Optics and the exascale data center

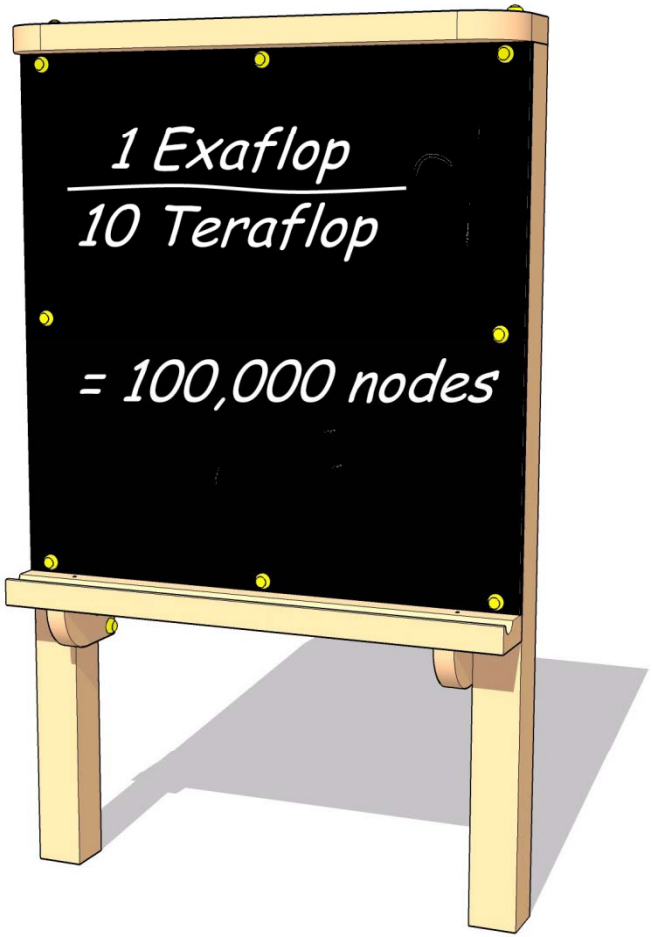
Moray McLaren

Exascale Computing Lab

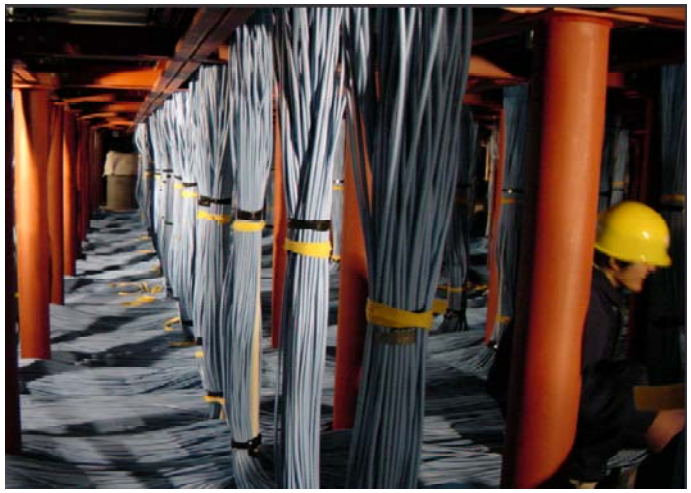
Optical interconnects – gold plated plumbing for computers?

Exascale systems arithmetic…

- 10 Teraflop manycore processor by 2017
  - 256 cores per socket
  - ~200W
- Total system power 20MW
- How do we connect them all up?



