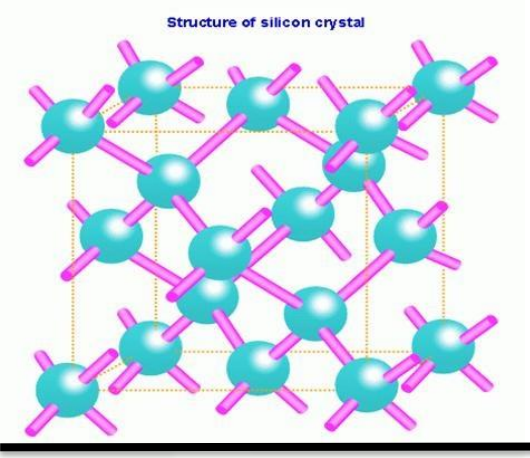
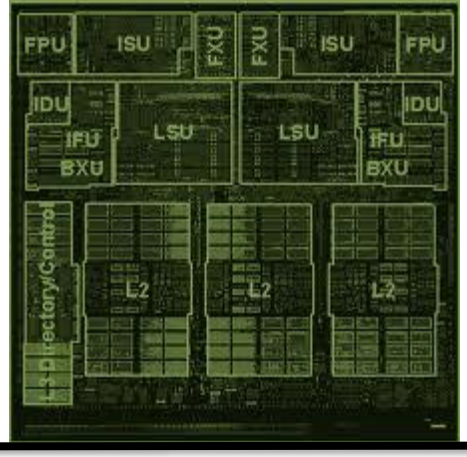
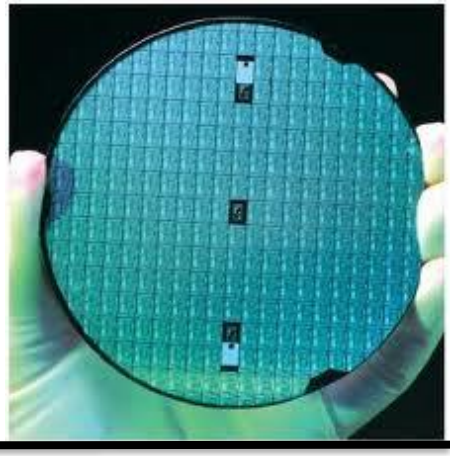






ECE 225

Lecture 4

MOS & MOSFET Review

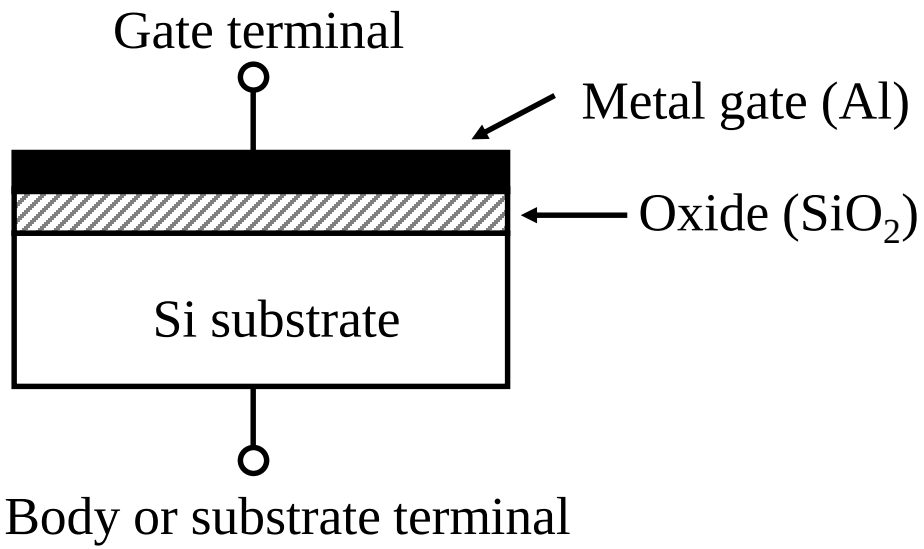
Prof. Kaustav Banerjee
Electrical and Computer Engineering
University of California, Santa Barbara

Outline

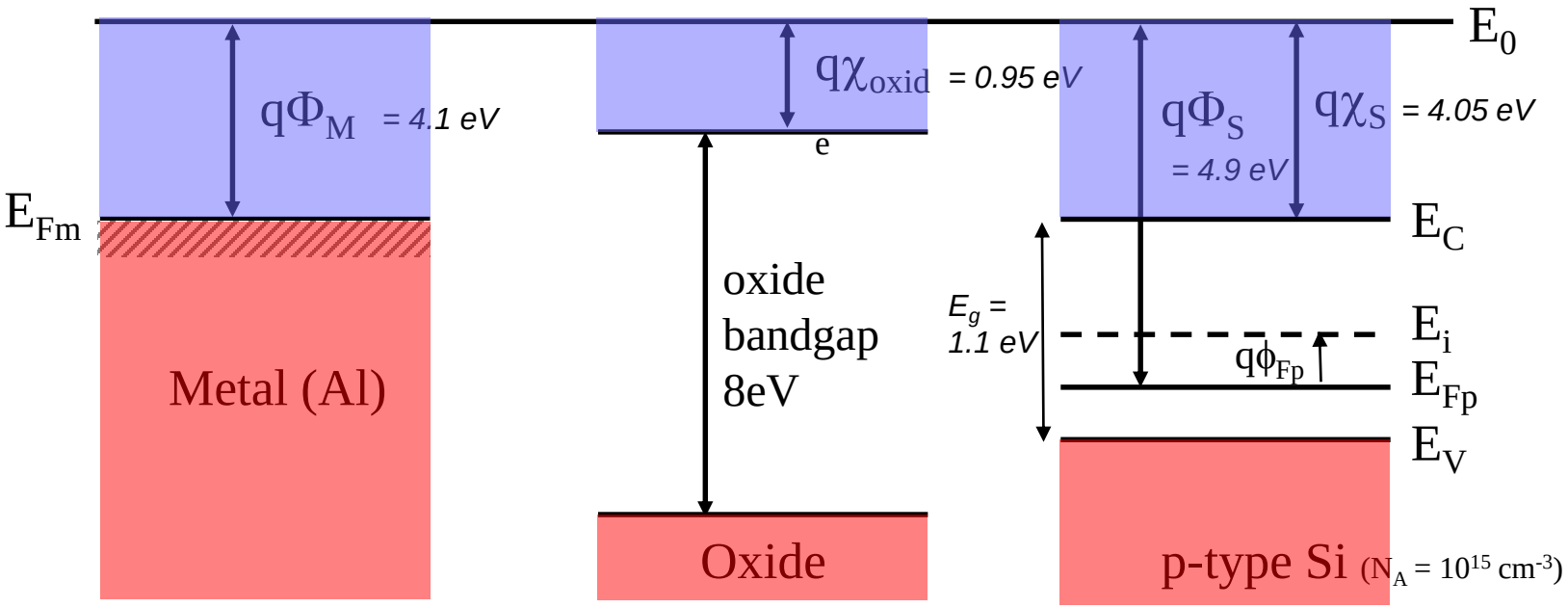
- MOS Structure
- Derivation of Flat-band condition
- Derivation of Surface-accumulation condition
- Derivation of Surface-depletion condition
- Derivation of Threshold condition and threshold voltage
- Derivation of Strong-inversion (beyond V_T) condition
- MOS/MOSFET C-V characterization
- MOSFET Scaling

MOS Structure

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



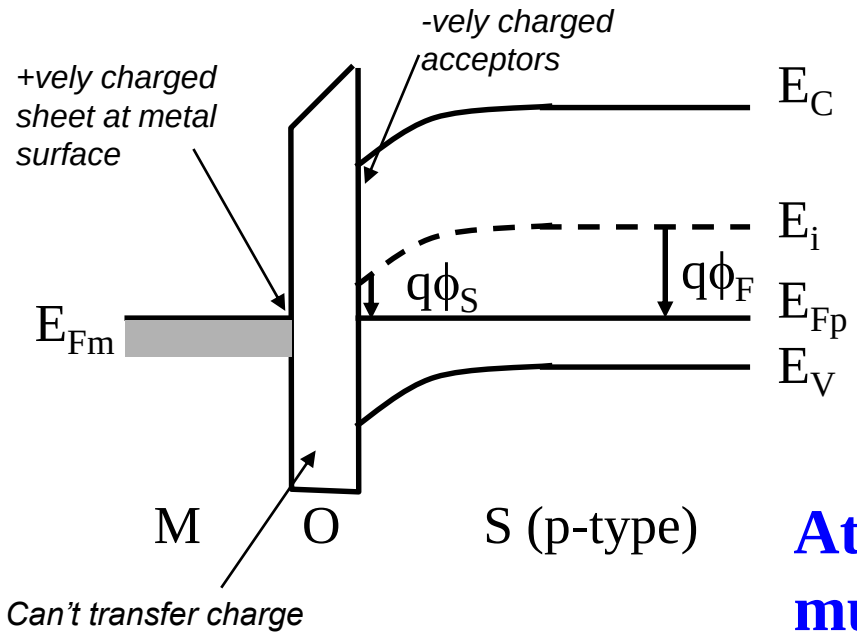
MOS Energy Band Diagram



- Work function ($q\Phi_M$, $q\Phi_S$): energy required to take electron from Fermi level to free space
- Electron affinity is the energy required to move an electron from conduction band to free space (E_0) = qX_s
- Work function difference between Al and Si = 0.8V

