EE3310 Class notes – Part 2

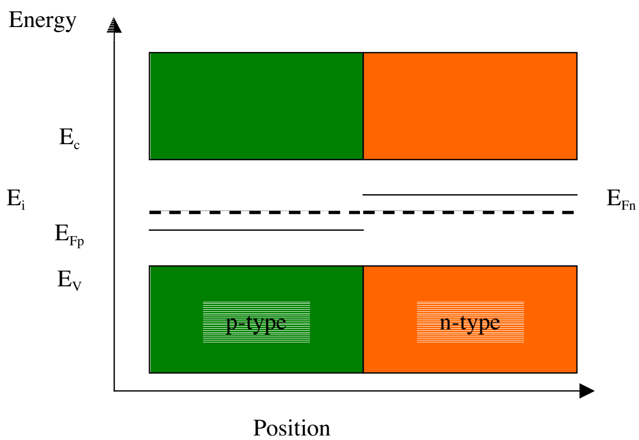
Version: Fall 2002
These class notes were originally based on the handwritten notes of Larry Overzet. It is expected that they will be modified (improved?) as time goes on. This version was typed up by Matthew Goeckner.

Solid State Electronic Devices - EE3310

Class notes

p-n junctions

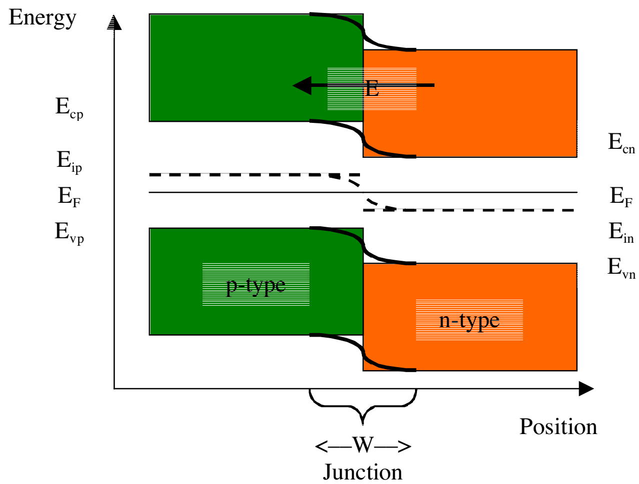
Up to this point, we have looked at single materials that are not touching anything else. Such a system does not make a 'device'. We now want to put a p-type and an n-type material together and see what happens. **This is our first true device!**



What will happen?

1) Holes in the p-type will diffuse into the n-type (because of the density gradient).
2) Electrons in the n-type will diffuse into the p-type (because of the density gradient).
3) The Fermi levels in the two materials have to match as we are in equilibrium

=> There will be an electric field that forms between the materials to OPPOSE diffusion and also so that the Fermi level across the device is constant.



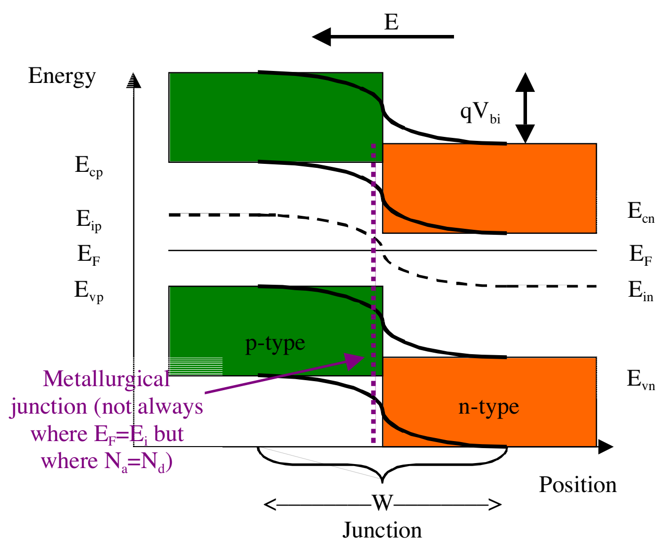
We know that this shift in the intrinsic energy level must correspond to an electric field. Further, we know that the change in density of the charge carriers will result in diffusion across the boundary. Thus, we find that the particles will initially move around in the junction and this will result in the formation of an electric field across the boundary. We can look at what each of the 'movements' do…

Particle flux density, Γ	Mechanism	Particle current density, J
<—	Hole drift (**E**)	<—
—>	Diffusion	—>
—>	Electron drift (**E**)	<—
<—	Diffusion	—>

Experimentally, we usually do not have a system in which we simply place a piece of n-type material on top of a piece of p-type. (Such a system can be made but it is very rarely used because of the cost.) Rather what is typically done is that the semiconductor is doped, via ion implantation, in one area to be p-type and then it is doped to be n-type in an adjacent location. (For Si technology, B is the most common n-type dopant while As is the most common p-type dopant.) Ion implantation is an in-exact process, for a number of physical reasons, and thus we cannot get 'perfectly' abrupt (step) dopant change. (These reasons are discussed in other classes and are not particularly important here.) The best possible is known as 'ultra shallow junctions' and is the result of very low energy ion implantation and very special 'activation' steps. On the other hand, we can certainly get the dopant to change 'slowly' or 'graded' in the area of the junction, through high energy implants and/or long (time) activation steps.

The upshot of the above discussion is that we can have 'step' junctions as well as 'graded' junctions. Both are important in the production of electronic devices. For the purposes of the class, we will only consider step junctions.

Let us go back and look at our picture of the junction.



(Note that the junction is also known as the 'Depletion Region' and the 'Transition Region'.)

Looking at the figure, begs the question, what are W, V_{bi} and **E**?

To get at this, lets us look at the areas outside of the junction. First, we know the location of the Fermi energy with respect to the intrinsic energy on both sides of the junction.

$$n_0 = n_i \exp\left[\frac{(\mathcal{E}_F - \mathcal{E}_i)}{kT}\right]$$

$$p_0 = n_i \exp\left[\frac{-(\mathcal{E}_F - \mathcal{E}_i)}{kT}\right]$$

Thus,

$$\Delta\mathcal{E}_{Fn} = (\mathcal{E}_{Fn} - \mathcal{E}_i) = kT\ln\left(\frac{n(x \ge x_{n0})}{n_i}\right)$$

$$\Delta\mathcal{E}_{Fp} = (\mathcal{E}_i - \mathcal{E}_{Fp}) = kT\ln\left(\frac{p(x \le -x_{p0})}{n_i}\right)$$

Where we have defined x_{n0} and $-x_{p0}$ as the edge of the junction on the n side and p side respectively. When the Fermi energy on each side shifts such that

$$\mathcal{E}_{Fp} = \mathcal{E}_{Fn}$$

then the intrinsic energy must shift on each side. This gives rise to the voltage shift across the junction.

$$\Downarrow$$

$$qV_{bi} = \Delta\mathcal{E}_{Fn} + \Delta\mathcal{E}_{Fp}$$

$$= kT\ln\left(\frac{n(x_{n0})}{n_i}\right) + kT\ln\left(\frac{p(-x_{p0})}{n_i}\right)$$

$$= kT\ln\left(\frac{n(x_{n0})p(-x_{p0})}{n_i^2}\right)$$

Now what are $n(x_{n0})$ and $p(-x_{p0})$? They are simply

$n(x_{n0}) = N_D$ (really $N_D - N_A$ on the n side)

$p(-x_{p0}) = N_A$ (really $N_A - N_D$ on the p side)

so…

$$qV_{bi} = kT\ln\left(\frac{N_D N_A}{n_i^2}\right)$$

Now taking into account

$$p(x_{n0}) = \frac{n_i^2}{n(x_{n0})}$$

and

$$n(-x_{p0}) = \frac{n_i^2}{p(-x_{p0})}$$

so…

$$qV_{bi} = kT\ln\left(\frac{n(x_{n0})p(-x_{p0})}{n_i^2}\right)$$

$$= kT\ln\left(\frac{p(-x_{p0})}{p(x_{n0})}\right)$$

$$= kT\ln\left(\frac{n(x_{n0})}{n(-x_{p0})}\right)$$

or!

$$\frac{p(-x_{p0})}{p(x_{n0})} = \frac{n(x_{n0})}{n(-x_{p0})} = e^{qV_{bi}/kT}$$

We now need to find W and **E**. To get them requires that we know the currents in the junction. To get at the currents, we need to make some approximations, known as the ‘depletion approximation’.

To understand why solving these is tough to do exactly, we need to consider the equation describing the electric field across the boundary.

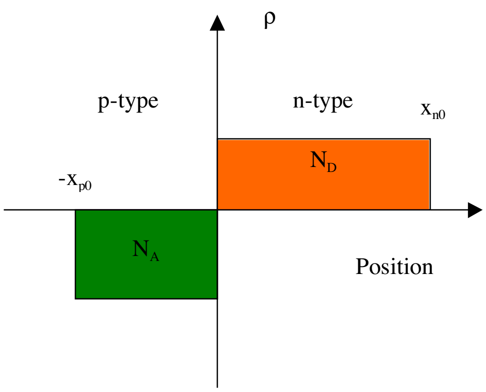
$$\nabla \cdot \mathbf{E} = \frac{\partial E_x}{\partial x} = \frac{\rho}{\varepsilon}$$
$$= \frac{1}{\varepsilon}\left(p(x) - n(x) + N_D(x) - N_A(x)\right)$$

(Remember that while our charge carriers are p and n, we still have the bound charges, N_D and N_A! The donor sites are positively charged while the acceptor sites are negatively charged.) However the hole and electron densities are set in part by the electric field – and we have no clue as to what the functional form is for those densities. Our approximation is related to those densities.

First, outside of the junction, the charge carriers cover the bound charges and thus the charge density is zero. If we assume that the hole and electron densities are zero inside the junction, then we can come up with an approximation that we can solve. (This is not a bad approximation, as the electric field should push all of the charge carriers out of the junction region.) This is because we can assume that the numbers of donor and acceptor sites are constant on each side of the metallurgical junction. Thus

$$\nabla \cdot \mathbf{E} = \frac{\partial E_x}{\partial x} = \frac{\rho}{\varepsilon}$$
$$= \begin{cases} \frac{1}{\varepsilon}\left(N_D(x) - N_A(x)\right) & -x_{p0} \le x \le x_{n0} \\ 0 & \text{elsewhere} \end{cases}$$

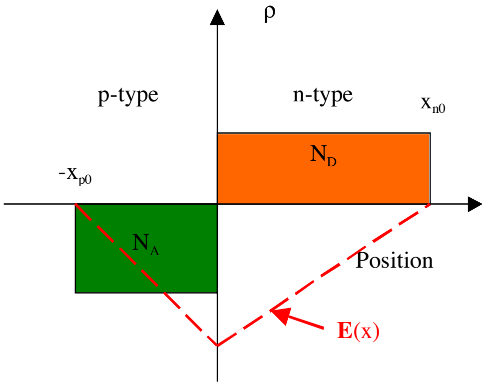
where $N_D(x)$ and $N_A(x)$ are graphically given by:



(This is actually very close to reality in most step p-n junctions. Also note that we have depicted a step junction. A graded junction will have $N_D(x)$ and $N_A(x)$ vary in the region.)
We can integrate on each side of the junction to get:

$$E_{xp} = \int_{-x_{p0}}^{x} -\frac{qN_A}{\varepsilon}dx$$
$$= -\frac{qN_A}{\varepsilon}\left(x + x_{p0}\right)$$
$$E_{xn} = \int_{x}^{x_{n0}} \frac{qN_D}{\varepsilon}dx$$
$$= \frac{qN_D}{\varepsilon}\left(x_{n0} - x\right)$$

Graphically this looks like:



Note that the maximum electric field is at x = 0 and that it goes to zero at the edges of the junction.

We can now get the potential at all points inside the junction (which will lead us to our junction width!)

$$V = -\frac{\partial E}{\partial x}$$
$$\Downarrow$$
$$V = -\int_{x_1}^{x_2} \mathbf{E}dx$$
$$\Downarrow$$
$$V(x) = -\int_{-x_{p0}}^{x} \mathbf{E}dx$$
$$= \frac{qN_A}{2\varepsilon}\left(x + x_{p0}\right)^2 - V\left(-x_{p0}\right) \quad x_{p0} \le x \le 0$$
$$V(x) = -\int_{x}^{x_{n0}} \mathbf{E}dx$$
$$= V\left(x_{n0}\right) - \frac{qN_D}{2\varepsilon}\left(x_{n0} - x\right)^2 \qquad 0 \le x \le x_{n0}$$

By looking at our picture, we see that we can set

$$V\left(x_{p0}\right) = 0$$
$$V\left(x_{n0}\right) = V_{bi}$$

Thus,

$$V(x) = \frac{qN_A}{2\varepsilon}\left(x + x_{p0}\right)^2 \qquad x_{p0} \le x \le 0$$

$$V(x) = V_{bi} - \frac{qN_D}{2\varepsilon}\left(x_{n0} - x\right)^2 \qquad 0 \le x \le x_{n0}$$

At the metallurgical, x = 0, these must match.

$$V_{bi} = \frac{qN_D}{2\varepsilon}\left(x_{n0}\right)^2 + \frac{qN_A}{2\varepsilon}\left(x_{p0}\right)^2$$

At this point, we note that the total charge inside the junction must be zero. (In electromagnetism, one learns that if the total change in a region is not zero, then there is an electric field outside of that region. We have setup our definition of the junction such that 'all' of the electric field is inside.) Thus,

$$Q_p = Aqx_{n0}N_D$$

$$Q_n = -Aqx_{p0}N_A$$

$$Q_p + Q_n = 0$$

$$\Downarrow$$

$$x_{n0}N_D = x_{p0}N_A$$

We can now plug this into our above equation to get

$$V_{bi} = \frac{qN_D}{2\varepsilon}\left(x_{n0}\right)^2 + \frac{qN_A}{2\varepsilon}\left(\frac{x_{n0}N_D}{N_A}\right)^2$$

$$= \left(\frac{qN_D}{2\varepsilon} + \frac{qN_D^2}{2\varepsilon N_A}\right)\left(x_{n0}\right)^2$$

$$= \left(\frac{q\left(N_AN_D + N_D^2\right)}{2\varepsilon N_A}\right)\left(x_{n0}\right)^2$$

$$\Downarrow$$

$$x_{n0} = \left(\frac{2\varepsilon N_A V_{bi}}{qN_D\left(N_A + N_D\right)}\right)^{1/2}$$

$$x_{p0} = \frac{x_{n0}N_D}{N_A} = \left(\frac{2\varepsilon V_{bi} N_D}{qN_A\left(N_A + N_D\right)}\right)^{1/2}$$

So… the width is the sum of x_{n0} and x_{p0}. Or,

$w = x_{n0} + x_{p0}$

$$= \left(\frac{2\varepsilon N_A V_{bi}}{qN_D(N_A + N_D)}\right)^{1/2} + \left(\frac{2\varepsilon V_{bi} N_D}{qN_A(N_A + N_D)}\right)^{1/2}$$

$$= \left(\frac{2\varepsilon V_{bi}}{q(N_A + N_D)}\right)^{1/2}\left(\left(\frac{N_D}{N_A}\right)^{1/2} + \left(\frac{N_A}{N_D}\right)^{1/2}\right)$$

$$= \left(\frac{2\varepsilon V_{bi}}{q(N_A + N_D)}\right)^{1/2}\left[\left(\left(\frac{N_D}{N_A}\right)^{1/2} + \left(\frac{N_A}{N_D}\right)^{1/2}\right)^2\right]^{1/2}$$

$$= \left(\frac{2\varepsilon V_{bi}}{q(N_A + N_D)}\right)^{1/2}\left[\left(\left(\frac{N_D}{N_A}\right) + \left(\frac{N_A}{N_D}\right) + 2\right)\right]^{1/2}$$

$$= \left(\frac{2\varepsilon V_{bi}}{q(N_A + N_D)}\right)^{1/2}\left[\frac{N_D^2 + N_A^2 + 2N_A N_D}{N_A N_D}\right]^{1/2}$$

$$= \left(\frac{2\varepsilon V_{bi}(N_A + N_D)}{qN_A N_D}\right)^{1/2}$$

We can rearrange these to show

$$x_{n0} = w\frac{N_A}{(N_A + N_D)}$$

$$x_{p0} = w\frac{N_D}{(N_A + N_D)}$$

$$V_{bi} = \frac{1}{2}\frac{q}{\varepsilon}\frac{N_A N_D}{(N_A + N_D)}w^2$$

$$= \frac{kT}{q}\ln\left(\frac{n(x_{n0})p(-x_{p0})}{n_i^2}\right) = \frac{kT}{q}\ln\left(\frac{N_A N_D}{n_i^2}\right)$$

Example:

A p-type GaAs layer is grown on n-type GaAs:

$N_D = 10^{14}$ cm^{-3}, $N_A = 2X10^{15}$ cm^{-3}, $n_i = 2X10^6$ cm^{-3}, $e_r = 13.2 \Rightarrow e = e_r e_0$.

