ECE 225
High Speed Digital IC Design
Lecture 5

Interconnect Technology and Scaling Issues-I

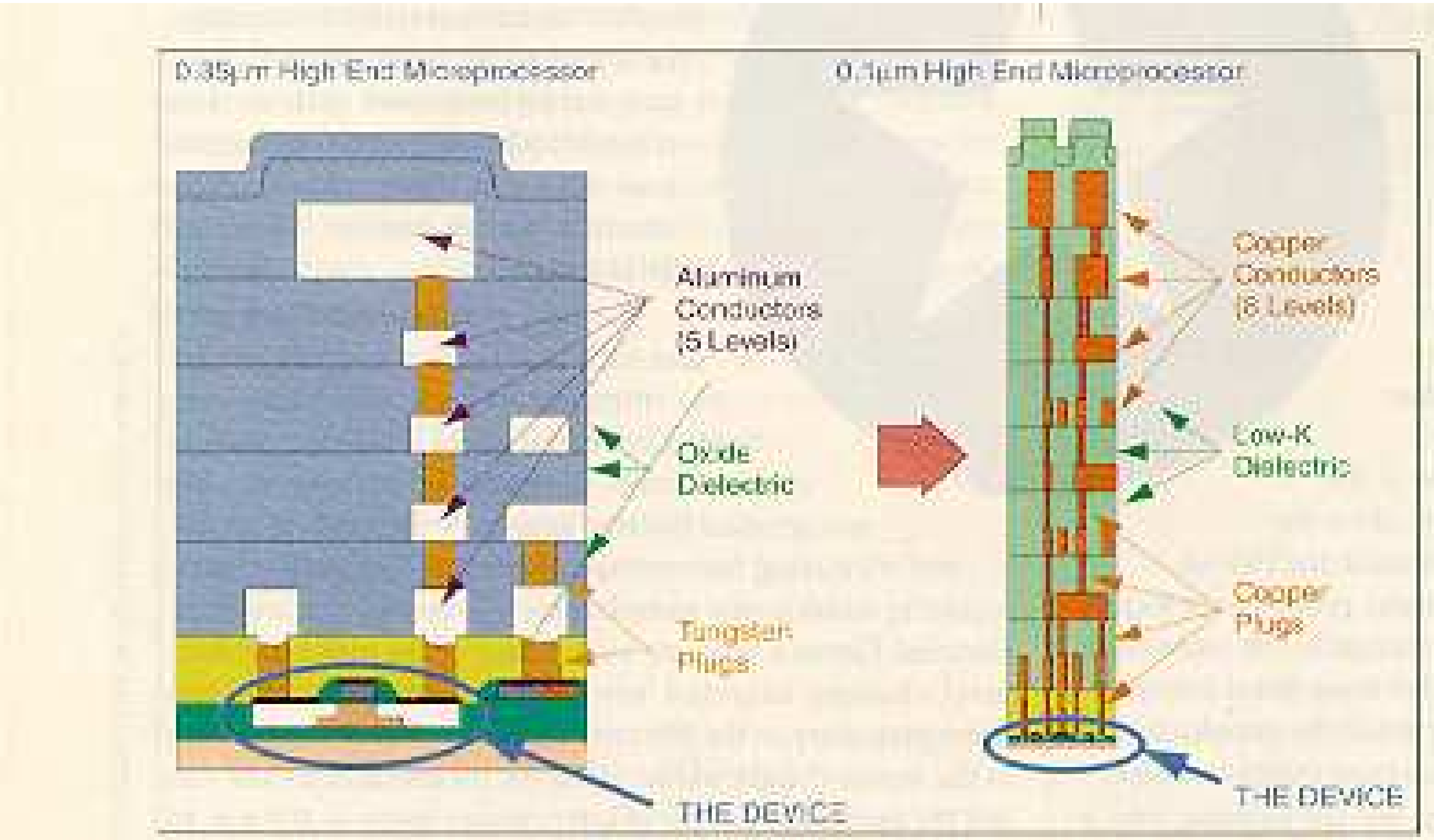
Prof. Kaustav Banerjee
Electrical and Computer Engineering
E-mail: kaustav@ece.ucsb.edu

Outline

- Interconnect and Technology Scaling
  - ➔ Interconnect technology: background and trends
  - Interconnect parasitics: R-L-C
  - Elmore delay formalism
  - Interconnect scaling: implications for technology and design

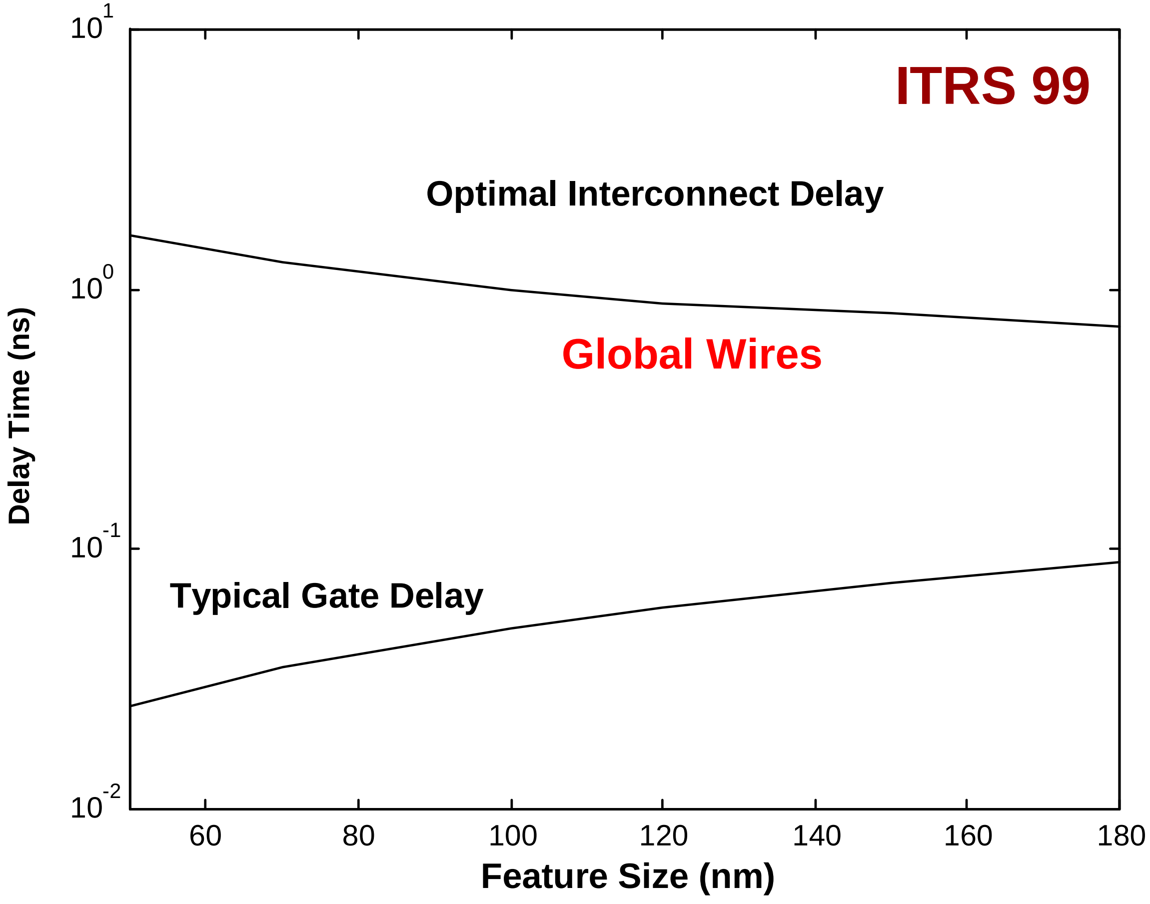
Interconnect Technology: Past and Present