LABS hp



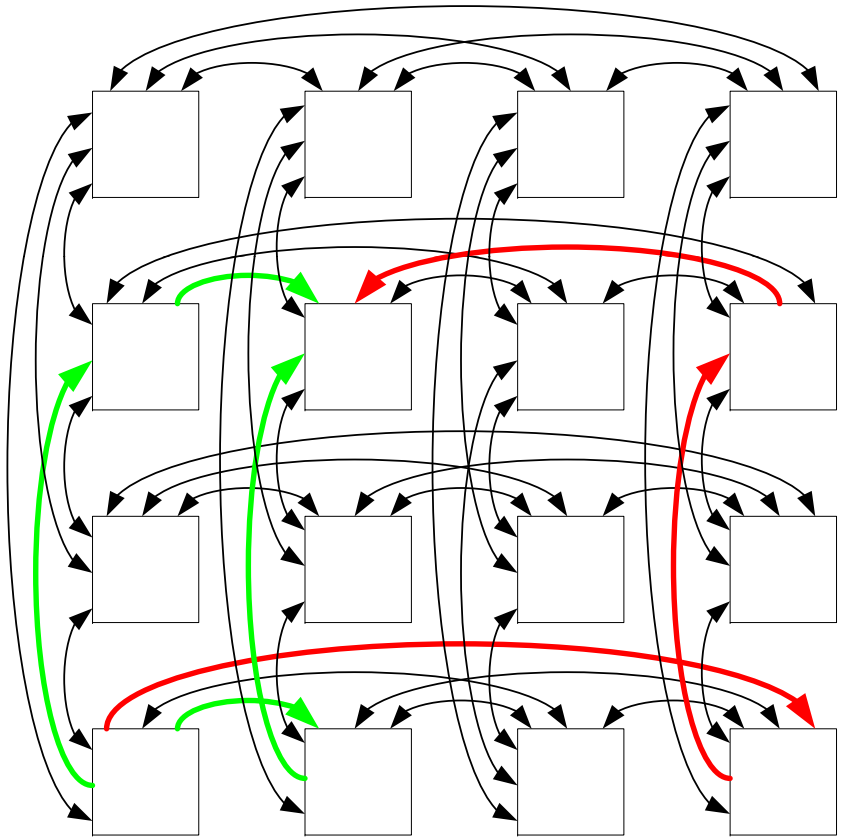
Not copper! – too bulky

Not centralized – too difficult to wire

Not a grid or torus – too many hops

Fully connected sub-networks in multiple dimensions*

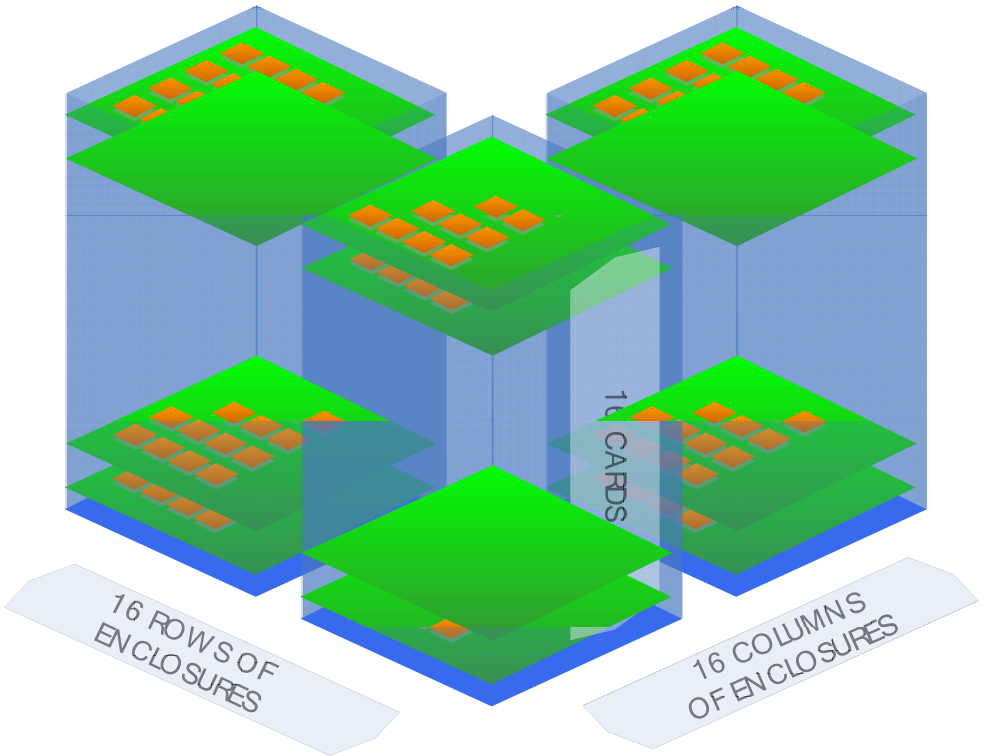
- Direct network – switch is embedded in processor
  - Avoids wiring complexity of central switches (fat trees)
- Much lower hop count than grids and torus
  - But many different interconnect lengths
- Low hop count means:-
  - improved latency
  - lower power
  - less connectors



