CMOS Transistor and Circuits

Instructed by Shmuel Wimer
Eng. School, Bar-Ilan University

Credits: David Harris
Harvey Mudd College

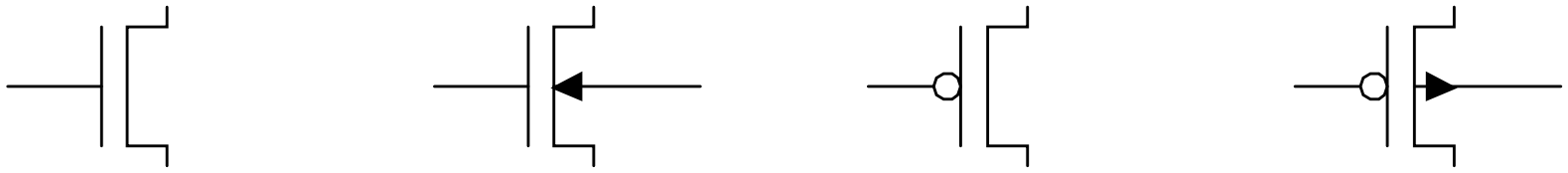
(Some materials copied/taken/adapted from Harris' lecture notes)

Outline

- MOS Capacitor
- nMOS I-V Characteristics
- pMOS I-V Characteristics
- DC characteristics and transfer function
- Noise margin
- Latchup
- Pass transistors
- Tristate inverter

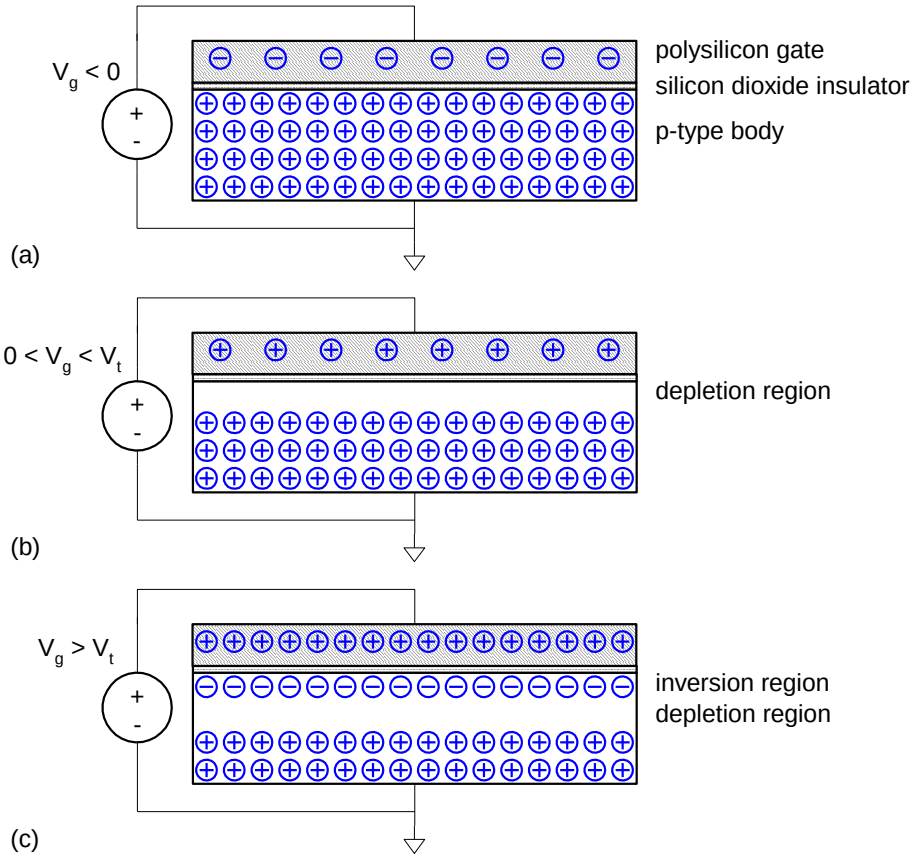
Introduction

- 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



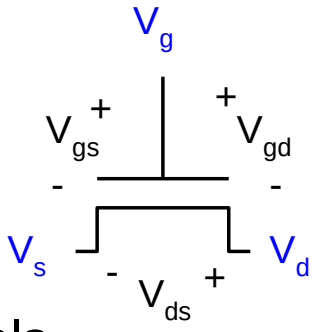
MOS Capacitor

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



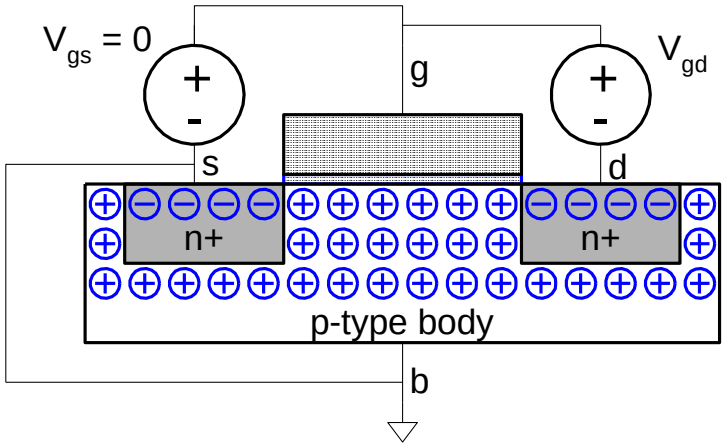
Terminal Voltages

- 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*



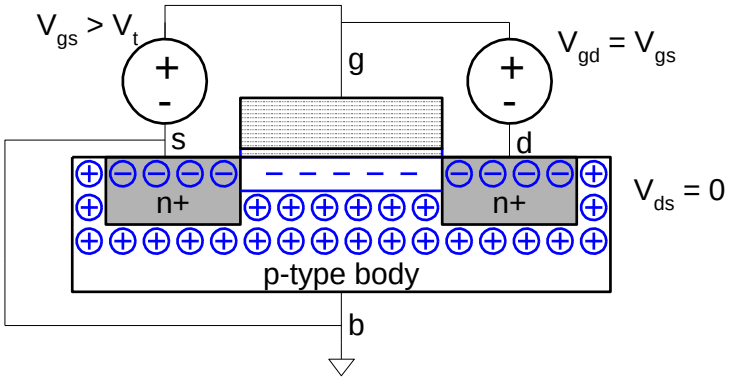
nMOS Cutoff

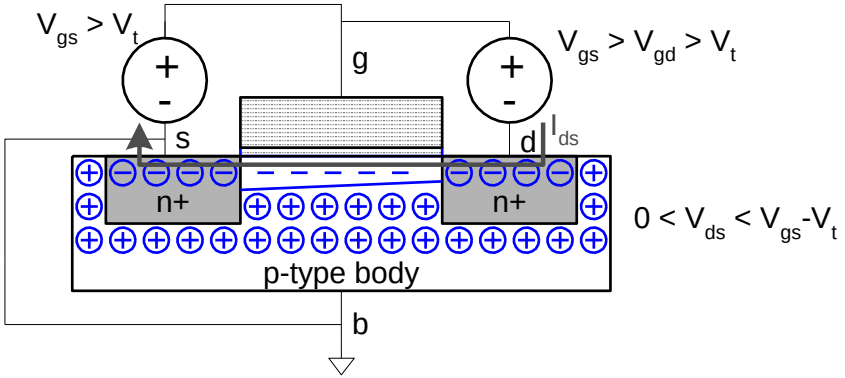
- No channel
- $I_{ds} = 0$



nMOS Linear

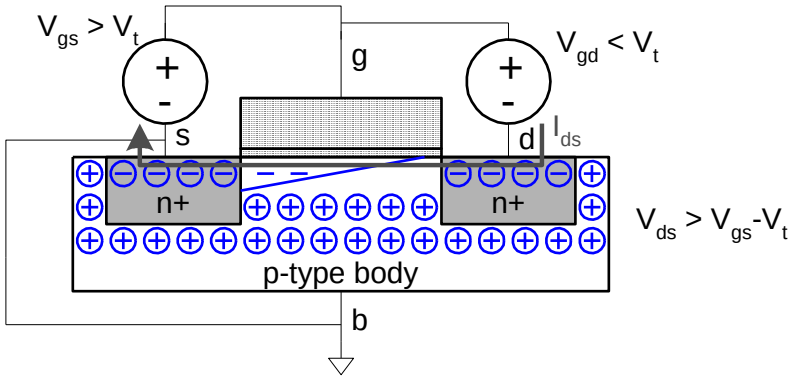
- Channel forms
- Current flows from d to s
  - e^- from s to d
- I_{ds} increases with V_{ds}
- Similar to linear resistor





nMOS Saturation

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

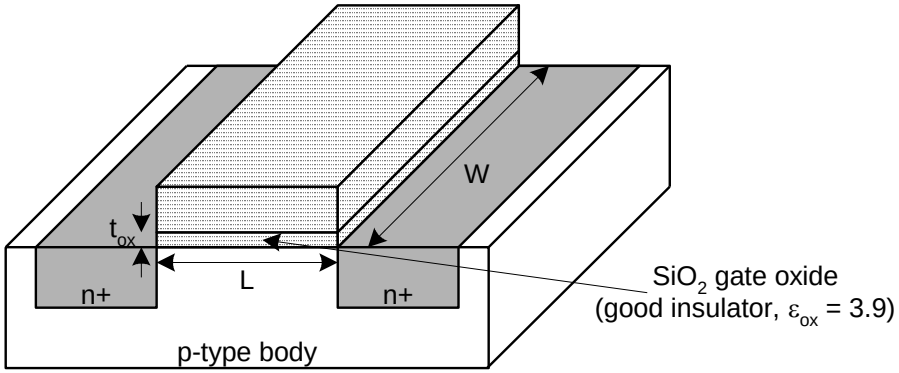


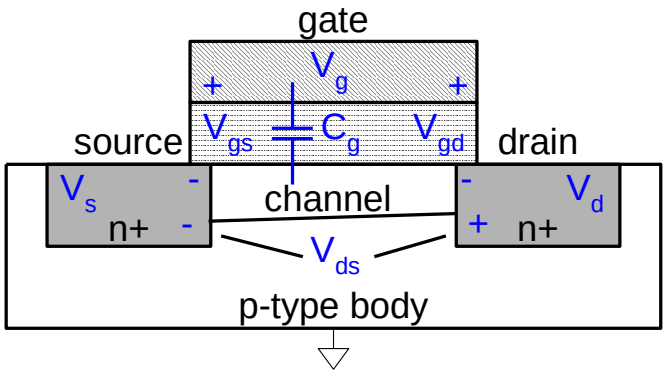
I-V Characteristics

- 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
- $Q_{channel} = CV$
- $C = C_g = \varepsilon_{ox}WL/t_{ox} = C_{ox}WL$
- $V = V_{gc} - V_t = (V_{gs} - V_{ds}/2) - V_t$ ($V_{gc} - V_t$ is the amount of voltage attracting charge to channel beyond the voltage required for inversion)





