Lab 1: DC Analysis of MOS Transistors

EECS 170B

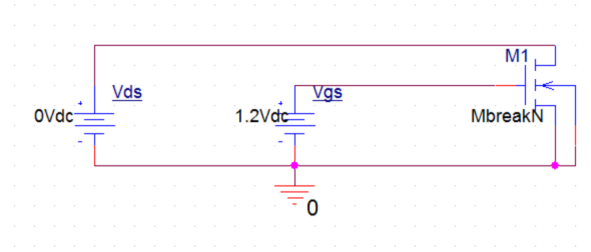### January 23, 2018

Roman Parise

1 NMOS Transistor

Figure 1: NMOS Circuit

In general, NMOS transistors have three operation modes. The first mode is cutoff, where $V_{GS} < V_T$, where V_T is the threshold voltage. The NMOS structure uses a MOS capacitor in the middle. Below the threshold voltage, 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.

By setting a voltage at the gate, the majority carrier holes in the p-type substrate are pushed away, exposing the negative acceptor ions beneath them. A depletion region forms as a result. Once the gate voltage exceeds the threshold voltage, valence band electrons in the depletion region acquire enough energy to jump the bandgap and become conduction electrons. As a result, an inversion layer develops in the bulk between the source and the drain. A drain current can now flow.

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:

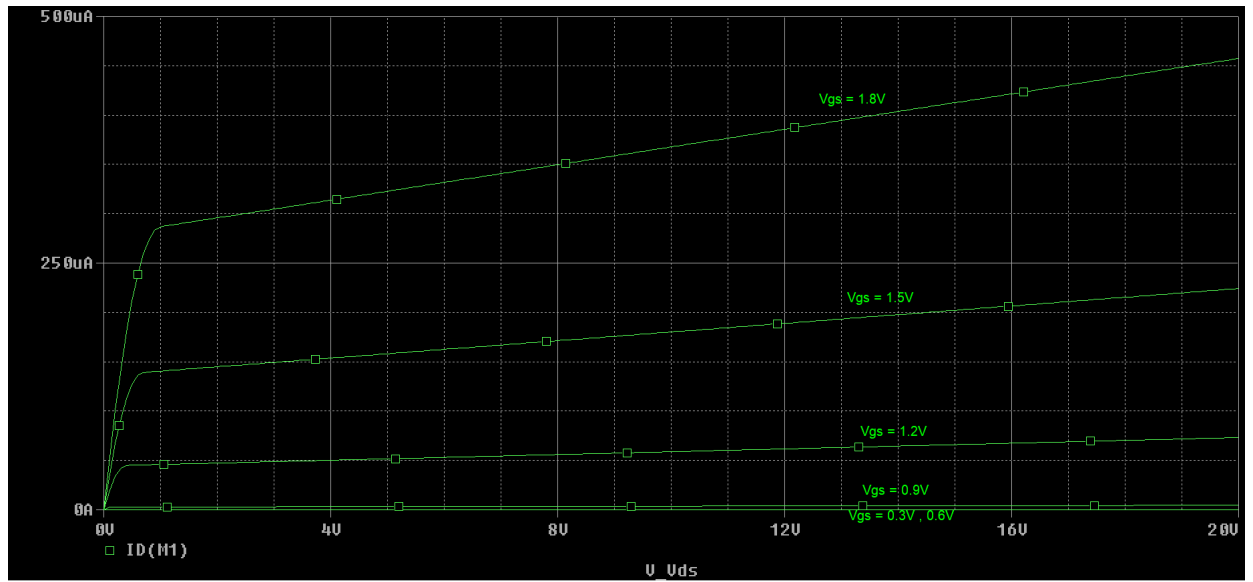$$i_D = \frac{k_n}{2}(V_{GS} - V_T)^2(1 + \lambda V_{DS}) \ (\lambda : constant) \tag{1}$$

