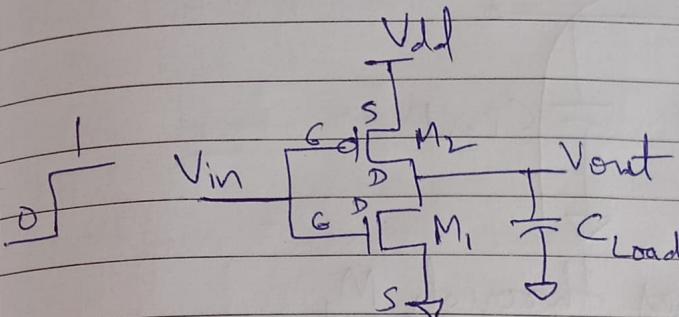


CMOS Inverter

→ Gain ↑ helps Analog design but also digital bcs it switches fast from 0 to 1 or viceversa fast



Vin	Vout
1	0
0	1

$$1 - V_{DD}$$

$$0 \rightarrow V_{GND}$$

Input in digital ckt's
are given mostly to
Gate and output mostly
taken at drain

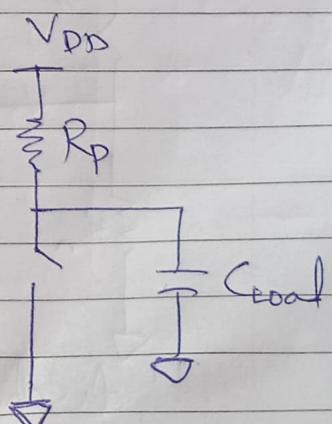
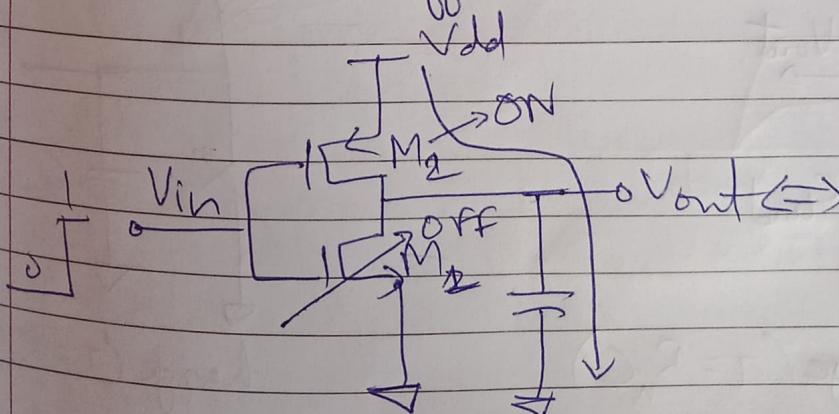
for PMOS

$$V_{GS} > |V_{tp}|$$

- There is a chance that V_{out} can be negative voltage
- M₁ source should be at most -ve terminal and
- M₂ source should be at most +ve terminal
- C_{load} takes care of Cas of M₁ & M₂ and next stages of the ckt's

when $V_{in} = 0$

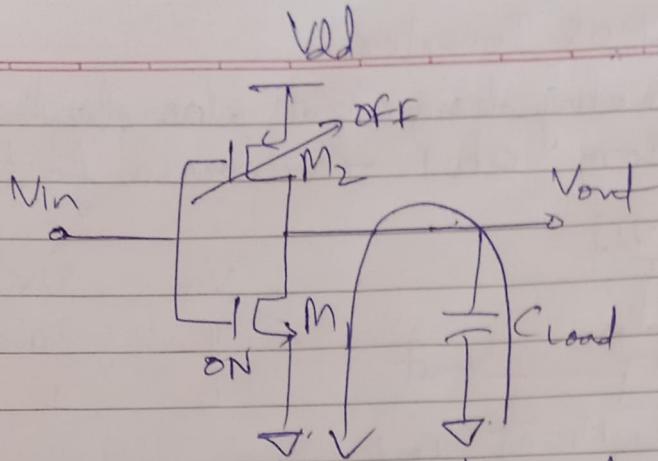
M₂ is ON (switched ON (saturation) acts as current source
M₁ is OFF (cut off)



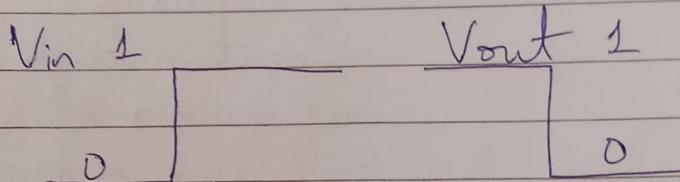
When $V_{in} = 1$ (Vdd).

→ M₂ is OFF (cutoff)

→ M₁ is ON (saturation) acts as a



All charge discharges through M_3

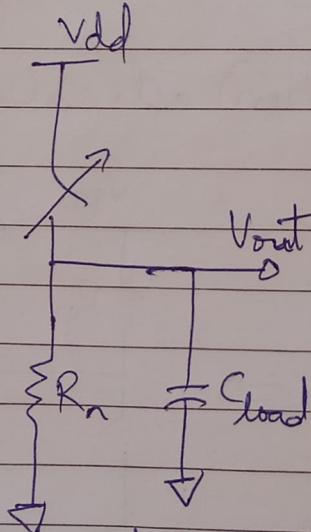


Switch Model

when $V_{in} = V_{DD}$.

