

CMOS VLSI Design

Lecture 02

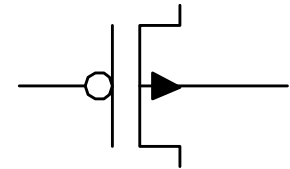
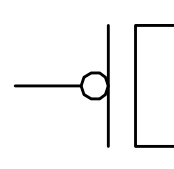
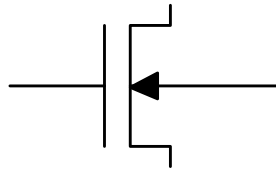
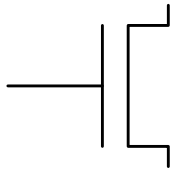
Sonoma State University

Outline

- Introduction
- MOS Capacitor
- nMOS I-V Characteristics
- pMOS I-V Characteristics
- Gate and Diffusion Capacitance
- Secondary effects
 - Carrier velocity saturation
 - Mobility degradation
 - Threshold voltage variation
 - Subthreshold conduction

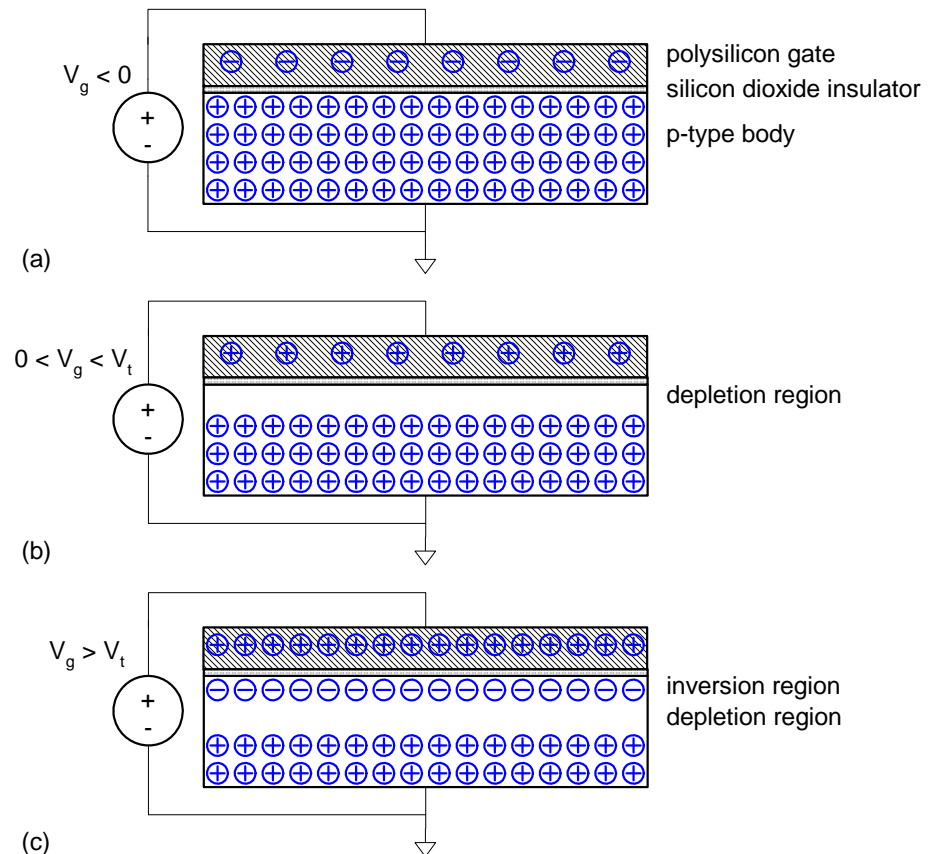
Introduction

- So far, we have treated transistors as ideal switches
- An ON transistor passes a finite amount of current
 - Depends on terminal voltages
 - Derive current-voltage (I-V) relationships
- Transistor gate, source, drain all have capacitance
 - $I = C (\Delta V / \Delta t) \rightarrow \Delta t = (C/I) \Delta V$
 - Capacitance and current determine speed
- Also explore what a “degraded level” really means



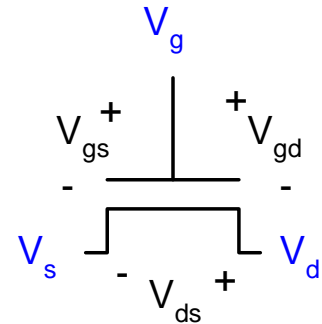
MOS Capacitor

- Gate and body form MOS capacitor
- Operating modes
 - Accumulation
 - Depletion
 - Inversion



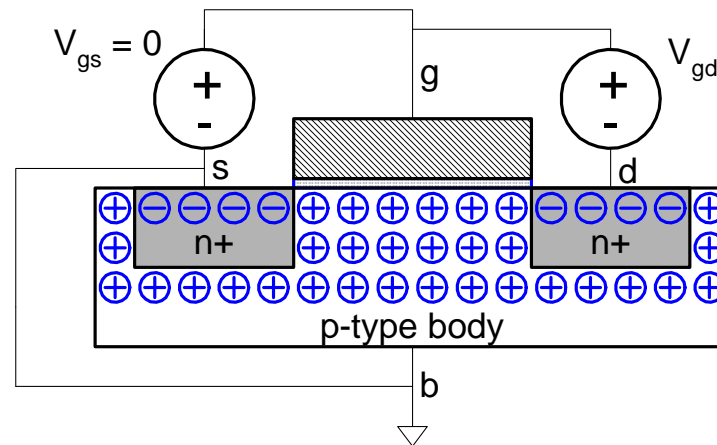
Terminal Voltages

- Mode of operation depends on V_g , V_d , V_s
 - $V_{gs} = V_g - V_s$
 - $V_{gd} = V_g - V_d$
 - $V_{ds} = V_d - V_s = V_{gs} - V_{gd}$
- Source and drain are symmetric diffusion terminals
 - By convention, source is terminal at lower voltage
 - Hence $V_{ds} \geq 0$
- nMOS body is grounded. First assume source is 0 too.
- Three regions of operation
 - *Cutoff*
 - *Linear*
 - *Saturation*



