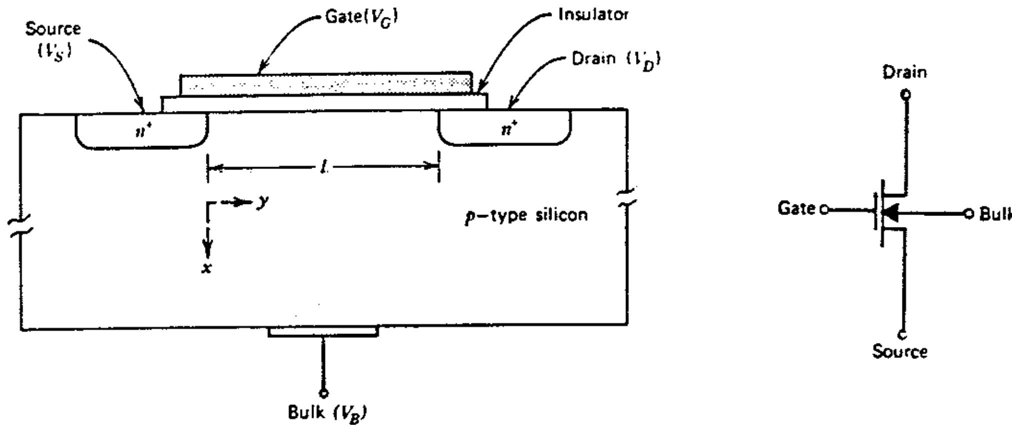
MOS Transistors

We can extend the analysis we did of the MOS capacitor system to the MOS transistor by considering adding source and drain regions doped with the opposite impurity type from the substrate.



As before V_G controls the charges in the channel region.

$\underline{V_G < V_{th}}$:

The structure consists of two back-to back diodes and only reverse leakage current flows between the source and drain.

$\underline{V_G > V_{th}}$:

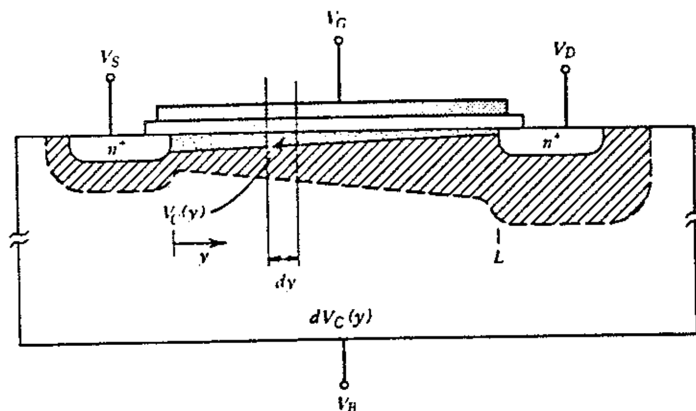
Inversion layer exists. This inversion layer forms a conducting channel between the source and drain allowing current to flow.

V_{th} is unchanged from that calculated for the MOS capacitor.

$$V_{th} = \Phi_{MS} - \frac{Q'_{ss}}{C'_{ox}} - \frac{1}{C'_{ox}} \int_0^{x_{ox}} \frac{x}{x_{ox}} \rho(x)\, dx - 2\phi_F \pm \frac{\sqrt{2q\epsilon_s N_B |2\phi_F|}}{C'_{ox}} \tag{1}$$

where N_B is the substrate doping and the term for depletion charge is + for n-channel (p-substrate) devices and − for p-channel (n-substrate) devices.

Current - Voltage Characteristics



1. The depletion region around the drain is larger than around the channel because the voltage V_D applied to the drain reverse biases the drain to substrate junction.

2. The potential along the channel varies from $V_S = 0$ to V_D between the source and drain.

3. The channel inversion charge Q'_I and the bulk depletion charge Q'_B are functions of distance along the channel since the surface potential varies along the channel.

The voltage drop across any elemental length of the channel is given by

$$dV = I_D dR = \frac{I_D dy}{W \mu'_n Q'_I(y)} \tag{2}$$

directly from the expression for the channel conductance of the MOS capacitor.

