An Overview:
The Next Generation of Ethernet

IEEE 802 Plenary
Atlanta, GA
November 12, 2007

IEEE 802.3 Higher Speed Study Group - TUTORIAL

Contributors

- Chris Cole, Finisar
- John D'Ambrosia, Force10 Networks
- Chris DiMinico, MC Communications
- Howard Frazier, Broadcom
- Adam Healey, LSI Corporation
- John Jaeger, Infinera
- Jack Jewell, JDSU
- Mark Nowell, Cisco
- Steve Trowbridge, Alcatel-Lucent

Agenda

- Introduction (John D'Ambrosia, HSSG Chair)
  - Objectives
  - Project Authorization Request
  - 5 Criteria
- Applications (Mark Nowell, Cisco)
- Architecture (Mark Nowell, Cisco)
- Fiber PMDs (Jack Jewell, JDSU)
  - MMF
  - SMF
- Cu PMDs (Chris DiMinico, MC Communications)
 - Cabling
  - Backplane
- Wrap-up

Introduction:
The Next Generation of Ethernet

John D'Ambrosia, Force10 Networks
Chair, IEEE 802.3 Higher Speed Study Group

The Ethernet Ecosystem

Broadband Access

Content Providers

Broadband
Access Networks

Internet
Backbone Networks

Content
Networks

Internet Backbone
Networks

Enterprise
Networks

Research
Networks

Internet eXchange and
Interconnection Points
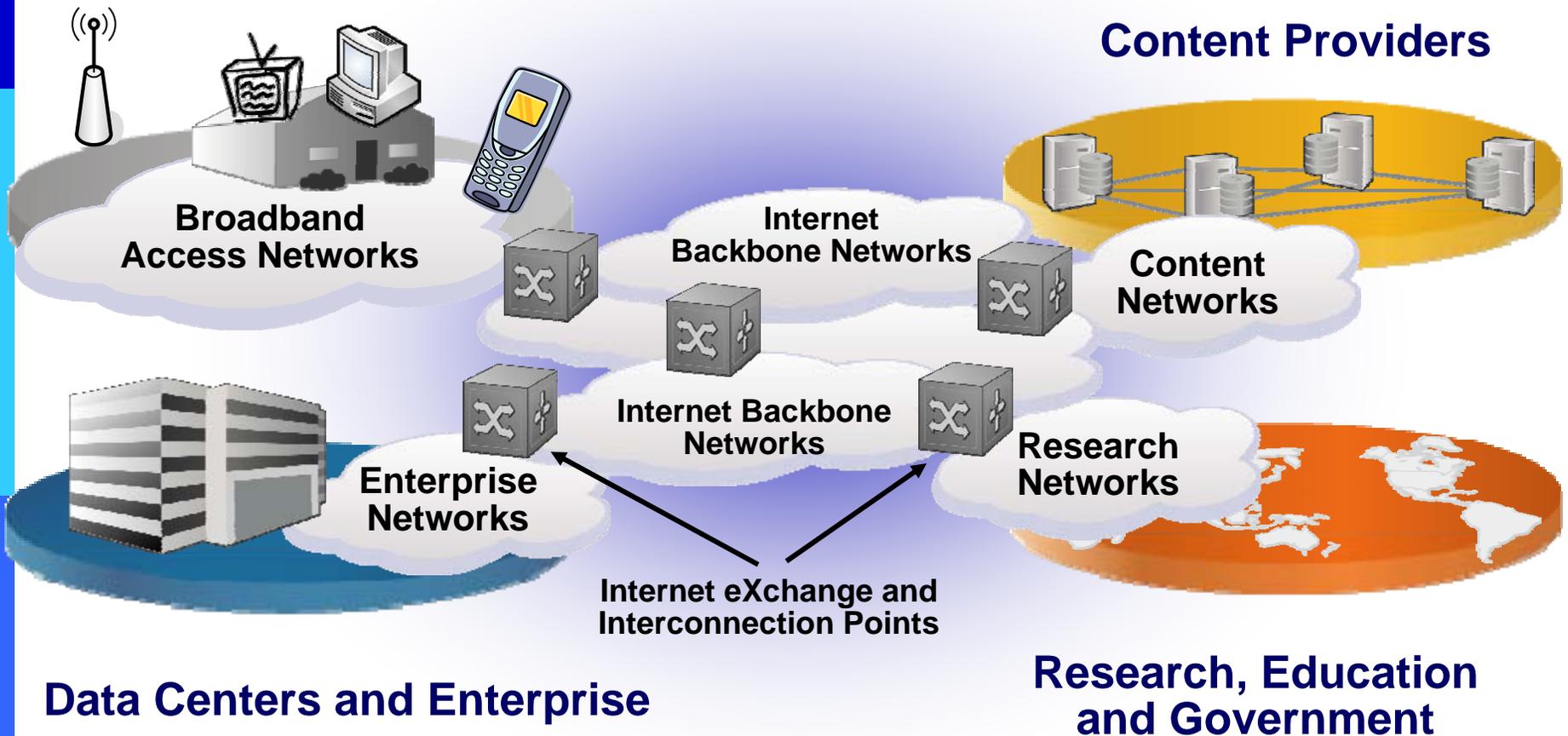
Data Centers and Enterprise

**Research, Education
and Government**

HSSG Objectives

- Support full-duplex operation only
- Preserve the 802.3 / Ethernet frame format utilizing the 802.3 MAC
- Preserve minimum and maximum FrameSize of current 802.3 standard
- Support a BER better than or equal to 10-12 at the MAC/PLS service interface
- Provide appropriate support for OTN
- Support a MAC data rate of 40 Gb/s
- Provide Physical Layer specifications which support 40 Gb/s operation over:
    - at least 100m on OM3 MMF
    - at least 10m over a copper cable assembly
    - at least 1m over a backplane
- Support a MAC data rate of 100 Gb/s
- Provide Physical Layer specifications which support 100 Gb/s operation over:
    - at least 40km on SMF
    - at least 10km on SMF
    - at least 100m on OM3 MMF
    - at least 10m over a copper cable assembly

Project Authorization Request
Key Elements (1 of 3)

- Project is an amendment to the IEEE 802.3 standard

- **Title:** Amendment: Media Access Control Parameters, Physical Layers and Management Parameters for 40 Gb/s and 100 Gb/s Operation

- **Target Completion Date: May 2010**

- **Scope:** Define 802.3 Media Access Control (MAC) parameters, physical layer specifications, and management parameters for the transfer of 802.3 format frames at 40 Gb/s and 100 Gb/s.

HSSG Project Authorization Request Key Elements (2 of 3)

- **Purpose:** The purpose of this project is to extend the 802.3 protocol to operating speeds of 40 Gb/s and 100 Gb/s in order to provide a significant increase in bandwidth while maintaining maximum compatibility with the installed base of 802.3 interfaces, previous investment in research and development, and principles of network operation and management. The project is to provide for the interconnection of equipment satisfying the distance requirements of the intended applications.

HSSG Project Authorization Request Key Elements (3 of 3)

- **Need:** The project is necessary to provide a solution for applications that have been demonstrated to need bandwidth beyond the existing capabilities. These include data center, internet exchanges, high performance computing and video-on-demand delivery.   Network aggregation and end-station bandwidth requirements are increasing at different rates, and is recognized by the definition of two distinct speeds to serve the appropriate applications.

Broad Market Potential (1 of 2)

- **Broad sets of applications**
- **Multiple vendors and numerous users**
- **Balanced cost (LAN versus attached stations)**

- **Bandwidth requirements for computing and core networking applications are growing at different rates, which necessitates the definition of two distinct data rates for the next generation of Ethernet networks in order to address these applications:**

  - Servers, high performance computing clusters, blade servers, storage area networks and network attached storage all currently make use of 1G and 10G Ethernet, with significant growth of 10G projected in '07 and '08. I/O bandwidth projections for server and computing applications indicate that there will be a significant market potential for a 40 Gb/s Ethernet interface.

  - Core networking applications have demonstrated the need for bandwidth beyond existing capabilities and the projected bandwidth requirements for computing applications. Switching, routing, and aggregation in data centers, internet exchanges and service provider peering points, and high bandwidth applications, such as video on demand and high performance computing environments, have demonstrated the need for a 100 Gb/s Ethernet interface.

Broad Market Potential (2 of 2)

- **Broad sets of applications**
- **Multiple vendors and numerous users**
- **Balanced cost (LAN versus attached stations)**

- **There has been wide attendance and participation in the study group by end users, equipment manufacturers and component suppliers. It is anticipated that there will be sufficient participation to effectively complete the standardization process.**

- **Prior experience scaling IEEE 802.3 and contributions to the study group indicates:**

  - **40 Gb/s Ethernet will provide approximately the same cost balance between the LAN and the attached stations as 10 Gb/s Ethernet.**

  - **The cost distribution between routers, switches, and the infrastructure remains acceptably balanced for 100 Gb/s Ethernet.**

- **Given the topologies of the networks and intended applications, early deployment will be driven by key aggregation & high-bandwidth interconnect points. This is unlike the higher volume end system application typical for 10/100/1000 Mb/s Ethernet, and as such, the initial volumes for 100 Gb/s Ethernet are anticipated to be more modest than the lower speeds. This does not imply a reduction in the need or value of 100 Gb/s Ethernet to address the stated applications.**

Compatibility

- IEEE 802 defines a family of standards. All standards shall be in conformance with the IEEE 802.1 Architecture, Management, and Interworking documents as follows: 802. Overview and Architecture, 802.1D, 802.1Q, and parts of 802.1f. If any variances in conformance emerge, they shall be thoroughly disclosed and reviewed with 802. Each standard in the IEEE 802 family of standards shall include a definition of managed objects that are compatible with systems management standards.

- **As an amendment to IEEE Std 802.3, the proposed project will remain in conformance with the IEEE 802 Overview and Architecture as well as the bridging standards IEEE Std 802.1D and IEEE Std 802.1Q.**

- **As an amendment to IEEE Std 802.3, the proposed project will follow the existing format and structure of IEEE 802.3 MIB definitions providing a protocol independent specification of managed objects (IEEE Std 802.1F).**

- **The proposed amendment will conform to the full-duplex operating mode of the IEEE 802.3 MAC.**

- **As was the case in previous IEEE 802.3 amendments, new physical layers specific to either 40 Gb/s or 100 Gb/s operation will be defined.**

- **By utilizing the existing IEEE 802.3 MAC protocol, this proposed amendment will maintain maximum compatibility with the installed base of Ethernet nodes.**

Distinct Identity

- **Substantially different from other IEEE 802 standards**
- **One unique solution per problem (not two solutions to a problem)**
- **Easy for the document reader to select the relevant specification**

- **The proposed amendment is an upgrade path for IEEE 802.3 users, based on the IEEE 802.3 MAC.**

- **The established benefits of the IEEE 802.3 MAC include:**
  – Deterministic, highly efficient full-duplex operation mode
  – Well-characterized and understood operating behavior
  – Broad base of expertise in suppliers and customers
  – Straightforward bridging between networks at different data rates

- **The Management Information Base (MIB) for IEEE 802.3 will be extended in a manner consistent with the IEEE 802.3 MIB for 10 / 100 / 1000 / 10000 Mb/s operation.**

- **The proposed amendment to the existing IEEE 802.3 standard will be formatted as a collection of new clauses, making it easy for the reader to select the relevant specification.**

- **Bandwidth requirements for computing and networking applications are growing at different rates. These applications have different cost / performance requirements, which necessitates two distinct data rates, 40 Gb/s and 100 Gb/s.**

Technical Feasibility

- **Demonstrated system feasibility**
- **Proven technology, reasonable testing**
- **Confidence in reliability**

- **The principle of scaling the IEEE 802.3 MAC to higher speeds has been well established by previous work within IEEE 802.3.**

- **The principle of building bridging equipment which performs rate adaptation between IEEE 802.3 networks operating at different speeds has been amply demonstrated by the broad set of product offerings that bridge between 10, 100, 1000, and 10000 Mb/s.**

- **Systems with an aggregate bandwidth of greater than or equal to 100 Gb/s have been demonstrated and deployed in operational networks.**

- **The proposed project will build on the array of Ethernet component and system design experience, and the broad knowledge base of Ethernet network operation.**
  - The experience gained in the development and deployment of 10 Gb/s technology is applicable to the development of specifications for components at higher speeds. For example, parallel transmission techniques allow reuse of 10 Gb/s technology and testing.
  - Component vendors have presented data on the feasibility of the necessary components for higher speed solutions. Proposals, which either leverage existing technologies or employ new technologies, have been provided.

- **The reliability of Ethernet components and systems can be projected in the target environments with a high degree of confidence. Presentations demonstrating this have been provided.**

Economic Feasibility

- Known cost factors, reliable data
- Reasonable cost for performance
- Consideration of installation costs

- The cost factors for Ethernet components and systems are well known. The proposed project may introduce new cost factors which can be quantified.

- Presentations indicate that for the server market and computing applications the optimized rate to provide the best balance of performance and cost is 40 Gb/s. For the network aggregation market and core networking applications, the optimized rate offering the best balance of performance and cost is 100 Gb/s.

- In consideration of installation costs, the project is expected to use proven and familiar media, including optical fiber, backplanes, and copper cabling technology.

- Network design, installation and maintenance costs are minimized by preserving network architecture, management, and software.

