layer increases, its energy also increases and at some point its energy will become large enough that will be energetically favorable for the strain in the layer to decrease

slightly (relax) and give off the energy to a crystal dislocation line. Strain relaxation means that the layer is now no longer perfectly lattice matched to the substrate. Not surprising, the critical thickness h_c depends on the degree of lattice mismatch between the grown layer and the substrate. In almost all optical device applications one does not want strain relaxation (since the accompanying defects have a harmful effect on the device performance). So it is important that thickness of strained layers be kept below the critical thickness h_c. The value of the critical thickness is given by the Mathews-Blakeslee formula,

$$h_c = \frac{b}{4\pi f}\left[\frac{1-\nu\cos^2\theta}{(1+\nu)\cos\lambda}\right]\ln\left(\frac{h_c}{b}\right)$$

where, for diamond and zinc blende lattices b is related to the lattice constant, $b = a/\sqrt{2}$, ν is the Poisson ratio (a material constant related to the elasticity of the material), f is the strain in the layer and given by,

$$f = \left|\frac{a_{sub} - a}{a_{sub}}\right|$$

and the values of both the angles, θ and λ, are 60-degress for diamond and zinc blende lattices.

2.5.2 Strain Compensation:

What if one has many different strained layers in a stack with strains $f_1, f_2, f_3, \ldots$ and thickness $h_1, h_2, h_3, \cdots$, respectively. Mathews -Blakeslee relation can still be used to calculate the critical thickness of the stack provided it is understood that h_c is the critical thickness of the stack and f is the average strain defined as $f = (f_1h_1 + f_2h_2 + f_3h_3 + \cdots)/(h_1 + h_2 + h_3 + \cdots)$. This is particularly useful since in many cases the limitations imposed by the critical thickness are difficult to meet. In such cases, a technique called strain compensation is used. Alternate layers are chosen with opposite signs of the strain (i.e tensile – compressive – tensile – compressive -…….) so that the average strain remains close to zero. With this technique many layers with large strain (but zero average strain) can be grown on a substrate.