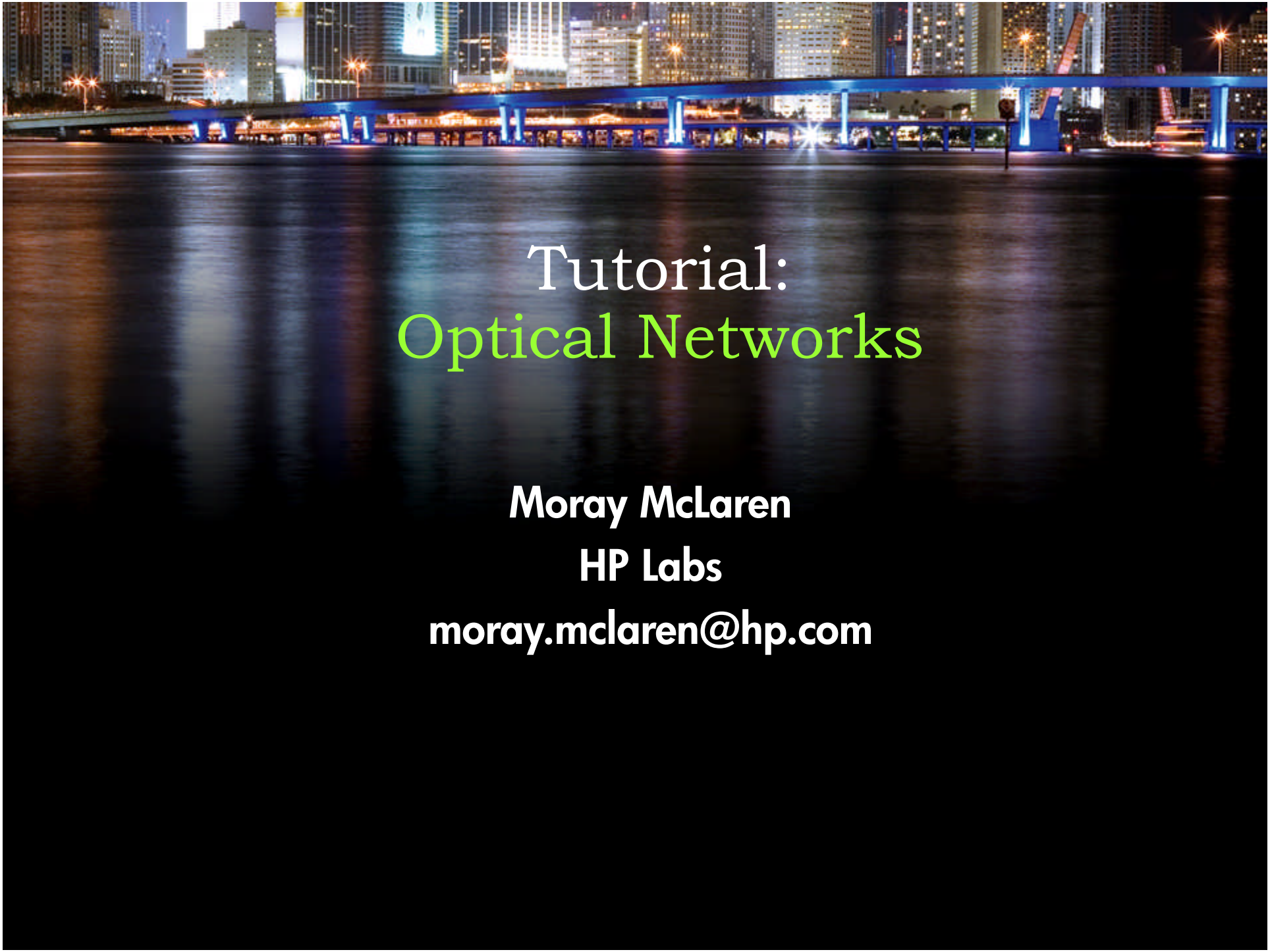
Optical Networks

Moray Mclaren



Tutorial:
Optical Networks

Moray McLaren
HP Labs
moray.mclaren@hp.com

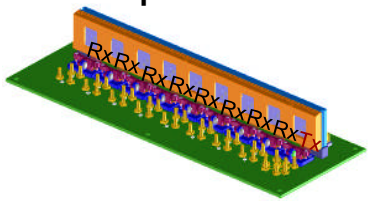
Overview

Topic
The Optical Roadmap
Optical busses and backplanes
Optic-based networking
Intrachip optics – multicore NOCs

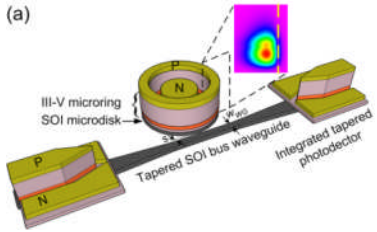
Tackling the bandwidth bottleneck with PHOTONICS

Active cable

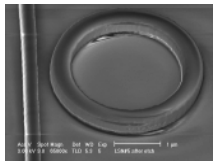
Optical Bus



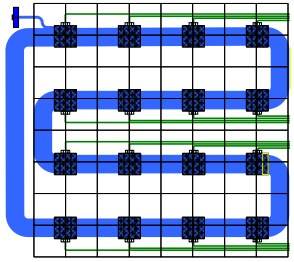
Hybrid laser cable



Silicon PIC



On-chip interconnect



Now | 1 Year | 3 Years | 5 Years | 7 Years | 10 Years

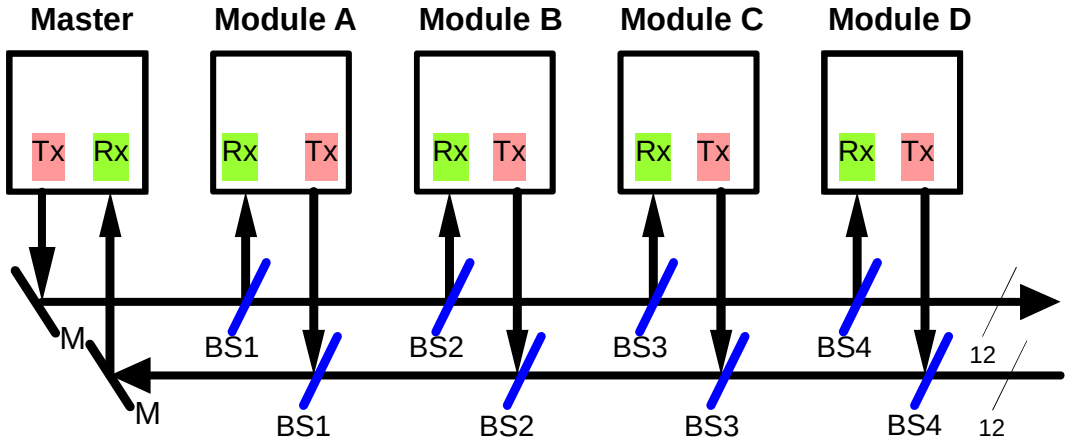
Single wavelength → CWDM → DWDM

100pJ/bit → pJ/bit → <.1

Optical buses and backplanes

Topic
The Optical Roadmap
Optical busses and backplanes
Optic-based networking
Intrachip optics – multicore NOCs

Optical Multidrop Bus

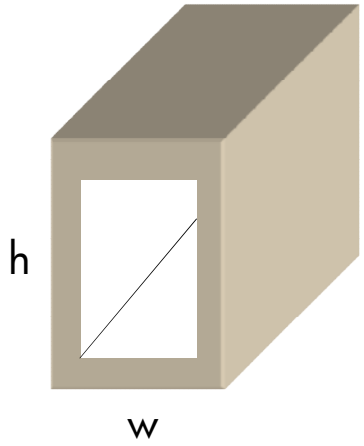


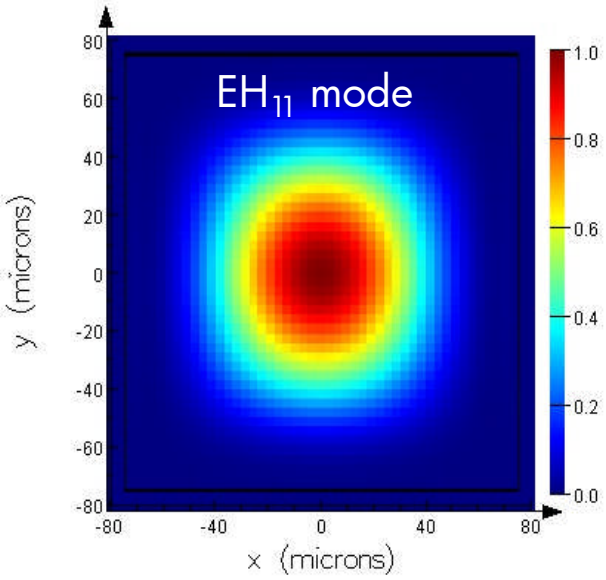
- Replace electrical transmission line with optical waveguides
- Replace electrical stubs with optical taps
- Two Unidirectional buses: 12 bit wide @ 10Gb/s = 30GB/s
  - Master broadcasts to each module on the bus;
  - Distribute optical power equally among modules
  - Each module sends data back to master at full bus bandwidth
- Lower latency with reduced power

Optical Waveguide

- Hollow Metal Waveguides[1] (HMWG)
  - Low propagation loss – light rays travel at near grazing angle to metal walls
  - Low numerical aperture
  - Prop delay 33psec/cm

Air core
Ag clad (n, k) = (0.15+i 5.68)
w =150μm, h =150μm
α = 0.0015 dB/cm
n_{eff} ~ 1
NA ~ 0.01