A brief review of the past interconnect technologies…..(Al, AlCu, W-vias, Al-vias, silicides, doped poly etc)



1. Process architecture challenge.

Wire Vs Gate Delay…..



Interconnect delay has become the dominant factor determining chip performance…

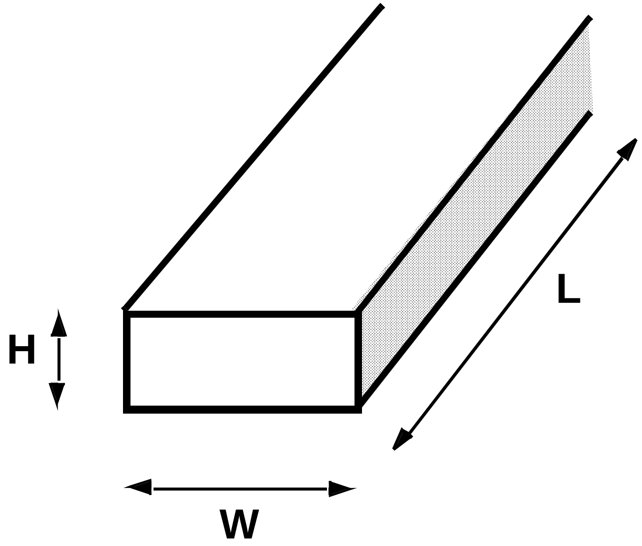
K. Banerjee et al., Proc. IEEE, May 2001.

Interconnect Parasitics

- Interconnect parasitics
  - affect performance and power consumption
  - affect reliability

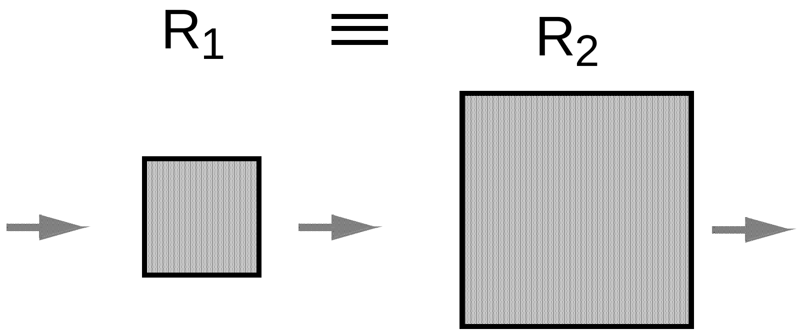
- Classes of parasitics
  - Capacitive
  - Resistive
  - Inductive

Wire Resistance



$$R = \frac{\rho \, L}{H \, W}$$

Sheet Resistance R_o

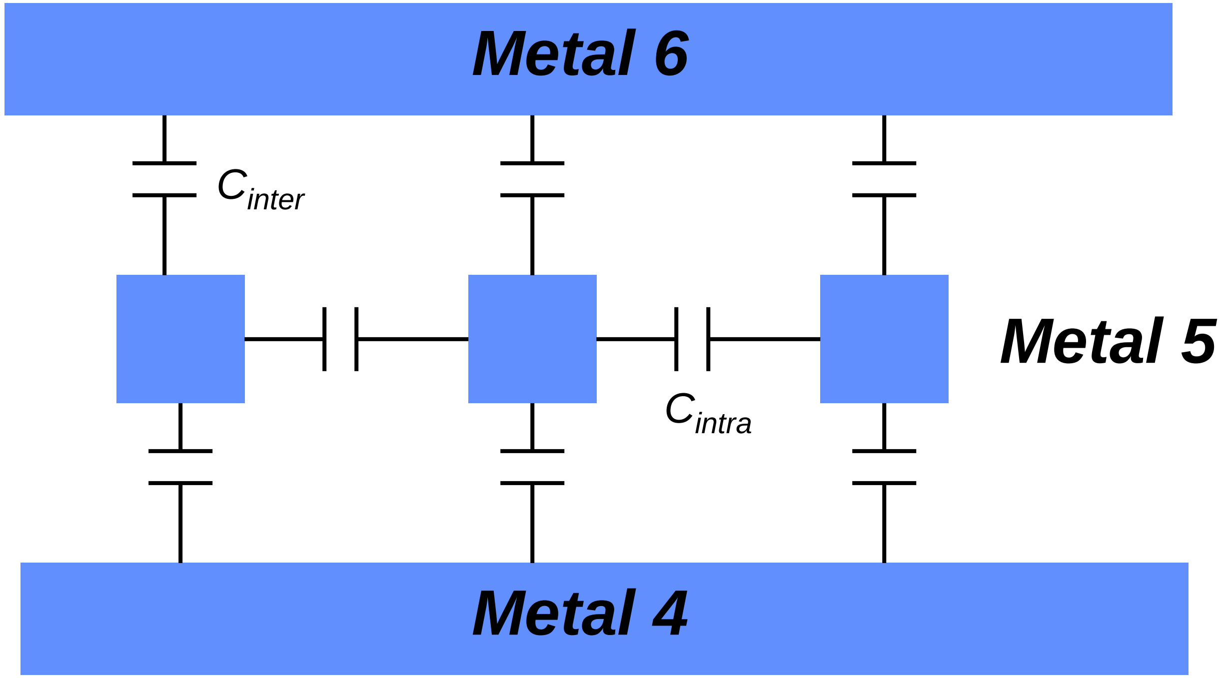


By adjusting L/W ratios…..

