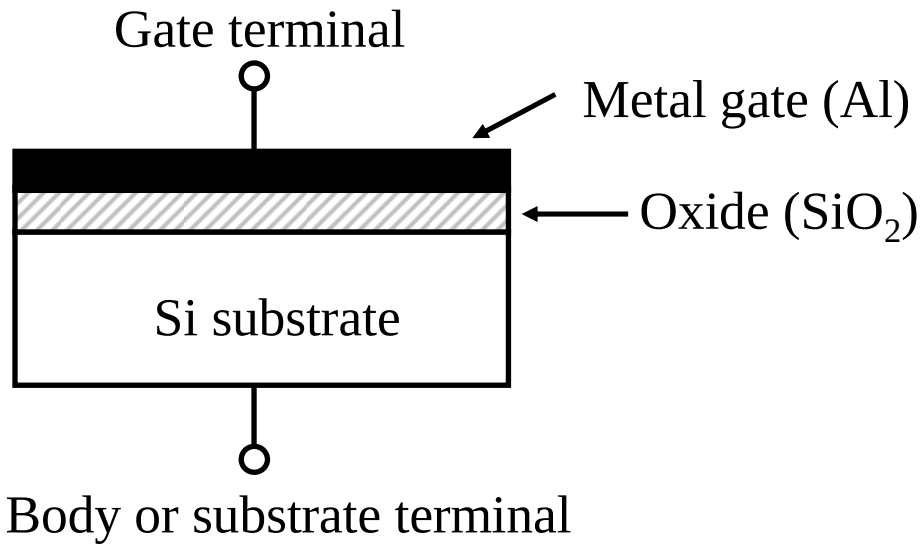
ECE 225
High-Speed Digital IC Design
Lecture 9

CMOS Devices: MOS Capacitor, Basic MOSFET Operation, Long-channel vs Short-channel Devices, Device Capacitances

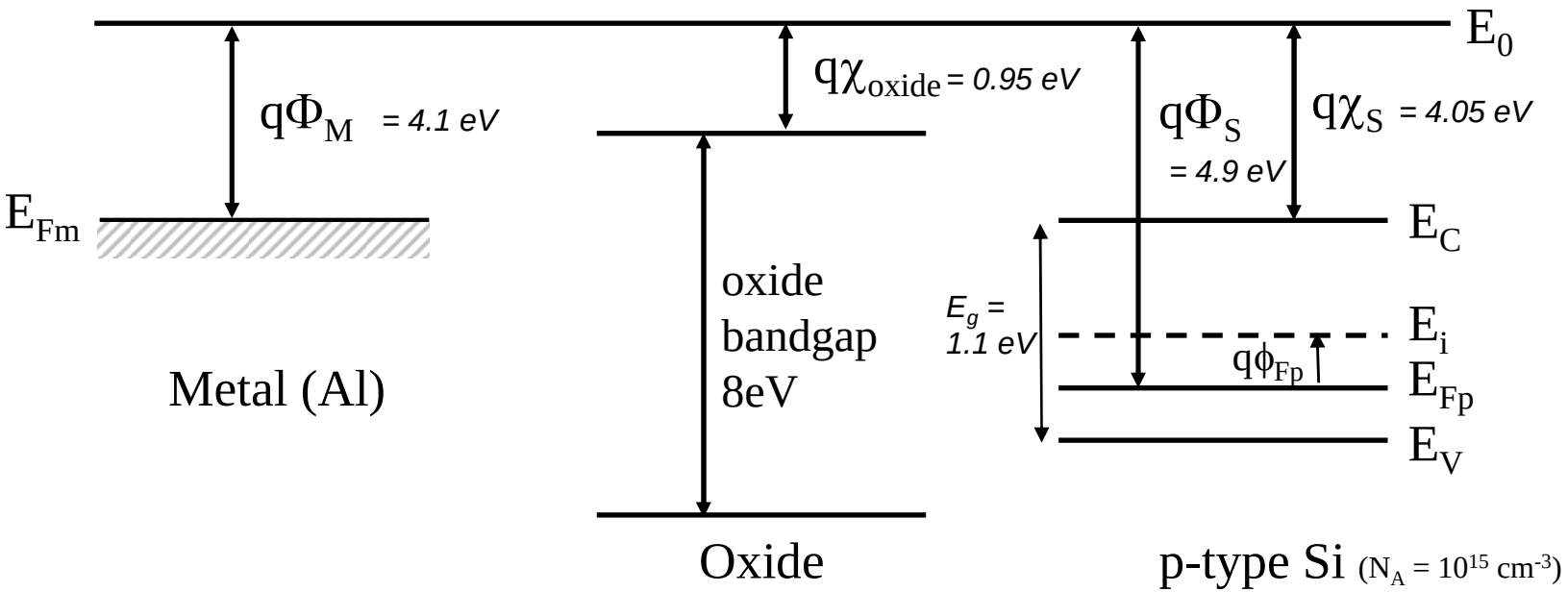
Prof. Kaustav Banerjee
Electrical and Computer Engineering
E-mail: kaustav@ece.ucsb.edu

MOS Structure

- MOS: Metal-oxide-semiconductor
  - Gate: metal (or polysilicon)
  - Oxide: silicon dioxide, grown on substrate
- MOS capacitor: two-terminal MOS structure



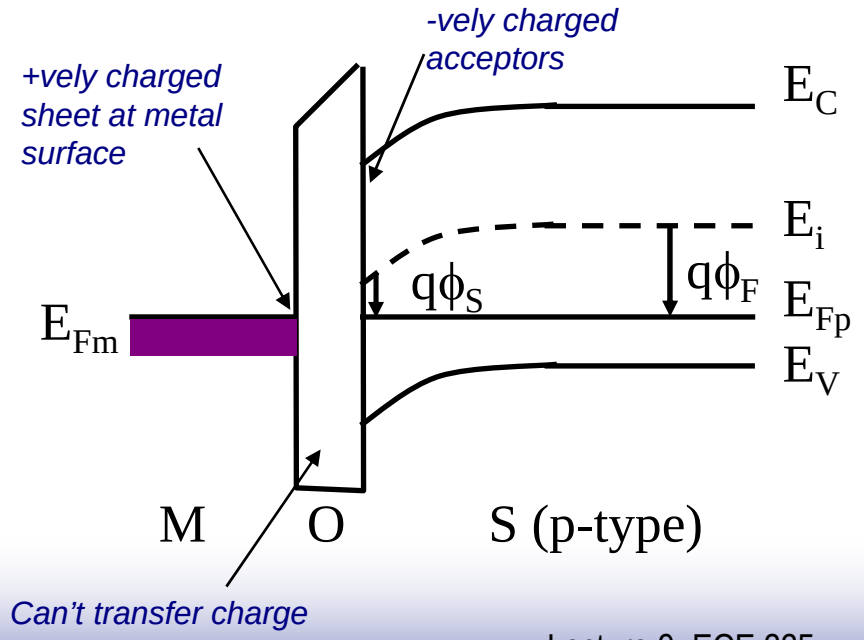
MOS Energy Band Diagram



- Work function ($q\Phi_M$, $q\Phi_S$): energy required to take electron from Fermi level to free space
- Electron affinity is the potential difference between the conduction band level and vacuum (free-space) level = qX_s
- Work function difference between Al and Si = 0.8V
- At equilibrium, Fermi levels must line up

MOS Energy Band Diagram

- Bands must bend for Fermi levels to line up
- Amount of bending is equal to work function difference: $q\Phi_M - q\Phi_S$
- Fermi levels equalized by transfer of –ve charge from materials with higher E_F (smaller work functions) across interfaces to materials with lower E_F
- Part of voltage drop occurs across oxide, rest occurs next to O-S interface



ϕ_F = Fermi potential (difference between E_F and E_i in bulk)

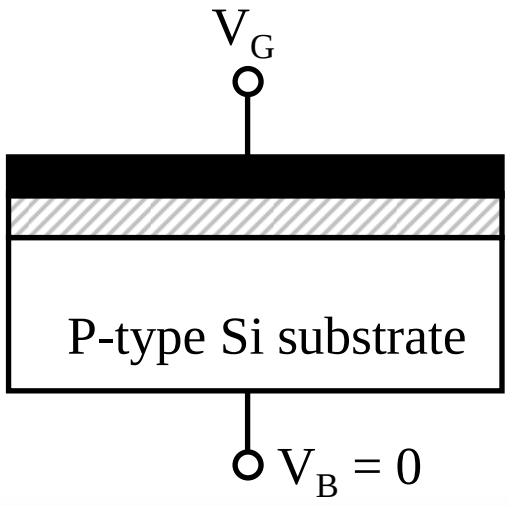
ϕ_S = surface potential

Flat-Band Voltage

- Flat-band voltage
  - Built-in potential of MOS system
  - Work function difference: $V_{FB} = \Phi_m - \Phi_S$
  - Apply this voltage to “flatten” energy bands
  - For the MOS system considered on the previous slide, a –ve voltage applied to the metal w.r.t the Si opposes the built-in voltage on the capacitor and tends to reduce the charge stored on the capacitor plates below its equilibrium value
- Example
  - P-type substrate: $q\phi_{Fp} = 0.2$ eV, $q\chi_S = 4.05$ eV
  - Aluminum gate ($q\Phi_m = 4.1$ eV)
  - What is flatband voltage?

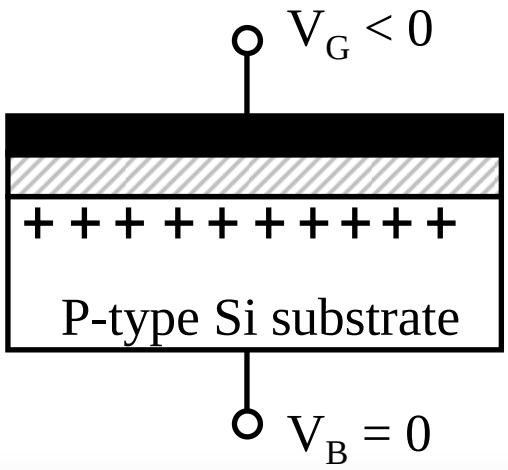
MOS Capacitor Operation

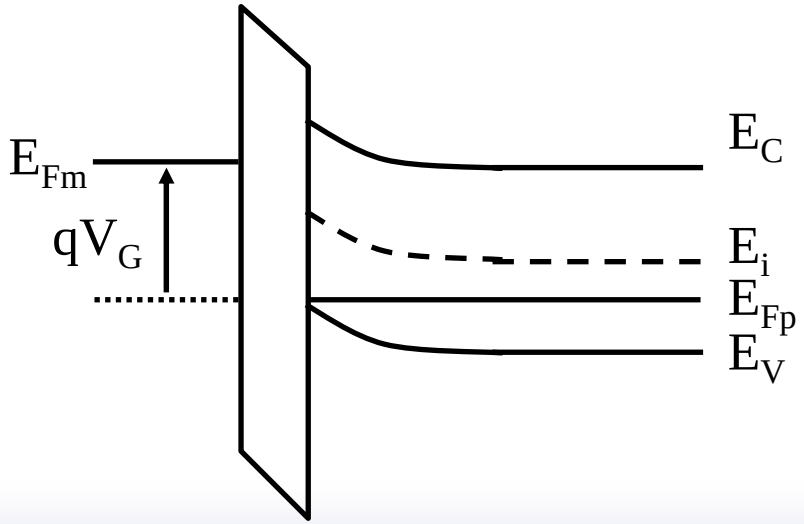
- Assume p-type substrate
- Three regions of operation
  - Accumulation ($V_G < 0$)
  - Depletion ($V_G > 0$ but small)
  - Inversion ($V_G >> 0$)



Accumulation

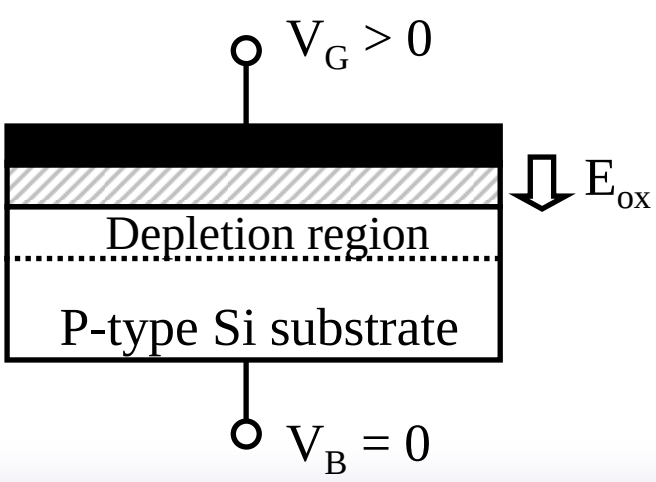
- Negative voltage on gate: attracts holes in substrate towards oxide
- Holes “accumulate” on Si surface (surface is more strongly p-type)
- Electrons pushed deeper into substrate