Find V_{bi}, w, E_{max}, x_{n0}, x_{p0}, and Q/A

a)

$$V_{bi} = \frac{kT}{q} \ln\left(\frac{N_A N_D}{n_i^2}\right)$$

$$= 0.0259\text{V} \ \ln\left(\frac{10^{14} \ \ 2X10^{15}}{\left(2X10^6\right)^2}\right)$$

$$= 0.996\text{V}$$

b-d)

$$w = \left(\frac{2\varepsilon V_{bi}(N_A + N_D)}{qN_A N_D}\right)^{1/2}$$

$$= \left(\frac{2 \ \ 13.2 \ \ 8.85X10^{-14} \ \ 0.996 \ \left(2X10^{15} + 10^{14}\right)}{1.6X10^{-19} \ \ 2X10^{15} \ \ 10^{14}}\right)^{1/2}$$

$$\approx 3.9 \ \ \mu\text{m} \ \ \text{(large!)}$$

$$x_{n0} = w \frac{N_A}{(N_A + N_D)}$$

$$= 3.9 \ \ \mu\text{m} \frac{2X10^{15}}{\left(2X10^{15} + 10^{14}\right)}$$

$$= 3.72 \ \ \mu\text{m}$$

$$x_{p0} = w \frac{N_D}{(N_A + N_D)}$$

$$= 0.19 \ \ \mu\text{m}$$

e)

$$E_{MAX} = \frac{qN_D}{4\pi\varepsilon}(x_{n0})$$

$$= 8.11 \ \ \text{kV / m}$$

f) Total charge on each side

$$\frac{Q}{A} = qx_{p0}N_A = qx_{n0}N_D$$

$$= 2.098X10^{-8} \ \ \text{C / cm}^{-3}$$

Now, we make our life even more complex by placing a voltage across it.

APPLIED VOLTAGE ACROSS A P-N JUNCTION

How we apply the voltage across the p-n junction makes a significant difference in how the device responds. To understand why, we first look at the device in a qualitative manner. (We will come back and do this quantitatively later.)

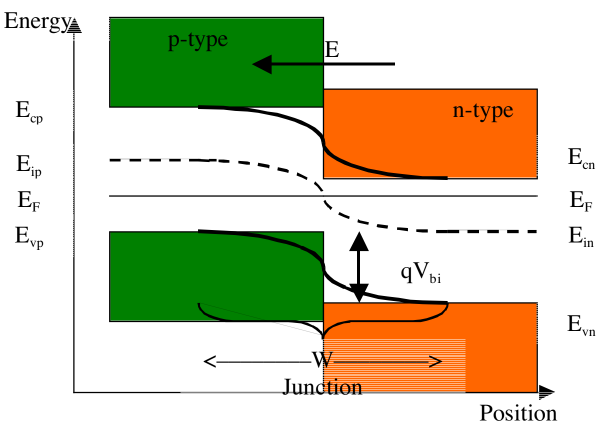
Bias and Junction

To get at the bias and junction parameters we need to understand that most of an applied bias will fall across the junction. This is because both the p-side and n-side have charge carriers that are mobile and thus will move so as to get rid of any electric field outside of the junction. Inside of the junction, the number of mobile charge carriers is very small and thus the electric field can be maintained. This means that to get the junction parameters we simply need to replace V_{bi} in the above derivation with V_{bi} - V_A. (We are taking V_A to be positive in the 'forward' direction. [Reverse and forward have to do with current flow.] This sign flip makes life easier when we look at diodes as a whole in circuits. For now, it seems backwards.)

Current flow

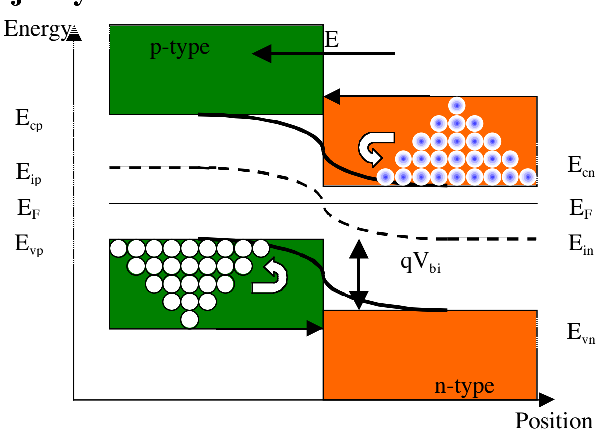
Qualitative analysis

Without any applied bias, and hence no current, we have our original energy band structure.



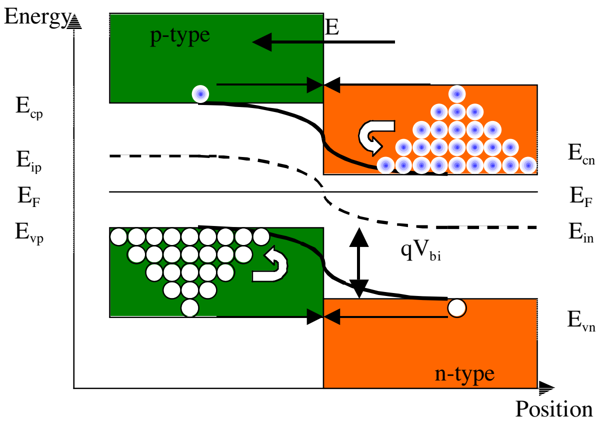
Let us put some electrons and holes on this diagram and examine where they go.

Majority carriers



The most energetic electrons and holes can overcome the electric field barrier and move to the other side of the junction. (This is the diffusion process at work.)

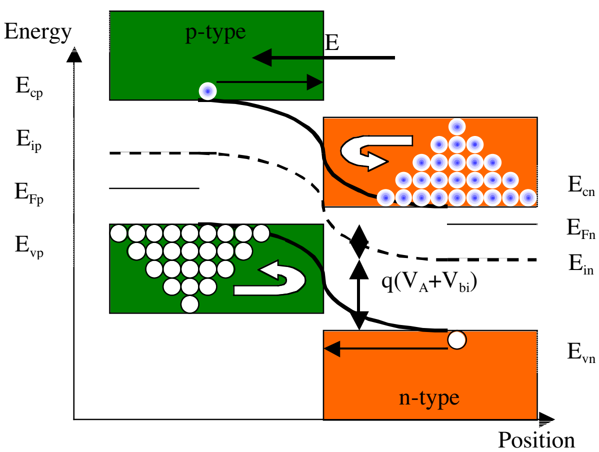
Minority carriers



Any minority carrier that wanders into the junction region, gets drawn by the electric field to the other side of the junction. In our unbiased system, these currents balance and thus the total current density is zero.

Reverse Bias

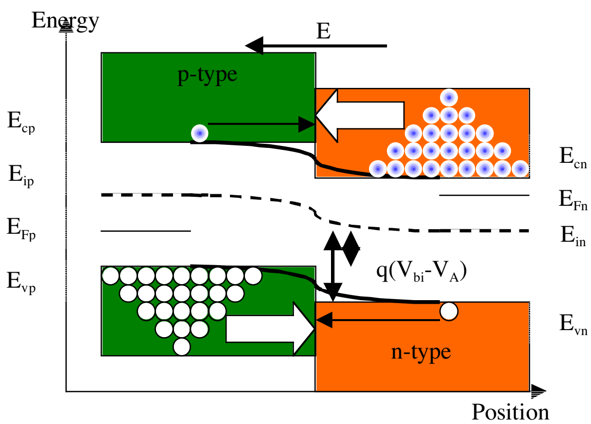
Reverse bias implies that we are adding a bias in the same direction as junction bias. It is known as reverse bias as because this is, as we will see, the low current direction. Now our energy diagram looks like:



Now the minority carriers can still move across the barrier but the majority carriers do not have enough energy to over come the barrier. Additionally, we see that the Fermi energies on the two sides are different, by qV_A, where $-V_A$ is the applied bias. Thus we would expect that the total current is set by any minority carriers that just happen to run into the junction. Thus we would expect a low level current that is effectively independent of the applied bias.

Forward Bias

Forward bias implies that we are adding a bias in the opposite direction to the junction bias. It is known as forward bias as because this is, as we will see, the high current direction. Now our energy diagram looks like:



Now the current is set by the number of majority carriers that have sufficient energy to overcome the reduced barrier. We know that the number of electrons/holes is

$n(\mathcal{E})d\mathcal{E} = f(\mathcal{E})N_c(\mathcal{E})d\mathcal{E}$ Electrons in the Conduction band

$$= \frac{1}{1+\exp\left[\frac{(\mathcal{E}-\mathcal{E}_F)}{kT}\right]}\frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}-\mathcal{E}_c}\,d\mathcal{E}$$

$$\approx \exp\left[\frac{-(\mathcal{E}-\mathcal{E}_F)}{kT}\right]\frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}-\mathcal{E}_c}\,d\mathcal{E}$$

$$\propto \exp\left[\frac{-(\mathcal{E}-\mathcal{E}_F)}{kT}\right]d\mathcal{E}$$

$p(\mathcal{E})d\mathcal{E} = (1-f(\mathcal{E}))N_v(\mathcal{E})d\mathcal{E}$ Holes in the Valance band

$$= \left(1-\frac{1}{1+\exp\left[\frac{(\mathcal{E}-\mathcal{E}_F)}{kT}\right]}\right)\frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}_v-\mathcal{E}}\,d\mathcal{E}$$

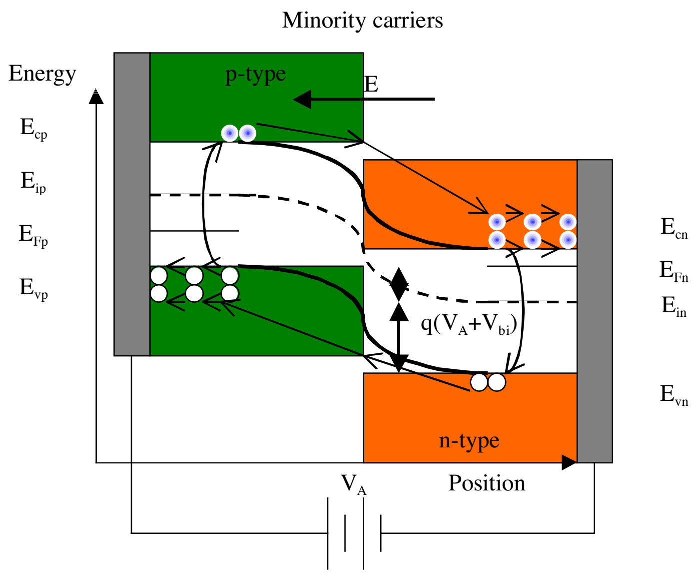
$$\approx \left(\exp\left[\frac{(\mathcal{E}-\mathcal{E}_F)}{kT}\right]\right)\frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}_v-\mathcal{E}}\,d\mathcal{E}$$

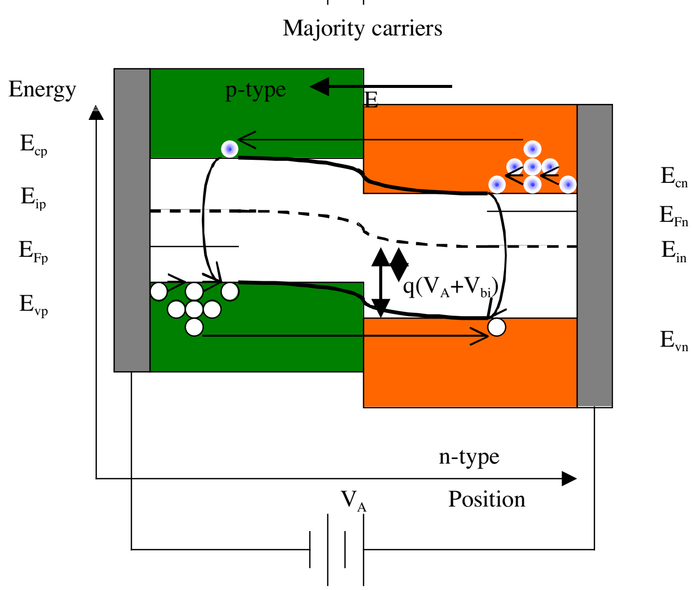
$$\propto \exp\left[\frac{(\mathcal{E}-\mathcal{E}_F)}{kT}\right]d\mathcal{E}$$

Thus we might expect that in forward bias, the current increases exponentially with voltage.

Total current flow

Before we go on, we want to look at the total path of the current. For this, we will use the reverse bias condition. The other conditions look similar.





There are two things to note:

1) For both the majority and minority carriers, the electrons flow in the opposite direction to the holes and thus, because of the difference in charge, produce current flow in the same direction.
2) The majority carriers flow in the opposite direction as the minority carriers. Thus switching the sign of V_A will change the direction of the current flow.

Quantitative Analysis of current flow.

To make this analysis, we need to make a few simplifying assumptions.

1) The diode is being operated under steady state conditions.

2) The junction can be modeled as a non-degenerately doped step junction. (More complex junctions can be modeled at the price of increased difficulty.)
3) The diode is one-dimensional. (Again, more complex diodes can be modeled at the price of increased difficulty.)
4) Low-level injection prevails in the quasi-neutral region. (This is the typical mode of operation.)
5) The only processes occurring in the system are drift, diffusion, thermal generation and recombination. (No light emission or absorption etc.)

Each of these assumptions are reasonable and are similar to what we assumed for electrostatics in an unbiased p-n junction. In fact, the only distinction that we have is that the bias across the junction is now V_A+V_{bi} and there is a net current flow.

