# Enabling Dense 200GbE and 400GbE

Kapil Shrikhande
Innovium

July 13, 2021
IEEE 802.3 Beyond 400Gb/s Ethernet Study Group, July 2021 meeting session

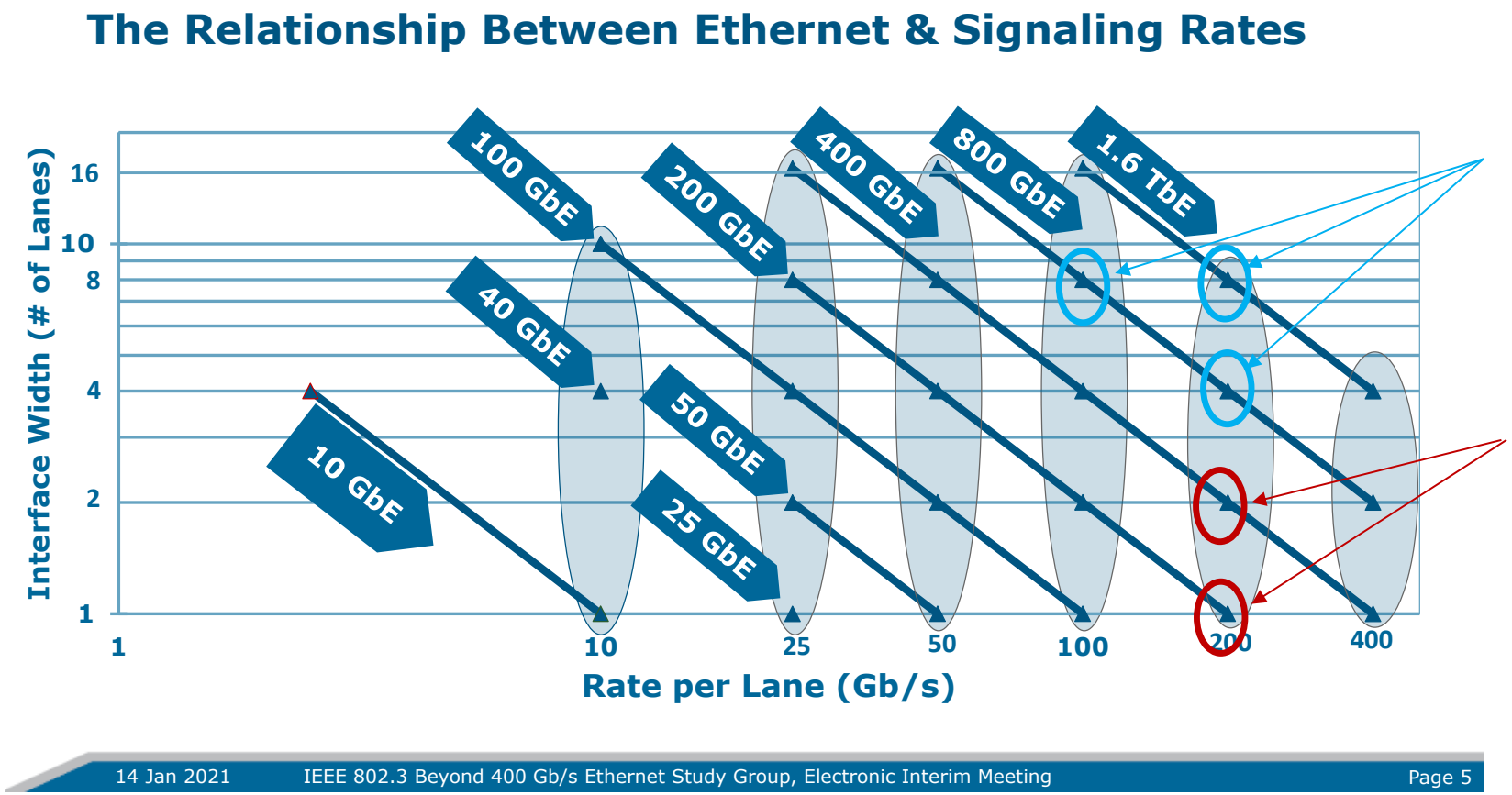
# Supporters

- Rob Stone, Facebook
- Brad Booth, Microsoft
- Mark Gustlin, Cisco
- Dave Ofelt, Juniper
- Chris Cole, II-VI
- Vipul Bhatt, II-VI
- Roberto Rodes, II-VI
- Phil Sun, Credo
- Osa Mok, Innolight
- Rich Mellitz, Samtec
- Ryan Latchman, Macom
- Nathan Tracy, TE Connectivity
- Mike Li, Intel
- Brian Welch, Cisco
- Adee Ran, Cisco
- Xinyuan Wang, Huawei
- Jeff Maki, Juniper
- Thananya Baldwin, Keysight
- Jerry Pepper, Keysight
- Matt Brown, Huawei
- Jeff Slavick, Broadcom
- Eugene Opsasnick, Broadcom
- Lenin Patra, Marvell
- Arash Farhoodfar, Marvell
- Scott Schube, Intel
- Scott Sommers, Molex
- Ali Ghiasi, Ghiasi Quantum
- Sam Kocsis, Amphenol

