ECE1388 VLSI Design Methodology

Lecture 8: Interconnect and Wire Engineering

Outline

❑ Introduction

❑ Interconnect Modeling

    – Wire Resistance

    – Wire Capacitance

❑ Wire RC Delay

❑ Crosstalk

❑ Wire Engineering

❑ Repeaters

Introduction

❑ 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

❑ 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
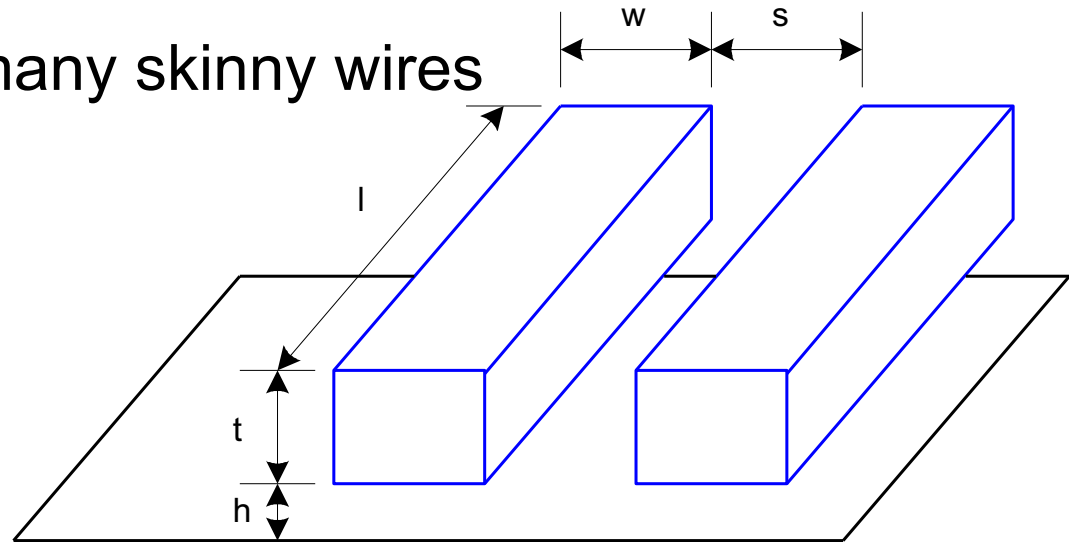
Layer Stack

❑ AMI 0.6 μm process has 3 metal layers
– M1 for within-cell routing
– M2 for vertical routing between cells
– M3 for horizontal routing between cells
❑ Modern processes use 6-10+ metal layers
– M1: thin, narrow (< 3λ)
• High density cells
– Mid layers
• Thicker and wider, (density vs. speed)
– Top layers: thickest
• For V_{DD}, GND, clk

