EEC 116 Lecture #3: CMOS Inverters MOS Scaling

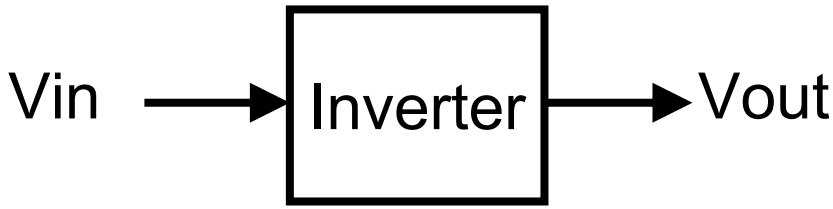
Rajeevan Amirtharajah
University of California, Davis
Jeff Parkhurst
Intel Corporation

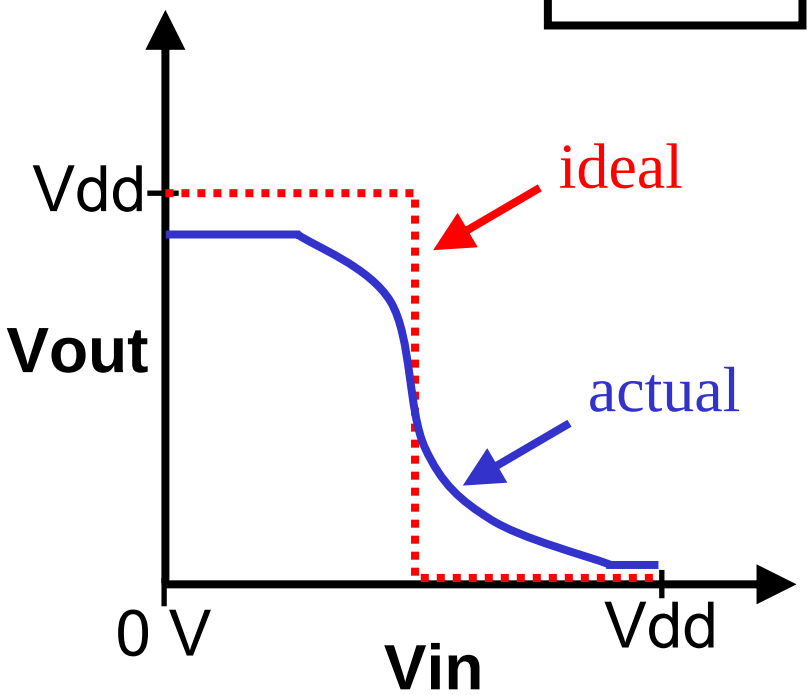
Outline

- **Review: Inverter Transfer Characteristics**
- **Lecture 3: Noise Margins, Rise & Fall Times, Inverter Delay**
- **CMOS Inverters: Rabaey 1.3.2, 5 (Kang & Leblebici, 5.1-5.3 and 6.1-6.2)**

Review: Inverter Voltage Transfer Curve

Voltage transfer curve (VTC): plot of output voltage Vout vs. input voltage Vin

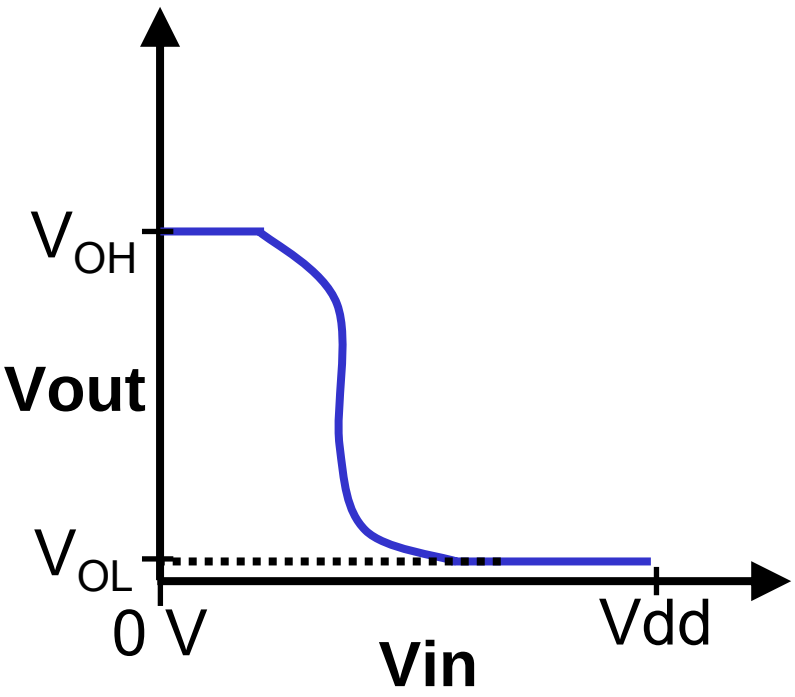




Ideal digital inverter:

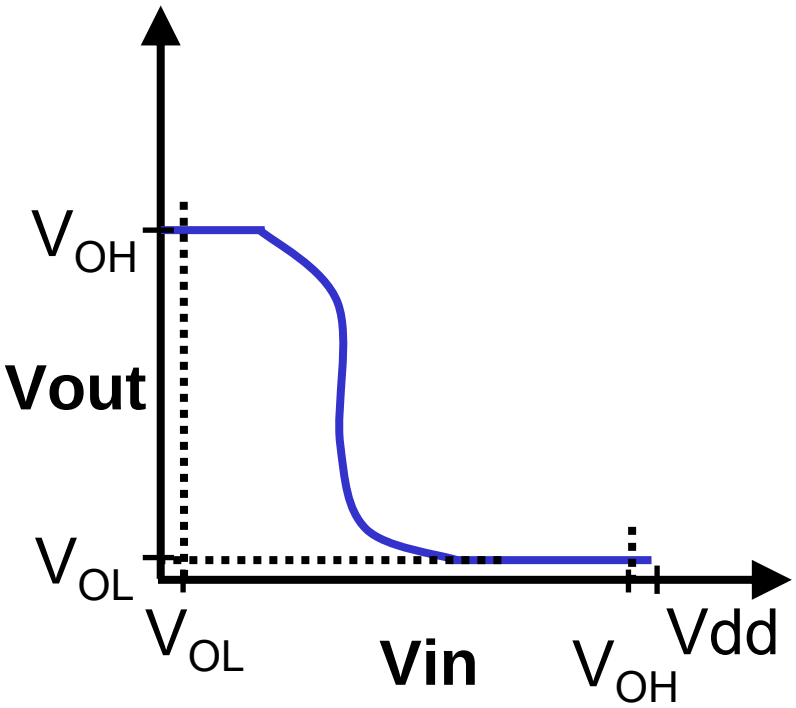
- When Vin=0, Vout=Vdd
- When Vin=Vdd, Vout=0
- Sharp transition region

