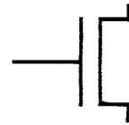
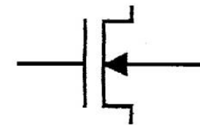
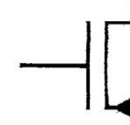
MOS Transistor Theory

Outline

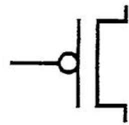
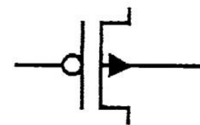
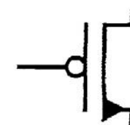
1. **Introduction**
2. Ideal I-V Characteristics
3. Nonideal I-V Effects
4. C-V Characteristics
5. DC Transfer Characteristics
6. Switch-level RC Delay Models

Introduction

- MOS Transistor symbols
  - nMOS symbols

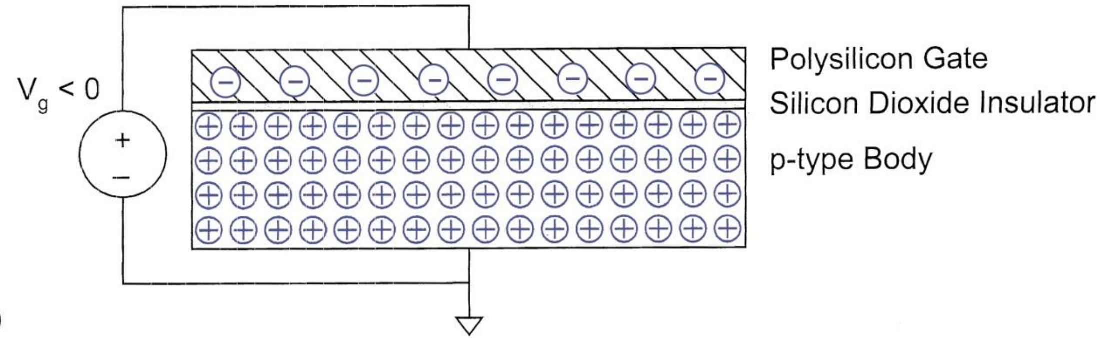




  - pMOS symbols

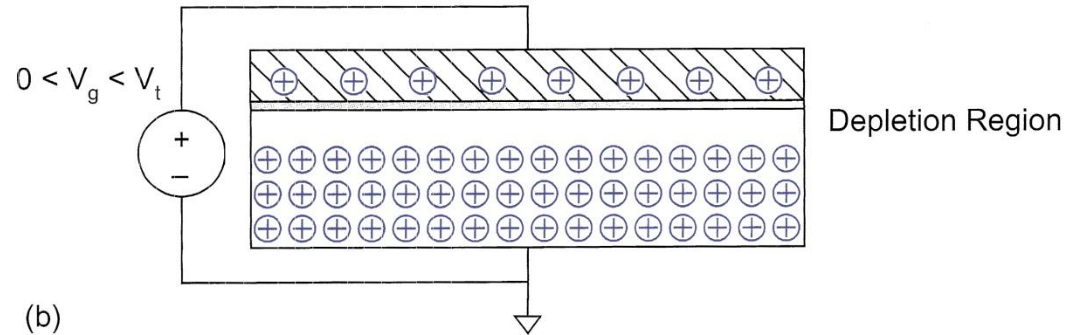




MOS Structure

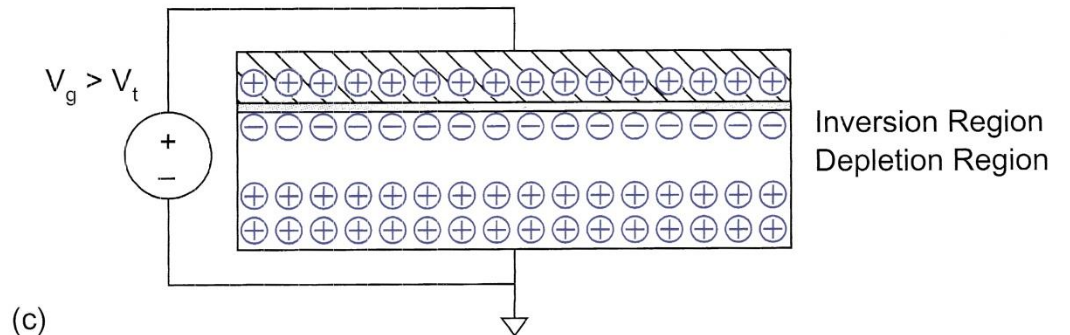
– Accumulation Mode



– Depletion Mode

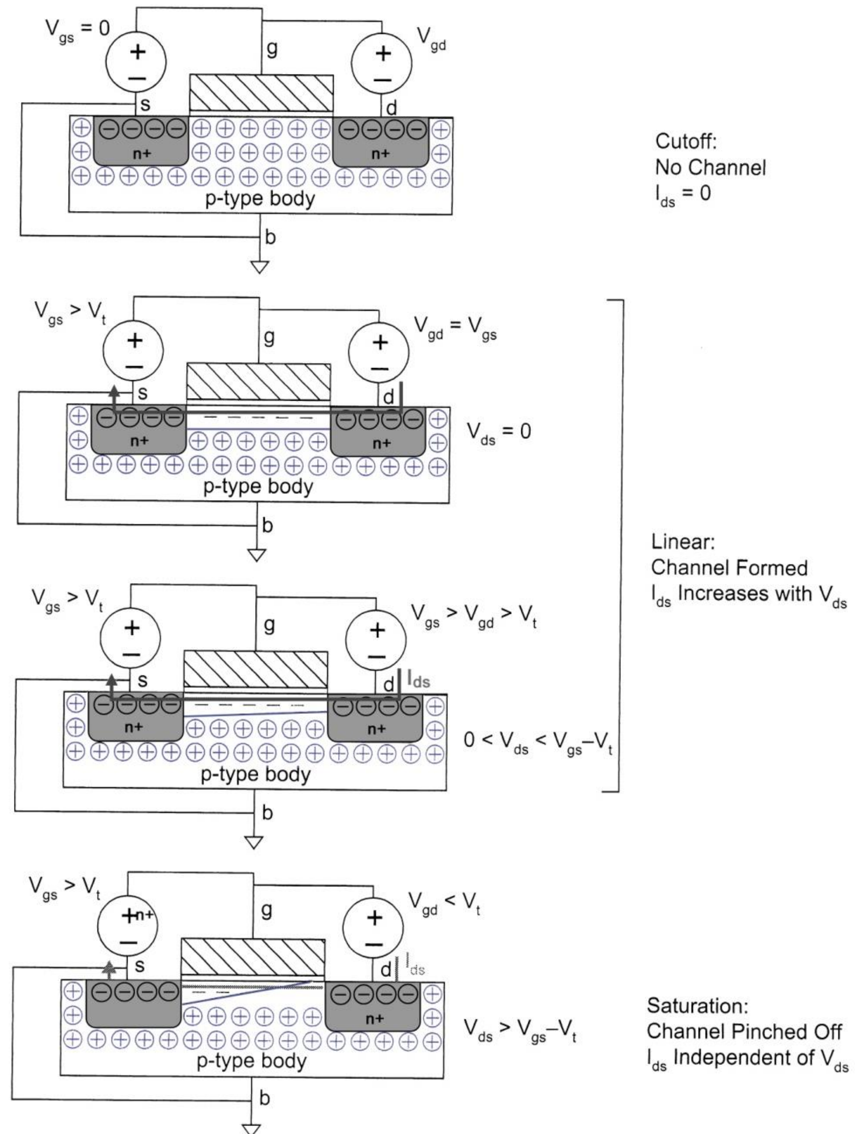


– Inversion Mode



nMOS Operation Regions

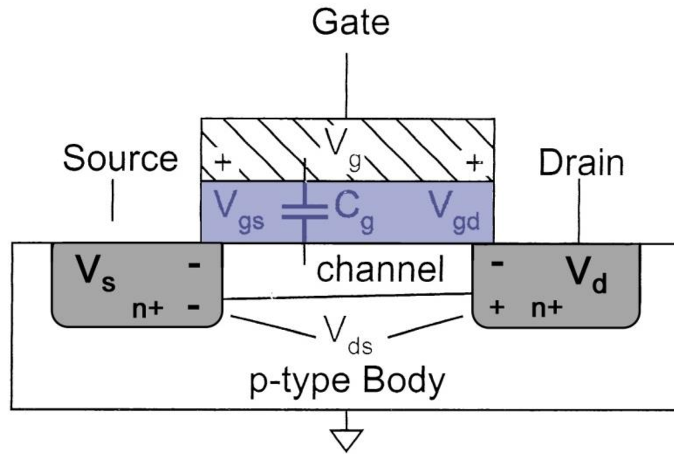
- Cutoff
  - $V_{gs} < V_t$: no channel
- Linear
  - $V_{gs} > V_t$, $V_{gd} > V_t$: linear resistor
- Saturation
  - $V_{gs} > V_t$, $V_{gd} < V_t$: channel pinchoff



Outline

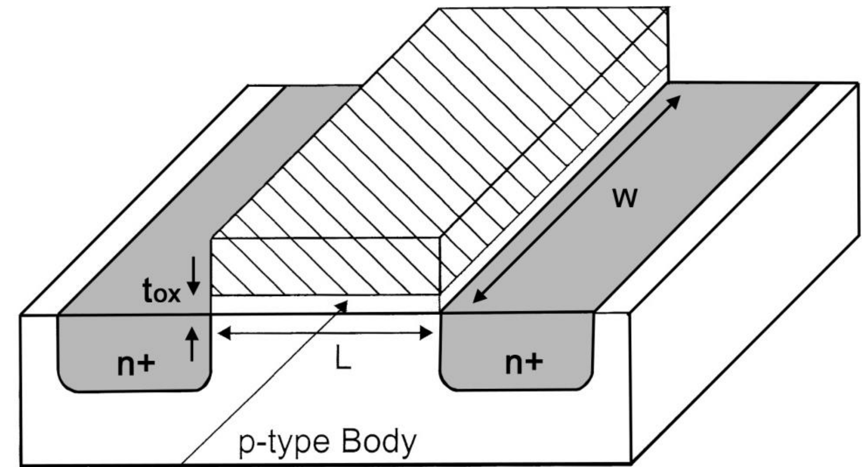
1. Introduction
2. **Ideal I-V Characteristics**
3. Nonideal I-V Effects
4. C-V Characteristics
5. DC Transfer Characteristics
6. Switch-level RC Delay Models

MOS Channel Charge



Average gate to channel potential:

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



SiO_2 Gate Oxide
(Good insulator, $\varepsilon_{ox} = 3.9\varepsilon_0$)

$$Q_{channel} = C_g\left(V_{gc} - V_t\right) = C_g\left(V_{gs} - \tfrac{1}{2}V_{ds} - V_t\right)$$

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

$$\upsilon = \mu E, E = V_{ds} / L, I_{ds} = \frac{Q_{channel}}{t_{channel}}, t_{channel} = \frac{L}{\upsilon}$$

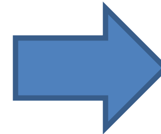
Ideal I-V Equations

$$Q_{channel} = C_{ox} WL\left(V_{gs} - \tfrac{1}{2}V_{ds} - V_t\right)$$

$$\upsilon = \mu E, E = V_{ds} / L,$$

$$\frac{L}{\upsilon} = \frac{L}{\mu E} = \frac{L^2}{\mu V_{ds}}$$

$$I_{ds} = \frac{Q_{channel}}{t_{channel}}, t_{channel} = \frac{L}{\upsilon}$$



- Ideal I-V equation

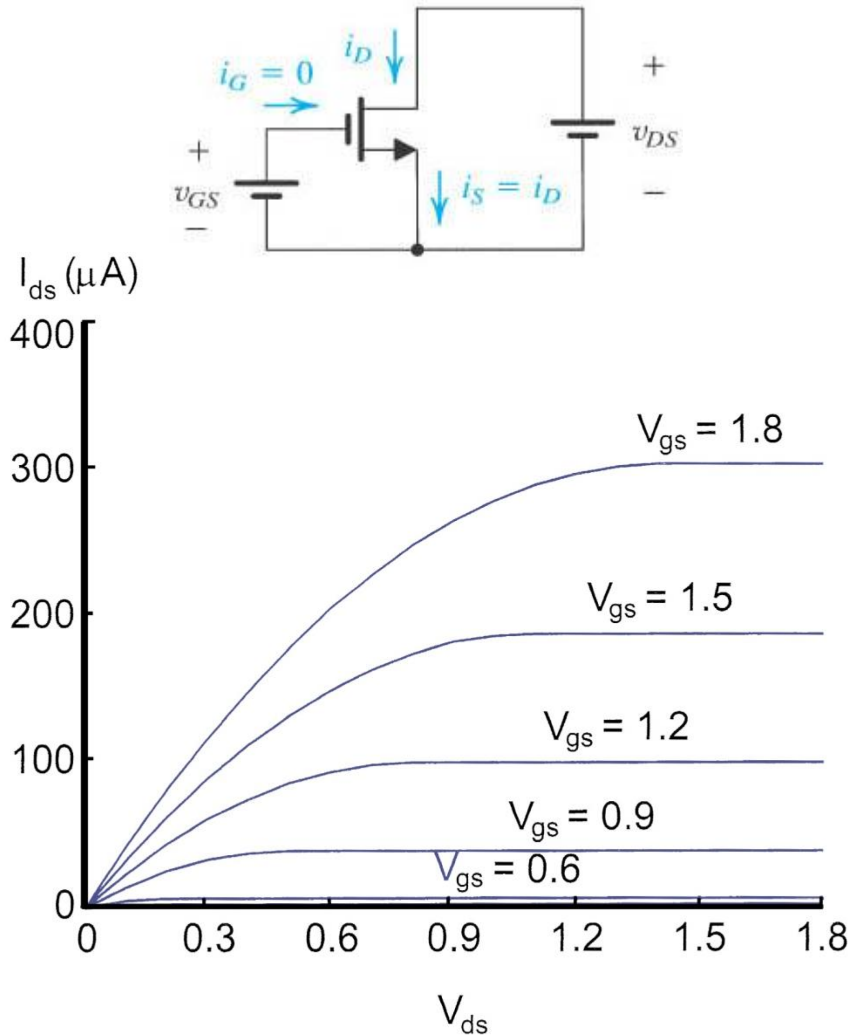
$$I_{ds} = \frac{Q_{channel}}{t_{channel}} = \frac{Q_{channel}}{L/\upsilon}$$

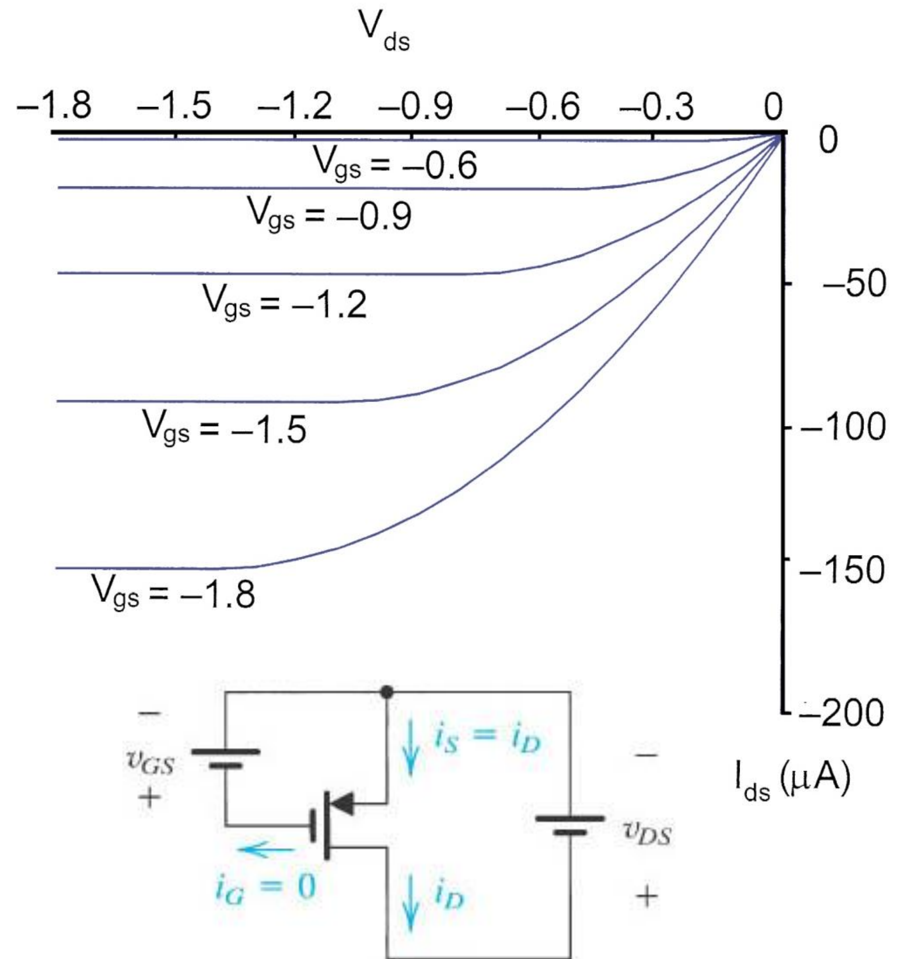
$$= \mu C_{ox} \frac{W}{L}\left(V_{gs} - V_t - \frac{V_{ds}}{2}\right)V_{ds}$$

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

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

Ideal I-V Characteristics





- nMOS I-V
- pMOS I-V

Outline

1. Introduction
2. Ideal I-V Characteristics
3. **Nonideal I-V Effects**
4. C-V Characteristics
5. DC Transfer Characteristics
6. Switch-level RC Delay Models

Nonideal I-V Effects

- **Velocity saturation**: at high V_{ds} ↗, the carrier velocity is not proportional to lateral field. I_{ds} decrease ↘.
- **Mobility degradation**: at high V_{gs} ↗, the carrier scatter more and mobility decrease. I_{ds} decrease ↘.
- **Channel length modulation effect**: at high V_{ds} ↗, depletion of S/D ↗ , effective L↘. I_{ds} increase ↗.
- **Subthreshold conduction**: $V_{gs} < V_t$, I_{ds} is exponentially dropoff instead of abruptly becoming zero.
- **Drain/Source leakage**: reverse diode junction leakage.
- **Non-zero gate current I_g**: carriers tunneling effect.
- **Body effect**: threshold voltage V_t is influence by V_{bs} (body-to-source voltage).

