Advanced VLSI Design : 2021-22
Lecture 13
Interconnects

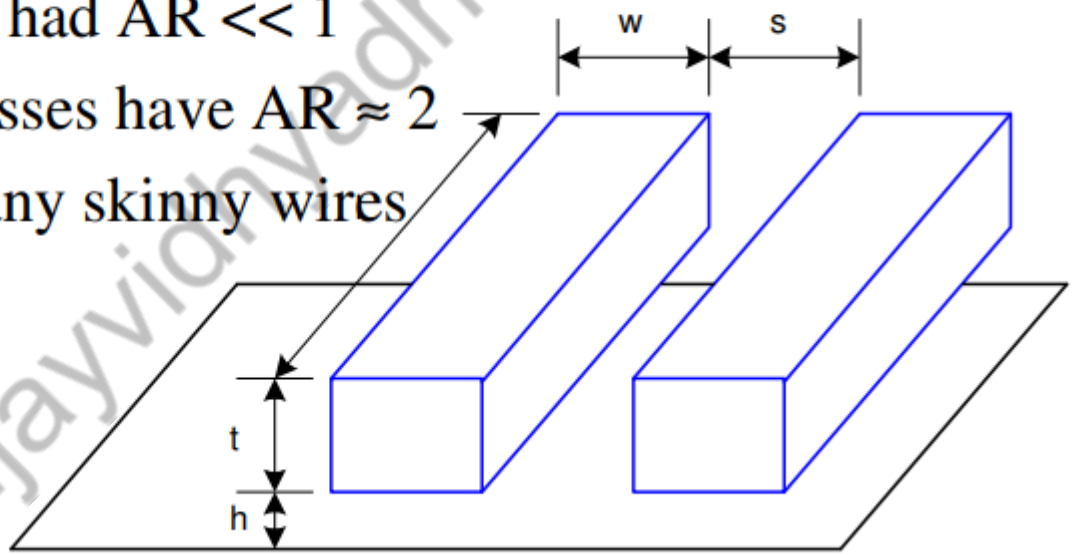
By Dr. Sanjay Vidhyadharan

Interconnects

- Chips are mostly made of wires called *interconnect*
  - In stick diagram, wires set size
  - Transistors are little things under the wires
  - Many layers of wires
- Wires are as important as transistors
  - Speed
  - Power
  - Noise
  - Area
- Alternating layers run orthogonally

Wire Geometry

- Pitch = w + s
- Aspect ratio: AR = t/w
  - Old processes had AR << 1
  - Modern processes have AR ≈ 2
    - Pack in many skinny wires

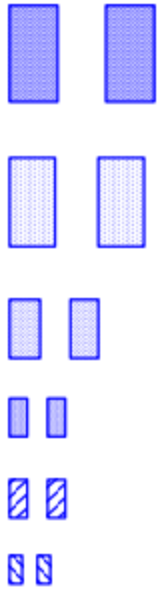


Layer Stack

- AMI 0.6 μm process has 3 metal layers
- Modern processes use 6-10+ metal layers
- Example:

  Intel 180 nm process
- M1: thin, narrow (< 3λ)
  - High density cells
- M2-M4: thicker
  - For longer wires
- M5-M6: thickest
  - For V_{DD}, GND, clk

Layer	T (nm)	W (nm)	S (nm)	AR
6	1720 1000	860	860	2.0
5	1600 1000	800	800	2.0
4	1080 700	540	540	2.0
3	700 700	320	320	2.2
2	700 700	320	320	2.2
1	480 800	250	250	1.9



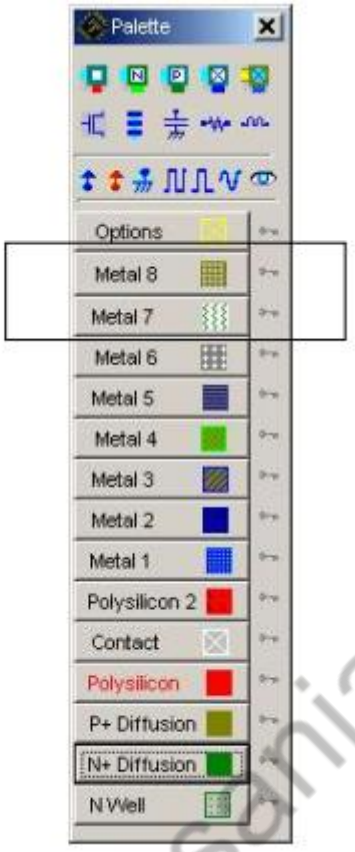
Choice of Metals

- Until the 180 nm generation, most wires were aluminum
- Modern processes often use copper
  - Cu atoms diffuse into silicon and damage FETs
  - Must be surrounded by a diffusion barrier