Possible Timeline

TASK FORCE Formed

HSSG Formed

PAR Approved

TF Review

WG Ballot

Sponsor Ballot

STD!!!

| 2006 | 2007 | 2008 | 2009 | 2010 |

JUL S E P N O V J A N M A R M A Y J U L S E P N O V J A N M A R M A Y J U L S E P N O V J A N M A R M A Y J U L S E P N O V J A N M A R M A Y J U L S E P N O V

D1 D2 D3

■ Plenary

◆ Interim
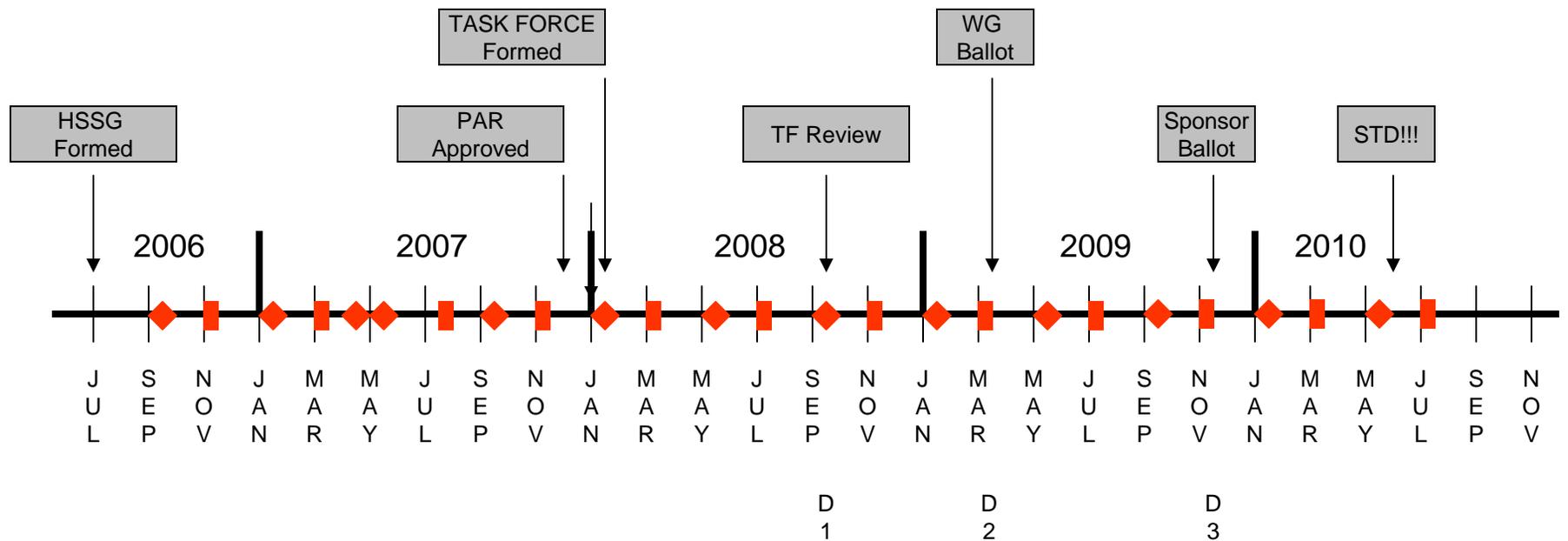
Applications for the Next Generation of Ethernet

Howard Frazier, Broadcom

Mark Nowell, Cisco

Steve Trowbridge, Alcatel-Lucent

Why Higher Speed Ethernet?

Fundamental bottlenecks are happening everywhere

| Increased # of users | + | Increased access rates and methods | + | Increased services | = | **Bandwidth explosion everywhere** |

Increased # of users	Increased access rates and methods	Increased services
As demonstrated by the number of ISPs: Comcast, AOL, YahooBB, NTT, Cox, EasyNet, Rogers, BT, ...	EFM, xDSL, WiMax, xPON, Cable, WiFi, 3G/4G...	YouTube, BitTorrent, VOD, Facebook, Kazaa, Netflix, iTunes, 2nd life, Gaming...

Higher Speed Ethernet Penetration

Campus/ Enterprise

Data Center

Service Provider

Core

Core

Core

Distribution

Aggregation

Wiring Closet

Access

Distribution

Clients

Servers

Residential

IEEE 802.3 Higher Speed Study Group - TUTORIAL

19
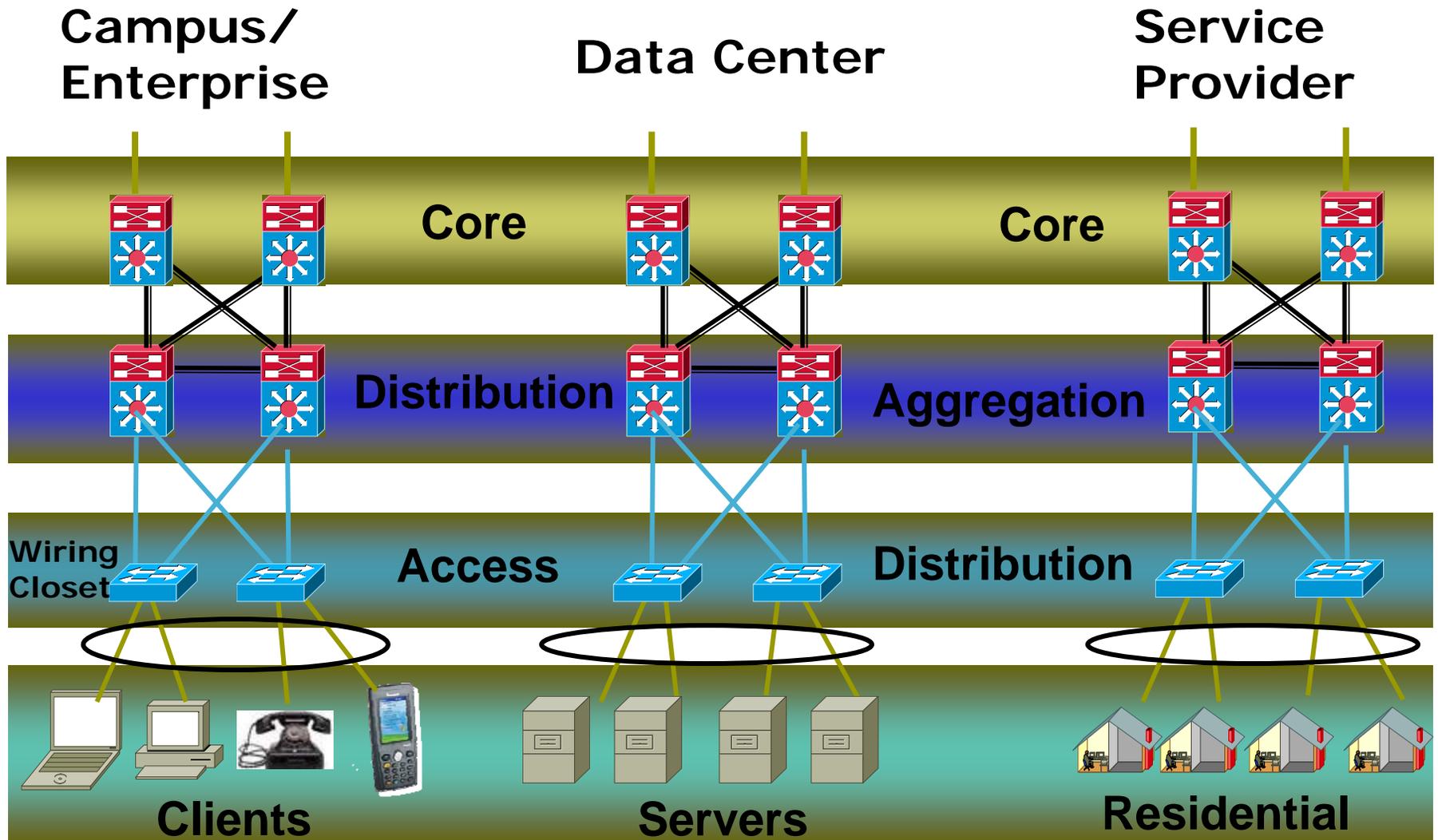
Example of Historical Correlation Between Speeds and Ports Growth
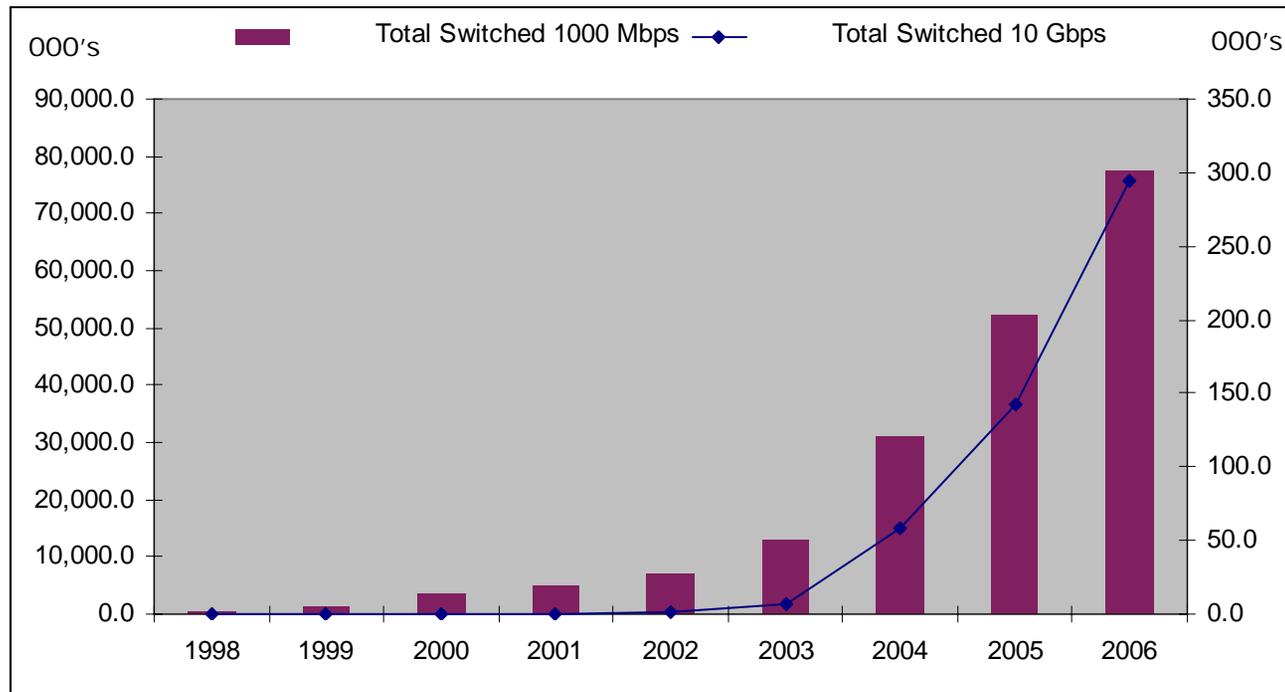
Source: Cisco (barbieri_01_0107.pdf)

- Between 2003 and 2006, GbE growth and 10GbE growth were correlated. Symbiotic relationship.

- 2007: 10GbE growth being constrained by lack of higher speed interface (Sources: Sprint, Yahoo, EDS, Amazon, AMS-IX, Cox, NTT, Equinix)

What is Driving Industry Needs?

- Computing: (system throughput doubles approximately every 2 yrs)
  - Ref:
    www.ieee802.org/3/hssg/public/jan07/muller_01_0107.pdf

- Networking: driven by the aggregation of data from multiple computing platforms
  - Number of computing platforms growing – multiplicative effect on networking

# 40GbE and 100GbE: Computing and Networking
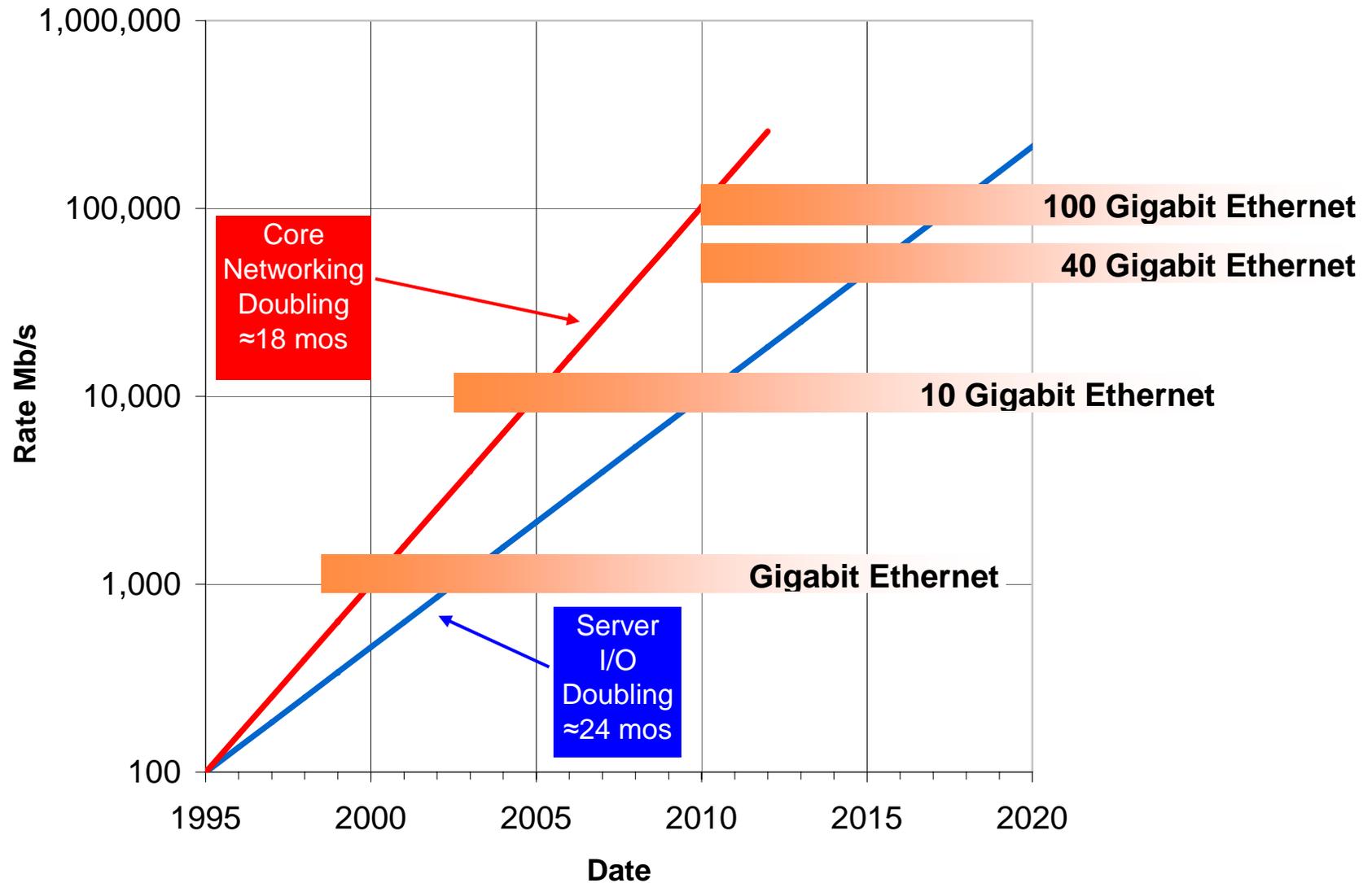
High Speed Ethernet

## Broad Market Applications

Market Drivers for More Bandwidth

Consumer &
Broadband
Access

Content
Providers

Server
Virtualization

Video on
demand

Blade Server

Service
Providers &
IXCs

Fabric
Consolidation

High Performance
Computing

Data Center

Research and
development

Higher Speed Ethernet

Video Example

Recommendation

List of favorites

Multi-channel broadcasting TV menu

Confirmation and change of service plan

VoD menu

Karaoke menu

Help

Start button of the introduction movie of newly arrived contents

Courtesy KDDI

- How many would go back to Black & White TV?
- Today's change to HDTV resolutions is minor compared to TV's change to interactivity via DVRs, iTV & VoD

