Large and Small Signals

Transfer characteristic

When we plot the output voltage as a function of the input voltage for the amplifier, this is also referred to as the transfer characteristic.
\( v_I \) small: \[ v_{GS} = v_I < V_{th_s} \quad \text{cutoff} \quad i_D = 0 \quad v_O = V_{DD} \]

\( v_I \) increases: \[ v_{GS} = v_I \geq V_{th_s} \quad \text{saturation} \quad i_D \uparrow \quad v_O \downarrow \]

\( v_I \) increases further: \[ v_{DS} < v_{GS} - V_{th_s} \quad \text{triode} \quad i_D \uparrow \quad v_O \downarrow \]

From this plot, we can find the small signal gain between input and output, the same way we did for a general function \( f(.) \). It is the tangent line to the curve at the operating point.

We observe the following:
- The transfer characteristic \( v_I - v_O \) is non-linear. This is expected, since one of the building blocks of the circuit (i.e. the MOSFET) is non-linear.
- Depending on the bias value, the MOSFET is in different operating modes.
- The small signal gain \( v_{out} / v_{in} \) depends on the bias point.
- The small signal gain is negative. This is not that important, since we are mainly interested in the absolute value of the gain.
- The small signal gain (in absolute value) is largest when the MOSFET is in saturation. This is typically true for any type of analog amplifier configuration that contains MOSFETs.

Note that the transfer characteristic for MOSFET circuits is less steep then the corresponding amplifier built with BJTs. The big advantage of MOSFETs is their lower power consumption and the zero gate current. For these reasons, almost all circuits these days are based on CMOS (complementary MOS – meaning with both nMOS and pMOS) technology.
Biasing

The goal of biasing is to set the bias conditions (i.e. large signals at DC) of the circuit such that the MOSFETs are in the desired operating mode, i.e. saturation. The bias settings impact the small signal behavior of the circuit, since the small signal model parameters depend on the large signal values.

However, properly biasing a circuit is non-trivial. The reason is that quite a few parameters of a MOSFET are often not well-defined due to variability during fabrication or temperature dependence. Spec sheets list average values, which may differ significantly from those of your actual device. Examples of parameters that are not well-defined are $V_{th}$, $\mu$, ... (they are temperature and process dependent). The goal of biasing is to have a robust way of setting the operating mode (typically saturation) and the desired bias voltages and currents of the MOSFET (i.e. large signals at DC).

Biasing Method 1: Degeneration resistor

First consider the most basic solution: simply setting a gate voltage. This will lead to a bad biasing scheme, and basically shows what not to do.

$$I_D = \frac{1}{2} \cdot \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_G - V_{th})^2 \cdot (1 + \lambda \cdot V_{OUT})$$

$$V_O = V_{DD} - I_D \cdot R_D$$

$$V_O = \frac{V_{DD} - R_D \cdot \frac{1}{2} \cdot \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_G - V_{th})^2}{1 + R_D \cdot \frac{1}{2} \cdot \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_G - V_{th})^2 \cdot \lambda}$$

However, since $V_{th}$, $\mu$, ... are not well-defined (and depend on temperature), this results in unpredictable values of $V_O$, such that the transistor might leave saturation and enter triode. This is undesirable.
Instead, it is better to use a degeneration resistor at the source. This resistor introduces negative feedback, which helps in making the configuration much less sensitive to not well-defined parameters, such as \( V_{th_a}, \mu, \ldots \). 

![Diagram of a MOSFET circuit with feedback resistors.]

In this case, the voltages and currents can be calculated from the following set of equations:

\[
\begin{align*}
I_D &= I_S = \frac{1}{2} \cdot \mu \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{GS} - V_{th_a})^2 \cdot (1 + \lambda \cdot V_{DS}) \\
V_G &= V_D = V_{dd} - I_D \cdot R_D \\
V_G &= V_{GS} + I_S \cdot R_S
\end{align*}
\]

The solution is non-trivial, but it is possible to gain the following intuition:

If \( V_{GS} < I_S \cdot R_S \) (so we have a lot of feedback), then the current is largely defined by the gate voltage (set by the resistive divider) and the resistor \( R_S \). In this case, \( I_S \approx V_G / R_S \). As a result, the biasing no longer depends on the MOSFET parameters.

