Support for an objective of 25 Gb/s over MMF

Draft 0.3d

September 2014, Ottawa, Canada

(team of many)

Contributors

- Alan Flatman, LAN Technologies
- Jonathan King, Finisar
- Scott Kipp, Brocade
- Paul Kolesar, Commscope
- Dale Murray, LightCounting
- John Petrilla, Avago Technologies
- George Zimmerman, CME Consulting/Commscope

Supporters

- John Abbot, Corning
- Chris Cole, Finisar
- Jack Jewell, independent
- Jonathan King, Finisar
- Scott Kipp, Brocade
- Paul Kolesar, Commscope
- Greg LeCheminant, Keysight Technologies
- Robert Lingle, OFS
- Dale Murray, LightCounting
- Rick Pimpinella, Panduit
- Steve Swanson, Corning
- George Zimmerman, CME Consulting/Commscope

Contents

- Why include 25 Gb/s over MMF in this project ?
  - Ethernet diversity
    - Top-of-Rack (ToR), cabinet-to-cabinet, Middle-of-Row (MoR), and End-of-Row (EoR), server to switch architectures
    - Link length distributions, ToR and EoR
    - Market data at 10G, AOCs and Transceivers
- So what does 25 Gb/s over MMF do for us?
- Incremental developments needed to standardize 25 Gb/s over MMF
- Proposed 25 Gb/s over MMF objective
- How an optical objective will augment the 5 Criteria responses
- Summary and Conclusion

…. & back up slides

Why include 25 Gb/s over MMF in this project ?

- History shows the success of co-developed electrical and optical specifications for 10 Gb/s lanes
  - SFF8431, 802.3ba.
- 10 Gb/s market data shows optics are connecting millions of servers today, supporting a broad range of switch-server topologies.
- Similar expectations for 25 Gb/s
  - Not just about a favorite leading application, but also about what will be adopted by the wider market over the next 1 to 3 years.
- Co-development of a PHY for 25 Gb/s over MMF within the 25GE project will significantly broaden the diversity of applications served by, and the market potential and economic feasibility of, 25 Gb/s Ethernet.

Diversity of DC architectures

"There is no single end-all cabling configuration for every data center"[a]

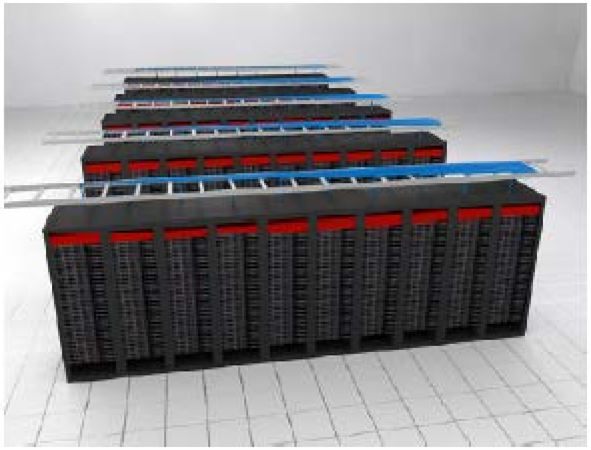
- Original CFI based on backplane and Top-of-Rack (ToR) server-to-switch architectures - ToR is not sufficient for all applications[b]
  - A single 1U switch can connect to multiple racks of servers[c]
  - Modular switches can connect to 1000's of servers over many tens of racks[c]
    - A 3 to 5 m PHY addresses just intra-rack cabling
- A longer reach PHY enables cabinet to cabinet, Middle of Row (MoR) and End of Row (EoR) data center architectures

- Each topology (ToR, MoR, EoR) has pros & cons, and exists in the market today for very sound reasons:
  - Cost, switch port use efficiency, thermal management, maintenance, scalability, support of mixed applications, reconfigurability, etc.
  - Abundant supporting material (appendix) ! [a,b,c,d,e,f]
- A 25 Gb/s MMF PHY will have the reach to support a much broader range of DC architectures.
  - *See Scott Kipp's presentation this meeting !*

ToR, MoR, EoR

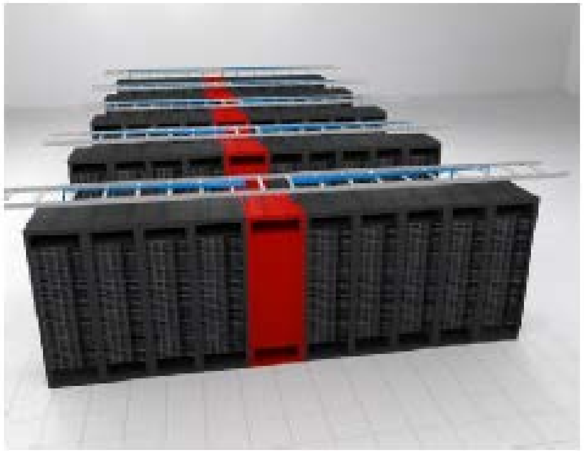
ToR

(up to 5 m)



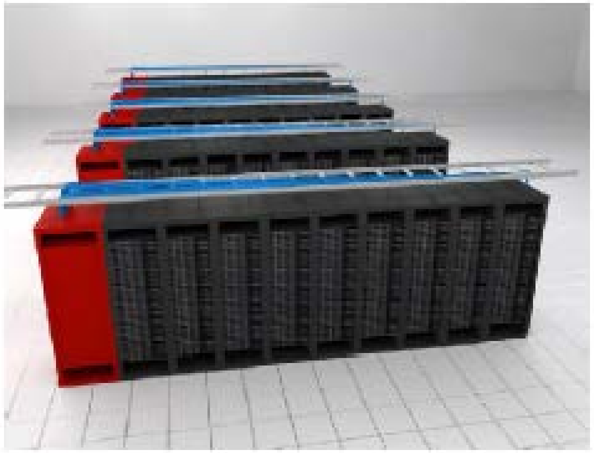
MoR

(up to 15 m)



EoR

(up to 50 m)



