# Chapter1. The Devices: Diode and MOS Transistors (Chapter 3 in textbook)

[Adapted from Rabaey's *Digital Integrated Circuits*, ©2002, J. Rabaey et al.]



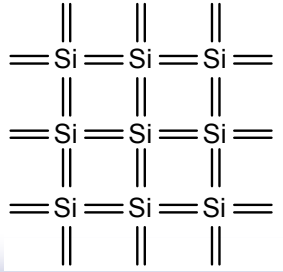
## Goal of this chapter

- Present intuitive understanding of device operation
- Introduction of basic device equations
- Introduction of models for manual analysis
- Introduction of models for SPICE simulation
- Analysis of secondary and deep-sub-micron effects
- Future trends



## Semiconductors

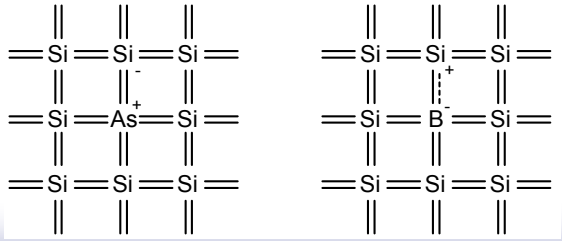
- Semiconductor: electrical conductivity greater than insulator, but less than conductor.
- Typical semiconductor materials: Si, Ge, GaAs, etc.
- Transistors are built on a silicon (Si) substrate
- Silicon is a Group IV semiconductor material
- Forms crystal lattice with bonds to four neighbors






## Silicon Material

- Pure silicon: very few free carriers, conducts poorly
- Intrinsic Si (undoped):
  $n=p=n_i=1.5\times10^{10}/cm^3$ (room temperature)
  where $n$ (p): free-electron(hole) concentration,
  $n_i$: intrinsic carrier concentration.
- Adding dopants increases the conductivity
  - ✓ Group V dopants: contribute extra electron (n-type)
  - ✓ Group III dopants: contribute hole (missing electron, p-type)



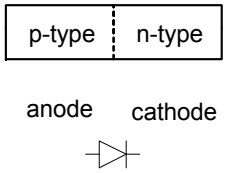
## Dopants

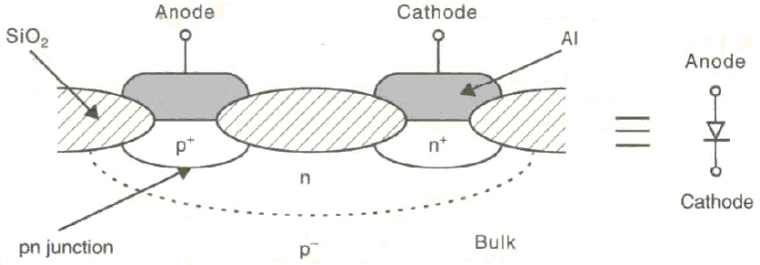
- Donor: impurity with valence ≥5 (e.g. P, As) => n-type with free electrons. Assume doping concentration of donor as $N_D$.
  1). Concentration of free electrons in n-type material ($n_n$):
  $n_n=N_D>>n_i$
  2). Concentration of free holes in n-type material ($p_n$)
  $p_n=n_i^2/N_D$
- Acceptor: impurity with valence ≤3 (e.g. B) => p-type with free holes. Assume doping concentration of acceptor as $N_A$.
  1). Concentration of free holes in p-type material ($p_p$):
  $p_p=N_A>>n_i$
  2). Concentration of free electrons in p-type material ($n_p$)
  $n_p=n_i^2/N_A$



## p-n Junctions

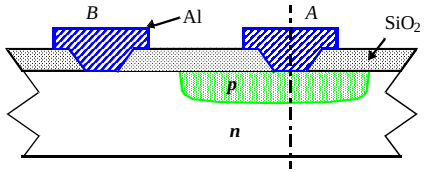
- A junction between p-type and n-type semiconductor forms a diode.
- Current flows only in one direction



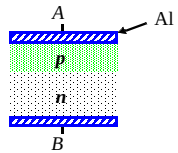





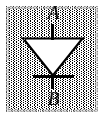

## The Diode




Cross-section of *pn*-junction in an IC process




One-dimensional representation

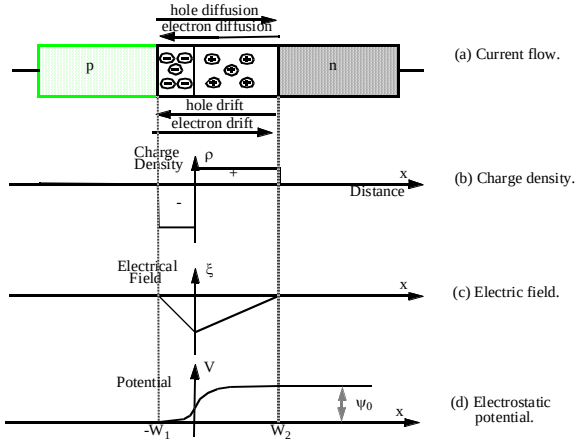


diode symbol

*Mostly occurring as parasitic element in Digital ICs*



## Depletion Region



## PN Junction

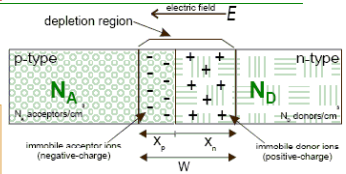
- Built-in potential of PN junction: $\Phi_0 = V_T \ln\left(\frac{N_A N_D}{n_i^2}\right)$

where the thermal voltage $V_T$: $V_T = kT/q$, in it,
k: Boltzmann constant, $k = 1.38\times10^{-23}$ m²·kg·s⁻²·K⁻¹,
q: electron charge, $q = 1.6\times10^{-19}$C.
At room temperature (T=300K), $V_T = 0.026$V.



## Reversed-biased Diode: Depletion Width

- Depletion Width
  - use Poisson's equation & charge neutrality
  - $W = x_p + x_n$




$$x_p = \left[\frac{2\varepsilon(\Psi_0 + V_R)N_D}{qN_A(N_D + N_A)}\right]^{1/2} \qquad x_n = \left[\frac{2\varepsilon(\Psi_0 + V_R)N_A}{qN_D(N_D + N_A)}\right]^{1/2}$$

  - where $V_R$ is applied reverse bias

$$W = \left[\frac{2\varepsilon(\Psi_0 + V_R)}{q}\frac{N_D + N_A}{N_D N_A}\right]^{1/2}$$

ε is the permittivity of Si
ε = 1.04x10⁻¹² F/cm
ε = $K_S\varepsilon_0$, where $\varepsilon_0 = 8.85\times10^{-14}$ F/cm
and $K_S$ = 11.8 is the relative permittivity of silicon

$$\Psi_0 = V_T \ln\left(\frac{N_A N_D}{n_i^2}\right)$$

- One-sided Step Junction
  - if $N_A \gg N_D$ (p+n diode)
    - most of junction on n-side

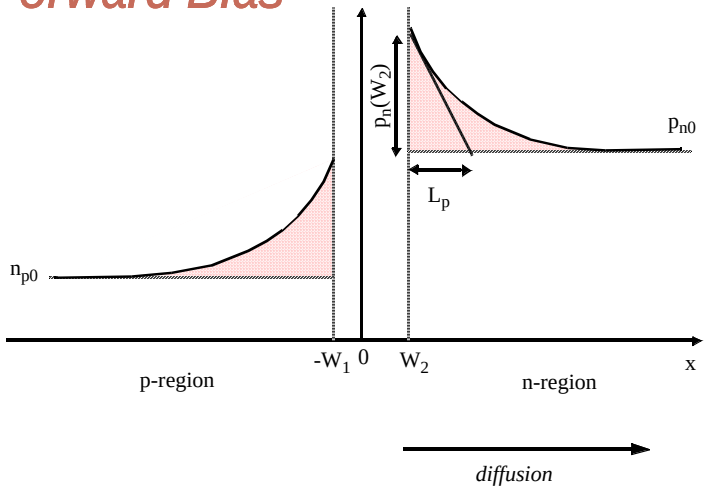
$$W \cong x_n = \left[\frac{2\varepsilon(\Psi_0 + V_R)}{qN_D}\right]^{1/2}$$

  - if $N_D \gg N_A$ (n+p diode)
    - most of junction on p-side

$$W \cong x_p = \left[\frac{2\varepsilon(\Psi_0 + V_R)}{qN_A}\right]^{1/2}$$



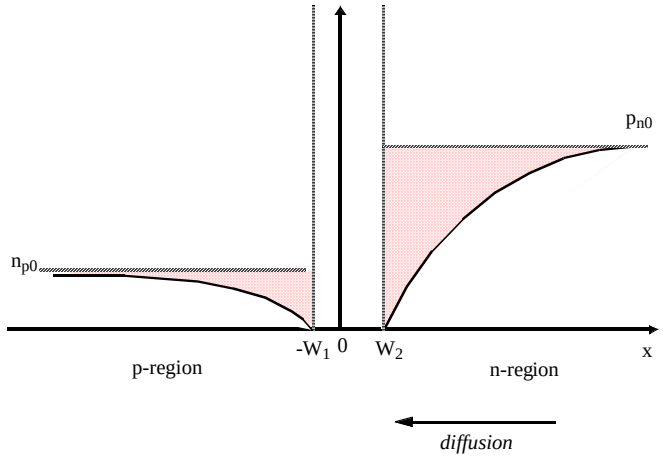
## Forward Bias




*Typically avoided in Digital ICs*



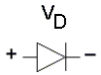
## Reverse Bias




*The Dominant Operation Mode*

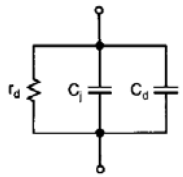
## Forward-biased Junction

- Drift current: current flow due to concentration gradient of the minority charge near junction
- Forward-bias current under forward-bias voltage $V_D$:




$$I_D = I_S\left(e^{V_D/V_T} - 1\right) \qquad I_S \propto A\left(\frac{1}{N_D} + \frac{1}{N_A}\right)$$

- Small-signal model of forward-biased diode




1). Incremental resistance $r_d$

$$\frac{1}{r_d} = \frac{dI_D}{dV_D} = I_S\frac{e^{V_D/V_T}}{V_T} = \frac{I_D}{V_T}$$

2). Depletion capacitance $C_j$: for depletion width change and immobile charge change in depletion region.

3). Diffusion capacitance $C_d$: for minority carrier concentration change close to junction.

