

EE3310 Class notes – Part 3

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.

Homework Sets

Set 8 Due Tuesday Nov. 19th, 2002

Chapter 6 #: 1, 2, 3, 4, 5, 10, 11

Set 9 Due Thursday Nov. 21st, 2002

Chapter 6 #: 12, 15, 19, 20

Notes: Do not do the following:

#12 do not do the Boron dose part

#19 do not do the substrate (bulk) bias = - 2.5 V part

Set 10 Due Tuesday Nov. 26th, 2002

Chapter 7 #: 3, 6, 7, 8, 10, 11, 24

Solid State Electronic Devices - EE3310

Class notes

Transistors

Now we will take our p-n junctions and use them to make transistors. We will start with the Field Effect Transistor, FET. We do this mostly for historical reasons. (The first FET was patented in the 1920's and 30's!) The first modern FET, a Junction FET – JFET, was that produced by William Shockley, in 1952. This started the rapid growth in the field that you are now studying.

Before we start, we need to establish a few terms.

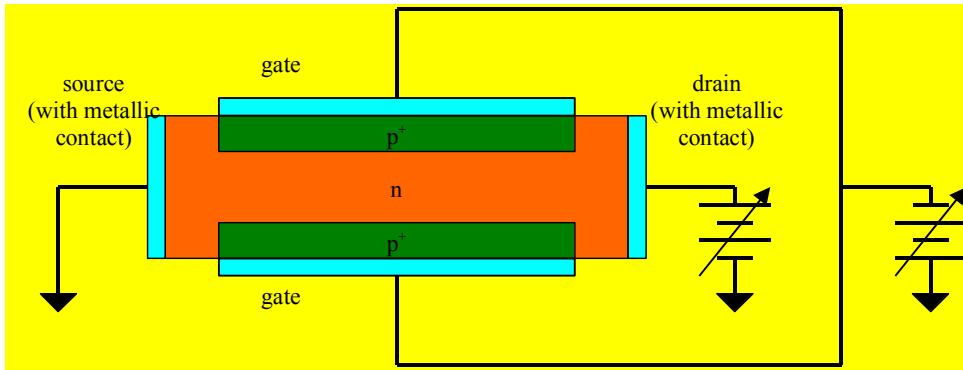
Source – This is where the charge carriers come from. By necessity this means the location at which majority carriers leave the metal contact and enter the device. Source also refers to the side of the device connected to that metal contact.

Drain – This is where the charge carriers leave the device. Again by necessity this is where the majority carriers leave the device.

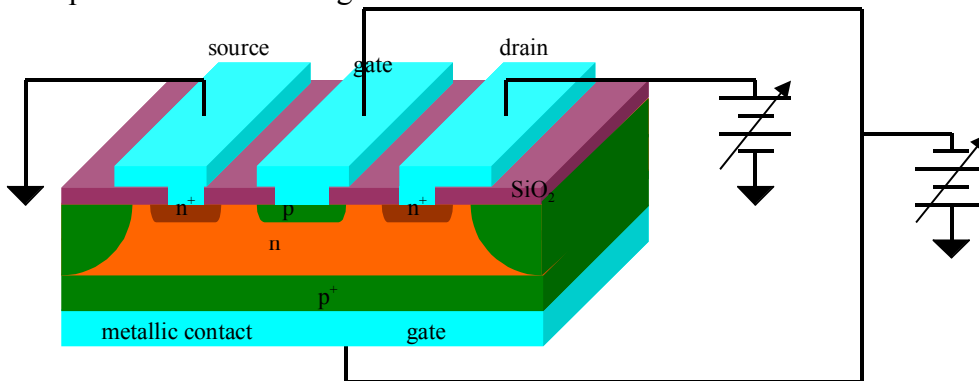
Gate – This is what controls the flow of charge carriers through the device. (If the gate is closed, nothing can go though...)

Channel – This is the region in which most of the charge carriers flow. The gate is used to open and close the channel.

Junction – Field Effect Transistor, J-FET



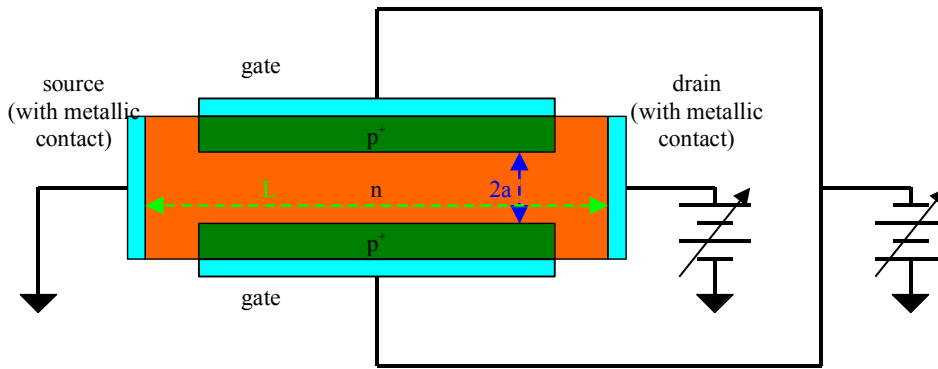
The original devices looked not unlike the picture above. It is with this picture that we will build an understanding of how they operate. [Understand that modern J-FETs can look very different! An example is shown in the figure below.]



Except for some geometric affects that we will not worry about in this class, this operates in much the same way that the original version worked.

Qualitative analysis of J-FETs

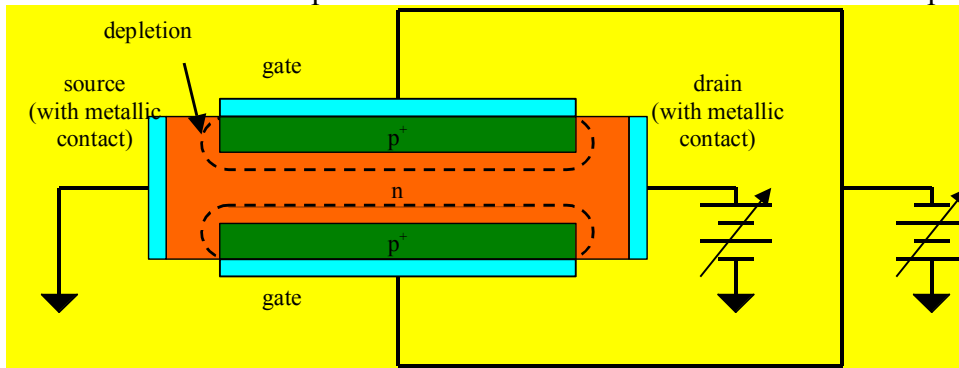
To establish the basic principles behind the J-FET we will first look qualitatively at what happens to the device.



What do we know.

- 1) If there is no bias applied to the gate, the depletion regions (or junctions in p^+-n devices) is fairly narrow.
- 2) Most of the depletion region around the gate is in the n-side of the junction. This is because the density of the n-side is much lower than the p^+ -side. Because the total charge inside the junction must be zero, more of the junction must be on the n-side.
- 3) If there is a small bias between the source and the drain, electrons will flow from the source to the drain.
- 4) The source side of the device is assumed to have a zero bias. (This is of course a relative value!)
- 5) The gate bias will be assumed to be equal to or negative relative to the source bias.
- 6) The drain bias will be assumed to be equal to or positive relative to the source bias.
 - a. A n^+-p-n^+ version of this will have the opposite polarity with holes being the dominate charge carrier.
- 7) The length of the device, L , is much larger than the width, $2a$.

Let us now draw a new picture of what the device looks like with the depletion region shown.

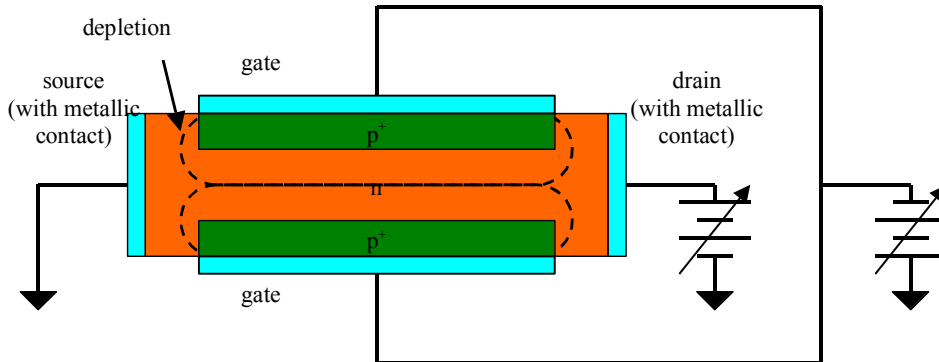


Because the depletion region lacks charge carriers, all of the current flow must go between the top and bottom depletion regions. (This is quite true. Remember that we can still have current flow through a p-n junction, which is simply a depletion region – this however is via the minority carriers! Here however, the area between the regions will have many more charge carriers as the majority carriers can and do play a major role.) This region is known as the **channel**.

Now we can do one of two things. We can turn up the bias on the drain or we can make the gate bias more negative. Let us change the gate first. When we do this, the depletion region grows, making our

Adjusting the gate

Making the gate bias more negative with respect to the source will cause the depletion region to grow in size. Once we have made the gate negative enough, the width of the channel will go to zero, as pictured below.



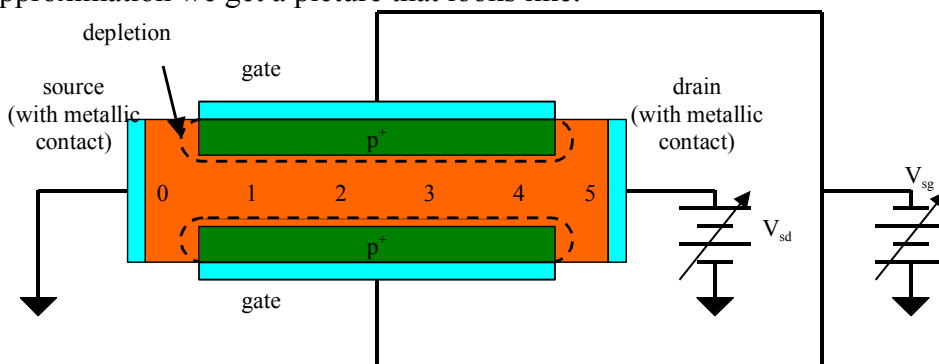
When this happens the majority carriers will no longer be present, and the current will be ‘shut-off’. (It is not a complete shut-off as we will see later.) **The voltage at which pinch-off occurs for no source-drain bias, $V_{sd} = 0$, is known as the pinch bias, V_p .** For some reason, Streetman has labeled this as a positive number. However, it must be a negative voltage for a J-FET like shown above. (If we had the opposite materials, it could and would be positive.)

Now let us go back to a case in which the gate bias is zero relative to the source bias. Before, we had a very small bias between the source and the drain, say less than 0.1 V. Now, let us raise the drain bias to +5 V relative to the source. This will cause significantly more current to flow. (We can understand this by realizing that the mobility of the electrons will remain about the same, and thus the velocity of the electrons must increase significantly.) Under these conditions, we have an electric field in the n-material that is approximately

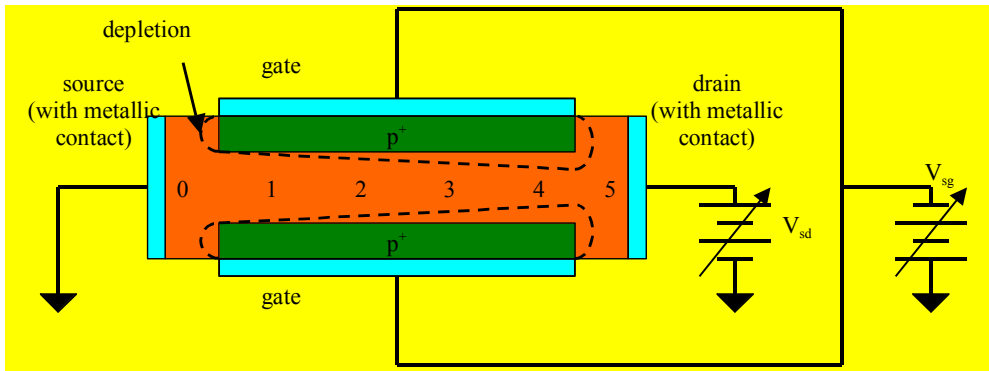
$$\mathbf{E} \approx \frac{V_{sd}}{L} \hat{\mathbf{y}}$$

(Some geometric affects will cause the electric field to be slightly different than this, but if $L \gg 2a$, then the difference is not large.)

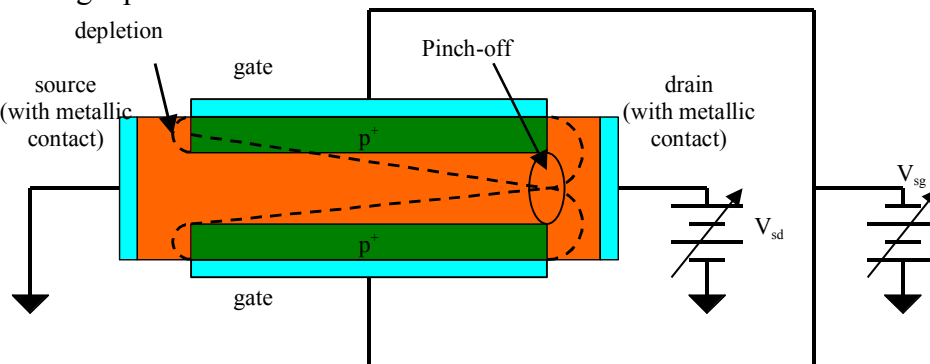
This means that the n-material has a gradient of the bias across the system. Assuming our linear approximation we get a picture that looks like:



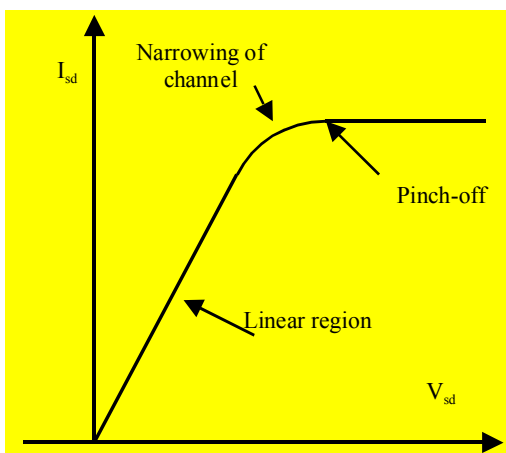
Now however we have a reverse bias between the channel and the gate. As this bias gets larger, the larger the depletion region becomes. This implies that we have growing depletion region width – and hence narrowing channel width – as we move toward the drain. This will look not unlike:



Let us continue to increase the source-drain voltage... At some point, the depletion regions will meet and we get pinch off.



What will happen now? Before, we expected the current to continue to grow as we increased the bias. In fact, we expected the current to grow in an approximately linear fashion with respect to the voltage. When pinch-off occurs, we expect the current to reach a constant value. The pinch-off can not stop the current flow as it requires that we have a bias inside the material – which in turn requires that we have a current flow. This means that we should have a current-voltage trace that looks like:



Side note on device resistance:

We know that the conductance is given by:

$$\sigma = qn\mu_n + qp\mu_p$$

$$\approx qn\mu_n$$

⇓

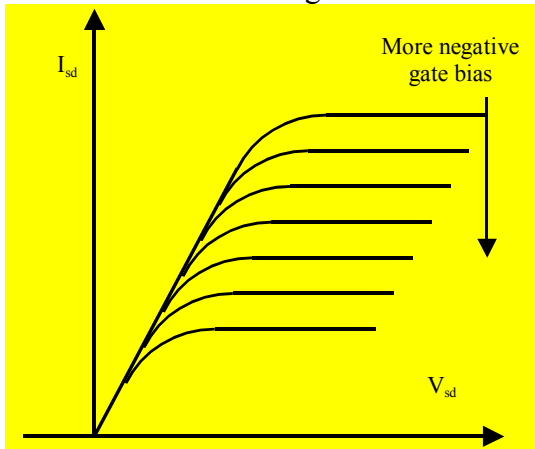
$$\rho = \frac{1}{\sigma} \approx \frac{1}{qn\mu_n}$$

⇓

$$R \approx \frac{\rho L}{A} = \frac{L}{Aqn\mu_n}$$

Since the area that the current flows through shrinks as we increase the gate bias, or drain bias, the resistance will increase. Thus, our device has a resistance that varies with our applied biases.

If we now use both the gate and the drain voltage, we find that we should have curves that look like:

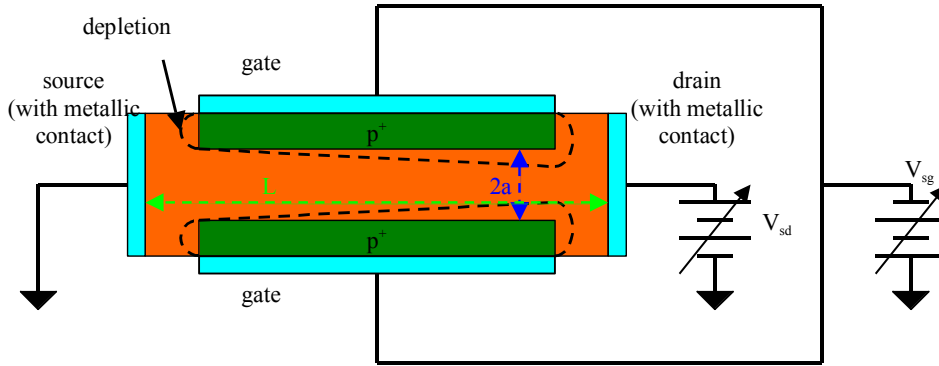


(This is because as we make the gate bias more negative, we will close (pinch) the channel at lower drain biases.)

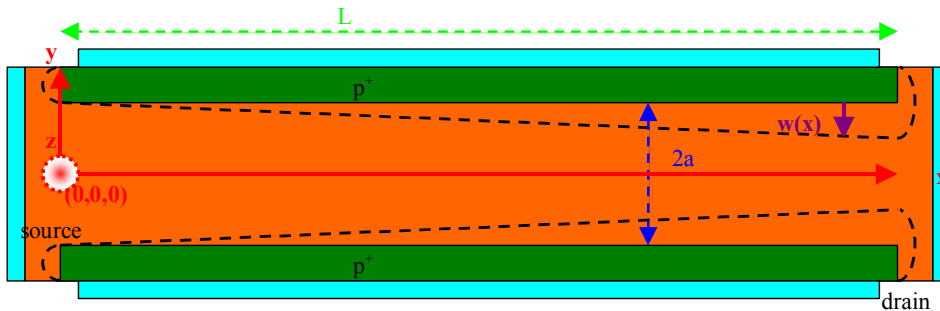
Now let us derive the current in a semi-analytical manner

Quantitative Analysis of the J-FET

Let us start with our picture of the system.



Now we need to blow this picture up and look at things in detail



Here we have added a coordinate system – with x along the channel and y in the direction between the two parts of the gate. (Thus z is out of the page via right-hand rule.) To get at an analytic solution we need to make a series of basic assumptions. (These are really just simplifying assumptions.) They are:

- 1) The junctions are p^+ - n step junctions. (Keep it simple!)
- 2) The materials are uniformly doped. (Keep it simple!)
- 3) The device is symmetric about the x -axis, with the p^+ material a constant distance ‘ a ’ from the axis. (Keep it simple!)
- 4) The top and bottom gate voltages are the same. (One does not need to have this... but it makes analysis tough – exactly what we do not want right now.)
- 5) Current flow is via majority carriers and it is confined to the non-depleted regions of the device. (Except when pinch occurs!)
- 6) Current flow is from the drain to source only. (- or the $-x$ direction but not in the y or z directions!)
- 7) The voltage drop from $x = 0$ to the source and from $x = L$ to the drain is negligible. (This means that the ends do not play a large role in the operation of the device.)
- 8) $L \gg a$
- 9) The voltage along the device is simply a function of the position x . Thus $V = V(x)$.
- 10) The depleted width, w , is set by the local voltage and thus is only a function of the position x .
 $\Rightarrow w = w(x)$.
- 11) The depleted width can grow to a maximum of a . (This means that the depleted regions cannot overlap.)
- 12) Breakdown does not occur – hence we are ignoring the fact that our device can become massively non-linear.

