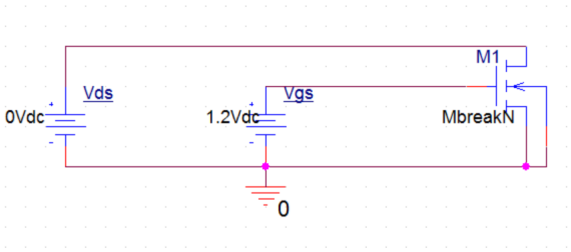

Figure 1: NMOS Circuit

In general, MOSFETs have three operation modes. The first mode is cutoff, where $V_{GS} < V_T$. The MOSFET structure uses a MOS capacitor in the middle. By setting a voltage at the gate, the MOS capacitor accumulates a high concentration of minority carriers in the semiconductor bulk. At a certain point, known as the threshold voltage, an inversion layer develops in the bulk between the source and the drain. Below this threshold voltage (at least in an NMOS), labeled V_T, not enough minority carriers exist in the bulk's conduction band to actually drive a substantial current. In the case of an NMOS transistor, the bulk is p-type, meaning minority carriers are electrons. Thus, when $V_{GS} < V_T$ in the NMOS, the drain current due to the minority carrier electrons is effectively 0mA.

When $V_{GS} > V_T$, a bifurcation occurs. The NMOS can either enter the triode or saturation mode. In the triode mode, the NMOS acts as an approximately linear resistor. The conductivity, and therefore the resistance, can be controlled by varying the number of minority carrier electrons by changing the gate voltage. By applying V_{DS}, a current is then driven through the somewhat linear resistance. For small values of V_{DS}, the resistance is basically linear. For larger values, the resistance begins to act less linear, but still exhibits similar behavior. The MOSFET remains in this triode mode for as long as $V_{DS} < V_{GS} - V_T$. However, once $V_{DS} > V_{GS} - V_T$, the MOSFET enters saturation. In the saturation mode, the NMOS does not have enough carrier electrons in the channel to support the desired current that should be present given V_{DS}. At this point, increasing V_{DS} leads to no additional current past the saturation current.

However, in the saturation mode, as V_{DS} is increased, more electrons concentrate toward the source than the drain. This is because the voltage in the semiconductor needs to reflect those applied at the metal electrodes. Thus, electrons concentrate toward the source to reflect the lower voltage that exists at that position. At saturation, no excess carrier electrons exist at the drain, and the channel "pinches-off" as a result. Past this point, the channel length decreases because more and more of the area closer to the drain becomes unoccupied by electrons as they move toward the source. As the channel becomes shorter, its resistance decreases, and the current increases linearly. This effect is known as channel length modulation. So, the saturation current looks like:

$$i_D = \frac{k_n}{2}(V_{GS} - V_T)^2(1 + \lambda V_{DS}) \ (\lambda : constant) \tag{1}$$