At any point the total charge in the silicon is

$$Q'_s(y) = Q'_I(y) + Q'_B(y) \tag{3}$$

Recall that

$$V_G = V_{FB} - \frac{Q'_s}{C'_{ox}} + (\phi_s - \phi_F) \tag{4}$$

where V_{FB} includes Φ_{MS} and oxide charges.

Combining (3) and (4),

$$Q_I'(y) = -[V_G - V_{FB} - (\phi_s(y) - \phi_F)] C_{ox}' - Q_B'(y) \tag{5}$$

Since the surface is inverted, the surface potential will differ from the bulk potential by approximately $2\phi_F$ plus any reverse bias which exists between the channel and the substrate (due to V_D or substrate bias)

$$\phi_s(y) - \phi_F \cong V_{CB}(y) - 2\phi_F \tag{6}$$

We also know from our analysis of the MOS capacitor that

$$Q_B'(y) = \mp qN_B x_{d\max}(y) = \mp\sqrt{2K_s\epsilon_0 qN_B|V_{CB}(y) - 2\phi_F|} \tag{7}$$

For p-substrate (n-channel), as we move from $S \to D$, $V_{CB}(y)$ increases due to IR drop in channel.

- $x_{d\max}$ increases as we move towards the drain.
- $|Q_B'|$ increases as we move towards the drain.

Substituting (6) and (7) into (5),

$$Q_I'(y) = -[V_G - V_{FB} - V_{CB}(y) + 2\phi_F)] C_{ox}' \pm \sqrt{2K_s\epsilon_0 qN_B|V_{CB}(y) - 2\phi_F|} \tag{8}$$

Note that

$$(V_G - V_{FB}) > 0,\ V_{CB}(y) > 0,\ \phi_F < 0,\ Q_B' < 0 \quad \text{for n-channel device}$$

$$(V_G - V_{FB}) < 0,\ V_{CB}(y) < 0,\ \phi_F > 0,\ Q_B' > 0 \quad \text{for p-channel device}$$

Since Q_I' is a function of channel voltage, Equation (2) can be rearranged and integrated from the source ($y = 0, V_{CB} = 0$) to the drain ($y = L, V_{CB} = V_D$).

$$\int_0^{V_D} W\mu_n' Q_I'(V)\,dV = \int_0^L I_D\,dy$$

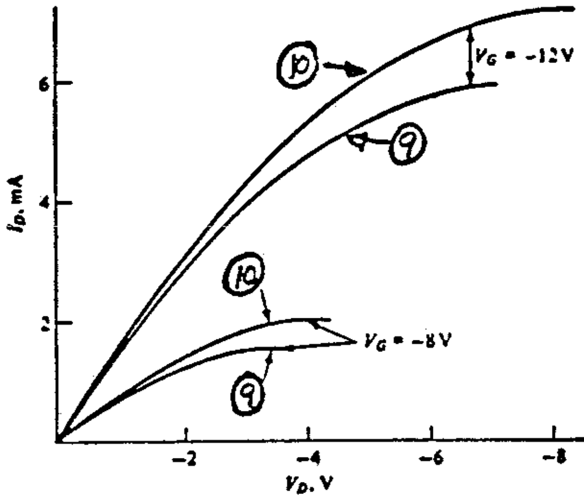
If equation (8) is used for Q'_I, we obtain

$$I_D = \frac{W}{L}\mu'_n C'_{ox}\left\{\left[V_G - V_{FB} + 2\phi_F - \frac{V_D}{2}\right]V_D \mp \frac{2}{3}\frac{\sqrt{2\epsilon_s q N_B}}{C'_{ox}}\left[|V_D - 2\phi_F|^{3/2} - |2\phi_F|^{3/2}\right]\right\} \quad (9)$$

If we had assumed $Q'_B(y)$ was just a constant given by its value without an applied drain voltage (ignored the influence of channel voltage on Q_B), we would have found

$$I_D \cong \frac{W}{L}\mu'_n C'_{ox}\left[V_G - V_{th} - \frac{V_D}{2}\right]V_D \quad (10)$$

where V_{th} includes the effects of ϕ_F, ϕ_{MS}, oxide charges and $Q'_B(y = 0)$.