For conditions in which we do not have pinch-off, the current density is given by:

$$\begin{aligned} \mathbf{J} &= \mathbf{J}_n \\ &= J_{nx} \hat{\mathbf{x}} \\ &= q\mu_n n \mathbf{E} + qD_n \nabla n \end{aligned}$$

Within the channel, the electron density is approximately uniform and thus the diffusion term of the current density is approximately zero. (This is not unlike flowing current through a small piece of material with no injected charges.) Thus

$$\begin{aligned} J_{nx} &\approx q\mu_n n E_x \\ &\approx q\mu_n N_D E_x \\ &= q\mu_n N_D (-\partial_x V(x)) \end{aligned}$$

Now the total current is just the current density integrated over the channel area:

$$\begin{aligned} I_x &= \int_{-(a-w(x))}^{(a-w(x))} dy \int_0^{Z_0} dz \left[q\mu_n N_D (-\partial_x V(x)) \right] \\ &= \int_{-(a-w(x))}^{(a-w(x))} dy \left[z q\mu_n N_D (-\partial_x V(x)) \right]_0^{Z_0} \\ &= \int_{-(a-w(x))}^{(a-w(x))} dy \left[Z_0 q\mu_n N_D (-\partial_x V(x)) \right] \\ &= 2 \int_0^{(a-w(x))} dy \left[Z_0 q\mu_n N_D (-\partial_x V(x)) \right] \quad \text{-- by symmetry!} \\ &= -2 \left[(a-w(x)) Z_0 q\mu_n N_D \partial_x V(x) \right] \\ &\equiv I_D \end{aligned}$$

where I_D is the total current through the device. (Note that the current is in the negative x direction – as make sense, we have electrons flowing in the positive x direction!) Now we need to note that the current through the device must be a constant – we are not building up charge in the device. We can set $I_D = a$ constant and try to solve our differential equation – or we can note that as I_D is a constant, integrating along the length of the channel is equivalent to multiplying it by the channel length. Thus

$$\begin{aligned} I_D L &= \int_0^L I_D dx \\ &= \int_0^L -2 \left[(a-w(x)) Z_0 q\mu_n N_D \partial_x V(x) \right] dx \\ &= \int_0^{V(L)=V_D} -2 \left[(a-w(V)) Z_0 q\mu_n N_D \right] dV \\ &= -2 \left[Z_0 q\mu_n a N_D \right] \int_0^{V_D} \left(1 - \frac{w(V)}{a} \right) dV \\ &\Downarrow \\ I_D &= -\frac{2 \left[Z_0 q\mu_n a N_D \right]}{L} \int_0^{V_D} \left(1 - \frac{w(V)}{a} \right) dV \end{aligned}$$

Now we need to find the depletion width as a function of the voltage – but we know this from our study of p-n junctions

$$\begin{aligned}
w(V) &= x_{n0} \approx w_{\text{junction}} = x_{n0} + x_{p0} \\
&= \left(\frac{2\epsilon(V_{bi} - v_A)(N_A + N_D)}{qN_A N_D} \right)^{1/2} \\
&\approx \left(\frac{2\epsilon(V_{bi} - v_A)}{qN_D} \right)^{1/2} \\
&= \left(\frac{2\epsilon(V_{bi} - (V_G - V(x)))}{qN_D} \right)^{1/2}
\end{aligned}$$

Further the maximum width will occur when we reach pinch-off voltage.

$$a = \left(\frac{2\epsilon(V_{bi} - V_p)}{qN_D} \right)^{1/2}$$

⇓

$$\frac{w(V)}{a} = \left(\frac{(V_{bi} - V_G + V(x))}{(V_{bi} - V_p)} \right)^{1/2}$$

Here, V_p is the local voltage difference (i.e. the voltage at some x) that is required to cause a pinch at that location. It can also be described as the voltage at which pinch occurs for $V_G = 0$ – or the gate voltage at which pinch occurs for $V_d = 0$. These are all equivalent definitions. Plugging the above equation into the integral we find that

$$I_D = -\frac{\overbrace{2[Z_0 q \mu_n a N_D]}^{G_0}}{L} \left\{ V_D - \frac{2}{3} (V_{bi} - V_p) \left[\left(\frac{V_D + V_{bi} - V_G}{V_{bi} - V_p} \right)^{3/2} - \left(\frac{V_{bi} - V_G}{V_{bi} - V_p} \right)^{3/2} \right] \right\}$$

for $V_{Dsat} > V_D > 0$ and $V_p < V_G < 0$

Above saturation – e.g. pinch-off – we need to replace all of the V_D 's with V_{Dsat} 's.

At this point, we need to figure out what the saturation voltage is. Saturation is the point at which pinch-off first occurs. We know that this will happen at the end nearest the drain. Further we know that this will happen when the depletion width is equal to a . Thus will need to set the depletion width to 'a' and determine the voltage from that.

$$\begin{aligned}
w(V) &= \left(\frac{2\varepsilon \left(V_{bi} - V_G + \sqrt{V(x=L)} \right)}{qN_D} \right)^{1/2} \\
&= a \\
&= \left(\frac{2\varepsilon (V_{bi} - V_p)}{qN_D} \right)^{1/2} \\
&\Downarrow \\
(V_{bi} - V_p) &= \left(V_{bi} - V_G + \sqrt{V(L)} \right) \\
&\Downarrow \\
V_{Dsat} &= V_G - V_p
\end{aligned}$$

Now we can plug this in for the drain voltage, getting the saturation current,

$$\begin{aligned}
I_{Dsat} &= -\frac{2[Z_0 q \mu_n a N_D]}{L} \left\{ V_G - V_p - \frac{2}{3} (V_{bi} - V_p) \left[\left(\frac{V_G - V_p + V_{bi} - V_G}{V_{bi} - V_p} \right)^{3/2} - \left(\frac{V_{bi} - V_G}{V_{bi} - V_p} \right)^{3/2} \right] \right\} \\
&= \frac{G_0}{L} \left\{ -2[Z_0 q \mu_n a N_D] \left[V_G - V_p - \frac{2}{3} (V_{bi} - V_p) \left[1 - \left(\frac{V_{bi} - V_G}{V_{bi} - V_p} \right)^{3/2} \right] \right] \right\}
\end{aligned}$$

for $V_D > V_{Dsat} > 0$ and $V_p < V_G < 0$

Comparison between this and experimental data shows that our analytical model does a reasonable job of predicting the experimental data. (Improvements can be made to the model by improving some of our assumptions. For example assumption 7 is clearly not accurate. Removing the assumption improves the agreement between the model results and the experimental data.)

One problem with the above equation is that it is tough to remember. Further, we generally do not have the design characteristics of the device when we 'pull it out of the box'. However, there is a useful approximation that is relatively easy to remember that gives a reasonable value for I_{Dsat} . That equation is:

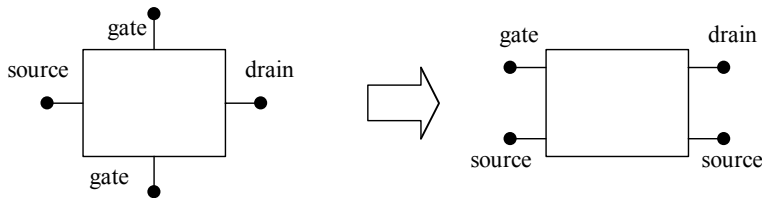
$$I_{Dsat} = I_{D0} \left(1 - \frac{V_G}{V_p} \right)^2$$

$$I_{D0} = I_{Dsat} \Big|_{V_G=0} = -\frac{2[Z_0 q \mu_n a N_D]}{L} \left\{ -V_p - \frac{2}{3} (V_{bi} - V_p) \left[1 - \left(\frac{V_{bi}}{V_{bi} - V_p} \right)^{3/2} \right] \right\}$$

Yes, mathematically it does not look the same, but it does give a reasonable approximation and it is easier to remember.

Small ac signals on J-FETs

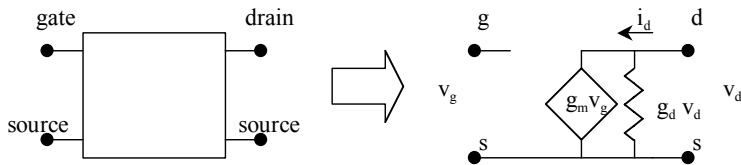
At this point it is time to look at small ac signals on our J-FET. (This is following our typical pattern – qualitative assessment of dc I-V traces, a quantitative assessment of dc I-V traces, and then a quantitative assessment of small ac I-V traces.) To do this, we need to look at how we use the device.



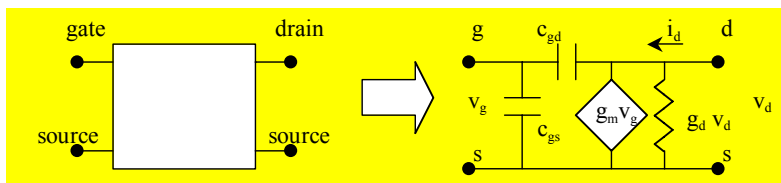
In general, we put our signal in on the gate line and pull our signal out on the drain line – leaving the source as our effective ground (not unlike we have done in our analysis above).

For low frequencies, the current that flows between the gate and the source is very low. This is because we have a reversed biased diode. Thus the gate to source acts like an open circuit.

The current between the drain and the source is set by the conductance, g , between the source and the drain. Additionally the gate voltage will control change the conductance. Let g_d be the conductance at zero gate voltage. Then we get an additional conductance, g_m , that is set by the gate voltage.



Now let us go high frequency. Here, the gate connects to both the drain and the source as if there were capacitors between the terminals. Thus we get



We can now use standard circuit analysis to arrive at our current and voltage relationships. The total drain current is given by the sum of the dc and ac currents. Thus,

$$I_D \left(\begin{matrix} \text{dc drain} & \text{ac drain} & \text{dc gate} & \text{ac gate} \\ \text{bias} & \text{bias} & \text{bias} & \text{bias} \\ \widehat{V}_D & + \widehat{v}_D & , \widehat{V}_G & + \widehat{v}_G \end{matrix} \right) = I_D(V_D, V_G) + \overset{\text{ac current}}{\widehat{i}_D}$$

⇓

$$i_D = I_D(V_D + v_D, V_G + v_G) - I_D(V_D, V_G)$$

At this point, we need to understand that the total current is simply a small signal variation from the dc current. Thus, we can use a Taylor expansion of the total current to arrive at the approximate total current.

$$I_D(V_D + v_D, V_G + v_G) \approx I_D(V_D, V_G) + v_D \left. \frac{\partial I_D}{\partial V_D} \right|_{\substack{v_D=0 \\ v_G=\text{const}}} + v_G \left. \frac{\partial I_D}{\partial V_G} \right|_{\substack{v_D=\text{const} \\ v_G=0}} + \dots$$

⇓

$$i_D \approx v_D \left. \frac{\partial I_D}{\partial V_D} \right|_{\substack{v_D=0 \\ v_G=\text{const}}} + v_G \left. \frac{\partial I_D}{\partial V_G} \right|_{\substack{v_D=\text{const} \\ v_G=0}}$$

$$= v_D g_d + v_G g_m$$

⇓

$$g_d = \left. \frac{\partial I_D}{\partial V_D} \right|_{\substack{v_D=0 \\ v_G=\text{const}}}$$

$$g_m = \left. \frac{\partial I_D}{\partial V_G} \right|_{\substack{v_D=\text{const} \\ v_G=0}}$$

We can now differentiate our drain current to get

Below pinch-off	Above pinch-off
$g_d = G_0 \left\{ 1 - \left(\frac{V_D + V_{bi} - V_G}{V_{bi} - V_p} \right)^{1/2} \right\}$	$g_d = 0$
$g_m = G_0 \left[\left(\frac{V_D + V_{bi} - V_G}{V_{bi} - V_p} \right)^{1/2} - \left(\frac{V_{bi} - V_G}{V_{bi} - V_p} \right)^{1/2} \right]$	$g_m = G_0 \left[1 - \left(\frac{V_{bi} - V_G}{V_{bi} - V_p} \right)^{1/2} \right]$

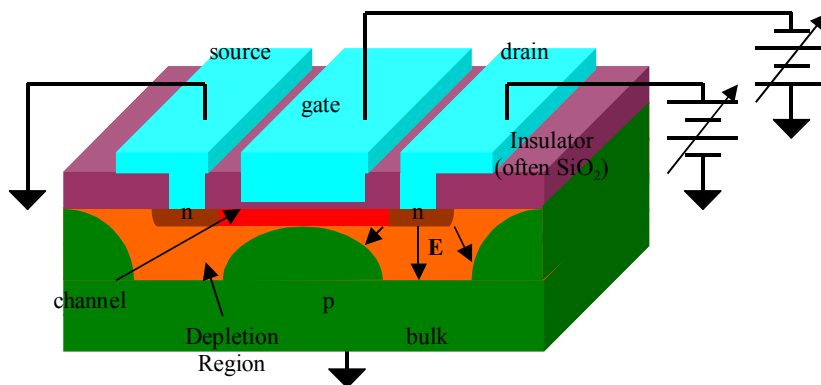
MESFETs

MESFETs are METal Semiconductor FETs. These are extremely fast J-FETs. The Major changes are:

- 1) The gate is a Schottky diode. (Usually a metal contact directly on the channel semiconductor.)
 - 2) Drain and Sources are Ohmic contacts
 - 3) No diffusion required => can be very small and very fast
- (In commercial applications, most MESFETs are made from GaAs.)

Insulated Gate Field Effect Transistors IGFETs.

Insulated gate FETs look somewhat similar to J-FETs – except there is a replacement of the heavily doped gate semiconductor with a thin layer of dielectric material between the metal pad and the channel.



We will find that the width of the channel is set by the gate voltage – the larger the voltage, the wider the channel. This is exactly the opposite to what we saw with J-FETs.

IGFETs come in a few different flavors:

MOSFET – Metal Oxide Semiconductor FET

The gate is metal (or heavily doped poly-Si) the dielectric is SiO_2 , and the rest of the device is Si.

MISFET – Metal Insulator Semiconductor FET

The gate is metal (or heavily doped poly-Si) the dielectric is something besides SiO_2 , and/or the rest of the device is something besides Si.

NOTE THAT MOSFET IS A SPECIAL VERSION OF MISFET

CMOS – Complimentary MOS FETs

This device consists of two MOSFETs – one p-channel and one n-channel

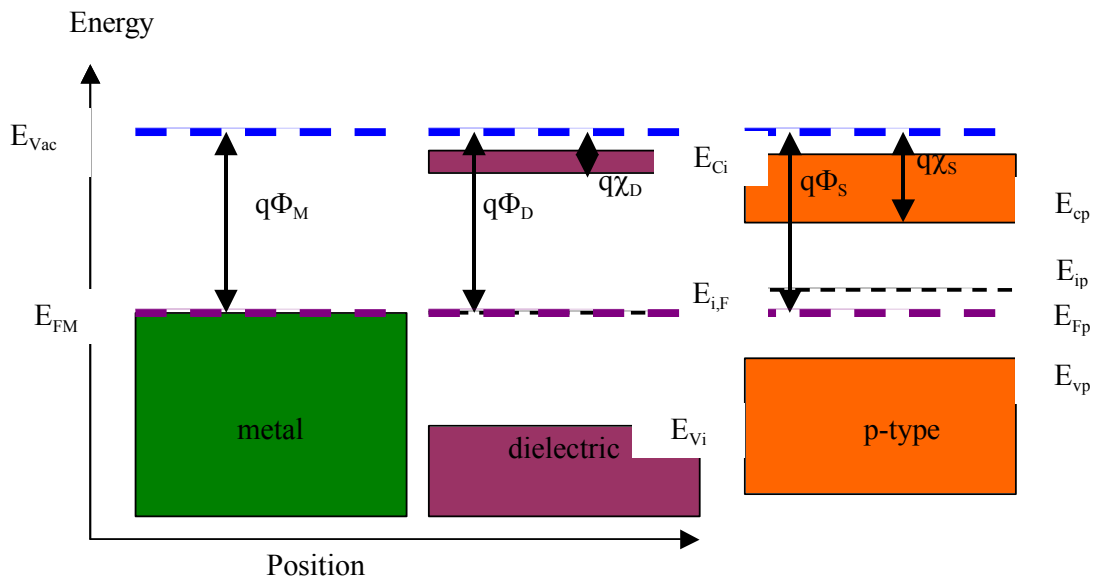
The operation of each of these revolves around what happens in the gate region. There are two things that we can quickly note:

- 1) There is no dc current flow from the gate. (Because of the insulator... In reality, there is always some ‘leakage’ current, just not much.)
- 2) The region between the gate and the substrate/bulk acts like a capacitor.

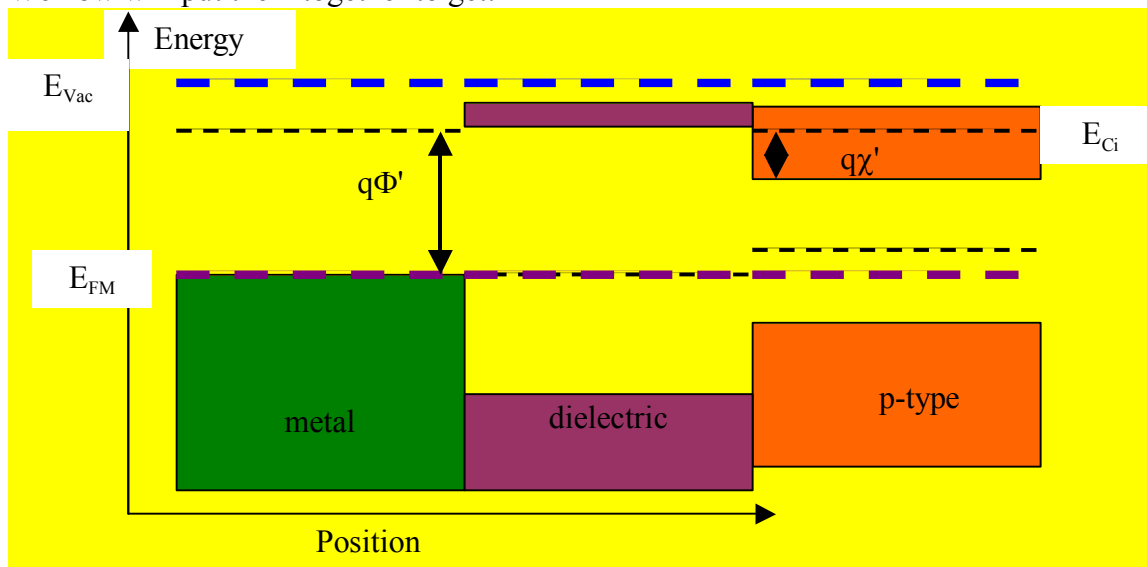
Given this simple description, we can begin to develop a picture of how these devices work. We have three energy diagrams, one each for the metal, dielectric and semiconductor. Before we draw them, we are going to make a few simplifying assumptions.

- 1) We will assume, for simplicity, that the work function for the metal, dielectric and the semiconductor are the same.
- 2) That there are no charges in the oxide or at the interfaces between the materials.

