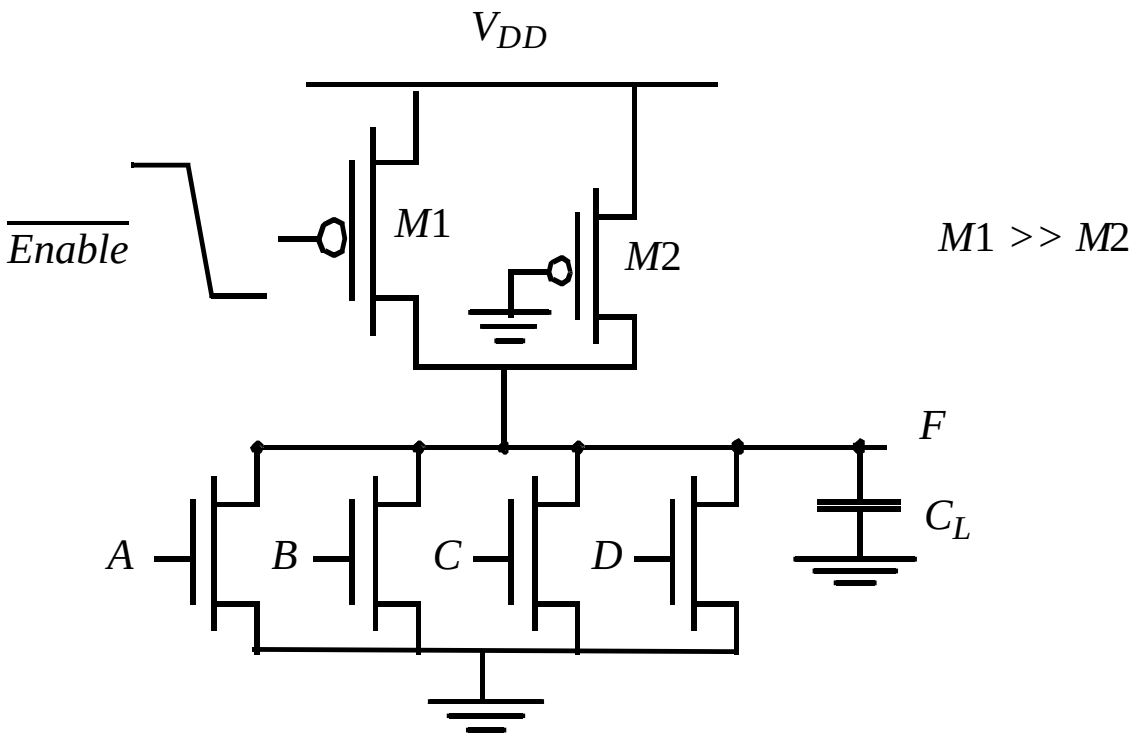
ECE 225
High-Speed Digital IC Design

Lecture 8

Prof. Kaustav Banerjee
Electrical and Computer Engineering
E-mail: kaustav@ece.ucsb.edu

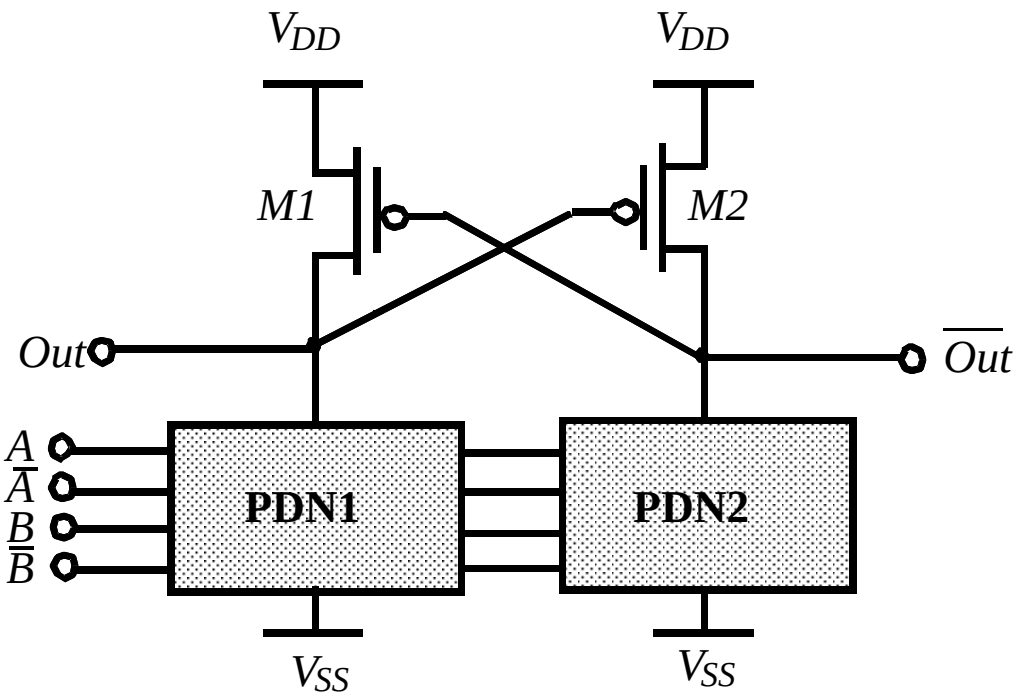
Designing Combinational Logic Circuits

Improved Loads



Adaptive Load

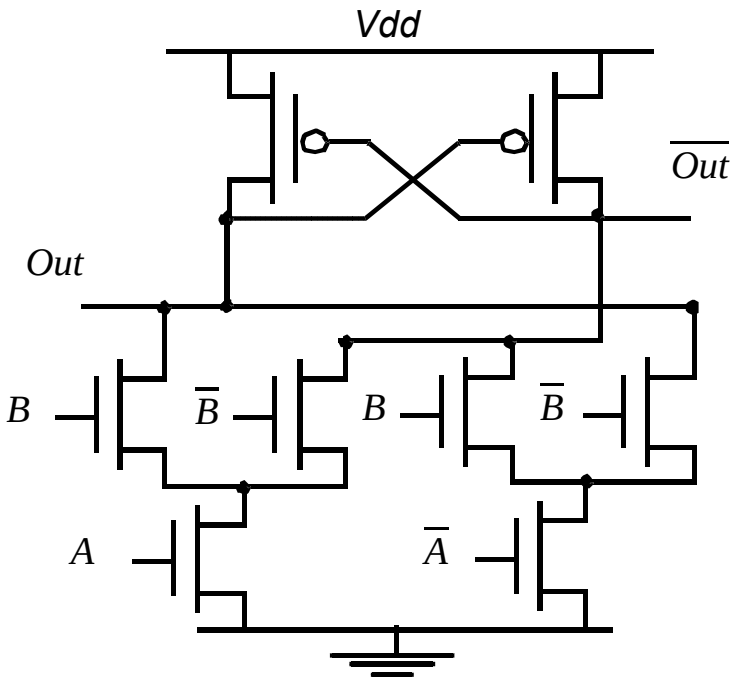
Improved Loads (2)



Differential Cascode Voltage Switch Logic (DCVSL)

- ***Completely eliminates static currents***
- ***Provides rail to rail swing***
- ***Differential Gate: each input is provided in a complementary format***
- ***Feedback Mechanism: ensures that the load device is turned off when not needed***
- ***PDNs are mutually exclusive: when PDN1 is ON, PDN2 is OFF and vice versa***
- ***If Out=1, $\overline{Out}$=0, PDN1=1 causes Out node to be pulled down with some contention between M1 and PDN1***
- ***As Out=Vdd-Vtp, M2 turns on and starts charging $\overline{Out}$ to Vdd and M1 is turned off enabling Out to discharge completely***

DCVSL Example (1)

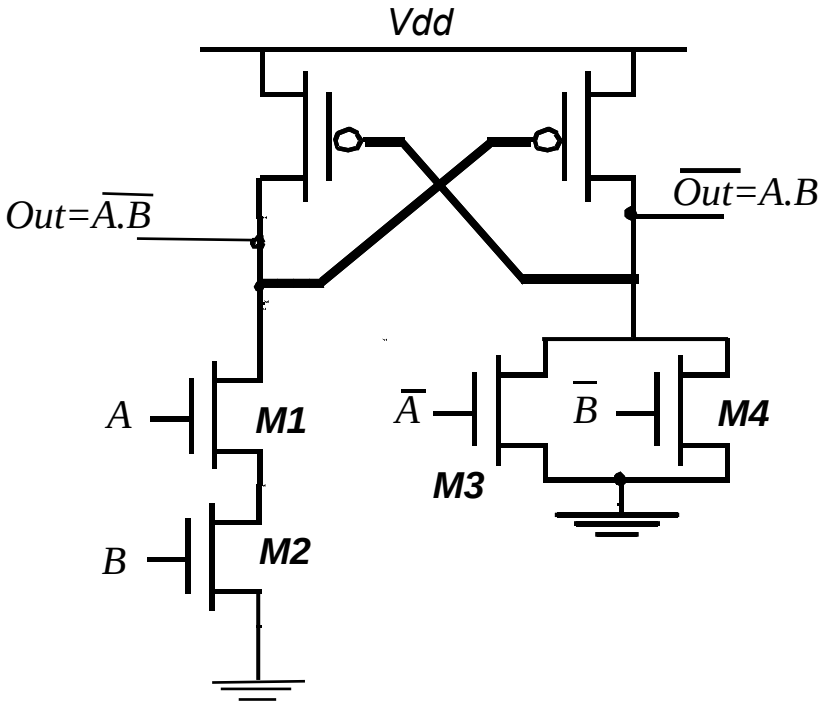


XOR-NXOR gate

A	B	Out=XOR	$\overline{Out}$=XNOR
1	0	1	0
0	1	1	0
0	0	0	1
1	1	0	1

Note: The circuit is still Ratioed!

