Architectural Thoughts – 25G Interconnect

BRAD BOOTH

MICROSOFT

Supporters

Bob Felderman, Google
Kapil Shrikhande, Dell
Yong Kim, Broadcom
Mike Andrewartha, Microsoft
Gary Nicholl, Cisco
Oded Wertheim, Mellanox
Greg McSorley, Amphenol
Thananya Baldwin, IXIA
Chengbin Wu, ZTE Corporation
Scott Kipp, Brocade
Daniel Koehler, MorethanIP GmbH
Dave Husak, Plexxi
Arthur Lee, MediaTek
Tom Ambrose, Emulex
Scott Jacobson, Cadence
Jaspreet Singh Gambhir, Synopsys
Rich Mellitz, Intel
Rob Stone, Broadcom
Kent Lusted, Intel
Qianggao Hu, Accelink
Benny Koren, Mellanox
Robert Chu, Alpha Networks
Eric Baden, Broadcom
Cedrik Begin, Cisco
Yi Jiang, Accelink
Andre Viera, Plexxi
Chloe Lin, Alpha Networks
Vincent Tseng, MediaTek
Adil Haque, Altera
Rochan Sankar, Broadcom
Y.Y. Tsai, Delta Networks
Peter Jones, Cisco
Yonatan Malkiman, Mellanox
David Chalupsky, Intel

Assumptions

Switch implementations likely to support multiple data rates; therefore likely to have all PCS and FEC versions available

Auto-negotiation is the available across most copper port types

- Exception: 10G SFP+ direct attach copper, doesn't exist in IEEE 802.3
- Optical ports will "force" the port type

10G/25G/40G NIC would have multi-lane capability due to 40G

- Could provide 4x10GBASE-R and n x25GBASE-R
- Likely to use QSFP28 module

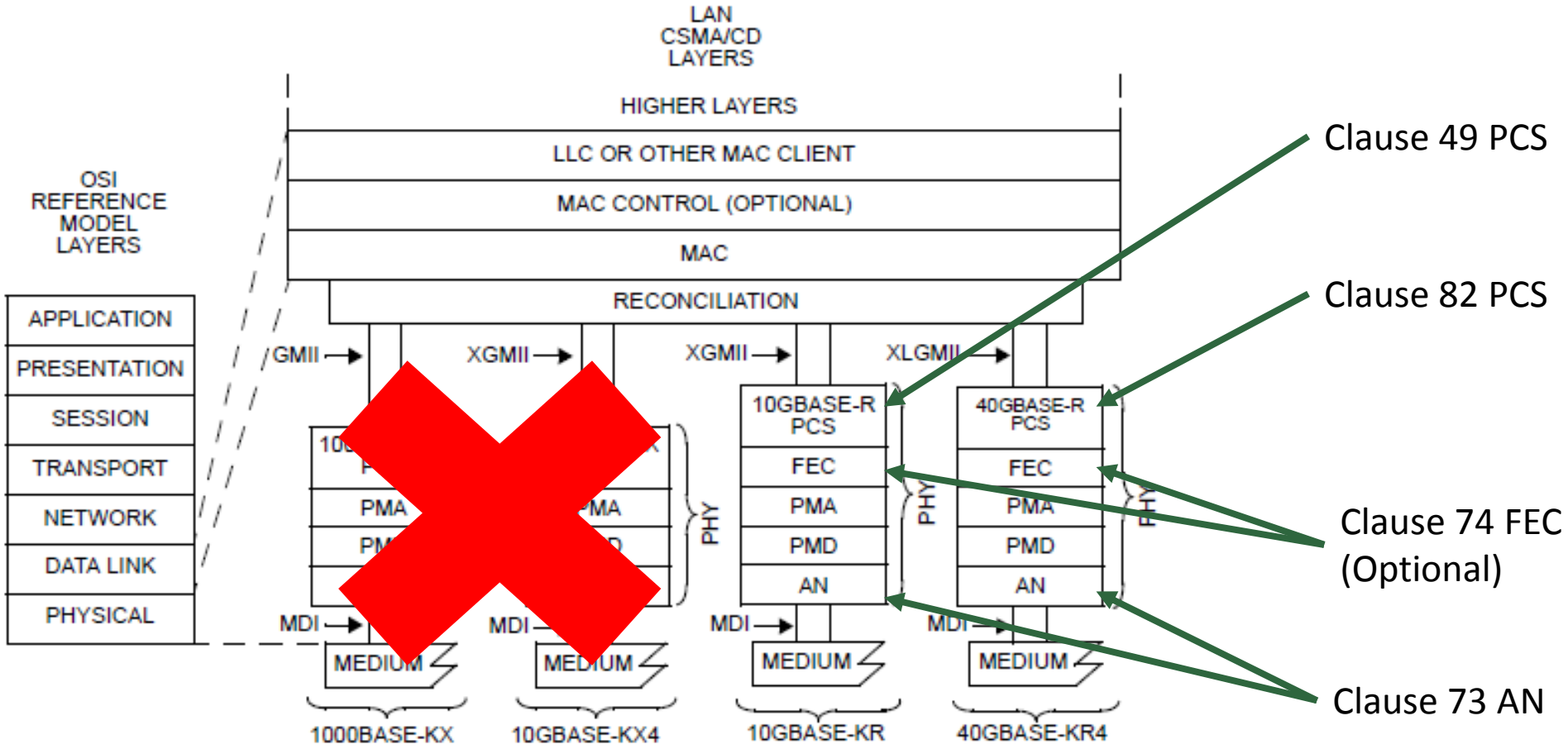
10G/25G NIC could be a single-port device