NMOS is ON,

PMOS is OFF

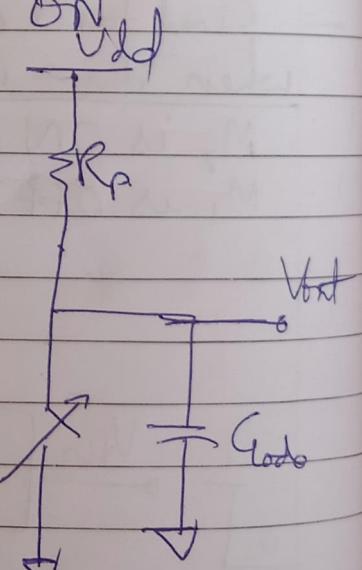


discharge ($T = R_n C$)

when $V_{in} = 0$;

NMOS is OFF

PMOS is ON

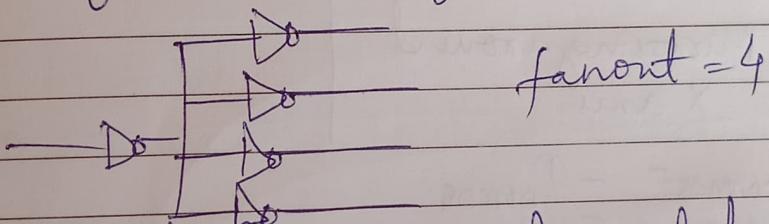


charge ($T = R_p C$)

Static Behaviors: (output Node is not floating Node)

- High Noise Margin, ideally equal to supply voltage
- Ratio-less logic level
- Low Input Output Impedance - less prone to noise
- High Impedance - improve fanout
- No static power consumption

- Static operation power consumption (No switching happens bcs V_{dd}, E_p, GND are not connected so the static power dissipation is very less)
- Gate has oxide layer so impedance is very high
- Fanout (Capability of a single device driving N number of devices) large devices



But Transient Response is degraded

- Low output Imp means high SNR which rejects noise inserted at output
- Ratio less logic level: Irrespective of aspect ratio $\frac{W}{L}$ the static behaviour is independent of that
- While charging we are going to V_{dd} and discharging we are going to ground. So we are able to get full swing of the signal and we are not wasting any of the signal in CMOS inverter

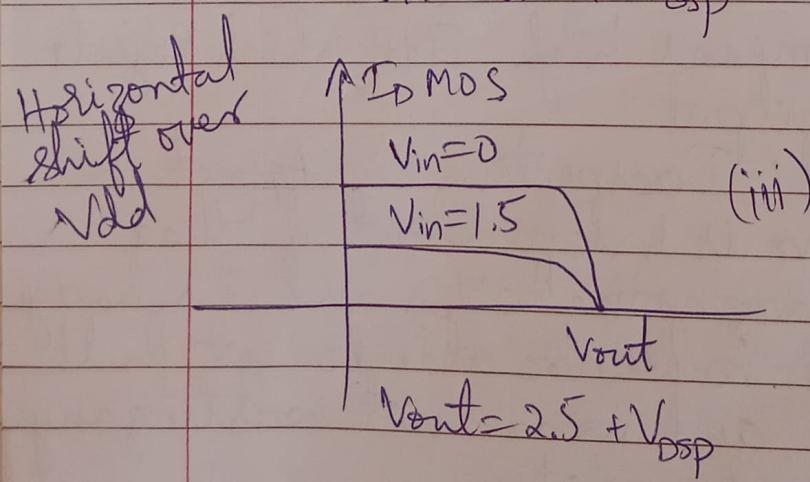
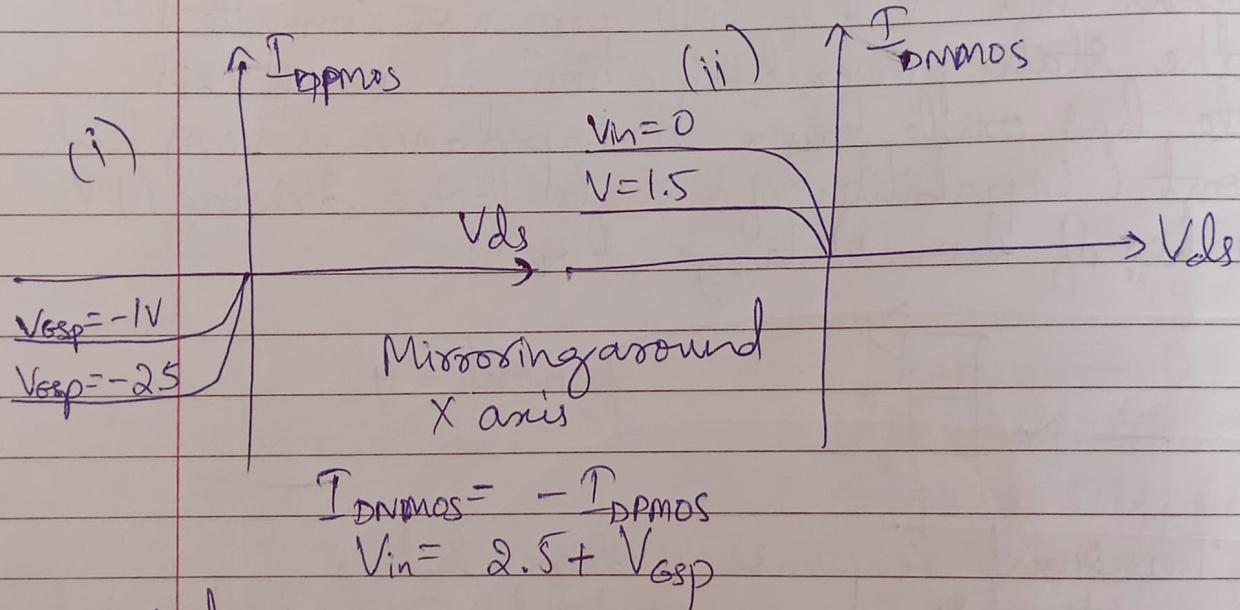
Since M₁ is responsible for pulling down the voltage at V_{out}, So M₁ → pull down device.

M₂ is responsible for pulling up the voltage at V_{out}, So M₂ → pull up device.

