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

Lecture - 05
MOS Capacitances

In general, different kinds of MOS capacitance are associated with the transistor model. Here its focused on NMOS transistor model. The performance of any circuit design in digital system is heavily dependent on the capacitance.

In any critical path to propagate a signal from input to output node, then the performance delay of that particular propagation signal will depend on the capacitance associated with it. This capacitance is the extraction of transistor which are designed for the system essentially. In addition to it the gate, the source and drain also have capacitances. These capacitances are not fundamental to operation of the devices, but do impact circuit performance and hence are called parasitic capacitance. These parasitic should be minimized in the transistor model. The majority carriers (electrons) in n^+ drain region will be attracted towards the positive side. Thus, towards drain region the n^+p junction acts as a reverse bias. Then the positive immobile ions doped in the silicon material moves closer to the interface of PN junction diode is shown in figure 5.1.

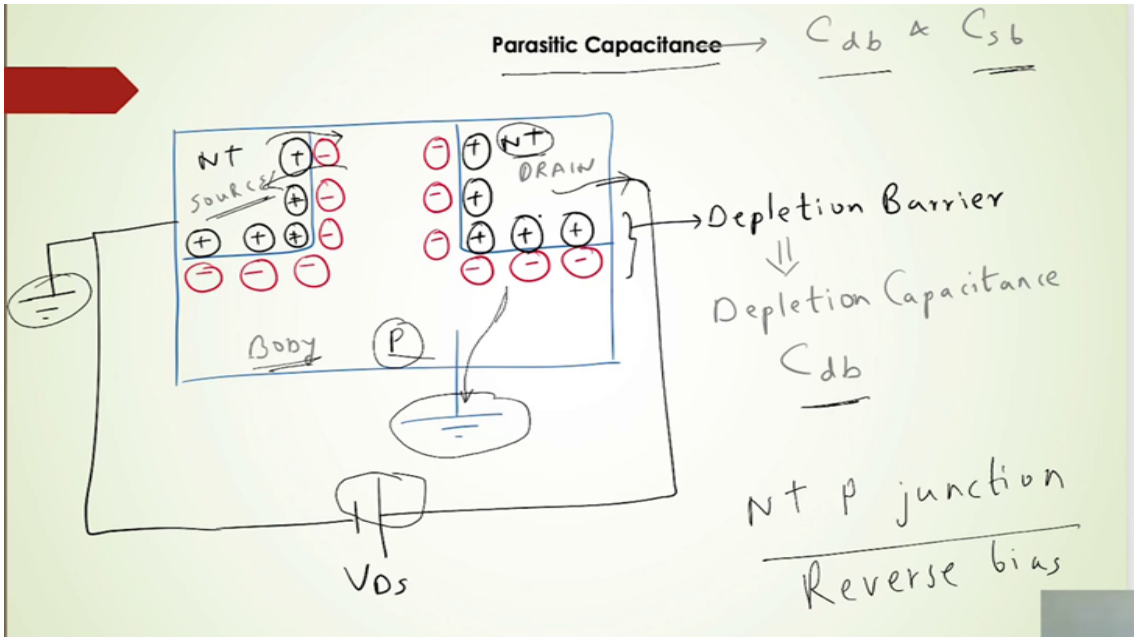

Figure 5.1: MOS capacitance structure

The negative immobile boron ions which are formed on the other side of the PN junction diode is attracted towards the ground potential. The majority carriers (electrons) at the drain region are attracted towards the positive potential and leaving this immobile ion. Therefore, the immobile ions from PN junction forms the junction capacitance. The junction capacitance is also known as depletion barrier or depletion capacitance and it's denoted by C_{db}, where d represents the drain and b represents the body.

At source region, as source and body potential are grounded due to drift the majority carriers in n^+ plus region move to the other side by leaving the positive ion. Similarly, the holes move to the other side leaving behind the negative immobile ions. Thereby this forms the depletion barrier or depletion capacitance.

