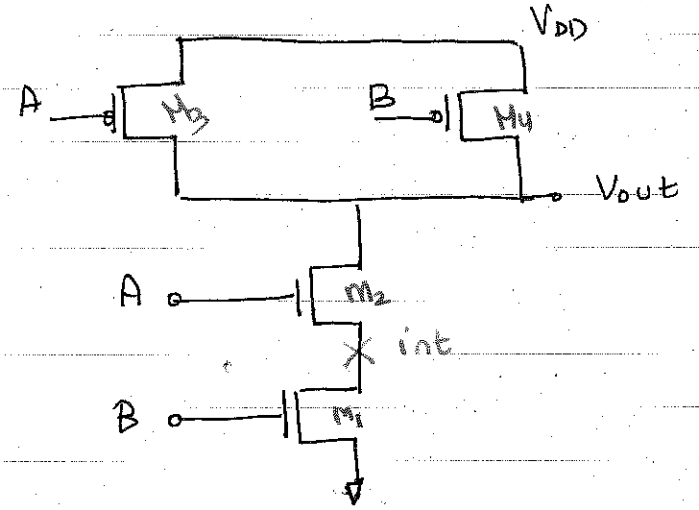
STATIC PROPERTIES OF NAND LOGIC GATE

Unlike the inverter, the switching threshold, V_m, of the NAND gate depends on the input pattern.

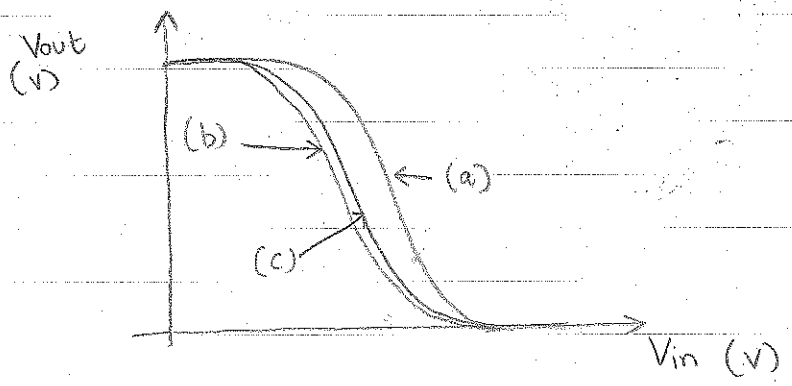


Consider three specific cases :-

(a) A = B = 0 → 1

(b) A = 1, B = 0 → 1

(c) B = 1, A = 0 → 1



CASE (a) When both inputs are initially "0", then both PMOS devices are initially ON. Hence, the pull-up in this case is quite strong.

When both A, B ⇒ 0 → 1 simultaneously, then the switching threshold will be close to V_{DD} since the pull-up is quite strong.

Cases (b) & (c)

When only one of the inputs is initially "0" and then it transitions to "1", the switching threshold will be closer to Ground, because in this case pull-up is not that strong.

The difference between (b) & (c) comes from which of the two transistors, M1 or M2, undergoes switching.

M_1 has $V_{BS} = 0$ since its source is always grounded.

M_2 has $V_{BS} \neq 0$ since its source labeled as "int" is not always at zero.

Therefore, V_T of M_2 is actually have higher than V_T of M_1. This means, "M_2" is a weak pull-down device.

Hence, when "M_2" switches, the switching threshold will be slightly higher (more toward V_{DD}) than when "M_1" switches.

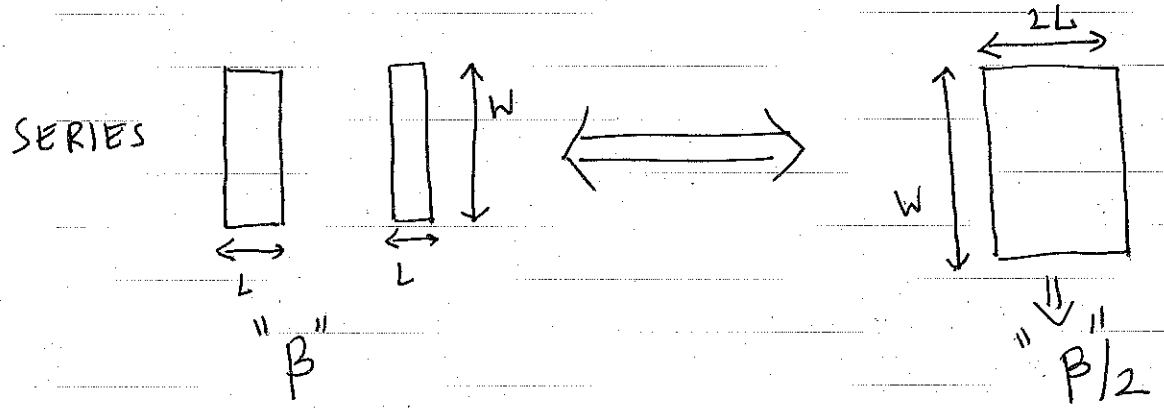
NOTE:- The strength of the transistor depends on $(V_{GS} - V_T)$. So when $V_T \uparrow$, the strength of the transistor $\downarrow$.

