THE 2015 ETHERNET ROADMAP

SC 2015 BOF Panel

Scott Kipp, Brocade, President of the Ethernet Alliance

David Chalupsky, Intel

Shreyas Shah, Xilinx

Brandon Hoff, Avago

November 18, 2015

ethernet alliance

Disclaimer

- Opinions expressed during this presentation are the views of the presenters, and should not be considered the views or positions of the Ethernet Alliance.

2015 ETHERNET ROADMAP

As shown on the long and winding road, Ethernet could have 12 speeds before 2020 with 6 new speeds introduced in the next 5 years. The progression of speeds is not in chronological order because 40GbE and 100GbE were primarily based on multiple lanes of 10Gb/s technology that was available before 25Gb/s serial technology enabled 25GbE. Lanes running at 25Gb/s are becoming practical in 2015 and will be used in 25GbE SFP+ and 4x25Gb/s 100GbE QSFP28. The next serial lane speed is expected to be 50Gb/s and enable 50GbE SFP28, 200GbE QSFP28 (4X50G) and 400GbE CFP2 (8X50G).

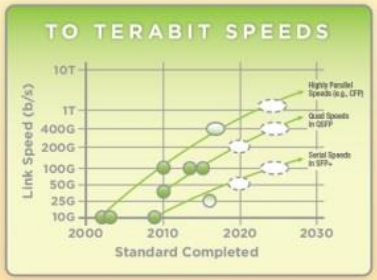
Beyond 400GbE, the map shows the unknown distant future that will become clearer as we approach 2020. Terabit links are expected when single lanes can be modulated at 100Gb/s and grouped into 10 or 16 lanes to form TbE or 1.6TbE. Significant investments in technology are needed before 100Gb/s lanes are economically feasible.

Low cost 100Gb/s lane technology that can fit in an SFP+ is not expected to be available until after 2020. The Ethernet Alliance will award the first company that produces a 100GbE SFP+ with the Holy Grail of the 100GbE SFP+.

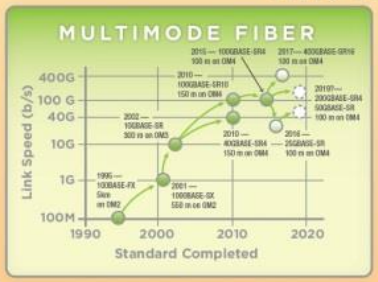
The twisted pair or BASE-T roadmap in the lower right corner of the map shows how 10GBASE-T technology is being used in 4 new speeds — 2.5, 5, 25 and 40Gb/s. All four of these speeds are expected to be standardized in 2016 but they are targeting different cabling infrastructure. 2.5 and 5GBASE-T are being designed for Cat 5e cabling up to 100 meters while 25 and 40GBASE-T are being designed for 30 meters of Cat 8 cabling.

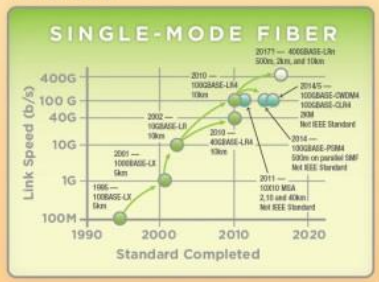
10 Tb/s
1.6 Tb/s >2020
6.4 Tb/s >>2020
1 Tb/s >2020
800 Gb/s >2020
400 Gb/s 2017 (est)
200 Gb/s 2018-2020 (est)
100 Gb/s 2010
50 Gb/s 2018-2020 (est)
40 Gb/s 2010
25 Gb/s 2016 (est)
10 Gb/s 2002
5 Gb/s 2016 (est)
2.5 Gb/s 2016 (est)
1 Gb/s 1998
100 Mb/s 1995
10 Mb/s 1983

Ethernet Speed
Speed in Development
Possible Future Speed



FIBER OPTIC LINKS





COPPER LINKS

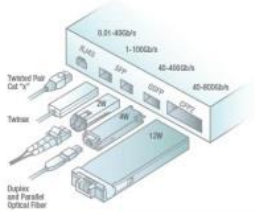
To get a pdf version of the roadmap and to find out more about the roadmap, please go to: **www.ethernetalliance.org/roadmap/**

MEDIA AND MODULES

Ethernet is wired technology and supports a variety of media including backplanes, twisted pair, twinax, multimode fiber and single-mode fiber. Most people know Ethernet by the twisted pair or Cat "x" cabling with RJ45 connectors because close to a billion ports a year are sold. Cat 8 is the latest generation of twisted pair cabling that will be used in 25GBASE-T and 40GBASE-T.

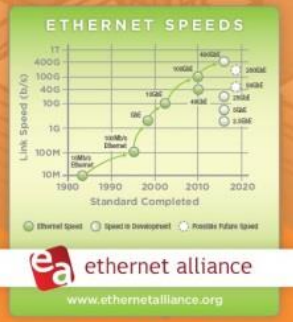
Another popular copper interface is Twinax copper cables that are also known as direct attach cables (DAC)s. DACs may be passive or active and provide very low cost connectivity to servers. Passive DACs are limited to 25 meters or less while active optical cables can go hundreds of meters.

For links longer than 100 meters, fiber optics are required and the graphic below shows three of many module types. The SFP family is the most popular module and supports a single channel or lane in each direction and duplex fibers. The QSFP family supports 4 channels while the CFP2 supports up to 10 channels and duplex or parallel fibers. For 40GbE and beyond, the electrical interface to the module is being defined in IEEE and supports a variety of optical interfaces from IEEE and other sources.



2015 ETHERNET ROADMAP

THE PAST, PRESENT AND FUTURE OF ETHERNET



ethernet alliance
www.ethernetalliance.org

Designed by Scott Kipp

$9.95

ETHERNET ECOSYSTEM

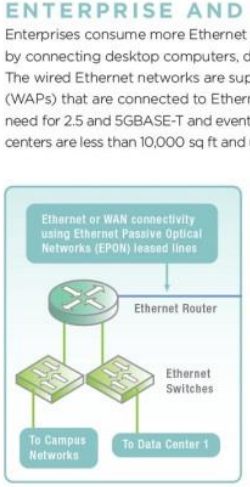
Represented as a city, the Ethernet Ecosystem is divided into four quadrants that are interconnected by multiple MANs that are typically not Ethernet. While each quadrant has overlapping technologies and requirements, this map organizes the environments with a broad brush. Specific implementations may vary considerably.

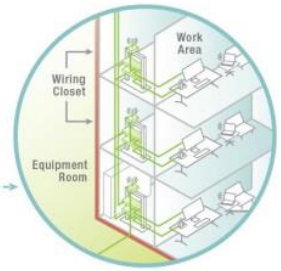
The top half of the map represents applications where cost and connectivity are driving concerns. In the home, small office and car, link distances are less than 100 meters and speeds are typically under 10Gb/s, so copper cabling and wireless are ideal. As enterprises scale in size and requirements, they shift towards fiber and 10Gb/s speeds and beyond.

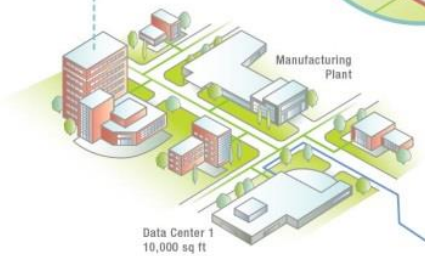
The lower half of the map captures applications that consistently push the bounds of Ethernet and require higher speeds and massive scalability. For example, service providers and hyperscale data centers will be the early adopters of 400GbE. These users may deploy hundreds of thousands of servers in data centers that span multiple football fields and consume hundreds of megawatts of power.

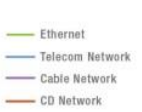
ENTERPRISE AND CAMPUS

Enterprises consume more Ethernet ports than the other environments by connecting desktop computers, devices and Voice over IP (VoIP) phones. The wired Ethernet networks are supplemented with wireless access points (WAPs) that are connected to Ethernet cables. 802.11ac WAPs are driving the need for 2.5 and 5GBASE-T and eventually 10GBASE-T. Most enterprise data centers are less than 10,000 sq ft and use Cat "x" cabling to connect to servers.



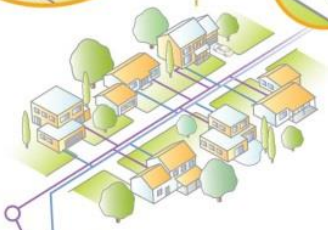


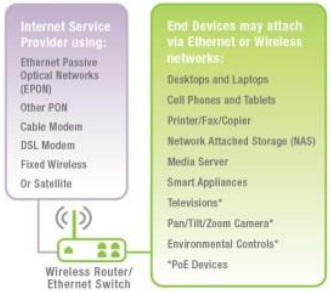




RESIDENTIAL AND CONSUMER

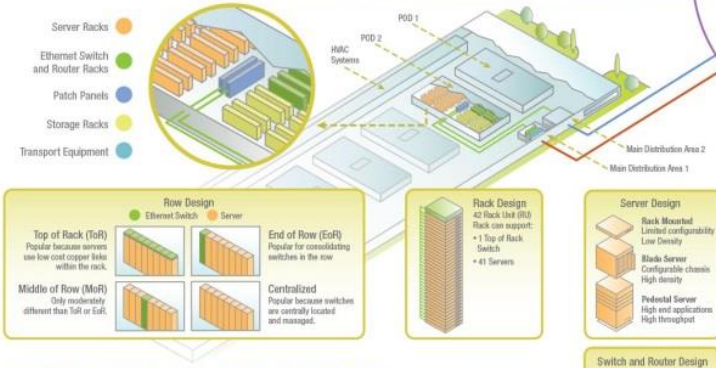
Ethernet Passive Optical Networking (EPON) delivers Internet service to millions of residential customers around the world. Regardless of how the Internet reaches the home, residents may wire their home with Ethernet or use wireless connectivity to connect devices. From cameras to cars, Ethernet provides the network to enable sharing resources and content.





BACKBONE TO OTHER CITIES

BACKBONE TO OTHER CITIES

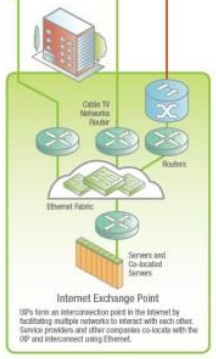


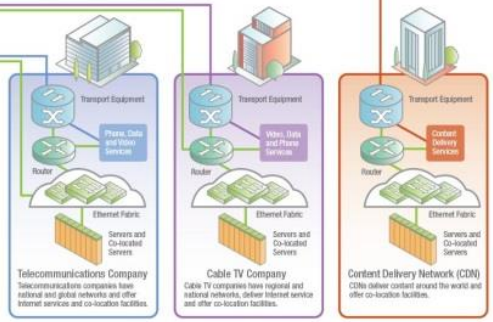
HYPERSCALE DATA CENTER

Hyperscale data centers, also known as warehouse scale computing and mega data centers, are known by their massive size and scalability. Cloud service providers, large enterprises and service providers pack over 100,000 servers that are often divided into several pods. Thousands of 25GbE servers and eventually 50GbE servers in these data centers drive the need for 400GbE to the MAN and WAN.

MANs

Metropolitan Area Networks (MANs) come in many varieties and deliver services to a variety of enterprises, organizations and consumers. Some MANs are based on Ethernet, but the largest MANs are based on Optical Transport Networks (OTN) technologies.

