ECT 201 SSD

MODULE 5

MOSFET scaling – need for scaling, constant voltage scaling and constant field scaling. Sub threshold conduction in MOS. Short channel effects- Channel length modulation, Drain Induced Barrier Lowering, Velocity Saturation, Threshold Voltage Variations and Hot Carrier Effects. Non-Planar MOSFETs: Fin FET –Structure, operation and advantages.

MOSFET SCALING

Scaling of a MOS transistor means reducing the critical parameters of the device in accordance with a given criterion in order to improve some performance features such as Speed, Application, Power Dissipation, and so on while keeping the basic operational characteristics unchanged.

scaling requires all device dimensions to reduce proportionally. The main device dimensions are the channel length, channel width, and oxide thickness.

Lateral dimensions such as channel length and width are reduced by a factor of k, so should the vertical dimensions such as source/drain junction depths and gate insulator thickness

Scaling of depleton width is achieved indirectly by scaling up doping conentrations.If we simply reduce the dimensions of the device and kept the power supply voltages same,the internal electric field in the device would increase.

Need for scaling

Scaling improves

1. Packing density: The packing density of the device improves as a result of scaling hence we can fit more transistors in the same space as before.
2. Speed
3. Power dissipation

Two types of scaling are common:

(i) constant field scaling

(ii) constant voltage scaling.

Full scaling (constant-field scaling) –

- All dimensions are scaled by k and the supply voltage and other voltages are so scaled
- Magnitude of internal electic field is kept constant

- Only lateral dimensions are changed
- Threshold voltage is also affected

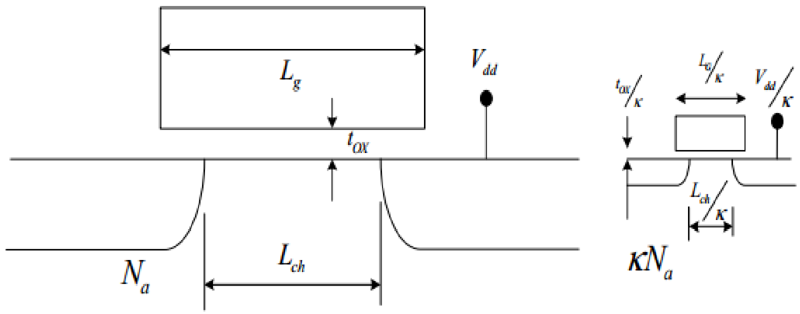


Figure 2: 5 Illustration of MOSFET miniaturisation. The sketch on the right hand is the scaled device according to the constant field rule. (Reference [2.15])

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.

For ideal scaling, power supply voltages should be reduced to keep the internal electric field reasonably constant from one technology generation to the next.But power supply voltages are not scaled hand in hand with the device dimensions, partly because of other system related constraints. The longitudinal electric field in the pinch off region and the transverse electric field across the gate oxide increase with MOSFET scaling which causes hot electron effects and short channel effects.

Table 6–1 Scaling rules for MOSFETs according to a constant factor K. The horizontal and vertical dimensions are scaled by the same factor. The voltages are also scaled to keep the internal electric fields more or less constant, and the hot carrier effects manageable.

	Scaling factor
Surface dimensions (L,Z)	1/K
Vertical dimensions (d, x_j)	1/K
Impurity Concentrations	K
Current, Voltages	1/K
Current Density	K
Capacitance (per unit area)	K
Transconductance	1
Circuit Delay Time	1/K
Power Dissipation	$1/K^2$
Power Density	1
Power-Delay Product	$1/K^3$

Constant voltage scaling does not have this problem and is therefore the preferred scaling method since it provides voltage compatibility with older circuit technologies.

Constant-voltage scaling: In Constant voltage scaling, all dimensions of the MOSFET are reduced by a factor of ‘k’, but power supply & terminal voltage remain unchanged.

The voltages are not scaled and, in some cases, dimensions associated with voltage are not scaled.

The disadvantage of constant voltage scaling is that the electric field increases as the minimum feature length is reduced.

