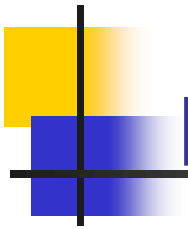
ECE 124a/256c
VLSI RC(L) Interconnect Models

Forrest Brewer

Wayne Burleson, Atul Maheshwari

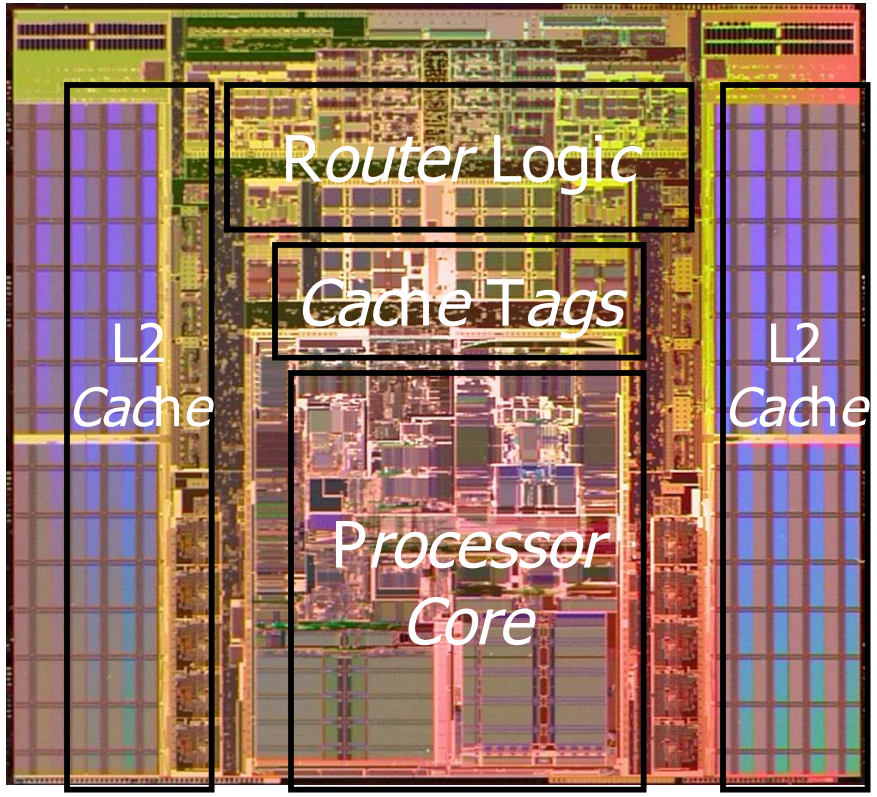


Readings

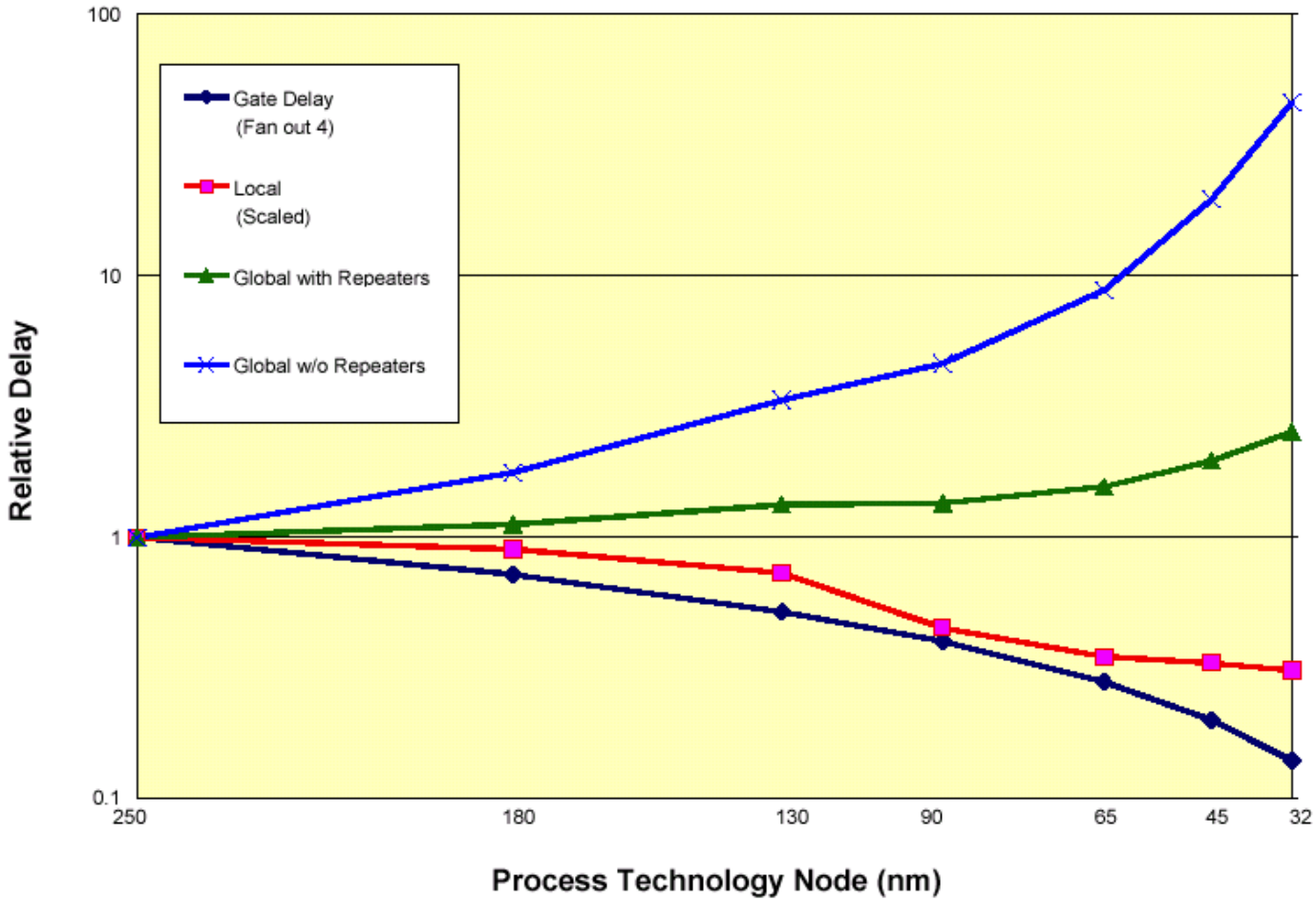
- H. B. Bakoglu, "*Circuits interconnects and packaging for VLSI*" , Addison Wesley
- W. J. Dally and J. W. Poulton, "*Digital Systems Engineering*" , Cambridge Press
- J. M. Rabaey, "*Digital Integrated circuits : A design perspective*" , Prentice Hall

Components of VLSI system

- Logic
  - Functional Block
  - Logic Gates
  - Transistors
- Interconnects
  - Power/ground and Clock
  - Inter-block Signals
  - Intra-block Signals



Delay with technology scaling



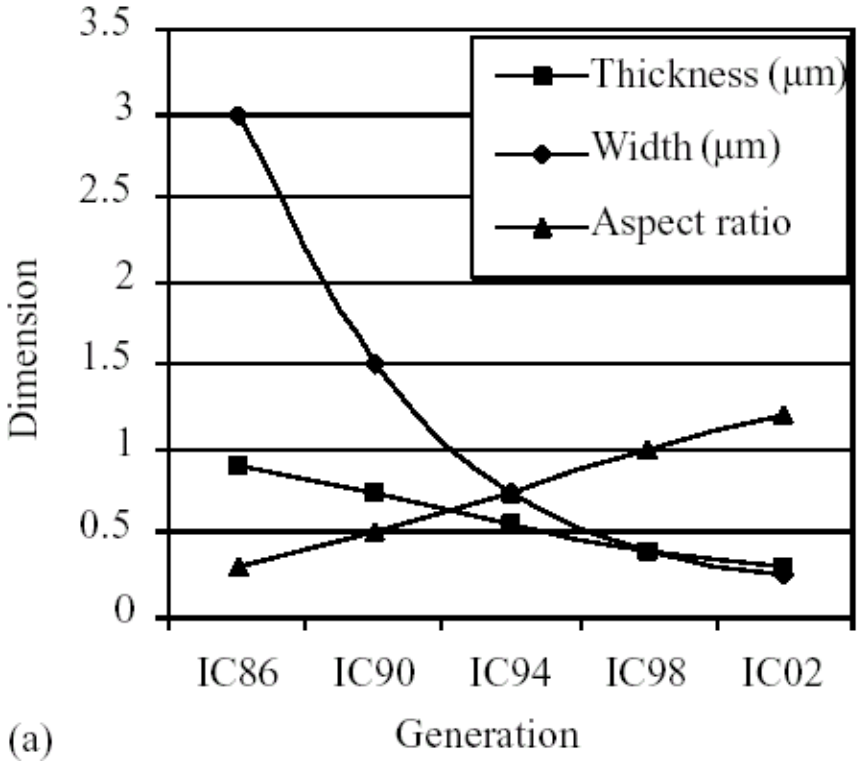
This figure is from the ITRS Roadmap on interconnects