- Implementations likely to use SFP28 module

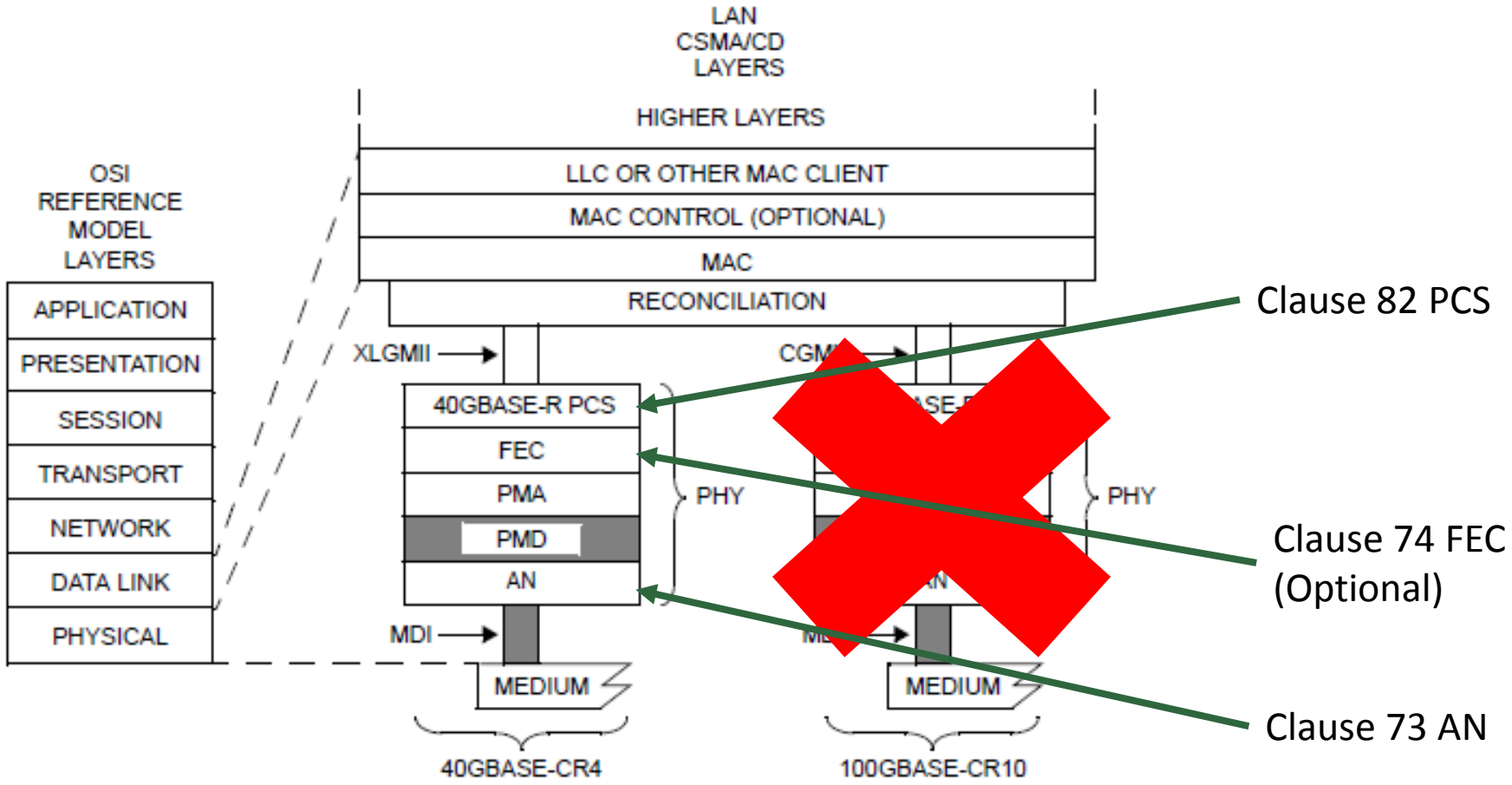
OTN, if used, not likely to be between server-to-switch connection; only switch-to-switch

