

Prof: Abishek Srivastava , CVEST

23 lectures

- 30 hours

Grading :

Assignment (3/4) : 10 %.

Course Project : 20 %.

Quiz : 10 % + 10 %.

Midsem : 20 %.

Endsem : 30 %.

## VLSI Design

Verilog Textbooks : Brown's , J.Bhaskar,  
Samir Patnikan .

Assignment Deadline : 6pm

# VLSI Design

→ Topics:-

1) Intro to VLSI Design

2) CMOS Inverter

3) Multi-stage logic Design and Optimization

4) Other logic styles

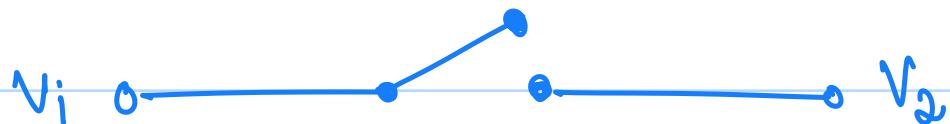
5) Intro. to HDL System Design.

• SCL - Semiconductor Complex Limited. Only fabrication facility in India. Can do upto 180 nm CMOS. Located in Mohali.  
↳ Considered a very old technology.

• We will be using 180 nm technology in this course, since it is now open-source.

→ High impedance state :-

• When a node is not connected to any well defined voltage, it is said to be existing in a high impedance state.

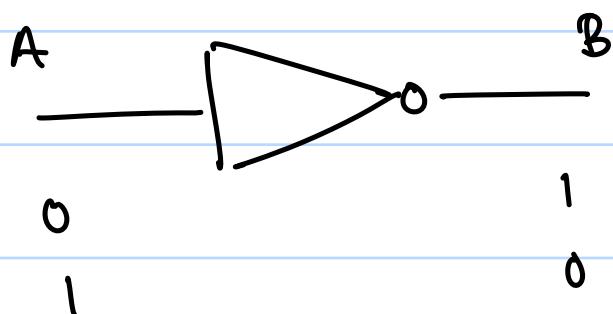


$V_2$  is in a high impedance state.

(Refer Weste & Harris MOS T Physics for  
today's lecture)

5/8/25

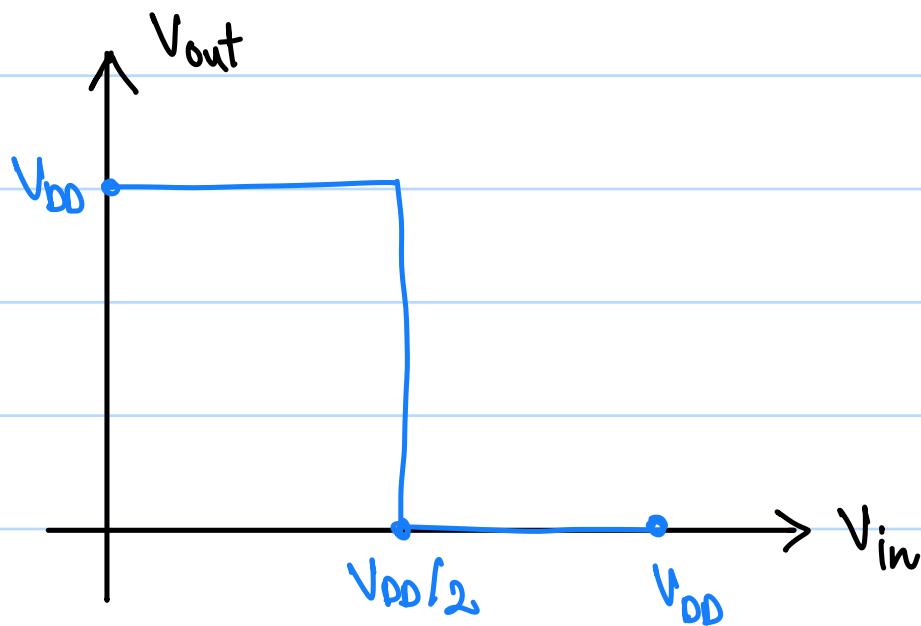
→ Inverter :-



A	B
0	1
1	0

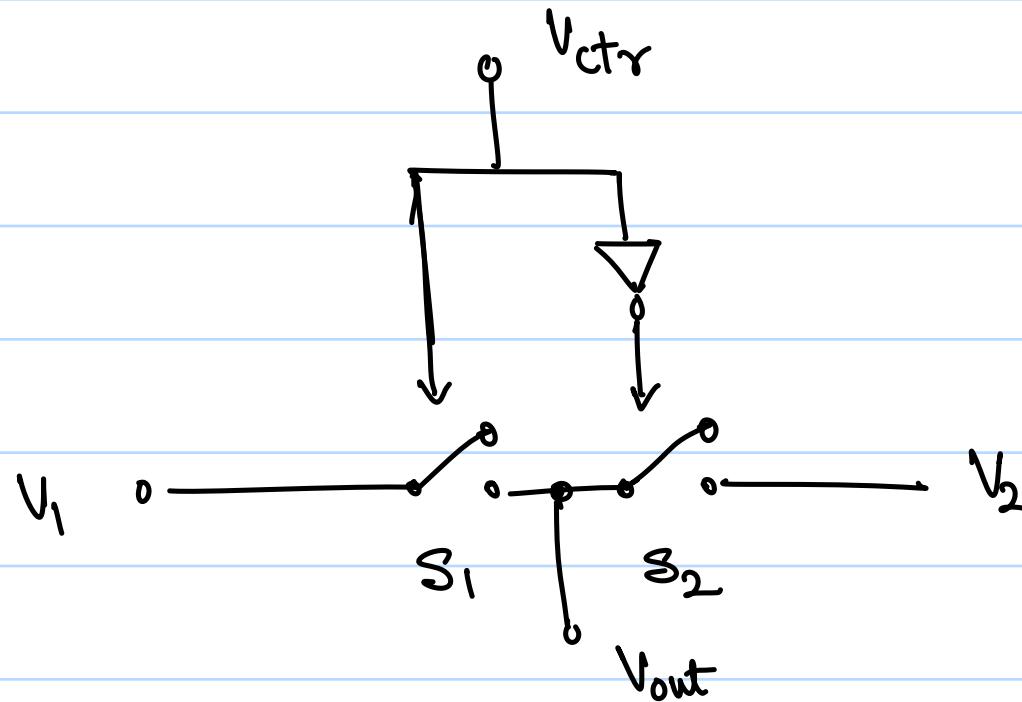
- 1 and 0 correspond to 2 different voltage levels.

• Ideally,



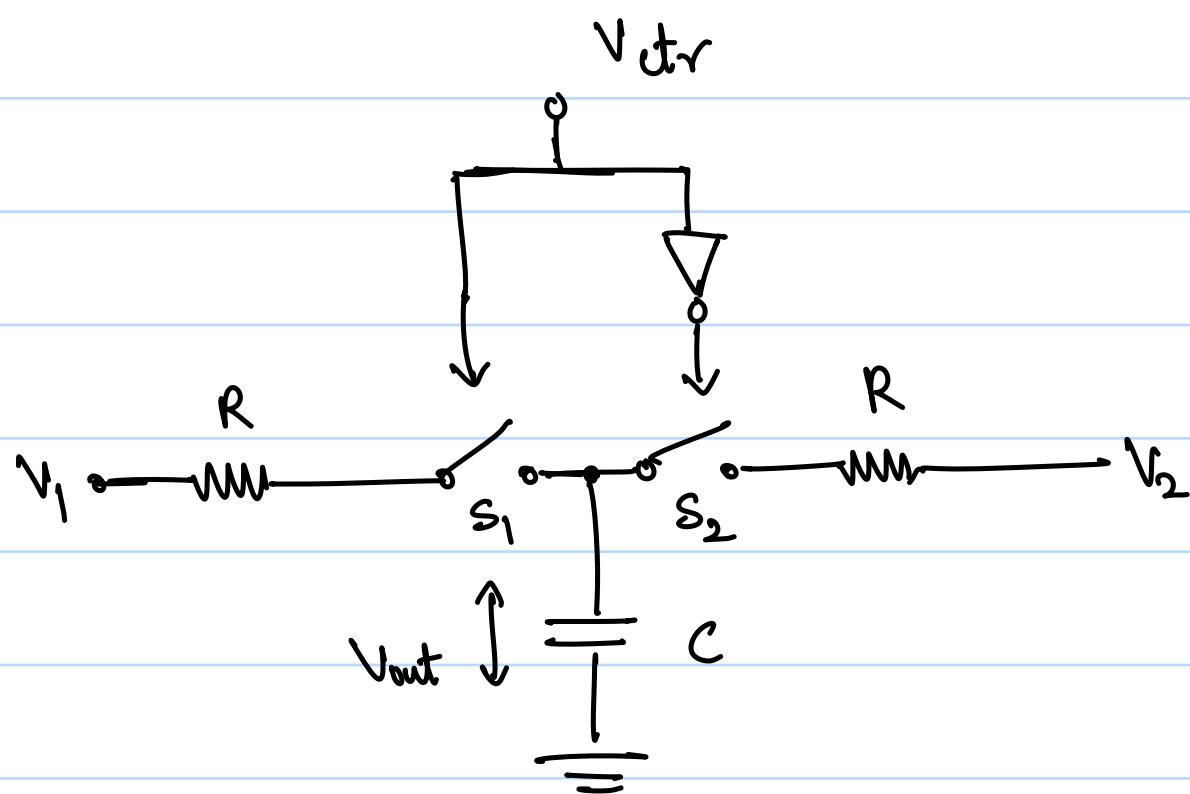
The above plot is termed as a VTC plot (Voltage Transfer Characteristics).

1)

Truth Table:

$V_{ctr}$	$V_{out}$
0	$V_2$
1	$V_1$

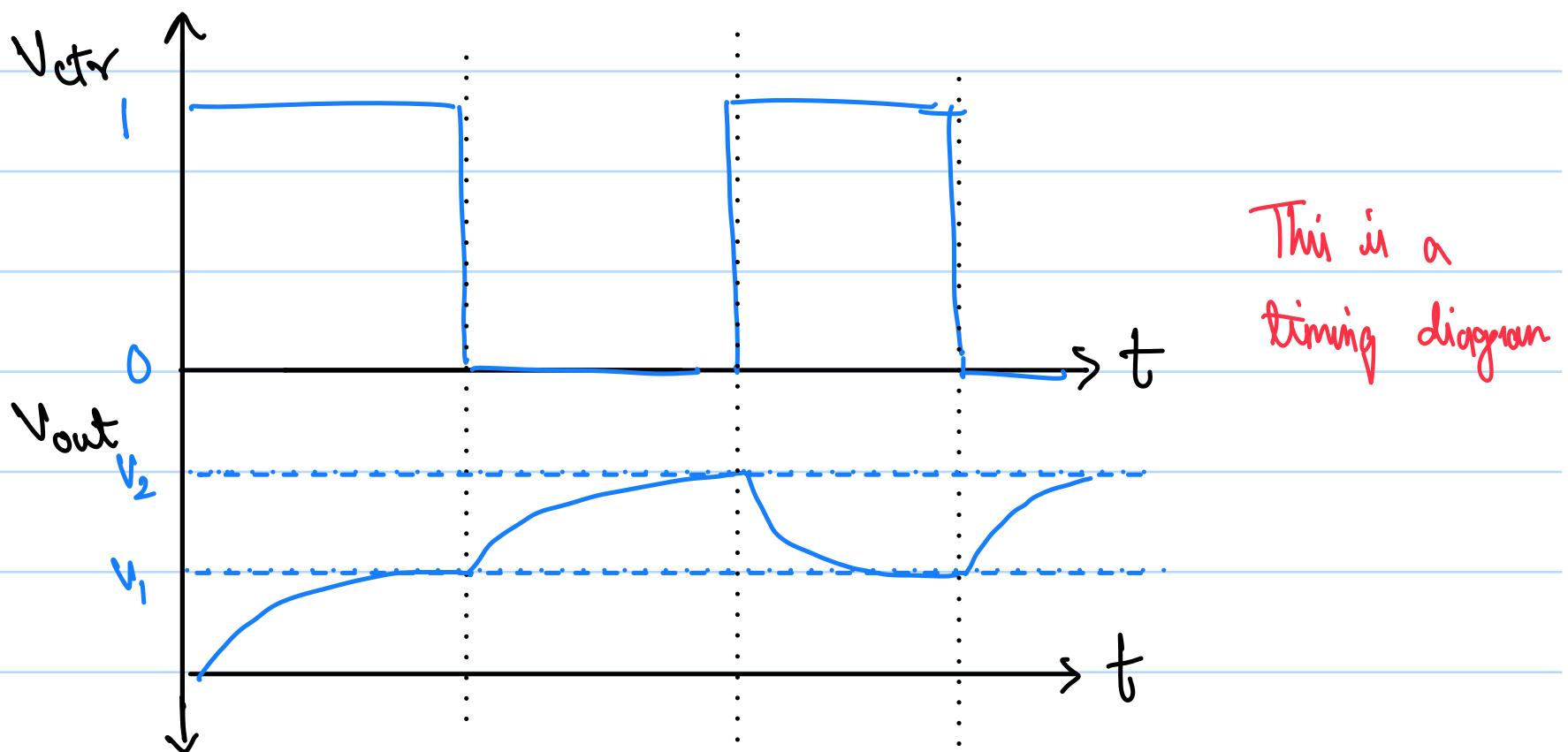
2)



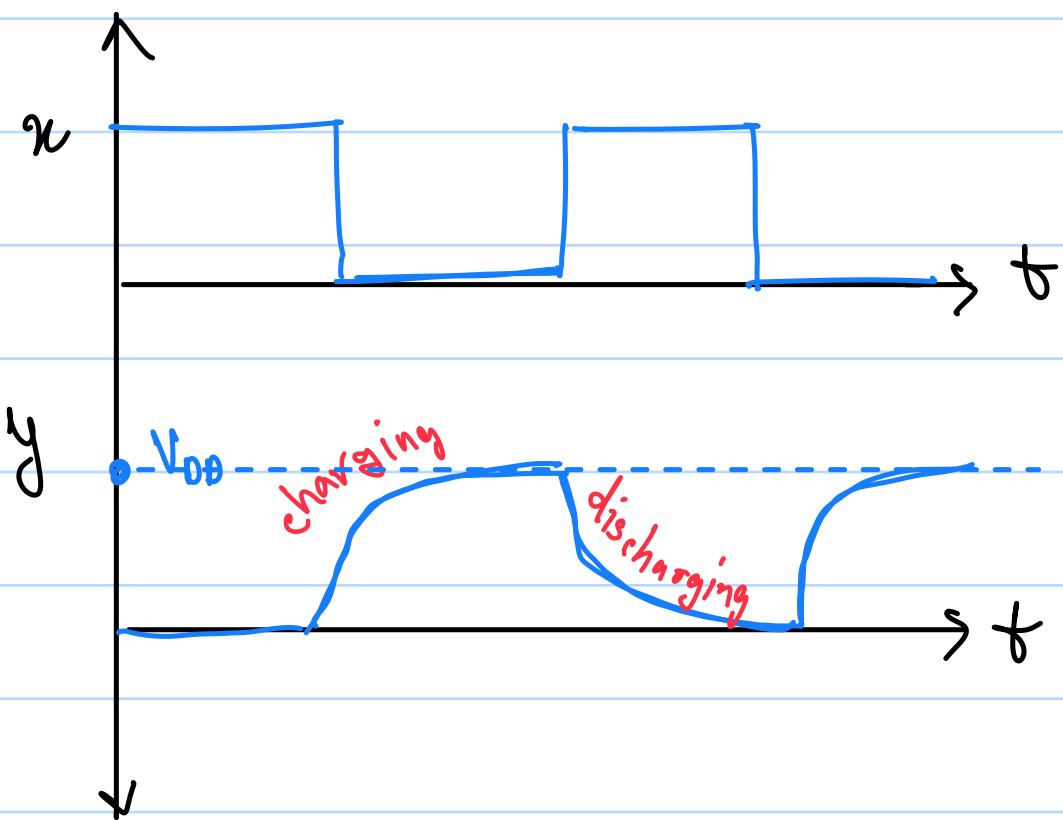
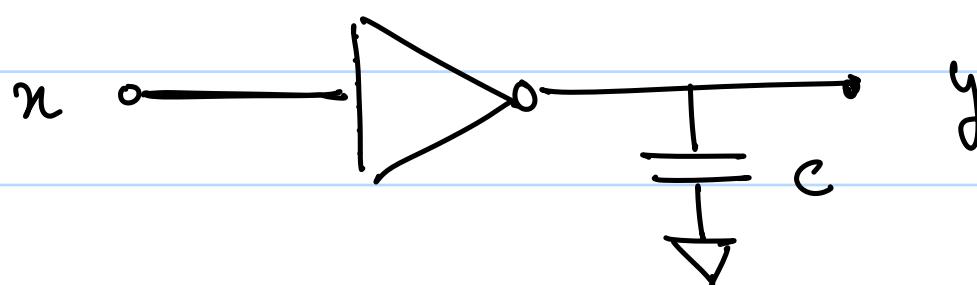
Here,  $V_{out} = \begin{cases} V_2(1 - e^{-t/RC}), & V_{ctr} = 0 \\ V_1(1 - e^{-t/RC}), & V_{ctr} = 1 \end{cases}$

If there is no resistance involved, the current through the capacitor will be a Dirac impulse, infinite for an infinitesimal time.

( $i = C \frac{dV}{dt}$ ,  $dt$  is very small)

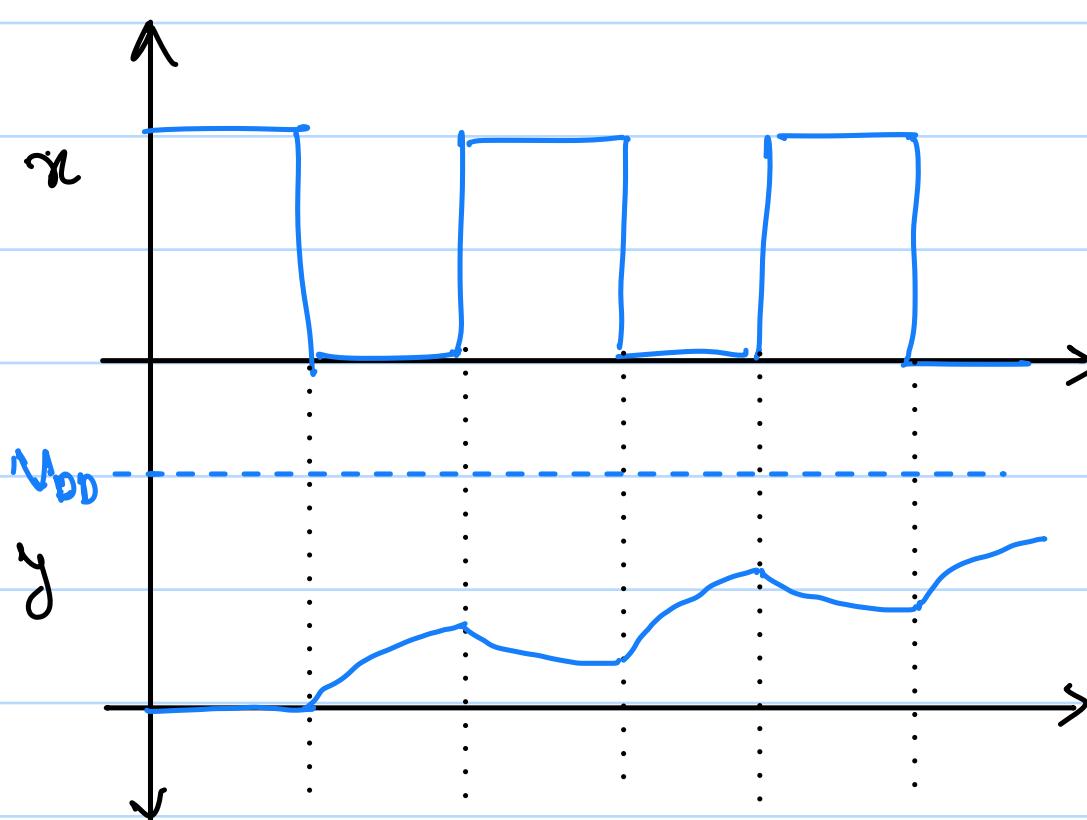


3)



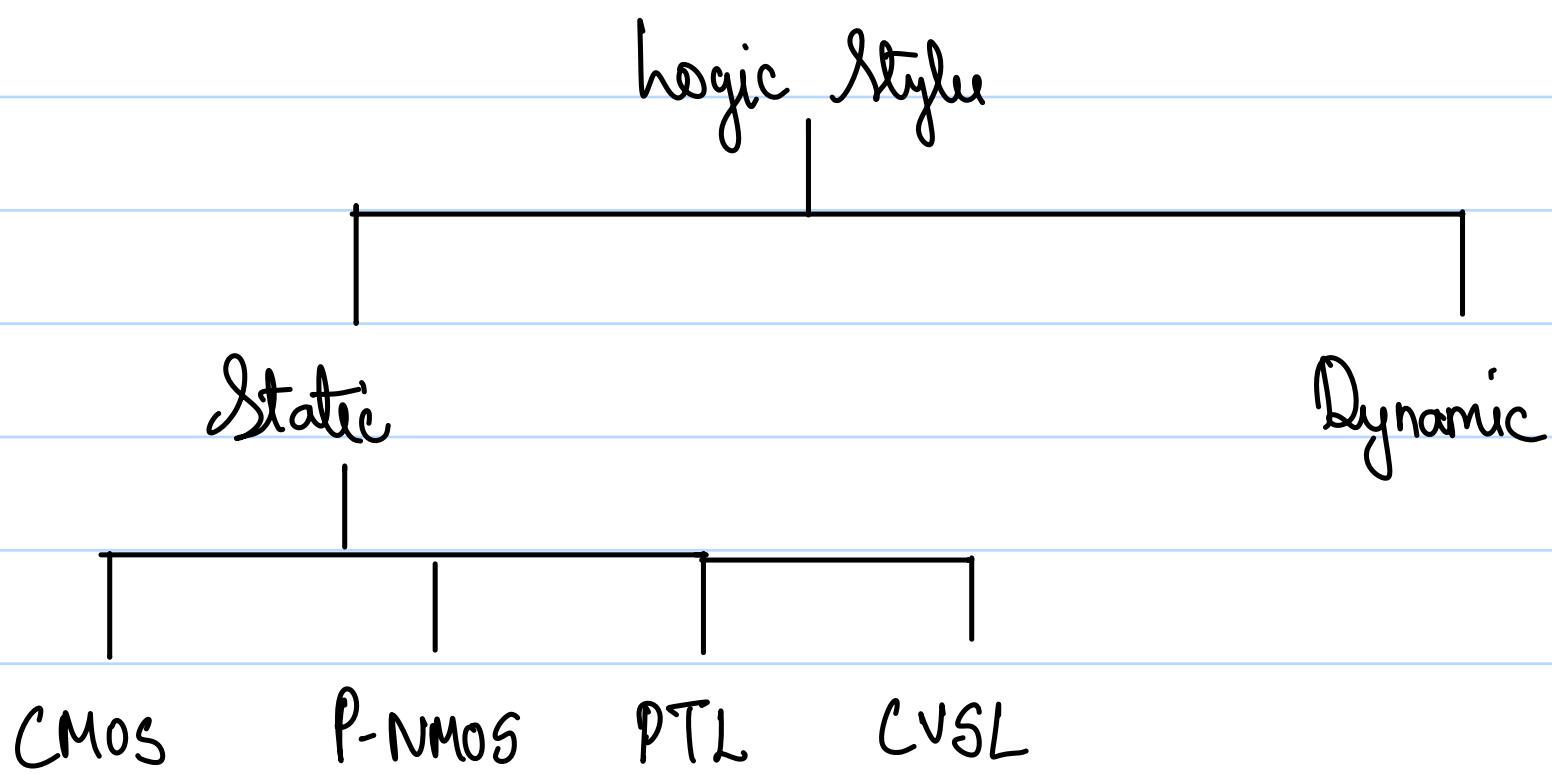
- The charging and discharging speeds of the capacitor are dependent on the RC / time constants of the circuit.

If  $RC$  is very large,



The capacitor is not fast enough to reach  $V_{DD}$  and zero, so information is lost.

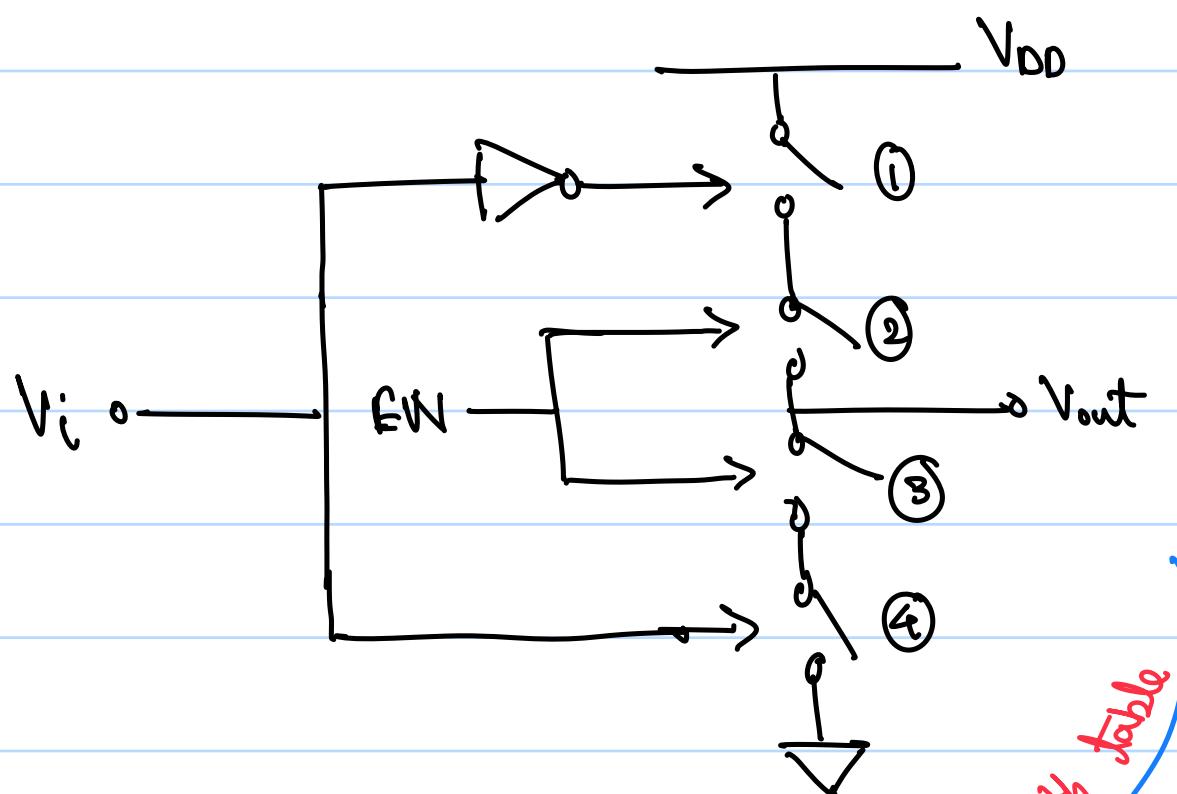
- Because of this, the RC parameters of the circuit must be fast enough to handle the input speed/frequency.



# Circuit Design

Combinational      Sequential      Memory

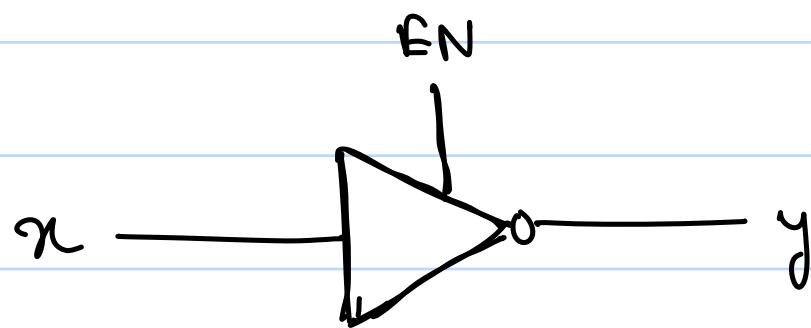
4)



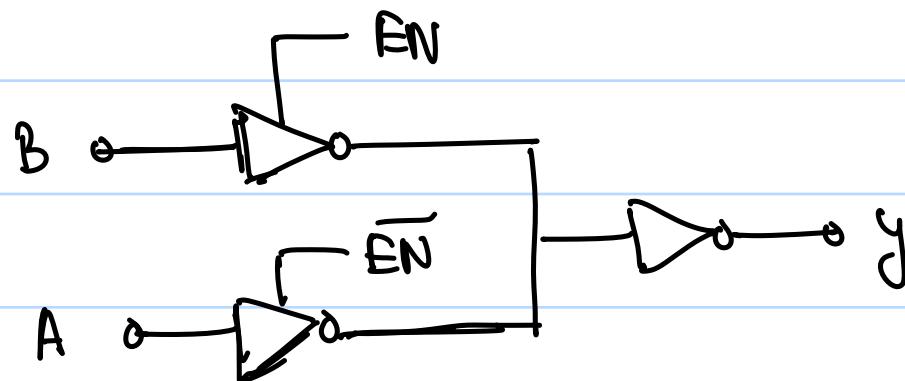
*truth table*

EN	V <sub>i</sub>	V <sub>o</sub>
0	0	Z
0	1	Z
1	0	V <sub>DD</sub>
1	1	0

The above device acts like an inverter if  $EN = 1$ . This is the design of a tri-state inverter.



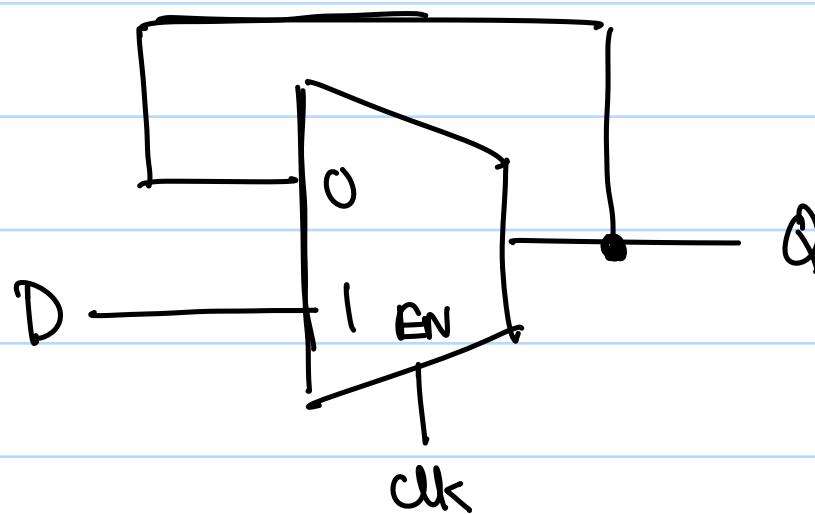
5)



Truth Table:

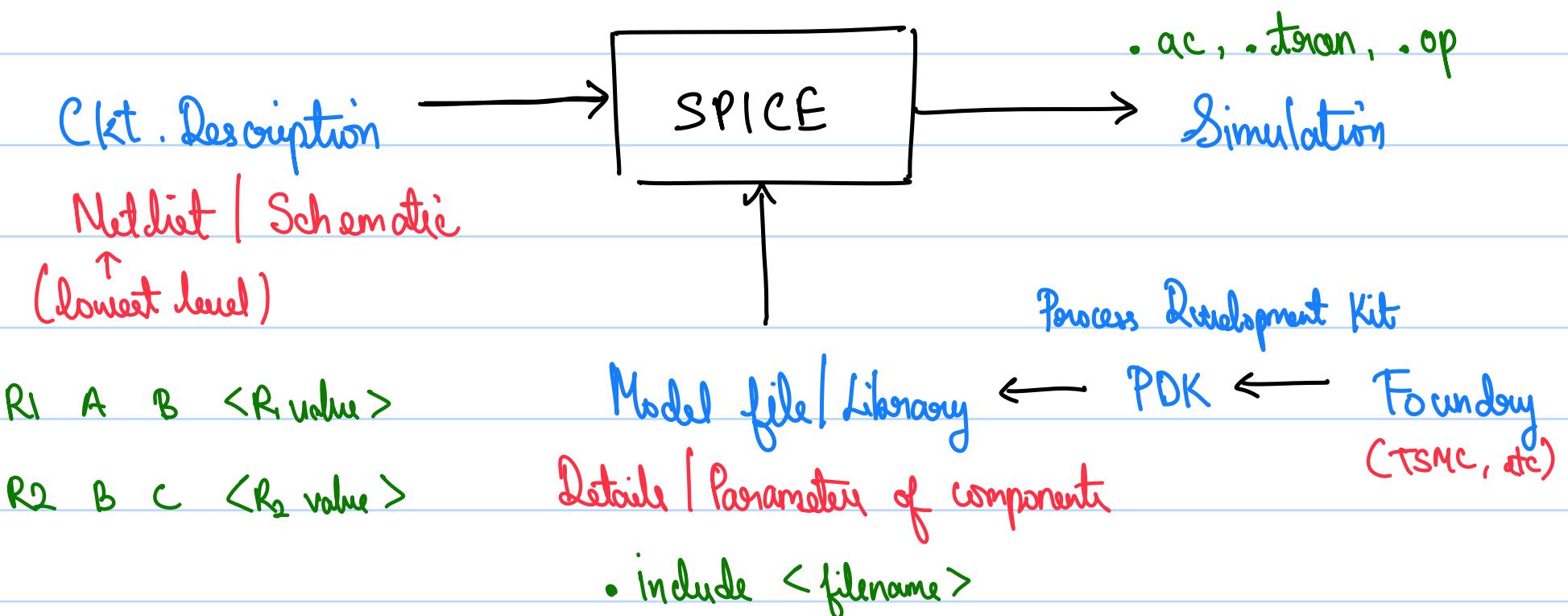
EN	Y
0	A
1	B

- Due to the high impedance, the output of one tri-state inverter will not be affected by the other, since one output is always  $z$ .
- The above device will act as a  $2 \times 1$  MUX, which can be used to design devices like gates and latches.



Level Triggered D-Latch

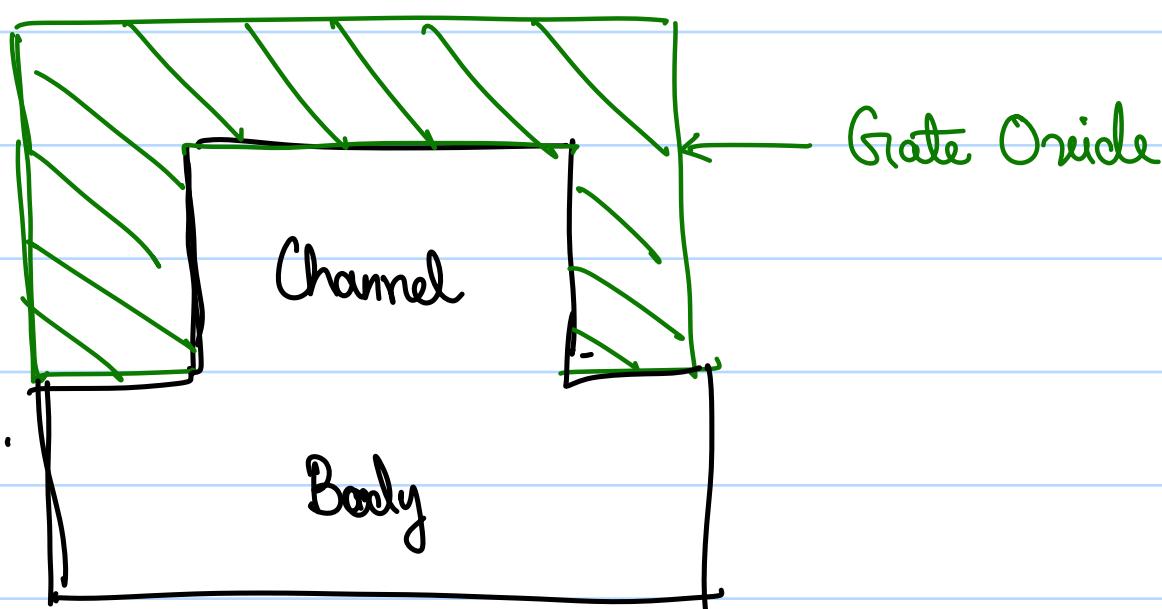
→ SPICE :-



→ Technology Node:

- To make a semiconductor chipset on a substrate, we use a "stencil" of the circuit layout.
- In the chipset, there are multiple metal layers on top of the substrate, to provide the interconnection between the MOSFETs.
- Within the MOSFETs, there are multiple dimensions that vary between them, depending on the usage of that MOSFET
- The minimum of those dimensions, ie, minimum feature length among the MOSFETs is deemed as the "Technology node" of that MOSFET. Usually minimum channel length.
- 3nm (N3) node - Retail products released  
2nm (N2) node - Will release early next year.

- < 22nm - FinFETs are used instead of MOSFETs.



View of FinFET through Source/Drain

In FinFETs, the oxide wraps around the channel, enabling better inversion at a lower gate potential, better than MOSFETs.

- A14 - 1.4nm node } In development.
- A7 - 0.7nm node }

- 180nm is still widely used, in devices that do not require extremely high competition.

- Lower transistor size  $\Rightarrow$  lower cost per transistor.

- Clock Speed:

- Recently the clock speed of processes has saturated.

- Since the packing density has increased by a lot, the heat and leakage of the transistors have increased by a lot.

- These factors made it difficult for clock speed to improve as fast as the technology node.