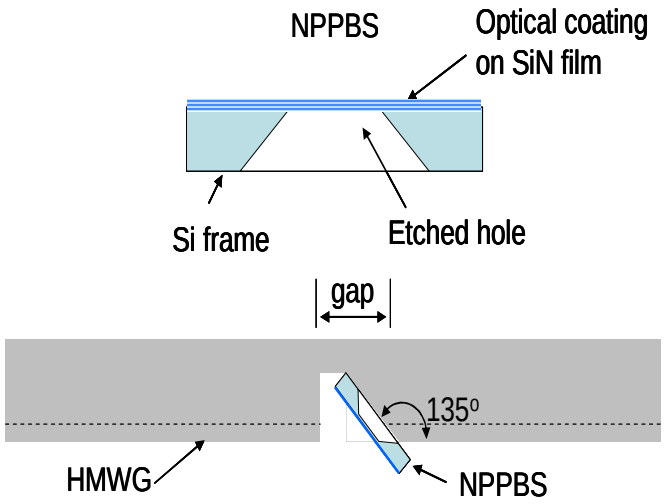




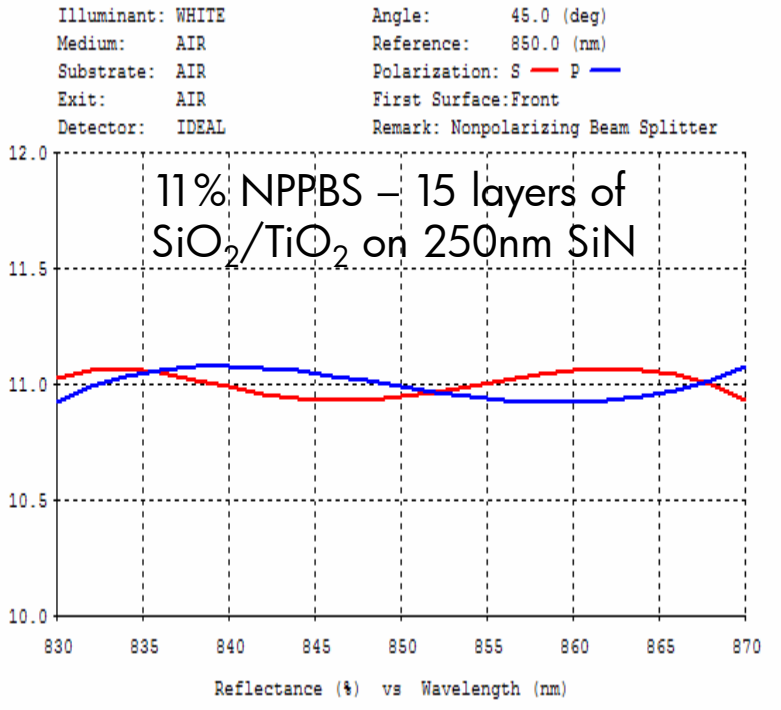
(1) E. Marcatili *et al., Bell Syst. Tech. J.* 43, 1783 (1964).

Optical Taps

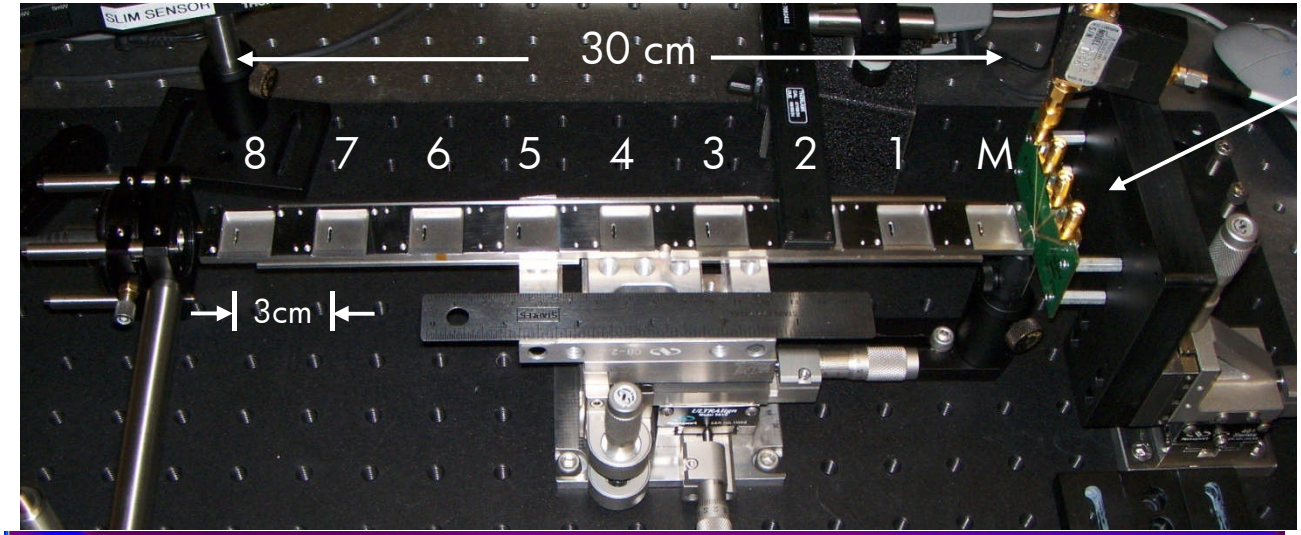
- Non-Polarizing Pellicle Beam Splitters
  - Low cost VCSELs randomly polarized
  - Negligible beam-walk off



