

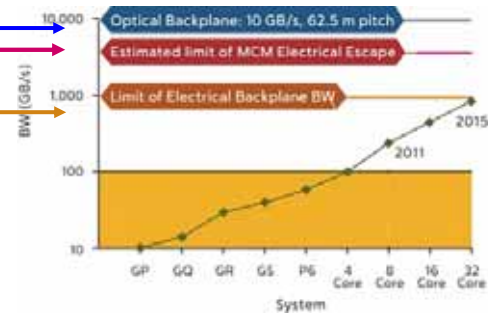
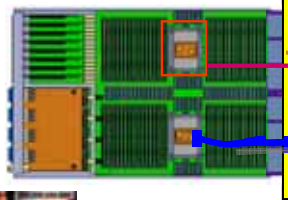
Optical PCB Overview

Jeff Kash, Dan Kuchta, Fuad Doany, Clint Schow,
Frank Libsch, Russell Budd, Yoichi Taira, Shigeru
Nakagawa, Bert Offrein, Marc Taubenblatt

November, 2009

Electrical BW Bottlenecks → Optics Opportunities

- **Electrical Buses become increasingly difficult at high data rates (physics):**
 - Increasing losses & cross-talk ; Frequency resonant affects
- **Optical data transmission:**
 - Power Efficiency , much less lossy, not plagued by resonant effects
- **Physical size of electrical connections (BGA, connector) limits number of connections**
 - Optical density ~10X higher



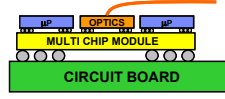
Rack



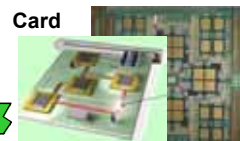
Backplane



Module



Card



Evolution of Rack-to-Rack Optics in Supercomputers

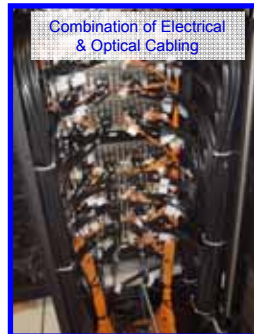
→ VCSEL-based Optics has displaced electrical cables today

2002



NEC Earth Simulator
• no optics

2005



Combination of Electrical & Optical Cabling

IBM Federation Switch for ASCI Purple (LLNL)
- Copper for short-distance links (≤ 10 m)
- Optical for longer links (20-40m)
~3000 parallel links 12+12@2Gb/s/channel

2008: 1PFLOP/sec

IBM Roadrunner (LLNL) Cray Jaguar(ORNL)



*<http://www.lanl.gov/roadrunner/>

- Introduced in 2008
- Still #1 as of June, 2009
- 4X DDR Infiniband (5Gb/s)
- 55 miles of Active Optical Cables