DIFFERENCE BETWEEN SERIES AND PARALLEL DEVICES

SERIES

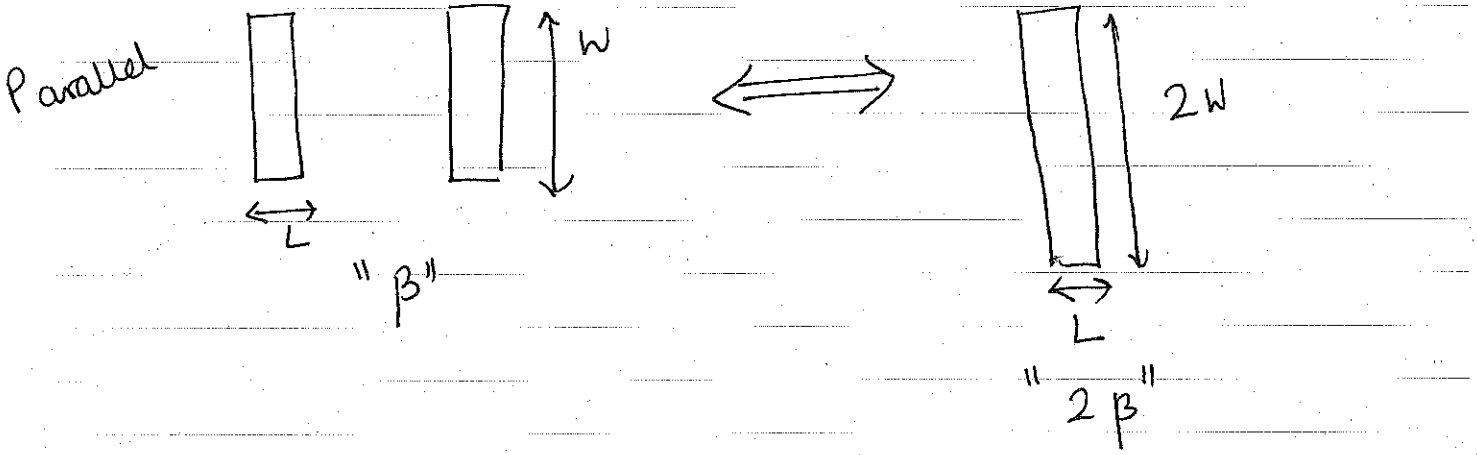
When devices are connected in series, resistance is more.

That is, two devices connected in series may be treated as one device with half the value of 'β' of individual devices.

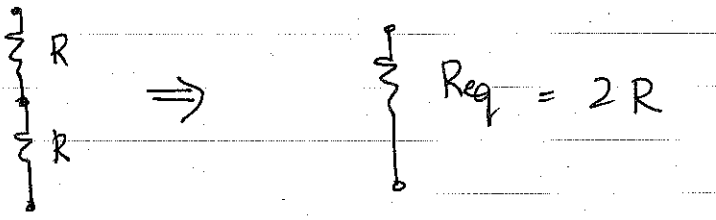


PARALLEL

When devices are connected in parallel, then the overall resistance is reduced.



The analogy with linear resistors is simple.



As you can see, series connected resistors have a higher equivalent resistance.

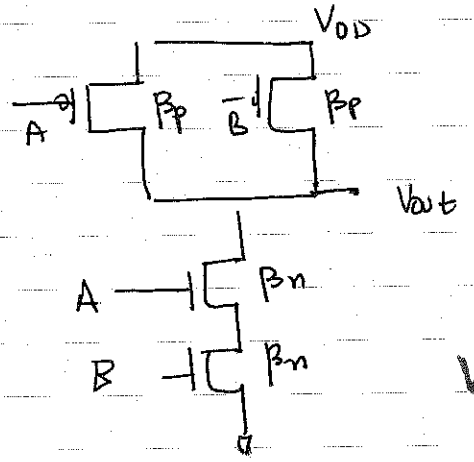
Parallel connected resistors have a lower equivalent resistance.

SWITCHING THRESHOLD OF NAND 2

While we know that Vm is dependent on the input pattern, we can easily calculate Vm when both inputs are switching simultaneously.

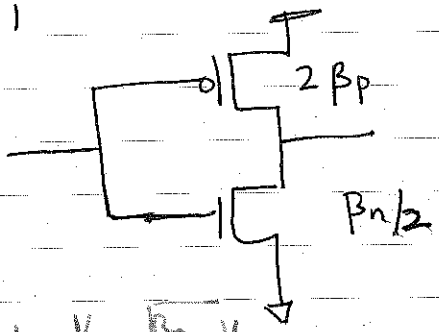
To do that, we can replace the parallel connected PMOS with one PMOS having higher "β" value.
Likewise, we can replace the series connected NMOS, with only one NMOS have a lower β.

[Recall: Parallel connected transistors have a lower resistance while series connected transistor have a higher effective resistance. ⇒ The caveat is that this explanation works when all the transistors (whether parallel or in series) are simultaneously ON.]