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:
  - $t = L / v$

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

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

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

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

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

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

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

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

nMOS I-V Summary

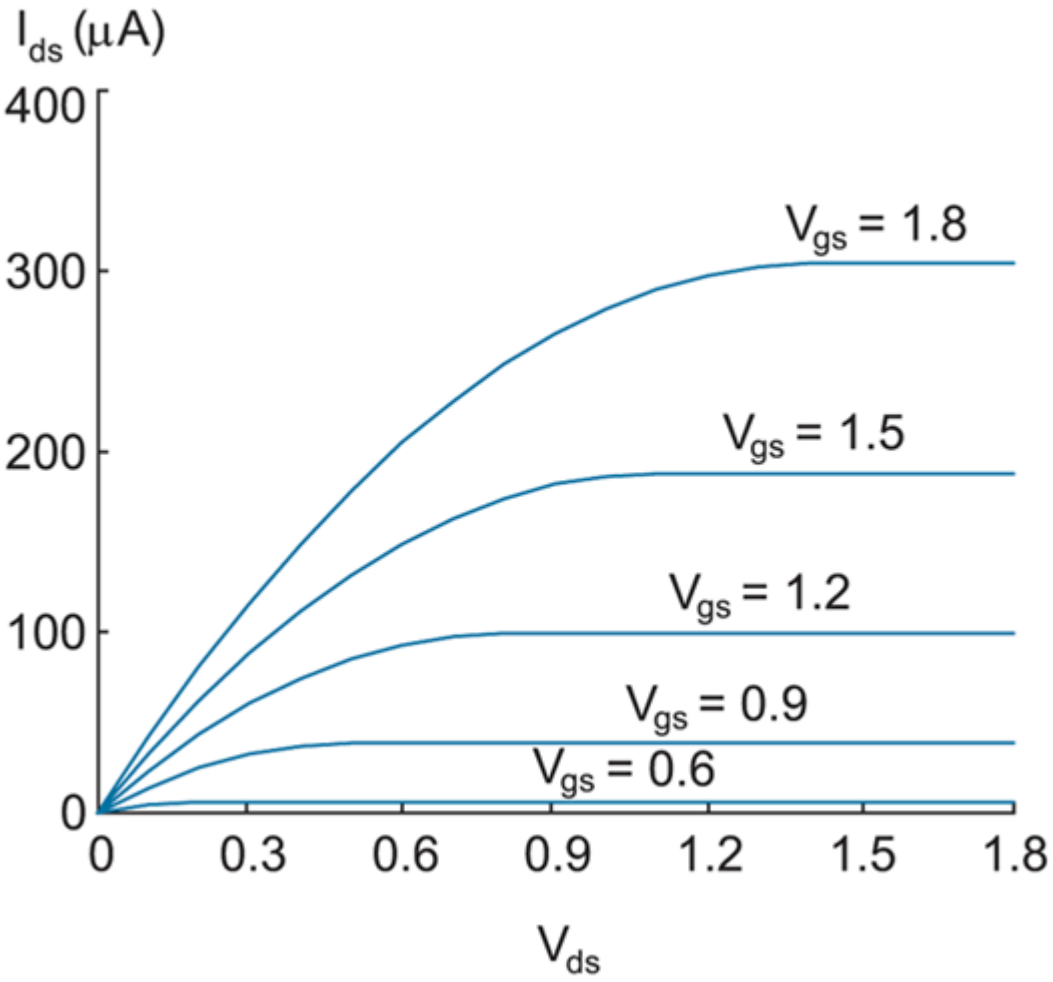
- *Shockley* 1st order transistor models

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

Example

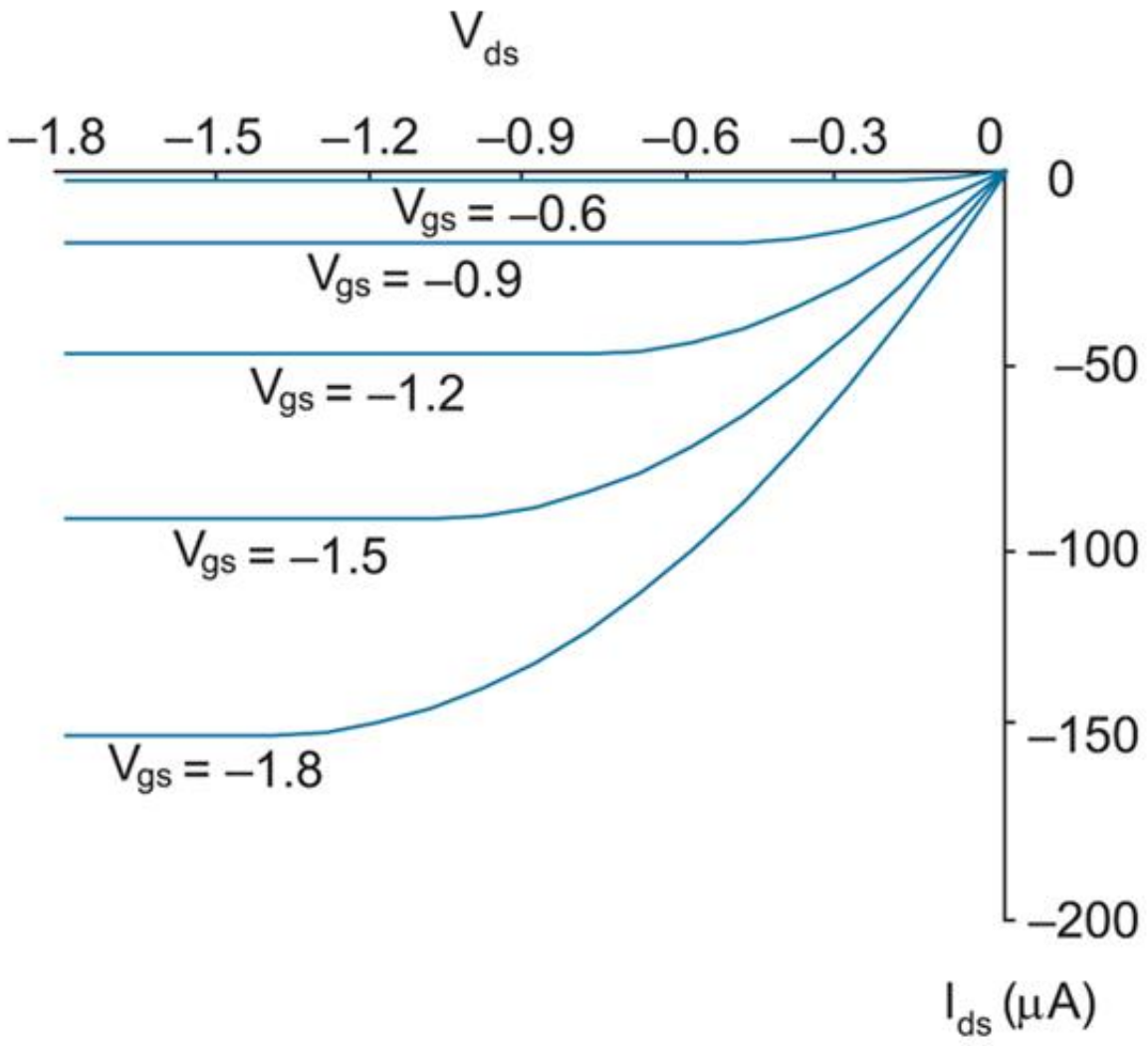
- We will be using a 0.18 μm process for your project
  - t_{ox} = 40 Å
  - μ = 180 $cm^2/V*s$
  - V_t = 0.4 V
- Plot I_{ds} vs. V_{ds}
  - V_{gs} = 0, 0.3, 0.6, 0.9, 1.2, 1.5 and 1.8V.
  - Use W/L = 4/2 λ

$$\beta = \mu C_{ox}\frac{W}{L} = (350)\left(\frac{3.9 \bullet 8.85 \cdot 10^{-14}}{100 \cdot 10^{-8}}\right)\left(\frac{W}{L}\right) = 155\frac{W}{L}\,\mu A/V^2$$


I_{ds} (μA)
400
300
200
100
0
V_{gs} = 1.8
V_{gs} = 1.5
V_{gs} = 1.2
V_{gs} = 0.9
V_{gs} = 0.6
0
0.3
0.6
0.9
1.2
1.5
1.8
V_{ds}

pMOS I-V

- All doping and voltages are inverted for pMOS
- Mobility μ_p is determined by holes
  - Typically 2-3x lower than that of electrons μ_n
- Thus pMOS must be wider to provide same current
  - In this class, assume $\mu_n / \mu_p = 2$

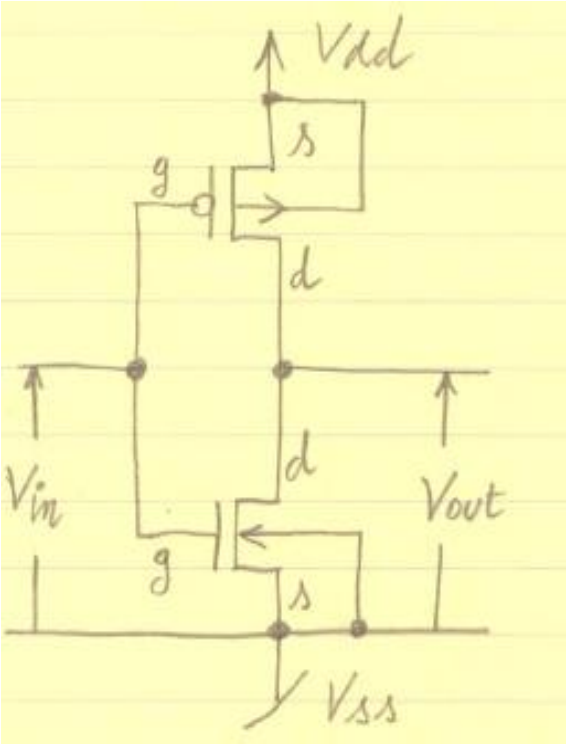

V_{ds}
–1.8
–1.5
–1.2
–0.9
–0.6
–0.3
0
0
–50
–100
–150
–200
V_{gs} = –0.6
V_{gs} = –0.9
V_{gs} = –1.2
V_{gs} = –1.5
V_{gs} = –1.8
I_{ds} (μA)

DC Transfer Characteristics

Objective: Find the variation of output voltage V_{out} for changes in input voltage V_{in}.

V_{tp} – Threshold voltage of p-device

V_{tn} – Threshold voltage of n-device



	CUTOFF	NON SATURATED	SATURATED
P-device	$V_{gsp} > V_{tp}$	$V_{gsp} < V_{tp}$	$V_{gsp} < V_{tp}$
		$V_{in} < V_{tp} + V_{DD}$	$V_{in} < V_{tp} + V_{DD}$
	$V_{in} > V_{tp} + V_{DD}$	$V_{dsp} > V_{gsp} - V_{tp}$	$V_{dsp} < V_{gsp} - V_{tp}$
		$V_{out} > V_{in} - V_{tp}$	$V_{out} < V_{in} - V_{tp}$
n-device	$V_{gsn} < V_{tn}$	$V_{gsn} > V_{tn}$	$V_{gsn} > V_{tn}$
		$V_{in} > V_{tn}$	$V_{in} > V_{tn}$
	$V_{in} < V_{tn}$	$V_{dsn} < V_{gsn} - V_{tn}$	$V_{dsn} > V_{gsn} - V_{tn}$
		$V_{out} < V_{in} - V_{tn}$	$V_{out} > V_{in} - V_{tn}$

Recall CMOS device

cutoff: $I_{ds} = 0 \quad V_{gs} \leq V_t$

linear: $I_{ds} = \beta\left[(V_{gs} - V_t)V_{ds} - \frac{V_{ds}^2}{2}\right] \quad 0 < V_{ds} < V_{gs} - V_t$

saturation: $I_{ds} = \beta\frac{(V_{gs} - V_t)^2}{2} \quad 0 < V_{gs} - V_t < V_{ds}$

CMOS inverter characteristics is derived by solving for $V_{inn}=V_{inp}$ and $I_{dsn}=-I_{dsp}$

