

TECHNOLOGY SCALING

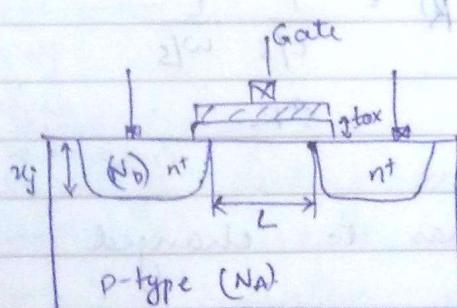
Effect of small size.

- Coupling, crosstalk b/w 2 interconnects as they are near.
- Reliability
- Power consumption
- Noise

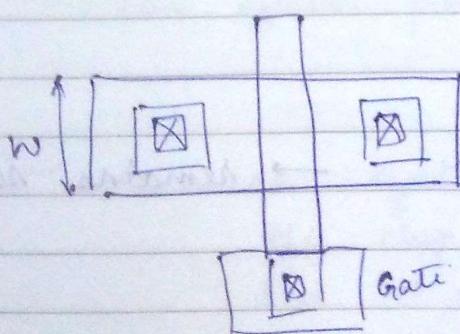
By scaling, device delay ↓ but interconnect delay ↑

Weak Inversion Region. → occurs when low voltage from nearby interconnect or is provided to other interconnect → device turns on.

Variation in supply voltage → Reduce signal integrity
 ↓ " power "



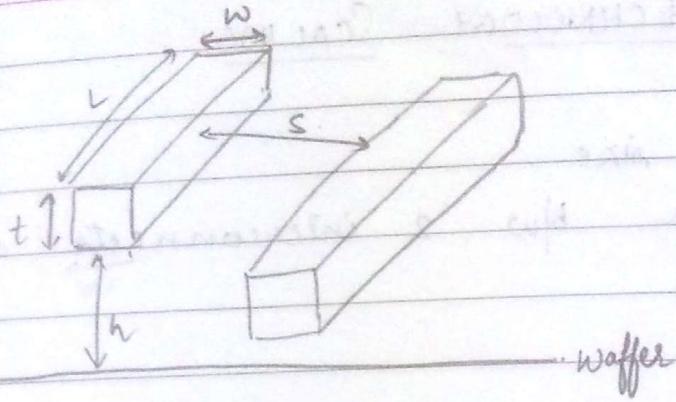
Mosfet size along with supply V_{DD} & V_{TO}
 $S > 1$.



- Constant field Scaling
- Constant voltage Scaling

$C_{ox} \rightarrow s C_{ox} \Rightarrow$ no. of charges
 should ↑ ⇒ N_A & N_D
 are scaled up by s .

If Voltage kept constant while scaling → then no.
 of carriers should be more ⇒ N_A & N_D are scaled
 up by s again.



	Before	After scaling by s .
w	w	w/s
Spacing s	s	s s/s
Thickness t	t	t/s
Oxide height h	h	h/s

- Resistance of interconnect (R) = $\frac{s L}{t w} \rightarrow$

sheet resistance R_0
(without scaling)

- Resistance after scaling (R) = $\frac{s}{t/s} \frac{L}{w/s}$ → this is physical
 $= s^2 R$; distance
 ∴ cannot be scaled.

- To scale L , design has to be changed.

- $C = \frac{\epsilon_{ox}}{h} w L$

$C = \frac{\epsilon_{ox}}{h/s} w/s L = C \rightarrow$ remains same

RC delay = $R \times C$

$= s^2 R C$

delay lies
as capacitance lies
↑

Page No.:		
-----------	--	--

- Low k dielectric \rightarrow used for interconnects
High " " " \rightarrow " " devices.
- Two field present - $\frac{V_{DS}}{L}$ (lateral field),
 $\frac{V_{GS}}{t_{ox}}$ (longitudinal field)
- In voltage scaling $\rightarrow E$ rises by S factor
so even if $V_{GS} < V_T$ due to E current flows
 \Rightarrow "Drain induced Barrier lowering" (DIBL) occurs.
 $V_{GS} > V_T$,
- When lateral & long. E rises \rightarrow mobility & velocity of e^- rises \rightarrow mobility degradation & velocity saturation occurs.
- Due to channel length modulation, resistance becomes finite. Ideally it's ∞ . as $R = \frac{\Delta V}{\Delta I} = 0$.