Thus, we find

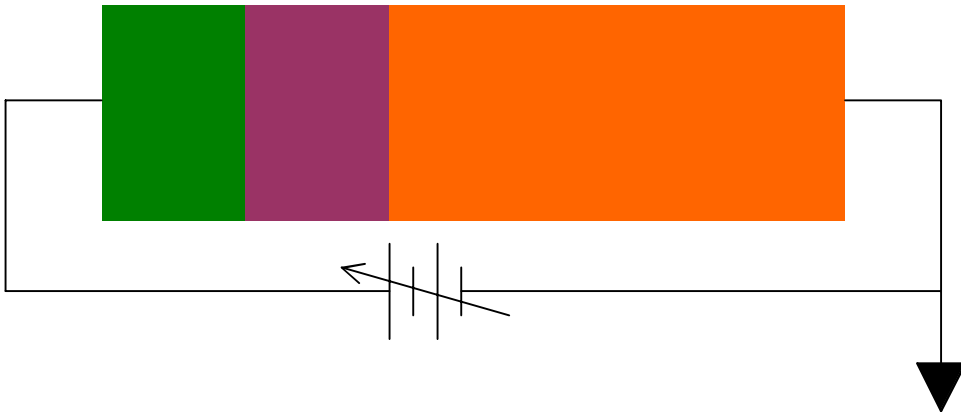


We now will put them together to get:



Note that we have now defined a new 'work function' and a new χ based on the level of the conduction band of the insulator.

Now what happens when the device is biased?



Can we decide what the energy levels will look like?

- 1) Let us assume that the bias applied to the metal is negative, thus $V_A < 0$. Thus we would assume that the Fermi energy of the metal will be above that for the semiconductor.
- 2) There will be no potential variation in the metal – e.g. no electric field inside the metal – and hence the Fermi level of the metal will be constant.
- 3) As there is no charge inside the dielectric, the electric field inside the dielectric will be a constant. This means that the potential inside the dielectric must change in a linear fashion.

Mathematically

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} = 0$$

↓

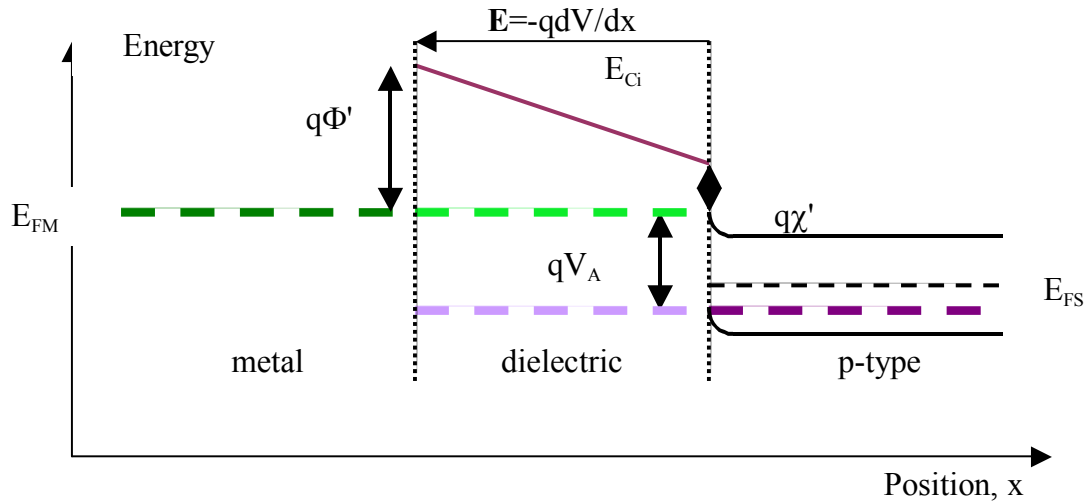
$$\mathbf{E} = \text{const} = -\nabla \phi$$

↓

$$\phi(x) = \phi_0 x$$

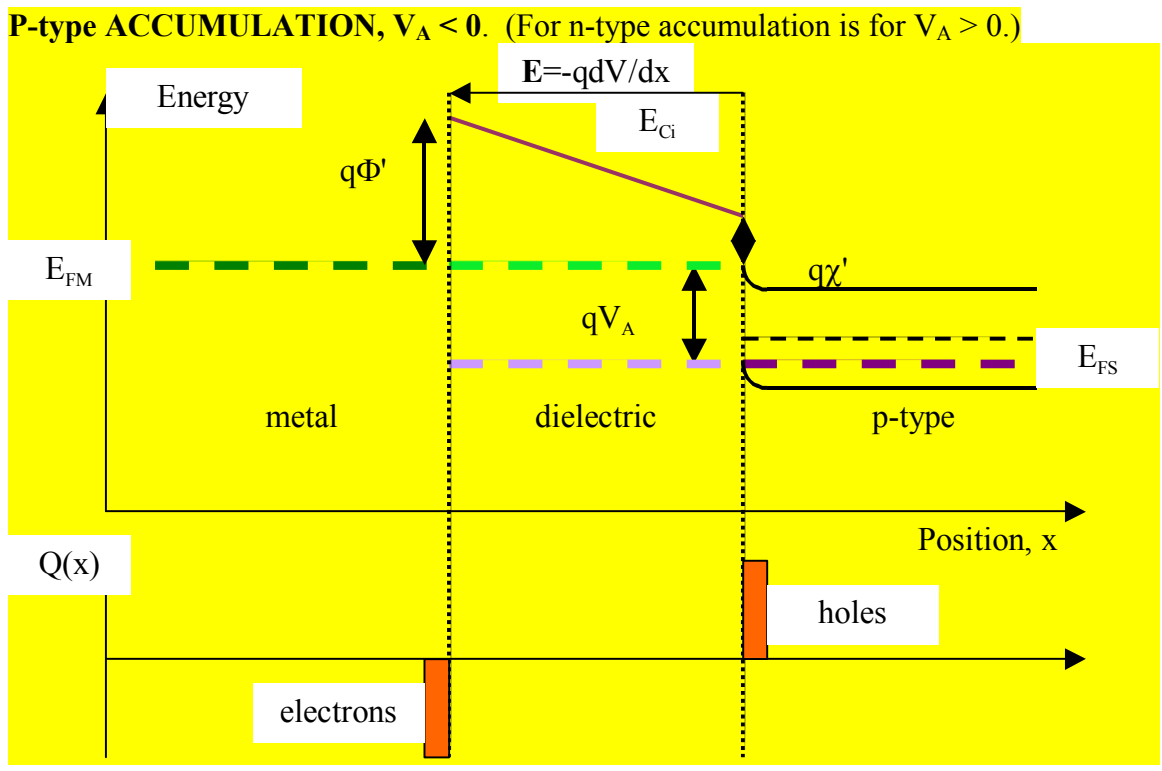
- 4) Finally, the gap between the conduction band of the insulator and the conduction band of the semiconductor must remain constant.

So we expect a picture that looks like

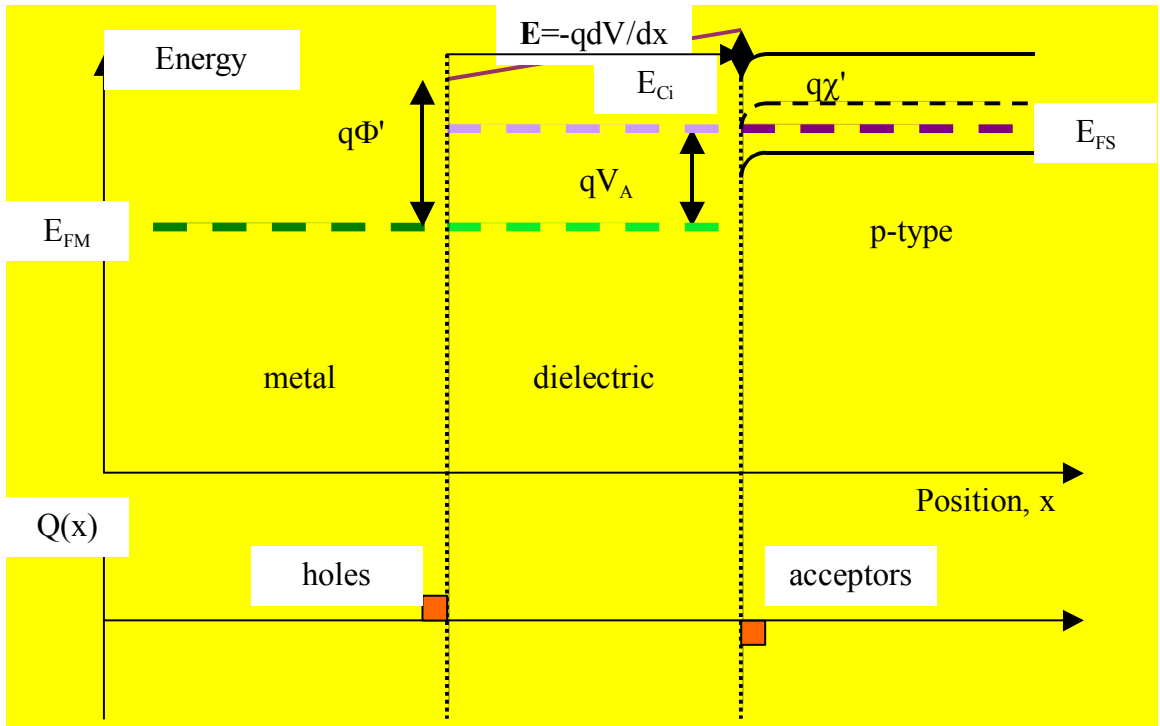


(We are now eliminating all non-useful energy levels) What does this look like in terms of charge carriers?

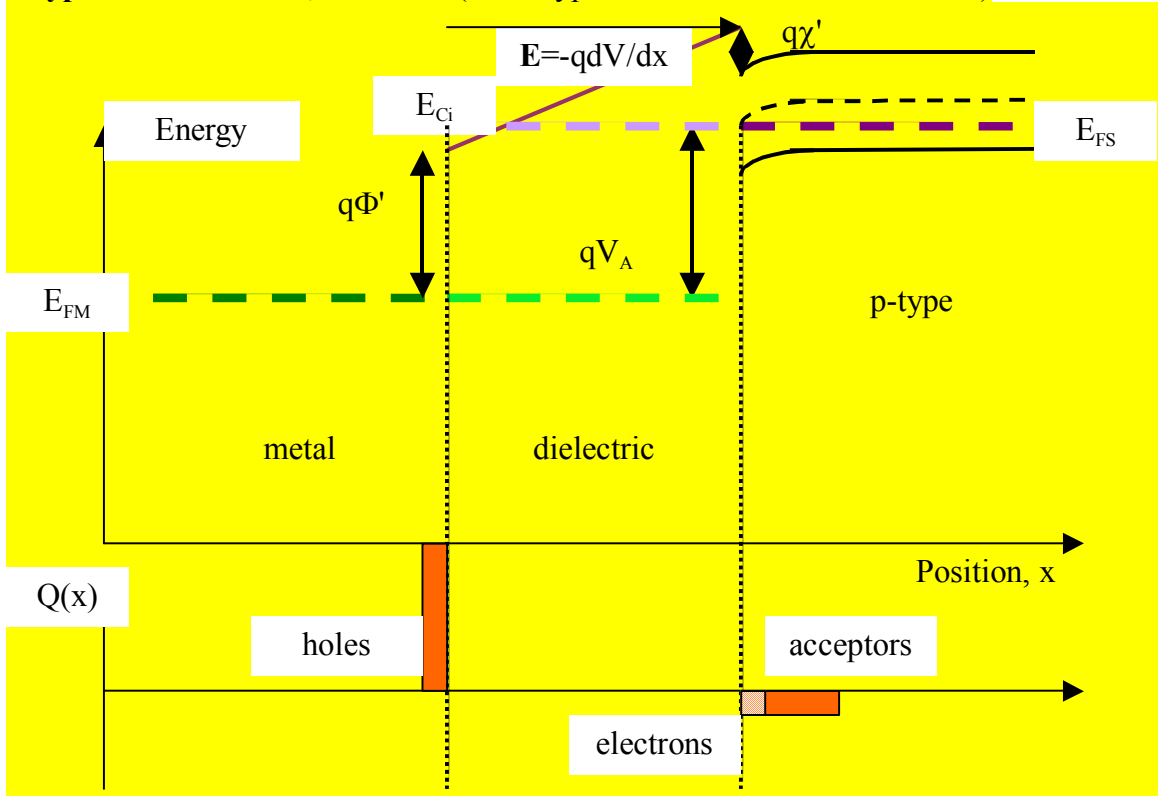
We know that when we have a structure like this on the semiconductor side, we are accumulating or removing (depleting) the majority charge carriers. In this particular case, we will have additional holes right near the semi-dielectric boundary. (This is because our majority carriers are holes and because the holes will be at a lower energy at the boundary.) Thus, we are ACCUMULATING holes at the edge. Because the electric field cannot penetrate into the metal, we must have an equal but opposite amount of charge build up at the metal- dielectric boundary. Thus, we get a charge distribution that looks like:



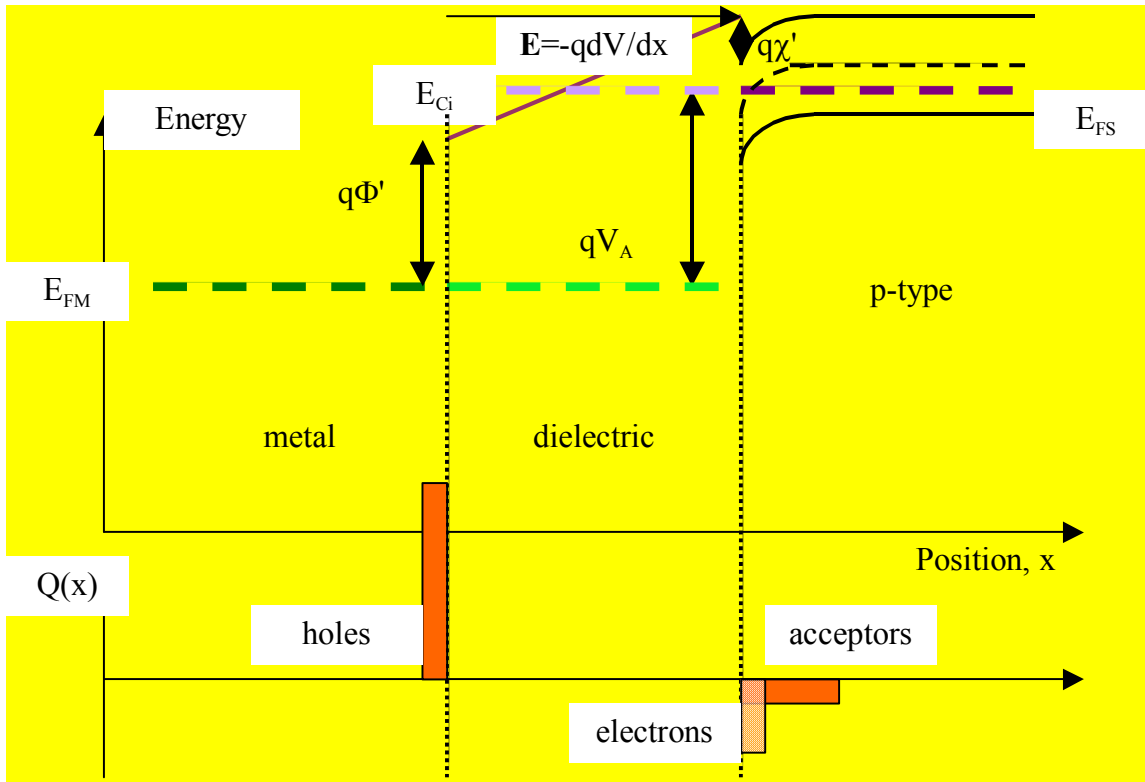
P-type small depletion, $V_A > 0$. (For n-type small depletion is for $V_A < 0$.)



P-type inversion onset, $V_A \sim V_T$. (For n-type accumulation is for $V_A \sim V_T$.)



P-type Inversion, $V_A > V_T$. (For n-type inversion is for $V_A < V_T$.)



We have drawn picture of what is going on but can we show the same thing using mathematics? The difference in the pictures has to do with the location of the intrinsic energy relative to the location of the Fermi energy. As we push the bias higher, the intrinsic energy sinks lower, until it drops below the Fermi energy at the oxide-Si interface (or dielectric-semiconductor interface). When this happens, the semiconductor begins to act like it is n-type. Hence we have converted a p-type material into an n-type – at least right next to the insulator.

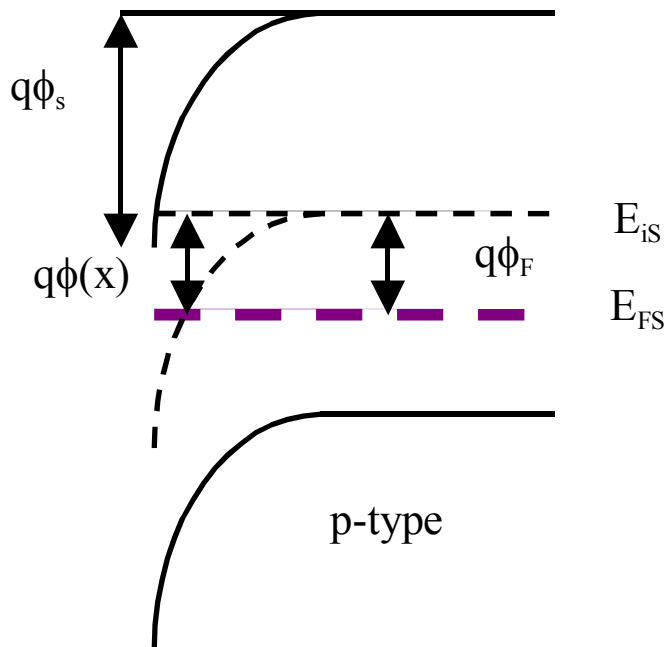
Let us define three more energy levels. The first will be the energy between the intrinsic energy and the Fermi energy, $q\phi_F$. The second will be the amount of band bending that occurs in the semiconductor as a function of position, $q\phi(x)$. The final energy is the total amount of bending at the interface or surface, $q\phi_s$. Mathematically this is:

$$q\phi_F = E_i(\text{bulk}) - E_F$$

$$q\phi_s = E_i(\text{bulk}) - E_i(\text{surface})$$

$$q\phi(x) = E_i(\text{bulk}) - E_i(x)$$

A picture looks like:



Note that ϕ_s is different than the semiconductor work function Φ_s .