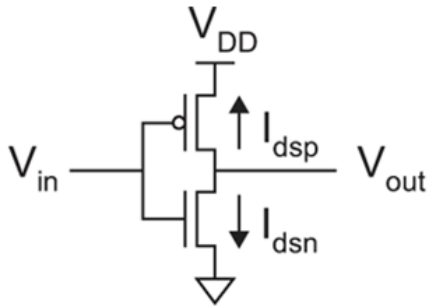


FIG 2.23 A CMOS inverter

CMOS inverter is divided into five regions of operation

Region A: $0 \le V_{in} \le V_{tn}$

n-device in cutoff, $I_{ds} = 0$

p-device in linear region, $I_{dsn} = -I_{dsp}$, $I_{dsp} = 0$

$V_{dsp} = 0 \Rightarrow V_{out} = V_{DD}$

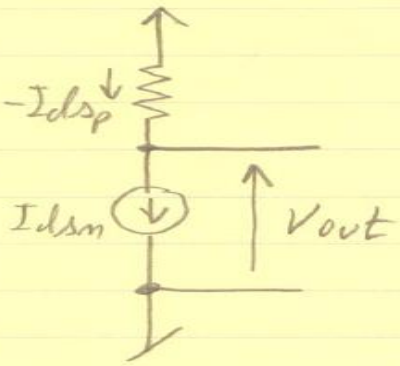
Region B: $V_{tn} \le V_{in} < V_{DD}/2$

n-device is saturated (current source)

p-device is nonsaturated (linear, resistor)

$$I_{dsn} = \beta_n \frac{(V_{in} - V_{tn})^2}{2}$$

$$\beta_n = \frac{\mu_n \varepsilon}{t_{ox}} \left(\frac{W_n}{L_n} \right)$$



In p-device $V_{gs} = (V_{in} - V_{DD})$

$$V_{DS} = (V_{out} - V_{DD})$$

$$I_{ds_p} = -\beta_p \left[(V_{in} - V_{DD} - V_{tp})(V_{out} - V_{DD}) - \frac{(V_{out} - V_{DD})^2}{2} \right]$$

$$\beta_p = \frac{\mu_p \varepsilon}{t_{ox}} \left(\frac{W_p}{L_p} \right)$$

Substitute $I_{dsp} = -I_{dsn}$

$$V_{out} = (V_{in} - V_{tp}) + \sqrt{(V_{in} - V_{tp})^2 - 2\left(V_{in} - \frac{V_{DD}}{2} - V_{tp}\right) - \frac{\beta_n}{\beta_p}(V_{in} - V_{tn})^2}$$

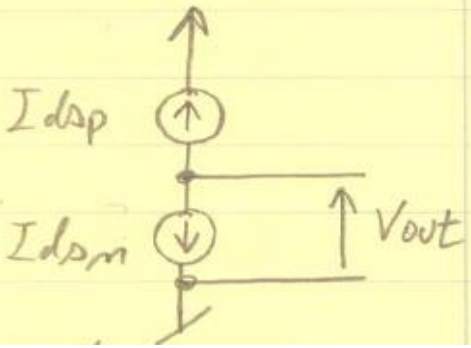
Region C:

n-device still in saturation

p-device enters saturation

$$I_{dsp} = -\frac{\beta_p}{2}(V_{in} - V_{DD} - V_{tp})^2$$

$$I_{ds_n} = \frac{\beta_n}{2}(V_{in} - V_{tn})^2$$



setting $\beta_n = \beta_p$ and $V_{tn} = -V_{tp}$ and using $I_{ds_n} = -I_{dsp}$, we obtain

$$V_{in} = \frac{V_{DD}}{2}$$

Region c exist only for one value of V_{in}.

Possible values of V_{out} are

n-device $V_{out} > V_{in} - V_{tn}$

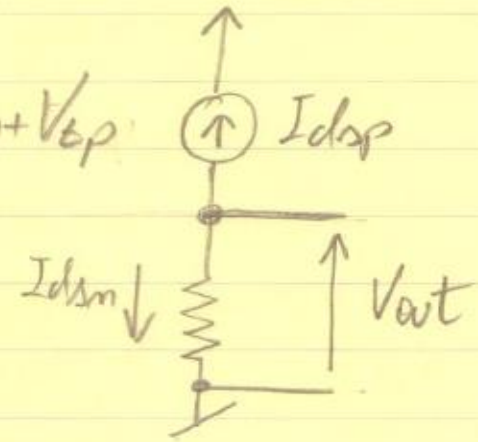
p-device $V_{out} < V_{in} - V_{tp}$

concluding $V_{in} - V_{tn} < V_{out} < V_{in} - V_{tp}$

Region D: $\frac{V_{DD}}{2} < V_{in} \leq V_{DD} + V_{tp}$

n-device **linear**

p-device **saturated**



$$V_{out} = (V_{in} - V_{tn}) - \sqrt{(V_{in} - V_{tn})^2 - \frac{\beta_p}{\beta_n}(V_{in} - V_{DD} - V_{tp})^2}$$

Region E: $V_{in} > V_{DD} + V_{tp}$

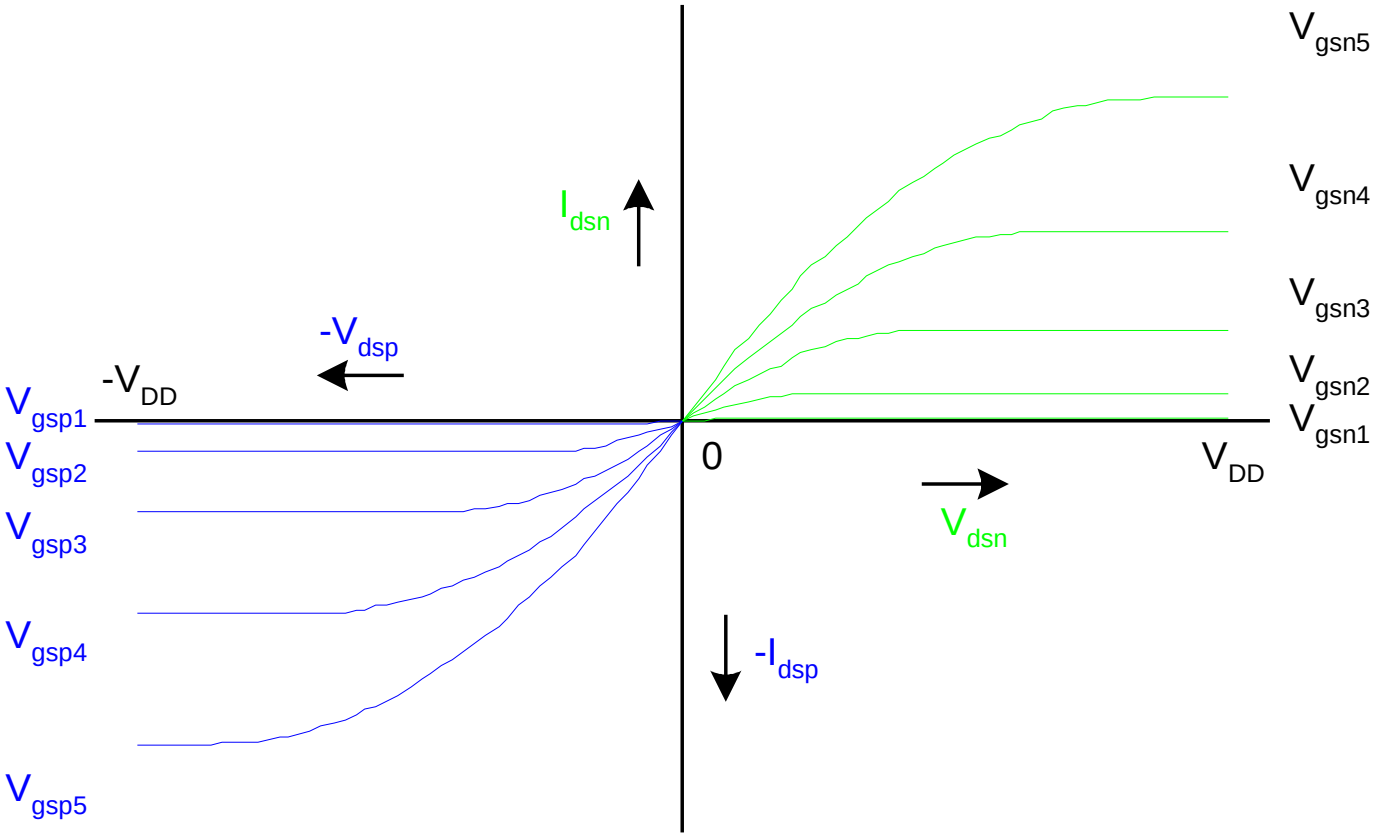
n-device linear

p-device cutoff

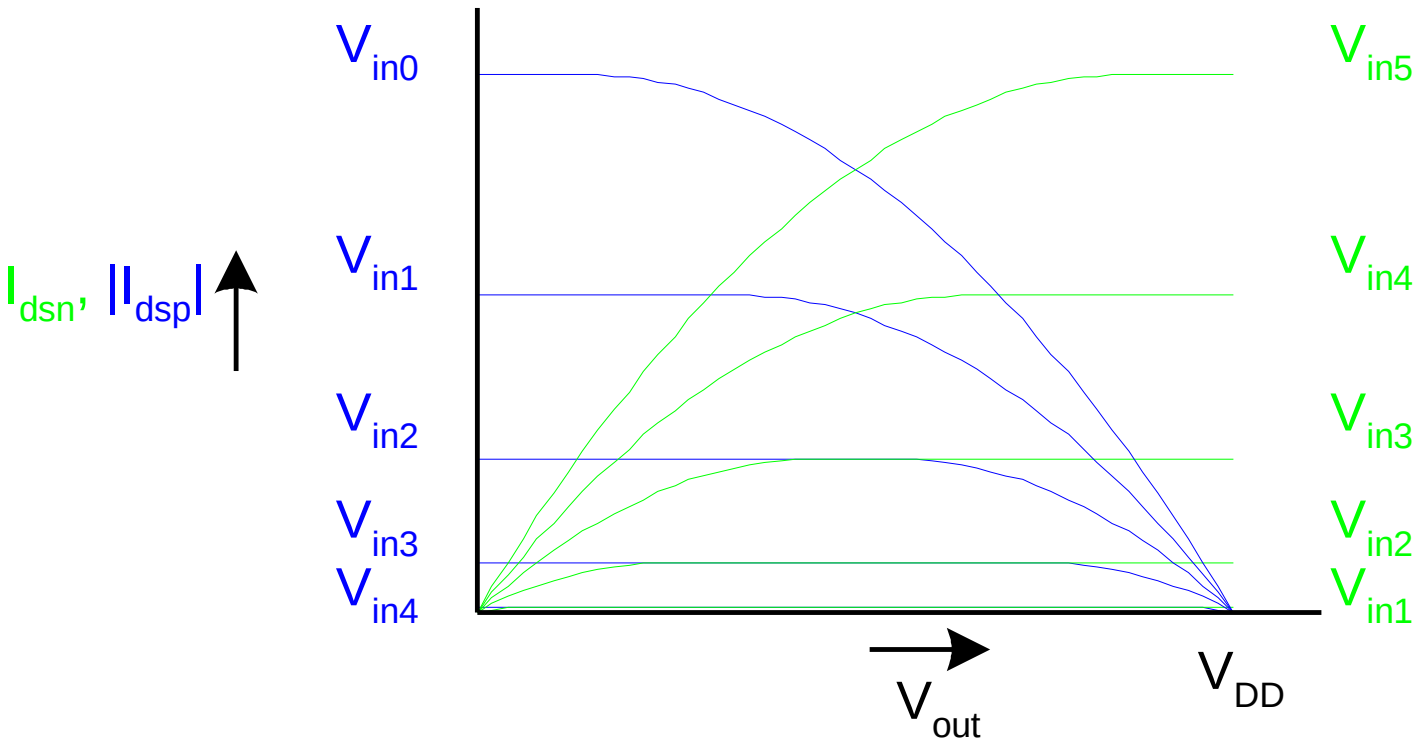
$V_{ds_n} = 0 \Rightarrow$ **$V_{out} = 0$**

