Lecture 12: pn junction breakdown and heterojunctions

Contents

1 Si pn junction

A pn junction acts as a rectifier since the current through the junction in forward bias is orders of magnitude larger than the reverse bias current. Consider a Si pn junction with N_A of 10^{17} cm^{-3} and N_D of 10^{16} cm^{-3}. Since the material is Si the intrinsic carrier concentration (n_i) at room temperature is 10^{10} cm^{-3}. When the pn junction is in equilibrium there is a built-in potential, given by

$$V_0 = \frac{k_B T}{e} \ln(\frac{N_A N_D}{n_i^2}) \tag{1}$$

For dopant concentrations given above, the value of V_0 is 0.78 V. Now consider the application of a forward bias potential of 0.5 V. This forward bias lowers the barrier between the p and n side and leads to an increased injection of minority carriers forming a current. This excess carrier injection is proportional to the applied forward bias voltage and is given by

$$p_n(0) = p_{p0} \exp[-\frac{e(V_0 - V_{ext})}{k_B T}] \tag{2}$$

Here, p_{p0} is the equilibrium concentration of holes in the p side (equal to N_A) and $p_n(0)$ is the excess hole concentration at the interface between the depletion region and n side. For a forward bias of 0.5 V, $p_n(0)$ is equal to 2.4×10^{12} cm^{-3}. This is more than 8 orders of magnitude higher than the equilibrium hole concentration in the n side (p_{n0}), which is 10^4 cm^{-3}. A similar calculation can be performed for excess electrons injected into the p side, using the equation

$$n_p(0) \ = \ n_{n0} \ \exp[-\frac{e(V_0 \ - \ V_{ext})}{k_B T}] \tag{3}$$

For a forward bias of 0.5 V this is equal to be 2.4×10^{11} cm^{-3}, which is again eight orders of magnitude higher than the equilibrium concentration $(10^3$ $cm^{-3})$!

These excess carriers diffuse through the p and n regions and ultimately recombine with the majority carriers. The diffusion lengths for these minority carriers are determined by their mobilities and the minority carrier lifetimes. The electron and hole mobility values for Si depend on the dopant concentration and are given by 120 $cm^2 V^{-1} s^{-1}$ and 440 $cm^2 V^{-1} s^{-1}$ respectively. The diffusivities can then be calculated from the mobilities using Einstein relation

$$
\begin{aligned}
D_e \ &= \ \frac{k_B T \mu_e}{e} \ = \ 3.10 \ cm^2 s^{-1} \\
D_h \ &= \ \frac{k_B T \mu_h}{e} \ = \ 11.39 \ cm^2 s^{-1}
\end{aligned}
\tag{4}
$$

The diffusion lengths can then be calculated from the diffusivities and the minority carrier lifetime.

$$
\begin{aligned}
L_h \ &= \ \sqrt{D_h \ \tau_h} \\
L_e \ &= \ \sqrt{D_e \ \tau_e}
\end{aligned}
\tag{5}
$$

The minority carrier lifetimes depend on the carrier type and also on the dopant concentration. For this particular pn junction τ_h is 417 ns and τ_e is 5 ns. So the hole and electron diffusion lengths, calculated using equation 5, are 21.8 μm and 1.24 μm. A device whose dimensions are longer than the diffusion lengths is a long diode. For this long diode, the diffusion current in forward bias is given by

$$
\begin{aligned}
J_D \ &= \ J_{s0} \ [\exp(\frac{eV}{k_B T}) \ - \ 1] \\
J_{s0} \ &= \ e \ n_i^2 (\frac{D_h}{L_h N_D} \ + \ \frac{D_e}{L_h N_A})
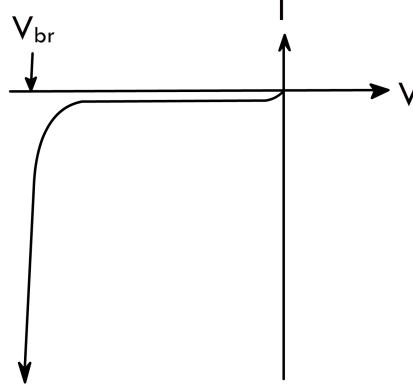\end{aligned}
\tag{6}
$$

