MOS Transistor

So far, we have treated MOS transistors as ideal switches.

An *ON* transistor passes a finite amount of current
- Depends on terminal voltages
- Need to derive current-voltage (I-V) characteristics.

Transistor gate, source and drain all have capacitance

$$I = C(\Delta V / \Delta t)$$

$$\Delta t = (C / I) \cdot \Delta V$$

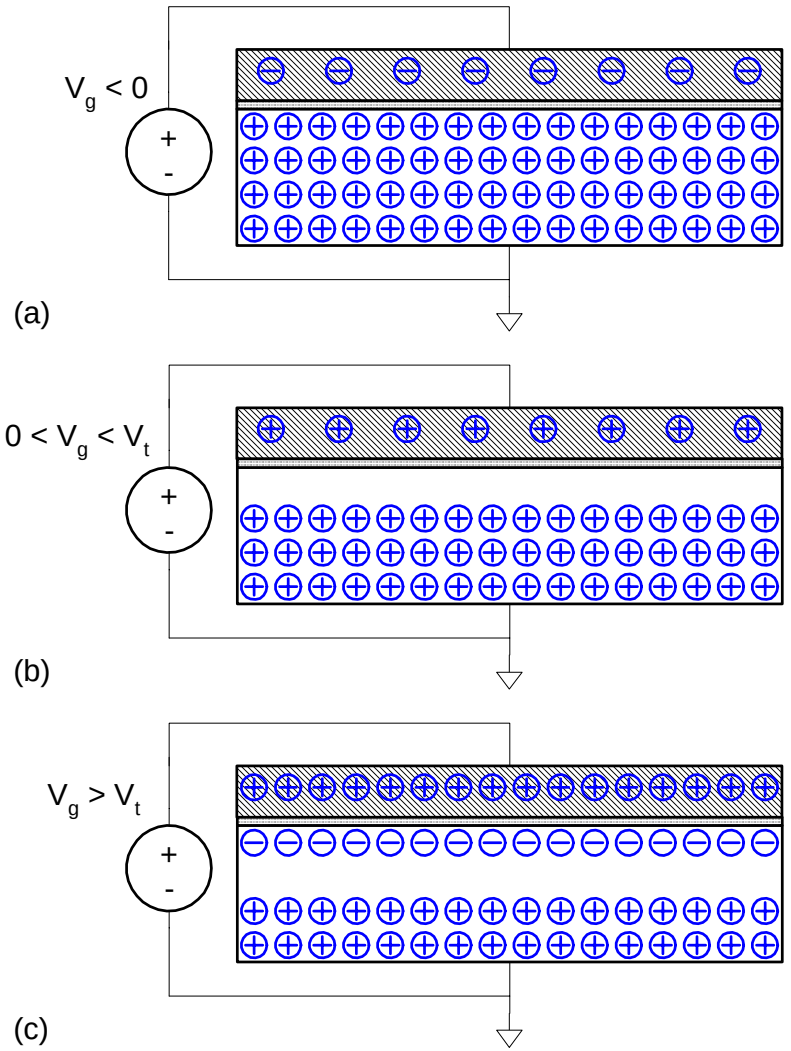
We will also look at what a *degraded level* really means.

Positive/negative voltage applied to the gate (with respect to substrate) enhances the number of electrons/holes in the channel and increases conductivity between source and drain.

V_t defines the voltage at which a MOS transistor begins to conduct. For voltages less than V_t (threshold voltage), the channel is cut off.

MOS Capacitor

Gate and body form a MOS capacitor



Operating Modes:

Accumulation

$V_g < 0$

Depletion

$0 < V_g < V_t$

Inversion

$V_g > V_t$

MOS Terminal Voltages and Modes of Operation

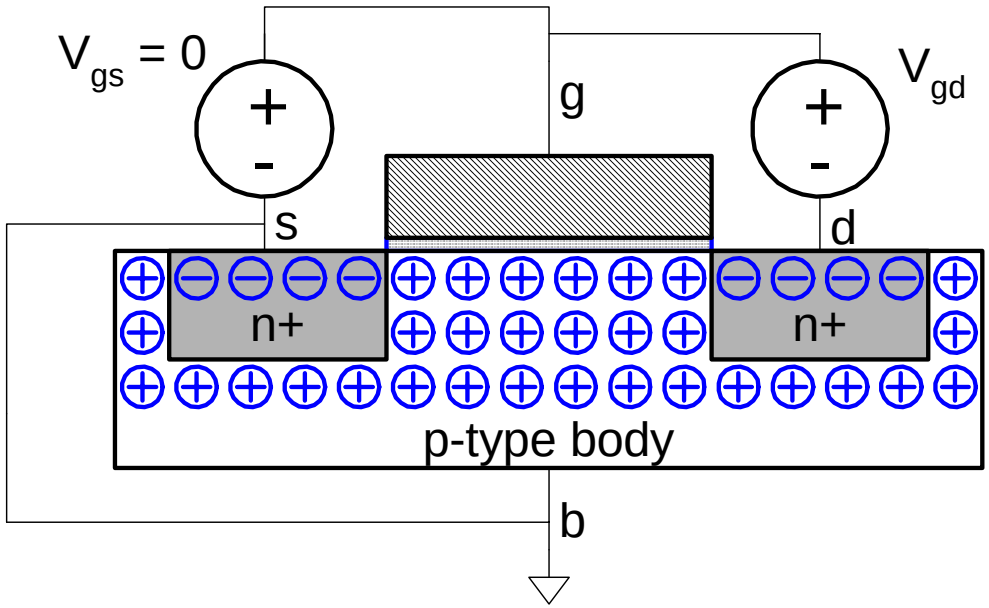
- Mode of operation depends on the terminal voltages. V_g, V_s, V_d
  - $V_{gs} = V_g - V_s$
  - $V_{gd} = V_g - V_d$
  - $V_{ds} = V_d - V_s = V_{gs} - V_{gd}$
- Source and Drain are symmetric diffusion terminals
  - By convention, source is terminal at lower voltage
  - Hence $V_{ds} > 0$
- NMOS body is grounded. First assume that source is 0 too.
- Three modes of operation
  - Cutoff
  - Linear
  - Saturation