Review: Actual Inverter Output Levels



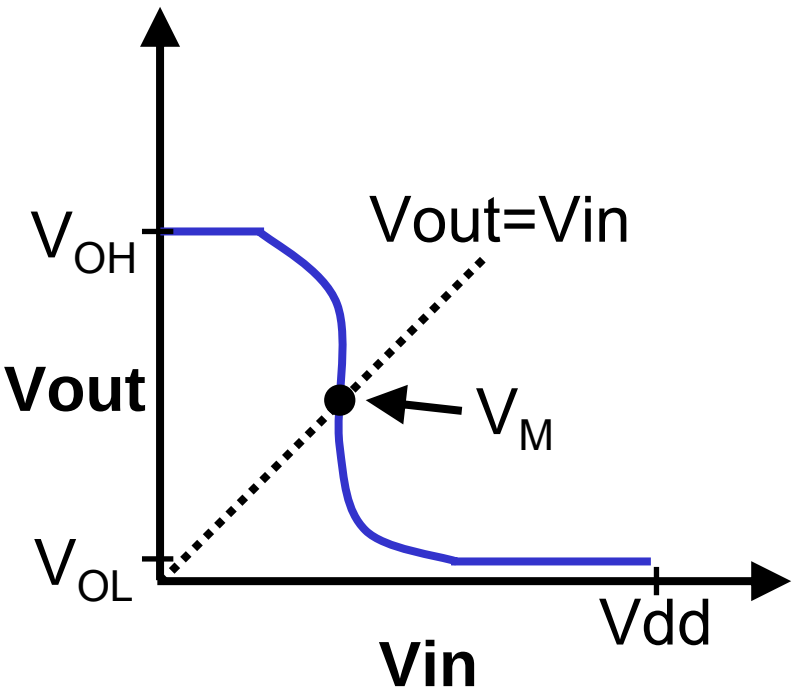
- **V_{OH} and V_{OL} represent the "high" and "low" output voltages of the inverter**
- **V_{OH} = output voltage when Vin = '0' (V Output High)**
- **V_{OL} = output voltage when Vin = '1' (V Output Low)**
- **Ideally,**
  - V_{OH} = Vdd
  - V_{OL} = 0 V

Review: VOL and VOH



- **In transfer function terms:**
  - $V_{OL} = f(V_{OH})$
  - $V_{OH} = f(V_{OL})$
  - f = inverter transfer function
- **Difference (V_{OH}-V_{OL}) is the *voltage swing* of the gate**
  - *Full-swing logic* swings from ground to Vdd
  - Other families with smaller swings

Review: Inverter Switching Threshold



Inverter switching threshold:

- Point where voltage transfer curve intersects line Vout=Vin
- Represents the point at which the inverter switches state
- Normally, $V_M \approx Vdd/2$
- Sometimes other thresholds desirable

VTC Mathematical Definitions

- **V_{OH} is the output high level of an inverter**

  $V_{OH} = VTC(V_{OL})$

- **V_{OL} is the output low level of an inverter**

  $V_{OL} = VTC(V_{OH})$

- **V_M is the switching threshold**

  $V_M = V_{IN} = V_{OUT}$

- **V_{IH} is the lowest input voltage for which the output will be ≥ the input (worst case '1')**

  $dVTC(V_{IH})/dV_{IH} = -1$

- **V_{IL} is the highest input voltage for which the output will be ≤ the input (worst case '0')**

  $dVTC(V_{IL})/dV_{IL} = -1$

Noise Margin and Delay Definitions

- **NM_L is the difference between the highest acceptable '0' and the lowest possible '0'**

  $$NM_L = V_{IL} - V_{OL}$$

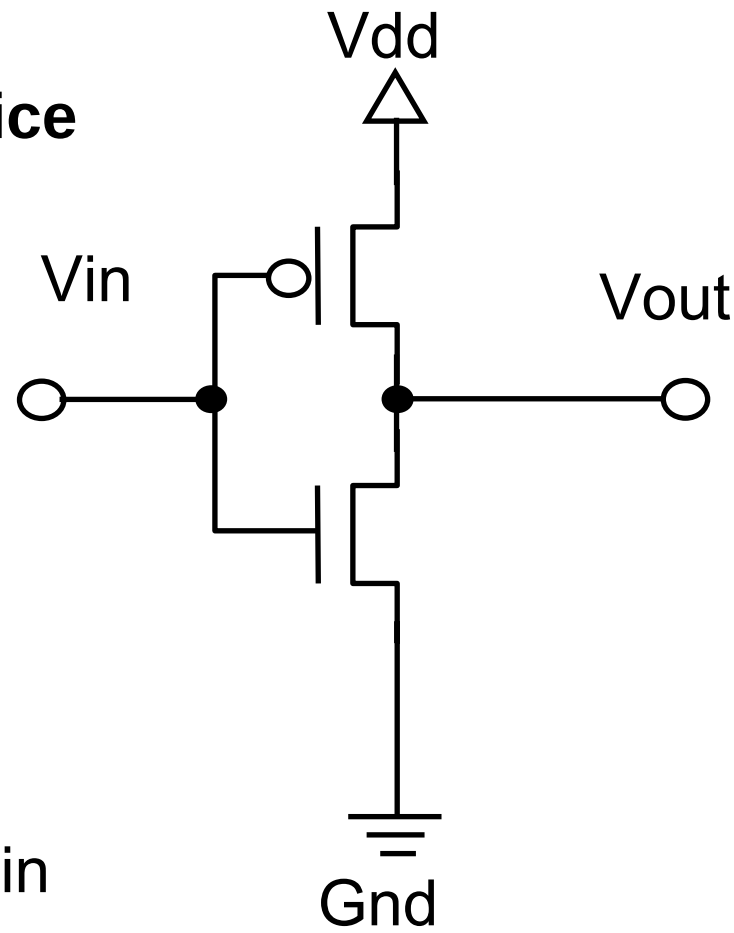
- **NM_H is the difference between the lowest acceptable '1' and the highest possible '1'**