A, B = 0 → 1

⇒



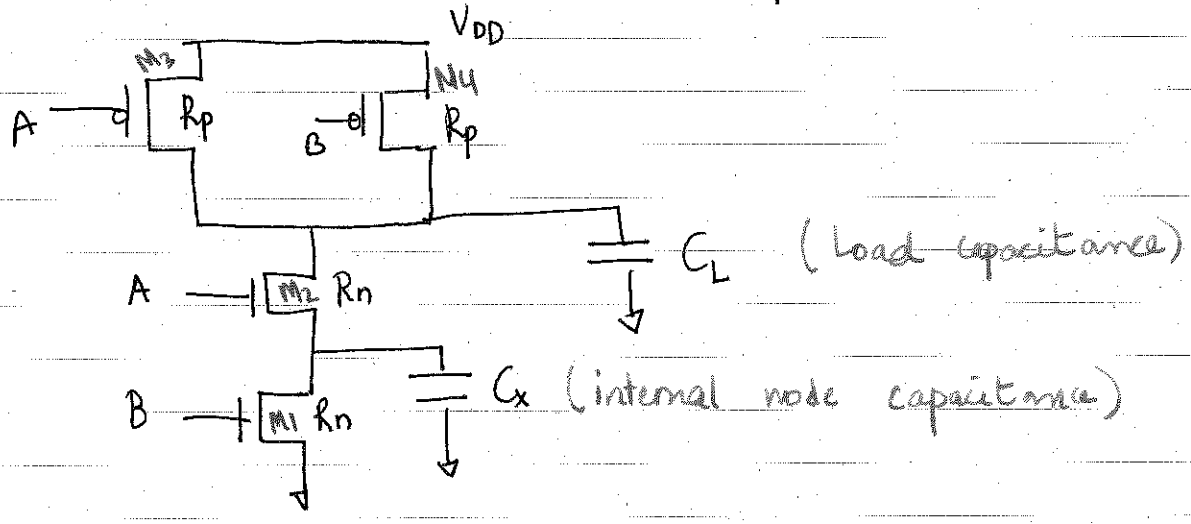
$$V_m = \frac{V_{DD} - |V_{tp}| + \frac{1}{2}\sqrt{\frac{\beta_n}{\beta_p}}\,V_{DD}}{1 + \frac{1}{2}\sqrt{\frac{\beta_n}{\beta_p}}}$$

We already know the switching threshold of inverter with a given 'β' For NMOS and PMOS devices. All we have to do is replace appropriate 'β' in the expression.

SWITCHING CHARACTERISTICS OF NAND-2

Switching characteristics are also dependent on input patterns.

But let's say, we have sized our transistors such that each PMOS device has eff. resistance = R_p while each NMOS device has eff. resistance = R_n



Where does "C_L" come from?

C_L comes From a) C_{db} of M_3, M_4, M_2.
b) C_{miller} of m_2, m_3, m_4
c) $C_{Fan\text{-}out}$
d) C_{wire}

C_{db} :- drain-body junction capacitance
C_{miller} :- $\sum 2(C_{gd})$ of m_2, m_3, m_4 → Overlap capacitance.
$C_{Fan\text{-}out}$:- This corresponds to the next-stage input cap.
C_{wire} :- This is the wire cap.

Where does C_x come From?

C_x comes only because of the parasitics of M_1 and M_2.

The key contributions to C_x are from

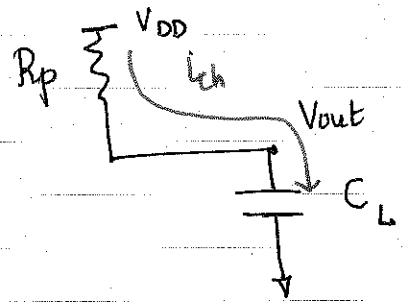
a) C_{miller} of M_1, M_2

b) C_{sb} of M_2

c) C_{db} of M_1

C_{miller} = 2 C_{gs} of M_2 + 2 C_{gd} of M_1 } C_{gs}, C_{gd} ⇒ overlap capacitances.

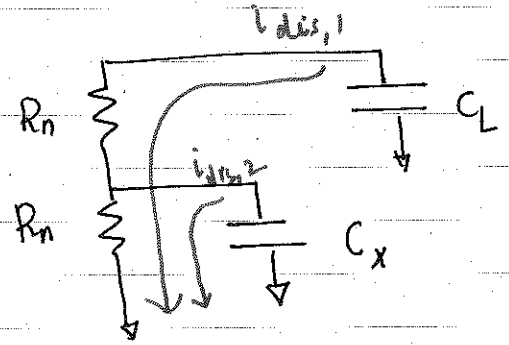
Low-to-high transition



i_{ch} = charging current

$$t_{pLH} = 0.69 R_p C_L$$

High-to-Low transition (Both NFETs must be ON)



$$t_{pHL} = 0.69 [2R_n C_L + R_n C_x]$$

For high-to-low, both NFETs must be ON.

And two capacitances have to be discharged.

(a) The Load capacitance, C_L

(b) The internal node capacitance, C_x

'C_L' gets discharged through both NFETs, while 'C_x' gets discharged through the lower NFET ("M_1").