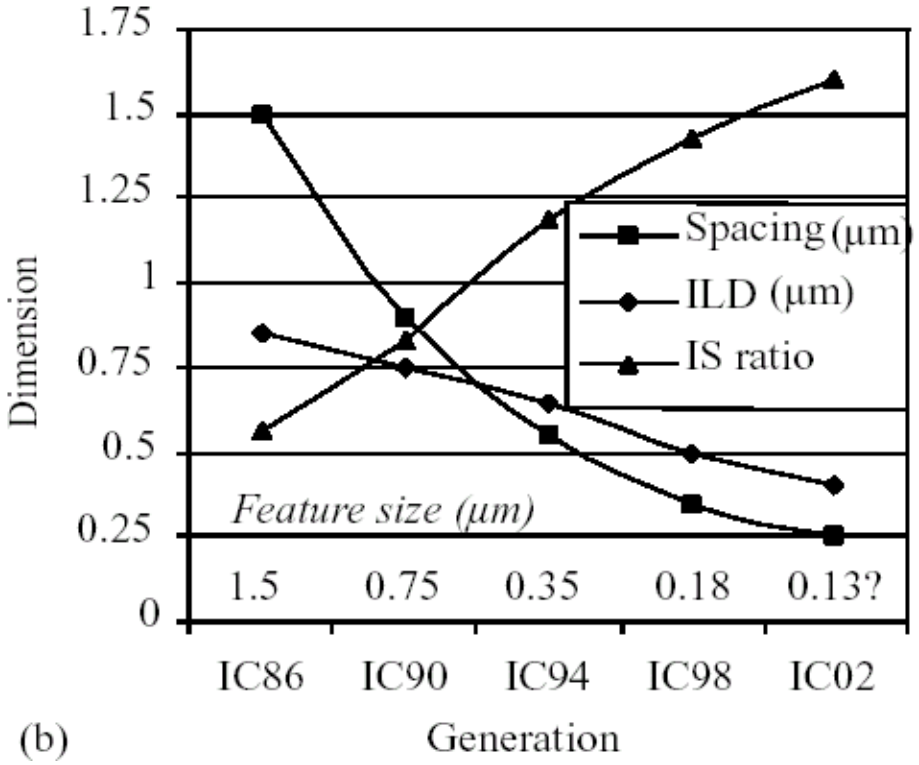
NTRS Roadmap

Year / Parameter	2003	2004	2005	2008	2011	2014
Technology(nm)	120	110	100	70	50	35
# of Transistors	95.2M	145M	190M	539M	1523M	4308M
Clock Frequency	1724 MHz	1857 MHz	2000 MHz	2500 MHz	3000 MHz	3600 MHz
Chip Area (mm^2)	372	372	408	468	536	615
Wiring Levels	8	8	8-9	9	9-10	10
Pitch(L/I/G)(nm)	330/420/690	295/375/620	265/340/560	185/240/390	130/165/275	95/115/190
A/R (L/I/G)	1.6/2.2/2.8	1.6/2.3/2.8	1.7/2.4/2.8	1.9/2.5/2.9	2.1/2.7/3.0	2.3/2.9/3.1
Dielectric Const.	2.2-2.7	2.2-2.7	1.6-2.2	1.5	<1.5	<1.5

This data is from the ITRS Roadmap on interconnects

Interconnect dimension trends





These figures are derived from *Design of High-Performance Microprocessor Circuits*, A. Chandrakasan, W. Bowhill, F. Fox, IEEE, 2001

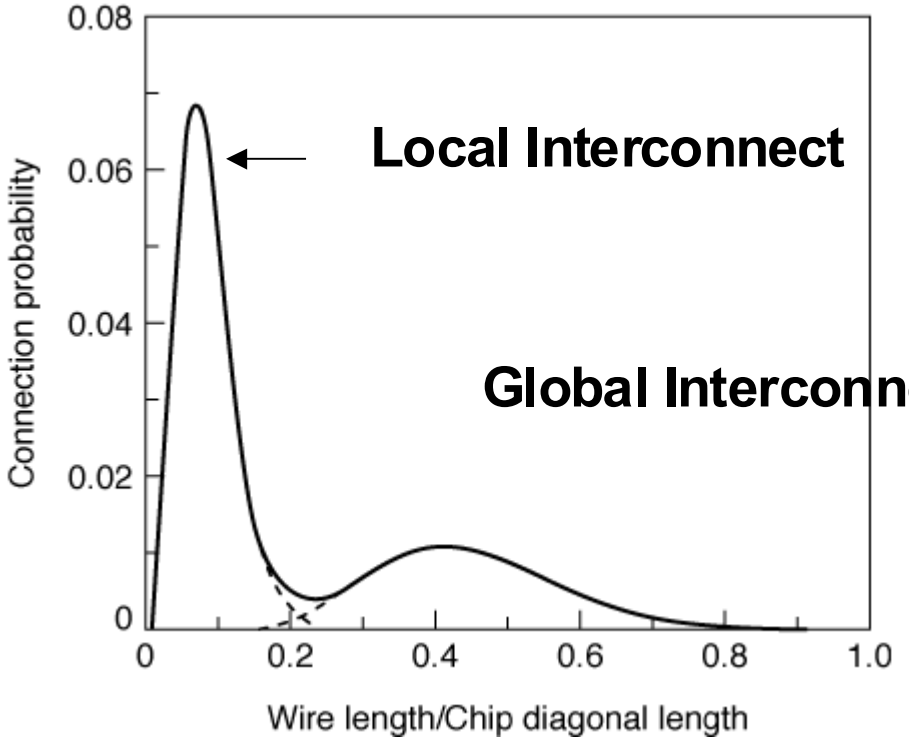
Rent's rule

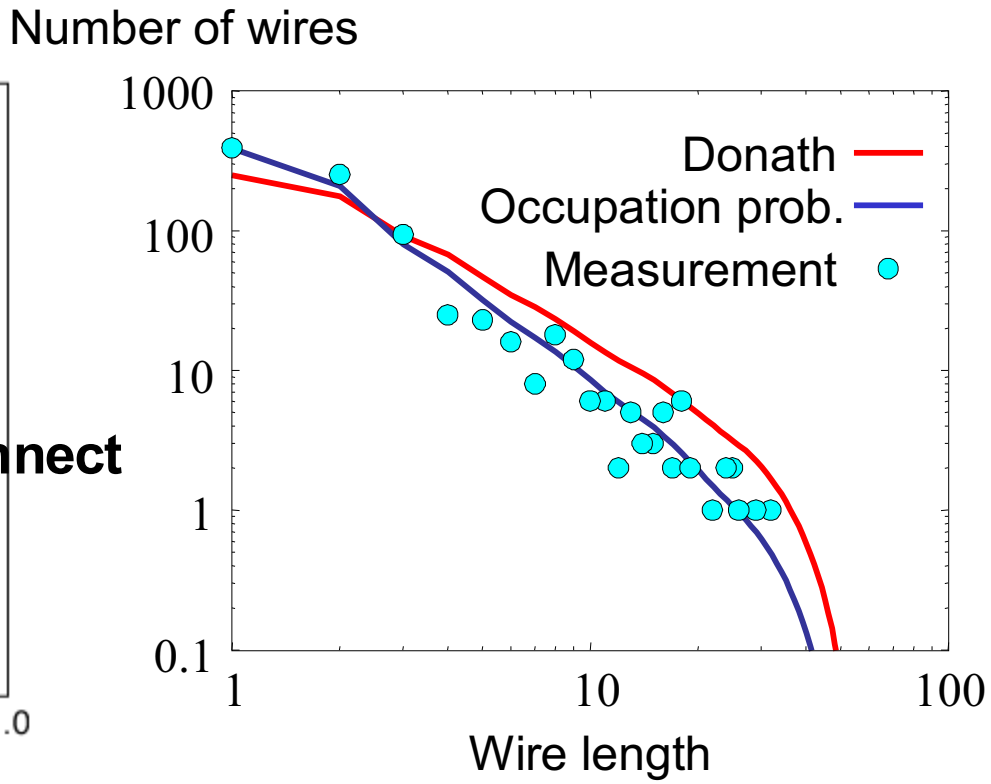
- Rent's rule relates the I/O requirement to the number of gates as :