Interconnect Material Resistivity

Material	ρ (Ω-m) at room temp.
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}

VLSI Interconnect Structure

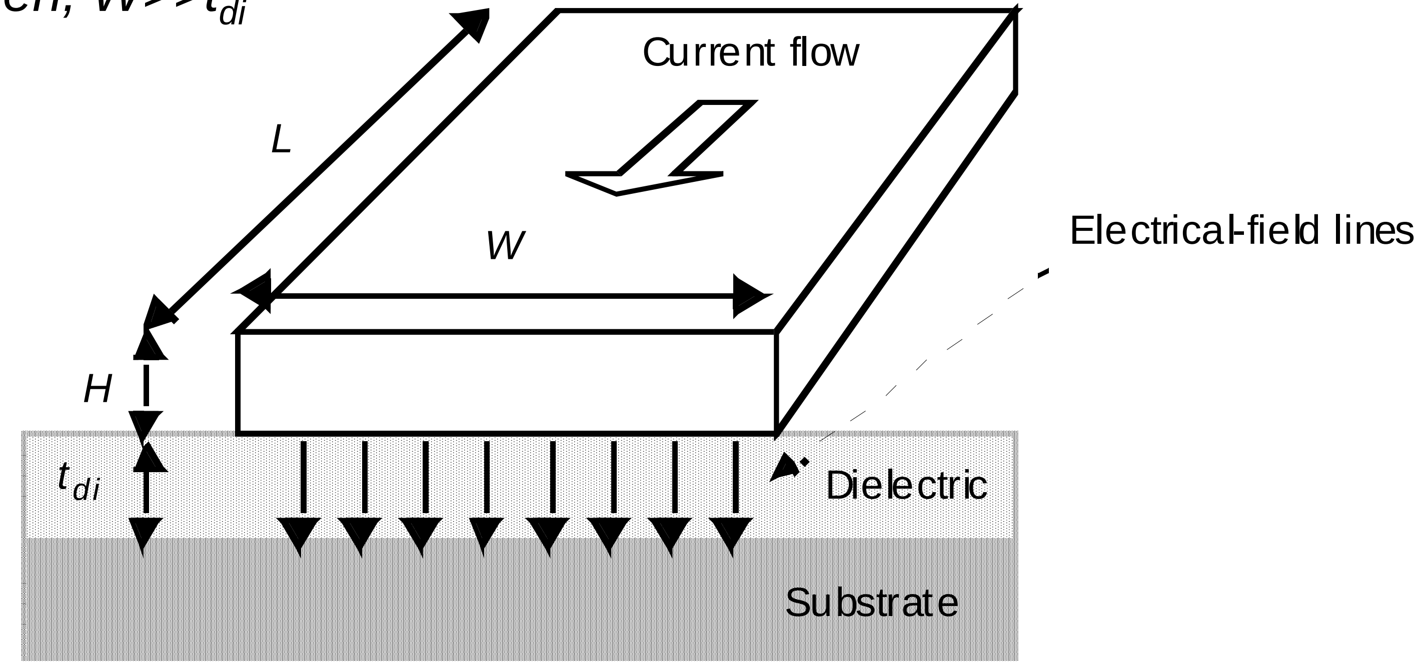


- Wires on adjacent metal layers are orthogonal
- C_{intra} is the dominating capacitance

