400 Gigabit Ethernet Call-For-Interest Consensus

IEEE 802.3 Ethernet Working Group
IEEE 802 March 2013 Plenary, Orlando, FL

Objective for this Meeting

- To ***measure the interest*** in starting a study group for ***400 Gb/s Ethernet***

- We don't need to
  - Fully explore the problem
  - Debate strengths and weaknesses of solutions
  - Choose any one solution
  - Create PAR or five criteria
  - Create a standard or specification

- Anyone in the room may speak / vote
- RESPECT… give it, get it

Contributors

- Pete Anslow, Ciena
- Chris Cole, Finisar
- Kai Cui, Huawei
- John D'Ambrosia, Dell
- Mark Gustlin, Xilinx
- Jeff Maki, Juniper
- Andy Moorwood, Infinera
- Gary Nicholl, Cisco
- Mark Nowell, Cisco
- David Ofelt, Juniper
- Brian Teipen, ADVA
- Steve Trowbridge, Alcatel-Lucent
- IEEE 802.3 Higher Speed Ethernet Consensus Ad Hoc

Agenda

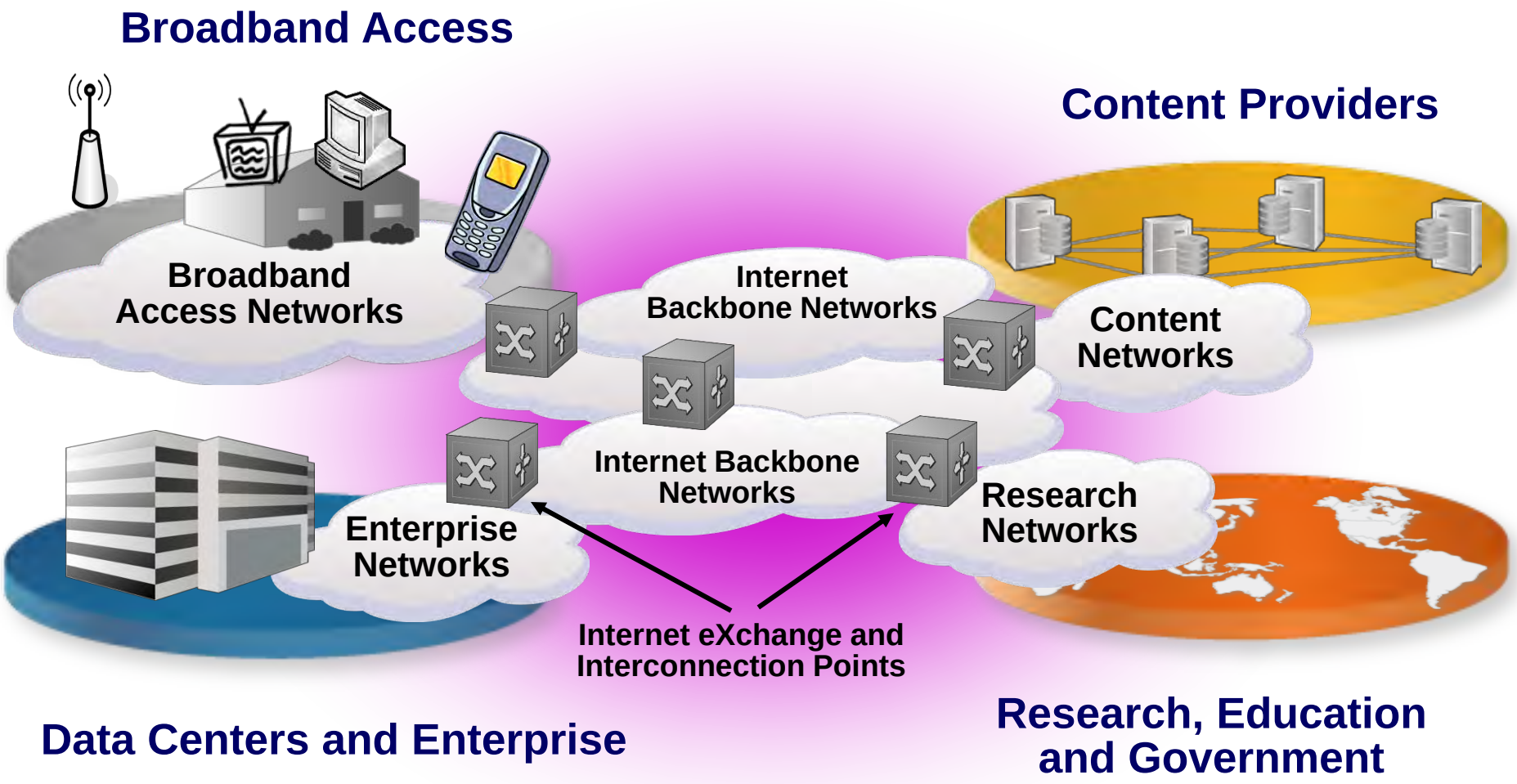
- **Presentations**
  - **“The Bandwidth Explosion,” John D’Ambrosia**
  - **“Beyond 100 Gigabit Ethernet,” TBD.**
  - **“Technical Viability Beyond 100 Gigabit Ethernet,” TBD.**
  - **“400 Gigabit Ethernet- Why Now,” John D’Ambrosia.**
- **Discussion**
- **Call for Interest**
- **Future Work**

Presented by

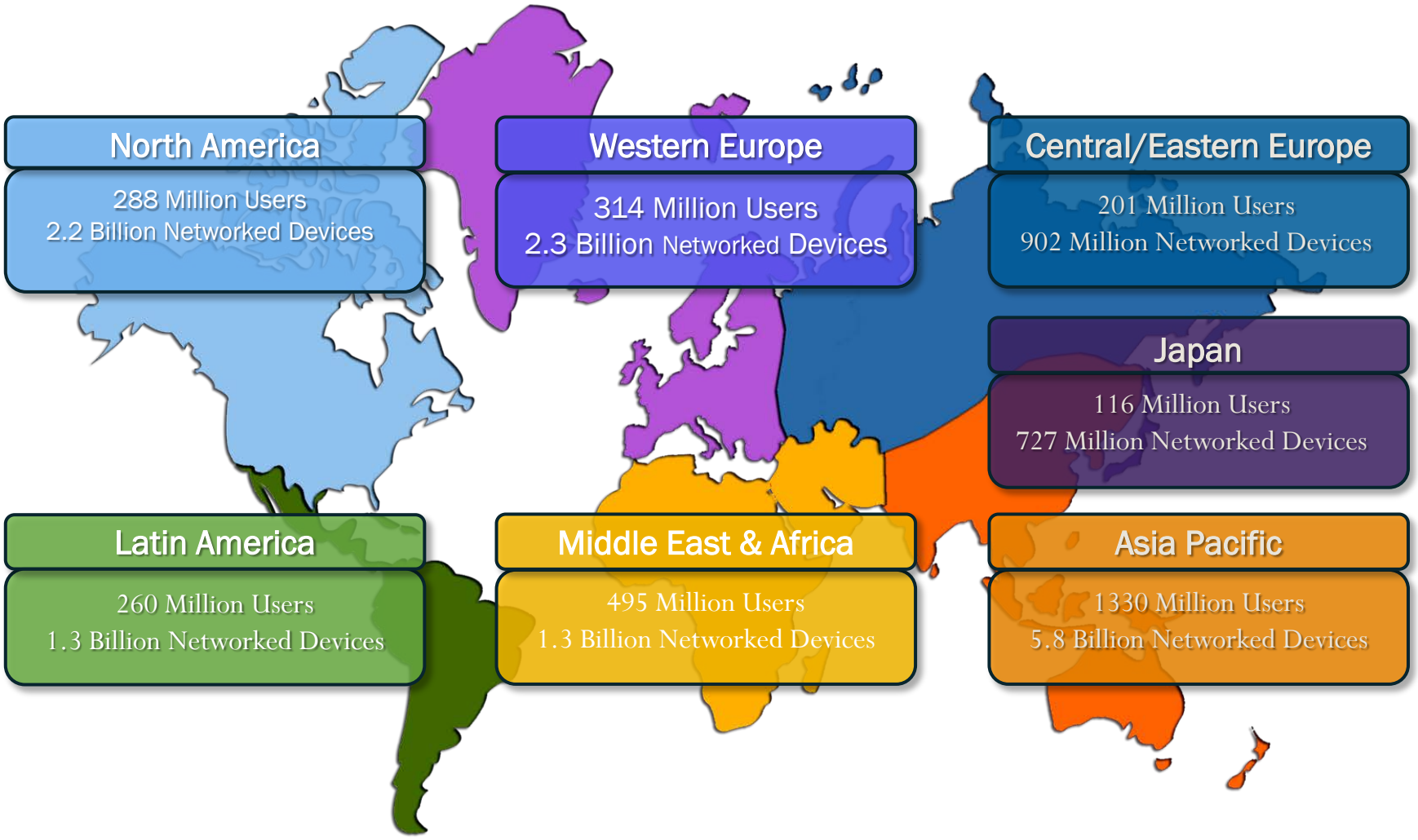
IEEE 802.3 Working Group
Orlando, FL, USA
March 19, 2013

THE BANDWIDTH EXPLOSION

The Ethernet Eco-System (2007 HSSG)



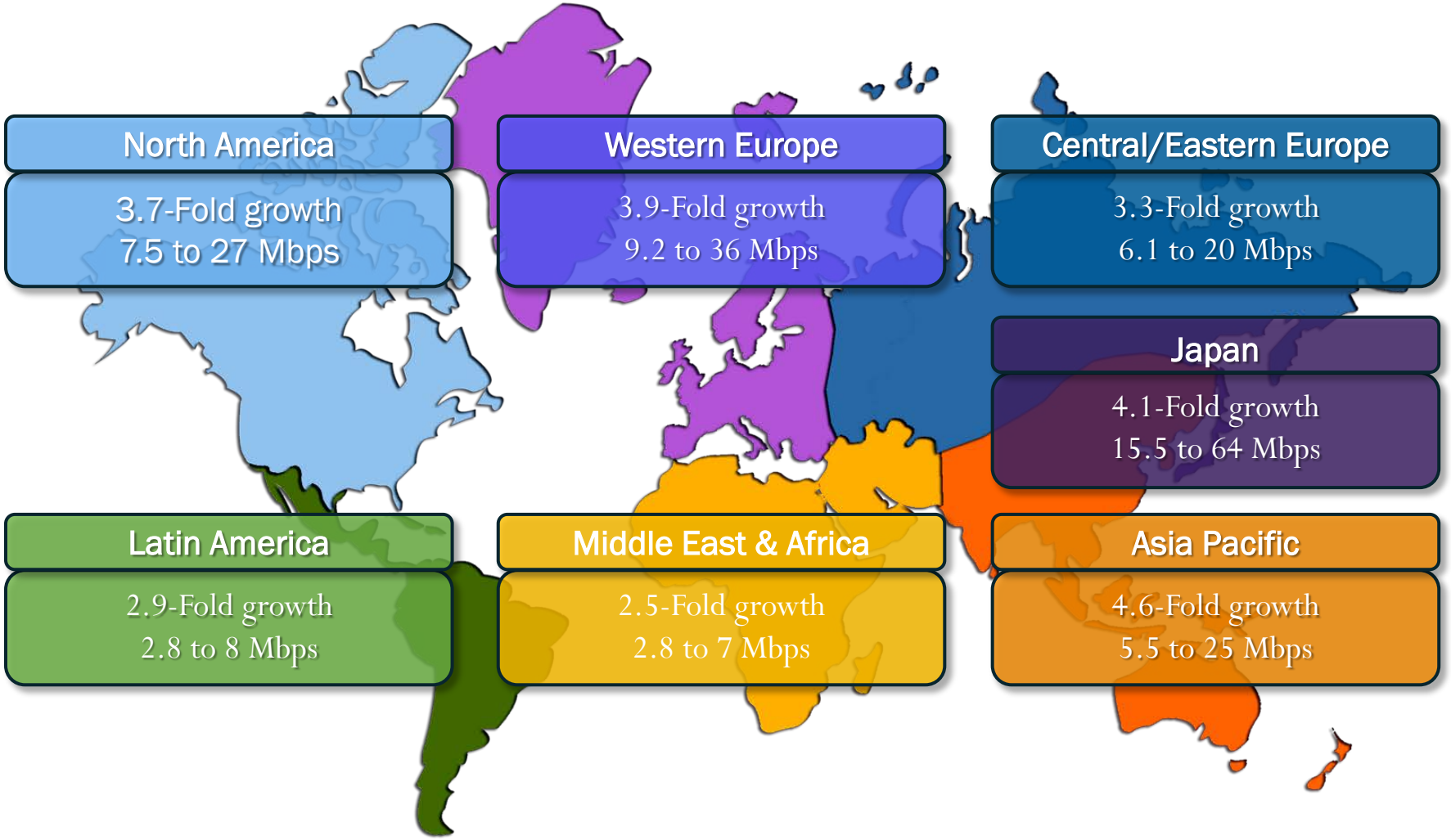
2015 Global Users and Network Connections



