The Future of Electrical I/O for Microprocessors

Frank O’Mahony
frank.omahony@intel.com
Intel Labs, Hillsboro, OR USA
Outline

- 1TByte/s I/O: motivation and challenges
- Circuit Directions
- Channel Directions
- Tool Directions
- 470Gb/s Prototype
Microprocessor Bandwidth Needs

- As CPU core count increases, I/O bandwidth (BW) requirements will increase for all segments
- Current system bandwidth requirements (Y2010)
  - Client BW = ~50GB/s
  - Server BW = ~100GB/s
  - High-end Server BW = ~200GB/s
Microprocessor Bandwidth Trends

Bandwidth Drivers:
- CPU↔Memory
- CPU↔CPU
- CPU↔Peripheral
- CPU↔I/O bridge

High-end microprocessors are expected to need ~1TB/s during coming decade
Microprocessor I/O Power

- Current system I/O power efficiency is 20-40pJ/bit

<table>
<thead>
<tr>
<th>System</th>
<th>BW</th>
<th>I/O Pwr. Eff.</th>
<th>I/O Pwr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client</td>
<td>~50GB/s</td>
<td>20pJ/bit</td>
<td>8W</td>
</tr>
<tr>
<td>Server</td>
<td>~100GB/s</td>
<td>20pJ/bit</td>
<td>16W</td>
</tr>
<tr>
<td>High-End Server</td>
<td>~200GB/s</td>
<td>20pJ/bit</td>
<td>32W</td>
</tr>
</tbody>
</table>

- If I/O power efficiency doesn’t improve during the next decade, then:

\[ 1\text{TB/s} \times 20\text{pJ/bit} = 160\text{W} \]
I/O Energy Efficiency Trends

Issue: ~20% per year power reduction while bandwidth increasing 40-70% per year

Ref: R. Palmer, ISSCC ‘07
Energy Efficiency and Channel Loss Tradeoff

- Power efficiency is strongly correlated to channel loss
- Simply scaling per-pin BW will not meet power budget
- Low power interfaces should be “wider” not faster

Based on transceivers reported 2006-2009 in 65-130nm CMOS
Channel/Interconnect Density

• Conventional package/socket density does not scale with process
• “Width” of interfaces is limited by routing congestion

![Graph showing C4 bump area and Circuit area over time]

Flip-chip Package

C4 pitch << Pkg. pin pitch
Problem Statement Summary

- Bandwidth needs are quickly approaching 1TB/s
- Energy efficiency is not scaling as aggressively as bandwidth
- The channel limits our ability to increase per-pin data rate and/or increase the width of an interface
How Will Electrical I/O scale to 1TB/s?

1. Co-design the interconnects and I/O circuitry to meet bandwidth, scalability and power efficiency demands

2. Scale the channel by transitioning to new channel configurations and materials

3. Use accurate, statistical link design tools to identify balanced architectures.
Outline

• 1TByte/s I/O: motivation and challenges
• Circuit Directions
• Channel Directions
• Tool Directions
• 470Gb/s Prototype
Low Active Power Techniques

Power Optimized Links
- Simple equalization
- Low TX swing
- Sensitive RX sampler
- Low-power clocking
Minimize analog circuit complexity

- Lowest power links find ways to simplify equalization and clocking circuitry to reduce power

- Equalization examples:
  - Constrain equalization range by known channel characteristics
  - Continuous-time linear Rx equalizer

Ref: G. Balamurugan, JSSC 4/08
Power Management: Scalable supplies

• Adapt supply to frequency, process, temperature (f,P,T)
  – Digital: Power $\propto V_{\text{SUPPLY}}^2 \cdot f$
  – Analog: Power $\propto V_{\text{SUPPLY}} \cdot I_{\text{bias}}$

• Removes excess circuit BW and headroom
Power Management: Scalable supplies

- Power efficiency improves with adaptive supply/biasing

Refs: G. Balamurugan, JSSC 4/08 and B. Casper, ISSCC ‘06
Aggressive Power Management

- Don’t spend power doing nothing!
- Rapidly adapt to bandwidth demand
  - Requires fast, granular bandwidth adaptation
Aggressive Power Management

Conventional (fixed Bandwidth)

- Don’t spend power doing nothing!
- Rapidly adapt to bandwidth demand
  - Requires fast, granular bandwidth adaptation
• Don’t spend power doing nothing!
• Rapidly adapt to bandwidth demand
  – Requires fast, granular bandwidth adaptation
Device Variation in Scaled CMOS

- Device manufacturing tolerances are improving
- …but area scaling still causes higher variation
- Fundamental power/area to variation tradeoff is not acceptable

\[ \sigma V_T = \frac{1}{\sqrt{2}} \left( \frac{c_2}{\sqrt{W_{\text{eff}} \cdot L_{\text{eff}}}} \right) \]

Need circuit architectures that fundamentally change this tradeoff.