Wire Capacitance: The Parallel Plate Model

1-dimensional model

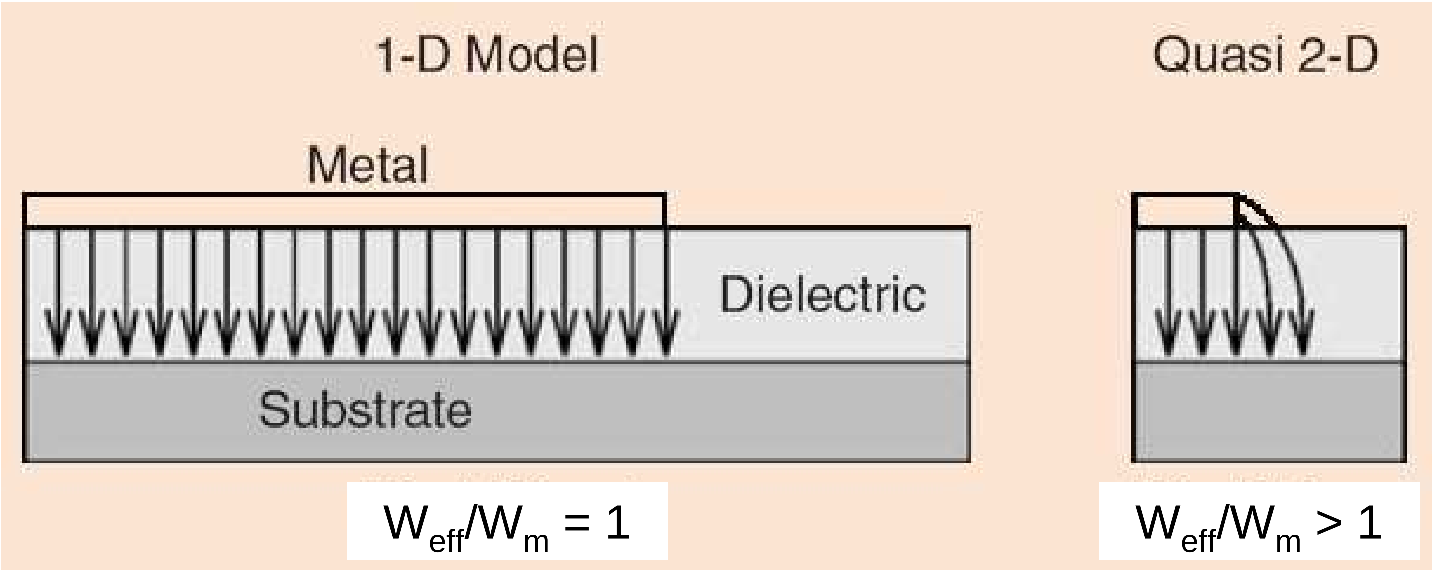
Valid when, $W \gg t_{di}$



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

Simplest structure; known closed form expression

Wire Capacitance: Fringing Fields

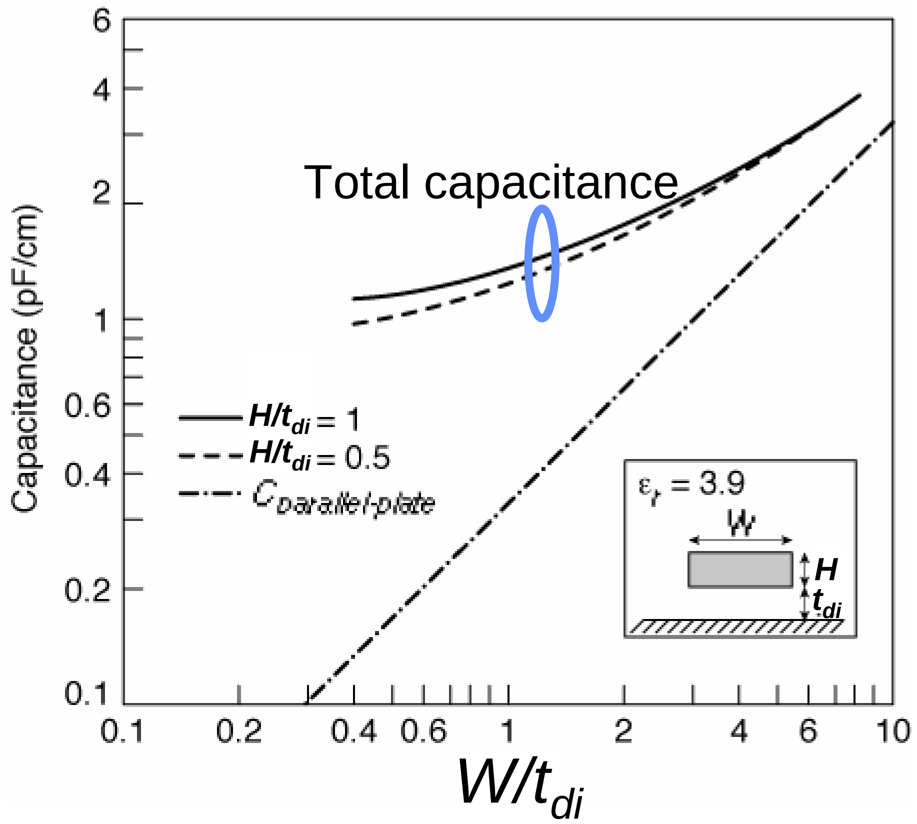


Fringing fields from side walls

Increasingly important as aspect ratio *H/W > 1*

No closed form expressions for capacitance

Wire Capacitance: Fringing Vs Parallel Plate

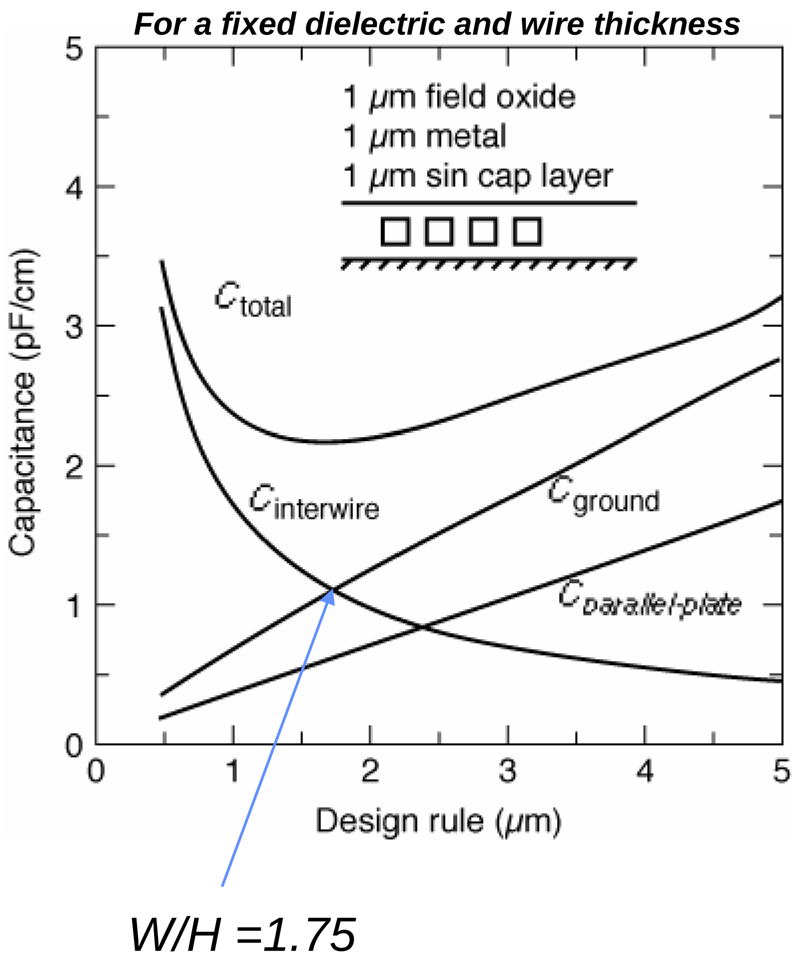


Saturates to ~1pF/cm due to fringing fields

Note: aspect ratio = H/W

(from [Bakoglu89])

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. is important as H>W for most wires.

