

# Chapter 2

# MOS Transistor Theory

Jin-Fu Li

Advanced Reliable Systems (ARES) Lab.  
Department of Electrical Engineering  
National Central University  
Jhongli, Taiwan

# Outline

- Introduction
- I-V Characteristics of MOS Transistors
- Nonideal I-V Effects
- Pass Transistor
- Summary

# MOS Transistor

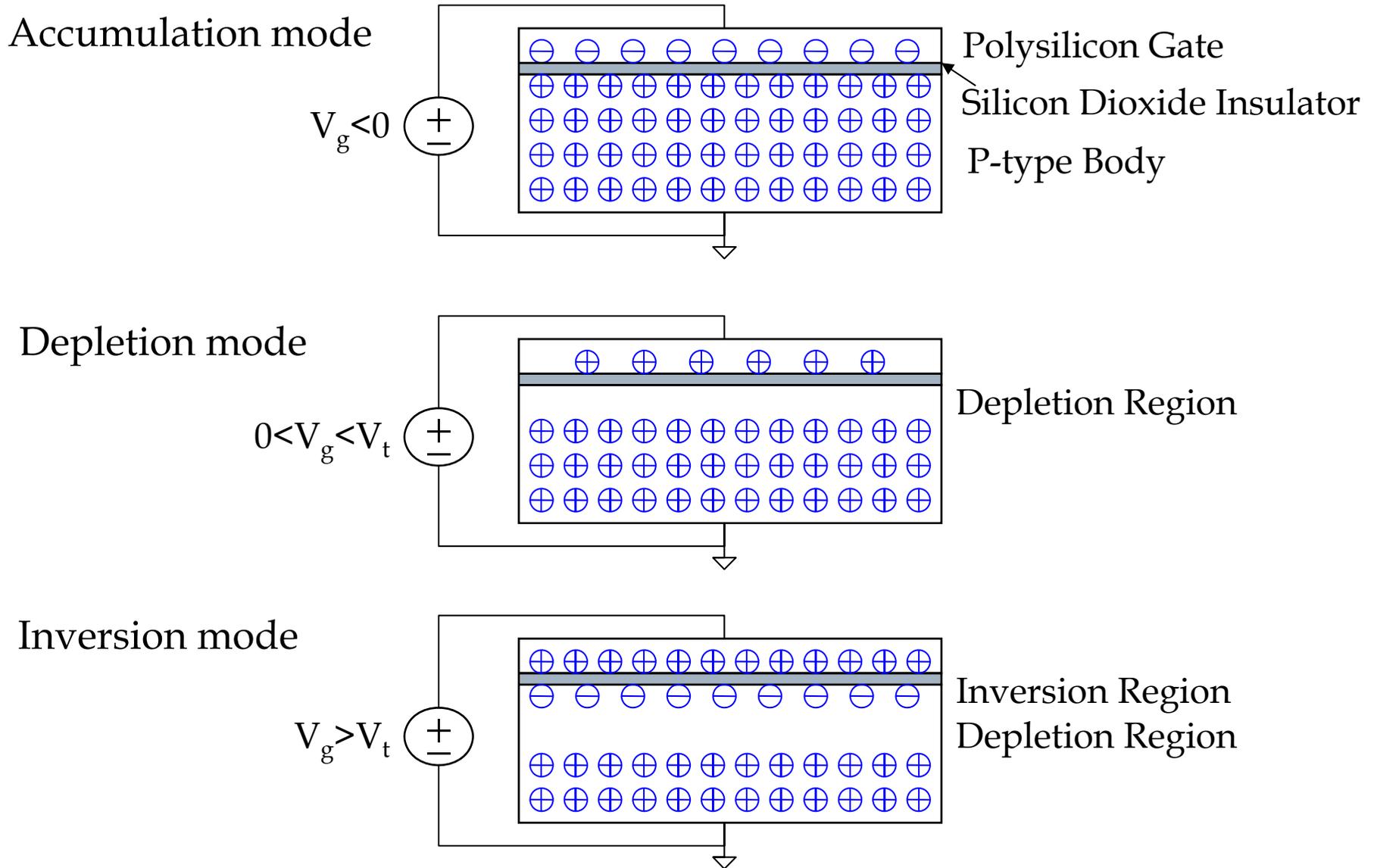
- MOS transistors conduct electrical current by using an applied voltage to move charge from the *source* side to the *drain* side of the device
- An MOS transistor is a *majority-carrier* device
- In an *n-type* MOS transistor, the majority carriers are *electrons*
- In a *p-type* MOS transistor, the majority carriers are *holes*
- Threshold voltage
  - It is defined as the voltage at which an MOS device begins to conduct ("turn on")
- MOS transistor symbols



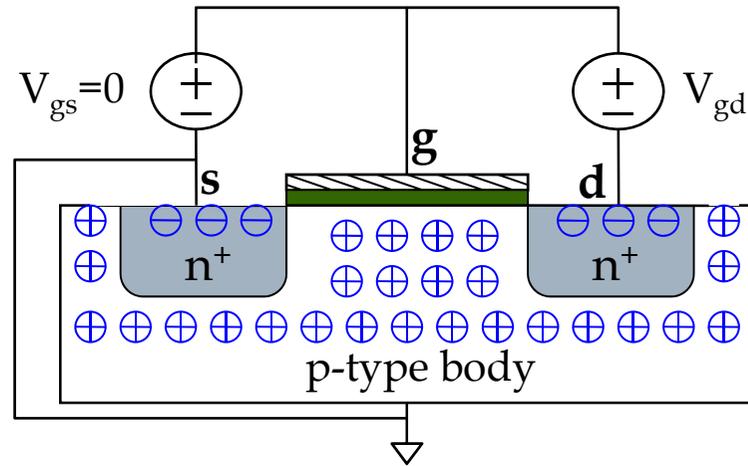
# MOS Transistor

- 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
- The structure of a MOS transistor is symmetric
  - Terminals of source and drain of a MOS can be exchanged