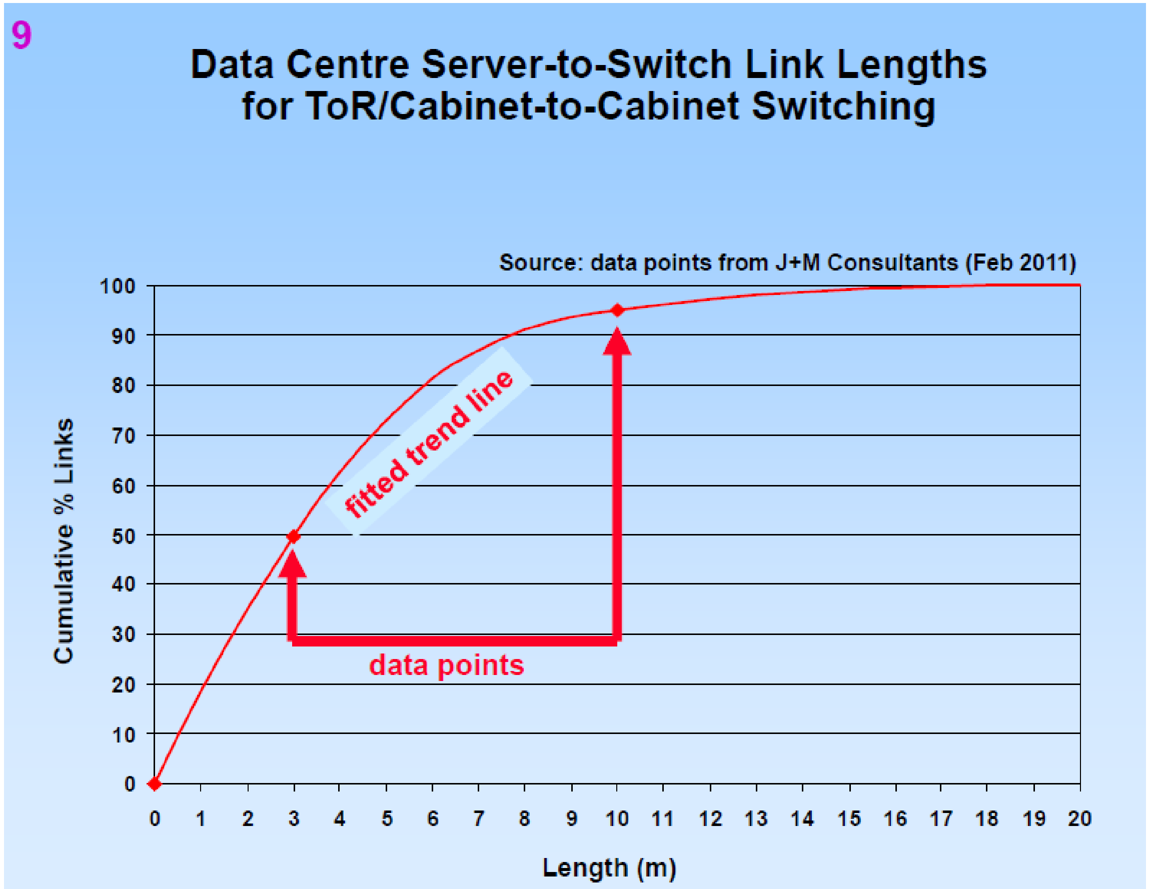
Intra-rack addressed by 25Gb/s copper direct attach

Not addressed by 25Gb/s copper direct attach

Pictures from jiminez_3bq_01_0711.pdf, 802.3bq

ToR link distributions

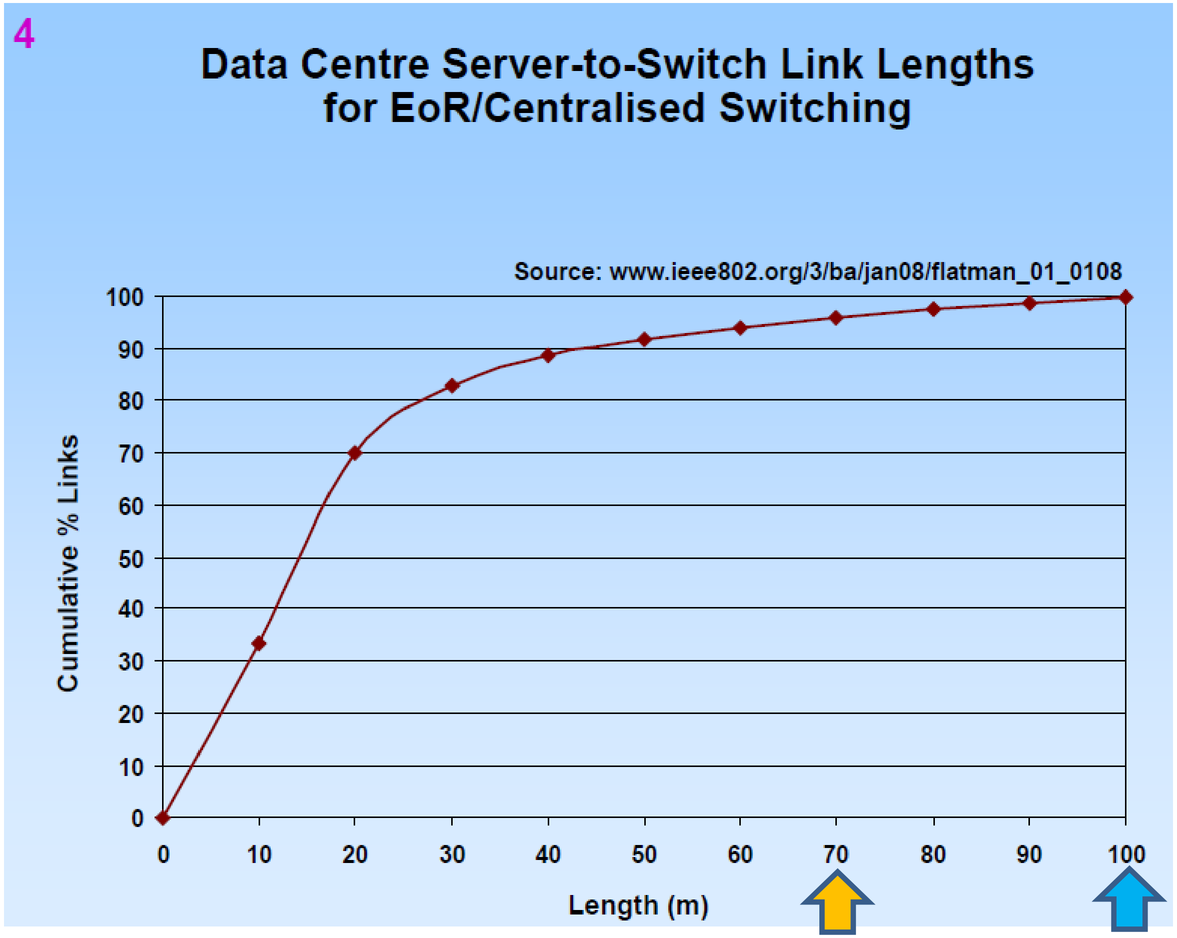
- Includes cabinet-to-cabinet links
  - Note: slide 8 of *CFI_01_0714*, “The term “TOR” has become synonymous with server access switch, even if it is not located “top of rack”, acknowledging that a 3 m reach may not be sufficient for all ‘TOR’ server to switch links.
- Link lengths: ~50% > 3 m
  - From: *flatman_01_0911_NG100GOPTX.pdf*, reproduced with kind permission of Alan Flatman



http://www.ieee802.org/3/100GNGOPTX/public/sept11/flatman_01_0911_NG100GOPTX.pdf

EoR link distributions

- Link lengths: ~90% > 3 m, ~ 90% < 50 m
  - From: *flatman_01_0911_NG100GOPTX.pdf*, reproduced with kind permission of Alan Flatman