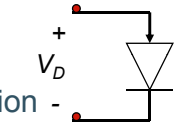
$$C_d = \frac{\tau_T}{r_d}$$

$\tau_T$: transition time of diode (specified for a given technology)



## Reverse Bias Diode

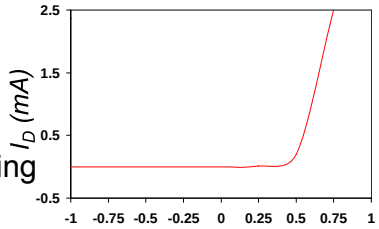
- The *ideal diode equation* (for both forward and reverse-bias conditions) is:




$$I_D = I_S(e^{V_D/\phi_T} - 1)$$

where $V_D$ is the voltage applied to the junction

- a forward-bias lowers the potential barrier allowing carriers to flow across the diode junction
- a reverse-bias raises the potential barrier and the diode becomes non-conducting

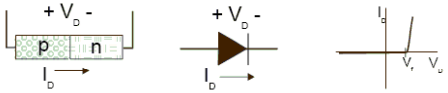


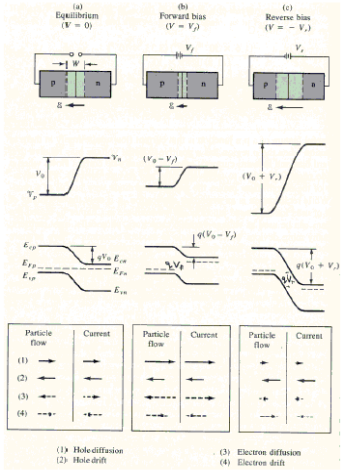

$\phi_T = kT/q = 26mV$ at 300K

$I_S$ is the saturation current of the diode



## Diode Biasing and Current Flow



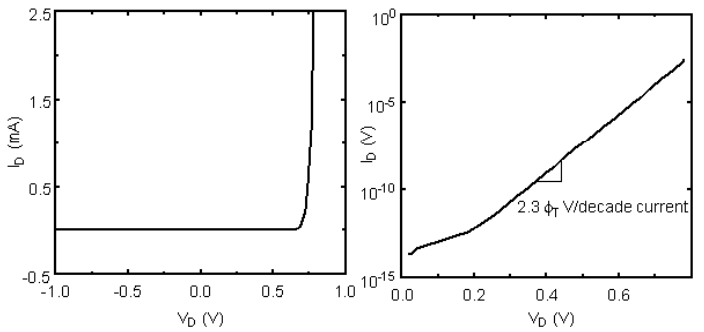

- Forward Bias; $V_D > \Psi_0$
  - acts against built-in potential
  - depletion width reduced
  - diffusion currents increase with $V_D$
    - minority carrier diffusion

$$I_D = I_S\left(e^{V_D/V_T} - 1\right) \qquad I_S \propto A\left(\frac{1}{N_D} + \frac{1}{N_A}\right)$$

- Reverse Bias; $V_R = -V_D > 0$
  - acts to support built-in potential
  - depletion width increased
  - electric field increased
  - small drift current flows
    - considered leakage
    - small until $V_R$ is too high and breakdown occurs

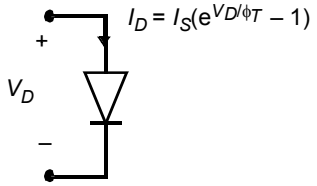


## Diode Current

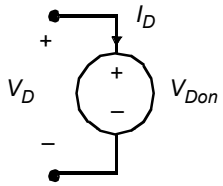



$$I_D = I_S\left(e^{V_D/\phi_T} - 1\right)$$

# Models for Manual Analysis




(a) Ideal diode model




(b) First-order diode model



# Junction Depletion Capacitance

- *Depletion capacitance: capacitance due to varying charge storage in depletion regions.*
- *Reason: bias voltage changes =>junction widths change =>charge in depletion region changes => capacitance effect.*
- *Depletion capacitance:*

$$C_j = \frac{dQ^+}{dV_R} = \left[\frac{qK_s\varepsilon_0}{2(\Phi_0 + V_R)}\frac{N_A N_D}{N_A + N_D}\right]^{1/2} = \frac{C_{j0}}{\sqrt{1 + \frac{V_R}{\Phi_0}}} \quad C_{j0} = \sqrt{\frac{qK_s\varepsilon_0}{2\Phi_0}\frac{N_A N_D}{N_A + N_D}}$$

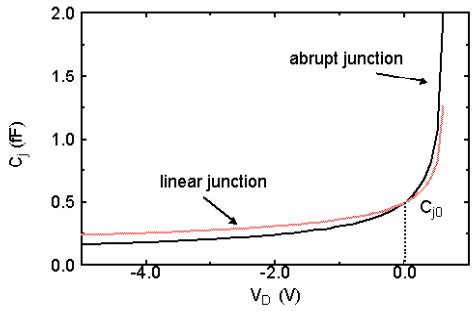
*$C_{j0}$: depletion capacitance per unit area at $V_R$=0.*

*If $N_A$>>$N_D$ (one-side diode):*

$$C_j = \left[\frac{qK_s\varepsilon_0 N_D}{2(\Phi_0 + V_R)}\right]^{1/2} = \frac{C_{j0}}{\sqrt{1 + \frac{V_R}{\Phi_0}}} \quad \text{where} \quad C_{j0} = \sqrt{\frac{qK_s\varepsilon_0 N_D}{2\Phi_0}}$$



# Junction Capacitance

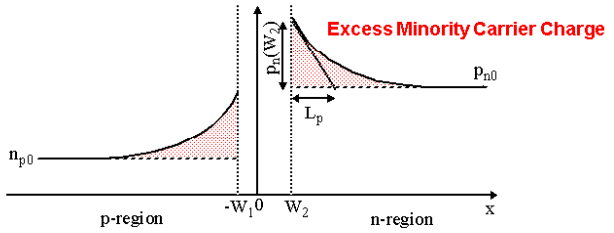



$$C_j = \frac{C_{j0}}{(1 - V_D/\phi_0)^m}$$

m = 0.5: abrupt junction
m = 0.33: linear junction

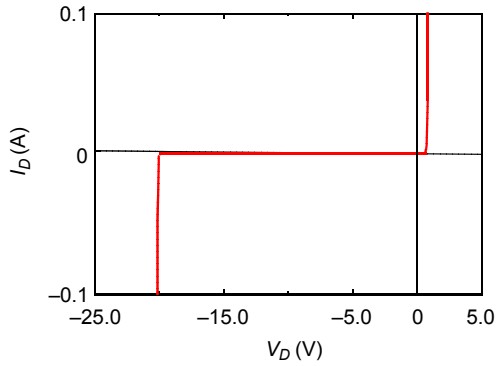


# Diffusion Capacitance




$$C_d = \frac{dQ_D}{dV_D} = \tau_T \frac{dI_D}{dV_D} \approx \frac{\tau_T I_D}{\phi_T}$$

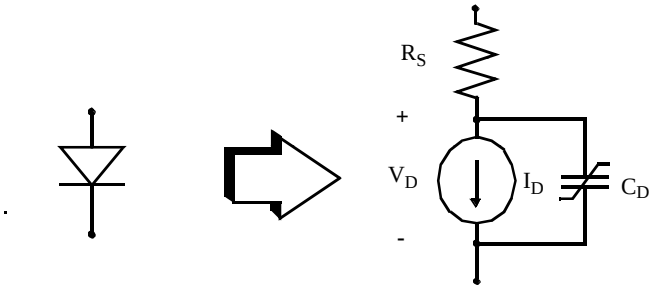
# Secondary Effects




**Avalanche Breakdown**



# Diode Model





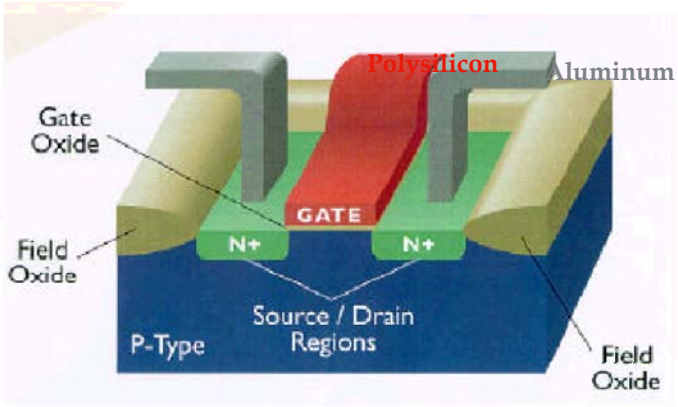

# SPICE Parameters

| Parameter Name | Symbol | SPICE Name | Units | Default Value |
|---|---|---|---|---|
| Saturation current | $I_S$ | IS | A | 1.0 E-14 |
| Emission coefficient | $n$ | N | - | 1 |
| Series resistance | $R_S$ | RS | Ω | 0 |
| Transit time | $\tau_T$ | TT | sec | 0 |
| Zero-bias junction capacitance | $C_{j0}$ | CJ0 | F | 0 |
| Grading coefficient | $m$ | M | - | 0.5 |
| Junction potential | $\phi_0$ | VJ | V | 1 |

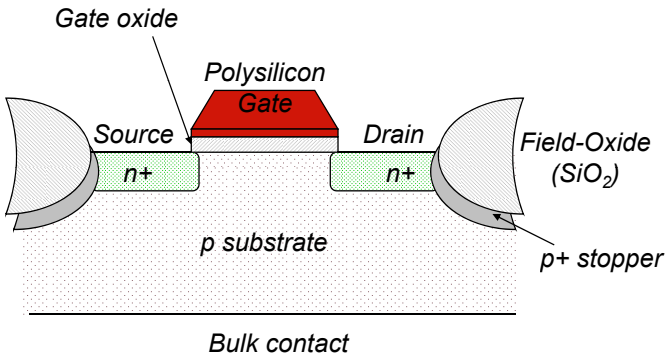
First Order SPICE diode model parameters.



# Device: The MOS Transistor



# Device: The MOS Transistor

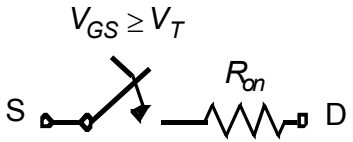


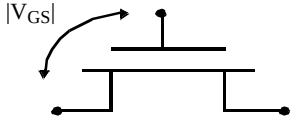

CROSS-SECTION of NMOS Transistor



# What is a Transistor?

A Switch! ⟷ An MOS Transistor

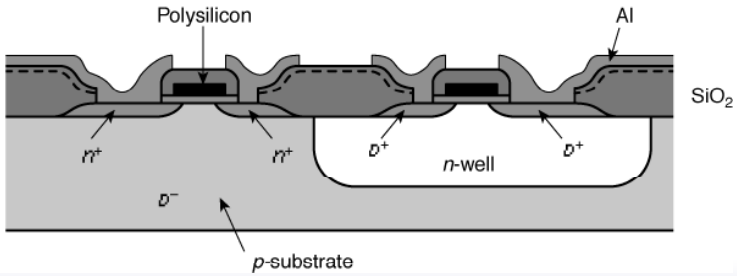




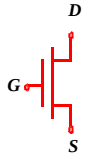

# Power Supply Voltage for CMOS

- GND = 0 V
- In 1980's, $V_{DD}$ = 5V
- $V_{DD}$ has decreased in modern processes
  - High $V_{DD}$ would damage modern tiny transistors
  - Lower $V_{DD}$ saves power
- $V_{DD}$ = 3.3V, 2.5V, 1.8V, 1.5V, 1.2V, 1.0V, …

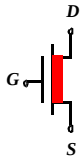




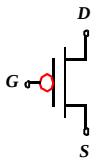

# MOS Transistors - Types and Symbols



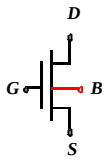

NMOS Enhancement




NMOS Depletion

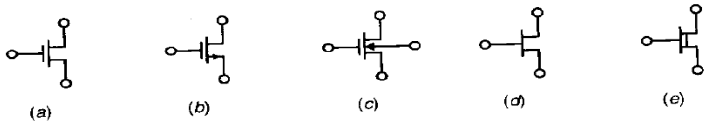



PMOS Enhancement

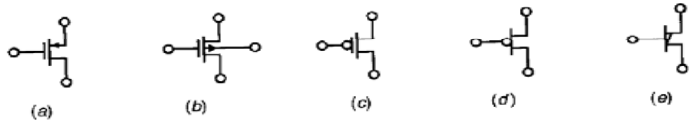



NMOS with Bulk Contact

## MOS transistors: Types and Symbols




Commonly used symbols for n-channel transistor




- *Bulk connection:*

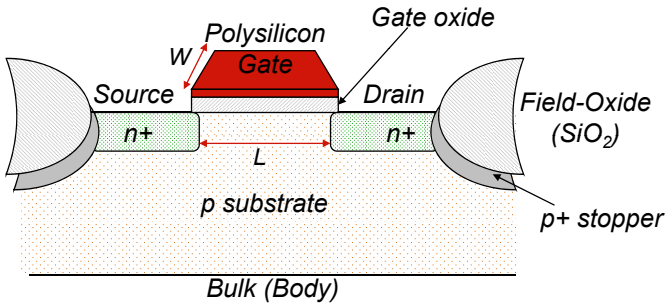
✓ *NMOS: bulk is generally connected to the lowest voltage level available in the circuit (Gnd for digital VLSI, $-V_{ss}$ for analog VLSI).*

