UNIVERSITY OF CAMBRIDGE

Optical Interconnects for Backplane and Chip-to-chip Photonics

I H White* and R V Penty

*** van Eck Professor of Engineering**

University of Cambridge, Electrical Engineering Division,
9 JJ Thomson Avenue, Cambridge CB3 0FA, United Kingdom

Acknowledgements:
J Beals, N Bamiedakis, University of Cambridge
Dr D Cunningham, Avago Technologies
Dr T Clapp and Dr J De Groot, Dow Corning
UK EPSRC

Outline

1 Introduction to Datacommunications

2 Background – the LAN/Server Networks

- GbE and 10 GbE systems
- The importance of MultiMode optical Fibre (MMF)

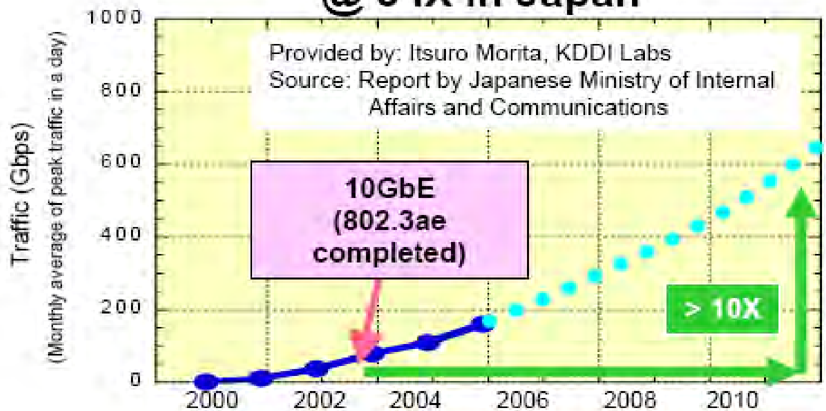
3 The Need for Optical Interconnects

- Cluster Computing, Chip to Chip and on-Chip
- PCB Optical Circuits

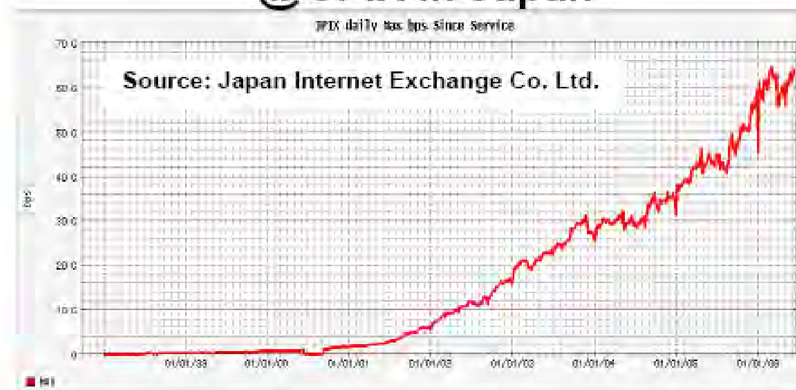
4 Conclusions

The Challenge is Bandwidth –
Traffic patterns at major Internet exchanges

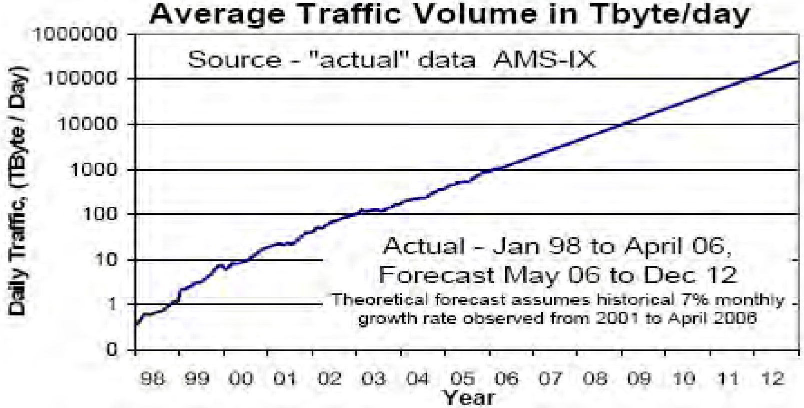
Monthly Average Peak Traffic @ 3 IX in Japan



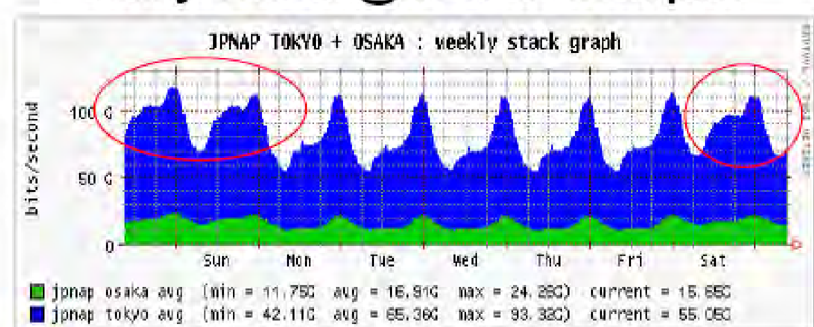
Daily Maximum Gbps @ JPIX in Japan



AMS-IX in Amsterdam
Average Traffic Volume in Tbyte/day



Daily Traffic @ JPNAP in Japan



Average traffic growth rate since 2003 6% per month

Used with permission from Internet Multifeed Co.

Source: J. Cain, Cisco Systems, July 2006

UNIVERSITY OF CAMBRIDGE

Trends in Optics … and Bandwidth

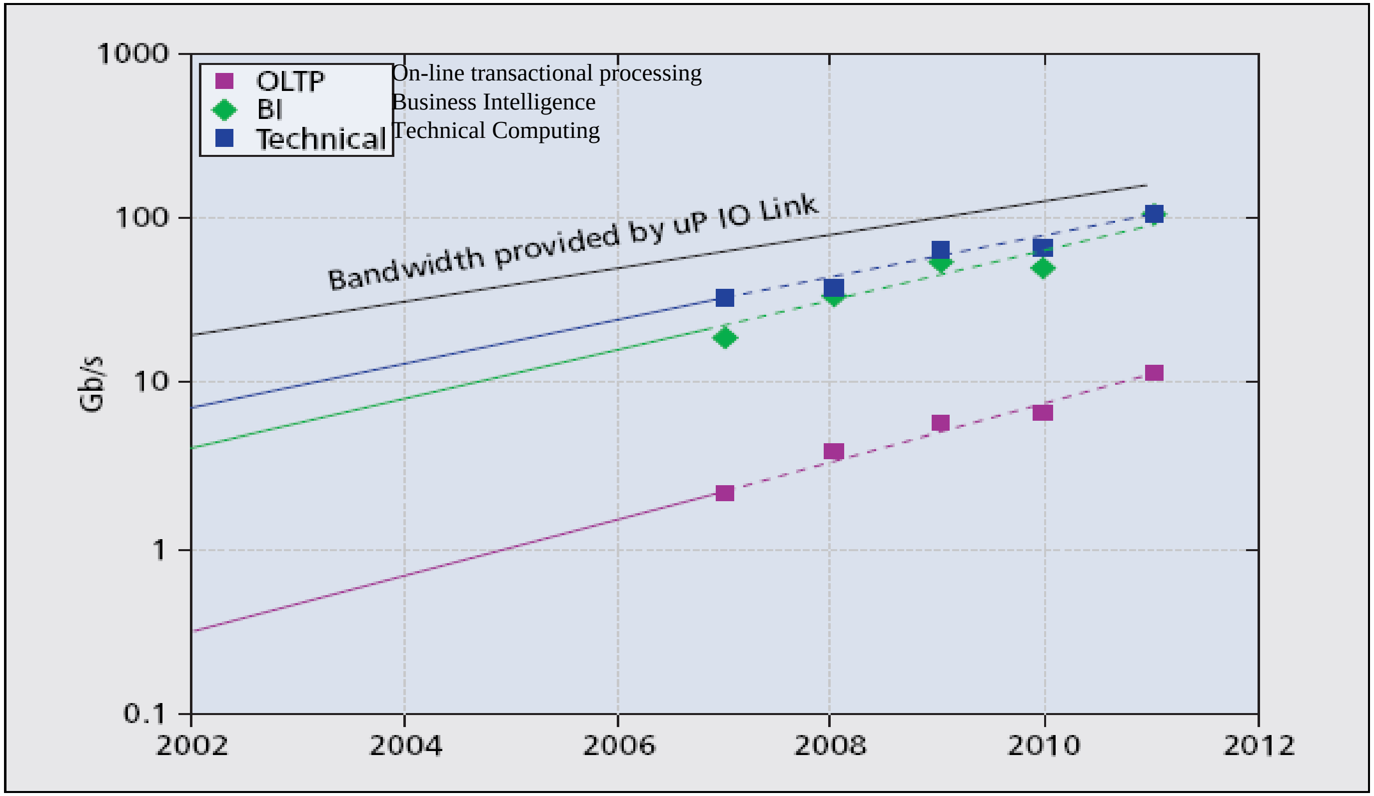


Figure 9. *Technical and Business Intelligence computing will require 100 Gb/s rates in the near future.*

A.F. Benner, P. Pepeljugoski, R. Recio, *IEEE Apps and Practice* (2007)

Relentless increase in bandwidth requirements across computing applications …

Trends in Optics …

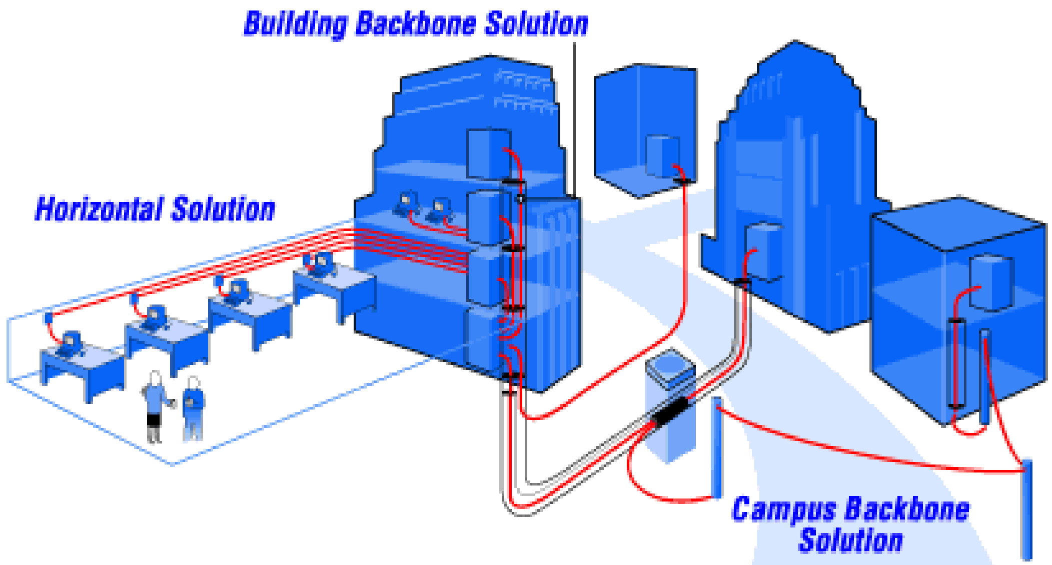
	MAN/WAN	Cables–long	Cables–short	Card-to-card	Intra-card	Intra-module	Intra-chip
Length	Multi-km	10–300 m	1–10 m	0.3–1 m	0.1–0.3 m	5–100 mm	0–20 mm
No. of lines per link	One	One to tens	One to tens	One to hundreds	One to hundreds	One to hundreds	One to hundreds
No. of lines per system	Tens	Tens to thousands	Tens to thousands	Tens to thousands	Thousands	Approximately ten thousand	Hundreds of thousands
Standards	Internet Protocol, SONET, ATM	LAN/SAN (Ethernet, InfiniBand, Fibre Channel)	Design-specific, LAN/SAN (Ethernet, InfiniBand)	Design-specific and standards (PCI, backplane InfiniBand and Ethernet)	Design-specific, generally	Design-specific	Design-specific
Use of optics	Since the 1980s	Since the 1990s	Present time, or very soon	2005–2010 with effort	2010–2015	Probably after 2015	Later

Optical links becoming

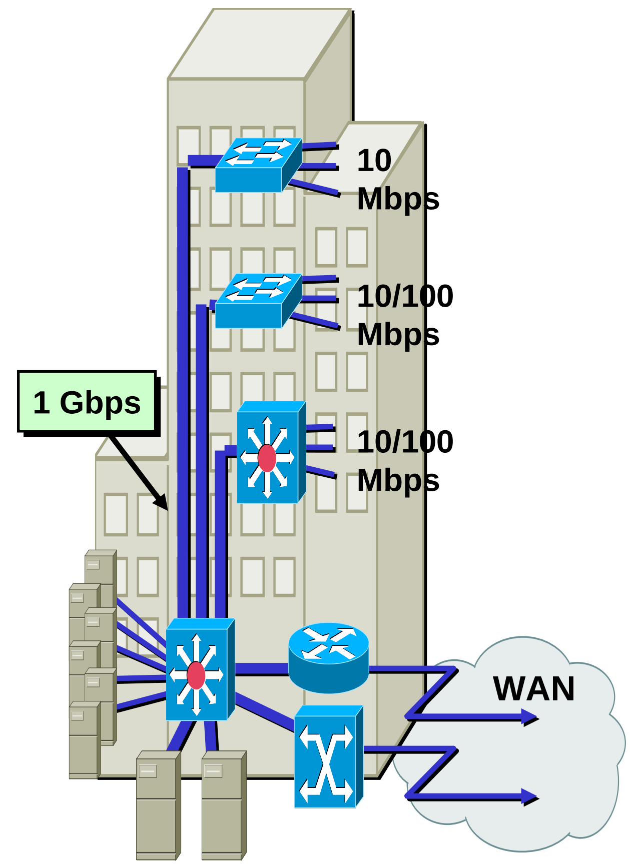
- shorter
- denser
- higher bandwidth
- application specific
- cheaper!

A.F. Benner et al. *IBM J. Res. & Dev*. 49 (2005)

2 Datacommunication Scenarios



Building Backbone "Gbps between Floors and in the Building Data Center"



Hierarchies of Datacommunication Links

1. Horizontal cabling from telecommunications closest to workstations (100 m)
2. Intra-building (inside) backbone from telecom closet to equipment room (500 m)
3. Combined campus and building backbone (2000 m)

Recent developments in 10 Gigabit Ethernet

Phase 1: 1999-2002 Fibre port types required for the early market

Name	Description	Fibre Type	Reach
10GBASE-LR	1310 nm serial LAN PHY	SMF	10 km
10GBASE-ER	1550 nm serial LAN PHY	SMF	30 or 40 km
10GBASE-SR	850 nm serial LAN PHY	MMF (OM3)	300 m
10GBASE-LX4	1310 nm WDM LAN PHY	OM1, OM2 & OM3 MMF	300 m

Phase 2: 2002-2006 Copper port types required for the mature market

Name	Description	Media Type	Reach
10GBASE-CX4	Copper Serial LAN PHY	Cable	15 m
10GBASE-T	Twisted Pair Serial LAN	Cat 6 or better cable	100 m

Phase 3: 2003-2006 Fibre port type required for the mature market

Name	Description	Media Type	Reach
10GBASE-LRM	1310 Serial LAN PHY Multimode Fiber	OM1, OM2 & OM3 MMF	220 m

Transceivers for Datacommunications

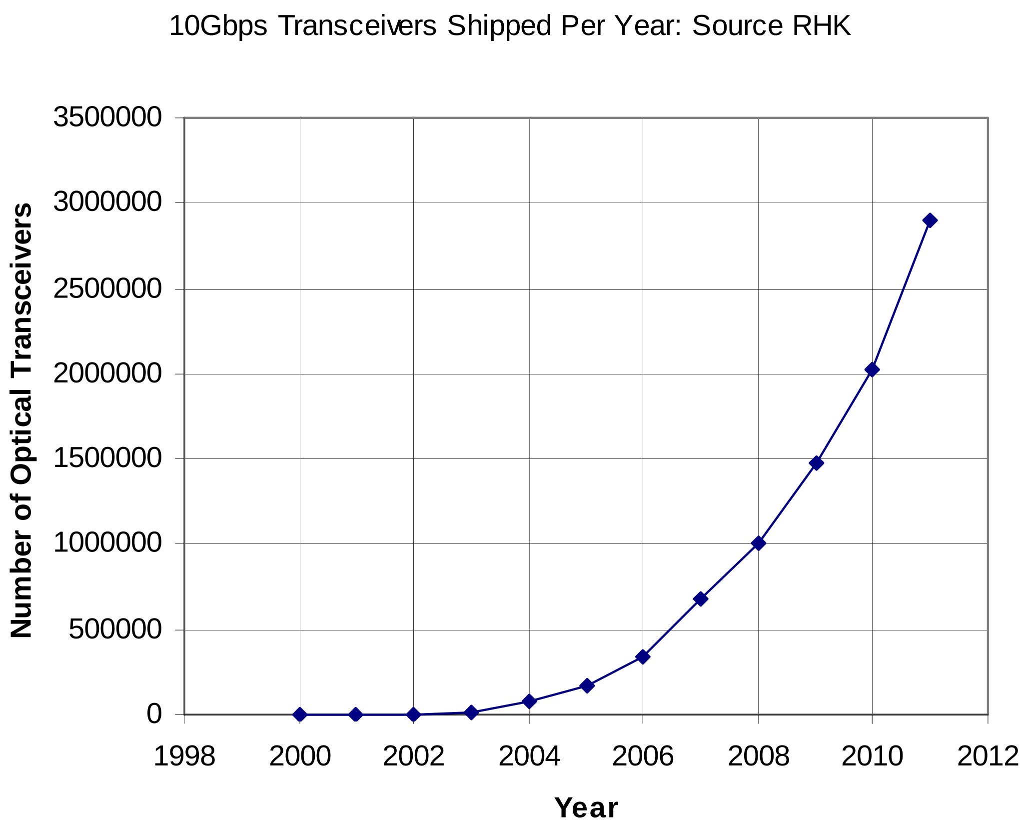
Components:

- Smaller Size (Discretes/Optics/ICs)
- Higher Speed (100 Mb/s to 1+ Gb/s)
- 3.3 V Operation
- Surface Mount Packages
- Shielded for EMI Compliance

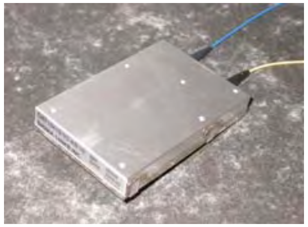
Systems:

- Higher Density Component Loading
- High Bandwidth Capability (Terabit)
- Lower Power Requirement
- Lower per port solution cost $
- Larger Chassis Designs