# $V_g$ & Channel for P-Type Body



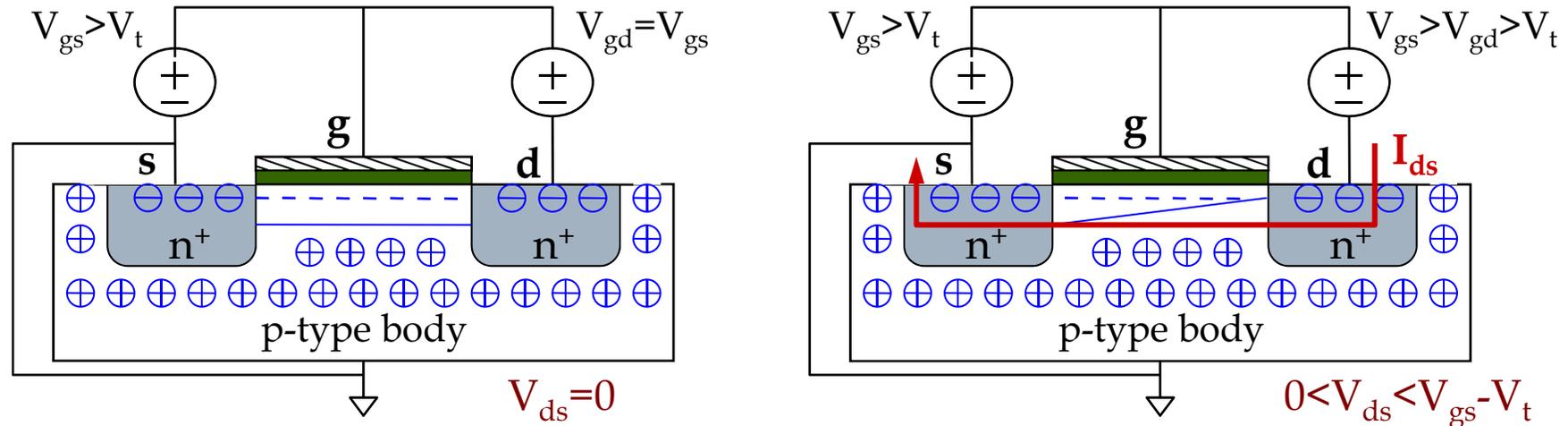
# NMOS Transistor in Cutoff Mode



## ➤ Cutoff region

- ✓ The source and drain have free electrons
- ✓ The body has free holes but no free electrons
- ✓ The junction between the body and the source or drain are reverse-biased, so almost zero current flows

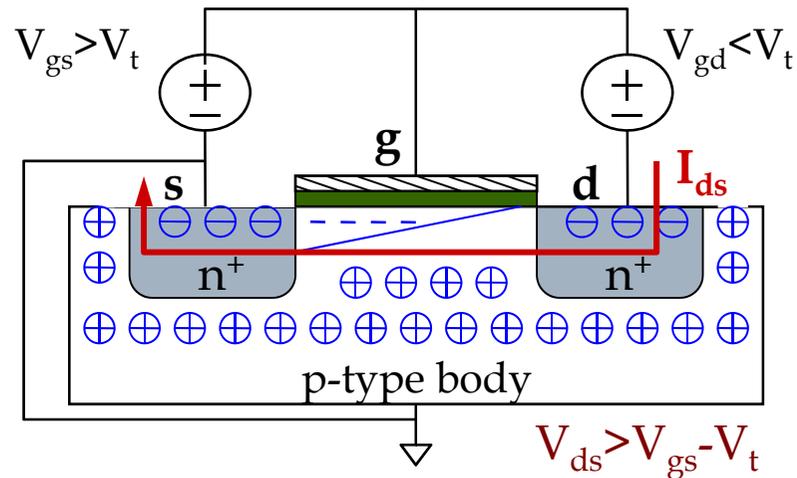
# NMOS Transistor in Linear Mode



## ➤ Linear region

- ✓ A.k.a. resistive, nonsaturated, or unsaturated region
- ✓ If  $V_{gd} = V_{gs}$ , then  $V_{ds} = V_{gs} - V_{gd} = 0$  and there is no electrical field tending to push current from drain to source
- ✓ If  $V_{gs} > V_{gd} > V_t$ , then  $0 < V_{ds} < V_{gs} - V_t$  and there is a small positive potential  $V_{ds}$  is applied to the drain, current  $I_{ds}$  flows through the channel from drain to source
- ✓ The current increases with both the drain and gate voltage

