



CS150 - EE141/241A

Fall 2014

**Digital Design and
Integrated Circuits**

Instructors:

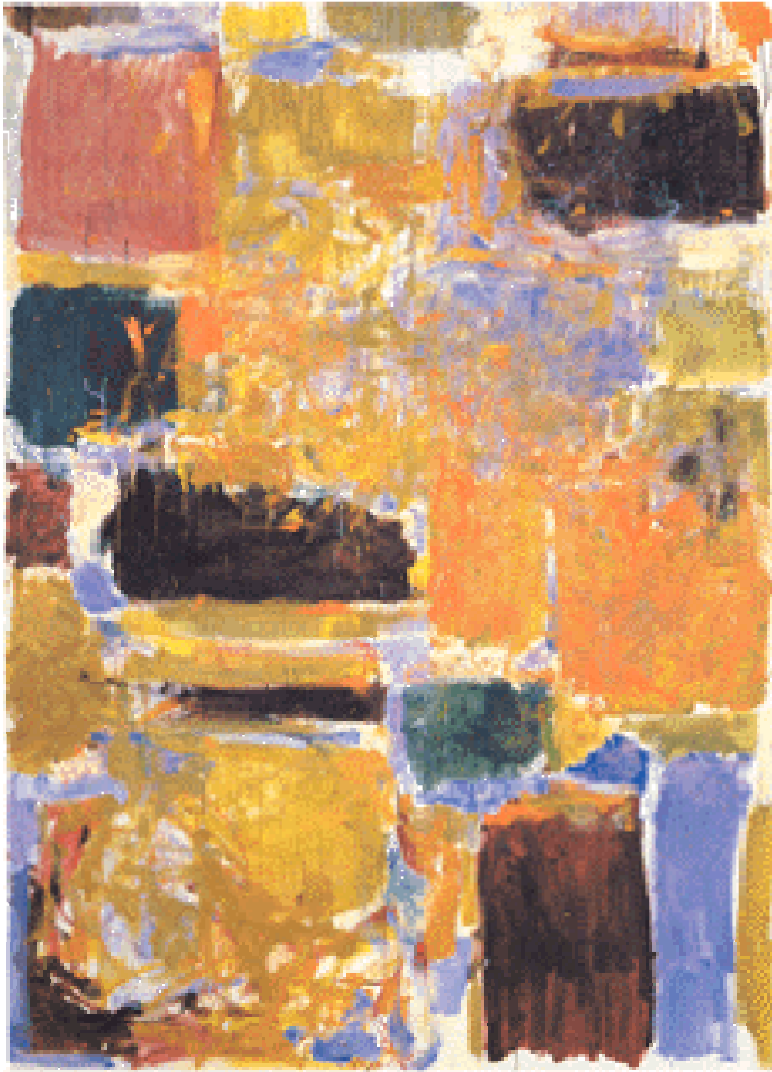
John Wawrzynek and Vladimir Stojanovic

Lecture 10

Outline

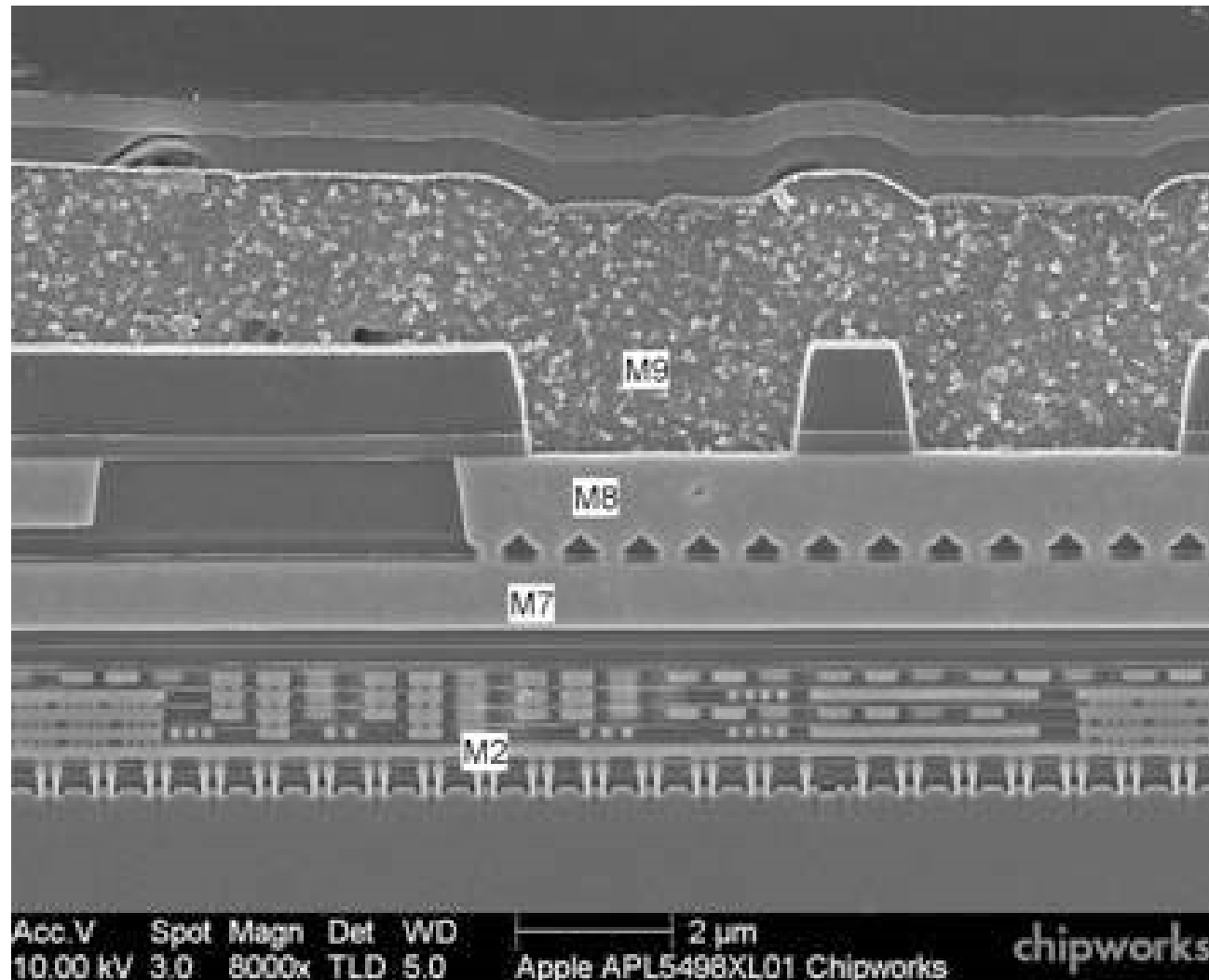


- Wires
 - Capacitance
 - Resistance
- Interconnect Modeling
 - Elmore Delay
- Repeaters



Wires

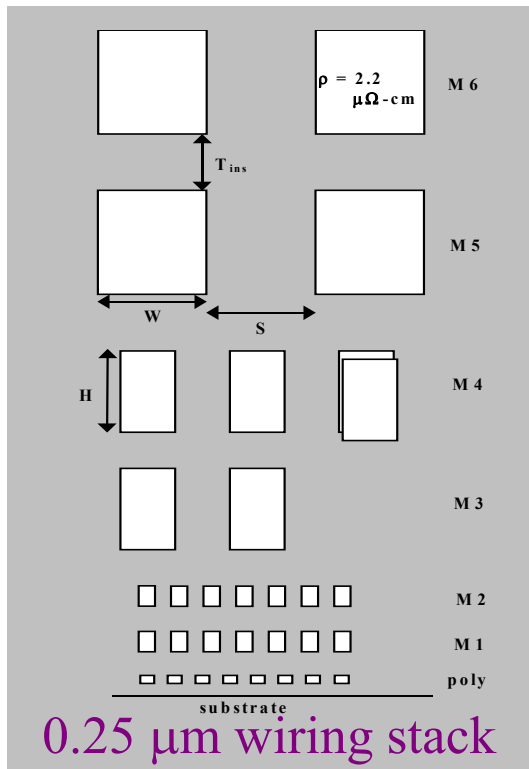
The importance of wires



Samsung 45nm process

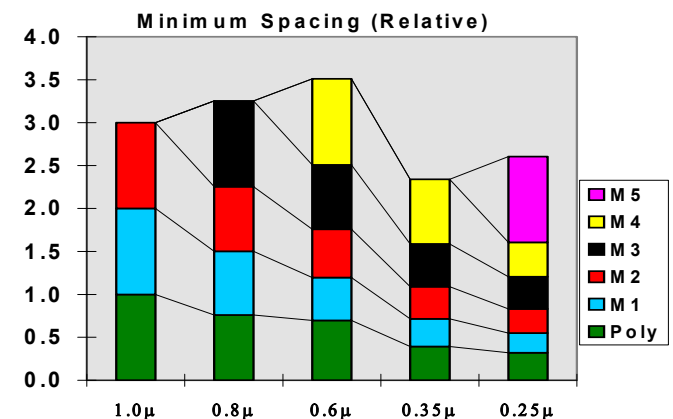
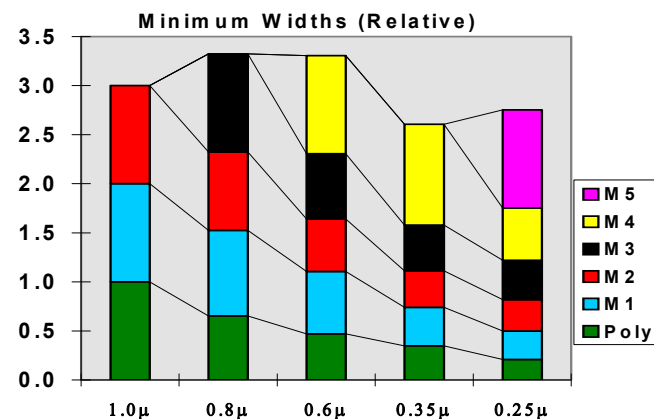
Apple iPad A5X processor (Chipworks)

Interconnect: # of Wiring Layers

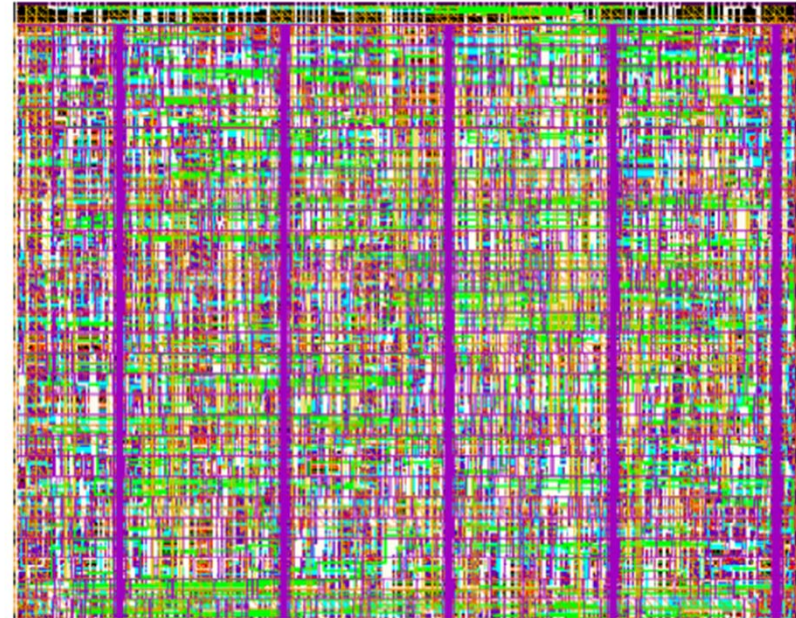
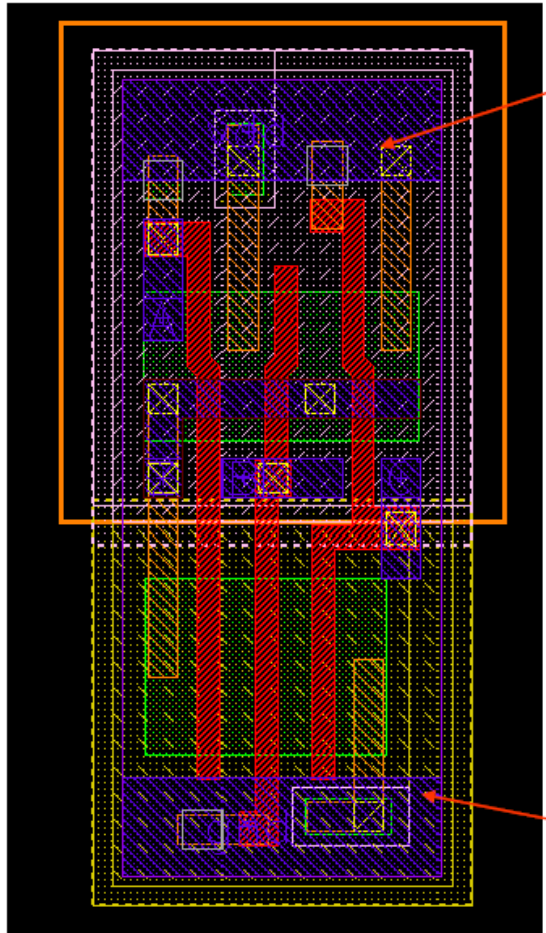


of metal layers is steadily increasing due to:

- Increasing die size and device count: we need more wires and longer wires to connect everything
- Rising need for a hierarchical wiring network; local wires with high density and global wires with low RC



Wires dominate in modern designs



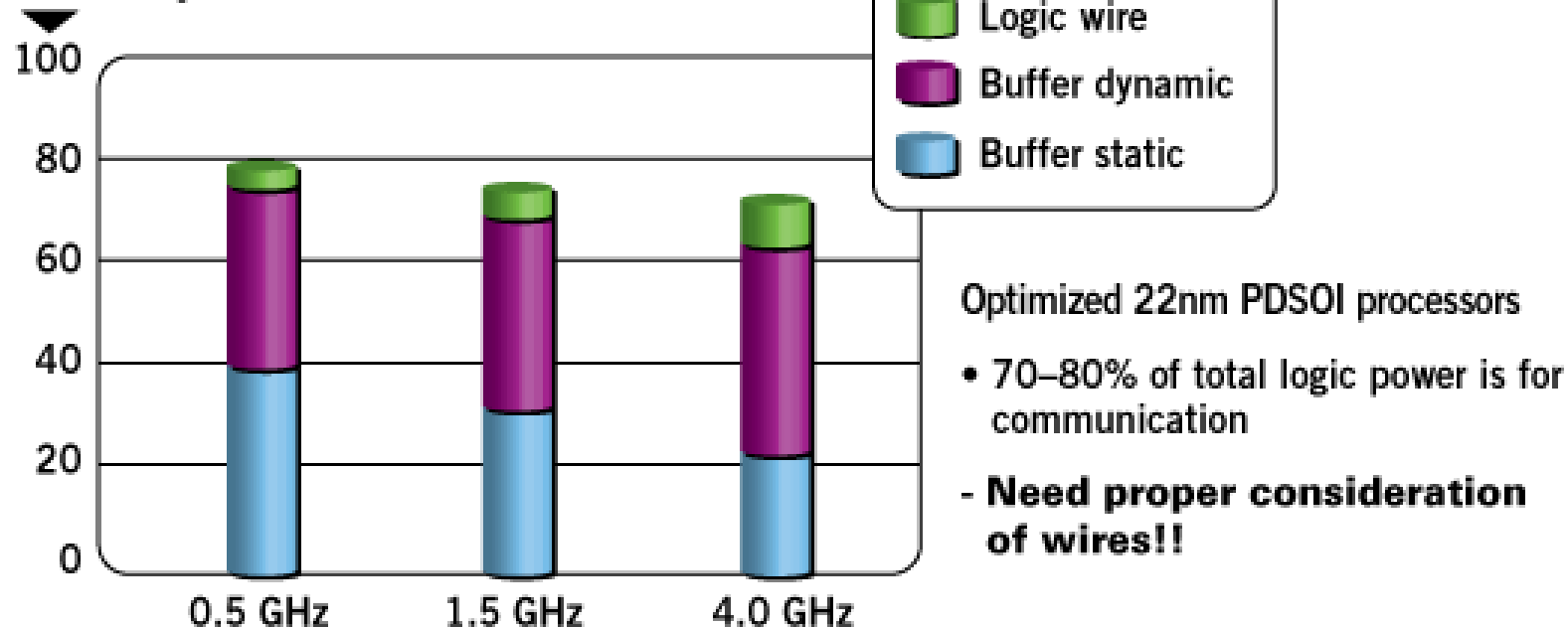
*Modern sub-100nm processes
“Transistors are free things
that fit under wires”*

- ❑ Cells sized by the number of routing tracks they can pass

Wires dominate energy as well

Communication dominates power

% of total power



Optimized 22nm PDSOI processors

- 70–80% of total logic power is for communication

- **Need proper consideration of wires!!**

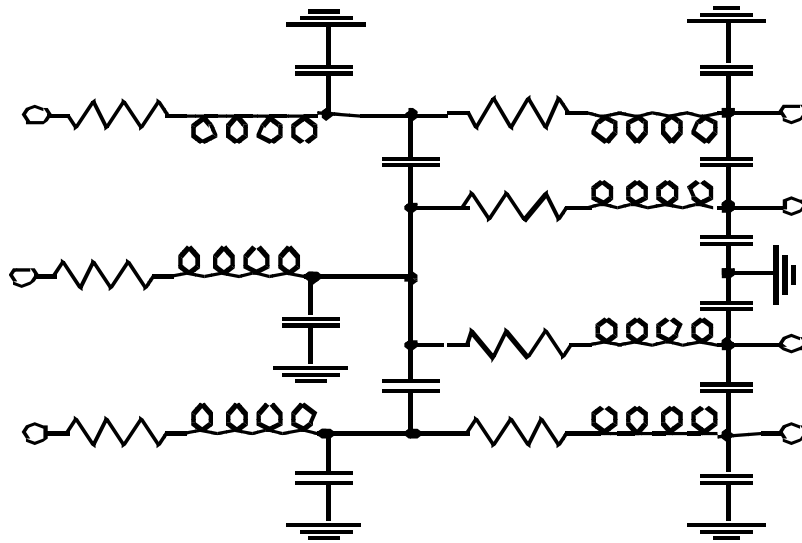
Source: L. Chang, D.J. Frank

IEDM 2012 Short Course

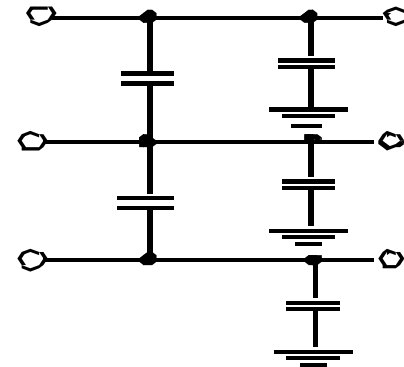
IBM T.J. Watson Research Center

- ❑ Most of the energy goes to driving the wires

Wire Models



All-inclusive model

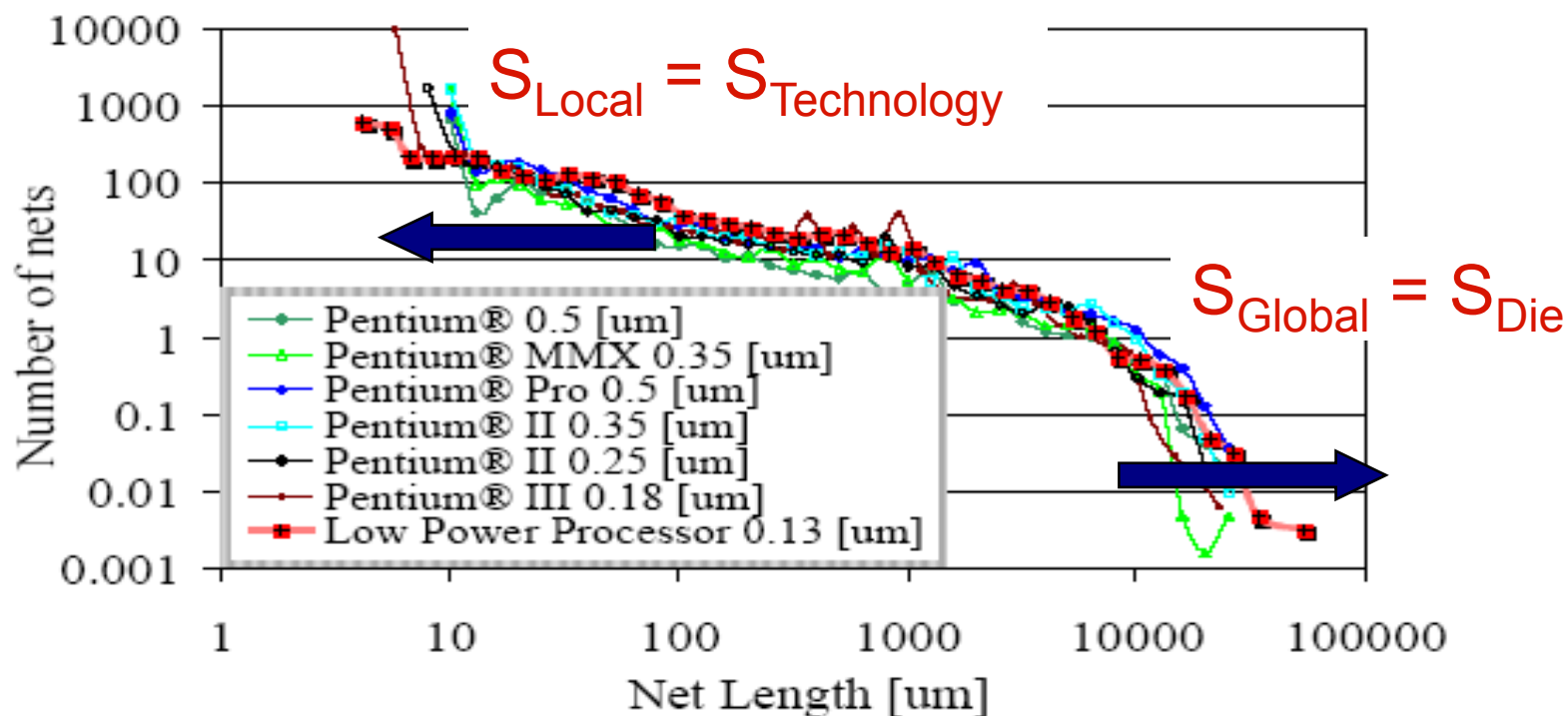


Capacitance-only

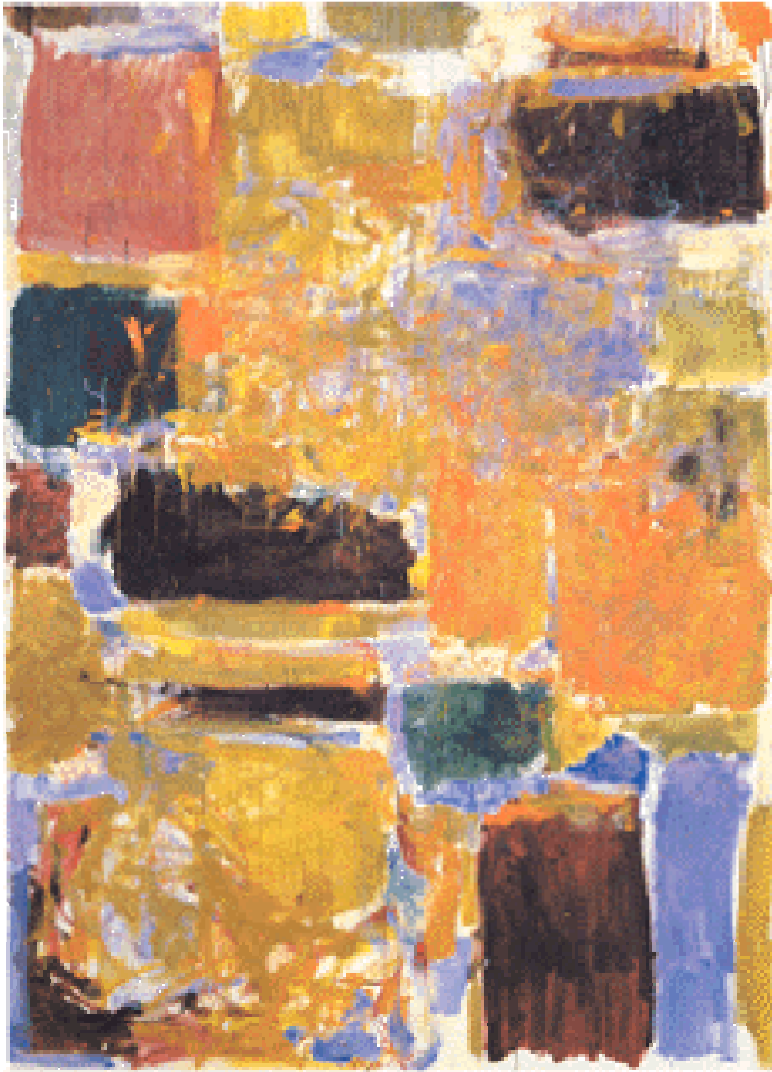
Impact of Interconnect Parasitics

- ❑ Interconnect and its parasitics can affect all of the metrics we care about
 - Cost, reliability, performance, power consumption
- ❑ Parasitics associated with interconnect:
 - Capacitance
 - Resistance
 - Inductance