The source and drain capacitances arise from the p–n junctions between the source or drain diffusion and the body and hence are also called diffusion capacitance C_{sb} and C_{db} as shown in figure 5.1. A depletion region with no free carriers forms along the junction. The depletion region acts as an insulator between the conducting p- and n-type regions, creating capacitance across the junction. The capacitance of these junctions depends on the area and perimeter of the source and drain diffusion, the depth of the diffusion, the doping levels, and the voltage.

For a 65nm technology node the length of the gate or the length of the channel which is 2λ. From figure the diffusion length l_d is 5λ and width is 4λ. Where 'd' is represented as diffusion depletion depth. If l_d increases, then the area of the depletion region will keep increasing. This will have a directly effect on the C_{db}. As C_{db} increases the area of the drain surface region increases thereby the depletion capacitance also increases. That is as l_d increases the resistance diffused in that region also increases.

Considering the l_d is fixed to small value to retain the diffusion resistance R_{diff} and C_{db} value is low and that is standardized to 5λ. The minimum channel length is 2λ. Then the source to body or the drain to body or the gate capacitances is almost equal to $1fF/\mu m$ of the width. Thus, for 65nm process node the width value is 4λ. If lambda is equal to 25nm and width is 4λ then the total width is 100nm. Therefore C_{sb}, C_{db} and C_g is equal to $0.1fF$.

In the nmos gate capacitance, the gate capacitance varies with respect to different mode of the transistor. At accumulation layer, if V_{gs} is negative the majority carriers are holes in p-type body are attracted to the interface in first condition. Then the distance of two plates will be the thickness of oxide. Thereby the capacitance

$$C_{gb} = C_o = \frac{\varepsilon_{ox} WL}{t_{ox}} \tag{5.1}$$

In second condition, when V_{gs} is positive but less than V_t it enters into depletion region. Depletion region represents that this positive potential is going to move away the holes. The overall parallel plate capacitance thickness is going to increase then the distance between the parallel plate capacitance also increases and hence C_{gb} starts decreasing.

Third condition, in the inversion region the gate is at positive potential it move away all the holes, but the minority carriers are able to reach the interface and thereby the channel is formed between the source and the drain end. Once the channel is formed, it carries the charge and thereby this forms the capacitance. The distance between the two parallel plates between the gate and the channel that is being formed is nothing but at the t_{ox}. Then,

$$C_{gb} = C_o = \frac{\varepsilon_{ox} WL}{t_{ox}}$$

In these three different modes when V_{gs} was negative and V_{gs} was greater than V_t, the capacitances values will be same. In depletion mode when $V_{gs} = 0V$ the holes are moving away from the oxide interface as shown in figure 5.2.

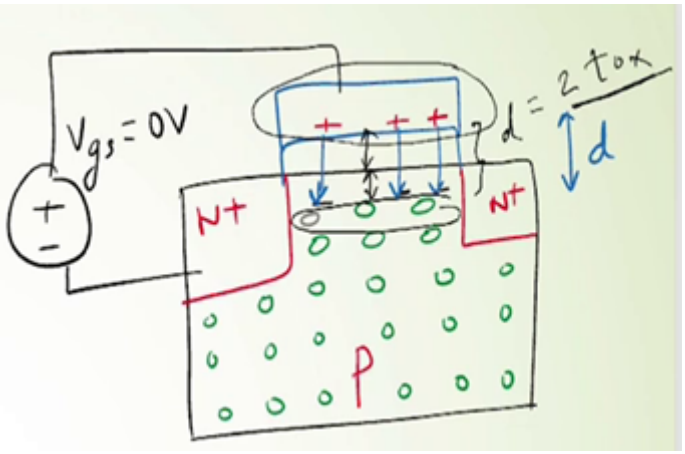

Figure 5.2: Depletion mode $V_{gs} = 0V$

The effective distance between the two parallel plates turns out to be twice that of the t_{ox}.