*Dally & Kim, Flattened Butterfly

Example exascale photonic interconnect network

- Embedded high radix routers allow the construction of direct networks with very low hop counts.
- Example network – 64K nodes arranged in 256 enclosures of 16 cards
- Maximum 4 hops between any two nodes
- Chip links must span the entire data center



Exaflop interconnect requirements

- How many connections?
  - Assume 4 dimensions (board, rack, X, Y)

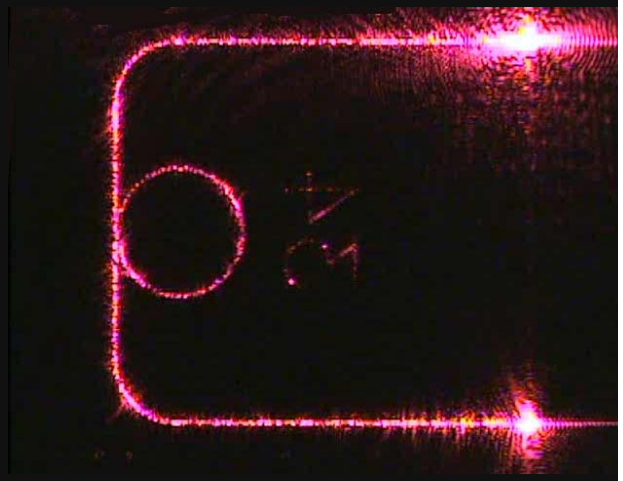
$$= (\sqrt[4]{100{,}000} - 1) \times 4 \approx 68$$

- How much bandwidth?
  - Assume compute communication ratio of 0.1Bytes/flop

$$= \frac{10 \times 1000 \times 0.1 \times 8 \times 4}{68} = 470 Gbit/s$$

  - 47 lambdas at10Gbit/s modulation
- How much power?
  - Assume IO power budget of 5%

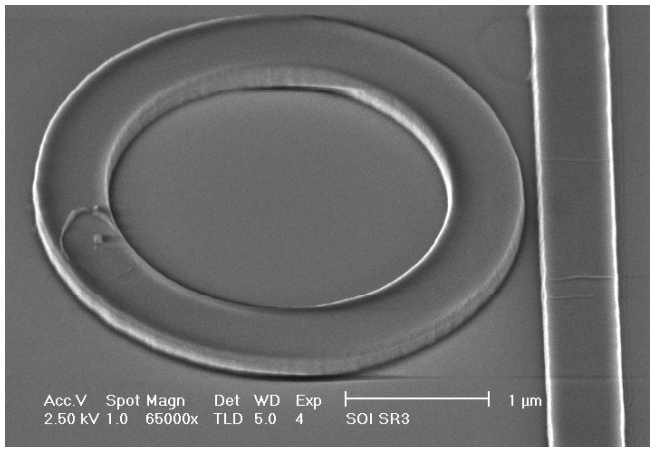
$$= \frac{200 \times 0.05 \times 1000}{470 \times 68} = 0.312 mW/Gbit/s$$