✓ *PMOS: bulk is generally connected to the highest voltage level in the circuit ($V_{dd}$ for digital VLSI, $+V_{ss}$ for analog VLSI)*



## The NMOS Transistor Cross Section

*n areas have been doped with donor ions (arsenic) of concentration $N_D$ - electrons are the majority carriers*

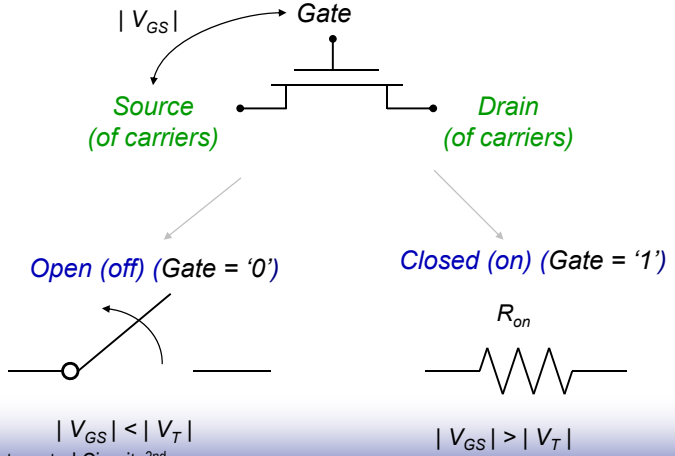



*p areas have been doped with acceptor ions (boron) of concentration $N_A$ - holes are the majority carriers*



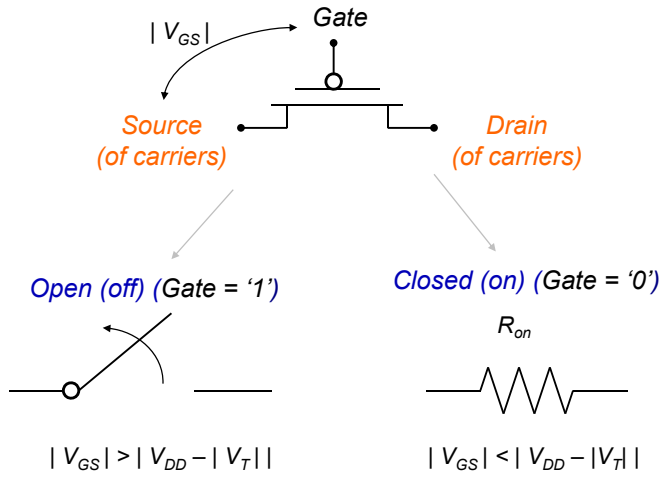
## Switch Model of NMOS Transistor

- We can view MOS transistors as electrically controlled switches
- Voltage at gate controls path from source to drain






## Switch Model of PMOS Transistor

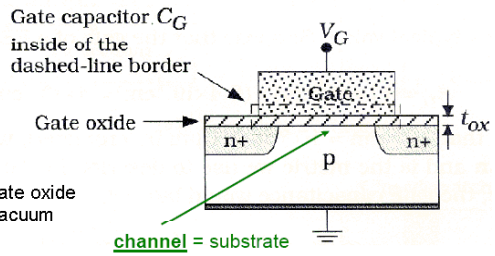


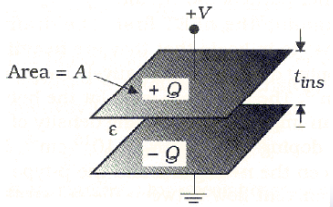
# MOSFET Gate Operation

- Gate Capacitance $C_G$
  - gate-substrate parallel plate capacitor
  - $C_G = \varepsilon_{ox} A_G / t_{ox}$ [F]
    - $\varepsilon_{ox}$: dielectric constant of gate oxide
    - $\varepsilon_o$ : dielectric constant of vacuum

      $\varepsilon_{ox} = 3.9\ \varepsilon_o$

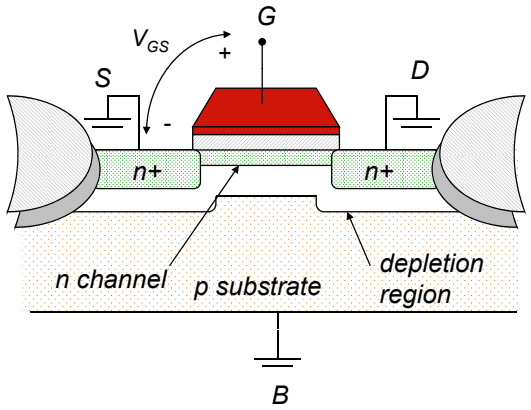
      $\varepsilon_o = 8.85 \times 10^{-14}$ [F/cm]
    - $A_G$=gate area = L·W [cm²]
    - $t_{ox}$: gate oxide thickness
- Oxide capacitance per unit area - $C_{ox}$
  - $C_{ox} = \varepsilon_{ox}/t_{ox}$ [F/cm²]
  - $C_G = C_{ox} A_G$ [F]






# Threshold Voltage Concept




- *Depletion region: area devoid of mobile carriers (holes)*
- *Inversion region: n-channel region under oxide*
- *The value of $V_{GS}$ where strong inversion occurs is called the threshold voltage, $V_T$*



# The Threshold Voltage

*where* $V_T = V_{T0} + \gamma(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|})$

*$V_{T0}$ is the threshold voltage at $V_{SB} = 0$ and is mostly a function of the manufacturing process*

- *Difference in work-function between gate and substrate material, oxide thickness, Fermi voltage, charge of impurities trapped at the surface, dosage of implanted ions, etc.*

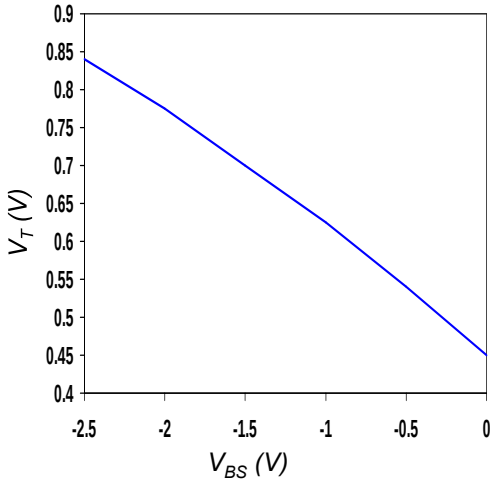
*$V_{SB}$ is the source-bulk voltage*

*$\phi_F = -\phi_T \ln(N_A/n_i)$ is the Fermi potential ($\phi_T = kT/q = 26$mV at 300K is the thermal voltage; $N_A$ is the acceptor ion concentration; $n_i \approx 1.5 \times 10^{10}$ cm$^{-3}$ at 300K is the intrinsic carrier concentration in pure silicon)*

*$\gamma = \sqrt{(2q\varepsilon_{si}N_A)/C_{ox}}$ is the body-effect coefficient (impact of changes in $V_{SB}$) ($\varepsilon_{si}=1.053 \times 10^{-10}$F/m is the permittivity of silicon; $C_{ox} = \varepsilon_{ox}/t_{ox}$ is the gate oxide capacitance with $\varepsilon_{ox}=3.5 \times 10^{-11}$F/m)*

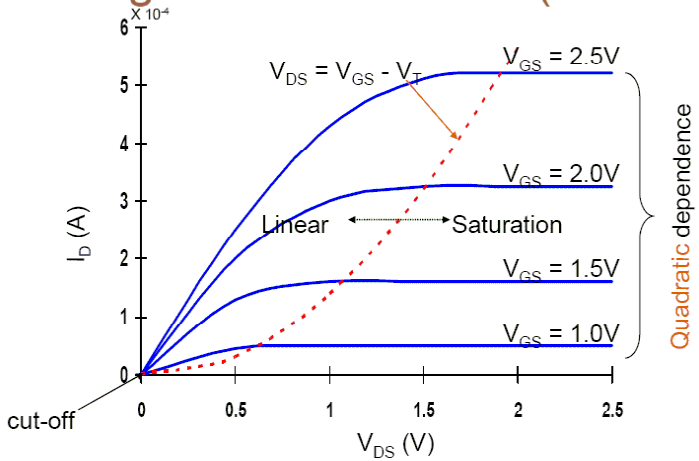


# The Body Effect




- *$V_{SB}$ is the substrate bias voltage (normally positive for n-channel devices with the body tied to ground)*
- *A negative bias causes $V_T$ to increase from 0.45V to 0.85V*

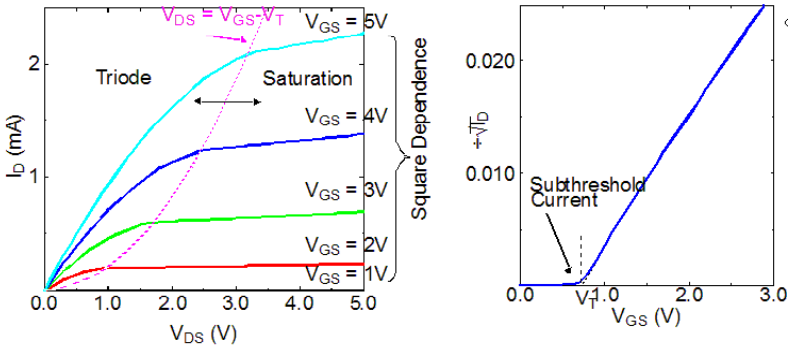
# Long Channel I-V Plot (NMOS)




NMOS transistor, 0.25um, $L_d$ = 10um, W/L = 1.5, $V_{DD}$ = 2.5V, $V_T$ = 0.4V



# Long Channel I-V Relation (NMOS)




(a) $I_D$ as a function of $V_{DS}$

(b) $\sqrt{I_D}$ as a function of $V_{GS}$ (for $V_{DS}$ = 5V)

NMOS Enhancement Transistor: W = 100 μm, L = 20 μm



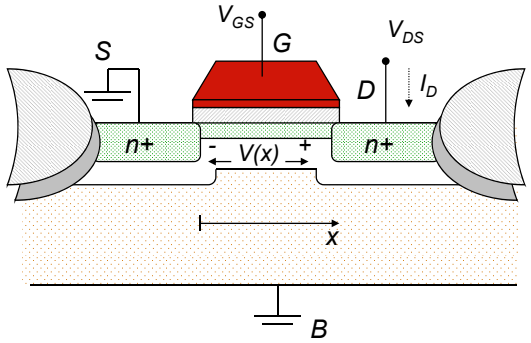
# MOSFET Operating Regions

- Strong inversion when $V_{GS} > V_T$
  - Linear (resistive) when $V_{DS} < V_{DS\text{-}sat}$
  - Saturated (constant current) when $V_{DS} \geq V_{DS\text{-}sat}$
- Weak inversion (sub-threshold) when $V_{GS} \leq V_T$
  - Exponential in $V_{GS}$ with linear $V_{DS}$ dependence
- $V_{DS\text{-}sat}$: Drain-source voltage when the channel becomes pinched off. $V_{DS\text{-}sat} = V_{GS} - V_T$
- Transconductance
  - process transconductance, $k'_n = \mu_n C_{ox}$ (for NMOS), $k'_p = \mu_p C_{ox}$ (for PMOS).
    - constant for a given fabrication process
  - device transconductance, $\beta_n = k'_n W/L$ (for NMOS), $\beta_p = k'_p W/L$ (for PMOS).