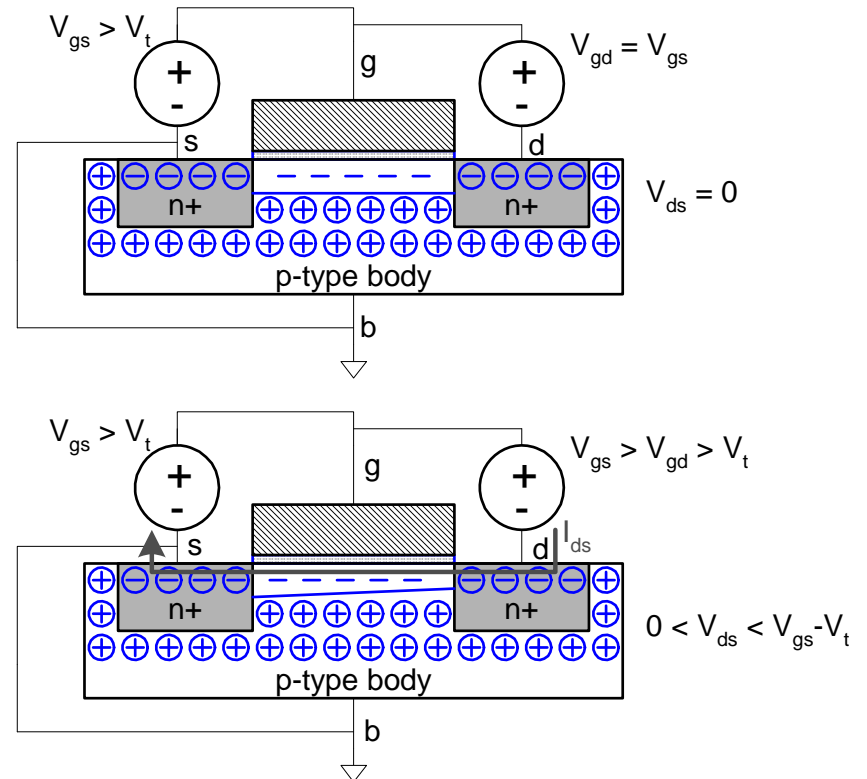
nMOS Cutoff

- No channel
- $I_{ds} = 0$



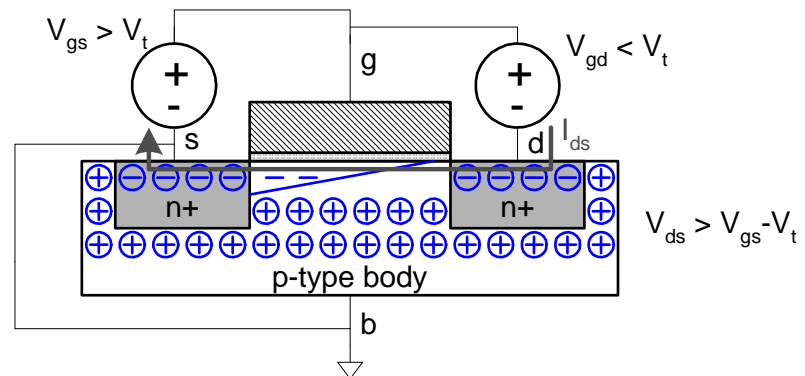
nMOS Linear

- Channel forms
- Current flows from d to s
 - e^- from s to d
- I_{ds} increases with V_{ds}
- Similar to linear resistor



nMOS Saturation

- Channel pinches off
- I_{ds} independent of V_{ds}
- We say current saturates
- Similar to current source

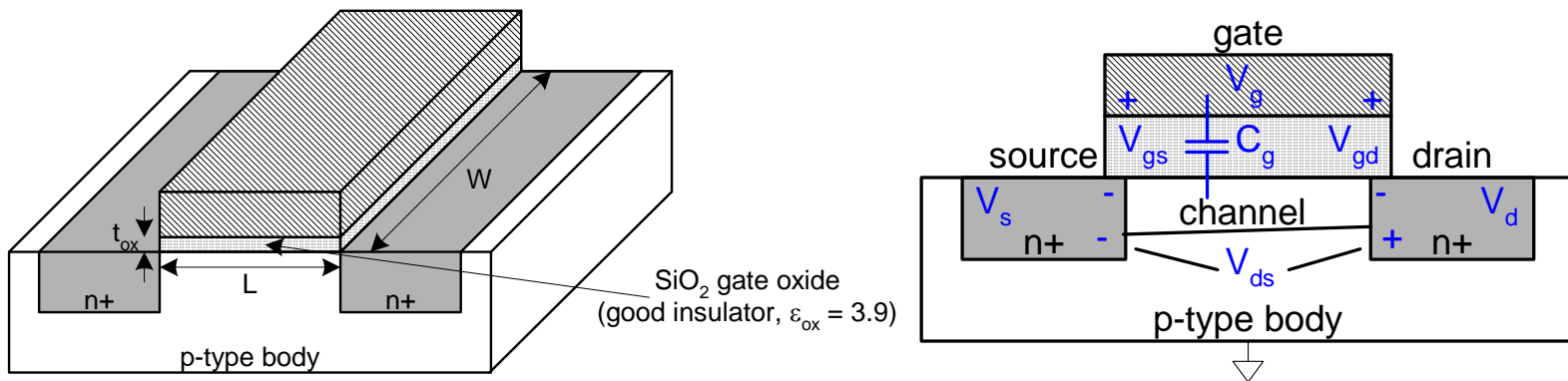


I-V Characteristics

- In Linear region, I_{ds} depends on
 - How much charge is in the channel?
 - How fast is the charge moving?

Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
 - Gate – oxide – channel
- $Q_{\text{channel}} =$

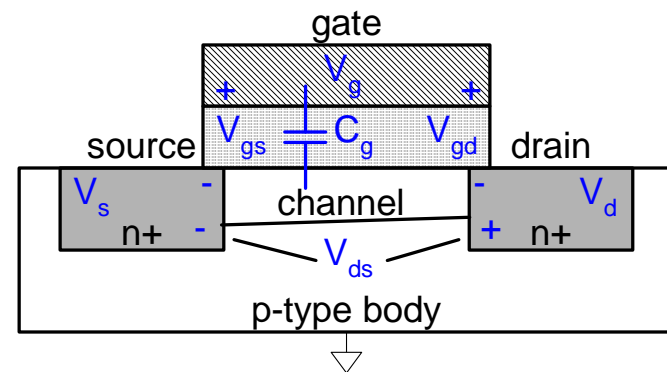
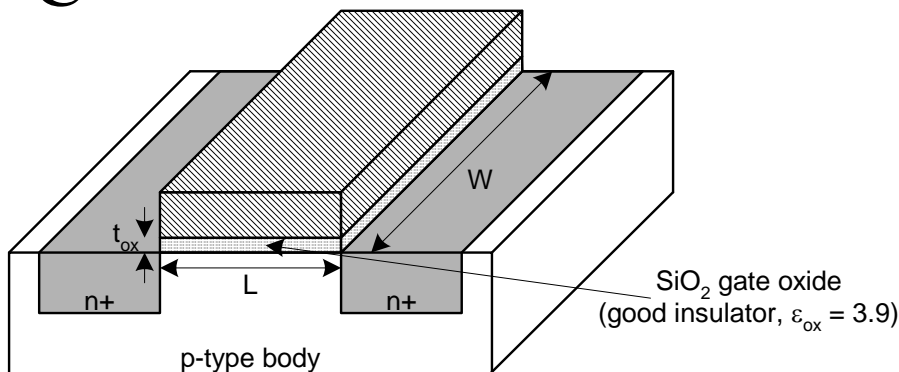


Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
 - Gate – oxide – channel

- $Q_{\text{channel}} = CV$

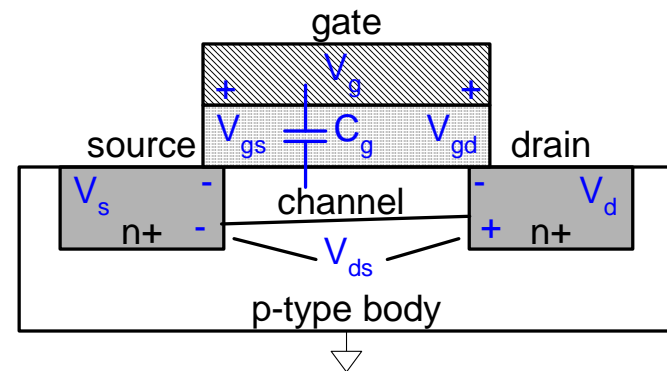
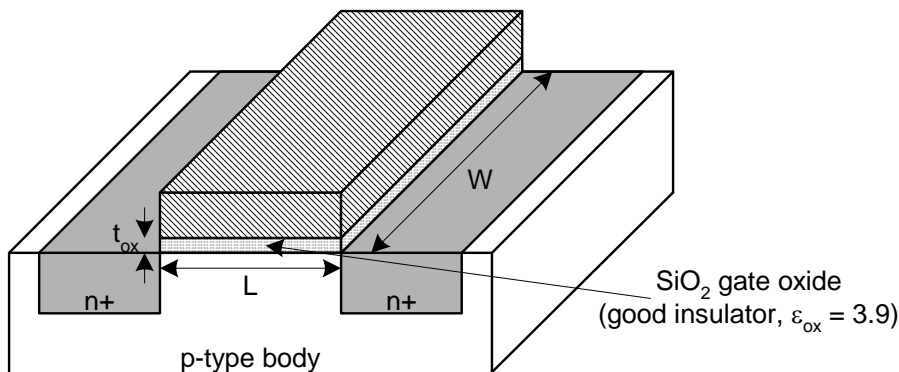
- $C =$



Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
 - Gate – oxide – channel
- $Q_{\text{channel}} = CV$
- $C = C_g = \epsilon_{\text{ox}} WL/t_{\text{ox}} = C_{\text{ox}} WL$
- $V =$

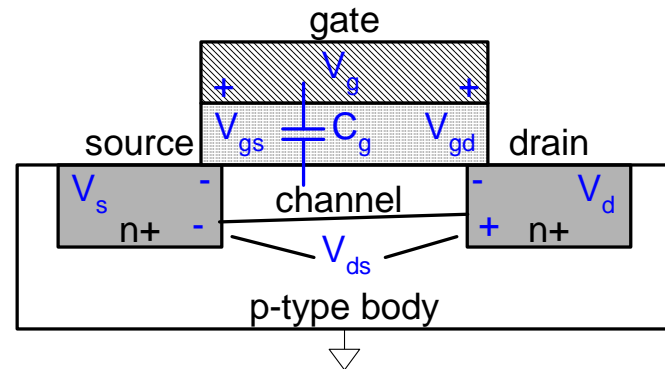
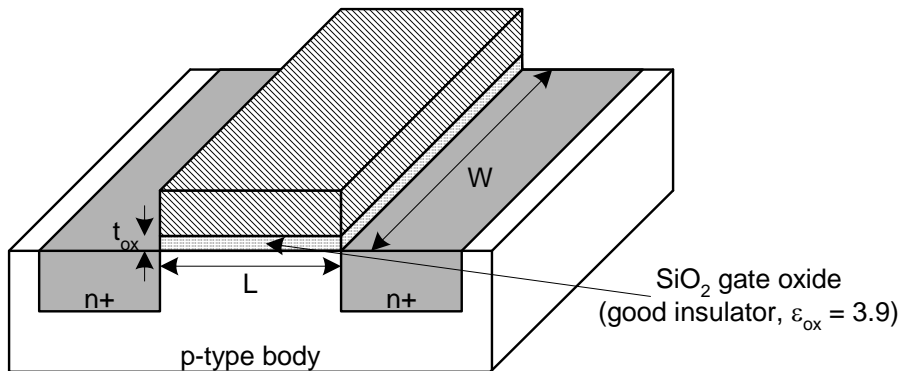
$$C_{\text{ox}} = \epsilon_{\text{ox}} / t_{\text{ox}}$$



Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
 - Gate – oxide – channel
- $Q_{\text{channel}} = CV$
- $C = C_g = \epsilon_{\text{ox}} WL / t_{\text{ox}} = C_{\text{ox}} WL$
- $V = V_{\text{gc}} - V_t = (V_{\text{gs}} - V_{\text{ds}}/2) - V_t$

$$C_{\text{ox}} = \epsilon_{\text{ox}} / t_{\text{ox}}$$



Carrier velocity

- Charge is carried by e-
- Carrier velocity v proportional to lateral E-field between source and drain
- $v =$

Carrier velocity

- Charge is carried by e-
- Carrier velocity v proportional to lateral E-field between source and drain
- $v = \mu E$ μ called mobility
- $E = V_{ds}/L$
- Time for carrier to cross channel:
 - $t =$