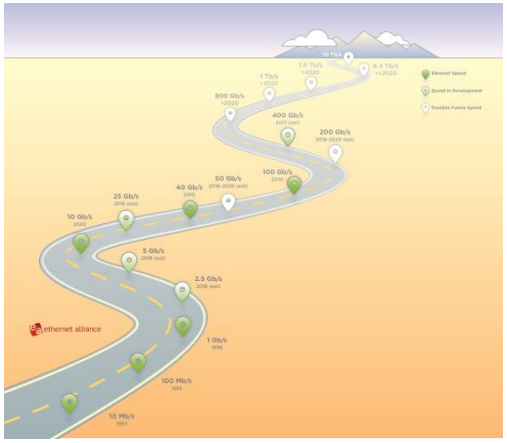


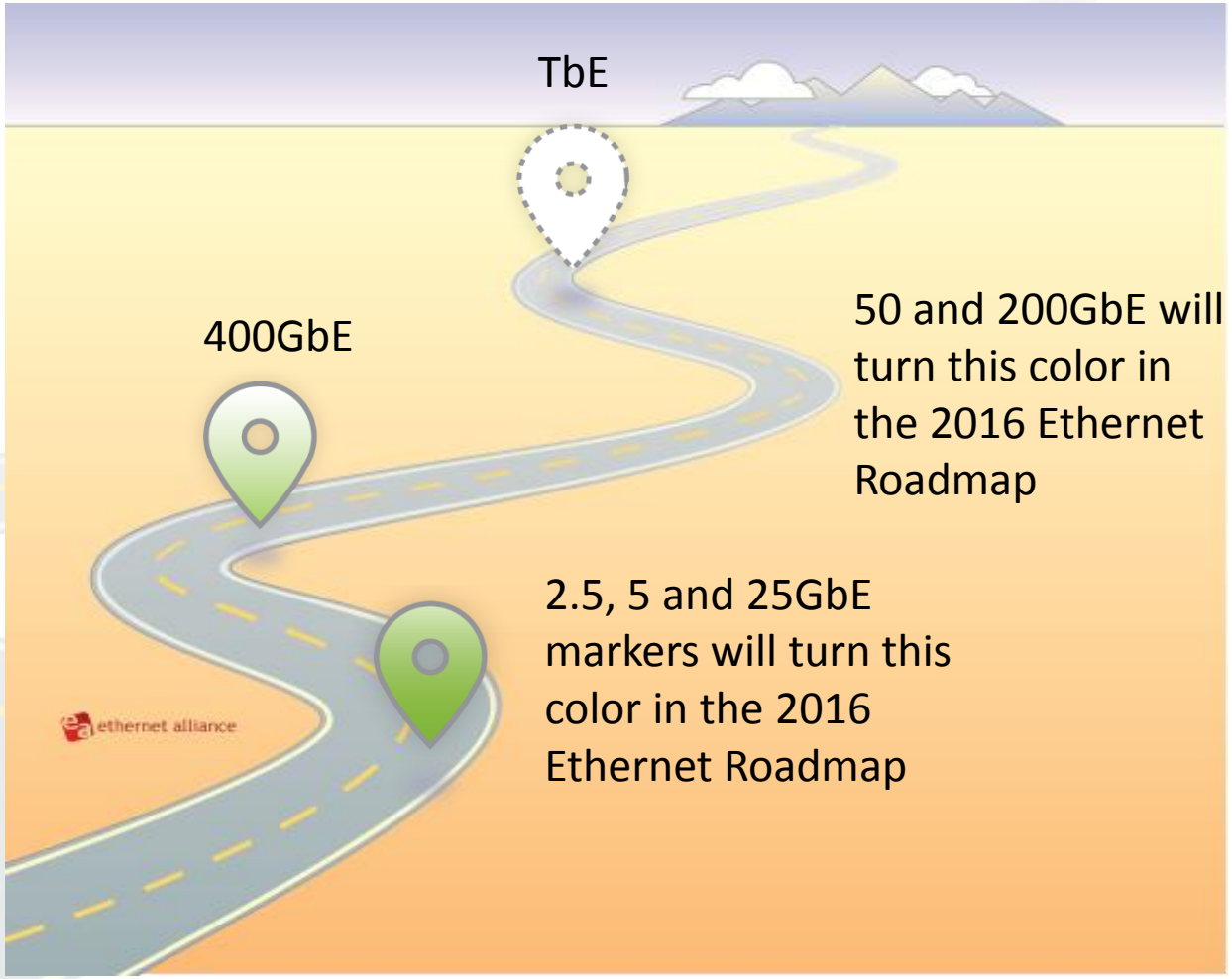


SERVICE PROVIDERS

Service providers deploy MANs and Wide Area Networks (WANs) to deliver a variety of services including Carrier Ethernet. Service providers may use Ethernet Passive Optical Networks (EPON) and cable companies send EPON Protocol over Coax (EPOC). Service providers use routers to interconnect various networks.