Example
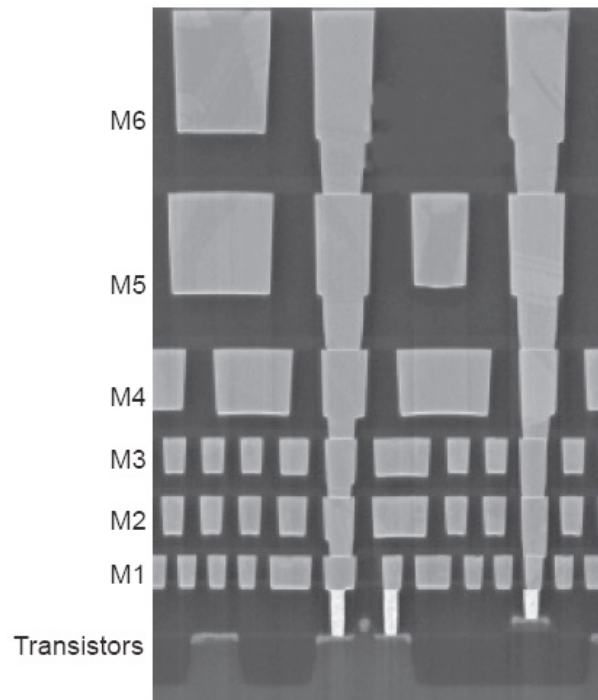
Intel 90 nm Stack

[Thompson02]
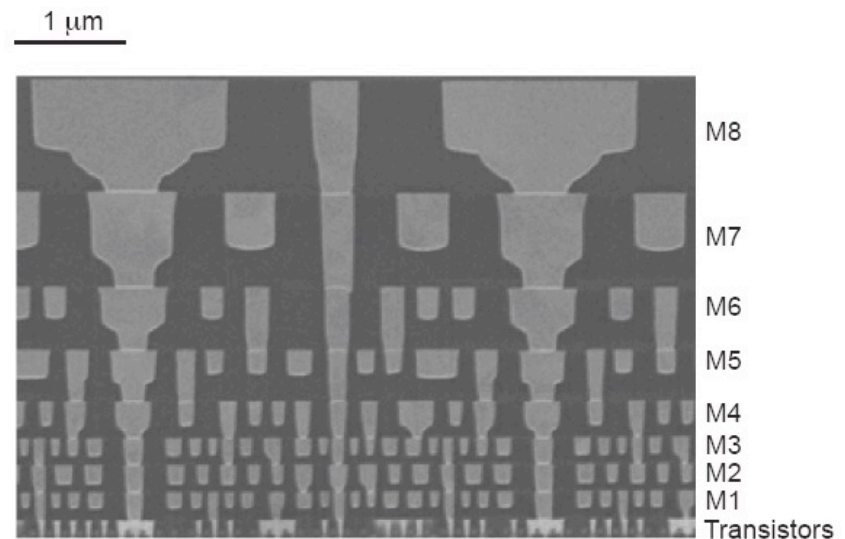
Intel 45 nm Stack

[Moon08]

Interconnect Modeling

❑ Current in a wire is analogous to current in a pipe

– Resistance: narrow size impedes flow

– Capacitance: trough under the leaky pipe must fill first

– Inductance: paddle wheel inertia opposes changes in flow rate

• Negligible for most wires

Water Wheel: Inductance

Pipe: Resistance

Trough: Capacitance
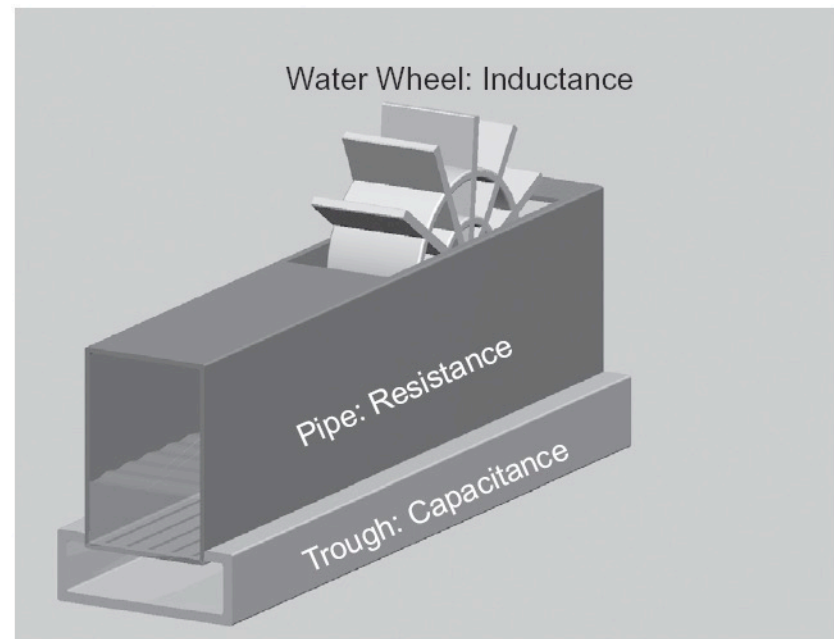
Lumped Element Models