# Transistor in Linear Mode

*Assuming $V_{GS} > V_T$*




- *When $V_{GS} > V_T$, and $V_{DS} < V_{GS} - V_T$, a current flows from drain to source*
- *The current is a linear function of both $V_{GS}$ and $V_{DS}$*

## Voltage-Current Relation: Linear Mode

For long-channel devices (L > 0.25 micron)

- When $V_{DS} \leq V_{GS} - V_T$

$$I_D = k'_n \, W/L \, [(V_{GS} - V_T)V_{DS} - V_{DS}^2/2]$$

where

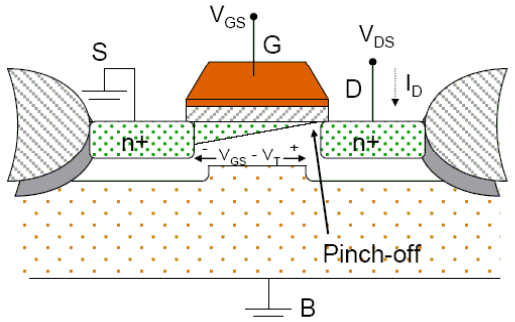
$k'_n = \mu_n C_{ox} = \mu_n \varepsilon_{ox}/t_{ox}$ = is the process transconductance parameter ($\mu_n$ is the carrier mobility ($m^2$/Vsec))

$k_n = k'_n \, W/L$ is the gain factor of the device

For small $V_{DS}$, there is a linear dependence between $V_{DS}$ and $I_D$, hence the name resistive or linear region



## Transistor in Saturation Mode




Assuming $V_{GS} > V_T$
$V_{DS} > V_{GS} - V_T$

- When charge in the inversion layer=0, number of mobile electrons at drain reduces drastically. This is pinch-off
- Number of carriers arriving at pinch-off point from source remain the same → current flowing from drain to source remains constant.
- Increasing $V_{DS}$ beyond $V_{GS}$-$V_T$ causes the depletion region at drain to grow → effective channel length L decreases
- Since ID is inversely proportional to L, as L decreases, ID increases, but compared to linear mode, it is relatively constant



## Voltage-Current Relation: Saturation Mode

For long channel devices

- When $V_{DS} \geq V_{GS} - V_T$

$$I_D' = k'_n/2 \, W/L \, [(V_{GS} - V_T)^2]$$

since the voltage difference over the induced channel (from the pinch-off point to the source) remains fixed at $V_{GS} - V_T$

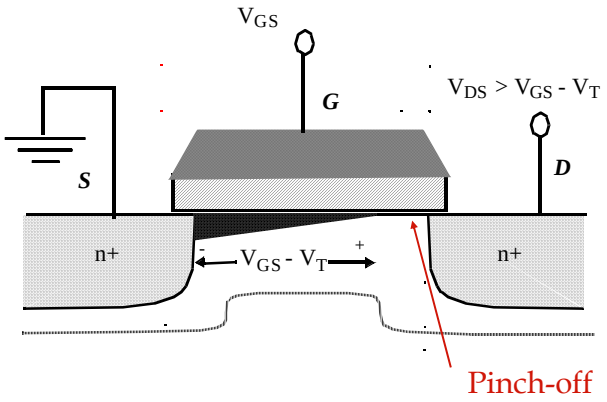
- However, the effective length of the conductive channel is modulated by the applied $V_{DS}$, so

$$I_D = I_D' \, (1 + \lambda V_{DS})$$

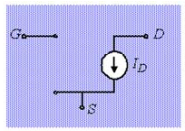
where $\lambda$ is the channel-length modulation (varies with the inverse of the channel length)



## Transistor in Saturation



# Current-Voltage Relations




$$V_T = V_{T0} + \gamma\left(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|}\right)$$

$$V_{DS\text{-}sat} = V_{GS} - V_T = V_{eff}$$

**1). Linear (Triode) Region: $V_{GS} > V_T, V_{DS} \le V_{GS} - V_T$**

$$I_D = k'_n \frac{W}{L}\left[(V_{GS} - V_T)V_{DS} - \frac{V_{DS}^2}{2}\right]$$

with $k'_n = \mu_n C_{ox} = \dfrac{\mu_n \varepsilon_{ox}}{t_{ox}} \Rightarrow$ Process Transconductance Parameter

**2). Saturation (Active) Region : $V_{GS} > V_T, V_{DS} \ge V_{GS} - V_T$**

Channel Length Modulation

$$I_D = \frac{k'_n}{2}\frac{W}{L}(V_{GS} - V_T)^2[1 + \lambda(V_{DS})]$$

**3). Sub-threshold (Cutoff) Region: $V_{GS} \le V_T$**

$$I_D \doteq 0$$

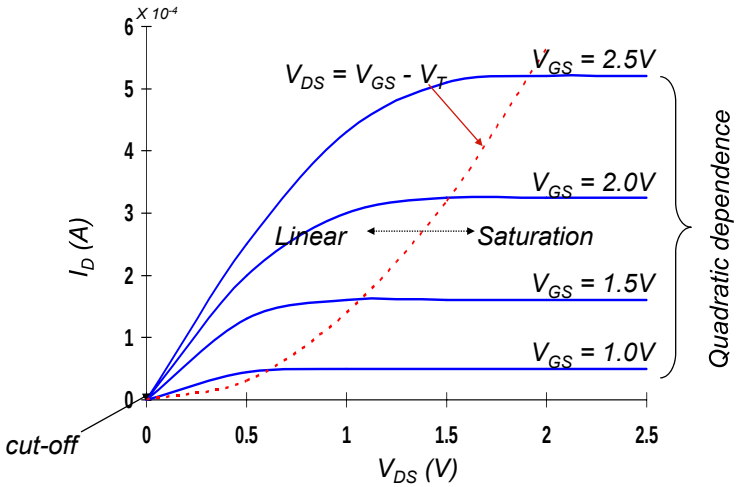


# Current Determinates

- For a fixed $V_{DS}$ and $V_{GS}$ (> $V_T$), $I_{DS}$ is a function of
  - the distance between the source and drain – L
  - the channel width – W
  - the threshold voltage – $V_T$
  - the thickness of the $SiO_2$ – $t_{ox}$
  - the dielectric of the gate insulator ($SiO_2$) – $\varepsilon_{ox}$
  - the carrier mobility
    - for nfets: $\mu_n$ = 500 cm²/V-sec
    - for pfets: $\mu_p$ = 180 cm²/V-sec



# Long Channel I-V Plot (NMOS)




*NMOS transistor, 0.25um, $L_d$ = 10um, W/L = 1.5, $V_{DD}$ = 2.5V, $V_T$ = 0.4V*



# Second Order Effects

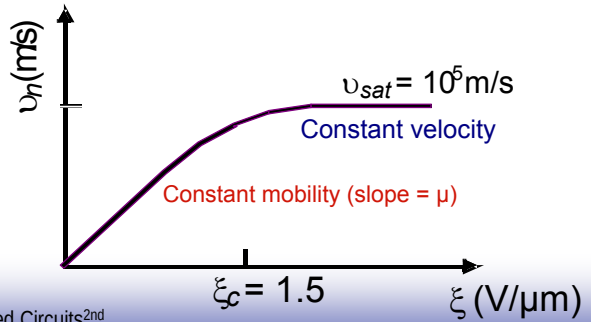
- Channel Length Modulation
  - Square Law Equation predicts $I_D$ is constant with $V_{DS}$
  - However, $I_D$ actually increases slightly with $V_{DS}$
    - due to effective channel getting shorter as $V_{DS}$ increases
    - effect called channel length modulation
  - Channel Length Modulation factor, $\lambda$
    - models change in channel length with $V_{DS}$
  - Corrected $I_D$ equation

$$I_D = \frac{\mu_n C_{OX}}{2}\frac{W}{L}(V_{GS} - V_t)^2\left(1 + \lambda(V_{DS} - V_{eff})\right)$$

    - Veff = $V_{GS}$ - Vtn
- Body Effect
  - so far we have assumed that substrate and source are grounded
  - if source not at ground, source-to-bulk voltage exists, $V_{SB} > 0$
  - $V_{SB} > 0$ will increase the threshold voltage, Vtn = f($V_{SB}$)
  - called Body Effect, or Body-Bias Effect

# Velocity Saturation

- *For long channel device: carrier velocity increases linearly with E-field:* $v_n=-\mu_n\xi(x)=\mu_n(dV/dx)$
- *For transistors with very short channel length (short-channel devices), when E-field along the channel reaches a critical value $\xi_c$, the velocity of carriers saturates due to scattering effects (collisions suffered by carriers).*



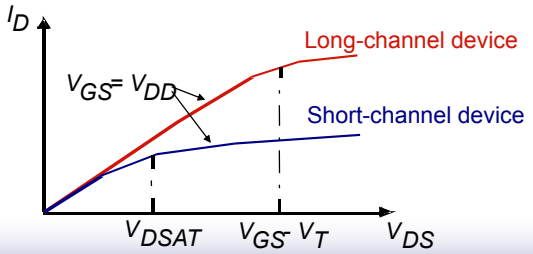



# Perspective

- *For short channel transistor, due to velocity saturation:*

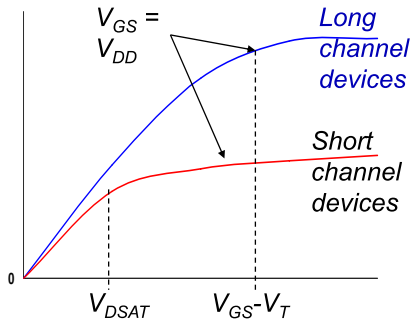
✓ *Transistor enters saturation before $V_{DS}$ reaches $V_{GS}$-$V_T$. Short-channel devices experience extended saturation region, and tend to operation more often in saturation conditions.*

✓ *Saturation current $I_{DSAT}$ shows linear dependence to $V_{GS}$ instead of squared dependence in long-channel device. This reduces the amount of current a transistor can deliver for a given control voltage.*






# Velocity Saturation Effects




*For short channel devices and large enough $V_{GS} - V_T$*

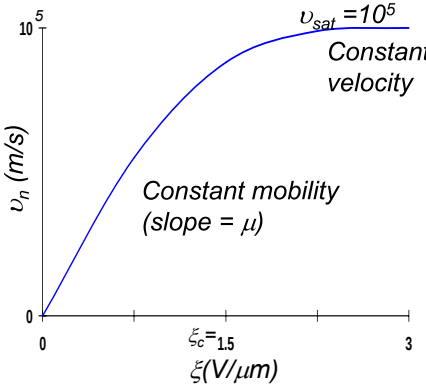
- *$V_{DSAT} < V_{GS} - V_T$ so the device enters saturation before $V_{DS}$ reaches $V_{GS} - V_T$ and operates more often in saturation*

- *$I_{DSAT}$ has a linear dependence wrt $V_{GS}$ so a reduced amount of current is delivered for a given control voltage*



# Short Channel Effects

- *Behavior of short channel device mainly due to*




- *Velocity saturation – the velocity of the carriers saturates due to scattering (collisions suffered by the carriers)*

- *For an NMOS device with L of .25μm, only a couple of volts difference between D and S are needed to reach velocity saturation*

