Optical Interconnects For Supercomputers.

- Pradheep Khanna Kaliraj
- Manoj Prashanth Yuvaraj.
Outline

- Motivation.

- Photonic interconnects – Super computers (SC).

- Towards Optical Interconnects.
  1. VCSEL
  2. Various optical switching tech.
  3. Polymer based waveguides.

- Future of interconnects in SC.
Motivation.

1. Exponential growth of performance in SC—about 60% per yr.
2. Power scales up with the performance.
3. Power consumption will be in “gig watts” for Exascale SC (year 2020)
4. Ideal Scalability
   - Ten fold increase in performance
   - Two fold increase in power consumption.
5. Requires extreme power take downs.
   - Optical links.
## Estimated Performance

- Power and Cost estimate.

<table>
<thead>
<tr>
<th>Year</th>
<th>Peak Performance</th>
<th>Bidirectional Optical Bandwidth</th>
<th>Optics Power Consumption</th>
<th>Optics Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>10PF</td>
<td>1PB/s (10^7 Gb/s)</td>
<td>0.5MW</td>
<td>$22M</td>
</tr>
<tr>
<td>2016</td>
<td>100PF</td>
<td>20PB/s (2x10^8 Gb/s)</td>
<td>2MW</td>
<td>$68M</td>
</tr>
<tr>
<td>2020</td>
<td>1000PF (1EF)</td>
<td>400PB/s (4x10^9 Gb/s)</td>
<td>8MW</td>
<td>$200M</td>
</tr>
</tbody>
</table>
Energy is the Bottleneck in data transfer.

1. 200 times more energy is required to transfer data from a node to nearest neighbor.
2. Data transfer on card – 0.05–0.1 pj/bit
3. 1000x for intersystem data transfer across a distance of 50m.
4. Data transfer can no longer use traditional systems.
5. BW should scale along with the system performance. 
   ~1B/FLOP.
6. Requires exponential increase in bandwidth at all levels of the communication.
Electrical links

– Terrible Power efficiency.
– High frequency loss.
– Lower Bandwidth
– Size and weight of Copper cables.

Future interconnects?

– Power efficient
– Density of interconnects should be improving.
– Cost, Should be less
– Fast.

Optical interconnects became ubiquitous.
Advantages of Optical interconnects.

1. High speed.
   Speed does not depend on the number of receiving components.
   i.e. Zero loading capacitance.

2. Large Bandwidth
   Multimode fibers can provide data rate of several Gbps.
   Low cost and Easy to couple.

3. Less interference
   Stray capacitance is absent, hence low interference.
   Scattering of light is still a problem.
   Less interference == Reduced power requirement.

4. Flexibility in Routing signals.
   Does not encounter cross-coupling.
   Major advantage over Electrical interconnects.
Optical interconnects in a Supercomputer.
- Campus level, LAN, SAN.
- System level, Intra rack, Rack to Rack.
- Board level, Module to Module, Chip to chip
- Chip level, Intra chip interconnects.

#. Some of the interconnects are already implemented in the SCs.

Inter system optical interconnects –

1. **Road Runner**.
   All the node to node communications are through optical links in the Road runner super computer.

2. **POWER&IH**.
   Board to Board communication is through optical interconnects.

Machines with a peak performance closer to 10PF, where optics play an important role in it.
Inter System Optical Interconnects.

– The optical interconnects, today, are used for inter-rack communication.

– The optic cables are placed at the edge of the card far from the Host.

– Active optical cable provide the interconnection with optics hidden inside cable connector housing.

– The cables are mounted on the edge of the system like electric cables.

– IBM Road runner, active optical cable assemblies that runs for 92 km.

– The presence of electrical cables to connect the CPU with the optical lines has its effect on power consumption and Density.
- Long electrical data lines limit the data rate due to attenuation.

- Requires equalization at both ends, and intermediate clock and data recovery circuits.

- These increase the power consumption in a Electro – optical link.

- Small packaging factor advantage of the optical links cannot be tapped because of the need for compatibility between the Electrical and optical links.

- Active optical cable technology is a success for today’s SCs but they cannot continue to deliver the same success in the future SCs.

- The answer : Optics has to be placed closer to the host adaptor or CPU.

- Requires Tighter integration of the optics with the system.
Towards Optical Interconnects.

- Current Photonic Link Technologies
Vertical Cavity Surface Emitting Laser

“The vertical-cavity surface-emitting laser (VCSEL) is a type of semiconductor laser diode with laser beam emission perpendicular from the top surface”.

➢ The laser resonator consists of two distributed Bragg reflector (DBR) mirrors parallel to the wafer surface with an active region.