Even if the feedback is not as strong, it still has a stabilizing effect on the biasing.

**Biasing Method 2: Current source**

An extremely popular and powerful method to bias a MOSFET circuit is using a current source. This is the most common way of biasing in an integrated circuit.

\[
I_D = I_S = I
\]

Now, how do we build a current source? The answer lies in a circuit called a current mirror, which is common in many designs.
Current mirror

First, consider the circuit on the right (BTW, notice the resemblance with a diode-connected BJT).

The nMOS is always in saturation (check this), and thus:

\[ I_D = \frac{1}{2} \cdot \mu_n \cdot C_{ox} \cdot \left( \frac{W}{L} \right)_1 \cdot (V_{GS} - V_{th_n})^2 \cdot (1 + \lambda \cdot V_{DS_1}) \]

\[ i_i = \frac{1}{2} \cdot \mu_n \cdot C_{ox} \cdot \left( \frac{W}{L} \right)_1 \cdot (v_i - V_{th_n})^2 \cdot (1 + \lambda \cdot v_i) \]

If the current through this device is set, this determines the voltage \( v_i \) over it. When that same voltage \( v_i \) is applied to the gate of another nMOS \( Q_2 \), the configuration is called a current mirror. It effectively mirrors the current through one branch (with the G-D connected device \( Q_1 \)) into the other branch (with device \( Q_2 \)). If we assume \( Q_2 \) is in saturation:

\[
\begin{align*}
I_{REF} &= \frac{1}{2} \cdot \mu_n \cdot C_{ox} \cdot \left( \frac{W}{L} \right)_1 \cdot (v_i - V_{th_n})^2 \cdot (1 + \lambda \cdot v_i) \\
I_{OUT} &= \frac{1}{2} \cdot \mu_n \cdot C_{ox} \cdot \left( \frac{W}{L} \right)_2 \cdot (v_i - V_{th_n})^2 \cdot (1 + \lambda \cdot V_{OVER}) \\
\Rightarrow \quad \frac{I_{OUT}}{I_{REF}} &= \left( \frac{W}{L} \right)_2 \cdot \frac{1 + \lambda \cdot V_{OVER}}{1 + \lambda \cdot v_i}
\end{align*}
\]

If channel modulation is ignored (\( \lambda = 0 \)), the current mirroring factor only depends on the sizes of the devices. The mirror factor can therefore be adjusted through device sizing (typically by changing the width \( W \) of the device).

\[
\frac{I_{OUT}}{I_{REF}} = \left( \frac{W}{L} \right)_2 = \text{Const}
\]

It was assumed \( Q_2 \) is in saturation. However, if \( V_{OVER} < v_i - V_{th} \), the transistor enters triode and the circuit no longer behaves as a current mirror. Note that \( V_{OVER} \) is determined by the rest of the circuit that is fed by \( I_{OUT} \).
Current source

A MOSFET current source can be built by using a current mirror and setting $I_{REF}$ through a resistor.

$$I_{REF} = \frac{1}{2} \cdot \mu_n \cdot C_{ox} \cdot \left(\frac{W}{L}\right)_1 \cdot (v_1 - V_{th})^2 \cdot (1 + \lambda \cdot v_1)$$

To set $I_{REF}$:

$$V_{DD} = v_1 + I_{REF} \cdot R$$

Solve for $R$ and $V_{GS}$.

Again, it is possible that $Q_2$ enters triode, depending on $V_{OVER}$. If that happens, the circuit no longer behaves as a current source.

