# Section 3: PN Junctions and diodes

The simplest semiconductor device is a PN junction, this involves a p-type semiconductor connected to an n-type semiconductor. Consider a PN junction with doping $N_a$ on the p-side, $N_d$ on the n-side.

## Depletion zone - built in voltage

p - side has an excess of holes, with negatively charged acceptors to cancel this charge out
n - side has an excess of electrons, with positively charged donors to cancel this charge out

When the p - side and n - side are brought together, the holes and electrons move across the join and combine. This leaves behind bare acceptors and donors, with their negative and positive charges. The p - side becomes negatively charged, the n - side becomes positively charged. This creates an electric field from the n - side towards the p - side. This electric field opposes the flow of holes and electrons across the junction, resulting in a region of fixed width with a depletion of holes and electrons. This region is called the depletion zone.

There is a voltage produced across the depletion zone, we call this the built in voltage $V_{\text{bi}}$. The n-type side is at a higher potential than the p-type side. To calculate $V_{\text{bi}}$ we consider the equilibrium situation when the diffusion current of holes across from the p-side is equal to the drift current of holes due to the electric field.

So we have $J_p = {J_p}^{\text{drift}} + {J_p}^{\text{diff}} = 0$

$$\therefore e\,\mu_p\,p(x)\,E(x) - e\,D_p\,\frac{dp}{dx} = 0$$

$$\therefore \frac{\mu_p}{D_p}E = \frac{1}{p}\frac{dp}{dx}$$

Using the Einstein relations and $E = -\frac{dV}{dx}$, we get $\frac{-q}{kT}\frac{dV}{dx} = \frac{1}{p}\frac{dp}{dx}$. Integrating this from $-\infty$ to $\infty$ we get:

$$V_{\text{bi}} = \frac{kT}{e}\ln\left(\frac{p_p}{p_n}\right)$$

Where $p_p$ is the number of holes on the p-side, $p_n$ is the number of holes on the n-side. The same argument can be performed for the electrons on the n-side to give $V_{\text{bi}} = \frac{kT}{e}\ln\left(\frac{n_n}{n_p}\right)$. Sometimes $kT/e$ is called the thermal voltage $V_T$. So for holes we have $p_p = N_a$, $p_n = {n_i}^2 / N_d$ and so

$$V_{\text{bi}} = V_T\ln\left(\frac{N_a N_d}{{n_i}^2}\right) = V_T\ln\left(\frac{N_a N_d}{N_v N_c}\right) + \frac{E_g}{e}$$

As doping increases so does $V_{\text{bi}}$.

## Depletion zone - width

We expect the total charge of the depletion zone to be zero, otherwise electrons or holes will be attracted towards it. Assuming that we have a width $x_p$ into the p-type, a width $x_n$ into the n-type and that the acceptors and donors in this region are fully ionised this gives us: $x_p N_a = x_n N_d$.
If the doping in the n-side is large $N_d >> \text{N}_a$, this means $x_p >> \text{x}_n$, so large doping on one side relative to the other increases the depletion width on the other side.

To calculate the total depletion width $w = x_p + x_n$, we need to solve the Poisson equation in one dimension across the junction. Doing so results in the following width:

$$w = \sqrt{\frac{2\epsilon}{q} \frac{(N_a + N_d)}{N_a N_d} V_{\text{bi}}}$$

## Biased PN Junctions - width

Forward biasing is when the p-type is connected to a positive terminal, and the n-type to the negative terminal. This forward biasing $V_{\text{bias}}$ reduces the potential difference between the p and n sides. We will take $V_{\text{bias}} < V_{\text{bi}}$ otherwise we require a different type of analysis to proceed. The width of the depletion zone shrinks as less charge is required to sustain the potential difference.

$$w = \sqrt{\frac{2\epsilon}{q} \frac{(N_a + N_d)}{N_a N_d} (V_{\text{bi}} - V_{\text{bias}})}$$

The decreased potential also means the drift current is smaller than the diffusion current, there is a net positive current of holes across the junction.

Reverse bias is when the voltage is connected the other way around, the depletion zone becomes wider, with effectively $V_{\text{bias}} < 0$.