# NMOS Transistor in Saturation Mode



## ➤ Saturation region

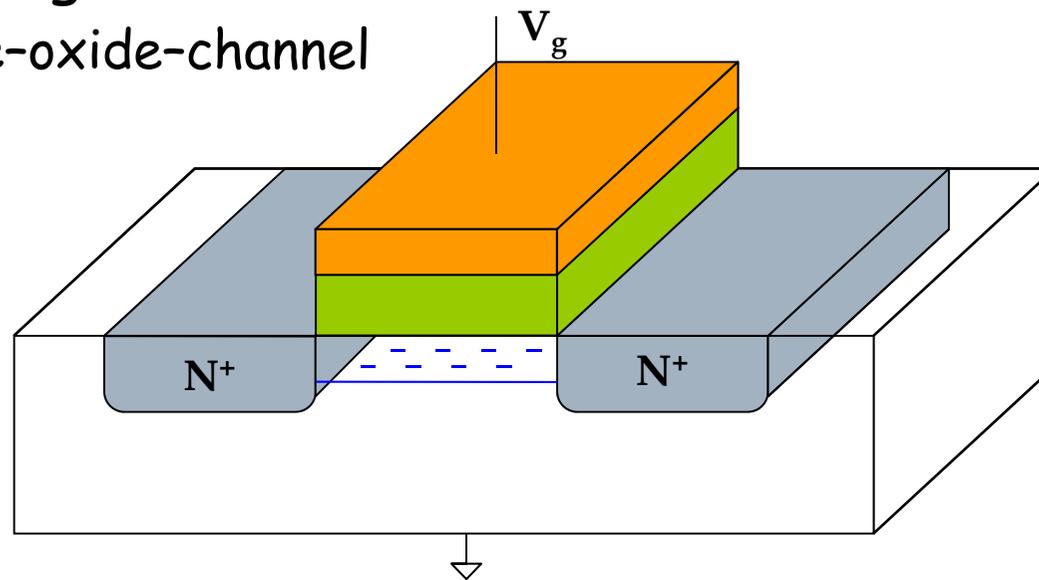
- ✓ The  $V_{ds}$  becomes sufficiently large that  $V_{gd} < V_t$ , the channel is no longer inverted near the drain and becomes pinched off
- ✓ However, conduction is still brought about by the drift of electrons under the influence of the positive drain voltage
- ✓ As electrons reach the end of the channel, they are injected into the depletion region near the drain and accelerated toward the drain
- ✓ The current  $I_{ds}$  is controlled by the gate voltage and ceases to be influenced by the drain

# NMOS Transistor

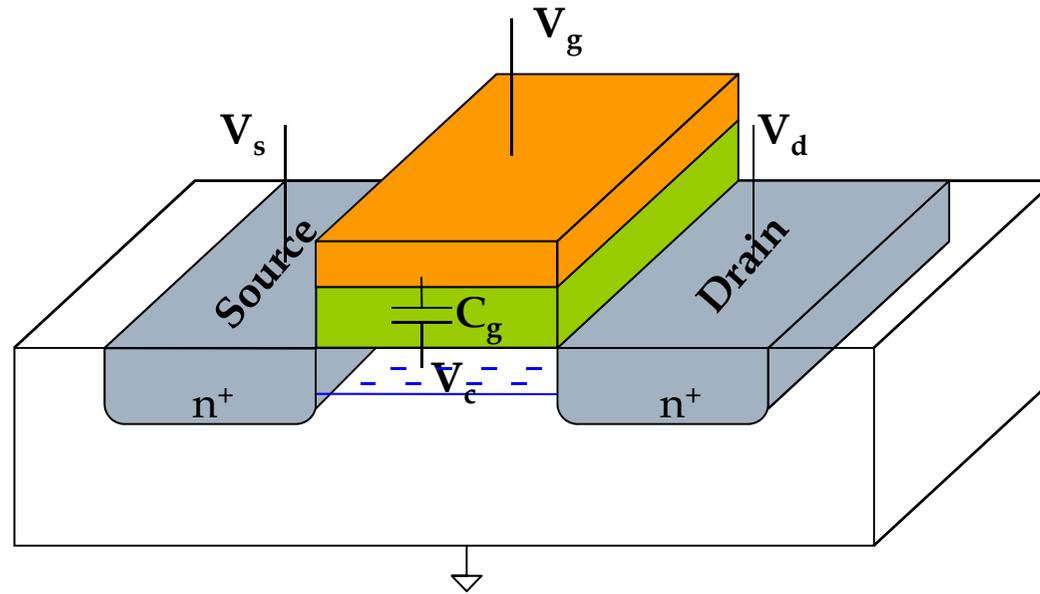
- In summary, the NMOS transistor has three modes of operations
  - If  $V_{gs} < V_t$ , the transistor is cutoff and no current flows
  - If  $V_{gs} > V_t$  and  $V_{ds}$  is small, the transistor acts as a linear resistor in which the current flow is proportional to  $V_{ds}$
  - If  $V_{gs} > V_t$  and  $V_{ds}$  is large, the transistor acts as a current source in which the current flow becomes independent of  $V_{ds}$
- The PMOS transistor operates in just the opposite fashion

# I-V Characteristics of MOS

- In linear and saturation regions, the gate attracts carriers to form a channel
- The carriers drift from source to drain at a rate proportional to the electric field between these regions
- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate-oxide-channel



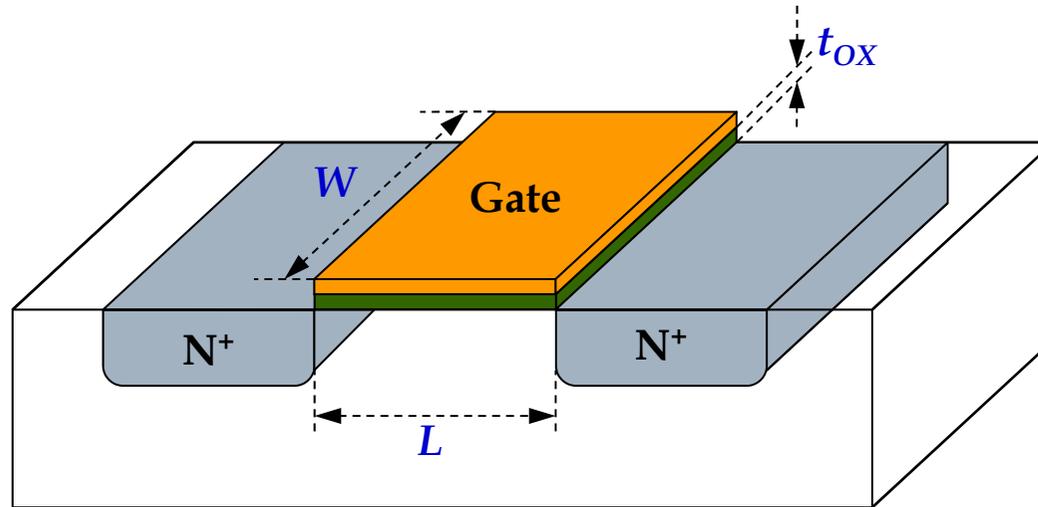
# Channel Charge



- $Q_{\text{channel}} = C_g(V_{g_c} - V_t)$ , where  $C_g$  is the capacitance of the gate to the channel and  $V_{g_c} - V_t$  is the amount of voltage attracting charge to the channel beyond the minimal required to invert from p to n
- $V_c = (V_s + V_d)/2 = V_s + V_{ds}/2$
- Therefore,  $V_{g_c} = (V_{gs} + V_{gd})/2 = V_{gs} - V_{ds}/2$