The Long and Winding Road



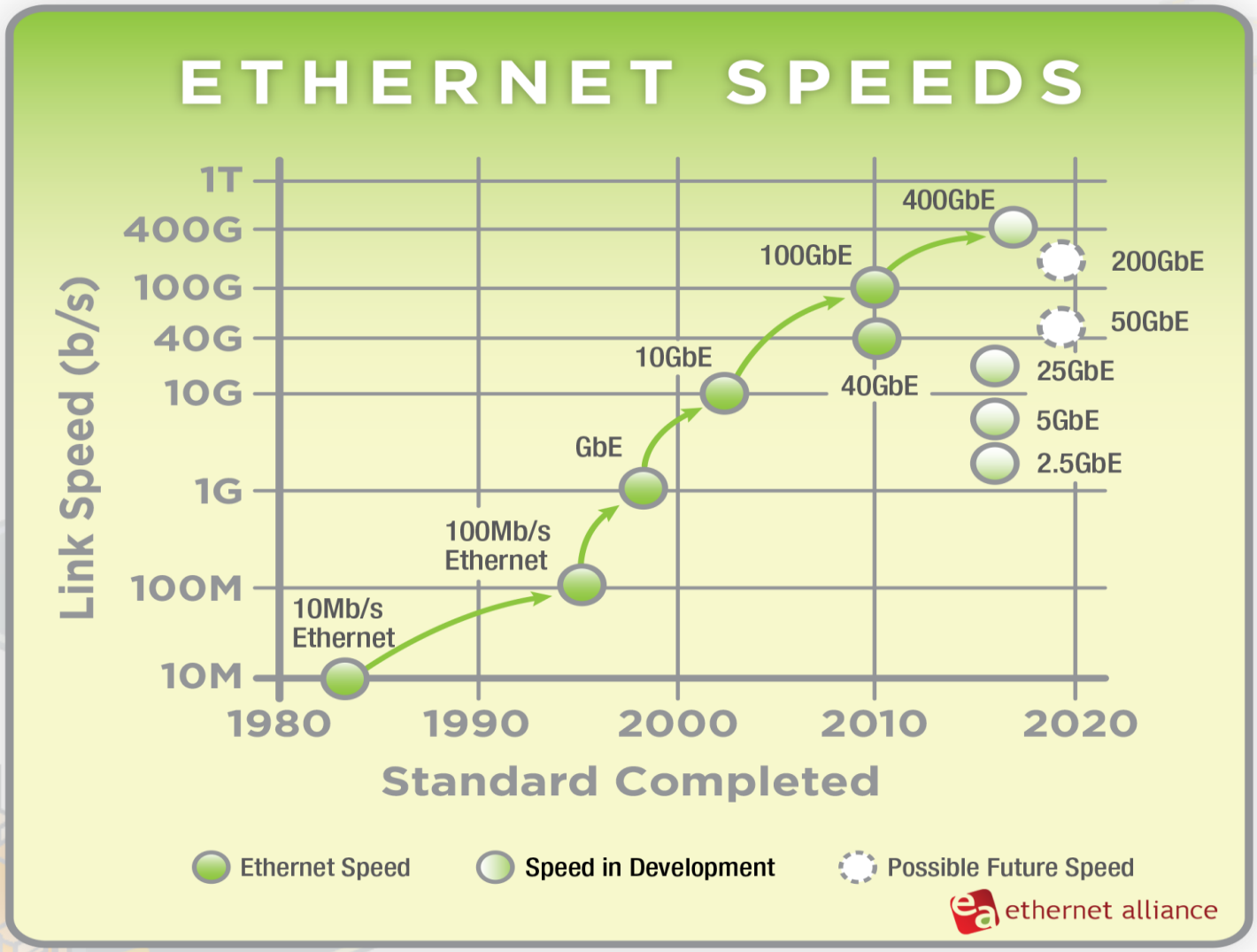


Ethernet Speed

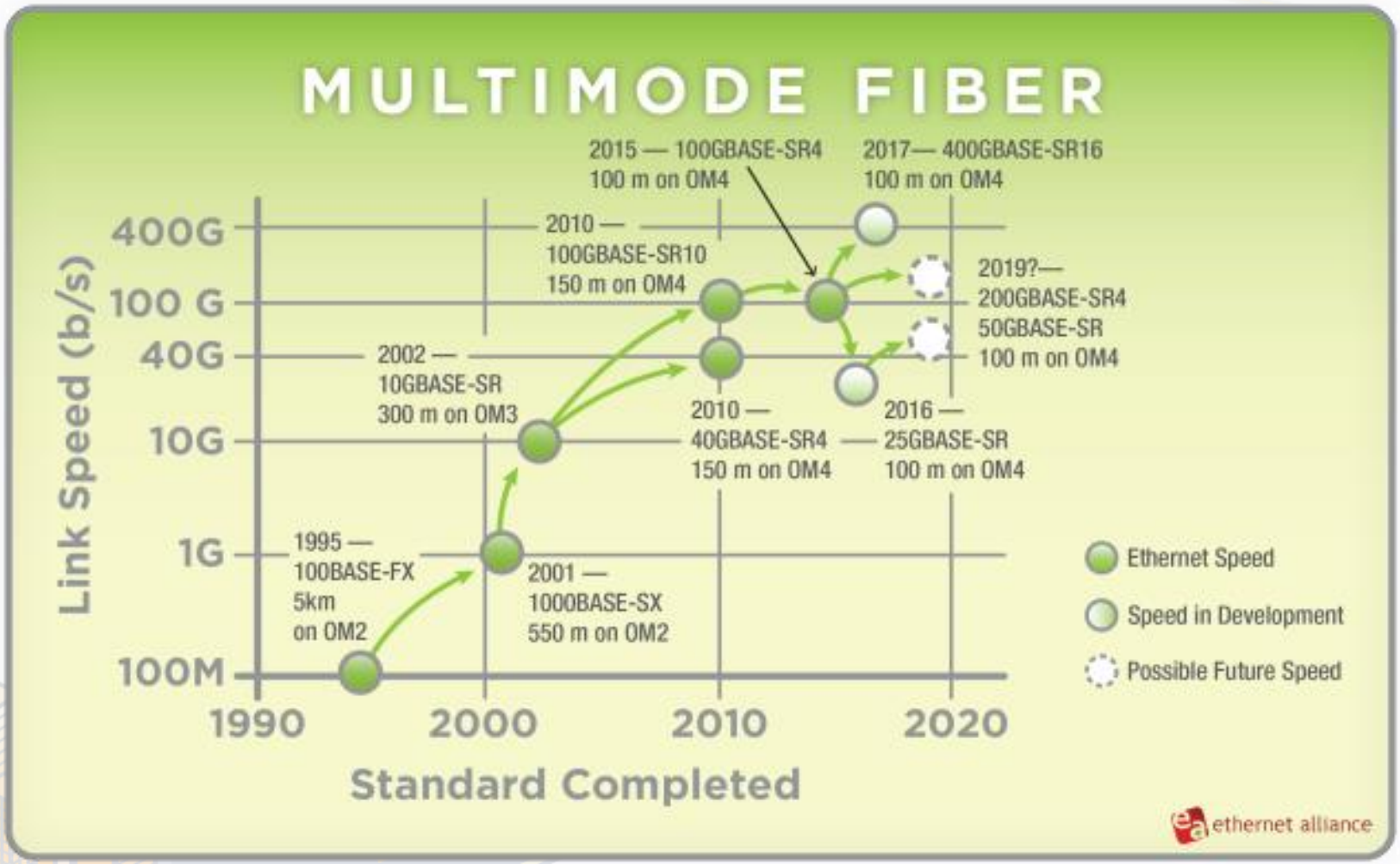
Speed in Development

Possible Future Speed

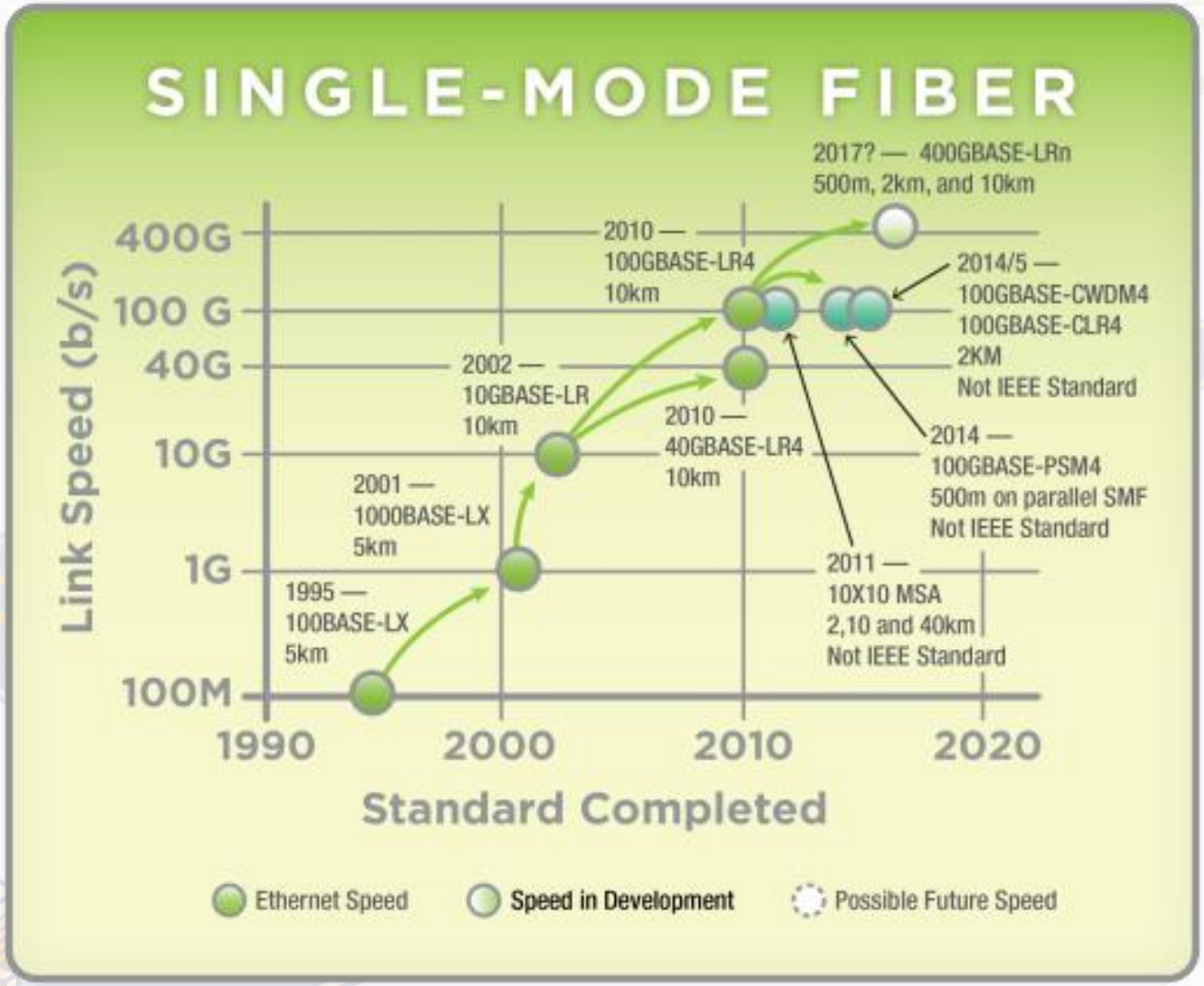
A Dozen Ethernet Speeds



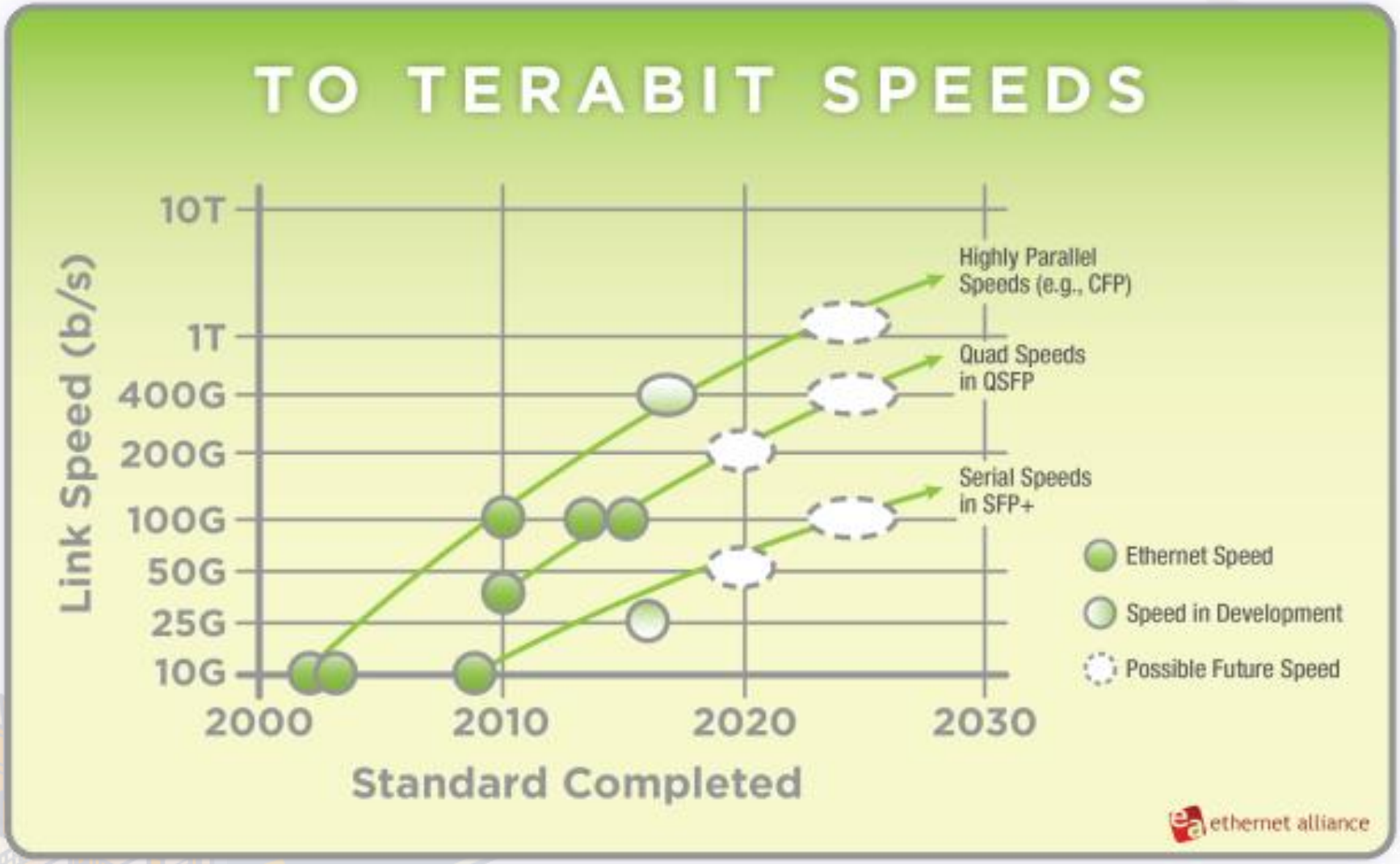
Optical Fiber Roadmaps



Optical Fiber Roadmaps



What's Really Going on?



More Roadmap Information

- See me for a free map after the panel
- Free downloads at www.ethernetalliance.org/roadmap/
  - Pdf of map
  - White paper
  - Presentation with graphics for your use
- Free maps and T-shirts at Booth #1219

COPPER CONNECTIVITY IN THE 2015 ETHERNET ROADMAP

David Chalupsky

SC 2015 BOF Panel

November 18, 2015

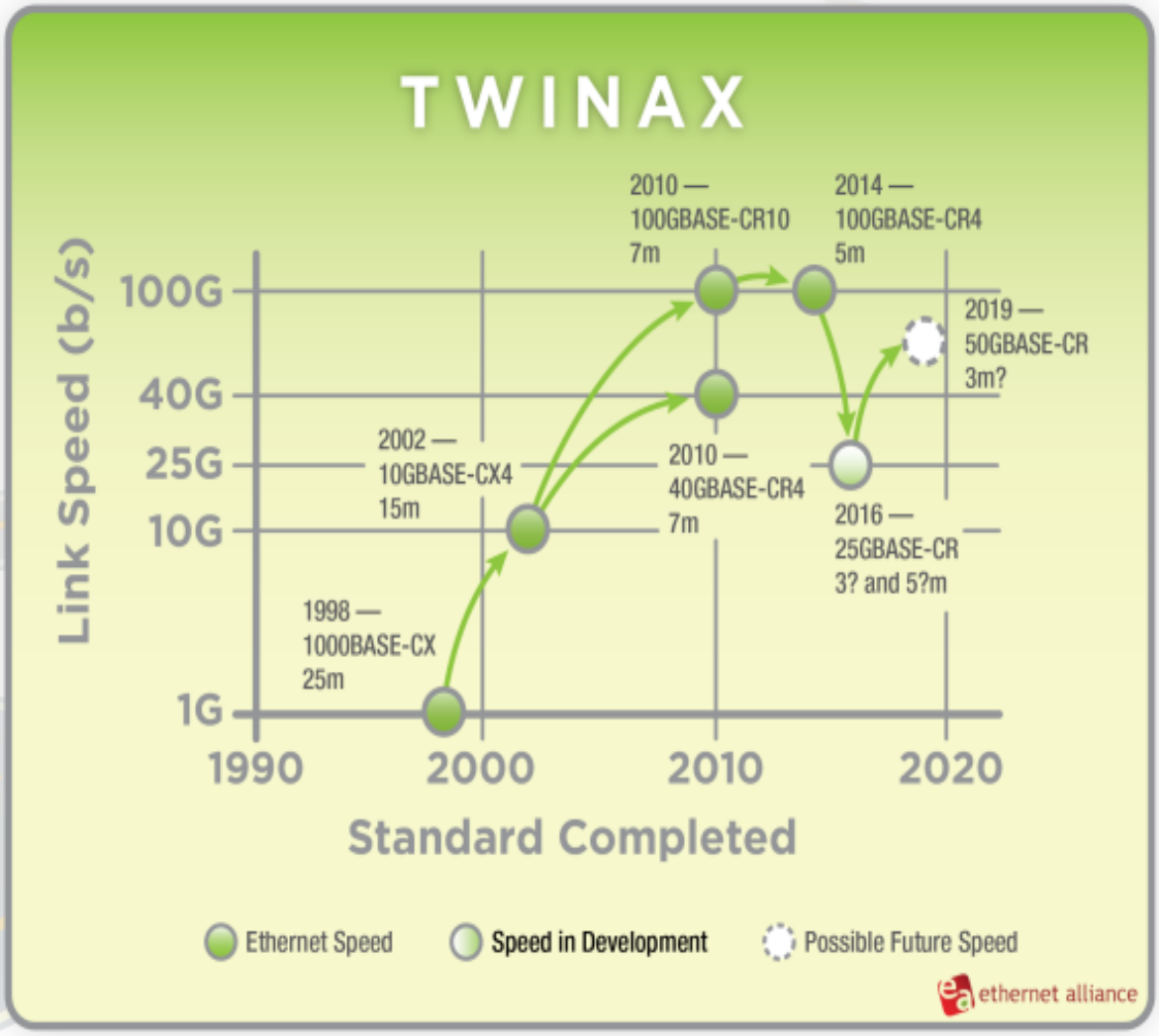
ethernet alliance

Current IEEE 802.3 Copper Activity

- High Speed Serial
  - P802.3by 25Gb/s TF: twinax, backplane, chip-to-chip or module.
  - P802.3bs 400Gb/s TF: 50Gb/s lanes for chip-to-chip or module. (PAM4)
  - P802.3cb 2.5G/ 5Gb/s TF: Backplane and Copper Cables
  - 50 Gb/s Ethernet over a single lane Study Group
  - Next generation 100 Gb/s and 200 Gb/s Ethernet Study Group
- Twisted Pair (4-pair)
  - P802.3bq 25G/40GBASE-T
  - P802.3bz 2.5G/5GBASE-T
- Single twisted pair for automotive
  - P802.3bp 1000BASE-T1
  - P802.3bw 100BASE-T1
- PoE
  - P802.3bt – 4-pair PoE
  - P802.3bu – 1-pair PoE

Twinax Copper Roadmap

- 10G SFP+ Direct Attach: highest volume 10GbE server port today
- 40GBASE-CR4 and 100GBASE-CR4 entering the market
- 25GBASE-CR products entering the market ahead of standard completion
- 50Gb/lane standards development starting now
  - 50GbE in SFP
  - 200GbE in QSFP

