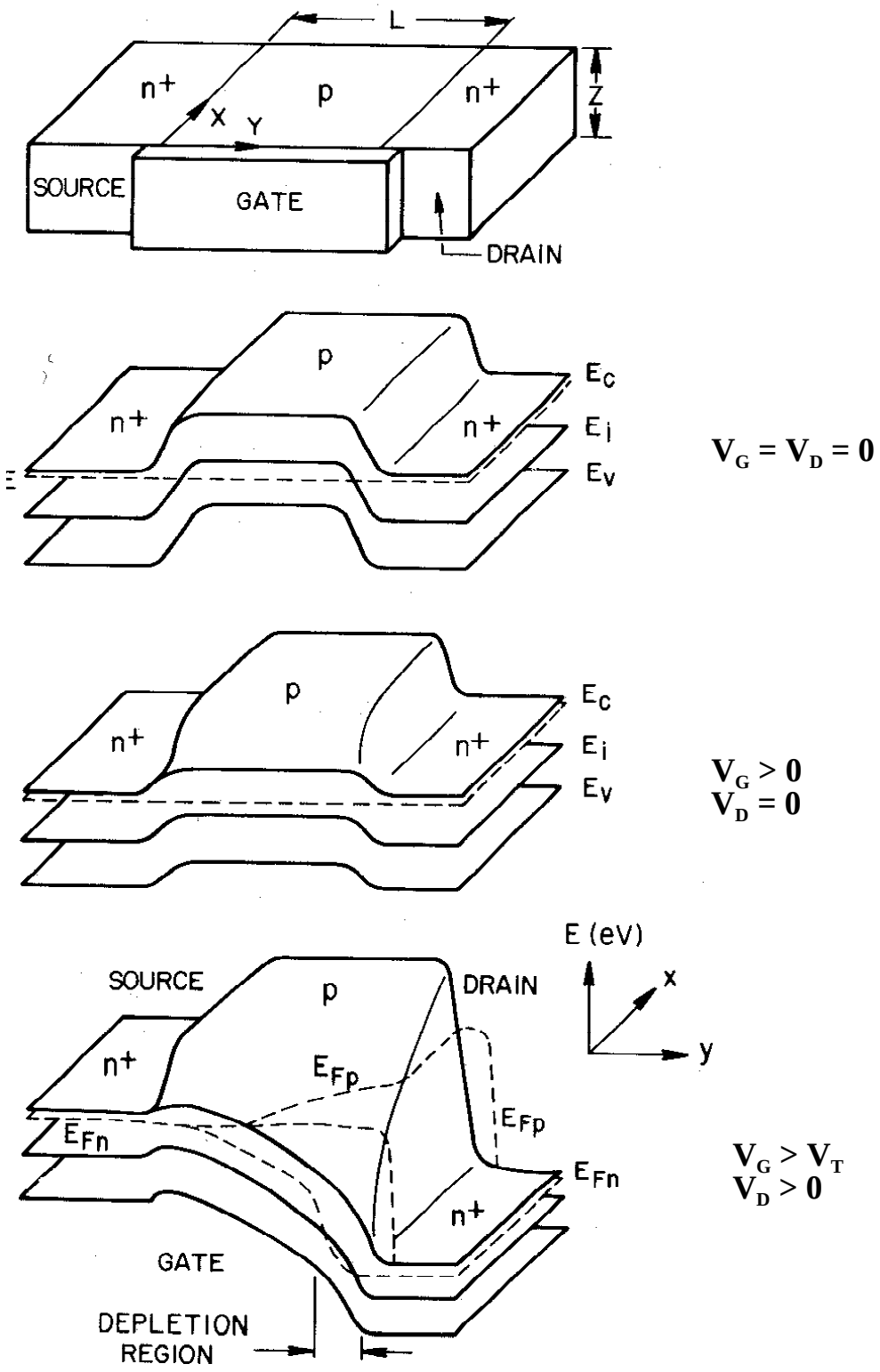
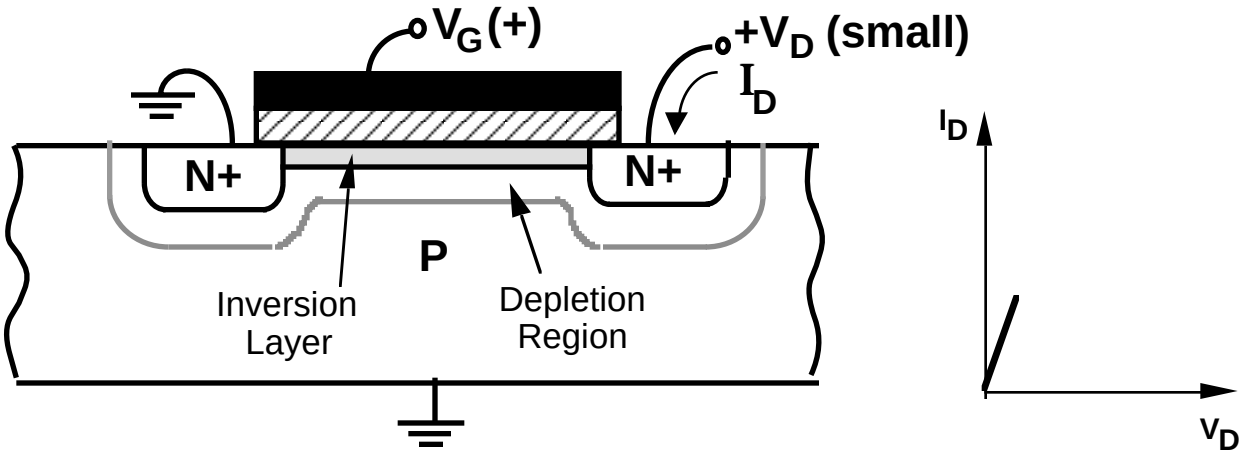
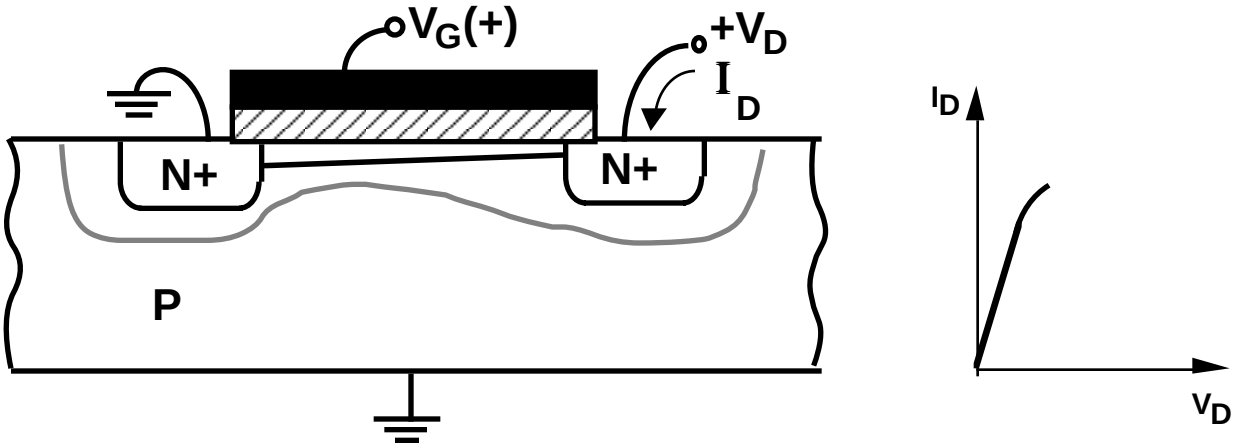
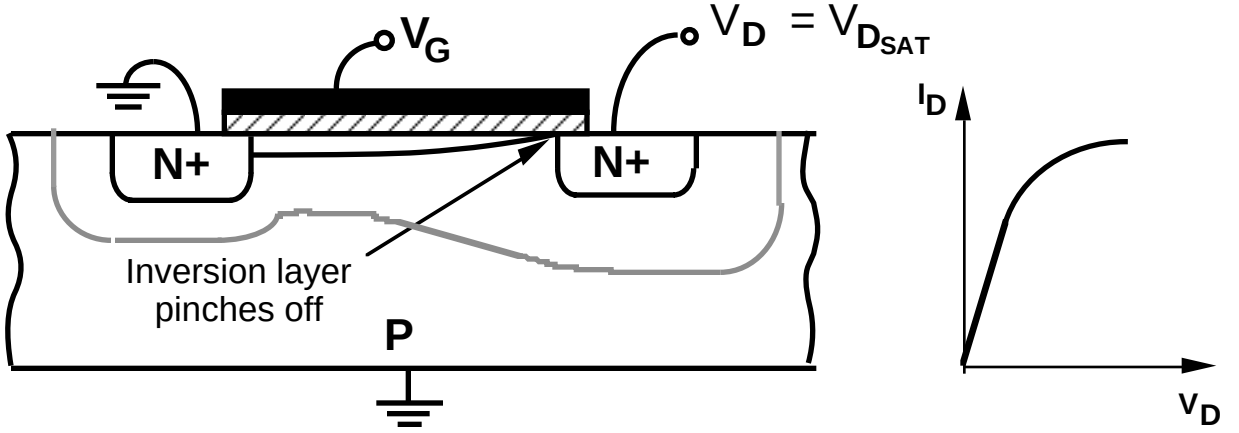
MOS TRANSISTOR REVIEW