DCVSL Example (2)



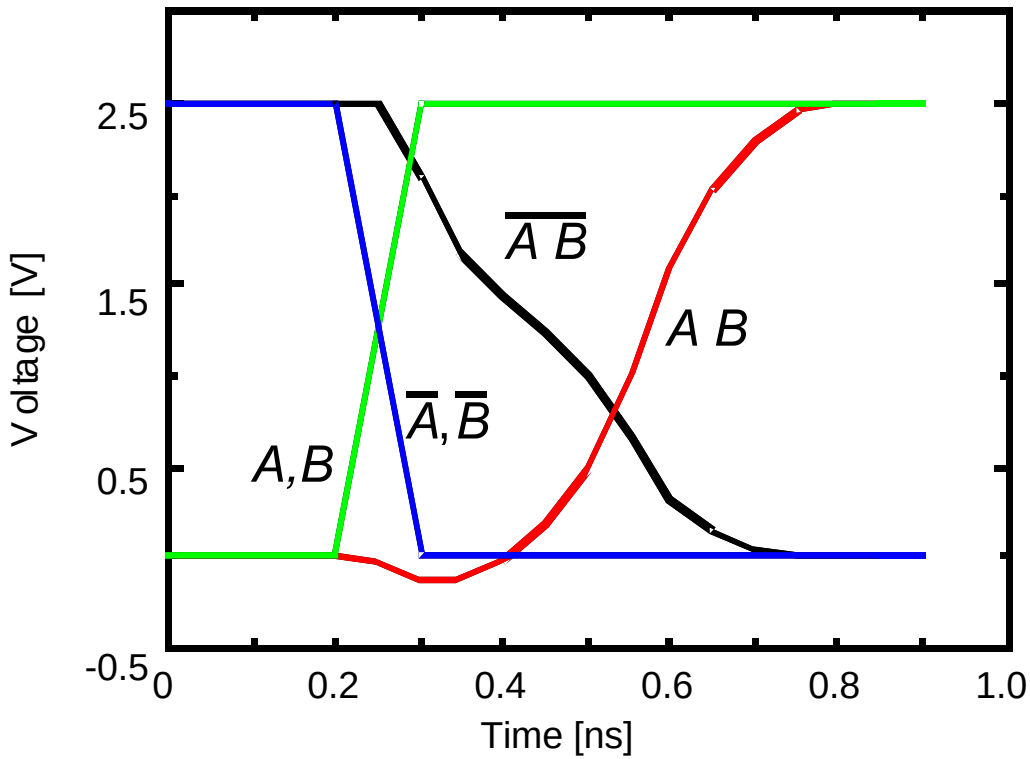
AND-NAND gate

As Out=Vdd-Vtp, $\overline{Out}$ starts to charge up to Vdd

A	B	Out=NAND	$\overline{Out}$=AND
1	0	1	0
0	1	1	0
0	0	1	0
1	1	0	1

Note: The circuit is still Ratioed!

Transient Response: AND/NAND DCVSL Gate



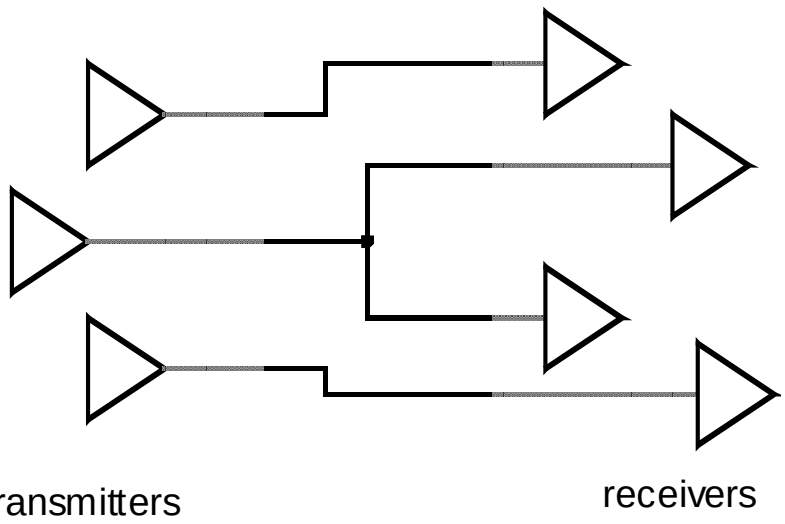
Delay from Input (A, B) to Out ($\overline{AB}$) is 197 ps and to $\overline{Out}$ (A.B) is 321 ps

In comparison a static CMOS AND gate: NAND+INV has a delay of 200 ps

Design Issues with DCVSL Gates

- Pros:
  - Provides differential (or complementary outputs)
  - Reduces number of required gates (for complex function) by almost a factor of 2!
  - Number of gates in the critical timing path is often reduced
  - Prevents some time differential problems introduced by the use of additional inverters (when both a signal and its complement are needed simultaneously)
- Cons:
  - Increased design complexity
  - Doubles the number of interconnects
  - High dynamic power dissipation: more power dissipation due to cross-over (short-circuit) currents during the transition