10 Gb/s optical transceiver market



300 pin

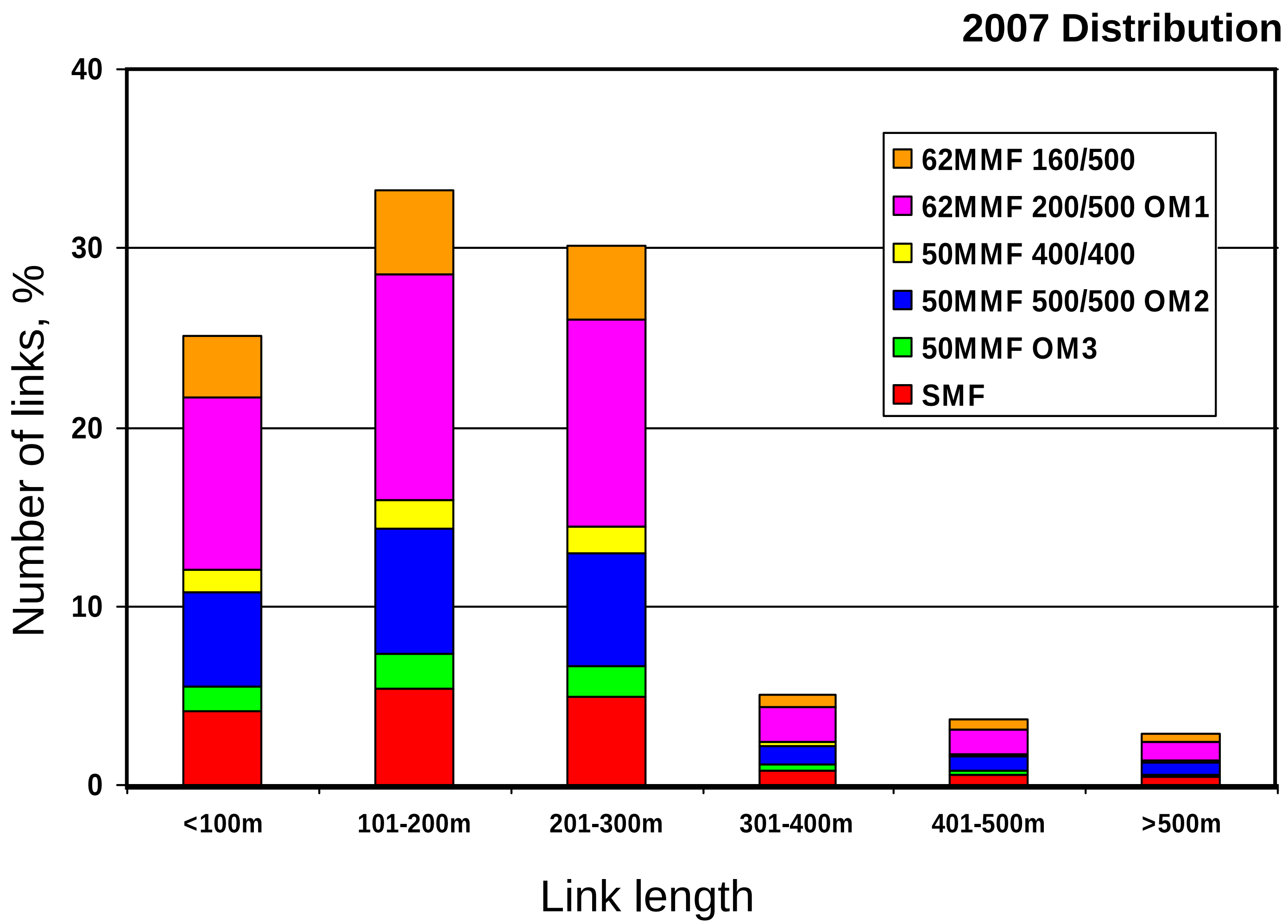


XENPAK

X2

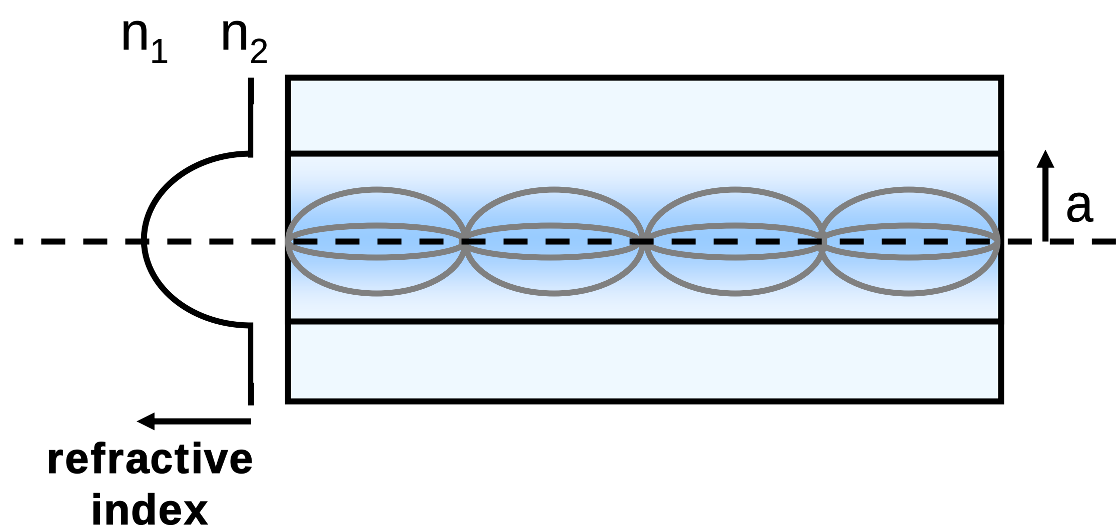
XFP

Installed link length distribution



Graph based on: In-Premises Optical Fibre Installed Base Analysis to 2007, Alan Flatman,
http://grouper.ieee.org/groups/802/3/10GMMFSG/public/mar04/flatman_1_0304.pdf

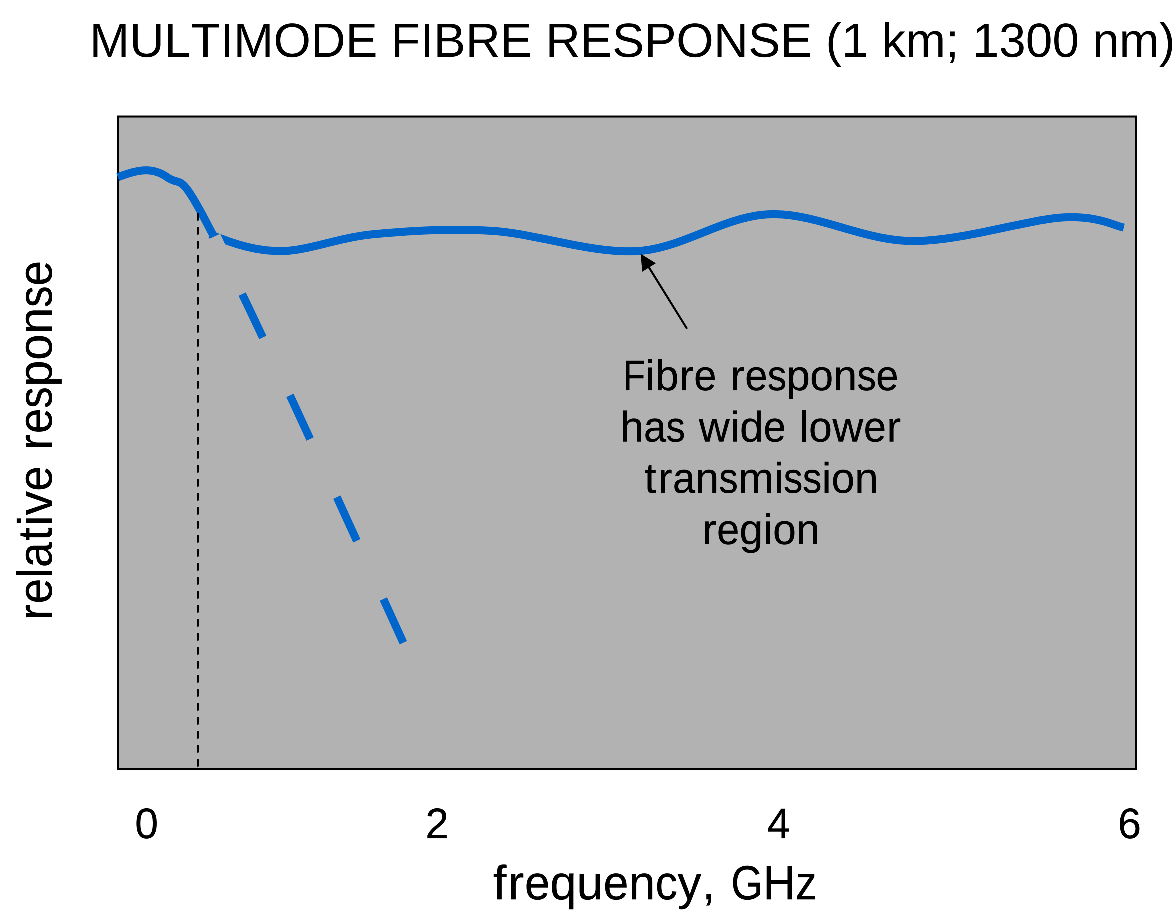
Why is Graded Index MMF Challenging?



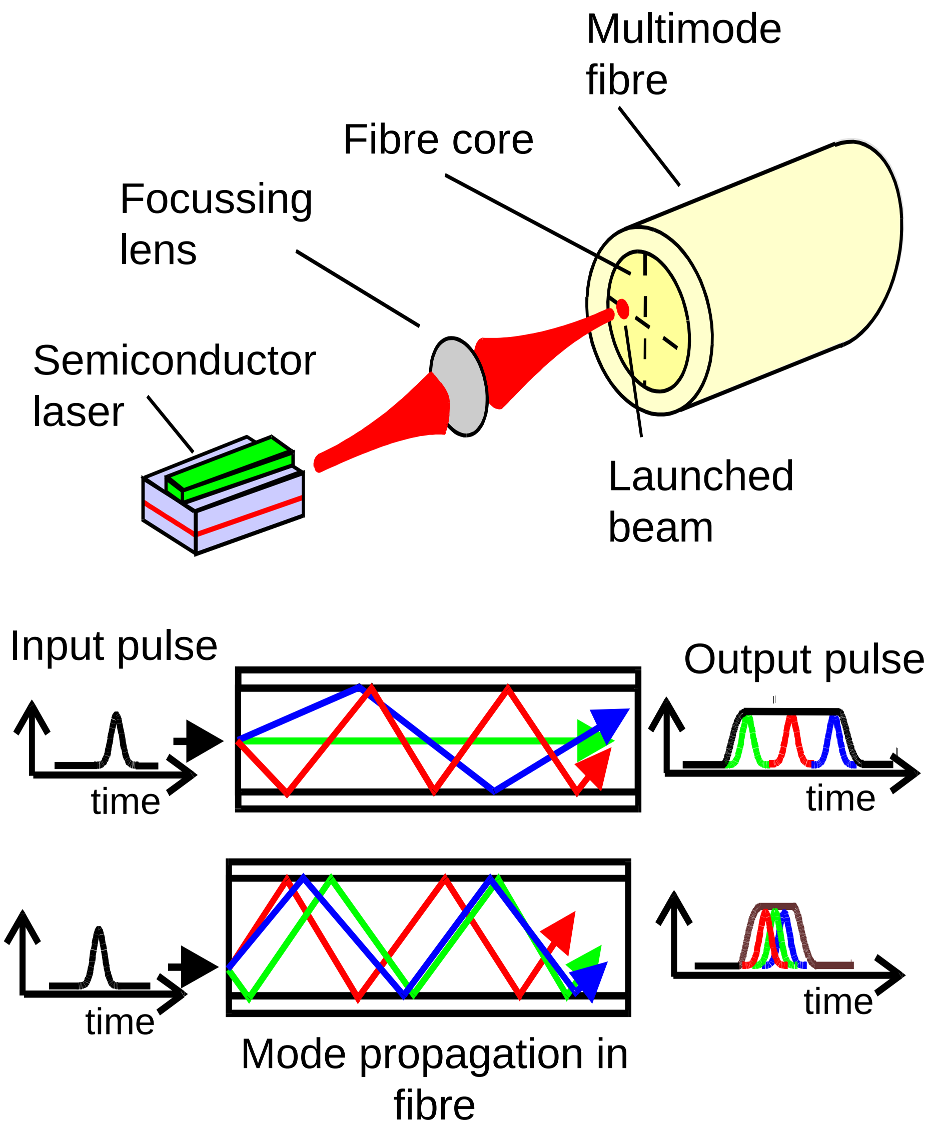
MMF Bandwidth Specifications

	62.5 µm MMF	50 µm MMF
850 nm	160 MHz.km	400 MHz.km
1300 nm	500 MHz.km	500 MHz.km

Techniques for enhancing the bandwidth of MMF links



Offset Launch for Ethernet Links



Offset launch has been standardised within IEEE 802.3 Gigabit Ethernet

Used with 1000BASE-LX GbE transceivers

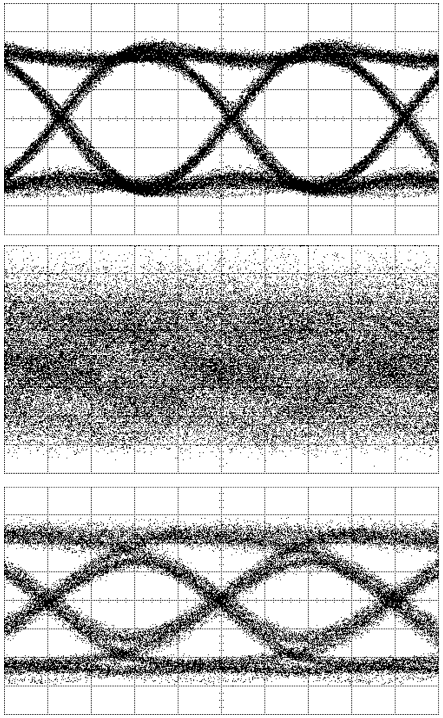
Offset launch patchcord implementation - Example

2.5 Gb/s over 3 km of standard MMF

Back-to-back

Standard launch

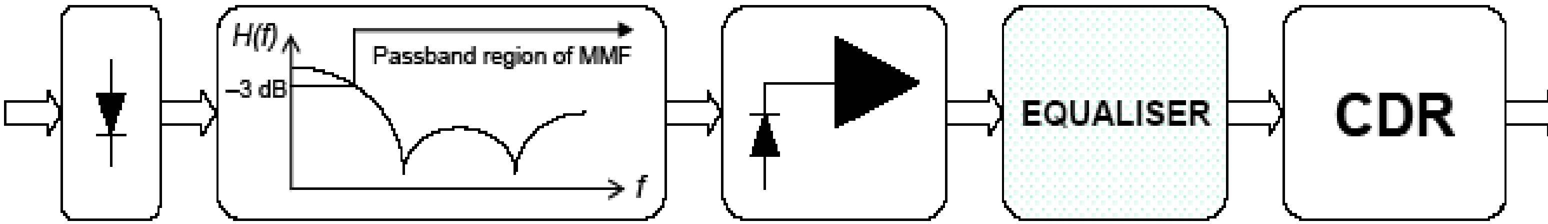
Offset launch



Link contains 7 connectors / 3 splices - *offset launch is robust in presence of multiple connectors and patch panels*

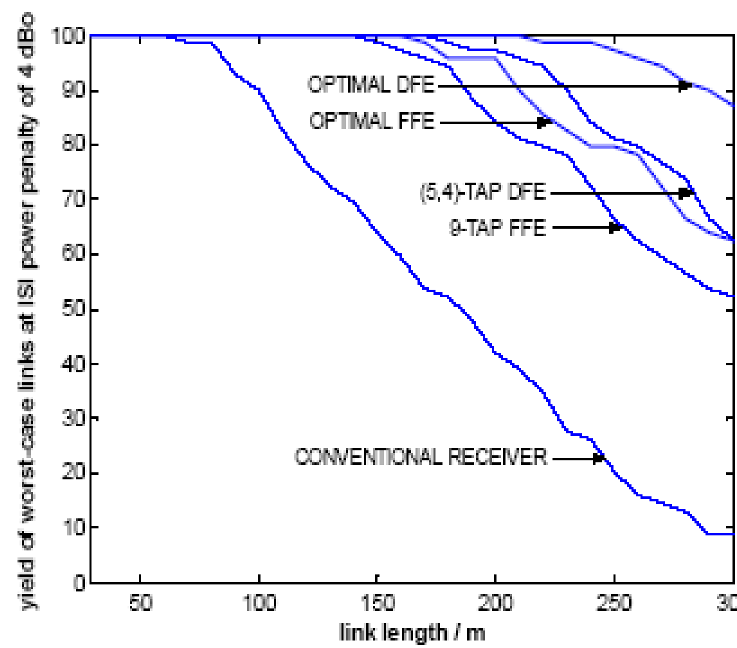
Mode-conditioning patchcord (MCP)

Multimode Fibre Transmission: Electronic Compensation

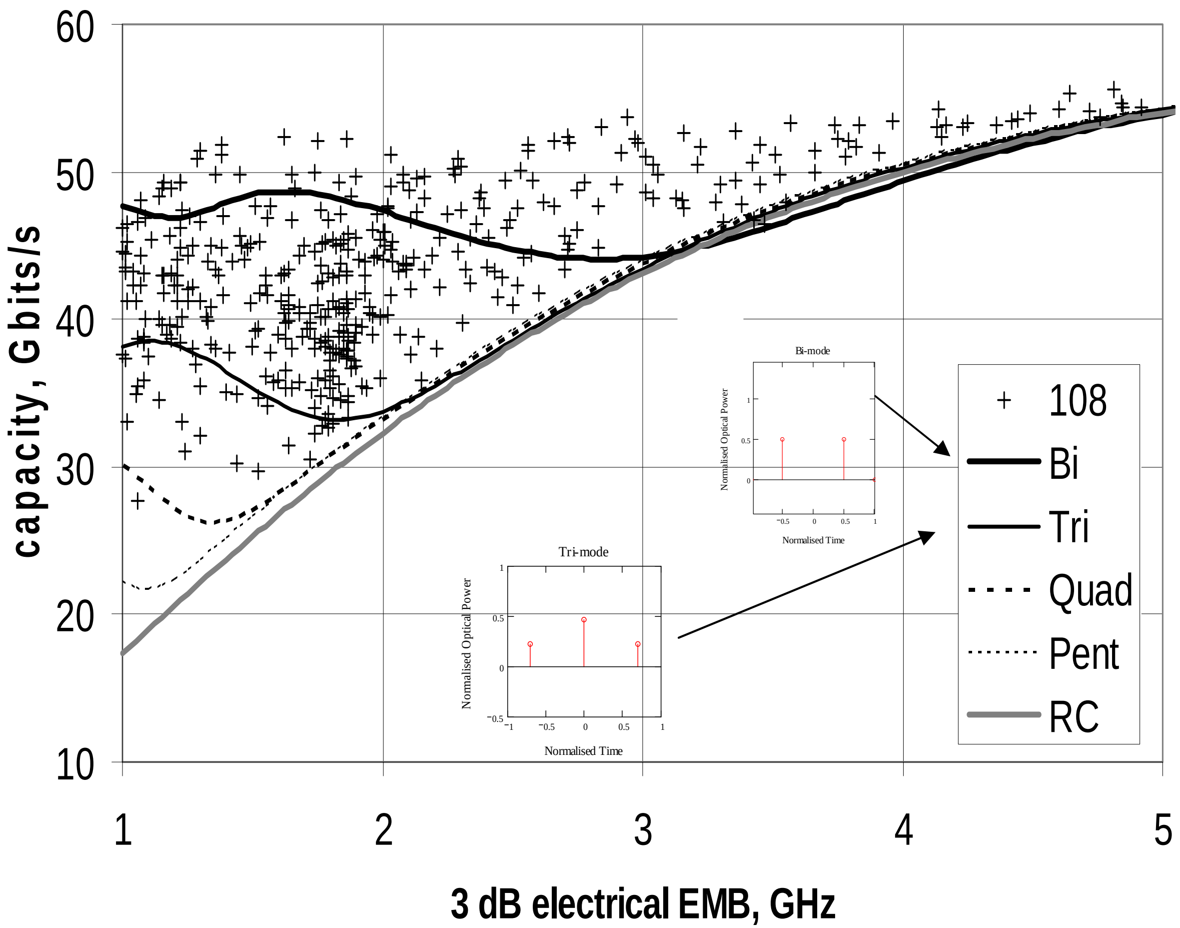


Signal impairment due to fibre properties may be compensated after the receiver, using emerging electronic signal processing techniques

Instrumental in emerging 10 GbE standards in MMF



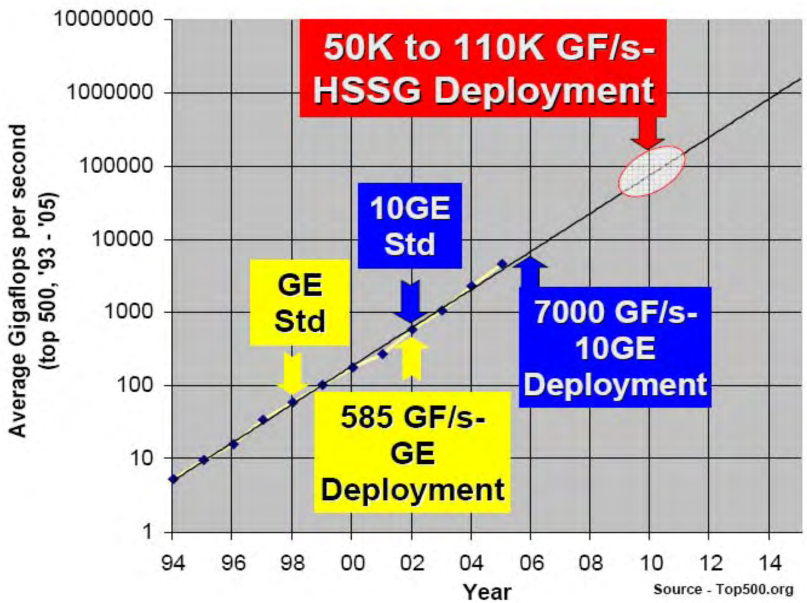
How far can we push MMF?