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**



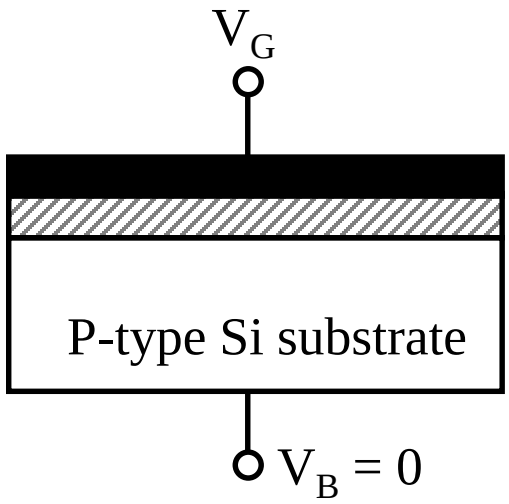
ϕ_F = Fermi potential (difference between E_F and E_i in bulk)

ϕ_S = surface potential

At equilibrium, Fermi levels must line up!!

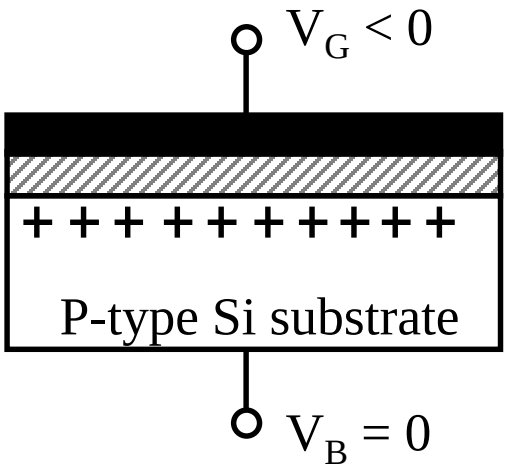
MOS Capacitor Operation

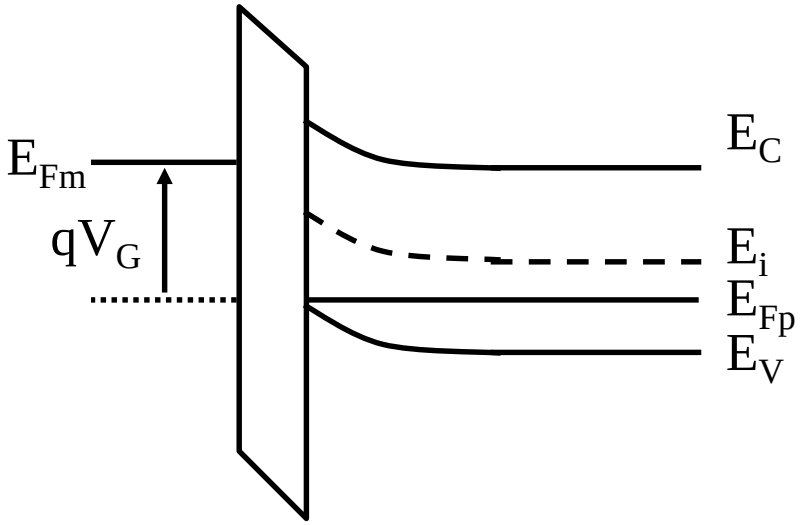
- Assume p-type substrate
- Three regions of operation
  - Accumulation ($V_G < 0$)
  - Depletion ($V_G > 0$ but small)
  - Inversion ($V_G >> 0$)



Accumulation

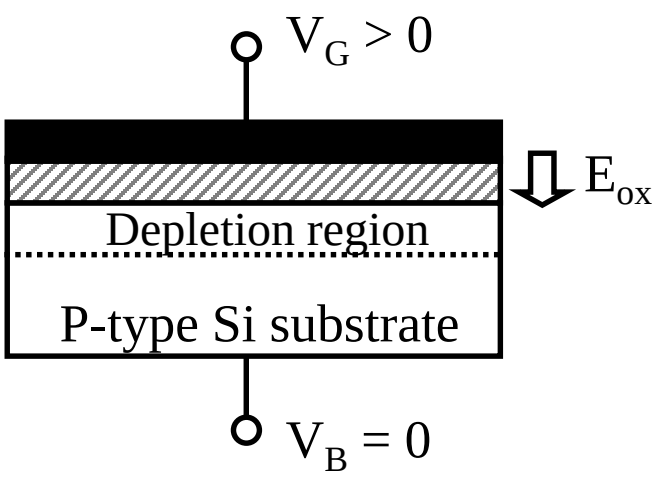
- Negative voltage on gate: attracts holes in substrate towards oxide
- Holes “accumulate” on Si surface (surface is more strongly p-type)
- Electrons pushed deeper into substrate

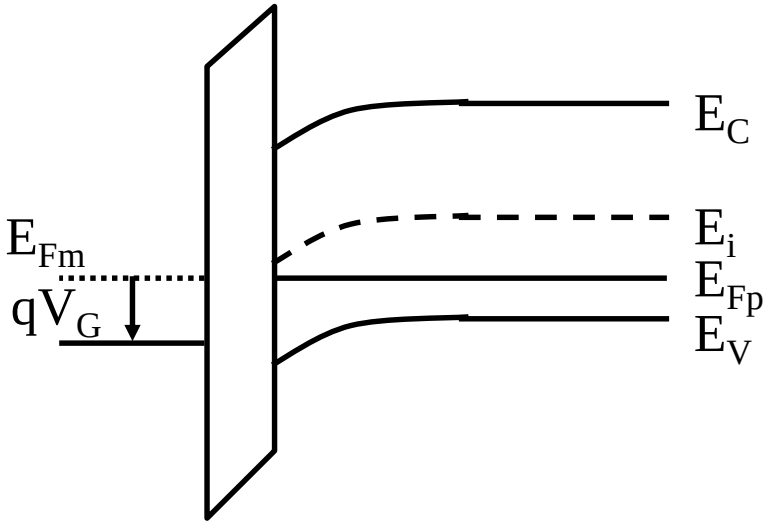




Depletion

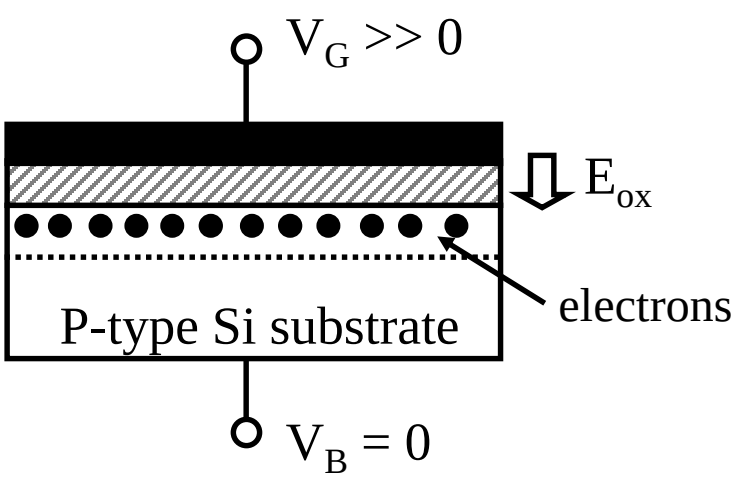
- Positive voltage on gate: repels holes in substrate
  - Holes leave behind negatively charged acceptor ions
- Depletion region forms: devoid of carriers
  - Electric field directed from gate to substrate
- Bands bend downwards near surface
  - Surface becomes <u>less</u> strongly p-type (E_F close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