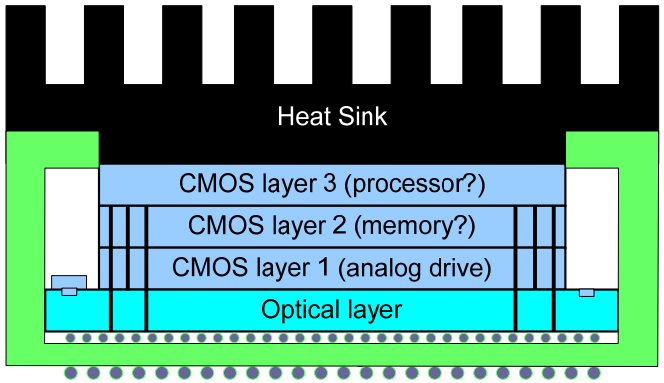


Power and bandwidth targets require integrated photonics

Ring resonator based CMOS Integrated Photonics

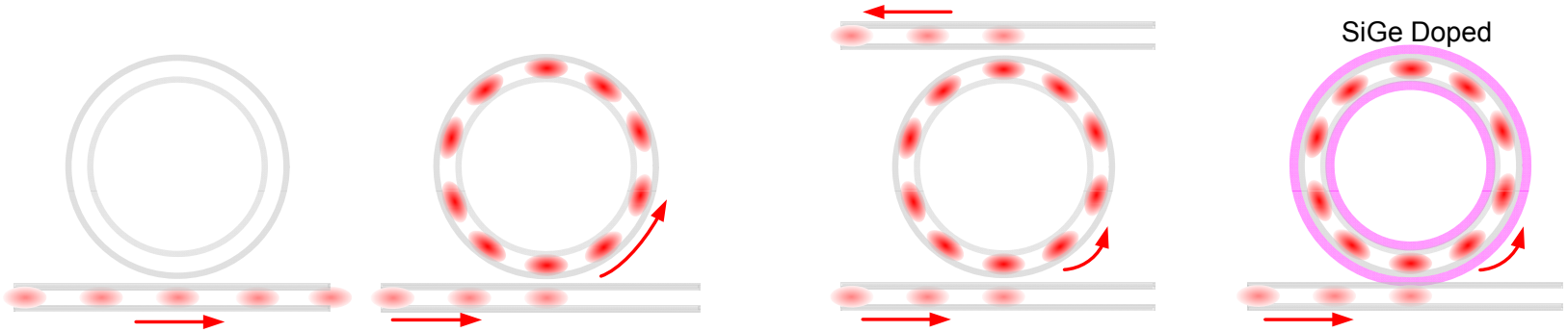
- Why ring resonators?
  - Inherently DWDM
  - Potential for very low power operation
  - Small silicon area
- Use stacking to avoid modifying CMOS processes
- Tuning and temperature control issues…





