Building manycore processor-to-DRAM networks using monolithic silicon photonics

Ajay Joshi†, Christopher Batten†, Vladimir Stojanović†, Krste Asanović‡

†MIT, 77 Massachusetts Ave, Cambridge MA 02139

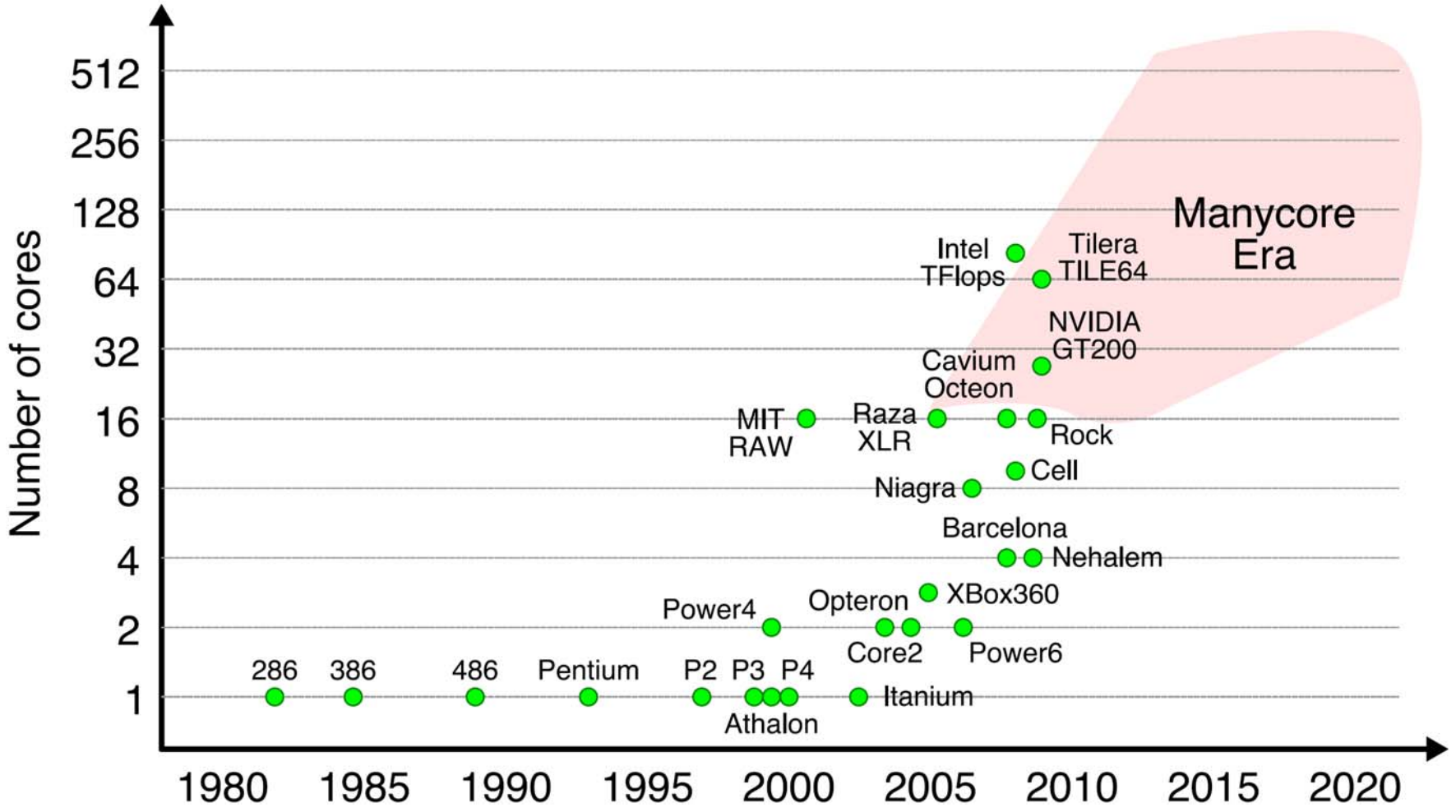
‡UC Berkeley, 430 Soda Hall, MC #1776, Berkeley, CA 94720

{joshi, cbatten, vlada}@mit.edu, krste@eecs.berkeley.edu

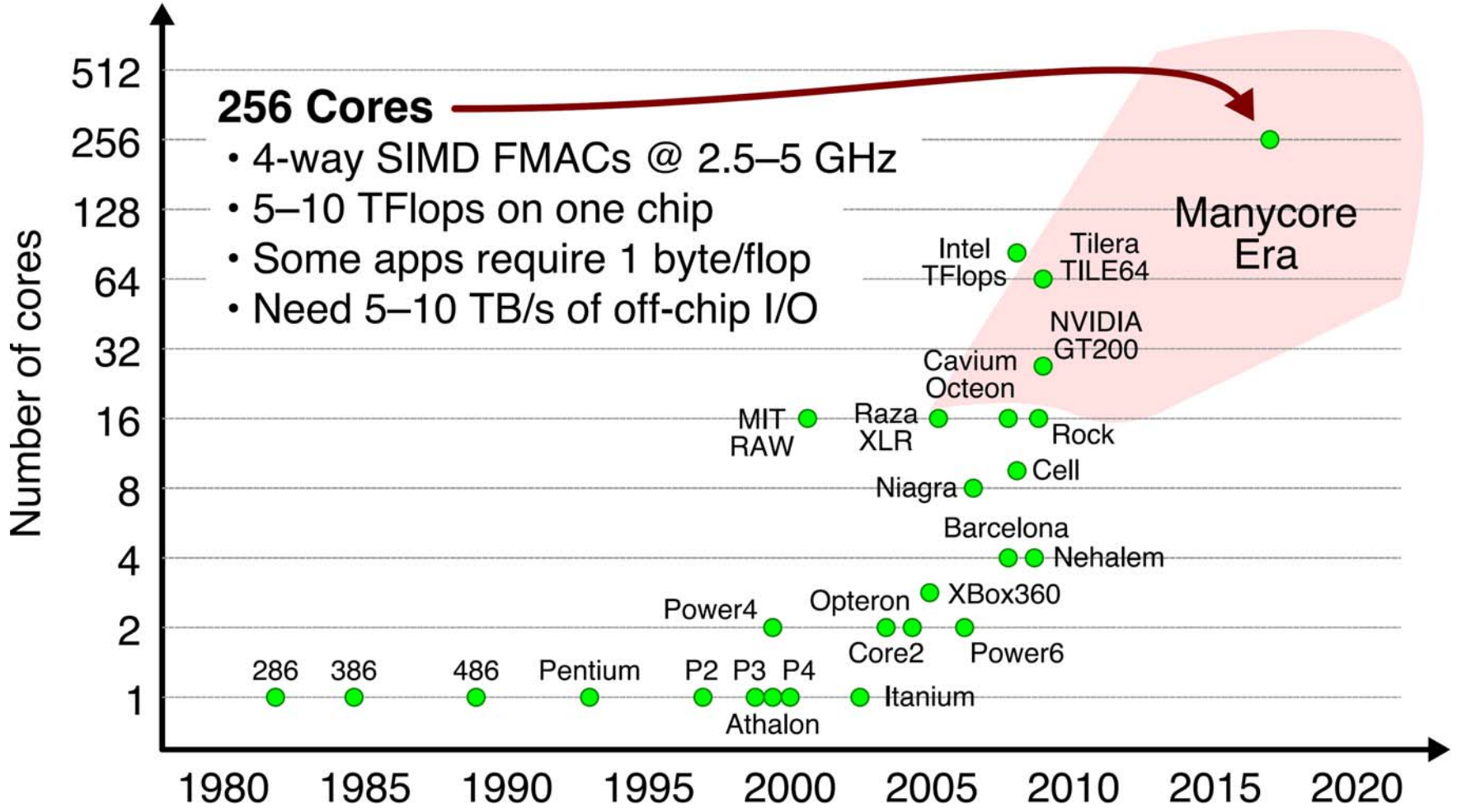
High Performance Embedded Computing (HPEC) Workshop

