6.772 - Compound Semiconductor and Heterostructure Devices
Lecture 10 - Heterojunction Bipolar Transistors I

- i-v model for HMETs and HIGFETs
  DC model - without velocity saturation
  - with velocity saturation
  Small signal linear equivalent circuit
  High frequency limits - $\omega_T$, $\omega_{\text{max}}$

- Bipolar junction transistors (BJTs)
  Review of homojunction BJTs: i-v characteristics
  small signal linear equiv. ckts.
  high-f performance

- Heterojunction bipolar transistors (HBTs)
  Historical perspective: Shockley and Kroemer, motivation
  A methodical look at heterojunction impacts:
  - emitter issues
  - base issues
  - collector issues
  Summary - putting this together in a single device
HFET i-v: characteristics for HEMTs and HIGFETs

In the HEMT and HIGFET the channel carriers are localized in a two-dimensional electron gas at the interface between the wide bandgap gate "dielectric" and the narrow bandgap layer beneath it:

\[ qN_{ch}(y) = -\frac{\varepsilon_{WBG}}{t_{WBG}} \left[ v_{GS} - v_{CS}(y) - V_T \right] \]
**HFET i-v: for HEMT and HIGFET, cont.**

As before, we write the drain current at y as:

\[ i_D = -\left[ -qN_{ch}(y) \cdot W \cdot s_y(y) \right] \]

We just found \( N_{ch}(y) \), and we can write the average net carrier velocity in the low- to moderate field region as:

\[ s_y(y) = -\mu_e F_y(y) = \mu_e \frac{dv_{CS}(y)}{dy} \]

Putting this together yields:

\[ i_D = W \frac{\varepsilon_{WBG}}{t_{WBG}} \left[ v_{GS} - v_{CS}(y) - V_T \right] \mu_e \frac{dv_{CS}(y)}{dy} \]

Integrating with respect to y from 0 to L, or equivalently, with respect to \( v_{CS} \) from 0 to \( v_{DS} \), we have the result given on the next foil....
HFET i-v: for HEMT and HIGFET, cont.

The result of the integration is:

\[ i_D = \frac{W}{L} \mu_e \frac{\varepsilon_{WBG}}{t_{WBG}} \left[ (v_{GS} - V_T)v_{DS} - \frac{v_{DS}^2}{2} \right] \]

This expression is valid as long as \( v_{DS} < (v_{GS} - V_T) \). When \( v_{DS} > (v_{GS} - V_T) \), \( i_D \) saturates at:

\[ i_D = \frac{1}{2} \frac{W}{L} \mu_e \frac{\varepsilon_{WBG}}{t_{WBG}} (v_{GS} - V_T)^2 \]

These expressions should look familiar from MOSFETs. They are identical to the commonly used MOSFET expression.

Finally, the small signal transconductance, \( g_m \), in saturation is:

\[ g_m \equiv \frac{\partial i_D}{\partial v_{GS}} \bigg|_Q = \frac{W}{L} \mu_e \frac{\varepsilon_{WBG}}{t_{WBG}} (V_{GS} - V_T) \]
HFET i-v: for HEMT and HIGFET, cont.

One last thing to look at for these devices is their high frequency performance. A good way to do this is to look at $\omega_T$, which we recall can be written:

$$\omega_T = \frac{g_m}{C_{gs}}$$

We have not found $C_{gs}$ yet, but we in fact already know it because $C_{gs}$ for these devices will be the same as it is for a MOSFET:

$$C_{gs} \approx \frac{2}{3} W L \frac{\varepsilon_{WBG}}{t_{WBG}}$$

Using this, we find:

$$\omega_T = \frac{3}{2} \frac{\mu_e (V_{GS} - V_T)}{L^2}$$

without velocity saturation
The common-source short-circuit current gain is:

\[ \beta_{sc}(j\omega) = \frac{i_d(j\omega)}{i_g(j\omega)} = \frac{g_m - j\omega C_{gd}}{j\omega(C_{gs} + C_{gd})} \]

there is one pole at \( \omega = 0 \), and one zero, \( \omega_z \):

\[ \omega_z = \frac{g_m}{C_{gd}} \]

The short circuit current gain, \( \beta_{sc} \), is infinite at DC (\( \omega = 0 \)), and its magnitude decreases linearly with increasing frequency.
The magnitude of $\beta_{sc}$ decreases with $\omega$, but it is still greater than one for a wide range of frequencies.