I-V Characteristics

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

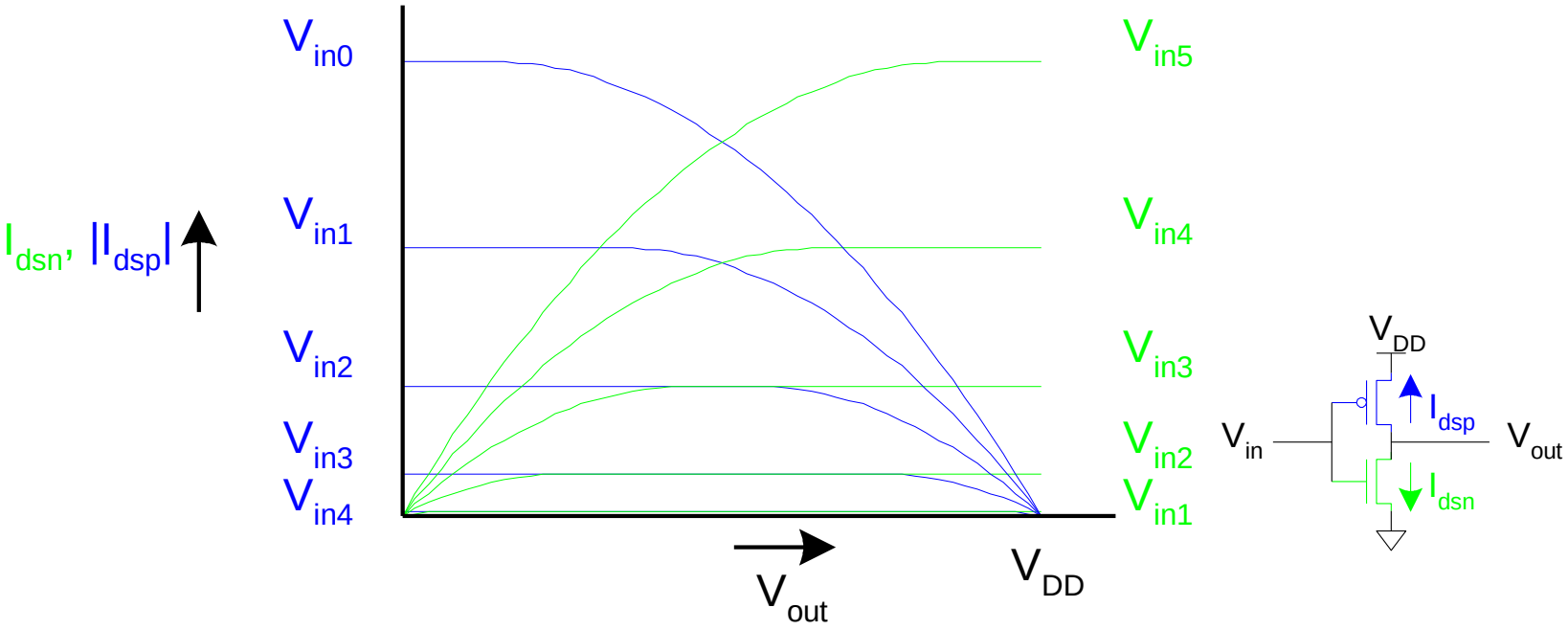


Current vs. V_{out}, V_{in}

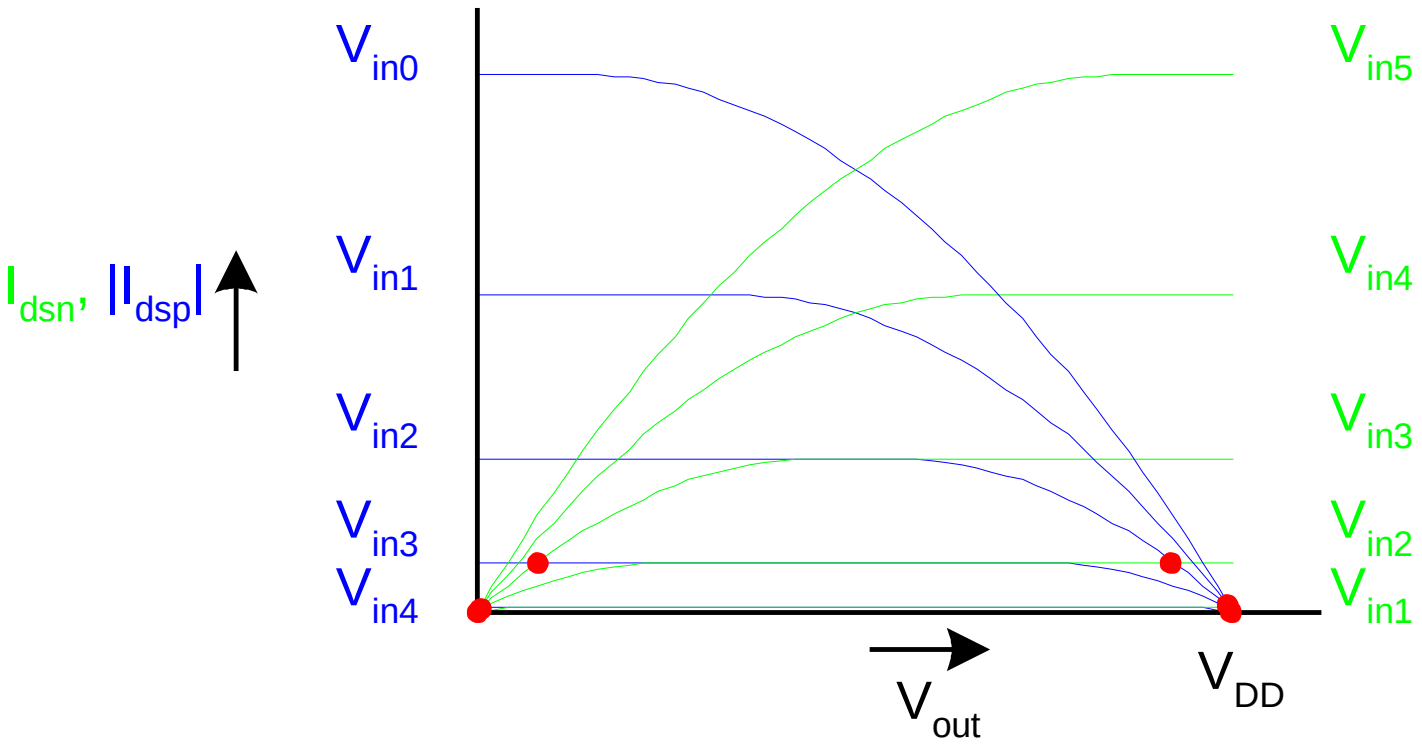


Load Line Analysis

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

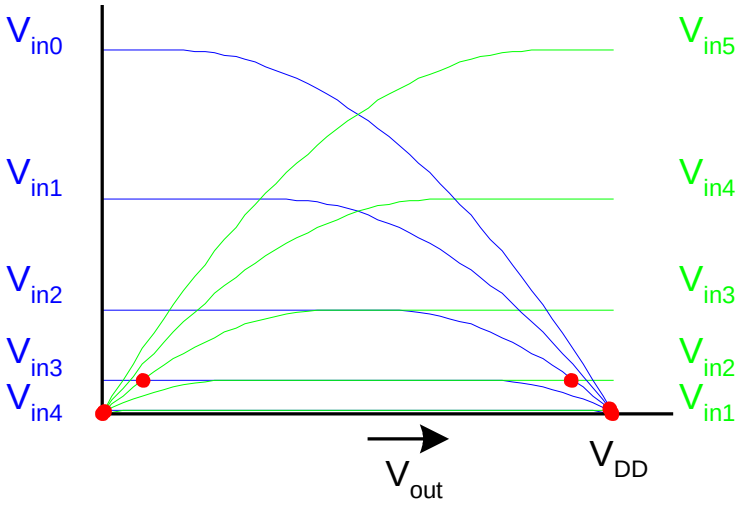


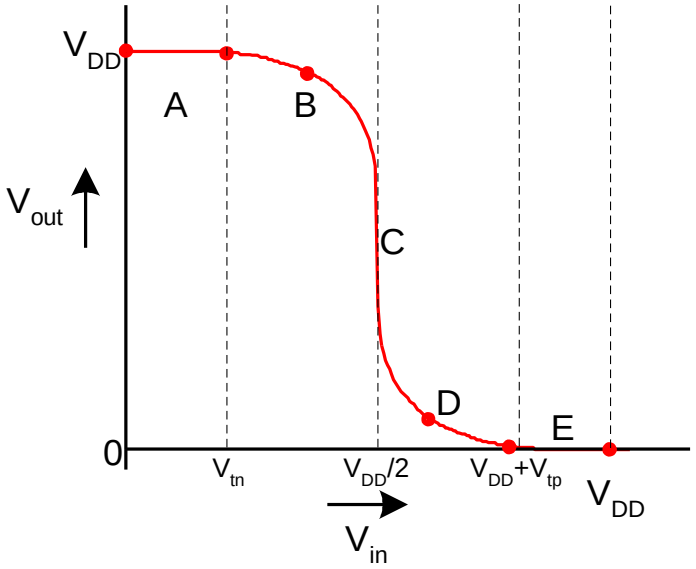
Load Line Summary



DC Transfer Curve

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

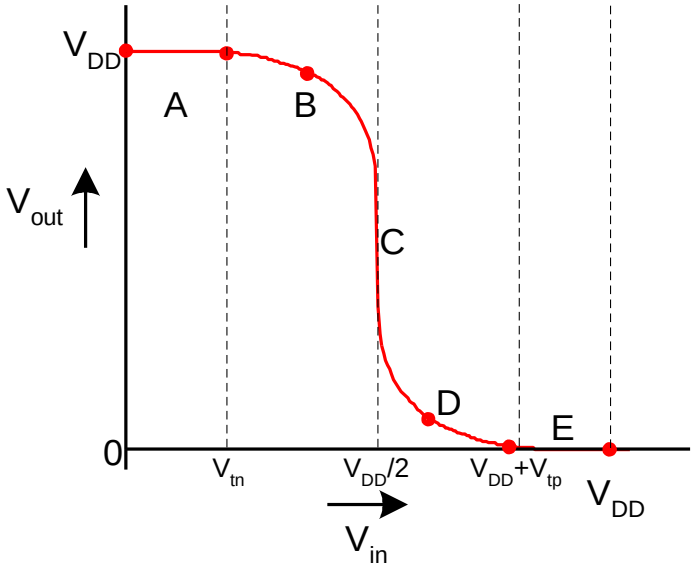




Operating Regions

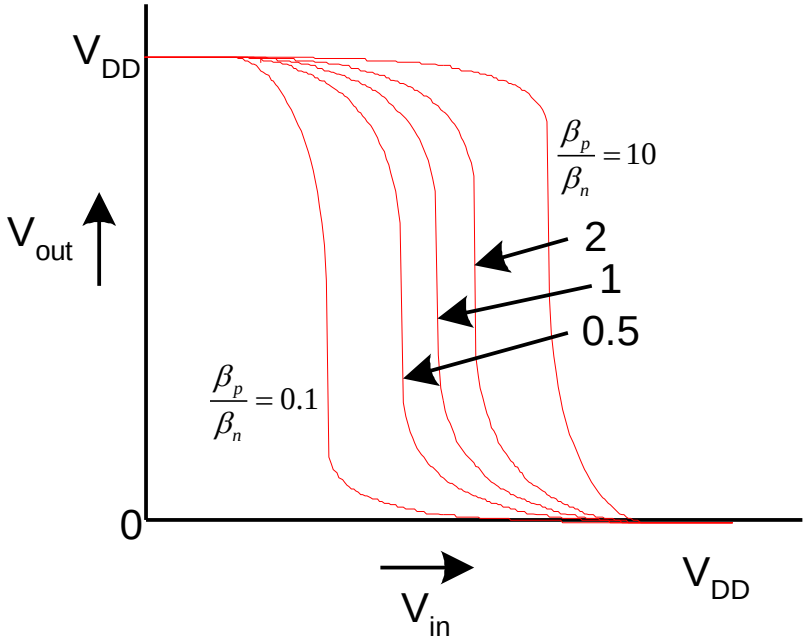
- Revisit transistor operating regions

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

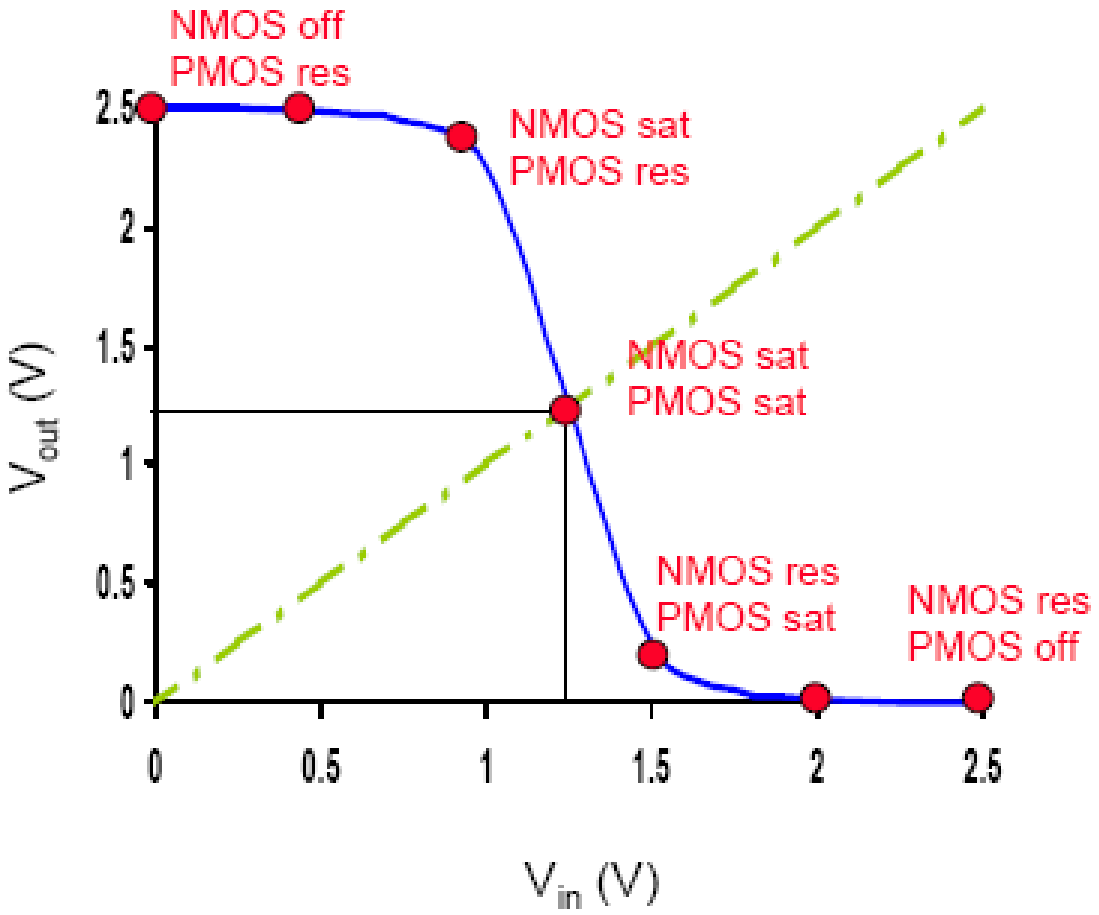


Beta Ratio

