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

---

Two to three order of magnitude delay difference

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

\[ R_{\text{line}} = Rl \quad l \quad C_{\text{line}} = Cl \]
Modeling Interconnect Inductance

- 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

\[ C_{line} = Cl \quad R_{line} = Rl \quad L_{line} = Ll \]
## Interconnect Models

<table>
<thead>
<tr>
<th>Technology</th>
<th>≥ 1 μm</th>
<th>1.0 ~ 0.5 μm</th>
<th>0.5 ~ 0.25 μm</th>
<th>&lt; 0.25 μm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resistance</strong></td>
<td>Ignored</td>
<td>Lumped</td>
<td>Distributed</td>
<td>Distributed</td>
</tr>
<tr>
<td><strong>Capacitance</strong></td>
<td>Area Capacitance</td>
<td>Area, 2-D fringing</td>
<td>2/3-D fringing, Lateral coupling</td>
<td>Lateral coupling dominant</td>
</tr>
<tr>
<td><strong>Inductance</strong></td>
<td>Ignored</td>
<td>Ignored</td>
<td>Ignored</td>
<td>Important</td>
</tr>
<tr>
<td><strong>Circuit model</strong></td>
<td>Lumped C</td>
<td>Lumped RC</td>
<td>Distributed RC</td>
<td>Distributed RLC</td>
</tr>
<tr>
<td><strong>Circuit Symbol</strong></td>
<td><img src="image1.png" alt="Circuit Symbol" /></td>
<td><img src="image2.png" alt="Circuit Symbol" /></td>
<td><img src="image3.png" alt="Circuit Symbol" /></td>
<td><img src="image4.png" alt="Circuit Symbol" /></td>
</tr>
</tbody>
</table>

- 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
• A historical perspective

⇒ Where are we now ?

• Where are we going ?

• Conclusions
# Problems in On-Chip Interconnect

<table>
<thead>
<tr>
<th>Extraction</th>
<th>Figures of Merit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling/Simulation</td>
<td>Design</td>
</tr>
</tbody>
</table>
Problems in On-Chip Interconnect

<table>
<thead>
<tr>
<th>Extraction</th>
<th>Figures of Merit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling/Simulation</td>
<td>Design</td>
</tr>
</tbody>
</table>
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 \textit{resistance} of the paths
  - High frequency — determined by the \textit{inductance} of the paths

\[ I_1 \approx I_0 \frac{R_2}{R_1 + R_2}, \quad I_2 \approx I_0 \frac{R_1}{R_1 + R_2} \]

\[ I_1 \approx I_0 \frac{L_2}{L_1 + L_2}, \quad I_2 \approx I_0 \frac{L_1}{L_1 + L_2} \]

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

<table>
<thead>
<tr>
<th>Extraction</th>
<th>Figures of Merit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling/Simulation</td>
<td>Design</td>
</tr>
</tbody>
</table>
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

\[
\frac{Rl}{2} \sqrt{\frac{C}{L}} > 1 \quad \text{or} \quad t_r > 2l\sqrt{LC} = 2T_O
\]

Problems in On-Chip Interconnect

<table>
<thead>
<tr>
<th>Extraction</th>
<th>Figures of Merit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling/Simulation</td>
<td>Design</td>
</tr>
</tbody>
</table>
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
<table>
<thead>
<tr>
<th>Extraction</th>
<th>Figures of Merit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling/Simulation</td>
<td>Design</td>
</tr>
</tbody>
</table>
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

• **Dynamic power** increases with line width

• **Short-circuit power** may decrease in underdamped highly inductive lines

• An optimum interconnect width exists
  – Minimum **transient power**

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

Vertical interplane interconnects (vias)