*<http://www.nccs.gov/jaguar/>

- #2 as of June, 2009
- Infiniband
- 3 miles of Optical Cables, longest = 60m

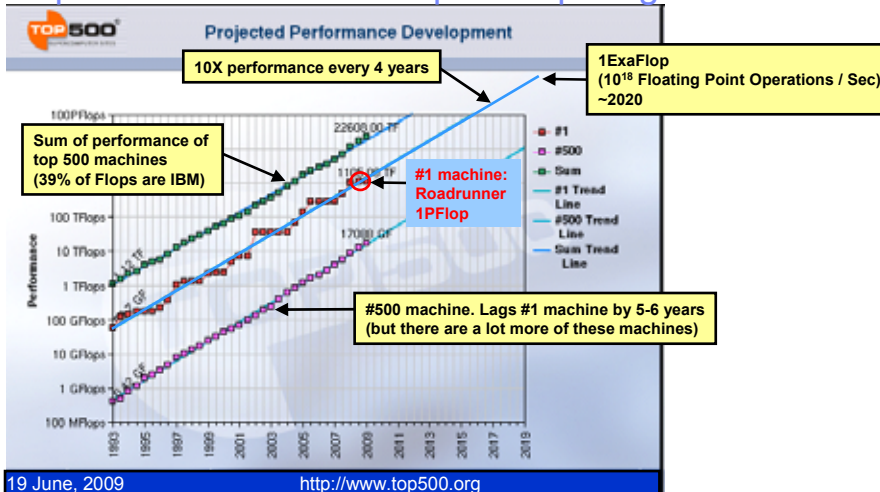


3

November 2009

© 2009 IBM Corporation

Exponential Growth in Supercomputing Power



- BW requirements must scale with System Performance, ~1B/FLOP (memory & network)
- **Requires exponential increases in communication bandwidth at all levels of the system** → Inter-rack, backplane, card, chip

4

November 2009

© 2009 IBM Corporation

The Road to Exascale:

Cost and power of a supercomputer

Year	Peak Performance	Machine Cost	Total Power Consumption
2008	1PF	\$150M	2.5MW
2012	10PF	\$225M	5MW
2016	100PF	\$340M	10MW
2020	1000PF (1EF)	\$500M	20MW

- **Assumptions: Based on typical industry trends –**
(See, e.g., top500.org and green500.org)
 - 10X performance / 4yrs (from top500 chart)
 - 10X performance costs 1.5X more
 - 10X performance consumes 2X more power

5

November 2009

© 2009 IBM Corporation

The Road to Exascale:

Total bandwidth, cost and power for optics in a machine

Year	Peak Performance	(Bidi) Optical Bandwidth	Optics Power Consumption	Optics Cost
2008	1PF	0.012PB/s (1.2x10 ⁵ Gb/s)	0.012MW	\$2.4M
2012	10PF	1PB/s (10 ⁷ Gb/s)	0.5MW	\$22M
2016	100PF	20PB/sec (2x10 ⁸ Gb/s)	2MW	\$68M
2020	1000PF (1EF)	400PB/sec (4x10 ⁹ Gb/s)	8MW	\$200M

- **Require >0.2Byte/FLOP I/O bandwidth, >0.2Byte/FLOP memory bandwidth**
 - 2008 optics replaces electrical cables (0.012Byte/FLOP, 40mW/Gb/s)
 - 2012 optics replaces electrical backplane (0.1Byte/FLOP, 10% of power/cost)
 - 2016 optics replaces electrical PCB (0.2Byte/FLOP, 20% of power/cost)
 - 2020 optics on-chip (or to memory) (0.4Byte/FLOP, 40% of power/cost)

6

November 2009

© 2009 IBM Corporation

In HPC space, increased need for and use of optics → cost and power must decrease (per bit unidirectional is shown)

Year	Peak Performance	number of optical channels	Optics Power Consumption	Optics Cost
2008	1PF	48,000 (@5Gb/s)	50mW/Gb/s (50pJ/bit)	\$10/Gb/s
2012	10PF	2x10 ⁶ (@10Gb/s)	25mW/Gb/s	\$1.1/Gb/s
2016	100PF	4x10 ⁷ (@10Gb/s)	5mW/Gb/s	\$0.17/Gb/s
2020	1000PF (1EF)	8x10 ⁸ (@10Gb/s)	1mW/Gb/s	\$0.025/Gb/s

Industry trend derived roadmap, not IBM product plans

- Table is based on historical trends for HPCs
- To get optics to millions of units in HPC, need ~\$1/Gb/s unidirectional
 - **Cost targets continue to decrease with time below that**
- Power is OK for 2012, then sharp reductions will be needed