We know that inversion will start when the intrinsic energy drops below the Fermi energy. Thus we reach the threshold of inversion when

$$\phi_s \geq \phi_F$$

Before ϕ_s is greater than ϕ_F , we are in depletion mode. Once ϕ_s is at least equal to ϕ_F , we are beginning to build up a supply of free electrons at the oxide-Si interface. If we continue to increase our external bias, eventually we will reach a point at which our **interface material (or channel)** is as 'n-type' as we had originally doped it to be p-type. **This is known as Strong inversion.** Mathematically, this is when the shift in the intrinsic energy is twice the difference between the Fermi energy and the intrinsic energy in the Bulk – or:

$$\phi_s \geq 2\phi_F$$

Putting this together with our charge carrier densities, we find:

$$\begin{aligned} n(x) &= n_i e^{(E_F - E_i)/kT} \\ &= n_i e^{(E_F - E_i(x))/kT} \\ &= n_i e^{(-q\phi_F + q\phi(x))/kT} \quad \text{but } n_0 = n_i e^{(-q\phi_F)/kT} = n_i e^{(E_F - E_i)/kT} \\ &= n_0 e^{(q\phi(x))/kT} \end{aligned}$$

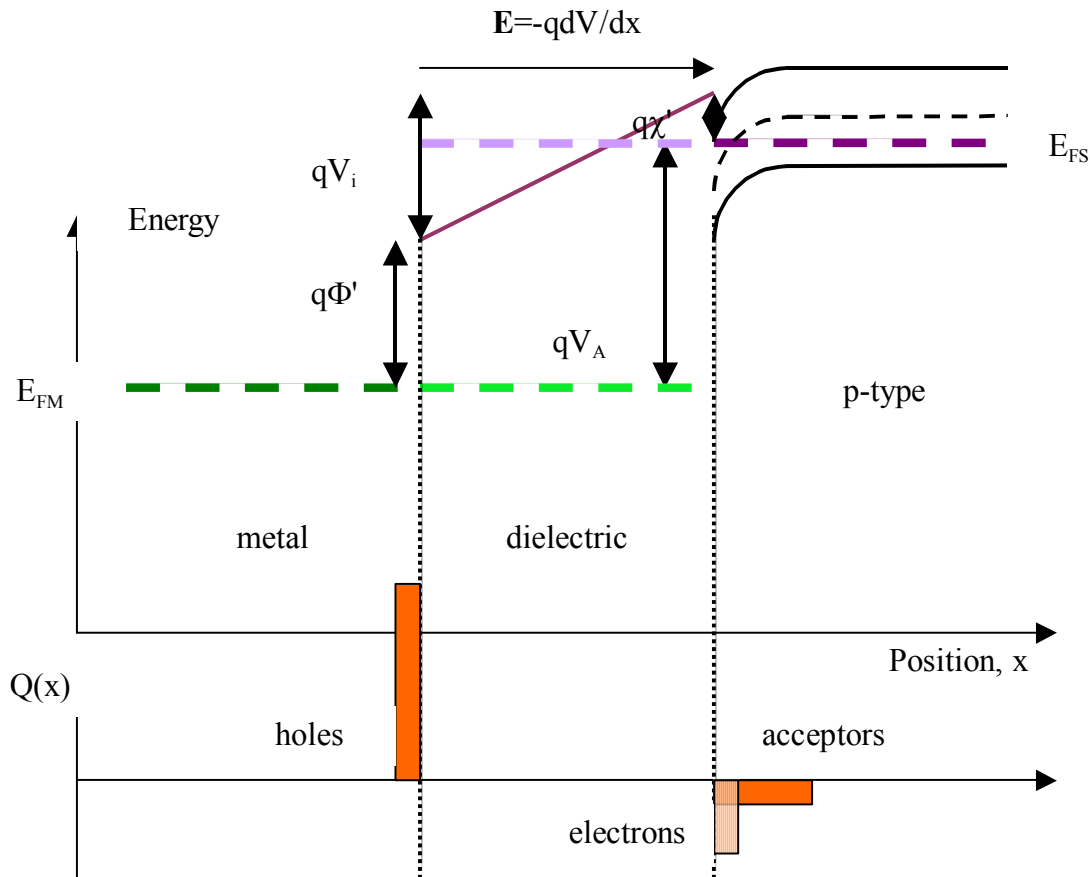
$$\begin{aligned} p(x) &= n_i e^{(q\phi_F - q\phi(x))/kT} \\ &= p_0 e^{(-q\phi(x))/kT} \end{aligned}$$

Note that

$$n(x)p(x) = n_i^2$$

Now we have n and p versus $\phi(x)$. What is $\phi(x)$?

To understand this, let us go back to our picture.



First, the charge that is built up on the metal side is in a very thin region. It is so thin that we will assume that it is effectively a delta function distribution for its width. Likewise the free charge carriers (majority or minority) that we build up on the semiconductor side are also in a thin region, typically less than 100 Å wide. These we can also assume are a delta function distribution. Finally, any bound charges that are built up on the semiconductor side must be over a fairly wide region. The width of that region is given by an equation that is similar to our junction width from our study of p-n junctions.

However, once inversion begins to happen, the width of the depletion region does not grow significantly. This can be understood in the following manner. The electric field will extend only as far as required to balance the charge that builds up on the opposite side of the insulator. As we have free charge carriers now on the semiconductor side, they will move in the direction required by an electric field. Thus, they will serve to shorten the distance that the electric field would normally penetrate into the semiconductor material in a p-n junction.

The fundamental equation is:

Poisson's Equation:

$$\nabla \cdot \mathbf{E} = \frac{\rho(x)}{\epsilon_r \epsilon_0}$$

$$\nabla^2 \phi(x) = \partial_x^2 \phi(x) \leftarrow \text{we are assuming only a single dimension}$$

$$= -\frac{\rho(x)}{\epsilon_r \epsilon_0}$$

During accumulation or depletion, the charge build up is primarily on the surface of the metal and that of the semiconductor. Thus the electric field in the oxide is approximately a constant and is given by

$$\mathbf{E}(x) = \frac{\overbrace{\rho_m}^{\text{surface charge on metal}}}{\epsilon_r \epsilon_0}$$

⇓

$$V_{\text{ins}}(x) = -\frac{\rho_m}{\epsilon_r \epsilon_0} x + \text{const} \leftarrow \text{we will pick the constant later to be } \phi_s$$

$$V_{\text{ins}}(x) = -\frac{\rho_m x}{\epsilon_r \epsilon_0} + \phi_s$$

This can be obtained by integrating Poisson's equation from just inside to metal to just inside the dielectric. In the metal, the electric field is zero. In addition, the only charge in the region is the surface charge – a delta function. Thus, we get the above terms.

During depletion, the unknown electric field is that in the semiconductor. That can be found again from Poisson's equation

$$\partial_x \mathbf{E}(x) = \frac{\overbrace{\rho_s}^{\text{charge in semi}}}{\epsilon_r \epsilon_0} \approx \frac{qN_A}{\epsilon_r \epsilon_0} \leftarrow \text{assumes p-type material}$$

⇓

$$\mathbf{E}(x) \approx \frac{qN_A}{\epsilon_r \epsilon_0} (W - x)$$

⇓

$$\phi \approx \frac{qN_A}{2\epsilon_r \epsilon_0} (W - x)^2$$

where W is the width – and the point at which the electric field goes to zero in the semiconductor.

We can get the width by noting that the intrinsic energy has shifted ϕ_s at the interface, $x = 0$. (Note that this just assumed 'ground' to be the bulk semiconductor.) Thus,

$$\phi_s \approx \frac{qN_A}{2\epsilon_r \epsilon_0} W^2$$

⇓

$$W = \sqrt{\frac{2\epsilon_r \epsilon_0 \phi_s}{qN_A}}$$

The widest this will become is the width at which a significant number of free electrons begin to accumulate at the semi-insulator interface. This occurs approximately when we reach the inversion-depletion transition or threshold point, $\phi_s = 2\phi_F$. Replacing ϕ_s in the above equation, we get the approximate maximum width – known as the threshold width.

$$W_T = \sqrt{\frac{4\epsilon_r \epsilon_0 \phi_F}{qN_A}}$$

Gate Voltage relationship:

Now we know that we are applying a total voltage across the system that is simply the sum of the voltage shift across the oxide plus the voltage shift across the semiconductor, which is simply the gate voltage. Thus,

$$\begin{aligned} V_G &= \Delta V_{\text{ins}} + \Delta V_{\text{Semi}} \\ &= -\frac{\rho_m x_0}{\epsilon_r \epsilon_0} + \phi_s \end{aligned}$$

Now we do not know what the width of the oxide, x_0 , is so that we cannot determine which part belong to the semiconductor and which part belongs to the oxide. (However, once we have built a device, x_0 is a constant.) There is however, one more equation that we can use. That equation relates the electric field strength across a boundary to the charge density on the boundary

$$(\epsilon_{rs} \epsilon_0 \mathbf{E}_s|_{\text{normal}} - \epsilon_{ri} \epsilon_0 \mathbf{E}_i|_{\text{normal}}) = Q_{si}$$

where Q_{si} is the charge at the semiconductor – insulator interface. Assuming that this is zero, we get

$$\epsilon_{rs} \mathbf{E}_s|_{\text{normal}} = \epsilon_{ri} \mathbf{E}_i|_{\text{normal}}$$

From above, the electric field in the semiconductor at the interface is:

$$\mathbf{E}_s \approx \frac{qN_A}{\epsilon_{rs} \epsilon_0} (W)$$

so

$$\mathbf{E}_i = \frac{\epsilon_{rs}}{\epsilon_{ri}} \mathbf{E}_s \approx \frac{qN_A}{\epsilon_{ri} \epsilon_0} (W)$$

⇓

$$\Delta V_{\text{ins}} = \frac{\epsilon_{rs}}{\epsilon_{ri}} x_0 \mathbf{E}_s$$

⇓

$$\begin{aligned} V_G &= \Delta V_{\text{ins}} + \Delta V_{\text{Semi}} \\ &= \frac{\epsilon_{rs}}{\epsilon_{ri}} x_0 \mathbf{E}_s + \phi_s \end{aligned}$$

If we now assume that the free carriers on the semiconductor side are delta function, then

$$\partial_x \mathbf{E}(x) = \frac{\overbrace{\rho_s}^{\text{charge in semi}}}{\epsilon_r \epsilon_0} \approx \frac{qN_A}{\epsilon_r \epsilon_0} \leftarrow \text{assumes p-type material}$$

⇓

$$\mathbf{E}(x) \approx \frac{qN_A}{\epsilon_r \epsilon_0} (W - x)$$

⇓

$$\phi \approx \frac{qN_A}{2\epsilon_r \epsilon_0} (W - x)^2$$

$$\mathbf{E}(x) \approx \frac{qN_A}{\epsilon_{rs} \epsilon_0} (W - x); \quad W = \sqrt{\frac{2\epsilon_{rs} \epsilon_0 \phi_s}{qN_A}}$$

⇓

$$\mathbf{E}_s \approx \frac{qN_A}{\epsilon_{rs} \epsilon_0} \left(\sqrt{\frac{2\epsilon_{rs} \epsilon_0 \phi_s}{qN_A}} - 0 \right)$$

⇓

$$\mathbf{E}_s \approx \sqrt{\frac{2qN_A \phi_s}{\epsilon_{rs} \epsilon_0}}$$

This equation provides a reasonable approximation for ($0 \leq \phi_s \leq 2\phi_F$) and can be plugged into our equation above, to get,

$$V_G = \Delta V_{\text{ins}} + \Delta V_{\text{Semi}}$$

$$= \frac{\epsilon_{rs}}{\epsilon_{ri}} x_0 \mathbf{E}_s + \phi_s$$

$$\approx \frac{\epsilon_{rs}}{\epsilon_{ri}} x_0 \sqrt{\frac{2qN_A \phi_s}{\epsilon_{rs} \epsilon_0}} + \phi_s$$

If we remember that we reach threshold (effective device turn-on) at

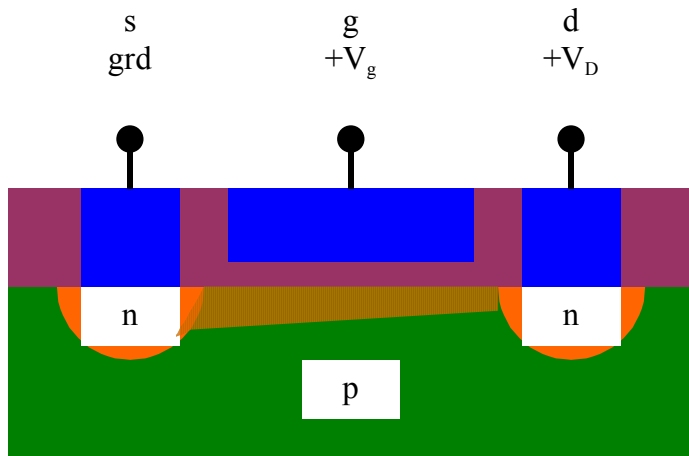
$$\phi_s = 2\phi_F$$

then the threshold voltage is simply the voltage at which turn-on occurs or

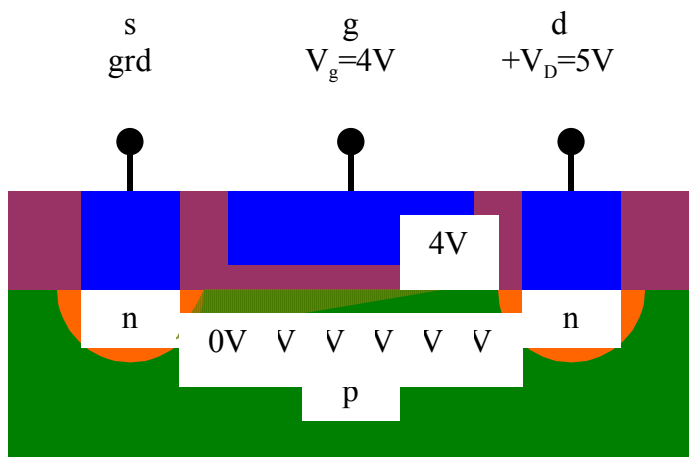
$$V_T \approx \frac{\epsilon_{rs}}{\epsilon_{ri}} x_0 \sqrt{\frac{4qN_A \phi_F}{\epsilon_{rs} \epsilon_0}} + 2\phi_F$$

Current in IGFETS

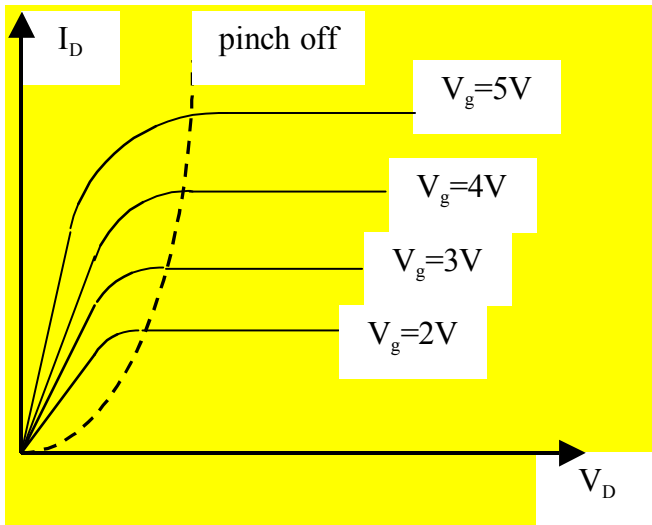
Now we really need to know what the current will be like in our device. Let us look at a picture of the System



When there is no gate bias, there is no current channel and hence no current. Therefore the gate controls the output current just like in a JFET – except the higher the voltage the WIDER the channel and hence the higher the current. Also, like a JFET, if V_D becomes too large, the current can be pinched off. This can be seen in the following figure.



We see in this case, the gate is biased at 4 V but the drain voltage has pulled the p-type material up to 5 V. Under such a situation, the voltage difference is not large enough and thus we will not have any free electrons – but rather we will have free holes near the drain (in the p-material). (This is because the drain pulls it positive.) Hence we will not reach inversion. From this, we would expect that the current voltage traces look like:



In general, the device must have some minimum gate voltage to conduct. Typically, it is a slightly negative gate voltage that will allow conduction but sometimes it takes a gate voltage above zero to get conduction. Also note that the pinch off voltage increases with increasing gate voltage – just the opposite of what happens in a JFET.

Capacitance in IGFETS

In addition to the current flow we will also briefly consider the capacitance of an ideal MOS device. (This is a simplified version of reality. We will return to the concept later, after we have developed a more realistic version of the MOS capacitor.)

In general, we can consider the MOS capacitor as the combination of two capacitors, the capacitor made up of the oxide layer and the capacitor made up of the depletion layer. We know that the voltage across the system is simply the voltage across the oxide plus the voltage across the semiconductor. Thus,

$$V_G = \Delta V_{\text{ins}} + \Delta V_{\text{Semi}}$$

Now if we were to just consider the section across the oxide, we know that

$$\Delta V_{\text{ins}} = -\frac{Q_{\text{si}}}{C_i}$$

where Q_{si} and the C_i are the free charge at the semiconductor-insulator interface and insulator capacitance per unit area. We know from a number of other subjects that

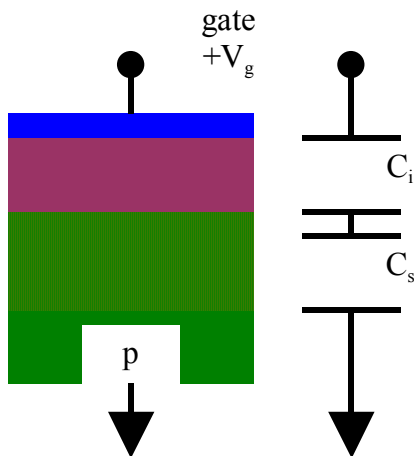
$$C_i \approx \frac{\epsilon_{\text{ri}} \epsilon_0}{x_0}$$

We know from our discussion of p-n junctions, that the depletion region can also have a capacitance associated with it. That is given by

$$\begin{aligned} C_s &= \frac{dQ_s}{dV_s} \\ &= \frac{dQ_s}{d\phi_s} \\ &\approx \frac{\epsilon_{\text{rs}} \epsilon_0}{W} \end{aligned}$$

Where again our values, C and Q, are ‘per unit area’.

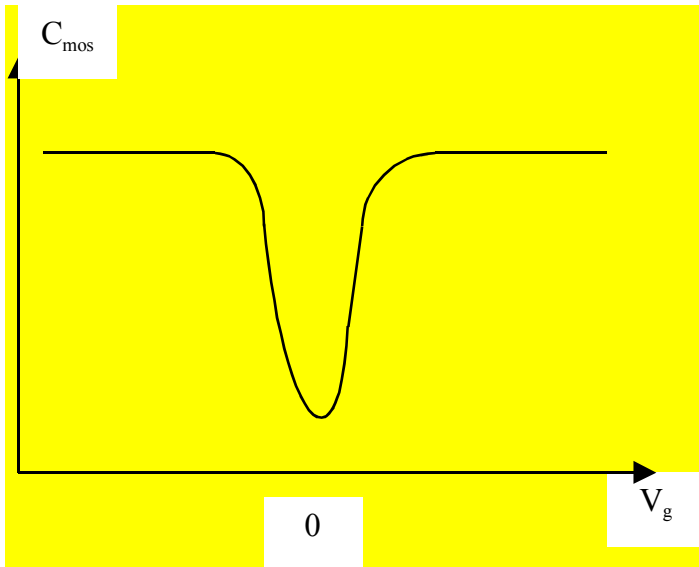
Graphically the system that we are describing looks like:



Thus we have a series of capacitors. Combining these together, we get the capacitance of the whole structure.

$$\begin{aligned}
 C_{\text{mos}} &= \frac{C_s C_i}{C_s + C_i} \\
 &\approx \frac{\frac{\epsilon_{rs} \epsilon_0}{W} \frac{\epsilon_{ri} \epsilon_0}{x_0}}{\frac{\epsilon_{rs} \epsilon_0}{W} + \frac{\epsilon_{ri} \epsilon_0}{x_0}} \\
 &= \frac{\frac{\epsilon_{rs}}{W} \frac{\epsilon_{ri} \epsilon_0}{x_0}}{\frac{\epsilon_{rs} x_0 + \epsilon_{ri} W}{W x_0}} \\
 &= \frac{\epsilon_{rs} \epsilon_{ri} \epsilon_0}{\epsilon_{rs} x_0 + \epsilon_{ri} W}
 \end{aligned}$$

When the depletion region is very narrow, i.e. very little applied voltage, then the total capacitance is primarily due to the capacitance in the depletion region. When the depletion region is wide, i.e. either strong accumulation or strong inversion, the total capacitance is primarily due to the capacitance on the insulator. Thus we would expect that the capacitance as a function of gate voltage looks like:



(Note that the C-V curve is not necessarily symmetric – and is typically not – around zero gate bias.)