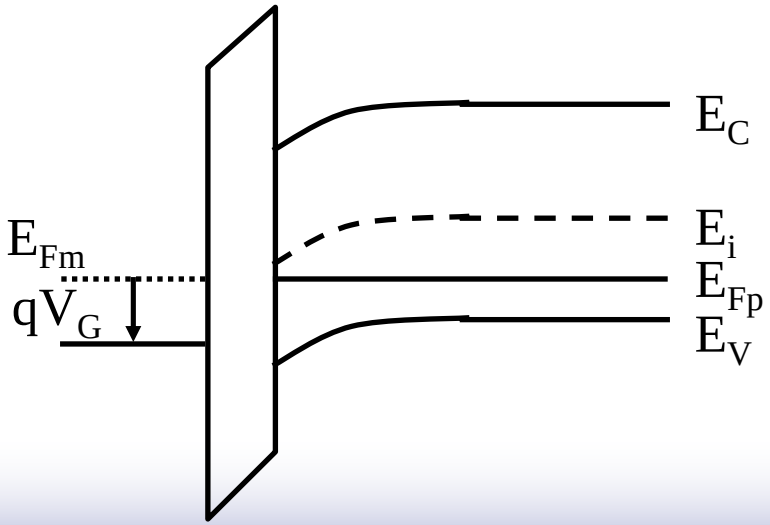




Depletion

- Positive voltage on gate: repels holes in substrate
  - Holes leave behind negatively charged acceptor ions
- Depletion region forms: devoid of carriers
  - Electric field directed from gate to substrate
- Bands bend downwards near surface
  - Surface becomes <u>less</u> strongly p-type (E_F close to E_i)

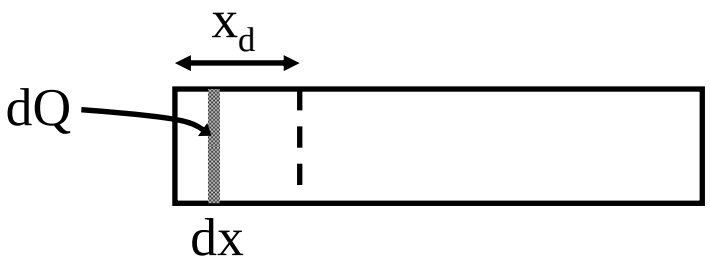




Depletion Region Depth

- Calculate thickness x_d of depletion region
  - Find charge dQ in small slice of depletion area

$$dQ = -qN_A dx$$



  - Find change in surface potential to displace dQ by distance x_d from the surface (Poisson equation):

$$d\phi = -x\frac{dQ}{\varepsilon_{Si}} \qquad d\phi = xqN_A\frac{dx}{\varepsilon_{Si}}$$

Depletion Region Depth (cont.)

- Integrate perpendicular to surface

$$\int_{\phi_F}^{\phi_S} d\phi_S = \int_0^{x_d} \frac{qN_A x}{2\varepsilon_{Si}} dx$$

$$\phi_S - \phi_F = \frac{qN_A x_d^2}{2\varepsilon_{Si}}$$

- Result:

$$x_d = \sqrt{\frac{2\varepsilon_{Si}\left|\phi_S - \phi_F\right|}{qN_A}}$$

Depletion Region Charge

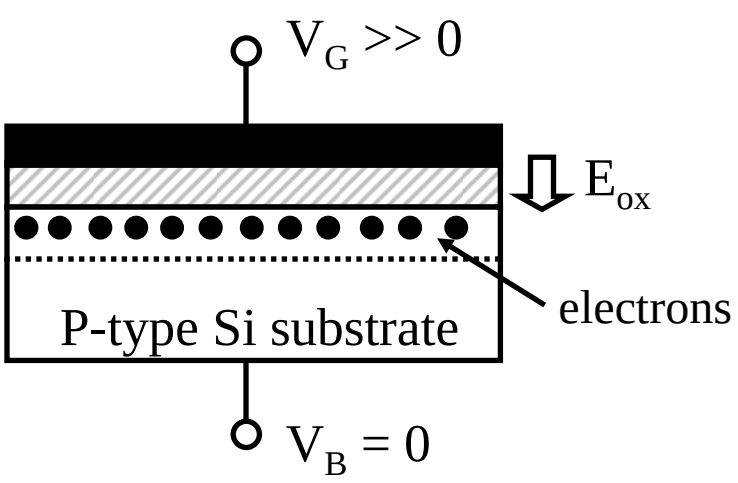
- Depletion region charge density
  - Due only to fixed acceptor ions
  - Charge per unit area

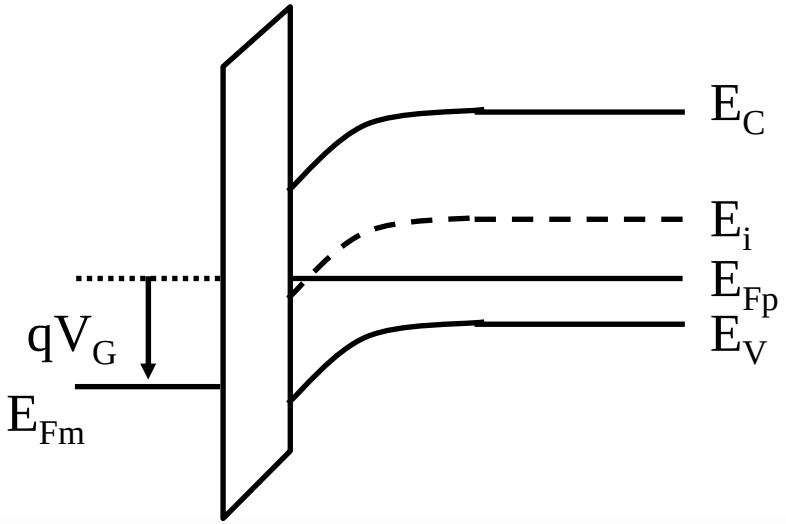
$$Q = -qN_A x_d$$

$$Q = -\sqrt{2qN_A \varepsilon_{Si} \left| \phi_S - \phi_F \right|}$$

Inversion

- Increase voltage on gate, bands bend more
- Additional minority carriers (electrons) attracted from substrate to surface
  - Forms “inversion layer” of electrons
- Surface becomes n-type





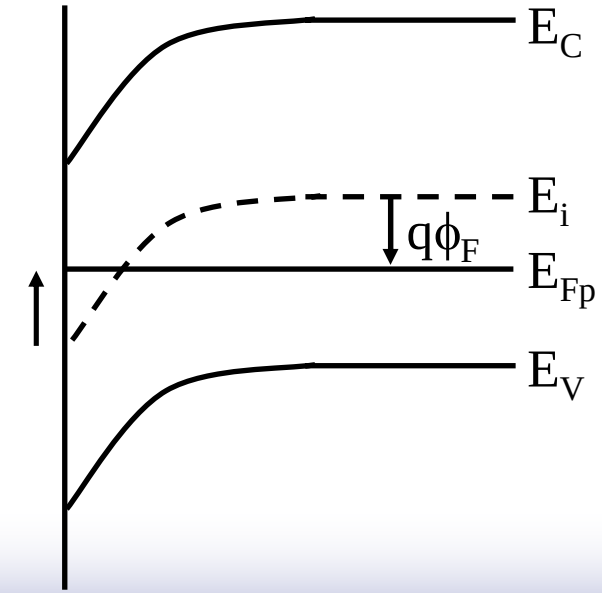
Inversion

- Definition of inversion
  - Point at which density of electrons on surface = density of holes in bulk
  - Surface potential is same as ϕ_F, but different sign

Remember:

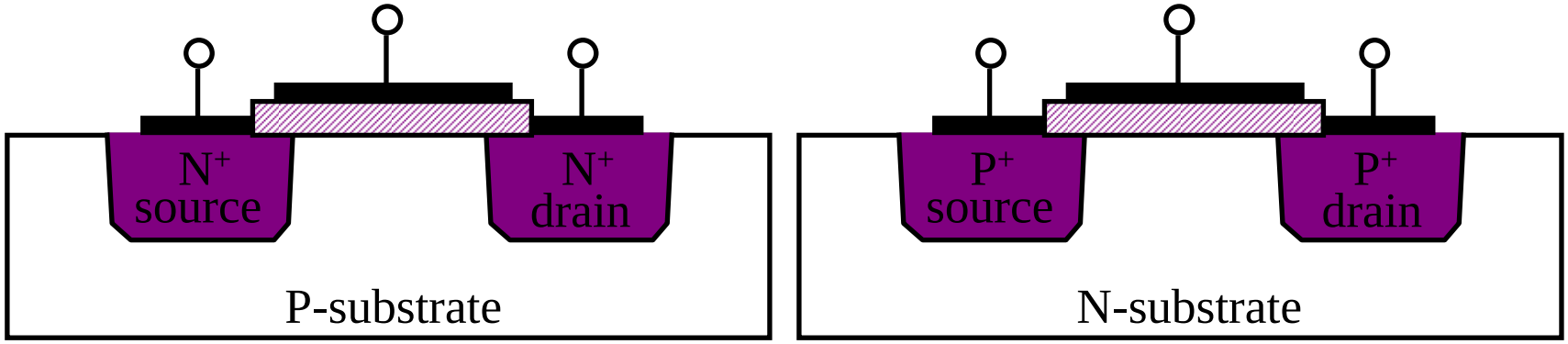
$q\phi_F = E_F - E_i$

$q\phi_S = -q\phi_F$



MOS Transistor

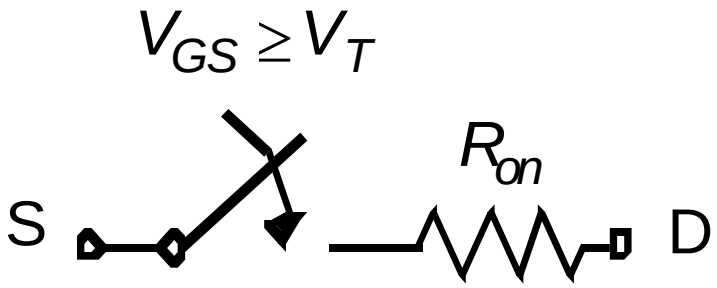
- Add “source” and “drain” terminals to MOS capacitor
- Transistor types
  - NMOS: p-type substrate, n^+ source/drain
  - PMOS: n-type substrate, p^+ source/drain

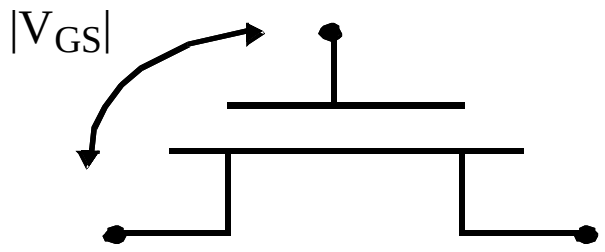


NMOS PMOS

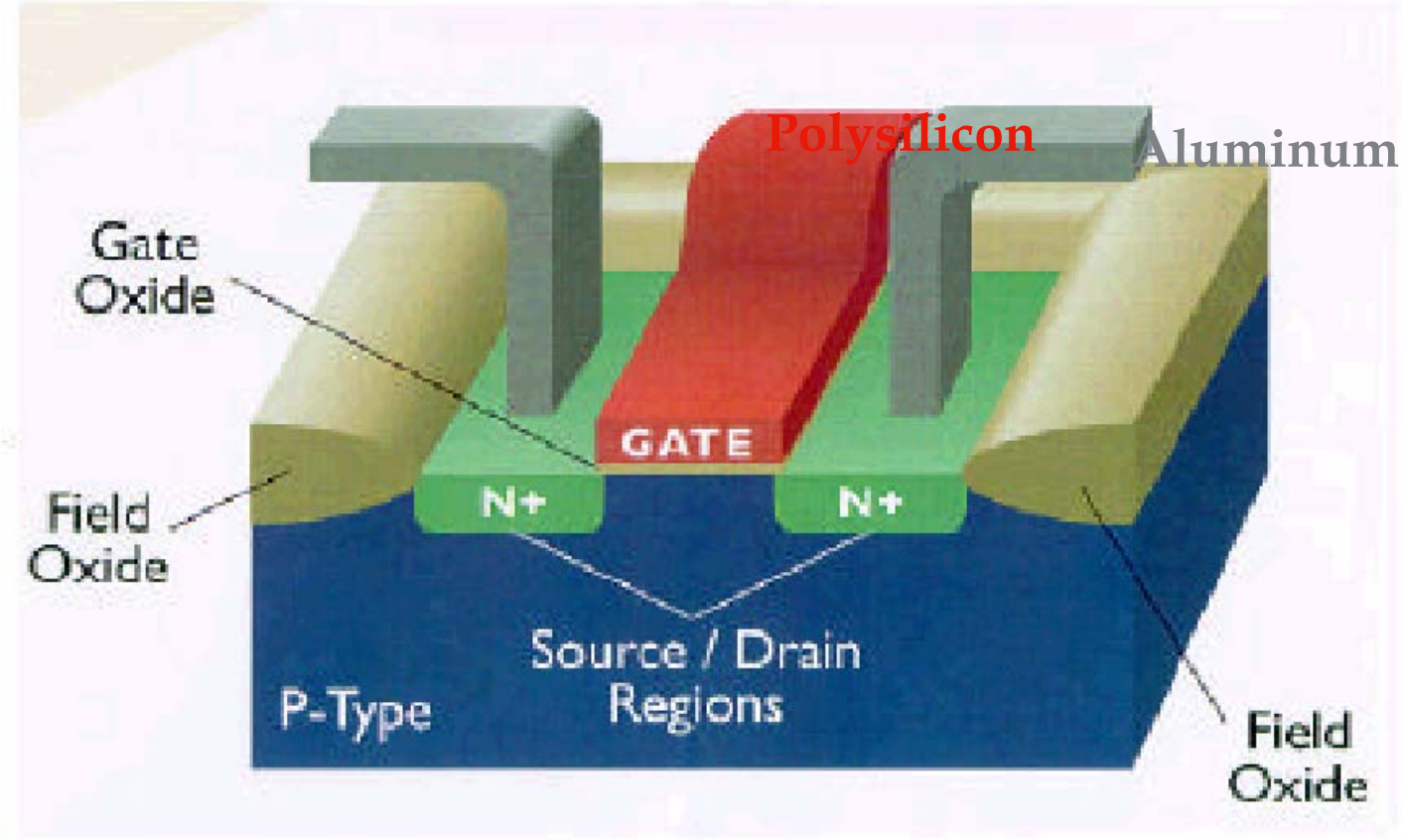
What is a Transistor?