$$N_p = K_p N_g^{\beta}$$

- As technology scales number of gates in a given area is increasing.
- More routing is required as technology scales.

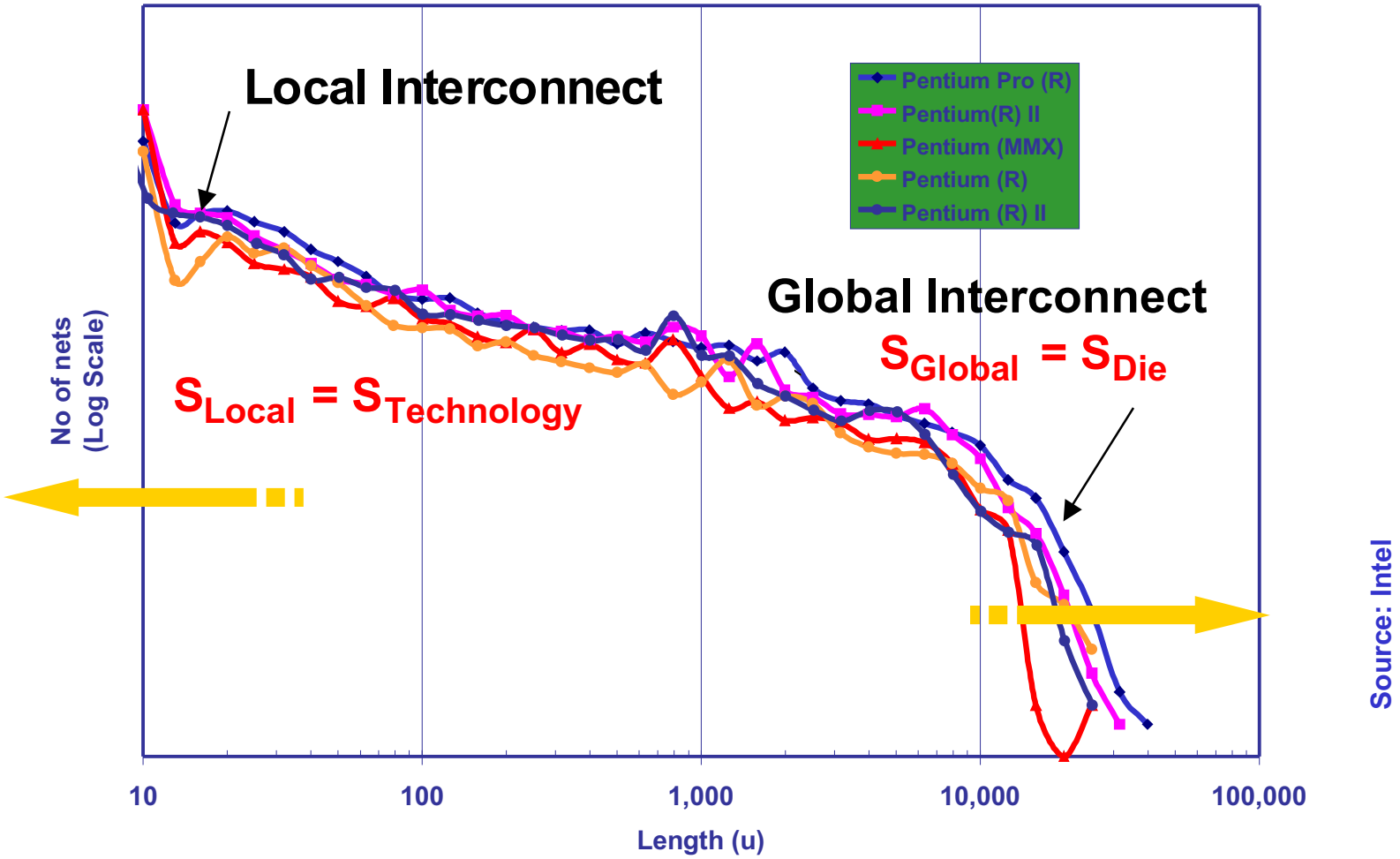
Nature of the interconnect



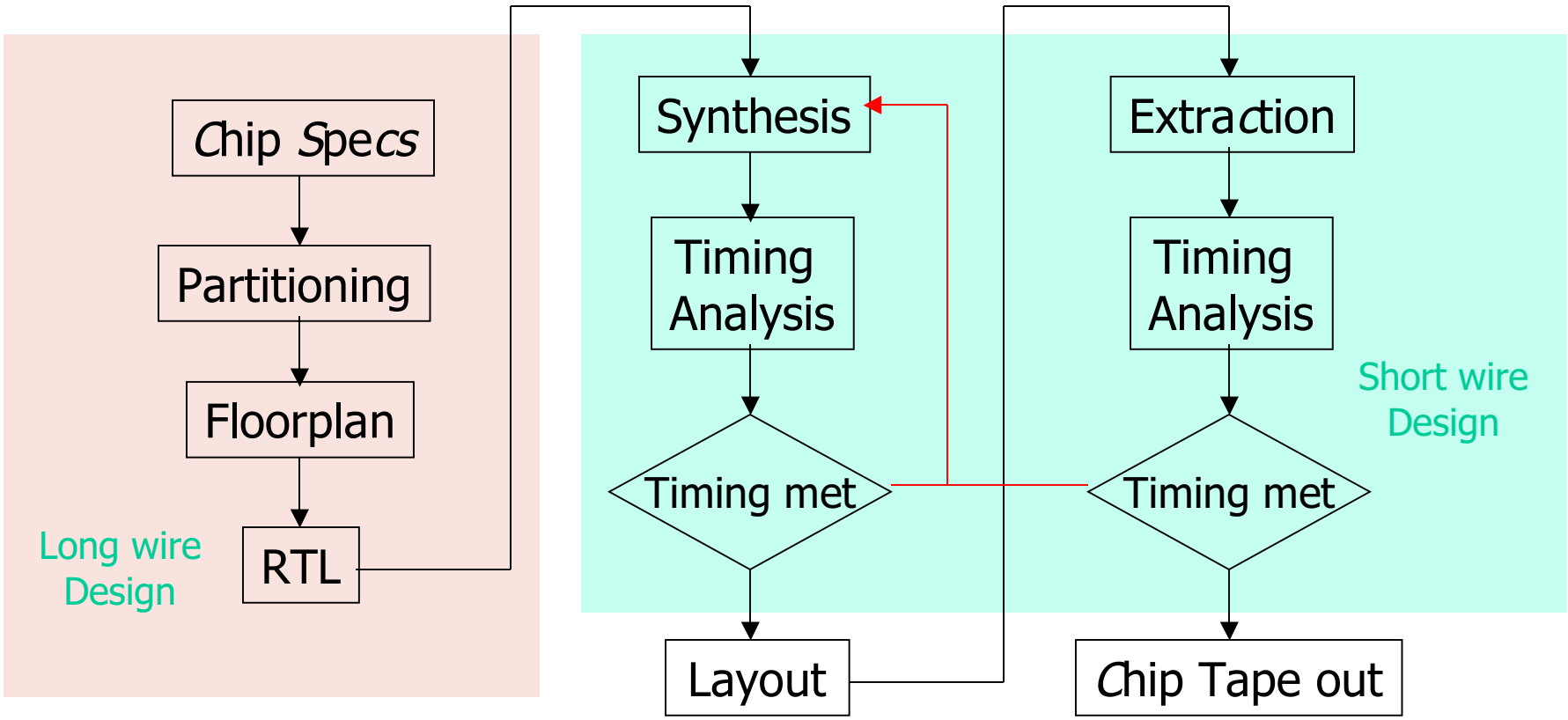


These figures are derived from *Digital integrated circuit – a design perspective*, J. Rabaey Prentice Hall and a tutorial in SLIP by Dirk *St*roobandt respectively