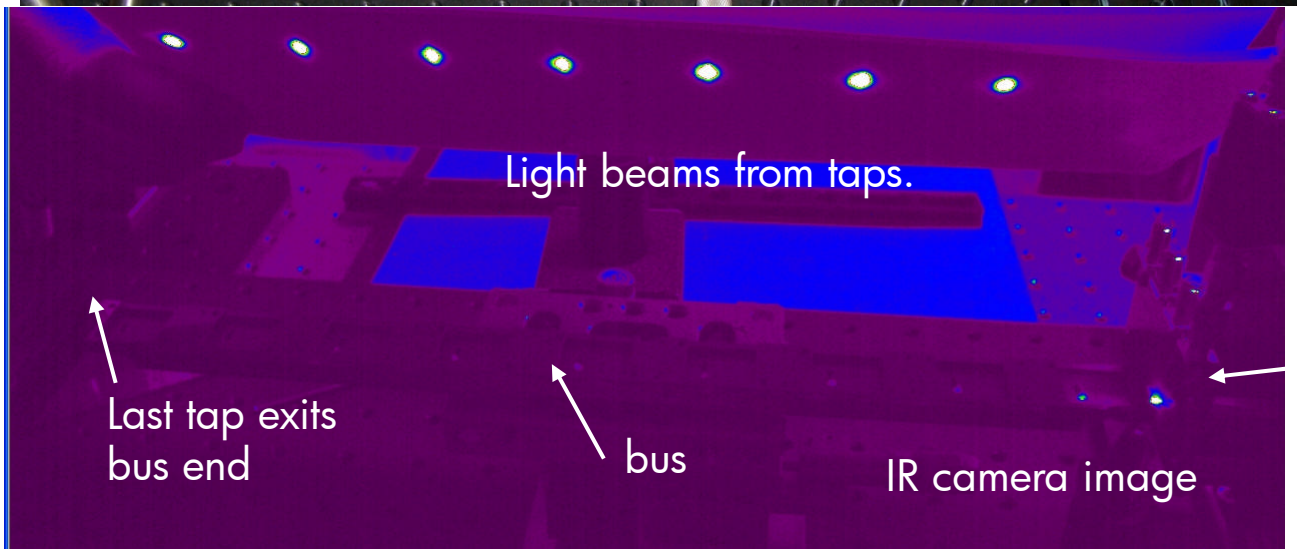
Insertion into HWMG @ 45°



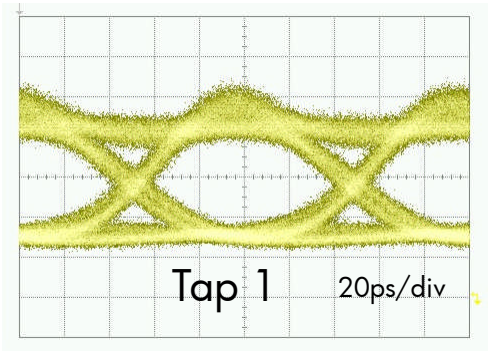
1x8 Fanout

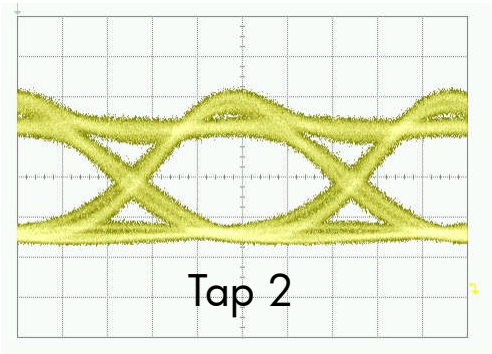


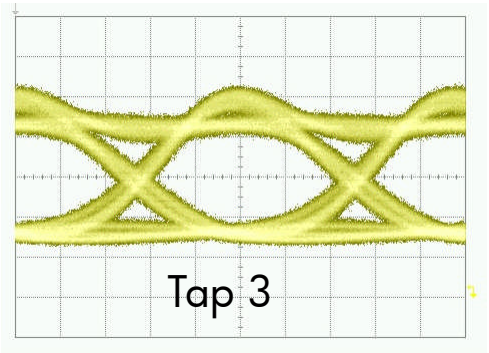
VCSEL driven from BERT thru bias-tee

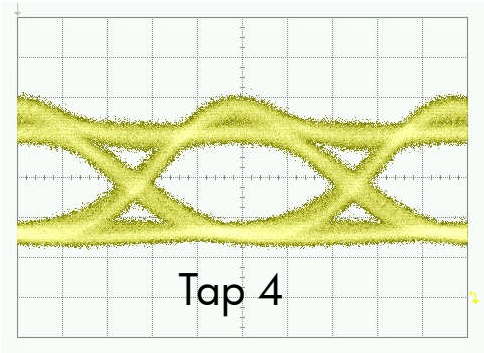


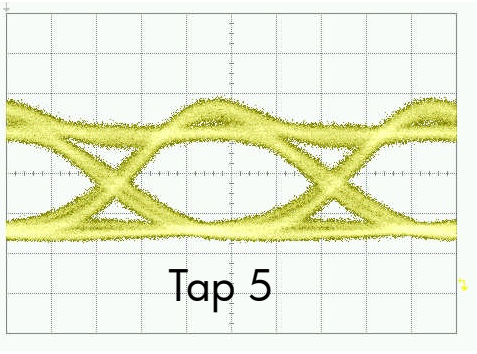
1x8 Fanout @ 10.3125Gbps

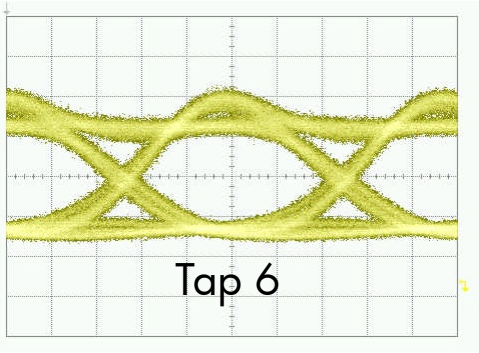


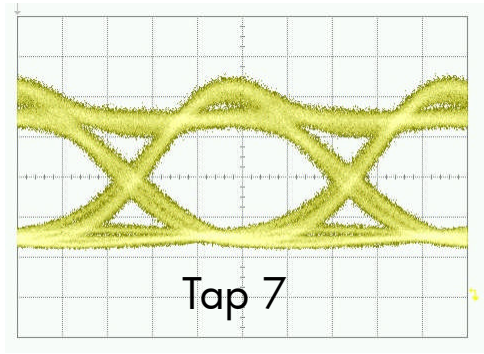


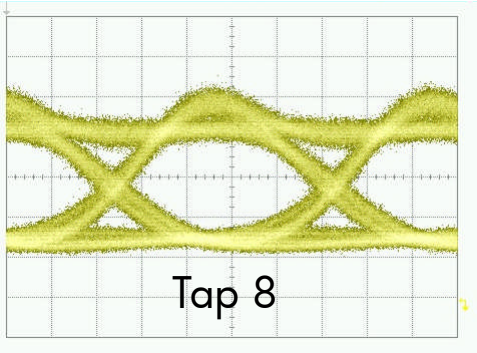










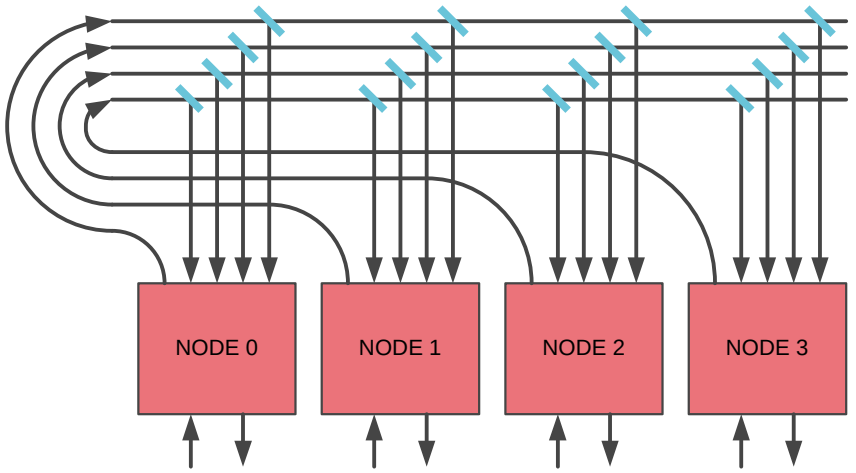


Application to core routers

- Bandwidth scaling
  - Core switch requirement doubling every 18 months
  - Electronic technologies can no longer keep up
  - EMI abatement a big problem
- Power
  - High % of switch power in interconnect and fabric ASIC
- Cost
  - Can we achieve equivalent cost for optical solutions?

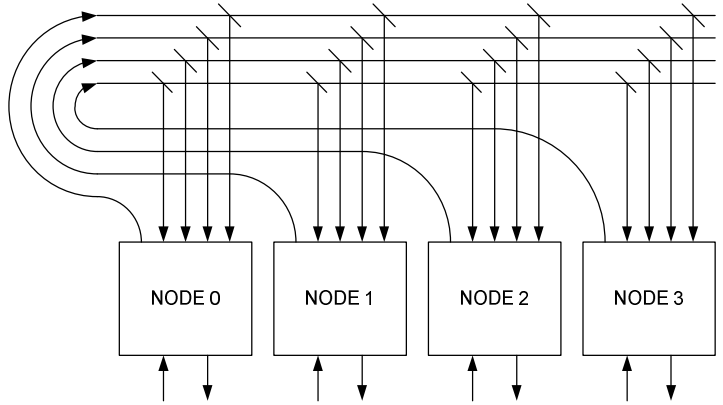
Optically Connected Switch

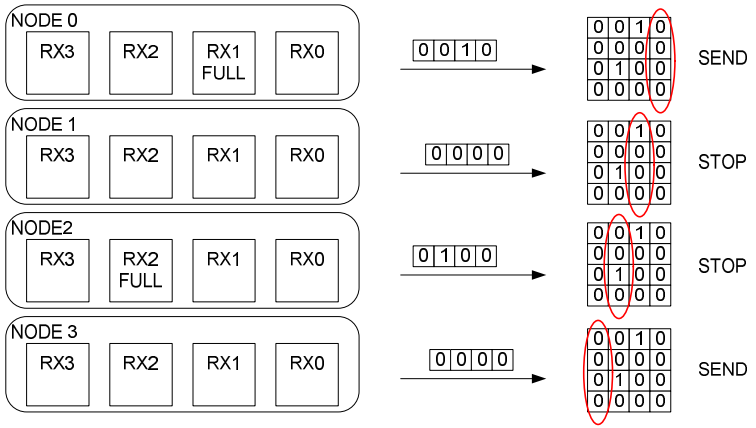
- Multibus fabrics
  - Remove requirement for fabric switches
  - Passive optical backplane
  - Linear cost scaling
- Key technologies
  - Alignment tolerant connectors
  - Low cost optical engines
  - Plastic molded waveguides



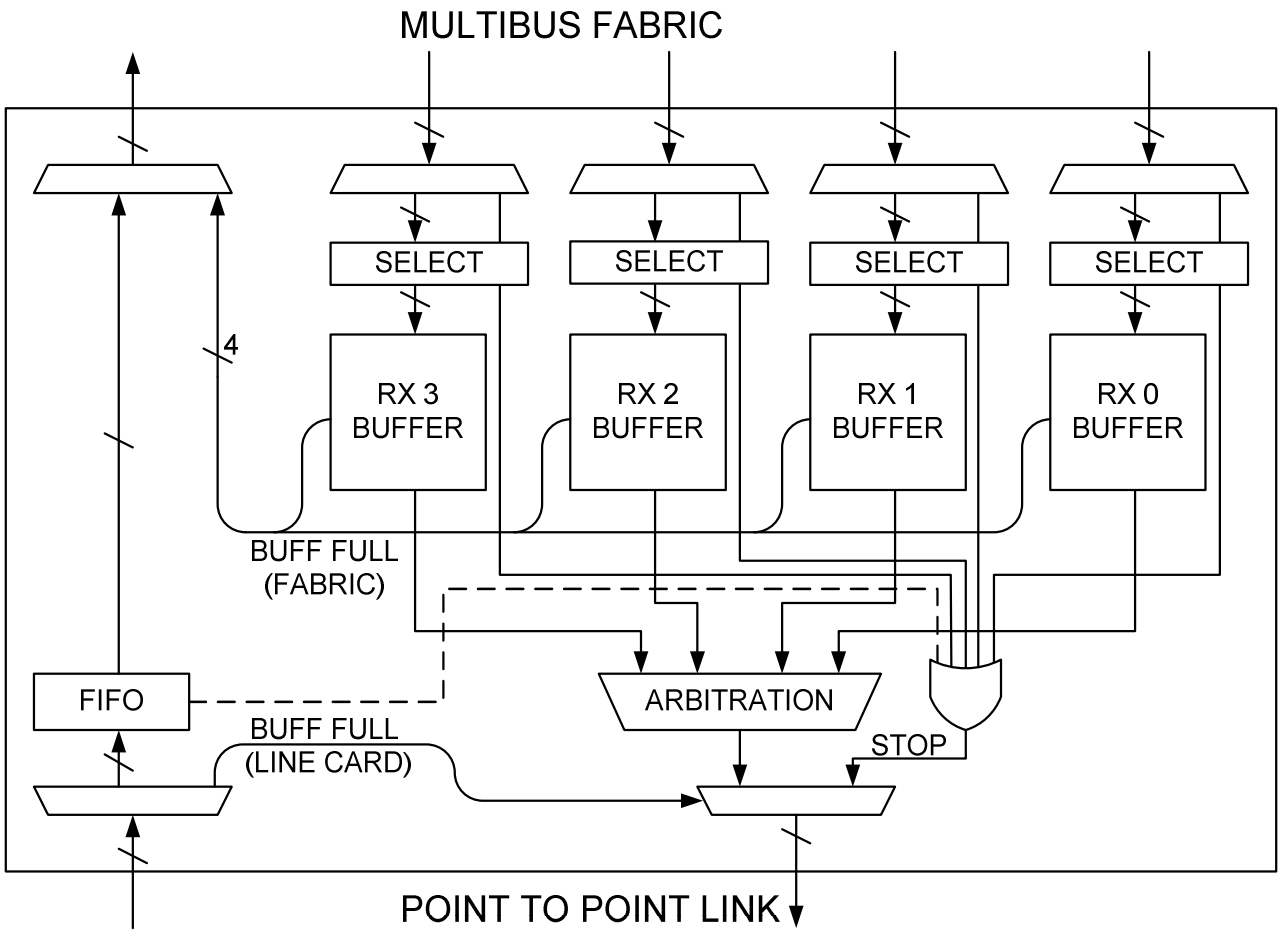
Multibus flow control

- Problem
  - All nodes can send to one node at the same time
- Each node:
  - Broadcasts the status of all its buffers
  - Receives the status of all N^2 buffers
  - Selects the relevant set of connected buffers.
  - Halt if any connected buffer is full
- Implementation
  - Roundtrip delay determines buffer size