$$|\beta_{sc}(j\omega)| = \sqrt{\frac{g_m^2 + \omega^2 C_{gd}^2}{\omega^2 (C_{gs} + C_{gd})^2}}$$

The transistor is useful until $|\beta_{sc}|$ is less than one. The frequency at which this occurs is called $\omega_t$. Setting $= 1$ and solving for $\omega_t$ yields:

$$\omega_t = \sqrt{\frac{g_m^2}{\left(C_{gs} + C_{gd}\right)^2 - C_{gd}^2}} \approx \frac{g_m}{C_{gs} + C_{gd}}$$
FET short-circuit current gain, $\beta_{sc}(j\omega)$, cont.

Unity gain point, $\omega_t$: $\omega_t = g_m/(C_{gs} + C_{gd})$

Zero, $\omega_z$: $\omega_z = g_m/C_{gd}$

Note: $\omega_z > \omega_t$

Low frequency value: infinity

No 3dB point, $\omega_\beta$.  

$\log |\beta_{sc}|$ vs $\log \omega$ graph.
FET short-circuit current gain, $\beta_{sc}(j\omega)$, cont.

Can we bias to maximize $\omega_t$?

$$\omega_t (\text{FET}) = \frac{g_m}{(C_{gs} + C_{gd})} \approx \frac{g_m}{C_{gs}}$$

$$= \frac{W}{L} \mu_{Ch} C_{ox}^* |V_{GS} - V_T|$$

$$= \frac{2}{3} \frac{W L C_{ox}^*}{3}$$

Maximize $V_{GS}$.

What is the ultimate limit?

$$\omega_t (\text{FET}) = \frac{3}{2} \frac{\mu_{Ch} |V_{GS} - V_T|}{L^2} = \frac{3}{2L} \mu_{Ch} \frac{|V_{DS}|}{L} = \frac{3}{2L} \mu_{Ch} \frac{E_{Ch}}{L} = \frac{3}{2} \frac{s_{Ch}}{L} = \frac{1}{\tau_{Ch}}$$

Lessons: Bias at large $I_D$; make $L$ small, use n-channel.

Channel transit time!
These expressions should look familiar from MOSFETs. They are identical to the commonly used MOSFET expression.

When we have severe velocity saturation, we have \( s(y) = s_{sat} \), and we have simply

\[
i_D = W \frac{\varepsilon_{WBG}}{t_{WBG}} \left[ V_{GS} - V_T \right] s_{sat}
\]

Calculating the small signal transconductance now, we have:

\[
g_m \equiv \left. \frac{\partial i_D}{\partial V_{GS}} \right|_Q = W s_{sat} \frac{\varepsilon_{WBG}}{t_{WBG}}
\]

These expressions should look familiar from MOSFETs. They are identical to the commonly used MOSFET expression.

Finally, the small signal gate-to-source capacitance, \( C_{gs} \), in saturation is simply the gate capacitance, \( WL\varepsilon_{WBG}/t_{WBG} \) and thus:

\[
\omega_T \approx \frac{g_m}{C_{gs}} = \frac{s_{sat}}{L}
\]

with velocity saturation.
High frequency performance metrics: \( \omega_T, \omega_{\text{max}} \)

We introduced \( \omega_T \), which we defined as the frequency at which the magnitude of the short circuit current gain was unity, and showed that:

\[
\omega_T \approx g_m/C_{gs}
\]

It turns out that \( \omega_T \) is a good metric for high speed switching, but for microwave applications, a better metric is \( \omega_{\text{max}} \), the unity power gain with matched source and load impedances. We call this frequency \( \omega_{\text{max}} \), and find that it is roughly given by:

\[
\omega_{\text{max}} \approx \sqrt{\frac{\omega_T}{4R_gC_{gd}}} \approx \sqrt{\frac{g_m}{4R_gC_{gd}C_{gs}}}
\]

This later metric shows the importance in microwave amplifier applications of FETs of reducing the gate resistance, \( R_g \), and the gate-to-drain capacitance, \( C_{gd} \).