If we now plug in our oxide capacitance into our total voltage term, we find

$$V_G = \Delta V_{\text{ins}} + \Delta V_{\text{Semi}}$$

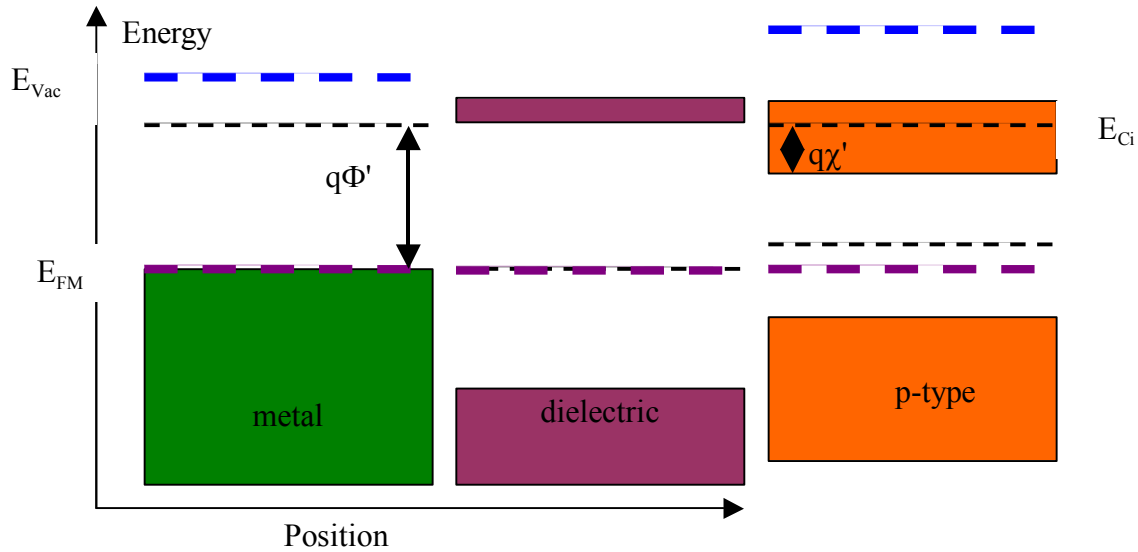
$$= -\frac{Q_{\text{si}}}{C_i} + \Delta V_{\text{Semi}}$$

Threshold – and hence large-scale conduction will occur when the voltage drop across the semiconductor is twice ϕ_F . So

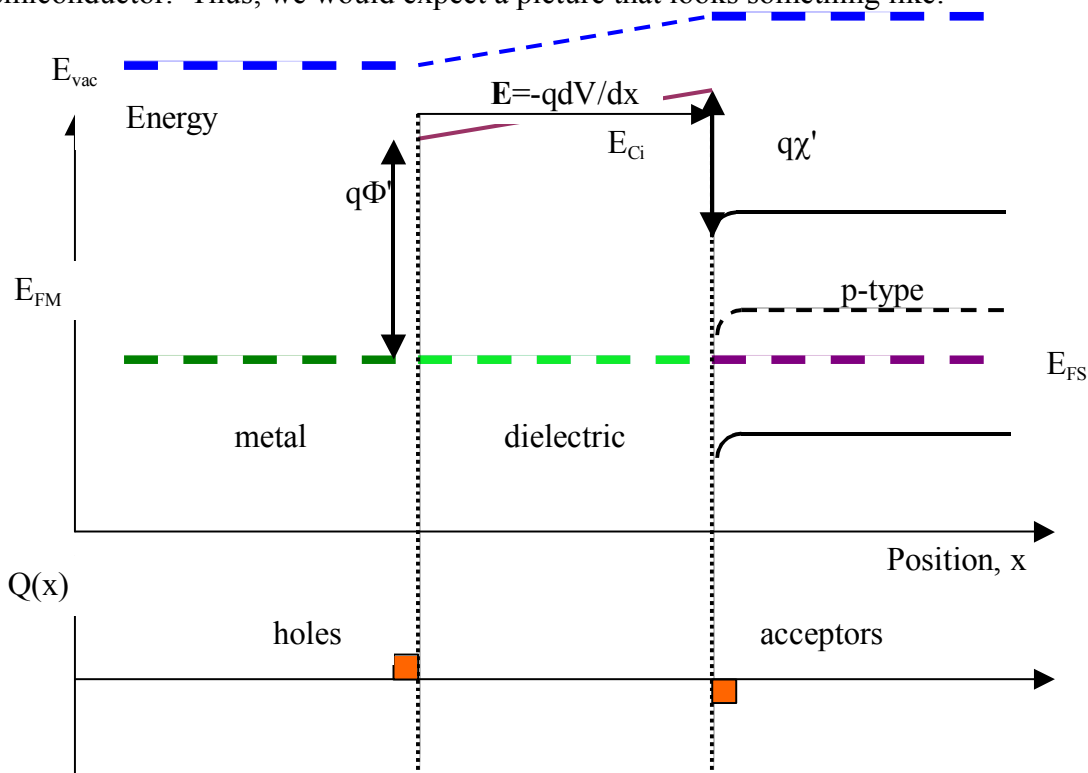
$$V_T = -\frac{Q_{\text{si}}}{C_i} + 2\phi_F$$

‘Real’ devices

Real devices vary in a number of ways from the assumptions we made when we first described an IGFET. Notably, we assumed that the work functions of the metal, semiconductor and insulators were identical. (When no external bias is applied, this situation is known as ‘FLAT BAND’.) The reality of the situation is that typically the work function of the semiconductor is often larger than that for the metal. (Si is an example of this.) Additionally, the work function for the semiconductor often is a function of the dopant concentration. (Actually, I do know of any case in which the dopant type and dopant concentration does not influence the work function of the semiconductor. Contaminate materials will also influence the work function.) Thus, we would expect that the energy diagrams for a true MOS capacitor should look like:



We know that the Fermi energy must align if we do not apply an external bias, further the energy difference between the metal Fermi and Oxide conduction band must be the same as before we put them together. This also applies to the difference between the conduction bands of the oxide and the semiconductor. Thus, we would expect a picture that looks something like:



We note that because the work function of both the metal and semiconductor are spatially constants, the dielectric must have a shift in the potential across it. This in turn gives rise to charge buildup on the metal surface (holes) and a corresponding charge buildup in the semiconductor - as a small depletion region - these will be either acceptors for p-type or excess electrons for n-type materials. (If the difference between the metal work function and the semiconductor work function is large enough, we can even reach inversion - but this is not that common.)

The important distinction between what we were doing before and now, is the shift in the work functions. We give this shift a special symbol:

$$\Phi_{ms} = \Phi_m - \Phi_s$$

Note that it is typically negative for many systems of interest – but not all systems.

Interface traps and oxide defects

In addition to the ‘minor’ fact that the work functions of various materials are not equal, insulators also can have charges trapped inside them (defects). Often, these come from some metal ion (hence a positive charge) that is incorporated into the dielectric during the growth stage. (Oxide/nitride layers are typically grown in a low-pressure chemical vapor deposition tool (thicker layers) – or a plasma enhanced chemical vapor deposition tool (thinner layers). While these tools are very clean compared to previous generations of processing tools, one can still get contamination from the walls of the processing tool. - Understanding the mechanisms involved in this contamination process is an area of significant research.) Additionally, because of the growth mechanisms, interface traps can be produced at the boundary between the Oxide and the semiconductor. These are typically dangling bonds that were broken during growth of the oxide. (A Si atom is stripped from the surface to make SiO₂ – leaving a ‘hole’ in the bond structure.) Additionally – and not talked about in the book – interface traps increase as the amount of total (time integrated) current – hence charge – increases. There is a reasonably well-quantified total charge, known as the ‘charge-to-breakdown’ at which the oxide fails. This is also an issue in processing of devices and is the subject of a yearly conference.

The important thing about these defects is that they add charge to the system, further shifting the energy levels away from our ideal ‘flat band’ condition. The voltage shift is given by:

$$V_{\text{trap}} = \frac{Q_{\text{trap}}}{C_i}$$

which, based on the location of the charge, is a positive value.

Thus our total voltage shift between the metal and the bulk semiconductor is given by:

$$\begin{aligned} \Phi_{\text{total}} &= -\Phi_{ms} + V_{\text{trap}} \\ &= -\Phi_m + \Phi_s + \frac{Q_{\text{trap}}}{C_i} \end{aligned}$$

Now comes the question of how we deal with these extra terms.

If one looks at the energy diagram above, one notes that we can recover most of our ‘flat band’ diagram (except the constant Fermi energy) by applying a voltage to the gate that simply shifts the energies in the semiconductor until we reach our flat band condition. This shift is obviously just the negative of Φ_{total} . If we do that, then we have effectively the same system that we started this subject with and hence we have all of our equations. We define the flat band voltage as:

$$\begin{aligned} V_{\text{fb}} &= \Phi_{ms} - V_{\text{trap}} \\ &= \Phi_m - \Phi_s - \frac{Q_{\text{trap}}}{C_i} \end{aligned}$$

Now our threshold voltage for inversion is shifted by the same amount and thus:

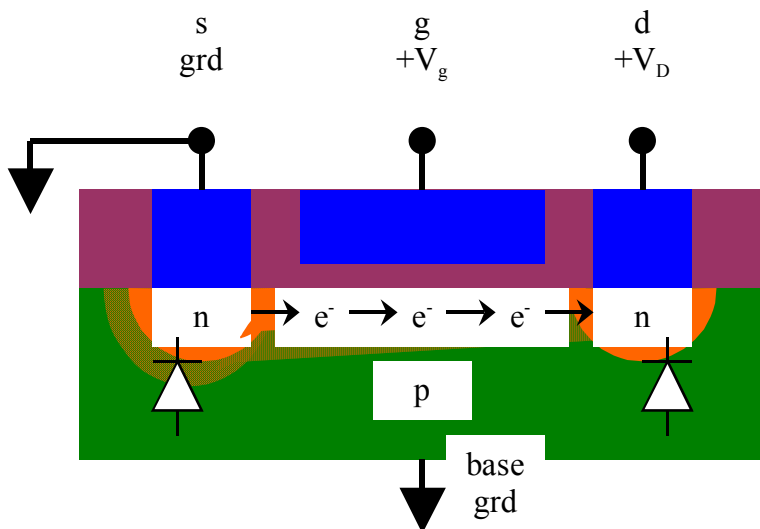
$$V_T = \phi_{ms} - \frac{Q_{si}}{C_i} - \frac{Q_{trap}}{C_i} + 2\phi_F$$

Current-voltage characteristics at the drain

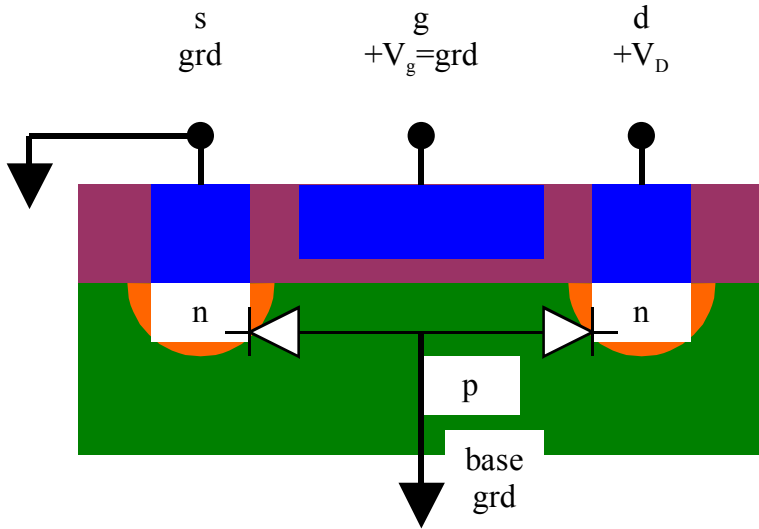
Now we are going to try to look at the drain current in a little detail. (Note that what we are doing is still an approximation. There are better approximations – but they are more difficult to arrive at than what we will do here.)

The basic idea of a MOS FET is fairly simple:

- 1) V_D is reversed biased
- 2) V_G is biased to attract minority carriers to the oxide interface.

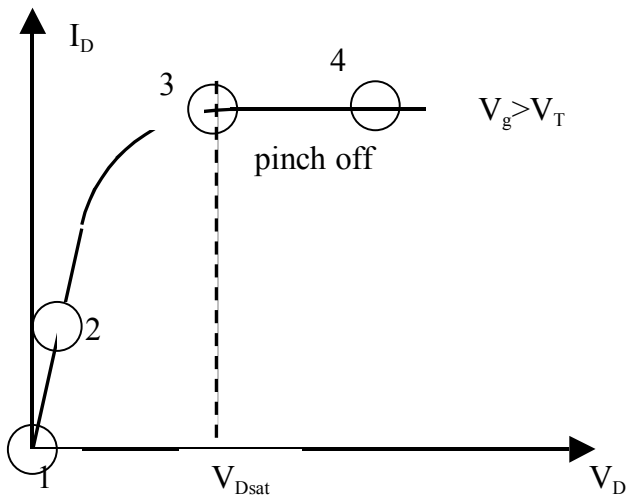


Now, if the channel does not exist, we get a device that looks like:

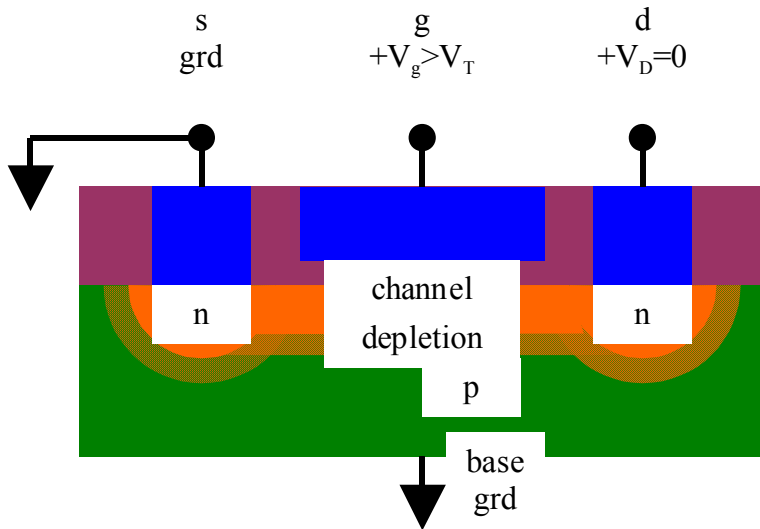


- 1) We see that no matter how we bias the drain relative to the source, we always have a junction that is reversed biased – and hence only a small amount of current can penetrate.
- 2) If a channel does exist, the channel simply acts like a variable resistor.

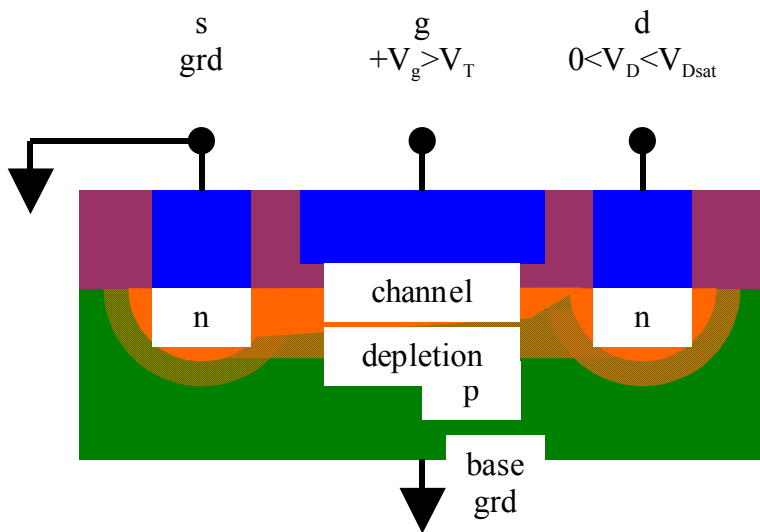
We know that the current voltage trace looks something like:



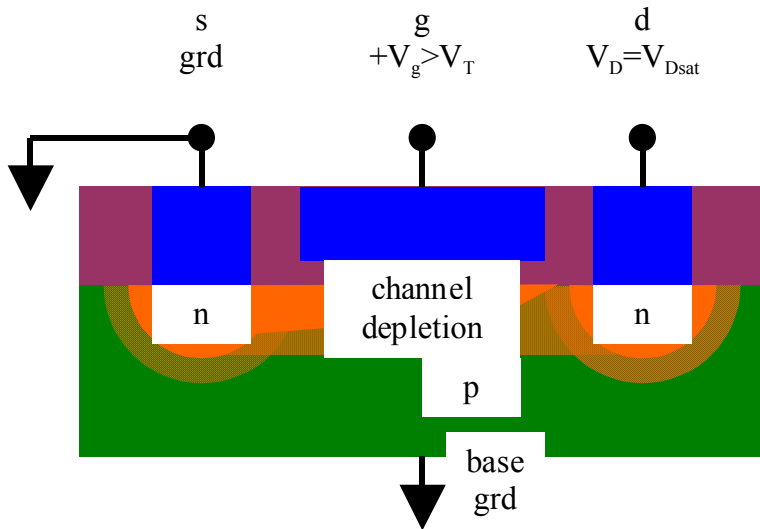
We have four important situations to look at in the above figure. They are labeled 1, 2, 3 and 4. Starting with '1'. This is under the condition that $I_{DS} = 0$ because $V_{DS} = 0$. Our picture looks like:



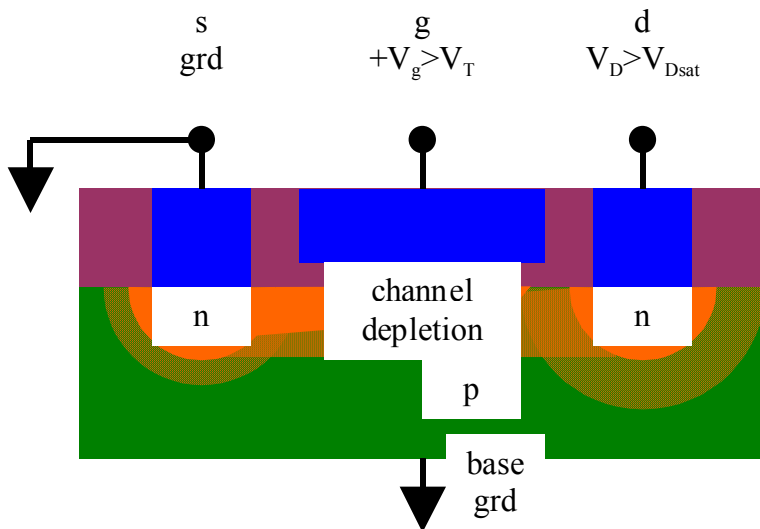
- 1) Wide open channel but no drain current because no drain voltage. **[Strictly speaking $V_g \geq V_T$ in all of these figures but MS\$ will not let me use that without crashing on print (for MacOS 9). It reads fine for Mac OSX but it does not even read the '≥' in Windows 2000.]**



- 2) Drain current because drain voltage not zero. Channel still open to current flow. Note that the channel width is being constricted at the end because the drain source voltage reduces the gate bulk voltage.



- 3) Pinch off occurs at this voltage. From now on the electric field in the channel remains \sim constant because the drain voltage is dropped across the depletion region. (This is the long channel approximation. If the channel is short, then the electric field across the channel changes as the drain voltage increases – the inversion length gets notably shorter but the voltage drop is the same. This is known as the ‘short channel effect’.)



- 4) Here all additional drain voltage get dropped across the depletion region and does not have much impact on the source-drain current.

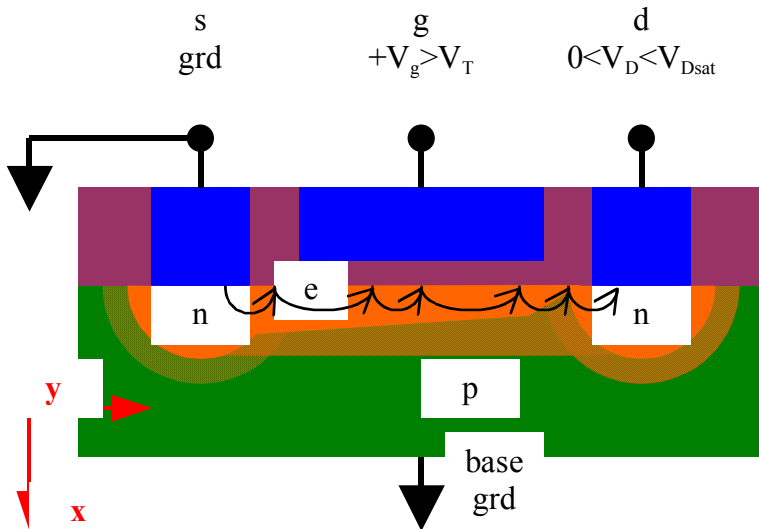
At this point, we need to look at which equations do we have.

