Transistor Models

Section 2

Outline
I. Diodes - Reverse Bias
II. - Forward Bias
III. BJT - Large Signal
IV. - Small Signal
V. JFET - Large Signal
VI. - Small Signal

I. Diodes - Reverse Biased

The depletion capacitance for a reverse biased pn diode is found by starting with Poisson’s equation. The coordinate system is chosen so that x = 0 at the metallurgical junction. The depletion region in p side extends to x = - \( w_p \) to the left and in the n side extends to x = \( w_n \) to the right. In the depletion region, Poisson’s equation is

\[
\frac{d^2 \phi}{dx^2} = -\frac{dE}{dx} = \frac{-q}{\epsilon_s} (N_a - N_d)
\]

Consider now the n side where \( N_a = 0 \). The electric field is found by integration.

\[
\int_x^{x_n} dE = \frac{-qN_d}{\epsilon_s} \int_x^{x_n} dx
\]

This results in

\[
E(x_n) - E(x) = \frac{-qN_d}{\epsilon_s} (x - x_n).
\]

At the edge of the space charge region, there exists lots of free charge. Consequently, \( E(x_n) = 0 \). This abrupt change from a region completely free of mobile charge to a region with lots of charge is clearly an approximation to physical reality, but a good approximation that makes life easier. The maximum electric field occurs at x = 0, since

\[
E(x) = \frac{-qN_d}{\epsilon_s} (x_n - x). \quad 0 < x < x_n
\]

Similarly, in the p region

\[
E(x) = \frac{-qN_a}{\epsilon_s} (x + x_p). \quad -x_p < x < 0
\]
At \( x = 0 \), the electric field is the same. This will give the space charge neutrality condition
\[ N_d x_n = N_a x_p. \]

Now the Poisson equation is integrated again to obtain the potential, \( \phi \).

\[
\int_{x}^{x_n} E \, dx = -\frac{q N_d}{\epsilon_s} \int_{x}^{x_n} (x_n - x) d(x - x_n)
\]

\[
-\phi_n + \phi(x) = \frac{q N_d}{2 \epsilon_s} (x_n - x)^2
\]

where at the edge of the depletion region at \( x = x_n, \phi(x_n) = \phi_n \). Since the charge density is related to the potential by an exponential function, \( n = n_i e^{-\frac{\phi}{kT}} \), and since in this case \( \phi = \phi_n \) and \( n = N_d \),

\[
\phi_n = \frac{kt}{q} \ln \frac{N_d}{n_i}.
\]

A similar expression is obtained for the p side.

\[
\phi_p = -\frac{kt}{q} \ln \frac{N_a}{n_i}.
\]

The built-in potential is the difference between \( \phi_p \) and \( \phi_n \).

\[
\phi_n - \phi_p = \phi_i = \frac{kt}{q} \ln \left[ \frac{N_d}{n_i} + \ln \frac{N_a}{n_i} \right] = \frac{kt}{q} \ln \left[ \frac{N_d N_a}{n_i^2} \right]
\]

The next step is to find the depletion width. At \( x = 0 \),

\[
\phi(0) = \phi_n - \frac{q N_d}{2 \epsilon_s} (x_n - 0)^2 = \phi_p + \frac{q N_a}{2 \epsilon_s} (0 + x_p)^2
\]

\[
\phi_i = \frac{q}{2 \epsilon_s} \left( N_a x_p^2 + N_d x_n^2 \right)
\]

\[
\frac{2 \epsilon_s \phi_i}{q} = x_n^2 N_d \left( \frac{N_a x_p^2}{N_d x_n^2} + 1 \right)
\]

Dividing the square of the space charge neutrality expression by \( N_a \),

\[
\frac{(N_a x_p)^2}{N_a} = \frac{(N_d x_n)^2}{N_a}
\]

the previous expression reduces to

\[
\frac{2 \epsilon_s \phi_i}{q} = x_n^2 N_d \left( \frac{N_d}{N_a} + 1 \right).
\]
This can be solved for $x_n$

$$x_n = \sqrt{\frac{2 \varepsilon_s \phi_i}{q N_d^2 (1/N_a + 1/N_d)}}$$

A similar procedure can be carried out to find $x_p$. The sum of these two distances is the depletion width, $W$.

$$W = x_n + x_p = \sqrt{\frac{2 \varepsilon_s \phi_i}{q} \left( \frac{1}{N_a} + \frac{1}{N_d} \right)}$$

The depletion width depends mostly on the more lightly doped region. This derivation applies to abrupt junction diodes where there is an abrupt change from the p doping to the n doping.