Velocity Saturation

- **Carrier velocity** : non-linearly proportional to lateral electrical field before velocity saturation

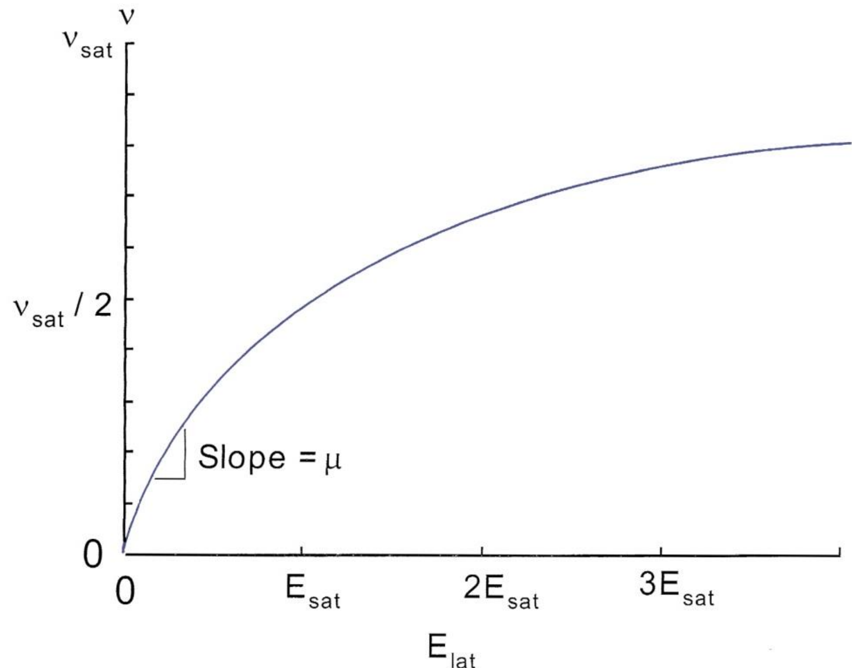
$$\upsilon = \mu E_{lat} / \left(1 + E_{lat} / E_{sat}\right)$$

υ : carrier velocity

μ : mobility

$E_{lat} = V_{ds} / L$: lateral electrical field

$E_{sat} = \upsilon_{sat} / \mu$



- I_{ds} will saturate due to velocity saturation, it depends on channel length **L** and applied $\mathbf{V_{ds}}$.

$$I_{ds} = \frac{Q_{channel}}{t_{channel}} = \frac{Q_{channel}}{L / \upsilon_{sat}} = C_{ox} W \left(V_{gs} - V_t\right) \upsilon_{sat}$$

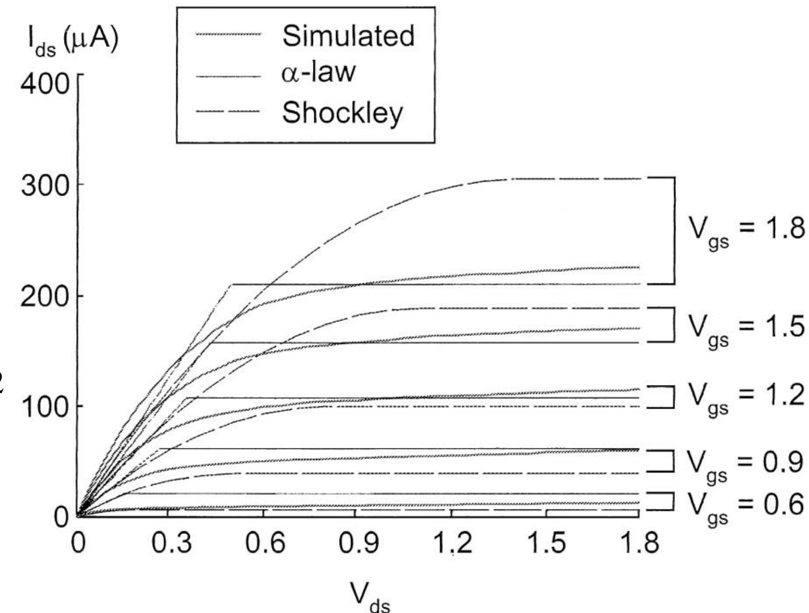
α-Power Law Model

- **α-Power Law Model**: piece-linear model to illustrate MOSs I-V characteristic with velocity saturation.

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t & \text{cutoff} \\ I_{dsat}\dfrac{V_{ds}}{V_{dsat}} & V_{ds} < V_{dsat} & \text{linear} \\ I_{dsat} & V_{ds} > V_{dsat} & \text{saturation} \end{cases}$$

$$I_{dsat} = P_c \frac{\beta}{2}\left(V_{gs} - V_t\right)^{\alpha}, \quad V_{dsat} = P_\upsilon \left(V_{gs} - V_t\right)^{\alpha/2}$$

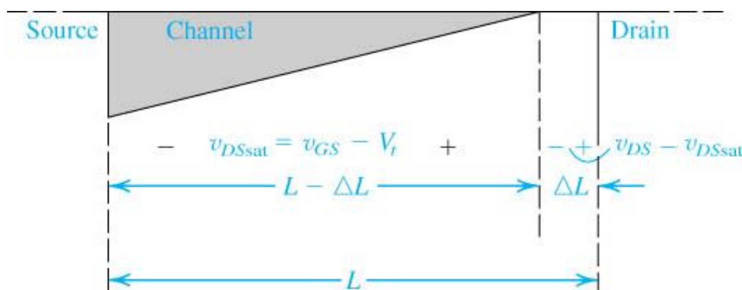
Empirical parameters: P_c, P_υ, α



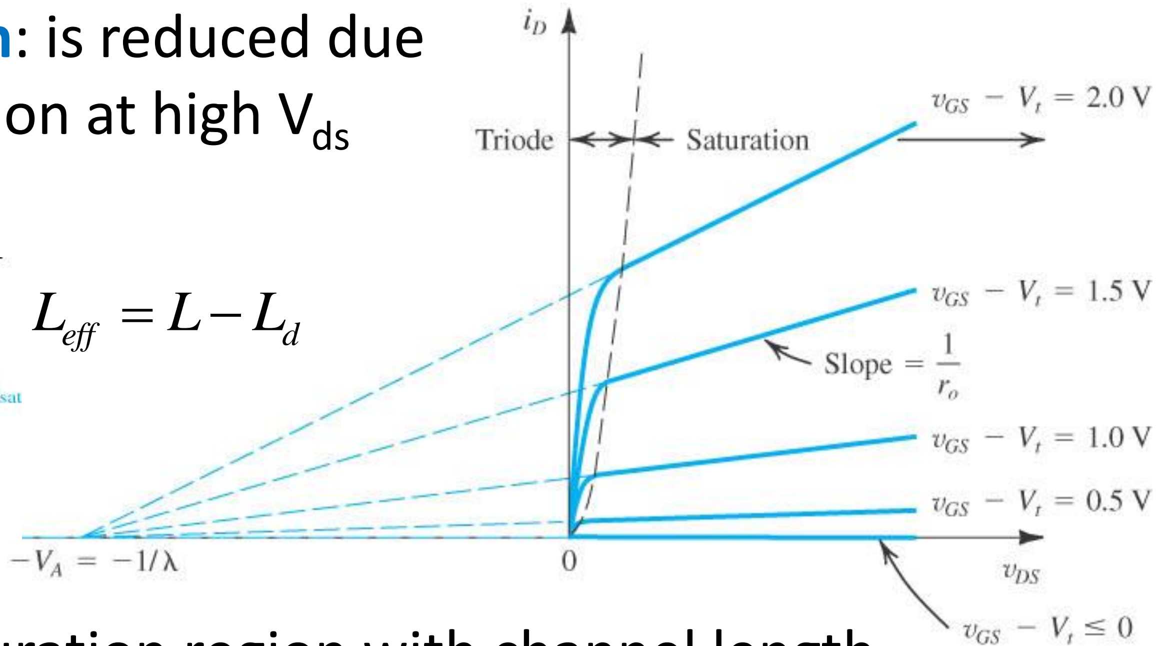
- **Because of $\mu_p < \mu_n$**, pMOS experience less velocity saturation than nMOS, therefore $\alpha_p > \alpha_n$
- **Mobility degradation** is modeled by a $\mu_{eff} < \mu$, and the it can be included in the parameter α

Channel Length Modulation

- **Effective channel length**: is reduced due the wider depletion region at high V_{ds}



$$L_{eff} = L - L_d$$



- The I-V equation at saturation region with channel length modulation effect. λ' is the empirical parameter

$$I_{ds} = \beta \frac{\left(V_{gs} - V_t\right)^2}{2}\left(1 + \lambda V_{ds}\right), \lambda = \lambda' / L$$

- With shorter **L**↘, **λ**↗, result in **output resistance**↘, then MOSFET's **intrinsic gain**↘

Body Effect

- The threshold voltage V_t is increased by positive V_{sb}.
- $V_{sb} < 0$, V_t ↘, OFF leakage ↗ ➔ Design tradeoff

$$V_t = V_{t0} + \gamma[\sqrt{2\varphi_s + V_{sb}} - \sqrt{2\varphi_s}],$$

$$\varphi_s = 2\upsilon_T \ln\frac{N_A}{n_i}$$

V_{t0} : *the threshold voltage for* $V_{sb} = 0$

φ_f : *fabrication-process parameter*

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

γ : *Body-effect parameter (fabrication-process parameter)*

N_A : *Doping concentration of p-type substrate*

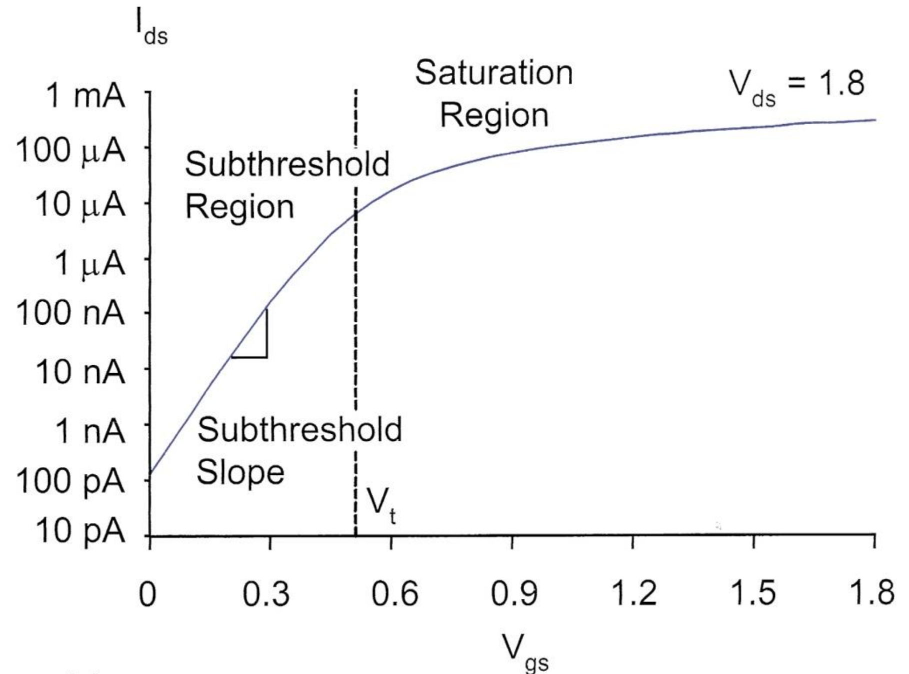
ε_s : *permittivity of silicon=11.7* ε_0

Subthreshold Conduction

- Leakage current at subthreshold region $V_{gs} < V_t$: weak inversion

$$I_{ds} = I_{ds0} e^{\frac{V_{gs}-V_t}{n\upsilon_T}} \left(1 - e^{\frac{-V_{ds}}{\upsilon_T}}\right)$$

$$I_{ds0} = \beta \upsilon_T^2 e^{1.8}$$

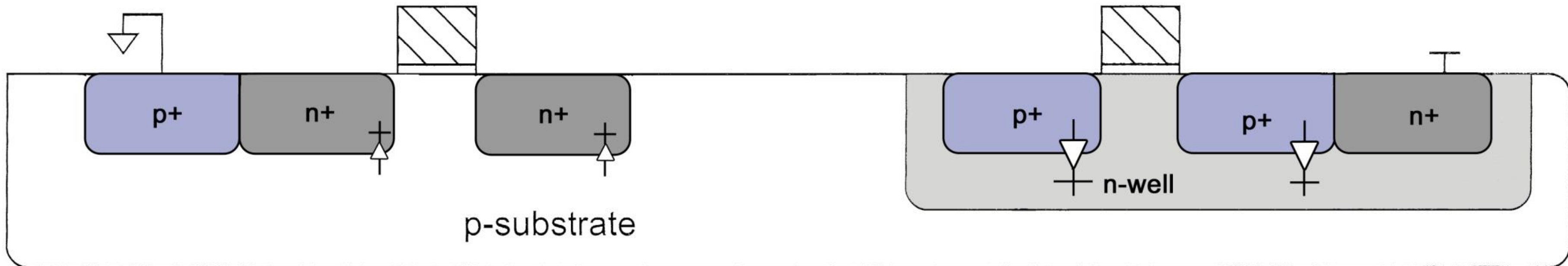


- Leakage current = 0, when V_{ds} = 0. It increase exponentially with V_{gs}.

- **Drain-Induced Barrier Lowering (DIBL)** : the V_t will be reduced by a positive V_{ds}. It will worse the leakage at subthreshold. It acts like channel length modulation effect at active mode.

$$V_t' = V_t - \eta V_{ds}$$

Junction Leakage

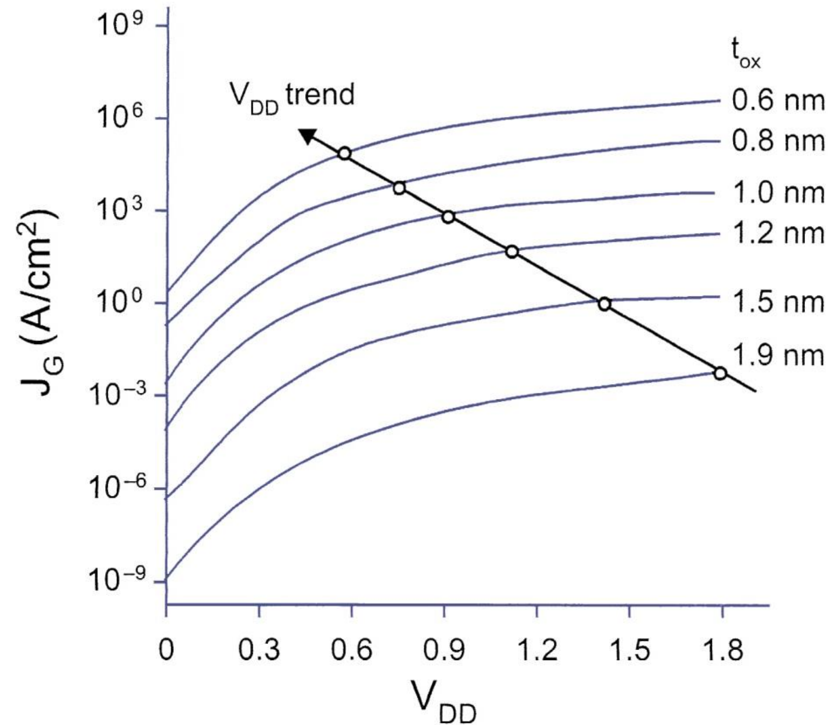


- S/D junction leakage from a reverse-biased diode

$$I_D = I_S\left(e^{\frac{V_D}{\upsilon_T}} - 1\right)$$

- Junction leakage used to be the storage time limitation. In modern transistor with shorter length, subthreshold leakage far exceed junction leakage.

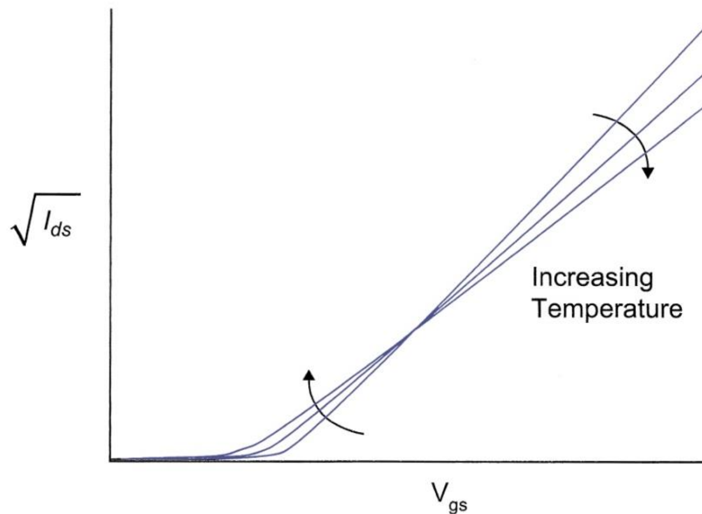
Tunneling

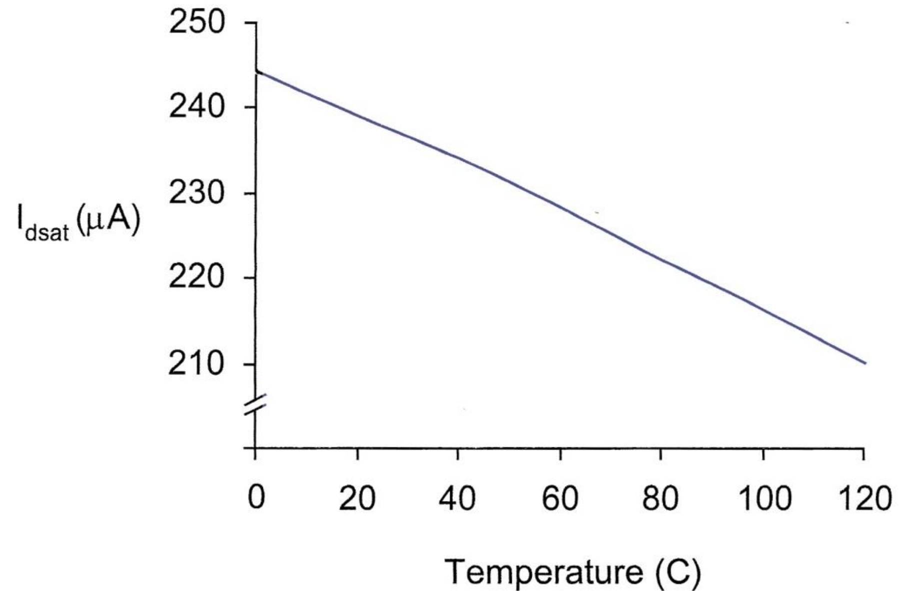


- Gate leakage: from carriers' tunneling through gate oxide. It is exponentially reverse proportional to gate oxide.
- Hi-K (dielectric constant) gate insulator is used to reduce it.

