Silicon photonics and memories

Vladimir Stojanović
Integrated Systems Group, RLE/MTL
MIT
Acknowledgments

- Krste Asanović, Christopher Batten, Ajay Joshi
  - Scott Beamer, Chen Sun, Yon-Jin Kwon, Imran Shamim
- Rajeev Ram, Milos Popovic, Franz Kaertner, Judy Hoyt, Henry Smith, Erich Ippen
  - Hanqing Li, Charles Holzwarth
  - Jason Orcutt, Anatoly Khilo, Jie Sun, Cheryl Sorace, Eugen Zgraggen
  - Michael Georgas, Jonathan Leu, Ben Moss

- Dr. Jag Shah – DARPA MTO
- Texas Instruments
- Intel Corporation
Processors scaling to manycore systems

- 64-tile system (64-256 cores)
  - 4-way SIMD FMACs @ 2.5 – 5 GHz
  - 5-10 TFlops on one chip
  - Need 5-10 TB/s of off-chip I/O
  - Even larger bisection bandwidth
Bandwidth, pin count and power scaling

- 8 Flops/core @ 5GHz
- Need 16k signal pins in 2017 for HPC
- 1 Byte/Flop
- 256 cores
- 2 TFlop/s signal pins
- 2,4 cores
- Package pin count
- Limit for entire mobile client chip
- Limit for entire server or HPC chip
Electrical Baseline in 2016

Node Board
10 TFlop/s
512 GB DRAM
80 Tb/s mem BW

CPU Power 1kW -> 100W
Energy-efficiency
100 pJ/Flop -> 10pJ/Flop

Memory Power 1kW

200 W
400 W
1kW

Cross-chip
I/O
Compute

200 W
400 W
400 W

Request
Response

64 memory channels
(controllers)
1.28 Tb/s per controller
160 Gb/s per chip
(16 x 10 Gb/s) @ 5pJ/b

512 x 1GB DRAM chips
8 chips per DIMM
1DIMM per memory channel
Need at least 16 banks/chip to sustain BW

Processors
Router
Memory
Controller

I/O
Cross-chip
Activate
Monolithic CMOS-Photonics in Computer Systems

Si-photonics in advanced bulk CMOS, thin BOX SOI and DRAM process NO costly process changes

Embedded apps

Bandwidth density – need dense WDM
Energy-efficiency – need monolithic integration
### CMOS photonics density and energy advantage

#### Metric | Energy (pJ/b) | Bandwidth density (Gb/s/μ)
--- | --- | ---
Global on-chip photonic link | 0.1-0.25 | 160-320
Global on-chip optimally repeated electrical link | 1 | 5
Off-chip photonic link (100 μ coupler pitch) | 0.1-0.25 | 6-13
Off-chip electrical SERDES (100 μ pitch) | 5 | 0.1

Assuming 128 10Gb/s wavelengths on each waveguide
But, need to keep links fully utilized ...

Fixed and static energy increase at low link utilization!
Core-to-Memory network: Electrical baseline

- Both cross-chip and I/O costly
Aggregation with Optical LMGS* network

* Local Meshes to Global Switches

Ci = Core in Group i, DM = DRAM Module, S = Crossbar switch

- Shorten cross-chip electrical
- Photonic both part cross-chip and off-chip
Network layout optimization significantly affects the component requirements.

64-tile system w/ 16 groups, 16 DRAM Modules, 320 Gbps bi-di tile-DRAM module BW

[Tile]

32 Waveguides (64λ/direction)

[Joshi et al – PICA 2009]
Photonic LMGS - U-shape

64-tile system w/ 16 groups, 16 DRAM Modules, 320 Gbps bi-di tile-DRAM module BW
Photonic LMGS - U-shape

64-tile system w/ 16 groups, 16 DRAM Modules, 320 Gbps bi-di tile-DRAM module BW
Photonic LMGS - U-shape

64-tile system w/ 16 groups, 16 DRAM Modules, 320 Gbps bi-di tile-DRAM module BW
Photonic LMGS - U-shape

- 64 tiles
- 64 waveguides (for tile throughput = 128 b/cyc)
- 256 modulators per group
- 256 ring filters per group
- Total rings > 16K $\rightarrow$ 0.32W (thermal tuning)
Photonic device requirements in LMGS - U-shape

Through loss (dB/ring)

Waveguide loss (dB/cm)

Through loss limits for 2 W optical laser power

Optical Laser Power

Die Area Overhead

Waveguide loss and Through loss limits for 2 W optical laser power
Photonic LMGS – ring matrix vs u-shape

LMGS – ring matrix

- 0.64 W power for thermal tuning circuits
- 2 W optical laser power
  - Waveguide loss < 0.2 dB/cm
  - Through loss < 0.002 dB/ring

LMGS – u-shape

- 0.32 W power for thermal tuning circuits
- 2 W optical laser power
  - Waveguide loss < 1.5 dB/cm
  - Through loss < 0.02 dB/ring

[Batten et al – Micro 2009] [Joshi et al – PICA 2009]
Power-bandwidth tradeoff

![Graphs showing latency and power consumption for different bandwidths and processing configurations.](image)

- 2-3x better
- 8-10x better

**Electrical with grouping and over-provisioning**

**Optical with grouping and over-provisioning**
System Organization – Defragmentation

Example 256 core node – with 64 core dies

[Beamer et al – ICS 2009]
System Organization – Die view

8 Waveguides w/ 64λ each
8-Core Cluster
Cluster
Cluster
Cluster
Cluster
Cluster
Cluster
Cluster

Power Fiber

16 MC Ribbon
16 Cluster Ribbon

Modulators

Shared L2

Ring Filters

Photo-detectors

64 core die supporting 256 core node
Electrical DRAM is also Limited

- Pin-bandwidth on the compute chip
- I/O energy to move between chips
- Cross-chip energy within DRAM chip
- Activation energy within DRAM chip
Solution: Silicon Photonics