For the first time:

- Calculated the capacity of MMF
- Derived an analytical worst case model
- Further 7x speed enhancement possible over 10GbE using single laser

Need for 100 Gb/s – High performance computing



Historically:

12X increase in average GF/s needs

10X increase in Ethernet interconnect

What routes for higher speeds – Go parallel (with help from serial)

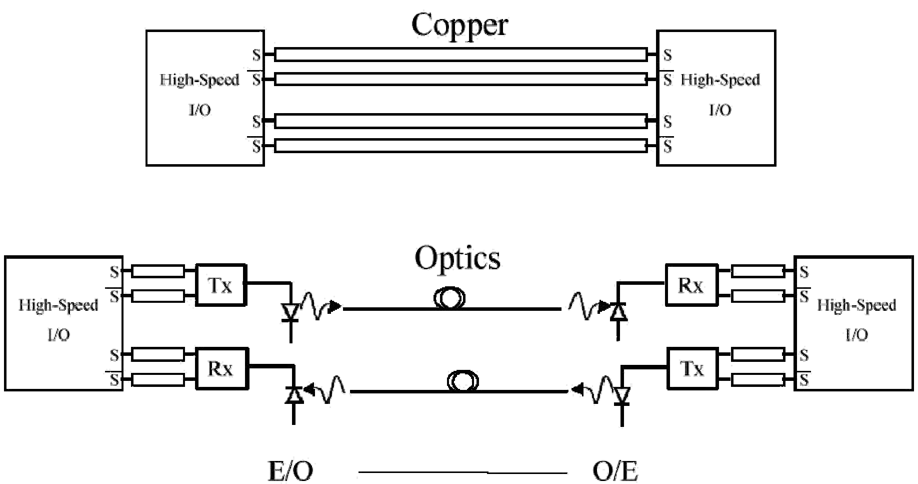
- Parallel Fibre

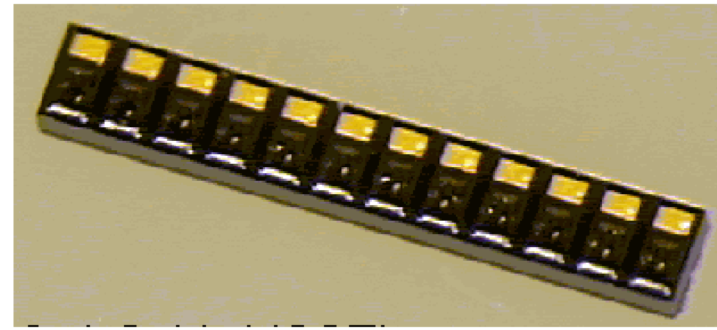
(as long as we have integration)

- Wavelength Division Multiplexing

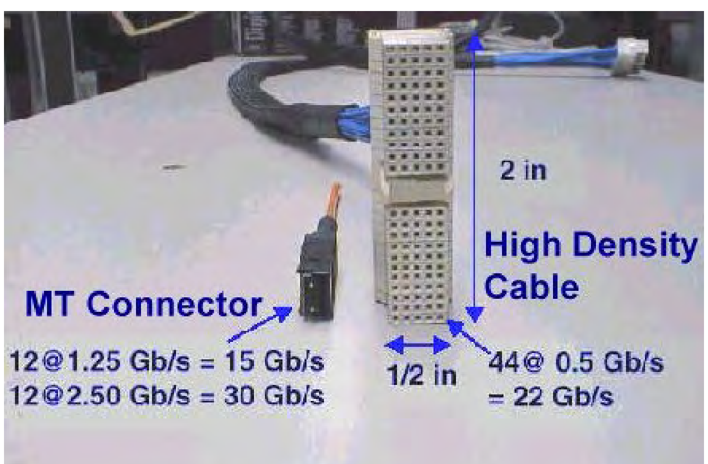
(as long as we have integration)

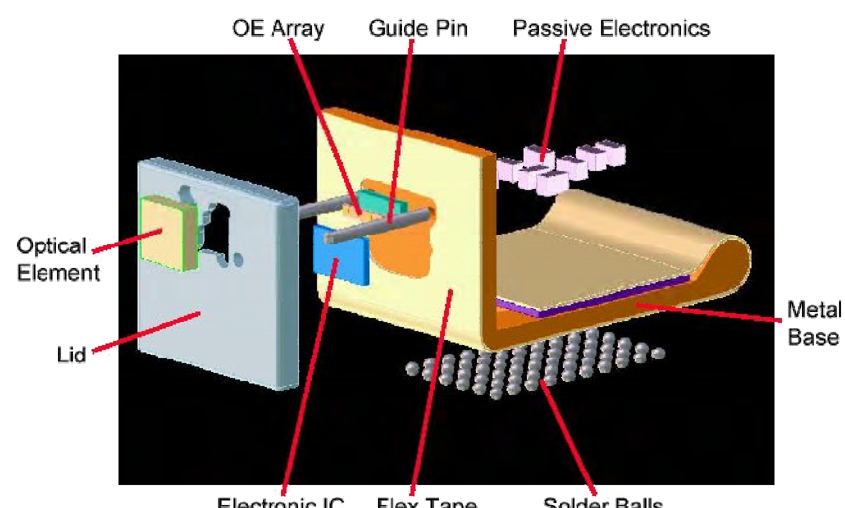
Parallel Optics – Always Watch Copper!



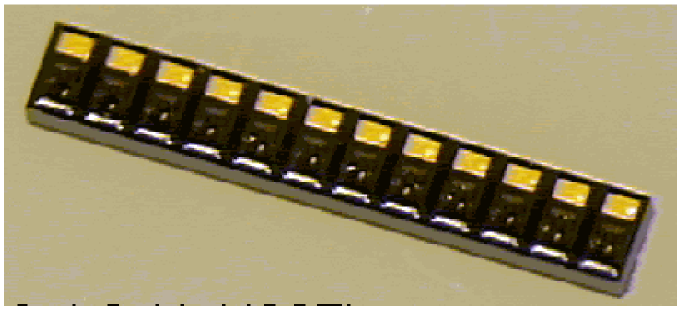


850nm VCSEL 1X12 Array



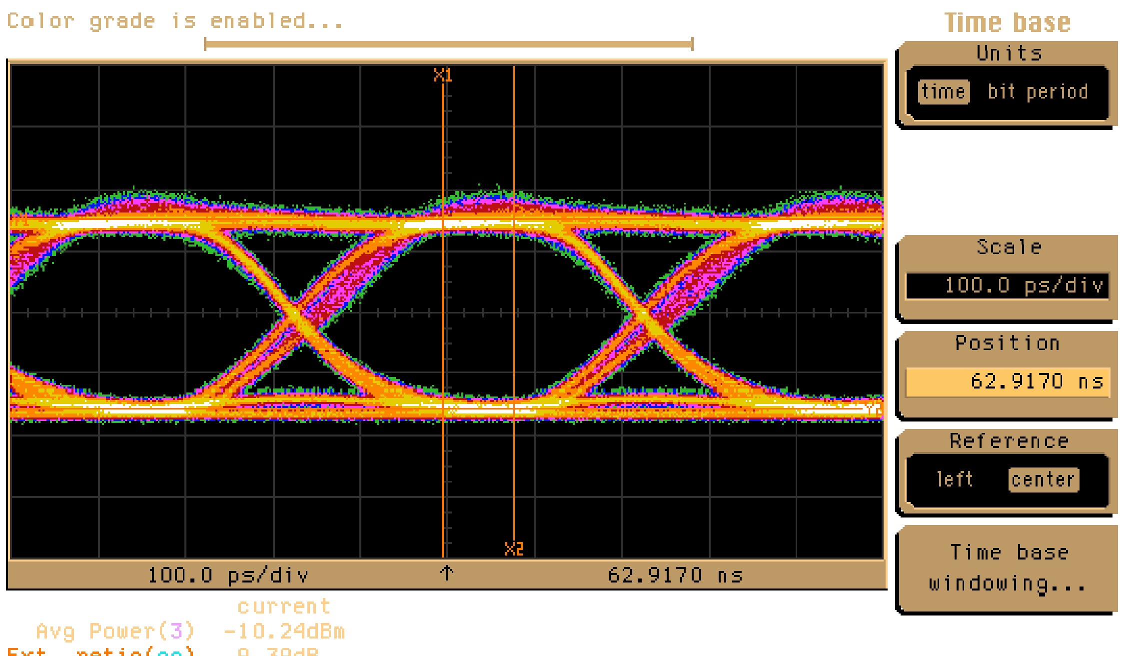


2.5 Gb/s/ch 850 nm VCSEL Array



2.5 Gb/s per channel

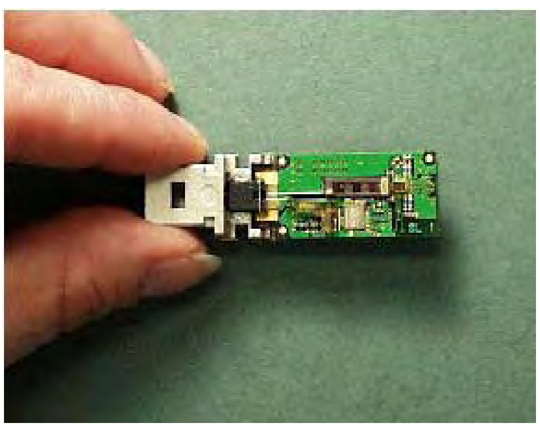
(30Gb/s per array)

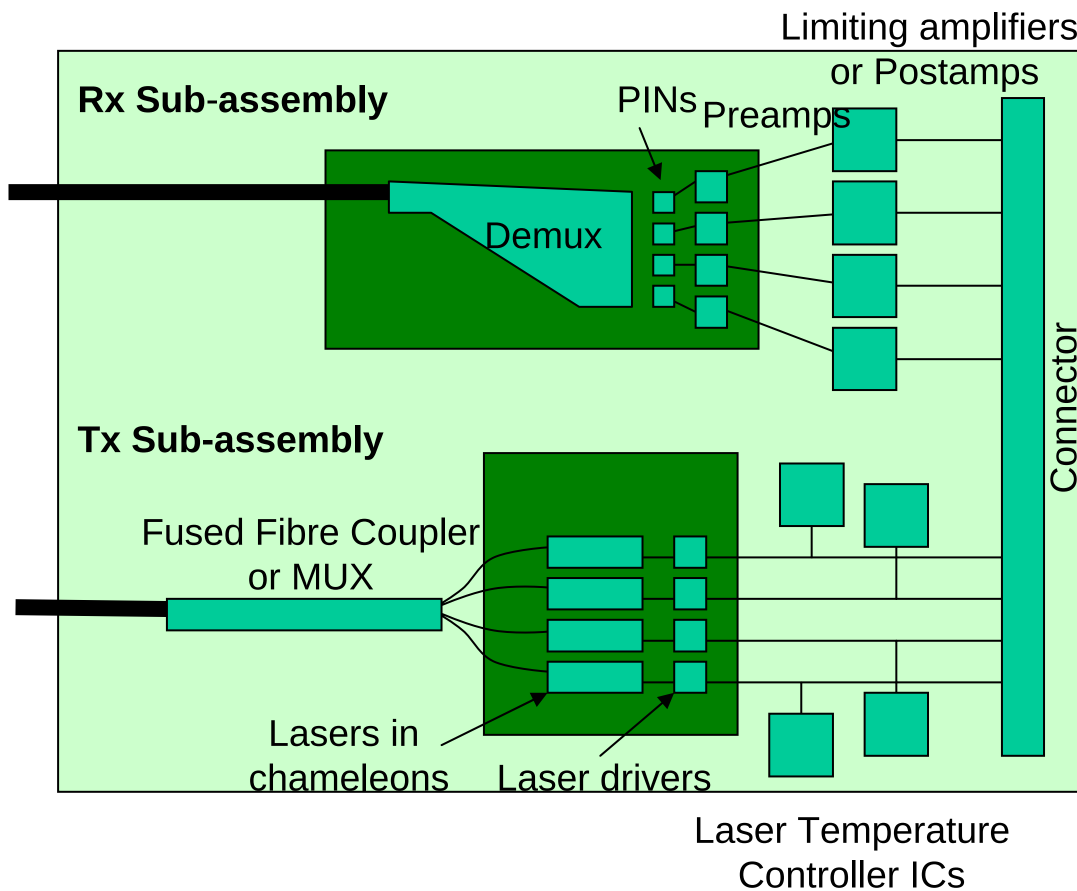


Power ~0dBm, Ext. Ratio=9dB

Low Cost Wavelength Division Multiplexed Systems

- 4 wavelengths
- Low cost
- Potential future 100 Gb/s capacity



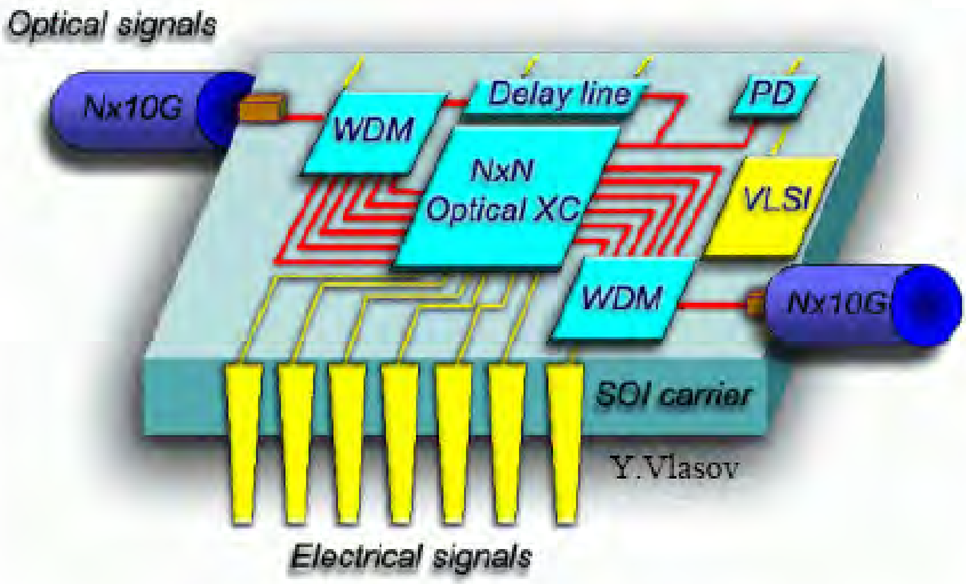


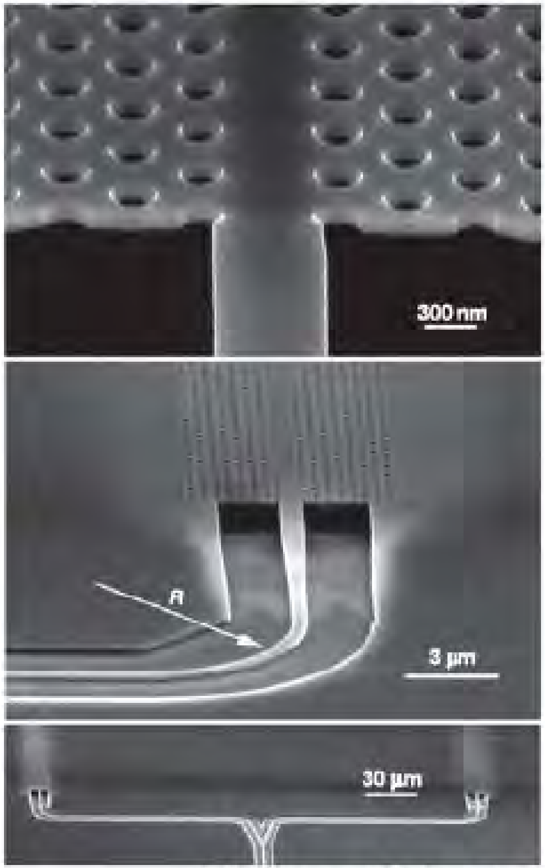
Agilent Technologies

Source:

LA Buckman et al., IEEE PTL, Vol.14, pp 702-704, 2002

Chip Level Integrated Photonics





Y.Vlasov, Nature 11/05

- Low cost IC paradigm
  - Fab costs
  - Yield
  - Volumes

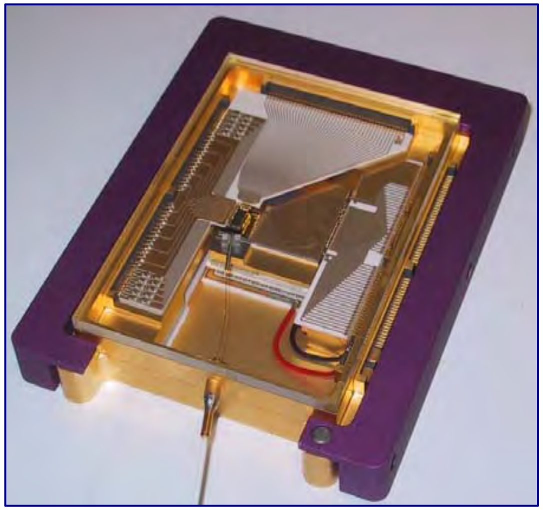
- Higher levels of functionality
- New exploitable phenomena

