Global
Foundation
Services

Microsoft

Global Networking Services

Objectives to Support Cloud Scale Data Center Design

Brad Booth, Tom Issenhuth

IEEE 802.3 400Gb/s Ethernet Study Group
IEEE 802 November 2013 Plenary
Dallas, TX

Supporters

- Andreas Bechtolsheim, Arista
- Anshul Sadana, Arista
- Chris Bergey, Luxtera
- Brian Welch, Luxtera
- Radha Nagarajan, Inphi
- Sudeep Bhoja, Inphi
- Jeff Maki, Juniper
- Dave Chalupsky, Intel
- Ryan Latchman, Mindspeed
- Dan Dove, Dove Networking Solutions
- David Warren, HP
- Brian Teipen, ADVA Optical Networking
- Rick Rabinovich, Alcatel-Lucent
- Adit Narasimh, Molex
- Scott Sommers, Molex
- Joe Dambach, Molex
- Kapil Shrikhande, Dell
- Scott Schube, NeoPhotonics
- Winston Way, NeoPhotonics
- John Petrilla, Avago Technologies
- Nathan Tracy, TE Connectivity
- Piers Dawe, Mellanox Technologies
- Arlon Martin, Mellanox Technologies
- Kent Lusted, Intel
- Gary Nicholl, Cisco
- Arash Farhood, Cortina Systems

Cloud Scale Data Centers

- There is no single design or size for a cloud data center
  - Topologies continue to evolve with technology advancements and cost structures
  - Differences are driven by generation of design, location and scale
- While the overall traffic flow within different data centers is similar the design differences drive different link requirements
- Data center development/growth
  - Three phases: design, build-out and operational (often simultaneously)
  - Three year refresh cycle
    - New colo* may come online as old one is being refreshed
    - Infrastructure should last at least 4-6 generations of refresh
  - New data centers and colos being added to meet growing demand

* Colo = colocation (8 MW deployment area)

Optical Ethernet Historical – Data Center*

Inside Data Center

1 Gigabit Ethernet – 500m MMF (SX) STD 550m

10 Gigabit Ethernet – 300m MMF (SR) OM4

40 Gigabit Ethernet – 150m MMF (SR4)

100 Gigabit Ethernet – 150m MMF (SR10)

100 Gigabit Ethernet – 100m MMF (SR4)

400 Gigabit Ethernet – ??m SMF

Outside Data Center

1 Gigabit Ethernet – 3km SMF

10 Gigabit Ethernet – 2km & 10km SMF

40 Gigabit Ethernet – 10km SMF

100 Gigabit Ethernet – 10km SMF

400 Gigabit Ethernet – ??km SMF

* Objectives

Inside the Data Center is ≤ 500m

Optical Ethernet Historical – Campus*

Inside Campus	Outside Campus
1 Gigabit Ethernet – 3km SMF (LX) — STD (5km)	
10 Gigabit Ethernet – 2 & 10km SMF (LR)	10 Gigabit Ethernet – 40km SMF (ER)
40 Gigabit Ethernet – 10km SMF (LR4)	40 Gigabit Ethernet – 40km SMF (ER4)
100 Gigabit Ethernet – 10km SMF (LR4)	100 Gigabit Ethernet – 40km SMF (ER4)
400 Gigabit Ethernet – ??km SMF	400 Gigabit Ethernet – ??km SMF

