On-Chip Interconnect: The Past, Present, and Future

Professor Eby G. Friedman

Department of Electrical and Computer Engineering
University of Rochester

URL: http://www.ece.rochester.edu/~friedman

Future Interconnects and Networks on Chip

1st NoC Workshop – DATE '06

March 10, 2006

Agenda

- A historical perspective
- Where are we now ?
- Where are we going ?
- Conclusions

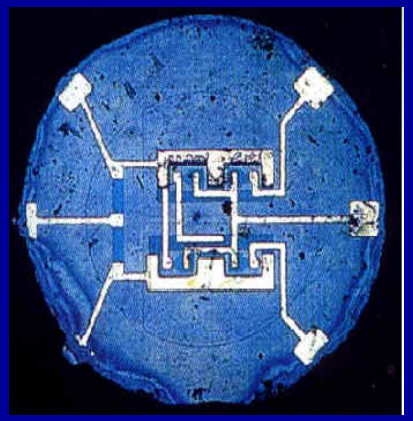
Agenda

⇒ A historical perspective

- Where are we now ?
- Where are we going ?
- Conclusions

Advances in IC Technologies

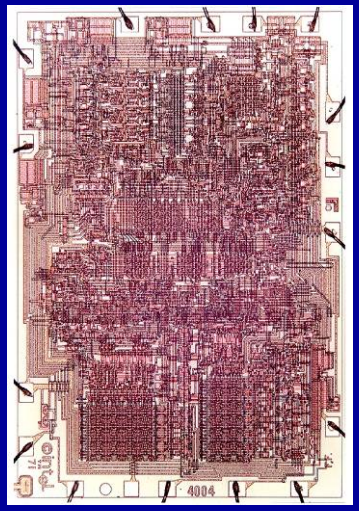
- A journey that started in 1959



First integrated circuit

Fairchild Semiconductor

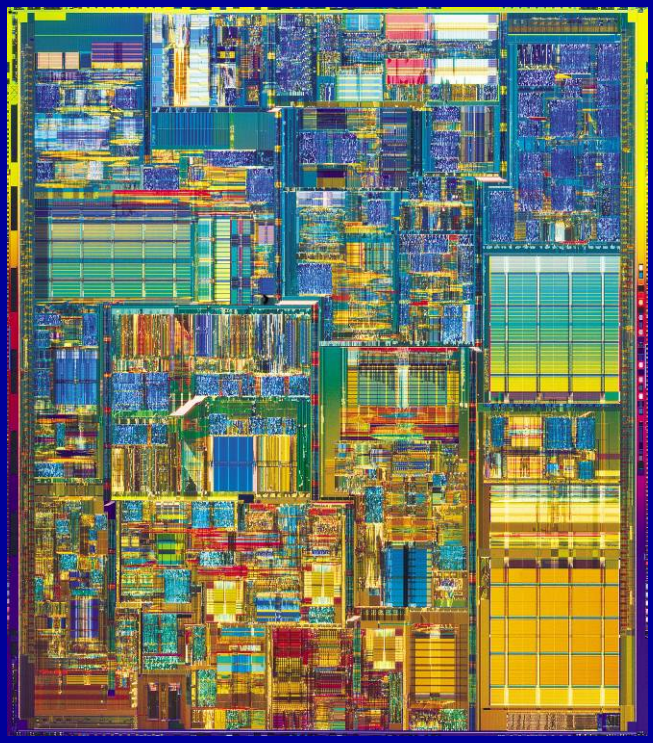
1959



First microprocessor

Intel 4004

1971



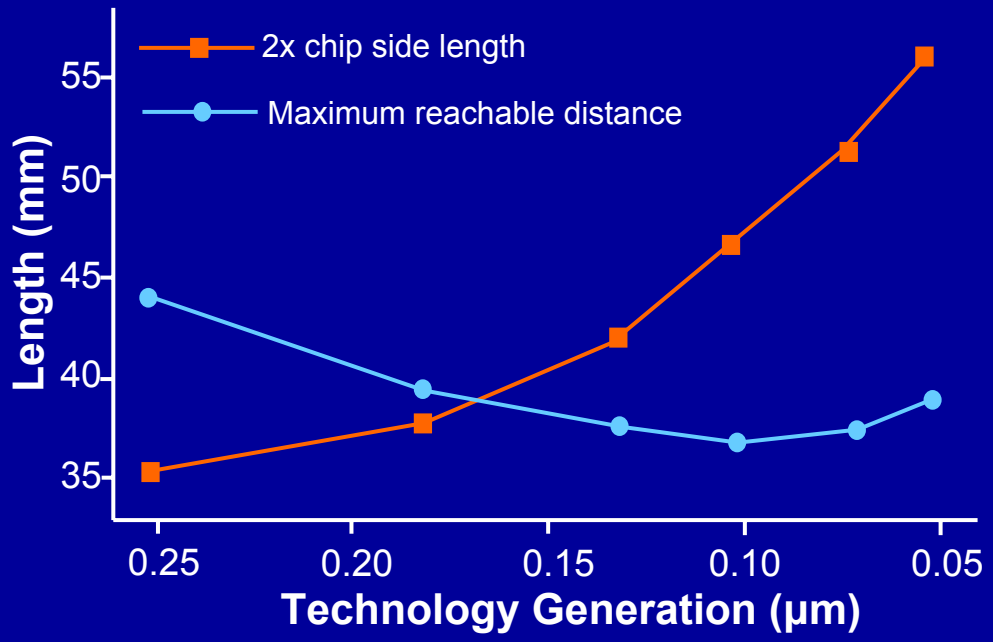
Pentium 4

Intel Corporation

2002

Die Size Constraints

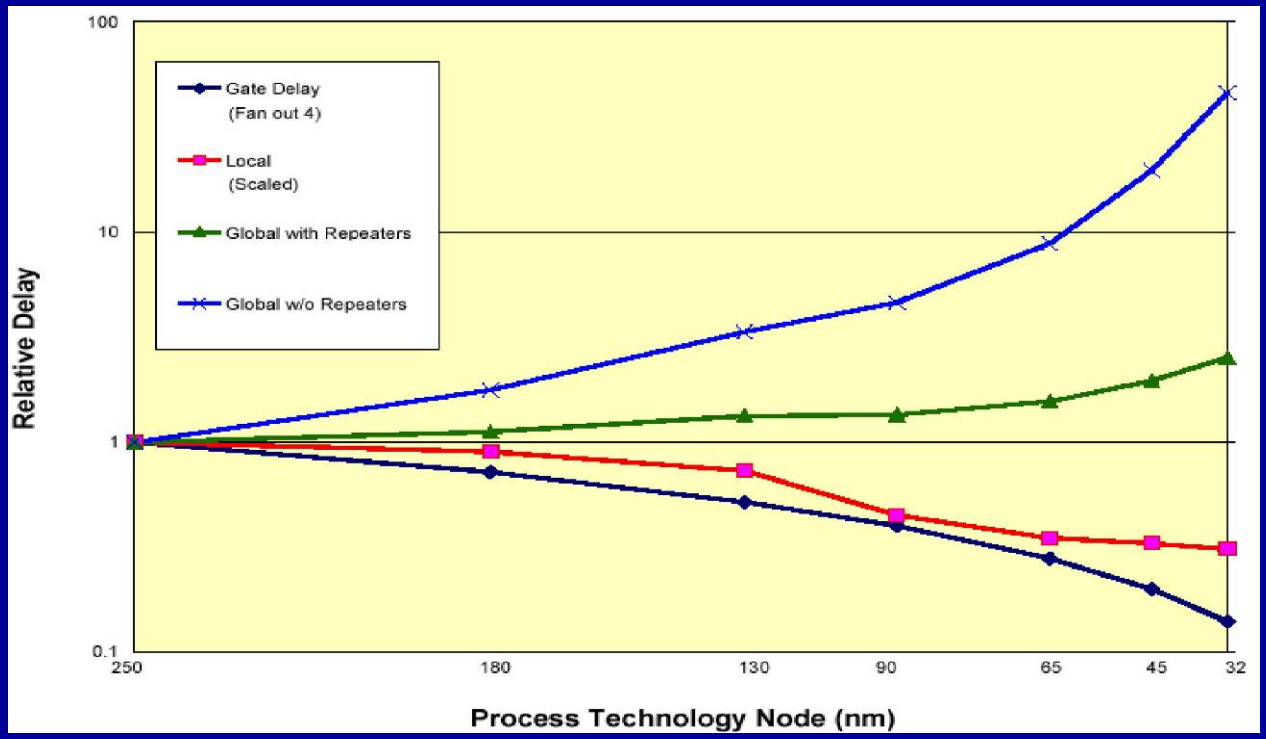
- Signals cannot travel across the entire die
  - Within a global clock cycle
- Example
  - 1.5 μm wire thickness
  - Reachable distance
    - Distance that can be traveled in 80% of the global clock cycle



- Constraints in the routing distance of signal lines

*D. Sylvester and K. Keutzer, "A Global Wiring Paradigm for Deep Submicron Design," *IEEE Tran. on CAD*, February 2000

Trends in Interconnect Delay



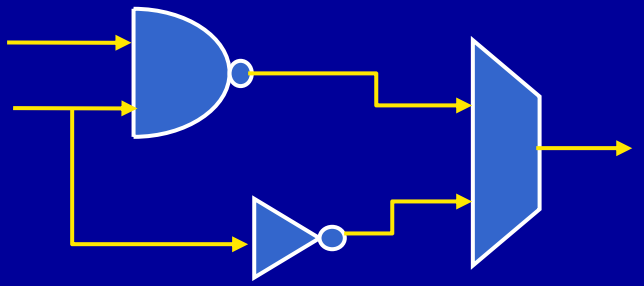
Two to three order of magnitude delay difference

2001 International Technology Roadmap for Semiconductors

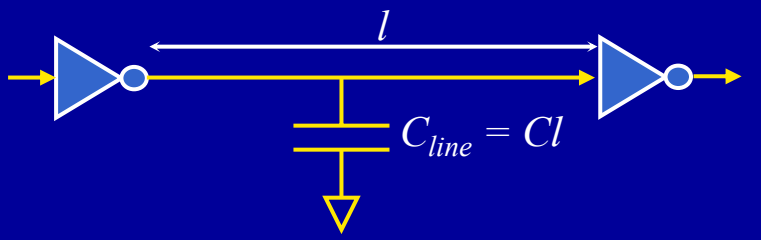
- Interconnect delay dominates gate delay
  - Global interconnect delay continuously increasing
  - Need multiple clock cycles to cross chip die
  - Limits the performance of microprocessors

History of Interconnect Modeling

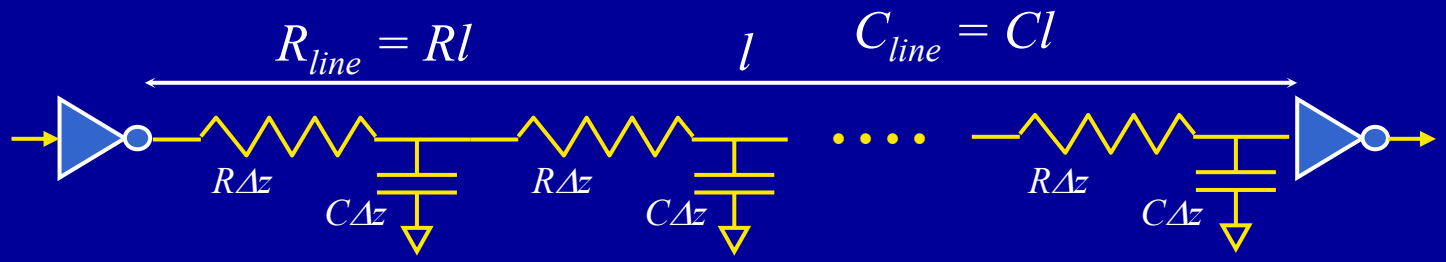
- Gate delay was dominant
  - Interconnect was modeled as short-circuit



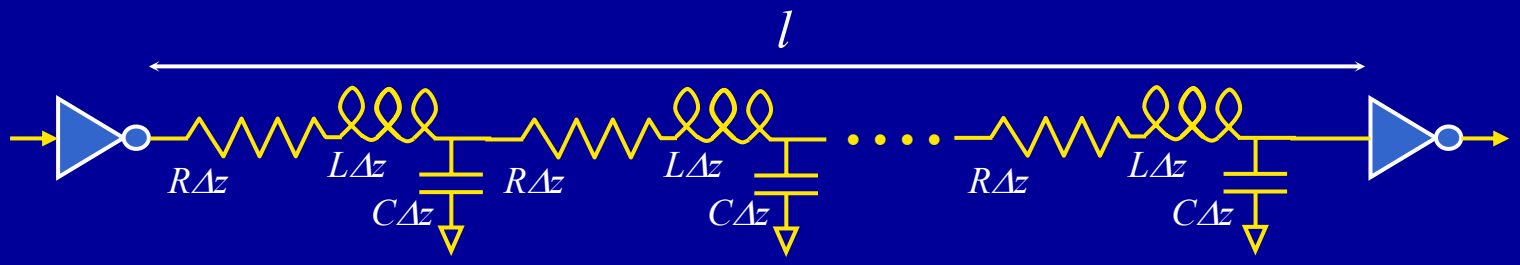
- Interconnect capacitance became comparable to gate capacitance



- Interconnect resistance became comparable to gate resistance



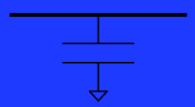
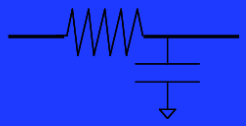
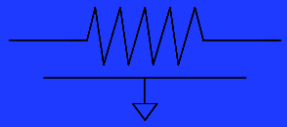
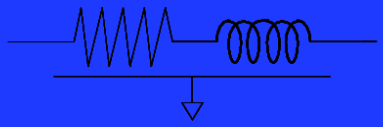
Modeling Interconnect Inductance



$$C_{line} = Cl \qquad R_{line} = Rl \qquad L_{line} = Ll$$

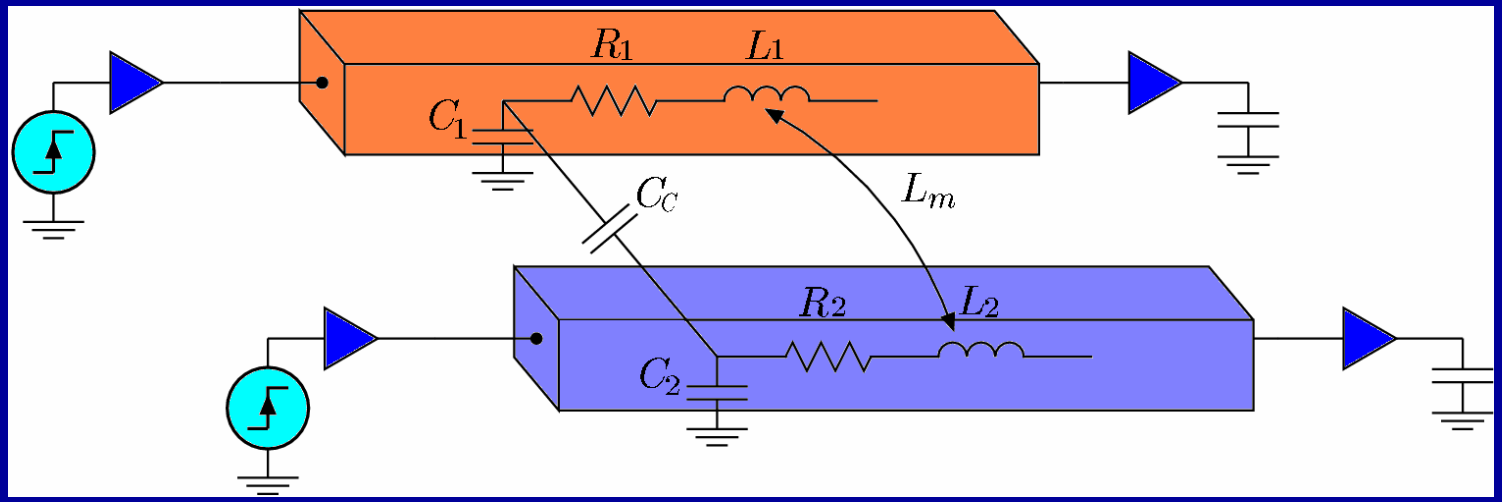
- Factors that make inductance effects important
  - Signal transition times are much shorter
    - Comparable to the signal time of flight
    - Faster devices
  - Reduction in interconnect resistance
    - Wide lines at higher metal layers
    - Introduction of low resistance materials for interconnect

Interconnect Models

Technology	> 1 µm	1.0 ~ 0.5 µm	0.5 ~ 0.25 µm	< 0.25 µm
Resistance	Ignored	Lumped	Distributed	Distributed
Capacitance	Area Capacitance	Area, 2-D fringing	2/3-D fringing, Lateral coupling	Lateral coupling dominant
Inductance	Ignored	Ignored	Ignored	Important
Circuit model	Lumped C	Lumped RC	Distributed RC	Distributed RLC
Circuit Symbol				

- Significant inductance effects with technology scaling

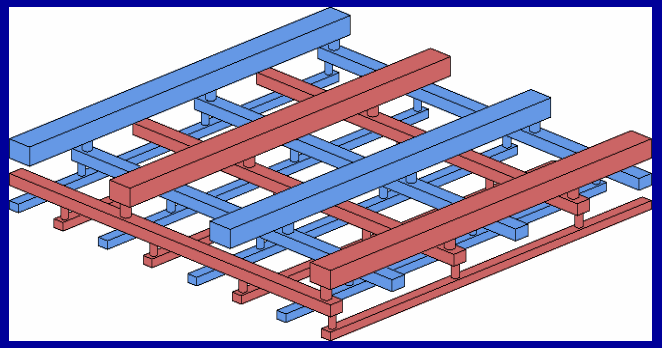
Crosstalk Between Coupled *RLC* Interconnects

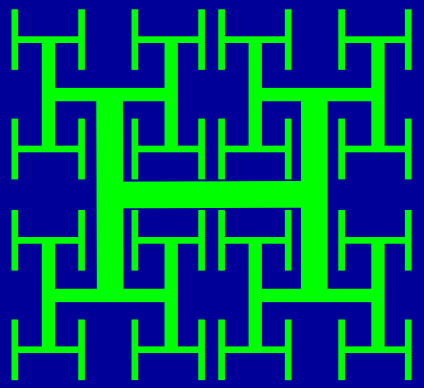


- Significant inductive effects in the upper metal layers
  - Faster rise times
  - Lower resistance
    - Wider lines
    - Copper interconnect
- Capacitive/inductive coupling degrades the signal integrity
  - Crosstalk noise increases

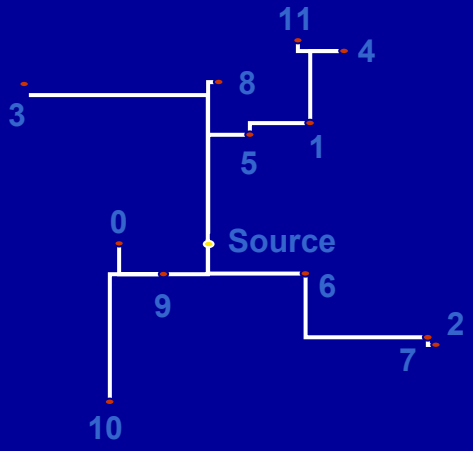
Interconnect Networks

- Power distribution networks
  - Consume about 30% on-chip metal
  - *IR, Ldi/dt* noise
  - *RLC* resonances
- Clock distribution networks
  - Consume up to 70% of the total power
  - Clock skew, jitter
- Signals with multiple fan-out
- Large global busses





- Interconnect networks have become increasingly complicated with greater integration
  - Accurate and efficient models are required



Agenda

- A historical perspective

⇒ Where are we now ?