A Switch! ⟷ An MOS Transistor



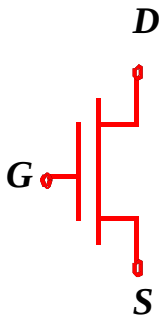


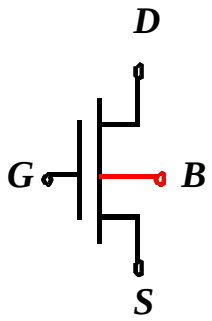
The MOS Transistor



MOS Transistors - Types and Symbols

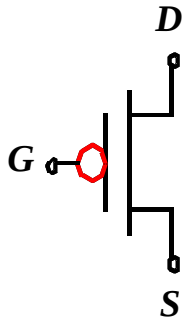
NMOS





NMOS with Body Contact

PMOS



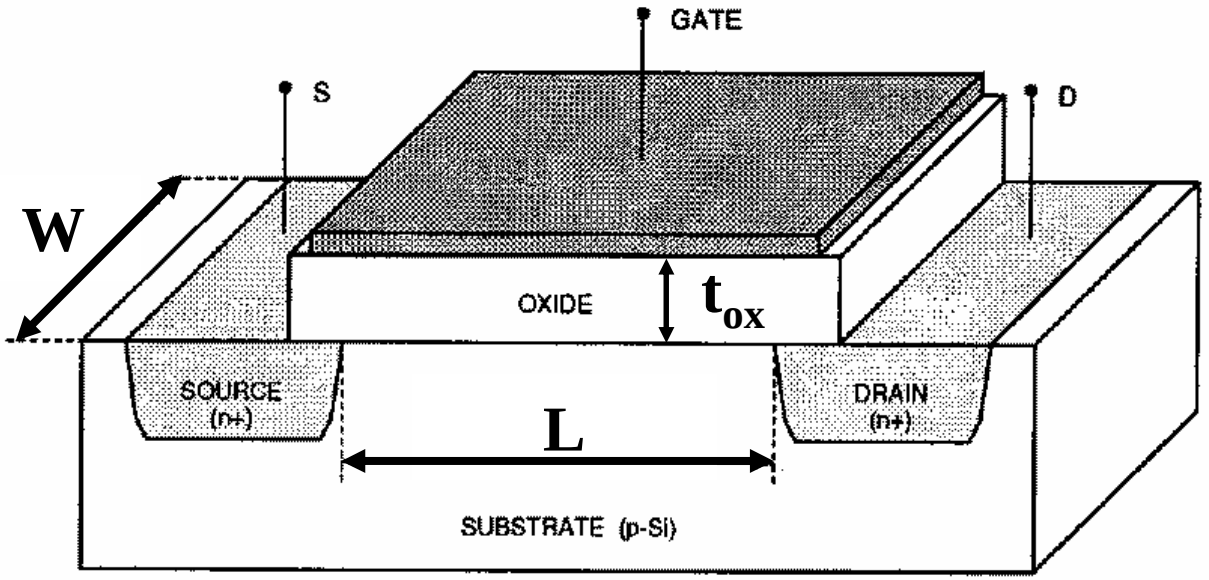
For NMOS: Body tied to Gnd

For PMOS: Body tied to Vdd

Why?

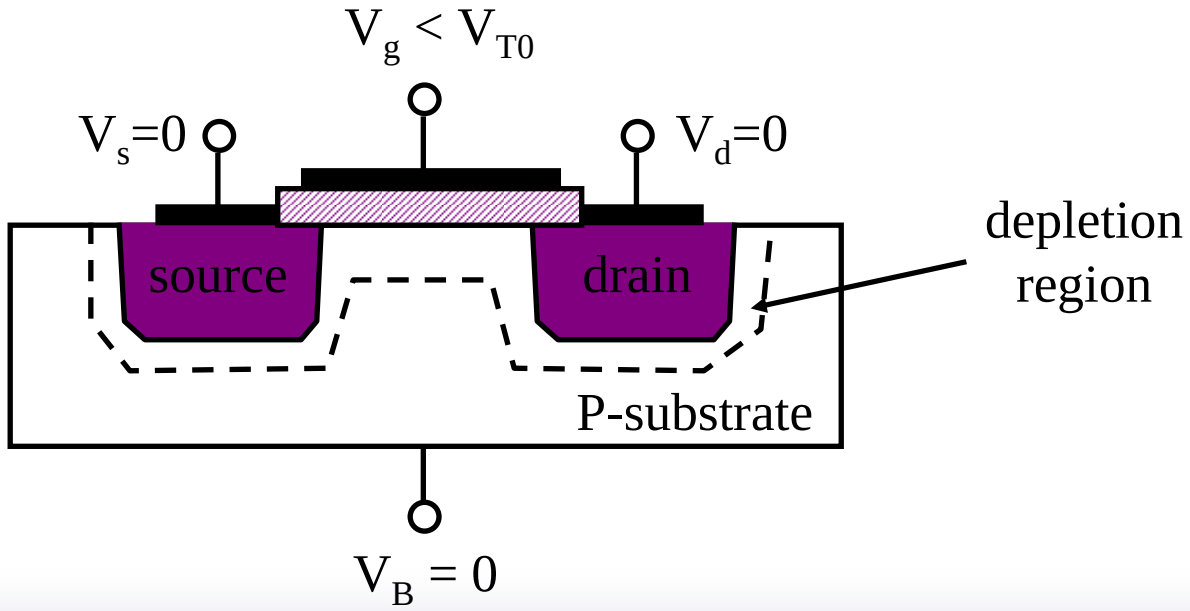
MOS Transistor

- Important transistor physical characteristics
  - Channel length L
  - Channel width W



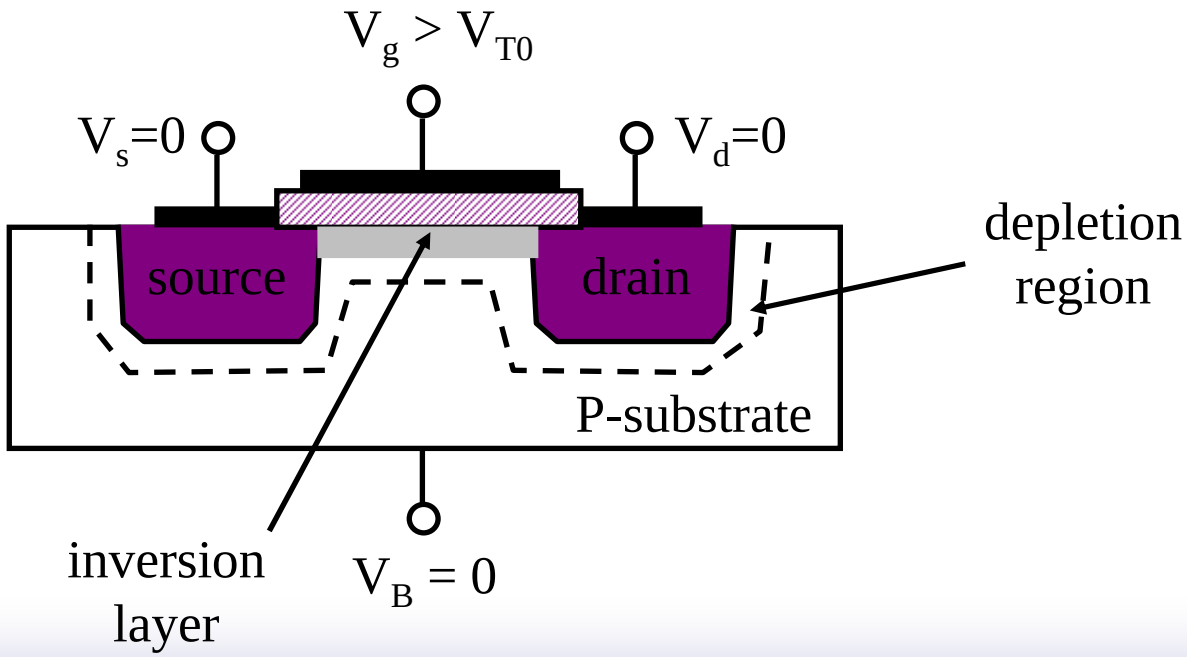
MOS Transistor Operation

- Simple case: $V_D = V_S = V_B = 0$
  - Operates as MOS capacitor
- When $V_{GS}<V_{T0}$ (but positive) , depletion region forms
  - No carriers in channel to connect S and D
- V_{T0} is known as the *threshold voltage*

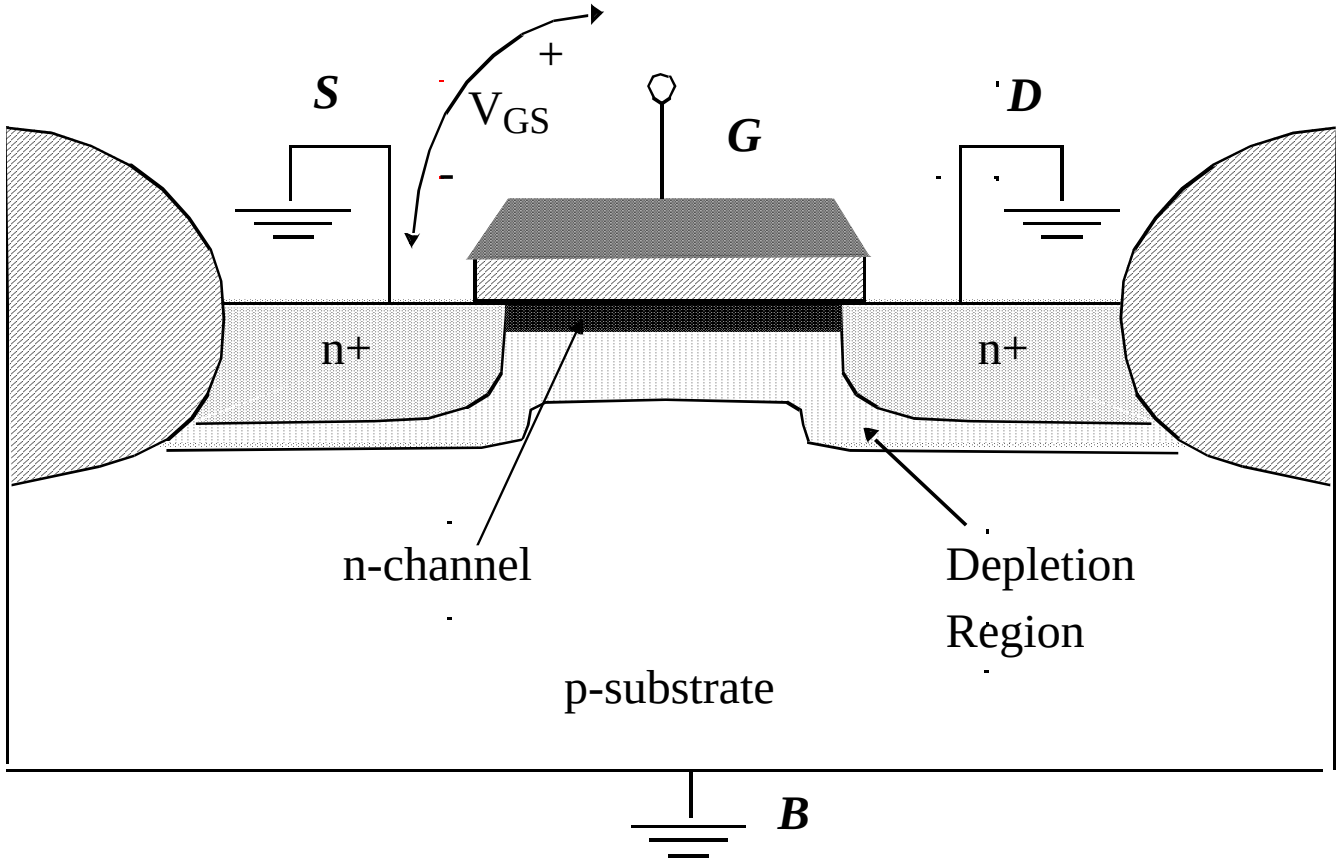


MOS Transistor Operation

- When $V_{GS} > V_{T0}$, inversion layer forms
- Source and drain connected by conducting n-type layer (for NMOS)



Threshold Voltage (V_{T0}): Concept



Physical Parameters that Affect V_{T0}

- Threshold voltage (V_{T0}): voltage between gate and source required for inversion
  - NMOS Transistor is “off” when $V_{GS} < V_{T0}$
- Components:
  - Work function difference between gate and channel (Flat-band voltage)
  - Gate voltage to change surface potential
  - Gate voltage to offset depletion region charge
  - Gate voltage to offset fixed charges in gate oxide and in silicon-oxide interface

Threshold voltage (1)

- Work function difference Φ_{GC} between gate and channel
  - Represents built-in potential of MOS system
  - For metal gate: $\Phi_{GC} = \Phi_F(\text{substrate}) - \Phi_M(\text{gate}) = \Phi_{ms}$
  - For poly gate: $\Phi_{GC} = \phi_F(\text{substrate}) - \phi_F(\text{gate})$

$$V_{T0} = \Phi_{GC} + \cdots$$

Threshold voltage (2)

- First component accounts for built-in voltage drop
- Now apply additional gate voltage to achieve inversion: change surface potential by $-2\phi_F$

$$V_{T0} = \Phi_{GC} - 2\phi_F + \cdots$$

Threshold voltage (3)

- Offset depletion region charge, due to fixed acceptor ions
- Calculate charge at inversion ($\phi_S = -\phi_F$)
  - From before: $Q = -\sqrt{2qN_A\varepsilon_{Si}\left|\phi_S - \phi_F\right|}$
  - So: $Q_{B0} = -\sqrt{2qN_A\varepsilon_{Si}\left|-2\phi_F\right|}$
  - For non-zero substrate bias ($V_{SB} \neq 0$):

$$Q_B = -\sqrt{2qN_A\varepsilon_{Si}\left|-2\phi_F + V_{SB}\right|}$$

Due to larger depletion region

Threshold voltage (3, cont.)