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.
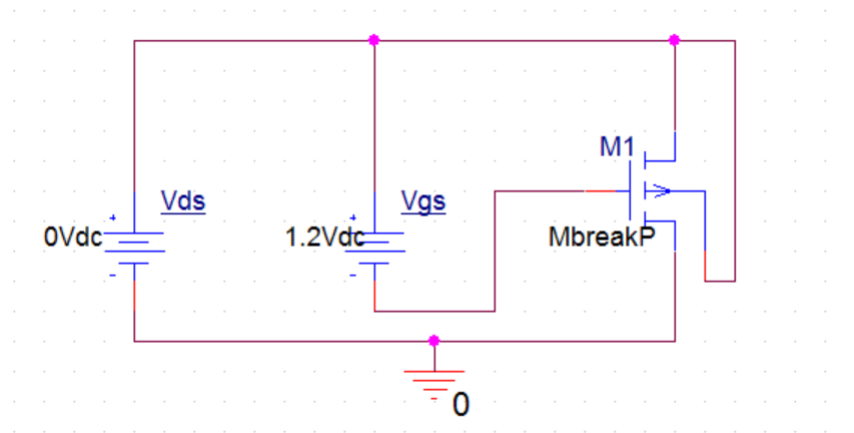
2 PMOS Transistor

Figure 4: PMOS Circuit

The PMOS transistor has behavior similar to the NMOS, but with a few slight deviations. When the MOS capacitor is charged a voltage lower than the source, holes are attracted to form the channel. The terminal from which carriers flow shall be called the source. So, for the PMOS, the source terminal should be higher than the drain terminal so that holes can flow from high to low potential from the source. Therefore, current flows into the drain, unlike in an NMOS transistor, where the current flows away from the drain. So, the NMOS current plot should be flipped over the x-axis for the PMOS since the current is negative.

Figure 5: PMOS i_D versus V_{SD}

If only the magnitude is observed, the plots should turn out to be the same. The voltages applied to the transistor are essentially the same, with source and drain terminals flipped, giving rise to the factor of -1 difference between the NMOS and PMOS current plots. For the SPICE models used, the PMOS has a process transconductance parameter $k_p \approx \frac{k_n}{2}$. The PMOS's process transconductance parameter is given by:

$$k_p' = \mu_p C_{ox} \tag{4}$$

The NMOS's parameter is:

$$k_n' = \mu_n C_{ox} \tag{5}$$

Here, C_{ox} is the oxide capacitance density per area, μ_n is the electron mobility, and μ_p is the hole mobility. k_p' is about half the magnitude of k_n' because the hole mobility μ_p is about half the magnitude of μ_n. Carrier electrons are simply

electrons that move around freely in the conduction band. Carrier holes arise from electrons moving around in the valence band. When an electron moves in the valence band, the hole moves in the opposite direction in the conduction band. In the valence band, electrons are much closer to the nuclei of the semiconductor material, meaning the electrostatic force is much stronger and prevents the electrons from moving as quickly as they would in the conduction band. Because the electrons in the valence band move more slowly, the corresponding holes in the conduction band also move more slowly (for the same applied electric field as electrons in the conduction band). This is what leads to the lower hole mobility and thus a lower process transconductance parameter.

The drain currents in transistors depend on the MOSFET transconductance parameter, which for a PMOS is given by:

$$k_p = k_p' \frac{W}{L} \tag{6}$$

W is the transistor's channel width, and L is the channel length. Making the width W twice the width of the corresponding NMOS model compensates for the lower process transconductance parameter. Thus, the PMOS's $|i_D|$ curve should be the same as the NMOS's. The factor of -1 difference for the i_D plots should be the only difference between the two.
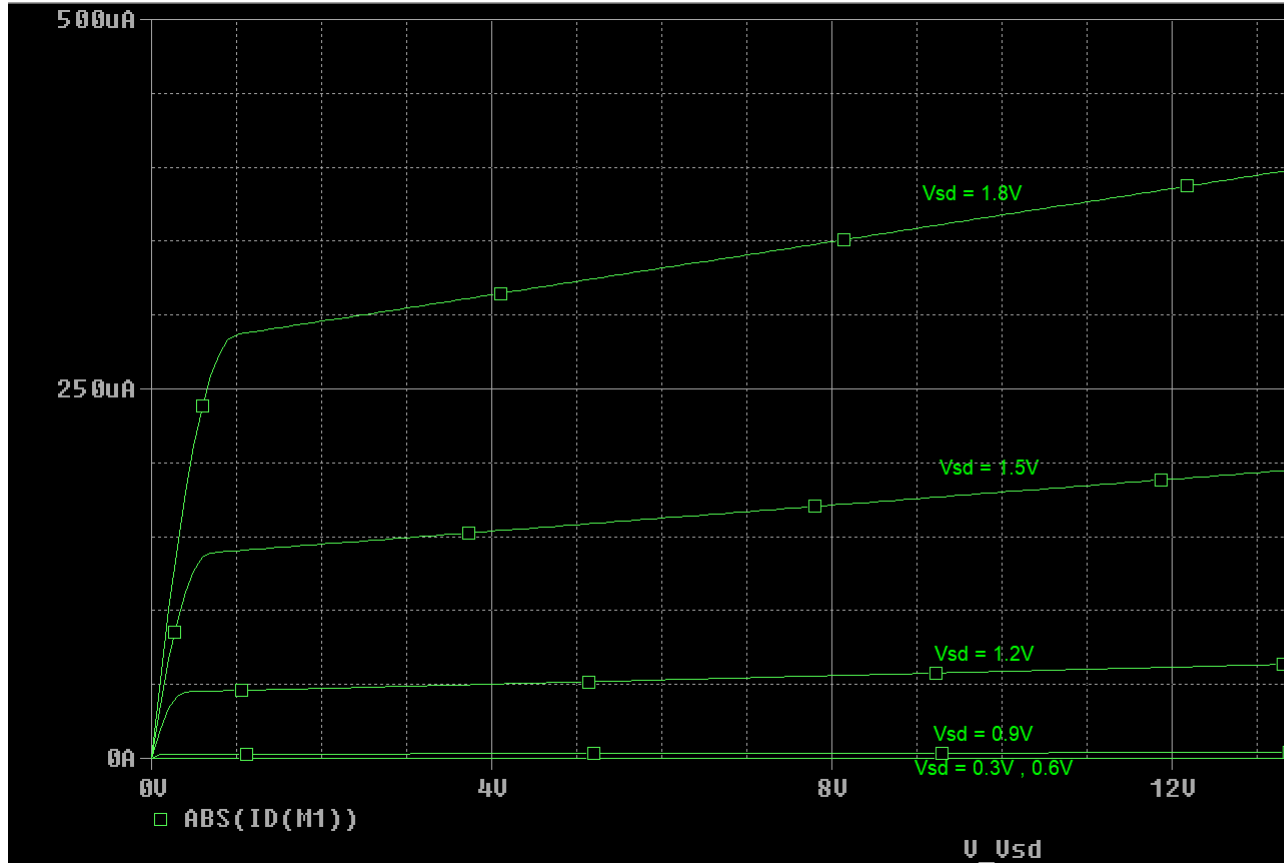
Figure 6: PMOS $|i_D|$ vs V_{SD}

When sweeping V_{SD}, the mechanics of the PMOS transistor are nearly identical besides the fact that holes are the charge carrier. As a result, the drain current differs by a factor of -1. The $|i_D|$ curve is essentially the same as for the NMOS for the same reasons as above.
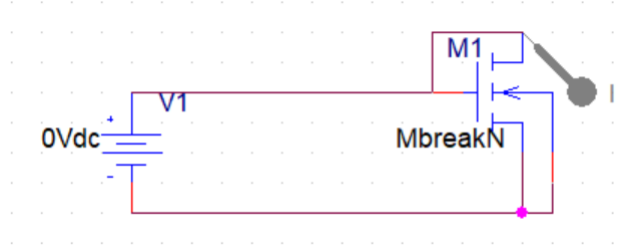
3 Diode-Connected NMOS

Figure 7: Diode-Connected NMOS Circuit

Two operation modes are possible for the circuit in figure (7). In the first case, consider $V_{GS} < V_T$. At this point, the transistor is in cutoff mode, and no current flows. For the second case, $V_{GS} > V_T$. Since $V_G = V_D$ and $V_{GS} > V_{GS} - V_T$, $V_{DS} > V_{GS} - V_T$. This satisfies the condition for saturation. So, the transistor is either in cutoff or in saturation. So, the transistor's drain current is given by:

$$i_D = \begin{cases} \frac{k_n}{2}(V_{GS} - V_T)^2 & V_{GS} > V_T \\ 0 & V_{GS} < V_T \end{cases} \tag{7}$$