- Where are we going ?
- Conclusions

Problems in On-Chip Interconnect

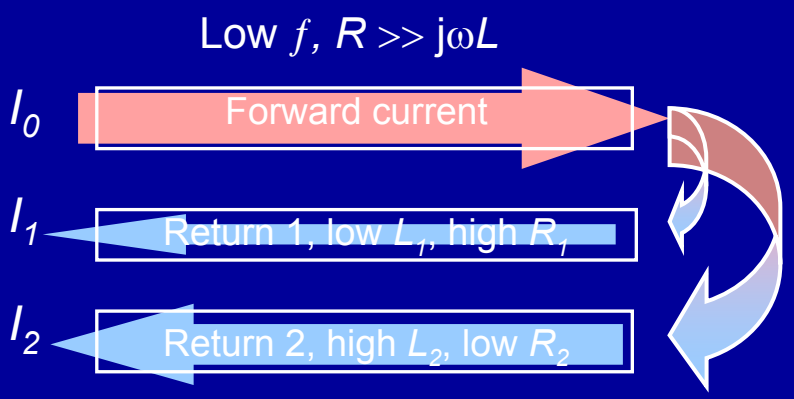
Extraction	**Figures of Merit**
Modeling/Simulation	**Design**

Problems in On-Chip Interconnect

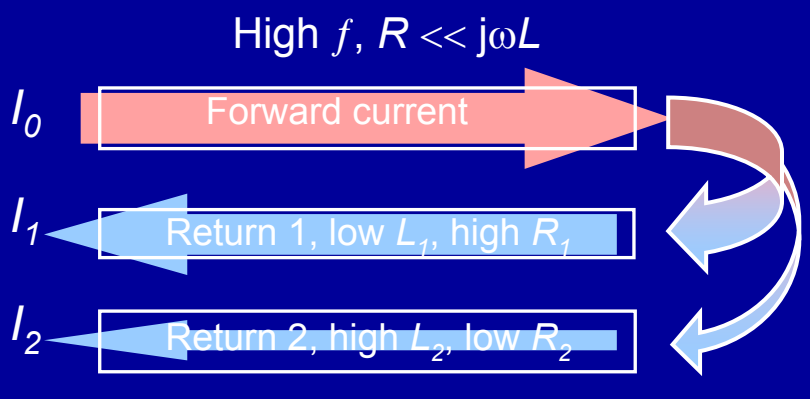
Extraction	**Figures of Merit**
Modeling/Simulation	**Design**

Dependence of Impedance on Frequency: Multi-path Current Redistribution

- In a circuit with multiple current paths the distribution of the current flow is frequency dependent
  - Low frequency — determined by the *resistance* of the paths
  - High frequency — determined by the *inductance* of the paths



$$I_1 \approx I_0 \frac{R_2}{R_1 + R_2} \qquad I_2 \approx I_0 \frac{R_1}{R_1 + R_2}$$



$$I_1 \approx I_0 \frac{L_2}{L_1 + L_2} \qquad I_2 \approx I_0 \frac{L_1}{L_1 + L_2}$$

- This effect is the primary source of inductance variation with frequency in integrated circuits

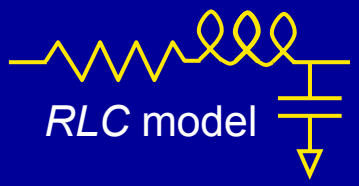
* A. V. Mezhiba and E. G. Friedman, "Impedance Characteristics of Power Distribution Grids in Nanoscale Integrated Circuits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* Vol. 12, No. 11, pp. 1148-1155, November 2004.

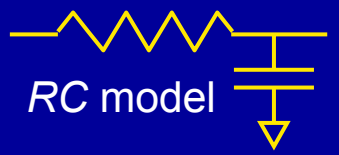
Problems in On-Chip Interconnect

Extraction	**Figures of Merit**
Modeling/Simulation	**Design**

Figures of Merit to Characterize On-Chip Inductance

- Compare a distributed *RLC* model to a distributed *RC* model

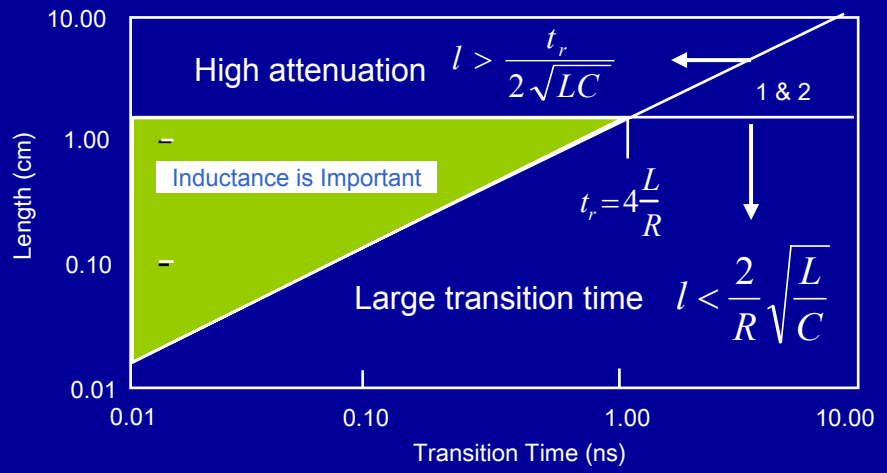




- *RC* model is sufficient if:
  - Attenuation is sufficient large to make reflections negligible
  - Waveform transition is slower than twice the time of flight

$$\frac{Rl}{2}\sqrt{\frac{C}{L}} > 1$$

$$t_r > 2l\sqrt{LC} = 2T_O$$



*Y. I. Ismail, E. G. Friedman, and J. L. Neves, "Figures of Merit to Characterize the Importance of On-Chip Inductance," *IEEE Trans.VLSI Systems*, December 1999

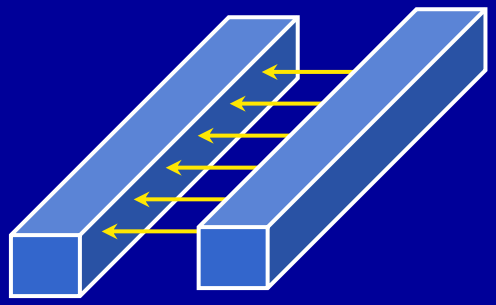
Problems in On-Chip Interconnect

Extraction	**Figures of Merit**
Modeling/Simulation	**Design**

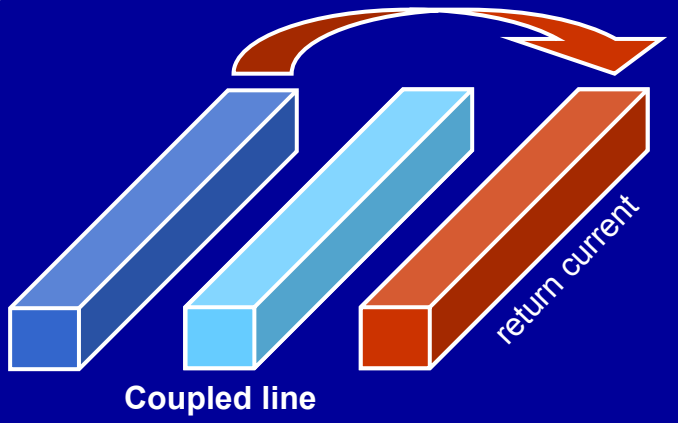
Interconnect and Substrate Coupling

- Crosstalk noise (line-to-line coupling)
  - Voltage variations
  - Delay uncertainty
  - Clock jitter
  - Depends upon wire layout and signal switching pattern

- Capacitive coupling
  - Short range effects



- Inductive coupling
  - Long range effects
  - Depend upon the current return path



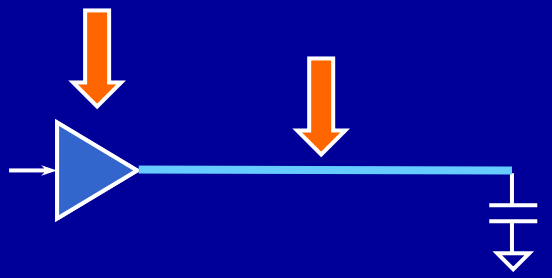
- Substrate coupling
  - Significant issue in mixed-signal circuits
  - Developing issue in digital circuits

Problems in On-Chip Interconnect

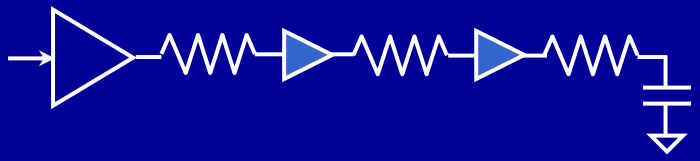
Extraction	**Figures of Merit**
Modeling/Simulation	**Design**

Interconnect Design

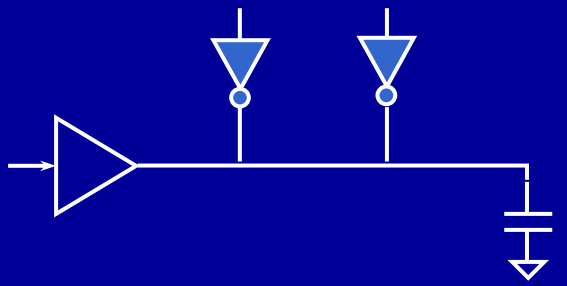
- Simultaneous driver and wire sizing
  - Optimize
    - Delay
    - Signal transition time
    - Dynamic and short circuit power



- Repeater insertion
  - Linear delay with wire length
  - Tapering factor - buffers
  - Optimal repeater sizing and spacing

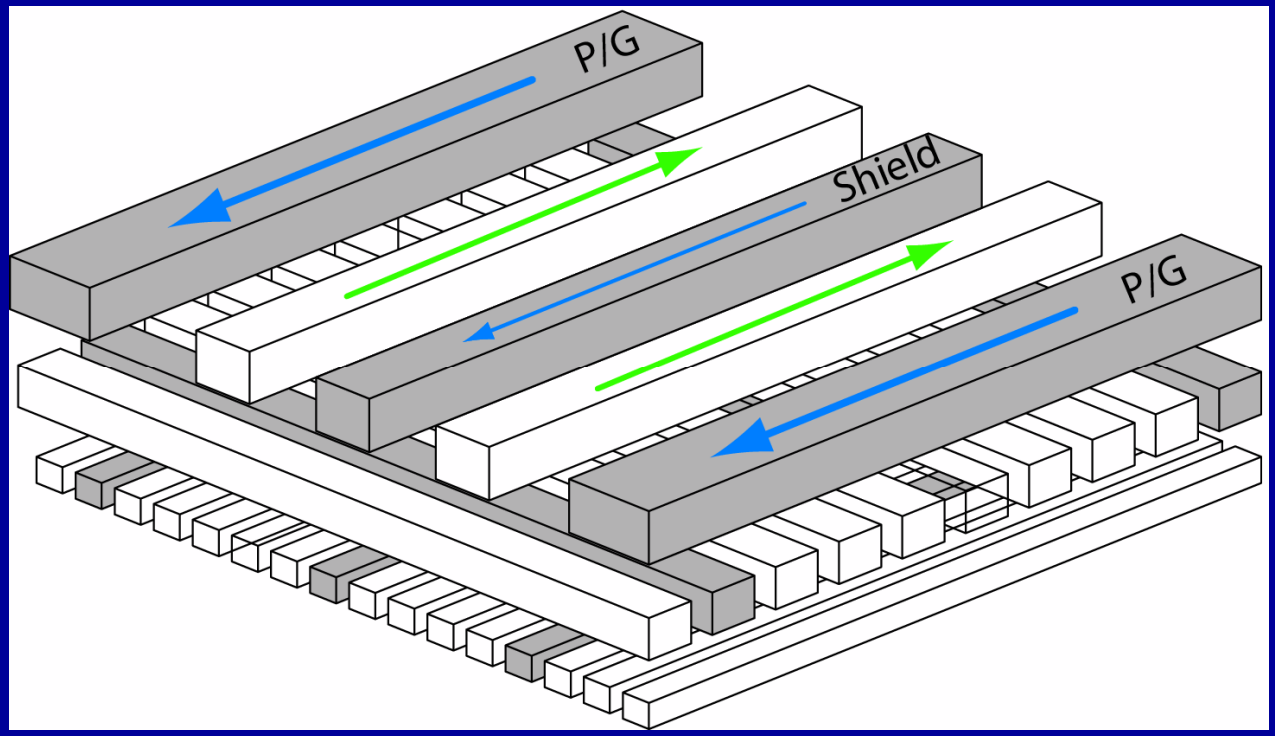


- Active regenerators - boosters
  - Support bi-directional wire behavior
  - No spacing constraints
  - High power dissipation



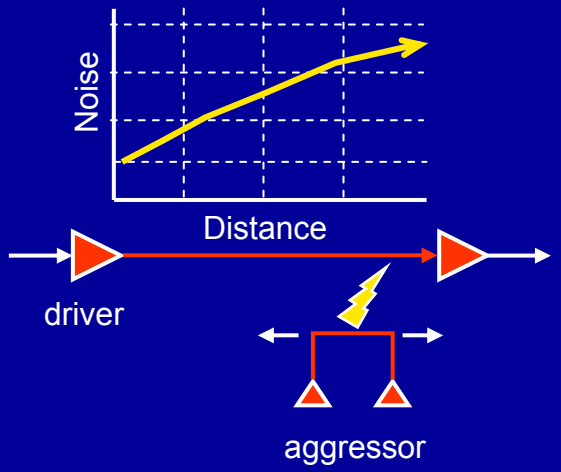
Shield Line Insertion for Coupled *RLC* Interconnects

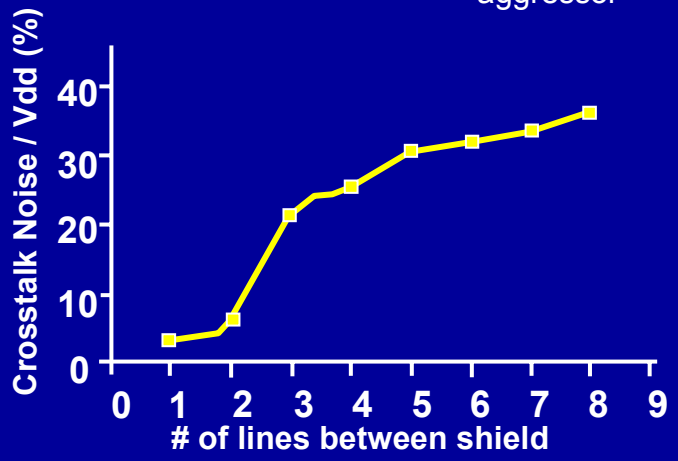
- Shield line insertion can also control inductive crosstalk noise
  - In addition to reducing capacitive crosstalk noise
  - Provides a current return path through the shield line for both the aggressor and victim lines