Note: We'll introduce a third high frequency performance metric, \( \omega_v \), when we talk about HBTs.
BJT/FET Comparison

Controlled by $v_{EB}$ in a BJT; by $v_{GS}$ in an FET

Potential energy of majority carriers in emitter

Emitter/Source

Base/Channel

Collector/Drain

$x$

0 $w_B$ or $L$

Potential energy of majority carriers in emitter

$x$

0 $w_B$ or $L$

$v_{CE}$ or $v_{DS}$
**BJT/FET Comparison - cont.**

OK, that's nice, but there is more to the difference than how the barrier is controlled.

<table>
<thead>
<tr>
<th></th>
<th><strong>BJT</strong></th>
<th></th>
<th><strong>FET</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Charge carriers</strong></td>
<td>minority in base</td>
<td>majority in channel</td>
<td></td>
</tr>
<tr>
<td><strong>Flow mechanism</strong></td>
<td>diffusion in base</td>
<td>drift in channel</td>
<td></td>
</tr>
<tr>
<td><strong>Barrier control</strong></td>
<td>direct contact made to base</td>
<td>change induced by gate electrode</td>
<td></td>
</tr>
</tbody>
</table>

The nature of the current flow, minority diffusion vs majority drift, is perhaps the most important difference.
Basic Bipolar Junction Transistor (BJT) - cross-section

An npn BJT
Adapted from Fig. 8.1 in Text

The heart of the device, and what we are modeling
Bipolar Junction Transistors: basic operation and modeling...
... how the base-emitter voltage, $v_{BE}$, controls the collector current, $i_C$
Assume a well designed npn BJT in FAR:

\[ N_{DE} >> N_{AB}, \ w_E << L_{hE}, \ w_B << L_{eB} \]

**Excess Carriers:**

\[
\begin{align*}
(n_i^2/N_{DE})(e^{qV_{BE}/kT} - 1) \\
0 \ (\text{ohmic}) \\
0 \ (V_{BC} = 0) \\
0 \ (\text{ohmic})
\end{align*}
\]

**Currents:**

\[
\begin{align*}
i_E &= i_E(1 + \delta_E) \\
i_hE &= \delta_E i_E \\
i_C &= i_E(1 - \delta_B) \approx i_E \\
i_B &= i_E - (-i_C) \\
&= i_E(\delta_E + \delta_B) \approx i_E\delta_E
\end{align*}
\]

**Our task is to determine:**

Given a structure, what are \( i_E(v_{BE}, v_{CE}), i_C(v_{BE}, v_{CE}), \) and \( i_B(v_{BE}, v_{CE})? \)
npn BJT, cont.: F.A.R. model

With $v_{BE} > 0$, $v_{BC} \leq 0$ (and neglecting $I_{CS}$), we have:

$$i_E = -I_{ES}[e^{qV_{BE}/kT} - 1]$$

$$i_C = -\alpha_F i_E = \beta_F i_B$$

$$i_B = i_C/\beta_F$$

where the defects are:

$$\delta_E \equiv \frac{i_{hE}}{i_{eE}} = \frac{D_h \cdot N_{AB}}{D_e \cdot N_{DE}} \cdot \frac{w_{B,eff}}{w_{E,eff}}$$

and

$$\delta_B \approx \frac{w_{B,eff}^2}{2D_e \tau_e} = \frac{w_{B,eff}^2}{2L_e^2} \ll \delta_E$$

The circuit model that fits this behavior is the following:

![Circuit diagram](image)

Note: $i_F = -i_E$.

With:

$$I_{ES} = A q n_i^2 \left( \frac{D_h}{N_{DE} w_{E,eff}} + \frac{D_e}{N_{AB} w_{B,eff}} \right)$$

and with $\alpha_F$ and $\beta_F$ as defined above.

Remember that $\alpha_F$ and $\beta_F$ are related:

$$\alpha_F = \frac{\beta_F}{(\beta_F + 1)}, \quad \beta_F = \frac{\alpha_F}{(1 - \alpha_F)}$$

so the model is totally specified by knowing $I_{ES}$ and either $\alpha_F$ or $\beta_F$. 

In modern Si BJTs.
**BJT's:** Looking at the characteristics over the whole voltage range

**The Ebers-Moll model**

Our F.A.R. results are a special case of an important general model that describes the BJT in all of its operating regions. It is the "Ebers-Moll" model.