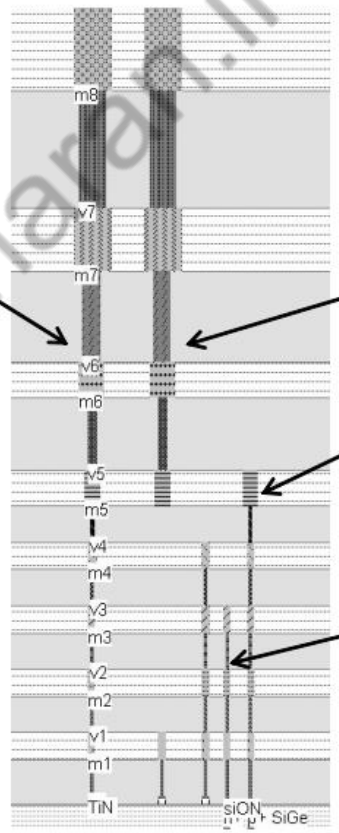
Metal	**Bulk Resistivity** $(\mu\Omega * cm)$
Silver (Ag)	1.6
Copper (Cu)	1.7
Gold (Au)	2.2
Aluminum (Al)	2.8
Tungsten (W)	5.3
Molybdenum (Mo)	5.3

Metal Layers 45 nm Technology

New layers introduced in 65-nm technology node



Contact poly-metal1-..-metal8



Contact Metal5..metal8

Contact P+diff/metal1..metal5

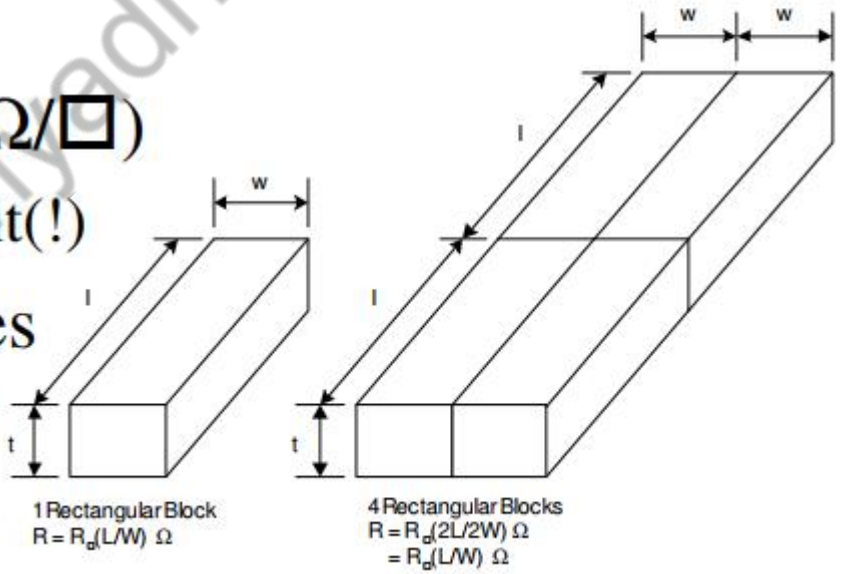
Contact poly/metal1..metal3

Wire Resistance

ρ = *resistivity* (Ω*m)

$$R = \frac{\rho}{t}\frac{l}{w} = R_{\square}\frac{l}{w}$$

- $R_{\square}$ = *sheet resistance* ($\Omega/\square$)
  - $\square$ is a dimensionless unit(!)
- Count number of squares
  - $R = R_{\square} * (\#\ of\ squares)$