Temperature Dependence

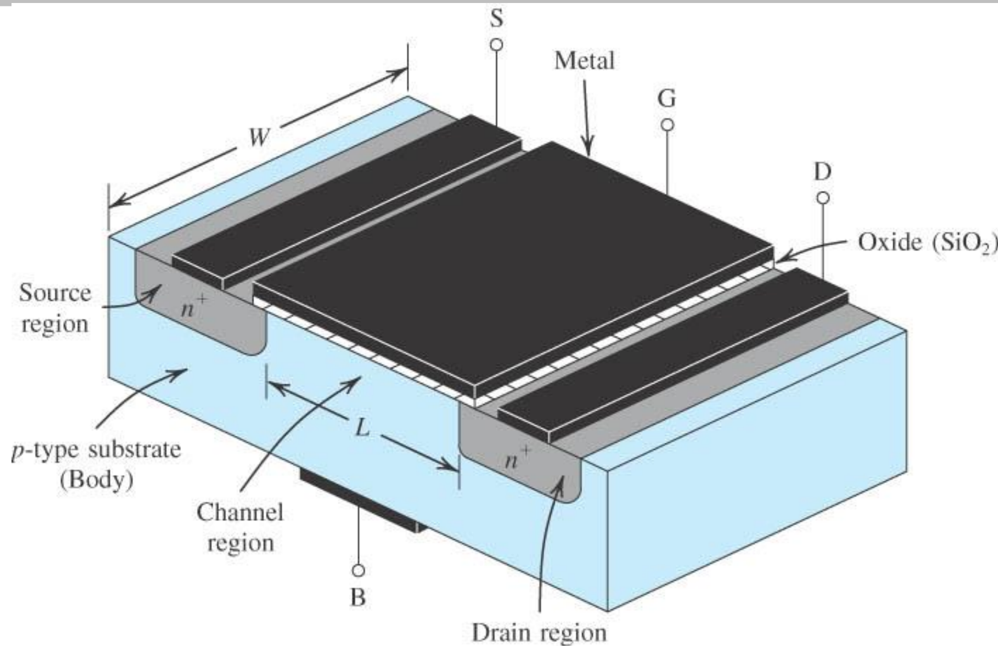
- T↗ μ↘: $\mu(T) = \mu(T_r)\left(\frac{T}{T_r}\right)^{-k_\mu}$
- T↗ V_t↘: $V_t(T) = V_t(T_r) - k_{vt}(T - T_r)$

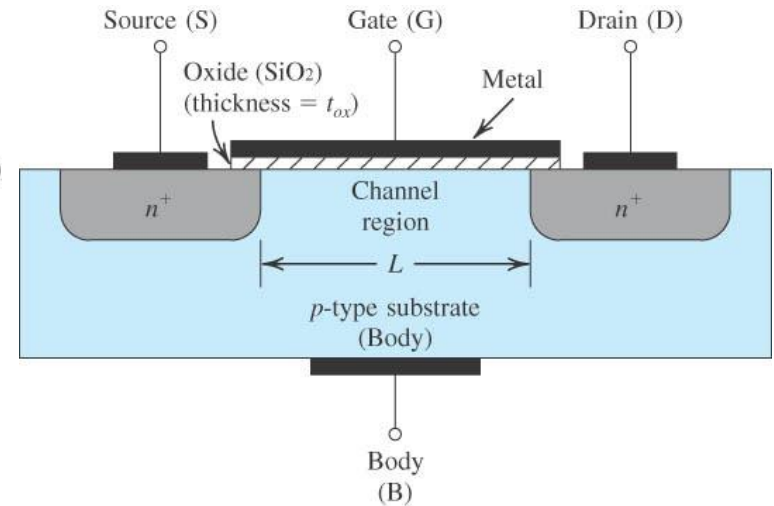




- T↗ I_{OFF} ↗ I_{ON}↘
- T↗ I_{dsat}↘
- T↘ The circuit performance is improved: subthreshold leakage ↘, saturation velocity ↗, mobility ↗, junction capacitance ↘. But breakdown voltage ↘.

Geometry Dependence





- Effective channel length and width

$$L_{eff} = L_{drawn} + X_L - 2L_D \qquad X_L, X_W : \text{Poly over-etch}$$

$$W_{eff} = W_{drawn} + X_W - 2W_D \qquad L_D, W_D : \text{Source-drian lateral diffusion}$$

- Use identical and same orientation MOSFET to get a good matching. ex. Current mirror

Outline

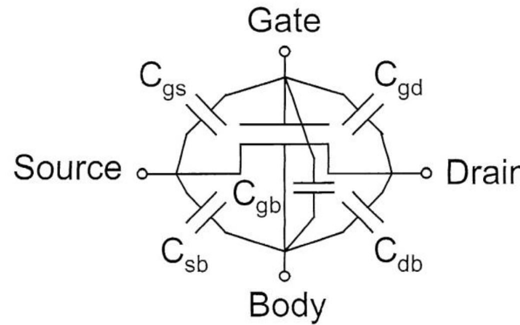
1. Introduction
2. Ideal I-V Characteristics
3. Nonideal I-V Effects
4. **C-V Characteristics**
5. DC Transfer Characteristics
6. Switch-level RC Delay Models

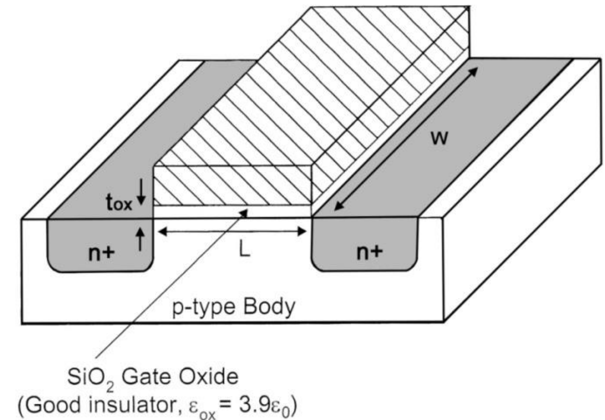
C-V Characteristics

- **Gate capacitance**: with advance technology, t_{ox}↘, L↘, $C_{permicron}$ keeps constant.

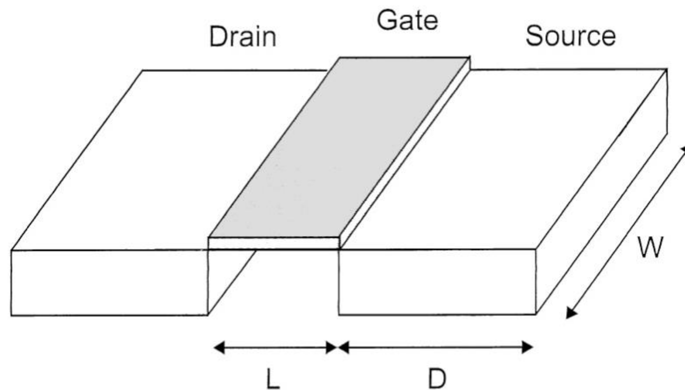
$$C_g = C_{ox}WL = C_{permicron}W$$

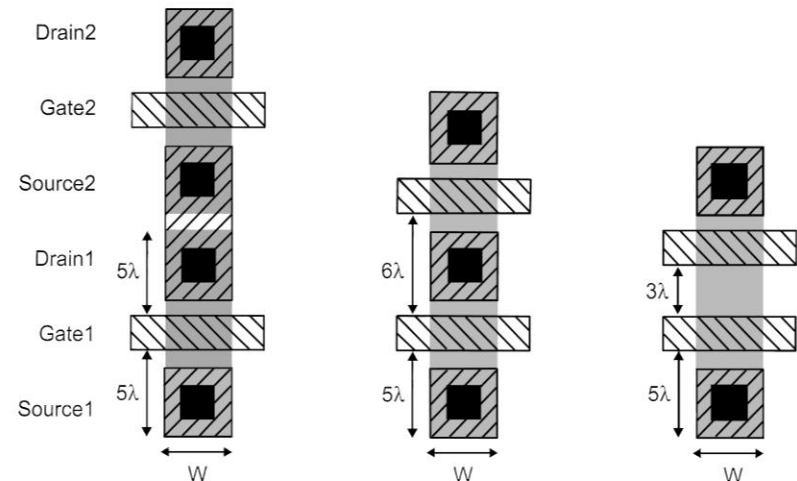
$$C_{permicron} = C_{ox}L = \frac{\varepsilon_{ox}}{t_{ox}}L$$





- **Parasitic capacitance**: C_{db} and C_{sb} are from reverse p-n junction and proportional to S/D area.





MOS Gate Capacitance Model

- **Gate capacitance**: it varies with channel behavior at different operation region

$$C_0 = C_{ox}WL$$

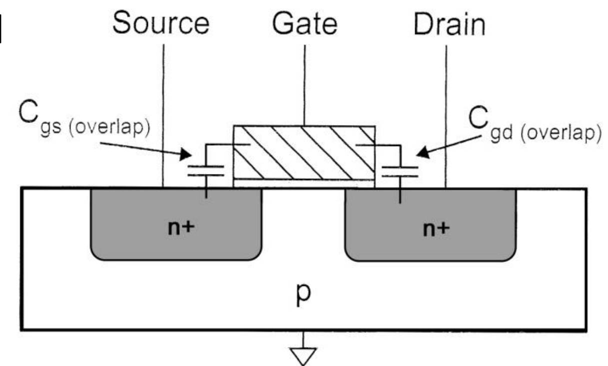
Table 2.1 Approximation of intrinsic MOS gate capacitance

Parameter	Cutoff	Linear	Saturation
C_{gb}	C_0	0	0
C_{gs}	0	$C_0/2$	$2/3\ C_0$
C_{gd}	0	$C_0/2$	0
$C_g = C_{gs} + C_{gd} + C_{gb}$	C_0	C_0	$2/3\ C_0$

- **S/D overlap capacitance**: $C_{gs(overlap)}$ & $C_{gd(overlap)}$ are from S/D lateral diffusion, don't confused with C_{gs} & C_{gd}

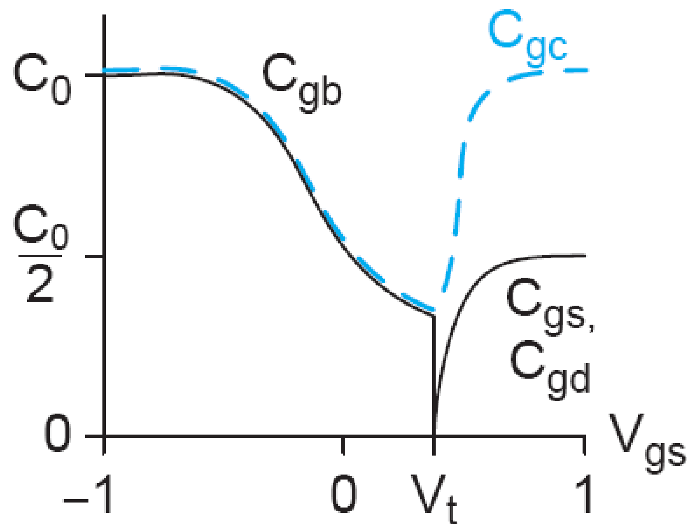
$$C_{gs(overlap)} = C_{gsol}W$$

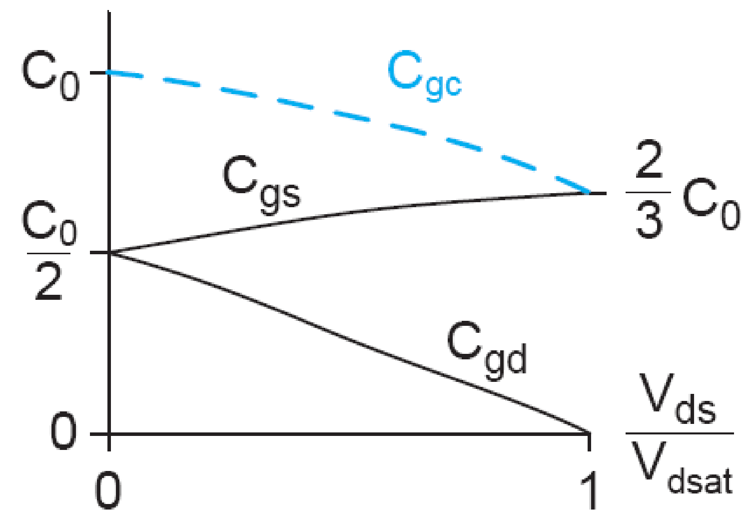
$$C_{gd(overlap)} = C_{gdol}W$$



C_{gc} vs. V_{gs} & V_{ds}

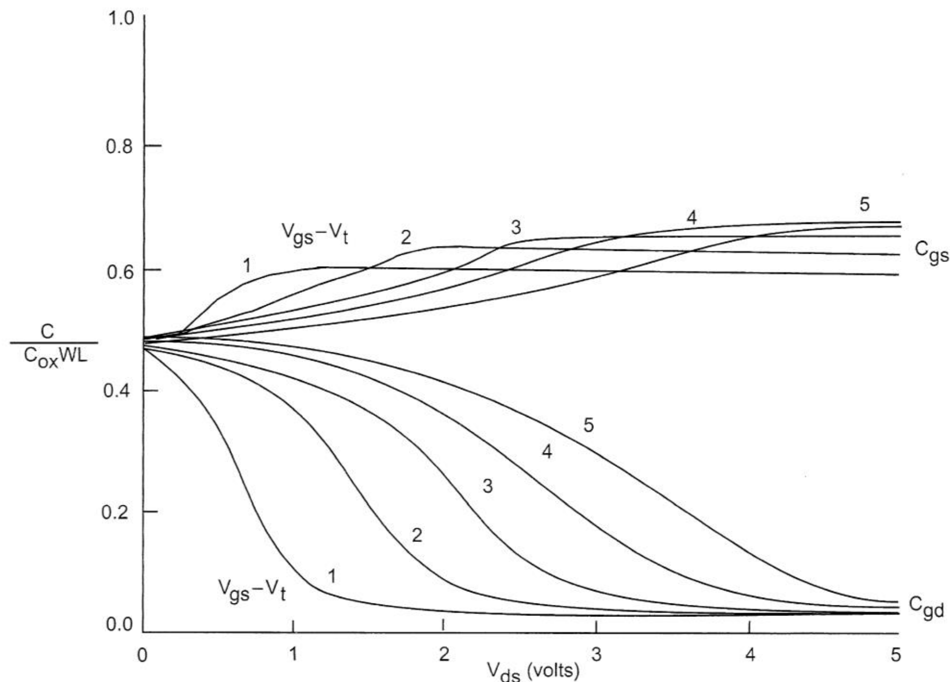
- C_{gc} vs. V_{gs}
- C_{gc} vs. V_{ds}

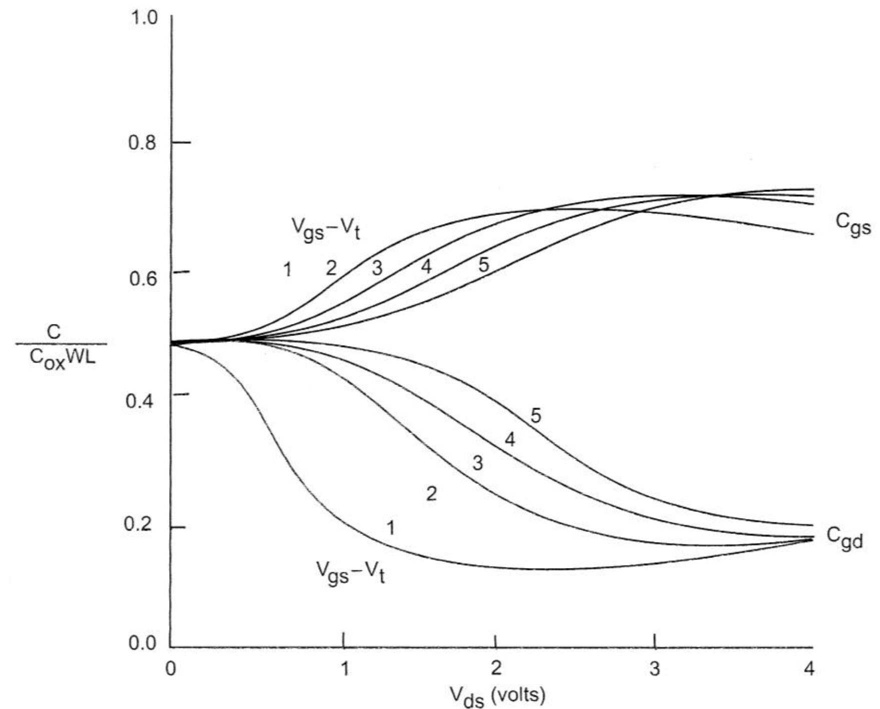




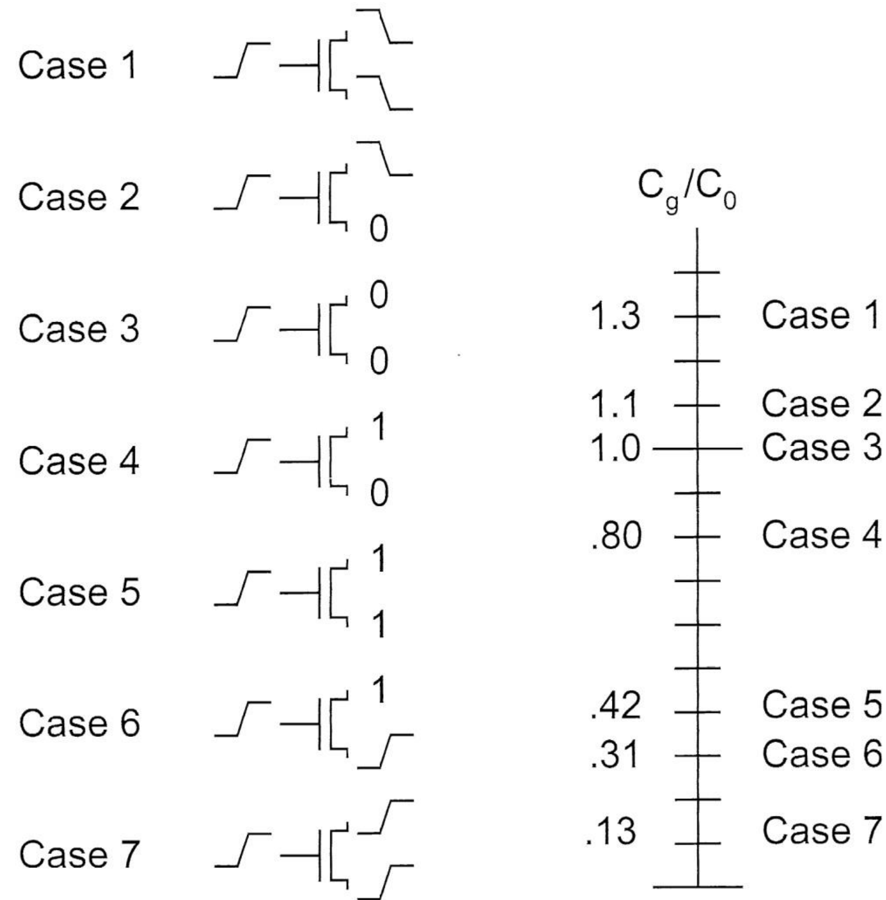
Gate Capacitance vs. V_{ds}

- **Long channel device:** C_{gd} will becomes near 0 at saturation
- **Short channel device:** more $C_{gd(overlap)}$ and $C_{gs(overlap)}$ factor