Copper

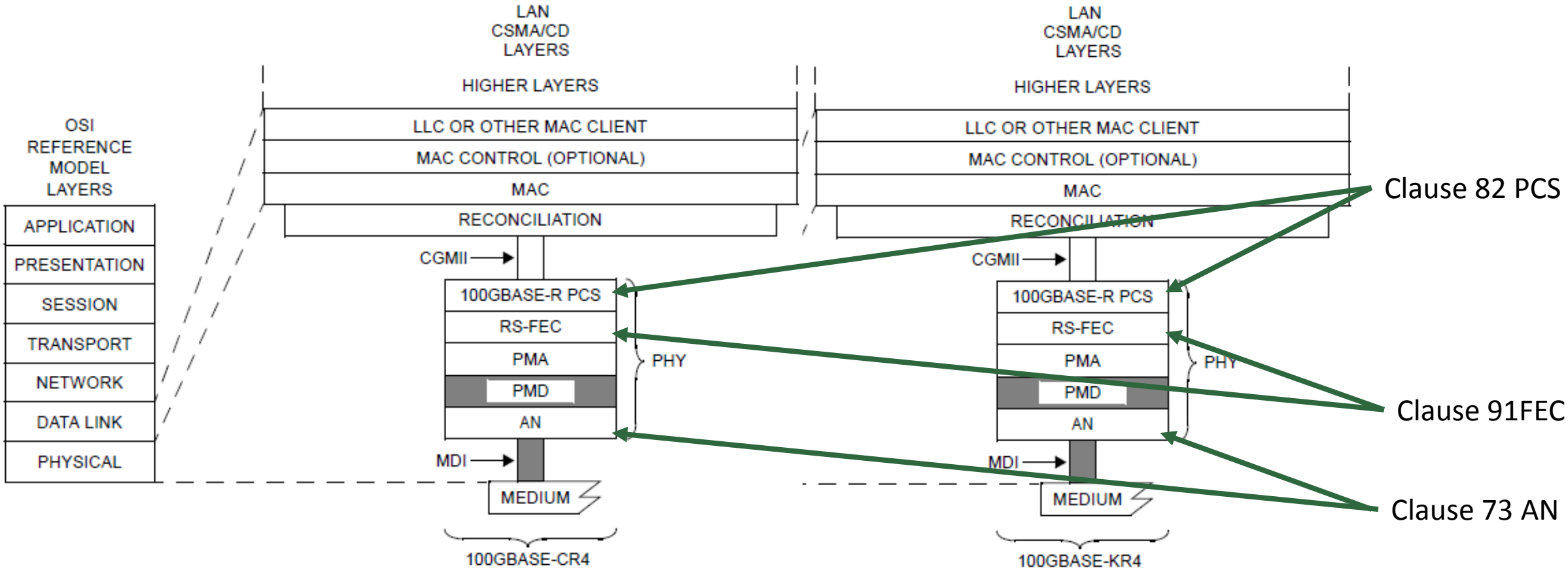
802.3-2012 10G & 40G Backplane



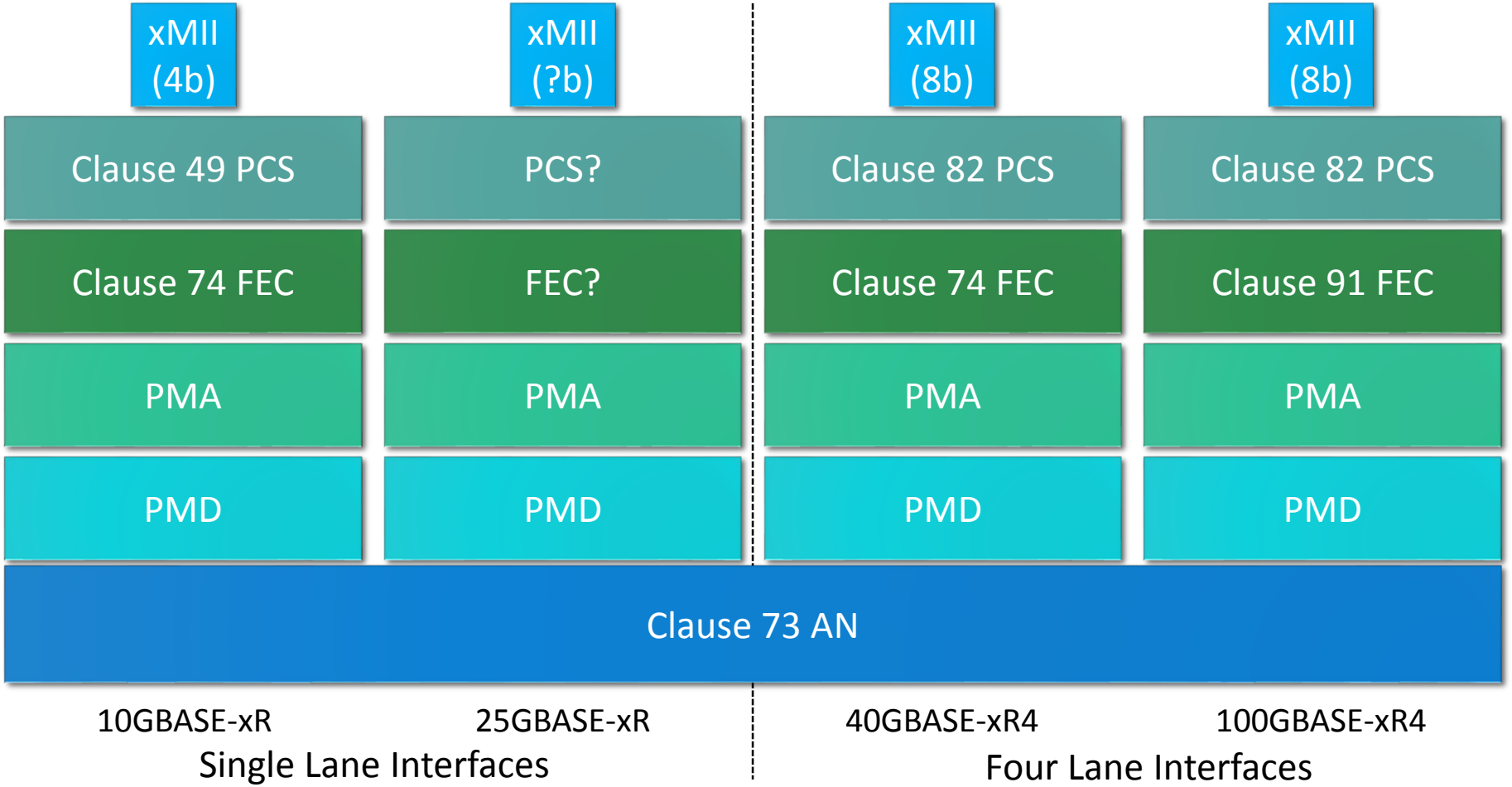
802.3-2012 40G Twinax