i.e., $$C_{gb} \cong \frac{\varepsilon_{ox} WL}{2t_{ox}} \quad (5.2)$$

When, $V_{gs} = V_t$, in figure 5.3,

$$C_{gb} < \frac{\varepsilon_{ox} WL}{2t_{ox}} \quad (5.3)$$

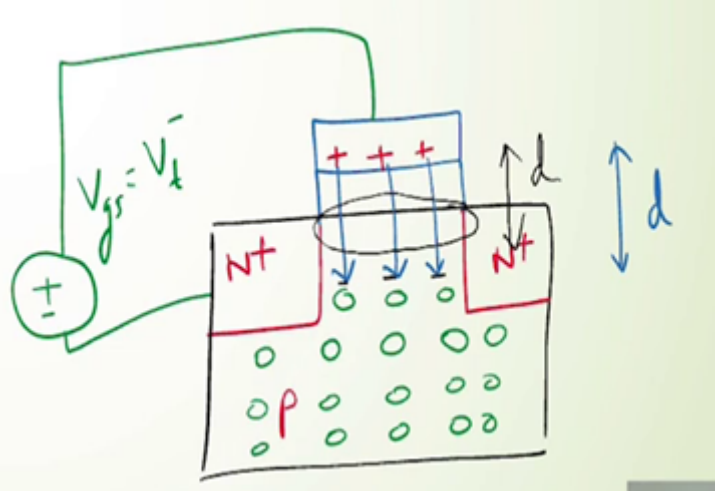

Figure 5.3: Depletion mode $V_{gs} = V_t$

Plotting the capacitance graph for V_{gs} vs C_{gb}. For negative V_{gs}, the transistor is in accumulation and $C_{gb} = C_o$. As V_{gs} increases but remains below a threshold value, a depletion region is formed at the surface. This effectively moves the bottom plate downward from the oxide, reducing the capacitance, as shown in Figure 5.4.