23-25 September 2008

Manycore systems design space

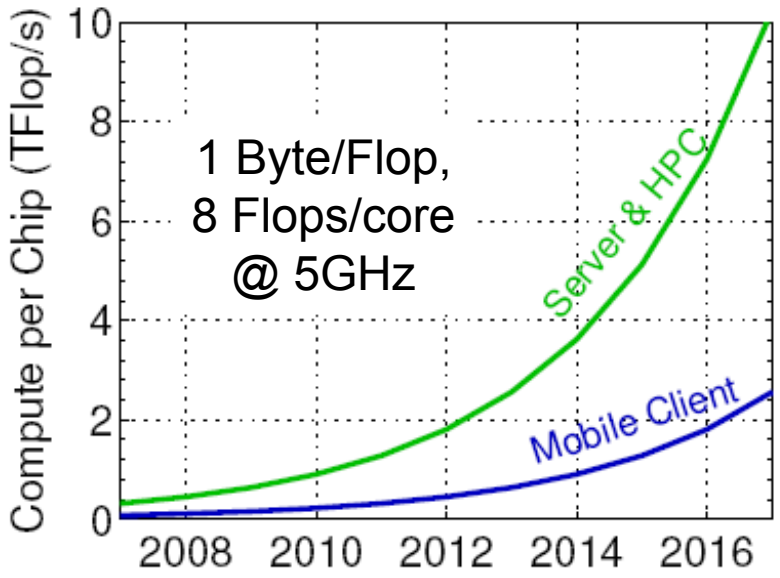


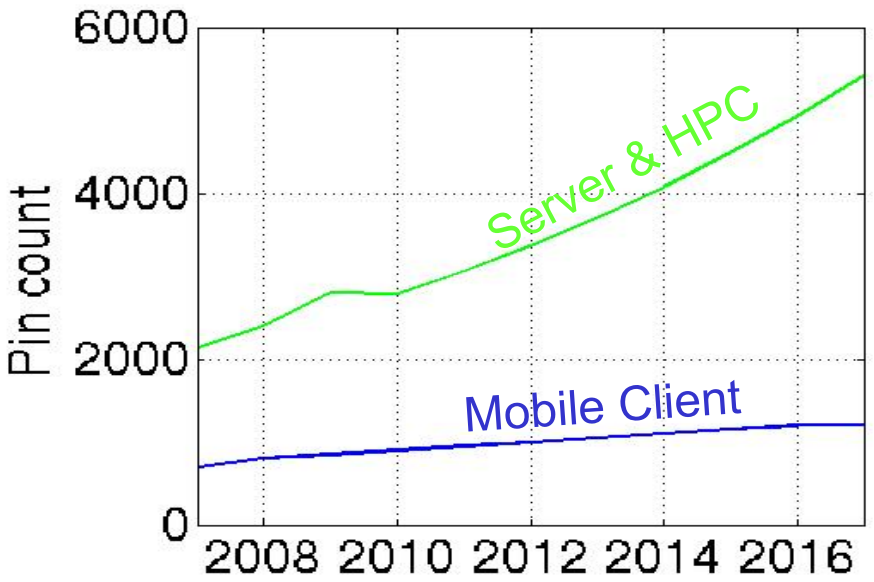
Manycore system bandwidth requirements

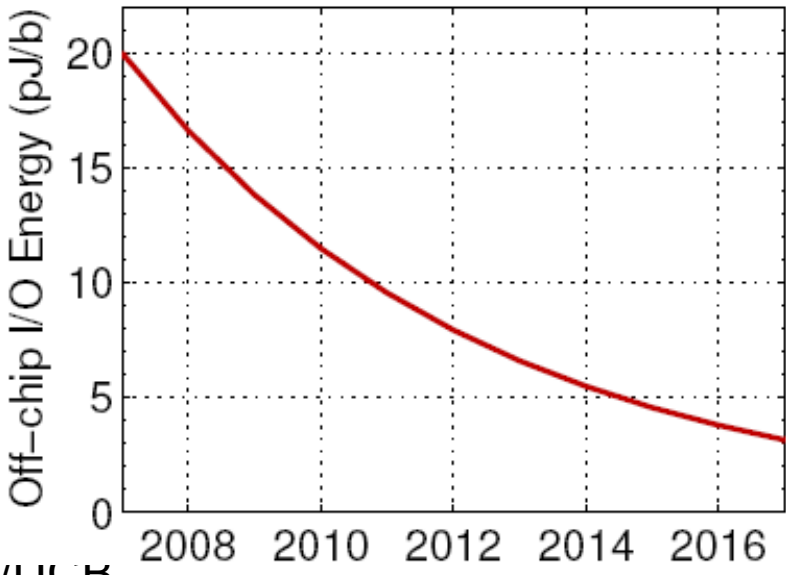


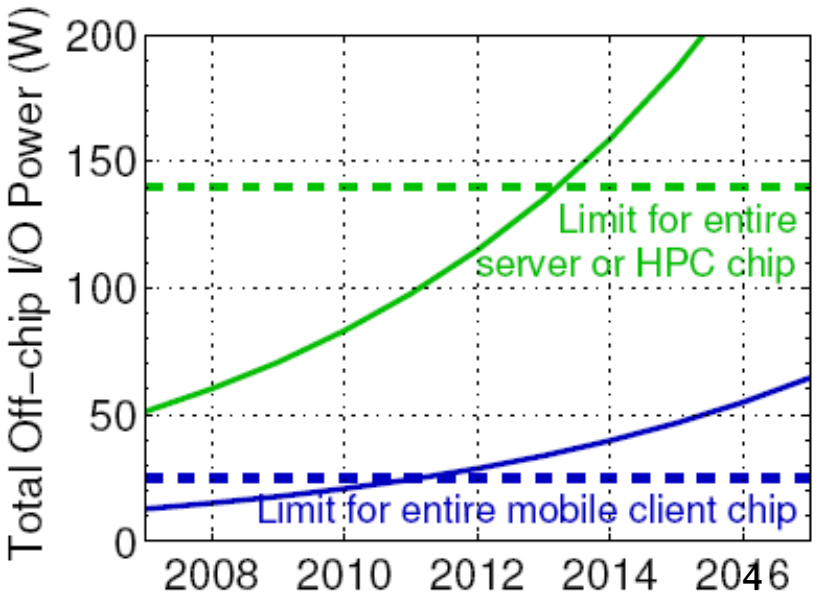
MIT/UCB

Manycore systems – bandwidth, pin count and power scaling

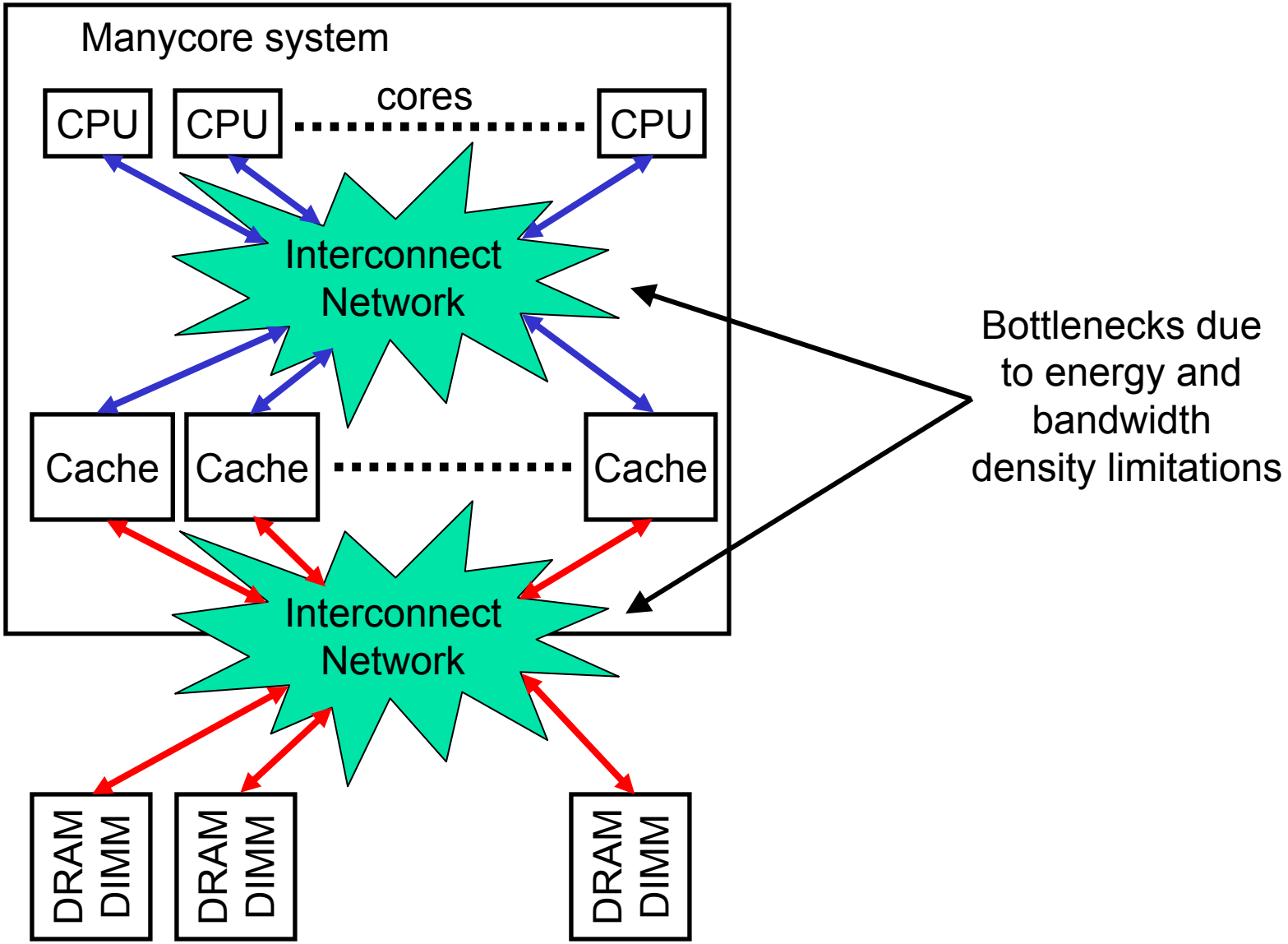




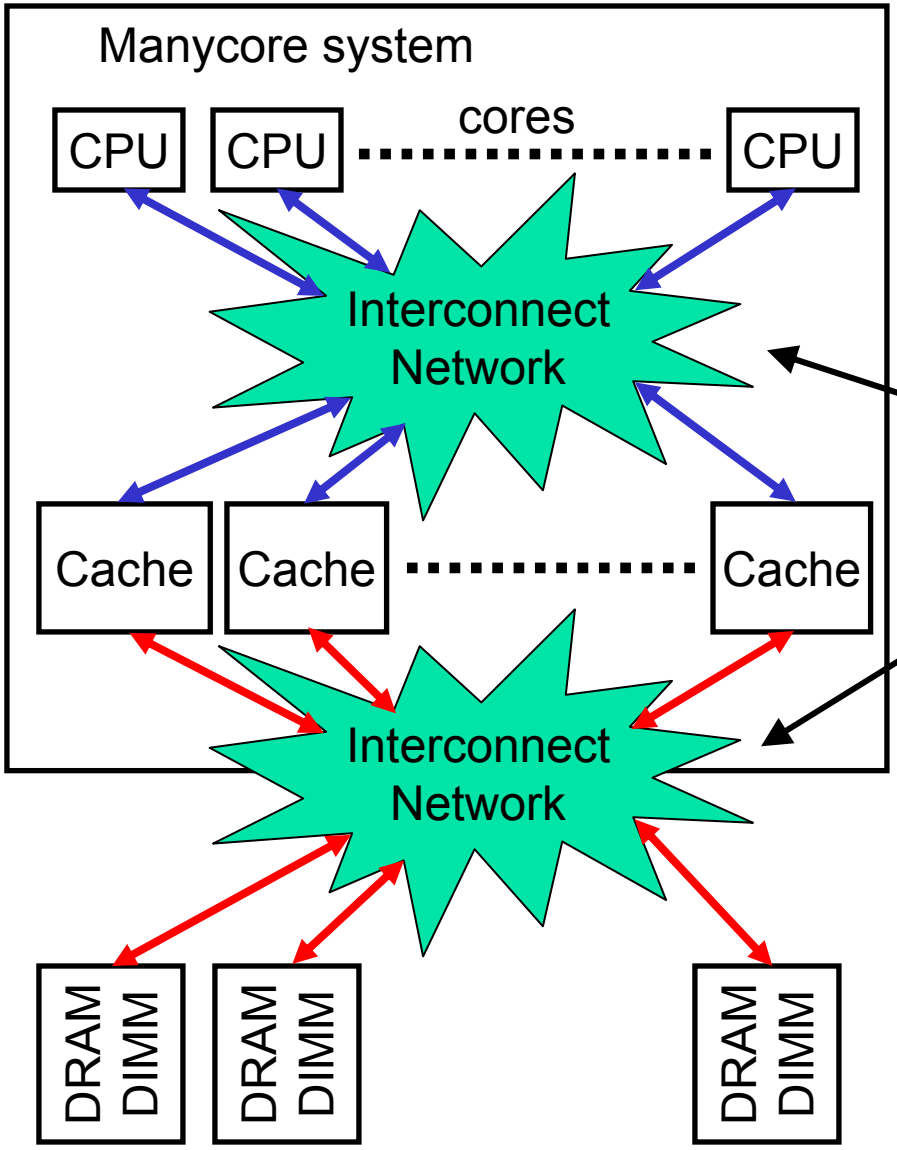




Interconnect bottlenecks



Interconnect bottlenecks



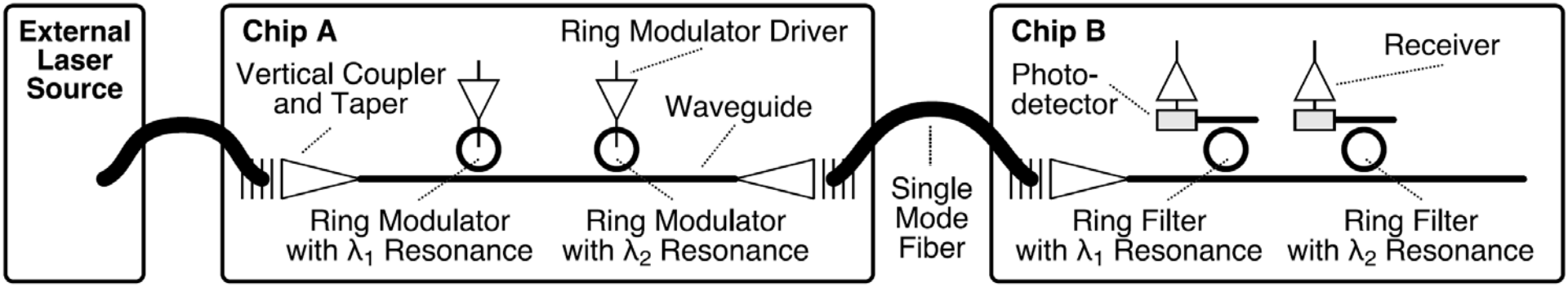
Bottlenecks due to energy and bandwidth density limitations

Need to **jointly** optimize on-chip and off-chip interconnect network

Outline

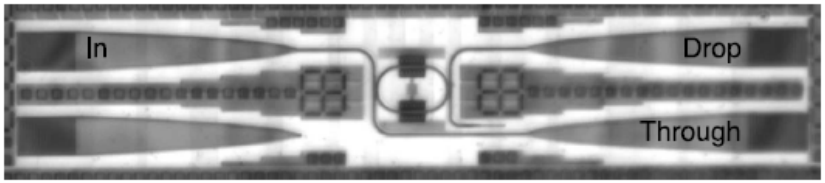
- Motivation
- Monolithic silicon photonic technology
- Processor-memory network architecture exploration
- Manycore system using silicon photonics
- Conclusion

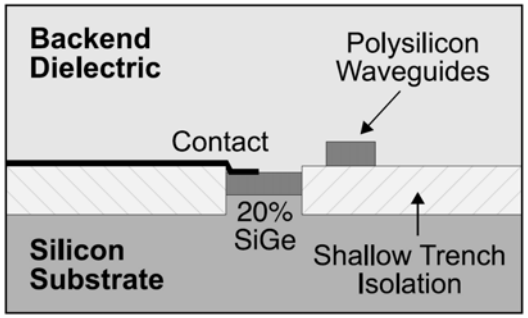
Unified on-chip/off-chip photonic link

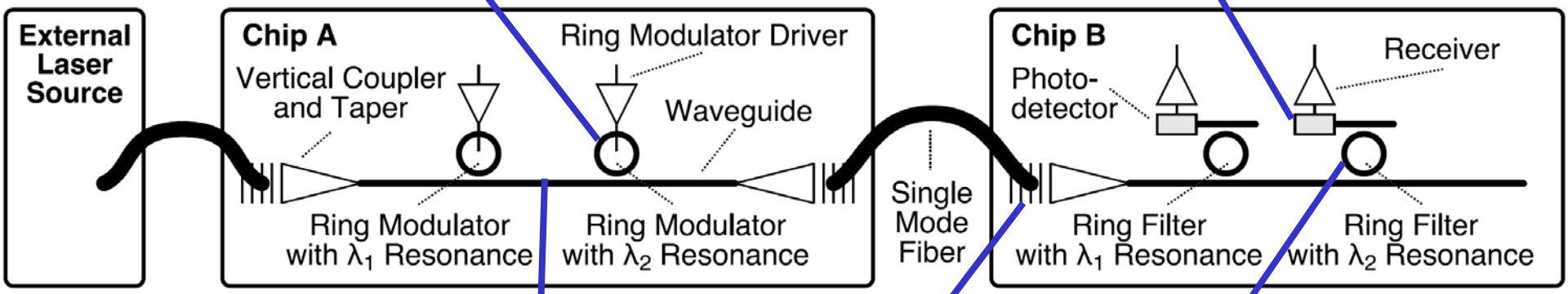


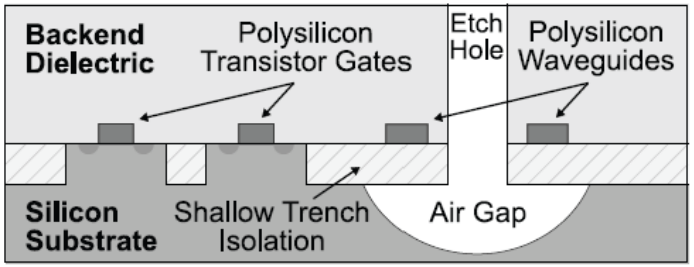
- Supports dense wavelength-division multiplexing that improves bandwidth density
- Uses monolithic integration that reduces energy consumption
- Utilizes the standard bulk CMOS flow

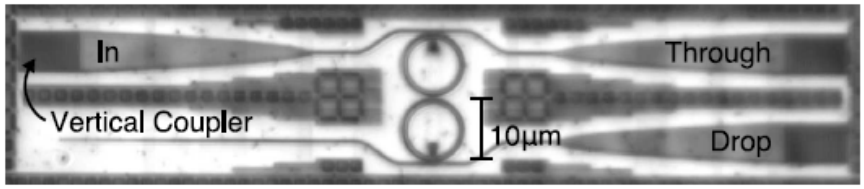
Optical link components





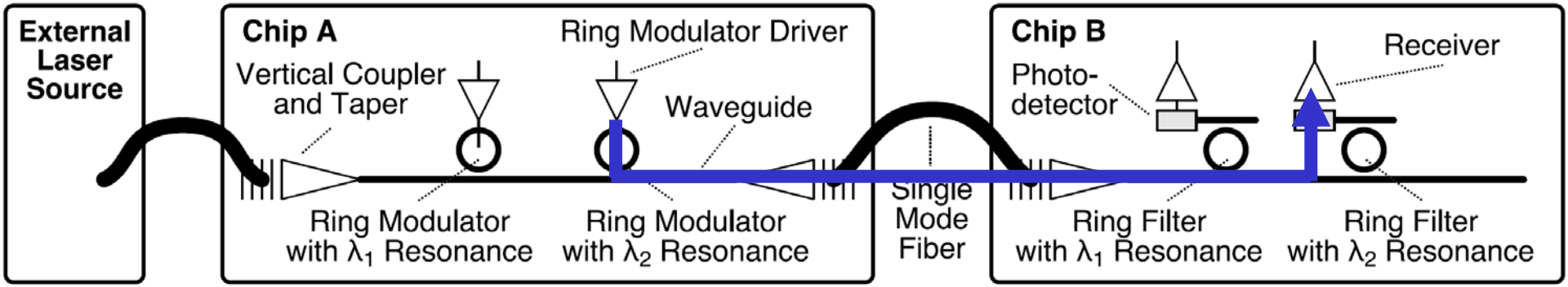






65 nm bulk CMOS chip designed to test various optical devices

Silicon photonics area and energy advantage



Metric	Energy (pJ/b)	Bandwidth density (Gb/s/μ)
Global on-chip photonic link	0.25	160-320
Global on-chip optimally repeated electrical link	1	5
Off-chip photonic link (50 μ coupler pitch)	0.25	13-26
Off-chip electrical SERDES (100 μ pitch)	5	0.1
On-chip/off-chip seamless photonic link	0.25	