19/4

Device Scaling

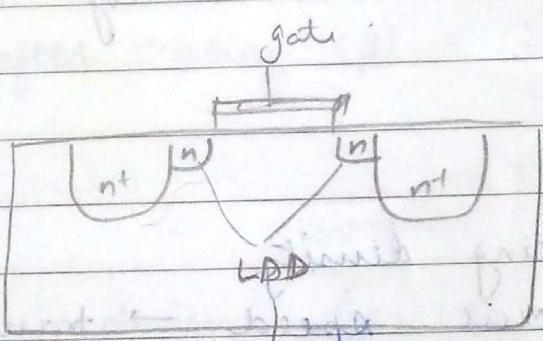
- high packing density
 - " circuit speed \rightarrow transistor performance \uparrow
 - " system speed \rightarrow small chip-to-chip delay
 - low system power consumption (static power)
- $$P = \frac{1}{2} \frac{V_{DD}}{S} \frac{I_{DS}}{S} = \frac{P}{S^2}$$
- enhanced system reliability — few leads & off chip connection as connectors distance rises \Rightarrow noise \downarrow
 - Heterogeneous system integration

- Deep sub-micron technology — below $100\text{ }\mu\text{m}$
 - Threshold roll-off → extra voltage req. at gate to reduce depletion region at drain.
 - $\frac{I_{on}}{I_{off}}$ → should be high \Rightarrow leakage current is less.

Improve performance beyond 90 nm.

- Stain engineering → change gate ~~oxide~~ material.
 - Improving the device structure — addition of extra gate → if 1 does not work other can be used
 - high k and metal gate to reduce leakage.
 - Non-uniform channel doping:
 - | LDD (lightly doped drain) structure

↓ Drain Induced Barrier lowering (DIBL)



- collect hot carrier
 - decrease leakage current
 - improves channel quality
 - ↓ depletion region width

⇒ ↓ 4th roll-off.

INTERCONNECTS

- Problems with Interconnects

- (1) IR drop
- (2) $\frac{dI}{dt}$ noise.
- (3) crosstalk

- More the voltage → more will be the width of interconnect
 This is the reason ~~why~~ for global interconnects to
 be of larger width.

~~Range~~ width ranges from μm to nm .

- Wire capacitance $C_{\text{tot}} = C_{\text{top}} + C_{\text{bottom}} + 2C_{\text{adjacent}}$

- Control cross-talk by -

- (1) Increase spacing b/w adjacent lines.
- (2) Shield wires → 2 signal carrying wires are not kept adjacent to each other.
- (3) Ensure neighbour switch at different time → if 2 signal lines are adj. then ~~only~~ only 1 is ON at a time.
- (4) Cross-talk cancellation.

Pitch → wire width + spacing

Scaling types :-

- ① Ideal
- ② Quasi-ideal
- ③ Scaling based on constant resistance
- ④ " " " thickness

- Die size & wafer size has scaled up by S_c factor due to increase in no. of MOS — as SOC are now fabricated.
- Global interconnects ^{length} has scaled up by S_c factor
 - connects 2 functional units, power supply lines, timing & control lines
 - ⇒ delay has increased.

26/2

Interconnect model — Short circuit

C model

LC model

RLC model

These models are considered bcoz speed of clock ^{has} increased from 3M to 3GHz. Also the ~~the~~ length of interconnect has to be comparable to λ .

Width of interconnect $\propto \frac{1}{\text{delay}}$ ~~delays~~.