Modes of Operation

NMOS Cutoff

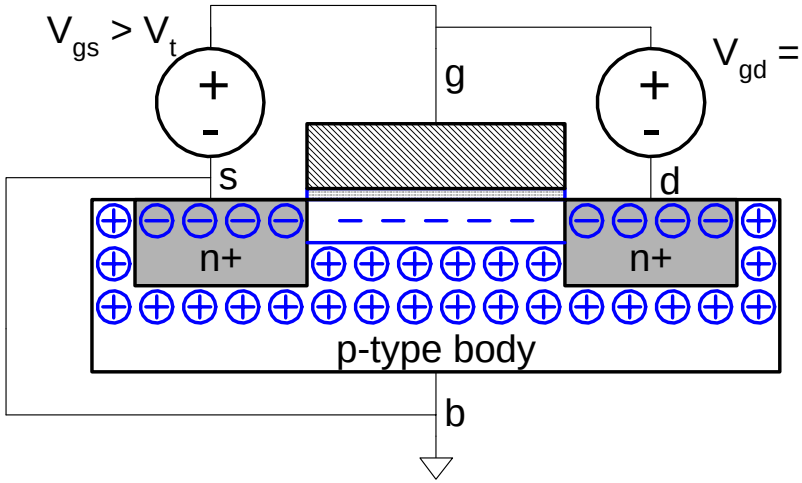
No channel

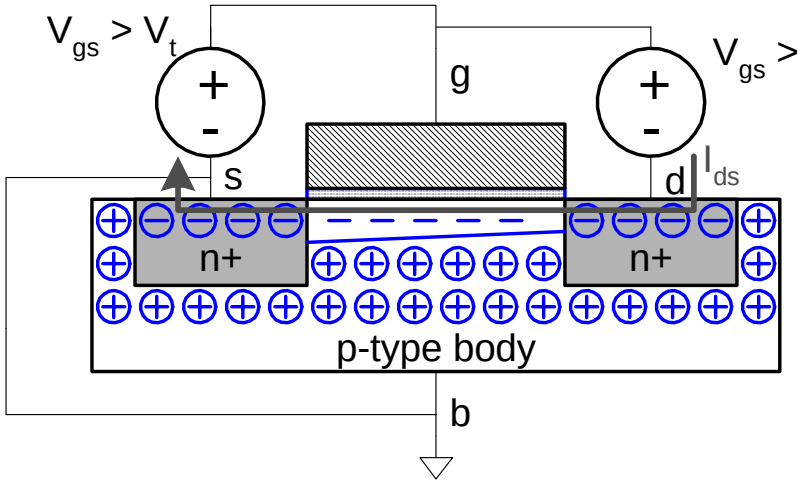
I_{ds} is 0



Modes of Operation

NMOS Linear (resistive or non-saturated region)





Channel is formed and extends from the source to the drain

Current flow from drain to source (electrons)

I_{ds} increases with V_{ds}

Similar to a linear resistor

Modes of Operation

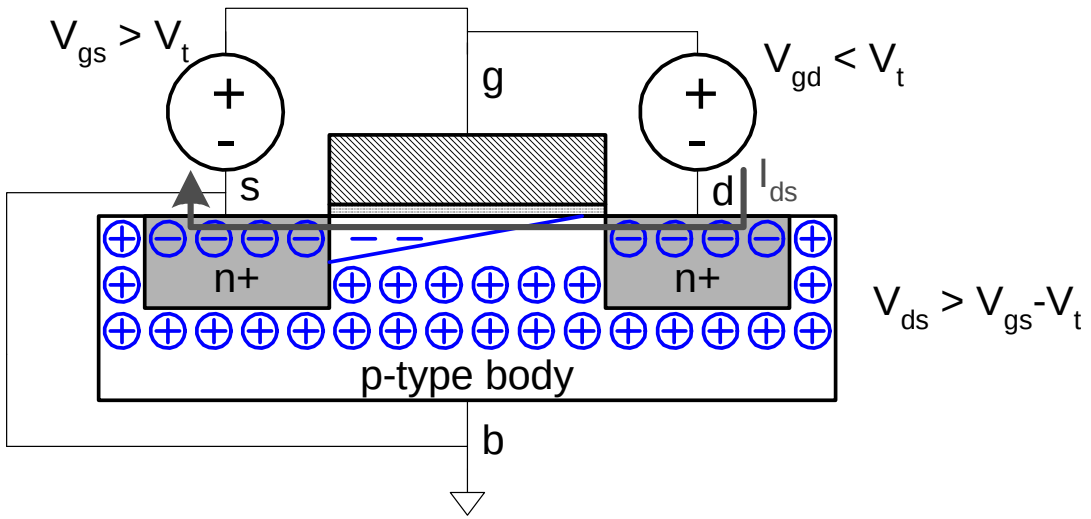
NMOS Saturation

Channel is pinched off near the drain

I_{ds} independant of V_{ds}. I_{ds} is a function of V_{gs} only

We refer to it as current saturates

Similar to a current source



MOS I-V Characteristics

MOS transistors can be modeled as a voltage controlled switch. I_{ds} is an important parameter that determines the behavior, e.g., the speed of the switch.

What are the parameters that effect the magnitude of I_{ds}? (Assume V_{gs} and V_{ds} are fixed).

- The distance between source and drain (channel length).
- The channel width.
- The threshold voltage.
- The thickness of the gate oxide layer.
- The dielectric constant of the gate insulator.
- The carrier (electron or hole) mobility.

Summary of normal conduction characteristics:

- ***Cut-off*:** accumulation, I_{ds} is essentially zero.
- ***Nonsaturated*:** weak inversion, I_{ds} dependent on both V_{gs} and V_{ds}.
- ***Saturated*:** strong inversion, I_{ds} is ideally independent of V_{ds}.

