6.012 - Electronic Devices and Circuits
Fall 2003

The Gradual Channel Approximation for the MOSFET:

We are modeling the terminal characteristics of a MOSFET and thus want $i_D(v_{DS}, v_{GS}, v_{BS})$, $i_B(v_{DS}, v_{GS}, v_{BS})$, and $i_G(v_{DS}, v_{GS}, v_{BS})$. We restrict our model to $v_{DS} \geq 0$ and $v_{BS} \leq 0$, so the diodes at the source and drain are always reverse biased; in this case $i_B \approx 0$. Because of the insulating nature of the oxide beneath the gate, we also have $i_G = 0$, and our problem reduces to finding $i_D(v_{DS}, v_{GS}, v_{BS})$.

The model we use is what is called the "gradual channel approximation", and it is so named because we assume that the voltages vary gradually along the channel from the drain to the source. At the same time, they vary quickly perpendicularly to the channel moving from the gate to the bulk semiconductor. In the model we assume we can separate the problem into two pieces which can be worked as simple one-dimensional problems. The first piece is the x-direction problem relating the gate voltage to the channel charge and the depletion region; this is the problem we solved when we studied the MOS capacitor. The second piece is the y-direction problem involving the current in, and voltage drop along, the channel; this is the problem we will consider now. To begin we assume that the voltage on the gate is sufficient to invert the channel and proceed.

Notice that $i_D(v_{DS}, v_{GS}, v_{BS})$ is the current in the channel; this is a drift current. There is a resistive voltage drop, $v_{CS}(y)$, along the channel from $v_{CS} = v_{DS}$ at the drain end of the channel, $y = L$, to $v_{CS} = 0$ at the source end of the channel, y = 0. At any point, y, along the channel we will have:

$$i_D = -q_N^*(y) \quad s_y(y)\, W$$

The current is not a function of y, $-q_N^*(y)$ is the channel sheet charge density at y,

$$-q_N^*(y) = -C_{ox}^* \quad [v_{GB} - V_T(y)]$$

with $C_{ox}^* \equiv \varepsilon_o / t_o$, and $s_y(y)$ is the net velocity of the charge carriers in the y-direction at y, which for modest electric fields is linearly proportional to the field:

$$s_y(y) = -\mu_e E_y(y) = -\mu_e \frac{-dv_{CS}(y)}{dy}$$

The current is then

$$i_D = W \mu_e C_{ox}^* \quad [v_{GB} - V_T(y)] \frac{dv_{CS}(y)}{dy}$$

To proceed, we must examine the factor $[v_{GB} - V_T(y)]$. We are referencing our voltages to the source so we first write $v_{GB} = v_{GS} - v_{BS}$. Next we look at $V_T(y)$; why is it a function of y? To answer this question we must note that the picture is a bit different in the MOSFET than it was before with the isolated MOS structure because now the channel (inversion layer) can have a voltage relative to the substrate. It is reverse biased by an amount $-v_{CB}(y)$ and so now the potential drop across the depletion region is $-2\phi_p + v_{CB}(y)$. Thus in our expression for V_T, $-2\phi_p$ is replaced by $-2\phi_p + v_{CB}(y)$. We have:

$$V_T(y) = V_{FB} - 2\phi_p + v_{CB}(y) + \frac{1}{C_{ox}^*}\sqrt{2\varepsilon_{Si} q N_A[-2\phi_p + v_{CB}(y)]}$$

