Electronic Circuits I (ELE 404): Notes

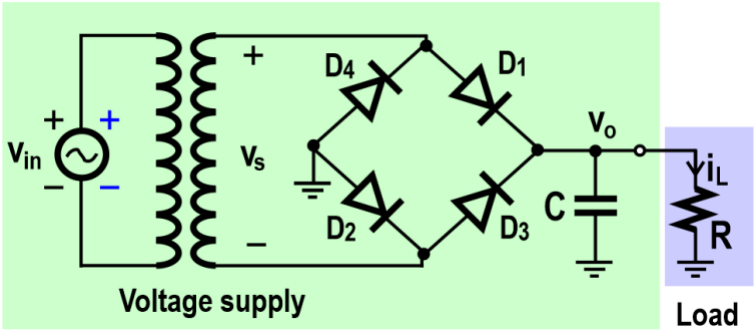
Adam Szava

Winter 2022

Introduction

This document is a compilation of my notes from Electronic Circuits I (ELE 404) from Ryerson University. All information comes from my professor's lectures, the course textbook *Microelectronic Circuits by A. Sedra, et all*, and online resources.

The beginning of this course is review of PCS 224. I will choose to skip it in this document since I have another note for it.



Module 1: Diodes and their Applications

Current in a PN-Junction

In a *pn-junction*, the *saturation current* or *leakage current* is defined as this combination of unchangeable parameters based on the physical properties of the semiconductor:

$$I_s = Aqn_i^2 \left(\frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A} \right)$$

Where:

- I_s is the saturation current.
- A is the area of the *pn-cross section.*
- q is the charge of an electron.
- n_i is the intrinsic carrier concentration.
- D_p is the diffusivity of holes.
- D_n is the diffusivity of electrons.
- L_p is the diffusion length of holes (the distance that holes travel through the n-type before being neutralized).
- L_n is the diffusion length of electrons (the distance that electrons travel through the p-type before being neutralized).
- N_D is the doping concentration of donors.
- N_A is the doping concentration of acceptors.

From this definition of I_s, the current through a pn-junction by diffusion is:

$$I_{diff} = I_s \left(e^{\frac{V_F}{V_t}} - 1 \right)$$

Where:

- I_s is the saturation current.
- I_{diff} is the current due to diffusion.
- V_F is the applied voltage across the junction.
- V_t is the thermal voltage defined as:

  $$V_t = \frac{k_B T}{e}$$

  Where:

  - V_t is thermal voltage.
  - k_B is Boltzmann's constant:

    $$1.38 \times 10^{-23} \frac{J}{K}$$

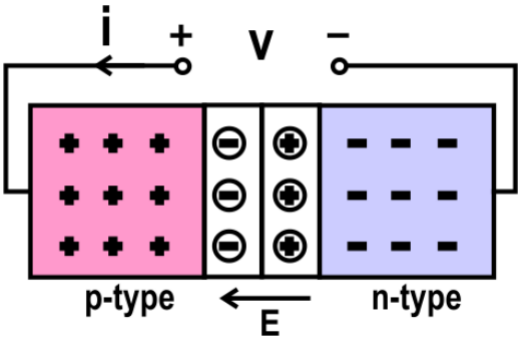
  - e is the charge of an electron.

Reverse Junction Breakdown

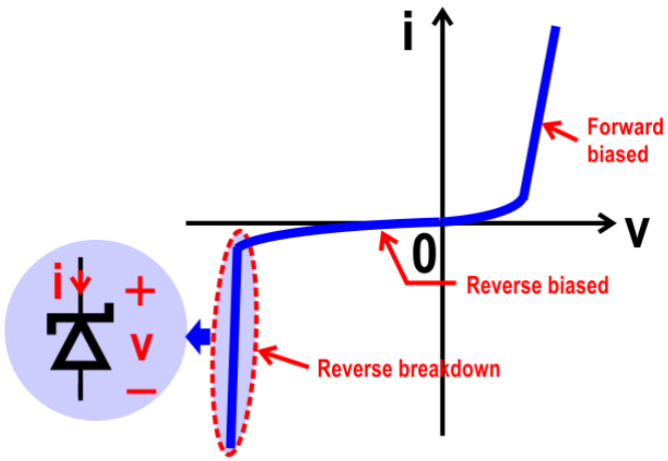
Once the pn-junction applied voltage in reverse bias reaches a critical level, covalent bonds are broken, electron hole pairs are generates, and current sharply increases while the change in voltage is negligible. This process happens in two steps:

1. **Zener Breakdown** is when the first covalent bonds are broken.
2. **Avalanche Breakdown** is when the Zener breakdown causes exponential breakdown of the covalent bonds in the crystal.

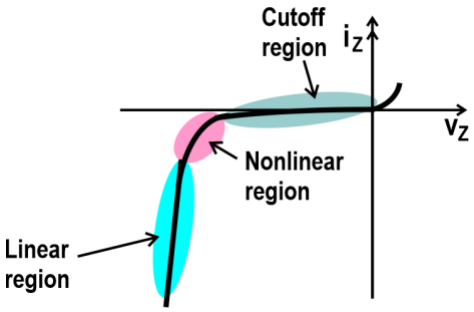
The following image is of a standard pn-junction:



The following is the $i - v$ characteristic of a pn-junction undergoing breakdown:

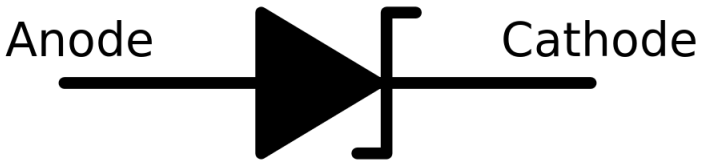


We break this reverse bias breakdown into three regions:



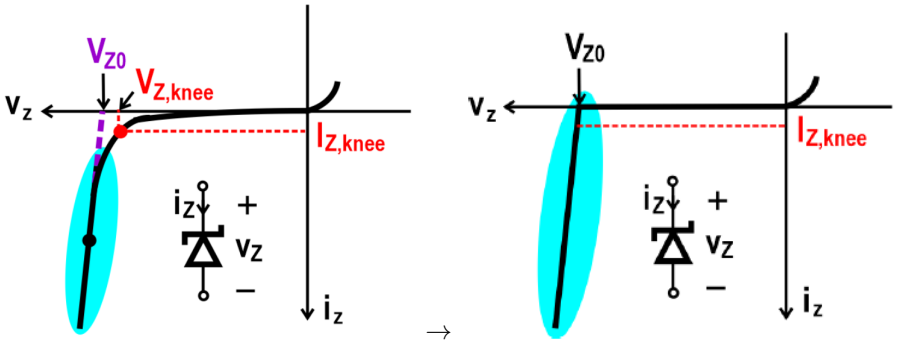
1. **Cutoff region** is where the current is approximately 0 even while the applied voltage varies greatly.
2. **Nonlinear region** is where both the current and the voltage vary greatly but their relation in non-linear.
3. **Linear region** is where the voltage is nearly constant while the current increases dramatically in a linear relation with very steep slope.

The **Zener Diode** is a diode (defined later for some reason) which operates in the breakdown configuration. The Zener diode is denoted:



Model of the Zener Diode

In order to approximate the curve of the Zener diode, in order to make an equivalent linear circuit, we want to treat the cutoff region as 0, the non-linear region as a sharp corner, and the linear region as it is. This is done as in the following image:



Where:

- $V_{Z,knee}$ and $I_{Z,knee}$ are the value of the current and voltage at the "knee" of the linear region.
- V_{z0} is the "activation voltage" for the Zener diode, and is the amount of voltage consumer by the diode when active.

Remember, Zener diodes operate in reverse bias.

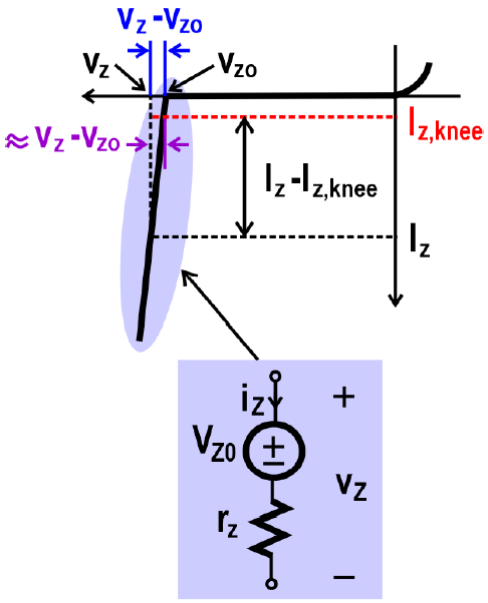
The linear region of the Zener diode can be modelled by a linear equation in terms of i_z (r_z is the resistance of the semiconductor):

$$v_z = V_{z0} + r_z(i_z - i_{z,knee})$$

This can be approximated since $i_z \gg i_{z,knee}$:

$$v_z = V_{z0} + r_z i_z$$

Which is equivalently modelled by the following circuit:

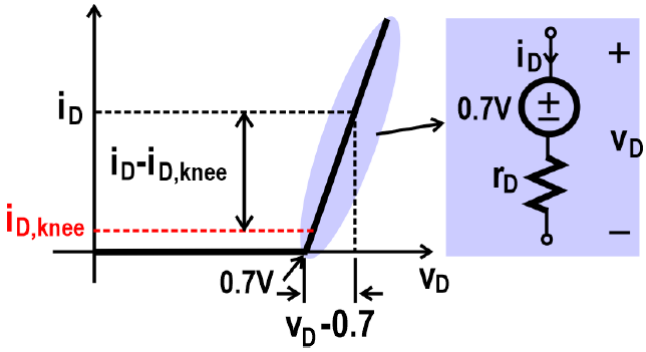


This linear model only applies when the diode is on, allowing current to pass. Otherwise the equivalent model is an open circuit.

Conventional Diodes

Model of a Conventional Diode

A conventional diode operates in a very similar way, except no breakdown is occurring and the diode is operating in forward bias. For conventional diodes we can also make an equivalent circuit diagram, as follows:

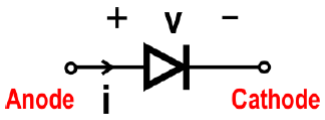


$$v_z = 0.7 + r_D i_D$$

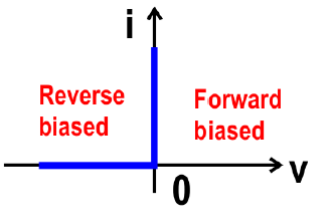
The number 0.7 is an approximation which is standard for all conventional diodes and is the voltage required to activate and power the diode. The diode consumed $0.7V$ when operating, which is why in the equivalent model we have a constant DC voltage source which is decreasing the potential difference by that amount.

Ideal Conventional Diodes

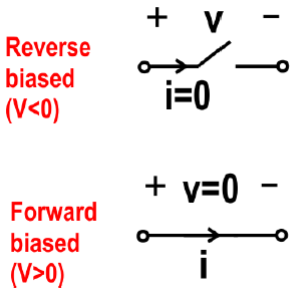
The diode is denoted with the following symbol:



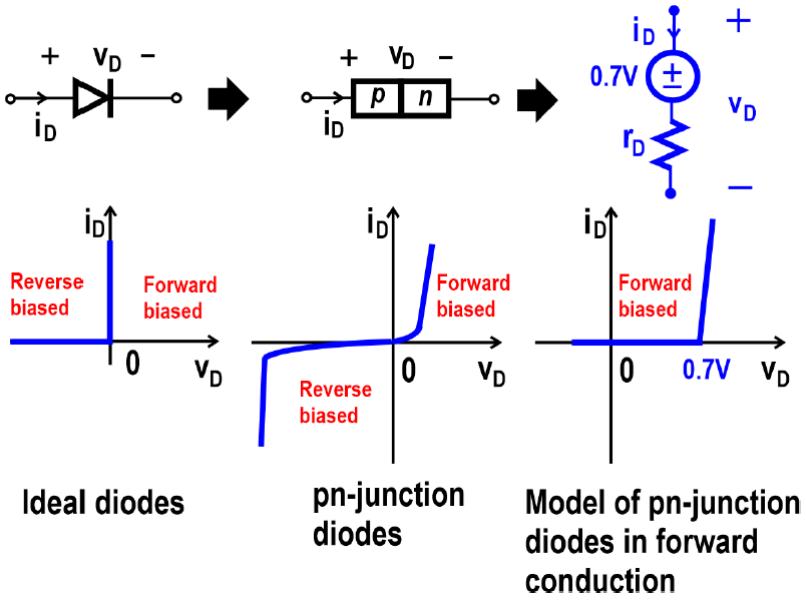
... and has the following **ideal** characteristics:



One way to think of an ideal diode is like a switch:



Diodes are realized (non-ideally) using pn-junctions, whose characteristics we then linearise into the equivalent circuit:

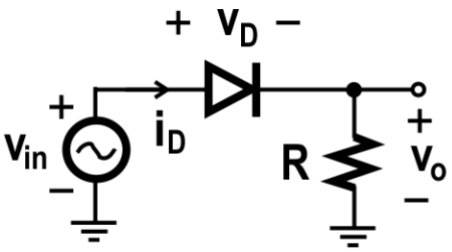


Applications of Diodes

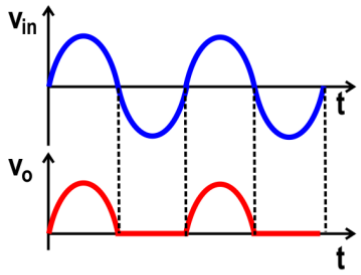
Half-Wave Rectifiers with No-Capacitor

Half-wave rectifiers turn half of the sinusoidal AC current into just DC $0V$. This is typically the negative half of the sinusoid. The other half stays the same. Voltage rectifiers are used in many other circuit applications.

The following circuit is a half-wave rectifier with a load (begin the resistor).



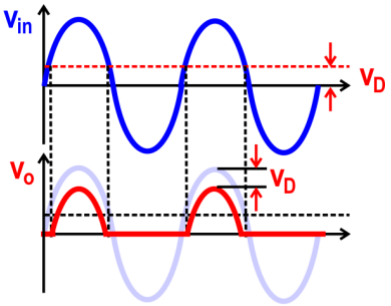
Ideally this circuit operates resulting in the following output:



(a) Ideal diodes

As you can see, the negative parts of the wave are turned to $0V$ and the positive parts are the normal AC signal. This is because when the potential is negative across the diode, the diode is off and current is not permitted through it.

An **actual** implementation of this circuit using *pn-junction diodes* would have the following output:



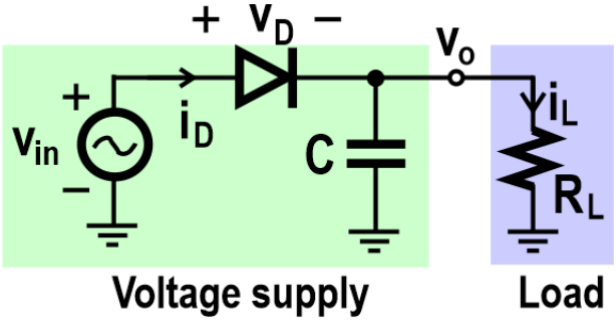
(b) pn-junction diodes

This is because the diode has some activation voltage v_D which will need to be reached before the diode is turned on. For this reason we effectively shift the input signal down by v_D as well as rectify its negative parts.

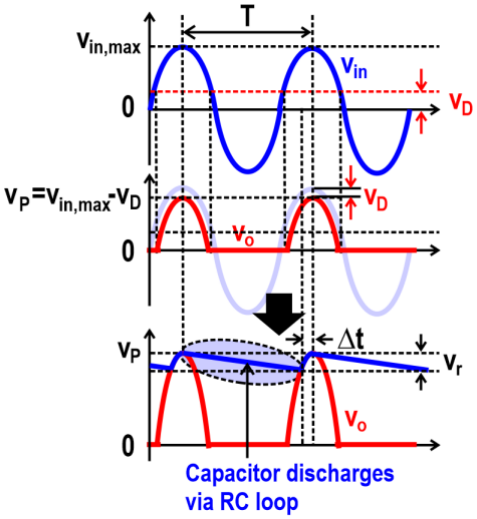
A half-wave rectifier is not efficient since half of the energy of the ave is lost in the process.

Half-Wave Rectifiers with Capacitors

If we want to be able to get a DC signal out of this AC input signal, we need to be able to smooth out the curves. We do this using a capacitor. Recall that a capacitor always gradually discharges its stored voltage in its electric field. We can use this to our advantage by placing a capacitor in the half-wave rectifier circuit as follows:



Here, we still have a resistive load, but now the power supply will output a steadier output as the following:



The actual output v_o is the blue curve on the bottom graph. This happens because as the voltage in the AC signal rises, the voltage across the capacitor matches it until the peak. At that point the AC signal will decrease faster than the voltage across the capacitor since the capacitor stores voltage and dissipates it slowly. For this reason the potential across the diode will go below $0.7V$ and so it will shut off. At that point the source is disconnected from the circuit, and so the capacitor will dissipate its energy governed by:

$$v_o = V_p e^{-\frac{t}{\tau}}$$

Until the next wave of from the half-wave rectifier comes to recharge the capacitor. The discharging process ends when $v_{in} > v_o$. Δt which is the time it takes to recharge the capacitor, is typically negligible compared to the period of the oscillation.

The capacitor gets charged to a maximum voltage (V_p) of:

$$V_p = v_{in,max} - v_D$$

V_r **voltage ripple** defined as the difference between V_p (peak voltage) and the voltage when discharging ends. It is essentially how much the blue output voltage in the last graph varies. We see that:

$$V_r = \frac{T}{\tau} V_p = \frac{V_p}{f\tau}$$

Since at this point v_o is nearly constant, we can see that the current i_L is:

$$i_L \approx \frac{V_p}{R_L}$$

... and so:

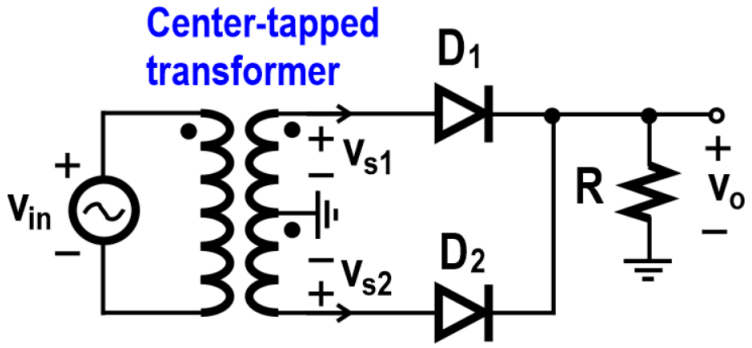
$$V_r \approx \frac{i_L}{fC}$$

The **conduction time** Δt is typically negligibly small but we do have a formula to compute it:

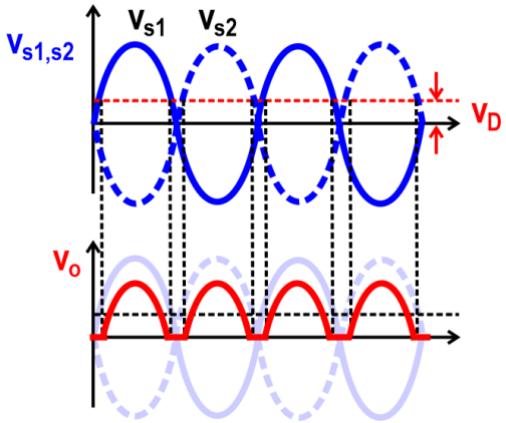
$$\Delta t \approx \frac{1}{\omega}\sqrt{\frac{2V_r}{V_p}}$$

Full-Wave Rectifiers with Transformers

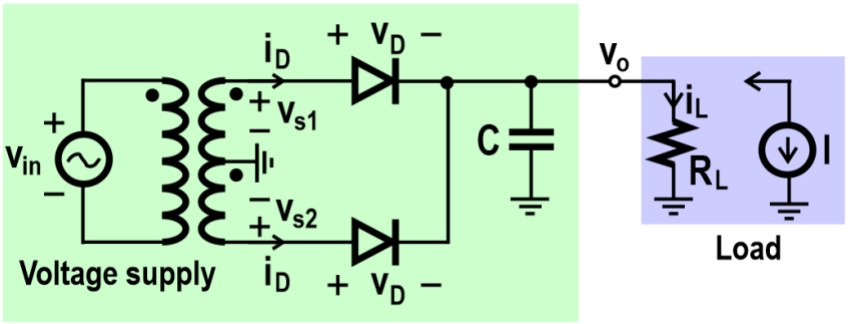
A full wave rectifier does not waste half the sinusoid, but instead flips the negative portion to become positive instead.One type of full wave rectifier is built with a transformer as follows:



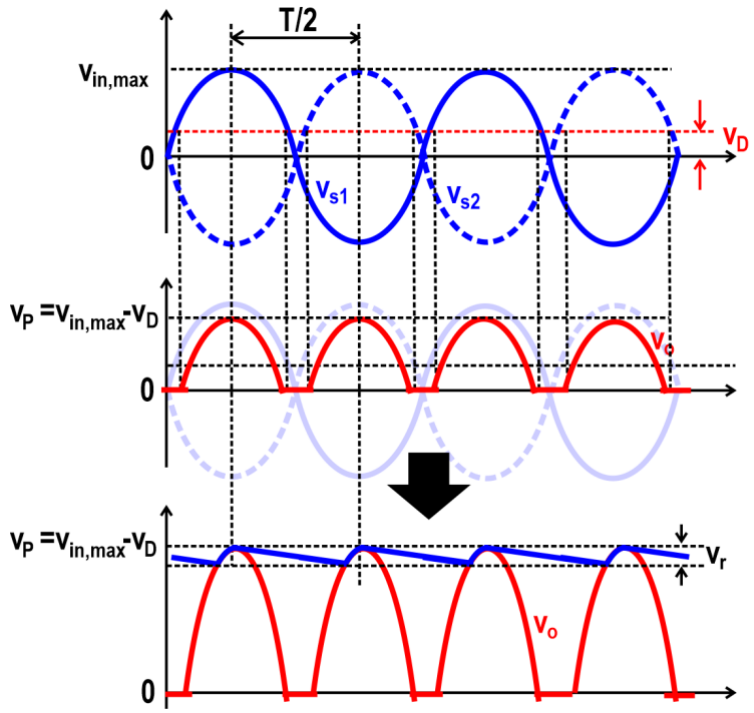
In this circuit, while V_{in} is positive, then V_{s1} will also be positive, causing a voltage drop across the diode $D1$ which turns it on. While V_{in} is negative, then V_{s2} will be positive, causing voltage drop across the diode $D2$ which turns it on. Either way the voltage drop across the resistor will always be positive. This circuit has the following input/output:



Now if we have a capacitor we can use this rectifier as an AC to DC converter:



Which would have the following output characteristic:



With a resistor load, the voltage ripple would be:

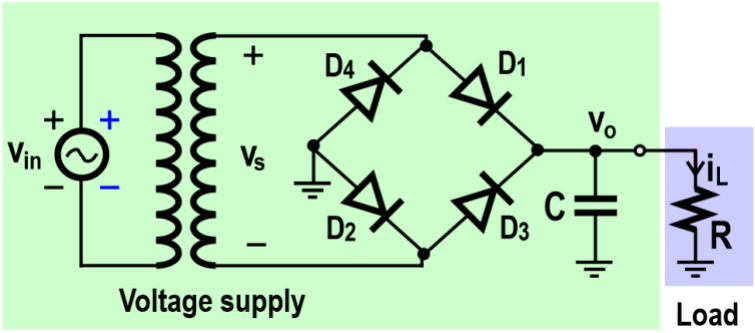
$$V_r = \frac{V_P}{2f\tau} \approx \frac{i_L}{2fC}$$

With a current load:

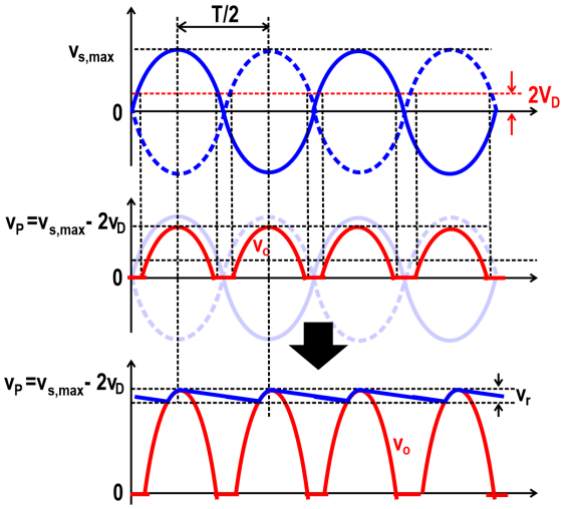
$$V_r = \frac{I}{2fC}$$

Full-Wave Bridge Rectifiers

Full wave bridge rectifiers are built as in the following circuit schematic:



This circuit has an output voltage using the capacitor of:



This works because when a positive voltage is applied to the transformer as input, the same side will be positive on the output of the transformer. This causes $D1$ to turn on, and allow the current through to the output. The current returns to the negative terminal of the transformer by the ground node which turns on $D2$. $D4$ and $D3$ are off during this process. The opposite is true when a negative voltage is applied. This guarantees that there is always positive voltage at the output, as in a full wave rectifier which can then be smoothed out by the capacitors.

The pros to using this type of rectifier is that it is less expensive since you do not need a center tapped transformer, however you do lose $2V_D$ of voltage because you need to power two diodes. With a resistor load, the voltage ripple would be:

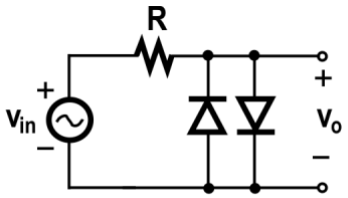
$$V_r = \frac{V_P}{2f\tau} \approx \frac{i_L}{2fC}$$

With a current load:

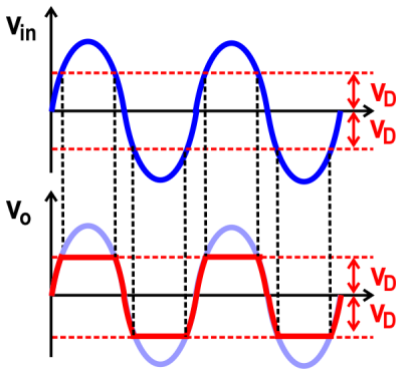
$$V_r = \frac{I}{2fC}$$

Voltage Clippers

Voltage clippers limit a signal to be below some threshold. The circuit which implements this is:



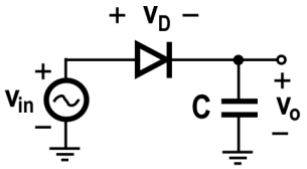
This circuit has an output graph as follows:



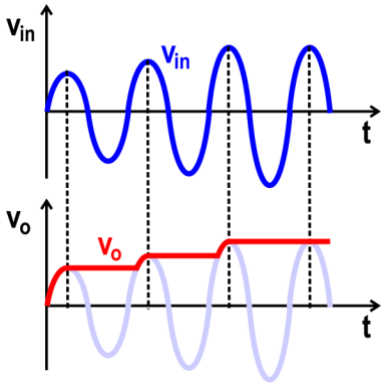
Diodes operate when $|V_{in}| > V_D$, and always one is operational within that range (either positive or negative). This limits the maximum voltage to be V_D.

Peak Detectors

If you take the half wave rectifier circuit with a capacitor, but do not give it a load to discharge onto, then the voltage of the capacitor will stay constant unless it is being raised by a new peek in the input signal. We call this circuit a *Peek Detector*:

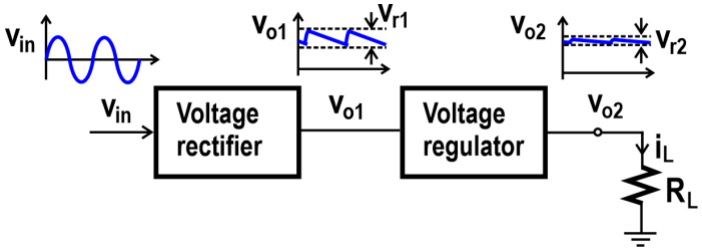


This circuit has an output graph as follows:



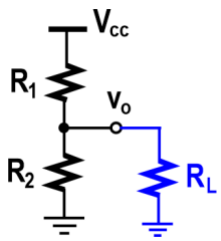
Voltage Regulators with Resistors

A voltage regulator is used to reduce the ripple voltage of a signal after coming from a voltage rectifier, as in the following image:



Voltage regulators provide stable output voltage from a less-stable supply voltage, and provide stable output voltage even in the case of a varying load. As the load changes, the current will change but the voltage must be stable. We will often consider the rate of change of the voltage from the regulator with respect to change in the load.

We will now discuss resistor based voltage regulators. Consider the circuit below:



V_{CC} is voltage from a voltage rectifier, and voltage v_o is the voltage we want out from the resistor based regulator.

In this circuit:

$$v_o = \frac{R_L \parallel R_2}{R_L \parallel R_2 + R_1} V_{CC}$$

Implying:

$$\frac{dv_o}{dV_{cc}} = \frac{R_L \parallel R_2}{R_L \parallel R_2 + R_1}$$

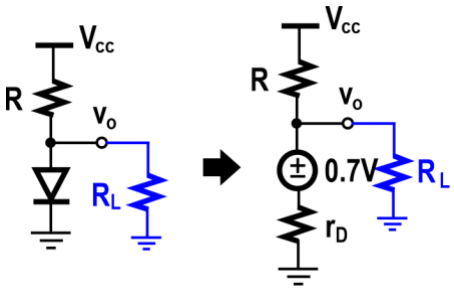
Which isn't good since v_o is still sensitive to the load resistance R_L. Even if we set $R_L \gg R_1, R_2$ the equation simplifies to:

$$v_o = \frac{R_2}{R_2 + R_1} V_{cc}$$

V_o is now independent of R_L but still sensitive to fluctuations in V_{cc}. We need a better kind of regulator.

Voltage Regulators with Diodes

Consider the following circuit:



By doing KCL at the output node we have:

$$v_o = 0.7 \left(\frac{1 + \frac{GV_{cc}}{0.7 g_D}}{1 + \frac{G_L + G}{g_D}} \right)$$

Recall that permittivity is the reciprocal of resistivity:

$$G = \frac{1}{R}, g_D = \frac{1}{r_D}, G_L = \frac{1}{R_L}$$

If the diode is ideal then $v_o = 0.7V$ which is constant and not dependent on R_L or fluctuations V_{cc} but not attainable. If $r_D \ll R, R_L$ then we have:

$$v_o \approx 0.7 \left(1 + \frac{GV_{cc}}{0.7 g_D} - \frac{G_L + G}{g_D} \right)$$

From this we see:

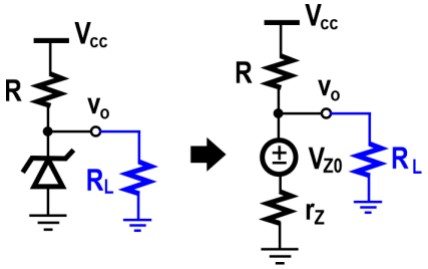
$$\frac{dv_o}{dR_L} = \frac{0.7}{g_D}$$

$$\frac{dv_o}{dV_{cc}} = \frac{G}{g_D}$$

And since g_D is very large, v_o is not sensitive to R_L or fluctuations in V_{cc}.

Voltage Regulators with Zener Diodes

Consider the following circuit:



The use of Zener diodes provides an advantage as g_z is higher than g_D on a conventional diode.

Once again by doing KCL at the output node we have:

$$v_o = V_{Z0}\left(\frac{1+\frac{GV_{cc}}{V_{Z0}g_z}}{1+\frac{G_L+G}{g_z}}\right)$$

If $r_Z \gg R, R_L$ then we have:

$$v_o \approx V_{Z0}\left(1+\frac{GV_{cc}}{V_{Z0}g_z}-\frac{G_L+G}{g_z}\right)$$

From this we see:

$$\frac{dv_o}{dR_L} = -\frac{V_{Z0}}{g_z}$$
$$\frac{dv_o}{dV_{cc}} = -\frac{G}{g_z}$$

Which makes this voltage regulator even less sensitive to the load resistance and fluctuations in V_{cc}.

I have decided at this point to base my notes off of textbook chapters instead of lecture modules, but I don't want to get rid of all the pages so far. For this reason I will now begin by going through *Chapter 4: Diodes* from the textbook which will repeat a lot of content but hopefully make a lot more sense. I will leave out *Chapter 3: Semiconductors* since it is all PCS 224 Review.

I will still divide I will still divide *Chapter 7, 8, and 10* into two portions, to better follow the flow of the course.

Chapter 4: Diodes (Module 1)

4.1: The Ideal Diode

The ideal diode is the mathematical object which we wish to implement in a real circuit. Before we learn how a real diode works, it is worth our time to look at idea diodes.

An idea diode is denoted by the following symbol:

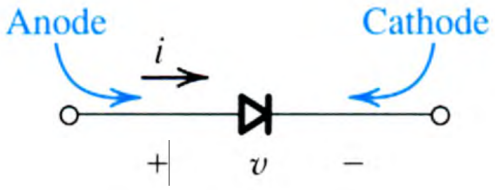


Figure 1: Circuit symbol for a diode with reference polarity and current direction.

This circuit element behaves like an open circuit when a negative voltage is applied (use the diagram as reference for polarity) (This is called reverse bias). If a positive current is applied in the reference direction then a zero voltage drop appears across the diode and the idea diode behaves like a short circuit (this is called forward bias). The magnitude of the current is then determined by the external circuit.

Based on this description, we see that the $i-v$ characteristic must be:

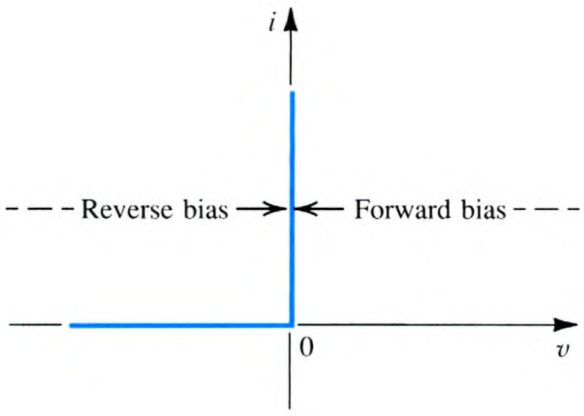


Figure 2: $i-v$ characteristic of the **ideal diode**.

We can think of the diode like a switch which is off (meaning no current flow) if the potential across it is negative, and on (meaning current flow) if the current through it is positive and the potential across it is $0V$.

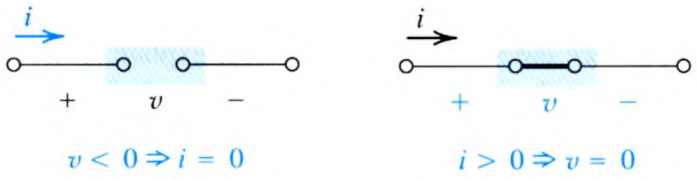


Figure 3: Thinking of the diode as a switch.

The following example shows how the diode acts in a simple circuit:

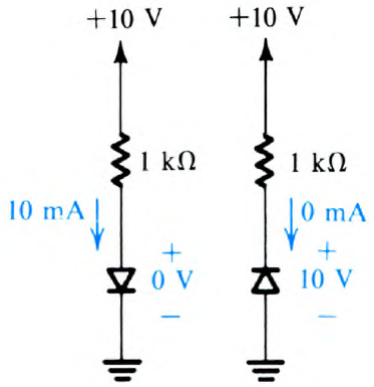
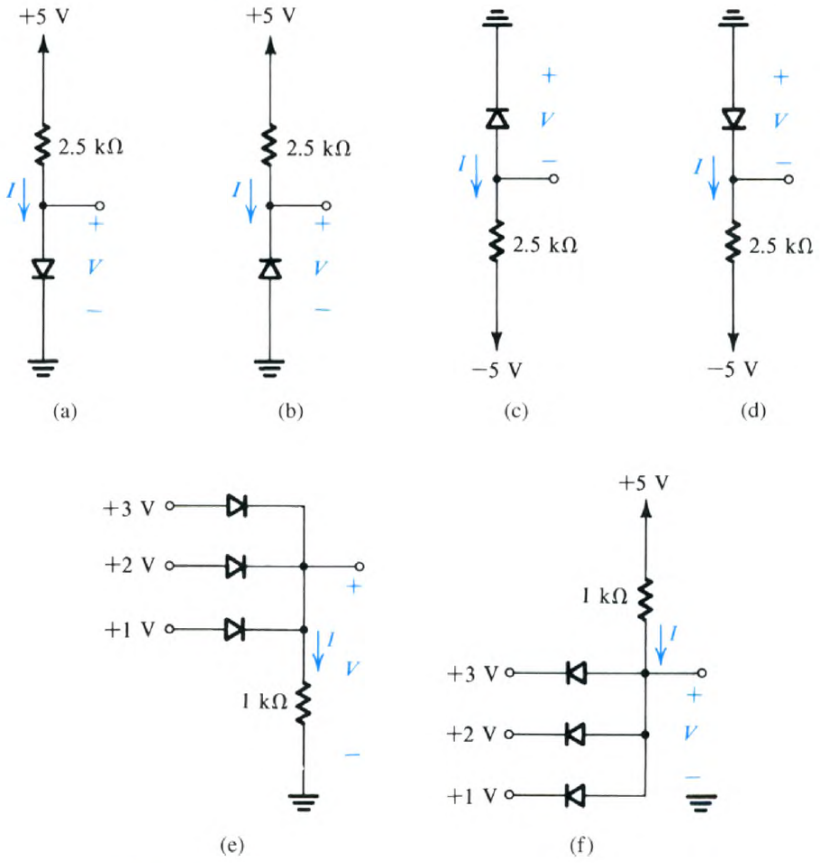


Figure 4: Two examples of the diode in use. One in forward bias, and one in reverse bias.

As you can tell, the diode is a non-linear circuit element. It is however, made of two linear segments, and so it is **piecewise linear**. If a device is piecewise linear, and we can somehow limit the the of the device to just one of the linear pieces, then we can consider it a linear circuit element in that range.

Example 4.1.1 Ideal Diodes

Assuming the following diodes are ideal, find I and V for each circuit.



For circuit a) we see that a positive current is going into the diode, meaning the potential across it is 0 ($V = 0V$). That current can be found using Ohm's law:

$$I = \frac{V}{R} = \frac{5V}{2.5k\Omega} = 2mA$$

For circuit b) we see that a negative potential difference is across the diode, which means the current through the diode is 0 ($I = 0A$). The voltage drop across the resistor can be found by Ohm's law:

$$V = IR = (0A)(2.5k\Omega) = 0V$$

Therefore the voltage at the node above the diode must be $5V$ and so $V = 5V$.
For circuit c) we see that a negative potential difference is across the diode, which means the current through the diode is 0 ($I = 0A$). The voltage drop across the diode is $0V$ for the same reasoning as **circuit b**, so the voltage on the bottom node of the diode is $-5V$. Notice the polarity of V, $V = 5V$.
For circuit d) we see that a positive current is going through it, meaning the potential across the diode is 0 ($V = 0V$). We can find the current using Ohm's law:

$$I = \frac{V}{R} = \frac{0V - (-5V)}{2.5k\Omega} = 2mA$$

For the next two circuits the following video was helpful:

Click here for helpful video explaining the last two circuits.

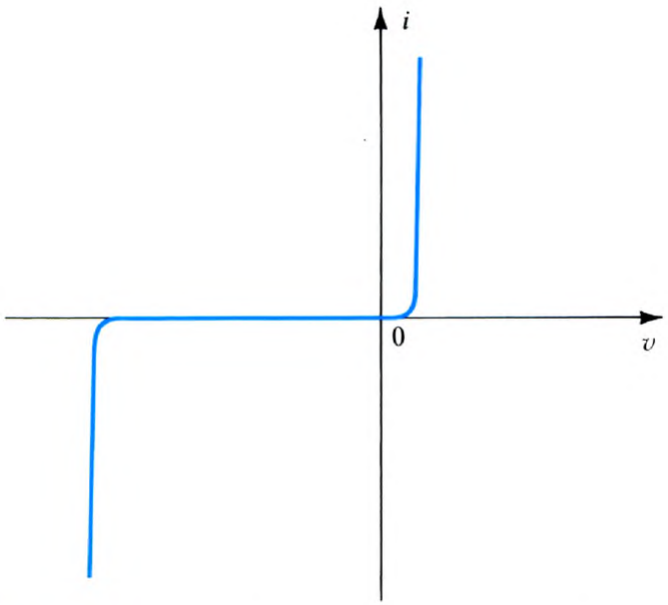
For circuit e) we solve this by considering the voltage V to be initially 0 before the connection to the diode begins. Begin by attaching the first $1V$ diode, V will also rise to $1V$ since the $1V$ diode is forward biased. But since the $2V$ diode is also forward biased, the voltage V will increase past $1V$ shutting off the $1V$ diode. Now the $2V$ diode is forward biased and will bring V to $2V$. The $3V$ diode is also forward biased and this will let V raise past $2V$ and reverse bias the $2V$ diode. V will raise up to $3V$ and stay there, so $V = 3V$.