Area of interconnect $\propto \frac{1}{\cancel{\text{delays}}} \text{BW}$

Reduce leakage — by changing device structure
using no. of gates.

$$\text{dynamic power dissipation} = f C V^2$$

- length of wire is multiple of wavelength of signal
⇒ inductance effect occurs.

3/17.
Reducing device size

⇒ gate capacitance C_{gss}
interconnect capacitance C_{ics} .

- Crossover capacitance — metal to metal
- Parallel plate cap — substrate to metal.
- fringe capacitance
- Lateral capacitance

- Interconnect resistance

$R_{on} \rightarrow$ drain to source resistance (R_{ds})

$$= \frac{V_{ds}}{I_d} \text{ (for MOS device)}$$

- Copper has more reliability than aluminium as its melting point is high.

- 2/18.
- Interconnects which are broader — inductor is considered — in global interconnects

- Interconnect Resistance

- " Inductance — cause, effects, trends.

- Options for wired interconnect
 - Optical Interconnect
 - Carbon Nanotube
 - Current mode

CHAPTER - 4

Electrical size

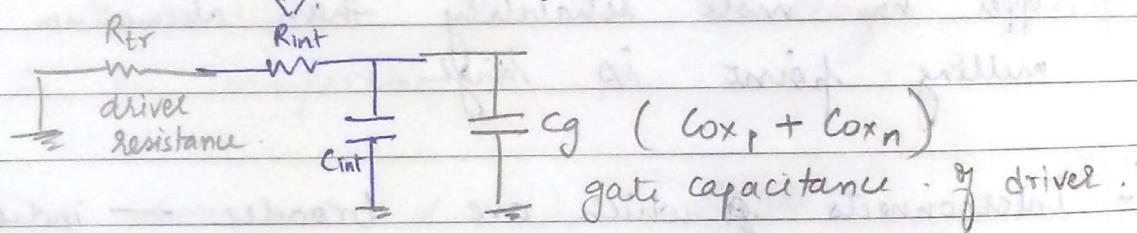
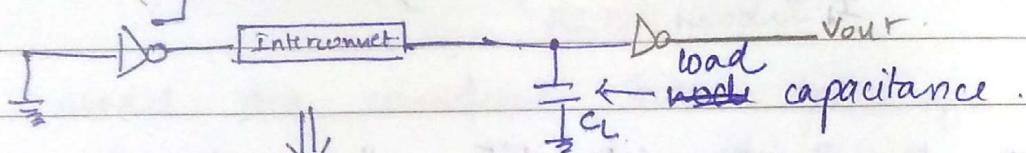
$$\text{of Interconnect } (L_e) = \frac{l}{\lambda} \quad \begin{matrix} \text{(physical length of interconnect)} \\ \text{(wavelength of signal propagating)} \end{matrix}$$

$\kappa = 0.1 \rightarrow$ Lumped RC model can be used

$> 0.1 \rightarrow$ RLC distributed model

\Rightarrow Lumped RC model for Interconnect.

→ driver (always generate 0P - step)



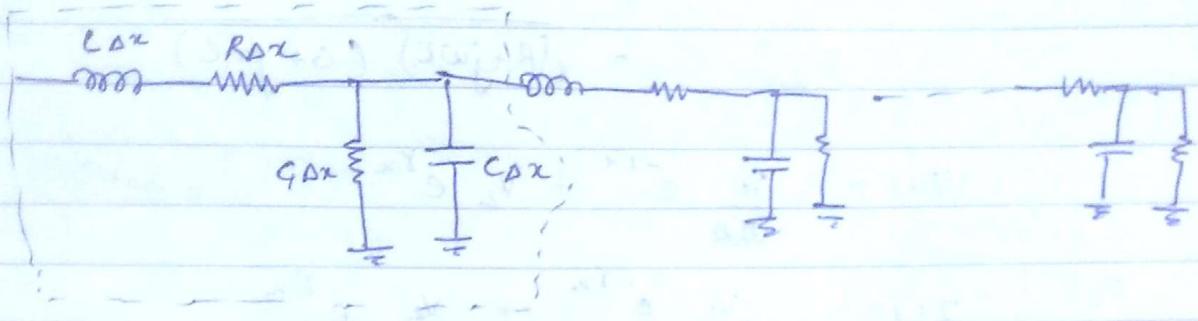
$$\tau_{RC} = (R_{br} + R_{int}) (C_g + C_{int})$$