Microprocessor Interconnect

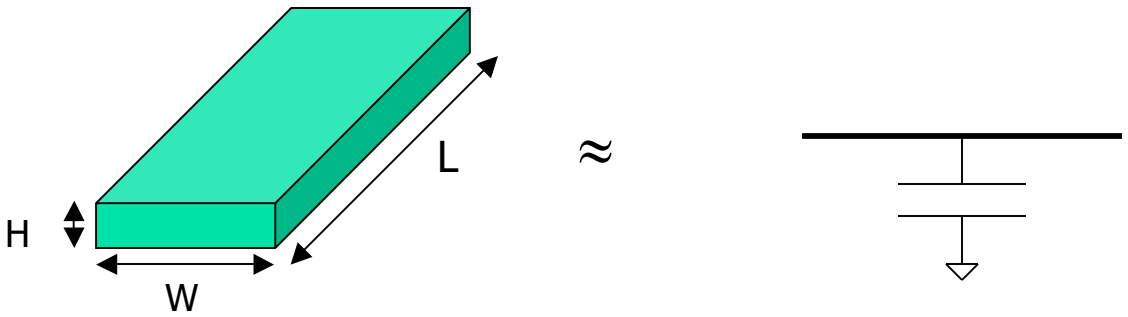


VLSI Design Cycle



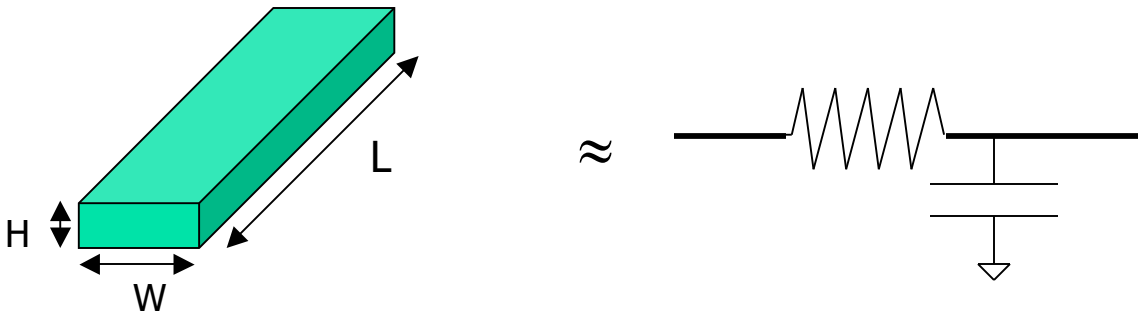
Early models

- Wire width $\propto$ feature size
- Older technology had wide wires
- More cross-section area implies less resistance and more capacitance.
- Model wire only with capacitance

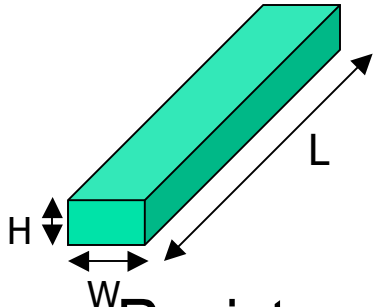


However…

- With scaling, width of wire reduced.
- Resistance of the wire no longer negligible.
- Wire not very long and a lumped RC is good enough approximation.



Interconnect Resistance



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

- Ohm's Law: Resistance of wire $\propto$ wire length (L) and $1/\propto$ cross-section(HW)
- ρ (resistivity) is the property of the material.

Sheet Resistance

- Wire height (H) is constant for a technology.
- Sheet resistance (R_q) is constant for each metal layer.
- Calculation of wire resistance is easy : multiply R_q by L/W

$$R = R_q \frac{L}{W}$$

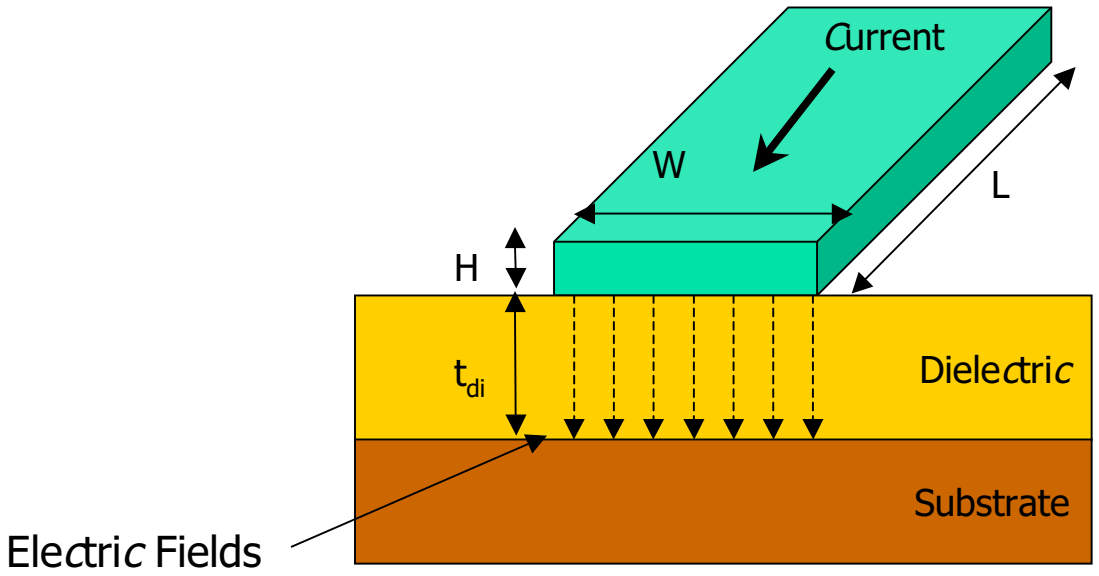
with

$$R_q = \frac{\rho}{H}$$

Interconnect Capacitance

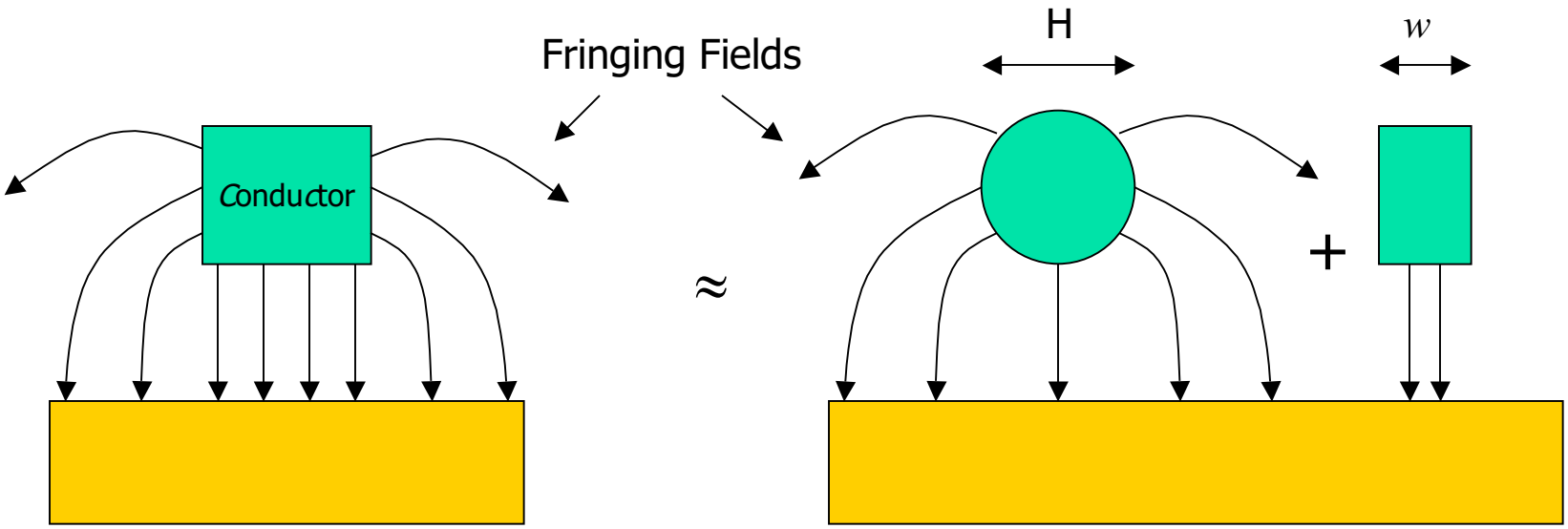
- Capacitance of a wire = f(Shape, Distance to surrounding wires, Distance to the substrate)
- Estimating Capacitance is a matter of determining where the field lines go.
- To get an accurate estimate electric field solvers (2D or 3D) are used. E.g. Fastcap or Rafael
- When in doubt, typical wires have self capacitance between 1 and 3 pf/cm