How should we size transistors for a NAND2 gate?

First, we focus only on the worst-case scenario for delay.

The worst case scenario for delay is when only one of the two inputs switches From 1 to 0 for a 0 to 1 output transition.

For "1" to "0" output transition, both inputs must switch from "0" to "1".

$t_{pLH} = 0.69 R_p C_L$

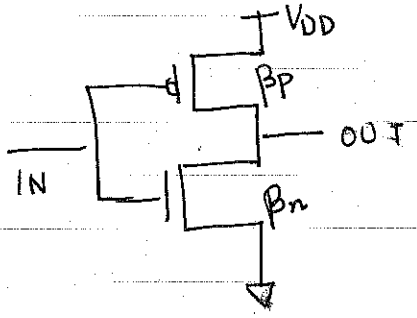
$t_{pHL} = 0.69 (2 R_n C_L)$ {ignore C_x}.

↑ consider only worst case delays.

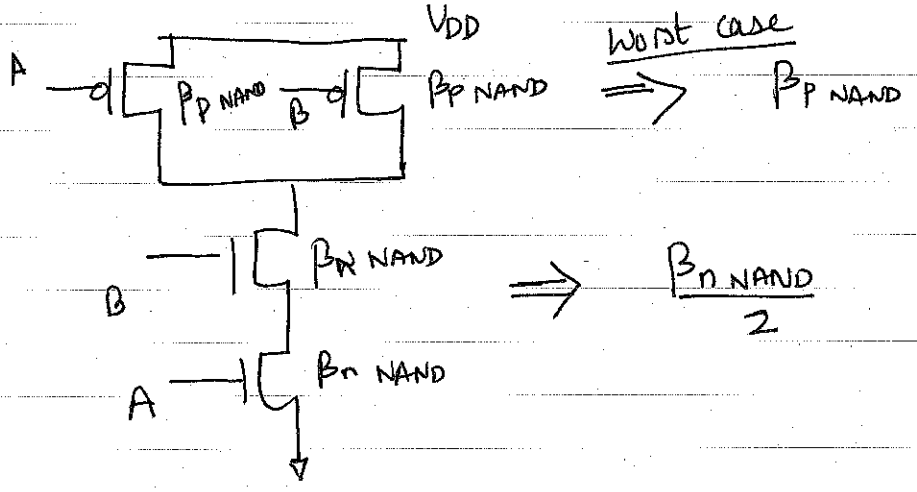
FOR SIZING, a) find the size of the PFET in NAND-2 that gives the same t_{PLH} as in the minimum sized inverter.

b) find the size of the NFET in NAND-2 that gives the same t_{PHL} as in the minimum sized inverter.

INVERTER SIZING



NAND-2 SIZING

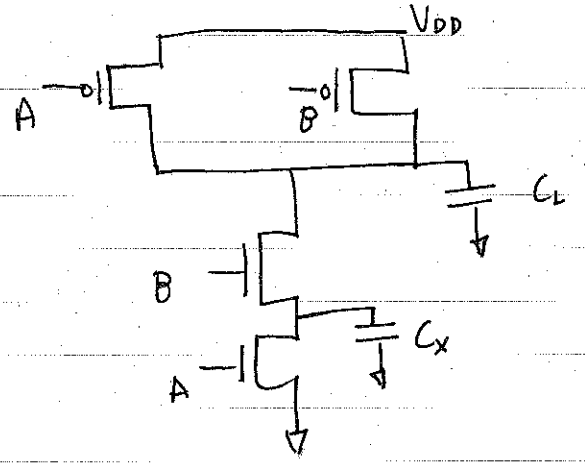


$$\frac{\beta_{n,NAND}}{2} = \beta_n \Rightarrow \boxed{\beta_{n,NAND} = 2\beta_n}$$

$$\beta_{p,NAND} = \beta_p \Rightarrow \boxed{\beta_{p,NAND} = \beta_p}$$

DEPENDENCE OF INPUT PATTERN ON DELAY

To find out the dependence of t_{pHL} and t_{pLH}, we must find the input patterns that need to charge or discharge the internal node capacitance, C_x.



Cases For LOW to High @ OUTPUT

(a) A, B $\Rightarrow 1 \rightarrow 0$

(b) A = 1, B $\Rightarrow 1 \rightarrow 0$

(c) B = 1, A $\Rightarrow 1 \rightarrow 0$

Case (a)

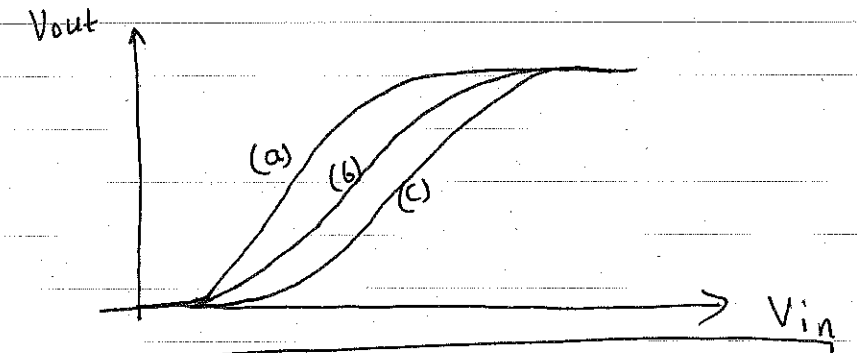
When A, B = $1 \rightarrow 0$ then the pull-up is "strong" as both PMOS devices turn on simultaneously. Hence, t_{pLH} will be the fastest in this case.