Carrier velocity

- Charge is carried by e-
- Carrier velocity v proportional to lateral E-field between source and drain
- $v = \mu E$ μ called mobility
- $E = V_{ds}/L$
- Time for carrier to cross channel:
 - $t = L / v = L/(\mu V_{ds}/L) = L^2/(\mu V_{ds})$

nMOS Linear I-V

- Now we know
 - How much charge Q_{channel} is in the channel
 - How much time t each carrier takes to cross

$$I_{ds} =$$

nMOS Linear I-V

- Now we know
 - How much charge Q_{channel} is in the channel
 - How much time t each carrier takes to cross

$$I_{ds} = \frac{Q_{\text{channel}}}{t}$$
$$=$$

nMOS Linear I-V

- Now we know
 - How much charge Q_{channel} is in the channel
 - How much time t each carrier takes to cross

$$I_{ds} = \frac{Q_{\text{channel}}}{t}$$

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

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

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

nMOS Saturation I-V

- If $V_{gd} < V_t$, channel pinches off near drain
 - When $V_{ds} > V_{dsat} = V_{gs} - V_t$
- Now drain voltage no longer increases current

$$I_{ds} =$$

nMOS Saturation I-V

- If $V_{gd} < V_t$, channel pinches off near drain
 - When $V_{ds} > V_{dsat} = V_{gs} - V_t$
- Now drain voltage no longer increases current

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

nMOS Saturation I-V

- If $V_{gd} < V_t$, channel pinches off near drain
 - When $V_{ds} > V_{dsat} = V_{gs} - V_t$
- Now drain voltage no longer increases current

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

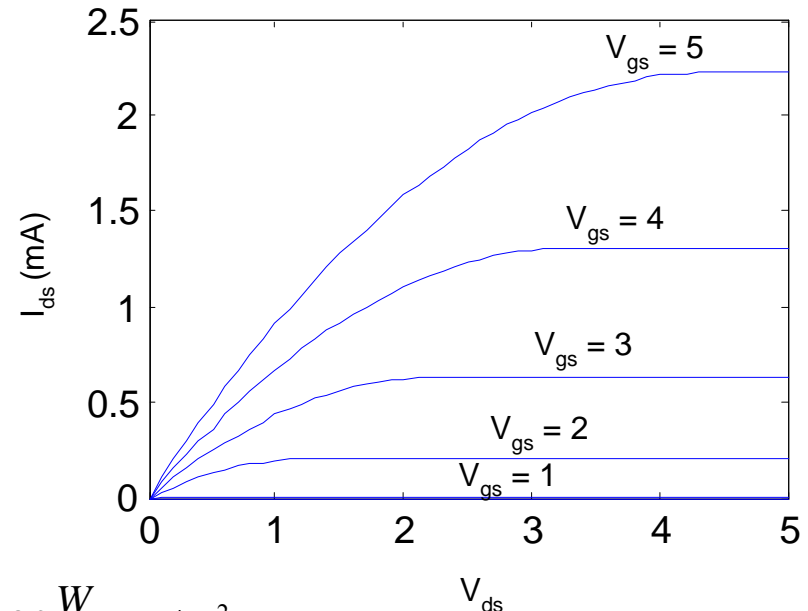
nMOS I-V Summary

- *Shockley* 1st order transistor models

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t & \text{cutoff} \\ \beta \left(V_{gs} - V_t - \frac{V_{ds}}{2} \right) V_{ds} & V_{ds} < V_{dsat} & \text{linear} \\ \frac{\beta}{2} (V_{gs} - V_t)^2 & V_{ds} > V_{dsat} & \text{saturation} \end{cases}$$

Example

- Consider a 0.6 μm process
 - From AMI Semiconductor
 - $t_{\text{ox}} = 100 \text{ \AA}$
 - $\mu = 350 \text{ cm}^2/\text{V}\cdot\text{s}$
 - $V_t = 0.7 \text{ V}$
- Plot I_{ds} vs. V_{ds}
 - $V_{\text{gs}} = 0, 1, 2, 3, 4, 5$
 - Use $W/L = 4/2 \lambda$



$$\beta = \mu C_{\text{ox}} \frac{W}{L} = (350) \left(\frac{3.9 \cdot 8.85 \cdot 10^{-14}}{100 \cdot 10^{-8}} \right) \left(\frac{W}{L} \right) = 120 \frac{W}{L} \mu\text{A}/\text{V}^2$$

pMOS I-V

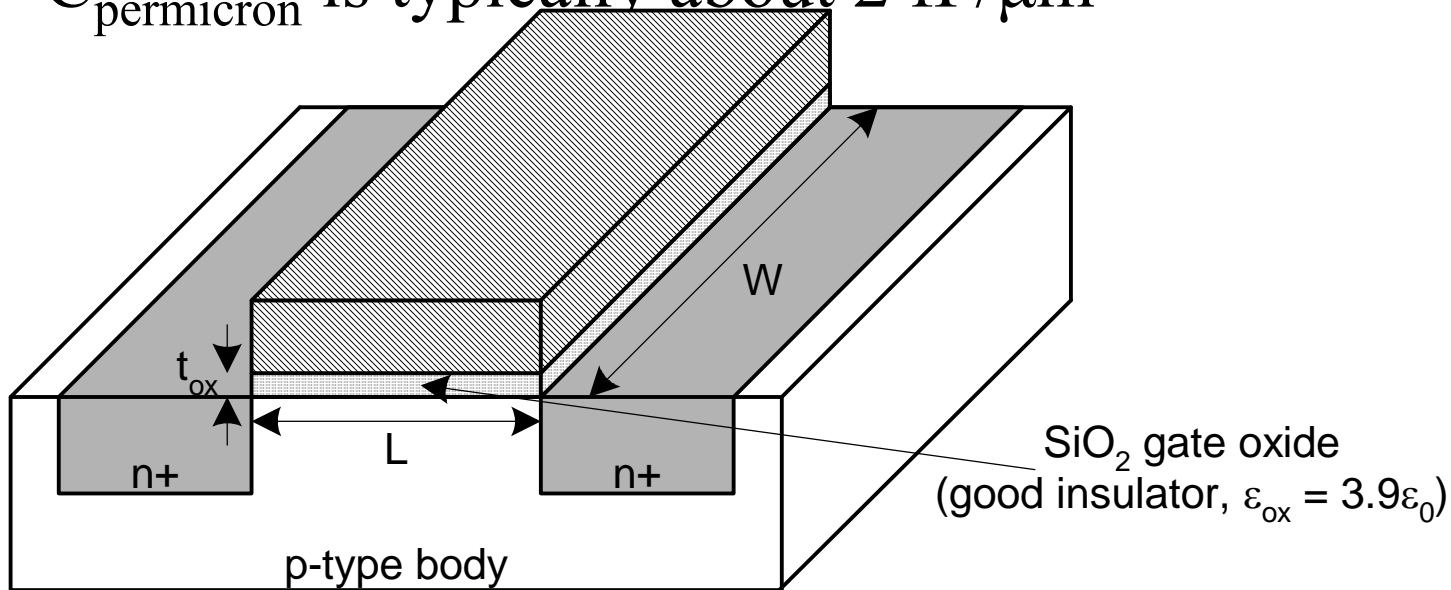
- All dopings and voltages are inverted for pMOS
- Mobility μ_p is determined by holes
 - Typically 2-3x lower than that of electrons μ_n
 - 120 cm²/V*s in AMI 0.6 μ m process
- Thus pMOS must be wider to provide same current
 - In this class, assume $\mu_n / \mu_p = 2$
 - *** plot I-V here