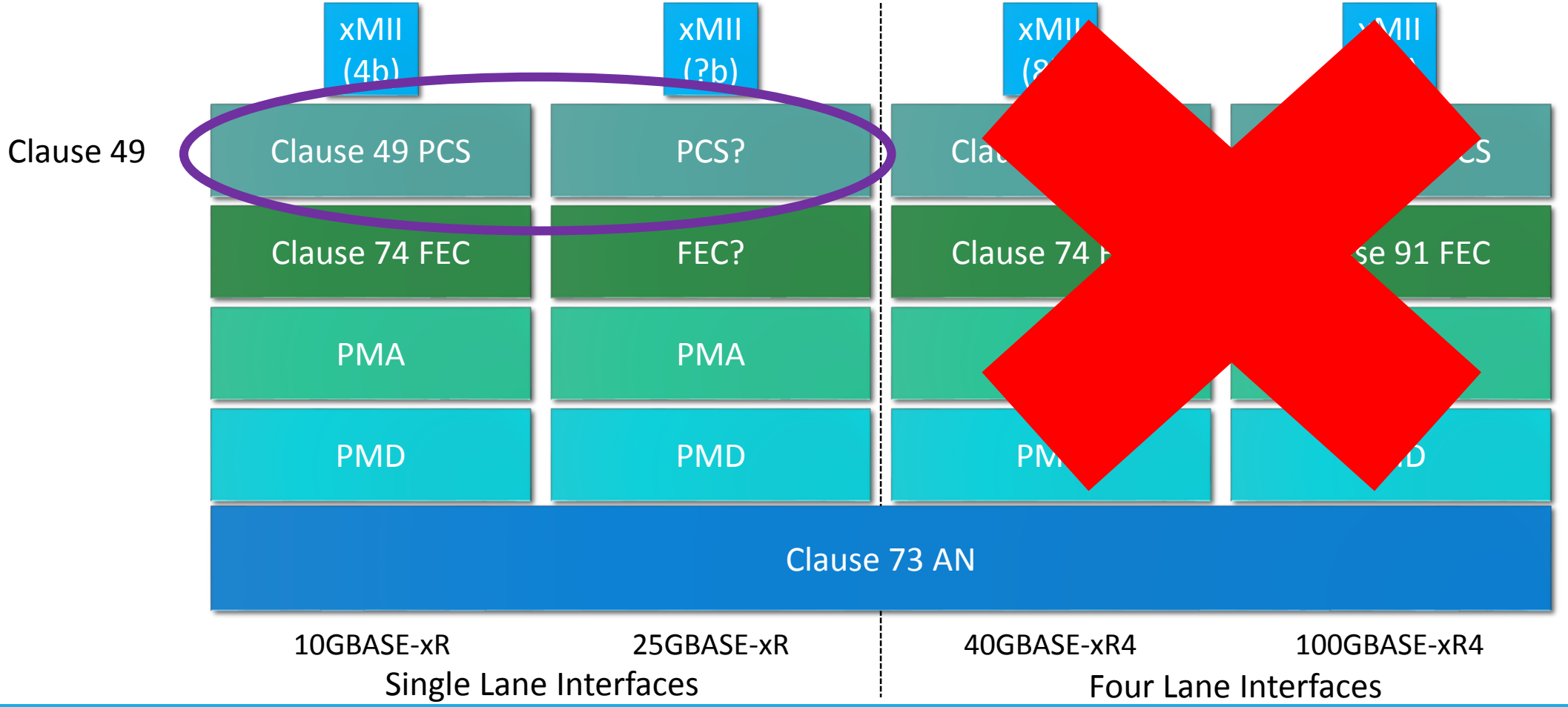
802.3bj-2014 100G Twinax & Backplane



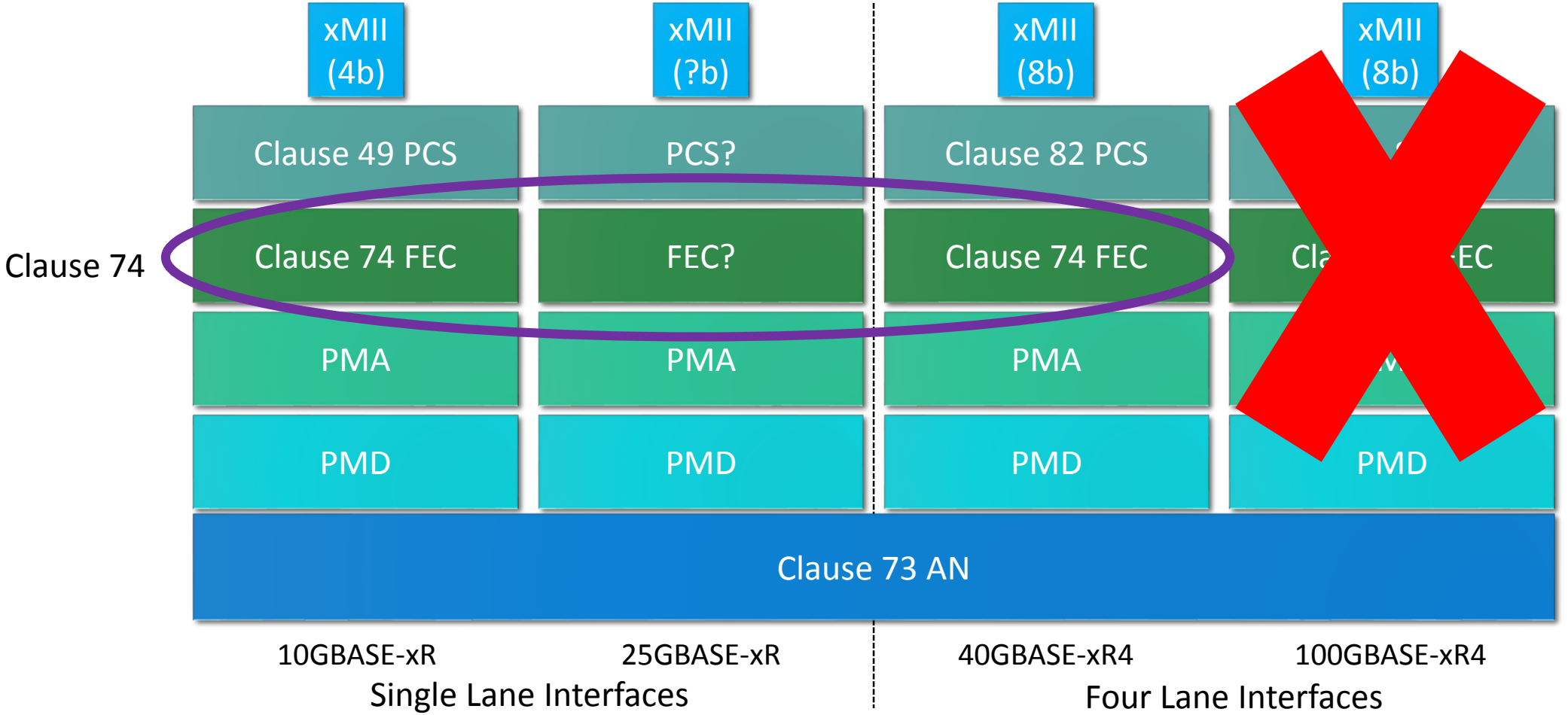
Architectural Overview – What FEC and PCS for 25G Copper?