Constant voltage scaling

Parameter	Scaled parameter
Channel length (L)	$1/\alpha$
Junction depth (x_j) Substrate doping (N_A) Depletion layer thickness (d)	$1/\alpha$ α $1/\alpha$
Transconductance (g_m)	α
Static power dissipation (P_{stat})	α
Dynamic power dissipation (P_{dyn})	α
Current (I)	α
Gate delay (τ_p)	$1/\alpha^2$
Load capacitance (C_L)	$1/\alpha$
Channel width (W)	$1/\alpha$
Supply voltage (V)	1
Gate oxide thickness (t_{ox})	$1/\alpha$
Current density (J)	α^3

Comparison

Quantity	Sensitivity	Constant Field	Constant Voltage
Scaling Parameters			
Length	L	$1/S$	$1/S$
Width	W	$1/S$	$1/S$
Gate Oxide Thickness	t_{ox}	$1/S$	$1/S$
Supply Voltage	V_{dd}	$1/S$	1
Threshold Voltage	$\boldsymbol{V_{T0}}$	$\boldsymbol{1/S}$	$\boldsymbol{1}$
Doping Density	N_A, N_D	S	S^2
Device Characteristics			
Area (A)	WL	$1/S^2$	$1/S^2$
D-S Current (I_{DS})	$\beta(V_{dd} - v_T)^2$	$1/S$	S
Gate Capacitance (C_g)	WL/t_{ox}	$1/S$	$1/S$
Power Dissipation (P)	$I_{DS}V_{dd}$	$1/S^2$	S
Power Dissipation Density (P/A)	P/A	1	S^3

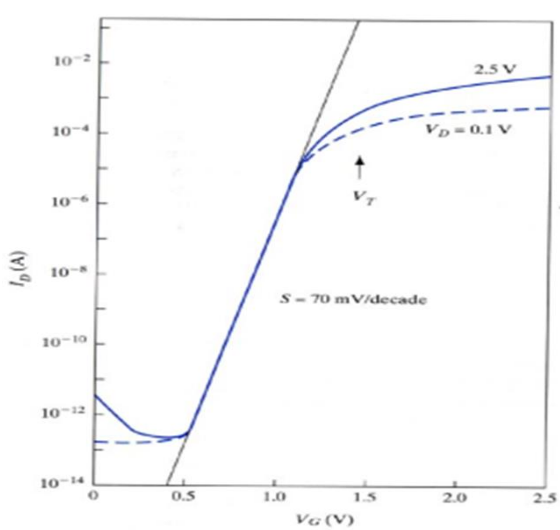
Subthreshold conduction / Subthreshold Characteristics

For the conduction to happen in the MOSFET; we need 'V_G' to be greater than the threshold voltage V_T. But this threshold voltage is calculated at the point where the region below the oxides has entered into strong inversion.

From experimental results, one can observe that there is still some non-zero current flowing from drain to source even when we are operating at a region with $V_G < V_T$ (sub-threshold region). This happens because, for the subthreshold region, the substrate near oxide-interface is in "Weak-Inversion". At this point, if we apply a positive V_{DS}, there will be a small current I_D flowing. This effect is plotted in the transfer characteristics in figure below. We have,

$$I_D = \frac{W}{L} \cdot \mu_n C_{OX} \left[(V_{GS} - V_T \ ) V_{DS} - \frac{V_{Ds}^2}{2} \right]$$

In this equation, current abruptly goes to zero as soon as V_G is reduced to V_T. In reality there is still some drain conduction below threshold, and is known as subthreshold conduction. This current is due to weak inversion in the channel between flat band and threshold which leads to a diffusion current from source to drain.



Subthreshold Swing,

$$S = \log \frac{d(V_G)}{d(\log I_D)}$$

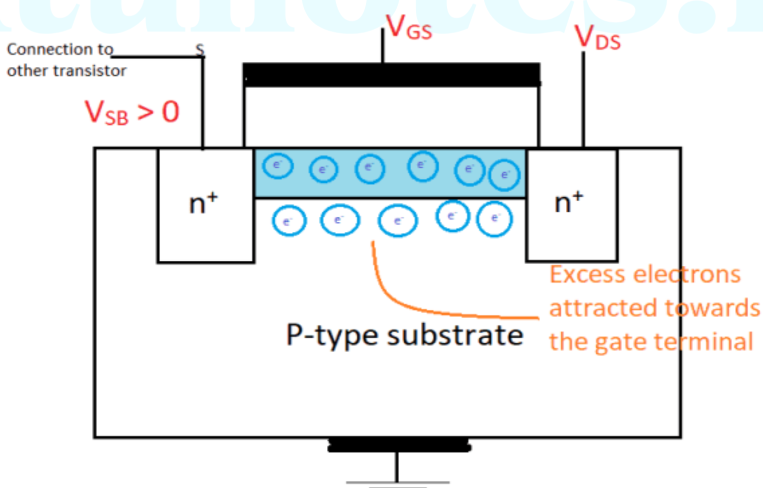
- Generally, in order to improve the performance and reduce the cost of production, one would prefer to scale down the size of the transistors.
- This scaling down also eliminates many stray capacitances that are present in the overall device. Ultimately increasing the speed of operation.
- But when the channel length is scaled down to the order of the depletion layer, a certain number of non-ideal effects come into play. These are called **second-order effects**