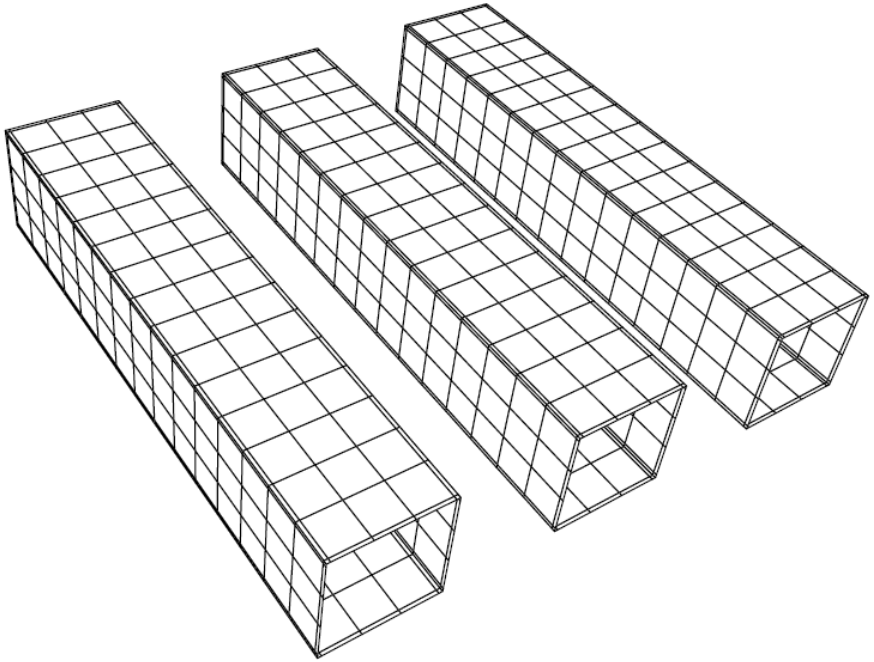
Interwire cap. main source of crosstalk noise (signal traveling on one wire causes voltage fluctuations on a neighboring wire)

(from [Bakoglu89])

Accurate Interconnect Capacitance Extraction

- Requires full 3-D electromagnetic simulation
- FASTCAP (from MIT) is a widely used tool
  - See K. Nabors and J. White, "FastCap: a multipole accelerated 3-D capacitance extraction program," IEEE Trans. Computer-Aided Des. Integr. Circuits Syst., vol. 10, no. 11, pp. 1447–1459, Nov. 1991.
- Can be time consuming for complex interconnect topologies
- FASTCAP is available for download at:
  - http://www.rle.mit.edu/cpg/research_codes.htm

Accurate Capacitance Extraction: Field Solvers



CAPACITANCE MATRIX, picofarads

		1	2	3
GROUP1	1	206.2	-95.93	-18.59
GROUP2	2	-95.93	253.3	-95.98
GROUP3	3	-18.59	-95.98	206.2

- Discretize each conductor surface into panels
- Assign excitation voltage to one conductor at a time
- Compute charges on all conductor panels
- Charge on excited conductor gives self capacitance
- Charge on other conductors gives mutual capacitance

Permittivity of Materials

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

RC Delay: Design Rules of Thumb

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

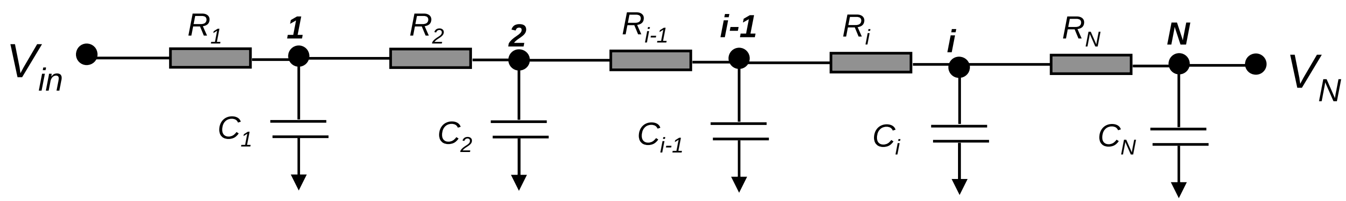
$$L_{crit} >> \sqrt{t_{pgate}/0.38rc}$$

- 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

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 ($R_{seg}= rL/N$, $C_{seg}= cL/N$)

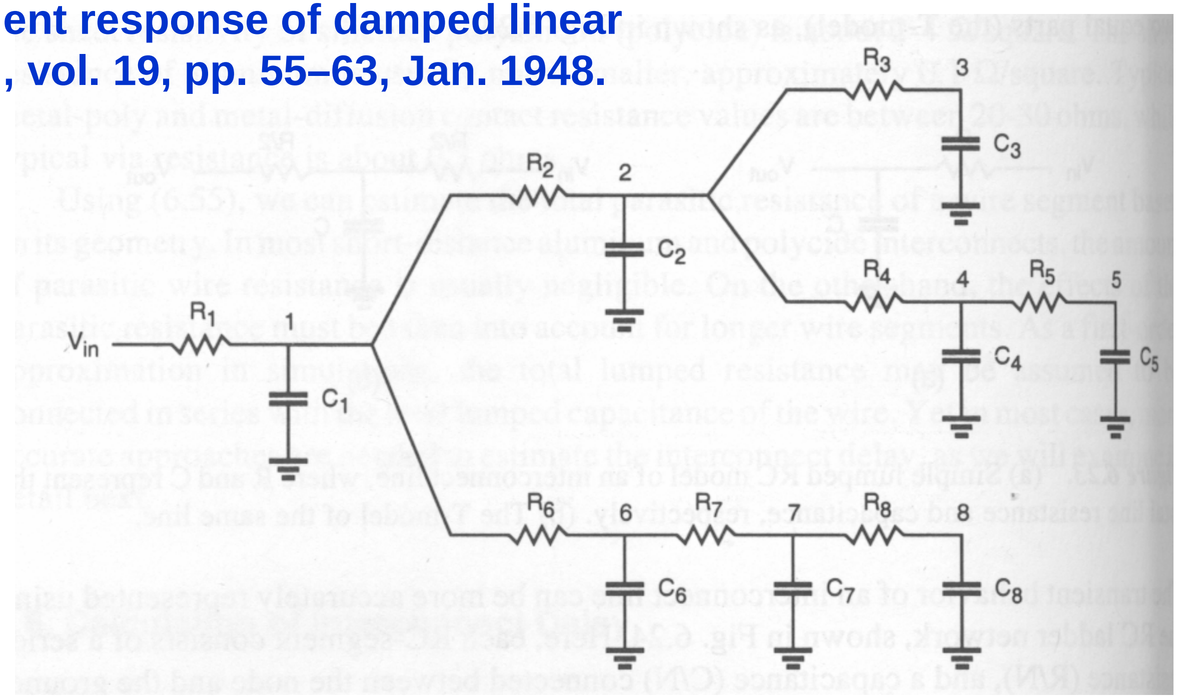
$$\tau_{DN} = \left(\frac{L}{N}\right)^2 (rc + 2rc + \ldots + Nrc) = (rcL^2)\frac{N(N+1)}{2N^2} = RC\frac{N+1}{2N}$$