Multibus flow control - 2



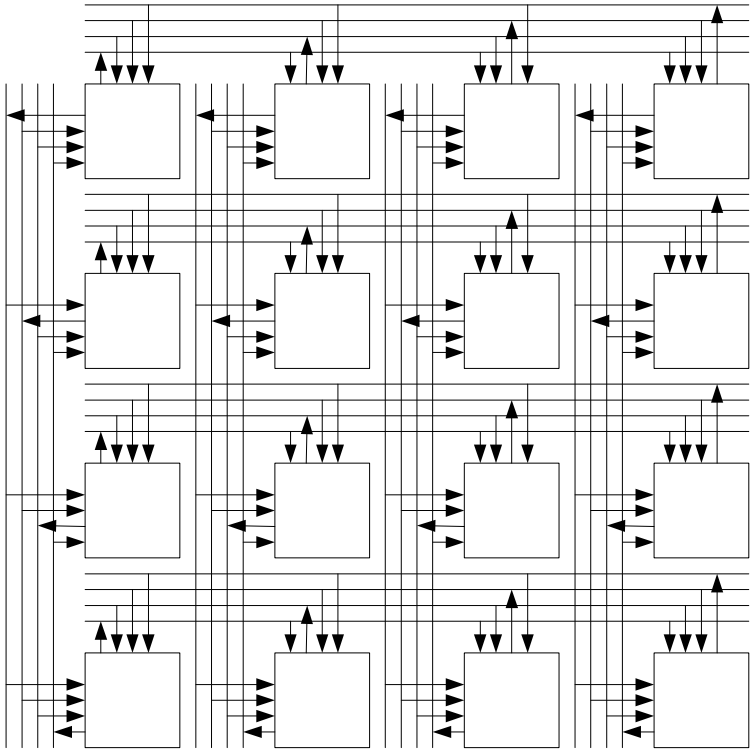
VCSEL based Bandwidth Scaling

- VCSEL BW scaling 10G → 25G
- Single λ → CWDM 2 λ → 4 λ
- Optical Fabric unchanged

	4x10G ports 12 Line card	12x10G ports 12 Line cards	6100G ports 12 Line cards	12x100G ports 12 Line cards
System Throughput	0.5 Tb/s	1.5Tb/s	7.2Tb/s	14.4Tb/s
No. of Waveguides	12x4	12x12	12x12	12x12
No. of colors	1	1	2	4
Base OE rate	10G	10G	25G	25G
Power Tx/Rx	5pJ/bit	2.5pJ/bit	1pJ/bit	1pJ/bit

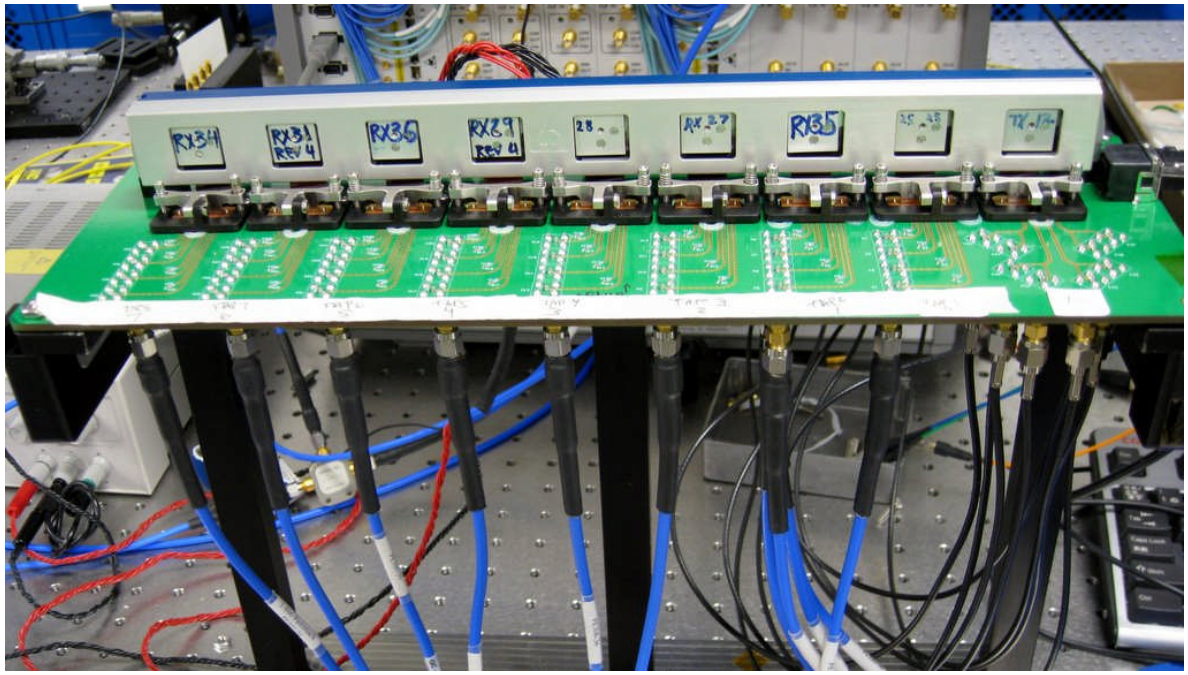
Node count scaling

- Limits to scalability
  - Broadcast bus fanout
  - N^2 in growth in receivers
- Solution
  - Multiple dimensions of multibus structures
  - Connections now go via intermediate node so must consider resiliency



Optical Bus Summary

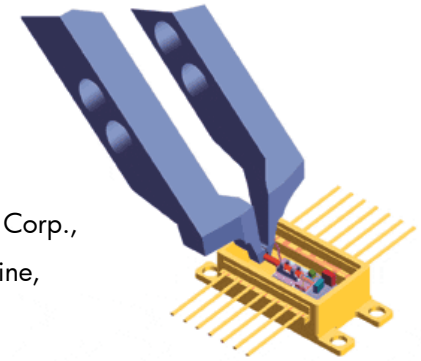
- Can build today
- Provides good fan-in and fan-out (>8)
- Distance not an issue
- Composite structures (e.g., crossbars) possible



Topic
The Optical Roadmap
Optical busses and backplanes
Optic-based networking
Intrachip optics – multicore NOCs

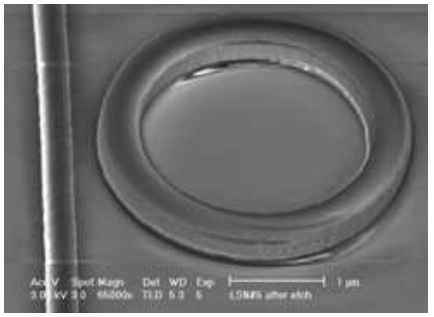
Integrated Photonics

- The 2000 telecom bubble based on discrete optics
  - Think pre-Noyce/Kilby era in electronics
  - Components are measured in mm
  - Hand alignment
  - Expensive and not scalable



Source: Newport Corp., Assembly Magazine, September 2001

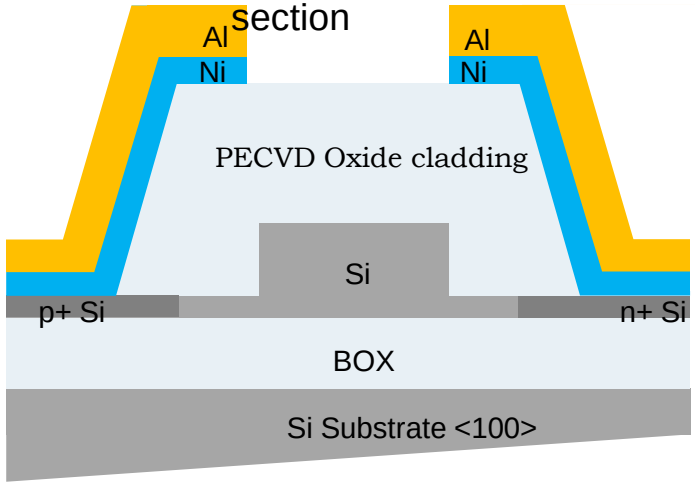
- Recent research is on integrated photonics
  - Think post-Noyce/Kilby era in electronics
  - Components are measured in a few μm
  - Manufacture many thousands per die
  - Advances in lithography -> better devices



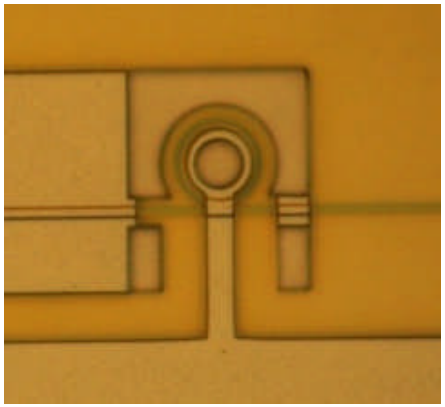
Silicon integrated circuits

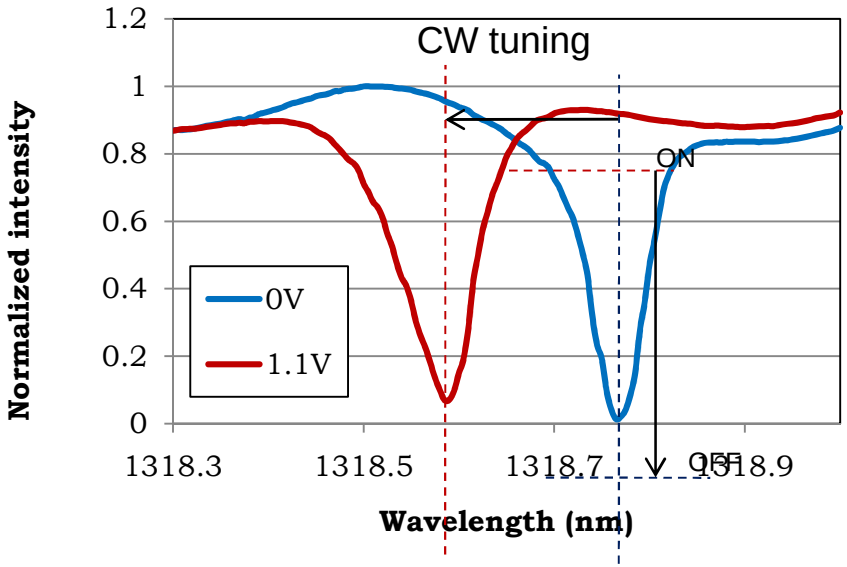
- 10 µm silicon ring resonators
  - Charge injection
  - 1310 nm (compatible with Ge detectors)
- Experimental Results
  - Q ~ 10,000
  - 0.18 nm shift
  - 18 dB extinction
  - 3 Gbps modulation
  - 54 fJ/bit