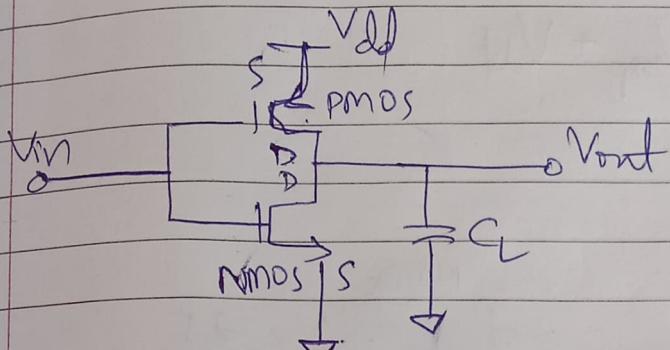
Here we vary V_{in} from 0 to V_{dd}
 We need to plot IV of $M_1 \oplus M_2 \oplus$ combination
 of both

CMOS INVERTER Voltage transfer char

→ CMOS Voltage transfer characteristics can be achieved by superimposing of drain current of NMOS & PMOS onto a common coordinate plot



Governing Eqn's



$$I_{DSN} = -I_{DSP}$$

$$V_{GSN} = V_{in}$$

$$V_{GSP} = V_{in} - V_{dd}$$

$$V_{DSN} = V_{out}$$

$$V_{DSP} = V_{out} - V_{dd}$$

$$V_{GSN} = (V_G - V_S)_N \\ V_{GSN} = (V_G - 0)_N = V_G \Rightarrow \underline{V_{in}} = V_{GSN}$$

$$V_{DSN} = (V_D - V_S)_N = (V_D - 0)_N = V_D = V_{out}$$

$$V_{DSN} = \underline{V_{out}}$$

$$V_{GSP} = (V_G - V_S)_P = (V_G - V_{dd}) = (V_{in} - V_{dd})_P \\ V_{GSP} = \underline{V_{in} - V_{dd}}$$

$$V_{DSP} = (V_D - V_S)_P = (V_{out} - V_{dd}) = \underline{V_{out} - V_{dd}}$$

$$\text{In (ii)} \quad V_{in} = V_{GSP} + V_{dd}$$

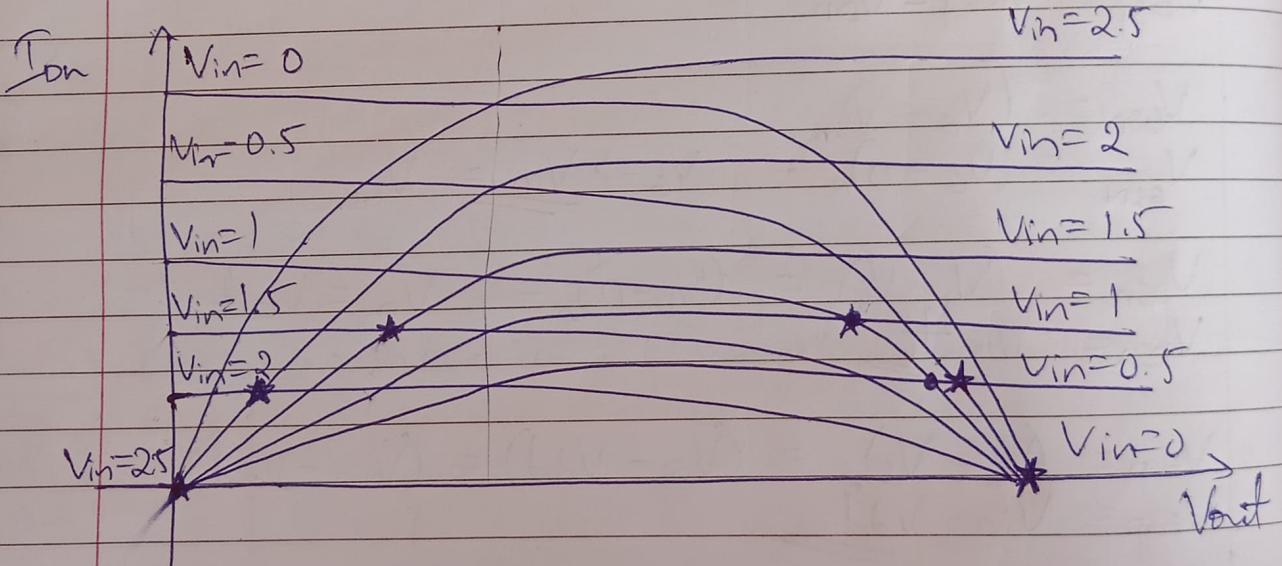
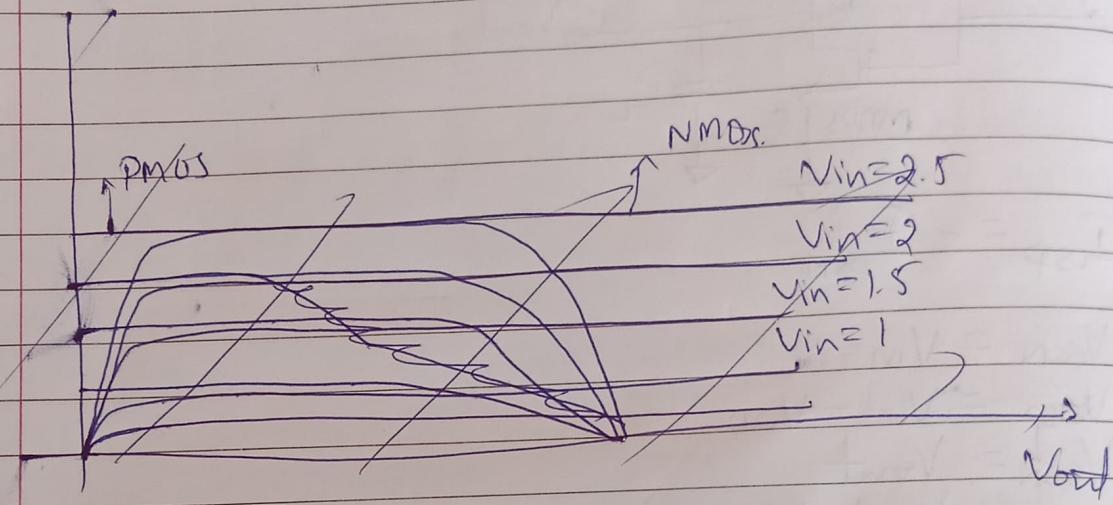
$$\text{at } V_{dd} = 2.5 \quad V_{GSP} = -2.5$$