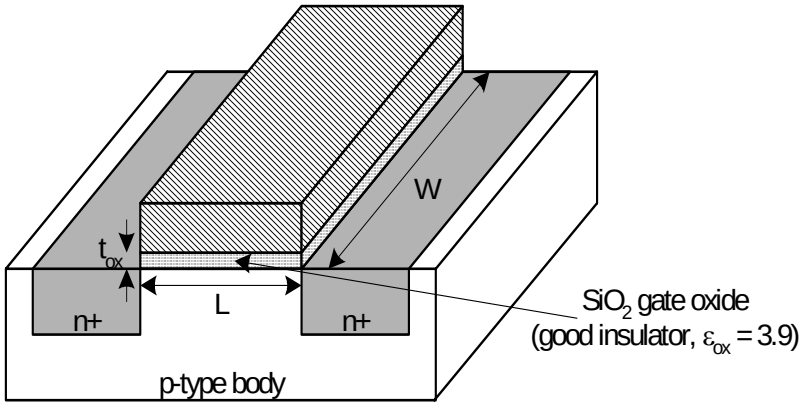
MOS I-V Characteristics (Linear)

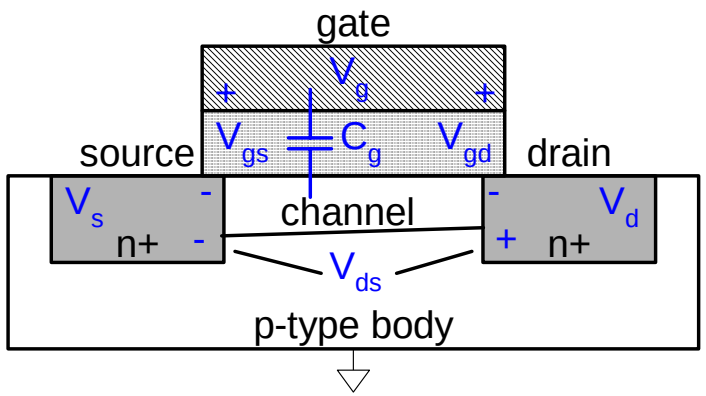
In Linear Region, I_{ds} depends on

How much charge is in the channel?

How fast is the charging moving?

Channel Charge





MOS structure looks like parallel plate capacitor while operating in inversion *(Gate-Oxide-channel)*

$$Q_{channel} = CV$$

$$C = C_g = \varepsilon_{ox}\left(\frac{WL}{t_{ox}}\right) = C_{ox}(WL)$$

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$$

$$V = V_{gc} - V_t = \left(V_{gs} - \frac{V_{ds}}{2}\right) - V_t$$

MOS I-V Characteristics (Linear)

Carrier Velocity

Charge is carried by an electron

Carrier velocity v is proportional to lateral E-field between source and drain

$$v = \mu E \qquad \mu \text{ called mobility}$$

$$E = V_{ds}/L$$

Time of carrier to cross channel: $t = L/V$

Now we know

How much charge $Q_{channel}$ is in the channel

How much time t each carrier takes to cross from source to drain

$$I_{ds} = Q_{channel}/t$$

$$= \mu C_{ox}\frac{W}{L}\left((V_{gs} - V_t) - \frac{V_{ds}}{2}\right)V_{ds}$$

$$= \beta\left((V_{gs} - V_t) - \frac{V_{ds}}{2}\right)V_{ds} \qquad \beta = \mu C_{ox}\frac{W}{L}$$

MOS I-V Characteristics (Saturation, linear, cutoff)

MOS I-V characteristics (saturation)

If $V_{gd} < V_t$ channel pinches off near the drain

When $V_{ds} > V_{dsat} = V_{gs} - V_t$

Now drain voltage no longer increases current

$$I_{ds} = \beta\left((V_{gs} - V_t) - \frac{V_{dsat}}{2}\right)V_{dsat} = \frac{\beta}{2}(V_{gs} - V_t)^2$$

Shockley 1st order transistor models

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \quad \textbf{Cutoff} \\ \beta\left((V_{gs} - V_t) - \frac{V_{ds}}{2}\right)V_{ds} & V_{ds} < V_{dsat} \quad \textbf{Linear} \\ \frac{\beta}{2}(V_{gs} - V_t)^2 & V_{ds} > V_{dsat} \quad \textbf{Saturation} \end{cases}$$