# Goals of this presentation

- Highlight lack of objectives for Dense 200 and 400GbE
- Discuss specifications needed for future devices and applications
- Discuss possible objectives for the SG to consider

# Enabling denser Ethernet using higher signaling rates




Higher Speeds (800GbE, 1.6TbE) addressed by B400G SG

Reducing Lane counts of existing MACs (200 and 400GbE) ?
***-- Not yet addressed by B400G SG***



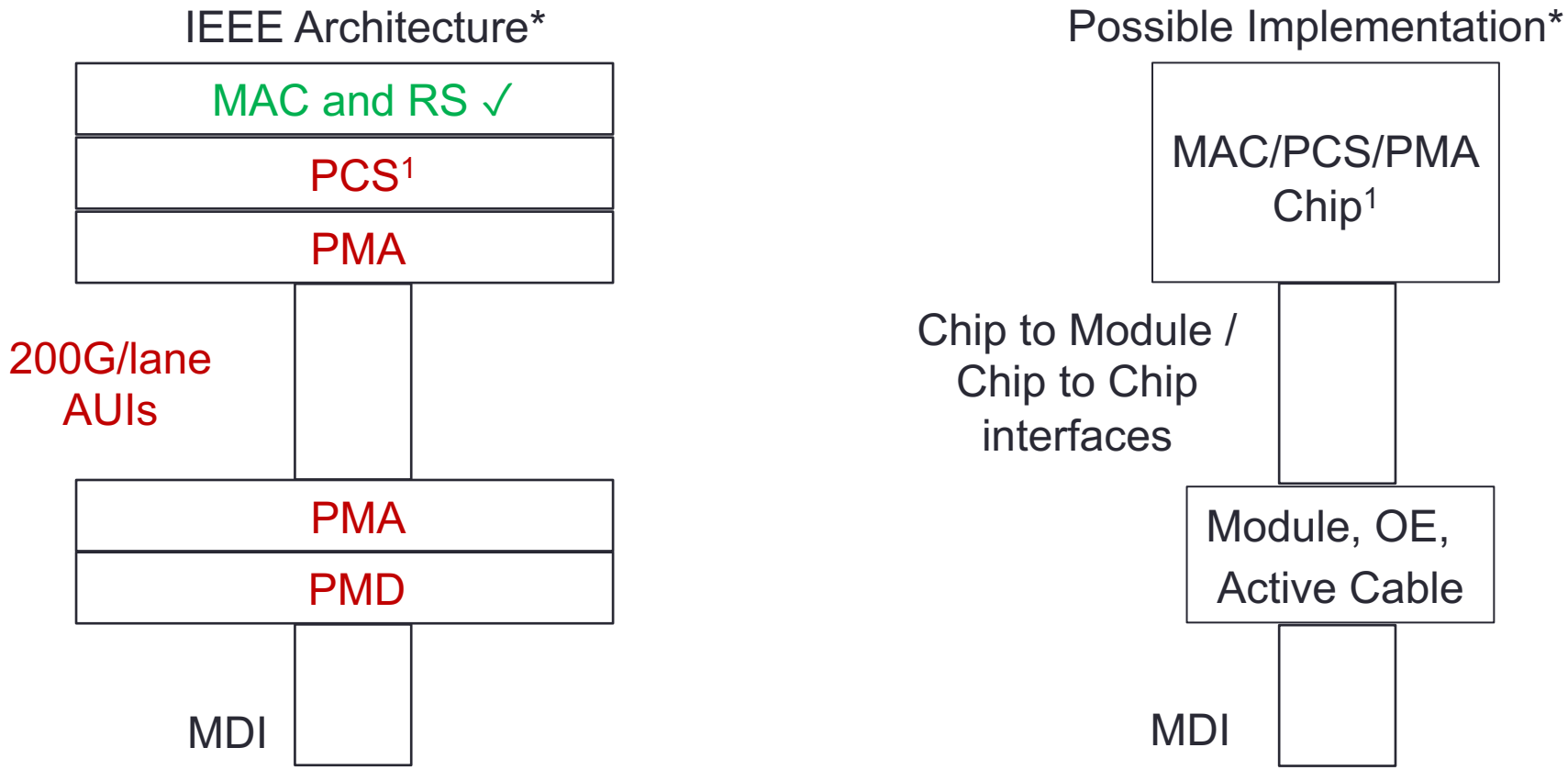
*Source of graph : John D'Ambrosia, presentation at Jan'21 B400G SG meeting*