Interconnect Length Distribution

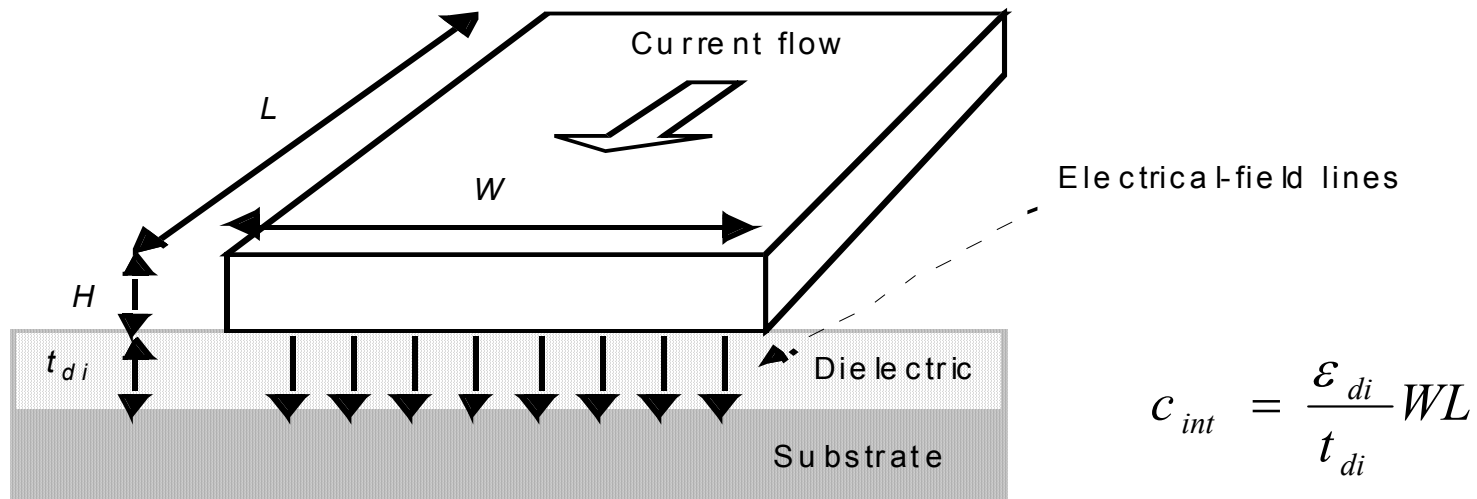


From Magen et al., "Interconnect Power Dissipation in a Microprocessor"



Capacitance

Capacitance: The Parallel Plate Model



□ Cap scaling (s – process scaling factor)

- Local wires
 - W , L , t_{di} all decrease ($\sim s$)
 - Cap decreases linearly with feature size ($\sim s$)
- Global wires
 - W , t_{di} decrease ($\sim s$), L constant
 - Cap \sim constant

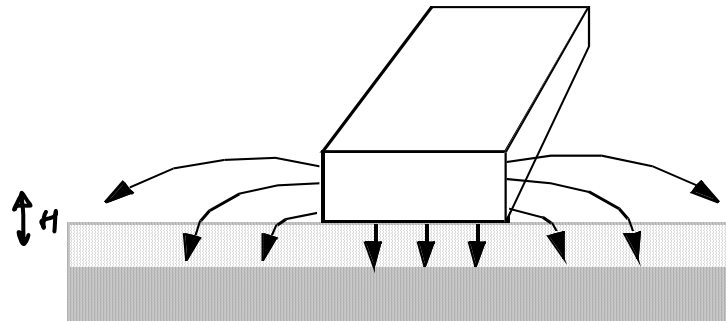
Permittivity

Material	ϵ_r
Free space	1
Aerogels	~ 1.5
Polyimides (organic)	3-4
Silicon dioxide	3.9
Glass-epoxy (PC board)	5
Silicon Nitride (Si_3N_4)	7.5
Alumina (package)	9.5
Silicon	11.7

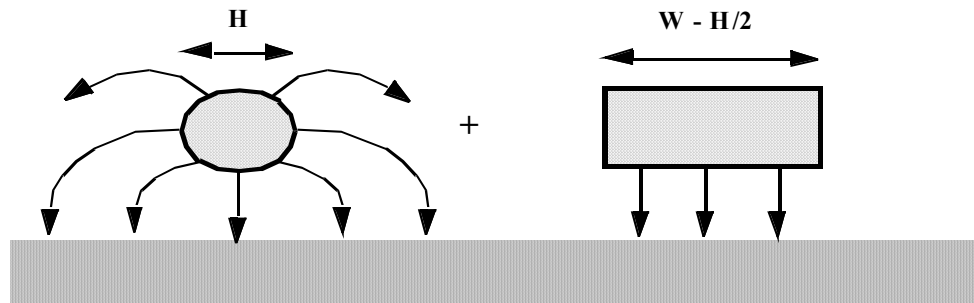
- Low-k dielectrics used sub-130nm
 - Carbon-doped oxide
 - Fabs also looking at air-gaps

Fringing Capacitance

$$C_{wire} = C_{pp} + C_{fringe} = \frac{w\epsilon_{di}}{t_{di}} + \frac{2\pi\epsilon_{di}}{\log(t_{di}/H)}$$



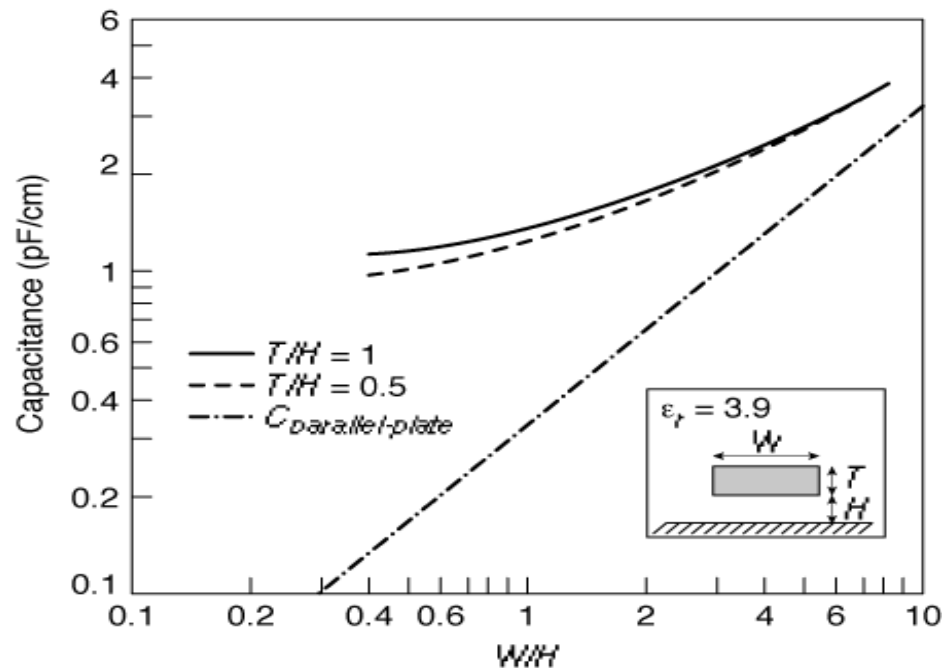
(a)



(b)

- Fringe cap per unit length \sim const
(good rule of thumb 0.2fF/ μ m)

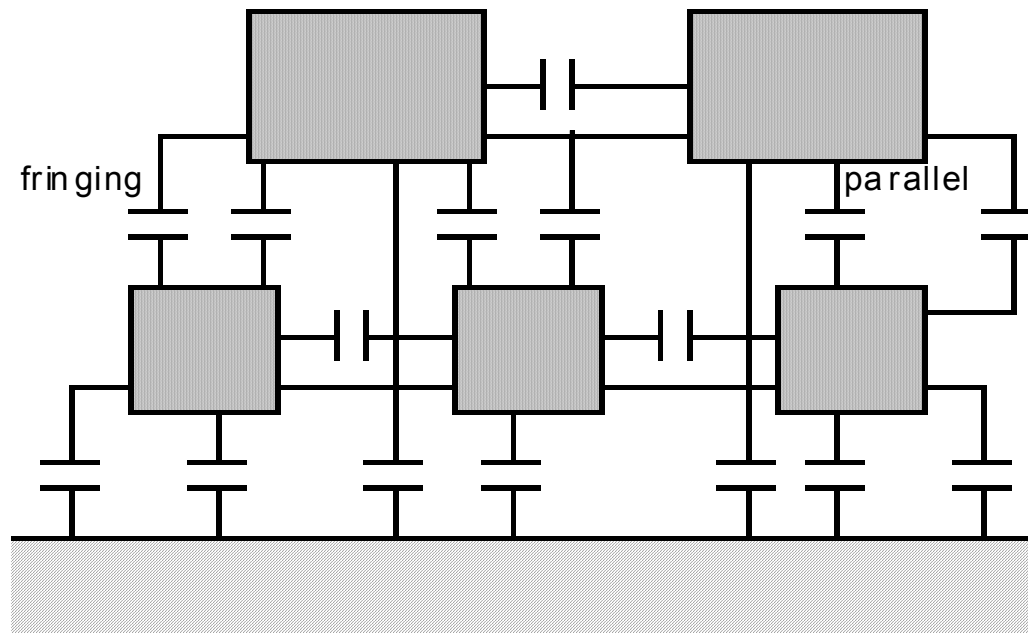
Fringing versus Parallel Plate



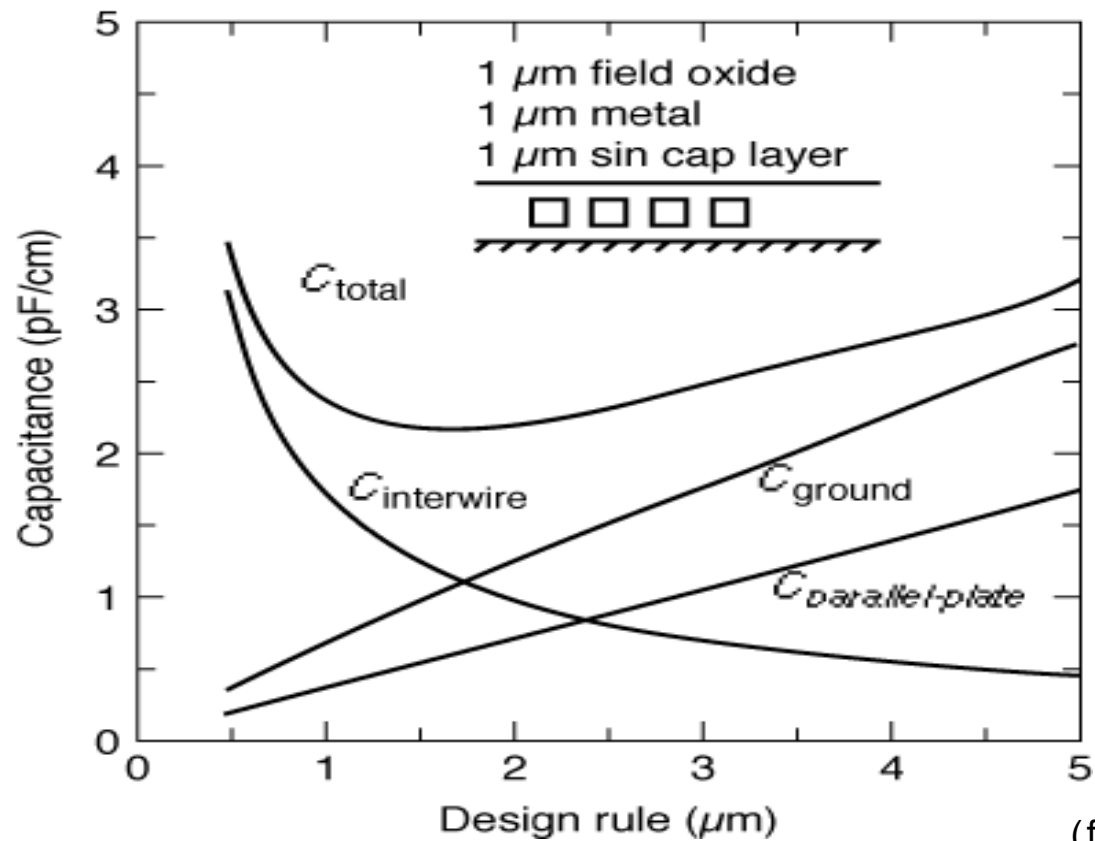
(from [Bakoglu89])