## Voltage-Current Relation: Velocity Saturation

For short channel devices

- Linear: When $V_{DS} \leq V_{GS} - V_T$

$$I_D = \kappa(V_{DS})\, k'_n\, W/L\, [(V_{GS} - V_T)V_{DS} - V_{DS}^2/2]$$

where

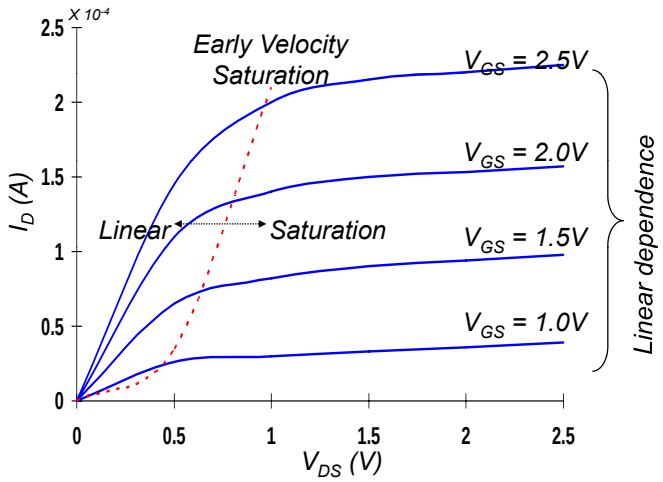
$\kappa(V) = 1/(1 + (V/\xi_c L))$ is a measure of the degree of velocity saturation

- Saturation: When $V_{DS} = V_{DSAT} \geq V_{GS} - V_T$

$$I_{DSat} = \kappa(V_{DSAT})\, k'_n\, W/L\, [(V_{GS} - V_T)V_{DSAT} - V_{DSAT}^2/2]$$



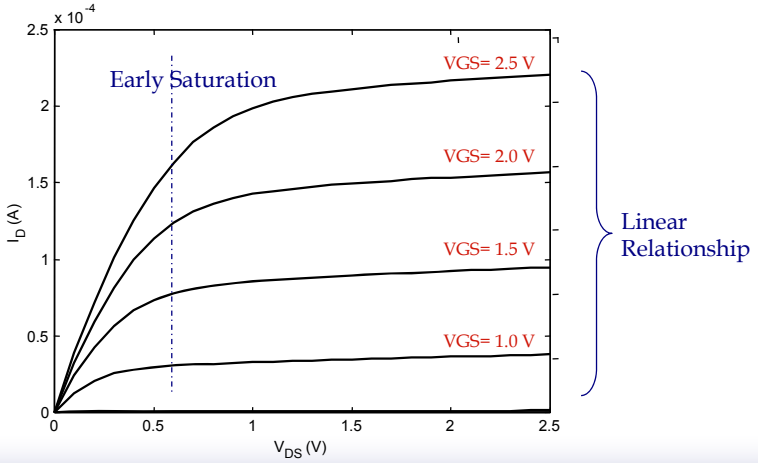
## Short Channel I-V Plot (NMOS)




*NMOS transistor, 0.25um, $L_d$ = 0.25um, W/L = 1.5, $V_{DD}$ = 2.5V, $V_T$ = 0.4V*

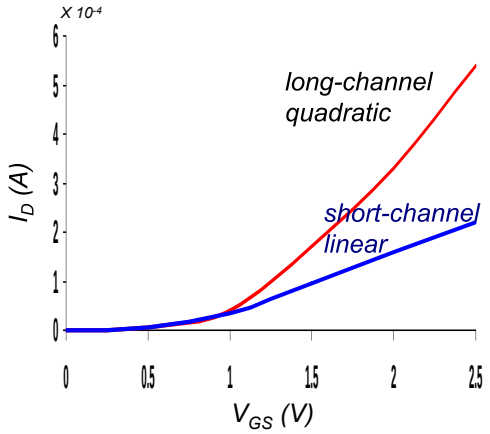


## Current-Voltage Relations The Deep-Submicron Era





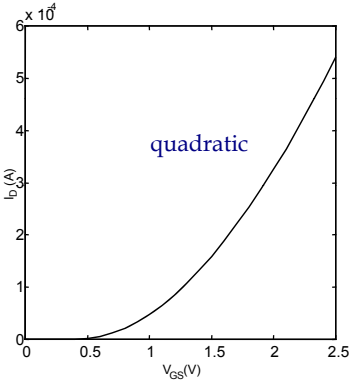

## MOS $I_D$-$V_{GS}$ Characteristics



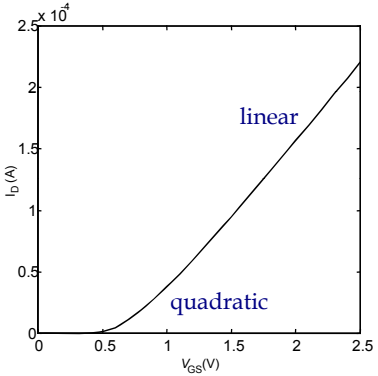

- *Linear (short-channel) versus quadratic (long-channel) dependence of $I_D$ on $V_{GS}$ in saturation*
- *Velocity-saturation causes the short-channel device to saturate at substantially smaller values of $V_{DS}$ resulting in a substantial drop in current drive*

*(for $V_{DS}$ = 2.5V, W/L = 1.5)*

# $I_D$ versus $V_{GS}$



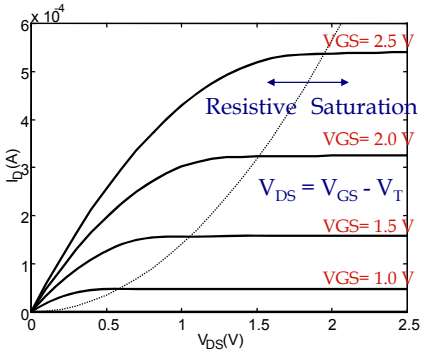

Long Channel



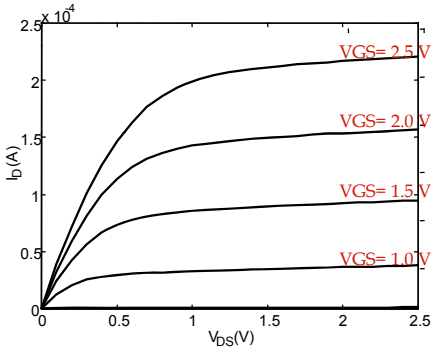

Short Channel



# $I_D$ versus $V_{DS}$




Long Channel

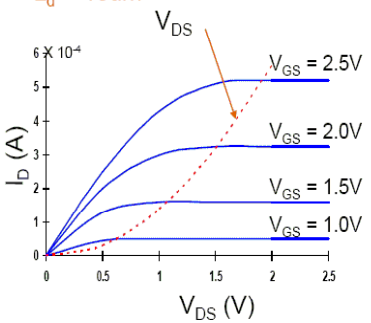



Short Channel



# Short Channel I-V Plot (NMOS)

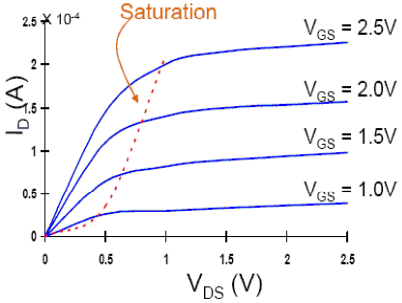
$L_d$ = 10um




Quadratic dependence on VGS

Early Velocity Saturation

$L_d$ = 0.25um

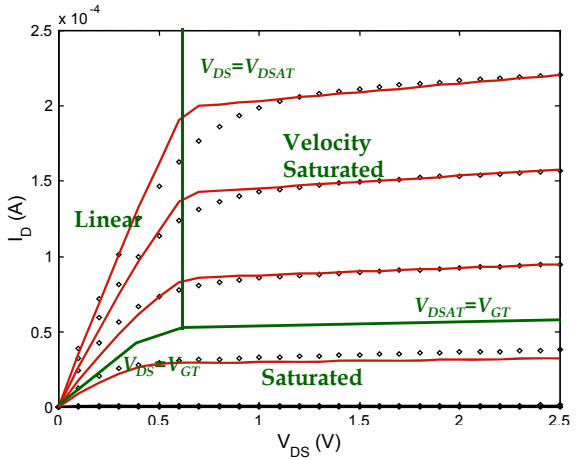



Linear dependence on VGS

- $I_{DSAT}$ has a linear dependence with respect to $V_{GS}$ (as opposed to quadratic) so a reduced amount of current is delivered for a given control voltage



# Simple Model versus SPICE



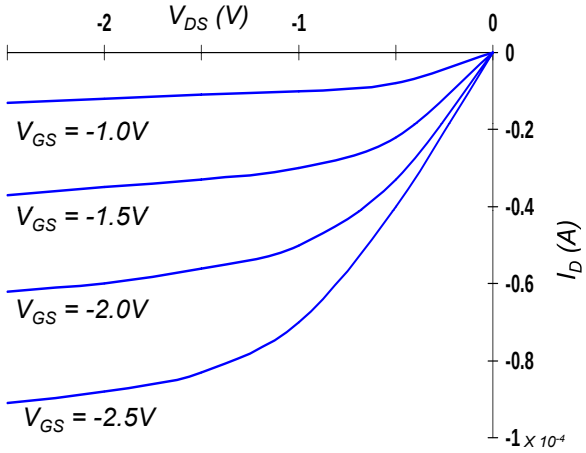
## pMOS Equations

- Analysis of nMOS applies to pMOS with following modifications
  - physical
    - change all n-tpye regions to p-type
    - change all p-type regions to n-type
      - substrate is n-type (nWell)
    - channel charge is positive (holes) and (+)ve charged ions
  - equations
    - change $V_{GS}$ to $V_{SG}$ ($V_{SG}$ typically = VDD - $V_G$)
    - change $V_{DS}$ to $V_{SD}$ ($V_{SD}$ typically = VDD - $V_D$)
    - change Vtn to |Vtp|
      - pMOS threshold is negative, nearly same magnitude as nMOS
  - other factors
    - lower surface mobility, typical value, $\mu_p$ = 220 cm²/V-sec
    - body effect, change $V_{SB}$ to $V_{BS}$



## Short Channel I-V Plot (PMOS)

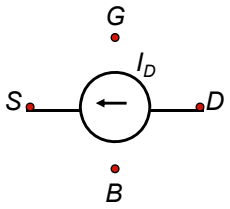
- *All polarities of all voltages and currents are reversed*




*PMOS transistor, 0.25um, $L_d$ = 0.25um, W/L = 1.5, $V_{DD}$ = 2.5V, $V_T$ = -0.4V*



## The MOS Current-Source Model




$I_D = 0$ for $V_{GS} - V_T \leq 0$

$I_D = k' W/L [(V_{GS} - V_T)V_{min} - V_{min}^2/2](1+\lambda V_{DS})$
for $V_{GS} - V_T \geq 0$

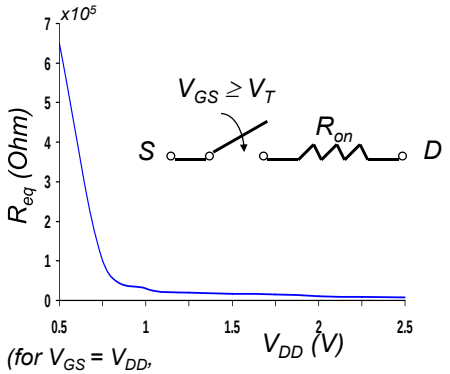
with $V_{min} = min(V_{GS} - V_T, V_{DS}, V_{DSAT})$
and $V_{GT} = V_{GS} - V_T$