Sheet Resistance

Typical sheet resistances in 180 nm process

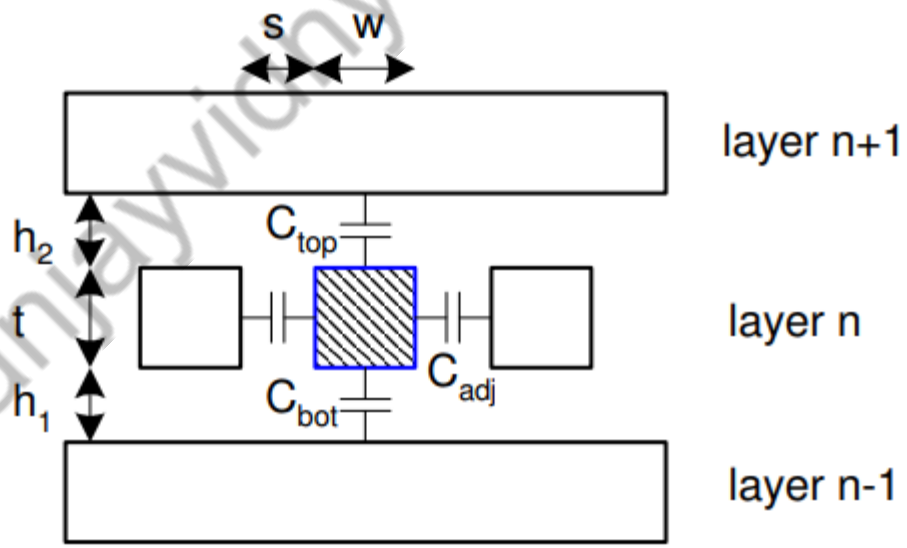
Layer	**Sheet Resistance** ($\Omega/\square$)
Diffusion (silicided)	3–10
Diffusion (no silicide)	50–200
Polysilicon (silicided)	3–10
Polysilicon (no silicide)	50–400
Metal1	0.08
Metal2	0.05
Metal3	0.05
Metal4	0.03
Metal5	0.02
Metal6	0.02

Contacts Resistance

- Contacts and vias also have 2-20 Ω
- Use many contacts for lower R
  - Many small contacts for current crowding around periphery

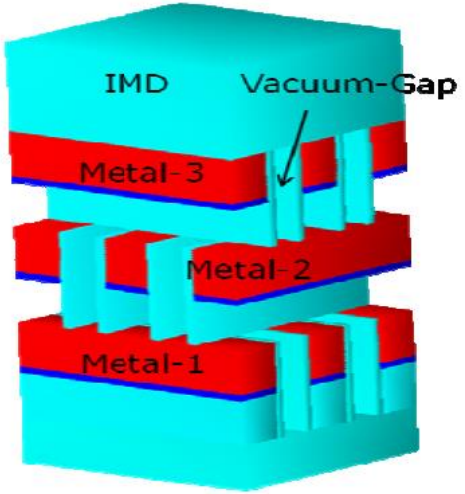
Wire Capacitance

- Wire has capacitance per unit length
  - To neighbors
  - To layers above and below
- $C_{total} = C_{top} + C_{bot} + 2C_{adj}$

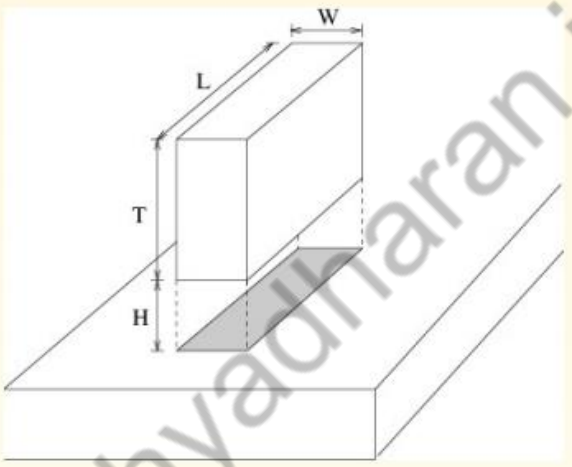


Capacitance Trends

- Parallel plate equation: $C = \varepsilon A/d$
  - Wires are not parallel plates, but obey trends
  - Increasing area (W, t) increases capacitance
  - Increasing distance (s, h) decreases capacitance
- Dielectric constant
  - $\varepsilon = k\varepsilon_0$
- $\varepsilon_0 = 8.85 \times 10^{-14}$ F/cm
- k = 3.9 for SiO_2
- Processes are starting to use low-k dielectri
  - $k \approx 3$ (or less) as dielectrics use air pockets
- Typical (M2) wires have ~ 0.2 fF/µm
  - Compare to 2 fF/µm for gate capacitance

