Silicon Photonics for Exascale Computing

Columbia University
Keren Bergman
Lightwave Research Laboratory
TOP500 (June 2011): Performance Trend

Expected Exascale performance ~ 2019
Requires >120× performance increase in about 8 years
# TOP500 (June 2011): Energy Efficiency

<table>
<thead>
<tr>
<th>Rank</th>
<th>System</th>
<th>Rating (MW, GF/W)</th>
<th>Country/Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>K Computer</td>
<td>9.9 (0.82)</td>
<td>Japan (Fujitsu)</td>
</tr>
<tr>
<td>#2</td>
<td>Tianhe-1A</td>
<td>4.0 (0.64)</td>
<td>China (NUDT)</td>
</tr>
<tr>
<td>#3</td>
<td>Jaguar</td>
<td>7.0 (0.25)</td>
<td>United States (Cray Inc.)</td>
</tr>
<tr>
<td>#4</td>
<td>Nebulae</td>
<td>2.6 (0.49)</td>
<td>China (Dawning)</td>
</tr>
<tr>
<td>#5</td>
<td>TSUBAME 2.0</td>
<td>1.4 (0.85)</td>
<td>Japan (NEC/HP)</td>
</tr>
<tr>
<td>#6</td>
<td>Cielo</td>
<td>4.0 (0.28)</td>
<td>United States (Cray Inc.)</td>
</tr>
<tr>
<td>#7</td>
<td>Pleiades</td>
<td>4.1 (0.27)</td>
<td>United States (SGI)</td>
</tr>
<tr>
<td>#8</td>
<td>Hopper</td>
<td>2.9 (0.36)</td>
<td>United States (Cray Inc.)</td>
</tr>
<tr>
<td>#9</td>
<td>Tera-100</td>
<td>4.6 (0.23)</td>
<td>France (Bull SA)</td>
</tr>
<tr>
<td>#10</td>
<td>Roadrunner</td>
<td>2.3 (0.44)</td>
<td>United States (IBM)</td>
</tr>
</tbody>
</table>

No current Petascale system above 1 GF/W
Exascale system at 1 GF/W requires 1 GW of power
TOP500 (June 2011): Energy Efficiency

Three systems with energy efficiency above 1 GF/W
Highest two are IBM Blue Gene/Q systems (2.10 and 1.68 GF/W)
Interconnects significant fraction of power budget (~20%-40%)
Why Photonics is a good idea?

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.
But Why Really Photonics?

![Graph showing memory pin count, power consumption, and double precision computation over years.](image)

- ITRS Projection
- Memory Pin Count
- Power (W)
- Double Precision Computation (GFLOPS)
- Data points for UltraSPARC T2, Nehalem, Athlon64FX, Power XCell, SandyBridge, and POWER7.
Silicon Photonics

Silicon-on-insulator (SOI) platform produces valuable photonic building blocks
High index contrast enables high confinement, low-loss propagation, virtually lossless bending
CMOS compatibility allows monolithic integration with advanced microelectronics
Many active and passive functionalities have already been demonstrated
Optically Interconnected High-Performance Embedded Board

3D stack with CMPs, memory, and photonic NoC

Silicon Photonic Interconnection Network

Memory Stack

CMPs

DRAM

CMPs 3D Stack
Crystalline Silicon Waveguides

1.28 Tb/s (32 x 40 \(\lambda\))

- 1.28 Tb/s through 5-cm waveguide
- 32 40-Gb/s wavelength channels
- 24 90°-bends (6.5-\(\mu\)m bending radii)
- Inverse-taper mode converters with index-matching polymer
- Error-free transmission

Propagation loss: 3 dB/cm
Aggregate data rate: 1.28 Tb/s

B. G. Lee, PTL 2008
Waveguide Crossings

W. Bogaerts (2007); Ghent: waveguide broadening and double-etch technique
Insertion loss: 0.16 dB/crossing
Crosstalk: < –40 dB

F. Xu (2008); UST: adiabatic tapers and multi-mode-interference method
Insertion loss: 0.12 dB/crossing
Crosstalk: < –40 dB
Back-reflection: –45 dB

M. A. Popović (2007); MIT: non-adiabatic tapers and periodically-matched multi-mode focusing
Insertion loss: 0.045 dB/crossing
Crosstalk: –40 dB
Microring Resonator Modulator