Source: nowell_01_0911.pdf citing Cisco Visual Networking Index (VNI) Global IP Traffic Forecast, 2010–2015, http://www.ieee802.org/3/ad_hoc/bwa/public/sep11/nowell_01_0911.pdf

Global Broadband Speed 2010-2015

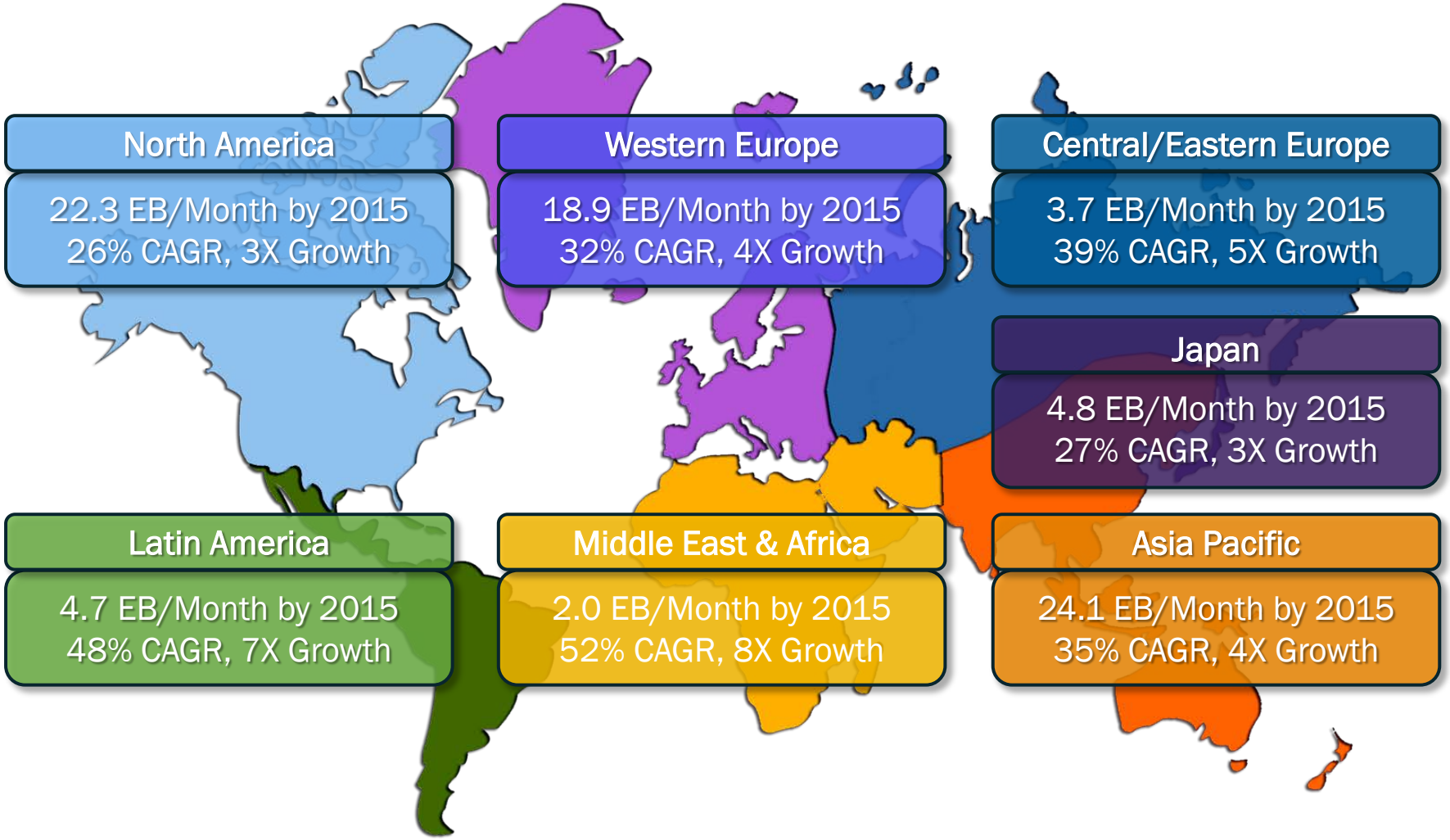
Average broadband speed will grow 4X; from 7 to 28 Mbps



Source: nowell_01_0911.pdf citing Cisco Visual Networking Index (VNI) Global IP Traffic Forecast, 2010–2015, http://www.ieee802.org/3/ad_hoc/bwa/public/sep11/nowell_01_0911.pdf

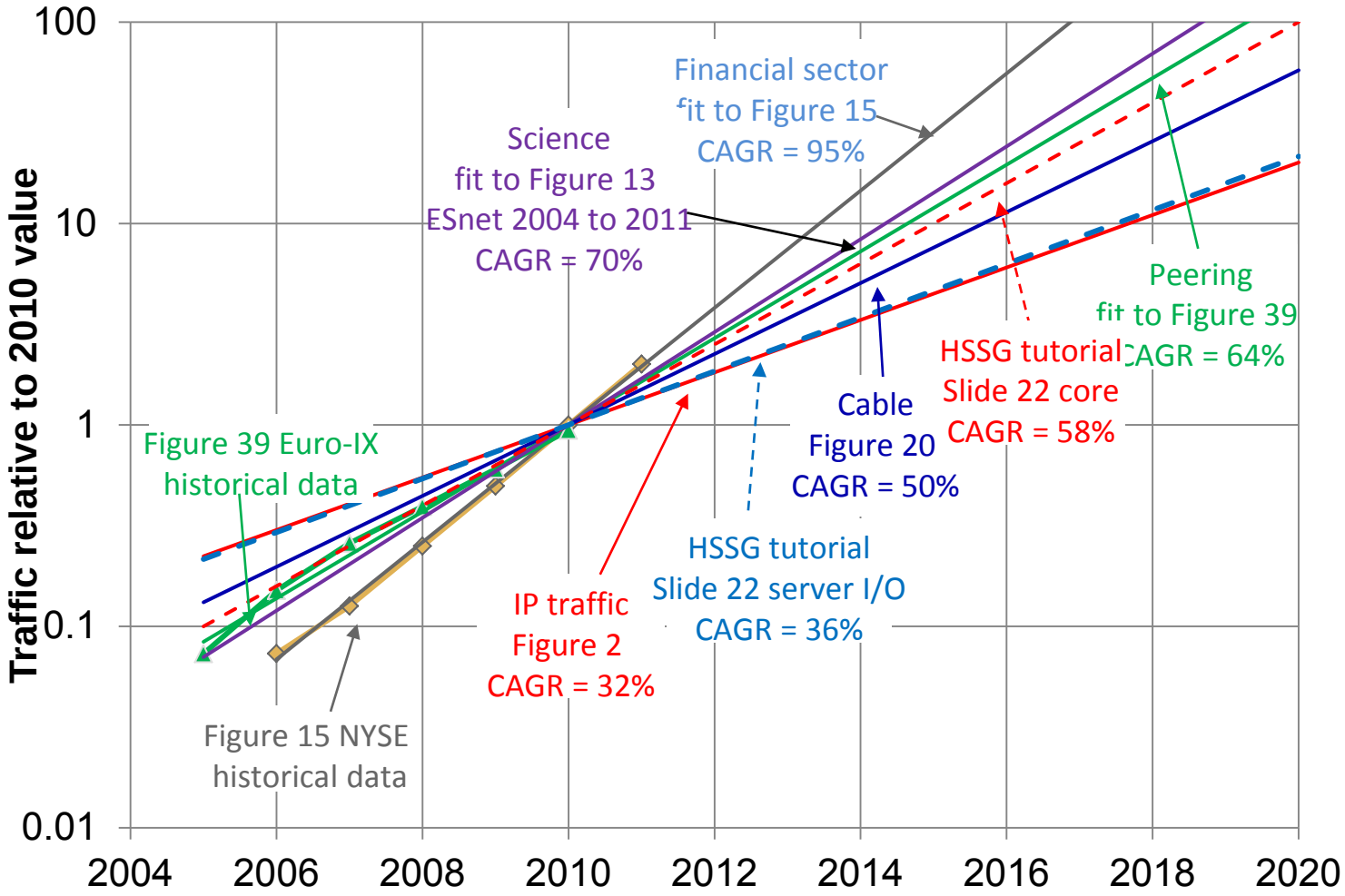
Global IP Traffic Growth, 2010–2015

Regional contributions to the Zettabyte journey



Source: nowell_01_0911.pdf citing Cisco Visual Networking Index (VNI) Global IP Traffic Forecast, 2010–2015, http://www.ieee802.org/3/ad_hoc/bwa/public/sep11/nowell_01_0911.pdf

Findings of IEEE 802.3 BWA Ad Hoc



Source: http://www.ieee802.org/3/ad_hoc/bwa/BWA_Report.pdf

IEEE 802.3 BWA Findings

- **On average - 58% CAGR**
  - **In 2015, 10x requirements of 2010 – Terabit capacities**
  - **In 2020, 100x requirements of 2010 – 10 Terabit capacities**