From there you can find I by Ohm's law to be $3mA$.
For circuit f) we solve this by considering the source voltage increasing to $5V$. If the source voltage begins at $0V$ all diodes will be reverse biased. As the voltage source increases to $1V$, the $1V$ diode will turn on. This established the voltage of V to be $1V$. As the source voltage continues to increase, the voltage V stays at $1V$ ($V = 1V$), and the current I increases to account for the increased voltage. The other two diodes stay in reverse bias. We can determine I using Ohm's law to be $4mA$.

4.2: Terminal Characteristics of (Real) Junction Diodes

We now begin our study of non-idea diodes. The most common type of non-ideal implementation of a diode is the *pn-junction*.

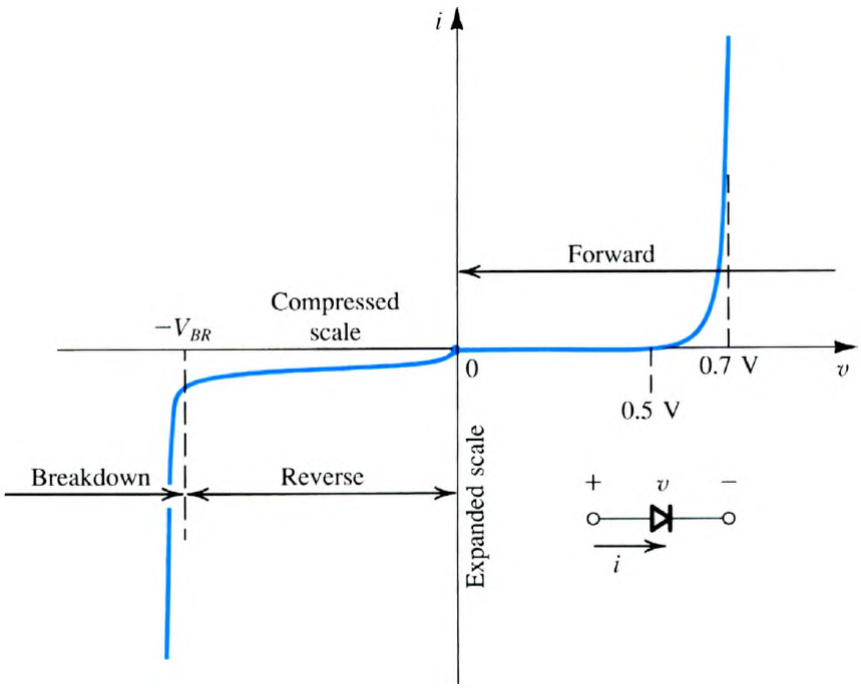
The following is the $i - v$ characteristic of the *pn-junction diode* (in this chapter I will now just call it a diode and it is implied that it is a non-ideal pn-junction diode):



We can break this graph into three operating modes:

1. The **forward-bias** region when $v > 0$.
2. The **reverse-bias** region when $v < 0$.
3. The **breakdown** region when $v < -V_{BR}$.

V_{BR} is called the breakdown voltage and we will discuss it more later. The following graph shows all of the regions of the *pn-junction*:



The Forward Bias-Region

In the forward bias region the characteristic is closely approximated by:

$$i_D = I_s(e^{\frac{v}{nV_T}} - 1)$$

Where:

- i_D is the current through the diode.
- I_s is the saturation current (also called scale current) of the diode (determined by manufacturing process).
- v is the voltage across the diode.
- n is the constant which is also affected by the physical structure of the diode. This constant varies from 1 to 2. For simplicity we assume $n = 1$ for all diodes unless stated otherwise.
- V_T is the thermal voltage:

  $$V_T = \frac{k_B T}{q}$$

  ... which is $\approx 26mV$ at $300K$. For other temperatures:

  $$V_T = 0.0862TmV$$

For current i in the forward direction being $i \gg I_s$, we can approximate the equation as:

$$i_D = I_s e^{v/V_T}$$

For a fully functioning diode, the voltage drop lies in a narrow range between $0.6V$ and $0.8V$, in this course we assume the voltage drop to be $0.7V$. This is a simple model of the diode which we will discuss more of later.

The Reverse-Bias Region

The diode enters the reverse bias region when the voltage across it is negative. In this case:

$$i \approx -I_s$$

This reverse current is very small (nA scale)

The Breakdown Region

In the breakdown region, a large reverse current begins due to the large negative potential breaking bonds within the semiconductor. The diode breakdown is typically will not destroy the diode unless the power dissipated by the diode increases past some specified safe level, done by limiting the current.

4.3 + 4.4: Modelling the Diode + Small Signal Model

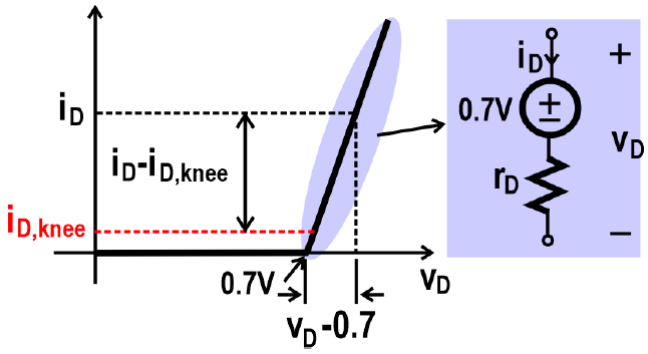
There are two types of analysis you can do of a circuit which contains diodes:

1. **Large-signal analysis** which looks at some input signal as a DC signal.
2. **Small-signal analysis** which looks at only the ripples in the input signal as an AC signal.

Total-signal analysis is the combination of these two types of analysis.

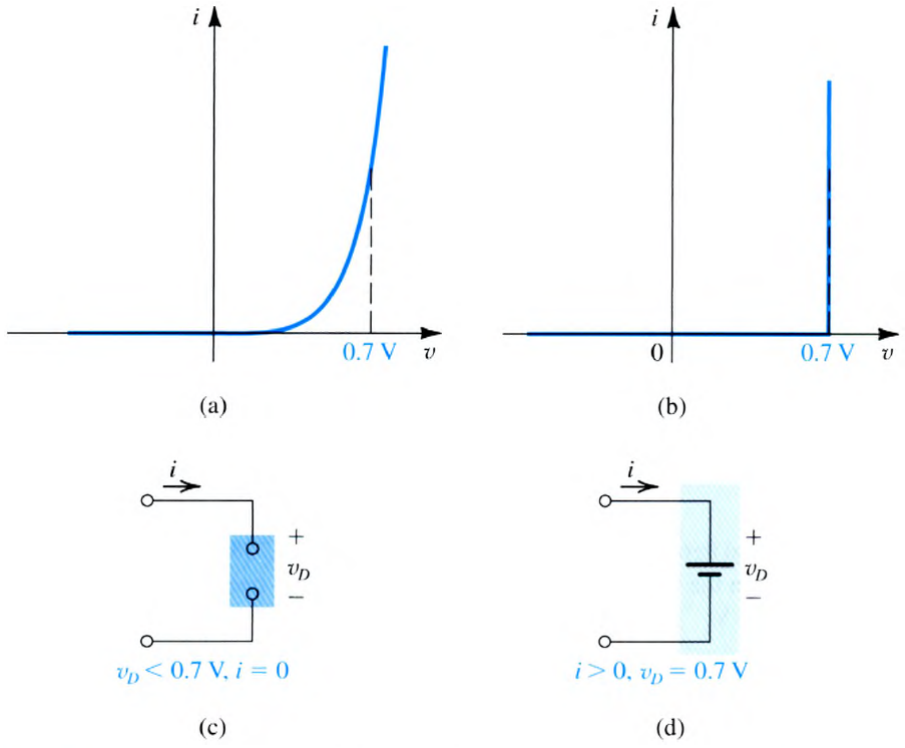
Total-signal Analysis

For **total-signal analysis** you can model the diode as a constant voltage drop in series with a resistor r_d which accounts for the internal resistance of the diode:



Large-signal Analysis

For **large-signal analysis** you can model the diode as a constant voltage drop:



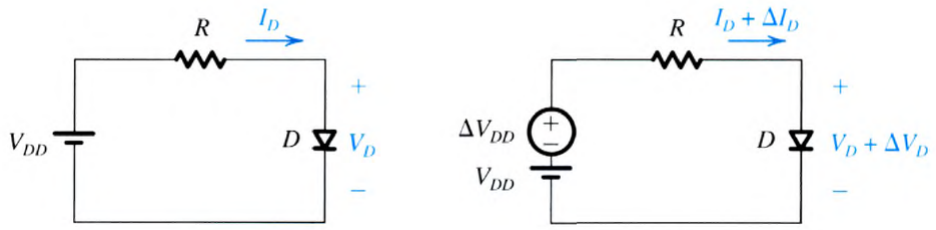
In this case we neglect the internal resistance r_d of the diode since it is significantly smaller than the resistance of the other parts of the circuit.

Small-signal Analysis

Small-signal Analysis is the most detailed type of analysis because the signals are small in magnitude (less than approximately $5mV$ for a voltage signal). For this reason we need to use the exact model for the diode, being:

$$i_D = I_s e^{v/V_T}$$

To aide in the explanation of small-signal analysis, use the following circuit as an example:



We are interested in studying the changes in the current and diode voltage values, as a result in the ripple of the input signal. We use the word signal to emphasize that ΔV_{DD} is typically time varying.

Given that the voltage across the diode is:

$$v_D = V_D + \Delta V_D$$

Where:

- v_D denotes the total voltage across the diode.
- V_D denotes the DC **nominal** voltage across the diode.
- ΔV_D denotes the fluctuations in the V_D signal.

We can substitute this value into our model of current through the diode:

$$i_D = I_s e^{(V_D + \Delta V_D)/V_T}$$

$$i_D = I_s e^{(V_D)/V_T} e^{(\Delta V_D)/V_T}$$

$$i_D = (I_s e^{(V_D)/V_T}) e^{(\Delta V_D)/V_T}$$

This first bracket is by definition the **nominal current** through the diode:

$$i_D = I_D e^{(\Delta V_D)/V_T}$$

Given that:

$$\frac{\Delta V_D}{V_T} \ll 1$$

(Satisfied when $\Delta V_D < 5mV$), we can use the Taylor series to approximate it as:

$$i_D \approx I_D \left(1 + \frac{\Delta V_D}{V_T}\right)$$

$$\boxed{i_D = I_D + \frac{I_D}{V_T} \Delta V_D}$$

Where:

- i_D denotes the total current through the diode.
- I_D denotes the DC nominal current through the diode.
- V_T denotes the thermal voltage.
- ΔV_D denotes the fluctuations in V_D.

The magic of this is that we can think of this as the total current being some nominal current plus the fluctuations of I_D being ΔI_D, as in:

$$i_D = I_D + \Delta I_D$$

Meaning:

$$\Delta I_D = \frac{I_D}{V_T}\Delta V_D$$

Writing it in the following form:

$$\frac{V_T}{I_D}\Delta I_D = \Delta V_D$$

Which is in the form of Ohm's law:

$$R \cdot I = V$$

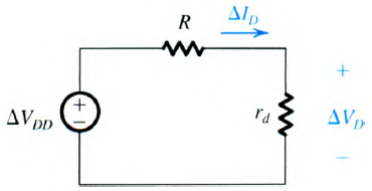
Which means we can think of it like there is a resistor which resists the flow of current during small signal analysis:

$$\boxed{r_d = \frac{V_T}{I_D}}$$

... and so we get the following small signal approximation:

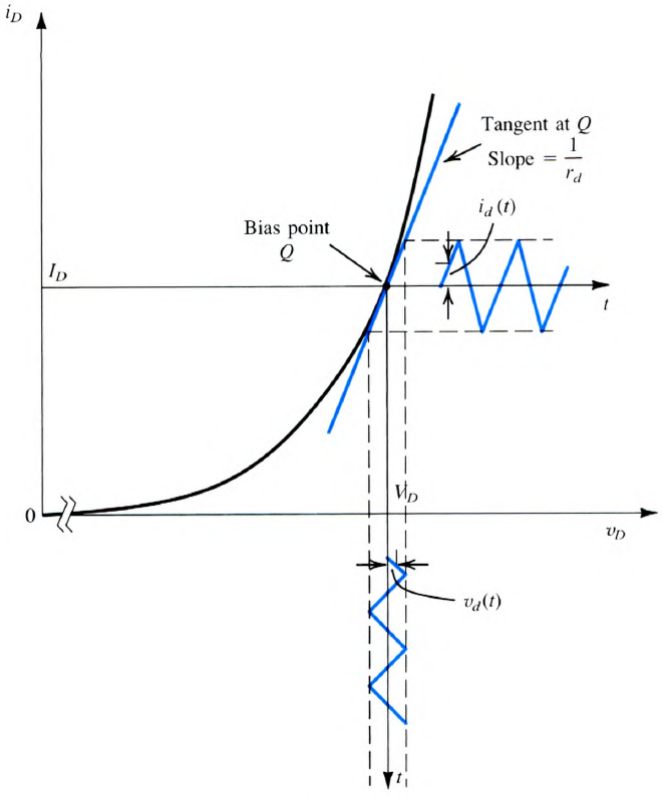
$$\boxed{\Delta V_D = r_d \Delta I_D}$$

Which is represented by the following **equivalent small-signal circuit**:



Note that we have replaced the diode with **just** the resistance. We have not included the constant voltage drop in this analysis because it is already accounted for in the *large-signal analysis*. Additionally we only consider the fluctuations in the value and not the nominal values them selves for the same reason.

The following diagram visually represents what we are doing by this small-signal approximation:



We call the point of operation Q for **quiescent point** which is always around $0.7V$. We approximate the non-linear curve as a linear relation.

We should be clear that there is no actual resistor r_d inside the diode. It is simply a way to represent the linear behaviour of our diode approximation using a linear circuit element.

Note that to even do small signal analysis, you need to know the nominal current through the diode in order to find r_d.

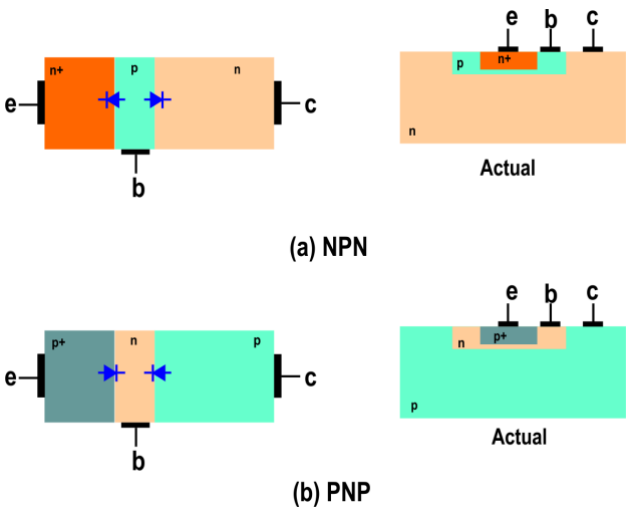
Chapter 6: Bipolar Junction Transistors (BJTs) (Module 2)

Three terminal devices work by using the voltage between two terminals to control the current through the other one, in a sense that third terminal is a controlled source. This chapter begins our study of the first three-terminal device we use called the **bipolar junction transistor** or (BJT).

6.1: Device Structure and Physical Operation

Structure

A BJT consists of 3 semiconductor regions. There are two configurations that the BJT can be in, called *npn* and *pnp* which reflect the order of the semiconductors. The following figure shows the semiconductor profile of the two configurations (note that we also show the *actual* structure of the BJT, however we almost always use the diagram given on the left for analysis):



Evidently, the BJT is a 3 terminal device, with one terminal called the *emitter* (e), the *base* (b), and the *collector* (c). As we study the operation of the BJT these names will make sense.

There are 2 pn junctions within the BJT, and the biasing of the junctions determines the modes of operation of the BJT. Note that **EBJ** is the *emitter-base junction* and **CBJ** is the *collector-base junction*:

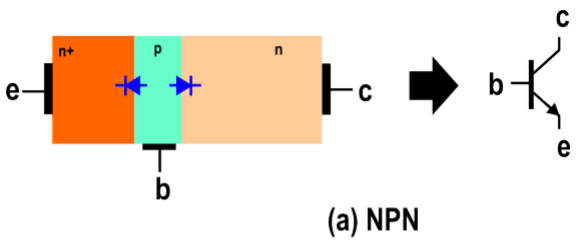
Mode	EBJ	CBJ
Cutoff	Reverse	Reverse
Active	Forward	Reverse
Saturation	Forward	Forward

Active mode is the mode we are concerned with in this course for the most part. This is where the BJT acts as an amplifier. For digital logic applications we use the other two modes

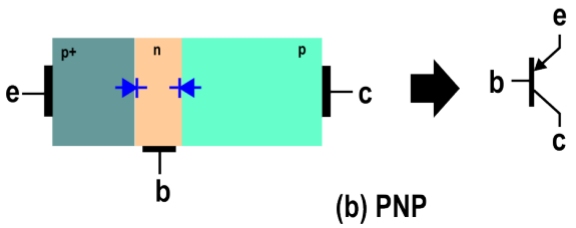
The emitter is always heavily doped, and the base is a very thin layer of semiconductor (in the nm scale).

Both holes and electrons contribute to the flow of current through the BJT, which is where the word *bipolar* comes into the name.

In a circuit, these elements get the following symbols (arrows indicating the direction of current):

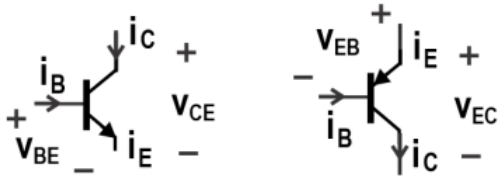


(a) NPN



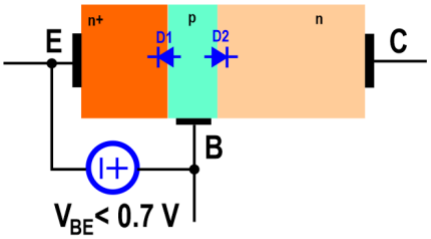
(b) PNP

The following circuit symbols are labeled with the standard direction and polarity for the voltage and current values (*npn* then *pnp*):

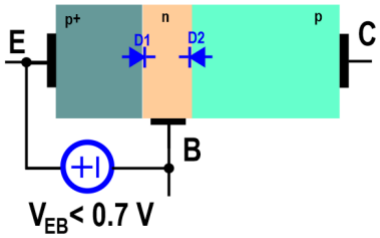


Operation

Given a potential difference between the emitter and base which is insufficient to forward bias the EBJ, no electrons can flow through the device, as in:



(a) NPN

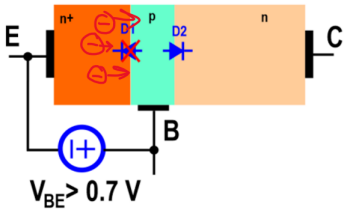


(b) PNP

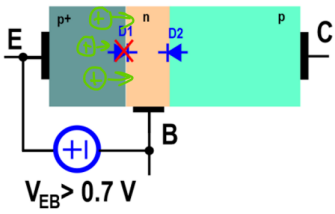
From here we see that:

$$i_E = i_C = i_B = 0A$$

Given sufficient voltage ($0.7V$) in the correct direction based on the configuration, we see that the majority charge carrier of the emitter can begin to flow out of the emitter and into the base/collector. This happens due to the diffusion of majority charge carriers from the heavily doped region.