Substrate Bias Effect

For the ideal I-V characteristics, the biasing scheme we used had the source and the body both connected to ground.

But in practical design applications, Source is connected to substrate(body) so that there is a voltage V_{SB}

In such scenarios, the difference in potential between the body and the source terminal causes a change in the threshold voltage of the MOSFET. This effect of change in threshold voltage is called the "Body Effect" or the "Back Gate Effect".



When V_{SB} is positive, there is reveres bias between source and bulk. This causes depletion layer to widen.

The electrons in the bulk are repelled by the body terminal and are now attracted by the gate toward the oxide layer.

The threshold voltage of the MOS is also proportional to the density of electrons in the depletion layer.

Hence as we accumulate more and more electrons in the depletion layer below the oxide interface, there will be an increase in the value of threshold voltage.

As depletion region is widened, larger charge density is occupied. Therefore, the threshold required to achieve inversion increases.

$$V_{TN} = V_{T0} + \gamma\left(\sqrt{2|\phi_F| + V_{SB}} - \sqrt{2|\phi_F|}\right)$$
$$V_{T0} = \text{ zero-substrate-bias } V_T$$
$$\gamma = \text{ body effect parameter}$$
$$|\phi_F| = \text{ surface potential parameter}$$

Short Channel Effects

Short-channel effects occur in MOSFETs in which the channel length is comparable to the depletion layer widths of the source and drain junctions.

A MOSFET device is considered to be short channel device when the channel length is the same order of magnitude as the depletion-layer widths (x_{dD}, x_{dS}) of the source and drain junction. (That is, the effective channel length L_{eff} is approximately equal to the source and drain junction depth x).

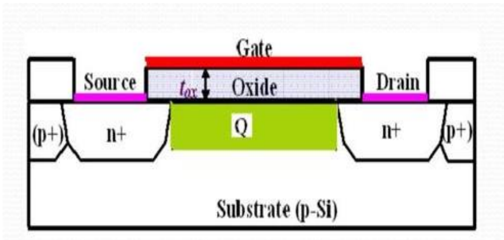
As the channel length L is reduced to increase both the operation speed and the number of components per chip, the so-called short-channel effects arise.

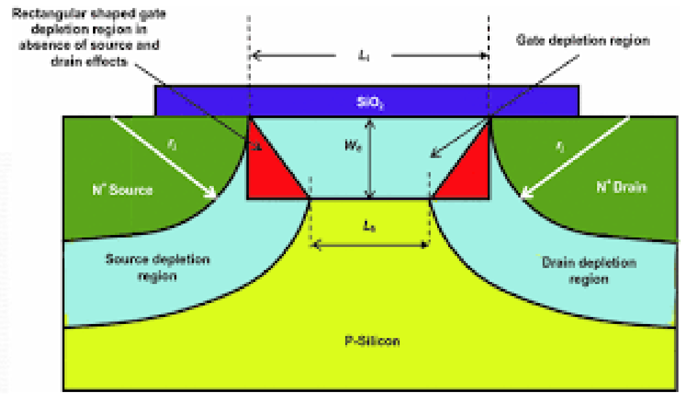
The short-channel effects are attributed to two physical phenomena:

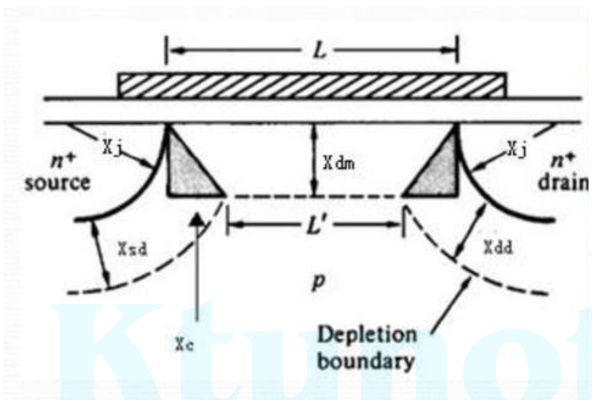
1. The limitation imposed on electron drift characteristics in the channel
2. The modification of the threshold voltage due to the shortening channel length.