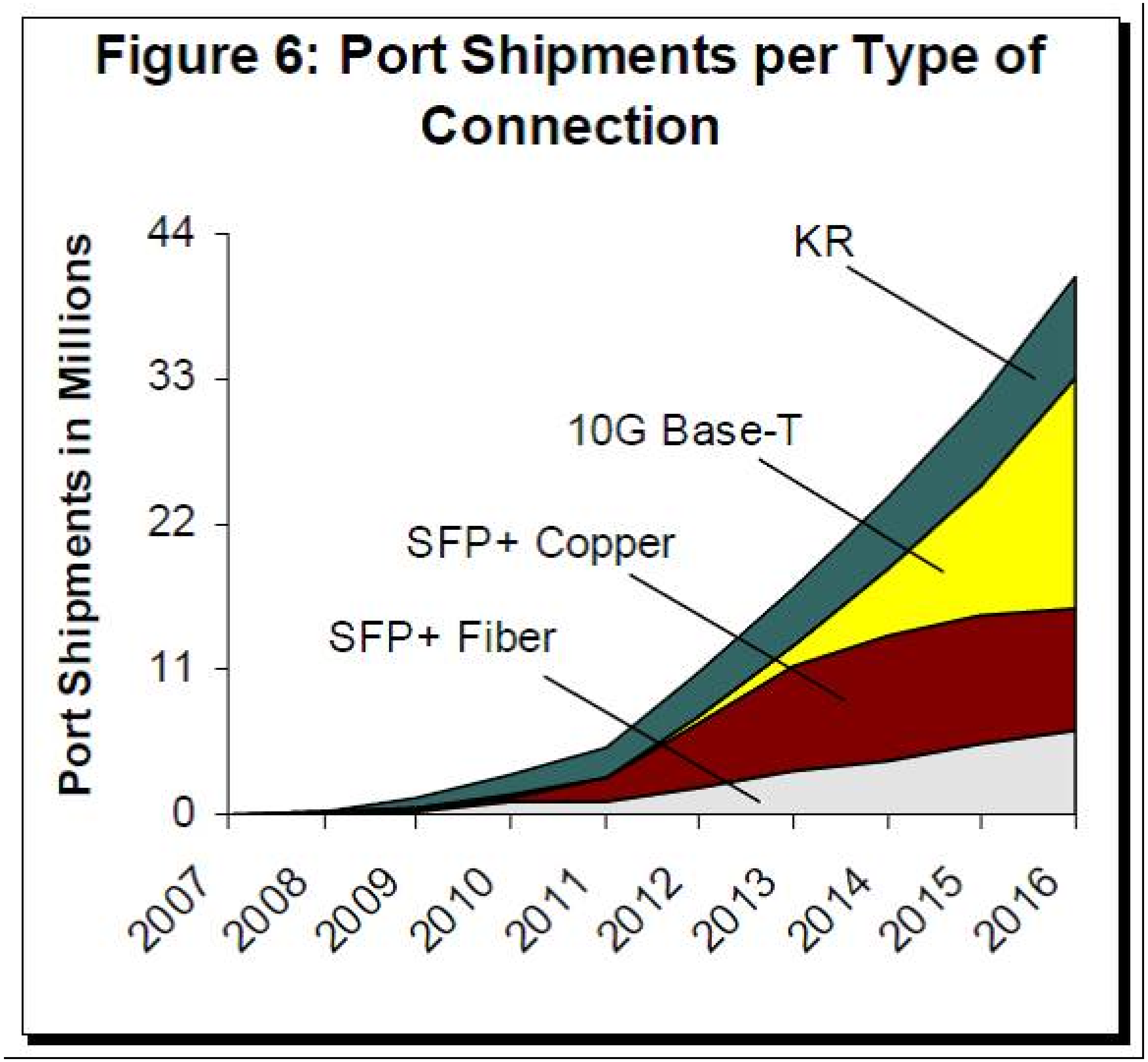
32GFC &100GBASE-SR4
OM4 🟦 OM3 🟨

http://www.ieee802.org/3/100GNGOPTX/public/sept11/flatman_01_0911_NG100GOPTX.pdf

10G port shipments by type, market data & projection

- Millions of 10GbE servers today connecting to switches with optics
- 5 million ports/year by 2016

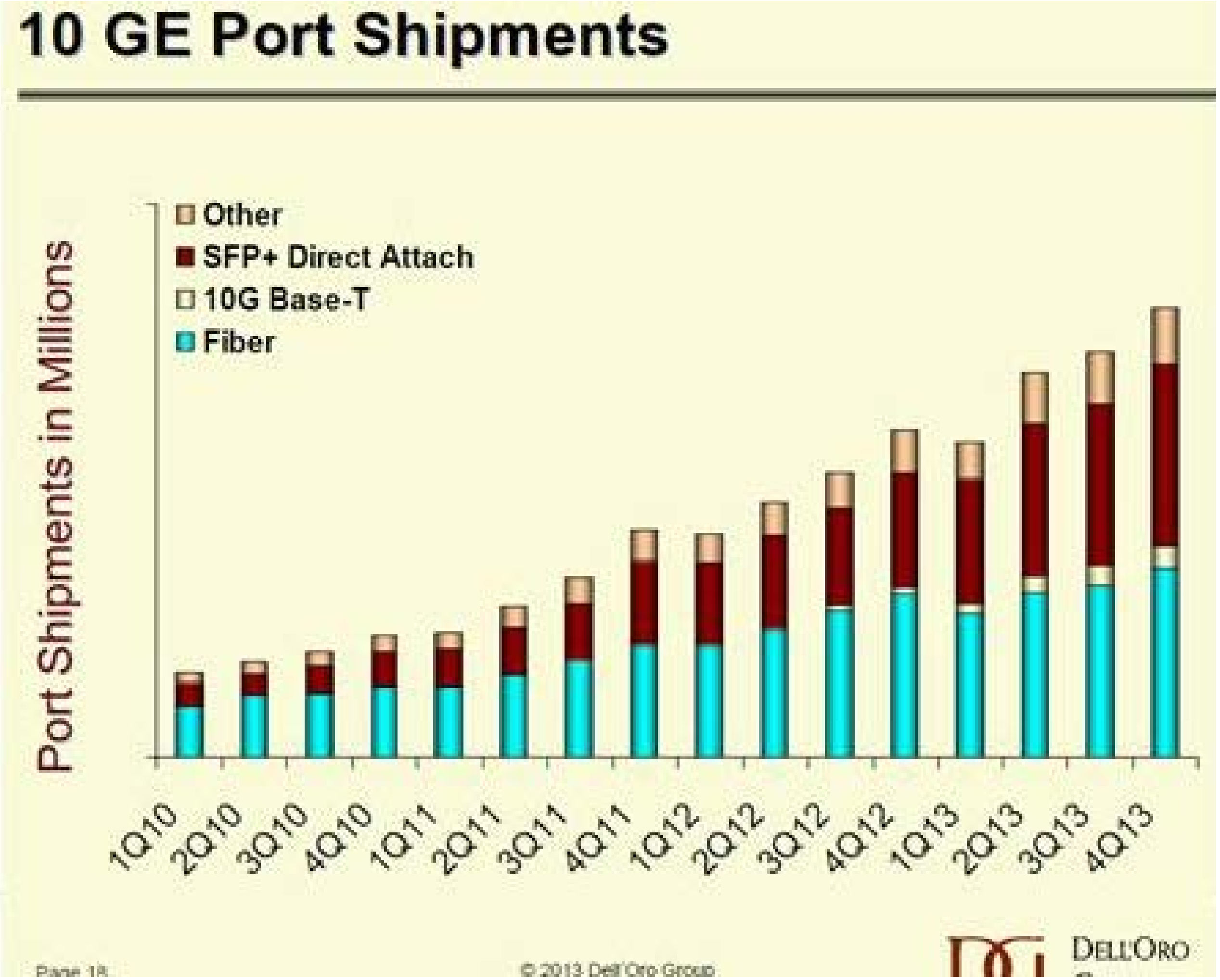
- SFP+ fiber and SFP+ direct attach copper were both fast to market and equally important in terms of volume.



