Design and Analysis of VLSI Subsystems
Dr. Madhav Rao
Department of Electronics and Communication Engineering
Indian Institute of Technology, Bangalore

Lecture - 03
NMOS Transistor Working

The threshold voltage value is defined by the formation of transistor or from the doping concentration. The n^+ dopant concentration depends on thickness of transistor as shown in figure 3.1. The thickness of the oxide will have the electric field forming the positive ions on the gate side to that of the negative minority carriers that has been stacked up at the interface side. This threshold voltage V_t is the function of p-type doping concentration N_D and oxide thickness t_{ox} when source and body terminal as same potential. If source and body terminal is grounded then threshold voltage is the function of source dopant concentration and potential difference between source to body terminal.

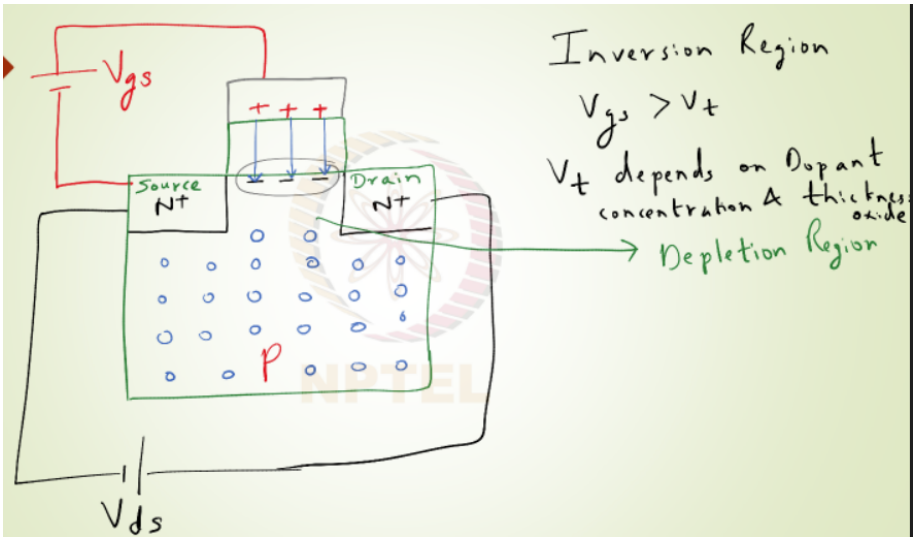

Figure 3.1: NMOS transistor – Inversion Region

Considering the source and the body has the same potential, let the V_t value be $0.3V$ and $V_{gs} = 1V$. As gate-to-source voltage V_{gs} is less than the threshold voltage. The source and drain region will have free electrons. The body has free holes but no free electrons. Suppose the source is grounded. The junctions between the body and the source or drain are zero-biased hence no current flows. Therefore, the transistor is OFF, and this mode of operation is called cutoff.

As V_{gs} increases from $0.3V$, $0.4V$ and so on to 1 volts, a strong channel is formed from source to drain side. Strong channel means a long channel, this channel accommodates a more majority carriers from source side to flow through the channel and then collect at the drain side. If $V_{gs} > 0.3V$ an inversion region of electrons (majority carriers) called the channel connects the source and drain, creating a conductive path and turning the transistor ON. The number of carriers and the conductivity increases with the gate voltage. The potential difference between drain and source is $V_{ds} = V_{gs} - V_t$.

If $V_{ds} = 0$ (i.e., $V_{gs} = V_{gd}$), there is no electric field tending to push the current from drain to source. When a small positive potential V_{ds} is applied to the drain ($V_{ds} > 0$), current I_{ds} flows through the channel from drain to source. This mode of operation is known as linear mode in case 1 as shown in figure 3.2.