Area Capacitance



$$C_{\text{int}} = \frac{\varepsilon_{di}}{t_{di}} WL$$

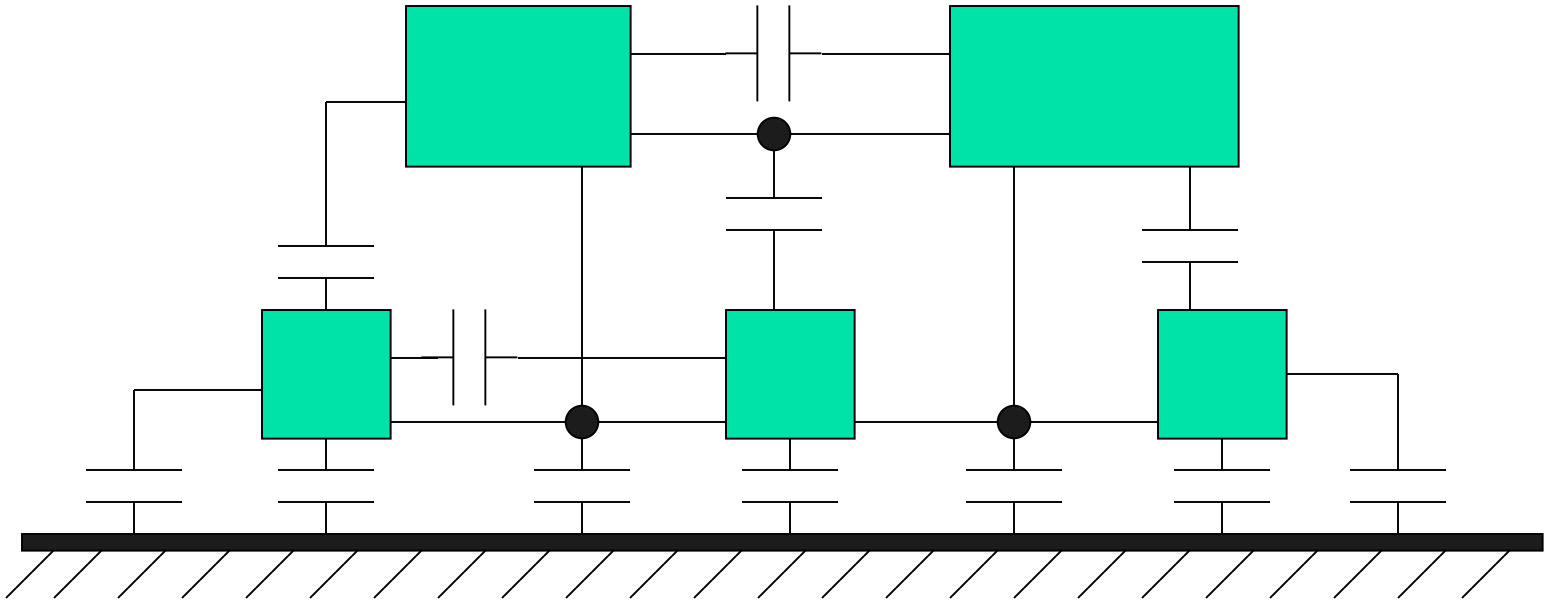
Fringing Capacitance



$$C_{wire} = C_{pp} + C_{fringe} = \frac{2\pi\varepsilon_{di}}{\log(4t_{di}/H + 2)} + \frac{w\,\varepsilon_{di}}{t_{di}}$$

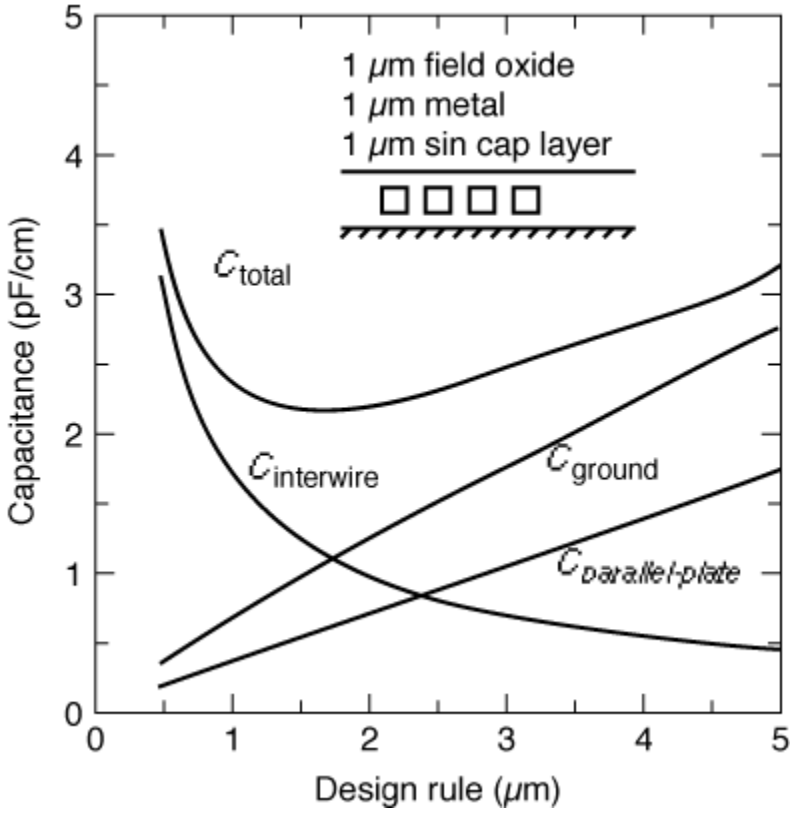
$w \approx$ W-H/2

Detailed Picture



Is this much of detail required… How to *c*ompute this?

Interwire Capacitance



(from [Bakoglu89])

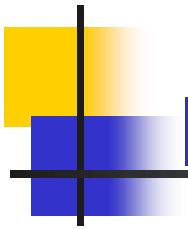
Wiring Capacitances (0.18μm)

Capacitance	N+	P+	poly	m1	m2	m3	m4	m5	m6
substrate	998+244	1152+201	103+21	39+38	19+61	13+55	9+43	8+25	3
N+ active			8655	54	21	14	11	10	9
P+ active			8324						
poly				64+69	18+39	10+29	7+24	6+21	5+19
m1					44+61	16+35	10+31	7+23	5+21
m2						38+54	15+37	9+27	7+24
m3							40+56	15+34	9+31
m4								37+58	14+40
m5									36+61

Units: First number is area *c*omponent (af/μm^2), se*c*ond is fringing *c*omponent (af/μm)

How to use fringe capacitance tables

- Estimation of wire Capacitance
  - Where do field lines terminate?
  - What fraction go where?
- E.g. 1cm of M1 over substrate:39af/μm^2, 38af/μm fringe
  - If 200nm wide = 0.2um, 0.2um*10,000um=2,000*39af=78fF
  - 1cm = 10,000um, fringe on both sides: 2*38af*10,000 = 760fF
  - Total = 848fF/cm
- Over Poly 64aF, 69aF – nearly doubles (half the distance to conductor)

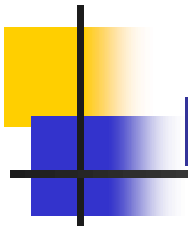


Importance of Resistance

- Delay of wire $\propto$ to the resistance of the wire.
- Resistance means ohmic (IR) drop along the wire, reduces noise margin.
- IR drop a significant problem in the power lines where current density if high.
- Keep wires short, to reduce resistance.
- Contact resistance makes them vulnerable to electromigration.