* Objectives

Inside the Campus is ≤ 2km

Why Outside Not Used Inside

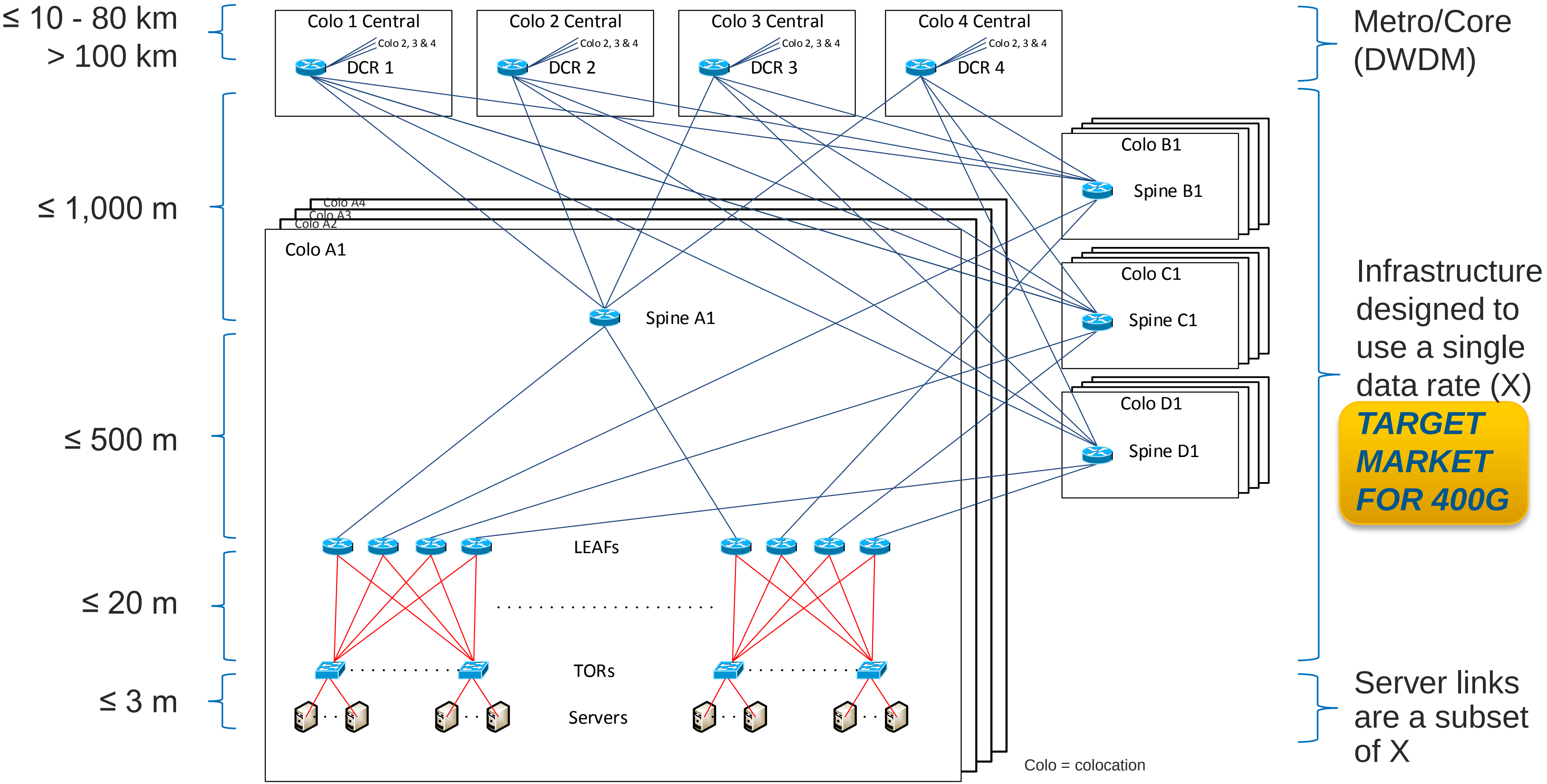
A lot of bucks…

Examples of Missing the Market

- 40GBASE-SR4 is a good solution for a row, but the 150m reach doesn't cross the data center
  - Industry created a 300m version based off of 10GBASE-SR
- 100GBASE-SR10 is not cost effective (# of fibers required)
- 100GBASE-SR4 reach doesn't cross the data center
  - .3bm unable to adopt SMF solution
- 100GBASE-LR4 "Lite" solution developed for ≤ 2km
- New 1300nm optimized MMF for data center applications

Study Group Has Opportunity to Develop Standard to Meet Market Need

Cloud Data Center Campus Interconnections



Interconnection Volume

- Four sections per colo & multiple colos (≥ 4) per data center
- Volumes below are per section (except DCR to Metro)

A End	Z End	Volume	Reach (max)	Medium	Cost Sensitivity	Market Space
Server ǂ	TOR	10k – 100k	3 m	Copper	Extreme	LAN
TOR	LEAF	1k – 10k	20 m	Fiber (AOC)	High	
LEAF	SPINE	1k – 10k	400 m	SMF	High	
SPINE	DCR	100 – 1000	1,000 m	SMF	Medium	Campus
DCR	Metro	100 – 300	10 - 80 km	SMF	Low	WAN