→ Shannon's Expansion :-

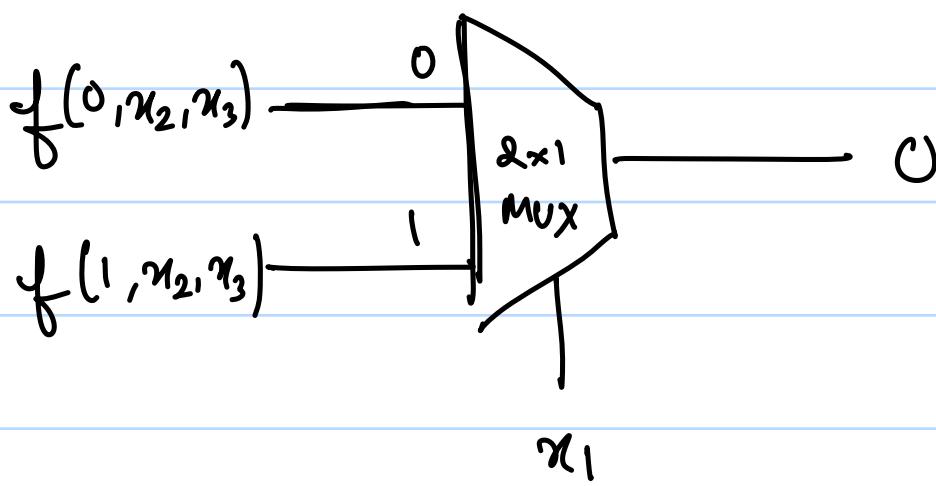
- For any function  $f(x_1, x_2, x_3 \dots x_n)$ ,

$$f(x_1, x_2, x_3 \dots x_n) = x_1 f(1, x_2, x_3 \dots x_n) + \bar{x}_1 f(0, x_2, x_3 \dots x_n)$$

Example: Expand  $f(x_1, x_2, x_3) = \bar{x}_1 \bar{x}_3 + x_1 x_2 + x_1 x_3$ . w.r.t  $x_1$

$$\Rightarrow f(x_1, x_2, x_3) = x_1 (x_2 + x_3) + \bar{x}_1 (\bar{x}_3)$$

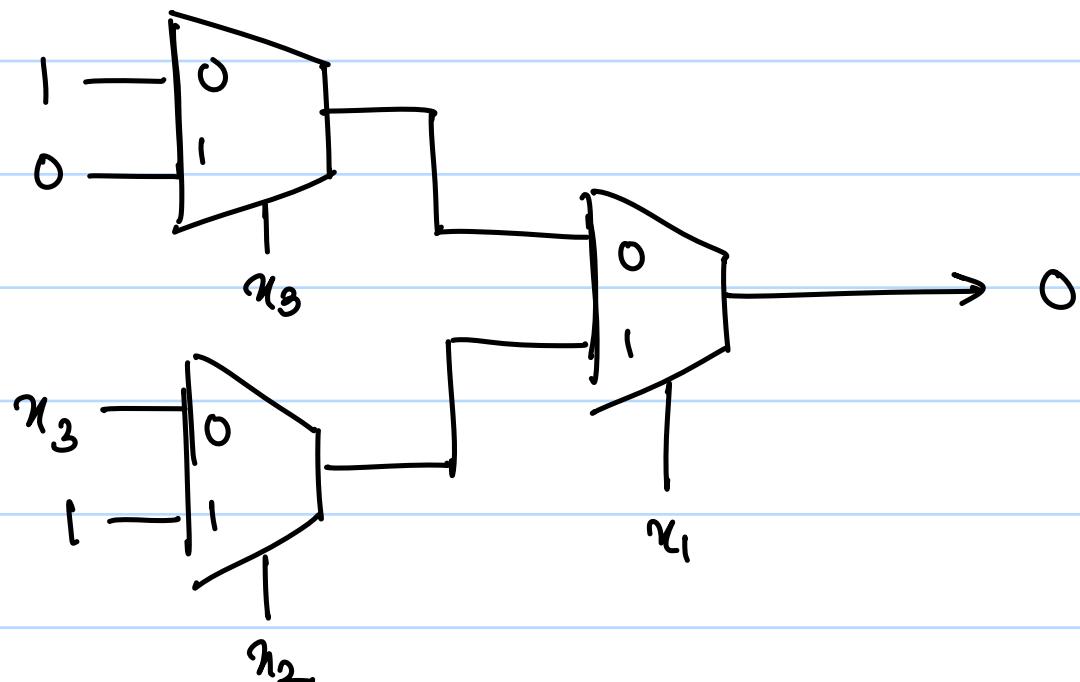
- If we expand the function in this way, we can use a  $2 \times 1$  MUX to implement the function, using  $x_1$  as the selection variable.



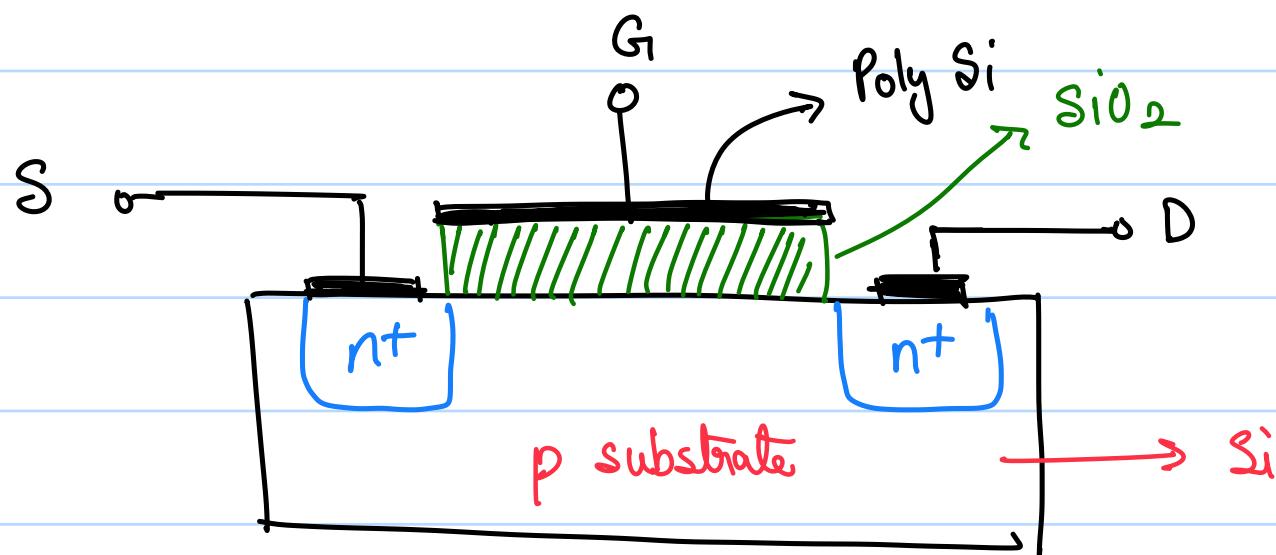
Example: Express the prev. function only using 2x1 MUXes.

$$f(x_1, x_2, x_3) = \bar{x}_1 \bar{x}_3 + x_1 x_2 + x_1 x_3 \\ = x_1 (x_2 + x_3) + \bar{x}_1 (\bar{x}_3)$$

$$f(x_2, x_3) = x_2 + x_3 \\ = x_2 (1) + \bar{x}_2 (x_3)$$



→ MOSFETs :-

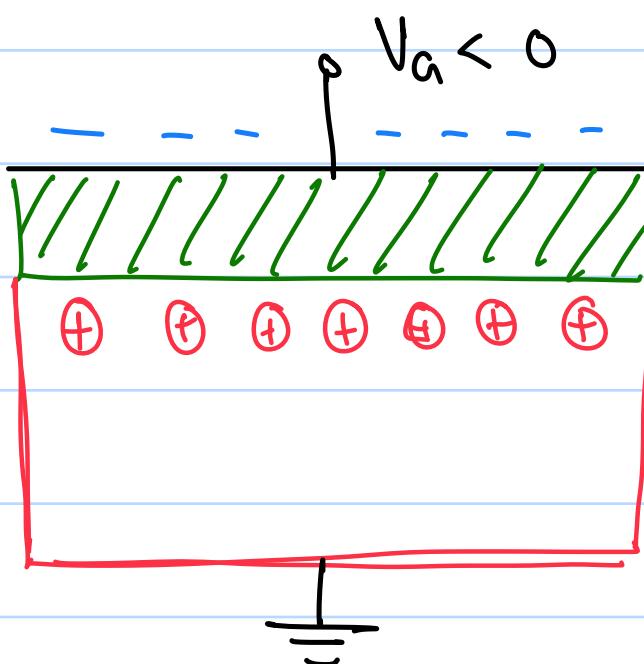


NMOS Structure

- Poly silicon is used in the gate terminal to simplify the manufacturing process, since silicon is the major element of a MOSFET.

- Let S and D be grounded and ,

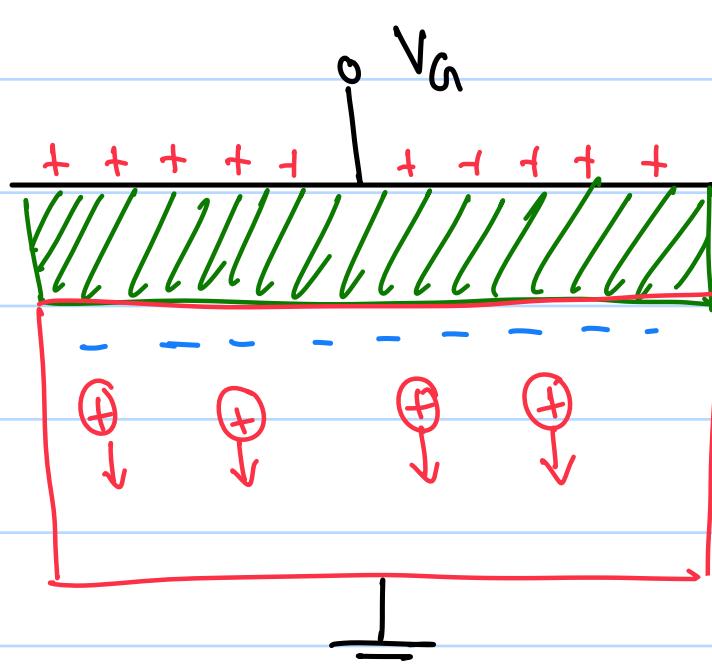
1)  $V_G < 0$



holes are attracted towards the gate

- This is accumulation mode of operation .

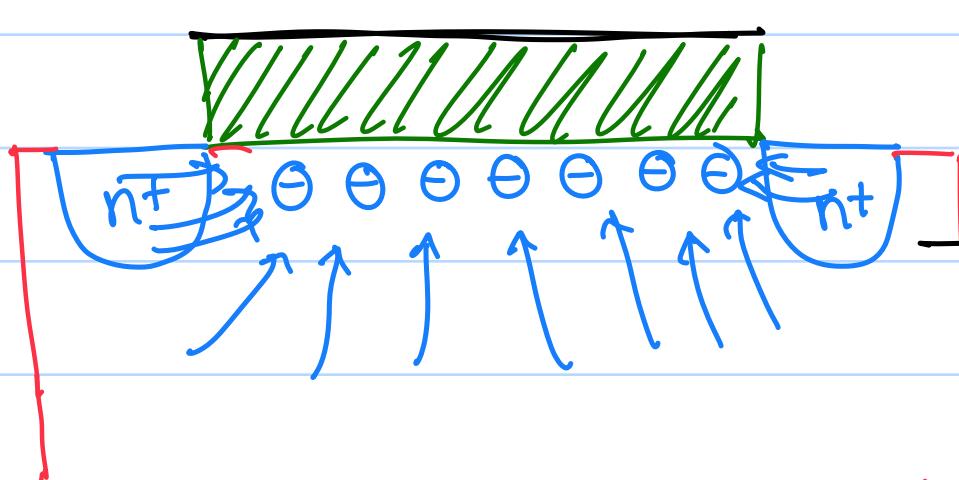
2)  $V_G > 0$



holes go away from gate, leaving -ve charge .

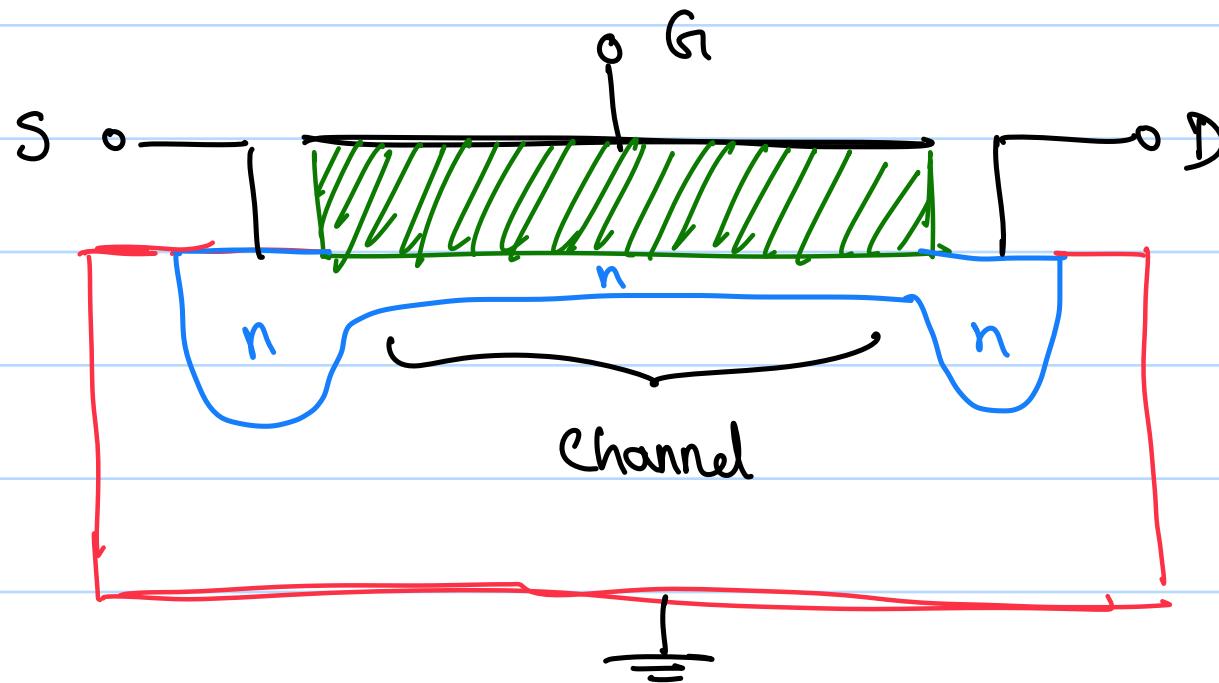
- This is depletion mode of operation .

3)  $V_G \gg 0$



electrons are attracted from the substrate and (majority) from S and D

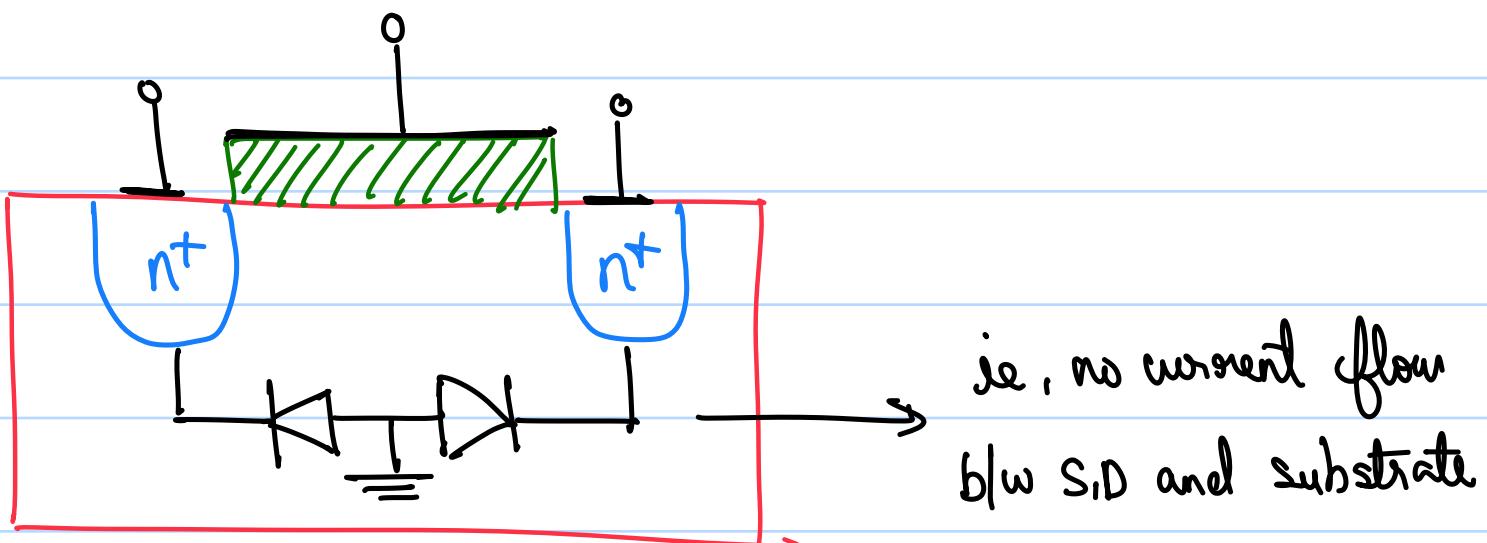
- The surface of the p-substrate becomes dominated by electrons, ie, becomes n-type.



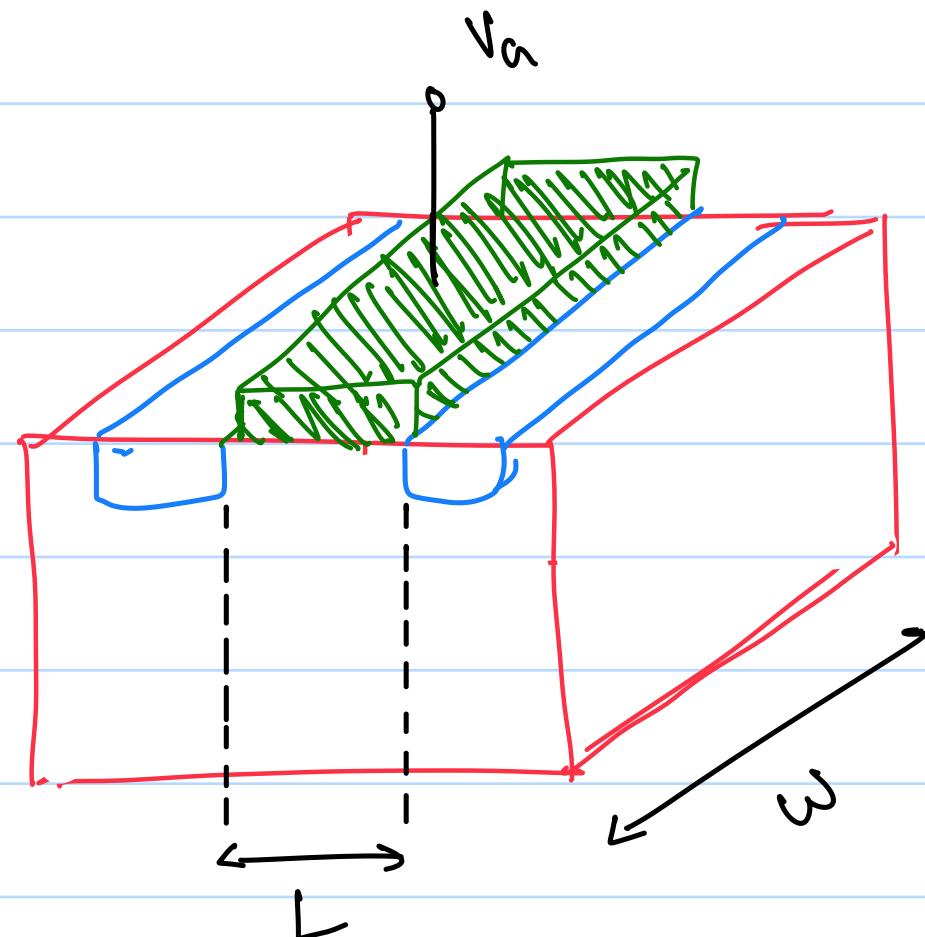
- This is inversion mode of operation.

- If  $n_e$  at surface  $< N_A$  (hole conc. of substrate), it is weak inversion / subthreshold mode.

- If  $n_e$  at surface  $\geq N_A$ , it is strong inversion. The min. gate voltage to attain strong inversion is  $V_{TH}$  (threshold voltage)

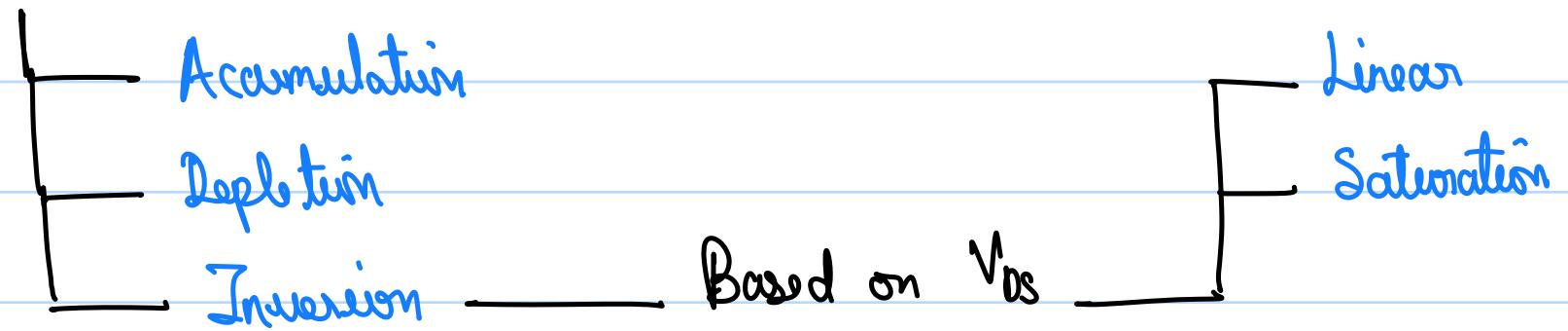


• 3D View:-



Since the n-channel is not perfectly conductive, it has some resistance to it, depending on the dimensions L and w.

Based on  $V_{AS}$



1)  $V_{AS} \leq 0 \Rightarrow I_D = 0 + V_{DS}$ . Accumulation mode has zero channel.

2)  $0 < V_{AS} < V_{TH} \Rightarrow I_D \approx I_{D0} e^{\frac{V_{AS}-V_{TH}}{nVt}}$ .  $V_t$  - thermal voltage.

Subthreshold conduction.

3)  $V_{AS} > V_{TH}$  &  $V_{DS} < V_{AS} - V_{TH} = V_{ov}$ . (Overshoot voltage)

$$I_D = \frac{1}{2} \mu_n C_o \frac{w}{L} \left( 2V_{ov}V_{DS} - \frac{V_{DS}^2}{2} \right) \rightarrow \text{Linear Mode of Opn.}$$

If  $V_{DS} \ll V_{OV}$ ,

$$I_D = \left( \mu_n C_o x \frac{W}{L} V_{OV} \right) V_{DS}$$

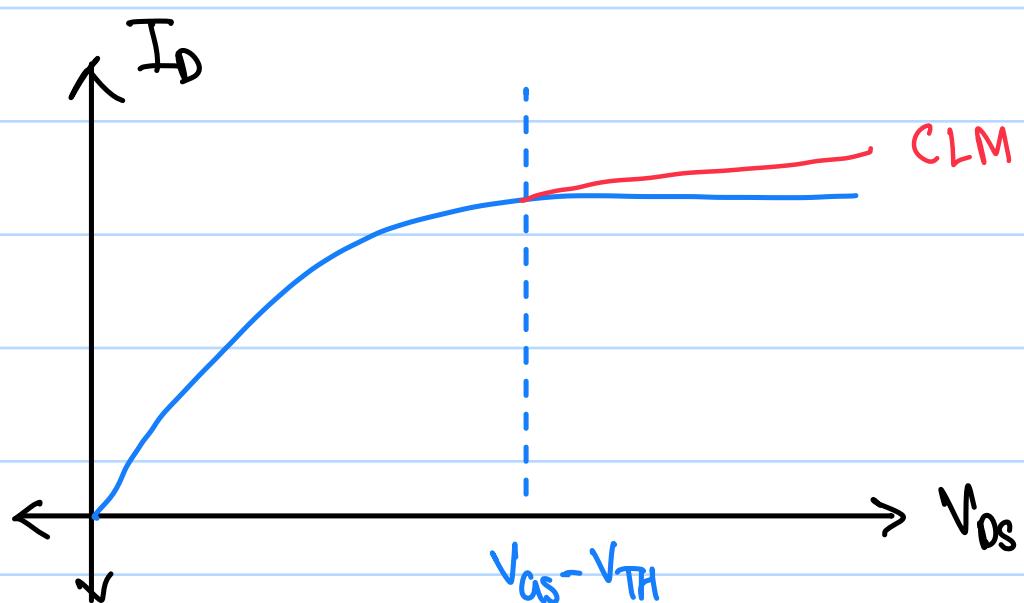
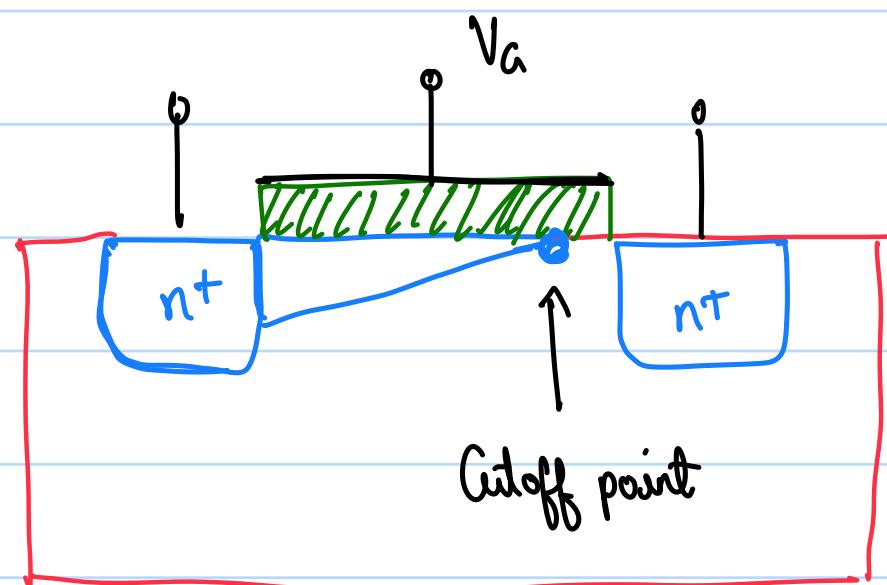
$$\Rightarrow R = \frac{1}{\mu_n C_o x \frac{W}{L} V_{OV}}$$

4)  $V_{GS} > V_{TH}$ ,  $V_{DS} \geq V_{GS} - V_{TH}$

$$\Rightarrow V_{GS} - V_{DS} \leq V_{TH}$$

$$\Rightarrow \underline{V_{DS} \leq V_{TH}}$$

Since  $V_{DS} \leq V_{TH}$ , the channel will cease to exist near the drain, since potential at the point will be lesser than threshold.



Beyond the cutoff point, the charge density is very low (depletion mode). Let linear charge density there be  $Q_d$ .  $Q_d \approx 0$ .

$$dQ = Q_d \cdot dx$$

$$\frac{dQ}{dt} = Q_d \frac{dx}{dt}$$

$$= I = Q_d V_d$$

$$= V_d = \frac{I}{Q_d} \Rightarrow \lim_{Q_d \rightarrow 0} \frac{I}{Q_d}$$

$\therefore V_d \rightarrow \infty$  for finite  $I$ .

$\therefore$  The electrons are swept almost instantaneously across the cutoff point into the drain.

$$I_{DS} = \begin{cases} \mu n C_o \frac{W}{L} (2(V_{GS} - V_{TH})V_{DS} - \frac{V_{DS}^2}{2}), & V_{GS} > V_{TH}, V_{DS} < V_{ov} \\ \frac{1}{2} \mu n C_o \frac{W}{L} (V_{GS} - V_{TH})^2, & V_{GS} > V_{TH}, V_{DS} > V_{ov} \\ 0, & V_{GS} \ll V_{TH} \\ I_{DSS} e^{\frac{V_{GS}-V_T}{nVt}} [1 - e^{\frac{V_{DS}}{4}}], & V_{GS} < V_{TH} \end{cases}$$

(Subthreshold leakage)

• PMOS:

$$I_{SD_{PSI}} = \frac{1}{2} \mu p C_o \frac{W}{L} [V_{SA} - |V_T|]^2$$

$$V_{SA} \geq |V_T| \quad \&$$

$$V_{SD} \geq V_{SA} - |V_T|$$

$$I_{SD\text{plin}} = \mu_p C_{ox} \frac{W}{L} \left[ (V_{SD} - |V_T|) V_{SD} - \frac{V_{SD}^2}{2} \right]$$

• Second-Order Effects :-

1) Channel Length Modulation :-

Since the cutoff point moves further towards the source, as  $V_{DS}$  increases,  $I_D$  actually increases with  $V_{DS}$  even during saturation.

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{DS} - V_{TH})^2 (1 + \lambda V_{DS})$$

$\lambda$  CLM Parameter

$$\lambda V_{DS} = \frac{\Delta L}{L}$$

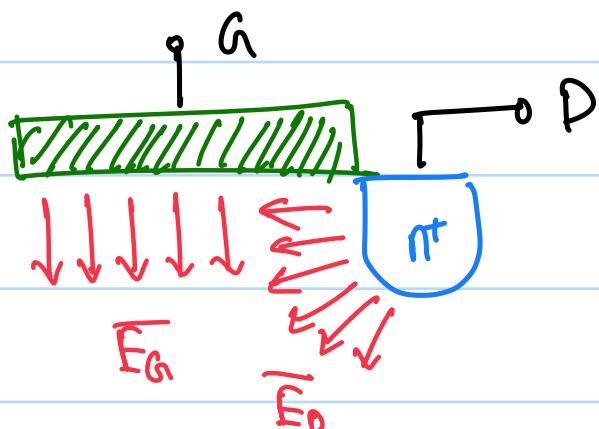
2) Body Biasing :-

If  $V_{BS}$  is increased, the threshold voltage of the MOSFET is decreased.

$$V_{TH} = V_{TH0} + \sqrt{2\psi_s + V_{SB}} - \sqrt{2\psi_s}$$

3) Drain Induced Barrier Lowering :-

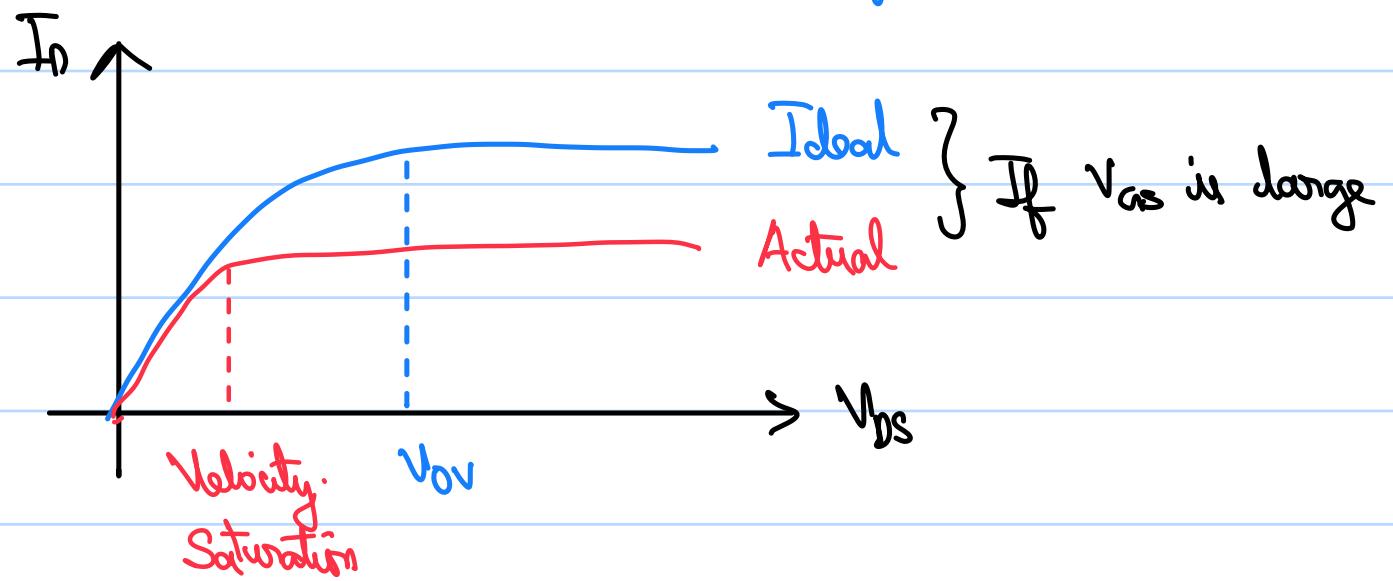
In short channel MOSFETs, the drain voltage will aid in the inversion of the channel.



Because of this aid, the threshold voltage of the MOSFET will be lower.

#### 4) Velocity Saturation :-

For each channel, there is a certain  $V_d$  at which mobility degradation occurs due to scattering of the charge carriers. Prominent if  $V_{as}$  is large.

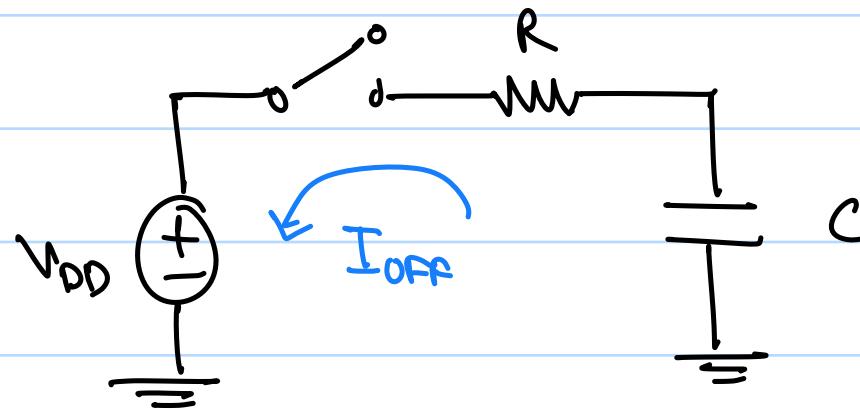


If  $V_{as}$  is low, overdrive voltage is attained before velocity saturation, so the effect is not visible.

#### • Switch $I_{ON}, I_{OFF}$ :-

- $I_{ON}$  : Current through the switch when it is in the ON position.
- $I_{OFF}$  : Current through the switch when it is in the OFF position.
- Ideally  $I_{OFF}$  should be zero, but it is non-zero due to effects like leakage.