Broadband Access

Broadband access
Fixed access traffic growth

Global subscriber access traffic

IPTV Assumption:
50 Million IPTV subscribers in 2011
4 hours/day – 5 Mbps (from 2007)

(y-axis: TB per Year; values: 0, 50 000 000, 100 000 000, 150 000 000, 200 000 000, 250 000 000, 300 000 000)

IPTV

Internet

(x-axis: 2005, 2006, 2007, 2008, 2009, 2010, 2011)
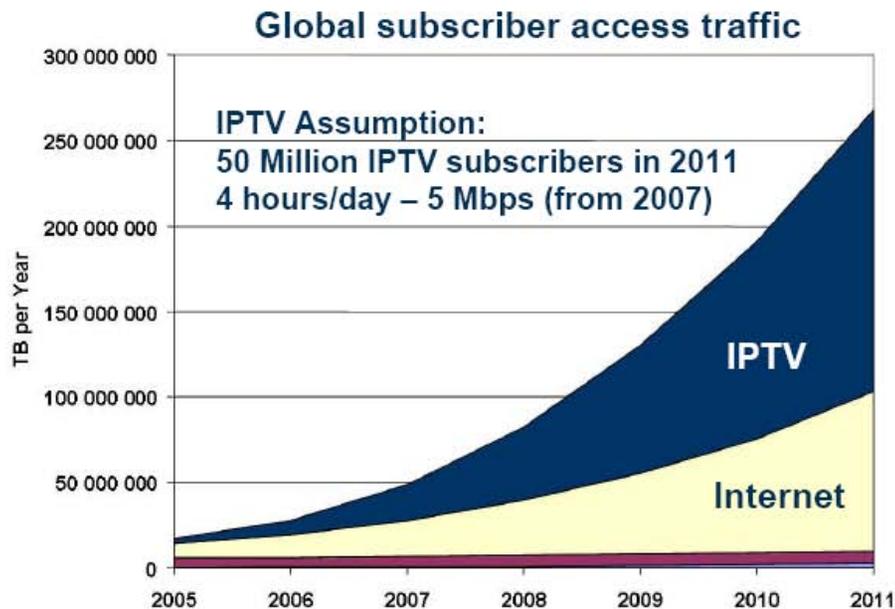
Access traffic growth through increased:

- broadband penetration
- bandwidth demanding services

Service	Bandwidth
VoIP	100 kbps
IMS/Video conf	0.7 - 1.5 Mbps
Internet	0.2 - 5.0 Mbps
Gaming	0.2 - 0.5 Mbps
SDTV (MPEG-2)	6 Mbps
SDTV (MPEG-4)	3 Mbps
HDTV (MPEG-2)	20 Mbps
HDTV (MPEG-4)	10 Mbps

Typical service portfolio (2010):
1 HDTV, 2 SDTV, gaming, voice, high-speed internet → 25 – 30 Mbps (MPEG-4)

Reference: Alping_01_1106.pdf

Massive Data Center – Server Virtualization

Today

Next

Nx10GbE

10GbE

GbE

Nx100GbE

100GbE

10GbE
&
40GbE

1000's of servers

IEEE 802.3 Higher Speed Study Group - TUTORIAL

27
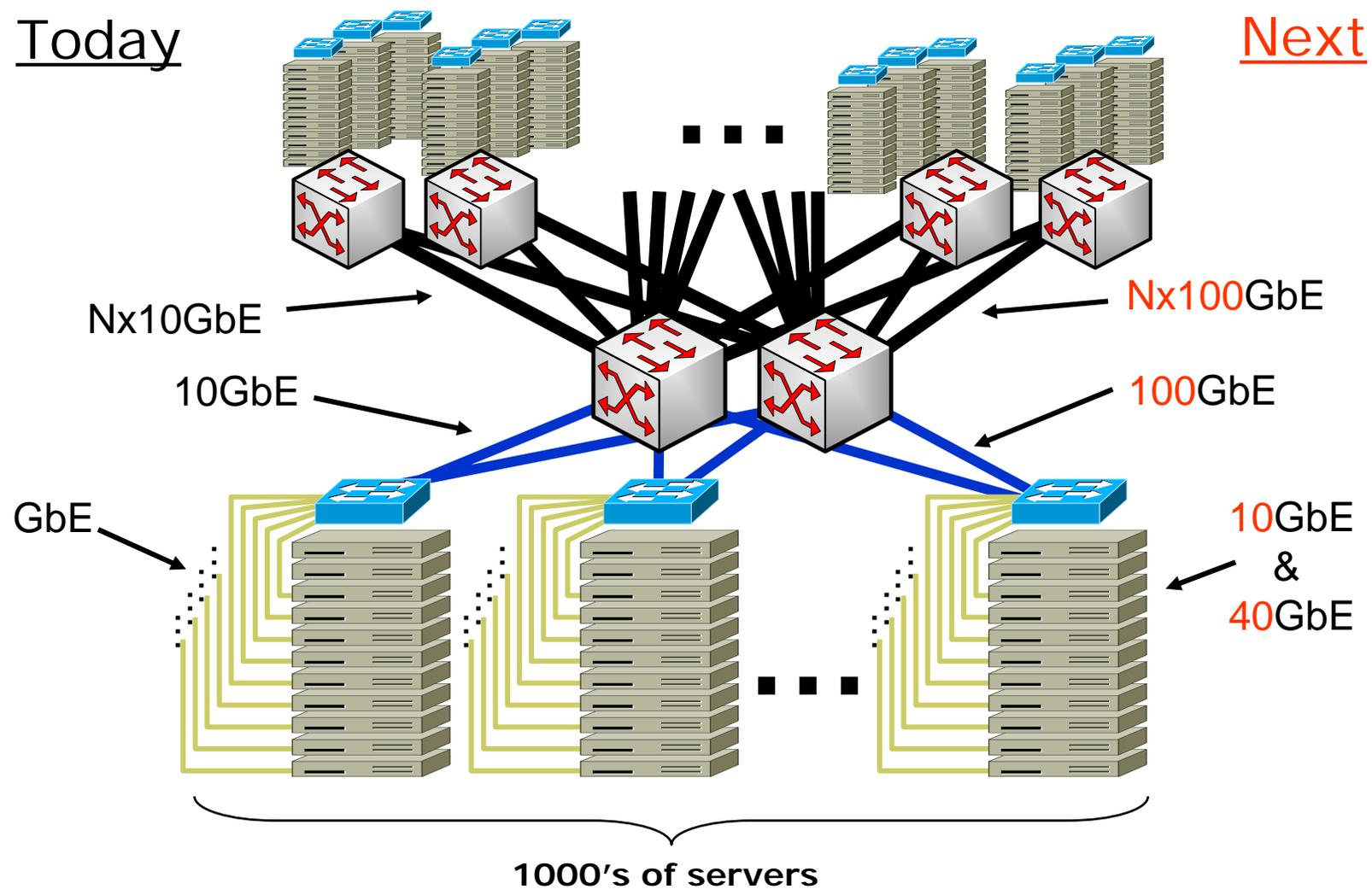
Vision: A Single, Flexible Ethernet Server I/O

Separate LAN & SAN Fabrics (Today)

SAN

LAN

FC Switch

Ethernet Switch

HBA

NIC

Server

Ethernet LAN & SAN Fabric (Tomorrow)

SAN

LAN

Ethernet Switch

NIC
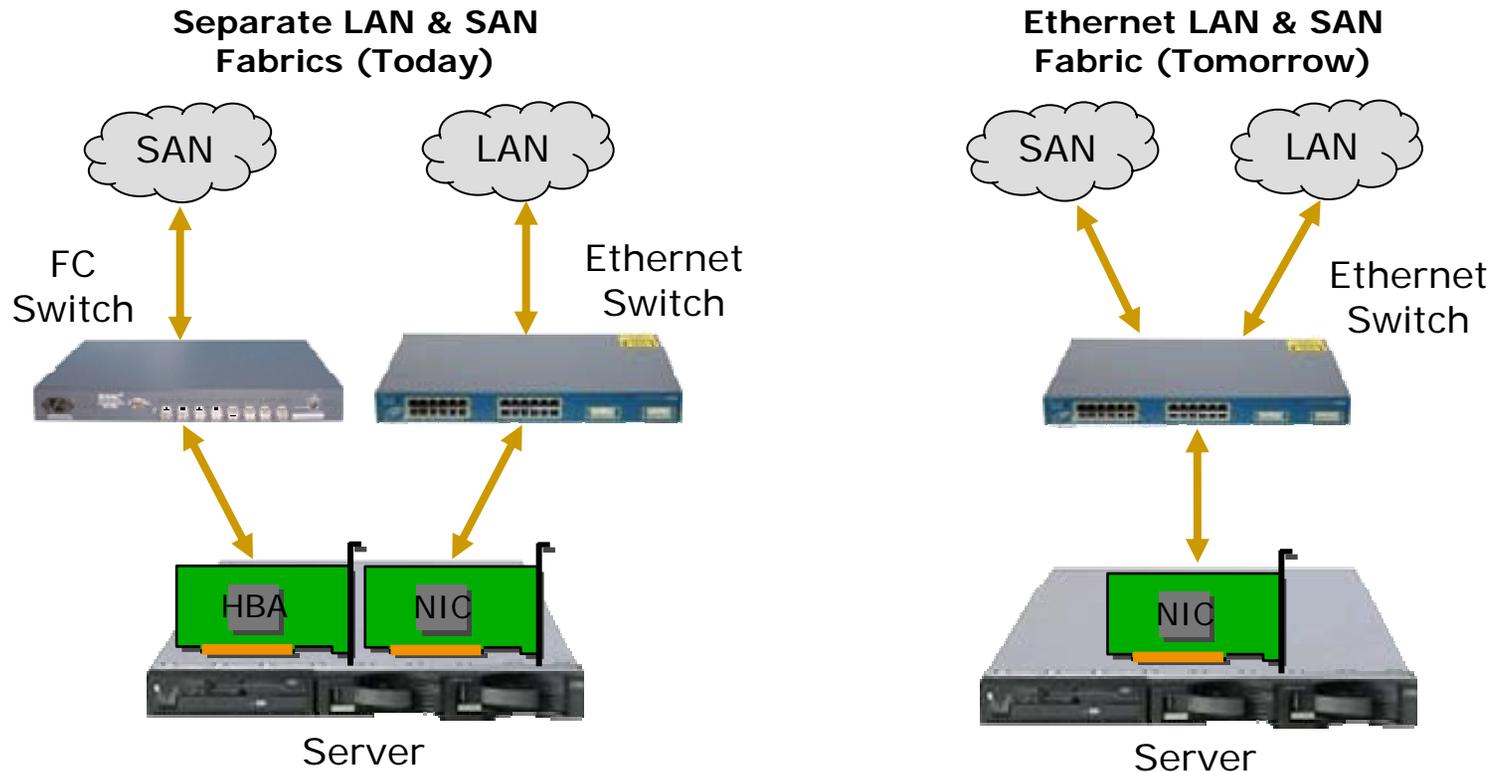
Server

- **Goal:** Reduce the cost, size, power, and complexity of servers by eliminating the need for multiple I/O connections & fabrics. Plug and play simplicity.

from http://www.ieee802.org/3/hssg/public/apr07/hays_01_0407.pdf

40G Market Potential for Server Networking
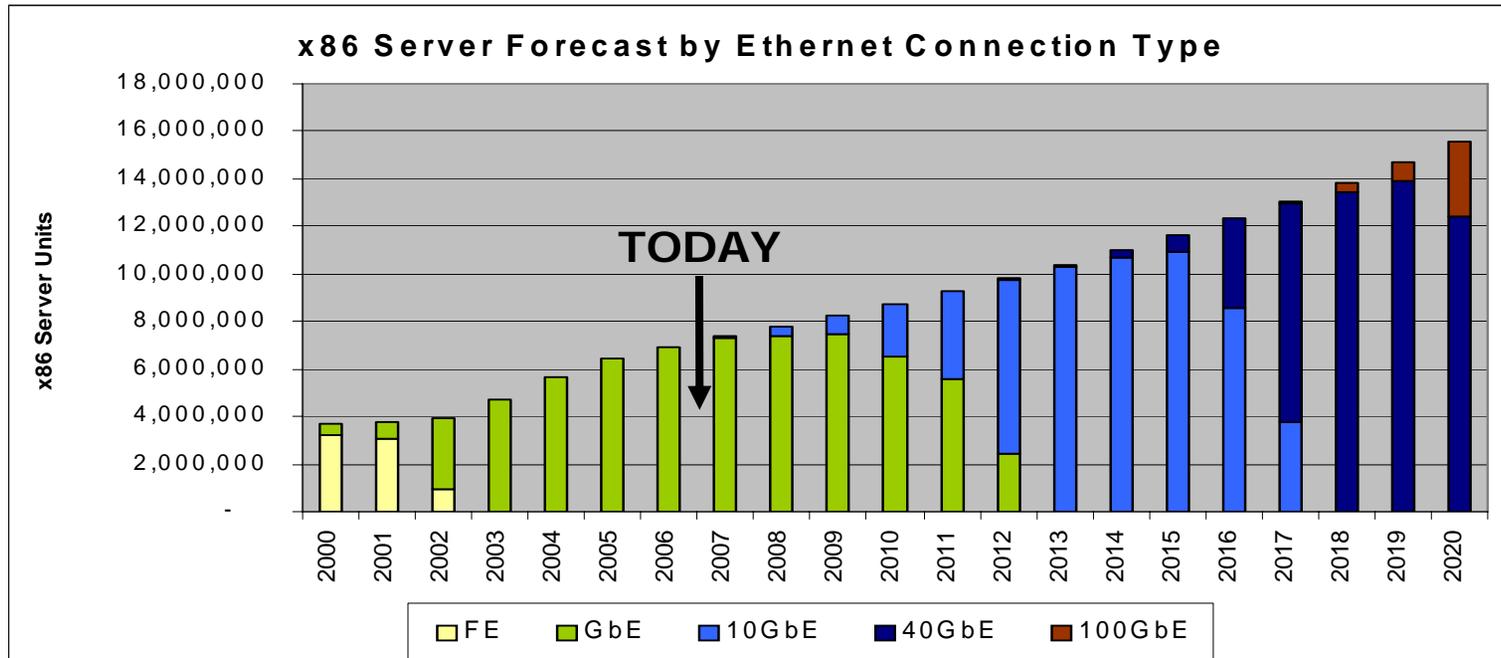
updated:
July 2007