It is common practice to name the factor $\frac{1}{C_{ox}^{*}}\sqrt{2\varepsilon_{Si}qN_A}$ the body factor, and call it γ, so we can then write $V_T(y)$ as

$$V_T(y) = V_{FB} - 2\phi_p + v_{CB}(y) + \gamma\sqrt{[-2\phi_p + v_{CB}(y)]}$$

Using this in the factor $[v_{GB} - V_T(y)]$ in the i_D expression, we have

$$[v_{GB} - V_T(y)] = v_{GB} - V_{FB} + 2\phi_p - v_{CB}(y) - \gamma\sqrt{[-2\phi_p + v_{CB}(y)]}$$

which, after using $v_{GB} = v_{GS} - v_{BS}$ and $v_{CB} = v_{CS} - v_{BS}$, and rearranging terms somewhat, is

$$[v_{GB} - V_T(y)] = v_{GS} - v_{CS}(y) - V_{FB} + 2\phi_p - \gamma\sqrt{[-2\phi_p + v_{CS}(y) - v_{BS}]}$$

The $v_{CS}(y)$ factor under the square root turns out to complicate the subsequent mathematics annoyingly and it has been found that it is better (and possible) to linearize this term before proceeding. We write the term as

$$\sqrt{[-2\phi_p + v_{CS}(y) - v_{BS}]} = \sqrt{[-2\phi_p - v_{BS}]}\sqrt{1 + \frac{v_{CS}(y)}{[-2\phi_p - v_{BS}]}}$$

and approximate the factor involving v_{CS} by expanding it and retaining only the first (linear term):

$$\approx \sqrt{[-2\phi_p - v_{BS}]}\,\left(1 + \frac{v_{CS}(y)}{2[-2\phi_p - v_{BS}]}\right)$$

which upon multipying becomes

$$= \sqrt{[-2\phi_p - v_{BS}]} + \frac{v_{CS}(y)}{2\sqrt{[-2\phi_p - v_{BS}]}}$$

Finally, giving the factor $1/2\sqrt{[-2\phi_p - v_{BS}]}$ the symbol δ, we write our linear approximation to the troublesome term as:

$$\sqrt{[-2\phi_p + v_{CS}(y) - v_{BS}]} \approx \sqrt{[-2\phi_p - v_{BS}]} + \delta\, v_{CS}(y)$$

Making this replacement, we have

$$[v_{GB} - V_T(y)] \approx v_{GS} - (1 + \gamma\delta)\, v_{CS}(y) - V_{FB} + 2\phi_p - \gamma\sqrt{[-2\phi_p - v_{BS}]}$$

Defining $V_T(v_{BS})$ as,

$$V_T(v_{BS}) \equiv V_{FB} - 2\phi_p + \gamma\sqrt{[-2\phi_p - v_{BS}]}$$

and giving the factor $(1 + \gamma\delta)$ the symbol α, we can write

$$[v_{GB} - V_T(y)] \approx [v_{GS} - V_T(v_{BS}) - \alpha\, v_{CS}(y)]$$

Putting this back into our expression for i_D, we find:

$$i_D = \frac{\varepsilon_o}{t_o}\, \mu_e W \,[\, v_{GS} - V_T(v_{BS}) - \alpha\, v_{CS}(y)]\, \frac{dv_{CS}(y)}{dy}$$

Multiplying both sides by "dy" yields

$$i_D dy = W\, \mu_e\, C_{ox}^{*} \,[\, v_{GS} - V_T(v_{BS}) - \alpha\, v_{CS}]\, dv_{CS}$$

We can now integrate both sides from $y = 0$ and $v_{CS} = 0$ to $y = L$ and $v_{CS} = v_{DS}$. We have

$$\int_0^L i_D\, dy = i_D \int_0^L dy = i_D L$$

and

$$\int_0^{v_{DS}} [\, v_{GS} - V_T - \alpha\, v_{CS}]\, dv_{CS} = [(v_{GS} - V_T)v_{DS} - \frac{\alpha\, v_{DS}^2}{2}\,]$$

Setting these two integals equal, and dividing both sides by L yields the expression for i_D we are looking for:

$$i_D(v_{DS}, v_{GS}, v_{BS}) = \frac{W}{L} \mu_e C^*_{ox} \left[\{v_{GS} - V_T(v_{BS})\} v_{DS} - \frac{\alpha v_{DS}^2}{2}\right]$$

It is worth reminding ourselves that arriving at this result we assumed that $v_{GS} > V_T$; otherwise i_D is zero because their is no channel. We also specified $v_{DS} \geq 0$ and $v_{BS} \leq 0$.