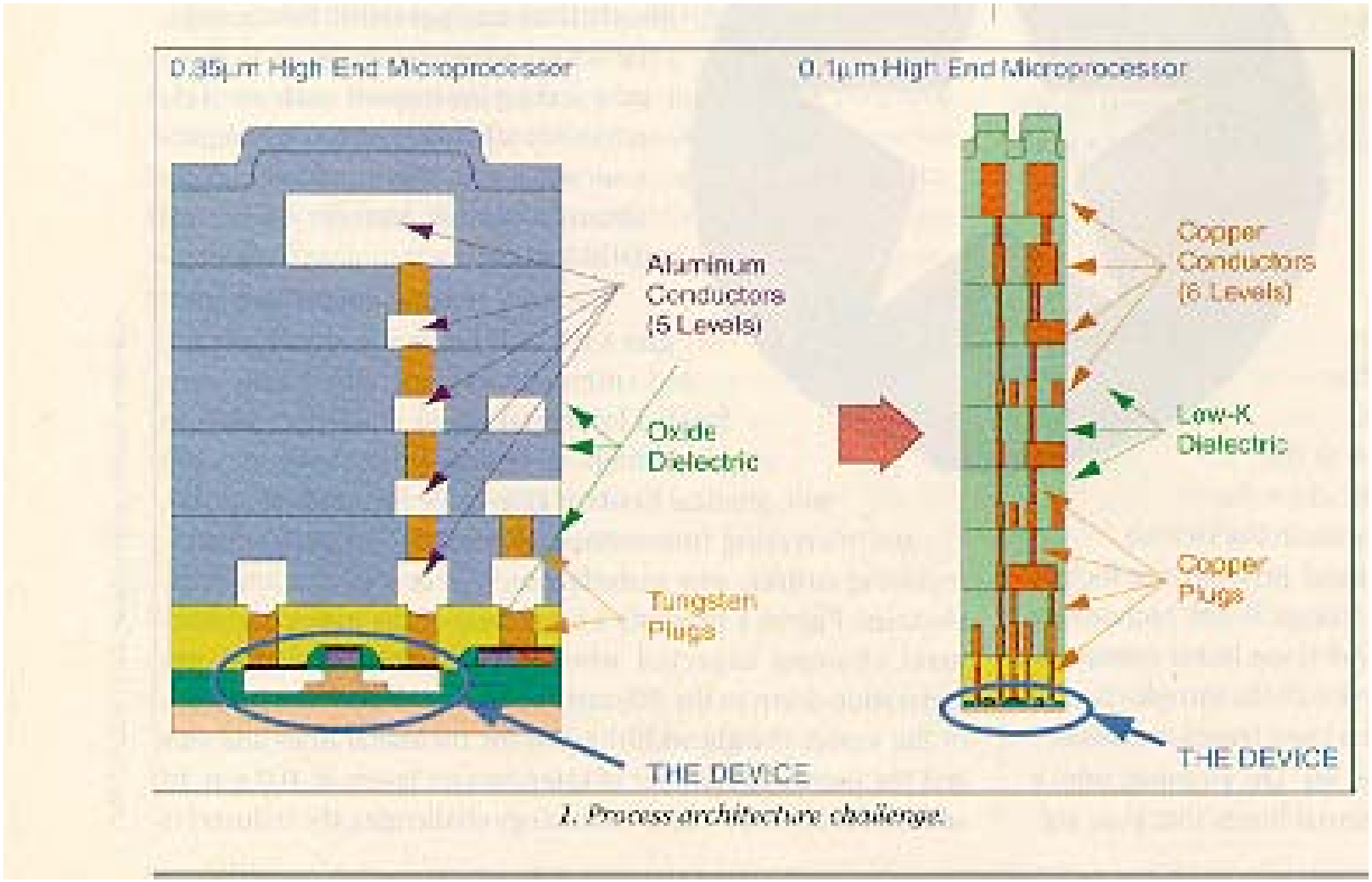
The Wire



schematics

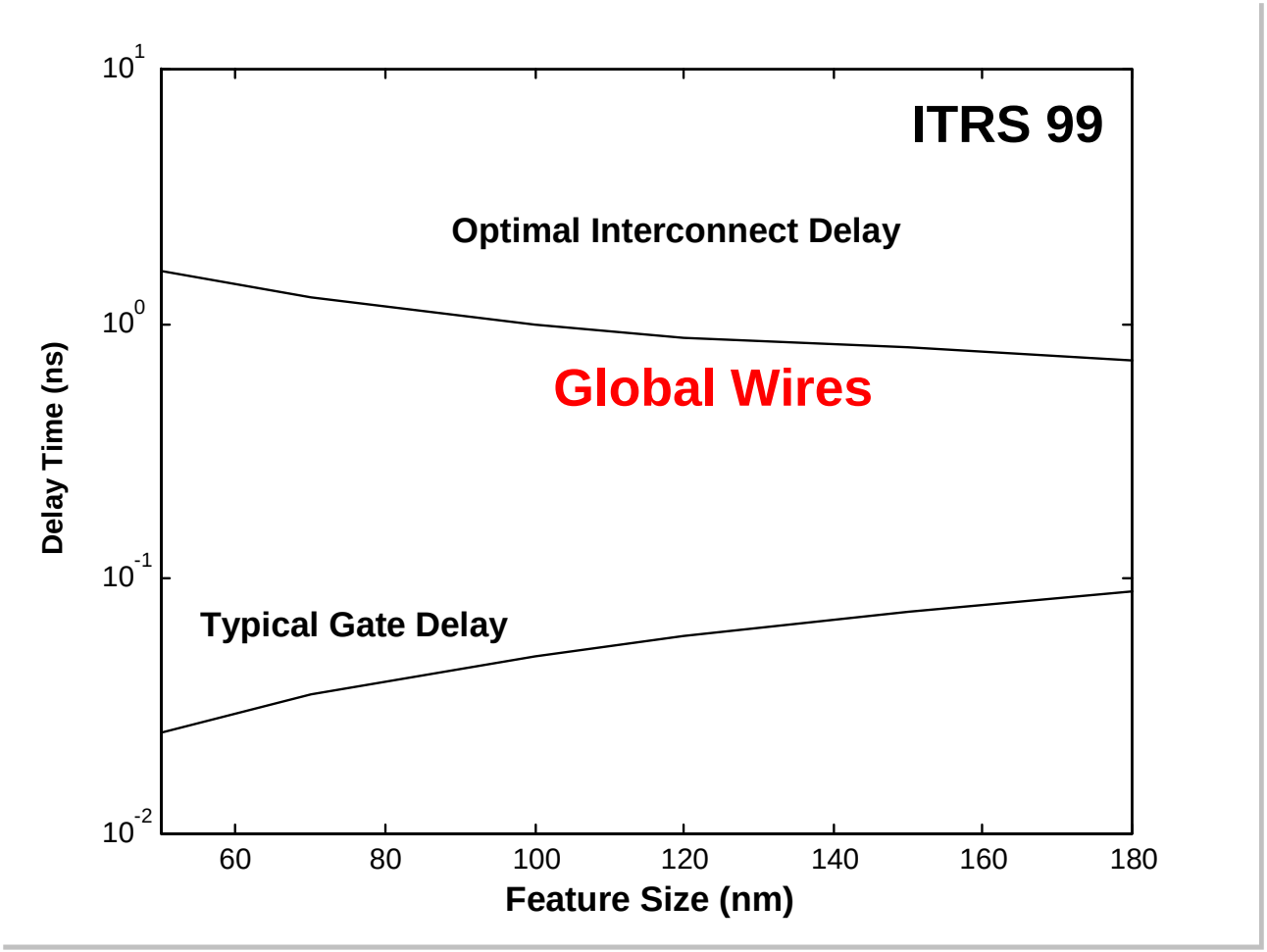
physical

Interconnect Impact on Chip



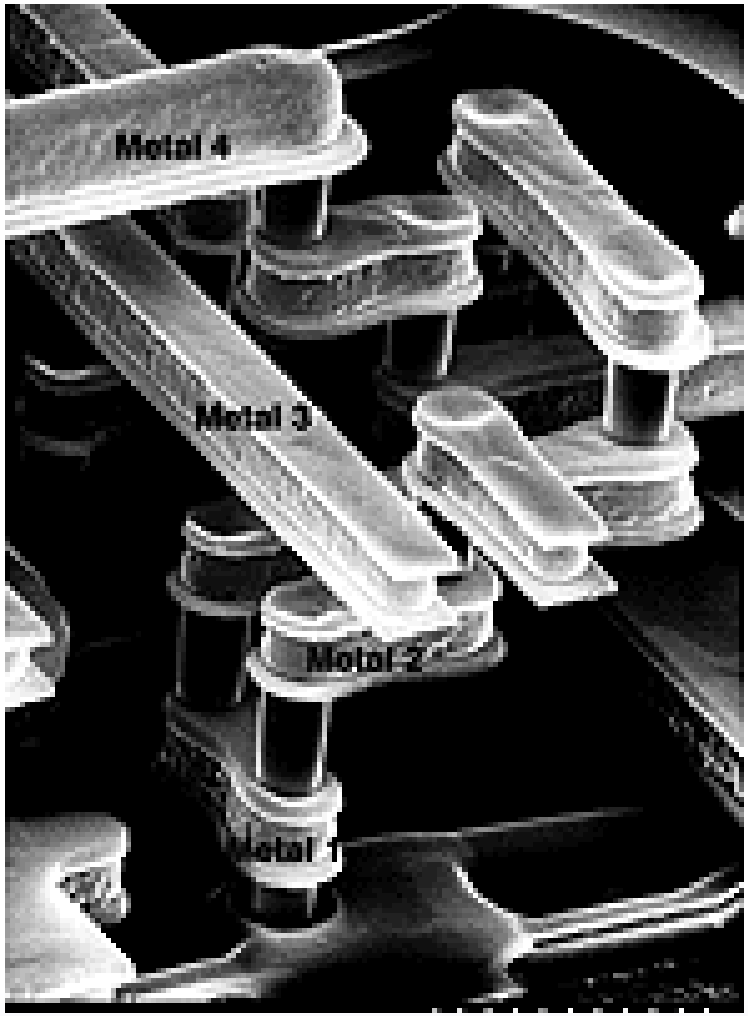
Interconnect delay has become the dominant factor determining chip performance……..

IC performance is being dominated by interconnects…..



K. Banerjee et al., Proc. IEEE, May 2001.

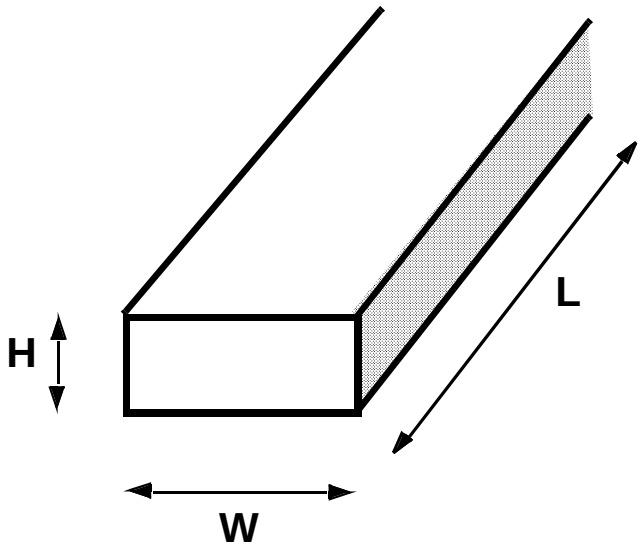
Modern Interconnect



Impact of Interconnect Parasitics

- Interconnect parasitics
  - affect performance and power consumption
  - affect reliability
- Classes of parasitics
  - Capacitive
  - Resistive
  - Inductive

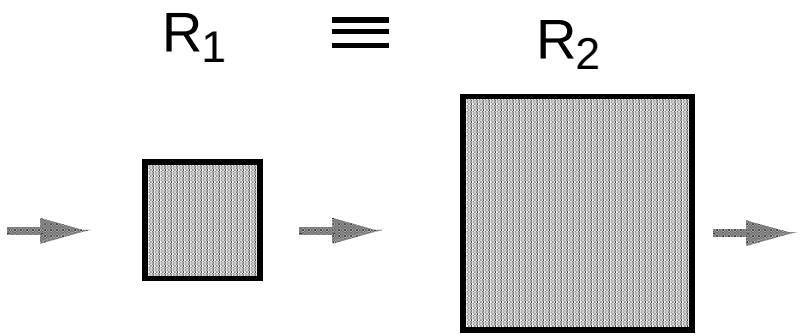
Wire Resistance



$$R = \rho \frac{L}{HW}$$

Sheet Resistance

R_o



Interconnect Resistance

Material	ρ (Ω-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}

Dealing with Resistance

- **Selective Technology Scaling**
- **Use Better Interconnect Materials**
  - reduce average wire-length
  - e.g. copper, silicides
- **More Interconnect Layers**
  - reduce average wire-length

Sheet Resistance

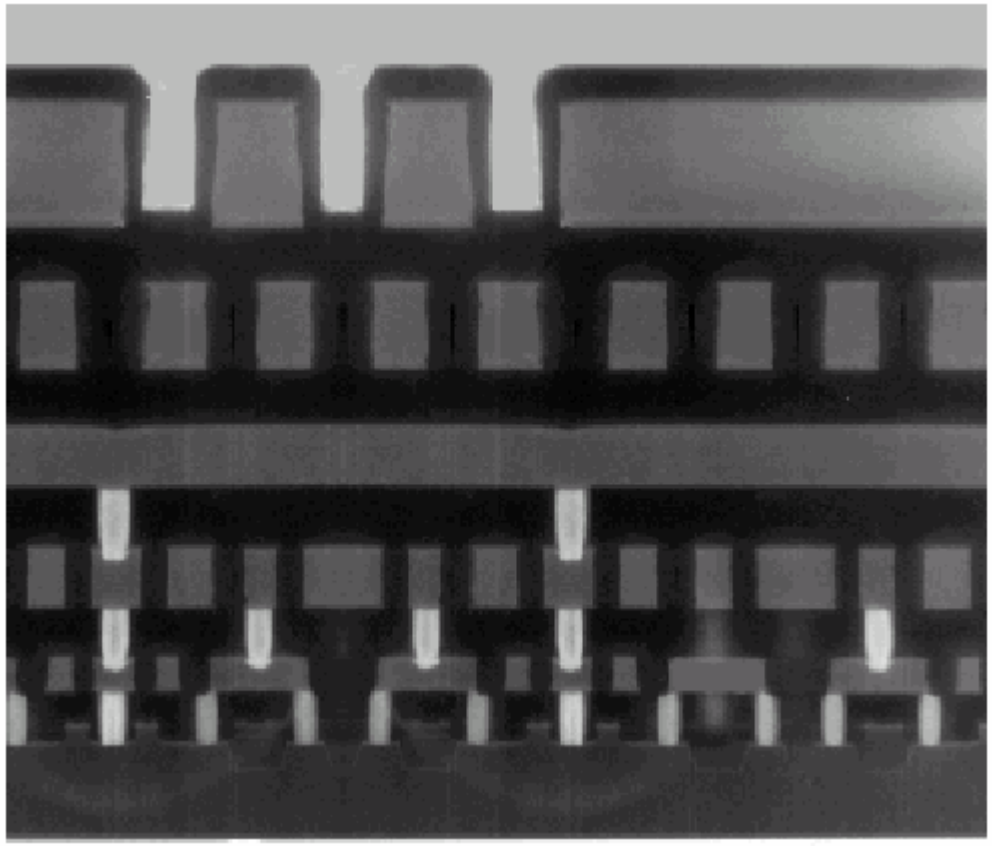
Material	Sheet Resistance (Ω/❑)
n- or p-well diffusion	1000 – 1500
n^+, p^+ diffusion	50 – 150
n^+, p^+ diffusion with silicide	3 – 5
n^+, p^+ polysilicon	150 – 200
n^+, p^+ polysilicon with silicide	4 – 5
Aluminum	0.05 – 0.1