- To offset this charge, need voltage $-Q_B/C_{ox}$
- Cox = gate capacitance per unit area
  - $C_{ox}=\varepsilon_{ox}/t_{ox}$
  - t_{ox} = thickness of gate oxide (normally in Å)

$$V_{T0} = \Phi_{GC} - 2\phi_F - \frac{Q_B}{C_{ox}} + \cdots$$

Threshold voltage (4)

- Finally, correct for non-ideal fixed charges
  - Fixed positive charged ions at boundary between oxide and substrate. Density = N_{OX}
  - Due to impurities, lattice imperfections at interface
  - Positive charge density $Q_{ox} = qN_{ox}$
  - Correct with gate voltage = $-Q_{ox}/C_{ox}$

- Final threshold voltage formula (for NMOS):

$$V_{T0} = \Phi_{GC} - 2\phi_F - \frac{Q_{B0}}{C_{ox}} - \frac{Q_{ox}}{C_{ox}}$$

Threshold voltage, summary

- If $V_{SB} = 0$ (no substrate bias):

$$V_{T0} = \Phi_{GC} - 2\phi_F - \frac{Q_{B0}}{C_{ox}} - \frac{Q_{ox}}{C_{ox}}$$

- If $V_{SB} \neq 0$ (non-zero substrate bias)

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

- Body effect (substrate-bias) coefficient:

$$\gamma = \frac{\sqrt{2qN_A\varepsilon_{Si}}}{C_{ox}}$$

+ for NMOS
- for PMOS

- Threshold voltage increases as V_{SB} increases!

Threshold Voltage (NMOS vs. PMOS)

	NMOS	PMOS
Substrate Fermi potential	$\phi_F < 0$	$\phi_F > 0$
Depletion charge density	$Q_B < 0$	$Q_B > 0$
Substrate bias coefficient	$\gamma > 0$	$\gamma < 0$
Substrate bias voltage	$V_{SB} > 0$	$V_{SB} < 0$
Threshold voltage (enhancement devices)	$V_{T0} > 0$	$V_{T0} < 0$

Threshold voltage adjustment

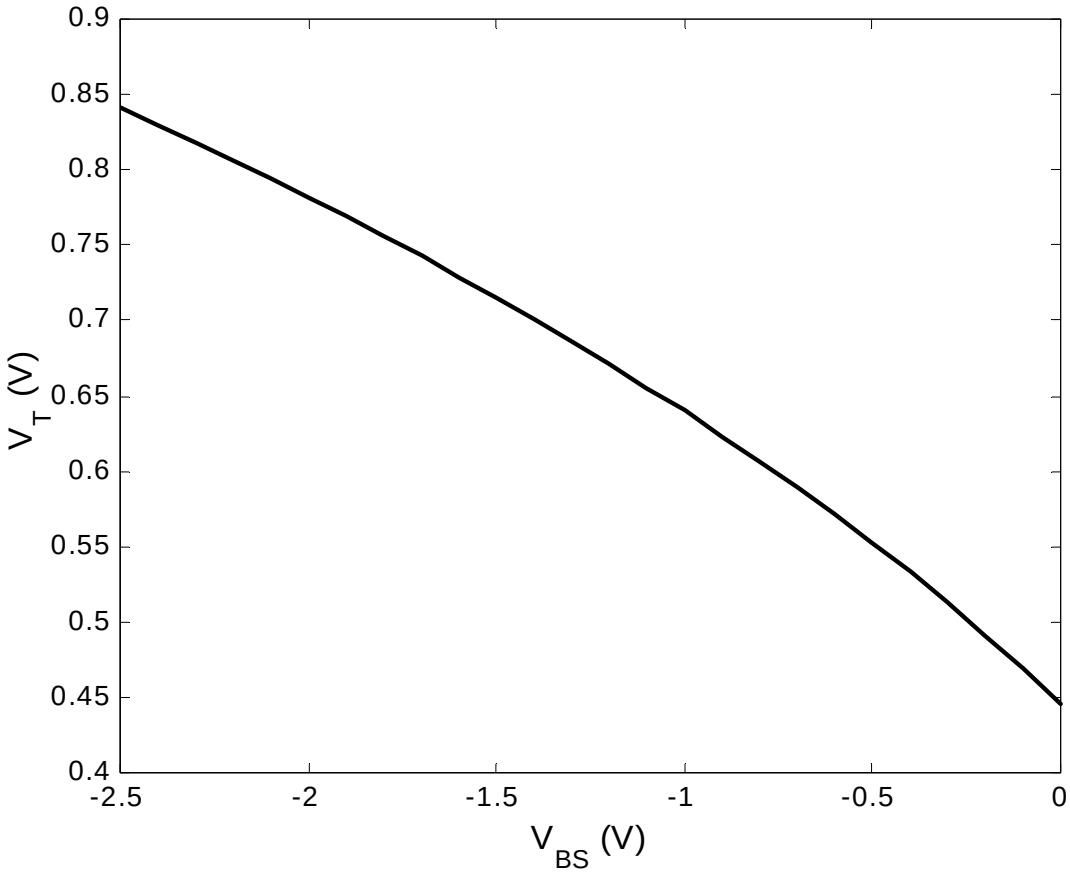
- Threshold voltage can be changed by doping the channel region with donor or acceptor ions
- For NMOS:
  - V_T increased by adding acceptor ions (p-type)
  - V_T decreased by adding donor ions (n-type)
  - Opposite for PMOS
- Approximate change in V_{T0}:
  - Density of implanted ions = N_I [cm^{-2}]
  - Assume all implanted impurities are ionized

$$\Delta V_{T0} = \frac{qN_I}{C_{ox}}$$

Example: V_{T0} Adjustment

- Consider an NMOS device:
  - P-type substrate: $N_A = 2 \times 10^{16}$ cm^{-3}
  - Polysilicon gate: $\Phi_{GC} = -0.92$V
  - $t_{ox} = 600$ Å (1Å = 1×10^{-8} cm)
  - $N_{ox} = 2 \times 10^{10}$ cm-2
  - $\varepsilon_{Si} = 11.7\ \varepsilon_0$, $\varepsilon_{ox} = 3.97\ \varepsilon_0$
- (a) Find V_{T0}
- (b) Find amount and type of channel implant to get $V_{T0} = 0.4$ V

The Body Effect



Transistor Currents (NMOS)

Cutoff Region: $I_{ds} = 0,\ V_{gs} < V_t$

$I_{ds} = Q_{channel}$ /carrier velocity(v)

Linear Region: $V_{gs} > V_t,\ V_{ds} < V_{gs} - V_t$

$I_{ds} = \mu C_{ox} W/L\ (V_{gs} - V_t - V_{ds}/2) V_{ds}$

$v = \mu E$

$E = V_{ds}/L$

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

Saturation Region: $V_{gs} > V_t,\ V_{ds} > V_{gs} - V_t$

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

Note: for PMOS $V_{tp} \neq V_{tn}$

$\mu_p < \mu_n$, *hence* $(W/L)_{PMOS} \sim 2\ (W/L)_{NMOS}$

NMOS Characteristics

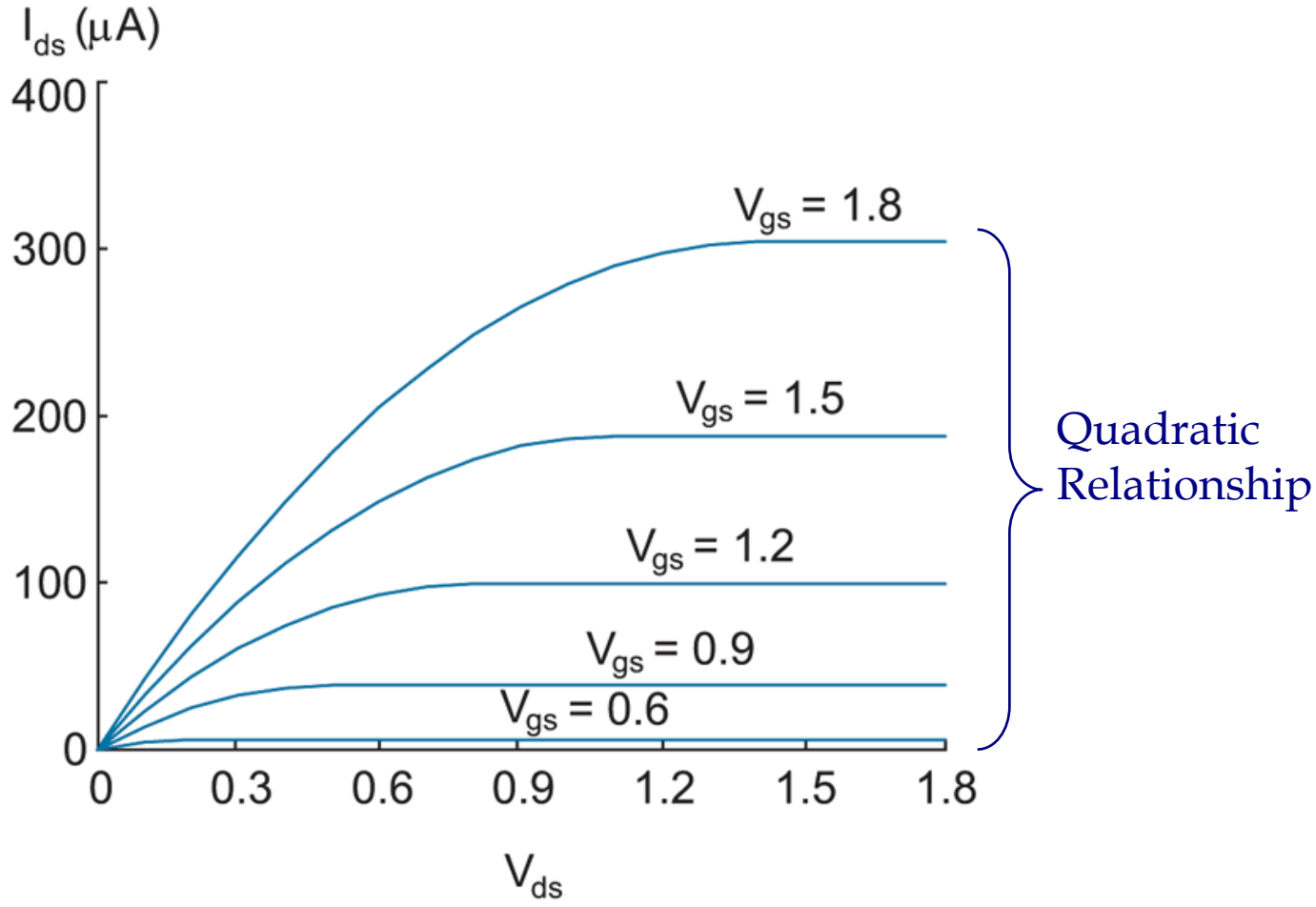


FIG 2.7 I-V characteristics of ideal nMOS transistor

PMOS Characteristics

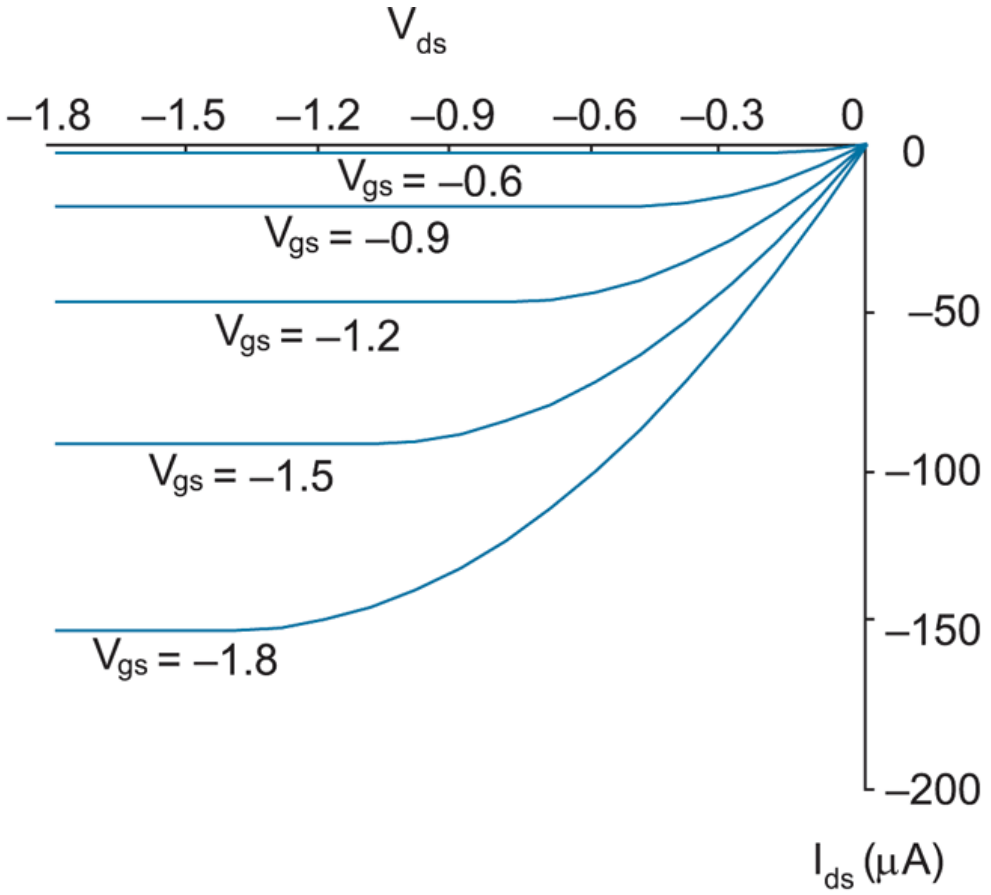


FIG 2.8 I-V characteristics of ideal pMOS transistor

Channel Length Modulation

- In saturation, pinch-off point moves
  - As V_{DS} is increased, pinch-off point moves closer to source
  - Effective channel length becomes shorter
  - Current increases due to shorter channel