Figure 1: Reverse bias I-V characteristics of a *pn* junction. Typical reverse bias current is very low until a certain voltage, when the current increases exponentially. This voltage is called the breakdown voltage and can be a reversible or irreversible process. Adapted from *Principles of Electronic Materials - S.O. Kasap.*

J_{s0} is the reverse saturation current and is calculated to be 1.23×10^{-11} Acm^{-2}. This is the current in reverse bias. For a forward bias of 0.5 V, the current is 3.03×10^{-3} Acm^{-2}. This is more than *8 orders* of magnitude higher than the reverse bias current, making a *pn* junction a rectifier.

2 Junction breakdown

Consider the I-V characteristics of a *pn* junction in reverse bias, as shown in figure 1. There is initially a small reverse saturation current due to thermally generated electron and holes in the depletion region. This current is called *drift current*, since this is due to movement of the thermally generated carriers under the applied electric field. With increase in voltage there is a particular value, called the **breakdown voltage**, beyond which the current increases rapidly. This is called **junction breakdown**. There are two main mechanisms of junction breakdown, depending on the dopant concentration levels.
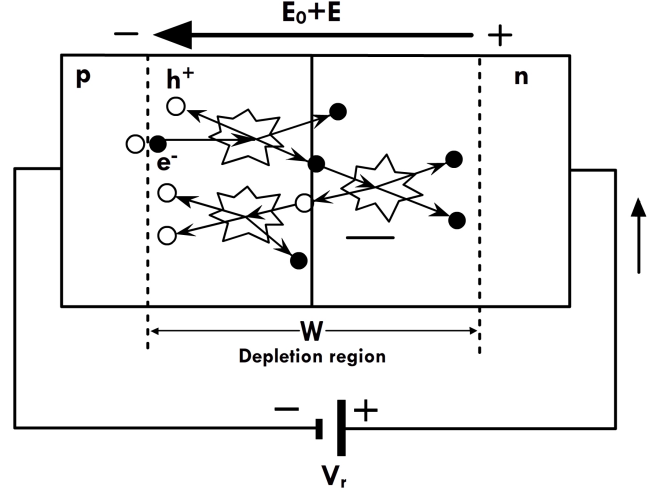
Figure 2: Avalanche breakdown in a lightly doped *pn* junction. Thermally generated carriers in the depletion region get accelerated by the electric field and cause ionization in other Si atoms producing a cascade of electrons and a rapidly increasing current. Adapted from *Principles of Electronic Materials - S.O. Kasap.*

2.1 Avalanche breakdown

Avalanche breakdown occurs in moderately and lightly doped *pn* junctions with a wide depletion region. The schematic of the process is shown in figure 2. Electron hole pairs thermally generated in the depletion region are accelerated by the external reverse bias. Electrons are accelerated towards the n side and holes towards the p side. These electron can interact with other Si atoms and if they have sufficient energy can knock off electrons from these Si atoms. This process is called **impact ionization** and leads to production of a large number of electrons. This causes the rapid rise in current. The breakdown voltage decreases with increase in dopant concentration, as shown in figure 3. The large current during breakdown can be modeled by a multiplication factor (M) that relates the current to the breakdown voltage

$$J = M J_{s0}$$
$$M = \frac{1}{1 - (\frac{V_r}{V_{br}})^n} \tag{7}$$

where J_{s0} is the reverse saturation current, V_r is the applied voltage and V_{br} is the breakdown voltage. n is a constant of value between 3-5. Closer the
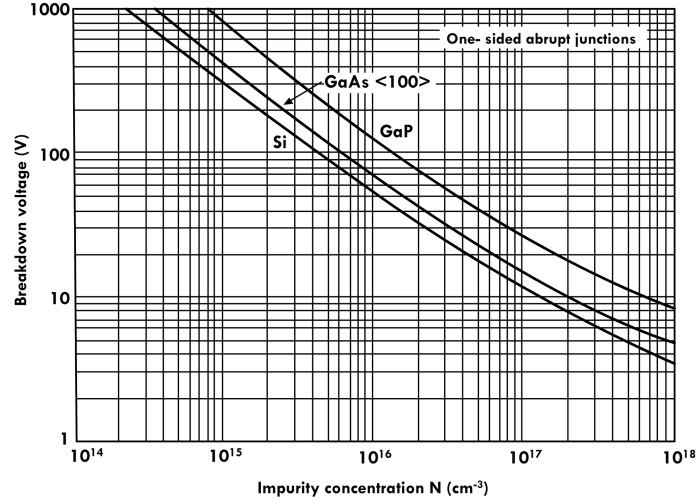
Figure 3: Avalanche breakdown voltage as a function of dopant concentration for different semiconductors. With higher dopant concentrations the mechanism changes to a Zener breakdown process. Adapted from *Physics of semiconductor devices - S.M. Sze.*