We can readily solve the general flow problem and get a general expression for the BJT characteristics, following the process we used to look at the F.A.R., but in the Ebers-Moll model we make use of superposition and solve the problem with each "excitation" applied separately, and then combine the results:

\[
i_E(v_{BE}, v_{BC}) = i_E(v_{BE}, 0) + i_E(0, v_{BC})
\]

\[
i_C(v_{BE}, v_{BC}) = i_C(v_{BE}, 0) + i_C(0, v_{BC})
\]

Flow problems are linear so we can use superposition like this, but we do have to be a bit careful because the boundary conditions at the junctions are non-linear functions of \(v_{BE}\) and \(v_{CE}\). The solution is to put a non-zero bias on only one junction at a time.
**BJT's**: The Ebers-Moll model, cont.

Using superposition, we first find the currents when $v_{BC} = 0$ and $v_{BE}$ is arbitrary, and repeat the process with $v_{BE} = 0$ and $v_{CE}$ arbitrary.

We call the solution for $v_{BC} = 0$ the "forward" solution:

$$i_E(v_{BE},0) = -I_{ES}(e^{\frac{qv_{BE}}{kT}} - 1) \quad \text{and} \quad i_C(v_{BE},0) = -\alpha_F i_E(v_{BE},0)$$

with

$$I_{ES} = Aq n_i^2 \left( \frac{D_h}{N_{DE}w_{E,\text{eff}}} + \frac{D_e}{N_{AB}w_{B,\text{eff}}} \right), \quad \alpha_F = \frac{(1 - \delta_B)}{(1 + \delta_E)}, \quad \delta_E = \frac{D_h}{D_e} \cdot \frac{N_{AB}}{N_{DE}} \cdot \frac{w_{B,\text{eff}}}{w_{E,\text{eff}}}, \quad \delta_B \approx \frac{w_{B,\text{eff}}^2}{2L_e^2}$$

And the solution for $v_{BE} = 0$ is the "reverse" solution:

$$i_C(0,v_{BC}) = -I_{CS}(e^{\frac{qv_{BC}}{kT}} - 1) \quad \text{and} \quad i_E(0,v_{BC}) = -\alpha_R i_C(0,v_{BC})$$

with

$$I_{CS} = Aq n_i^2 \left( \frac{D_h}{N_{DC}w_{C,\text{eff}}} + \frac{D_e}{N_{AB}w_{B,\text{eff}}} \right), \quad \alpha_R = \frac{(1 - \delta_B)}{(1 + \delta_C)}, \quad \delta_C = \frac{D_h}{D_e} \cdot \frac{N_{AB}}{N_{DC}} \cdot \frac{w_{B,\text{eff}}}{w_{C,\text{eff}}}, \quad \delta_B \approx \frac{w_{B,\text{eff}}^2}{2L_e^2}$$

Superimposing these results gives us the full Ebers-Moll model:

$$i_E(v_{BE},v_{BC}) = -I_{ES}(e^{\frac{qv_{BE}}{kT}} - 1) + \alpha_R I_{CS}(e^{\frac{qv_{BC}}{kT}} - 1)$$

$$i_C(v_{BE},v_{BC}) = \alpha_F I_{ES}(e^{\frac{qv_{BE}}{kT}} - 1) - I_{CS}(e^{\frac{qv_{BC}}{kT}} - 1)$$

The best way to remember the Ebers-Moll model is as a circuit; this is shown on the next foil.
BJT's: The Ebers-Moll model, cont.

Schematically, the forward and backward portions are shown below:

**Forward:**

$$I_{ES} = Aqni^2\left(\frac{D_h}{N_{DE}w_{E,\text{eff}}} + \frac{D_e}{N_{AB}w_{B,\text{eff}}}\right)$$

$$\beta_F = \frac{(1-\delta_B)}{(\delta_E + \delta_B)}$$

$$\delta_E = \frac{i_{BE}}{i_{EE}} = \frac{D_h}{D_e} \cdot \frac{N_{AB}}{N_{DE}} \cdot \frac{w_{B,\text{eff}}}{w_{E,\text{eff}}}$$

**Reverse:**

$$I_{CS} = Aqni^2\left(\frac{D_h}{N_{DC}w_{C,\text{eff}}} + \frac{D_e}{N_{AB}w_{B,\text{eff}}}\right)$$

$$\beta_R = \frac{(1-\delta_B)}{(\delta_C + \delta_B)}$$

$$\delta_C = \frac{i_{BC}}{i_{EC}} = \frac{D_h}{D_e} \cdot \frac{N_{AB}}{N_{DC}} \cdot \frac{w_{B,\text{eff}}}{w_{C,\text{eff}}}$$

**Combined they form the full Ebers-Moll model:**

Note: $i_F = -i_E(v_{BE}, 0)$ and $i_R = -i_C(0, v_{BC})$. 