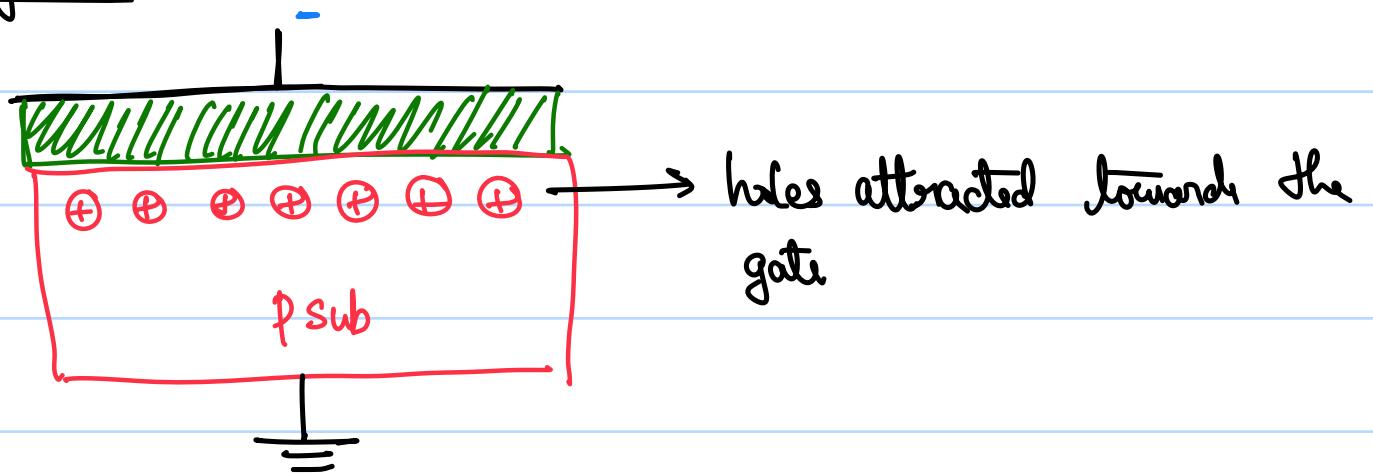
-  $t_{\text{hold}}$  : Time taken for a capacitor to discharge in an RC circuit when the switch is OFF.



$$\text{For a good circuit, } t_{\text{hold}} \geq \frac{1}{2} \frac{C V_{DD}}{I_{\text{off}}}$$

- o MOS Capacitor:

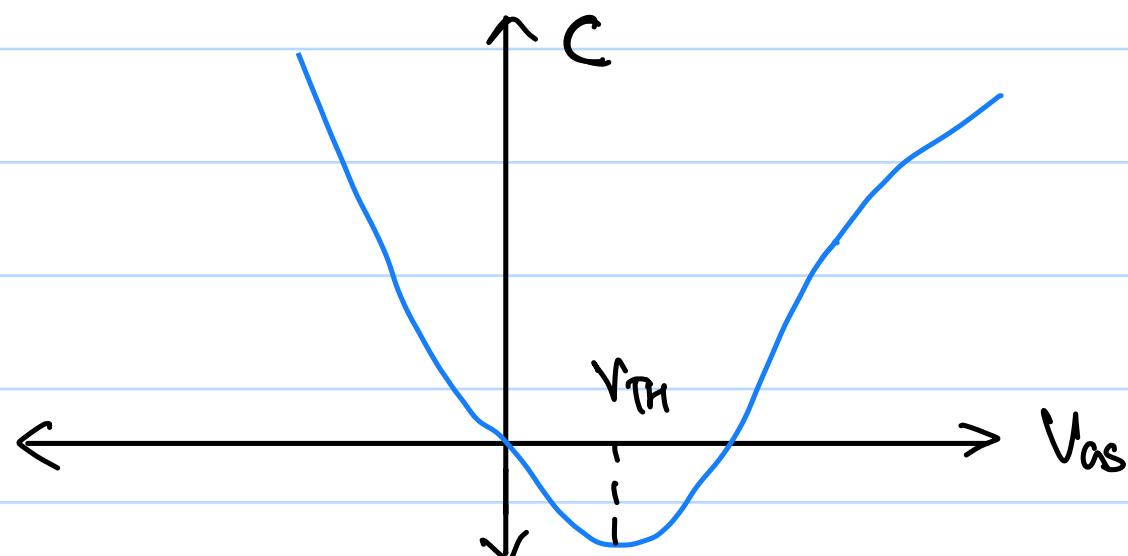
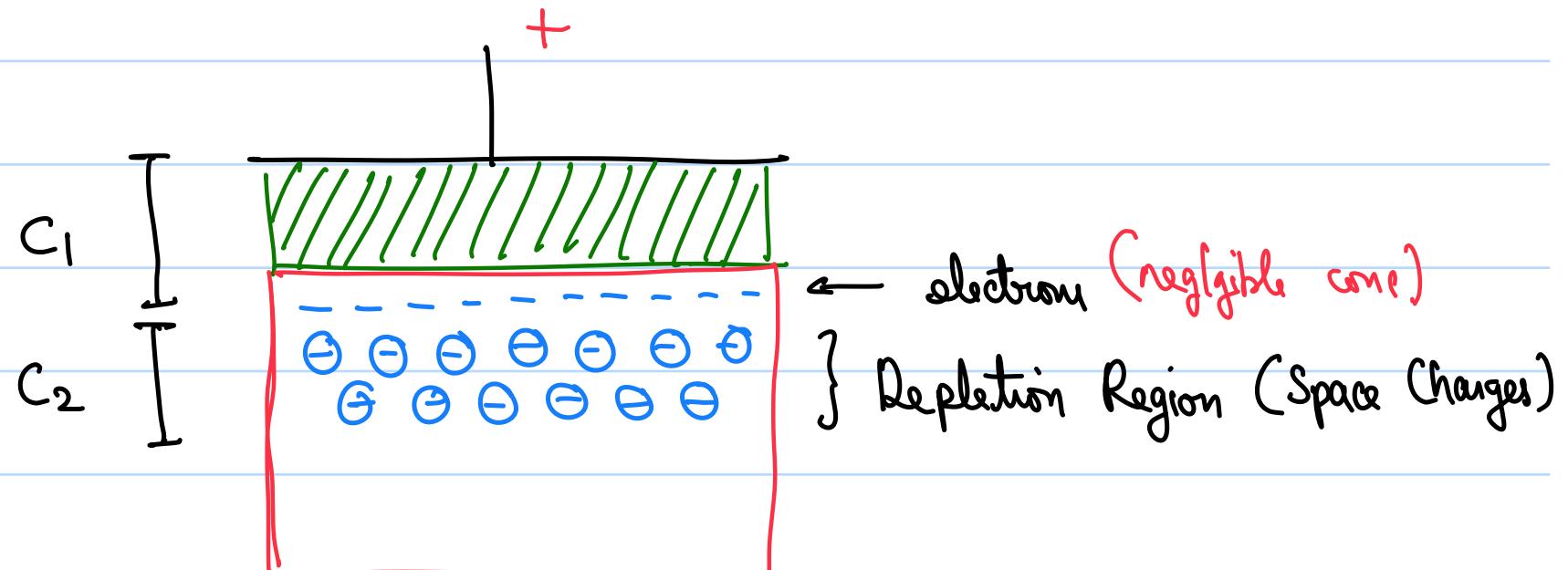
- i) Accumulation Region:-



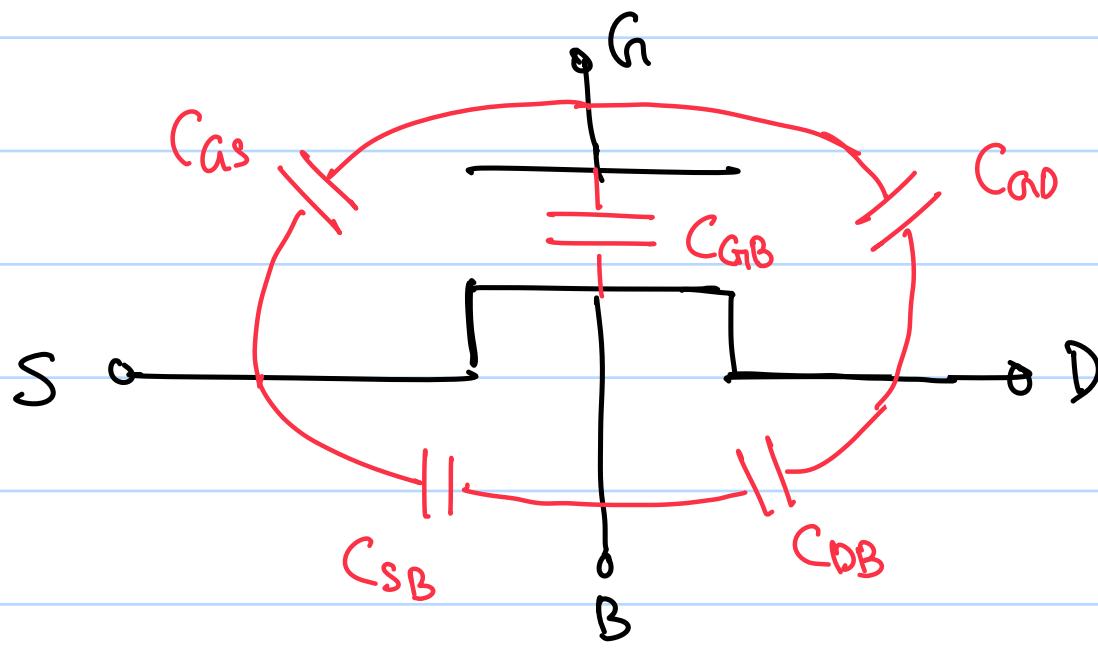
The capacitor acts like a normal parallel plate capacitor.

$$C = \frac{\epsilon A}{d} = \frac{\epsilon W L}{t_{ox}}$$

## 2) Depletion and Inversion :-



## • Capacitance in the Modes of Operation :-



1) In Cutoff:  $C_{GS} = C_{GD} = wC_{ov} \rightarrow$  Overlap Capacitance

$$C_{ov} = WL C_{ox}$$

2) In Linear :-  $C_{GD} = C_{GS} = \frac{WLCon}{2}$

$$C_{GD_{ov}} = C_{GS_{ov}} = WLCon$$

3) In Saturation :-  $C_{GS} = WLCon + \frac{2}{3}WLCon$

5) Mobility Degradation :-

The mobility of the channel depends on the electric field inside the channel as follows,

$$\mu \propto E^0 \text{ for } E < 10^3 \text{ V/cm}$$

$$\mu \propto \frac{1}{\sqrt{E}} \text{ for } E \approx 10^3 - 10^4 \text{ V/cm}$$

$$\mu \propto \frac{1}{E} \text{ for } E > 10^4 \text{ V/cm}$$

Note: Short Channel MOSFETs are those MOSFETs whose channel length is of the same magnitude as the dimensions of the source / drain depletion regions.

6) Tunneling :

Tunneling is the conduction of current through the gate terminal. Significant in short channel MOSFETs.

## → Parametric Extraction :-

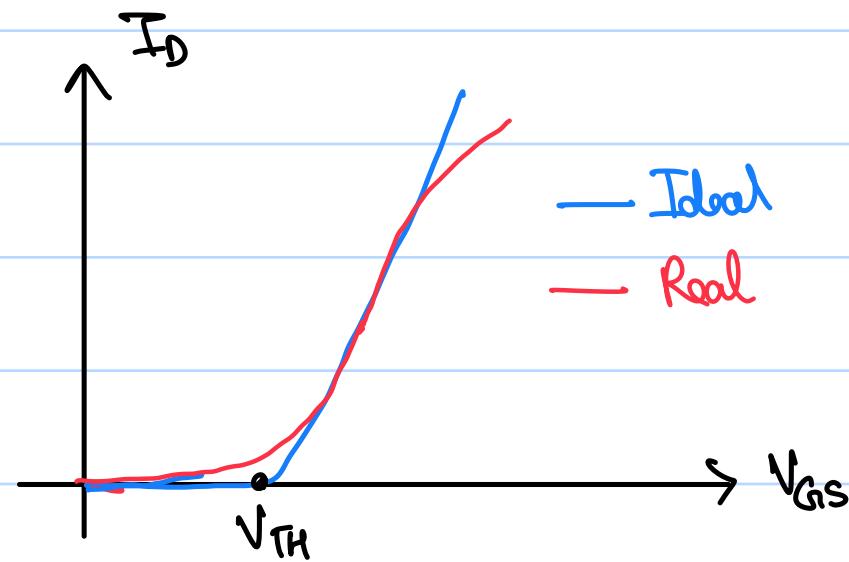
- $V_{GS} \geq V_T$  and  $V_{DS} \ll (V_{GS} - V_T)$  [Deep Trickle]

$$\Rightarrow I_D = \mu_n C_o \frac{w}{L} (V_{GS} - V_{TH}) V_{DS}$$

is of the form  $y = m(x - x_0)$  [ $I_D \propto V_{GS}$  sweep]

$x_0$  is the  $x$  intercept of the line

$\therefore V_{TH}$  is the  $x$  intercept of the above line



We take the point of maximum slope to construct the line equation, to avoid mobility reducing second order effects.

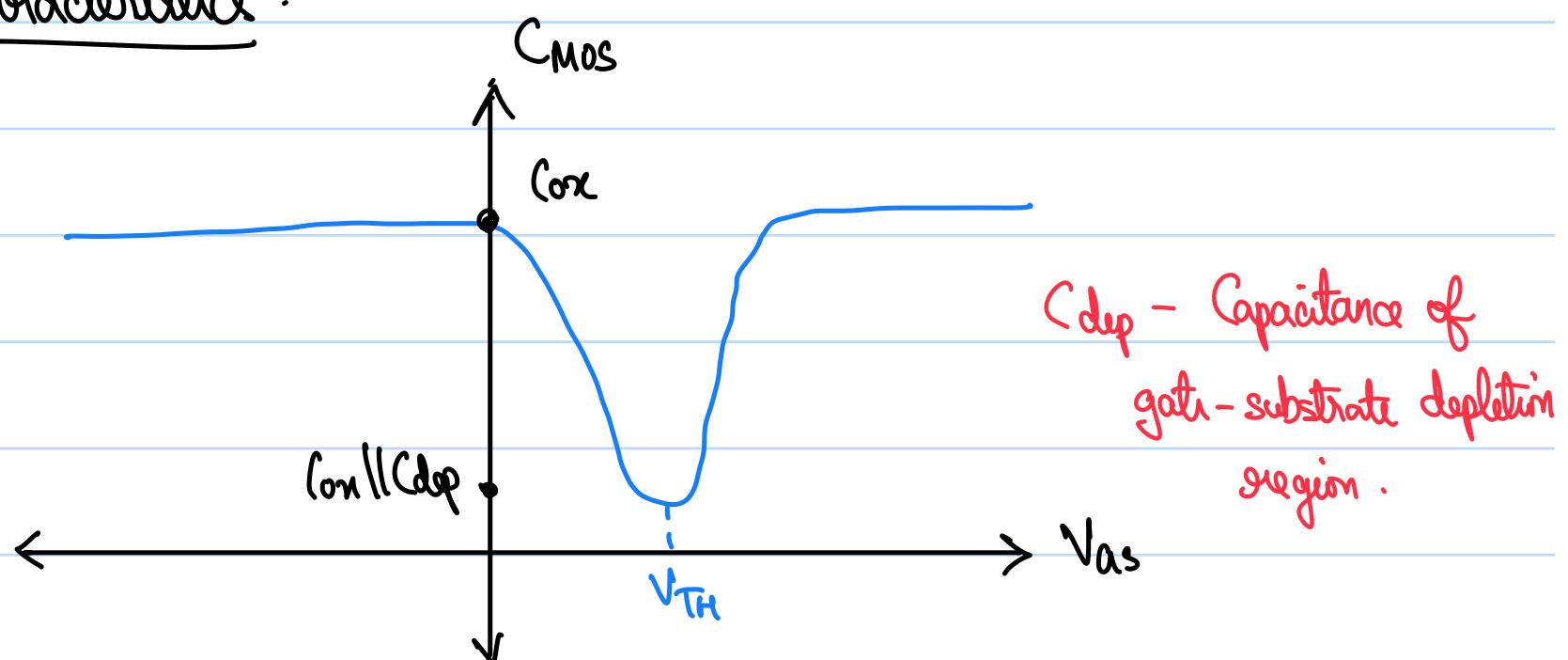
$$\mu = \frac{\mu_0}{1 + \Theta(V_{GS} - V_{TH})}$$

If  $V_{GS} \approx V_{TH}$

$$\mu = \frac{\mu_0}{1 + \Theta(V_{GS} - V_{TH})} \approx \underline{\underline{\mu_0}} \quad [\text{Maximum value}]$$

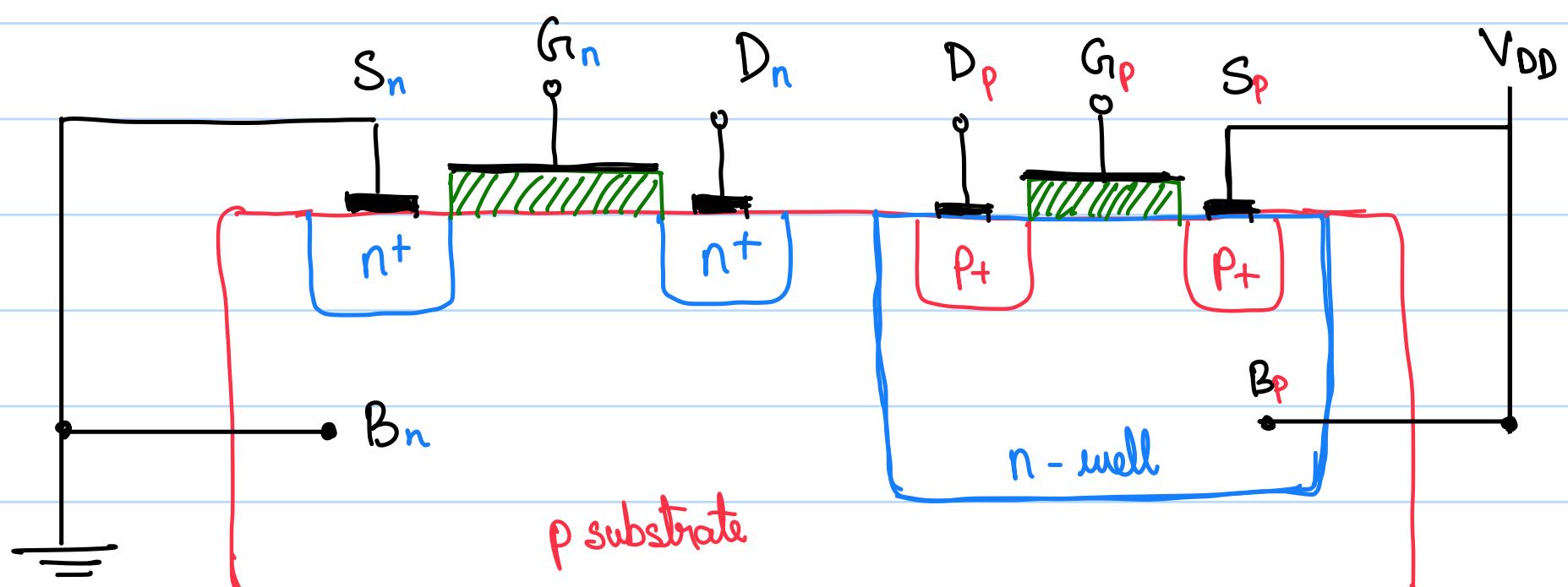
The effect of all the second order effects is modelled using  $\Theta$ . So when  $V_{GS} \approx V_{TH}$ , ie at max mobility,  $\Theta$ 's effect is low.

- C-V Characteristics :-



- CMOS Static Logic Devices :-

- CMOS - Complementary MOS (Combination of PMOS and NMOS)



CMOS Structure

(Well settled)

- Static - Well defined op states : Either 0 (gnd) or 1 ( $V_{DD}$ )

- CMOS Devices will have a Pull-up and Pull-down networks in each logic stage.

Pulls op and ip to 0,1 levels

- Purpose of Logic Stage : To evaluate i/p and pull the o/p up or down to match  $V_{DD}$  or GND.
- No static power consumption , ie, while i/p's are unchanged the power consumption is zero. Circuit still draws power when the i/p's are changed , ie, switching power is non-zero. (dynamic power)

- CMOS Inverter :- ( Rabaray's CMOS Design textbook & Weste )

1) Transfer Characteristic (VTC) (output vs input)

2) Noise Margin

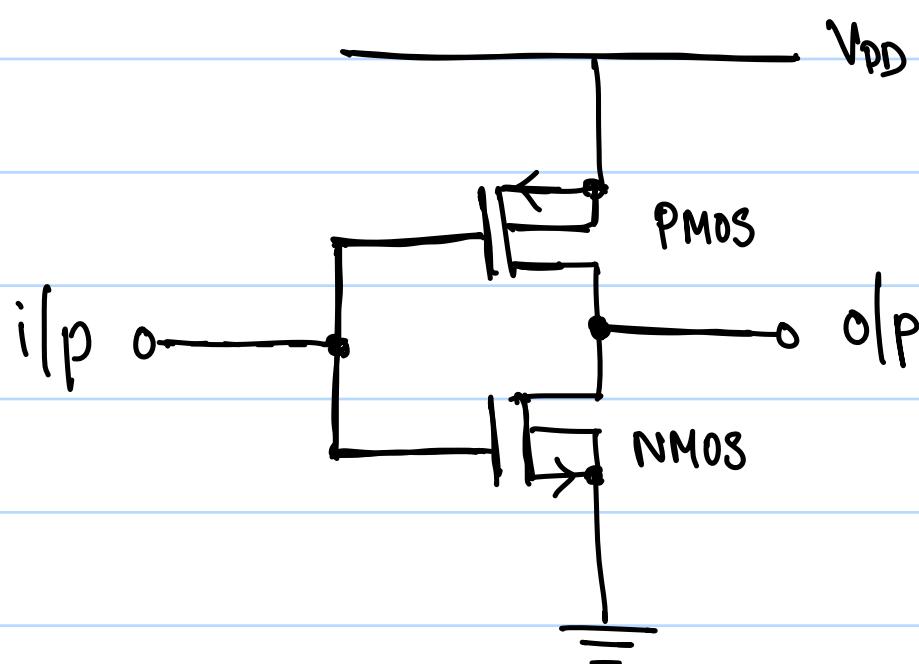
3) Dynamic Characteristics (Rise/fall, time delay)

4) Inverter Design Flow

6) Efficacy

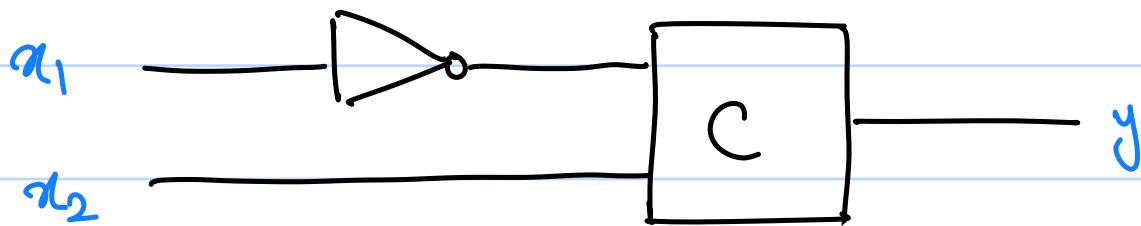
5) Inverter in Other Logics

- Circuit:



- The design procedure of the inverter needs to consider things like time delay, load impedance, output load type, etc.

Example:

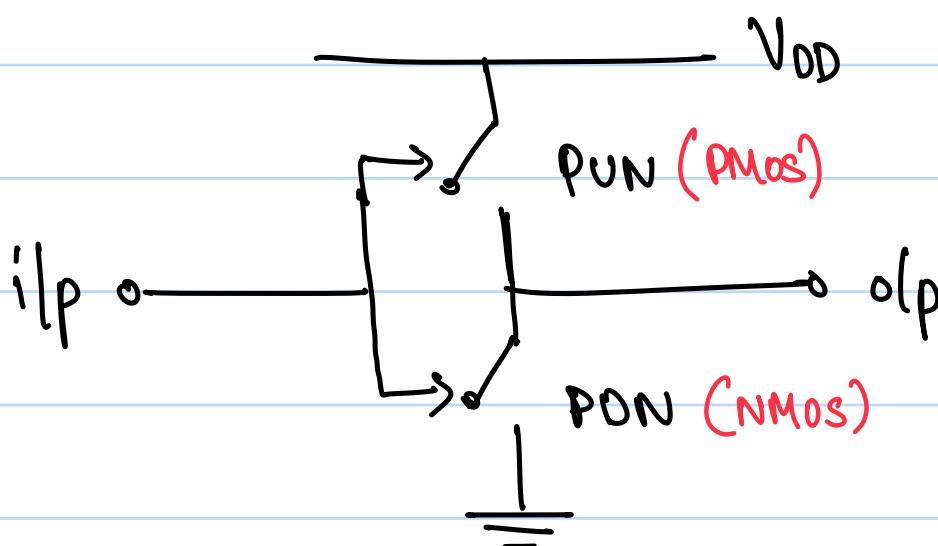


Since the inverter adds some delay,  $x_1$  and  $x_2$  will not reach C at the same time.

- Pull Up and Pull Down Network :-

- Pull Up : Pulls ilp to VDD

- Pull Down : Pulls ilp to GND



PUN and PDN are never on simultaneously.

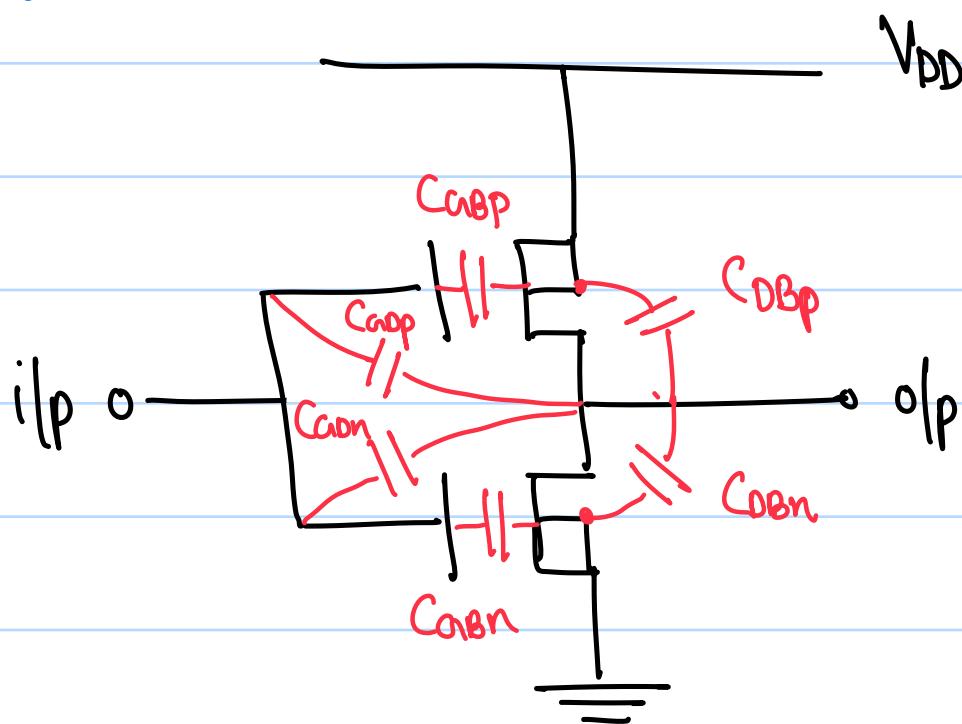
- VTC of CMOS Inverter :-

- The VTC of a CMOS Inverter can be analyzed by plotting the different modes of the PMOS and NMOS.

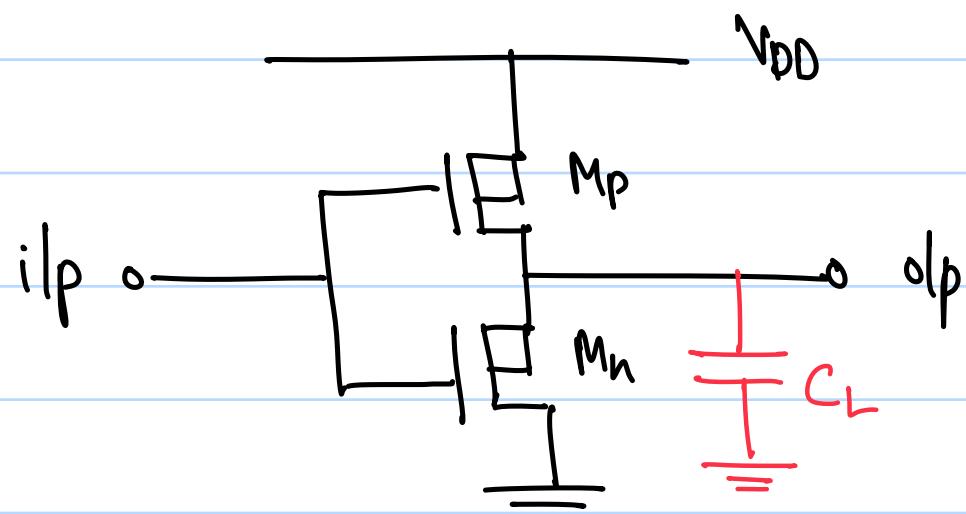
- On condition for  $M_n$  :  $V_{GSn} = V_{ip} \geq V_{Tn}$
- " " " "  $M_p$  :  $V_{Sop} = V_{DD} - V_{ip} \geq |V_{Tp}|$   
 $= V_{ip} \leq V_{DD} - |V_{Tp}|$

Condition	$M_n$	$M_p$	$o/p$
$V_{ip} = 0$	OFF	ON	$V_{DD}$
$V_{ip} = V_{DD}$	ON	OFF	0 } Inverter Behavior $V_{ip} \geq V_{Tn}$ $V_{ip} \leq V_{Tp} + V_{DD}$ $(V_{Tp} \leq 0)$

- Modelling the capacitances in the inverter



Since at the settled state, the voltages are all constant, they can be considered as AC grounds. Therefore the total capacitive impedance can be modelled as a capacitive load.



## Modes of Operation of MOS

Operation Mode	Type	Condition	Equation
Cut-Off	NMOS	$V_{GS} < V_{TH}$	$I_D = 0$
	PMOS	$ V_{GS}  <  V_{TH} $	$I_D = 0$
Linear	NMOS	$V_{GS} \geq V_{TH}$ $V_{DS} \leq V_{GS} - V_{TH}$	$I_D = K_n(2(V_{GS} - V_{TH}) V_{DS} - V_{DS}^2)$
	PMOS	$ V_{GS}  \geq  V_{TH} $ $ V_{DS}  \leq  V_{GS}  -  V_{TH} $	$I_D = K_p(2(V_{GS} - V_{TH}) V_{DS} - V_{DS}^2)$
Saturation	NMOS	$V_{GS} \geq V_{TH}$ $V_{DS} \geq V_{GS} - V_{TH}$	$I_D = K_n(V_{GS} - V_{TH})^2(1 + \lambda V_{DS})$
	PMOS	$ V_{GS}  \geq  V_{TH} $ $ V_{DS}  \geq  V_{GS}  -  V_{TH} $	<del><math>I_D = K_p(V_{GS} - V_{TH})^2(1 - \lambda V_{DS})</math></del>

$$I_D = K_p(V_{GS} - V_{TH})^2(1 + \lambda V_{DS})$$

Case 1:  $0 \leq V_{ilp} < V_{Th}$

$$M_n : V_{asn} = V_{ilp}$$

$$V_{ilp} \leq V_{Th} \rightarrow M_n \text{ is in cutoff}$$

$$M_p : |V_{asp}| = |V_{ilp} - V_{DD}| = V_{DD} - V_{ilp} \quad (V_{ilp} < V_{DD})$$

$$V_{DD} - V_{ilp} \geq |V_{Tp}| \quad [\text{Assuming } V_{Th} \approx |V_{Tp}|] \quad \text{--- (1)}$$

$$\Rightarrow |V_{asp}| \geq |V_{Tp}|$$

$$|V_{asp}| \approx |V_{DD} - V_{DD}| = 0$$

$$|V_{asp}| - |V_{Tp}| = V_{DD} - V_{ilp} - |V_{Tp}| > 0 \quad [\text{Assuming } V_{DD} > 2|V_{Tp}|]$$

$$\Rightarrow |V_{asp}| \leq |V_{asp}| - |V_{Tp}| \rightarrow M_p \text{ is in linear mode}$$

Case 2:  $V_{Th} \leq V_{ilp} < V_{DD}/2$   $\rightarrow V_{ilp} < V_{DD}$  [Take  $V_{ilp} \approx V_{ilp}$ ]

$$M_p : |V_{asp}| = |V_{ilp} - V_{DD}| = V_{DD} - V_{ilp}$$

$$V_{DD} - V_{ilp} \geq |V_{Tp}| \quad [\text{By Assn. (1)}]$$

$$V_{DS} = V_{ilp} - V_{DD} < 0 \quad [V_{ilp} < V_{DD}]$$

$$\Rightarrow |V_{DS}| \approx 0 \Rightarrow |V_{DS}| < |V_{asp}| - |V_{Tp}| \quad \xrightarrow{\text{Mp is linear}}$$

$$M_n : V_{asn} = V_{ip}$$

$$V_{ip} \geq V_{Tn}$$

$$V_{bsn} \approx V_{DD} \Rightarrow V_{bsn} \geq V_{asn} - V_{Tn} \rightarrow M_n \text{ is in Saturation}$$

$$\text{Case 3: } V_{in} = V_{DD}/2 \rightarrow V_{ip} \approx V_{DD}/2$$

$$M_n : V_{asn} = \frac{V_{DD}}{2} \geq V_{Tn} \quad [\text{Assn ②}]$$

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

$$\frac{V_{DD}}{2} \geq \frac{V_{DD}}{2} - V_{Tn} \rightarrow M_n \text{ is in Saturation}$$

$$M_p : |V_{asp}| = \left| \frac{V_{DD}}{2} - V_{DD} \right| = \frac{V_{DD}}{2} > V_{Tp}$$

$$|V_{asp}| = \left| \frac{V_{DD}}{2} - V_{DD} \right| = \frac{V_{DD}}{2}$$

$$\frac{V_{DD}}{2} > \frac{V_{DD}}{2} - |V_{Tp}| \Rightarrow |V_{asp}| \geq |V_{asp}| - |V_{Tp}| \rightarrow M_p \text{ is in Saturation}$$

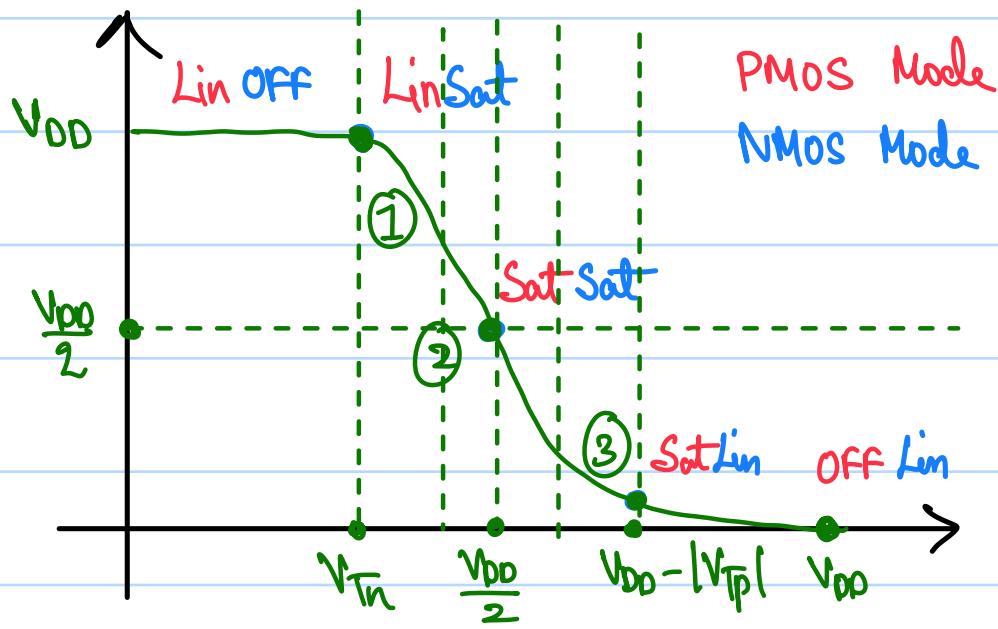
III proof can be stated for  $V_{DD}/2 < V_{ip} < V_{DD} - |V_{Tp}|$  and  $V_{DD} - |V_{Tp}| < V_{ip}$ .

Once the voltages are settled

$\therefore$  The VTC of the CMOS Inverter can be summarized as,

Condition	M <sub>p</sub>	M <sub>n</sub>	O/p
$0 \leq V_{in} \leq V_{Tn}$	Lin	OFF	$V_{DD}$
$V_{Tn} \leq V_{in} < \frac{V_{DD}}{2}$	Lin	Sat	$< V_{DD}$
$V_{in} = \frac{V_{DD}}{2}$	Sat	Sat	$\frac{V_{DD}}{2}$
$\frac{V_{DD}}{2} < V_{in} < V_{DD} -  V_{Tp} $	Sat	Lin	$< \frac{V_{DD}}{2}$
$V_{DD} -  V_{Tp}  < V_{in}$	OFF	Lin	0

This switching point is valid if  $V_{in} = |V_{Tp}|$



Note: By KCL, currents through both the MOSFETs will always be equal in the 3 intermediate points (LinSat SatSat SatLin)

At ① (LinSat)

$$I_{Dp\text{ Lin}} = I_{Dn\text{ Sat}}$$

$$= K_p \left[ (V_{SD} - |V_{Tp}|) V_{SD} - \frac{V_{SD}^2}{2} \right] = \frac{1}{2} K_n (V_{DS} - V_{Tn})^2$$

$$= K_p \left[ (V_{DD} - V_{in} - |V_{Tp}|)(V_{DD} - V_{Dp}) - \frac{(V_{DD} - V_{Dp})^2}{2} \right] = \frac{K_n}{2} (V_{Dp} - V_{Tn})^2$$

On Solving, we will get, (Take  $V_{DD} - V_{Dp} = y$ ,  $V_{Dp} - V_{Tn} = x$  easy)  
and  $K_n = K_p$

$$V_{Dp} = (V_{Dp} + |V_{Tp}|) \pm \sqrt{(V_{DD} - V_{Tn} - |V_{Tp}|)(V_{DD} - 2V_{Dp} + V_{Tn} - |V_{Tp}|)}$$

Since  $V_{Dp} \approx V_{DD}$  near ①,

$$V_{Dp} = (V_{Dp} + |V_{Tp}|) + \sqrt{(V_{DD} - V_{Tn} - |V_{Tp}|)(V_{DD} - 2V_{Dp} + V_{Tn} - |V_{Tp}|)}$$

$$\text{Condition : } V_{DD} \geq V_{Tn} + |V_{Tp}|, V_{Dp} \leq \frac{V_{DD} + V_{Tn} - |V_{Tp}|}{2}$$

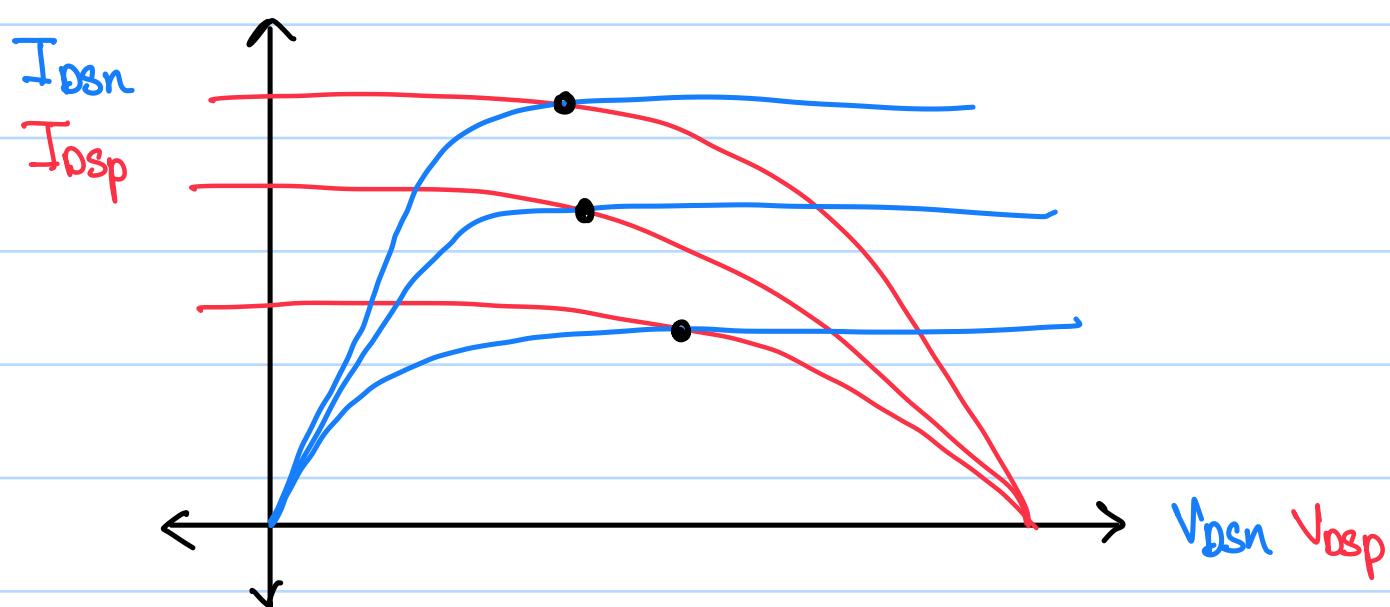
Therefore, the operating requirements of a CMOS inverter are:

$$V_{DD} \geq V_{Th} + |V_{Tp}| \quad 2) \quad V_{ilp} \leq \frac{V_{DD} + V_{Th} - |V_{Tp}|}{2} \approx \frac{V_{DD}}{2}$$

min. power supply

max. ilp voltage

||| by we can calculate for all the other points , by equating PMOS current and NMOS current .



$V_{DD} \approx 1.8\text{ V}$  for 180 nm technology node.