Metal Resistivity

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}

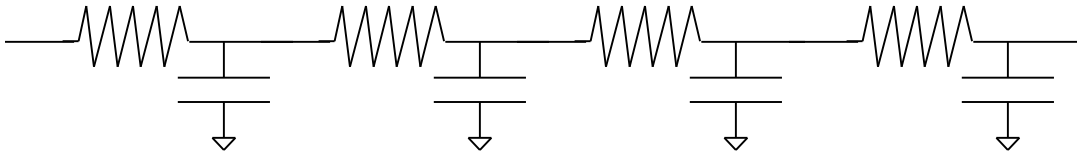


Importance of capacitance

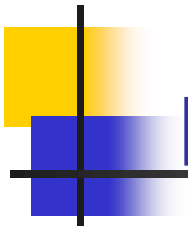
- Delay of the wire is proportional to the capacitance charged.
- More capacitance means more dynamic power.
- Capacitance an increasing source of noise (coupling).
- Coupling make delay estimation hard.

Distributed model

- Wire can be modeled as a distributed RC line.



- As the number of elements increase distributed model becomes more accurate.
- For practical purposes wire-models with 5-10 elements are used to model the wire.

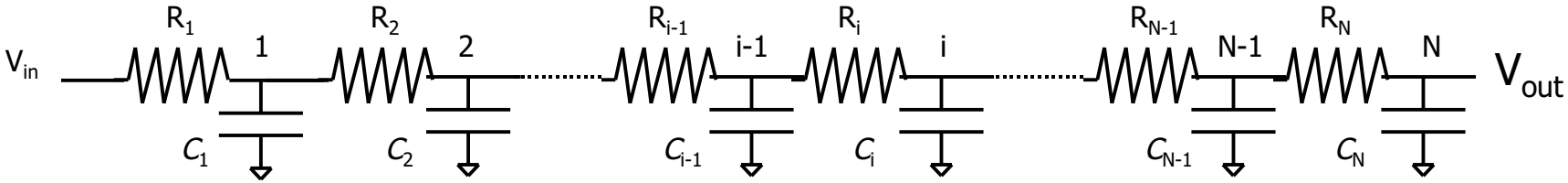


Elmore Delay…

- First order time constant at node is a sum of RC components.
- All the upstream resistances are taken into account.
- Thus each node contributes to the delay.
- Amount of contribution is the product of the cap at the node and the amount of resistance from source to the node.

Delay in distributed RC line

- Elmore analyzed the distributed model and came up with the figures for delay.



$$\tau_N = \sum_{i=1}^{N} R_i \sum_{j=i}^{N} C_j = \sum_{i=1}^{N} C_i \sum_{j=1}^{i} R_j$$

Elmore derived this equation in 1948 way before VLSI !!!

Wire Model

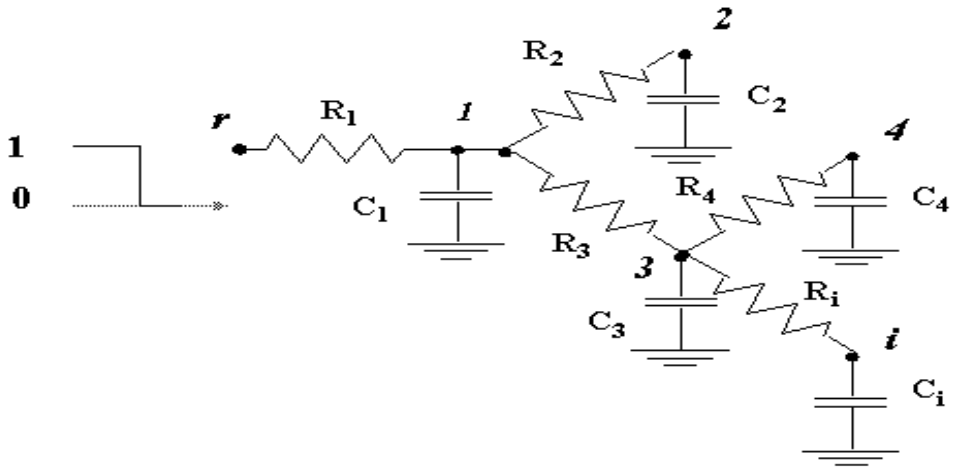
Assume: Wire modeled by N equal-length segments

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

For large values of N:

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

Generalized Elmore delay



$$\int_0^\infty v_i(t)dt = \sum_{k=1}^{N} C_k V_k(0) R_{i,k}$$

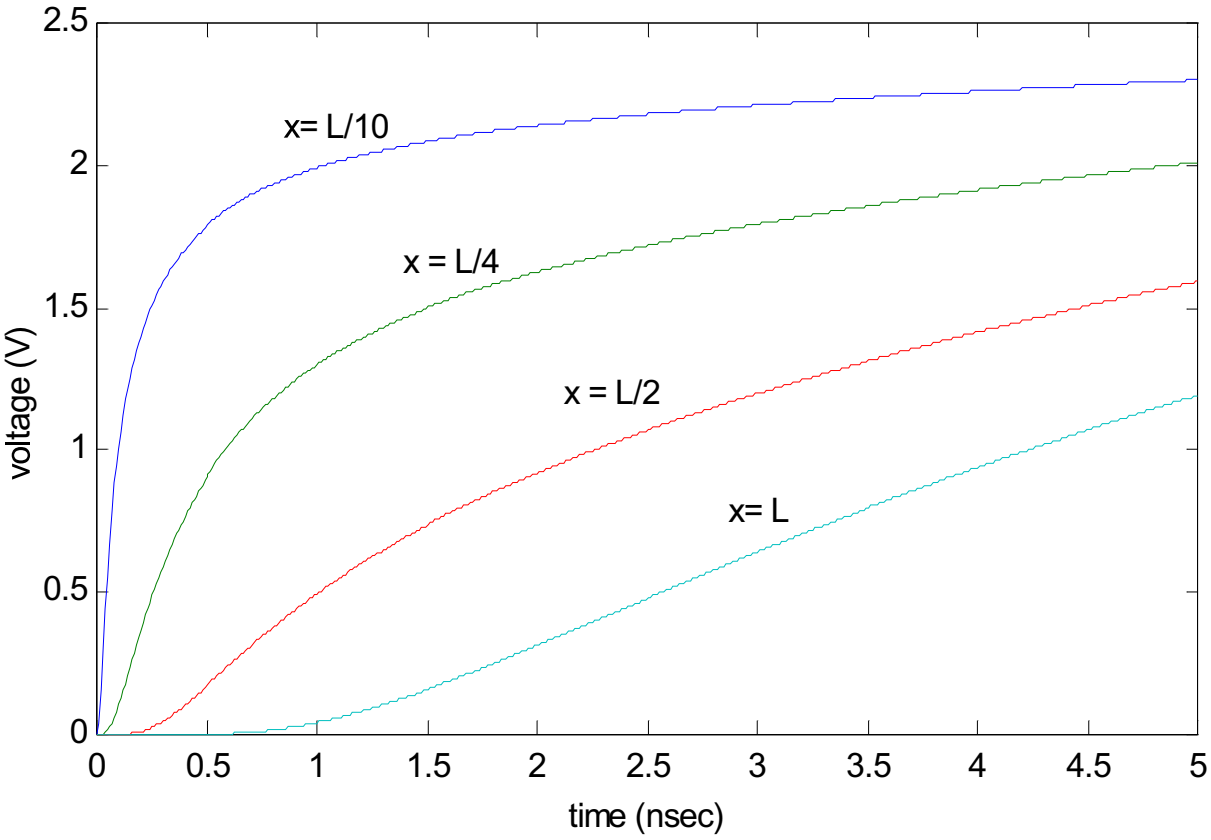
with

$$R_{i,k} = \sum R_j \Rightarrow \left(R_j \in [path(i \to r) \cap path(k \to r)]\right)$$