- Narrow, tall wires in modern processes
 - Trying to keep resistance from increasing
 - Comes at the expense of fringe cap

Interwire Capacitance

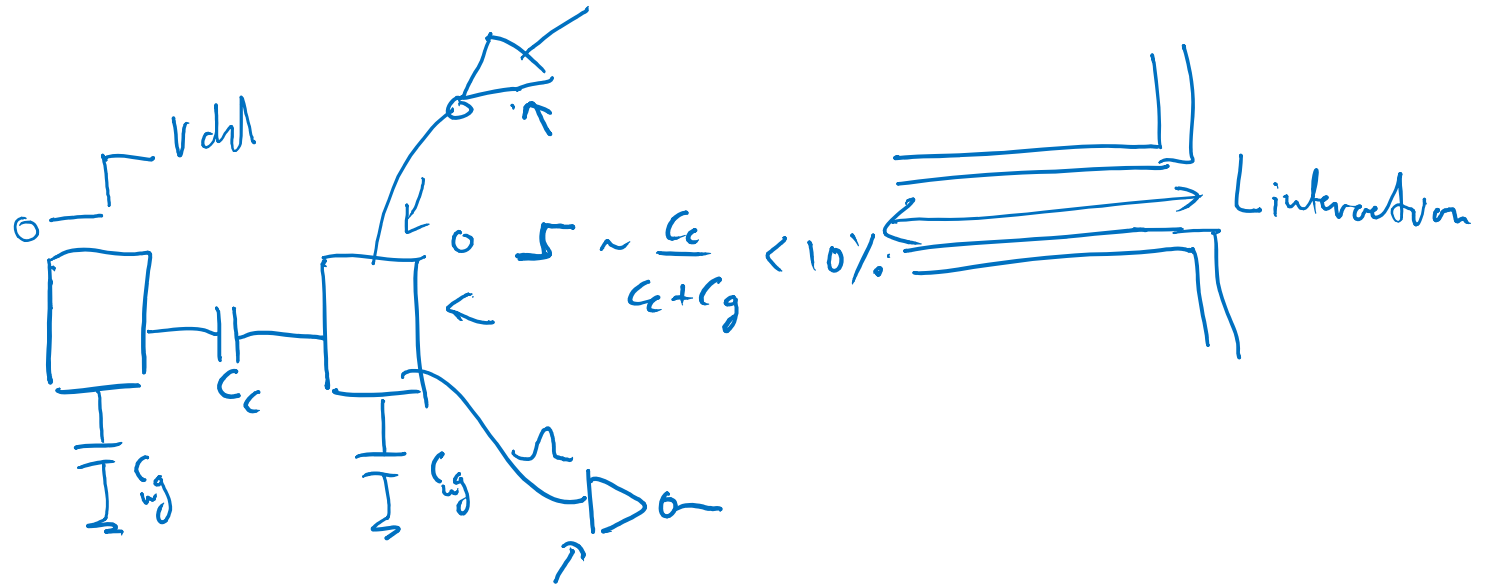


Impact of Interwire Capacitance

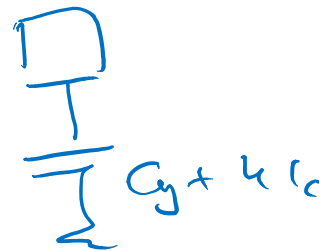
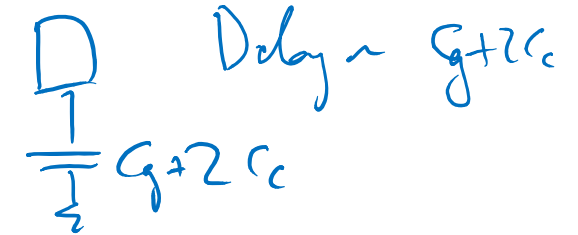
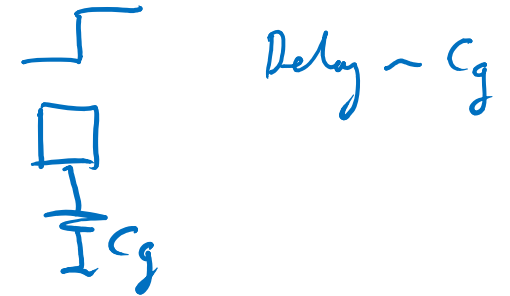
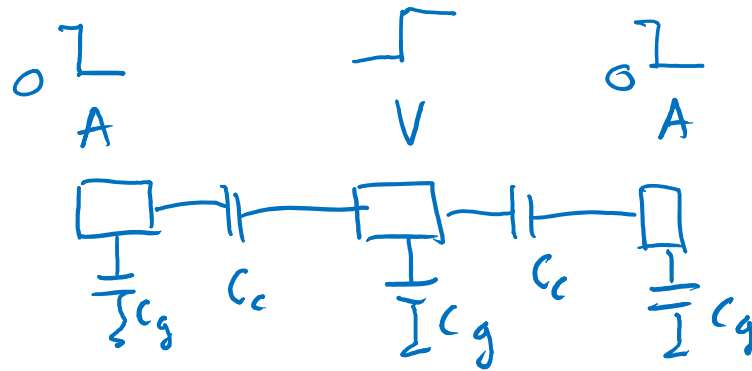


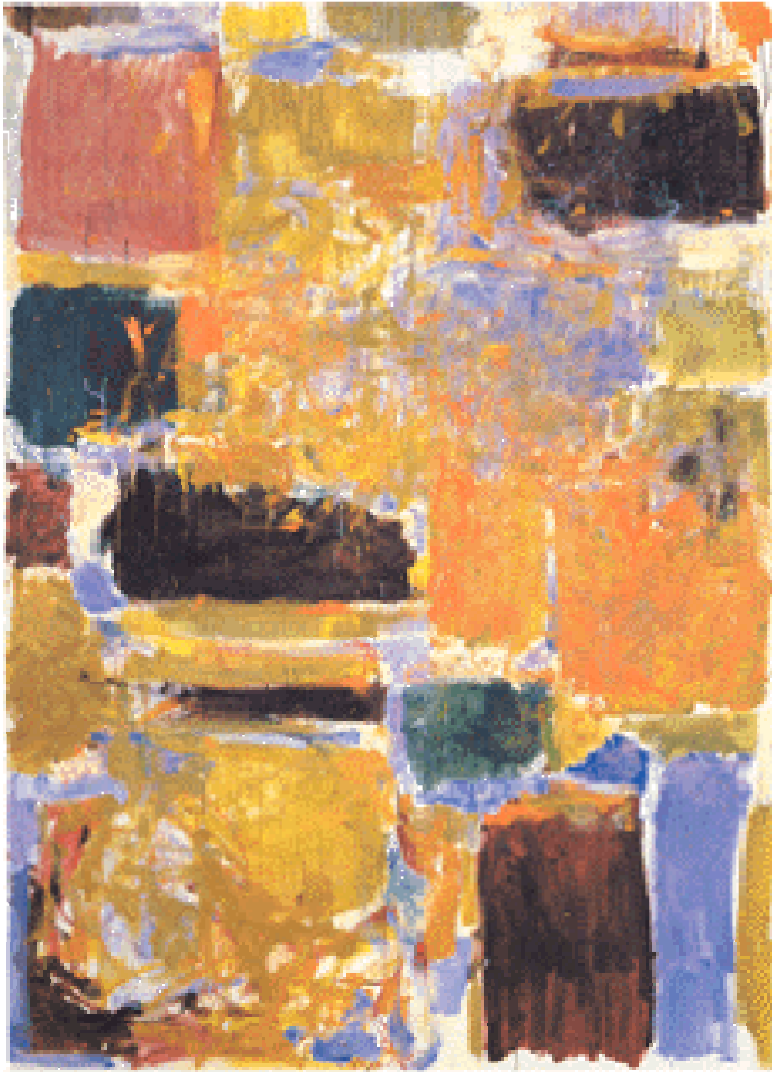
(from [Bakoglu89])