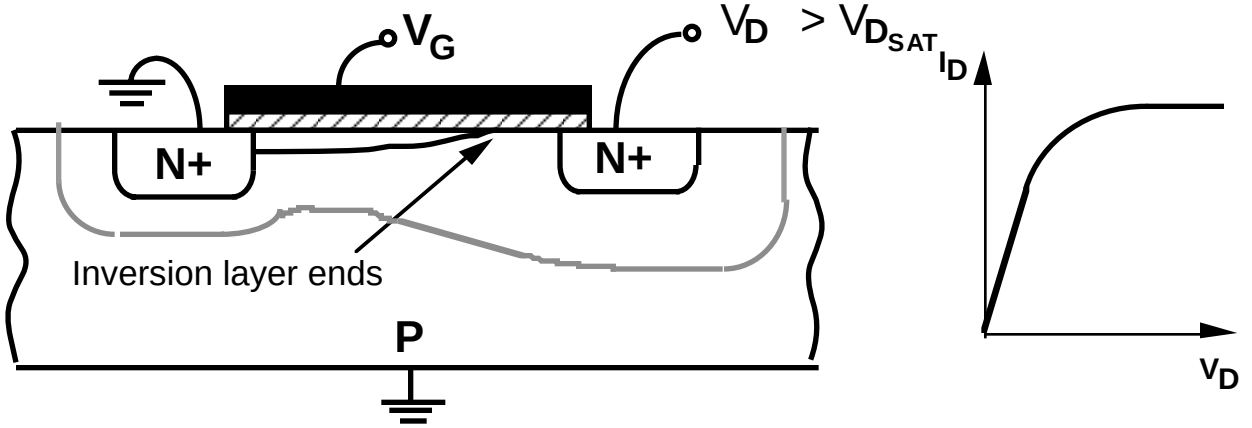


3D band diagram of a long channel enhancement mode NMOS transistor


$V_G(+)$
$+V_D$ (small)
I_D
N+
N+
P
Inversion Layer
Depletion Region
I_D
V_D


$V_G(+)$
$+V_D$
I_D
N+
N+
P
I_D
V_D


V_G
$V_D = V_{D_{SAT}}$
N+
N+
Inversion layer pinches off
P
I_D
V_D


V_G
$V_D > V_{D_{SAT}}$
N+
N+
Inversion layer ends
P
I_D
V_D

From EE216 notes:

Drain current

$$I_D = \iint J_x dydz = -W \int Q_I \mu_n E_x dy$$

Charge density in the channel:

$$Q_I(y) = - C'_{ox} \left[V_G - V_T(y)\right]$$

Gate voltage required to induce inversion under the influence of V_D

$$V_T(y) = V_{FB} + \frac{1}{C'_{ox}} \sqrt{2\ \varepsilon_s\ q\ N_a \left[-2\phi_p - \left(V_B - V(y)\right)\right]} - 2\phi_p + V(y)$$

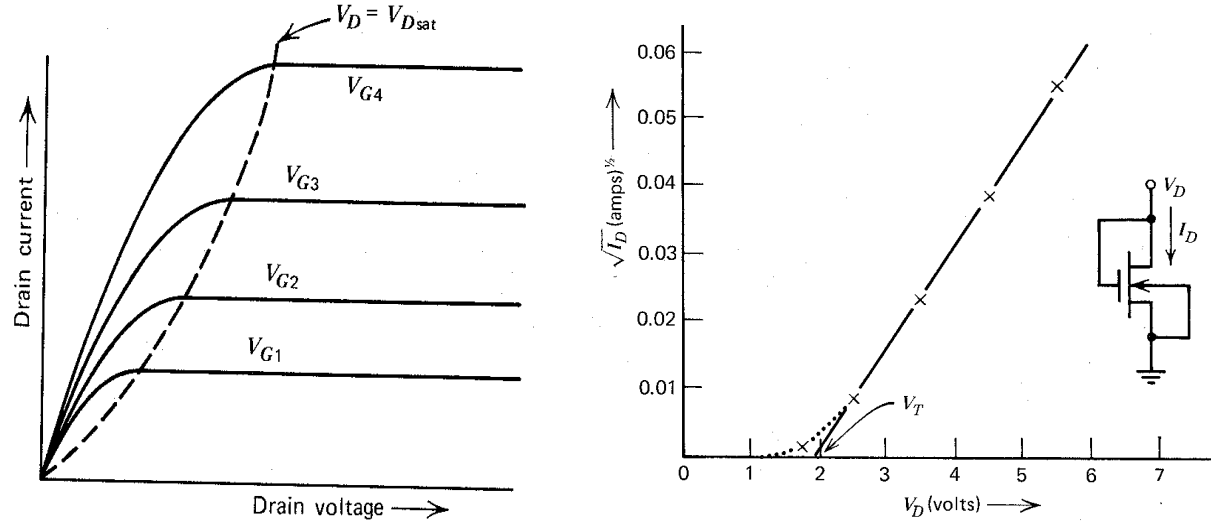
Solving the above 3 equations we get $I_D - V_D$ characteristics

Linear Region

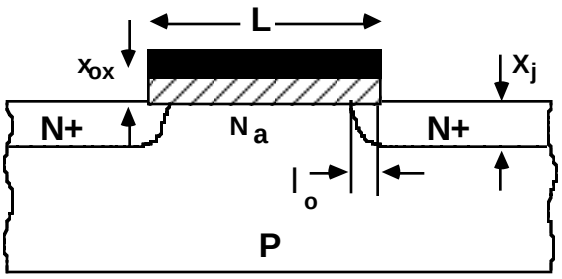
$$I_D = \frac{W}{L} \mu_n\ C'_{ox} \left[V_G - V_T\right] V_D$$

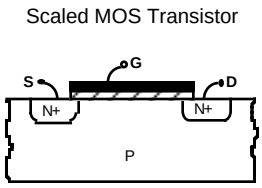
Saturation Region

$$I_{D_{SAT}} \approx \frac{W}{2\ L} \mu_n \frac{\varepsilon_{ox}}{t_{ox}} \left(V_G - V_T\right)^2$$



MOS Device Scaling





Why do we scale MOS transistors?

1. Increase device packing density
2. Improve frequency response (transit time) $\alpha \quad \frac{1}{L}$
3. Improve current drive (transconductance g_m)

$$g_m = \left.\frac{\partial I_D}{\partial V_G}\right|_{V_D = const}$$

$$\approx \frac{W}{L}\ \mu_n\ \frac{K_{ox}}{t_{ox}}\ V_D \qquad \text{for } V_D < V_{D_{SAT}},\ \textit{linear region}$$

$$\approx \frac{W}{L}\ \mu_n\ \frac{K_{ox}}{t_{ox}}\ (V_G - V_T) \quad \text{for } V_D > V_{D_{SAT}},\ \textit{saturation region}$$

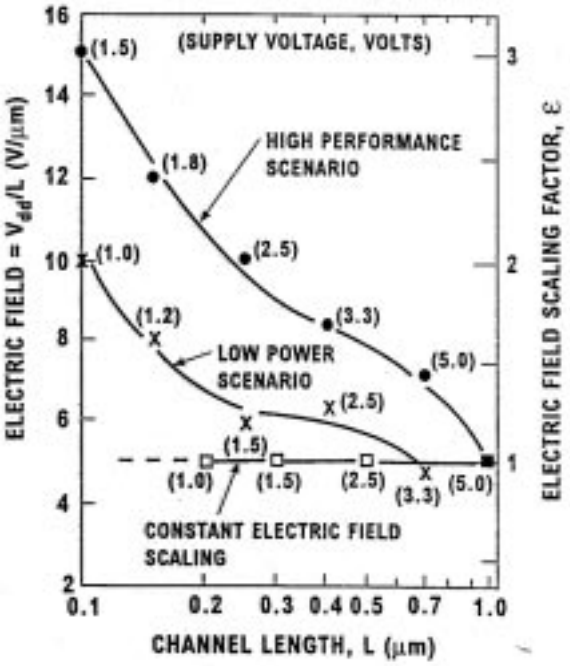
Decreasing the channel length and gate oxide thickness increases g_m, i.e., the current drive of the transistor. Much of the scaling is therefore driven by decrease in L and t_{ox}. However if only these two parameters are scaled many problems are encountered, e.g., increased electric field.