❑ Wires are a distributed system
   – Approximate with lumped element models
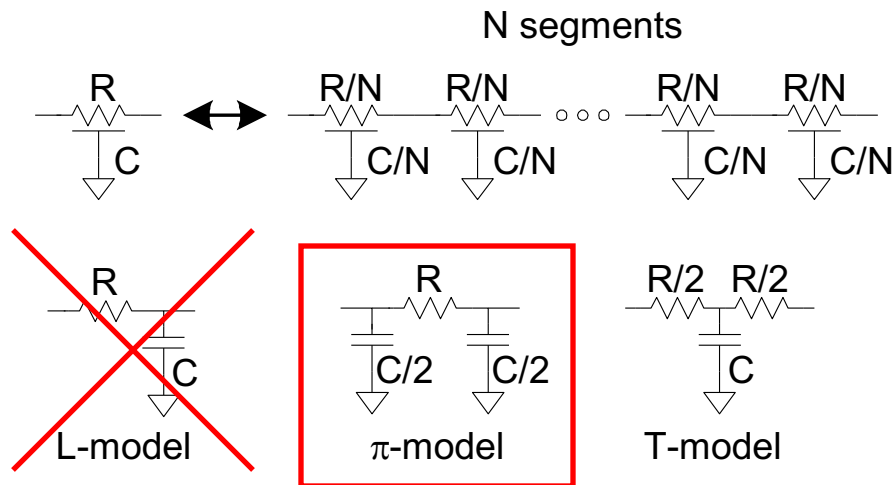
❑ 3-segment π-model is accurate to 3% in simulation
❑ L-model needs 100 segments for same accuracy!
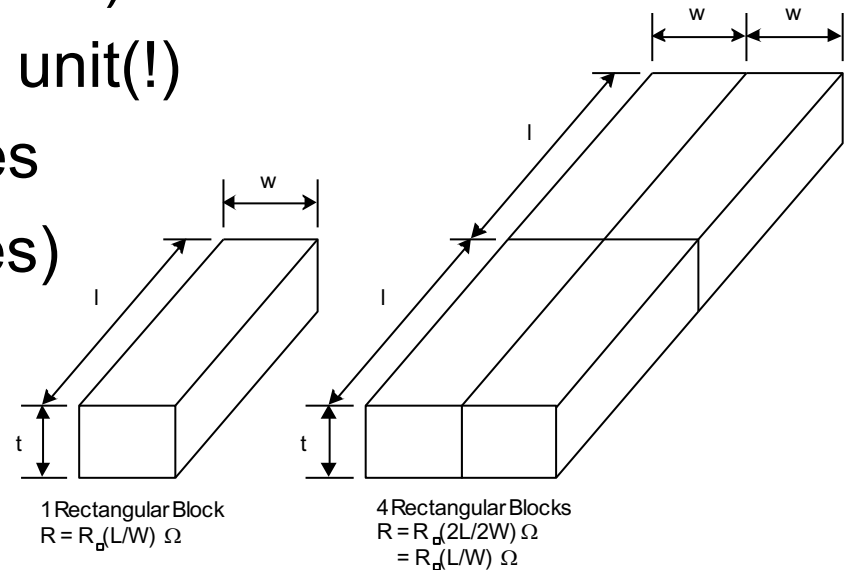❑ Use single segment π-model for Elmore delay

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

  – $R = R_{\square} * (\# \text{ of squares})$

1 Rectangular Block
$R = R_{\square}(L/W) \ \Omega$

4 Rectangular Blocks
$R = R_{\square}(2L/2W) \ \Omega$
$= R_{\square}(L/W) \ \Omega$

Choice of Metals

❑ Until 180 nm generation, most wires were aluminum

❑ Contemporary processes normally use copper

  – Cu atoms diffuse into silicon and damage FETs

  – Must be surrounded by a diffusion barrier

| Metal	Bulk resistivity ($\mu\Omega \cdot cm$)
Silver (Ag)	1.6
Copper (Cu)	1.7
Gold (Au)	2.2
Aluminum (Al)	2.8
Tungsten (W)	5.3
Titanium (Ti)	43.0

