Hybrid On-chip Data Networks

Gilbert Hendry
Keren Bergman
Chip-Scale Interconnection Networks

- Chip multi-processors create need for high performance interconnects
- Performance bottleneck of on-chip networks and I/O
- Power dissipation constraints of the chip package
  - > 50% of total power comes from interconnects*

Motivation

• CMPs of the future = 3D stacking
• Lots of data on chip
• Photonics offers key advantages
Why Photonics?

Photonics changes the rules for Bandwidth, Energy, and Distance.

**ELECTRONICS:**
- Buffer, receive and re-transmit at every router.
- Each bus lane routed independently. \( P \propto N_{\text{LANES}} \)
- Off-chip BW is pin-limited and power hungry.

**OPTICS:**
- Modulate/receive high bandwidth data stream once per communication event.
- Broadband switch routes entire multi-wavelength stream.
- Off-chip BW = On-chip BW for nearly same power.
Hybrid Network Premise

Optical processing difficult and limited

Source, destination routing inefficient

Use electronics for routing, optics for switching and transmission

Hybrid Circuit-Switching
Hybrid Circuit-Switched Networks

Step 1: Path SETUP request

Electronic SETUP Msg

Photonic Plane

Electronic Plane

Processing Plane

Source core

Destination Core
Step 2: Path ACK
Hybrid Circuit-Switched Networks

Step 3: Transmit Data

Photonic Switch Use Information
Hybrid Circuit-Switched Networks

Meanwhile: Path Contention

Path BLOCKED Msg (Backoff)
Hybrid Circuit-Switched Networks

Step 4: Path TEARDOWN

Electronic SETUP Msg

Source core

Destination Core
Hybrid Circuit-Switched Networks

Pros:

- Energy-efficient end-to-end transmission
- High bandwidth through WDM
- Electronic network still available for small control messages*
- Network-level support for secure regions

Cons:

- Path setup latency
- Path setup contention (no fairness)

* [G. Hendry et al. Analysis of Photonic Networks for a Chip Multiprocessor Using Scientific Applications. In NOCS, 2009]
Programming and Communication
“… [OpenMP on large systems] often performs worse than message passing due to a combination of false sharing, coherence traffic, contention, and system issues that arise from the difference in scheduling and network interface moderation”

~ Exascale Report
### Partitioned Global Address Space

<table>
<thead>
<tr>
<th>Access</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local Read</td>
<td>Optical Receive</td>
</tr>
<tr>
<td>Local Write</td>
<td>Optical send</td>
</tr>
<tr>
<td>Remote Read</td>
<td>Electronic request, optical receive</td>
</tr>
<tr>
<td>Remote Write</td>
<td>Optical send</td>
</tr>
<tr>
<td>Shared R/W</td>
<td>?</td>
</tr>
</tbody>
</table>

[See Hendry et al. *Circuit-Switched Memory Access in Photonic Interconnection Networks for HPEC.* In Supercomputing, Nov. 2010]
Message Passing

- Complex, dynamic access patterns
- Relatively larger blocks of data
- Scientific computing →

Streaming

- Embedded / specialized systems (Graphics, Image + Signal Proc.)
- Execution mode of general-purpose systems (Cell Processor)

```
1 -> 2
1 -> 4
3 -> 1
4 -> 3
```

Input Data

Output Data

Persistent optical circuits

Shared Memory

PGAS

MPI

Explicit Communication

Implicit Communication

Streaming
Electronic Plane
Electronic Router

- Low frequency operation (~ 1GHz)
- 1 VC (typically)
- Small buffers (64-28)
- Narrow Channels (8-32)
Network Gateway

External Concentration

[P. Kumar et al. Exploring concentration and channel slicing in on-chip network router. In NOCS, 2009]
The Photonic Plane
Silicon photonic waveguide technology provides low-power optical interconnects in CMOS-compatible platforms.