Data-Dependent Gate Capacitance



Outline

1. Introduction
2. Ideal I-V Characteristics
3. Nonideal I-V Effects
4. C-V Characteristics
5. **DC Transfer Characteristics**
6. Switch-level RC Delay Models

CMOS Inverter DC Characteristic I

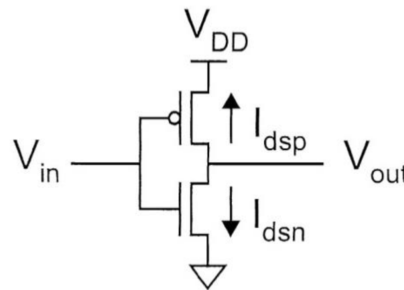
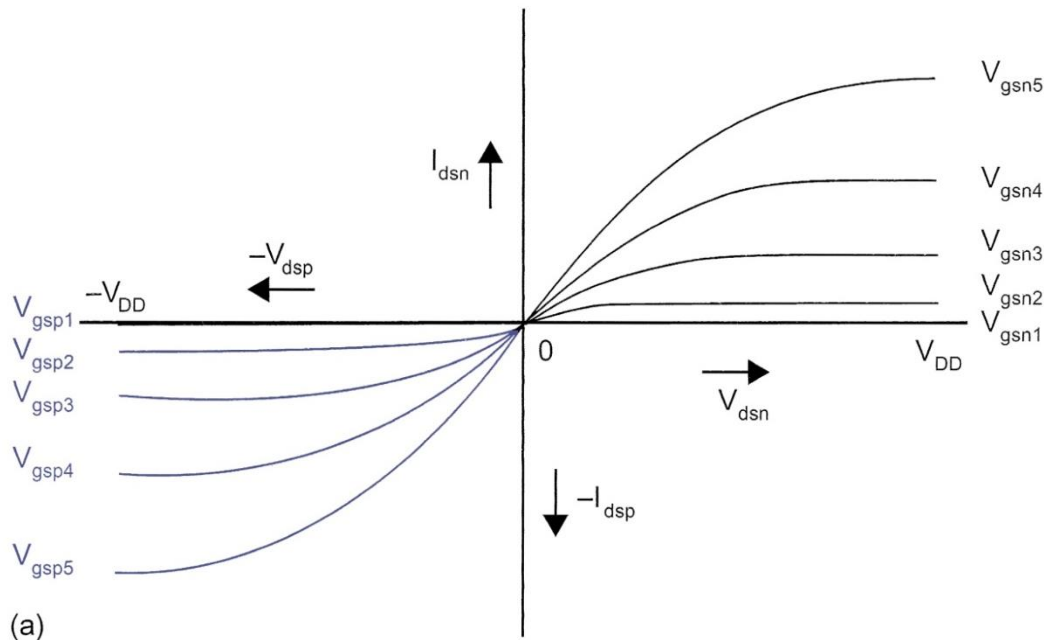
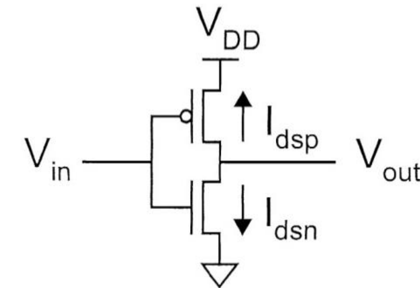


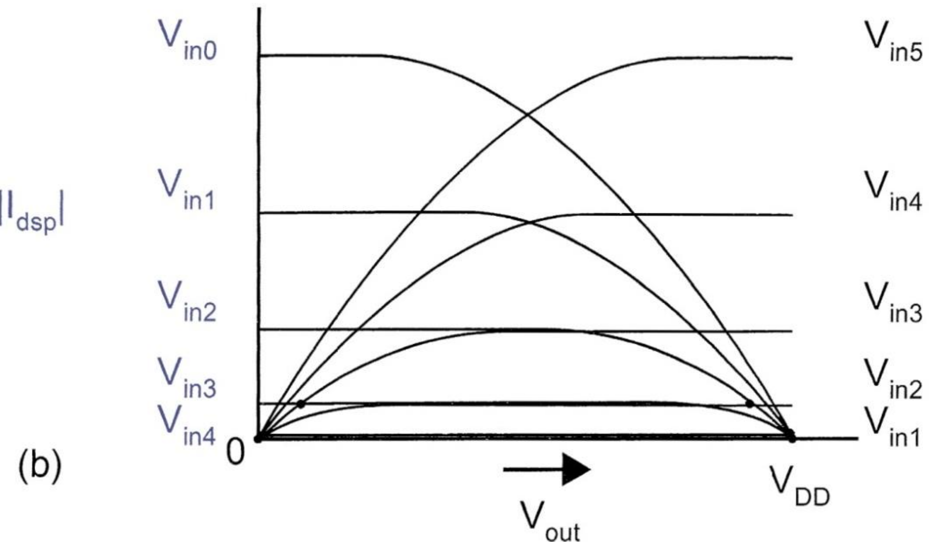
Table 2.2 Relationships between voltages for the three regions of operation of a CMOS inverter

	Cutoff	Linear	Saturated
nMOS	$V_{gsn} < V_{tn}$	$V_{gsn} > V_{tn}$	$V_{gsn} > V_{tn}$
	$V_{\text{in}} < V_{tn}$	$V_{\text{in}} > V_{tn}$	$V_{\text{in}} > V_{tn}$
		$V_{dsn} < V_{gsn} - V_{tn}$	$V_{dsn} > V_{gsn} - V_{tn}$
		$V_{\text{out}} < V_{\text{in}} - V_{tn}$	$V_{\text{out}} > V_{\text{in}} - V_{tn}$
pMOS	$V_{gsp} > V_{tp}$	$V_{gsp} < V_{tp}$	$V_{gsp} < V_{tp}$
	$V_{\text{in}} > V_{tp} + V_{DD}$	$V_{\text{in}} < V_{tp} + V_{DD}$	$V_{\text{in}} < V_{tp} + V_{DD}$
		$V_{dsp} > V_{gsp} - V_{tp}$	$V_{dsp} < V_{gsp} - V_{tp}$
		$V_{\text{out}} > V_{\text{in}} - V_{tp}$	$V_{\text{out}} < V_{\text{in}} - V_{tp}$

CMOS Inverter DC Characteristic II





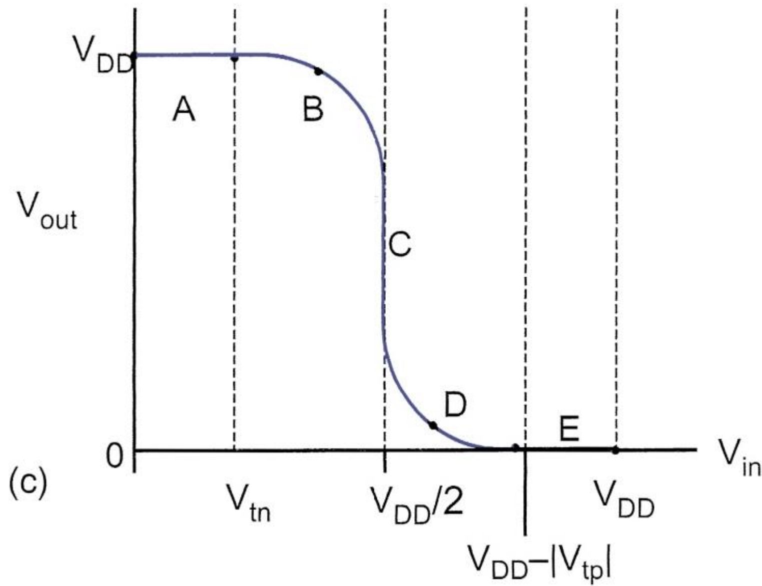


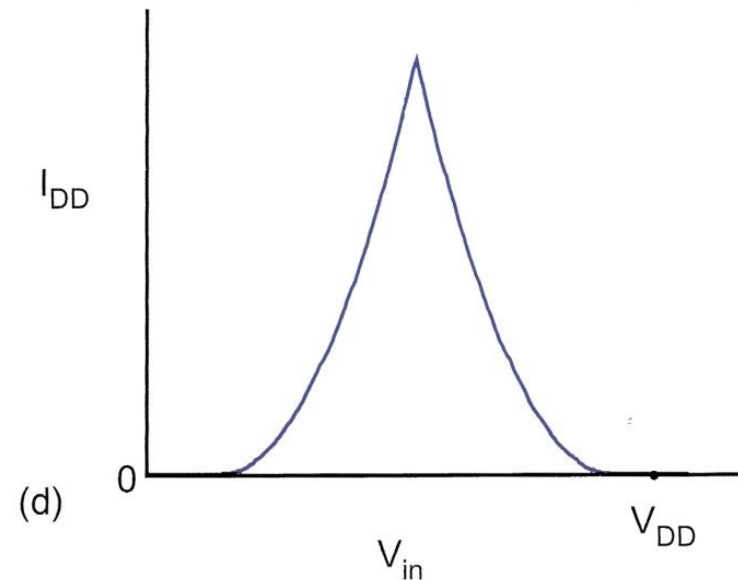
- I_{ds} vs. V_{ds} (nMOS & pMOS)
- For pMOS, I_d, V_{gs}, V_{ds}, V_{th} <0
- pMOS I-V as load-line of nMOS input device.

CMOS Inverter DC Characteristic III

- $V_{in} - V_{out}$ DC transfer curve
- Rail-to-rail operation

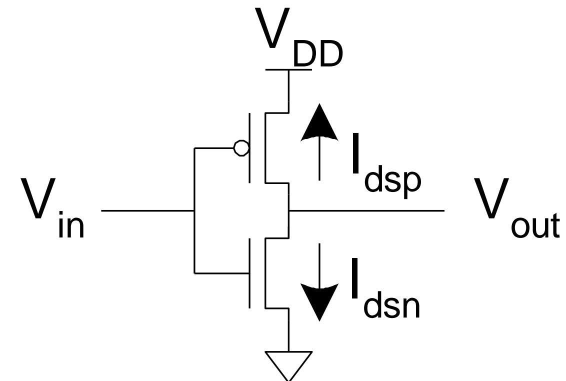
- $V_{in} - I_{DD}$ DC transfer curve
- Dynamic power dissipation





DC Response

- DC Response: V_{out} vs. V_{in} for a gate
- Ex: Inverter
  - When $V_{in} = 0$ -> $V_{out} = V_{DD}$
  - When $V_{in} = V_{DD}$ -> $V_{out} = 0$
  - In between, V_{out} depends on transistor size and current
  - By KCL, must settle such that $I_{dsn} = |I_{dsp}|$
  - We could solve equations
  - But graphical solution gives more insight

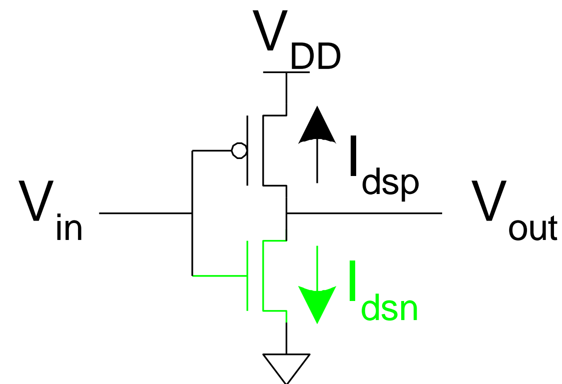


Transistor Operation

- Current depends on region of transistor behavior
- For what V_{in} and V_{out} are nMOS and pMOS in
  - Cutoff?
  - Linear?
  - Saturation?

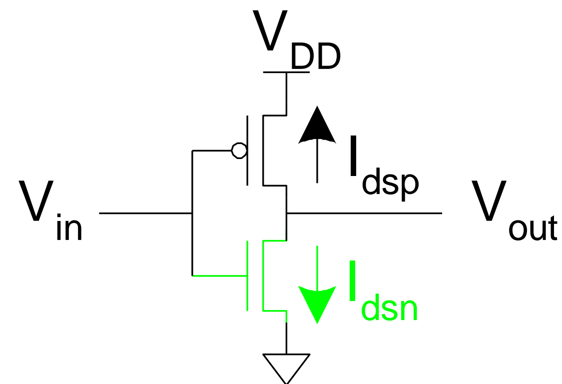
nMOS Operation

Cutoff	Linear	Saturated
$V_{gsn} <$	$V_{gsn} >$ $V_{dsn} <$	$V_{gsn} >$ $V_{dsn} >$



nMOS Operation

Cutoff	Linear	Saturated
$V_{gsn} < V_{tn}$	$V_{gsn} > V_{tn}$ $V_{dsn} < V_{gsn} - V_{tn}$	$V_{gsn} > V_{tn}$ $V_{dsn} > V_{gsn} - V_{tn}$

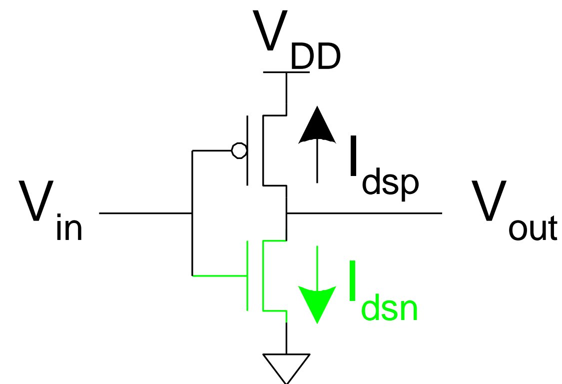


nMOS Operation

Cutoff	Linear	Saturated
$V_{gsn} < V_{tn}$	$V_{gsn} > V_{tn}$ $V_{dsn} < V_{gsn} - V_{tn}$	$V_{gsn} > V_{tn}$ $V_{dsn} > V_{gsn} - V_{tn}$

$V_{gsn} = V_{in}$

$V_{dsn} = V_{out}$

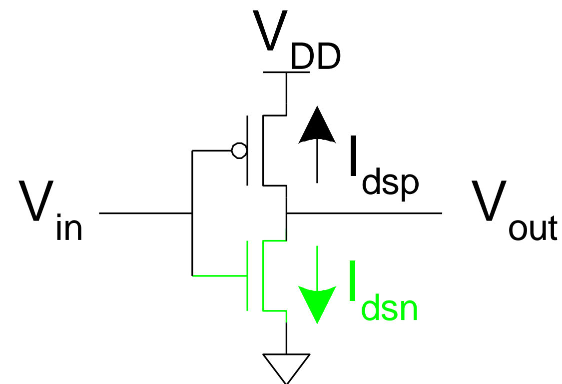


nMOS Operation

Cutoff	Linear	Saturated
$V_{gsn} < V_{tn}$ $V_{in} < V_{tn}$	$V_{gsn} > V_{tn}$ $V_{in} > V_{tn}$ $V_{dsn} < V_{gsn} - V_{tn}$ $V_{out} < V_{in} - V_{tn}$	$V_{gsn} > V_{tn}$ $V_{in} > V_{tn}$ $V_{dsn} > V_{gsn} - V_{tn}$ $V_{out} > V_{in} - V_{tn}$

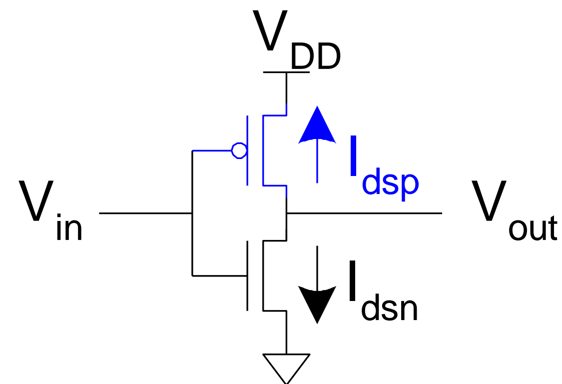
$V_{gsn} = V_{in}$

$V_{dsn} = V_{out}$



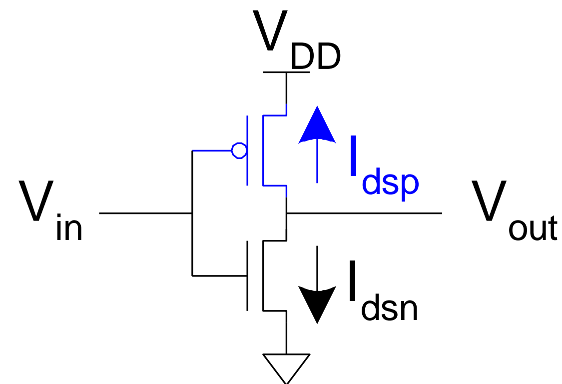
pMOS Operation

Cutoff	Linear	Saturated
$V_{gsp} >$	$V_{gsp} <$ $V_{dsp} >$	$V_{gsp} <$ $V_{dsp} <$



pMOS Operation

Cutoff	**Linear**	**Saturated**
$V_{gsp} > V_{tp}$	$V_{gsp} < V_{tp}$ $V_{dsp} > V_{gsp} - V_{tp}$	$V_{gsp} < V_{tp}$ $V_{dsp} < V_{gsp} - V_{tp}$



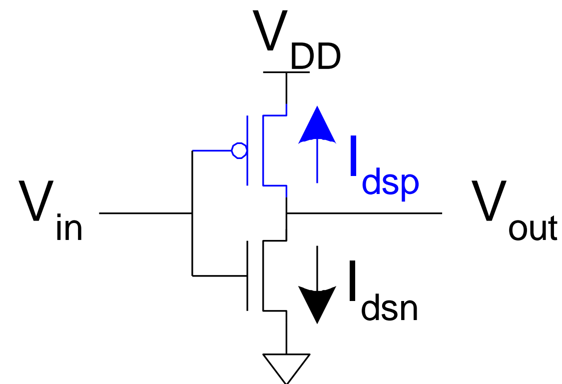
pMOS Operation

Cutoff	Linear	Saturated
$V_{gsp} > V_{tp}$	$V_{gsp} < V_{tp}$ $V_{dsp} > V_{gsp} - V_{tp}$	$V_{gsp} < V_{tp}$ $V_{dsp} < V_{gsp} - V_{tp}$