The advantages over edge emitted laser are

1. This can be tested at certain stages.
2. Can be produced simultaneously.
3. The yield can be controlled.
**Characteristics**

- The fabrication cost is less because they emit only from the top surface.
- Large output aperture, low divergence angle of output, high coupling efficiency.
- Wavelength can be tuned by adjusting the thickness of reflective layers.
Optical Switching

- Inter processor communication for super computers require very large bandwidth and degree of parallelism.

- Crossbar based on space division is costly and are not economically not feasible.
Guided waves

- Optical wave guides used in large space division switches.
- The increase in the optical path does not reduce bandwidth.
- There is negligible attenuation on optical fiber for short distances.
- The switching network has dense fabrication of optical waveguides.
A 2 x 2 optical switch can be implemented directly or crossed.

Connection is controlled with electric voltage or beam of light.

The latter is an intersecting waveguide where the channels intersect at a small angle.
A Crossbar network for matrix vector multiplication.

- The inputs are light sources. These are focused on the crossbar mask. The mask has optical shutters.
- It has optical shutters which are dynamically controlled.
- The mask provides desired interconnection pattern.
TDM and WDM

- Currently electronic buses are used for interconnection.
- This provides limited bandwidth.
- Instead the use of optical network as backbone will provide better bandwidth.
- The recent development if multiplexing schemes have helped implement the optical networks.
In TDM systems the data is concentrated and transmitted on fiber using TDM.

They usually use star coupler which has passive beam splitting.

Signals from each transmitter are multiplexed and combined by coupler.

The clock synchronization must be maintained properly.
In WDM, the problem of clock synchronization is eliminated.

A fixed wavelength is assigned to a transmitter.

The receiver can tune to any desired wavelength from the coupler.
Chip to chip switching

- The interconnection must support high data rates.
- It must also have high degree of parallelism.
- The interconnection must be flexible to support various board layouts.
- The use of electronic buses increases the impedance which results in high power which leads to increase in size of chip.
Holographic Switching

- The communication on chip are of short distances. The free space is used for effective communication.

- This allows the use of the third dimension where are all the beams of light can cross without interference.

- These are implemented using non coherent white light and bulk optics.
The optical signal distribution can be improved by using this technique.

The hologram is used to change the direction of the light rays.

The main element is the Holographic optical element, which allows focusing of sources on to different detectors.

This provides interconnections to various chips located on board as well as various CPU’s within a chip.
The interconnection can be established by lining the sources and detectors in the same plane.

The hologram reflector is suspended above this plane at small distance.
Non coherent light

- The use of non coherent light is possible by using a combination of prisms, lenses and masks.

- A perfect shuffle network can serve as an interconnection network.

- It is used due to high flexibility and simple routing schemes.
- The input are split into two halves and stretched so that each are equal to size of original signal.

- Then a shift is applied which allows interleaving of the inputs. This gives the output a shuffled sequence.
Interconnection using guided waves

- The free space interconnection requires precise alignment to ensure the beams of light strike at the right detector.

- The coupler and gates can also be used for module to module communication but becomes a problem when the number of chips increase.

- The use of an optical bus will be a simple solution.

- An optical bus can provide high data rate and the fan outs are greater when compared to that of electronic buses.
Polymer Waveguide technology.

1. Hybrid interconnects overcome the drawback of current technologies.
2. High density optical transceivers are required.
3. Polymer optical waveguides will act as the optical transport.
4. Polymer waveguides are compatible with fiber based optical links.
5. Bandwidth in excess of 40GH.m
6. Propagation loss – 0.05 db/cm.
7. 2X improvement in performance.
Future of Super computer interconnects.

1. Optical interconnects will be closer to the CPU.

Now: Board to Board connections through electro-optic links.
In Future:

1. Optical links will be closer to the CPU.
2. Bandwidth will not be limited by the number of pins.
3. Equalization is not required.
4. Packet switches capable of switching 2.5TB of data has been demonstrated.
Growth:

1. Yr. 2002
   NEC Earth Simulator.

No optics were used.
Year 2005:
IBM Federation switch for ASCI purple.

Copper Cables for shorter distance. <10m
Optical links for longer distance. (20 – 40m)
Year: 2008

Road Runner:
55 miles of optic links

*http://www.lanl.gov/roadrunner/*
Summary:

1. Optics are fast replacing copper cables.
2. Drive is to save power in communications.
3. Still at early stage, building blocks are being developed.
4. Use of optics have proven to increase the Bandwidth.

Optical interconnects for SC will be growing in volumes of 10X every four years if the cost for every GB transmitted is reduced by 3X every 4 years.

Future Is Bright...........

Questions??????
Reference:

1. Low Power and High density optical interconnects for Future super Computers.
2. Towards ExaFlop Servers and Super Computers: The Road Map for Low Power and High density optical interconnects.
5. Leveraging Optical interconnects in Future Super Computers and Servers