Note that for a typical device, the approximate expression (10) works well for low V_D, but predicts too high a current at larger V_D. This makes sense since some of the gate charge must support the extra depletion charge and therefore the inversion layer charge is corresponding reduced, an effect which was ignored in the approximation.

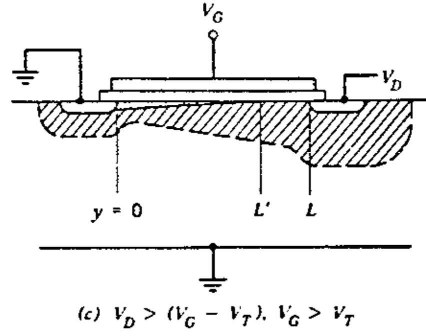
- Both equations (9) and (10) are only valid when an inversion layer exists all the way across the channel from the source to the drain.

- As V_D increases, the voltage between the gate and the channel near the drain decreases until it is less than V_{th}, pinching off the channel. This occurs when

$$V_D > V_G - V_{th} = V_{D\,sat} \quad (11)$$

When V_D exceeds the saturation voltage, a channel no longer exist all the way to the drain. This is known as the saturation or "pinch-off" region.

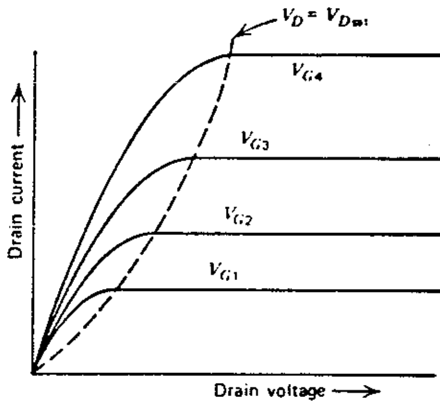
$\underline{V_D > V_{D\,sat}}$



(c) $V_D > (V_G - V_T)$, $V_G > V_T$

- Electrons will travel along the inversion layer and then become injected into the depletion region where the high $\mathcal{E}$ field pulls them into the drain.
- Further increase of V_D does not change I significantly.

$$I_D \cong \text{constant for } V_D > V_{D\,sat}$$



The boundary between the linear and saturation regions is described by

$$V_D = V_G - V_{th}$$

More accurately, we can solve equation (8) for $Q_I'(y = L) = 0$ (no inversion layer),

$$V_{D\,sat} = V_G - V_{FB} + 2\phi_F \pm \frac{K_s \epsilon_0 q N_B}{C_{ox}^2} \left[1 - \sqrt{1 + \frac{2C_{ox}^2 |V_G - V_{FB}|}{K_s \epsilon_0 q N_B}} \right] \qquad (12)$$

Linear Region Conductance

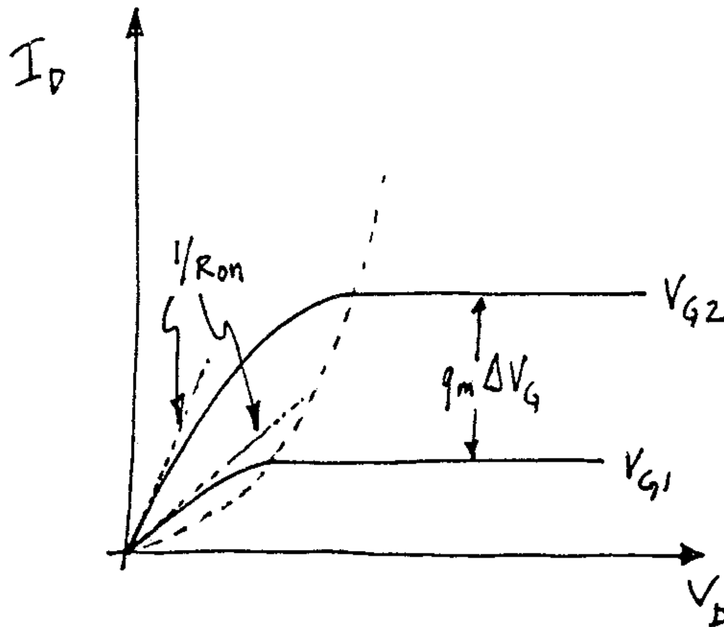
If $V_D \ll (V_G - V_{th})$, then (10) reduces to