"Dell'Oro Controller and Adapter Forecast Summary – Jan 2012"

10G port shipment market data

- Dell’Oro, April 2014, Ethernet Summit

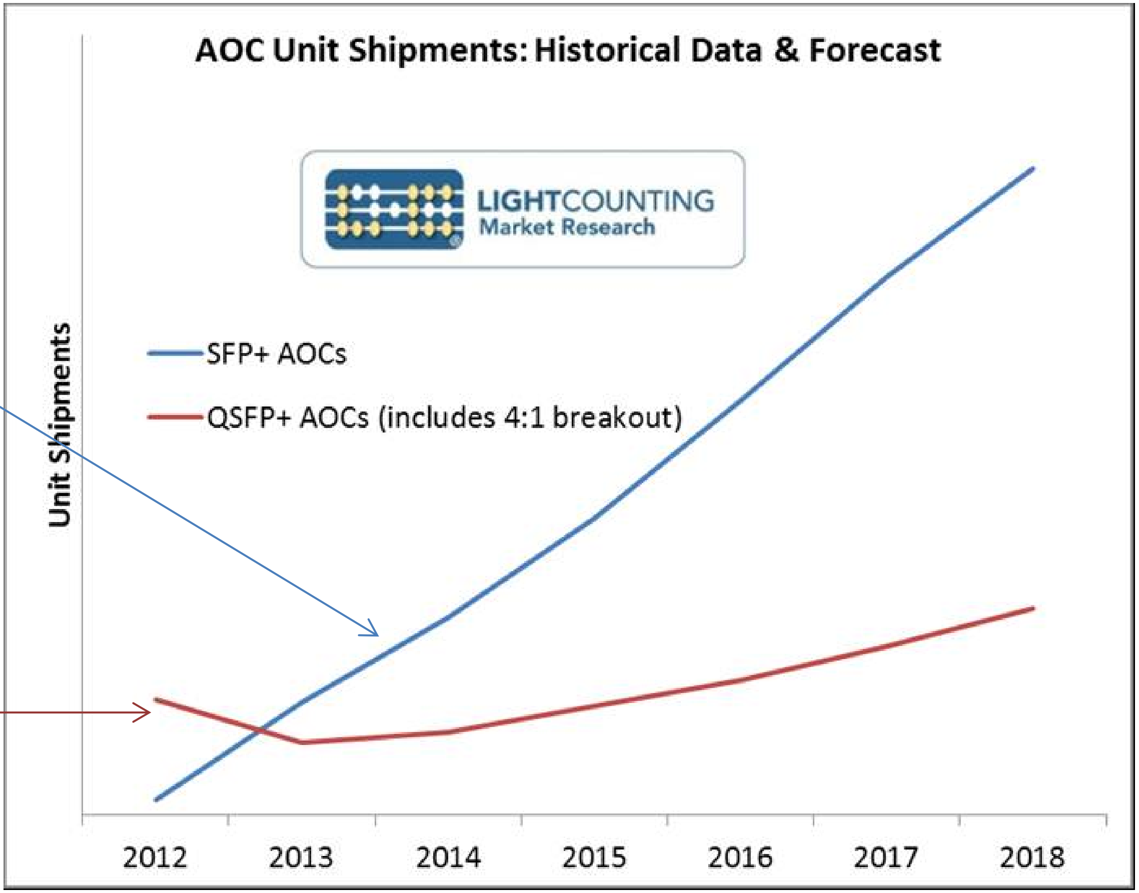


http://www.ethernetsummit.com/English/Collaterals/Proceedings/2014/20140430_Keynote2_Weckel.pdf

Market data: 10G per lane AOCs

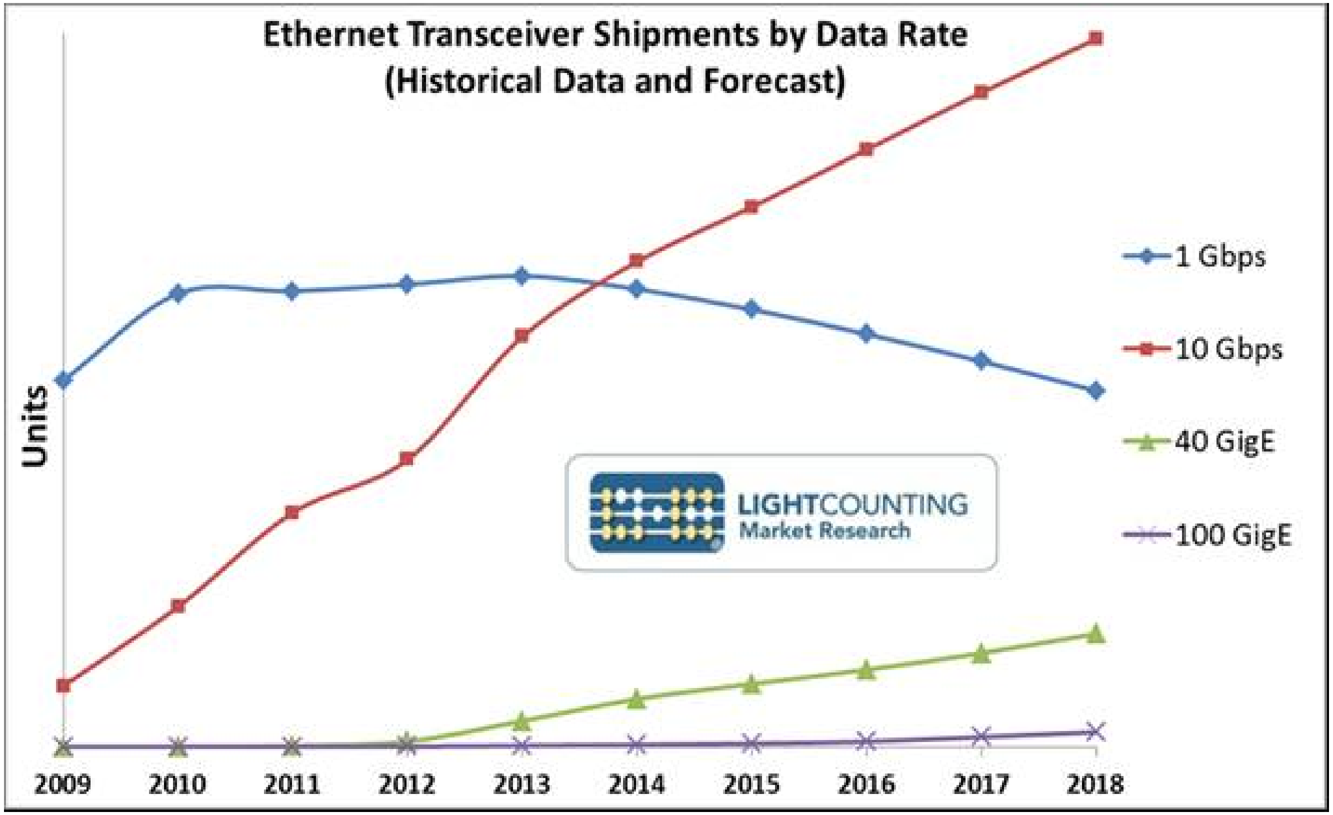
- SFP+ AOCs mostly for server connections as direct replacements for DAC cables
  - interoperability
  - optical assemblies easier to manage

- Majority of the QSFP+ AOCs are for HPC clusters using InfiniBand, some 4:1 to connect servers to ToR switches, split unknown.