- If $\beta_p / \beta_n \neq 1$, switching point will move from $V_{DD}/2$
- Called *skewed* gate
- Other gates: collapse into equivalent inverter



DC Transfer function is symmetric for $\beta_n=\beta_p$



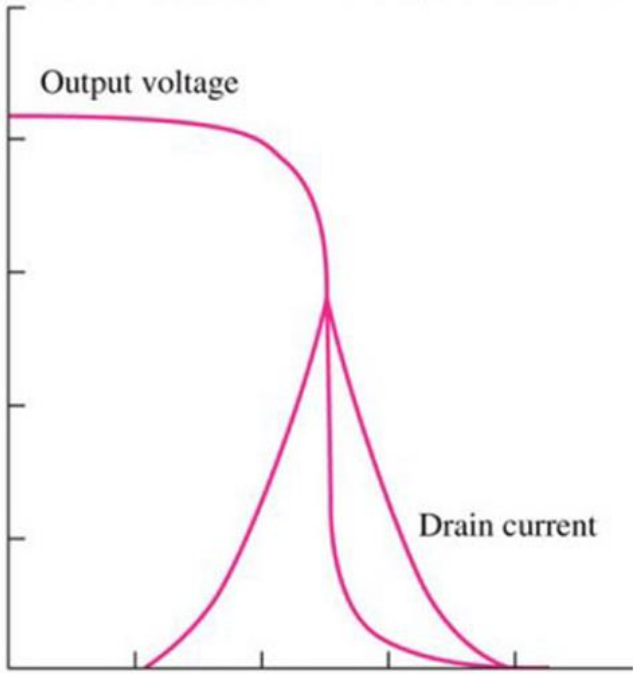


Table 2.3 Summary of CMOS inverter operation

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	$V_{out} = 0$

Noise Margin

It determines the allowable noise at the input gate (0/1) so the output (1/0) is not affected

Noise margin is closely related to input-output transfer function

It is derived by driving two inverters connected in series

Output Characteristics Input Characteristics

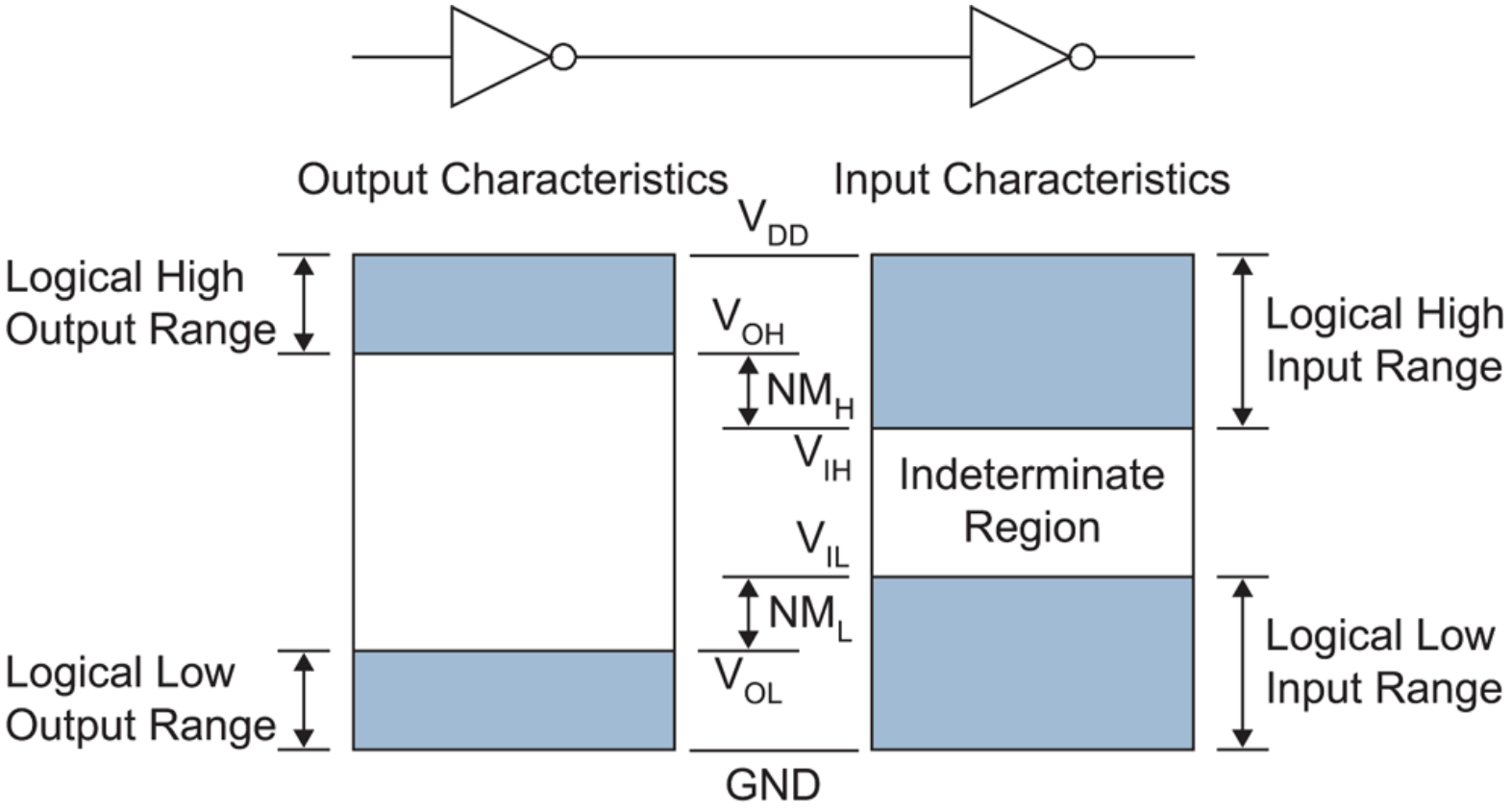
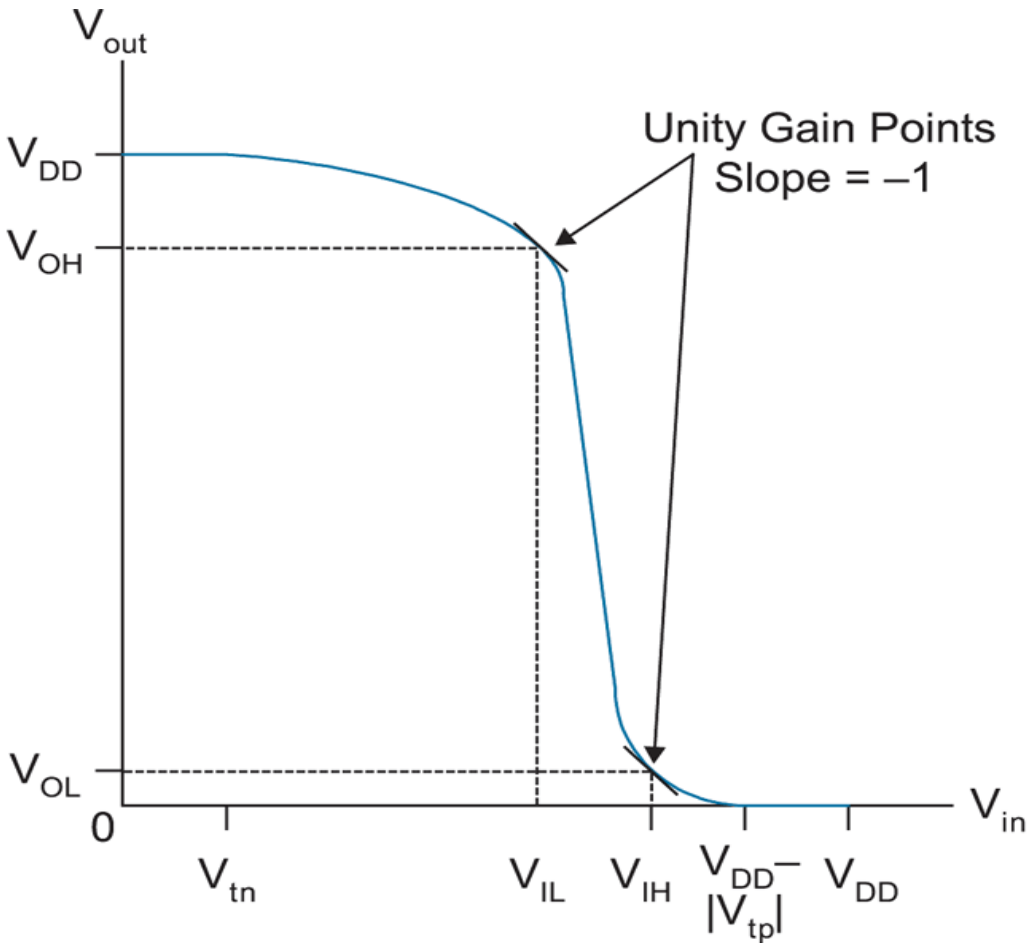