Shielding Efficiency

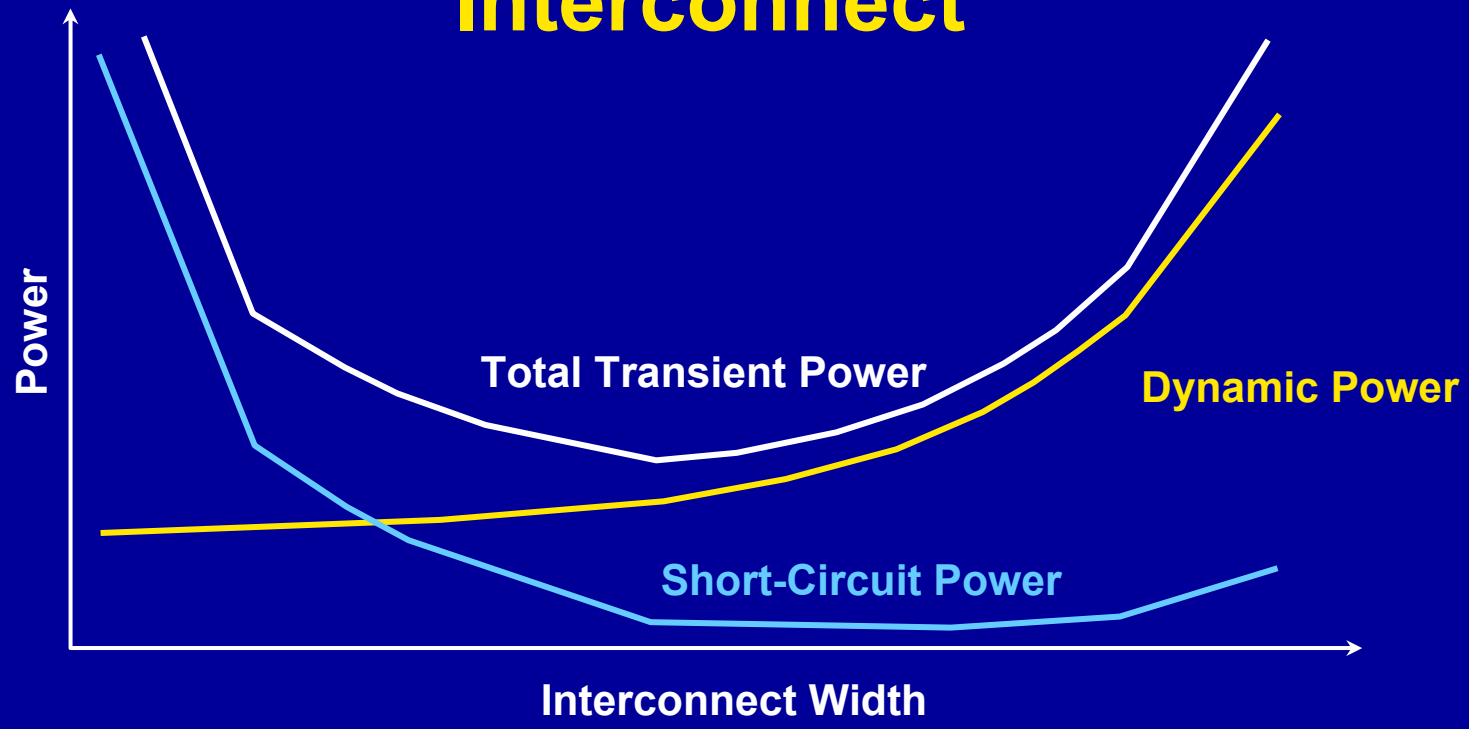
- Achieves a target reduction in noise
  - Uses minimal metal line resources
- Shielding close to the driver may be redundant
  - When crosstalk occurs farther from the wire driver
  - Peak noise increases
- Shielding line density
  - Tradeoff between
    - Noise reduction
    - Wire routing area





* X. Huang *et al.* "RLC Signal Integrity Analysis of High-Speed Global Interconnects," *IEEE International Electron Devices Meeting*, 2000

Transient Power Tradeoff in Inductive Interconnect



- Dynamic power increases with line width
- Short-circuit power may decrease in underdamped highly inductive lines
- An optimum interconnect width exists
  - Minimum transient power

* M. A. El-Moursy and E. G. Friedman, “Power Characteristics of Inductive Interconnect," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* Vol. 12, No. 10, pp. 1295-1306, December 2004

Research Problems

- Develop methodologies to characterize interconnect impedances
- Co-design interconnect drivers with wires
  - Optimize
    - Signal delay
    - Signal transition time
    - Reduction in power dissipation
- Unusual physical structures
  - Next generation packaging – chip interfaces
  - 3 - D architectures

Agenda

- A historical perspective
- Where are we now ?

⇒ Where are we going ?

- Conclusions

Agenda

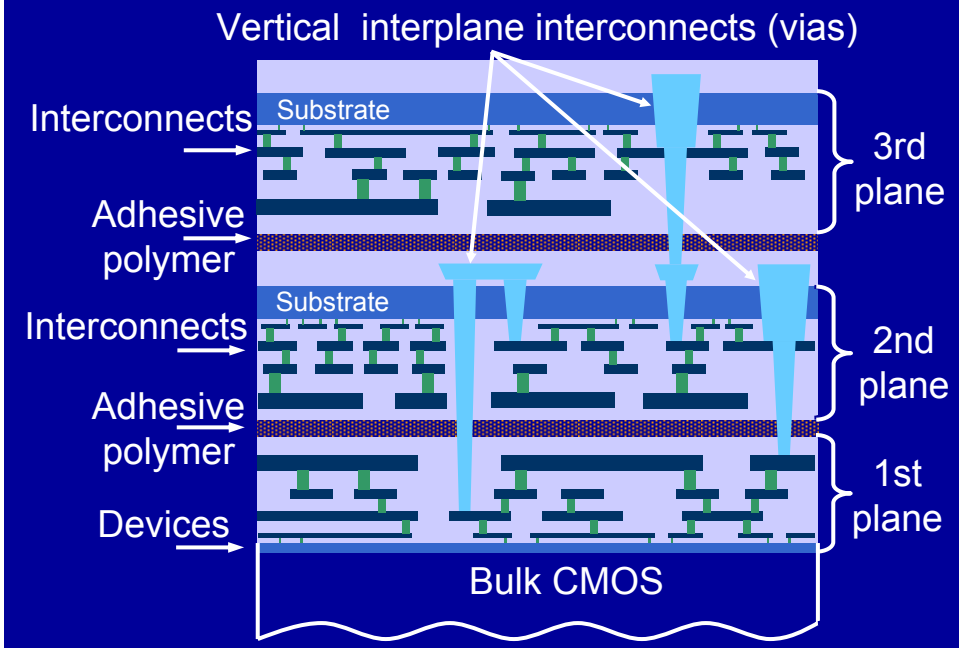
- A historical perspective
- Where are we now ?

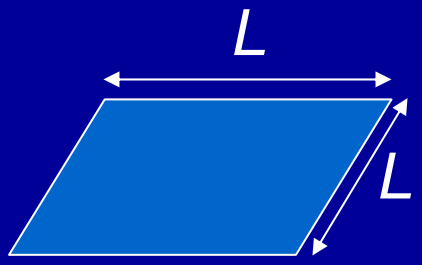
⇒ Where are we going ?

  - One possible solution ➔ 3 – D integration

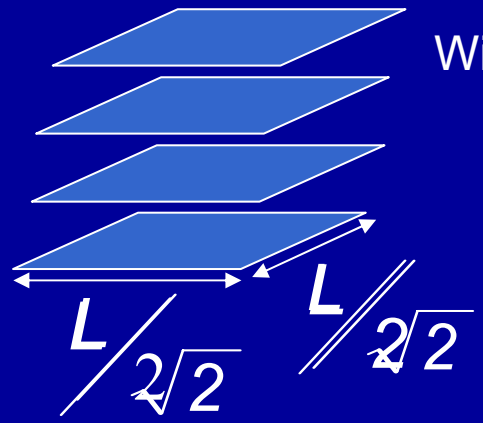
- Conclusions

Three-Dimensional Integration





Area = L^2, Corner to corner distance = $2L$



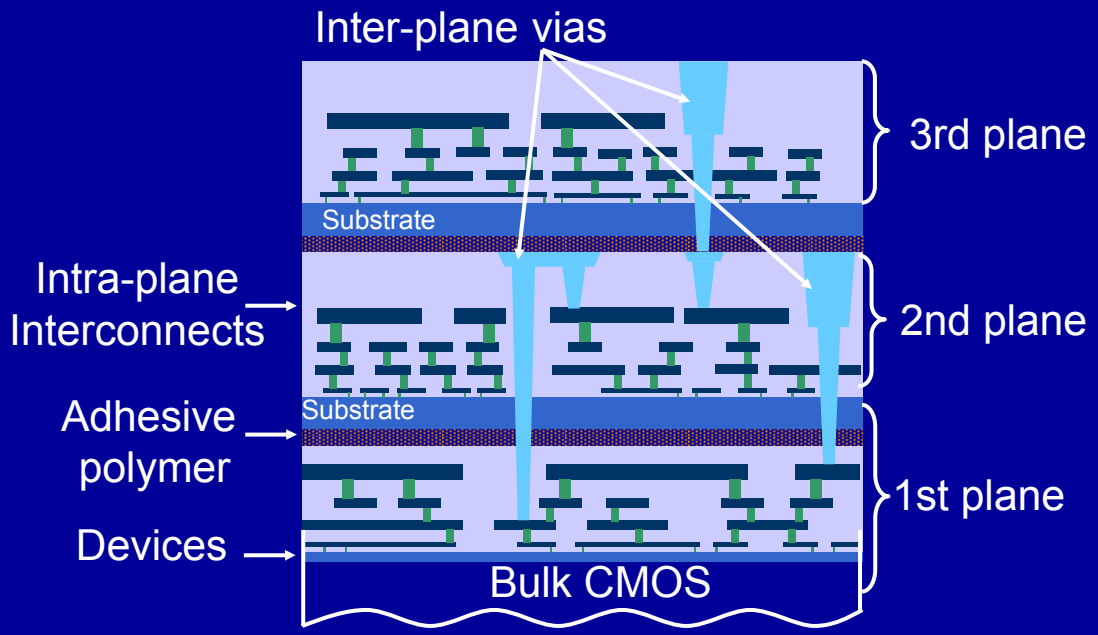
Wirelength reduction
2 planes ~ 30%
4 planes ~ 50%

Area = L^2, Corner to corner distance ≈ $1.2L$

*R. J. Gutmann *et al*., "Three-dimensional (3D) ICs: A Technology Platform for Integrated Systems and Opportunities for New Polymeric Adhesives," *Proceedings of the Conference on Polymers and Adhesives in Microelectronics and Photonics*, pp. 173-180, October 2001.

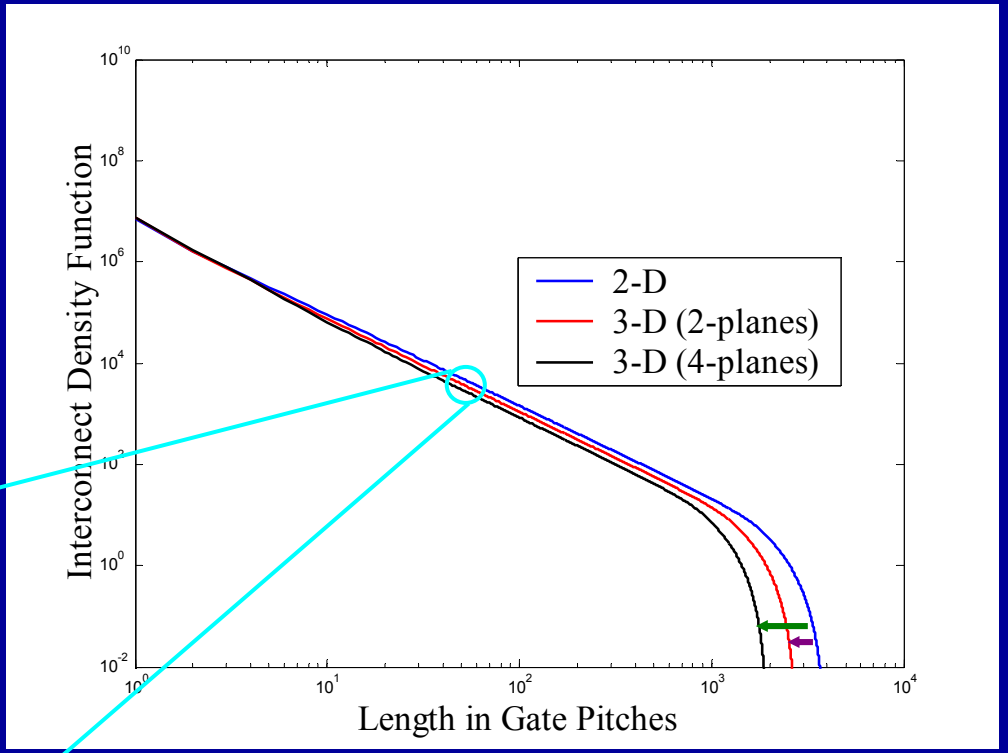
Three-Dimensional Integrated Circuits

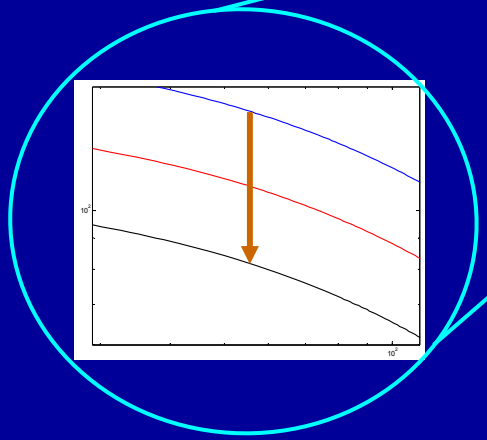
- Pros
  - Reduced interconnect delay
  - Disparate technologies
    - Non-silicon
  - No yield compromise
- Cons
  - Difficult to model interconnects
  - Crosstalk noise
  - Inter-plane via density
  - Thermal density



Three-dimensional Integration Pros

- Offers greater performance
- Substrate isolation between analog and digital circuits



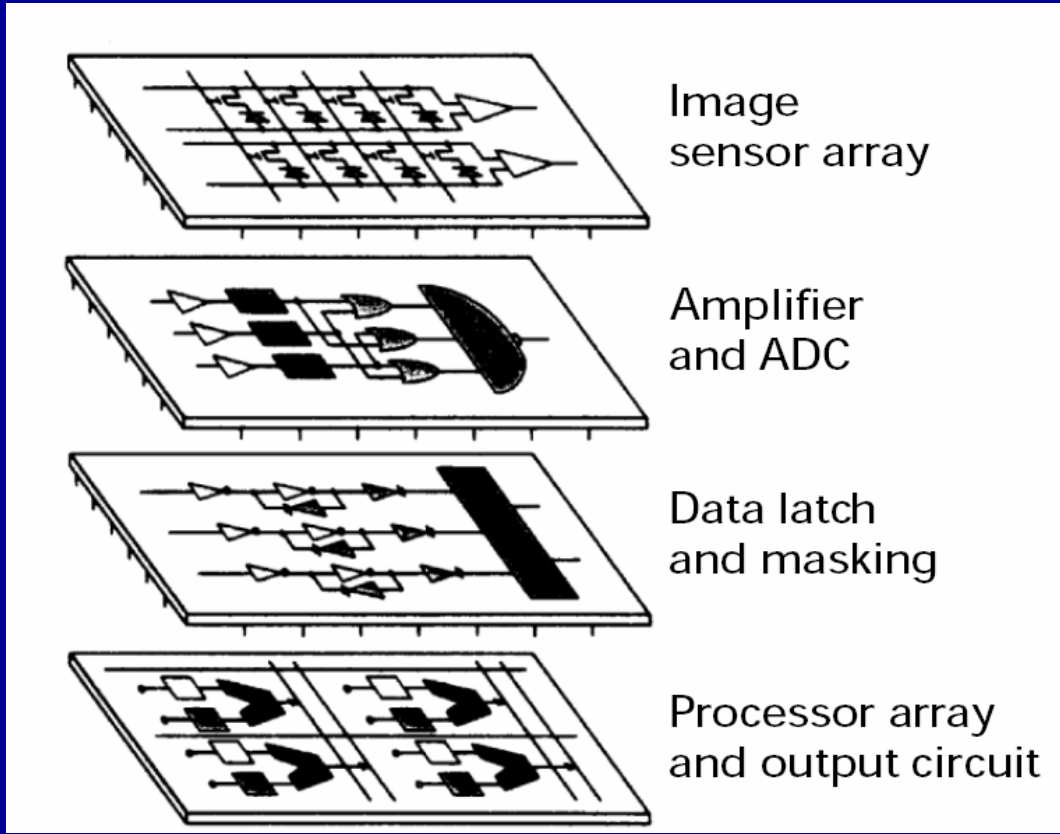


- Reduction of the length of the longest global interconnects
- Decrease in the number of the global interconnects
  - The number of local interconnects increases

*J. Joyner *et al.*, "Impact of Three-Dimensional Architectures on Interconnects in Gigascale Integration," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* Vol. 9, No. 6, pp. 922-927, December 2001

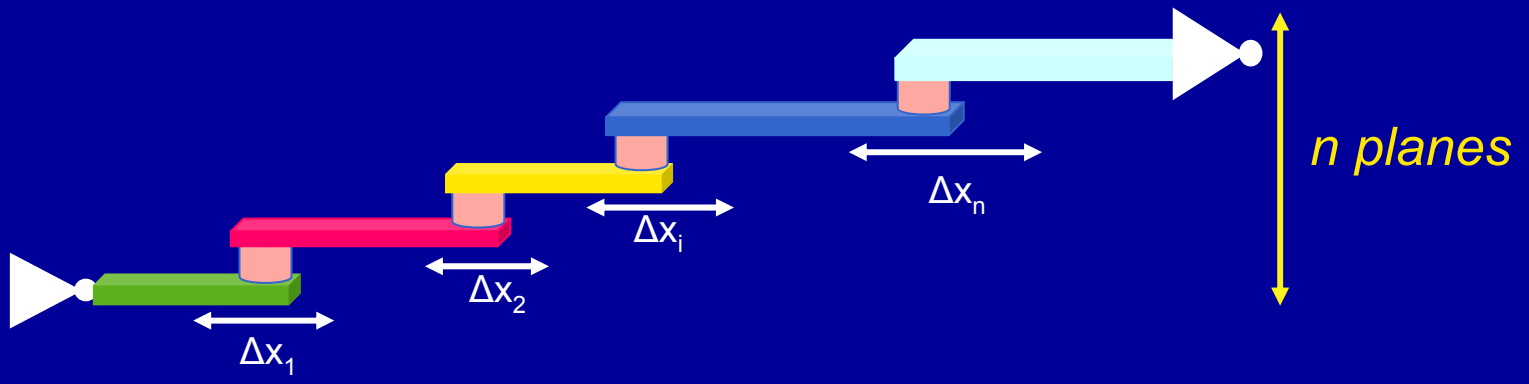
Novel System-on-Chip Designs

- Integrating
  - Circuits from different fabrication processes
  - Non-silicon technologies
  - Non-electrical systems



* M. Koyanagi *et al.*, "Future System-on-Silicon LSI Chips," *IEEE Micro*, Vol. 18, No. 4, pp. 17-22, July/August 1998

Delay and Power Improvement from Optimum Via Placement



- Δx_i's: available region for via placement between the *i* and *i+1* physical plane
- Average delay improvement is 6% for $n = 4$ and 18% for $n = 5$
  - As compared to equally spaced interconnect segments
- Reduced power dissipation
  - Due to decrease in interconnect capacitance
    - 5% to 6% lower power
      - Independent of the number of planes

* V. Pavlidis and E. G. Friedman, "Via Placement for Minimum Interconnect Delay in Three-Dimensional (3-D) Circuits," *Proceedings of the International Symposium on Circuits and Systems,* May 2006.