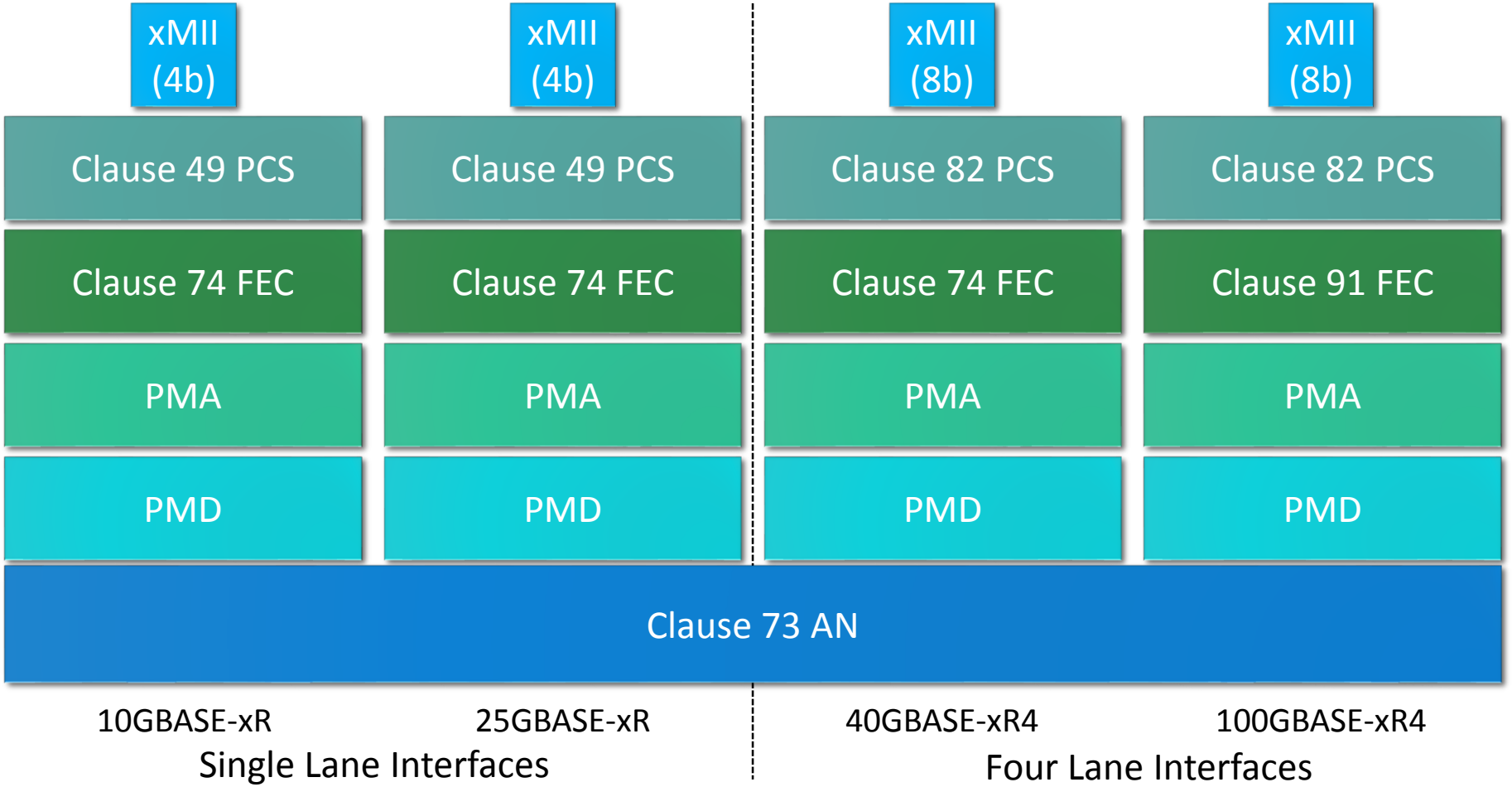
10G and 25G Network Interface Card



10G/25G/40G Network Interface Card



Resulting Block Diagram – Copper



Copper Observations

Auto-negotiation is a very helpful option

- If not used, the port will be managed to be a specific port type

10G/25G NIC

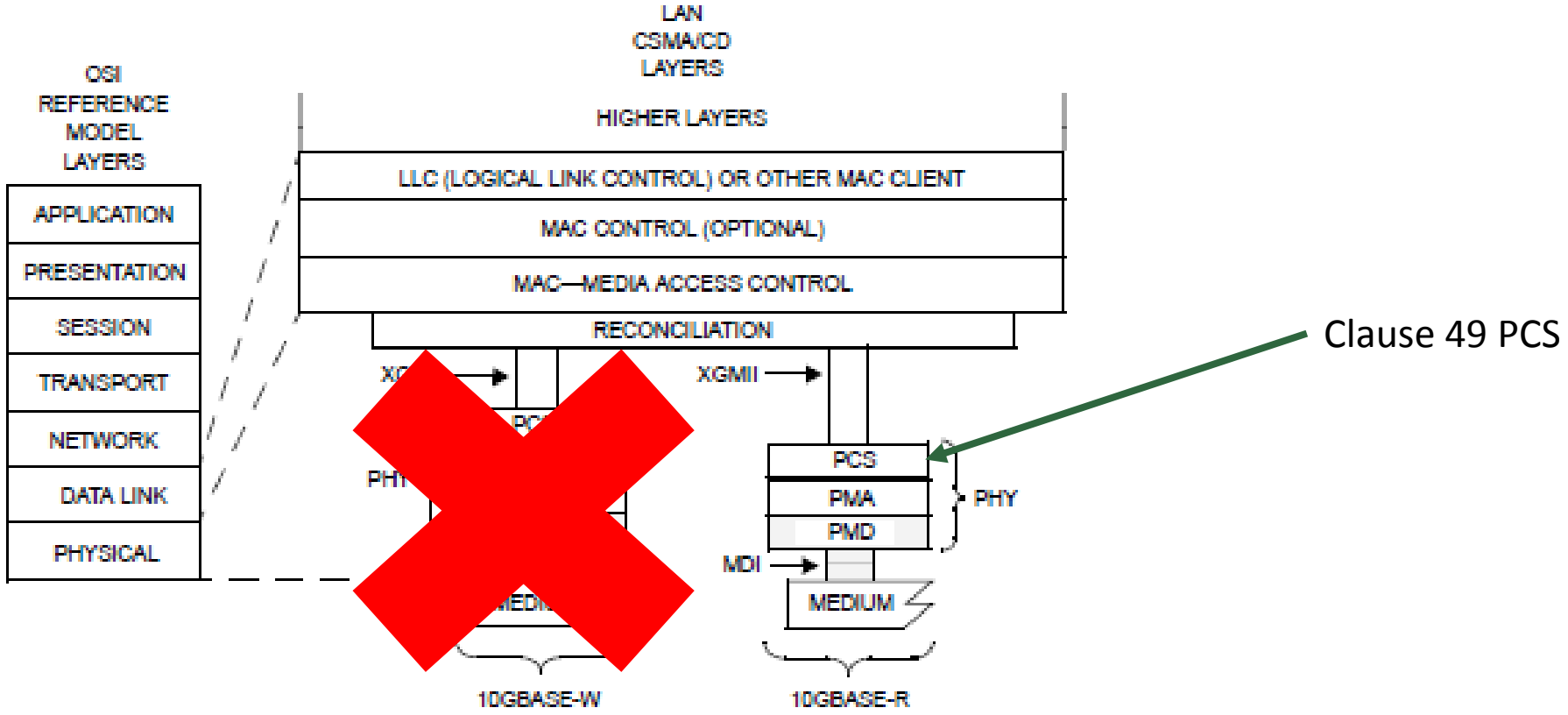
- Support of Clause 49 PCS and Clause 74 FEC would be the simplest from an implementation and silicon cost point of view