In general, current mirrors and current sources are circuits that are very common in analog integrated CMOS design.
Biasing method 3: Cascading

The large signal output of one stage is used as the large signal input of the next stage.

Coupling small signals

The biasing schemes discussed above, all deal with the large signals. Small signals are present on top of these bias conditions.

One way of adding these small signals is using coupling capacitors. This assumes that the small signals of interest are at sufficiently high frequencies, such that the (large) capacitors can be approximated as shorts. These capacitors can be added at input and output.

The coupling capacitors can also be present between different stages, see (a) below. In that case, each stage is biased independently. However, the use of coupling capacitors is only possible when discrete components are used. In integrated circuits, it is difficult to create capacitors with sufficiently large values. Therefore, in this case, stages are typically coupled directly together for both large and small signals, see (b) below.
Small Signal Equivalent Model

\[ x_A = X_A + x_a \]

large signal \hspace{1cm} small signal

\[ x_A \rightarrow f(\cdot, \cdot) \rightarrow z_A \]

Taylor series

\[ z_A = f(x_A, y_A) \]
\[ = f(X_A, Y_A) + \frac{\partial f(x, y)}{\partial x} \bigg|_{x_A, y_A} (X_A - x_A) + \frac{\partial f(x, y)}{\partial y} \bigg|_{x_A, y_A} (Y_A - y_A) + \cdots \]
\[ \approx f(X_A, Y_A) + \frac{\partial f(x, y)}{\partial x} \bigg|_{x_A, y_A} x_a + \frac{\partial f(x, y)}{\partial y} \bigg|_{x_A, y_A} y_a \]
\[ = f(X_A, Y_A) + g(x_a, y_a) \]

\[ \begin{cases} 
Z_A = f(X_A, Y_A) & \text{large signal behavior} \\
z_a = \frac{\partial f}{\partial x} \bigg|_{X_A, Y_A} x_a + \frac{\partial f}{\partial y} \bigg|_{X_A, Y_A} y_a & \text{small signal behavior} 
\end{cases} \]

For MOSFETs, there are two inputs (e.g. for the pMOS, any two of the voltages \(v_{SG}, v_{SD}, v_{GD}\)). Each current (\(i_s, i_d, i_g\)) is a different function of these two inputs. We can apply the above theory on each one of these functions separately.

\[ i_s = f_1(v_{SG}, v_{SD}) \]
\[ i_d = f_2(v_{SG}, v_{SD}) \]
\[ i_g = f_3(v_{SG}, v_{SD}) \]

We will assume the MOSFET is in saturation. In principle, you could also derive the small signals models for any of the other modes, although MOSFETs are almost always biased in saturation.
Small signals (PMOS)

**Drain current**

\[ i_D = f(x, y) = \frac{1}{2} \cdot \mu_p \cdot C_{ox} \cdot \frac{W}{L} \cdot (x - V_{thp})^2 \cdot (1 + \lambda \cdot y) \]

\[ x = v_{SG} \]
\[ y = v_{SD} \]

\[ i_d = \frac{\partial f}{\partial x}\bigg|_{v_{SG}, v_{SD}} \cdot v_{sg} + \frac{\partial f}{\partial y}\bigg|_{v_{SG}, v_{SD}} \cdot v_{sd} \]

\[ = \left[ \frac{1}{2} \cdot \mu_p \cdot C_{ox} \cdot \frac{W}{L} \cdot 2 \cdot (x - V_{thp}) \cdot (1 + \lambda \cdot y) \right]_{v_{SG}, v_{SD}} \cdot v_{sg} \]
\[ + \left[ \frac{1}{2} \cdot \mu_p \cdot C_{ox} \cdot \frac{W}{L} \cdot (x - V_{thp})^2 \cdot \lambda \right]_{v_{SG}, v_{SD}} \cdot v_{sd} \]