Contacts Resistance

❑ Contacts and vias also have 2-20 Ω resistance

❑ Use many contacts for lower R

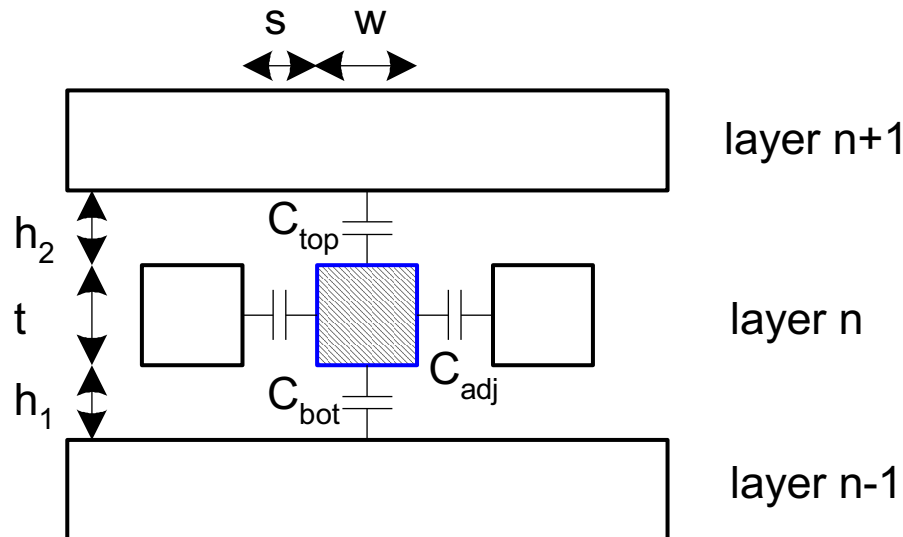– Many small contacts for current crowding around periphery

Copper Issues

❑ Copper wires diffusion barrier has high resistance
❑ Copper is also prone to *dishing* during polishing
❑ Effective resistance is higher

$$R = \frac{\rho}{\left(t - t_{\text{dish}} - t_{\text{barrier}}\right)} \frac{l}{\left(w - 2t_{\text{barrier}}\right)}$$
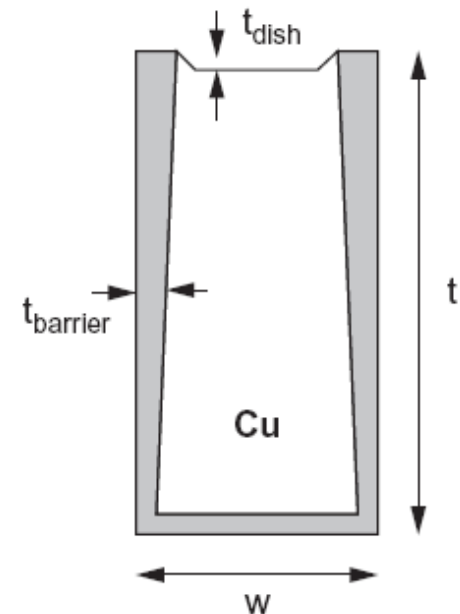
Wire Capacitance

❑ 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: $C = \varepsilon_{ox}A/d$