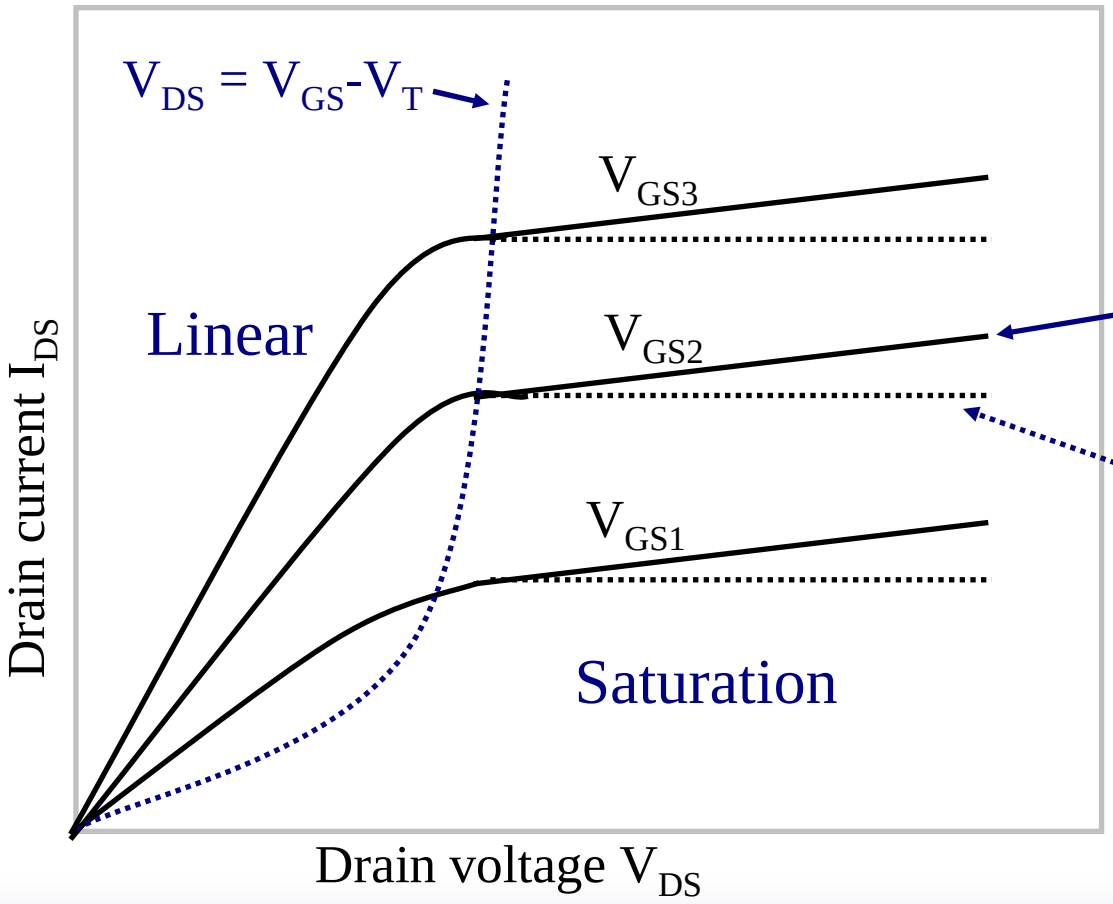
$$L' = L - \Delta L$$

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

λ = channel length modulation coefficient

Summary: MOS I/V

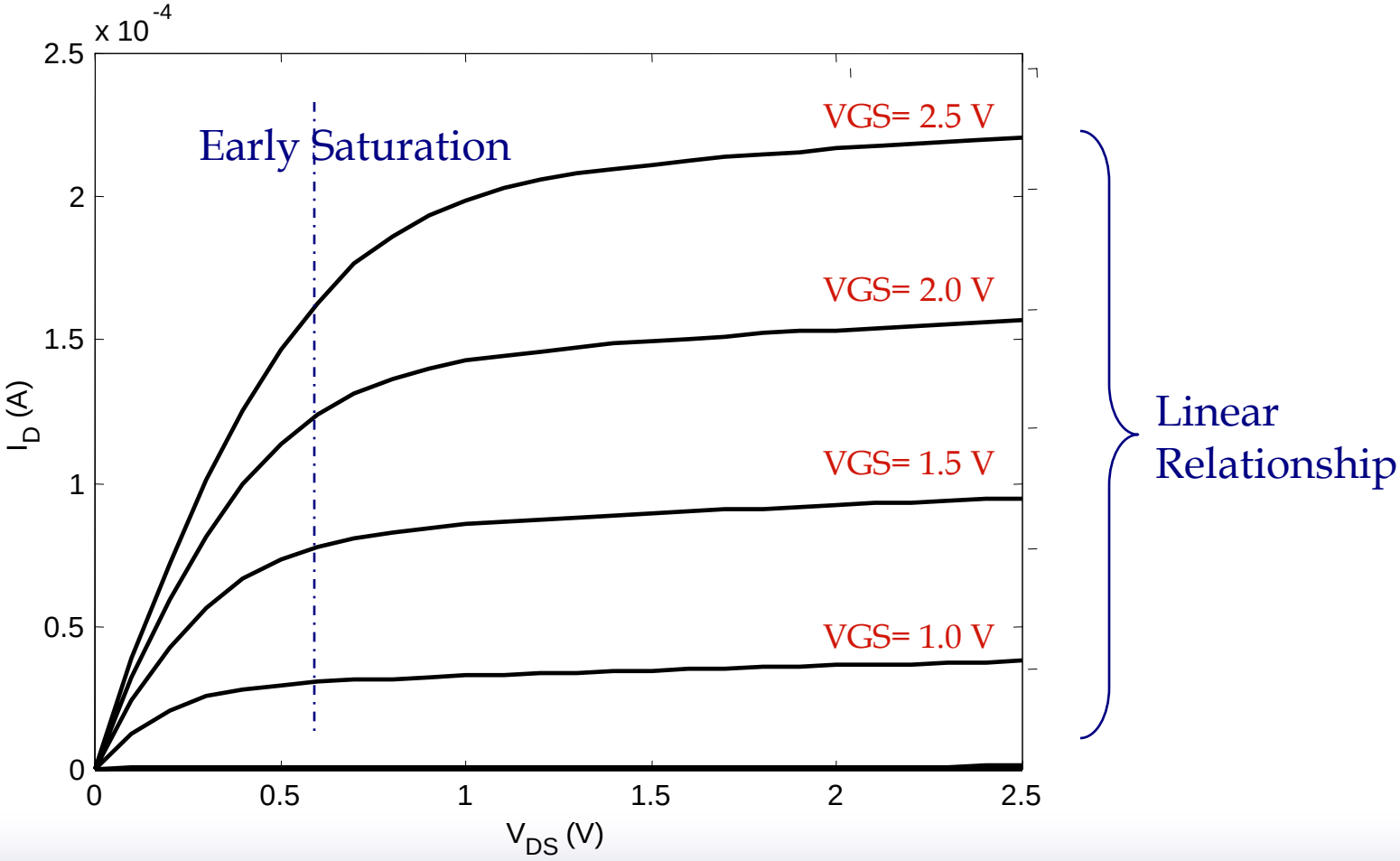
I/V curve for NMOS device:



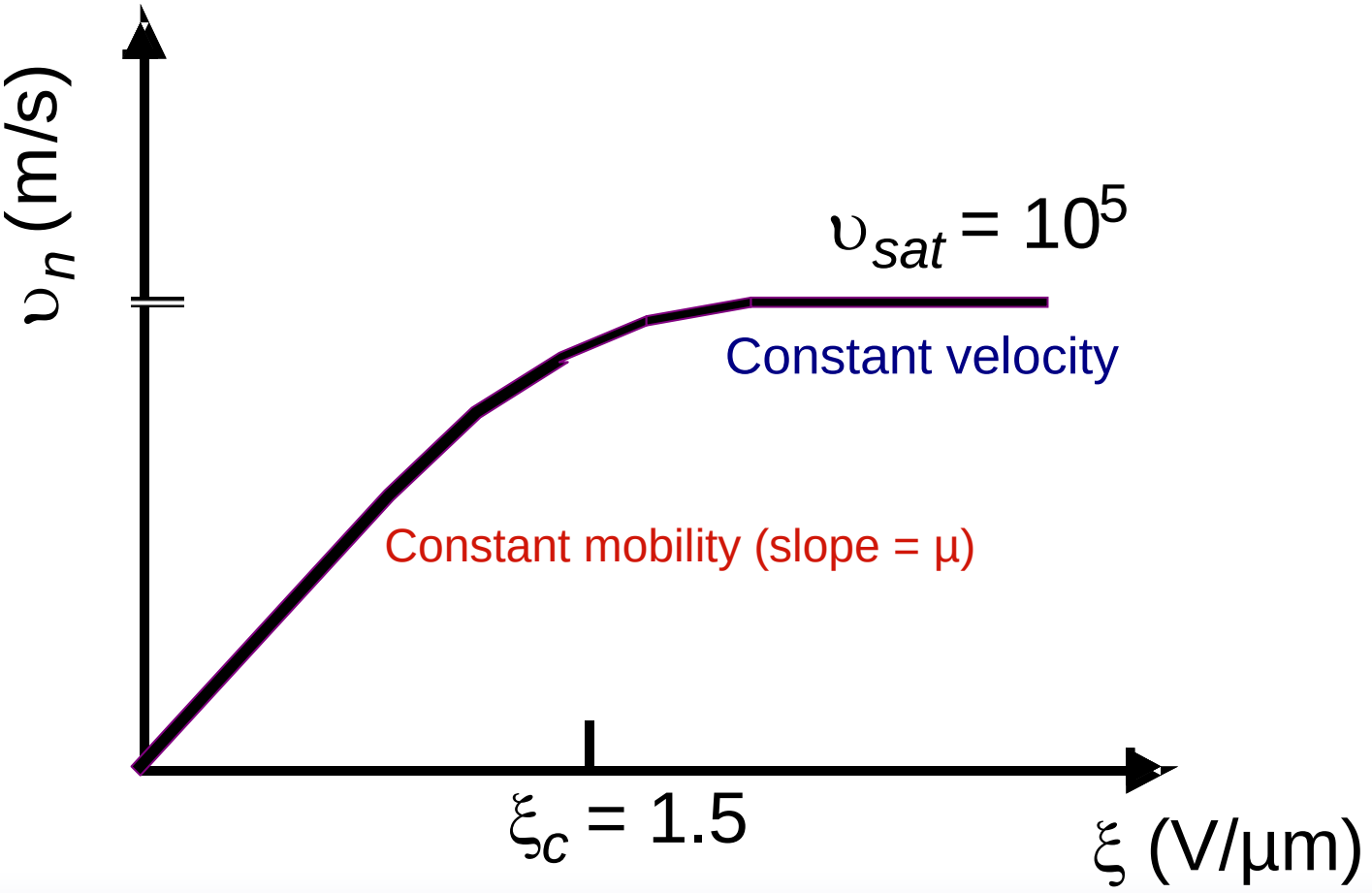
with channel-length modulation

without channel-length modulation ($\lambda=0$)