MOS I-V Characteristics

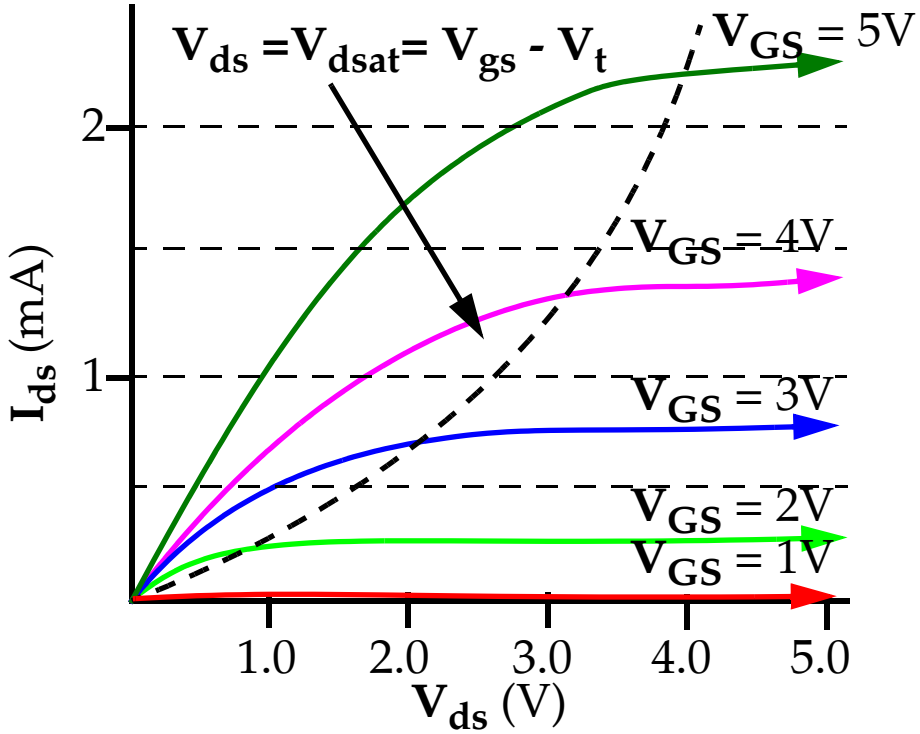
Following are the parameters on the previous slide

$$\beta = \mu C_{ox} \frac{W}{L}$$

Process dependent parameters μ and $C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$

Geometry dependent factors (designer controlled) $\frac{W}{L}$

Voltage-Current Characteristics for NMOS



MOS I-V Characteristics

All dopings and voltages reversed for PMOS.

- Mobility μ_p is determined by holes
- Typically it is 2-3 times lower than that of electrons

Typical values for AMI 0.6 μm technology that we use in the lab

$$\mu_n = 350\text{cm}^2/\text{V-sec} \qquad \mu_p = 120\text{cm}^2/\text{V-sec} \qquad t_{ox} = 10\text{nm}$$

$$\varepsilon = 3.9\varepsilon_0 = 3.9 \times 8.85\times10^{-14}\ \text{F/cm (permittivity of silicon dioxide)}$$

β for NMOS and PMOS

$$\beta_n = \frac{350 \times 3.9 \times 8.85\times10^{-14}}{0.1 \times 10^{-5}}\frac{W}{L} = 120.8\frac{W}{L}\mu A/V^2$$

$$\beta_p = \frac{120 \times 3.9 \times 8.85\times10^{-14}}{0.1 \times 10^{-5}}\frac{W}{L} = 41.1\frac{W}{L}\mu A/V^2$$

NMOS gain approximately 2-3 times higher than PMOS. Thus W/L for PMOS needs to be higher to provide same amount of current (same rise and fall times).

Threshold Voltage

V_t is also an important parameter. What effects its value?

Most are related to the material properties. In other words, V_t is largely determined at the time of fabrication, rather than by circuit conditions, like I_{ds}.