This current behavior mimics that of a pn-junction diode, but with a current that varies quadratically instead of exponentially with the applied voltage V_{GS}. The transistor, like a diode, requires a certain voltage in order to "turn-on", specifically V_T. Once this voltage is exceeded, the transistor is activated and immediately enters the saturation region. From this, increasing V_{GS} leads to quadratically larger drain currents, just like how increasing the applied voltage leads to exponentially growing diode currents.
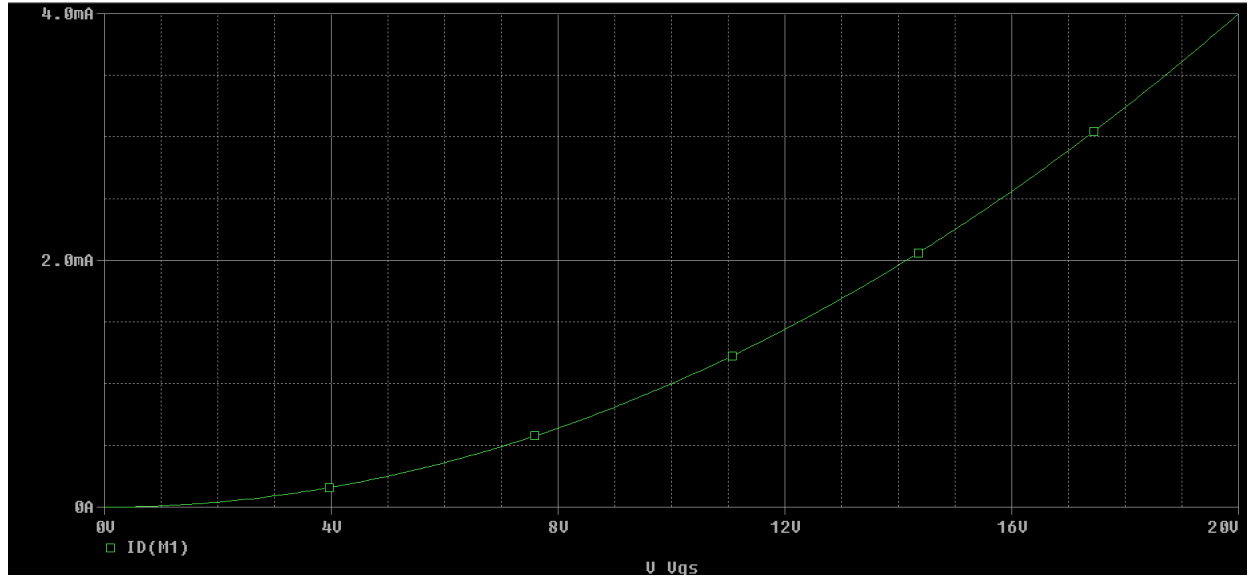
Figure 8: DC Sweep of Circuit in Figure (7)

4 Current Mirror
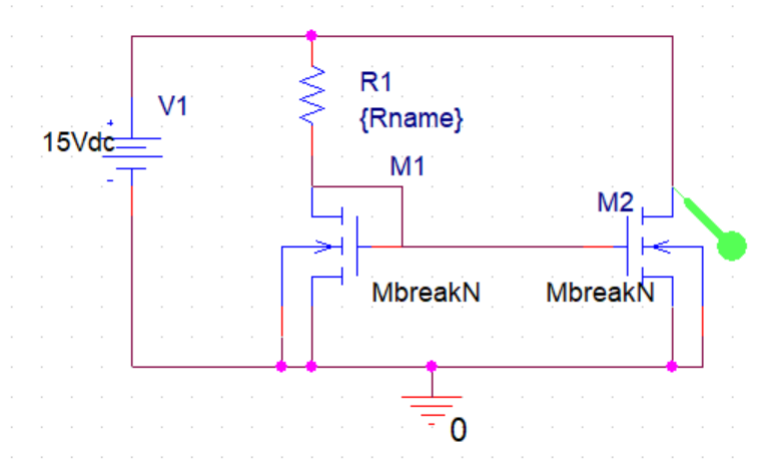
4.1 First Current Mirror Circuit

Figure 9: First Current Mirror Circuit

Assume transistors M_1 and M_2 in figure (9) are identical. Because M_1 is a diode-connected MOSFET, if $V_{DS} = V_{GS} > V_T$, then M_1 operates in satura-

9

tion. If R_1 is sufficiently small, then V_R can be small enough that $V_{DS} > V_T$. M_1's drain current shall be called i_{ref}, and M_2's i_{out}.