  $$NM_H = V_{OH} - V_{IH}$$

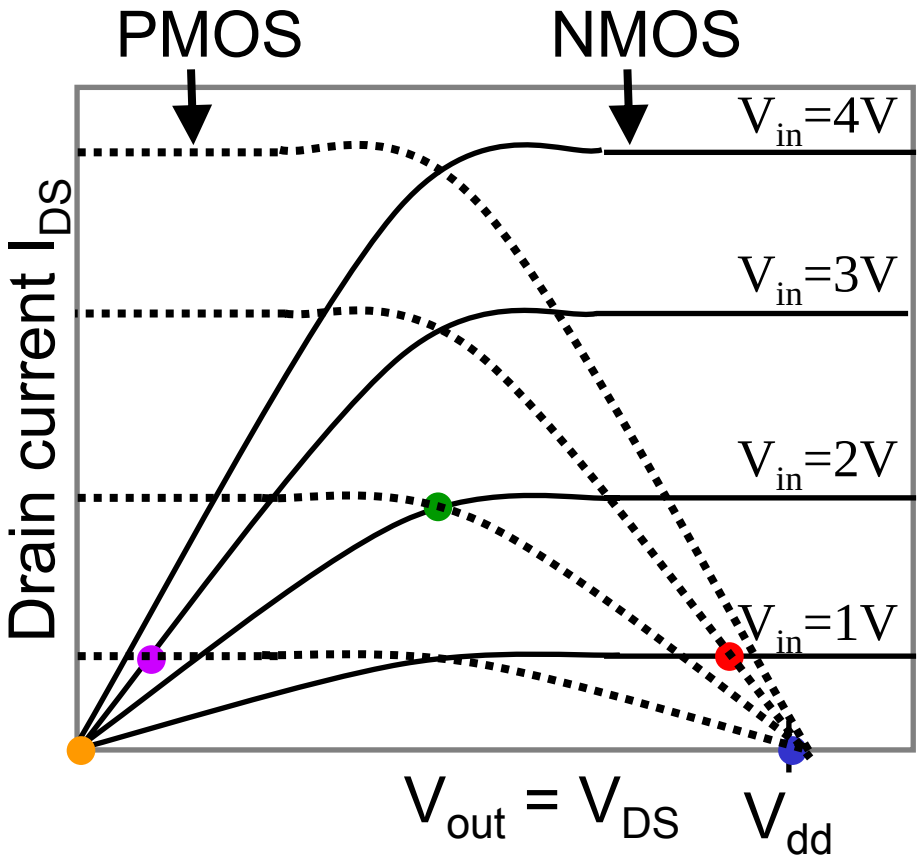
- **t_{PHL} is the propagation delay from the 50% point of the input to the output when the output goes from high to low**
- **t_{PLH} is the propagation delay from the 50% point of the input to the output when the output goes from low to high**
- **t_P is the average propagation delay**
- **t_R is the rise time (usually 10% to 90%)**
- **t_F is the fall time (usually 90% to 10%)**

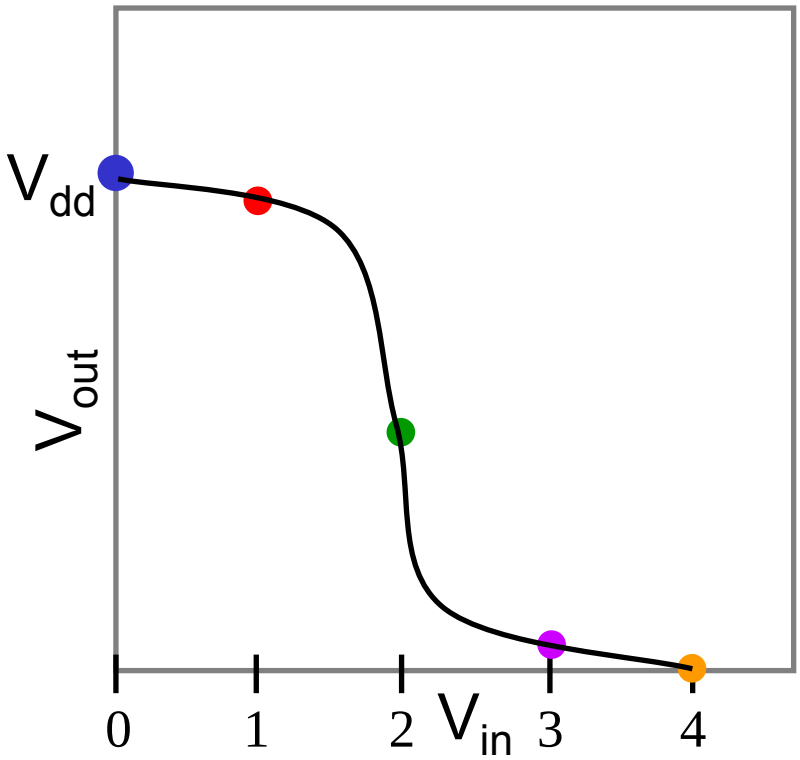
CMOS Inverter

- **Complementary NMOS and PMOS devices**
- **In steady-state, only one device is on (no static power consumption)**
- **Vin=1: NMOS on, PMOS off**
  - Vout = V_{OL} = 0
- **Vin=0: PMOS on, NMOS off**
  - Vout = V_{OH} = Vdd
- **Ideal V_{OL} and V_{OH}!**
- Ratioless logic: output is independent of transistor sizes in steady-state