Now what is our current?

$$I = JA$$

$$J = J_N(x) + J_P(x) = \text{constant}$$

$$J_N = q\mu_n n(x)\mathbf{E}(x) + qD_n \frac{\partial n(x)}{\partial x}$$

$$J_P = q\mu_p n(x)\mathbf{E}(x) - qD_p \frac{\partial p(x)}{\partial x}$$

where A is the cross section of the diode. Note that because we do not have source/sink of charge and we are in equilibrium, then the total current flowing across the diode must be constant.

Now how does quasi-neutrality influence our results? Under steady state condition, the drift of the minority carrier must be balanced by the diffusion. Thus,

Repeat of derivation

⇓

$$\frac{\partial n}{\partial t} = \frac{1}{q_n}\nabla \Sigma \mathbf{J}_n + \overbrace{g_{opt}}^{\text{optical excitation}} + \overbrace{\alpha_R n_i^2}^{\text{thermal excitation}} - \overbrace{\alpha_R n(t)p(t)}^{\text{recombination}}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q_p}\nabla \Sigma \mathbf{J}_p + \overbrace{g_{opt}}^{\text{optical excitation}} + \overbrace{\alpha_R n_i^2}^{\text{thermal excitation}} - \overbrace{\alpha_R n(t)p(t)}^{\text{recombination}}$$

Now, however, we have shown that the last two terms can be combined to leave a simpler single term

$$\frac{\partial n}{\partial t}=\frac{1}{q_n}\nabla\Sigma\mathbf{J}_n+g_{opt}+\alpha_R n_i^2-\alpha_R(n_0+\Delta n)(p_0+\Delta p)$$
$$=\frac{1}{q_n}\nabla\Sigma\mathbf{J}_n+g_{opt}-\Delta n\alpha_R(n_0+p_0)$$
$$=\frac{1}{q_n}\nabla\Sigma\mathbf{J}_n+g_{opt}-\frac{\Delta n}{\tau}$$
$$\frac{\partial p}{\partial t}=-\frac{1}{q_p}\nabla\Sigma\mathbf{J}_p+g_{opt}+\alpha_R n_i^2-\alpha_R(n_0+\Delta n)(p_0+\Delta p)$$
$$=-\frac{1}{q_p}\nabla\Sigma\mathbf{J}_p+g_{opt}-\Delta p\alpha_R(n_0+p_0)$$
$$=-\frac{1}{q_p}\nabla\Sigma\mathbf{J}_p+g_{opt}-\frac{\Delta p}{\tau}$$

At this point, we can now put in what we have derived for our currents

$$\mathbf{J}_n=q_n n\mu_n\mathbf{E}+q_n D_n\nabla n$$
$$\mathbf{J}_p=q_p p\mu_p\mathbf{E}-q_p D_p\nabla p$$

Plugging these in gives

⇓

$$\frac{\partial n}{\partial t}=\frac{1}{q_n}\nabla\Sigma(q_n n\mu_n\mathbf{E}+q_n D_n\nabla n)+g_{opt}-\frac{\Delta n}{\tau}$$
$$=n\mu_n\nabla\Sigma\mathbf{E}+D_n\nabla^2 n+g_{opt}-\frac{\Delta n}{\tau}$$
$$\frac{\partial p}{\partial t}=-\frac{1}{q_p}\nabla\Sigma(q_p n\mu_p\mathbf{E}-q_p D_p\nabla p)+g_{opt}-\frac{\Delta p}{\tau}$$
$$=-p\mu_p\nabla\Sigma\mathbf{E}+D_p\nabla^2 p+g_{opt}-\frac{\Delta p}{\tau}$$

finally, we need to look at the left-hand side of the equation and note that the density is made up of the equilibrium part and the injected part and further note that the equilibrium part is a constant; thus

⇓

$$\frac{\partial\Delta n}{\partial t}=n\mu_n\nabla\Sigma\mathbf{E}+D_n\nabla^2\Delta n+g_{opt}-\frac{\Delta n}{\tau}$$
$$\frac{\partial\Delta p}{\partial t}=-p\mu_p\nabla\Sigma\mathbf{E}+D_p\nabla^2\Delta p+g_{opt}-\frac{\Delta p}{\tau}$$

Thus for the minority carriers on the p-side ($x\le -x_{p0}$)

$$\frac{\partial \Delta n_p}{\partial t} = 0 = \overbrace{n\mu_n \nabla \cdot \mathbf{E}}^{\mathbf{E}=0} + D_n \nabla^2 \Delta n_p + \overbrace{g_{opt}}^{=0} - \frac{\Delta n_p}{\tau}$$

$$\Downarrow$$

$$\frac{\Delta n_p}{\tau} = D_n \nabla^2 \Delta n_p$$

and on the n side ($x \geq x_{n0}$)

$$\frac{\partial \Delta p_n}{\partial t} = 0 = -\overbrace{p\mu_p \nabla \cdot \mathbf{E}}^{\mathbf{E}=0} + D_p \nabla^2 \Delta p_n + \overbrace{g_{opt}}^{=0} - \frac{\Delta p_n}{\tau}$$

$$\Downarrow$$

$$\frac{\Delta p_n}{\tau} = D_p \nabla^2 \Delta p_n$$

Likewise, we can examine the minority current on each side of the junction. Again, $\mathbf{E} = 0$, so

$$J_N = qD_n \frac{\partial n_p(x)}{\partial x} = qD_n \frac{\partial \Delta n_p(x)}{\partial x} \qquad x > x_{n0}$$

$$\Downarrow$$

$$\nabla \cdot J_N = qD_n \nabla^2 \Delta n_p(x)$$

$$= \frac{q\Delta n_p}{\tau}$$

and

$$J_P = -qD_p \frac{\partial p_n(x)}{\partial x} = -qD_p \frac{\partial \Delta p_n(x)}{\partial x} \qquad x < -x_{p0}$$

$$\Downarrow$$

$$\nabla \cdot J_P = -qD_p \nabla^2 \Delta p_n(x)$$

$$= -\frac{q\Delta p_n}{\tau}$$

As we move away from the junction, we should expect the minority carrier density to drop toward the equilibrium densities. Thus at a long distance from the junction, e.g. the contact location, the minority carrier density drops to the equilibrium density.

$$\Delta n_p(x = -\infty) = 0$$

$$\Delta p_n(x = +\infty) = 0$$

Because the total current in each side is constant, these equations imply that the current density shifts from the minority carrier to the majority carrier as one moves away from the junction toward the contact. We will return to this concept later.

Junction region

Now we need to look at what happens in the junction region. Here, we cannot set the electric field to zero. Thus

$$\frac{\partial \Delta n_p}{\partial t} = 0 = \overbrace{n\mu_n \nabla \Sigma \mathbf{E} + D_n \nabla^2 \Delta n_p}^{=\frac{1}{q}\nabla \Sigma \mathbf{J}_N} + \overbrace{g_{opt}}^{=0} - \frac{\Delta n_p}{\tau}$$

$$\frac{\partial \Delta p_n}{\partial t} = 0 = \overbrace{-p\mu_p \nabla \Sigma \mathbf{E} + D_p \nabla^2 \Delta p_n}^{=-\frac{1}{q}\nabla \Sigma \mathbf{J}_P} + \overbrace{g_{opt}}^{=0} - \frac{\Delta p_n}{\tau}$$

If the thermal recombination is small inside the junction then the currents must be constant across the junction. Physically, the recombination mechanism requires both hole and electrons to be plentiful for the rate to be high. In the junction, the number of holes and electrons is relatively low, so the recombination rate is relatively slow. This make the last term in the above equations approximately zero. Thus

$$\frac{1}{q}\nabla \Sigma \mathbf{J}_N = 0 \Rightarrow \mathbf{J}_N\left(-x_{p0} \le x \le x_{n0}\right) \approx \mathbf{J}_N\left(-x_{p0}\right)$$

$$\frac{1}{q}\nabla \Sigma \mathbf{J}_P = 0 \Rightarrow \mathbf{J}_P\left(-x_{p0} \le x \le x_{n0}\right) \approx \mathbf{J}_P\left(x_{n0}\right)$$

$$\Downarrow$$

$$\begin{aligned} J &= J_N(x) + J_P(x) = \text{constant} \\ &= \mathbf{J}_N\left(-x_{p0}\right) + \mathbf{J}_P\left(x_{n0}\right) \end{aligned}$$

Now we need to match the boundary conditions. (We cannot have a discontinuous solution.)
To do this, we will make an assumption for which we do not have a good a-priori reason – other then it works. That assumption is that the quasi-Fermi energies are constant through the junction and equal to the Fermi energy on the appropriate side. Thus

$$F_n = E_{Fn}$$

$$F_p = E_{Fp}.$$

Now the quasi-Fermi energy is related to the carrier densities by

$$\begin{aligned} np &= n_i e^{(F_n - E_i)/kT} * n_i e^{-\left(F_p - E_i\right)/kT} \\ &= n_i^2 e^{\left(F_n - F_p\right)/kT} \end{aligned}$$

This holds everywhere. If we assume that the quasi-Fermi energies are constant inside the junction then

$$np = n_i^2 e^{qV_A/kT} \qquad -x_{p0} \le x \le x_{n0}$$

At the p-edge

$$n(-x_{p0})p(-x_{p0}) = n_i^2 e^{(F_n - F_p)/kT} = n_i^2 e^{qV_A/kT}$$
$$= n(-x_{p0})N_A$$
$$\Downarrow$$
$$n(-x_{p0}) = \frac{n_i^2}{N_A} e^{(F_n - F_p)/kT} = \frac{n_i^2}{N_A} e^{qV_A/kT}$$
$$= n_{p0} + \Delta n(-x_{p0})$$
$$= \frac{n_i^2}{N_A} + \Delta n(-x_{p0})$$
$$\Downarrow$$
$$\Delta n(-x_{p0}) = \frac{n_i^2}{N_A}\left(e^{(F_n - F_p)/kT} - 1\right) = \frac{n_i^2}{N_A}\left(e^{qV_A/kT} - 1\right)$$

At the n-edge

$$n(x_{n0})p(x_{n0}) = n_i^2 e^{(F_n - F_p)/kT} = n_i^2 e^{qV_A/kT}$$
$$= p(x_{n0})N_D$$
$$\Downarrow$$
$$p(x_{n0}) = \frac{n_i^2}{N_D} e^{(F_n - F_p)/kT} = \frac{n_i^2}{N_D} e^{qV_A/kT}$$
$$= p_{n0} + \Delta p(x_{n0})$$
$$= \frac{n_i^2}{N_D} + \Delta p(x_{n0})$$
$$\Downarrow$$
$$\Delta p(x_{n0}) = \frac{n_i^2}{N_D}\left(e^{(F_n - F_p)/kT} - 1\right) = \frac{n_i^2}{N_D}\left(e^{qV_A/kT} - 1\right)$$

Now, we can solve our equation for the minority carriers subject to the boundary conditions. First, we will start with $x \le -x_{p0}$, (The p-side).

$$D_n \nabla^2 \Delta n_p = \frac{\Delta n_p}{\tau}$$
$$\Delta n_p(x = -\infty) = 0$$
$$\Delta n(-x_{p0}) = \frac{n_i^2}{N_A}\left(e^{qV_A/kT} - 1\right)$$

integrating

$$\Downarrow$$

$$\Delta n_p = Ae^{+(x+x_{p0})/L_n} + Be^{-(x+x_{p0})/L_n}$$

where $L_n = \sqrt{D_n \tau_n}$

Plugging in our boundary conditions, we find that

$$A = \frac{n_i^2}{N_A}\left(e^{qV_A/kT} - 1\right)$$

and

$$B = 0$$

Thus

$$\Delta n_p = \frac{n_i^2}{N_A}\left(e^{qV_A/kT} - 1\right)e^{+(x+x_{p0})/L_n}$$

Likewise for the n-side (in a repeat of the above derivation)

$$D_p \nabla^2 \Delta p_n = -\frac{\Delta p_n}{\tau}$$

$$\Delta p\left(x_{n0}\right) = \frac{n_i^2}{N_D}\left(e^{qV_A/kT} - 1\right)$$

$$\Delta p_n\left(x = +\infty\right) = 0$$

integrating

$$\Downarrow$$

$$\Delta p_n = Ae^{+(x-x_{n0})/L_n} + Be^{-(x-x_{n0})/L_p}$$

where $L_p = \sqrt{D_p \tau_p}$

Plugging in, we find

$$B = \frac{n_i^2}{N_D}\left(e^{qV_A/kT} - 1\right)$$

and

$$A = 0$$

Thus

$$\Delta p_n = \frac{n_i^2}{N_D}\left(e^{qV_A/kT} - 1\right)e^{-(x-x_{n0})/L_p}.$$

Now, we can get the currents

$$J_N = qD_n \frac{\partial \Delta n_p(x)}{\partial x}$$
$$= qD_n \frac{\partial}{\partial x}\left[\frac{n_i^2}{N_A}\left(e^{qV_A/kT}-1\right)e^{+(x+x_{p0})/L_n}\right]$$
$$= qD_n \frac{n_i^2}{N_A}\left(e^{qV_A/kT}-1\right)\frac{\partial}{\partial x}\left[e^{+(x+x_{p0})/L_n}\right]$$
$$= \frac{qD_n}{L_n}\Delta n_p(x)$$

and

$$J_P = -qD_p \frac{\partial \Delta p_n(x)}{\partial x}$$
$$= -qD_p \frac{\partial}{\partial x}\left[\frac{n_i^2}{N_D}\left(e^{qV_A/kT}-1\right)e^{-(x-x_{n0})/L_p}\right]$$
$$= -qD_p \frac{n_i^2}{N_D}\left(e^{qV_A/kT}-1\right)\frac{\partial}{\partial x}\left[e^{-(x-x_{n0})/L_p}\right]$$
$$= \frac{qD_p}{L_p}\Delta p_n(x)$$

At this point, we have assumed that the currents of electrons and holes are constant through the junction. Thus, the total current is simply the sum of the minority currents at the respective junction edges