Fonstad/Palacios 3/10/09

Lecture 10 - Slide 20
**BJT's: The Gummel-Poon model.**

Another common model can be obtained from the Ebers-Moll model is the Gummel-Poon model:

**Forward:**

\[ I_S = \frac{\beta_F}{(\beta_F + 1)} I_{ES} = \frac{\beta_R}{(\beta_R + 1)} I_{CS} \]

\[ = \alpha_F I_{ES} = \alpha_R I_{CS} \]

**Combined they form the Gummel-Poon model:**

- Aside from the historical interest, another value this has for us in 6.012 is that it is an interesting exercise to show that the two forward circuits above are equivalent.
**BJT's:** Looking at the characteristics over the whole voltage range

**Regions of operation:**

The Ebers-Moll and the Gummel-Poon large signal BJT models cover positive and negative biases on both junctions. They thus cover more than just the forward active region that we care about most.

There is a reverse active region, making a total of four regions of operation.

Bipolar junction transistors can be operated in reverse, i.e. with the base-emitter junction reverse biased and the base-collector junction forward biased, but it is usually not optimized for this mode of operation and typically $\beta_R \ll \beta_F$. 
BJT's: FAR characteristics

\[ i_B \approx I_{BS} e^{qV_{BE}/kT} \]
\[ V_{CE} > 0.2 \text{ V} \]

- **Forward Active Region (FAR):**
  - \( V_{BE} > 0.6 \text{ V} \)
  - \( V_{CE} > 0.2 \text{ V} \) (i.e. \( V_{BC} < 0.4 \text{ V} \))
  - \( i_R \) is negligible

- **Saturation:**
  - \( V_{CE} < 0.2 \text{ V} \)

**E-M in FAR**

- **Forward active region:**
  - \( V_{BE} > 0.6 \text{ V} \)
  - \( V_{CE} > 0.2 \text{ V} \)
  - \( i_R \) is negligible

- **Other regions (Cutoff):**
  - \( V_{BE} < 0.6 \text{ V} \)

- **Saturation:**
  - \( V_{CE} < 0.2 \text{ V} \)
**npn BJT:** Connecting with the n-channel HFET

A very similar behavior, and very similar uses.
nnp BJT Output IV Plot

\[ I_B = 5 \mu A/\text{Step} \]
BJT's, review: Limitations of the large signal model

Limitations of our junction model - their impact on BJT characteristics

- **Base width modulation, the Early effect and Early voltage:**
  The width of the depletion region at the B-C junction increases as $v_{CE}$ increases and the effective base width, $w_{B,\text{eff}}$, gets smaller, thereby increasing $\beta$ and, in turn, $i_C$.

- **Punch through:** *base width modulation taken to the limit*
  When the depletion region at the B-C junction extends all through the base all the way to the collector. Punch through has a similar effect on the characteristics as does B-C junction reverse breakdown.
BJT's, review: Limitations of the large signal model
p-n junction non-idealities - the impact on BJT characteristics

Non-idealities of p-n junctions directly impact BJT performance:

• Large forward bias: High level injection (c) Series voltage drop (d)
• Large reverse bias: Reverse breakdown
• Very low bias levels: SCL generation and recombination (a, e)

In a BJT these lead to $\beta_F$ decrease at high and at low current levels

Reverse breakdown; in a BJT this limits $|V_{CE}|$

Ref: Figure 18 in S. M. Sze, "Physics of Semiconductor Devices" 1st. Ed (Wiley, 1969)
BJT's, review: p-n junction non-idealities
- the impact on BJT characteristics

• Beta roll-off at high and low collector currents

• B-C junction breakdown; base punch through

B-E junction recombination

High Level Injection Series resistance

B-C junction breakdown
Base punch through
nnp BJT $\beta_F$ vs $I_C$ Plot

$\beta_F \approx 160$
npn BJT Gummel Plot
Linear equivalent circuit for **BJTs in FAR** (low f):