$V_{gsp} = V_{in} - V_{DD}$

$V_{dsp} = V_{out} - V_{DD}$

$V_{tp} < 0$



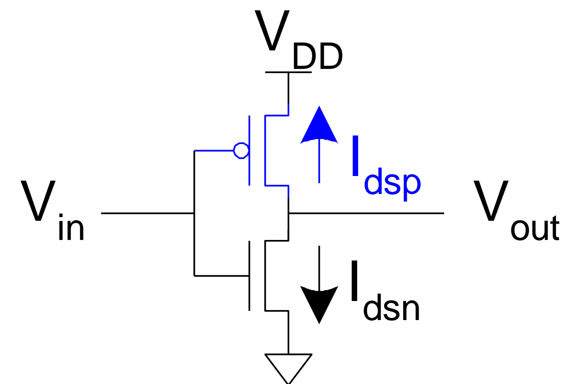
pMOS Operation

Cutoff	Linear	Saturated
$V_{gsp} > V_{tp}$ $V_{in} > V_{DD} + V_{tp}$	$V_{gsp} < V_{tp}$ $V_{in} < V_{DD} + V_{tp}$ $V_{dsp} > V_{gsp} - V_{tp}$ $V_{out} > V_{in} - V_{tp}$	$V_{gsp} < V_{tp}$ $V_{in} < V_{DD} + V_{tp}$ $V_{dsp} < V_{gsp} - V_{tp}$ $V_{out} < V_{in} - V_{tp}$

$V_{gsp} = V_{in} - V_{DD}$

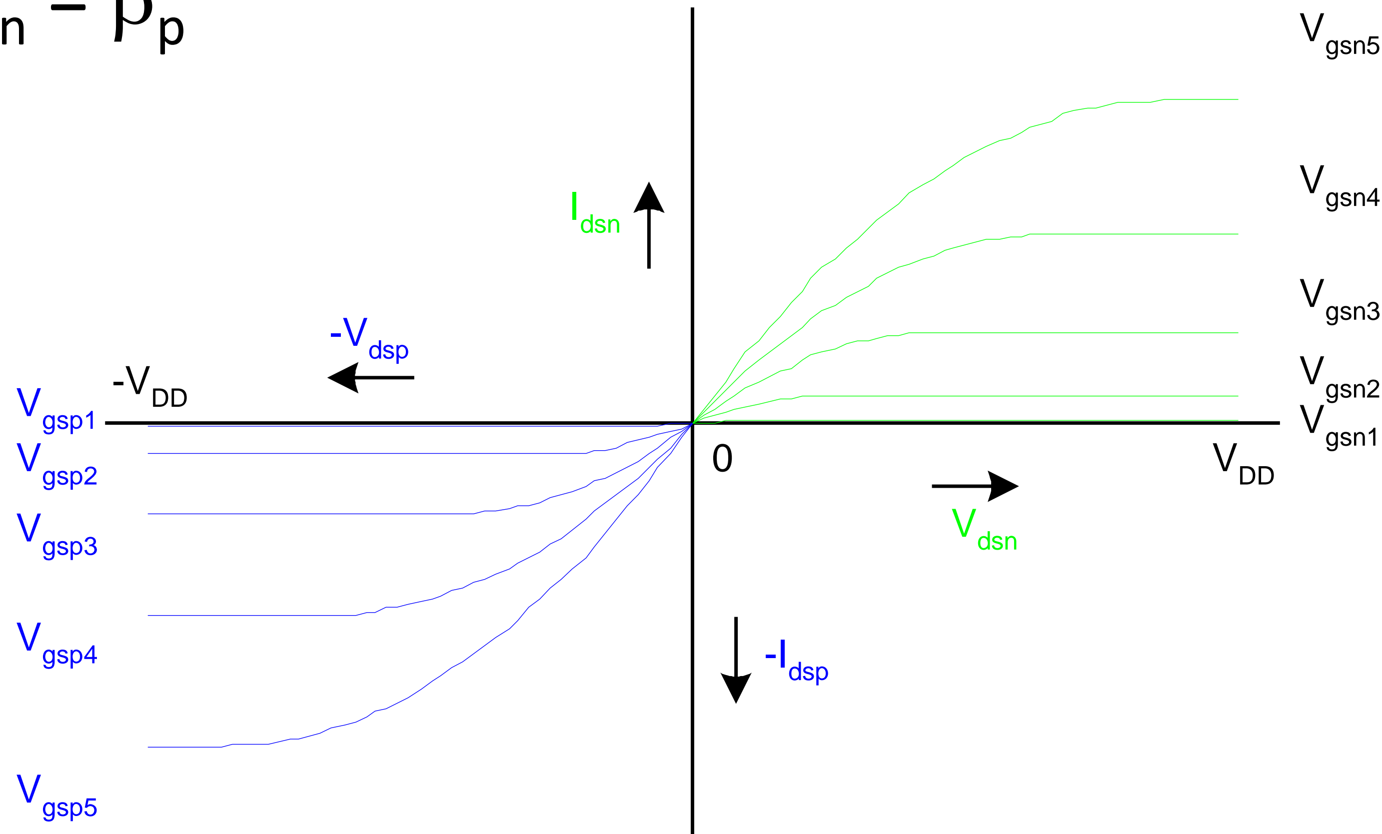
$V_{dsp} = V_{out} - V_{DD}$

$V_{tp} < 0$

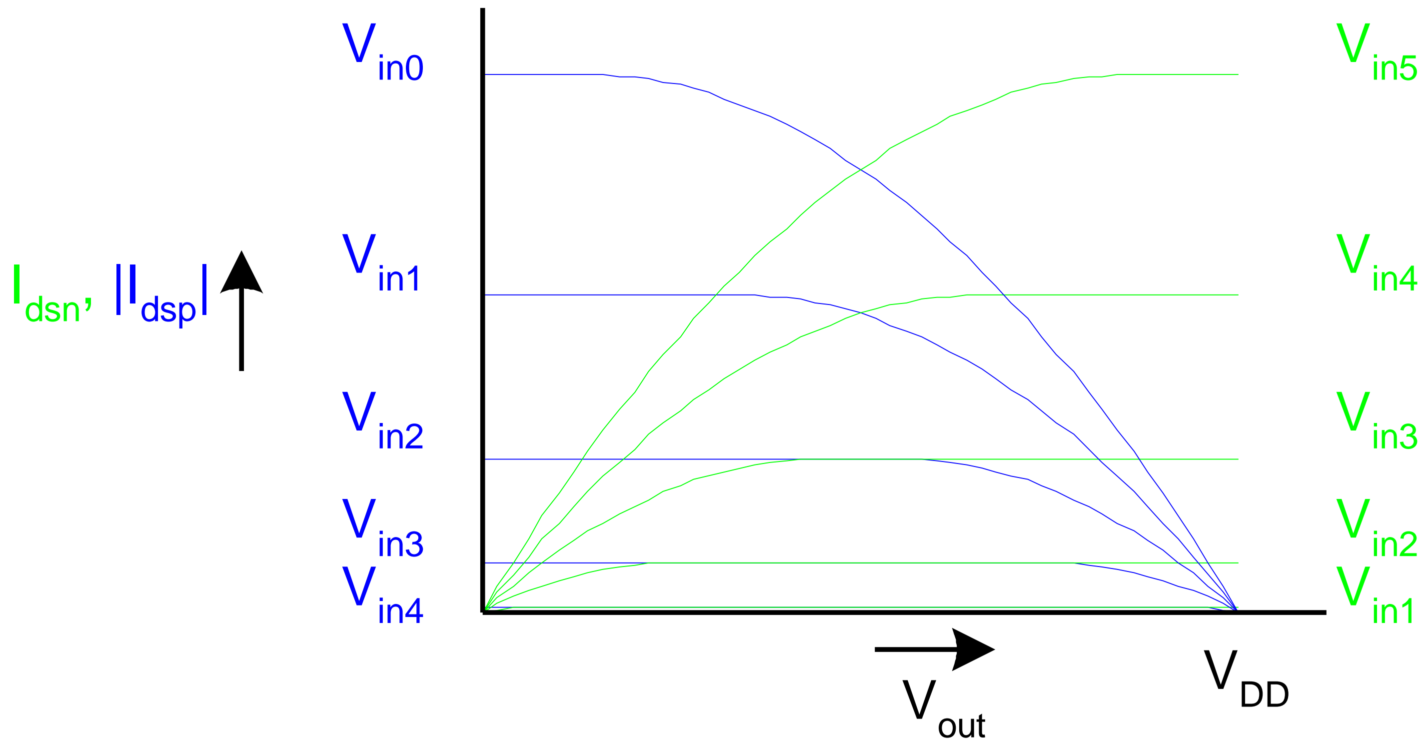


I-V Characteristics

- Make pMOS is wider than nMOS such that $\beta_n = \beta_p$

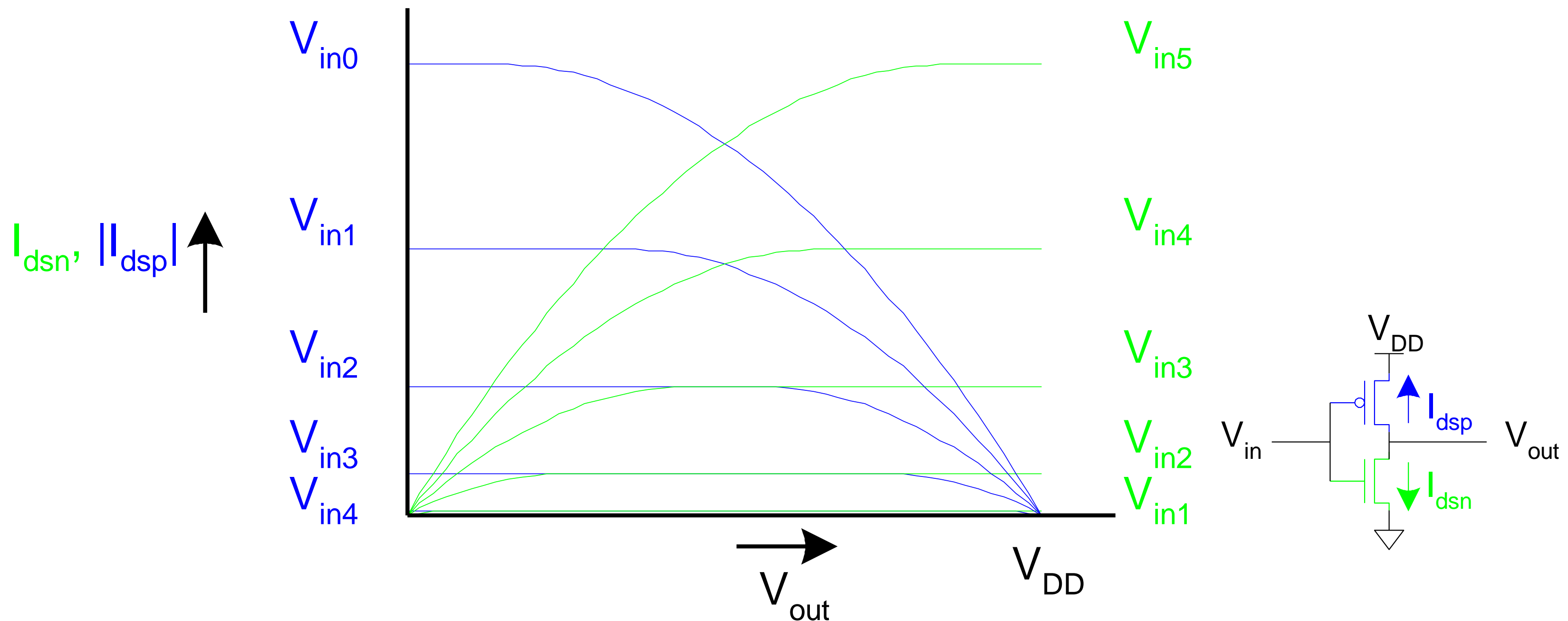


Current vs. V_{out}, V_{in}



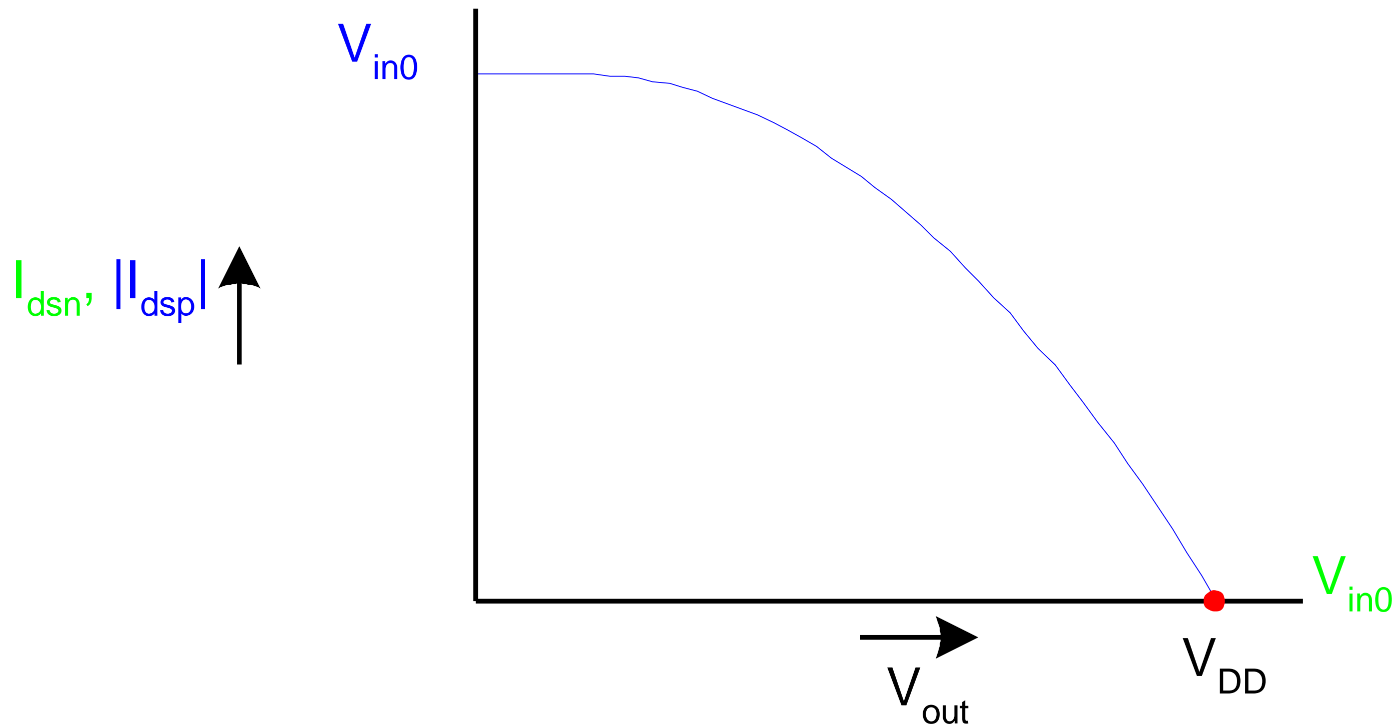
Load Line Analysis

- For a given V_{in}:
  - Plot I_{dsn}, I_{dsp} vs. V_{out}
  - V_{out} must be where |currents| are equal in