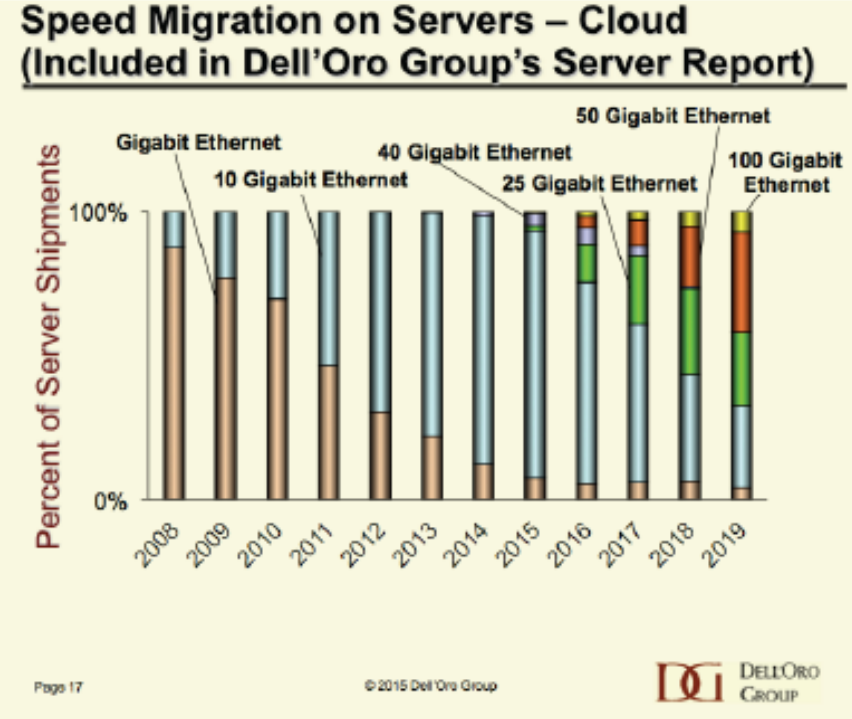


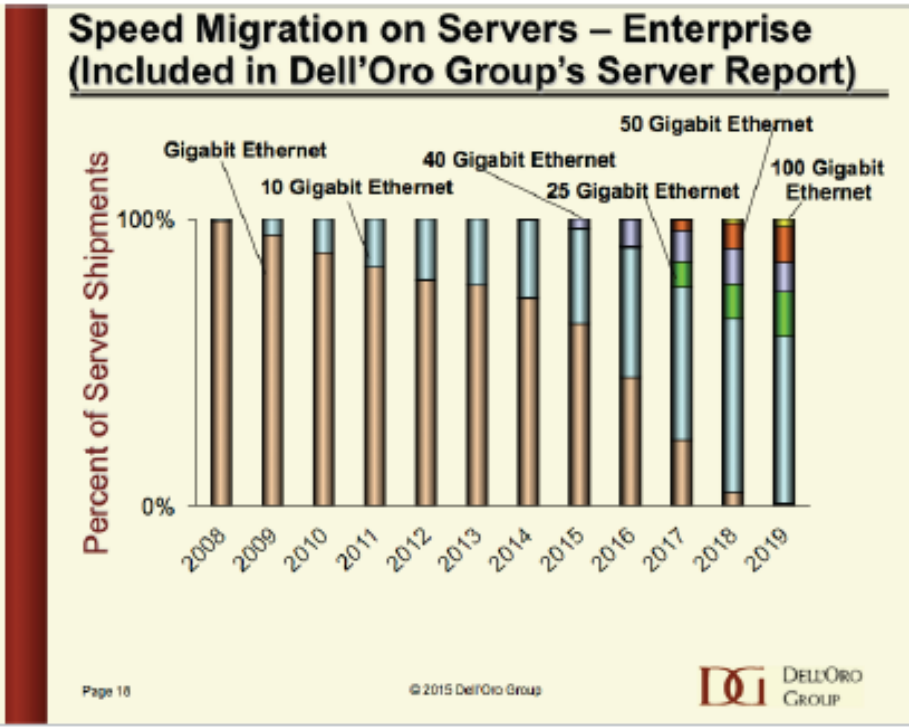
Server Diversification and Port Speed

- Divergence between Cloud and Enterprise
  - Cloud transition to 10GbE has passed, pushing fast to 25G, 50G and more
  - Enterprise servers still making the transition to 10GbE
- Accelerated investment into higher speed Ethernet for Cloud will make low cost 50G/100G/200GbE available to HPC
  - Memory and Storage bandwidth increases need a faster network
  - PCIe Gen4 will enable 200GbE in a single slot

Server Port Speed Forecasts

50G Server forecasts



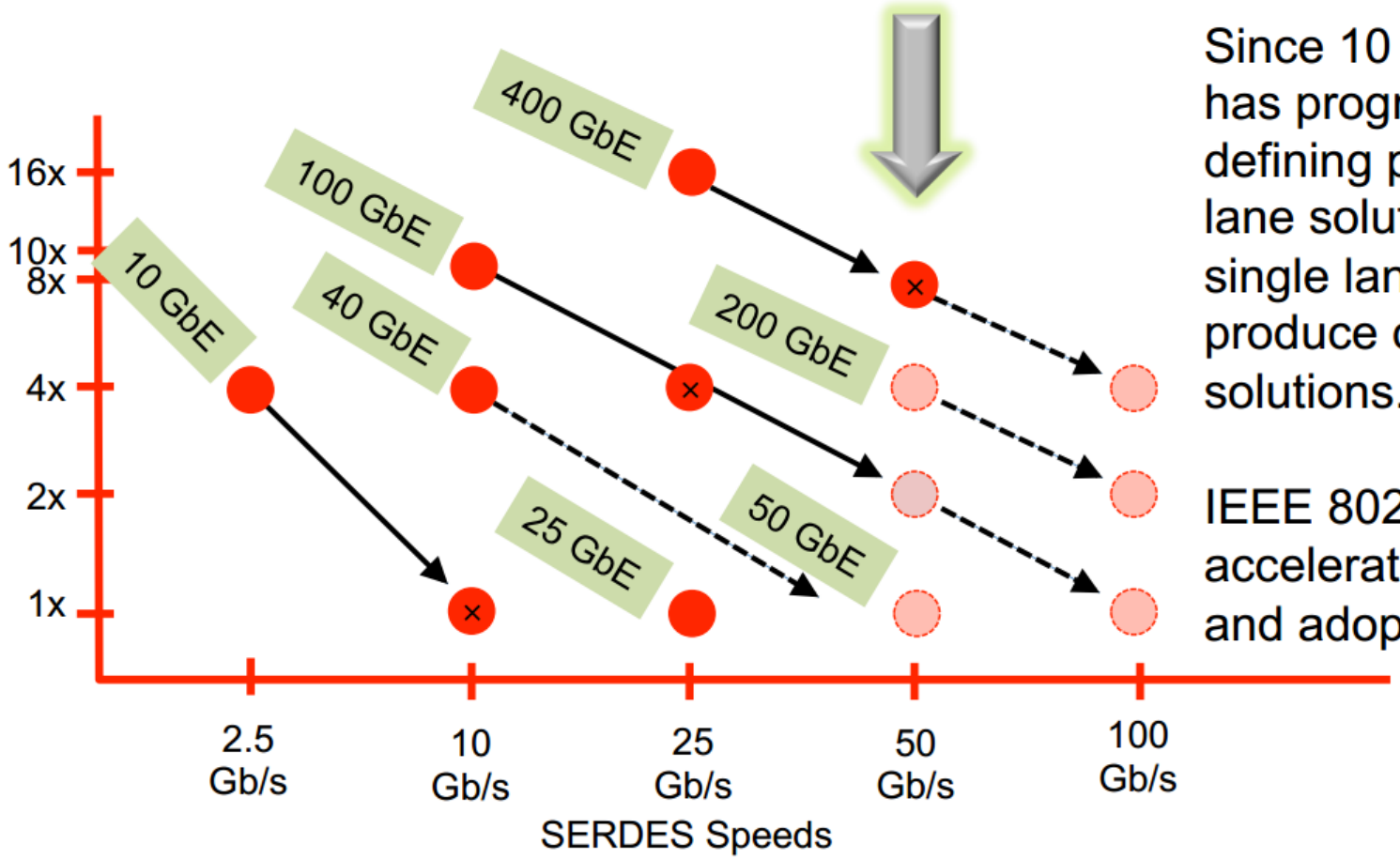


Estimated 50G server port forecast for 2019 > 7M

Courtesy: Dell'Oro

Source: http://www.ieee802.org/3/cfi/1115_1/CFI_01_1115.pdf

The new normal – multi-lane and re-use



Since 10 GbE, Ethernet has progressed by defining pragmatic multi-lane solutions and fastest single lane technologies to produce cost-effective solutions.

IEEE 802 definition accelerates market focus and adoption.

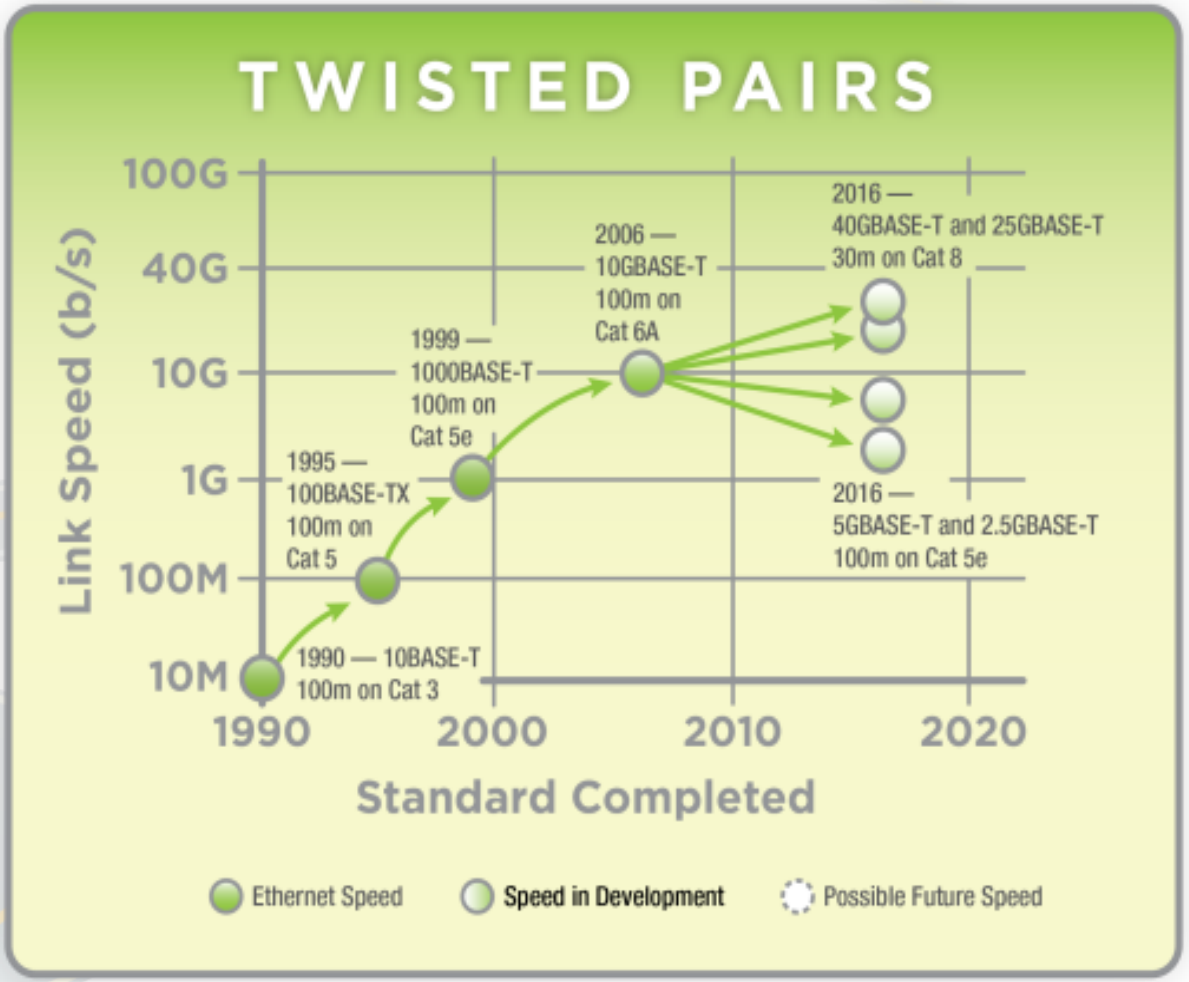
Source: http://www.ieee802.org/3/cfi/1115_1/CFI_01_1115.pdf