- *Determined by the voltages at the four terminals and a set of five device parameters*

| | $V_{T0}$(V) | $\gamma$($V^{0.5}$) | $V_{DSAT}$(V) | k'($A/V^2$) | $\lambda$($V^{-1}$) |
|---|---|---|---|---|---|
| NMOS | 0.43 | 0.4 | 0.63 | 115 x $10^{-6}$ | 0.06 |
| PMOS | -0.4 | -0.4 | -1 | -30 x $10^{-6}$ | -0.1 |



## The Transistor Modeled as a Switch




(for $V_{GS} = V_{DD}$, $V_{DS} = V_{DD} \rightarrow V_{DD}/2$)

*Modeled as a switch with infinite off resistance and a finite on resistance, $R_{on}$*

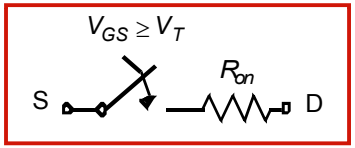
- *Resistance inversely proportional to W/L (doubling W halves $R_{on}$)*
- *For $V_{DD} >> V_T + V_{DSAT}/2$, $R_{on}$ independent of $V_{DD}$*
- *Once $V_{DD}$ approaches $V_T$, $R_{on}$ increases dramatically*

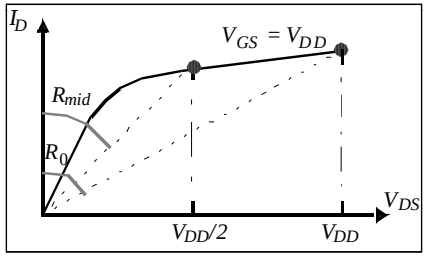
| $V_{DD}$(V) | 1 | 1.5 | 2 | 2.5 |
|---|---|---|---|---|
| NMOS(kΩ) | 35 | 19 | 15 | 13 |
| PMOS (kΩ) | 115 | 55 | 38 | 31 |

*$R_{on}$ (for W/L = 1)*
*For larger devices divide $R_{eq}$ by W/L*

# The Transistor as a Switch

- *In digital VLSI, transistor is treated as a switch with infinite "off" resistance, and a finite "on" resistance $R_{on}$.*
- *$R_{on}$ is time varying, nonlinear and dependent on transistor operating point. We can use the average value of resistances at the end points of the transition.*
- *$R_{eq}$ is inversely proportional to (W/L) ratio of transistor. Doubling the transistor width cuts the resistance in half. (IDSAT is linear to (W/L))*



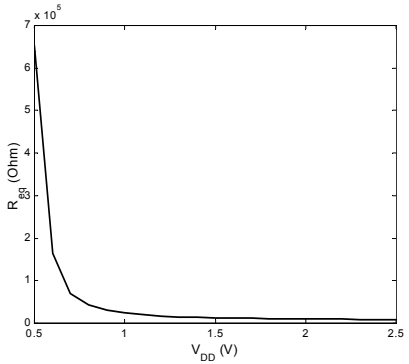




$$R_{eq} = \frac{1}{2}\left(\frac{V_{DD}}{I_{DSAT}(1+\lambda V_{DD})} + \frac{V_{DD}/2}{I_{DSAT}(1+\lambda V_{DD}/2)}\right) \approx \frac{3}{4}\frac{V_{DD}}{I_{DSAT}}\left(1-\frac{5}{6}\lambda V_{DD}\right)$$



# The Transistor as a Switch

- *For $V_{DD} >> V_T + V_{DSAT}/2$, $R_{eq}$ is virtually independent of $V_{DD}$. On a minor increase on $R_{eq}$ due to channel length modulation can be observed when raising $V_{DD}$.*
- *Once $V_{DD} \rightarrow V_T$, $R_{eq}$ increases dramatically.*



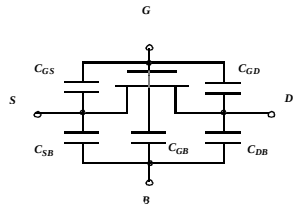



# MOS Capacitances Dynamic Behavior



## Capacitances of MOS Transistor

- *MOS capacitances according to physical mechanisms*
  - *Overlap capacitances: $C_{gso}$, $C_{gdo}$*
  - *Channel capacitances: $C_{gs}$, $C_{gd}$, $C_{gb}$*
  - *Diffusion capacitance: $C_{Sdiff}$, $C_{Ddiff}$*
- *Lumped capacitance model of MOS transistor: capacitances between terminals without considering physical mechanisms*
  - $C_{GS}=C_{gs}+C_{gso}$
  - $C_{GD}=C_{gd}+C_{gdo}$
  - $C_{GB}=C_{gb}$
  - $C_{SB}=C_{Sdiff}$
  - $C_{DB}=C_{Ddiff}$

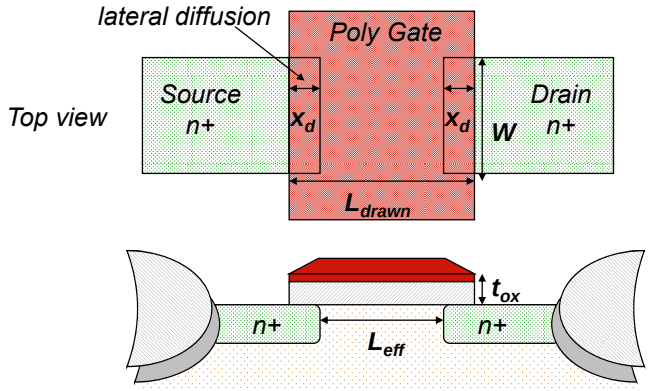


# MOS Intrinsic Capacitances

- Structure capacitances
- Channel capacitances
- Depletion regions of the reverse-biased *pn*-junctions of the drain and source



# MOS Structure Capacitances




*Overlap capacitance (linear)*

$$C_{GSO} = C_{GDO} = C_{ox}\, x_d\, W = C_o\, W$$



## MOS Capacitances: Structure Capacitances

- *MOS structure capacitance: originate from MOS structure*

*1). Gate capacitance $C_g$*

$C_g = C_{ox} W \cdot L$, where $C_{ox} = \varepsilon_{ox}/t_{ox}$

*Lateral diffusion: during to fabrication, both source and drain tends to extend somewhat below the oxide by $x_d$.*

*Effective channel length:*

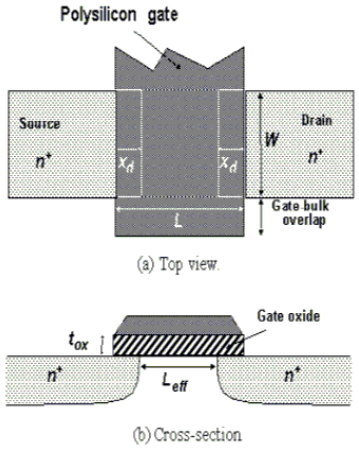
$L_{eff} = L - 2x_d$

*Gate capacitance considering $x_d$:*

$C_g = C_{ox} W \cdot L_{eff}$

*2). Overlap capacitance: $C_{gso}$, $C_{gdo}$*

$C_{gso} = C_{gdo} = C_{ox} \cdot x_d \cdot W$

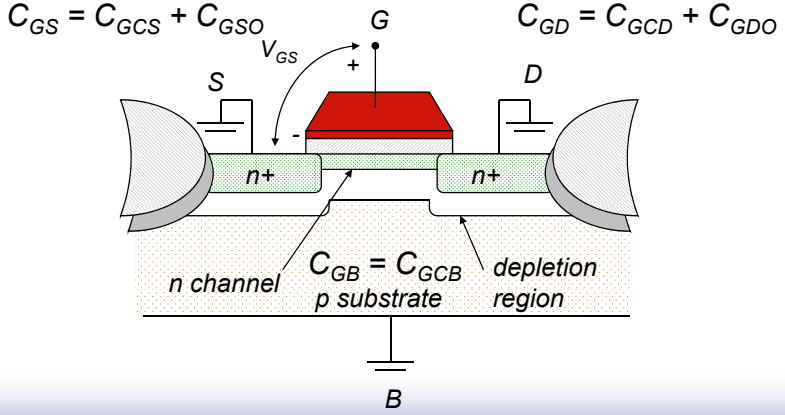





# MOS Channel Capacitances

- *The gate-to-channel capacitance depends upon the operating region and the terminal voltages*

$C_{GS} = C_{GCS} + C_{GSO}$ $\qquad$ $C_{GD} = C_{GCD} + C_{GDO}$



## Review: Summary of MOS Operating Regions

- Cutoff (really subthreshold) $V_{GS} \leq V_T$
  - Exponential in $V_{GS}$ with linear $V_{DS}$ dependence

    $I_D = I_S\, e^{(qV_{GS}/nkT)} (1 - e^{-(qV_{DS}/kT)}) (1 - \lambda V_{DS})$ where $n \geq 1$
- Strong Inversion $V_{GS} > V_T$
  - Linear (Resistive) $V_{DS} < V_{DSAT} = V_{GS} - V_T$

    $I_D = k' W/L [(V_{GS} - V_T)V_{DS} - V_{DS}^2/2] (1+\lambda V_{DS}) \kappa(V_{DS})$
  - Saturated (Constant Current) $V_{DS} \geq V_{DSAT} = V_{GS} - V_T$

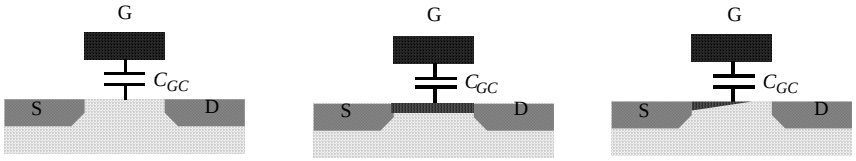
    $I_{DSat} = k' W/L [(V_{GS} - V_T)V_{DSAT} - V_{DSAT}^2/2] (1+\lambda V_{DS}) \kappa(V_{DSAT})$

| | $V_{T0}$(V) | $\gamma(V^{0.5})$ | $V_{DSAT}$(V) | $k'(A/V^2)$ | $\lambda(V^{-1})$ |
|---|---|---|---|---|---|
| NMOS | 0.43 | 0.4 | 0.63 | 115 x $10^{-6}$ | 0.06 |
| PMOS | -0.4 | -0.4 | -1 | -30 x $10^{-6}$ | -0.1 |



## Channel Capacitance

- *MOS channel capacitances: originate from channel charge*
- *MOS gate-to-channel capacitances: $C_{gb}$, $C_{gs}$, $C_{gd}$*
- *Distribution of gate capacitance among channel capacitances depends on operating conditions.*




**MOS gate-to-channel capacitances in different operation regions**

| Operation Region | $C_{gb}$ | $C_{gs}$ | $C_{gd}$ |
|---|---|---|---|
| Cutoff | $C_{ox}WL_{eff}$ | 0 | 0 |
| Triode | 0 | $C_{ox}WL_{eff}/2$ | $C_{ox}WL_{eff}/2$ |
| Saturation | 0 | $(2/3)C_{ox}WL_{eff}$ | 0 |