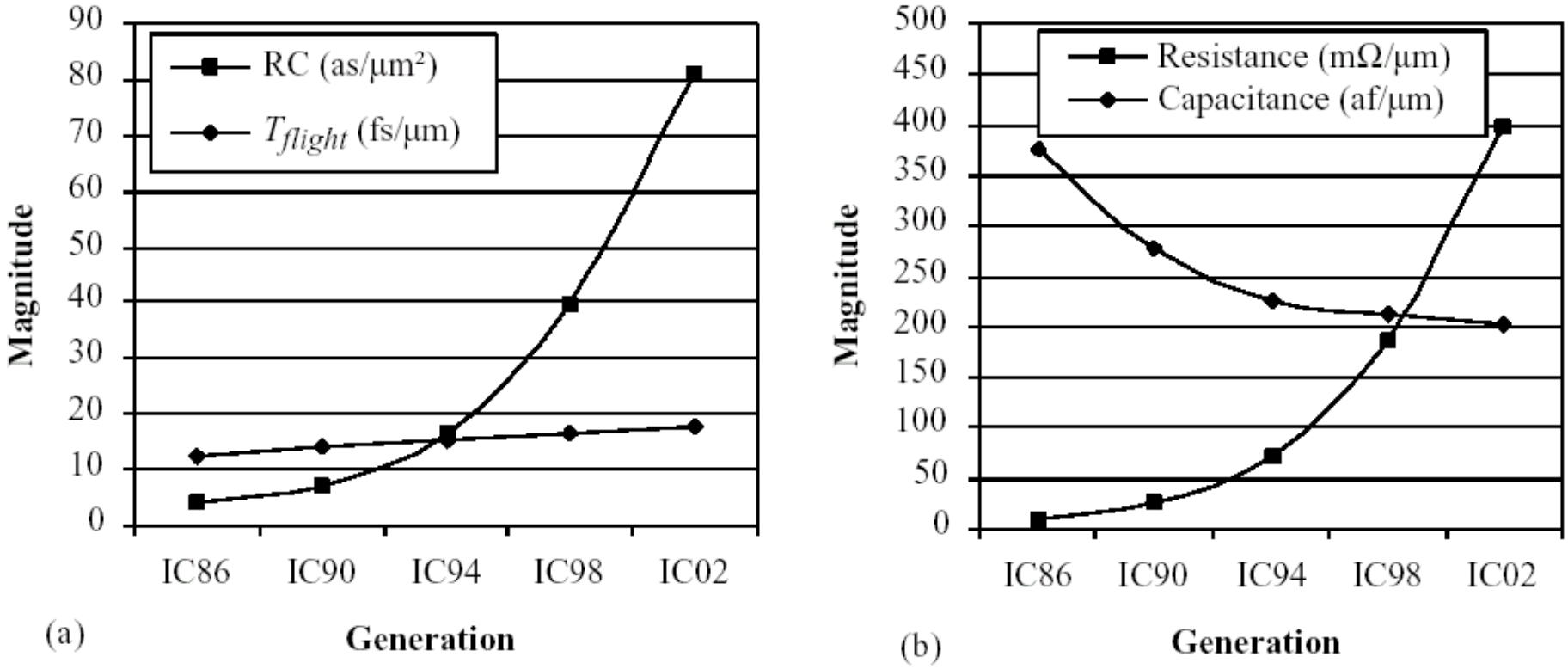
Rubinstein, Penfield and Horowitz generalized Elmore delay

This figure is derived from *Digital integrated circuit – a design perspective*, J. Rabaey Prentice Hall

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



RC and flight-time

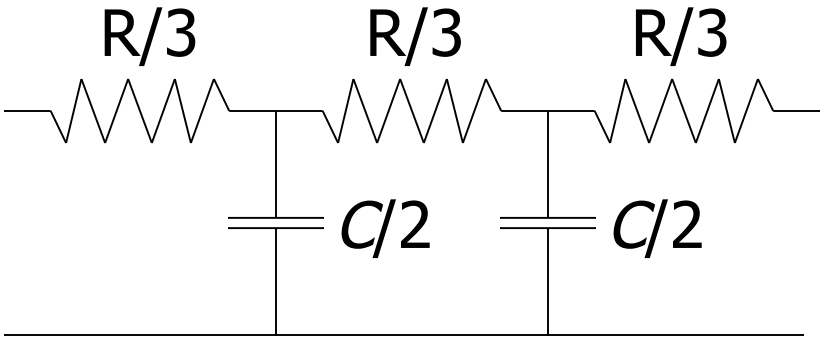


for a wide bus above a plane and beneath orthogonally routed layer

These figures are derived from *Design of High-Performance Microprocessor Circuits*, A. Chandrakasan, W. Bowhill, F. Fox, IEEE, 2001

Pi Model

- Pi Model of wire:



- Elmore Delay = RC/3+RC/6 = RC/2 agrees with distributed model RC
- Pi Model is often used in Spice instead of large number of segments as a reasonable approximation of distributed RC

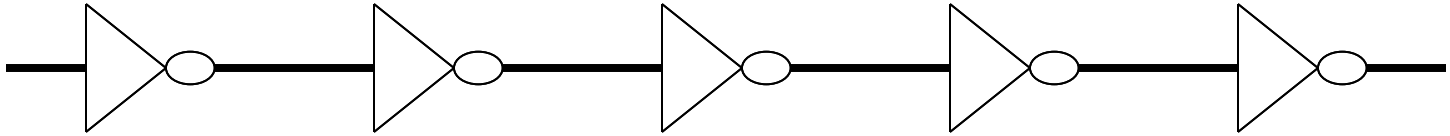
Driving an RC-line

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

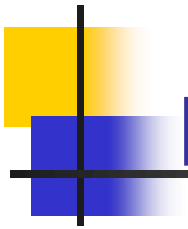
$$t_p = 0.69 R_s C_w + 0.38 R_w C_w$$

- Delay for FET driven distributed $\mathrm{RC} - \mathrm{R_s}$ is equivalent source resistance (usually assumed $\mathrm{R_s} = \mathrm{V_{dd}}/(2\ \mathrm{I_{sat}})$)
- RC gives delay for $\exp^{(-1)}$ change in output
- Scale time by $\ln(2) = 0.69$ to get typical 50% CMOS gate threshold

Repeaters



- Repeaters are buffers or inverters inserted at regular intervals.
- Delay linearly proportional to the wire length
- Questions to be answered – Where and how big should the repeaters be ?



Repeater placement

- Delay of the interconnect is typically optimum when

$$Delay_{wire} = Delay_{buffer}$$

- Closed form solutions for Repeater Number and Sizing
  - Bakoglu and Meindl, 1985 (Classical)
  - Adler and Friedman, 1998 (considering inductance)
  - Nalamalpu and Burleson, 2000 (ramped waveforms)
  - Chen Marek-Sadowska, Brewer, 2003 (short channel timing)
  - Cong, 2004 (tapered wires)

Bakoglu and Meindl

- For a wire with k repeaters each of size h times minimum size inverter is given by:

$$T_{50\%} = k\left[0.7\ \frac{R_o}{h}\left(\frac{C_{\text{int}}}{k} + hC_o\right) + \frac{R_{\text{int}}}{k}\left(0.4\frac{C_{\text{int}}}{k} + 0.7hC_o\right)\right]$$

Bakoglu and Meindl…

- By setting $\mathrm{dT/dk} = 0$ and $\mathrm{dT/dh} = 0$, optimal values for k and h are obtained

$$k = \sqrt{\frac{0.4R_{\text{int}}C_{\text{int}}}{0.7R_oC_o}} \qquad h = \sqrt{\frac{R_oC_{\text{int}}}{R_{\text{int}}C_o}}$$