## Biased PN Junctions - ideal diode equation

In an unbiased PN junction there is no net current. This means the diffusion currents from the majority sides are almost totally blocked by the built-in voltage, so that they equal the diffusion current from the minority side. If we define a diffusion lengths $L_p = \sqrt{D_p \tau_p}$ and $L_n = \sqrt{D_n \tau_n}$ where $\tau_p$ and $\tau_n$ are the lifetimes of holes and electrons before they re-combine, then $L_p$ and $L_n$ give us length scales over which the diffusion occurs.

If we consider holes to begin with, there is a diffusion current of holes from the n-side to the p-side, any holes that do diffuse into the depletion zone will be carried into the p-side by the potential difference. We can work this out:

$$\overleftarrow{J_p} = q F_p = -e D_p \frac{\text{dp}}{\text{dx}} = -e D_p \frac{p_n}{L_p} = -e \frac{D_p n_i^2}{L_p N_d}$$

Similarly for electrons diffusing from the minority p-side to the majority n-side:

$$\overrightarrow{J_n} = q F_n = e D_n \frac{\text{dn}}{\text{dx}} = -e D_n \frac{n_p}{L_n} = -e \frac{D_n n_i^2}{L_n N_a}$$

The sum of these gives us the saturation current $J_s$.

$$-\left(\overleftarrow{J_p} + \overrightarrow{J_n}\right) = J_s = e\left(\frac{D_p}{L_p} p_n + \frac{D_n}{L_n} n_p\right) = e\left(\frac{D_p n_i^2}{L_p N_d} + \frac{D_n n_i^2}{L_n N_a}\right).$$

Now for the majority carriers, these have much larger diffusion currents, which are opposed by the built in voltage, until they equal the diffusion currents of the minority carriers. So we know that with no bias for holes:

$$\overrightarrow{J_p} + \overleftarrow{J_p} = 0, \quad \therefore \ \overrightarrow{J_p} = -\overleftarrow{J_p} = D_p \frac{p_n}{L_p}$$

We can re-arrange the expression for the built-in voltage to get: $p_n = e^{\frac{-e V_{\text{bi}}}{k T}} p_p$, subbing this to our expression for $\overrightarrow{J_p}$ we have $\overrightarrow{J_p} = D_p \frac{p_p}{L_p} e^{\frac{-e V_{\text{bi}}}{k T}}$. The factor of $e^{\frac{-e V_{\text{bi}}}{k T}}$ represents the built-in voltage $V_{\text{bi}}$ cutting down the forward majority diffusion current $\overrightarrow{J_p}$. If we now apply a forward bias $V_{\text{bias}}$ then this becomes

$$\overrightarrow{J_p} = D_p \frac{p_p}{L_p} e^{\frac{-e(V_{\text{bi}}-V_{\text{bias}})}{kT}} = D_p \frac{p_p}{L_p} e^{\frac{-e V_{\text{bi}}}{kT}} e^{\frac{e V_{\text{bias}}}{kT}} = \left(-\overleftarrow{J_p}\right) e^{\frac{e V_{\text{bias}}}{kT}}$$

For electrons this is: $\overleftarrow{J_n} = \left(-\overrightarrow{J_n}\right) e^{\frac{e V_{\text{bias}}}{kT}}$. In total we have

$$J = \overrightarrow{J_p} + \overleftarrow{J_n} + \overleftarrow{J_p} + \overrightarrow{J_n} = \left(-\overleftarrow{J_p}\right) e^{\frac{e V_{\text{bias}}}{kT}} + \left(-\overrightarrow{J_n}\right) e^{\frac{e V_{\text{bias}}}{kT}} + \overleftarrow{J_p} + \overrightarrow{J_n} = -\left(e^{\frac{e V_{\text{bias}}}{kT}} - 1\right)\left(\overleftarrow{J_p} + \overrightarrow{J_n}\right) = \left(e^{\frac{e V_{\text{bias}}}{kT}} - 1\right) J_s$$

This is the ideal diode equation: $J = \left(e^{\frac{V_{\text{bias}}}{V_T}} - 1\right) J_s$. Multiplying by the cross-sectional area we get it in terms of current: $I = \left(e^{\frac{V_{\text{bias}}}{V_T}} - 1\right) I_s$. The saturation current is $I_s = A\, e\left(\frac{D_p n_i^2}{L_p N_d} + \frac{D_n n_i^2}{L_n N_a}\right)$. $V_T = kT/e$ is the thermal voltage, at 300K, $V_T = 25.85$ mV.