Crystalline Silicon Microring Resonator Electro-Optic Modulator

- Waveguide (450-nm × 260-nm)
- Input
- Output
- p-Region
- n-Region
- Microring Resonator (12-μm Diameter)
- Slab (50-nm Tall)
- 2 μm
WDM Transmission with Microring Resonator Modulator Array

Crystalline Silicon Microring Resonator Modulator Array

Continuous-Wave Light Wavelength Channels $\lambda_1 - \lambda_M$

Optical Data Stream Wavelength Channels $\lambda_1 - \lambda_M$

Encode $\lambda_1$  
Encode $\lambda_2$  
Encode $\lambda_3$  
Encode $\lambda_M$
Multiwavelength Switching Building Blocks

1×2 Switch

Through State

Drop State

2×2 Switch

Cross State

Bar State
Crystalline Silicon Photonic Switch

Diagram showing a Crystalline Silicon Photonic Switch with Through Port, Slab, and Drop Port. The diagram includes labels for Left Cavity, Right Cavity, and Input Port with a scale of 10 μm.

Graphs showing optical traces at different speeds: 5 Gb/s, 10 Gb/s, 20 Gb/s, and 40 Gb/s.
Example Electro-Optic Silicon Photonic Switch

High-Speed Optical Domain Routing (< 0.4dB Power Penalty)


Courtesy M. R. Watts, MIT
4×4 Microring Resonator Broadband Router

Wavelength-Parallel Switching Configuration

A. Biberman, PTL 2010
Example Silicon Photonic Germanium Photodetector

Germanium Detectors

Moore's Law
Silicon Photonics
The Remaining Challenges

Filtering
Polarization
Modulation, Switching, and Control
Detection and Integration

Diagram showing responsivity and transmission as functions of applied voltage and time.

Courtesy M. R. Watts, MIT
Full Photonic Link

Scanning-Electron-Microscope Image of Silicon Photonic Link

- Crystalline silicon microring resonator electro-optic modulator
- Crystalline silicon waveguide
- Germanium metal-semiconductor-metal photodetector

Error-free operation at 3-Gb/s, 180fJ/bit photonic transmission

N. Ophir, OFC 2011
Example Silicon Photonic Thermal Stabilization

Challenge III: Thermal Stabilization

Achieving and Maintaining Resonance Position Challenging

- **Thermal Sensitivity:** Thermo-optic coefficient $\frac{df}{dT} \approx 10 \text{ GHz/K}$
- **Requirements:** Systems require a 0-to-50°C operating range
- **Challenge:** Can we overcome uniformity and thermal control with one integrated and electronic solution?
Electrical Thermo-Optic Microring Resonator Tuning Functionality

N. Sherwood-Droz, OE 2008
Inverse-tapered crystalline silicon waveguide with CMOS-compatible V-groove structure. 7.5-dB coupling loss, spanning over 70 nm of the spectrum (may be reduced to about 1 dB with improved optical mode matching).

J. V. Galan, GFP 2009
Packaging Silicon Photonics: A Challenge
Solution 1: Later Couplers

CMOS-compatible SiON, efficient spot-size converting waveguide couplers
1-dB coupling loss, −35-dB crosstalk with eight-channel coupling
8-Tb/s/mm bandwidth densities with a 20-µm pitch

IBM: F. E. Doany, JLT 2011
Packaging Silicon Photonics: A Challenge
Solution 2: Vertical Couplers

Grating structures for vertical coupling
Commercial fiber array connected to silicon photonic chip with many grating coupler ports
Packaged with electrical connections

Luxtera: C. Gunn, Micro 2006
J. V. Galan, GFP 2009; Zimmermann, JSTQE 2011
Example Silicon Photonic Integration with CMOS

Direct Integration with CMOS

Results

- **Approach:** Integrated modulators directly into Sandia’s CMOS process enabling dense, low-capacitance drive.
- **Results:** Successful integration with “error-free” operation at 2Gb/s, limited by CMOS speed.