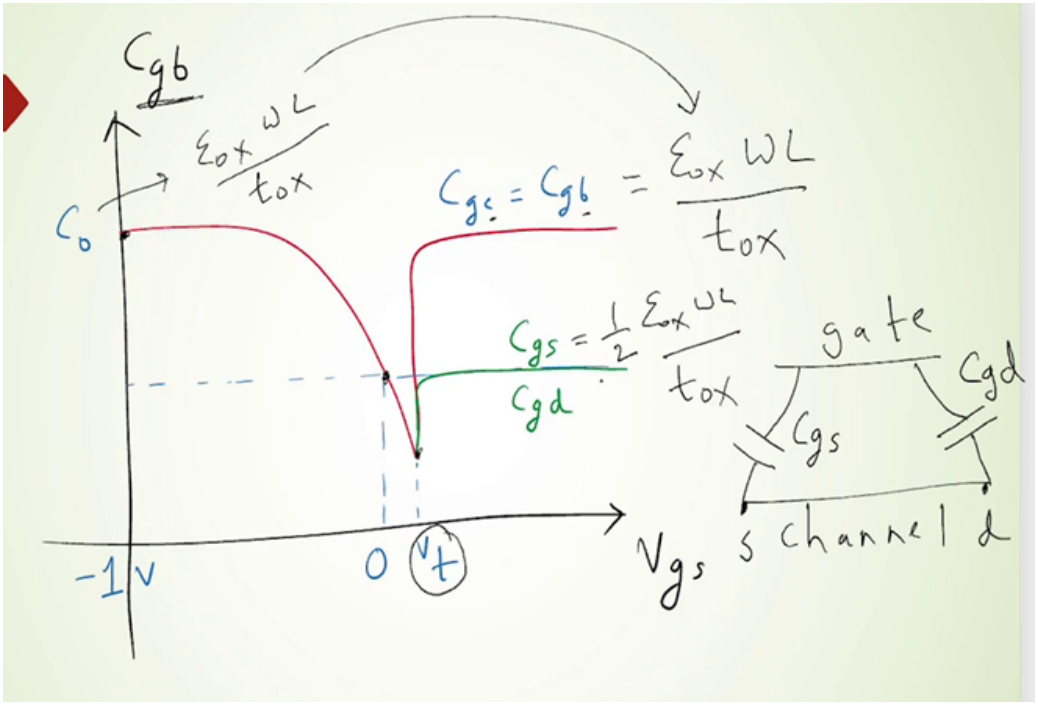

Figure 5.4: C_{gd} *versus* V_{gs} *characteristic*

At low values of V_{ds}, the channel charge is roughly shared between source and drain, then $C_{gs} = C_{gd} = C_o/2$. As V_{ds} increases, the region near the drain becomes less inverted, thus a greater fraction of the capacitance is attributed to the source and a smaller fraction to the drain, as shown in Figure 5.5.

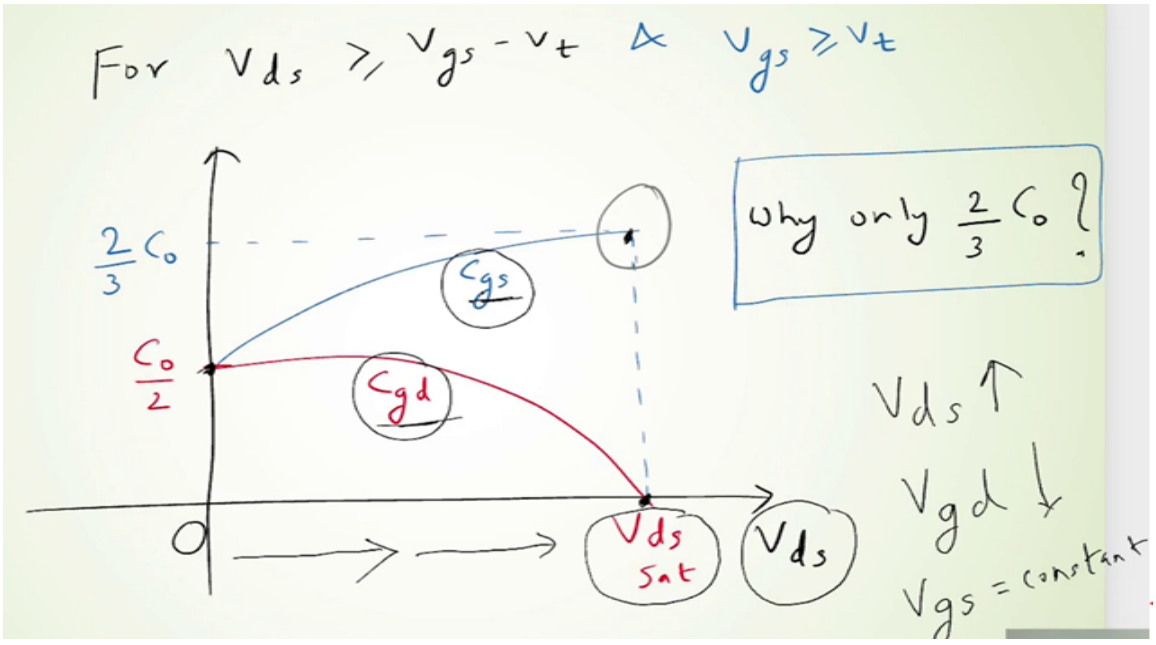

Figure 5.5: Inversion region characteristic

As $V_{ds} > V_{ds,sat}$, the transistor saturates and the channel is pinched off. At this point, all the intrinsic capacitance is to the source, as shown in Figure 5.5. Because of pinch-off, the capacitance in saturation reduces to $C_{gs} = 2/3\ C_o$. Therefore, as V_{ds} increases, V_{gd} starts decreasing for constant V_{gs}.