$$V_{in} = 2.5 - 2.5 = \underline{0}$$

$$V_{in} = V_{GSP} + V_{dd} = -1 + 2.5 = \underline{+1.5}$$

In (iii) we shift (ii) horizontally

$$\text{but } V_{\text{out}} = V_{\text{Dsp}} + V_{\text{dd}}$$

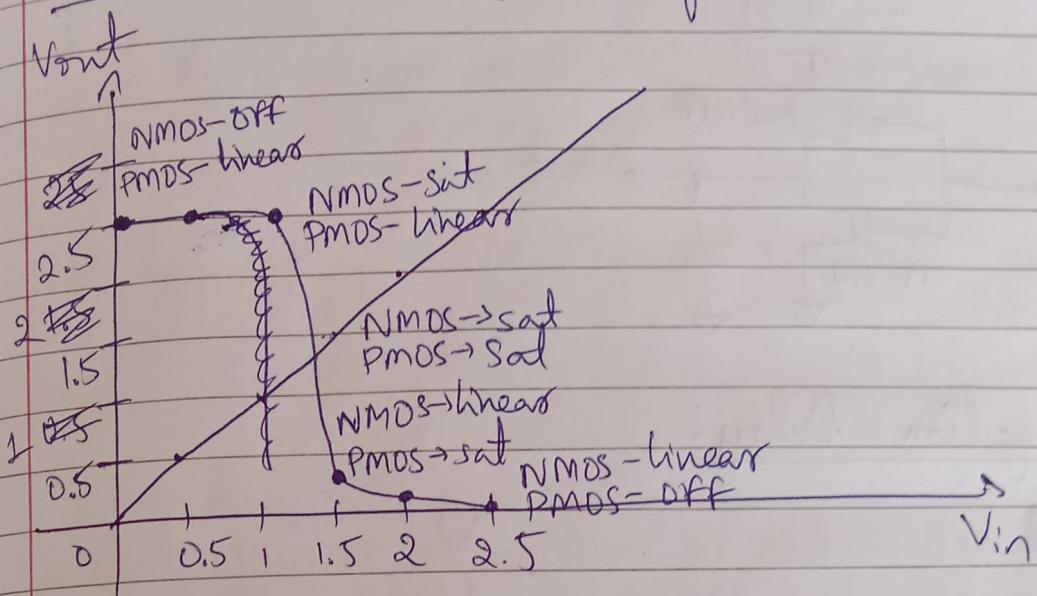


→ at * the G-S voltage for both NMOS & PMOS will be same and current will also be same

$$I_{\text{Dsp}} = -I_{\text{DSn}}$$

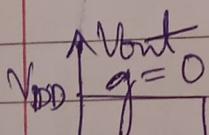
→ We have only those voltages either low or either very high

CMOS Inv - Voltage transfer characteristics

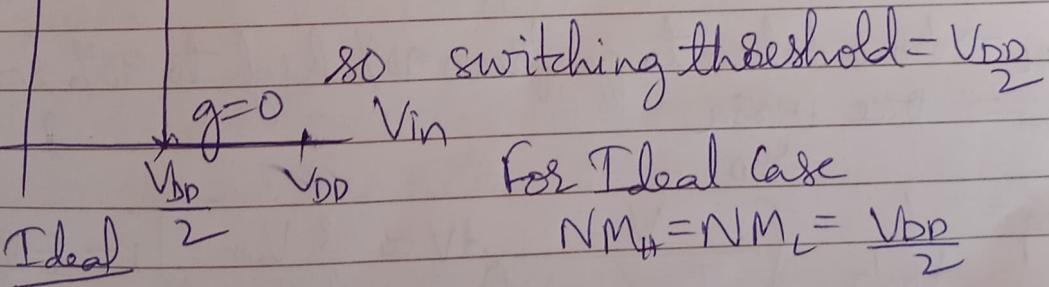


Noise Margin

$NM_H \Rightarrow$ low I/p & high o/p. How much noise can be added
 ~~NM_L~~ so as to not change the output



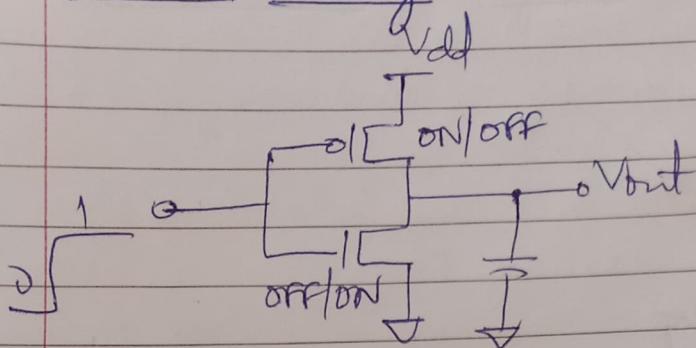
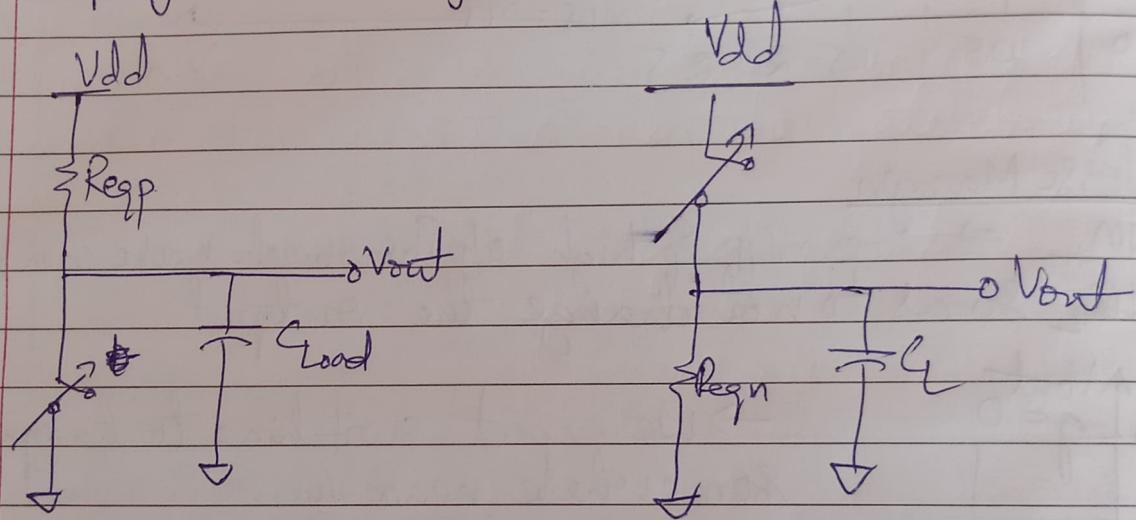
\rightarrow We expect switching to happen at somewhere near $\frac{V_{DD}}{2}$