"In spite of a significant disadvantage in cost vs. DACs, AOCs are finding use in server connections and this is growing quickly at 10G. At 25G, AOCs and optics in general will enjoy additional advantages in distance, bulk, weight, cable management and plug-and-play vs. DAC assemblies", Dale Murray, LightCounting

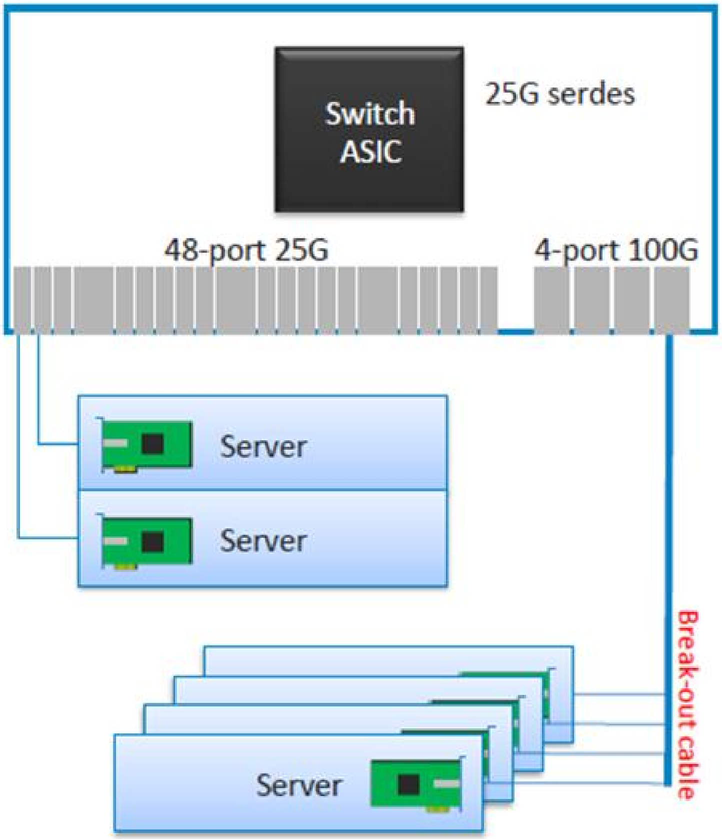
Market data: 10G per lane Transceivers



- “Very strong growth of 40G QSFP+ modules, especially short reach… much of it has not been for 40GbE but rather a density play for 10GbE between servers and ToR switches”, Dale Murray, LightCounting

25 Gb/s over MMF Ethernet Connectivity

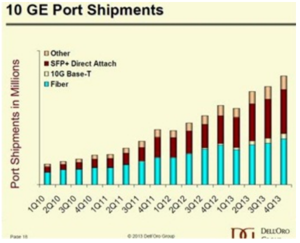
- Enables similar topology as 40 Gb/s and 10 Gb/s over longer reach
  - Single 25 Gb/s SFP28 port implementation or Quad 25 Gb/s
  - QSFP28 breakout implementation possible
- Maximizes ports and bandwidth in switch faceplate for MoR and EoR architectures
- Dense rack server

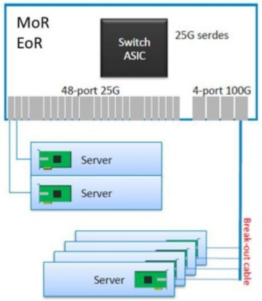


Picture from *CFI_01_0714.pdf*

So what does 25 Gb/s over MMF do for us?

- A PHY for 25 Gb/s over MMF
  - Enables optimization of switch port usage over a broad range of server to switch architectures (cabinet to cabinet, MoR, EoR), which make up a substantial fraction of total server interconnects.
  - Supports structured-cable installations
  - Enables optimized port utilization of very high port count modular switches
  - Allows utilization of otherwise stranded ports
  - Enables existing form factors (SFP and QSFP) to support multiple lanes of similar electrical and optical interfaces to provide high port density options





… just like 10 Gb/s over MMF....

Incremental work needed to define a PHY for 25 Gb/s over MMF

- Single lane FEC
  - FEC option required for backplane, re-use for optics (as in 100GBASE-SR4/KR4)
- Chip-to-module interface
  - Needed for AOCs and for pluggable optics.
  - Technology re-use of 25 Gb/s lane standards e.g. clause 83E chip-to-module specs (slide 18 of *CFI_01_0714*)
- Electrical connector
  - Re-use copper twin-ax cables port interfaces & form factors: SFP28, QSFP28, CFP4
- Optical interface specs
  - Re-use 32GFC and 100GBASE-SR4, both of which include applicable ~25 Gb/s optical lane specifications.
  - No new component developments.
  - <1 Watt SFP+ form factor has been demonstrated (32GFC samples)
- Optical MDI
  - Same MDI as SFP+ and QSFP optical modules: LC and MPO connectors

No technical risk + extensive industry experience + full suite of existing standards near completion to draw from = rapid standardization

Proposed objective

- Define a single-lane 25 Gb/s PHY for operation over MMF consistent with IEEE P802.3bm Clause 95

How the 5 Criteria responses are augmented by a 25 Gb/s over MMF objective

- No change in proposed wording needed.
  - Broad Market Potential
    - Lower cost, size and power for server interconnects in data centers, internet exchanges, co-location services, services provider and operator networks.
    - Enables optimized switch port usage over broad range of server to switch architectures (Cabinet-to-Cabinet, Middle-of-Row, End-of-Row).
    - Allows otherwise stranded ports to be utilized.
    - Enables large modular switches with high port counts.
  - Economic Feasibility
    - 25GBASE-SR (single lane) will be lower cost than 40GBASE-SR4 (four lanes)
    - 25GBASE-SR increases bit-rate/fiber (vs 10GBASE-SR and 40GBASE-SR4)
    - 25GBASE-SR has the reach to enable higher port use efficiency for large modular switches
  - Technical feasibility - 32G Fibre Channel and 802.3bm standards
  - Distinct Identity - No other PHYs for 25 Gb/s over MMF
  - Compatibility - No change.

Summary

- The server interconnect world is very diverse.
  - ToR is right for some users, but for many applications and user facilities, EoR or MoR architectures enable better cost effectiveness, floor planning and support of mixed applications.
- Based on market data at 10G, a 25G MMF PHY is expected to be a significant solution serving 25G Ethernet data centers.
    - Enables optimized switch port usage over broad range of server to switch architectures.
    - Enables large modular switches with high port counts.
    - Allows otherwise stranded ports to be utilized.
- An objective to define a single-lane 25 Gb/s PHY for operation over MMF:
  - Standardizes chip-to-module interface and optical interface specs.
    - Enables AOC and transceiver implementations.
    - Market data for 10 Gb/s shows both will be important for 25 Gb/s server interconnect.

Conclusion

An objective to define a single-lane 25 Gb/s PHY for operation over MMF should be supported

Thank you !