   – Wires are not parallel plates, but obey trends

   – Increasing area (W, t) increases capacitance

   – Increasing distance (s, h) decreases capacitance

❑ Dielectric constant

   – $\varepsilon_{ox} = k\varepsilon_0$

     • $\varepsilon_0 = 8.85 \times 10^{-14}$ F/cm

     • $k = 3.9$ for SiO_2

❑ Processes are starting to use low-k dielectrics

   – $k \approx 3$ (or less) as dielectrics use air pockets
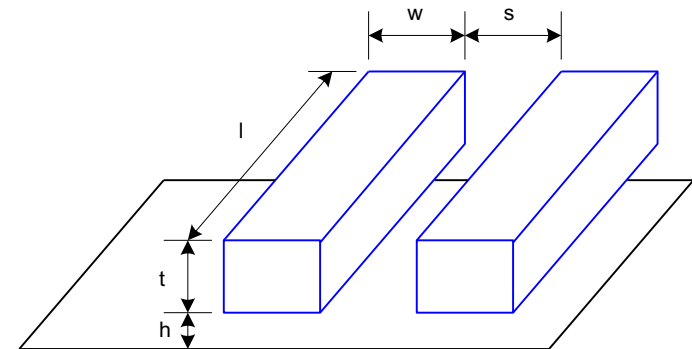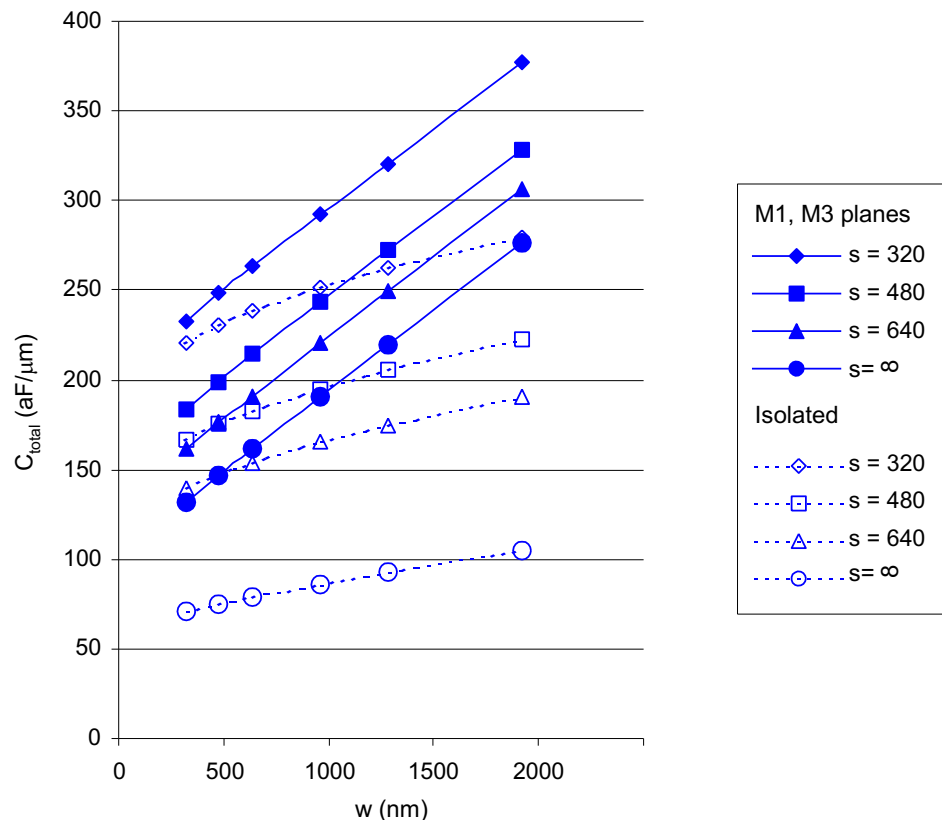
Capacitance Formula

❑ Capacitance of a line without neighbors can be approximated as

$$C_{tot} = \varepsilon_{ox} l \left[\frac{w}{h} + 0.77 + 1.06 \left(\frac{w}{h} \right)^{0.25} + 1.06 \left(\frac{t}{h} \right)^{0.5} \right]$$

❑ This empirical formula is accurate to 6% for AR < 3.3

M2 Capacitance Data

❑ Typical dense wires have ~ 0.2 fF/μm
  – Compare to 1-2 fF/μm for gate capacitance

Diffusion & Polysilicon

❑ Diffusion capacitance is very high (1-2 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

Crosstalk

❑ A capacitor does 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:
    • Miller Coupling Factor (MCF)
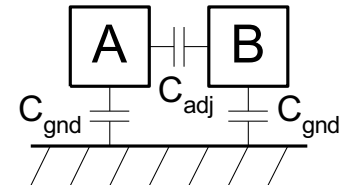
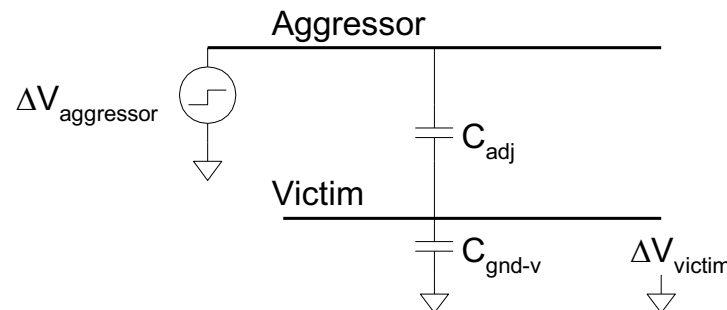| B	ΔV	$C_{eff(A)}$	MCF
Constant	V_{DD}	$C_{gnd} + C_{adj}$	1
Switching with A	0	C_{gnd}	0
Switching opposite A	$2V_{DD}$	$C_{gnd} + 2\,C_{adj}$	2