$$V_{out} = (1 - e^{-t/\tau_{RC}})$$

$$\text{Propagation delay} = 0.693 \tau_{RC}$$

- for wide interconnect, this model is not accurate

⇒ RLC Distributed Model



- Change in voltage and current

- $dV = -L \frac{dI}{dx} dx - RI dx$ —①

(+ve sign bcoz voltage drop occurs)

- $dI = -c \frac{dV}{dx} dx - GV dx$ —②.

- from ① $\frac{dV}{dx} = -L \frac{dI}{dx} - IR$.

- from ② $\frac{dI}{dx} = -c \frac{dV}{dt} - VG$.

- $\frac{dV}{dx}(x,s) = -(R+SL) I(x,s)$

- $\frac{dI}{dx}(x,s) = -(G + SC) V(x,s)$

$R+SL = Z$ → impedance per unit length

$G + SC = Y$ → admittance per unit length.

$$\frac{dV^2}{dx^2}(x,s) = ZY V(x,s) = \gamma^2 V(x,s)$$

$$\frac{dI^2}{dx^2}(x,s) = ZY I(x,s) = \gamma^2 I(x,s)$$

$$\gamma = \sqrt{ZY}$$

$$\gamma(s) = \alpha + j\beta = \sqrt{ZC}$$

signal propagates in
+ve x direction = $\sqrt{(R+j\omega L)(G+j\omega C)}$

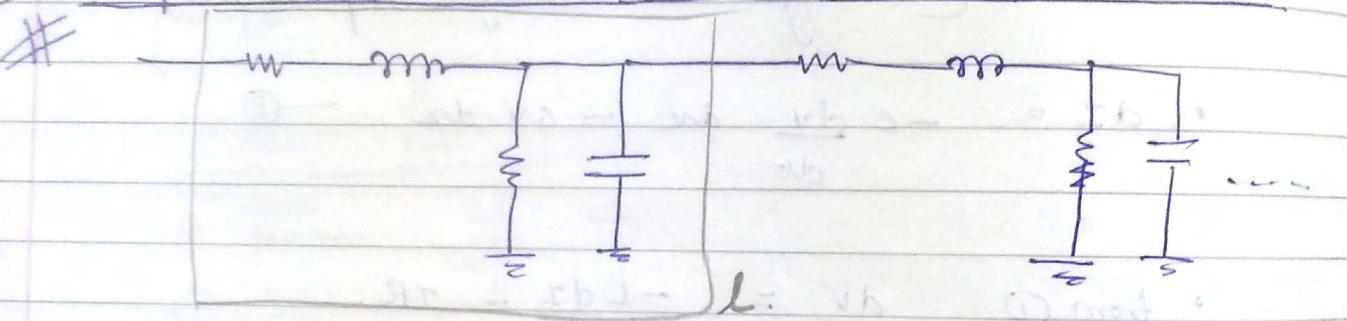
$V(x) = \frac{V_1}{Z_0} e^{-\gamma x} + \frac{V_2}{Z_0} e^{\gamma x}$ → propagates in -ve x-direction

direction of propagation

 $I(x) = \frac{V_1}{Z_0} e^{-\gamma x} + \frac{V_2}{Z_0} e^{\gamma x}$

$$Z_0 = \sqrt{\frac{R+j\omega L}{G+j\omega C}}$$

9.8. Lumped Representation of Distributed Interconnect



Driver → [Interconnect] → Load

$n \leftarrow$ no. of sections to be considered for a specific length of interconnect so that accuracy is not affected
 ↴ difference in delay is very less

