Chapter 9
Advanced CMOS Devices

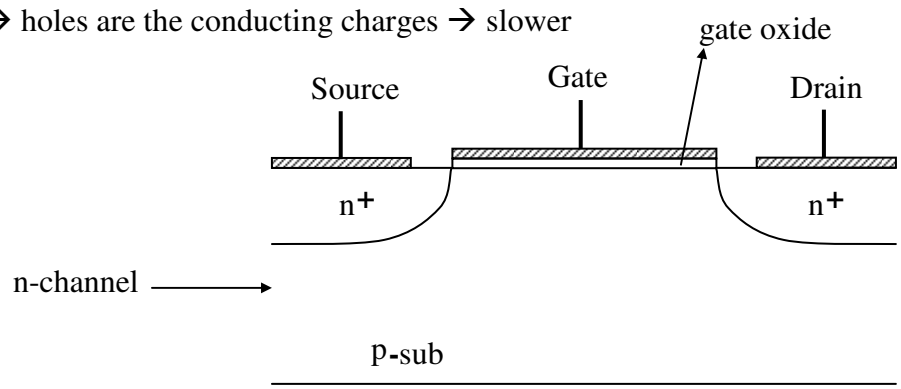
In this chapter we study the basics of metal oxide semiconductor field effect transistors (MOSFETs) and their state of the art performance. It is important to realize that they are two distinctive models for MOS transistor. In long channel limit (typically L>1μm) a simple quadratic equation can model device performance. Deviations from the quadratic model are captures as secondary effects. For submicron devices (typically L<0.5μm) which includes almost all current processes, the simple quadratic equation does not hold. Such devices are referred to as short channel devices as short channel effects dominate their performance. There are several approaches to model short channel transistors. Advanced short channel MOS models involve more than 150 parameters, out of which only a handful are physical parameters. Therefore, modeling MOS transistor becomes a tedious task.

As the gate length of the transistor shortens by scaling to deep submicron technology, not only one gets the benefit of higher packing density and lower supply voltage (low power application) but also there are benefits in terms of the speed of the transistor. f_T and f_{max} increase to above 200GHz for 90nm MOS technology. All of a sudden, MOS transistor seems to be a reasonable choice for RF and microwave application. In this chapter, we will study the implications of MOS transistors as RF and microwave devices.

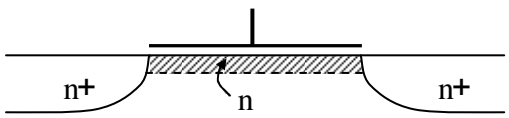
Basic Principles of Long Channel MOSFET

N-channel → electrons are the conducting charges → faster

P-channel → holes are the conducting charges → slower



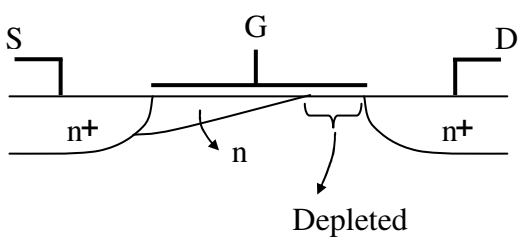
Weak inversion → behaves like a resistor



$$I_D = \mu_n C_{ox} \frac{W}{L} \underbrace{(V_{GS} - V_{th})}_{V_{eff}} V_{DS} \tag{10.1}$$

As you now increase V_{DS} → $$I_D = \mu_n C_{ox} \frac{W}{L}\left[(V_{GS} - V_{th})V_{DS} - \frac{V_{DS}^{\ 2}}{2}\right] \tag{10.2}$$

As you now increase V_{DS} → your channel narrows down close to drain until it pinches off.



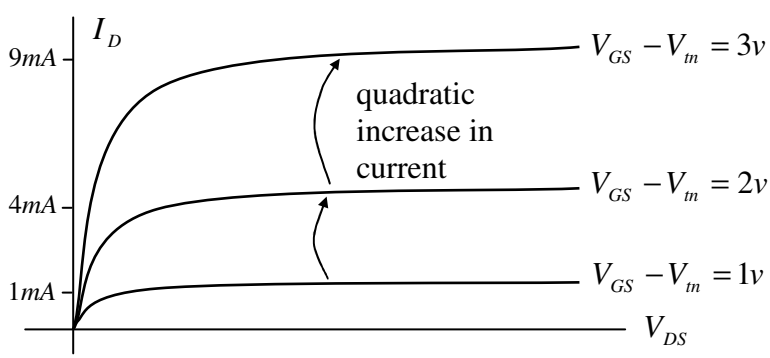
→ current saturates.