For large values of N:

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

Elmore Delay-RC Chain (Branched)

W. C. Elmore, "The transient response of damped linear networks," *J. Appl. Phys.*, vol. 19, pp. 55–63, Jan. 1948.



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

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

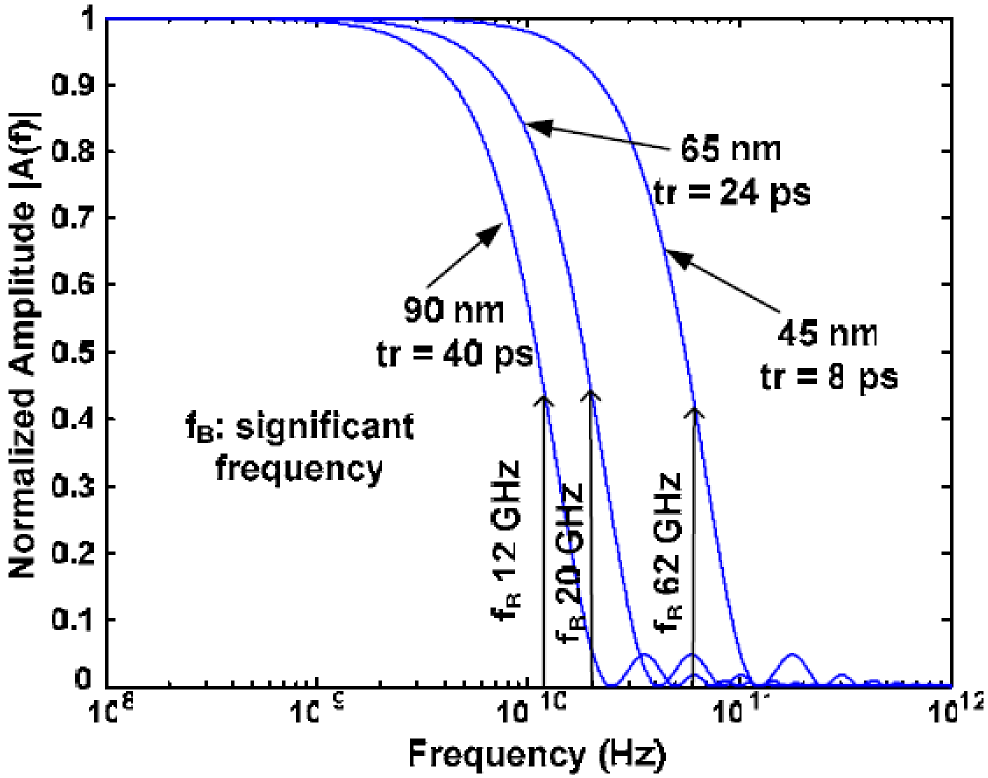
From Maxwell's Laws:

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

v = propagation speed of EM wave

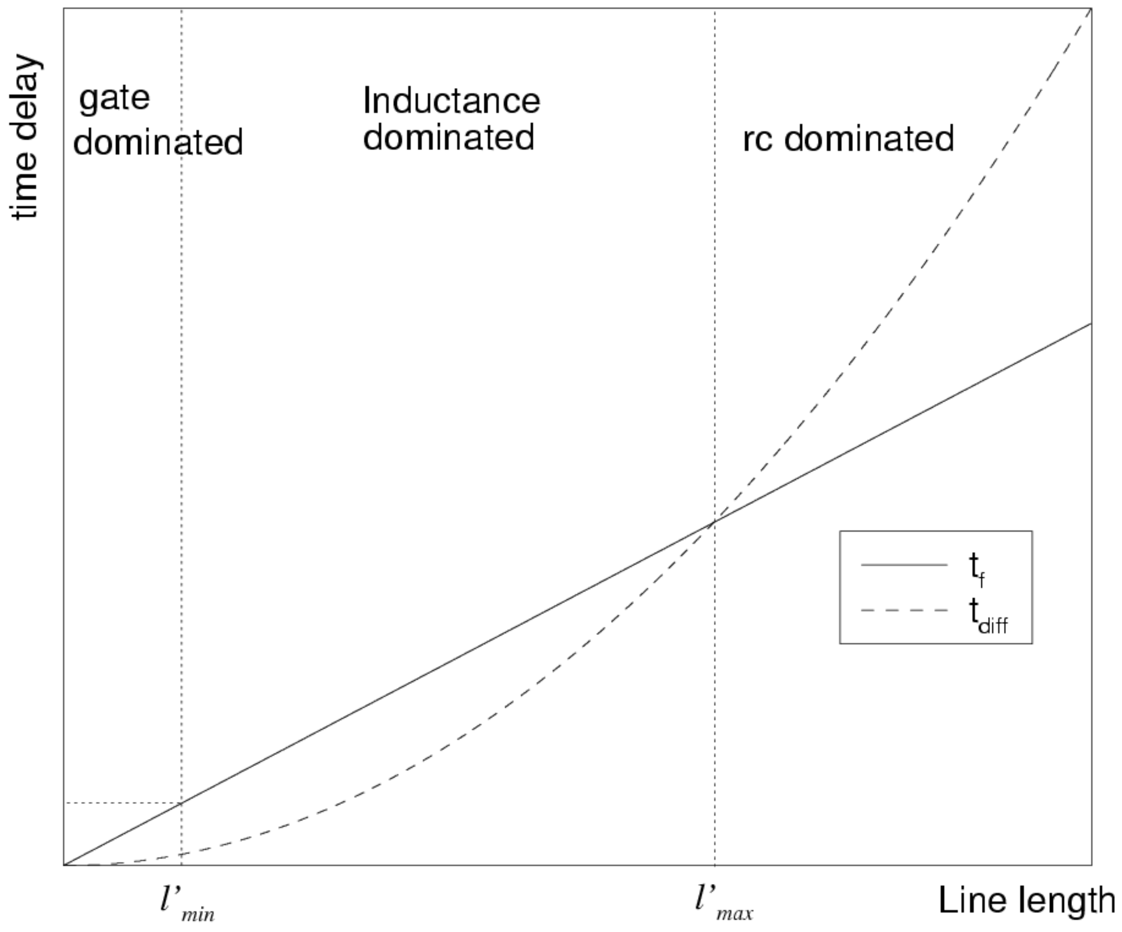
c_0 = speed of light in vacuum = 30 cm/ns