1.28 Tb/s Data Transmission Experiment (occupies small slice of available WG BW)

Silicon photonic waveguides offer low-loss (1.7 dB/cm), high-bandwidth (> 200 nm) silicon photonic waveguides that can be fabricated in commercial CMOS process.
Silicon Photonic Modulator and Detector Technology

- 18 Gb/s demonstrated

[Ge on Si Detectors:]
- 40-GHz bandwidths
- 1 A/W responsivities

 Receivers (detectors w/ CMOS amplifiers):
- 1.1 pJ/bit demonstrated at 10 Gb/s
- Scalable to < 50 fJ/bit

- 85 fJ/bit demonstrated at 10 Gb/s
- Scalable to < 25 fJ/bit

[M Lipson, Optics Express (2007)]

[M Watts, Group Four Photonics (2008)]

[S Koester, J. Lightw. Technol. (2007)]
Silicon Photonic Micro-Ring Switch Explanation

fast control of resonance wavelength via carrier injection

Transmission \((\text{in}_i \rightarrow \text{out}_i)\)

bar state
no current, on-resonance

cross state
current, off-resonance
Higher Order Switch Designs
On-Chip Topology Exploration

- Photonic Torus
  - [A. Shacham et al., Trans. on Comput., 2008]

- Nonblocking Photonic Torus
  - [M. Petracca et al. IEEE Micro, 2008]
On-Chip Topology Exploration

- TorusNX

- Square Root

[J. Chan et al. JLT, May 2010]
Photonic Plane Characteristics

- Insertion Loss
- Noise
- Power
Insertion Loss and Optical Power Budget

Nonlinear Effects

Total Power of WDM Signal

WDM Factor

Worst-case Insertion Loss

Detector Sensitivity

Optical Power Budget

Waveguide crossing
-0.15 dB each

1 dBm

Waveguide propagation
-1.5 dB/cm

0.68 dBm

Passing by a ring
-0.005 dB each
Simulation Results

Torus Topology

<table>
<thead>
<tr>
<th>Topology Size (nodes)</th>
<th>Insertion Loss (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4x4</td>
<td>20.6</td>
</tr>
<tr>
<td>6x6</td>
<td>25.6</td>
</tr>
<tr>
<td>8x8</td>
<td>31.2</td>
</tr>
<tr>
<td>10x10</td>
<td>37.0</td>
</tr>
<tr>
<td>12x12</td>
<td>42.8</td>
</tr>
<tr>
<td>14x14</td>
<td>48.6</td>
</tr>
<tr>
<td>16x16</td>
<td>54.5</td>
</tr>
<tr>
<td>18x18</td>
<td>60.3</td>
</tr>
</tbody>
</table>

Non-Blocking Torus Topology

<table>
<thead>
<tr>
<th>Topology Size (nodes)</th>
<th>Insertion Loss (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4x4</td>
<td>18.7</td>
</tr>
<tr>
<td>6x6</td>
<td>25.3</td>
</tr>
<tr>
<td>8x8</td>
<td>31.5</td>
</tr>
<tr>
<td>10x10</td>
<td>38.0</td>
</tr>
<tr>
<td>12x12</td>
<td>44.1</td>
</tr>
<tr>
<td>14x14</td>
<td>50.6</td>
</tr>
<tr>
<td>16x16</td>
<td>56.8</td>
</tr>
<tr>
<td>18x18</td>
<td>63.2</td>
</tr>
</tbody>
</table>

TorusNX Topology

<table>
<thead>
<tr>
<th>Topology Size (nodes)</th>
<th>Insertion Loss (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4x4</td>
<td>15.8</td>
</tr>
<tr>
<td>6x6</td>
<td>19.5</td>
</tr>
<tr>
<td>8x8</td>
<td>23.2</td>
</tr>
<tr>
<td>10x10</td>
<td>27.1</td>
</tr>
<tr>
<td>12x12</td>
<td>31.0</td>
</tr>
<tr>
<td>14x14</td>
<td>34.9</td>
</tr>
<tr>
<td>16x16</td>
<td>38.8</td>
</tr>
<tr>
<td>18x18</td>
<td>42.7</td>
</tr>
</tbody>
</table>