Because the gates of M_1 and M_2 are connected and both of their sources are grounded, V_{GS} is the same for each. Since M_1 and M_2 are identical transistors, their V_T values are identical as well. So, if current can flow through M_1's drain, then M_2 must also be active because $V_{GS} >= V_T$ in both cases. M_2's V_{DS} certainly exceeds M_1's V_{DS} since M_2's drain is directly connected to the supply voltage rather than an intermediary resistor. Since M_1 is in saturation, $V_{DS} >= V_{GS} - V_T$ is the condition for saturation, and M_2's V_{DS} exceeds M_1's V_{DS}, then M_2 must also be in saturation.

Since both transistors are in saturation, have the same V_{GS} value, and have identical structures, the following must be true:

$$\frac{i_{ref}}{i_{out}} = \frac{\frac{k_n}{2}(V_{GS} - V_T)^2}{\frac{k_n}{2}(V_{GS} - V_T)^2} = 1 \rightarrow i_{ref} = i_{out} \tag{8}$$

This circuit is called a "current mirror" because of the property demonstrated in equation (8). Given any input current i_{ref}, the output current i_{out} must be the same. Changing the dimensions of the transistors relative to one another can alter the output current by a constant factor [1].
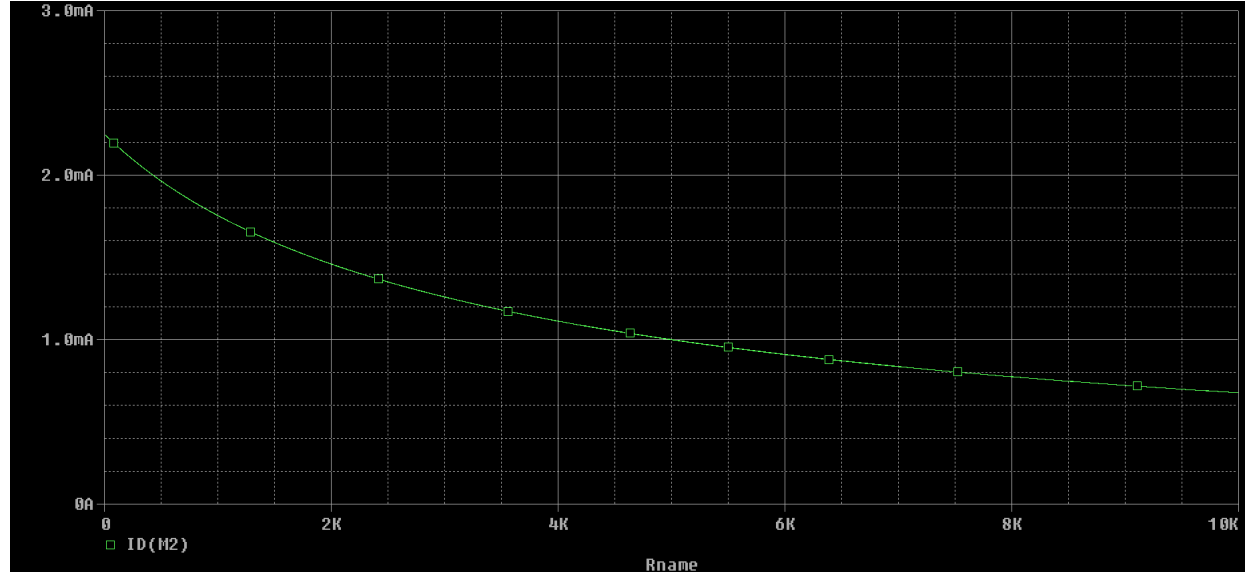

Figure 10: i_{out} versus R_1 for Current Mirror in Figure (9)

In the extreme case of an open circuit in the place of R_1, the open circuit absorbs the supply voltage, leaving none for M_1. So, $V_{DS} = V_{GS} = 0$. For

these NMOS transistors, $V_T > 0$ (confirm detail with a reference). Therefore, $V_{GS} < V_T$. So, M_1 is operating in the cutoff region, and no current flows.