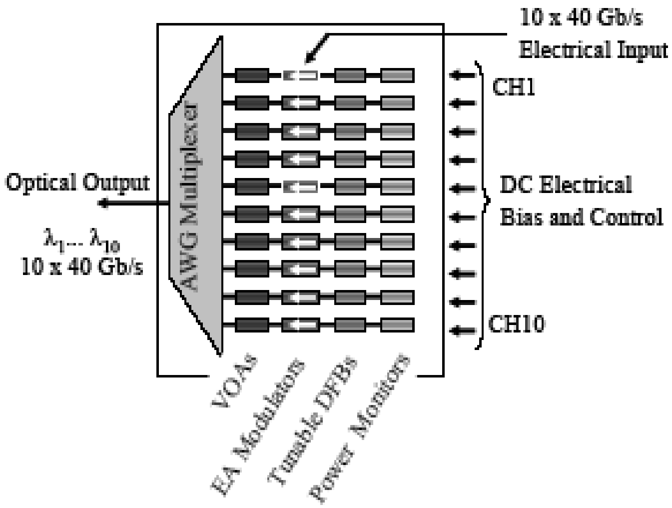
III-V Integrated PIC Transmitters for 400Gb/s

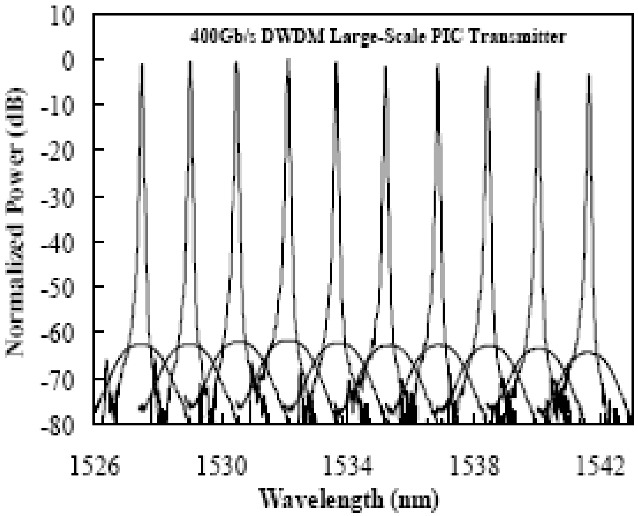
High performance components using advanced integration concepts

“400 Gb/s (10-channel x 40 Gb/s) DWDM Photonic Integrated Circuits”, Infinera, OFC 2005

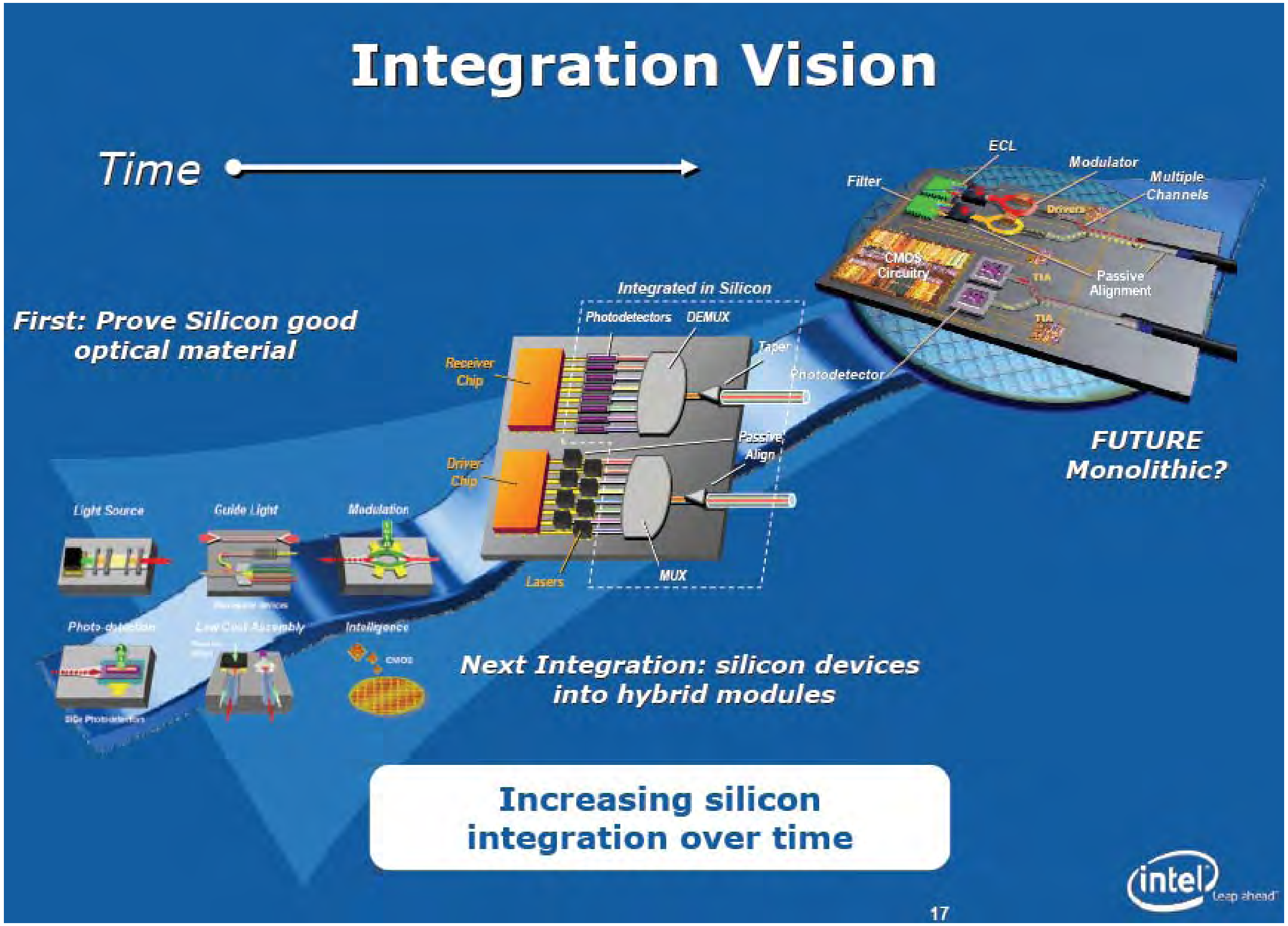
New generations of ultra-high speed integrated WDM transmitters emerging



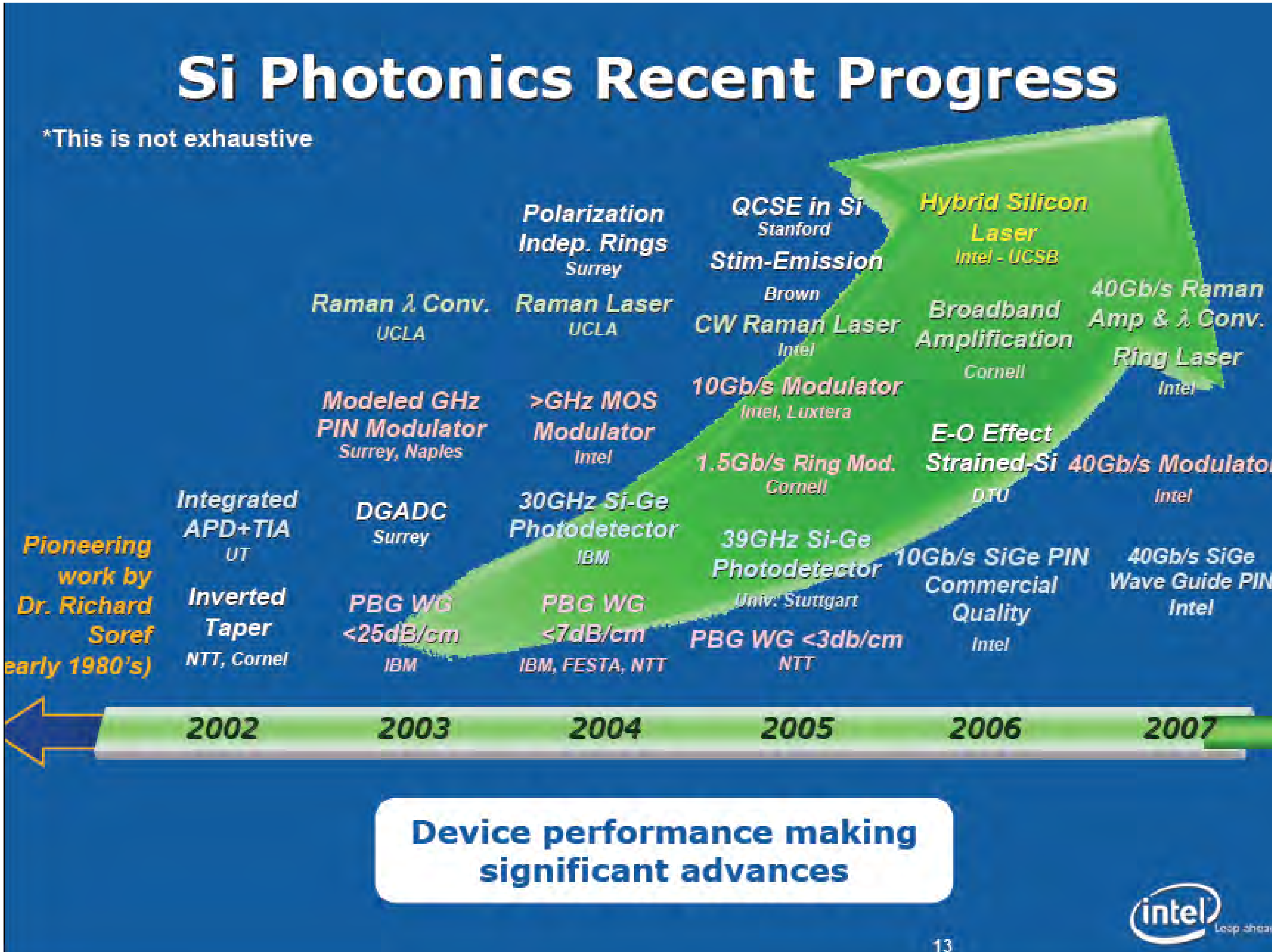




Silicon Photonic and Electronic Integration

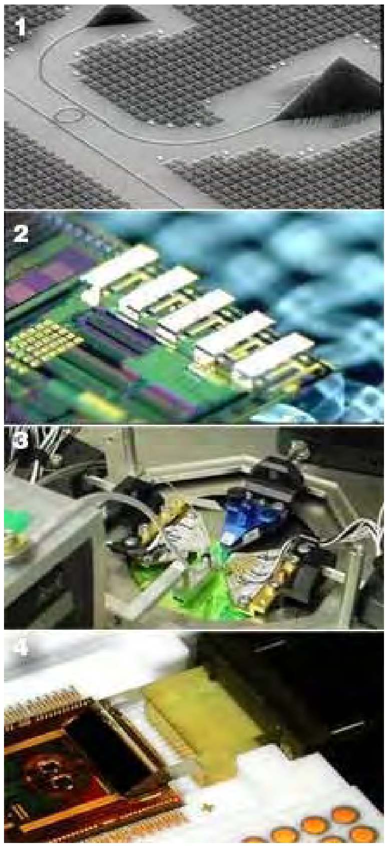


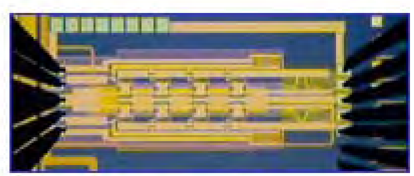
M Paniccia, 2007



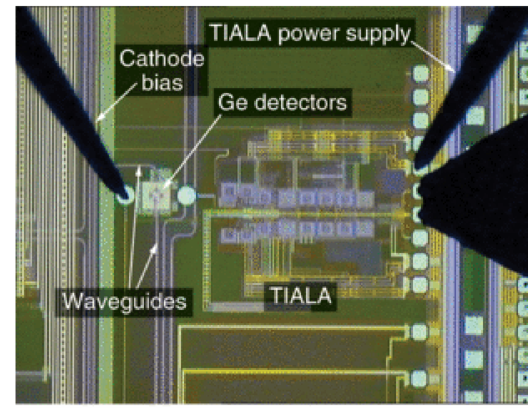
M Paniccia, 2007

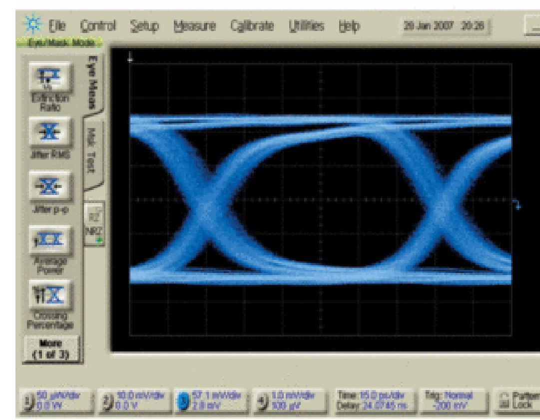
4 x 10G Optical Cable using Integrated Silicon Chip





High speed optical modulator
realized in CMOS-SOI

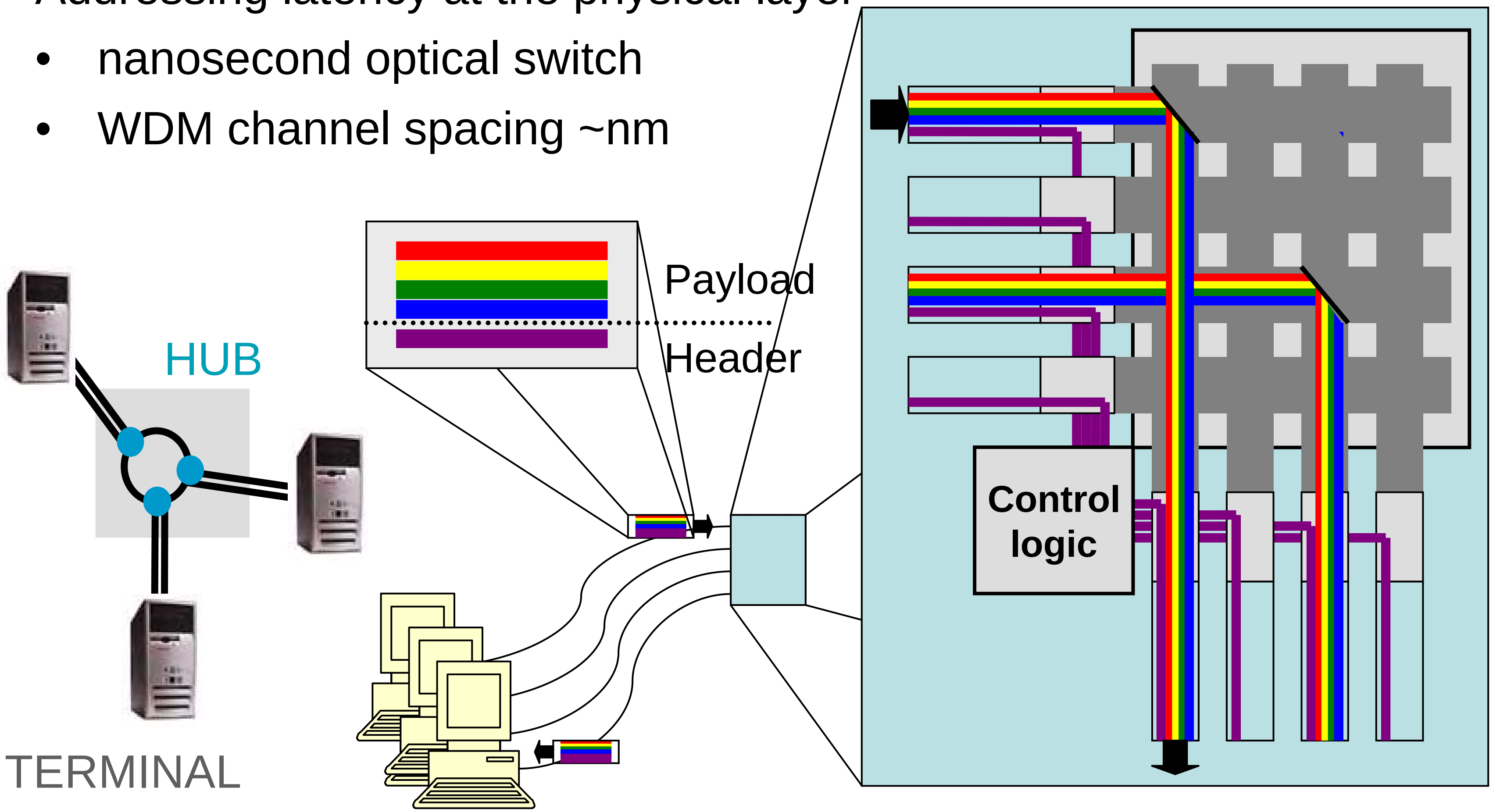




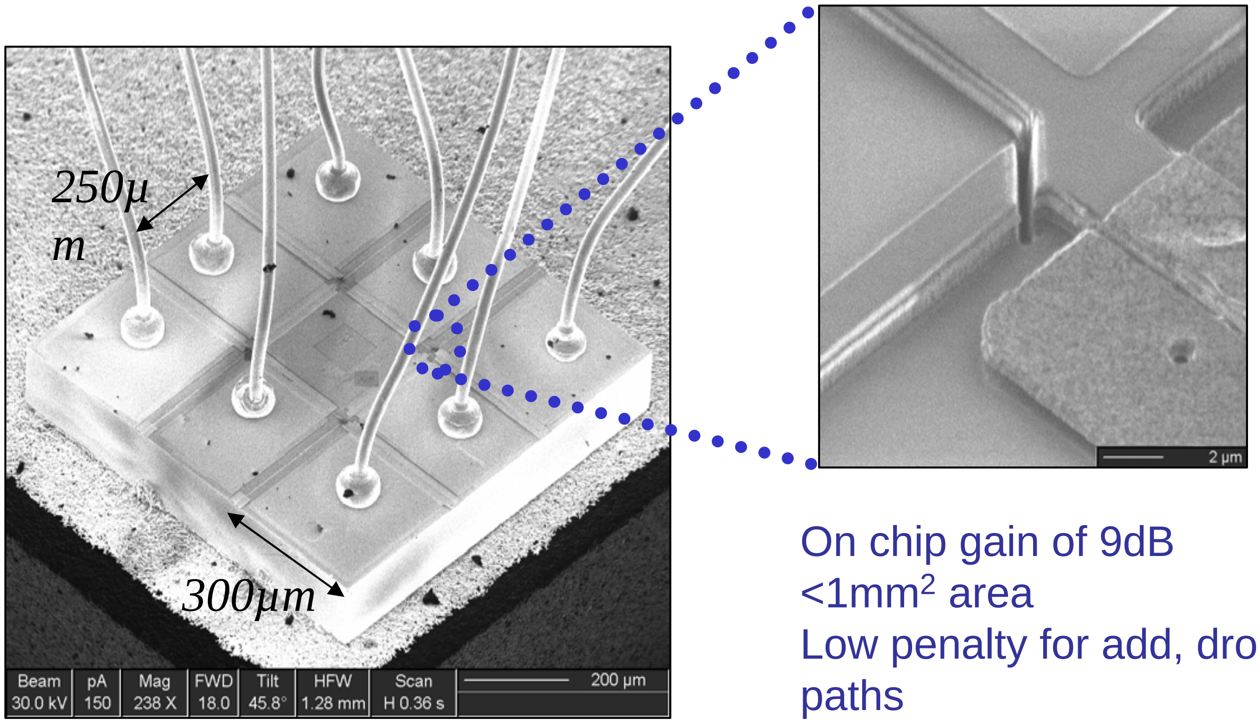
Optical Routing of Datacommunications Signals: Wavelength Striped Semisynchronous LAN