Cross section

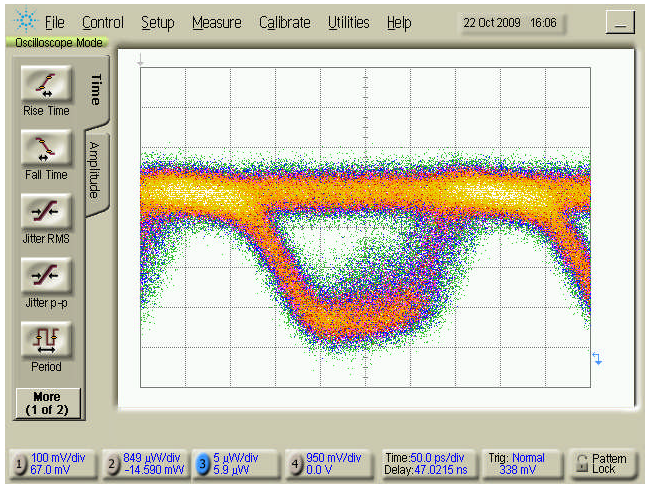


Top view



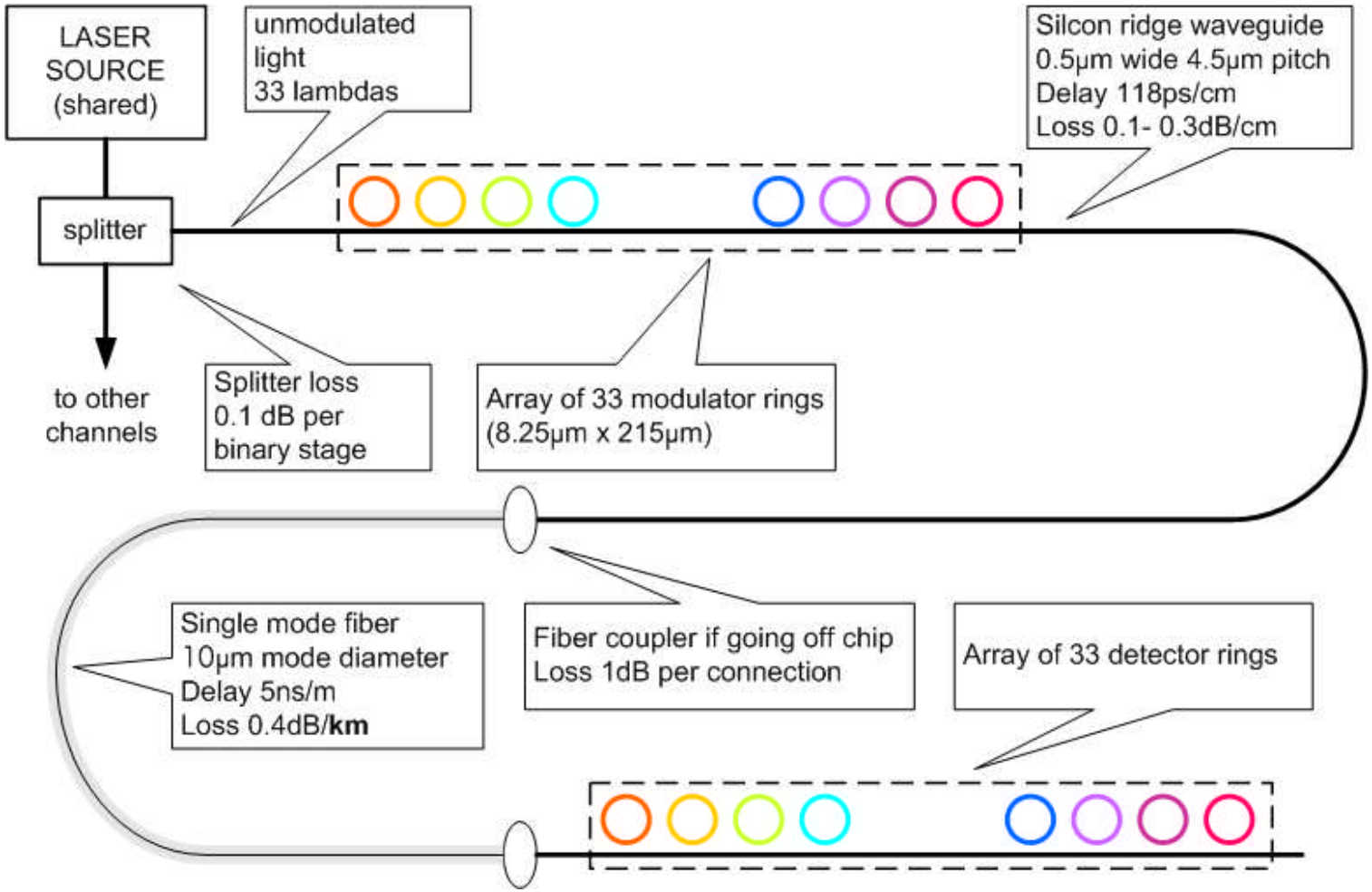


Eye diagram RZ 3 Gbps



Zhen Peng – HP labs

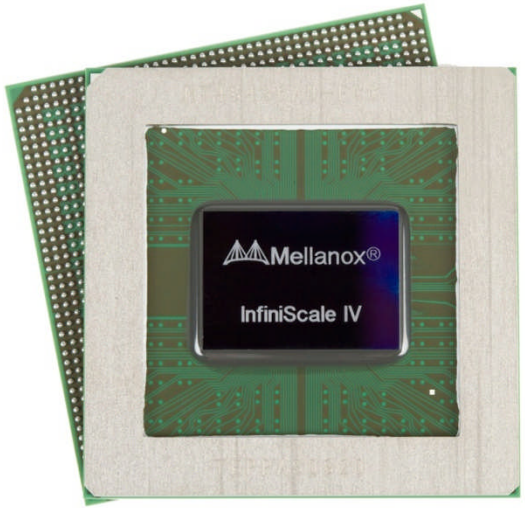
DWDM POINT TO POINT PHOTONIC LINK



High performance network switch

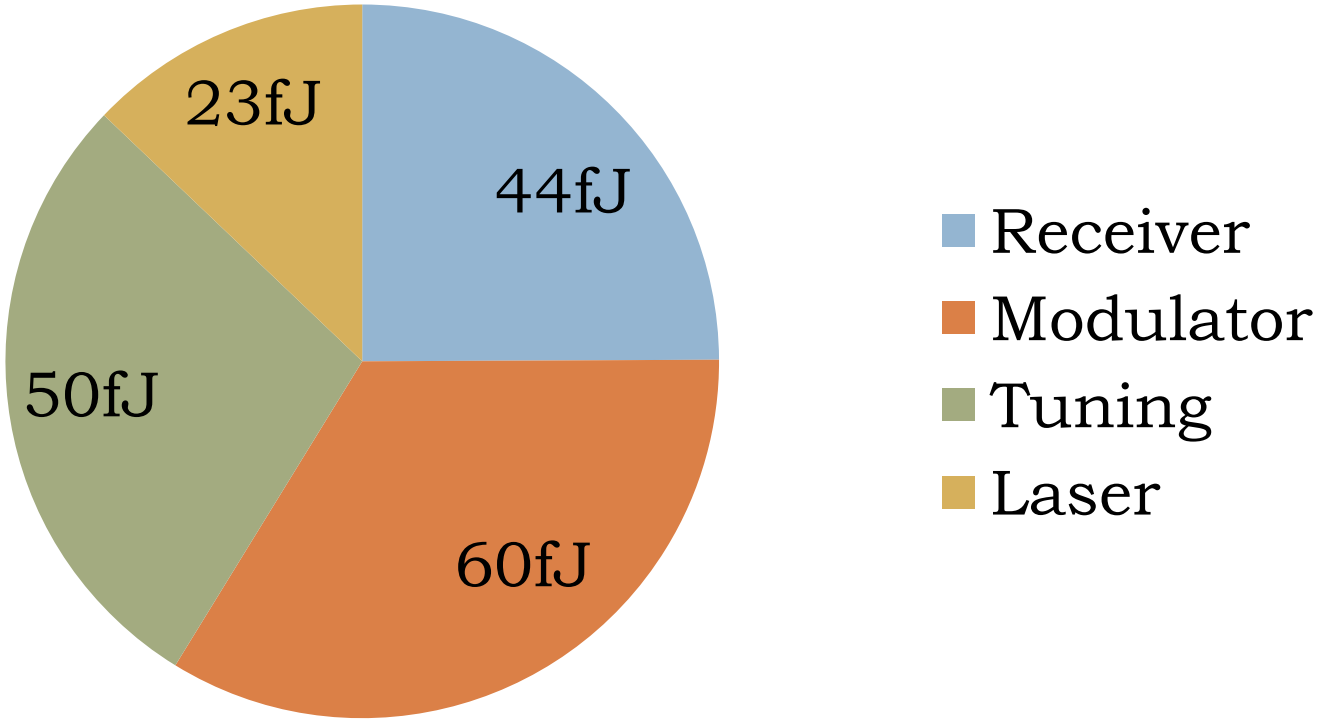
(electronic state of the art)

- Mellanox Infiniswitch IV
  - 36 ports @ 40Gbps or 12 ports @ 120Gbps.
  - 10Gbps per diff pair
  - 576 signal pins
  - 90W, 30% of which is IO



- Issues
  - Switch port count limited by pin count & IO power
  - Additional external transceivers needed to drive >0.7m FR4 or 6m cable
  - Increasing port bandwidth decreases port count
  - EMI & signal integrity problematic

Point-to-point power budget

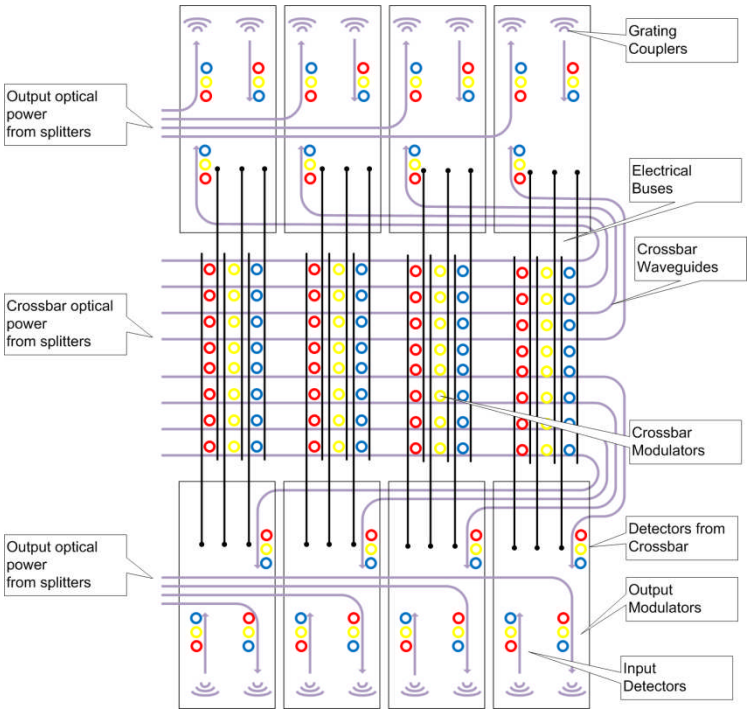


- 10Gbit/s per wavelength
- 177fJ/bit assuming 32nm process
- No clock recovery and latching - not directly comparable to electronic numbers