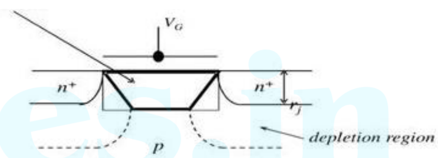
This occurs due to the charge sharing between source/drain and gate. A triangle region forms at both ends

Hence the rectangular area under the gate becomes Trapizoid









Different short-channel effects include

1. Channel Length Modulation

2. Drain-induced barrier lowering and “Punch through”

3.Velocity saturation

4.Threshold voltage variations

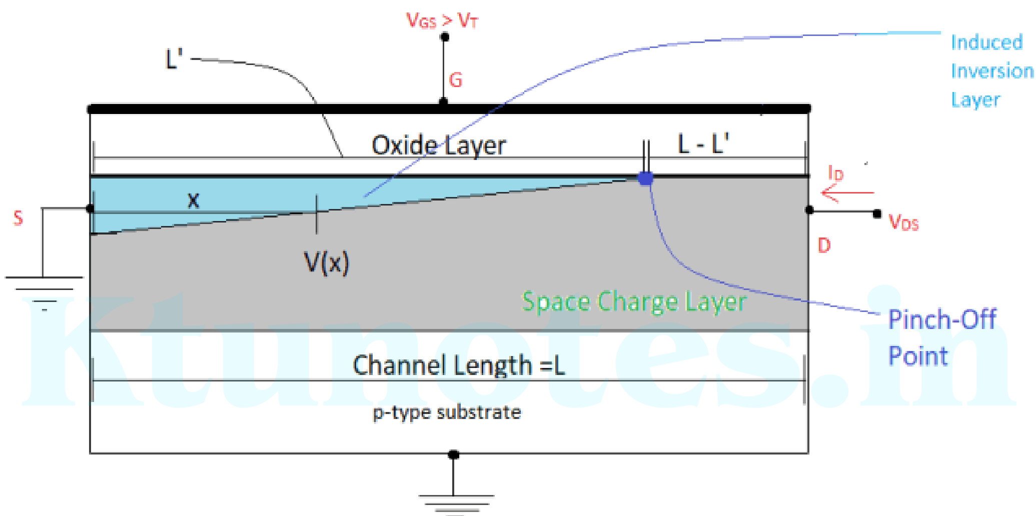
5. Hot carrier effects

Channel Length Modulation(CLM)

As we keep on increasing V_{DS}, the region for which the inversion charge is zero keeps on increasing for a constant value of V_{GS} maintained. Thus channel length keeps on decreasing. This phenomenon is called Channel Length Modulation.

This is a similar to "Base Width Modulation" Thus we get a V_{DS} term in the expression for I_D even when we are operating in the saturation region.

Generally, the fabrication of the MOSFET devices is done in a way such that the change in length given by $\Delta L = L - L'$ is low with a change in V_{DS}.



Drain Induced Barrier Lowering (DIBL)

When the depletion regions surrounding the drain extends to the source, so that the two-depletion layer merge (i.e., when $x_{dS} + x_{dD} = L$), punch through occurs.

Punch through can be minimized with thinner oxides, larger substrate doping, shallower junctions, and obviously with longer channels.

The current flow in the channel depends on creating and sustaining an inversion layer on the surface.

If the gate bias voltage is not sufficient to invert the surface ($V_{GS} < V_T$), the carriers (electrons) in the channel face a potential barrier that blocks the flow. Increasing the gate voltage reduces this potential barrier and, eventually, allows the flow of carriers under the influence of the channel electric field.

In small-geometry MOSFETs, the potential barrier is controlled by both the gate-to-source voltage V_{GS} and the drain-to-source voltage V_{DS}.