Case (b)

When A = 1, B $\Rightarrow 1 \rightarrow 0$, then the PMOS only needs to charge C_L.

Case (c)

When B = 1, A $\Rightarrow 1 \rightarrow 0$, then PMOS needs to charge both C_L and C_x.



$t_{pLH}(a) < t_{pLH}(b) < t_{pLH}(c)$

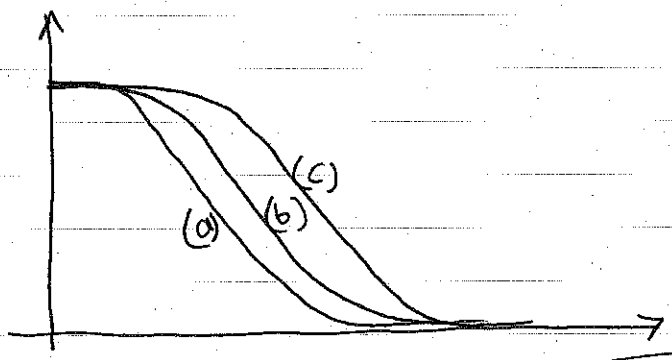
Likewise, we construct cases for high to low output transition.

Cases for high to low transition @ O/P

(a) $A = 1$, $B \Rightarrow 0 \rightarrow 1$ → internal node already discharged

(b) $B = 1$, $A \Rightarrow 0 \rightarrow 1$ → internal node needs to be discharged.

(c) $A, B \Rightarrow 0 \rightarrow 1$ → strong initial pull-up so maximum delay.

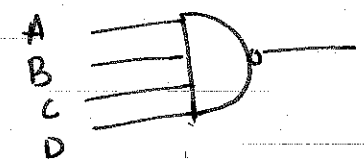


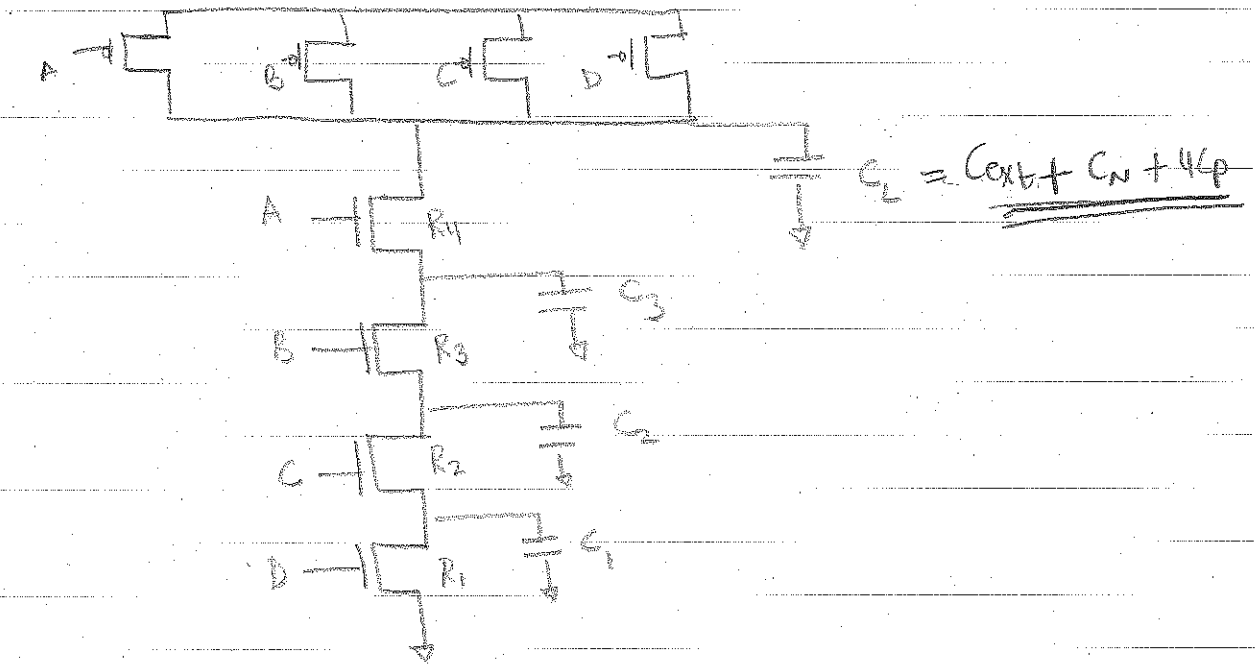
$t_{pHL}(a) < t_{pHL}(b) < t_{pHL}(c)$

FAN - IN CONSIDERATION

Ques:- How does the delay of a complex logic gate depend on the no. of inputs or Fan-in?