Capacitive coupling and noise



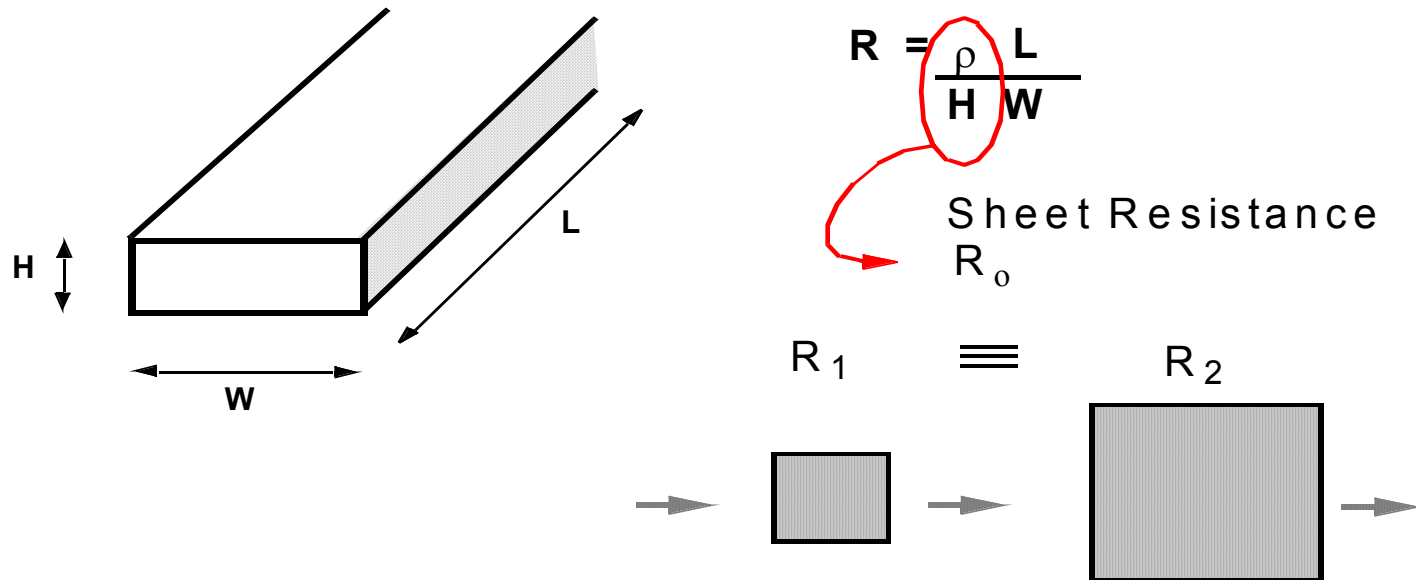
Capacitive Coupling and Delay





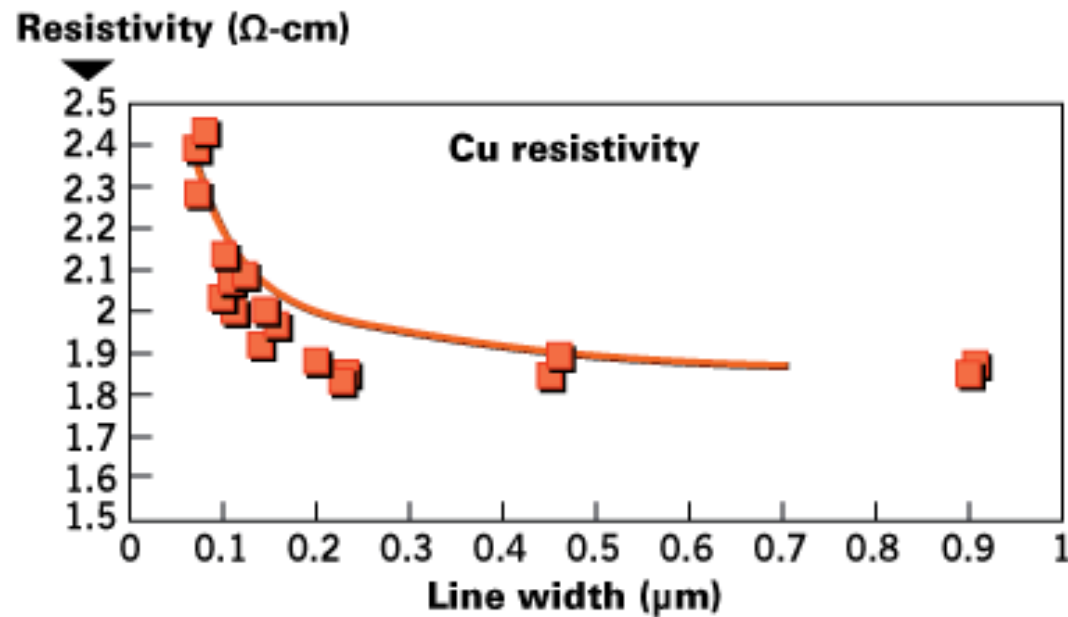
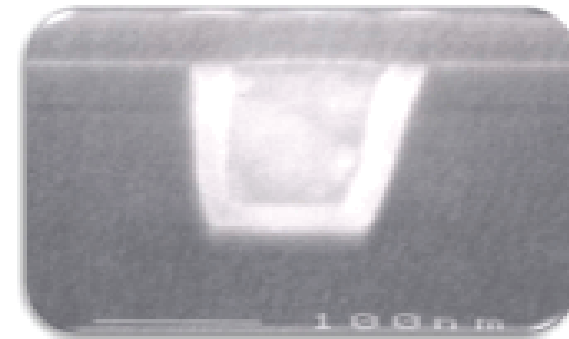
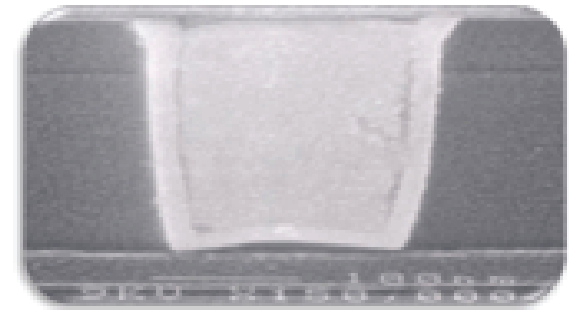
Resistance

Wire Resistance



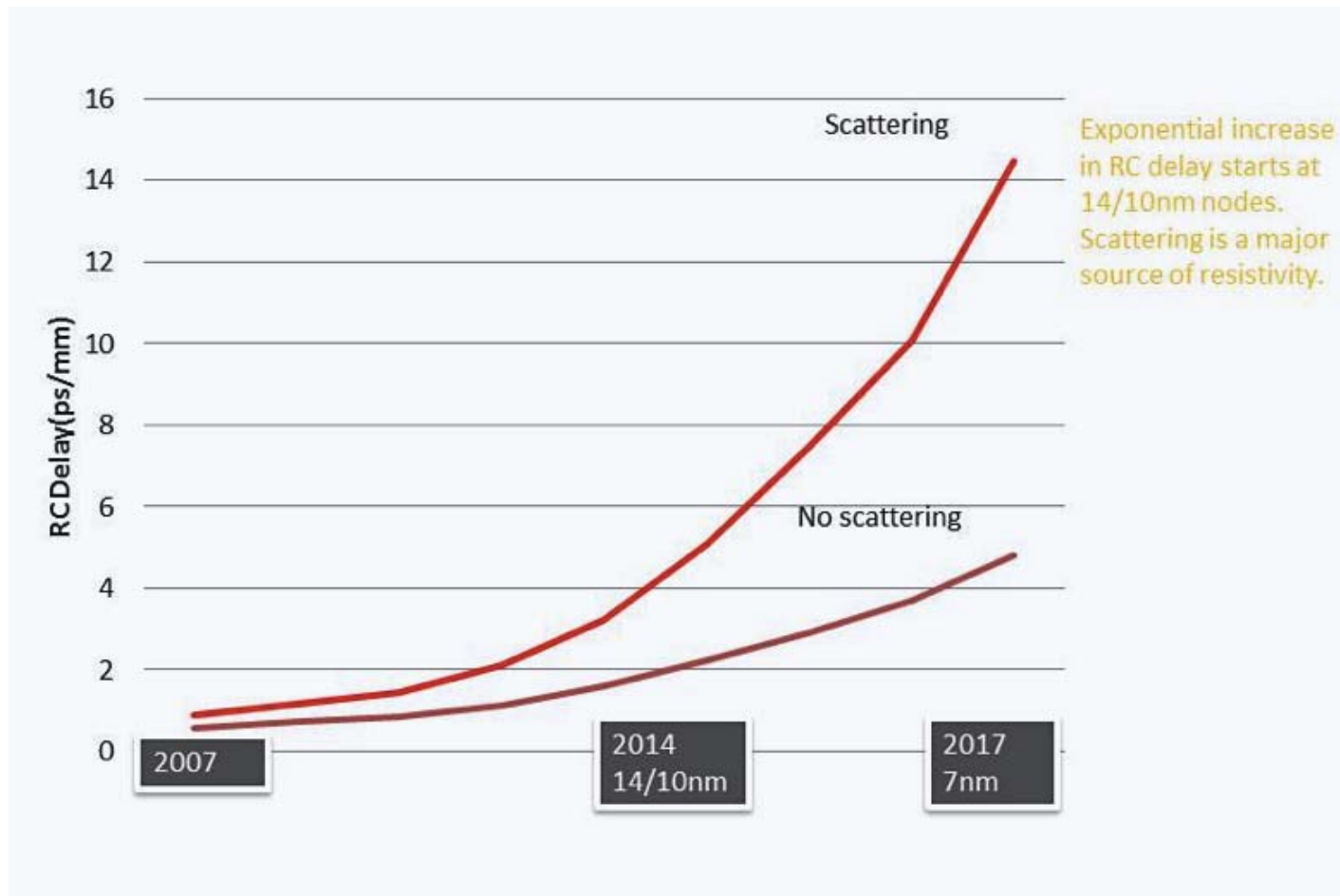
Interconnect Resistance

Material	ρ ($\Omega\text{-m}$)
Silver (Ag)	1.6×10^{-8}
Copper (Cu)	1.7×10^{-8}
Gold (Au)	2.2×10^{-8}
Aluminum (Al)	2.7×10^{-8}
Tungsten (W)	5.5×10^{-8}

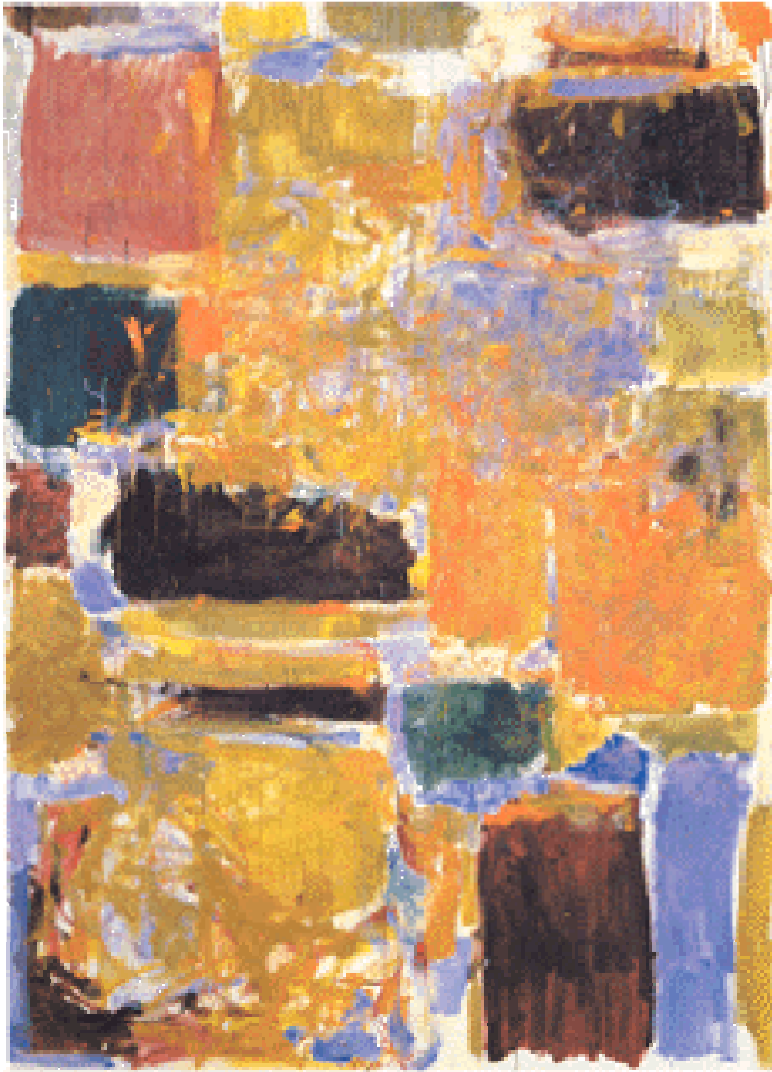


Source: S. Naffziger, AMD, VLSI 2011

Impact on Delay

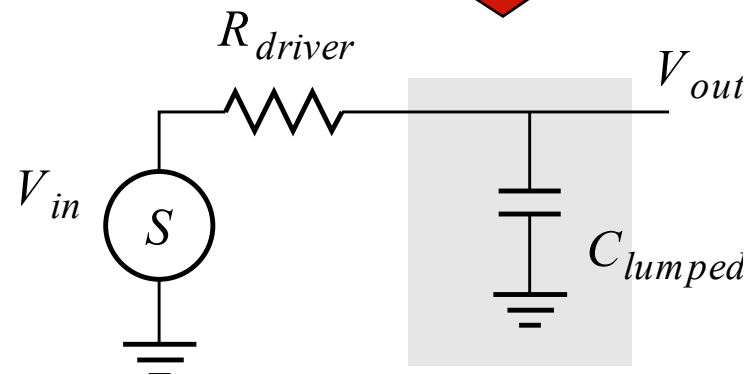
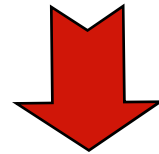
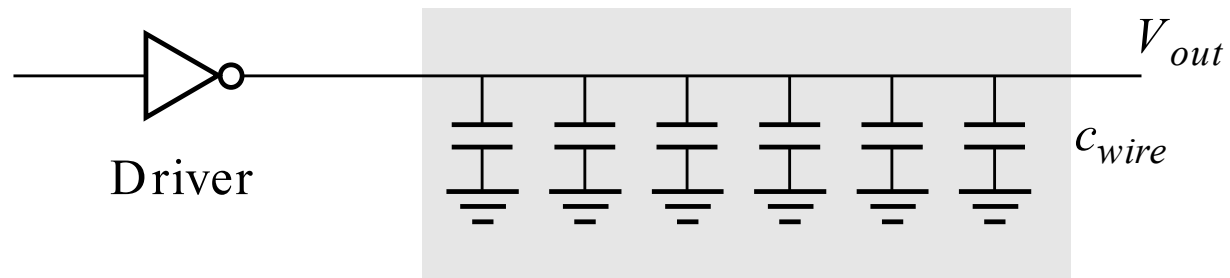