value of V_r to V_{br}, higher is the reverse bias current. This equation is valid only for values of V_r smaller than V_{br}.

2.2 Zener breakdown

With increase in doping concentration the breakdown mechanism, changes from Avalanche to a tunneling mechanism. This is called a Zener breakdown. This is because the depletion width decreases with dopant concentration. Also, the reverse bias causes an offset in the bands such that it is possible for carriers to tunnel across the narrow depletion region. This tunneling process is shown schematically in figure 4, where electrons tunnel from the valence band on the p side to the conduction band on the n side, driven by the externally applied reverse bias. Tunneling also leads to a large increase in current. The transition from avalanche to Zener as the primary breakdown mechanism with dopant concentration is shown in figure 5. Zener diodes are primarily used as surge protectors in circuits, since there is a rapid increase in current with a small change in voltage. Prior to breakdown there is a high resistance (small reverse saturation current) but after breakdown the resistance is very small. This can be used as voltage regulators in circuits.
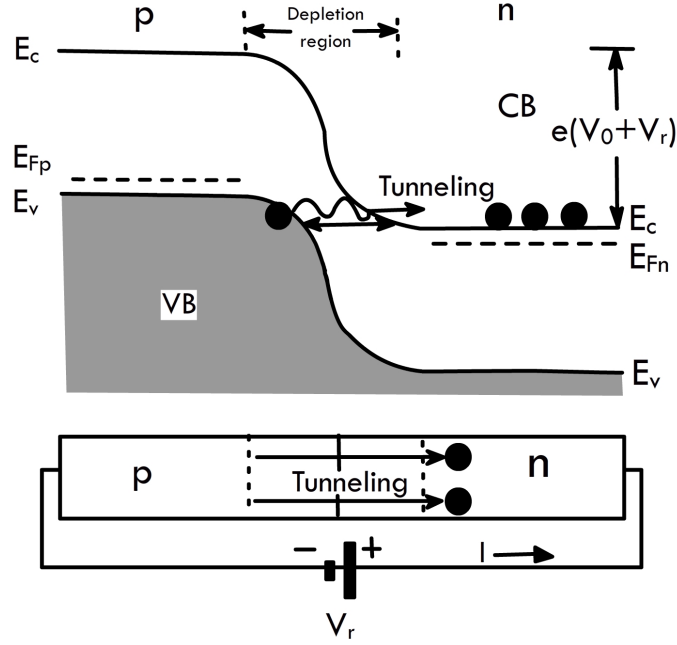
Figure 4: Tunneling of electrons from p to n side in a heavily doped pn junction under reverse bias. The bias causes band bending such that the valence band on the p side is aligned with the conduction band on the n side. Adapted from *Principles of Electronic Materials - S.O. Kasap.*
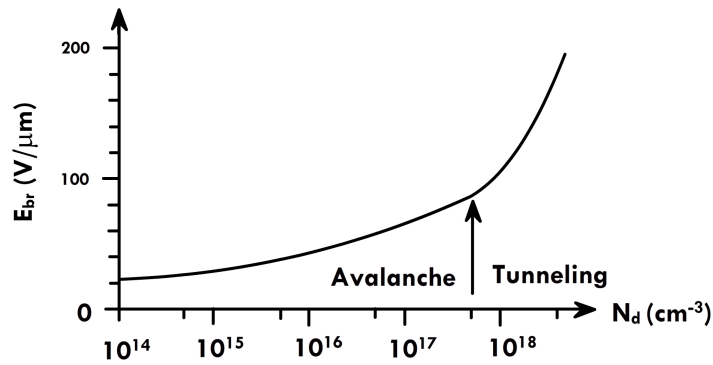