Appendix: Diversity of DC architectures - references

a. ‘Navigating the Pros and Cons of Structured Cabling vs. Top of Rack in the Data Center’, CCCA , May 2013, *http://www.cccassoc.org/files/2114/0138/9928/WP-SCS_vs_ToR_WP_Final_May_2013.pdf*

b. ‘40GBASE-T advantages and use cases’, Andy Jiminez, 802.3bq, July 2014, jiminez_3bq_01_0711.pdf

c. ‘25GE server to switch architectures’, Scott Kipp, 25G Ethernet Study Group Meeting, Ottawa, Sept 2014, kipp_25GE_01_0914.pdf

d. ‘Deploying 10GBASE-T as the low cost path to the cloud’, Panduit, Intel & Cisco joint white paper, *http://www.panduit.com/ccurl/413/996/zcan02_10gbase_tecosystem.pdf*

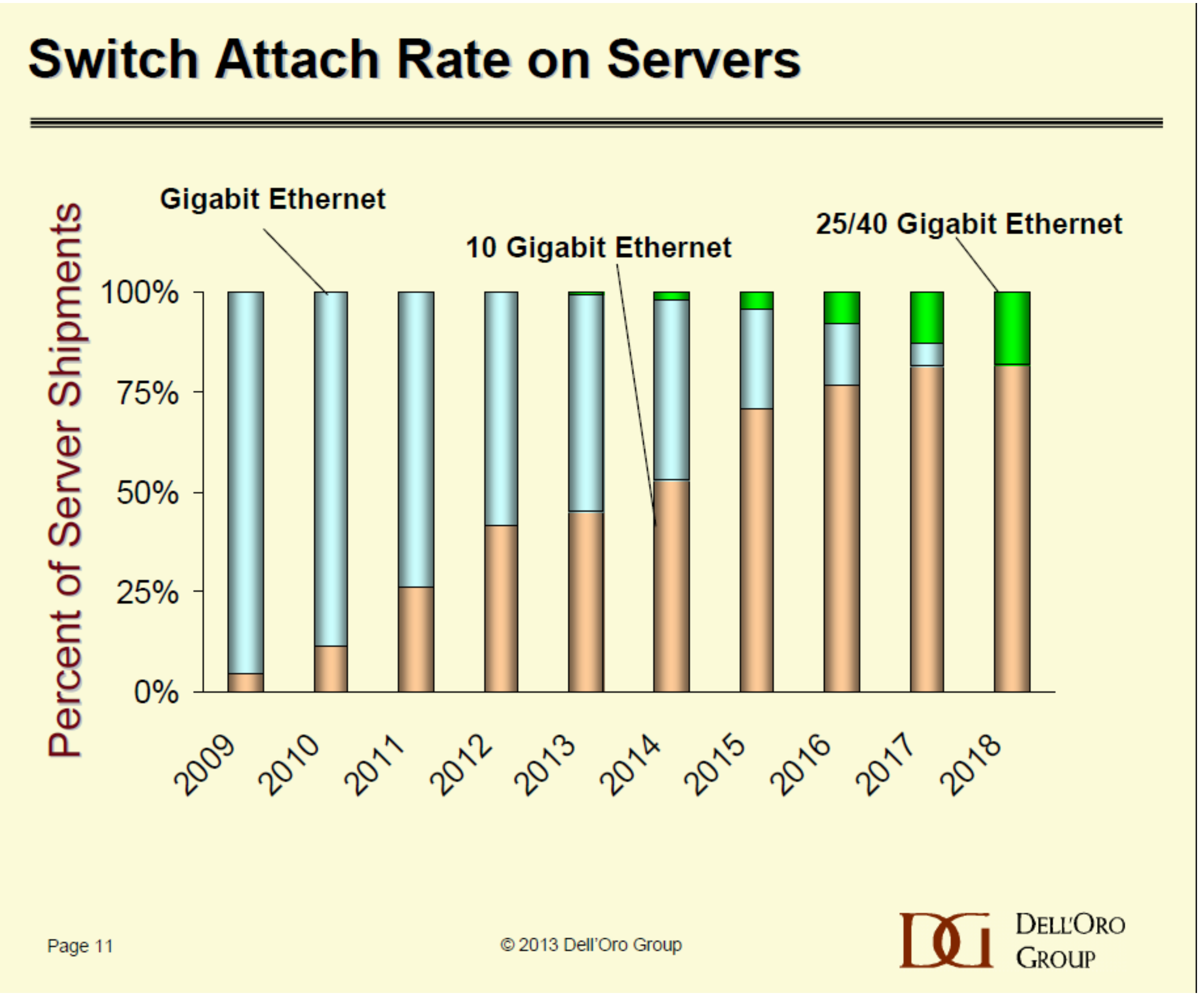
e. ‘Reconsidering Physical Topologies with 10GBASE-T’, Broadcom white paper, May 2013, *http://www.broadcom.com/collateral/wp/84848-WP100-R.pdf*

f. ‘Link distance and server connectivity’, Scott Kipp, Next Generation Base T study group, Nov 2012, *http://www.ieee802.org/3/NGBASET/public/nov12/kipp_01a_1112_ngbt.pdf*

MoR & EoR topologies allow fewer stranded ports on servers and on switches.
Substantial cost-effectiveness improvements.
Better flexibility for thermal planning in MoR and EoR architectures.
Support of mixed applications.
Support of reconfiguration.