Courtesy M. R. Watts, MIT
Example Silicon Photonic Integration with CMOS

- A) CMOS front-end process flow
- B) CMOS die with integrated CMOS and silicon photonics

IBM: W. M. J. Green, IPRSN 2011; S. Assefa OFC 2011
Example Silicon Photonic Integration with CMOS

- A) CMOS digital circuits (ring oscillator)
- B) CMOS analog circuits (transimpedance amplifier)
  - C) Modulator driver amplifier
  - D) Cascaded Mach-Zehnder WDM

IBM: W. M. J. Green, IPRSN 2011; S. Assefa OFC 2011
Example: Photonic Interconnect Design
Exascale System – Single Rack

Lightwave Interconnection Broadband Routing Architecture (LIBRA)
Power Calculations

Rack performance: 1.28 PF
Bisectional bandwidth: 2.2 Pb/s
Example: Photonic Interconnect Design

Lightwave Interconnection Broadband Routing Architecture (LIBRA) Power Calculations

On-Chip Transceiver: Transmitter

- Light from laser 128 wavelengths
- On-Chip Electrical Router
- I/O
- Waveguide < 0.5 dB/cm

128 Modulator Bank
- 22-Gb/s Each
- < 25 fJ/bit
- 3-µm Diameter

128 modulated wavelengths
- 2.816 Tb/s
Example: Photonic Interconnect Design

Lightwave Interconnection Broadband Routing Architecture (LIBRA) Power Calculations

On-Chip Transceiver: Receiver

- 128 wavelength channels
  - 2.816 Tb/s
- Waveguide
  - < 0.5 dB/cm
- On-Chip Electrical Router
- I/O
- 128 Photodetectors
  - 22-Gb/s Each
  - < -30 dBm
  - 3-µm Diameter
Example: Photonic Interconnect Design

Lightwave Interconnection Broadband Routing Architecture (LIBRA) Power Calculations

Broadband Circuit Switch

128 wavelength channels: 2.816 Tb/s

Logic

128 wavelength channels: 2.816 Tb/s

One Output Port Selected

128 wavelength channels: 2.816 Tb/s

Broadband Packet Switch

128 wavelength channels: 2.816 Tb/s

Delay Module

Logic

Rx

128 wavelength channels: 2.816 Tb/s

One Output Port Selected

128 wavelength channels: 2.816 Tb/s
Example: Photonic Interconnect Design

Lightwave Interconnection Broadband Routing Architecture (LIBRA) Power Calculations

Photonic Broadband Switching

Max Insertion Loss: 0.75 dB  2.25 dB  5.25 dB  6.75 dB  11.25 dB

Max Propagation Latency: 100 ps  300 ps  700 ps  900 ps  1.5 ns
**Example: Photonic Interconnect Design**

Lightwave Interconnection Broadband Routing Architecture (LIBRA) Power Calculations

<table>
<thead>
<tr>
<th>Assumptions</th>
<th>Data Rate (1 λ) (Gb/s)</th>
<th>Packet Size (bits/transition)</th>
<th>Total Switches</th>
<th>Wavelength Channels</th>
<th>Switch Duty Cycle</th>
<th>Bandwidth per Transciever (Gb/s)</th>
<th>Spatial Channels</th>
<th>Ring Utilization Probability</th>
<th>Unidirectional Bisectonal Bandwidth (Gb/s)</th>
<th>Data Rate (WDM) (Gb/s)</th>
<th>Macro Switch Size Radix</th>
<th>Bidirectional Bisectonal Bandwidth (Gb/s)</th>
<th>Modulator Duty Cycle</th>
<th>Number of 1x2 Switches in Macro Switch</th>
<th>Peak PFLOPS</th>
<th>Message Size (B)</th>
<th>Total Transcievers</th>
<th>Byte/Second per Peak FLOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>8192</td>
<td></td>
<td>512</td>
<td>192</td>
<td>0.25</td>
<td>4224</td>
<td>1</td>
<td>0.5</td>
<td>1081344</td>
<td>4224</td>
<td>26</td>
<td>2162688</td>
<td>0.5</td>
<td>768</td>
<td>1.28</td>
<td>1024</td>
<td>640</td>
<td>0.2112</td>
</tr>
</tbody>
</table>

All assumptions are variables to understand all the design trade-offs
### Example: Photonic Interconnect Design

Lightwave Interconnection Broadband Routing Architecture (LIBRA) Power Calculations