Crosstalk Noise

❑ Crosstalk causes noise on nonswitching wires
❑ If victim is floating:
   – model as capacitive voltage divider

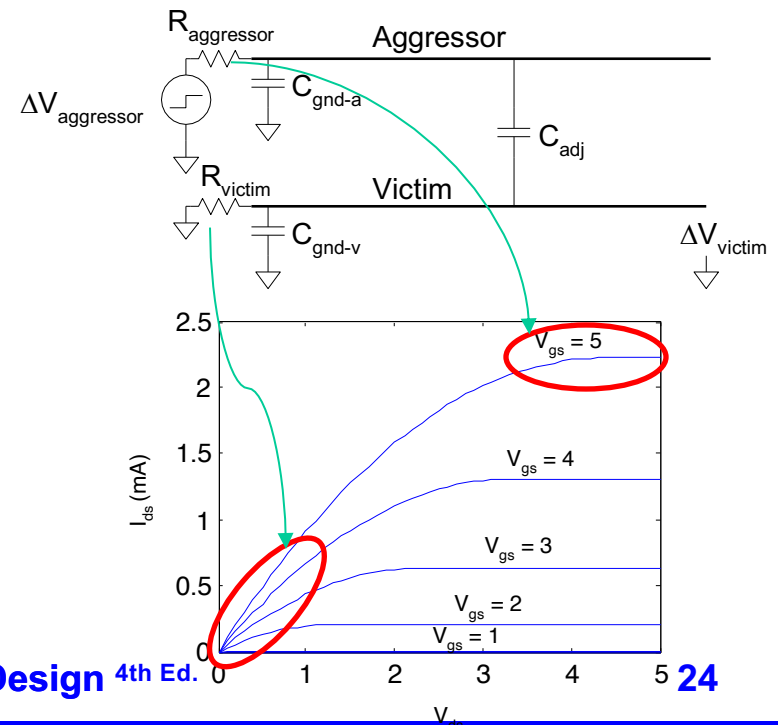$$\Delta V_{victim} = \frac{C_{adj}}{C_{gnd-v} + C_{adj}} \Delta V_{aggressor}$$