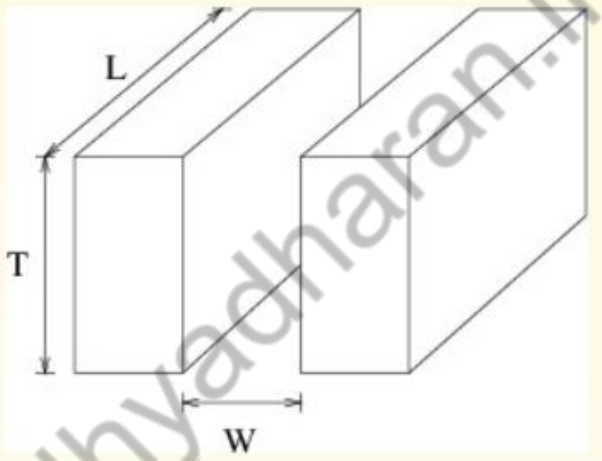


Capacitance Trends



- $W >> H \Rightarrow$ Parallel Plate Model
  - $C = k \cdot \epsilon_0 \cdot \frac{W \cdot L}{H}$
- $W \leq H \Rightarrow$ Fringing Model
  - $C \; \alpha \; log(W)$
- For Deep Sub-Micron (DSM) (or nanoscale) processes, fringing model applies

Capacitance Trends



- $T >> W \Rightarrow$ Parallel Plate Model
  - $C = k \cdot \epsilon_0 \cdot \frac{T \cdot L}{W}$
- $T \leq W \Rightarrow$ Fringing Model
  - $C\ \alpha\ log(T)$
- For DSM processes, parallel plate model applies

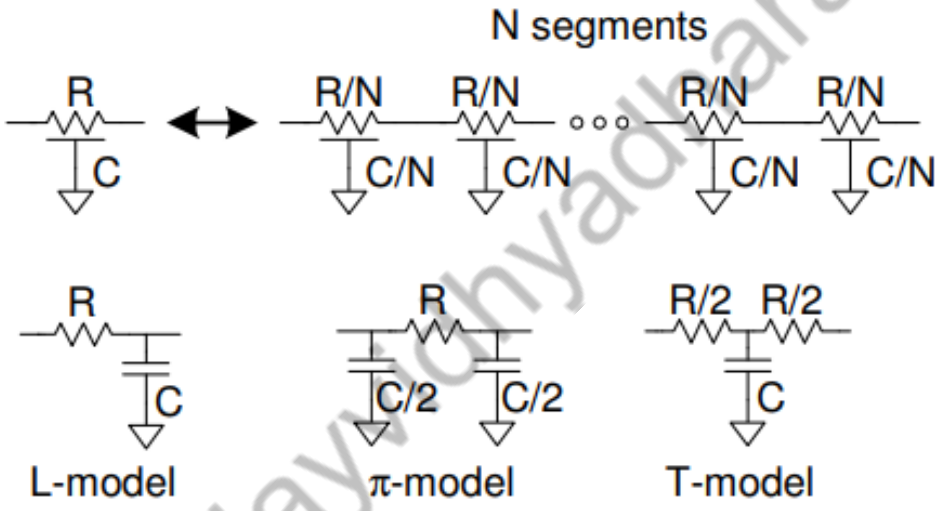
Polysilicon

- Diffusion capacitance is very high (about 2 fF/μm)
  - Comparable to gate capacitance
  - Diffusion also has high resistance
  - Avoid using diffusion *runners* for wires!

- Polysilicon has lower C but high R
  - Use for transistor gates
  - Occasionally for very short wires between gates

Lumped Element Models

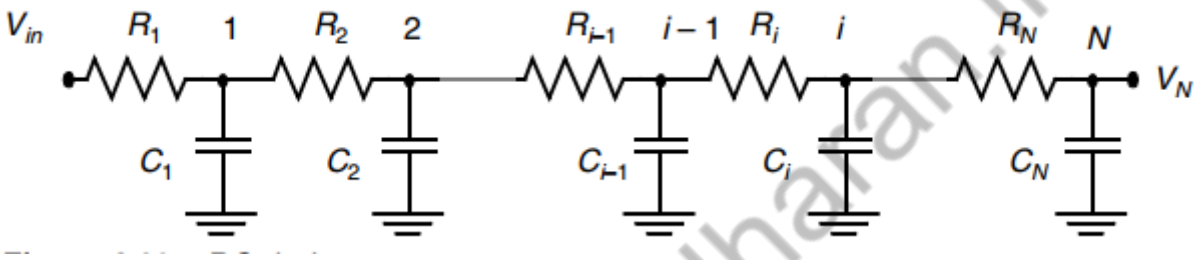
- Wires are a distributed system
  - Approximate with lumped element models