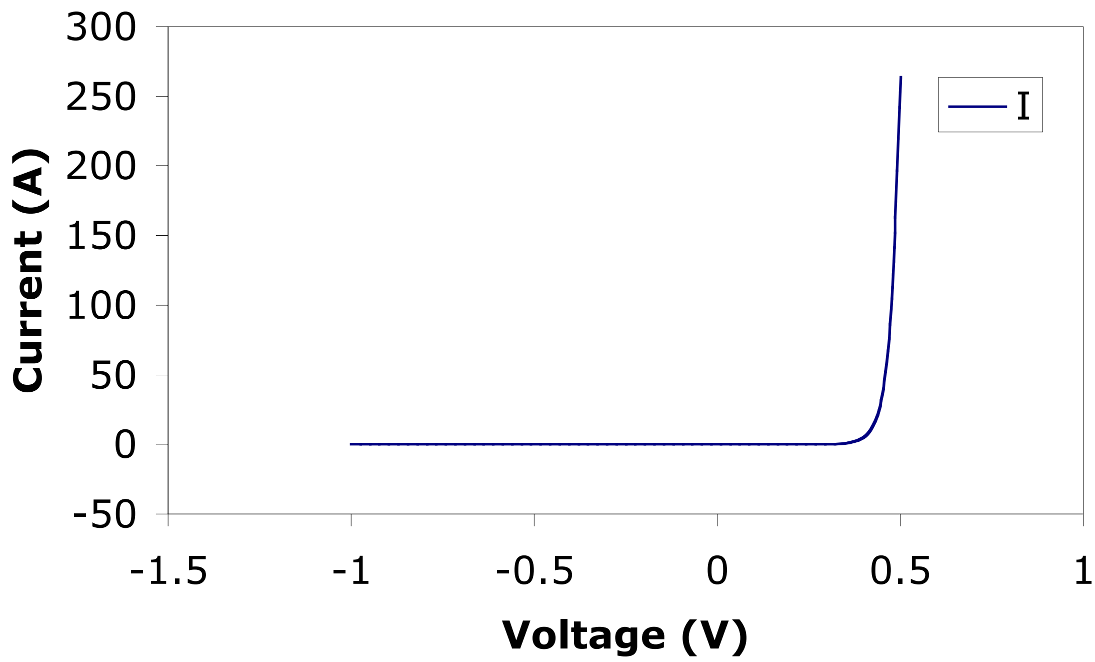
$$J_{total} = J_N(x) + J_P(x)$$
$$= J_N\left(x = -x_{p0}\right) + J_P\left(x = x_{n0}\right)$$
$$= q\frac{D_n n_i^2}{L_n N_A}\left(e^{qV_A/kT}-1\right) + q\frac{D_p n_i^2}{L_p N_D}\left(e^{qV_A/kT}-1\right)$$
$$= \left(\frac{D_n}{L_n N_A} + \frac{D_p}{L_p N_D}\right) q n_i^2\left(e^{qV_A/kT}-1\right)$$
$$= J_0\left(e^{qV_A/kT}-1\right)$$
$$\Downarrow \quad (I = JA)$$
$$I = I_0\left(e^{qV_A/kT}-1\right)$$
$$I_0 = q n_i^2 A\left(\frac{D_n}{L_n N_A} + \frac{D_p}{L_p N_D}\right)$$

On a side note, we can show an additional relationship between the current and the total excess charge on each side. From above:

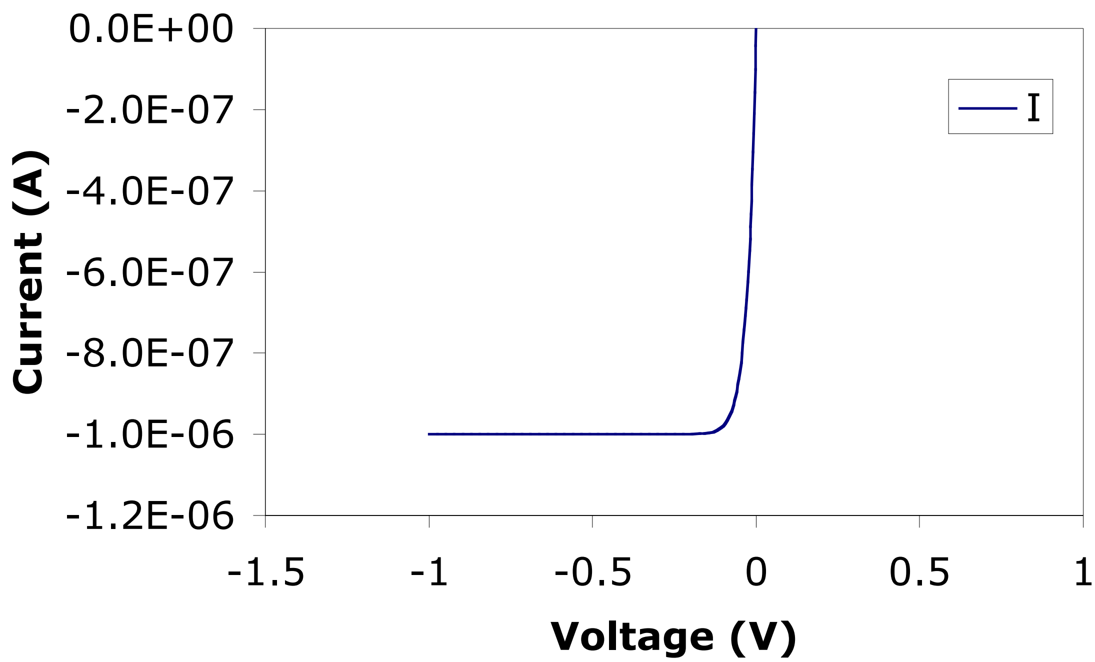
$$I_{P_{\min or}}(x) = A\frac{qD_p}{L_p}\Delta p_n(x)$$
$$= A\frac{qD_p}{L_p}\frac{n_i^2}{N_D}\left(e^{qV_A/kT}-1\right)\left[e^{-(x-x_{n0})/L_p}\right]$$
but the total charge is
$$Q_p = qA\int_{x_{n0}}^{\infty}\Delta p_n(x)dx$$
$$= qA\frac{n_i^2}{N_D}\left(e^{qV_A/kT}-1\right)\int_{x_{n0}}^{\infty}\left[e^{-(x-x_{n0})/L_p}\right]dx$$
$$= -qA\frac{L_p n_i^2}{N_D}\left(e^{qV_A/kT}-1\right)\left[e^{-(x-x_{n0})/L_p}\right]_{x_{n0}}^{\infty}$$
$$= qA\frac{L_p n_i^2}{N_D}\left(e^{qV_A/kT}-1\right)$$
plugging this into our total hole current, e.g. the minority current at the junction edge, we find
$$I_{P_{total}} = I_{P_{\min or}}(x = x_{n0}) = A\frac{qD_p}{L_p}\Delta p_n(x = x_{n0})$$
$$= A\frac{qD_p}{L_p}\frac{n_i^2}{N_D}\left(e^{qV_A/kT}-1\right)$$
$$= \frac{D_p}{L_p^2}\left(qA\frac{L_p n_i^2}{N_D}\left(e^{qV_A/kT}-1\right)\right)$$
$$= \frac{D_p}{D_p\tau_p}(Q_p)$$
$$= \frac{Q_p}{\tau_p}$$
Likewise, we can show
$$I_{n_{total}} = \frac{Q_n}{\tau_n}$$
This means that we have a couple of equivalent methods of calculating the current, first from the slope of the minority carrier concentration, the method we used first, or the total charge and the carrier lifetimes, our second method, shown just now.

So what does this look like and what does this mean?

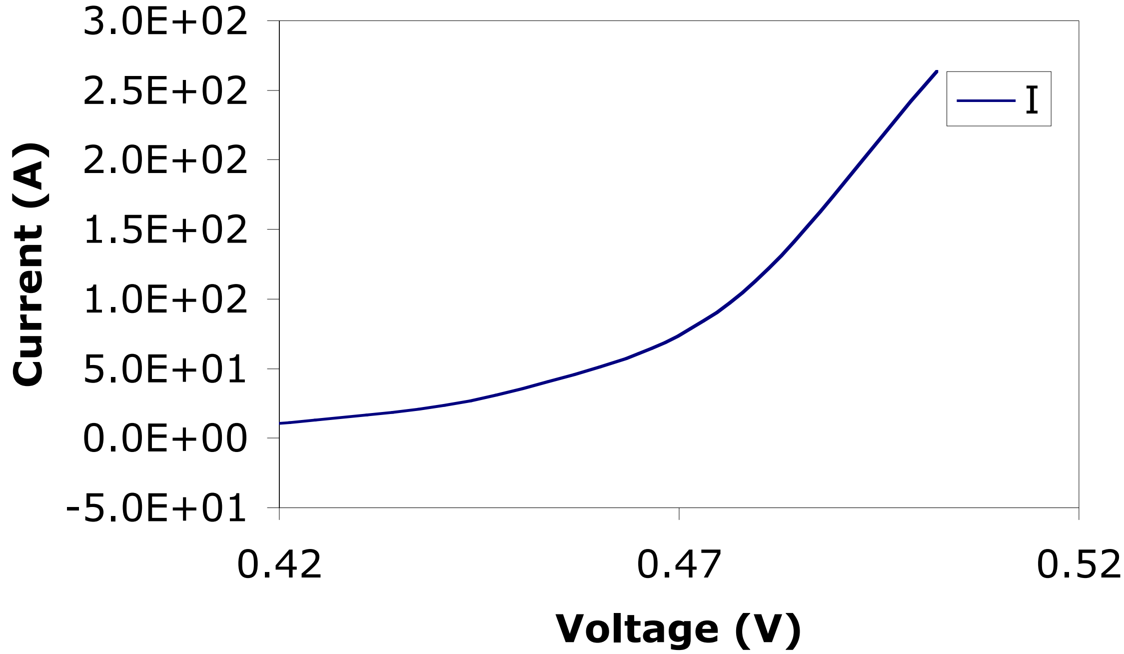
A basic plot of the I-V character looks like:



On this scale, we see that the diode has very little current below $V_A = 0.5$ V. Above that, the current spikes. We can blow this picture up to get a better view of the extremes.



Here we have blown the up the area of negative current. From this, we can see that the diode begins to turn on about 0 V not 0.5. (It is just that the first picture is not on the correct scale.)



Now we are blowing up the area near 0.5 V. This gives us a reasonable picture of how the current changes as a function of V; it is approximately logarithmic. None of these results are surprising.

Can we gain an understanding of what will happen to similar diodes with different material characteristics? Yes
First a single diode is usually only dependent on the material characteristic of one side of the diode.
How can we tell this?

$$I = I_0\left(e^{qV_A/kT} - 1\right)$$

$$I_0 = qn_i^2 A\left(\frac{D_n}{L_n N_A} + \frac{D_p}{L_p N_D}\right)$$

$$= qn_i^2 A\left(\frac{\sqrt{D_n/\tau_n}}{N_A} + \frac{\sqrt{D_p/\tau_p}}{N_D}\right)$$

The important part is the equation for I_0. We see in that equation that the contribution to the current depends linearly on the dopant concentration and weakly on the diffusion coefficient and the mean collision frequency. (The dependence is in a square root and thus in not even linear.) For any given single type of material, the diffusion coefficients and collision frequencies of the holes and electrons are very similar – though not exactly the same. Thus our only true dependence is in the dopant concentrations. Thus our current is highly dependent on the dopants. If one of the concentration is significantly higher than the other – i.e. by an order of magnitude (factor of ten) – then the contribution to the current from that side is insignificant. Thus,

$$I_0 = \begin{cases} qn_i^2 A\left(\dfrac{\sqrt{D_n/\tau_n}}{N_A}\right) & N_A < N_D \Leftrightarrow p-n^+ \text{ type diode} \\ qn_i^2 A\left(\dfrac{\sqrt{D_n/\tau_n}}{N_A} + \dfrac{\sqrt{D_p/\tau_p}}{N_D}\right) & N_D \approx N_A \Leftrightarrow p-n \text{ type diode} \\ qn_i^2 A\left(\dfrac{\sqrt{D_p/\tau_p}}{N_D}\right) & N_D < N_A \Leftrightarrow p^+-n \text{ type diode} \end{cases}$$

Often, our diodes are either p^+-n type or p-n^+ type.

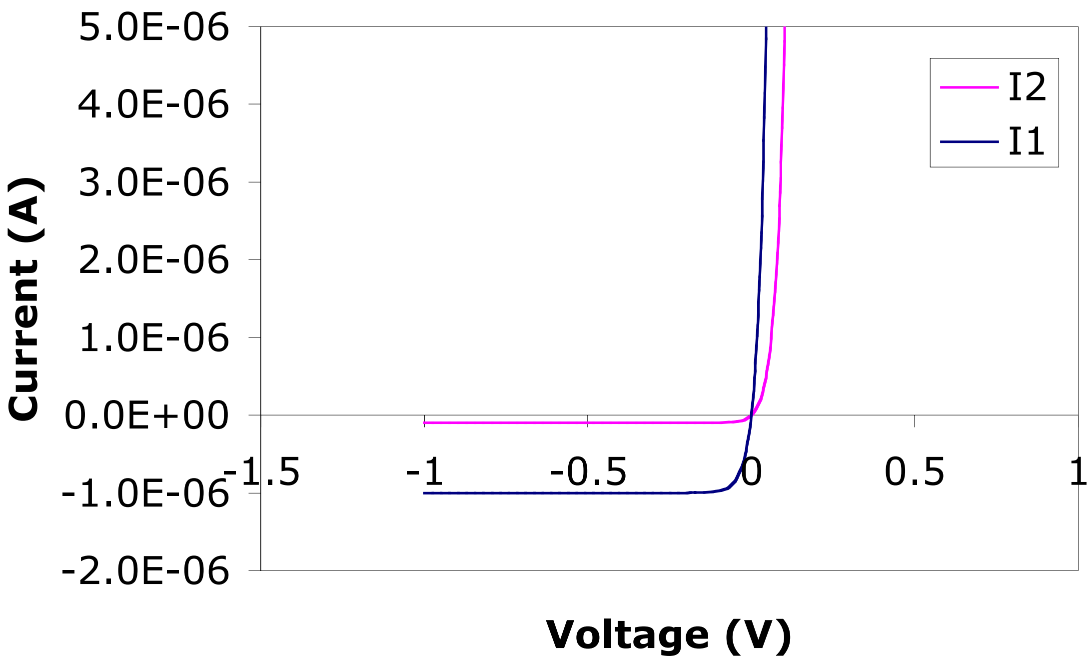
Example:
Given two identical Si based p^+-n diodes, except for the donor concentration on the n-side, plot the currents of the two diode. Use N_{D1}=1E15 cm^{-3} and N_{D2}=1E16 cm^{-3}.
Answer:

$$\frac{I_{01}}{I_{02}} \approx \frac{qn_i^2 A\left(\dfrac{\sqrt{D_{p1}/\tau_{p1}}}{N_{D1}}\right)}{qn_i^2 A\left(\dfrac{\sqrt{D_{p2}/\tau_{p2}}}{N_{D2}}\right)}$$

$$= \frac{\sqrt{D_{p1}/\tau_{p1}}}{\sqrt{D_{p2}/\tau_{p2}}} \frac{N_{D2}}{N_{D1}}$$

now the diffusion and collision frequencies will be almost identical in the two devices. Thus,

$$\frac{I_{01}}{I_{02}} \approx \frac{N_{D2}}{N_{D1}}$$



Second, if we change the type of base material, this can have a major impact on the current from the diode. For example, if we have equally doped and sized GaAs and Si based diodes, can we tell which will have the higher currents? While this depends weakly on the diffusion and collision rates, it is primarily set by the intrinsic densities of the materials. For Si $n_i = 1.5$ E10 cm^{-3}, while for GaAs $n_i = 2$ E6 cm^{-3}. Thus the current in the Si diode will have significantly higher currents – by ~8 orders of magnitude!

Carrier currents

We can also use our equations to look at the carrier currents.