# Gate Capacitance ( $C_g$ )

- Transistor dimensions



- The gate capacitance is

- $C_g = \epsilon_{ox} \frac{WL}{t_{ox}}$

# Carrier Velocity

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

# NMOS Linear I-V

□ Now we know

- How much charge  $Q_{\text{channel}}$  is in the channel
- How much time  $t$  each carrier takes to cross

□

$$\begin{aligned} I_{ds} &= \frac{Q_{\text{channel}}}{t} \\ &= \mu C_{\text{ox}} \frac{W}{L} \left( V_{gs} - V_t - \frac{V_{ds}}{2} \right) V_{ds} \\ &= \beta \left( V_{gs} - V_t - \frac{V_{ds}}{2} \right) V_{ds} \end{aligned}$$

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

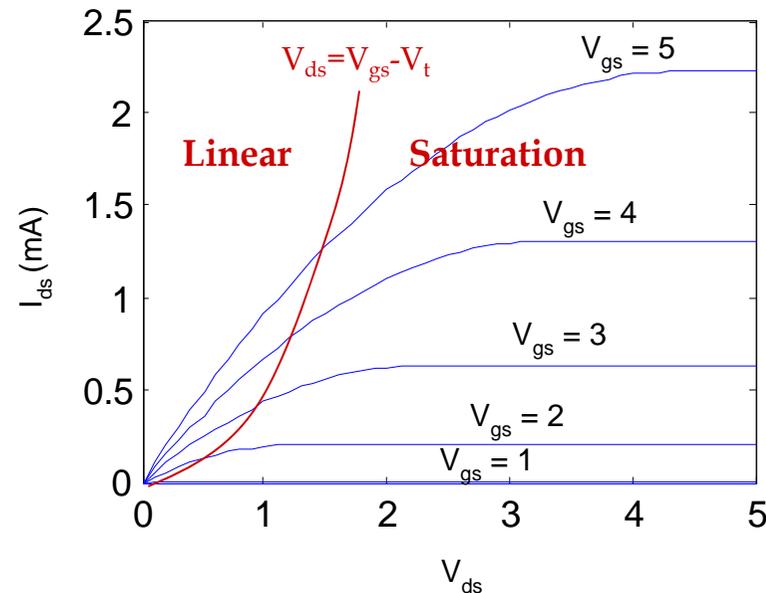
# NMOS Saturation I-V

- 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

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

# Summary of NMOS I-V Characteristics

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



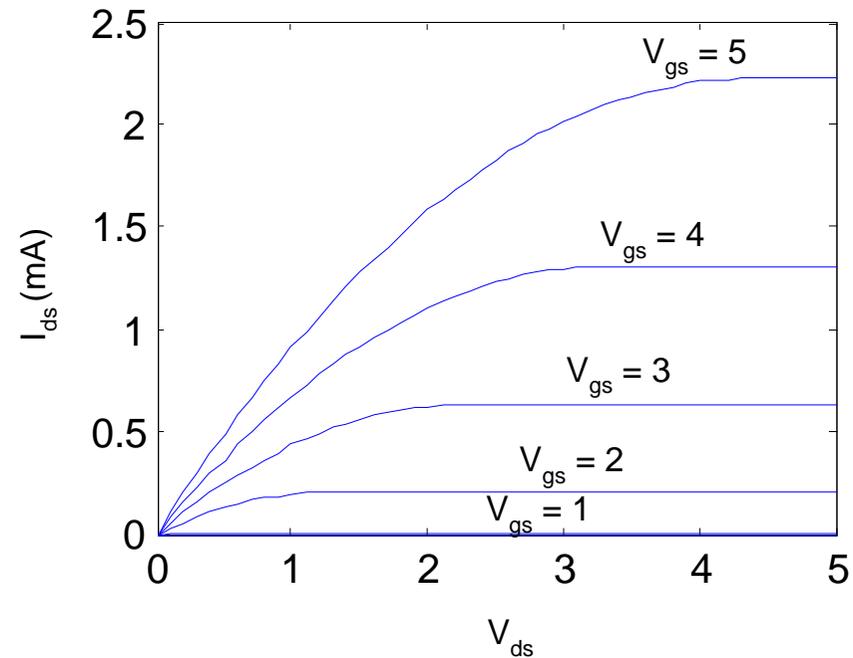
# Example

□ Assume that the parameters of a technology are as follows

- $t_{ox} = 100 \text{ \AA}$
- $\mu = 350 \text{ cm}^2/\text{V}\cdot\text{s}$
- $V_t = 0.7 \text{ V}$

□ Plot  $I_{ds}$  vs.  $V_{ds}$

- $V_{gs} = 0, 1, 2, 3, 4, 5$
- Use  $W/L = 4/2 \lambda$



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

# Nonideal I-V Effects

## □ Nonideal I-V effects

- Velocity saturation, mobility degradation, channel length modulation, subthreshold conduction, body effect, etc.