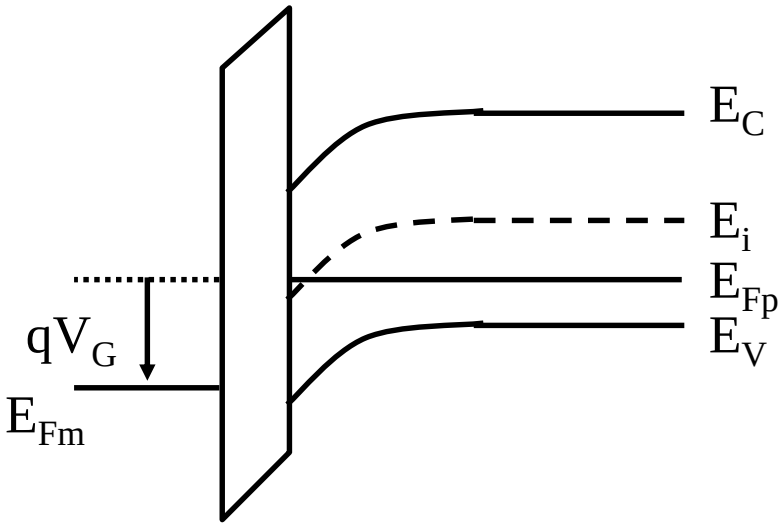




Inversion

- 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



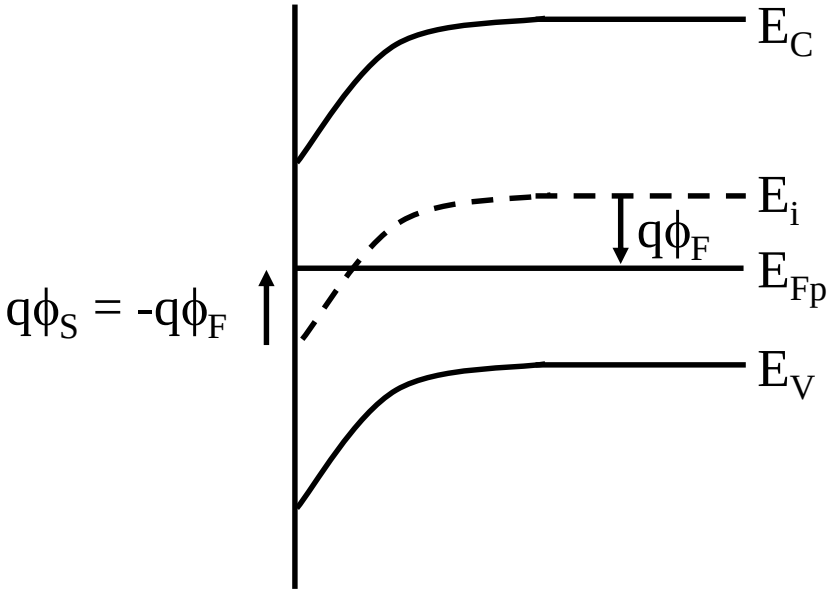


Inversion

- 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:

$q\phi_F = E_F - E_i$



Summary…for p-type Si substrate with n+poly-Si gate electrode

- Flat-band condition: $V_{fb} = \Psi_g - \Psi_s$
- Surface-accumulation condition:

  $V_{ox} = -Q_{acc}/Cox$; $Q_{acc} = -C_{ox}\,(V_g - V_{fb})$
- Surface-depletion condition:

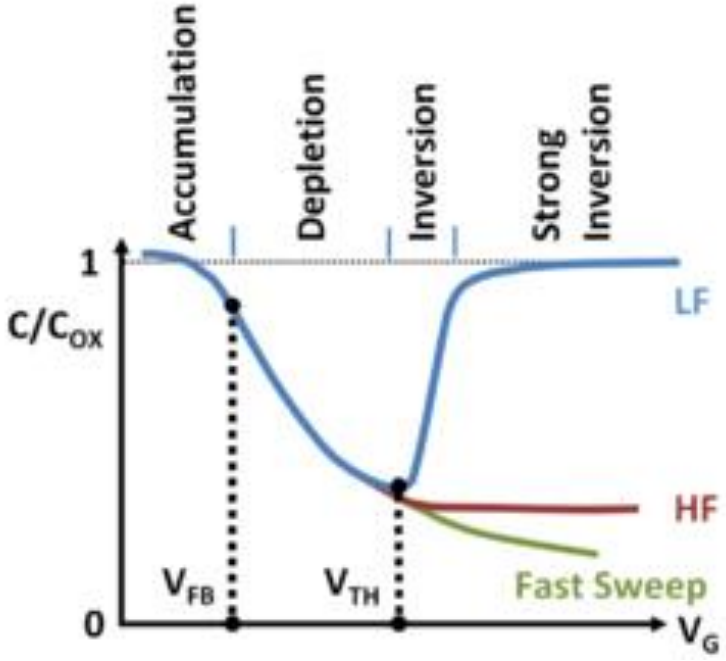
  $V_g = V_{fb} + (qN_aW_{dep}^2)/2\varepsilon_s + (qN_aW_{dep})/C_{ox}$
- Threshold condition: $\phi_{st} = 2\phi_b = 2kT/q\ \ln(N_a/n_i)$
- Threshold Voltage: $V_t = V_{fb} + \phi_s + V_{ox}$

  $= V_{fb} + 2kT/q\ \ln(N_a/n_i) + \{(qN_a2\varepsilon_s2\phi_b)^{1/2}\}/C_{ox}$
- Strong Inversion: $Q_{inv} = -C_{ox}\,(V_g - V_t)$

Summary…MOS CV Characteristics

C-V measurements useful for determining:

t_{ox}, N_a or N_d, V_t, and V_{fb}



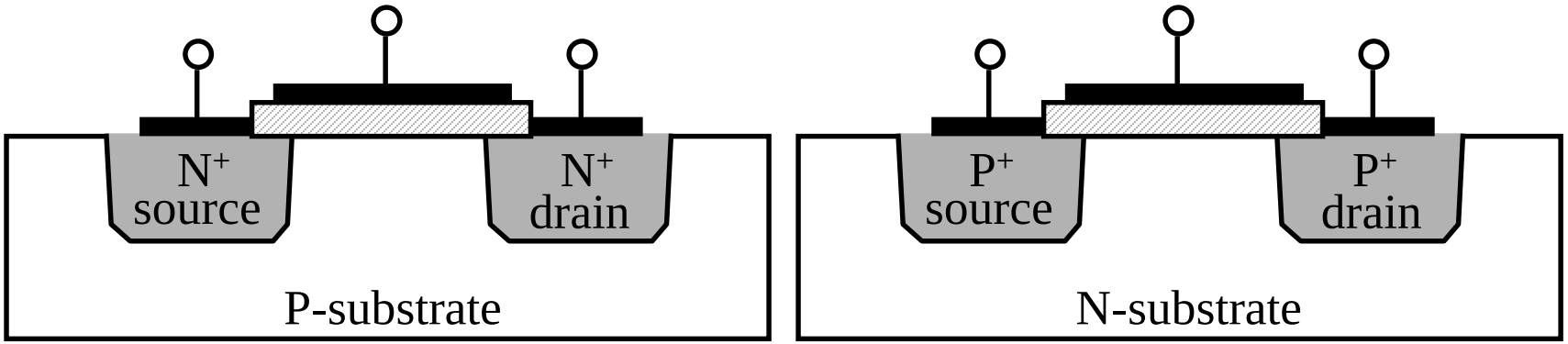
Typical MOS Capacitor C-V

$$1/C = \{(1/C_{ox}^2 + 2(V_g - V_{fb})/qN_a\varepsilon_s)\}^{1/2}$$

Recall the main differences w.r.t the C-V for a MOSFET…at any frequency!!

MOS Transistor

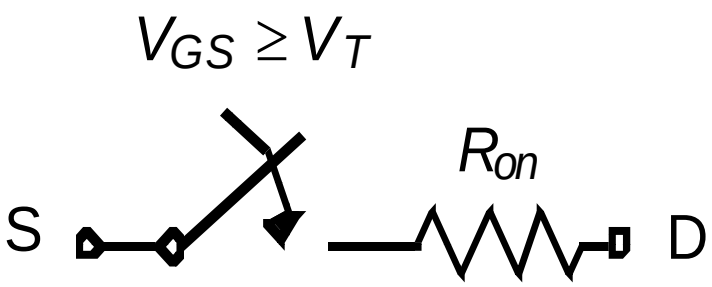
- Add “source” and “drain” terminals to MOS capacitor
- Transistor types
  - NMOS: p-type substrate, n^+ source/drain
  - PMOS: n-type substrate, p^+ source/drain