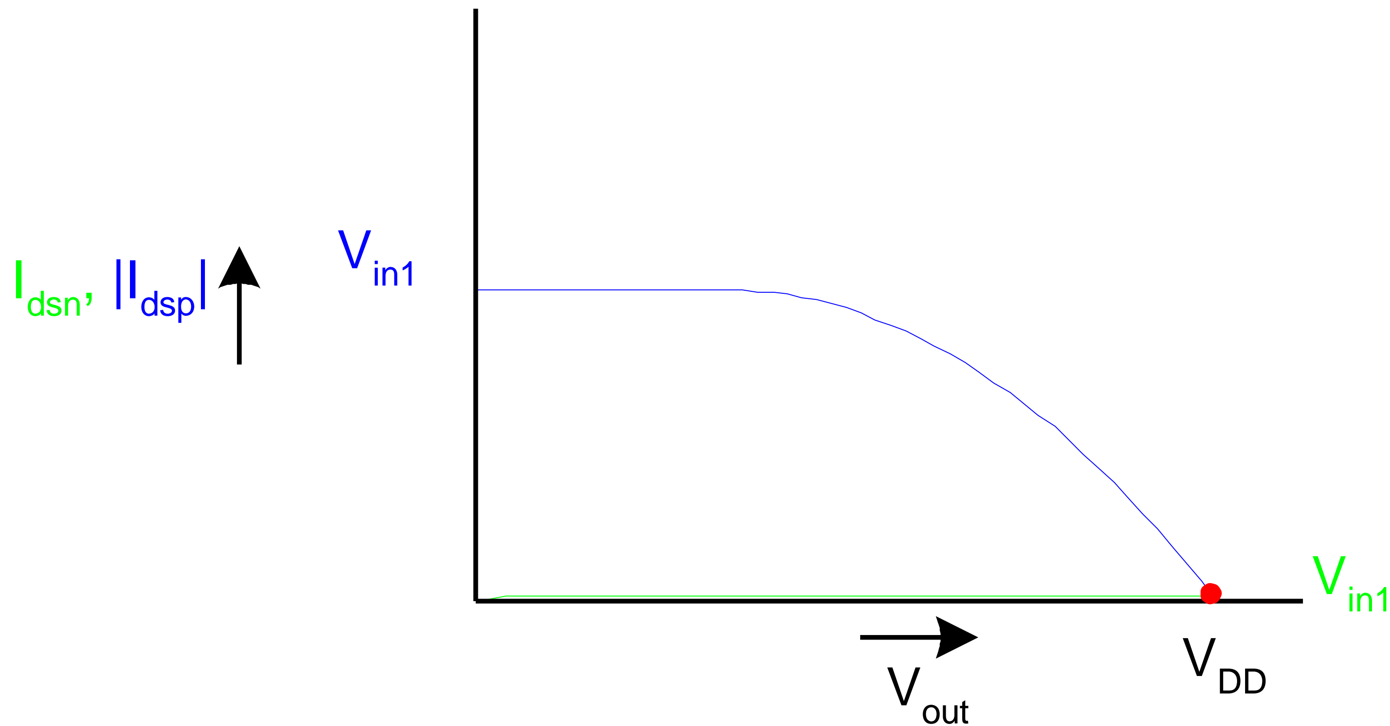
Load Line Analysis

- $V_{in} = 0$



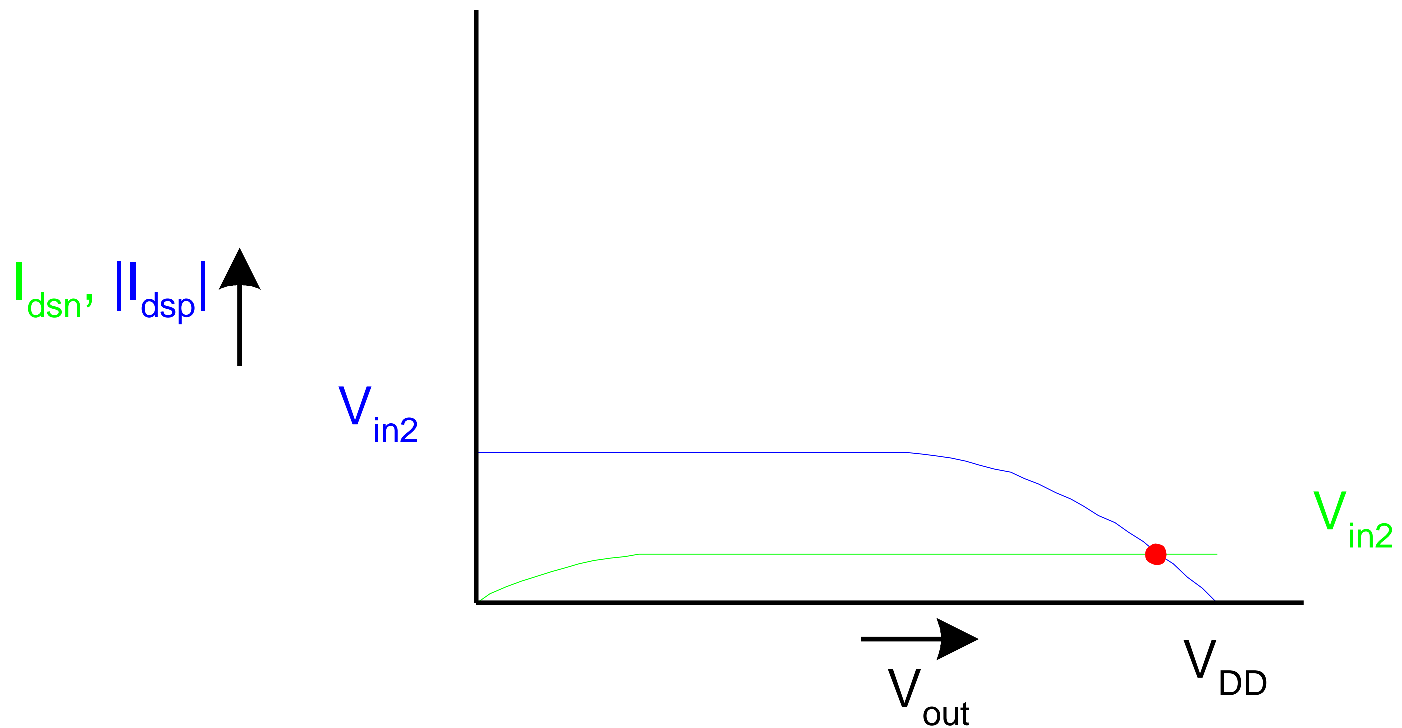
Load Line Analysis

- $V_{in} = 0.2V_{DD}$



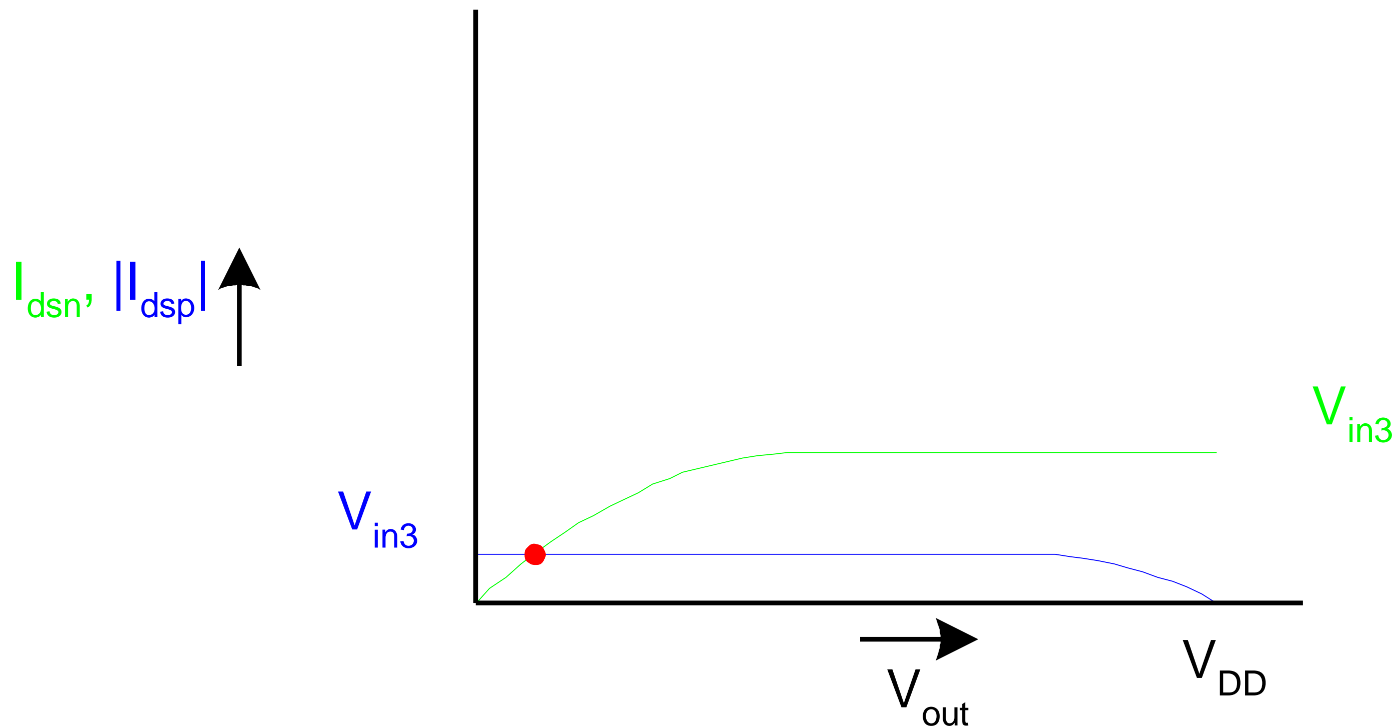
Load Line Analysis

- $V_{in} = 0.4V_{DD}$



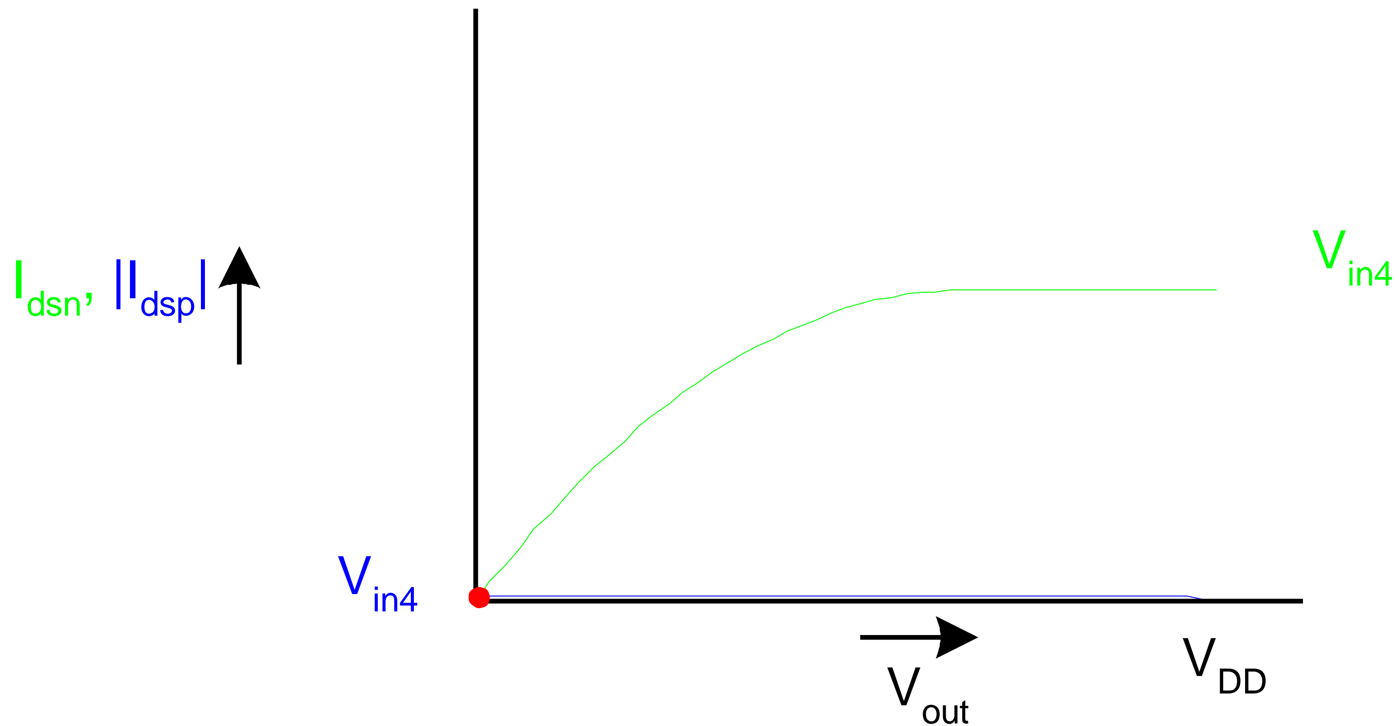
Load Line Analysis

- $V_{in} = 0.6V_{DD}$



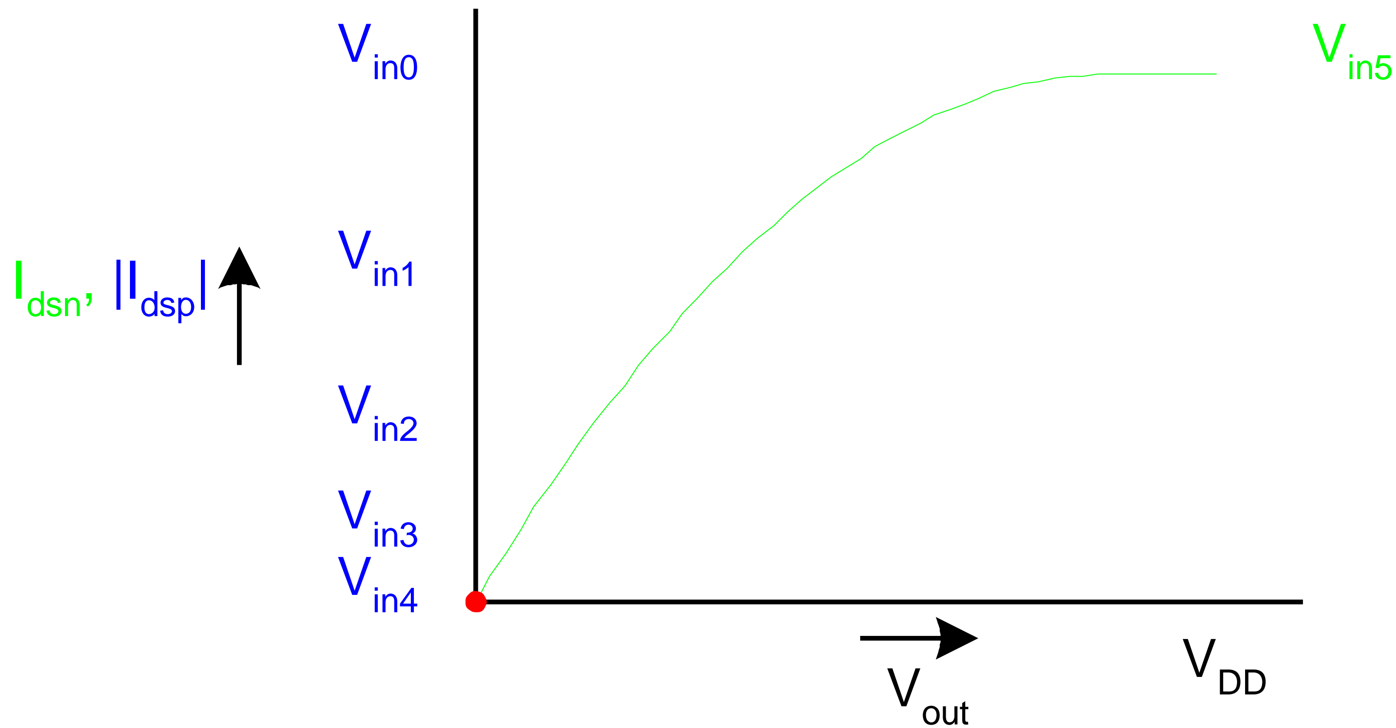
Load Line Analysis

- $V_{in} = 0.8V_{DD}$

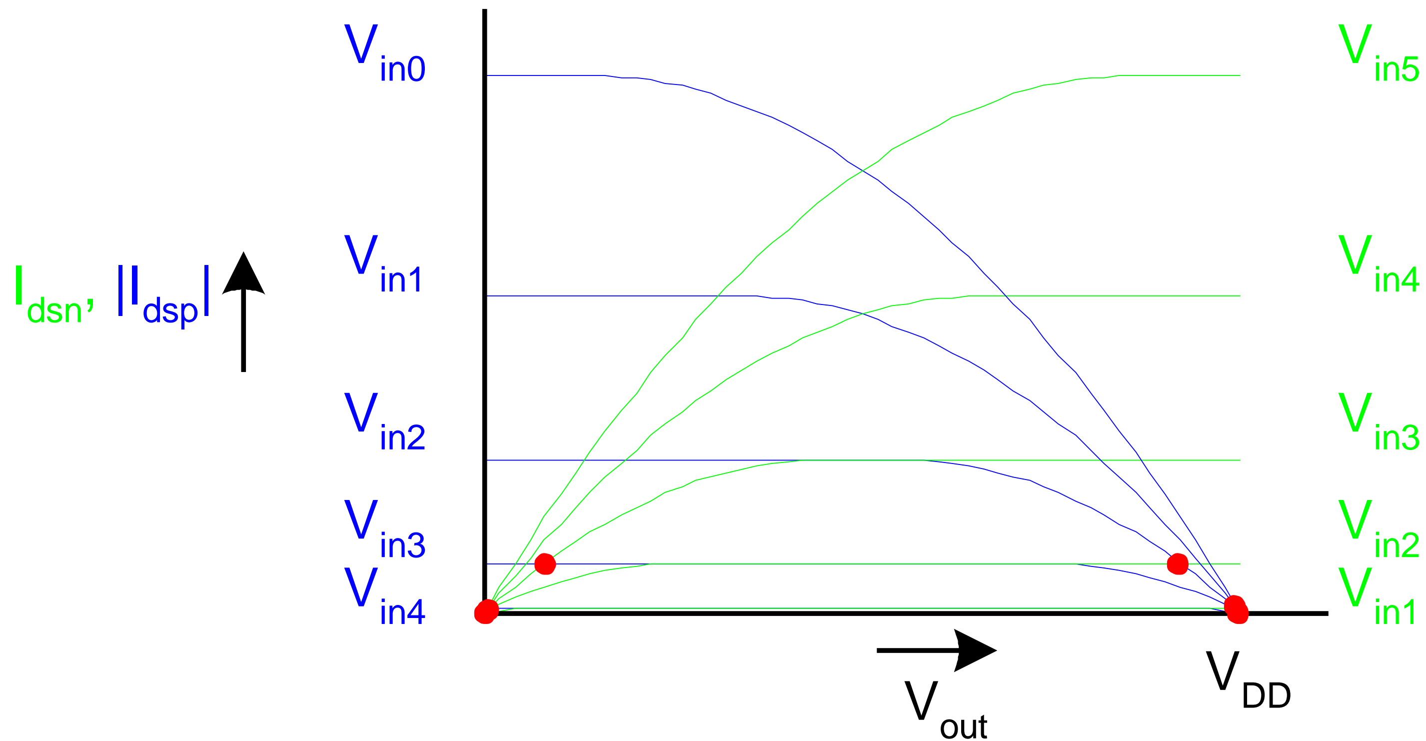


Load Line Analysis

- $V_{in} = V_{DD}$

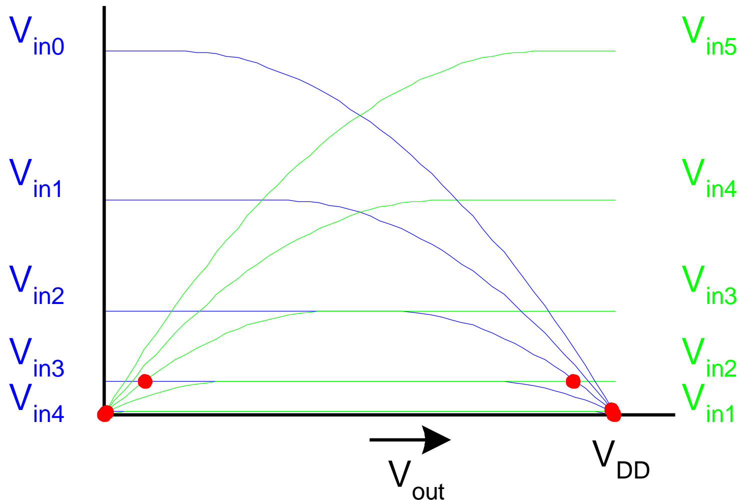


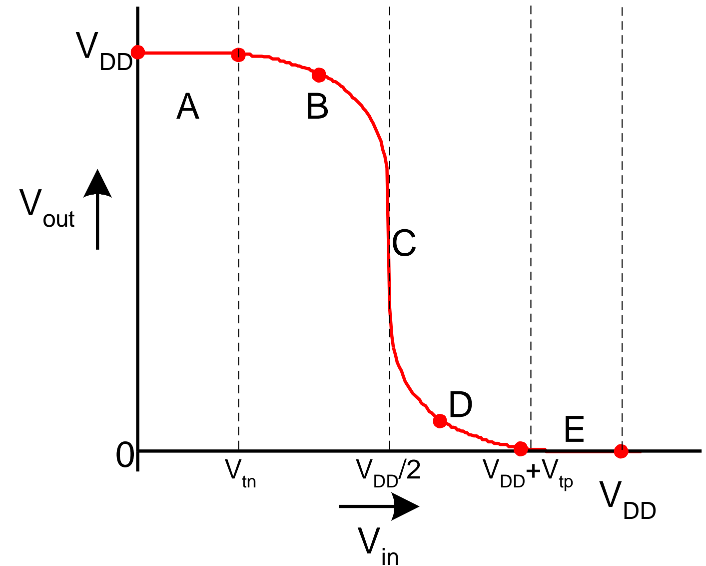
Load Line Summary



DC Transfer Curve

- Transcribe points onto V_{in} vs. V_{out} plot

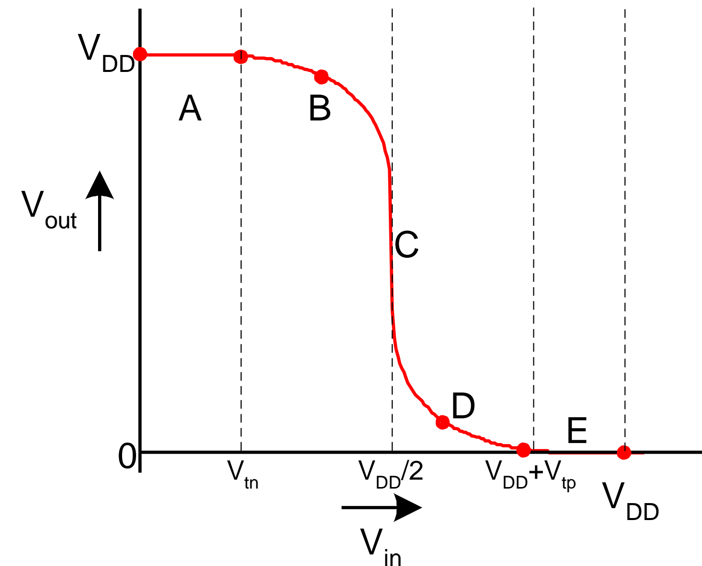




Operating Regions

- Revisit transistor operating regions

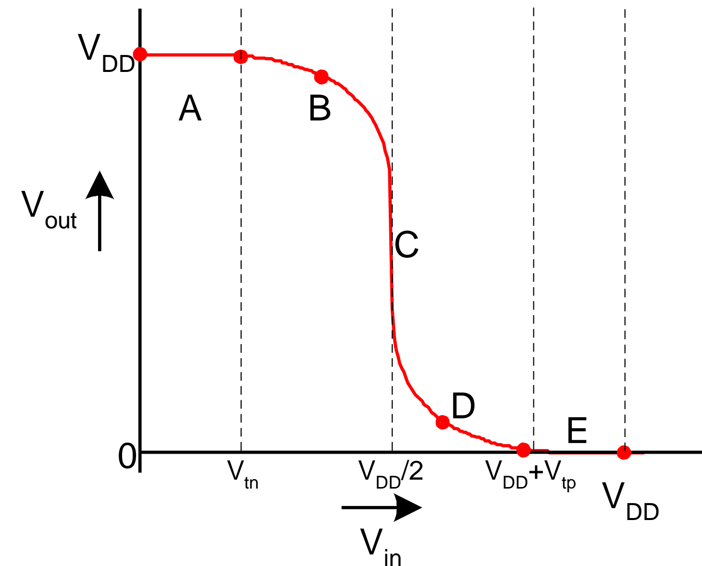
Region	nMOS	pMOS
A		
B		
C		
D		
E		



Operating Regions

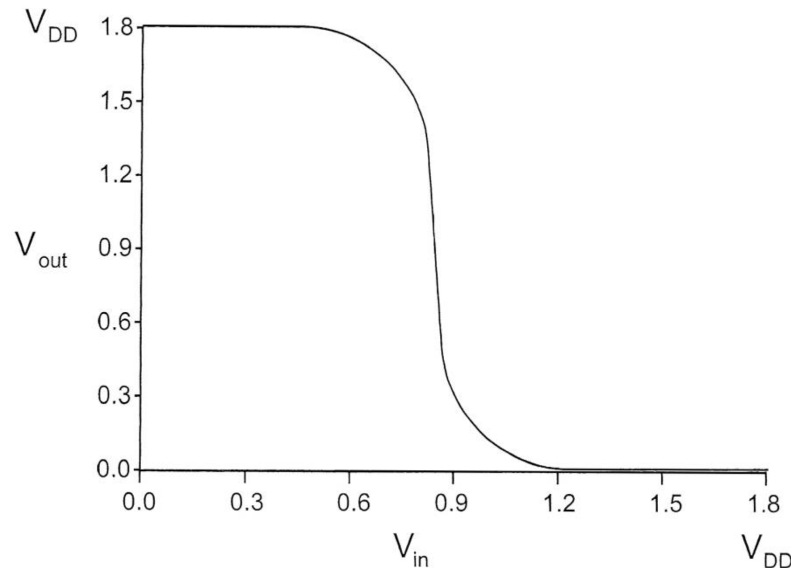
- Revisit transistor operating regions

Region	nMOS	pMOS
A	Cutoff	Linear
B	Saturation	Linear
C	Saturation	Saturation
D	Linear	Saturation
E	Linear	Cutoff



CMOS Inverter Operation

Table 2.3	Summary of CMOS inverter operation			
Region	Condition	p-device	n-device	Output
A	$0 \leq V_{in} < V_{tn}$	linear	cutoff	$V_{out} = V_{DD}$
B	$V_{tn} \leq V_{in} < V_{DD}/2$	linear	saturated	$V_{out} > V_{DD}/2$
C	$V_{in} = V_{DD}/2$	saturated	saturated	V_{out} drops sharply
D	$V_{DD}/2 < V_{in} \leq V_{DD} - \lvert V_{tp} \rvert$	saturated	linear	$V_{out} < V_{DD}/2$