Addressing latency at the physical layer

- nanosecond optical switch
- WDM channel spacing ~nm



Integrated Photonic Switch Fabric



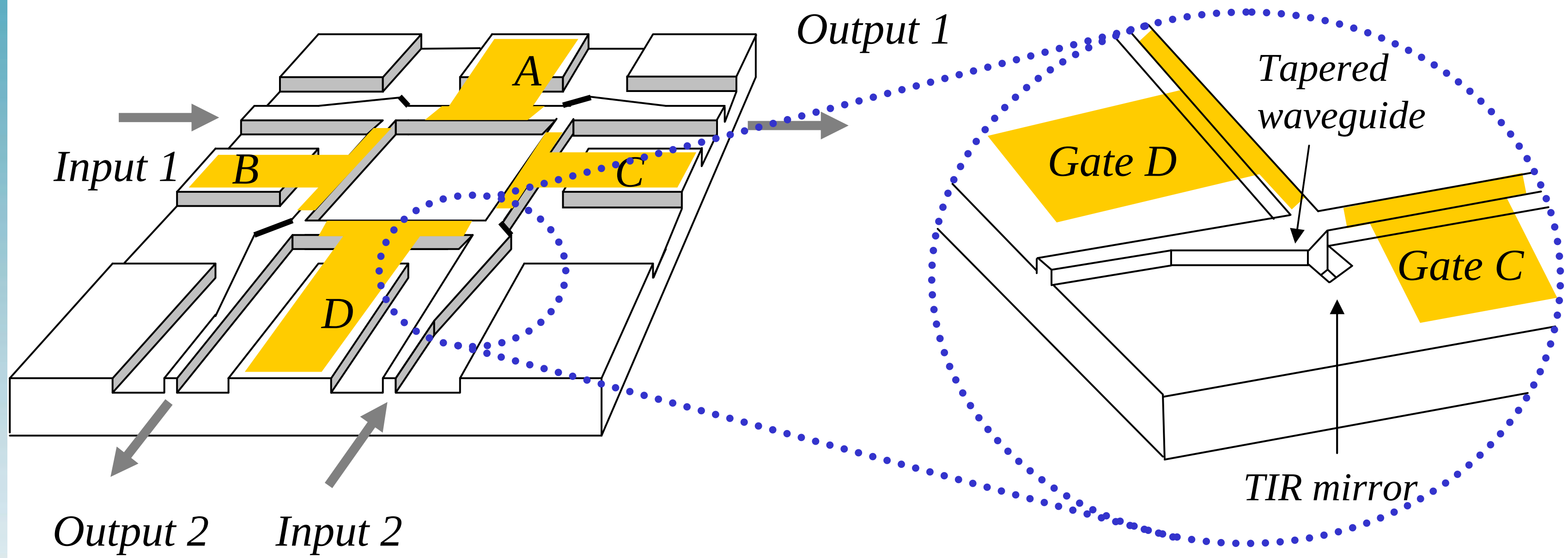
On chip gain of 9dB
<1mm^2 area
Low penalty for add, drop and through paths

InP based semiconductor optical amplifier technology

Conventional ridge waveguide fabrication processes with mirror etch

UNIVERSITY OF CAMBRIDGE

Integrated Photonic Switch Fabric



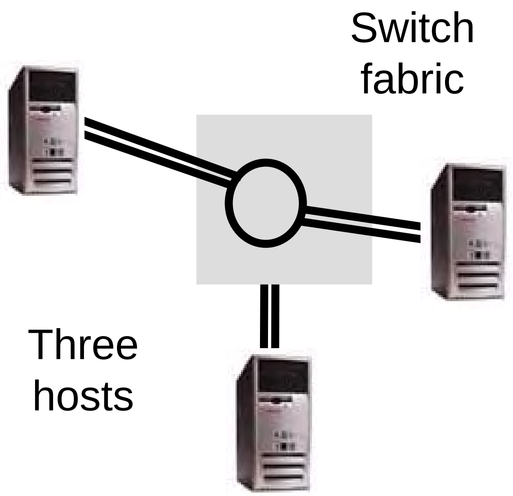
2 input - 2 output SOA optical switch configured

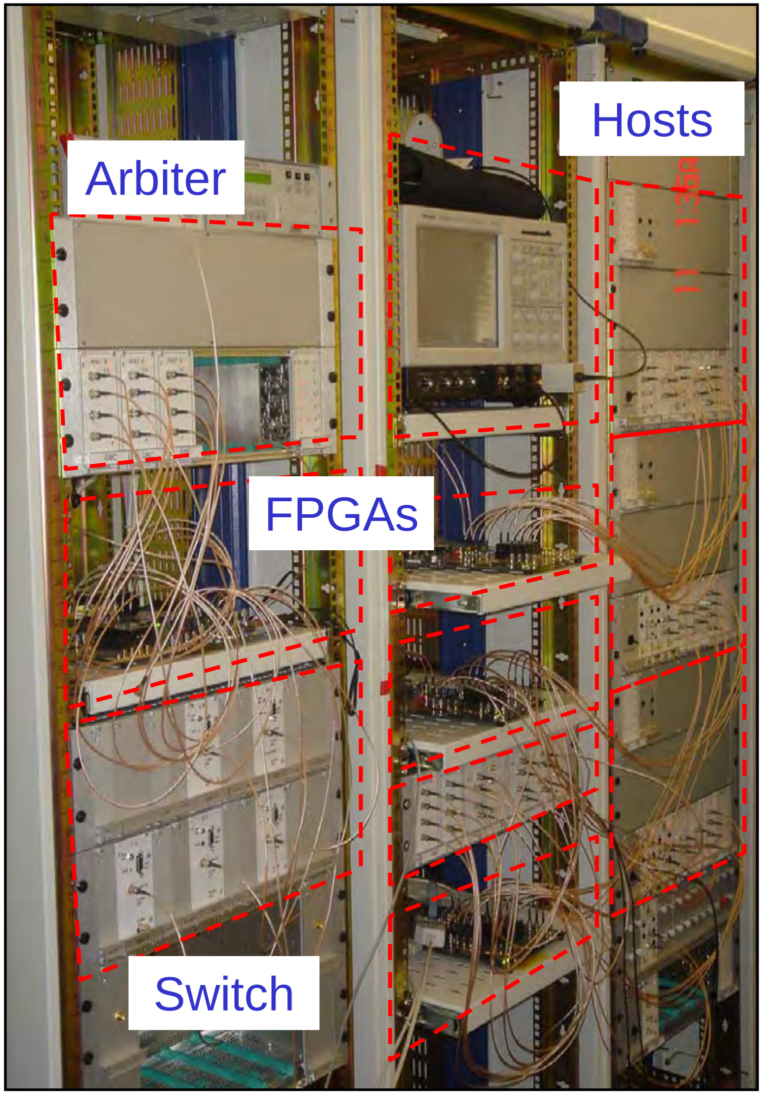
Implemented using 4 integrated SOA gates and 4 amplifying splitters

Nanosecond switching time

Low operating power: on state 1V, tens mA

Switched Wavelength-striped Test-bed





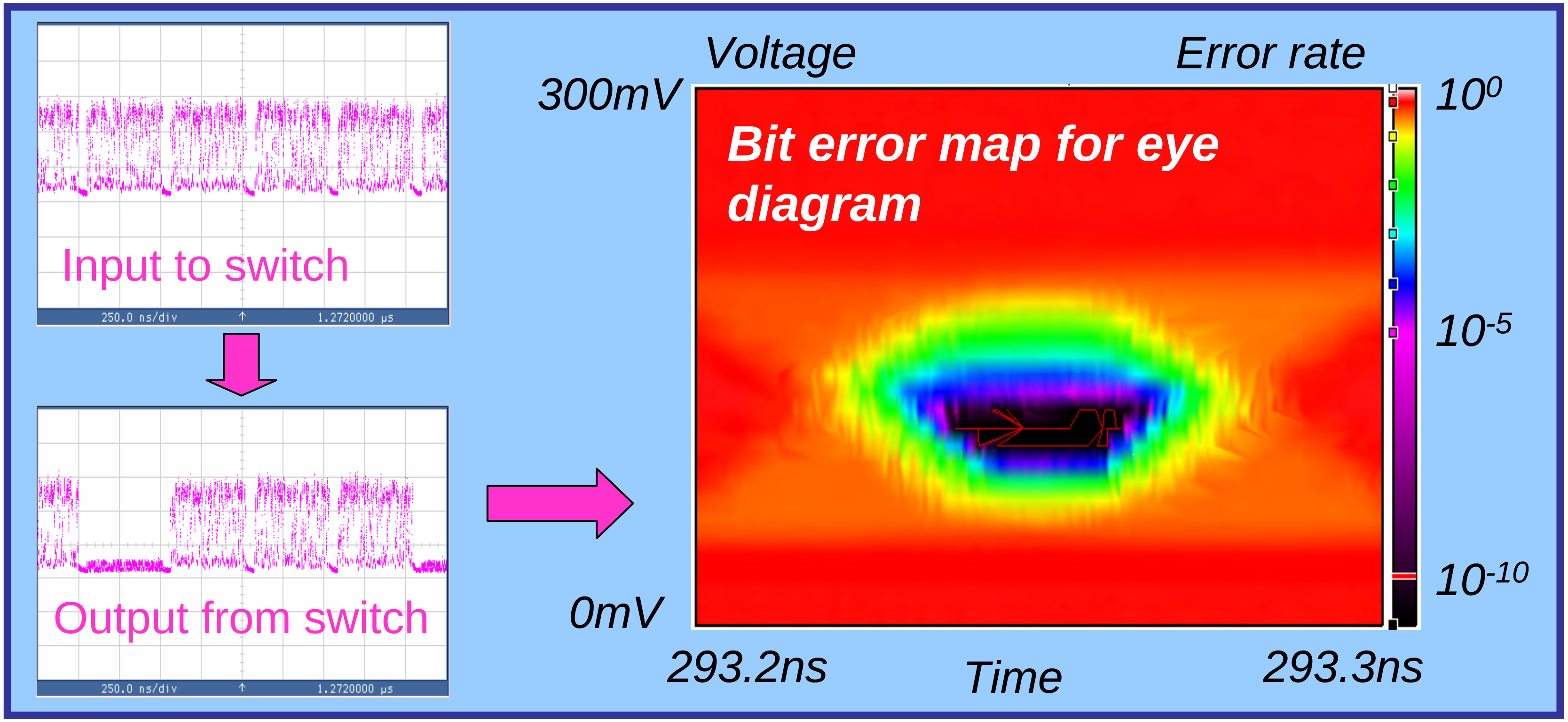
Media access control via 1.3 mm control wavelength

High capacity data within 1.5 mm band

Three FPGAs interface custom wavelength striped protocols to GbE and PC line-card

Fourth FPGA control SOA based switch

100 Gb/s Routing Performance for 2x2 Switch



Time resolved data packets and routed data packets (left) with three packets in four analysed

Bit error map (right) with open eye mask for one of ten 10 Gb/s data channels routed by 2x2 integrated switch

Outline

1 Introduction to Datacommunications

2 Background – the LAN/Server Networks

- GbE and 10 GbE systems
- The importance of MultiMode optical Fibre (MMF)

3 The Need for Optical Interconnects

- Cluster Computing, Chip to Chip and on-Chip

- PCB Optical Circuits

4 Conclusions

Optics in Interconnects

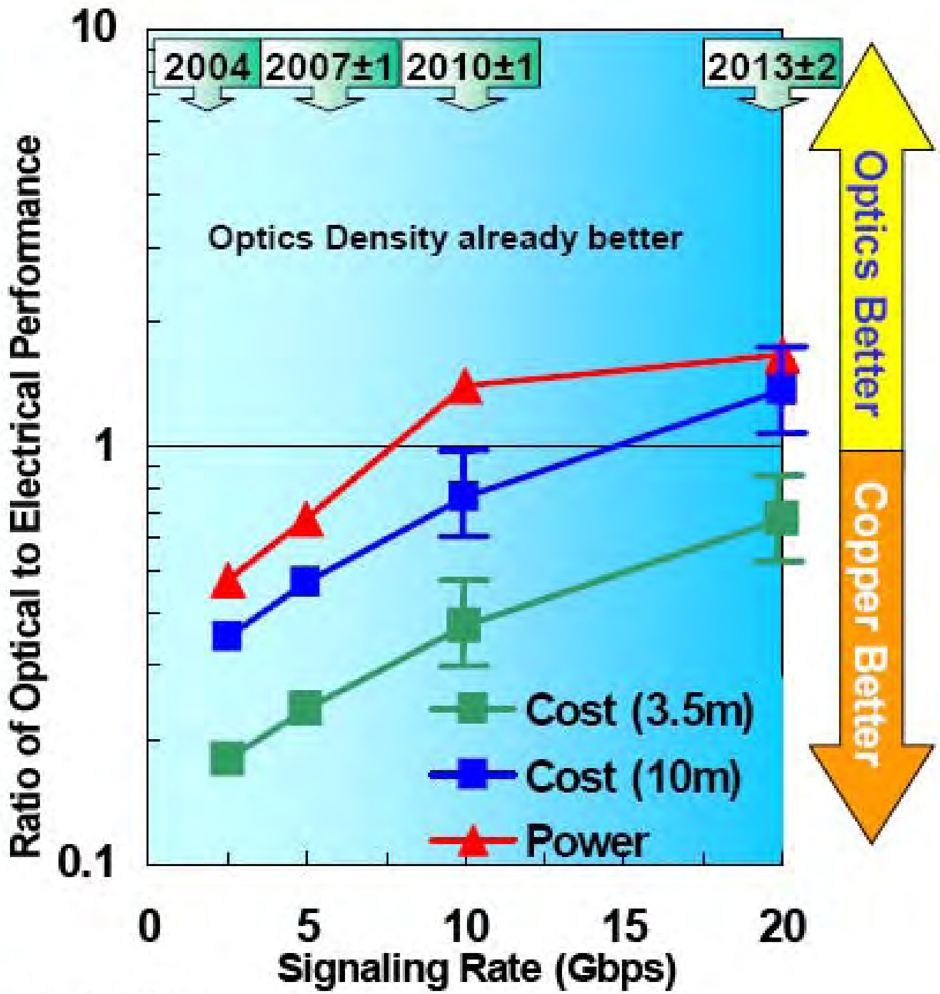
- Growing demand in optical interconnects driven by need for *high-capacity, short-reach* interconnections for future systems operating at data rates > 10 Gb/s.

- Existing interconnection technology:
  - Uses metal wiring architectures - sophisticated electronic techniques
  - Imposes a bottleneck to system performance due to inherent disadvantages such as
    - electromagnetic interference
    - size/density issues
    - power/thermal dissipation issues

- Optics - a promising solution as long as it:
  - is cost effective
  - has potential for integration into existing architectures
  - can be manufactured without significant capital expenditure (i.e. utilizes existing manufacturing processes and equipment)

Benefits for Optical Interconnects are Increasing for Shorter Distances as Signaling Rate Increases

2004 IBM Federation HPC Cluster



- Optics replacing electrical cable for >10m links @ 2.5Gbps
- Benefits for optics are increasing for shorter distances (<10m)
- Optical interconnects at less than 10m emerging in 2010

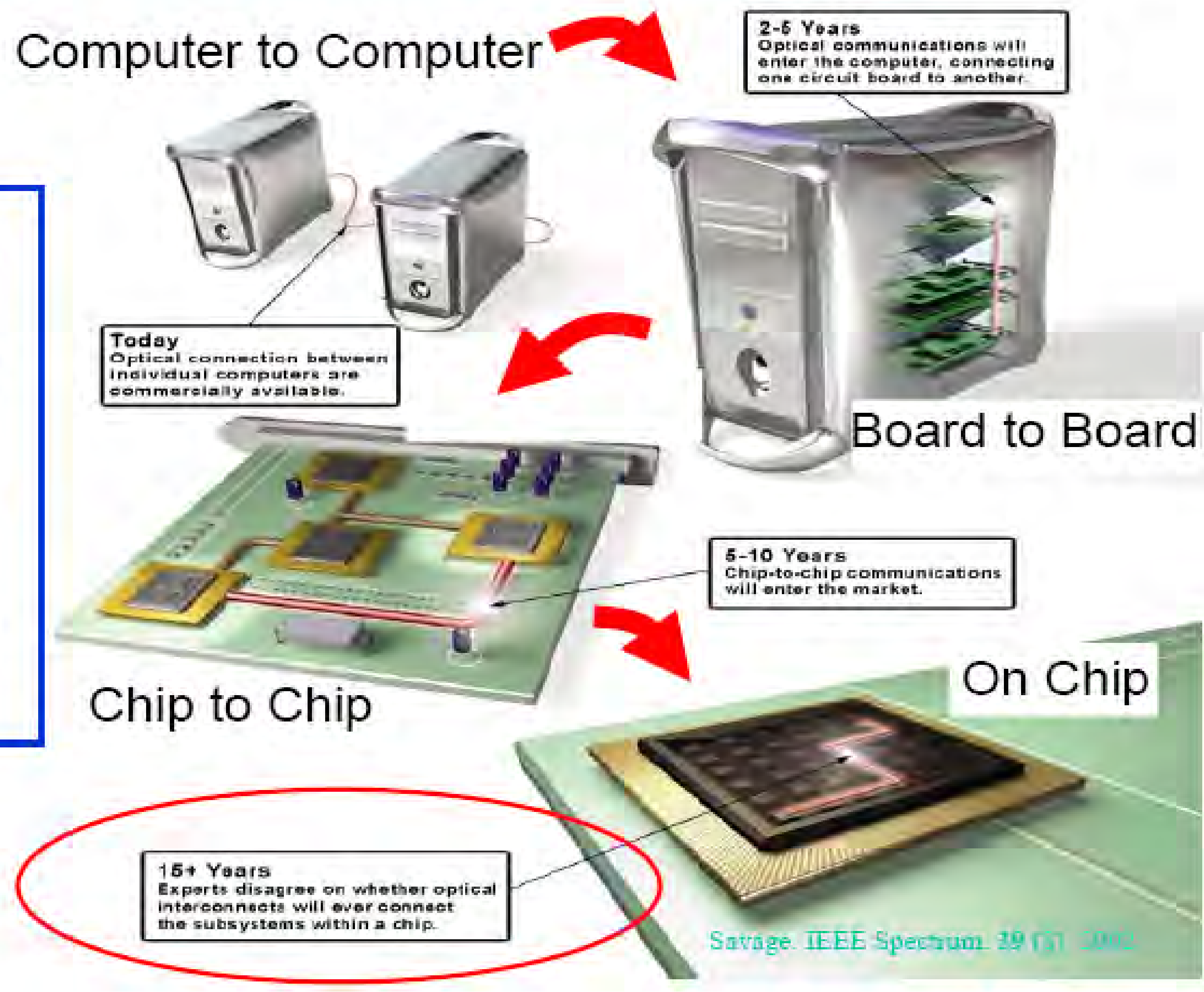
M.A.Taubenblatt 2006

© 2006 IBM Corporation

Optical Interconnect Roadmap

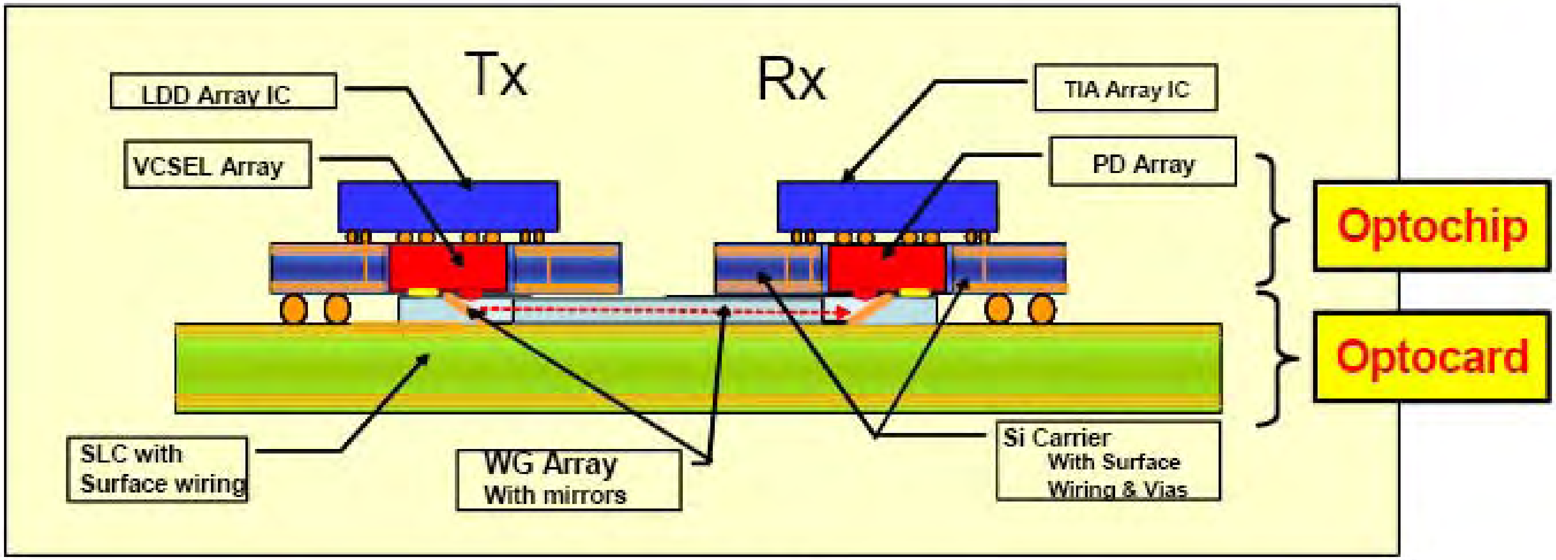
Electronic interconnection
DSL : 10 Mbps-km
Ethernet : 10 Mbps-km
Backplane : 10 Mbps-km

Optical interconnection
Board: 100 Mbps-km
Metro: 1 Tbps-km
Long haul: 10 Tbps-km



Rick Clayton, Clayton & associates,
Roadmapping exercise
for the MIT Microphotonics
Industry Consortium

Options for Chip to Chip (and Board to Board)



Program Goals

- Hybrid integration of optics into a server environment
- **Explore limits of:**
  - speed: 10-20Gb/s/Ch
  - density: 48 channels, 4x12
  - power: <10mW/Gbps/Ch (Ph I)
    - <5mW/Gbps/Ch (Ph II)

Major Efforts

- Optical Components: **2D arrays of 980nm VCSELs and PIN photodiodes, backside lenses**
- IC's: **Low Power CMOS designs**
- Si Carrier: **Wiring and mechanical packaging**
- Circuit Card: **Integrate waveguides onto SLC (buildup wiring) card for high speed surface**
- Packaging and Assembly: **Electrical interconnections, mechanical tolerances, thermal and mechanical issues, solder hierarchies**

Funded by Darpa

Is there another way?