INTEGRATED CMOS PHOTONIC SWITCH

Characteristics

- 64-128 DWDM ports
- <400fJ/bit IO power
- 160 - 640 Gbps per port

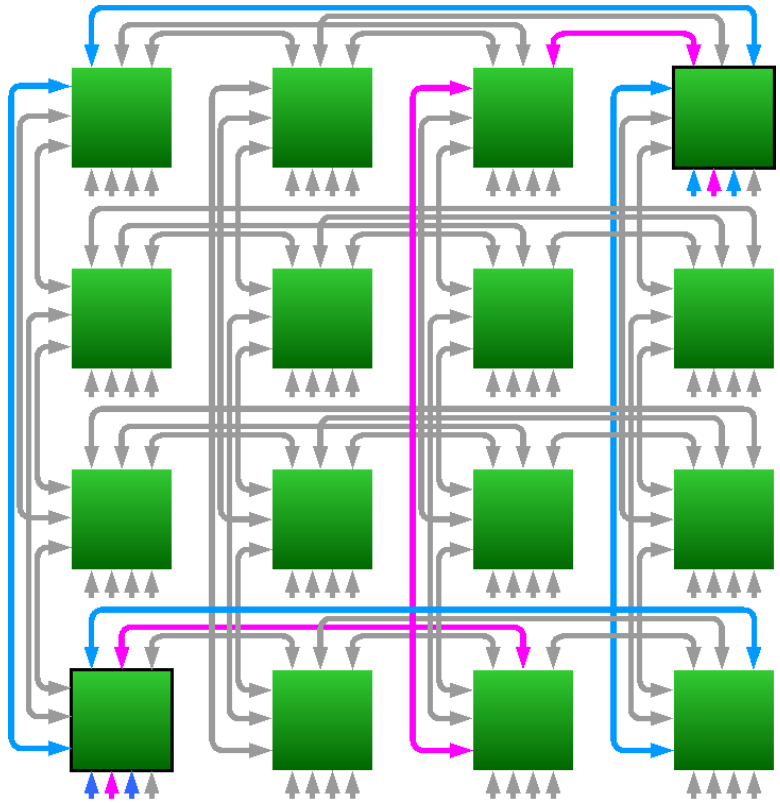


ADVANTAGES

- Switch size unconstrained by device IO limits
- Port bandwidth scalable by increasing number of wavelengths
- Optical link ports can directly connect to anywhere within the data centre
- Greatly increased connector density, reduced cable bulk

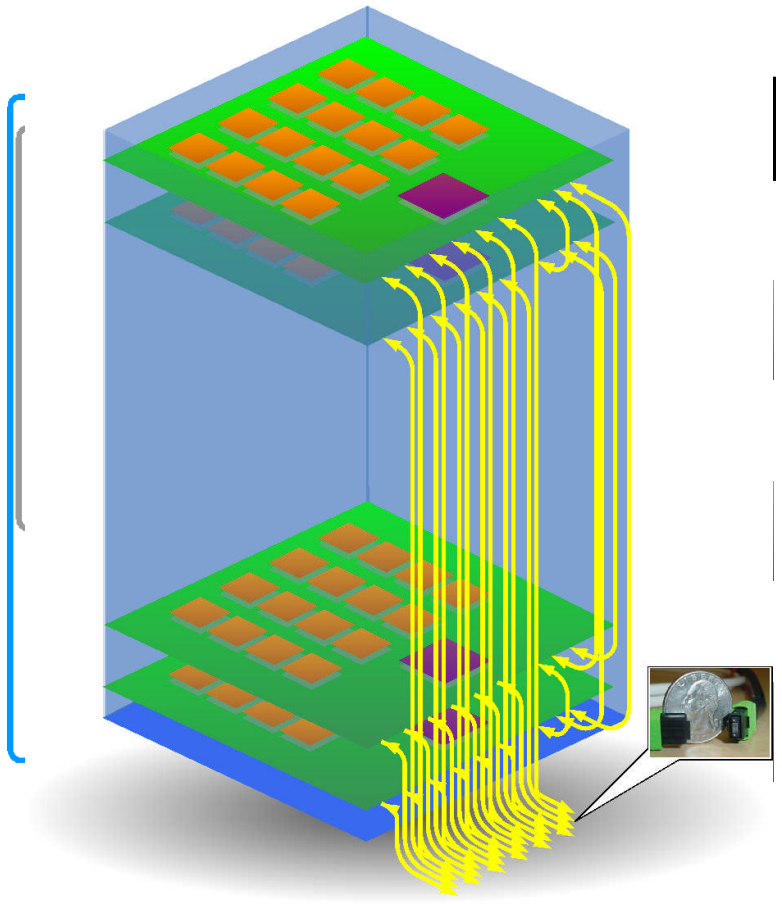
New network topologies - HyperX

- Direct network – switch is embedded with processors
  - Avoids wiring complexity of central switches (fat trees)
- Much lower hop count than grids and torus
  - But many different interconnect lengths
  - Higher interconnect density at module edges



New network topologies - HyperX

- Low hop count means:-
  - improved latency
  - lower power
  - less connectors
- Huge packaging simplification
- Anywhere in the data center in <1µs

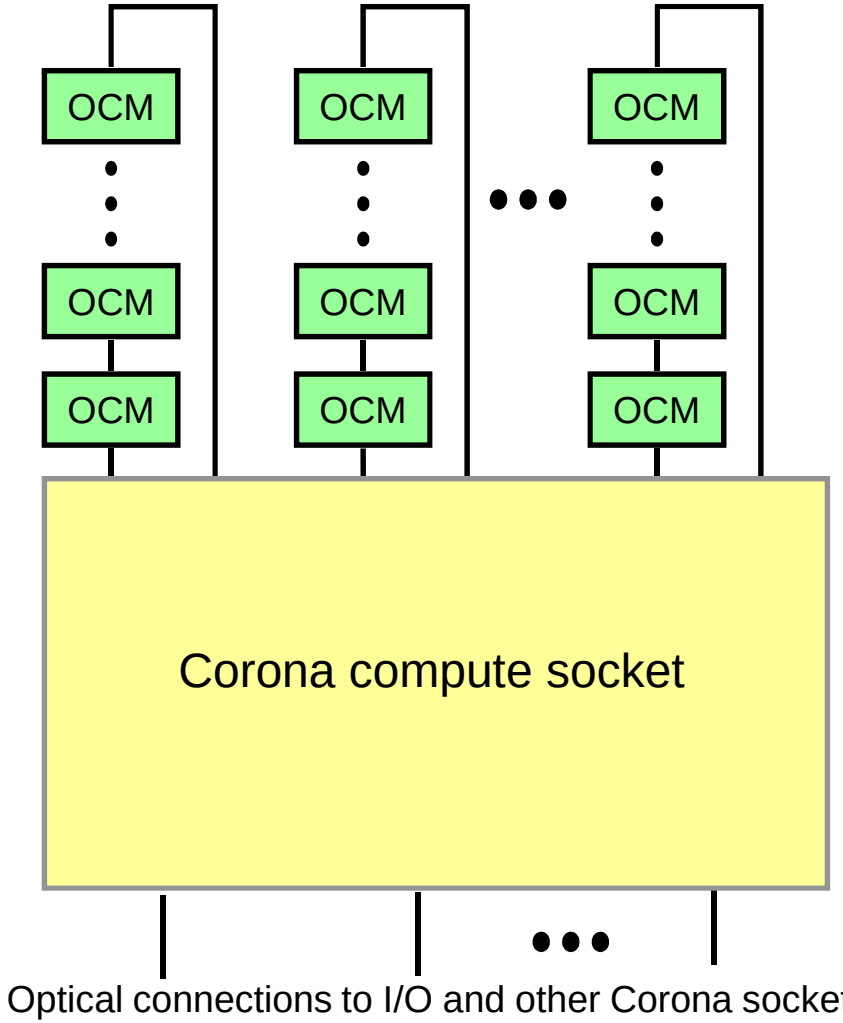


Topic
The Optical Roadmap
Optical busses and backplanes
Optic-based networking
Intrachip optics – multicore NOCs

The *Corona Manifesto*

- Take full advantage of nanophotonics
  - Don't just replace today's wires with optics
  - Redesign the multi-core processor from the ground up
  - No off-chip or cross-chip electrical wires
  - Restore balance: memory bandwidth scales with cores
  - All memory readily reachable from all cores