FIG 2.27 Noise margin definitions


V_out
V_DD
V_OH
V_OL
0
Unity Gain Points
Slope = –1
V_in
V_tn
V_IL
V_IH
V_DD – |V_tp|
V_DD

Impact of skewing transistor size on noise margin

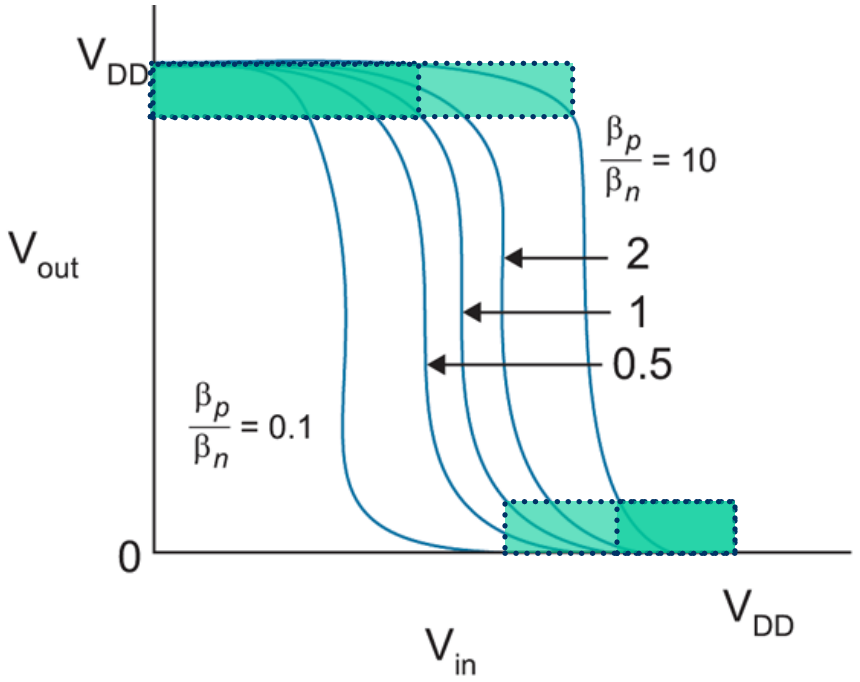
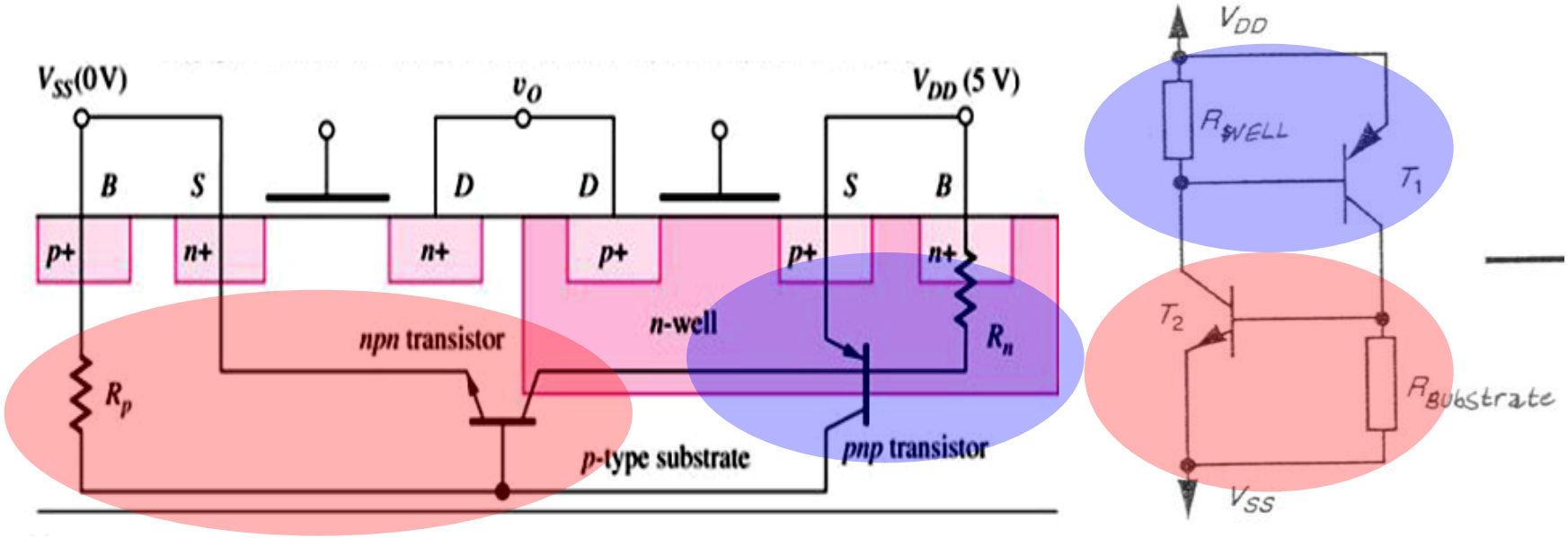


FIG 2.26 Transfer characteristics of skewed inverters

Increasing (decreasing) P / N ratio increases (decreases) the low noise margin and decreases (increases) the high noise margin

Latchup in CMOS Circuits



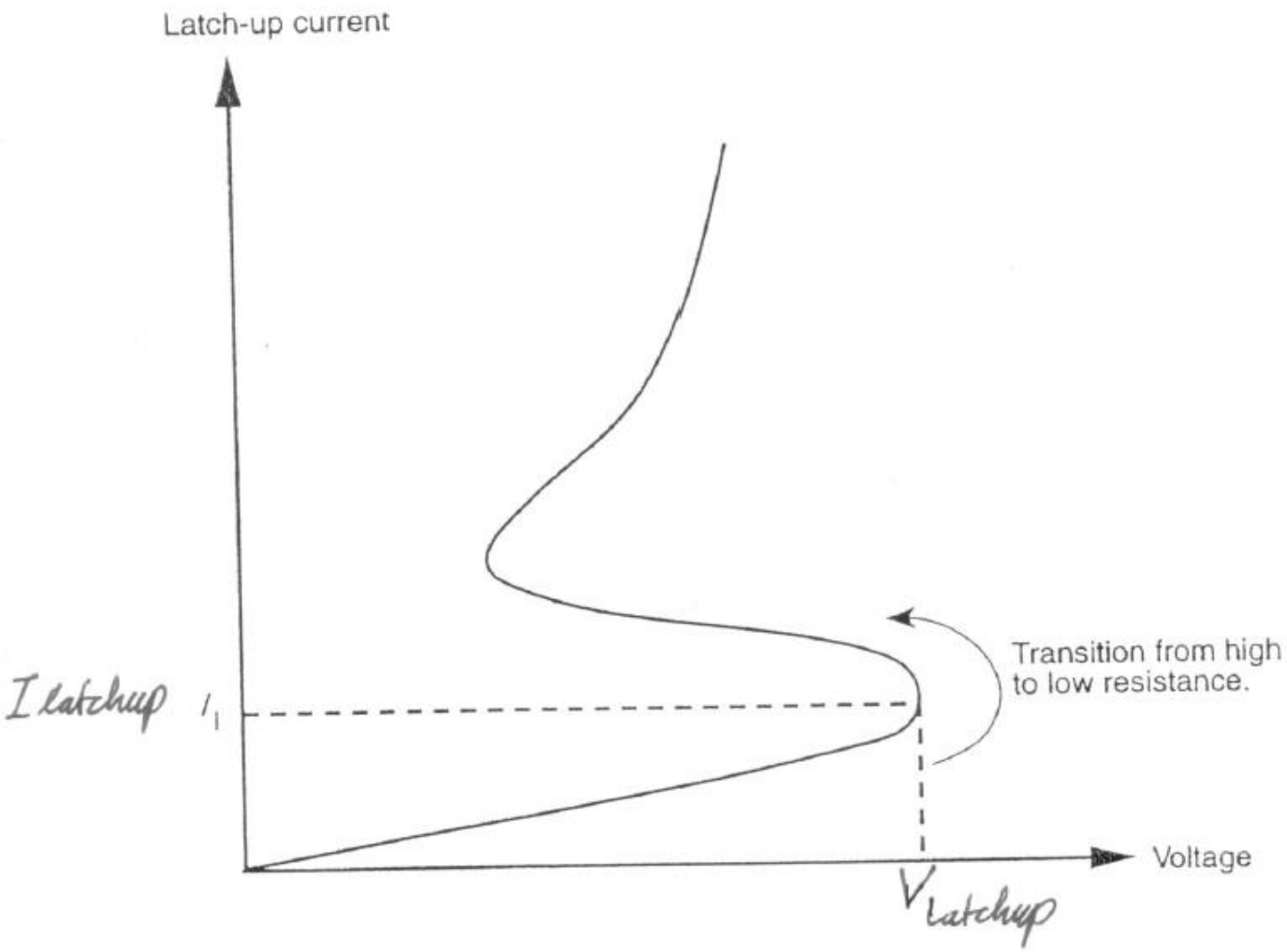
Parasitic bipolar transistors are formed by substrate and source / drain devices

Latchup occurs by establishing a low-resistance paths connecting V_{DD} to V_{SS}

Latchup may be induced by power supply glitches or incident radiation

If sufficiently large substrate current flows, V_{BE} of NPN device increases, and its collector current grows.

This increases the current through R_{WELL}. V_{BE} of PNP device increases, further increasing substrate current.


Latch-up current
I latchup
I_1
Transition from high to low resistance.
Voltage
$V_{latchup}$

If bipolar transistors satisfy $\beta_{PNP} \times \beta_{NPN} > 1$, latchup may occur.

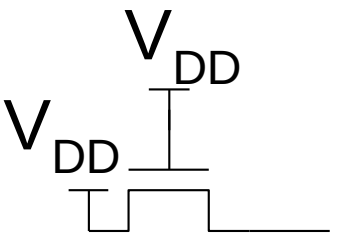
Operation voltage of CMOS circuits should be below $V_{latchup}$.