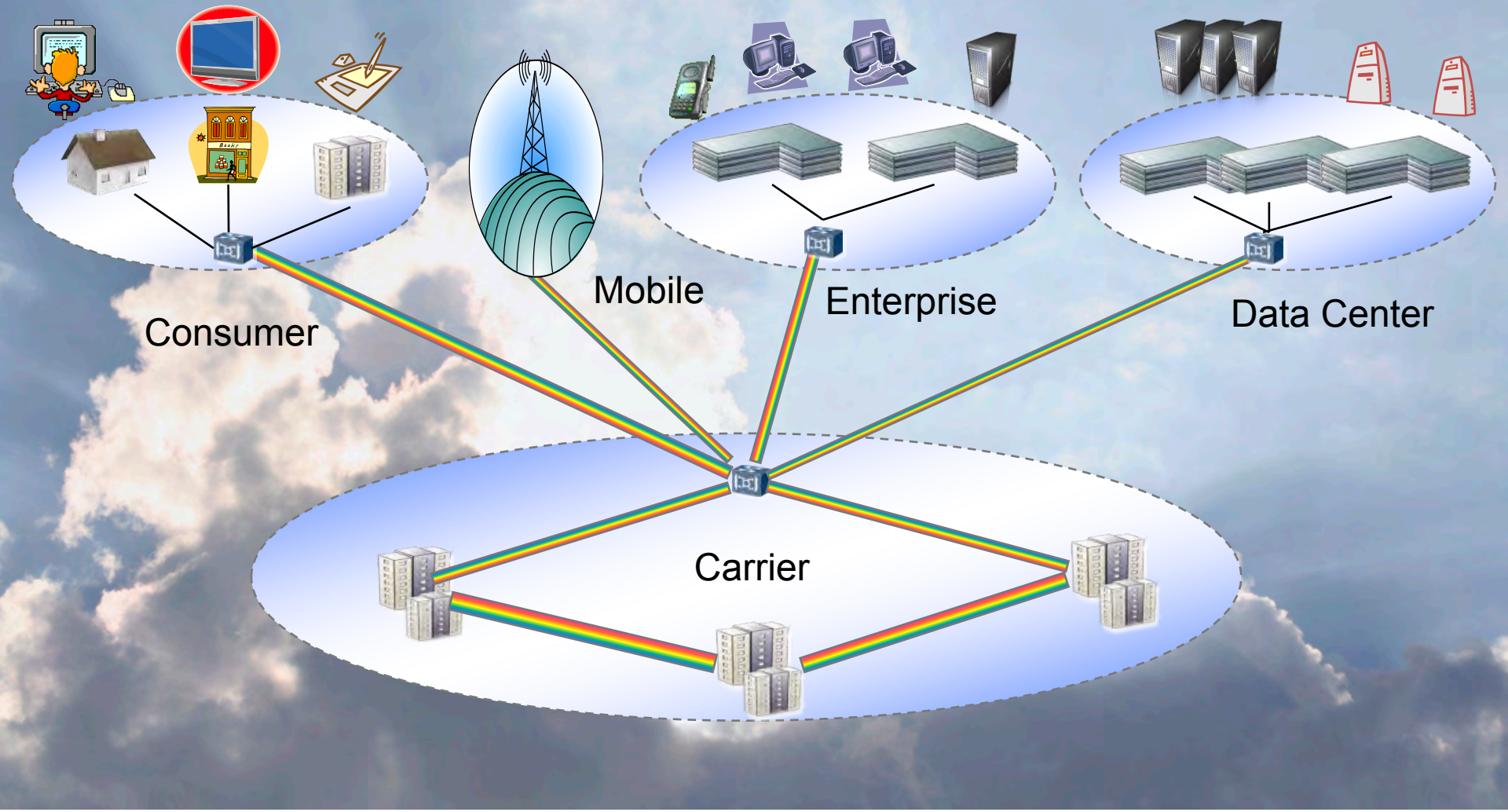
- **This growth is a predictor of the future only if downward cost per bit trend is continued**
  - **Ethernet cost per bit has to fall with time or the predicted exponential rise in traffic will result in unsupportable costs**

Presented by

IEEE 802.3 Working Group
Orlando, FL, USA
March 19, 2013

"BEYOND 100 GIGABIT ETHERNET"

The Ethernet Eco-System Today



Changes– Infrastructure / Devices

- **Since 2007 HSSG**
  - **Smart Phones**
    - **iPhone – 2007-01**
    - **Android – 2007-11**
  - **Tablets**
    - **iPad – 2010-04**
  - **Wi-Fi Deployments**
  - **3G / 4G / LTE**
  - **10G Server Deployment**
  - **Future 40G Server Deployment**
  - **Internet Enabled TV**
  - **The “Cloud”**

Changes - Applications

- **Cloud Businesses**
  - **Amazon (AWS)– (beta) 2006-08**
  - **MicroSoft Azure – 2008-10**
- **Practical Cloud Storage**
  - **Dropbox – 2007-06**
- **Ubiquitous Video Streaming**
  - **NETFLIX streaming – 2007**
  - **HULU – 2007-03**
- **Social Media**
  - **Facebook**
  - **Twitter**
  - **LinkedIn**

Science: Big Data Sources

CERN is "the tip of the iceberg"

Today

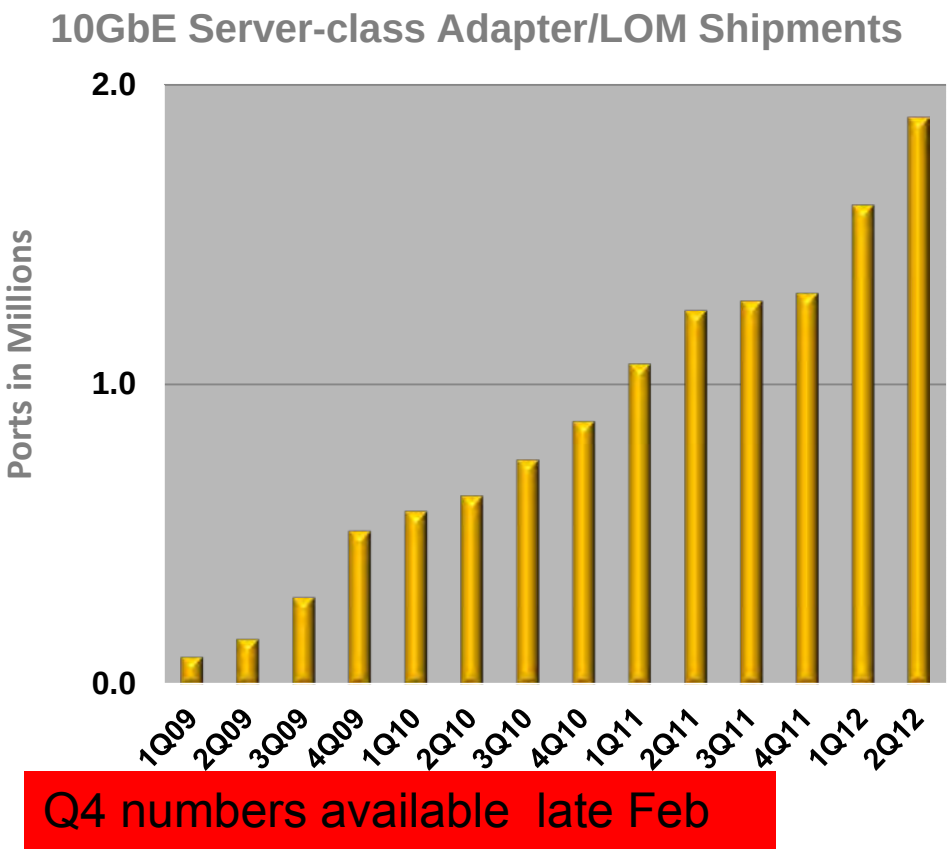
- CERN
  - Atlas detector in LHC (Large Hadron Collider) generates ~1 petabyte/sec
  - Trigger farm reduces to 450MB/sec
    - Tens of Gb/s of outbound traffic to analysis centers
- Genome sequencing
  - Per-instrument data rate strongly ↗ (~10x over 5 years)
  - Data costs plummeting → vastly increased data volume
  - http://www.genome.gov/sequencingcosts/

Future

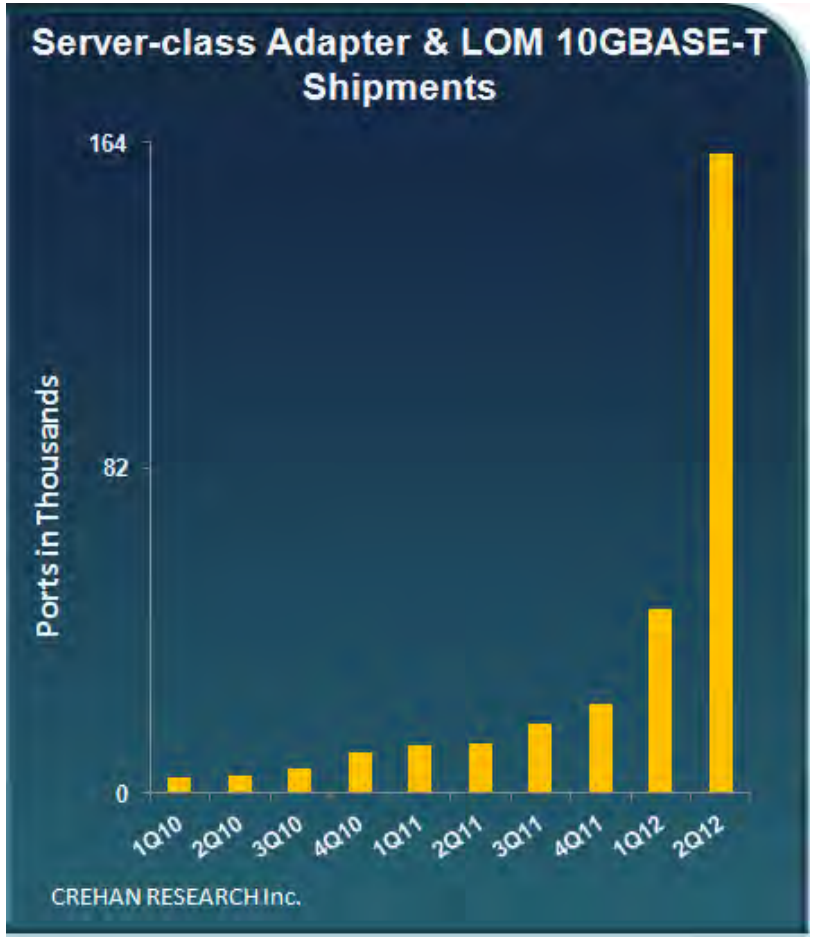
- Belle-II
  - 250PB of experimental data in first 5 years of operation
- Square Kilometer Array (SKA)
  - ~2800 receivers in telescope array
  - 2 petabytes/sec to central correlator
    - sending @ ~100 Gb/s to analysis centers

Source: http://www.ieee802.org/3/ad_hoc/bwa/public/dec11/dart_01_1211.pdf (updated: interview Eli Dart, August 29, 2012)