# 200 and 400GbE in Silicon devices using 200G SerDes

- Silicon devices driving 200G SerDes IO:
  - Switch Devices will use 200G SerDes to bring MAC bandwidth in/out of > 100 Tb/s devices
  - AI/ML and other End-point Devices will use 200G SerDes for higher capacity and denser MACs
- **Devices will use 200G SerDes to save area, power and cost of 200 and 400GbE**
  - 400GbE : 4 x 100G SerDes ➔ 2 x 200G SerDes
  - 200GbE : 2 x 100G SerDes ➔ 1 x 200G SerDes
- **Switches will use 200G SerDes to increase radix of 200/400GbE MACs using 512 SerDes lanes**
  - 102.4T (512 x 200G SerDes) Switch Device Radix : 256 x 400GbE, 512 x 200GbE
- **Combined with 200G Optical Lane technology provides efficient end-end solutions**

***Silicon Devices will use 200G SerDes to enable denser and higher radix 200 and 400GbE Provide efficient end-end solutions matched with 200G optical lanes***

# Specifications to enable Dense 200 and 400GbE




[1] PCS includes FEC in these diagrams; but architectures where FEC is a separate sub-layer is possible too

* Representation of IEEE Arch. and Possible implementation -- concept borrowed from M. Gustlin

# Current B400G SG Objectives : no 200 and 400GbE yet

| | |
|---|---|
| * | Adopted by B400G SG, Apr 2021 |
| ** | Adopted by B400G SG Apr 26, 2021 |
| *** | Adopted by B400G SG May 3, 2021 |
| **** | Adopted by B400G SG May 17, 2021 |
| # | Adopted by B400G SG Jun 3, 2021 |
| | Approval by 802.3 WG Pending |

- Non-Rate Specific
  - Support full-duplex operation only *
  - Preserve the Ethernet frame format utilizing the Ethernet MAC *
  - Preserve minimum and maximum FrameSize of current IEEE 802.3 standard *
  - Support a BER of better than or equal to 10 -13 at the MAC/PLS service interface (or the frame loss ratio equivalent) **
  - Provide support to enable mapping over OTN ***
- 800 Gb/s Related
  - Support a MAC data rate of 800 Gb/s *
  - Support optional eight-lane 800 Gb/s attachment unit interfaces for chip-to-module and chip-to-chip applications ****
  - Support optional four-lane 800 Gb/s attachment unit interfaces for chip-to-module and chip-to-chip applications ****
  - Define a physical layer specification that supports 800 Gb/s operation:
    - over 8 pairs of MMF with lengths up to at least 50 m *
    - over 8 pairs of MMF with lengths up to at least 100 m *
    - over 8 pairs of SMF with lengths up to at least 500 m *
    - over 8 pairs of SMF with lengths up to at least 2 km #
    - over 4 pairs of SMF with lengths up to at least 500 m *
    - over 4 pairs of SMF with lengths up to at least 2 km *
    - over 4 wavelengths over a single SMF in each direction with lengths up to at least 2 km *
    - over a single SMF in each direction with lengths up to at least 10 km *
    - over a single SMF in each direction with lengths up to at least 40 km *
- 1.6 Tb/s Related
  - Support a MAC data rate of 1.6 Tb/s #
  - Support optional eight-lane 1.6 Tb/s attachment unit interfaces for chip-to-module and chip-to-chip applications #
  - Define a physical layer specification that supports 1.6 Tb/s operation:
    - over 8 pairs of SMF with lengths up to at least 500 m #
    - over 8 pairs of SMF with lengths up to at least 2 km #