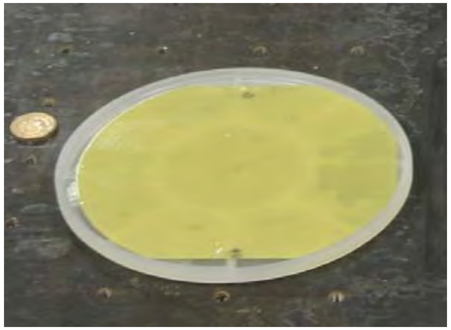
– Waveguides (and components on the PCB)

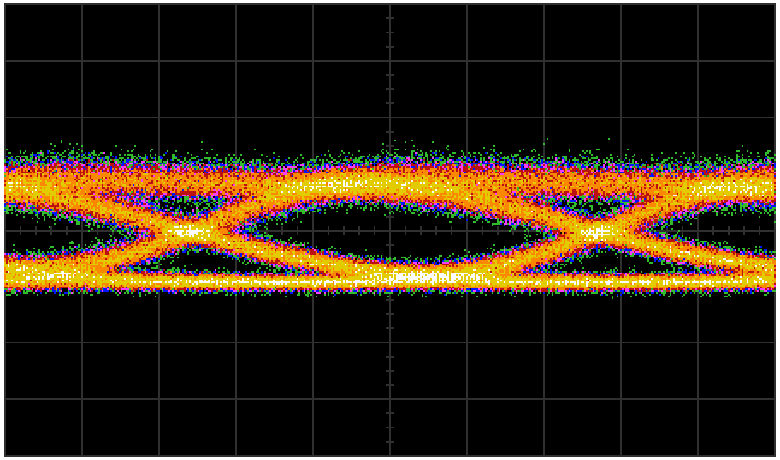
- Optical Interconnects today
  - We buy modules
- Electrical Interconnects today
  - Mostly assembled from subcomponents

- Need to move Optics to mass manufacturing from sub-components
  - Polymer waveguides on pcb

Multimode Polymer Waveguides

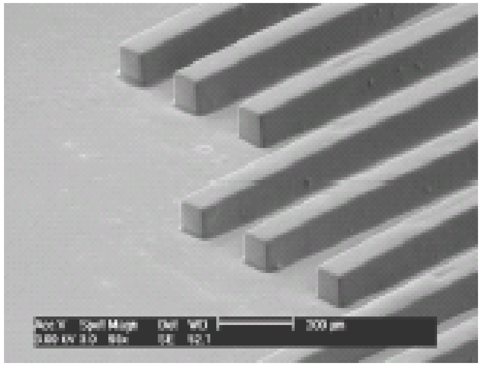
- Waveguides fabricated by conventional photolithographic techniques onto various substrates: FR4, silicon, glass.
- Waveguide cross-section is typically 50 µm x 50 µm, with waveguide separation of 250 µm to match conventional ribbon fiber, VCSEL and photodiode array spacing.
- Waveguides are effectively bit-rate transparent





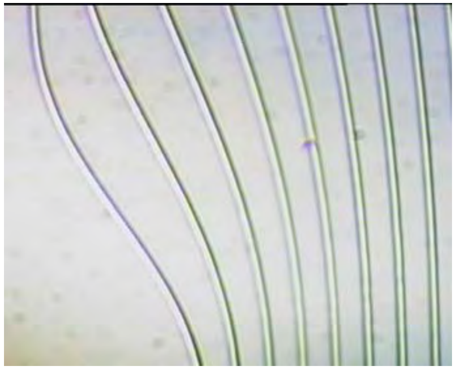
Eye from 10 Gb/s data transmission in 1.4 meter long spiral waveguide

Multimode Polymer Waveguides

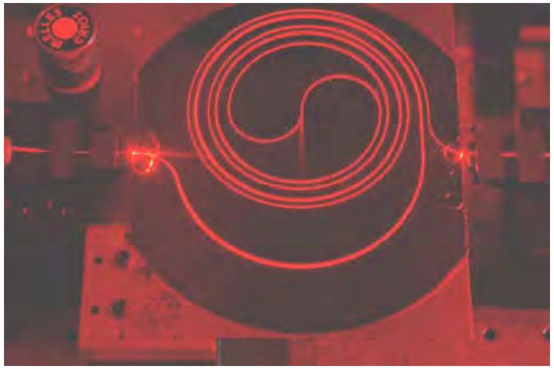


Straight waveguides

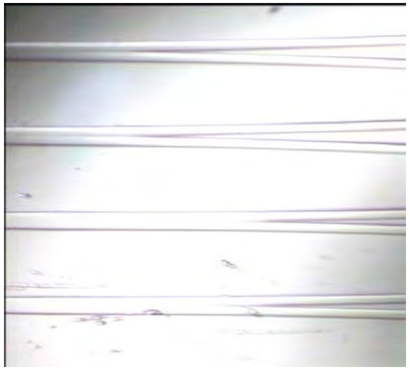
90° bends



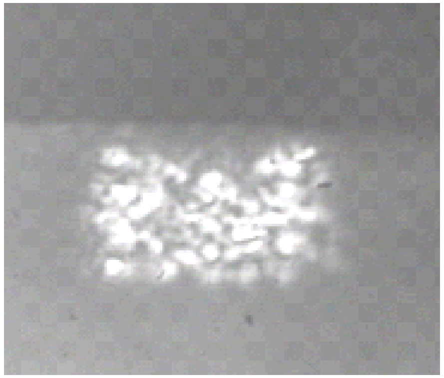
S-bends



Spiral waveguides



Y- splitters/combiners



Waveguide facet

Polymer Waveguides
based on Dow Corning PDMS polymer

Siloxane based polymer waveguides

→ meet key requirements for successful integration into existing architectures and manufacturing processes

Siloxane polymer materials exhibit:

- excellent mechanical and thermal properties.
- withstand > 250^{o}C required for lead-free solder reflow.
- can be deposited directly onto standard FR4 substrate.
- low intrinsic loss at 850 nm wavelength → 0.03-0.05 dB/cm.
- readily patterned by photolithography or embossing techniques

UNIVERSITY OF CAMBRIDGE

Application Space: Backplanes

- Blade servers are a popular method of increasing packing density in IT environments.

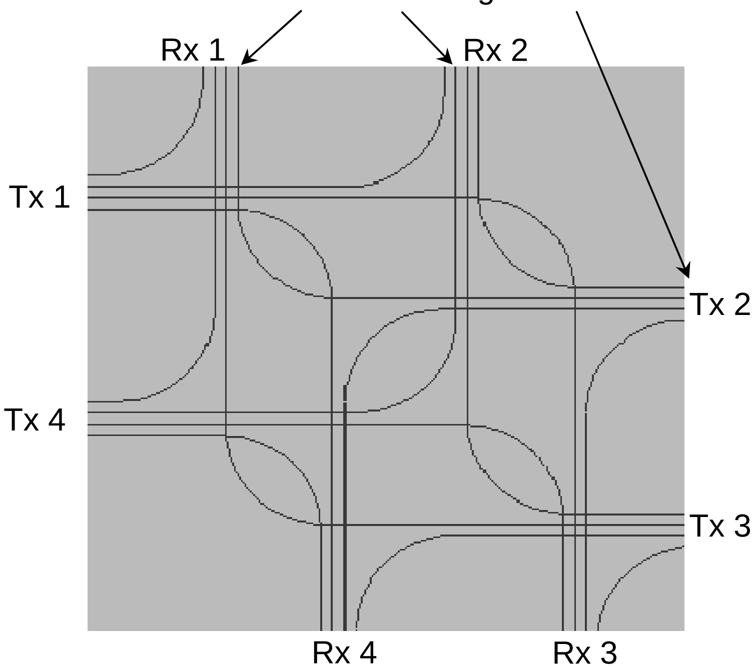
- Network connectivity is currently provided by an electrical backplane capable of providing several Gb/s total throughput.

- Blade servers typically have 14 blades and another 2 external network connections, making a total of 16 backplane connections.

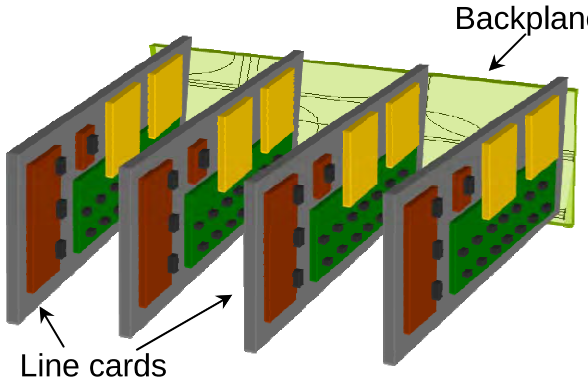
- There is a perceived need for a low cost next generation backplane which will enable one blade to talk to any other in the chassis at ~1Gb/s.

Polymer Backplane: Design Approach

Ribbon fibers connect at board edges and run to line cards.



Current implementation uses standard ribbon fibres to link backplane to transmit and receive arrays on line-cards.

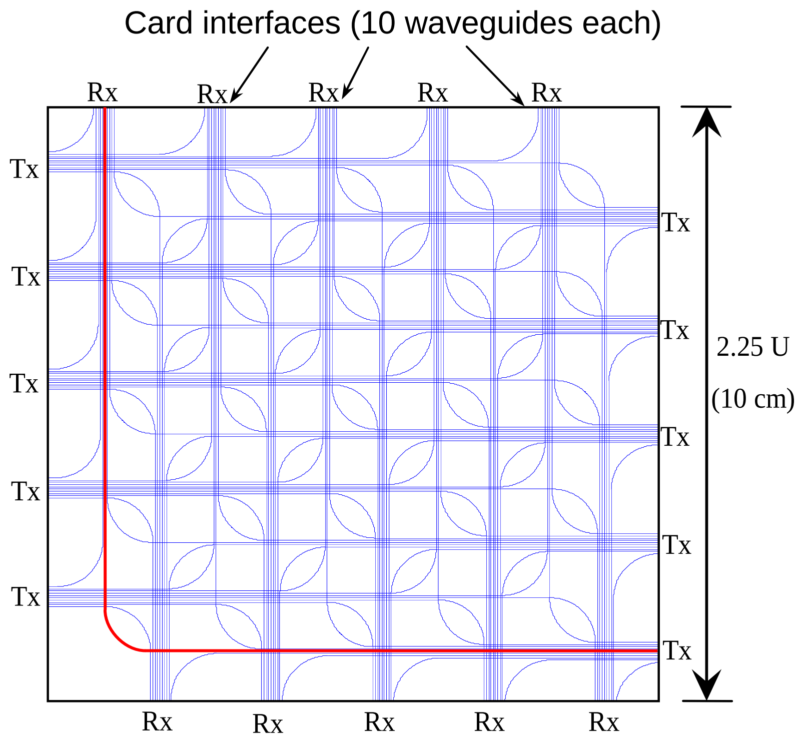


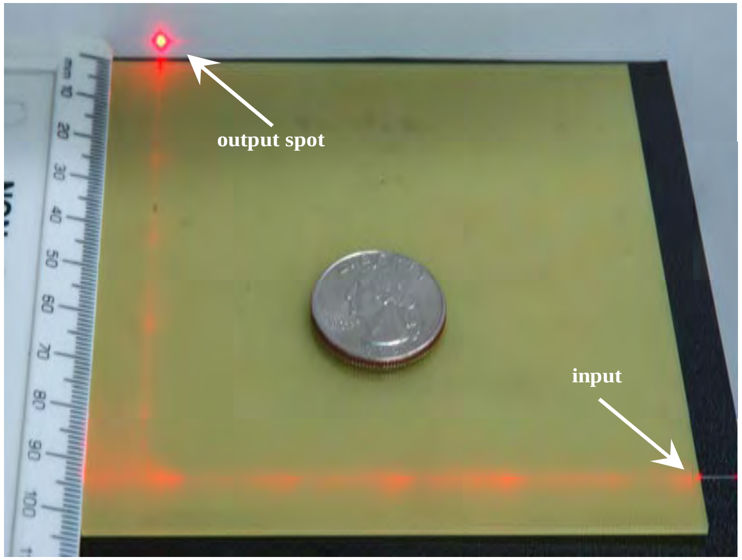
Schematic of conventional electrical backplane with pluggable line cards.

Polymer Backplane: Design Details

- **simple 90° bends** rather than corner mirrors
  - bend loss ~ 1 dB for 8mm RoC bend
- **90° waveguide crossings** – all structures in single plane
  - crossing loss ~0.01 dB/crossing with MMF input
- **crosstalk** < 30 dB
- **waveguide spacing of 250μm** – matches ribbon fiber

Demonstrated 10 Card Optical Backplane





Schematic of 10-card backplane layout and

- 100 waveguides
- single 90° bend per waveguide
- 90 crossings or less per waveguide

Photograph of FR4 based backplane with red light tracing the link illustrated at left. Note output spot visible at top.

Insertion Loss and Crosstalk Measurements

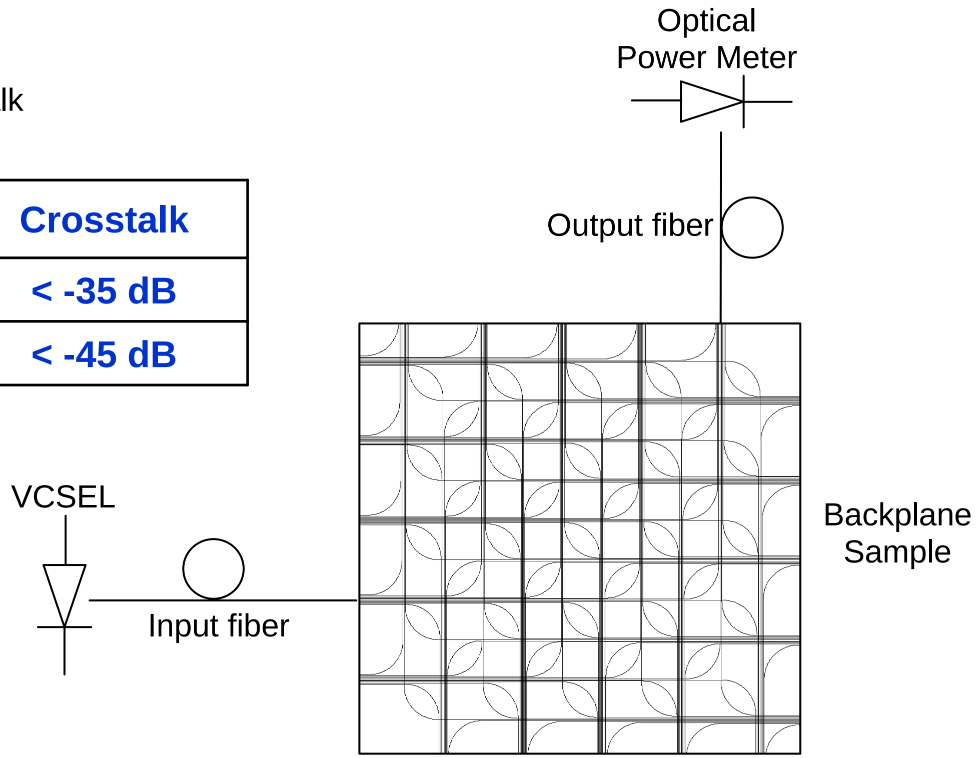
Worst-case values

- **longest** links
- links **most susceptible** to crosstalk

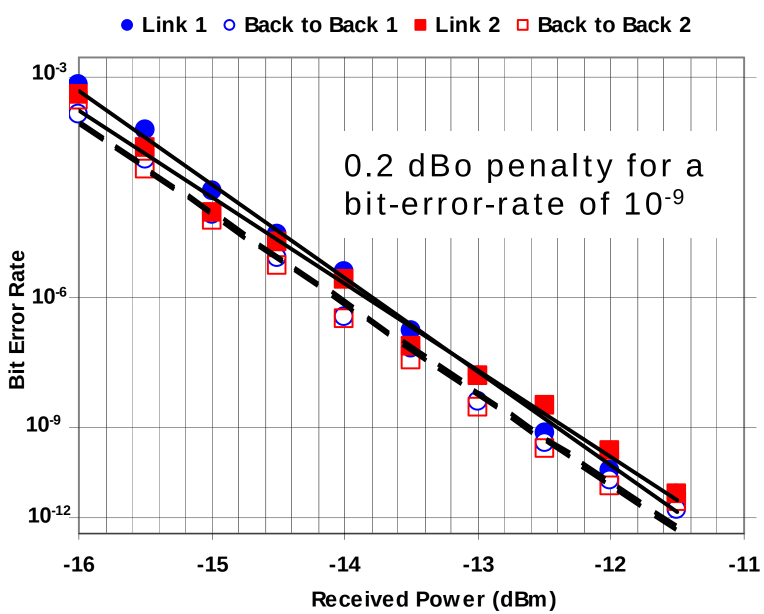
Input Type	Insertion Loss	Crosstalk
50 µm MMF	2 to 8 dB	< -35 dB
SMF	1 to 4 dB	< -45 dB

As anticipated from previous work, *crosstalk from bends an crossings not a problem.*

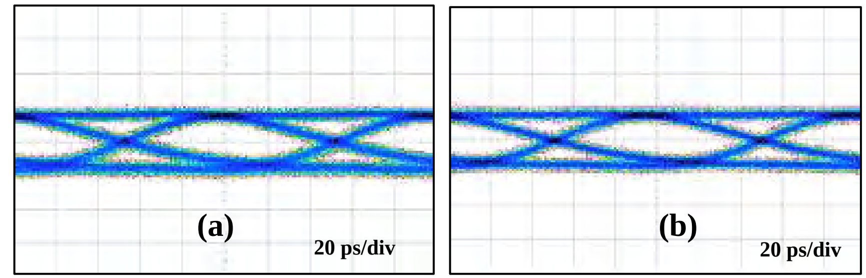
Crosstalk contribution primarily due to coupling between long adjacent parallel waveguides.