## Average Distribution of Channel Capacitance

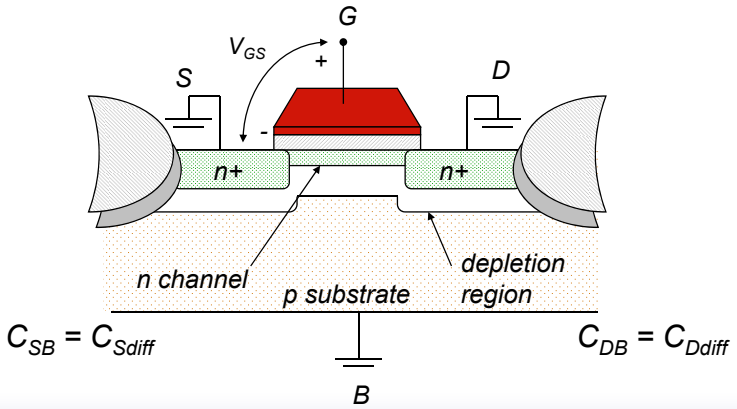
| Operation Region | $C_{GCB}$ | $C_{GCS}$ | $C_{GCD}$ | $C_{GC}$ | $C_G$ |
|---|---|---|---|---|---|
| Cutoff | $C_{ox}WL$ | 0 | 0 | $C_{ox}WL$ | $C_{ox}WL + 2C_oW$ |
| Resistive | 0 | $C_{ox}WL/2$ | $C_{ox}WL/2$ | $C_{ox}WL$ | $C_{ox}WL + 2C_oW$ |
| Saturation | 0 | $(2/3)C_{ox}WL$ | 0 | $(2/3)C_{ox}WL$ | $(2/3)C_{ox}WL + 2C_oW$ |

- *Channel capacitance components are nonlinear and vary with operating voltage*
- *Most important regions are cutoff and saturation since that is where the device spends most of its time*



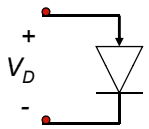
## MOS Diffusion Capacitances

- *The junction (or diffusion) capacitance is from the reverse-biased source-body and drain-body pn-junctions.*



# Review: Reverse Bias Diode

- All diodes in MOS digital circuits are reverse biased; the dynamic response of the diode is determined by depletion-region charge or junction capacitance




$$C_j = C_{j0}/((1 - V_D)/\phi_0)^m$$

where $C_{j0}$ is the capacitance under zero-bias conditions (a function of physical parameters), $\phi_0$ is the built-in potential (a function of physical parameters and temperature)

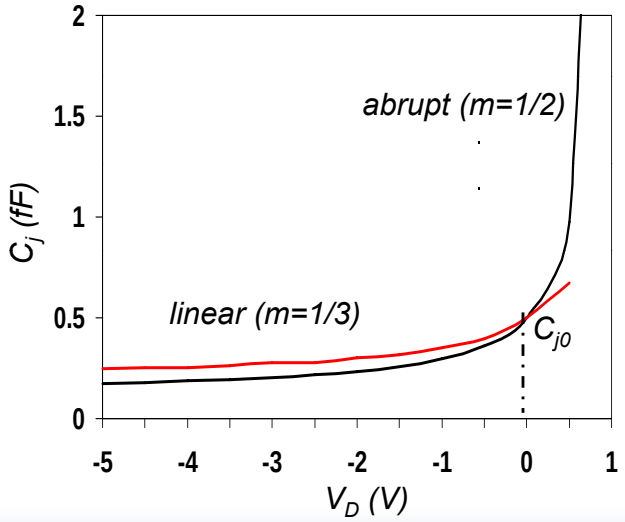
and m is the grading coefficient

- m = ½ for an abrupt junction (transition from n to p-material is instantaneous)
- m = 1/3 for a linear (or graded) junction (transition is gradual)

- Nonlinear dependence (that decreases with increasing reverse bias)



# Reverse-Bias Diode Junction Capacitance






# Linearizing the Junction Capacitance

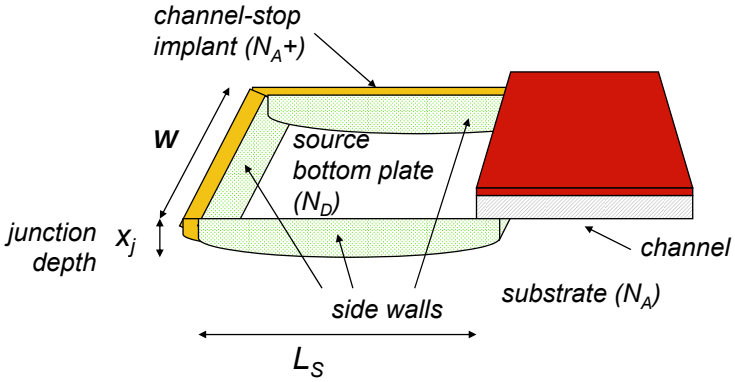
Replace non-linear capacitance by
large-signal equivalent linear capacitance
which displaces equal charge
over voltage swing of interest

$$C_{eq} = \frac{\Delta Q_j}{\Delta V_D} = \frac{Q_j(V_{high}) - Q_j(V_{low})}{V_{high} \quad V_{low}} = K_{eq} C_{j0}$$

$$K_{eq} = \frac{-\phi_0^m}{(V_{high} - V_{low})(1-m)}[(\phi_0 - V_{high})^{1-m} - (\phi_0 - V_{low})^{1-m}]$$



# Source Junction View




$$C_{diff} = C_{bp} + C_{sw} = C_j AREA + C_{jsw} PERIMETER$$

$$= C_j L_S W + C_{jsw}(2L_S + W)$$

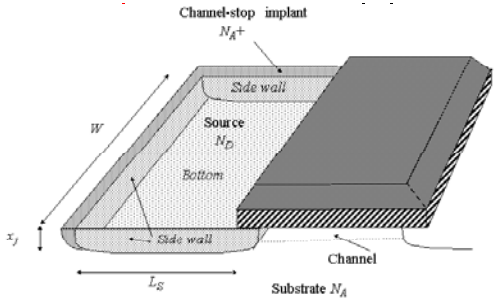
## MOS Capacitances: Junction/Diffusion Capacitances

- MOS junction/diffusion capacitances ($C_{Sdiff}$, $C_{Ddiff}$): originate from depletion regions of reversed-biased pn-junctions of drain and source
  - $C_{Sdiff}$: for source-bulk pn-junction
  - $C_{Ddiff}$: for drain-bulk pn-junction
- Let's define
  - $C_j$: junction capacitance per unit area
  - $C_{jsw}$': sidewall junction capacitance per unit area
  - $x_j$: junction depth.
  - $W$: channel width
  - $L_S$: source length
  - $L_D$: drain length



## MOS Capacitances: Junction/Diffusion Capacitances

- For $C_{Sdiff}$:
  1). Bottom plate junction: $C_{S\_bottom}=C_jWL_S$
  2). Side-wall junctions (3 sidewalls): $C_{S\_sw}=C_{jsw}' x_j(W+2L_S)$
  ➡ Total $C_{Sdiff}=C_{S\_bottom}+C_{S\_sw}=C_jWL_S+C_{jsw}(W+2L_S)$
- Total $C_{Ddiff}=C_jL_DW+C_{jsw}(W+2L_D)$

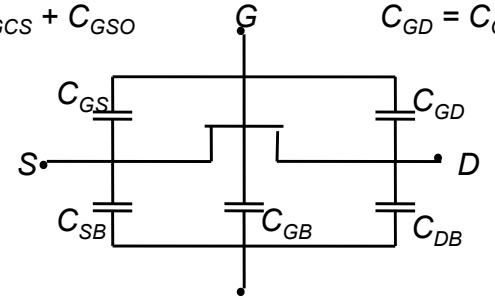



$$C_{diff} = C_{bottom} + C_{sw} = C_j \times AREA + C_{jsw} \times PERIMETER$$
$$= C_jL_SW + C_{jsw}(2L_S + W)$$



## MOS Capacitance Model

$C_{GS} = C_{GCS} + C_{GSO}$ $\quad$ $C_{GD} = C_{GCD} + C_{GDO}$




$C_{SB} = C_{Sdiff}$ $\quad$ $C_{DB} = C_{Ddiff}$

$C_{GB} = C_{GCB}$



## Transistor Capacitance Values for 0.25$\mu$

Example: For an NMOS with $L = 0.24\ \mu m$, $W = 0.36\ \mu m$, $L_D = L_S = 0.625\ \mu m$

$C_{GSO} = C_{GDO} = C_{ox}\, x_d\, W = C_o\, W = 0.11\ fF$

$C_{GC} = C_{ox}\, WL = 0.52\ fF$

so $C_{gate\_cap} = C_{ox}WL + 2C_oW = 0.74\ fF$

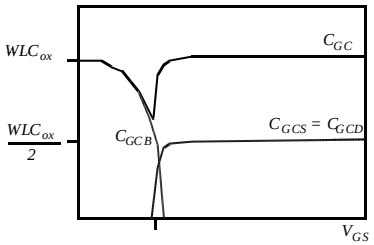
$C_{bp} = C_j\, L_S\, W = 0.45\ fF$

$C_{sw} = C_{jsw}\,(2L_S + W) = 0.45\ fF$

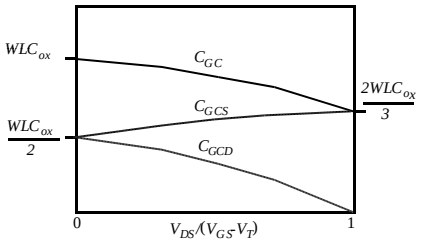
so $C_{diffusion\_cap} = 0.90\ fF$

| | $C_{ox}$ (fF/$\mu m^2$) | $C_o$ (fF/$\mu m$) | $C_j$ (fF/$\mu m^2$) | $m_j$ | $\phi_b$ (V) | $C_{jsw}$ (fF/$\mu m$) | $m_{jsw}$ | $\phi_{bsw}$ (V) |
|---|---|---|---|---|---|---|---|---|
| NMOS | 6 | 0.31 | 2 | 0.5 | 0.9 | 0.28 | 0.44 | 0.9 |
| PMOS | 6 | 0.27 | 1.9 | 0.48 | 0.9 | 0.22 | 0.32 | 0.9 |

# Gate Capacitance



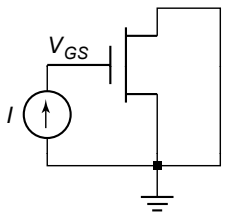

Capacitance as a function of VGS (with VDS = 0)

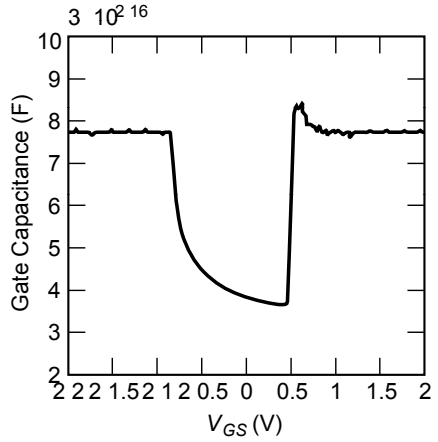



Capacitance as a function of the degree of saturation



# Measuring the Gate Cap

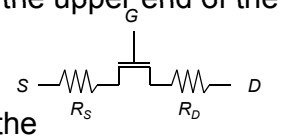





# Other (Submicon) MOS Transistor Concerns

- Velocity saturation
- Subthreshold conduction
  - Transistor is already partially conducting for voltages below $V_T$
- Threshold variations
  - In long-channel devices, the threshold is a function of the length (for low $V_{DS}$)
  - In short-channel devices, there is a drain-induced threshold barrier lowering at the upper end of the $V_{DS}$ range (for low L)
- Parasitic resistances
  - resistances associated with the source and drain contacts
- Latch-up



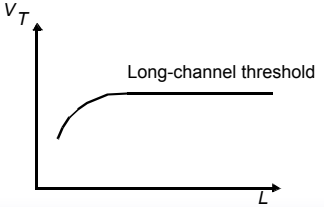



# Threshold Variations

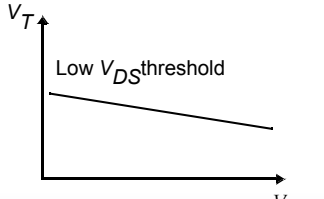
- For short-channel transistors:

✓ The depletion regions of source and reverse-biased drain junction, previously was ignored in long-channel device. But for short-channel device, they become relatively more important with shrinking channel length. Since a part of the region below gate is already depleted, a smaller VT suffices to cause strong inversion. Thus, VTO decreased with L for short-channel device.