Remedies of latchup problem:

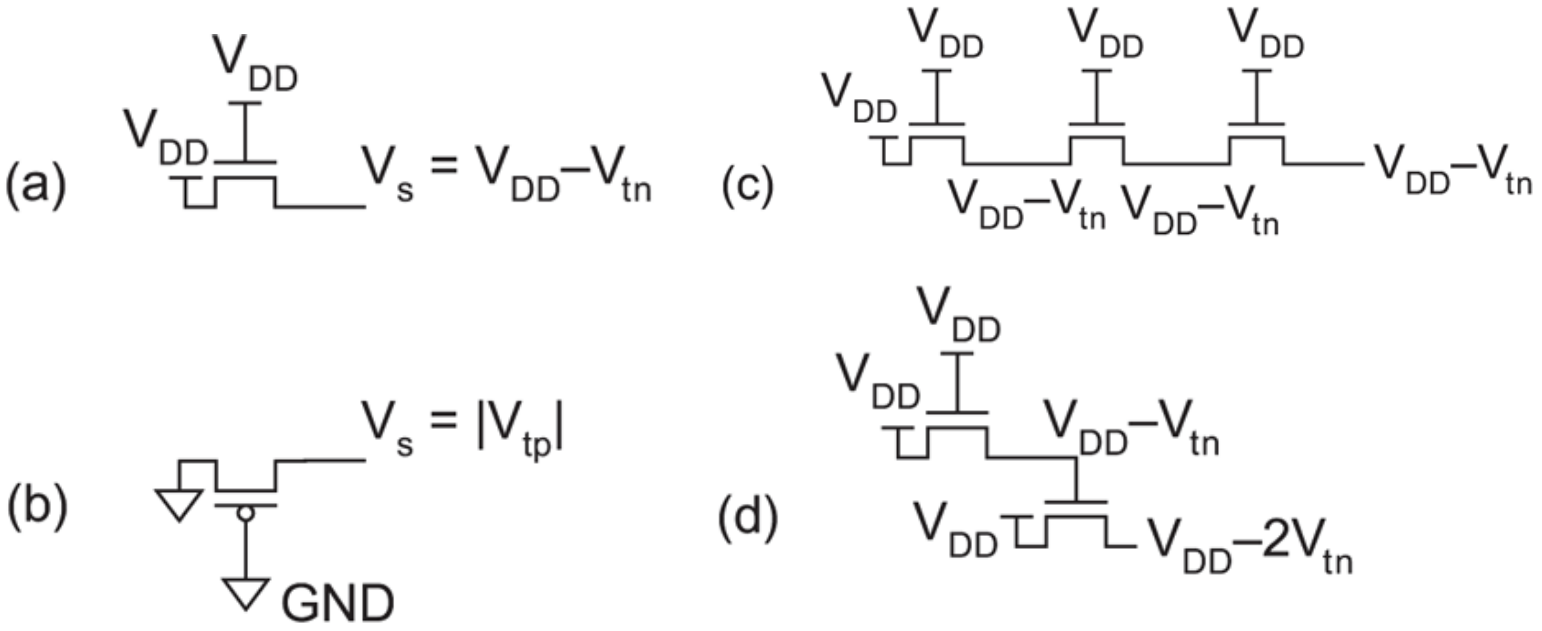
1. Reduce $R_{substrate}$ by increasing P doping of substrate by process control.
2. Reducing R_{WELL} and resistance of WELL contacts by process control.
3. Layout techniques: separation of P and N devices, guard rings, many WELL contacts (at design).

Pass Transistors

- We have assumed source is grounded
- What if source > 0?
  - e.g. pass transistor passing V_{DD}
- $V_g = V_{DD}$
  - If $V_s > V_{DD}-V_t \Rightarrow V_{gs} < V_t$
  - Hence transistor would turn itself off
- nMOS pass transistors pull no higher than $V_{DD}-V_{tn}$
  - Called a degraded “1”
  - Approach degraded value slowly (low I_{ds})
- pMOS pass transistors pull no lower than V_{tp}



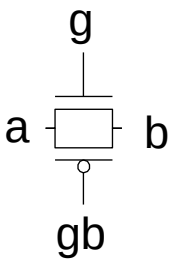
Pass Transistor CKTs

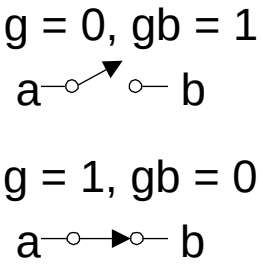


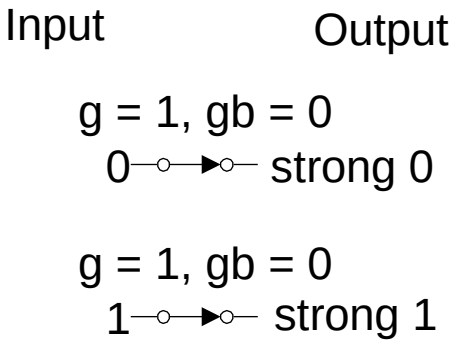
As the source can rise to within a threshold voltage of the gate, the output of several transistors in series is no more degraded than that of a single transistor.

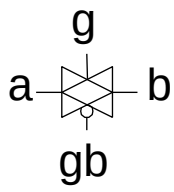
Transmission Gates

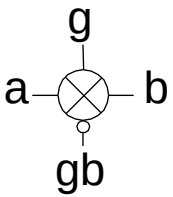
- Single pass transistors produce degraded outputs
- *Complementary Transmission gates* pass both 0 and 1 well

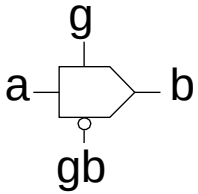






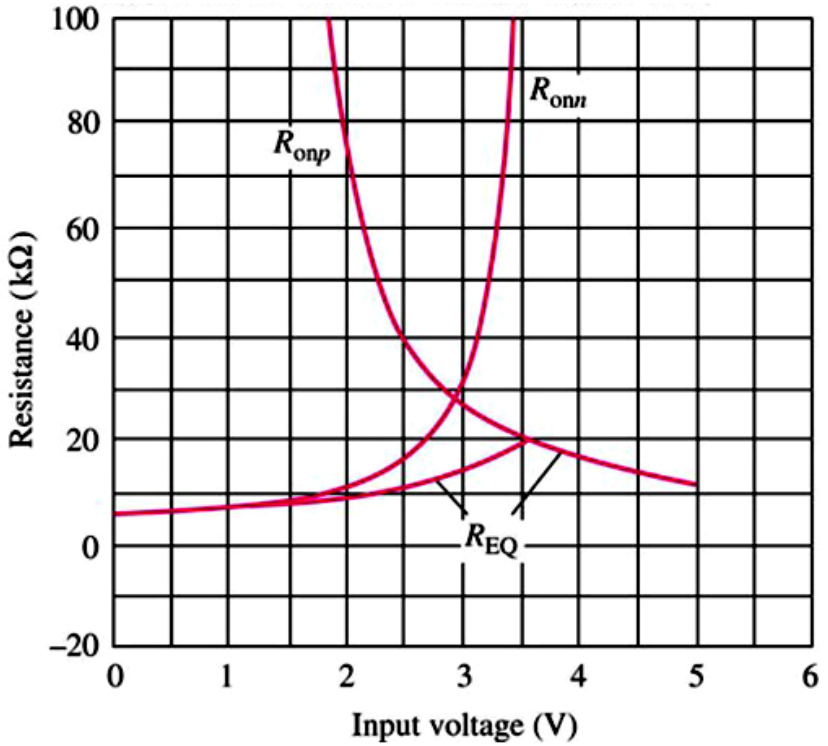






Transmission gate ON resistance as input voltage sweeps from 0 to 1(V_{SS} to V_{DD}), assuming that output follows closely.

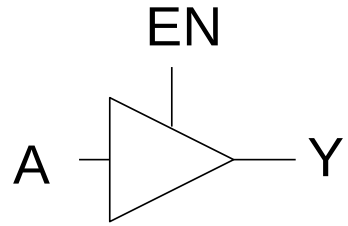
$$R_{EQ} = \frac{R_{onp} R_{onn}}{R_{onp} + R_{onn}}$$

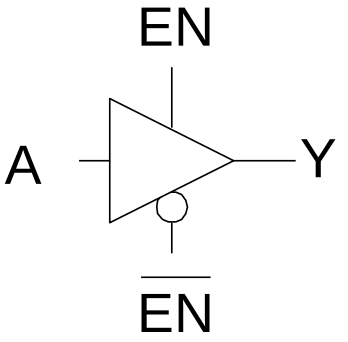


Tristates

- *Tristate buffer* produces Z when not enabled

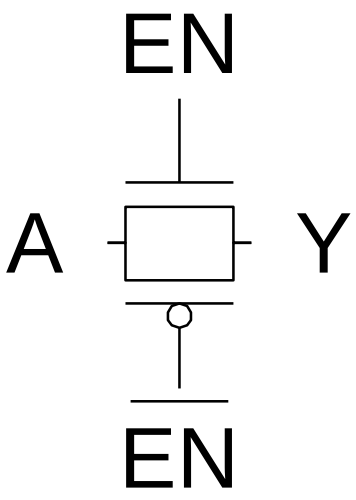
EN	A	Y
0	0	Z
0	1	Z
1	0	0
1	1	1





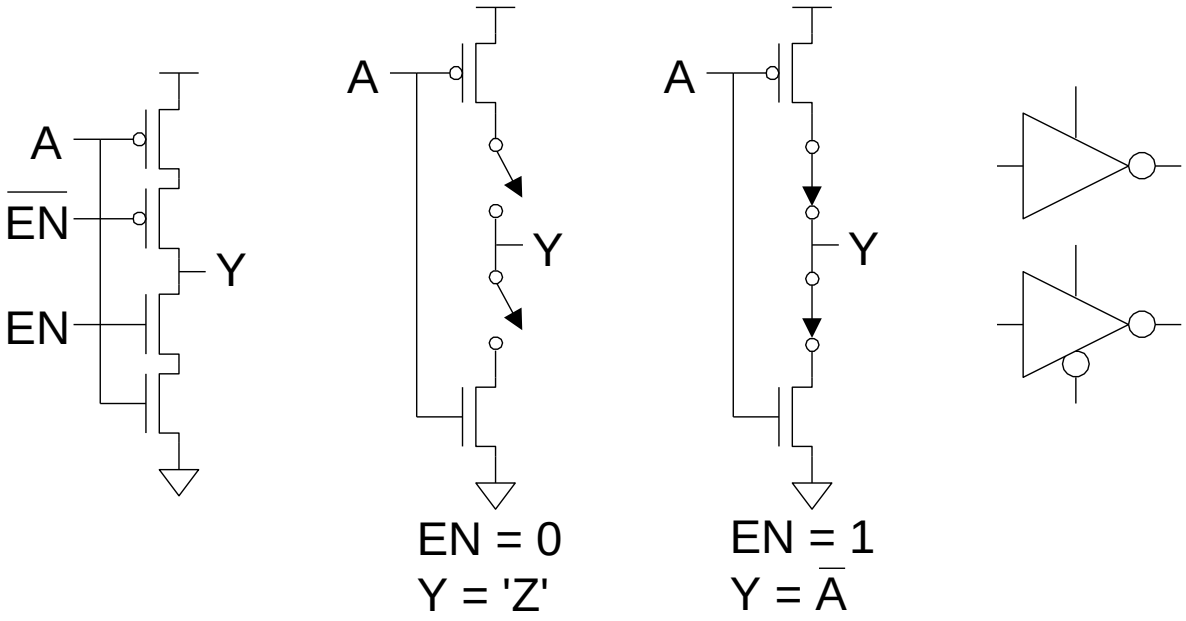
Nonrestoring Tristate

- Transmission gate acts as tristate buffer
  - Only two transistors
  - But *nonrestoring*
    - Noise on A is passed on to Y