Research Problems in 3-D Interconnect Design

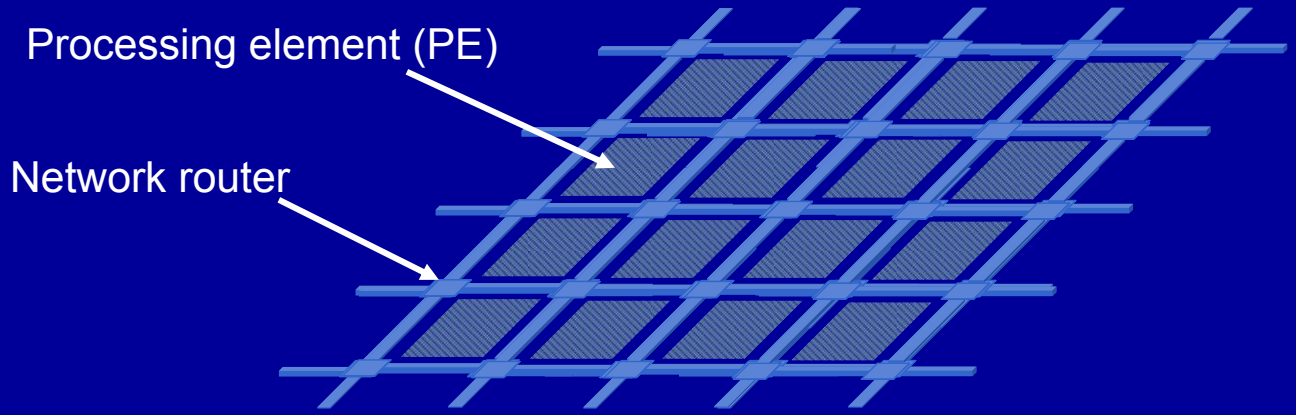
- Interplane via placement to efficiently minimize delay
  - Interconnect tree structures across multiple planes
- Design expressions which consider inductance
- Optimum via location for lines with repeaters
- Clock distribution networks for 3 – D ICs
- Power distribution structures for 3 – D ICs
- Satisfy heating constraints without compromising performance
- 3 – D NoC

Agenda

- A historical perspective
- Where are we now ?
- ⇒ Where are we going ?
  - One possible solution ➔ 3 – D integration
  - Another solution ➔ Networks on Chip
- Conclusions

Network on Chip

- Pros
  - Canonical interconnect structure
  - Shared interconnect bandwidth
  - Increased flexibility
- Cons
  - Intra-PE interconnect delay
  - PE yield limitations
  - Constrained to CMOS



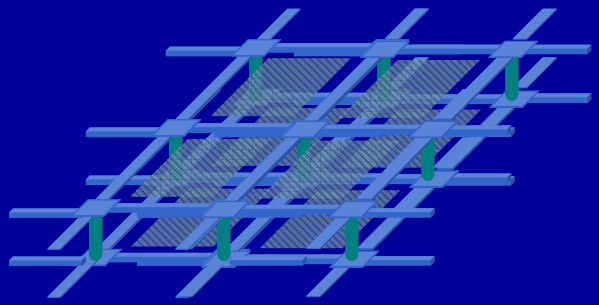
Agenda

- A historical perspective
- Where are we now ?
- ⇒ Where are we going ?
  - One possible solution ➔ 3 – D integration
  - Another solution ➔ Networks on Chip
  - Why not 3 – D NoC ?
- Conclusions

3-D NoCs

- Pros
  - Reduced interconnect delay
  - Canonical interconnect structure
  - Integration of dissimilar systems and technologies
  - No yield limitations
  - Flexibility

- Design Methodologies

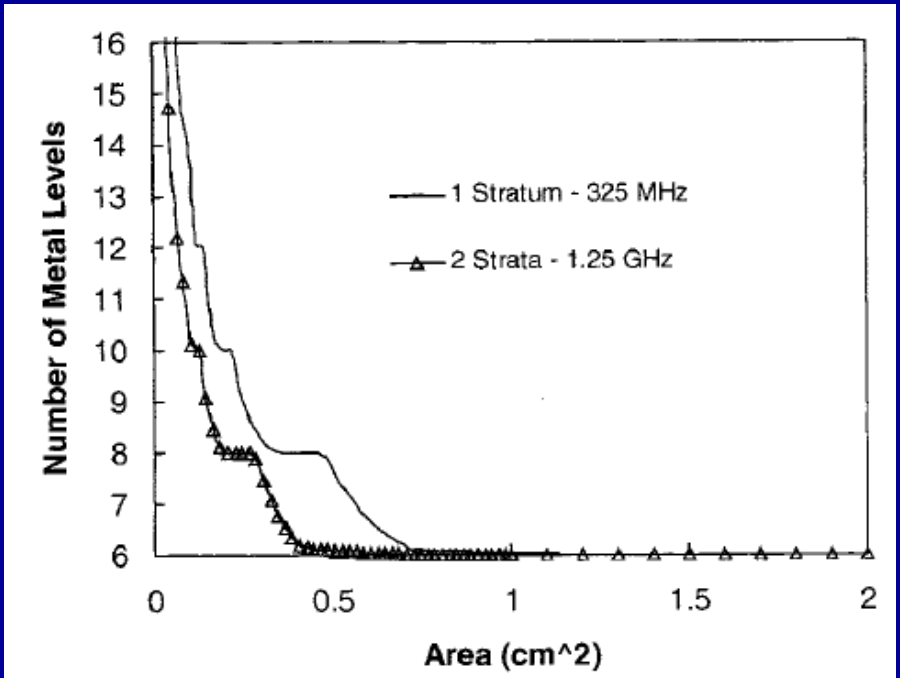


3 – D NoC

Three Dimensional IC – NoC Design

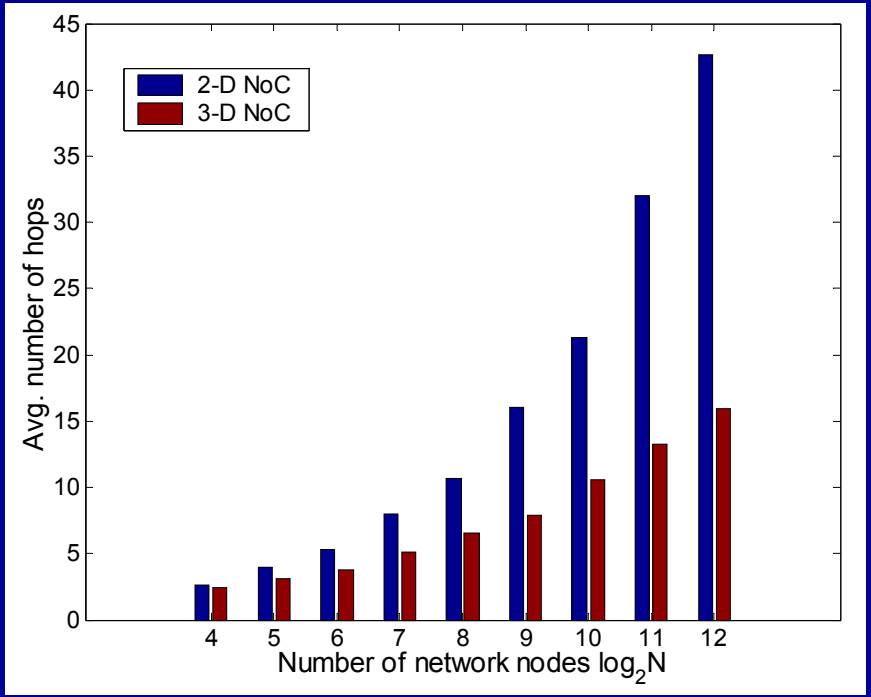
3-D IC

- Reduced interconnect delay



3-D NoC

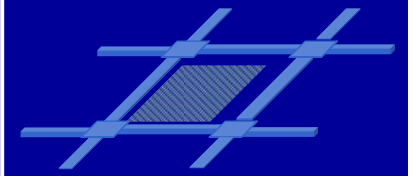
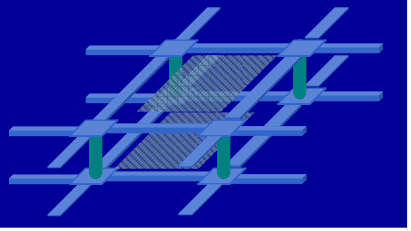
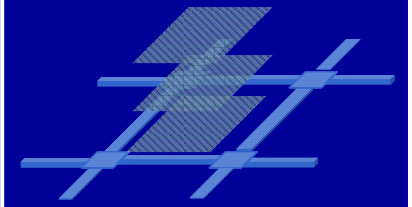
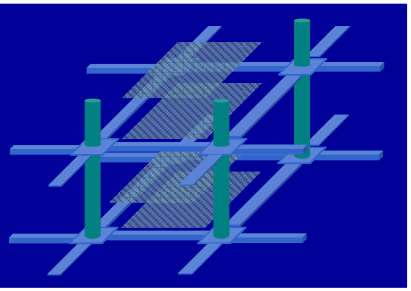
- Decreased number of hops



*J. Joyner, et al., "Impact of Three-Dimensional Architectures on Interconnects in Gigascale Integration," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* Vol. 9, No. 6, pp. 922-927, December 2001

- Maximum number of physical planes = 16

Various Topologies for 3-D Mesh IC - NoC

IC \ NOC	2-D	3-D
2-D		
3-D		

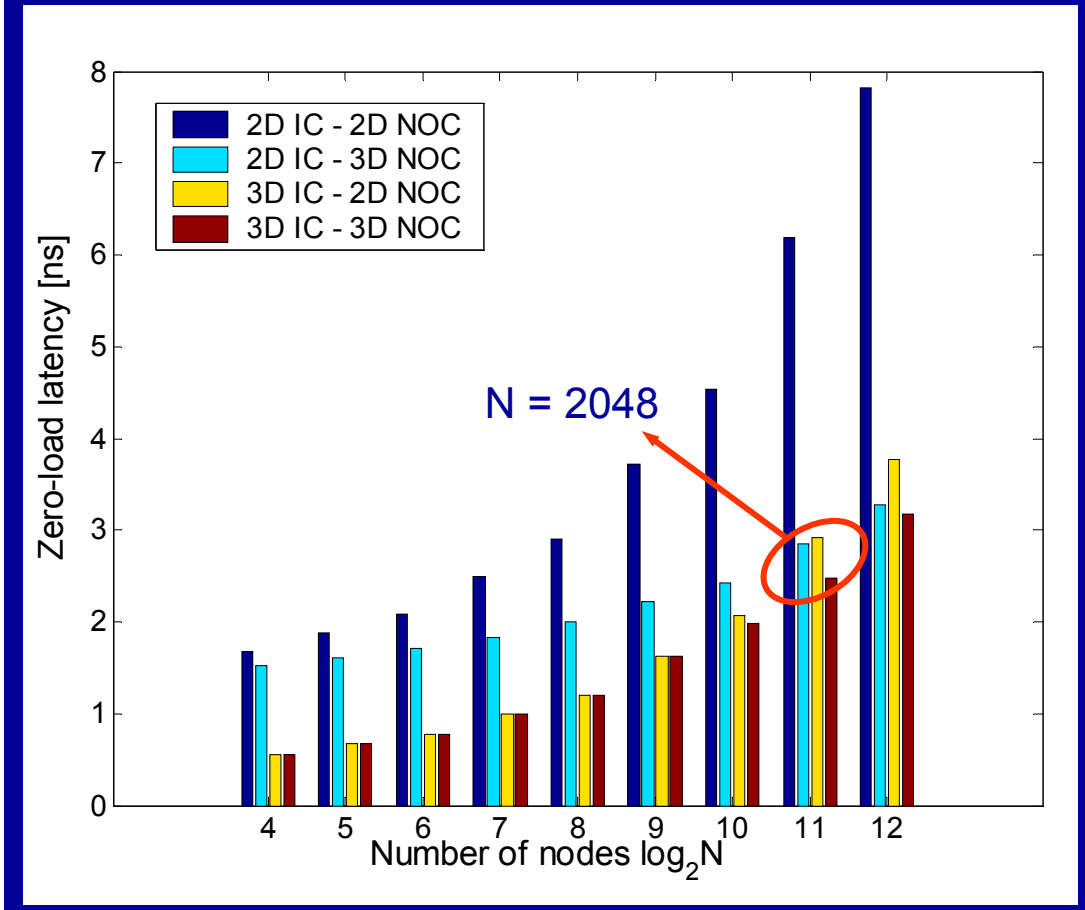
Reduced number of hops

Best of both Number of hops and Channel length

Shorter channel length

- Smaller number of hops
- 2-D network topology, same number of hops
- Asymmetric channel length
- Additional PE planes lead to reduced channel length
  - Vertical channel ➔ Short interplane vias
  - Horizontal channel ➔ Sidelength of the PE

Zero-Load Latency for 3-D NoC

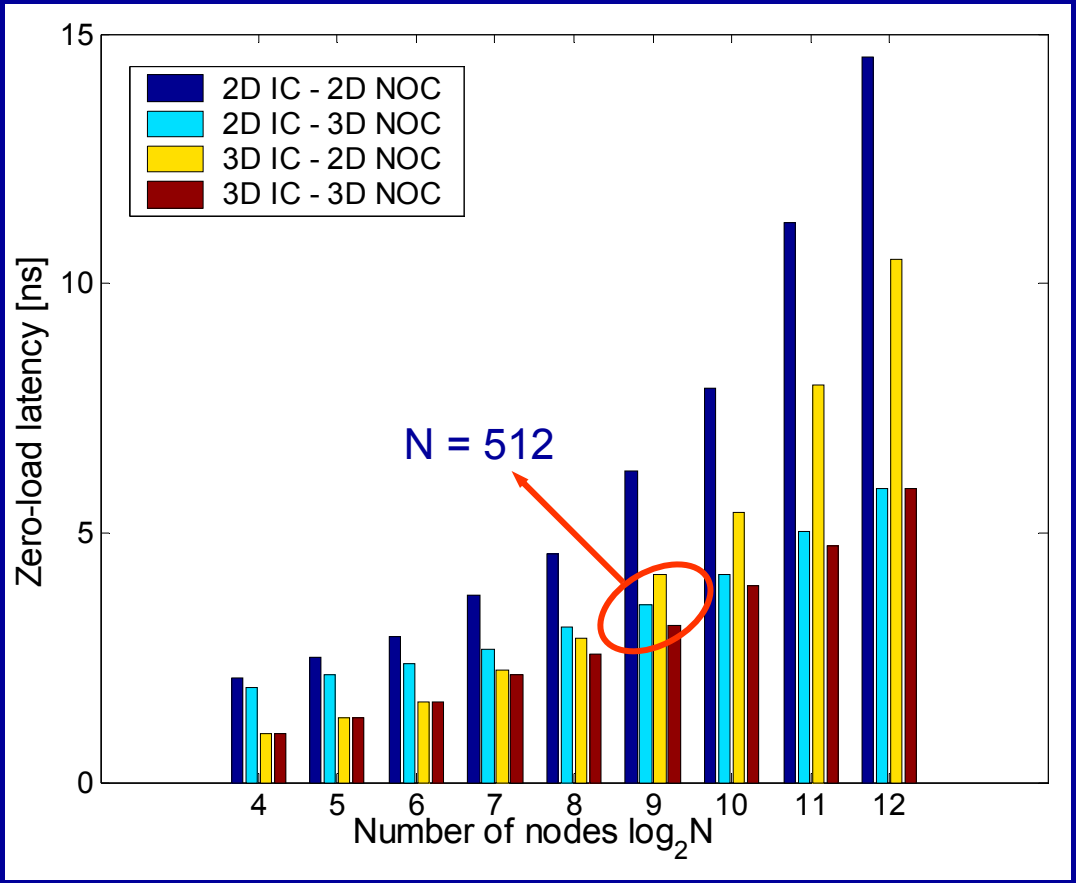


- $t_r = 0.5\ t_c$
- L_h = 1 mm
- L_v = 20 μm
- Wormhole routing
- Uniform traffic
- Max. number of planes = 16
- Packet size = 10 flits