Figure 5: Breakdown field vs. dopant concentration. At lower concentrations the breakdown is due to avalanche mechanism, which shifts to Zener mechanism at higher values. Adapted from *Principles of Electronic Materials - S.O. Kasap.*

3 Heterojunctions

Typical *pn* junctions are formed between *p* and *n* types of the same material, with the only difference being the dopant type and concentration. These junctions are called **homojunctions** and they can be grown by *compensation doping* of certain regions in a semiconductor. It is also possible to form *pn* junctions between different semiconductors. These are called **heterojunctions**. In such systems there is also a chemical change across the interface along with the change in dopant type and concentration. Because of the presence of different materials with different band gaps, these can lead to localized quantized energy states at the interface during band alignment.

3.1 Lattice mismatch

Heterojunctions are characterized by a lattice mismatch between the two semiconductors. This is important from a growth perspective since it restricts the types and thickness of materials that can be grown together. The primary consequence of lattice mismatch is that this will lead to *defects at the interface* which can act as traps for the carriers. These can affect the electrical properties of the device. The lattice mismatch also dictates the maximum defect-free layer thickness that can be grown. For epitaxial growth, the maximum layer thickness that can be grown is *inversely proportional* to the lattice mismatch. If a_s and a_e are the lattice constants of the substrate and the epitaxially grown layer then the lattice mismatch (Δ) is defined as

$$\Delta = \frac{|a_e - a_s|}{a_e} \tag{8}$$

Given this lattice mismatch, the critical thickness (t_c) of the epitaxial layer that can be grown defect free is approximately given by

$$\boxed{t_c \approx \frac{a_e}{2\Delta} \approx \frac{a_e^2}{2|a_e - a_s|}} \tag{9}$$

This leads to the formation of a *strained heteroepitaxial layer*. As thickness increases beyond the critical thickness the strain is relieved by formation of dislocations at the interface. Consider the example of a SiGe heterojunction. The lattice constants are 5.431 $\mathring{A}$ and 5.658 $\mathring{A}$ for Si and Ge, respectively. For Ge on Si, this leads to a lattice mismatch (Δ) of 0.04 (4%). The critical layer thickness can be calculated approximately using equation 9 and gives a value of t_c of 7.07 nm. This is approximately 15 atomic layers of Ge. SiGe is complicated by the fact that Ge films breaks down into islands before
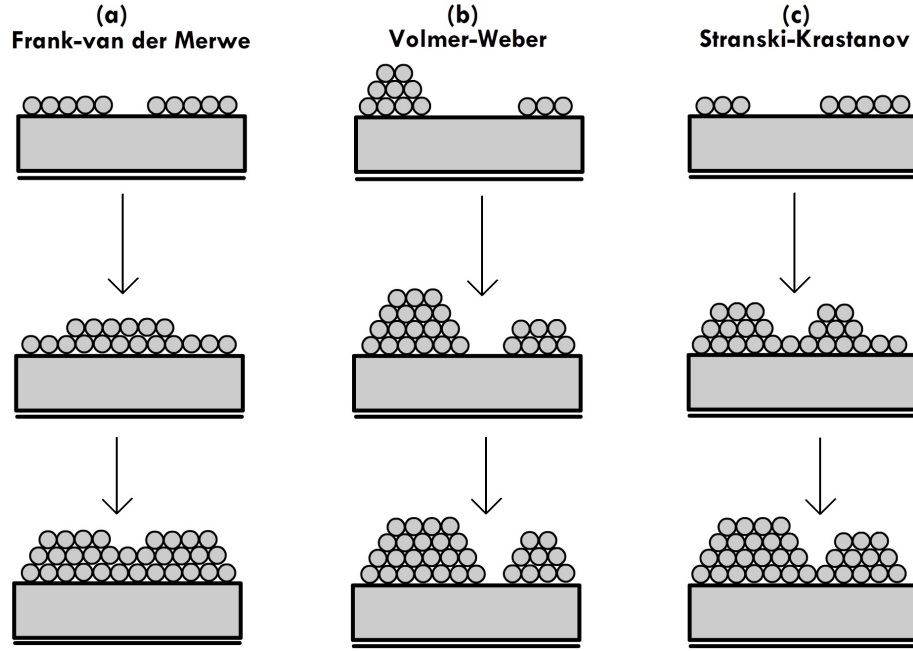
7