(a) NPN



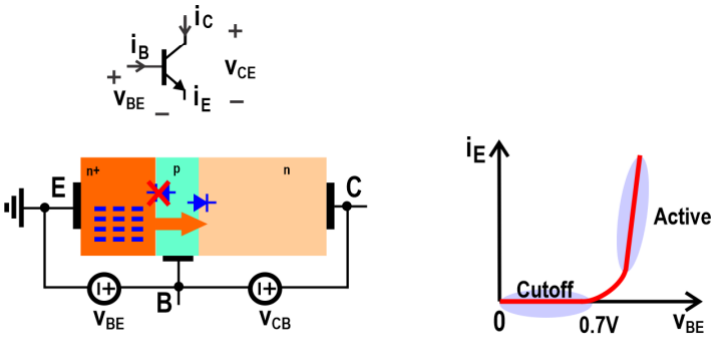
(b) PNP

Recall that current is opposite the direction of the movement of electrons.

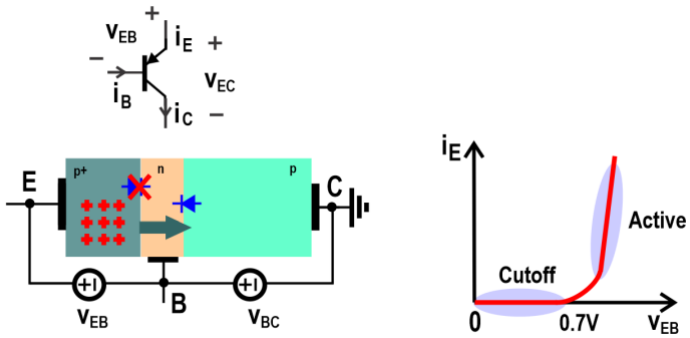
We know from semiconductor physics (and chapter 3 of this course) that:

$$i_E \approx I_s e^{v_{BE}/V_t}$$

Recall that the base is very thin, and so even though CBJ is reverse biased, the electrons can still make their way through, helped by an additional voltage difference added between the base and the collector:



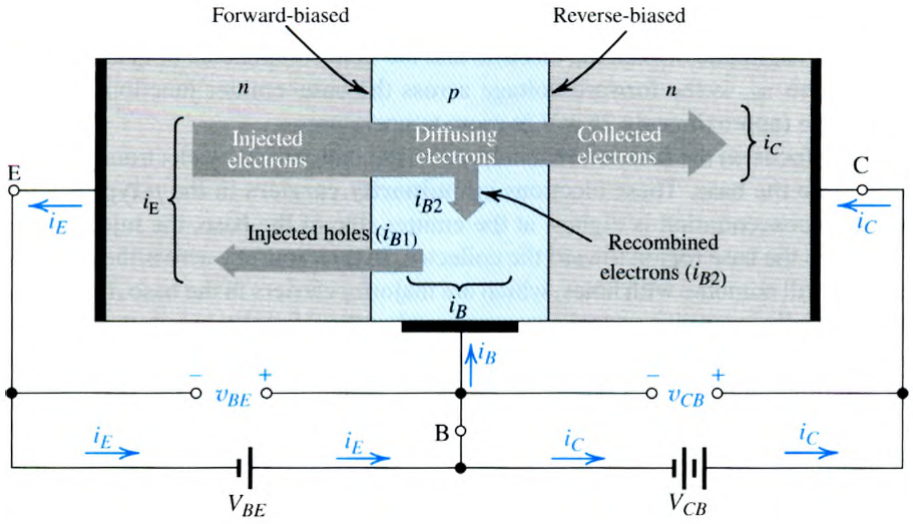
(a) NPN



(b) PNP

This movement of electrons towards the collector will generate the current i_C. Some electrons will reach the base terminal and not go through the CBJ, they then form i_B, but this is minimal. For this reason $i_B \approx 0A$. This set-up is called **active mode**. In the active region the slope of the curve can be thought of as a kind of resistance r_E

Focusing in on the *npn* transistor, we can see where the electrons are going in **active mode**:



Evidently:

$$\boxed{i_E = i_B + i_C}$$

If $i_B \approx 0A$ then:

$$i_C \approx i_E \approx I_s e^{v_{BE}/V_t}$$

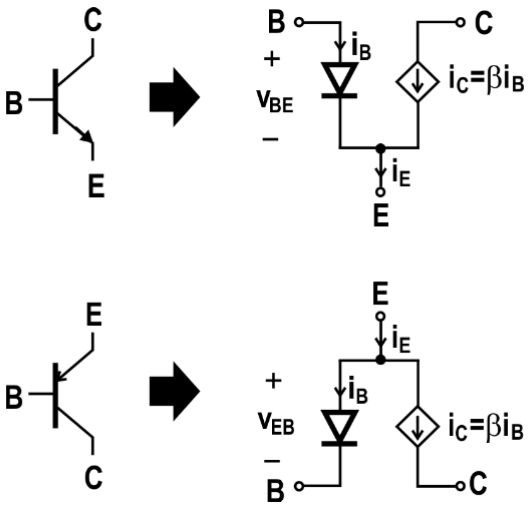
Note that the collector current is independent of the voltage V_{CB} (in an *npn*) or V_{BC} (in a *pnp*).

We can think of the rate at which the emitter current changes with respect to v_{BE} as a kind of resistance (or conductance), as in:

$$\frac{di_E}{dv_{BE}} = \boxed{\frac{i_E}{V_t} = g_E}$$

Large-Signal Equivalent Circuit

The following is the large signal equivalent circuit for the BJT (npn then pnp):



Note that the large signal equivalent circuit takes into account the nominal values of the input, as well as the AC fluctuations.

Current gain (β) is defined as:

$$\beta \equiv \frac{i_C}{i_B}$$

... and is a physical property of the BJT, and so does not vary with current/voltage values. For this reason we can use it in the equivalent circuit.

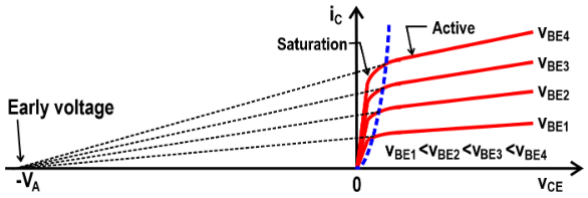
Using β we get the following (very useful) relationships between the terminal currents:

- $$i_C = \beta i_B$$

- $$i_E = (1+\beta) i_B$$
- $$i_C = \left(\frac{\beta}{\beta+1}\right) i_E$$

Base Width Modulation

If you raise v_{CE} for the same v_{BE} you will see an increase in the collector current i_C, as in:



This happens due to the width of the base region being changed by this raise in v_{CE}:

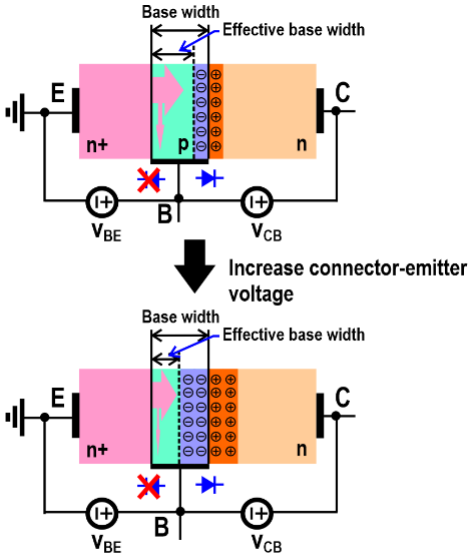


Figure 12: Base-width modulation.

If v_{CB} goes up (which causes v_{CE} to go up as well) then the width of $D2$ goes up, using some of the base's region. This makes the effective base width go down, and makes more of the electrons go to the collector.

Using a linear approximation in the linear region of the above graph, we can see that the collector current, with base width modulation taken into account is:

$$I_{c,total} = I_s e^{\frac{v_{BE}}{V_t}} \left(1 + \frac{V_{CE}}{V_A}\right)$$

... where V_A is called the *Early voltage* and is a property of the device.

You can then distribute to get:

$$I_{c,total} = I_s e^{\frac{v_{BE}}{V_t}} + I_s e^{\frac{v_{BE}}{V_t}} \frac{V_{CE}}{V_A}$$

$$I_{c,total} = I_s e^{\frac{v_{BE}}{V_t}} + \frac{I_s e^{\frac{v_{BE}}{V_t}}}{V_A} V_{CE}$$

$$I_{c,total} = I_s e^{\frac{v_{BE}}{V_t}} + \frac{I_c}{V_A} V_{CE}$$

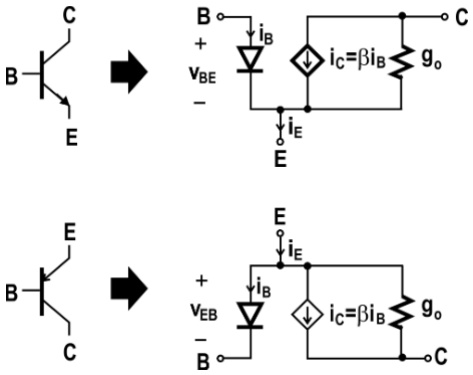
... and so you can think of this like a resistor:

$$I_{c,total} = I_s e^{\frac{v_{BE}}{V_t}} + g_o V_{CE}$$

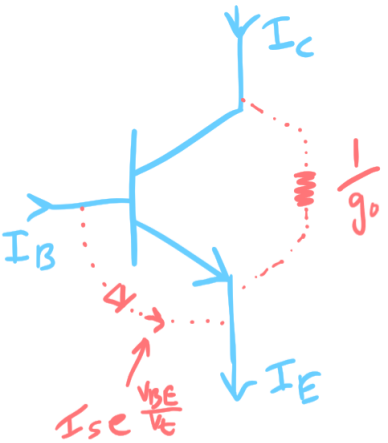
... where:

$$g_o \equiv \frac{I_c}{V_A}$$

This resistance r_o is used to quantify the effect of base width modulation. It is a parameter which functions as a resistor. It is not a real resistor. The equivalent circuit can thus be updated to:

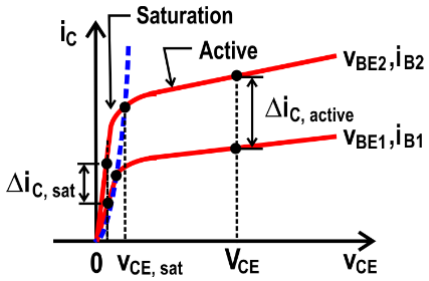


We can then think of the npn transistor symbol a little more accurately as:

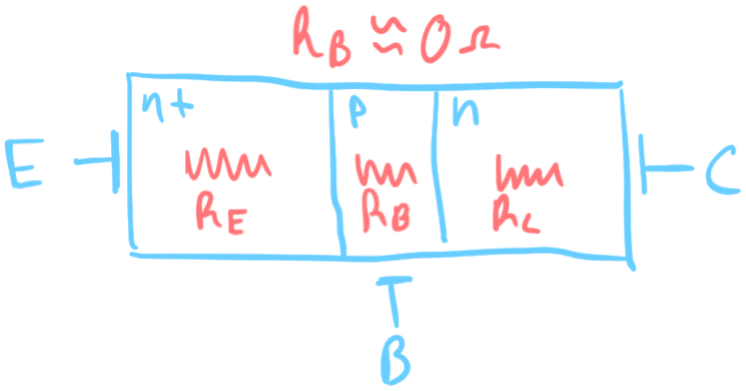


Saturation Mode of a BJT

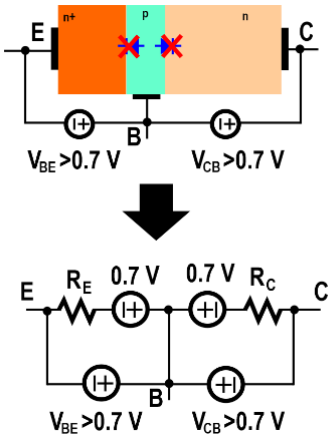
If both diodes within the BJT are on, then the BJT is in saturation mode. This means that a large current flows between emitter and collector. In the following graph, then the BJT operates to the left of the blue line, it is in saturation mode and acts like a linear resistor. For this reason we think of the Saturation mode BJT as a resistor r_{CE}.



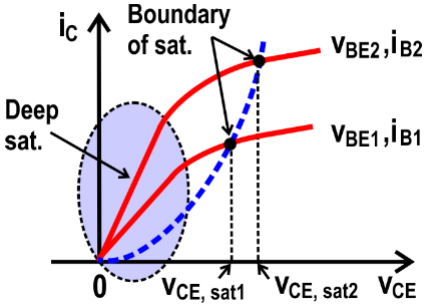
The resistance r_{CE} is due to the resistance of the emitter and collector, and is a physical resistor ($\approx 10\Omega$).



In saturation mode, the BJT can be modelled by two diodes, which can also be modelled by linear circuits, and so:

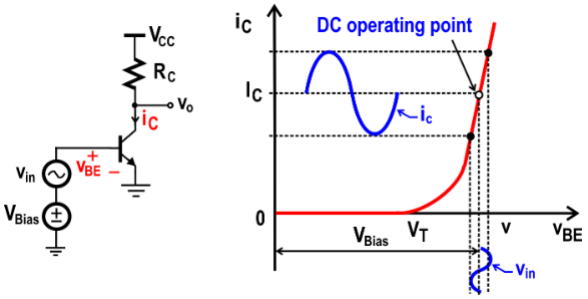


If we take a closer look at the boundary between saturation and active modes, we see the boundary is non-linear, and there is a segment called **Deep Sat.** in which we use the linear approximation.



Small-Signal Equivalent Circuit

The small signal circuit is how the BJT acts with small fluctuations of the input signal.

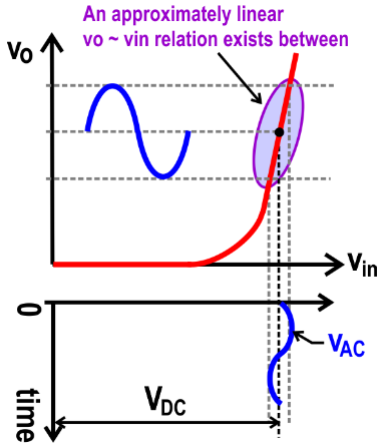


As you can see in the image, we bias the circuit to some DC value, and then apply some v_{in} which is time-dependant and small. So:

$$v_{BE} = V_{Bias} + v_{in}$$

V_{bias} must be greater than $0.7V$ to activate the diode within the BJT. We want the fluctuations to be small so that the BJT always stays in active mode.

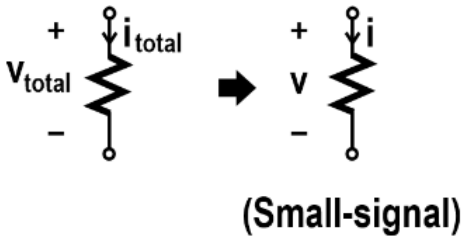
Most electronic devices are non-linear. We want to be able to study them using linear approximations and so we bias the device in a linear region, and use small-signal analysis in that linear region:



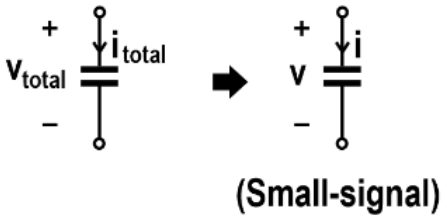
The small signal equivalent circuit depicts the localized behaviour of a nonlinear circuit.

All of the linear circuit element we know have small-signal direct equivalents (note that we denote the small signal components with lower-case letters with no subscript):

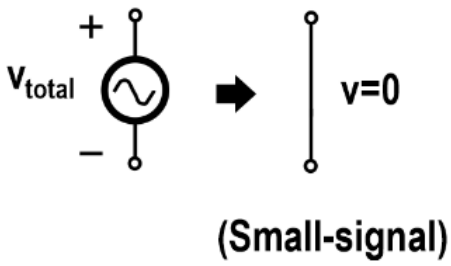
- **Linear Resistors**



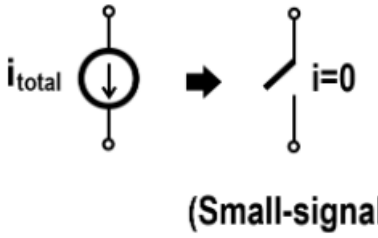
- **Linear Capacitors**



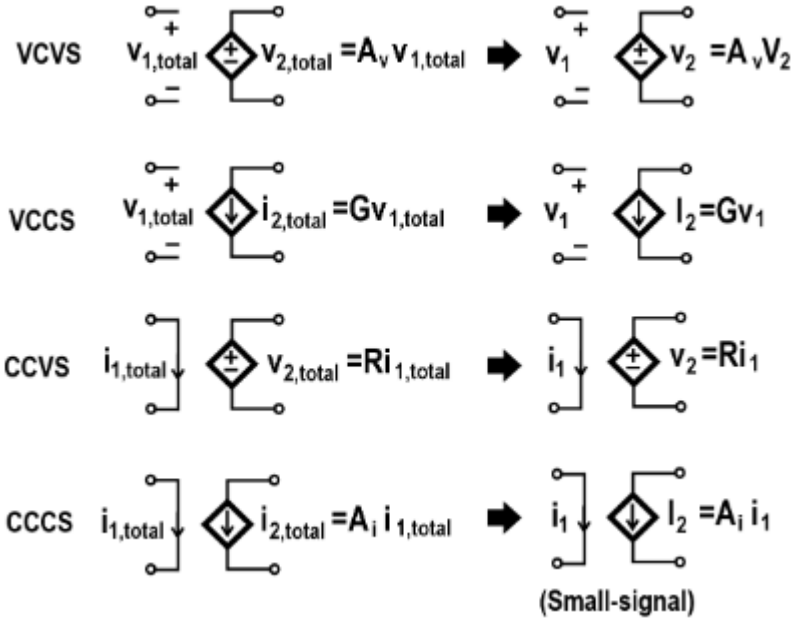
- **Independent Voltage Sources**



- **Independent Current Sources**



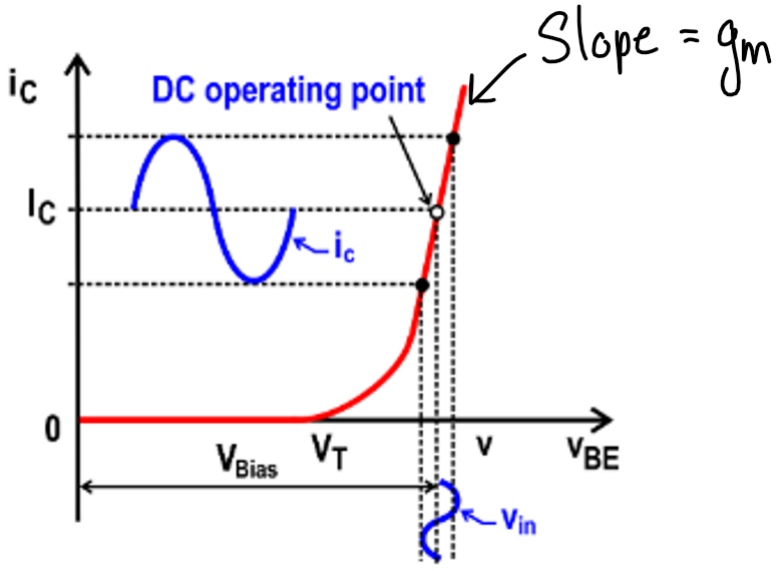
- **Dependent Sources**



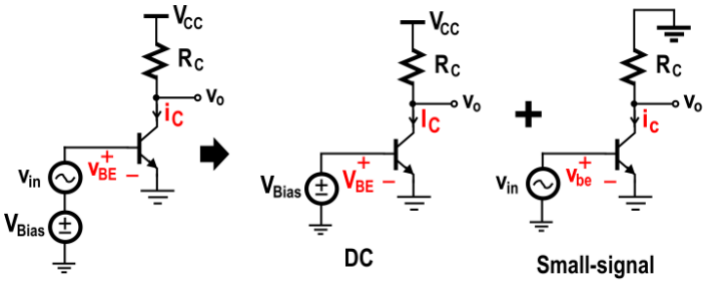
Back to considering a BJT, the slope of the curve at the DC biasing point is called the *transconductance* g_m, meaning:

$$g_m = \frac{di_C}{dv_{BE}}_{v_{BE}=V_{Bias}} = \frac{I_C}{V_t}$$

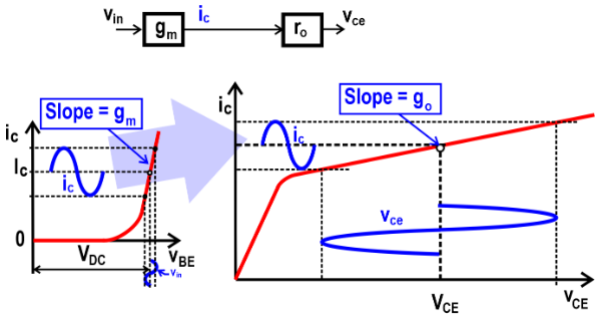
The higher the power consumption (the further the bias point is from boundary of active mode) the higher the transconductance. A high transconductance means that a small fluctuation v_{in} will cause a large fluctuation in output current.