<table>
<thead>
<tr>
<th></th>
<th>With Maximum Load Data, 1024-B Message Size</th>
<th>Total Device (fJ/bit)</th>
<th>Total Device (W)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transmitters</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Serializer</strong></td>
<td>4 mW</td>
<td></td>
<td>180</td>
</tr>
<tr>
<td><strong>Laser (per Wavelength)</strong></td>
<td>Static (mW)</td>
<td>Thermal (mW)</td>
<td>1</td>
</tr>
<tr>
<td><strong>Modulator</strong></td>
<td>Dynamic (fJ/bit) Static (µW)</td>
<td>Thermal (mW)</td>
<td>25</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td>433</td>
</tr>
<tr>
<td><strong>Photodetector</strong></td>
<td>Dynamic (fJ/bit) Thermal (mW)</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td><strong>Receivers</strong></td>
<td></td>
<td></td>
<td>180</td>
</tr>
<tr>
<td><strong>Deserializer</strong></td>
<td>4 mW</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td>321</td>
</tr>
<tr>
<td><strong>Switches</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1x2 Switch</strong></td>
<td>Dynamic (fJ/transition) Static (µW)</td>
<td>Thermal (mW)</td>
<td>550</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td>694</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td>1448</td>
</tr>
</tbody>
</table>

Total switch fabric power budget is a ~1.5 pJ/bit
Includes all transceivers and a 26x26 switch fabric
Example: Photonic Interconnect Design
Lightwave Interconnection Broadband Routing Architecture (LIBRA) Power Calculations

Total Power Versus Message Size

Message Size (B)

Total System Power (pJ/bit)
Example: Photonic Interconnect Design

Lightwave Interconnection Broadband Routing Architecture (LIBRA) Power Calculations

Total Power Versus Data Rate Per Wavelength Channel

- X-axis: Data Rate Per Wavelength Channel (Gb/s)
- Y-axis: Total System Power (pJ/bit)

The graph shows a trend where the total system power decreases as the data rate per wavelength channel increases.
Total optical interconnect power at 1.8 pJ/bit, is 3.8 kW
Rack performance: 1.28 PF
Bisectional bandwidth: 2.2 Pb/s

Total system power budget: 57 kW
Optical interconnect ~6.7% of total system power
Optically-interconnected Exascale System
Optically-Interconnected Exascale System

Number of clusters: 25
Number of racks: 400
Number of blades: 25600
Number of boards: 102400
Summary: Gaps and Challenges

- Silicon photonic – chip scale technology can deliver system wide high-bandwidth density interconnect performance with extreme energy efficiency

- Technology under development in small batch industry/academic labs – critical need for scaled fabrication of integrated CMOS-photonics

- Major gap in packaging, high bandwidth density off-chip interconnect, switch fabric, thermally robust
### Summary Table of Silicon Photonic Materials

<table>
<thead>
<tr>
<th>Material</th>
<th>Propagation Loss</th>
<th>Electrically-Active</th>
<th>Deposit Capability</th>
<th>Stable</th>
<th>CMOS Compatibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>c-Si</td>
<td>1.7 dB/cm</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Poly-Si</td>
<td>6.45 dB/cm</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Si₃N₄</td>
<td>0.1 dB/cm</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>a-Si:H</td>
<td>2 dB/cm</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

- Crystalline silicon (c-Si)
- Polycrystalline silicon (poly-Si)
- Silicon nitride (Si₃N₄)
- (Hydrogenated) Amorphous Silicon (a-Si:H)

A. Biberman, OFC 2011
Cross-Sectional View of 3D Silicon Photonic Stack

- Silicon Nitride
- Polycrystalline Silicon and Germanium
- Silicon Dioxide
- Metal

A. Biberman, JETC 2011
I. A. Young (2010); Intel: silicon nitride waveguides as carrier medium in high-speed communication links; monolithic integration of high-performance photonic and electrical elements using standard CMOS processes

**Propagation loss:** 1 dB/cm (at 1310 nm)

**Data rate:** > 20 Gb/s
Polycrystalline Silicon Photonic Devices

electrically-active polycrystalline silicon microring resonator modulators coupled to polycrystalline silicon waveguides

**Quality factors:** 3,400; **extinction ratios:** 16 dB

**Modulation rate:** 2.5 Gb/s; **modulation depth:** 10 dB

K. Preston, Optics Letters 2011
Multi-Layer Silicon Nitride Microring Resonator

- 60-µm-diameter microring resonator; 6.7-nm FSR
- 1-µm wide and 400-nm tall waveguides, 1-dB/cm propagation losses

A. Biberman, CLEO 2011
Multi-Layer Silicon Nitride Microring

A. Biberman, CLEO 2011
Multi-Layer Deposited Silicon Photonics

Switch Matrix for Both Types of Integration

- Silicon Nitride provides low-loss propagation of optical signals.
- Polycrystalline provides active devices for modulation, routing, and detection of optical signals.

A. Biberman, JETC 2011