- **>200M port** opportunity for 40GbE in Servers & Switches
  - 40GbE server connections drive demand for 100GbE switch uplinks
  - Without 40GbE, higher-speed IB/FC/PCI-E may displace 10GbE before 100GbE is cost-effective, potentially reducing the long-term Ethernet TAM in the data center
- Drives **>26M port** 100GbE uplink opportunity

IEEE 802.3 Higher Speed Study Group - TUTORIAL

Server Form Factors & Physical Networking

Network

100GbE

40GbE

Network

100GbE

40GbE

Network

100GbE

40GbE
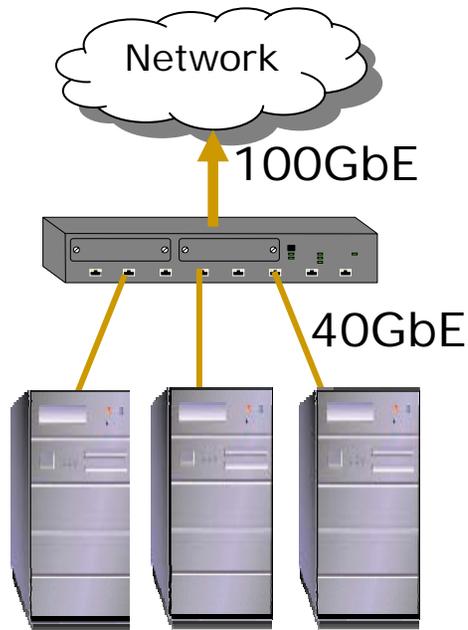
Pedestal Server:
- <100m to switch across data center
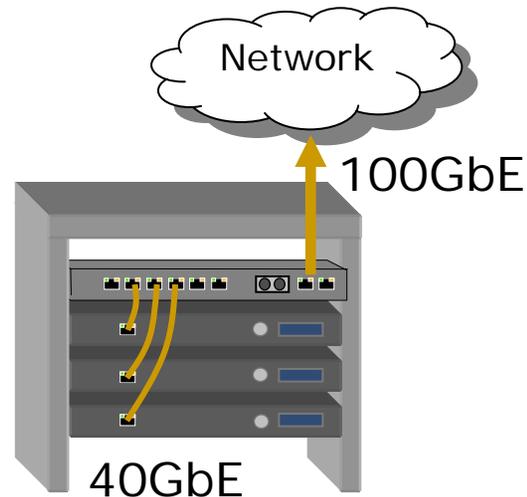
Rack Server:
- <10m to switch at top-of-rack or in-row
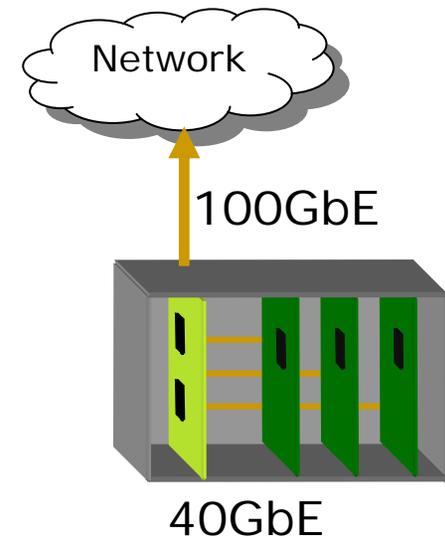
Blade Server:
- <1m to switch in back of chassis

from http://www.ieee802.org/3/hssg/public/apr07/hays_02_0407.pdf

IEEE 802.3 Higher Speed Study Group - TUTORIAL

Conclusions

- **Numerous applications are driving the need for higher speed Ethernet**

- **Bandwidth requirements for computing and network aggregation applications are growing at different rates**

- **The next generation of Ethernet requires both 40Gb/s and 100Gb/s to address these applications**

HSSG Tutorial
# MAC / PHY
Architecture

John Jaeger, Infinera

Overview – 40G & 100G MAC & PHY

LLC
MAC
Reconciliation
PCS
PMA
PMD
Medium
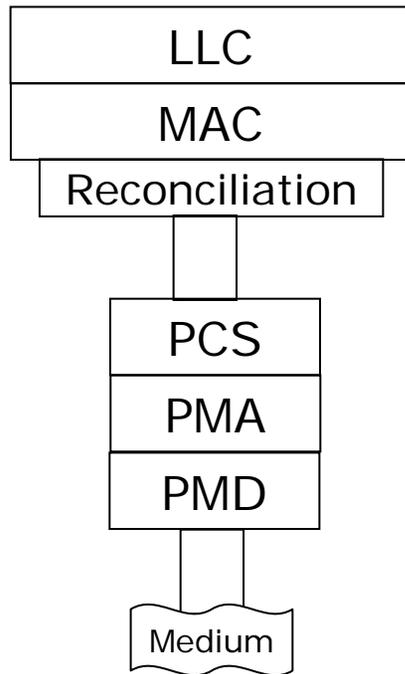
Generalized LAN
CSMA/CD Layers

- Consistent with previous Ethernet rates, extension to 40Gb/s & 100Gb/s data rates
  - Frame format; Services; Management attributes
- MAC
  - No changes to the MAC operation
  - Technical feasibility established:
    - CRC checker, general MAC functions, gear-boxes, host interfaces
    - No issues or concerns identified
- PCS
  - Specified PCS(s) need to accommodate:
    - 40G backplane PHY; 40 & 100G copper cable PHYs; 40 & 100G MMF PHYs and 100G SMF PHYs
  - Commonality, leveraging existing 10G technology, and working with the specified multi-channel/cable/fiber/wavelength PMDs will be examined
  - Two example approaches reviewed (next slides)

"CTBI" Technical Overview

Tx PCS   Tx PMA/PMD   Rx PMA/PMD   Rx PCS

| 100G 64B66B Encode | 100G X^58 Scrm | CTBI I/F (Inv Mux) | 10:n | ... | n:10 | CTBI I/F (deskew, reorder, mux) | 100G X^58 De-scrm | 100G 64B66B Decode |

n lanes
(fiber or wavelengths)
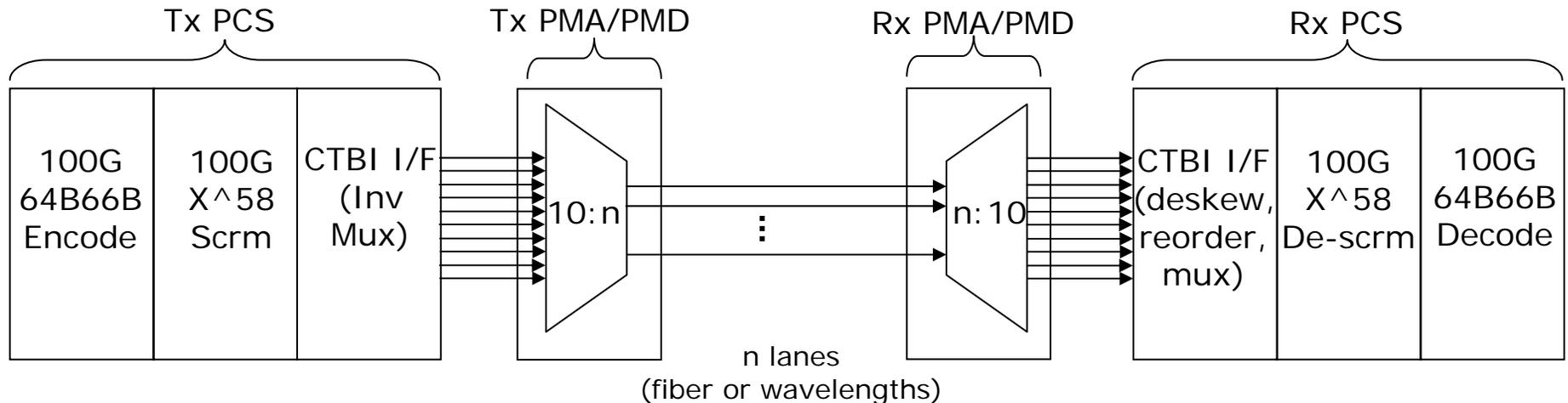
- Standard 64B/66B PCS (running up to 10x faster)
- 40G/100G Example: 4/10 Lane Electrical PCS to PMA/PMD Interface (CTBI)
  - 64B/66B aggregate is inverse mux'ed to Virtual Lanes
  - A unique identifier is added to each Virtual Lane on a periodic basis
  - Virtual lanes are bit mapped to/from the 4/10 CTBI electrical lanes
  - Virtual lane alignment and skew compensation is done in Rx PCS only
- PMA maps m lane CTBI to n lane PMD
  - PMA is simple bit level muxing and demuxing
  - no realignment in PMA (for either electrical or optical skew)
- PCS and Virtual Lane overhead is very low, and independent of frame size

"APL" Technical Overview

Existing 10GBASE PHY

Multi-wavelength PHY Example (nxXAUI)
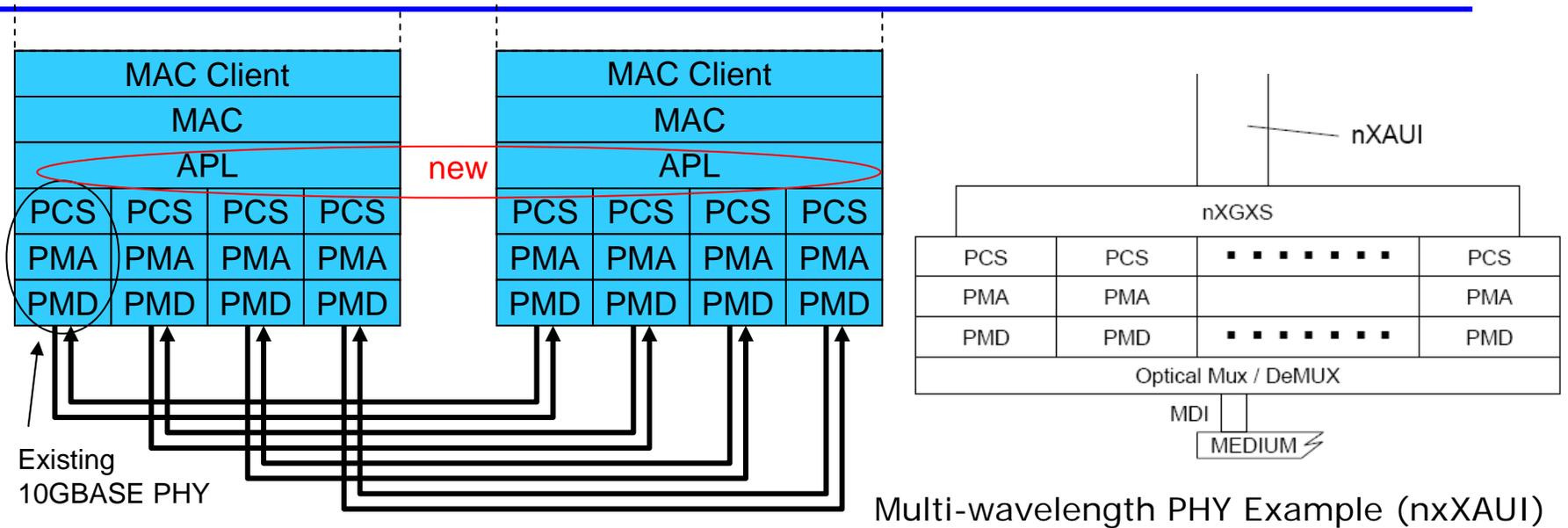
- The PME aggregation concept from 802.3ah can be used with existing 10GBASE PHYs – Aggregation at the Physical Layer (APL)
- Variety of fragment format considerations discussed (header, size, fragment CRC…)
- APL:
  - Assumes equal speed links, point-to-point, & full duplex links
  - Resilient and scalable
  - Ensures ordered delivery and detects lost or corrupted fragments
  - Minimal added latency
  - Fits well with multi-port (quad/octal) PHYs