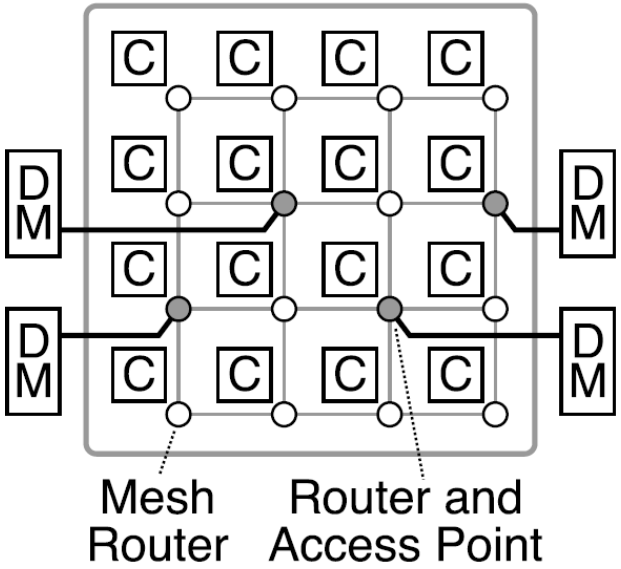
MIT/UCB

Outline

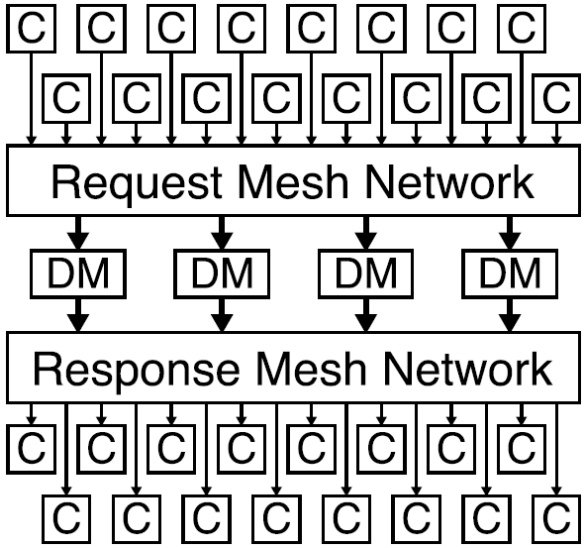
- Motivation
- Monolithic silicon photonic technology
- Processor-memory network architecture exploration
  - Baseline electrical mesh topology
  - Electrical mesh with optical global crossbar topology
- Manycore system using silicon photonics
- Conclusion

Baseline electrical system architecture

Mesh physical view



Mesh logical view

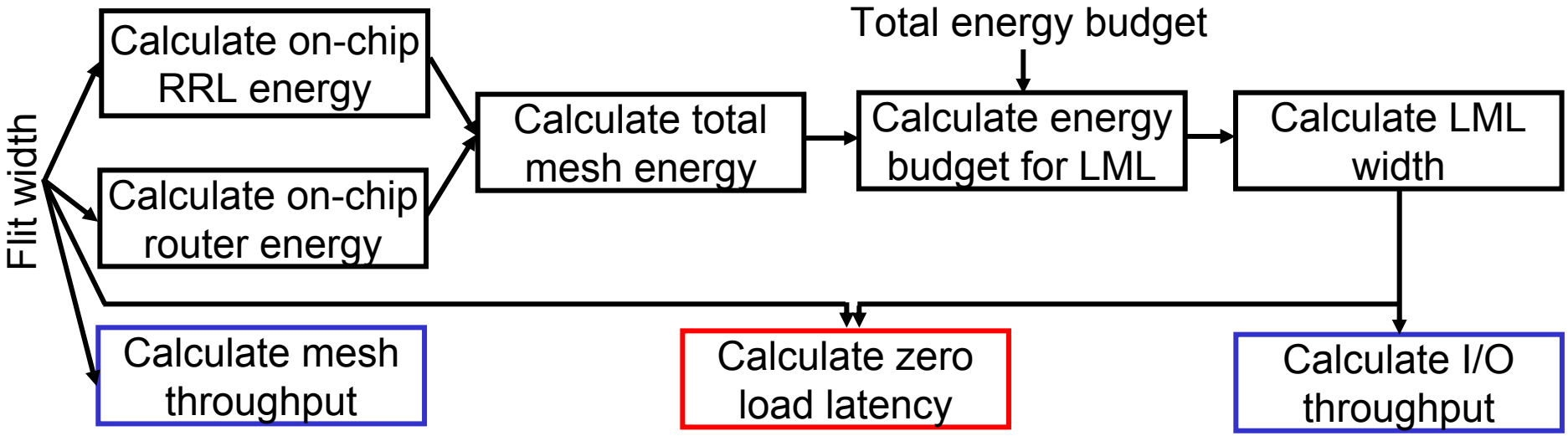


C = core, DM = DRAM module

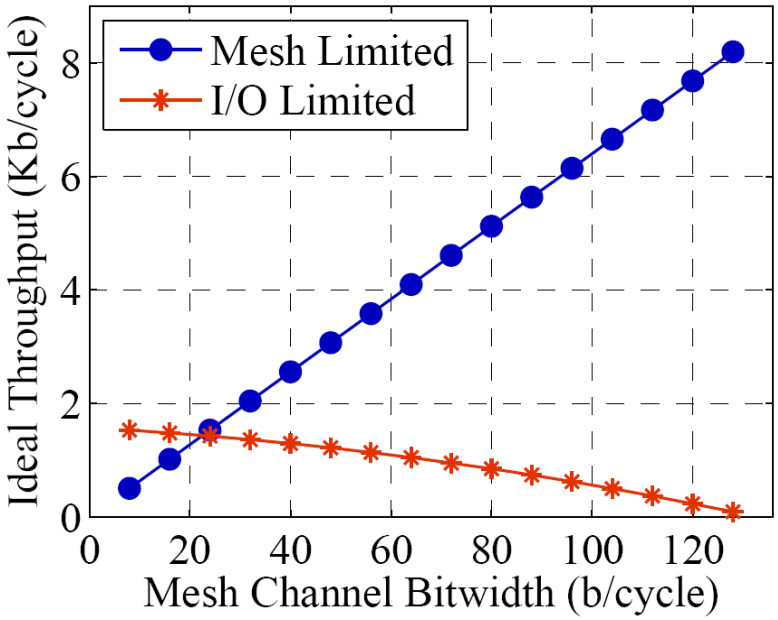
- Access point per DM distributed across the chip
- Two on-chip electrical mesh networks
  - Request path – core → access point → DRAM module
  - Response path – DRAM module → access point → core

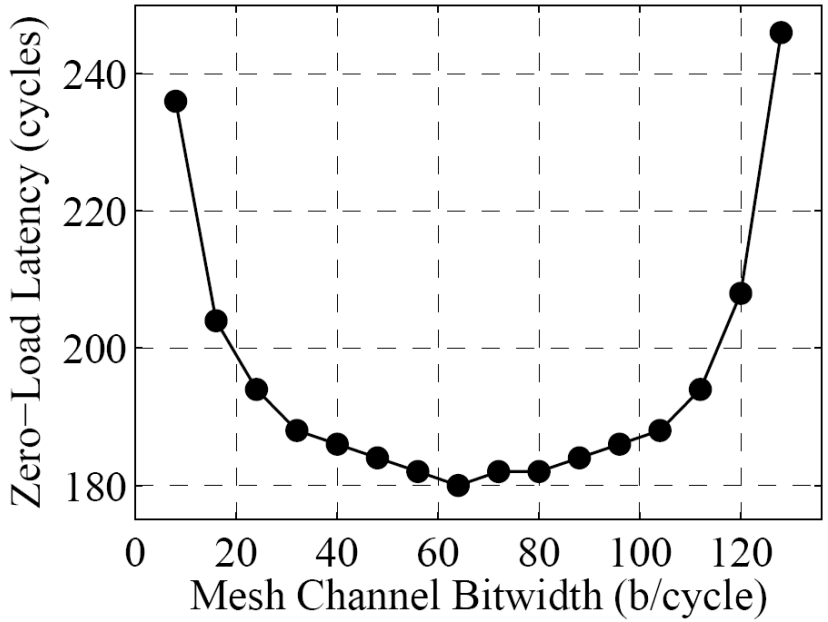
Interconnect network design methodology

- Ideal throughput and zero load latency used as design metrics
- Energy constrained approach is adopted
- Energy components in a network
  - Mesh energy (E_m) (router-to-router links (RRL), routers)
  - IO energy (E_{io}) (logic-to-memory links (LML))



Network throughput and zero load latency

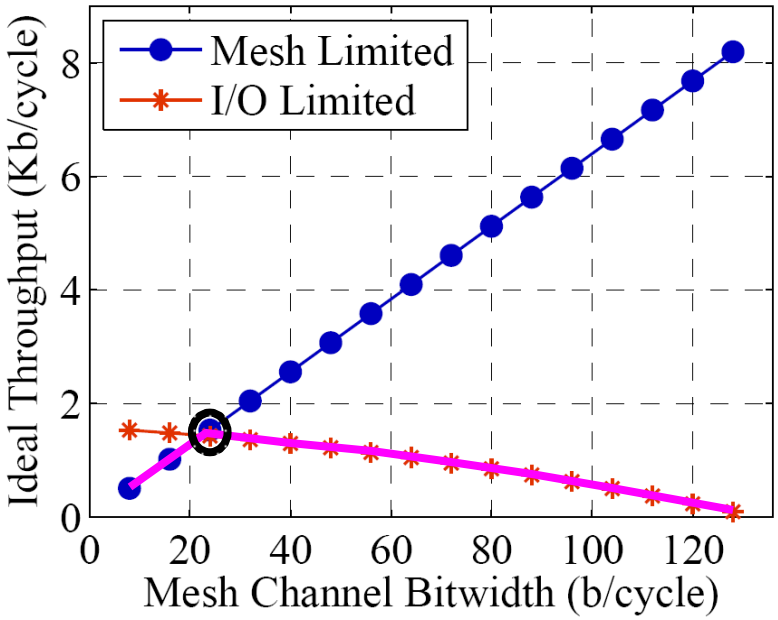


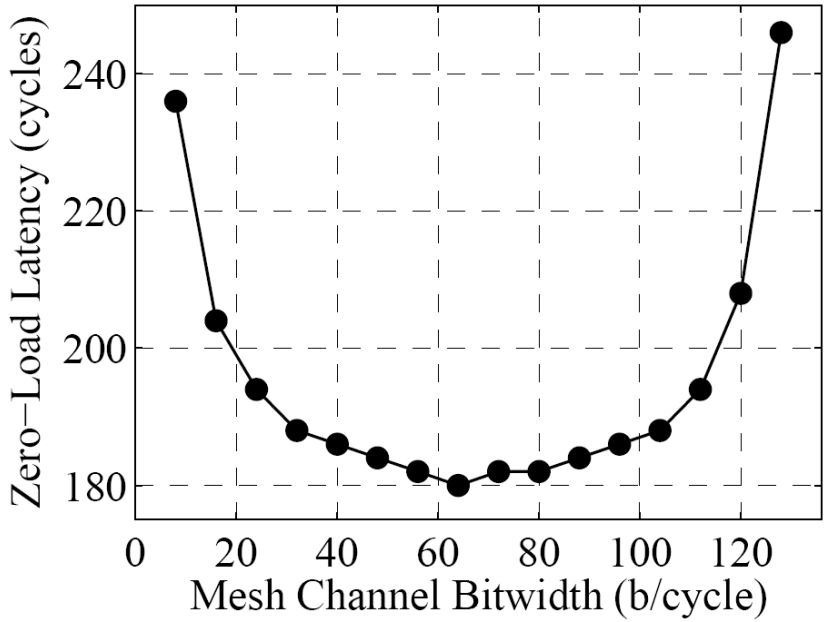


(22nm tech, 256 cores @ 2.5 GHz, 8 nJ/cyc energy budget)

- System throughput limited by on-chip mesh or I/O links
- On-chip mesh could be over-provisioned to overcome mesh bottleneck
- Zero load latency limited by data serialization

Network throughput and zero load latency

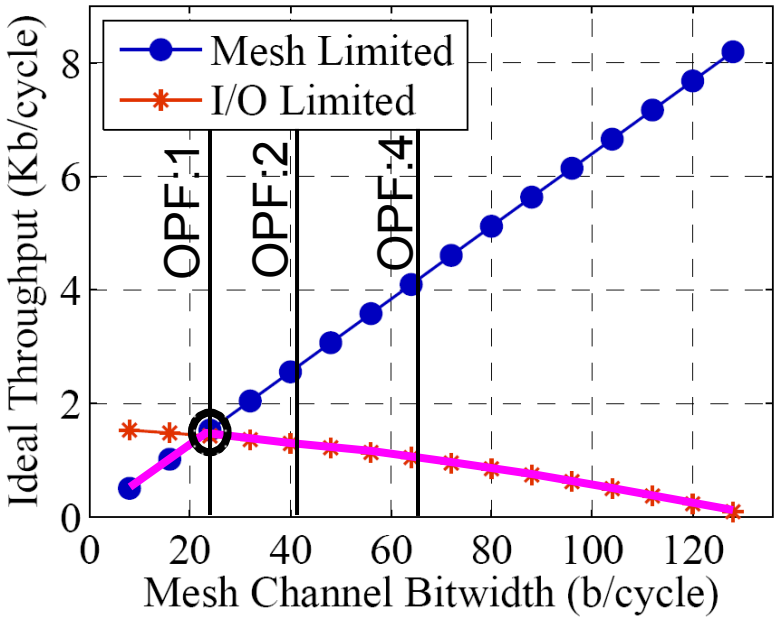


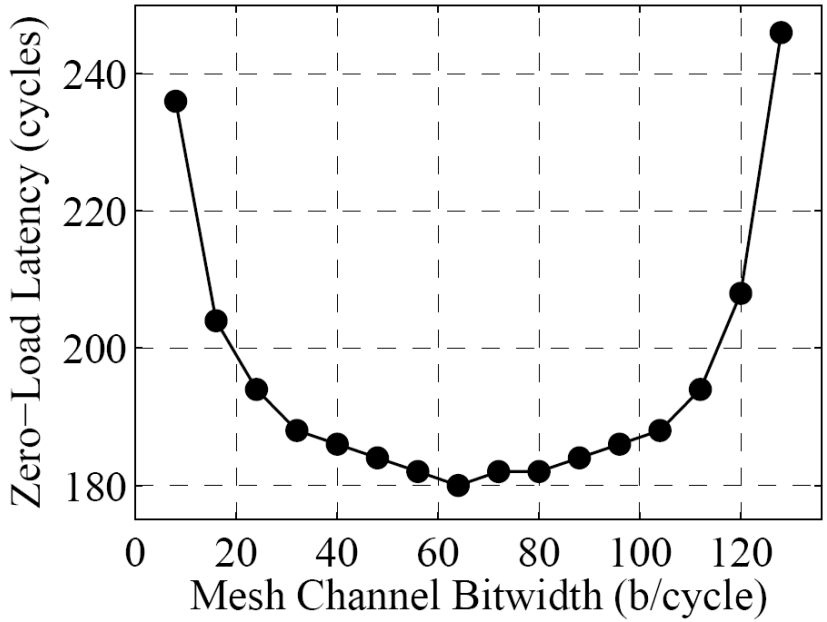


(22nm tech, 256 cores @ 2.5 GHz, 8 nJ/cyc energy budget)