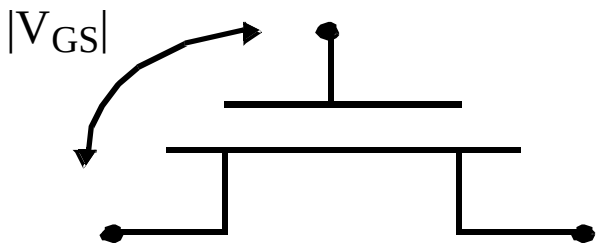


NMOS PMOS

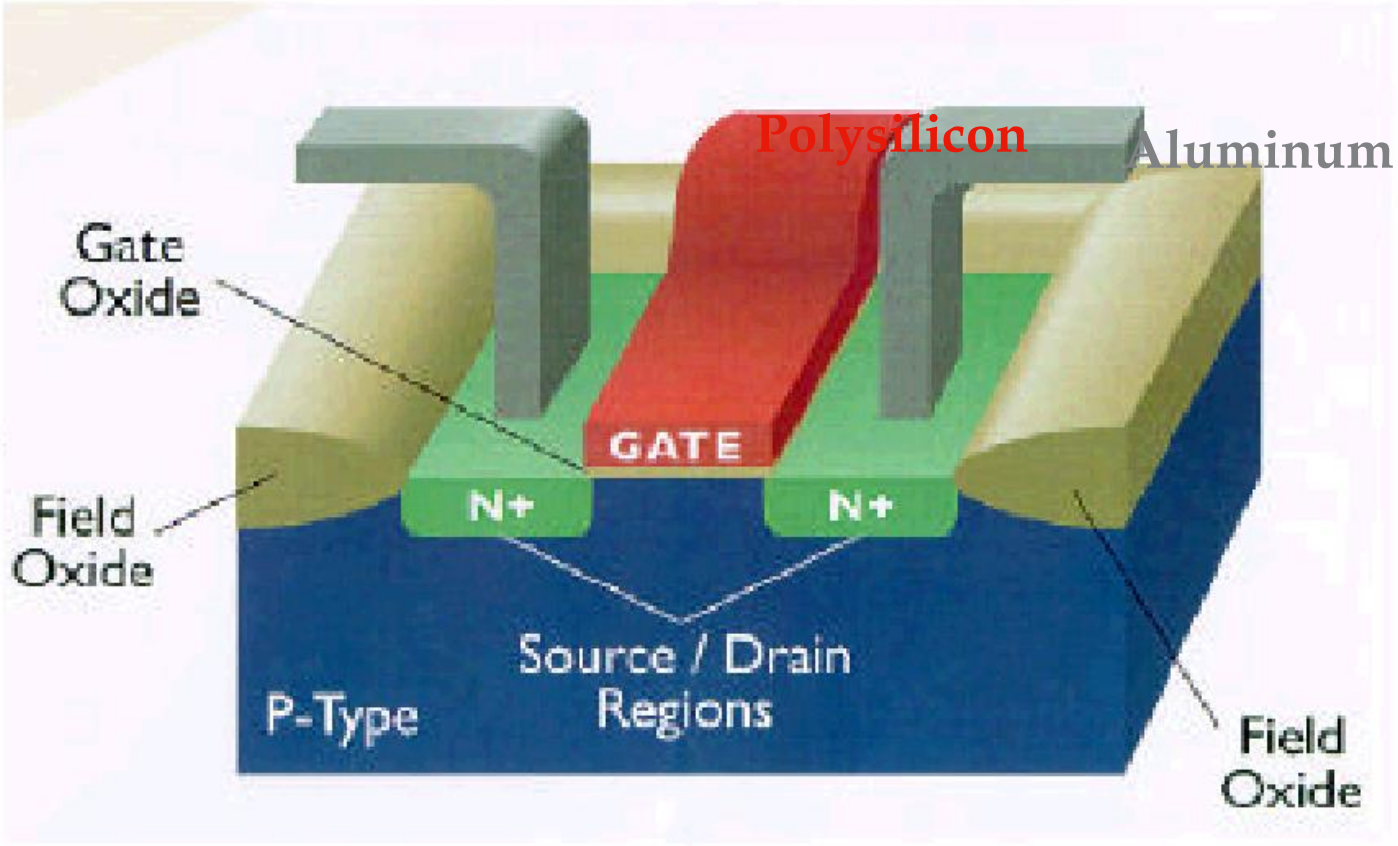
What is a Transistor?