Ref: K. Kuhn, IEDM 2007
Mitigating Device Variation

- Calibration greatly improves the power/variation tradeoff
  - Receiver offset calibration
  - Duty cycle correction
  - Adaptive equalizers
  - Clock recovery (or deskew)

- Simple calibration doesn’t alleviate all variation issues (e.g. PSRR)

Circuit derivatives ($gm$, $ro$) are not calibrated by offset calibration $\rightarrow$ PSRR is not calibrated
Mitigating Device Variation

- Calibration greatly improves the power/variation tradeoff
  - Receiver offset calibration
  - Duty cycle correction
  - Adaptive equalizers
  - Clock recovery (or deskew)
- Simple calibration doesn’t alleviate all variation issues (e.g. PSRR)
- Possible solutions:
  - “Dynamic” calibration (e.g. auto-zero)
  - Redundancy/reconfigurability
  - Better “correct by design” circuits

Ref: P. Hanumolu, JSSC 2/08.
Outline

- 1TByte/s I/O: motivation and challenges
- Circuit Directions
- Channel Directions
- Tool Directions
- 470Gb/s Prototype
Channel scaling

- Circuit innovation alone will probably not be enough to reach the 1TB/s target → the channel needs to scale too!
- **Better signal integrity:** Improved electrical characteristics mean less power in clocking and equalization
- **Higher density:** More lanes allow each lane to operate at lower data rate → better power efficiency
Channel vs. Equalization tradeoffs: Backplane example

Ref: B. Casper, CICC ‘07.
Improve channel signal integrity

[Graph showing signal strength over frequency for different connector types: 5mm Stubbed-via BP, Drilled-via BP, 19” Flex cable.]

- 19” Flex
- 5mm Stubbed-via BP
- Drilled-via BP

- Frequency ranges: 0GHz, 5GHz, 10GHz, 15GHz
- Signal strength in dB: 0dB, -20dB, -40dB, -60dB, -80dB

- Components: CPU Socket, Flex Connector, 19” Flex Cable
High density channels

Contact pitch

Routing pitch
High density channels

Approx. Contact/Routing Pitch (µm)

- Contact pitch
- Routing pitch

- FR4
- Flex
- Package
- Si interposer
- Proximity
High density channels

![Contact/Routing Pitch (µm)](image)

- **Contact pitch**
- **Routing pitch**

Approx. Contact/Routing Pitch (µm):
- FR4: ~1000 µm
- Flex: ~100 µm
- Package: ~10 µm
- Si interposer: ~1 µm
- Proximity: ~1 µm

Intel logo
High density channels

Contact pitch
Routing pitch

Approx. Contact/Routing Pitch (µm)

FR4  | Flex  | Package | Si interposer | Proximity

Contact pitch
Routing pitch
Outline

• 1TByte/s I/O: motivation and challenges
• Circuit Directions
• Channel Directions
• Tool Directions
• 470Gb/s Prototype
What is the “Right” Link Architecture?

- Designers need the ability to quickly and accurately compare architecture options.
Empirical Approach

- Simulate system with random data
- This doesn’t provide adequate accuracy ($\text{BER} < 10^{-12}$)
Full System Statistical Analysis

- Specify high-level architecture and block characteristics
- Enables fast evaluation of link sensitivities
Maximum Data Rate Comparison: Backplane vs. Flex

- Statistical system analysis provides designers with real performance tradeoffs and “brick walls”
Statistical system analysis provides designers with real performance tradeoffs and “brick walls”
Outline

• 1TByte/s I/O: motivation and challenges
• Circuit Directions
• Channel Directions
• Tool Directions
• 470Gb/s Prototype
47x10Gb/s, 1.4pJ/bit Interface (45nm CMOS)
Bundled Architecture

Conventional: Independent clocking
Bundled Architecture

- Clocking innovation ➔ Bundle clocking

Conventional: Independent clocking
Bundled Architecture

- Clocking innovation → Bundle clocking

**Conventional:** Independent clocking

**Optimized:** Bundle clocking

Bundled clocking reduces I/O power
Fast RX Power States

• RX bundle power reduced by 93% in standby
• All RX lanes return to reliable operation in <5ns
Silicon Area Compression

Conventional: I/O floor plan

- Power
- Ground
- I/O layout
- I/O signals
Silicon Area Compression

- Floor plan optimization $\rightarrow$ minimize I/O area

**Conventional: I/O floor plan**
Silicon Area Compression

- Floor plan optimization → minimize I/O area

**Conventional: I/O floor plan**

**Optimized: Bundle layout**

**I/O circuitry**
Interface Floorplan

- Active circuit area is reduced with TL routing.
Interface Configuration

- Within-bundle lanes matched to <100µm
  - Dense LGA connector minimizes breakout area
  - Bundles share the same routing layer
  - 2X density on stripline layers due to reduced Xtalk
Silicon and Interconnect Prototypes

0.5m flex interconnect

3m twinax cable
Electrical Interconnect Scaling Challenges

(Based on transceivers reported 2006-2009 in 65-130nm CMOS)

This work: 45nm CMOS Prototype @10Gb/s data rate
I/O Power Efficiency Measurements

Link data rate = 10Gb/s

- HDI*
- LCP flex
- 32AWG - twinax

*high density interconnect (HDI)
Summary

• Bandwidth needs are quickly approaching 1TB/s

• Extending electrical I/O to 1TB/s requires balance between power, data rate, density and cost

• Evaluate alternate channel configurations and materials

• Recent results indicate that electrical will be up to the task for “in-box” I/O