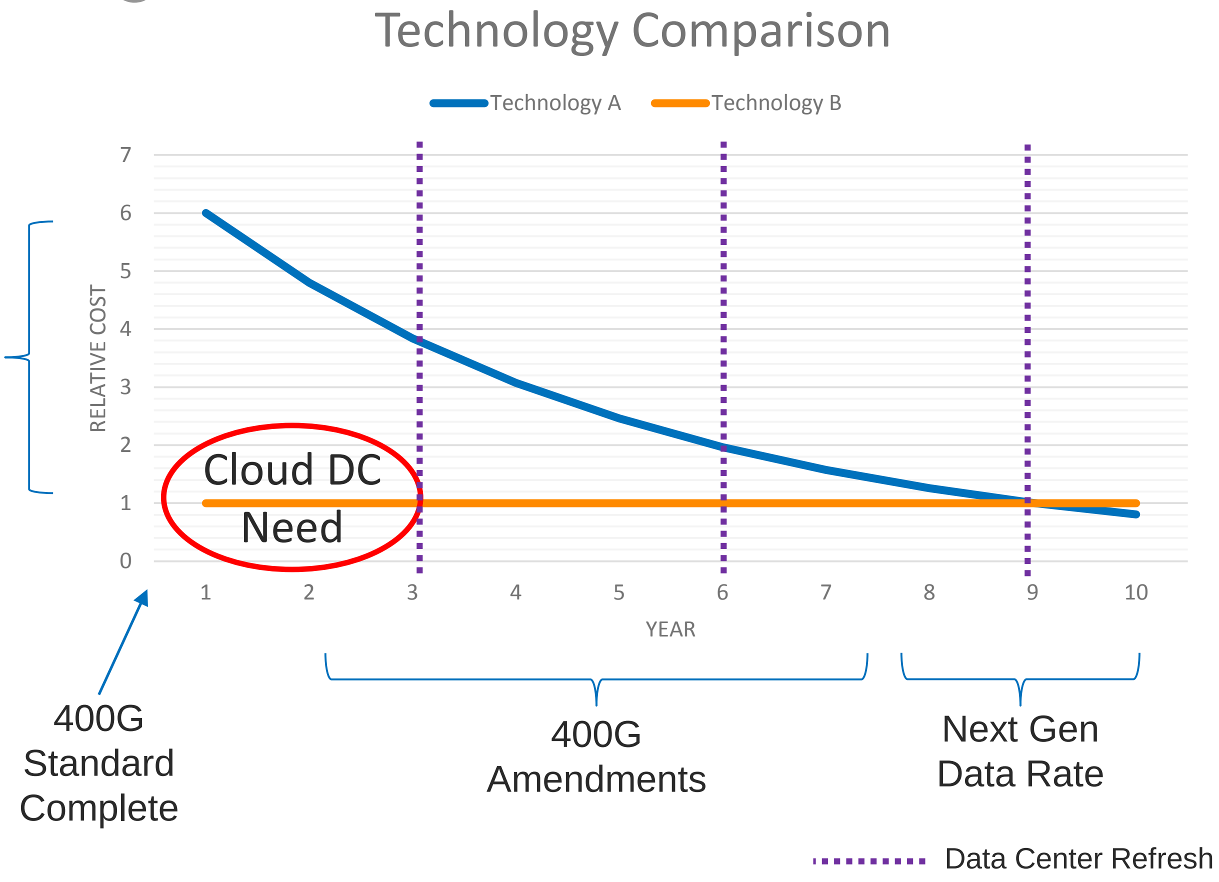
ǂ Server-TOR links may be served by breakout cables

Technology Adoption

- 40G
  - Growing for server connectivity
  - Strong in TOR to DC router
- 100G
  - Starting in TOR to DC router (non-standard)
  - Missing components: low-cost 300-400m optics, switch silicon
- 400G
  - Targeting TOR to DC router
  - 40G servers will increase the need to reduce over-subscription
  - Need to supply components that slowed 100G adoption

Technology Timing Considerations

Early high-volume (cloud) adopters not willing to pay premium for Technology A



Cloud Data Center Reach Considerations

- LAN links ≤ 500 m
  - Very cost sensitive due to high volume of links being used
  - Typically assume a 3-4 dB loss budget
- Campus links ≤ 2 km
  - Decreased cost sensitivity due to lower volume and technical trade-offs
  - Loss budget typically in 4-6 dB range
- Metro/Core is typically DWDM (outside of IEEE 802.3 scope)
- Links ≤ 20 m
  - MMF module is a possibility, but needs to be cost competitive with AOCs
  - Copper still being used intra-rack – breakout is of interest

Recommendation

- Adopt objectives to support the high-volume Cloud Data Center reach requirementǂ
  - Provide physical layer specifications which support 400 Gb/s operation over:
    - At least 500 m of single-mode fiber
    - At least 2 km of single-mode fiber
- Electrical interface specification
  - Objective should enable AOC implementations
  - Sufficient for direct-attach copper (DAC) implementations?

ǂ Task Force may decide a single PMD can satisfy both objectives.

Thank you!