We know that for an ideal MOSFET inversion occurs when the gate voltage reaches threshold voltage

$$V_T \approx 2\phi_F + \frac{\epsilon_{rs}}{\epsilon_{ri}} x_0 \sqrt{\frac{4qN_A\phi_F}{\epsilon_{rs}\epsilon_0}} \quad \Leftarrow \text{n-channel devices (as drawn above)}$$

$$V_T \approx 2\phi_F - \frac{\epsilon_{rs}}{\epsilon_{ri}} x_0 \sqrt{\frac{4qN_A(-\phi_F)}{\epsilon_{rs}\epsilon_0}} \quad \Leftarrow \text{p-channel devices (exact opposite to what is drawn above).}$$

Clearly when this occurs, the charge carriers, electrons for the n-channel, holes for the p-channel, move in the channel near the semiconductor-insulator interface. Additionally, the electric field in that area is such that it tries to pull the charge carriers closer to the interface. Remember that we have the field due to the drain and the field due to the gate! This means that we will have a path that looks like:



If the electron path were entirely in the Si, then we would already have model of its movement, via its mobility. Remember:

$$\mu = \frac{v_{\text{drift}}}{E}$$

The mobility is determined by the numbers and types of collisions in the bulk lattice. Now we also have collisions with the interface region, which we know has more damage sites than the regular lattice. (This is due to how the device is made and due to the fact that the semi and insulators rarely (never) match exactly causing stresses and strains. Therefore we expect that the mobility of the electrons will be lower than in the bulk material. In fact, we expect that the mobility is a function of position in the channel. Thus,

$$\mu \Rightarrow \mu(x, y)$$

Why $\mu(x)$?

- The mobility near the surface is less because of collisions with the interface.

Why $\mu(y)$?

- The channel width narrows as we move from source to drain – meaning in general that the electrons spend more time near the surface as we move from source to drain.
- The channel doping can change across the channel – and often is setup to do so.
- The electric field varies across the channel (This is more important in short channels)

Because of this, we need to define a new mobility, known as the ‘channel’ mobility or ‘effective’ mobility. It is simply the average mobility/area of all electrons (or holes) in the channel

$$\bar{\mu}_n = \langle \mu_n \rangle \equiv \frac{\int_0^{x_c(y)} \mu_n(x,y) n(x,y) dx}{\int_0^{x_c(y)} n(x,y) dx}$$

where $x_c(y)$ is the width of the channel at y .

Now the denominator of the above equation is essentially the number of charge carriers/area in the channel.

$$Q_n(y) = -q \int_0^{x_c(y)} n(x,y) dx$$

so we can rewrite this as

$$Q_n(y) \bar{\mu}_n = -q \int_0^{x_c(y)} \mu_n(x,y) n(x,y) dx$$

Now we have the mobility as a function of position along the length of the channel. If we think about things for a minute, we will realize that the channel is never very wide, and in general is approximately a constant. (This is at least true for long channels.) Thus we can make an approximation that the mobility is a constant along the channel. However, we do know that the mobility is a function of the applied gate voltage – relative to the threshold voltage. An empirical fit, based on experimental measurements gives

$$\bar{\mu}_n \approx \begin{cases} 0 & V_{gs} < V_T \\ \frac{\bar{\mu}_{n0}}{1 + \theta(V_{gs} - V_T)} & V_{gs} \geq V_T \end{cases}$$

where $\bar{\mu}_{n0}$ is the mobility at $V_{gs} = V_T$, and θ is a voltage fitting parameter.

Now the current density in the channel is

$$\begin{aligned} \mathbf{J} &= \mathbf{J}_n = J_{ny} \hat{y} \\ &= qn\mu_n \mathbf{E}_y + qD_n \nabla n \\ &= qn\mu_n \frac{-d\phi(y)}{dy} \hat{y} + qD_n \frac{\approx 0}{\nabla n} \\ &= -qn\mu_n \frac{d\phi(y)}{dy} \end{aligned}$$

The total source-drain current is the integral over the area of the channel

$$\begin{aligned} I_{DS} &= - \int \int \mathbf{J} dA \\ &= + \int_0^{x_c(y)} \int_0^Z qn\mu_n \frac{d\phi(y)}{dy} dx dz \\ &= + Z \int_0^{x_c(y)} qn\mu_n \frac{d\phi(y)}{dy} dx \end{aligned}$$

If we now assume that the electric field does not change in the channel, then we can pull it out of the integral. (Again, we note that this really only works for long channels.)

$$\begin{aligned}
I_{DS} &= +Zq \frac{d\phi(y)}{dy} \int_0^{x_c(y)} n \mu_n dx \\
&= -Z \frac{d\phi(y)}{dy} Q_n(y) \bar{\mu}_n
\end{aligned}$$

We can get rid of our y-dependence by noting that the total current must be a constant along the junction. Then if we integrate from the source to the drain, we should have $I_{DS}L$. (We have used this trick before!)

$$\begin{aligned}
\int_0^L I_{DS} dy &= I_{DS}L \\
&= \int_0^L -Z \frac{d\phi(y)}{dy} Q_n(y) \bar{\mu}_n dy \\
&= \int_0^{V_{DS}} -Z Q_n(\phi) \bar{\mu}_n d\phi
\end{aligned}$$

Now all that we need is the local charge as a function of local voltage. We know

$$Q_n(y) = -q \int_0^{x_c(y)} n(x,y) dx$$

If we think about where the charges are located at, the charge on the gate must equal the charge on the semiconductor-insulator interface (e.g. the channel) plus the depletion charge.

$$Q_{mi} = -Q_{channel} - Q_{depletion}$$

If our gate voltage is above the threshold then any change in the gate voltage will give rise to changes only at the metal-insulator and semi-insulator interfaces.

$$\Delta Q_{mi} = -\Delta Q_{channel}$$

We know what this charge is from the capacitance across the insulator

$$-\Delta Q_{channel} = C_i V_G$$

⇓

$$-\Delta Q_{channel} = C_i (V_G - V_T)$$

⇓

$$\Delta Q_{channel} = -\frac{\epsilon_{ri} \epsilon_0}{x_0} (V_G - V_T)$$

But this is for a MOS capacitor. We know that the voltage will vary as we move across the device. The amount that it reduces is $\phi(y)$. Thus, we arrive at:

$$\Delta Q_{channel} = -\frac{\epsilon_{ri} \epsilon_0}{x_0} ((V_G - \phi(y)) - V_T)$$

we can now plug this into our equation for the current to arrive at:

$$\begin{aligned}
\int_0^L I_{DS} dy &= I_{DS} L \\
&= \int_0^{V_{DS}} -Z Q_n(\phi) \bar{\mu}_n d\phi \\
&= \int_0^{V_{DS}} -Z \left[-\frac{\epsilon_{ri} \epsilon_0}{x_0} ((V_G - \phi(y)) - V_T) \right] \bar{\mu}_n d\phi \\
&\Downarrow \\
I_{DS} &\approx \frac{Z \bar{\mu}_n C_i}{L} \left[(V_G - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] \quad \text{for } V_G \geq V_T, \quad 0 \leq V_{DS} \leq V_{Dsat}
\end{aligned}$$

So where does pinch off occur? When $Q(L) = 0 \dots$ From above

$$\begin{aligned}
Q_{\text{channel}}(L) &= -\frac{\epsilon_{ri} \epsilon_0}{x_0} ((V_G - \phi(L)) - V_T) \\
&\approx -\frac{\epsilon_{ri} \epsilon_0}{x_0} ((V_G - V_{DS}) - V_T) \\
&= 0 \\
&\Downarrow
\end{aligned}$$

$$V_{Dsat} = (V_G - V_T)$$

Plugging this into the above equation, we find the saturation current

$$I_{DS} \approx \frac{Z \bar{\mu}_n C_i}{2L} [(V_G - V_T)]^2 \quad \text{for } V_G \geq V_T, \quad 0 \leq V_{DS} \leq V_{Dsat}$$

Note that the saturation current does not depend on the drain voltage but does depend on the gate voltage less the threshold in a square-law term.

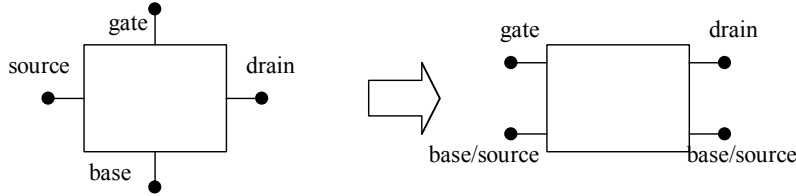
Like most things we have done in this class, we have used approximations to arrive at simple models of our devices. This means that the results are subject to errors and hence can be improved on. For example, if we had included the fact that the depletion region varies with location across the device, we would have arrived at a better model of the current-voltage relationship. This model is known as the 'bulk-charge theory'. This distinction gives

$$\begin{aligned}
\Delta Q_{\text{channel}} &= -\frac{\epsilon_{ri} \epsilon_0}{x_0} (V_G - V_T - \phi(y)) + q N_A [w(y) - w_T] \\
I_{DS} &\approx \frac{Z \bar{\mu}_n C_i}{L} \left[(V_G - V_T) V_{DS} - \frac{V_{DS}^2}{2} - \frac{4}{3} \frac{q N_A w_T}{C_i} \phi_F \left(\left(1 + \frac{V_{DS}}{2\phi_F} \right)^{3/2} - \left(1 + \frac{3V_{DS}}{4\phi_F} \right) \right) \right] \\
V_{Dsat} &= (V_G - V_T) - \frac{q N_A w_T}{C_i} \left\{ \left[\frac{V_G - V_T}{2\phi_F} + \left(1 + \frac{q N_A w_T}{4 C_i \phi_F} \right)^2 \right]^{1/2} - \left[1 + \frac{q N_A w_T}{4 C_i \phi_F} \right] \right\}
\end{aligned}$$

These are considerably harder to deal with but more accurate.

Small ac signals on MOSFETs

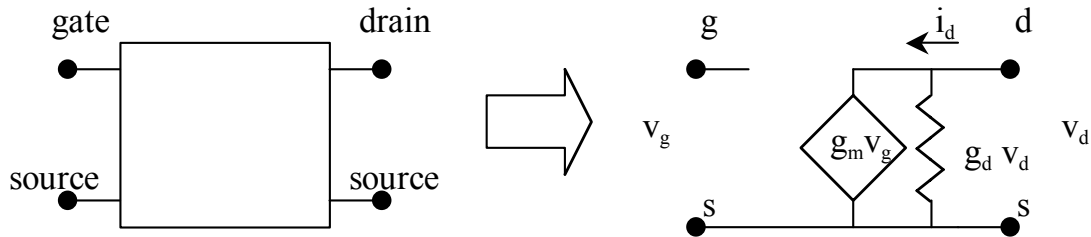
At this point it is time to look at small ac signals on our MOSFET. (This is following our typical pattern – qualitative assessment of dc I-V traces, a quantitative assessment of dc I-V traces, and then a quantitative assessment of small ac I-V traces.) To do this, we need to look at how we use the device.



In general, we put our signal in on the gate line and pull our signal out on the drain line – leaving the source/base as our effective ground (not unlike we have done in our analysis above).

For low frequencies, the current that flows between the gate and the source is very low. This is because we have a reversed biased diode. Thus the gate to source acts like an open circuit.

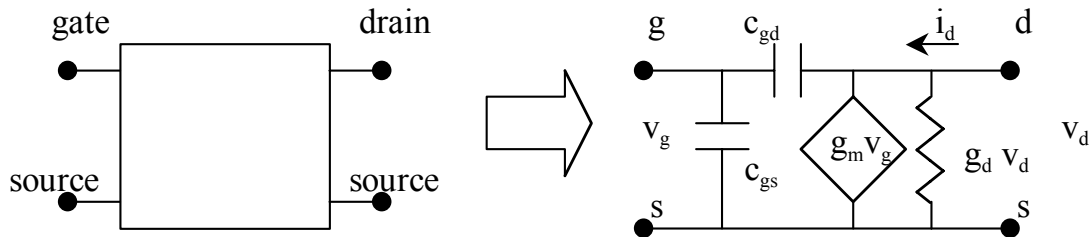
The current between the drain and the source is set by the conductance, g , between the source and the drain. Additionally the gate voltage will control change the conductance. Let g_d be the conductance at zero gate voltage. Then we get an additional conductance, g_m , that is set by the gate voltage.



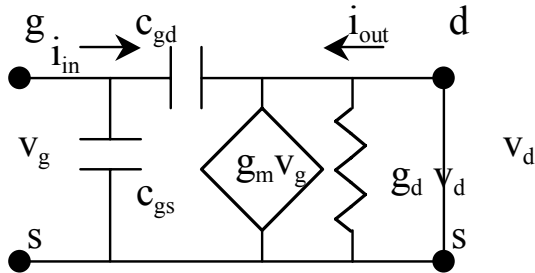
$$g_d \equiv \left. \frac{\partial I_{DS}}{\partial V_{DS}} \right|_{V_G = \text{const}} \equiv \text{channel conductance}$$

$$g_m \equiv \left. \frac{\partial I_{DS}}{\partial V_G} \right|_{V_{DS} = \text{const}} \equiv \text{transconductance}$$

Now let us go high frequency. Here, the gate connects to both the drain and the source as if there were capacitors between the terminals. Thus we get



We can now use standard circuit analysis to arrive at our current and voltage relationships. However, this is exactly like what we found for our JFET so we will not repeat it. We can however try to determine what the maximum operational frequency of the device might be. We will define the maximum frequency as the frequency at which the current in is equal to the current out. This means that we have an ac 'short' between source and drain. Our circuit is now



Thus

$$i_{in} = \omega(C_{gs} + C_{gd})v_g$$

$$\approx \omega(C_i A_g)v_g$$

$$i_{out} = g_m v_g$$

⇓

$$\frac{i_{out}}{i_{in}} = 1$$

$$= \frac{g_m v_g}{\omega(C_i A_g)v_g}$$

$$= \frac{g_m}{\omega(C_i A_g)}$$

⇓

$$\omega_{max} \approx \frac{g_m}{(C_i A_g)} = \frac{1}{\underbrace{C_i A_g}_{=ZL}} \left[\frac{Z\mu_n C_i}{L} (V_g - V_T) \right] \leftarrow \text{from JFETs}$$

⇓

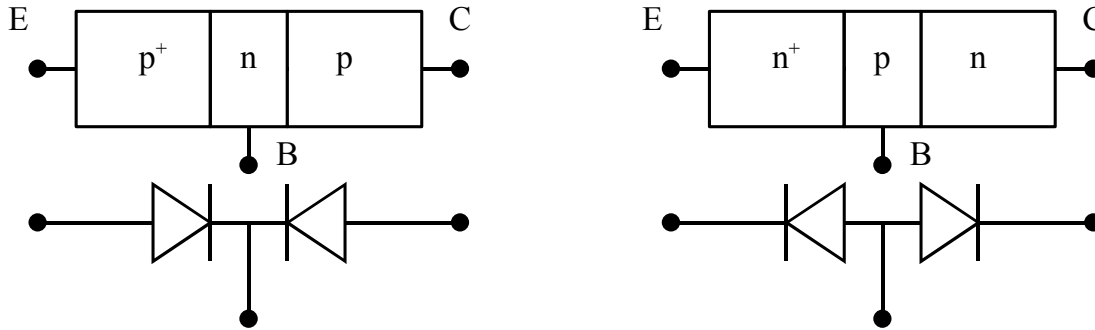
$$f_{max} = \frac{1}{2\pi} \left[\frac{\mu_n}{L^2} (V_g - V_T) \right]$$

This means the faster we want to operate, the shorter we need the channel to be...

BJT – Bipolar Junction Transistors

This is the last topic of this class.

The Bipolar Junction transistor is a relatively simple device. In essence, one takes a p⁺-n junction and adds a p-type material on the other side of the n material. (One can also do this with a n⁺-p junction.) They look like:



Here ‘E’ stands for **emitter**, ‘C’ stands for **collector** and ‘B’ stands for **base**. This terminology will become clear shortly.

To understand this device we need to consider (or rather remember) what happens with just a p-n junction. Reaching back into the notes and looking at our diodes, we find that:

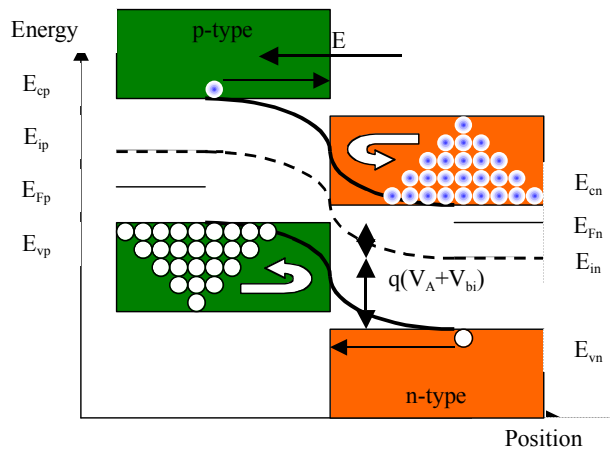
$$I = I_0 \left(e^{qV_A/kT} - 1 \right)$$

$$I_0 = qn_i^2 A \left(\frac{D_n}{L_n N_A} + \frac{D_p}{L_p N_D} \right) \quad \text{where} \quad L = \sqrt{D\tau}$$

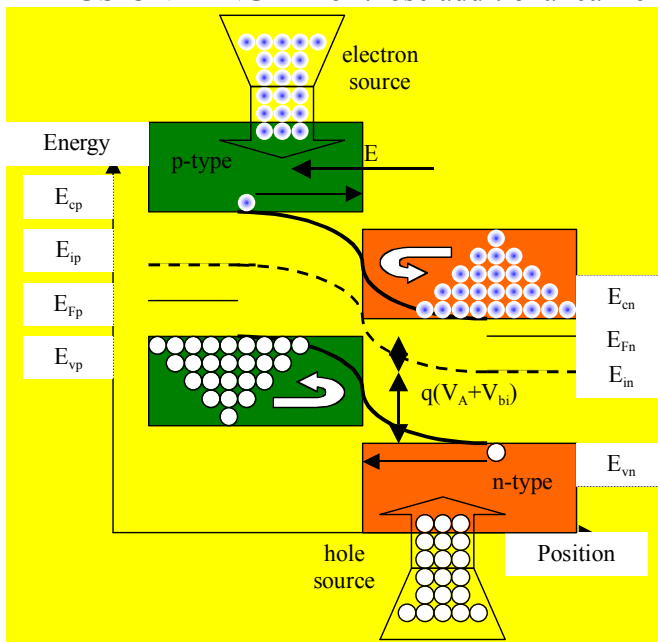
We see that in the forward biased conditions, we get an exponentially increasing current. However, we see from our picture of a BJT that at least one of the junctions is always reversed biased. Thus we really need to examine what happens in the reversed bias conditions. Under such a condition, the current is

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

This means that the current is independent of the applied bias but it is dependent on the diffusion length of the carriers – and hence the drift of electrons/holes across the junction. Since these are the minority carriers, very few of them are around. So I_0 depends mostly on the production rate of the minority carriers **near the junction**. This last part is due the fact that the minority carriers really need to be generated within a diffusion length of the junction for them to ‘fall’ down the potential hill. (Some minority carriers will come from further away, while some minority carriers created near the junction will recombine before they fall down the hill. On average, many of the minority carriers within L will make it while those outside that length do not make it.) Our picture of the situation was:

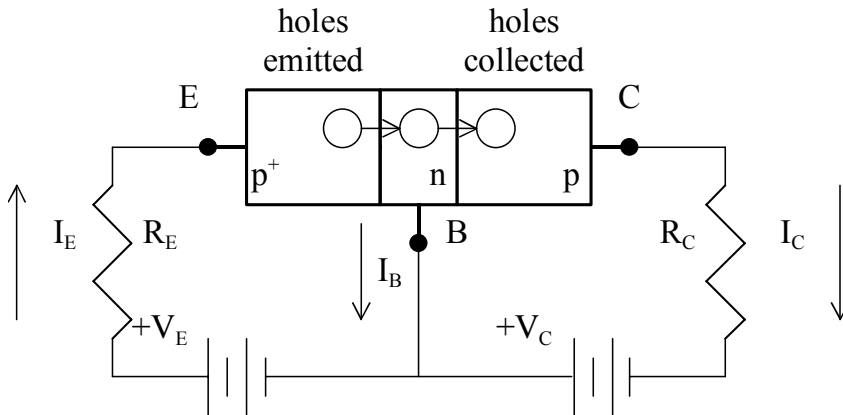


If somehow we could place a source of minority carriers near the junction, we could then increase the current across the junction. By near, we mean that the minority carriers MUST BE WITHIN A DIFFUSION LENGTH for those additional carriers to make a difference.