In 1974 Dennard proposed a scaling methodology which maintains the electric field in the device constant. (Dennard et al., IEEE JSSCC, 9, pp. 256-268, 1974)

Device/Circuit Parameter		**Constant Field Scaling Factor**
Dimension :	x_{ox}, L, W, X_j,	1/K
Substrate doping :	N_a	K
Supply voltage :	V	1/K
Supply current :	I	1/K
Gate Capacitance :	W L/x_{ox}	1/K
Gate delay :	C V / I	1/K
Power dissipation :	C V^2 / delay	$1/K^2$

In reality constant field scaling has not been observed strictly. Since the transistor current is proportional to the gate overdrive (V_G-V_T), high performance demands have dictated the use of higher supply voltage. However, higher supply voltage implies increased power dissipation (CV^2f). In the recent past low power applications have become important and have required a scaling scenario with lower supply voltage.

PARAMETER	**1970**	**1980**	**1990**	**2000**	**2006**
Channel length (μm)	10	4	1	0.18	0.1
Gate oxide (nm)	120	50	15	4	1.5
Junction depth (μm)	>1	0.8	0.3	0.08	0.02-0.03
Power supply voltage	12	5	3.3 - 5	1.5-1.8	0.6-0.9



Ref: Davri, et al. Proc. IEEE, April 1995

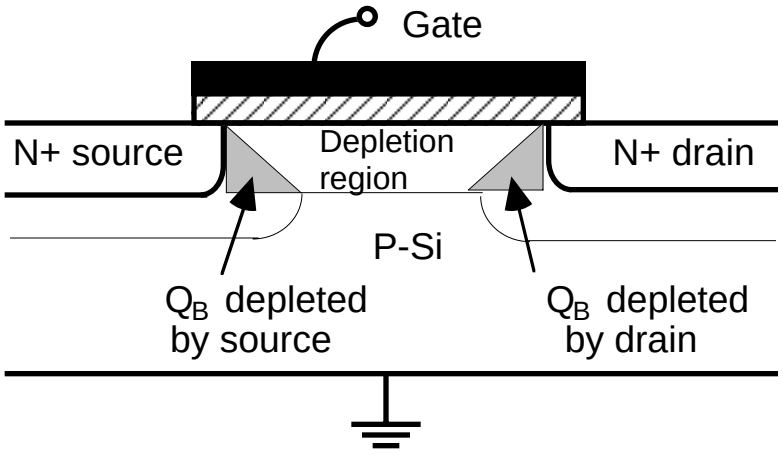
Device/Circuit Parameter		**Quasi Constant Voltage Scaling (K>B>1)**
Dimension :	L, W, x_{ox}, X_j,	1/K
Substrate doping :	N	K
Supply voltage :	V	1/B

Limitations of Scaled MOSFET

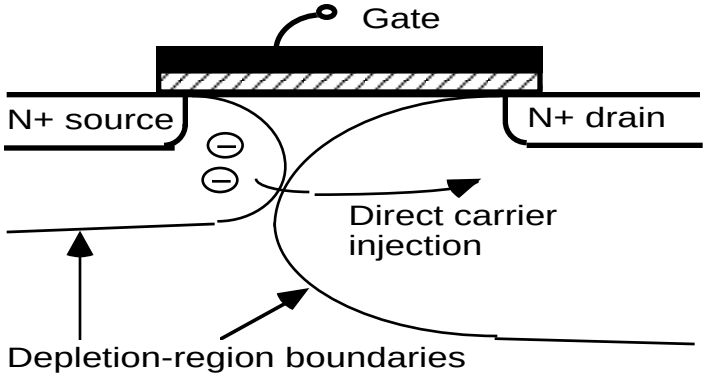
Effect of Reducing Channel Length: Drain Induced Barrier Lowering (DIBL)

In devices with long channel lengths, the gate is completely responsible for depleting the semiconductor (Q_B). In very short channel devices, part of the depletion is accomplished by the drain and source bias

Since less gate voltage is required to deplete Q_B, $V_T\downarrow$ as $L\downarrow$. Similarly, as $V_D\uparrow$, more Q_B is depleted by the drain bias, and hence $V_T\downarrow$. These effects are particularly pronounced in lightly doped substrates.

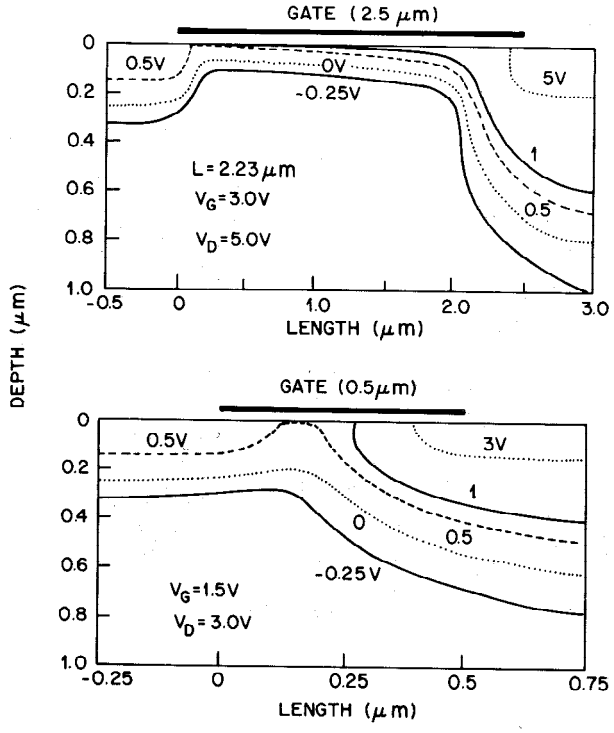


If the channel length becomes too short, the depletion region from the drain can reach the source side and reduces the barrier for electron injection. This is known as punch through.