- 3-segment π-model is accurate to 3% in simulation

Lumped Element Models

L Model



$$\tau_{Di} = C_1R_1 + C_2(R_1 + R_2) + \ldots + C_i(R_1 + R_2 + \ldots + R_i)$$

The wire with a total length of L is partitioned into N identical segments, each with a length of L/N. The resistance and capacitance of each segment are hence given by rL/N and cL/N, respectively.

$$\tau_{DN} = \left(\frac{L}{N}\right)^2 (rc + 2rc + \ldots + Nrc) = (rcL^2)\frac{N(N+1)}{2N^2} = RC\frac{N+1}{2N}$$

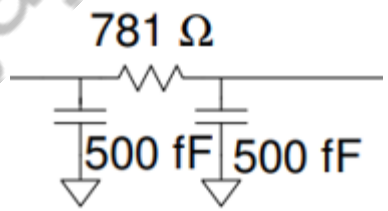
$$\tau_{DN} = \frac{RC}{2} = \frac{rcL^2}{2}$$

The delay of a wire is **a quadratic function of its length**! This means that doubling the length of the wire quadruples its delay

Lumped Element Models

Example

- Metal2 wire in 180 nm process
  - 5 mm long
  - 0.32 μm wide
  - Number of squares = 5000/0.32 = 15625
- Construct a 3-segment π-model
  - $R_{\square} = 0.05\ \Omega/\square$ => R = 15625 * 0.05 = 781 Ω
  - $C_{permicron} = 0.2$ fF/μm => C = 0.2 fF/μm * 5000 μm = 1 pF

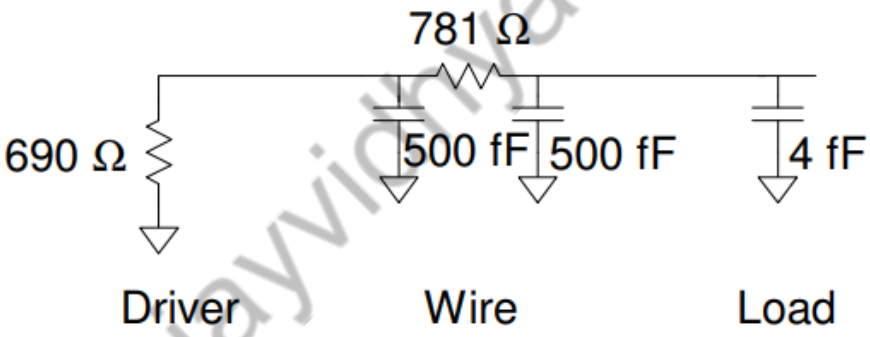


Lumped Element Models

- Estimate the delay of a 10x inverter driving a 2x inverter at the end of the 5mm wire from the previous example.
  - R = 2.5 kΩ*µm for gates
  - Unit inverter: 0.36 µm nMOS, 0.72 µm pMOS
  - Unit inverter has 4λ = 0.36µm wide nMOS, 8λ = 0.72µm wide pMOS
  - Unit inverter: effective resistance of (2.5 kΩ*µm)/(0.36µm) = 6.9 kΩ
  - Capacitance: (0.36µm + 0.72 µm) * (2fF/µm) = 2fF