- System throughput limited by on-chip mesh or I/O links
- On-chip mesh could be over-provisioned to overcome mesh bottleneck
- Zero load latency limited by data serialization

Network throughput and zero load latency

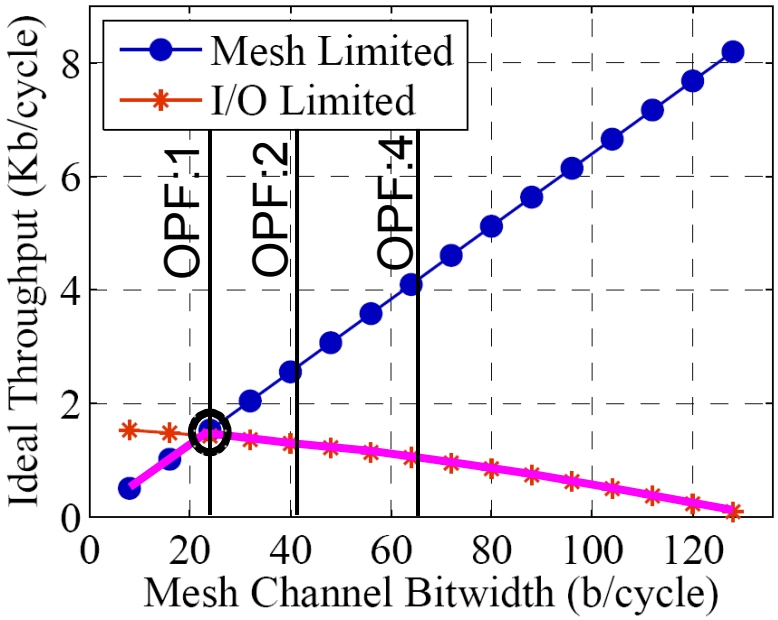


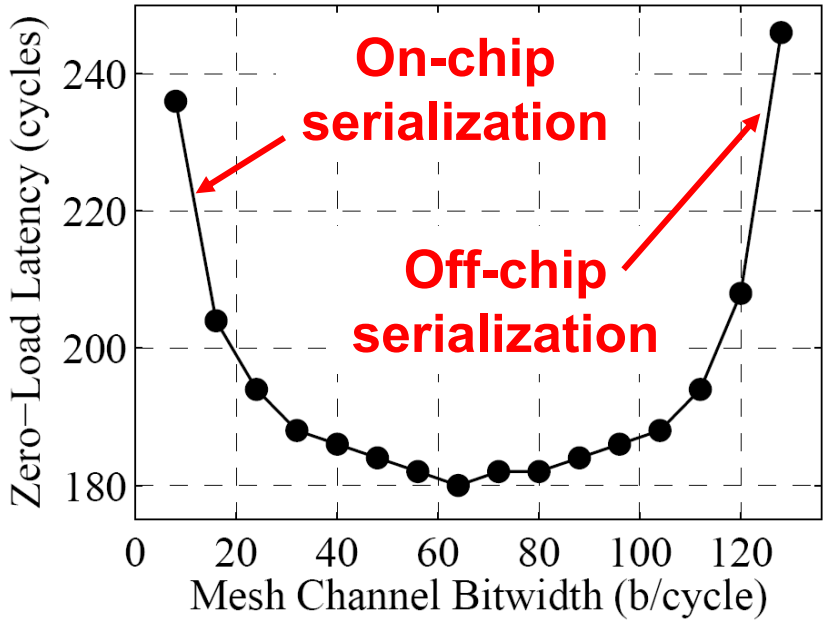


(22nm tech, 256 cores @ 2.5 GHz, 8 nJ/cyc energy budget)

- System throughput limited by on-chip mesh or I/O links
- On-chip mesh could be over-provisioned to overcome mesh bottleneck
- Zero load latency limited by data serialization

Network throughput and zero load latency





(22nm tech, 256 cores @ 2.5 GHz, 8 nJ/cyc energy budget)

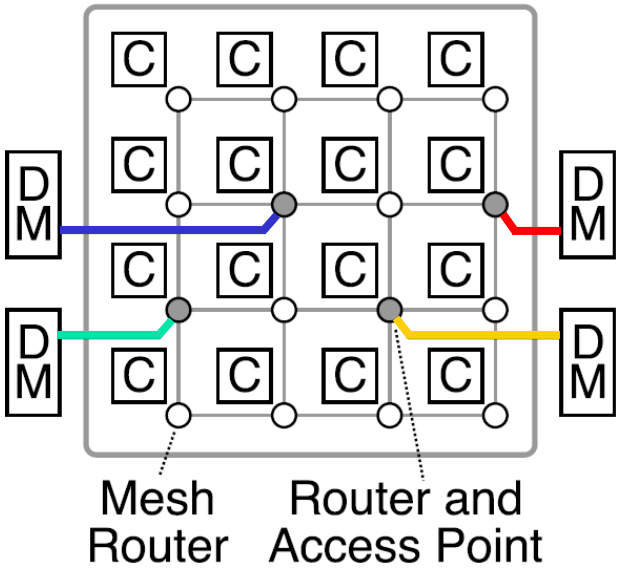
- System throughput limited by on-chip mesh or I/O links
- On-chip mesh could be over-provisioned to overcome mesh bottleneck
- Zero load latency limited by data serialization

Outline

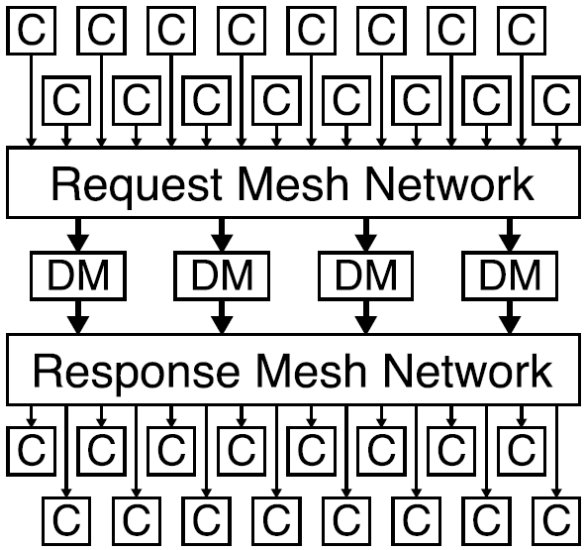
- Motivation
- Monolithic silicon photonic technology
- Processor-memory network architecture exploration
  - Baseline electrical mesh topology
  - Electrical mesh with optical global crossbar topology
- Manycore system using silicon photonics
- Conclusion

Optical system architecture

Mesh physical view



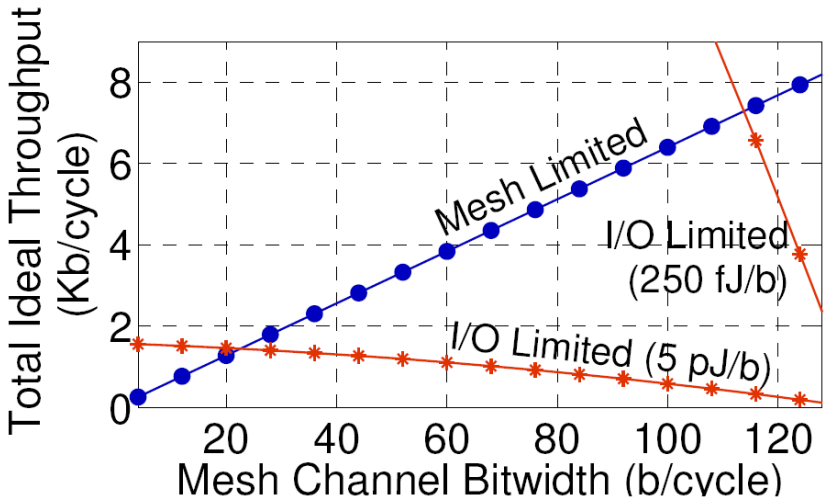
Mesh logical view

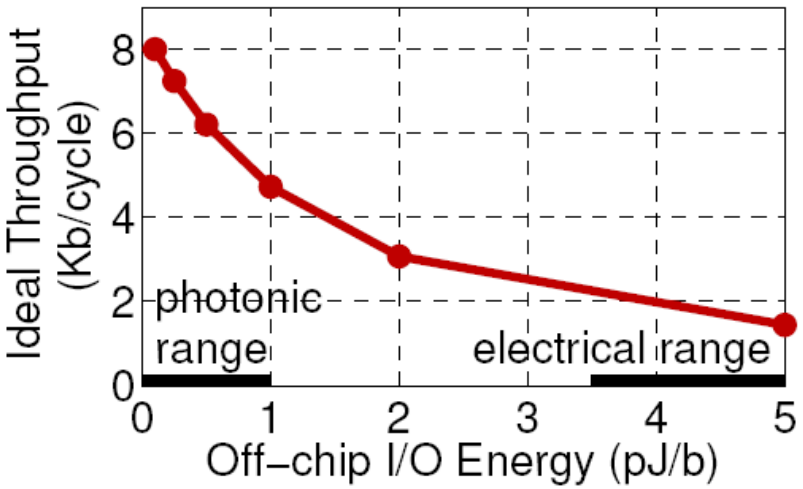


C = core, DM = DRAM module

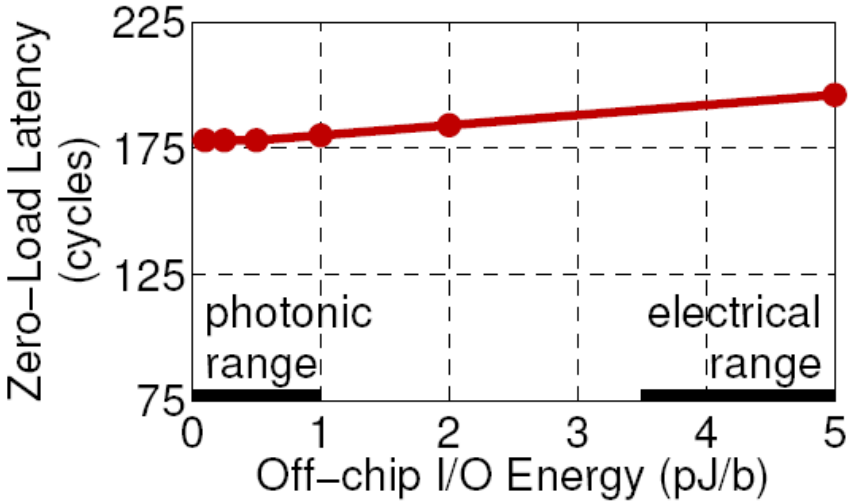
- Off-chip electrical links replaced with optical links
- Electrical to optical conversion at access point
- Wavelengths in each optical link distributed across various core-DRAM module pairs

Network throughput and zero load latency

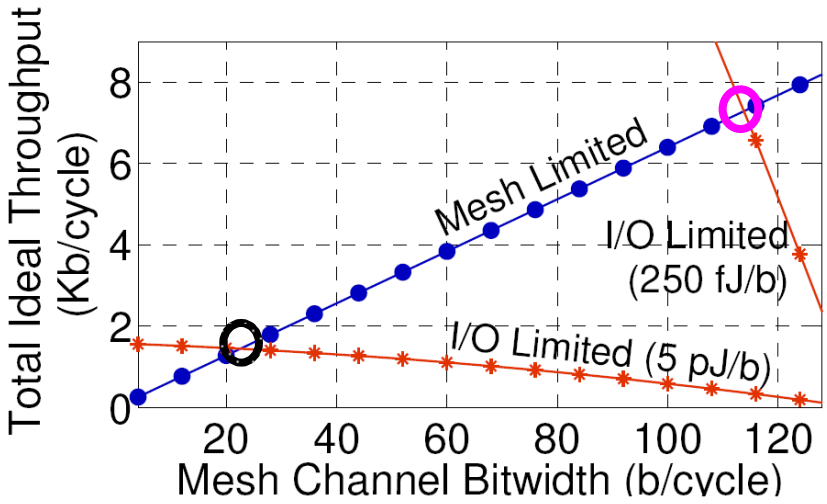


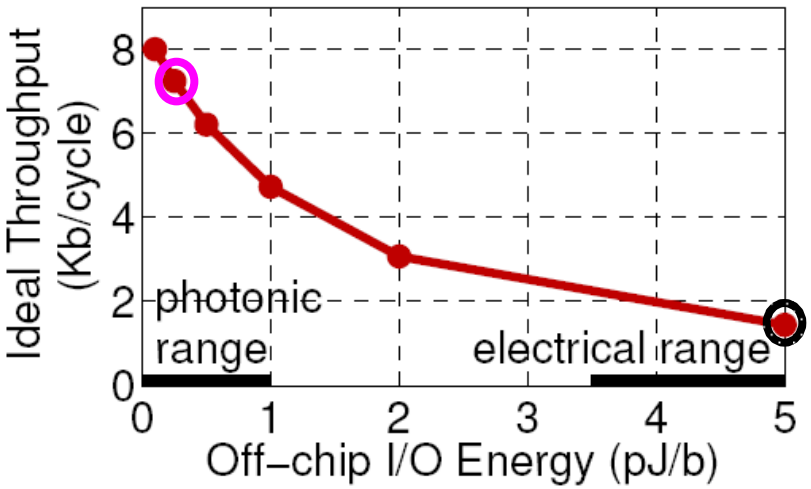


- Reduced I/O cost improves system bandwidth
- Reduction in latency due to lower serialization latency
- On-chip network is the new bottleneck