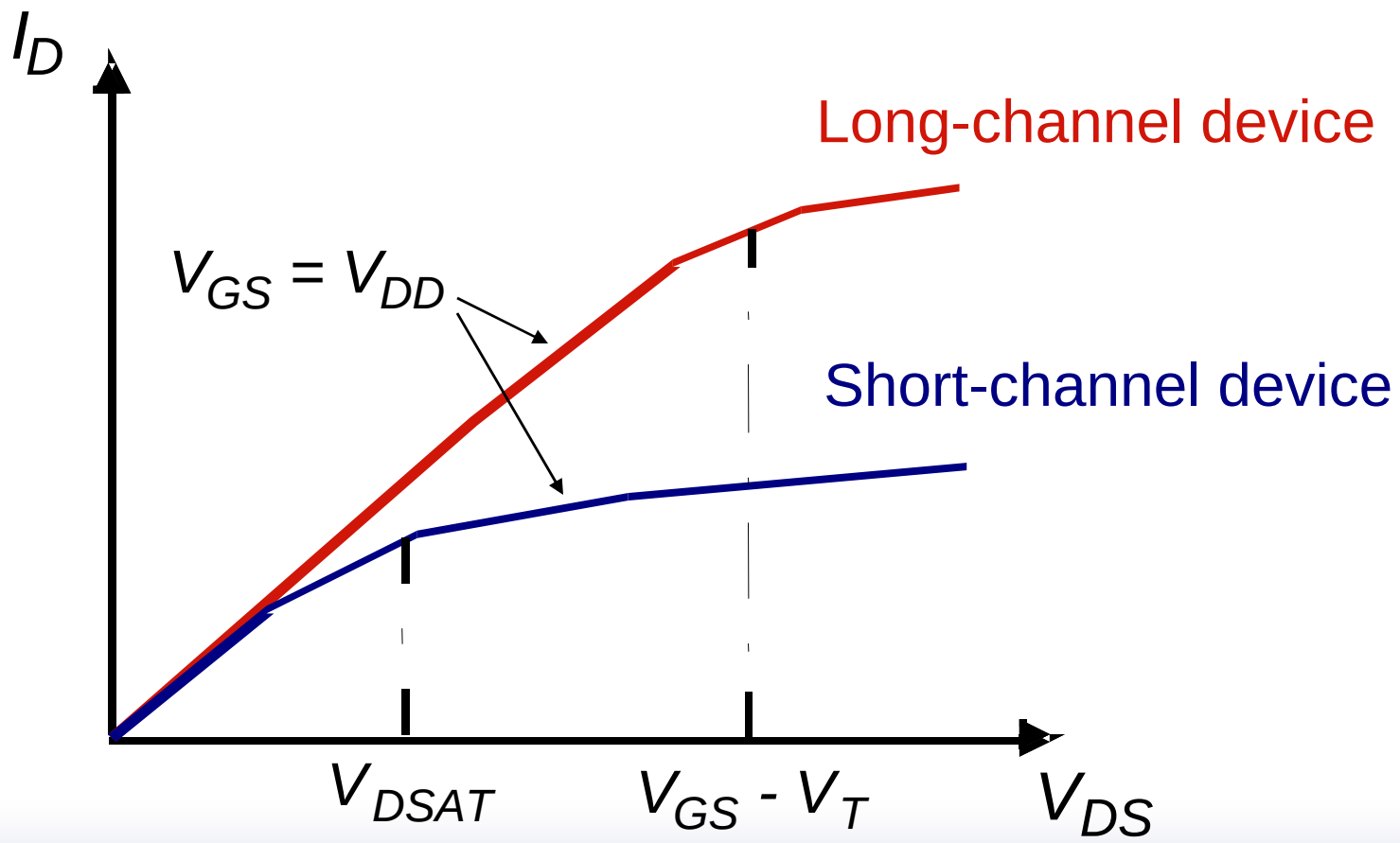
Current-Voltage Relations Short-Channel Transistors