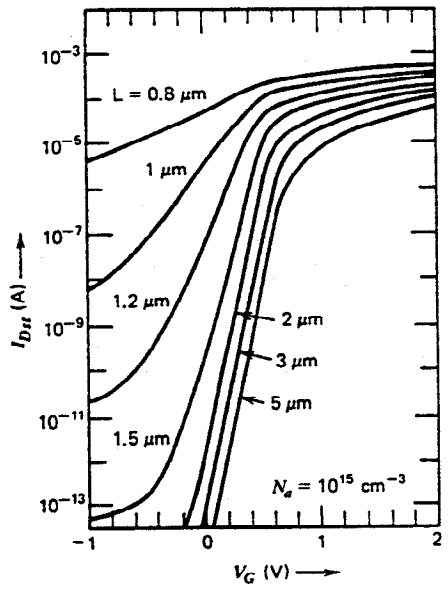
In devices with long channel lengths, the gate is completely responsible for depleting the semiconductor (Q_B). In very short channel devices, part of the depletion is accomplished by the drain and source bias. Since less gate voltage is required to

deplete Q_B, the barrier for electron injection from source to drain decreases. This is known as *drain induced barrier lowering (DIBL).*

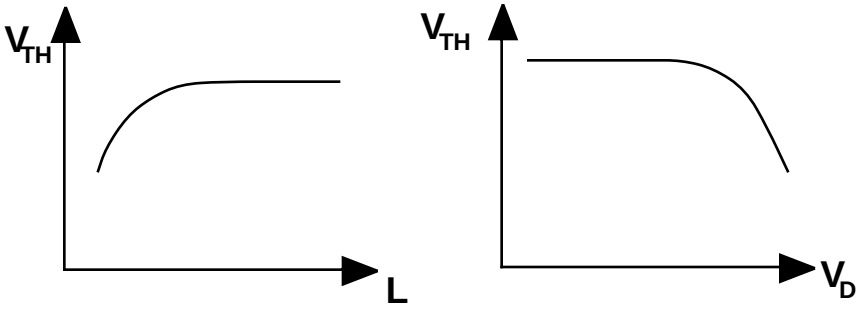


Potential variation along the channel illustrating drain induced barrier lowering (DIBL).

DIBL results in an increase in drain current at a given V_G. Therefore $V_T\downarrow$ as $L\downarrow$. Similarly, as $V_D\uparrow$, more Q_B is depleted by the drain bias, and hence $I_D\uparrow$ and $V_T\downarrow$.

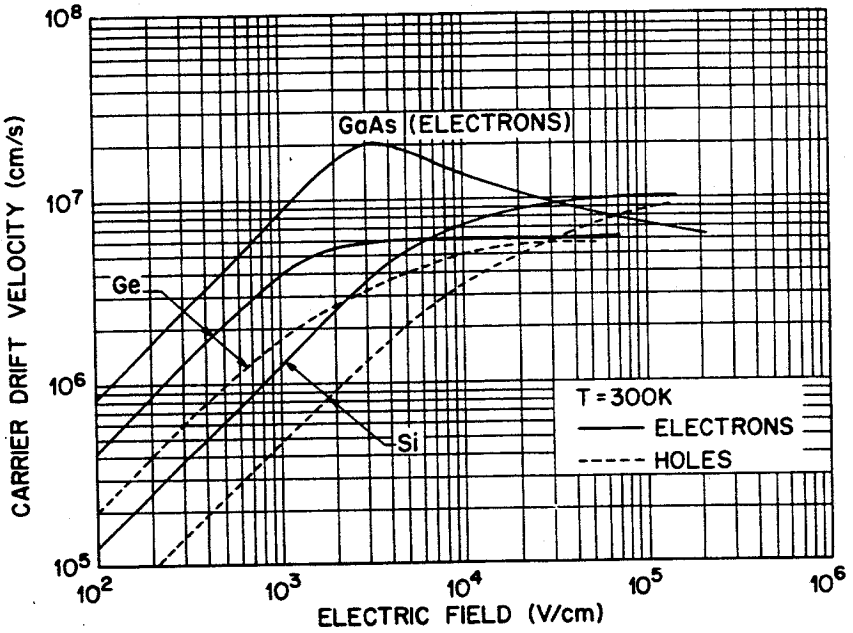


Effect of drain induced barrier lowering on drain current.



Carrier Mobility: Velocity Saturation

The mobility of the carriers reduces at higher electric fields normally encountered in small channel length devices due to velocity saturation effects.



Velocity as a function of electric field

As the channel length, L, is reduced while the supply voltage is not, the tangential electric field will increase, and the carrier velocity may saturate. $\varepsilon_C \approx 10^4$ V/cm for electrons. Hence for N-channel MOSFET with L < 1 μm, velocity saturation causes the channel current to reach saturation before $V_D = V_G - V_T$. Instead of I_{DSAT} being proportional $(V_G - V_T)^2$ it is linearly proportional to $(V_G - V_T)$ and is approximately given by

$$I_{Dsat} = WC_{ox}(V_G - V_T)v_{sat}$$

$e_C \approx 5 \times 10^4$ V/cm for holes, hence velocity saturation for P-channel MOSFET will not become important until L < 0.25 μm.