Elmore delay model

- t_{pd} = 1.1 ns



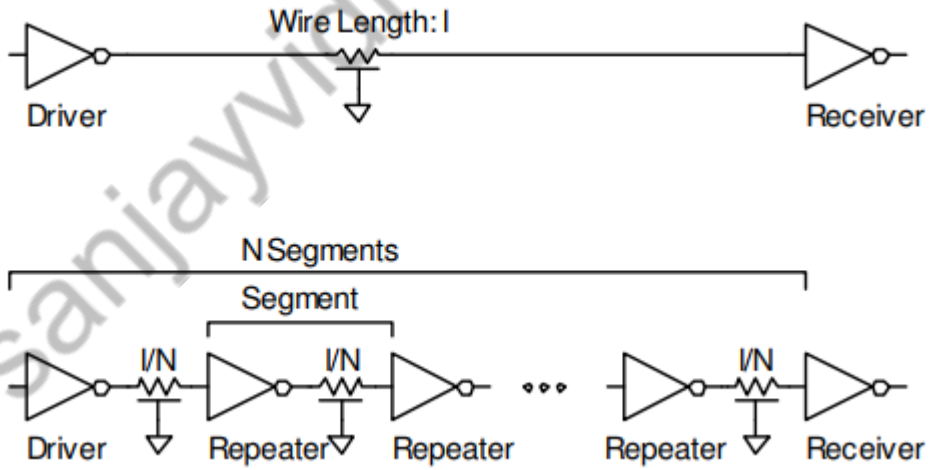
Empirical values have been found for R_N and R_P,

$$R_N = \frac{12.5}{(W/L)_n} \text{ k}\Omega$$

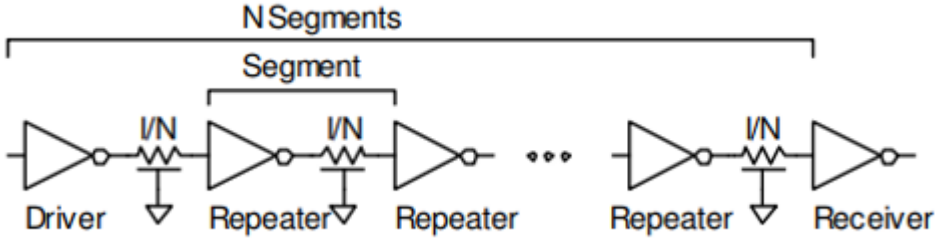
$$R_P = \frac{30}{(W/L)_p} \text{ k}\Omega$$

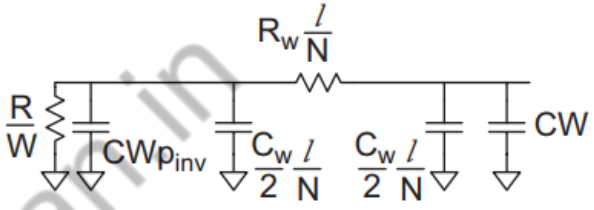
Repeaters

- R and C are proportional to l
- RC delay is proportional to l^2
  - Unacceptably great for long wires
- Break long wires into N shorter segments
  - Drive each one with an inverter or buffer



Repeaters





$$t_{pd} = N\left[\frac{R}{W}\left(C_w \frac{l}{N} + CW\left(1 + p_{\text{inv}}\right)\right) + R_w \frac{l}{N}\left(\frac{C_w}{2}\frac{l}{N} + CW\right)\right]$$

$$\frac{l}{N} = \sqrt{\frac{2RC\left(1 + p_{\text{inv}}\right)}{R_w C_w}}$$

FO4 inverter is $5RC$. Assuming $p_{\text{inv}} \approx 0.5$

$$\frac{l}{N} = 0.77\sqrt{\frac{\text{FO4}}{R_w C_w}}$$

The delay per unit length of a properly repeated wire is

$$\frac{t_{pd}}{l} = \left(2 + \sqrt{2\left(1 + p_{\text{inv}}\right)}\right)\sqrt{RCR_w C_w} \approx 1.67\sqrt{\text{FO4}\, R_w C_w}$$

To achieve this delay, the inverters should use an nMOS transistor width of

$$W = \sqrt{\frac{RC_w}{R_w C}}$$

Crosstalk

A capacitor does not like to change its voltage instantaneously

A wire has high capacitance to its neighbor

When the neighbor switches from 1→0 or 0→1, the wire tends to switch too Called capacitive **coupling** or **crosstalk**

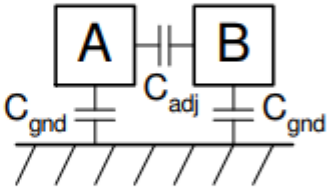
Crosstalk effects

Noise on nonswitching wires

Increased **delay** on switching wires

Crosstalk Delay

- Assume layers above and below on average are quiet
  - Second terminal of capacitor can be ignored
  - Model as $C_{gnd} = C_{top} + C_{bot}$
- Effective C_{adj} depends on behavior of neighbors
  - *Miller effect*

