Digital Integrated Circuit Design

CMPE 530/630

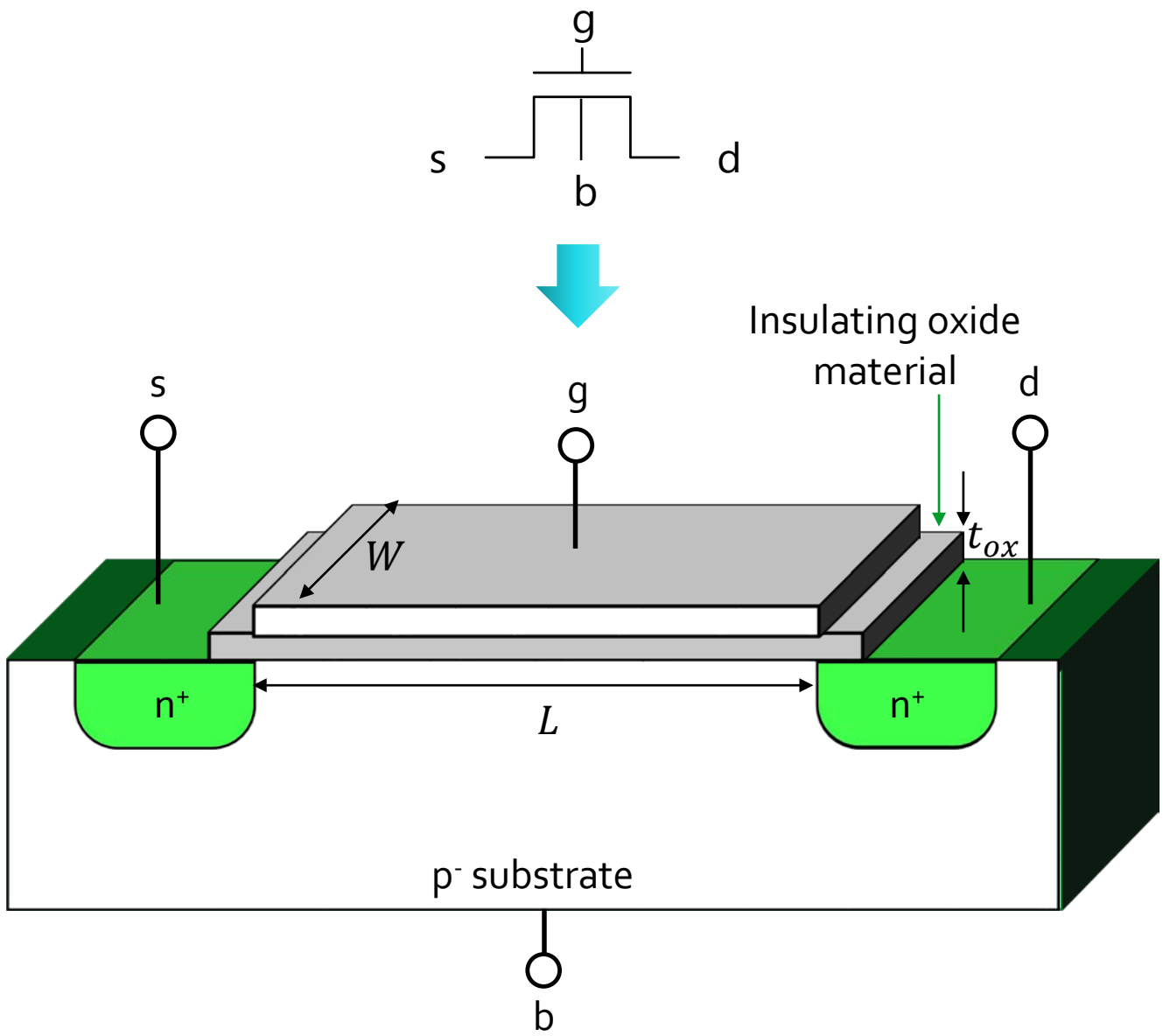
MOSFET Modeling

Dr. Cory Merkel

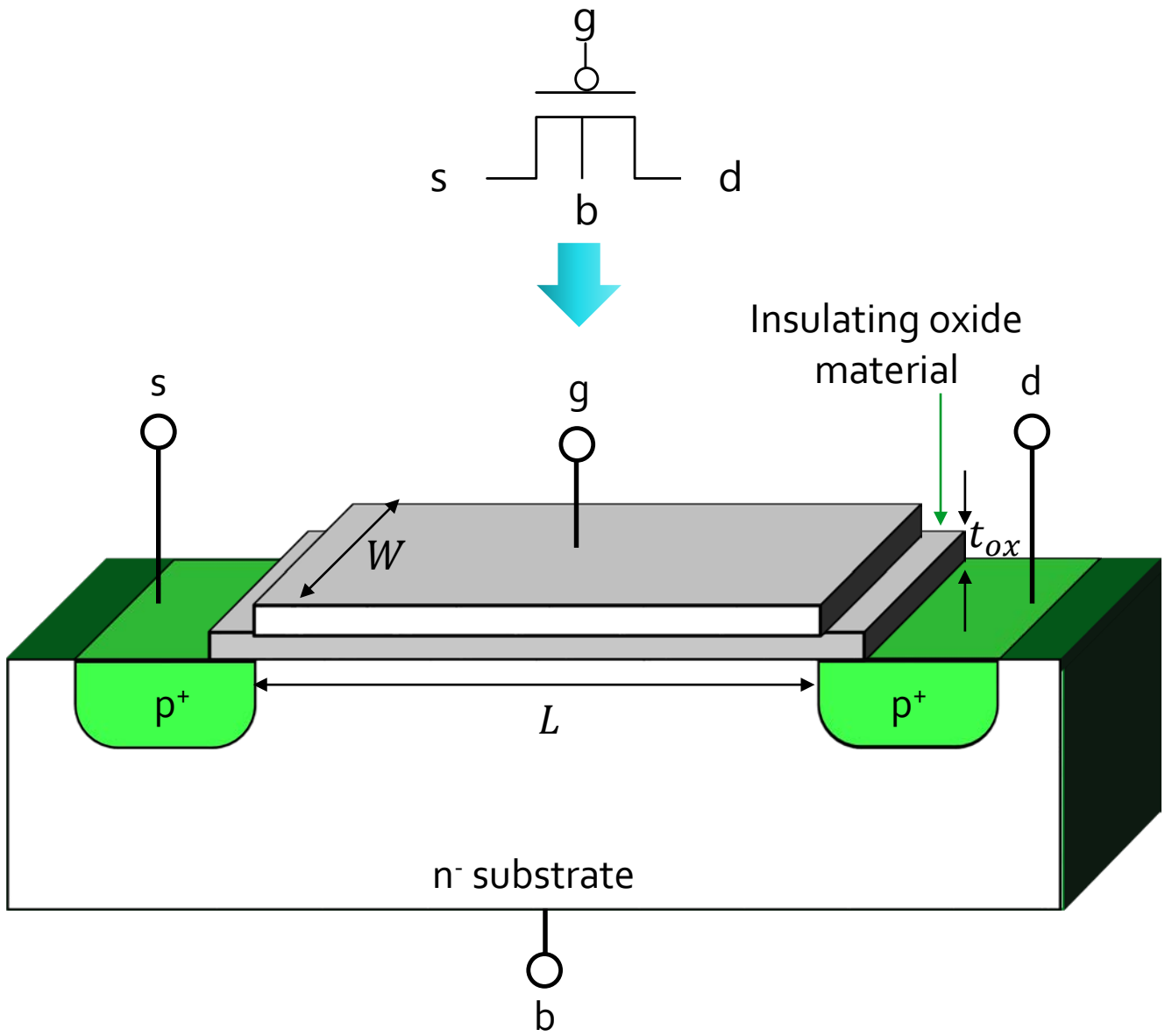
Learning Objectives

- Understand the physical structure of NMOS and PMOS transistors
- Understand the drain current equations for cutoff, linear, and saturation regions of the devices, as well as some of the "non-ideal" effects (e.g. the body effect)
- Be able to apply the drain current equations to understand the DC and transient behavior of complementary CMOS and pass transistor circuits