10GbE Server Deployments



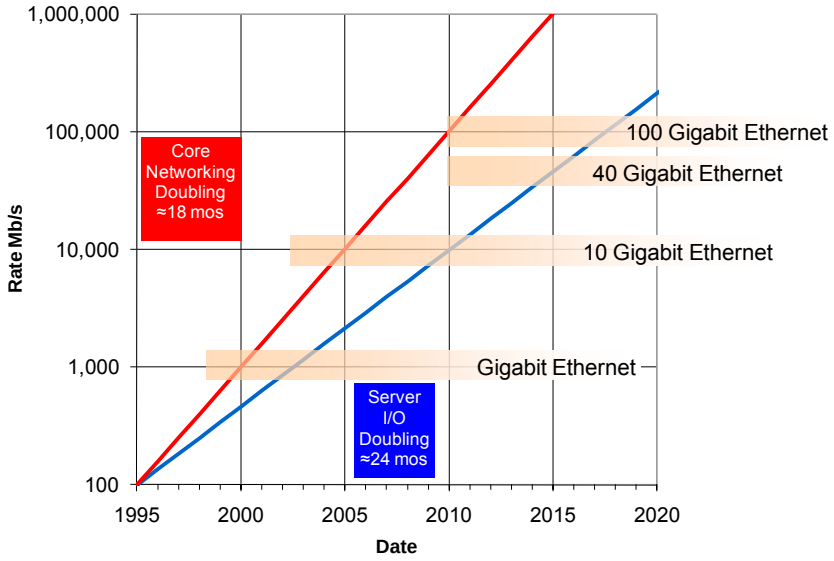
Q4 numbers available late Feb



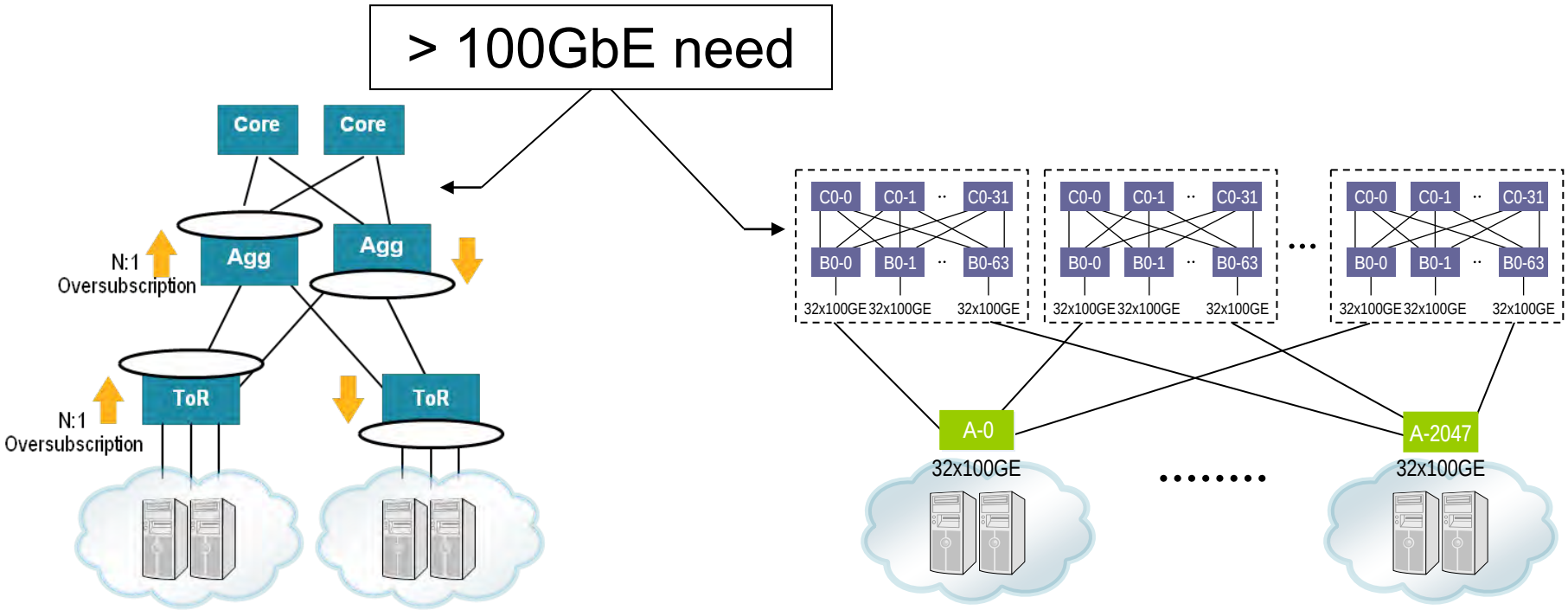
All data used with permission Seamus Crehan, Crehan Research.

The Server Roadmap

- Let's not forget 40GbE servers in 2014!
- Upgrade of 10GbE Blade Servers to 40GbE
- Introduction of 40GbE blade Servers for migration path to 100GbE blade servers
- IEEE P802.3bj
- Next Gen BASE-T Study Group



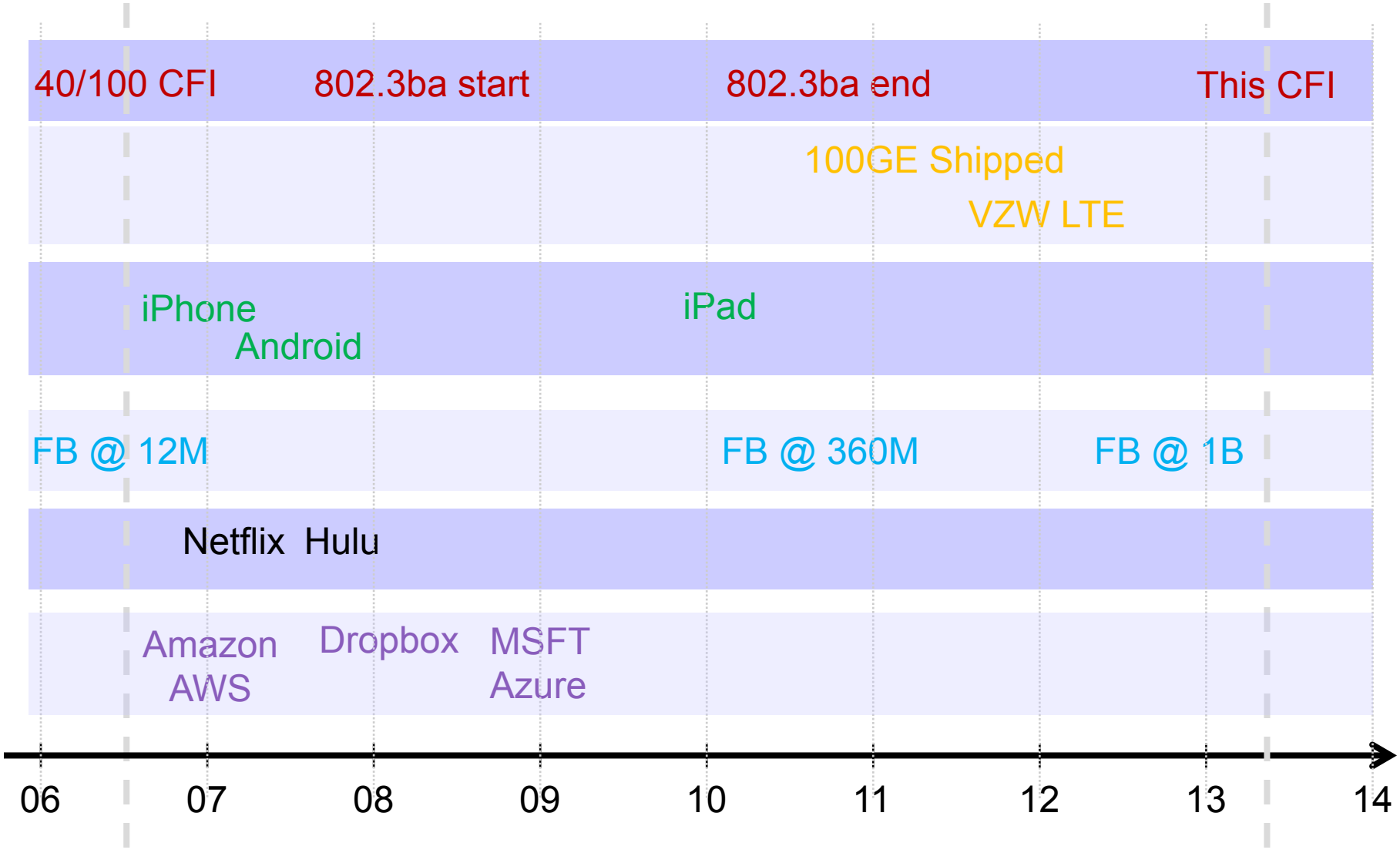
Data Center Architectures



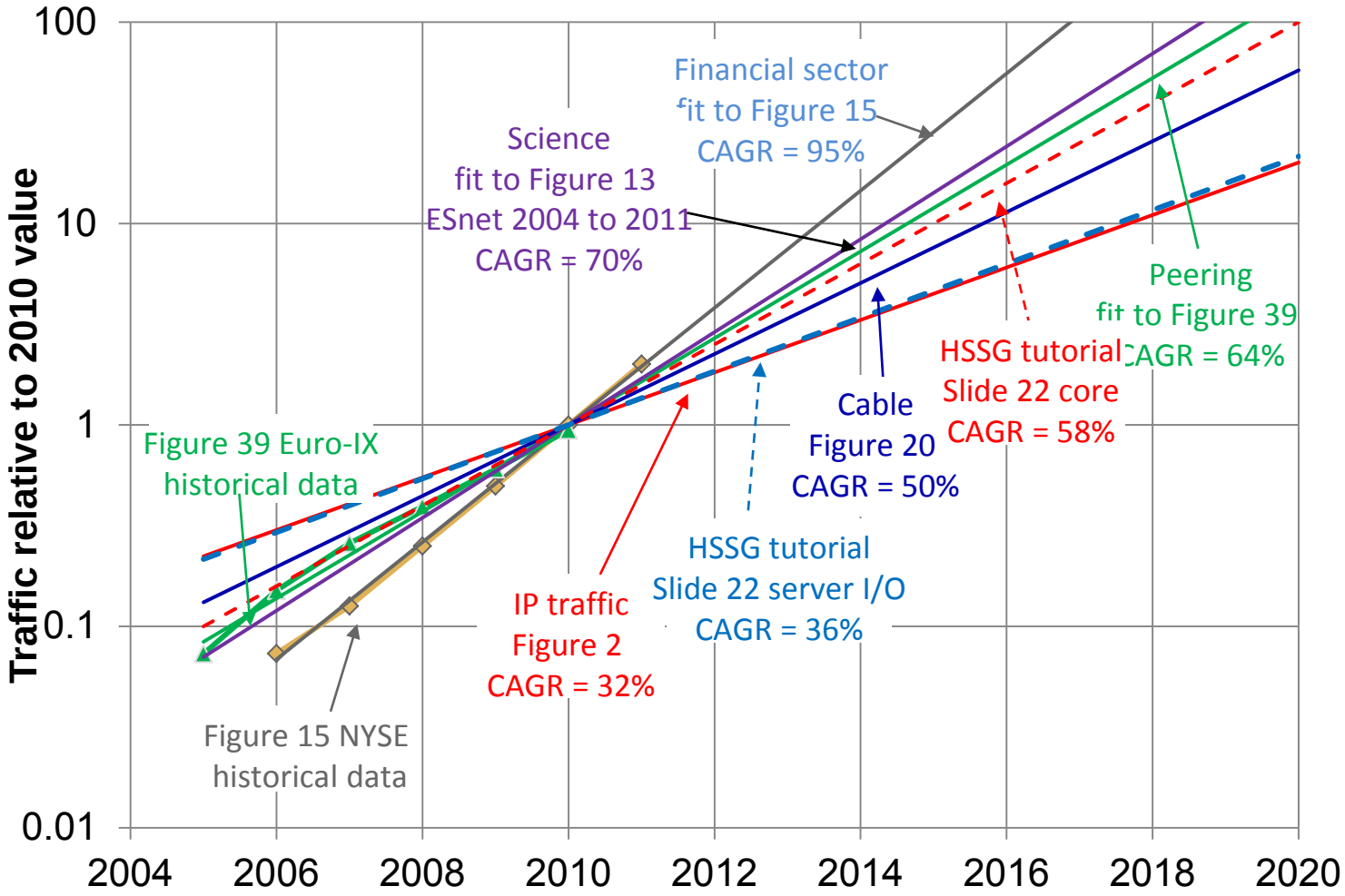
Hierarchical Fat Tree architecture

Non-blocking architecture

Life since IEEE P802.3ba



Findings of IEEE 802.3 BWA Ad Hoc



Source: http://www.ieee802.org/3/ad_hoc/bwa/BWA_Report.pdf

Aggregation Driving Near-term Applications

- Core ⇔ Transport (400Gb/s Transport demonstrated)
- Core ⇔ Core
- Datacenter ⇔ Datacenter
- Datacenter upper layer switch interconnect (shown on previous slide)

End User Quotes (???)