Corona System Overview



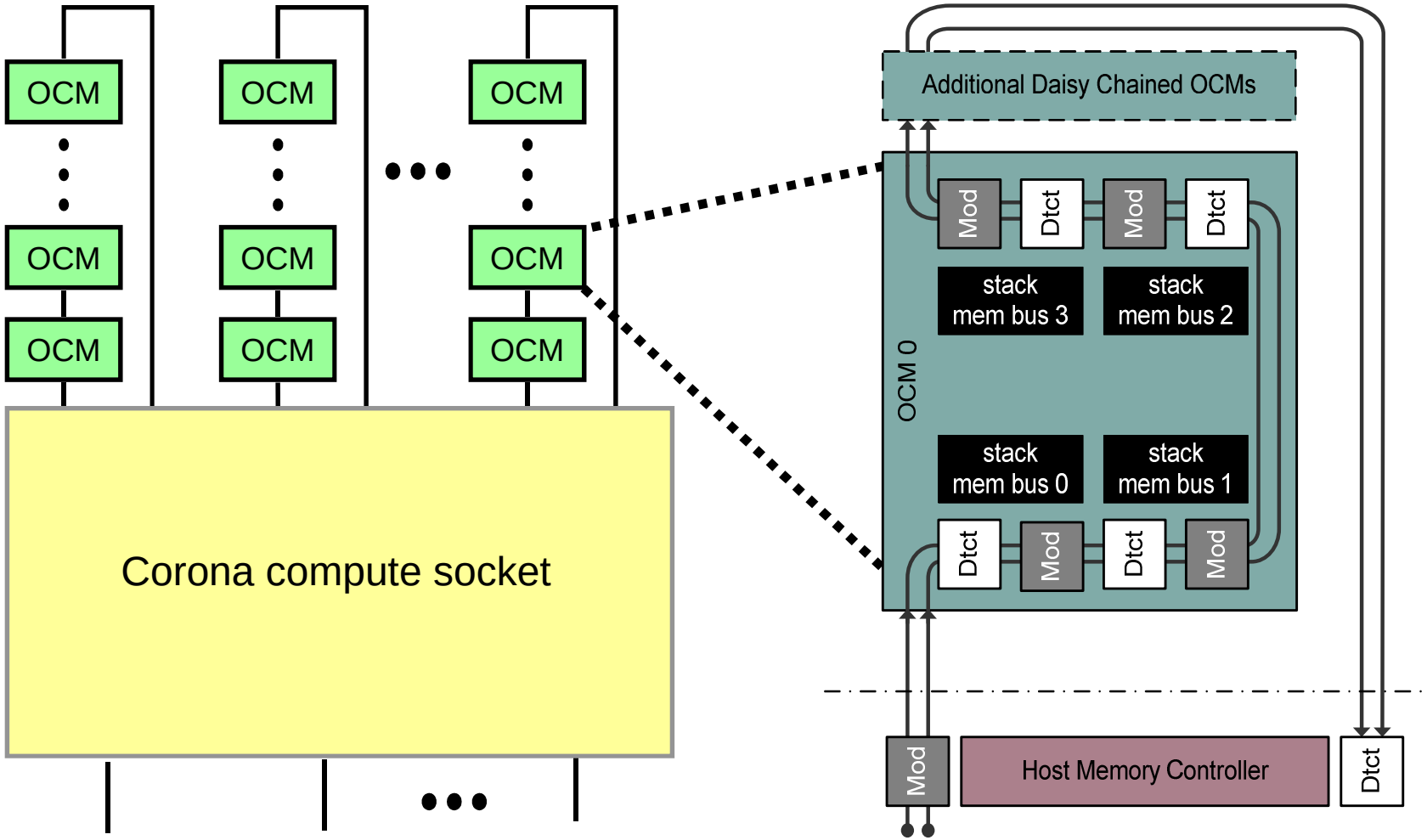
10 teraflops compute performance

10 terabytes/s memory bandwidth

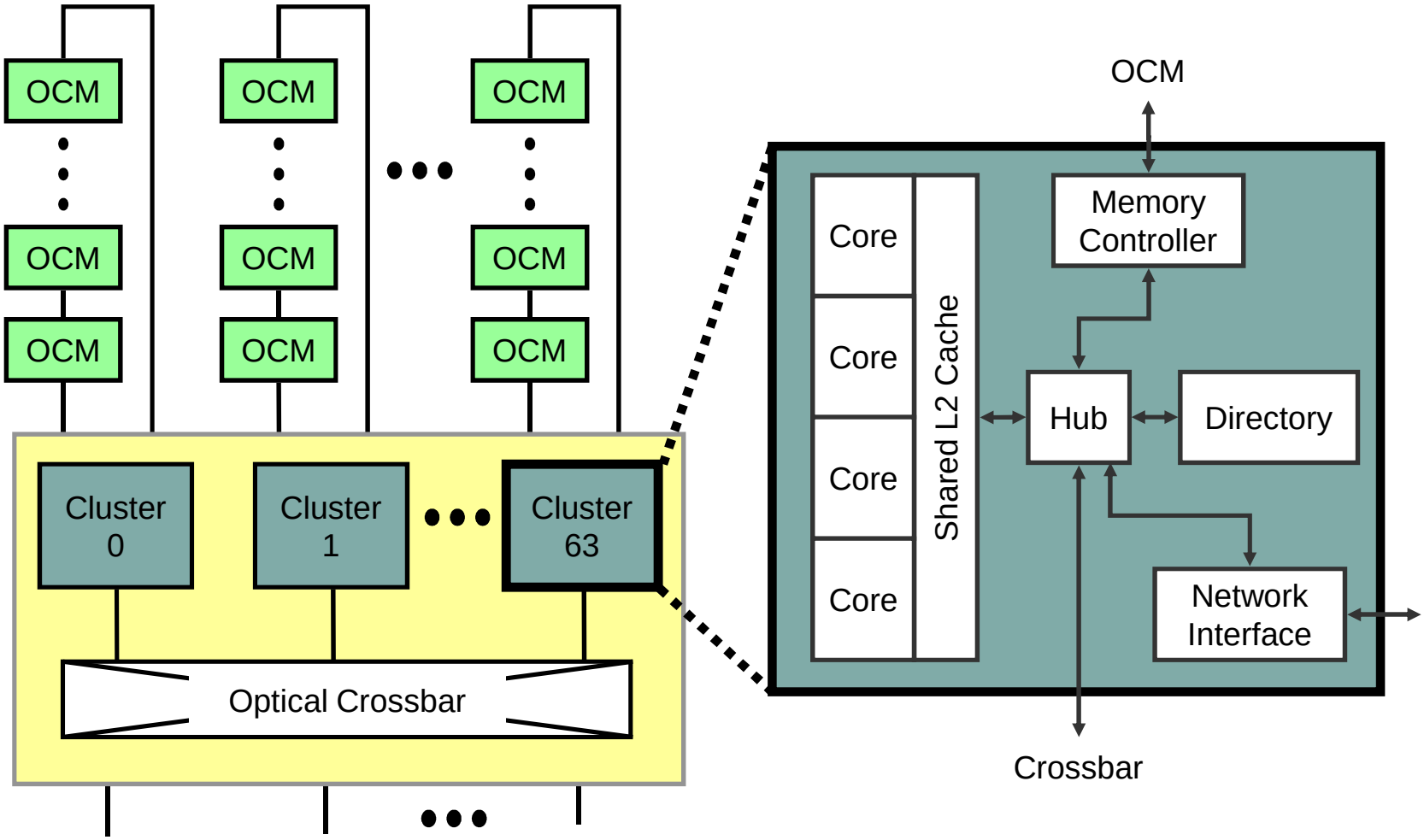
20 terabytes/s on-chip interconnect

All off-socket and cross-socket communication is optical

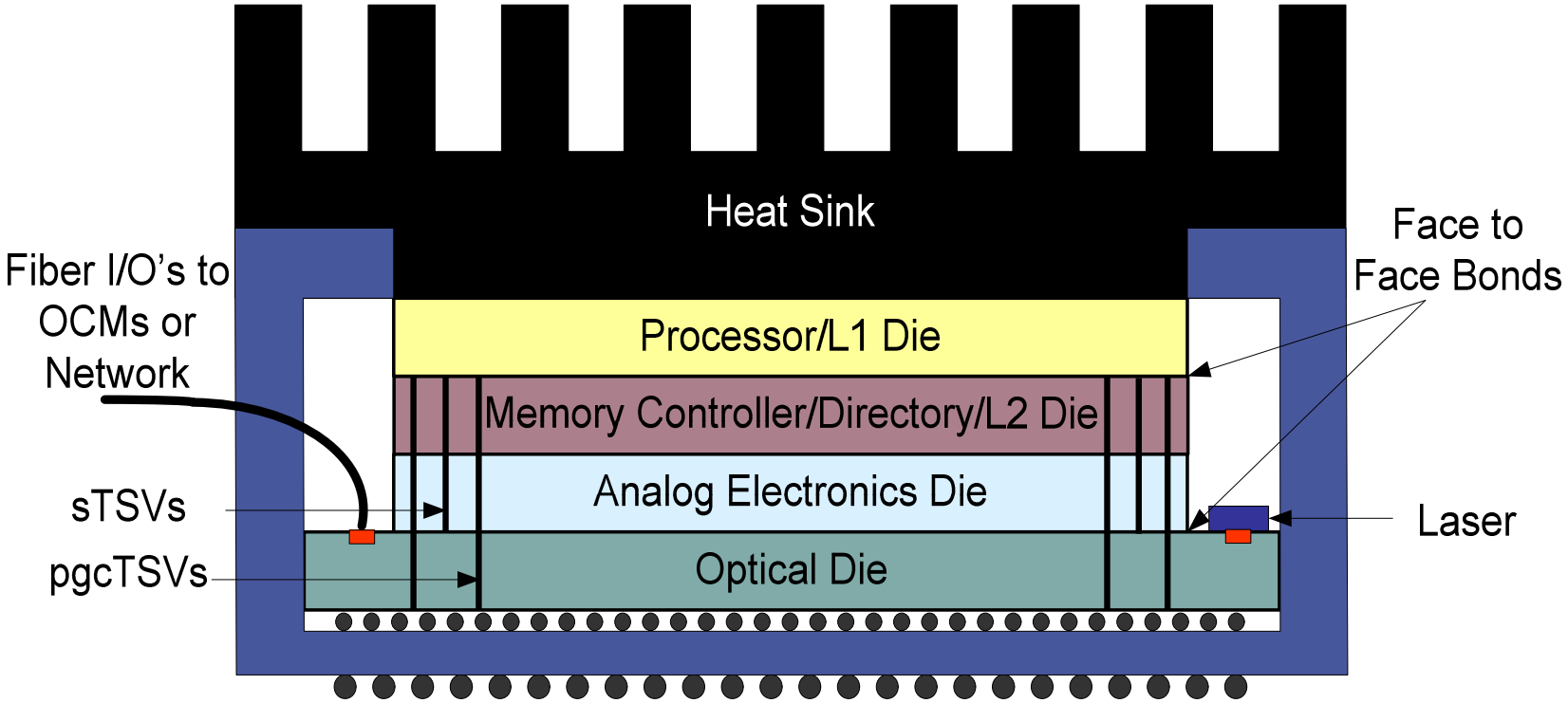
Optically Connected Memory (OCM)