At ② ( Sat Sat ),

$$\begin{aligned}
 I_{DSnSut} &= I_{DspSut} \\
 &= K_n (V_{asn} - V_{Tnn})^2 = K_p (V_{sap} - |V_{Thpl}|^2)^2 \\
 &= K_n (V_{ilp} - V_{Tnn})^2 = K_p (V_{od} - V_{ilp} - |V_{Thpl}|^2)^2
 \end{aligned}$$

$$\Rightarrow V_{ilp} = \frac{V_{DD} + \sqrt{B} V_{Thn} - |V_{Thpl}|}{1 + \sqrt{B}} \quad B = \frac{k_n}{k_p}$$

$$V_{DSN} \geq V_{ASN} - V_{THN}$$

$$\Rightarrow V_{Olp} \geq V_{ILP} - V_{THN}$$

$$V_{SDP} \geq V_{SBN} - |V_{THP}|$$

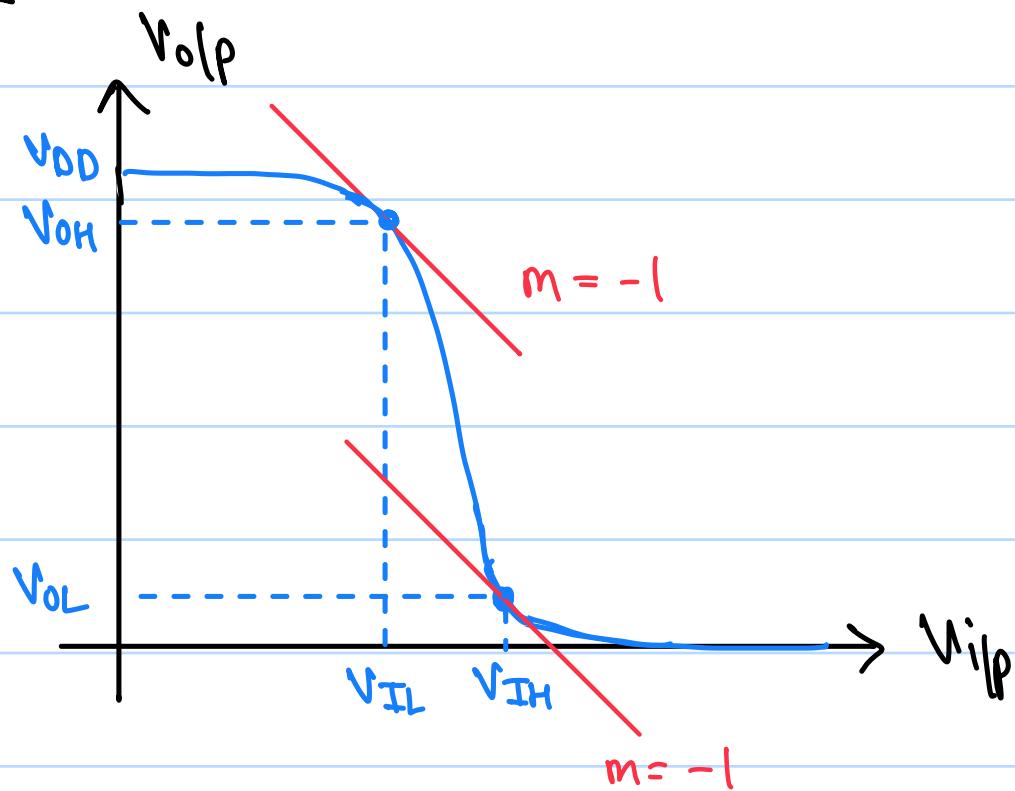
$$V_{DD} - V_{Olp} \geq V_{DD} - V_{ILP} - |V_{THP}|$$

$$\Rightarrow V_{Olp} \leq V_{ILP} + |V_{THP}|$$

$$\therefore V_{Olp} \in [V_{ILP} - V_{THP}, V_{ILP} + |V_{THP}|]$$

In this region,  $V_{Olp}$  vs  $V_{ILP}$  is approximately a straight line of very high slope. Therefore it shows amplifying action.

- Noise Margin :-



- To avoid the amplification behavior, we set

$$V_{out} \in [V_{OH}, V_{OD}] \longrightarrow \text{Logic 1}$$

$$V_{out} \in [0, V_{OL}] \longrightarrow \text{Logic 0}$$

$\Rightarrow V_{OH} = \text{Minimum o/p for o/p logic 1}$

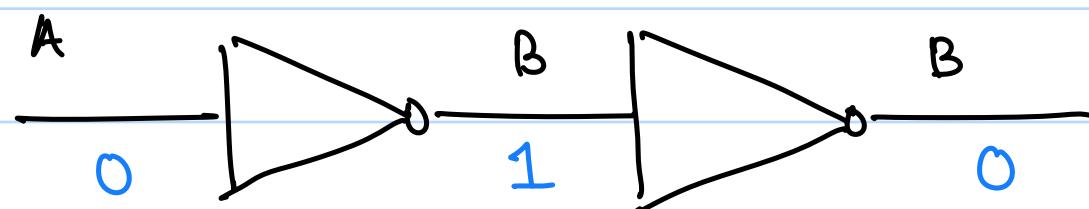
$V_{OL} = \text{Maximum o/p for o/p logic 0}$

$V_{IH} = \text{Minimum i/p for o/p logic 0}$

$V_{IL} = \text{Maximum i/p for o/p logic 1}$

- If  $V_{i/p} \in (V_{IL}, V_{IH})$ , the Inverter acts like an amplifier, which is unwanted since it **amplifies noise as well, creating undefined behavior.**

- If we have 2 inverters in series as,



$$V_A \in [0, V_{IL}] \quad V_B \in [V_{OH}, V_{OD}] \quad V_C \in [0, V_{OL}]$$

$$[V_{OH}, V_{OD}] \subseteq [V_{IH}, V_{OD}]$$

- From the VTC, we can say that  $V_{OH} > V_{IH}$  and  $V_{OL} < V_{IL}$ . So, we define the noise margin of our inverter as,

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

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

- But since  $V_{IH} < V_{OH}$ , even if  $V_B < V_{OH}$ , as long as  $V_B > V_{IH}$  the inverter C will output 0. Hence the noise margin  $NM_H$ . Only for  $V_B$  low, we have  $NM_L$ .

- Also if the 2 inverters are not the same, then it must be ensured that,

$$V_{OH1} > V_{IH2}$$

- The parameter extraction of  $V_{OH}$ ,  $V_{OL}$ ,  $V_{IH}$ ,  $V_{IL}$ , can be done as follows,

For point  $(V_{IL}, V_{OH})$ , Inv. is in region of LinSat

For point  $(V_{IH}, V_{OL})$ , Inv. is in region of SatLin

We know the current equation at both these regions (or atleast how to derive them). The required points will have  $\frac{dV_{Op}}{dV_{Ip}} = -1$

We will get,

$$V_{IL} = \frac{3V_{DD} + 5V_{Tn} - 8|V_{Tp}|}{8}$$

$$V_{OH} = \frac{7V_{DD} + V_{Tn} + |V_{Tp}|}{8}$$

$$V_{IH} = \frac{5V_{DD} + 3V_{Tn} - 5|V_{Tp}|}{8}$$

$$V_{OL} = \frac{V_{DD} - V_{Tn} + |V_{Tp}|}{8}$$

Assume

$$K_n = K_p$$

$$(or) \beta = 1$$

$$\Rightarrow NM_H = \frac{V_{DD} - V_{Th} + 3|V_{Tp}|}{4}$$

$$NM_L = \frac{V_{DD} + 3V_{Th} - |V_{Tp}|}{4}$$

if  $V_{Th} = |V_{Tp}| = V_T$

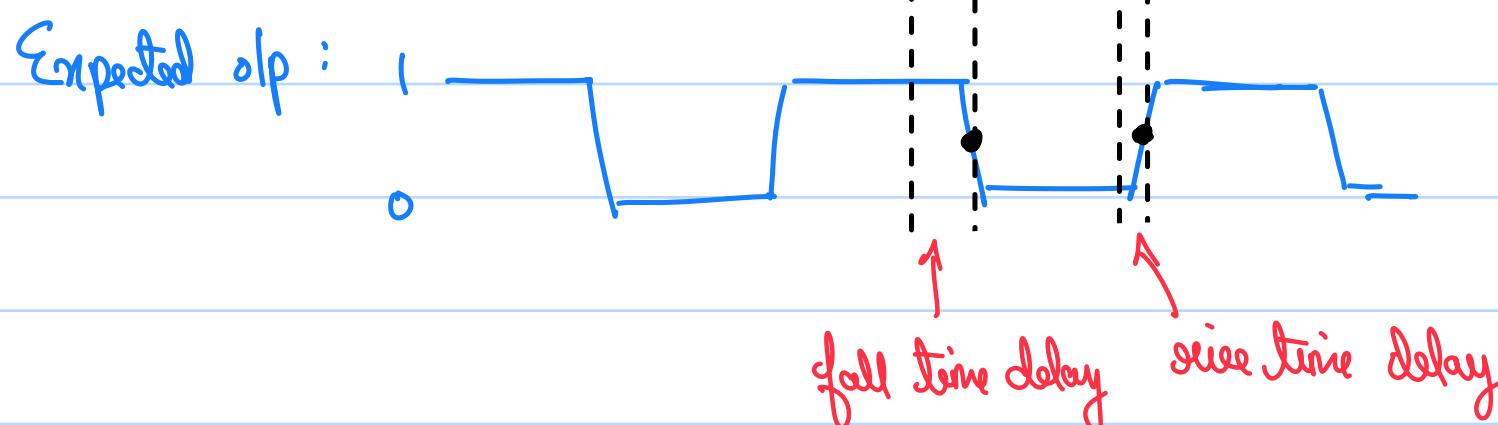
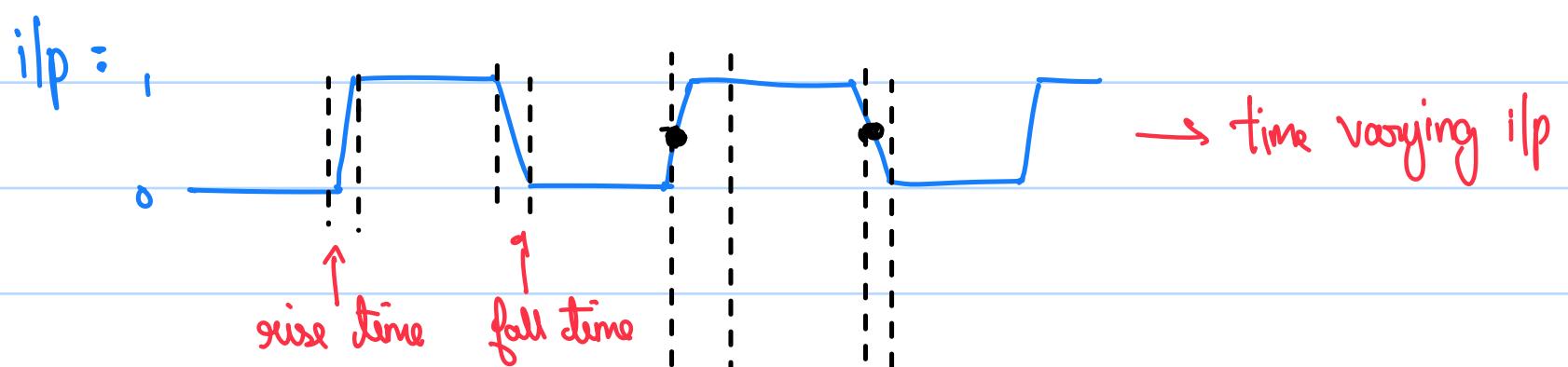
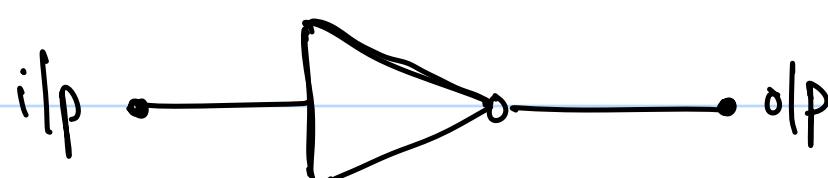
- From the  $NM_H/L$  equations we can see that increasing  $V_{DD}$  and  $V_{Th}$  will give us better noise margin.

Trade-offs :

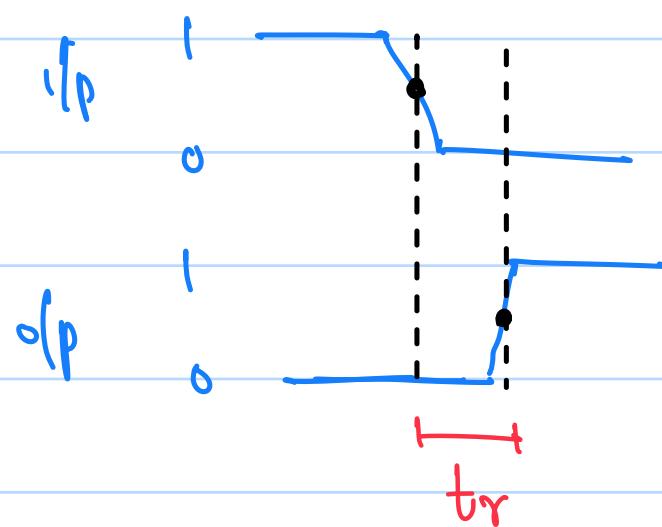
- $V_{DD} \uparrow$  increases power consumption
- $V_T \uparrow$  increases min.  $V_{DD}$ , lower current causing delays and lower speed.

Higher current charge the capacitances faster, reducing delay / increasing bandwidth

### Dynamic Characteristics :-



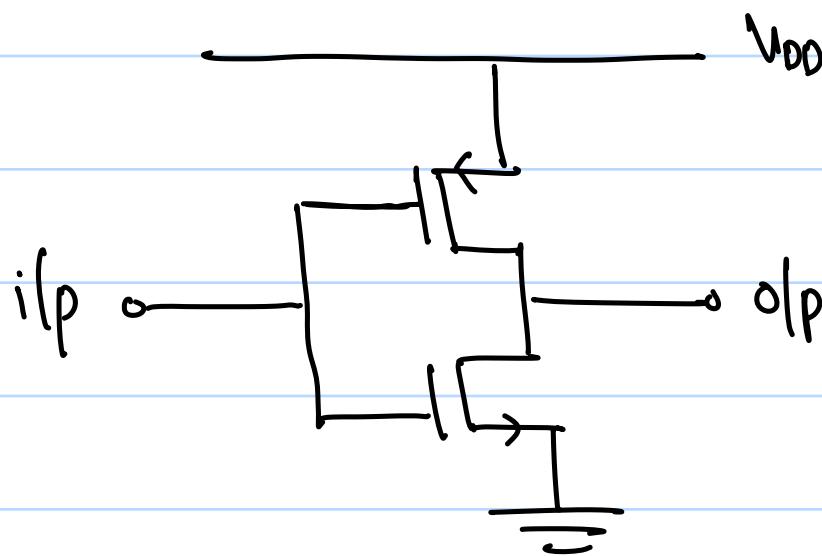
- The delay between the i/p and its respective o/p is due to the internal capacitances of the device and is unavoidable.
- Ideally, we want fall time delay and rise time delay to be the same, to get symmetric behavior.
- We assume that one of the MOSFETs is ON and the other is OFF always, to develop an intuition for the system.
- Rise Time Delay :-



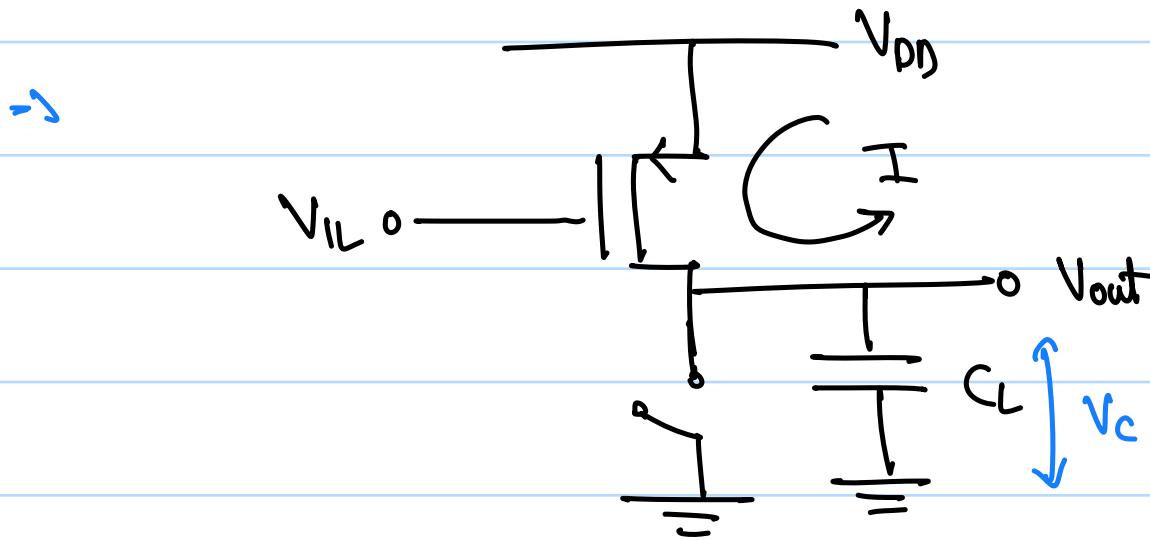
### Assumptions:

$$V_{ilp} = V_{IL}$$

$V_{olp}$  = Grows from 0 to  $V_{OH}$ .



~ In rise time, NMOS is OFF and PMOS is ON



$$I = I_{SD}, \Rightarrow I_{SD} = C \frac{dV_c}{dt}$$

$$\Rightarrow dt = \frac{C dV_c}{I_{SD}}$$

$$\Rightarrow t_r = \int_0^{V_{on}} \frac{C dV_c}{I_{SD}}$$

The mode of operation of the MOSFET,

$$V_{SDp} = V_{DD} - V_c$$

$$V_{on} = V_{SDp} - |V_{Tp}| = V_{DD} - V_{IL} - |V_{Tp}|$$

For linear,

$$V_{SDp} < V_{on}$$

$$\Rightarrow V_{DD} - V_c < V_{DD} - V_{IL} - |V_{Tp}|$$

$$\Rightarrow V_c > V_{IL} + |V_{Tp}|$$

$\therefore$  For  $V_c \in [0, V_{IL} + |V_{Tp}|]$ , PMOS is in Saturation mode

For  $V_c \in [V_{IL} + |V_{Tp}|, V_{on}]$ , PMOS is in Linear mode

$$\rightarrow t_r = \int_0^{V_{IL} + |V_{Tp}|} \frac{dV_c}{I_{SO\text{ sat}}} + \int_0^{V_{IL} + |V_{Tp}|} \frac{dV_c}{I_{SO\text{ lin}}}$$

↑ constant current changing      ↑ RC charging

$$I_{SPP} = \begin{cases} \frac{1}{2} K_p (V_{SDP} - |V_{Tp}|)^2, & \text{saturation} \\ K_p (V_{SDP} - |V_{Tp}|) V_{SDP} - \frac{V_{SDP}^2}{2}, & \text{linear} \end{cases}$$

Where,

$$V_{SDP} = V_{DD} - V_{IL}, \quad V_{SDP} = V_{DD} - V_C$$

On solving, we will get,

$$t_r = \frac{\alpha C [V_{IL} + |V_{Tp}|]}{K_p [V_{DD} - V_{IL} - |V_{Tp}|]^2} + \frac{C}{K_p [V_{DD} - V_{IL} - |V_{Tp}|]} \ln \left( \frac{V_{DD} + V_{OH} - 2V_{IL} - 2|V_{Tp}|}{V_{DD} - V_{OH}} \right)$$

↑ constant current charging time      ↑ RC charging time

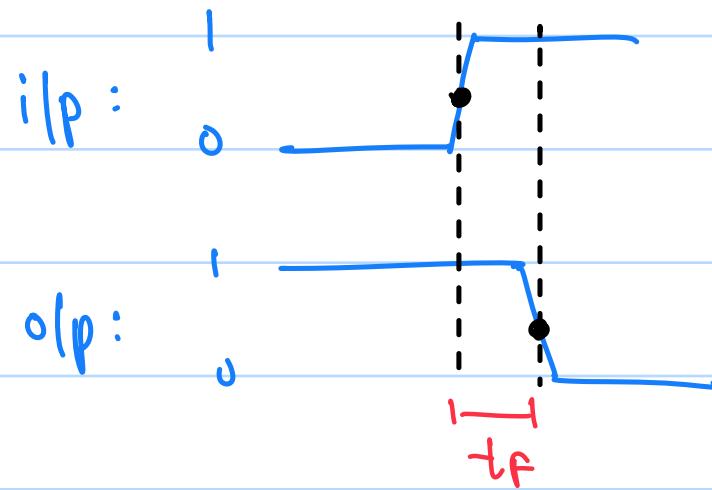
is the apparent  
ON resistance of  
the circuit

~ We can see that  $t_r$  is dependent on the ON resistance of the circuit

~ Also, the voltage terms resemble the Noise margin parameters of the inverter, i.e., for given Noise margin parameter,

$$\frac{K_p t_r}{C} \approx \text{const}$$

## - Fall Time Delay :-

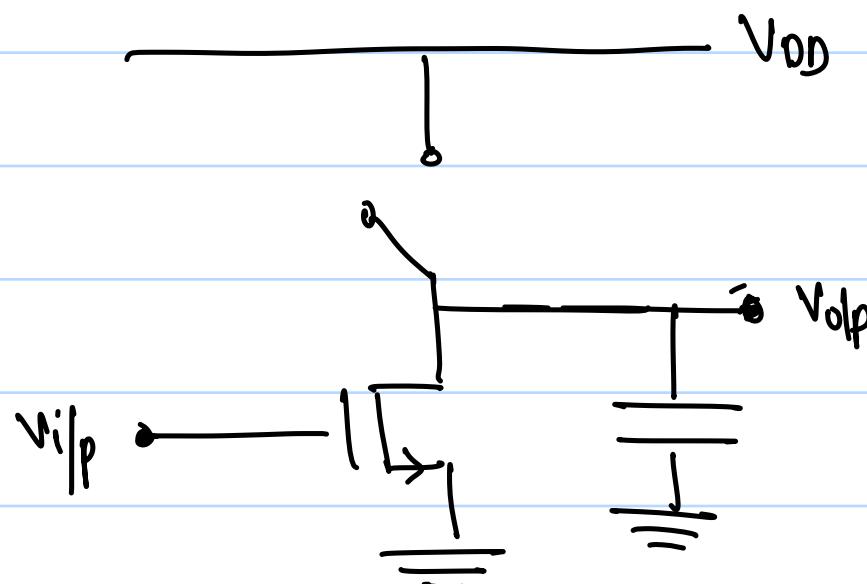


### Assumptions:

$$V_{lp} = V_{IH}$$

V<sub>lp</sub> = Falls from V<sub>DD</sub> to V<sub>OH</sub>

PMOS is OFF, NMOS is ON



$$I_{DS} = -C \frac{dV_C}{dt} \Rightarrow dt = -\frac{C dV_C}{I}$$

$$\Rightarrow t_f = -C \int_{V_{lp}}^{V_{O2}} \frac{dV_C}{I} = C \int_{V_{lp}}^{V_{O2}} \frac{dV_C}{I}$$

$$\Rightarrow t_f = \int_{V_{DD}}^{\frac{V_{IH} - V_{TN}}{I_{DSSat}}} \frac{dV_C}{I_{DSSat}} - \int_{\frac{V_{IH} - V_{TN}}{I_{DSSat}}}^{\frac{V_{O2}}{I_{DSlin}}} \frac{dV_C}{I_{DSlin}}$$

$$\Rightarrow \frac{K_{n\text{tf}}}{c} = \frac{2(V_{DD} - V_{IH} + V_{Th})}{(V_{IH} - V_{Th})^2} + \frac{1}{V_{IH} - V_{Th}} \ln \left( \frac{2(V_{IH} - V_{Th}) - V_{OL}}{V_{OL}} \right)$$

$$\Rightarrow \frac{K_{n\text{tf}}}{c} = \text{const for a given NM}$$

- Delay: -

The total delay of the inverter is given by,

$$\tau = \frac{t_r + t_f}{2}$$

$$\Rightarrow \tau \propto \frac{c}{k_p + k_n}$$

~ We know,

$$C_{IN} = C_{asn} + C_{asp} + C_{aon} + C_{aop}$$

$$\approx w_n L_n C_{oxn} + w_p L_p C_{oxp} + C_{ovn} + C_{ovp}$$

$$+ w_n C_{ovn} + w_p C_{ovp}$$

Since  $K_{p/n} \propto W$ , we can say  $C_{IN} \propto K_p / n$

$$\Rightarrow \tau \propto \frac{C_L}{C_{IN}}$$

~ Also,  $C_L = C_{ext} + C_{self}$

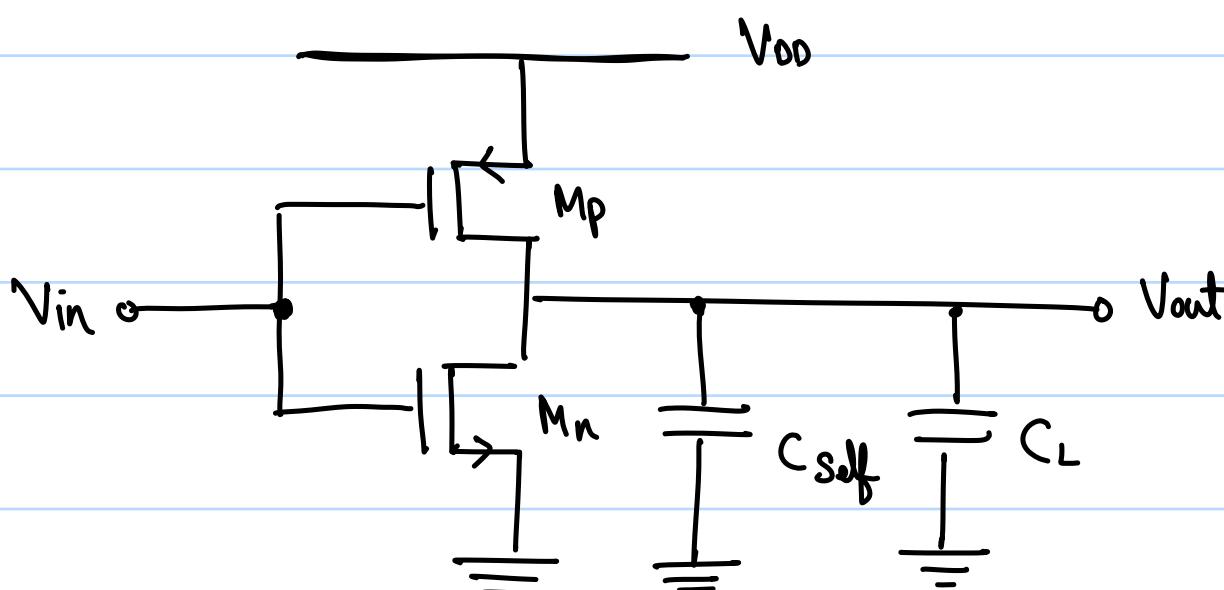
$$C_{self} = C_{DBn} + C_{DBp} + C_{aon} + C_{aop}$$

$$\Rightarrow C_{self} \propto K \Rightarrow C_{self} \propto C_{IN}$$

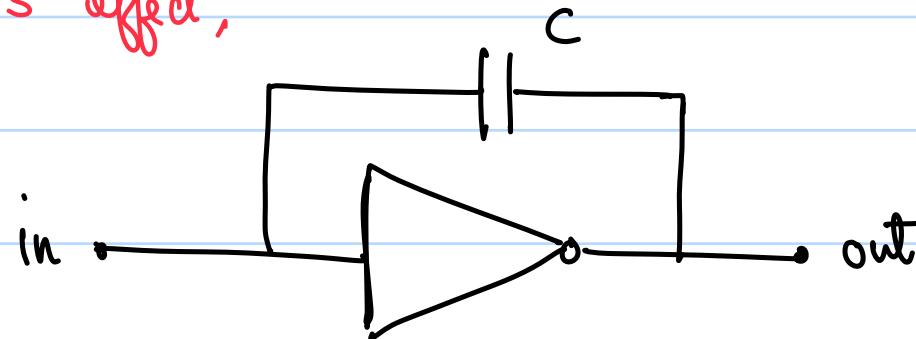
$$\Rightarrow T \propto \frac{C_{ext}}{C_{IN}} + \text{const}$$

~ Because of the constant in the equation, if we increase  $C_{IN}$  to a very large number, any further increase will be dominated by the constant.

~ Also,  $C_{IN}$  will act as  $C_{ext}$  for any logic circuit in the previous stage.



Note: Miller's effect,



This is high for  
inv. in the Ampl.

$$C_{out} = C(1 + A) \quad A - \text{Gain mode}$$

+ Any DC potential acts like a ground for AC responses in capacitances.

$$\sim C_{self} = C_{Aop} + C_{GDn} + C_{DBn} + C_{DB}$$

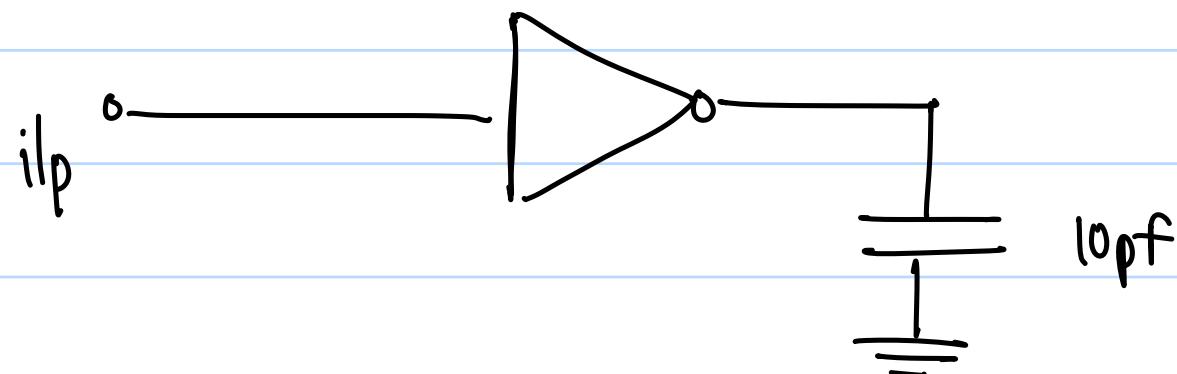
~ Define  $\frac{C_{ext}}{C_{in}} = h$ , since Delay =  $\frac{C_{ext}}{C_{in}} + \text{const.}$

$$\Rightarrow \text{Delay} = (h + p)\tau \quad p - \text{Constant Term}$$

$\tau$  - Proprietary Const.

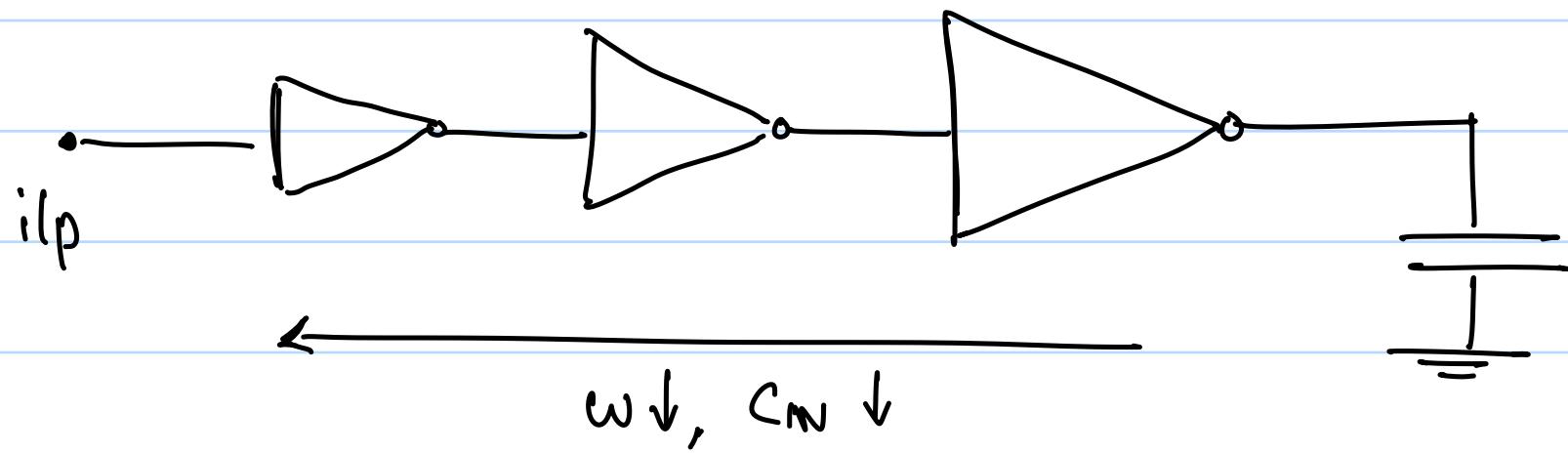
$h$  - Electrical effort of the gate.

~ If we have a big capacitor on the output side.



Since a huge driving current is required to quickly charge the capacitor, the width of the inverter must be high, which increases  $C_{IN}$ .

If  $C_{IN}$  must be low, we can chain multiple inverters together.



Since  $C_{IN} < C_{ext}$ , smaller inverters are used to drive the

smaller i/p capacitance.

~ If  $K_n = K_p$ , then  $\frac{W_p}{W_n} = \frac{M_n}{M_p} \Rightarrow W_p = \frac{M_n}{M_p} W_n$

If we say an inverter has width  $W$ , we usually mean the width  $W_n$ .  $W_p$  is calculated by  $\frac{M_n}{M_p} W_n$ . ( $\approx 2W_n$ )

~ Having such a chain of inverters will actually be faster than a single inverter, ie, the delay is lower.

Min. size transistor :  $L = L_{min} = \text{Tech.node}$ ,  $W = W_{min} > L_{min}$   
 $0.18\mu\text{m}$ ,  $0.27\mu\text{m}$  in 180nm TSMC.

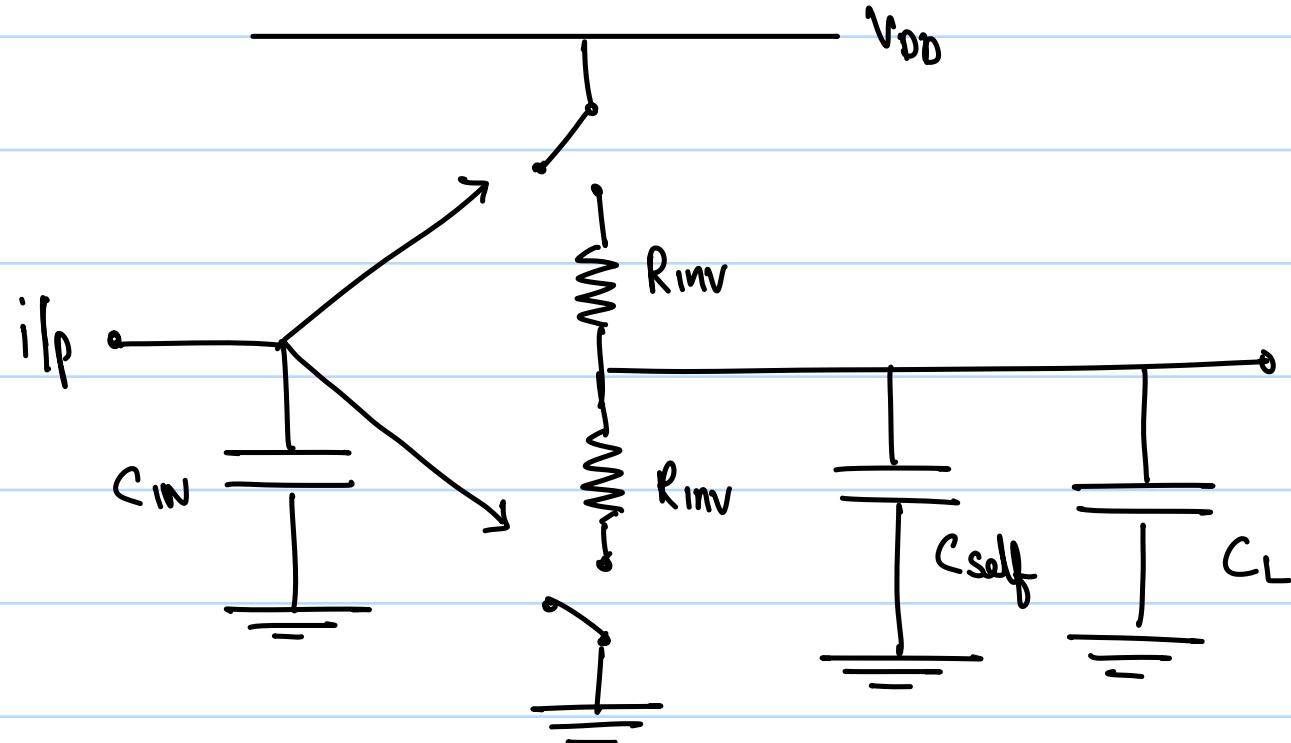
If we directly use a min. size transistor, the delay will be extremely high. (low  $W$ ). But  $C_{IN}$  will be good.

So, using a series of inverters with increasing widths will be faster, since the smaller transistors drive a smaller capacitance, leading to smaller delays.

This is termed as the buffer chain / tapered buffer design.