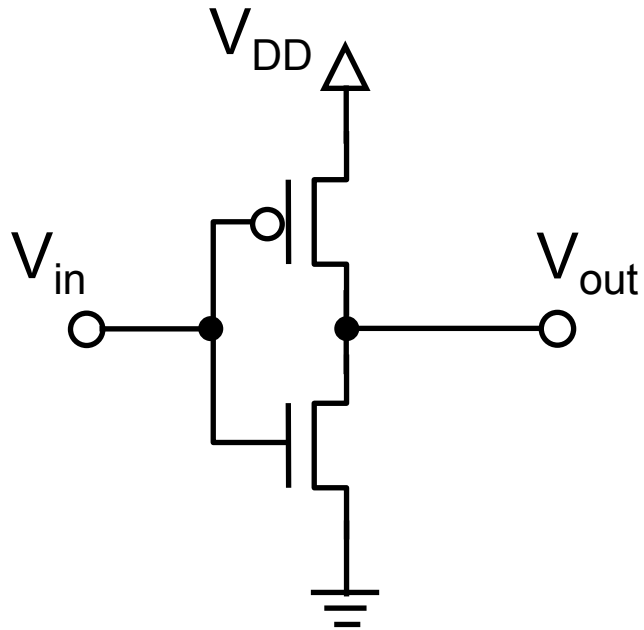
CMOS Inverter: VTC





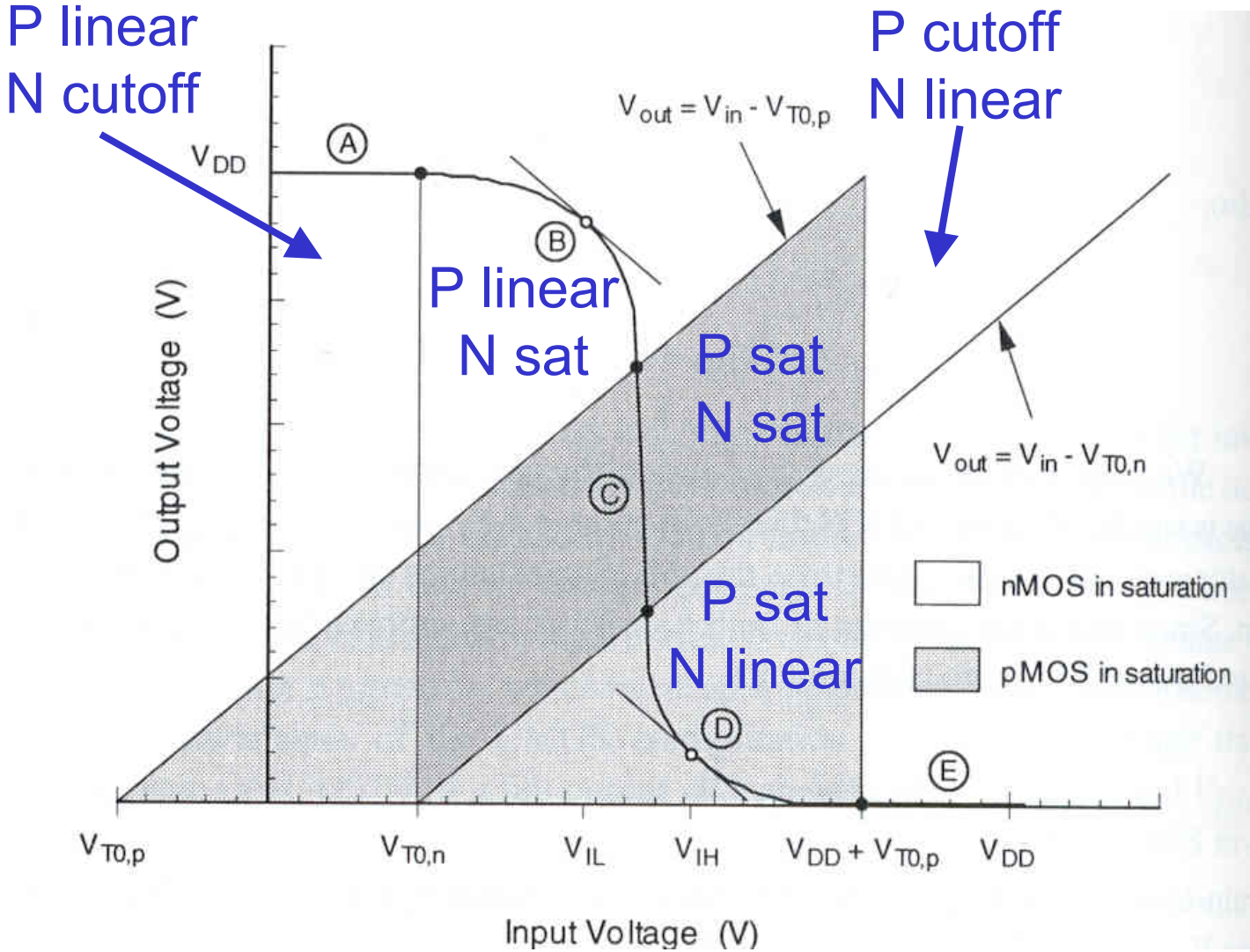
- Output goes completely to Vdd and Gnd
- Sharp transition region