Set $V_{DS} = V_{GS} - V_{th}$ in equation (I)

→ $$I_D = \mu_n C_{ox} \frac{W}{2L}(V_{GS} - V_{th})^2 \tag{10.3}$$

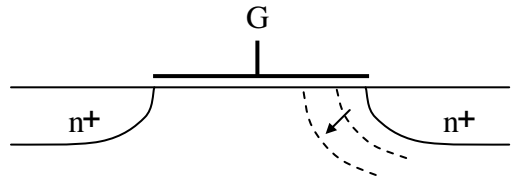
I_D is now independent of V_{DS}.

Condition for pinch off: $V_{DS} > V_{GS} - V_{th}$.



Secondary Effects

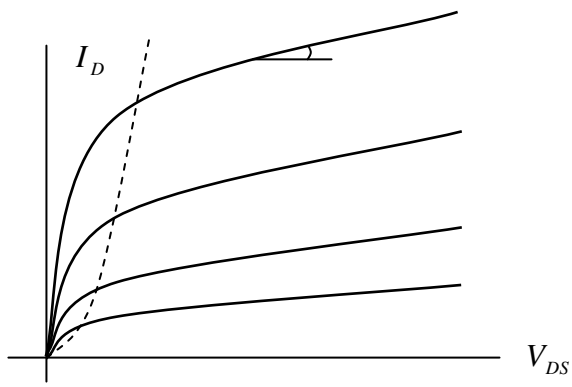
1. Channel length modulation



By changing the V_{DS} you effectively shorten the channel length → current increases.

$$I_D = \mu_n C_{ox} \frac{W}{2L} (V_{GS} - V_{th})^2 \left(1 + \lambda \left[V_{DS} - (V_{GS} - V_{th})\right]\right) \tag{10.4}$$

$\lambda \equiv \left(V^{-1}\right)$ → Channel length modulation coefficient



$$\rightarrow \quad 1/R_{out} = \frac{\partial I_D}{\partial V_D} = \lambda I_D \qquad R_{out} = \frac{1}{\lambda I_D} \tag{10.5}$$

2. Body effect

Body effect influences both DC and AC performance of the transistor:

For DC one observes a change in the threshold voltage as one biases the substrate different from Source voltage.

$$V_{th} = V_{th0} + \gamma\left(\sqrt{V_{SB} + |2\Phi_F|} - \sqrt{|2\Phi_F|}\right) \tag{10.6}$$

γ: Body effect constant

V_{th} increases as you increase the source-body voltage

V_{SB} is the source body potential that causes the body effect

Φ_F is a constant fitting parameter around 0.7V

We will discuss the AC effects of body biasing when we talk about small signal modeling of MOSFETs.

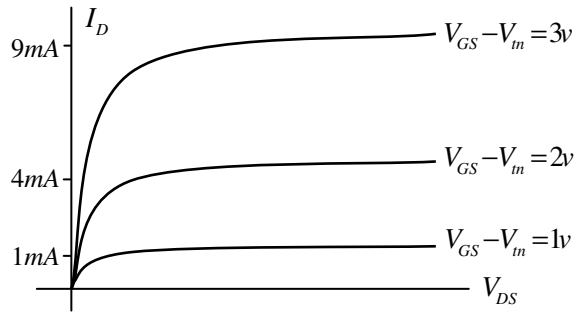
3. Velocity Saturation

→ presents itself in short channel devices. This short channel effect exist in devices with length typically smaller than 0.5µm and is due to a premature saturation of current due to velocity saturation of the carriers (replica of what happens n MESFETs during current saturation)

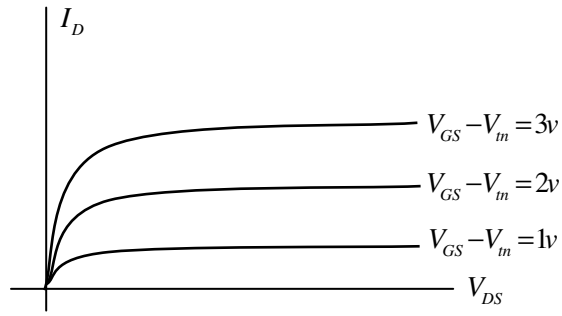
Take triode region equation and substitute V_{DS} with $V_{DS\,sat}$ (constant ~ 0.3v).

$$I_D = \mu_n C_{ox} \frac{W}{L}\left[(V_{GS} - V_{th})V_{DS\,sat} - \frac{V_{DS\,sat}^{\;2}}{2}\right] \tag{10.7}$$

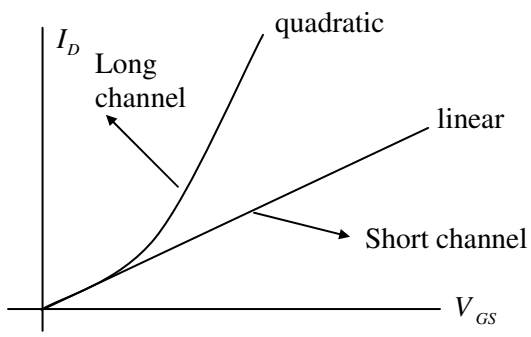
→ The above equation for I_D is no longer a function of V_{DS} .