## - Buffer Chain Design :

~ Simplified Model of an Inverter,



Case 1: If  $\rho = 0$  (ie  $C_{self} = 0$ ), then  $T_1 = R_{inv} C_L$

Case 2:  $T_2 = R_{inv} (C_{self} + C_L) \rightarrow$  General Case

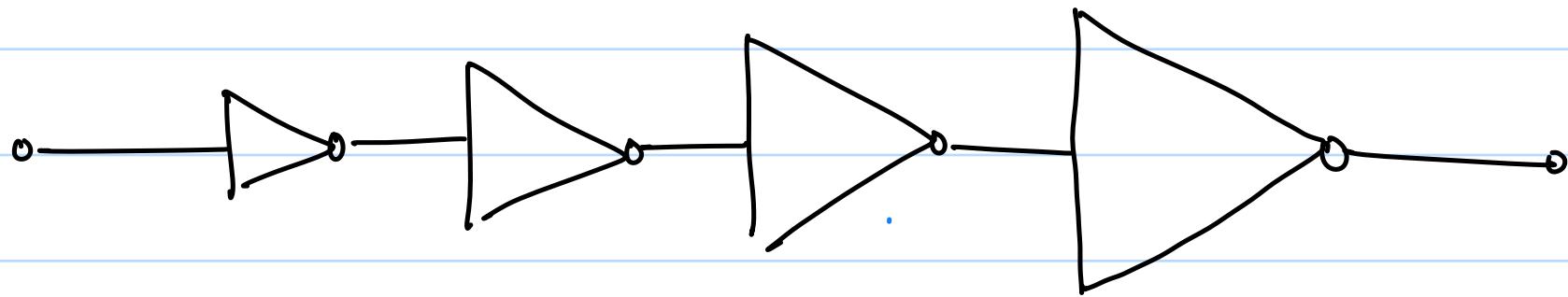
$$\frac{T_2}{T_1} = \frac{C_{self} + C_L}{C_L} = 1 + \frac{C_{self}}{C_L}$$

In buffer chain,  $C_L = i_ip$  capacitance of next inverter. ( $C_{in-inv}$ )

$$\Rightarrow \frac{T_2}{T_1} = 1 + \frac{C_{self}}{C_{in-inv}}$$

$$\Rightarrow T_2 = \left(1 + \frac{C_{self}}{C_{in-inv}}\right) T_1$$

So, given  $C_L$  and  $C_{in}$  limitations we can design a buffer chain for the lowest delay, through the following algorithm.



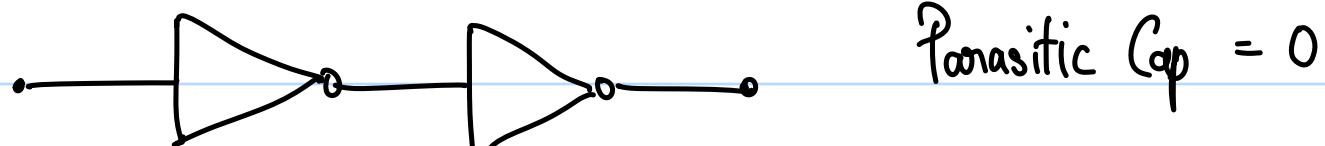
Total delay =  $\sum$  Delay in each inverter.

$$\text{Also, in } \tau_2 = \tau_1 \left( 1 + \frac{C_{self}}{C_{in-inv}} \right)$$

Since  $C_{self} \propto C_{in-inv}$ ,  $\frac{C_{self}}{C_{in-inv}} = \text{const} = f$

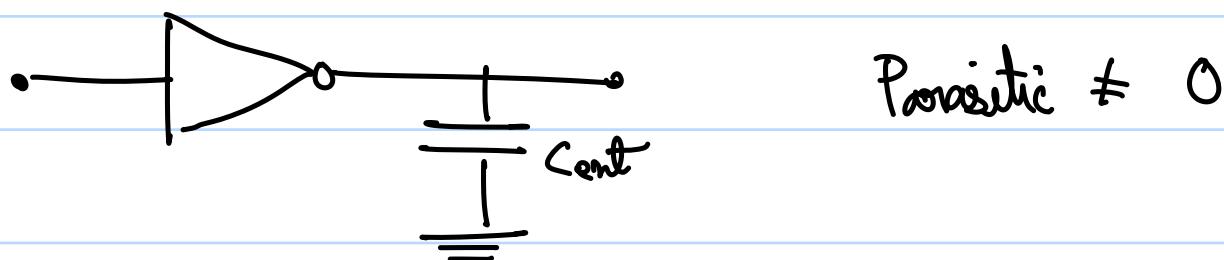
$$\Rightarrow \tau_2 = \tau_1 (1 + p)$$

Case 1:



$$\tau = R_{inv} \cdot C_{in-inv}$$

Case 2:

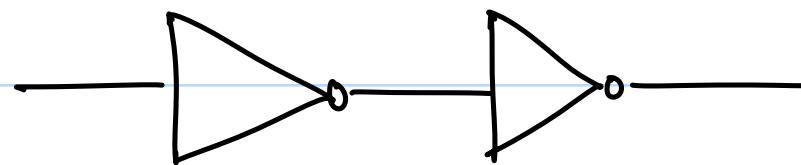


$$D = R_{inv} (C_{ext} + C_{self})$$

$$\frac{D}{\tau} = \frac{C_{out} + C_{self}}{C_{in-inv}}$$

$$\Rightarrow D = \tau \left( \frac{C_{out}}{C_{in-inv}} + p \right)$$

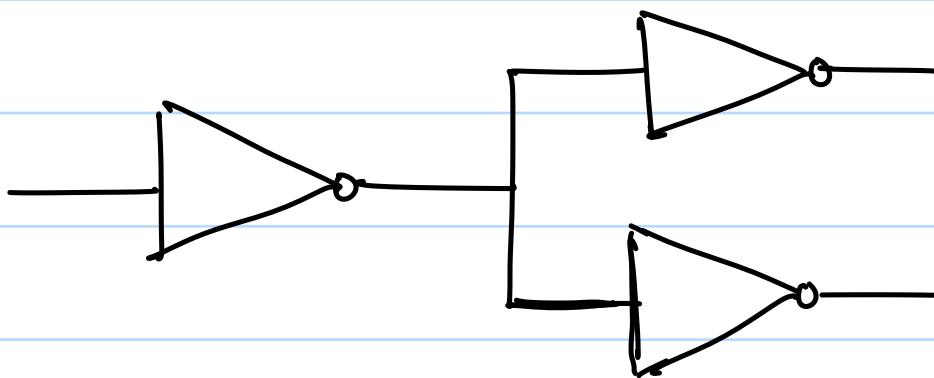
Using this formula, if the inverter is driving 1 other inverter,



$$C_{out} = C_{in-inv}$$

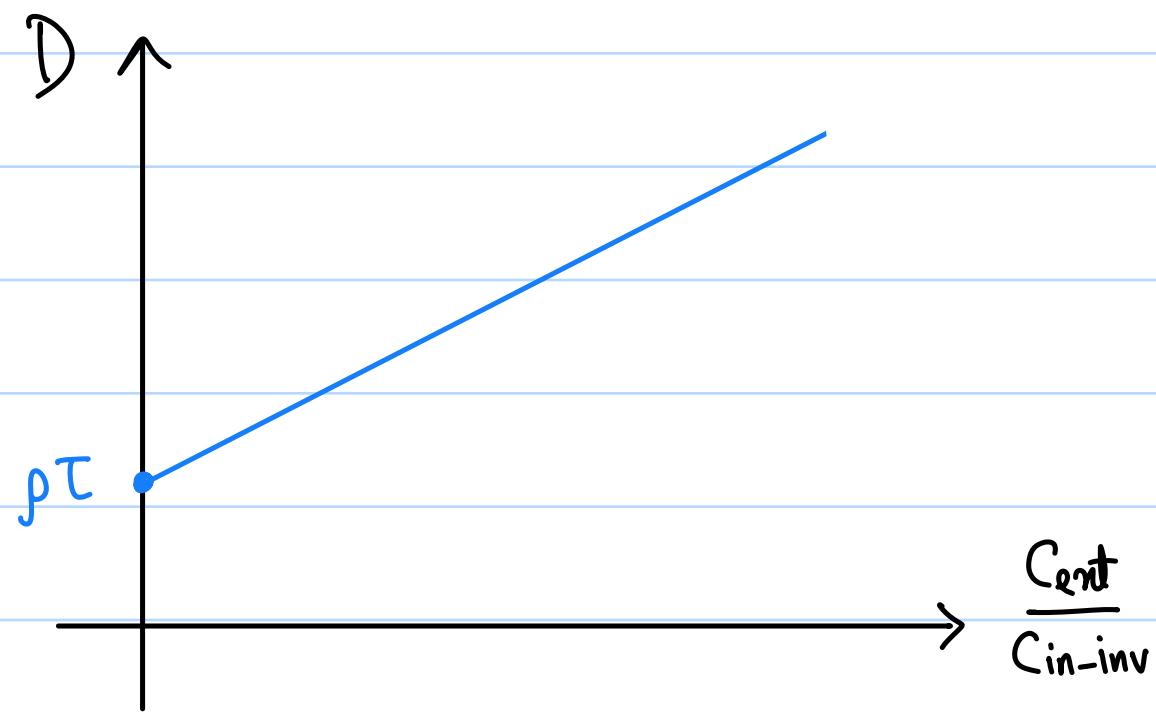
$$\Rightarrow D_1 = \tau(1+p)$$

If the inverter is driving 2 other inverters in parallel,

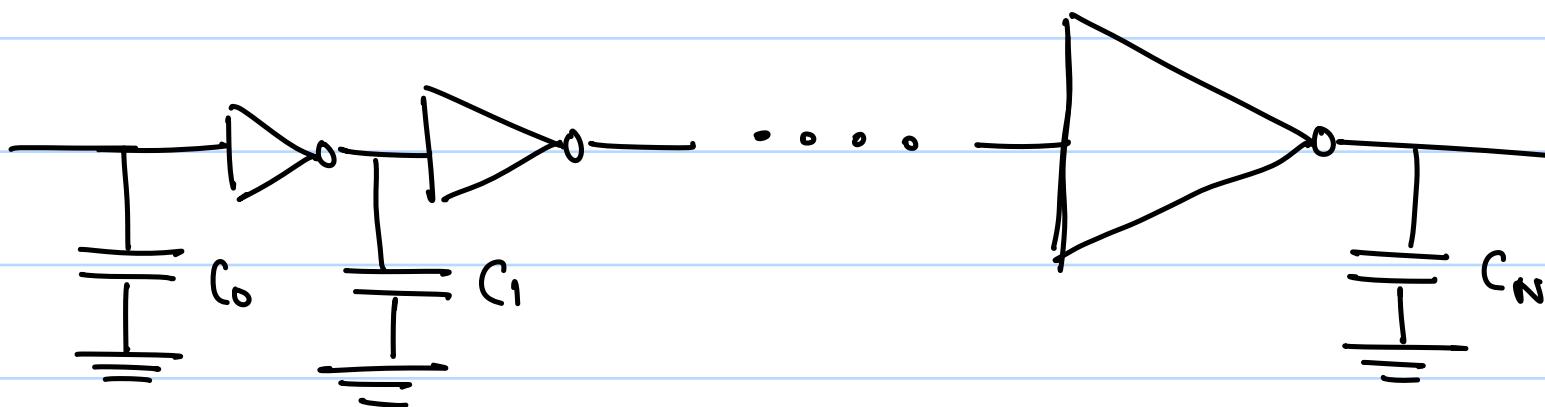


$$C_{out} = 2C_{in-inv}$$

$$\Rightarrow D_2 = \tau(2+p)$$



~ In an  $N$ -stage buffer chain,



$$D = D_1 + D_2 + D_3 + \dots + D_N$$

$$= \left( \left( \frac{C_1}{C_0} + \rho \right) + \left( \frac{C_2}{C_1} + \rho \right) + \dots + \left( \frac{C_N}{C_{N-1}} + \rho \right) \right) \tau$$

$$\Rightarrow D = \tau \left( \sum_{i=1}^N \frac{C_i}{C_{i-1}} + N\rho \right)$$

$$\Rightarrow D = \tau \left( \sum_{i=1}^N h_i + N\rho \right) \quad h_i - \text{Electrical Effort} = \frac{C_{\text{out}}}{C_{\text{in}}}$$

~ To optimize delay,

$$1) h_1 = h_2 = h_3 = \dots = h_N$$

$$\frac{SD}{SC_i} = \frac{S}{SC_i} \left( \frac{c_i}{c_{i-1}} + \beta + \frac{c_{i+1}}{c_i} + \beta \right) = 0$$

$$= \frac{1}{c_{i-1}} + c_{i+1} \left( \frac{-1}{c_i^2} \right) = 0$$

$$= \frac{1}{c_{i-1}} = \frac{c_{i+1}}{c_i^2}$$

$$= c_i = \sqrt{c_{i+1} c_{i-1}} \longrightarrow \text{For Min. delay}$$

$$h_i = \frac{c_i}{c_{i-1}} = \sqrt{\frac{c_{i+1}}{c_{i-1}}}$$

$$\Rightarrow h_1 = h_2 = h_3 = h_4 \dots \dots = h_N \quad (\text{On simplifying})$$

$$\text{Overall electrical effect } H = \frac{C_{ext}}{C_{in}}$$

$$\frac{C_N}{C_0} = \frac{c_1}{c_0} \times \frac{c_2}{c_1} \times \frac{c_3}{c_2} \times \dots$$

$$H = h_1 \times h_2 \times h_3 \times \dots$$

$$\Rightarrow h = H^{1/N}$$

$$D = \tau \left( \sum_{i=1}^N h_i + N_p \right)$$

$$= N \tau (h + \beta)$$

$$\Rightarrow D = N \tau (h^{1/N} + \beta)$$

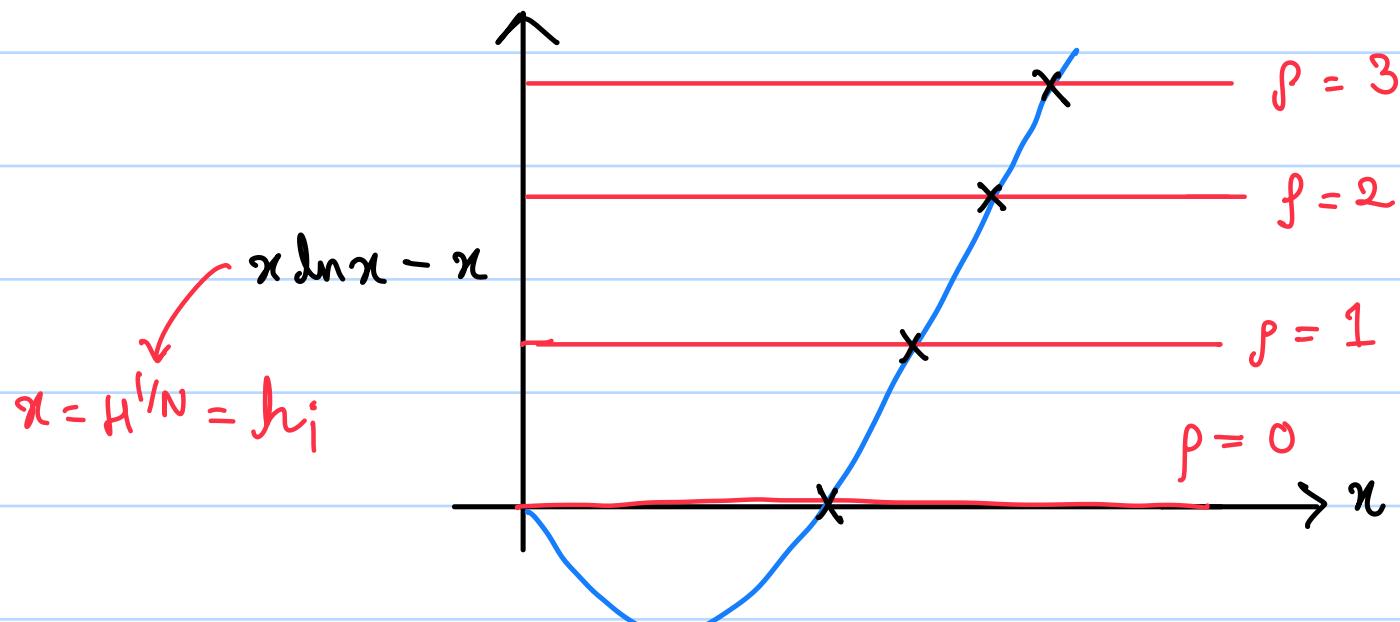
$$\text{Take } x = H^{1/N}, \Rightarrow \ln x = \frac{1}{N} \ln H \quad x = h_i$$

$$D(x) = \tau N(x + p)$$

$$\Rightarrow D(x) = \tau \frac{\ln H}{\ln x} (x + p)$$

$$D'(x) = \tau \frac{\ln H}{\ln x} \left[ \frac{x \ln x - (x + p)}{x \ln x} \right] = 0$$

$$\Rightarrow x \ln x - x = p$$



The intersection points of  $x$  give us the value of  $H^{1/N}$  and thereby, the value of  $N$  ( $H$  is known) needed for the tapered buffer.

$p$	$h_i$
0	2.73
1	3.59
2	4.31

Now for each inverter,

$$h_i = \frac{C_{out}}{C_{in}}$$

$$\Rightarrow C_{in} = \frac{C_{out}}{h_i}$$

Since  $C_{in} = WL C_{ox,n} + W C_{ov,p} = w [const]$ , we can determine the width of each inverter in the chain.

~ Newton Raphson Method to find  $h_i$  :-

Newton Raphson is an iterative method to find the zeros of any function

Let  $f(x)$  be our function and let the approximate zero be  $m$ , then,

$$m_{next} = m - \frac{f(m)}{f'(m)}$$

Eventually,  $f(m_{next})$  will start to converge to zero.

Here, our  $f(x) = x \ln x - x - p$

$$m_{next} = m - \frac{m \ln m - m - p}{-\ln m}$$

$$\Rightarrow m_{next} = \frac{m + p}{\ln m}$$

Example:  $H = \frac{C_L}{C_{in}} = 10^3$ ,  $\beta = 1 \text{ or } 2$  Find the no. of stages required for an optimized non-inverting buffer.

$$f(x) = x \ln x - x - p$$

$$\text{If } p = 1, x = 3.59$$

$$x = H^{1/N} \Rightarrow N = \frac{\ln H}{\ln x} = \frac{\ln(1000)}{\ln(3.59)} \approx 5.404$$

$$\Rightarrow N = 5.404$$

$$\begin{aligned} D &= N\tau(H^{1/N} + \beta) \\ &= 4\tau(1000^{1/4} + 1) \quad \text{or} \quad 6\tau(1000^{1/6} + 1) \\ &\approx \underline{26.5\tau} \qquad \qquad \qquad \approx \underline{\underline{24.9\tau}} \end{aligned}$$

$N = 6$  has lower delay than  $N = 4$

$\therefore$  Our buffer chain should be 6 stage with  $\chi_i = \underline{\underline{3.59}}$ .

$$\text{If } p = 2, x = 4.319$$

$$N = \frac{\ln H}{\ln x} = \frac{\ln(1000)}{\ln(4.319)} = 4.72$$

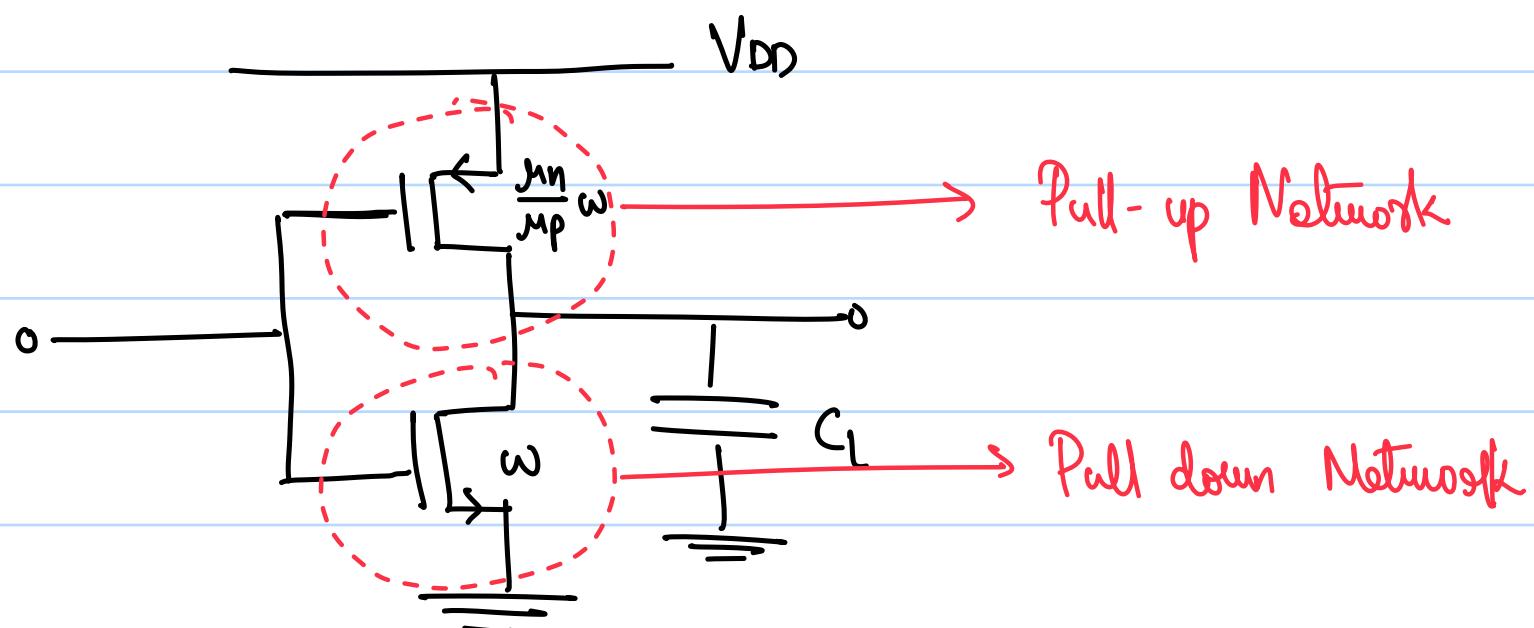
$$\begin{aligned}
 D &= N\tau(H^{1/N} + f) \\
 &= 4\tau(1000^{1/4} + 2) \quad \text{or} \quad 6\tau(1000^{1/6} + 2) \\
 &= 30.49\tau \quad = 30.97\tau
 \end{aligned}$$

$\Rightarrow$  4 stage is more optimized.

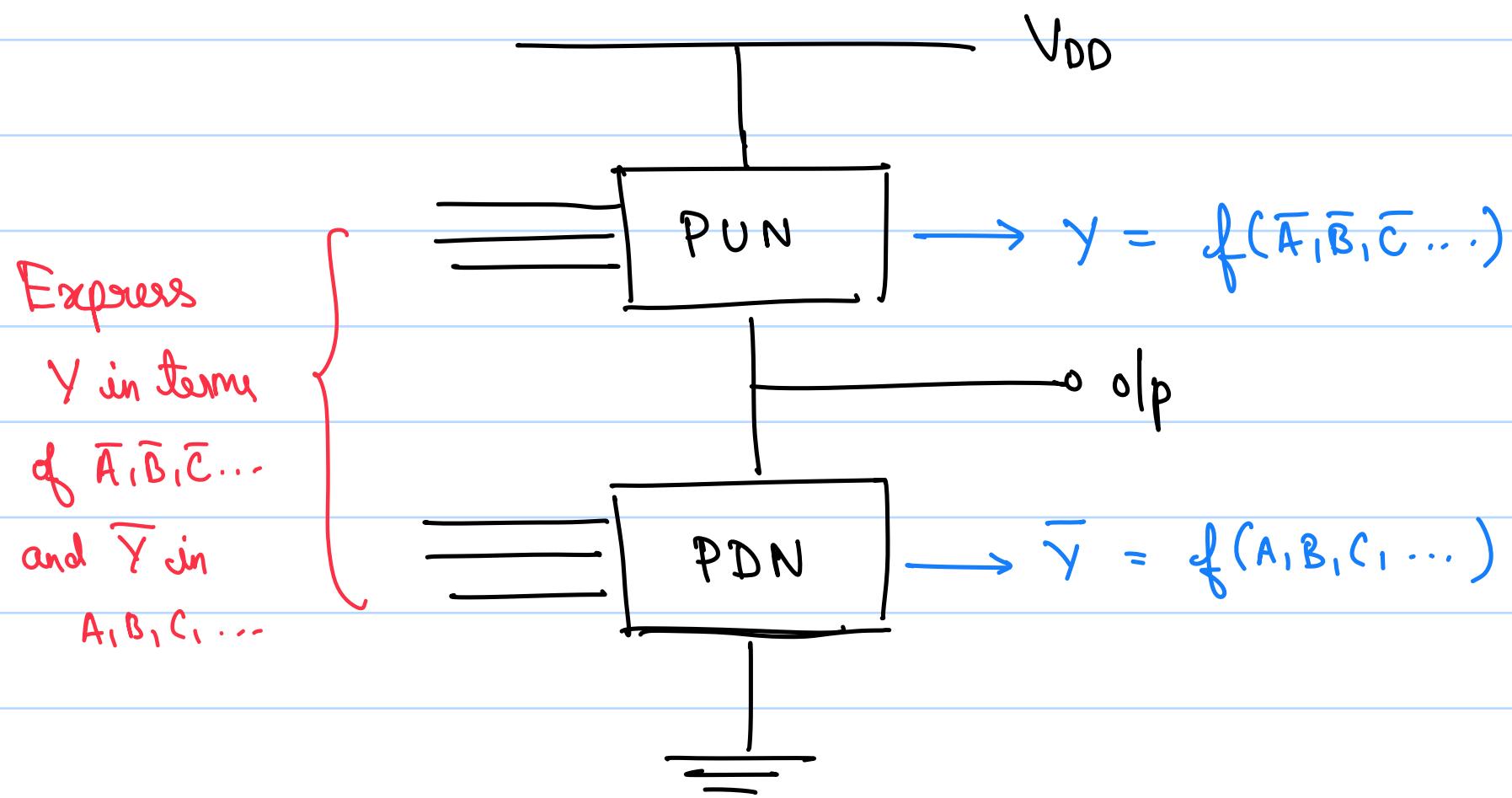
### \* Method Of Logical Effort :-

Ref: Sutherland | Sproull | Kaviraj  
Designing fast CMOS logic

- In our CMOS Inverter,



- If we want to implement the logical circuit  $y = \overline{A \cdot B}$ , since the inverter is the most basic design in CMOS logic, we try to equate the delay of any logic gate to that of an inverter.
- Any CMOS logic device is a combination of a pull up network of PMOS's and a pull down network of NMOS's.
- The binary operations of a logic circuit are designed as follows,

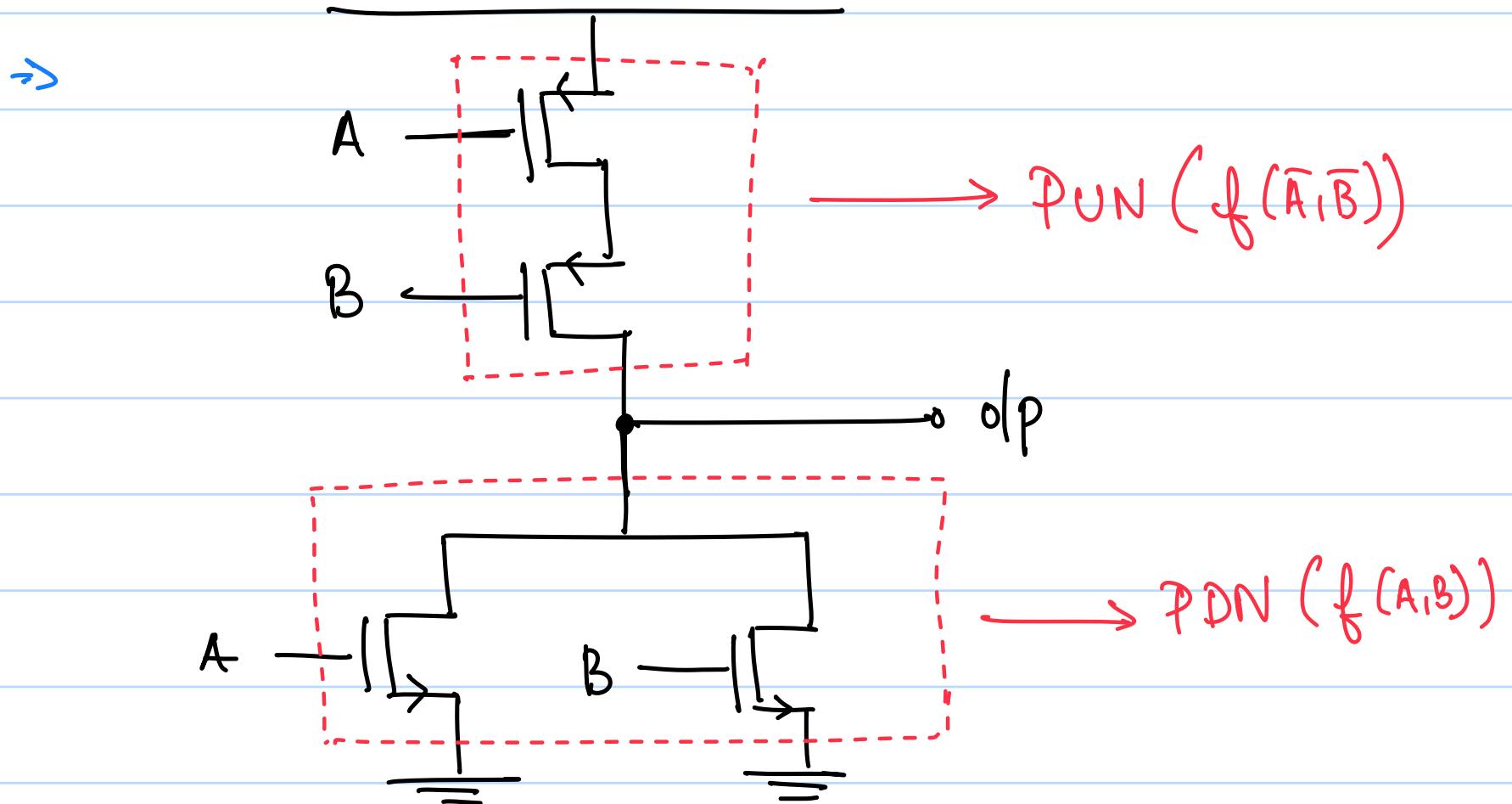


AND operation - MOSFETs in Series

OR operation - MOSFETs in Parallel

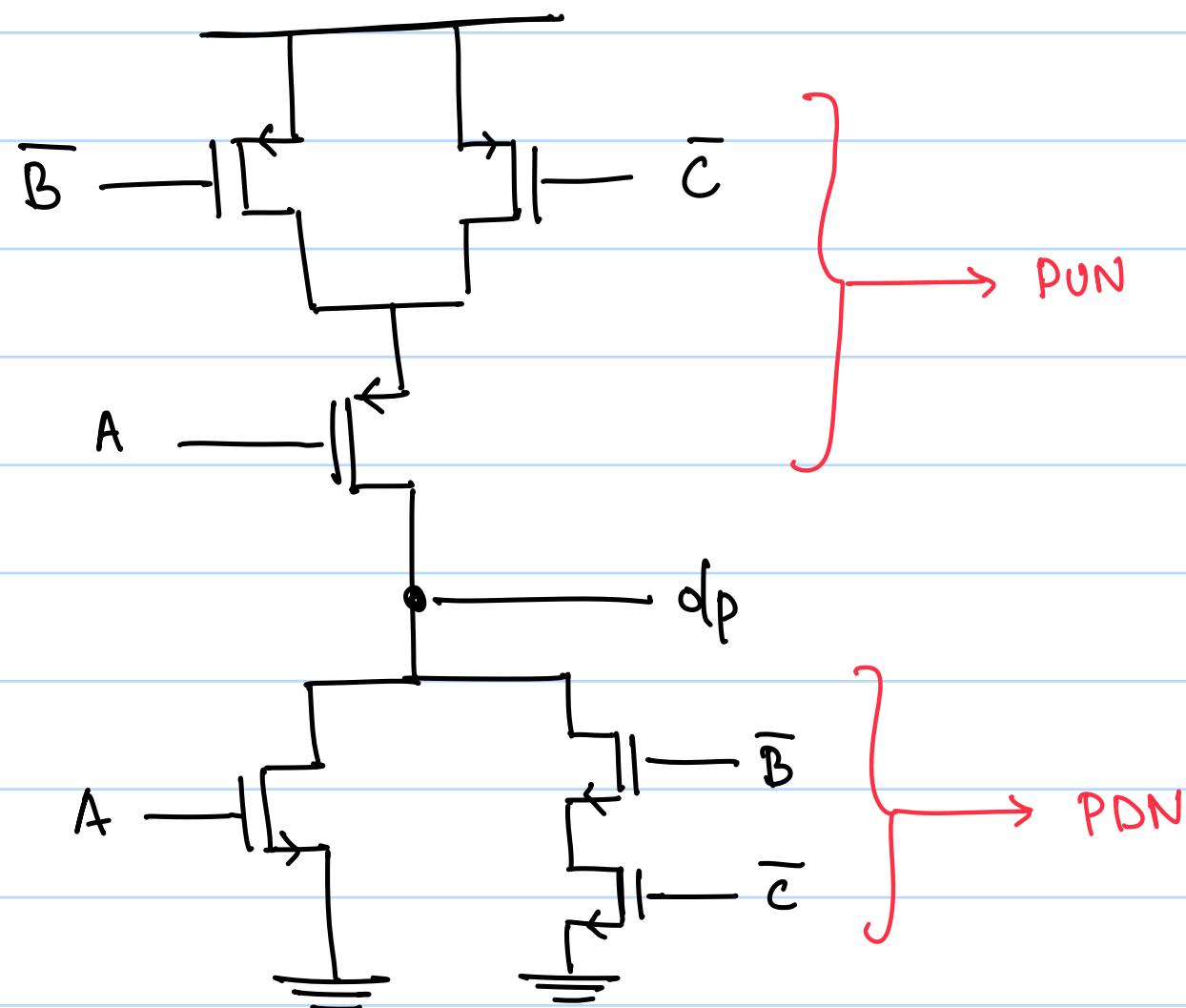
For logic  $Y = \overline{A+B}$   
 $\Rightarrow Y = \bar{A} \cdot \bar{B}, \bar{Y} = A + B$

(De Morgan's)



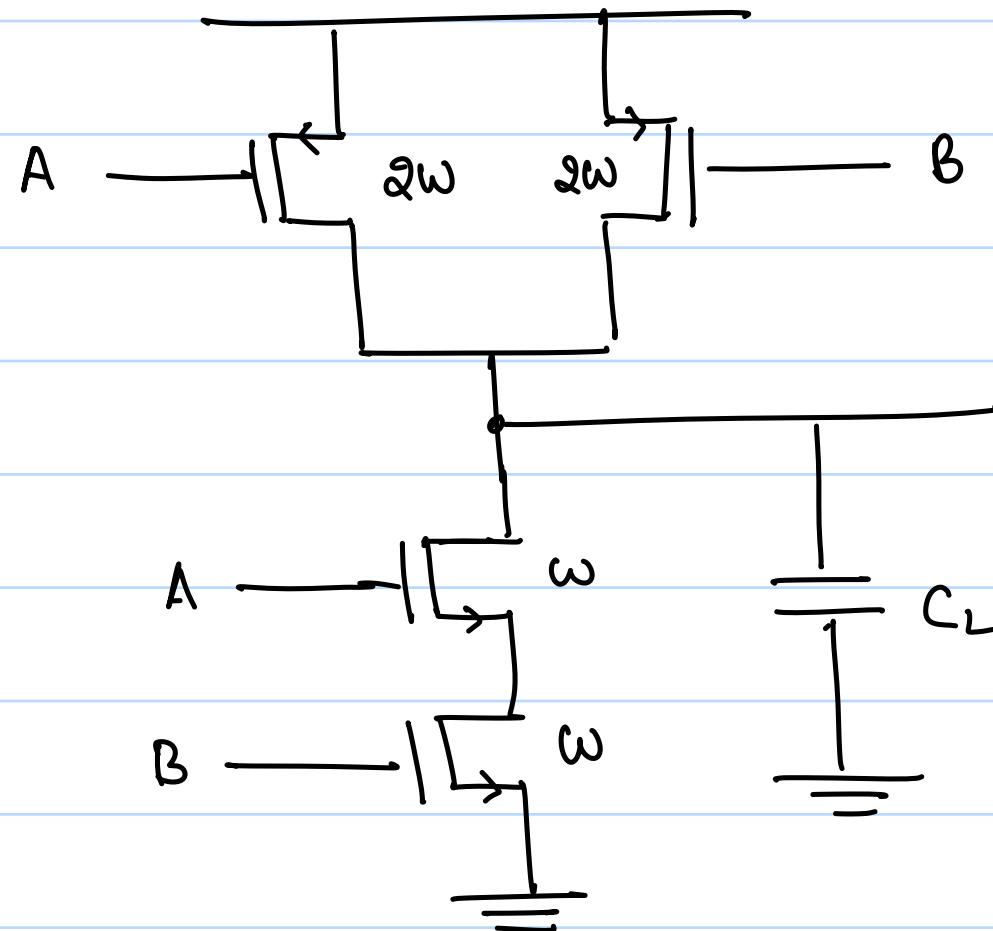
$$\text{For function } f = \bar{A}(B+C) \quad \bar{f} = A + (\bar{B} + \bar{C})$$

$$= \bar{A}(\bar{\bar{B}} + \bar{\bar{C}}) \quad = A + \bar{B}\bar{C}$$



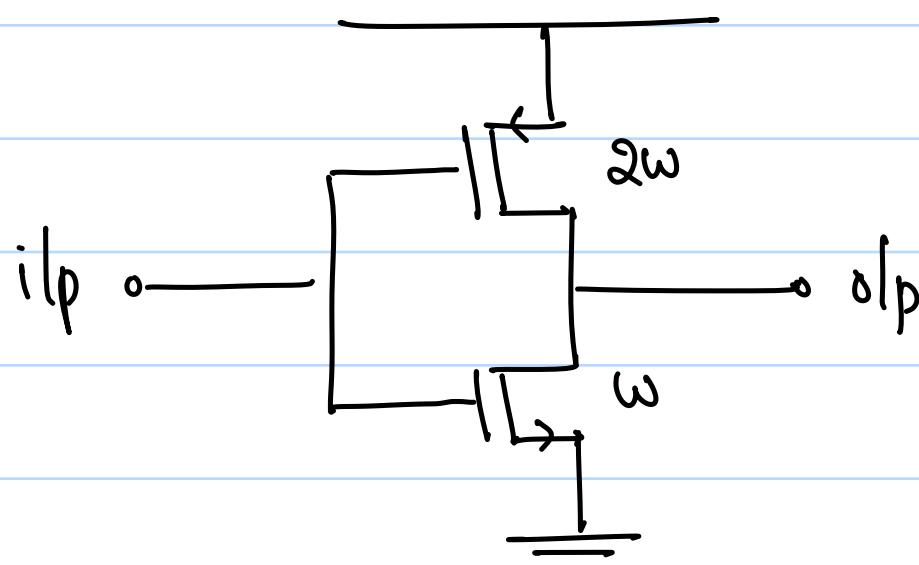
→ Delay of Combinational Circuits :-

- Take a NAND gate, with the following widths,



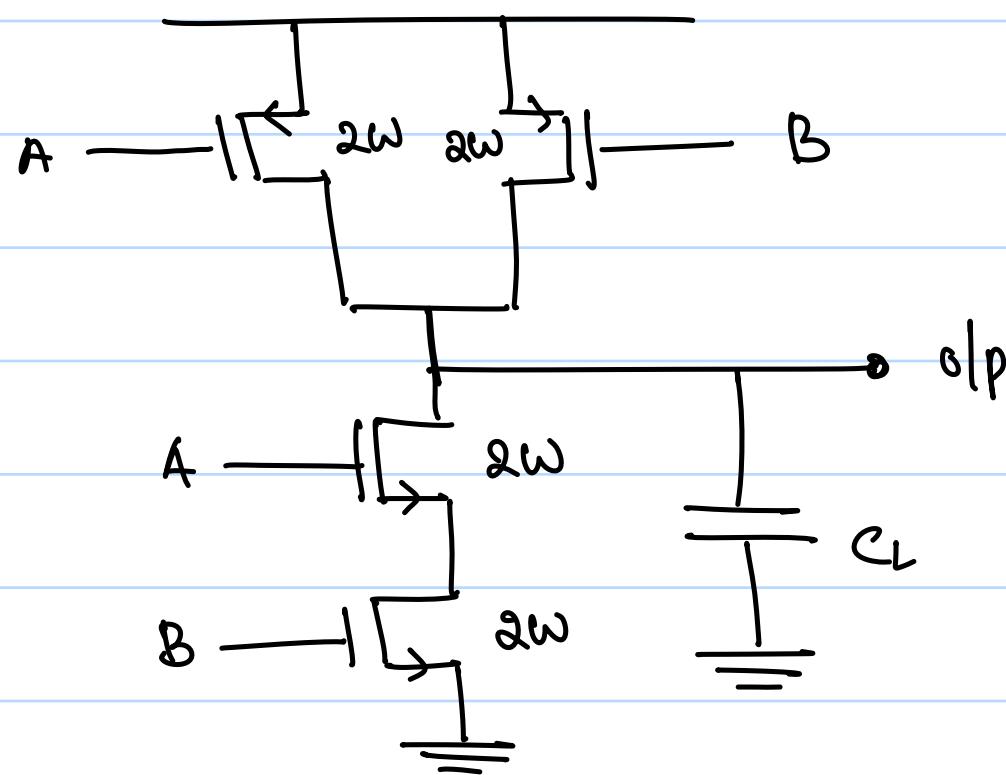
- We know that  $R_{on} \propto \frac{1}{w}$  and greater the  $R_{on}$ , lower the driving capability of the device, since o/p current will be smaller.