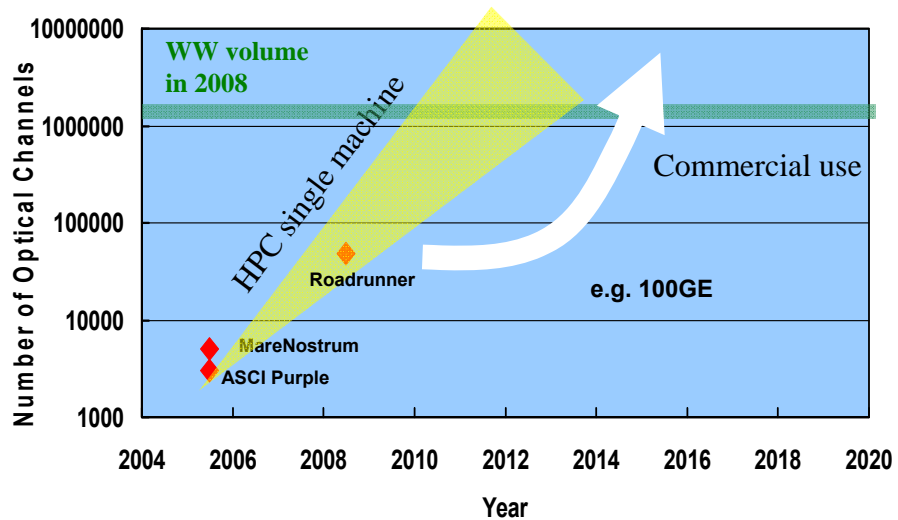
7

November 2009

© 2009 IBM Corporation

HPC driving volume optics → Higher volumes → lower Cost

A Single machine in the next few years could be similar to today's WW parallel optics



8

November 2009

© 2009 IBM Corporation

Optical Printed Circuit Boards and Components: Enabling mass manufacturing

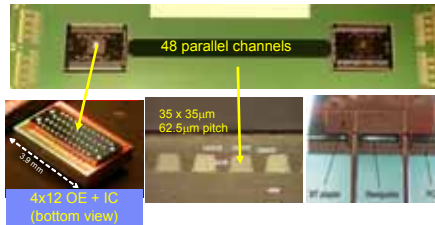
Electronics: Wires and discretes ...



Optics: Fibers and modules...



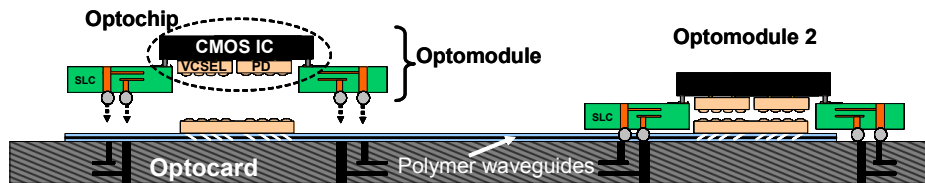
... to Printed Circuit Boards with electrical components



... to integrated waveguides on PCBs with optical components

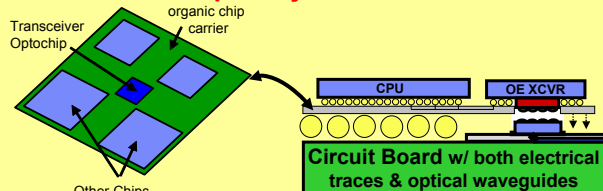
Optical waveguide interconnects:

The Terabus project and related work



Dense Hybrid Integration: demonstrate a low-cost packaging approach compatible with conventional PCB manufacturing and surface-mount board assembly

Future Vision: optically-enabled MCM's



- Low-density, conventional electrical interface for power & control
- High-density, wide and fast optical interfaces for data I/O
- Much higher off-module bandwidth at low cost in \$\$ and power

Circa 2014-2016

Power	<10mw/Gbps (EOE)
Cost	<\$0.25/Gbps (TRx + on-card optics)
Datarate	20Gbps/channel
Density	2 Tbps/cm ² (on module)
Reliability	< 10 FIT per channel

Full Terabus Link (985-nm): 2 Transceiver Optomodules on Optocard

**TRX1:
16TX + 16RX**

4x4 VCSEL Array **4x4 PD Array**

16 Channels TRX1 → TRX2 at 10Gb/s + 16 Channels TRX1 ← TRX2 at 10Gb/s

11 November 2009 © 2009 IBM Corporation

Optical interconnect density exceeds electrical:

Another reason for optics –
Electrical Packaging is Running out of Pins – Optics Can Help