Figure 6: Three commonly founds modes of thin film growth. (a) Growth occurs layer by layer. (b) Growth occurs directly in the form of islands. (c) Growth occurs due to a mixture of layers and then islands. Adapted from *http://www.webzeest.com/article/511/thin-film-growth*

the critical thickness is reached. This is related to the difference in the free energies of the two materials. Based on the free energy difference there are generally three growth modes

1. Island formation or Volmer-Weber growth mode

2. Layer-by-layer or Frank van de Merwe growth mode

3. Layer-island or Stranski Krastanov growth mode

This is summarized in figure 6. Epitaxial growth is normally by layer-by-layer or Frank van de Merwe mode.

Heterojunctions can be formed by substitutional doping of the substrate. The lattice mismatch in this case is controlled by the amount of doping. This also controls other parameters like the band gap and electrical properties like mobility. In substitutional doping, lattice mismatch is smaller than when we have two dissimilar layers, so that thicker films can be grown. Consider the heterojunction formed between GaAs substrate and $Al_xGa_{1-x}As$. $x = 0$ corresponds to GaAs with a band gap of 1.42 eV and lattice constant of 5.65
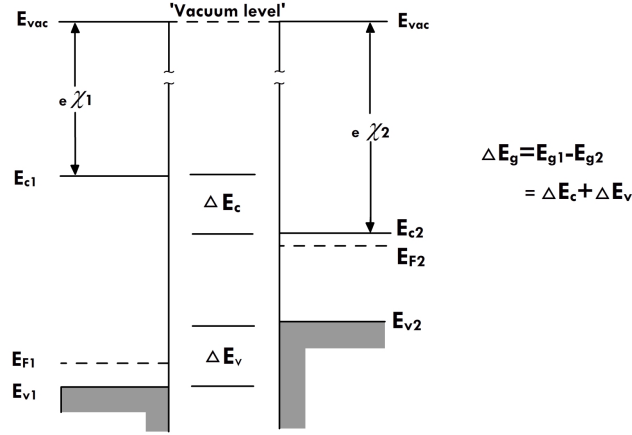
Figure 7: Energy band diagram of two dissimilar materials before junction formation. Before joining the vacuum levels are aligned but the electron affinity and work functions are different so that the valence and conduction bands are not aligned. Adapted from *Solid state electronic devices - Streetman and Banerjee.*

$Å$. $x = 1$ corresponds to AlAs with a larger band gap value of 2.17 eV and a lattice constant of 5.66 $Å$. The lattice mismatch in this case is less than 0.2% so that t_c is much larger than the SiGe case. The band gap is a smooth function between GaAs and AlAs for different values of x.

3.2 Band gap alignment

Consider a pn junction formed between two dissimilar materials. The junction is ideal, so there are no defects at the interface. Achieving this practically is hard and not possible for all combinations of materials. The band picture of the two materials before the junction is formed is shown in figure 7. In this particular case the band gap of the p type material is higher than the band gap of the n type material. Correspondingly, the electron affinity (distance of conduction band from the vacuum level) of the n type material is higher than the p type material. There is also an energy difference between the valence bands and the conduction bands, depicted as ΔE_c and ΔE_v in figure 7.