- MAC/RS/xMII would be closest between 10G and 25G

10G/25G/40G NIC

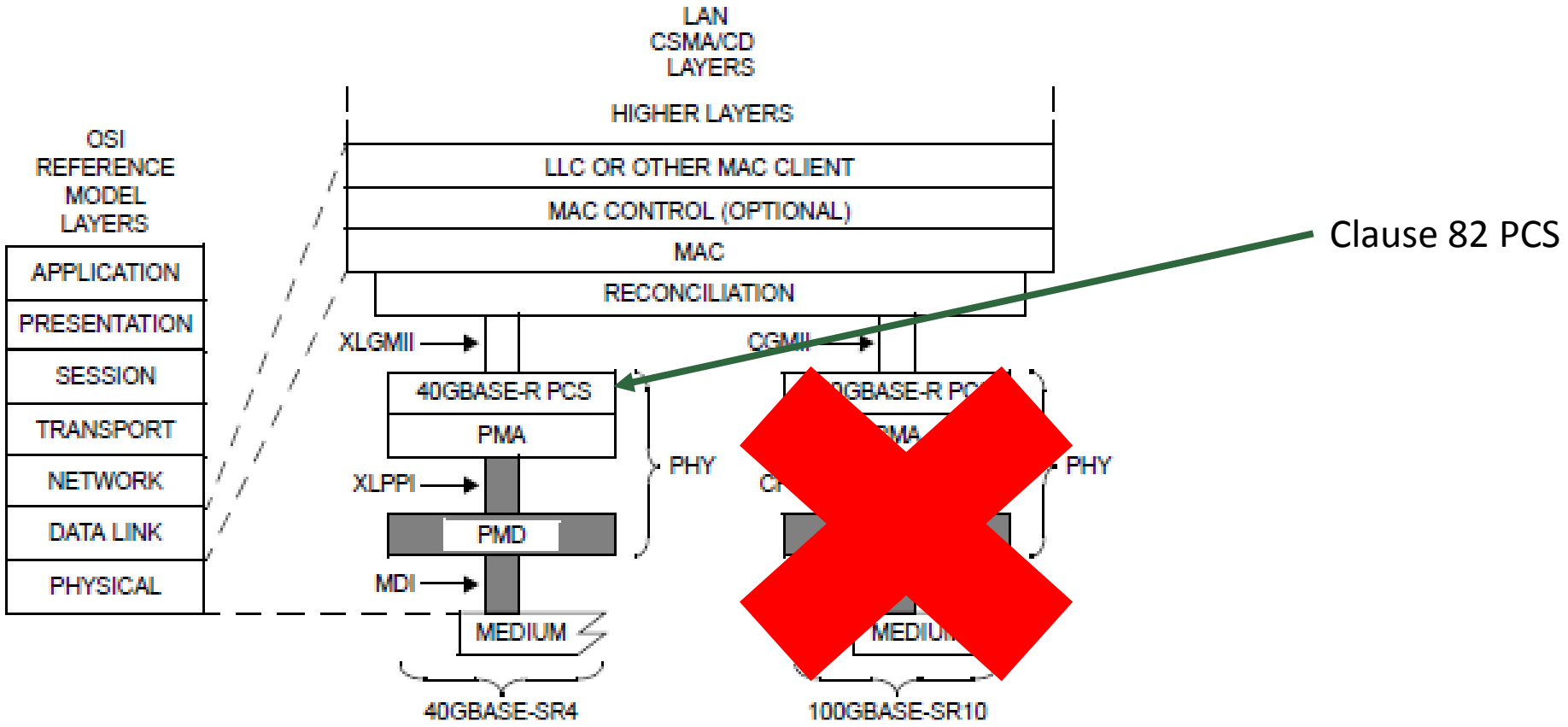
- Both Clause 49 and Clause 82 PCS would need to be supported
  - No different that existing 10G/40G implementations available today
- Clause 74 FEC (optional) could be shared across all port types
- Clause 91 FEC would be an addition to existing 10G/40G implementations