Section Summary

- Bigger fat pipe
- Time to market
- Cost

- 400 GbE meets this
- Terabit Ethernet doesn't

The Ethernet EcoSystem Now

- Update with a figure that shows today's Applications
- Smart pad
- Smart phone
- HDTV
- Wi-Fi
- Euro-IX (JD)
- Data Center (40GBASE-T)
- OTN (Andy)
- R&D (JD>Bennett)
- Content Providers (Andy / Dave O)
- Cloud (Amazon, Google)

Today's Trends

- Cloud slide – anytime anywhere anyhow(Business / consumer)
- Mobile
- Outsourcing of data center / IT
- Thin Client is here
- Flat Networks
- Data Centers

Presented by

IEEE 802.3 Working Group
Orlando, FL, USA
March 19, 2013

TECHNICAL VIABILITY BEYOND 100 GIGABIT ETHERNET

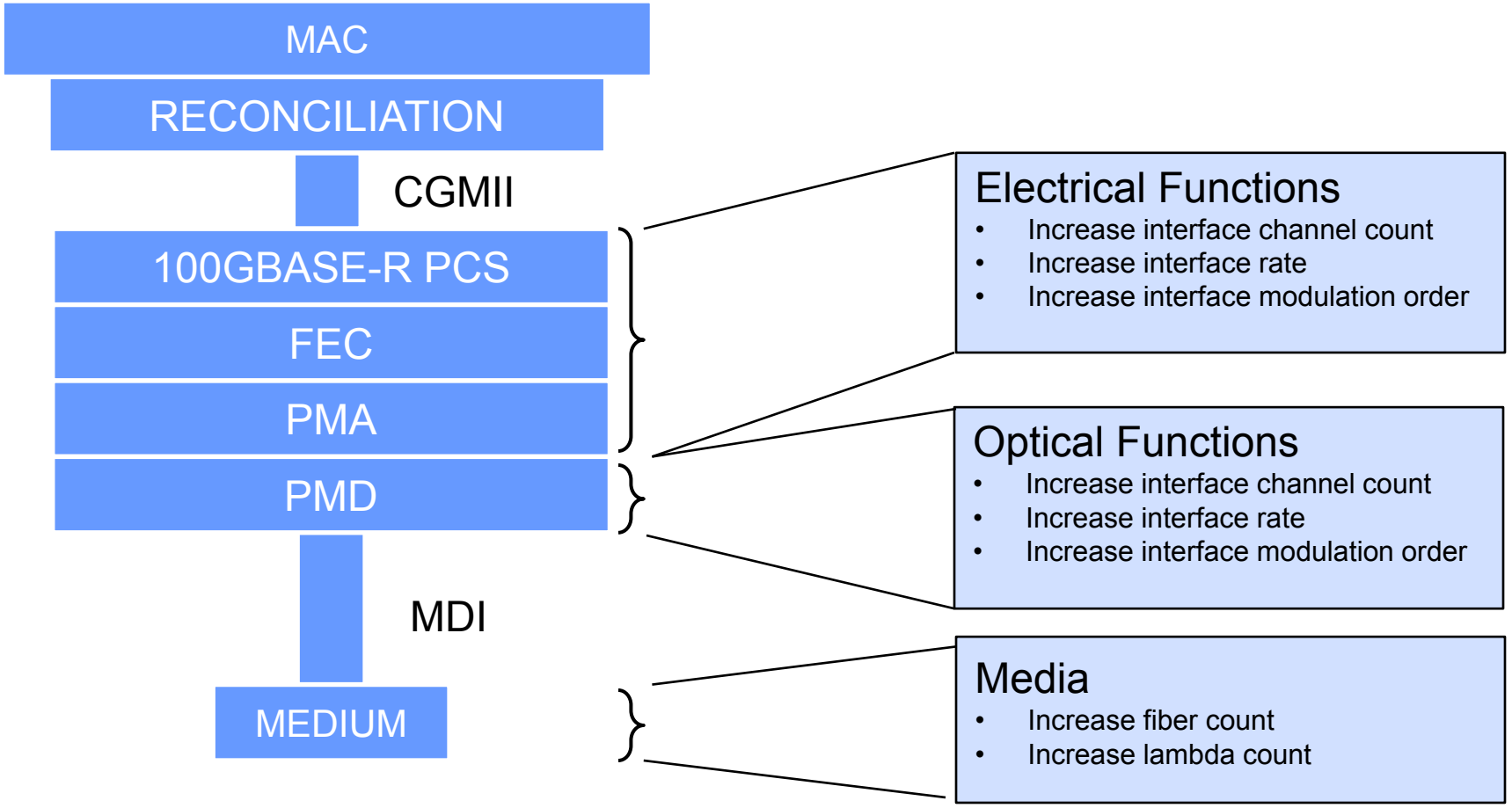
Link Aggregation

- Link Aggregation: an aggregation of lower speed links to create a virtual higher speed link
  - A way to address bandwidth requirements between the release of faster links
- But…
  - Traffic is often trunked into large tunneled flows
    - Insufficient entropy to do hashing efficiently
  - Load imbalance due to differences in flow sizes
  - Exponential bandwidth growth implies exponential growth in number of links
    - Management / operational issues
  - Fastest flow limited to individual link speed
- Faster links address these issues!
  - … and they will be lagged also!

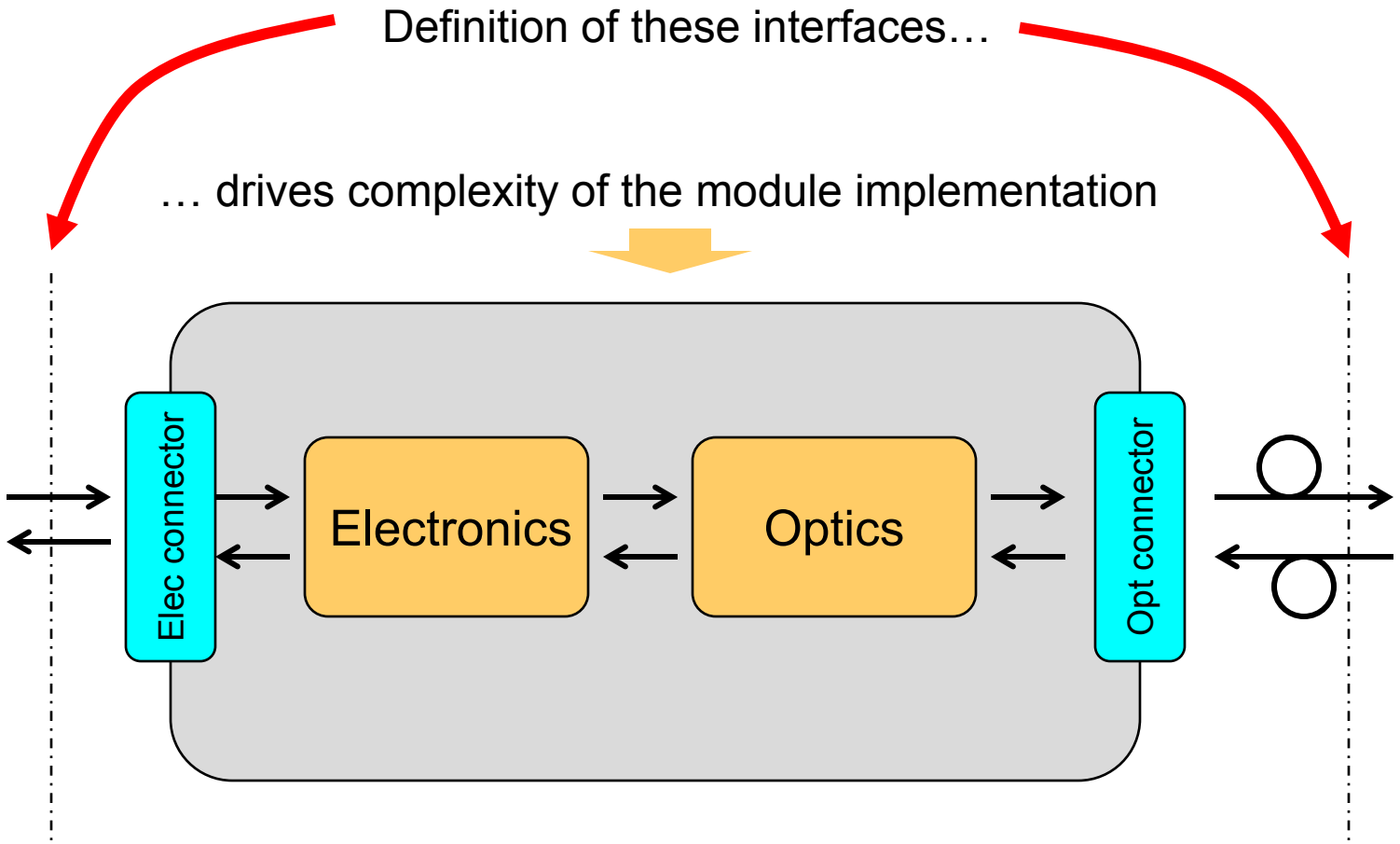
400Gb/s vs. 4 x 100Gb/s Link Aggregation

- Traffic is often trunked into large tunneled flows
  - Insufficient entropy to do hashing efficiently
  - Link Aggregation (LAG) is inefficient
  - BW not considered which leads to flow imbalance
  - A faster interface provides predictable performance
- Sources of large flows:
  - Content distribution
  - Secure traffic
- Fewer items to manage provides operational efficiency
  - Bandwidth is growing exponentially
  - Without faster links, link count grows exponentially therefore management pain grows exponentially