Consider a **four input NAND gate**





What does C_L consist of?

$C_L = C_{ext} + C_N + 4C_P$

C_N = Parasitic From NFET.

C_P = Parasitics From PFET.

C_{ext} = external load capacitance.

$$t_{pHL} \propto R_1 C_1 + (R_1+R_2) C_2 + (R_1+R_2+R_3) C_3 + (R_1+R_2+R_3+R_4) C_L$$

If there are "N" inputs and each NFET resistance is denoted as R_{NMOS} and each parasitic capacitance is denoted as C_N, we can write

$$t_{pHL} \propto R_{NMOS} (C_N + 2C_N + 3C_N + \cdots (N-1)C_N) + N R_{NMOS} C_L$$

$$\Downarrow$$

$$R_{NMOS} C_N (1+2+3+ \cdots + (N-1)) + N R_{NMOS} C_L$$

$$\Downarrow$$

$$0.5 N (N-1)$$

$$C_L = C_N + N C_p + C_{ext}$$

- C_N: parasitic from NMOS
- $N C_p$: parasitic from PMOS
- C_{ext}: external load

$$t_{pHL} \propto 0.5N(N-1) C_N R_{NMOS} + N R_{NMOS} (C_N + N C_p + C_{ext})$$

$$t_{pHL} \propto 0.5 N (N-1) C_N R_{NMOS} + N R_{NMOS} C_N + N^2 R_{NMOS} C_p + N R_{NMOS} C_{ext}$$

$$t_{pHL} = a_1 N + a_2 N^2$$

N = Fan-in

so t_{pHL} varies quadratically with Fan-in.

C_{ext} depends on Fan-out.

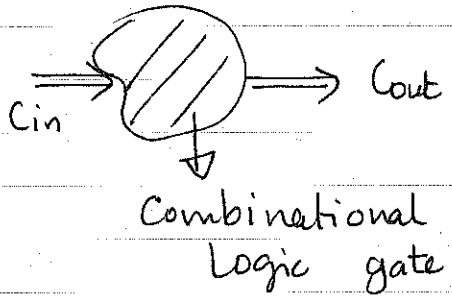
That is $C_{ext} \propto$ F.O.

so overall $t_{pHL} = a_1 F.I. + a_2 (F.I.)^2 + a_3 (F.O.)$

Quadratic dependence on Fan-in.

Linear dependence on Fan-out.

Logical Effort and Electrical effort



Logical effort is defined as $\boxed{g = \frac{C_{in}}{C_{ref}}}$

C_{in} = input capacitance of the gate

C_{ref} = input capacitance of the "reference" inverter.

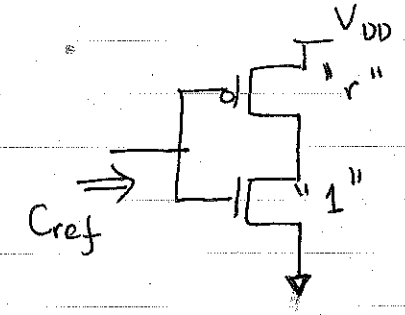
Electrical effort is defined as $\boxed{h = \frac{C_{out}}{C_{in}}}$

Note, we had previously called electrical effort as effective fan-out of the logic.

Since logical effort needs the capacitance of a reference inverter, we must first define the ref. inverter.

A REFERENCE INVERTER HAS ITS NMOS and PMOS SIZED SUCH THAT IT HAS SAME Resistance FOR THE PULL-UP AND PULL-DOWN.

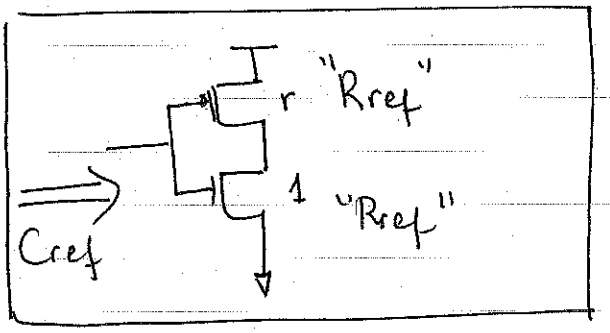
It is also the minimum sized inverter for the given technology.



We size the NMOS as "1" which means $W_N = W_{min}$ allowed by the technology.

"r" is chosen such that $\beta_p = \beta_n$ or $R_n = R_p = R_{ref}$.

Typically $r \simeq 2$ for 0.25 μm tech.

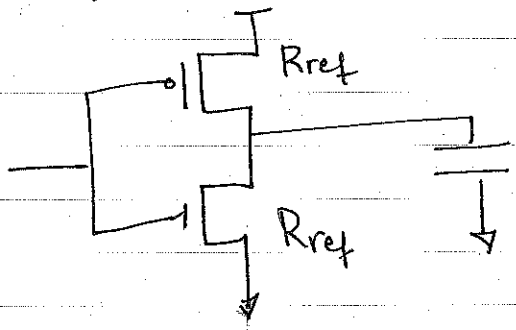


REFERENCE INVERTER.

$$C_{in} = C_{ref} = (1 + r) C_n$$

Note :- The logical effort of the reference inverter is UNITY. That, in fact, comes from the definition of the reference inverter.