- This can be thought of as the below inverter, for each input



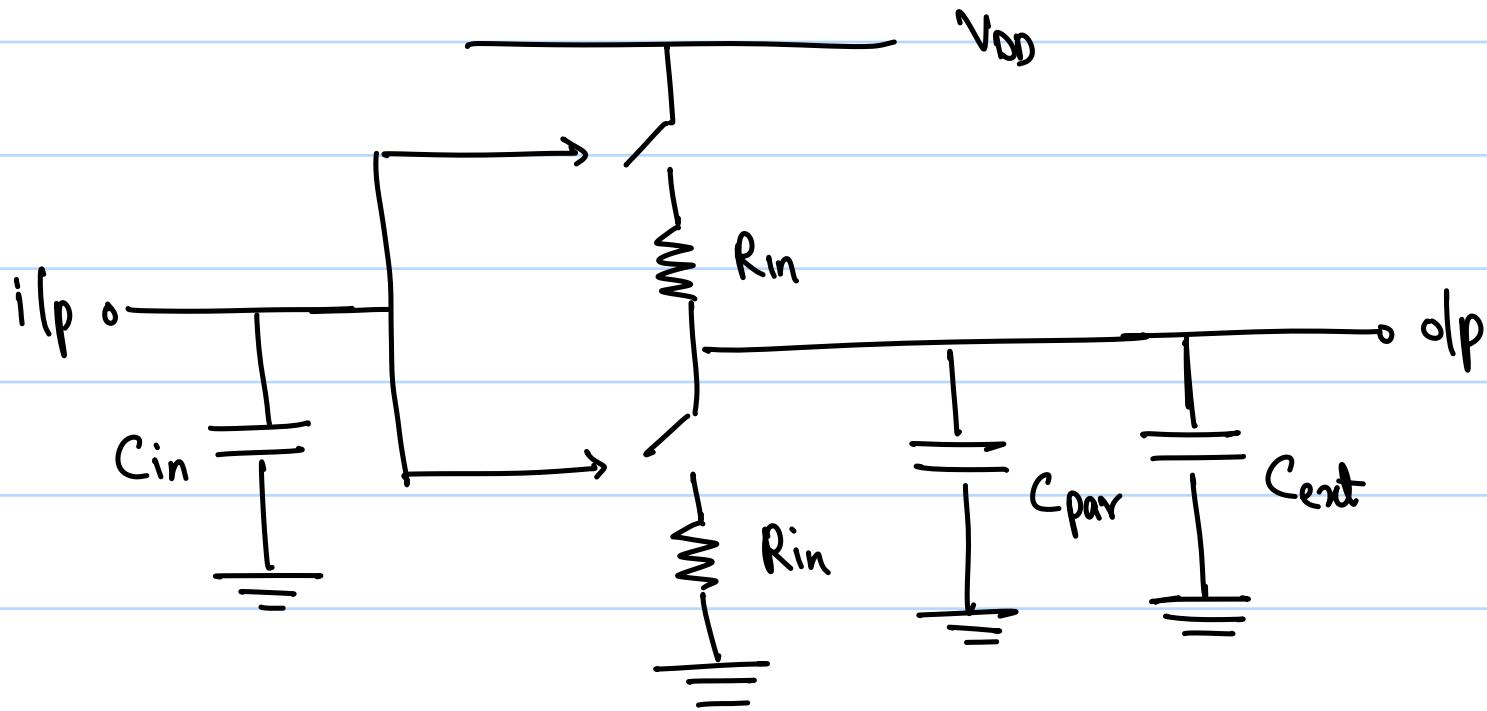
Each ilp in the NAND gate 'sees'  $3w$  width of MOSFETs. In the above inverter, the ilp also 'sees'  $3w$  width of MOSFETs.

- But in the NAND gate, when both the ilps are active, the PDN network's delay will be higher ( $2Rc$ ) than the reference inverter ( $Rc$ ). So to equalize delays with the reference inverter, we increase the NMOS widths to  $2w$ .



Here, the delays are similar to the reference inverter, but the widths 'seen' by the i/p's are higher  $\Rightarrow C_{in}$  is higher.

- In our reference minimum sized inverter,



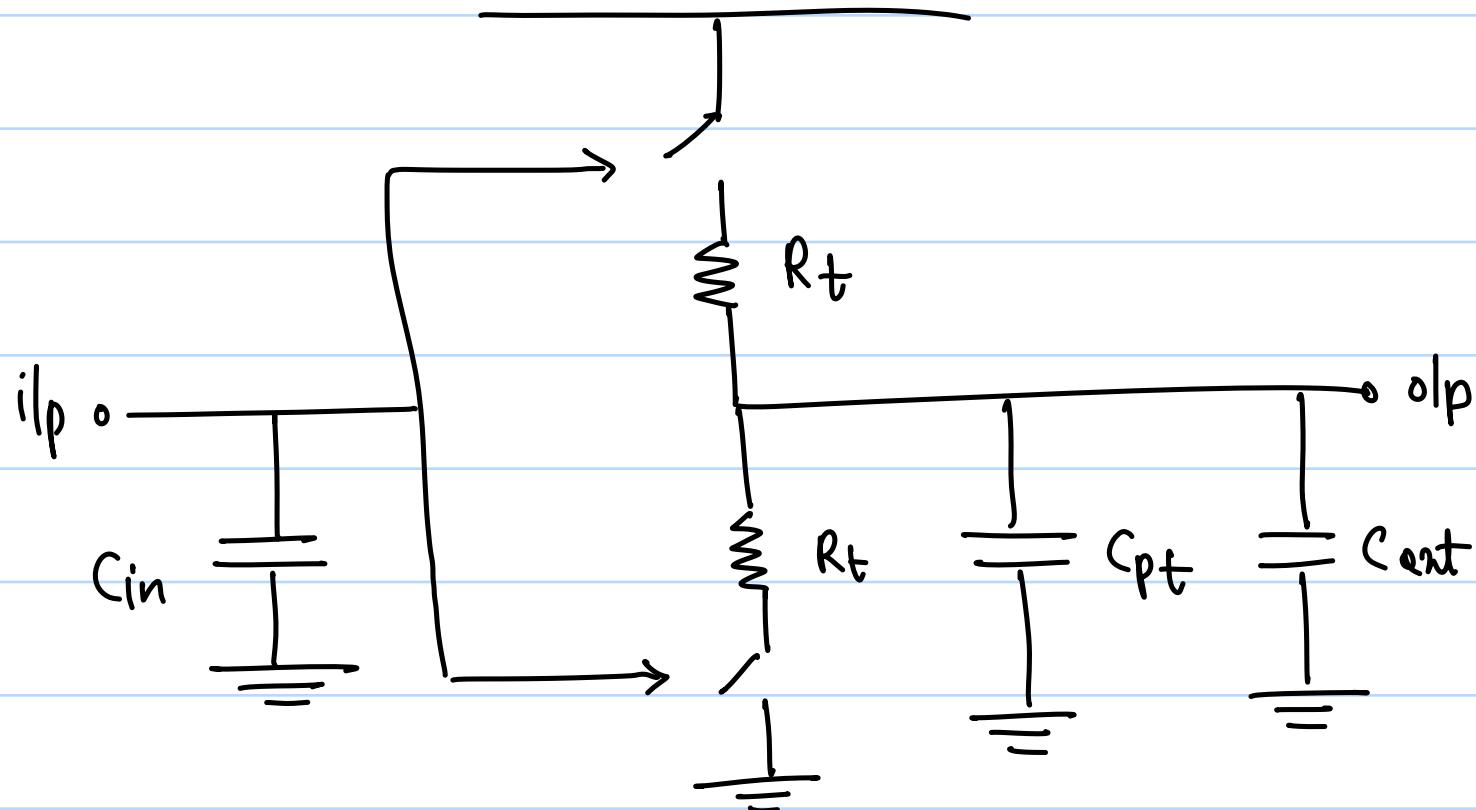
$$d_{inv} \propto R_{inv} (C_{par} + C_{ent})$$

$$\Rightarrow d_{inv} \propto R_{inv} C_{in} \left( \frac{C_{par}}{C_{in}} + \frac{C_{ent}}{C_{in}} \right)$$

$$\Rightarrow d_{inv} = \tau \left( \frac{C_{par}}{C_{in}} + \frac{C_{ent}}{C_{in}} \right)$$

$$\Rightarrow d_{inv} = \tau (p+1) \quad \left[ \frac{C_{par}}{C_{in}} = p, \frac{C_{ent}}{C_{in}} = 1 \right]$$

- To understand the delay of any CMOS device, we use the following model of the PUN and PDN, for each single input.



$$d_{\text{logic}} \propto R_t (C_{pt} + C_{ent})$$

$$= d_{\text{logic}} \propto R_t C_t \left( \frac{C_{pt}}{C_t} + \frac{C_{ent}}{C_t} \right)$$

$$= d_{\text{logic}} \propto R_{\text{inv}} C_{\text{inv}} \left( \frac{R_t C_t}{R_{\text{inv}} C_{\text{inv}}} \cdot \frac{C_{pt}}{C_t} + \frac{R_t C_t}{R_{\text{inv}} C_{\text{inv}}} \cdot \frac{C_{ent}}{C_t} \right)$$

$$= d_{\text{logic}} \propto R_{\text{inv}} C_{\text{inv}} \left( \frac{R_t C_t}{R_{\text{inv}} C_{\text{inv}}} \cdot \frac{C_{ent}}{C_t} + \frac{R_t C_{pt}}{R_{\text{inv}} C_{\text{inv}}} \right)$$

$$\Rightarrow d_{\text{logic}} = \tau \left( \underbrace{\frac{R_t C_t}{R_{\text{inv}} C_{\text{inv}}} h}_{g} + \rho_{\text{logic}} \right) \quad [ \tau \text{ is the same as in inverter. } ]$$

$$\Rightarrow d_{\text{logic}} = \tau (gh + \rho_{\text{logic}})$$

Where  $g = \text{Logical Effort} = \frac{R_t C_t}{R_{\text{inv}} C_{\text{inv}}}$

$\downarrow$   
 $\tau$  is the same since it is a technology-dependent parameter.

The term  $gh$  is denoted as the stage effort.

- In a multi-stage device,  $d = \tau(gh + f)$  b - branching effort
- Logical Effort 'g':
- We know that if my reference inverter is seeing  $3W$  switch of MOSFET, my NAND gate must be seeing  $4W$  switch of MOSFET.
- $g = \frac{R_t C_t}{R_{inv} C_{inv}}$

In our NAND gate example, we tried to make  $R_t = R_{inv}$

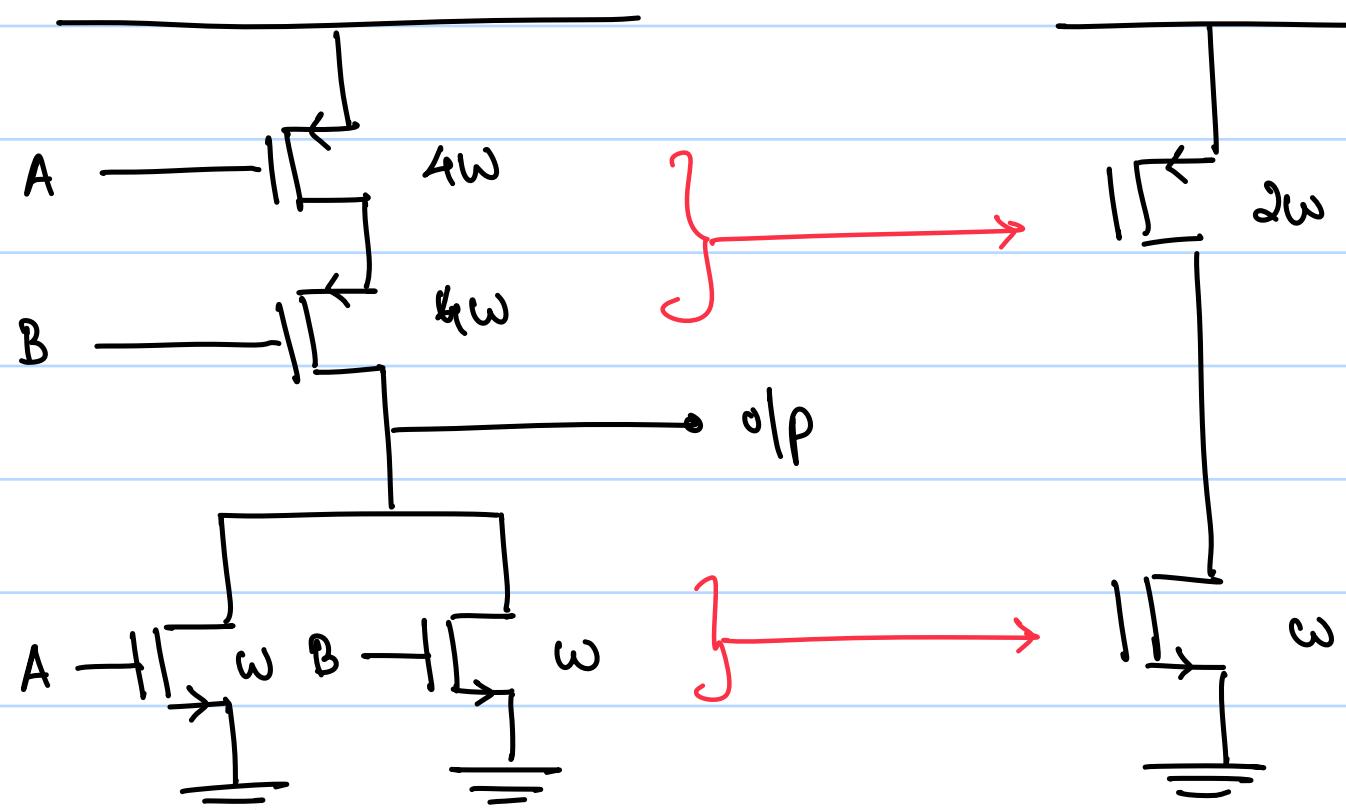
$$\Rightarrow g = \frac{C_t}{C_{inv}}, \quad C_t \propto 4W \text{ and } C_{inv} \propto 3W,$$

$$\Rightarrow g = \underline{\underline{\frac{4}{3}}}$$

$\therefore g$  can be thought of as the ratio required between the CMOS device and the reference inverter, in order to have similar driving capabilities.

- The widths must be calculated for worst case scenario, when only one of the ilp's is active. (lessen driving current through the device)

- For a NOR gate,



$$g = \frac{4w + w}{3w} = \underline{\underline{\frac{5}{3}}}$$

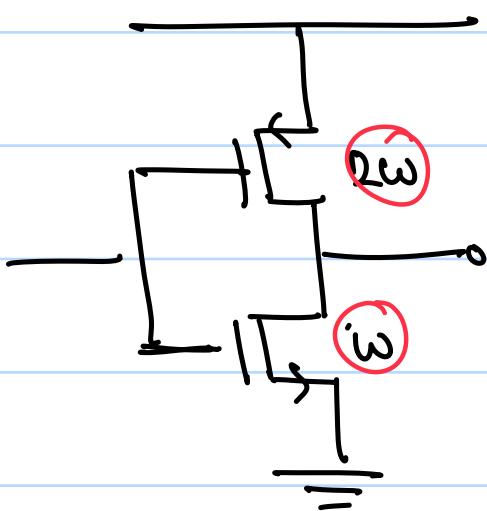
- For n-ilp NAND gate,  $g = \frac{n+2}{3}$
- " " NOR gate,  $g = \frac{2n+1}{3}$

- The parasitic delay is proportional to the width seen at the o/p

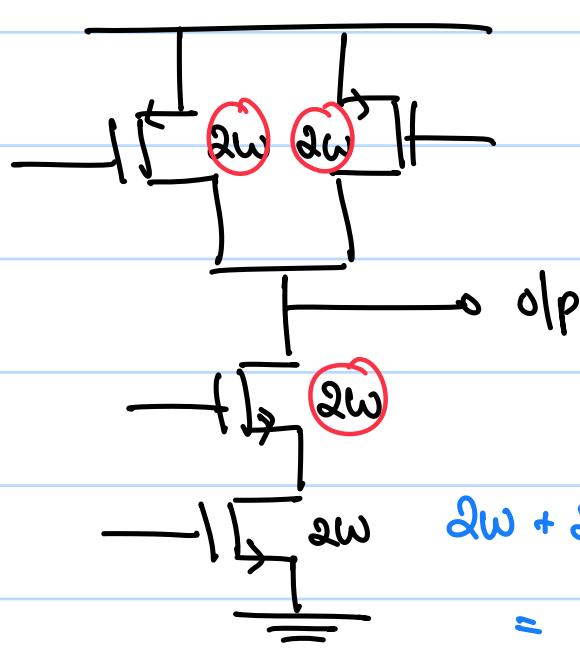
$$\tau_{\text{inv}} \propto 3w$$

$$\tau_{\text{NAND}} \propto 6w$$

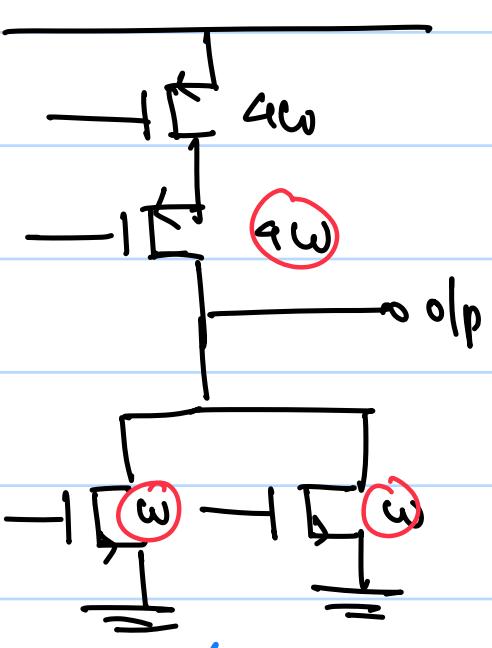
$$\tau_{\text{NOR}} \propto 6w$$



$$2w + w = 3w$$



$$2w + 2w + 2w = 6w$$

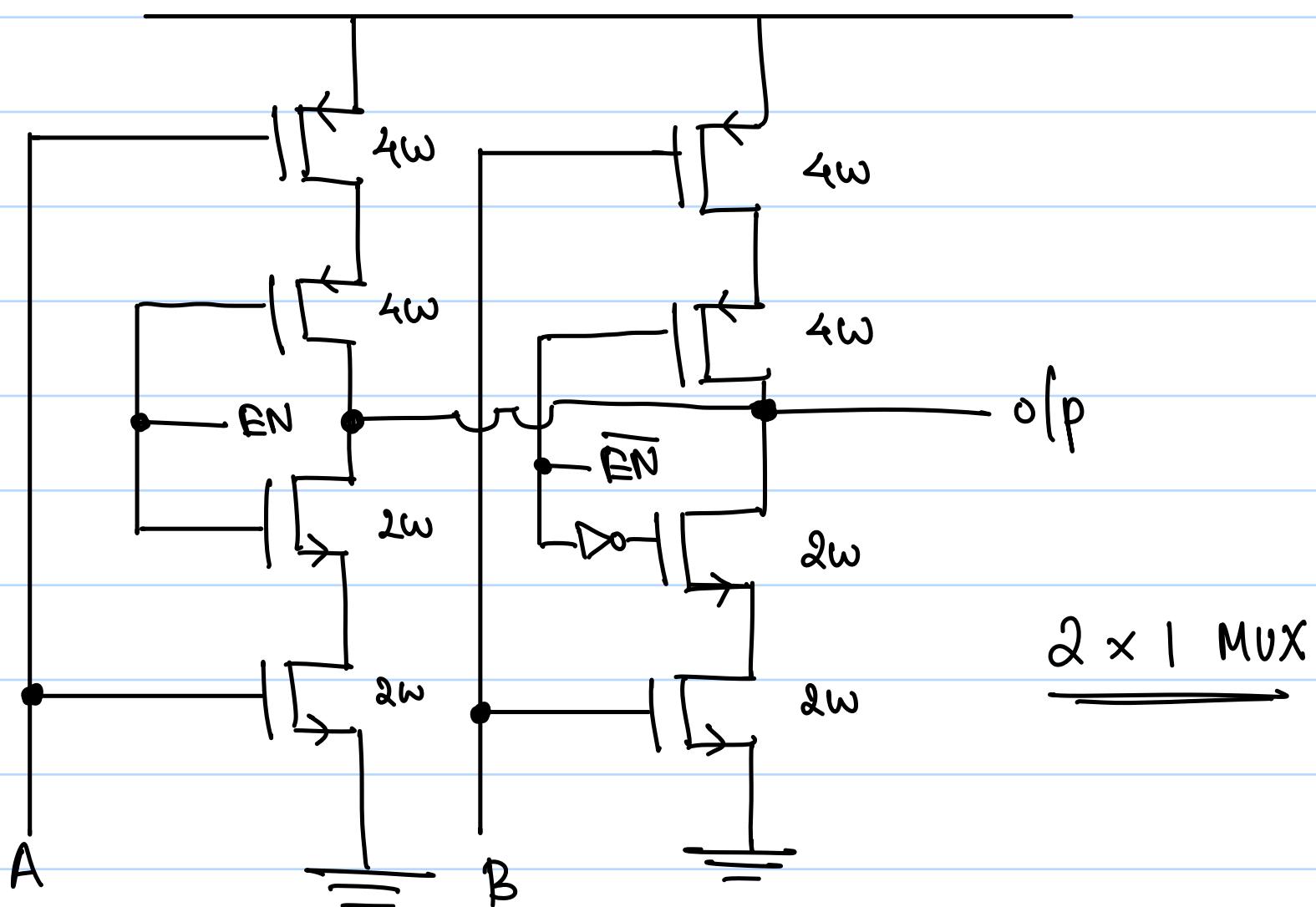


$$4w + w + w = 6w$$

$$f_{N-NOR} = N \cdot f_{inv}$$

$$f_{N-NAND} =$$

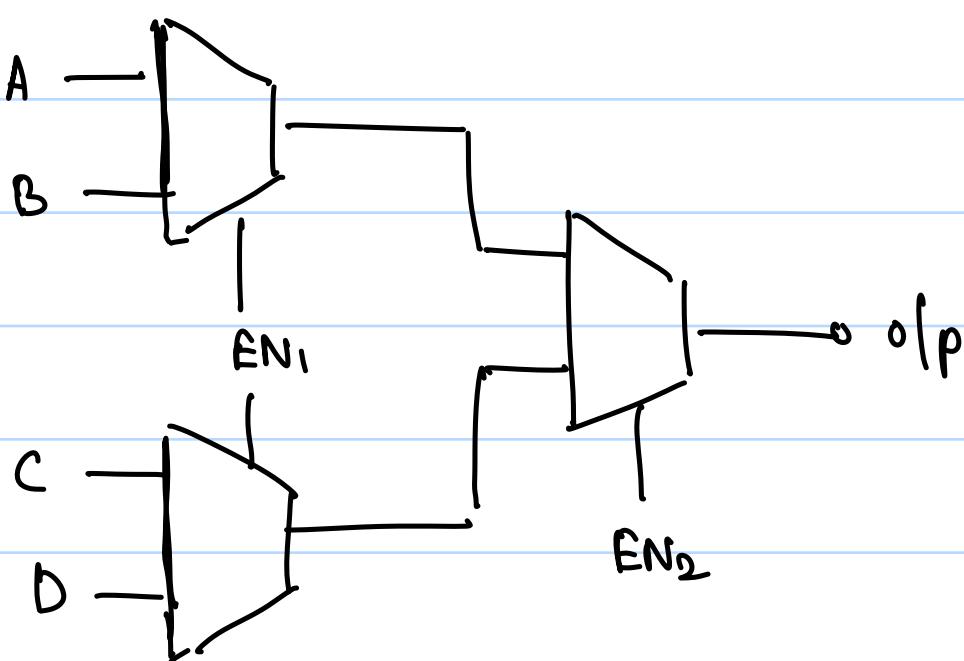
o



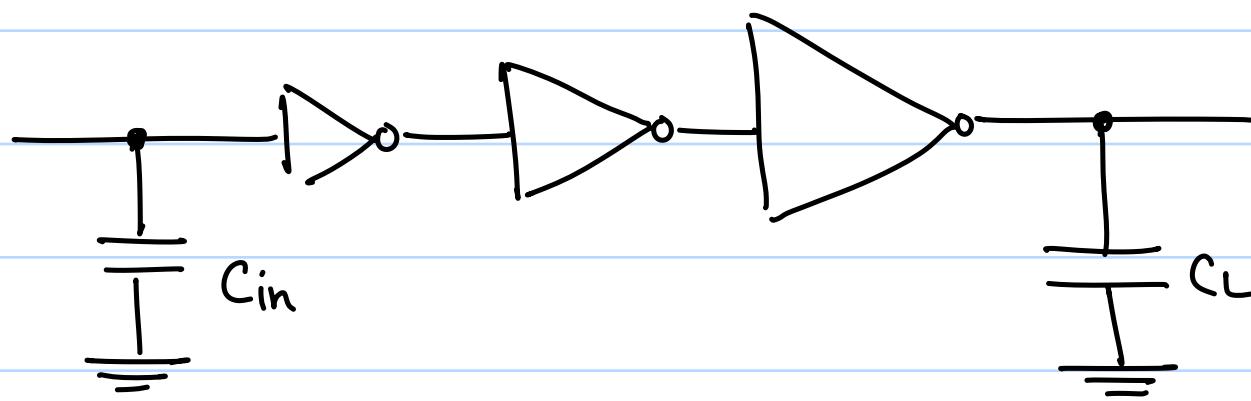
$$g_{2 \times 1 \text{ MUX}} = \frac{4w + 4w}{2w + 2w} = \underline{\underline{2}}$$

$$f_{2 \times 1 \text{ MUX}} \propto (4w + 2w)2 = 12w = 4f_{inv}$$

o We can make a 4x1 MUX by,



Recap:



- $\frac{K\tau}{c} = \text{const} \Rightarrow \tau \propto \frac{c}{K}$

- $d_i = \tau(h_i + \rho) = \tau \left( \frac{\text{Const}}{C_{in-i}} + \frac{C_{self}}{C_{in}} \right)$   
 $= \tau \left( \frac{C_{in-inv2}}{C_{in-inv1}} + \rho_{inv} \right)$

- $C_{in-inv} = C_{GDP} + C_{ADn} + C_{DBn} + C_{DBp} \propto C_{ox}WL$

- $D = N\tau(H^{1/N} + \rho), H = \frac{C_L}{C_{in}}$

- To minimize delay,  $h_1 = h_2 = h_3 \dots = h_N = H^{1/N}$ ,

$h_i$  is the solution of  $x \ln x - x = \rho \rightarrow h_i = \frac{w_{i+1}}{w_i}$

$$N = \ln H / \ln x = \ln H / \ln h_i$$

- If  $N \uparrow, H^{1/N} \downarrow \Rightarrow h_i \downarrow$ , but this decrease is not shown in delay due to the dominance of the  $N\rho$  term.

$$D = N(H^{1/N} + \rho)\tau$$

- For multi-input combinational circuits, we analyze the delay by studying the critical path of any one of the inputs.

Critical path - path of longest delay.

- $D = (gh + p_{\text{logic}}) \tau$  All wrt to a reference min.sized inverter.

$$g = \frac{R_t C_t}{R_{\text{inv}} C_{\text{inv}}} \rightarrow \text{keep } R_t = R_{\text{inv}} \text{ and find ratio of width}$$

to get  $C_t / C_{\text{inv}}$ .

$$h = \frac{C_{\text{ext}}}{C_t}, p_{\text{logic}} = \frac{R_t C_{\text{pt}}}{R_{\text{inv}} C_{\text{inv}}} \rightarrow \text{Proportional to the width}$$

"seen" at the op when  $R_t = R_{\text{inv}}$

- Controlling delay is important when designing sequential clock-based circuits.

- $D = (gh + p_{\text{logic}}) \tau$

$\underbrace{\hspace{10em}}$  numbers       $\downarrow$  time unit

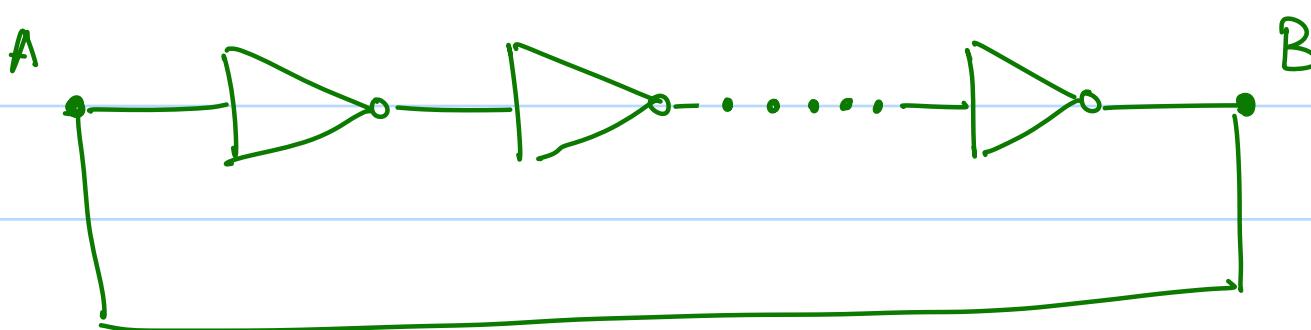
$g$  depends on the topology of the circuit, ie, how the MOSFETs are connected.

- If we want to drive the same  $C_{\text{ext}}$  as the inverter, with the same delay as the inverter, then  $C_{\text{in}} = g \times C_{\text{in-inv}}$

- If we want to keep  $C_{in} = C_{in-inv}$ , then the delay of the circuit will be increased.

- Overall logical effort =  $\sum$  logical effort of each flip-flop

Example: There is an  $N$ -stage Ring Oscillator, where  $N$  is odd and the inverters are identical.



Find the frequency of the ring oscillator.

$$d_i = \tau(1 + \rho)$$

$$\Rightarrow D = \sum_{i=1}^N \tau(1 + \rho) = (1 + \rho) N \tau$$

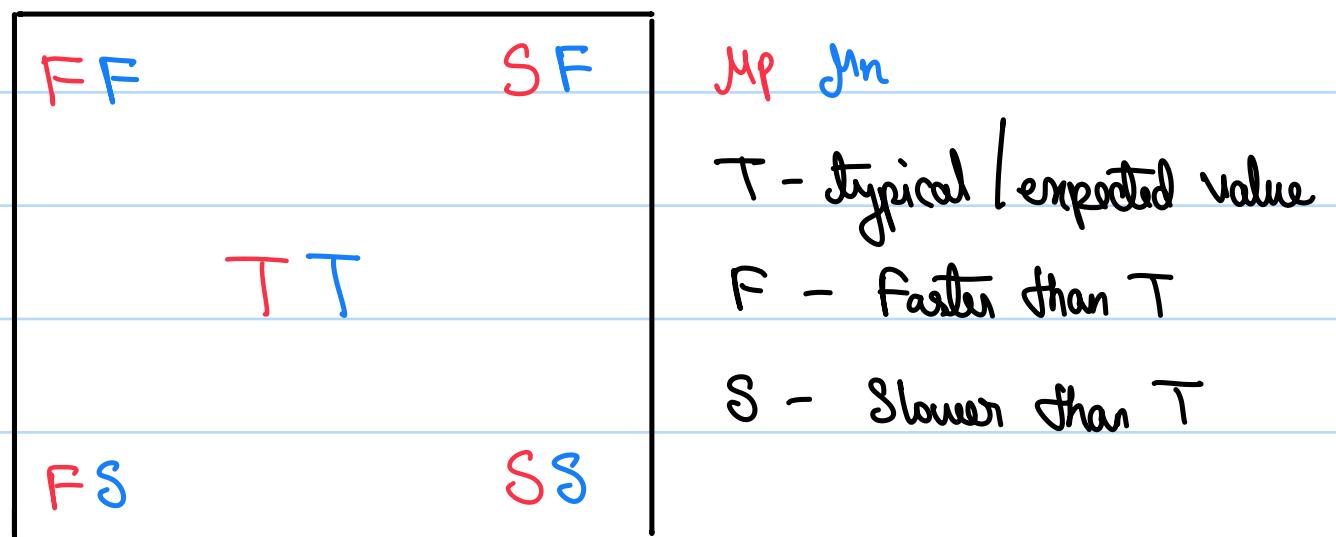
$$\Rightarrow f_{osc} = \frac{1}{T} = \frac{1}{2D} = \frac{1}{2N\tau(1 + \rho)} = \frac{1}{2Nd_i}$$

Ring Oscillator is used for parameter extraction, to get the delay of each inverter.

Large  $N$  is used since at low  $N$ , technology parameters will start to affect the frequency measured.

## ◦ Process Corners:

- Due to fabrication inefficiencies and inaccuracies, the  $\mu_n$  and  $\mu_p$  of the CMOS used may change, irrespective of the widths used.