Capacitance

- Any two conductors separated by an insulator have capacitance
- Gate to channel capacitor is very important
 - Creates channel charge necessary for operation
- Source and drain have capacitance to body
 - Across reverse-biased diodes
 - Called diffusion capacitance because it is associated with source/drain diffusion

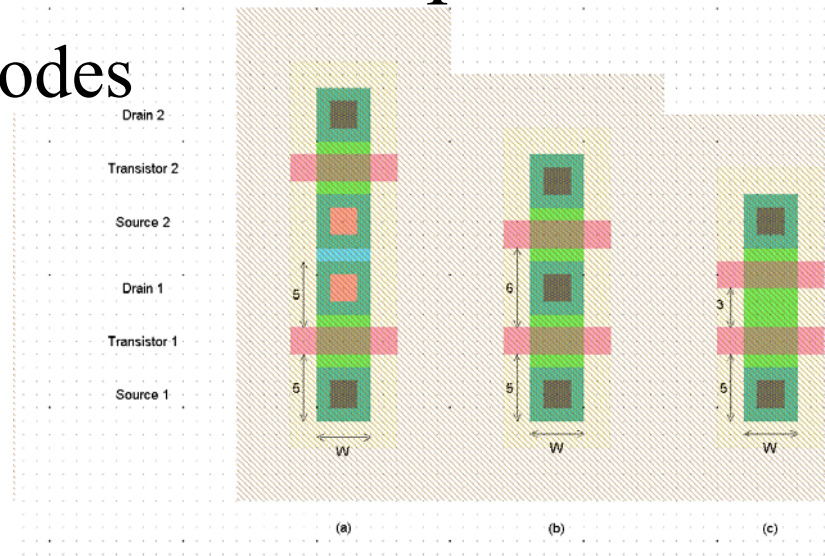
Gate Capacitance

- Approximate channel as connected to source
- $C_{gs} = \epsilon_{ox} WL/t_{ox} = C_{ox} WL = C_{permicron} W$
- $C_{permicron}$ is typically about 2 fF/ μm



Diffusion Capacitance

- C_{sb} , C_{db}
- Undesirable, called *parasitic* capacitance
- Capacitance depends on area and perimeter
 - Use small diffusion nodes
 - Comparable to C_g for contacted diff
 - $\frac{1}{2} C_g$ for uncontacted
 - Varies with process



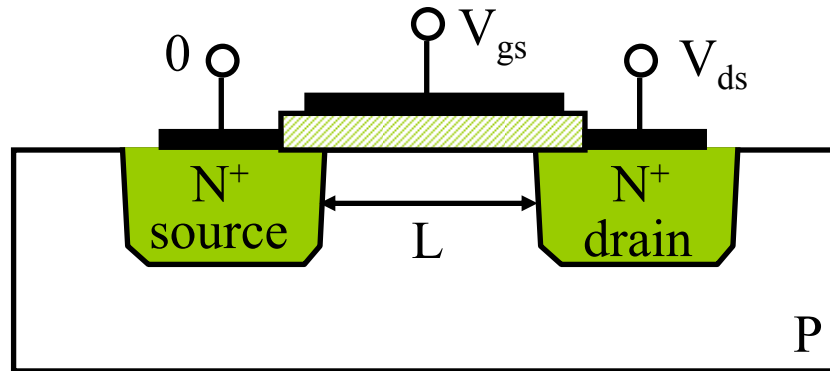
Secondary effects

- **Short-channel effects:**
 - Short channel device has channel length comparable to depth of drain and source junctions and depletion width
 - Causes threshold voltage and I/V curve variations
- **Narrow-channel effects:**
 - Narrow channel device has small channel width
- **Subthreshold conduction** (leakage current)

Secondary effects

- **Short-channel device**: channel length is comparable to depth of drain and source junctions and depletion width
 - In general, effects are visible when $L \sim 1\mu\text{m}$ and below
- **Short-channel effects**:
 - Carrier velocity saturation
 - Mobility degradation
 - Threshold voltage variation

Carrier velocity saturation



- Electric field E_y exists along channel
 - As channel length is reduced, electric field increases (if voltage is constant)
- Electron drift velocity v_d is proportional to electric field
 - only for small field values
 - for large electric field, velocity saturates

Effects of High fields

- **Vertical field** - The vertical field occurs in the y-direction from the gate to the channel
($E_Y = V_{DD}/t_{ox}$)
- **Horizontal field** - The horizontal field occurs in the x-direction from the drain to the source
($E_X = V_{DS}/L$)

Carrier velocity saturation

- Effect of velocity saturation:

- Current saturates before “saturation region”
- V_{DSAT} = voltage at which saturation occurs
- Drain current is reduced:

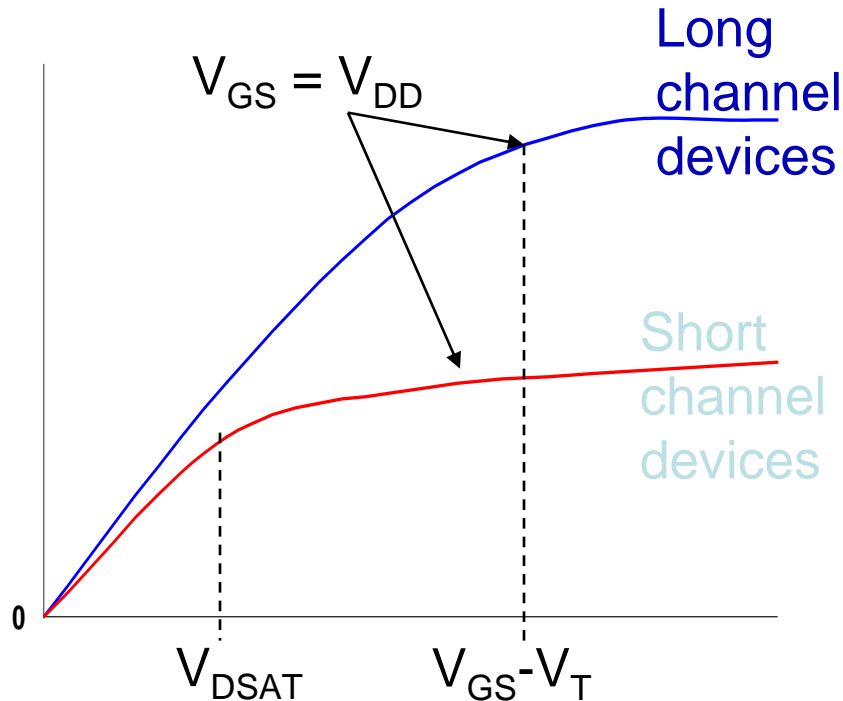
$$I_D(sat) = Wv_d(sat)C_{ox}(V_{GS} - V_T - \frac{1}{2}V_{DSAT})$$