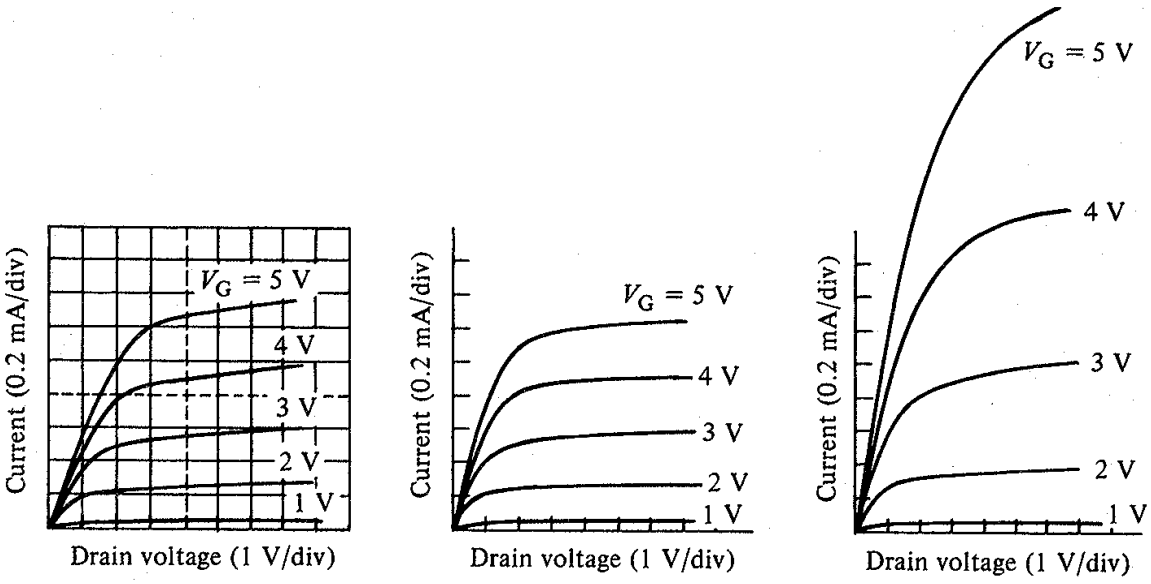
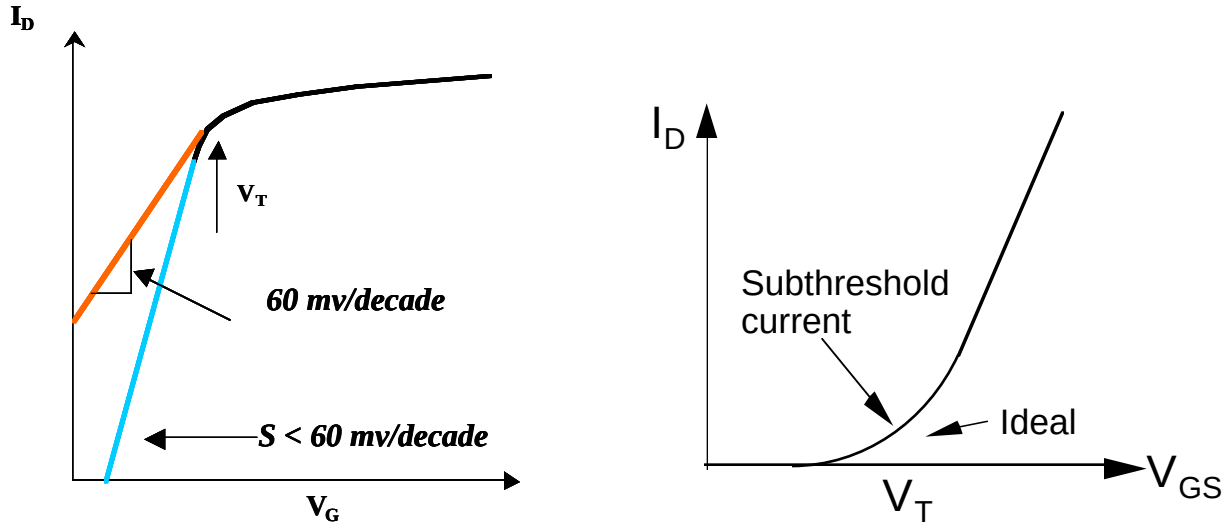


Figure. Effects of velocity saturation on the MOSFET I-V characteristics. (a)Experimental characteristics of a MOSFET with L = 2.7 μm, x_0 =0.05 μm, Comparative theoretical characteristics computed (b) including velocity saturation and (c) ignoring velocity saturation.

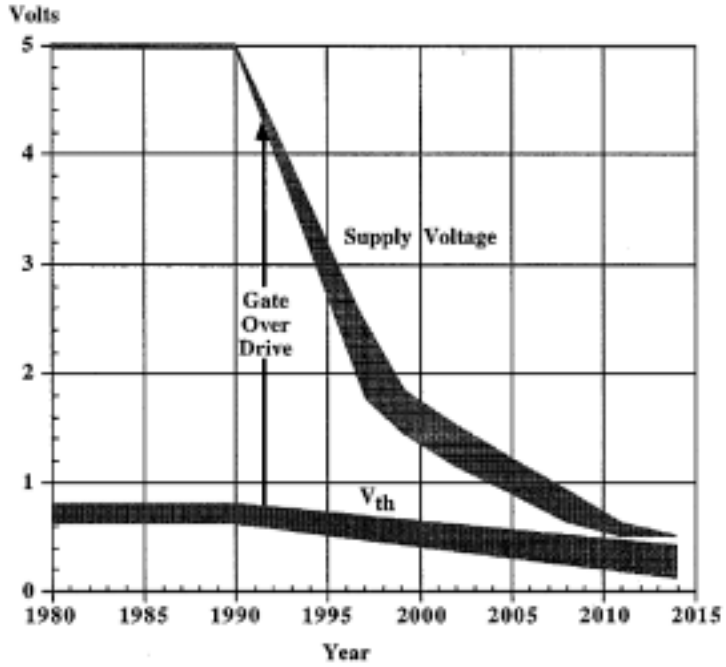
Subthreshold Conduction

When the surface is in weak inversion (i.e., o < ϕ_S < -ϕ_p, $V_G < V_T$), a conducting channel starts to form and a low level of current flows between source and drain.