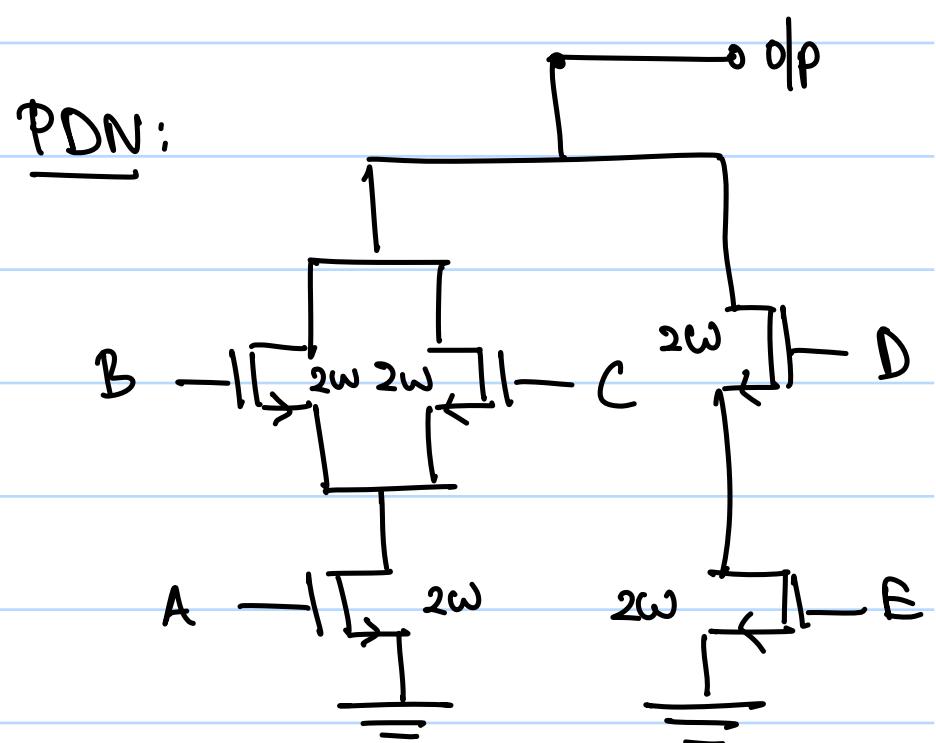
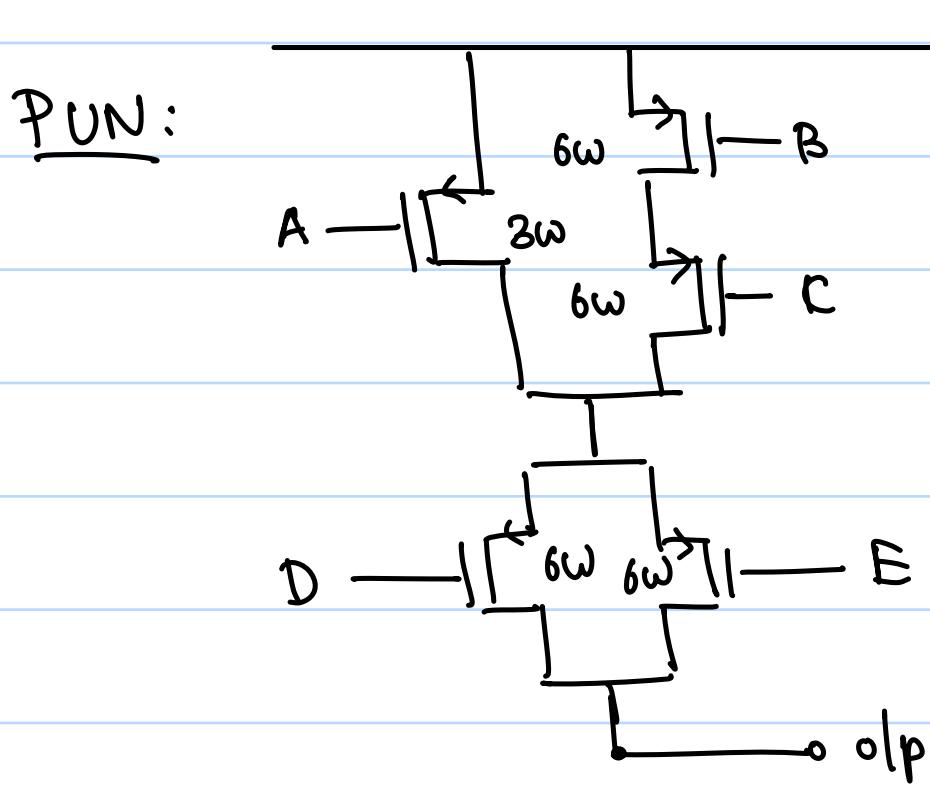


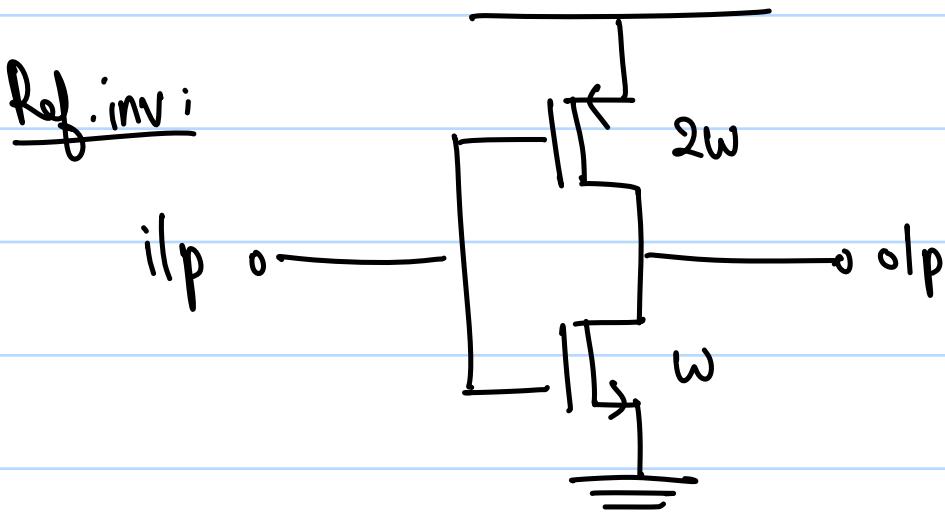
- Depending on the value of  $M_n, M_p$  measured, the CMOS is placed into one of the 5 regions of the above grid.

Example: Find the logical effect of each input in  $y = \overline{A(B+C)} + DE$

$$Y = \overline{A(B+C) + DE} = (\bar{A} + \bar{B} \cdot \bar{C})(\bar{D} + \bar{E})$$

$$\bar{Y} = A(B+C) + DE$$





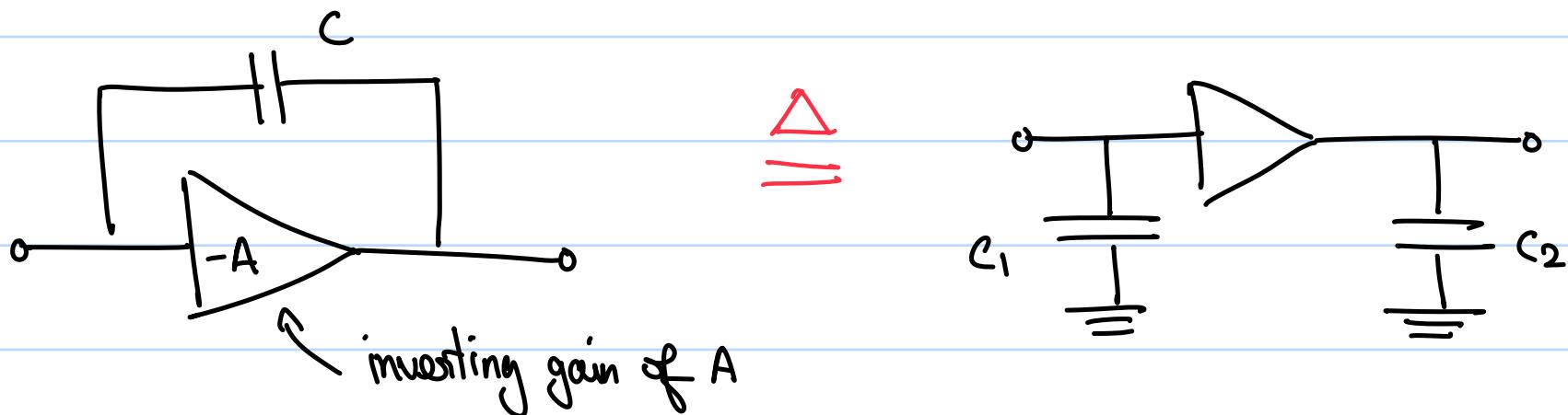
$$g_A = \frac{2w + 2w}{3w} = \frac{5}{3} \quad g_C = \frac{6w + 2w}{3w} = \frac{8}{3}$$

$$g_B = \frac{6w + 2w}{3w} = \frac{8}{3} \quad g_D = \frac{6w + 2w}{3w} = \frac{8}{3}$$

$$g_E = \frac{6w + 2w}{3w} = \frac{8}{3}$$

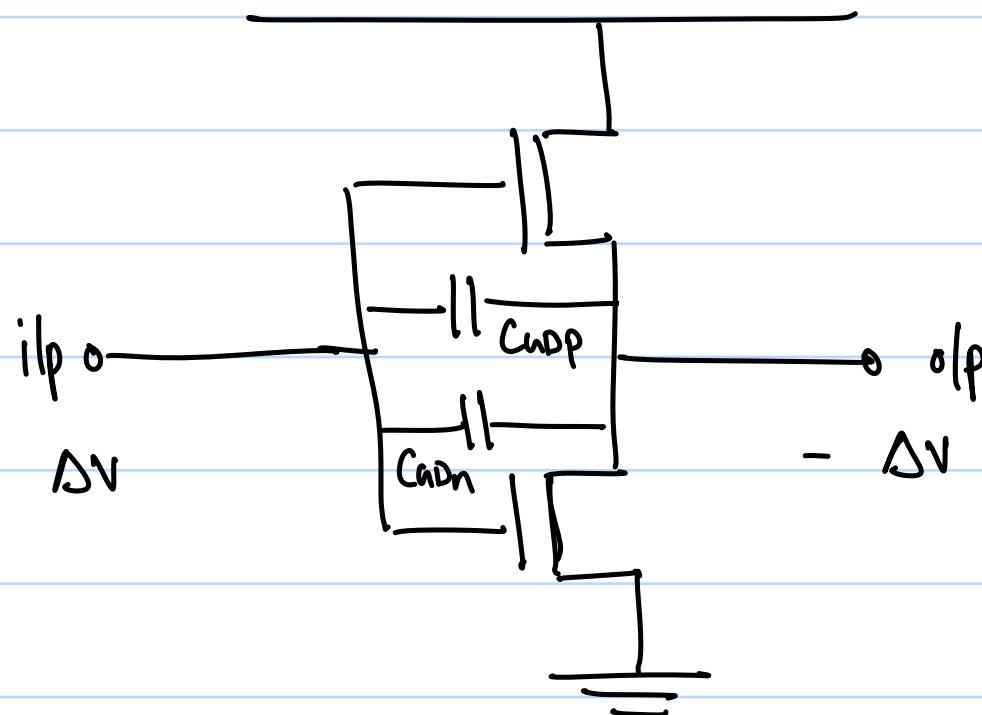
$f_{\text{logic}} \propto 18w$

Note: Miller Effect:



$$C_1 = C(1+A), \quad C_2 = C\left(1+\frac{1}{A}\right)$$

Applying Miller's effect in CMOS inverter,



$$A = \frac{\Delta V - (-\Delta V)}{\Delta V} = 2$$

$$\Rightarrow C_{ilp} = 2C_{aoN} + 2C_{aoP} \quad \left. \begin{array}{l} \\ \end{array} \right\} + \text{other terms}$$
$$C_{olp} = \frac{3}{2}C_{aoN} + \frac{3}{2}C_{aoP} \quad \left. \begin{array}{l} \\ \end{array} \right\} + \text{other terms}$$

→ Upcoming Topics:-

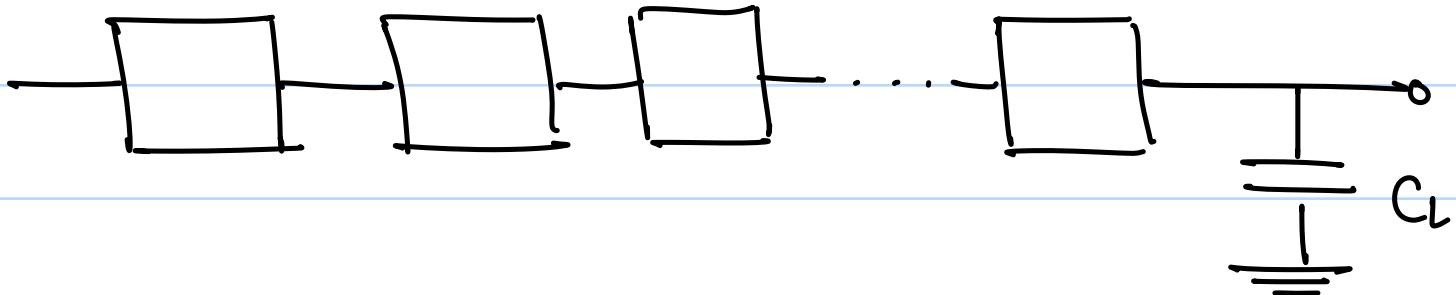
1) Logic Optimization

2) Other logic types : P-NMOS, PTL

3) Sequential Circuit fundamentals : Timing Analysis, Latches, Flip flops, memory.

→ Multi Stage Logic :-

- Suppose we have realised the topology for a certain logic function, the algorithm to minimize delay is as follows.



$$D_i = (j_i + p_i) \tau, \quad j_i = g_i \cdot b_i \cdot h_i$$

$g_i$  - Logical effort of  $i$

$h_i$  - Electrical effort of  $i$

$b_i$  - Branching effort of  $i$  =  $\frac{\text{Con-the-path} + \text{Off-the-path}}{\text{Con-the-path}}$

$$D = D_1 + D_2 + D_3 \dots = \tau \left( \sum_i f_i + \sum_i p_i \right)$$

- The total delay is minimized when  $f_1 = f_2 = f_3 \dots = f_i$

- However, the delay of each gate won't be the same as like in the buffer chain design, since  $f_i$  is not equal  $f_i$ .

Define  $G_i = \prod_i g_i$ ,  $H = \prod_i h_i$ ,  $B = \prod_i b_i$ , then

$$\Rightarrow GBH = f^N = F \rightarrow \text{path effect}$$

↓ stage effect

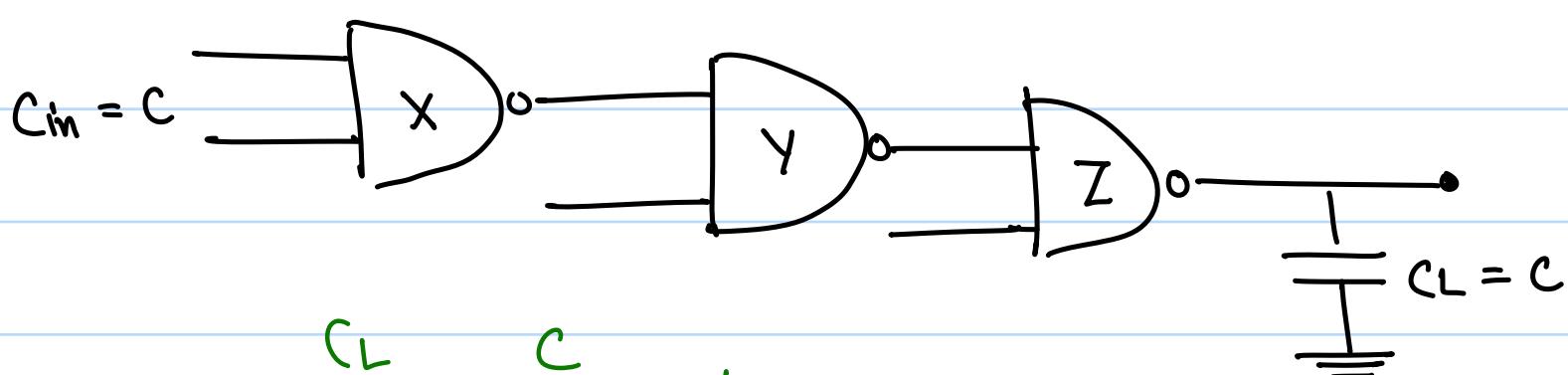
$$\Rightarrow D = (\sum f + \sum p_i) \tau$$

$$= (\hat{Nf} + \sum p_i) \tau$$

$$\Rightarrow D = (N^{1/N} + \sum p_i) \tau$$

$$\Rightarrow h_i = \frac{F^{1/N}}{g_{ibi}} \longrightarrow \text{gives us the size for each gate.}$$

Example:



Optimize delay across  
this cascade of  
NAND gates.

$$H = \frac{C_L}{C_{in}} = \frac{C}{C} = 1$$

$$G = \prod g_i = \frac{4}{3} \times \frac{4}{3} \times \frac{4}{3} = \frac{64}{27}$$

$$B = 1$$

$$\Rightarrow F = \frac{64}{27}, \quad f = \left(\frac{64}{27}\right)^{1/3} = \underline{\underline{\frac{4}{3}}}$$

$$h_i = \frac{4/3}{4/3 \cdot 1} = 1 + i \Rightarrow \text{All the gates have the same width}$$

$$D = (N^{1/N} + \sum p_i) \tau$$

$$= (3(\frac{4}{3}) + 6g_{inv}) \tau$$

$$\Rightarrow D = (4 + 6g_{inv}) \tau$$

Suppose  $C_L = 8C$

$$H = \frac{C_L}{C_{in}} = 8$$

$$\Rightarrow F = 8 \times \frac{64}{27} \times 1 = \frac{512}{27} \Rightarrow f = \left(\frac{512}{27}\right)^{1/3} = \frac{8}{3}$$

$$h_i = \frac{\frac{8}{3}}{\frac{4}{3} \times 1} = 2 \Rightarrow \omega_y = 2\omega_x, \omega_z = 4\omega_x$$

$\hookrightarrow$  ref. inverter size

$$D = \left(3 \times \frac{8}{3} + \sum p_i\right) \tau = \left(8 + 6p_{inv}\right) \tau$$

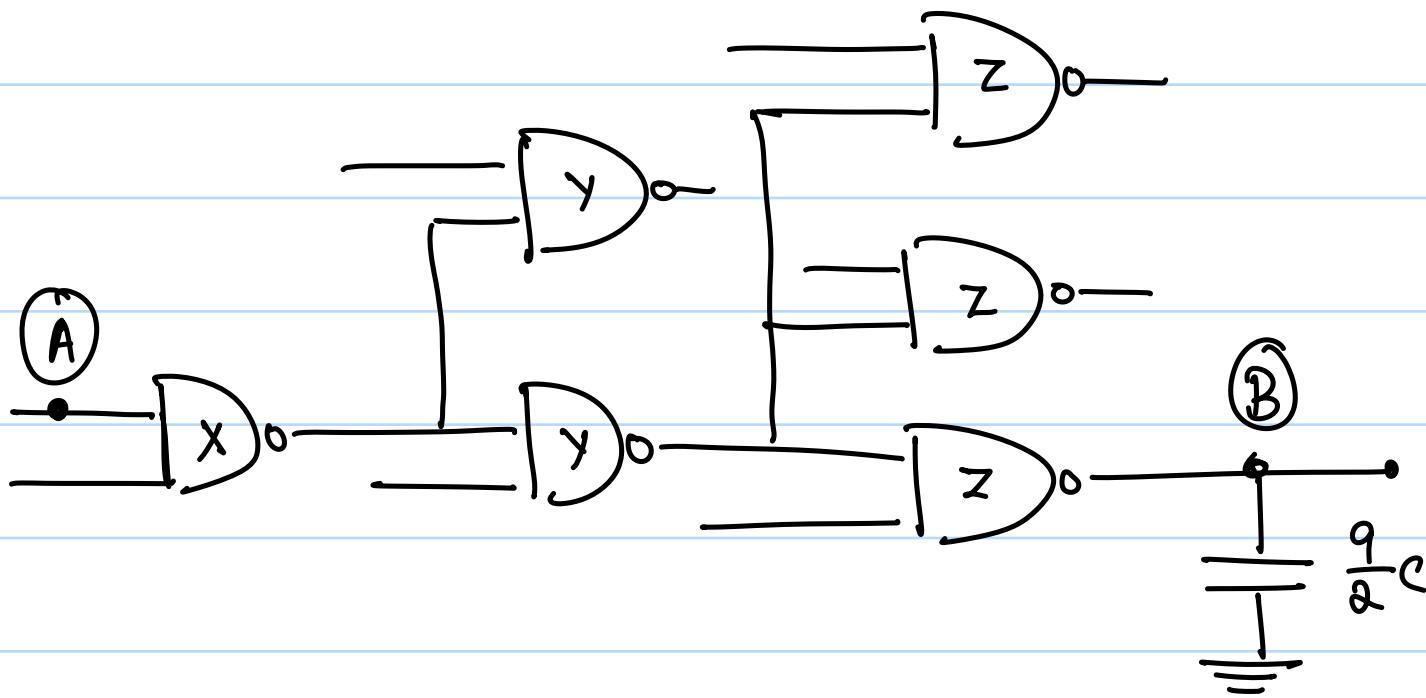
$$f = h_z \times g_z \times b_z \Rightarrow \frac{8}{3} = \frac{C_L}{C_{in-z}} \times \frac{4}{3}$$

$$C_{in-z} = \frac{C_L}{2} = 4C$$

$$h_y \times g_y = f$$

$$C_{in-y} = \frac{C_z}{2} = \frac{C_z}{2} = 2C$$

Example:



Optimize delay across A, B.

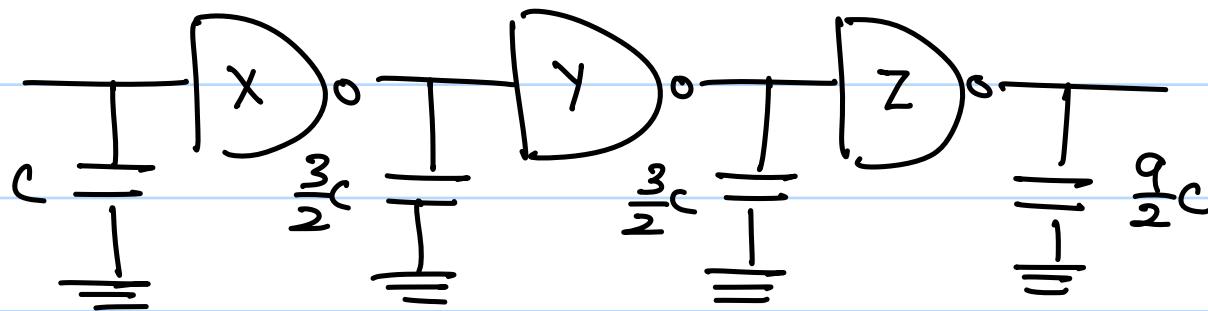
$$G_1 = \left(\frac{1}{3}\right)^3, B = 2 \times 3 \times 1 = 6, H = \frac{9}{2}$$

$$\Rightarrow F = \left(\frac{4}{3}\right)^3 \times 6 \times \frac{9}{2} = \frac{4^3 \times 9}{2^2} = \underline{\underline{4^3}}$$

$$\Rightarrow \hat{f} = 4$$

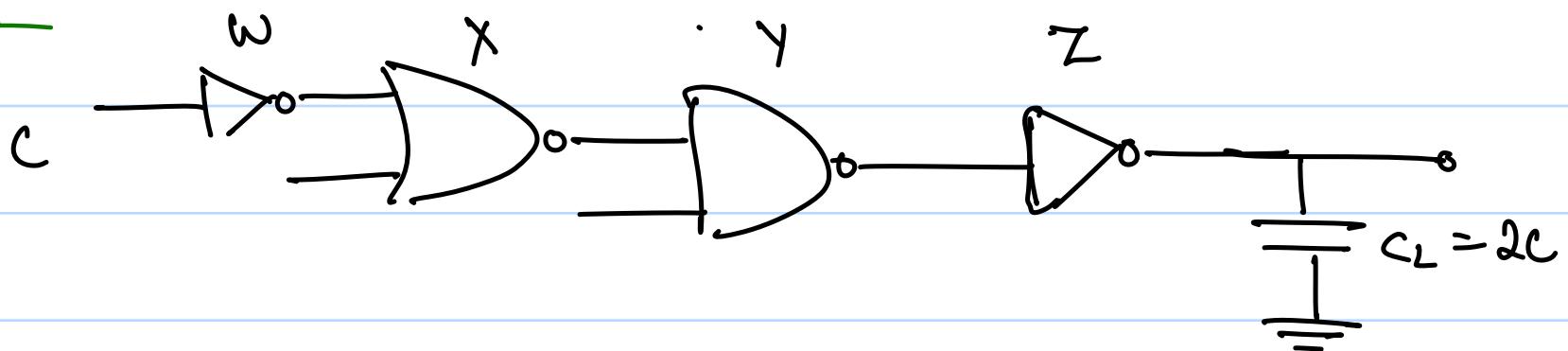
$$h_x = \frac{4}{\frac{4}{3} \times 2} = \frac{3}{2}, \quad h_y = \frac{4}{\frac{4}{3} \times 3} = 1, \quad h_z = \frac{4}{\frac{4}{3} \times 1} = 3$$

Let ref. inv of the sized  $\omega$ ,



$$\therefore \omega_y = \frac{3}{2} \omega, \quad \omega_z = \frac{3}{2} \omega$$

Example:



$$H = 2$$

$$G = 1 \times \frac{5}{3} \times \frac{4}{3} \times 1 = \frac{20}{9} \quad h_w = \frac{1.452}{1 \times 1} = 1.452$$

$$B = 1 \quad h_x = \frac{1.452}{\frac{5}{3} \times 1} = 0.871$$

$$\Rightarrow F = \frac{40}{9} \Rightarrow f = \underline{1.452}, \quad h_y = \frac{1.452}{\frac{4}{3} \times 1} = 1.089$$

$$D = \left( 4 \left( \frac{40}{9} \right)^{1/4} + \sum p_i \right) \tau \quad h_z = \frac{1.452}{1 \times 1} = 1.452$$

$$= \left( 4 \times 1.452 + 6 f_{inv} \right) \tau \Rightarrow \omega_x = 0.871 \omega$$

$$\omega_y = 0.949 \omega$$

$$\omega_z = 1.378 \omega$$

Example: Given  $C_{in} = C$ ,  $C_{out} = 25C$ ,  $f_{pin} = 1$ , then for values of  $N$  will the buffer chain be minimized.  $N = 1, 3, 5$ .

$$D = N(F)^{1/N} + Nf_{pin} = N(F^{1/N} + 1)$$

$$F = 25, \Rightarrow D = N(25^{1/N} + 1)$$

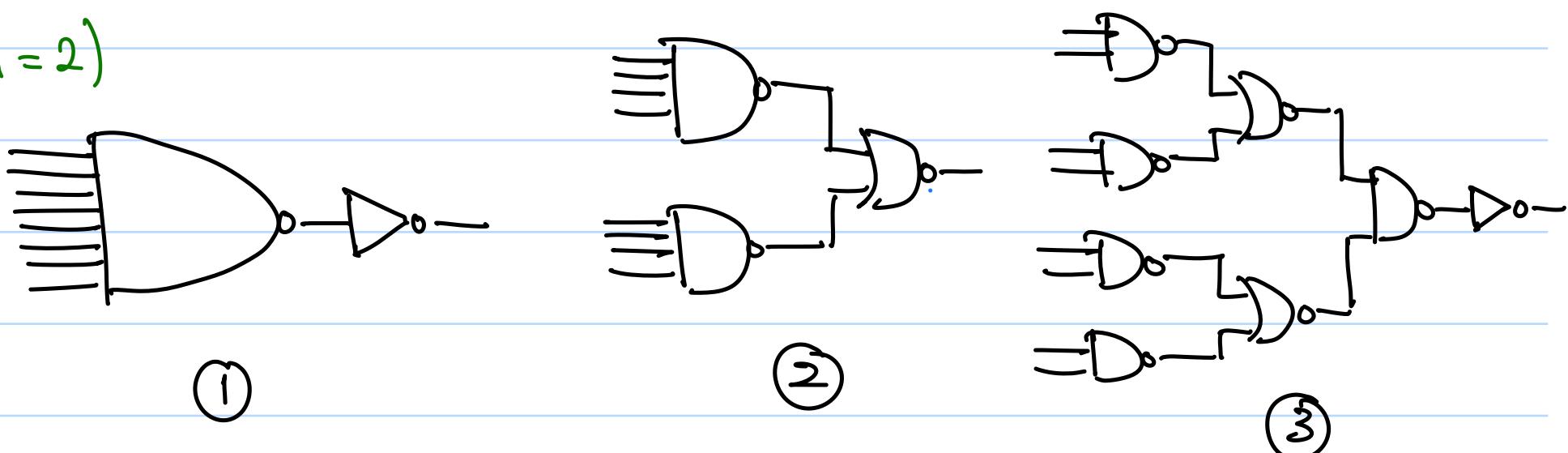
$$D_1 = 26$$

$$D_3 = 3(25^{1/3} + 1) = 3(3.924) = 11.772$$

$$D_5 = 5(25^{1/5} + 1) = 5(2.904) = 14.52$$

Example: 3 implementation of an 8 ilp AND gate are as below.

(Let  $m = 2$ )



Check which implementation has the lowest delay for  $H = 1, 10, 100$  and discuss the sizing of the gates.

Arrangement 1,

$$F = GBH \Rightarrow F = \frac{10}{3}H$$

$$G = \frac{8+2}{3} = \frac{10}{3}$$

$$B = 1,$$

$$D = \left(\frac{10}{3}H\right)^{1/2} + \sum f_i = \frac{10}{3}H + 9$$

Arrangement 2,

$$G = \left(\frac{4+2}{3}\right) \left(\frac{5}{3}\right) = \frac{30}{9}$$

$$B = 1$$

$$\Rightarrow F = \frac{30}{9} H \quad \Rightarrow D = 2 \left(\frac{30}{9} H\right)^{1/2} + 2(4+2)$$

$$D = 2 \left(\frac{30}{9} H\right)^{1/2} + 12$$

Arrangement 3,

$$G = \frac{4}{3} \times \frac{5}{3} \times \frac{4}{3} = \frac{80}{27}$$

$$B = 1$$

$$\Rightarrow F = \frac{80}{27} H \quad \Rightarrow D = 4 \left(\frac{80}{27} H\right)^{1/4} + 4(2+2+2+1)$$

$$= 4 \left(\frac{80}{27} H\right)^{1/4} + 28$$

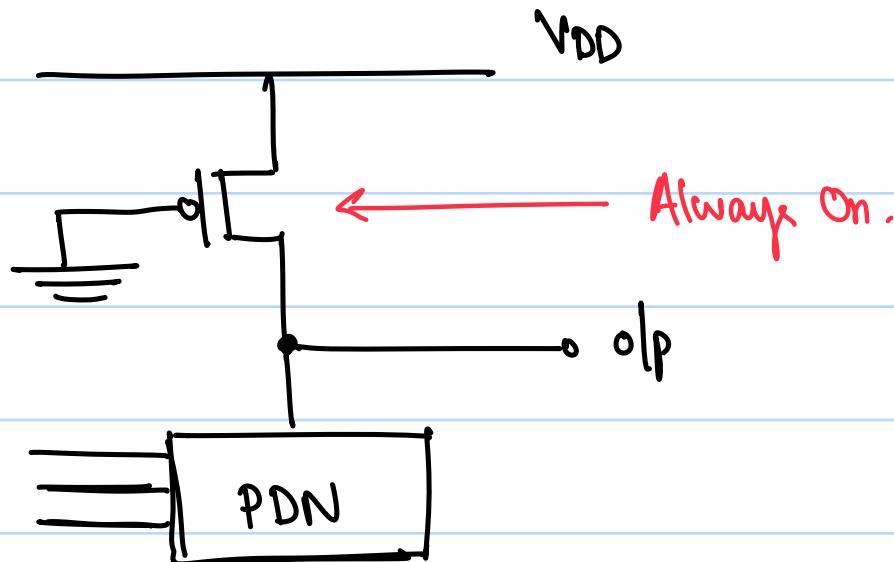
D/H	1	10	100
①	10.82	14.77	27.25
②	15.65	23.54	48.51
③	33.24	37.33	44.59

\* Other Logic Styles :-

Static Logic Styles

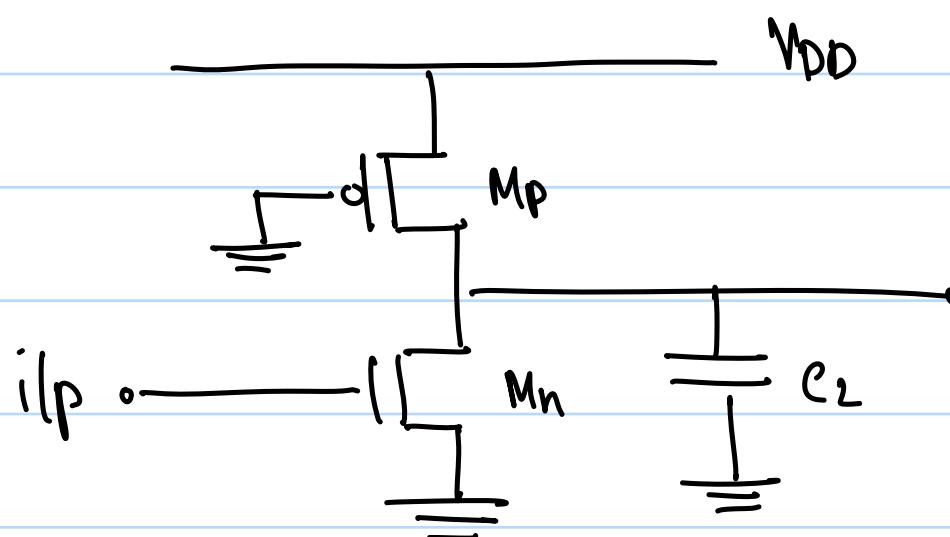
Cmos      Pseudo-NMOS      Pass Transistor Logic      Cascaded Voltage Sw. Logic  
pNMOS      (or Complement PTL)  
PTL, CPTL/CPL      CNVL

→ Pseudo - NMOS :-

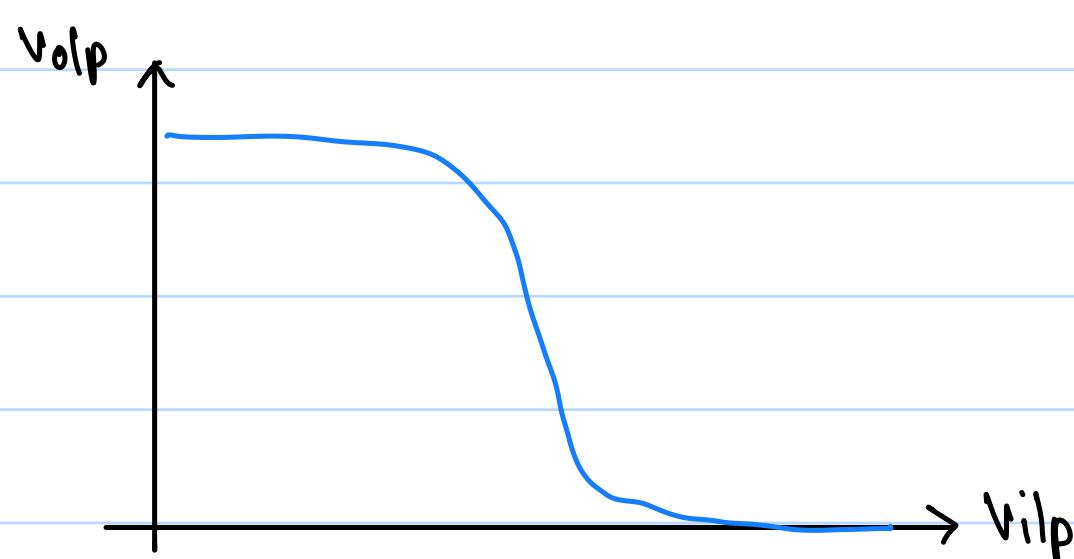


- The PDN is just one PMOS transistor which is always on.

• p-NMOS Inverter :-



- When  $i_{lp}$  is low,  $M_n$  is off and  $C_L$  will get charged through  $M_p$ , giving us a high  $o_{lp}$  at the end at steady state.
- When  $i_{lp}$  is high and  $M_n$  can give large current draw, the load  $C_L$  is discharged, giving us a low  $o_{lp}$  at the end at steady state.

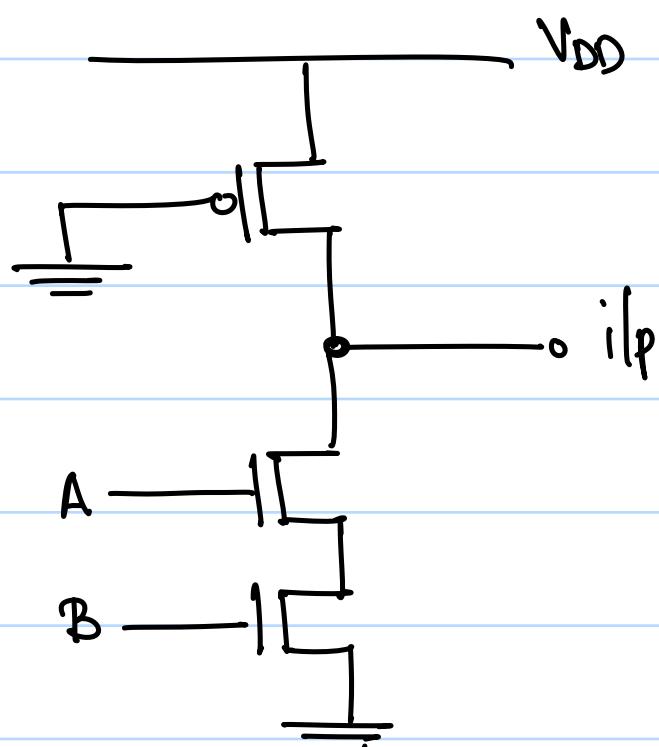


- This inverter is similar to a CMOS inverter, so parameters like Noise Margin, Delay, etc can be extracted through a similar process as in static CMOS style.
- In static CMOS design, the widths of the PMOS and NMOS can be anything, if symmetric delay is not required. There will be no difference in the logical operation. So CMOS style is a staticless style.

However in p-NMOS style, there is a strict lower bound on the width of the NMOS, with respect to the width of the PMOS. ( $w_n > w_p$ ), so p-NMOS is gated style.

- The design of the PDN network follows the same procedure as in CMOS style.

### p-NMOS 2inp NAND gate :



- Since the PDN network is just one PMOS transistor, the area of our circuit decreases, due to the absence of the PMOS stack.

- But due to the much higher width of the NMOS,  $C_{in}$  can be higher than in CMOS style.

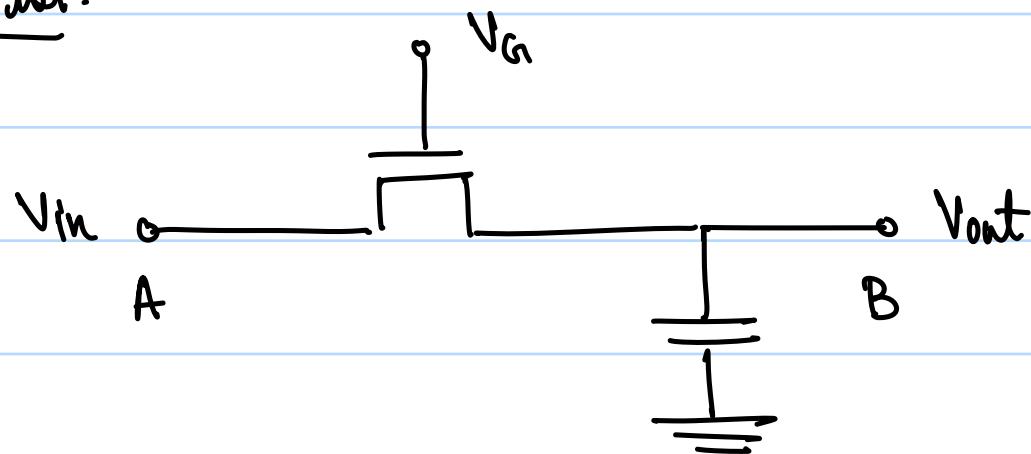
Also, since the PMOS is always on, static power consumption is non-zero. (when  $i_{lp}$  is high)

- p-NMOS is used in devices that stay at a certain logic level most of the time, ex: Reset circuit.

- Pass Transistor Logic / Complementary Pass Transistor Logic :-

- A MUX based logic style.