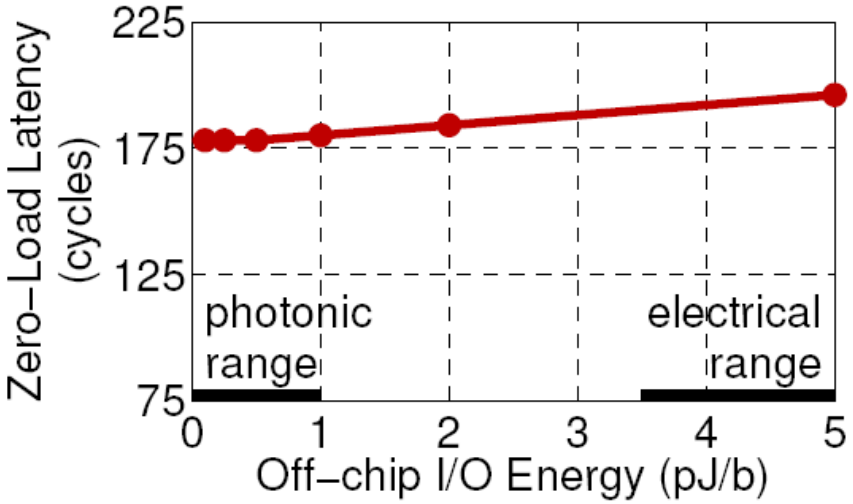


Network throughput and zero load latency

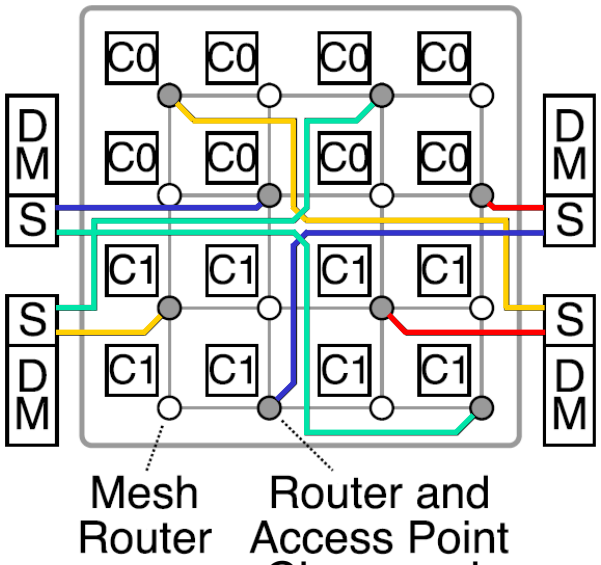


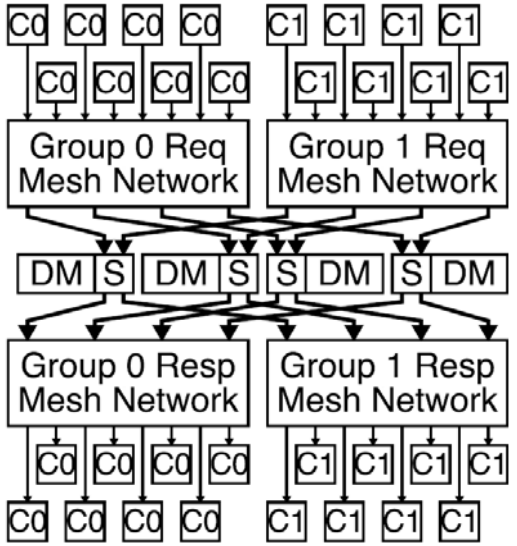


- Reduced I/O cost improves system bandwidth
- Reduction in latency due to lower serialization latency
- On-chip network is the new bottleneck



Optical multi-group system architecture

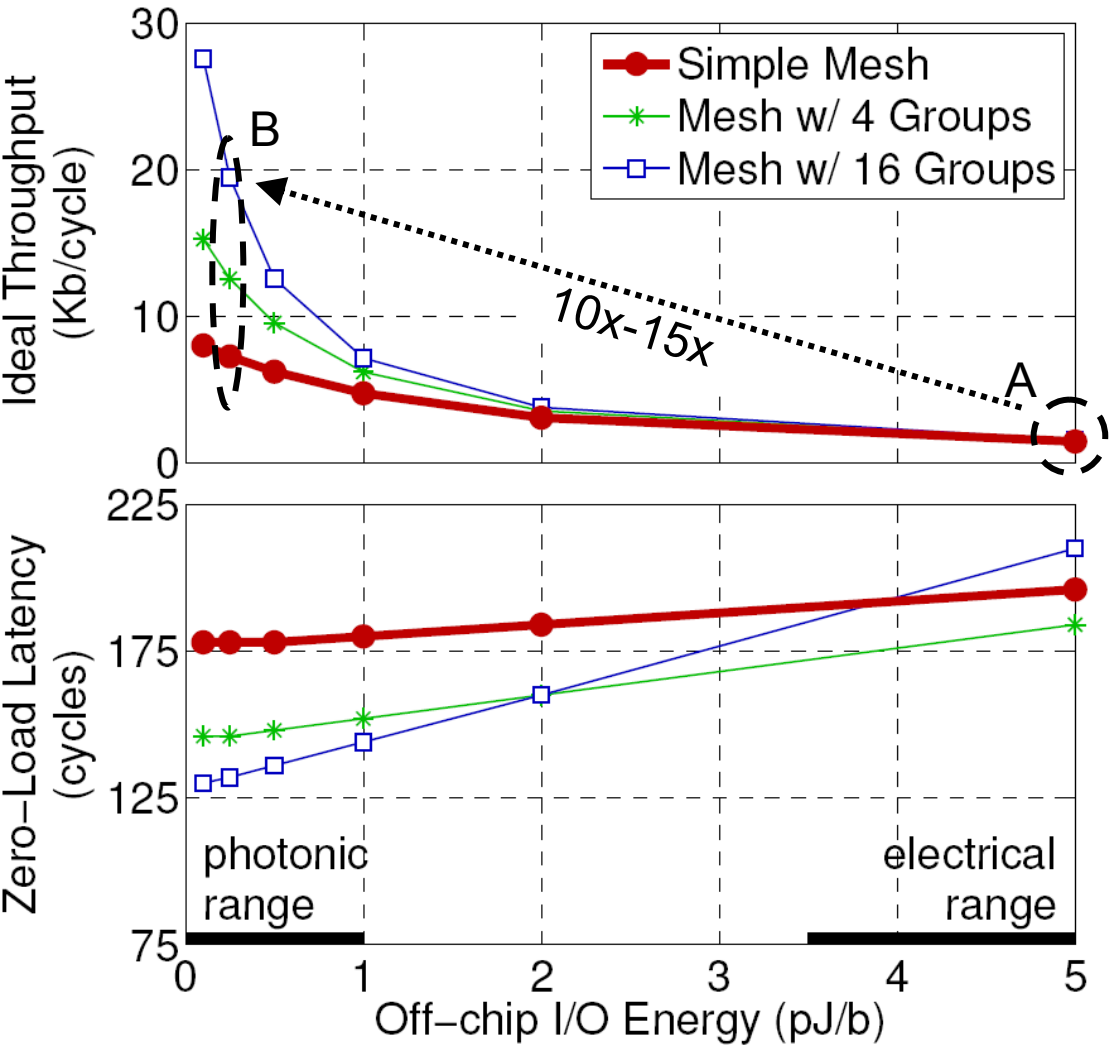




Ci = core in group *i*, DM = DRAM module, S = global crossbar switch

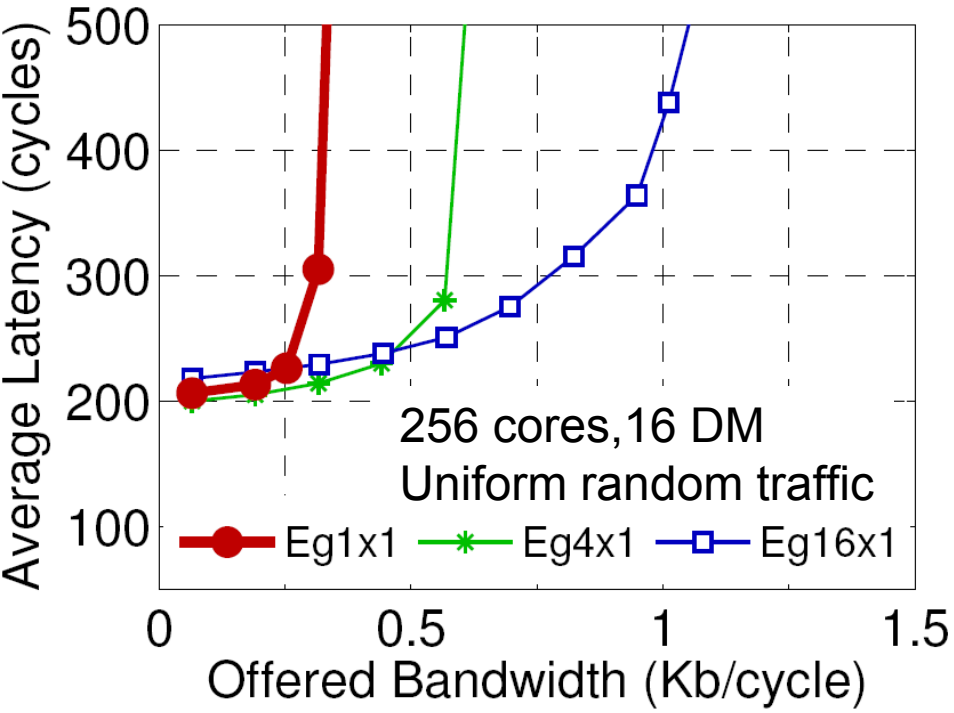
- Break the single on-chip electrical mesh into several groups
  - Each group has its own smaller mesh
  - Each group still has one AP for each DM
  - More APs → each AP is narrower (uses less λs)
- Use optical network as a very efficient global crossbar
- Need a crossbar switch at the memory for arbitration

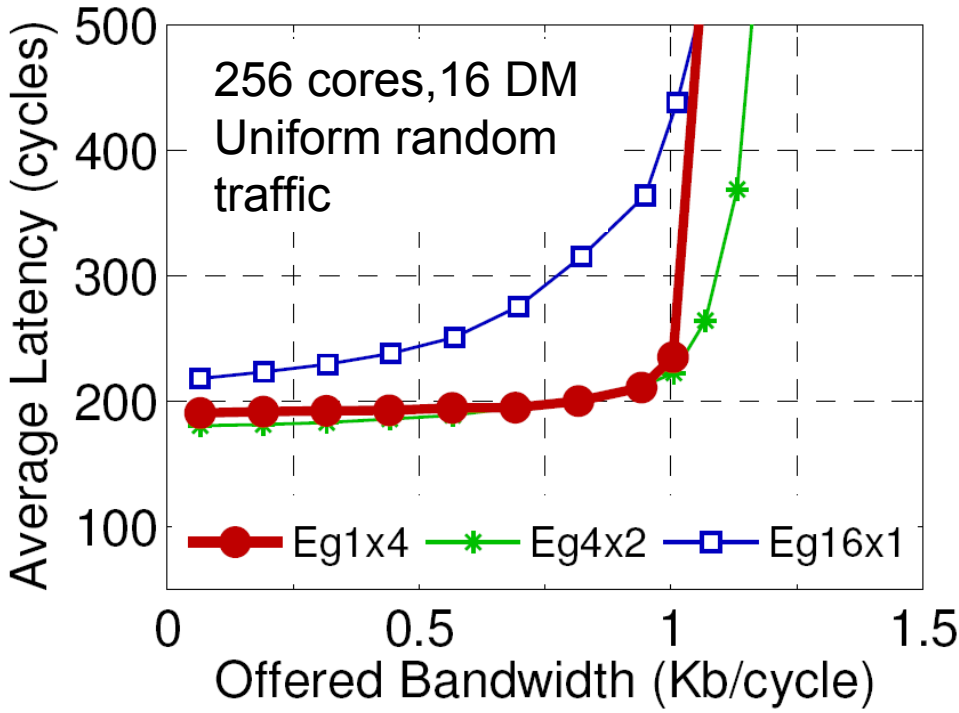
Network throughput vs zero load latency



- Grouping moves traffic from energy-inefficient mesh channels to energy-efficient photonic channels
- Grouping and silicon photonics provides 10x-15x throughput improvement
- Grouping reduces ZLL in photonic range, but increases ZLL in electrical range