When problem solving we break a circuit into its DC component and AC component, like the following:



Using the concept of transocnductance, we can see how a BJT amplifies a voltage, by converting it to an amplified current, then back to an amplified voltage somewhere else on the BJT:



Small Signal Model of the NPN BJT

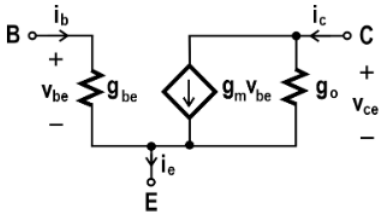
Using some first order approximations of exponentials, and neglecting small terms, we can go from:

$$i_{c,total} \approx I_s e^{\frac{V_{BE}+v_{be}}{V_t}} \left(1 + \frac{V_{CE} + v_{ce}}{V_A}\right)$$

...to:

$$\boxed{i_c = g_o v_{ce} + g_m v_{be}}$$

Which shows that the collector current comes from the transconductance and base width modulation. The small signal equivalent circuit for this is:



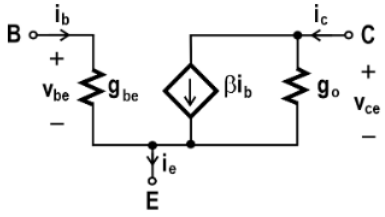
The current gain can also be written as:

$$\beta = \frac{i_c}{i_b} = g_m r_{be}$$

The collector current can then be written as:

$$i_c = \beta i_b + g_o v_{ve}$$

So the small signal circuit can also be written as:



Small signal Parameters

- **Base-emitter resistance** r_{be}:

$$r_{be} = \frac{\beta V_t}{I_C} = \frac{\beta}{g_m} = \frac{V_t}{I_B}$$

- **Emitter resistance** r_e:

$$r_e = \frac{r_{be}}{\beta + 1}$$

- **Transconductance** g_m:

$$g_m = \frac{I_C}{V_t}$$

- **Output Conductance** g_o:

$$g_o \approx \frac{I_C}{V_A}$$

Module 3: BJT Amplifiers

Please forgive the next 14 pages, something made me lose a bunch of progress on this note, and I only have a backup of it in PDF form, so the following is that PDF, as an image, in this PDF. What a mess.

Small signal Parameters

- **Base-emitter resistance** r_{be}:

$$r_{be} = \frac{\beta V_t}{I_C} = \frac{\beta}{g_m} = \frac{V_t}{I_B}$$

- **Emitter resistance** r_e:

$$r_e = \frac{r_{be}}{\beta + 1}$$

- **Transconductance** g_m:

$$g_m = \frac{I_C}{V_t}$$

- **Output Conductance** g_o:

$$g_o \approx \frac{I_C}{V_A}$$

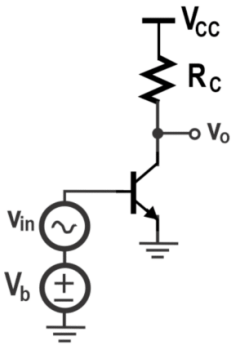
Module 3: BJT Amplifiers

This section of my notes will mostly be from the lecture slides with some addition from the textbook.

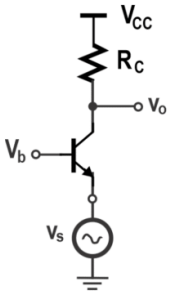
A BJT amplifier is a circuit configuration using a BJT which amplifies an input signal to some output. Amplifiers in general are a widely used circuit element, and BJT amplifiers is just one way we do it.

There are 3 types of BJT amplifier circuits:

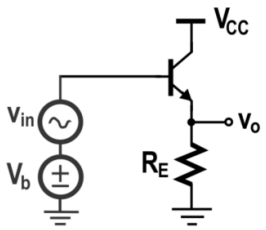
- Common-emitter (CE) amplifiers:



- Common-base (CB) amplifiers:



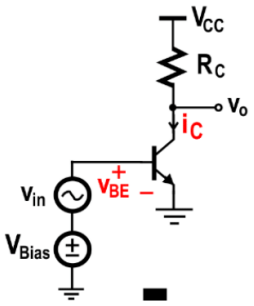
- Common-base (CB) amplifiers:



Before we start talking about the first kind of amplifier, we first have to understand a little better what's happening at the DC biasing point.

Load Lines

We know that in the following circuit, V_{Bias} provides some biasing voltage for the input signal:



In the form of V_{BE}. Given this voltage, we may also consider what is the V_{CE} value at this biasing point. Knowing this value will help us to understand what will happen at the output, and where we are in the $i_C \sim V_{CE}$ graph.

KCL at the output node yields:

$$i_C + \frac{v_o - V_{CC}}{R_C} = 0 \implies I_C = \frac{V_{CC} - v_o}{R_C}$$

Notice that $v_o = V_{CE}$ and so:

$$\boxed{I_C = \frac{V_{CC} - V_{CE}}{R_C}}$$

We call this the load line and it can be viewed in the following diagram:

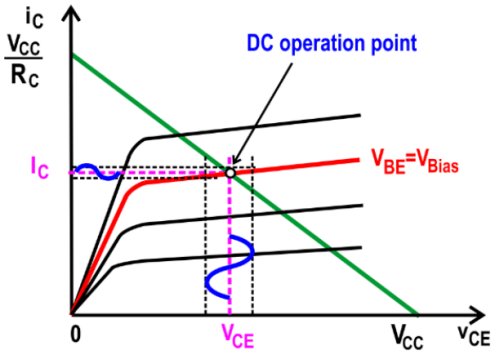
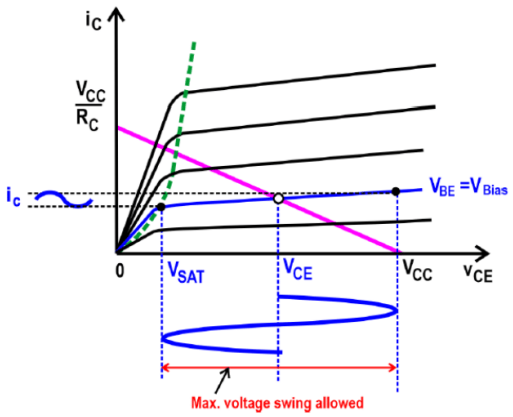


Figure 1: Load line.

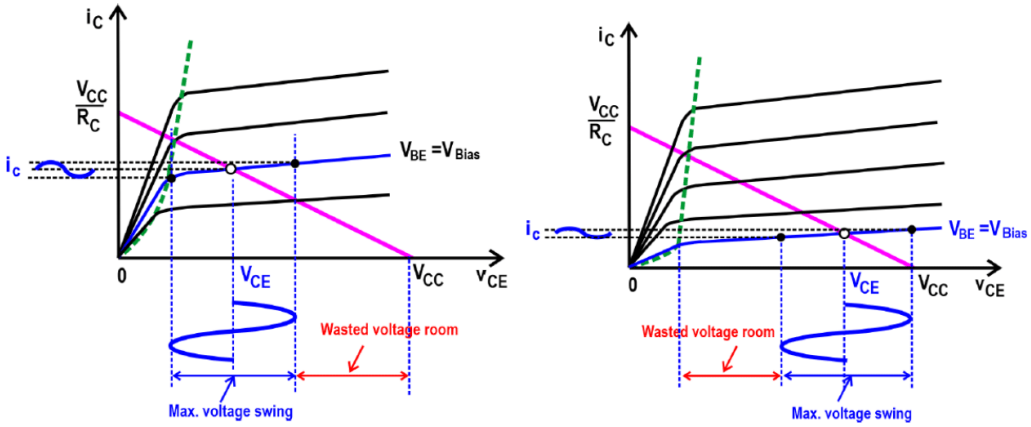
The intersection of the load line and the $i_C \sim V_{CE}$ curve with $V_{BE} = V_{Bias}$ defines the DC operating point. You can control this *DC* operating point by either adjusting the biasing voltage (which determine which curve we use) or by adjusting R_C. Typically this is done with the DC biasing voltage because adjusting the resistance affects properties of the circuit.

Now we want to determine where the best DC operating point is for a circuit. Recall that we need the amplitude of the output voltage signal to be small enough so that it does not pass into the saturation region of the BJT, furthermore it should not pass above V_{CC}.

For example the following choice for V_{CE} is optimal (the lowest is V_{sat} and the max is V_{CC}):



While the following two are either too close to the saturation region, or V_CC:



Since the optimal AC signal can vary at max between V_{SAT} and V_{CC}, then the maximum swing of the output signal v_o of an amplifier circuit is:

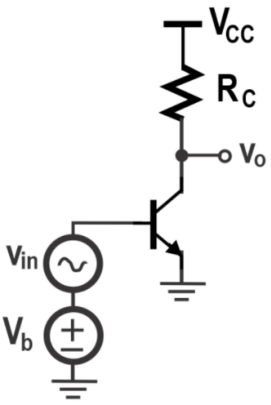
$$\boxed{v_{o,AC,max} = \frac{V_{CC} - V_{sat}}{2}}$$

Now how we actually set the V_{CE} to be at this point is done by controlling the R_C and V_{Bias}.

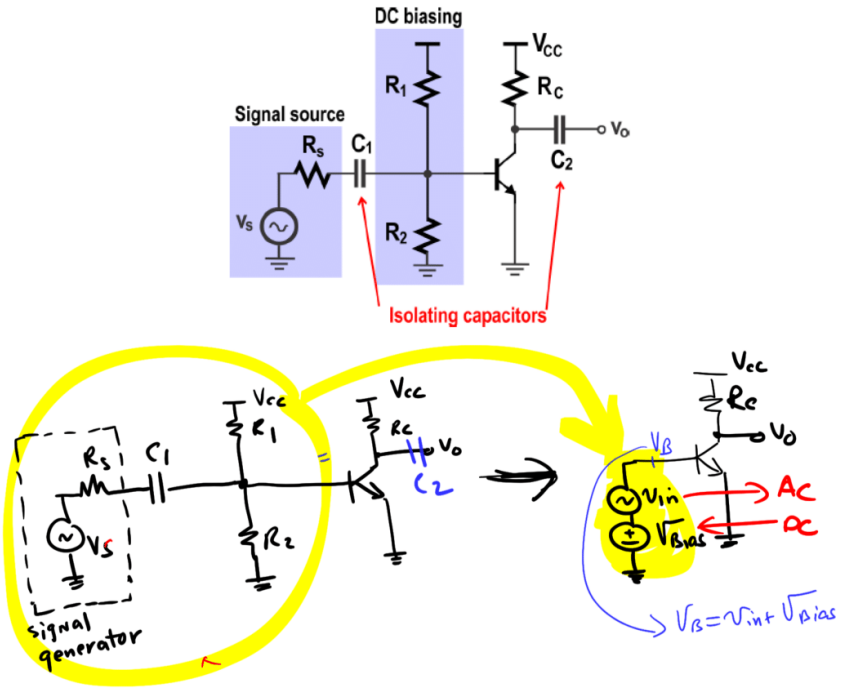
Common-Emitter (CE) Amplifiers

In this subsection we look at the CE configuration of the BJT amplifier, but we also look in detail at exactly how the biasing is done using a voltage divider.

The following circuit is the CE BJT amplifier circuit:



But practically, the circuit would be built like:



Consider the following remarks:

- The top of R_1 is connected to V_{CC}.
- The signal source is what is being amplified, it is required to be small compared to the DC biasing voltage as usual.
- R_s is the internal resistance of the signal source v_s.
- The two resistors, and connected in series as a voltage divider circuit, the values of R_1 and R_2 can be chosen to establish V_b (we will get to this later).
- The capacitors are called *isolating capacitors*. We will get to the calculations as to why they "*isolate*" below, but for now just remember they are large capacitors which block DC current from flowing into the signal source, and provides a low resistance path for the v_s to reach the BJT.

Recall that the impedance of a capacitor is:

$$\bar{Z}_C = \frac{1}{\hat{j}\omega C}$$

- At DC $\omega = 0 \rightarrow Z_c \rightarrow \infty$. Meaning at DC the capacitor is just an open circuit, and so it isolates it the two sides.
- At AC if we make C sufficiently large $Z_C \rightarrow 0$. Meaning at AC with a high enough capacitance, the signal passes through with no impedance.

Now we want to determine the DC operating point of the circuit, we do this by KCL at the base:

$$\frac{V_B}{R_2} + \frac{V_B - V_{CC}}{R_1} + I_B = 0$$

$$\frac{V_B}{R_2} + \frac{V_B - V_{CC}}{R_1} + \frac{I_C}{\beta} = 0$$

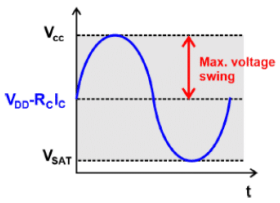
$$\frac{V_B}{R_2} + \frac{V_B - V_{CC}}{R_1} + \frac{I_s e^{\frac{V_{BE}}{V_t}}}{\beta} = 0$$

$$\boxed{\frac{V_B}{R_2} + \frac{V_B - V_{CC}}{R_1} + I_s \frac{e^{\frac{V_B}{V_t}}}{\beta} = 0}$$

Provided I_s and β, V_B can be determined.

Optimal V_B

We now want to determine the optimal value for V_B so that the output voltage has maximum swing. Consider the following graph of $V_o \sim t$:



Recall:

$$v_{o,AC,max} = \frac{V_{CC} - V_{sat}}{2}$$

So we want to set:

$$V_{CE} = V_C = V_{sat} + v_{o,AC,max}$$

$$\boxed{V_C = V_{sat} + \frac{V_{CC} - V_{sat}}{2} = \frac{1}{2}(V_{CC} + V_{sat})}$$

Since we also know:

$$V_C = V_{CC} - R_C I_C$$

... we get:

$$I_c = \frac{1}{2R_C}(V_{CC} - V_{sat})$$

$$I_s e^{\frac{V_{BE}}{V_t}} = \frac{1}{2R_C}(V_{CC} - V_{sat})$$

$$\boxed{V_B = V_t \operatorname{Ln}\left(\frac{V_{CC} - V_{sat}}{2I_s R_C}\right)}$$

Now we want to use this information to determine the values of R_1 and R_2 in the biasing circuit. There are two possibilities:

1. R_1 **and** R_2 **are small:** In this case we can neglect the base current as the current trough the divider will be large, therefore:

$$V_B = \frac{R_2}{R_1 + R_2} V_{CC}$$

$$\frac{R_2}{R_1 + R_2} V_{CC} = V_t \operatorname{Ln}\left(\frac{V_{CC} - V_{sat}}{2I_s R_C}\right)$$

This equation has infinite solutions for R_1 and R_2... any pair works.

2. R_1 **and** R_2 **are large:** In this case we cannot neglect the base current. We do KCL at the base:

$$\frac{V_B}{R_2} + \frac{V_B - V_{CC}}{R_1} + I_B = 0$$

$$\frac{V_B}{R_2} + \frac{V_B - V_{CC}}{R_1} + \frac{I_C}{\beta} = 0$$

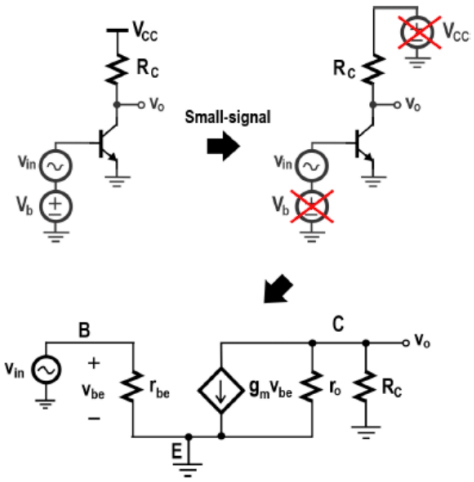
$$V_B = \frac{G_1 V_{CC} - \frac{I_C}{\beta}}{G_1 + G_2}$$

$$\frac{G_1 V_{CC} - \frac{I_C}{\beta}}{G_1 + G_2} = V_t \,\mathrm{Ln}\left(\frac{V_{CC} - V_{sat}}{2 I_s R_C}\right)$$

This equation has infinite solutions for R_1 and R_2... any pair works.

Operation of the CE Amplifier

We are now ready to see the amplification property of the CE BJT configuration. Given the following circuit:



We can do KCL at the output node to get:

$$G_C v_o + g_o v_o + g_m v_{in} = 0$$

Therefore the voltage gain (A_v) is:

$$\boxed{A_v = \frac{v_o}{v_{in}} = -g_m(r_o \parallel R_C)}$$

Note that:

- CE is an inverting amplifier (signals will be 180° out of phase)
- A_v goes up with g_m
- A_v goes up with R_c

Input/Output Resistance of the BJT CE Amplifier

Recall that to find the input/output resistance in a port we:

1. Remove all independent sources.
2. Apply a test voltage V_x at the input/output node.
3. Find the corresponding i_x.

4.

$$R_{eq} = \frac{V_x}{i_x}$$

Using this method:

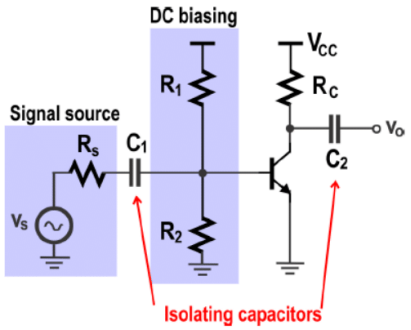
$$R_{in} = r_{be}$$
$$R_{out} = r_o \parallel R_C$$

Superposition to Find V_B

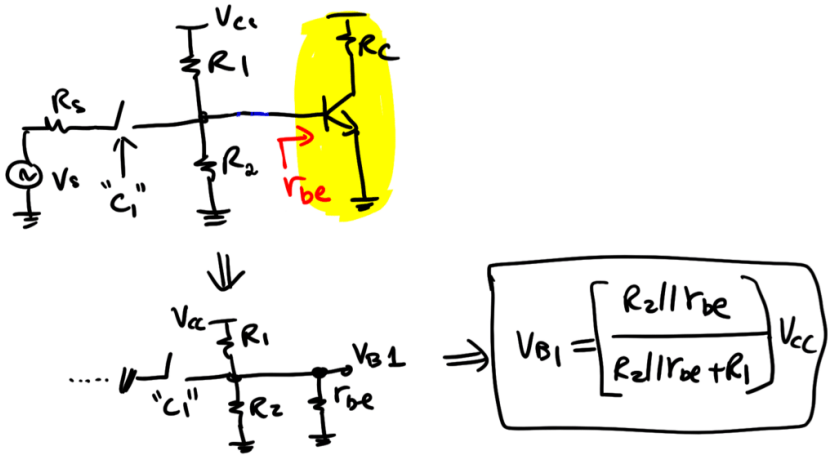
Recall:

$$V_B = V_{Bias} + V_{in}$$

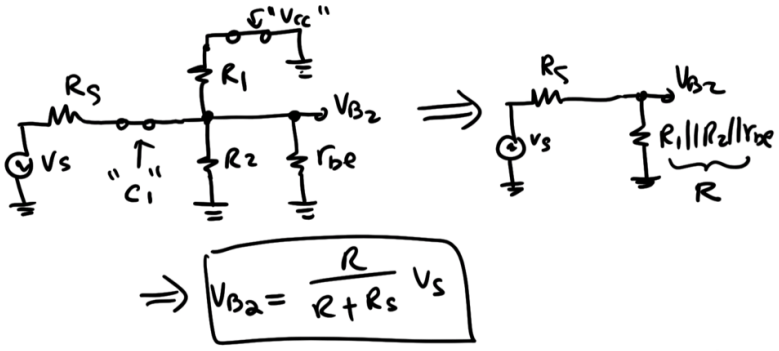
We can actually see this in action by using superposition on the following circuit, and knowing the input resistance:



1. **Consider only V_{CC}:**



2. **Consider only V_s:**



As you can see V_B is made of two components. Note:

- We want $R >> R_s$ so that $V_{B2} = V_{in} = V_s$. To make this happen, we need R_1 and R_2 to be in the $k\Omega$ range, as that is the range of r_{be}.
- If R_1 and R_2 are too small, the biasing circuit will consume too much DC power.