$$I_D \cong \frac{W}{L}\mu_n' C_{ox}'(V_G - V_{th})V_D \tag{13}$$

The device therefore looks like a voltage controlled resistor with a conductance given by

$$\begin{aligned} g_D &= \left.\frac{\partial I_D}{\partial V_D}\right|_{V_G} \\ &= \frac{1}{R_{on}} = \frac{W}{L}\mu_n' C_{ox}'(V_G - V_{th}) \end{aligned} \tag{14}$$

This is the same relationship which was derived for the MOS capacitor.



Saturation Region Transconductance

The transconductance or gain of the device is defined as

$$g_m \equiv \left.\frac{\partial I_D}{\partial V_G}\right|_{V_D} \tag{15}$$

Differentiating (9) with respect to V_G yields

$$g_m = \frac{W}{L}\mu_n' C_{ox}' V_D \qquad (V_D < V_{D\,\mathrm{sat}}) \tag{16}$$

In the saturation region, I is approximately constant, so we can evaluate the transconductance at the edge of the saturation region and that will hold throughout the saturation region. We can therefore substitute $V_{D\,\mathrm{sat}}$ as given in (12) into (15).

$$g_{m\,\mathrm{sat}} = \frac{W}{L}\mu_n' C_{ox}' \left\{ V_G - V_{FB} + 2\phi_F \pm \frac{K_s \epsilon_0 q N_B}{C_{ox}^2}\left[1 - \sqrt{1 + \frac{2C_{ox}^2 |V_G - V_{FB}|}{K_s \epsilon_0 q N_B}}\right]\right\} \tag{17}$$

Making the same approximation we made in going from (9) to (10) (ignore change in Q_B' with y)

$$g_{m\,\mathrm{sat}} \cong \frac{W}{L}\mu_n' C_{ox}'(V_G - V_{th}) \tag{18}$$

$$R_{on} \cong \frac{1}{\frac{W}{L}\mu_n' C_{ox}'(V_G - V_{th})}$$

$$g_m \cong \frac{1}{R_{on}}$$

The on resistance and the saturation transconductance (gain) are inversely related.

Threshold Control

It is usually desirable to make enhancement mode devices. (V_{th}(n-channel) > 0, V_{th}(p-channel) < 0)

$$\text{p-channel:} \qquad V_{th} = \underbrace{\Phi_{MS}}_{-} \underbrace{-2\phi_n}_{-} - \underbrace{\frac{Q'_B}{C'_{ox}}}_{-} - \underbrace{\frac{Q'_{ss}}{C'_{ox}}}_{-} \qquad < 0$$

$$\text{n-channel:} \qquad V_{th} = \underbrace{\Phi_{MS}}_{-} \underbrace{-2\phi_p}_{+} - \underbrace{\frac{Q'_B}{C'_{ox}}}_{+} - \underbrace{\frac{Q'_{ss}}{C'_{ox}}}_{-} \qquad ??$$

In order to ensure that $V_{thn} > 0$, need:

- $|\phi_p|$ large $\Rightarrow$ high doping $\Rightarrow$ large C_s, low junction BV, low μ'_n
- C'_{ox} small $\Rightarrow$ thick oxide $\Rightarrow$ low gain ($g_m \propto C'_{ox}$)