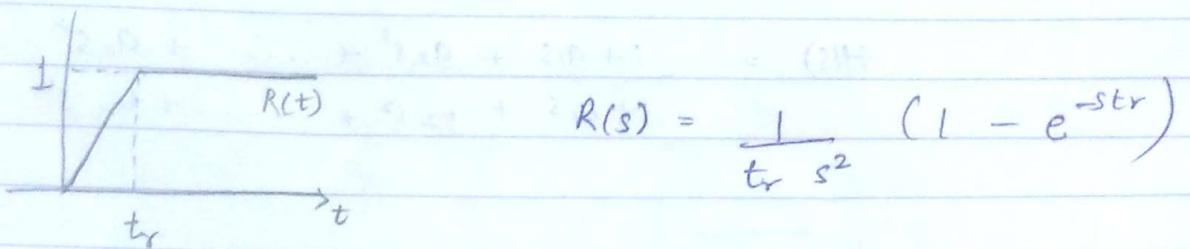
$$n \geq \frac{5l}{tr \sqrt{LC}}$$

tr — shortest rise time

l — length of interconnect

L, C — inductance & capacitance of Interconnect

Determining highest frequency of Interconnect.

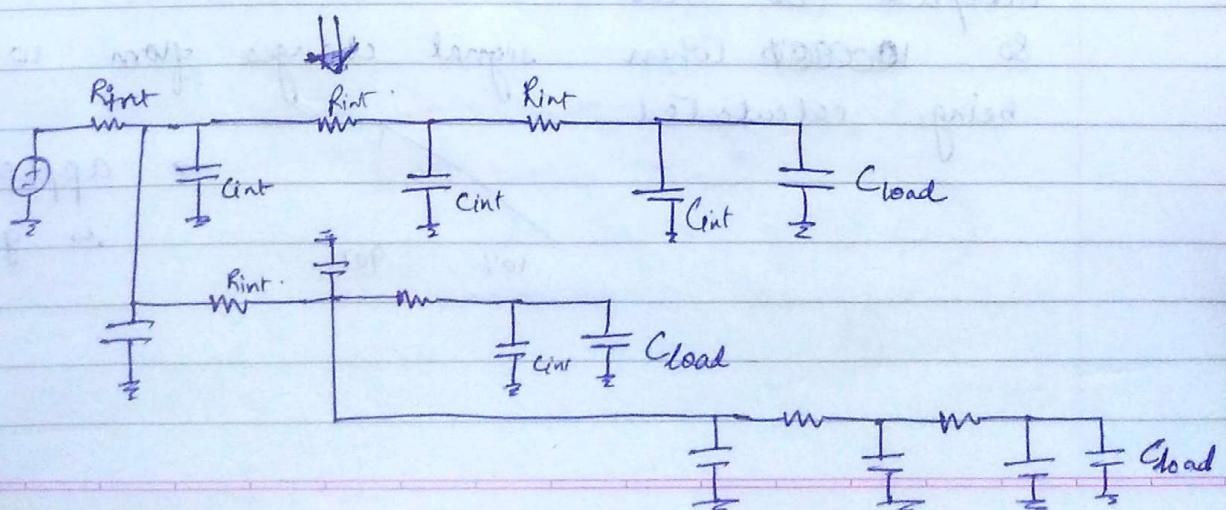
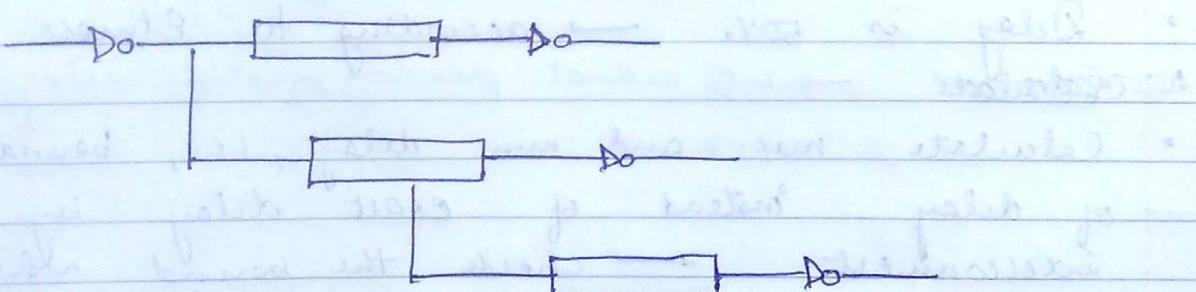