Active Load

So far we have considered only the CE amplifier with a resistive load R_C. From this we derived the voltage gain of:

$$A_v = -g_m(r_o \parallel R_C)$$

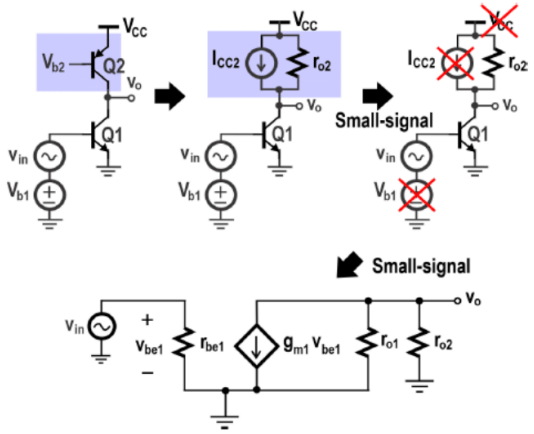
... since $R_{out} = r_o \parallel R_C$:

$$A_v = -g_m R_{out}$$

There are thus two ways to increase the gain of a CE BJT amplifier:

- Increase g_m which would mean increasing I_c causing more DC power consumption.
- Increase R_{out} by increasing R_C which is costly in terms of money and space on a chip.

The solution is to use another BJT as the load and use its internal resistance to augment the amplification of the original BJT.



In this configuration note the following:

- Q1 operates in active mode in order to amplify the input signal using g_m.
- Q2 operates in active mode for its large r_{o2}.
- We assume V_{B2} is fixed to some constant amount which biases Q2 in active mode. Therefore the CCCS is has a constant independent value of I_{CC2}.
- We do not show the base current for Q2 as it is not relevant to the amplifier and it is constant.

- In small-signal the second BJT entirely acts like a resistor with a large resistance r_{o2}.
- Q2 is a pnp BJT (explained below).

Evidently, the new voltage gain is the same except now $R_C = r_{o2}$:

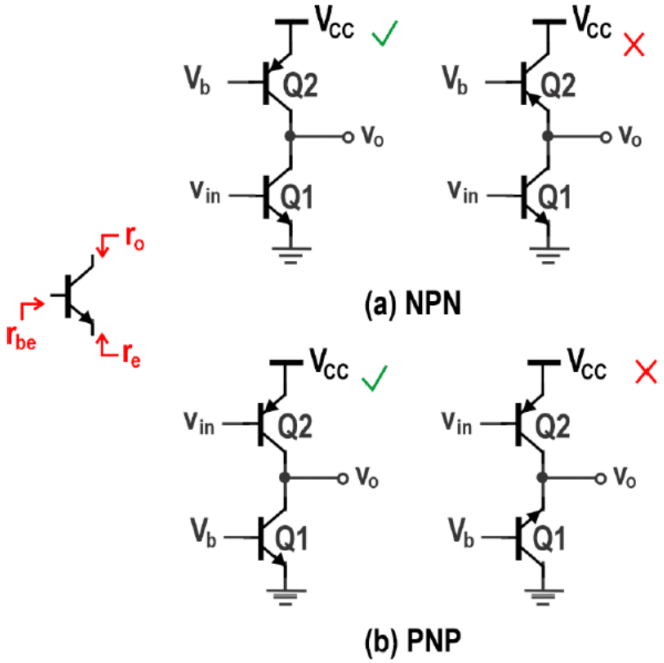
$$A_v = -g_m(r_{o1} \parallel r_{o2})$$

Since both resistances in parallel are large, we get a large voltage gain without using a large resistor. The only cost to doing this is that you must bias Q2 as well.

Once again:

$$R_{out} = r_{o1} \parallel r_{o2}$$
$$A_v = -g_m R_{out}$$

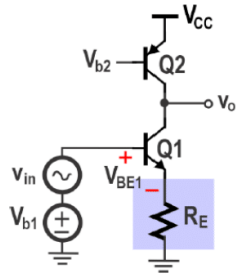
Now, why did we use an npn and then a pnp? The idea is that from the perspective of the output voltage, in both directions, we must be seeing r_o as the equivalent resistance to get the gain we want:



If you look into the wrong end of the BJT, you will get r_e which is typically less than $1k\Omega$.

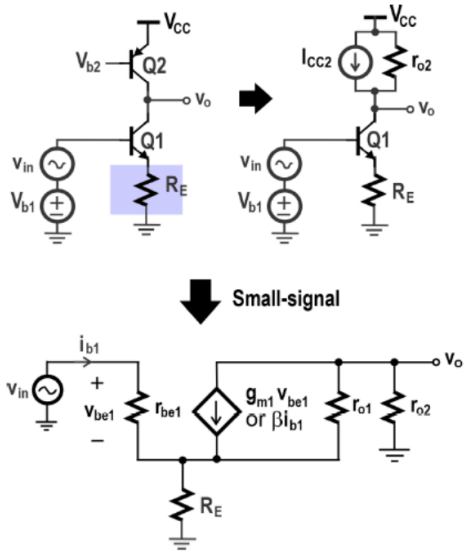
Emitter Degeneration

Emitter degeneration is a technique used to stabilize I_{c1} through means of a negative feedback loop, the following is the circuit and the feedback loop:



$$V_{b1} \uparrow \rightarrow V_{BE1} \uparrow \rightarrow I_{c1} \uparrow \rightarrow V_{R_E} = R_E I_{c1} \uparrow \rightarrow V_{BE1} \downarrow \rightarrow I_{c1} \downarrow$$

Let's look at the effect this has on input resistance:



Using the fact that v_{in} is the sum of the voltage drops across r_{be1} and R_E we get:

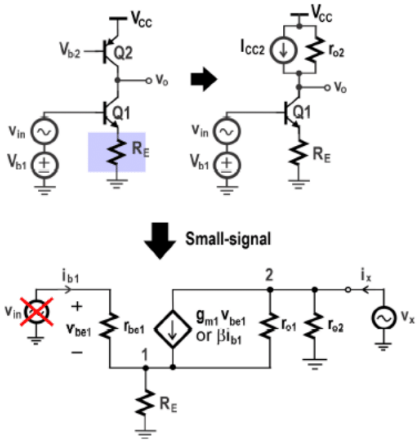
$$R_{in} = r_{be1} + g_{m1}R_E r_{be1} + R_E \approx (1 + g_{m1}R_E)r_{be1}$$

As you can see input resistance from the perspective of the source has increased. The voltage gain is:

$$A_v \approx \frac{-g_m(r_{o1} \parallel r_{o2})}{1 + g_m R_E}$$

Which you can think of at the voltage gain from before, but now due to the emitter degeneration we have another term in the denominator. The larger the R_E the more loss of voltage gain.

Let's look at the effect this has on input resistance:



$$R_{out} = (r_{o1} \parallel r_{o2})(1 + g_{m1}R_E)$$

Which shows that emitter degeneration increases output resistance.

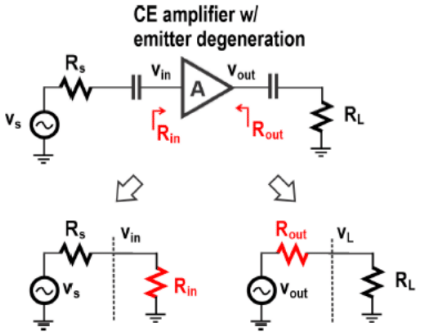
There are four characteristics of CE amplifiers which use emitter degeneration:

- Increased input resistance. This is desirable as it reduces the loading effect of the amplifier on the input source.
- Reduced voltage gain.

- **Increased output resistance.** This is understandable since it worsens the loading effect of the load resistor on the gain.
- Reduced power consumption.

Note that the loading effect is how much the actual value of the load affects the properties of the amplifier.

We can think of this amplifier element as just a two terminal circuit element with input resistance and output resistance as noted above:



Note that:

$$v_{in} = \frac{R_{in}}{R_S + R_{in}} v_s \approx \left(1 - \frac{R_S}{R_{in}}\right) v_s$$

Meaning we want a high input resistance so that $v_{in} \approx v_s$. Since emitter degeneration increases input resistance, this is one of the reasons we use it.

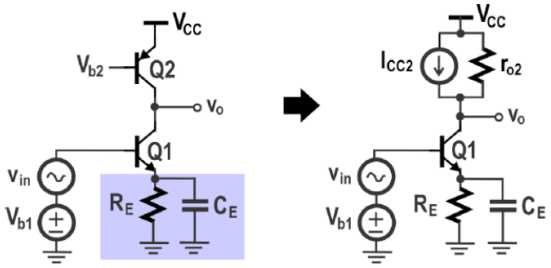
Similarly note that:

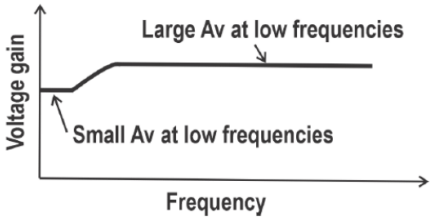
$$V_L = \frac{R_L}{R_L + R_{out}} v_{out} \approx \left(1 - \frac{R_{out}}{R_L}\right) v_{out}$$

Meaning we want a low output impedance so that $V_L \approx v_{out}$. Emitter degeneration lowers the voltage delivered to the load resistor R_L.

Frequency Dependent Emitter Degeneration

We would like some way to preserve the voltage gain at some specified frequency which we would have without the emitter degeneration, while still using some of its benefits for DC. We do this by placing as shunt capacitor which behaves as a short circuit when an AC signal passes through it:



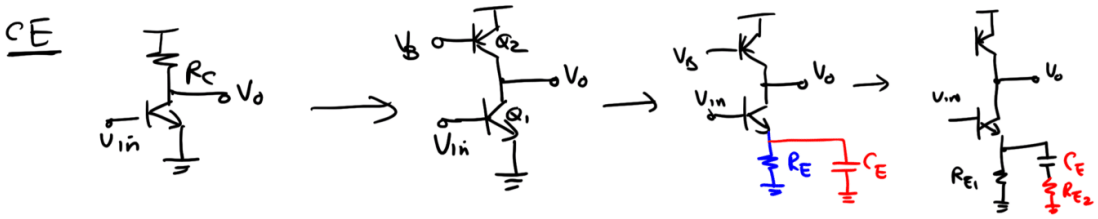


At low frequencies, C_E behaves as an open circuit and so the lowered gain is present because of the emitter degeneration resistor. At high frequencies that resistor is being shorted and so we have no emitter degeneration. DC power consumption also goes down.

This means that:

- Emitter degeneration is active at DC.
- Emitter degeneration is inactive at AC.

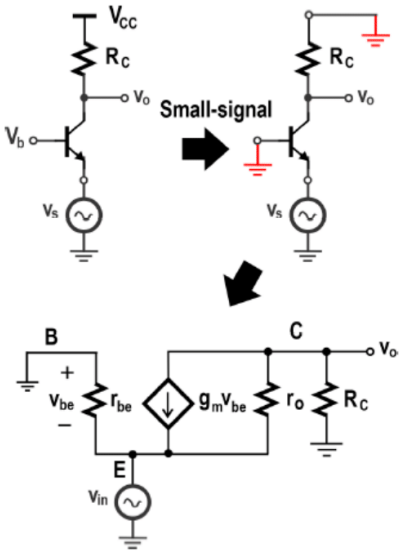
Sometimes (like in lab 5) we also put a resistor passed that shunt capacitor to have *some* emitter degeneration at AC. The following summarizes all the ways we have discussed to use a CE BJT amplifier:



If you make $R_{E2} << R_{E1}$ then at AC the emitter degeneration is approximately just R_{E2}.

Common-Base (CB) Amplifiers

The following is the general configuration of a CB BJT amplifier, with its small signal equivalent circuit:



The base is held at a constant DC biasing voltage and functions as a small signal ground. This biasing voltage must maintain the BJT in active mode for all v_{in} values. The signal is placed into the emitter of the BJT.

To find the voltage gain of this circuit, we do KCL at the output node:

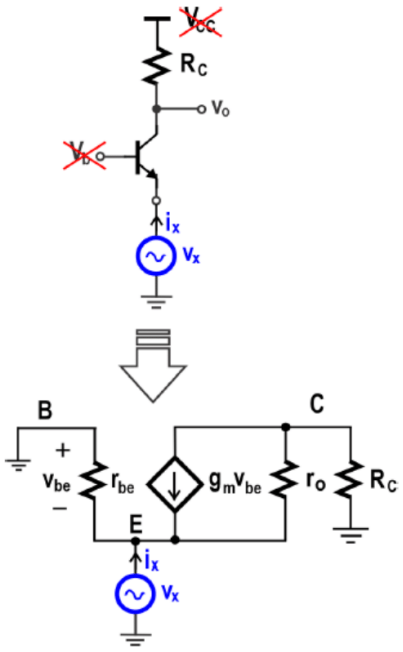
$$\frac{v_o}{R_C} + \frac{v_o - v_{in}}{r_o} - g_m v_{in} = 0$$

$$\boxed{A_v = \frac{v_o}{v_{in}} \approx g_m(r_o \parallel R_C)}$$

Note that the voltage gain in a CB amplifier is the same as in a CE amplifier, however now the amplifier is non-inverting.

Input/Output Impedance of CB Amplifier

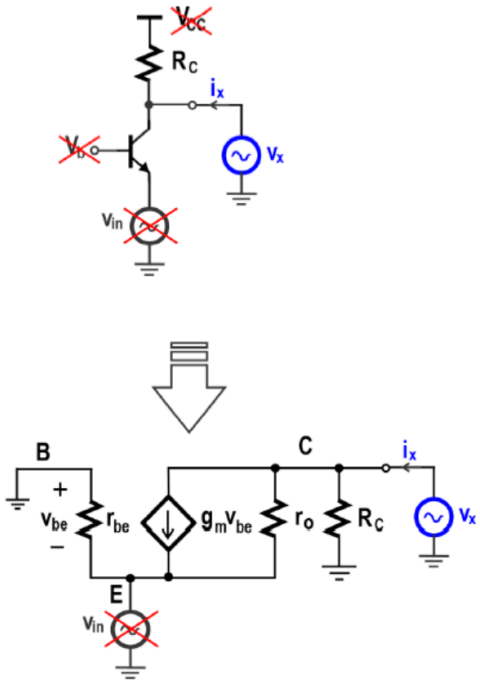
We can find the input impedance in the same way we did before:



It can be shown that:

$$\boxed{R_{in} \approx r_e}$$

Similarly for output impedance:



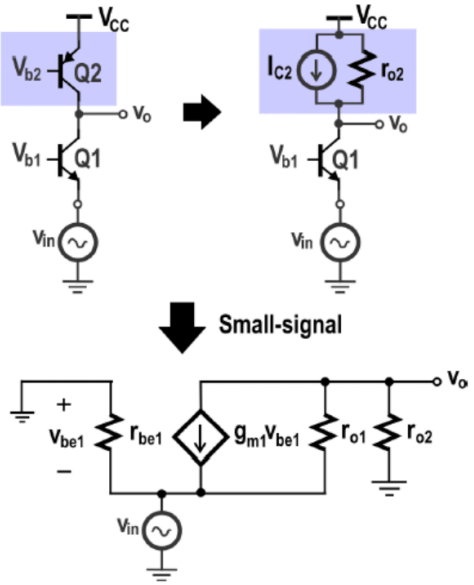
It can be shown that:

$$\boxed{R_{out} = r_o \parallel R_C}$$

Evidently, CB amplifiers have the same output impedance, but much smaller input impedances.

CB Amplifier with Active Load

You can also use an active load on a CB amplifier to increase the gain:

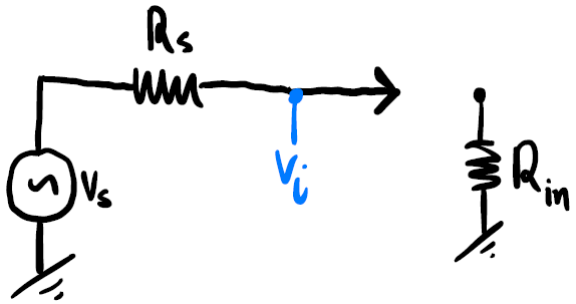


In which case:

$$A_v = g_{m1}(r_{o1} \parallel r_{o2})$$

CE amplifiers have much larger input impedances that CB amplifiers which means CB amplifiers should not be used as voltage amplifier circuits in general.

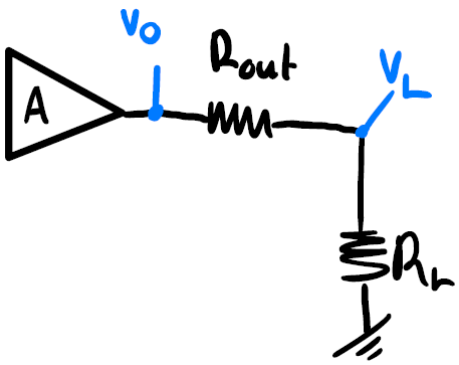
Why do we care about input impedance? The reason is for the following circuit:



We see that the input impedance is like a resistor in series with the source resistance, and to ground. We want $v_s = v_i$ but this is not possible since there must be internal resistance from the signal source. To determine the ratio of $\frac{v_i}{v_s}$ we can use voltage division:

$$\frac{v_i}{v_s} = \frac{R_{in}}{R_{in} + R_s} \approx 1 \text{ (for } R_{in} >> R_s)$$

Why do we care about output impedance? The output impedance affects how much of the output signal from the amplifier gets to the load. This depends on the load resistance R_L. Consider the following circuit:



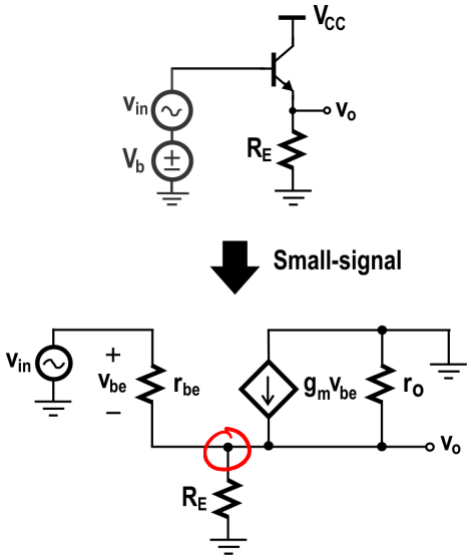
We want $V_L = V_{out}$, but there are two notes about this:

- If $R_L >> R_{out}$ then:
$$\frac{V_L}{V_{out}} = \frac{R_L}{R_L + R_{out}} = \left(1 - \frac{R_{out}}{R_L}\right) \approx 1$$
- If $R_L < R_{out}$ then:
$$\frac{V_L}{V_{out}} = \frac{R_L}{R_L + R_{out}} = \left(1 - \frac{R_{out}}{R_L}\right) < 1$$
This means we loose some of the output signal before it gets to the load. This can be fixed by increasing the output resistance using the following type of circuit.

CC Amplifiers

In this subsection we discuss common collector BJT amplifiers.

Here, the collector is connected to constant voltage (AC small signal ground).

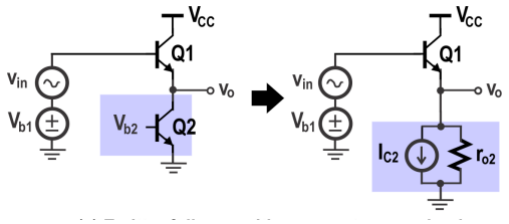


By doing KCL at emitter and some approximations, it can be shown that:

$$A_v \approx 1$$

In a sense, the output (emitter) follows the input, giving this circuit the name *Emitter Follower*.

This fact is also true if you use an active load for this amplifier:

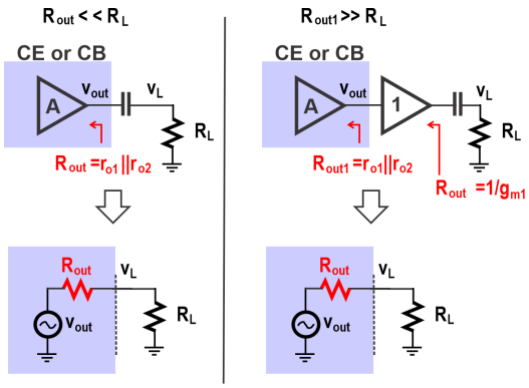


For both resistor load CC amplifiers, and active load CC amplifiers, we have:

$$R_{out} = \frac{1}{g_{m1}}$$

(Where g_{m1} is the transconductance of the BJT getting the input signal).