A Switch! ⟷ An MOS Transistor



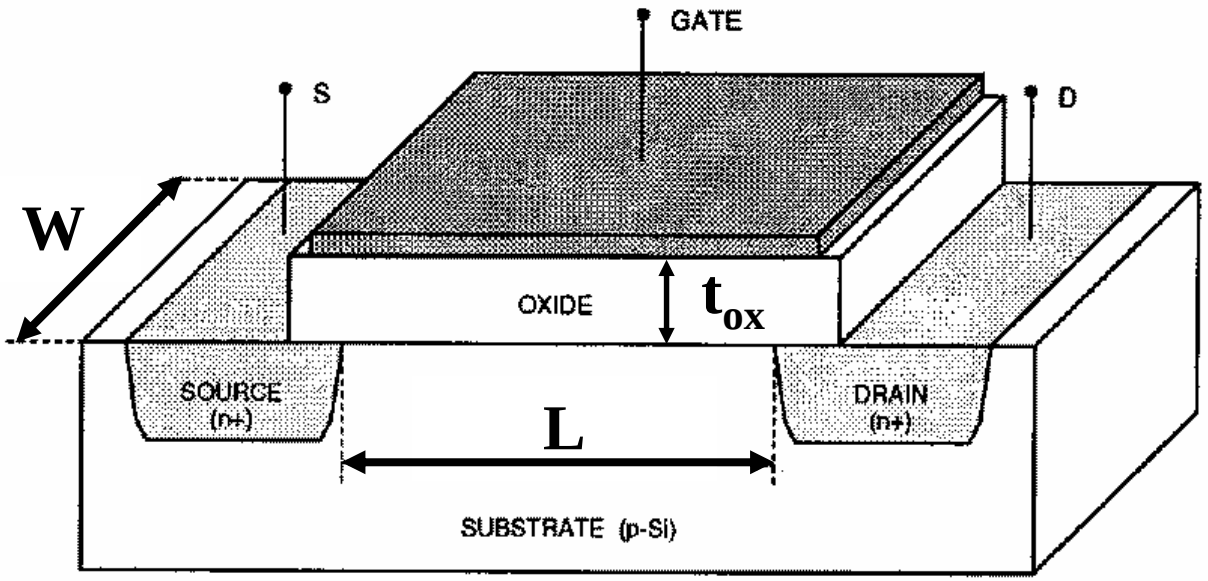


The MOS Transistor



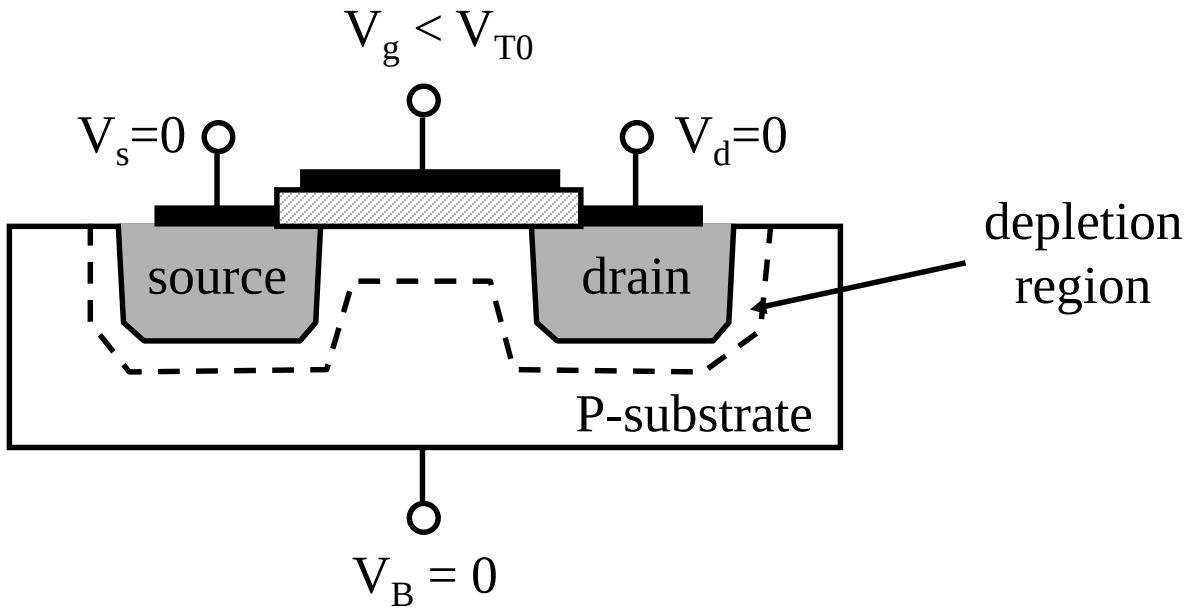
MOS Transistor

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



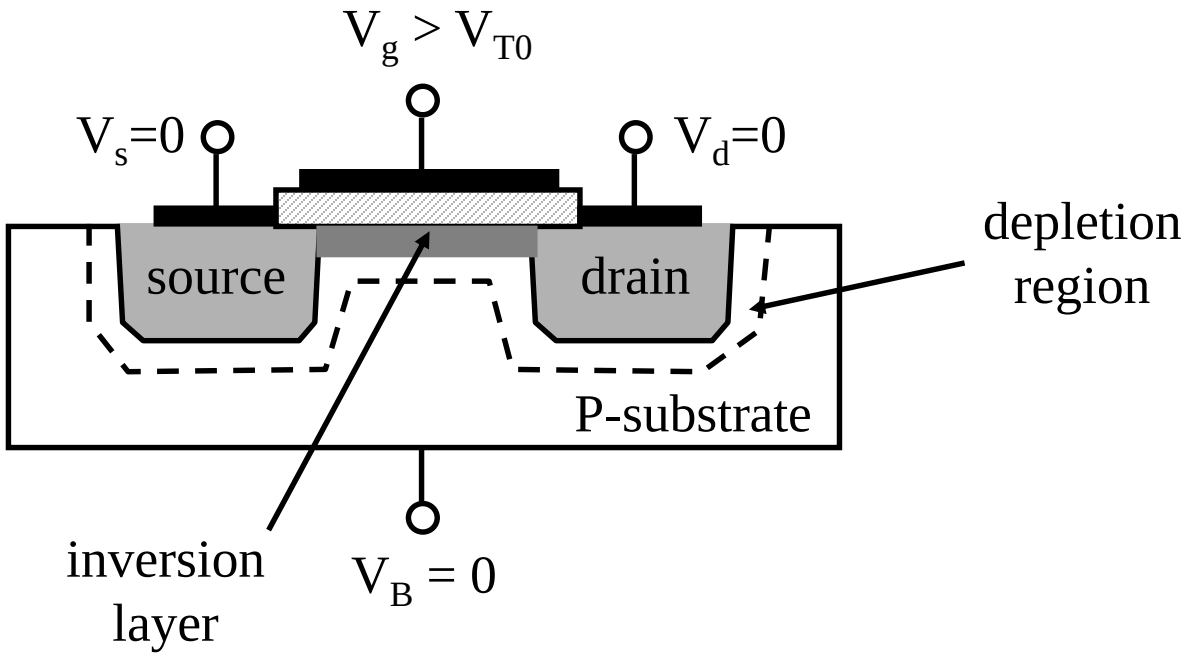
MOS Transistor Operation

- 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*

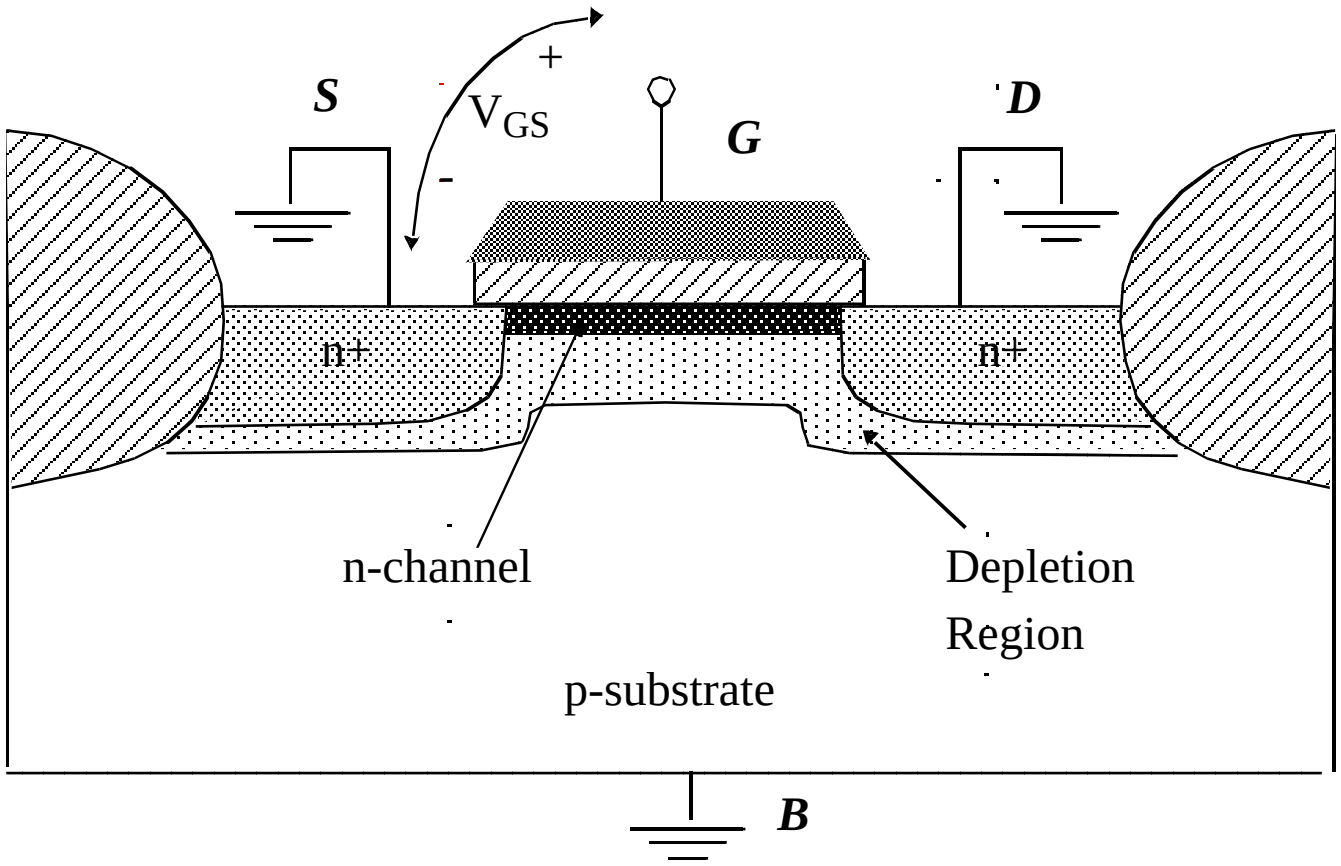


MOS Transistor Operation

- 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:

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

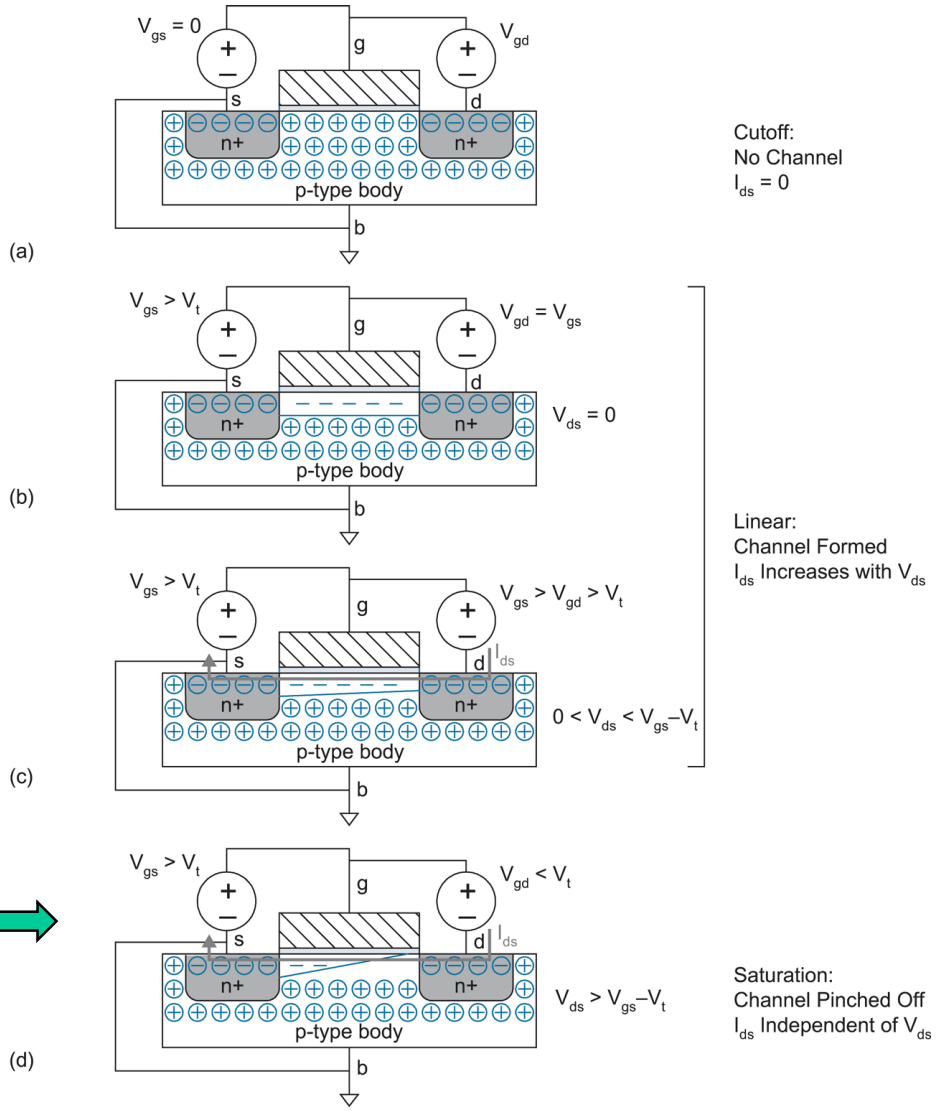
+ for NMOS
- for PMOS

- 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)	$V_{T0} > 0$	$V_{T0} < 0$

MOSFET Operation (NMOS)



Channel voltage no longer > V_t all along the channel….