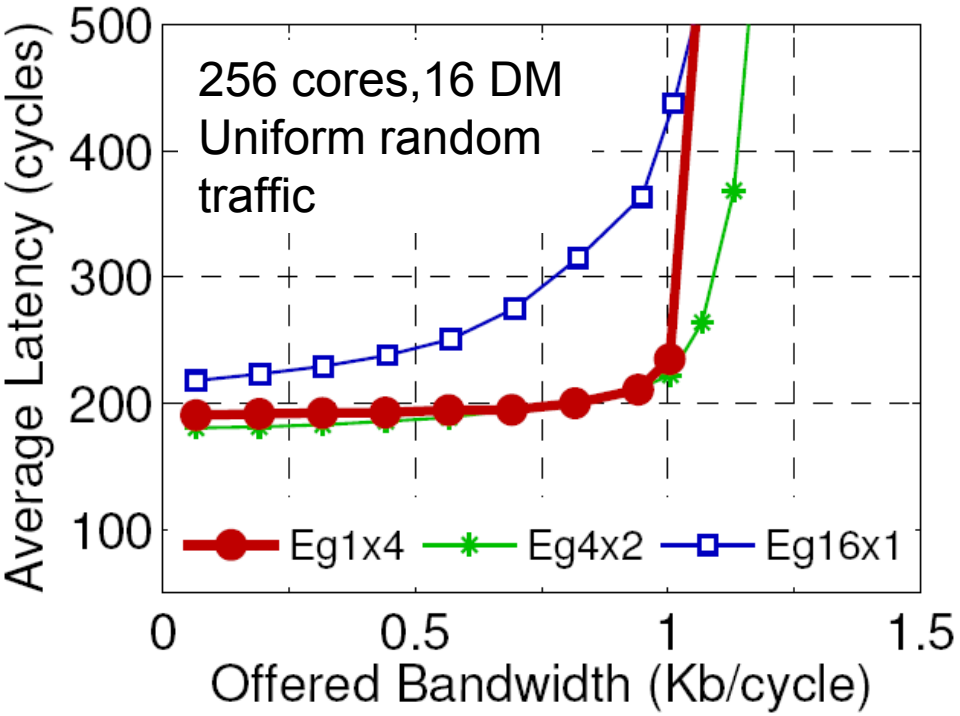
Simulation results

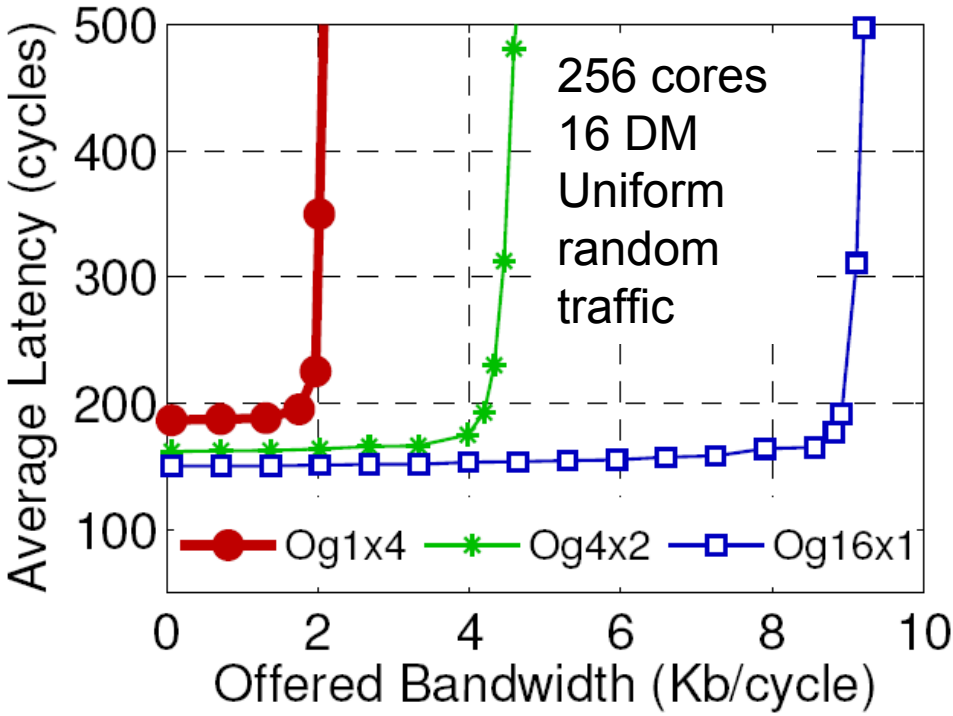




- Grouping
  - 2x improvement in bandwidth at comparable latency
- Overprovisioning
  - 2x-3x improvement in bandwidth for small group count at comparable latency
  - Minimal improvement for large group count

Simulation results



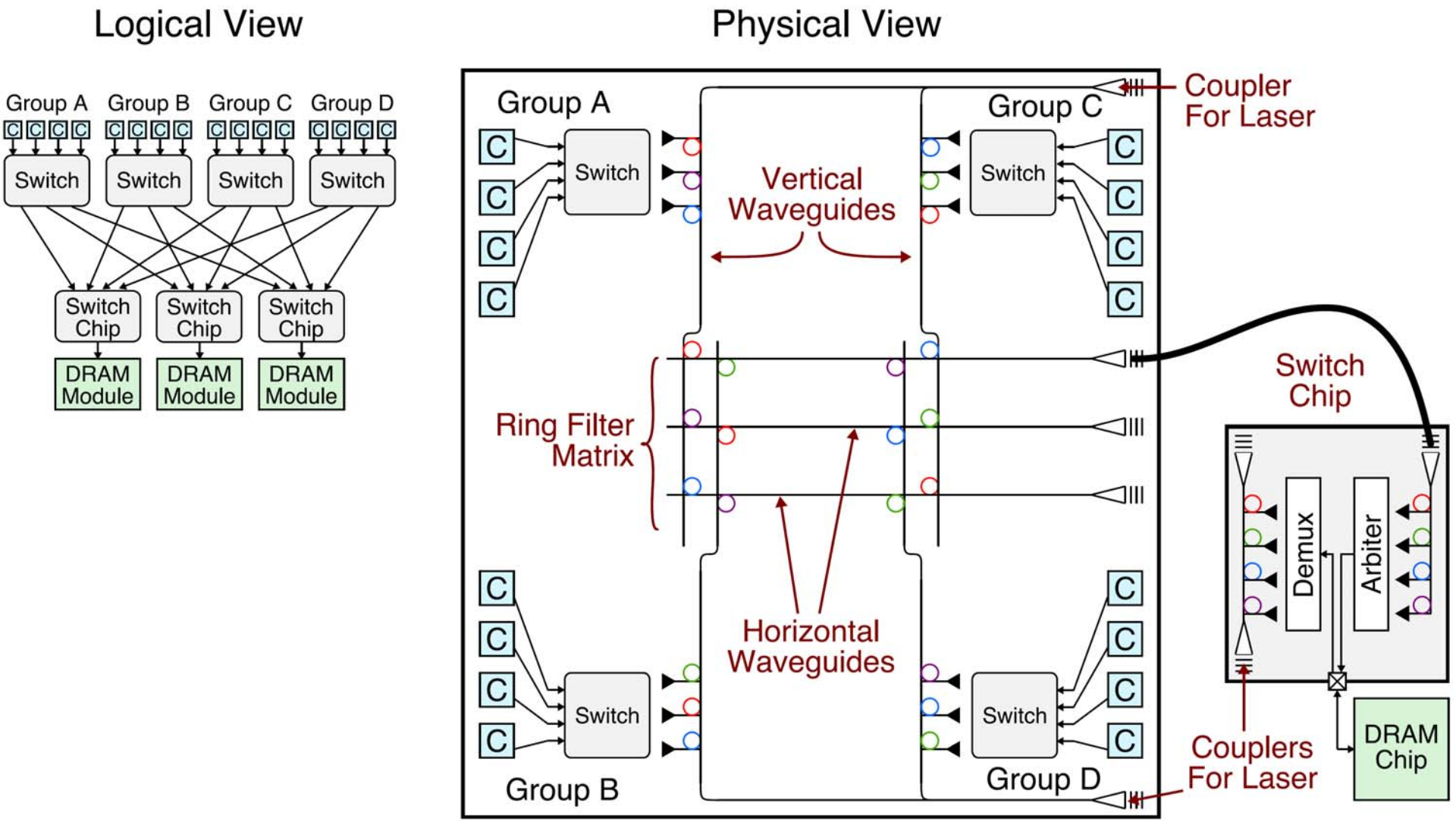


- Replacing off-chip electrical with photonics (Eg1x4 → Og1x4)
  - 2x improvement in bandwidth at comparable latency
- Using opto-electrical global crossbar (Eg4x2 → Og16x1)
  - 8x-10x improvement in bandwidth at comparable latency

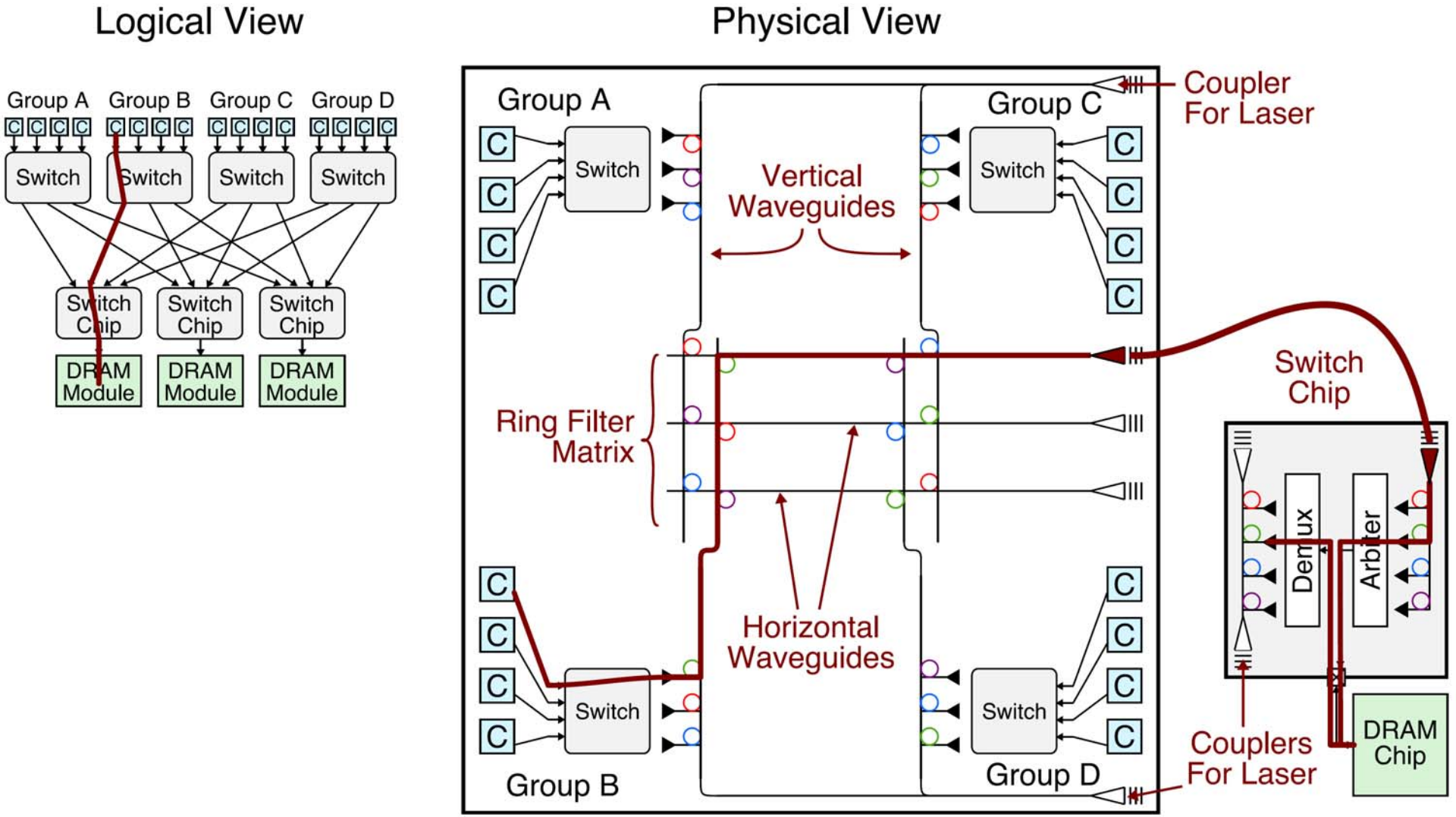
Outline

- Motivation
- Monolithic silicon photonic technology
- Processor-memory network architecture exploration
- Manycore system using silicon photonics
- Conclusion