(no longer quadratic function of V_{GS})

- Saturation region is extended:

$$V_{DSAT} < V_{GS} - V_T$$

Velocity Saturation Effects

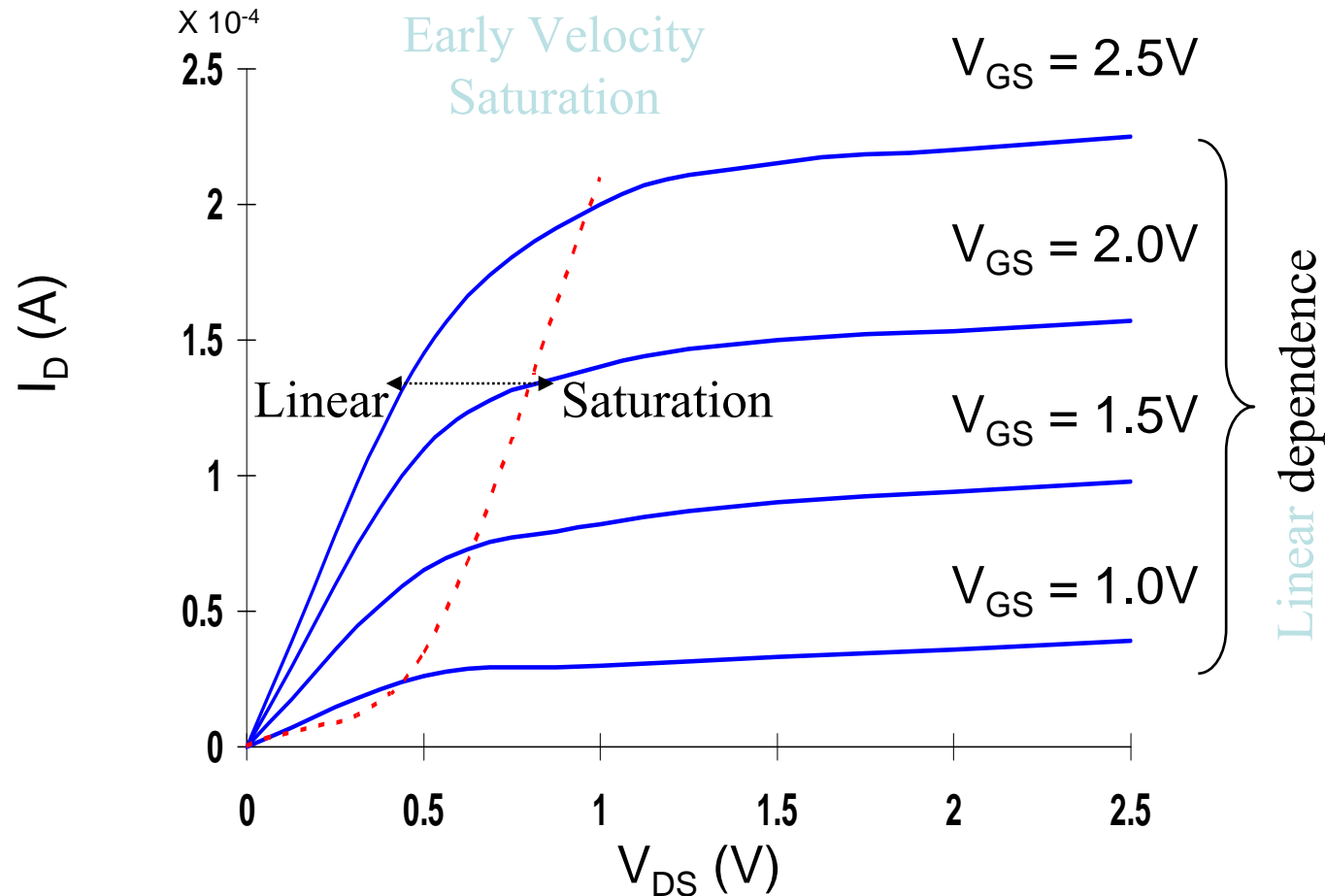


For short channel devices and large enough $V_{GS} - V_T$

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

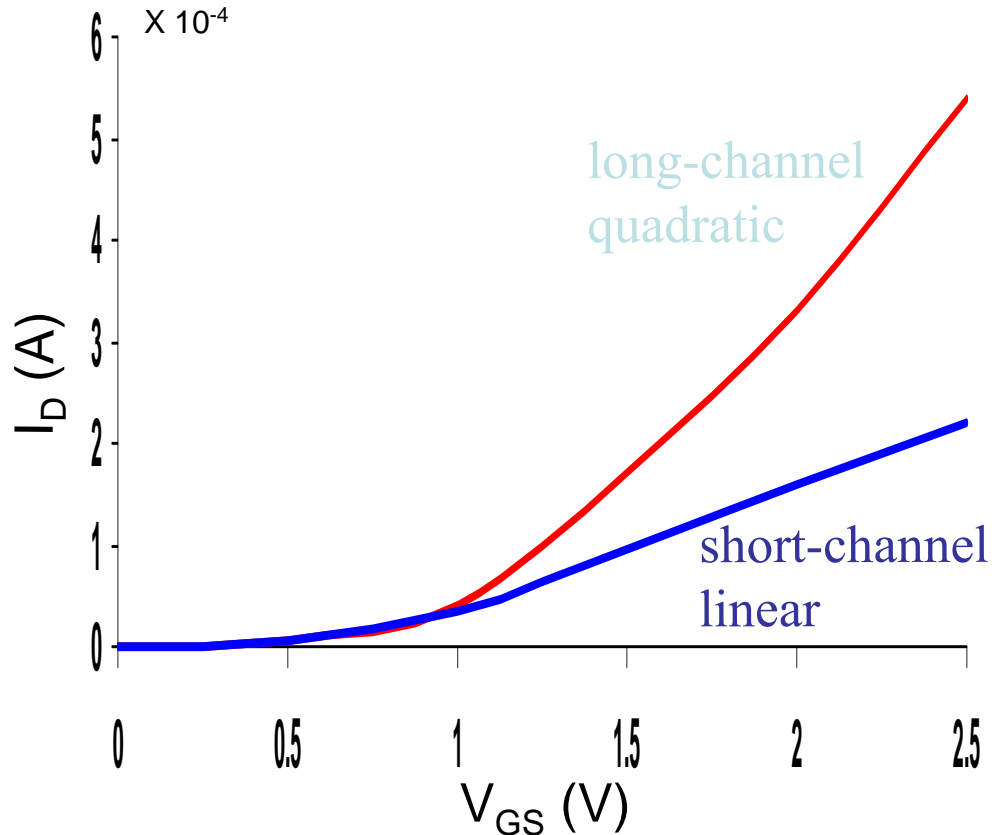
- I_{DSAT} has a linear dependence wrt V_{GS} so a reduced amount of current is delivered for a given control voltage