IEEE 802.3 Higher Speed Study Group - TUTORIAL

Providing Support for OTN

Ethernet Ecosystem

10 Gbps / λ DWDM

Broadband Access Networks

Internet Backbone Networks

Content Networks

Internet Backbone Networks

Enterprise Networks

Research Networks

40 Gbps / λ DWDM

**Future –
100 Gbps / λ DWDM**
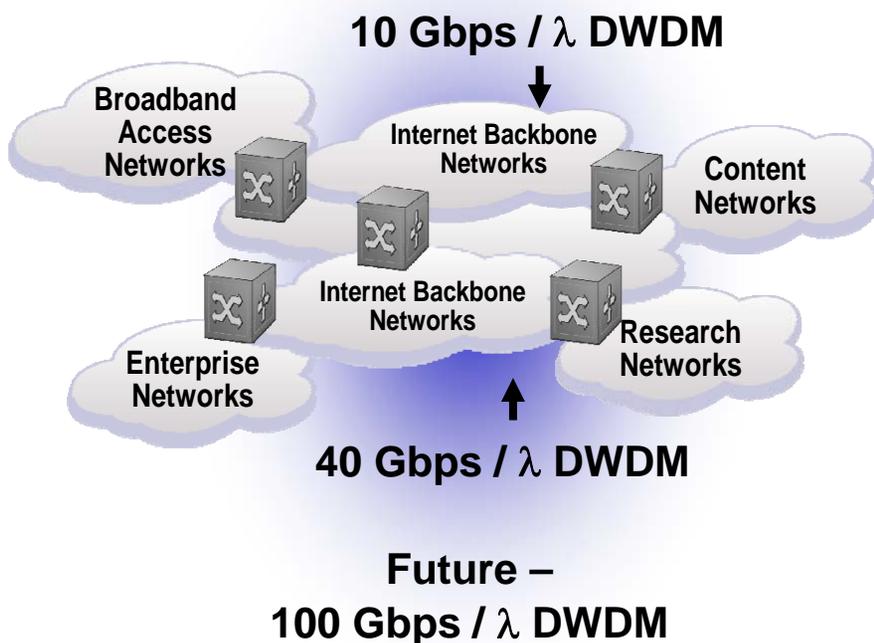
- OTN Networks
  - 10 Gbps / λ DWDM
  - 40 Gbps / λ DWDM
  - **100 Gbps transport network under development; work in progress @ ITU-T to accommodate 100GbE**

- Issues for Next Gen Ethernet
  - if 40GbE: line rate > Payload of standard ODU3

Purpose of objective - ensure ability to appropriately transport 40GbE and 100GbE
- HSSG has identified potential solutions for transparent 40GbE over ODU3
- Transparent 100GbE mapping into ODU4 will be defined by ITU-T

Physical Layer Specifications to be Defined

	40GbE	100GbE
At least 1m backplane	✓	
At least 10m cu cable	✓	✓
At least 100m OM3 MMF	✓	✓
At least 10km SMF		✓
At least 40km SMF		✓

HSSG Tutorial
MMF/SMF PMDs

Jack Jewell – JDSU
Chris Cole – Finisar

Physical Layer Specifications to be Defined

	40GbE	100GbE
At least 1m backplane	✔	
At least 10m cu cable	✔	✔
At least 100m OM3 MMF	✔	✔
At least 10km SMF		✔
At least 40km SMF		✔

MMF Overview

- Objective: At least 100m on OM3 MMF (40GbE and 100GbE)
- Platform: Parallel fibers; ~10Gbps/ch; 850nm VCSEL arrays
  - Combines existing 10GbE serial and 12x2.7G (or (4+4)x2.5G) parallel product technologies - both are technically sound and economical
  - Considering 10x (4x)10.3Gbaud (64/66), 12x10Gbaud (8B/10B), etc
  - Considering WDM (e.g. 2 λ's) to reduce fiber count/cost
- Advantages:
  - Low power for ~100m reach - <3W for 100GbE and <1.5W for 40GbE
  - Small footprint for high-density interconnects
  - Low cost
- Early Demo: 12x10GbE; >300m by IBM/Picolight; OFC 2003
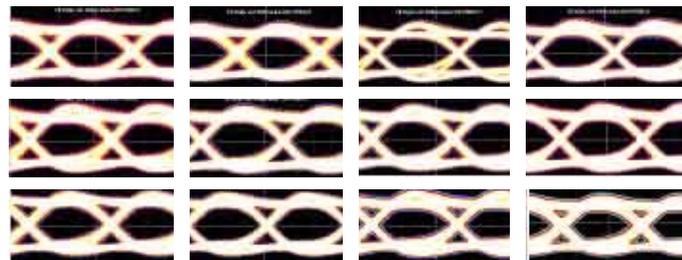
12 eyes, each 10Gbps, from a SNAP12 Parallel Module

IEEE 802.3 Higher Speed Study Group - TUTORIAL

100G and 40G Form Factors

100G Proven Technologies

10 Gb/s SFP+, XFP XENPAK

12x2.7Gb/s SNAP12

10x10Gb/s SNAP12

10 Gb/s SFP+, XFP XENPAK
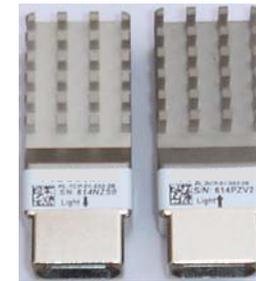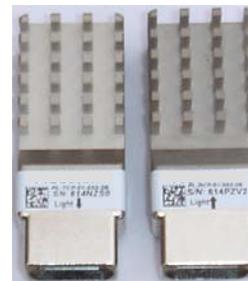
40G Proven Technologies

(4+4)x2.5 Gb/s POP4

(4+4)x10 Gb/s POP4

QSFP
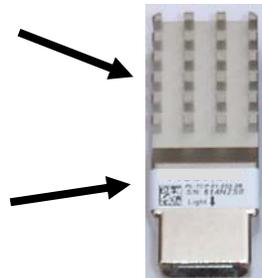
Technical Feasibility - Array Reliability

Wearout Time (Depends on uniformity)
- **Perfect uniformity → 12x lifetime same as 1x**
- **<u>EXAMPLE:</u> 12x lifetime ~1/2 as long as 1x**
 **4x lifetime ~3/4 as long as 1x**

Random Failures
- **Virtually all "random" failures are eliminated through burn-in**

ESD-Related Failures
- **Above-threshold ESD events damage 1x, 4x and 12x about equally**

Non-Hermetic Packaging
- **12x3G VCSELs robust to harsh environments**

Power @ 7mA (mW) 85C, 85% Humidity

Time (hours)

Delta Threshold

Failure Criterion

0   Time   1/2   3/4   1
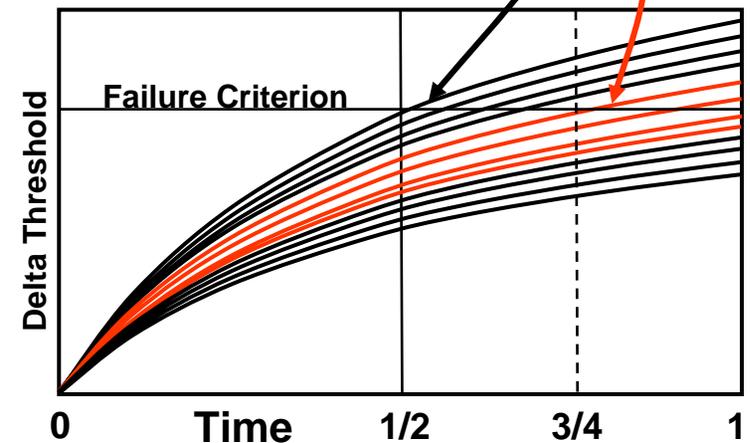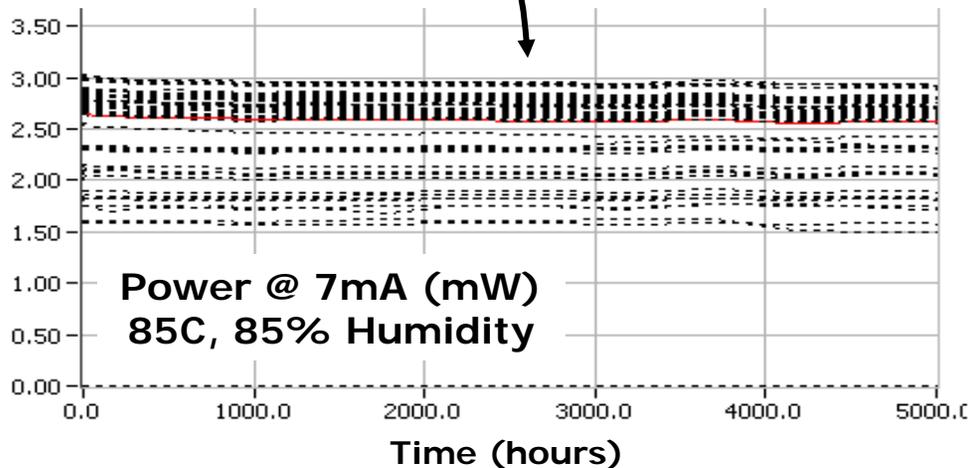
# Economic Feasibility - Array Yield

10x10G and 4x10G VCSEL array yields are cost feasible

- **VCSEL cost is only a small portion of a module cost**
- **Areal-dependent performance yield:**

    Penalty is slight for 12x, very slight for 4x
- **Random (microscale) yield:**

    Fallout is very small (<1%) → penalty is small
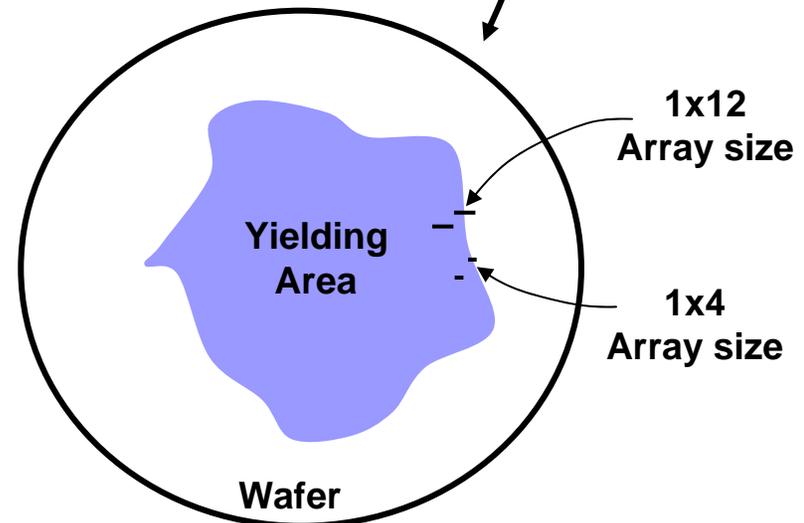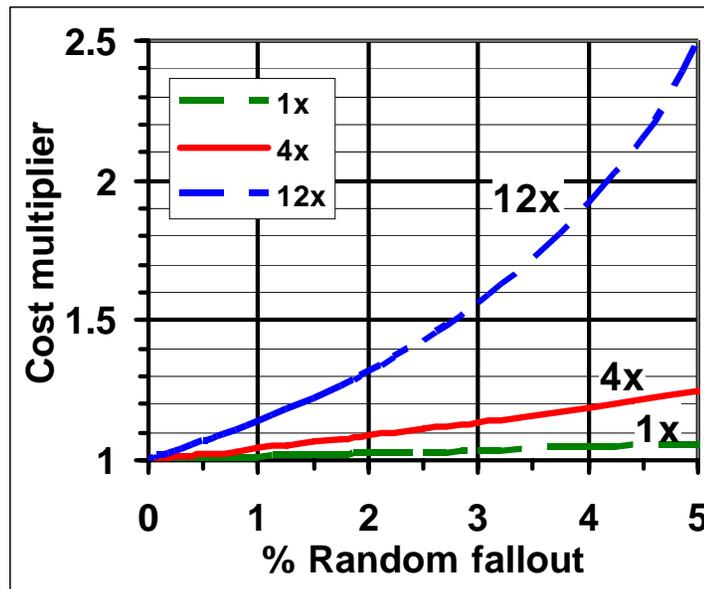
Array yield will not be a significant cost factor

Physical Layer Specifications to be Defined

	40GbE	100GbE
At least 1m backplane	✓	
At least 10m cu cable	✓	✓
At least 100m OM3 MMF	✓	✓
At least 10km SMF		✓
At least 40km SMF		✓

SMF PMD Technical Alternatives

SMF Technology	10km 1310nm	40km 1310nm	10km 1550nm	40km 1550nm
10x10G DML		OA	**OA** clairardin_01_0107 hartman_01_0107	**OA** schrans_01_0107 tsumura_foah_1206
10x10G EAML		OA	**OA** martin_01_0307	**OA** jaeger_01_0107
4x25G / 5x20G DML	cole_01_1106 traverso_02_0307	**OA** cole_01_0407		OA + DC
4x25G / 5x20G EAML	cole_01_0307 jiang_01_0407 traverso_02_0307	**OA** cole_01_0507 jiang_01_0507 traverso_01_0307		OA + DC
2x50G DQPSK MZML	schell_foah_0207 takeda_01_0107	OA + DC	DC	**OA + DC** duelk_01_0107 takeda_01_0907
1x100G NRZ MZML	schell_foah_0207	OA + DC	DC	**OA + DC** fisher_01_0107

Green shading designates alternatives studied by HSSG contributors.