In the forward active region, our static model says:

\[ i_B(v_{BE}, v_{CE}) = I_{BS} \left[ e^{qv_{BE}/kT} - 1 \right] \]
\[ i_C(v_{BE}, v_{CE}) = \beta_o \left[ 1 + \lambda v_{CE} \right] i_B(v_{BE}, v_{CE}) = \beta_o I_{BS} \left[ e^{qv_{BE}/kT} - 1 \right] \left[ 1 + \lambda v_{CE} \right] \]

We begin by linearizing \( i_C \) about \( Q \):

\[ i_C(v_{be}, v_{ce}) = \left. \frac{\partial i_C}{\partial v_{BE}} \right|_Q v_{be} + \left. \frac{\partial i_C}{\partial v_{CE}} \right|_Q v_{ce} = g_m v_{be} + g_o v_{ce} \]

We introduced the transconductance, \( g_m \), and the output conductance, \( g_o \), defined as:

\[ g_m = \left. \frac{\partial i_C}{\partial v_{BE}} \right|_Q \]
\[ g_o = \left. \frac{\partial i_C}{\partial v_{CE}} \right|_Q \]

Evaluating these partial derivatives using our expression for \( i_C \), we find:

\[ g_m = \frac{q}{kT} \beta_o I_{BS} e^{qV_{BE}/kT} \left[ 1 + \lambda V_{CE} \right] \approx \frac{q I_C}{kT} \]
\[ g_o = \beta_o I_{BS} \left[ e^{qV_{BE}/kT} + 1 \right] \lambda \approx \lambda I_C \left( \text{or} \quad \frac{I_C}{V_A} \right) \]
LEC for BJT (low f), cont.:

Turning next to $i_B$, we note it only depends on $v_{BE}$ so we have:

$$i_b(v_{be}) = \frac{\partial i_B}{\partial v_{BE}}|_Q v_{be} = g_{\pi} v_{be}$$

The input conductance, $g_{\pi}$, is defined as:

$$g_{\pi} \equiv \frac{\partial i_B}{\partial v_{BE}}|_Q$$

However, to evaluate $g_{\pi}$ we use $i_B = i_C / \beta_o$, not our equation for $i_B(v_{BE})$:

$$g_{\pi} \equiv \frac{1}{\beta_o} \frac{\partial i_C}{\partial v_{BE}}|_Q = \frac{g_m}{\beta_o} = \frac{q I_C}{kT \beta_o}$$

(Notice that we do not define $g_{\pi}$ as $qI_B/kT$)

Representing this as a circuit we have:

![Circuit Diagram]

(Notice that $v_{be}$ is also called $v_{\pi}$)
LEC for **BJTs** (high f):

To extend the model to high frequency we linearize the charge stores associated with the junctions and add them.

The **base-collector junction** is reverse biased so the charge associated with it, $q_{BC}$, is the depletion region charge. The corresponding capacitance is labeled $C_{\mu}$.

$$q_{BC}(v_{BC}) \approx -A \sqrt{2q\varepsilon_{Si} \left[\phi_{b,BC} - v_{BC}\right] N_{DC}}$$

$$C_{\mu}(V_{BC}) = \left. \frac{\partial q_{BC}}{\partial v_{BC}} \right|_Q = A \sqrt{\frac{q\varepsilon_{Si} N_{DC}}{2\left[\phi_b - V_{BC}\right]}}$$

The **base-emitter junction** is forward biased and its dominant charge store is the excess charge injected into the base; the base-emitter depletion charge store less important.

$$q_{BE}(v_{BE}) \approx Aq_n^2 \frac{D_e}{N_{AB} w_{B,eff}} \left[e^{qV_{BE}/kT} - 1\right] \approx \frac{w_{B,eff}^2}{2D_e} i_C(v_{BE})$$

The linear equivalent B-E capacitance is labeled $C_{\tau}$. 
Intrinsic $\omega_{Hi}$'s for BJTs - short-circuit current gain

The common-emitter short-circuit current gain is:

$$\beta_{sc}(j\omega) \equiv \frac{i_c(j\omega)}{i_b(j\omega)} = \frac{g_m - j\omega C_{\mu}}{g_{\pi} + j\omega(C_{\pi} + C_{\mu})}$$

there is one pole, call it $\omega_p$, and one zero, $\omega_z$:

$$\omega_p = \frac{g_{\pi}}{(C_{\pi} + C_{\mu})}, \quad \omega_z = \frac{g_m}{C_{\mu}}$$

Of these two, $\omega_p$ is much smaller and this is the 3dB point of the common-emitter short-circuit current gain. We give it the name $\omega_\beta$:

$$\omega_\beta = \frac{g_{\pi}}{(C_{\pi} + C_{\mu})}$$
Intrinsic $\omega_{HI}$'s for BJTs - short-circuit current gain, cont.