MultiChip Module 1 **ICs** **MultiChip Module 2** **ICs**

All-electrical circuit board

For each 10Gb/s signal you need:

- 3 or 4 electrical pins (differential pair) on 800um pitch
- OR
- 1 OE array element (250um pitch)

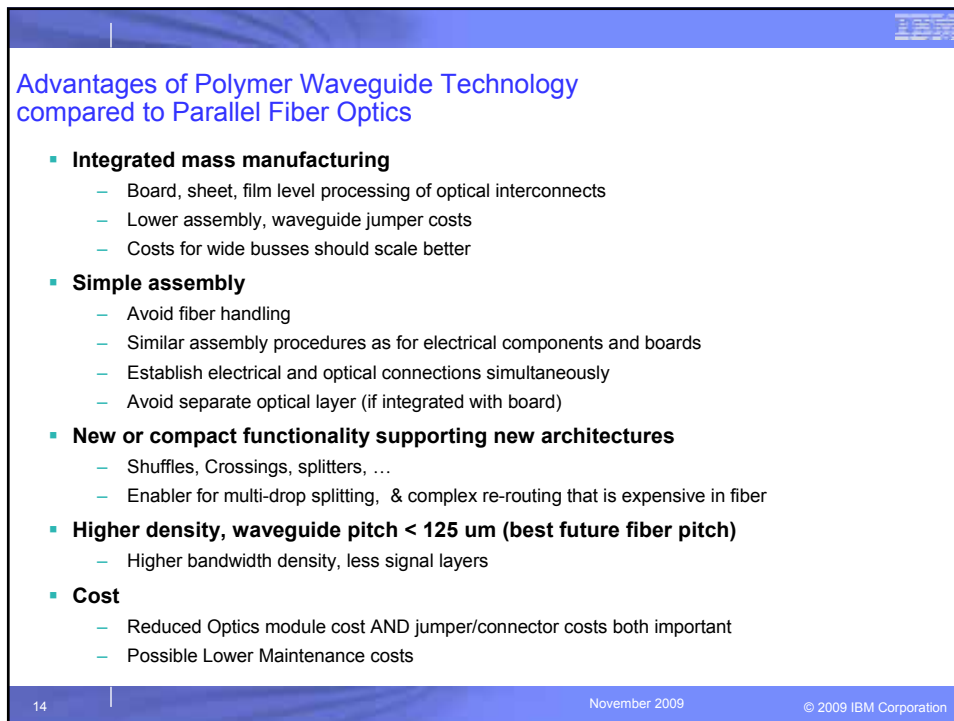
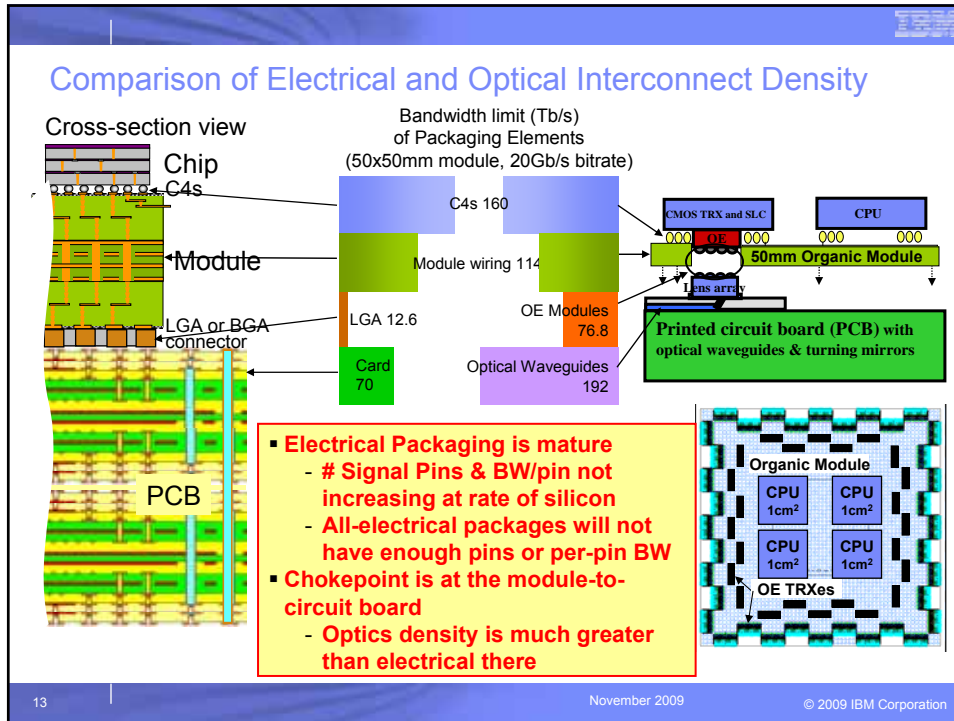
→ Optics density exceeds electrical by more than 10X