Other doping profiles can be modeled. This time assume that the n side of the junction is modeled as $N_d = ax^b$. Then integration twice would give

$$\frac{d\phi}{dx} = \frac{-qa}{\varepsilon_s (b+1)} (x^{b+1} - x_n^{b+1})$$

$$\phi(x) = \frac{-qa}{\varepsilon_s (b+1)} \left[ \frac{x^{b+2}}{b+2} - xx_n^{b+1} \right]$$

where the boundary conditions for the electric field and voltage are $= 0$ at $x = x_n$ and 0 respectively. The expression for depletion region in the n region is

$$x_n = \left[ \frac{(\phi_i - V_{app}) \varepsilon_s (b+2)}{qa} \right]^{1/(b+2)}$$

Here, the $V_{app}$ is added to the expression to indicate the affects of the application of an external voltage. The maximum electric field is located at $x = 0$. This is found by substituting in the value for $x_n$ into the maximum electric field expression.

$$E_{max} = \frac{b+2}{b+1} \left[ \frac{(\phi_i - V_{app})^{b+1} qa}{\varepsilon_s (b+2)} \right]^{1/(b+2)}$$

The depletion capacitance is then found from the ratio of the charge to voltagewhere $Q = AE_{max}$.

$$C = -\frac{dQ}{dV} = \left[ \frac{qa \varepsilon_s^{b+1}}{b+2} \right]^{1/(b+2)} A(\phi_i - V_{app})^{-1/(b+2)}$$

The physical constants can be lumped into one parameter, so that the functional relationship for the depletion capacitance is as follows.

$$C = \frac{C_{j0}}{(1 - \frac{V_{app}}{\phi_i})^\gamma}$$

For the abrupt junction where $b = 0$, $\gamma = 1/2$, for $b = 1$, $\gamma = 1/3$, or for $b = -1$, $\gamma = 1$. 

3
II. Diodes – Forward Biased
Shockley diode equation.

\[ I_d = I_s (e^{qV_d/kT} - 1) = I_s (e^{V_d/V_T} - 1) \approx I_s e^{V_d/V_T} \]

\[ V_d = V_T \ln \frac{I_d}{I_s} \]

III. BJT – Large Signal
The common base current gain consists of three parameters: the emitter efficiency \( \gamma \), the base transport factor, \( \alpha_t \), and the avalanche multiplication \( M \).

The emitter efficiency is

\[ \gamma = \frac{I_{nE} \rightarrow}{I_{nE} \rightarrow +I_{pB} \leftarrow}. \]

The base transport factor is

\[ \alpha_T = \frac{\text{minority carriers leaving base}}{\text{minority carriers entering base}}. \approx 1 - \frac{W_B^2}{2D_T}. \]

Avalanche multiplication is

\[ M = \frac{\text{carriers entering collector}}{\text{minority carriers arriving at the collector – base junction}}. \]

\[ = \left[ 1 - \left( \frac{V_{CB}}{B V_{CB0}} \right) \right]^{-1}. \]

In the expression for \( M \), \( m = 2 \) for npn and \( m = 4 \) for pnp devices.

In making use of the current in a reverse biased junction, there are actually two possible current mechanisms: avalanche multiplication and Zener breakdown. Zener breakdown occurs when both sides of the junction are heavily doped and quantum mechanical tunneling occurs. Typically the Zener breakdown voltages are less than 5 V.