If the drain voltage is increased, the potential barrier in the channel decreases, leading to drain-induced barrier lowering (DIBL).

The reduction of the potential barrier eventually allows electron flow between the source and the drain, even if the gate-to-source voltage is lower than the threshold voltage.

The channel current that flows under this condition ($V_{GS} < V_T$)is called the sub-threshold current

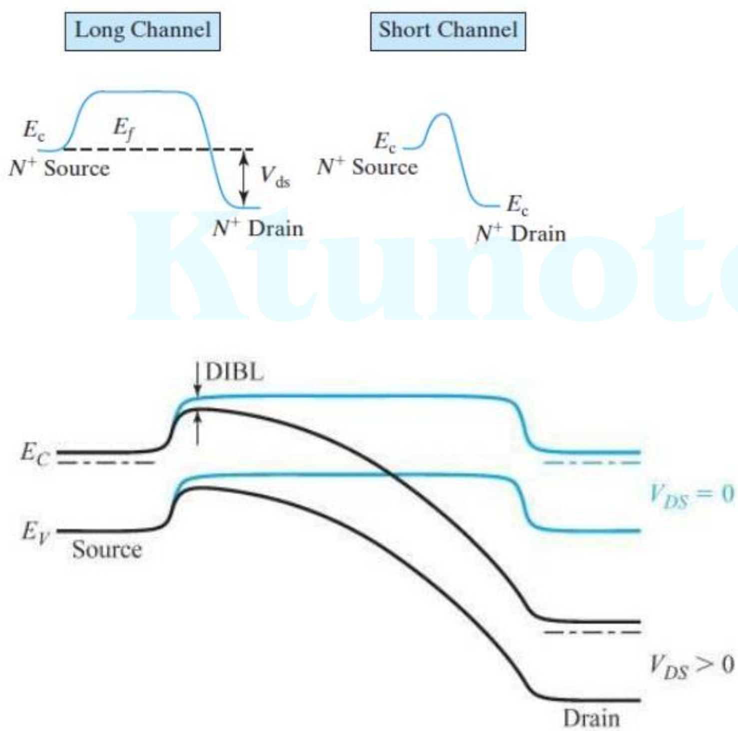


Figure 8.15 Energy-band diagrams along the channel of a MOSFET illustrating the drain-induced barrier lowering (DIBL).

<u>**Velocity Saturation**</u>

The velocity of charge carriers, such as electrons or holes, is proportional to the electric field that drives them, but that is only valid for small fields.

As the field gets stronger, their velocity tends to saturate. That means that above a critical electric field, they tend to stabilize their speed and eventually cannot move faster.

Velocity saturation is specially seen in short-channel MOSFET transistors, because they have higher electric fields

The drift velocity of the electrons in the inversion layer to be proportional to the lateral electric field applied. The proportionality constant was given by μ_n.

The key point to understand the effect of velocity saturation is that the linearity of the drift velocity only holds true for low values of the applied electric field. The actual variation of drift velocity with respect to the applied electric field is shown in figure 6.

The exact formula for the drift velocity can be given as:

$$v_d = \frac{\mu E}{1+E/E_c}$$

The term E_c is called the critical electric field. Here the electric field E is equal to $\frac{V_{DS}}{L}$, i.e. the lateral voltage applied across the channel divided by the effective channel length. We can see that for large channel devices, the drift velocity formula simplifies to $v_d = \mu E$. Hence this is also a short channel effect because the lateral electric field is higher in case of short channel devices for similar range of drain-to-source voltage applied.