Example: Intel 0.25 micron Process

5 metal layers
Ti/Al - Cu/Ti/TiN
Polysilicon dielectric

LAYER	PITCH	THICK	A.R.
Isolation	0.67	0.40	-
Polysilicon	0.64	0.25	-
Metal 1	0.64	0.48	1.5
Metal 2	0.93	0.90	1.9
Metal 3	0.93	0.90	1.9
Metal 4	1.60	1.33	1.7
Metal 5	2.56	1.90	1.5
	µm	µm	

Layer pitch, thickness and aspect ratio =*h/w*



Interconnect (RC) Delay

On-Chip VLSI interconnects can be modeled as RC elements

- *R is the wire resistance =*

$$\rho \frac{L}{A} = \rho \frac{L}{w \cdot h}$$

ρ is the resistivity of the metal

L is the wire length

A is the cross sectional area = wh (w is the width and h is the height of the wire)

- *C is the wire capacitance. For a parallel plate capacitor*

$$C = \varepsilon_d \frac{A}{d}, \ \varepsilon_d = k \, \varepsilon_0$$

A is the area of the plate

d is the distance between the plates

k is the dielectric constant of the insulating material between the plates

ε_0 is the permittivity of free space

If we can reduce both R and C, we can reduce wire delay…..

Will better materials like copper and low-k dielectrics solve the interconnect problem?

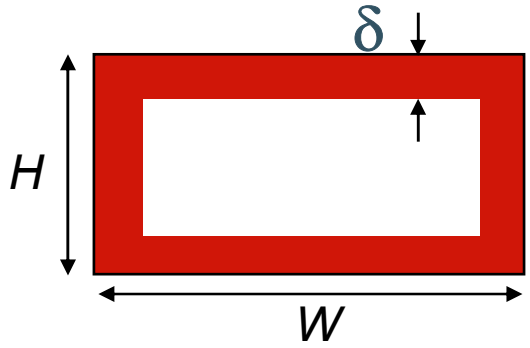
Cu has lower resistivity than Al, and is more robust (reliable) than Al…….

Changing Interconnect Materials

- Replace Al wires by Cu wires
- Resistivity of Al = 2.65 $\mu\Omega$-cm
  @ Room Temp. (20-25 ^{0}C)
- Resistivity of Cu = 1.67 $\mu\Omega$-cm
  @ Room Temp. (20-25 ^{0}C)
- Reduction in R is not even a factor of 2…..

Frequency Dependence of R

- At high enough frequencies, R can increase due to ***skin effect***
- Current flow constrained in a thin layer along the surface of the conductor: **R increases**
- The current falls off exponentially with depth into the interconnect
- **Skin depth**, δ is defined as the depth at which current falls of to a value of 1/e of its nominal value



$$\delta = \sqrt{\frac{\rho}{\pi f \mu}}$$

μ is the permeability of the surrounding dielectric

Resistance per unit length:

$$r(f) = \frac{\sqrt{\pi f \mu \rho}}{2(H+W)}, \quad for \ W, H >> \delta$$

Skin Effect

- Onset of skin effect: at $f=f_s$, δ =1/2 x smallest dimension

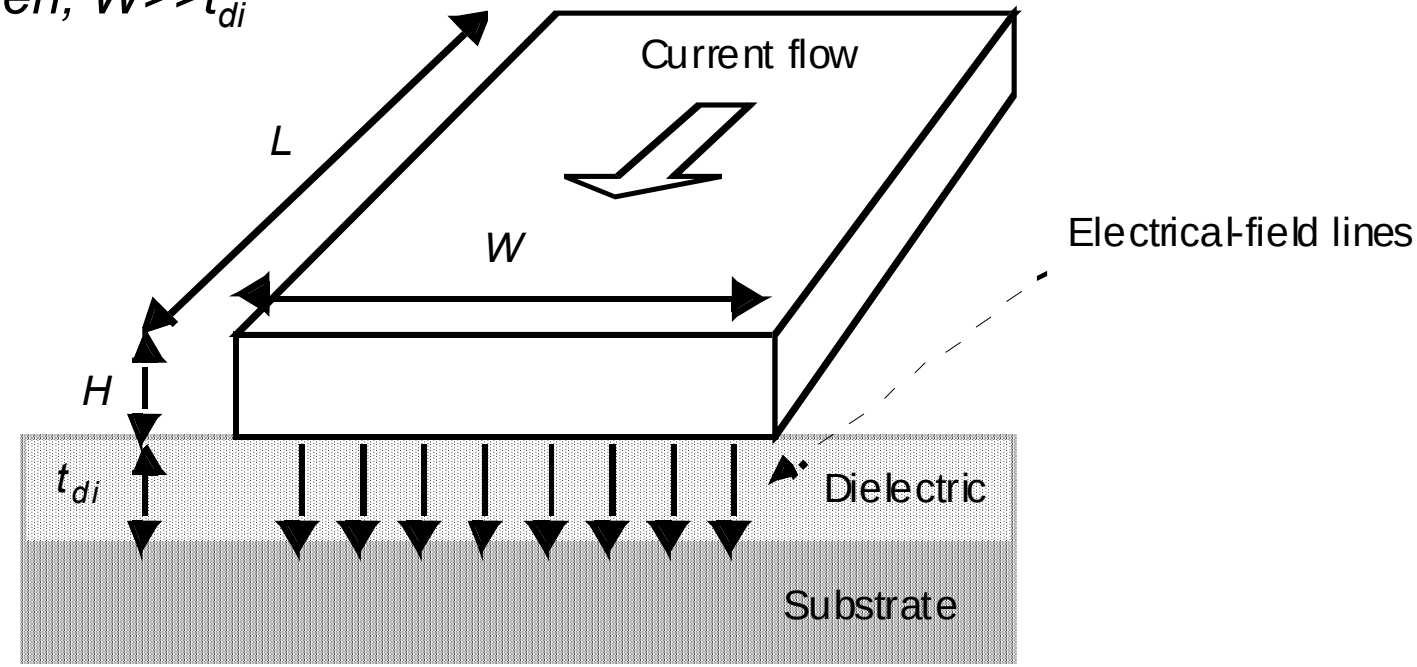
$$f_s = \frac{4\rho}{\pi\mu\left(\min\left(W,H\right)\right)^2}$$

- For Cu wires:

Freq	1 GHZ	5 GHZ	10 GHZ	15 GHZ	20 GHZ
Skin depth (μm)	2.08	0.93	0.66	0.53	0.46

Capacitance: The Parallel Plate Model

Valid when, $W >> t_{di}$

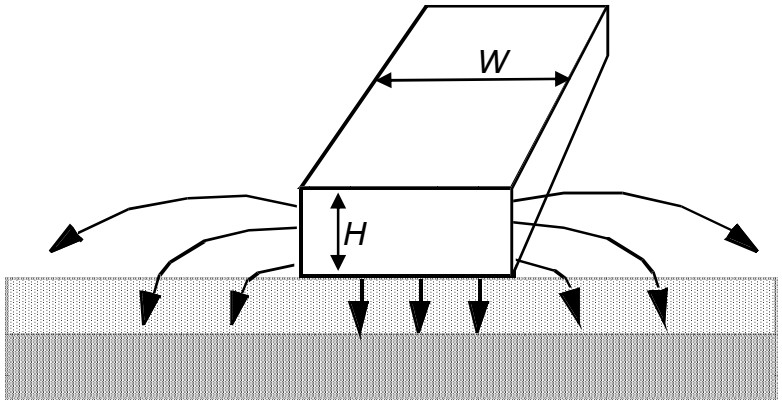


$$c_{int} = \frac{\varepsilon_{di}}{t_{di}} WL$$

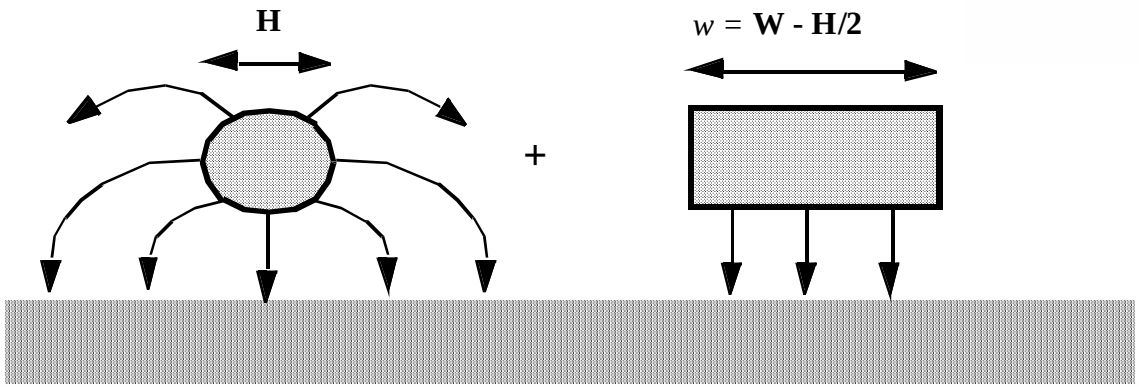
Permittivity

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

Fringing Capacitance Model

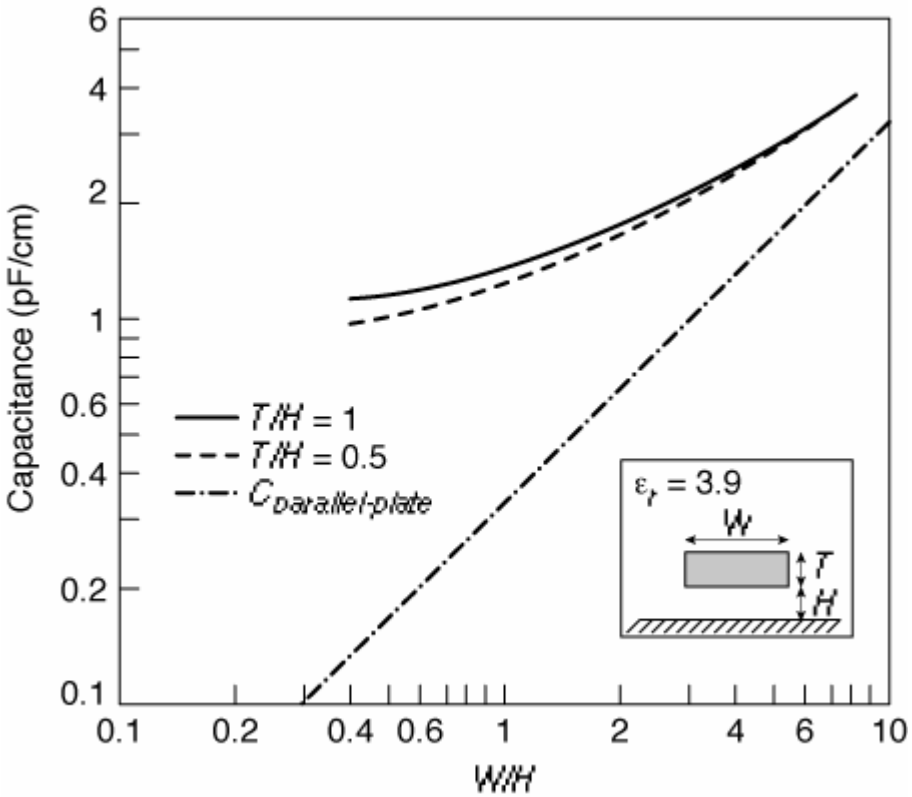


$$c_{wire} = c_{pp} + c_{fringe} = \frac{w\varepsilon_{di}}{t_{di}} + \frac{2\pi\varepsilon_{di}}{\log(t_{di}/H)}$$



Fringing Cap. ***Parallel plate Cap.***

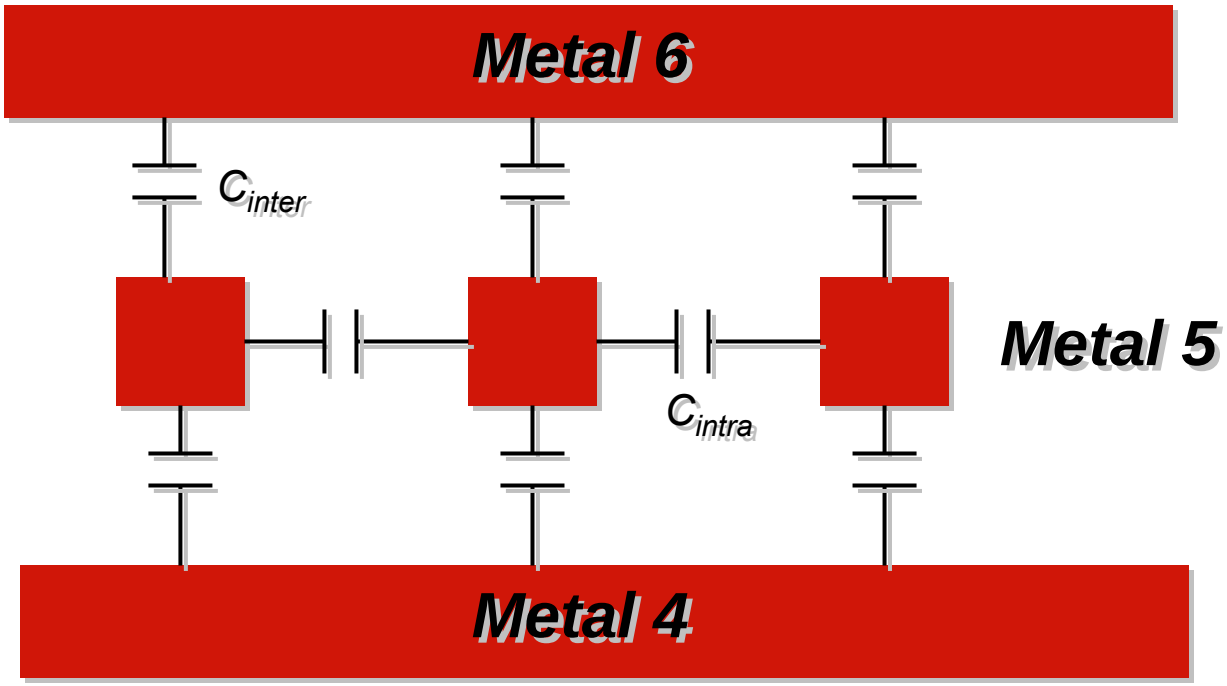
Fringing versus Parallel Plate



Saturates to ~1pF/cm due to fringing fields

(from [Bakoglu89])

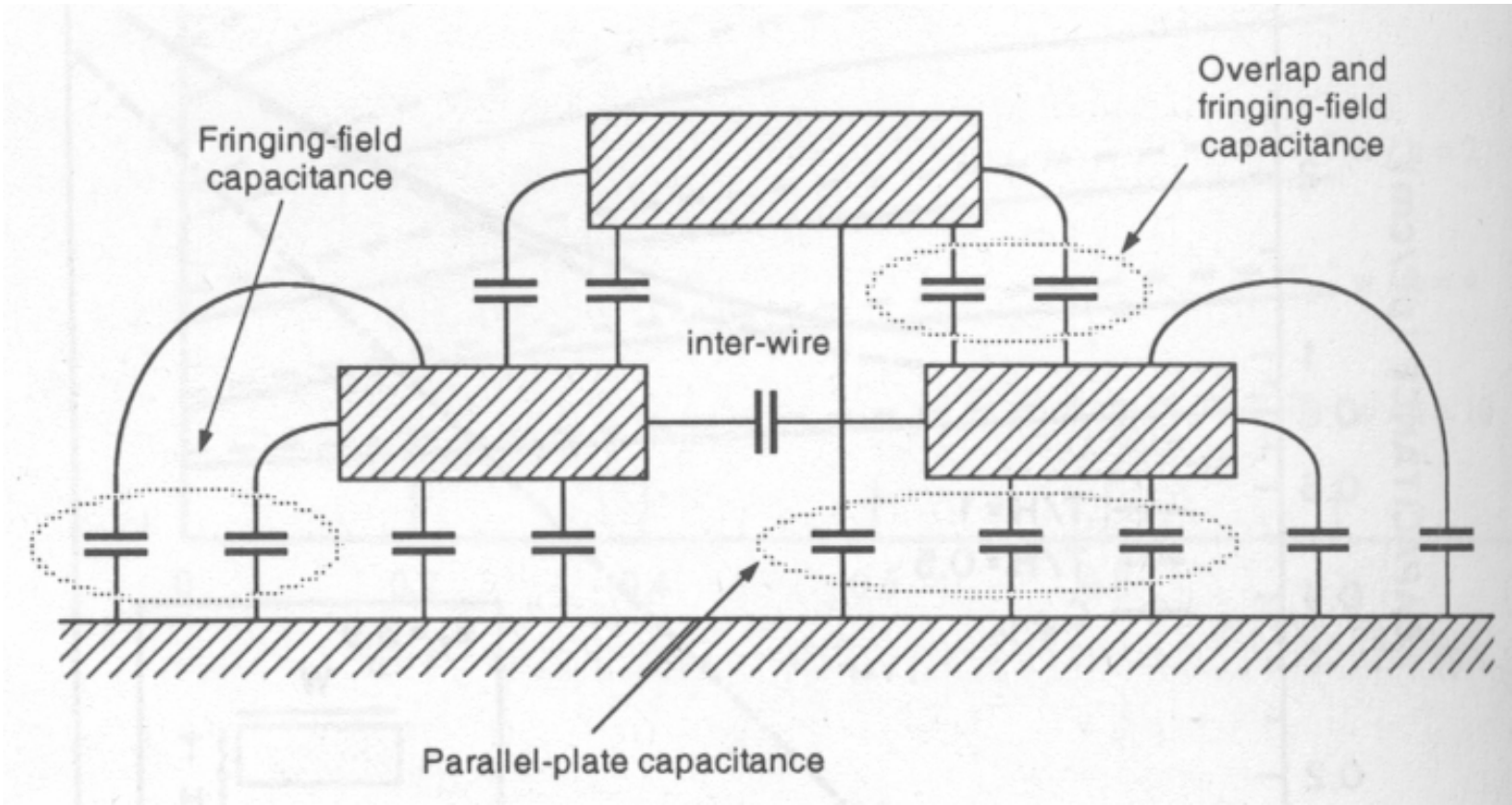
VLSI Interconnect Structure



Adjacent wires are orthogonal….why?