The main purpose of this circuit is to lower the output resistance of an amplifier circuit when dealing with small loads:



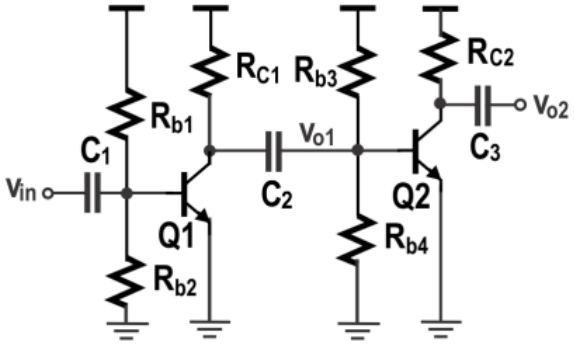
What you do is you add the CC amplifier after either a CE or CB amplifier to lower the output resistance for the load. This is called cascaded amplifiers or multi-stage amplifiers.

Multi-Stage Amplifiers

Multistage amplifiers are where you combine one or more amplifiers in a row between your input signal and output node. There are a few notes about doing this:

- Each stage must be properly DC biased such that the BJTs are in active mode and the circuit achieve the correct gain.
- Isolation capacitors are used between stages to isolate the amplifiers to DC bias each stage independently.
- Since the stages are isolated, the first stage loads the input signal, and the second stage load the output from the first stage, and so on.

For example, the following are two CE amplifiers cascaded together:

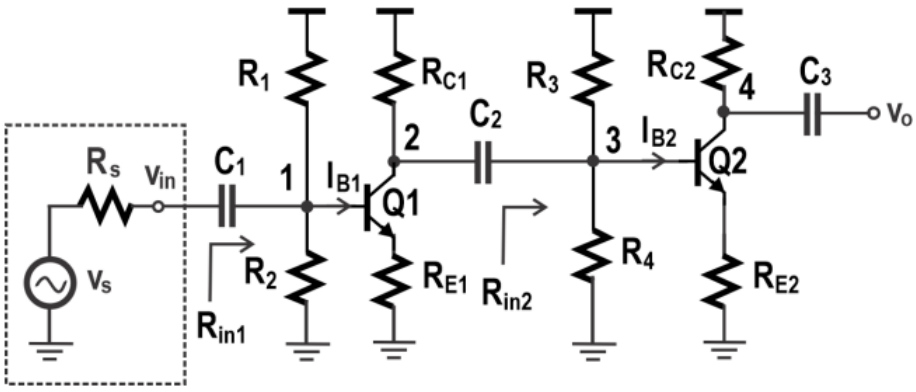


We can determine DC biasing conditions by solving the following $4-$variable system:

- KCL at node 1:
$$g_1(V_1 - V_{cc}) + g_2 V_1 + I_{B1} = 0$$
- KCL at node 3:
$$g_3(V_3 - V_{cc}) + g_4 V_3 + I_{B2} = 0$$
- KVL around the base-emitter of Q1:
$$V_1 = V_{BE1} + R_{E1}(1+\beta) I_{B1}$$
- KVL around the base-emitter of Q2:
$$V_3 = V_{BE2} + R_{E2}(1+\beta) I_{B2}$$

Analysis of a Multistage Amplifier

We will use the following circuit as an example of how to analyse a multistage amplifier:



We want to determine A_{vs} for this circuit which is defined as:

$$A_{vs} \equiv \frac{v_o}{v_s}$$

Looking at the circuit we can see that there are two amplifiers, and so the gain must be the product of the individual gains between nodes:

$$A_{vs} = \frac{v_1}{v_s} \cdot \frac{v_3}{v_1} \cdot \frac{v_o}{v_3} = \frac{v_1}{v_s} \cdot A_{v1} \cdot A_{v2}$$

We can find each of these gain factors independently:

- $$\frac{v_1}{v_s} = \frac{R_{in1}}{R_{in1} + R_s}$$
- For A_{v1} we use the formula for a CE amplifier with load and emitter degeneration:

$$A_{v1} = \frac{-g_{m1}(R_{c1} \parallel R_{in2})}{1 + g_{m1}R_{E1}}$$

- For A_{v2} we use the formula for a CE amplifier with no-load and emitter degeneration:

$$A_{v2} = \frac{-g_{m2}(R_{c2})}{1 + g_{m2}R_{E2}}$$

Therefore the voltage gain from the source A_{vs} is:

$$A_{vs} = \frac{v_1}{v_s} \cdot A_{v1} \cdot A_{v2} = \frac{R_{in1}}{R_{in1} + R_s} \cdot \frac{-g_{m1}(R_{c1} \parallel R_{in2})}{1 + g_{m1}R_{E1}} \cdot \frac{-g_{m2}(R_{c2})}{1 + g_{m2}R_{E2}}$$

Current Mirrors

Module 4: MOSFET

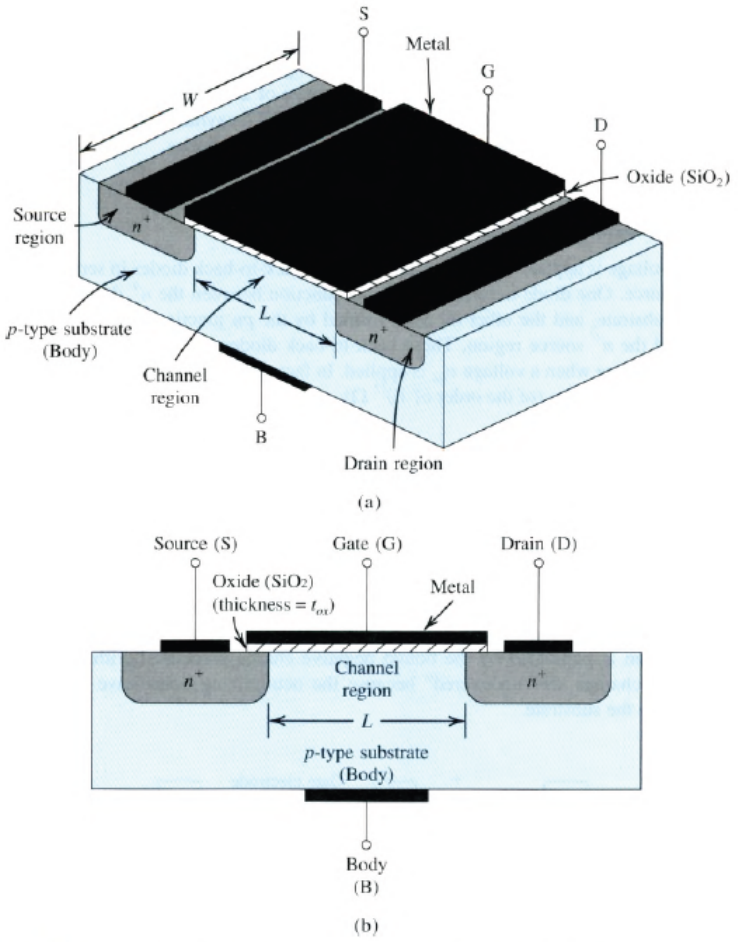
Metal Oxide Semiconductor Field Effect Transistors (MOSFETs) are another type of transistor we study in this course. MOSFETs are used in digital logic and operate similar to a BJT, as in they use the voltage between two terminals to control the current flowing in the third terminal. MOSFETs are actually four terminal devices, however one terminal can usually be omitted (more on this later).

MOSFETs can be made quite small compared to BJTs, are simpler to manufacture, and require less power. In most applications MOSFETs are the preferable transistor to use.

In this course we discuss the *enhancement-type MOSFET* (hereby referred to a MOSFET). There are two types of MOSFETs: PMOS and NMOS (very similar to NPN, and PNP for BJTs).

NMOS Transistors

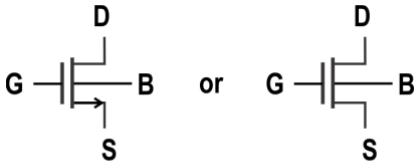
The following is the general structure of a NMOS transistor:



The NMOS has 4 terminals:

1. **Source (S)** is a heavily doped $n-type$ region which has a conductive metal contact.
2. **Drain (D)** is a heavily doped $n-type$ region which has a conductive metal contact. The drain is typically held at a higher potential than the source.
3. **Gate (G)** is a metal contact which is separated from the substrate by an **Oxide** which is a thin insulator blocking all current flow (in reality current is not 0 however for this course we assume no current).
4. **Substrate (B)** also called the **Bulk** or **Body** is a lightly doped $p-type$ connected to the ground via a back plate. For all purposes in this course, the bulk is grounded and so it can be omitted from circuit diagrams as it is irrelevant.

There are two circuit diagrams for the NMOS transistor, remember that in both cases you can omit the B terminal:

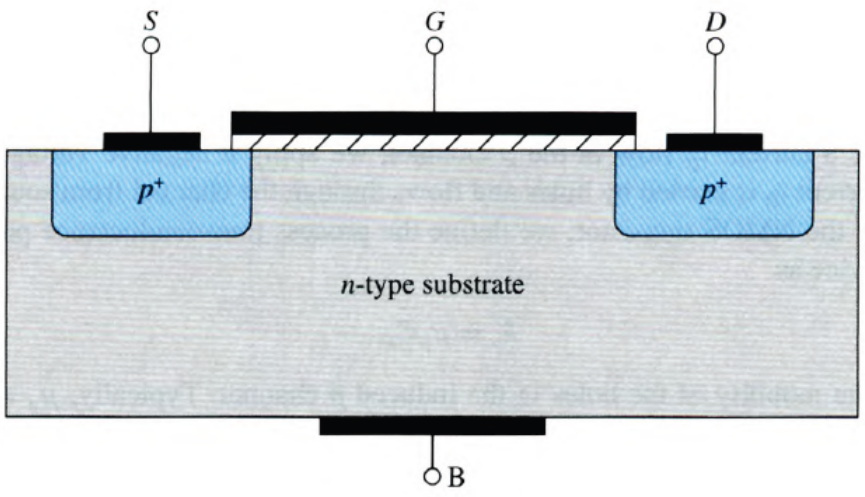


(a) NMOS

Note that the transistor is named N-MOS since the two highly doped regions are $n-type$. This type of transistor is also called **n-channel MOSFET** because of how it operates (more on operation later).

PMOS Transistors

The following is the general structure of a PMOS transistor:

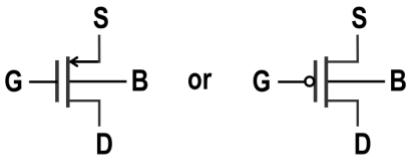


(a)

The PMOS has 4 terminals:

1. **Source (S)** is a heavily doped $p-type$ region which has a conductive metal contact. The source is typically held at a higher potential than the drain.
2. **Drain (D)** is a heavily doped $p-type$ region which has a conductive metal contact.
3. **Gate (G)** is a metal contact which is separated from the substrate by an **Oxide** which is a thin insulator blocking all current flow (in reality current is not 0 however for this course we assume no current).
4. **Substrate (B)** also called the **Bulk** or **Body** is a lightly doped $n-type$ connected to the V_{CC} (power supply) via a back plate. For all purposes in this course, the bulk is connected to the power supply and so it can be omitted from circuit diagrams as it is irrelevant.

There are two circuit diagrams for the PMOS transistor, remember that in both cases you can omit the B terminal:

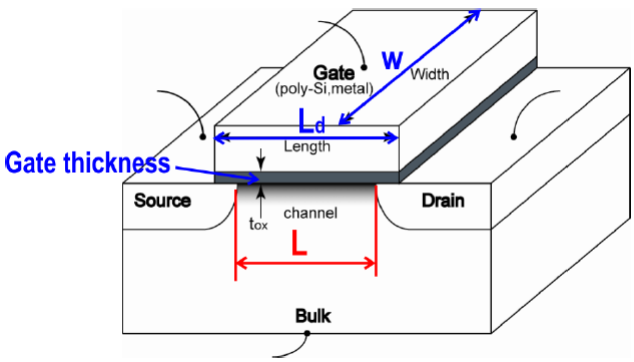


(b) PMOS

Note that the transistor is named P-MOS since the two highly doped regions are $p-type$. This type of transistor is also called **p-channel MOSFET** because of how it operates (more on operation later).

Dimensions of a MOSFET

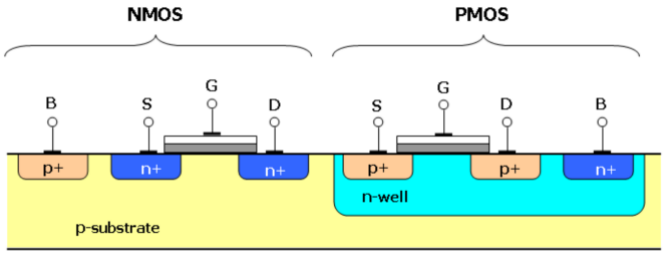
We have some standard ways to measure the dimensions of a MOSFET, all of which apply to both PMOS and NMOS. These dimensions affect the current through the terminals.



- Drawn Length (L_d).
- Effective length (L or L_{eff}).
- Width (W).
- Gate Oxide thickness (t_{ox}).

Now is a good time to note that the MOSFET is entirely symmetrical and so it is arbitrary which one is called the source and which one is called the drain.

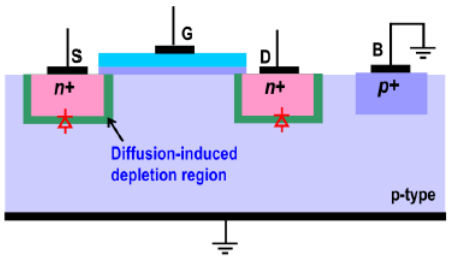
Directly comparing NMOS and PMOS transistors, we can manufacture them in the same piece of silicone. The PMOS transistor needs to be placed within a *n-well* as seen in the following diagram:



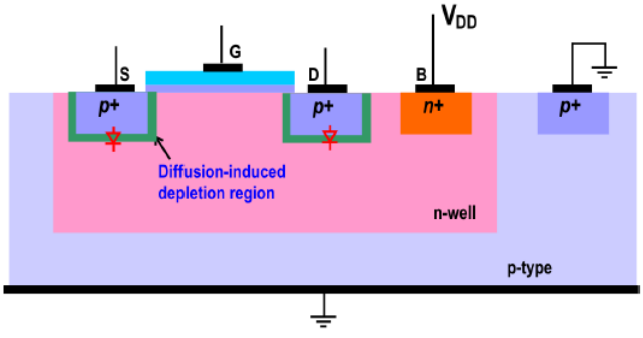
Notice that we have a backplate separate from the **body** terminal. In reality there are many **body** terminals all over the silicon that the MOSFET is made from, and the back plate it always grounded.

Operation of MOSFETs

To begin, recall that a pn-junction is formed between a $p-type$ and an $n-type$ semiconductor region. From this we are able to see that *diffusion-induced diodes* with *diffusion-induced depletion regions* are formed at the source and the drain with the substrate:



(a) NMOS



(b) PMOS

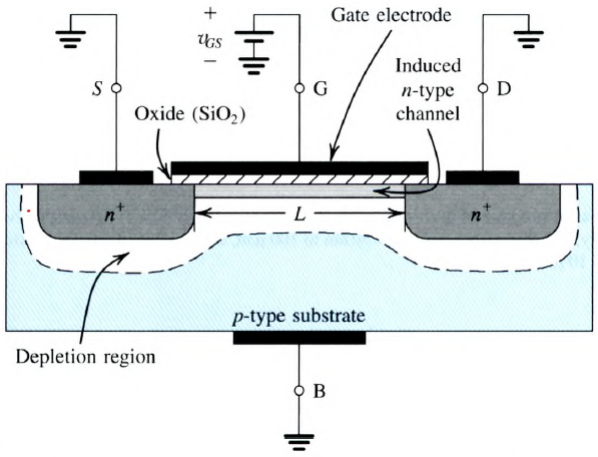
It is important for these diodes to be **reverse-biased** at all time to prevent the majority current carriers from migrating from the source to the **body**. We can do this by connecting the source and body by a short circuit. For this reason we treat the MOSFET as a three terminal device as the body has no effect on the operation. Notice that there are many ground around the **substrate** and that the **body** of the PMOS is to V_{DD}.

From here, we will explain in detail the operation of an $NMOS$ transistor, and then afterwards discuss $PMOS$.

Operation of NMOS Transistors

Positive Voltage at Gate

We can apply a positive voltage to the gate as seen in the following diagram:

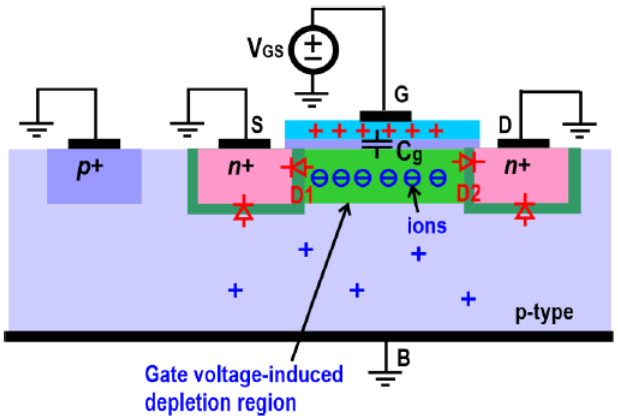


The positive voltage on the gate causes the free holes to be repelled from the area of the substrate just under the gate. These holes are pushed down into the substrate, leaving behind a carrier-depletion region populated by bound negative acceptor atoms.

Further, the positive gate voltage attracts electrons from the drain and source (as they are heavily doped) and electrons fill the depletion region. When a sufficient number of electrons accumulate near the surface of the substrate under the gate, and $n-type$ channel is effectively created connecting the source and the drain. We will see later that if you apply a voltage difference between the source and drain a current will flow through the channel.

We call the gate voltage v_{GS} as the source is usually grounded. The value of v_{GS} as which a sufficient number of mobile electrons accumulate under the gate to form a conducting channel is called the **threshold voltage** V_t or V_{tn}. For an NMOS the threshold voltage is positive. This voltage typically lies between $0.3V$ and $1.0V$. You can think of the threshold voltage as the voltage required to forward bias the source-substrate and drain-substrate diodes **only** in the region directly underneath the gate.

The gate and the channel form a **parallel-plate capacitor** with the oxide layer acting as the insulator. The positive gate voltage causes positive charge to accumulate on the top plate, and the corresponding negative charge accumulates in the channel, and so we have an electric field in the vertical direction. This field controls the charge in the channel, and in turn the current which is why these transistors are called *field effect*. The capacitor can be seen in the following diagram:



The capacitance can be described by:

$$C_g = C_{ox}(WL)$$

Where ($\epsilon_{ox} = 3.45 \times 10^{-11} F/m$):

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$$

Evidently, the dimensions of the oxide plate affect the capacitance of the transistor.

When $v_{GS} > V_t$ the diodes under the gate region become forward biased and the channel is formed. This is also referred to as the inversion layer. If no drain voltage is applied, then the voltage in the inversion layer is uniform.

Zero Voltage at Gate

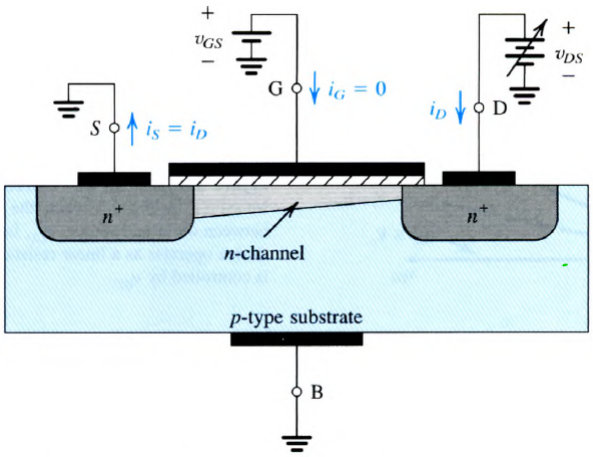
When source, drain, and gate are grounded all diodes are reverse biased, and the circuit acts as a resistor with very high resistance ($\approx 10^{12}\Omega$). Both diodes are reverse biased, and no channel is created.