CMOS Inverter Operation

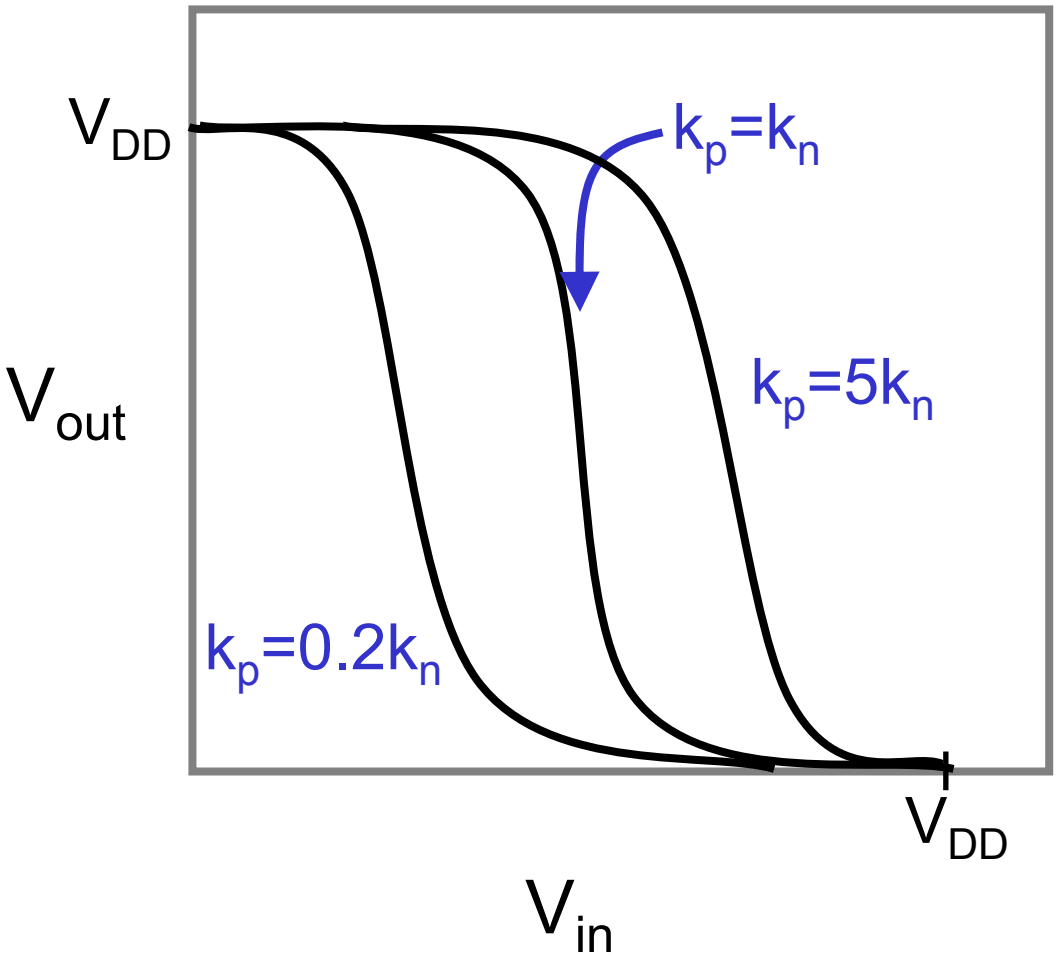


- **NMOS transistor:**
  - Cutoff if $V_{in} < V_{TN}$
  - Linear if $V_{out} < V_{in} - V_{TN}$
  - Saturated if $V_{out} > V_{in} - V_{TN}$
- **PMOS transistor**
  - Cutoff if $(V_{in}-V_{DD}) > V_{TP} \rightarrow V_{in} > V_{DD}+V_{TP}$
  - Linear if $(V_{out}-V_{DD})>V_{in}-V_{DD}-V_{TP} \rightarrow V_{out}>V_{in} - V_{TP}$
  - Sat. if $(V_{out}-V_{DD})<V_{in}-V_{DD}-V_{TP} \rightarrow V_{out} < V_{in}-V_{TP}$

CMOS Inverter VTC: Device Operation



CMOS Inverter VTC: Device Sizing



- Increase W of PMOS
  - k_p increases
  - VTC moves to right
- Increase W of NMOS
  - k_n increases
  - VTC moves to left
- For $V_M = V_{DD}/2$
  - $k_n = k_p$
  - $2W_n \approx W_p$

Effects of V_M adjustment

- **Result from changing k_p/k_n ratio:**
  - Inverter threshold $V_M \neq V_{DD}/2$
  - Rise and fall delays unequal
  - Noise margins not equal
- **Reasons for changing inverter threshold**
  - Want a faster delay for one type of transition (rise/fall)
  - Remove noise from input signal: increase one noise margin at expense of the other
  - Interfacing other types of logic (with different swings)

CMOS Inverter: V_{IL} Calculation

- **KCL (NMOS saturation, PMOS linear):**

$$\frac{k_n}{2}\left(V_{GS,n}-V_{T0,n}\right)^2=\frac{k_p}{2}\left[2\left(V_{GS,p}-V_{T0,p}\right)V_{DS,p}-V_{DS,p}{}^2\right]$$

$$\frac{k_n}{2}\left(V_{in}-V_{T0,n}\right)^2=\frac{k_p}{2}\left[2\left(V_{in}-V_{DD}-V_{T0,p}\right)\left(V_{out}-V_{DD}\right)-\left(V_{out}-V_{DD}\right)^2\right]$$

- **Differentiate and set dVout/dVin to –1**

$$k_n\left(V_{in}-V_{T0,n}\right)=k_p\left[\left(V_{in}-V_{DD}-V_{T0,p}\right)\frac{dV_{out}}{dV_{in}}+\left(V_{out}-V_{DD}\right)-\left(V_{out}-V_{DD}\right)\frac{dV_{out}}{dV_{in}}\right]$$

$$k_n\left(V_{IL}-V_{T0,n}\right)=k_p\left(2V_{out}-V_{IL}+V_{T0,p}-V_{DD}\right)$$

$$V_{IL}=\frac{2V_{out}+V_{T0,p}-V_{DD}+k_R V_{T0,n}}{1+k_R}\qquad k_R=\frac{k_n}{k_p}$$

- **Solve simultaneously with KCL to find V_{IL}**

CMOS Inverter: V_{IH} Calculation

- **KCL:** $\frac{k_n}{2}\left[2\left(V_{GS,n}-V_{T0,n}\right)V_{DS,n}-V_{DS,n}{}^2\right]=\frac{k_p}{2}\left(V_{GS,p}-V_{T0,p}\right)^2$

$$\frac{k_n}{2}\left[2\left(V_{in}-V_{T0,n}\right)V_{out}-V_{out}{}^2\right]=\frac{k_p}{2}\left(V_{in}-V_{DD}-V_{T0,p}\right)^2$$

- **Differentiate and set dVout/dVin to –1**

$$k_n\left[\left(V_{in}-V_{T0,n}\right)\frac{dV_{out}}{dV_{in}}+V_{out}-V_{out}\frac{dV_{out}}{dV_{in}}\right]=k_p\left(V_{in}-V_{DD}-V_{T0,p}\right)$$

$$k_n\left(2V_{out}-V_{IH}+V_{T0,p}\right)=k_p\left(V_{IH}-V_{DD}-V_{T0,p}\right)$$

$$V_{IH}=\frac{V_{DD}+V_{T0,p}+k_R\left(2V_{out}+V_{T0,n}\right)}{1+k_R}\qquad k_R=\frac{k_n}{k_p}$$

- **Solve simultaneously with KCL to find V_{IH}**

CMOS Inverter: V_M Calculation

- **KCL (NMOS & PMOS saturated):**

$$\frac{k_n}{2}\left(V_{GS,n}-V_{T0,n}\right)^2=\frac{k_p}{2}\left(V_{GS,p}-V_{T0,p}\right)^2$$

$$\frac{k_n}{2}\left(V_{in}-V_{T0,n}\right)^2=\frac{k_p}{2}\left(V_{in}-V_{DD}-V_{T0,p}\right)^2$$