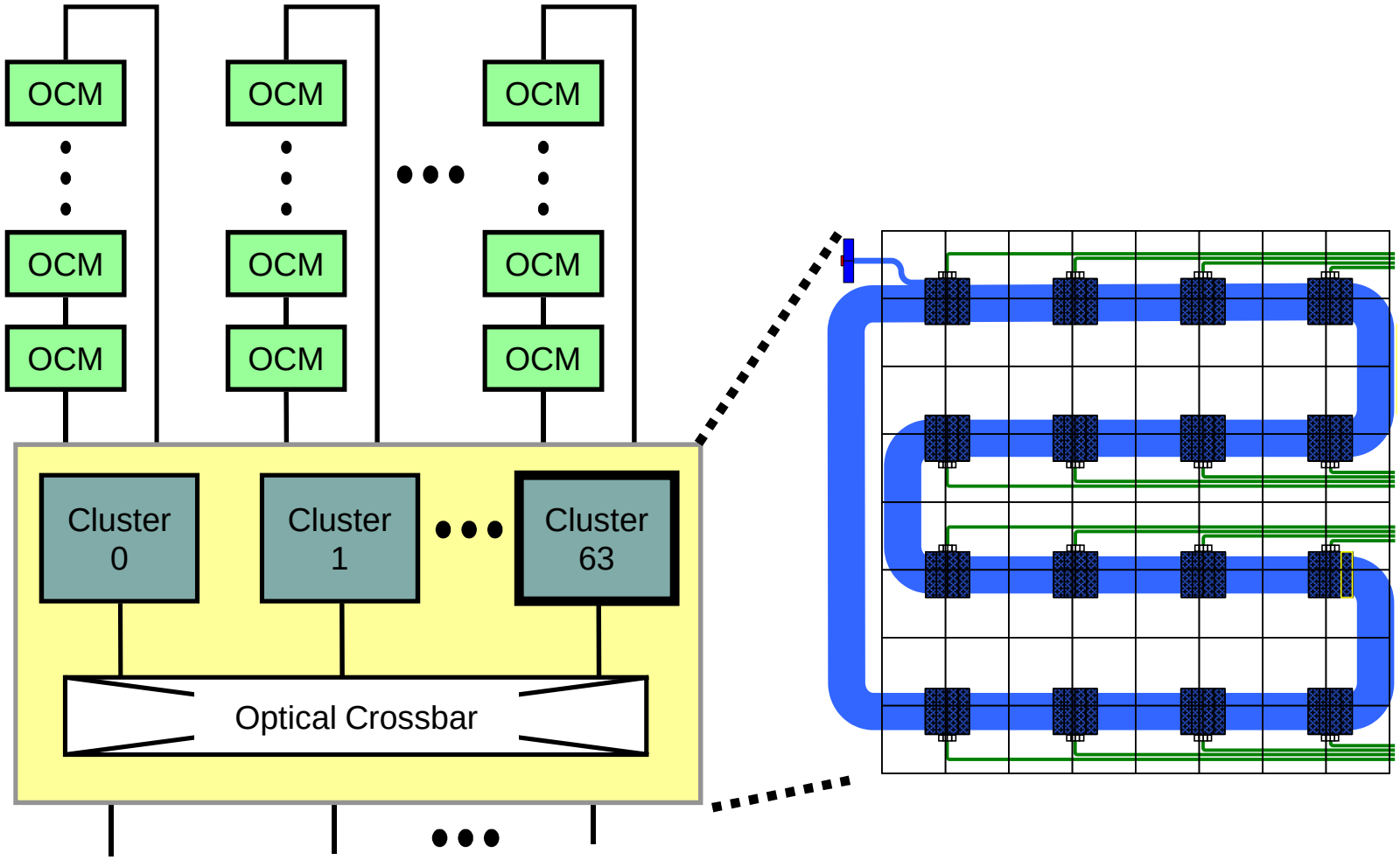
Corona Compute Socket



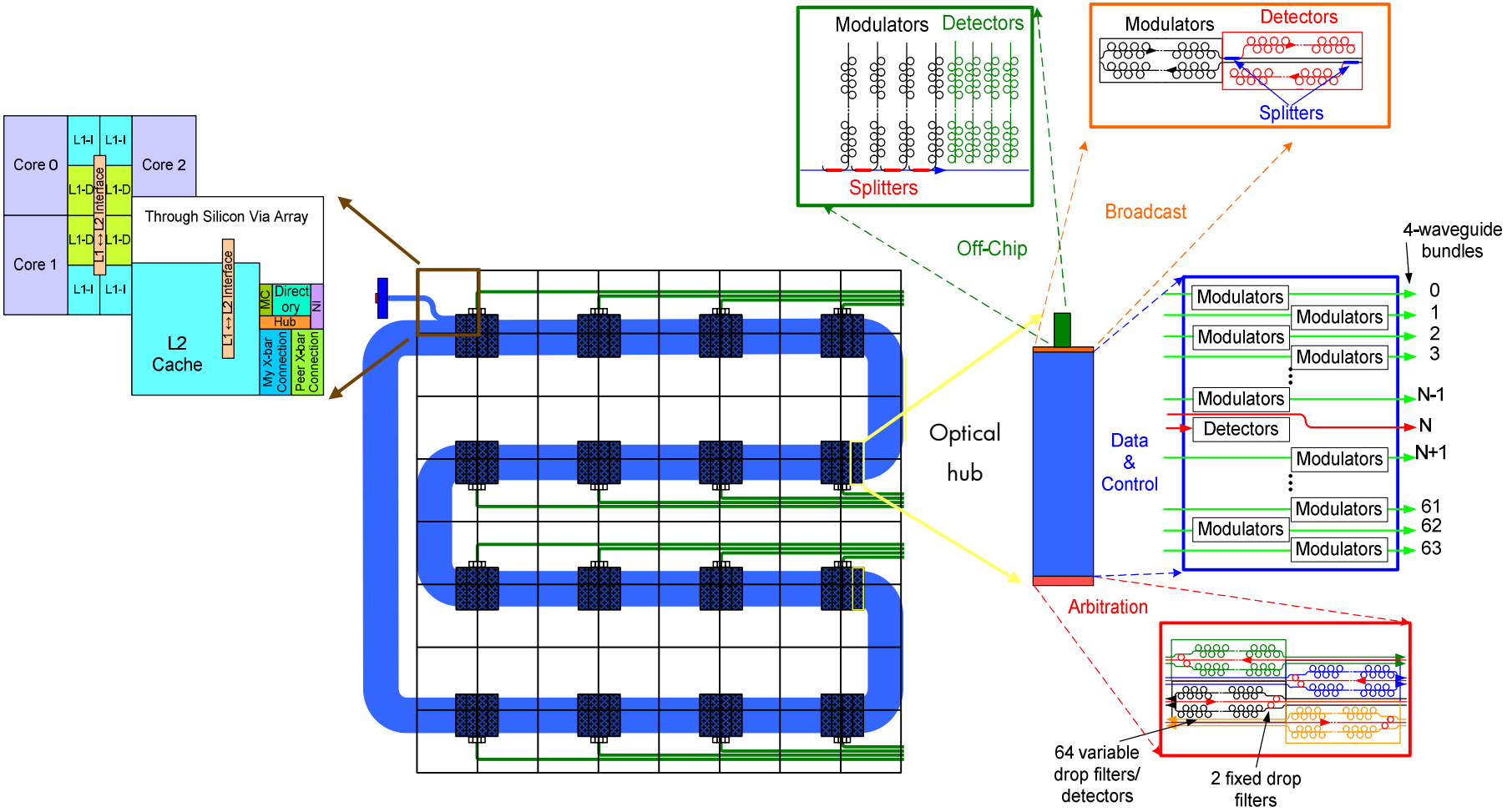
Corona Chip Stack



On-chip Interconnect

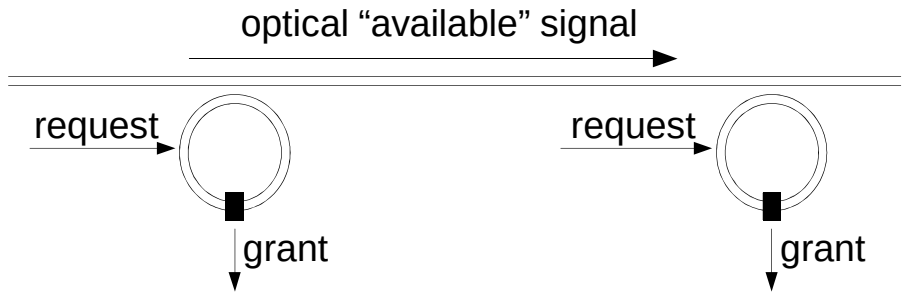


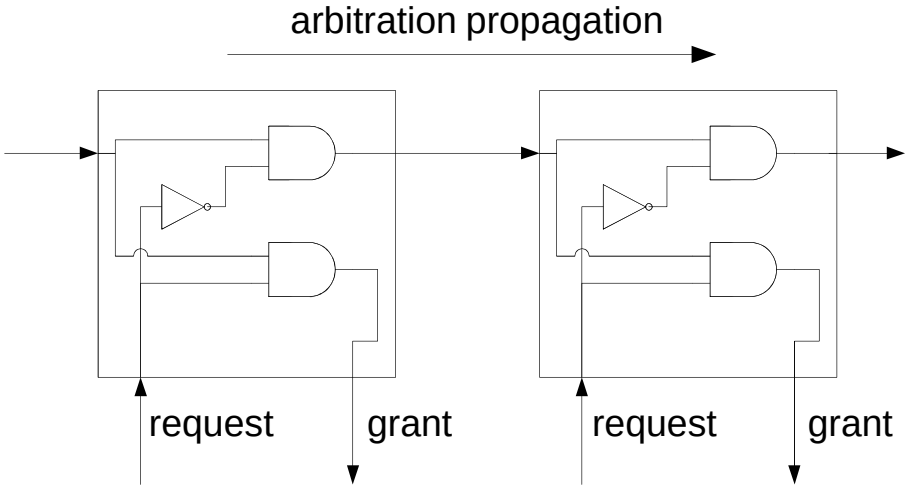
The Optical Crossbar



All-optical Arbitration

- A single micro-ring both asserts request and detects success or failure
- Requester tries to divert one wavelength
  - Detected power: success/failure
- Off resonance micro-rings add no delay and negligible loss –> highly scalable
- Arbitration time is light propagation time
- DWDM –> many concurrent arbitrations





equivalent electronic circuit

Conclusions

- Optical interconnect opportunity is greater than just “better wires”
- Distance crossover point at which optics becomes interesting is rapidly reducing
- Need to reconsider system architecture to best exploit optics

Questions?