Data Transmission Studies at 10 Gb/s (1 Tb/s Aggregate)



BER plot for two typical waveguides at 10Gb/s, 2^{31}-1 PRBS. Solid line denotes BER for link, dashed line BER for corresponding back-to-back.



Recorded eye diagrams for (a) back-to-back and (b) waveguide link for 10Gb/s, 2^{31}-1 PRBS.

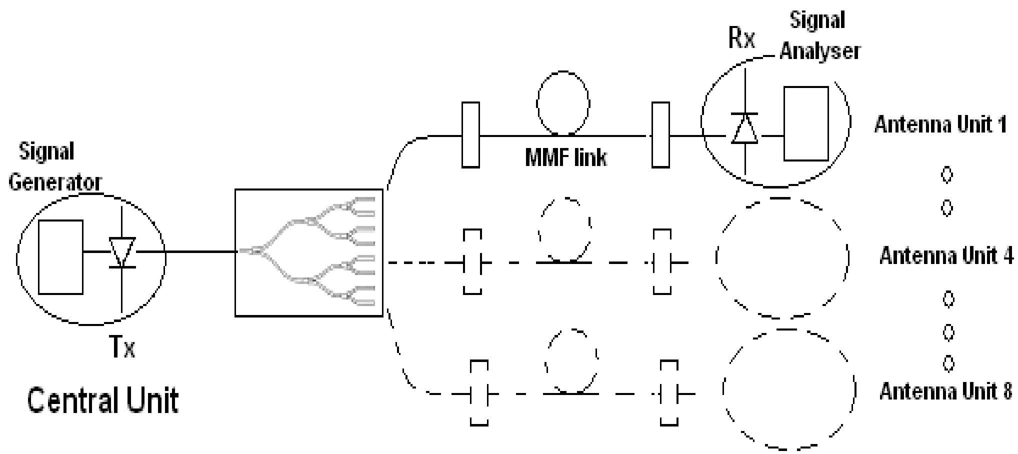
Dell PowerEdge 2850 servers for GbE tests

Gigabit Ethernet Demonstrated Across Backplane

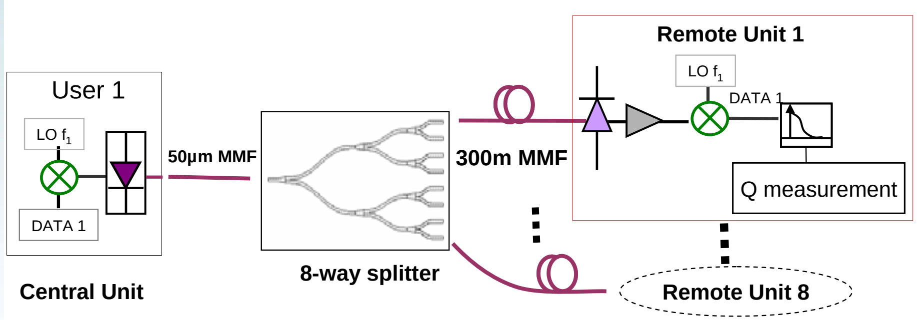
- full line-rate data transmission with no dropped packets
- transmission across waveguides with highest loss and greatest crosstalk

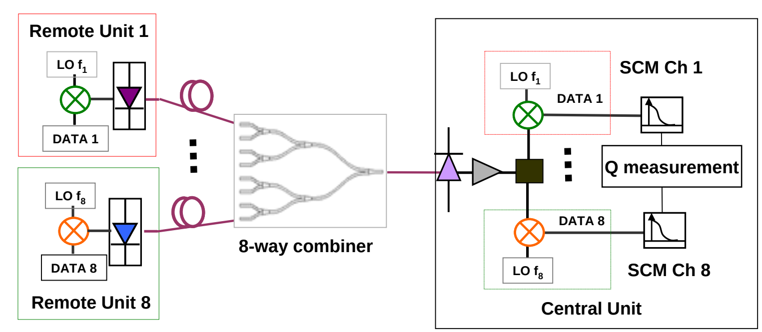
Demonstrated Application of Y-splitters/combiners

Devices used to demonstrate: RoF multicasting/Multimode PON architecture



Downlink of RoF network





MM PON Downlink **MM PON Uplink**

4 Conclusions

High performance *low cost* photonic transceivers can deliver transmission bandwidth for a range of LAN applications

MMF remains the dominant in-building fibre type

Recent advances in transmission have led to high performance demonstrations – > 10 GbE

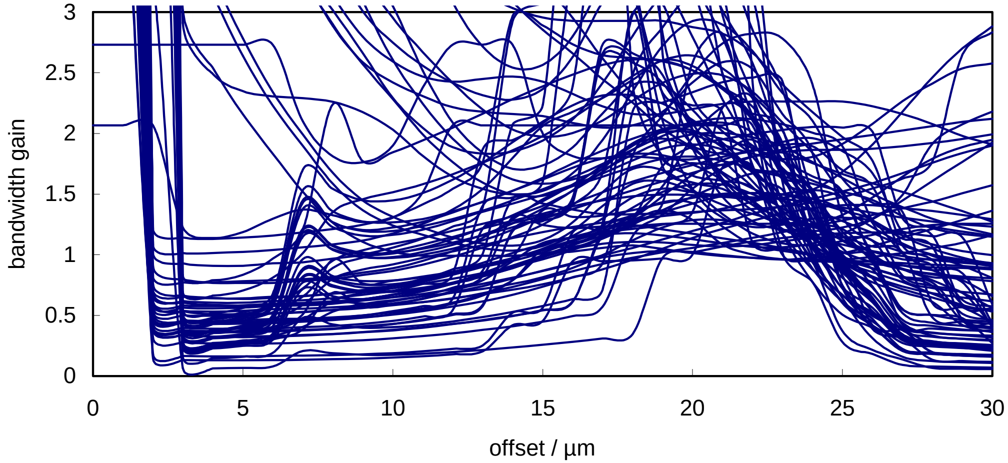
However MMF data links have the potential to be useful for interconnect applications also

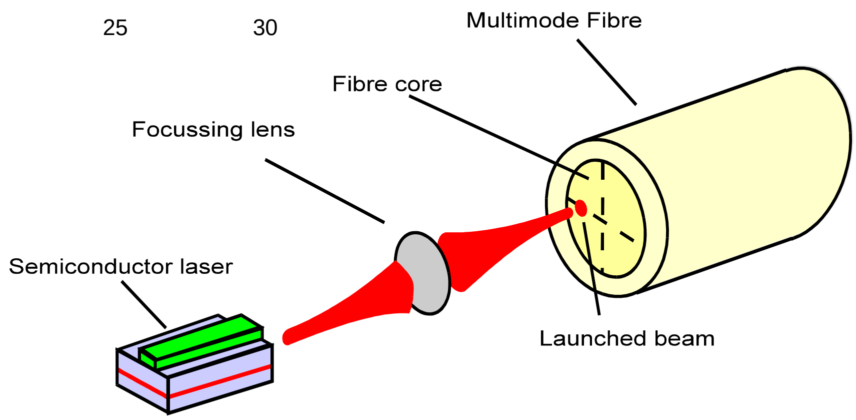
Simple low cost backplane is implemented with 1 Tb/s capacity

Background Slides

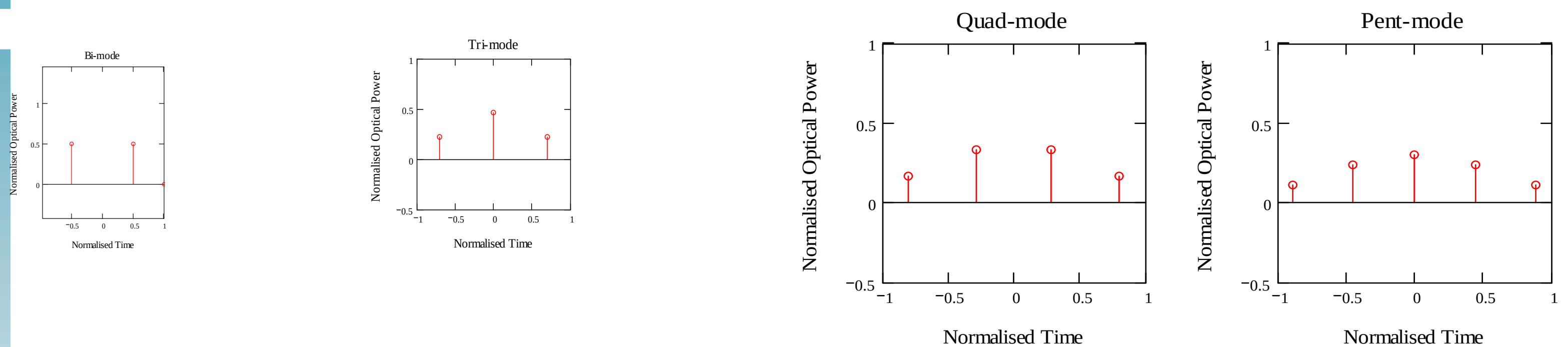
Gigabit Ethernet statistical model results

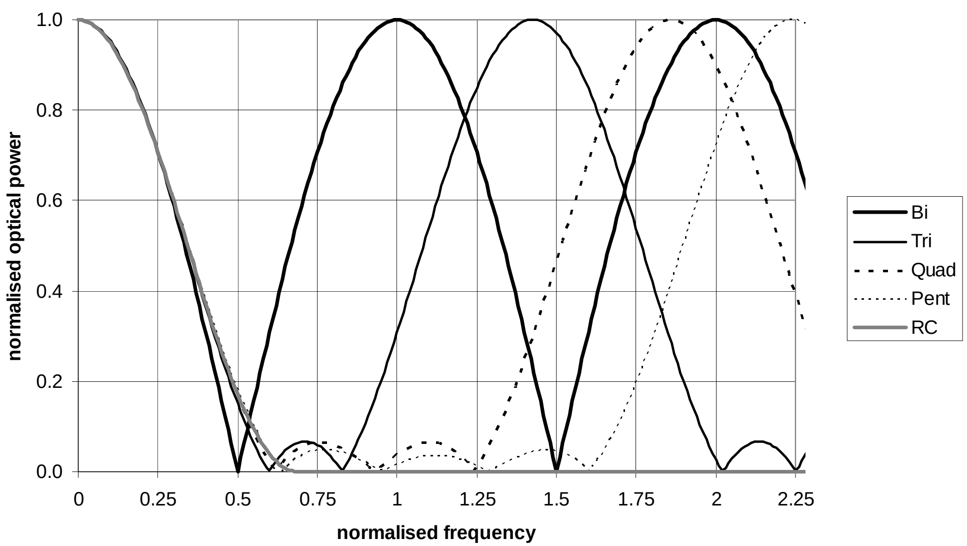
Calculate –3-dBo bandwidths of the MMF links, which is the key indicator of performance when using conventional receivers. For example:





A new approach:
Normalised worst case impulse and frequency responses





- The worst case discrete impulse response (IPR) and frequency response (FR) for the first four worst case IPR are plotted.
- The responses are normalised such that they have the same 3 dB electrical (1.5 dB optical) effective modal bandwidth (EMB)

Silicon Optoelectronics

Silicon photonics can satisfy distance x bandwidth needs of emerging volume applications.

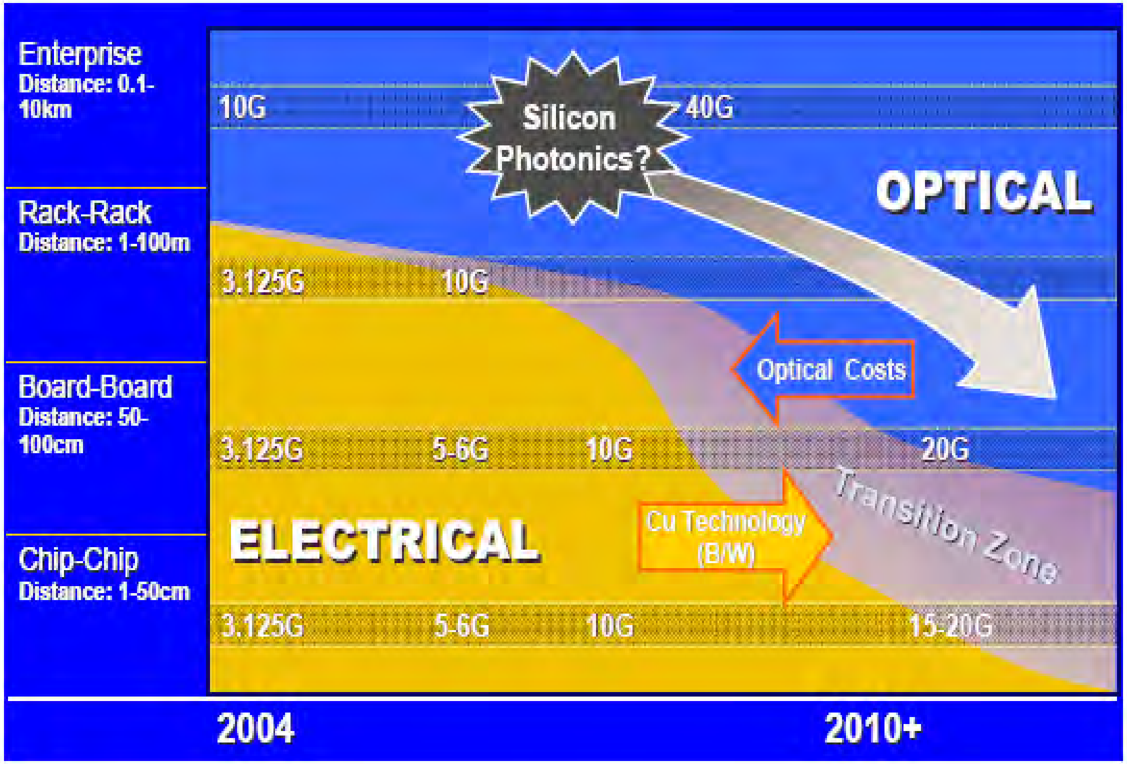
Key market driver is reduced cost and growing edge bandwidth requirement

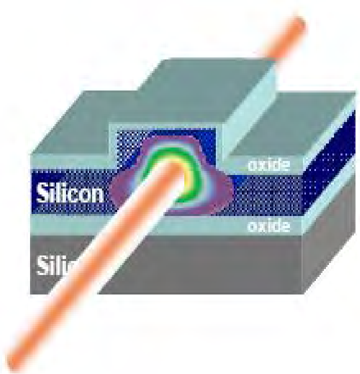
Key to reduced cost

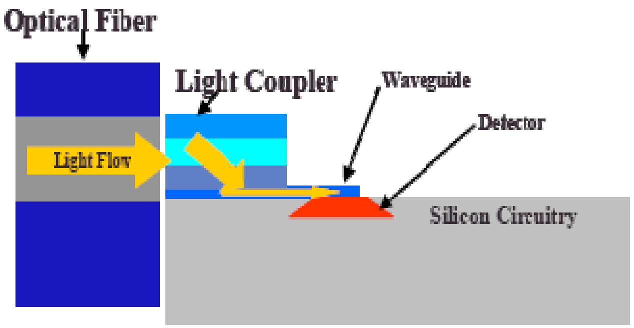
Monolithic integration of selected technologies

Standardization of processes and form factors

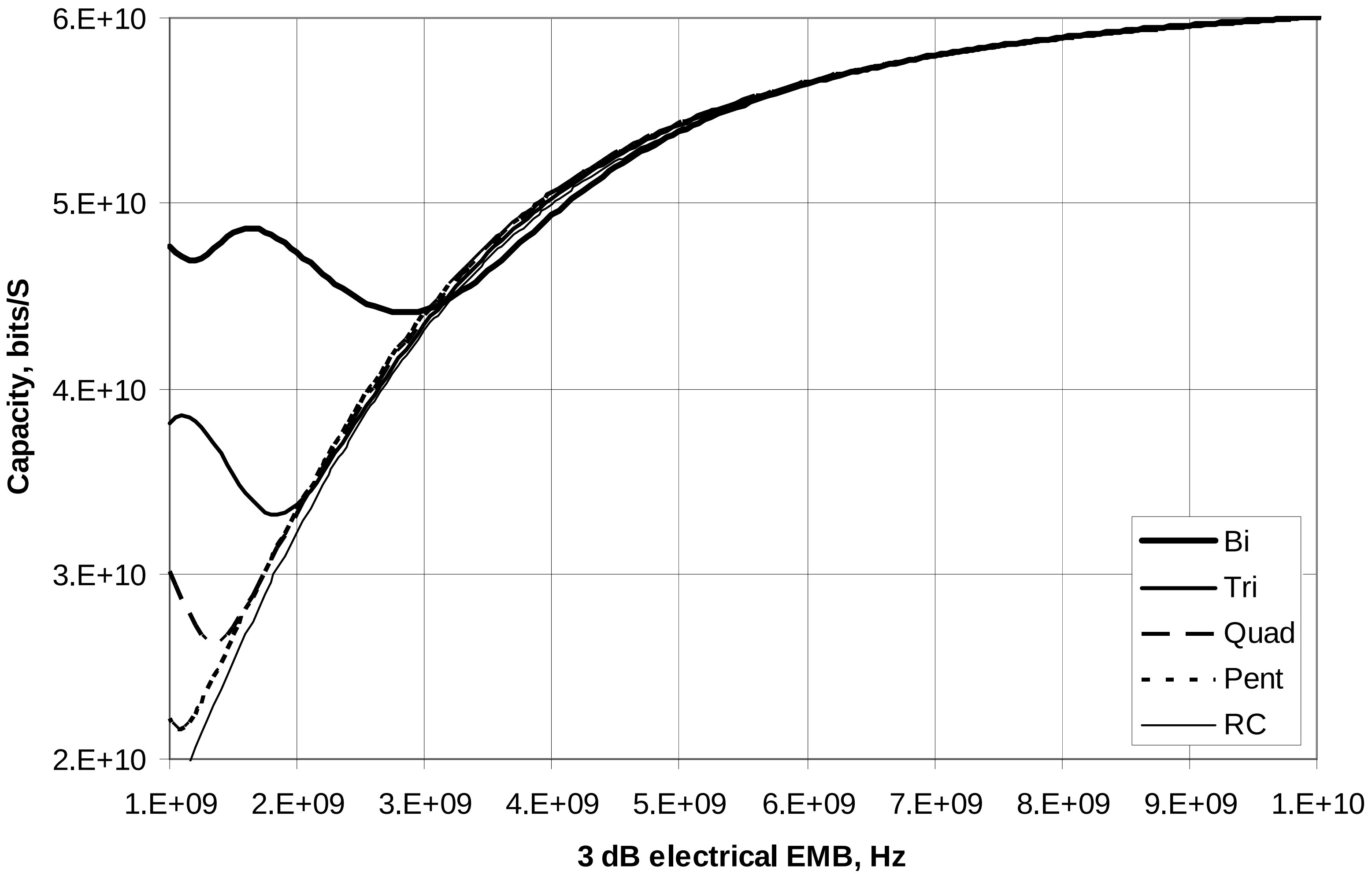
Opportunity for a $2G business by 2010







Shannon Capacity versus EMB



Systems Work using Multimode Y-Splitters/Combiners

No fundamental 3 dB loss as in single-mode combiners.