For Ideal Case

$$NM_H = NM_L = \frac{V_{DD}}{2}$$

- \rightarrow If we lower V_{DD} , NM will also be lower so its property rejection of noise will be compromised
- \rightarrow In digital application we bias near $\frac{V_{DD}}{2}$ or 0 but at analog we do it b/w $\frac{V_{DD}}{2}$ & 0
- \rightarrow At $V_{out} = V_{DD}$ and $V_{out} = 0$ the gain is 0.
 $V_{IL} \rightarrow V_{in}$ low $V_{OL} \rightarrow V_{out}$ low
 $V_{IH} \rightarrow -U$ high $V_{OH} \rightarrow -U$ high
- $\rightarrow V_{OH} - V_{IL} = NM_H$

Inverter DelayPropagation Delay:

t_{PHL} → time taken to go from 90% high to 90% low
 t_{PLH} → Vice Versa

V_{DD} .

$$R_{eq} = \frac{1}{V_{DD}/2} \int_{V_{DD}/2}^V \frac{V}{T_{dsat}(1+\lambda V)} dV \approx \frac{3}{4} \frac{V_{DD}}{T_{dsat}} \left[1 - \frac{7}{9} \frac{V_{DD}}{2} \right]$$

$$t_{PHL} = \ln(2) R_{eq} C_{eq} = 0.69 R_{eq} C_L$$

Simplifying

$$t_{PLH} = \ln(2) R_{eq} C_{eq} = 0.69 R_{eq} C_L$$

Overall propagation delay.

$$t_p = \frac{t_{pH+} + t_{pA+}}{2} = 0.69 C_L \left(R_{eqn} + R_{eqp} \right) \frac{1}{2}$$

For $t_{pH+} = t_{pH+L}$ then $R_{eqp} = R_{eqn}$

$$\text{But } R = \frac{V}{I} \quad \text{but} \quad I = \mu_n C_{ox} \frac{W}{L} (V_{gs} - V_t)^2$$

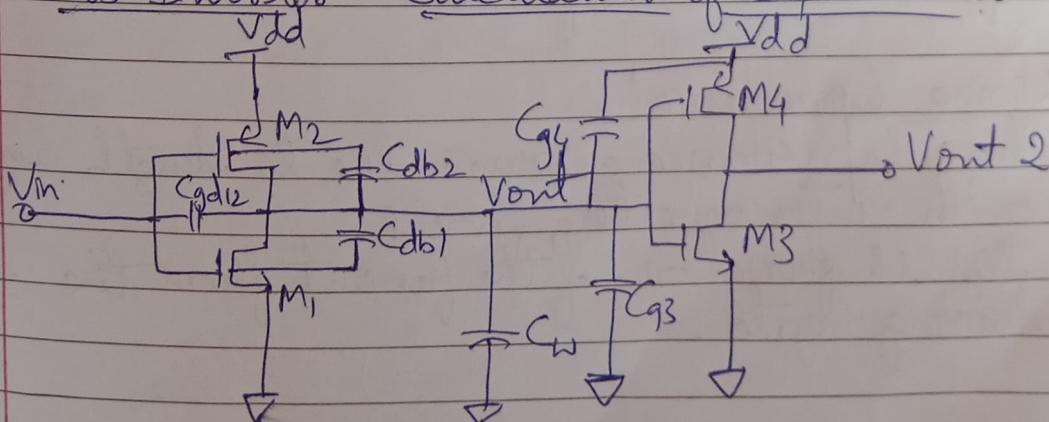
If we keep $\frac{W}{L}$ same but majority charge carriers for PMOS are holes which have less mobility (μ_p is less than μ_n) so the $R_{eqp} \neq R_{eqn}$, which is a problem

For $t_{pH+} = t_{pH+L}$

for that make width = $3L$ so Resistance reduced by 3 to R_{eqp} so the $R_{eqp} = R_{eqn}$. So we need to make 3 PMOS width 3 times

- for skewed transistor ($t_{pH+} = t_{pH+L}$)
- Non-skewed transistor ($t_{pH+L} \ll t_{pH+}$)

Cmos Inverter - Calculation of Capacitance



- For simplicity, we assume all capacitance are lumped together into one single capacitor C_L
- C_L capacitance can be broken down to
 - C_{gd12} → Gate drain Capacitance
 - C_{db12} → drain diffusion Capacitance
 - C_W → Wiring Capacitance
 - C_{g3}, C_{g4} → Gate Capacitance of fanout

Gate drain Cap. (C_{gd12}):

- M_1, M_2 are either in saturation or in cutoff region.
Under this condition only the gate drain Capacitance is C_{gd1}
- Because the signal swing is opposite in both terminal, effective Capacitance $C_{gd} = 2C_{GDOH}$.
Where, $C_{GDO} = \text{overlap Capacitance per unit width}$

Diffusion Cap (C_{db12}):

- Capacitance between drain and bulk is due to the reverse biased PN junction
- Simplified representation of diffusion cap, $C_d = k_{eq} C_{j0}$ where $C_{j0} = \text{junction Cap under zero-bias condition}$
- k_{eq} = multiplication factor

Wiring Capacitance!

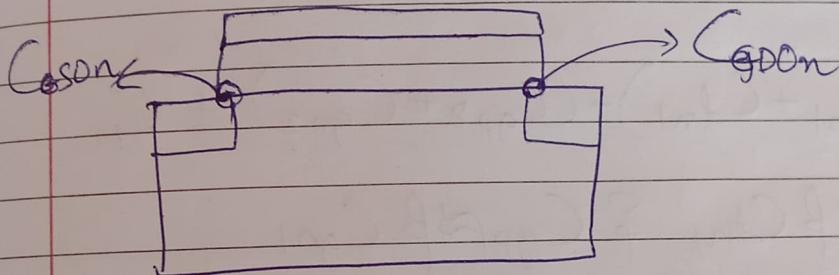
- This Capacitance appears due to length and width interconnecting wires
- This is function of the fanout from the driving gate.