When the two materials are brought together the Fermi lines must line up at equilibrium (no external bias). This causes a band bending in the p and n side, as shown in figure 8. To understand the formation of the band alignment the process of drawing the band alignment can be broken into in-
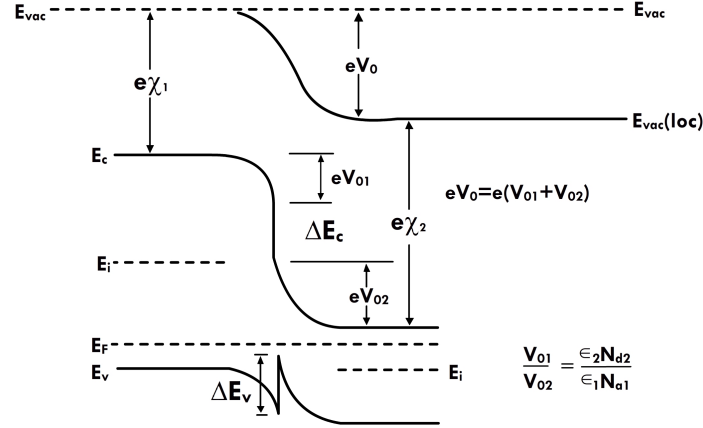
Figure 8: Energy band diagram of two dissimilar semiconductors at equilibrium. The Fermi levels line up but the energy difference between the two valence bands and conduction bands is maintained. Adapted from *Solid state electronic devices - Streetman and Banerjee.*

dividual steps. The same procedure can be adopted for homojunctions, but the arguments there are simpler since $\Delta E_v = \Delta E_c = 0$.

1. Align the Fermi levels. Also realize that far away from the depletion region the band picture will be similar to figure 7, i.e. the individual semiconductors. This is shown in figure 9.

2. The depletion region will be wider in the lightly doped region. This can be used to specify the junction interface between the p and n sides. This will also help to determine the curvature of the band bending. At the junction, the difference in energies between the two valence and conduction bands, ΔE_0 and ΔE_1, in figure 10, must be maintained.

3. In a pn junction electrons are transferred from the n to the p side (vice versa for the holes). This leads to an electric field going from n to p (field goes from positive to negative). So bands bend up from n to p and down from p to n, illustrated in figure 11.

4. When the discontinuities are joined the final band diagram is obtained. The sequence of steps and the final band alignment is summarized in figure 12.

The series of steps in figure 12 show that finally a quantum well like situation is manifest in the conduction band, while there is an energy discontinuity in
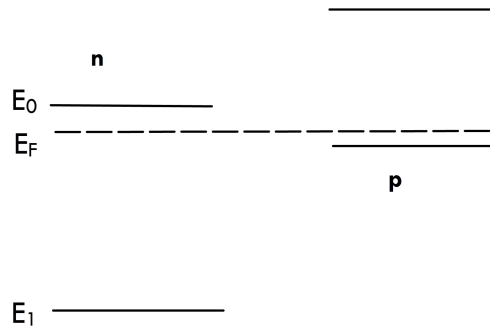
Figure 9: First step in drawing the band diagram of a heterojunction. The Fermi levels are aligned and far away from the depletion region the individual p and n junctions are marked. Adapted from *Solid state electronic devices - Streetman and Banerjee.*
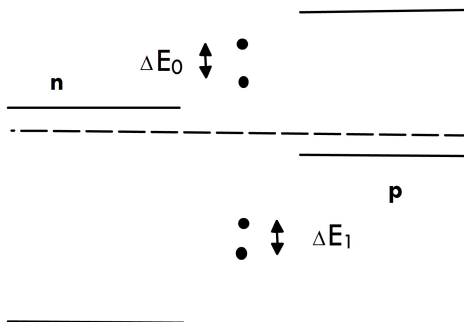

Figure 10: Second step in drawing the band diagram of a heterojunction. The junction location is marked, depending on the dopant concentration. The amount of band bending can be calculated and the energy differences between the valence and conduction bands are marked. Adapted from *Solid state electronic devices - Streetman and Banerjee.*
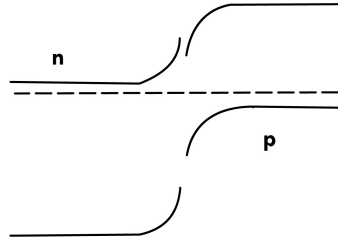
Figure 11: Third step in drawing the band diagram of a heterojunction showing the band bending. Bands bend up in the direction of the field, from n to p, and down from p to n. The energy difference at the junction is still maintained Adapted from *Solid state electronic devices - Streetman and Banerjee.*

