Scalable Nanophotonic Interconnect for Cache Coherent Multicores

Randy W. Morris, Jr. and Avinash K. Kodi
Department of Electrical Engineering and Computer Science
Ohio University, Athens, OH 45701
E-mail: rmorris@cs.ohiou.edu, kodi@ohio.edu
Website: http://oucsace.cs.ohiou.edu/~avinashk/

WINDS 2010: Workshop on the Interaction between Nanophotonic Devices and Systems
Atlanta, GA
December 5, 2010
Talk Outline

• Section I: Motivation & Background

• Section II: Dual Sub-Network for Snoopy Cache Coherent Nanophotonic Architecture

• Section IV: Performance Analysis

• Section V: Future Work
Why Nanophotonics?

- Power consumption of Network-on-Chips (NoCs) using metallic interconnects is projected to exceed expectation by a factor of 10

Nanophotonic Technology

- Low Power
- Small Footprint (10 – 15 μm)
- High Bandwidth (10 – 20 Gbps)
- CMOS Compatibility

**Micro-ring Resonators**

Resonant wavelength ($\lambda_0$)

$$\lambda_0 \times m = n_{\text{eff}} \times 2\pi R$$

$m \rightarrow$ an integer

$n_{\text{eff}} \rightarrow$ effective refractive index

$R \rightarrow$ radius of the ring resonator

Cache Coherence

- **Write propagation** (write by any processor should become visible to all other processors)
- **Write serialization** (all writes from same or different processors are seen in the same order by all processors)

**Snoopy Protocols**

- Easy to Program
- Not Easily Scalable

**Directory Protocols**

- Scalable
- High miss latency

**Interconnection Network**

- Point-to-Point
Problems with Snoopy Networks

Two major problems with snoopy cache coherent networks

(1) Interconnect bandwidth for broadcasting of memory requests
- Bus Networks: Limits one request per cycle
- Multiple Buses: Increases cache controllers
- Point-to-Point Networks: Selective multicasting & Ordering

(2) Cache Access Rate
- Cache tag lookup (latency)
- Increased power consumption
Related Work (to name a few)

**Electrical**
- Split Transactional Bus
- Sun Fireplane (SC 2001)
- Timestamp Snooping (ASPLOS 2000), Multicast Snooping (ISCA 2001)
- Jetty (HPCA 2001), Region Scout (ISCA 2005), Intel QPI
- Broadcasting on Ordered Networks (HPCA 2009, MICRO 2009)

**Optical/Nanophotonic**
- SYMNET (Trans on Parallel & Dist Systems 2004)
- Shared Bus (MICRO 2006), Wavelength Routed Oblivious Network (ASPLOS 2010)
- Spectra (ISPLED 2009), ATAC (PACT 2010)
Advantages of the proposed architecture

– Dual sub-networks for memory request
  • Broadcast & Multicast networks
– Broadcast network used by all tiles to fetch the missed block
  • Network access implemented using tokens
  • Determines the sharing pattern
– Multicast network to be shared between nodes to send selective requests
  • Reduces the broadcast requirement
  • Simultaneous transient requests in progress to different memory locations
– Reducing the external laser power by unique power guiding techniques
Proposed Broadcast Sub-Network Architecture: CC-NPA
Power Guiding

As only one core can transmit, route power to a column of cores.

- Reduction in optical power (~75%)

The active column is determined by the circulating optical tokes

2 dB optical loss
Optical Token System (1/3)

Requests a token

Received Token

Control Center

Inject token

Power

Return

Tile 0  Tile 1  Tile 2  Tile 3

Tile 4  Tile 5  Tile 6  Tile 7

Tile 8  Tile 9  Tile 10  Tile 11

Tile 12  Tile 13  Tile 14  Tile 15
Optical Token System (2/3)

Requests a token

Tile 0
Tile 1
Tile 2
Tile 3
Tile 4
Tile 5
Tile 6
Tile 7
Tile 8
Tile 9
Tile 10
Tile 11
Tile 12
Tile 13
Tile 14
Tile 15

Token Re-Injected

Inject token

Power

Return

Inject

Power
Optical Token System (3/3)

Fairness can be insured with additional techniques (Fair slot, Two pass)
Proposed Multicast Sub-Network