Driven Victims

❑ In practice, the voltage divider is not purely capacitive

❑ Usually victim is driven by a gate that fights noise
  – Noise depends on relative resistances
  – Victim driver is in linear region, agg. in saturation
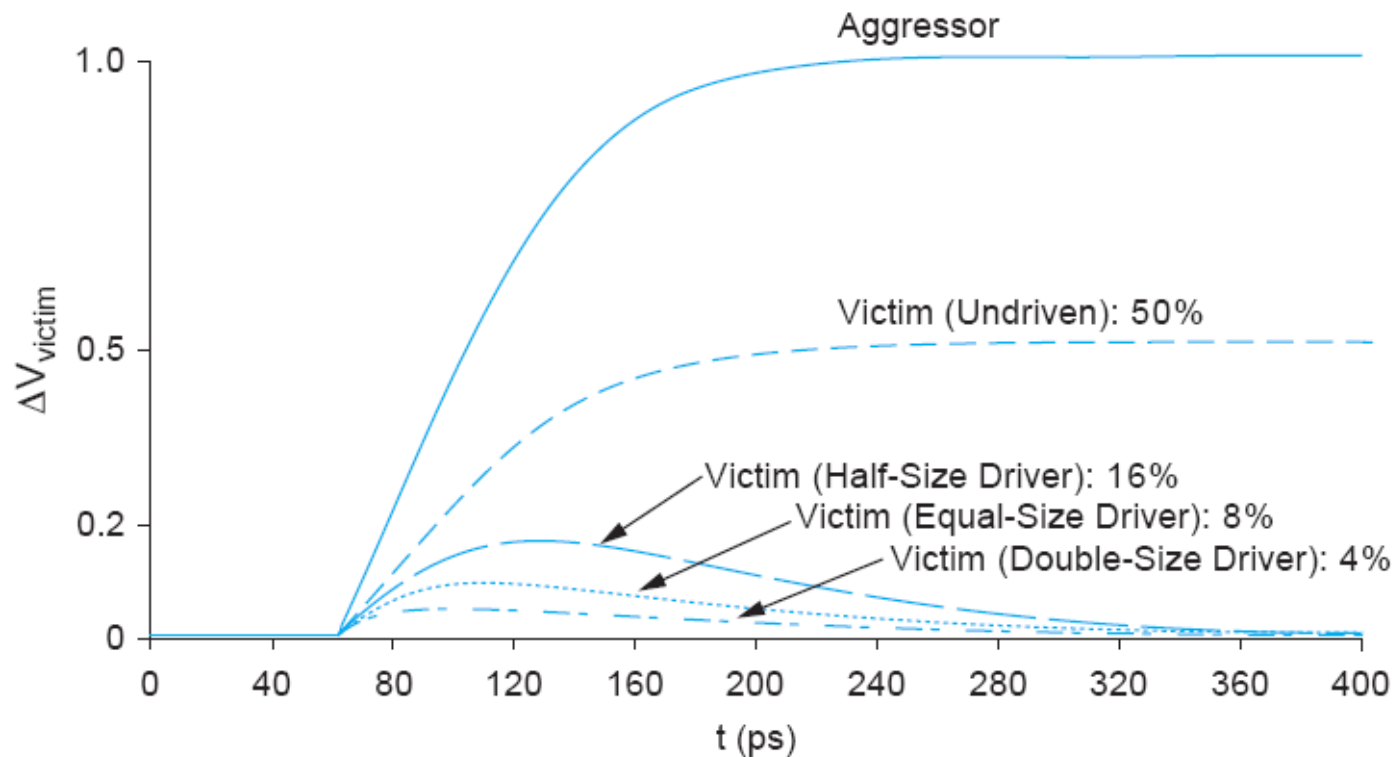 – If sizes are same, $R_{aggressor}$ = 2-4 x R_{victim}

$$\Delta V_{victim} = \frac{C_{adj}}{C_{gnd-v} + C_{adj}} \boxed{\frac{1}{1+k}} \Delta V_{aggressor}$$

$$k = \frac{\tau_{aggressor}}{\tau_{victim}} = \frac{R_{aggressor}\left(C_{gnd-a} + C_{adj}\right)}{R_{victim}\left(C_{gnd-v} + C_{adj}\right)}$$

Coupling Waveforms

❑ Simulated coupling for $C_{adj} = C_{victim}$

Aggressor

Victim (Undriven): 50%

Victim (Half-Size Driver): 16%
Victim (Equal-Size Driver): 8%
Victim (Double-Size Driver): 4%

Noise Implications

❑ *So what* if we have noise?

❑ 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

Wire Engineering

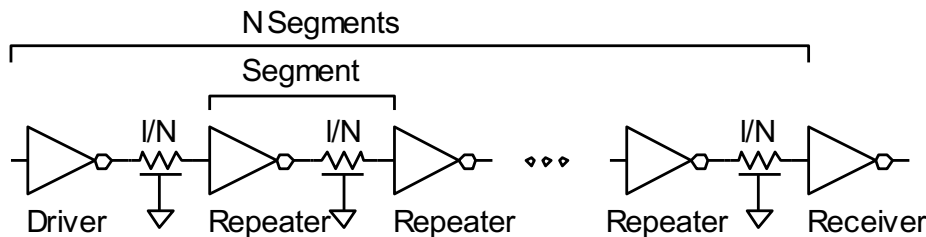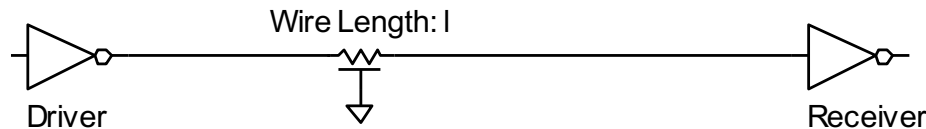❑ Goal: achieve delay, area, power goals with acceptable noise