Ring Resonators

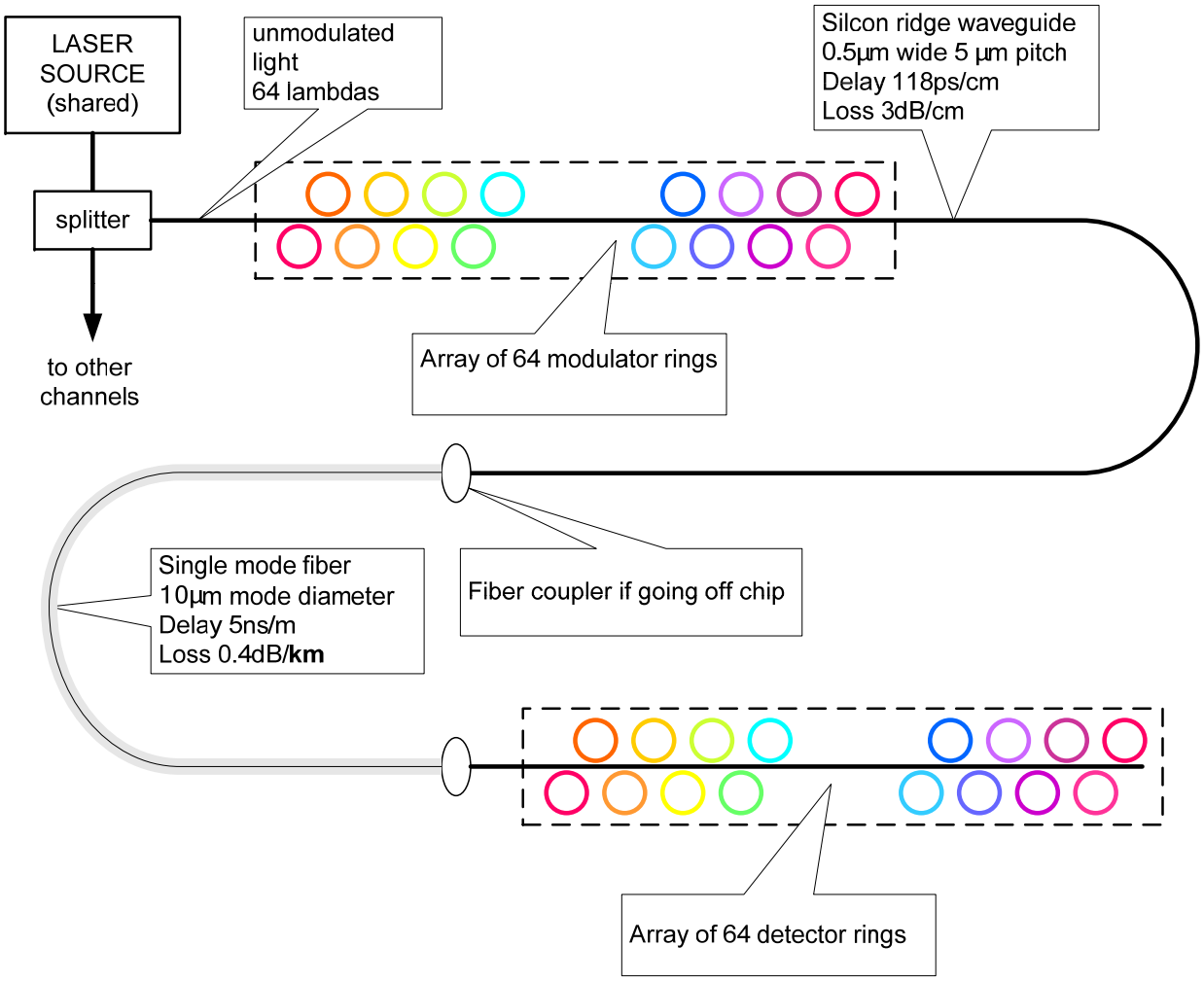
One basic structure, 3 applications



- A modulator – move in and out of resonance to modulate light on adjacent waveguide
- A switch – transfers light between waveguides only when the resonator is tuned
- A wavelength specific detector - add a doped junction to perform the receive function