Delay of the Reference inverter



$$C_{LOAD} = \left(C_{FET\_ref} + C_{out} \right)$$

Lets say the net parasitic capacitance at the output node is C_{FET_ref}.

$$C_{load} = C_{Fet\_ref} + C_{out}$$

$$d_{abs} = k R_{ref} \left(C_{Fet\_ref} + C_{out} \right)$$

$$= k R_{ref} C_{Fet\_ref} + k R_{ef} C_{out}$$

$$= \underbrace{k R_{ref} C_{ref}}_{\tau_{ref}} \left(\frac{C_{Fet\_ref}}{C_{ref}} \right) + \underbrace{k R_{ref} C_{ref}}_{\tau_{ref}} \left(\frac{C_{out}}{C_{ref}} \right)$$

$$d_{abs} = \tau_{ref} \left(\frac{C_{Fet\_ref}}{C_{ref}} \right) + \tau_{ref} \left(\frac{C_{out}}{C_{ref}} \right)$$

$\left(\frac{C_{Fet\_ref}}{C_{ref}} \right) \Rightarrow p$

$\left(\frac{C_{out}}{C_{ref}} \right) \Rightarrow h$

$$d = \frac{d_{abs}}{\tau_{ref}} = p + h$$

p: parasitic

h: electrical effort

Note $\tau_{ref} = k R_{ref} C_{ref}$ is NOT the delay of the reference inverter.

Lets say we have an inverter which is "s" times larger than the min. sized inverter given in the technology. What will the delay of this inverter be?

$$d_{abs} = k R \left(C_{FET} + C_{out} \right)$$

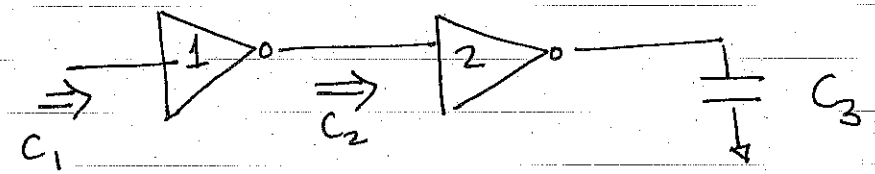
$$= k \frac{R_{ref}}{s} \left(s C_{Fet\_ref} + C_{out} \right)$$

$$= k R_{ref} C_{Fet\_ref} + k R_{ref} \frac{C_{out}}{s}$$

$$= \underbrace{k R_{ref} C_{ref}}_{\tau_r} \underbrace{\left[\frac{C_{Fet\_ref}}{C_{ref}} \right]}_{p} + \underbrace{k R_{ref} C_{ref}}_{\tau_r} \underbrace{\left[\frac{C_{out}}{s C_{ref}} \right]}_{h}$$

$$\boxed{d = \frac{d_{abs}}{\tau_r} = p + h}$$

2 inverter chain



$D = d_1 + d_2$

$d_1 = p_1 + h_1$

$d_2 = p_2 + h_2$

$D = (p_1 + h_1) + (p_2 + h_2)$

Path electrical effort :- $H = \frac{C_3}{C_1} = \left(\frac{C_3}{C_2}\right)_{h_1} \left(\frac{C_2}{C_1}\right)_{h_2}$

$H = h_1 h_2$

So $h_2 = (H/h_1)$

Substituting 'h_2' in the definition of D, we obtain

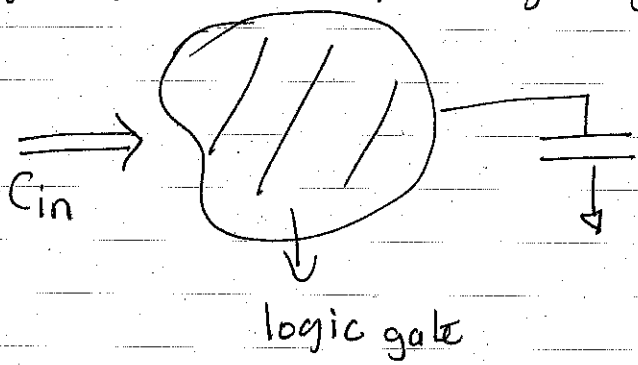
$D = (p_1 + h_1) + (p_2 + H/h_1)$

What is the optimal value of h_1?