For example, material parameters that effect V_t include:

- The gate conductor material (poly vs. metal).
- The gate insulation material (SiO_2).
- The thickness of the gate material.
- The channel doping concentration.

However, V_t is also dependent on

- V_{sb} (the voltage between source and substrate), which is normally 0 in digital devices.
- Temperature: changes by -2mV/degree C for low substrate doping levels.

Threshold Voltage

The expression for threshold voltage is given as:

$$V_t = \underbrace{2\phi_b + \frac{\sqrt{2\varepsilon_{Si} q N_A 2\phi_b}}{C_{ox}}}_{\text{Ideal threshold voltage}} + \underbrace{V_{fb}}_{\text{Flat band voltage}}$$

where $\phi_b = \frac{kT}{q}\ln\left(\frac{N_A}{N_i}\right)$ Bulk potential

and

N_A: Density of the carriers in the doped semiconductor substrate.

N_i: The carrier concentration of intrinsic (undoped) silicon.

$$N_i = 1.45\times10^{10} cm^{-3} \text{ (at 300 degrees K)}$$

k: Boltzman's constant. T: temperature. q: electronic charge.

$$\frac{kT}{q} = 25mV \text{ (at 300 degrees K)}$$

e_{Si}: permittivity of silicon $\varepsilon_{Si} = 1.06\times10^{-12}$ Farads/cm

C_{ox}: gate-oxide capacitance. $C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$

Threshold Voltage

Threshold voltage (cont.):

$$V_t = \underbrace{2\phi_b + \frac{\sqrt{2\varepsilon_{Si} q N_A 2\phi_b}}{C_{ox}}}_{\text{Ideal threshold voltage}} + \underbrace{V_{fb}}_{\text{Flat band voltage}}$$

and

$$V_{fb} = \phi_{ms} - \frac{Q_{fc}}{C_{ox}}$$

where Q_{fc} represents the fixed charge due to imperfections in silicon-oxide interface and doping.

and ϕ_{ms} is work function difference between gate material and silicon substrate (ϕ_{gate}- ϕ_{Si}).

Typical values of V_{fb} for n/p transistor is **-0.9V** (with $N_A = 10^{16}$ cm^{-3}) and **-0.2V**.

Typical values of V_t for n and p-channel transistors are +/- 700mV.

Threshold Voltage

From equations, threshold voltage may be varied by changing:

- The doping concentration (N_A).
- The oxide capacitance (C_{ox}).
- Surface state charge (Q_{fc}).

As you can see, it is often necessary to adjust V_t. Two methods are common:

- Change Q_{fc} by introducing a small doped region at the oxide/substrate interface via ion implantation.

- Change C_{ox} by using a different insulating material for the gate.

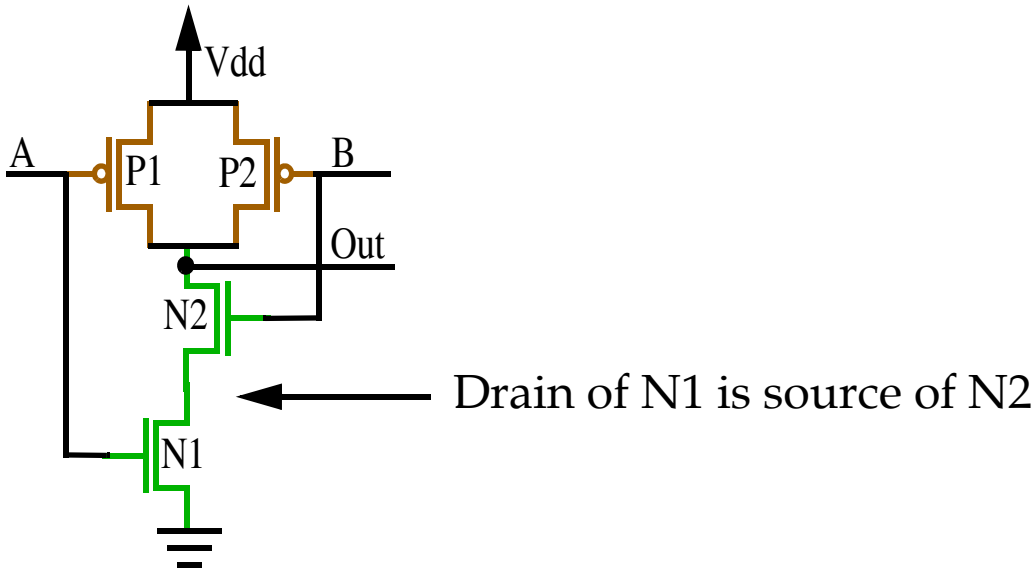
A layer of Si_3N_4 (silicon nitride) with a relative permittivity of 7.5 is combined with a layer of silicon dioxide (relative permittivity of 3.9).
This results into a relative permittivity of 6.