From the sodini reference paper, the drain to source current is given as

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

$$V_{gs} - \frac{V_{ds}}{2} = V_g - V_s - \left(\frac{V_d - V_s}{2}\right) = V_g - \left(\frac{V_d + V_s}{2}\right) = V_g - V(y)$$

$$I_{ds} = \mu W C_{ox}(V_{gs} - V_t - V(y))dV(y)/dy$$

(5.4)

The total charge is

$$Q_T = C_T V_T = C_{ox} WL\left(V_T\right) = WC_{ox}(V_T L) \tag{5.5}$$

From figure 5.6 at drain the y is equal to L and at source the y is equal to 0. The midpoint between source and drain y is equal to l/2. The potential at the point A is the source potential, potential at point B is the drain potential. Take the integral range from 0 to L. Also, represent the cross-section transistor source and drain region.

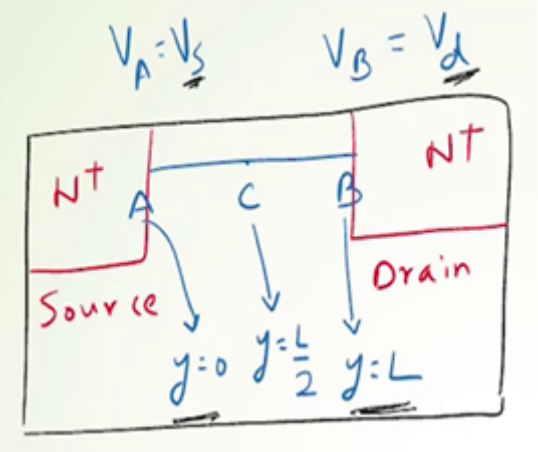

Figure 5.6: NMOS structure

$$Q_T = WC_{ox}\int_0^L \left(V_{gs} - V_t - V(y)\right)dy \tag{5.6}$$

Substitute equation 5.5 in 5.6,

$$Q_T = \frac{\mu W^2 C_{ox}^2}{I_{ds,sat}} \int_{V_s}^{V_d} \left(V_{gs} - V_t - V(y)\right)^2 dy$$

$$Q_T = \frac{\mu W^2 C_{ox}^2}{\mu \frac{W}{L} C_{ox} \frac{(V_{gs}-V_t)^2}{2}} \left[\frac{\left(V_{gs}-V_t-V(y)\right)^3}{-3}\right]_{Vs}^{Vdsat}$$

$$Q_T = \frac{-2}{3}\frac{WLC_{ox}}{(V_{gs}-V_t)^2}\left[(V_g - V_t - V_{ds,sat})^3 - (V_{gs} - V_t)^3\right]$$

To estimate the capacitance at saturation point, $V_g - V_t - V_{ds,sat} = 0$

$$Q_T = \frac{2}{3}WLC_{ox}(V_{gs} - V_t)$$

$$Q_T = \frac{2}{3}C_o(V_{ds,sat})$$

(5.7)

$$Q = C_{eff}V_{ds}$$

The total charge turns out to be $\frac{2}{3}C_o$ into V_{ds} saturation.

The gate capacitor can be viewed as a parallel plate capacitor with the gate on top and channel on bottom with the thin oxide dielectric between. Therefore, the gate capacitance is

$$C_g = C_{ox}WL \tag{5.8}$$

The bottom plate of the capacitor is the channel, which is not one of the transistor's terminals. When the transistor is on, the channel extends from the source (and reaches the drain if the transistor is unsaturated, or stops in saturation). Thus, this often approximate the gate capacitance as terminating at the source and call the capacitance C_{gs}. Thus, taking this minimum L as a constant for a particular process, can be defined as

$$C_g = C_{permicron} W \tag{5.9}$$

Where,

$$C_{permicron} = WL$$

For 65nm technology node, $C_{permicron}$ is $1fF/\mu m$, if w is 100nm then C_g turns out to be $0.1fF$.