If $v_{GS} < V_t$ then the MOSFET will enter **cut-off mode**. This mode is characterised by:

$$i_{DS} = 0$$

Positive Gate Voltage + Applied Drain Voltage

Given a $v_{GS} > V_t$ we can apply a voltage at the drain v_{DS} which will force electrons to move from the source to the drain. On the source side of the channel, the potential difference is v_{GS}, however now the potential is not uniform throughout the channel as on the drain side it is smaller (because we have increased the potential at the drain so the *difference* is smaller) this is shown in the following diagram:



We say that the channel is **tapered**, as in it is thicker at the source and thinner at the drain. In this case the NMOS is operating in **triode mode**. We are assuming here that the the potential at the drain is not enough to bring the difference between the gate and the drain to less than V_t near the drain which would *pinch-off* the inversion channel. In symbol that assumption is (for triode):

$$v_{GD} - V_t > 0$$

Which is held when:

$$\text{NMOS Triode: } v_{DS} < V_{GS} - V_t$$

If v_{DS} were to increase past this upper limit we would enter the pinch-off mode, and then saturation mode of the MOSFET (more on this later).

We want to know the current through the channel in *triode mode*. We do this using the equation:

$$i_{DS} = \frac{Q}{L} \cdot v_n$$

Which is the charge per unit length times the velocity of free electrons. The charge Q can be found by using the *average* voltage throughout the channel and the capacitance:

$$C_g = \frac{Q}{V} \rightarrow Q = C_g \cdot \Delta V$$

$$Q = C_g \left(\frac{1}{2}(v_S + v_D) \right) = C_g \left(v_{GS} - V_t - \frac{1}{2} v_{DS} \right)$$

Since we want the charge per unit length we just divide by L:

$$\frac{Q}{L} = \frac{C_g \left(v_{GS} - V_t - \frac{1}{2} v_{DS} \right)}{L}$$

Now for the velocity:

$$v_n = \mu_n \vec{E}$$
$$v_n = \mu_n \frac{v_{DS}}{L}$$

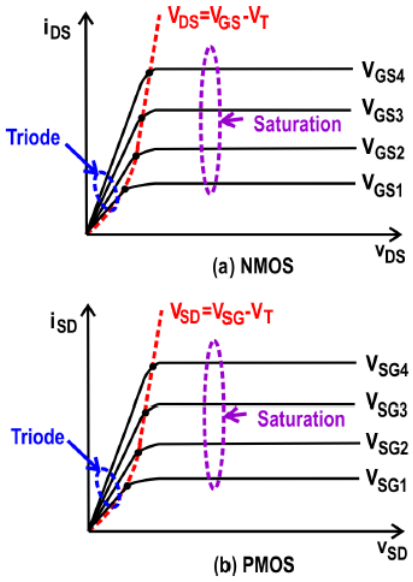
So all together, after taking the product and simplifying, we get (for NMOS):

$$\boxed{i_{DS,triode} = \mu_n C_{ox} \frac{W}{L} \left((v_{GS} - V_t) v_{DS} - \frac{1}{2} v_{DS}^2 \right)}$$

... and although we will discuss it more later, the following is the same equation for a PMOS transistor:

$$\boxed{i_{DS,triode} = \mu_n C_{ox} \frac{W}{L} \left((v_{SG} - |V_{tp}|) v_{SD} - \frac{1}{2} v_{SD}^2 \right)}$$

The following is the graph of this current function:



As you can see the current reaches a maximum and then flattens out. This region before the flattened area is what we are discussing right now as the *triode mode*.

Channel Conductance

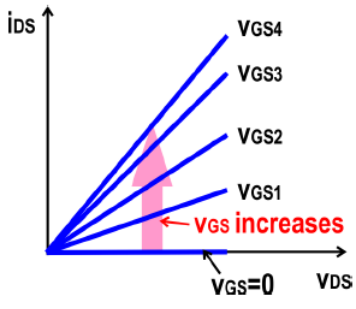
Incremental channel conductance is defined as:

$$g_{DS} \equiv \frac{di_{DS}}{dv_{DS}}_{v_{GS}}$$

In triode mode:

$$g_{DS} \approx \mu_n C_{ox} \frac{W}{L} (v_{GS} - V_t)$$

Which is notably only dependent on manufacturing parameters and v_{GS} meaning in triode mode the device operates as a variable resistor whose resistance varies with v_{GS}:



A few notes about channel conductance:

- Increasing W increases channel conductance.
- Increasing v_{GS} increases channel conductance.
- g_{DS} is linear with v_{GS} meaning the device acts as a linear resistor in triode mode with voltage controlled conductance.

Operations of NMOS Transistors Continued

Pinch-off Mode

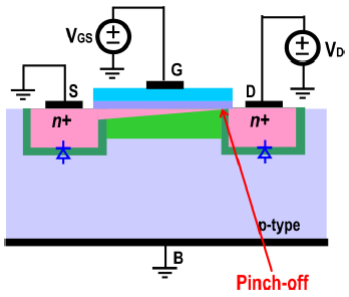


Figure 20: Pinch-off.

When v_{DS} is sufficiently large (recall S is grounded so we mean the voltage at the drain) then $V_{GD} = V_t$ may occur. At this point at the drain there is just barely enough potential difference to cause inversion. In fact at the drain there is no inversion. The condition for this to happen is:

$$\text{NMOS Pinch-off: } v_{DS} = v_{GS} - V_t$$

To be clear, **current still flows**. This is because while there is no *inversion* near the drain, the is *depletion*, meaning the region is still depleted of its holes, however there is not an effective $n - type$ region. This is fine for current because the electrons are accelerated through the channel and just have to make it through a short *depletion-but-not-inversion* region. The resistance of this region is negligible.

Using a very similar method to what was used before to determine current, we find that (for NMOS):

$$\boxed{i_{DS,\text{pinch-off}} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(v_{GS} - V_t)^2}$$

To increase the current at *pinch-off* you can either increase W or v_{GS}.

From this equation we also have:

$$g_m = \frac{di_{DS}}{dv_{GS}}_{DC} = \mu_n C_{ox}\frac{W}{L}(v_{GS} - V_t) = \frac{2I_{DS}}{v_{GS} - V_t}$$

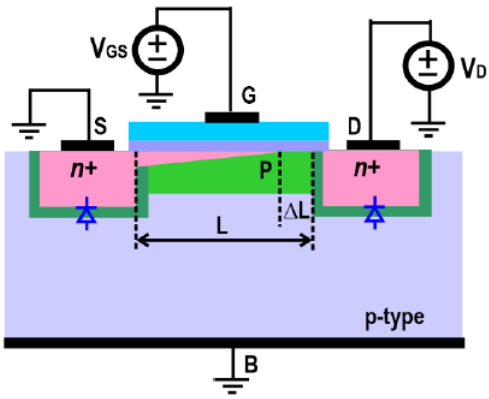
Saturation Mode

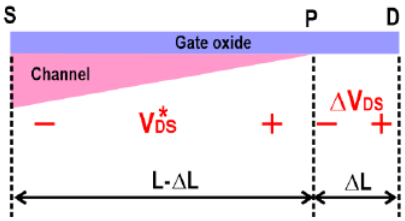
If you continue to raise v_{DS} past pinch-off (v^*_{DS}) to some value:

$$v_{DS} = v^*_{DS} + \Delta v_{DS}$$

... then the pinch off point will move towards the source from the drain as more and more of the depletion region will be *un*-inverted. Effectively the channel length is decreasing to:

$$L_{eff} = L - \Delta L$$





Using a very similar method to before, we can find that:

$$i_{DS,\text{sat}} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(v_{GS} - V_t)^2\left(1 + \frac{2\Delta L}{L}\right)$$

We know however that:

$$\frac{\Delta L}{L} \propto v_{DS}$$

We use the channel length modulation coefficient (λ) to describe this relationship, so we say:

$$\lambda v_{DS} = \frac{2\Delta L}{L}$$

So:

$$\text{NMOS Saturation: } \boxed{i_{DS,\text{sat}} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(v_{GS} - V_t)^2\left(1 + \lambda v_{DS}\right)}$$

This coefficient is there to quantify the impact of channel length modulation. We can rewrite this equation as (by distributing):

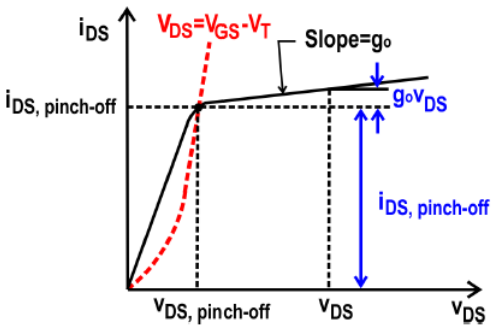
$$i_{DS,\text{sat}} = i_{DS,\text{pinch-off}} + i_{DS,\text{pinch-off}}\lambda v_{DS}$$

We can define:

$$g_o = i_{DS,\text{pinch-off}}\lambda$$

Then we can think of the saturation current as:

$$\text{NMOS Saturation: } \boxed{i_{DS,\text{sat}} = i_{DS,\text{pinch-off}} + g_o v_{DS}}$$



Summary of NMOS Modes

1. **Cut-off Mode**
   - $$i_{DS} = 0$$
   - Open switch.
   - $v_{GS} = 0 = v_{DS}$
2. **Triode Mode**
   - Voltage controlled resistor with conductance:
   $$g_{DS} = \mu_n C_{ox} \frac{W}{L}(v_{GS} - V_t)$$
   - $v_{GS} > V_t$
   - $v_{DS} < v_{GS} - V_t = V_{sat}$
3. **Pinch-off Mode**
   - $v_{GD} = V_t$
   - $v_{DS} = v_{GS} - V_t = V_{sat}$
   - No inversion near drain.
   - $$i_{DS,\text{pinch-off}} = \frac{1}{2}\mu_n C_{ox} \frac{W}{L}(v_{GS} - V_t)^2$$
4. **Saturation Mode**
   - $v_{GD} < V_t$
   - $v_{DS} > v_{GS} - V_t = V_{sat}$
   - No inversion near drain, and channel is getting smaller.
   - $$i_{DS,\text{sat}} = \frac{1}{2}\mu_n C_{ox} \frac{W}{L}(v_{GS} - V_t)^2 + g_o v_{DS}$$
   - $$i_{DS,\text{sat}} = \frac{1}{2}\mu_n C_{ox} \frac{W}{L}(v_{GS} - V_t)^2(1 + \lambda v_{DS})$$

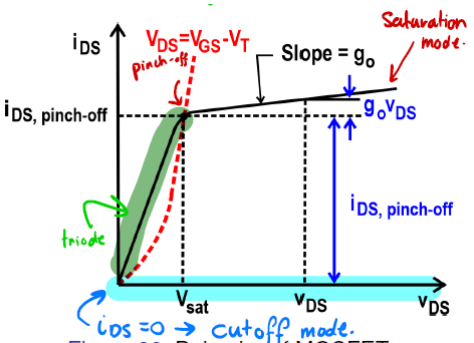
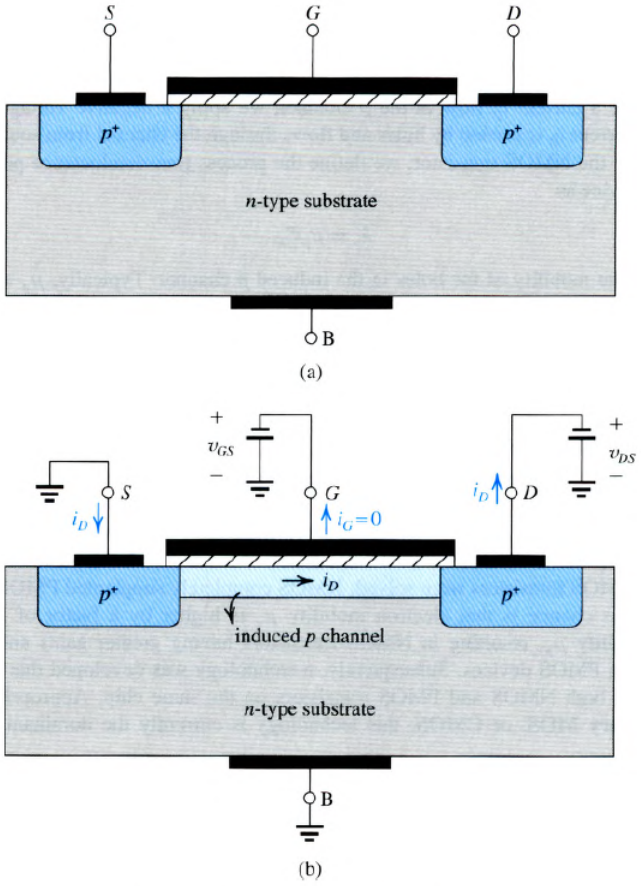


Figure 30: Behavior of MOSFET.

Differences in a PMOS



To induce current flow in a PMOS transistor, a negative voltage is applied to the gate. This voltage has to go below V_{tp} which is the threshold of a PMOS (V_{tp} is negative).

The condition for a *p-channel* to be created is:

$$V_{GS} \leq V_{tp}$$

... or equivalently:

$$|V_{sg}| \geq |V_{tp}|$$

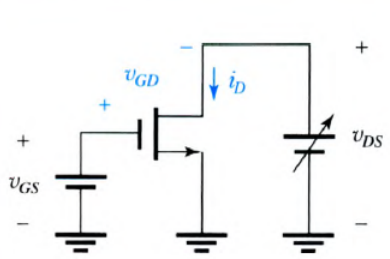
To induce current flow (carried by holes) in the *p-channel* a negative voltage is applied at the drain.

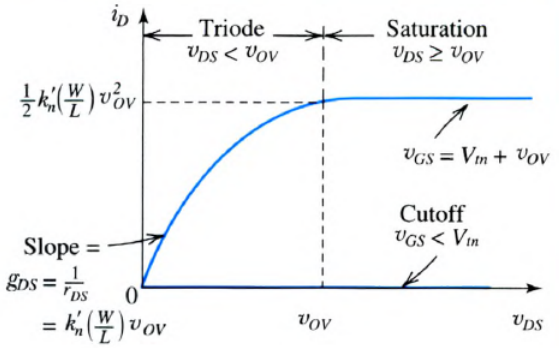
In general all the polarities of the voltages are flipped when considering a PMOS transistor. The following two table summarize these differences.

Note that in the tables:

- $$k'_n = \mu_n C_{ox}$$
- $$k'_p = \mu_p C_{ox}$$
- $$\text{NMOS: } v_{ov} = V_{GS} - V_{tn}$$
- $$\text{PMOS: } v_{ov} = V_{SG} - |V_{tp}|$$

Table 5.1 Regions of Operation of the Enhancement NMOS Transistor





- $v_{GS} < V_{tn}$: no channel; transistor in cutoff; $i_D = 0$
- $v_{GS} = V_{tn} + v_{OV}$: a channel is induced; transistor operates in the triode region or the saturation region depending on whether the channel is continuous or pinched off at the drain end;

Triode Region

Continuous channel, obtained by:

$$v_{GD} > V_{tn}$$

or equivalently:

$$v_{DS} < v_{OV}$$

Then,

$$i_D = k_n'\left(\frac{W}{L}\right)\left[(v_{GS} - V_{tn})v_{DS} - \frac{1}{2}v_{DS}^2\right]$$

or equivalently,

$$i_D = k_n'\left(\frac{W}{L}\right)\left(v_{OV} - \frac{1}{2}v_{DS}\right)v_{DS}$$

Saturation Region

Pinched-off channel, obtained by:

$$v_{GD} \leq V_{tn}$$

or equivalently:

$$v_{DS} \geq v_{OV}$$

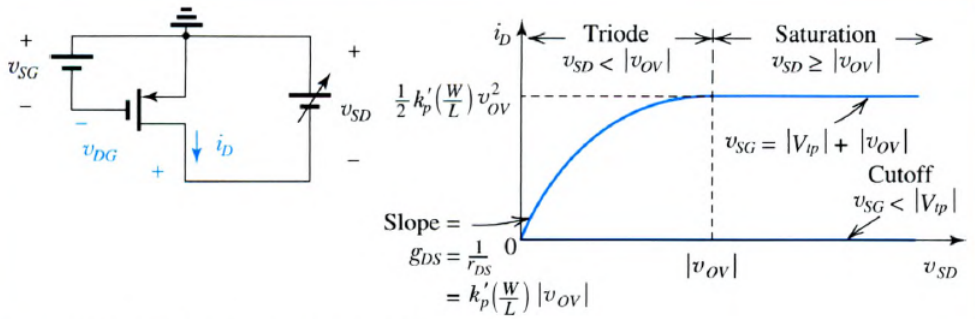
Then

$$i_D = \frac{1}{2}k_n'\left(\frac{W}{L}\right)(v_{GS} - V_{tn})^2$$

or equivalently,

$$i_D = \frac{1}{2}k_n'\left(\frac{W}{L}\right)v_{OV}^2$$

Table 5.2 Regions of Operation of the Enhancement PMOS Transistor



- $v_{SG} < |V_{tp}|$: no channel; transistor in cutoff; $i_D = 0$
- $v_{SG} = |V_{tp}| + |v_{OV}|$: a channel is induced; transistor operates in the triode region or in the saturation region depending on whether the channel is continuous or pinched off at the drain end;

Triode Region

Continuous channel, obtained by:

$$v_{DG} > |V_{tp}|$$

or equivalently

$$v_{SD} < |v_{OV}|$$

Then

$$i_D = k_p'\left(\frac{W}{L}\right)\left[(v_{SG} - |V_{tp}|)v_{SD} - \frac{1}{2}v_{SD}^2\right]$$

or equivalently

$$i_D = k_p'\left(\frac{W}{L}\right)\left(|v_{OV}| - \frac{1}{2}v_{SD}\right)v_{SD}$$

Saturation Region

Pinched-off channel, obtained by:

$$v_{DG} \leq |V_{tp}|$$

or equivalently

$$v_{SD} \geq |v_{OV}|$$

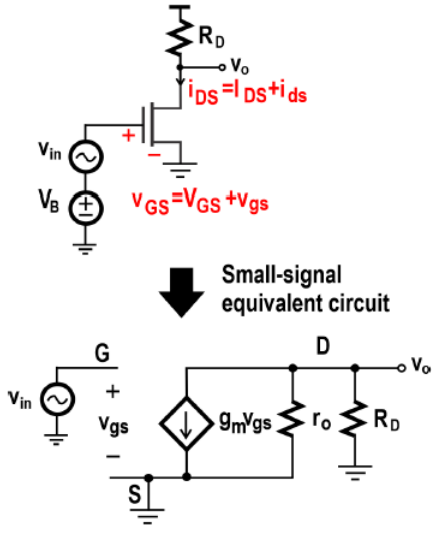
Then

$$i_D = \frac{1}{2}k_p'\left(\frac{W}{L}\right)(v_{SG} - |V_{tp}|)^2$$

or equivalently

$$i_D = \frac{1}{2}k_p'\left(\frac{W}{L}\right)v_{OV}^2$$

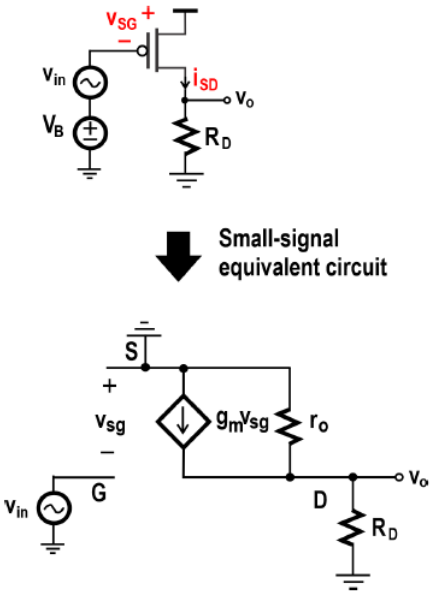
NMOS Small-Signal Equivalent Circuit



$$i_{ds} = g_m v_{gs} + g_o v_{ds}$$

$$g_m = \frac{2I_{DS}}{V_{GS} - V_t}$$

PMOS Small-Signal Equivalent Circuit



$$i_{ds} = g_m v_{sg} + g_o v_{sd}$$

$$g_m = \frac{2I_{DS}}{V_{GS} - V_t}$$