Square Root Topology

<table>
<thead>
<tr>
<th>Topology Size (nodes)</th>
<th>Insertion Loss (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4x4</td>
<td>12.2</td>
</tr>
<tr>
<td>8x8</td>
<td>21.5</td>
</tr>
<tr>
<td>16x16</td>
<td>30.6</td>
</tr>
</tbody>
</table>

Legend:
- Black: Propagation
- Red: Crossing
- Blue: Dropping Into a Ring
*Original* is based on the IL results from previous slide, *Improved* is based on a hypothetical improvement in crossing loss from 0.15 dB to 0.05 dB.
Photonic Plane Characteristics

- Insertion Loss
- Noise
- Power
Noise and Crosstalk

Laser Noise

Modulation Noise

Inter-Message Crosstalk

Coherent noise

Intra-Message Crosstalk

Incoherent noise

Crosstalk

Filter
Effects of Noise

- Optical SNR vs. Network Size
- Number of $\lambda$ vs. Network Load

Diagrams show the relationship between different parameters.
Simulation Results

**Results**

- Results are plotted for network size of $8 \times 8$ at saturation, at the detectors.
- Maximum OSNR = $\sim 45$ dB (due to laser noise)
- Minimum OSNR < 17 dB (due to message-to-message crosstalk)
- Variations between networks due to varying likelihood of two message intersecting on network topology.

**System Performance**

- SNR measures the likelihood of error-free transmission.
- Lower SNR designs will require additional retransmission, resulting in lower throughput performance.

The line at OSNR=16.9 dB is where a bit-error-rate of $10^{-12}$ can be achieved, assuming an ideal binary receiver circuit and orthogonal signaling.
Photonic Plane Characteristics

- Insertion Loss
- Noise
- Power
Power Usage

- Laser Power
- Active Power
  - Modulating
  - Detecting
  - Broadband
- Static Power
  - Thermal tuning
- Tx\Rx Power
  - Drivers
  - TIAs
Energy Per Bit

The graph illustrates the energy per bit (J/bit) as a function of message size (bit). The x-axis represents the message size in bits, ranging from 10^0 to 10^7. The y-axis shows the energy per bit from 10^-13 to 10^-7. Four graph lines are shown:

- **Torus**: Represented by black squares.
- **Non-blocking Torus**: Represented by red circles.
- **TorusNX**: Represented by blue triangles.
- **Square Root**: Represented by green diamonds.

As the message size increases, the energy per bit decreases for all graph lines, indicating an efficient energy consumption pattern with increasing data size.
Power Breakdown

• 12 wavelengths @ 10 Gbps/each
• Power Dissipation = 4.31 W

• 7 wavelengths @ 10 Gbps/each
• Power Dissipation = 1.59 W

• Results based on randomly generated traffic with message sizes of 100 kbit, with network in saturation.
• Data was collected on 64 nodes topologies constrained to a total surface area of 2 cm × 2 cm.
Power Breakdown

Square Root Topology

- Router Logic: 34%
- Router Buffer: 31%
- Modulator: 14%
- Detector: 8%
- Electronic Wire: 7%
- PSE: 2%
- Thermal: 4%

TorusNX Topology

- Router Logic: 37%
- Router Buffer: 31%
- Modulator: 17%
- Detector: 10%
- Electronic Wire: 1%
- PSE: 1%
- Thermal: 3%

- 27 wavelengths @ 10 Gbps/each
- Power Dissipation = 1.89 W

- 38 wavelengths @ 10 Gbps/each
- Power Dissipation = 3.22 W
Performance
Other Interesting Issues
Memory Access

Other Arbitration Means - TDM

[Image of network gateway and switch controller]

Wavelength Granularity

- Original

- Re-design

- Scalable number of WDM channels
Conclusion

• Some applications / programming models definitely well-suited to a circuit-switched photonic network
• Interesting tradeoffs and design space