For a \( p^+n \) diode where \( n < 10^{18} \text{ cm}^{-3} \), avalanche multiplication can occur. The breakdown voltage is determined by the impurity concentration of the lightly doped side. \( BV_{e0} \) is determined by the base and \( BV_{CB0} \) is determined by the collector. Since \( BV_{e0} \approx 6 – 8 \text{ Volts} \), it is sometimes used as a voltage reference.

III. BJT – Large Signal
There are four operating regions for a Bipolar Junction Transistor (BJT).

1. Forward active \( V_{BE} > 0 \ \ V_{BC} < 0 \)
2. Saturation \( V_{BE} > 0 \ \ V_{BC} > 0 \)
3. Cutoff \( V_{BE} < 0 \ \ V_{BC} < 0 \)
4. Reverse active \( V_{BE} < 0 \ \ V_{BC} > 0 \)

The Ebers–Moll Equations for npn transistors is

\[ I_C = I_S \left( e^{V_{BE}/V_T} - 1 \right) - \frac{I_S}{\alpha_R} \left( e^{V_{BC}/V_T} - 1 \right) \]

\[ I_E = -\frac{I_S}{\alpha_F} \left( e^{V_{BE}/V_T} - 1 \right) + I_S \left( e^{V_{BC}/V_T} - 1 \right) \]
IV. BJT – Small Signal

The circuit elements for the small signal $\pi$ equivalent circuit can be estimated.

**Transconductance**

$$g_m = \frac{\partial I_C}{\partial V_{BE}} = \frac{I_C}{V_T}$$

**Base charging capacitance**

$$C_b = \frac{\Delta Q_p}{V_i} = \tau_F g_m$$

where $\tau_F$ is the base transit time. $\tau_F = 10$ to $500$ ps (NPN) and 1 to $40$ ns (PNP).

**Input Resistance**

At DC, $\beta_F = I_C/I_B$ and for AC $\beta_0 = i_c/i_b$. The short circuit common emitter current gain, $\beta_F$ does itself vary with collector current and also is temperature sensitive. In doing designs, this should be taken account, but for hand calculations, it is normally considered a constant. The input resistance is

$$r_\pi = \frac{v_i}{i_b} = \frac{V_i}{i_c} \beta_0 = \frac{\beta_0}{g_m}$$

**Output Resistance**

Because the collector–base junction is reverse biased, there is a depletion capacitance there. When bias is changed, the depletion width changes. As the voltage increases, this depletion width approaches the total base width, $W_B$. When this happens, “punch through” occurs. The slope of the $I_C$ vs. $V_{CE}$ curve increases as $V_{BE}$ increases base–emitter depletion width decreases. This slope extrapolates to $V_{CE} = V_A$, the Early voltage. The slope of the $I_C$ vs. $V_{CE}$ curve is $1/r_o$. As a consequence of the finite Early voltage, the Shockley equation is modified.

$$I_C = I_{C0} e^{V_{BE}/V_T} \left( 1 + \frac{V_{CE}}{V_A} \right)$$

In addition the slope is given below.

$$r_o = \frac{\Delta V_{CE}}{I_C} = \frac{V_A}{I_C} = \frac{V_A}{g_m V_T}$$

**Collector – Base Resistance**
\[ r_\mu = \frac{\Delta V_{CE}}{\Delta I_B} = \frac{\Delta V_{CE}}{\Delta I_C} \cdot \frac{\Delta I_C}{\Delta I_B} = r_o \beta_0 \]

Typically, \( r_\mu > 10r_o \beta_0 \).

Capacitances

\( C_{je} \approx 2C_{je0} \) because this junction is typically forward biased. The capacitances, \( C_\mu \) and \( C_{CS} \) use the usual depletion capacitance formula.

Parasitic Resistances

The remaining resistances are given typical numbers and are often determined empirically. For the base, emitter, and collector, these are \( r_b = 50 \text{ to } 500 \, \Omega \), \( r_{ex} = 1 \text{ to } 3 \, \Omega \), and \( r_c = 20 \text{ to } 500 \, \Omega \).