Source: Applied Materials

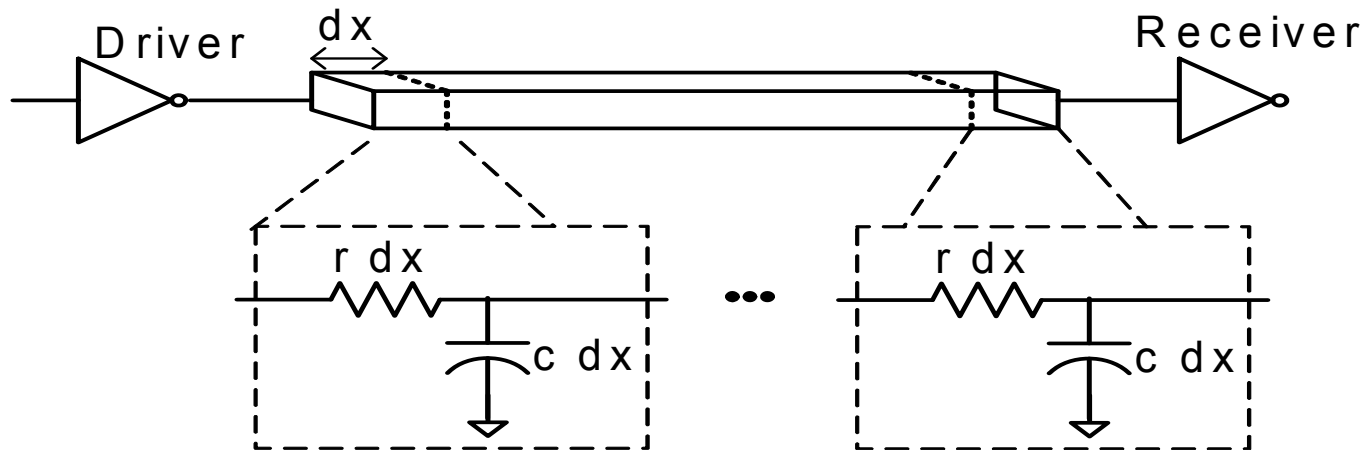


Interconnect Modeling

The Lumped Model



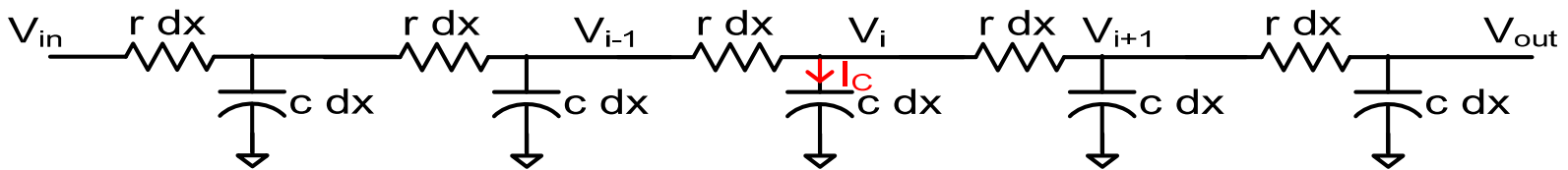
The Distributed RC-line



□ Analysis method:

- Break the wire up into segments of length dx
- Each segment resistance ($r dx$)
- Capacitance ($c dx$)

The Distributed RC-line



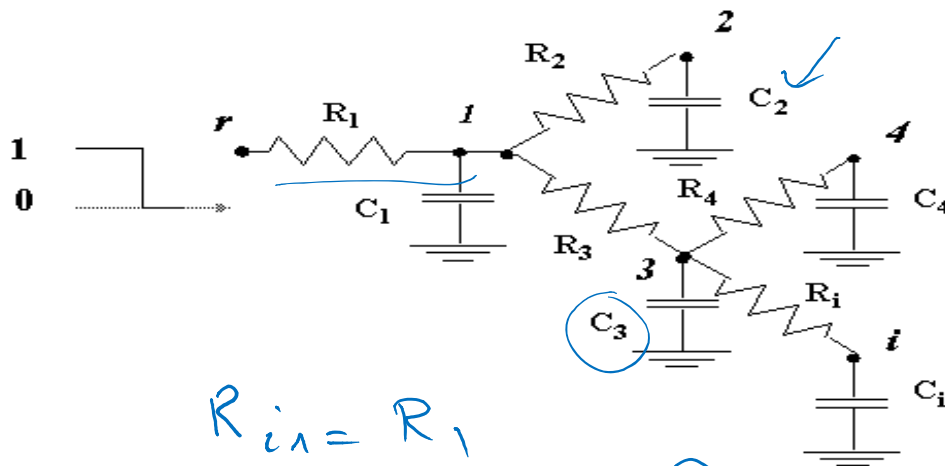
$$I_C = c \Delta L \frac{\partial V}{\partial t} = \frac{(V_{i-1} - V_i) - (V_i - V_{i+1})}{r \Delta L} \longrightarrow \boxed{rc \frac{\partial V}{\partial t} = \frac{\partial^2 V}{\partial x^2}}$$

The diffusion equation

Intermezzo – Delay of RC-networks



Delay Model of RC Networks: The Elmore Delay



$$\tau_{Di} = \sum_{k=1}^N C_k R_{ik}$$

$$R_{i1} = R_1$$

$$R_{i2} = R_1$$

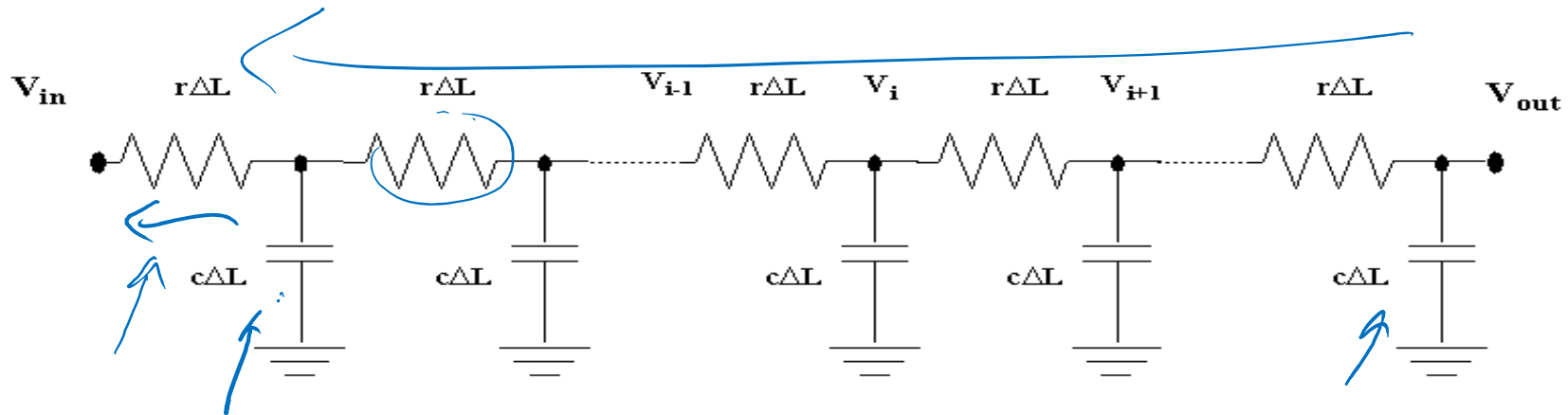
$$R_{i3} = R_1 + R_3$$

$$R_{i4} = R_1 + R_3$$

$$R_{ii} = R_1 + R_3 + R_i$$

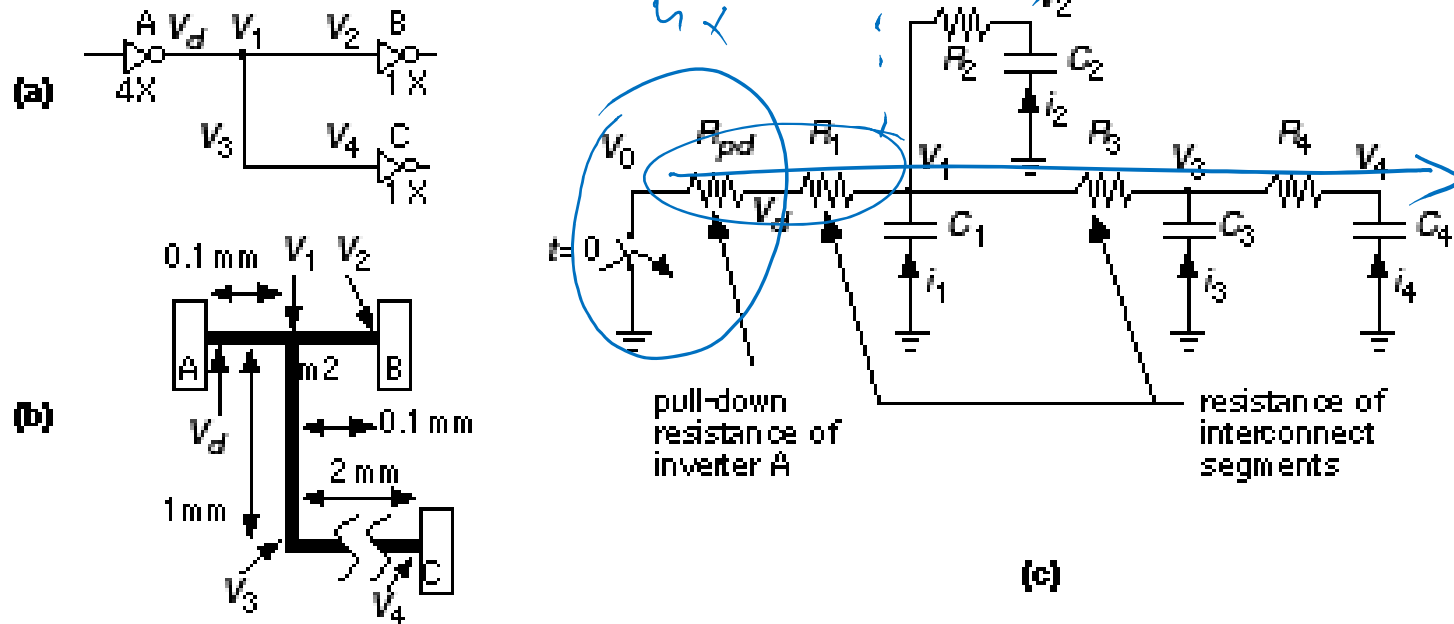
$$R_{ik} = \sum R_j \Rightarrow (R_j \in [path(s \rightarrow i) \cap path(s \rightarrow k)])$$

The Elmore Delay - RC Chain



$$\tau_N = \underbrace{\sum_{i=1}^N R_i \sum_{j=i}^N C_j}_{\text{Elmore delay}} = \underbrace{\sum_{i=1}^N C_i \sum_{j=1}^i R_j}_{\text{Elmore delay}}$$

Elmore Delay Example



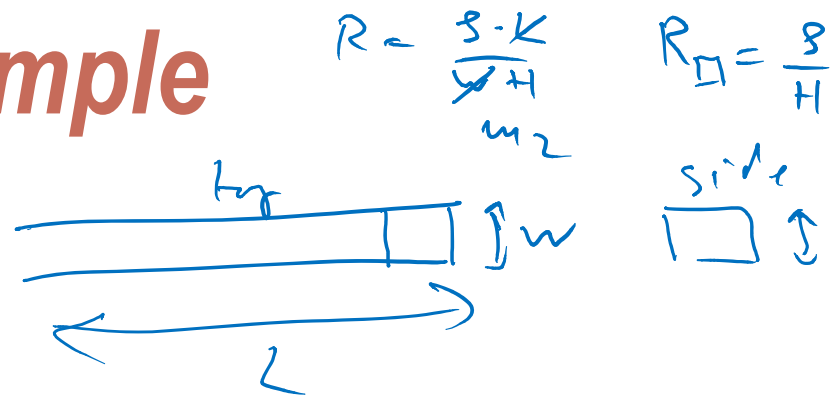
$$\tau_{D4} = \sum_{k=1}^4 C_k R_{4k} = C_1(R_{pd} + R_1) + C_2(R_{pd} + R_1) + C_3(R_{pd} + R_1 + R_3) + C_4(R_{pd} + R_1 + R_3 + R_4) =$$

$$(\underbrace{R_{pd} + R_1}_{\text{blue underline}})(C_1 + C_2 + C_3 + C_4) + R_3(C_3 + C_4) + R_4 C_4$$

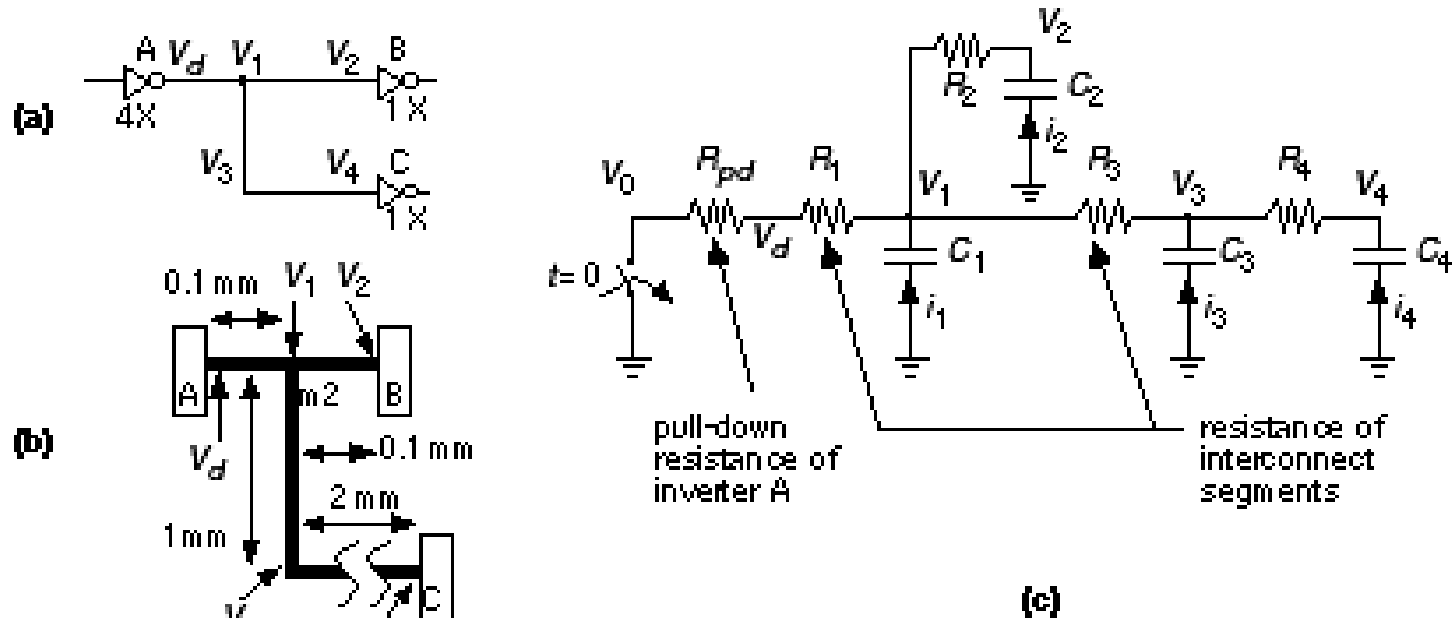
Elmore Delay Example

□ Assume:

- m2 r is 50 mΩ/□
- m2 c (for a minimum-width line) is 0.2fF/um
- m2 minimum width is 90nm
- 4x inverter $R_{pd} = 500\Omega$
- 1X inverter input capacitance is 20fF (a standard load)



Elmore Delay Example



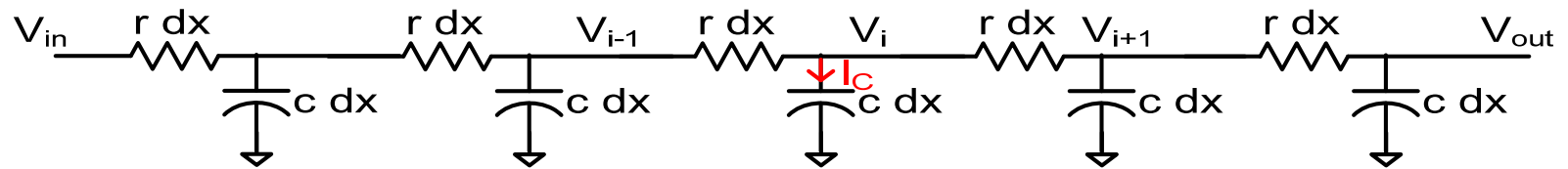
$$R_1 = R_2 = \frac{0.1\text{mm}}{90\text{nm}} 50\text{m}\Omega = 6\Omega, R_3 = 56\Omega, R_4 = 112\Omega$$

$$C_1 = \frac{0.1\text{mm} \cdot 0.2\text{fF}}{\text{um}} = 20\text{fF}, C_2 = C_1 + 20\text{fF} = 40\text{fF}$$

$$C_3 = 200\text{fF}, C_4 = 420\text{fF}$$

$$\tau_{D4} = 340\text{ps}$$

Back to Wire Delay



Wire Model

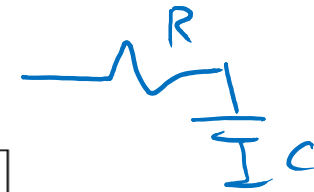
Model the wire with N equal-length segments:

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

For large values of N:

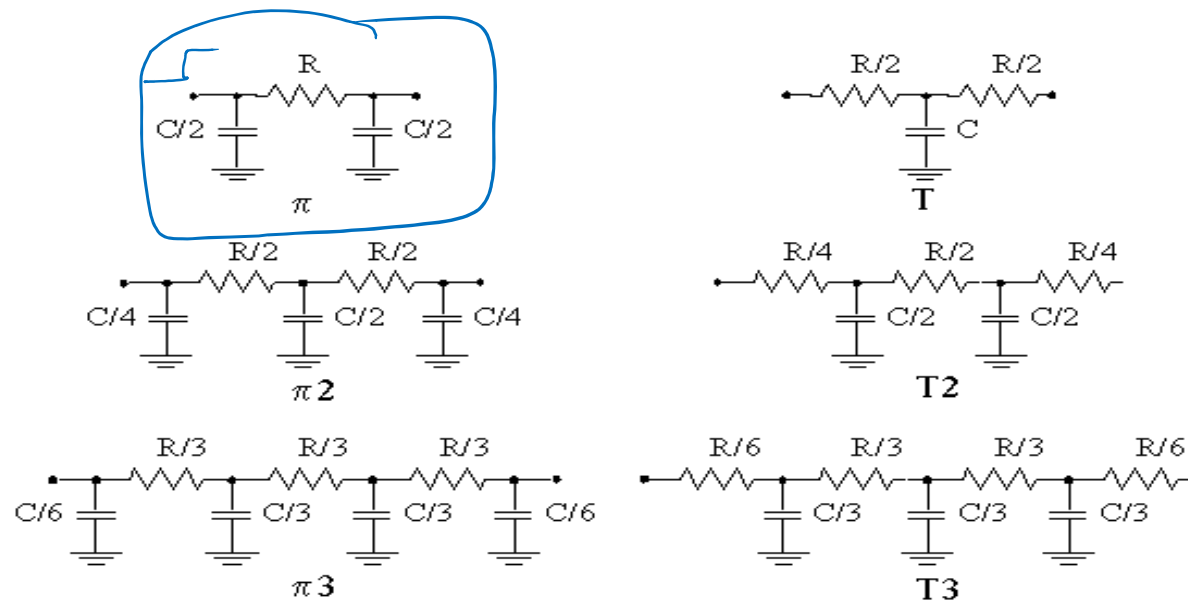
$$\tau_{DN} = \frac{RC}{2} = \frac{rcL^2}{2}$$

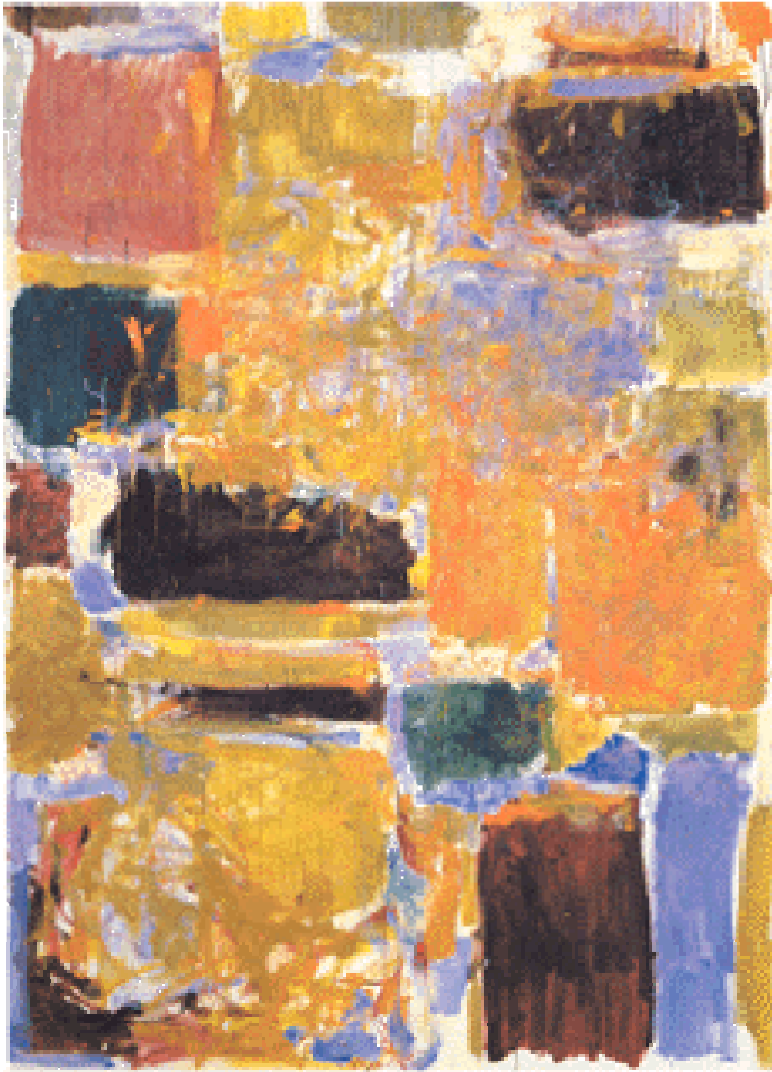
RC-Models



Voltage Range	Lumped RC-network	Distributed RC-network
0→50% (t_p)	0.69 RC	0.38 RC
0→63% (τ)	RC	0.5 RC
10%→90% (t_r)	2.2 RC	0.9 RC

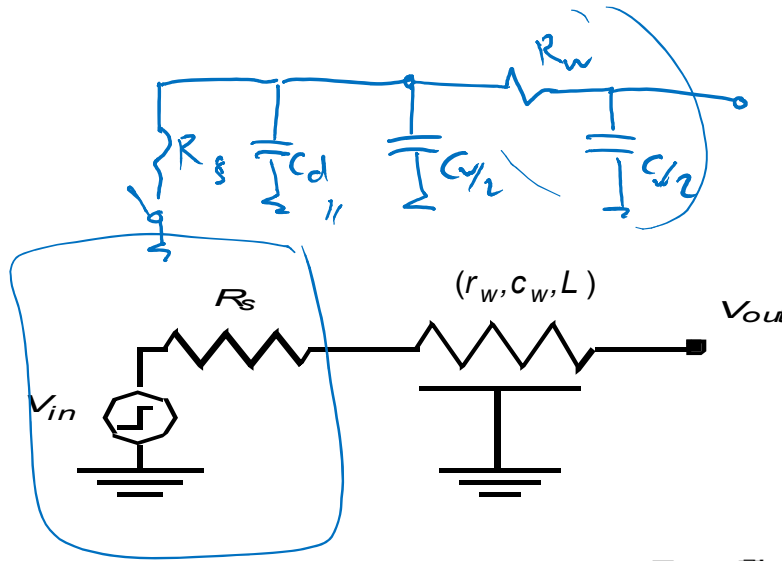
Step Response of Lumped and Distributed RC Networks:
Points of Interest.