For larger networks, snoopy-based cache coherence reduces performance
- Broadcasting data to all shared tiles, consuming more address bandwidth
- Consumes more latency and power at the caches

- Wavelength routed second multicast sub-network
- Filter and route cache requests to nodes that hold the cache data
- Reduction in required bandwidth and power dissipation
- Potential for simultaneous multiple requests (could lead to race conditions)
Initial Performance Analysis

• Performance Comparison
  – Simics with Gems Memory Module
  – FFT, LU, Radiosity, Ocean, Radix, & Water

• Area & Power Analysis

Simics Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1/L2 coherence</td>
<td>MOSI</td>
<td>Core Frequency</td>
<td>5 GHz</td>
</tr>
<tr>
<td>L2 cache size/accoc</td>
<td>256 KB/16-way</td>
<td>Threads (core)</td>
<td>2</td>
</tr>
<tr>
<td>L1 cache/accoc</td>
<td>64KB/4-way</td>
<td>Issue policy</td>
<td>In-order</td>
</tr>
<tr>
<td>Cache line size</td>
<td>64B</td>
<td>Memory Size (GB)</td>
<td>4</td>
</tr>
<tr>
<td>Memory Controllers</td>
<td>16</td>
<td>Address Bandwidth (opt)</td>
<td>640 GBps</td>
</tr>
<tr>
<td>Address Bandwidth (elec)</td>
<td>320 GBps</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Splash-2 Speed up (16-cores)

- CC-NPA increases performance by about 25%
Splash-2 Speed up (64-cores)

- CC-NPA increases performance of up to 2x
# Power Analysis

<table>
<thead>
<tr>
<th>Device</th>
<th>Loss (dB)</th>
<th>Device</th>
<th>Loss (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coupler (Lc)</td>
<td>1</td>
<td>Filter drop (Lf)</td>
<td>1</td>
</tr>
<tr>
<td>Non-Linearity (Ln)</td>
<td>1</td>
<td>Bending (Lb)</td>
<td>1</td>
</tr>
<tr>
<td>Photo-detector (Lp)</td>
<td>1</td>
<td>Waveguide Crossing (Lwc)</td>
<td>0.05</td>
</tr>
<tr>
<td>Modulator Insertion (Li)</td>
<td>1</td>
<td>Receiver (Lrs) Sensitivity</td>
<td>-20 dBm</td>
</tr>
<tr>
<td>Waveguide (per cm) (Lw)</td>
<td>1.3</td>
<td>Splitter (Ls)</td>
<td>3</td>
</tr>
<tr>
<td>Laser Efficiently</td>
<td>30%</td>
<td>Ring modulation</td>
<td>150 fJ/b</td>
</tr>
<tr>
<td>Ring Heating</td>
<td>100 fJ/b</td>
<td>TIA/ voltage amp.</td>
<td>1.1 pJ/b – 100 fJ/bit</td>
</tr>
</tbody>
</table>

\[ 5 \times L_S + 7 \times L_W + L_C + L_N + 3 \times L_I + L_F + 8 \times L_B + 100 \times L_{WC} \]

-43.1 dB (per wavelength)

Total Power (opt) = 5.44 W (8 wavelengths)
Area Analysis

<table>
<thead>
<tr>
<th>Device</th>
<th>Area ($\mu m^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waveguide (pitch)</td>
<td>5.5 $\mu m$</td>
</tr>
<tr>
<td>Micro-ring resonator</td>
<td>100</td>
</tr>
<tr>
<td>Photo-detector</td>
<td>100</td>
</tr>
<tr>
<td>TIA/ Limiting Amplifier</td>
<td>0.02625 (mm$^2$)</td>
</tr>
</tbody>
</table>

Broadcast Sub-Network: 24 mm$^2$ (optical)  
51 mm$^2$ (electrical)
Conclusion & Future Work

• CC-NPA is both a low power & high bandwidth network for future cache coherent many-core processors

• CC-NPA combines the benefits of snoopy cache coherent protocols and nanophotonics

• CC-NPA provides scalable bandwidth using two sub-networks (broadcast and multicast)

• Future work will involve designing and optimizing the multicast sub-network