Signal Frequency on Interconnects



- Transient signal composed of a wide range of frequencies
- Related to rise time of input signal
- Much higher than clock frequency (today 1-5 GHz)

Wire Length Dependency of L Effects



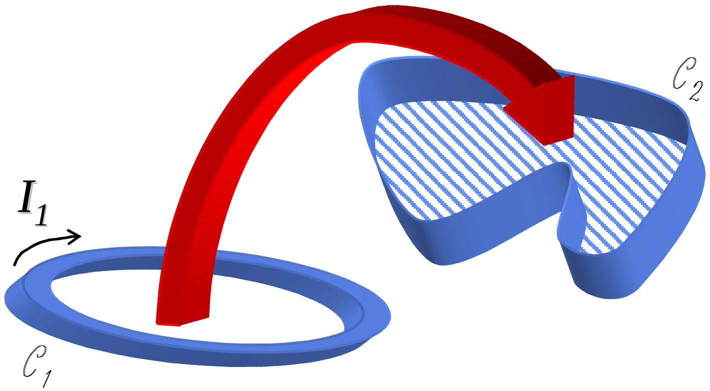
$$l'_{min} = \frac{T_{rise} v}{2}$$

$$l'_{max} = \frac{2Z_0}{r}$$

R. Suaya et al., Advanced Metallization Conf. 2006

Definition of Inductance (L)

L_{ij} = geometrical constant of proportionality between the magnetic flux through victim loop *j* and the current running on aggressor loop *i* causing it



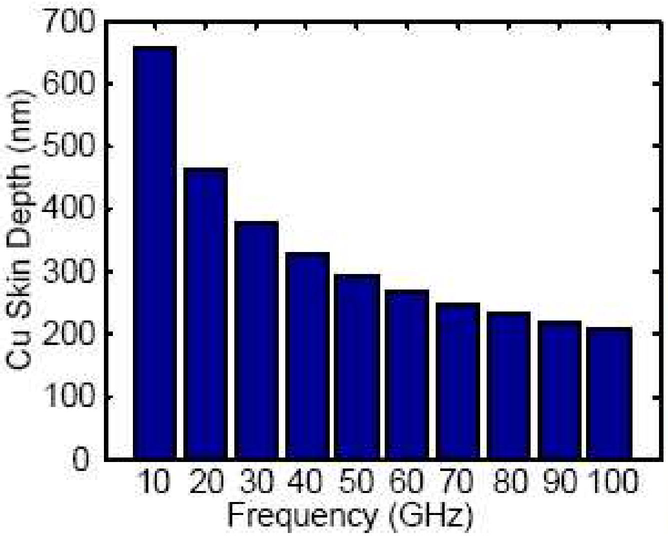
$$L_{ij} = \frac{\Psi_{i \to j}}{I_i} = \frac{1}{I_i} \int_{S_j} \mathbf{B}^{(\mathbf{i})} \cdot \mathbf{dS_j} \Longrightarrow L_{ij} = \frac{\mu_0}{4\pi} \oint_{C_j} \oint_{C_i} \frac{\mathbf{d}\ell_j \cdot \mathbf{d}\ell_i}{|\mathbf{r_j} - \mathbf{r_i}|}$$

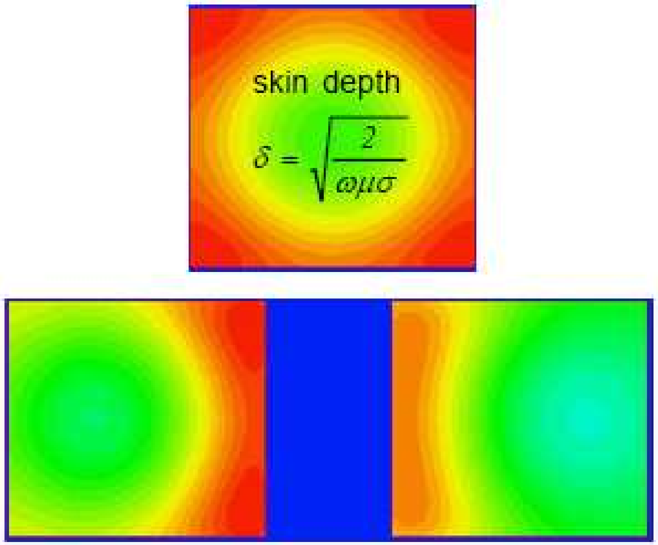
See A. E. Ruehli, "Inductance calculations in a complex integrated circuit environment," *IBM J. Res. Develop.*, pp. 470–481, Sept. 1972.

Interconnect Inductance Extraction Models

- Compact models are known for regular cross-section wires
- See: F. Grover, *Inductance Calculations: Working Formulas and Tables*, New York, Dover, 1962.
- These models assume uniform current distribution through wire
- Inductance is computed for current LOOPs – a wire carrying current and another carrying "return" current in opposite direction
- For on-chip wires, power/ground wires are present everywhere and carry the return current for signal lines

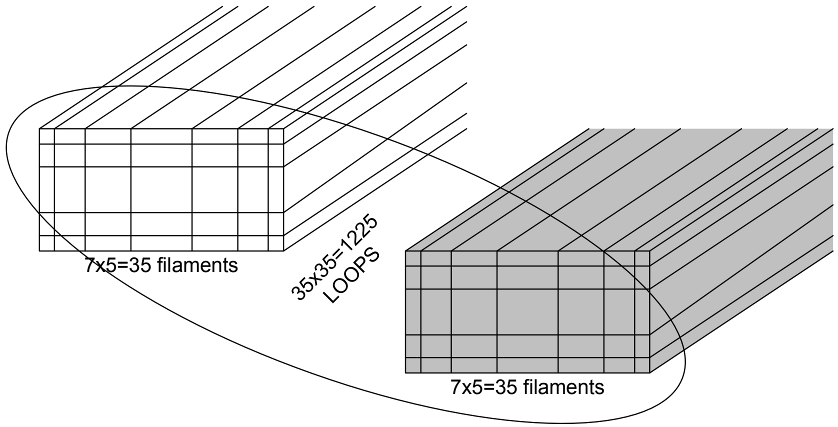
More Electromagnetic Effects at High Frequencies