- 2D IC – 3D NoC
  - Latency decreases due to hop reduction
- 3D IC – 2D NoC
  - Latency decreases due to channel length reduction

- For small N
  - Latency reduction due to hop decrease is smaller than that due to channel length reduction
  - 3D IC – 2D NoC outperforms 2D IC – 3D NoC
- For large N
  - Latency reduction due to hop decrease is greater than that due to channel length reduction
  - 2D IC – 3D NoC outperforms 3D IC – 2D NoC

Zero-Load Latency for 3-D NoC

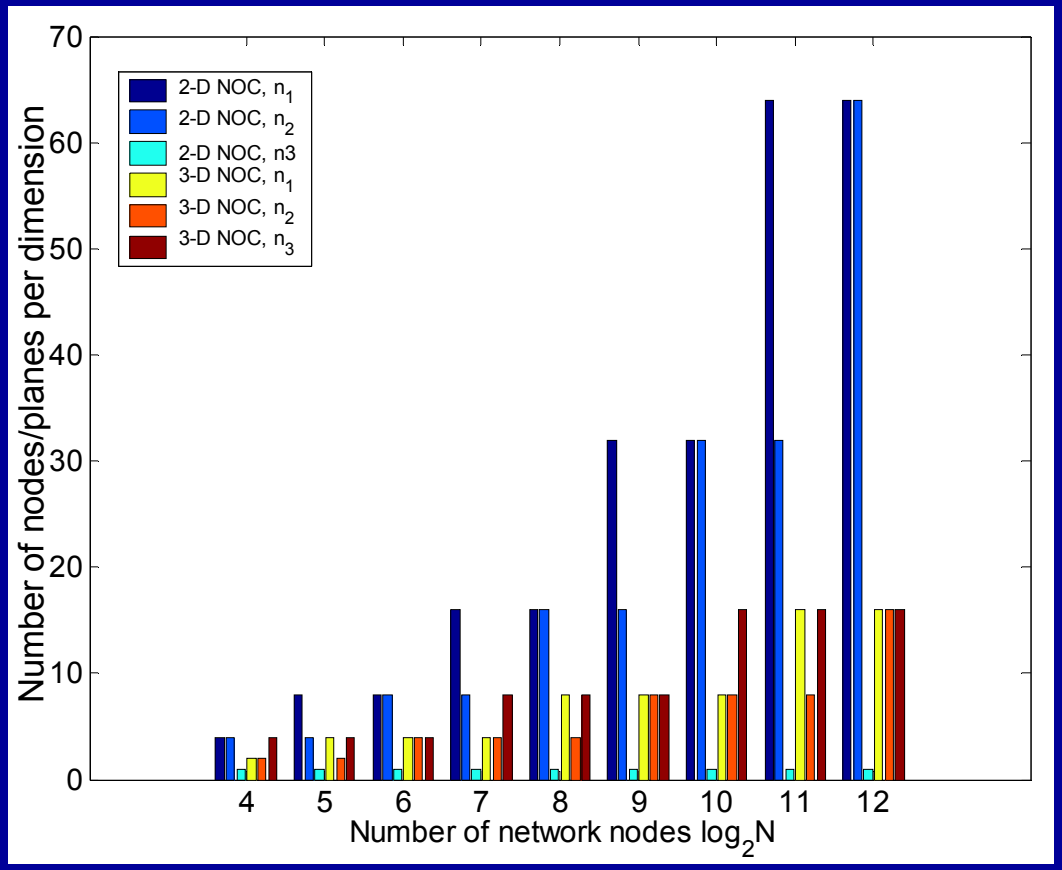


- $t_r = 2\ t_c$
- L_h = 1 mm
- L_v = 20 µm
- Wormhole routing
- Uniform traffic
- Max. number of planes = 16
- Packet size = 10 flits

- t_r / t_c determines at which *N*, 2D IC – 3D NoC outperforms 3D IC – 2D NoC
  - $t_r = 0.5\ t_c$, N = 2048
  - $t_r = 2\ t_c$, N = 512

- Optimum topology 3D IC – 3D NoC
  - Available physical planes of the stack are optimally distributed to
    - Reduce number of hops
    - Reduce communication channel length

Optimum Number of Nodes per Dimension for 3-D NoC



- $t_r = 2\ t_c$
- $L_h = 1$ mm
- $L_v = 20$ µm
- Wormhole routing
- Uniform traffic
- Maximum number of planes = 16
- Packet size = 5 flits

- Optimum number of nodes per dimension
  - As close as possible to $N^{1/3}$
    - A cube

- Infeasible for some network sizes
  - Due to discrete nature of the variables
  - n_3 should be greater than $n_{1,2}$
    - Greater n_3 exploits smaller communication channel delay

Agenda

- A historical perspective
- Where are we now ?
- Where are we going ?

⇒ Conclusions

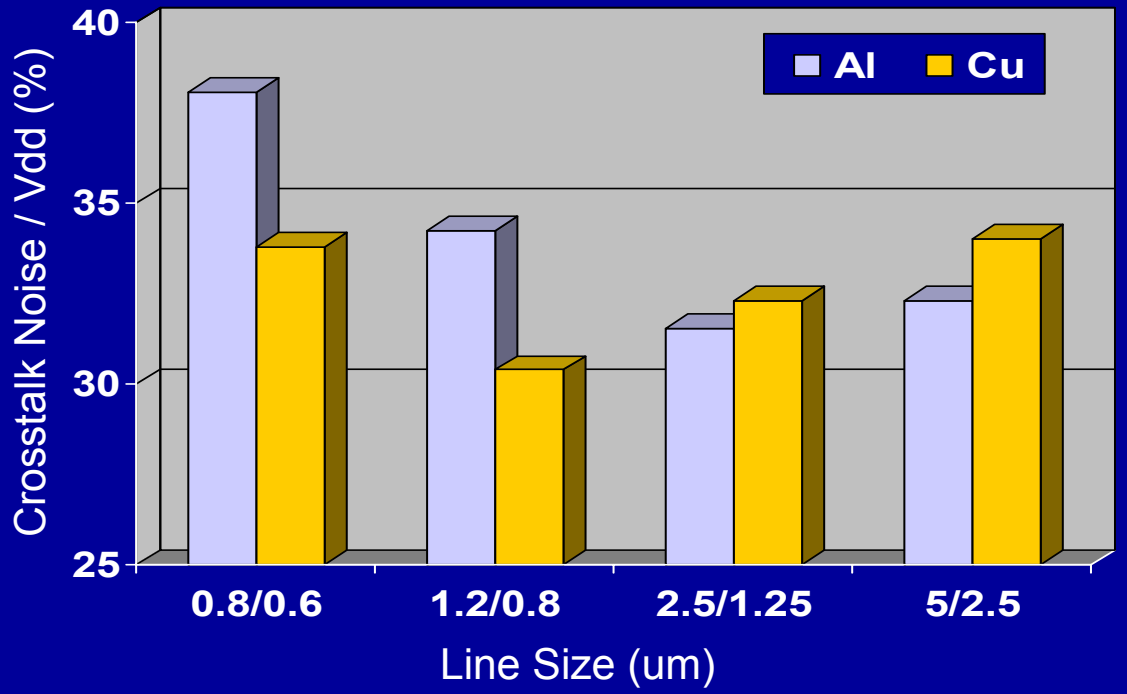
Conclusions

- The trends are clear
  - Interconnect now dominates the design process
- Different aspects of the interconnect design process
  - Figures of merit
  - Extraction
  - Modeling and simulation
  - Interconnect-aware design methodologies
- 3 – D integration is coming
  - NoC is here and important for on-chip global data transfer
- 3 – D NoC
  - A natural evolution

Any Questions

Back-Up Viewgraphs

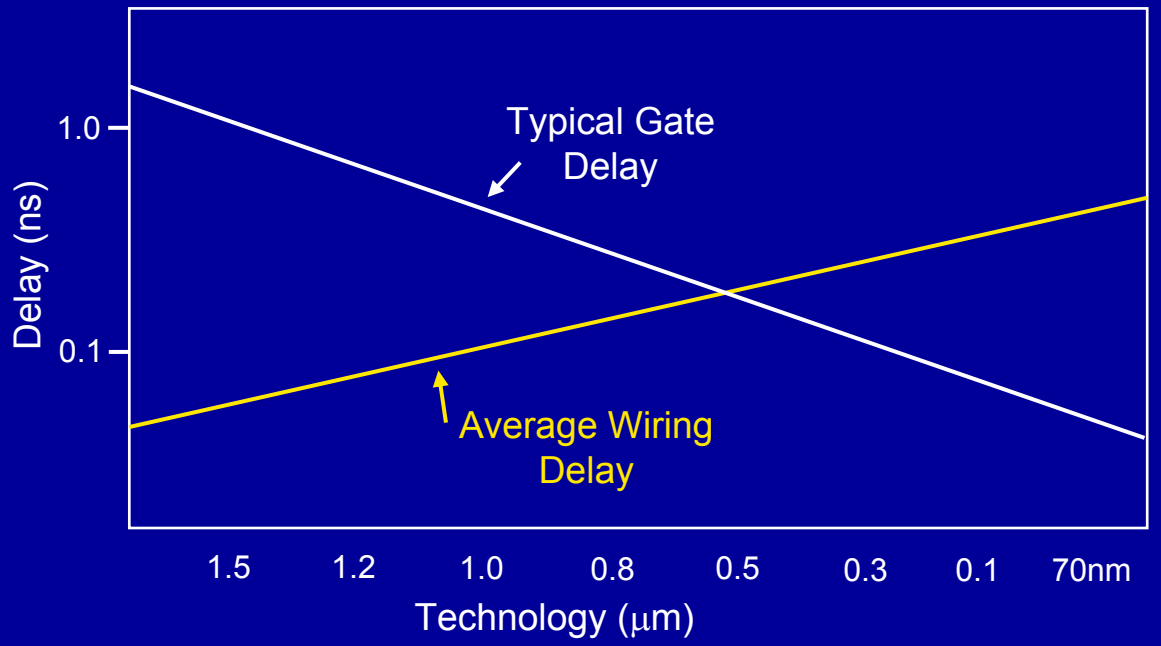
Aluminum vs Copper Characteristics



- Aluminum lines
  - Larger coupling capacitance
  - Larger coupling noise
- Copper lines
  - Lower resistance
  - Inductance effects are more significant in wider lines

* X. Huang *et al.,* "RLC Signal Integrity Analysis of High-Speed Global Interconnects," *IEEE International Electron Devices Meeting*, 2000

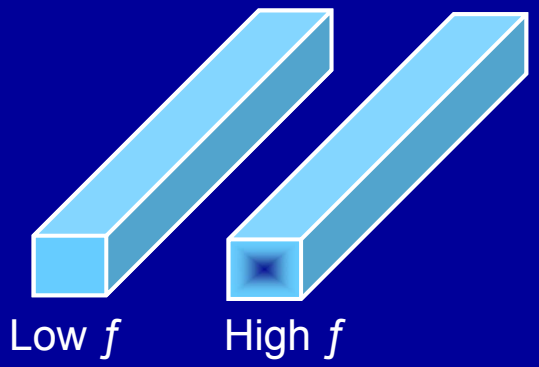
Importance of Interconnect



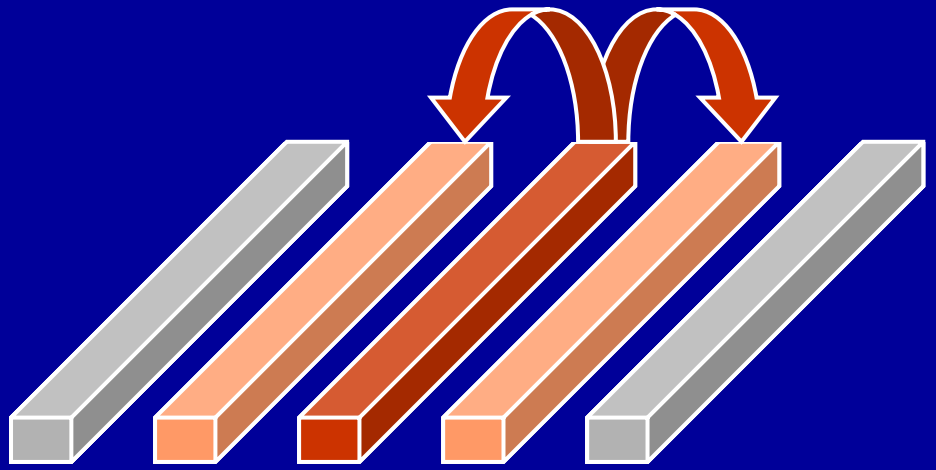
- With technology scaling
  - Gate delay decreases
  - Wire length increases
  - Wire cross sectional area decreases
- Wire delay increases polynomially with technology scaling

Dependence of Inductance on Frequency

- Skin Effect
  - Low frequency
    - Current flow is uniformly distributed within the wire
  - High frequency
    - Current flow concentrates at the wire surface

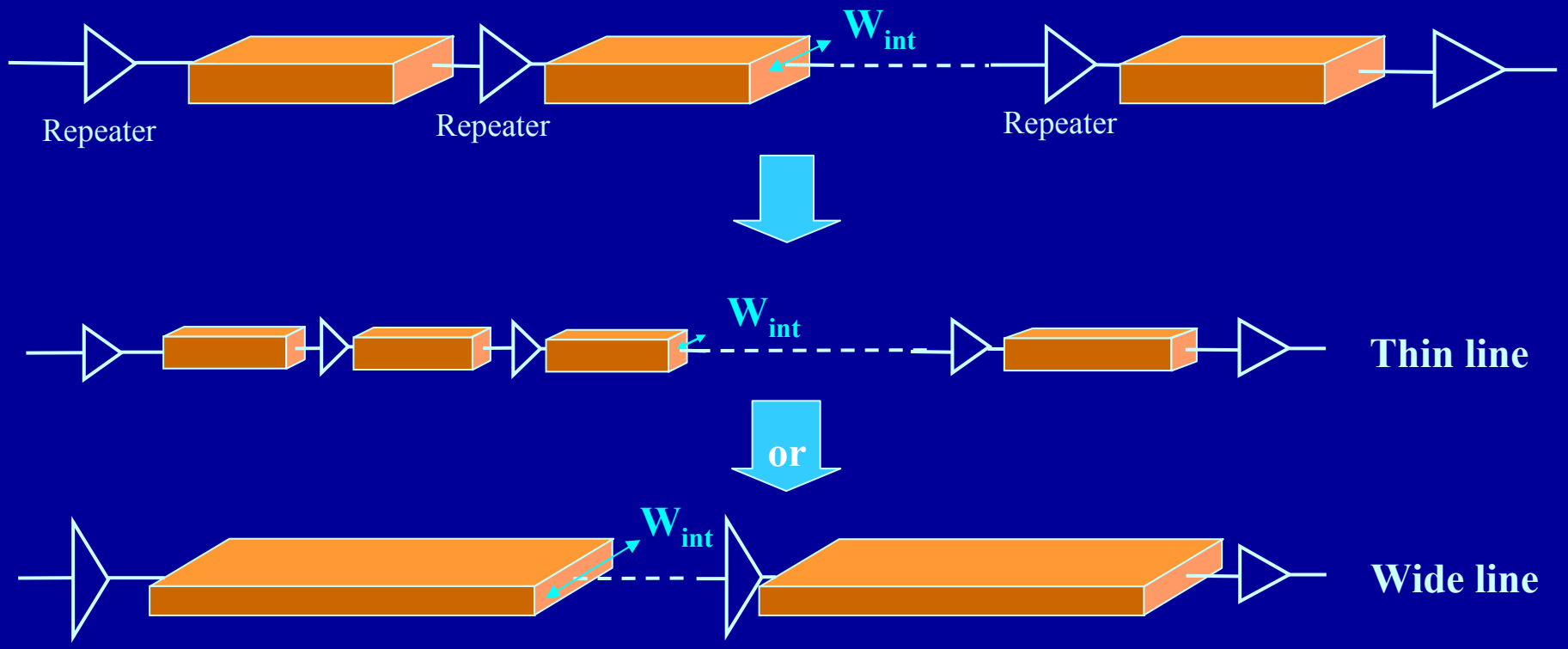


- Proximity effect
  - No effect at low frequency
  - High frequency
    - Return current concentrates along the edges