Potential enablers for more Gb/s/lane

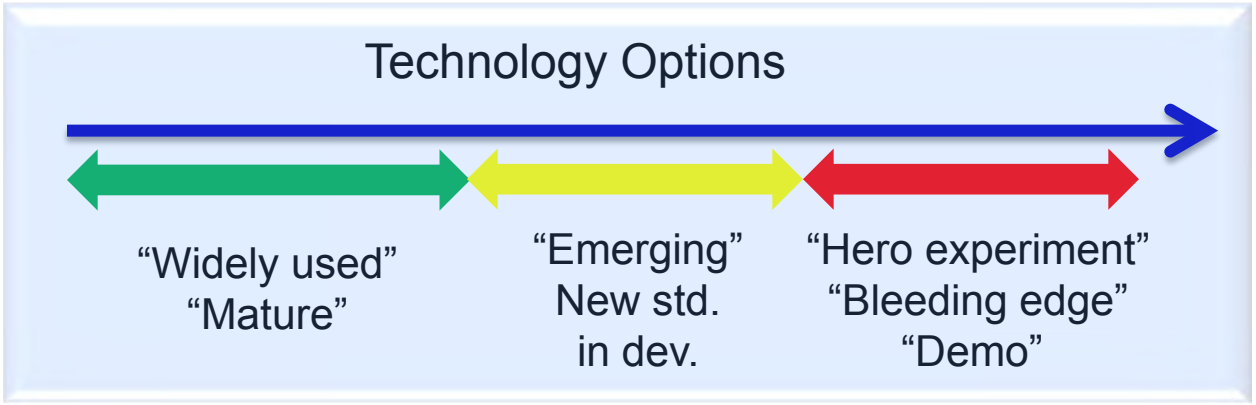


Example: Anatomy of an Optical Module Implementation

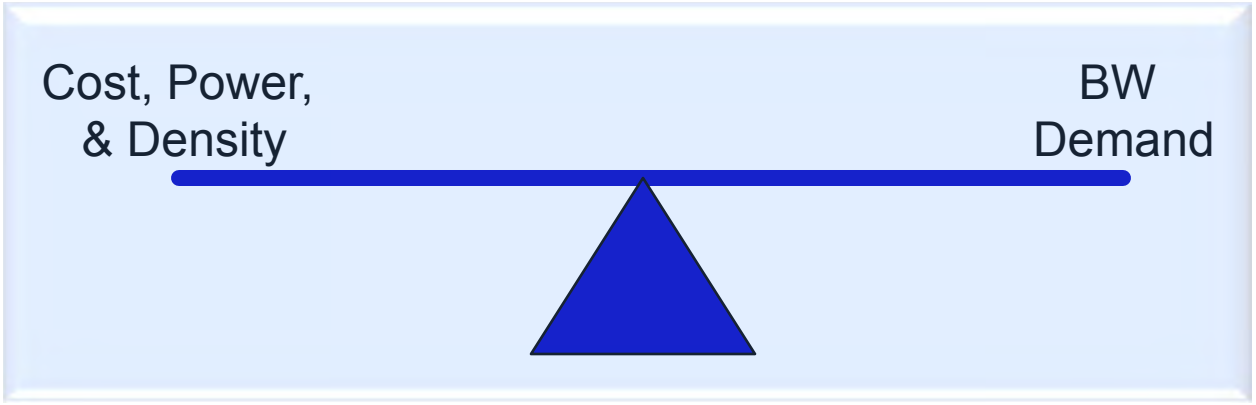


Matching Needs with Capabilities

It's all going to change with time

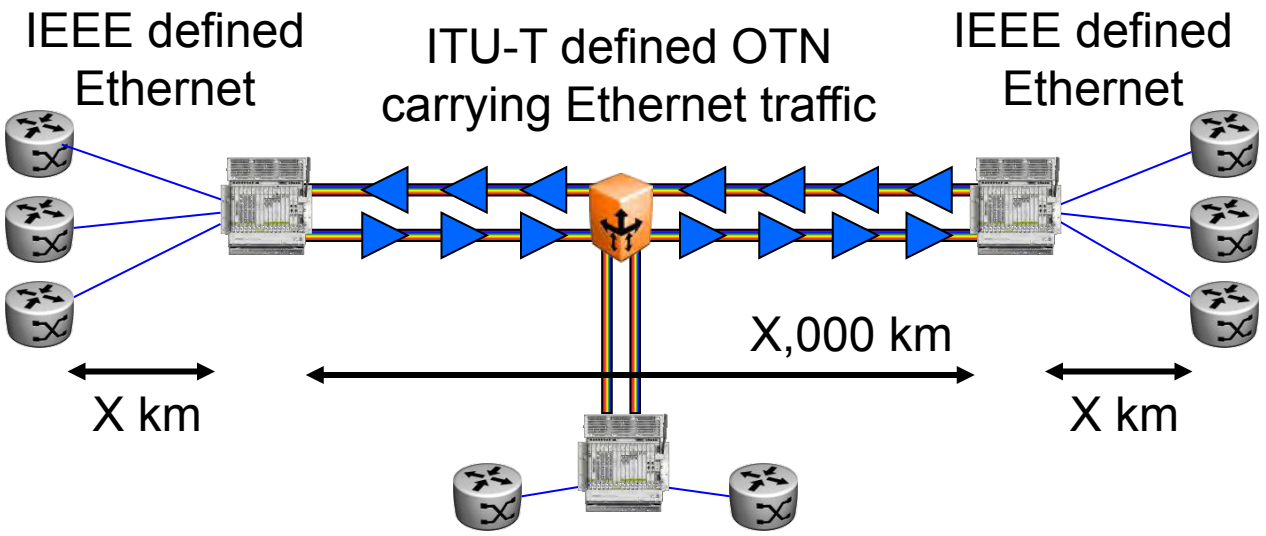


The never ending balancing act!

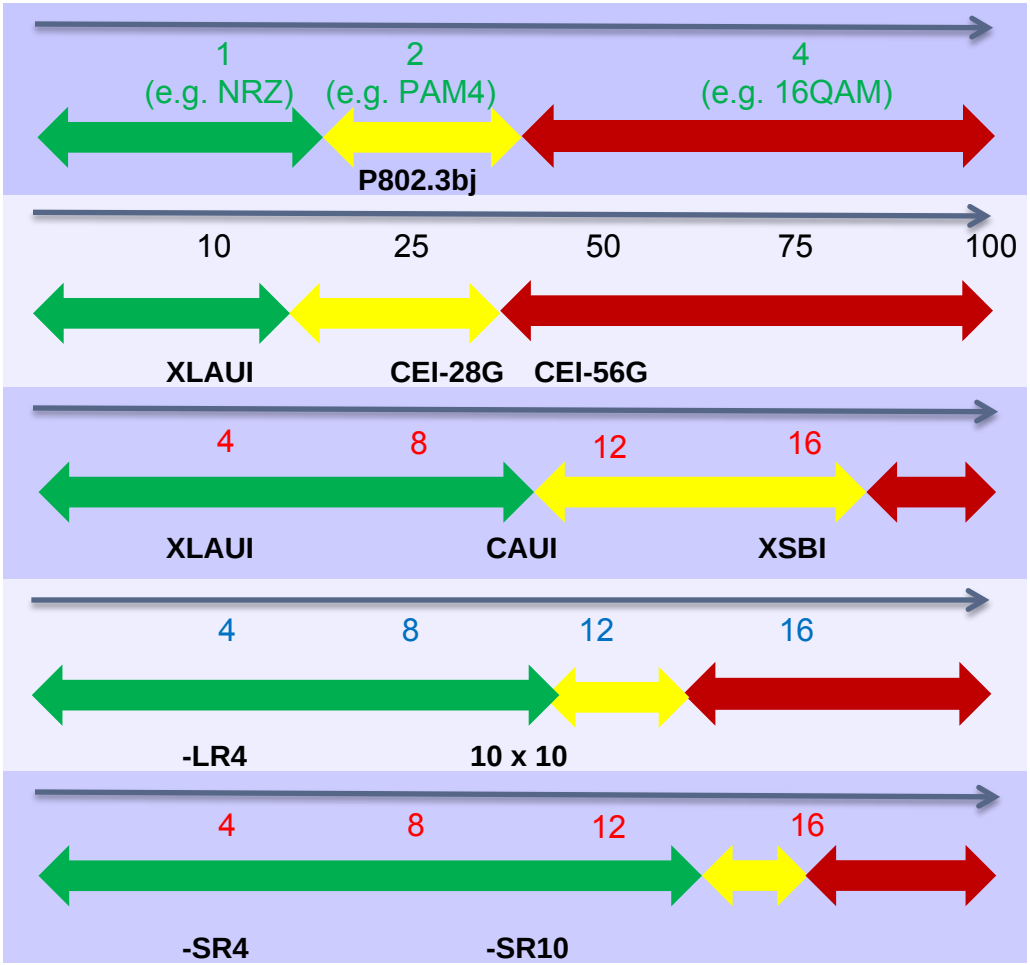


Line versus Client

- **At highest rates Ethernet is becoming dominant traffic for both client and line**
  - **Interdependent problems**
  - **But economics of the line and client are different**
  - **Optimum choice for Ethernet different from that for line**



Potential Technology Axes for increasing Gbit/s in the Optical and Electrical Domains



Modulation
(i.e. Bits per Hz)

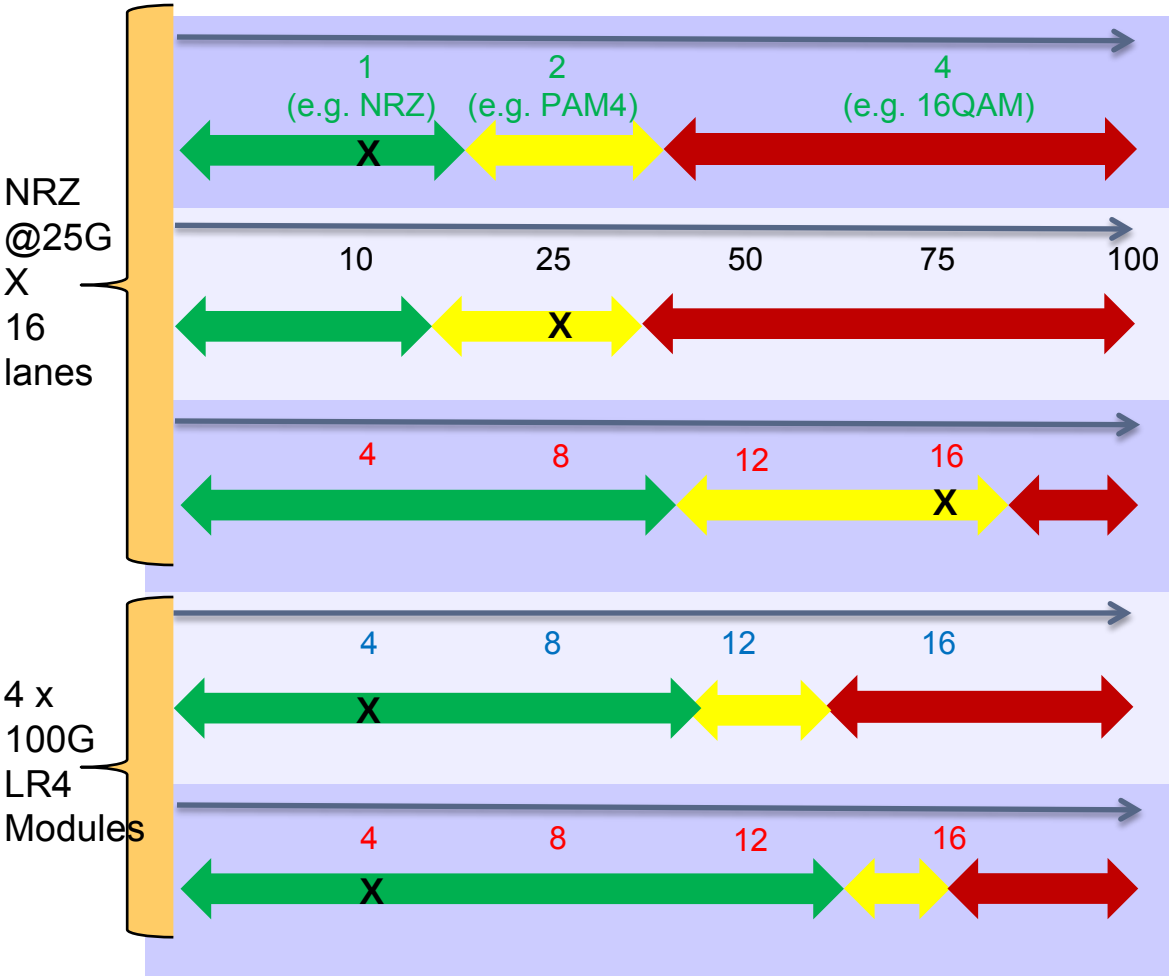
Time Division Multiplexing
(i.e. Baud Rate, Gbps)