The magnitude of $\beta_{sc}$ decreases above $\omega_b$, but it is still greater than one initially:

$$|\beta_{sc}(j\omega)| = \sqrt{\frac{g_m^2 + \omega^2 C^2_\mu}{g^2_\pi + \omega^2 (C_\pi + C_\mu)^2}}$$

The transistor is useful until $|\beta_{sc}|$ is less than one. The frequency at which this occurs is called $\omega_t$. Setting $= 1$ and solving for $\omega_t$ yields:

$$\omega_t = \sqrt{\frac{(g^2_\pi + g^2_m)}{(C_\pi + C_\mu)^2 - C^2_\mu}} \approx \frac{g_m}{(C_\pi + C_\mu)}$$
BJT short-circuit current gain, $\beta_{sc}(j\omega)$, cont.

Note: $\omega_z > \omega_t >\!
\!
\!
\!\!\!,\omega_\beta (= \omega_t / \beta_F)$

Low frequency value: $\beta_F$

Zero, $\omega_z$: $\omega_z = \frac{g_m}{C_\mu}$

3dB point, $\omega_\beta$: $\omega_\beta = \frac{g_m}{(C_\pi + C_\mu)}$

Unity gain point, $\omega_t$: $\omega_t @ \frac{g_m}{(C_\pi + C_\mu)}$
BJT short-circuit current gain, $\beta_{sc}(j\omega)$, cont.

Can we bias to maximize $\omega_t$?

$$\omega_t \approx \frac{g_m}{C_\pi + C_\mu} = \frac{\frac{qI_C}{kT}}{\left(\frac{qI_C}{kT}\right)\tau_b + C_{eb,dp} + C_{cb,dp}}$$

Maximize $I_C$.

Used $C_\pi = g_m \tau_b + C_{eb,dp}$

In the limit of large $I_C$:

$$\lim_{I_C \to \infty} \omega_t \approx \frac{1}{\tau_b} = \frac{2D_{\text{min,B}}}{W_B^2} = \frac{2\mu_{\text{min,B}} V_{\text{thermal}}}{W_B^2}$$

Lessons: Bias at large $I_C$; make $W_B$ small, use npn.
High frequency performance metrics: $\omega_T$, $\omega_{\text{max}}$, $\omega_v$

We have $\omega_T$ for a BJT, and it is interesting to write it in terms of the bias point and to then take the limit as $I_C$ becomes very large:

$$\omega_T = \frac{g_m}{(C_\pi + C_\mu)} = \frac{qI_C/kT}{C_{eb,depl} + (qI_C/kT)\tau_{tr} + C_{cb,depl}}$$

Thus

$$\lim_{I_C \to \infty} \omega_T = 1/\tau_{tr}$$

We found a similar result for FETs, that is that $\omega_T$ is approximately the inverse of the transit time through the active part of the device. In our simple model for the BJT the transit time that appears here is the transit time through the base, but in fact we find that it should in fact be the transit time from the emitter to the collector, which includes the time to transit the emitter-base and base-collector depletion regions:

$$\tau_{tr} = \tau_{tr,B} + \tau_{tr,EB} + \tau_{tr,CB} \approx \frac{W_b^* 2}{2D_{eB}} + \tau_{tr,EB} + \tau_{tr,CB}$$
High frequency performance metrics: $\omega_T$, $\omega_{\text{max}}$, $\omega_v$

As we said with FETS, $\omega_T$ is a good metric for high speed switching, but for microwave applications, a better metric is $\omega_{\text{max}}$, the unity power gain with matched source and load impedances. For a BJT, we find that $\omega_{\text{max}}$ is roughly given by:

$$\omega_{\text{max}} \approx \sqrt{\frac{\omega_T}{4 R_b C_\mu}} \approx \sqrt{\frac{g_m}{4 R_b C_\mu C_\pi}}$$

This metric shows the importance in microwave amplifier applications of HBTs of reducing the base series resistance, $R_b$, and the base-to-collector capacitance, $C_\mu$.