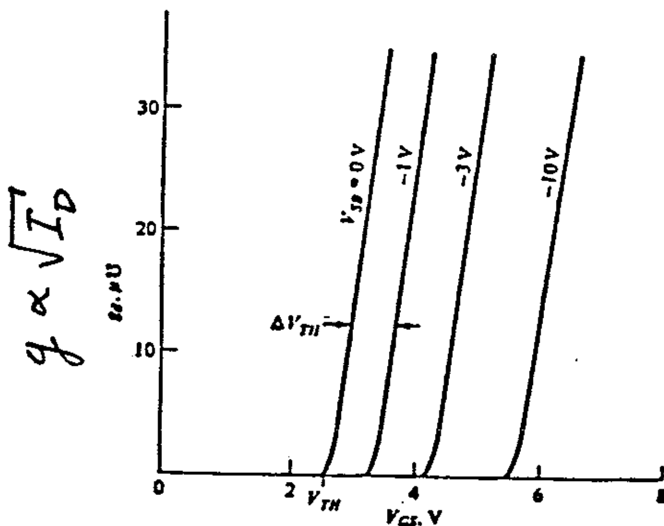
Body Bias (Back Gate Bias)

In the analysis to this point we have assumed that $V_B = 0$, i.e. $V_{SB} = 0$. If this is not true and the substrate is reverse biased, i.e. $V_{SB} > 0$ for an n-channel device, then $|Q'_B|$ will increase.

$$Q'_B = \mp\sqrt{2K_s\epsilon_0 q N_B |V_{SB} - 2\phi_F|}. \tag{13}$$

The change in Q'_B due to V_{SB} is supported by a change in the gate voltage, i.e. a change in V_{th}.

$$V_{th} = \Phi_{MS} - \frac{Q'_{ss}}{C'_{ox}} - 2\phi_F \pm \frac{\sqrt{2q\epsilon_s N_B |V_{SB} - 2\phi_F|}}{C'_{ox}} \tag{14}$$



The equations we derived for I-V relations etc., are still valid provided we replace V_{th} by the above expression.

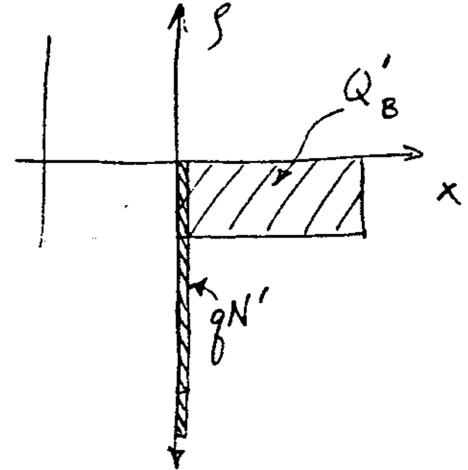
Therefore, the threshold voltage can be increased by reverse biasing the substrate. The disadvantage of this approach is the requirement for an additional voltage source.

This back gate bias is also used in modern MOS ICs to reduce junction capacitances.

Channel Implants

The most widespread approach used for control of threshold voltage in MOS transistors is the shallow-implantation of a surface layer of dopants in the channel.

If we assume that the implanted layer is very thin and close to the surface, then the region can be assumed to be depleted yet have no impact on the depletion region width (the electric field is changed only over a neglible distance).



Therefore the depletion charge is is changed from Q'_B to $Q'_B + qN'$ where N' is the implanted dose per unit area and

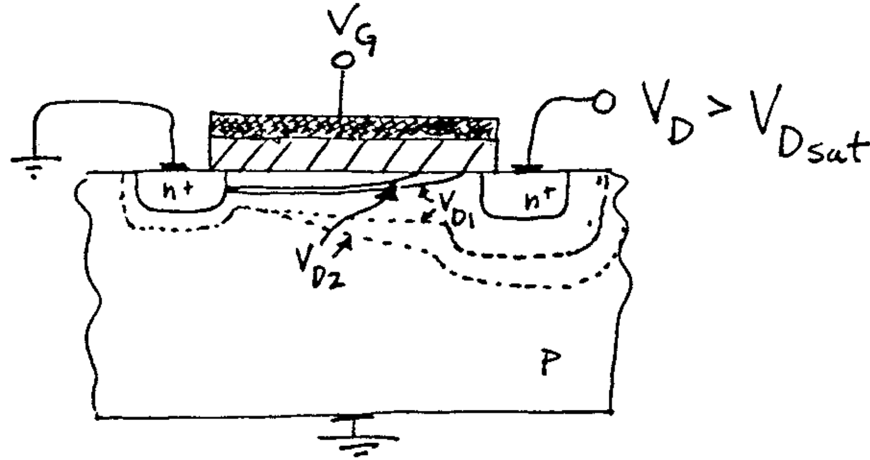
$$V_{th} = V_{FB} - 2\phi_F - \frac{Q'_B}{C'_{ox}} - \frac{qN'}{C'_{ox}} \tag{15}$$