Short Channel I-V Plot (NMOS)



NMOS transistor, $0.25\mu\text{m}$, $L_d = 0.25\mu\text{m}$, $W/L = 1.5$, $V_{DD} = 2.5\text{V}$, $V_T = 0.4\text{V}$

MOS I_D - V_{GS} Characteristics



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

- 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

Mobility degradation

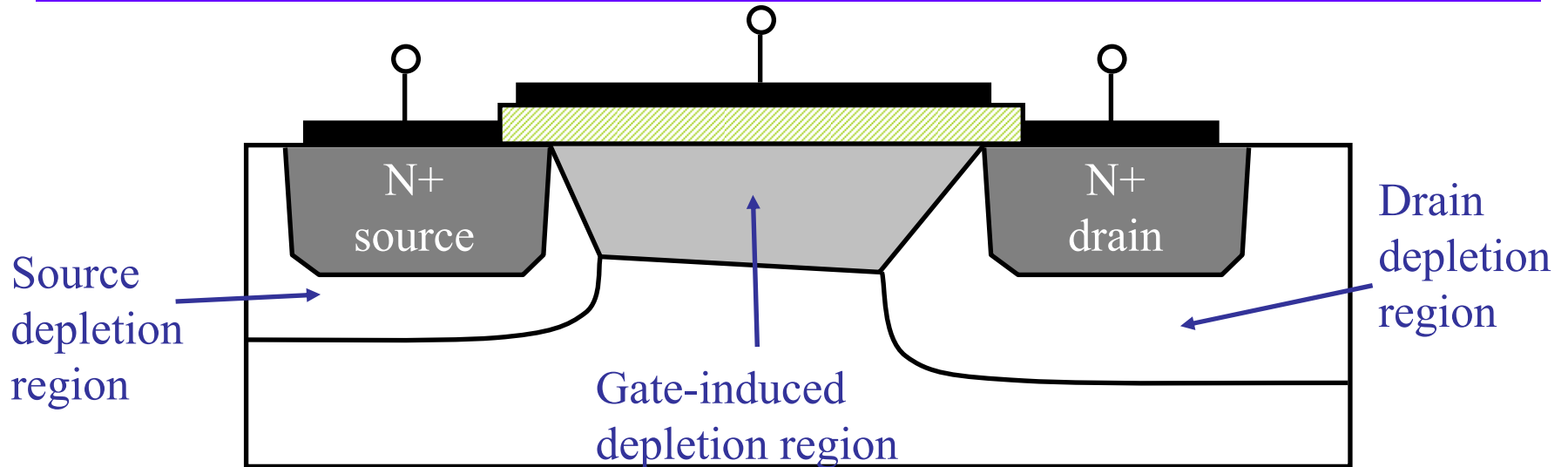
- MOS I/V equations depend on surface mobility μ_n (or μ_p)
- In short-channel devices, μ_n and μ_p are not constant
 - As vertical electric field E_Y increases, surface mobility decreases
 - μ_0 = low-field mobility, η is empirical constant
 - As V_{GS} increases, surface mobility decreases

$$\mu = \frac{\mu_0}{1 + \eta(V_{GS} - V_T)}$$

Threshold voltage variation

- Until now, threshold voltage assumed constant
 - V_T changed only by substrate bias V_{SB}
- In threshold voltage equations, channel depletion region assumed to be created by gate voltage only
- Depletion regions around source and drain neglected: **valid if channel length is much larger than depletion region depths**
- In short-channel devices, depletion regions from drain and source extend into channel

V_T Roll Off



- Even with $V_{GS}=0$, part of channel is already depleted
- Bulk depletion charge is smaller in short-channel device $\rightarrow V_T$ is smaller

Threshold voltage variation

Short-channel effects cause threshold voltage variation:

- V_T roll off
 - As channel length L decreases, threshold voltage decreases
- Drain-induced barrier lowering
 - As drain voltage V_{DS} increases, threshold voltage decreases
- Hot-carrier effect
 - Threshold voltages drift over time

Drain-induced barrier lowering (DIBL)

- Drain-induced barrier lowering (DIBL)
 - Drain voltage V_{DS} causes change in threshold voltage
 - As V_{DS} is increased, threshold voltage decreases
- Cause: depletion region around drain
 - Depletion region depth around drain depends on drain voltage
 - As V_{DS} is increased, drain depletion region gets deeper and extends further into channel
 - For very large V_{DS} , source and drain depletion regions can meet \rightarrow *punch-through!*
- Issue: results in uncertainty in circuit design

Hot Carrier Effects

- Hot-carrier effect
 - increased electric fields causes increased electron velocity
 - high-energy electrons can tunnel into gate oxide
 - This changes the threshold voltage (increases V_T for NMOS)
 - Can lead to long-term reliability problems

Impact Ionization

- Hot electrons
 - High-velocity electrons can also impact the drain, dislodging holes
 - Holes are swept towards negatively-charged substrate → cause substrate current-
 - Called *impact ionization*
 - This is another factor which limits the process scaling → voltage must scale down as length scales