the valence band. It is possible to reverse this situation by choosing semiconductors with the appropriate band gap values, as summarized in figure 13. The difference between a heterojunction and a homojunction is that the difference in band gaps leads to the formation of a energy minimum (either for electron or holes) at the junction. A close-up of this energy minimum is shown for a junction formed between n-AlGaAs and GaAs, in figure 14. For sufficiently thin interfaces this energy minimum can be quantized so that there are distinct energy levels. This is especially useful for optical transitions since quantized levels have smaller energy spreads. The energy minimum at the interface can also lead to carrier accumulation of specific polarity. This will also affect carrier transport properties across the junction.
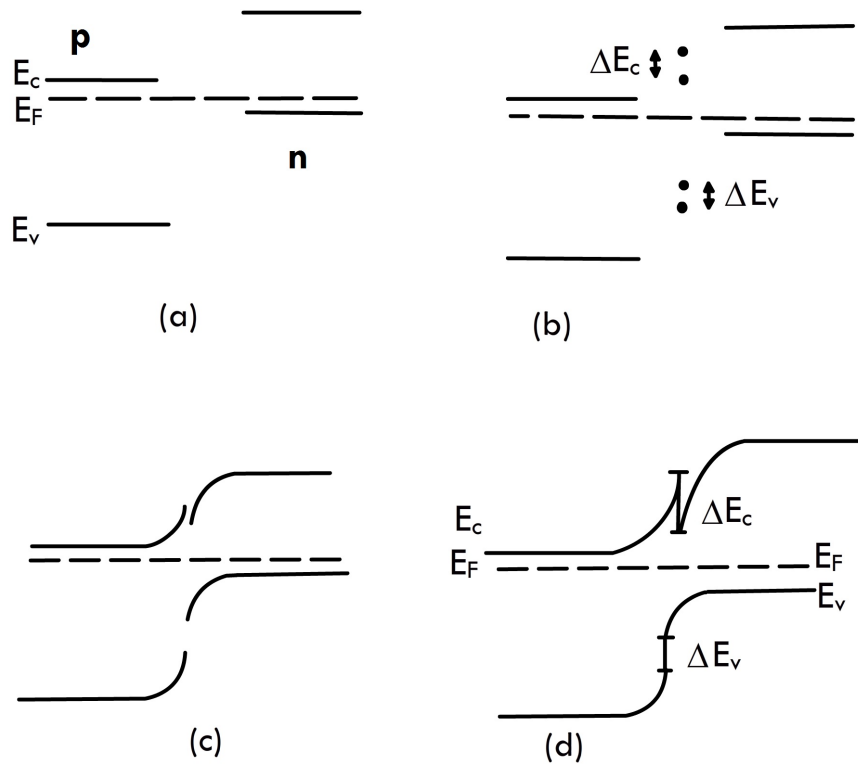