For the same thickness dielectric layer, C_{ox} is larger using the combined material, which lowers V_t.

Body Effect

In digital circuits, the substrate is usually held at zero.

The sources of n-channel devices, for example, are also held at zero, except in cases of series connections, e.g.,



The source-to-substrate (V_{sb}) may increase at this connections

e.g. $V_{sbN1} = 0$ but $V_{sbN2} \neq 0$.

V_{sb} adds to the channel-substrate potential:

$$V_t = 2\phi_b + \frac{\sqrt{2\varepsilon_{Si} q N_A |2\phi_b + V_{sb}|}}{C_{ox}} + V_{fb}$$

Non-Ideal I-V effects

The I-V characteristics designed so far neglect many effects that are important in modern deep-submicron processes.

Some of these effects include:

- Velocity Saturation and Mobility Degradation
- Channel length modulation
- Subthreshold conduction
- Tunneling
- Junction leakage
- Body Effect (discussed previously)
- Temperature and Geometry dependence

Velocity Saturation and Mobility Degradation

The equation for carrier velocity $v = \mu E$ predicts that carrier drift velocity and hence current increase linearly with the lateral electric field $E_{lat} = V_{ds}/L$ between source and drain.

Only true for weak electric fields, at high electric fields drift ***velocity rolls off and saturates*** to $\upsilon_{sat} = \mu E_{sat}$ where E_{sat} is determined empirically.

Thus the saturation current without velocity saturation

$$I_{ds} = \frac{\beta}{2}(V_{gs} - V_t)^2$$

changes to the equation below if the transistor were completely velocity saturated

$$Ids = C_{ox} W (V_{gs} - V_t) \upsilon_{sat}$$

Thus current is linearly dependent rather than quadratically dependent

For moderate supply voltages, transistors operate in a region where the velocity no longer increases linearly with field, but also is not completely saturated.

Velocity Saturation and Mobility Degradation

The *α-power law model*, given below provides a simple approximation

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \quad \textbf{Cutoff} \\ I_{dsat}\dfrac{V_{ds}}{V_{dsat}} & V_{ds} < V_{dsat} \quad \textbf{Linear} \\ I_{dsat} & V_{ds} > V_{dsat} \quad \textbf{Saturation} \end{cases}$$

where,

$$I_{dsat} = P_c \frac{\beta}{2}(V_{gs} - V_t)^{\alpha}$$

$$V_{dsat} = P_v (V_{gs} - V_t)^{\alpha/2}$$

where the parameters, βP_c, α and P_v are obtained by curve-fitting the I-V characteristics.

Transistors with long channels or low V_{DD} have $\alpha=2$, and as they become completely velocity saturated, increasing V_{gs} has less effect on current and α decreases to 1.

For short channel devices, the lateral field increases and transistors become more velocity saturated (α closer to 1) if the supply is held constant.

E.g. 2µm device velocity saturated at $V_{DD} = 4V$, and a 0.18µm device above $V_{DD} = 0.36V$

Strong vertical electric fields resulting from large V_{gs} reduce carrier mobility µ. This effect is called *mobility degradation* and is captured by α in the α-power law model.

Channel Length Modulation

Ideally, I_{ds} is independent of V_{ds} in saturation making the transistor a perfect current source.

The reverse-biased pn junction between the drain and the substrate forms a depletion region with a width L_d that increases V_{db}.