$V_{GS} - V(x) < V_t$
Conduction dominated by DRIFT

Pinch-off: conduction still takes place from Source to Drain due to drift of electrons under the influence of the +ve drain voltage

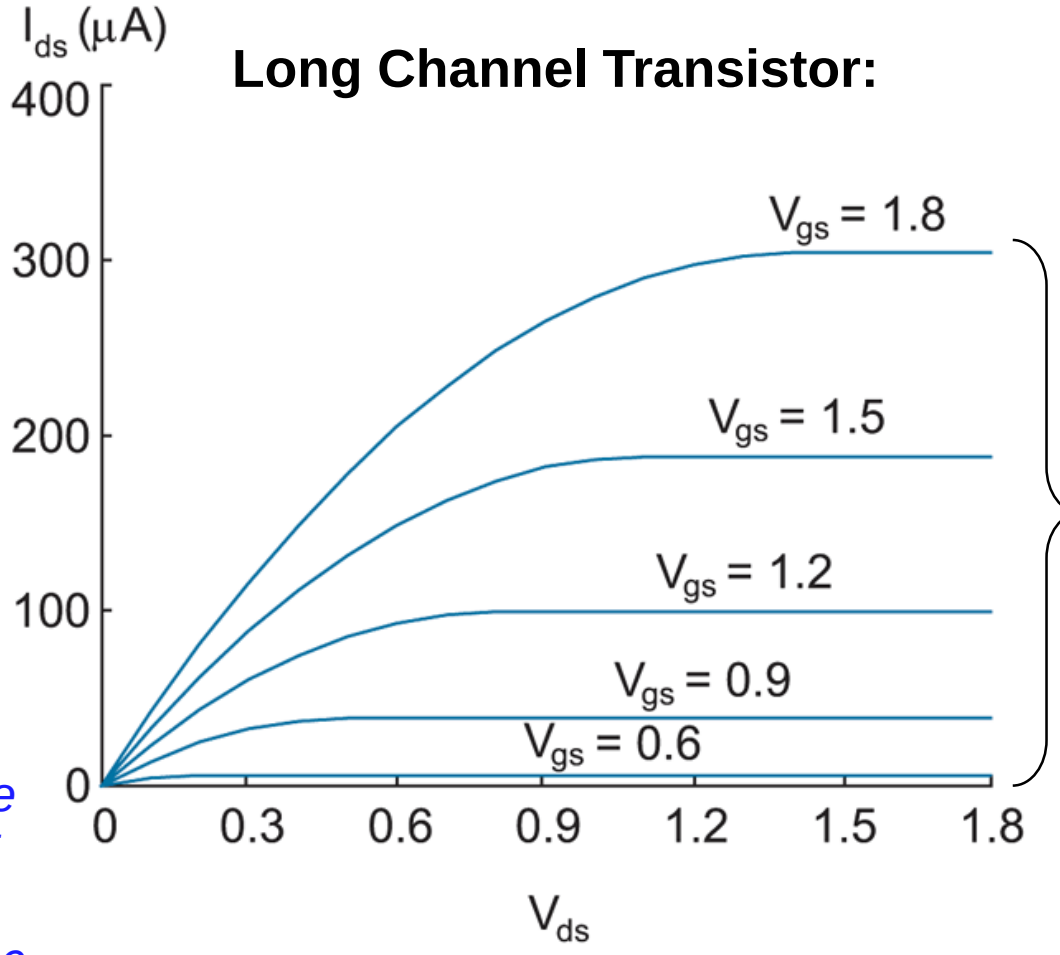
FIG 2.3 nMOS transistor demonstrating cutoff, linear, and saturation regions of operation

NMOS Characteristics

Recall that the surface channel vanishes at the drain end of the channel….when current saturates…known as "pinch-off"

Any voltage $V_{DS} > V_{DSAT}$ is dropped across the high-field pinch-off region…where inversion charge =0

Long Channel Transistor:



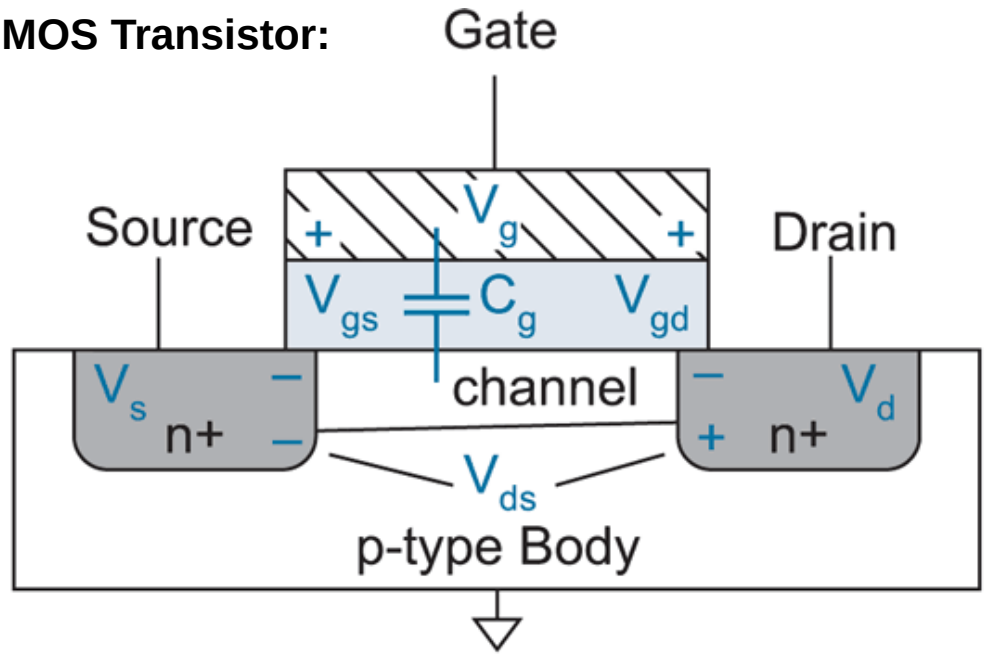
Quadratic Relationship

The saturation voltage V_{DSAT} can be estimated by equating dI_{ds}/dV_{ds} to zero

FIG 2.7 I-V characteristics of ideal nMOS transistor

Channel Charge

NMOS Transistor:



$V_{ds} = V_{gs} - V_{gd}$

Average gate to channel potential:

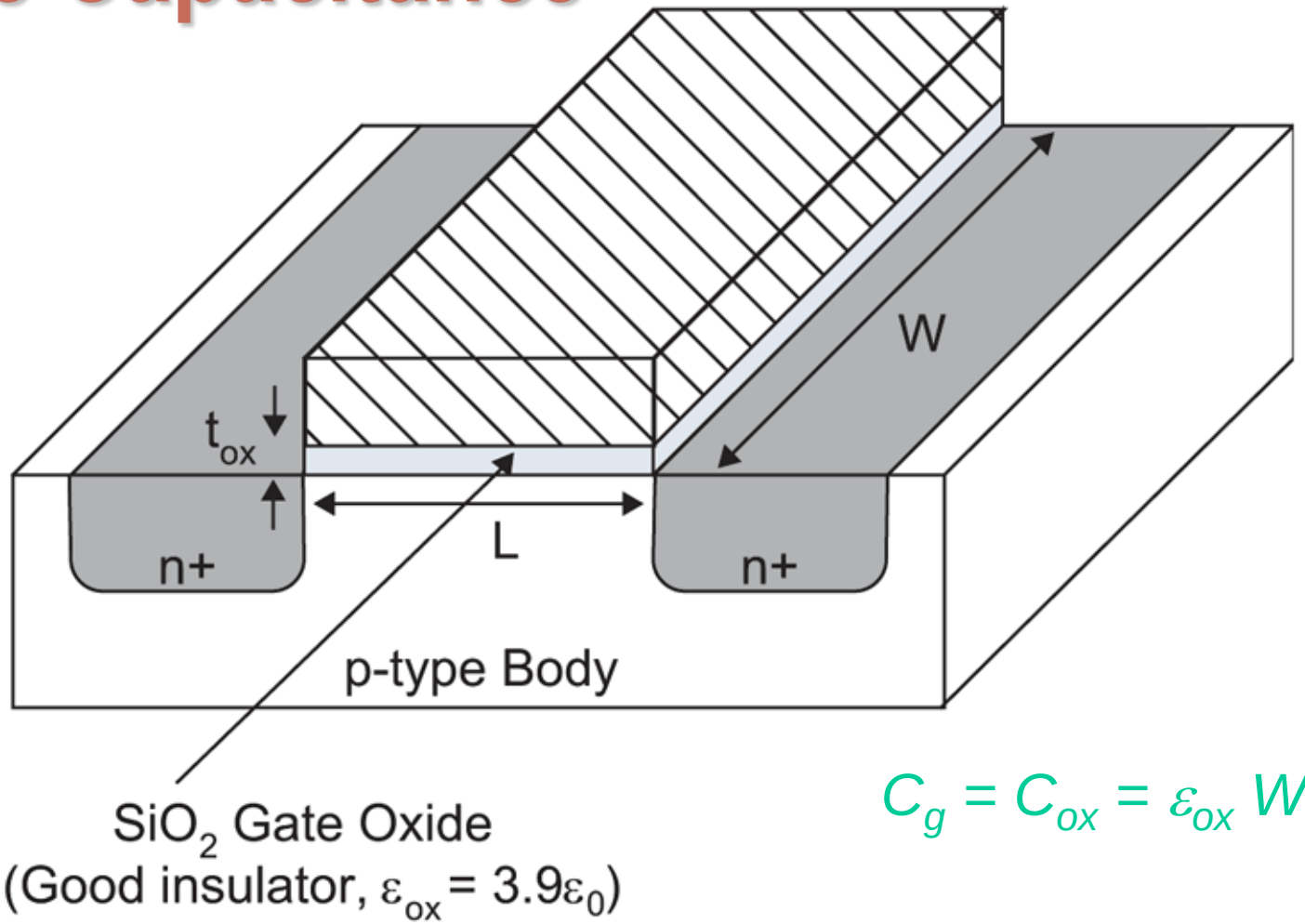
$$V_{gc} = (V_{gs} + V_{gd})/2 = V_{gs} - V_{ds}/2$$