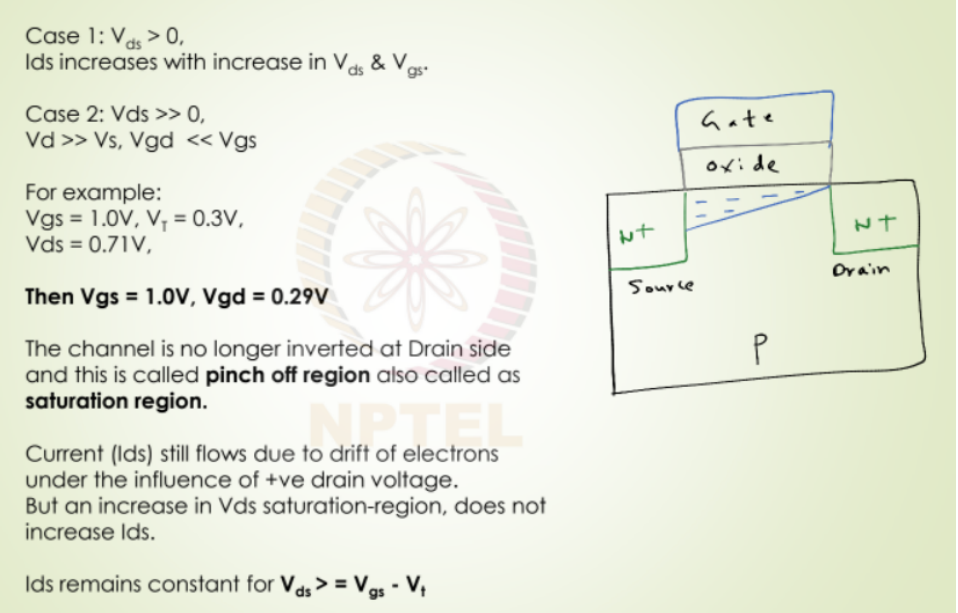

Figure 3.2: NMOS transistor working

Further increase in $V_{ds} \gg 0$ and $V_{gd} < V_t$, the channel is no longer inverted near the drain and becomes pinched off also called as saturation region. For example, when $V_{gs} = 1V$ fixed and $V_t = 0.3V$ and $V_{ds} = 0.7\text{V}$ then $V_{gd} = 0.29V$, which is less than V_t, thereby the channel is tapered at the drain end in case 2.

The reason being called as the saturation region is that the current which is extracted at the drain side or collected at the drain side is constant. As V_{ds} value increase beyond 0.7 volts like $0.71V$, 0.72, 0.79, 0.81, 0.89 and so on, the I_{ds} current remains constant and no channel exist for $V_{ds} \geq V_{gs} - V_t$. During this moment a strong lateral electric field is generated from drain to source and large number of majority carriers are stacked up at the drain side. This stacked up electrons is going to flow to the drain side not because of the channel being present, but because of the drift of the electrons which are caused by this lateral electric field. In summary, when $V_{ds} \geq V_{gs} - V_t$, the V_{gs} value tend to be less than V_t thereby the channel is pinch off and I_{ds} current extracted from the drain to source becomes less.

To estimate the average DC current I_{ds} for the NMOS transistor when channel is absent. Consider the average charge accumulated at the channel side is the capacitance multiplied by the voltage $Q = CV$. Thus, the charge in the channel $Q_{channel}$ is

$$Q_{channel} = C_g(V_{gc} - V_t) \tag{3.1}$$

Where, C_g = capacitance of the gate to the channel,

$V_{gc} - V_t$ = amount of voltage attracting charge to the channel beyond the minimum required to invert from p to n.

The gate voltage is referenced to the channel, which is not grounded. If the source is at V_s and the drain is at V_d , the average is $V_c = \frac{V_s + V_d}{2}$. Therefore, the mean difference between the gate and channel potentials V_{gc} is

$$V_g - V_c = V_{gs} - \frac{V_{ds}}{2} \tag{3.2}$$

For example, consider the linear resistor model, where drain is $5V$ and source is $3V$is shown in figure 3.3. The potential difference at two extreme points is $5 + 3/2 = 4V$. This linear resistor model in the channel is being formed by the minority carriers of the p-type body.

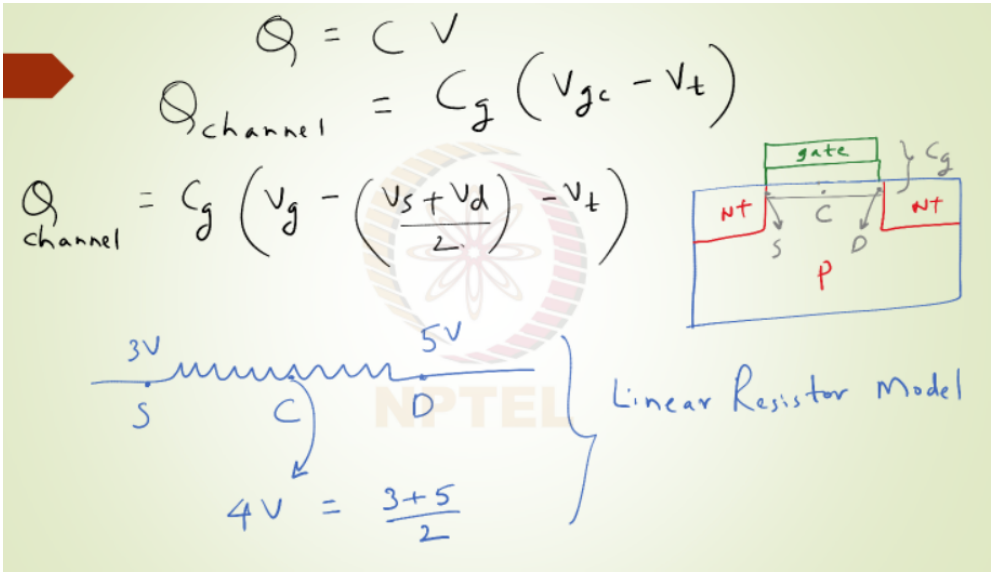

Figure 3.3: Linear resistor model

Substitute in equation 3.1 and 3.2 by adding and subtracting $\frac{V_s}{2}$, then the charge accumulated at channel is