E	$V_{in} > V_{DD} - \lvert V_{tp} \rvert$	cutoff	linear	$V_{out} = 0$



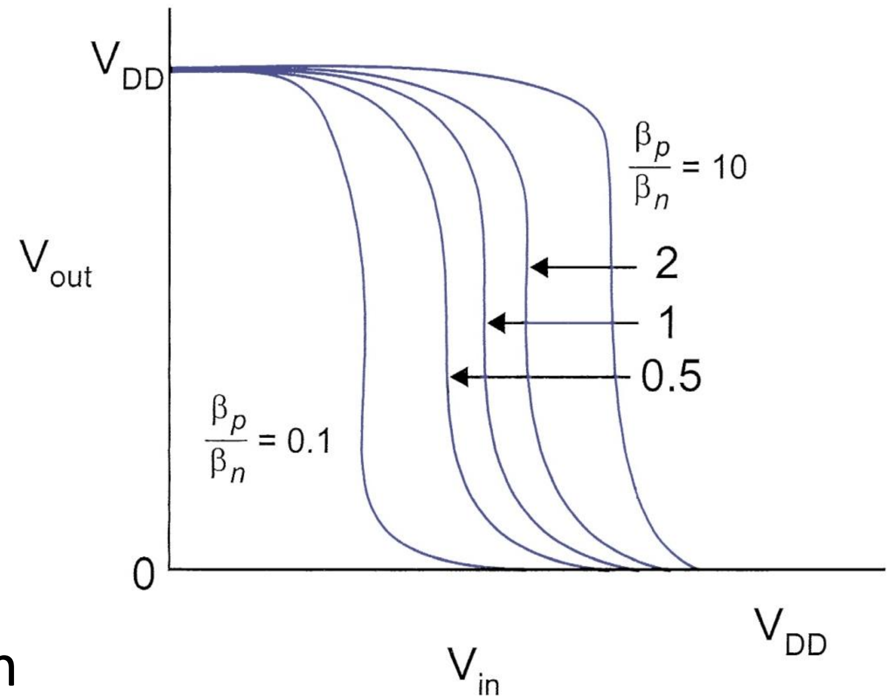
Beta Ratio Effect

- β parameters and ratio

$$\beta_p = \mu_p C_{ox} (W/L)_p$$

$$\beta_n = \mu_n C_{ox} (W/L)_n$$

$$\beta \text{ ratio} = \beta_p / \beta_n$$



- β ratio = 1: largest noise margin
  - ∵ $\mu_n > \mu_p$, choose $(W/L)_p > (W/L)_n$ to make **β ratio = 1**
- β ratio > 1: HI-skewed inverter, switching threshold > $0.5V_{DD}$
- β ratio < 1: LO-skewed inverter, switching threshold < $0.5V_{DD}$

Noise Margin I

- Noise margin definition
  - The allowable noise voltage on the input that the output wont be corrupted.

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

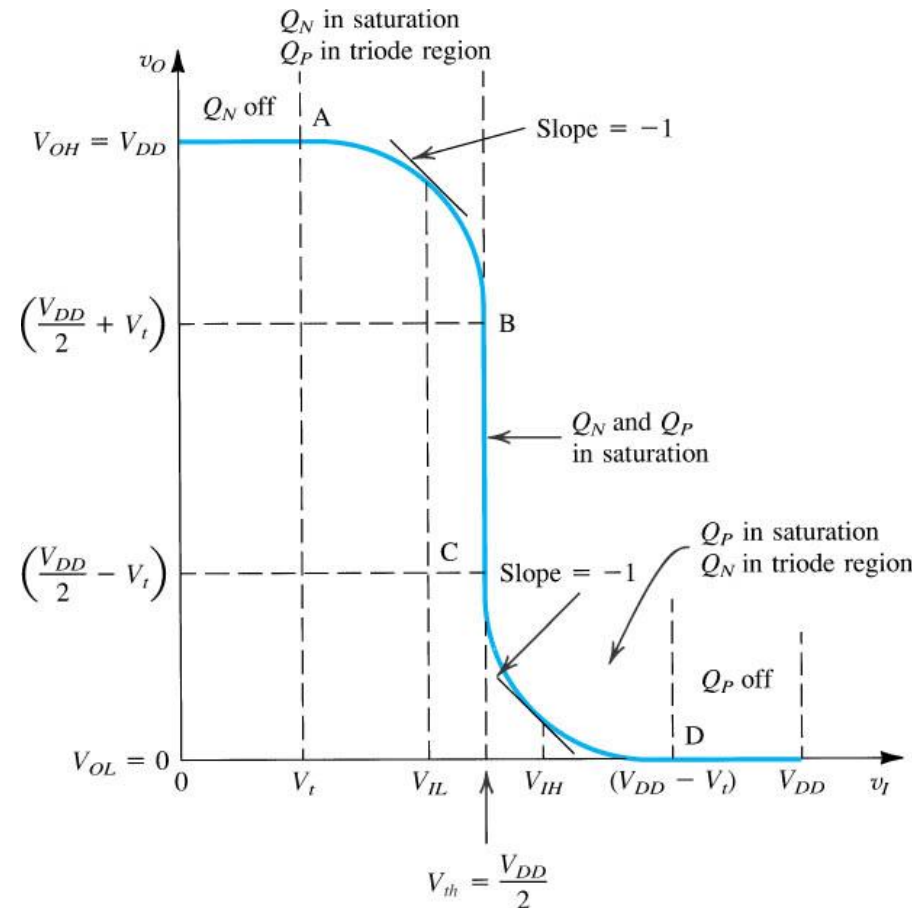
$$NM_H = V_{OH} - V_{IH}$$

V_{IH} = minimum HIGH input voltage

V_{IL} = maximum LOW input voltage

V_{OH} = minimum HIGH output voltage

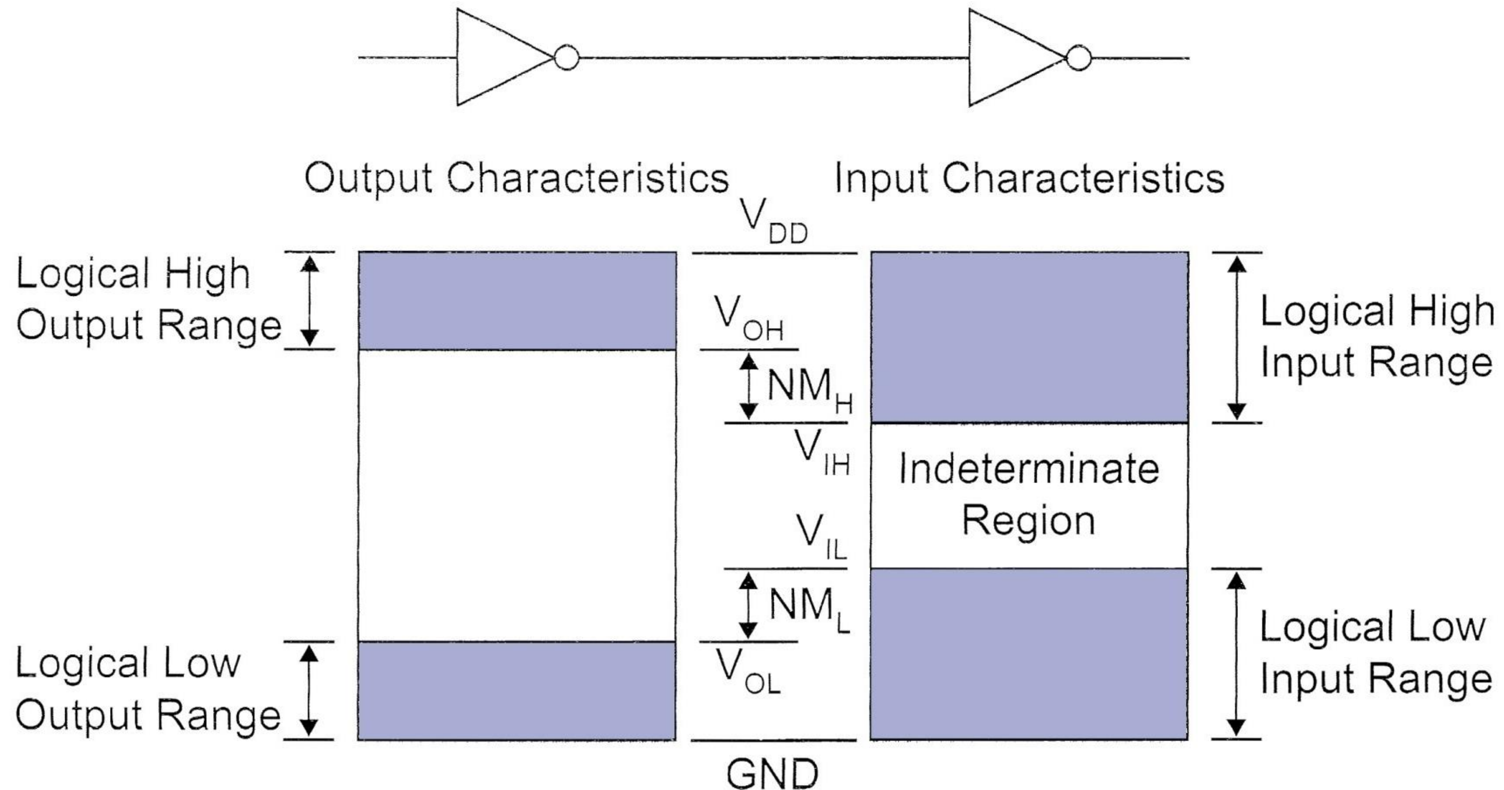
V_{OL} = maximum LOW output voltage



Noise Margin II

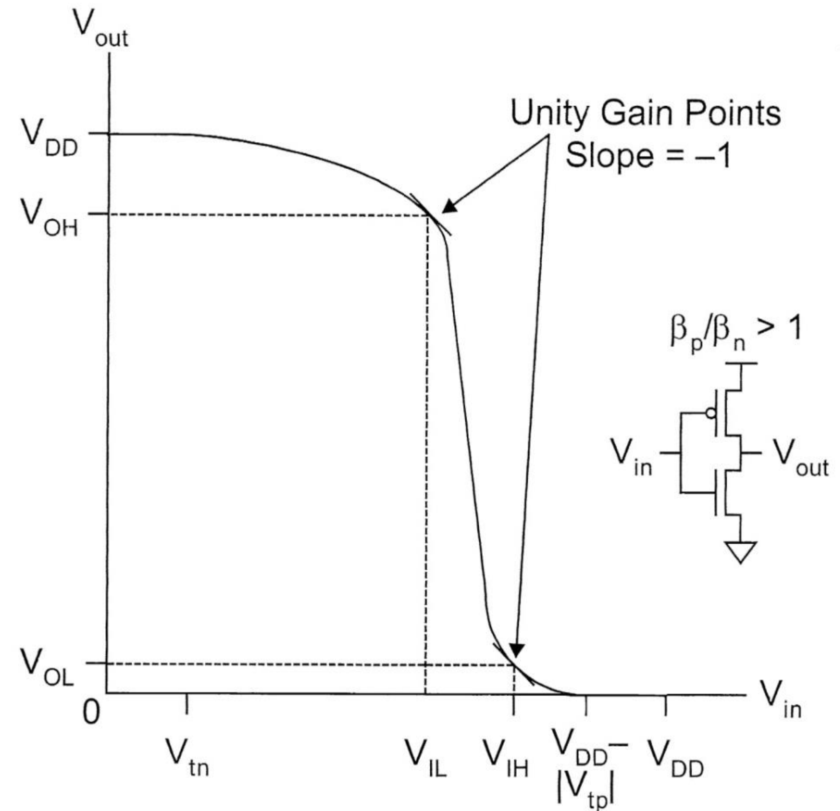
- *Indeterminate region (forbidden zone)*

$$V_{IL} < V_{in} < V_{IH} : output = \text{unknown logic level}$$



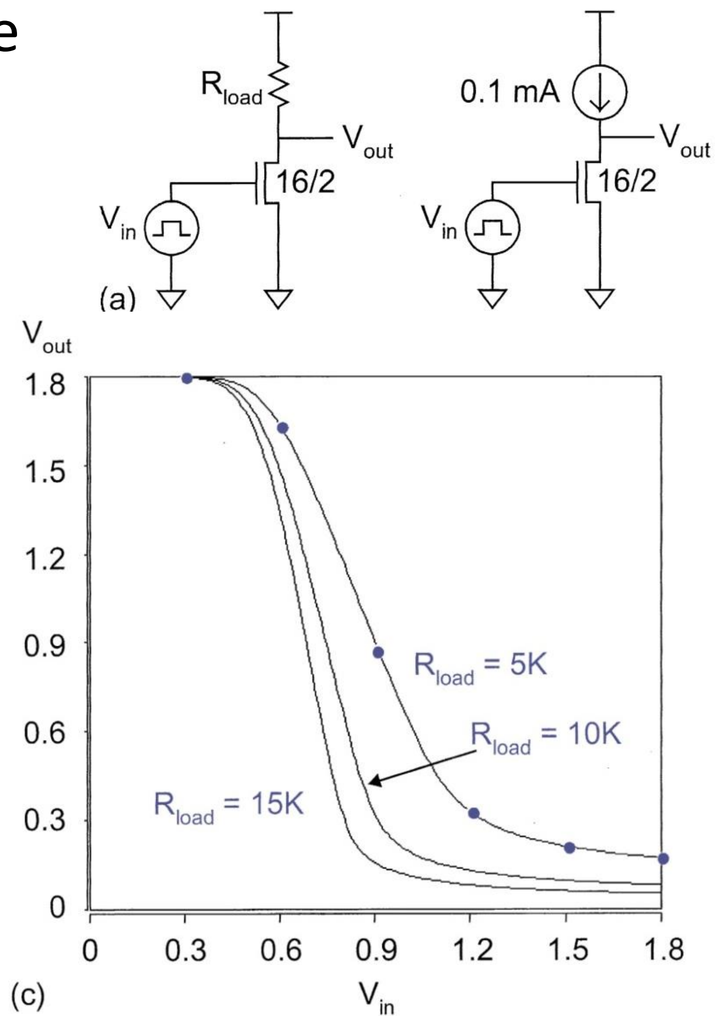
β Ratio and Noise Margin

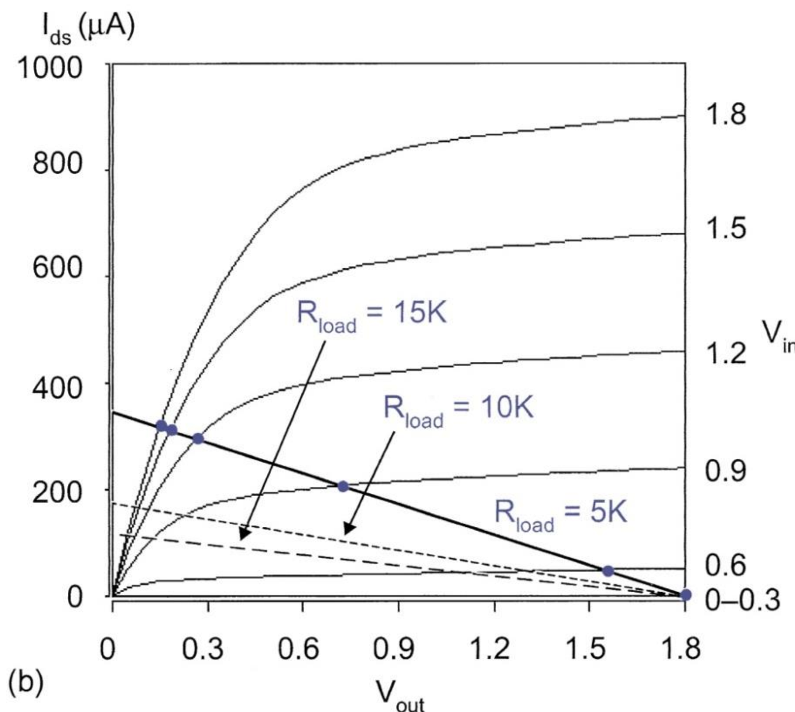
- β ratio > 1
  - Switching threshold > $0.5V_{DD}$
  - V_{IH} ↗ NM_H ↘
  - V_{IL} ↗ NM_L ↗
- β ratio < 1
  - Switching threshold < $0.5V_{DD}$
  - V_{IH} ↘ NM_H ↗
  - V_{IL} ↘ NM_L ↘
  - Noise is scaled with V_{DD}
  - VDD ↘, smaller NM is acceptable.