✓ Raising drain-source (bulk) voltage also increases the width of drain-junction depletion region, hence VTO decreases with increasing VDS: drain-induced barrier lowering (DIBL).

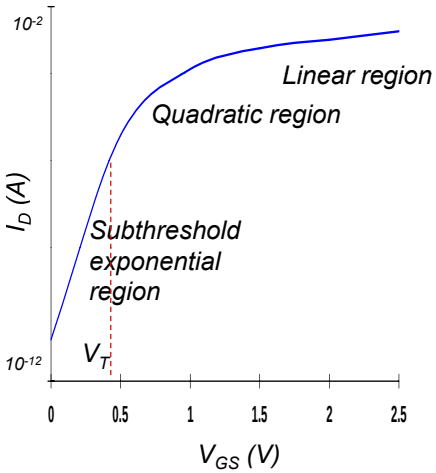



Threshold as a function of the length (for low $V_{DS}$)




Drain-induced barrier lowering (for low $L$)

# Subthreshold Conductance




$I_D \sim I_S\, e^{(qV_{GS}/nkT)}$ where $n \geq 1$

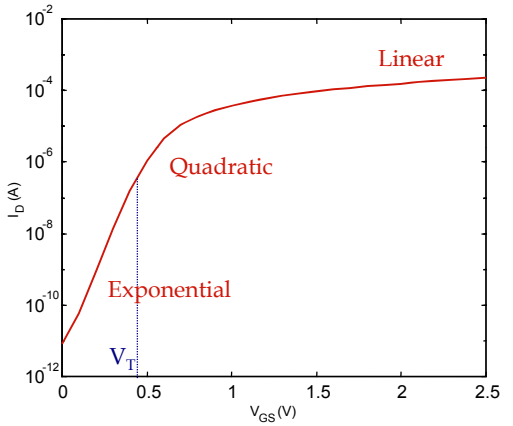
- Transition from ON to OFF is gradual (decays exponentially)
- Current roll-off (slope factor) is also affected by increase in temperature

$S = n\,(kT/q)\,\ln(10)$
*(typical values 60 to 100 mV/decade)*

- Has repercussions in dynamic circuits and for power consumption



# Sub-Threshold Conduction




**The Slope Factor**

$$I_D \sim I_0 e^{\frac{qV_{GS}}{nkT}}, \quad n = 1 + \frac{C_D}{C_{ox}}$$

$S$ is $\Delta V_{GS}$ for $I_{D2}/I_{D1}$ =10

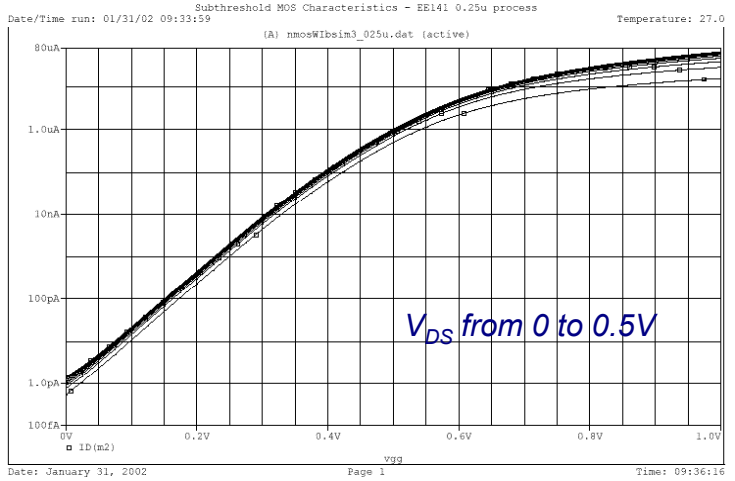
$$S = n\left(\frac{kT}{q}\right)\ln(10)$$

Typical values for S:
60 .. 100 mV/decade



# Subthreshold $I_D$ vs $V_{GS}$

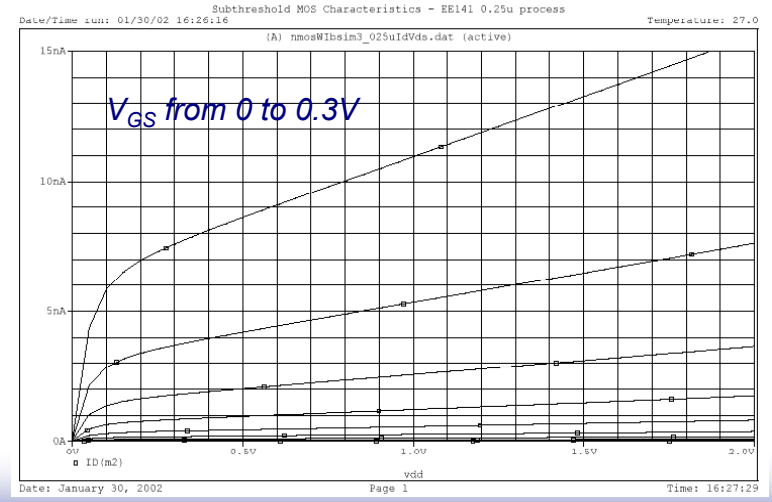
$I_D = I_S\, e^{(qV_{GS}/nkT)}\ (1 - e^{-(qV_{DS}/kT)})(1 + \lambda V_{DS})$






# Subthreshold $I_D$ vs $V_{DS}$

$I_D = I_S\, e^{(qV_{GS}/nkT)}\ (1 - e^{-(qV_{DS}/kT)})(1 + \lambda V_{DS})$

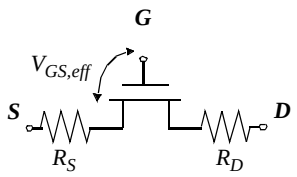


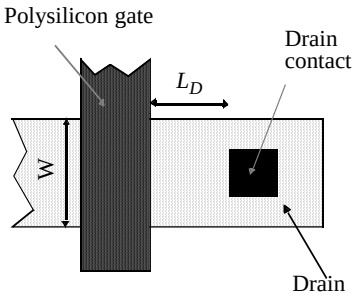
# Summary of MOSFET Operating Regions

- Strong Inversion $V_{GS} > V_T$
  - Linear (Resistive) $V_{DS} < V_{DSAT}$
  - Saturated (Constant Current) $V_{DS} \geq V_{DSAT}$
- Weak Inversion (Sub-Threshold) $V_{GS} \leq V_T$
  - Exponential in $V_{GS}$ with linear $V_{DS}$ dependence



# Parasitic Resistances

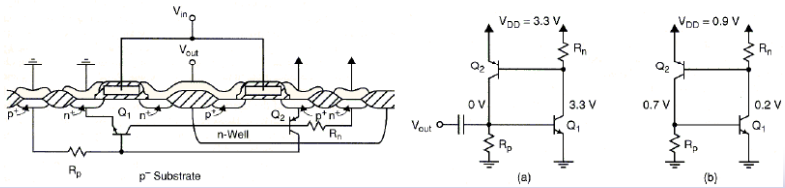





# Latch-Up

- Latch-up is a very real, very important factor in circuit design that must be accounted for
- Due to (relatively) large current in substrate or n-well
  - create voltage drops across the resistive substrate/well
    - most common during large power/ground current spikes
  - turns on parasitic BJT devices, effectively shorting power & ground
    - often results in device failure with fused-open wire bonds or interconnects
  - hot carrier effects can also result in latch-up
    - latch-up very important for short channel devices
- Avoid latch-up by
  - including as many substrate/well contacts as possible
    - rule of thumb: one "plug" each time a tx connects to the power rail
  - limiting the maximum supply current on the chip






# SPICE MODELS

**Level 1: Long Channel Equations - Very Simple**

**Level 2: Physical Model - Includes Velocity Saturation and Threshold Variations**

**Level 3: Semi-Emperical - Based on curve fitting to measured devices**

**Level 4 (BSIM): Emperical - Simple and Popular**

# MAIN MOS SPICE PARAMETERS

| Parameter Name | Symbol | SPICE Name | Units | Default Value |
|---|---|---|---|---|
| SPICE Model Index | | LEVEL | - | 1 |
| Zero-Bias Threshold Voltage | VT0 | VT0 | V | 0 |
| Process Transconductance | k' | KP | A/V2 | 2.E-5 |
| Body-Bias Parameter | g | GAMMA | V0.5 | 0 |
| Channel Modulation | l | LAMBDA | 1/V | 0 |
| Oxide Thickness | tox | TOX | m | 1.0E-7 |
| Lateral Diffusion | xd | LD | m | 0 |
| Metallurgical Junction Depth | xj | XJ | m | 0 |
| Surface Inversion Potential | 2\|fF\| | PHI | V | 0.6 |
| Substrate Doping | NA,ND | NSUB | cm-3 | 0 |
| Surface State Density | Qss/q | NSS | cm-3 | 0 |
| Fast Surface State Density | | NFS | cm-3 | 0 |
| Total Channel Charge Coefficient | | NEFF | - | 1 |
| Type of Gate Material | | TPG | - | 1 |
| Surface Mobility | m0 | U0 | cm2/V-sec | 600 |
| Maximum Drift Velocity | umax | VMAX | m/s | 0 |
| Mobility Critical Field | xcrit | UCRIT | V/cm | 1.0E4 |
| Critical Field Exponent in Mobility Degradation | | UEXP | - | 0 |
| Transverse Field Exponent (mobility) | | UTRA | - | 0 |



# SPICE Parameters for Parasitics

| Parameter Name | Symbol | SPICE Name | Units | Default Value |
|---|---|---|---|---|
| Source resistance | $R_S$ | RS | Ω | 0 |
| Drain resistance | $R_D$ | RD | Ω | 0 |
| Sheet resistance (Source/Drain) | $R_\square$ | RSH | $\Omega/_\square$ | 0 |
| Zero Bias Bulk Junction Cap | $C_{j0}$ | CJ | $F/m^2$ | 0 |
| Bulk Junction Grading Coeff. | $m$ | MJ | - | 0.5 |
| Zero Bias Side Wall Junction Cap | $C_{jsw0}$ | CJSW | F/m | 0 |
| Side Wall Grading Coeff. | $m_{sw}$ | MJSW | - | 0.3 |
| Gate-Bulk Overlap Capacitance | $C_{gbO}$ | CGBO | F/m | 0 |
| Gate-Source Overlap Capacitance | $C_{gsO}$ | CGSO | F/m | 0 |
| Gate-Drain Overlap Capacitance | $C_{gdO}$ | CGDO | F/m | 0 |
| Bulk Junction Leakage Current | $I_S$ | IS | A | 0 |
| Bulk Junction Leakage Current Density | $J_S$ | JS | $A/m^2$ | 1E-8 |
| Bulk Junction Potential | $\phi_0$ | PB | V | 0.8 |



# SPICE Transistors Parameters

| Parameter Name | Symbol | SPICE Name | Units | Default Value |
|---|---|---|---|---|
| Drawn Length | L | L | m | - |
| Effective Width | W | W | m | - |
| Source Area | AREA | AS | m2 | 0 |
| Drain Area | AREA | AD | m2 | 0 |
| Source Perimeter | PERIM | PS | m | 0 |
| Drain Perimeter | PERIM | PD | m | 0 |
| Squares of Source Diffusion | | NRS | - | 1 |
| Squares of Drain Diffusion | | NRD | - | 1 |