$\frac{\partial D}{\partial h_1} = 0 \Longrightarrow h_{1opt}$ (to minimize delay)

$h_{1,opt} = \sqrt{H} \Longrightarrow \boxed{h_{1,opt} = h_{2,opt} = \sqrt{H}}$

Delay of a complex logic gate with $R_{up} = R_{down} = R_{ref}$



Assume we have sized the transistors such that the pull-up and pull-down resistances match R_{ref}

$$d_{abs} = k R_{ref} \left(C_{FET} + C_{out} \right)$$

$$= \underbrace{k R_{ref} C_{ref}}_{\tau_{ref}} \left(\underbrace{\frac{C_{FET}}{C_{ref}}}_{p} \right) + \underbrace{k R_{ref} C_{ref}}_{\tau_{ref}} \left[\frac{C_{out}}{C_{ref}} \right]$$

$$(?)$$

$$\frac{C_{out}}{C_{ref}} = \frac{C_{out}}{C_{in}} * \frac{C_{in}}{C_{ref}}$$

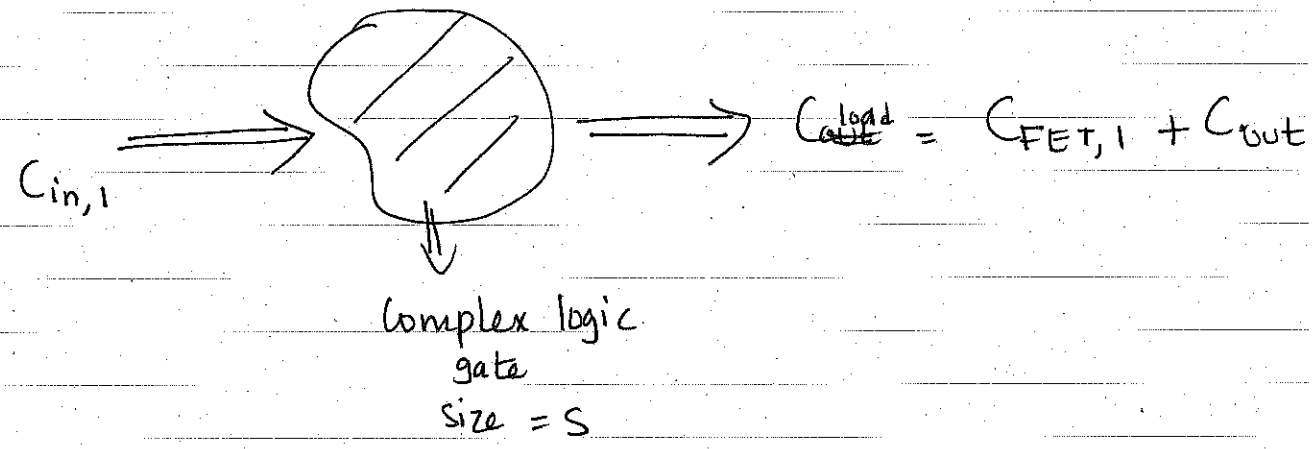
$$\Downarrow \qquad \Downarrow$$

$$h \qquad g$$

$$d_{abs} = \tau_{ref} p + \tau_{ref} (gh)$$

$$\boxed{d = \frac{d_{abs}}{\tau_{ref}} = p + gh}$$

Delay of a complex logic gate with a bigger size "S"



$$C_{in,1} = S\,C_{in}$$

$$C_{FET,1} = S\,C_{FET}$$

$$R_1 = \frac{R_{ref}}{S}$$

$$d_{abs} = k\,R_1\left(C_{Fet,1} + C_{out}\right)$$

$$= k\,\frac{R_{ref}}{S}\left(S\,C_{FET} + C_{out}\right)$$

$$= \underbrace{k\,R_{ref}\,C_{ref}}_{\tau_{ref}}\left(\underbrace{\frac{C_{FET}}{C_{ref}}}_{p}\right) + \underbrace{k\,R_{ref}C_{ref}}_{\tau_{ref}}\left(\underbrace{\frac{C_{out}}{S\,C_{ref}}}_{(?)}\right)$$

$$\frac{C_{out}}{S\,C_{ref}} = \left(\frac{C_{out}}{S\,C_{in}}\right)\cdot\left(\frac{C_{in}}{C_{ref}}\right)$$

$\frac{C_{out}}{S\,C_{in}} \Rightarrow h$, $\frac{C_{in}}{C_{ref}} \Rightarrow g$

$$d_{abs} = \tau_{ref}\,p + \tau_{ref}(g\,h)$$

$$\boxed{d = \frac{d_{abs}}{\tau_{ref}} = p + gh}$$

POINTS TO REMEMBER

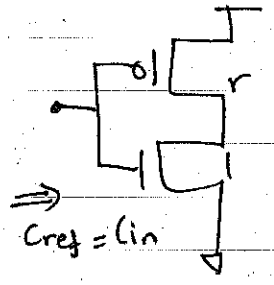
DELAY OF A COMPLEX LOGIC GATE IS ALWAYS

$$d = \frac{d_{abs}}{\tau_{ref}} = p + gh$$

$\tau_{ref} = k R_{ref} C_{ref}$

R_{ref} → reference resistance of a minimum-sized inverter.

C_{ref} → reference i/p capacitance of a minimum-sized inverter.



$C_{ref} = (1+r) C_{gn}$

$C_{gn} = C_{ox} W_n L_n$

Note:- The delay of a minimum-sized inverter is NOT τ_{ref}.

$$d_{min.\ inverter} = \tau_{ref}(p+h)$$

Also ~~ke~~ note, 'p' will vary w/ the logic circuit you're looking at.

For example, $p(\text{NAND2}) > p(\text{INV.})$

Logical effort, g, does not depend on the size. Size only plays a role in the electrical effort calculation.