$$Q_{channel} = C_g(V_{gs} - \frac{V_{ds}}{2} - V_t) \tag{3.3}$$

The gate as a parallel plate capacitor with capacitance proportional to area over thickness. If the gate has length L and width W and the oxide thickness is t_{ox} is shown in figure 3.3. The gate capacitance is

$$C_g = \frac{K_{ox}\varepsilon_o WL}{t_{ox}} = \varepsilon_{ox}\frac{WL}{t_{ox}} = C_{ox}WL \tag{3.4}$$

Where, ε_o is the permittivity of free space, $8.85 \times 10^{-14}\ F/cm$, and the permittivity of $Si0_2$ is $K_{ox} = 3.9$ times as great. The term $\frac{\varepsilon_{ox}}{t_{ox}}$ term is called C_{ox} is the capacitance per unit area of the gate oxide.

The average current or dc current is defined as the overall charge accumulated divide by time.

$$I_{avg}\ or\ I_{dc} = \frac{Q_{channel}}{time} \tag{3.5}$$

The time is defined as the time taken for the majority carriers to start from source end and reaches to drain end. Its dependent on channel length and velocity of charge carriers is given as

$$time = \frac{L}{V} = \frac{L}{\mu E_{ds}} = \frac{L^2}{\mu V_{ds}}$$

(3.6)

Where E_{ds} = lateral electric field generated after applying the voltage at drain to source side.

The electric field E is the voltage difference between drain and source V_{ds} divided by the channel length is $E_{ds} = \frac{V_{ds}}{L}$.

Therefore, the current between source and drain is the total amount of charge in the channel divided by the time is

$$I_{ds} = \frac{Q_{channel}}{time} \tag{3.7}$$

Substitute equation 3.3, 3.4 and 3.6 in 3.7 is given as

$$I_{ds} = \frac{C_{ox} WL}{\frac{L^2}{\mu V_{ds}}} (V_{gs} - V_t - \frac{V_{ds}}{2})$$

$$I_{ds,linear} = \mu C_{ox} \frac{W}{L} (V_{gs} - V_t - \frac{V_{ds}}{2}) V_{ds}$$

$$I_{ds,linear} = \beta (V_{gs} - V_t - \frac{V_{ds}}{2}) V_{ds} \tag{3.8}$$

$$where\ \beta = \mu C_{ox} \frac{W}{L}$$

The equation 3.8 is the current expression of NMOS transistor in linear region.

During saturation region, when $V_{ds} = V_{gs} - V_t$, then the drain to source current is

$$I_{ds,sat} = \frac{\beta}{2} (V_{gs} - V_t)^2$$

(3.9)

The channel is pinched off at the drain end. Thus, this saturation current is independent of the V_{ds}. As V_{ds} increases in linear and saturation region by this V_{gs} increases and hence I_{ds} of linear and saturation region current also increases.

Example 3.1: For 65nm technology given $\frac{W}{L} = \frac{4}{2}$, $t_{ox} = 1.05nm$, $V_t = 0.3V$, and $\mu = 80cm^2/V - sec$. Calculate β and $I_{ds,sat}$.

Solution: $\beta = \mu C_{ox} \frac{W}{L} = \frac{80\ x\ 8.854\ x \frac{10^{-19}F}{cm} x\ 3.9}{1.05\ x\ 10^{-7} cm} x \frac{4}{2}$

$\beta = 5.26\ x\ 10^{-4}$

$I_{ds,sat} = \frac{\beta}{2}\left(V_{gs} - V_t\right)^2 = \frac{5.26\ x\ 10^{-4}}{2}(1 - 0.3)^2 = 128.9\mu A\ or\ 0.128mA$

When $V_{gs} = 1V$

For 1.05nm of oxide thickness, a small dimension of 0.128mA of current is extracted. Hence the channel length of 65nm technology the dimension of width is 100nm and effective channel length is 50nm. With this small dimension, it can extract 1/10th mA of current. If width increases 10 times more, then the current also increases. This current I or β is directly proportional to W and thereby a more current is produced. A 1-micron width of transistor gives 1.28mA of current, which is approaching from the doped silicon material and then it is used to construct a transistor from that particular doped silicon material.