*800GbE Related*

*1.6TbE Related*



***Without objectives, 200 and 400GbE will not be included in the PAR for the Task Force***

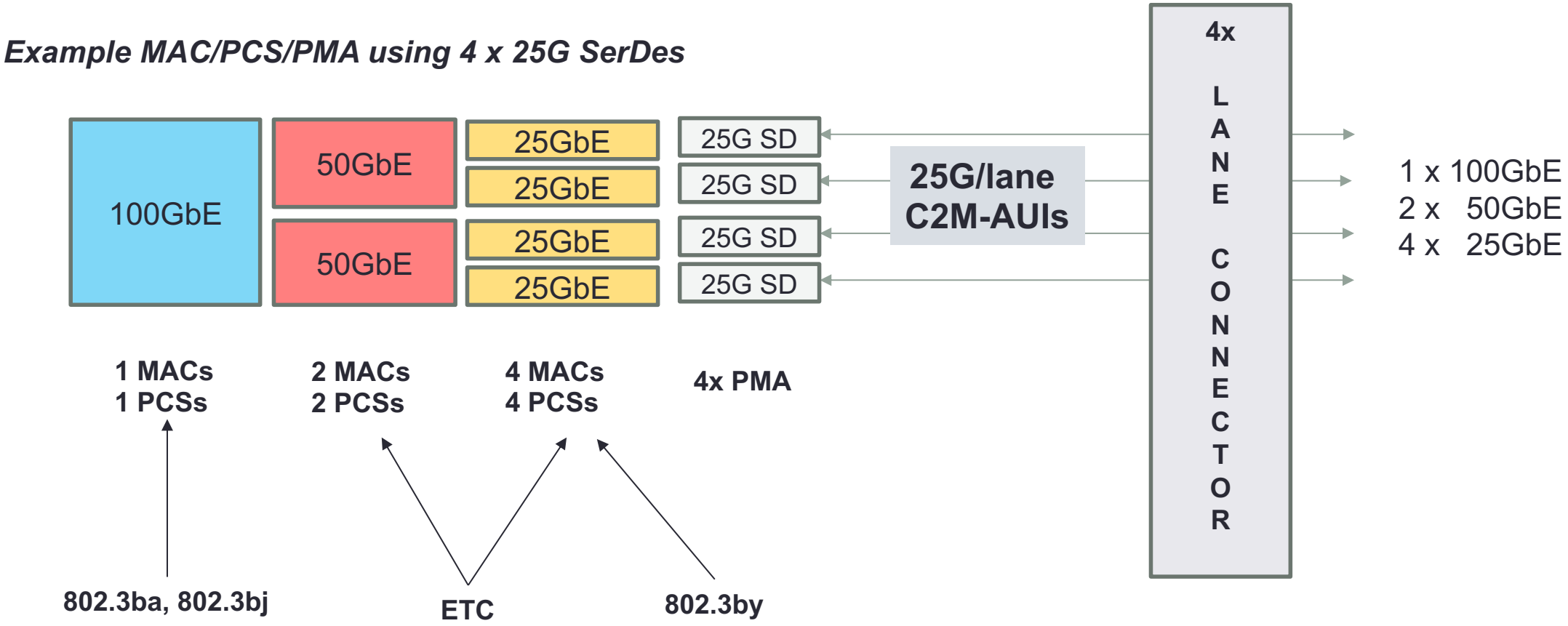
# Implications for the future Task Force

- PCSs for 1.6TbE, 800GbE PCSs using 200G/lane PMAs will be addressed
  - But PCSs for 200GbE and 400GbE using 200G PMA lanes will not be addressed
- PMAs for 1.6TbE, 800GbE using 200G/lane will be addressed
  - But 200GbE and 400GbE PMAs using 200G PMA lanes will not (even though same PMA technology)
- C2C and C2M AUIs using 200G AUI lanes will be specified for 1.6TbE, 800GbE
  - But not 200GAUI1 and 400GAUI-2 (even though these are 1 and 2 lanes of 800GAUI-4)
- Parallel SMF “DR” PHYs for 1.6TbE, 800GbE using 200G optical lanes will be addressed
  - But not 200GbE and 400GbE ”DR” PHYs (even though these are 1 and 2 lanes of 800G-DR4 PMD)

