Hybrid On-chip Data Networks

Gilbert Hendry
Keren Bergman

Lightwave Research Lab
Columbia University
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. ($P \propto N_{LANES}$)
- 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
Hybrid Circuit-Switched Networks

Step 2: Path ACK

Electronic ACK Msg
Step 3: Transmit Data

Photonic Switch Use Information
Hybrid Circuit-Switched Networks

Meanwhile: Path Contention

Path BLOCKED Msg (Backoff)
Step 4: Path TEARDOWN
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>

Message Passing

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

* G. Hendry et al. *Analysis of Photonic Networks for a Chip Multiprocessor Using Scientific Applications.* In NOCS, 2009

![Diagram showing the spectrum from implicit to explicit communication with Shared Memory, PGAS, MPI, and Streaming.]
Streaming

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

![Diagram showing Streaming concepts](image-url)

- Persistent optical circuits
- Input Data
- Output Data
- Shared Memory
- PGAS
- MPI
- Streaming
- Implicit Communication
- Explicit Communication
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
Wavelength Division Multiplexing

waveguide

\[ \lambda \]
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)

- C23 (1559 nm)
- C28 (1555 nm)
- C46 (1541 nm)
- C51 (1537 nm)

Before injection into waveguide

After 5-cm waveguide and EDFA

Low-loss (1.7 dB/cm), high-bandwidth (> 200 nm) silicon photonic waveguides can be fabricated in commercial CMOS process.

[Silicon photonic waveguide technology diagram]

[Vlasov and McNab, Optics Express 12 (8) 1622 (2004)]

[B. G. Lee et al., Photon. Technol. Lett. 20 (10) 767 (2008)]
Ring Resonator Operation

modulator/filter

Broadband spatial switch
Silicon Photonic Modulator and Detector Technology

- 18 Gb/s demonstrated

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

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
Higher Order Switch Designs

On-Chip Topology Exploration

- **Photonic Torus**

- **Nonblocking Photonic Torus**

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

[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

Passing by a ring
-0.005 dB each

Waveguide propagation
-1.5 dB/cm

1 dBm

0.68 dBm
Insertion Loss vs. Bandwidth

Number of λ vs. Network Size

Topologies
Simulation Results

Torus Topology

Non-Blocking Torus Topology

TorusNX Topology

Square Root Topology

Insertion Loss (dB)

Topology Size (nodes)

Propagation  |  Crossing  |  Dropping Into a Ring
Simulation Results

*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
- Intra-Message Crosstalk
- Coherent noise
- Incoherent noise
Effects of Noise

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

Results
• Results are plotted for network size of 8×8 at saturation, at the detectors.
  • Maximum OSNR = ~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

![Diagram showing power usage concepts]

- **Electronic Control**
  - p-region
  - n-region
- **Ohmic Heater**
- **Thermal Control**

- **Transmission**
  - Off-resonance profile
  - On-resonance profile
  - Injected Wavelengths
Energy Per Bit

![Graph showing energy per bit vs. message size for different network topologies: Torus, Non-blocking Torus, TorusNX, and Square Root. The x-axis represents message size in bits, ranging from $10^0$ to $10^7$. The y-axis represents energy per bit in joules per bit ($J$/bit), ranging from $10^{-13}$ to $10^{-7}$. The graph compares the energy efficiency of the different network configurations at various message sizes.]
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%

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

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

- 38 wavelengths @ 10 Gbps/each
- Power Dissipation = 3.22 W
Performance
Performance

- Uniform random traffic
- 256 cores, 64-node network
Scientific Applications

![Graph showing Execution Time vs Energy for Scientific Applications]

- Execution Time (s)
  - Cactus
  - GTC
  - MADbench
  - PARATEC

- Energy (J)
  - Cactus
  - GTC
  - MADbench
  - PARATEC

Bars represent E-Mesh and P-Mesh for each application.
Other Interesting Issues
Memory Access

- Processor Core
- Network Router
- Memory Access Point

[G. Hendry et al. Circuit-Switched Memory Access in Photonic Interconnection Networks for HPEC. In Supercomputing, Nov. 2010]
Other Arbitration Means - TDM

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
  - Photonic physical layout / design
  - System-level benefits from device improvement
  - Network-level improvements