BASE-T Copper Roadmap

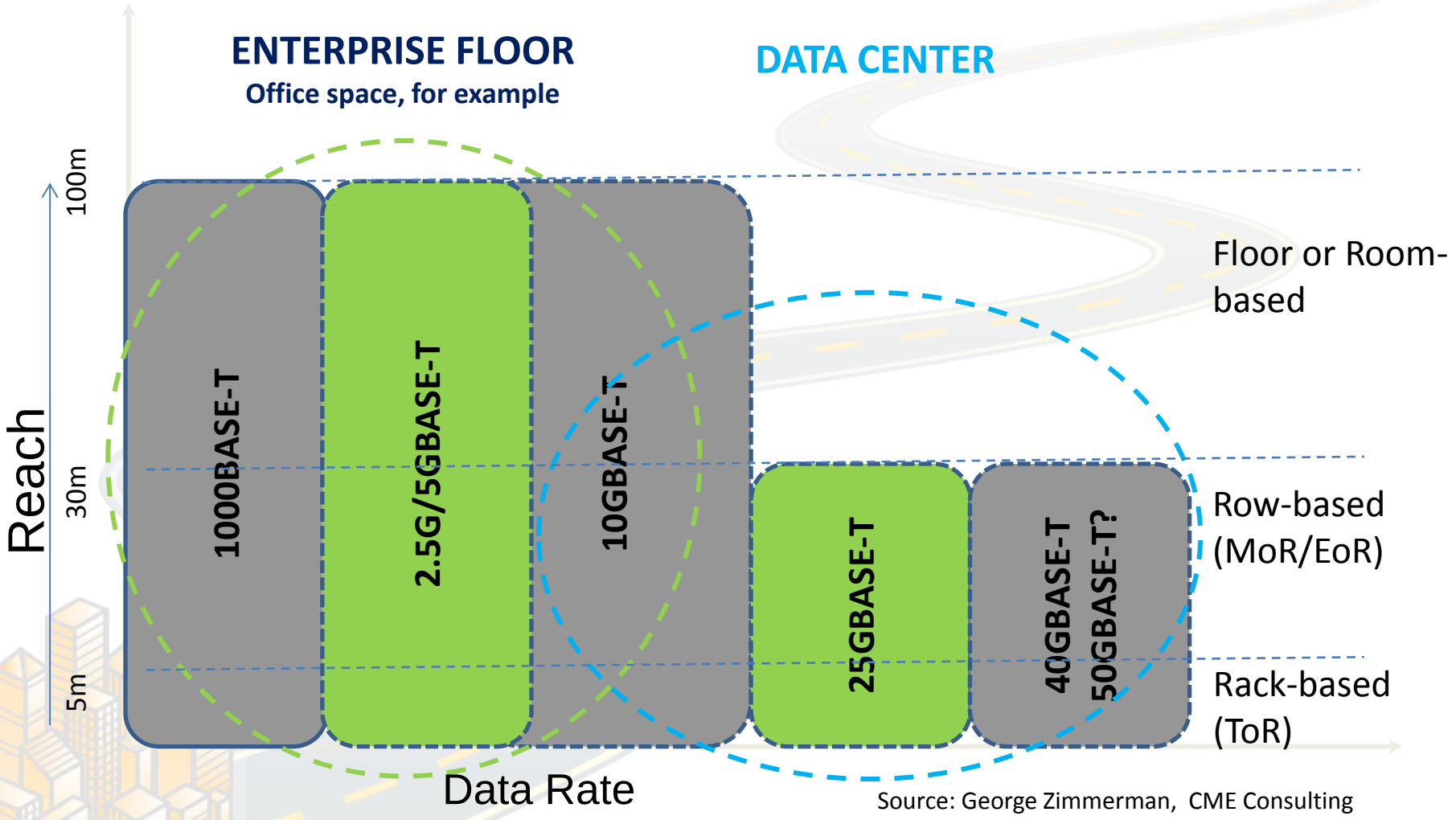
- 1000BASE-T still ~75% of the 47M server ports shipped in 2014

Optimizing for the Future

- Enterprise Data center:
  - 10G->25G->40GBASE-T roadmap on compatible infrastructure
- Enterprise horizontal spaces:
  - 2.5G/5GBASE-T squeeze more usable bandwidth from the 70B meters of Cat5e/6 cabling sold in the last 10 years



The Application Spaces of BASE-T



Source: George Zimmerman, CME Consulting

1 TB/S PORT IN 2025

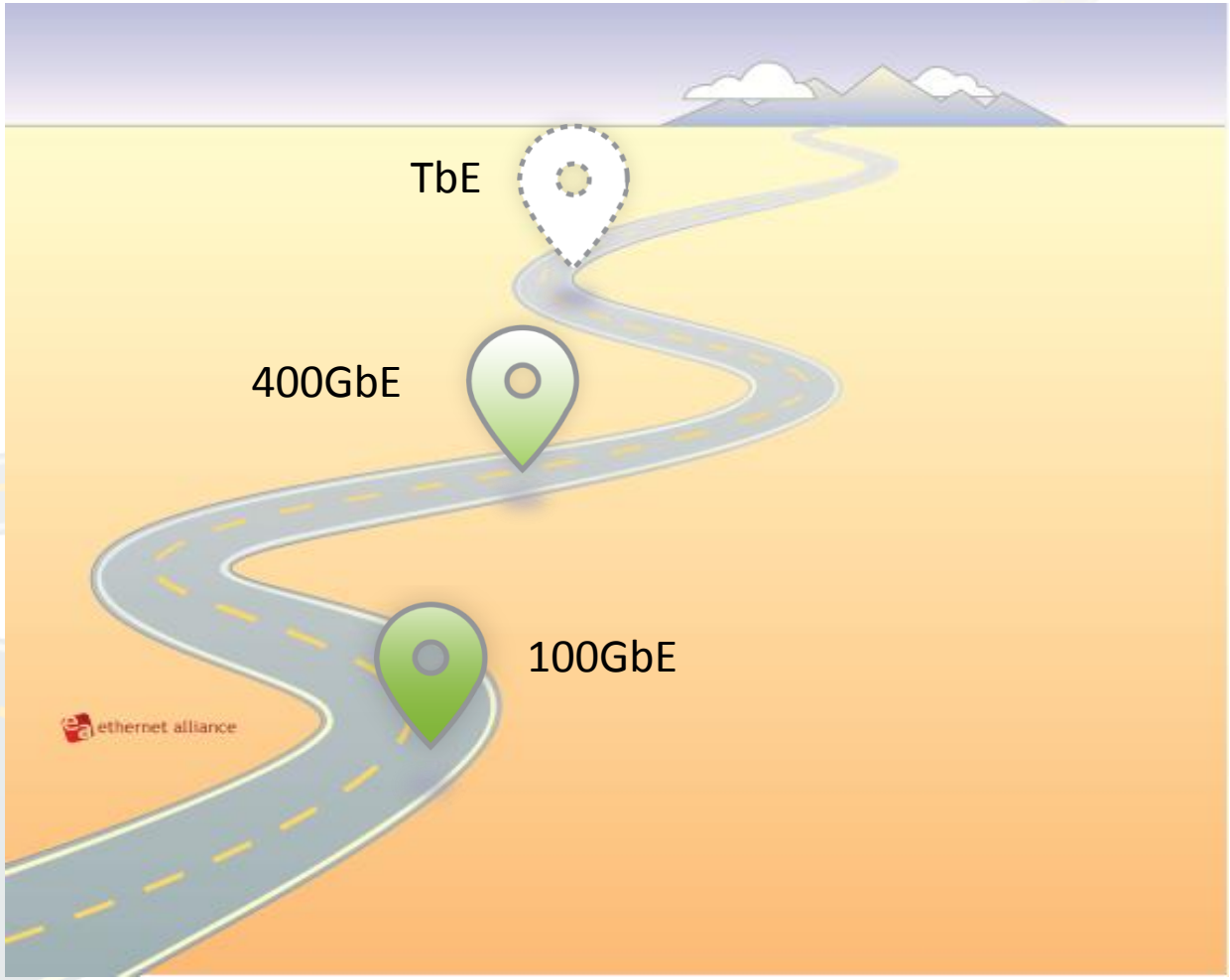
Shreyas Shah

SC 2015 BOF Panel

November 18, 2015

ethernet alliance

The Long and Winding Road



Ethernet Speed

Speed in Development

Possible Future Speed

Trends in the Market:

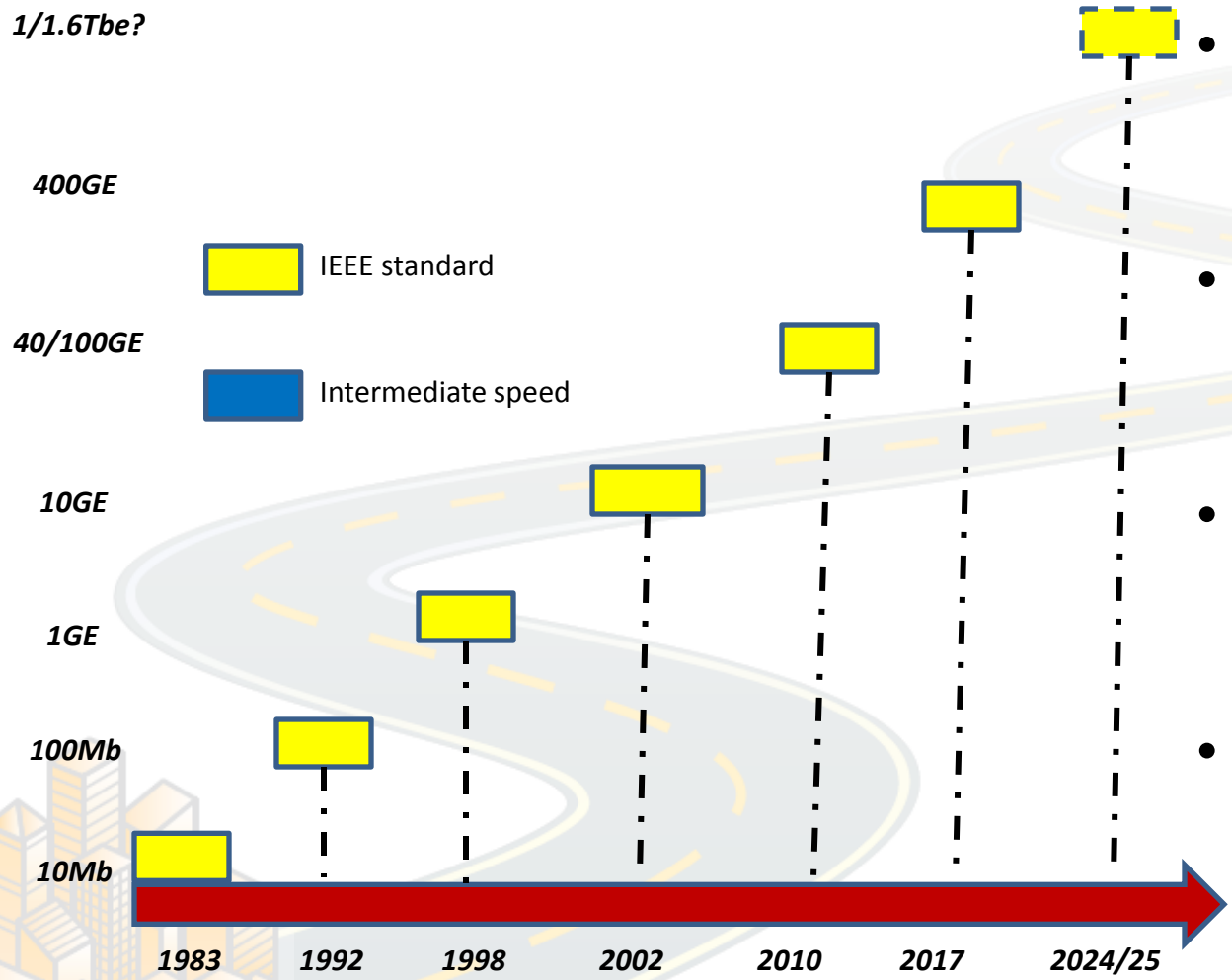
- Cloud computing & Virtualization
  - IAAS, PAAS, SAAS …
  - OPEX based Economy compare to CAPEX based
- Social Media & Mobile
  - Photo sharing, Chat
- Internet Traffic Increases due to
  - IPTV, Cloud computing, Social media …
- IOT, 5G and Big data
  - Infrastructure and DCI

Trends in Data Center:

- Heterogeneous computing : Machine learning
  - CPU, GPU and FPGA
- Big Data
  - Machine to Machine Traffic
  - Storage & compute closer
    - Scale out storage : SPARK and Hadoop
- Secured & Agile infrastructure
  - Application level monitoring
- IoT, 5G and C-RAN
  - Compute, storage and networking

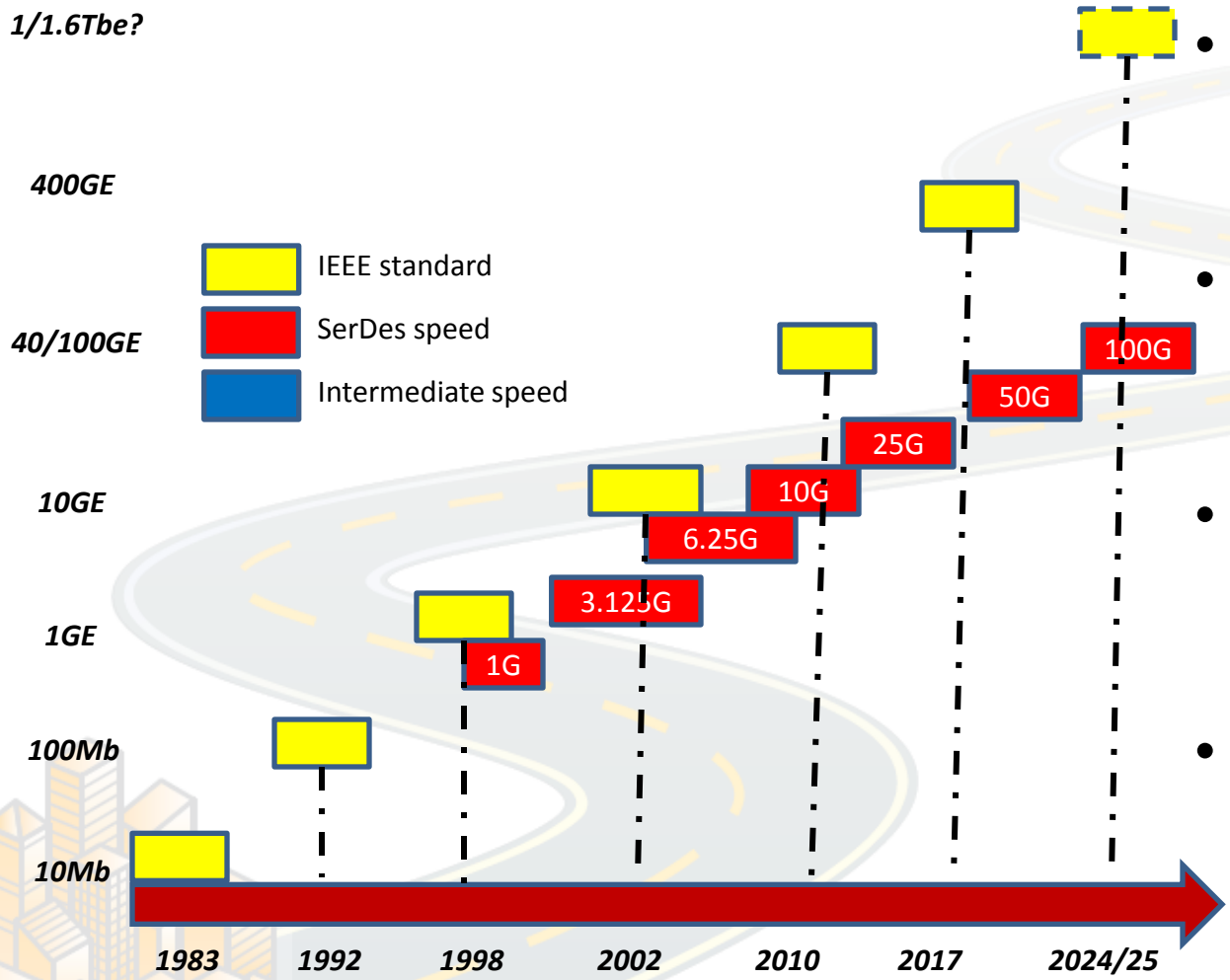
Network Ties All Together

Ethernet Port speed:



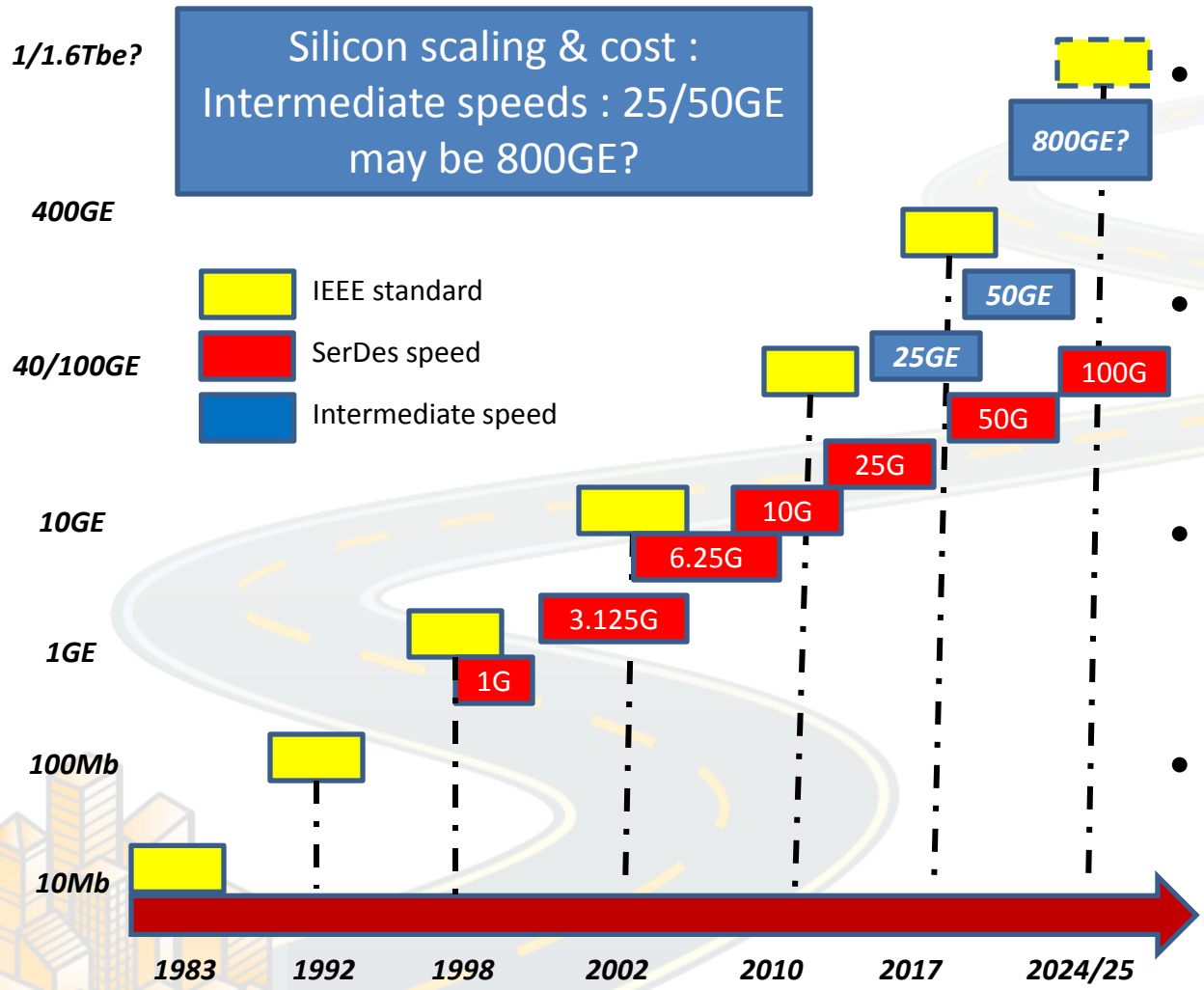
- Ethernet Adapts to Market Needs
- Ethernet speed : 10x/7 years
- Cost 4x : 10X speed
  - Massively deployed
- Cloud vs Enterprise
  - One : Two speed difference

Ethernet Port w SerDes speed:



- Ethernet Adapts to Market Needs
- Ethernet speed : 10x/7 years
- Cost 4x : 10X speed
  - Massively deployed
- Cloud vs Enterprise
  - One : Two speed difference

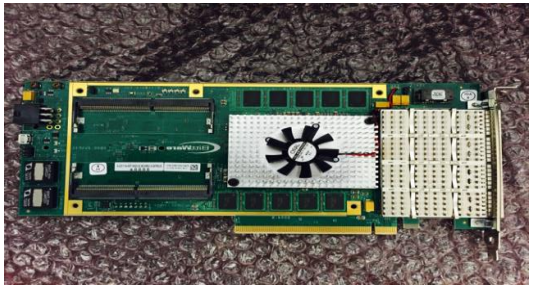
Ethernet Ports : Next speeds



- Ethernet Adapts to Market Needs
- Ethernet speed : 10x/7 years
- Cost 4x : 10X speed
  - Massively deployed
- Cloud vs Enterprise
  - Workload diverges

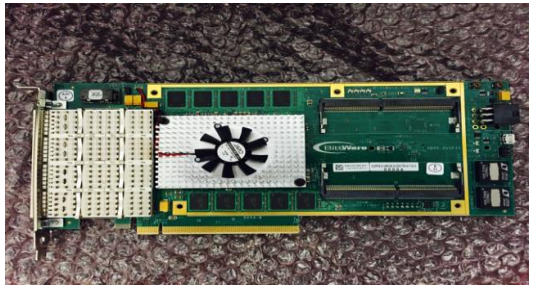
Where is 400GE NIC card today?

400GE Board



3 meters: Direct attached copper cable

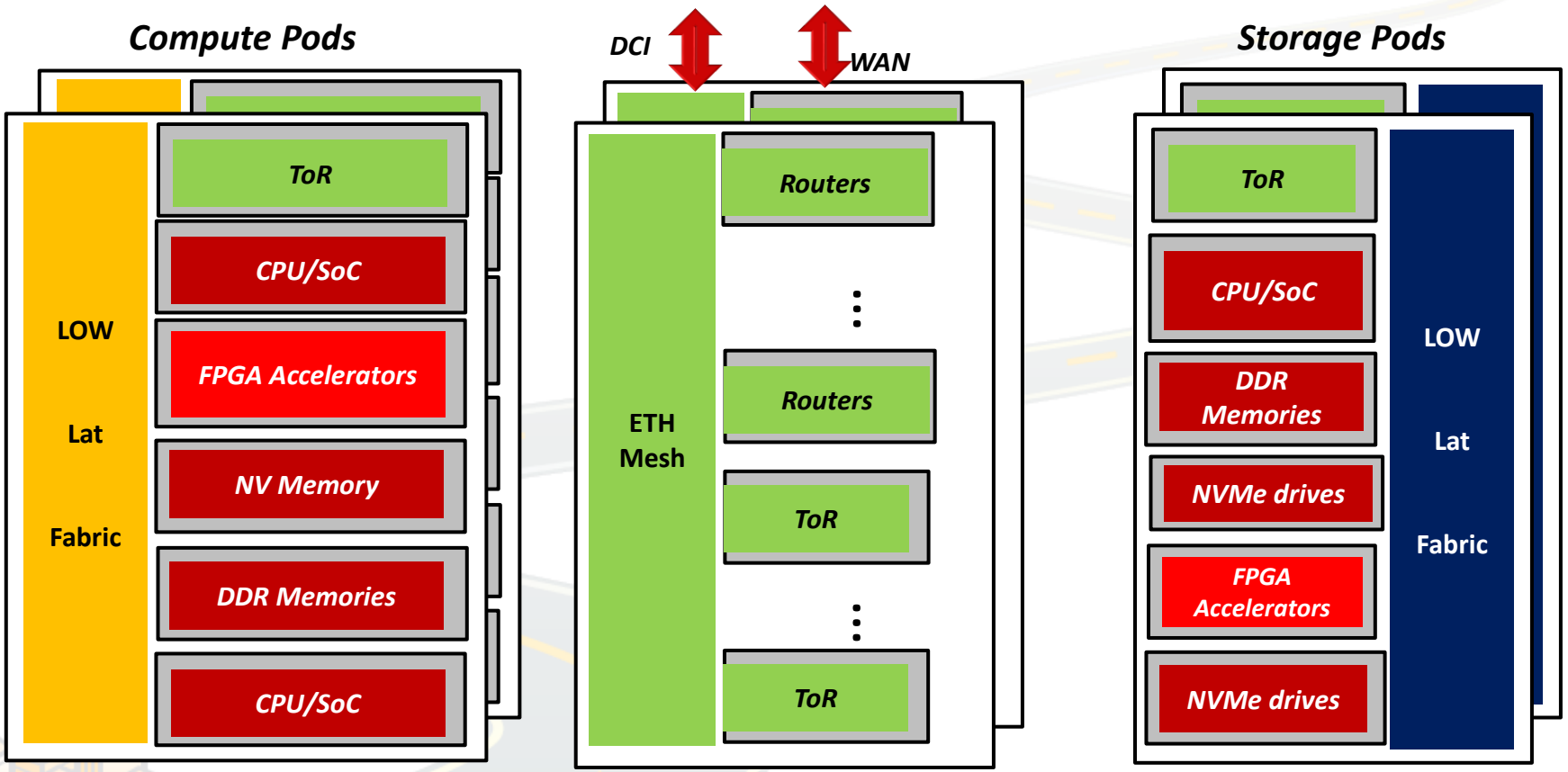
400GE Board



Component	Latency Number	Comment
MAC+PCS	<70 ns	Total w serdes <100 ns
SerDes	<28 ns	Measured in silicon
DMA	<2 us	From software ringing the doorbell to the packet out