❑ Degrees of freedom:
- Width
- Spacing
- Layer
- Shielding
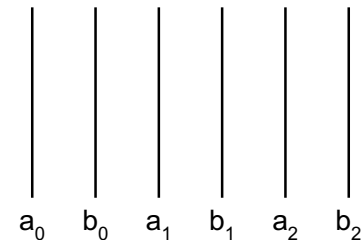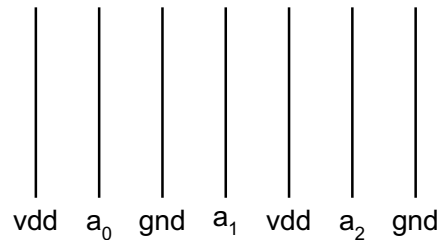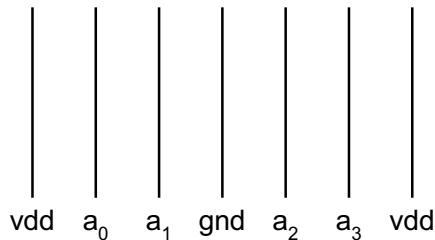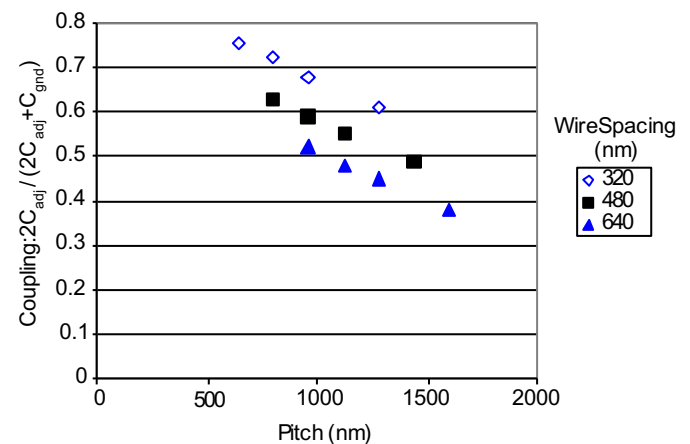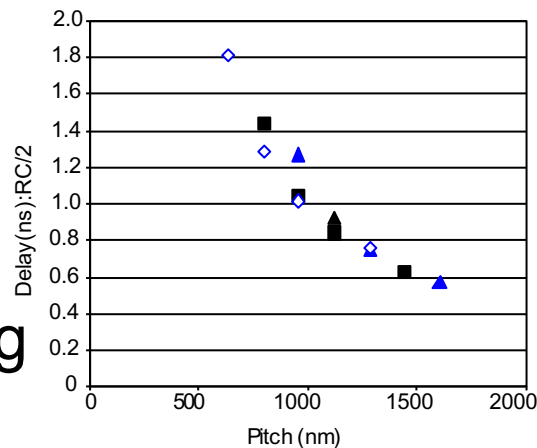
Repeaters

❑ R and C are proportional to length 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

Repeater Design

❑ How many repeaters should we use?

❑ How large should each one be?

❑ Equivalent Circuit

– Wire length *l/N*

• Wire Capacitance $C_w * l/N$, Resistance $R_w * l/N$

– Inverter width W (nMOS = W, pMOS = 2W)

• Gate Capacitance C′ *W, Resistance R/W

Repeater Results

❑ Write equation for Elmore Delay

   – Differentiate with respect to W and N

   – Set equal to 0, solve

$$\frac{l}{N} = \sqrt{\frac{2RC'}{R_w C_w}}$$ best length of wire between repeaters

$$\frac{t_{pd}}{l} = \left(2 + \sqrt{2}\right)\sqrt{RC'R_w C_w}$$

optimal delay per unit length of repeated wire (~60-80 ps/mm in 180 nm ~40 ps/mm in 65 nm)

$$W = \sqrt{\frac{RC_w}{R_w C'}}$$ optimal width of nMOS transistors

Repeater Energy

❑ Energy / length ≈ $1.87 C_w V_{DD}^2$

- – 87% premium over unrepeated wires
- – The extra power is consumed in the large repeaters

❑ If the repeaters are downsized for minimum EDP:

- – Energy premium is only 30%
- – Delay increases by 14% from min delay