

Module 2 : MOSFET

Lecture 7: Advanced Topics

Objectives

In this course you will learn the following

- Motivation for Scaling
- Types of Scaling
- Short channel effect
- Velocity saturation

7.1 Motivation for Scaling

The reduction of the dimensions of a MOSFET has been dramatic during the last three decades. Starting at a minimum feature length of 10 μm in 1970 the gate length was gradually reduced to 0.15 μm minimum feature size in 2000, resulting in a 13% reduction per year. Proper scaling of MOSFET however requires not only a size reduction of the gate length and width but also requires a reduction of all other dimensions including the gate/source and gate/drain alignment, the oxide thickness and the depletion layer widths. Scaling of the depletion layer widths also implies scaling of the substrate doping density.

In short, we will study simplified guidelines for shrinking device dimensions to increase transistor density & operating frequency and reduction in power dissipation & gate delays.

7.2 Types of Scaling

Two types of scaling are common:

- 1) constant field scaling and
- 2) constant voltage scaling.

Constant field scaling yields the largest reduction in the power-delay product of a single transistor. However, it requires a reduction in the power supply voltage as one decreases the minimum feature size.

Constant voltage scaling does not have this problem and is therefore the preferred scaling method since it provides voltage compatibility with older circuit technologies. The disadvantage of constant voltage scaling is that the electric field increases as the minimum feature length is reduced. This leads to velocity saturation, mobility degradation, increased leakage currents and lower breakdown voltages.

After scaling, the different Mosfet parameters will be converted as given by table below:

Before Scaling	After Constant Field Scaling	After Constant Voltage Scaling
----------------	------------------------------	--------------------------------

Before Scaling	After Constant Field Scaling	After Constant Voltage Scaling
L	$L' = L/s$	$L' = L/s$
W	$W' = W/s$	$W' = W/s$
t	$t'_{ox} = t_{ox}/s$	$t'_{ox} = t_{ox}/s$
x_i	$x'_i = x_i/s$	$x'_i = x_i/s$
V_{DD}	$V'_{DD} = V_{DD}/s$	$V'_{DD} = V_{DD}$
V_{Th}	$V'_{Th} = V_{Th}/s$	$V'_{Th} = V_{Th}$
N_a or N_d	$N'_a = N_a * s$ or $N'_d = N_d * s$	$N'_a = N_a * s^2$ or $N'_d = N_d * s^2$
C_{ox}	$C'_{ox} = C_{ox} * s$	$C'_{ox} = C_{ox} * s$
I_{DS}	$I'_{DS} = I_{DS}/s$	$I'_{DS} = I_{DS} * s$
P_D	$P'_D = P_D/s^2$	$P'_D = P_D * s$

Where s = scaling parameter of MOS

7.3 Short Channel Effect

So far our discussion was based upon the assumptions that channel was long and wide enough, so that "edge" effects along the four sides was negligible, longitudinal field was negligible and electric field at every point was perpendicular to the surface. So we could perform one-dimensional analysis using gradual channel approximation. But in devices where channel is short longitudinal field will not be negligible compared to perpendicular field. So in that case one-dimensional analysis gives wrong results and we will have to perform dimensional analysis taking into account both longitudinal and vertical fields. (which is out of the scope this course)

When is a channel called a short channel?

- When junction (source/drain) length is of the order of channel length.
- L is not much larger than the sum of the drain and source depletion width.

We have shown below the comparative graphs of **I-V** characteristics for both long channel and short channel length MOSFETs. From graph, it can be clearly concluded that when the channel becomes short, the current in saturation region becomes linearly dependent on applied drain voltage rather than being square dependent.

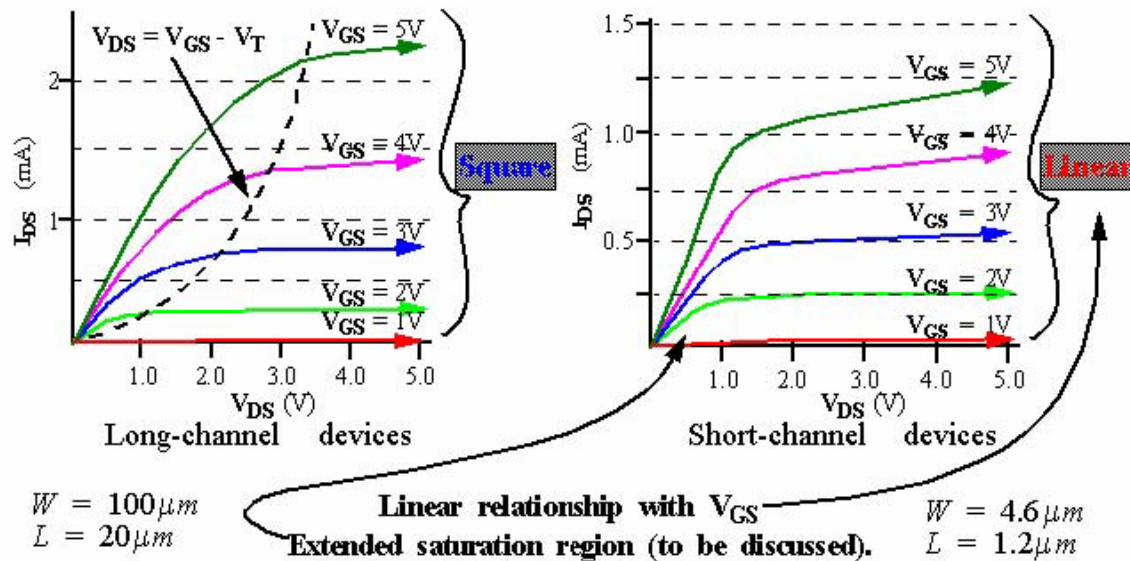
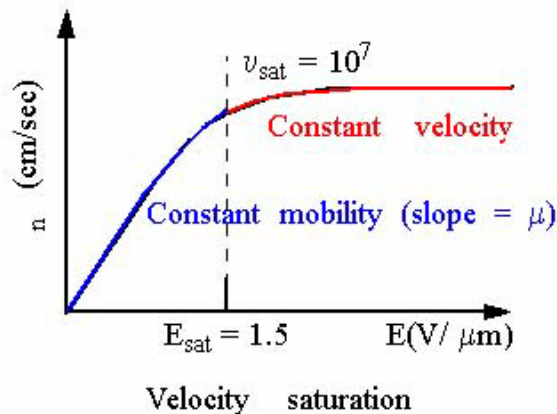


Figure 7.3: Comparison of I_D vs V_{DS} characteristics for long and short channel MOSFET devices

7.4 Velocity Saturation

As we were assuming longitudinal field to be very small in the channel, the magnitude of carrier velocity $|\mathbf{v_d}|$ was proportional to $|\mathbf{E_x}|$. But it has been observed that for high values of $|\mathbf{E_x}|$ carrier velocity tends to saturate. It is no more proportional to $|\mathbf{E_x}|$. This lack of proportionality at high $|\mathbf{E_x}|$ values is known as velocity saturation.



Electrons in p-type silicon:

$$E_{crit} = 1.5 \text{ V}/\mu m$$

$$v_{sat} = 10^7 \text{ cm/sec}$$

Therefore, only a couple volts are needed for NMOS devices with a channel length of $1\mu m$.

Holes in n-type silicon:

$$E_{crit} > 10 \text{ V}/\mu m$$

$$v_{sat} = 10^7 \text{ cm/sec (same)}$$

Recap

In this lecture you have learnt the following

- Motivation for Scaling
- Types of Scaling
- Short channel effect
- Velocity saturation