## □ The saturation current increases less than quadratically with increasing $V_{gs}$ . This is caused by two effects:

- Velocity saturation
- Mobility degradation

## □ Velocity saturation

- At high lateral field strengths ( $V_{ds}/L$ ), carrier velocity ceases to increase linearly with field strength
- Result in lower  $I_{ds}$  than expected at high  $V_{ds}$

## □ Mobility degradation

- At high vertical field strengths ( $V_{gs}/t_{ox}$ ), the carriers scatter more often
- Also lead to less current than expected at high  $V_{gs}$

# Channel Length Modulation

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

- $I_{ds} = \frac{1}{2} \mu \frac{W}{L} C_{ox} (V_{gs} - V_t)^2$

- Actually, the width  $L_d$  of the depletion region between the channel and drain is increased with  $V_{db}$ . To avoid introducing the body voltage into our calculations, assume the source voltage is close to the body voltage so  $V_{db} \sim V_{ds}$

- Thus the effective channel length is shorten to  $L_{eff} = L - L_d$

- Therefore, the  $I_{ds}$  can be expressed as

$$I_{ds} = \frac{1}{2} \mu \frac{W}{L_{eff}} C_{ox} (V_{gs} - V_t)^2 = \frac{1}{2} \mu \frac{W}{L} C_{ox} (V_{gs} - V_t)^2 \frac{1}{1 - \frac{L_d}{L}}$$

- Assume that  $\frac{L_d}{L} \ll 1$ , then

$$I_{ds} = \frac{1}{2} \mu \frac{W}{L} C_{ox} (V_{gs} - V_t)^2 \left(1 + \frac{L_d}{L}\right) = \frac{1}{2} \mu \frac{W}{L} C_{ox} (V_{gs} - V_t)^2 (1 + \lambda V_{ds})$$

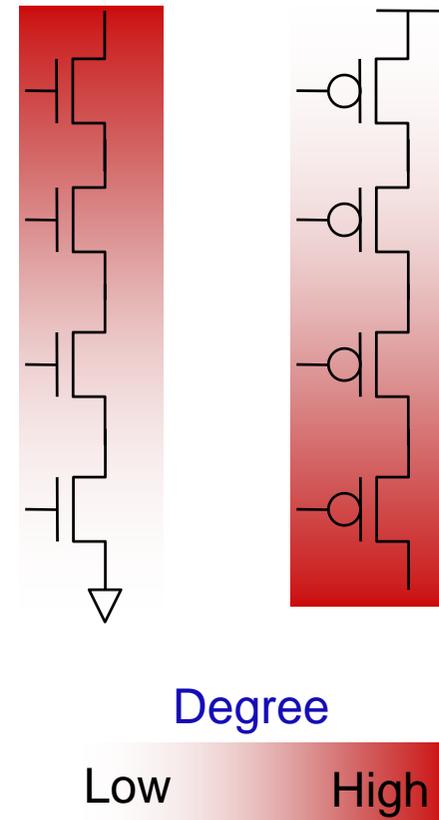
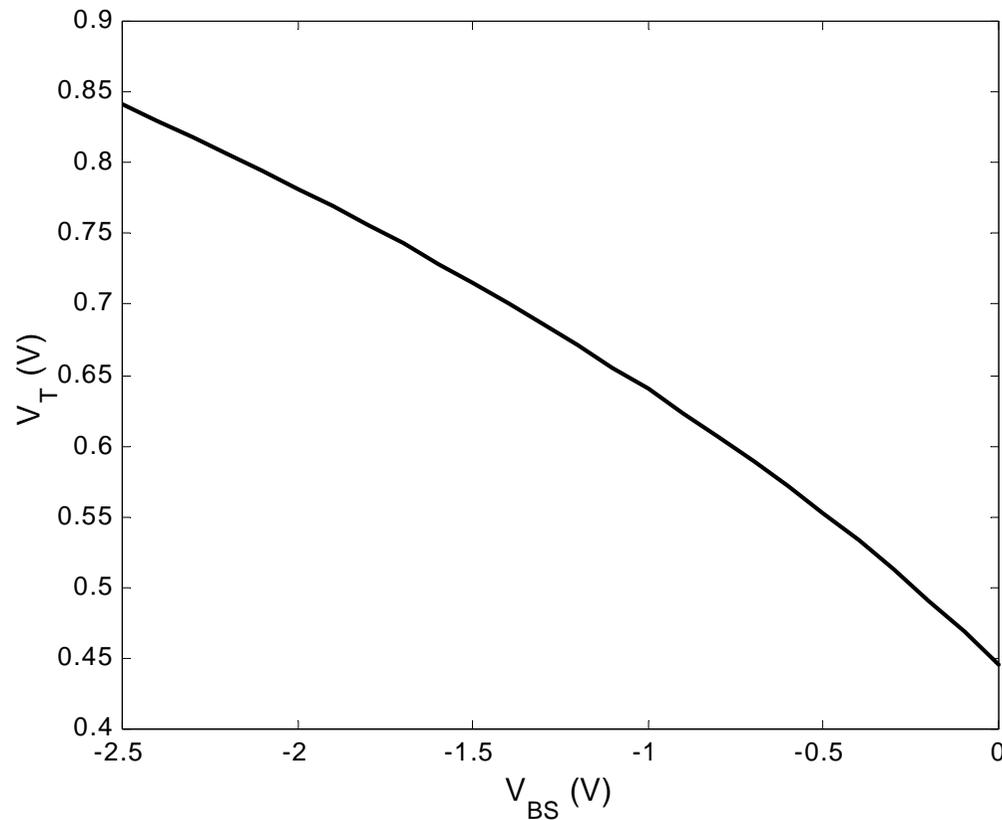
# Channel Length Modulation