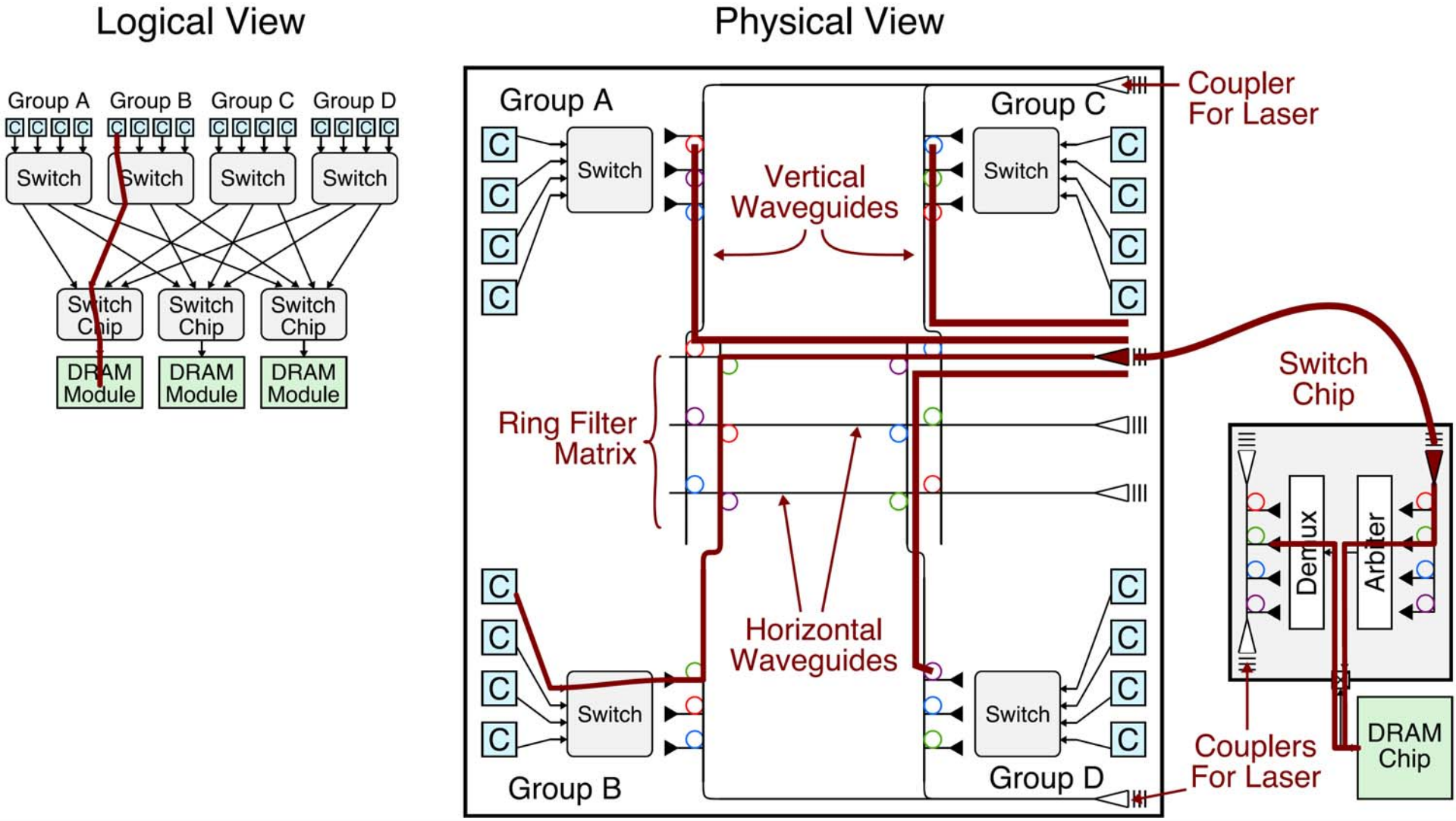
Simplified 16-core system design



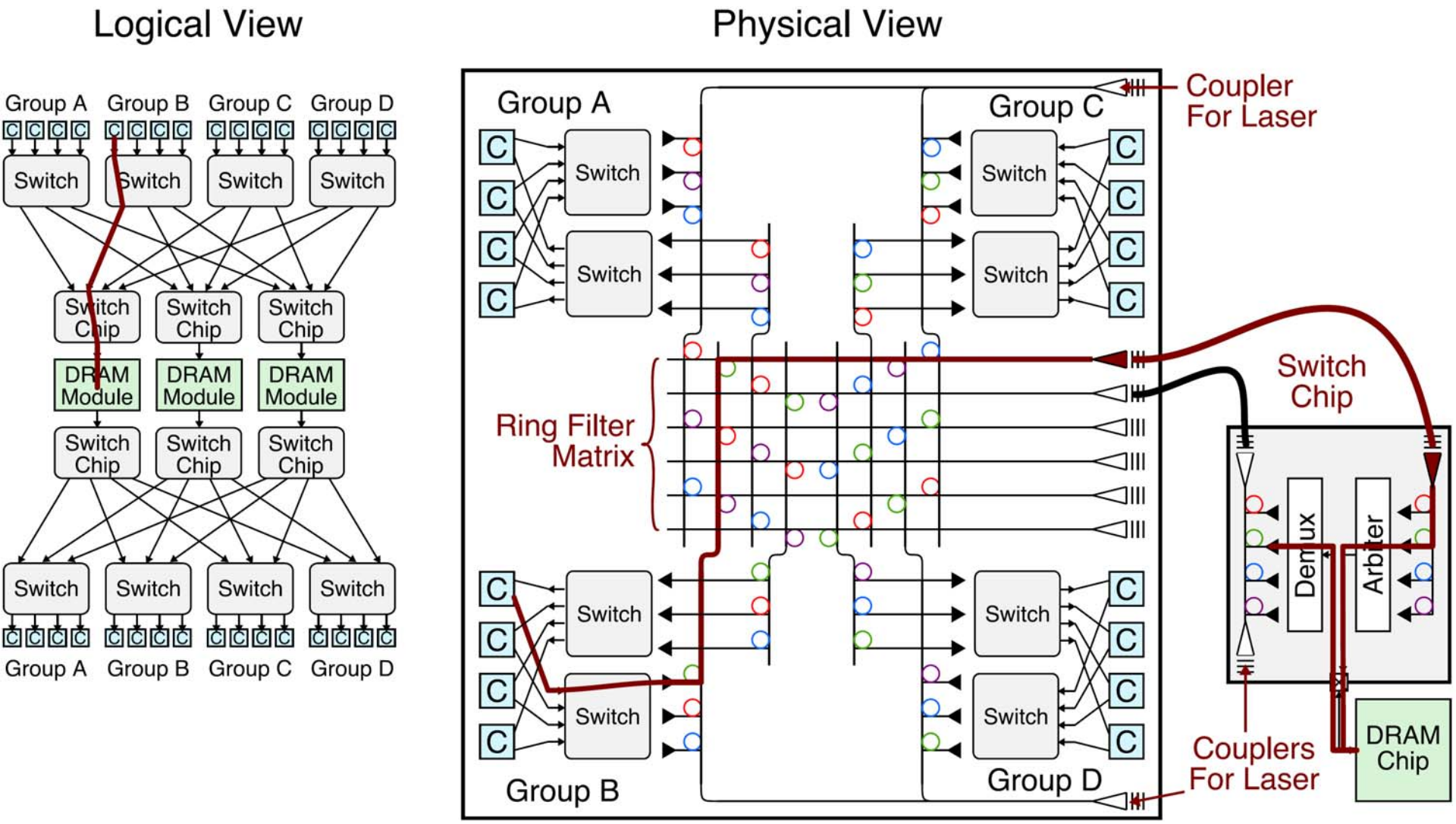
Simplified 16-core system design



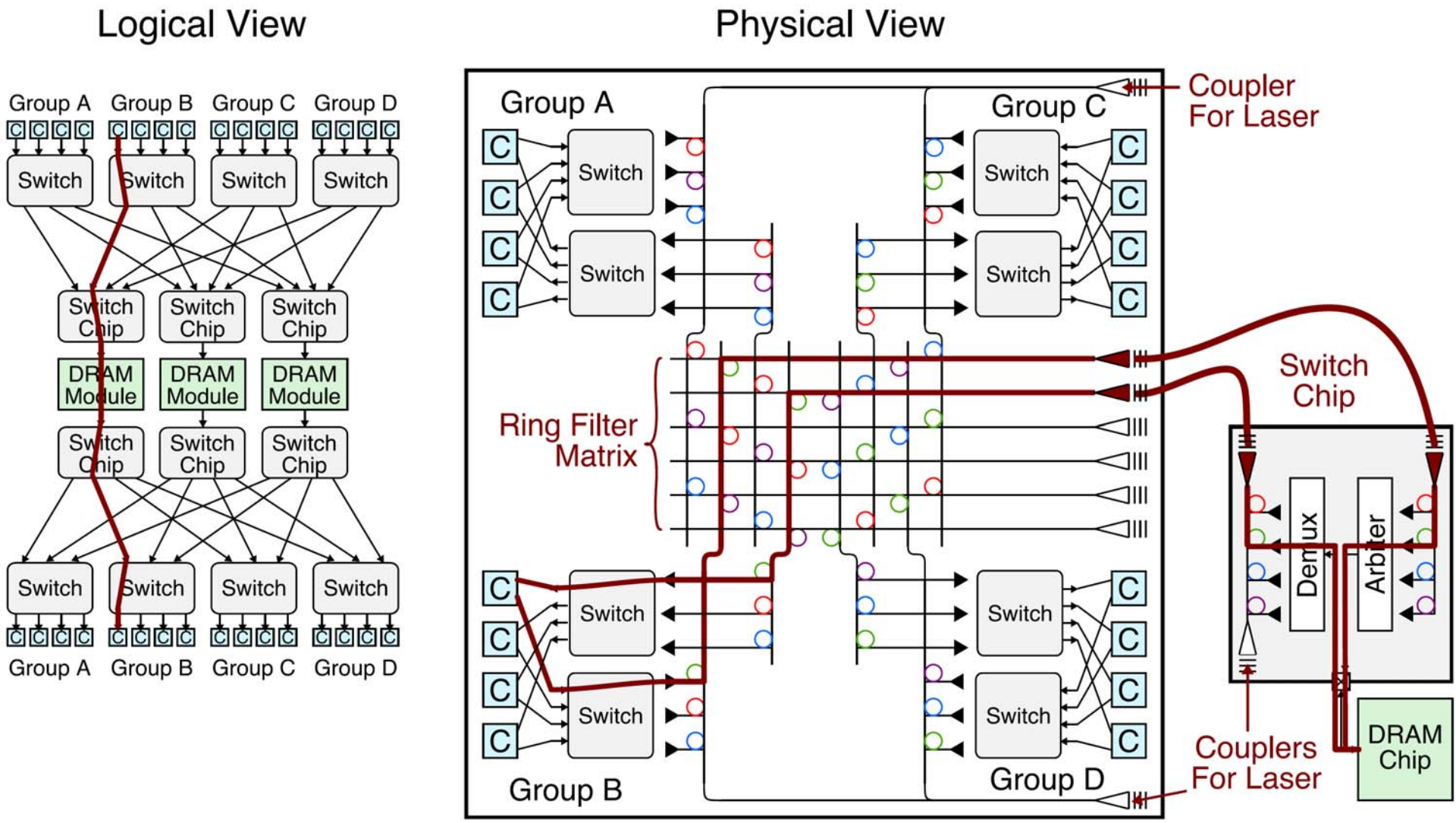
Simplified 16-core system design



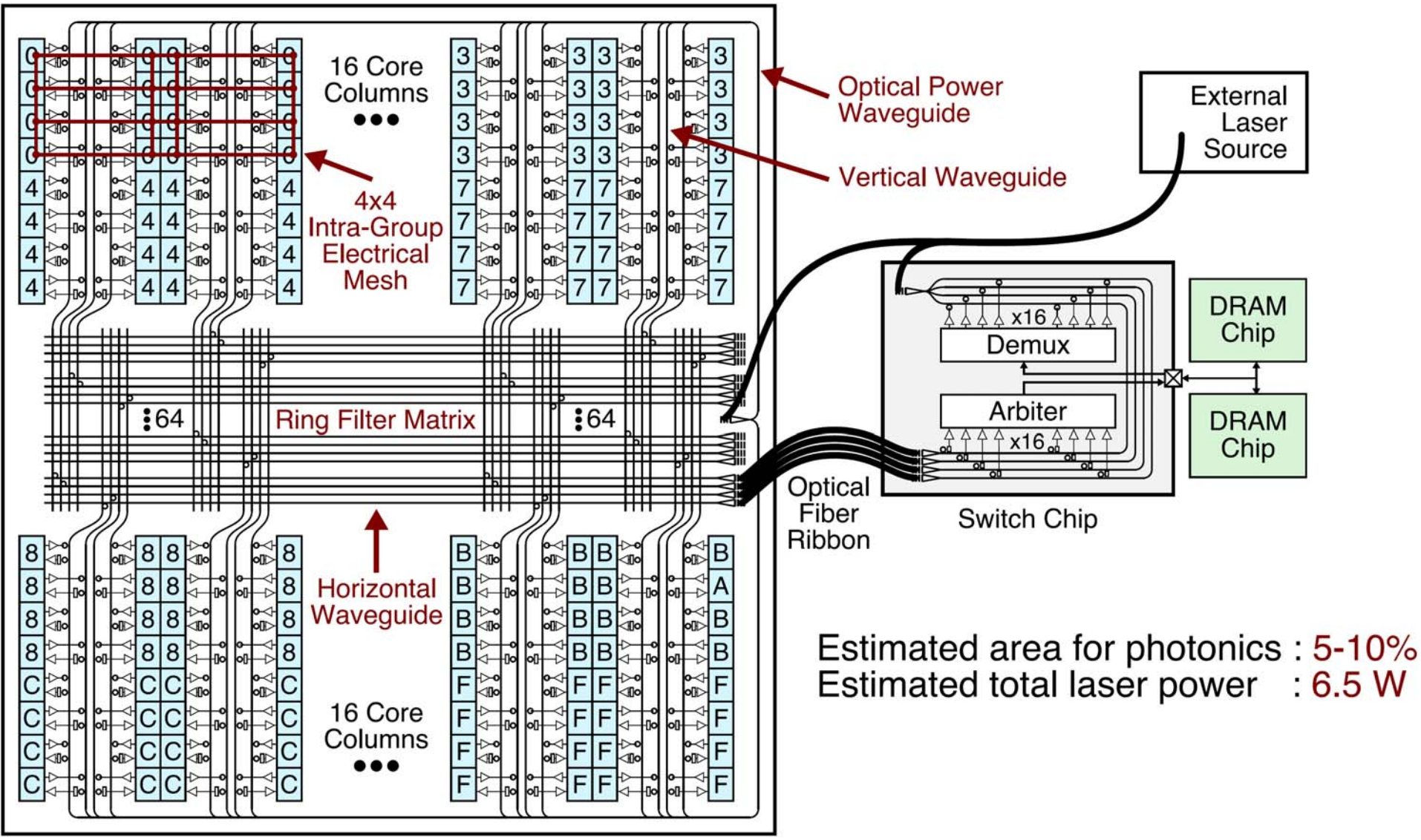
Simplified 16-core system design



Simplified 16-core system design



Full 256-core system design



Estimated area for photonics : 5-10%
Estimated total laser power : 6.5 W

Outline

- Motivation
- Monolithic silicon photonic technology
- Processor-memory network architecture exploration
- Manycore system using silicon photonics
- Conclusion