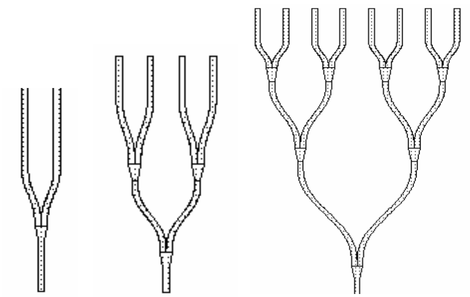


Fig. 1: Schematic of polymer Y-splitters

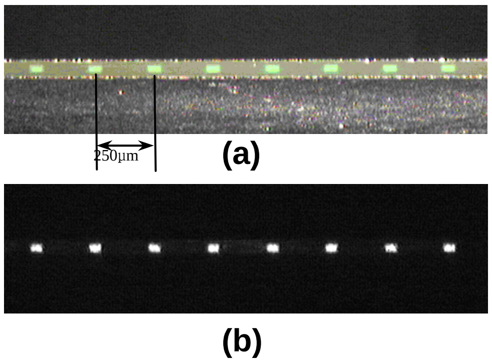
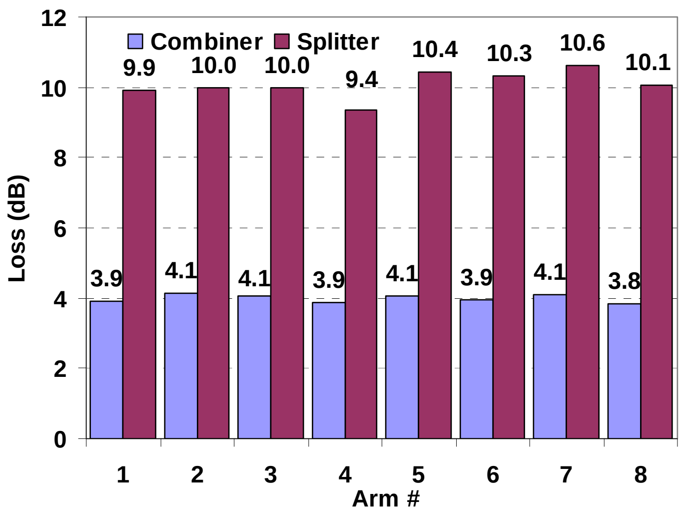


Fig. 2: Output facet of a 1x8 splitter
(a) photograph (b) IR image with an 850 nm

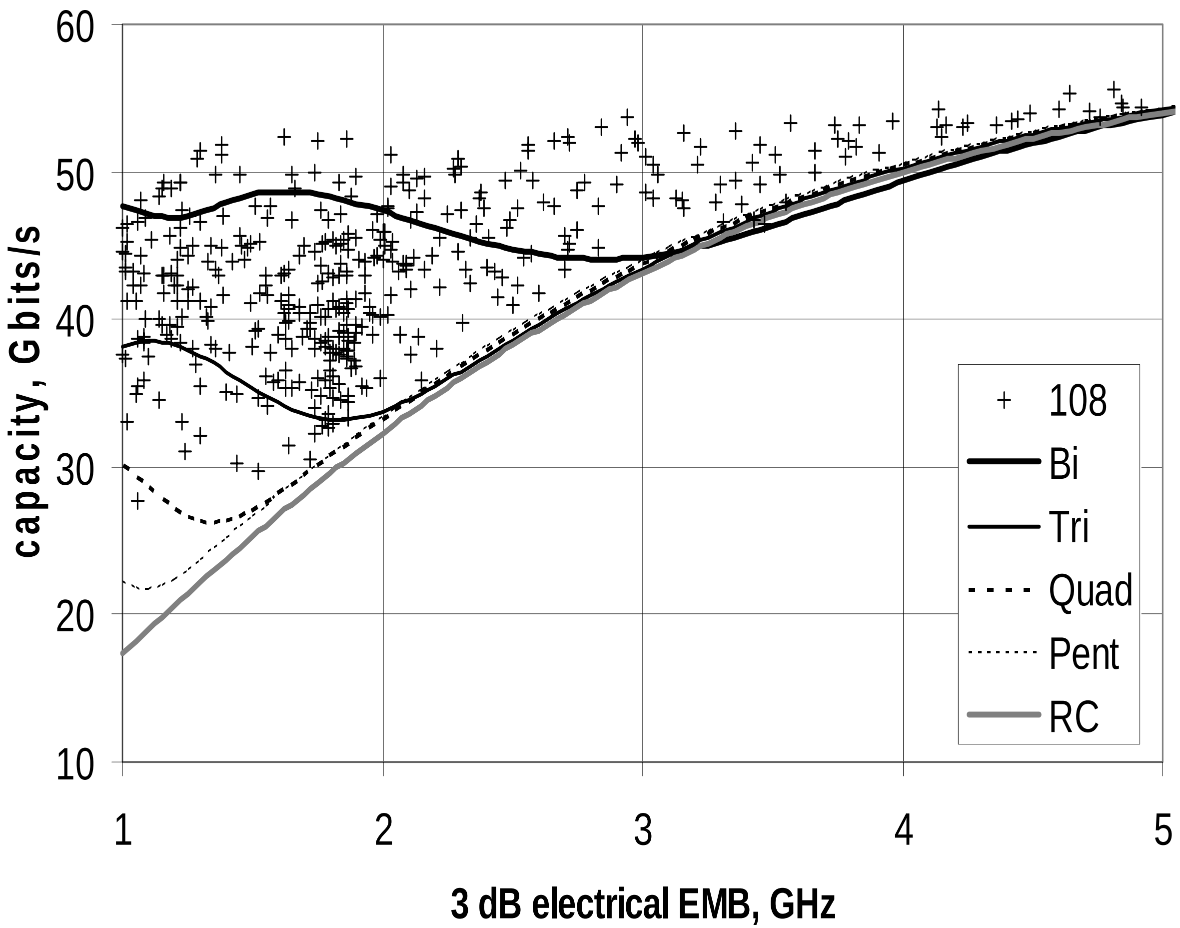
Uniformity of Splitting/Combining



Input	Splitter loss (dB)		
	1x2	1x4	1x8
SMF	3.4	6.6	10
50µm MMF	5	7.8	11
62.5µm MMF	5.7	9	12.5

Input	Combiner loss (dB)		
	1x2	1x4	1x8
SMF	0.9	1.5	4
50µm MMF	4	5.1	7
62.5µm MMF	4.7	6	9

Shannon Capacity Versus EMB: OM1 at 1300 nm



For the first time:

- Calculated the capacity of MMF
- Derived an analytical worst case model
- Eliminated the need for time consuming statistical models

Expectations for Server Optical Interconnects: 2010

- CMOS frequencies won't exceed 10 GHz (power efficiency)
  - Performance gain from parallelism & sophisticated design
  - System-level: Interconnects & optics get more important in time

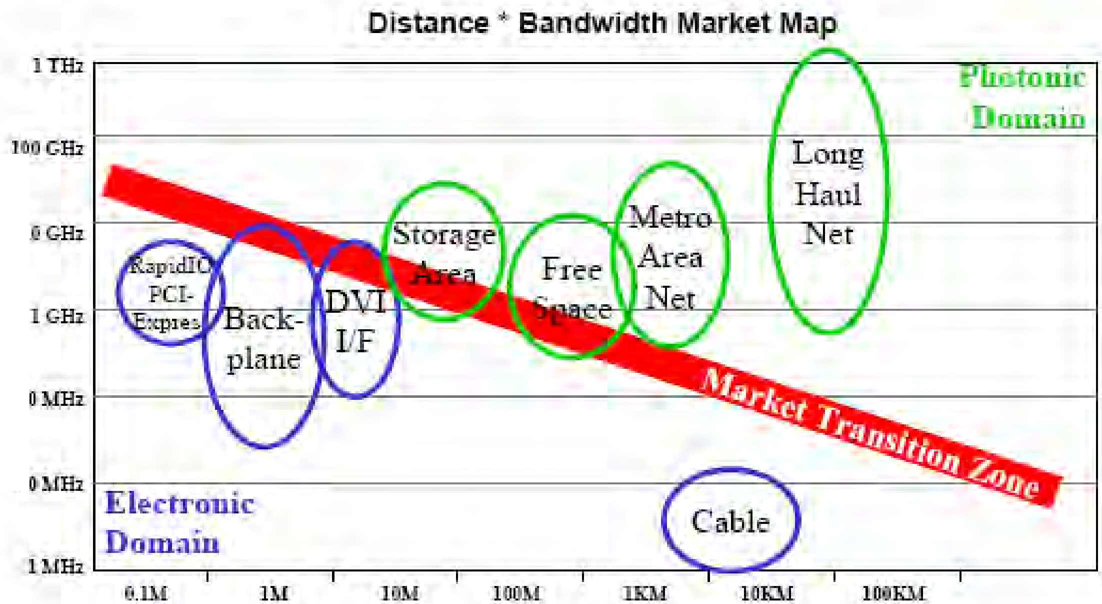
- Cost per Gb/s for 10–50 m optics must reach $1 per Gb/s
  - Optics for 25 Petabit/sec network cannot cost more than $25M, to balance remainder of system costs

- Longer-link methods will be used for shorter-length links
  - Telephone tech (FEC, NEXT/FEXT cancellation, PAM...) in 10GBASE-T
  - WDM (probably CWDM) for on-campus links

adapted from Alan Benner, IBM

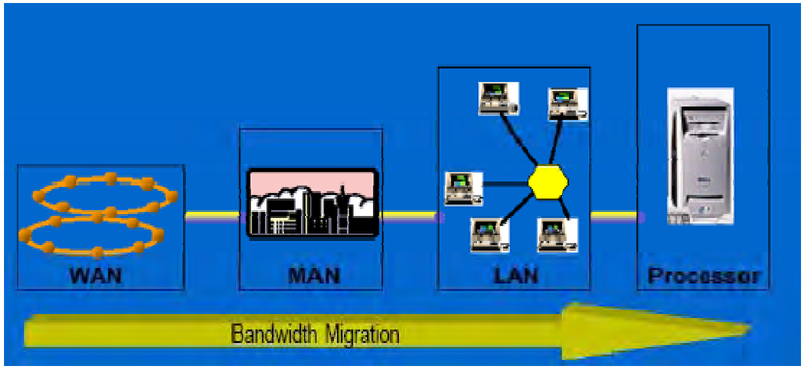
Market Technology Drivers

- **Optical/electrical transition point a moving target**
- **New applications emerging**



Industry	Current BW	Doubles every:	Electrical Limit	Transition Year
Serial Computer Bus	2.5 Ghz	4 Years	15-20 Ghz	12 Years
Backplane	3.6 Ghz	4 Years	10-20 Ghz	8 Years
DVI (Display I/F)	8 Ghz	4 Years	10 Ghz	Soon!
Cable	200 Mhz (Shared with up to 50 subscribers)	TBD	~200 Mhz (Shared with up to 50 subscribers)	TBD

BW and Volume Drivers



Current BW (hz)	1E12	1E11	1E10	1E9
Future BW	1E12	1E11	1E11	**1E12**
Interconnects	1E5	1E6	1E7	1E9
Wafers/week	2	20	200	**20,000**

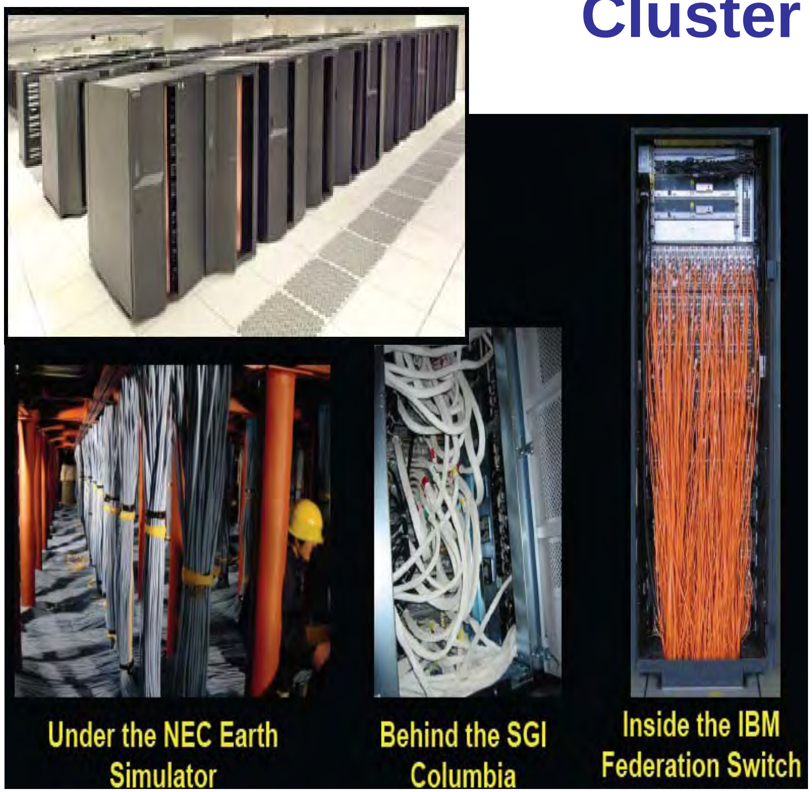
Volume drivers BW drivers

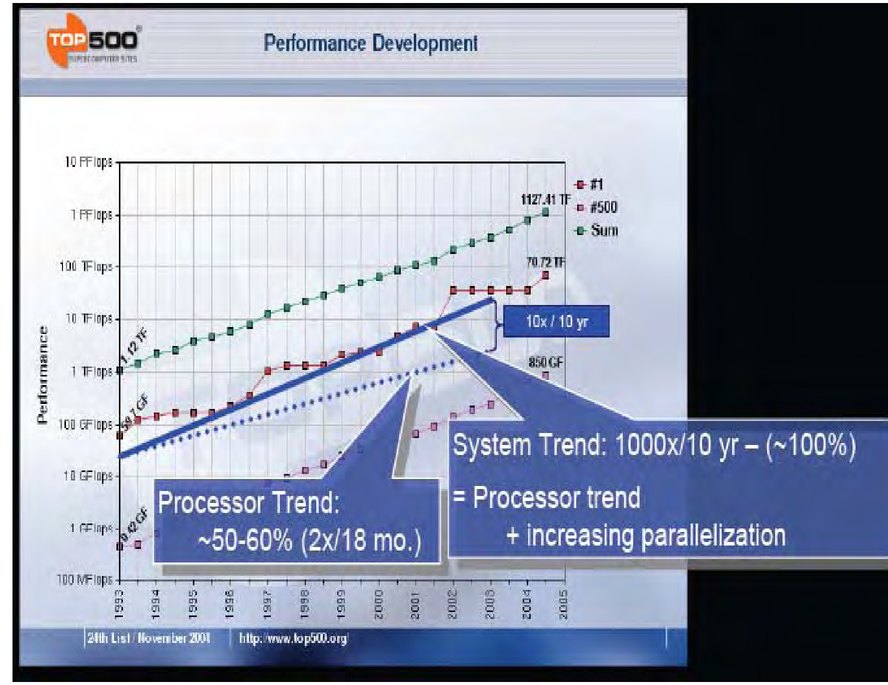
	Gbd	Cost/Gbd	Power/Gbd
Cu	20	$1.5	33mW
XFP	10	$5-10	200mW
Parallel Optics	245->1T	n/a	36mW
MAUI	240	n/a	3mW
Targets	>40	<$1	<25mW

- Requirements for investment:
  - Volume driven by network edge
  - Standardization in processes, form factors, etc

UNIVERSITY OF CAMBRIDGE

Cluster computing





- Computer performance is a function of internal architecture, processor speed, external architecture, data and I/O access …
- Cluster architectures provide value, and require lots of interconnect
  - now the most common architecture for top 500 machines
    - http://www.top500.org/lists/2005/06/PerformanceDevelopment.php

adapted from Alan Benner, IBM

Component requirements

Computing systems require

- Receivers
  - Optical coupling, light guiding, detector, circuit
- Transmitters
  - Optical coupling, light guiding, modulator/source, circuit
- Filters
- Packaging and Interconnect strategy
- Source strategy
- Integration strategy

Point to point interconnection does not address issues such as:

Fast reconfigurability; bandwidth on demand, low latency

Ease of redeployment

Ease of upgradeability

Optical Waveguides for On-Board Links

- **Multimode waveguides:**
  - relaxed alignment tolerances
  - simple fabrication process
    → potential manufacturing cost efficiency.

- **Successful on-board integration can be improved by**
  - forming components in the guides
    - obtain further cost reduction
    - achieve increased functionality.
  - designing complex optical paths
    - minimise link lengths
    - enable advanced on-board topologies.

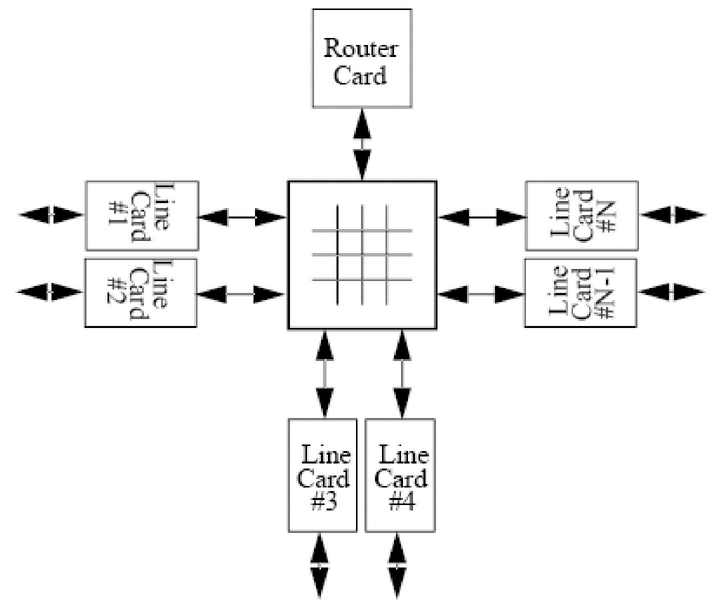
- **However, high-speed, on-board optical networks have stringent power budget requirements:**
  - low loss transmission
  - excellent crosstalk performance.

    (eg, the 10GbE standard only allows an 8 dB optical power budget.)

UNIVERSITY OF CAMBRIDGE

Issues with Electrical Switched Backplanes

- High throughput using a purely electrical backplane requires a very high performance electrical switch at its heart.
- Power dissipation is high – thermal management becomes a key concern.
- The backplane is not upgradeable in bit-rate without replacing the switch.
- High-bandwidth serial connections pose serious microwave engineering challenges at high bit rates, e.g. above 10 Gb/s
- Parallel electrical solutions require complex spatial routing



Schematic of a Fast Switched Backplane for a Gigabit Switched Router

- After Nick McKeown

Optical interconnects can **improve bandwidth-length products**, **eliminate electromagnetic interference effects** and **reduce thermal costs.**