We know from before that the minority carrier current is given by:

$$J_N = qD_n \frac{\partial \Delta n_p(x)}{\partial x}$$

$$= \frac{qD_n}{L_n} \Delta n_p(x)$$

$$J_P = -qD_p \frac{\partial \Delta p_n(x)}{\partial x}$$

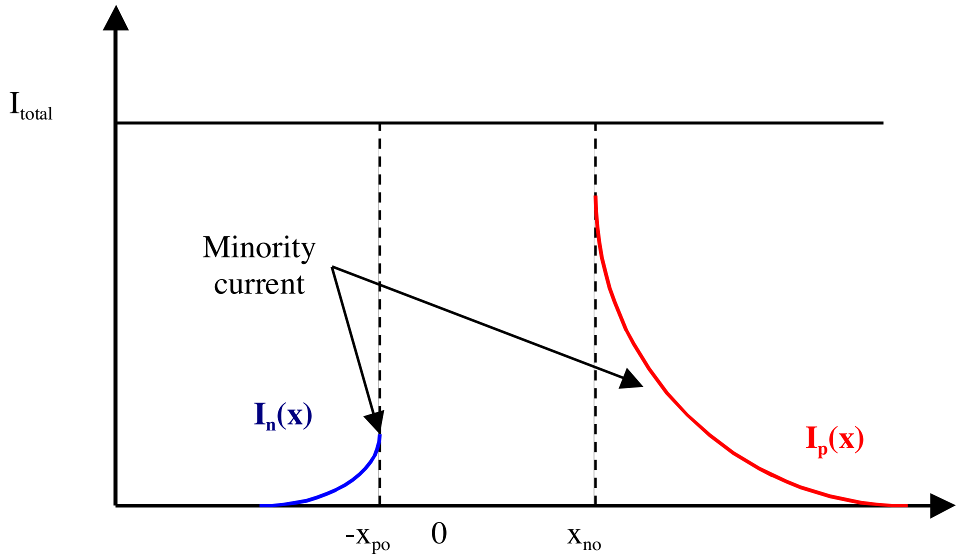
$$= \frac{qD_p}{L_p} \Delta p_n(x)$$

Further we know that the injected minority carrier concentrations drop off as:

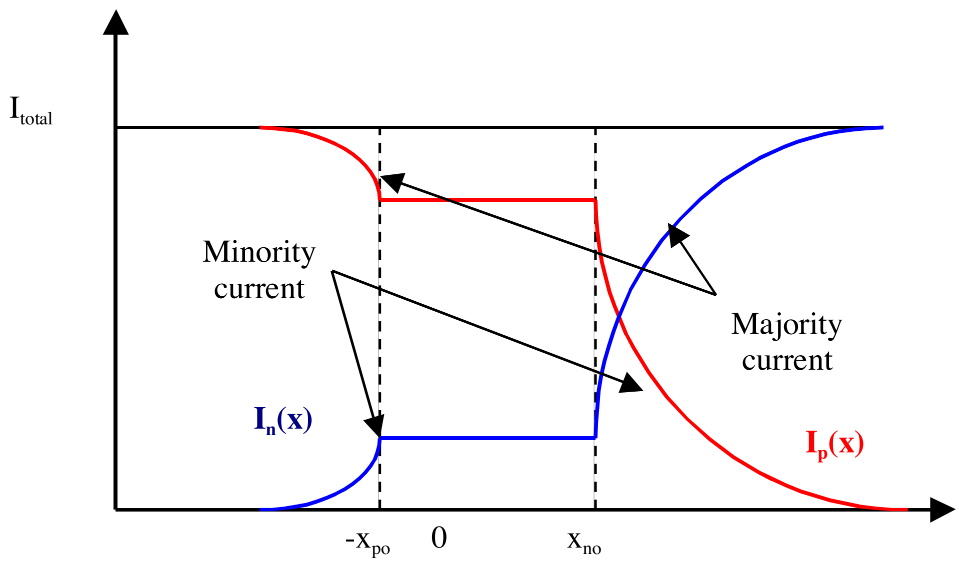
$$\Delta n_p = \frac{n_i^2}{N_A}\left(e^{qV_A/kT} - 1\right)e^{+(x+x_{p0})/L_n}$$

$$\Delta p_n = \frac{n_i^2}{N_D}\left(e^{qV_A/kT} - 1\right)e^{-(x-x_{n0})/L_p}$$

Thus we would expect a minority carrier current to look like:



Because the total current must be constant then we can come up with the majority carrier currents



This of course assumes a forward bias: If we instead look at reverse bias, what would we expect? Now, the minority carrier concentration must be greatly reduced – as it is only due to the diffusion across the junction. Further the direction of the current must be opposite to that which it had before. The pertinent equations are:

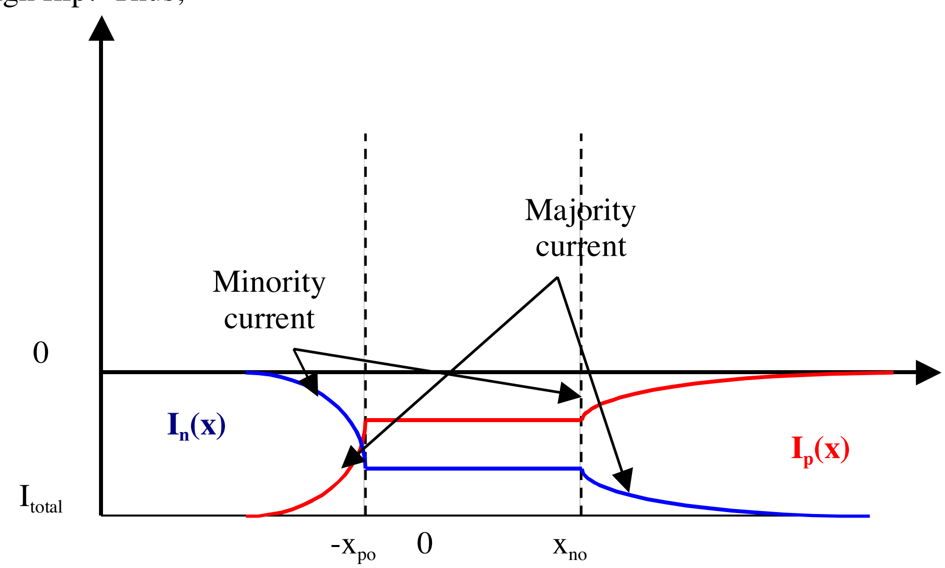
$$\Delta n_p = \frac{n_i^2}{N_A}\left(e^{qV_A/kT}-1\right)e^{+(x+x_{p0})/L_n}$$

$$\approx -\frac{n_i^2}{N_A}e^{+(x+x_{p0})/L_n}$$

$$\Delta p_n = \frac{n_i^2}{N_D}\left(e^{qV_A/kT}-1\right)e^{-(x-x_{n0})/L_p}$$

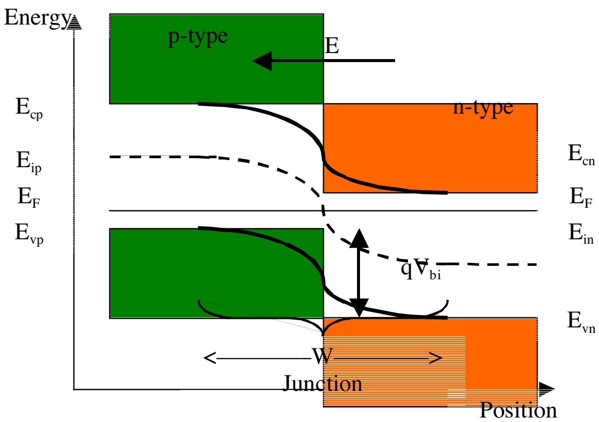
$$\approx -\frac{n_i^2}{N_D}e^{-(x-x_{n0})/L_p}$$

This means that outside of the change in size of the current (by a magnitude of $\sim e^{qV_A/kT}$) we have a sign flip. Thus,

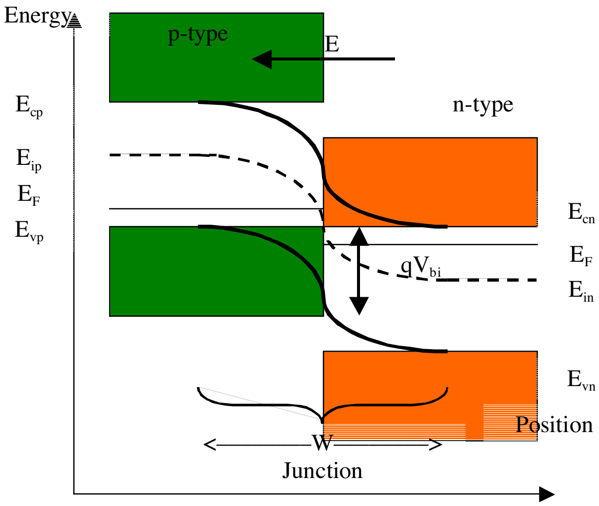


ZENER BREAKDOWN

If we reverse bias the diode sufficiently, it will go into a 'breakdown' mode. This breakdown mode does not mean that we have necessarily damaged the diode – which is possible – but rather we are causing another important process to occur. That process is known as tunneling. Zener diodes are created by heavily doping both sides of the junction. (If the diode is moderately or weakly doped then other things occur.) In such a situation we get an energy diagram that looks like:



Now we can apply a moderate reverse bias:



Now we have the situation in which the top valance band in the p-side is at the same – or similar – energy as the conduction band in the n-side. In such a case, we can have direct tunneling from the valance band of the p-side to the conduction band of the n-side. A second way to think about this is that the electric field in the junction is so strong that it ‘field’ ionizes an atom in the p-side. This requires an electric field on the order of 10^6 V/cm – hence the junction length needs to be very short.

Avalanche Breakdown

If we have not heavily doped the two sides then the junction, then it can be fairly wide. Remember:

$$w = x_{n0} + x_{p0}$$

$$= \left(\frac{2\varepsilon (V_{bi} - V_A)(N_A + N_D)}{qN_A N_D} \right)^{1/2}$$

This means that the distance that the electron needs to travel to get across the junction can be significant. The longer the distance the higher the likelihood that it will collide with lattice atoms. If the electron collides with sufficient energy, it can ionize a lattice atom, e.g. create and EHP. The old electron and hole are now accelerated by the electric field to high energies – until they collide with additional atoms, producing more EHPs. This can result in a significant ionization of the atoms that make up the junction – not unlike a lightning strike through air. This is known as Avalanche Breakdown.

Diodes under AC and transient conditions

An exact analysis of transient behavior across a diode is a non-trivial endeavor. It would take much more time than we have and require applications of non-trivial mathematical functions. Further, it would not help the students develop a physical understanding of the transient behavior of diodes. Thus, we will make a number of simplifying approximations that will lead us to approximately the correct solutions.

We know from our analysis of the steady state solution that the junction width is dependent on the applied bias.

$$w = x_{n0} + x_{p0}$$
$$= \left(\frac{2\varepsilon(V_{bi} - V_A)(N_A + N_D)}{qN_AN_D}\right)^{1/2}$$

Further we know that the injected charges depend on the applied bias:

$$\Delta n_p = \frac{n_i^2}{N_A}\left(e^{qV_A/kT} - 1\right)e^{+(x+x_{p0})/L_n}$$

$$\Delta p_n = \frac{n_i^2}{N_D}\left(e^{qV_A/kT} - 1\right)e^{-(x-x_{n0})/L_p}$$

Likewise the total inject charge depends on the applied bias:

$$Q_p = qA\int_{x_{n0}}^{\infty} \Delta p_n(x)dx$$
$$= qA\frac{L_pn_i^2}{N_D}\left(e^{qV_A/kT} - 1\right)$$
$$= qAL_pp_{n0}\left(e^{qV_A/kT} - 1\right)$$

$$Q_n = qA\int_{-x_{p0}}^{-\infty} \Delta n_p(x)dx$$
$$= qA\frac{L_nn_i^2}{N_A}\left(e^{qV_A/kT} - 1\right)$$
$$= qAL_nn_{p0}\left(e^{qV_A/kT} - 1\right)$$

Let us now define the current. We will break it up into the time-independent and time-dependent parts

i_D = Total current through the diode

I_D = DC current through the diode

i_d = AC current through the diode

$$i_D(t) = I_D + i_d(t)$$

We will do the same thing for the applied voltages

v_A = Total voltage across the diode
V_A = DC voltage across the diode
v_a = AC voltage across the diode
$v_A(t) = V_A + v_a(t)$

Now let us think about the current. The ac part of the current is simply the time derivative of the total charge in the diode. The dc part can be written a number of ways but one simple way is to use the charge divided by the recombination time. (This is derived above.)

From above:

$$I_{P_{\min or}}(x) = A\frac{qD_p}{L_p}\Delta p_n(x)$$
$$= A\frac{qD_p}{L_p}\frac{n_i^2}{N_D}\left(e^{qV_A/kT} - 1\right)\left[e^{-(x-x_{n0})/L_p}\right]$$

but the total charge is

$$Q_p = qA\int_{x_{n0}}^{\infty}\Delta p_n(x)dx$$
$$= qA\frac{n_i^2}{N_D}\left(e^{qV_A/kT} - 1\right)\int_{x_{n0}}^{\infty}\left[e^{-(x-x_{n0})/L_p}\right]dx$$
$$= -qA\frac{L_p n_i^2}{N_D}\left(e^{qV_A/kT} - 1\right)\left[e^{-(x-x_{n0})/L_p}\right]_{x_{n0}}^{\infty}$$
$$= qA\frac{L_p n_i^2}{N_D}\left(e^{qV_A/kT} - 1\right)$$

plugging this into our total hole current, e.g. the minority current at the junction edge, we find

$$I_{P_{total}} = I_{P_{\min or}}(x = x_{n0}) = A\frac{qD_p}{L_p}\Delta p_n(x = x_{n0})$$
$$= A\frac{qD_p}{L_p}\frac{n_i^2}{N_D}\left(e^{qV_A/kT} - 1\right)$$
$$= \frac{D_p}{L_p^2}\left(qA\frac{L_p n_i^2}{N_D}\left(e^{qV_A/kT} - 1\right)\right)$$
$$= \frac{D_p}{D_p\tau_p}(Q_p)$$
$$= \frac{Q_p}{\tau_p}$$

Likewise, we can show

$$I_{n_{total}} = \frac{Q_n}{\tau_n}$$

Noting of course that the above applies to the DC part of our current. (This is also known as the 'recombination' current.)

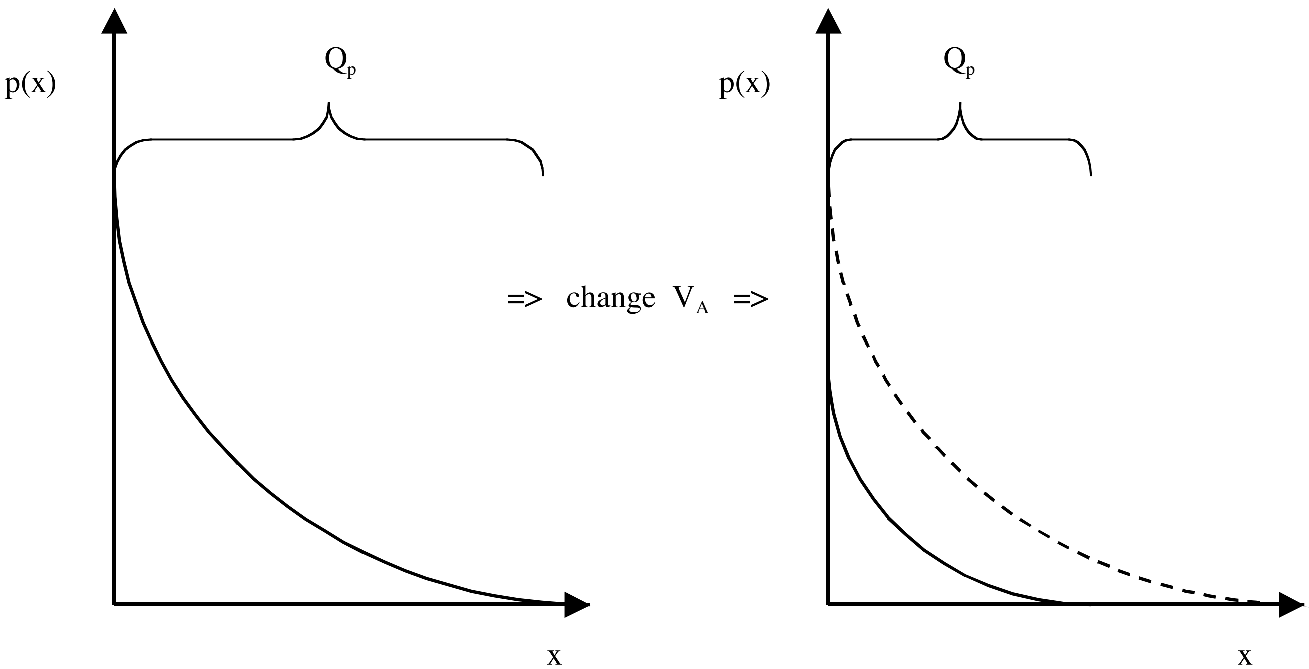
Thus, we find that:

$$i_D(t) = \frac{Q(t)}{\tau} + \frac{dQ(t)}{dt}$$

Example

Assume that we have a p^+-n junction ($N_A \gg N_D$).