→ Gate Capacitance of Fanout $C_{g3} \Sigma C_{g4}$
 This is equals to the total gate capacitance
 of the loading gates M3 & M4.

$$\begin{aligned} C_{\text{fanout}} &= C_{\text{gate(NMOS)}} + C_{\text{gate(PMOS)}} \\ &= [C_{GONP} + C_{GODN} + W_L n n_{ox}] + [C_{GOP} + C_{GODP} + W_P p_{ox}] \end{aligned}$$

C_{GONP} → Gate to Source Overlap n/p channel

C_{GODN} → Gate to Drain Overlap n/p channel



How to Improve design technique:-

- i) Reduce C_L : By Careful Layout design
- ii) Increase transistor sizing: Take care of self Loading as well
- iii) Increase V_{DD} : consider trade off energy dissipation

Self loading :- $W \uparrow$, Area \uparrow , Resistance \downarrow but we increase gate capacitance so we need care about that

Increase V_{DD} will have larger power dissipation

Optimal Value to PMOS to NMOS Ratio

- While improving the PMOS width improves $t_{P(H)}$ of the inverters by increasing the charging current, it also degrades the $t_{P(H)}$ by causing a large parasitic capacitance.
- If the optimum ratio β where

$$\beta = \left(\frac{L}{C}\right)_p / \left(\frac{L}{C}\right)_n$$

LKT

$$C_t = (C_{dp1} + C_{dn1}) + (C_{gp2} + C_{gn2}) + C_w$$

where $C_{dp1} \approx \beta C_{dn1}$ & $C_{gp2} \approx \beta C_{gn2}$

$$\text{so } C_t = [(1+\beta)(C_{dn1} + C_{gn2}) + C_w]$$

$$\Rightarrow t_p = \frac{0.69}{2} \left[(1+\beta)(C_{dn1} + C_{gn2}) + C_w \right] \left(\frac{R_{eqn}}{\beta} + \frac{R_{gp}}{\beta} \right)$$

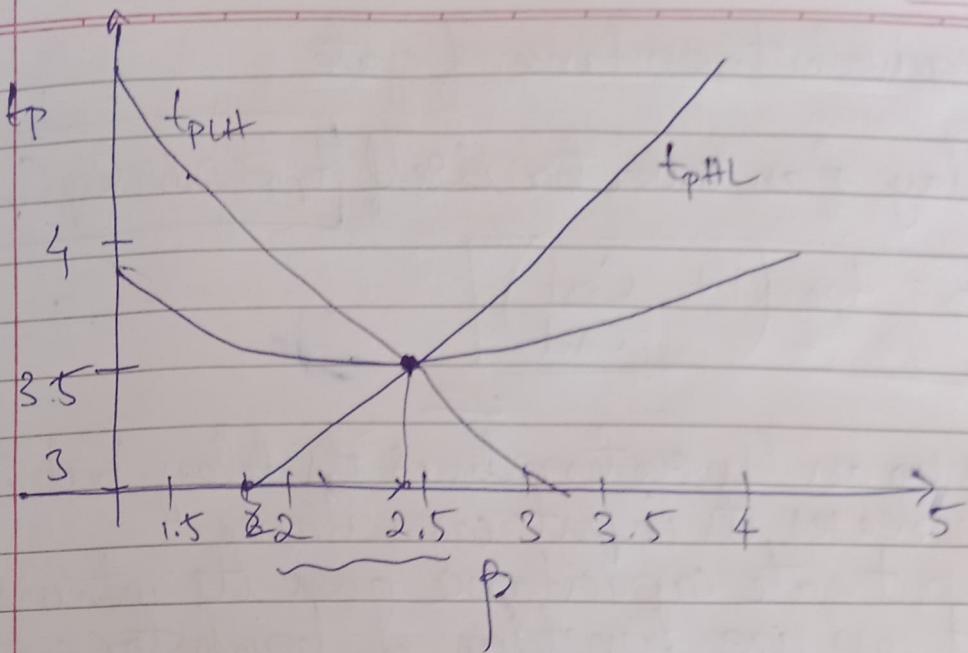
$$\text{take } \frac{\partial t_p}{\partial \beta} = 0$$

$$\frac{\partial t_p}{\partial \beta}$$

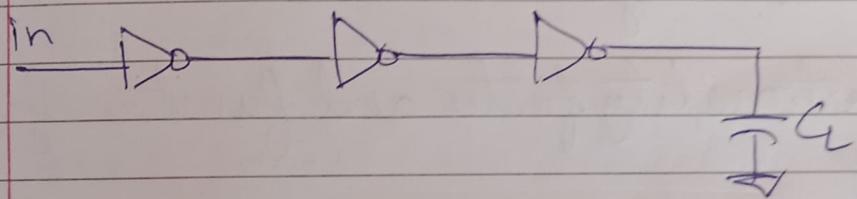
The optimal value of $t_p \beta$ would be

$$\beta_{opt} = \sqrt{\alpha \left(1 + \frac{C_w}{C_{dn1} + C_{gn2}} \right)}$$

$$\left[\alpha = \frac{R_{gp}}{R_{eqn}} \right]$$



Sizing of Inverter Chain



- For a given G_i , what is min t_p ?
- What is the no. of Inverter stages?
- What is optimum NMOS to PMOS ratio?