If the source voltage is same as the substrate voltage ($V_{db} \sim V_{ds}$), increasing V_{ds} decreases the effective channel length, resulting in higher currents with increasing V_{ds}.

Can be crudely modeled using

$$I_{ds} = \beta \frac{(V_{gs} - V_t)^2}{2}(1 + \lambda V_{ds})$$

The parameter λ is an empirical *channel length modulation factor* (not λ in the layout).

λ is inversely proportional to channel-length and so as transistors L's become shorter, this effect becomes relatively more important.

More important for analog designers (than digital designer) as it reduces gain of amplifiers.

Subthreshold Conduction

Ideally no current flows from source to drain when $V_{gs} < V_t$. In real transistors this is not true and current drops off exponentially.

Usually termed as ***leakage***, undesired current when transistor is OFF (testing problems).

$$I_{ds} = I_{ds0} e^{\frac{V_{gs} - V_t}{n v_T}} \left(1 - e^{\frac{-V_{ds}}{v_T}}\right)$$

$$I_{ds0} = \beta v_T^2 e^{1.8}$$

I_{ds0} is current at threshold
process and geometry dependent

$e^{1.8}$ found empirically

n process dependent affected by depletion region characteristics (usually 1.4-1.5)

v_T thermal voltage

The last term in the I_{ds} equation indicates leakage is 0 if $V_{ds} = 0$, but increases to its full value when V_{ds} is a few multiples of the thermal voltage (e.g. $V_{ds} > 50$mV).

Leakage increases exponentially as V_t decreases or as temperature rises. Impacted by ***drain-induced barrier lowering (DIBL)***, in which a positive V_{ds} effectively reduces V_t.

$$V_t' = V_t - \eta V_{ds}$$

η DIBL coefficient (typically 0.02-0.1)

Again, effect is more pronounced in short-channel transistors.

Junction Leakage

The p-n junctions between diffusion and the substrate or well form diodes.

The well to substrate junction is another diode.

The substrate and well are tied to GND or V_{DD} to ensure that these diodes remain reverse-biased. However, reverse-biased diodes still conduct a small amount of current.

$$I_D = I_s\left(e^{\frac{V_D}{v_T}} - 1\right)$$

I_s, diode reverse-biased saturation current, depends on the area and the perimeter of the diffusion region. V_d is the diode voltage (e.g. V_{sb} or V_{db}).

When junction is reverse-biased by significantly more that the thermal voltage, the leakage is just $-I_s$, generally in the 0.1-0.01 fA/μm^2 range.

Modern transistors (low thresholds), subthreshold conduction far exceeds junction leakage.

Tunneling

According to quantum mechanics, there is a finite probability that carriers will tunnel through the gate oxide.

This results in ***gate leakage*** current flowing into the gate.

Probability of tunneling drops off exponentially with oxide thickness, and so was negligible until recently.

For gate oxide thickness of 1.5-2 nm, tunneling current becomes a factor and may become comparable to substhreshold leakage in advanced technologies.

Experiments show that the gate oxide (SiO_2) thickness t_{ox}, must not be less that 0.8nm.

To keep dimensions in perspective, a SiO_2 atomic layer is about 0.3 nm !!!

High C_{ox}, is important for good transistors, so research has been focussed on using alternative gate insulator with a high dielectric constant. One contender is Si_3N_4.

Temperature and Geometry Dependence

Temperature influences many transistor characteristics.

- Carrier mobility decreases with temperature
- The magnitude of the threshold voltage decreases nearly linearly with temperature
- Junction leakage increases with temperature because I_s is strongly temperature dependent.

Net effect: ***negative temperature coefficient***.

- ON current decreases, OFF current increases, worse performance at high temperature

Layout designers draw transistors with some width and length, W_{drawn} and L_{drawn}.

The actual dimensions may differ, due to polysilicon overetching to provide shorter channels, lateral diffusion of source and drain under the gate, diffusion of the substrate.

Effective dimensions should be used rather than drawn dimensions for analysis or values can be significantly off.

Below 0.25µm, transistor orientation and amount of nearby poly affect the effective length.