Space Division Multiplexing
(i.e. Multiple Electrical Channels)

Wavelength Division
Multiplexing (i.e. λs)

Space Division Multiplexing
(i.e. Multiple Optical Fibers)

Example 1: Finding a Path to 400Gbit



Modulation
(i.e. Bits per Hz)

Time Division Multiplexing
(i.e. Baud Rate, Gbps)

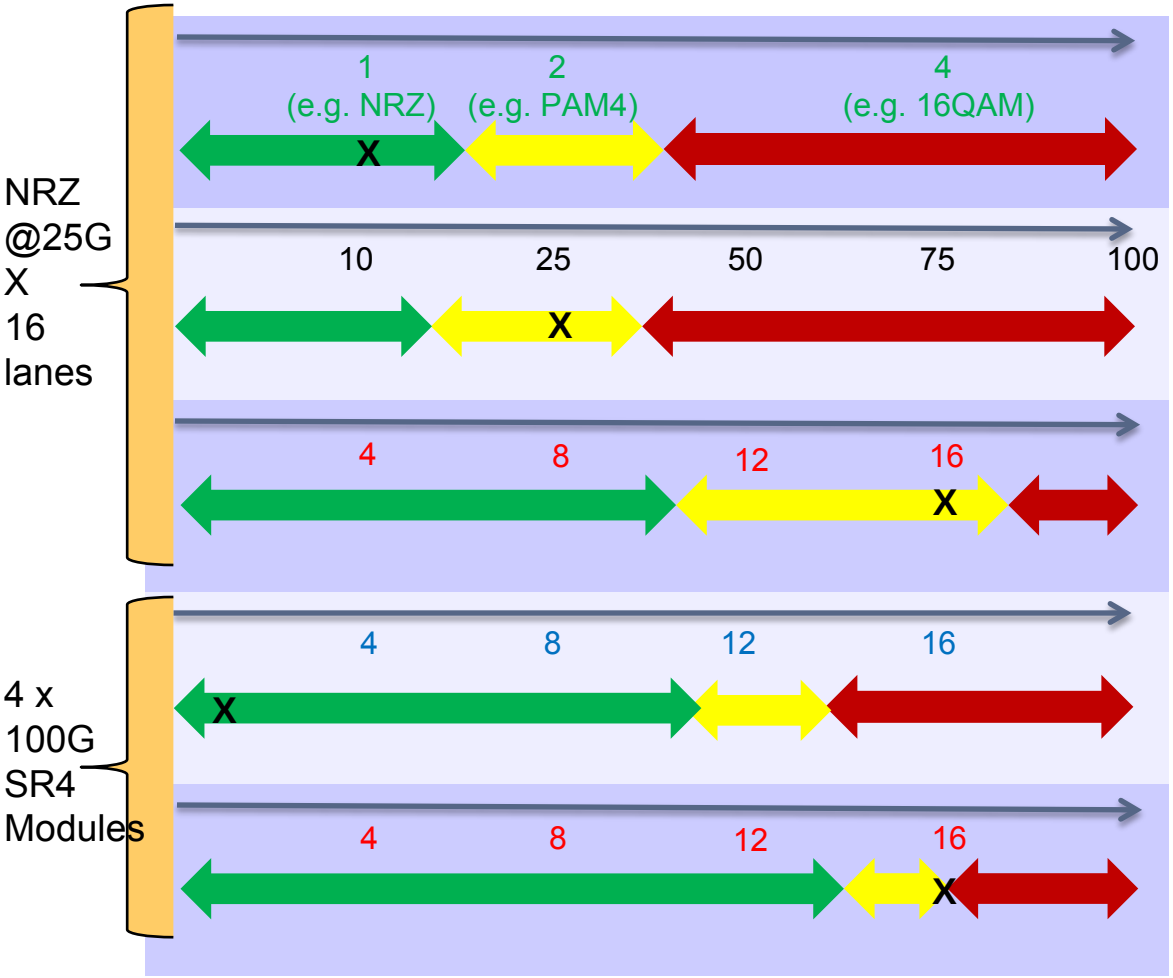
Space Division Multiplexing
(i.e. Multiple Electrical Channels)

Wavelength Division
Multiplexing (i.e. λs)

Space Division Multiplexing
(i.e. Multiple Optical Fibers)

- **Leverages 100GbE building blocks**
- **25 Gb/s : Industry Standards (bj, bm, OIF, IB, FC)**
- **Eye diagrams**
- **16 electrical channels – XSBI, 300 PIN MSA**
- **4 LAMBDA: LR4**
- **4 FIBERS: SR4, PSM4**

Example 2: Finding a Path to 400Gbit



Modulation
(i.e. Bits per Hz)

Time Division Multiplexing
(i.e. Baud Rate, Gbps)

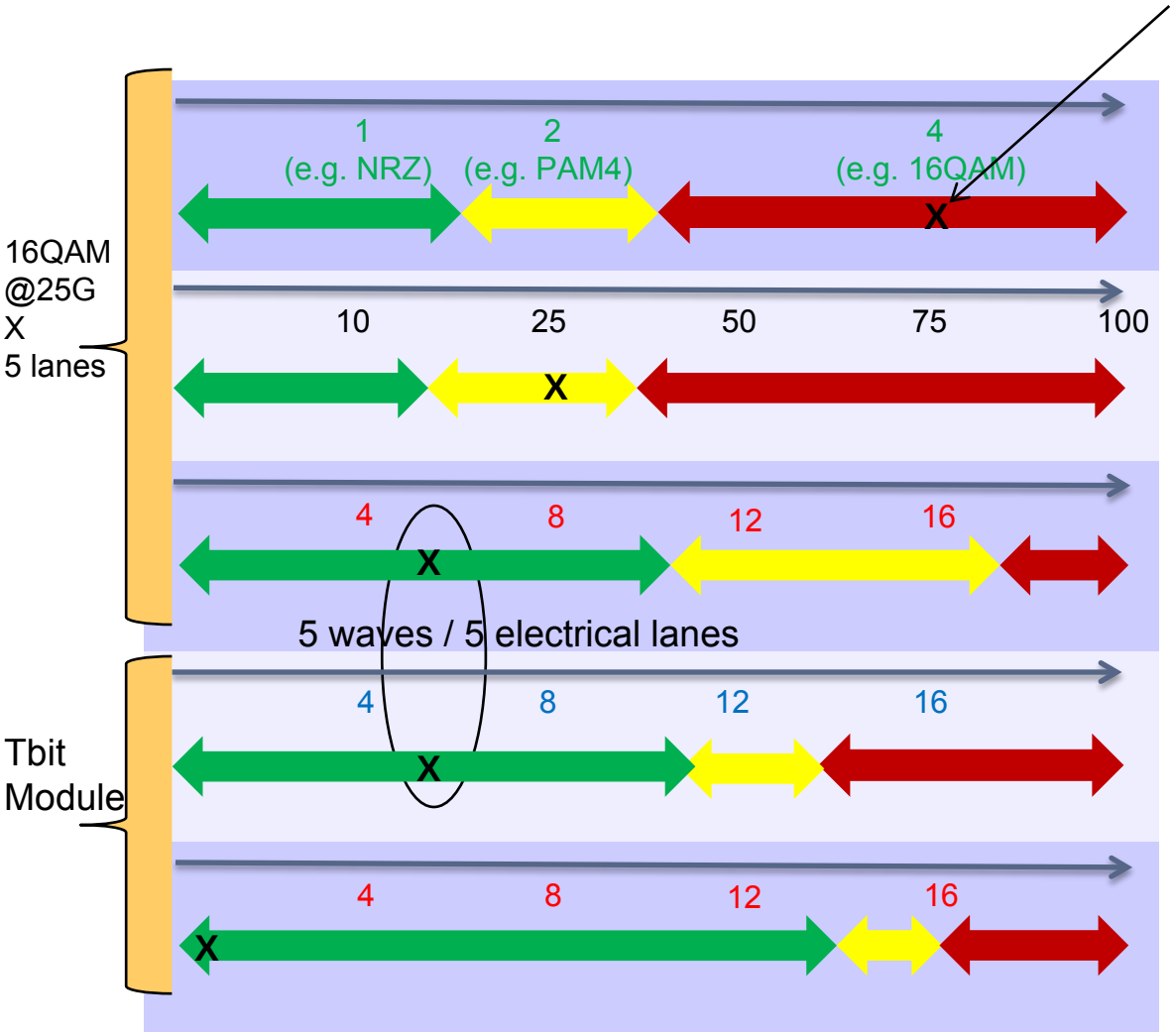
Space Division Multiplexing
(i.e. Multiple Electrical Channels)

Wavelength Division
Multiplexing (i.e. λs)

Space Division Multiplexing
(i.e. Multiple Optical Fibers)

- **Leverages 100GbE building blocks**
- **25g SEE SLIDE 34**
- **PSM-16**

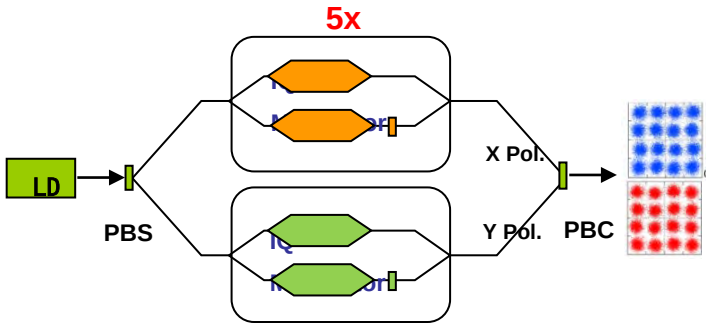
Example 3: Finding a Path to 1Tbit: Not so easy



Options for a 1Tb/s PMD

- **More sophisticated modulation format (amplitude and phase)**
  - ✓ **Re-use of line side technologies for 100G and 400G to minimize the number of lanes**
  - ✓ **Need advanced CMOS IC and PIC technology**

Alternative	Bits/Symbol	Rate, GBaud
5λx200G DP-16QAM	4	25G



400Gb/s MAC Technical Feasibility

- CMOS IC features have shrunk by ~2x since 100Gb/s MAC/PCS was defined in 802.3ba
- CMOS International Technology Roadmap for Semiconductors, 2011 Revision Overview:

- ITRS Sponsoring Industry Associations (IAs): European Semiconductor IA, Japan Electronics and Information Technology Association, Korea Semiconductor IA, Taiwan Semiconductor IA, (US) Semiconductor IA

400Gb/s MAC Technical (2 of 2)

- Typical 100Gb/s MAC/PCS ASIC:
  - 45/40nm CMOS
  - 160b wide bus
  - 644MHz clock
- Potential 400Gb/s MAC/PCS ASIC:
  - 28/20nm CMOS
  - 400b wide bus
  - 1GHz clock
- 400Gb/s MAC/PCS FPGA will be feasible with wider buses and slower clocks