Ratioed Inverter Transfer Function I

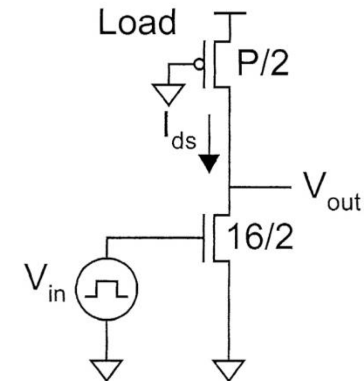
- nMOS inverters with resistive of constant current load
  - Transfer function depends on the ratio of pull-down to the pull-up transistor (*static load*).

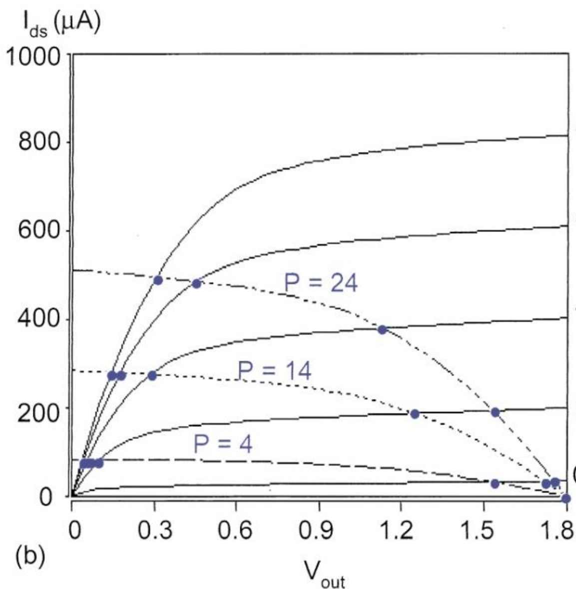


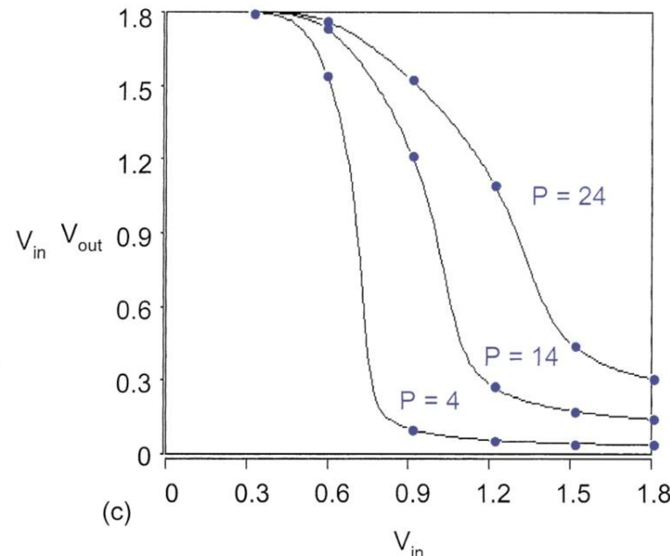


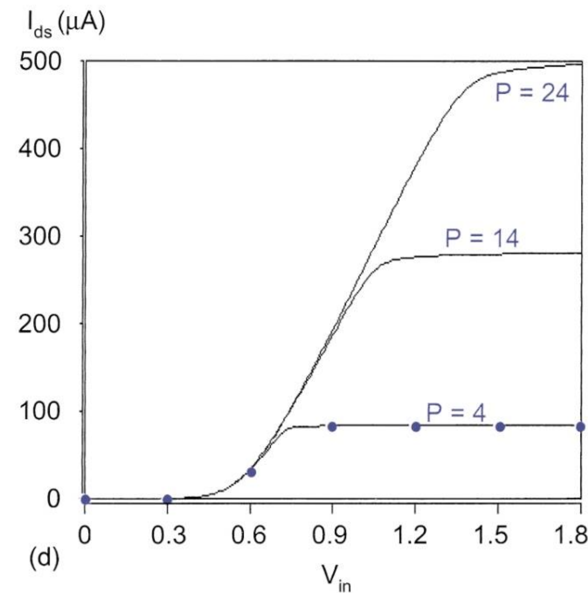
Ratioed Inverter Transfer Function II

- Pseudo-nMOS inverters with turn-ON pMOS as load
  - Turn-ON pMOS is made by a depletion mode nMOS in pure nMOS process.
  - **Dissipating static power when V_o = LOW**
  - **Poor NM but smaller area & input capacitance loading**



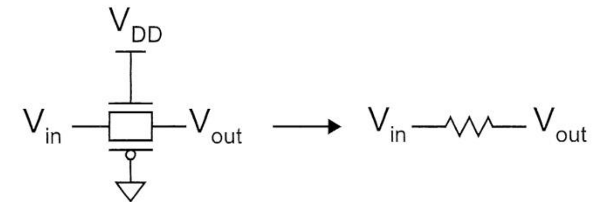


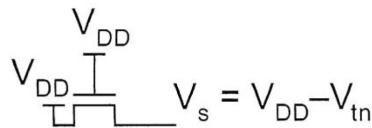


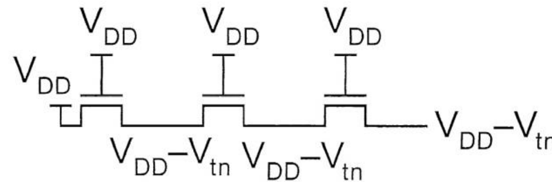


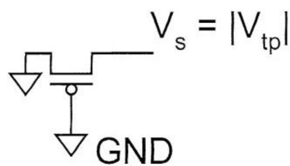
Pass Transistor DC Characteristic

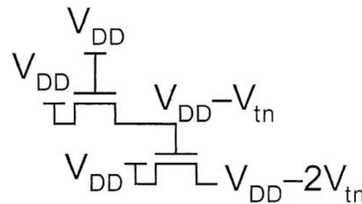
- Threshold drop
  - nMOS : **can NOT** pass "1"
  - pMOS : **can NOT** pass "0"
  - Need to consider **BODY EFFECT.**
- ON resistance depends on V_{in}
  - Need to **BOOST** gate voltage with small V_{DD}

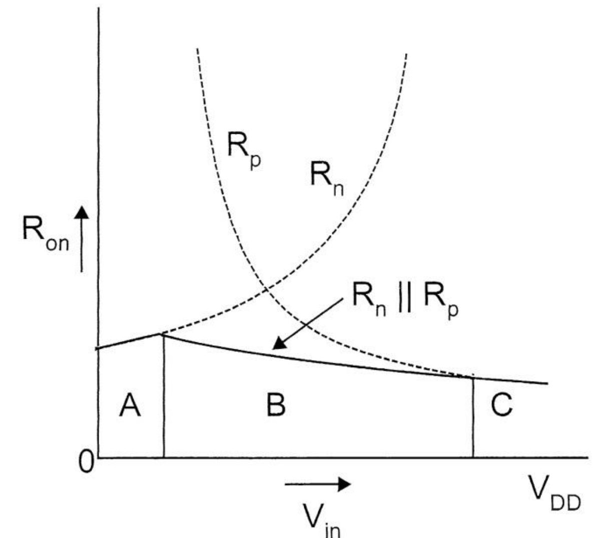






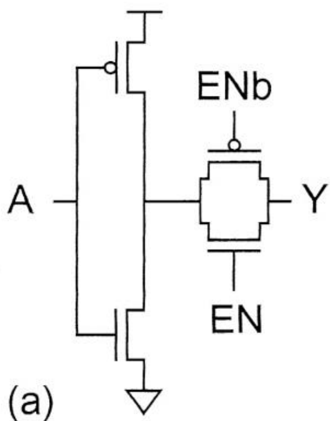


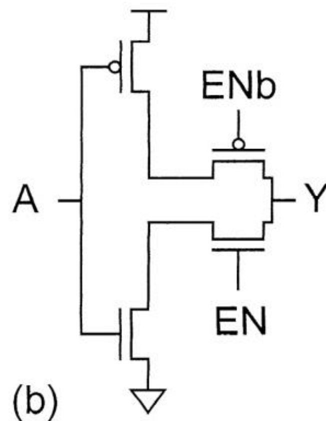


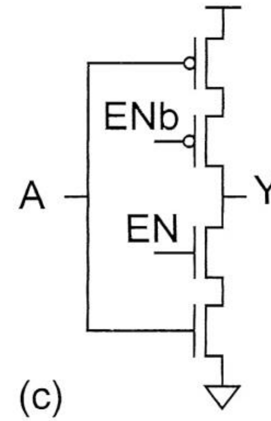


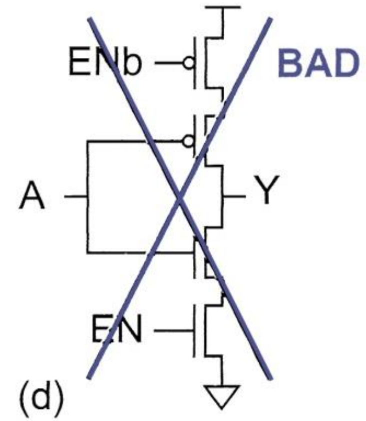
Tristate Inverter

- Inverter + transmission gate
  - For the same size n and p devices, it is approximately half the speed of complementary CMOS inverter.
  - The structure in (d) is suffered from A's toggle in tristate.
  - Need to consider **BODY EFFECT.**









Outline

1. Introduction
2. Ideal I-V Characteristics
3. Nonideal I-V Effects
4. C-V Characteristics
5. DC Transfer Characteristics
6. **Switch-level RC Delay Models**

Effective Resistance *R*

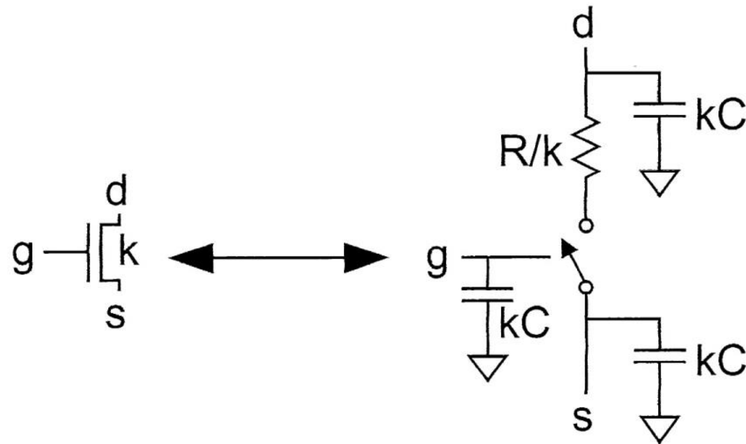
- ***R***: ***effective resistance*** of unit nMOS (minimum W & L)
  - An pMOS has resistance ***2R***(or ***3R***) because of smaller mobility
  - **In linear region**, it is inverse proportional to W/L and V_{gs}

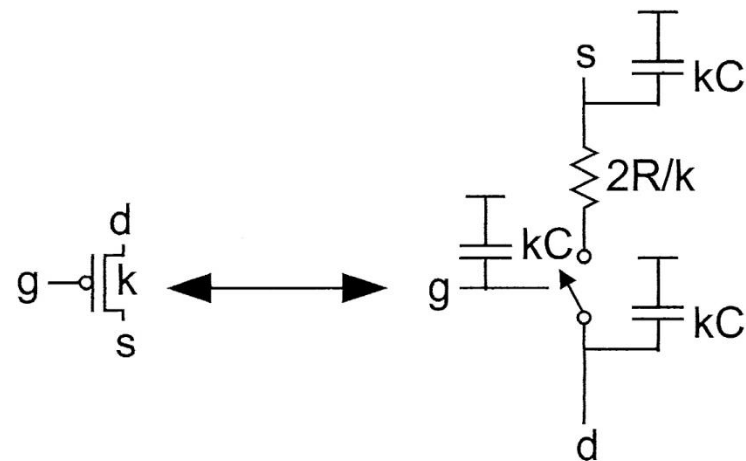
$$R = \left(\frac{\partial I_{ds}}{\partial V_{ds}}\right)^{-1} = \frac{1}{\beta\left(V_{gs}-V_t\right)} = \frac{1}{\mu C_{ox}}\frac{L}{W}\frac{1}{\left(V_{gs}-V_t\right)}$$

- ***C***: ***gate capacitance*** of unit transistor (nMOS & pMOS)
  - It is proportional to gate area W*L
- ***C***: ***junction capacitance*** of S/D of unit transistor
  - It is proportional to gate width W
- nMOS of ***k*** times unit width has resistance ***R/k***, gate capacitance ***kC*** and S/D capacitance ***kC***.

RC Circuit Model

- pMOS of ***k*** times unit width has resistance 2***R/k***, gate capacitance ***kC*** and S/D capacitance ***kC***
  - nMOS parasitic capacitors are referred to GND (p-body) and pMOS parasitic capacitors are referred to VDD (n-well)

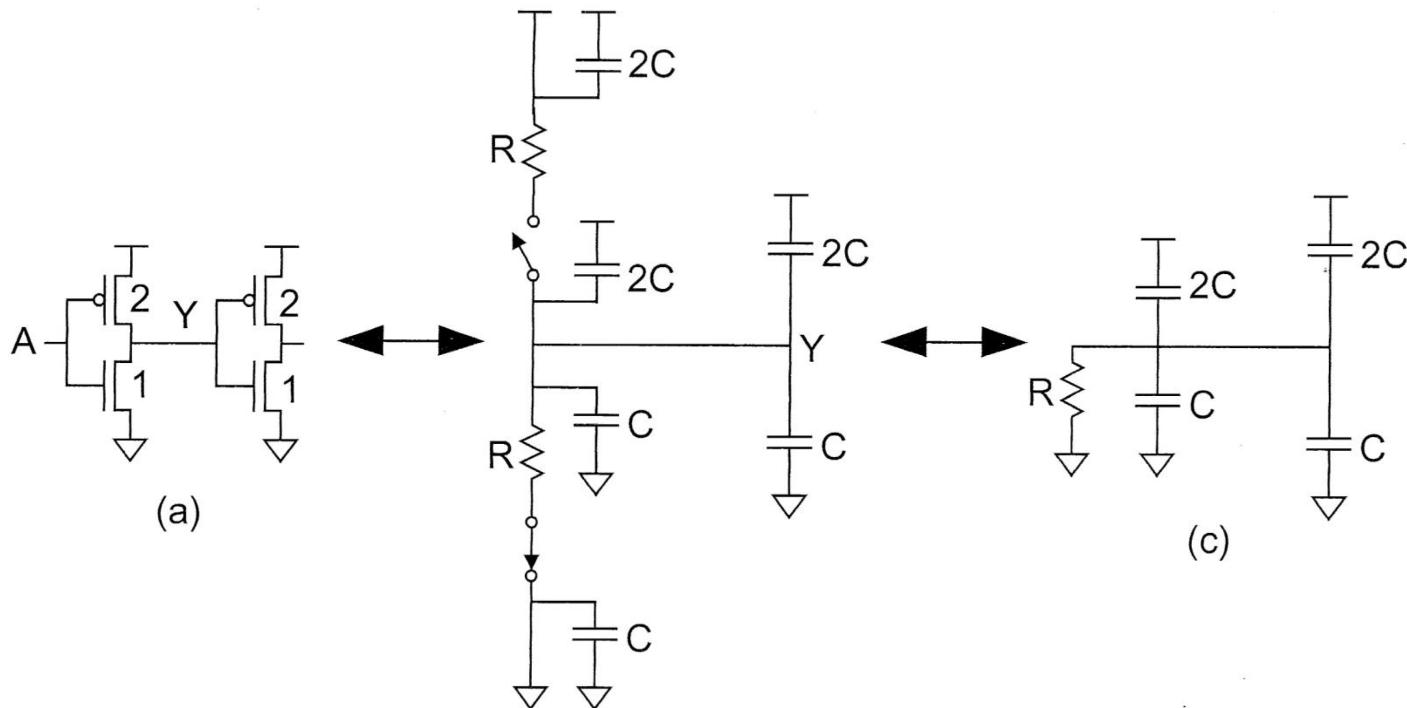




Inverter Propagation Delay

- Fanout-of-1 Inverter
  - Choose pMOS width size to be x2~x3 of that of nMOS

$$t_{pd} = R \bullet (6C) = 6RC$$



R of Transmission Gate

- Effective resistance of transmission gate is the parallel combination of nMOS and pMOS
  - It depends on the signal to be pass
  - pMOS pass 0 weakly with larger resistance 4R
  - nMOS pass 1 weakly with larger resistance 2R
  - Usually choose same size of nMOS/pMOS in transmission gate
  - Increase MOS size → ***R***↘ ***C***↗: need to check the tradeoff