- The parameter  $\lambda$  is an empirical channel length modulation factor
- As channel length gets shorter, the effect of the channel length modulation becomes relatively more important
  - Hence  $\lambda$  is inversely dependent on channel length
- This channel length modulation model is a gross oversimplification of nonlinear behavior and is more useful for conceptual understanding than for accurate device modeling
- Channel length modulation is very important to analog designers because it reduces the gain of amplifiers. It is generally unimportant for qualitatively understanding the behavior of digital circuits

# Body Effect

## □ Body effect

- $V_T$  is a function of voltage between source and substrate



# Mobility Variation

- Mobility  $\mu$ 
  - It describes the ease with which carriers drift in the substrate material
  - It is defined by
    - $\mu = (\text{average carrier drift velocity, } v) / (\text{electrical field, } E)$
- Mobility varies according to the type of charge carrier
  - Electrons have a higher mobility than holes
    - Thus NMOS has higher current-producing capability than the corresponding PMOS
- Mobility decreases with increasing doping-concentration and increasing temperature

# Drain Punchthrough & Hot Electrons

## □ Drain punchthrough

- When the drain voltage is high enough, the depletion region around the drain may extend to source. Thus, causing current to flow irrespective of the gate voltage

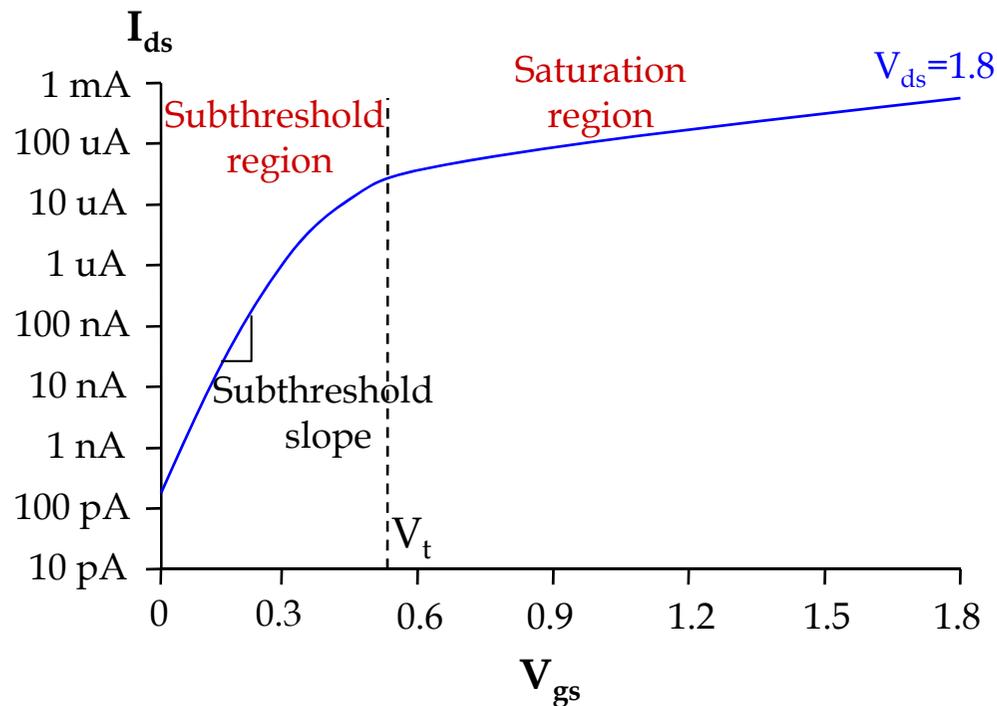
## □ Hot electrons

- When the source-drain electric field is too large, the electron speed will be high enough to break the electron-hole pair. Moreover, the electrons will penetrate the gate oxide, causing a gate current

# Subthreshold Conduction

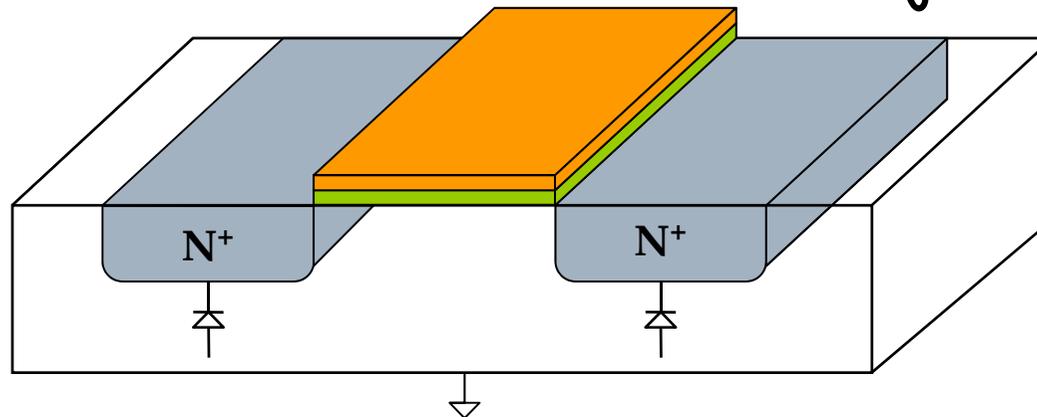
## □ Subthreshold region

- The cutoff region is also referred to as the subthreshold region, where  $I_{ds}$  increases exponentially with  $V_{ds}$  and  $V_{gs}$
- Observe in the following figure that at  $V_{gs} < V_t$ , the current drops off exponentially rather than abruptly becoming zero