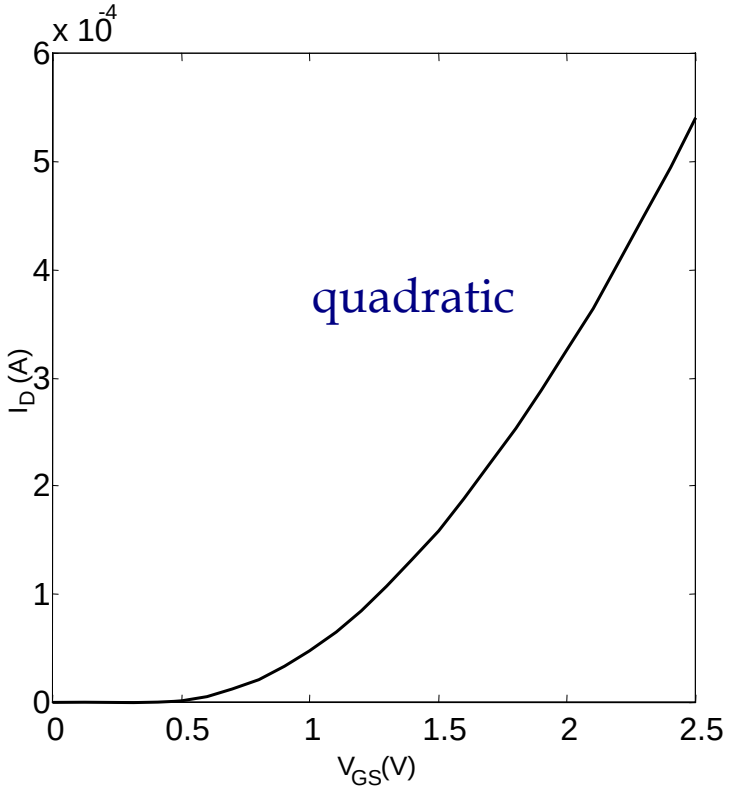
Velocity Saturation



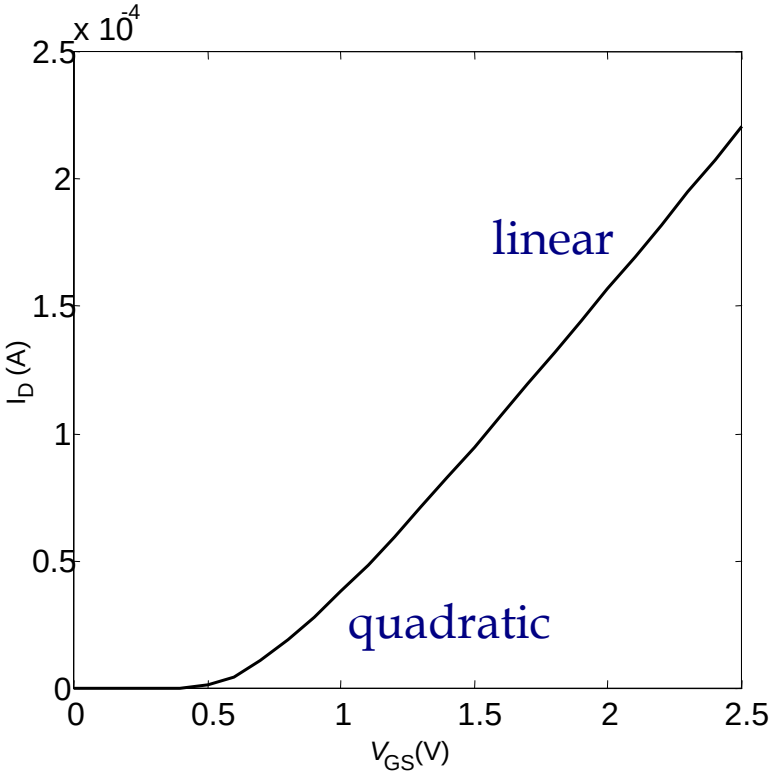
Perspective



I_D versus V_{GS}

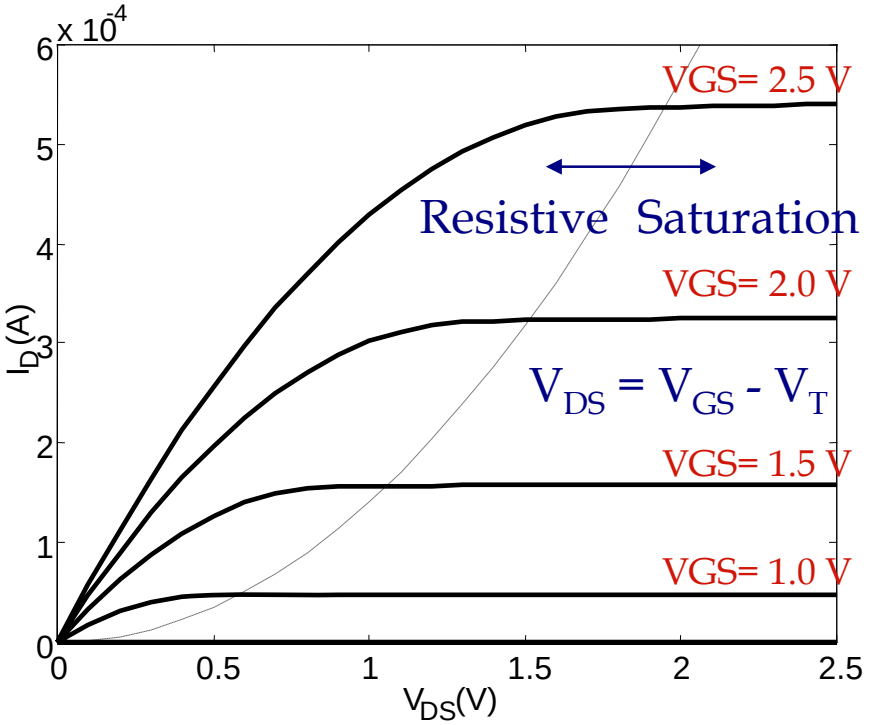


Long Channel

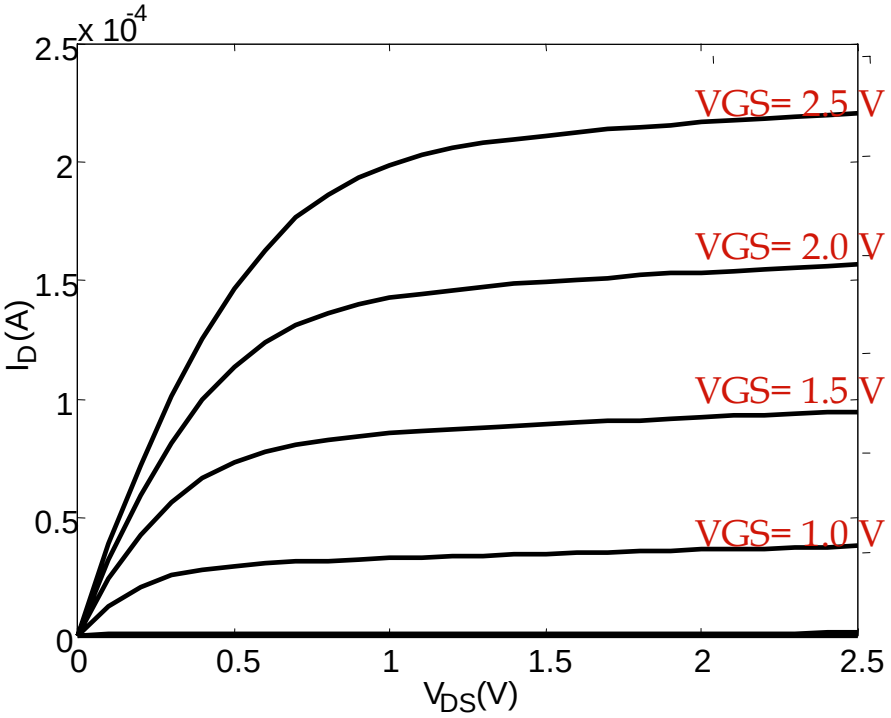


Short Channel

I_D versus V_{DS}

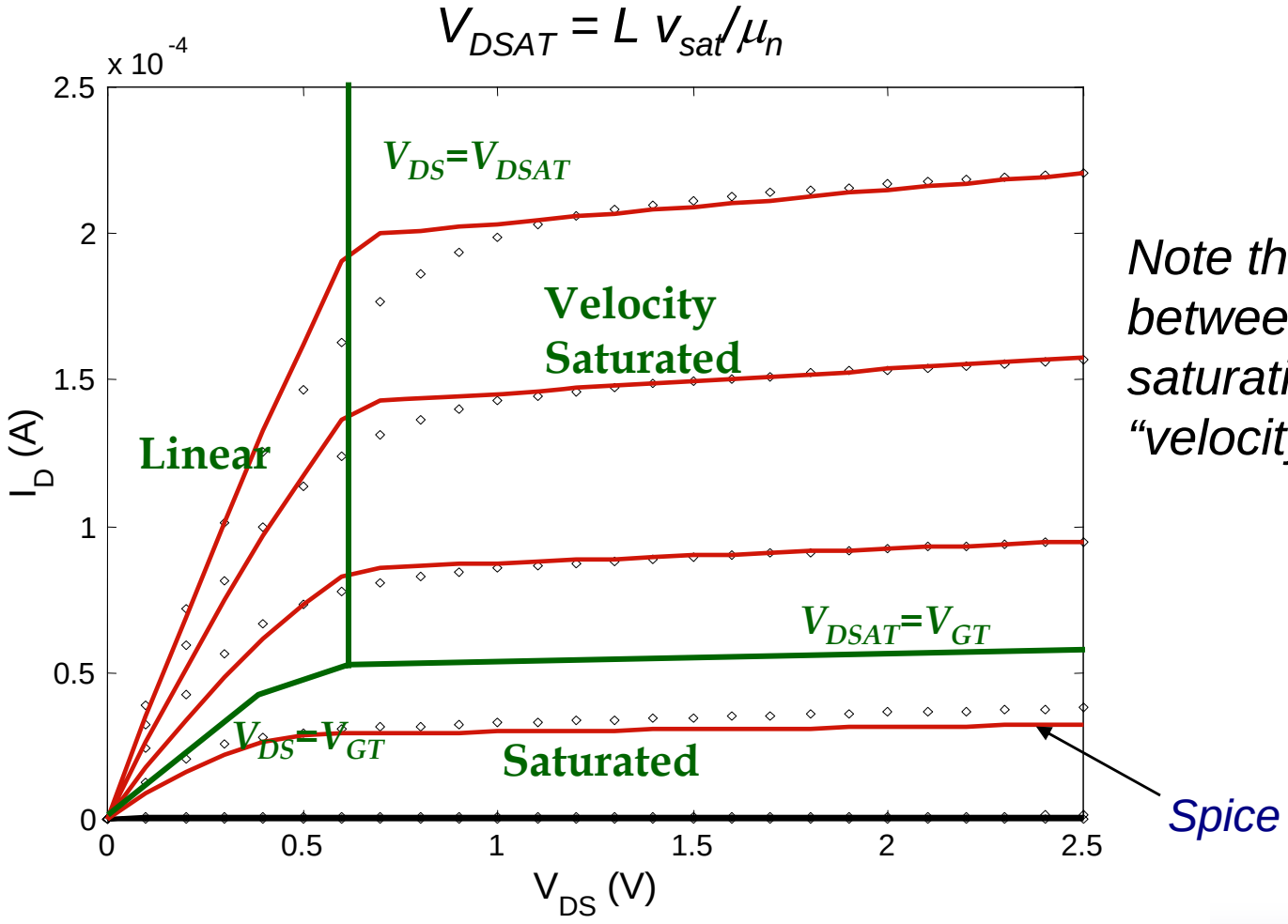


Long Channel



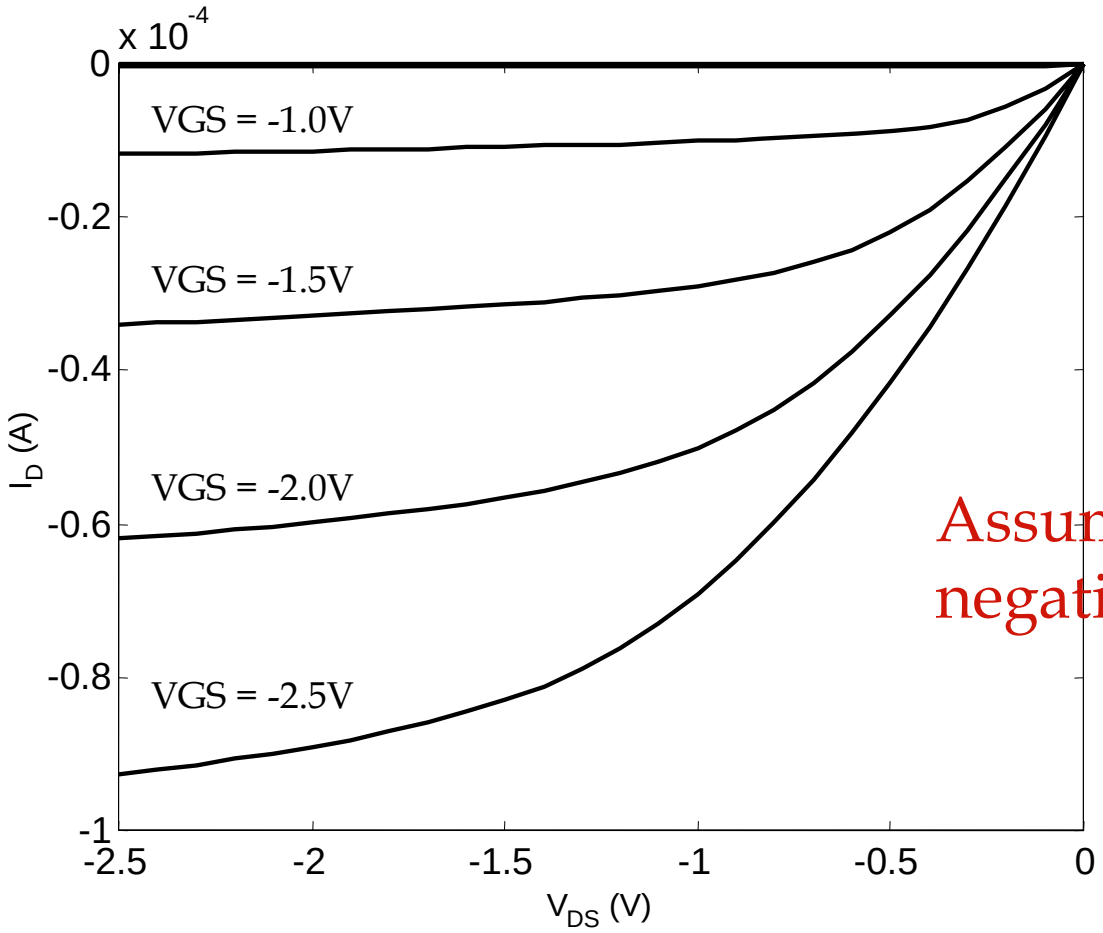
Short Channel

Simple Model versus SPICE



Note the difference between "transistor saturation" and "velocity saturation".

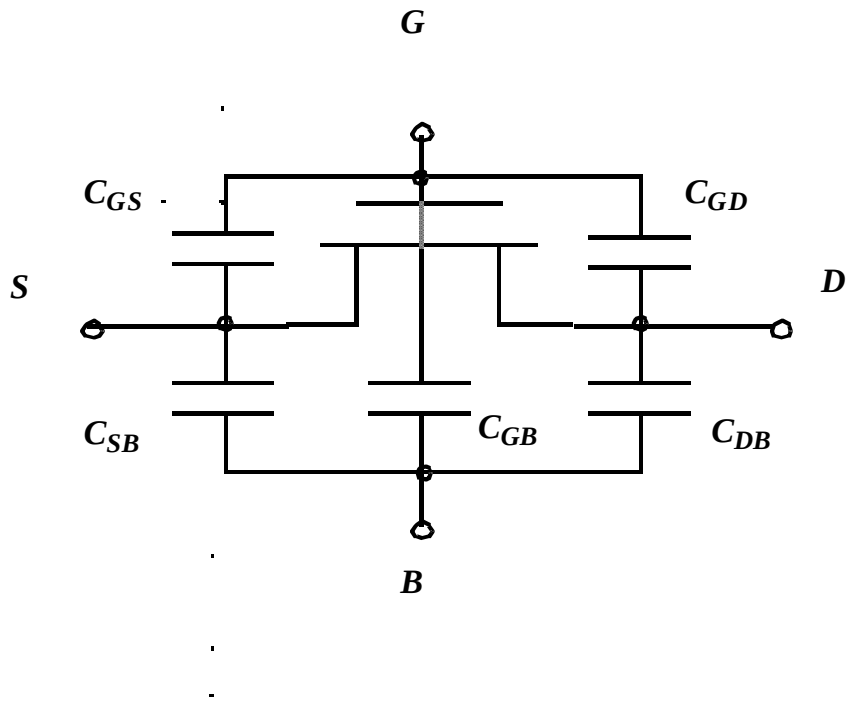
A PMOS Transistor (short-channel)



Assume all variables negative!

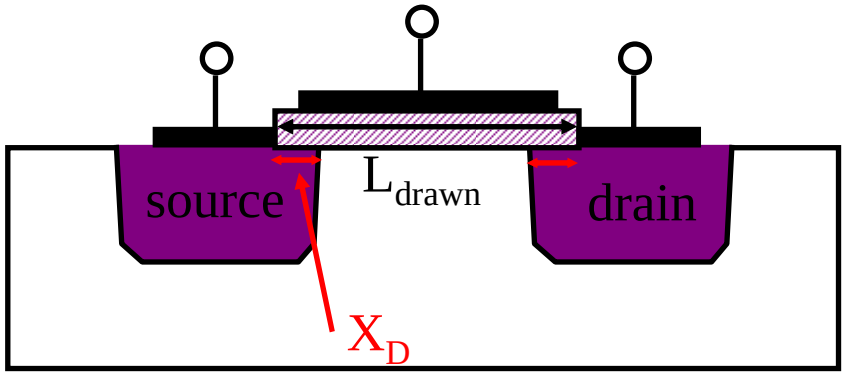
Dynamic Behavior of MOS Transistor

- Oxide Capacitance
  - Gate to Source overlap
  - Gate to Drain overlap
  - Gate to Channel/Bulk
- Junction Capacitance
  - Source to Bulk junction
  - Drain to Bulk junction



Oxide capacitances

Overlap

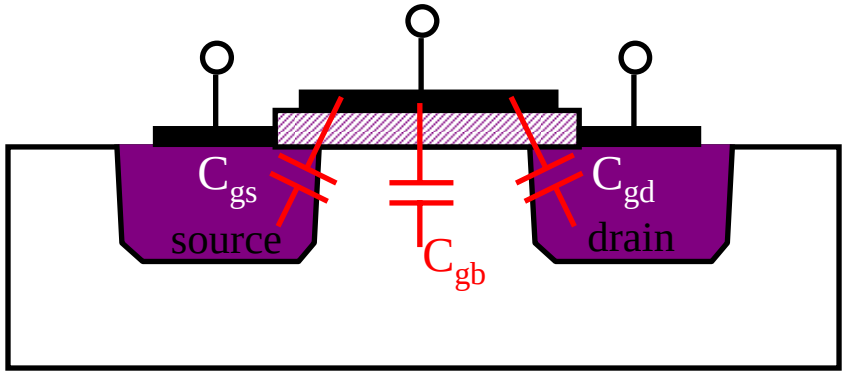


- Overlap capacitances
  - gate electrode overlaps source and drain regions
  - X_D is overlap length on each side of channel
  - $L_{eff} = L_d - 2X_D$
  - Total overlap capacitance:

$$C_{overlap} = C_{GSO} + C_{GDO} = 2C_{ox}WX_D$$

Oxide capacitances

Channel



- Channel capacitances
  - Gate-to-source: C_{gs}
  - Gate-to-drain: C_{gd}
  - Gate-to-bulk: C_{gb}

- Cutoff:
  - No channel connecting source and drain
  - $C_{gs} = C_{gd} = 0$
  - $C_{gb} = C_{ox}WL_{eff}$
  - Total channel capacitance $C_{GC} = C_{ox}WL_{eff}$

Oxide capacitances

Channel

- Linear mode
  - Channel spans from source to drain
  - Capacitance split equally between S and D

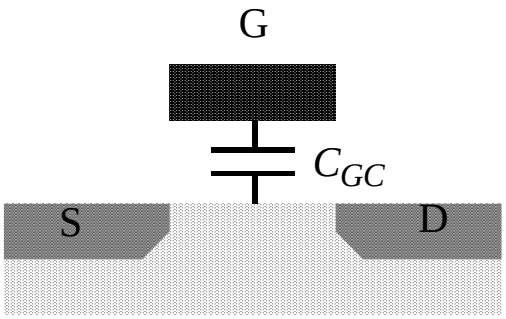
$$C_{GS} = \frac{1}{2} C_{ox} W L_{eff} \qquad C_{GD} = \frac{1}{2} C_{ox} W L_{eff} \qquad C_{GB} = 0$$

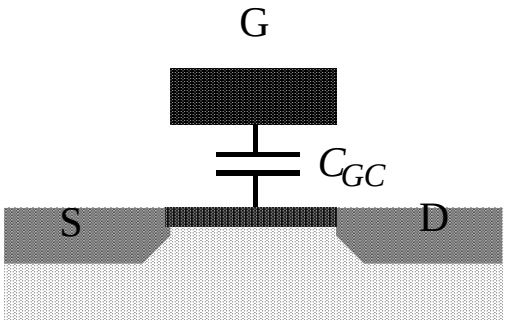
  - Total channel capacitance $C_{GC} = C_{ox} W L_{eff}$
- Saturation mode
  - Channel is pinched off:

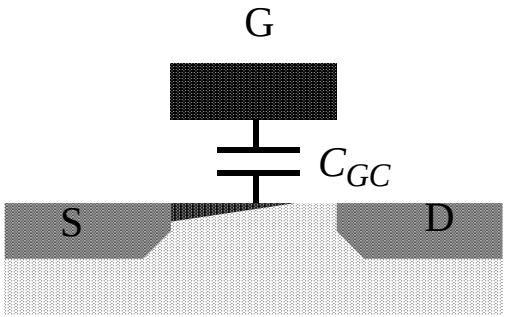
$$C_{GD} = 0 \qquad C_{GS} = \frac{2}{3} C_{ox} W L_{eff} \qquad C_{GB} = 0$$

  - Total channel capacitance $C_{GC} = 2/3\ C_{ox} W L_{eff}$

Gate-to-Channel Capacitance (summary)