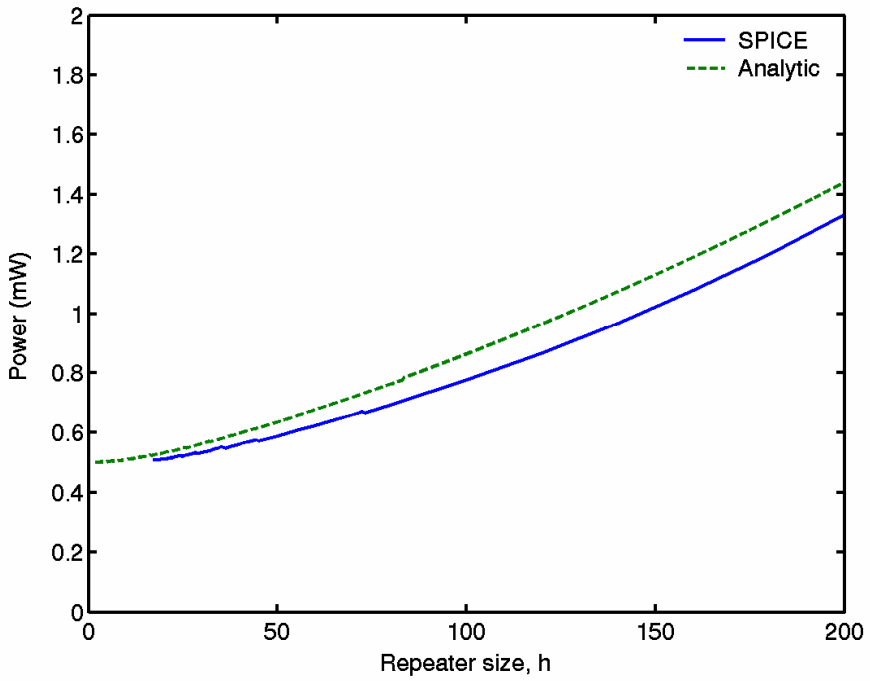


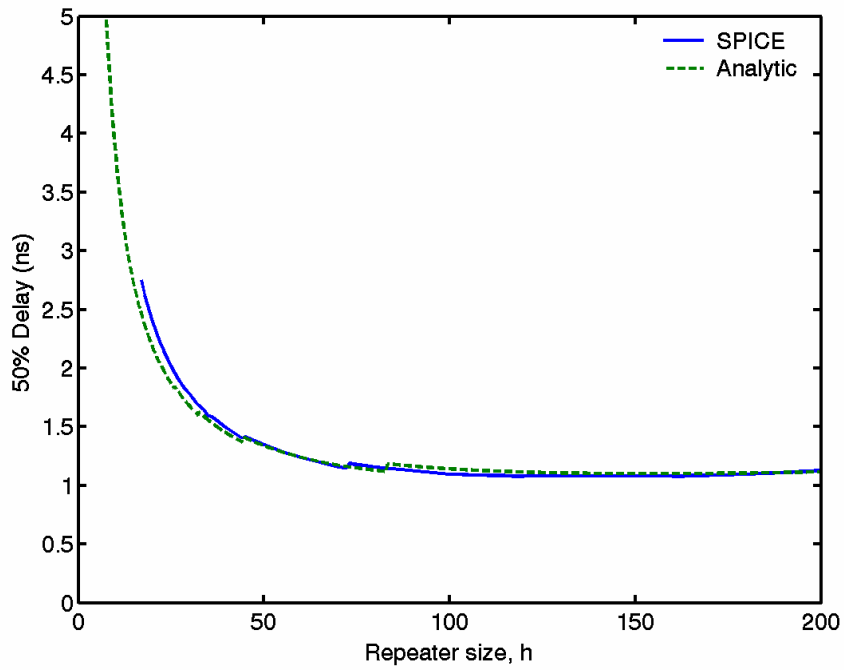
Effects of On-Chip Interconnect Inductance on Circuit Design Methodologies

- Optimum sizing of *RLC* line with repeaters



Minimum Interconnect Power with Bandwidth Constraints

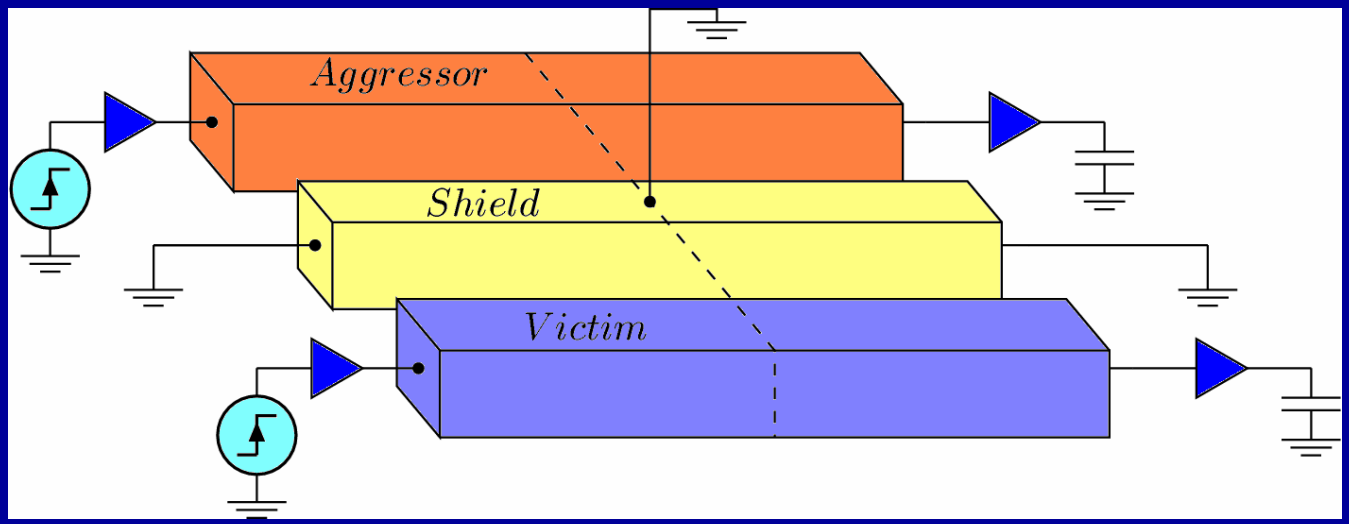




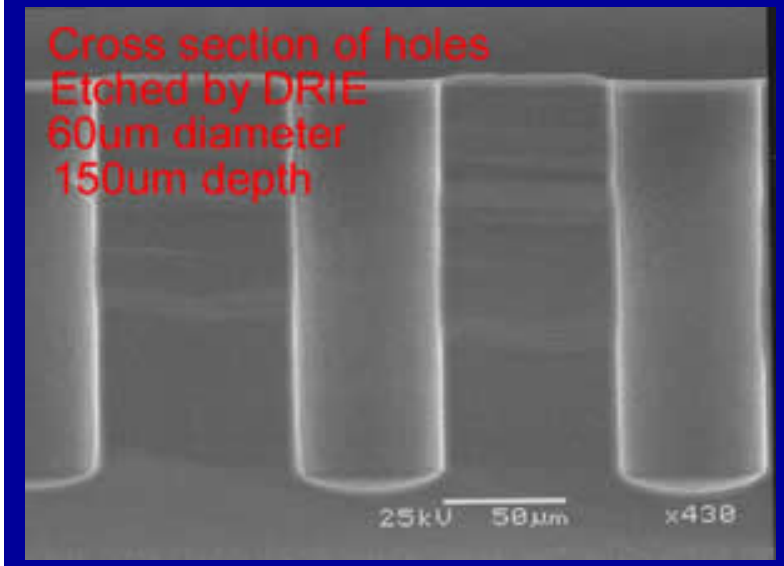
- Minimum power can be reached with minimum sized repeaters
  - Total delay and area, however, are unpractical large
- Multiple constraints should be considered

Current Research Plans

- Shielding for high speed interconnect
  - Develop crosstalk noise model and shielding techniques for coupled *RLC* interconnects
  - Develop shielding techniques to minimize power and delay under crosstalk noise constraints
  - Develop a shielding methodology to minimize delay and crosstalk noise