Recently an additional high frequency performance metric has been suggested, the 3dB breakpoint frequency of the effective transconductance, which is called $\omega_v$:

$$\omega_v = \frac{\omega_T}{g_m R_b} = \frac{1}{R_b C_\pi} \quad \left(\text{Note: in an FET, } \omega_v = \frac{1}{R_g C_{gs}}\right)$$

Using Heterojunctions to improve BJTs: general observations on heterojunction BJTs (HBTs)

**Historical Note:**

The first ideas for using heterojunctions in BJTs was directed at increasing beta by making the emitter defect smaller. This possibility was implied by Shockley in his original transistor patent (c. 1950), and it was proposed explicitly by Herb Kroemer in a 1958 paper.

Kroemer again proposed the HBT in another paper about 1980, when at last it was technologically possible to explore this idea, and that is when heterojunction bipolar transistors, HBTs, took off.

**Heterojunction Impacts:**

To investigate the full potential of heterojunctions in BJTs we will look at the issue somewhat more broadly.

There are three main pieces of a bipolar transistor, and we will look at issues involved with each one:

- Emitter issues
- Base issues
- Collector issues
**Emitter issues:**

In a homojunction transistor the lateral conductivity of the base is limited by the doping level and base width. It could be increased if the base doping level could be increased, but this impacts the emitter defect negatively.

With a wide bandgap emitter layer, the coupling another factor is introduced in the emitter defect which reduces the need to dope the base heavily.

**In a homojunction:**

\[ \delta_E \equiv \left( \frac{D_h}{D_e} \cdot \frac{w_B^*}{w_E^*} \cdot \frac{N_{AB}}{N_{DE}} \right) \]

**In a heterojunction:**

\[ \delta_E \equiv \left( \frac{D_h}{D_e} \cdot \frac{w_B^*}{w_E^*} \cdot \frac{N_{AB}}{N_{DE}} \right) e^{-\frac{(HB-EB)}{kT}} \]

**New factor:**  
HB = hole barrier  
EB = electron barrier

Note: We saw this when we talked about heterojunctions in Lecture 3.
Emitter issues: Hole barrier vs. electron barrier

The size of the barrier depends on whether or not there is an effective spike in one of the bands (usually in the conduction band)

NOTE: In an N-p+ junction, all of the band bending will be on the N-side.

If the spike is a barrier: $\text{HB} - \text{EB} \approx \Delta E_v$
Emitter issues: Hole barrier vs. electron barrier

If the spike can be made thin enough to tunnel, or (better) is eliminated by grading the composition at the heterojunction, then HB - EB can be increased significantly:

\[
EB \approx q \Delta \phi - \Delta E_c
\]

\[
HB \approx q \Delta \phi + \Delta E_v
\]

If the spike is not a barrier: HB - EB \approx \Delta E_c + \Delta E_v = \Delta E_g

NOTE: Grading removes the spike without requiring heavy N-doping.
Emitter issues: Dopant diffusion from base

In a heterojunction transistor the base is usually very heavily doped, while the emitter is more lightly doped. In such a case it is easy for dopant to diffuse into the emitter and dope it P-type. This is very bad:

Now we are back to a homojunction!

A solution is to introduce a thin, undoped narrow bandgap spacer.
Emitter issues: Using the barrier to advantage
Ballistic injection

The concept of ballistic injection is that the carriers going over the spike will enter the base with a high initial velocity:

This can reduce the base transit time.
A problem is that if the carriers are given too much initial energy they will scatter into a higher energy band and slow down.
Base issues: Reducing the base resistance
Reducing the base transit time

Base resistance:
dope very heavily

Base transit time:
make base very thin
grade the base to build in
a field for the electrons

Note: If the base is very thin it
may be difficult to grade the
composition very much.

As with ballistic injection, a problem is that if the carriers
get too much energy from the field they can scatter into
a higher energy band and slow down.
Collector issues:

**Base width modulation:** not an issue because of heavy base doping

**Avalanche breakdown:** an issue with InGaAs bases and collectors

*Solution:* wide band gap collector

**Turn-on/Saturation off-set:** can be significant with a homojunction collector

*Solution:* wide band gap collector
Collector issues, cont: **a caution about wide bandgap collectors**

A wide bandgap collector can be a problem if there is a conduction band spike at the interface. The collector won't collect at low reverse biases!!

The solution is to grade the base-collector interface.
HBT Processing: a typical cross-section

We'll look extensively at processes and circuits our next in lecture (#11).