LABS hp

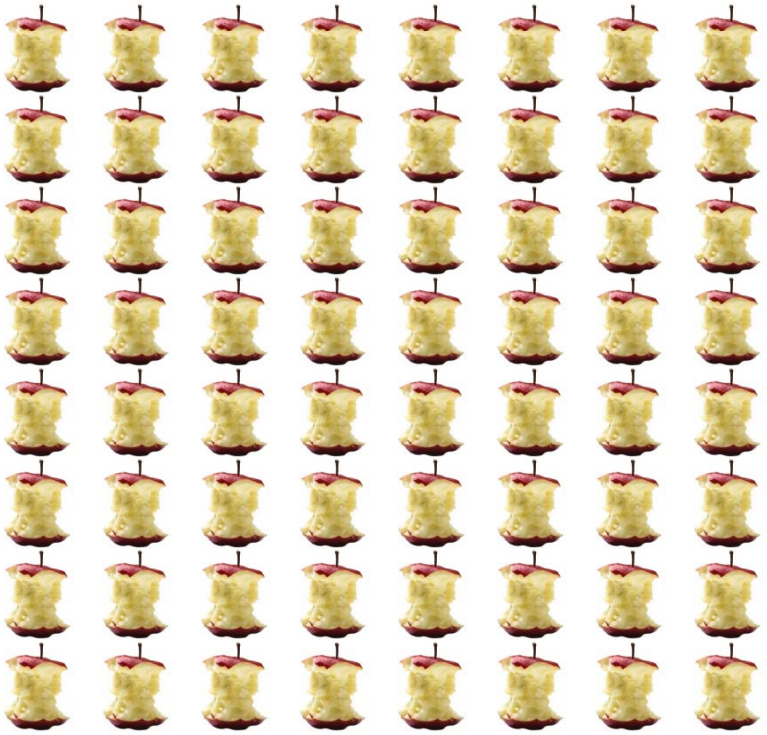
Point to point DWDM link



Wavelength (nm)	1310
Wavelengths	64
Channel Spacing (GHz)	80
Modulate Freq (GHz)	10
Data Rate (Gbytes/s)	80
Link Power (mW)	64
Energy (µW/Gb/s)	100

What about less extreme systems?

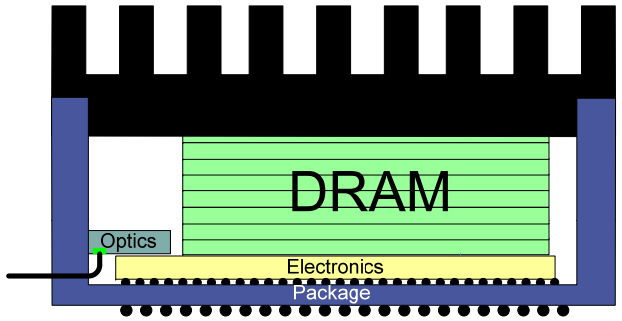
- Processor performance continuing to grow through core count scaling
- Memory bandwidth not scaling with core counts
- Ideally would like 1byte per flop memory systems
  - 10Tbytes/s in 2017
- Capacity scaling equally important



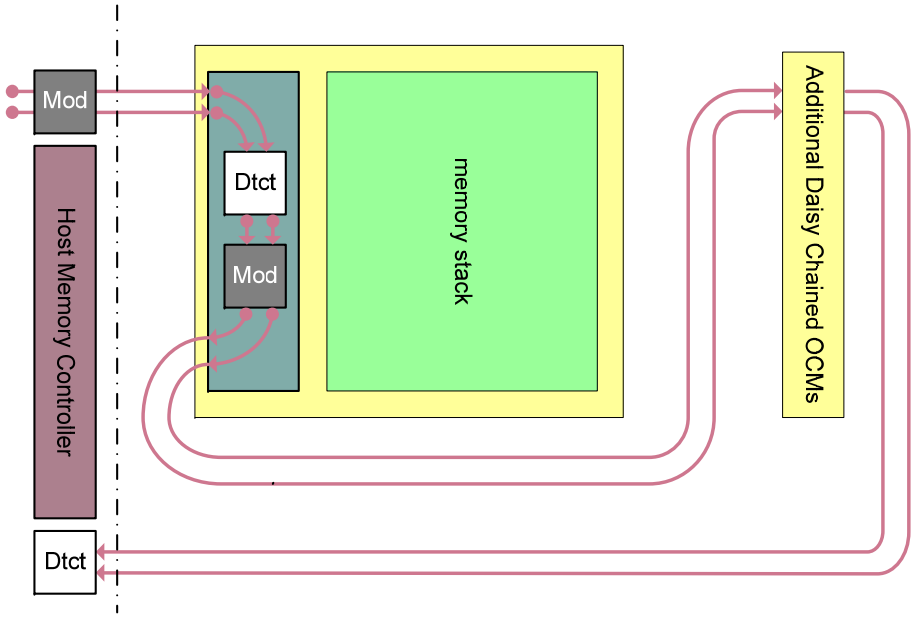
LABS hp

DWDM optics to the memory

- no laser on module, power from bus master
- controller/interface cost amortized over multiple DRAM stacks
- daisy-chaining to further modules for memory expansion