One way to create this source of minority carriers is to put a second junction, near our first junction, through which we supply our required minority carriers. This is the nature of a BJT device.

Let's now examine a 'typical' – or – 'standard' BJT. In the typical configuration it is wired as follows: (We will look at only p^+n-p devices right now.)



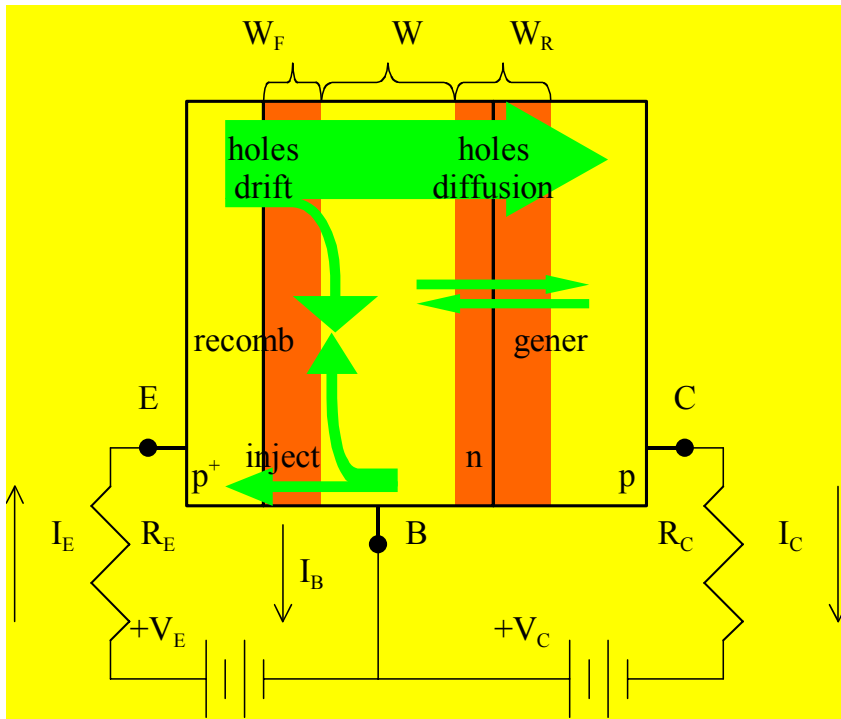
THIS IS WIRED CORRECTLY! - see the notes below

(NOTE that we can wire a BJT in a number of different ways. If you can dream it, it has probably been done.)

We see in this case that the emitter-base (E-B) junction is forward biased while the collector-base (C-B) junction is reverse biased. This means that holes are injected from the emitter into the base. These holes are a minority carrier in the n-type material. They are now available for transfer across the reverse biased C-B junction to the collector. We want these holes to make it across the device so:

- a) Make the E-B junction a p^+-n junction so that we have a lot of holes injected.
- b) Have the E-B junction forward biased under normal operation.
- c) Make the base region very narrow, such that the width, W , is less than L_p . (Here W is the width of the non-depleted region of the base material.)
- d) Strongly reverse bias the C-B junction.

Qualitatively what occurs is:



THIS IS WIRED CORRECTLY!

Notes:

- 1) The holes **drift** (follow the applied **E** field) **across the E-B junction** and **diffuse across the C-B junction**.
- 2)
 - a. p^+ - region 'emits' holes into the n-region (emitter)
 - b. p - region 'collects' holes out of the n-region (collector)
 - c. n - region called base because it used to be the physical base of the device – e.g. historical reasons
 - d. Base region acts a common for this particular configuration => Called 'Common Base' configuration. [Also can have 'common-emitter' and 'common-collector' <- last one is very rare.]
- 3) Since we want most of the hole to go from the emitter to the collector, we want
 - a. $W \ll L_p \Rightarrow$ small number of recombinations
 - b. τ_p is large in the base – hence the holes live a long time relative the time that it takes to get across the base region.
- 4) For efficiency choose p^+ -n-p type device
 - a. Hole injection dominates electron injections
- 5) Currents
 - a. I_E determined by external circuit. If strongly forward biased $I_E \sim V_E/R_E$.
 - b. I_C determined usually by I_E . $I_C \sim I_E$.
 - c. I_B is determined by
 - i. Electron recombination
 - ii. Electron injected into emitter
 - iii. Much small part due I_0 lower I_B by a very small amount.

BJT are used as amplification devices. Where does the amplification occur?

$I_C \sim I_E$ so this is not amplification
 $I_C \gg I_B$ so this might be amplification.
 So how does the BJT amplify?

From the figures above, we know that:

$$I_E = I_{Ep} + I_{En}$$

$$I_B = I_{Bp} + I_{Bn}$$

and

$$I_C = I_{Cp} + I_{Cn}$$

Note:

$$I_E \approx I_C \gg I_B$$

(E= emitter, B = base, C=collector, p=holes, n=electrons)

We can also see that the hole current in the collector is proportional to the hole current in the emitter.

$$I_{Cp} = \alpha_T I_{Ep} \quad 0 < \alpha_T < 1$$

where α_T is the base transport factor. (Basically how many holes make it across the base.)

We can also define the emitter injection efficiency: (Basically how large is the hole current compared to the background electron current.)

$$\gamma = \frac{I_{Ep}}{I_E} \quad 0 < \gamma < 1$$

so that

$$I_{Cp} = \alpha_T \gamma I_E$$

$$\equiv \alpha_{dc} I_E$$

where α_{dc} is the current transfer ratio or common base dc current gain

Now where do the currents go?

$$I_B = I_E - I_C$$

$$\approx I_E - \alpha_{dc} I_E$$

$$\approx \frac{1}{\alpha_{dc}} I_C - I_C$$

rearranging

$$I_C = \frac{\alpha_{dc}}{1 - \alpha_{dc}} I_B$$

$$\equiv \beta_{dc} I_B$$

where β_{dc} is the gain between the base and the collector currents. This is our amplification.

Note that β_{dc} can range over

$$0 \leq \beta_{dc} \leq \infty$$

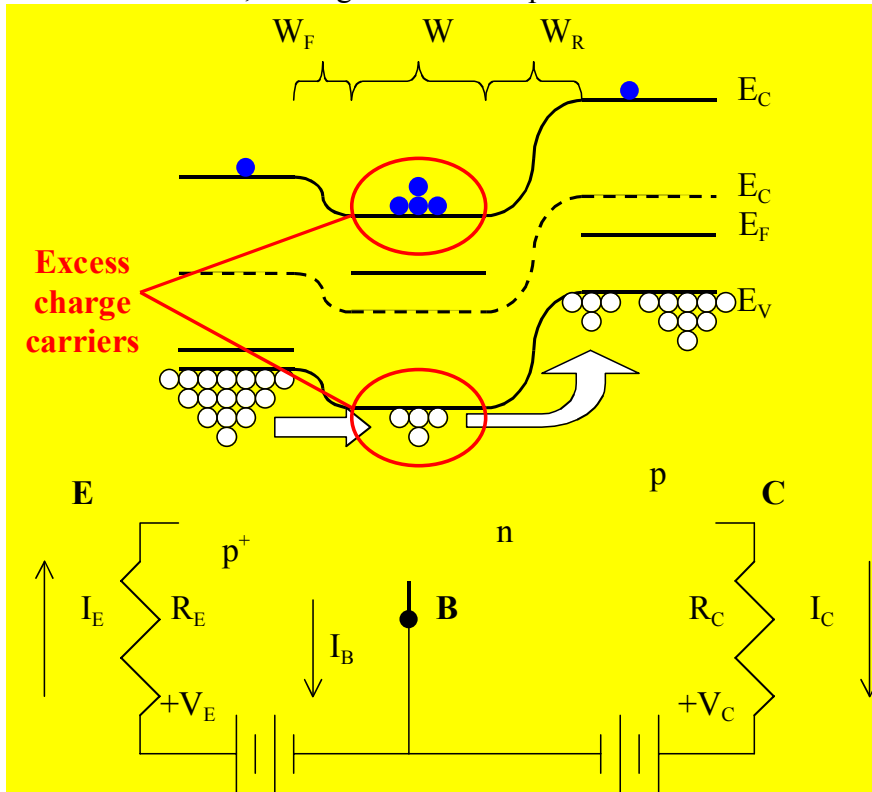
This leads to two questions.

- 1) How much larger is I_c than I_B ?
- 2) How is I_c controlled by I_B ?

To begin with, the base region (here we only consider the non-depletion zone) has to remain charge neutral or else the E-B junction voltage will change. Thus for every hole passing through, there must be an extra electron.

However, holes and electrons do not have to spend the same amount of time in the base region!

To understand this, let us go back to our picture of the device. Using our energy diagram we see



We see that the holes are not confined to the base region, and pass through quickly. (They are accelerated in to the region by the electric field on the E-B side and diffuse across the C-B junction.) The electrons on the other hand are trapped inside the base region. **(Note that we are only showing the excess charge carriers on the base region!)**

How long is each of the charge carriers there? The excess electrons will stay inside the base region until they are removed via recombination. Thus the electrons have a lifetime of τ_p . The excess holes, on the other hand, are only there long enough to pass through. We will call this time, τ_T . (This is kind of like measuring the difference in time people spend in a small town along a highway. Those that live there, stay for 70 years. Those that pass through stay for a few seconds at most.) From this, we would expect that for each excess electron that put into the base, τ_T/τ_p holes pass through the base. As the excess electrons can be gathered by the base contact, the lifetime is a little shorter than τ_p . However, this indicates that this is the source of the base current. Thus we find that

$$\beta_{dc} \approx \frac{\tau_T}{\tau_P}$$

This explains why
 $I_C \gg I_B$.

Now, why does the base control the collector?

- 1) Pretend that we lower I_B , what happens to I_C ?
 - a. The number of excess electrons trapped in the base is reduced
 - b. \Rightarrow There is a build up of holes in the base
 - c. \Rightarrow The current from the Emitter into the base must be reduced to balance the charge in the base. (e.g Reduce the E-B junction forward bias voltage)
 - d. \Rightarrow A reduction in the current from the emitter to the collector.
- 2) Now pretend that we raise I_B , what happens to I_C ?
 - a. The number of excess electrons trapped in the base gets increased
 - b. \Rightarrow There is a build up of extra electrons
 - c. \Rightarrow There needs to be more holes injected to balance the charge
 - d. \Rightarrow The E-B forward bias must increase to increase the number of holes moving into the base region.

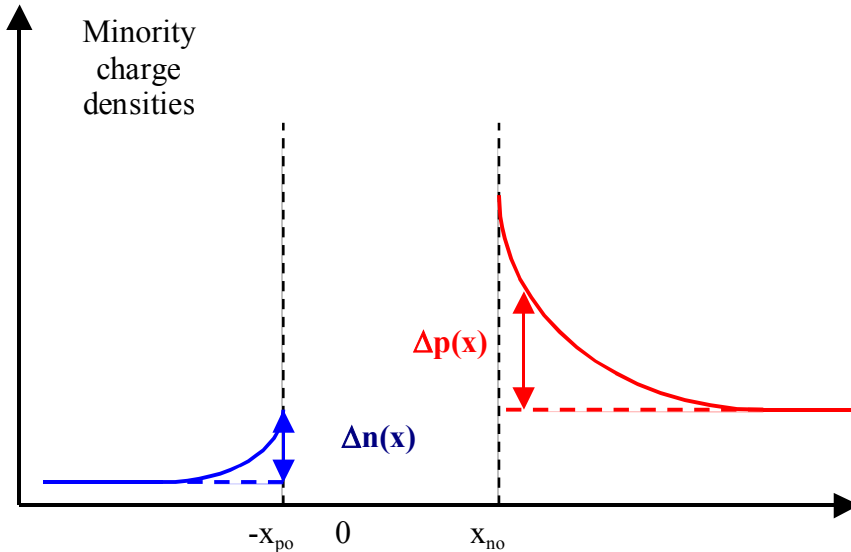
Let us now go through this in a little more detail.

To do this we need to examine the **Minority carrier distributions**, which will give us our **terminal currents**.

Let us start by making a few simple assumptions.

- 1) Ignore drift current in the base region. Why? Because the base is approximately neutral and thus the electric field will be very small.
- 2) Assume that the emitter current is dominated by the hole current. Why? Because we have a $p^+ - n$ junction and thus the hole will dominate the current across the junction.
- 3) Ignore the saturation current in the collector.
- 4) Assume steady state
- 5) Assume that the E-B area is the same as the C-B area. (Further assume that the areas are constant across the device.)

Pictorially this looks like:



Mathematically our excess charge carrier densities are given by:

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

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

$$L_{p/n} = \sqrt{D_{p/n} \tau_{p/n}}$$

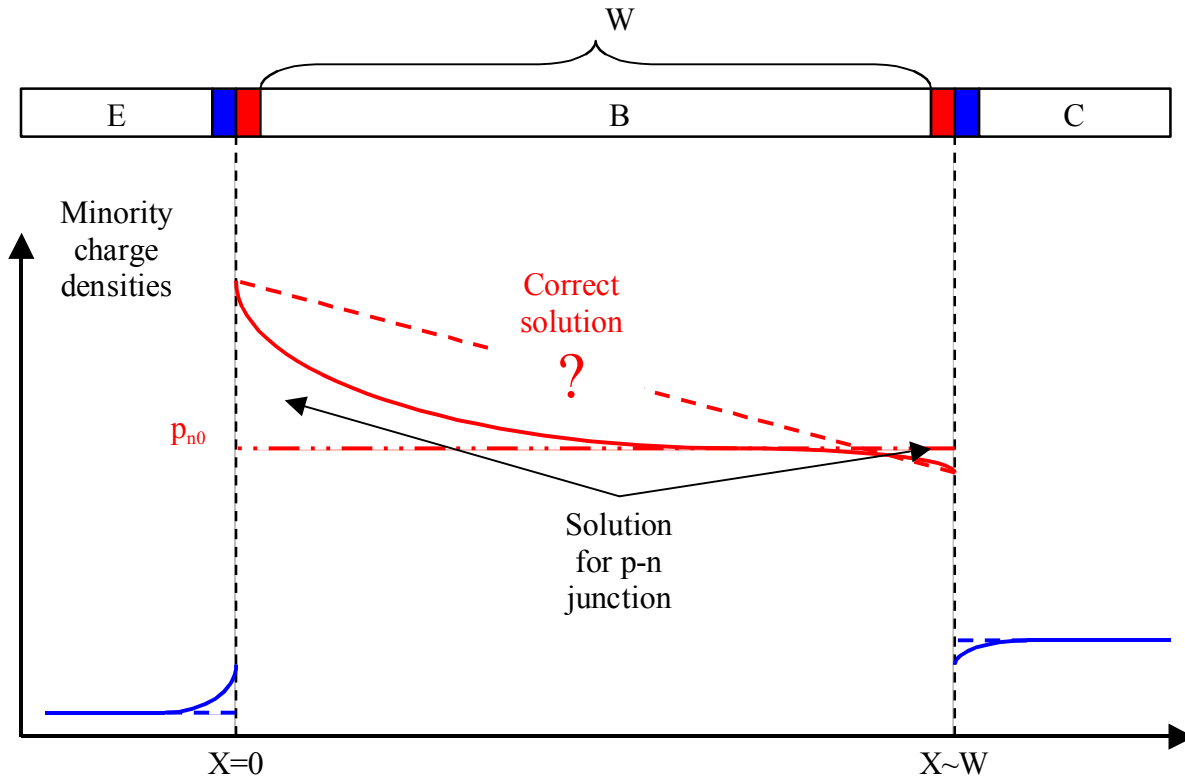
Typically the E-B junction is forward biased. Thus the number of excess charge carriers on the emitter side of the base is large.

$$\begin{aligned}\Delta p_{n\text{Emit}} &\equiv \Delta p_E = p_{n0} \left(e^{qV_A/kT} - 1 \right) \\ &\approx p_{n0} e^{qV_A/kT}\end{aligned}$$

Typically the C-B junction is reversed biased. Thus the number of excess charge carriers on the collector side of the base is small and negative.

$$\begin{aligned}\Delta p_{n\text{Collect}} &\equiv \Delta p_C = p_{n0} \left(e^{qV_A/kT} - 1 \right) \\ &\approx -p_{n0}\end{aligned}$$

This looks like:



In between, the holes move via diffusion. Thus to understand the current, we need to solve the diffusion equation.

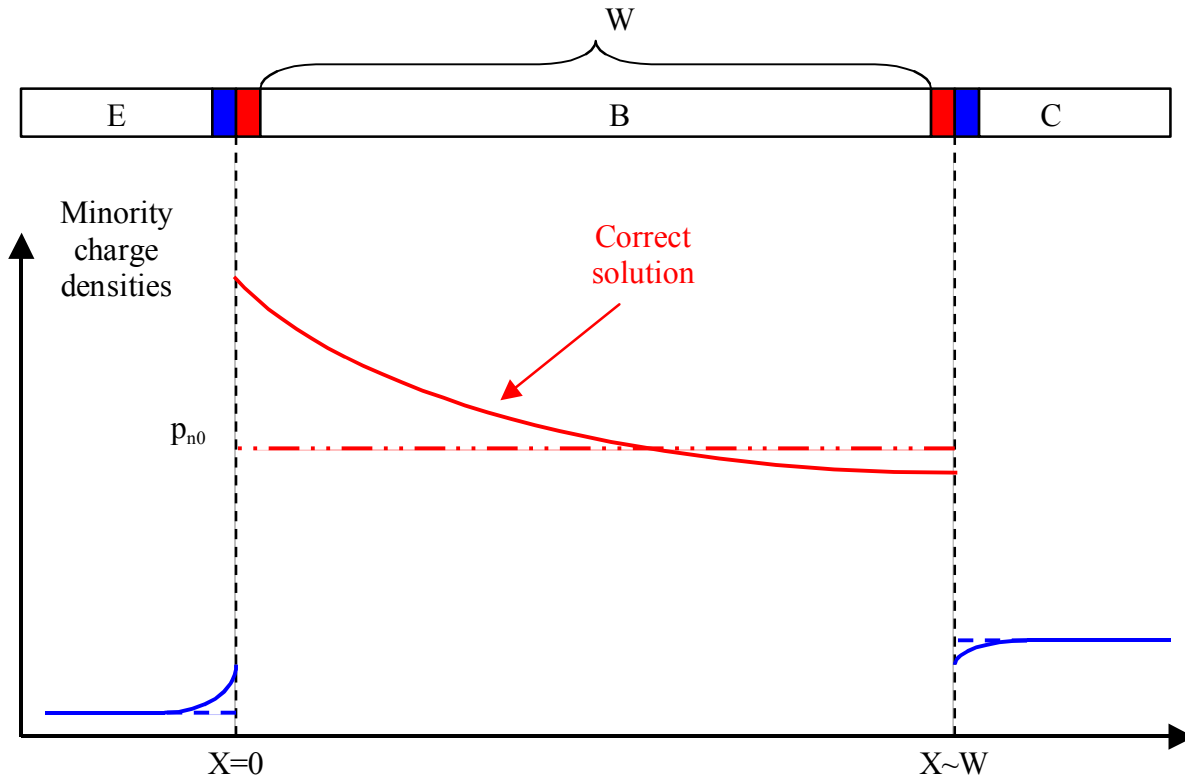
$$\frac{\partial^2 \Delta p(x)}{\partial x^2} = \frac{\Delta p(x)}{L_p^2}$$