Physical Realization of nMOSFET

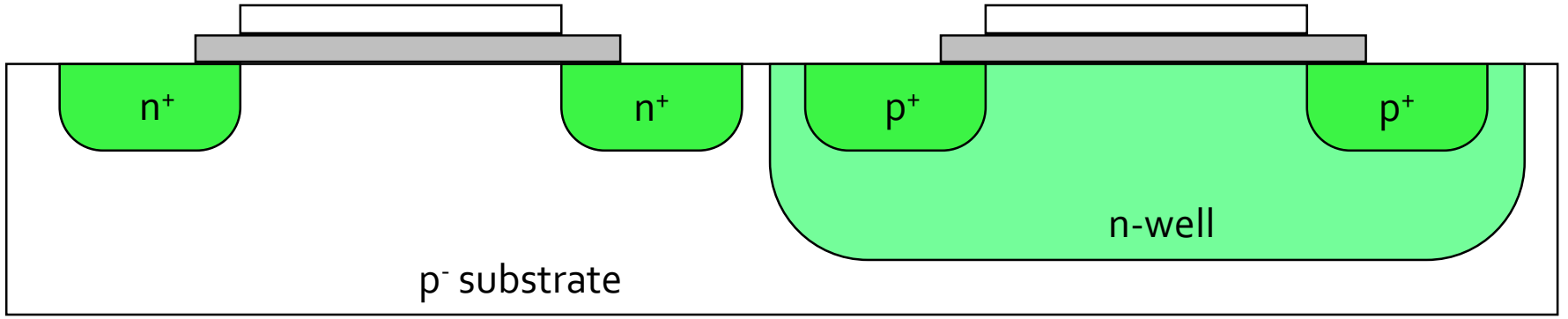


Physical Realization of pMOSFET



NMOS and PMOS Together

» In an *n-well process*, n-type dopants are diffused into the silicon to form a well that acts as the PMOS substrate:

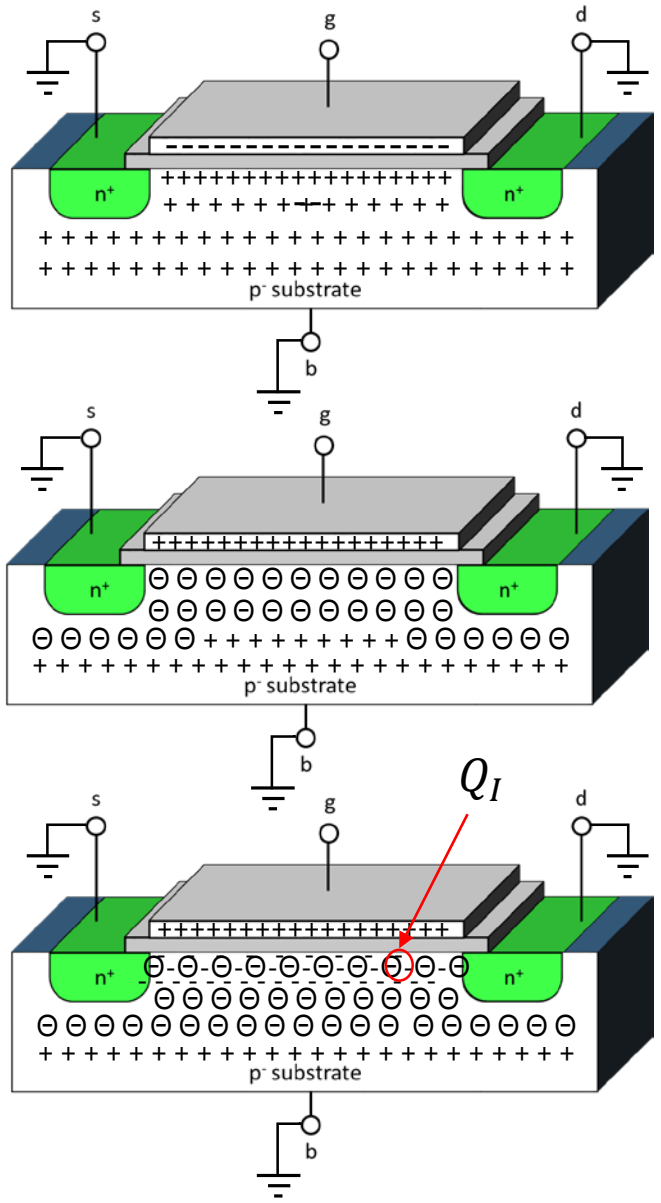


» We always connect NMOS bodies to ground and PMOS bodies to V_{dd} to avoid current flowing through the substrate