On the other hand, if R_1 is a short and $V_T < V_{DD}$ (as it should be), where V_{DD} is the supply voltage, then $V_{GS} = V_{DD} > V_T$. Thus, M_1 is enabled. Because M_1 is diode-connected, it operates in the saturation region. Thus, as R_1 is increased, it absorbs more and more of the supply voltage V_{DD} until M_1 operates in cutoff mode. Thus, the i_{ref} versus R_1 curve should be downward sloping.
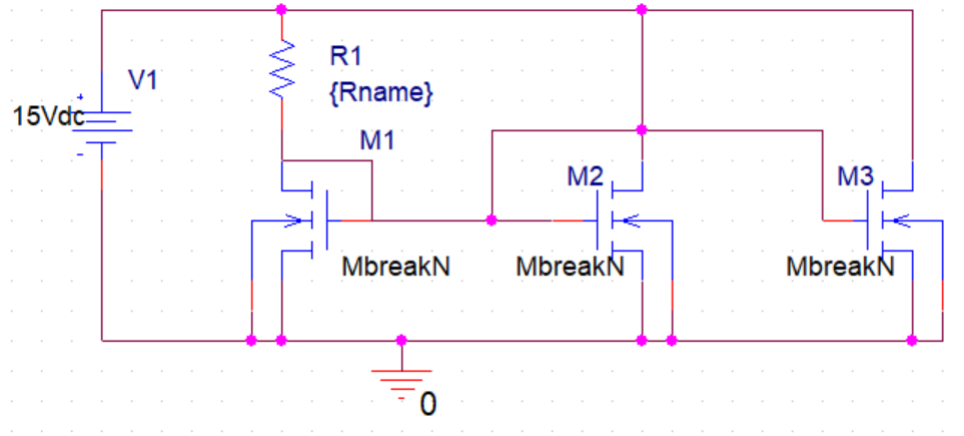
4.2 Second Current Mirror Circuit

Figure 11: Second Current Mirror Circuit

Consider the transistor M_1. $V_G = 15$V and $V_S = 0$V. Therefore, $V_{GS,M1}$ is likely much larger than $V_{T,M1}$. As a result, the transistor is active. Because M_1 is active and is diode-connected, M_1 must be in saturation. By the same arguments, M_2 and M_3 are also operating in saturation.

Both terminals of R_1 have the same potential due to the short path leading from the supply voltage lane to M_1's drain terminal. So, by Ohm's Law, no current flows through the resistor R_1. So, the circuit's operation is technically identical regardless of the value of R_1 in place. R_1 always acts as a short. Therefore, the plots of drain current versus R_1 should be constant for any of the transistors.

Because M_1 is in saturation, $i_{D,M1} = \frac{k_n}{2}(V_{GS,M1} - V_{T,M1})^2$. Because $V_{GS,M1}$ is so large in comparison to $V_{T,M1}$ due to M_1's terminals being directly connected to the supply voltage lane, $i_{D,M1}$ should not be 0mA. M_1 draws this

current from the supply voltage through the short path to its gate to its drain, bypassing the resistor R_1.

All of the transistors are identical in the sense that the voltages applied at the gate, source, and drain are the same. Thus, the drain current through each transistor is the same. So, the drain current through M_1 should be significant and constant when varied with R_1. The sum of the drain currents through M_2 and M_3 should also be double the drain current through M_1 and constant when varied with R_1. This is because the drain current curves are the same for M_2 and M_3 as they are for M_1. Therefore, adding them together simply doubles the constant drain current for all R_1 values.
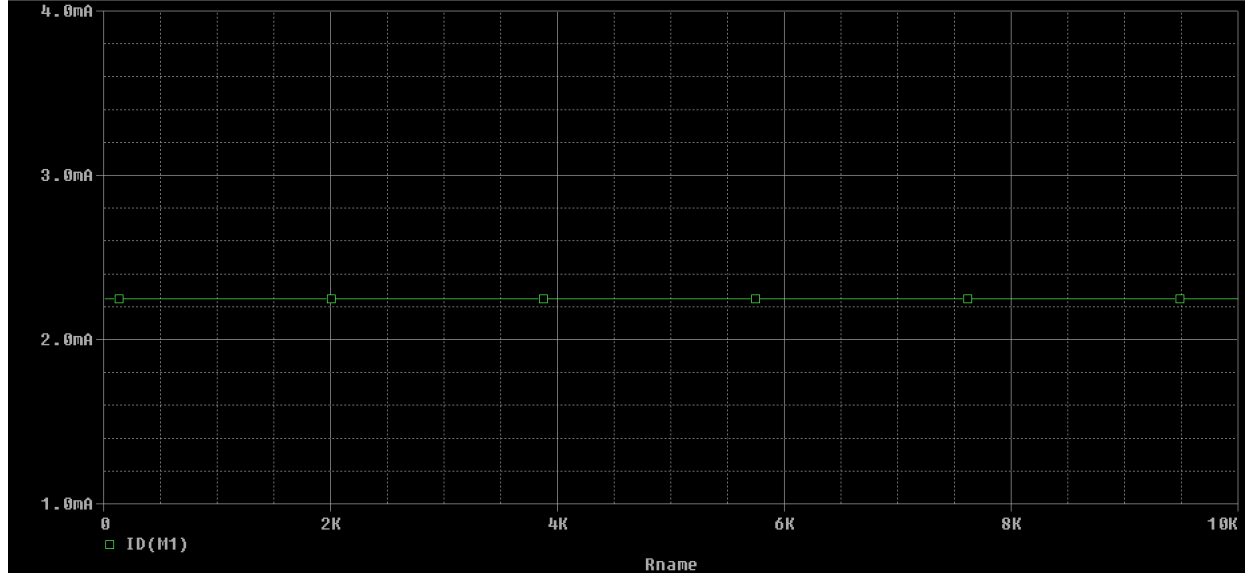