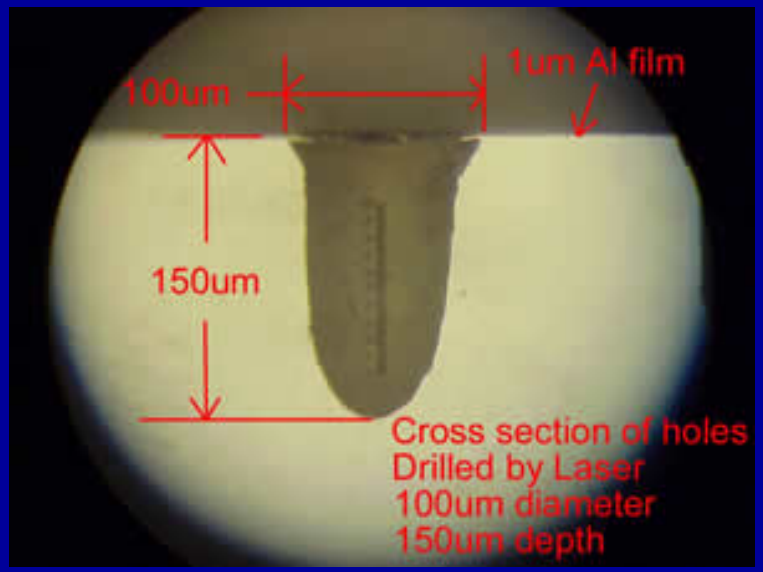


Via Drilling and Filling

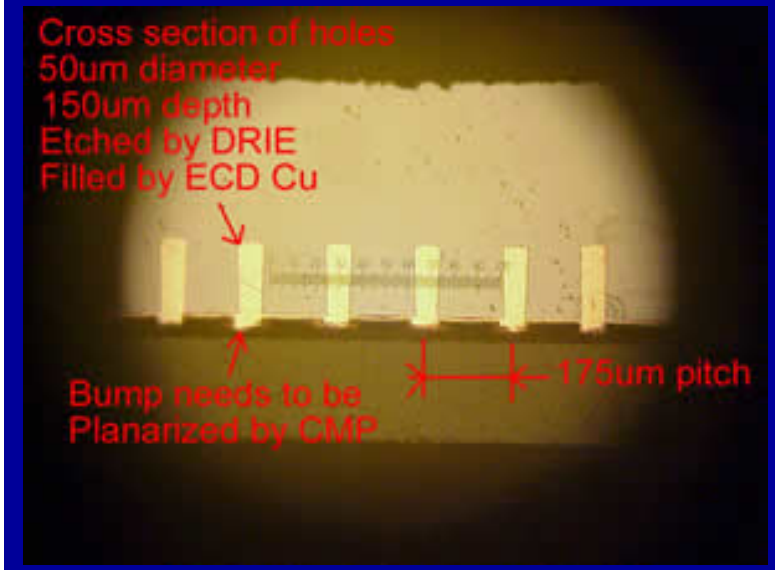


*Tru-Si Technologies, *www.trusi.com*

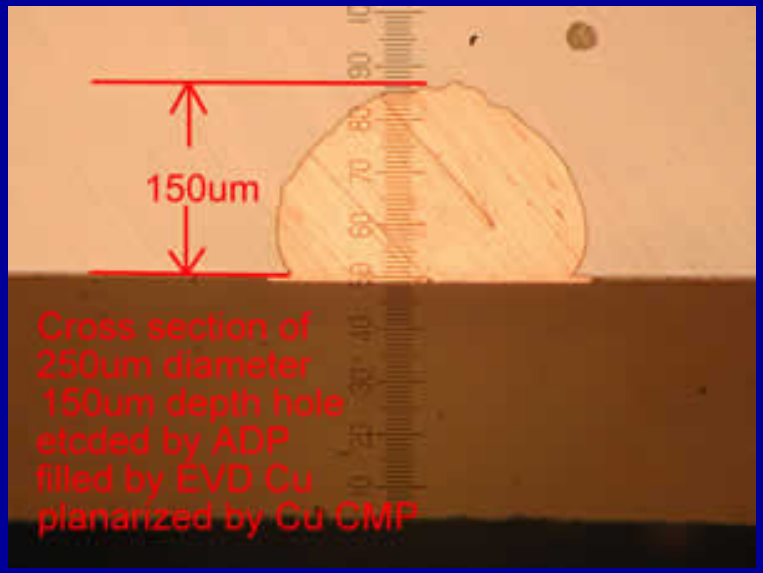
Drilling



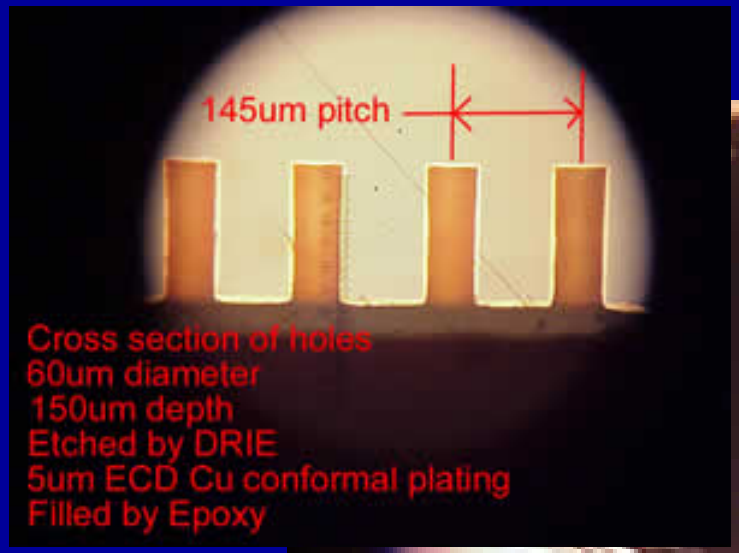
*Tru-Si Technologies, *www.trusi.com*



Filling

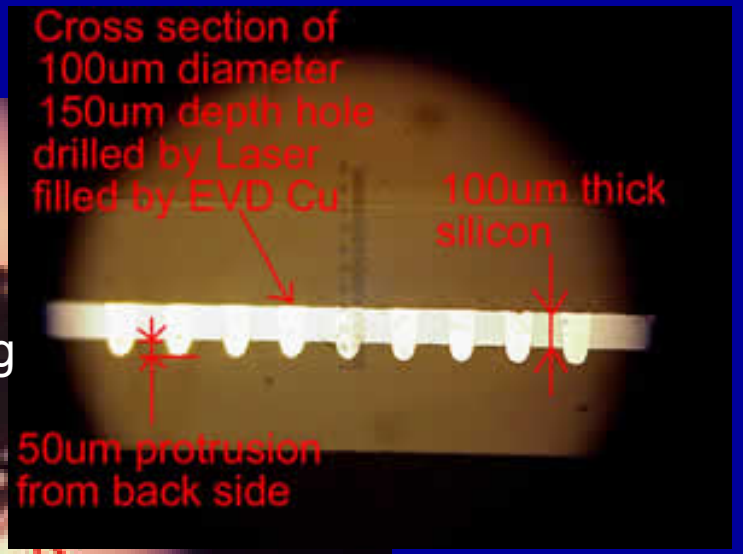


Grind, ADP Etching & Stacking



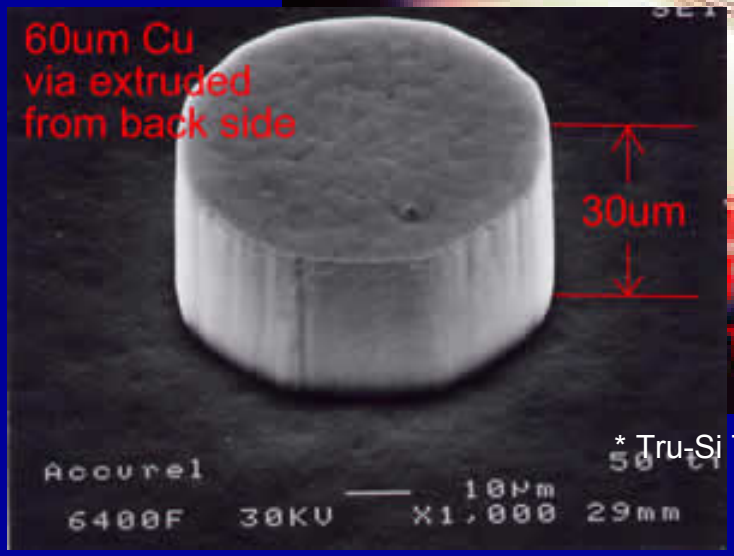
* Tru-Si Technologies, *www.trusi.com*

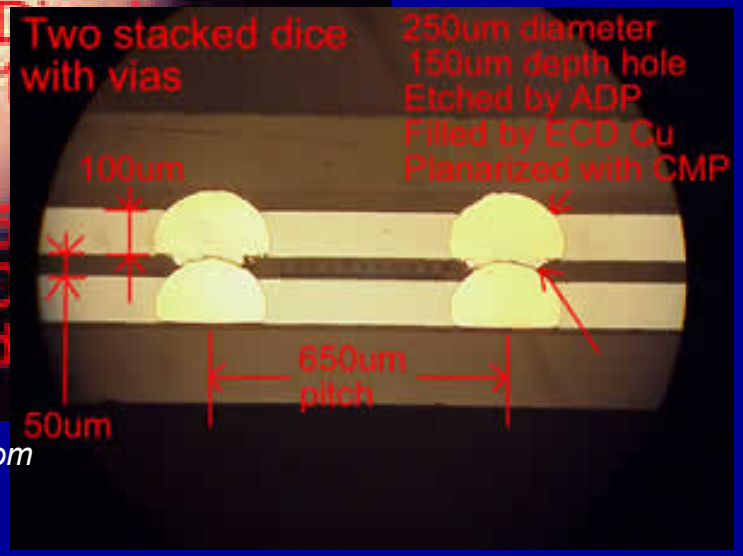
Grind and Etching



* Tru-Si Technologies, *www.trusi.com*

Stacking

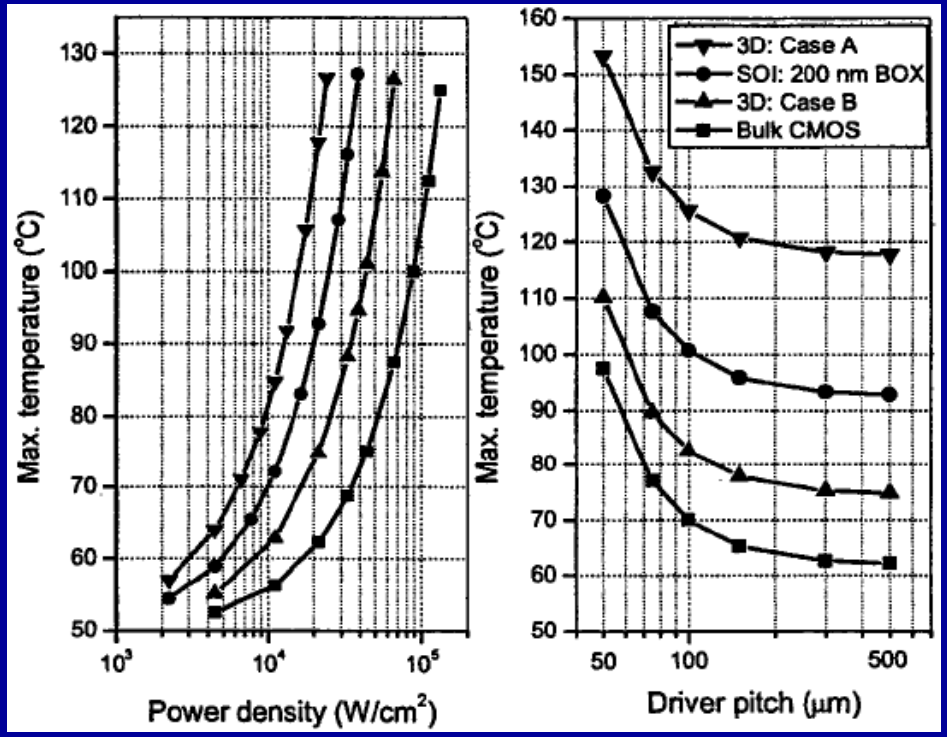




* Tru-Si Technologies, *www.trusi.com*

Existing Work in 3-D Integration

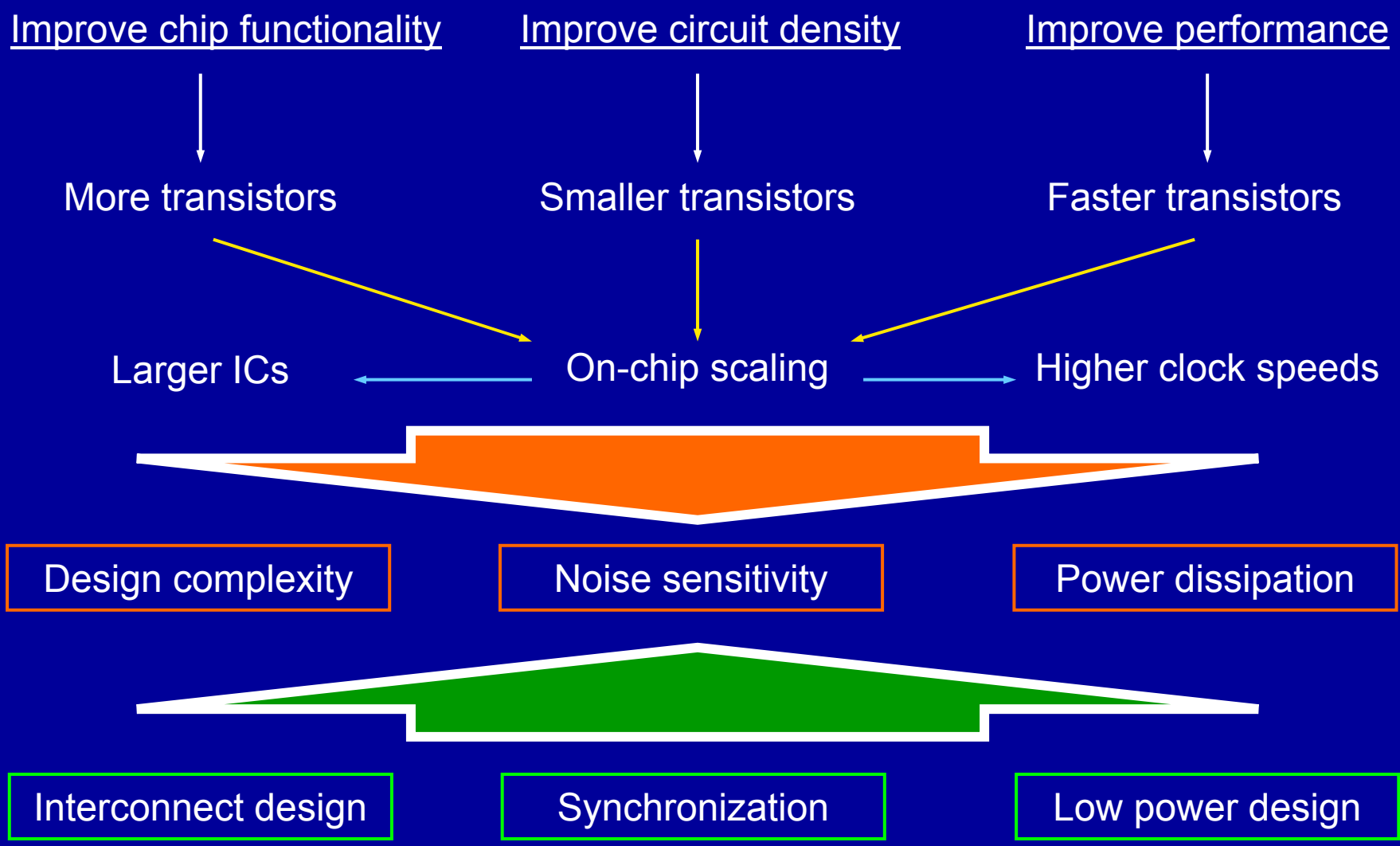
- Fabrication techniques
  - Copper wafer bonding
  - Adhesion with polymers
- Heating Effects and Constraints



* S. Tiwari et al., "Three-dimensional Integration in Silicon Electronics," *Proceedings of Lester Eastman Conference*, pp. 24-33, 2002

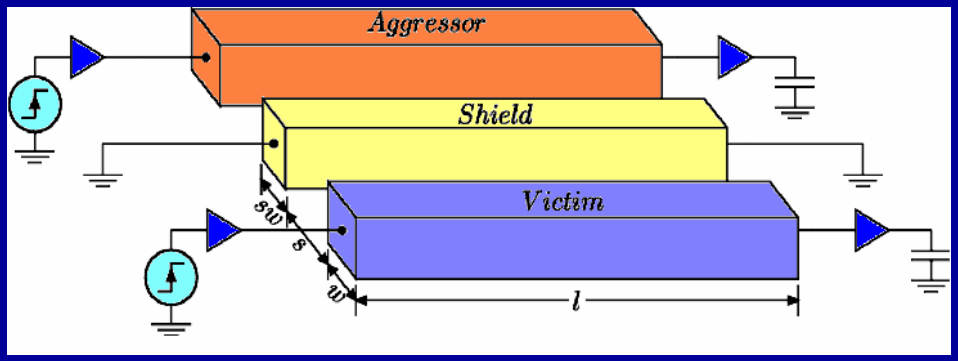
High Performance Digital IC Design Challenges

- Primary objectives

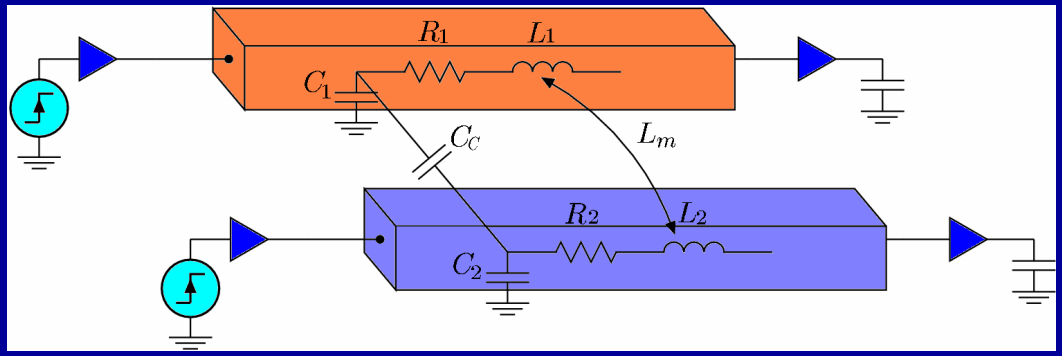


Shielding Design Methodologies for High Speed Interconnects

- Shielding design methodologies for coupled *RC* interconnects

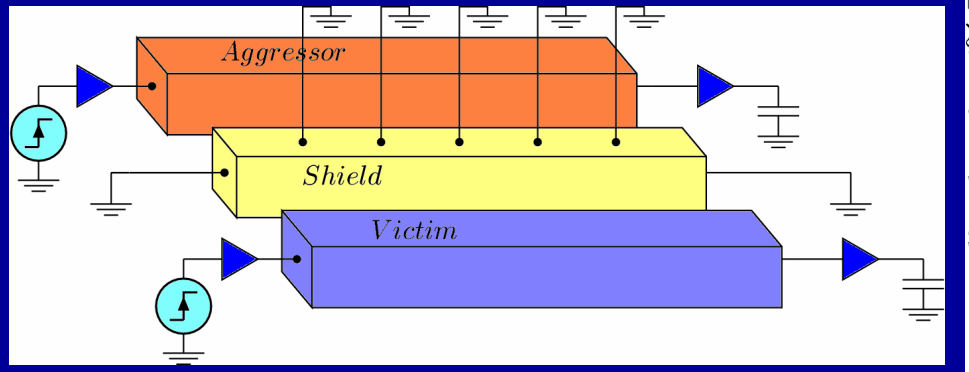


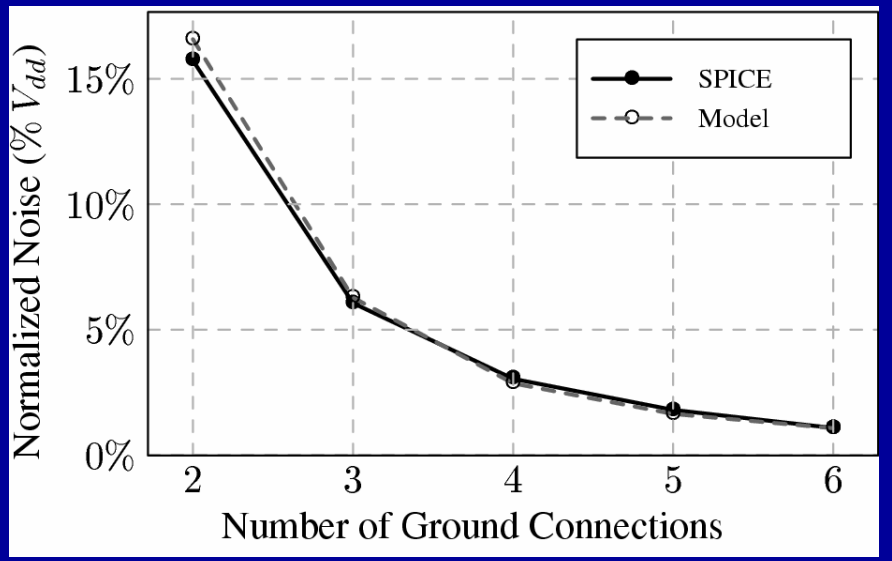
- Crosstalk analysis for coupled *RLC* interconnects



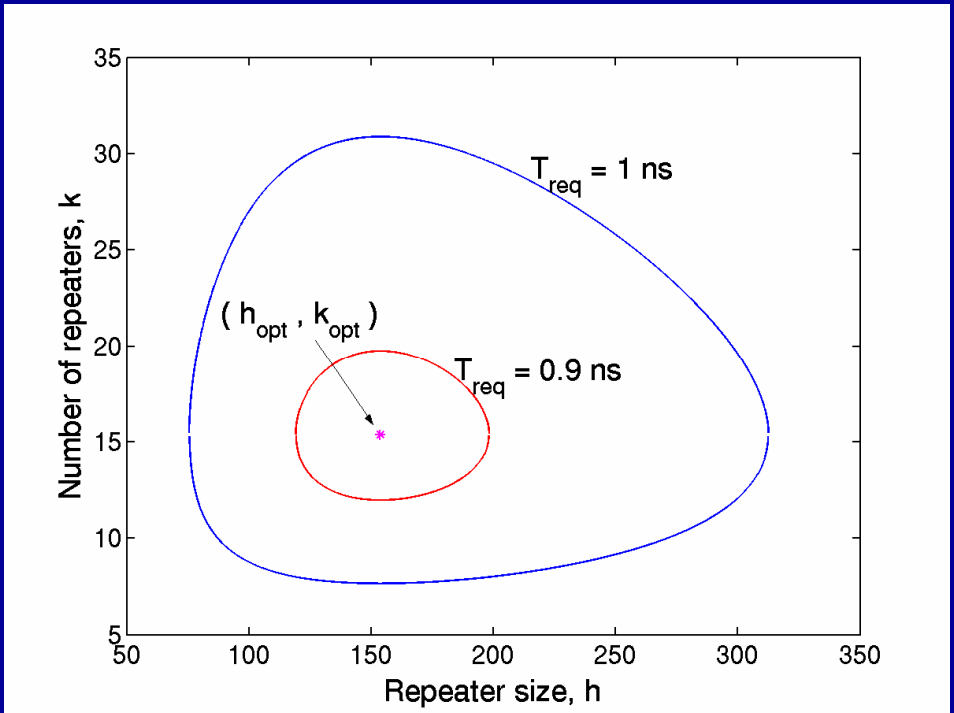
Effect of Ground Connections on Reducing Crosstalk

- The effect of the shield line on reducing crosstalk depends upon the number of ground connections tieing the shield line to the power/ground grid
- A shield line is not an ideal ground due to the interconnect resistance
  - Coupling noise to the signal line
- The more ground connections, the smaller the coupling noise on the victim signal





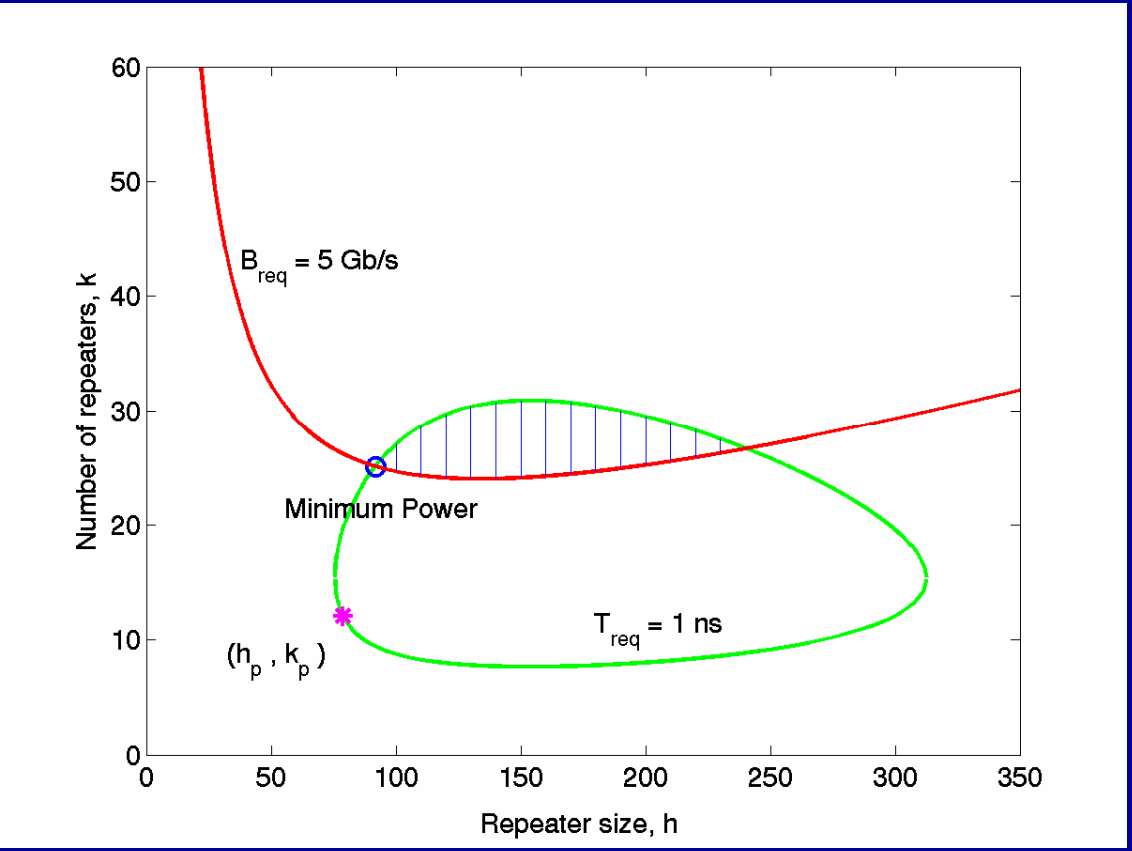
Repeater Insertion with Delay Constraints



- A design space of repeaters is determined by delay constraint
- With $T_{req} \to T_{min}$, design space converges to the delay optimum repeater system (h_{opt},k_{opt})

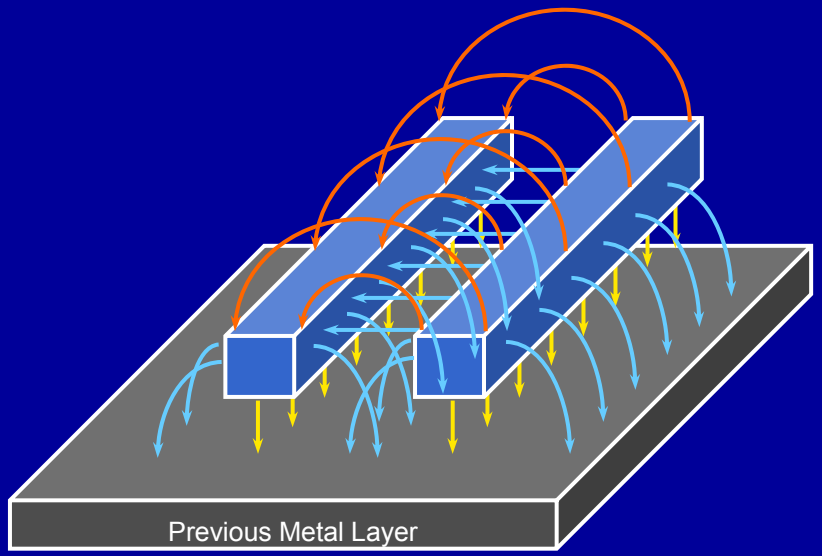
* G. Chen and E. G. Friedman, "Low Power Repeaters Driving *RC* Interconnects with Delay and Bandwidth Constraints," *Proceedings of the SOC Conference*, pp. 335-339, September 2004