- We want all of the p-n junctions to be reverse-biased

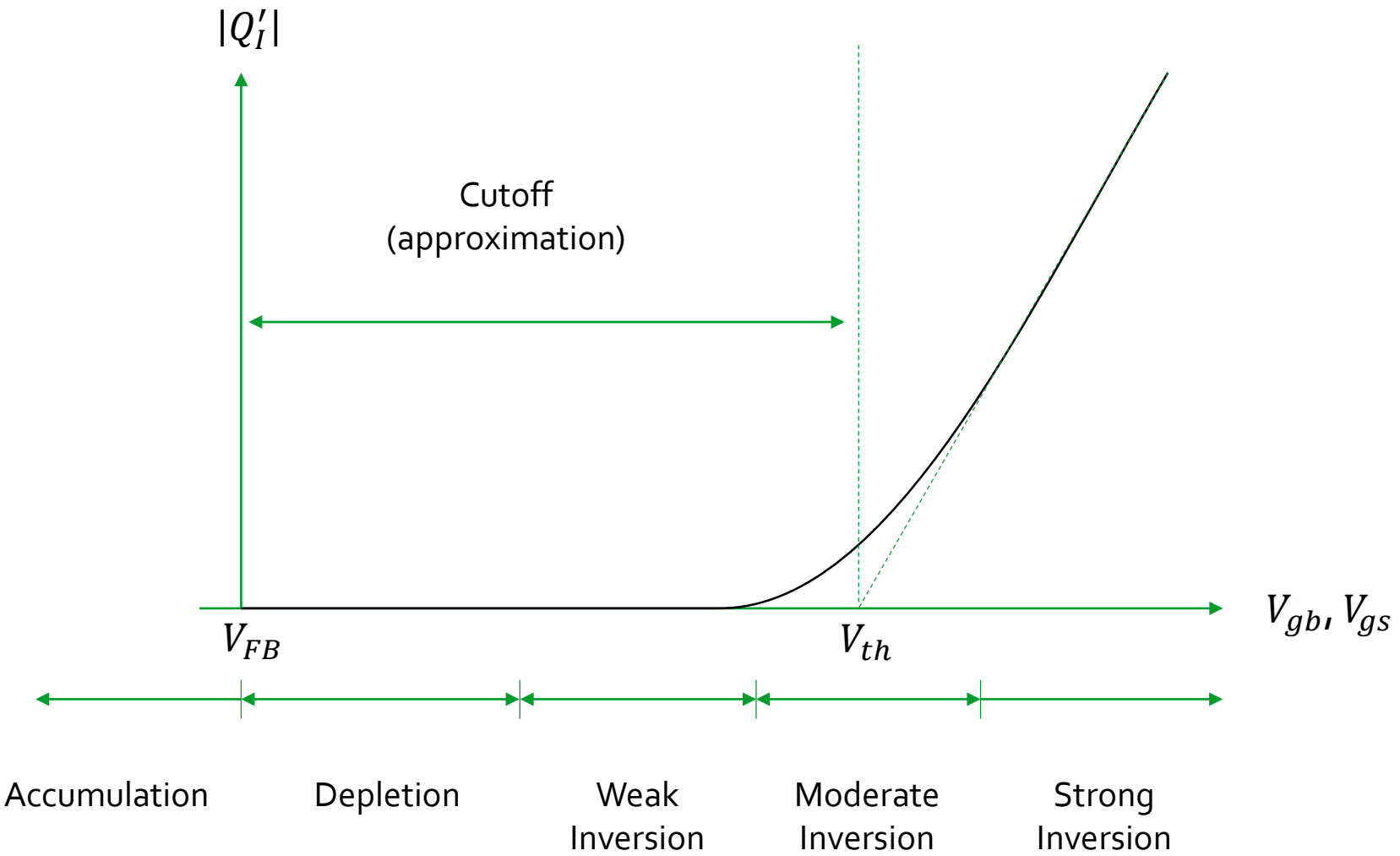
Regions of Operation

- Accumulation
  - $V_g \lesssim 0$
  - Holes accumulate at the surface
- Depletion (Weak/Moderate Inversion)
  - $0 \lesssim V_g < V_{th}$
  - Holes are repelled into bulk
  - Negatively charged dopant ions left behind
- (Strong) Inversion
  - $V_g \geq V_{th}$
  - Free electrons attracted to surface from source, inverting the channel
    - Becomes n-type



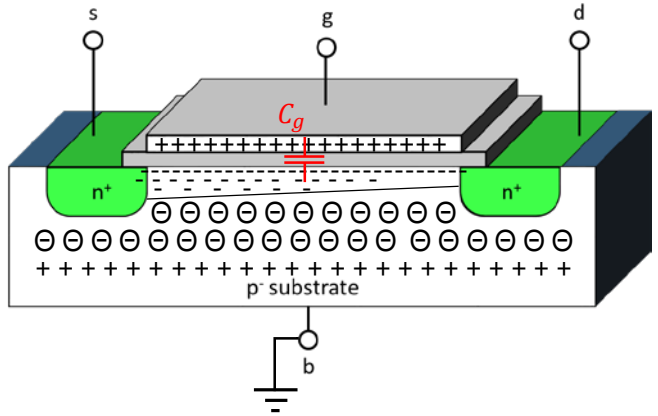
The Threshold Voltage

» Threshold voltage occurs at point where surface electron concentration ≈ hole concentration in p substrate (NMOS)



Channel Charge

- $Q_I = (V_{gc} - V_{th})C_g$
  $= (V_{gc} - V_{th})C_{ox}WL$
- $V_{gc} = V_g - V_c = V_g - (V_s + V_d)/2$
  $= V_{gs} - V_{ds}/2$



- C_{ox} is the oxide capacitance per unit area:

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = \frac{k_{ox}\epsilon_0}{t_{ox}}$$

- k_{ox} is the relative permittivity: ≈ 3.9 for SiO_2
- ϵ_0 is the permittivity of free space: 8.85×10^{-14}F/cm

Drain Current

- The velocity of electrons traveling in the channel depends on the lateral electric field E as

$$v = \mu_n E$$

- μ_n is a proportionality constant called the electron mobility
  - μ_n is measured in $\mathrm{cm}^2/\mathrm{Vs}$
- The average lateral electric field in the channel is

$$E = \frac{V_{ds}}{L}$$

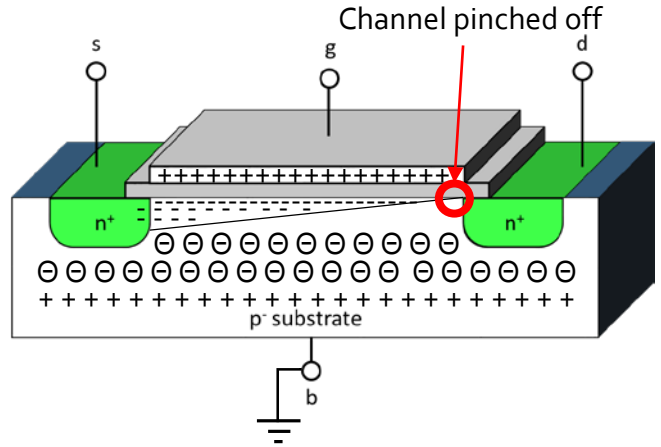
- $I_{ds} = \frac{Q_I}{L/v} = \frac{(V_{gs} - V_{th} - V_{ds}/2) C_{ox} W L}{L/(\mu_n V_{ds}/L)} = \beta_n (V_{eff} - V_{ds}/2) V_{ds}$

Where $\beta_n \equiv \mu_n C_{ox} W/L$, and $V_{eff} \equiv V_{gs} - V_{th0}$

Drain Current

» When $V_{ds} \ll V_{eff}$:

$$I_{ds} \approx \beta_n V_{eff} V_{ds}$$



» When $V_{ds} \geq V_{ds,sat} \equiv V_{eff}$,
The channel becomes "pinched off" at the drain end and the current becomes saturated at

$$I_{ds,sat} = I_{ds}(V_{gs} = V_{eff}) = \frac{1}{2}\beta_n V_{eff}^2$$

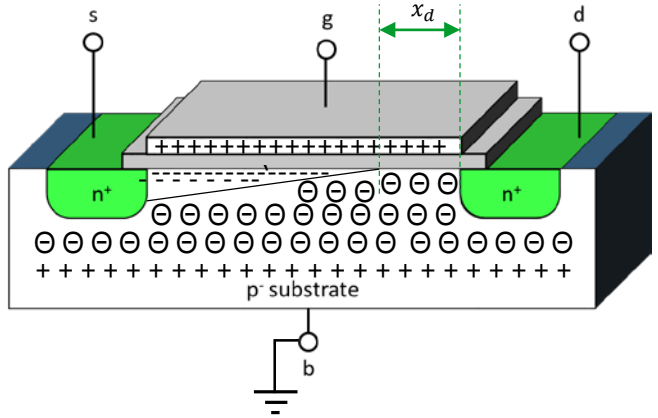
Summary

- PMOS devices
  - μ_n become μ_p (mobility of holes)
    - Usually a few times smaller than μ_n
  - β_n becomes β_p
  - V_{gs} becomes V_{sg} or $|V_{gs}|$
  - V_{th} becomes $|V_{th}|$
- Long channel (**Shockley Model**) characteristics:

Region	Conditions	NMOS	PMOS
Cutoff	$\|V_{gs}\| < \|V_{th}\|$	$I_{ds} = 0$	$I_{ds} = 0$
Linear	$\|V_{gs}\| \geq \|V_{th}\|, V_{ds} \ll V_{eff}$	$I_{ds} = \beta_n V_{eff} V_{ds}$	$I_{ds} = \beta_p V_{eff} V_{ds}$
Triode	$\|V_{gs}\| \geq \|V_{th}\|, V_{ds} < V_{eff}$	$I_{ds} = \beta_n \left(V_{eff} - \frac{V_{ds}}{2}\right) V_{ds}$	$I_{ds} = \beta_p \left(V_{eff} - \frac{V_{ds}}{2}\right) V_{ds}$
Saturation	$\|V_{gs}\| \geq \|V_{th}\|, V_{ds} \geq V_{ds,sat} = V_{eff}$	$I_{ds} = I_{ds,sat} = \frac{1}{2}\beta_n V_{eff}^2$	$I_{ds} = I_{ds,sat} = \frac{1}{2}\beta_p V_{eff}^2$

Channel Length Modulation

- In saturation, I_{ds} actually has small dependence on V_{ds}
- As V_{ds} increases, it counteracts the gate voltage, causing depletion around the drain
  - Effective channel length: $L - x_d$
- Taylor expand I_{ds} around V_{eff}:



$$I_{ds} \approx I_{ds,sat} + \frac{\partial I_{ds}}{\partial L}\frac{\partial L}{\partial V_{ds}}\Delta V_{ds} = I_{ds,sat} - \left(\frac{1}{2}\frac{\beta_n}{L}V_{eff}^2\right)\left(-\frac{\partial x_d}{\partial V_{ds}}\right)\Delta V_{ds}$$