In this case all of the action is dominated by what happens on the n-side of the junction. The excess hole density will look like:



Thus,

$$i_D(t) \approx i_p(t) = \frac{Q_p(t)}{\tau_p} + \frac{dQ_p(t)}{dt}$$

Now let us suppose that we start with a forward biased diode, ($I = I_F$) but at t = 0 we clamp the current to zero.

$$i_D(t) = \begin{cases} I_F & t < 0 \\ 0 & t \geq 0 \end{cases}$$

For t < 0, we are in steady state. Thus

$$
\begin{aligned}
i_D(t) &= I_F \\
&= \frac{Q_p(t)}{\tau_p} + \frac{dQ_p(t)}{dt} \\
&= \frac{Q_p}{\tau_p} \\
&\Downarrow \\
Q_p &= \tau_p I_F
\end{aligned}
$$

For $t \geq 0$, the total current is zero. Thus

$$
\begin{aligned}
i_D(t) &= 0 \\
&= \frac{Q_p(t)}{\tau_p} + \frac{dQ_p(t)}{dt} \\
&\Downarrow
\end{aligned}
$$

$$\frac{dQ_p(t)}{dt} = -\frac{Q_p(t)}{\tau_p}$$

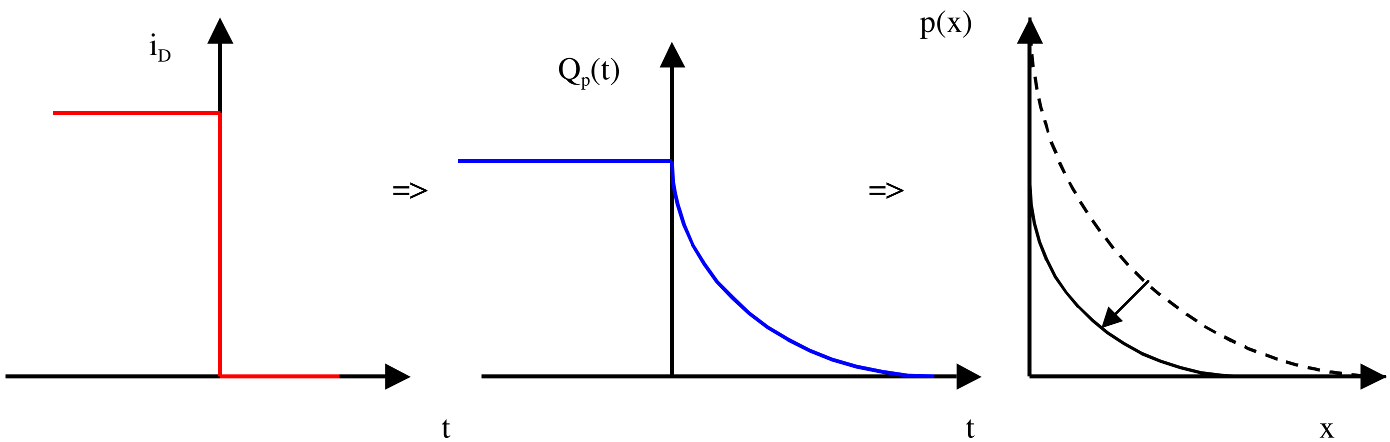
We can now integrate to get

$$
\begin{aligned}
Q_p(t) &= \int \frac{dQ_p(t)}{dt} dt \\
&= A \exp\left[-t / \tau_p\right]
\end{aligned}
$$

We now match the solutions at the boundary (t = 0) to find

$$Q_p(t) = \begin{cases} I_F \tau_p & t < 0 \\ I_F \tau_p \exp\left[-t / \tau_p\right] & t \geq 0 \end{cases}$$

A plot of this shows



While I have drawn the spatial distribution of the holes as an exponential decay (e^{-ax}) it does not look quite like that in a real device. The loss rate at a given point depends on both recombination and diffusion. (Our simple model only accounts for recombination.) The junction side of the distribution has greater diffusion than the contact side. (Remember, you had a much higher diffusion rate at the junction before but you also had the electric field to balance that lose rate. That electric field – and hence balancing current – is not there any more.) Thus one would find a non-exponential profile to the charge distribution. (In fact, you would find that the slope of the spatial distribution goes to zero right at the junction edge.)

What we have found is that even though we can force an instantaneous change in the diode current, we can not force an instantaneous change in the charge distribution. (This should come as no surprise, as in will take some time to move the charges to their new locations.) Because we can not move the charges instantaneously, the voltage across the junction can not change instantaneously either.

Thus the time varying junction voltage depends on the recombination rate of the charge carriers on both sides. We can calculate an approximate value for the temporal voltage profile from the total change.

$$Q_p(t) = \begin{cases} I_F\tau_p & t < 0 \\ I_F\tau_p \exp\left[-t/\tau_p\right] & t \geq 0 \end{cases}$$
$$= qAL_p\Delta p_n(x = x_{n0})$$

If we assume that the spatial profile of the holes follow an exponential profile then we get

$$Q_p(t > 0) = I_F\tau_p \exp\left[-t/\tau_p\right]$$
$$= qAL_p\Delta p_n(x = x_{n0})$$
$$= qAL_p p_{n0}\left(e^{qv_A/kT} - 1\right)$$

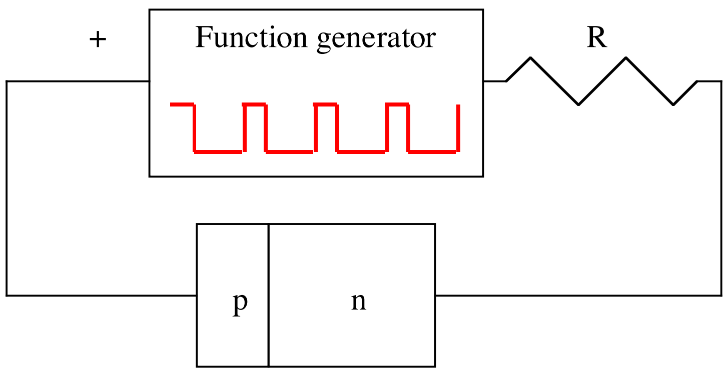
[The same approximation will hold for the electrons.] Thus,

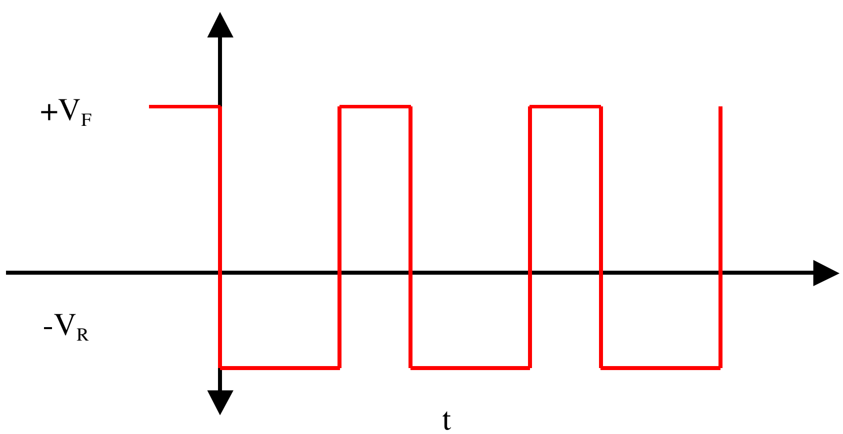
$$v_A = \frac{kT}{q}\ln\left(\frac{I_F\tau_p}{qAL_p p_{n0}}\exp\left[-t/\tau_p\right] + 1\right)$$

[Remember that this is a rough approximation!]

What we have found is that the voltage lags behind the current. This is indicative of a capacitive type of device! To try to understand this, we will first look at 'switching' in diodes. In particular, we will look at turn off and turn on. (These are known as 'turn-off transient' and 'storage decay time', for reasons that should be obvious soon.)

Let us assume that we are supplying our diode with a function generator. The bias supplied looks like





We will further assume that the applied biases are much larger than the self bias (V_{bi}).

Toward the end of the forward bias phase of the voltage, most of the bias is dropped across the resistor. This is because it takes very little forward voltage to allow us to draw a current. Further the current through the resistor must equal that through the diode. Then

$$\begin{aligned} V_F &= V_{F-diode} + V_{F-resistor} \\ &\approx V_{F-resistor} \end{aligned}$$

$$\Downarrow$$

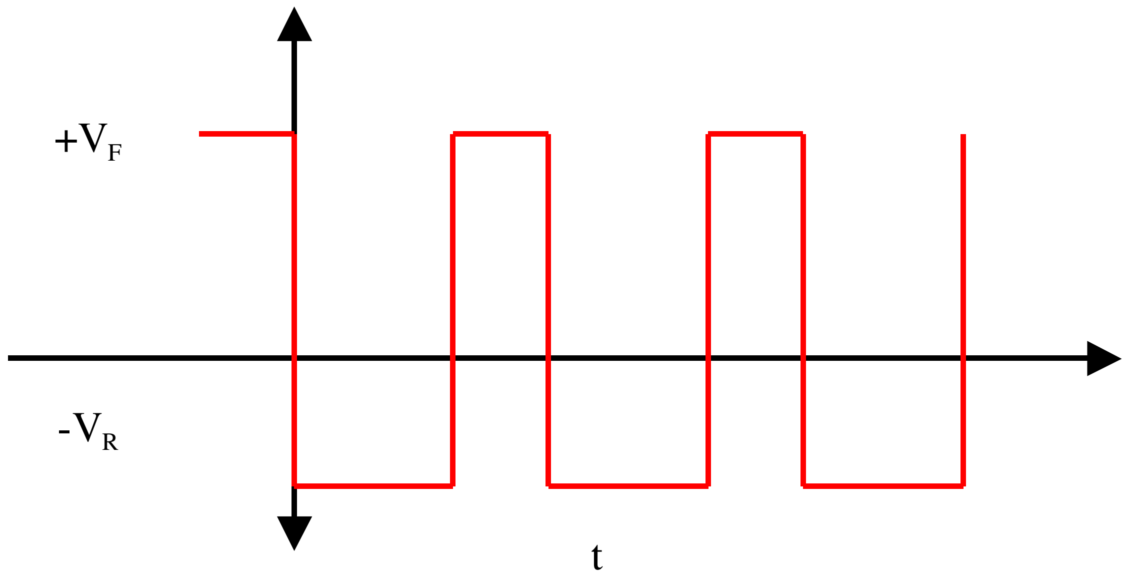
$$I_F = \frac{V_F}{R}$$

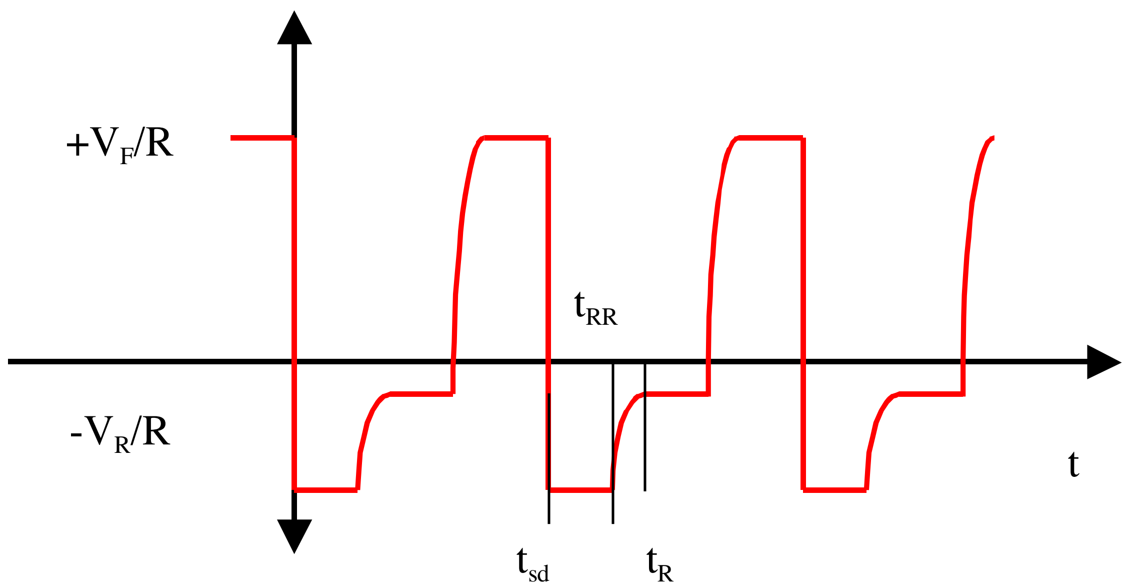
Further, the injected charge distributions will be near their steady state values, and thus spatially fall of as an exponential.

When the voltage source switches sign, a number things occur:

1) The excess holes/electrons do not go away instantaneously
2) The voltage across the junction remains ~ small until the excess hole concentration falls to the new equilibrium condition. [i.e. it acts as if it where still forward biased until the excess holes/electrons are used up.]
3) The current remains large – on the order of what it was before we switched the voltage

4) The excess holes decay by both recombination and by being swept back to the p^+ side.





Here t_{sd} is the storage delay time, t_r is the recovery time and t_{RR} is the total recovery time.

The book calculates the storage delay time as

$$t_{sd} = \tau_p \ln\left[\frac{I_f + I_r}{I_r}\right]$$

where I_r is the current in the reverse direction and I_f is the forward current. A more rigorous calculation shows that the storage delay time is

$$t_{sd} = \tau_p \left(\mathrm{erf}^{-1}\left[\frac{I_r}{I_f + I_r} \right] \right)^2$$

where erf is known as the 'error function' and erf^{-1} is the 'inverse error function'. These are rather ugly integrals that cannot be solved analytically.