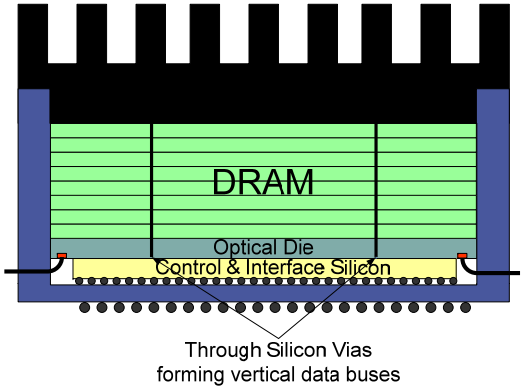


- standard stacked DRAM technology
- use widest interface possible for BW

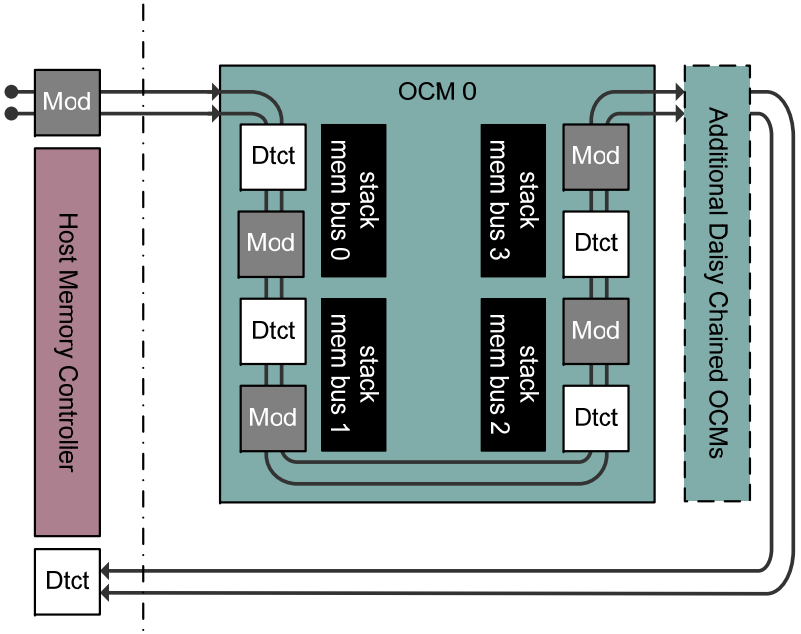


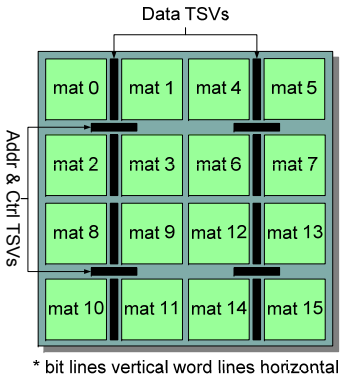
Optics to multiple points across a chip

- reduces requirement for DRAM global interconnect saving power and access time
- daisy-chaining to further modules for memory expansion



- smaller DRAM mats
  - ½ power consumption
  - ½ access time





Key requirements for integrated photonics

- Power - anything lower than 0.5mW/Gbit/s is interesting for chip to chip IO
- Further reducing power makes intrachip interconnect attractive
- Connectivity - ability to make 100s of optical connections per device to exploit integration

Conclusions

- Today – optics necessitated by distance
- Tomorrow – optics to create differentiated products
- Future – optics necessitated by power & bandwidth


LABS hp