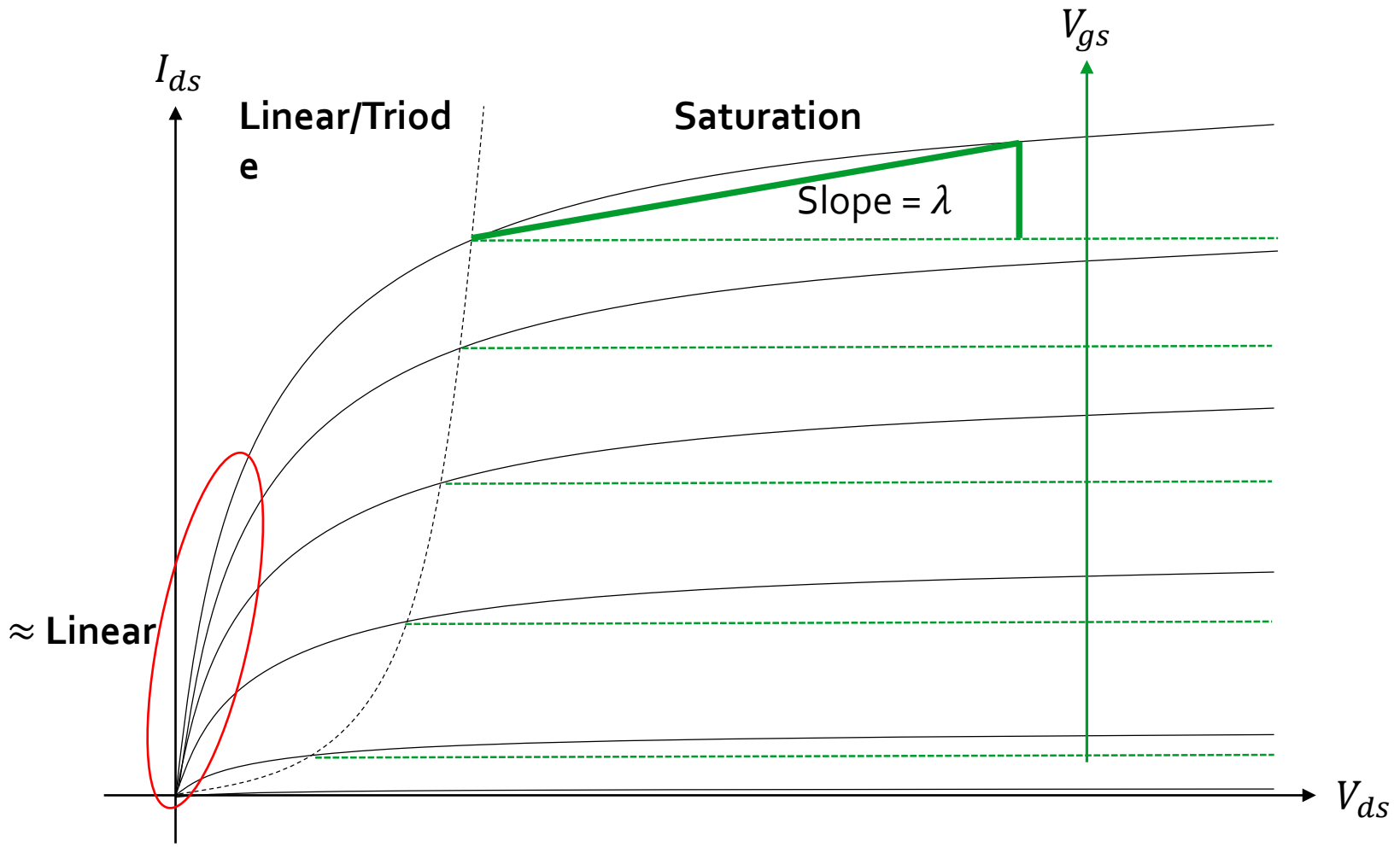
$$= I_{ds,sat}\left(1 + \frac{1}{L}\frac{\partial x_d}{\partial V_{ds}}\left(V_{ds} - V_{eff}\right)\right)$$

So,

$$I_{ds} = \frac{1}{2}\beta_n V_{eff}^2\left(1 + \lambda\left(V_{ds} - V_{eff}\right)\right)$$

λ is the channel length modulation constant with units of V^{-1}

MOSFET I-V Characteristics



Ideal saturation (I_{ds} constant)

Leakage

» Even when $V_{gs} < V_{th}$, MOSFETs still "leak" current
  - Carriers diffuse from source to drain

$$I_{ds} = \beta V_T^2 (n-1) e^{\frac{V_{eff}}{nV_T}} \left(1 - e^{-\frac{V_{ds}}{V_T}}\right)$$

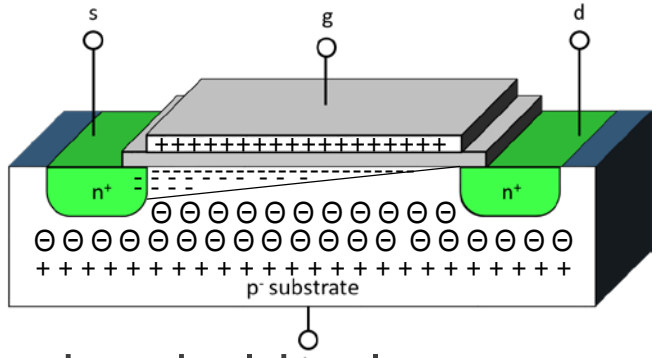
where n is the subthreshold slope factor and $V_T = k_B T/q$ is the thermal voltage

» Subthreshold slope:

$$S = \left[\frac{\mathrm{d}(\log_{10} I_{ds})}{\mathrm{d}V_{gs}}\right]^{-1} = nV_T \ln 10$$

Body Effect

- So far, we have ignored the MOSFET body
- Body acts as a second, weaker gate



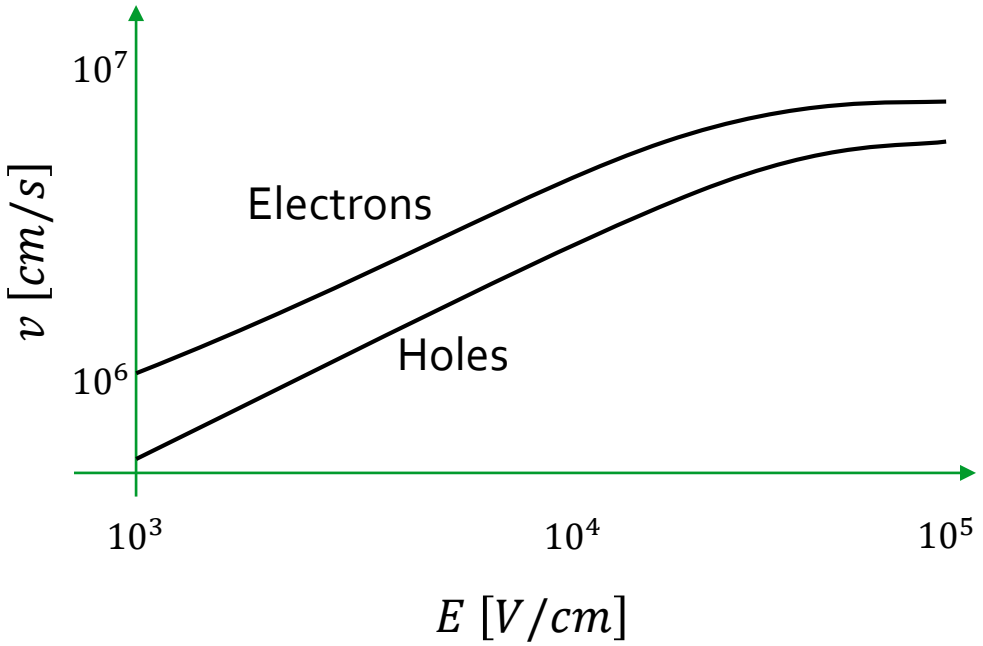
- As V_{sb} is increased, it masks V_{gs}, changing the effective threshold voltage:

$$V_{th} = V_{th0} + \gamma\left(\sqrt{V_{sb} + |2\phi_F|} - \sqrt{|2\phi_F|}\right)$$

where $\gamma = \frac{\sqrt{2qN_A k_s \epsilon_0}}{C_{ox}}$ is the body effect coefficient. k_s is the relative permittivity of silicon (≈ 11.8)

Velocity Saturation

- One of the main culprits in discrepancies between long channel model and behavior observed in short channel devices
- As lateral electric field becomes large, scattering causes carrier velocity to saturate:



Carrier velocity saturates around 10^5 V/cm.