→ For an inverter $G_i = C_{int} + C_{ext}$

C_{int} = intrinsic output capacitance of the inverter associated with diffusion capacitance

C_{ext} = represents the fan out and wire capacitance

$$t_p = 0.69 R_{eq} (C_{int} + C_{ext})$$

$$t_p = 0.69 R_{eq} C_{int} \left(1 + \frac{C_{ext}}{C_{int}} \right)$$

R_{eq} = equivalent resistance of gate

The delay of an inverter itself $[t_{po} = 0.69 R_{eq} C_{int}]$

$$\boxed{t_p = t_{po} \left(1 + \frac{C_{ext}}{C_{int}}\right)}$$

- Sizing up an inverter reduces its delay, but it also increases its input capacitance
- The input gate capacitance and its intrinsic output cap are function of transistor gate size

$$C_{in} = \gamma C_g$$

γ → proportionality factor and function of technology

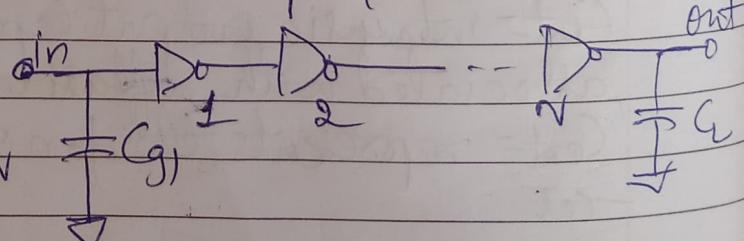
$$\boxed{t_p = t_{po} \left(1 + fH\right)}$$

where

$f = C_{ext}/C_{int}$ which is defined effective fan out

- The delay of the inverter is a function of the ratio between external load cap & its intrinsic cap

$$t_p = t_{p1} + t_{p2} + \dots + t_{pN}$$



$$t_p = \sum_{j=1}^N t_{pj} = t_{po} \sum_{j=1}^N \left[1 + \frac{C_{gin,j+1}}{\gamma C_{gin,j}} \right], C_{gin,N+1} = C_g$$

- Delay equation has $N+1$ unknowns
- Minimize the delay, find $N+1$ partial deri and solve for $S_{tp} = 0$

$$\delta C_{g,i}$$

$$\Rightarrow C_{g,j+1}/C_{g,i} = C_{g,i}/C_{g,j-1}$$

- Size of each stage is geometric Mean of 2 neighbours.

$$C_{g,i} = \sqrt{(C_{g,j+1})(C_{g,j-1})}$$

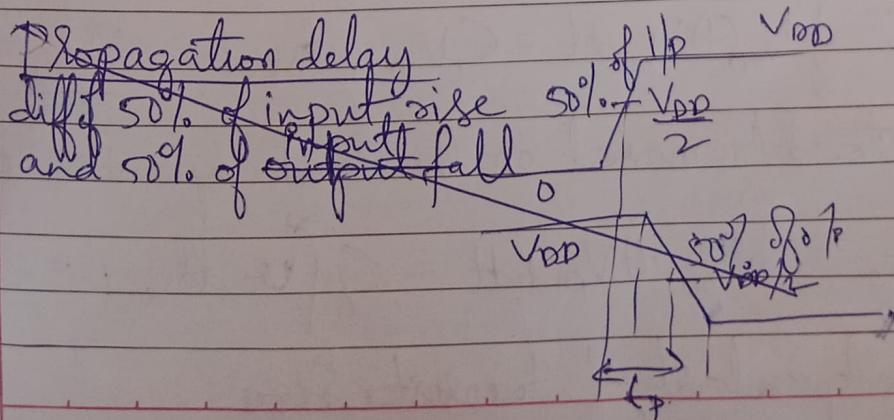
- Optimum size of each inverter is the geometric mean of its neighbours
- This means that each inverter is sized up by the factor 'f' w.r.t. the preceding gate, that the same effective fan-out ($f_i = f$)

$$f = \sqrt[N]{C_i/C_{g,1}} = \sqrt[N]{F}$$

The Minimum delay through the chain is

$$t_p = N t_{po} \left(1 + \sqrt[N]{F^N / r} \right)$$

$F \rightarrow$ represents overall clk Fanout

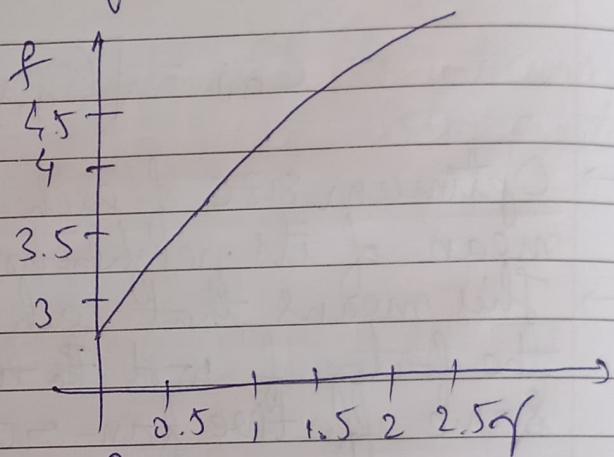


Choosing Right No of Stages.

→ The optimum value of right no of stages can be found by differentiating the min delay expression w.r.t. no. of stages and setting the result to 0

$$\text{We get } \gamma + \sqrt{F} - \frac{\sqrt{F}}{n} = 0$$

→ In common practice, optimum fanout could be selected as 4



Power Dissipation - Dynamic Power

→ When Inverter switches continuously it consumes dynamic power

1) Dynamic Power

2) Static Power

3) Short ckt Power

→ Power dissipation during switching activity

→ Energy taken from supply voltage E_{DD}

$$E_{DD} = \int_{-\infty}^{\infty} i_{V_{DD}}(t) V_{DD} dt = C_L V_{DD} \int_{-\infty}^{\infty} dV_{out} = C_L \underline{V_{DD}}^2$$

→ Energy stored/removed on the load capacitor = E_C

$$E_C = \int_{-\infty}^{\infty} i_{V_{DD}}(t) V_{out} dt = C \int_{-\infty}^{\infty} V_{out} dV_{out} = \underline{C} \underline{V_{DD}}^2$$