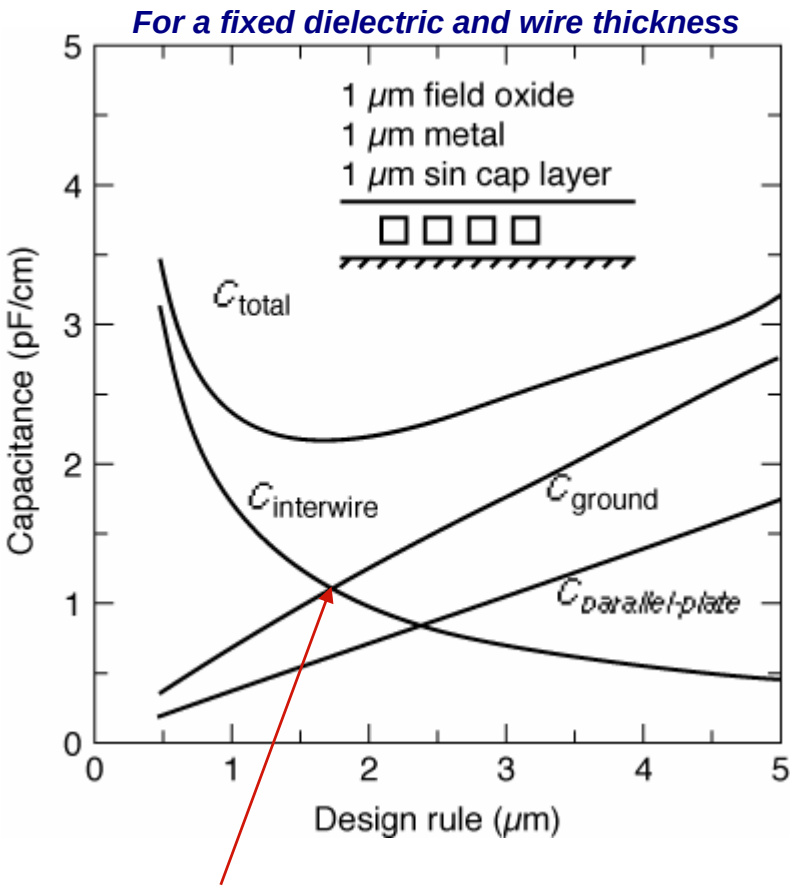
C_{intra} is the dominating capacitance……why?

Interconnect



- Usually interconnect layers alternate wire direction

Impact of Interwire Capacitance



- *Reduce wire width, w*
- *Reduce spacing (=w)*
- $C_{total} = C_{ground} + C_{interwire}$
- $C_{ground} = C_{p\text{-}p} + C_{fringing}$

Interwire cap. becomes important in a multi-level structure

(from [Bakoglu89])

Wiring Capacitances (0.25 μm CMOS)

Area cap (aF/um^2)

Fringing cap (aF/um)

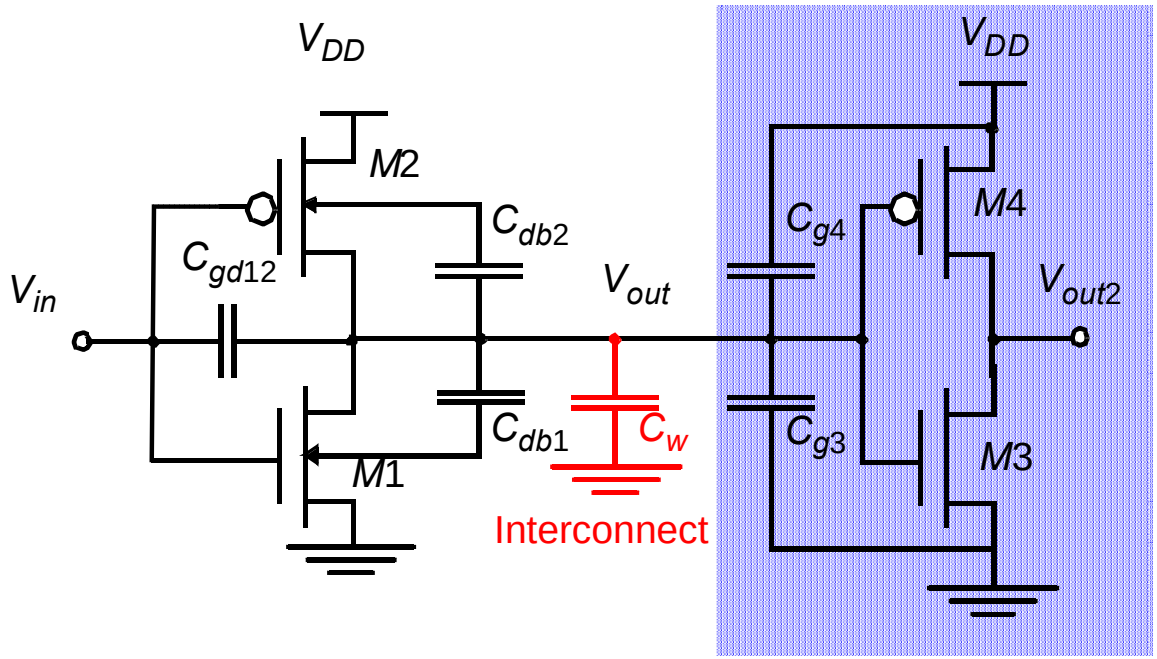
	Field	Active	Poly	Al1	Al2	Al3	Al4
Poly	88						
	54						
Al1	30	41	57				
	40	47	54				
Al2	13	15	17	36			
	25	27	29	45			
Al3	8.9	9.4	10	15	41		
	18	19	20	27	49		
Al4	6.5	6.8	7	8.9	15	35	
	14	15	15	18	27	45	
Al5	5.2	5.4	5.4	6.6	9.1	14	38
	12	12	12	14	19	27	52

Table rows represent the top plate of the capacitor and the columns represent the bottom plate

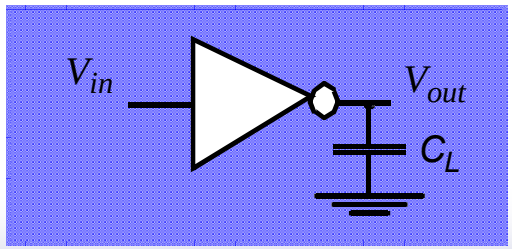
The wire at level 5 is twice as thick and is embedded in a dielectric of higher permittivity

Interconnect Modeling

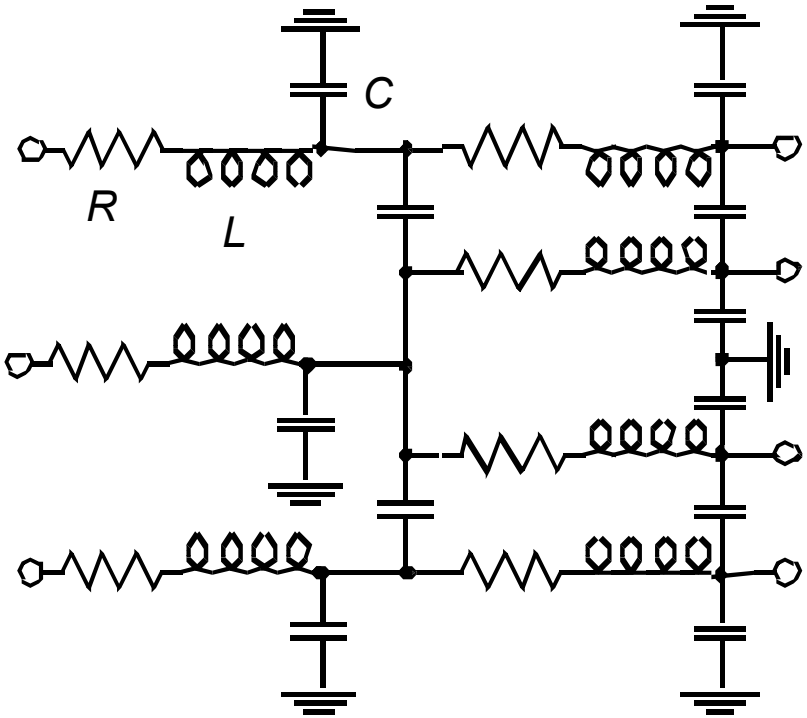
Capacitances of Driver-Wire-Load



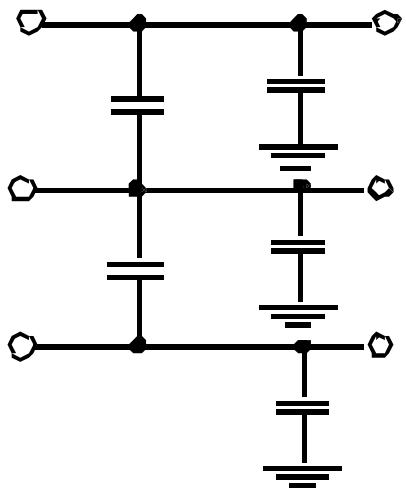
Simplified Model



Wire Models



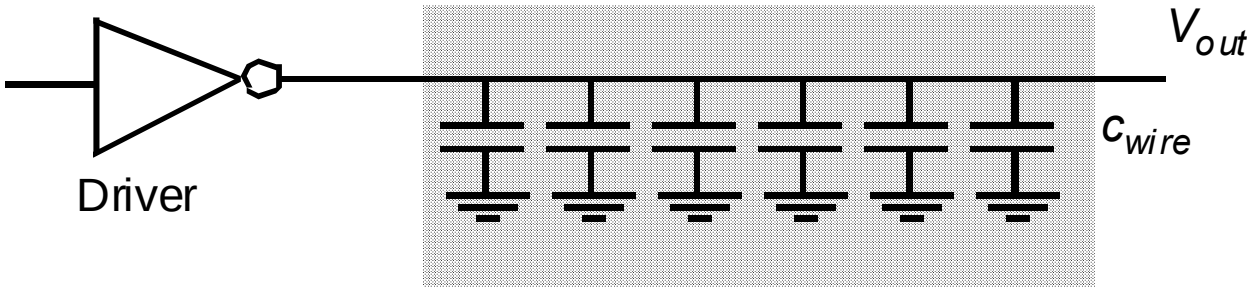
All-inclusive model



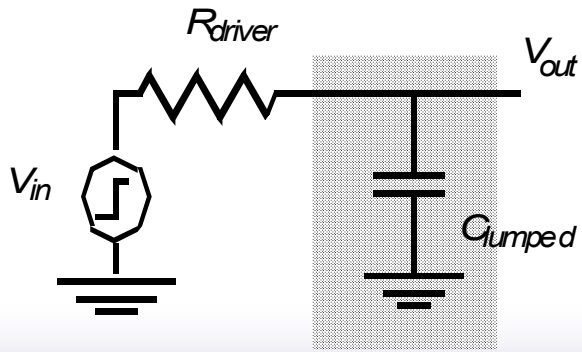
Capacitance-only

The Lumped Capacitor Model

Assuming no voltage drops along the wire….
i.e., wire is equipotential



If r is small and switching frequencies are not too high…



Advantage: Operation can be described by ordinary differential equation

The Lumped Capacitor Model

Operation of a simple RC circuit:

$$C_{lumped}\frac{dV_{out}}{dt}+\frac{V_{out}-V_{in}}{R_{driver}}=0$$

If a step input is applied from 0 to V:

$$V_{out}(t)=\left(1-e^{-t/\tau}\right)V$$

where $\tau = R_{driver}C_{lumped}$: *time-constant of network*

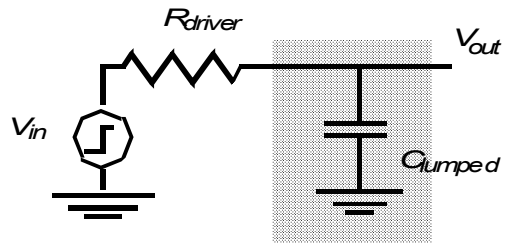
The time to reach 50% point can be calculated as:

$$t=\ln(2)\tau=0.69\tau$$

If $R_{driver}=10K\Omega$ *and* $C_{lumped}=11\,pF$:

$t = 76\,ns$ (*a very large number for high-performance digital circuits*)

Driver modeled as a voltage source and a source resistance (R_{driver}), $C_{lumped}=L\times c_{wire}$



Only impact on performance is due to the loading effect of the capacitor on the driving gate

The Lumped RC-Model

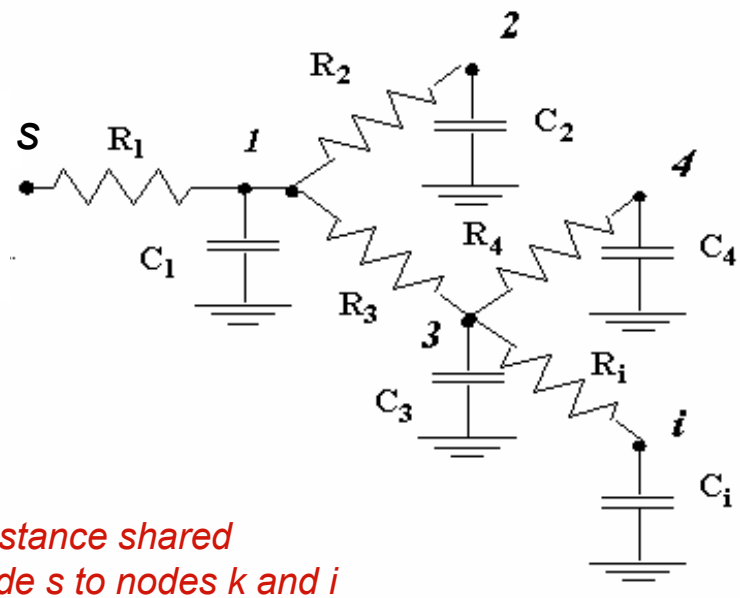
- On-chip metal wires at the semi-global and global tiers can be several millimeters long and can have significant resistance
- The equipotential assumption is no longer valid
- Need an RC model
- A simple approach: lump wire resistance of each segment into a single R and the global capacitance into a single C
  - This is called a lumped RC model….pessimistic and inaccurate for long wires
  - Should use a distributed RC model for such long wires

Then what is the use of the lumped RC model?

The Lumped RC-Model

- The distributed RC model is complex: involves partial differential equations, closed form solution does not exist
- However, the behavior of a distributed RC line can be adequately modeled by a simple RC network
- It is more convenient to reduce any driver-wire-load networks to an equivalent RC network and predict its first order response
- However, unlike the single R-C network analyzed earlier that had a single time constant (network pole), deriving the correct waveforms for a complex network becomes intractable
    - Need to solve a set of ordinary differential equations with many time constants (poles and zeroes)….
    - Or, run time-consuming SPICE simulations for the entire network….

Elmore Delay Comes to the Rescue…..



RC Tree: *unique resistive paths between s and any node i*

Total resistance along this path is called the ***path resistance*** R_{ii}

e.g., path resistance between s and node 4: $R_{44} = R_1 + R_3 + R_4$

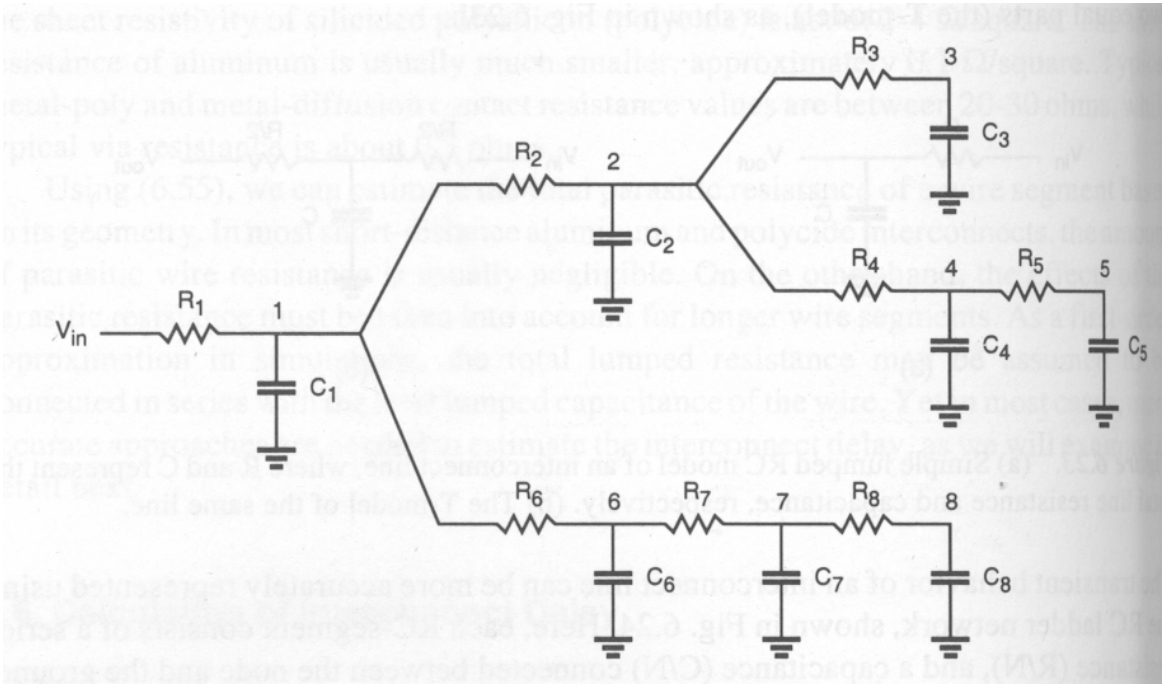
Shared path resistance: *resistance shared among the paths from root node s to nodes k and i*

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

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

$R_{i2}=R_1$

Elmore Delay-RC Chain (Branched)



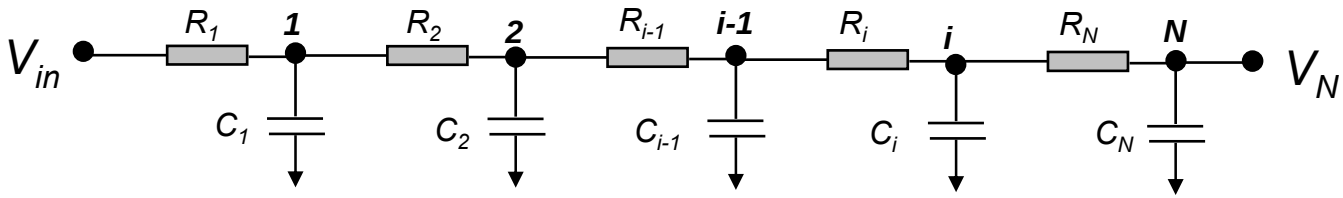
If a step input Vin is applied at t=0

Elmore Delay at node i: $\tau_{Di} = \sum_{k=1}^{N} C_k R_{ik}$

(First-order time constant of the network)

N = # of capacitances in the network

Wire Model



For a non-branched RC chain: $\tau_{DN} = \sum_{i=1}^{N} C_i R_{ii}$ $t_{Di} = C_1R_1 + C_2 (R_1+R_2)+\ldots\ldots+ C_i (R_1+R_2+\ldots R_i)$

If wire modeled by N equal-length segments ($R_{seg} = rL/N$, $C_{seg} = cL/N$)

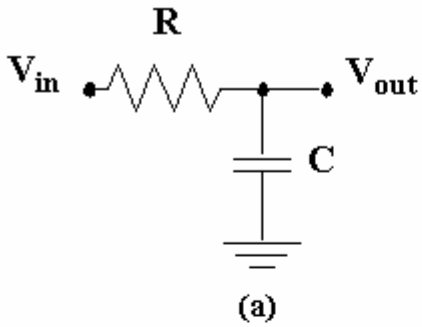
$$\tau_{DN} = \left(\frac{L}{N}\right)^2 (rc + 2rc + \ldots + Nrc) = (rcL^2)\frac{N(N+1)}{2N^2} = RC\frac{N+1}{2N}$$