In MOS subthreshold slope S is limited to kT/q (60mV/dec)

- I_D leakage ⇑
- Static power ⇑
- Circuit instability ⇑



V_{DD} is scaled for low power,delay, V_T must scale to maintain I_D (ON)

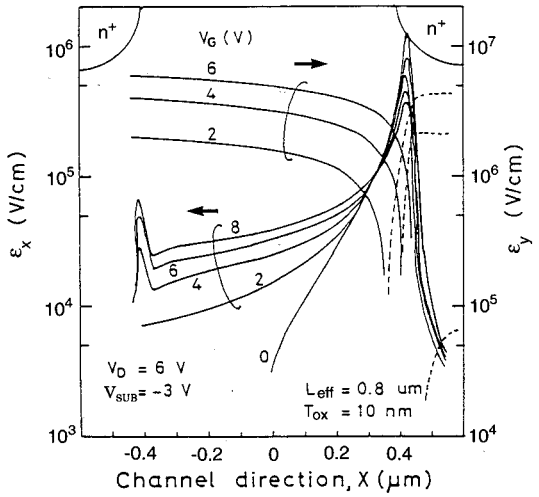
With subthreshold slope limited to 60mv/decade the dynamic range becomes limited.

Hot Carrier Effects

From our p-n junction discussion we remember that the maximum electric field intensity is near the junction itself and it increases with the reverse bias.

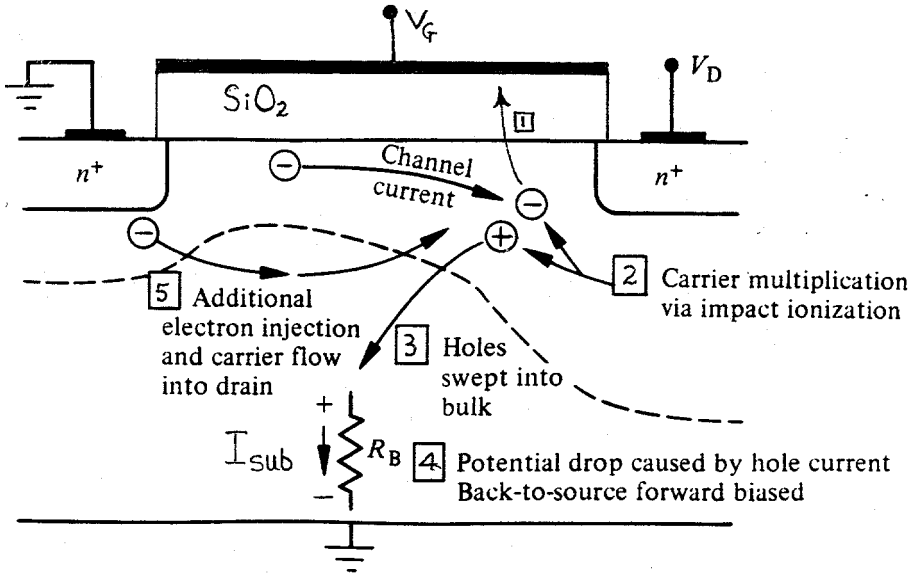
$$\xi_{max} = \sqrt{\frac{2qN_a(\phi_i - V_D)}{\varepsilon_{ox}}}$$

In the case of MOS transistor the maximum electric field is near the drain-substrate junction. The drain reverse bias has to be dropped from drain to source. As the channel length is reduced the electric field intensity in the channel near the drain increases more rapidly in comparison to the long channel case as ϕ_i does not scale.

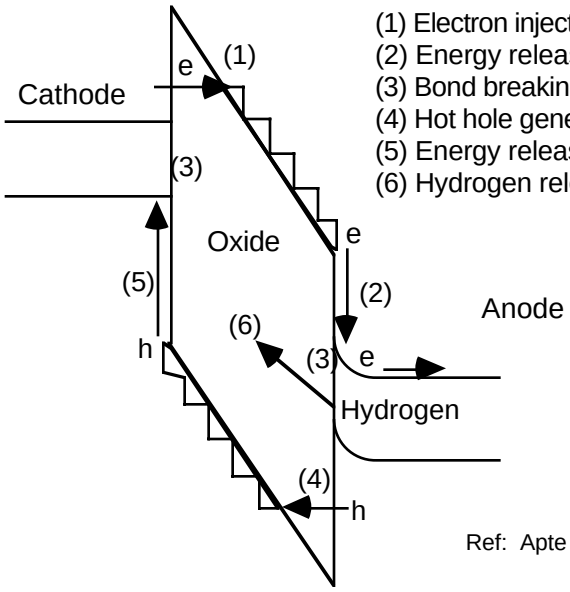


The free carriers passing through the high-field can gain sufficient energy to cause several *hot-carrier* effects. This can cause many serious problems for the device operation.

Hot carriers can have sufficient energy to overcome the oxide-Si barrier. They are injected from channel to the gate oxide (process 1) and cause gate current to flow. Trapping of some of this charge can change V_T permanently. Avalanching can take place producing electron-hole pairs (process 2). The holes produced by avalanching drift into the substrate and are collected by the substrate contact (process 3) causing I_{sub} IR drop due to I_{sub}(process 4) can cause substrate-source junction to be forward biased causing electrons to be injected from source into substrate (process 5). Some of the injected electrons are collected by the reversed biased drain and cause a parasitic bipolar action (process 5).