Back up

25G Ethernet servers are coming



Switch Attach Rate on Servers

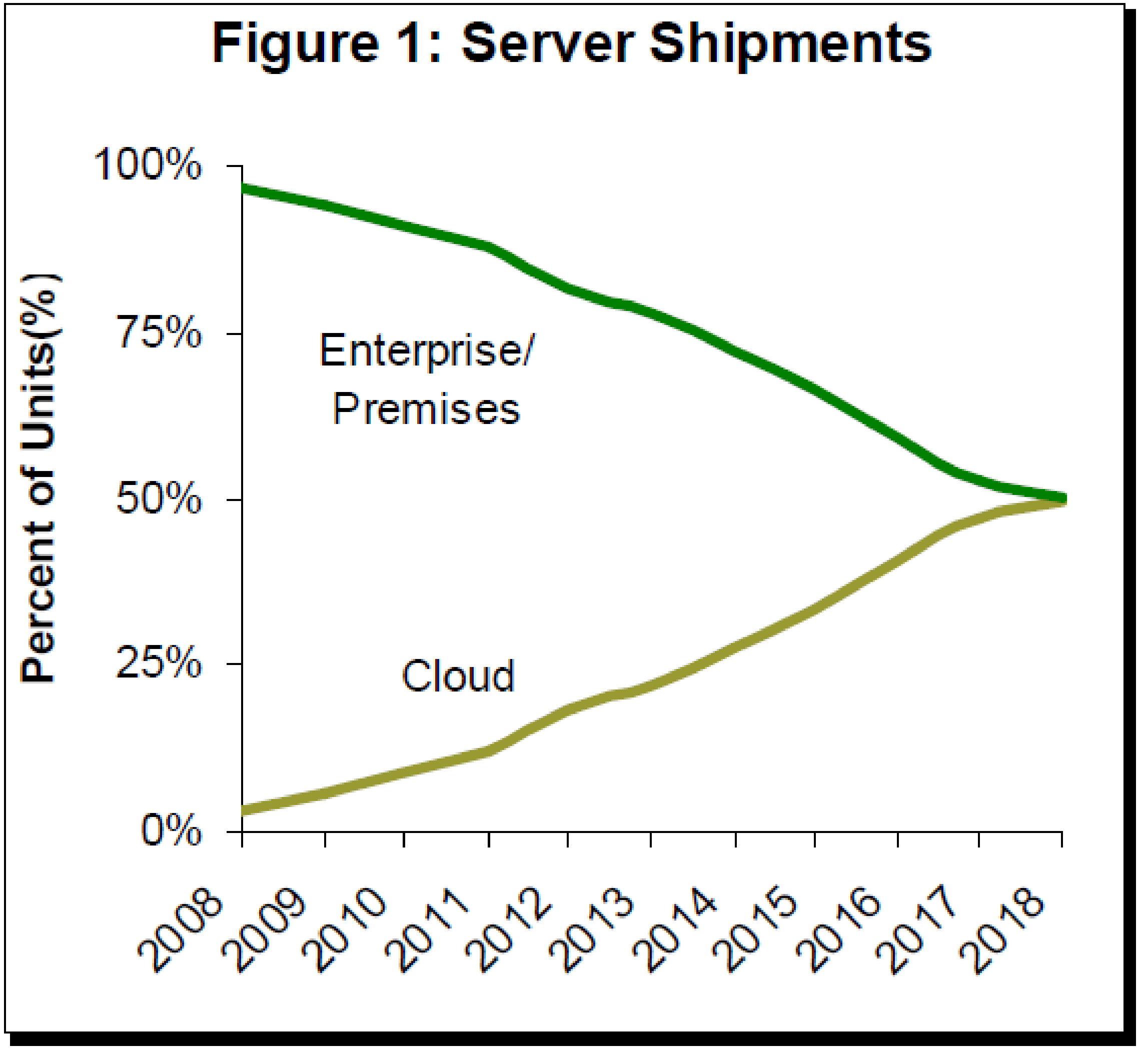
Page 11

© 2013 Dell'Oro Group

http://www.ethernetsummit.com/English/Collaterals/Proceedings/2014/20140430_Keynote2_Weckel.pdf

Most servers will be enterprise until at least 2018

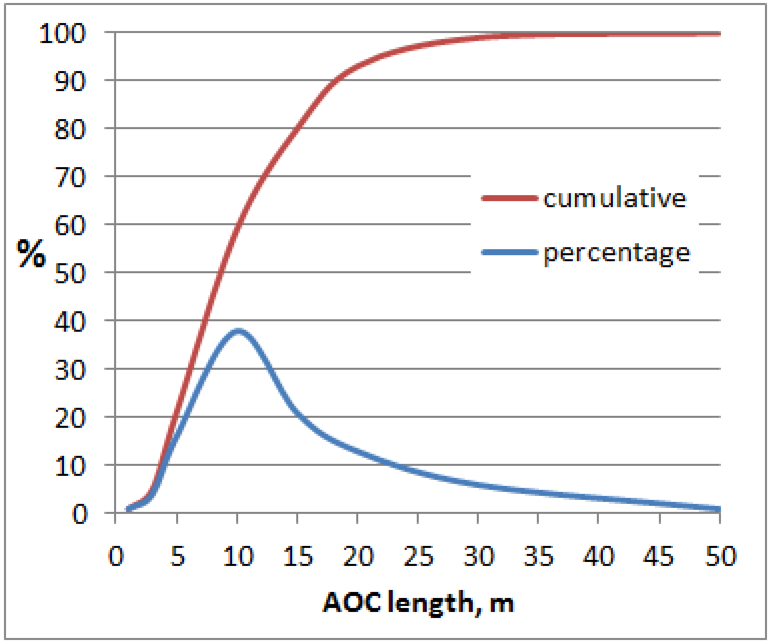
from "Dell'Oro Controller and Adapter Forecast Summary – July 2014"



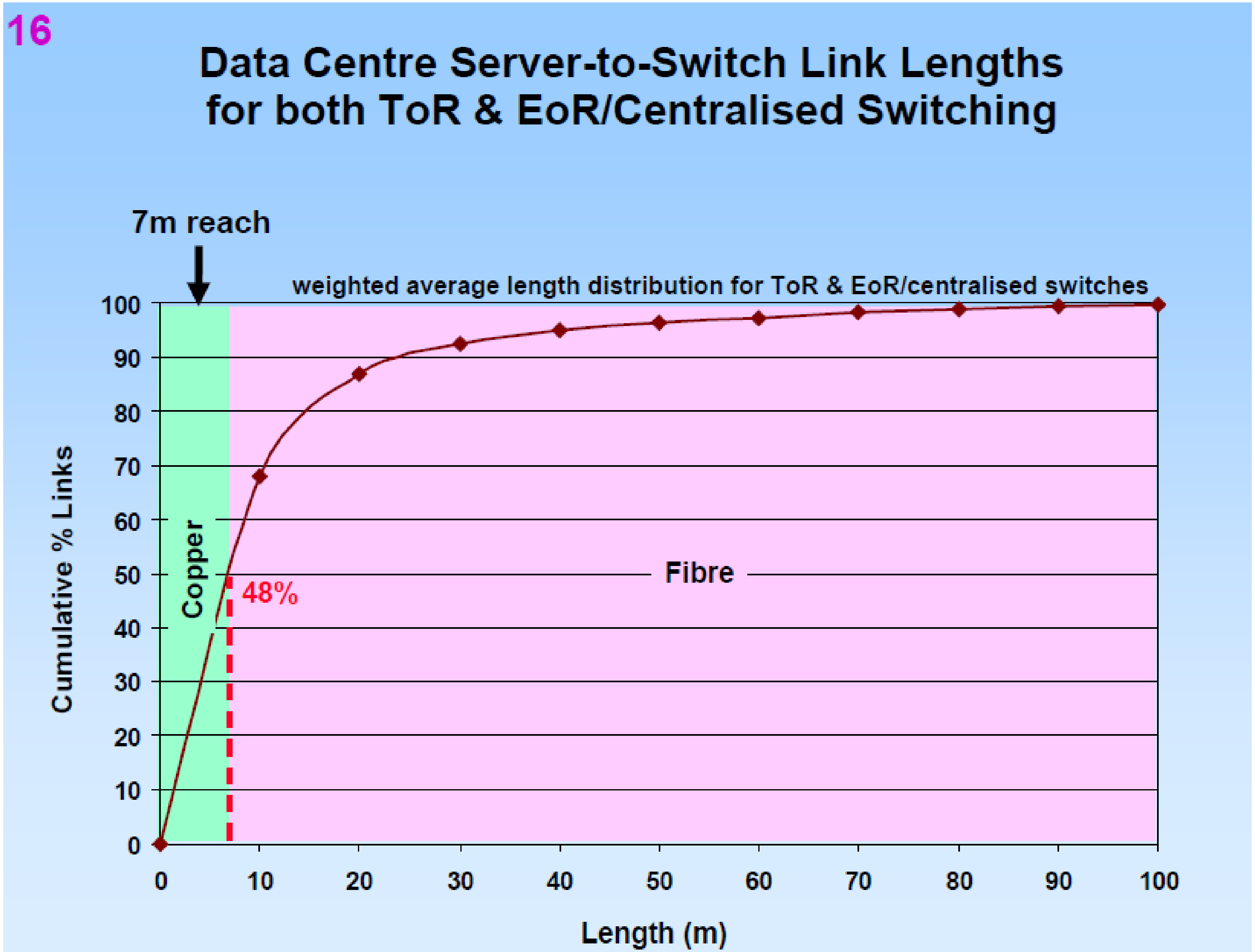
- Enterprise data centers will still be ~50% of market in 2018
- Largely follow TIA-942, and deploy a variety of access switch placements including MoR and EoR.
- Requires optics support: AOCs and pluggables.