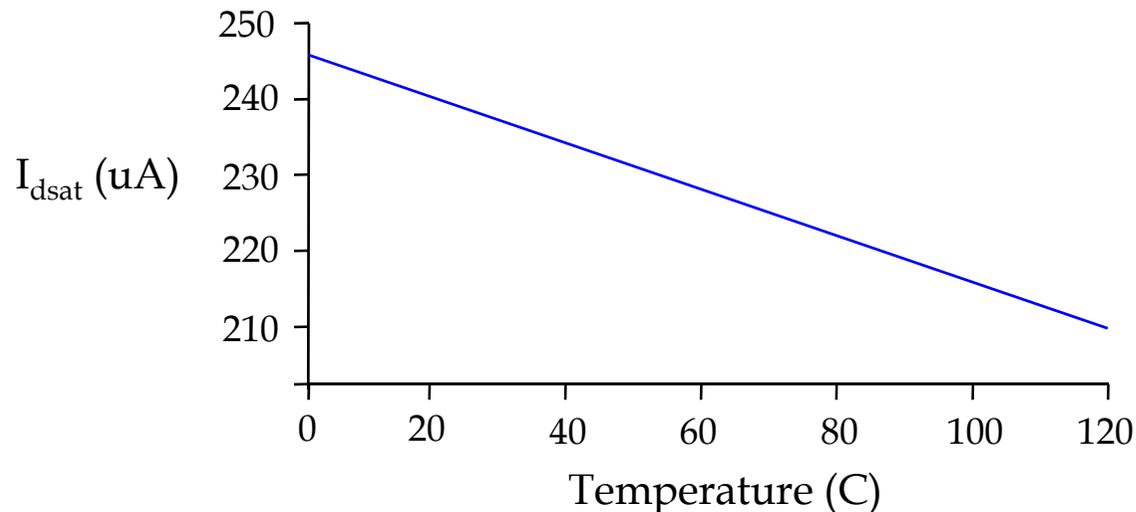
# Junction Leakage

- The p-n junctions between diffusion and the substrate or well form diodes
- The p-type and n-type substrates are tied to GND or  $V_{dd}$  to ensure these diodes remain reverse-biased
- However, reverse-biased diodes still conduct a small amount of current  $I_L$ 
  - $I_L = I_S (e^{\frac{V_D}{v_T}} - 1)$ ,  $V_D$ : diode voltage;  $v_T$ : thermal voltage (about 26mv at room temperature)
- In modern transistors with low threshold voltages, subthreshold conduction far exceeds junction leakage



# Temperature Dependence

- The magnitude of the threshold voltage decreases nearly linearly with temperature
- Carrier mobility decreases with temperature
- Junction leakage increases with temperature because  $I_s$  is strongly temperature dependent
- The following figure shows how the current  $I_{dsat}$  decreases with temperature

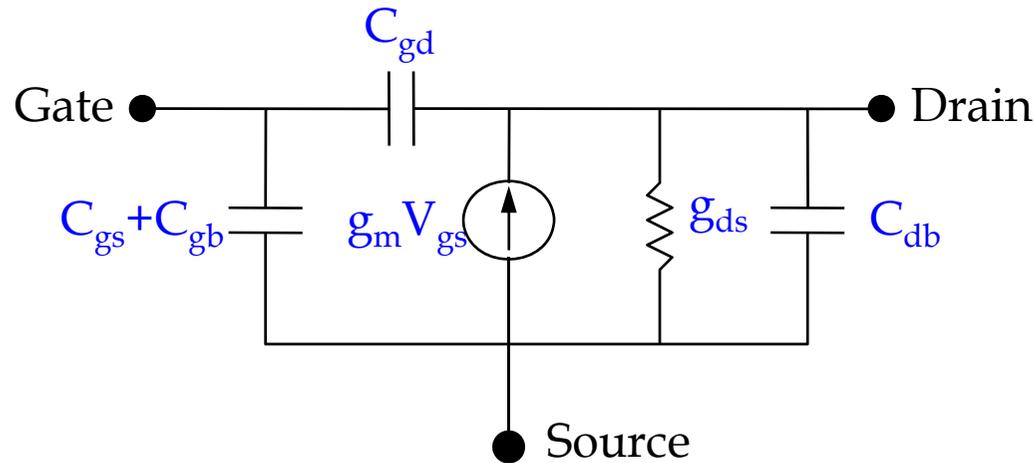


# Geometry Dependence

- The layout designer draws transistors with width and length  $W_{\text{draw}}$  and  $L_{\text{draw}}$ . The actual gate dimensions may differ by some factors  $X_W$  and  $X_L$ 
  - E.g., the manufacturer may create masks with narrower polysilicon or may overetch the polysilicon to provide shorter channels (negative  $X_L$ )
- Moreover, the source and drain tend to diffuse laterally under the gate by  $L_D$ , producing a shorter effective channel length that the carriers must traverse between source and drain. Similarly, diffusion of the bulk by  $W_D$  decreases the effective channel width
- Therefore, the actually effective channel length and width can be expressed as
  - $L_{\text{eff}} = L_{\text{draw}} + X_L - 2L_D$
  - $W_{\text{eff}} = W_{\text{draw}} + X_W - 2W_D$

# MOS Small Signal Model

$(V_{sb}=0)$



Linear region

$$I_{ds} = \mu \frac{W}{L} C_{ox} [(V_{gs} - V_t)V_{ds} - \frac{1}{2}V_{ds}^2]$$

$$g_{ds} = \frac{dI_{ds}}{dV_{ds}} = \mu \frac{W}{L} C_{ox} [(V_{gs} - V_t) - V_{ds}]$$

$$g_m = \left. \frac{dI_{ds}}{dV_{gs}} \right|_{(V_{ds} = \text{const.})} = \mu \frac{W}{L} C_{ox} V_{ds}$$