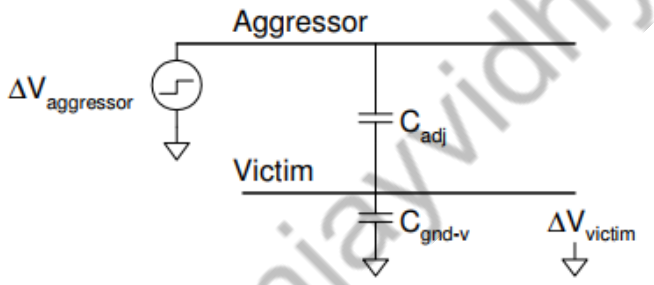


B	ΔV	$C_{eff(A)}$	MCF
Constant	V_{DD}	$C_{gnd} + C_{adj}$	1
Switching with A	0	C_{gnd}	0
Switching opposite A	$2V_{DD}$	$C_{gnd} + 2\,C_{adj}$	2

Crosstalk

- Crosstalk causes noise on non-switching wires
- If victim is floating:
  - model as capacitive voltage divider

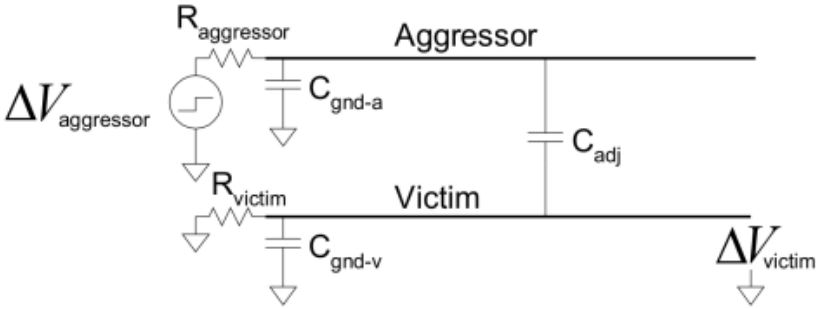
$$\Delta V_{victim} = \frac{C_{adj}}{C_{gnd-v} + C_{adj}} \Delta V_{aggressor}$$

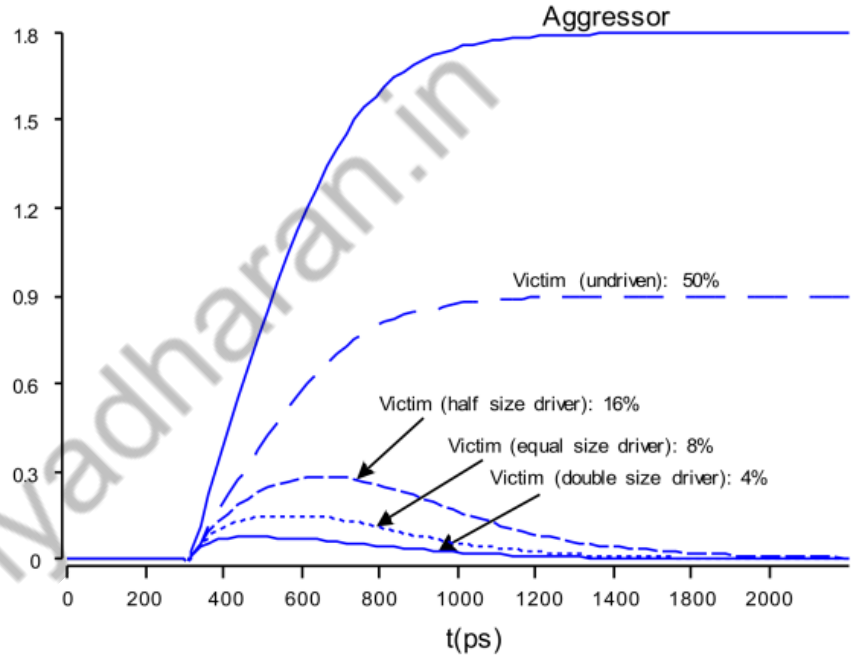


$$C_{adj}\left(\Delta V_{agg} - \Delta V_{vic}\right) = C_{gnd}\left(\Delta V_{vic}\right)$$

Crosstalk

Driven Victims





If the noise is less than the noise margin, nothing happens Static CMOS logic will eventually settle to correct output even if disturbed by large noise spikes

But glitches cause extra delay, also cause extra power from false transitions

Dynamic logic never recovers from glitches.

Memories and other sensitive circuits also can produce the wrong answer

Thank you