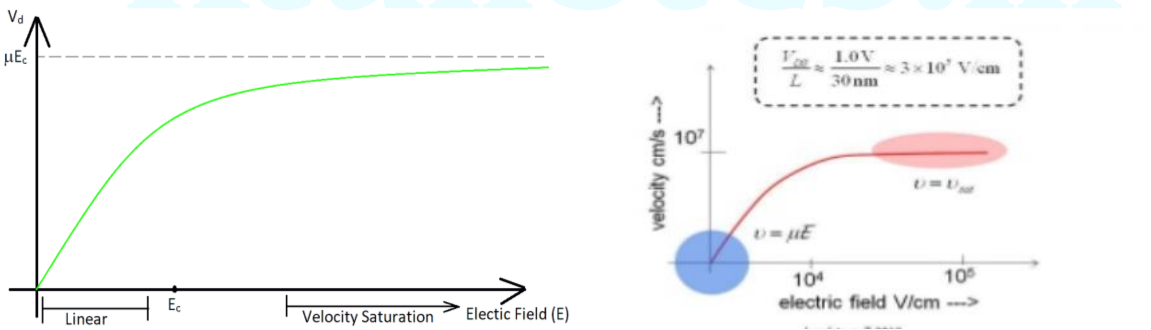


Figure 6: Variation of drift velocity of electron w.r.t. applied electric field

<u>Threshold Variations</u>

The threshold voltage is only a function of the manufacturing technology and the applied body bias V_{SB}.

The threshold can therefore be considered as a constant over all NMOS (PMOS) transistors in a design. As the device dimensions are reduced, this model becomes inaccurate, and the threshold potential becomes a function of L, W, and V_{DS}.

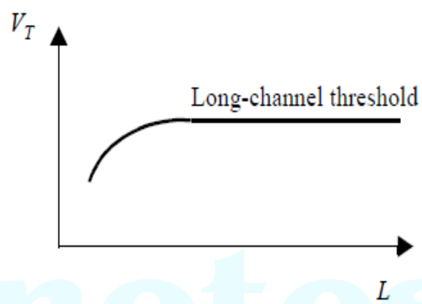
Two-dimensional second-order effects that were ignorable for long-channel devices suddenly become significant.

In the traditional derivation of the V_{TO}, for instance, it is assumed that the channel depletion region is solely due to the applied gate voltage and that all depletion charge beneath the gate originates from the MOS field effects.

This ignores the depletion regions of the source and reverse-biased drain junction, which become relatively more important with shrinking channel lengths.

Since a part of the region below the gate is already depleted (by the source and drain fields), a smaller threshold voltage suffices to cause strong inversion.

In other words, V_{T0} decreases with L for short-channel devices (Figure 3.35a).



(a) Threshold as a function of the length (for low V_{DS})

Figure 3.35 Threshold variations.

Hot-Carrier Effects

Another problem, related to high electric fields, is caused by so-called hot electrons. This high energy electrons can enter the oxide, where they can be trapped, giving rise to oxide charging that can accumulate with time and degrade the device performance by increasing V_T and affect adversely the gate’s control on the drain current.

Besides varying over a design, threshold voltages in short-channel devices also have the tendency to drift over time. This is the result of the hot-carrier effect

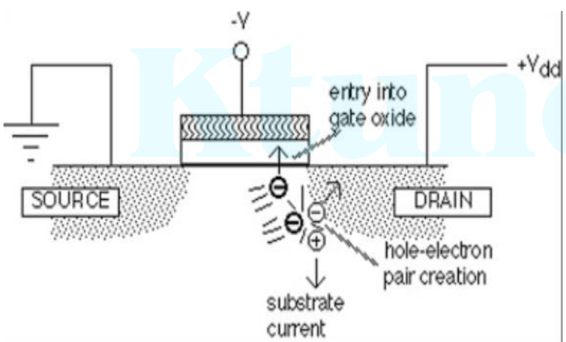
Over the last decades, device dimensions have been scaled down continuously, while the power supply and the operating voltages were kept constant. The resulting increase in the electrical field strength causes an increasing velocity of the electrons, which can leave the silicon and tunnel into the gate oxide upon reaching a high-enough energy level.

Electrons trapped in the oxide change the threshold voltage, typically increasing the thresholds of NMOS devices, while decreasing the V_T of PMOS transistors.

For an electron to become hot, an electrical field of at least 10^4 V/cm is necessary. This condition is easily met in devices with channel lengths around or below 1 mm.

The hot-electron phenomenon can lead to a long-term reliability problem, where a circuit might degrade or fail after being in use for a while. This is illustrated in Figure 3.36, which shows the degradation in the I-V characteristics of an NMOS transistor after it has been subjected to extensive operation.

MOSFET technologies, therefore use specially-engineered drain and source regions to ensure that the peaks in the electrical fields are bounded, hence preventing carriers to reach the critical values necessary to become hot.