Figure 12: (a)-(d) Steps in drawing a heterojunction staring from two different semiconductors. This is the summary of figures 9, 10, and 11. There is an energy discontinuity in the valence band while a quantum well is formed in the conduction band. Adapted from *Solid state electronic devices - Streetman and Banerjee.*
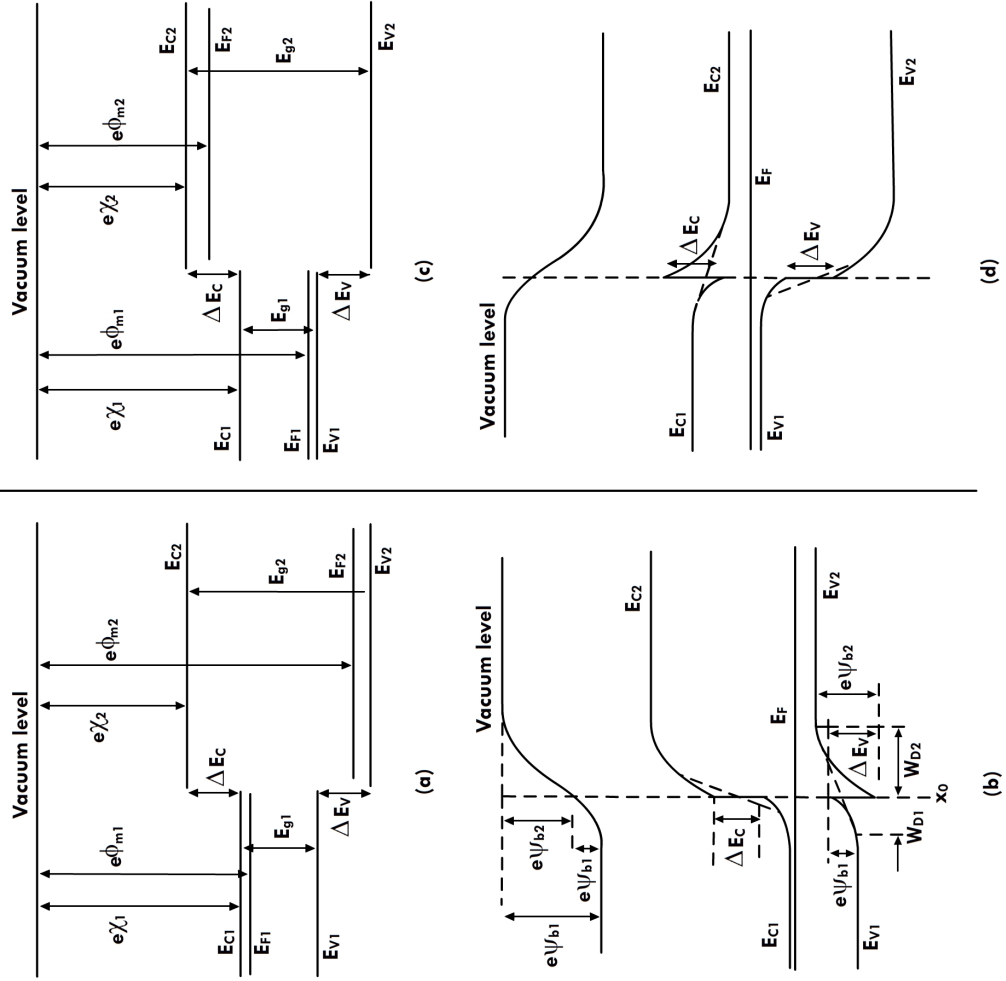
Figure 13: Summary of band bending in a heterojunction. (a)-(b) Initial and final alignment when the p side has a higher band gap than the n side. (c)-(d) Band alignment when the n side has a higher band gap than the p side. Adapted from *Physics of semiconductor devices - S.M. Sze.*
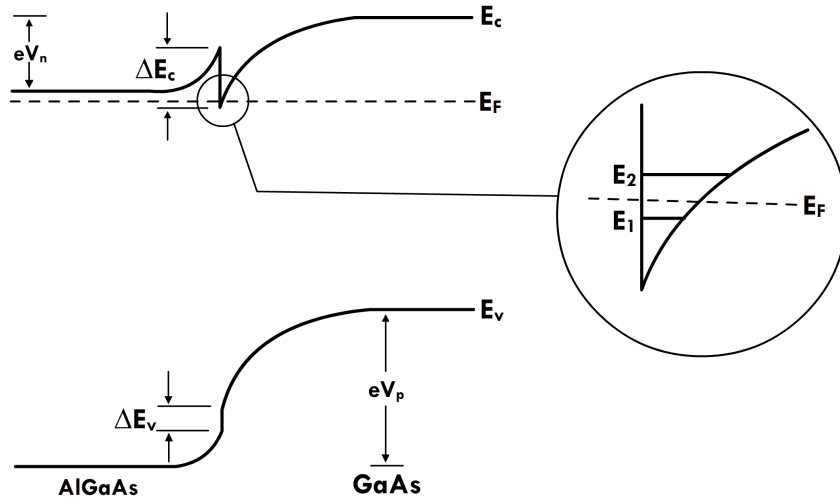
Figure 14: Energy minimum at the interface in the conduction band for a AlGaAs and GaAs heterojunction. When the depletion width is small, this energy minimum region is quantized, with specific discrete energy levels. Transitions from these discrete energy levels are sharp and can be used for opto-electronic devices. Adapted from *Solid state electronic devices - Streetman and Banerjee.*