→ This is independent of transistor size

If the switching activity is $f_{\text{to-1}}$ times per sec

$$P_{\text{dyn}} = \sum V_{\text{DD}}^2 f_{\text{to-1}}$$

Power dissipation \rightarrow Direct path currents

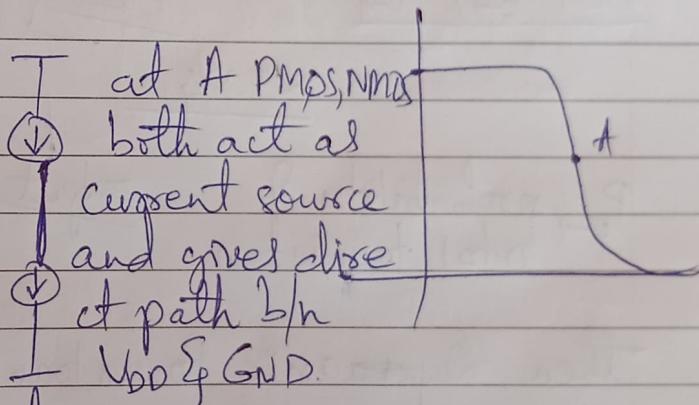
\rightarrow Finite value of rise time (t_r) & fall time (t_f) of logic signal
 - It causes a direct path between V_{DD} & GND while PMOS and NMOS are conducting simultaneously

$$E_{\text{dp}} = V_{\text{DD}} \frac{I_{\text{peak}} t_{\text{sc}}}{2} + V_{\text{DD}} \frac{I_{\text{peak}} t_{\text{sc}}}{2} = t_{\text{sc}} V_{\text{DD}} I_{\text{peak}}$$

\rightarrow Avg power consumption

$$P_{\text{dp}} = t_{\text{sc}} V_{\text{DD}} I_{\text{peak}} f$$

t_{sc} = time of both devices are conducting



Static Power Dissipation

This is the steady power dissipation, when no switching activity is present

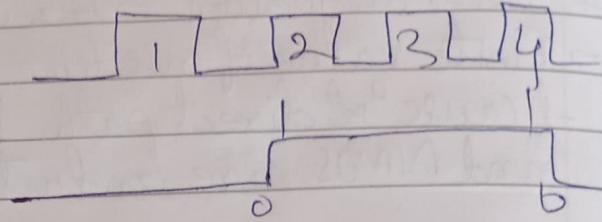
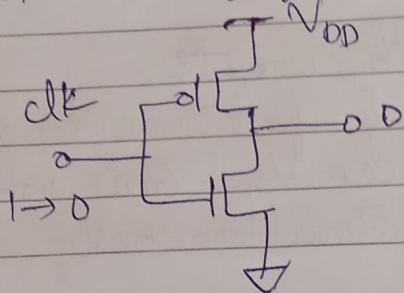
$$P_{\text{st}} = I_{\text{st}} V_{\text{DD}}$$

Sources of leakage current

- Reverse bias diode junction of transistors, located b/n source or drain & body. Thermally generated carriers may affect the junction leakage current
- Subthreshold current when $V_G < V_T$
- Choice of V_T is trade off between the performance and static Power consumption

$$P_{\text{tot}} = P_{\text{dyn}} + P_{\text{SC}} + P_{\text{stat}} = (C_L V_{DD}^2 + V_{DD} I_{\text{peak}} t_{SC}) f + V_{DD} T_{\text{lat}}$$

$\lambda_{0 \rightarrow 1}$ (Activity factor)



Here the output is showing transition from 2 clock Cycle and 1 to 0 transition at 4th clock Cycle

$$\text{then } \lambda = \frac{1}{4}$$

$$\Rightarrow f_{0 \rightarrow 1} = P_{b \rightarrow i} \cdot \lambda_{0 \rightarrow 1}$$



$$\lambda = \frac{3}{4}$$



$P_{0 \rightarrow 1}$ probability of output
0 to 1 transition

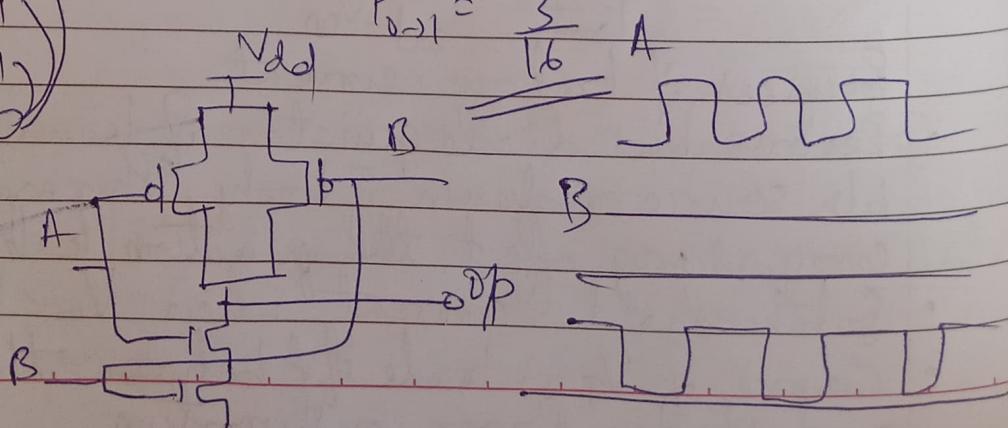
Then Dynamic Power transition = $C_L V_{DD}^2 \cdot P_{0 \rightarrow 1} \cdot \lambda_{0 \rightarrow 1}$

NAND-2

A	B	C
0	0	1
0	1	1
1	0	1
1	1	0

$$P_0 = \frac{1}{4}, P_1 = \frac{3}{4}$$

$$P_{0 \rightarrow 1} = \frac{3}{16}$$



(Contd.)

Power delay product and Energy delay product

→ Power delay Product :- It represents the average consumption per switching event

$$PDP = C_L V_{DD}^2 f_{max} t_p = \frac{C_L V_{DD}^2}{2} , \text{Bcs } f_{max} = \frac{1}{2 t_p}$$

This value can be arbitrary by changing the V_{DD} .

→ Energy - Delay Product :- This product performance metric combines the performance and energy

$$EDP = PDP \times t_p = \frac{C_L V_{DD}^2}{2} t_p$$

→ So to Reduce dynamic Power
Reduce V_{DD}

Reduce Power dissipation

→ Reduce Supply Voltage (V_{DD})

→ Reduce switching activity (α)

→ Reduce Physical capacitance (C_L)