For your information (do not memorize this!)

$$\mathrm{erf} = \frac{1}{2\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

To reduce the storage delay time one can do a number things

1) Reduce the recombination time by adding traps etc.
2) Make the currents smaller. What this really does is:
    a. Make the n-region short - thereby cutting off Q
    b. Don't forward bias too hard, thus reducing the current ratios

Turn on transient

For this, we assume that we are driving a fixed current through the device, I_F after t = 0. (Before t = 0 the device might be unbiased or reversed bias. Either way, that current level is so small that it does not play a major role in the result.) Here, after t = 0 we are not only driving our current through but we are also injecting holes, from the p^+ side in to the n-side. Therefore

$$\frac{dQ_p(t)}{dt} = I_F - \frac{Q_p(t)}{\tau_p}$$

This equation comes directly from

$$i_D(t) \approx i_p(t) = \frac{Q_p(t)}{\tau_p} + \frac{dQ_p(t)}{dt}$$

, where i_D is the total current through the device.

We will assume that I_F is a constant and is the final current. (In our system, as shown above, $I_F = V_F/R$.) Thus,

$$\frac{dQ_p(t)}{dt} = I_F - \frac{Q_p(t)}{\tau_p}$$

$$\Downarrow$$

$$\int_0^t dt = \int_0^Q \left(I_F - \frac{Q_p(t)}{\tau_p} \right)^{-1} dQ_p$$

$$\Downarrow$$

$$t = \tau_p \int_0^Q \left(\frac{1}{1 - Q_p(t)/\tau_p I_F} \right) d\left(Q_p/\tau_p I_F\right)$$

$$\Downarrow$$

$$t = -\tau_p \ln\left(1 - Q_p(t)/\tau_p I_F\right)$$

$$\Downarrow$$

$$e^{-t/\tau_p} = 1 - Q_p(t)/\tau_p I_F$$

$$\Downarrow$$

$$Q_p(t) = \tau_p I_F \left(1 - e^{-t/\tau_p}\right)$$

Now let us assume that the junction voltage can be determined from the total injected charge. (This is known as the quasi-static steady state approximation.) Thus

$$\begin{aligned} Q_p(t > 0) &= \tau_p I_F \left(1 - e^{-t/\tau_p}\right) \\ &= qAL_p \Delta p_n\left(x = x_{n0}\right) \\ &= qAL_p p_{n0}\left(e^{qv_A/kT} - 1\right) \\ &= I_0 \tau_p \left(e^{qv_A/kT} - 1\right) \end{aligned}$$

Rewriting this, we find

$$\Downarrow$$

$$I_F\left(1 - e^{-t/\tau_p}\right) = I_0\left(e^{qv_A/kT} - 1\right)$$

$$\Downarrow$$

$$e^{qv_A/kT} = \frac{I_F}{I_0}\left(1 - e^{-t/\tau_p}\right) + 1$$

$$\Downarrow$$

$$v_A = \frac{kT}{q} \ln\left[\frac{I_F}{I_0}\left(1 - e^{-t/\tau_p}\right) + 1\right]$$

(This happens to be very fast.)

Diode capacitance/resistance

Now that we know how the voltage changes with time, or with current, or with charge, we can begin to develop a capacitance/resistance model for diodes. We will have to do this for both forward and reverse biases.

1) Reverse bias:
    a. 'Depletion layer capacitance' – due to changes in the width of the junction, w, with the applied voltage, v_A. i.e. the change in the charge with v_A.
    b. Here conduction is effectively zero, realize that very little current flows in reverse bias, so conduction is very low.
2) Forward bias:
    a. "Charge storage capacitance" dominates, due to change in minority charge vs v_A at low frequencies. (Note that our model only real works at low frequencies – it assumes quasi-neutrality.)
    b. Conductance is primarily due to minority carrier injection.

Let us first look at Kirchoff's Law.

$$\begin{aligned} V &= L\ddot{Q} + R\dot{Q} + C^{-1}Q \\ &= L\ddot{Q} + G^{-1}\dot{Q} + C^{-1}Q \end{aligned}$$

where R is the resistance, G is the conduction and C is the capacitance.
Thus, if we only have capacitance,

$$C = \frac{dQ}{dV}$$

This what happens in reverse bias, so let us start with that:

Reverse Bias Capacitance

We know from several weeks ago that the total charge in the junction is zero and thus,

$$\begin{aligned} Q_p &= Aqx_{n0}N_D \\ &= AqN_D w \frac{N_A}{(N_A + N_D)} \\ &= -Q_n \\ &= Aqx_{p0}N_A \\ &= AqN_A w \frac{N_D}{(N_A + N_D)} \end{aligned}$$

where

$$\begin{aligned} w &= x_{n0} + x_{p0} \\ &= \left(\frac{2\varepsilon(V_{bi} - v_A)(N_A + N_D)}{qN_AN_D} \right)^{1/2} \end{aligned}$$

Now let us calculate the capacitance

$$C = \frac{dQ}{dV}$$

$$= \frac{d}{dV}\left[Aqw\frac{N_A N_D}{(N_A + N_D)}\right]$$

$$= Aq\frac{N_A N_D}{(N_A + N_D)}\frac{d}{dV}[w]$$

$$= Aq\frac{N_A N_D}{(N_A + N_D)}\frac{d}{dV}\left[\left(\frac{2\varepsilon(V_{bi} - v_A)(N_A + N_D)}{qN_A N_D}\right)^{1/2}\right]$$

$$= A\left(2\varepsilon\frac{qN_A N_D}{(N_A + N_D)}\right)^{1/2}\frac{d}{dV}\left[(V_{bi} - v_A)^{1/2}\right]$$

$$= \frac{A}{2}\left(2\varepsilon\frac{qN_A N_D}{(N_A + N_D)(V_{bi} - v_A)}\right)^{1/2}$$

$$= \varepsilon A\left(\frac{qN_A N_D}{2\varepsilon(N_A + N_D)(V_{bi} - v_A)}\right)^{1/2}$$

$$= \frac{\varepsilon A}{w}$$

$$\equiv C_j$$

(C_j is the junction capacitance.) Here we see that the capacitance varies weakly on the applied (reverse) voltage – but this is not significant. Now, let us look at the conductance:

$$G_0 = \frac{dQ}{dV} = \frac{dI}{dV}$$

$$= \frac{d}{dV}\left[I_0\left(e^{qv_A/kT} - 1\right)\right]$$

$$= \frac{q}{kT}I_0\left(e^{qv_A/kT}\right)$$

$$= \frac{q}{kT}(I + I_0)$$

$$\approx 0 \text{ in reverse bias}$$

(This of course assumes a 'perfect' or 'ideal' diode – usually this is a reasonable estimate.)

Forward Bias (again low frequency)

Now what is the charge? It is no longer <u>just</u> what we had before – in fact we now need to look at the charge outside of the junction. To make our life easier, let us assume that we have a p^+-n junction. For forward bias, the charge is

$$
\begin{aligned}
Q_p &= I\tau_p \\
&= \tau_p\left[I_0\left(e^{qv_A/kT}-1\right)\right] \\
&= \tau_p\left[\frac{qAL_p p_{n0}}{\tau_p}\left(e^{qv_A/kT}-1\right)\right] \\
&= qAL_p p_{n0}\left(e^{qv_A/kT}-1\right) \\
&\approx qAL_p p_{n0} e^{qv_A/kT}
\end{aligned}
$$

Now the charge storage capacitance is

$$
\begin{aligned}
C &= \frac{dQ}{dV} \\
&= \frac{d}{dV}\left[qAL_p p_{n0}\left(e^{qv_A/kT}-1\right)\right] \\
&= \frac{q}{kT} qAL_p p_{n0} e^{qv_A/kT} \\
&\approx \frac{q}{kT} Q_p \\
&= \frac{q}{kT} I\tau_p\left[= \frac{q}{kT}(I-I_0)\tau_p \text{ - if done without approximations}\right] \\
&\equiv C_S
\end{aligned}
$$

The conduction is the same as above, as we have the same equation for the current. Thus we see that $C_S \approx G_0\tau_p$

For high frequency biasing, the equations become approximately

$$G_{FB} = \frac{G_0}{\sqrt{2}}\left(\sqrt{1+\omega\tau_p}+1\right)^{1/2}$$

$$C_S = \frac{G_0}{\omega\sqrt{2}}\left(\sqrt{1+\omega\tau_p}-1\right)^{1/2}$$

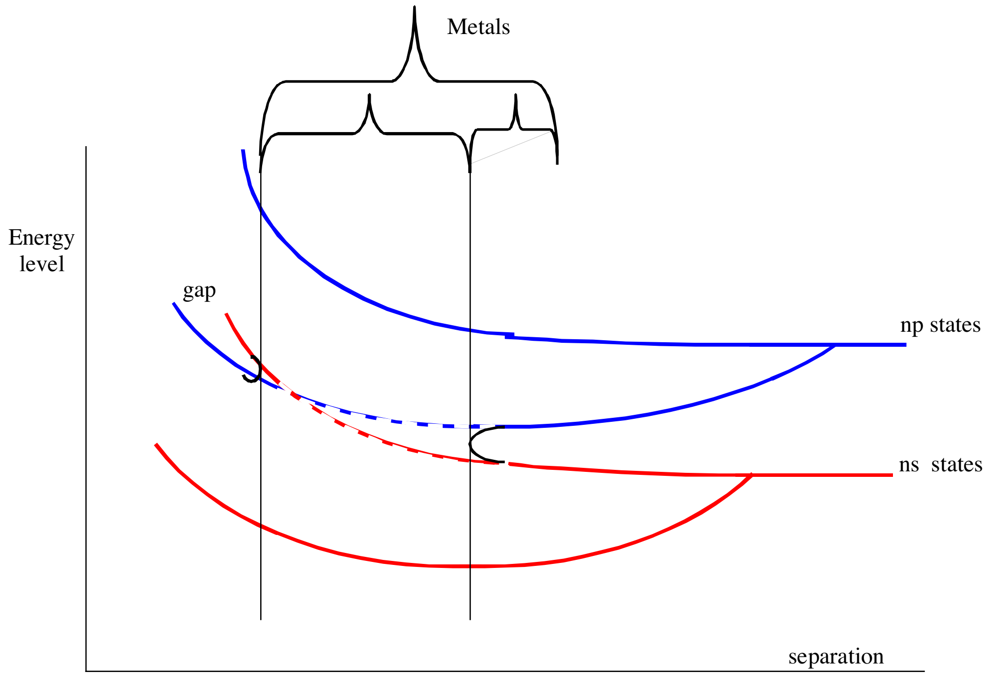
Finally, for small signal responses,

$$i_a = G_{FB}v_a + C_S\frac{dv_a}{dt}$$

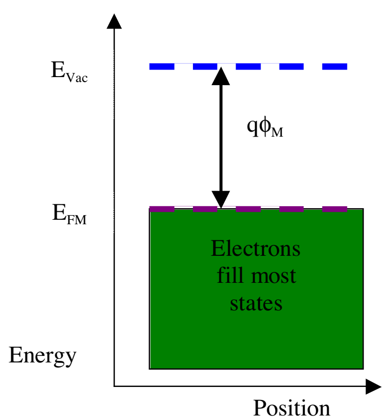
Metal-Semiconductor Junctions

Thus far, we have only considered the insides of our device. However, to use it we must connect it to the outside world. This means that we have additional junctions, notably the metal contact with each end of the device. What happens at these junctions and do they affect our results? (If they didn't, would I be asking the question?)

Let us first look at what the energy diagram might look like for a metal.

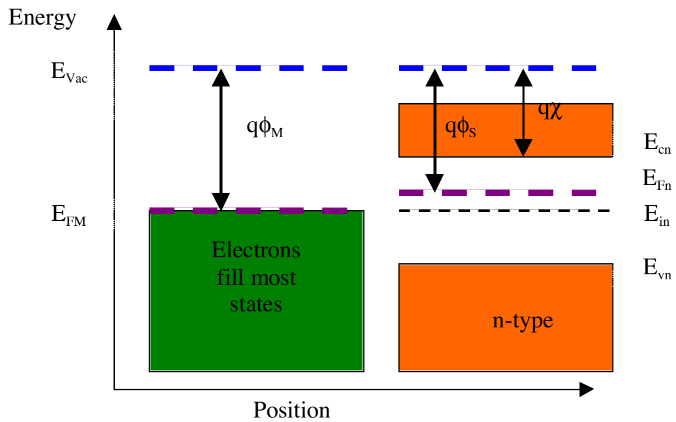


This implies that the Fermi level is energy level at which all of the states would be full at T = 0K. (As with the other materials, a higher temperature results in some of the electrons moving to higher energy levels.) To remove an electron from the metal, thus requires that we move an electron from the Fermi level to the vacuum potential. (It is common to set the vacuum potential to zero – but that is just a reference point. We will not do that here.) We know from very early experiments, that metals have a work function for removing an electron – this is what was found from shining a light on metals. Thus, we can draw the energy diagram for metal as:



where ϕ_M is the metal work function.

Now let us push our metal next to our semiconductor material.



As with p-n junctions, we need the Fermi level to be constant if there is no current flow. This means that we have a new junction at the contact point. There are two different types of junctions:

a) Shottky diode (rectification) (Shottky diodes are majority carrier devices)

$$\Downarrow$$

$$(\phi_M > \phi_S,\ \ n-\text{type})$$

$$(\phi_M < \phi_S,\ \ p-\text{type})$$

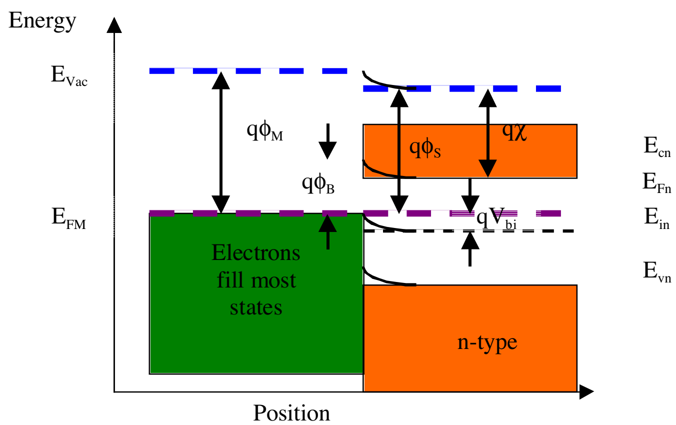
b) Ohmic contact (small resistance)

$$\Downarrow$$

$$(\phi_M < \phi_S,\ \ n-\text{type})$$

$$(\phi_M > \phi_S,\ \ p-\text{type})$$

Let us now put them in contact



There are two potentials that we need to consider. First, the bias between the two materials is given by the difference in the intrinsic energy – or the shift in the vacuum bias. This is simply

$$qV_{bi} = q\phi_M - q\phi_S.$$

The other energy is the amount of energy that an electron must gain (or lose) in order to move from the metal to the semiconductor. That energy is

$$q\phi_B = q(\phi_M - \chi).$$

The width of the junction is simply the width of the n-side – there is no p-side in this case. Thus

$$w = x_{n0} + \underbrace{x_{p0}}_{=0}$$

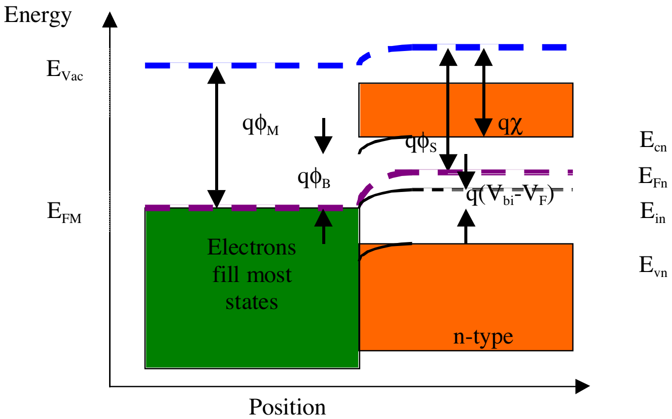
$$= \left(\frac{2\varepsilon\left(V_{bi} - v_A\right)}{qN_D} \right)^{1/2}$$

For metal-p-type junctions:

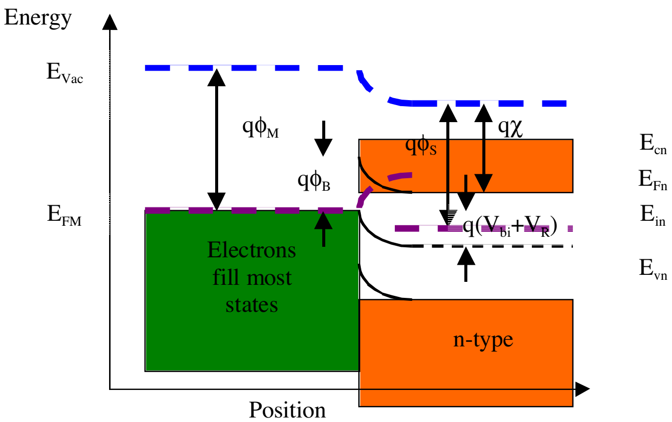
$$w = \underbrace{x_{n0}}_{=0} + x_{p0}$$

$$= \left(\frac{2\varepsilon\left(V_{bi} - v_A\right)}{qN_A} \right)^{1/2}$$

Now let us forward bias the junction



and now reverse bias it



Note that the edge of the conduction and valance band must follow the intrinsic energy level. Now let us look at what the minority and majority carriers do, i.e. how do they move from side to side.