- Pass Transistor:

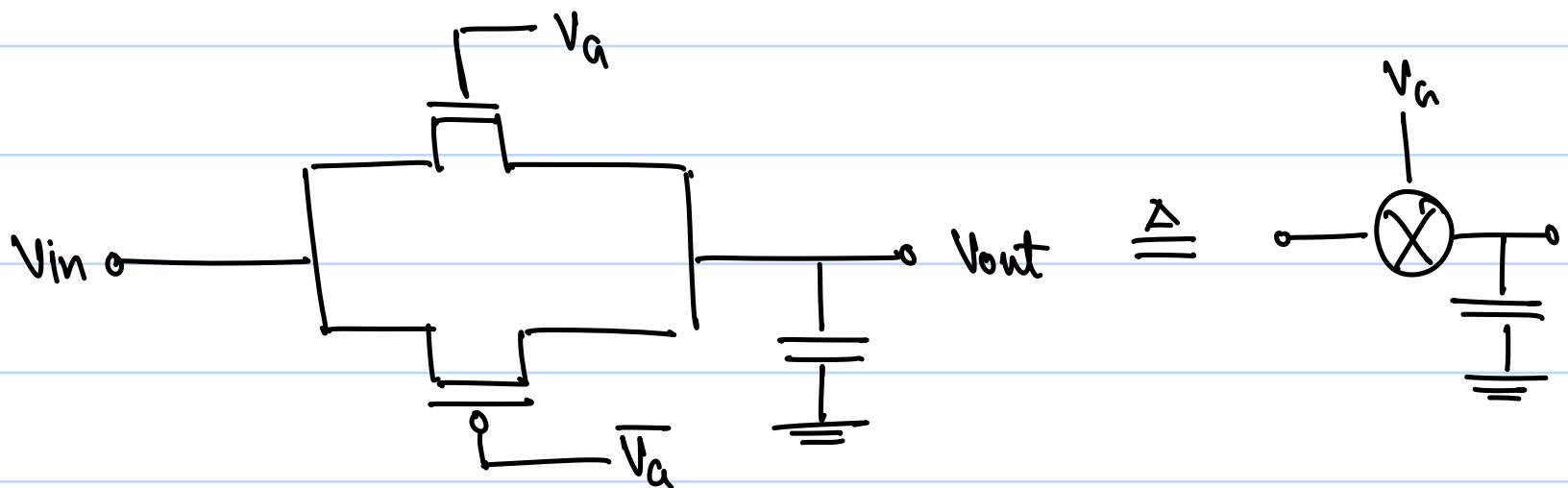


In the above circuit, at high  $V_G$ , if  $V_{in} = 0$ ,  $V_{out} = 0$  and if  $V_{in} = V_{DD}$ ,  $V_{out} = V_{DD} - V_{TN}$

Similarly if a PMOS is used, at low  $V_G$ , if  $V_{in} = V_{DD}$ ,  $V_{out} = 0$  and if  $V_{in} = 0$ ,  $V_{out} = |V_{TP|}$

So in a NMOS, PMOS pass transistor circuit, the logic levels are not proper high and low, either one of the MOS is non-ideal.

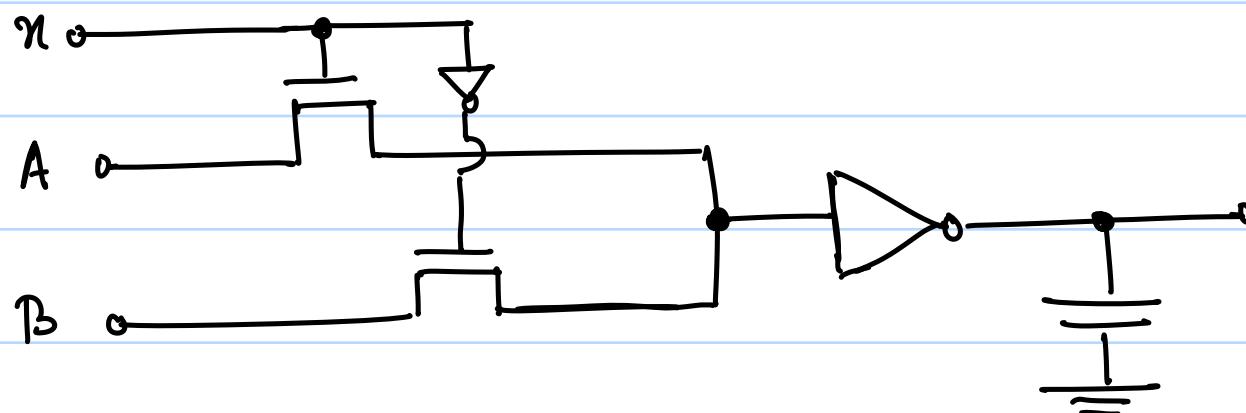
Therefore a Complementary Pair Transistor Gate is used,



Using a CPTL consumes more area than p-NMOS and CMOS, creating higher parasitics and increasing circuit complexity. Also, the usage of complementary controls ( $V_a$  and  $\bar{V}_a$ ) also adds complexity.

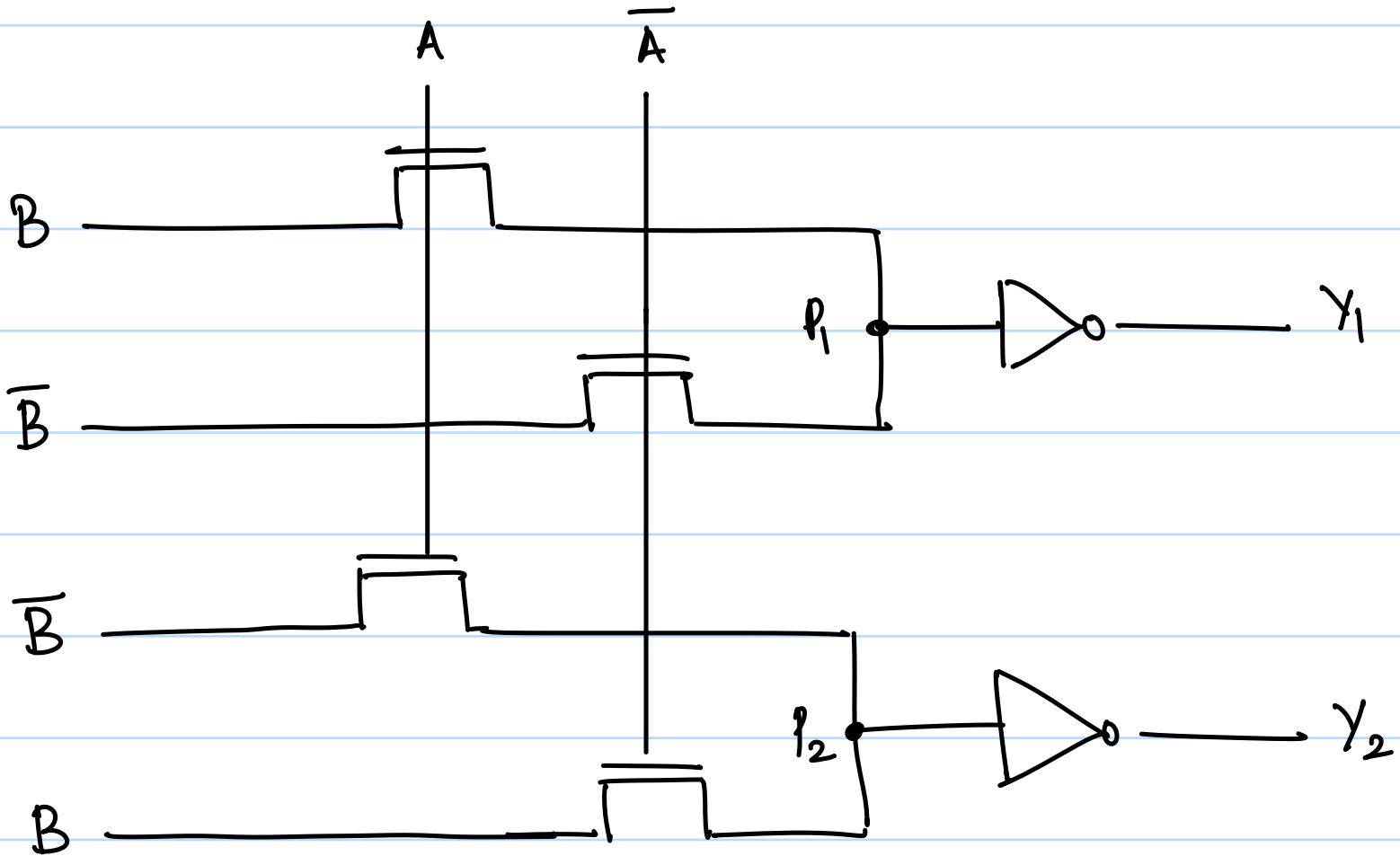
We can also use a CMOS inverter to invert the logic levels, in an NMOS / PMOS pass transistor, due to the idea of noise margin.

- Implementing logic functions in PTL :-



- The above circuit is an example of an inverting  $2 \times 1$  MUX, with inputs A, B and selection line X.

• By Shannon's expansion, any logic function can be implemented on a combination of MUXes.

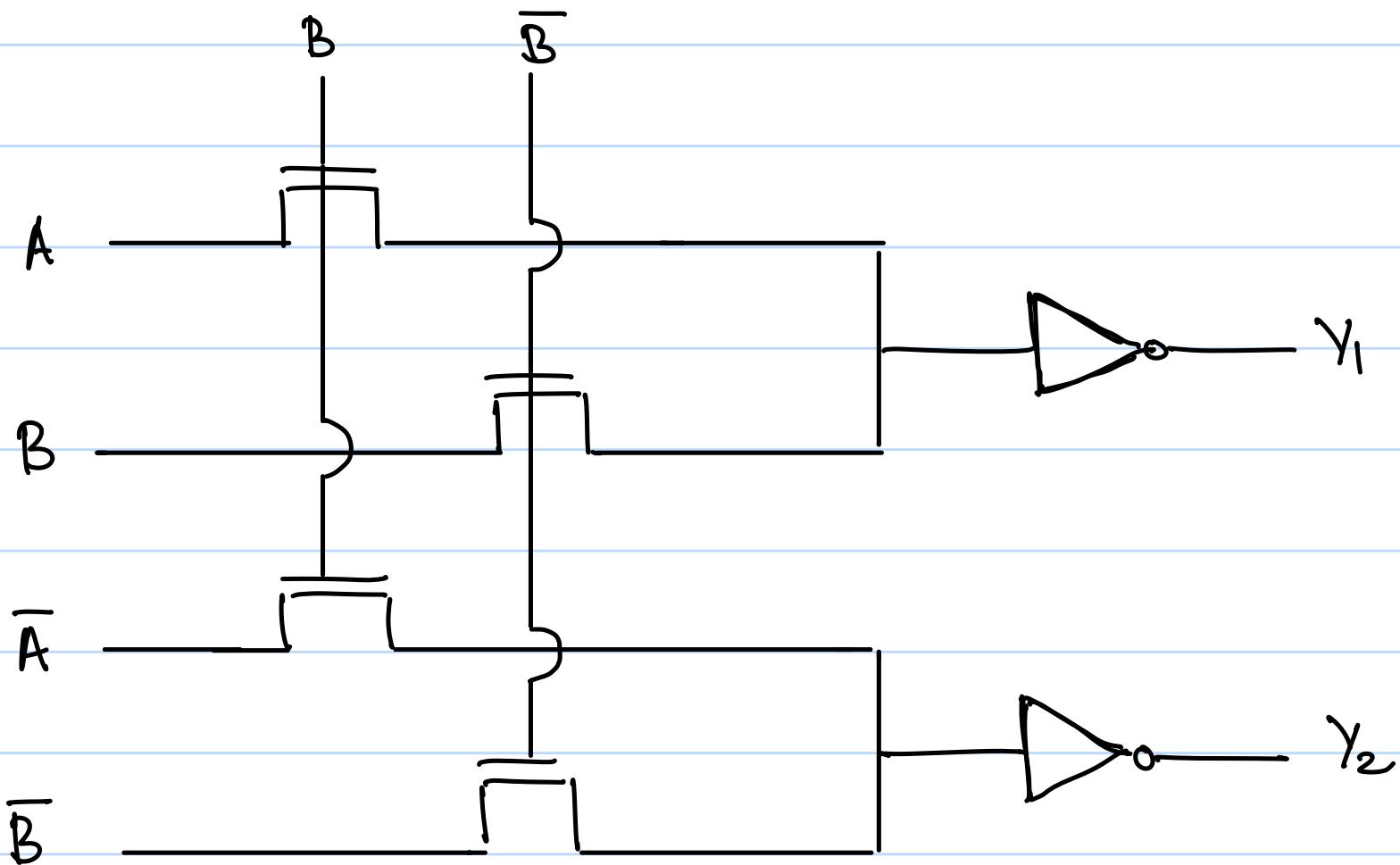


$$P_1 = AB + \bar{A}\bar{B} \xrightarrow{\text{XNOR}} Y_1 : \text{XOR}$$

$$P_2 = \bar{A}\bar{B} + \bar{A}B \xrightarrow{\text{XOR}} Y_2 : \text{XNOR}$$

This implementation of XOR and XNOR are much more efficient than CMOS topologies.

Since XOR and XNOR are the building blocks of arithmetic blocks like Adder and Multiplier, such topologies can vastly increase the efficiency of the design.



$$Y_1 = (AB + B\bar{B})^I = (AB)^I : \underline{\text{NAND}}$$

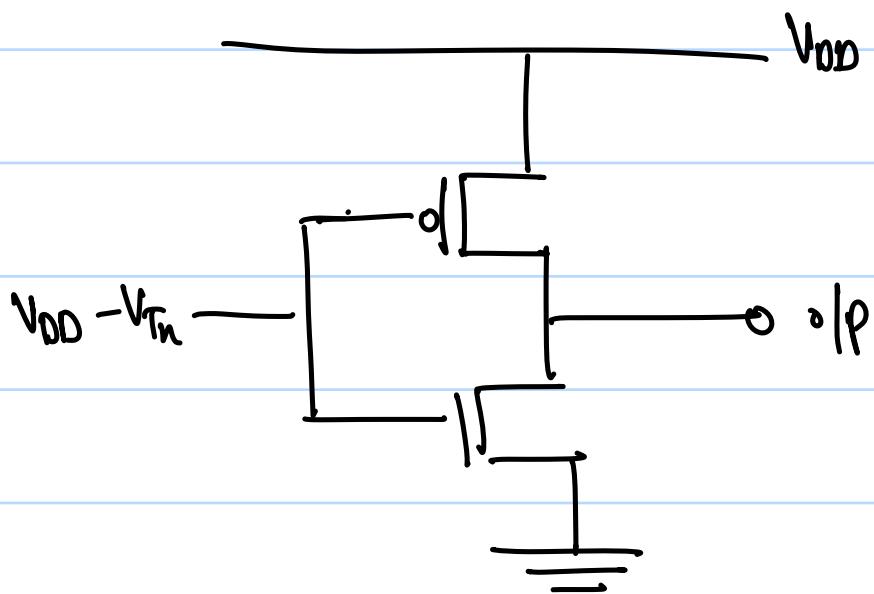
$$\begin{aligned} Y_2 &= (\bar{A}B + \bar{B}\bar{B})^I = (\bar{A}B + \bar{B})^I = (A + \bar{B})B \\ &= AB + \bar{B}B = AB : \underline{\text{AND}} \end{aligned}$$

Here the same topology is used for a different logic function by just changing the inputs.

Also, since the path of each ilp is identical, the delay of each ilp is the same for any logic function.

But static power dissipation is still present and the logic levels are not rail to rail, like in CMOS style.

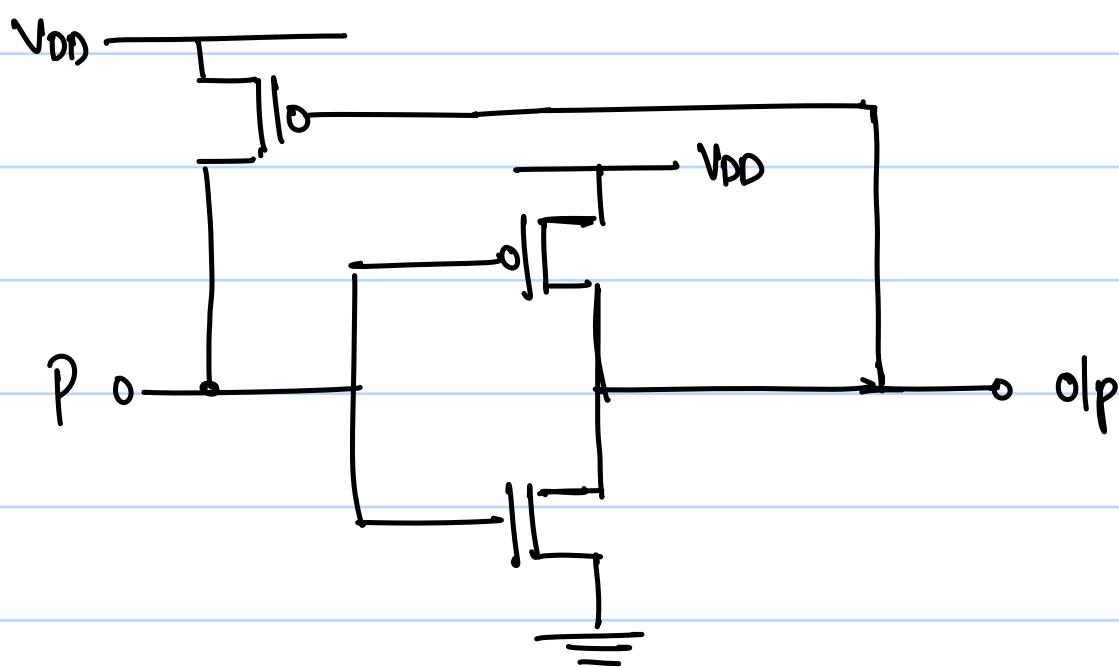
Since the high off of the NMOS PT is  $V_{DD} - V_{Th}$ , if this is given as ilp to the CMOS inv



- If  $V_{Tn} = |V_{Tp}|$ ,  $V_{sap} = V_{DD} - (V_{DD} - |V_{Tp}|)$   
 $= |V_{Tp}| \longrightarrow$  Edge of cutoff

Therefore the PUN is never fully turned off, causing current leakage and significant static power consumption in the CMOS inverter (like in p-NMOS)

To fix this, we use a keeper circuit, a kind of feedback circuit



If  $o/p = \text{low}$ , the keeper circuit is active and the voltage at P increases to  $V_{DD}$ , thus restoring the logic levels and reducing static power consumption.

If  $o_{lp} = \text{high}$ , the keeper circuit is inactive, which is necessary as keeper circuit is not required at that point.

However, if  $P: V_{DD} - V_{Th} \rightarrow 0$ , node P tries to discharge through the NMOS PT, but the keeper circuit is still momentarily charging.

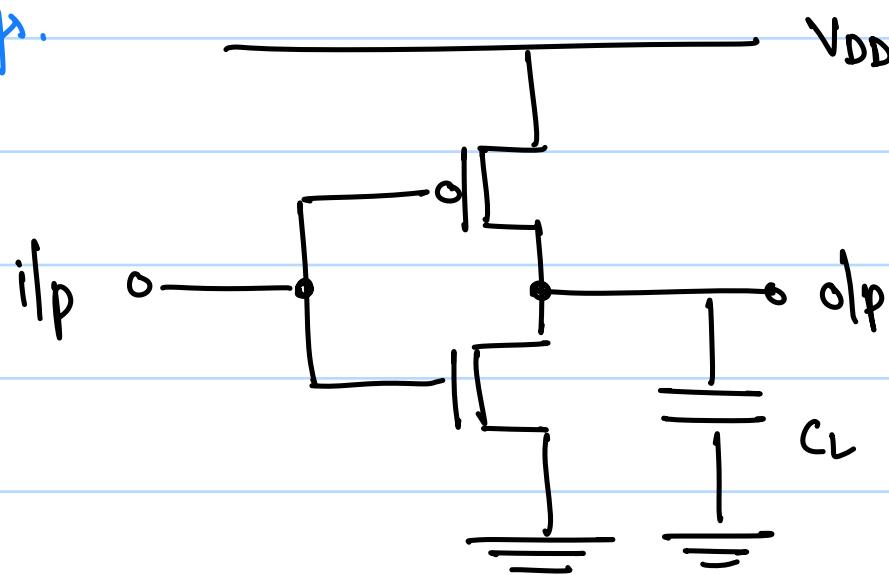
Therefore the NMOS must be stronger than the keeper PMOS, or the PMOS keeper should be weaker.

stronger  $\rightarrow$  higher current draw  $\rightarrow W \uparrow$

Making the NMOS stronger  $\rightarrow$  Increase  $C_{in}$ .

Making the keeper weak  $\rightarrow$  Increase delay.

- Say we cannot implement a keeper PMOS, and we want to equalize rise/fall delays.



Time taken for  $C_L$  to discharge at  $V_{DD} ip = t_{PHL1}$

" " " " "  $V_{DD} - V_{Th} ip = -t_{PHL2}$

$t_{PH1} < t_{PH2}$ , due to lesser current draw.

When  $V_{ilp} = V_{DD}$ ,  $I_1 \propto \frac{w}{L} (V_{DD} - V_{Tn})^2$   
 $V_{ilp} = V_{DD} - V_{Tn}$ ,  $I_2 \propto \frac{w}{L} (V_{DD} - 2V_{Tn})^2$

$\left. \begin{array}{l} \\ \end{array} \right\} \frac{w}{L} \rightarrow \text{Inv. NMOS aspect ratio}$

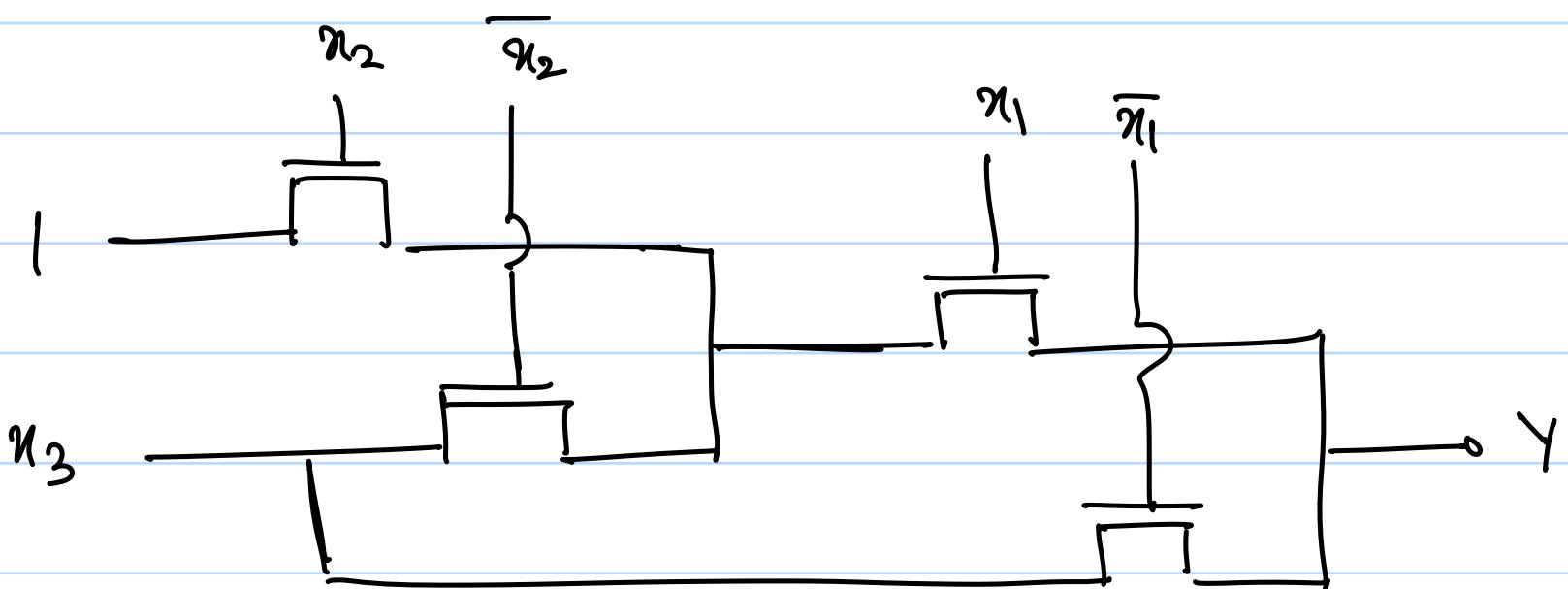
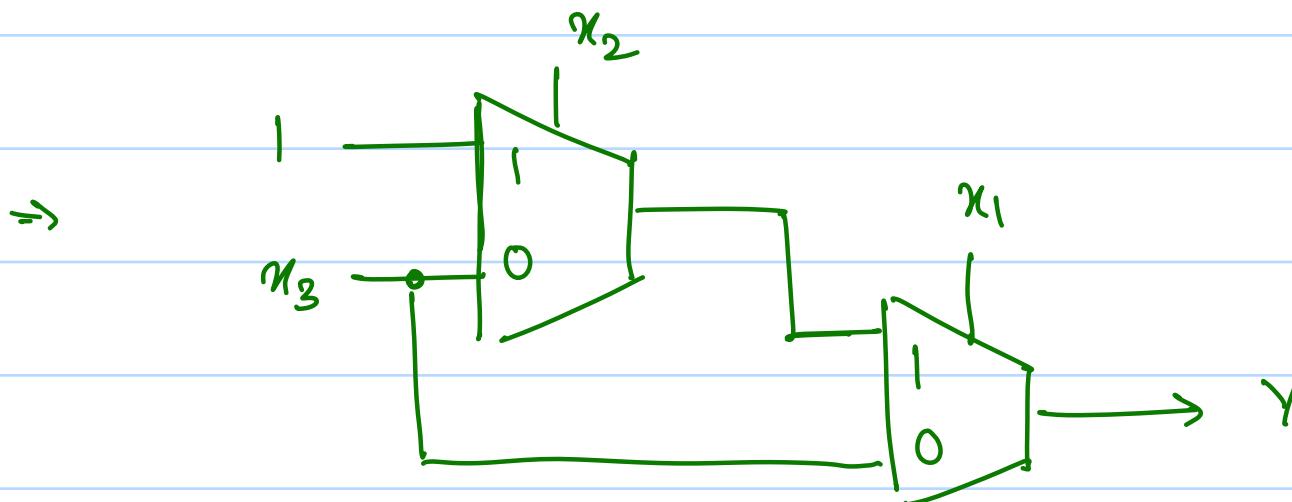
If we want to equalize the delays,  $I_1 = I_2$

$$\Rightarrow w_1 (V_{DD} - V_{Tn})^2 = w_2 (V_{DD} - 2V_{Tn})^2$$

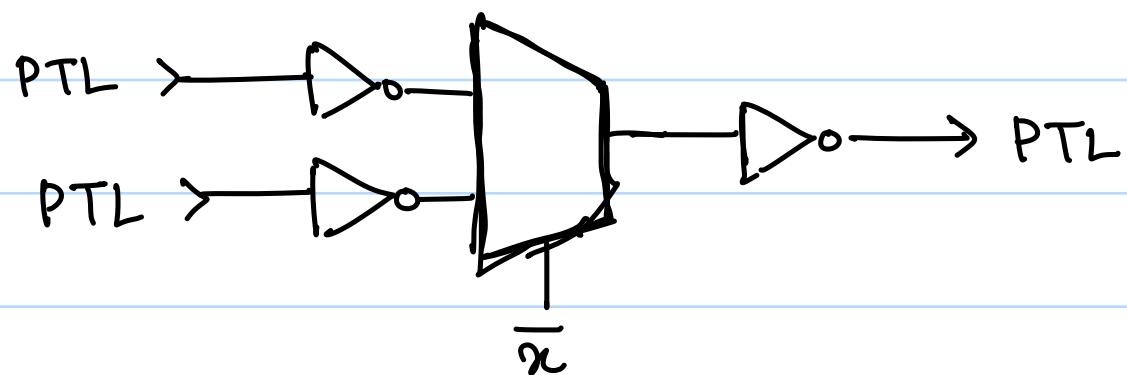
$$\Rightarrow w_2 = w_1 \frac{(V_{DD} - V_{Tn})^2}{(V_{DD} - 2V_{Tn})^2} > w_1 \rightarrow \text{Current draw is more.}$$

Example: Implement a PTL topology for  $f = x_1x_2 + \bar{x}_2x_3 + x_3\bar{x}_1$

$$\begin{aligned} f &= x_1x_2 + \bar{x}_2x_3 + x_3\bar{x}_1 \\ &= x_1(x_2 + \bar{x}_2x_3) + \bar{x}_1(\bar{x}_2x_3 + x_3) \\ &= x_1(x_2(1) + \bar{x}_2(x_3)) + \bar{x}_1(x_3) \end{aligned}$$

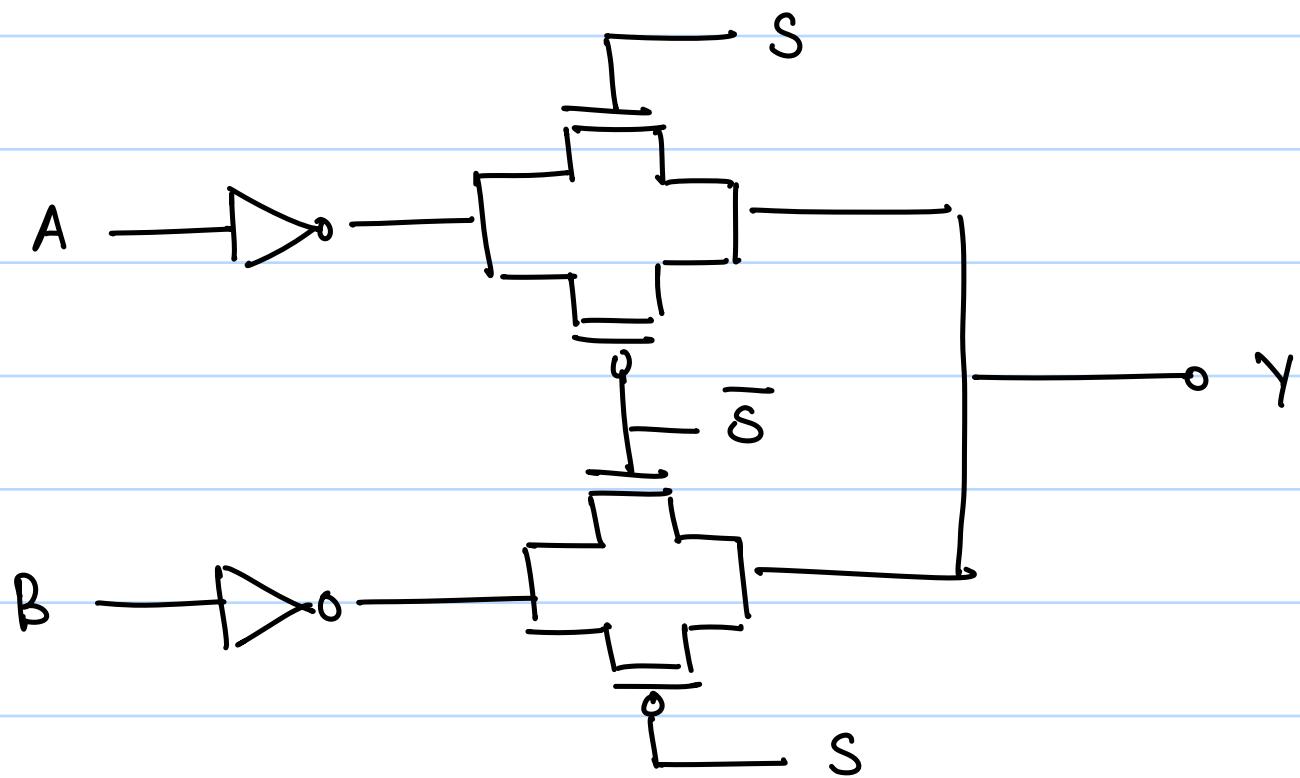


- When chaining multiple MUXes, since logic 1 is not fully passed by the pass NMOS, we can introduce inverters in between each of the MUXes, as "repeaters", ensuring that each MUX gets rail-to-rail inputs.



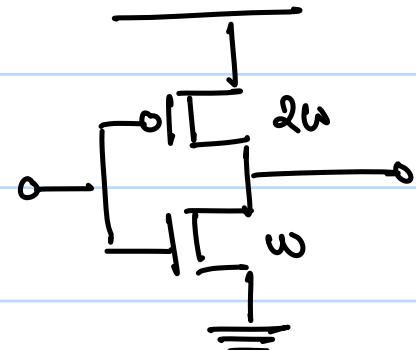
Note: The inverters used in PTL circuits is also termed as a decision circuit.

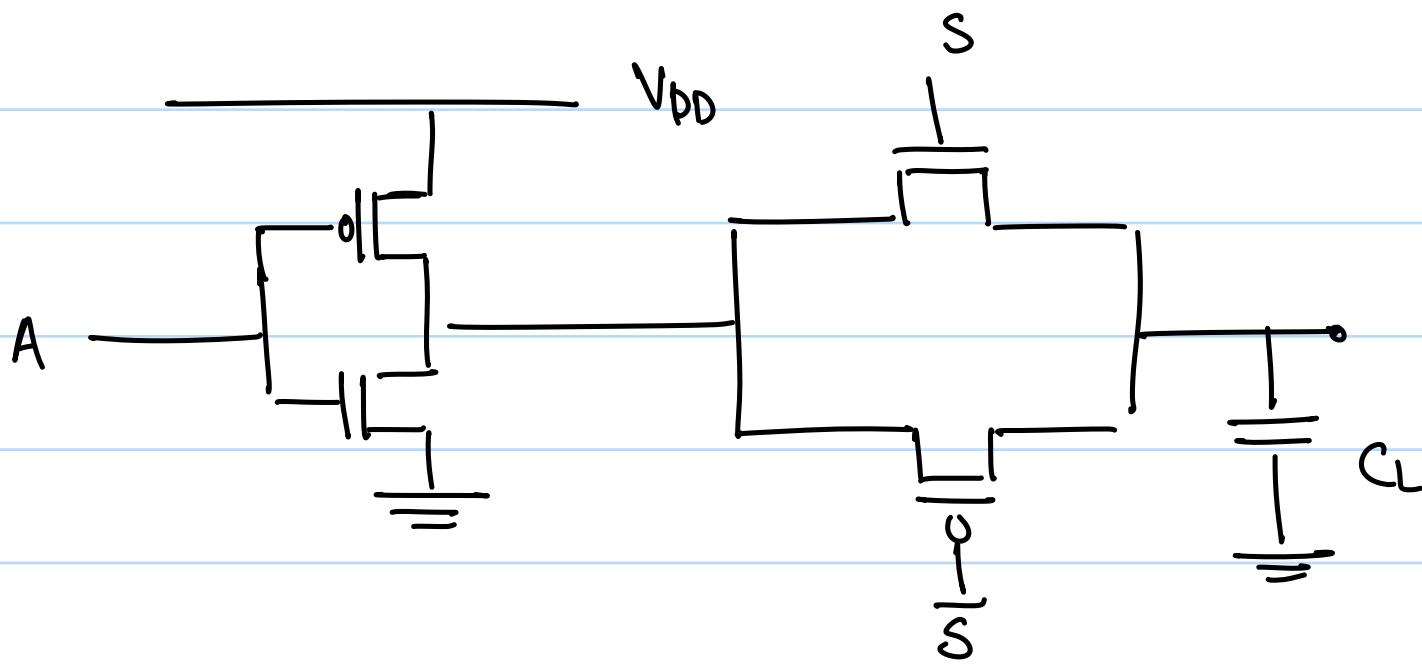
- We can remove the need of decision circuit by using a complementary pass transistor, at the cost of an higher area requirement.



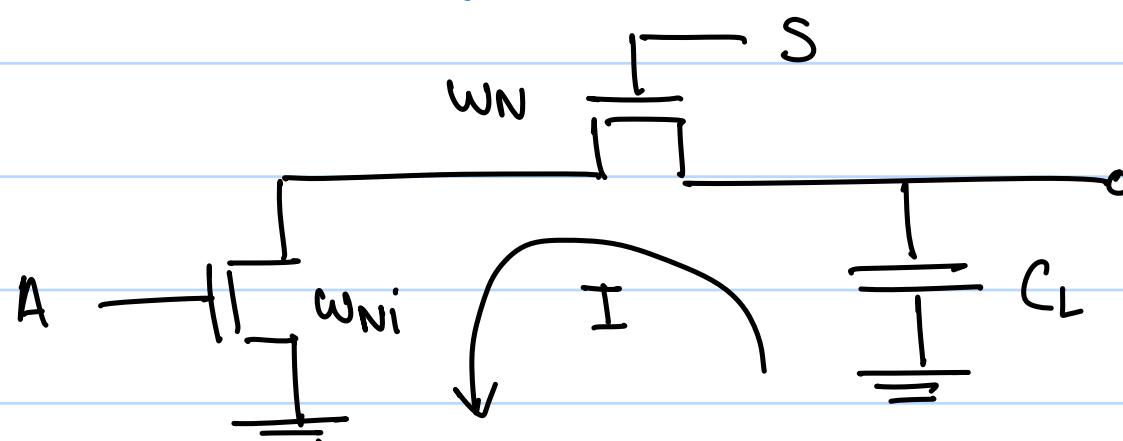
To size the pass transistors of the CPT, we will try to equalize the delay to that of a reference inverter.

(Let  $m = \frac{M_n}{M_p} = 2$  for the reference inverter)





A suggest new path for  $C_L$  to discharge will be,



This is like an RC circuit, with the resistance provided by the pair NMOS and inv. NMOS

$$\Rightarrow R \propto \frac{1}{w_N} + \frac{1}{w_{Ni}}$$

$$\text{If } w_N = w_{Ni} \Rightarrow R \propto \frac{2}{w_N} \Rightarrow \text{NMOS width : } 2w \quad (\text{Ref. inv w})$$

likly  $R \propto \frac{1}{w_p} + \frac{1}{w_{pi}}$  in the  $C_L$  charging path.

$$\text{If } w_p = w_{pi}, \text{ then } R \propto \frac{2}{w_p} \Rightarrow \text{PMOS width} = 4w$$

- Using repeaters in a cascade of pass transistors will reduce the charging/discharging path of the load capacitance, reducing delay.