For large values of N:

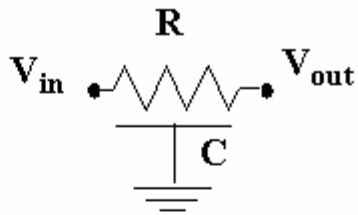
$$\tau_{DN} = \frac{RC}{2} = \frac{rcL^2}{2}$$

The Distributed RC-line

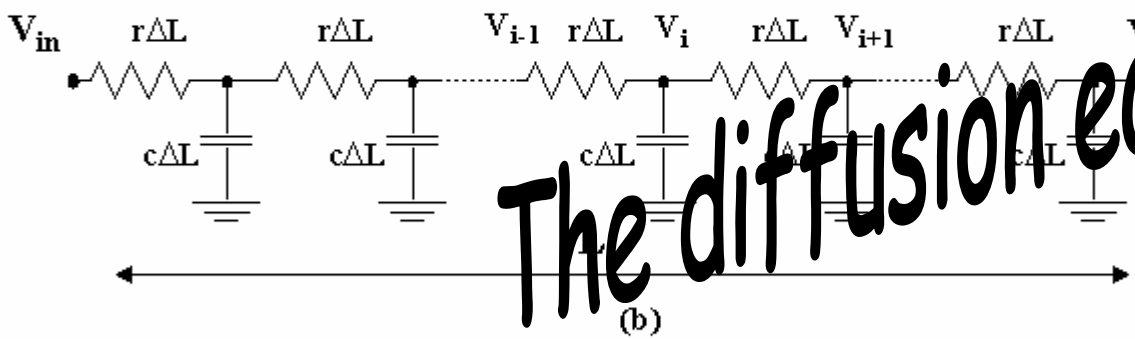
(can be approximated by the lumped RC network model)



(a)



(c)



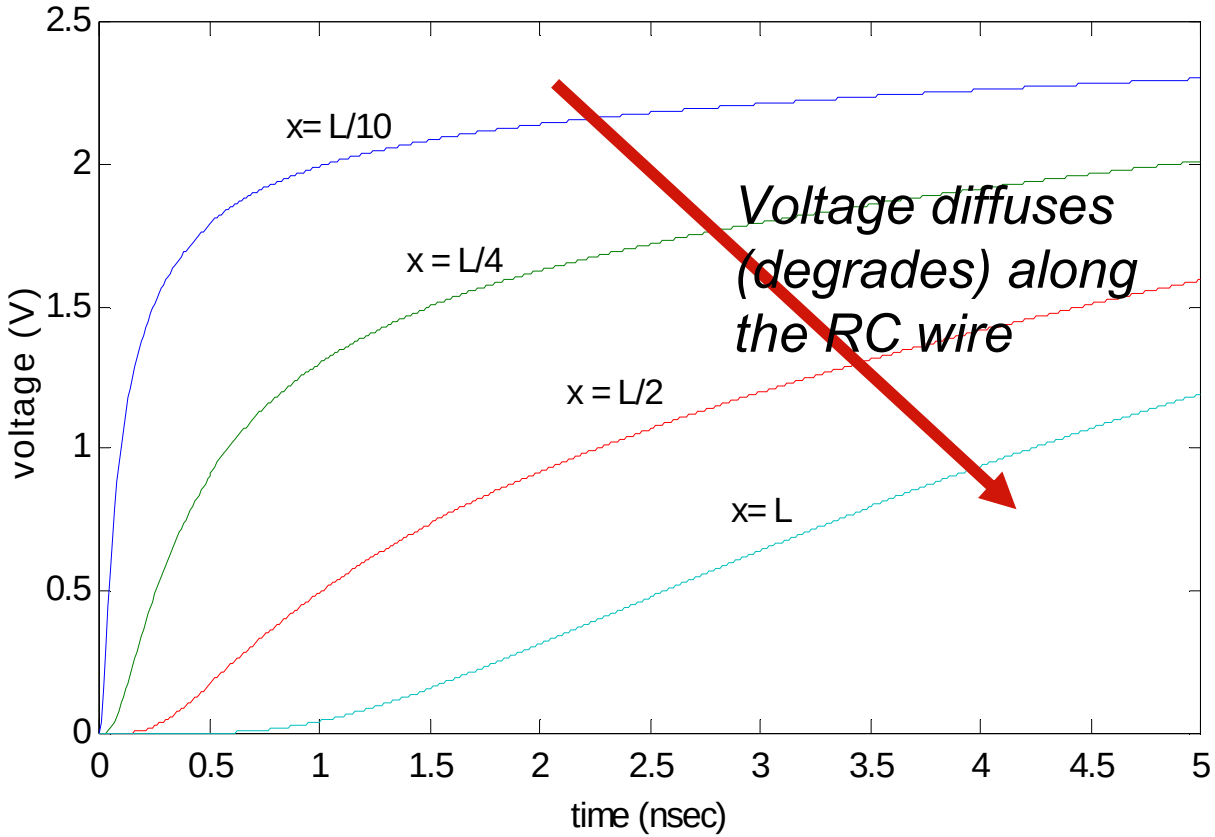
(b)

The diffusion equation

$$rc\frac{\partial V}{\partial t} = \frac{\partial^2 V}{\partial x^2}$$

$$\tau(V_{out}) = \frac{rc\,L^2}{2}$$

Step-response of RC wire as a function of time and space



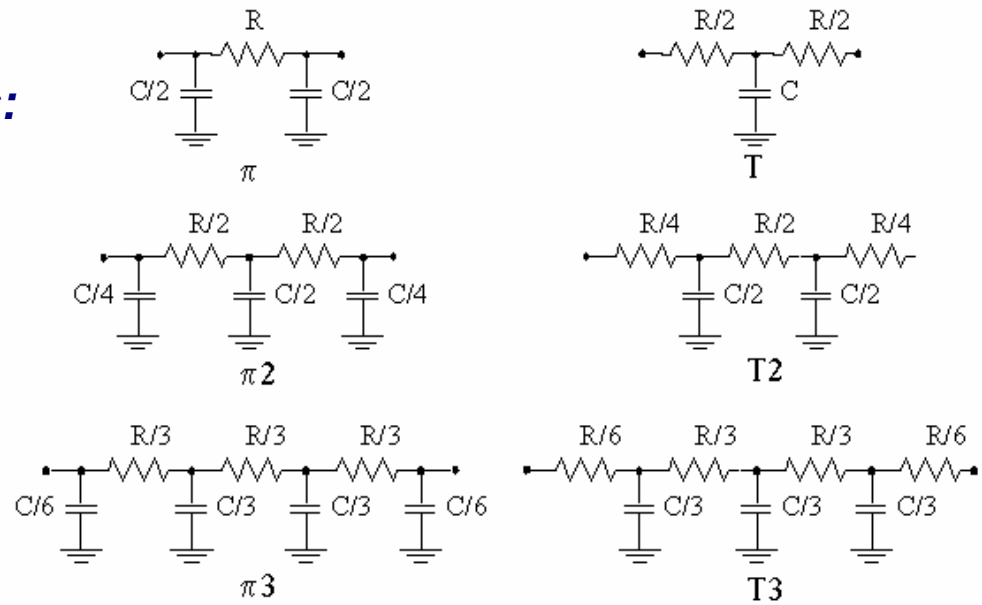
RC-Models

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.

SPICE Wire (RC) Models:

accuracy of the model determined by number of stages



Design Rules of Thumb

- rc delays should only be considered when $t_{pRC} >> t_{pgate}$ of the driving gate

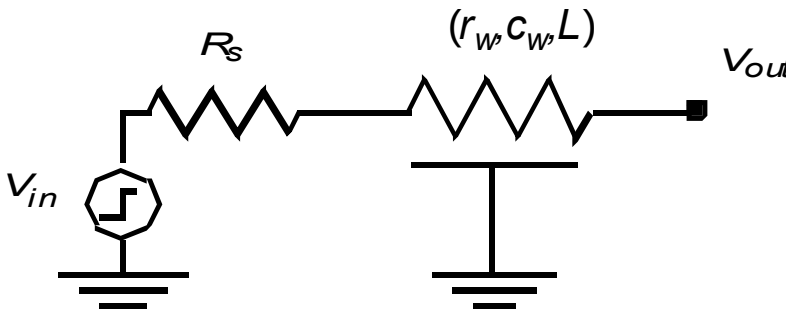
$$Lcrit >> \sqrt{t_{pgate}/0.38rc}$$

- rc delays should only be considered when the rise (fall) time at the line input is smaller than RC, the rise (fall) time of the line

$$t_{rise} < RC$$

  - when not met, the change in the signal is slower than the propagation delay of the wire

Illustration of Rule: Driving an RC-line



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

$R_w = r_w L$

$C_w = c_w L$

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

Delay due to wire resistance becomes important when $(R_w C_w/2) \geq R_s C_w$ or when $L \geq 2R_s/r_w$

On-Chip Inductance

- L starts to play a role at high GHz frequencies
- Effect increases for Cu: lower resistivity
- Can increase interconnect delay
- Can cause overshoot and undershoot effects
- Can cause switching noise: Ldi/dt voltage drops

$$Z = R + j\omega L$$

inductance of a wire surrounded by a uniform dielectric:

$$c\,\ell = \varepsilon\,\mu$$

c = cap. per unit length

l = inductance per unit length

From Maxwell's Laws:

$$v = \frac{1}{\sqrt{\ell\, c}} = \frac{1}{\sqrt{\varepsilon\,\mu}} = \frac{c_0}{\sqrt{\varepsilon_r\,\mu_r}}$$

v = propagation speed of EM wave

c_0 = speed of light in vacuum = 30 cm/ns