- **Solve for $V_M = V_{in} = V_{out}$**

$$V_M=\frac{V_{T0,n}+\sqrt{\frac{1}{k_R}}\left(V_{DD}+V_{T0,p}\right)}{1+\sqrt{\frac{1}{k_R}}} \qquad k_R=\frac{k_n}{k_p}$$

CMOS Inverter: Achieving Ideal V_M

$$V_{TH} = \frac{V_{T0,n} + \sqrt{\frac{1}{k_R}}\left(V_{DD} + V_{T0,p}\right)}{1 + \sqrt{\frac{1}{k_R}}} \qquad k_R = \frac{k_n}{k_p}$$

- **Ideally, $V_M = V_{DD}/2$** $\qquad k_{R,ideal} = \left(\frac{V_{DD}/2 + V_{T0,p}}{V_{DD}/2 + V_{T0,n}}\right)^2$

- **Assuming $V_{T0,n} = V_{T0,p}$,** $k_{R,ideal} = 1$

$$\frac{\left(\frac{W}{L}\right)_p}{\left(\frac{W}{L}\right)_n} = \frac{\mu_n}{\mu_p} \approx 2.5$$

CMOS Inverter: V_{IL} and V_{IH} for Ideal V_M

- **Assuming $V_{T0,n}$=-$V_{T0,p}$, and k_R = 1,**

$$V_{IL} = \frac{1}{8}\left(3V_{DD} + 2\left|V_{T0}\right|\right)$$

$$V_{IH} = \frac{1}{8}\left(5V_{DD} - 2\left|V_{T0}\right|\right)$$

$$V_{IL} + V_{IH} = V_{DD}$$

$$NM_L = V_{IL} - V_{OL} = V_{IL}$$

$$NM_H = V_{OH} - V_{IH} = V_{DD} - V_{IH} = V_{IL}$$

MOSFET Scaling Effects

- **Rabaey Section 3.5 (Kang & Leblebici Section 3.5)**
- **Scaling provides enormous advantages**
  - Scale linear dimension (channel length) by factor S > 1
  - Better area density, yield, performance
- **Two types of scaling**
  - Constant field scaling (full scaling)
    - $A' = A/S^2$; $L' = L/S$; $W' = W/S$; $I_D' = I_D/S$; $P' = P/S^2$; $V_{dd}' = V_{dd}/S$
    - Power Density P'/A' = stays the same
  - Constant voltage scaling
    - $A' = A/S^2$; $L' = L/S$; $W' = W/S$; $I_D' = I_D*S$; $P' = P*S$; $V_{dd}' = V_{dd}$
    - Power Density P'/A' = S^3*P (Reliability issue)

Change these two

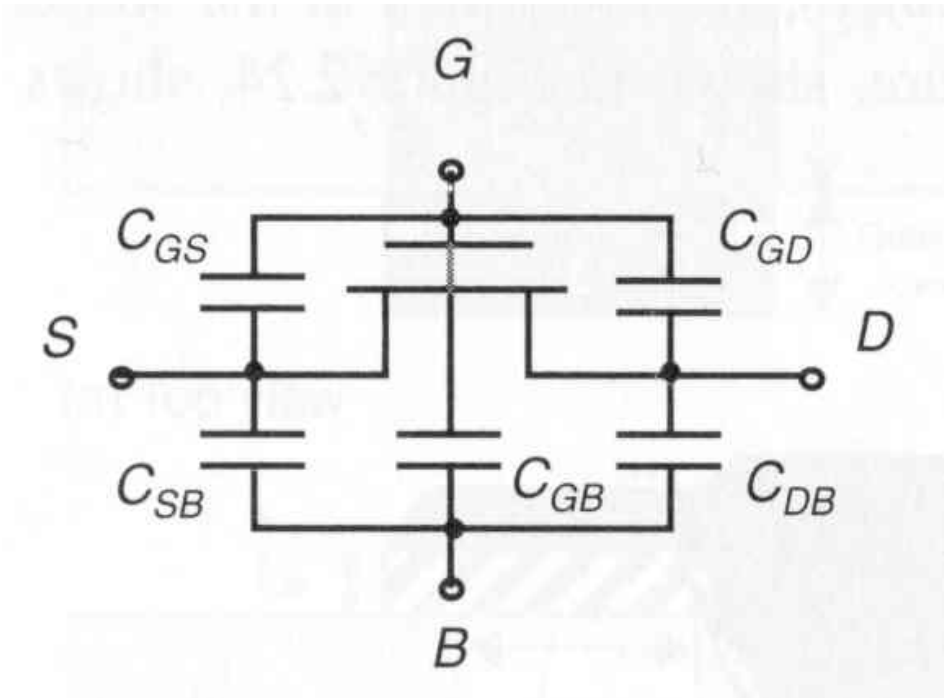
This changed as well

Short Channel Effects

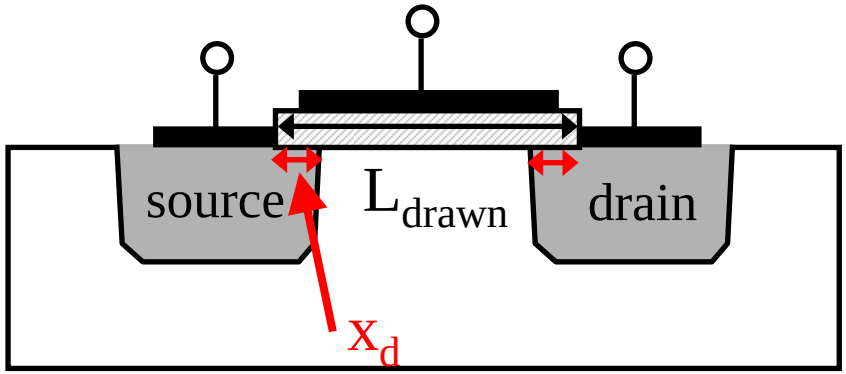
- **As geometries are scaled down**
  - V_T (effective) goes lower
  - Effective channel length decreases
  - Sub-threshold Ids occurs
    - Current goes from drain to source while Vgs < Vt
  - Tox is scaled which can cause reliability problems
    - Can't handle large Vg without hot electron effects
      - Changes the Vt when carriers imbed themselves in the oxide
  - Interconnects scale
    - Electromigration and ESD become issues

MOSFET Capacitances

- **Rabaey Section 3.3 (Kang & Leblebici Section 3.6)**
- **Oxide Capacitance**
  - Gate to Source overlap
  - Gate to Drain overlap
  - Gate to Channel
- **Junction Capacitance**
  - Source to Bulk junction
  - Drain to Bulk junction