Saturation region

$$I_{ds} = \frac{1}{2} \mu \frac{W}{L} C_{ox} (V_{gs} - V_t)^2$$

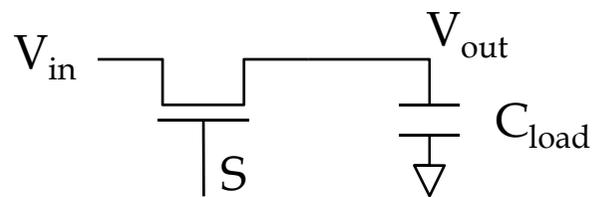
$$g_{ds} = 0$$

$$g_m = \mu \frac{W}{L} C_{ox} (V_{gs} - V_t)$$

# Pass Transistor

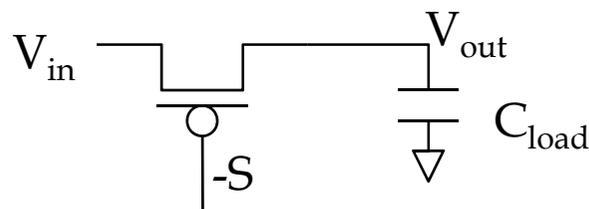
## □ NMOS pass transistor

- $C_{load}$  is initially discharged, i.e.,  $V_{out} = V_{ss}$
- If  $V_{in} = V_{dd}$  and  $V_S = V_{dd}$ , the  $V_{out} = V_{dd} - V_{tn}$
- If  $V_{in} = V_{ss}$  and  $V_S = V_{dd}$ , the  $V_{out} = V_{ss}$

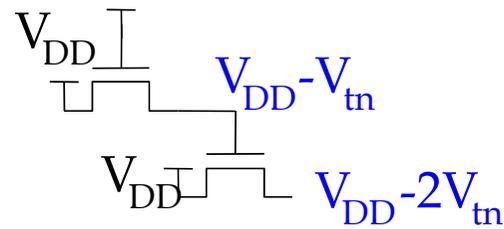
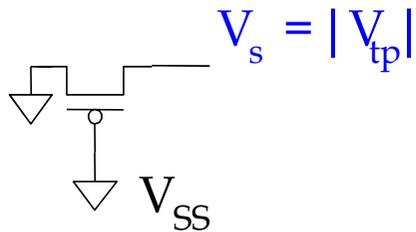
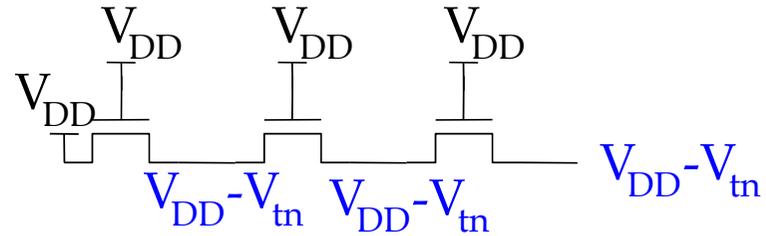
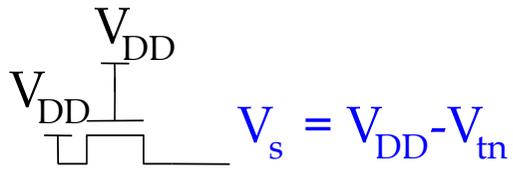


## □ PMOS pass transistor

- If  $V_{in} = V_{dd}$  and  $V_{-S} = V_{ss}$ , the  $V_{out} = V_{dd}$
- If  $V_{in} = V_{ss}$  and  $V_{-S} = V_{ss}$ , the  $V_{out} = V_{tp}$

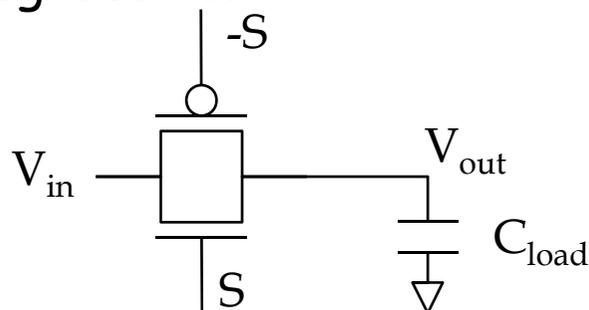


# Pass Transistor Circuits



# Transmission Gate

- By combining behavior of the NMOS and PMOS, we can construct a transmission gate
  - The transmission gate can transmit both logic one and logic zero without degradation



- The transmission gate is a fundamental and ubiquitous component in MOS logic
  - A multiplexer element
  - A logic structure,
  - A latch element, etc.

# Summary

## □ Threshold drops

- Pass transistors suffer a threshold drop when passing the wrong value: NMOS transistors only pull up to  $V_{DD}-V_{tn}$ , while PMOS transistors only pull down to  $|V_{tp}|$
- The magnitude of the threshold drop is increased by the body effect
- Fully complementary transmission gates should be used where both 0's and 1's must be passed well

## □ $V_{DD}$

- Velocity saturation and mobility degradation result in less current than expected at high voltage
- This means that there is no point in trying to use a high  $V_{DD}$  to achieve high fast transistors, so  $V_{DD}$  has been decreasing with process generation to reduce power consumption
- Moreover, the very short channels and thin gate oxide would be damaged by high  $V_{DD}$

# Summary

- Leakage current
  - Real gates draw some leakage current
  - The most important source at this time is subthreshold leakage between source and drain of a transistor that should be cut off
  - The subthreshold current of a OFF transistor decreases by an order of magnitude for every 60-100mV that  $V_{gs}$  is below  $V_{\dagger}$ . Threshold voltages have been decreasing, so subthreshold leakage has been increasing dramatically
  - Some processes offer multiple choices of  $V_{\dagger}$ ; low- $V_{\dagger}$  devices are used for high performance, while high- $V_{\dagger}$  devices are used for low leakage elsewhere
  - Leakage current causes CMOS gates to consume power when idle. It also limits the amount of time that data is retained in dynamic logic, latches, and memory cells
  - In modern processes, dynamic logic and latches require some sort of feedback to prevent data loss from leakage
  - Leakage increases at high temperature