MultiChip Module 1 **Optochip₁** **Multimode Polymer Waveguides** **Optochip₂** **MultiChip Module 2**

Circuit Board with multimode waveguides

Bottom
BGA
Equivalent bandwidths
Cutout with OEs-on-IC

12 November 2009 © 2009 IBM Corporation

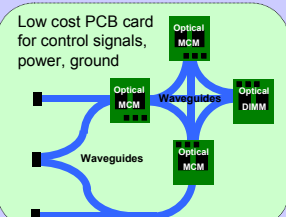
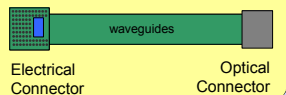


Optical PCB Roadmap

Initial FOCUS

2012 - 2014

Discrete Active Optical Flex assemblies are easily replaced, less disruptive



(All off-MCM links are optical)

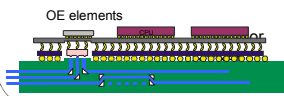
2014 - 2016

Optical waveguides/ modules on PCB (could be flex laminate) once technology matures to sufficient reliability



2018

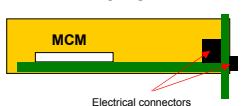
Optical Waveguides embedded in card w/optical vias are a complete replacement for all high speed copper lines



Use of active waveguide flex and fiber cable in systems

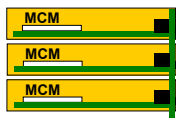
TODAY

Drawer



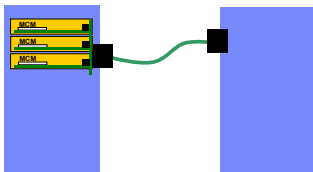
Electrical from MCM to back of drawer, electrical to backplane and to electrical connector for active optical cable

Backplane



Drawers attach to electrical backplane (PCB)

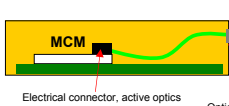
Rack to Rack



Active Optical fiber based cable: Rugged (may be routed under floor), 3-100m lengths

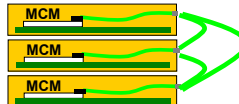
FUTURE with OPCB

Drawer



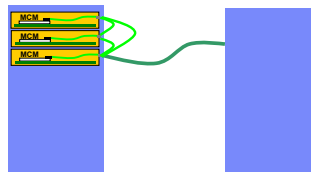
Active waveguide flex from MCM to back of drawer, may include channel shuffles

Backplane



Optical cable based optical backplane, may be passive waveguide flex cables or fiber cables (initially more likely). May include channel shuffles, typically < 1m-2m

Rack to Rack



Passive Optical fiber based cable: Rugged (may be routed under floor), 3-100m lengths

Optical Printed Circuit Boards

- IBM Research has invested heavily in the past 5 years in Optical printed circuit board technology based on multi-mode polymer waveguides
 - Partially funded by the US Government (Terabus program)
- We believe this technology will be needed to provide the needed BW for future server generations, allow highly integrated electrical-optical links and provide a path to much lower cost optical links.
- We are interested in establishing a market eco-system that will provide components, standards and specifications for this technology.