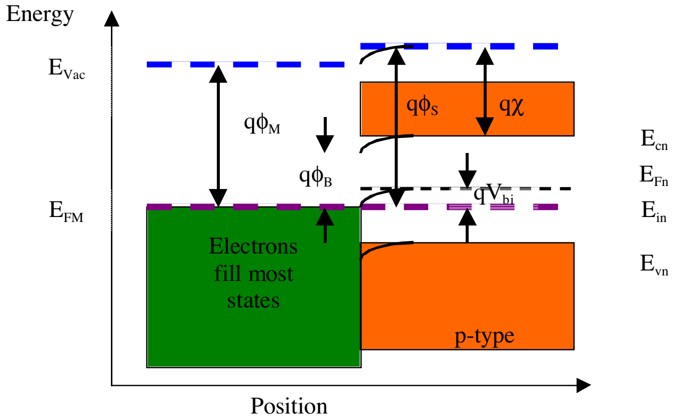
The majority carriers in the metal are the electrons.

1) Electron drift from the metal to the semiconductor is slow because of the voltage between the Fermi level in the metal and the conduction band energy in the semiconductor, ϕ_B.
2) Hole diffusion from the metal to the semiconductor is very small because there are very few holes in the metal.

The majority carriers in the Semiconductor are the electrons (But they could be holes – i.e. in p-type material)

3) For our example, n-type, the hole drift is very small because they are the minority carrier.
4) Electron diffusion can be significant.
   - ⇨ We have a majority carrier device

Now doing the same thing for p-type



5) For our example, p-type, the electron drift is very small because they are the minority carrier.
6) Hole diffusion can be significant.
   - ⇨ Again, we have a majority carrier device

If this does not make sense, draw stacks of charge carriers and look at it again.

Looking at the expected current flow we find

Flow	Equilibrium	Reverse	Forward	Set by
S -> M	←	◀	⟵	$q(V_{bi}-V_A)$
M -> S	→	→	→	$q(\phi_B)$

This very similar to what we saw for standard p-n junctions. For the same reasons, we would expect to see currents that are constant for reverse bias and exponentially increasing for forward bias. Thus, we would expect

$$I_{total} = I_{forward} + I_{reverse}$$
$$= I_f e^{qv_A/kT} + I_r$$

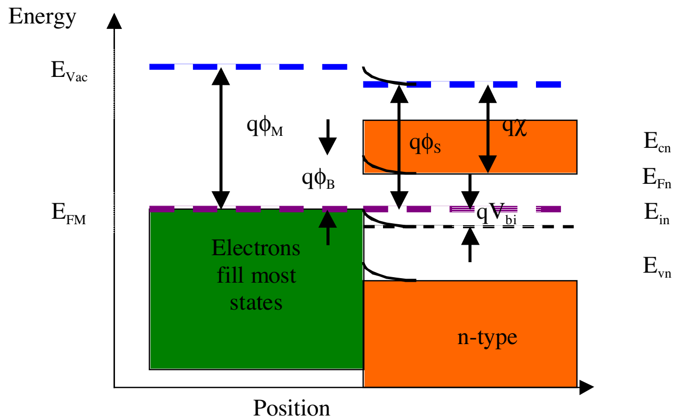
With no applied bias we would expect no current. Thus

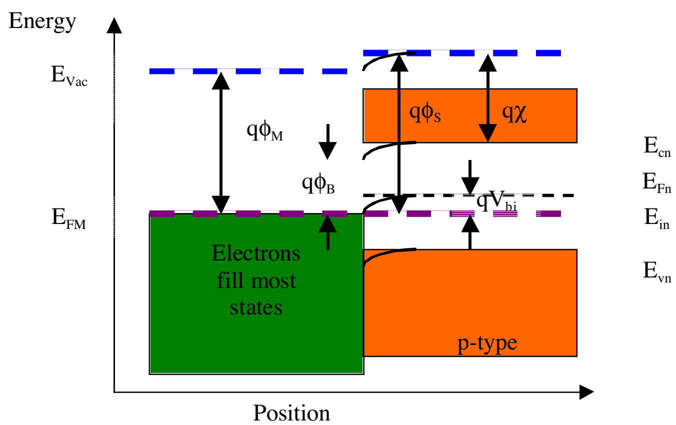
$$\Downarrow$$
$$I_f = -I_r \equiv I_0$$
$$\Downarrow$$
$$I_{total} = I_0\left(e^{qv_A/kT} - 1\right)$$

Pictures of all of the possible contacts are

$\left(\phi_M > \phi_S,\ \ n-\text{type}\right)$

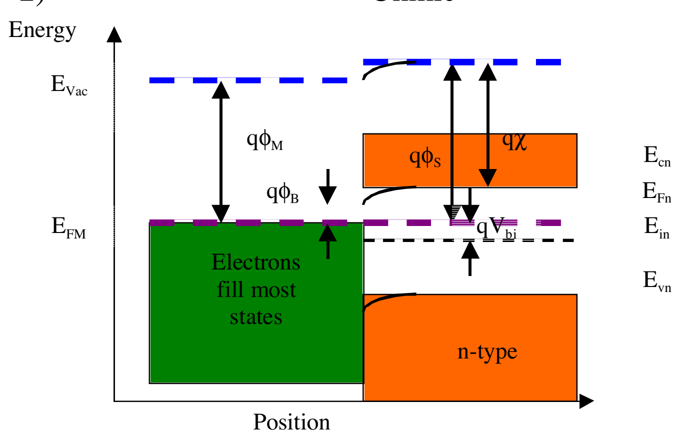
1) $\left(\phi_M < \phi_S,\ \ p-\text{type}\right)$ - Shottky



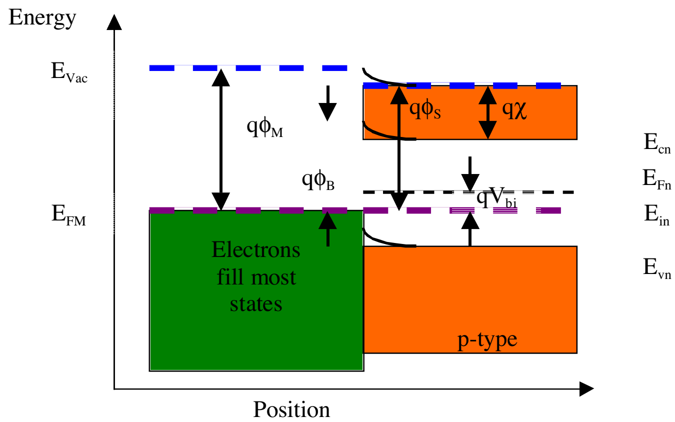


$\left(\phi_M < \phi_S,\ \ n-\text{type}\right)$

2) $\left(\phi_M > \phi_S,\ \ p-\text{type}\right)$ - Ohmic



Not a barrier to electron movement – just some small resistance -> Ohmic contact



No real barrier to hole movement – just some small resistance -> Ohmic contact.

Finally, there are two practical issues:

1) I can make any junction act ohmic. This is necessary as some materials make it tough to obtain ohmic contact otherwise. The trick is very heavy doping. Then the junction width is very narrow and tunneling occurs. Thus all junctions are typically heavily doped.
2) The surface effects often 'pin' E_F, which can alter the work function of the metal or χ for the semiconductor. So in general the values are empirically determined from the device characteristics.

Example

Si diode:

p-side
$N_A = 3E18\ cm^{-3}$
$\tau_n = 0.1\ \mu s = \tau_p$
$\mu_p = 100\ cm^2/V\text{-}s$
$\mu_n = 200\ cm^2/V\text{-}s$

n-side
$N_D = 4E15\ cm^{-3}$
$\tau_n = 12\ \mu s = \tau_p$
$\mu_p = 450\ cm^2/V\text{-}s$
$\mu_n = 1300\ cm^2/V\text{-}s$

Cross sectional area $A = 10^{-4}\ cm^2$
$n_i = 1.5E10\ cm^{-3}$.

1) Find V_{bi}.

$$V_{bi} = \frac{kT}{q}\ln\left[\frac{N_A N_D}{n_i^2}\right]$$
$$= 0.819V$$

2) Find the junction width at equilibrium

$$w = x_{n0} + x_{p0}$$
$$= \left(\frac{2\varepsilon\left(V_{bi} - v_A\right)\left(N_A + N_D\right)}{qN_D N_A}\right)^{1/2}$$
$$\varepsilon = \varepsilon_r \varepsilon_0 = 11.8(8,85E - 14F / cm)$$
$$\Downarrow$$
$$w = 5.17E - 5cm$$

3) Find I if the device is forward biased by 0.38 V

$$I = I_0\left(e^{qv_A/kT} - 1\right)$$
$$I_0 = qA\left[\frac{D_p}{L_p}p_{no} + \frac{D_n}{L_n}n_{po}\right]$$
$$= qA\left[\sqrt{\frac{D_p}{\tau_p}}p_{no} + \sqrt{\frac{Dn}{\tau_n}}n_{po}\right]$$
$$= qA\left[\sqrt{\frac{kT\mu_p}{q\tau_p}}p_{no} + \sqrt{\frac{kT\mu_n}{q\tau_n}}n_{po}\right]$$
$$= 9E - 16A$$
$$\Downarrow$$
$$I = 2.1E - 9A \text{ (or 2 nA)}$$

note that at 0.9 V the diode releases 1 A of current!

Useful equations

deBroglie momentum $p = h/\lambda = \hbar k$.

Heisenberg uncertainty principle

$$\Delta p \Delta x \geq \hbar$$

$$\Delta E \Delta t \geq \hbar$$

Photon energy $E_{photon} = h\nu$,

Bohr Model

$$r_n = \frac{K\hbar^2 n^2}{me^2} = 0.529\text{Å}n^2$$

$$E_{Bohr} = -\frac{1}{2}\frac{me^4}{K^2\hbar^2 n^2} = -13.56\text{eV}/n^2$$

Schrödinger's equation

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V\right)\Psi(\mathbf{r},t) = -\frac{\hbar}{j}\partial_t \Psi(\mathbf{r},t)$$

Maxwellian distribution

$$f(v) = n\left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left[\frac{-m(v)^2}{2kT}\right].$$

$$f(\mathcal{E}) = n\frac{1}{kT}\exp\left[\frac{-\mathcal{E}}{kT}\right]$$

Equation of motion

$$\mathbf{F}_n = -e\mathbf{E} = m_n^* \mathbf{a} \quad \text{(electron)}$$

$$\mathbf{F}_p = +e\mathbf{E} = m_p^* \mathbf{a} \quad \text{(hole)}$$

Fermi-Dirac function.

$$f(\mathcal{E}) = \frac{1}{1+\exp\left[\frac{(\mathcal{E}-\mathcal{E}_F)}{kT}\right]}$$

STATE DENSITY

$$N_c(\mathcal{E})d\mathcal{E} = \frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}-\mathcal{E}_c}\,d\mathcal{E} \quad \text{Conduction band}$$

$$N_v(\mathcal{E})d\mathcal{E} = \frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}_v-\mathcal{E}}\,d\mathcal{E} \quad \text{Valance band}$$

state distribution function.

$$n(\mathcal{E})d\mathcal{E} = f(\mathcal{E})N_c(\mathcal{E})d\mathcal{E} = \frac{1}{1+\exp\left[\frac{(\mathcal{E}-\mathcal{E}_F)}{kT}\right]}\frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}-\mathcal{E}_c}\,d\mathcal{E}$$ Electrons in the Conduction band

$$p(\mathcal{E})d\mathcal{E} = (1-f(\mathcal{E}))N_v(\mathcal{E})d\mathcal{E} = \left(1-\frac{1}{1+\exp\left[\frac{(\mathcal{E}-\mathcal{E}_F)}{kT}\right]}\right)\frac{\sqrt{2}}{\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\sqrt{\mathcal{E}_v-\mathcal{E}}\,d\mathcal{E}$$ Holes in the Valance band

total number of electrons and holes

$$n_0 \approx N_c \exp\left[\frac{-(\mathcal{E}_c-\mathcal{E}_F)}{kT}\right] = n_i \exp\left[\frac{(\mathcal{E}_F-\mathcal{E}_i)}{kT}\right]$$

$$p_0 \approx N_v \exp\left[\frac{(\mathcal{E}_v-\mathcal{E}_F)}{kT}\right] = n_i \exp\left[\frac{-(\mathcal{E}_F-\mathcal{E}_i)}{kT}\right]$$

$$n_i^2 = np$$

$$N_c = 2\left(\frac{m_n^* kT}{2\pi\hbar^2}\right)^{3/2}$$

$$N_v = 2\left(\frac{m_p^* kT}{2\pi\hbar^2}\right)^{3/2}$$

Intrinsic Energy

$$\mathcal{E}_i = \frac{(\mathcal{E}_v+\mathcal{E}_c)}{2} + \frac{kT}{2}\ln\left(\left(\frac{m_p^*}{m_n^*}\right)^{3/2}\right)$$

Mobility

$$\mu_n = -\frac{\langle v\rangle}{E} = \frac{\tau|q|}{m_n^*}$$

$$\mu_p = \frac{\langle v\rangle}{E} = \frac{\tau|q|}{m_p^*}$$

Conduction

$$\begin{aligned}\mathbf{J} &= nq\langle \mathbf{v}\rangle \\ &= q(n_0\mu_n + p_0\mu_p)\mathbf{E} \\ &= \sigma\mathbf{E} \\ &= \mathbf{E}/\rho\end{aligned}$$

Diffusion

$$D_n = \mu_n \frac{kT}{q_n}$$

$$D_p = \mu_p \frac{kT}{q_p}$$