$$R(\omega) = \frac{tr}{2} \frac{|\sin \omega|}{\omega^2}$$

$$f_{max} = \frac{0.24}{tr} \quad \text{or} \quad f_{max} = \frac{0.35}{tr}$$

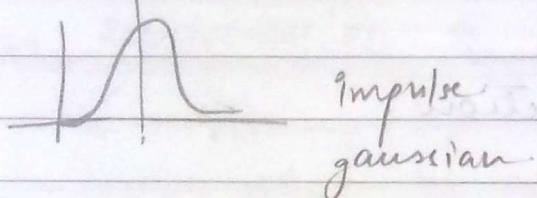
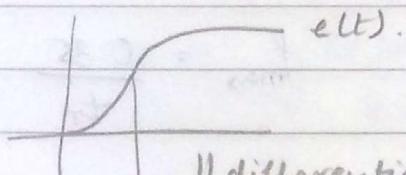
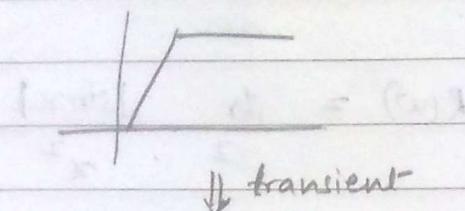
Model Order Reduction.

Why required? → ① Large number of interconnects present on chip.
 ② complex nature



Elmore delay Model → tried to calculate accurate delay.

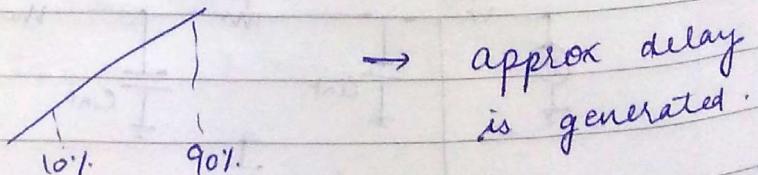
$$H(s) = \frac{1 + a_1 s + a_2 s^2 + \dots + a_n s^n}{1 + b_1 s + b_2 s^2 + \dots + b_n s^n}$$



$$T_{Elmore(max)} = -\left. \frac{\partial H(s)}{\partial s} \right|_{s=0} = b_1 - a_1$$

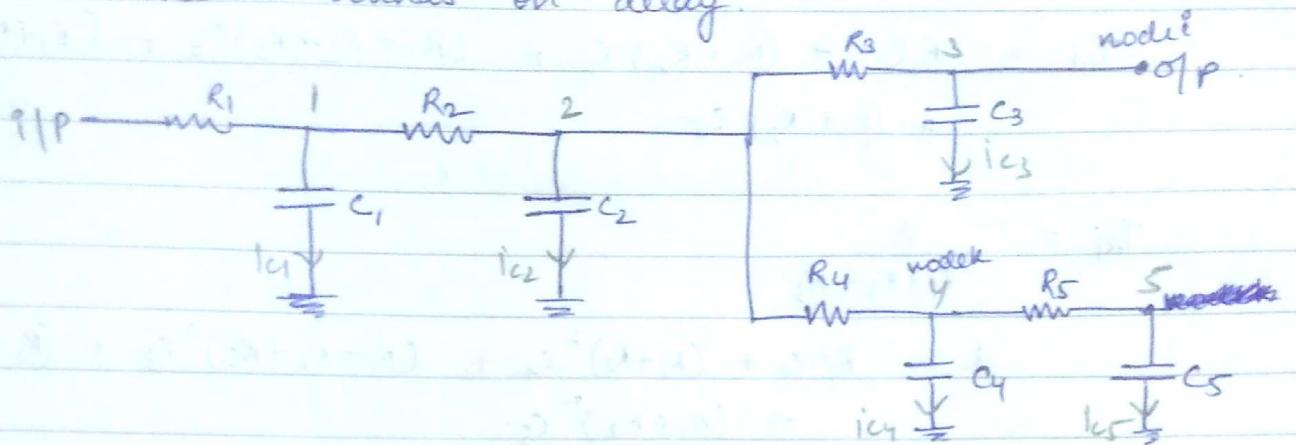
- Delay is 50% . → according to Elmore.
- ~~but did not provide accurate delay~~
- Calculate max. and min. delay, i.e., boundary of delay instead of exact delay by interconnect — check the bound whether accepted or not.