10x10G 1550nm CWDM DML Transceiver
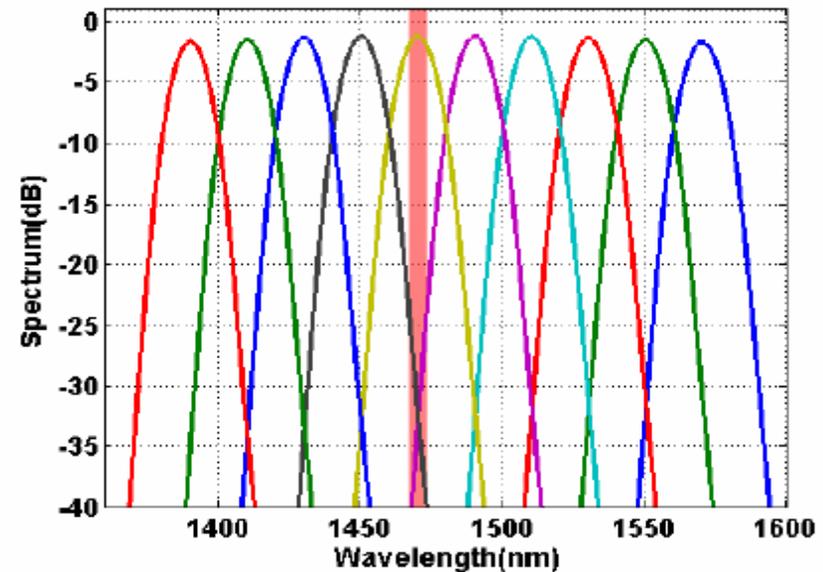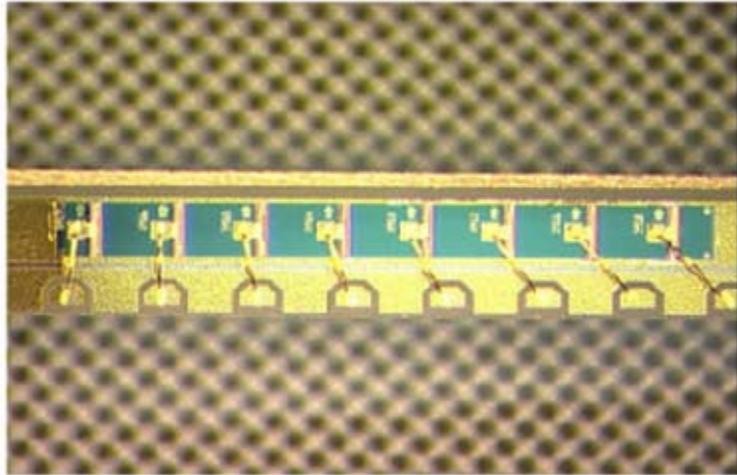
- 8x10G 1550nm DFB array showing 100G Transceiver feasibility; clairardin_01_0107
- 8x10G DFB array (left graphic: CyOptics,) Mux (right graphic: Kotura)

# 4x25G 1310nm EAML 10km Transceiver
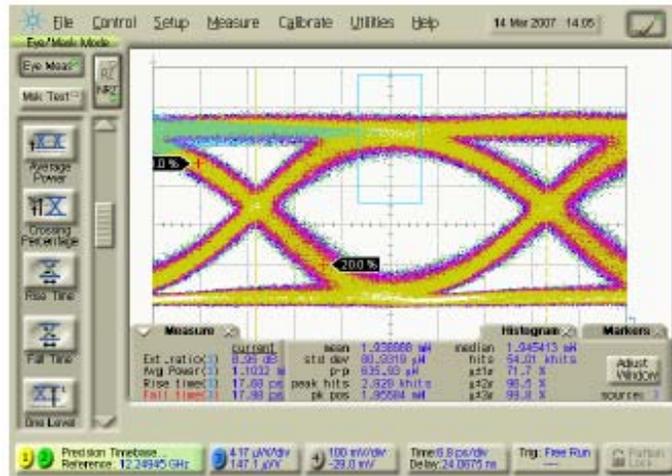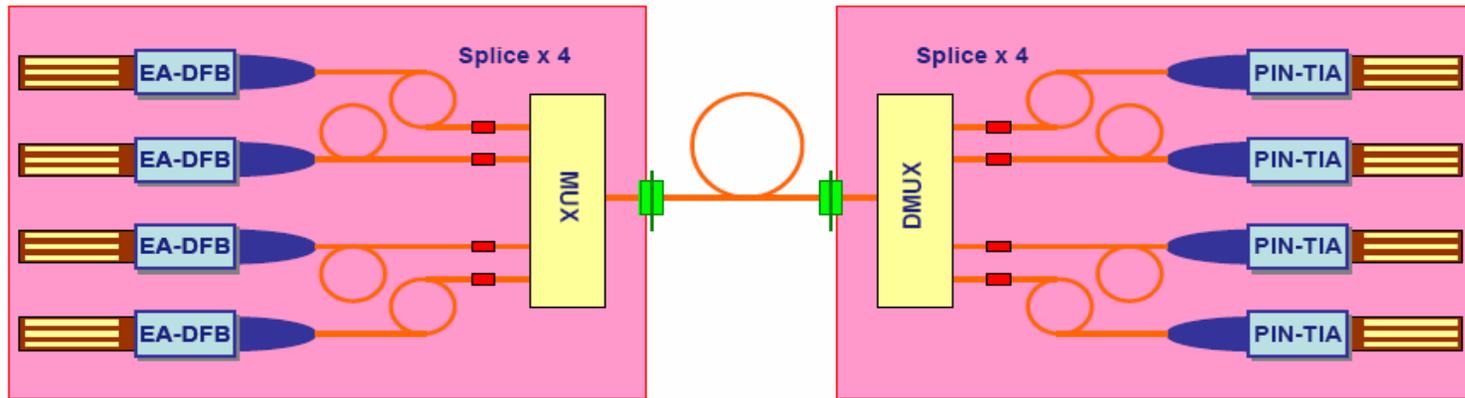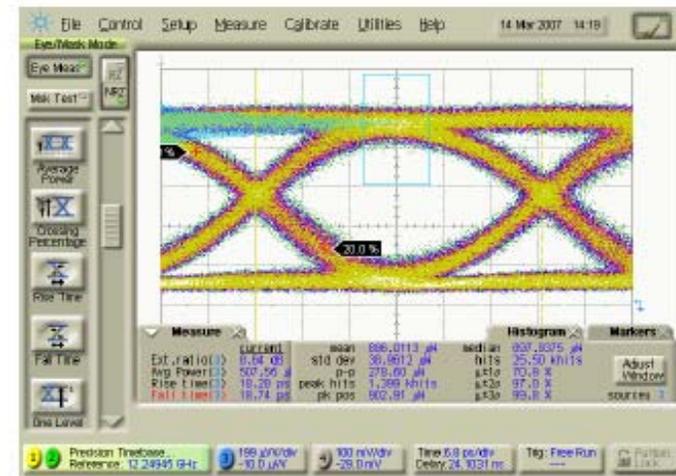
1m   10km

- 4x25G EAML Transceiver (upper graphic: Opnext); traverso_02_0307
- 25G EAML waveforms showing 100G Transceiver feasibility (lower graphic: JDSU); jiang_01_0407

40km 4x25G 1310nm EAML SOA Transceiver
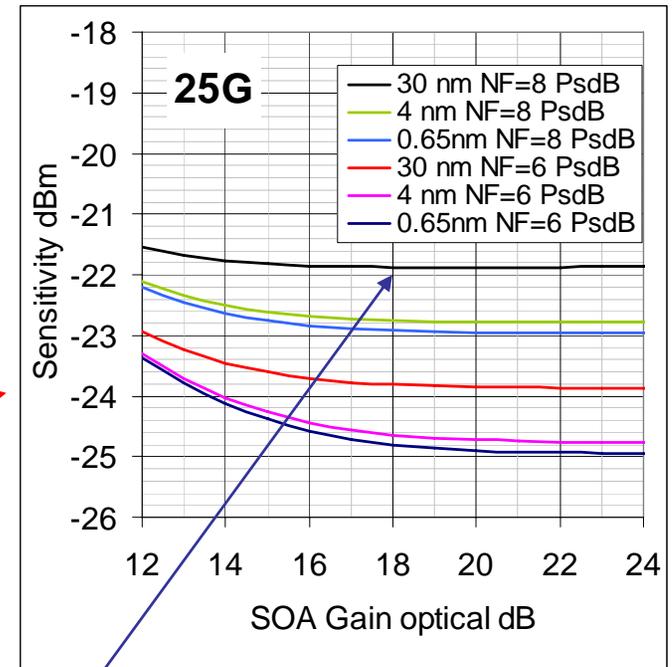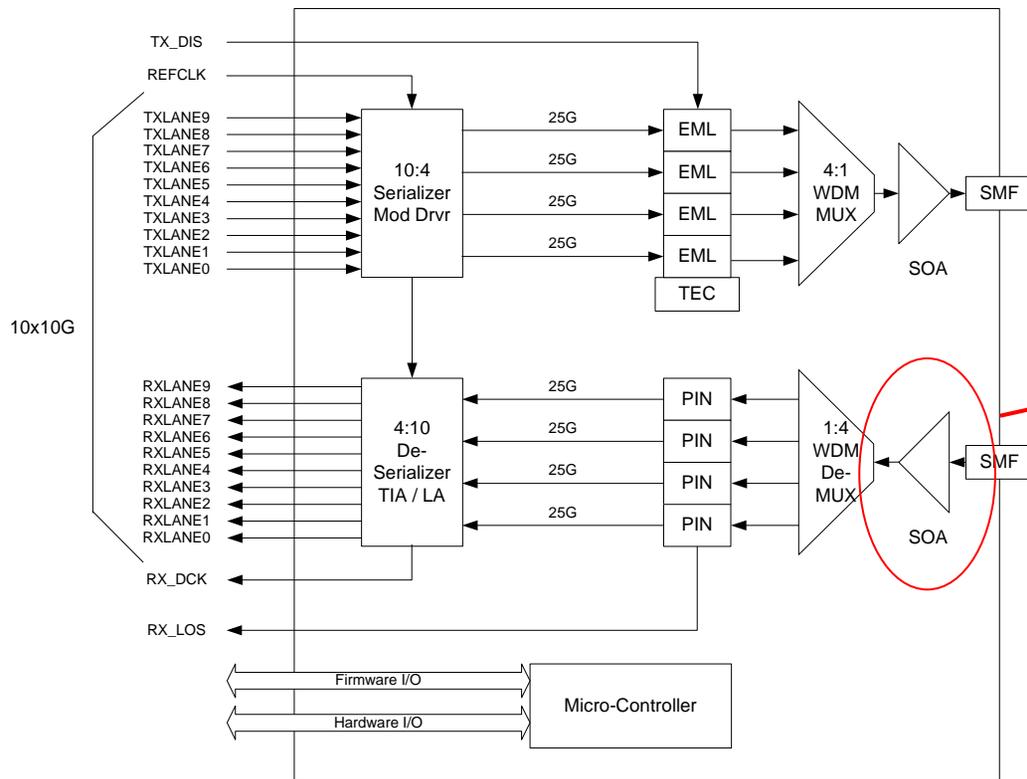
-22dBm SOA Receiver Sensitivity analysis showing 40km 100G Transceiver feasibility.

- 4x25G EAML + Optical Amplifier Transceiver (Finisar); cole_01_0507

10x10G 1550nm DWDM EAML Transceiver

10-λ DWDM mux
10-λ DWDM de-mux
10x10G modulators
10xDWDM lasers
10x10G receivers
100Gb/s Tx PIC
100 Gb/s Rx PIC

Modulator Driver Array
Transmitter PIC Chip
100Gb/s Tx Module
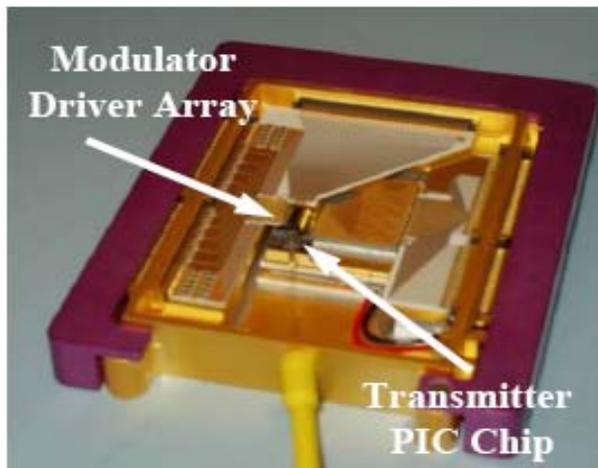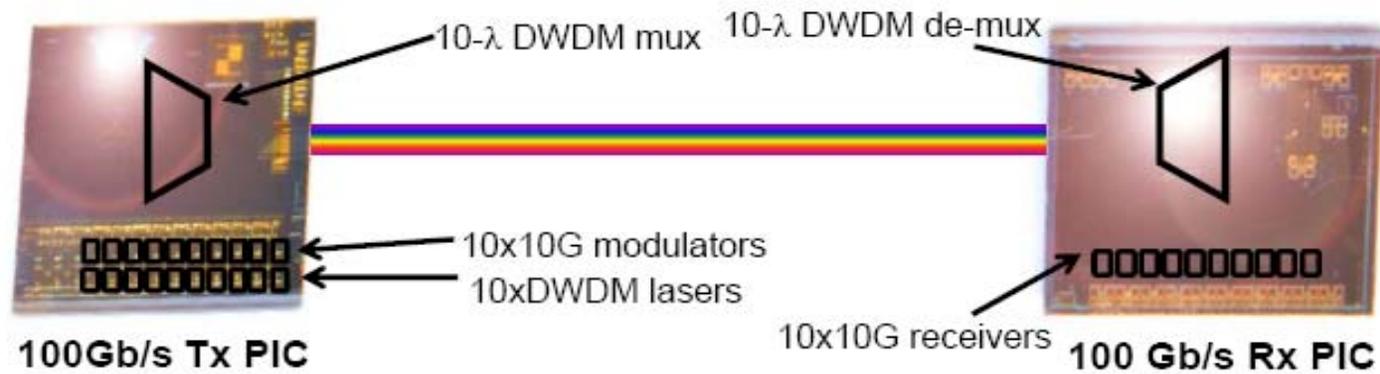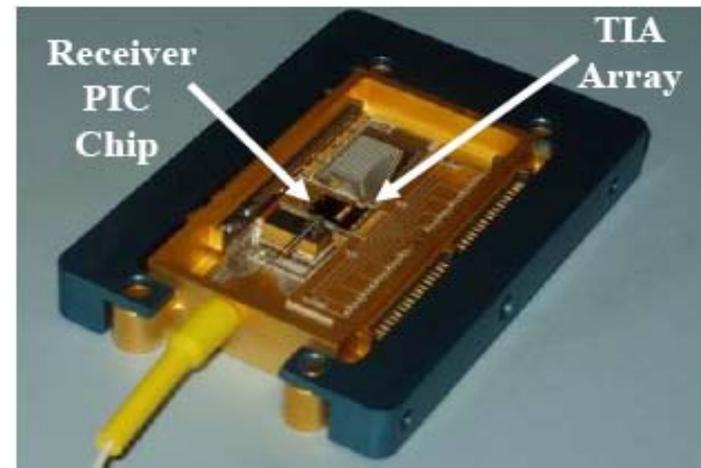
Receiver PIC Chip
TIA Array
100 Gb/s Rx Module

- 10x10G 1550nm PIC TX & RX components (Infinera) showing multi-channel integration feasibility (leading to low cost); jaeger_01_0107.