400Gb/s vs. Higher Rates

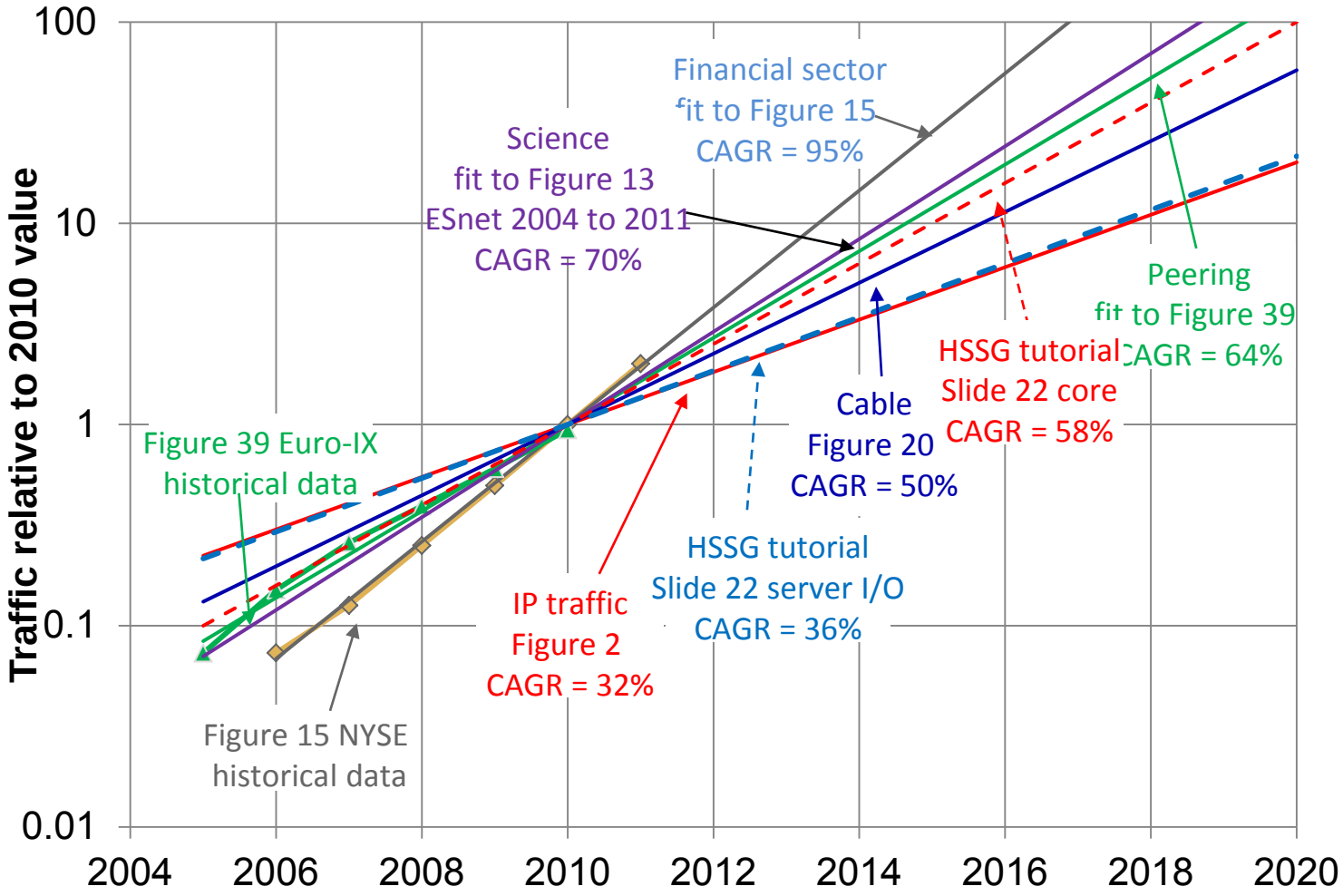
- Customers want parity OR BETTER in W/bit, $/bit, and bits/system
  - Addressing client side application space and cost targets. Must use appropriate solutions.
- Faster interface rates require exotic implementations
  - Not yet competitive per W, per $, or density
  - Higher R&D investment
  - Longer time to market
- 400GbE can reuse 100GbE building blocks
- 400GbE fits in the dense 100GbE system roadmap
- Data rates beyond 400Gb/s require an increasingly impractical number of lanes if 100GbE technology is reused

Presented by
John D'Ambrosia, Dell

IEEE 802.3 Working Group
Orlando, FL, USA
March 19, 2013

400 GIGABIT ETHERNET - WHY NOW?

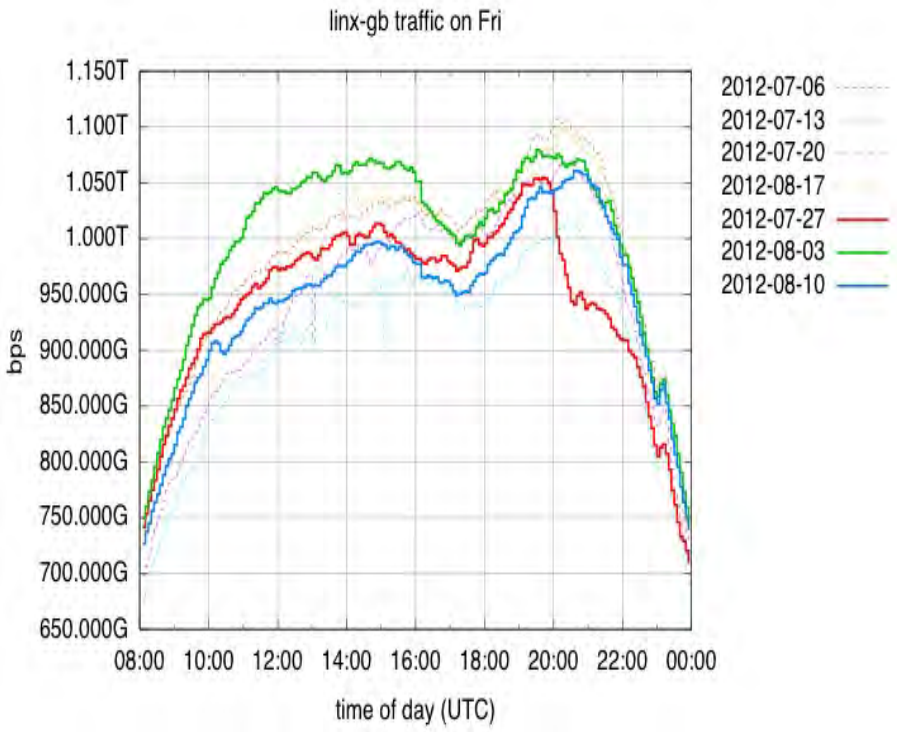
Findings of IEEE 802.3 BWA Ad Hoc



Source: http://www.ieee802.org/3/ad_hoc/bwa/BWA_Report.pdf

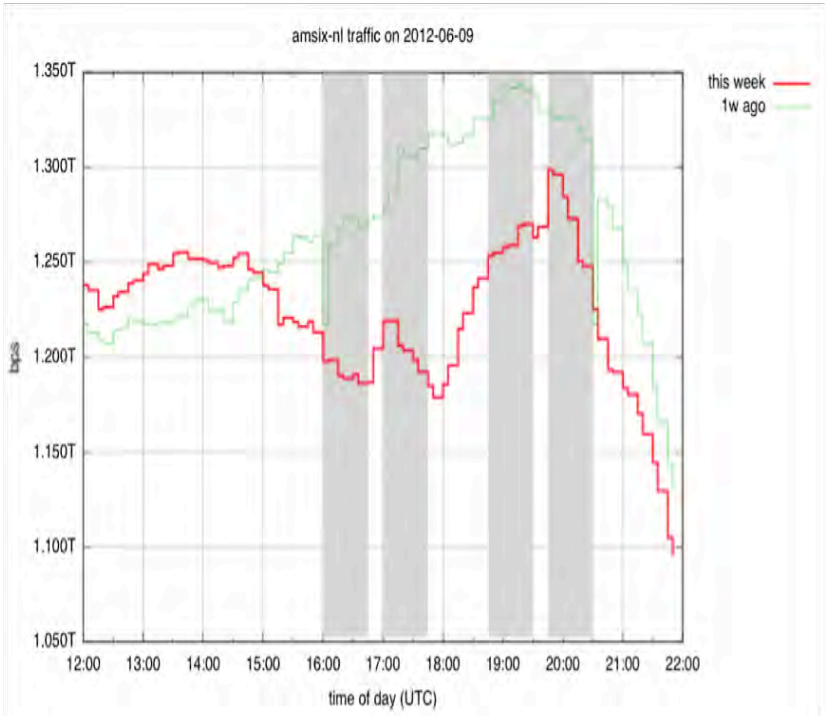
THE FUTURE IS HERE NOW

2012 Summer Olympics



Source: https://labs.ripe.net/Members/fergalc/internet-traffic-during-olympics-2012

After First Round of Euro 2012 Matches



Source: https://labs.ripe.net/Members/fergalc/internet-traffic-after-first-round-of-euro-2012-matches/AMSIXNL.png

Thanks to Bijal Sanghani, Euro-IX.

Other Data

The Need for Higher Speed

- **Traffic is growing everywhere**
  - **More Internet users**
  - **More ways to access the internet faster**
  - **Higher bandwidth content**
  - **New applications enabled**
- **IEEE 802.3 Bandwidth Assessment Adhoc Forecast, on average networks will need to support bandwidth capacities -**
  - **In 2015, 10x requirements of 2010 - Terabit**
  - **In 2020, 100x requirements of 2010 – 10 Terabit**
- **Multiple applications driving the “Bandwidth Tsunami”**

Summary

- The world is changing rapidly with new bandwidth generating applications constantly being introduced
- Bandwidth – exponential growth continues!
- Higher Speed @ lower cost per bit needed throughout the entire Ecosystem
- Past efforts took 3 to 4 years
  - 10 Gigabit Ethernet
  - Ethernet First Mile
  - 40 Gigabit and 100 Gigabit Ethernet
- We need to begin the process and study the problem

Supporters

- **Pete Anslow, Ciena**
- **Chris Cole, Finisar**
- **Kai Cui, Huawei**
- **John D'Ambrosia, Dell**
- **Mark Gustlin, Xilinx**
- **Jeff Maki, Juniper**
- **Andy Moorwood, Infinera**
- **Gary Nicholl, Cisco**
- **Mark Nowell, Cisco**
- **David Ofelt, Juniper**
- **Brian Teipen, ADVA**
- **Steve Trowbridge, Alcatel-Lucent**
- **xxxx**

STRAW POLLS

Call-For-Interest

- **Should a Study Group be formed for “ 400 Gb/s Ethernet”?**

  Y: **N:** **A:**

Participation

- **I would participate in the “400 Gb/s Ethernet” Study Group in IEEE 802.3.**

  Tally: xx

- **My company would support participation in the “400 Gb/s Ethernet” Study Group in IEEE 802.3**

  Tally: xx

Future Work

- **Ask 802.3 on Thursday**
  - **Form 400 Gb/s Ethernet SG**
  - **Pending approval of 400 Gb/s Ethernet SG, Request that 802.3 WG request 802 EC HSE Consensus Ad Hoc be disbanded.**
- **If approved, on Friday**
  - **Request 802 EC informed of 400 Gb/s Ethernet SG**
  - **Pending approval of 400 Gb/s Ethernet SG, Request that 802 EC HSE Consensus Ad Hoc be disbanded.**
  - **First 400 Gb/s Ethernet SG meeting, week of May 2013 IEEE 802.3 Interim.**

THANK YOU!