In order to increase the threshold voltage (as is required to make n-channel enhancement mode devices) acceptor impurities (−) must be implanted, while to reduce V_{th} donor impurities are implanted.

If the implant profile is not actually of neglible depth, an additional voltage drop will occur across the implanted layer reducing the effectiveness of the implant on V_{th}. In that case, Poisson's equation would have to be used to determine the threshold shift (see text).

Implantation is an effective method of controlling the threshold voltage because the critical parameter, total dose, is very well controlled ($\sim \pm 3\%$).

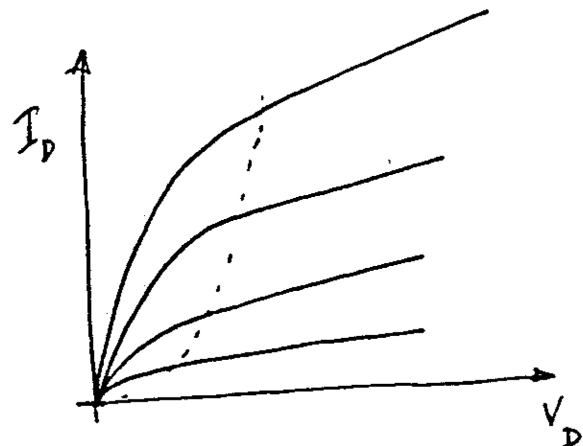
Channel Length Modulation



In the saturation region, as V_D increases, the pinch-off point of the channel moves back towards the source (the drain depletion region expands). Therefore the current I increases since it is $\propto 1/L_{\text{eff}}$. The depletion region expands as $\sqrt{V_D - V_{D\text{sat}}}$ assuming constant doping. Provided the device has a channel length $L \gg \Delta x_d$ and provided $x_{ox} \ll \Delta x_d$, then the change in channel length is approximately

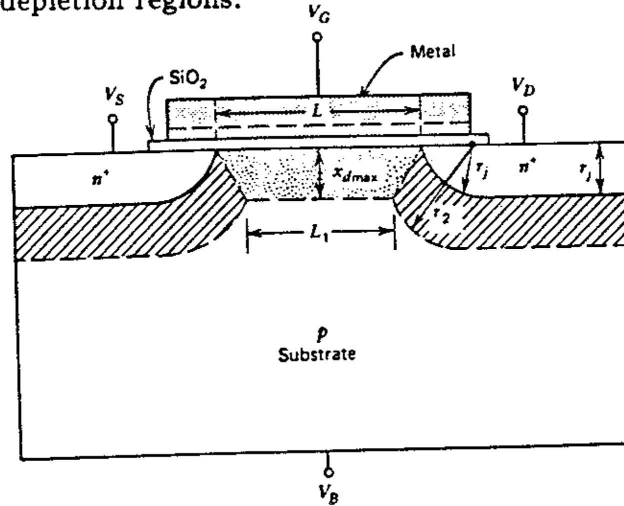
$$\Delta L \cong \sqrt{\frac{2\epsilon_s}{qN_B}|V_D - V_{D\,\text{sat}}|} \tag{16}$$

The decrease in L is responsible for an increase in I in the saturation region, therefore a finite output impedance results. Channel width modulation is most pronounced in devices with lightly doped substrates and short channels.