Tristate Inverter

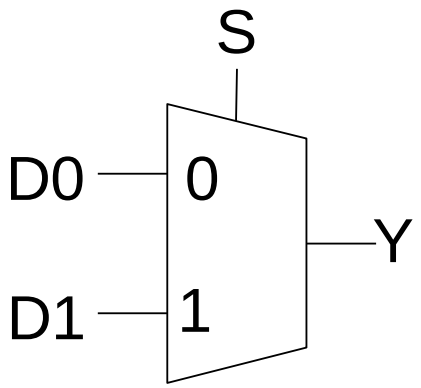
- Tristate inverter produces restored output
  - Violates conduction complement rule
  - Because we want a Z output



Multiplexers

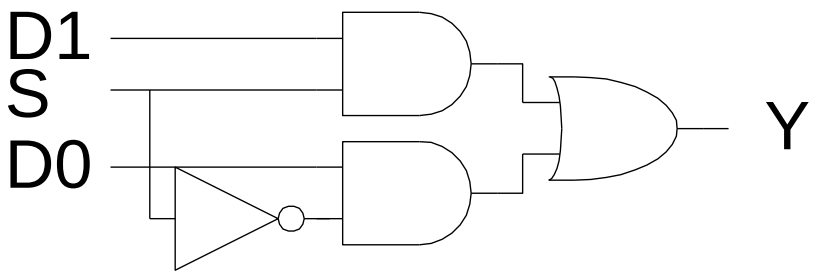
- 2:1 multiplexer chooses between two inputs

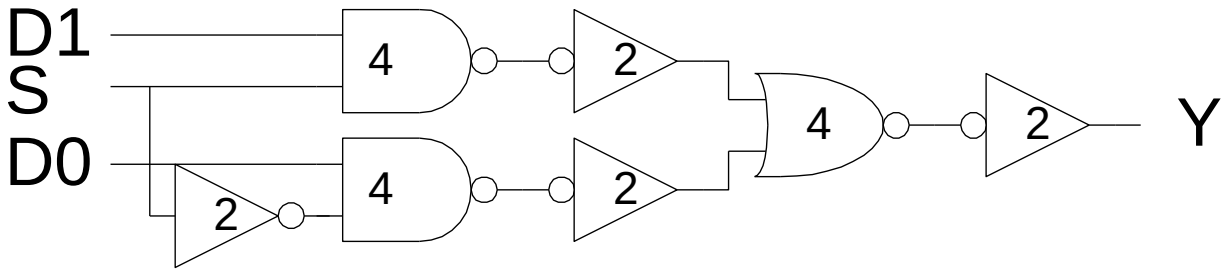
S	D1	D0	Y
0	X	0	0
0	X	1	1
1	0	X	0
1	1	X	1



Gate-Level Mux Design

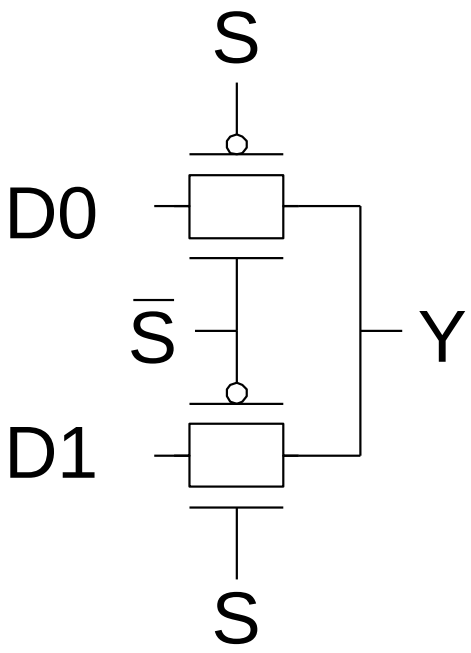
- $Y = SD_1 + \overline{S}D_0$ (too many transistors)
- How many transistors are needed? 20





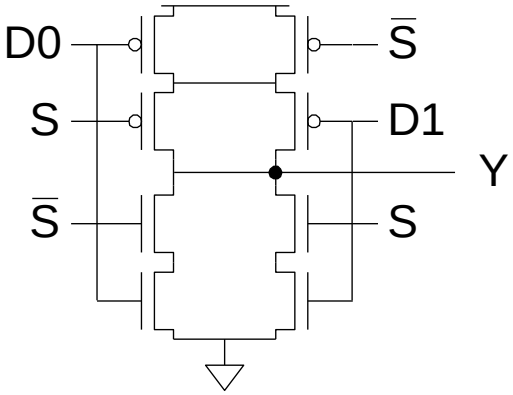
Transmission Gate Mux

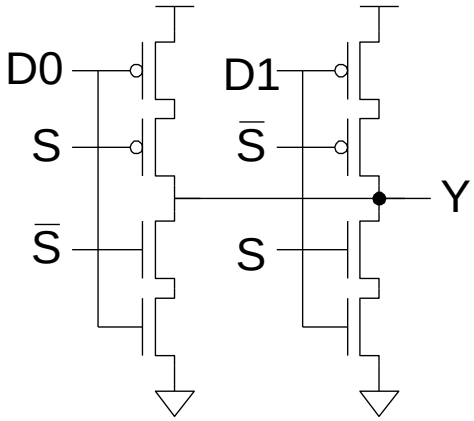
- Nonrestoring mux uses two transmission gates
  - Only 4 transistors

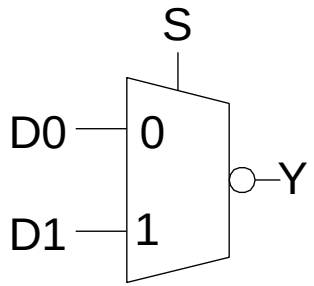


Inverting Mux

- Inverting multiplexer
  - Use compound AOI22
  - Or pair of tristate inverters
  - Essentially the same thing
- Noninverting multiplexer adds an inverter

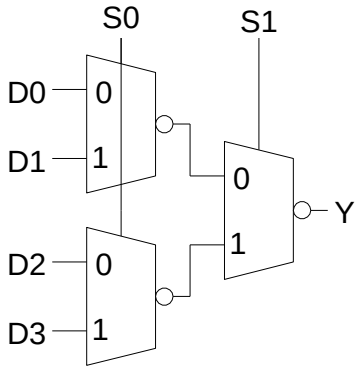


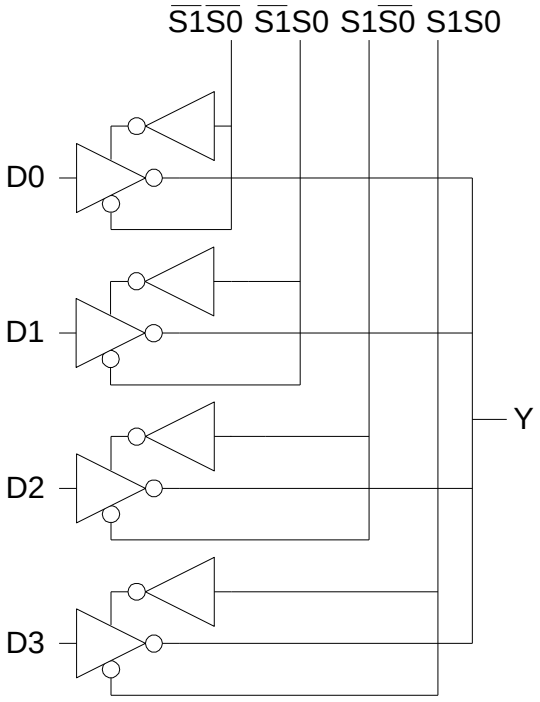




4:1 Multiplexer

- 4:1 mux chooses one of 4 inputs using two selects
  - Two levels of 2:1 muxes
  - Or four tristates





Sizing for Performance

C_{int} NMOS and PMOS diffusion + diffusion-gate overlap.

C_{ext} Fan-out (input gates) + interconnects.

R_{eq} Equivalent gate resistance.

$C_{\text{L}} = C_{\text{int}} + C_{\text{ext}}$ Capacitive load of an inverter.

$C_{\text{int}} = SC_{\text{iref}} \quad R_{\text{eq}} = R_{\text{ref}} / S$ S sizing factor.

Propagation delay: $t_{\text{p}} = 0.69 R_{\text{eq}} \left(C_{\text{int}} + C_{\text{ext}} \right) = t_{\text{p}_0} \left(1 + \frac{C_{\text{ext}}}{SC_{\text{int}}} \right)$

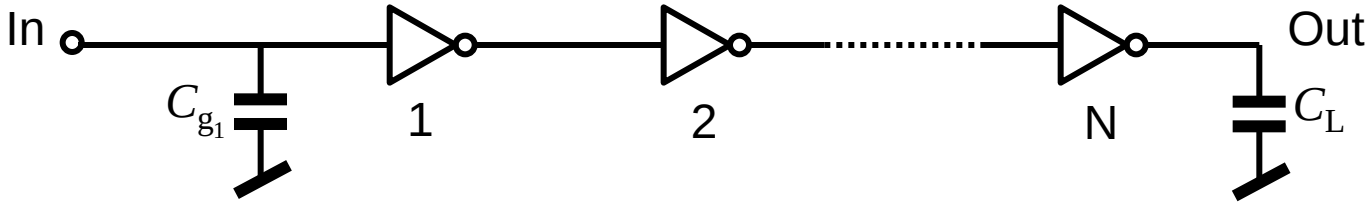
$t_{\text{p}_0} = 0.69 R_{\text{eq}} C_{\text{int}}$ Inverter delay loaded only by intrinsic.

$C_{\text{int}} = \gamma C_{\text{g}}$ Intrinsic cap to gate cap ratio ≈1.

$f = C_{\text{ext}} / C_{\text{g}}$ Effective fan-out.

$$t_{\text{p}} = t_{\text{p}_0}\left(1 + \frac{C_{\text{ext}}}{\gamma C_{\text{g}}}\right) = t_{\text{p}_0}\left(1 + \frac{f}{\gamma}\right)$$

The delay of an inverter is only a function of the ratio between its external load cap to its input cap



$$t_{\text{p}} = \sum_{j=1}^{N} t_{\text{p}_j} = \sum_{j=1}^{N} t_{\text{p}_0}\left(1 + \frac{C_{\text{g}_{j+1}}}{\gamma C_{\text{g}_j}}\right),\ C_{\text{g}_{N+1}} = C_{\text{L}}$$

$$\frac{\partial t_p}{\partial C_{g_j}} = 0,\ 1 \le j \le N-1 \quad \text{imply} \quad \frac{C_{g_{j+1}}}{C_{g_j}} = \frac{C_{g_j}}{C_{g_{j-1}}} = f,\ 2 \le j \le N-1$$

It implies that same sizing factor ***f*** is used for all stages.

The optimal size of an inverter is the geometric mean of its neighbor drives

$$C_{g_j} = \sqrt{C_{g_{j+1}} C_{g_{j-1}}}$$

Given C_{g_1} and C_L , and $F = C_L / C_{g_1}$ the optimal sizing factor is

$$f = \sqrt[N]{F}$$

The minimum delay through the chain is

$$t_p = N t_{p_0} \left(1 + \frac{\sqrt[N]{F}}{\gamma} \right)$$

What should be the optimal *N* ?

The derivative by *N* of t_p yields $\gamma + \sqrt[N]{F} - \frac{\sqrt[N]{F}\ln F}{N} = 0$

or equivalently $f = e^{(1+\gamma/f)}$ having a closed form solution

$f = e$ only for γ=0, a case where the intrinsic self load is ignored and only the fan-out is considered.