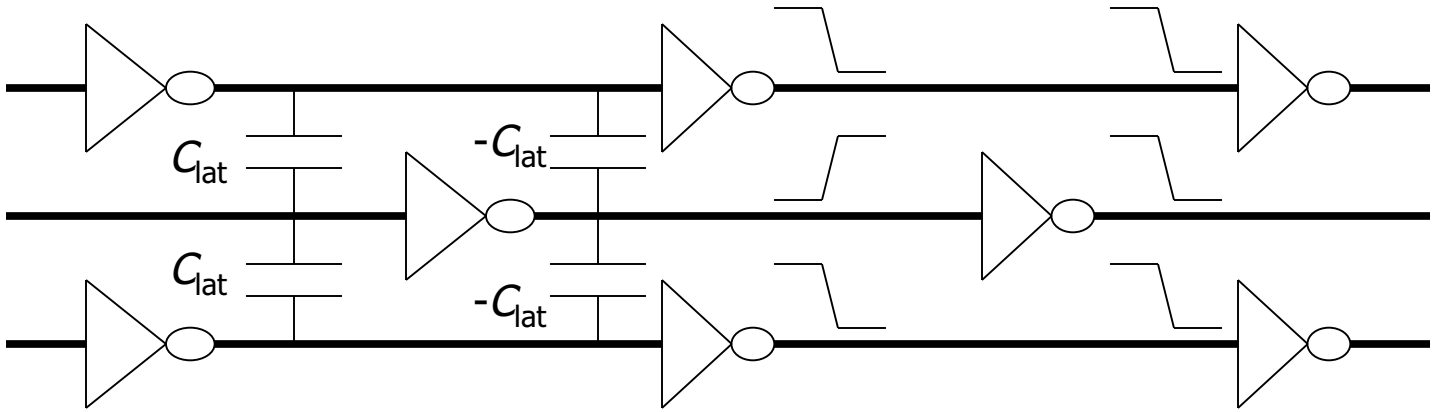
- Substituting these back, delay is given by

$$T_{50\%} = 2.5\sqrt{R_oC_oR_{\text{int}}C_{\text{int}}}$$

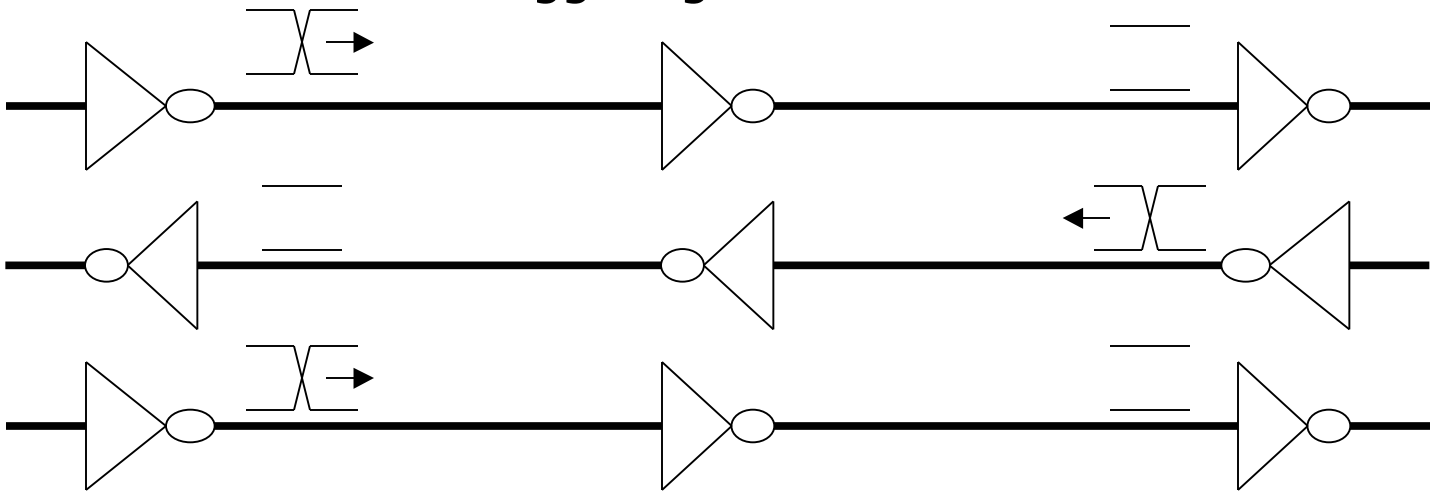
Optimization: Lagrange Multipliers

- A general technique for multi-dimensional optimization
- Problem: A function $f(x_1, x_2, \ldots, x_n)$ to maximize subject to several constraints: $g_1(x_1, x_2, \ldots, x_n) = 0$, $g_2(x_1, x_2, \ldots, x_n)=0$, $\ldots$, $g_m(x_1, x_2, \ldots,x_n) = 0$ where $m<n$.
- Solution: $\nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 + \ldots + \lambda_m \nabla g_m$
- The n-dimensional equation above plus the m constraints provide $n+m$ equations in $n+m$ variables (x_i's and λ_j's)
- Note: it is often useful to examine the functional forms of the lambdas – they are usually interesting.

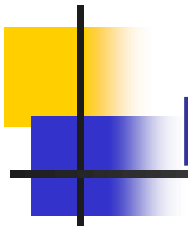
Best Placement for repeaters



Staggering the inverters



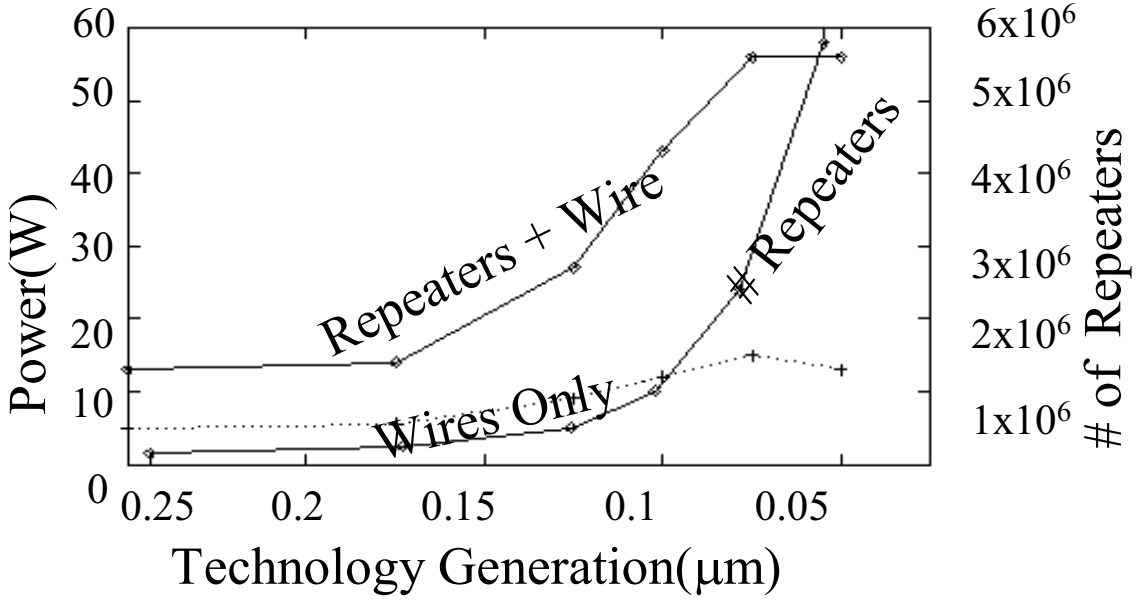
Avoiding the Miller *c*ap by opposite going signals



Repeater Design Issues

- Delay-optimal repeaters are area and power hungry – use of sub-optimal insertion
- Optimal placement requires accurate modeling of interconnect.
- Optimal placement not always possible.
- Performance limited due to significant interconnect resistance.
- Source of noise – Supply and Substrate

With Scaling …



- 1 million repeaters in a 100nm technology.
- Consuming about 30W (40%) in 100nm technology.
- Need to look at alternatives!!!

Differential Transmission

- Limiting swing saves significant amount of power.
- Rejects common-mode noise.
- Coupling is reduced due to dipole cancellation $O(n^3)$
- Doubled wire density --