Future DC Architecture :

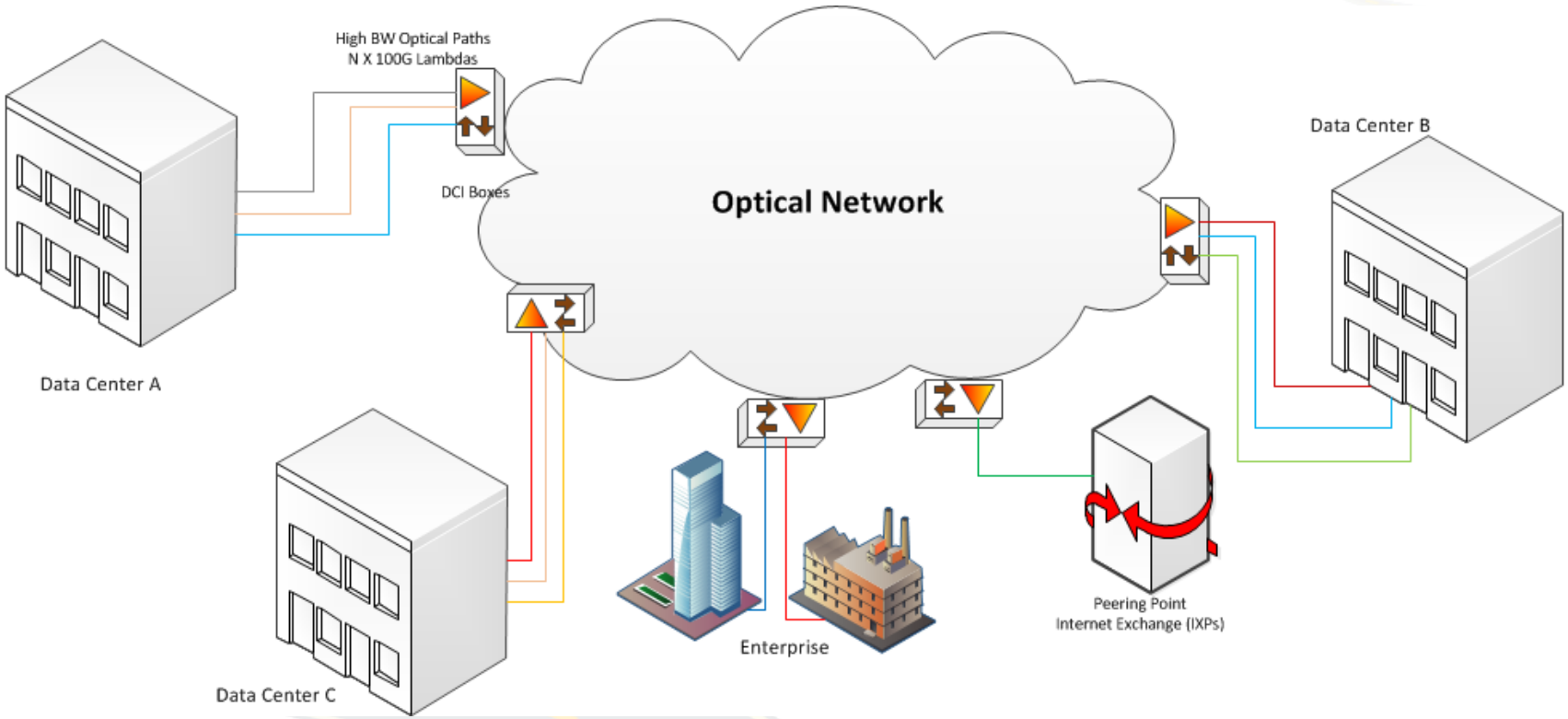
Composable components w different life cycle



DC Wide Management : Compute, Network Storage : Scale independently

Intra Pod : Copper/SiPhotonics/OBO : Inter Pods : Optics (SMF)

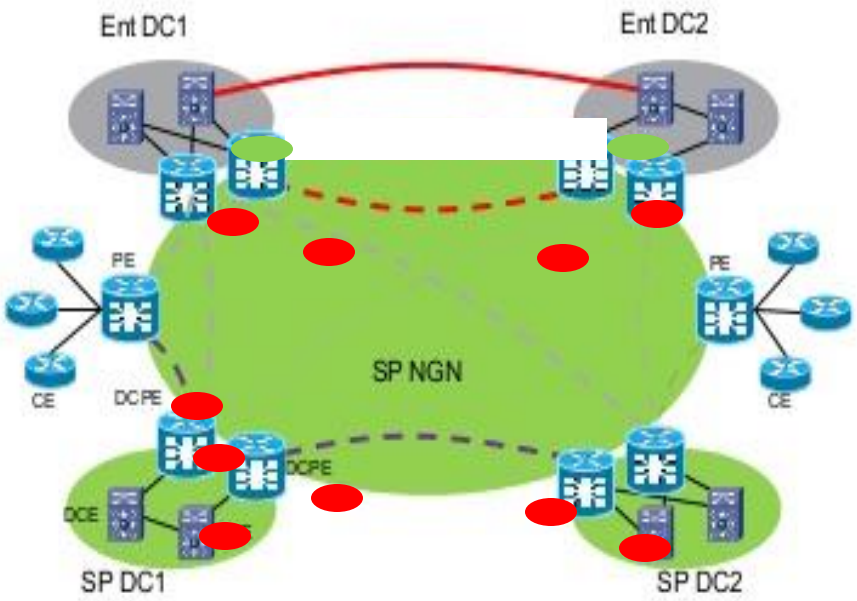
Example : The Data Center Interconnect World



- Data Centers : Communicate w Other Data Centers
- Data Centers : Communicate w Enterprises and Industry
- Data Centers : Communicate w Peering Points

Inter DC Traffic Growth:

DC WAN connectivity in the Cloud-era - *More than DCI*



- Enterprise Data Center inter-connect
- Enterprise Data Center to Provider Data Center
- Provider Data Center to Provider Data Center

"DCI" with varying requirements:

- Multiple 100G needs
- Higher Density Interconnect in metro
- Inter-DC architecture extend beyond metro

Customers shown interest to combine transport and switching gear
Efficient use of wire @ high traffic growth

Source : Cisco

Summary :

- Enterprises and consumers
  - Private, Public and Hybrid clouds
  - TCO reduction and addition of new services
- Telecom and service providers
  - White box : Replace specialized systems
    - NFV, 5G and C-RAN
- Intra-DC and DCI Traffic : Exponential Growth
  - Big data, Cloud computing, IAAS, PAAS, SAAS
- Ethernet : Grows stronger
  - Past: 10Mbps – 400GE
    - Future : 1Tbps+

ETHERNET FOR HIGH PERFORMANCE ENTERPRISE ENVIRONMENTS

Brandon Hoff

SC 2015 BOF Panel

November 18, 2015

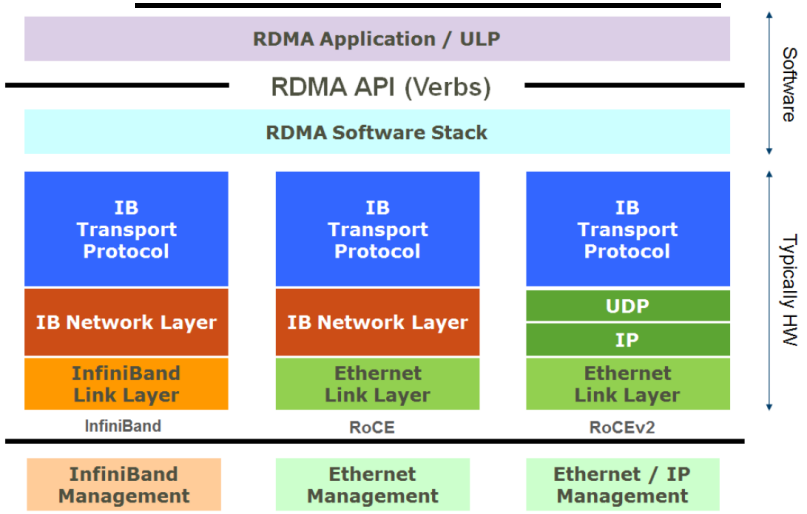
ethernet alliance

Benefits of RoCE

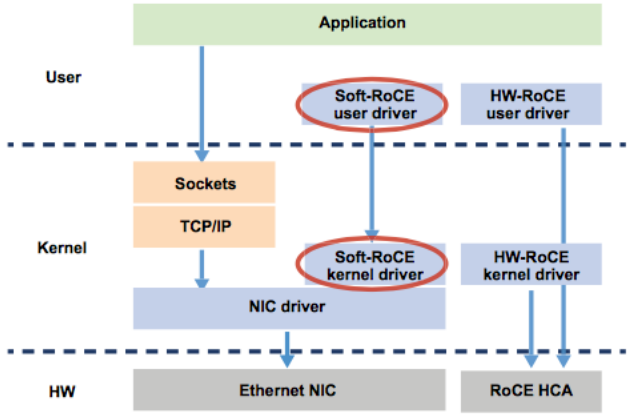
- RDMA over Converged Ethernet (RoCE) is the most commonly used RDMA technology for Ethernet networks
- The RoCE technology is the only RDMA protocol for Ethernet that is stanardized by IBTA, IEEE, IETF
- RoCE is simpler to implement because it uses the IB Transport Protocol

  … and RoCE doesn't add the protocol overhead and implementation difficulties of TCP, DDP, MPA, and RDMAP
- RoCE supports standard Verbs developed and validated in IB environments

IB to RoCEv2 evolution



Soft RoCE for Legacy NICs

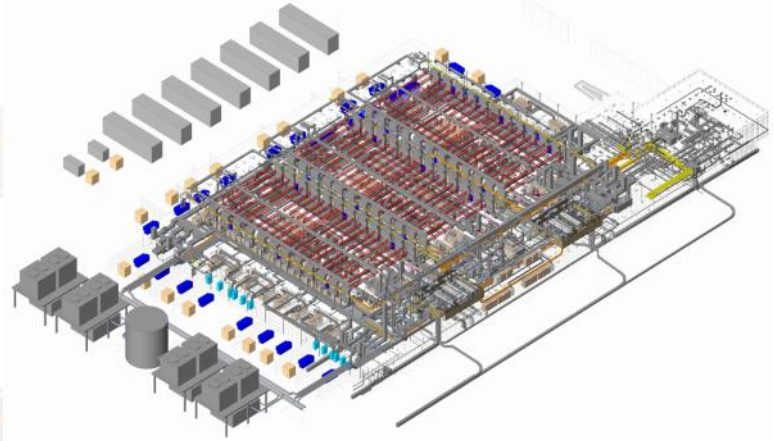


Routable RoCE (RoCEv2)

High performance RDMA that scales across the datacenter

Background	RoCE used Infiniband Network Layer which was not routable (i.e. no Layer 3 routing)
Protocol (new IBTA Standard)	RDMA over UDP – RoCE is now routable. RoCEv2 scales RDMA beyond a single rack
Benefit	Servers on separate subnets can leverage RDMA communications
Software Stack	Uses the same software stack as IB and RoCE
Networking	Generally, Data Center Bridging (DCB) Ethernet or lossless Ethernet is required for high performance on both TCP and UDP networks