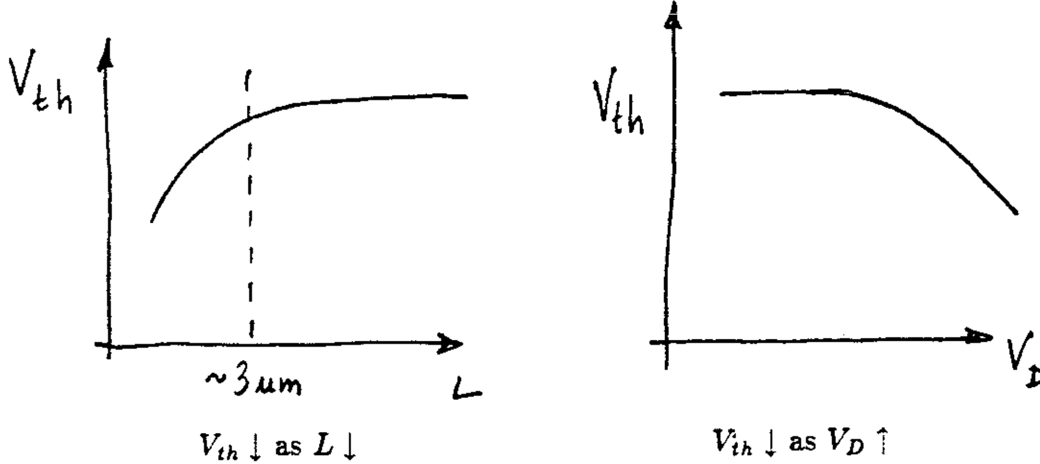


Short Channel Effects

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 depletion regions.



Therefore $V_{th} \downarrow$ as $L \downarrow$ since less gate voltage is required to support Q_B.

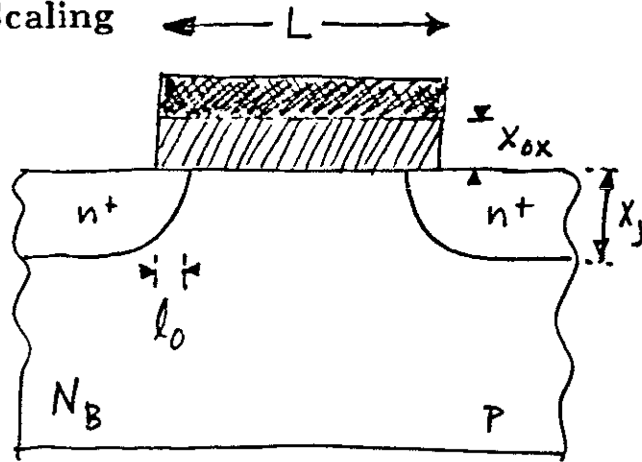


$V_{th} \downarrow$ as $L \downarrow$

$V_{th} \downarrow$ as $V_D \uparrow$

These effects are particularly pronounced in lightly doped substrates. Ion implanting a more heavily doped region under the gate (same type as substrate) can be used to minimize the threshold shifts.

MOS Device Scaling



The performance of the intrinsic MOS transistor may be expressed in terms of the following parameters.

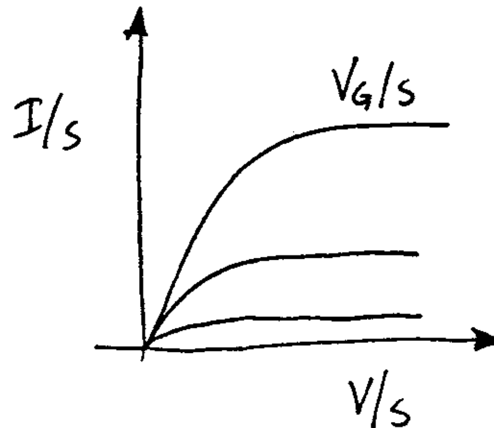
Device gain (g_m)	$\propto$	$\dfrac{1}{x_{ox}}$
Input capacitance	$\propto$	$\dfrac{WL}{x_{ox}}$
Miller capacitance	$\propto$	$\dfrac{Wl_0}{x_{ox}}$
Punch through voltage	$\propto$	$N_B L^2$
Threshold voltage	$\propto$	$\sqrt{N_B}x_{ox}$

In order to maintain the same I-V characteristics as size is shrunk,

$L, x_{ox}, x_j, l_0 \downarrow$ by S,

$N_B \uparrow$ by S,

where S = scaling factor.