\( \beta \) Frequency Response

If the \( \pi \) equivalent circuit for the common emitter BJT is excited with a current source at the base, \( i_i \), the resulting voltage across the input resistance, \( r_\pi \) is

\[ v_b = \frac{r_\pi}{1 + r_\pi s(C_\mu + C_{\pi})} \cdot i_i \]

If the current through \( C_\mu \) is neglected,

\[ i_c \approx g_m v_b = \frac{g_m r_\pi i_i}{1 + r_\pi s(C_\mu + C_{\pi})} \]

\[ \frac{i_c}{i_i} = \beta(\omega) = \frac{\beta_0}{1 + \beta_0 \left( \frac{C_{\mu} + C_{\pi}}{g_m} \right) j \omega} \]

At high frequencies the "1" in the denominator can be neglected so that a parameter \( \omega_T \) can be defined where \( \beta(\omega) \to 1 \).

\[ \omega_T = \frac{g_m}{C_{\mu} + C_{\pi}} \]

The "3 dB" frequency for \( \beta \) can be expressed as follows.

\[ \beta(\omega) = \frac{\beta_0}{1 + j \omega \beta_0 / \omega_T} = \frac{\beta_0}{1 + j \omega / \omega_\beta} = \frac{\beta_0 e^{-j \omega / \omega_\alpha}}{1 + j \omega / \beta_0} \]

This last expression includes an empirical factor that represents a change of phase at high frequencies.
When considering the pnp transistor, the base is the n-type epitaxial layer, which served as the collector for the npn lateral device. The carrier transport through the pnp base is most efficient near the surface where the emitter to collector distance is minimum. Masking tolerances limit the base width, \( W_B \). This dimension can be made much smaller in the vertical npn. The \( \beta \) for a pnp transistor is in the 5 to 50 range because of low \( \gamma \) which in turn is caused by low emitter p-type doping and small effective emitter area. In addition, some emitted holes go to the substrate rather than to the collector. In addition, the pnp transistor are two additional parasitic pnp transistors.

To fix the low \( \beta \) of the pnp transistor, the pnp can be cascaded with a npn transistor to effectively make one large pnp. The current handling capability of the transconductance for the total effective pnp is \( g_{mT} \) is the same as the pnp device, \( g_{mp} \). From the circuit shown in class,

\[
I_{Cp} = I_{Se}(V_{EB}/V_T) \\
\approx \beta_p I_{Bp} = \beta_p I_B \\
I_C = \beta_n I_{Bn} = \beta_n I_{Cp} = \beta_n \cdot \beta_p I_B \\
\beta_T = \beta_n \beta_p
\]

This cascade of two transistors is in effect one pnp transistor with high \( \beta \), but it can be unstable.

V. JFETs – Large Signal

Junction Field Effect Transistors (JFETs) are contrasted with the BJT.

<table>
<thead>
<tr>
<th>JFET</th>
<th>BJT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current conductance by majority carriers</td>
<td>Current conductance by minority carriers</td>
</tr>
<tr>
<td>Voltage controlled device</td>
<td>Current controlled device</td>
</tr>
<tr>
<td>High input impedance (reverse biased junction)</td>
<td>Low input impedance (forward biased junction)</td>
</tr>
<tr>
<td>n-channel</td>
<td>npn</td>
</tr>
<tr>
<td>p-channel</td>
<td>pnp</td>
</tr>
</tbody>
</table>

The gate contacts to a reverse biased pn junction, thereby producing a depletion region. The current flow through the drain to the source (n-channel) and electron from the source to the drain. The current produces a voltage gradient in the channel that makes the drain side more heavily reverse biased than the source side. At some \( V_{DS} \), the channel is pinched off. The pinch off voltage is \( V_p \). The drain current in the pinch off region (also known as the saturation region) is

\[
I_{DS} = I_{DSS} \left( 1 - \frac{V_{GS}}{V_p} \right)^2 .
\]
The non zero slope of $I_{DS}$ vs. $V_{DS}$ curve is incorporated into the IV characteristic in similar fashion as the BJT using the Early voltage, $V_A$.

\[ I_{DS} = I_{DSS} \left( 1 - \frac{V_{GS}}{V_p} \right)^2 \left( 1 - \frac{V_{DS}}{V_A} \right) \]

The current in the JFET is proportional to $V^2$ while in the BJT it is proportional to the exponential function $e^{V/V_T}$. If such a comparison can be made, the BJT is “more nonlinear” than the JFET.