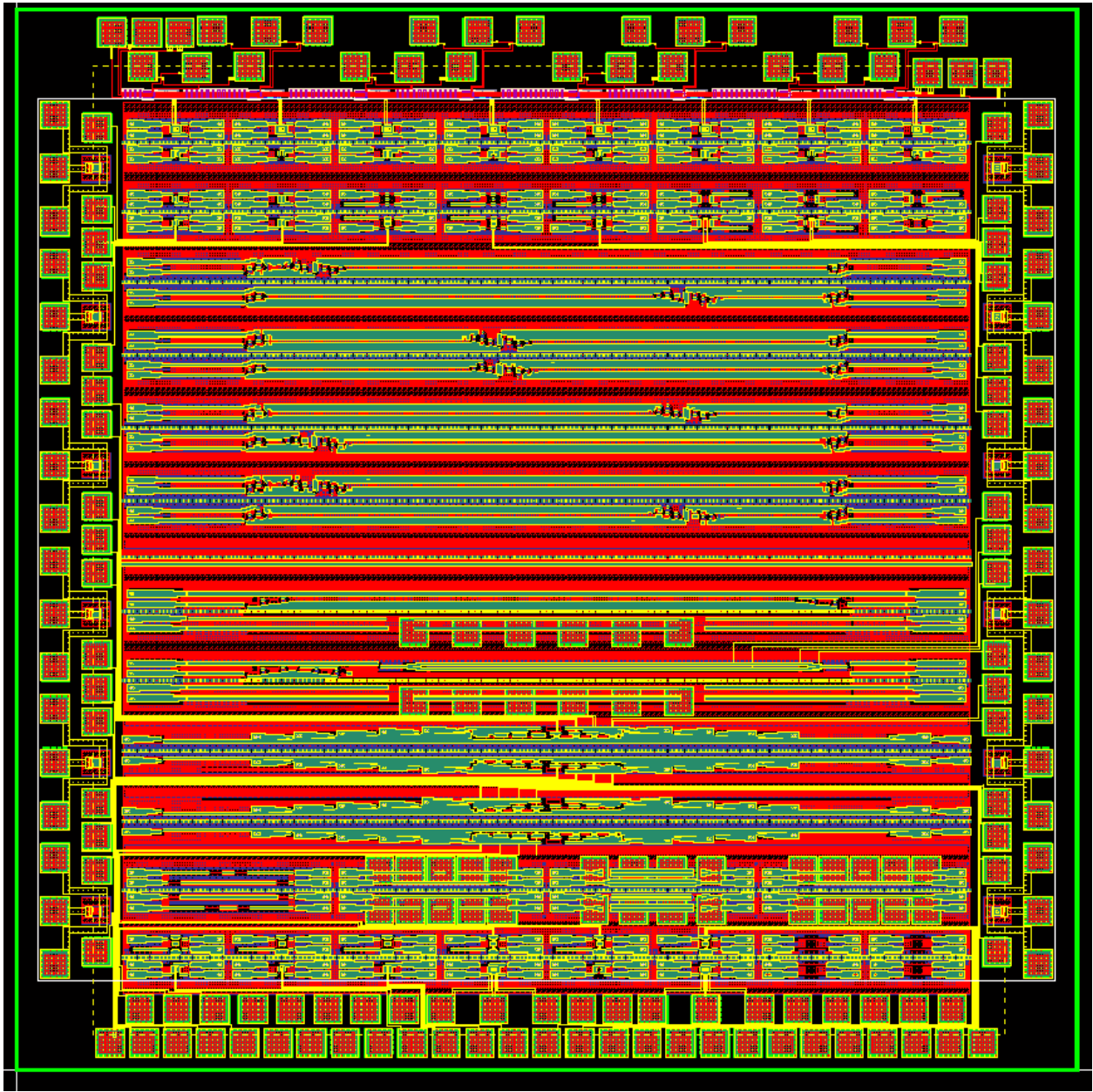
Conclusion

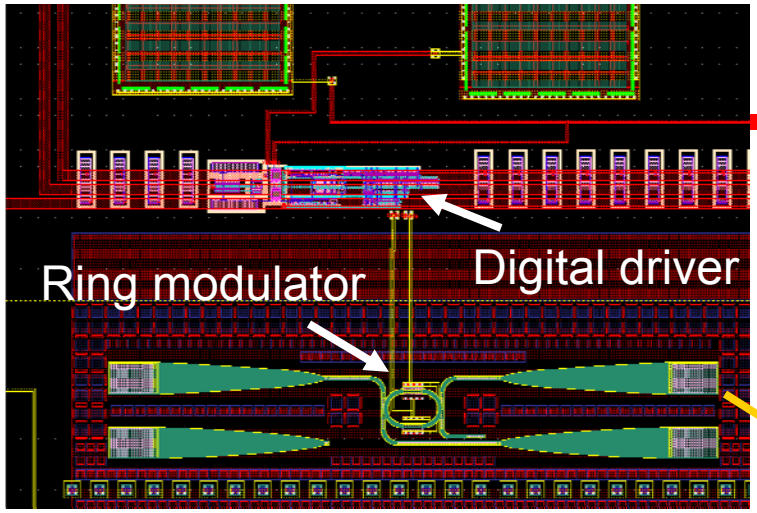
- On-chip network design and memory bandwidth will limit manycore system performance
- Unified on-chip/off-chip photonic link is proposed to solve this problem
- Grouping with optical global crossbar improves system throughput
- For an energy-constrained approach, photonics provide 8-10x improvement in throughput at comparable latency

Backup

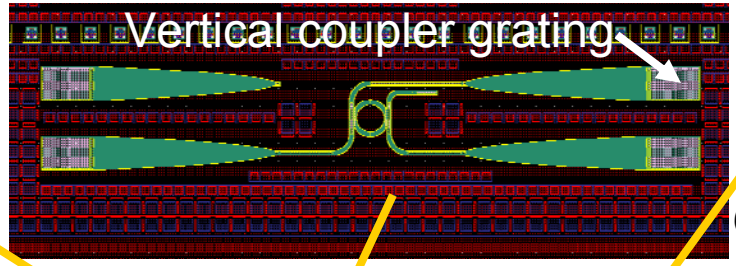
MIT Eos1 65 nm test chip

- Texas Instruments standard 65 nm bulk CMOS process
- First ever photonic chip in sub-100nm CMOS
- Automated photonic device layout
- Monolithic integration with electrical modulator drivers



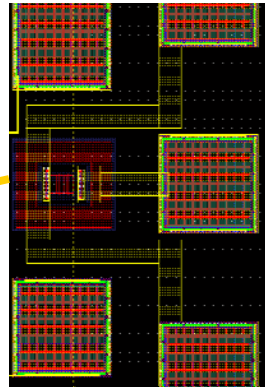


Two-ring filter

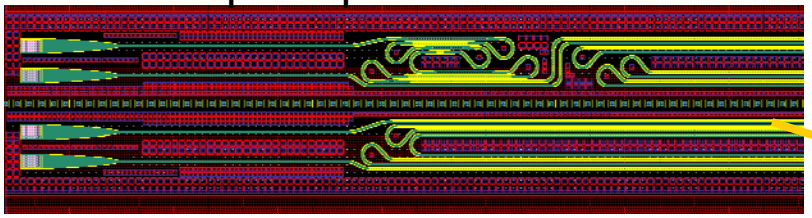


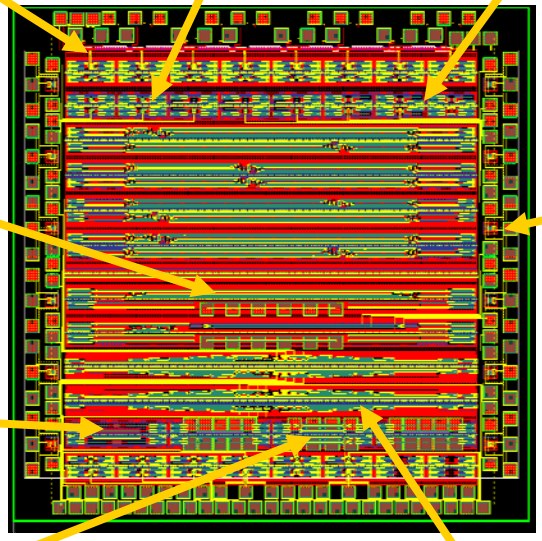
One-ring filter

Photo detector

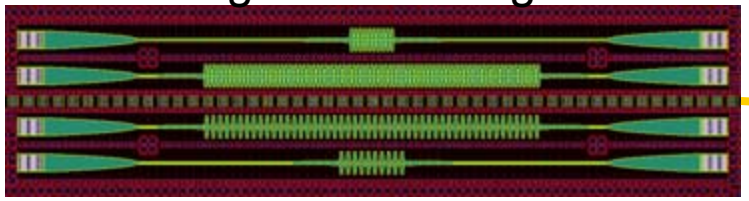


Paperclips

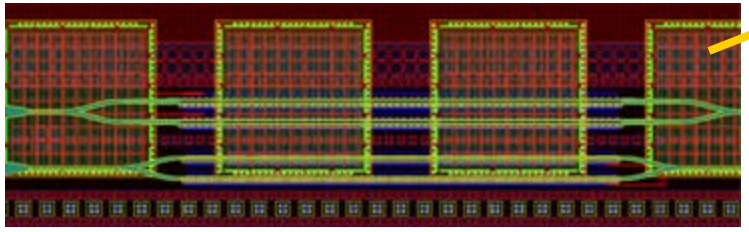




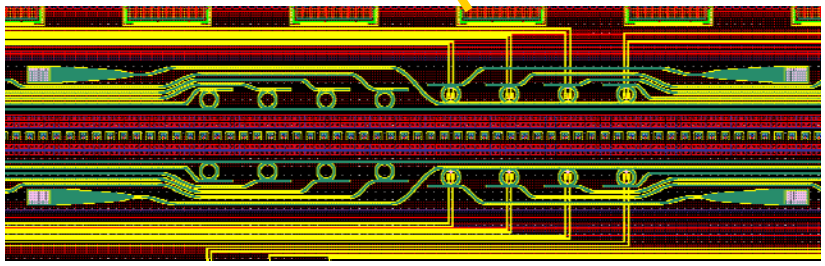
Waveguide crossings



M-Z test structures

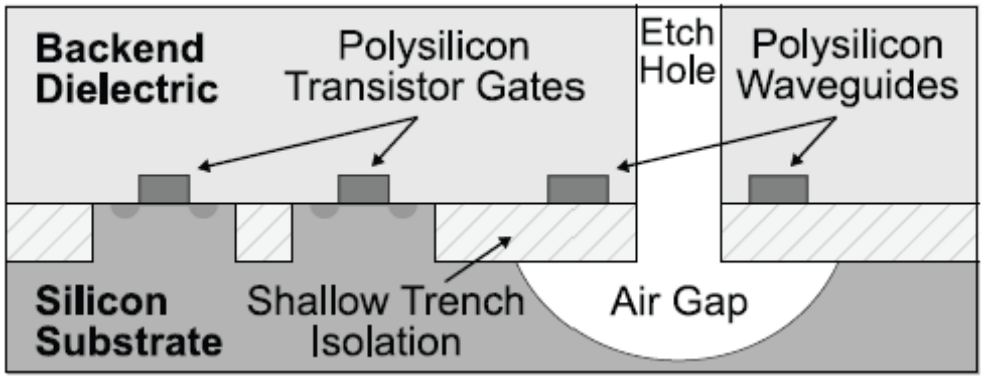


4 ring filter banks

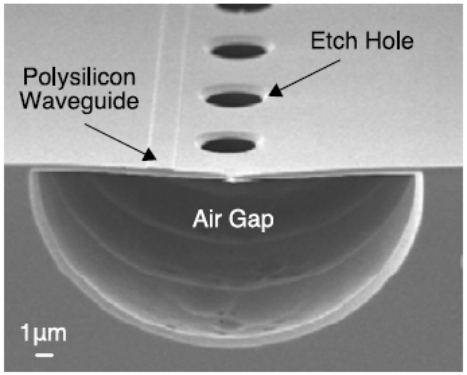


MIT/UCB

Optical waveguide



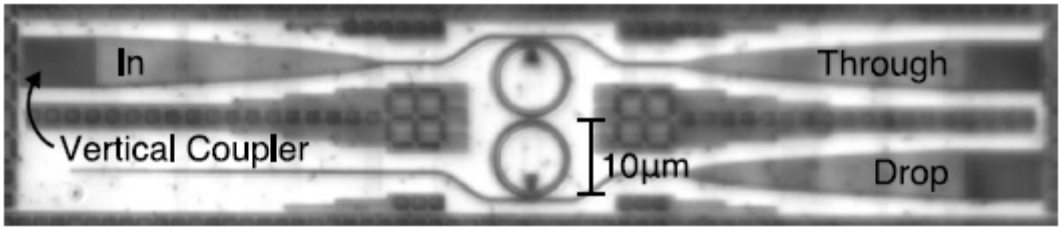
Cross-sectional view of a photonic chip



SEM image of a poly silicon waveguide

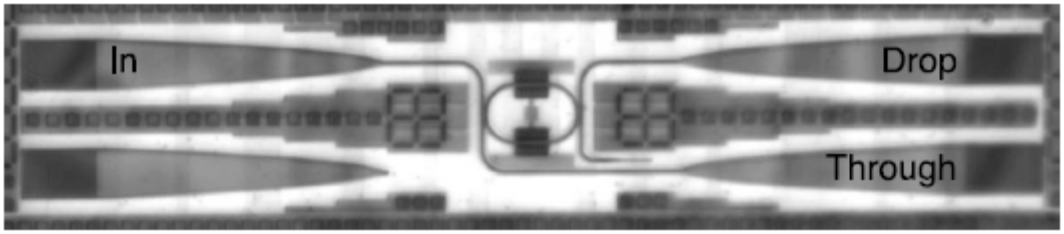
- Waveguide made of polysilicon
- Silicon substrate under waveguide etched away to provide optical cladding
- 64 wavelengths per waveguide in opposite directions

Modulators and filters



Double-ring resonant filter

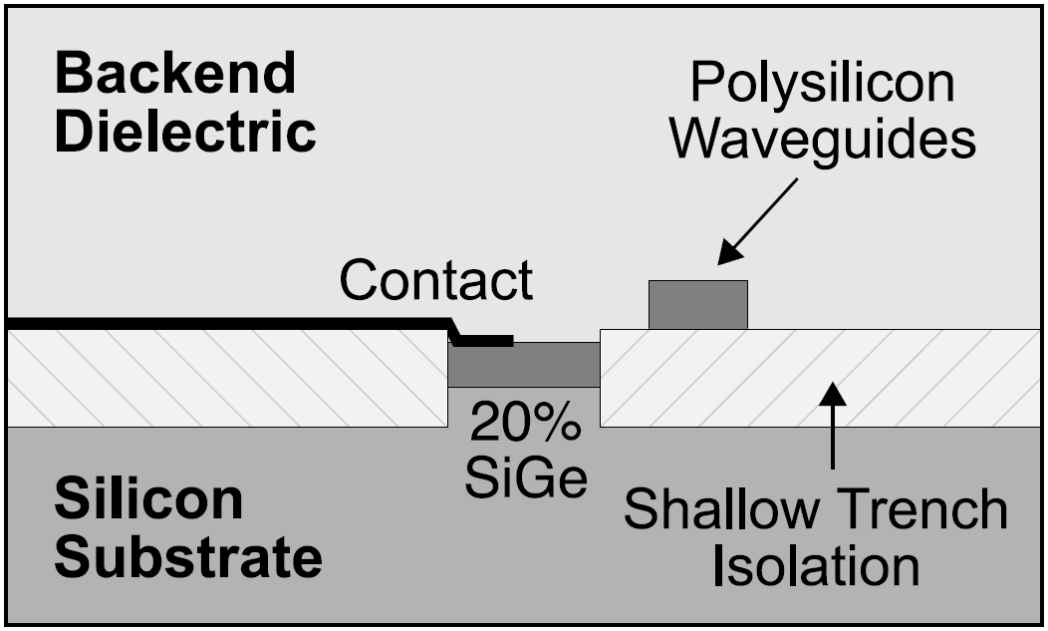
- 2nd order ring filters used
- Rings tuned using sizing and heating



Resonant racetrack modulator

- Modulator is tuned using charge injection
- Sub-100 fJ/bit energy cost for the modulator driver

Photodetectors



- Embedded SiGe used to create photodetectors
- Monolithic integration enable good optical coupling
- Sub-100 fJ/bit energy cost required for the receiver