Interconnects
Adhesive polymer
Interconnects
Adhesive polymer
Devices
Bulk CMOS

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

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

Area = $L^2$, Corner to corner distance $\approx \sqrt{2}L$

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

Delay and Power Improvement from Optimum Via Placement

- $\Delta x_i$'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

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

Maximum number of physical planes = 16

<table>
<thead>
<tr>
<th></th>
<th>NOC</th>
<th>2-D</th>
<th>3-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC</td>
<td>2-D</td>
<td><img src="image" alt="2-D 2-D" /></td>
<td><img src="image" alt="2-D 3-D" /></td>
</tr>
<tr>
<td></td>
<td>3-D</td>
<td><img src="image" alt="3-D 2-D" /></td>
<td><img src="image" alt="3-D 3-D" /></td>
</tr>
</tbody>
</table>

- **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

\[ t_r = 0.5 \, t_c \]
\[ L_h = 1 \, \text{mm} \]
\[ L_v = 20 \, \mu \text{m} \]
- Wormhole routing
- Uniform traffic
- Max. number of planes = 16
- Packet size = 10 flits

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

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

<table>
<thead>
<tr>
<th>N = 512</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimum topology 3D IC – 3D NoC</td>
</tr>
<tr>
<td>Available physical planes of the stack are optimally distributed to</td>
</tr>
<tr>
<td>Reduce number of hops</td>
</tr>
<tr>
<td>Reduce communication channel length</td>
</tr>
</tbody>
</table>
Optimum Number of Nodes per Dimension for 3-D NoC

- $t_r = 2 \ t_c$
- $L_h = 1 \ mm$
- $L_v = 20 \ \mu m$
- 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

- 3D integration is coming
  - NoC is here and important for on-chip global data transfer

- 3D 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

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

![Diagram showing the effects of on-chip interconnect inductance on circuit design methodologies.](image-url)
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

Drilling

Cross section of holes
Etched by DRIE
60μm diameter
150μm depth

*Tru-Si Technologies, www.trusi.com

Filling

Cross section of holes
50μm diameter
150μm depth
Etched by DRIE
Filled by ECD Cu
Bump needs to be Planarized by CMP

Cross section of holes
100μm
1μm Al film
150μm

100μm

Cross section of holes
Drilled by Laser
100μm diameter
150μm depth

Cross section of holes
250μm diameter
150μm depth hole etched by ADP
filled by EVD Cu planarized by Cu CMP
Existing Work in 3-D Integration

- Fabrication techniques
  - Copper wafer bonding
  - Adhesion with polymers

- Heating Effects and Constraints

High Performance Digital IC Design Challenges

- **Primary objectives**
  - Improve chip functionality
    - More transistors
    - Larger ICs
  - Improve circuit density
    - Smaller transistors
    - On-chip scaling
  - Improve performance
    - Faster transistors
    - Higher clock speeds

- **Additional challenges**
  - Design complexity
  - Noise sensitivity
  - Power dissipation
  - Interconnect design
  - Synchronization
  - Low power design
Shielding Design Methodologies for High Speed Interconnects

- Shielding design methodologies for coupled RC interconnects
- Crosstalk analysis for coupled RLC interconnects
The effect of the shield line on reducing crosstalk depends upon the number of ground connections tying 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.
A design space of repeaters is determined by delay constraint

With $T_{\text{req}} \geq T_{\text{min}}$, design space converges to the delay optimum repeater system $(h_{\text{opt}}, k_{\text{opt}})$

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

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

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

<table>
<thead>
<tr>
<th>No. of Shields</th>
<th>Effective Inductance (nH)</th>
<th>Crosstalk Noise (%$V_{dd}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mutual</td>
<td>Self</td>
</tr>
<tr>
<td></td>
<td>Analytic</td>
<td>FastHenry</td>
</tr>
<tr>
<td>No shield</td>
<td>2.655</td>
<td>2.655</td>
</tr>
<tr>
<td>One shield</td>
<td>0.086</td>
<td>0.089</td>
</tr>
<tr>
<td>Two shields</td>
<td>0.437</td>
<td>0.437</td>
</tr>
<tr>
<td>Three shields</td>
<td>-0.048</td>
<td>-0.045</td>
</tr>
</tbody>
</table>
## Delay Improvement from Optimum Via Placement

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

**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

---