Velocity Saturation

» The velocity saturation is modeled well with:

$$v = \begin{cases} \dfrac{\mu_{eff} E}{1 + E/E_c} & E < E_c \\ v_{sat} & E \geq E_c \end{cases}$$

Where $E_c = 2v_{sat}/\mu_{eff}$ is the critical field

» Then, the critical voltage is $V_c = E_c L$
- This is the drain-source voltage where the carrier velocity saturates

Velocity Saturation

» Now, the drain current in saturation can be approximated as (see if you can derive this):

$$I_{ds,sat} = W C_{ox} v_{sat} \frac{V_{eff}^2}{V_{eff} + V_c}$$

» Note that this is approximately **linear** in V_{eff} when $V_{eff} \gg V_c$

Other Short Channel/Non-Ideal Effects

- Drain-Induced Barrier Lowering (DIBL)
  - Drain voltage reduces threshold voltage
- Mobility Degradation
  - Carriers in channel are pulled against the oxide, reducing the mobility
- Velocity Saturation
  - Carriers "bump into" atoms in the channel
  - The faster they try to go, the more they bump
- Gate Leakage
  - Charge tunnels through the gate oxide

MOSFET Effective Resistance

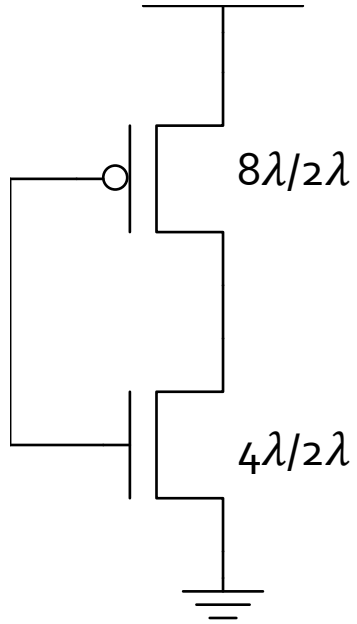
» MOSFET has effective channel resistance or drive strength:

$$R_{ds} \propto \frac{1}{\beta_n} = \frac{1}{\mu_n C_{ox} W/L}$$

» We want the PUN and PDN to have equal drive strength
  - This can be accomplished through sizing

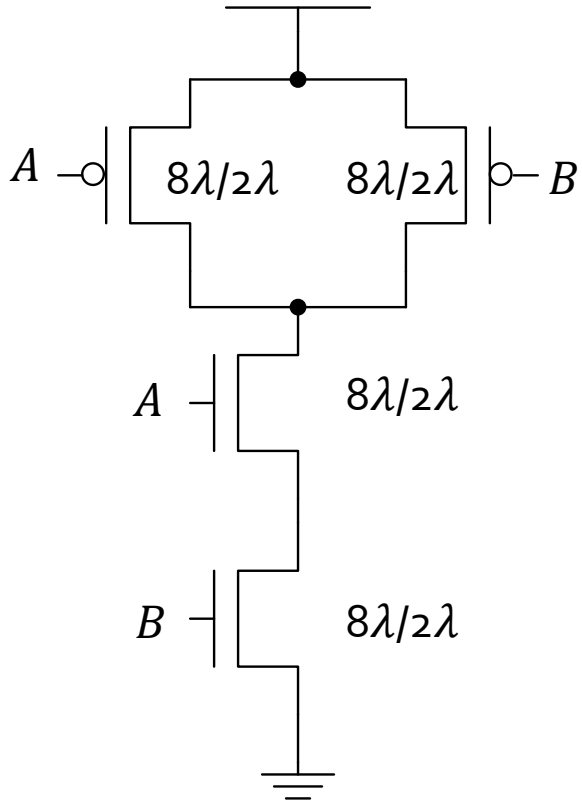
Example: CMOS Inverter

- Assume $\mu_n = 2\mu_p$
- We want $\beta_n = \beta_p$ for equal drive strength
- First, we size the NMOS length and width to have minimal sizing = 2λ, where λ is the *feature size*
- Second, size PMOS to have minimal length
- Third, size PMOS width so $\beta_p = \beta_n$
- This is called the **unit inverter**



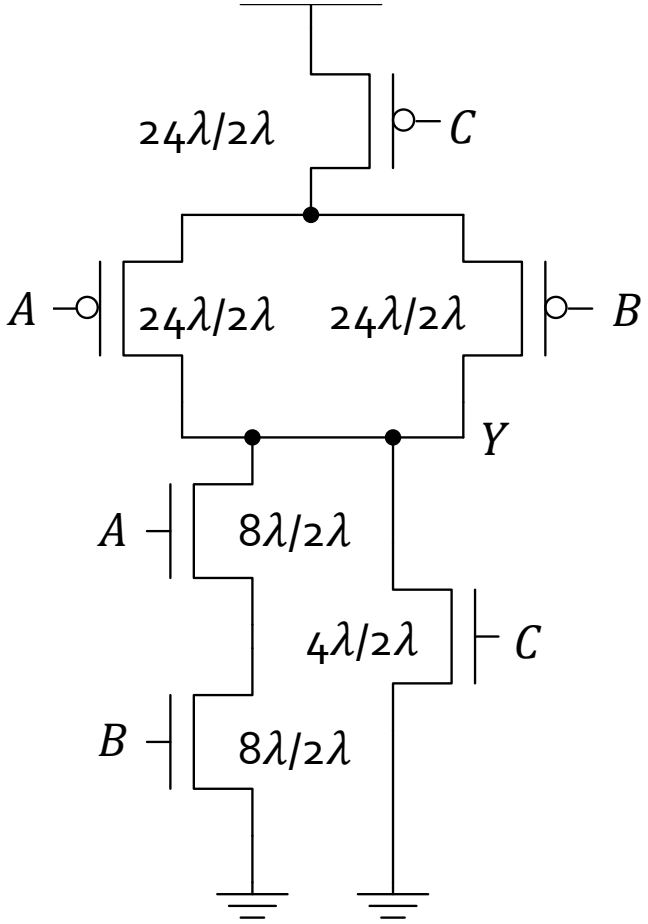
Example: NAND

- We want the *worst-case* drive strength to be identical to the unit inverter
  - Widths are doubled for series transistors
  - Widths are not changed for parallel transistors