**VI. JFETs – Small Signal**

Normally, operation is near the pinch off voltage. The transconductance is

\[
g_m = \frac{\partial I_{DS}}{\partial V_{GS}} = -2 \frac{I_{DSS}}{V_p} \left( 1 - \frac{V_{GS}}{V_p} \right)
\]

\[
g_m = g_{mo} \left( 1 - \frac{V_{GS}}{V_p} \right)
\]

For a n-channel device, $I_{DSS} > 0$ and $V_p < 0$. For a p-channel device, $I_{DSS} < 0$ and $V_p > 0$. In either case, $g_{mo} > 0$. The output resistance is

\[
\frac{1}{r_o} = \frac{\partial I_D}{\partial V_{DS}} = \frac{I_{DSS}}{V_A} \left( 1 - \frac{V_{GS}}{V_A} \right)^2
\]

\[
r_o \approx \frac{V_A}{I_D} = \frac{1}{\lambda I_D}
\]

The $C_{GS}$ varies as the square root of the voltage and the $C_{GD}$ varies as the cube root of the voltage.

**VII. MOS – Large Signal**

A n-channel MOSFET (Metal Oxide Semiconductor) uses p-type material in the channel and a p-channel device uses n-type material. An inversion layer is formed below the surface that conducts the current.

The Fermi voltage level lies between the valence and conduction bands. The Fermi-Dirac distribution function describes the probability of an electron with Fermi energy to be 0.5. From semiconductor physics, the Fermi level is

\[
\phi_f = \frac{kT}{q} \ln \left( \frac{N_A}{n_i} \right) \approx 0.3 \ V
\]

After the potential reaches $2\phi_f$, increases in the gate voltage produces no further change in the depletion layer.
For depletion mode devices, a channel exists at $V_{GS} = 0$. For enhancement mode devices, a channel does not exist at $V_{GS} = 0$. For the enhancement mode, increasing $V_{GS}$ above the threshold voltage, $V_t$, creates a local inversion layer under the gate. The voltage creates hole-electron pairs. The holes are swept away into the bulk and the electrons move toward the oxide layer. The applied voltage has created an inversion layer where there are more electrons than holes in the p material.

Enhancement mode transistors are preferred because:
1. Self isolation – All active regions are reverse biased. This allows for high packing density.
2. They do not require tight control on the diffusion.
3. They can be fabricated with a single diffusion step.

The gate capacitance per unit area is

$$C_{ox} = \frac{\epsilon_{ox}}{T_{ox}}.$$ 

The channel conductance is given by

$$g_c(x) = \mu_s C_{ox} (V_{GS} - V_t - V(x))$$

The gate voltage required to produce an inversion layer is the threshold voltage, $V_t$, $V(x)$ is the channel potential at distance x from the source, and $\mu_s$ is the surface mobility. There are three regions of operation: the triode or Ohmic region, the saturation region, and the cutoff region.

**VIII. MOS – Small Signal**

As before the transconductance is the ratio of the output current to the input voltage. For MOS devices,

$$g_m = \frac{2\mu_s C_{ox} W}{L} I_D$$

when $|V_{DS}/V_A| \ll 1$. When the substrate (body) is not grounded, an additional terminal is present.

$$g_{mb} = \frac{\partial I_D}{\partial V_{SB}} \approx \frac{W \gamma \sqrt{\mu_s C_{ox} I_D}}{\sqrt{2(2\phi_f + V_{SB})}}$$