Gates and Wires

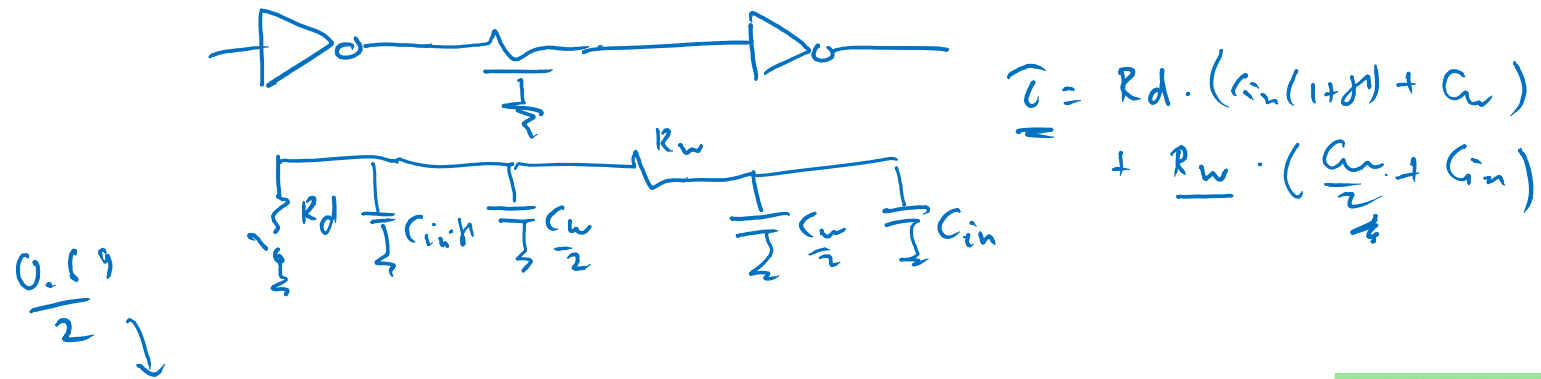
Driving an RC-line



$$\tau_D = R_s C_w + \frac{R_w C_w}{2} = R_s C_w + 0.5 r_w c_w L^2$$

$$t_p = 0.69 R_s C_w + 0.38 R_w C_w$$

The Global Wire Problem

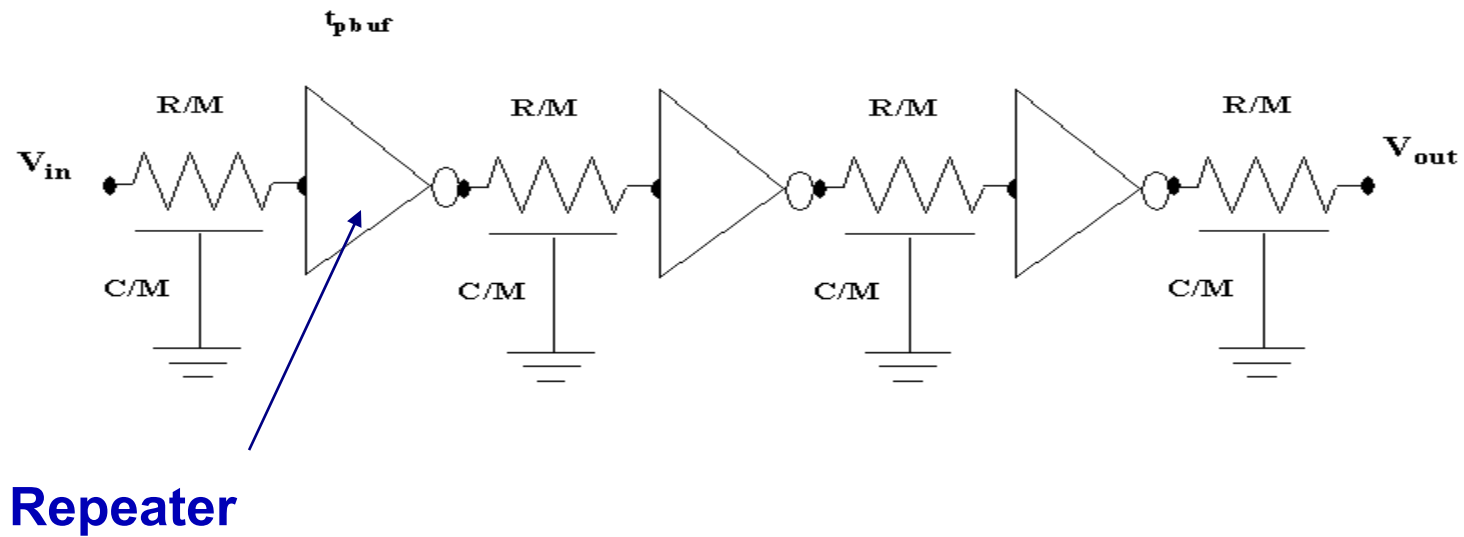


$$T_d = 0.38R_wC_w + 0.69(R_NC_{in} + R_NC_w + R_wC_{in})$$

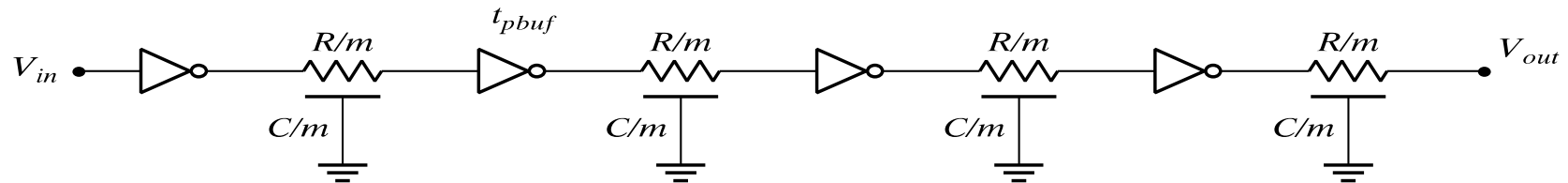
Challenges

- ❑ No further improvements to be expected after the introduction of Copper (superconducting, optical?)
- ❑ Design solutions
 - Use of fat wires
 - Efficient chip floorplanning
 - Insert repeaters

Reducing RC-delay Using Repeaters



Repeaters



$$t_p = 0.69m \left(\frac{R_N}{W} (W\gamma C_{in} + \frac{cL}{m} + WC_{in}) + \frac{rL}{m} (WC_{in} + 0.5 \frac{cL}{m}) \right)$$

$$m_{opt} = L \sqrt{\frac{0.38rc}{0.69R_N C_{in} (\gamma + 1)}} = \sqrt{\frac{t_{pwire(unbuffered)}}{t_{p1}}}$$

$$W_{opt} = \sqrt{\frac{R_N c}{r C_{in}}}$$

Repeater Insertion

$$m_{opt} = L \sqrt{\frac{0.38rc}{0.69R_N C_{in}(\gamma + 1)}} = \sqrt{\frac{t_{pwire(unbuffered)}}{t_{p1}}}$$

$$W_{opt} = \sqrt{\frac{R_N c}{r C_{in}}}$$

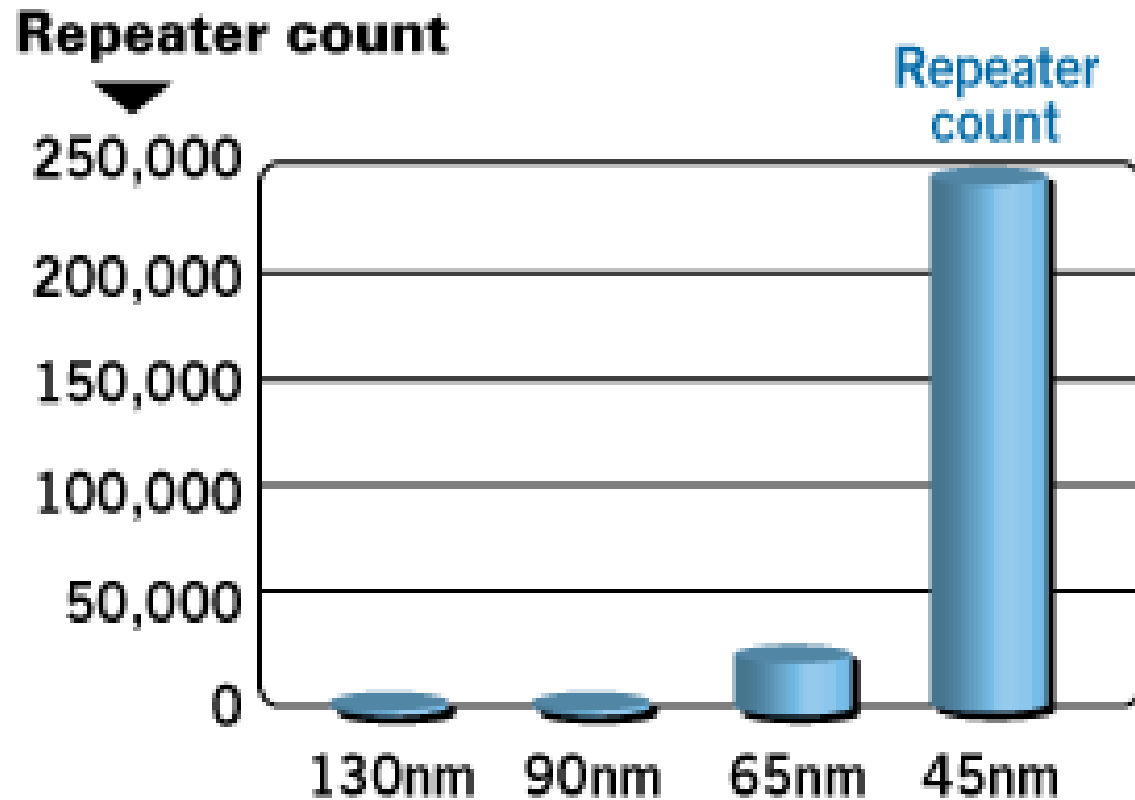
For a given technology and a given interconnect layer, there exists an optimal length of the wire segments between repeaters. The delay of these wire segments is **independent of the routing layer!**

$$L_{crit} = \frac{L}{m_{opt}} = \sqrt{\frac{t_{p1}}{0.38rc}}$$

$$t_{p, crit} = \frac{t_{p, min}}{m_{opt}} = 2 \left(1 + \sqrt{\frac{0.69}{0.38(1 + \gamma)}} \right) t_{p1}$$

From Elmore example: $rc = 0.1\text{fs}/\mu\text{m}^2$, $t_{p1} = 55\text{ps}$, $L_{crit} = 741\mu\text{m}$
(rule of thumb $\sim 0.5\text{-}1\text{mm}$)

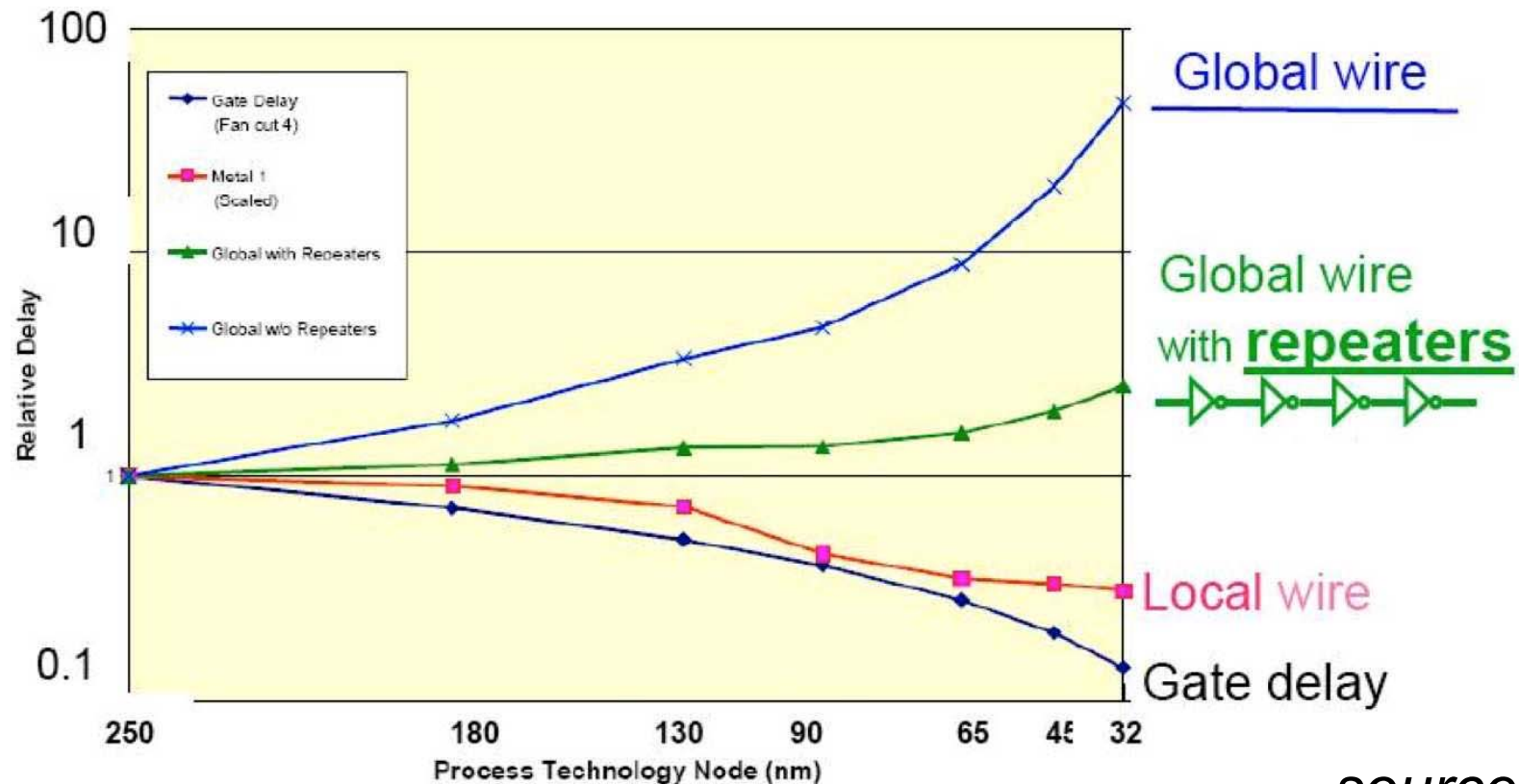
Importance of Repeaters



Source: IBM POWER processors,
R. Puri et al SRC Interconnect Forum 2006

- ❑ In modern designs the number of repeaters increases dramatically

Wire and Gate Delay Scaling



Delay for Metal 1 and Global Wiring versus Feature Size

□ Gate delay gets better, wire delay gets worse