If we plot this expression for i_D verses v_{DS} for fixed values of v_{GS} and v_{BS}, we find that i_D varies linearly with v_{DS} when v_{DS} is small, but increases sub-linearly as v_{DS} increases further, i.e., the curve bends over to the right. Physically, the amount of inversion decreases toward the drain end of the channel and the resistance of the channel increases. Still, i_D continues to increase until $v_{DS} = (v_{GS} - V_T)/\alpha$, at which point the equation says i_D starts to decrease. What happens physically is that the channel disappears near the drain when $v_{DS} = (v_{GS} - V_T)/\alpha$, i.e., the region under the gate is no longer inverted near the drain. For larger values of v_{DS} the current does not decrease, but stays saturated at the peak value. We find

$$i_D(v_{DS}, v_{GS}, v_{BS}) = \frac{1}{2\alpha} \frac{W}{L} \mu_e C^*_{ox} [v_{GS} - V_T(v_{BS})]^2$$

for $v_{DS} \geq (v_{GS} - V_T)/\alpha$ and $v_{GS} > V_T$.

This completes the gradual channel approximation model for the MOSFET. Summarizing the results, we have a model valid for $v_{DS} \geq 0$ and $v_{BS} \leq 0$, and it says that the gate and substrate currents are zero for this entire range, i.e.,

$$i_G(v_{DS}, v_{GS}, v_{BS}) = 0$$

and

$$i_B(v_{DS}, v_{GS}, v_{BS}) = 0$$

The drain current has three regions:

Cutoff:

$$i_D(v_{DS}, v_{GS}, v_{BS}) = 0 \qquad \text{for } (v_{GS} - V_T)/\alpha \le 0 \le v_{DS}$$

Saturation:

$$i_D(v_{DS}, v_{GS}, v_{BS}) = \frac{K}{2\alpha}[v_{GS} - V_T(v_{BS})]^2$$

$$\text{for } 0 \le (v_{GS} - V_T)/\alpha \le v_{DS}$$

Linear (or triode)**:**

$$i_D(v_{DS}, v_{GS}, v_{BS}) = K\left[\{v_{GS} - V_T(v_{BS})\}\, v_{DS} - \frac{\alpha\, v_{DS}^2}{2}\right]$$

$$\text{for } 0 \le v_{DS} \le (v_{GS} - V_T)/\alpha$$

where K, α, and $V_T(v_{BS})$ are defined as

$$K \equiv \frac{W}{L}\mu_e C_{ox}^{*}, \quad \alpha \equiv 1 + \frac{t_o \sqrt{2\varepsilon_{Si} q N_A}}{2\varepsilon_o \sqrt{[-2\phi_p - v_{BS}]}}, \qquad \text{and}$$

$$V_T(v_{BS}) \equiv V_{FB} - 2\phi_p + \frac{t_o}{\varepsilon_o}\sqrt{2\varepsilon_{Si} q N_A[-2\phi_p - v_{BS}]}$$

One last point: It is often convenient to write $V_T(v_{BS})$ in terms of $V_T(0)$, and a function of v_{BS}. We have

$$V_T(v_{BS}) \equiv V_{FB} - 2\phi_p + \gamma\sqrt{(-2\phi_p - v_{BS})}$$

and

$$V_T(0) \equiv V_{FB} - 2\phi_p + \gamma\sqrt{-2\phi_p}$$

so the expression we want is

$$V_T(v_{BS}) \equiv V_T(0) + \gamma\left[\sqrt{(-2\phi_p - v_{BS})} - \sqrt{-2\phi_p}\right]$$

This will be useful when we look at linear small signal models for the MOSFET.