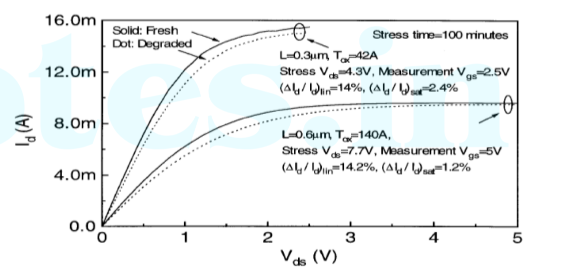


Figure 3.36 Hot-carrier effects cause the *I-V* characteristics of an NMOS transistor to degrade from extensive usage (from [McGaughy98]).

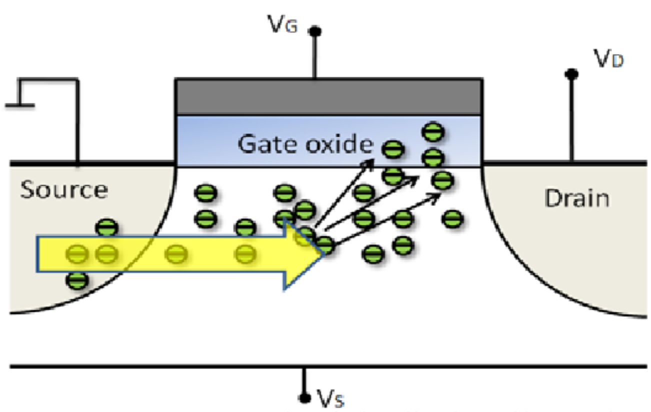


Figure 1. MOSFET cross-sectional visualization of hot carrier effect.

FinFET

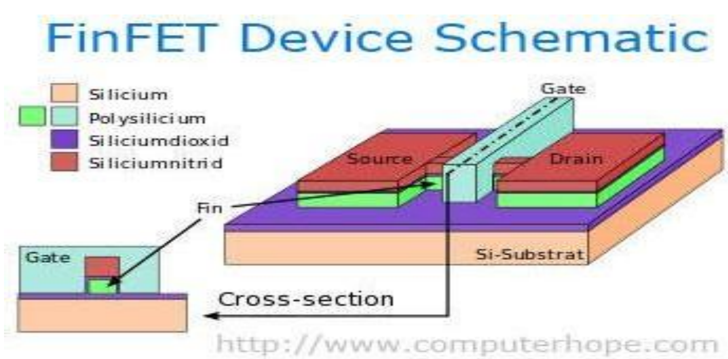
- FinFET, also known as Fin Field Effect Transistor, is a type of non-planar or "3D" transistor used in the design of modern processors
- FinFETs are new generation transistors which utilize tri-gate structure. In contrast to planar transistors where the Gate electrode was (usually) above the channel, the Gate electrode "wraps" the channel
- The distinguishing characteristic of the FinFET is that the conducting channel is wrapped by a thin silicon "fin", which forms the body of the device.
- The conducting channel is greatly controlled by the gate.
- The thickness of the fin (measured in the direction from source to drain) determines the effective channel length of the device.
- These effects make it harder for the voltage on a gate electrode to deplete the channel underneath and stop the flow of carriers through the channel – in other words, to turn the transistor Off. By raising the channel above the surface of the wafer instead of creating the channel just below the surface, it is possible to wrap the gate around up to three of its sides, providing much greater electrostatic control over the carriers within it.

Advantages

Reduced short-channel effects (SCEs) and **leakage current** .

To overcome the worst types of short-channel effect encountered by deep submicron transistors, such as drain-induced barrer lowering (DIBL).

This technique provides increased operating speed by low-threshold MOSFET and **reduced leakage** by high-threshold voltage.



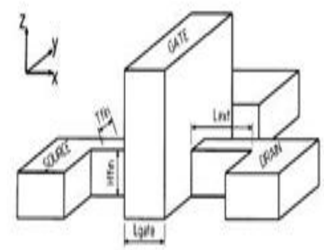
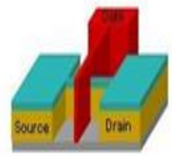
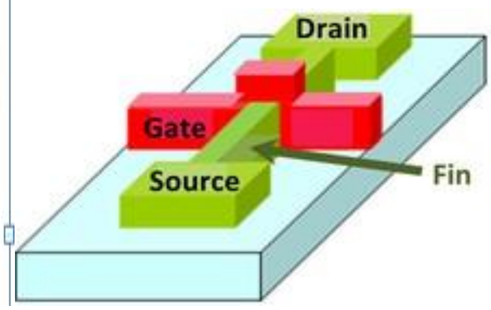