Threshold voltage variation

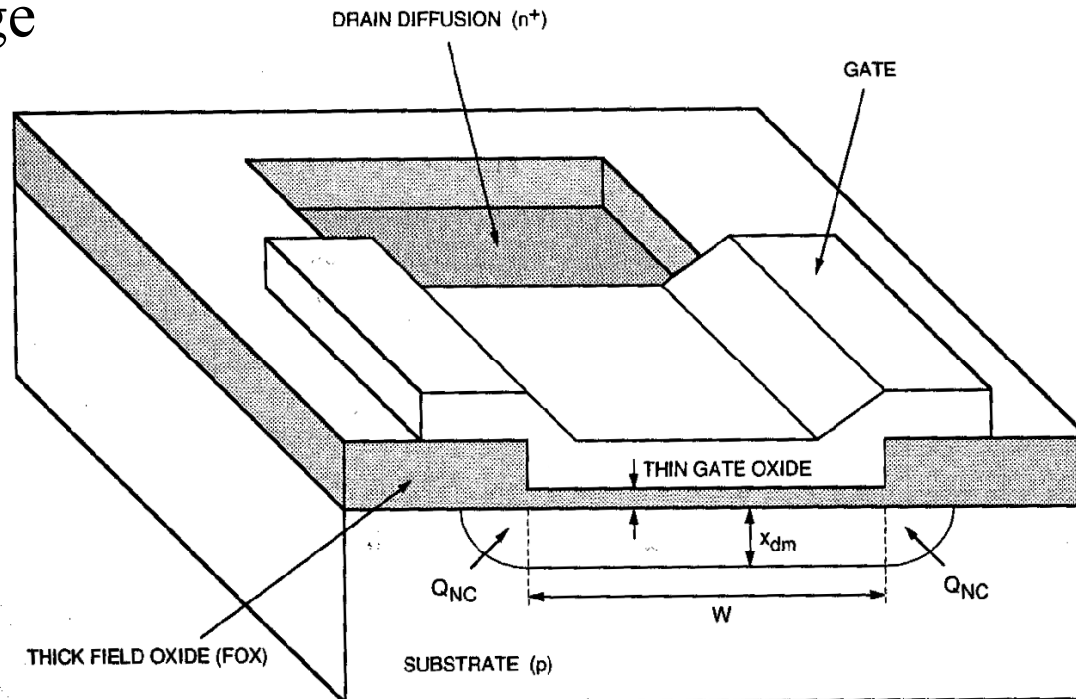
- Summary of threshold variations in short-channel devices
 - V_T roll off: threshold voltage reduces as channel length L reduces
 - **DIBL**: threshold voltage reduces as V_{DS} increases
 - **Hot-carrier effect**: threshold voltage drifts over time as electrons tunnel into oxide

Narrow-channel effects

- **Narrow-channel device:**
 - Channel width W is comparable to maximum depletion region thickness x_{dm}
- **Narrow-channel effect:**
 - Threshold voltage of narrow-channel device is *larger* than threshold of conventional device

Narrow-channel effect

- Cause of narrow-channel effect
 - Edges of gate metal are over field oxide (FOX)
 - This field oxide causes a small depletion region
 - Gate voltage must support this additional depletion region charge



Narrow-channel effect

- Change in threshold voltage:

$$V_{T0}(\text{narrow channel}) = V_{T0} + \Delta V_{T0}$$

$$\Delta V_{T0} = \frac{1}{C_{ox}} \sqrt{2q\epsilon_{Si}N_A|2\phi_F|} \bullet \frac{\kappa x_{dm}}{W}$$

- κ is empirical parameter: depends on shape of the **fringe depletion region**
- Change in V_{T0} proportional to (x_{dm}/W)

Subthreshold conduction

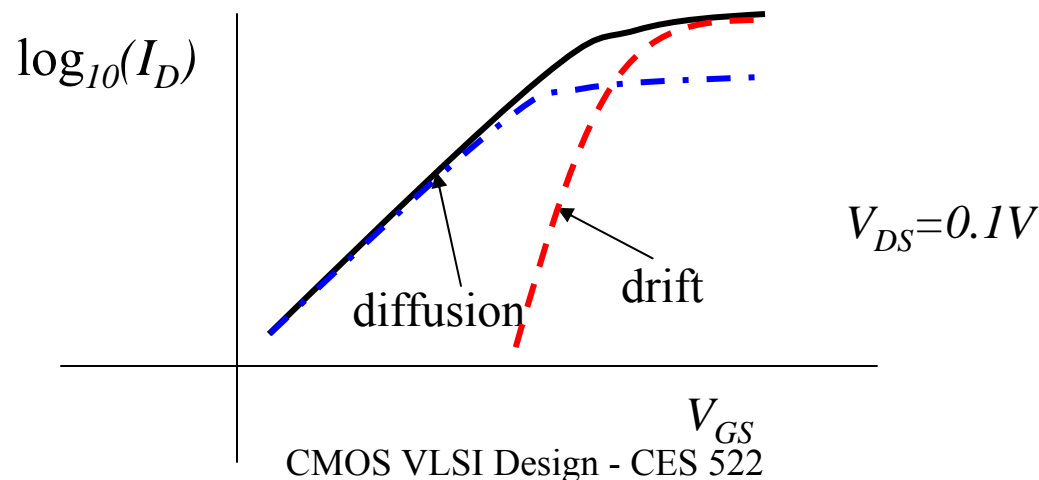
- When $V_{GS} < V_T$, transistor is “off”
 - However, small drain current I_D still flows
 - Called *subthreshold leakage* current
- Model for subthreshold current:

$$I_D(\textit{subthreshold}) = I_S W e^{\frac{q}{kT}(AV_{GS} + BV_{DS})}$$

- Increases as V_{GS} increases (potential barrier lowered)
- Increases as V_{DS} increases (DIBL)

Subthreshold Channel Conduction: Physical Origin

- Subthreshold current conduction is mainly caused by carrier diffusion, while above-threshold is mostly carrier drift.
- This transport mechanism is actually similar to BJT, and the channel current has an exponential dependence on V_{GS} .
- The slope of $\log_{10}(I_D)$ vs. V_{GS} , or required V_{GS} to reduce I_D for one decade, is called the subthreshold slope S , which is larger than 60mV for classical devices.



Leakage current (subthreshold)

- Effect of leakage current
 - “Wasted” power: power consumed even when circuit is inactive
 - Leakage power raises temperature of chip
 - Can cause functionality problem in some circuits: memory, dynamic logic, etc.
- Reducing transistor leakage
 - Long-channel devices
 - Small drain voltage
 - Large threshold voltage V_T

Leakage current issues

- Leakage vs. performance trade-off:
 - For high-speed, need small V_T and L
 - For low leakage, need high V_T and large L (to reduce DIBL and V_T roll-off)
- Process scaling
 - V_T reduces with each new generation technology (historically)
 - Leakage increases $\sim 10X!$
- One solution: dual- V_T process
 - Low- V_T transistors: use in critical paths for high speed
 - High- V_T transistors: use to reduce power