Be careful of units when calculating $V_T$, Volts are J/C. If you use $k$ in eV/K, then you need to multiply by $e = 1.6 \times 10^{-19}$ J/eV , this cancels with the $e$ in the denominator, so $V_T = 8.62 \times 10^{-5}\, T$ Volts.

## Biased PN Junctions - current

We can see from the ideal diode equation that a forward bias produces an exponential increase in forward current. The voltage across the diode is reduced, allowing a greater diffusion current of holes from the p-side to the n-side, and electrons from the n-side to the p-side. If the bias is increased to the point where it is larger than the built in voltage (on the order of 1 Volt, dependent on doping and temperature), then the depletion zone vanishes, after this the current trends towards an ohm's law relationship with voltage.

On the other hand if a reverse bias is applied then the forward (p to n) diffusion of holes and backward (n to p) diffusion of electrons is cut off sharply, this leaves just the saturation current $J = -J_s$, which is independent of voltage. The increased potential across the diode does not increase the saturation current, as any holes diffusing from the n-side into the depletion zone were always carried across to the p-side, similarly for electrons diffusing from the p-side to the n-side.

Thus the diode acts as you would expect, electric current is allowed to pass in the forward direction (positive bias), and is almost totally blocked in the reverse direction (negative bias). If a significant forward voltage is applied, then the diode acts like an ohmic conductor.

## Light emitting diodes (LEDs)

If a significant forward voltage is applied to a PN junction, then the depletion zone collapses so holes and electrons travel freely across the diode. If these holes and electrons meet they combine with the electron dropping from the conduction band into the valence band, producing a photon of energy equal to the band gap energy. This process is more efficient in direct semiconductors.

This process can work in reverse for photo-diodes. This is a PN junction without a strong forward bias, usually with either no bias or a reverse bias. If a photon with energy equal to or greater than the band gap energy collides with a valence electron in the depletion zone it produces a electron/-hole pair. This pair is quickly split up by electric field in the depletion zone and produces a small current. This current can be amplified by making use of a strong reverse bias and what is known as avalanche breakdown.

## Avalanche breakdown

The steady state voltage for reverse bias $J = -J_s$ only applies up until a certain reverse voltage. After this point the electric field in the depletion zone becomes so large that any electrons or holes are accelerated fast enough that they create electron/hole pairs by colliding with valence electrons. These pairs can then go on to collide and create more pairs resulting in an avalanche effect, where the diode starts to conduct rapidly. If the circuit in which the diode is placed allows for a large amount of reverse current to go through the diode, this can suddenly generate a lot of current, heat and damage to the diode.

In a properly designed circuit this effect can be useful, for instance in photo-diodes a single photon can be multiplied by an avalanche breakdown to produce a more significant current. This allows photo-diodes to be very sensitive. The circuit prevents the diodes being damaged by limiting the current that can flow, while maintaining the heavy reverse bias.

## Transistors

Transistors typically consist of either a PNP or NPN junctions. Let us consider a PNP type, the first p-type is called the emitter, the middle n-type is the base and the final p-type is the collector. The n-type is very thin, so holes and electrons diffuse across it quickly. We take the emitter to be at $V_e = 0$. We then have the base connected to a slight negative voltage, $V_b < 0$ and the collector connected to a significant negative voltage $V_c << \mathrm{V}_b$.

Let us consider the movement of holes across this transistor. Due to the negative voltage on the base, this acts like a forward bias in a PN junction, there is a significant current from the emitter to the base. Rather than the current all exiting through the base however, the fact that the n-type region is so thin means that the majority of the holes will diffuse into the final p-type region, the collector. There is a tiny current through the base and most of the current flows in from the emitter and out through the collector.

If we now vary the voltage $V_b$ at the base we cause a large change in the current that makes it from the emitter into the base. The fraction of this current that then moves into the collector does not change greatly due to the large negative potential on the collector. Thus as we vary $V_b$ we cause a large change in the current that flows in from the emitter and out from the collector. The transistor acts like an amplifier, a small current introduced into the base produces a large current, 100 times or so higher at the collector.

In this type of transistor we wish to minimise the electron current, this can be done by minimising the doping in the n-type base, by ensuring it is lower than the doping in the p-type regions. This results in the electron current being much smaller than the hole current. What we have described here is a Field Effect Transistor (FET), which operates based on varying voltages at the base and makes use of a single carrier type.