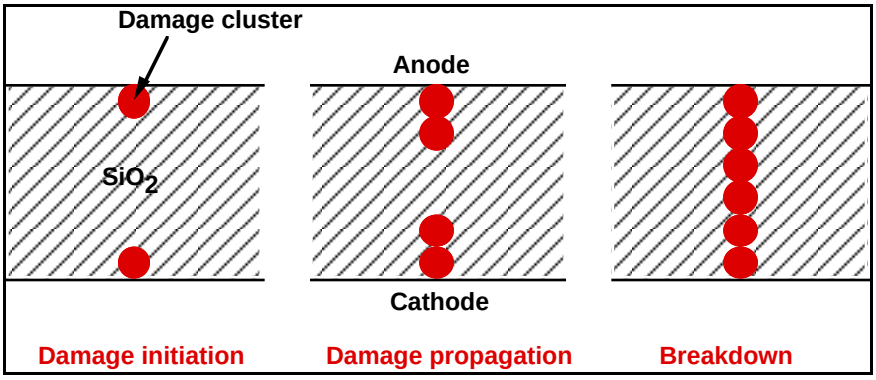


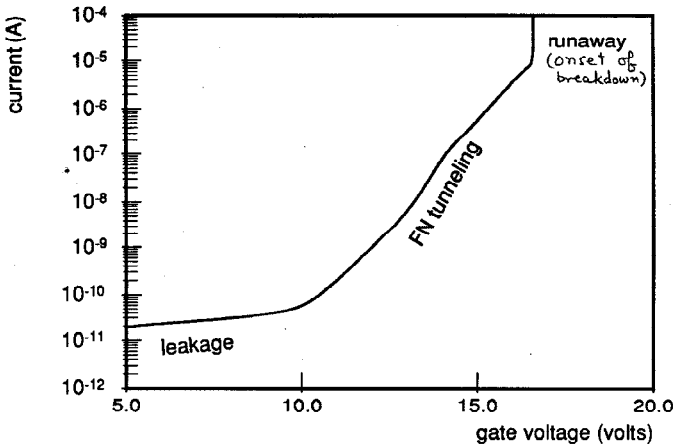
Gate Oxide Degradation and Breakdown



(1) Electron injection
(2) Energy released by hot electron
(3) Bond breaking at the interface - trap generation
(4) Hot hole generation by impact ionization and injection
(5) Energy released by hot hole - trap generation
(6) Hydrogen release - trap generation

Ref: Apte & Saraswat IEEE Trans. Electron Dev., Sept 1994

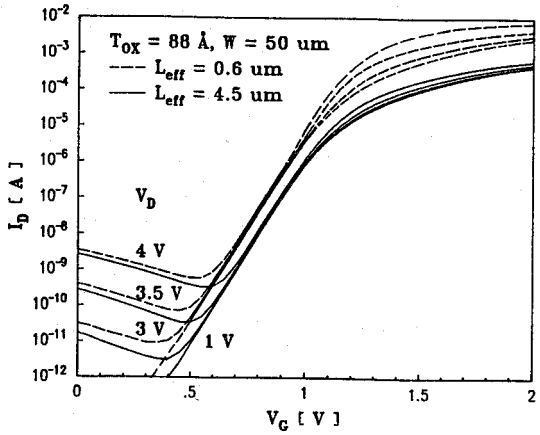




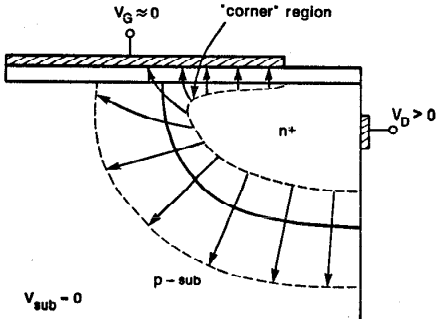
Breakdown field ~ 8 MV/cm for thick oxides and increases > 10 MV/cm for thinner oxides.

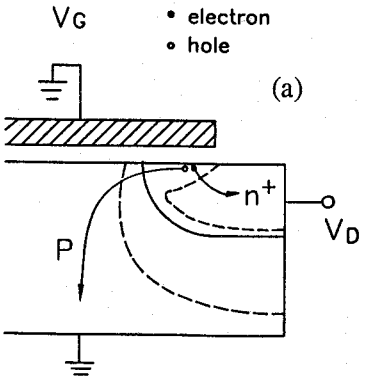
Band-to-Band Tunneling

For small gate bias at high drain bias a significant drain leakage can be observed, especially for short channel devices.

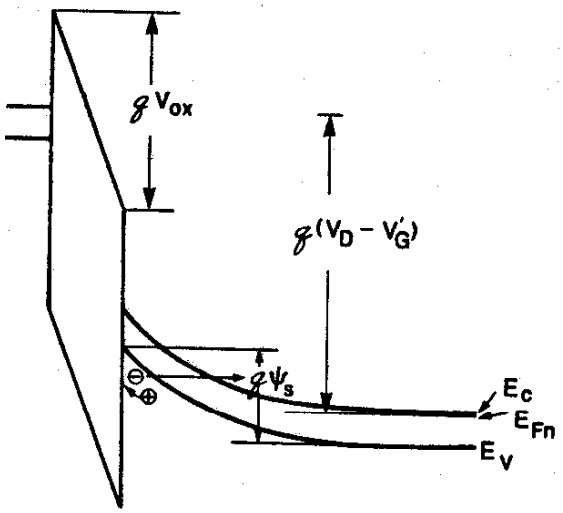


The electric field can be very high in the drain region for V_D high and V_G = 0. This can cause band-to-band tunneling. This will happen only if the electric field is sufficiently high to cause large band bending.





The figure below shows band-to-band tunneling



Effect of Reducing Channel Width on V_T

There are no diffusions on the side of the channel. Hence the depletion region extends sideways in areas lying outside the gate controlled region increasing the apparent channel width. As a result the V_T is increased. Note that the effect here is opposite to that of reducing channel length.