FIG 2.5 Average gate to channel voltage

Note: i) The inversion layer thickness is assumed to be zero: all charges are assumed to be located at the Si surface….like a sheet of charge….
ii) Hence, there is no potential drop or band bending across the inversion layer….

$$Q_{channel} = C_g (V_{gc} - V_t)$$

Gate Capacitance



$$C_g = C_{ox} = \varepsilon_{ox} WL/t_{ox}$$

FIG 2.6 Transistor dimensions

Transistor Currents (NMOS)

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

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

$I_{ds} = W\, Q_{channel}\, .carrier\ velocity(v)$

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

Since Vds is small, Vds/2 can be neglected…and Ids is linearly proportional to Vds….like a resistor

$Q_{channel} = C_g\, (V_{gc} - V_t)$

$V_{gc} = V_{gs} - V_{ds}/2)$

$v = \mu E$

$E_{lateral} = V_{ds}/L$

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

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

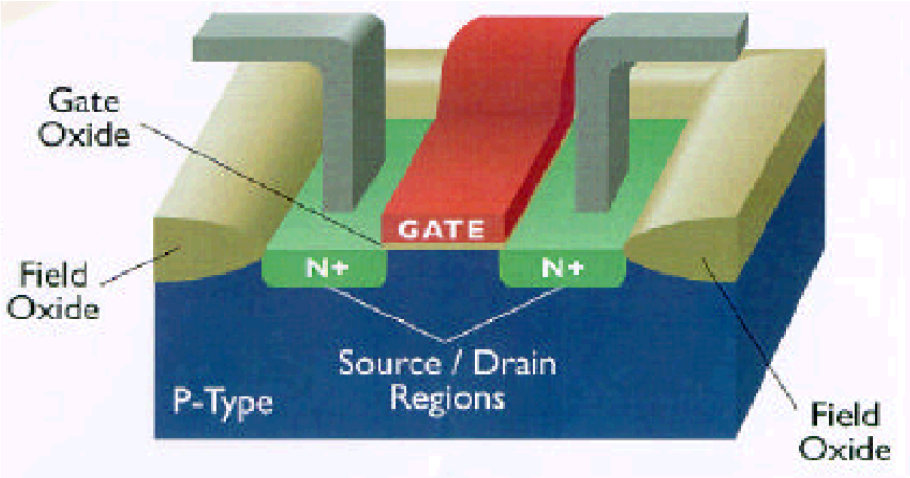
Note: as Vds increases, average $Q_{channel}$ *decreases…*

$dI_{ds}/dV_{ds} = 0$ *at* $V_{ds} = V_{dsat} = V_{gs} - V_t$

Substituting V_{ds} *with* V_{dsat} *above:* $I_{ds} = \beta/2\, (V_{gs} - V_t)^2$

Note: for PMOS $V_{tp} = V_{tn}$ $\qquad$ $\mu_p < \mu_n$, *hence* $(W/L)_{PMOS} \sim 2\, (W/L)_{NMOS}$

Goals of Technology Scaling



- Making things cheaper
  - Integrating more transistors per chip for the same money
  - Build same products cheaper, sell same part for less money
  - Price of fabrication of a transistor to be reduced
- But also want the transistor to be faster, smaller and low power

Technology Scaling

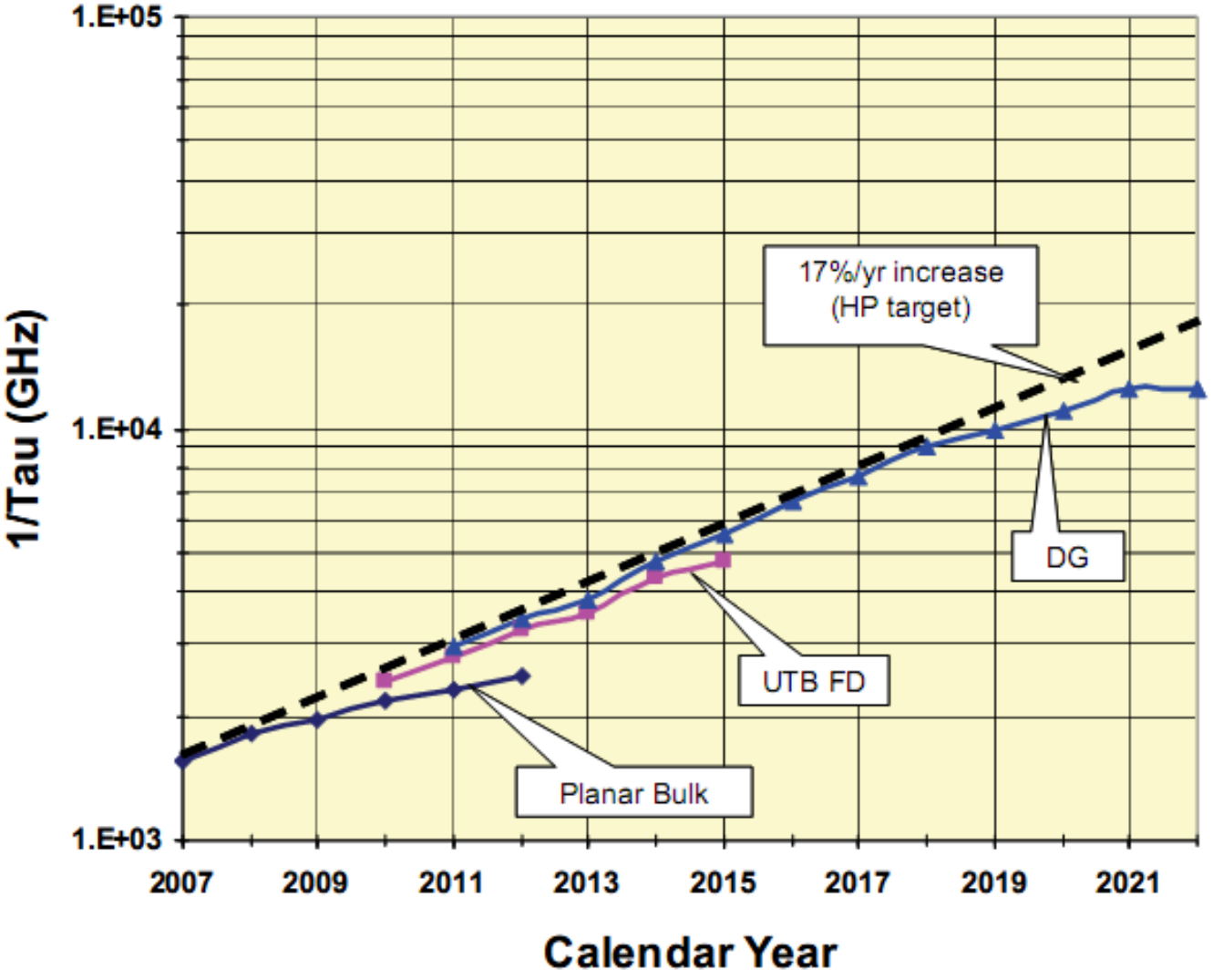
- Technology generation spans 2-3 years
- Benefits of scaling the dimensions by 30%:
  - Reduce gate delay by 30% (increase operating frequency by 43%)
  - Double transistor density
  - Reduce energy per transition by 65% (50% power savings @ 43% increase in frequency
- Die size used to increase by 14% per generation

Technology Roadmap

- International Technology Roadmap for Semiconductors 2002 data

Year	2001	2003	2005	2007	2010	2013	2016
DRAM ½ pitch [nm]	130	100	80	65	45	32	22
MPU transistors/chip	97M	153M	243M	386M	773M	1.55G	3.09G
Wiring levels	8	8	10	10	10	11	11
High-perf. phys. gate [nm]	65	45	32	25	18	13	9
High-perf. VDD [V]	1.2	1.0	0.9	0.7	0.6	0.5	0.4
Local clock [GHz]	1.7	3.1	5.2	6.7	11.5	19.3	28.8
High-perf. power [W]	130	150	170	190	218	251	288
Low-power phys. gate [nm]	90	65	45	32	22	16	11
Low-power VDD [V]	1.2	1.1	1.0	0.9	0.8	0.7	0.6
Low-power power [W]	2.4	2.8	3.2	3.5	3.0	3.0	3.0