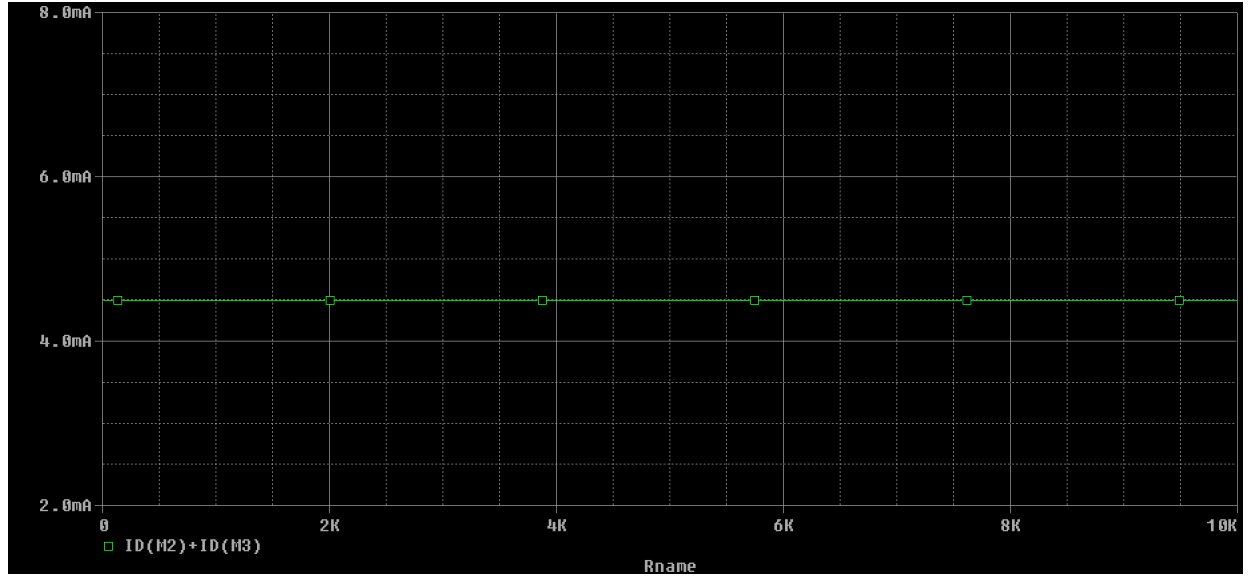Figure 12: $i_{D,M1}$ versus R_1

12

Figure 13: $i_{D,M2} + i_{D,M3}$ versus R_1

The results of the simulation are consistent with this analysis. In figure (12), the current value is essentially 2.25mA, a significant current value through the drain of M_1. In figure (13), the current value is nearly twice that, approximately 4.5mA. In both cases, the drain currents are not functions of the resistance R_1.

5   References

References

[1] http://inf-server.inf.uth.gr/courses/ce433/tutorials/mosfet%20current%20mirror.pdf.