Optics

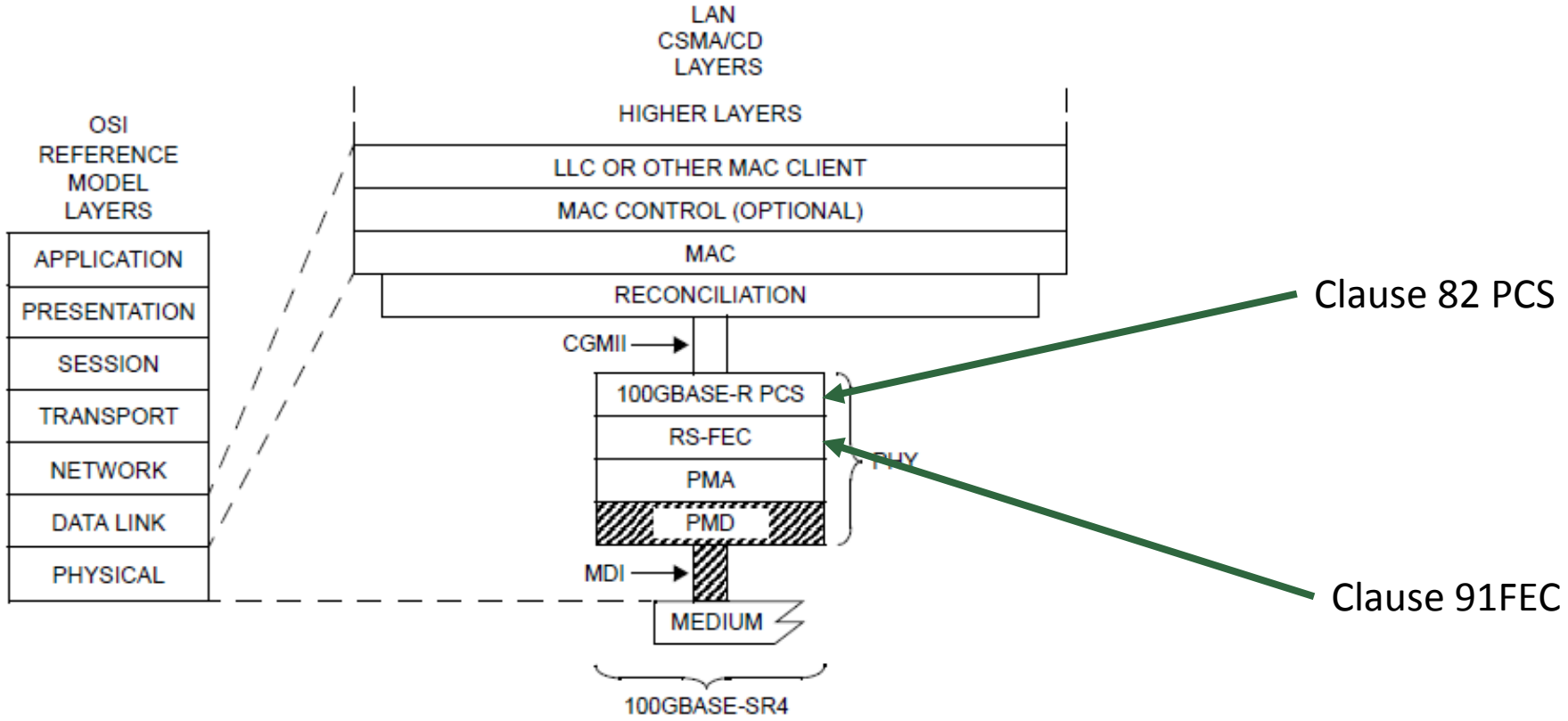
802.3-2012 10G SR Optics



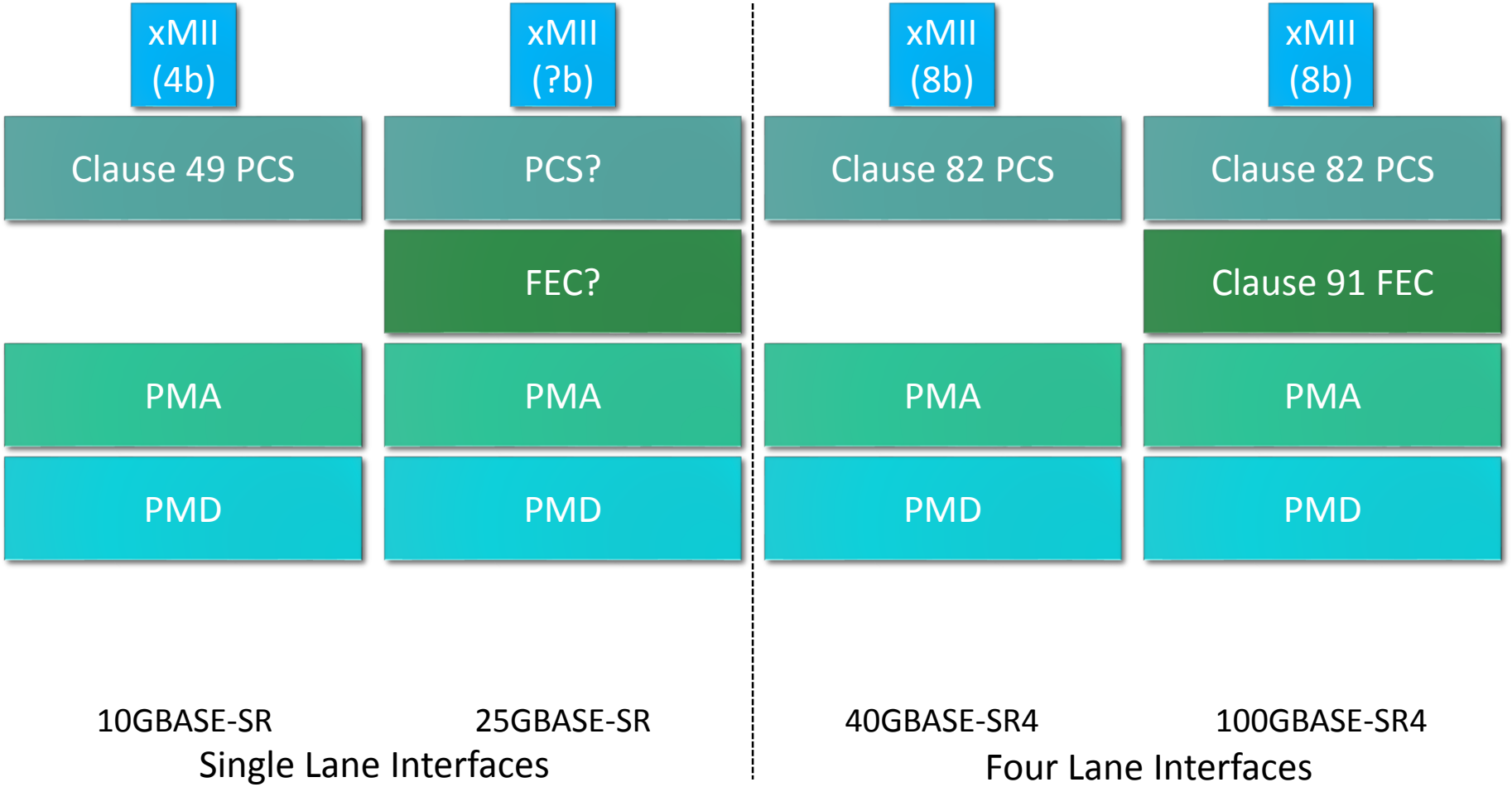