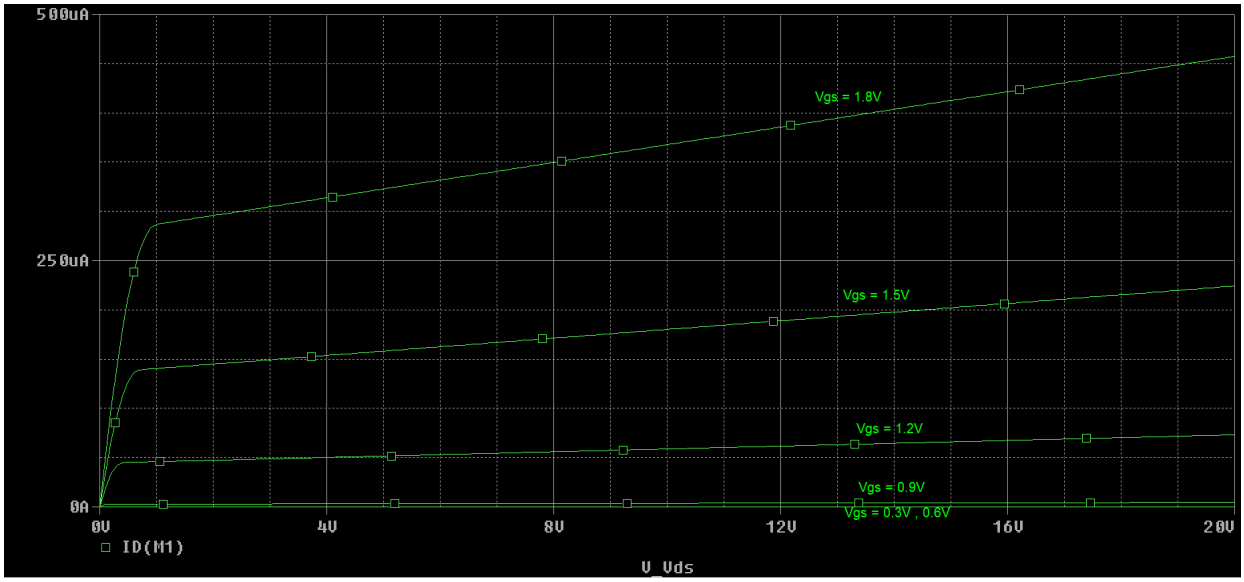

Figure 2: NMOS DC Sweep - i_D vs V_{DS}

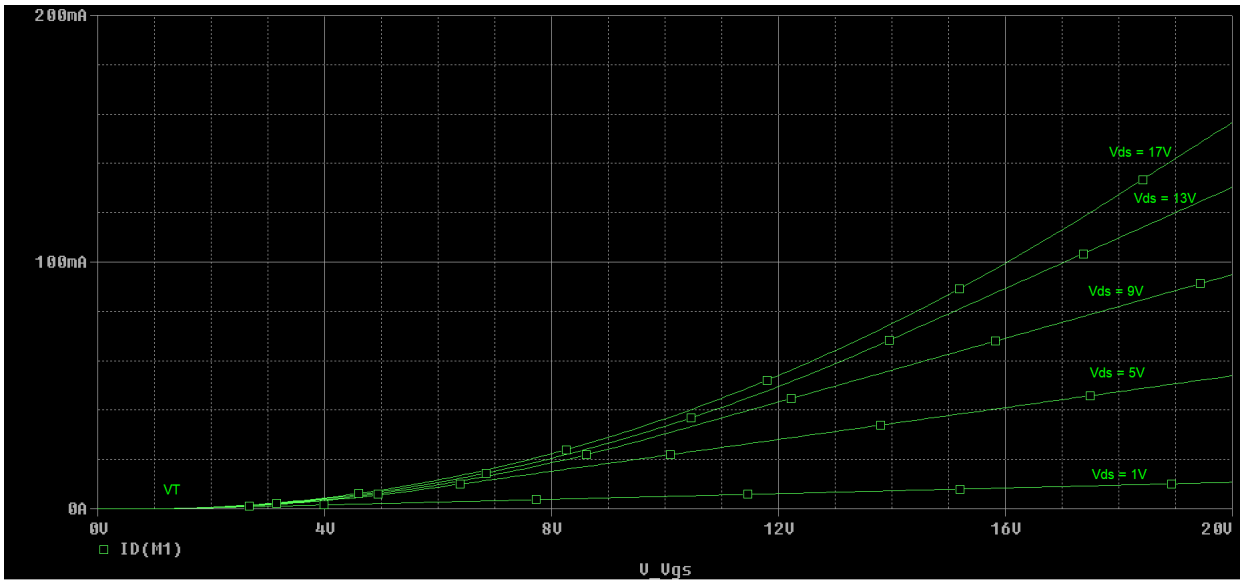

Figure 3: NMOS DC Sweep - i_D vs V_{GS}

When sweeping through increasing values of V_{GS}, the transistor goes from cutoff mode to saturation mode and eventually to triode mode. In the saturation mode, the current varies quadratically with V_{GS}:

$$i_D = \frac{k_n}{2}(V_{GS} - V_T)^2 \tag{2}$$

Once V_{GS} becomes sufficiently high, the transistor enters triode mode. At this point, the current has a linear dependence on V_{GS}:

$$i_D = k_n((V_{GS} - V_T) - \frac{V_{DS}}{2})V_{DS} \tag{3}$$

So, the transistor should remain at zero, then grow quadratically, and then linearly. This is precisely the behavior observed in figure (3). At higher V_{DS} values, the transistor remains in saturation for longer due to the fact that $V_{DS} > V_{GS} - V_T$ remains true for more values of V_{GS}. So, the drain attains higher values before entering the linear triode region. As a result, the slope of the curve when in the triode region increases with V_{DS}. However, as V_{DS} becomes large, the i_D versus V_{GS} curves asymptotically approach the curve $i_D = \frac{k_n}{2}(V_{GS} - V_T)^2$ for all V_{DS} values. When V_{DS} is large, the transistor may not enter the triode region in this V_{GS} range. This is why the curves begin to converge for larger V_{DS} values.