Minimum Power with Both Delay and Bandwidth Constraints



- Satisfactory design occurs at the intersection of the two design spaces with delay and bandwidth constraints

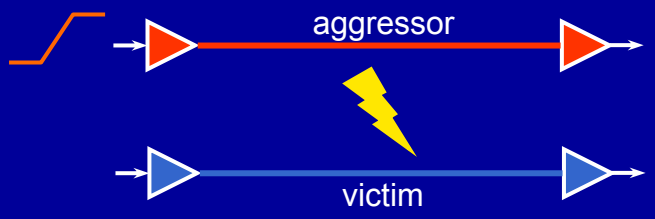
Interconnect Capacitive Coupling



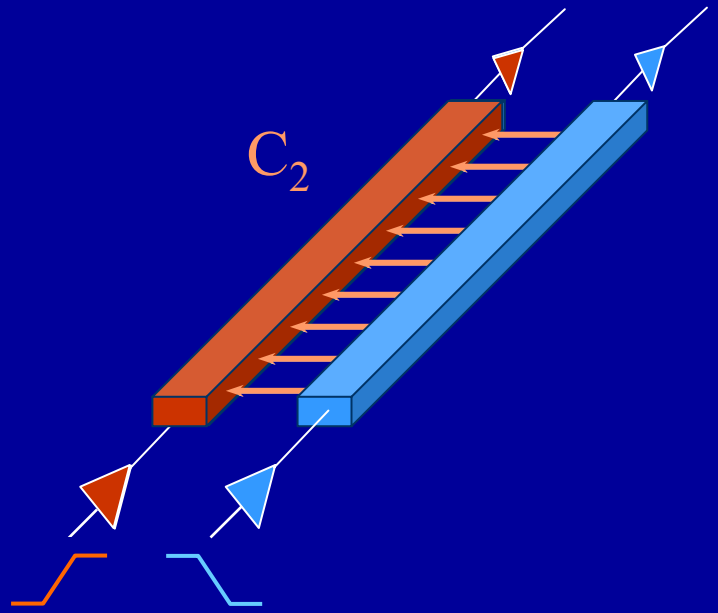
- Fringing capacitance increases with scaling
  - Spacing between lines decreases
  - Capacitive voltage divider creates coupling
    - Produces uncertainty in the signal delay

Capacitive Coupling Noise

- Signal coupling
  - Crosstalk
  - Aggressor – victim model
    - Aggressor: line generating noise
    - Victim: noise sensitive line



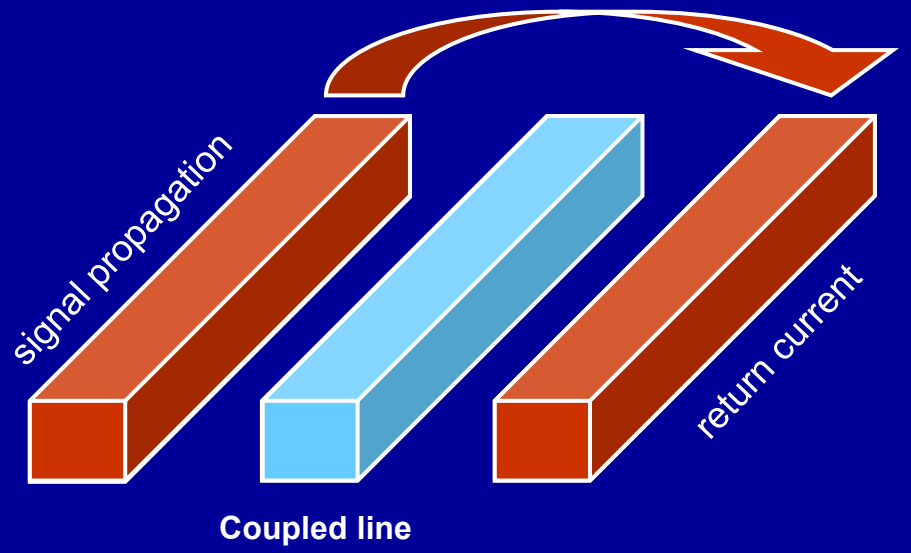
- Variations in signal delay
  - Simultaneous switching noise
  - Variations in effective coupling capacitance



*K. T. Tang and E. G. Friedman.," Delay and Noise Estimation of CMOS Logic Gates Driving Coupled Resistive-Capacitive Interconnections," *Integration*, September 2000

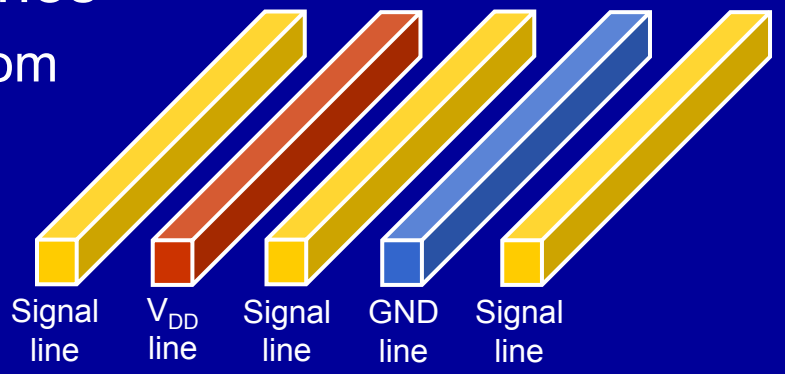
Inductive Coupling

- On-chip inductance effects
  - Increasing importance
    - Faster edge rates
    - Longer interconnect lengths on-chip
  - Mutual inductive coupling
    - Strongly depends upon the current return path
    - Return path can vary dynamically



Interconnect Shielding

- Insert power lines among signal lines
  - Isolates an aggressor (noisy) line from sensitive neighboring lines
  - Increases the noise tolerance of a sensitive line

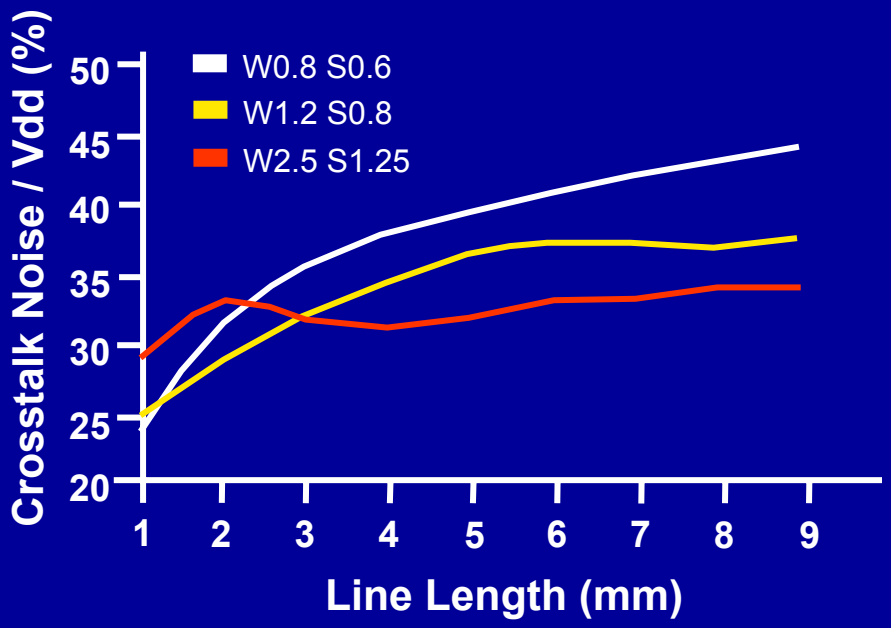


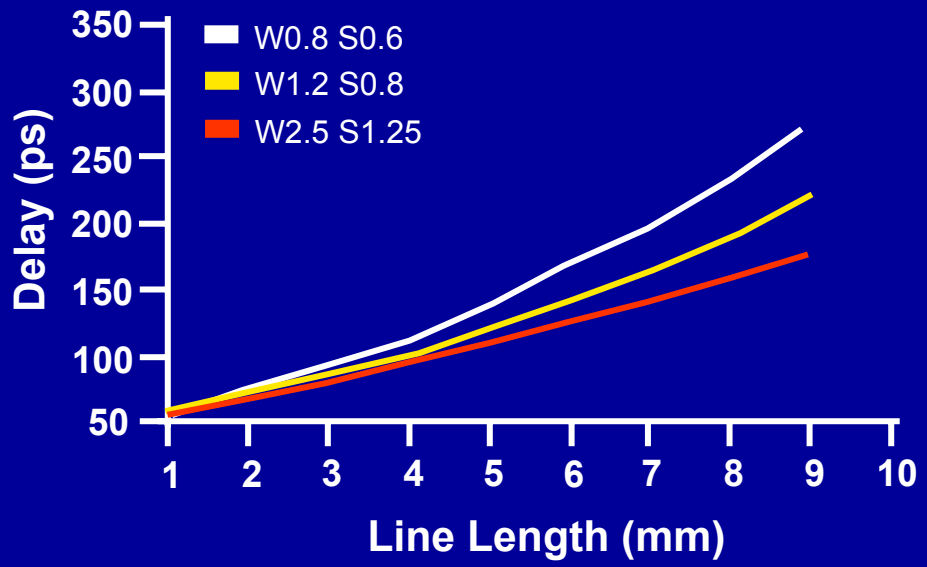
- Reduces capacitive coupling
  - The voltage of the shield lines typically does not switch
  - Reduces variations of the effective line capacitance
    - Reduced delay uncertainty

- Controls mutual inductance effects
  - The current return path is clearly determined

Geometric Wire Characteristics

- Narrow lines
  - RC dominant
  - Quadratic delay with line length
- Wide lines
  - Less noise at the far end
  - Delay is linearly dependent on line length
  - Inductive behavior

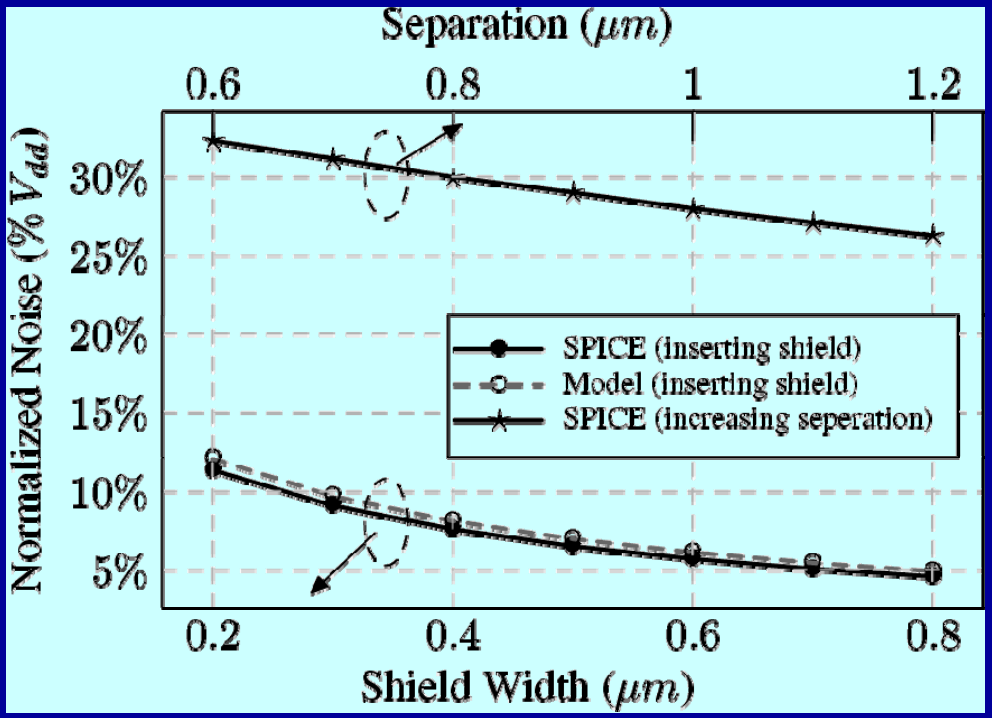




* X. Huang *et.al.* "RLC Signal Integrity Analysis of High-Speed Global Interconnects," *IEEE International Electron Devices Meeting*, 2000

Standard Techniques for Reducing Line-to-Line Crosstalk

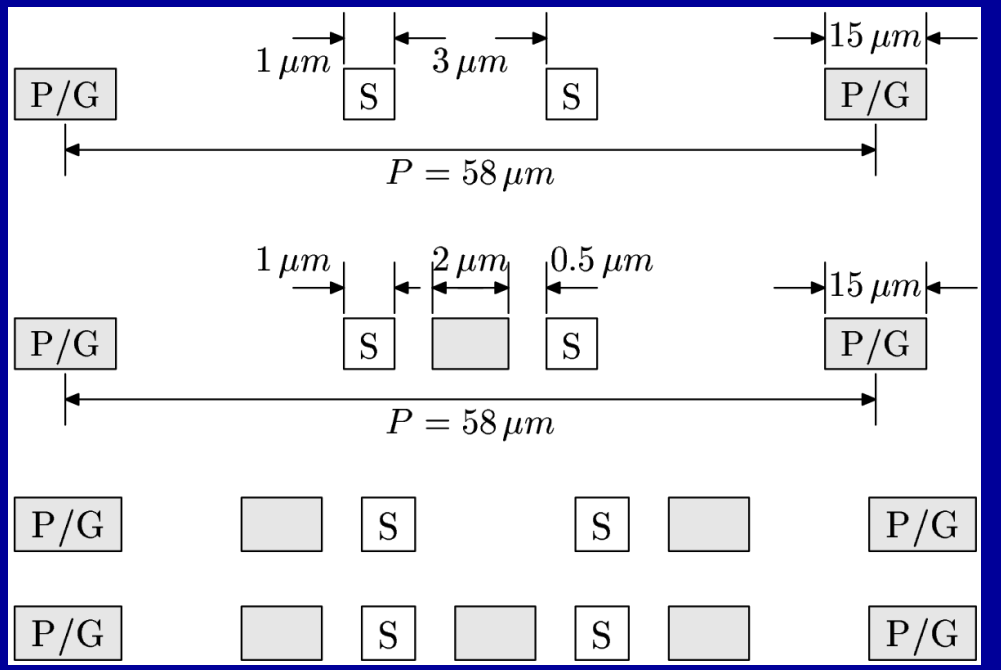
- Techniques to reduce crosstalk between coupled interconnects
  - Increase separation
  - Insert a shield line
- Inserting a shield line is more effective in reducing crosstalk than increasing the separation



* J. Zhang and E. G. Friedman, "Crosstalk Noise Model for Shielded Interconnects in VLSI-Based Circuits," *Proceedings of the IEEE International SOC Conference,* pp. 243-244, September 2003

Effect of Shield Insertion on Reducing Crosstalk Noise

- Three interconnect structures of one/two/three shield lines
  - Three shield structure has the least crosstalk noise
  - Two shield structure reduces the crosstalk noise to a level comparable to three shield structure
  - One shield structure has lower noise than two shield structure
    - Without considering other neighboring lines



No. of Shields	Effective Inductance (nH)						Crosstalk Noise ($\%V_{dd}$)		
	Mutual			Self					
	Analytic	FastHenry	Error	Analytic	FastHenry	Error	Analytic	SPICE	Error
No shield	2.655	2.655	0.00%	4.839	4.839	0.00%	38.66%	36.83%	4.73%
One shield	0.086	0.089	3.37%	2.269	2.273	0.18%	14.38%	15.15%	5.08%
Two shields	0.437	0.437	0.00%	2.115	2.118	0.14%	16.10%	15.06%	6.91%
Three shields	-0.048	-0.045	6.67%	1.634	1.636	0.12%	10.83%	9.61%	12.70%

Delay Improvement from Optimum Via Placement

Length [mm]	T_{el} (equally spaced) [ps]	T_{el}^{*} [ps]	Improvement [%]	n
1.560	56.11	52.40	7.08	10
1.609	58.78	10.59	10.59	10
2.383	75.00	64.11	16.99	10
1.665	60.36	50.10	20.48	10
1.749	63.10	52.42	20.37	10
1.233	35.92	33.11	8.49	7
1.167	34.97	31.43	10.12	7
1.132	34.26	30.28	11.62	7
1.933	45.23	34.65	23.39	7
1.716	46.18	37.18	19.49	7
2.428	29.03	23.89	17.71	4
1.875	44.28	38.73	12.53	4
2.121	34.56	25.62	25.86	4
3.429	56.11	42.47	24.31	4
1.701	27.34	25.69	6.03	4
Average Improvement			15.67	

- Interconnect Parameters
- r_v = 6.7 Ω/mm
- c_v = 6 pF/mm
- R_S = 15
- C_L = 50 fF
- l_v = 20 μm

V. Pavlidis and E. G. Friedman, “Interconect Delay Minimization through Interlayer Via Placement in 3-D ICs," *Proceedings of the ACM Great Lakes Symposium on VLSI,* pp. 20-25, April 2005.