MMF/SMF Conclusions

MMF

- We understand VCSEL and PIN Array yield, reliability, and cost scaling with size (1x→4x→12x) and rate (2Gb/s→10Gb/s)

- MMF PMDs using 4x10G and 12x10G arrays are feasible, reliable and cost effective for 40GbE and 100GbE

SMF

- We understand 10G DML and PIN technology and how to make 10x10G channel Transceivers

- We understand 25G EAML and PIN technology and how to make 4x25G channel Transceivers

- SMF PMDs supporting up to 40km are feasible, reliable and cost effective for 100GbE

HSSG Tutorial
Copper PMDs

Chris Di Minico - MC Communications
Adam Healey - LSI Corporation

Physical Layer Specifications to be Defined

	40G	100G
At least 1m backplane	✓	
At least 10m cu cable	✓	✓
At least 100m OM3 MMF	✓	✓
At least 10km SMF		✓
At least 40km SMF		✓

High Speed Copper Interconnect: Applications

1.5 m

2.4 m max
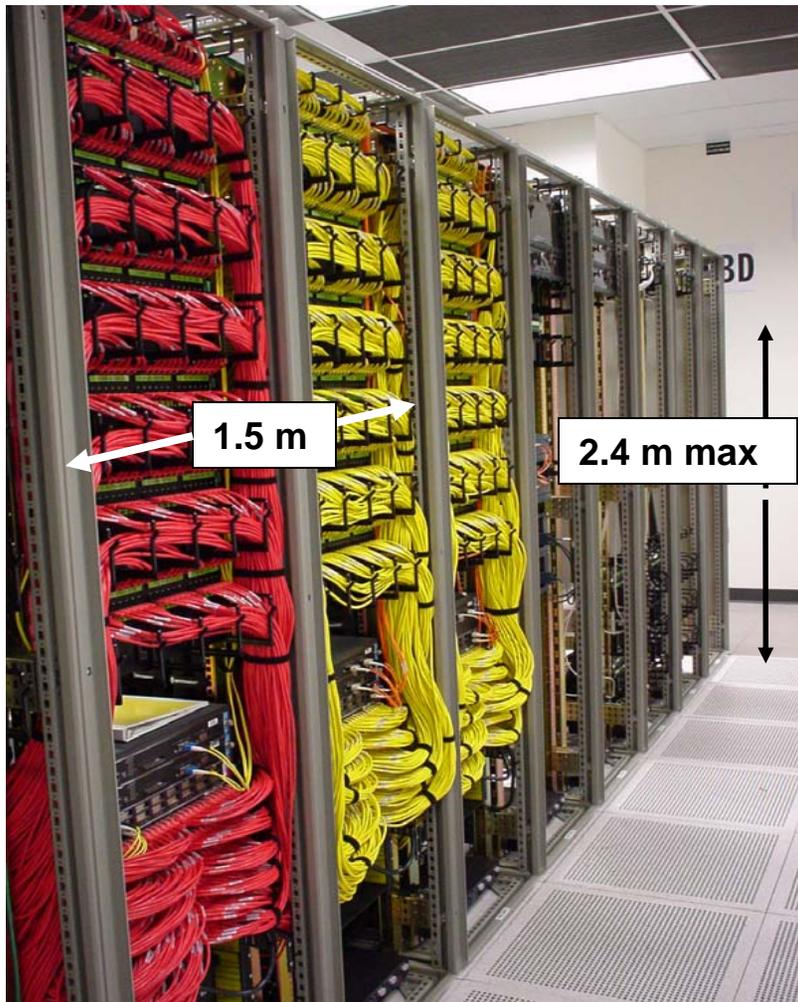
Intra/Inter rack/cabinet applications and High Performance Computing Interconnect

at least 10 meter copper objective

TIA-942 - Cabinet and rack height

• The maximum rack and cabinet height shall be 2.4 m (8 ft).

• Preferably no taller than 2.1 m (7 ft) for easier access to the equipment or connecting hardware installed at the top.

Twinaxial Copper Cable Assembly

- **100 ohm 8 pairs – 16 conductors**

 at least 10 meter interconnect
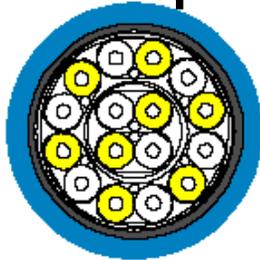
- **28 AWG - 5.6 mm-7.2 mm (0.220 in – 0.281 in)**

- **250 µm - 12-fibers groups for use with multifiber connectors**

  Flame-Retardant Outer Jacket
  250 µm Color-Coded Fiber
  Dielectric Strength Member

  Flame-Retardant Outer Jacket
  250 µm Color-Coded Fiber
  Dielectric Strength Member
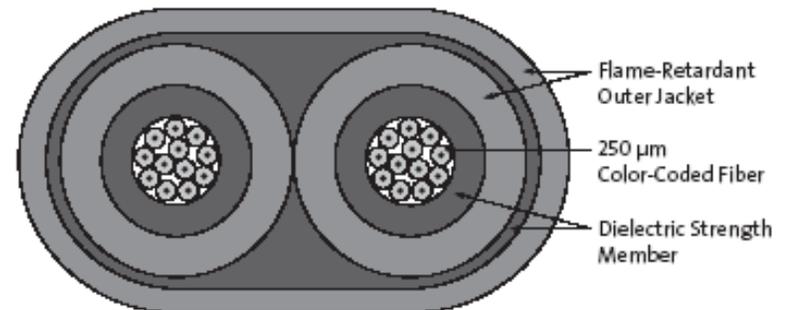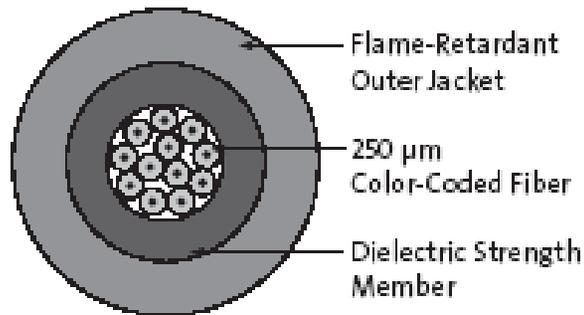
- **12-Fiber OFNP Fiber Optic Cable 4.4 mm (0.17 in)**
- **24-Fiber OFNP Fiber Optic Cable 8.3 mm (0.33 in)**

Source: Leoni High Speed Cables

Cable Assemblies – 28 AWG/24 AWG/28 AWG ADV-D 10 meter

MHz

Twinaxial cable:
advanced designs in comparison to
nominal CX4/Infiniband constructions
both electrical and mechanical

— worst case IL - 10 m - 28 AWG

— PSNEXT-4 disturber assembly

— PSFEXT-3 disturber-10 m assembly

— PS Total disturber Noise

— Insertion Loss - 10 m - scaled 24 AWG

— Insertion Loss ADV-D 28 AWG
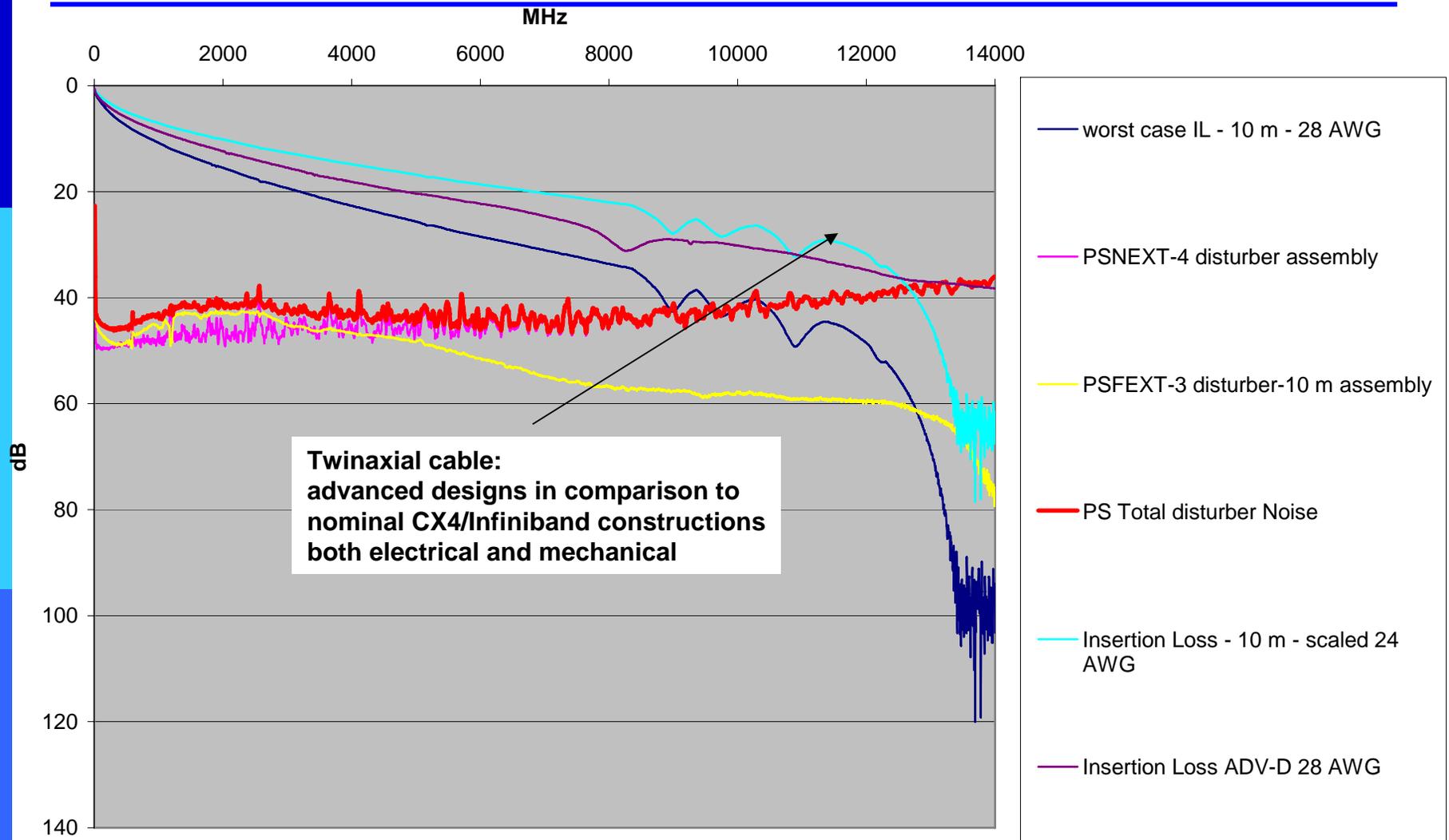
Source: Leoni High Speed Cables

802.3ap – Ethernet Operation over Electrical Backplanes

802.3 ap Insertion Loss vs twinaxial cable assembly IL

802.3ap – signaling 10.3125 GBd

Legend:
- IL 24 AWG equivlent - 10 m to 12 m
- ILmax(f)-KR
- Amax(f)-802.3ap Max Att
- CX4-IL-10 m
- IL-8mil-Strip Line-85 CM
- IL-IPass Cable assembly 26 AWG - 7m
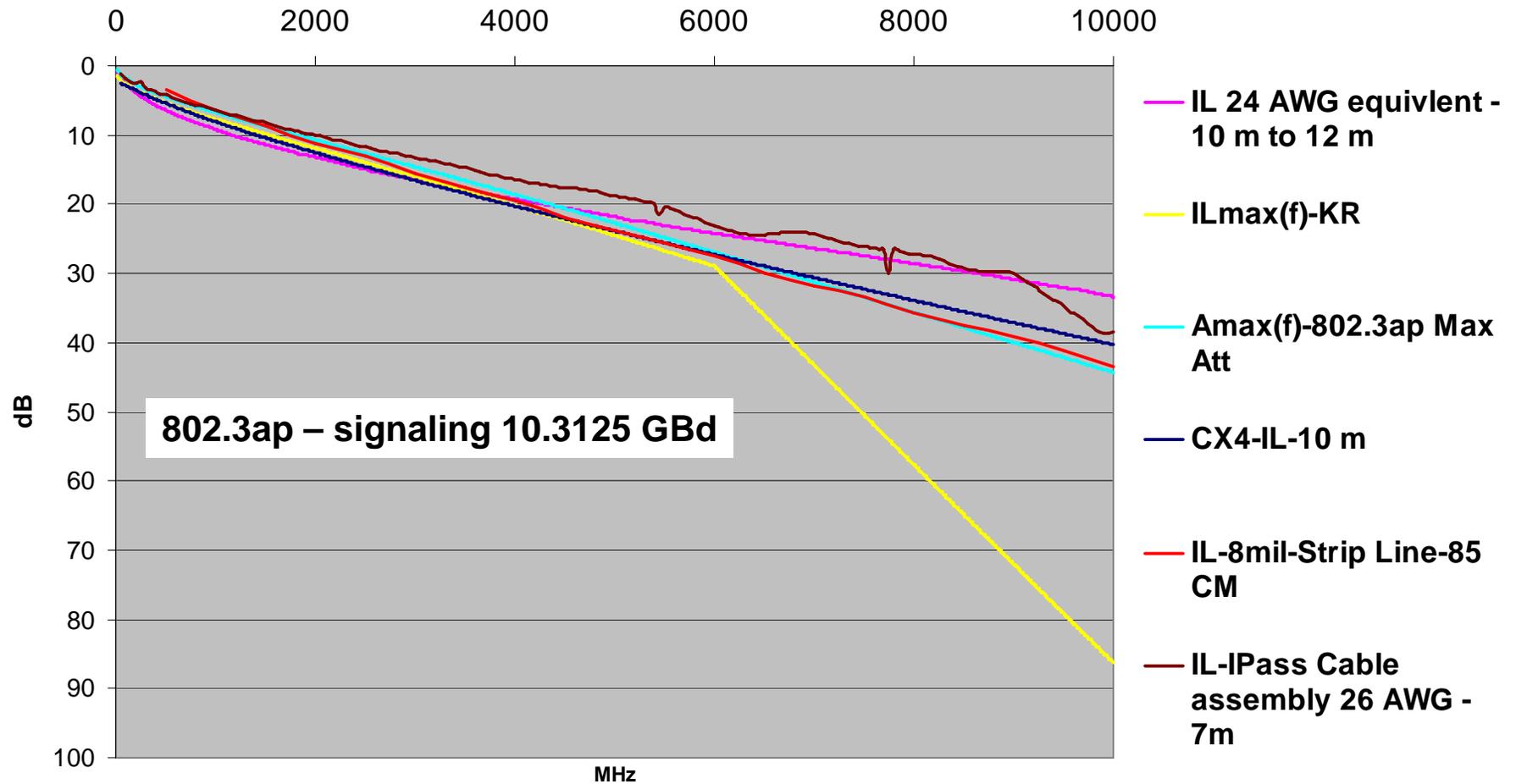
Eye Patterns for 24 AWG Cable - 10 to 12m