Fig. 1. Schematic of a FinFET structure



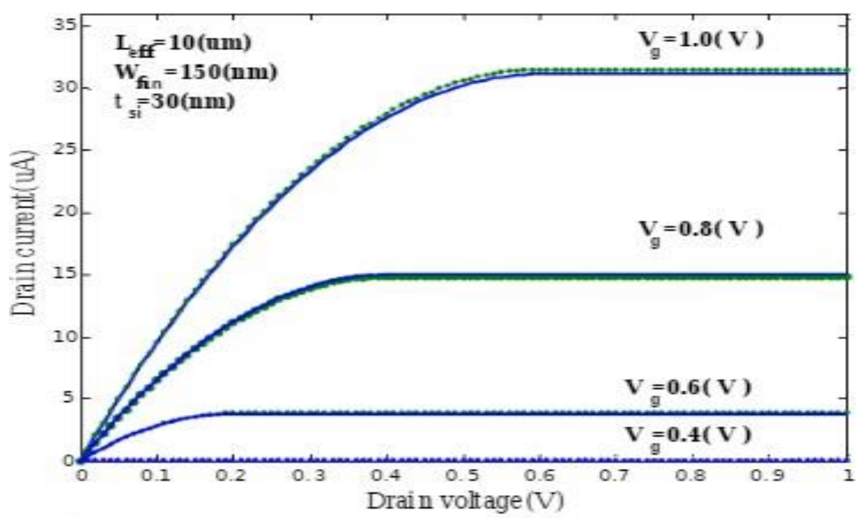
3D view of FinFET



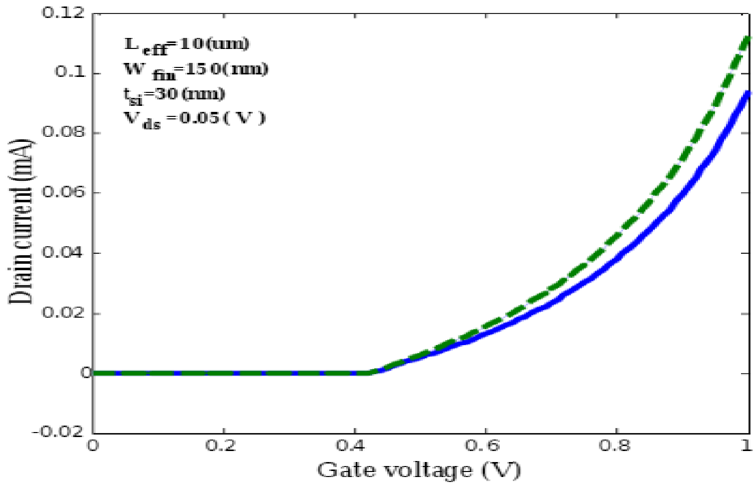
- The very first finFETs were manufactured on top of insulating layer.
- The fact that the current can't flow "underneath" the gate when the transistor is in OFF state reduces the leakage current.
- Alternative techniques for stopping leakage current from flowing in the bulk were introduced later, which allowed for manufacturing of Bulk finFETs.
- This technique utilizes very high doping gradients along the height of the fin in order to prevent the current from flowing in the bulk.

Output Characteristics

Drain Characteristics



Transfer Characteristics



PREVIOUS YEAR QUESTIONS

DECEMBER 2020

1. What is channel length modulation in MOSFETs? How does it affect the output characteristics of the MOSFET? (3)
2. Explain the principle of operation and advantage of FinFET.
3. What is meant by scaling in MOSFETs? Explain the challenges in device scaling? (7)
4. Explain the concept of constant voltage scaling and its limitations. (7)
5. What is meant by DIBL in MOSFETs? How does it affect the threshold voltage of a MOSFET? (7)
6. Explain the concepts of velocity saturation and hot carrier effects in a MOSFET. (7)

DECEMBER 2021

1. Explain Drain induced barrier lowering? (3)
2. Draw and label the structure of a FINFET (3)
3. Distinguish between constant voltage scaling and constant field scaling (8)
4. Illustrate the operation of FinFET (6)
5. Explain any four short channel effects in MOSFET (14)

****