Active Optical Cables (AOC) length distributions

- Pragmatic limitations to reach, reflected in reach distributions
- Average length < 10 m, 90% < 18 m
  - (Finisar: sales data)
- Note: Pluggable optics links exceed AOC volumes by ~ 3:1



- AOCs offer longer reach than direct attach copper
  - one of the solutions helping to maximize efficiency of server to access switch interconnect
    - but not compatible with structured cabling



16

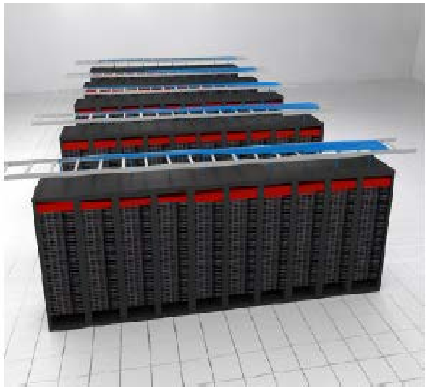
Data Centre Server-to-Switch Link Lengths for both ToR & EoR/Centralised Switching

From: flatman_01_0311.pdf, reproduced with kind permission of Alan Flatman

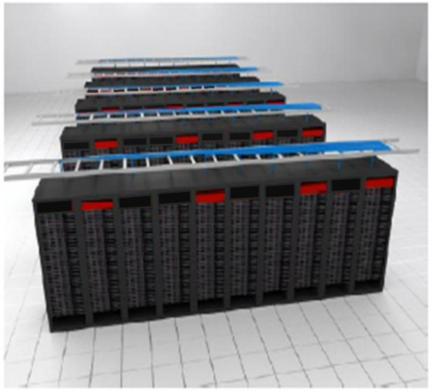
http://www.ieee802.org/3/100GCU/public/mar11/flatman_01_0311.pdf

Top of Rack, Adjacent Rack, Middle of Row, End of Row

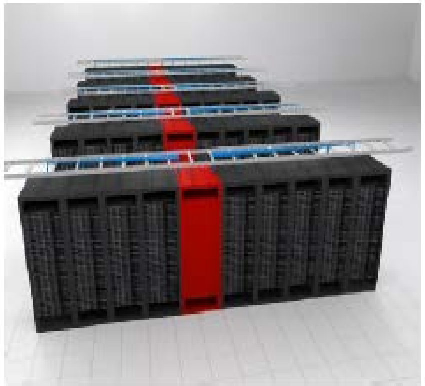
ToR
(up to 3 m)



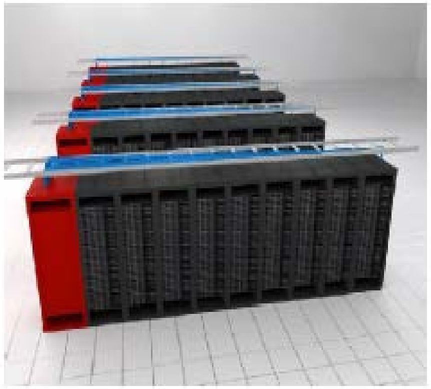
Cabinet
to cabinet



MoR
(up to 15 m)



EoR
(up to 50 m)

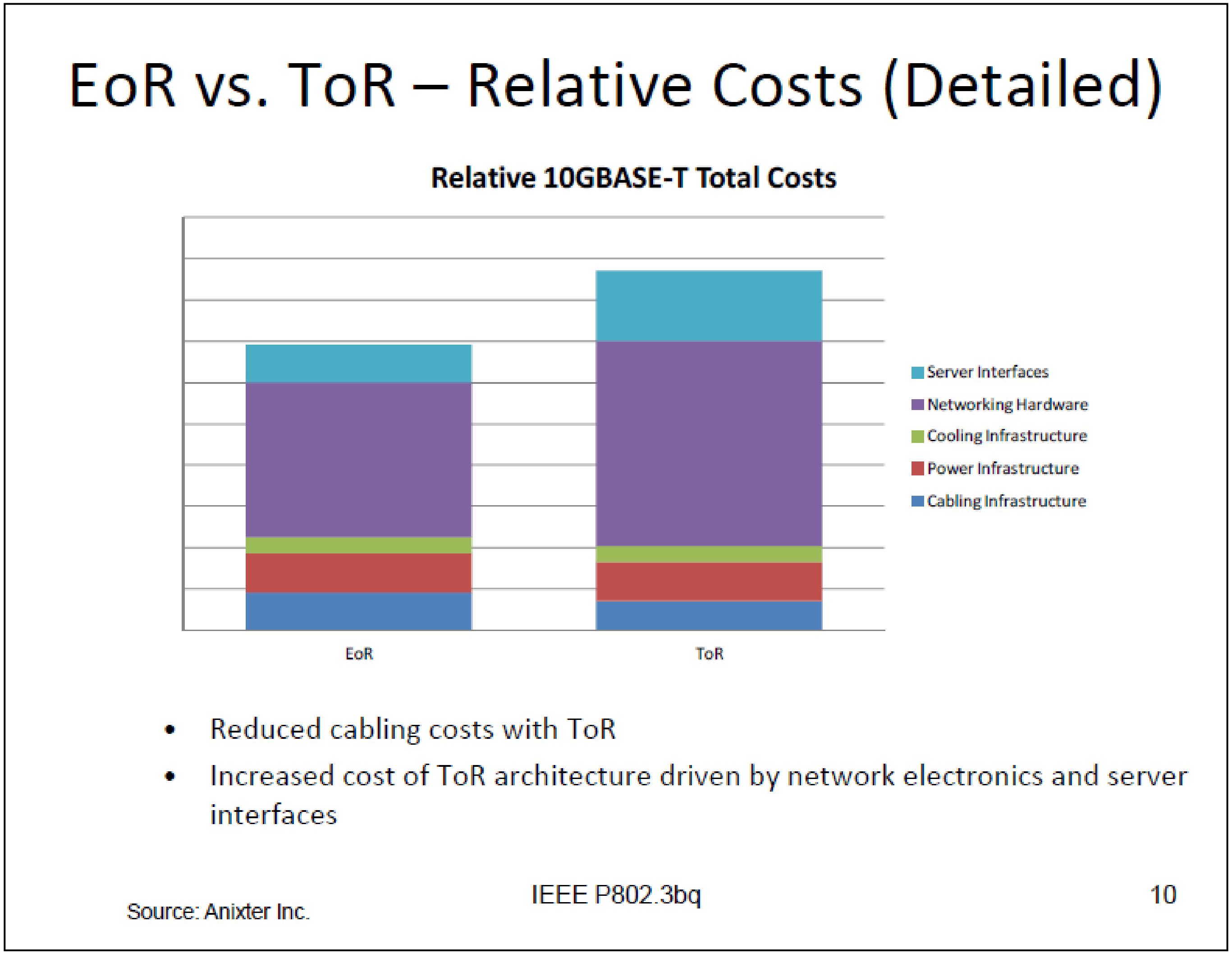


Intra-rack addressed by 25 Gb/s copper direct attach

Not addressed by 25 Gb/s copper direct attach

Pictures from jiminez_3bq_01_0711.pdf, 802.3bq

From “40GBASE-T advantages and use cases” (jiminez_3bq_01_0711.pdf, 802.3bq)



EoR vs. ToR – Relative Costs (Detailed)

- Reduced cabling costs with ToR
- Increased cost of ToR architecture driven by network electronics and server interfaces

Source: Anixter Inc.

IEEE P802.3bq 10

- Lower cost achieved by maximizing switch port utilization