Skin and proximity effects: non-uniform current distribution

Affects not only inductance but also resistance!



$$\begin{bmatrix} Z_{11}^{self} & Z_{12}^{mutual} & \ldots & Z_{1n}^{mutual} \\ Z_{21}^{mutual} & \ldots & \ldots & \ldots \\ \ldots & Z_{ij}^{mutual} & Z_{ii}^{self} & \ldots \\ Z_{n1}^{mutual} & \ldots & \ldots & Z_{nn}^{self} \end{bmatrix} \begin{bmatrix} I_1 \\ \ldots \\ I_i \\ I_n \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix};$$

Accurate Interconnect Inductance Extraction

- Requires full 3-D electromagnetic simulation
- FASTHENRY (from MIT) is a widely used tool
  - See M. Kamon, M. J. Tsuk, and J. K. White, “FASTHENRY:Amultipole-accelerated 3-D inductance extraction program,” IEEE Trans. Microwave Theory Techn., vol. 42, pp. 1750–1758, Sept. 1994.
  - Can be time consuming for complex interconnect topologies
  - Has limitations at very high frequencies…

RLC Delay: Design Rules of Thumb

Wires in RLC domain:

$$t_{50\%} = \sqrt{LC} + \frac{\tau_{FO4}}{2} + 0.7 Z_0 C_{load} \quad \text{if}$$

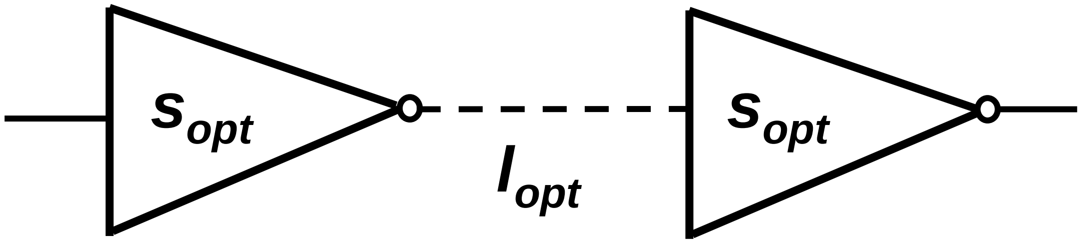
$$\frac{rL}{Z_0} < 2\ln\left(\frac{4Z_0}{R_{driver} + Z_0}\right), \quad R_{driver} < 3Z_0$$

$$\text{where} \quad Z_0 = \sqrt{l/c}$$

Also see: **A. Deutsch, et al., “When are transmission-line effects important for on-chip interconnections?,” IEEE Trans. Microwave Theory Tech., vol. 45, pp. 1836–1846, Oct. 1997.**

Repeater Insertion

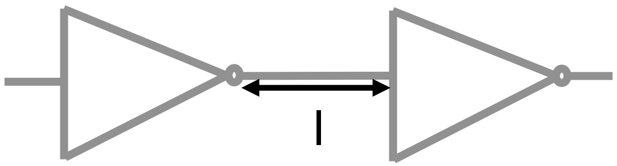
- RC-Delay of wire increases quadratically with its length
- Repeater (inverter) or buffer insertion makes the delay increase proportional to its length
- Repeater size and inter-repeater separation can be optimized to give minimal delay



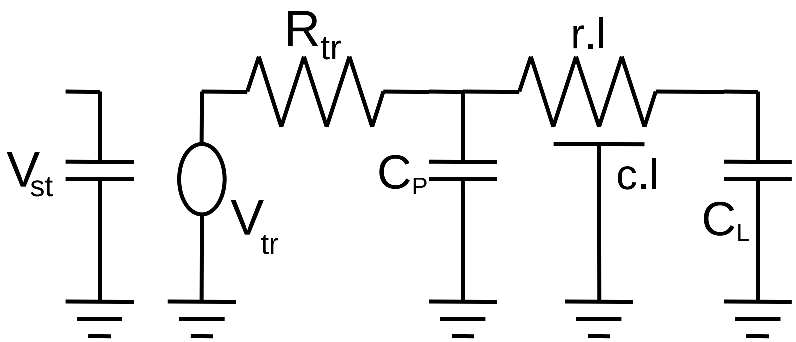
(from [Bakoglu89])

Driver-Interconnect-Load Equivalent Circuit

- Interconnect segment of length l between two inverters



Schematic



RC equivalent circuit

- V_{st} is the voltage at the input capacitance that controls the voltage source V_{tr}
- R_{tr} is the driver transistor resistance, C_p is output parasitic capacitance
- C_L is the load capacitance of next stage, *r* and *c* are interconnect resistance and capacitance per unit length respectively

Interconnect Scaling

Interconnect Scaling Scenarios

- Performance of a technology can be summarized by the time delay, t_d:

$$t_d = R_{tr}(C_P + C_L) + (R_{tr}c + rC_L)l + \frac{1}{2}rcl^2$$

intrinsic stage delay + load delay + interconnect delay

- If s is a scaling parameter: ratio of the feature size at a newer technology node to the feature size at an older node, then $s < 1$.
- Two simple scaling scenarios
  - scale metal pitch at constant thickness
  - scale metal pitch and the metal thickness

Interconnect Scaling

Implications of Technology Scaling on Technology Performance

- Scaling metal pitch at constant thickness
  - load delay, line delay, and current density will scale as $1/s$, $1/s^2$, and $1/s$, respectively
- Scaling metal pitch and the metal thickness
  - load delay, line delay, and current density will scale as $1/s^2$, $1/s^2$, and $1/s^2$, respectively
- Hence, interconnect delay begins to dominate technology performance and current density increases with scaling

Interconnect Technology Scaling Trends

- Interconnect scaling requirements drive several technology enhancements:
  - use of low-dielectric constant (low-k) materials lowers the interconnect capacitance per unit length (*c*) and therefore reduces the delay time
  - lower *c* helps in minimizing cross-talk noise
  - lower *c* also helps to reduce the dynamic power dissipation during switching of gates
  - lower interconnect resistance and higher current density requirements have driven the use of new metallization, such as Cu