We know, from our constantly solving this equation, that the solution is:

$$\Delta p(x) = A e^{+x/L_p} + B e^{-x/L_p}$$

Taking into account the hole densities at the two ends of the base material (e.g. at the material interface and not the edge of the 'base region') we get:

$$\Delta p(x) = \left\{ \frac{\Delta p_C - \Delta p_E e^{-W/L_p}}{e^{W/L_p} - e^{-W/L_p}} \right\} e^{+x/L_p} + \left\{ \frac{\Delta p_E e^{W/L_p} - \Delta p_C}{e^{W/L_p} - e^{-W/L_p}} \right\} e^{-x/L_p}$$



Our correct solution is very close to a straight-line solution. (The bow in the above drawing is greatly exaggerated.) We can now find the current from the slope of the minority carrier density

$$\begin{aligned}
 \frac{1}{q} \mathbf{J}_p &= p\mu_p \mathbf{E} - D_p \nabla \Delta p_n \\
 &= -D_p \nabla \Delta p_n \\
 &\Downarrow \\
 I &= A \mathbf{J}_p \\
 &= -AqD_p \nabla \Delta p_n \\
 &= -AqD_p \frac{\partial \Delta p_n}{\partial x}
 \end{aligned}$$

Applying this to both ends, we find:

$$\begin{aligned}
I_E &\approx I_{Ep} \\
&= -AqD_p \left. \frac{\partial \Delta p_n}{\partial x} \right|_{x=0} \\
&= -Aq \frac{D_p}{L_p} \Delta p(x=0) \\
&= -Aq \frac{D_p}{L_p} \frac{2\Delta p_C - \Delta p_E (e^{W/L_p} + e^{-W/L_p})}{e^{W/L_p} - e^{-W/L_p}} \\
&= Aq \frac{D_p}{L_p} \left\{ \Delta p_E \coth\left(\frac{W}{L_p}\right) - \Delta p_C \operatorname{csc}h\left(\frac{W}{L_p}\right) \right\}
\end{aligned}$$

$$\begin{aligned}
I_C &\approx I_{Cp} \\
&= -AqD_p \left. \frac{\partial \Delta p_n}{\partial x} \right|_{x=W} \\
&= -Aq \frac{D_p}{L_p} \Delta p(x=W) \\
&= Aq \frac{D_p}{L_p} \left\{ \Delta p_E \operatorname{csc}h\left(\frac{W}{L_p}\right) - \Delta p_C \coth\left(\frac{W}{L_p}\right) \right\}
\end{aligned}$$

Now, the base current is simply the difference between the two currents:

$$I_B = I_E - I_C$$

Note that these equations are functions of the bias voltages – notably the excess charge as the edge of the base material. As a consequence, we can rewrite the equations to give:

$$I_E = Aq \frac{D_p}{L_p} \left\{ p_{n0} (e^{qV_{EB}/kT} - 1) \coth\left(\frac{W}{L_p}\right) - p_{n0} (e^{qV_{CB}/kT} - 1) \operatorname{csc}h\left(\frac{W}{L_p}\right) \right\}$$

$$I_C = Aq \frac{D_p}{L_p} \left\{ p_{n0} (e^{qV_{EB}/kT} - 1) \operatorname{csc}h\left(\frac{W}{L_p}\right) - p_{n0} (e^{qV_{CB}/kT} - 1) \coth\left(\frac{W}{L_p}\right) \right\}$$

With the exception of the exponential voltage term, we have a lot of constants. Thus

$$I_E = I_{F0} \left(e^{qV_{EB}/kT} - 1 \right) - \alpha_R I_{R0} \left(e^{qV_{CB}/kT} - 1 \right)$$

$$I_C = \alpha_F I_{F0} \left(e^{qV_{EB}/kT} - 1 \right) - I_{R0} \left(e^{qV_{CB}/kT} - 1 \right)$$

$$\approx \alpha_R I_{R0} \left(e^{qV_{EB}/kT} - 1 \right) - I_{F0} \left(e^{qV_{CB}/kT} - 1 \right) \text{ (ignoring electron current)}$$

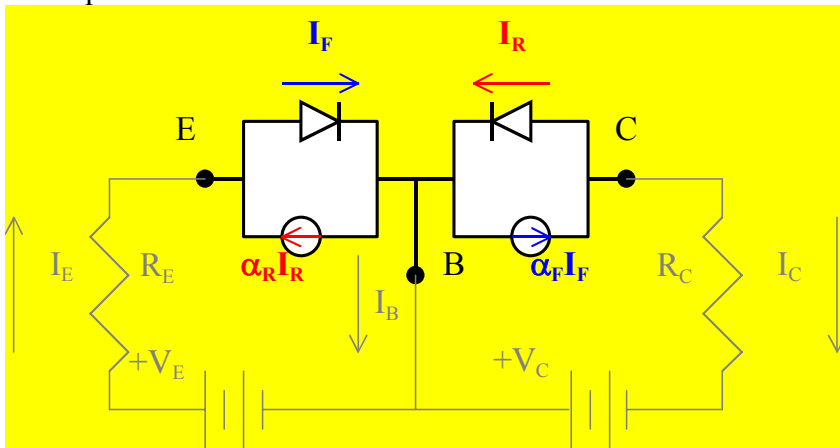
where

$$I_{F0} = Aq \frac{D_p}{L_p} p_{n0} \coth \left(\frac{W}{L_p} \right)$$

$$\alpha_R I_{R0} = Aq \frac{D_p}{L_p} p_{n0} \operatorname{csch} \left(\frac{W}{L_p} \right)$$

These are known as the Ebers-Moll Equations.

The equivalent circuit for this device is then:

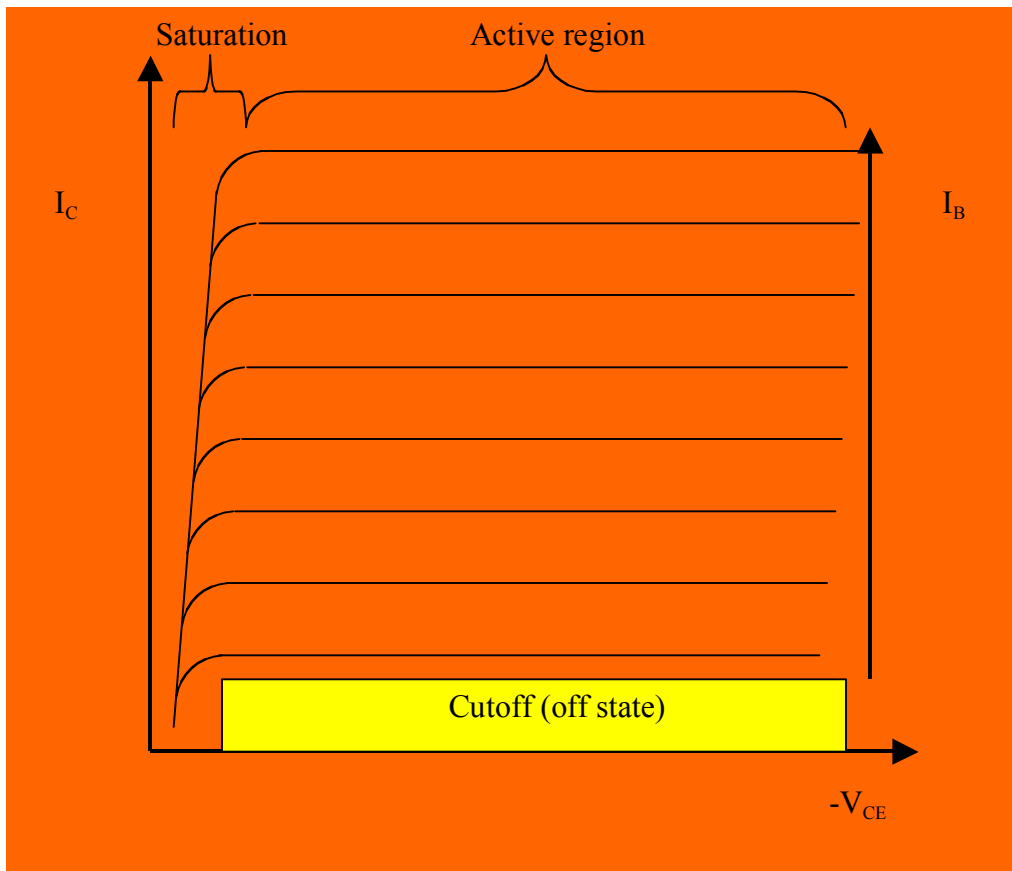


where

$$I_F = I_{F0} \left(e^{qV_{EB}/kT} - 1 \right)$$

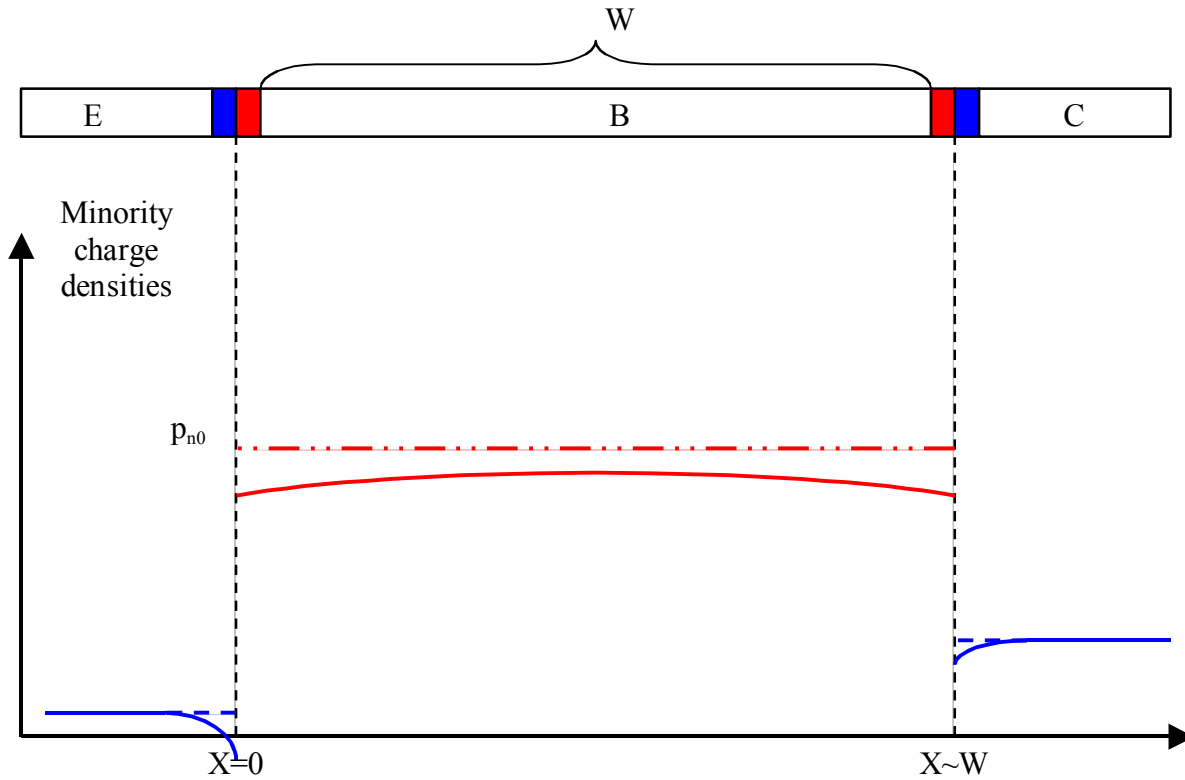
$$I_R = I_{R0} \left(e^{qV_{CB}/kT} - 1 \right)$$

What do these look like?



We have

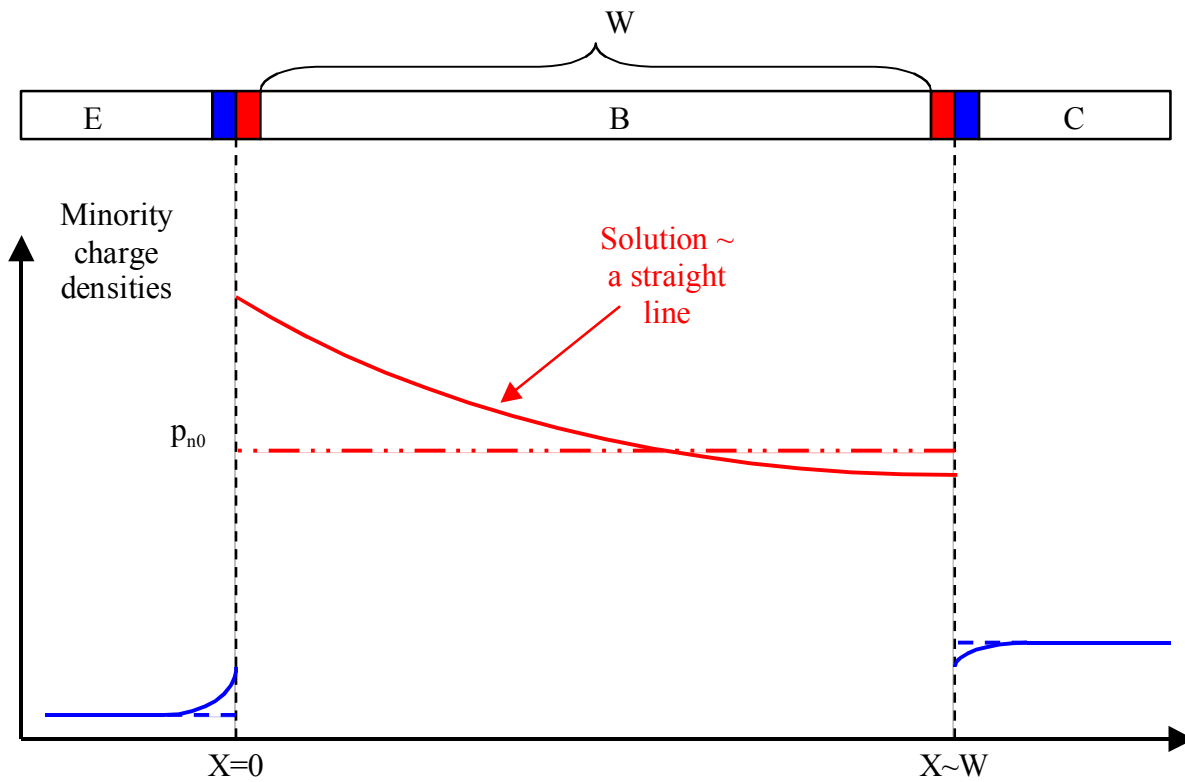
- a) Cutoff: What is happening?
 - a. $I_B = 0$ (or very small)
 - b. $I_C \sim 0$
 - c. I_B supplies electrons for recombination \Rightarrow no electrons are required \Leftrightarrow no excess holes
 - d. \Rightarrow We have reversed biased both junctions



b) Active Region: What is happening?

- a. V_{CE} is very large
- b. $I_B = \text{normal} \Rightarrow$ excess holes are present
- c. $I_C = \beta_{dc} I_B$

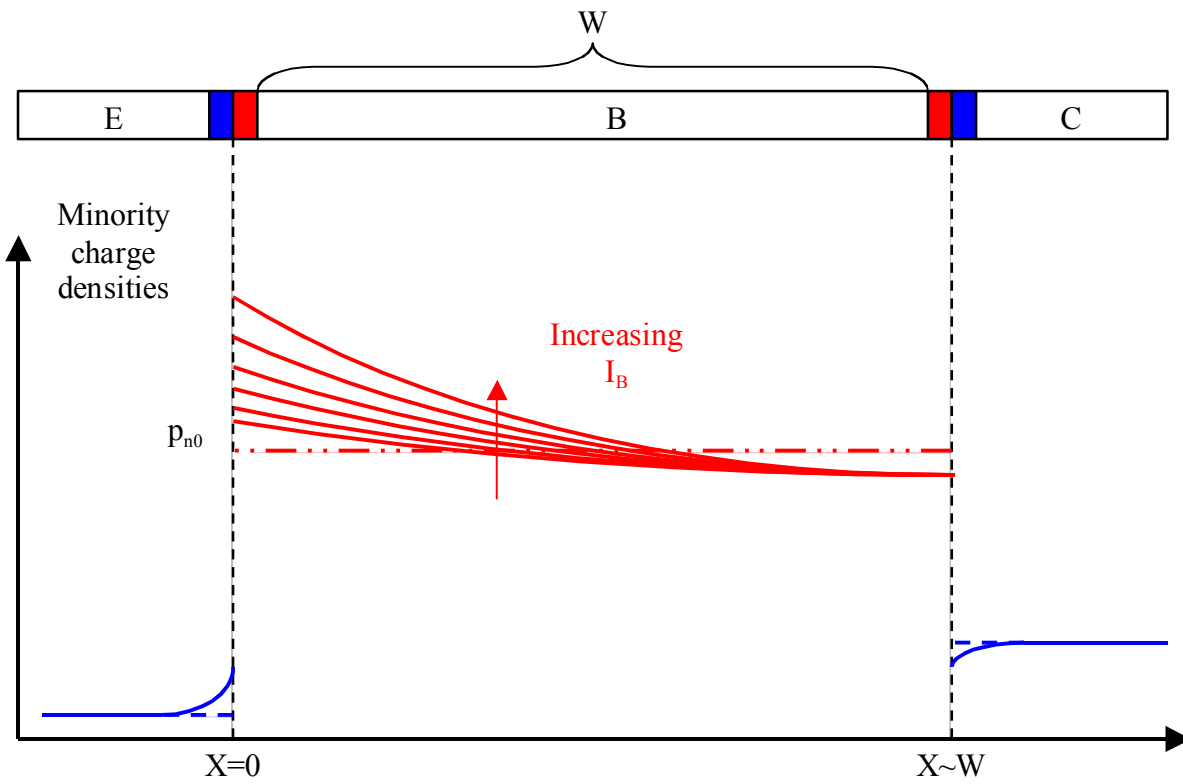
We have our normal operation and the minority charge carriers in the central region are as shown in the figure below



c) Saturation Region: What is happening?

- a. $V_{CE} \sim 0$ but not quite zero
- b. $I_B > 0$
- c. $I_C \neq \beta_{dc} I_B$

The very low V_{CE} means that we do not have enough voltage to reverse bias the C-B junction. In fact, we are someplace near the transition region to forward bias. The base is now more negative than either the emitter or collector.



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

$$E_{\text{photon}} = h\nu,$$

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

Bohr Model

$$= 0.529 \text{Å} n^2$$

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

$$= -13.56 \text{eV} / n^2$$

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

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

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

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.

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

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

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

STATE DENSITY

state distribution function.

$$n(E)dE = f(E)N_c(E)dE = \frac{1}{1 + \exp\left[\frac{(E - E_F)}{kT}\right]} \frac{\sqrt{2}}{\pi^2} \left(\frac{m^*}{\hbar^2}\right)^{3/2} \sqrt{E - E_c} dE \quad \text{Electrons in the Conduction band}$$

$$p(E)dE = (1 - f(E))N_v(E)dE = \left(1 - \frac{1}{1 + \exp\left[\frac{(E - E_F)}{kT}\right]}\right) \frac{\sqrt{2}}{\pi^2} \left(\frac{m^*}{\hbar^2}\right)^{3/2} \sqrt{E_v - E} dE \quad \text{Holes in the Valance band}$$

$$n_0 \approx N_c \exp\left[\frac{-(E_c - E_F)}{kT}\right] = n_i \exp\left[\frac{(E_F - E_i)}{kT}\right]$$

$$p_0 \approx N_v \exp\left[\frac{(E_v - E_F)}{kT}\right] = n_i \exp\left[\frac{-(E_F - E_i)}{kT}\right]$$

total number of electrons and holes

$$n_i^2 = np$$

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

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

$$E_i = \frac{(E_v + E_c)}{2} + \frac{kT}{2} \ln\left(\left(\frac{m_p^*}{m_n^*}\right)^{3/2}\right)$$

Intrinsic Energy

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

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

Mobility

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

Conduction

Diffusion

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

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