\[ = \left[ \mu_p \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{SG} - V_{thp}) \cdot (1 + \lambda \cdot V_{SD}) \right] \cdot v_{sg} + \left[ \frac{1}{2} \cdot \mu_p \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{SG} - V_{thp})^2 \cdot \lambda \right] \cdot v_{sd} \]

\[ \Rightarrow i_d = g_m \cdot v_{sg} + \frac{v_{sd}}{r_o} \]

\[ g_m = \mu_p \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{SG} - V_{thp}) \cdot (1 + \lambda \cdot V_{SD}) \approx \mu_p \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{SG} - V_{thp}) \]

\[ g_m = \frac{2 \cdot I_D}{V_{SG} - V_{thp}} \]

\[ g_m = \sqrt{2 \cdot \mu_p \cdot C_{ox} \cdot \frac{W}{L} \cdot I_D \cdot (1 + \lambda \cdot V_{SD})} \approx \sqrt{2 \cdot \mu_p \cdot C_{ox} \cdot \frac{W}{L} \cdot I_D} \]

Any of these three expressions for \( g_m \) can be used. Each one expresses \( g_m \) as a function of two of the following three (non-independent) design parameters: \( I_D, V_{SG} \) and \( W/L \) (we ignore \( V_{SD} \) as a design parameter, since the effect of channel modulation is typically small).

Only two of these design parameters can be chosen independently (e.g. if \( V_{SG} \) and \( W/L \) are chosen, this fixes the current \( I_D \)). As such, the above expressions for \( g_m \) can be used to evaluate the effect on \( g_m \) when one parameter is varied and a second one is kept fixed. E.g. the second expression tells you that \( g_m \) is inversely proportional to \( (V_{SG} - V_{thp}) \) for a fixed current \( I_D \) (note that this requires the third parameter \( W/L \) to be changed appropriately as well).

\[ \left\{ \begin{array}{l}
  r_o = \frac{1}{\lambda \cdot I_D'} \\
  \approx \frac{1}{\lambda \cdot I_D} \\
  I_D' = \frac{1}{2} \cdot \mu_p \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{SG} - V_{thp})^2 \\
  = \frac{I_D}{1 + \lambda \cdot V_{SD}}
\end{array} \right. \]
Source current
\[ i_s = i_d = g_m \cdot v_{sg} + \frac{v_{sd}}{r_o} \]

Gate current
\[ i_g = 0 \]

**Typical values:**
- \( V_{SG} - |V_{thp}| \sim 2 \text{ V} \)
- \( I_D \sim 1 \text{ mA} \)
- \( 1/\lambda \sim 25 \text{ V} \)

\[ \begin{cases} 
  i_s = i_d = g_m \cdot v_{sg} + \frac{v_{sd}}{r_o} \\
  i_g = 0 
\end{cases} \]
Small signals (NMOS)

The small signal model is exactly the same as that for the PMOS. You can derive this yourself.

\[
\begin{align*}
g_m &= \mu_p \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{GS} - V_{th}) \cdot (1 + \lambda \cdot V_{DS}) \\
g_m &= \frac{2 \cdot I_D}{V_{GS} - V_{th}} \\
g_m &= \sqrt{2 \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot I_D \cdot (1 + \lambda \cdot V_{DS})} \\
r_o &= \frac{1}{\lambda \cdot I_D} \quad \approx \quad \frac{1}{\lambda \cdot I_D} \\

I_D' &= \frac{1}{2} \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{GS} - V_{th})^2 \\
&= \frac{I_D}{1 + \lambda \cdot V_{DS}}
\end{align*}
\]

Body effect

As mentioned before, the MOSFET is really a 4-terminal device. When B is not connected to S, the source-body voltage impacts the effective threshold voltage. For small signals, this results in the modified equivalent model shown below (the same for both nMOS and pMOS). The derivation of these models can be found in the textbook.

Unless specifically stated otherwise, in this course we will assume B is connected to S. As a result, \(v_{bs} = 0\), and we can simply use the models discussed earlier that do not include the body effect.