Acceleration Continues



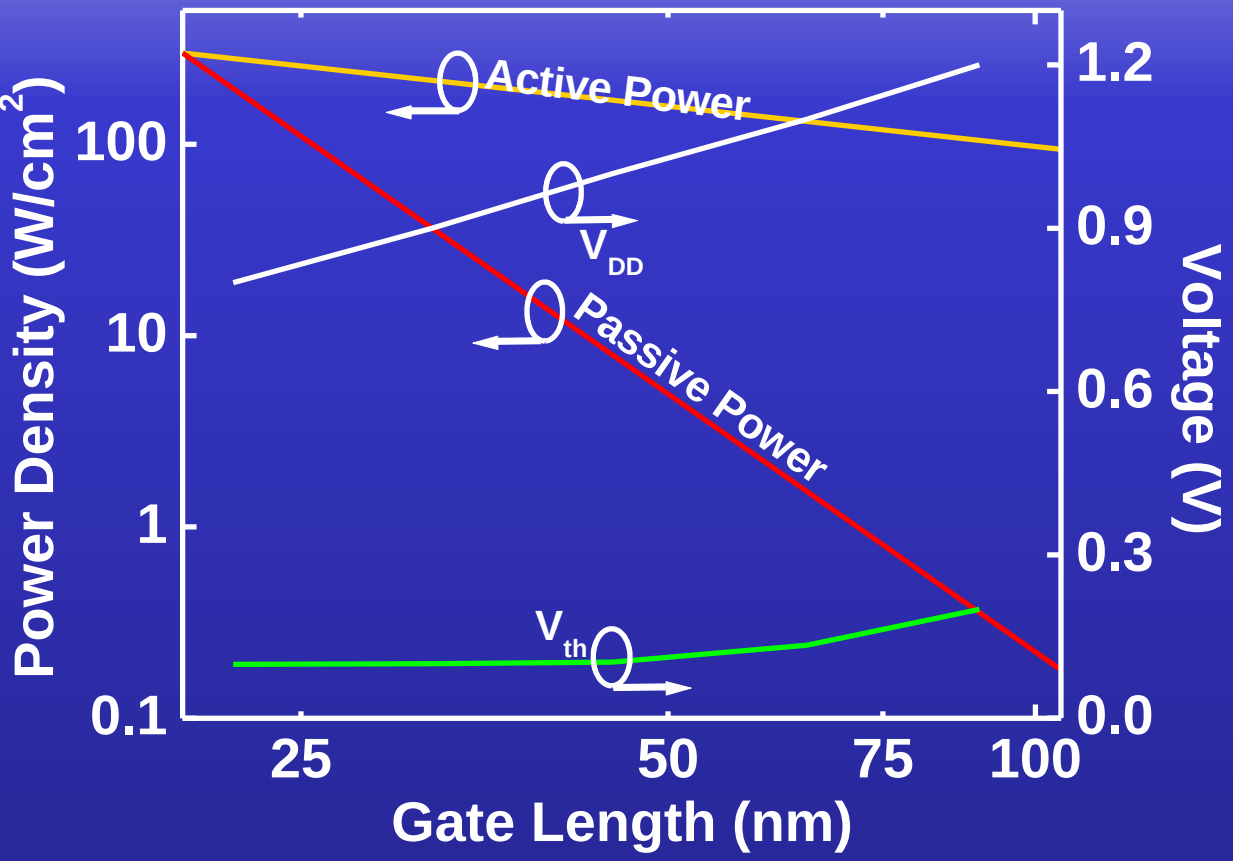
Technology Scaling Models

- Full Scaling (Constant Electrical Field)
  - ideal model, dimensions and voltage scale together by the same factor S
- Fixed Voltage Scaling
  - most common model until recently
  - only dimensions scale, voltages remain constant
- General Scaling
  - most realistic for todays situation
  - voltages and dimensions scale with different factors

Scaling Relationships for Long Channel Devices

Parameter	Relation	Full Scaling	General Scaling	Fixed Voltage Scaling
W, L, t_{ox}		1/S	1/S	1/S
V_{DD}, V_T		1/S	1/U	1
N_{SUB}	V/W_{depl}^2	S	S^2/U	S^2
Area/Device	WL	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S	S
C_L	$C_{ox}WL$	1/S	1/S	1/S
k_n, k_p	$C_{ox}W/L$	S	S	S
I_{av}	$k_{n,p} V^2$	1/S	S/U^2	S
t_p (intrinsic)	$C_L V / I_{av}$	1/S	U/S^2	$1/S^2$
P_{av}	$C_L V^2 / t_p$	$1/S^2$	S/U^3	S
PDP	$C_L V^2$	$1/S^3$	$1/SU^2$	1/S

Gate Length Scaling Trends



Puri *et. al., Nano Lett.,* 2003

Problems Faced

- Threshold voltage to be scaled aggressively
- Reduced Vth results in higher leakage current
- Temperature on the die increases → higher resistance
- Interconnect delay becomes significant
- Soft error rate increases with scaling

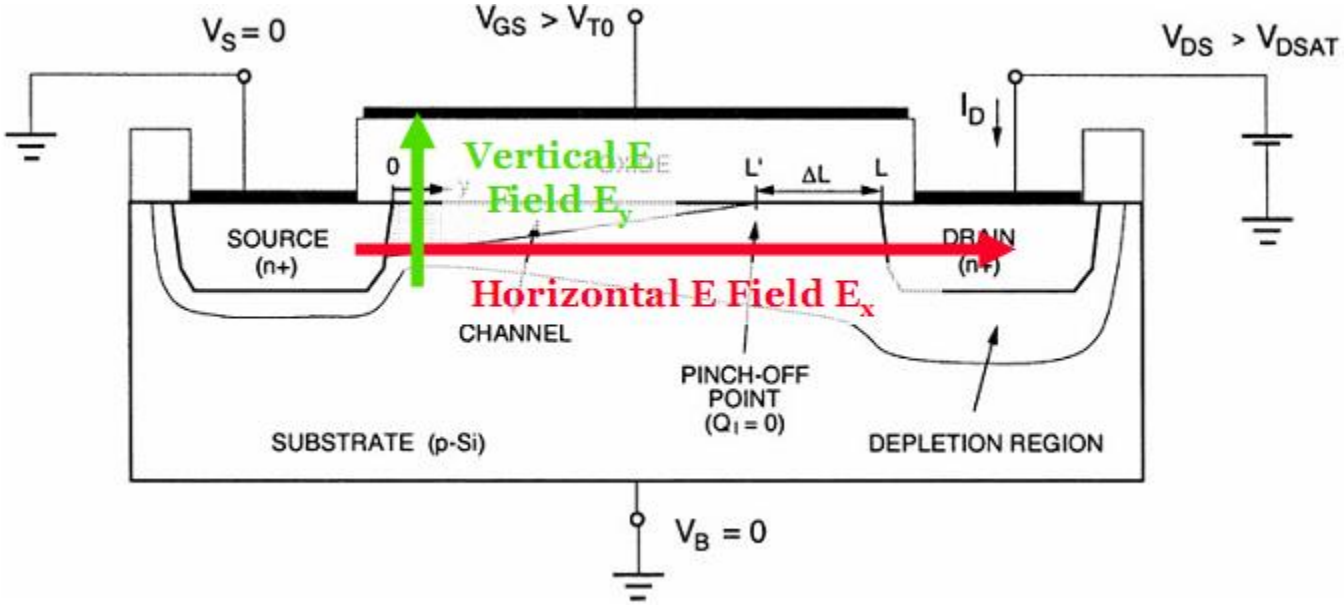
Short Channel Effects

- Short-channel device: channel length is comparable to depth of drain and source junctions and depletion width
  - In general, visible when L ~ 1μm and below
- Short-channel effects
  - Carrier velocity saturation
  - Mobility degradation
  - Threshold voltage variation
    - V_T Roll-Off
    - Drain-Induced Barrier Lowering
  - Oxide breakdown
  - Avalanche breakdown

Short Channel Effects

- Short-channel device: channel length is comparable to depth of drain and source junctions and depletion width
  - In general, visible when L ~ 1μm and below
- Short-channel effects
  - Carrier velocity saturation
  - Mobility degradation
  - Threshold voltage variation
    - V_T Roll-Off
    - Drain-Induced Barrier Lowering
  - Oxide breakdown
  - Avalanche breakdown

Velocity Saturation



Vertical Electric Field

1980 E_Y=5V/1000A° =50 x 10^4V/cm

1995 E_Y=3.3.V/75A°=4.4 x 10^6V/cm

2001 E_y=1.2V/22AO=5.5 x 10^6 V/cm

Horizantal Electric Field

1980 E_x=5V/5µm =10^4V/cm

1995 E_x=3.3.V/0.35µm=9.4 x 10^4V/cm

2001 E_x=1.2V/0.1µm=1.2 x 10^5 V/cm

Velocity Saturation

- As the horizontal E field grows, so does the velocity of the electrons
- As the gate shrinks, the velocity reaches the point where it cannot increase as much with increasing E field
- This is due to velocity saturation
  - Electrons become so energetic that they scatter (crash into the crystal lattice)