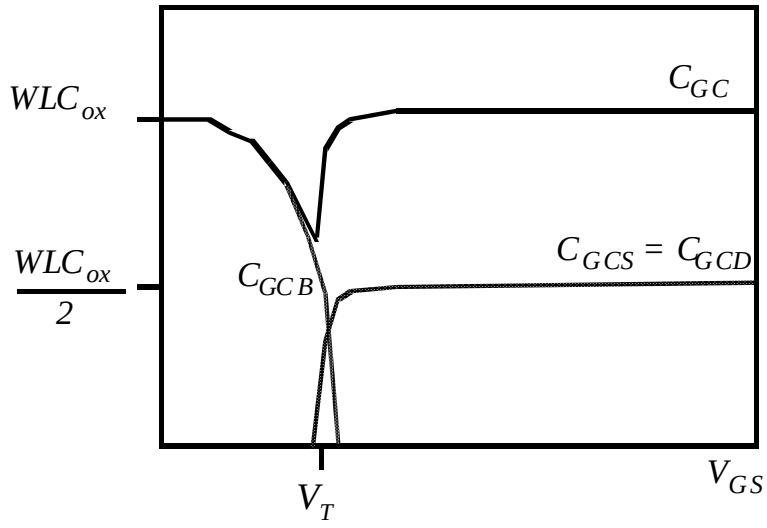


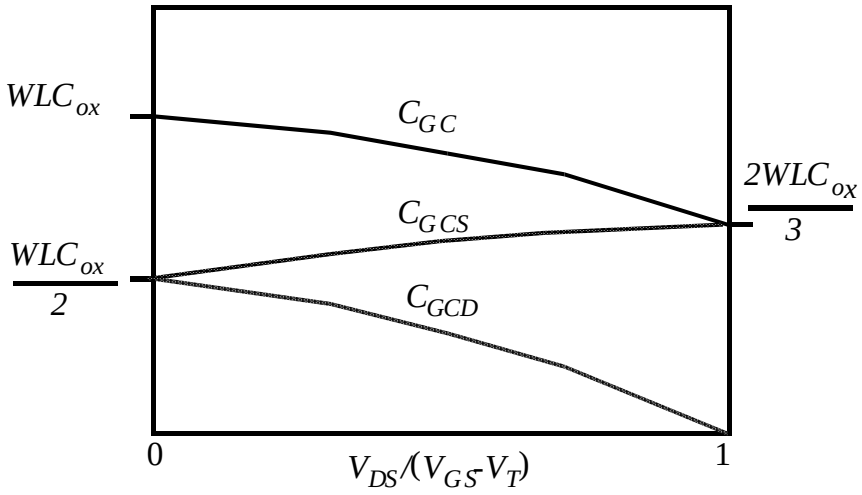


$$C_{GC} = C_{gb} + C_{gs} + C_{gd}$$

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

Gate-to-Channel Capacitance



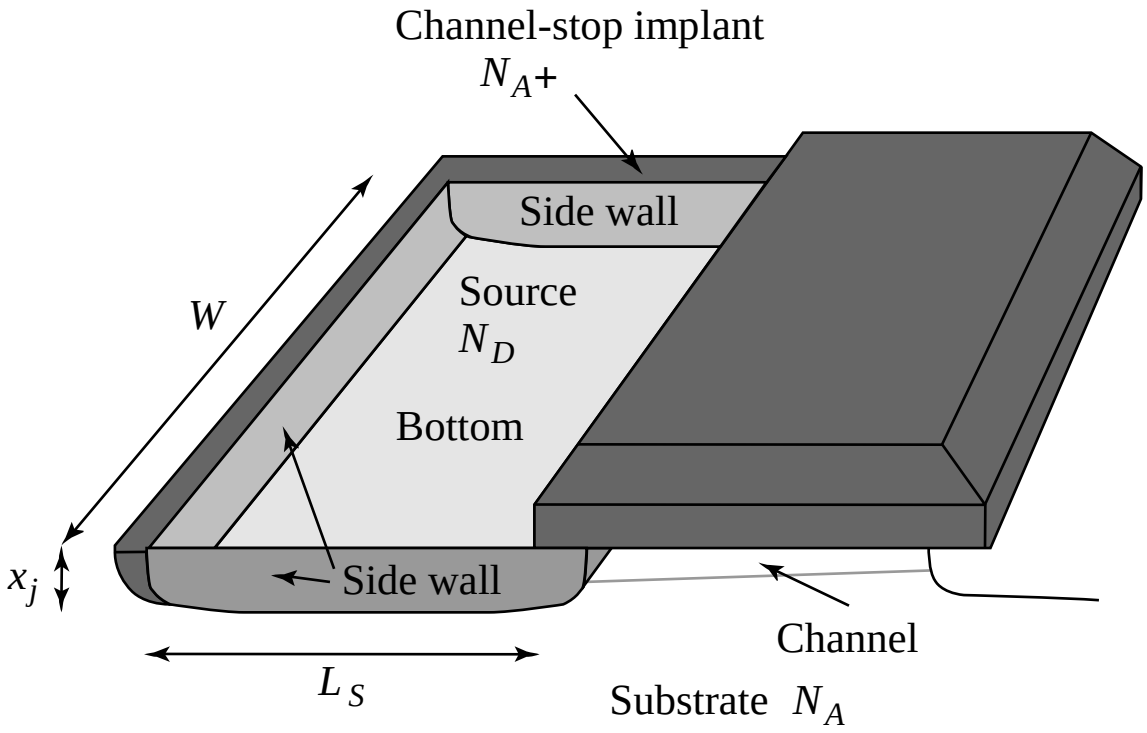


Capacitance as a function of VGS (with VDS = 0)

Capacitance as a function of the degree of saturation

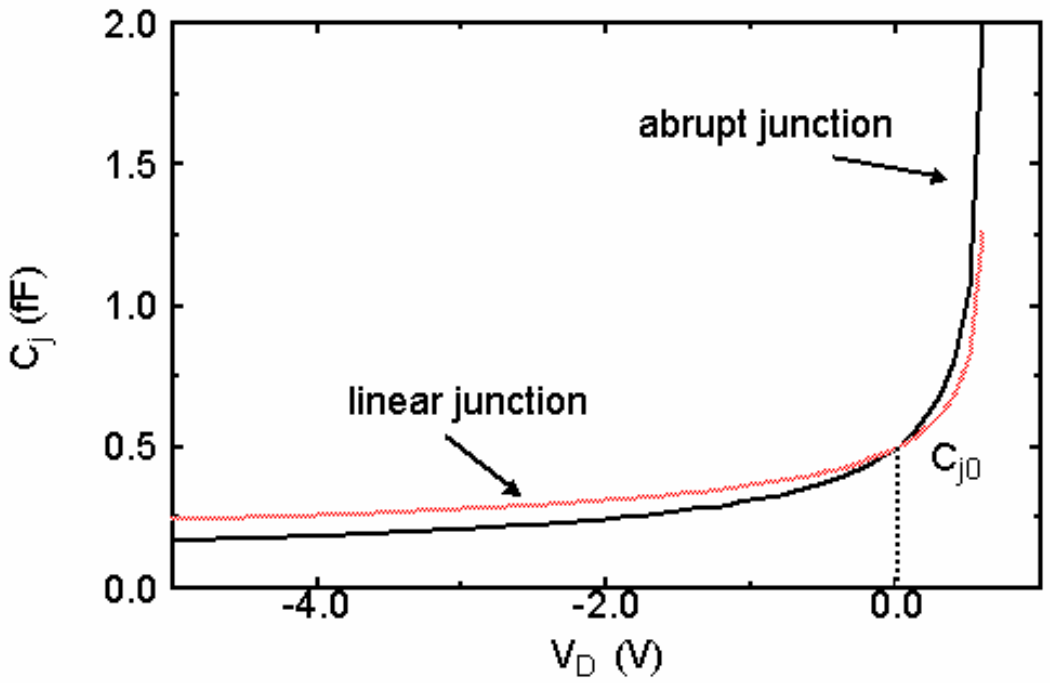
Bottom Line: Cap. components are non-linear

Diffusion Capacitance



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

Junction Capacitance



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

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

Linearizing the Junction Capacitance

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

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

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

Capacitances in 0.25 μm CMOS process

	C_{ox} (fF/μm^2)	C_O (fF/μm)	C_j (fF/μm^2)	m_j	ϕ_b (V)	C_{jsw} (fF/μm)	m_{jsw}	ϕ_{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

Data Dependency

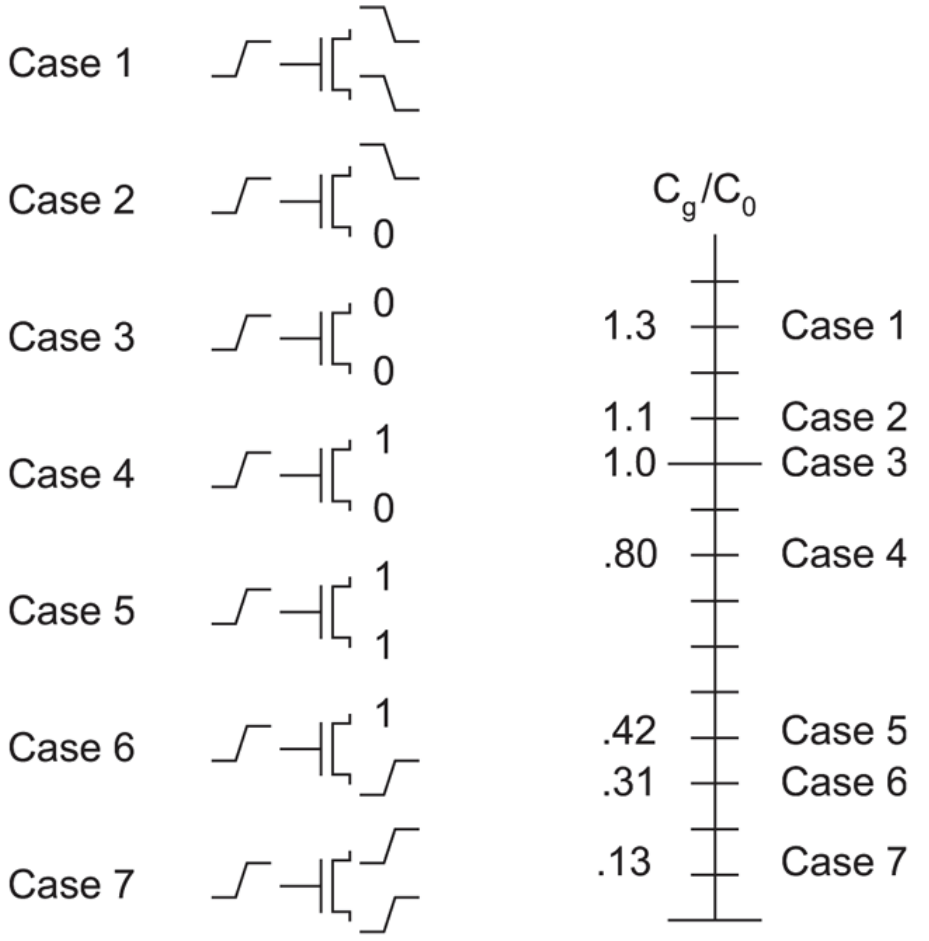


FIG 2.12 Data-dependent gate capacitance

MOS Cap. Summary

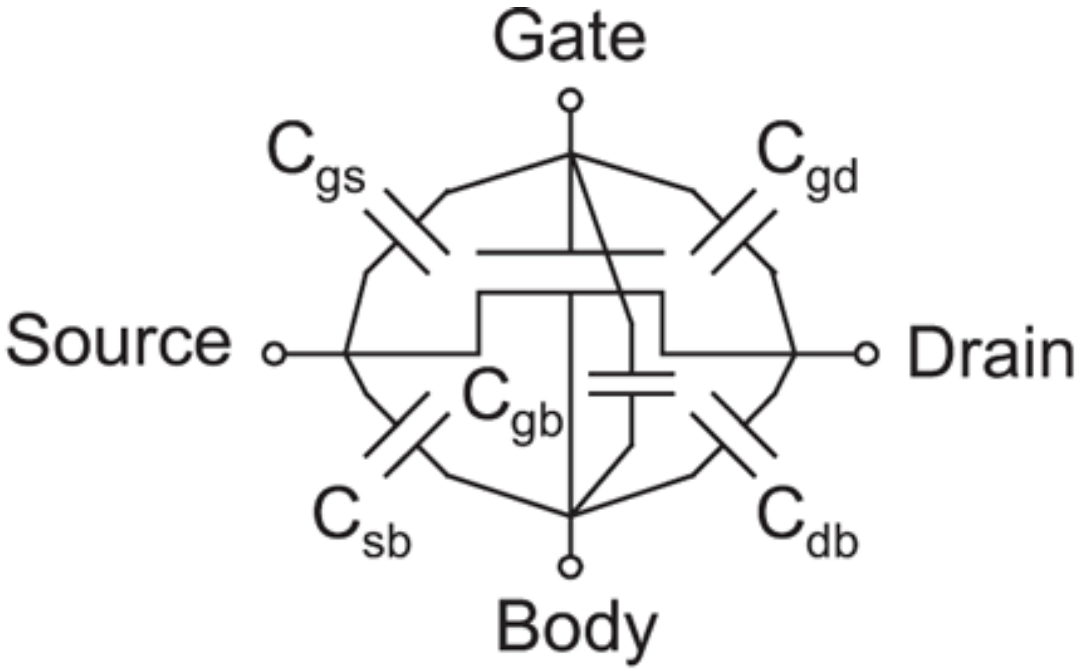


FIG 2.14 Capacitances of an MOS transistor