So ~~10000%~~ when signal changes from $10-90\%$ is being calculated



Algorithm by Penfield - Rubinstein

Determine 'bounds' on delay.



i = Index refer to o/p node in RC tree

k = Index cover every capacitor within circuit

C_k = lumped capacitor at node k

R_{ii} = Resistance of unique path b/w i/p & o/p i

R_{kk} = " " " " " " " " " " " " node k

R_{ik} = Resistance of portion i of unique path b/w i/p & o/p i that is common with unique path b/w o/p and node k.

$$R_{ii} = R_1 + R_2 + R_3 \quad R_{ik} \ll R_{kk}$$

$$R_{kk} = R_1 + R_2 + R_4 \quad R_{ik} \ll R_{ii}$$

$$R_{ik} = R_1 + R_2$$

T_p → upper bound

$$T_p = \sum_k R_{kk} C_k$$

T_{di} → delay on i

T_{ri} → lower bound

$$T_{di} = \sum_k R_{ik} C_k$$

$$T_{ri} = \frac{\left(\sum_k R_k^2 C_k \right)}{R_{ii}}$$

$$T_p = R_1 C_1 + (R_1 + R_2) C_2 + (R_1 + R_2 + R_3) C_3 + (R_1 + R_2 + R_3 + R_4) C_4 + (R_1 + R_2 + R_3 + R_4 + R_5) C_5$$

$$T_{di} = R_1 C_1 + (R_1 + R_2) C_2 + (R_1 + R_2 + R_3) C_3 + (R_1 + R_2) C_4 + (R_1 + R_2) C_5$$

\downarrow common b/w C_3 & C_4

$$T_{ri} = \frac{K}{R_1 + R_2 + R_3}$$

$$K = R_1^2 C_1 + (R_1 + R_2)^2 C_2 + (R_1 + R_2 + R_3)^2 C_3 + (R_1 + R_2)^2 C_4 + (R_1 + R_2)^2 C_5$$

$$\boxed{T_{ri} \ll T_{di} \ll T_p}$$

↑ Elmore's delay.

→ Voltage drop along the path.

$$1 - v_i(t) = \sum_k R_{ki} C_k \frac{d_v V_k}{dt}$$

node 3

$$1 - v_3(t) = (i_4 + i_{c2} + i_{c3} + i_{c4} + i_{c5}) R_1 + (i_{c2} + i_{c3} + i_{c4} + i_{c5}) R_2 + (i_{c3} R_3)$$

→ Node Delay.

$$T_{d1} = (C_1 + C_2 + C_3 + C_4 + C_5) R_1$$

$$T_{d2} = T_{d1} + (C_2 + C_3 + C_4 + C_5) R_2$$

$$T_{d3} = T_{d2} + C_3 R_3$$

$$T_{d4} = T_{d2} + (C_4 + C_5) R_4$$

$$T_{d5} = T_{d4} + C_5 R_5$$

All cap. are in parallel