10m of AWG 27 100 Ohm Twinax
Source: +/-400mV, 2^5-1 PRBS, 10.0 Gb/s

De-emphasis 65%

50 mV/div

20 pS/div

10m of AWG 27 100 Ohm Twinax
Source: +/-400mV, 2^5-1 PRBS, 10.0 Gb/s

De-emphasis 75%

50 mV/div

20 pS/div
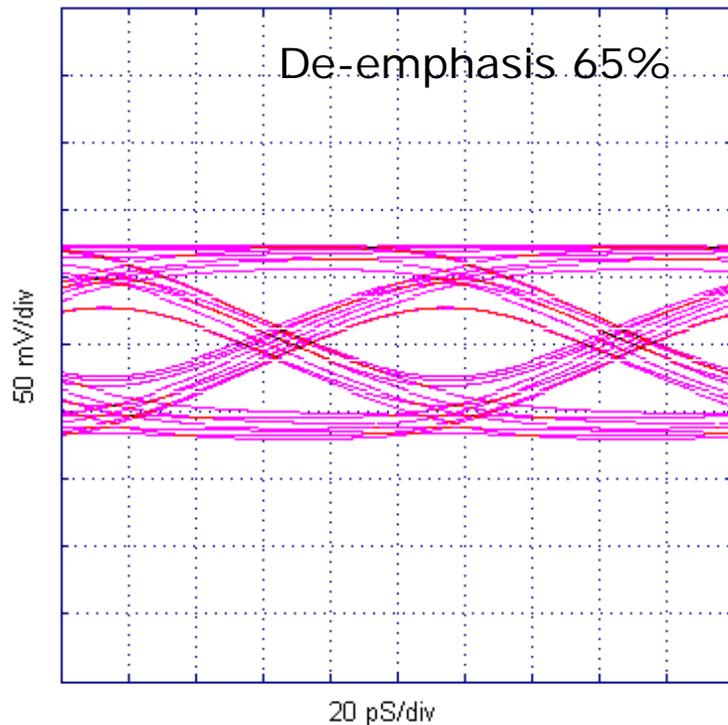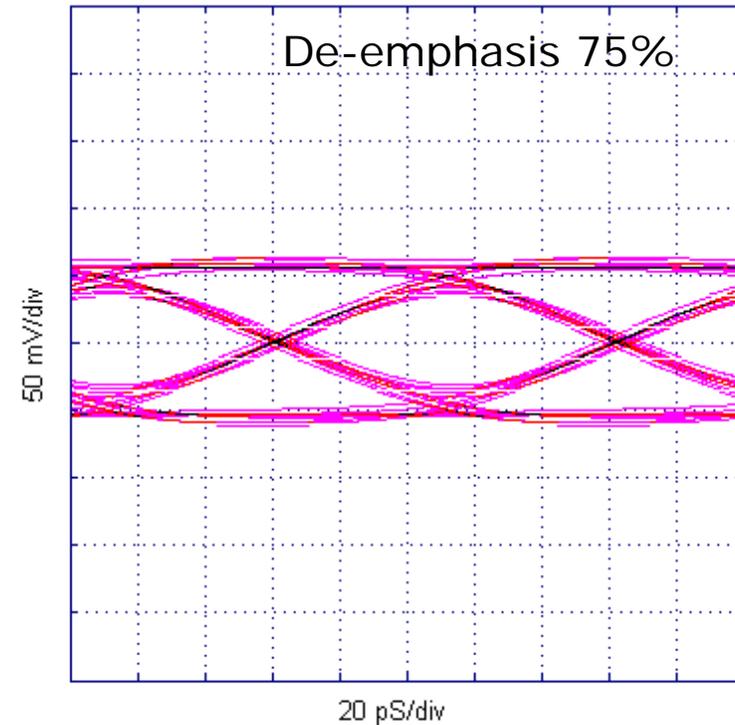
Various levels de-emphasis without cable equalization

Source: Herb Van Deusen, W.L. Gore

Lane Rate, Signaling Rate, Channel Bandwidth

10 m cable + connectors @ 6 dB Margin

maximum achievable lane rate for each coding gain

Maximum Lane rate	Maximum signaling rate	Info bits/baud/ dim	Modulated Info bits/baud/dim (2)	Channel bandwidth (1)	Copper Gauge	Code gain	Length	Modulation
Mb/s	Mbaud			MHz	AWG	dB	meters	
10889.28	8180.00	1.33	1.33	4090.00	28	0	10	PAM-3
13984.52	9476.64	1.38	1.48	4738.32	28	2	10	PAM-4
17555.24	10097.78	1.55	1.74	5048.89	28	4	10	PAM-4
24950.75	17290.00	1.44	1.44	8645.00	24	0	10	PAM-4
30727.91	16261.68	1.77	1.89	8130.84	24	2	10	PAM-4
36785.11	16888.89	1.94	2.18	8444.44	24	4	10	PAM-5
17212.31	13210.00	1.30	1.30	6605.00	28 ADVD	0	10	PAM-3
22044.24	13626.17	1.51	1.62	6813.08	28 ADVD	2	10	PAM-4
27147.01	13742.22	1.76	1.98	6871.11	28 ADVD	4	10	PAM-4

(1)Channel bandwidth=.5*(Maximum signaling rate)   (2)Infor/bits/baud/dim adjusted for coding gain

maximum achievable lane rate for each coding gain that yields 1 bit/baud

| Maximum Lane rate	Maximum signaling rate	Info bits/baud/ dim	Modulated Info bits/baud/dim	Channel bandwidth	Copper Gauge	Code gain	Length	Signaling
Mb/s	Mbaud			MHz	AWG	dB	meters	
10439.00	10439.00	1.00	1.00	5219.50	28	0	10	PAM-2/NRZ
12630.13	13580.00	0.93	1.00	6790.00	28	2	10	PAM-2/NRZ
14857.53	16690.00	0.89	1.00	8345.00	28	4	10	PAM-2/NRZ
21639.98	21639.98	1.00	1.00	10819.99	24	0	10	PAM-2/NRZ
23439.51	25170.00	0.93	1.00	12585.00	24	2	10	PAM-2/NRZ
24885.04	27950.00	0.89	1.00	13975.00	24	4	10	PAM-2/NRZ
16224.36	16224.36	1.00	1.00	8112.18	28 ADVD	0	10	PAM-2/NRZ
18748.43	20159.60	0.93	1.00	10079.80	28 ADVD	2	10	PAM-2/NRZ
21194.44	23813.98	0.89	1.00	11906.99	28 ADVD	4	10	PAM-2/NRZ

Source: George Zimmerman, Solarflare Communications, Chris DiMinico, MC Communications

IEEE 802.3 Higher Speed Study Group - TUTORIAL

Physical Layer Specifications to be Defined

	40G	100G
At least 1m backplane	✔	
At least 10m cu cable	✔	✔
At least 100m OM3 MMF	✔	✔
At least 10km SMF		✔
At least 40km SMF		✔

40 Gigabit Backplane Ethernet

- IEEE 802.3ap™-2007 defines GbE and 10GbE operation over a modular platform backplane (1 m objective)
  - 1000BASE-KX (GbE)
  - 10GBASE-KX4 (10 GbE, 4 x 3.125 GBd)
  - 10GBASE-KR (<u>serial</u> 10 GbE)

- Blade servers: 2nd generation backplane
  - Based on 10GBASE-KX4 architecture...
  - ...but satisfy 10GBASE-KR channel requirements
  - 40 Gb/s capability inherent

- "40GBASE-KR4" leveraged from 10GBASE-KR standard
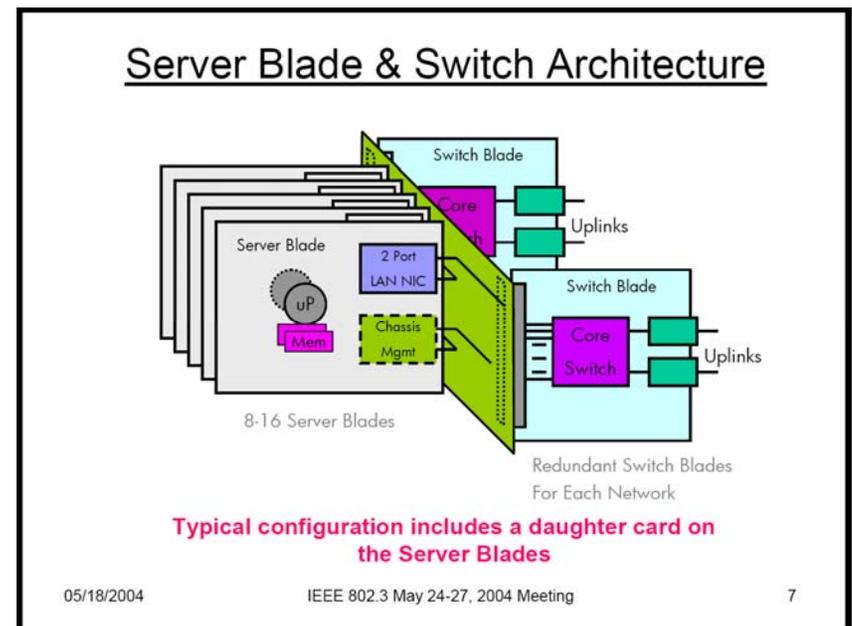
Source: Koenen, "Channel Model Requirements for Ethernet Backplanes in Blade Servers", May 2004.

Conclusions

- Copper cable assembly
  - Technical feasibility, economic feasibility, and market potential for a Higher Speed copper interconnect demonstrated.
  - Up to 10 meter reach consistent with intra/inter rack application and HPC cluster distances.
  - High speed study group copper interconnect objective of at least 10 meters to address intra/inter rack applications and high performance computing (HPC) interconnects.

- 40 Gb/s Backplane Ethernet
  - 40GBASE-KR4 leveraged from 10GBASE-KR standard
  - 40 Gb/s capability inherently demonstrated

Wrap-up

John D'Ambrosia, Force10 Networks

Summary

- Bandwidth requirements are growing for all applications in the Ethernet EcoSystem
  - 40GbE targets computing / server applications
  - 100GbE targets networking aggregation applications
- The project meets all of the 5 Criteria
- The project is technically feasible:
  - All Physical Layer Specifications
  - OTN Compatibility
- Estimated completion date: May 2010
- Website: http://www.ieee802.org/3/hssg/public/index.html

Questions?

Thank You!

Backup Slides

Abbreviations (1)

Amax – Maximum Fitted Attenuation

AWG - American Wire Gauge

BER – Bit Error Ratio

CWDM - Coarse Wavelength Division Multiplexing

DC – Dispersion Compensation

DFB – Distributed Feedback (laser)

DML – Direct Modulation Laser

DQPSK - Differential Quadrature Phase
       Shift Keying

DVR - Digitial Video Recorder

DWDM - Dense Wavelength Division Multiplexing

EAML - Electro Absorption Modulator Laser

EFM – Ethernet in the First Mile

EML – Externally Modulated Laser

FC – Fibre Channel

HBA - Host Bus Adapter

HDTV – High Definition Television

HPC – High Performance Computing

IB - Infiniband

IL – Insertion Loss

ILmax – Maximum Insertion Loss

IPTV – Internet Protocol Television

ISP – Internet Service Provider

ITU - International Telecommunication Union

iTV - Internet Television

IXC - Interexchange Carrier

LAN – Local Area Network

LLC – Logical Link Control

LWDM - LAN (Local Area Network)
       Wavelength Division Multiplexing

MAC - Media Access Control

MMF – Multimode Fiber

MSA - Multi-Source Agreement

MZML - Mach Zehnder Modulator Laser

NIC – Network Interface Card

NRZ - Non-return to Zero

IEEE 802.3 Higher Speed Study Group - TUTORIAL

Abbreviations (2)

OA – Optical Amplification

ODU3 - Optical Channel Data Unit-3 (~40G)

OFNP - Optical Fiber Nonconductive Plenum

OM2 – 850 / 1310nm Optimized MMF

OM3 – 850nm Laser Optimized MMF

OTN – Optical Transport Network

PCI-E - PCI Express (Peripheral Component Interconnect Express)

PCS – Physical Coding Sublayer

PHY- Physical Layer Device

PIC - Photonic Integrated Circuit

PIN - PIN diode (Positive Intrinsic Negative diode)

PLS – Physical Layer Signaling

PMA – Physical Medium Attachment

PMD – Physical Medium Dependent

PME - Physical Medium Entity

POP4 - Pluggable Optics 4-channel MSA module

PSFEXT - Power Sum Far-end Crosstalk

PSNEXT – Power Sum Near-end Crosstalk

QSFP - Quad Small Form-factor Pluggable

SAN – Storage Area Network

SDTV - Standard Definition Television

SFP+ - Small Form-factor Pluggable (8.5G & 10G optical modules)

SMF – Single Mode Fiber

SNAP12 - 12 channel Pluggable Optical Module MSA

SOA - Semiconductor Optical Amplifier

TAM - Total Available Market

TIA - Trans-Impedance Amplifier

VCSEL - Vertical-Cavity Surface-Emitting Laser

VOD – Video On Demand

WDM - Wavelength Division Multiplexing

XENPAK - 10 Gigabit Ethernet MSA

XFP - 10 Gigabit Small Form Factor Pluggable Module