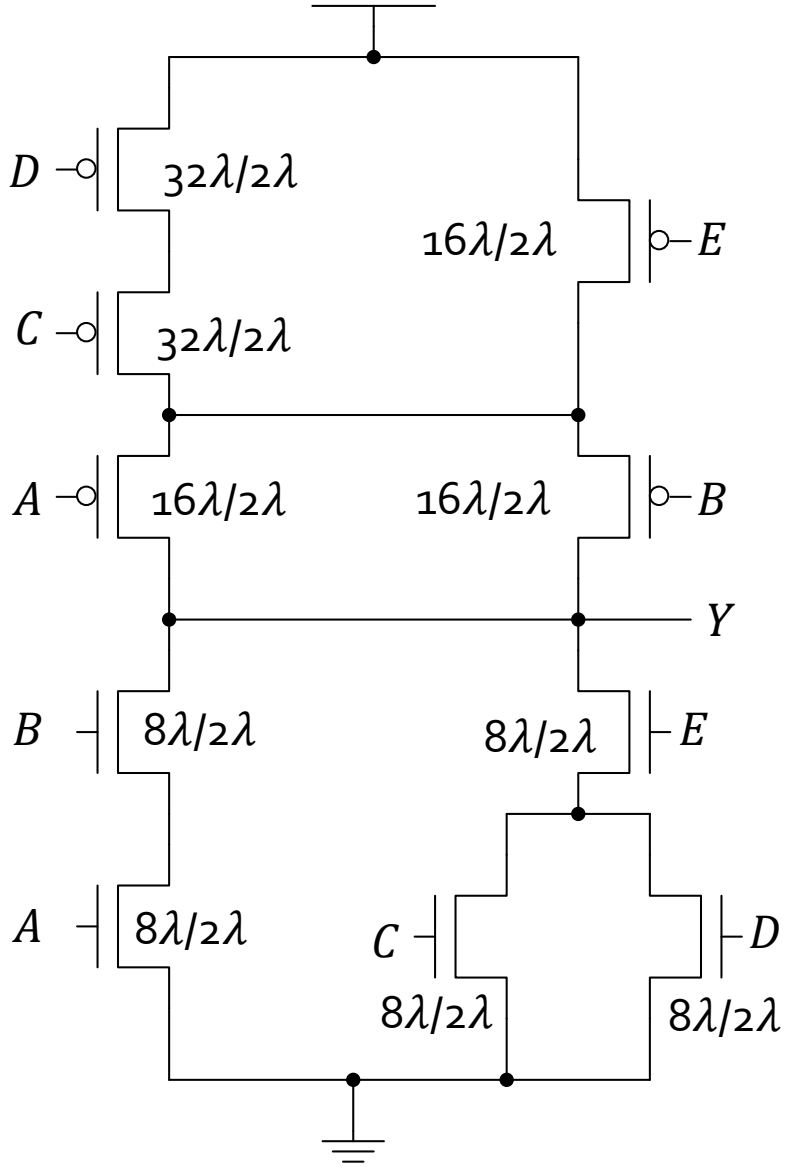
Example: $Y = \overline{AB + C}$

» This time we assume $\mu_n = 3\mu_p$

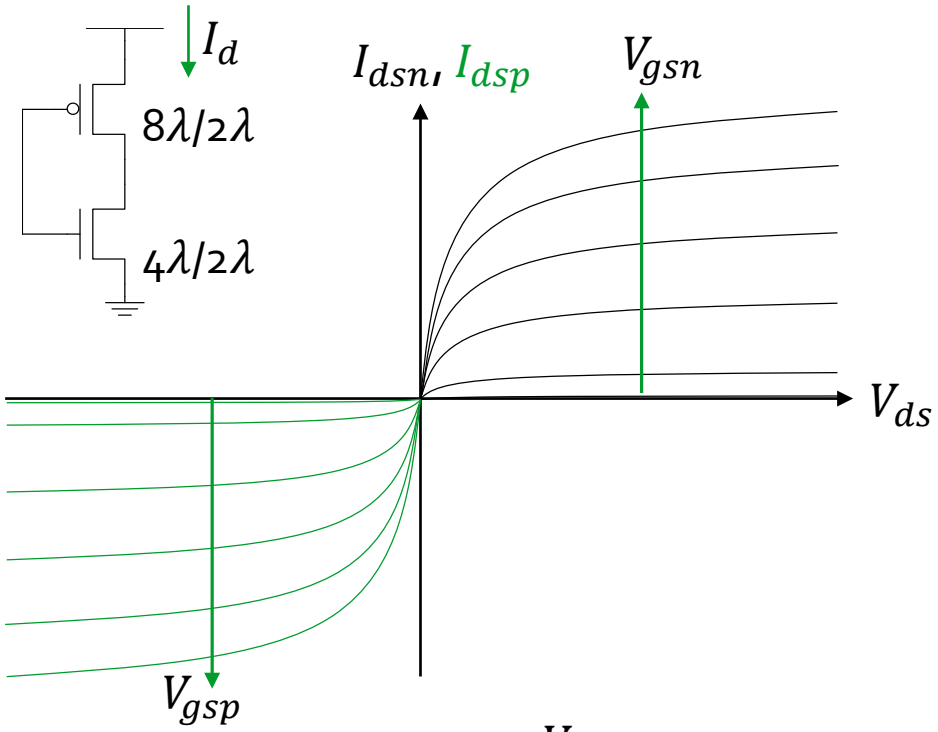


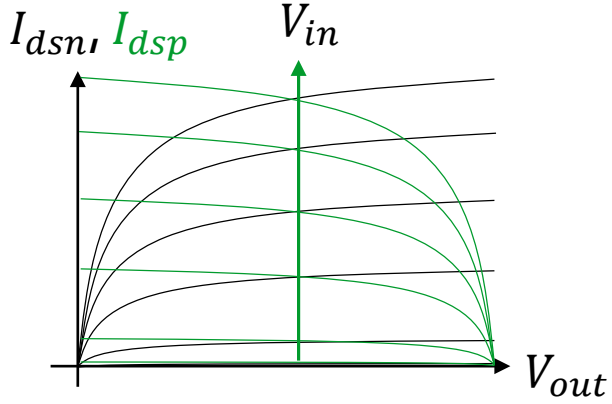
Example: $Y = \overline{AB + (C + D)E}$

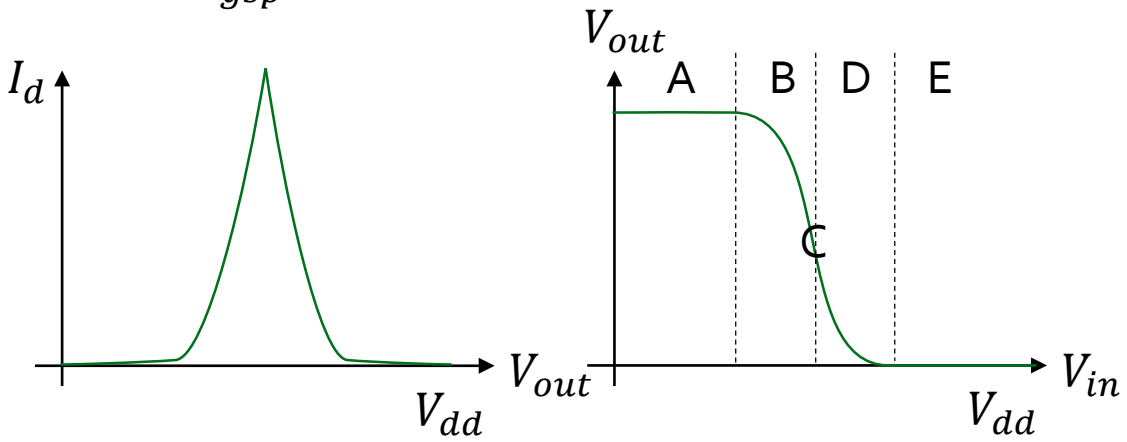
» $\mu_n = 2\mu_p$



CMOS Inverter DC Characteristics



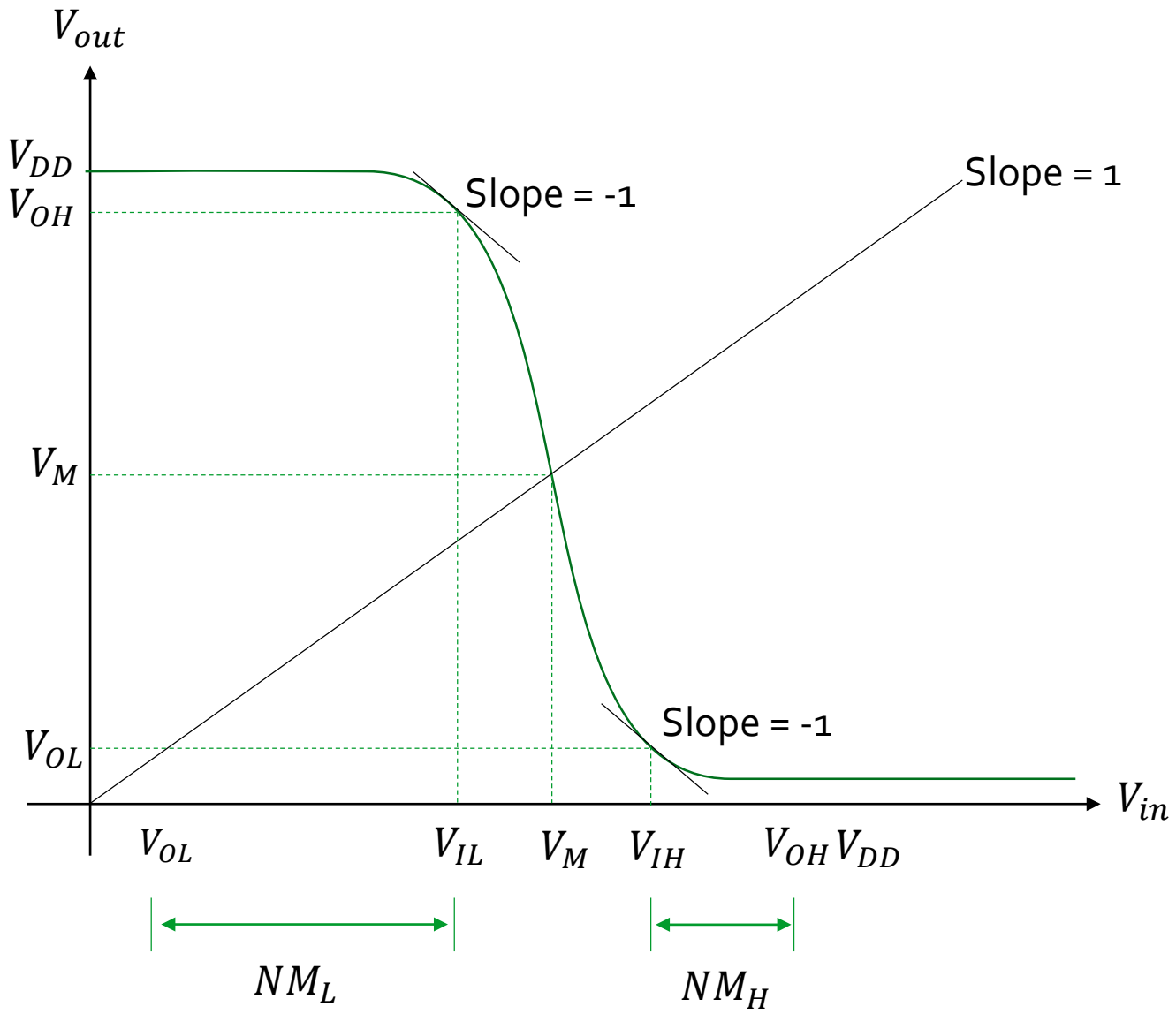




Operating Regions of Transistors in CMOS Inverter:

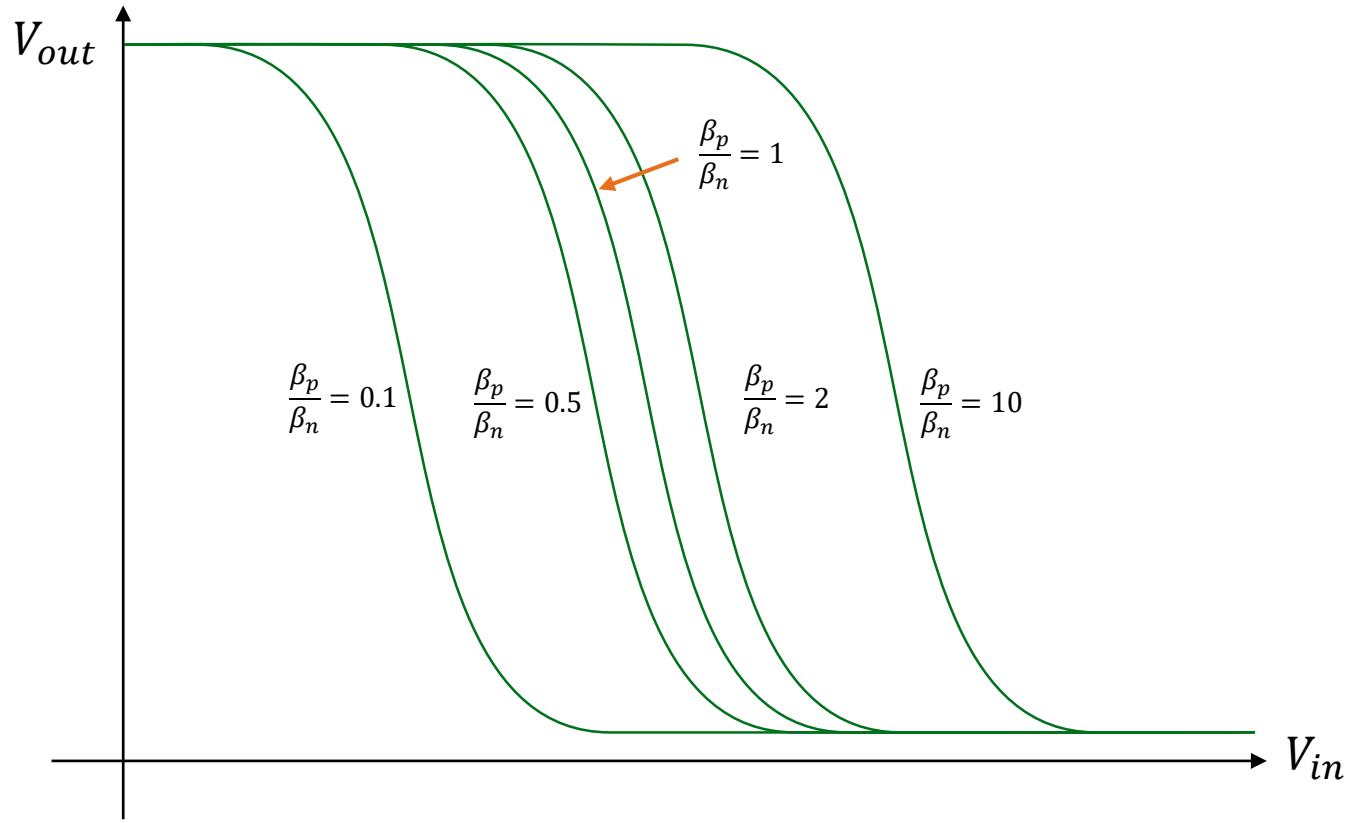
Region	NMOS Conditions	PMOS Conditions	NMOS	PMOS
A	$V_{gs} < V_{th}$	$V_{gs} > V_{th}$ $V_{ds} < V_{eff}$	Cutoff	Linear
B	$V_{gs} > V_{th}$ $V_{ds} \geq V_{eff}$	$V_{gs} > V_{th}$ $V_{ds} < V_{eff}$	Saturated	Linear
C	$V_{gs} > V_{th}$ $V_{ds} \geq V_{eff}$	$V_{gs} > V_{th}$ $V_{ds} \geq V_{eff}$	Saturated	Saturated
D	$V_{gs} > V_{th}$ $V_{ds} < V_{eff}$	$V_{gs} > V_{th}$ $V_{ds} \geq V_{eff}$	Linear	Saturated
E	$V_{gs} > V_{th}$ $V_{ds} < V_{eff}$	$V_{gs} < V_{th}$	Linear	Cutoff