***Address all foundational specifications at the same time, reuse where possible***

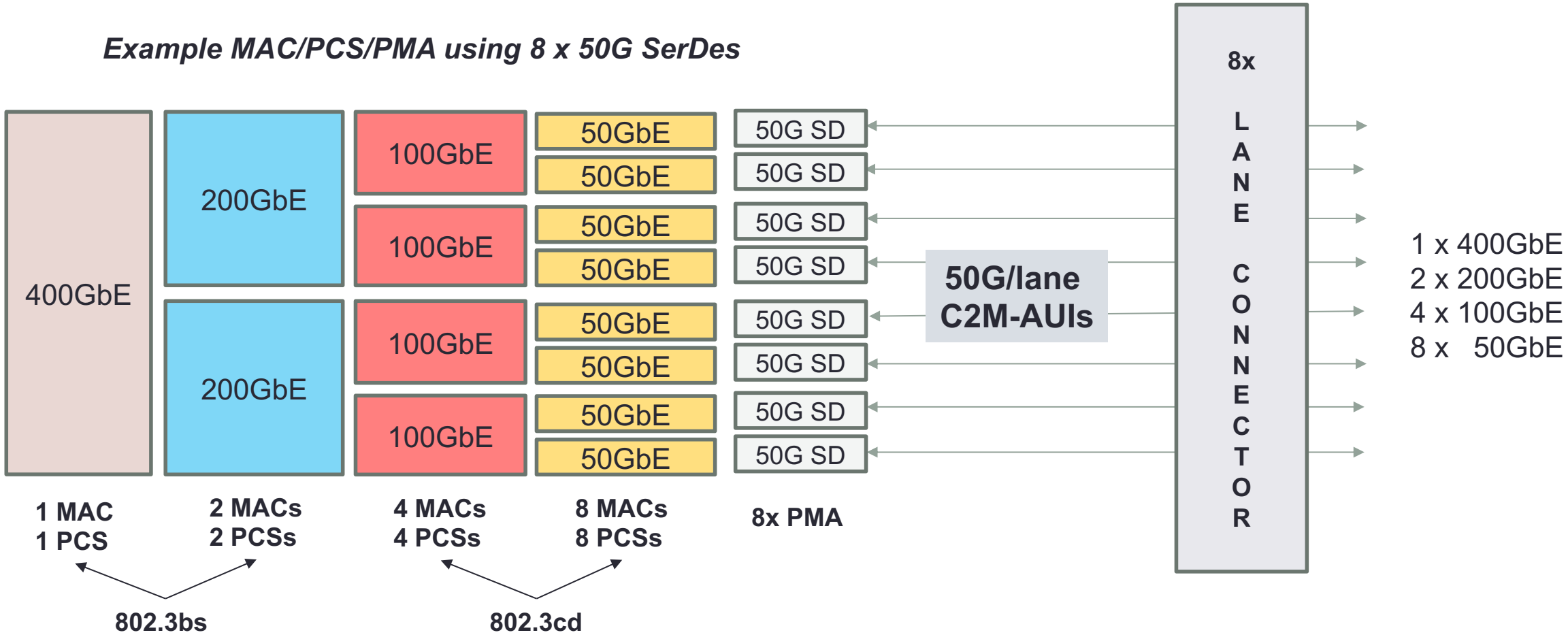
# Past and current generations

# Past : 25G/lane AUI generation



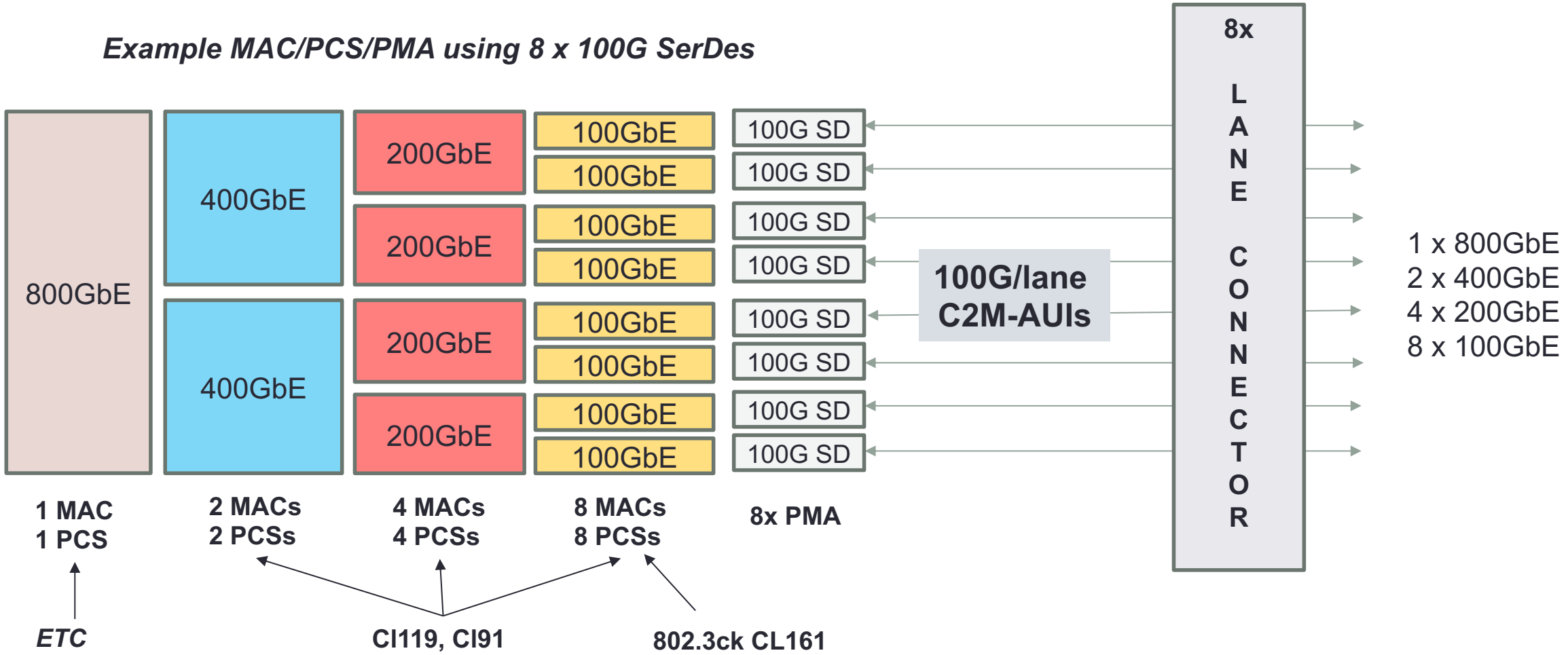

***IEEE 25GbE lagged ETC and MAC/PCS devices. IEEE did not specify 50GbE (2x25G) until later in 802.3cd***

# Past : 50G/lane AUI generation



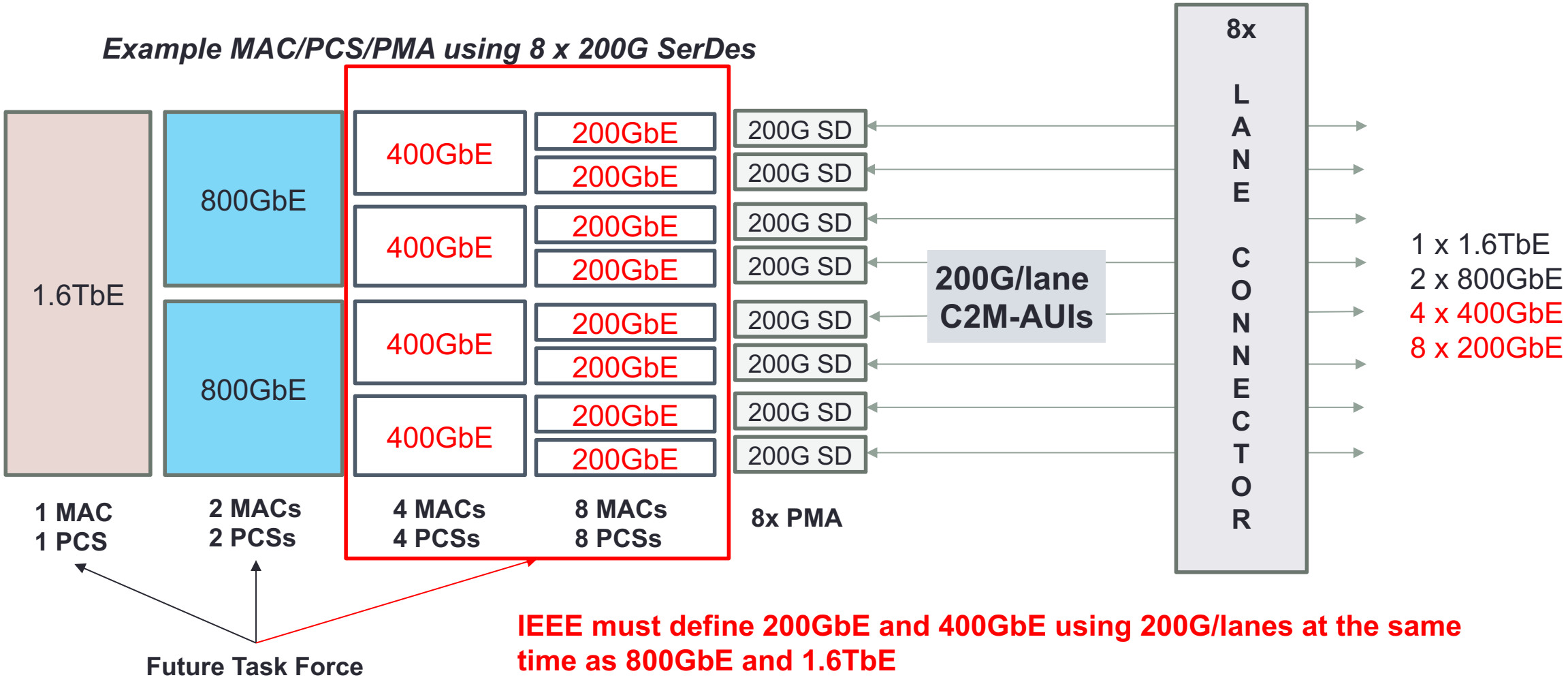

***802.3cd lagged .3bs by ~ 1 year, but 50 and 100GbE PCSs built on existing Cl91/Cl82 – completed on time***

# Current : 100G/lane AUI generation




***802.3ck specifying 100/200/400GbE using 100G/lane at the same time. IEEE will lag ETC on 800GbE***

# Future : 200G/lane AUI generation




**IEEE must define 200GbE and 400GbE using 200G/lanes at the same time as 800GbE and 1.6TbE**

# Applying lessons to 200G/lane AUI generation

- Devices using 200G SerDes will want to implement all MACs/PCSs using 1, 2, 4, 8 PMA lanes
- Systems will be built with 8 x 200G AUIs and 8-lane connectors (16-lane connectors possible too)
- Cannot fully anticipate uses of 200G/lane PMAs/AUIs and PMDs now -- applications are getting diverse
  - Modules built for 1600G-DR8 and 800G-DR4 can be re-used for 400G-DR2 and 200G-DR
  - IEEE 200G/lane AUIs built for 1.6TbE/800GbE can get reused for 200 and 400GbE MSA optics, and AOCs
- Industry will work to close any gaps that the IEEE leaves open, esp. in the PCSs
  - But there is time to discuss this and close the gaps

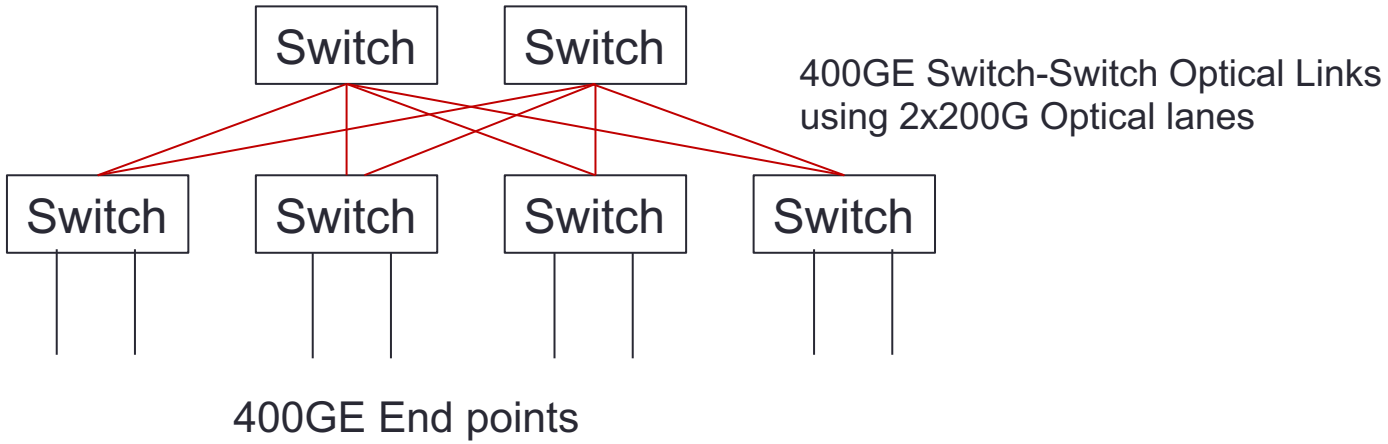
***Provide flexibility to implementors by specifying MAC/PCSs over 1, 2, 4, 8 PMA/AUI lanes***