Long channel device → no velocity saturation.



Short channel device



MOS Small Signal Model

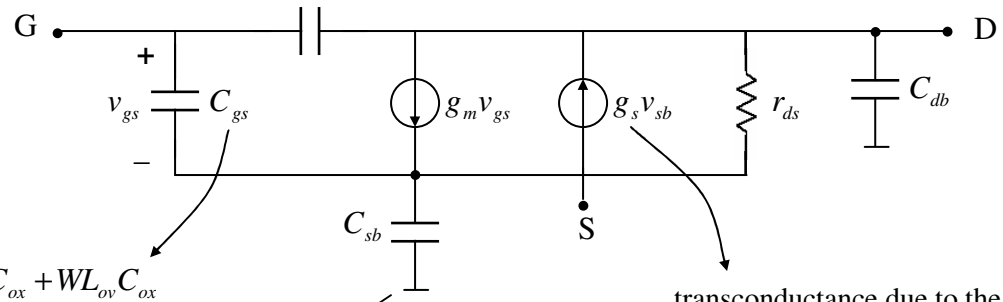
Let us study the transconductance of the device:

$$g_m = \mu_n C_{ox} \frac{W}{L} V_{DS\,sat}$$, for short channel device

(10.8)

$$g_m = \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{tn})$$, for long channel device

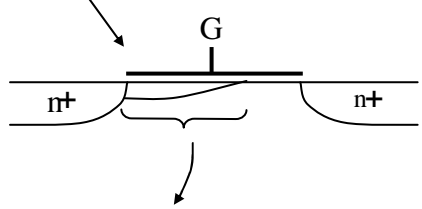
$C_{gd} = C_{ox}WL_{ov}$ → length of overlap region.



$\frac{2}{3}WLC_{ox} + WL_{ov}C_{ox}$

transconductance due to the body effect

body grounded



Approximately $2/3$ of L.

$$g_s = \frac{\partial I_D}{\partial V_{SB}} = \frac{\partial I_D}{\partial V_{tn}} \cdot \frac{\partial V_{tn}}{\partial V_{SB}} \rightarrow g_s = \frac{\gamma g_m}{2\sqrt{V_{SB} + 2\Phi_f}} \quad (10.9)$$

This is a parasitic effect.
If source is grounded do not calculate g_s (although the value from above equation is non-zero), but in reality $\Delta V_{SB} = 0 \rightarrow g_s = 0$.

g_s is the body effect transconductance.

Channel length modulation → $I_D = \mu_n C_{ox} \frac{W}{2L}(V_{GS} - V_{th})^2 \left(1 + \lambda[V_{DS} - (V_{GS} - V_{th})]\right)$

$$\frac{1}{r_{ds}} = \frac{\partial I_D}{\partial V_{DS}} = \lambda \mu_n C_{ox} \frac{W}{2L}(V_{GS} - V_{th})^2 \approx \lambda I_D$$
$$\rightarrow r_{ds} = \frac{1}{\lambda I_D} \quad (10.10)$$

MOSFET Noise and Linearity

We have already studied MOSFET noise and linearity in Chapters 2 and 3. It is worth mentioning here that MOSFETs although have achieved very high speed performance are still rather noisy devices. When a multi-GHz low noise amplifier with sub-dB noise figure

performance is needed, typically PHEMT, MHEMT or InP HBT technologies are the answer. However, in many today's application a noise figure of several dB is acceptable. In such applications it is possible to use CMOS RF circuits for lower cost as well as better integration of the RF circuit with the rest of the system.

In terms of linearity, MOS transistor is no worse than other devices. However, when you combine the facts that advanced MOS devices need to have very short gates and Si bandgap is rather small therefore Si devices experience lower breakdown voltage than their large bandgap counterparts (GaAs and GaN), it becomes very challenging to use MOSFETs for power amplifier applications. Despite this challenge, several groups have used creative circuit techniques to design power amplifiers based on CMOS for mobile phone applications paving the way for one chip solution for a mobile phone.

MOS transistors can be used as switches as well. When used as digital switch they are excellent as they don't consume any power. As RF switch, they are not as attractive as HEMT or MESFET or PIN diode switch due to high losses and nonlinearities that they introduce at RF frequencies.