Critical Points of Inverter VTC



Different Beta Ratios

» Relative drive strength can be expressed as beta ratios
  - Large ratio → Stronger PUN, Smaller ratio → Stronger PDN
  - Gates with beta ratios ≠ o are called *skewed*



CMOS Inverter Transient Characteristics

t_{pdf} - Propagation delay when output is falling

t_{pdr} - Propagation delay when output is rising

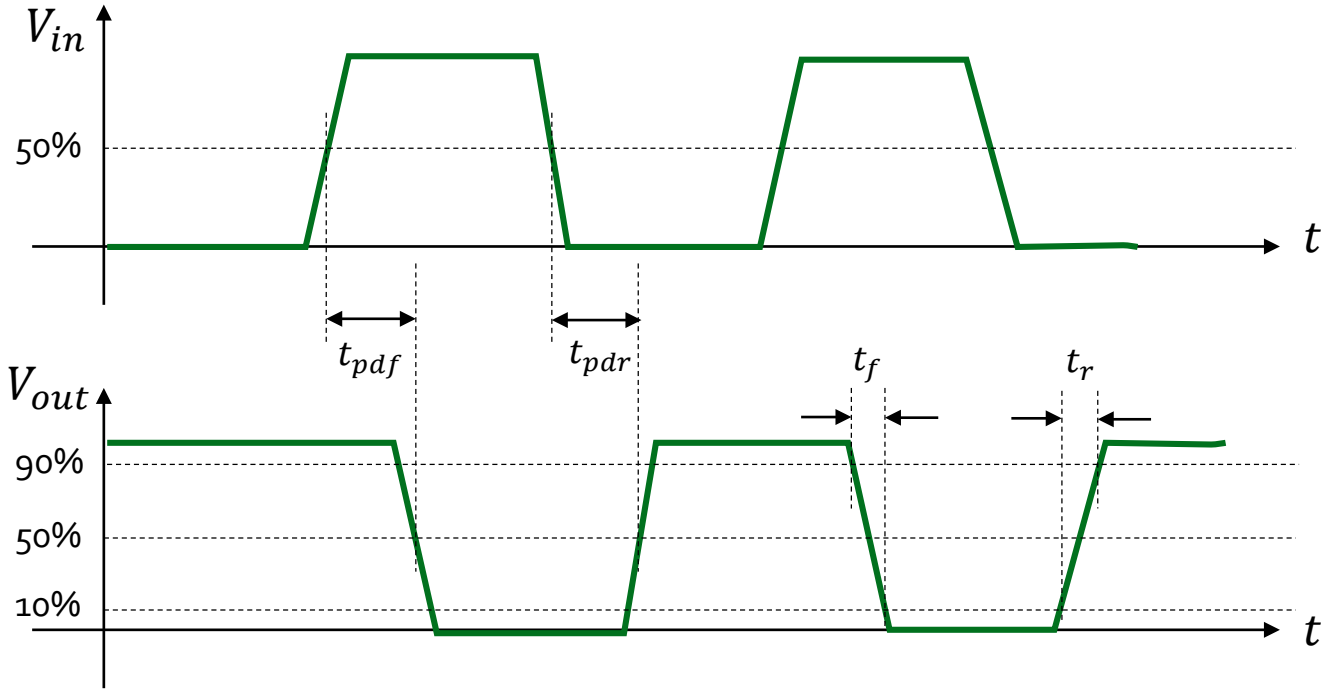
$t_{pd} = (t_{pdf} + t_{pdr})/2$ – Avg. Prop. delay

t_r - Rise time of output

t_f - Fall time of output

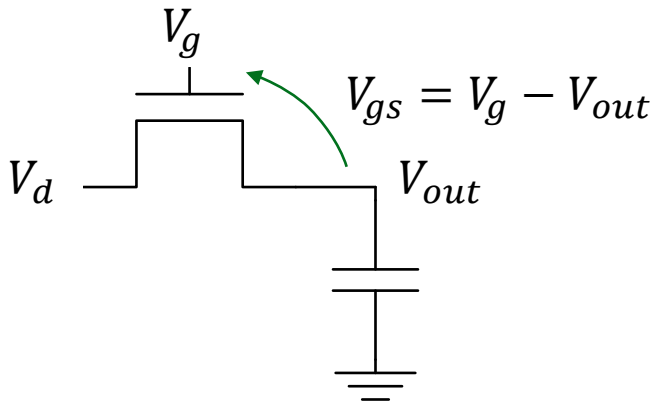
$f_{input,max} = \frac{1}{t_r + t_f}$ - Maximum input frequency

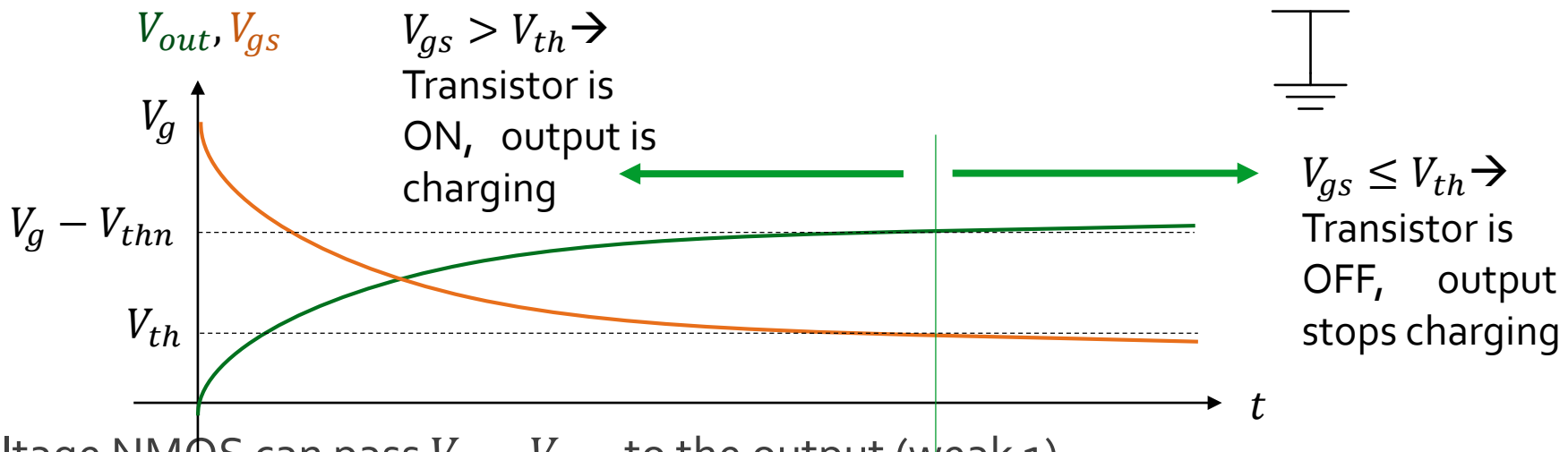
$f_{throughput,max} = \frac{1}{t_{pdf} + t_{pdr}}$ - Maximum output frequency



Pass Transistors Revisited

» Consider the circuit:
  • Assume $V_{out} = 0$ at $t = 0$





» Highest voltage NMOS can pass $V_g - V_{thn}$ to the output (weak 1)

» Similarly, lowest voltage PMOS can only pass is $V_g + V_{thp}$ (weak 0)

For More Information

- Excellent resource on detailed transistor modeling (beyond the scope of this course):
  - *Operation and Modeling of The MOS Transistor* by Yannis Tsividis