***Specify 200 and 400GbE PMDs based on 800GbE & 1.6TbE 200G per optical lane "DR" optics***

# Applications of Dense 200 and 400GbE

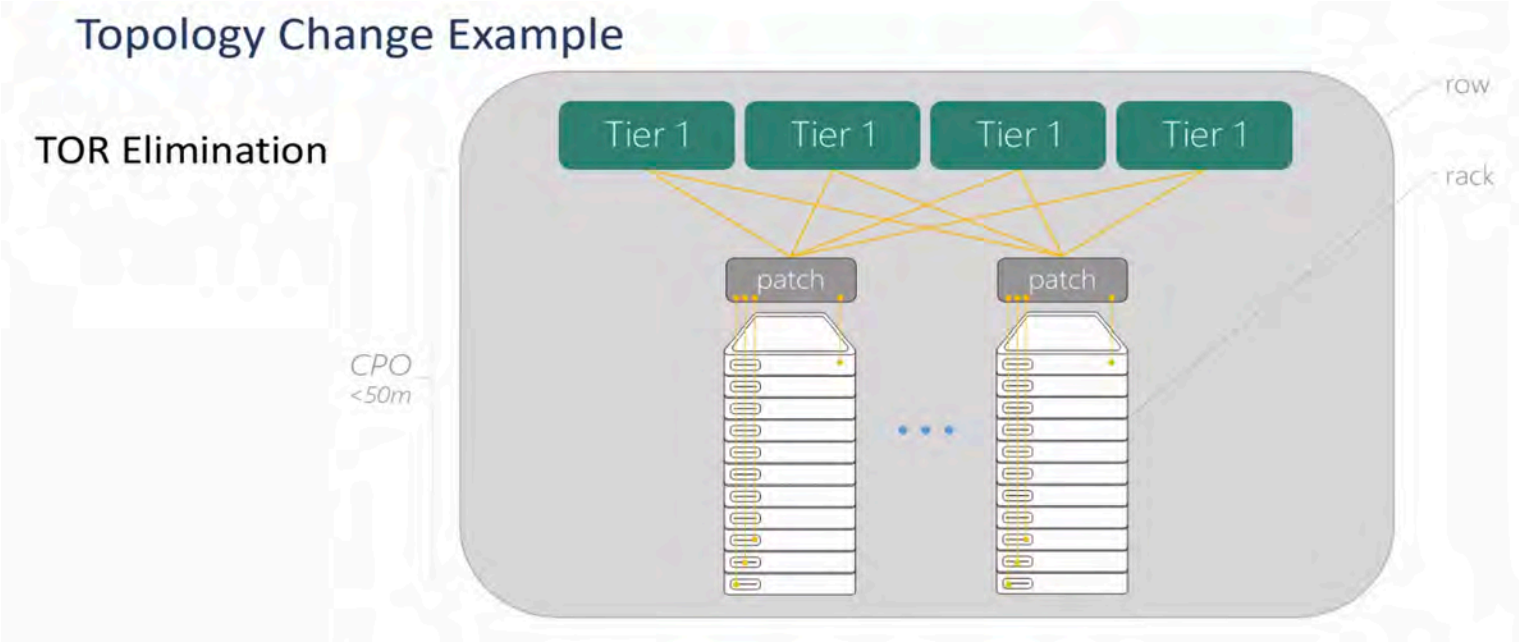
# 1. Switch-Switch

- Assume : 400GE end-points and 102.4T switches using 200G SerDes
- *Switch with 400GbE MAC/PCSs using 2x200G SerDes (PMA) lanes enables flatter networks*
  - In 128 x 800GbE configuration : can connect to ~ 8k end-points
  - In 256 x 400GbE configuration : can connect to ~ 32k end-points
- *Same benefits when using 200GbE end-points and 200GE (1x200G) MACs in Switches*
- *Switch-Switch Links : 200GE-DR, 400GE-DR2, 200GE and 400GE Active Optical Cables*



# 2. EOR : Switch-End Host

- *Optical links between Tier 1 switches and End-hosts in the rack*
  - *200G-DR, 400G-DR2 optical links*




*Source: Brad Booth, OFC 2021 – "Inside the Data Center"*

# New SG Objectives

# Objectives to address 200 and 400GbE applications

| Speed | Solution | Specifications required | New SG Objectives |
|---|---|---|---|
| 200GbE | 200GbE Active Cables | 200GbE PCS<br>200GbE PMA (1x200G)<br>200GAUI-1 | 200GAUI-1 C2C and C2M |
| 400GbE | 400GbE Active Cables | 400GbE PCS<br>400GbE PMA (2x200G)<br>400GAUI-2 | 400GAUI-2 C2C and C2M |
| | | | |
| 200GbE | 200GBASE-DR | 200GbE PCS<br>200GbE PMA (1x200G)<br>200GAUI-1<br>200GBASE-DR PMD | 200GBASE-DR PHY |
| 400GbE | 400GBASE-DR2 | 400GbE PCS<br>400GbE PMA (2x200G)<br>400GAUI-2<br>400GBASE-DR2 | 400GBASE-DR2 PHY |

***Adopt the following Objectives to enable Dense 200 and 400GbE applications !***

# Thank You