Oxide Capacitances: Overlap

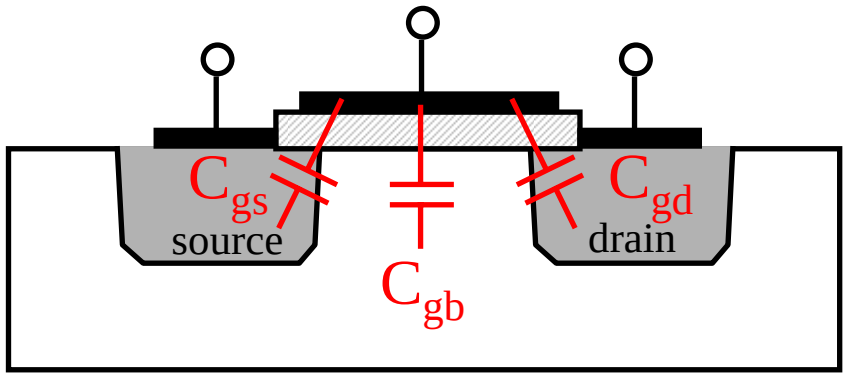


- **Overlap capacitances**
  - Gate electrode overlaps source and drain regions
  - x_d is overlap length on each side of channel
  - $L_{eff} = L_{drawn} - 2x_d$ (effective channel length)
  - Overlap capacitance:

$$C_{GSO} = C_{GDO} = C_{ox} W x_d$$ Assume x_d equal on both sides

Total Oxide Capacitance

- **Total capacitance consists of 2 components**
  - Overlap capacitance
  - Channel capacitance



- **Cutoff:**
  - No channel connecting to source or drain
  - $C_{GS} = C_{GD} = C_{ox}Wx_d$
  - $C_{GB} = C_{ox}WL_{eff}$
  - Total Gate Capacitance = $C_G = C_{ox}WL$

Oxide Capacitances: Channel

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

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

  - Total Gate capacitance $C_G = C_{ox} WL$

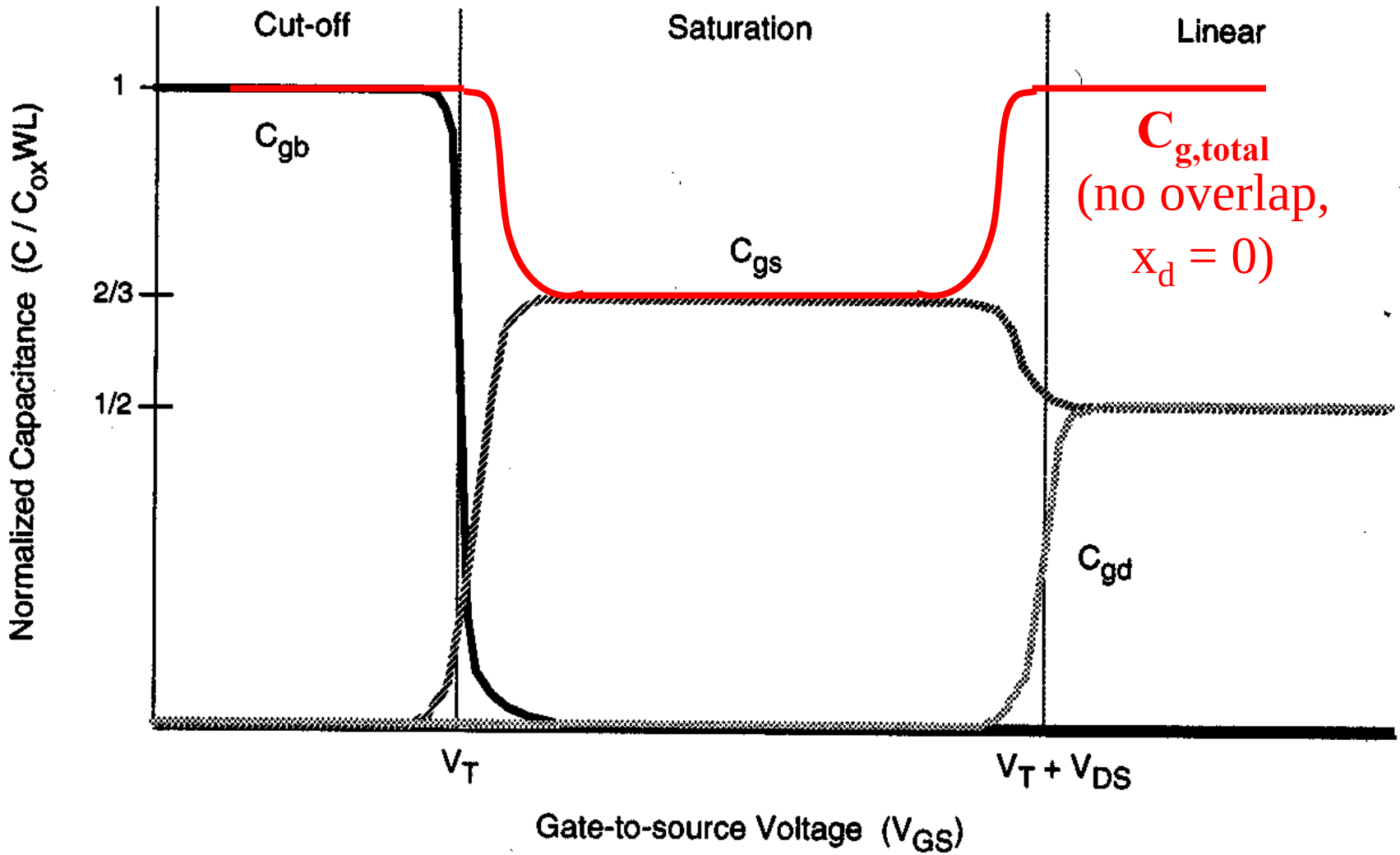
- **Saturation regime**
  - Channel is pinched off: Channel Capacitance --

$$C_{GD} = W x_d C_{ox} \quad C_{GS} = \frac{2}{3} C_{ox} W L_{eff} + C_{OX} W x_d \quad C_{GB} = 0$$