Advanced MOSFET Processing

One of the main challenges in deep submicron CMOS processing is to reduce the short channel effects. When the channel length becomes shorter, the magnitude of the lateral field (from drain to source) increases significantly. As the drain voltage increases, the depletion charge controlled by the gate field decreases below the long channel value. Even though the depletion region becomes slightly wider due to the application of drain voltage, the threshold voltage decreases.

In addition to the variation of threshold voltage with drain bias, another problem that is observed in short channel devices is that the on/off current ratio varies for different channel lengths. This is due to a varying leakage current when lateral field due to drain voltage becomes comparable to the horizontal field dictated by the gate voltage. To address these problems **HALO** or Pocket doping is introduced. For an NMOS, more highly p-type doped regions near the two ends of the channel known as HALO doping are beneficial to the suppression of the short channel effects

since they help compensate charge sharing effects from the source drain field. Fig. 10.1 depicts how HALO doping is introduced in an advanced MOS device. Halo doping is introduced by a moderate dose p-type implant carried out together with the n+ source and drain implants after gate patterning. This laterally non-uniform channel doping is a self-aligned process, which means it does not complicates the fabrication of the transistor.

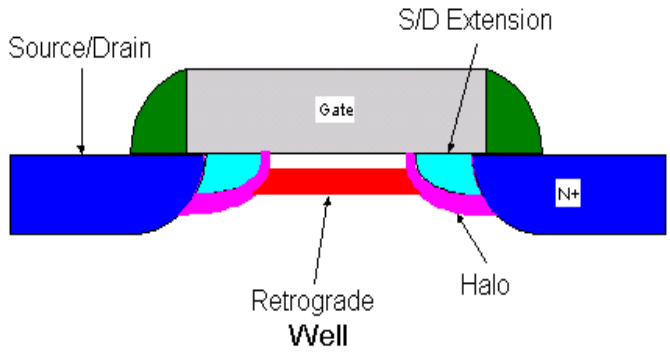


Fig. 10.1: Halo doping in submicron MOS

Quantum Effects, Device statistical variations and Leakage current

As the device critical dimensions are scaled down to less than 65nm, quantum effects and statistical variations from device to device starts to show up. The statistical variation is due to the fact that the channel size becomes so small that the number of dopants in the channel can vary significantly from device to device. As a result, even adjacent devices may not have an identical threshold voltage or current capability. A major effort is needed to estimate and model device statistical variations for circuit design purposes.

Another problem in deep submicron devices is the increase of gate leakage current. For short channel devices, in order to suppress the short channel effects, one needs to scale the gate oxide thickness with the gate length. That means very thin oxide layer, in the order of a few atoms is needed. While it is proven that such thin oxide layer does not introduce any reliability issues, there are problems associated with such thin oxide. Low gate breakdown voltage limits the gate source voltage swing. Moreover, higher leakage current due to quantum tunneling of carriers from gate to channel is expected. While this leakage current may seem very low for one transistor, it becomes a major source of power dissipation when $+10^6$ transistors are utilized.

Semiconductor on Insulator

An interesting approach in CMOS technology is to design MOS transistors on a very thin Si film that is floating on a layer of oxide. The oxide is formed by a shallow Oxygen implantation followed by an annealing step that leaves a top Si layer intact. SOI CMOS devices can be either partially or fully depleted, meaning that there is either little or no substrate except for the channel region. While this introduces kinks in the IV characteristics and makes the modeling of such devices a nightmare, it has a major benefit in reducing the power dissipation of logic gates as floating SOI devices see extremely low capacitances. Mobile CPUs are nowadays processed using SOI technology due to reduced power dissipation.

An additional benefit of CMOS SOI is for RF applications. Since the substrate in an SOI technology can be a high resistivity Si substrate, higher quality factor of passive components such as inductors and transmission lines are achieved. This translates to better RF and microwave circuits implemented on SOI substrate. To date, record high +100GHz circuits have been reported using SOI CMOS technology.

Ultimate Scaling Limit for Digital Circuits

The last word in this chapter is on the ultimate limit of scaling for digital circuits. Through scaling lower supply voltage and higher transistor transconductance per gate capacitance can be achieved. These two facts translate into lower power dissipation when transistor size is scaled. On the other hand an increase in the leakage current becomes an issue. The overall power dissipation of a digital microprocessor is expected to increase and one contributing factor is higher number of transistor, and higher speed of operation. Technologically, it is possible to further reduce the size of the transistors and therefore achieve more compact microprocessors with higher functionality. There is however a huge obstacle in the path. By packing more transistors per unit area, there is more heat generated per unit area of the digital chip. Despite the fact that Si is a very good heat conductor, there is a big challenge to remove so much heat from such a small area.