Benefits for Data Center Architecture



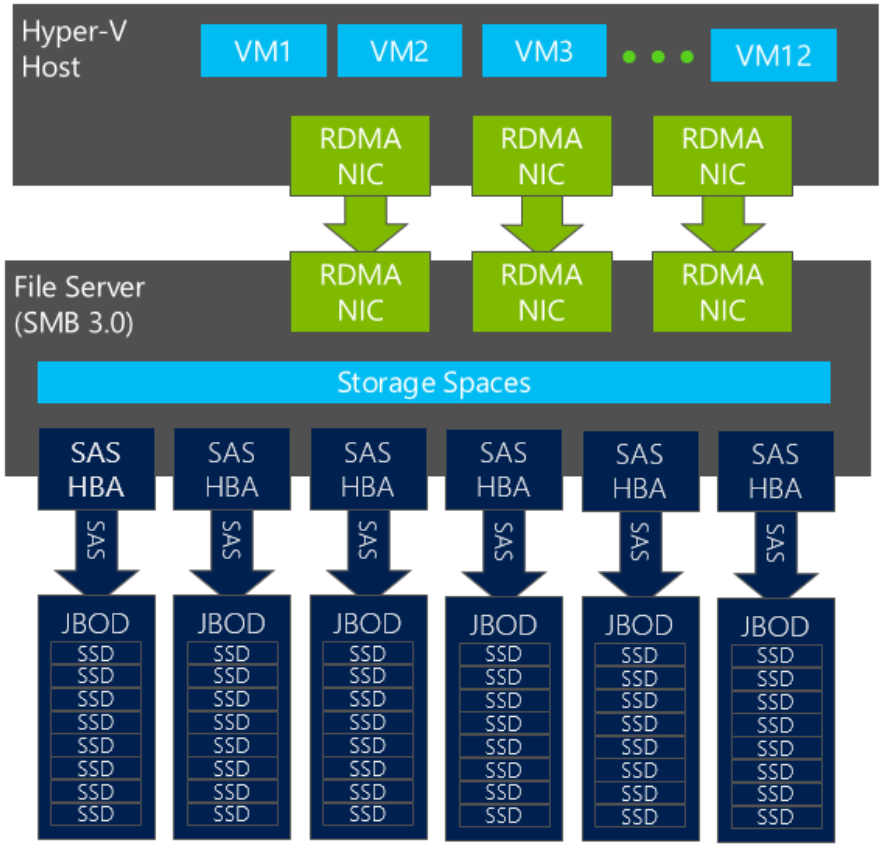
- Enterprise datacenters are now built on a scale-out architecture as pools of servers, networking and storage orchestrated by software
- Datacenter applications from Microsoft, Red Hat VMware are using RDMA as a new higher performance, less CPU intensive transport
- RoCEv2 enables applications to scale between racks and rows of servers
  - With low networking overhead
  - Utilizing traffic prioritization standards (DCB)

SMB Direct Support for RoCE

Hyper-V over SMB Direct

- SMB Direct offers significant performance improvements over standard SMB
  - 60% better bandwidth
  - 10x higher IOPs
- SMB Direct supports
  - Standard file access
  - HyperV remote storage
  - MSSQL clusters
- First production killer application for RoCE
- Demonstraits the value of RDMA beyond traditional HPC markets
- … and the benefits of a converged infrastructure

HyperV with SMB Direct

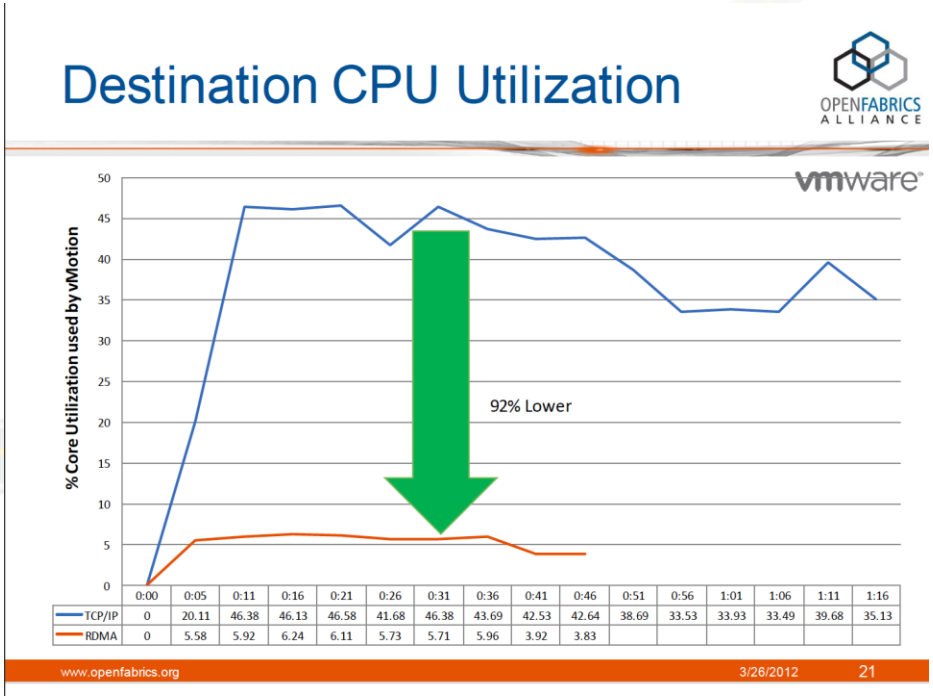


Source: http://blogs.technet.com/b/josebda/archive/2013/04/10/demo-hyper-v-over-smb-at-high-throughput-with-smb-direct-and-smb-multichannel.aspx

New RoCE Cluster Deployments

Clusters for traditional datacenter applications

- Traditional Data Center architectures are looking more and more like clusters
- VMware has been publishing results at the OFA conference
- Significant performance improvements for hypervisor features
  - Storage, VMotion, etc.
- Overall: New workloads looking to leverage RoCE



VMware PoC

92% less CPU cycles

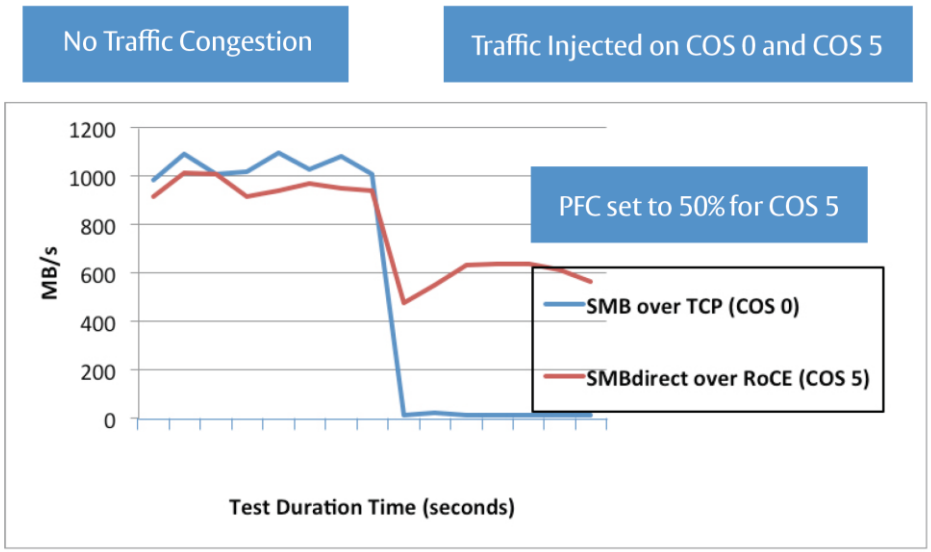
Half the time for a VM migration

PFC is Required for RDMA (RoCE and iWARP)

TCP is not sufficient for RDMA

- TCP congestion management is not sufficient for RDMA traffic
- Data from testing in Cisco's labs shows that TCP isn't for RDMA workloads
- PFC is required by iWARP and RoCE for congested networks

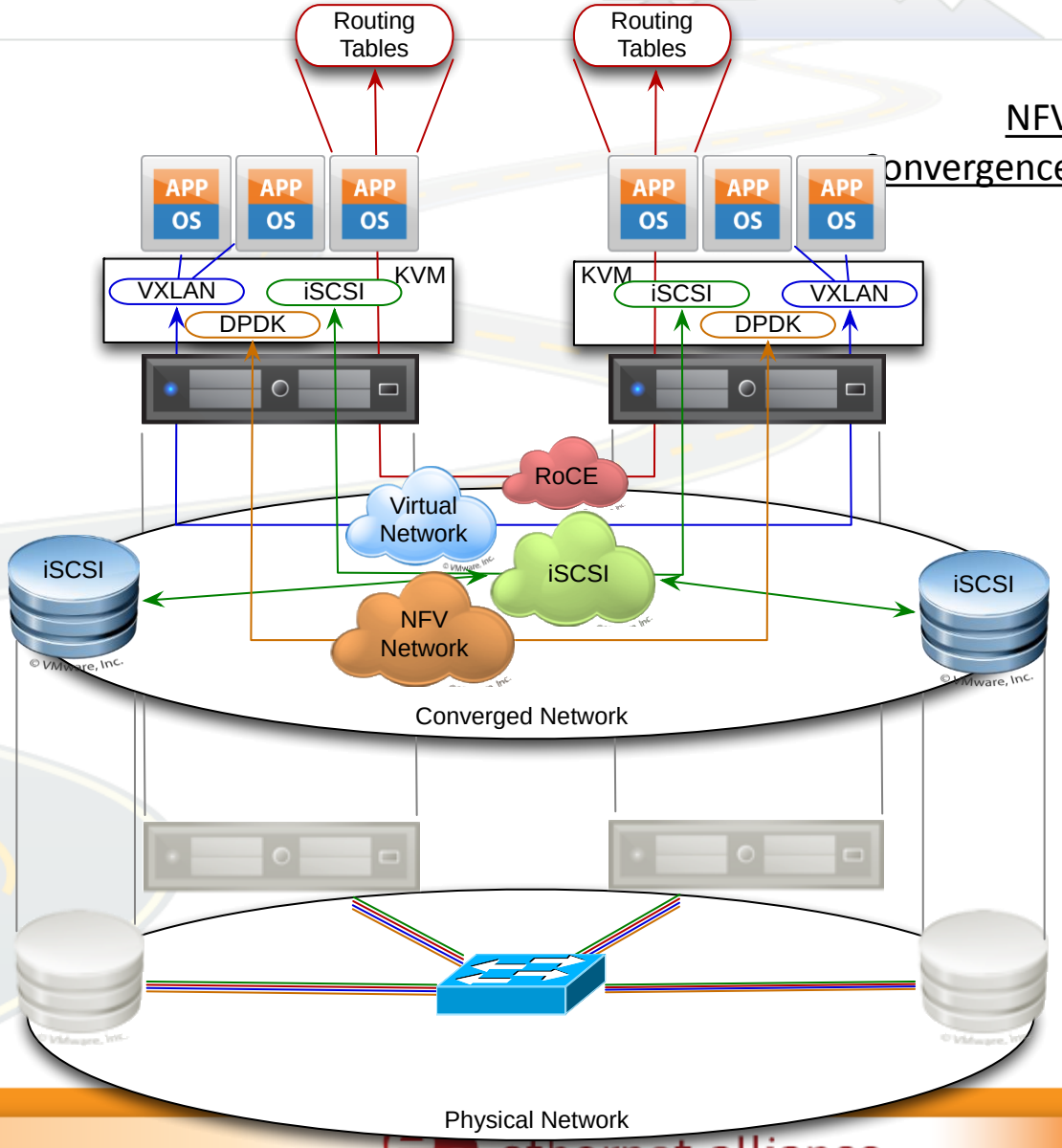
Performance with Network Congestion



The chart above shows TCP SMB performance drop significantly in the presence of congestion.
PFC is required for TCP traffic as well.

Converged NFV with RoCE

- Ethernet and IP are the leading connectivity solutions for NFV
- Some Telco applications are moving from proprietary hardware to RoCE enabled COTS hardware
- Delivers a converged infrastructure for cloud and carrier solutions



QUESTIONS AND SOME ANSWERS

Thank You!

If you have any questions or comments,
please email admin@ethernetalliance.org

Ethernet Alliance: visit www.ethernetalliance.org

Join the Ethernet Alliance LinkedIn group

Follow @EthernetAllianc on Twitter

Visit the Ethernet Alliance
on Facebook

ethernet alliance