Device / Circuit Parameter	Scaling Factor
Dimensions : x_{ox}, L, l_0, W, x_j	$\frac{1}{S}$
Substrate Doping : N_B	S
Supply voltage : V	$\frac{1}{S}$
Supply current : I	$\frac{1}{S}$
Parasitic Capacitance : $\frac{WL\epsilon_{ox}}{x_{ox}}$	$\frac{1}{S}$
Gate delay : $\frac{VC}{I}$	$\frac{1}{S}$
Power dissipation : IV	$\frac{1}{S^2}$
Power delay product : V^2C	$\frac{1}{S^3}$

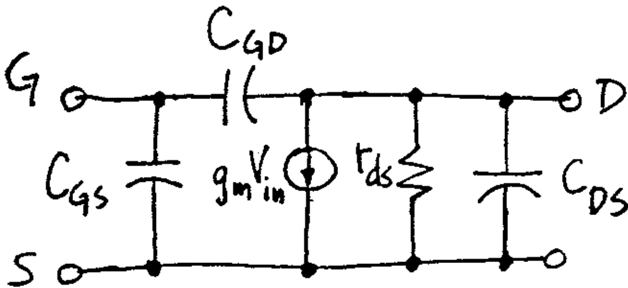
Parameter	1977	1980	1986
Layout density (gates/mm^2)	200	450	1000
Power delay product (pJ)	1	0.2	0.01
Gate delay (ns)	1 0	0.2	0.01
Power supply voltage	5	5	3
Channel length (μm)	3.5	2.0	0.7
x_{ox} (Å)	700	400	100

Circuit Models

The MOS transistor may be modeled in the following manner, by inspection of its physical structure.

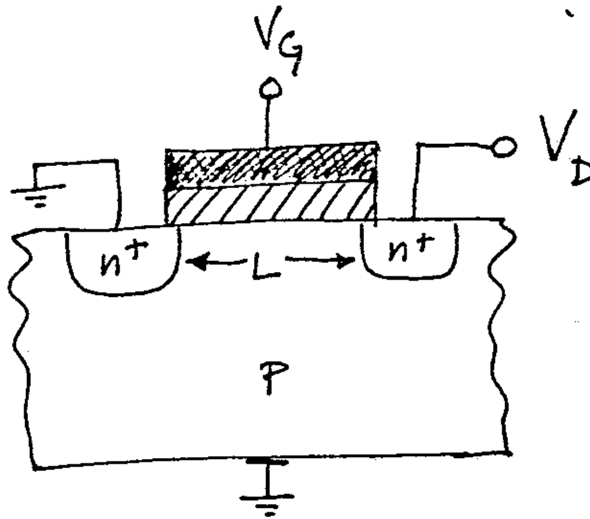
Of the elements in the model, only the gate to channel capacitance is essential; the rest are parasitic and degrade performance. Technology improvements are generally designed to reduce these parasitics.

In many cases, the equivalent circuit can be reduced to:



Note that many of the parameters in the model are voltage sensitive. Therefore accurate large signal modeling usually requires computer techniques.

Summary



1. MOS devices operate by controlling surface conductance.

2. Performance is greatly improved by shrinking dimensions (L, etc.). Therefore VLSI IC's will likely use this type of device because:

   (a) Small device size $\Rightarrow$ high performance.

   (b) Small device size $\Rightarrow$ lots of devices in a given chip area.

   (c) MOS devices are physically smaller for the same design rules than bipolar devices.

3. Small MOS devices ($L < 1\mu$m) require very tight process controls to achieve reproducible characteristics.