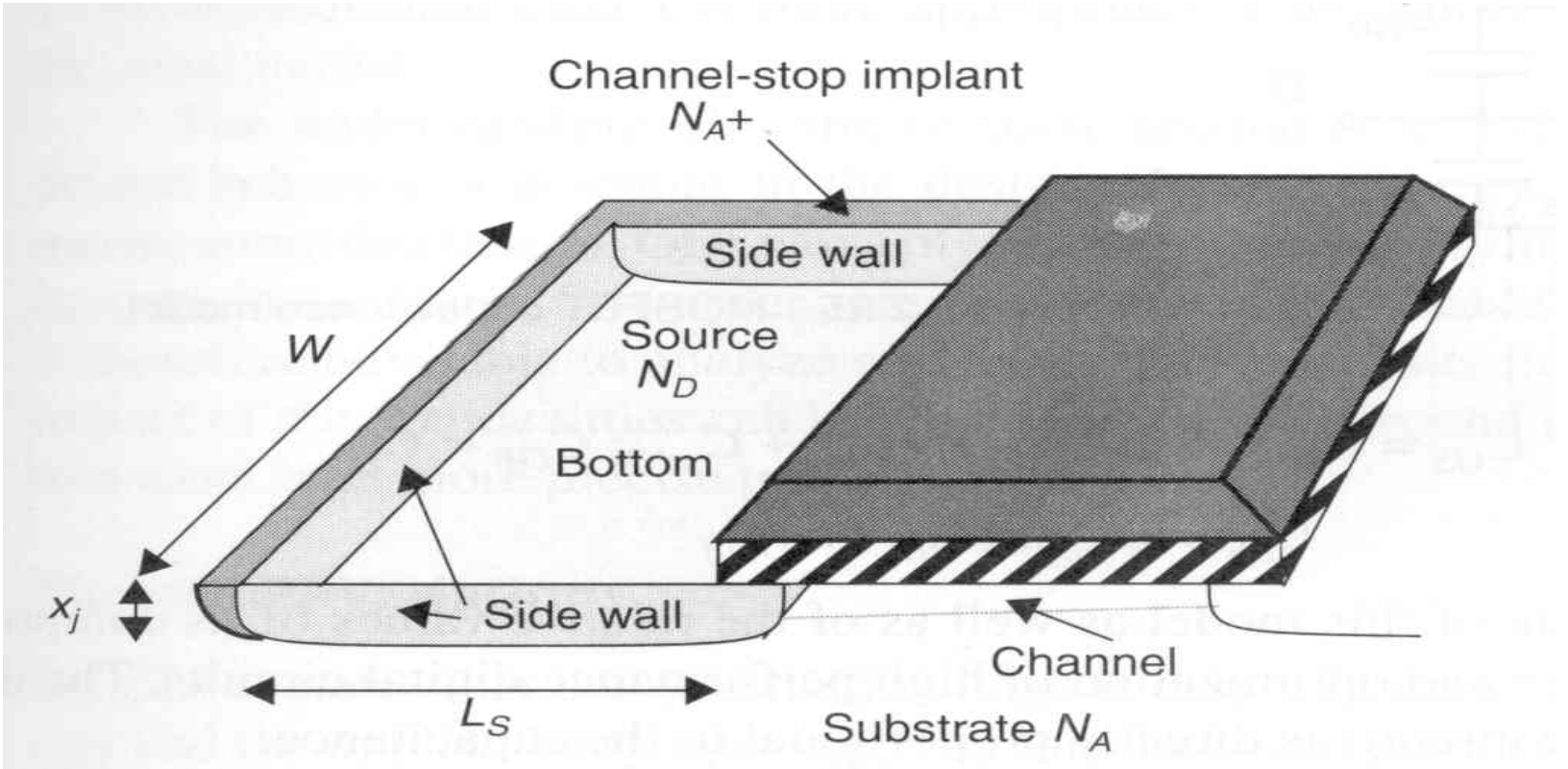
  - Total Gate capacitance:

$$C_G = 2/3\ C_{ox} W L_{eff} + 2 x_d W C_{OX}$$

Oxide Capacitances: Channel



Junction Capacitance



Reverse-biased P-N junctions!
Capacitance depends on reverse-bias voltage.

Junction Capacitance

For a P-N junction: $$C_j = \frac{A}{2}\sqrt{\frac{2q\varepsilon}{V_0 - V}\frac{N_d N_a}{N_d + N_a}}$$

If V=0, cap/area = $$C_{j0} = \sqrt{\frac{q\varepsilon_{Si}}{2V_0}\frac{N_d N_a}{N_d + N_a}}$$

General form: $$C_j = \frac{AC_{j0}}{\left(1 - \frac{V}{V_0}\right)^m}$$

m = grading coefficient (0.5 for abrupt junctions)
(0.3 for graded junctions)

Junction Capacitance

- **Junction with substrate**
  - Bottom area = W * L_S (length of drain/source)
  - Total cap = C_j
- **Junction with sidewalls**
  - "Channel-stop implant"
  - Perimeter = $2L_S + W$
  - Area = P * X_j
  - Total cap = C_{jsw}
- **Total junction cap C = $C_j + C_{jsw}$**

Junction Capacitance

- **Voltage Equivalence Factor**
  - Creates an average capacitance value for a voltage transition, defined as $\Delta Q/\Delta V$

$$C_{eq} = \frac{-AC_{j0}V_0}{(V_2 - V_1)(1-m)}\left(\left(1-\frac{V_2}{V_0}\right)^{1-m} - \left(1-\frac{V_1}{V_0}\right)^{1-m}\right) = AK_{eq}C_{j0}$$

$$K_{eq} = \frac{-2\sqrt{V_0}}{(V_2 - V_1)}\left(\sqrt{V_0 - V_2} - \sqrt{V_0 - V_1}\right) \quad \text{(abrupt junction only)}$$

$$C_{db} = AK_{eq}C_{j0} + PX_jK_{eqsw}C_{jsw0}$$

Example: Junction Cap

- **Consider the following NMOS device**
  - Substrate doping: $N_A = 10^{15}$ cm^{-3}
  - Source/drain doping: $N_D = 2 \times 10^{20}$ cm^{-3}
  - Channel-stop doping: 10X substrate doping
  - Drain length L_D = 1um
  - Transistor W = 10um
  - Junction depth Xj = 0.5um, abrupt junction

- **Find capacitance of drain-bulk junction when drain voltage = 3V**

Next Time: AC Characteristics

- **CMOS Inverters**
  - AC Characteristics: Designing for speed