- Great bandwidth density
- Great off-chip energy efficiency
- Costs little additional energy to use on-chip after off-chip
- Enables page size reduction

[Beamer et al – ISCA 2010]
Current DRAM Structure
Photonics to the Chip

Electrical Baseline (E1)

Photonics Off-Chip w/Electrical On-Chip (P1)
Photonics Into the Chip

2 Data Access Points per Column (P2)

8 Data Access Points per Column (P8)
Reducing Activate Energy

- Want to activate less bits while achieving the same access width
- Increase number of I/Os per array core, which decreases page size
  - Compensate the area hit by smaller photonic off-chip I/O
Methodology

- Photonic Model - aggressive and conservative projections
- DRAM Model - Heavily modified CACTI-D
- Custom C++ architectural simulator running random traffic to animate models
- Setup is configurable, in this presentation:
  - 1 chip to obtain 1GB capacity with >500Gbps of bandwidth provided by 64 banks
Energy for On/Off-Chip
Reducing Row Size

4 I/Os per Array Core

32 I/Os per Array Core
Latency Not a Big Win

- Latency marginally better
- Most of latency is within array core
- Since array core mostly unchanged, latency only slightly improved by reduced serialization latency
Area Neutral

4 I/Os per Array Core

32 I/Os per Array Core

- I/O Overhead
- Inter-Bank Overhead
- Intra-Bank Overhead
- Memory Cells
**Motivation:** allow the system to increase capacity without increasing bandwidth

**Disadvantage:** high path loss (grows exponentially) due to couplers and waveguide
Split Photonic Bus

- **Advantage**: much lower path loss
- **Disadvantage**: all paths lit
Guided Photonic Bus

- **Advantage:** only 1 low loss path lit
With Photonics...

- 10x memory bandwidth for same power
- Higher memory capacity without sacrificing bandwidth
- Area neutral
- Easily adapted to other storage technologies
Conclusion

- Computer interconnects are very complex micro-communication systems
- Cross-layer design approach is needed to solve the on-chip and off-chip interconnect problem
  - Most important metrics
    - Bandwidth-density (Gb/s/um)
    - Energy-efficiency (mW/Gb/s)
  - Monolithic CMOS-photonics can improve the throughput by 10-20x
  - But, need to be careful
    - Optimize network design (electrical switching, optical transport)
    - Use aggregation to increase link utilizations
    - Optimize physical mapping (layout) for low optical insertion loss
Backup Slides
Photonic Technology

- Monolithically integrated silicon photonics being researched by MIT Center for Integrated Photonic Systems (CIPS)

Orcutt et al., CLEO 2008

Holzwarth et al., CLEO 2008
Photonic Link

- Each wavelength can transmit at 10Gbps
- Dense Wave Division Multiplexing (DWDM)
  - 64 wavelengths per direction in same media

**Rough Comparison**

<table>
<thead>
<tr>
<th></th>
<th>Electrical</th>
<th>Photonic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off-Chip I/O Energy (pJ/bit)</td>
<td>5</td>
<td>0.150</td>
</tr>
<tr>
<td>Off-Chip BW Density (Tbps/mm²)</td>
<td>1.5</td>
<td>50.000</td>
</tr>
</tbody>
</table>
Resonant Rings

light not resonant

resonant light

resonant light w/ drop path

figures inspired by [Vantrease, ISCA ’08]
**Modulator** uses charge injection to change resonant wavelength

When resonant light passes it mostly gets trapped in ring
Ring Modulators

- **Modulator** uses charge injection to change resonant wavelength.
- When resonant light passes it mostly gets trapped in ring.

![Resonant Racetrack Modulator](image)

![Light Power in Ring](image)
Photonic Components
Why 5pJ/b for Electrical?

- Prior work has claimed lower than our forecasted 5pJ/b for off-chip electrical I/O
  - 2.24 pJ/b @ 6.25Gbps (Palmer et al., ISSCC 2007)
  - 1.4 pJ/b @ 10Gbps (O’Mahony et al., ISSCC 2010)

- Some important differences to consider:
  - We assume 20Gbps per pin
    - Otherwise will definitely be pin limited
  - At higher data rates it is hard to be as energy efficient: 8-13pJ/b @ 16Gbps (Lee et al., JSSC 2009)

- DRAM process has slower transistors leading to less energy efficient drivers

- Background energy averaged in (clocking, fixed energy, not 100% utilization)
Control Distribution

- Control distributed from the center of the chip
  - H-tree spreads out to banks
  - Can power gate control lines to inactive banks
Full Energy

- Aggressive
  - 64 Wavelengths, 4 I/Os
  - 64 Wavelengths, 32 I/Os
  - 8 Wavelengths, 32 I/Os

- Conservative
  - 64 Wavelengths, 4 I/Os
  - 64 Wavelengths, 32 I/Os
  - 8 Wavelengths, 32 I/Os

Energy (pJ/Bit)

Legend:
- Laser Write
- Laser Read
- Thermal Tuning
- Fixed Circuits
- Write
- Read
- Activate
Utilization

- Aggressive
  - 64 Wavelengths, 4 I/Os
  - 64 Wavelengths, 32 I/Os
  - 8 Wavelengths, 32 I/Os

- Conservative
Full Area

64 Wavelengths, 4 I/Os

64 Wavelengths, 32 I/Os

8 Wavelengths, 32 I/Os
Full Scaling

- Aggressive
- Conservative