802.3-2012 40G SR4 Optics



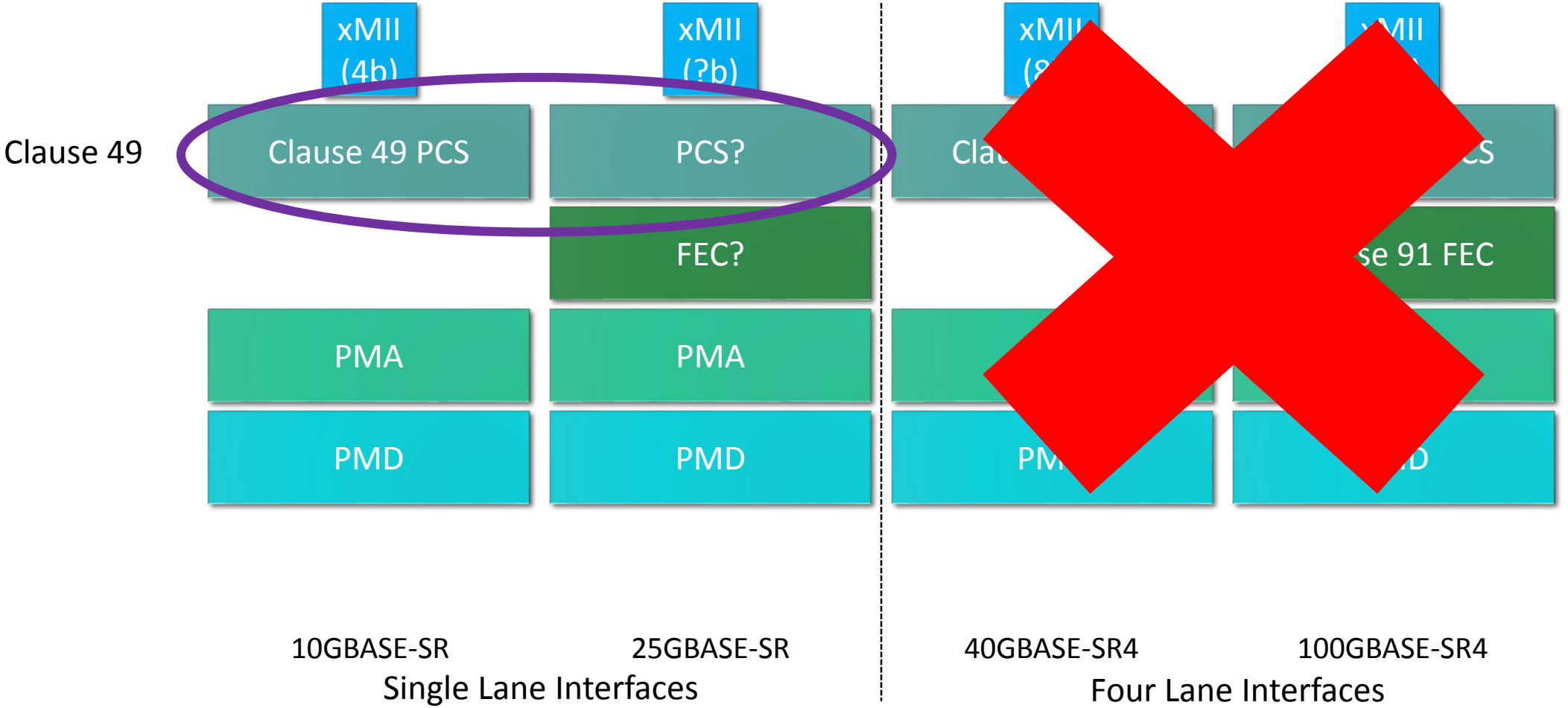
P802.3bm 100G SR4 Optics



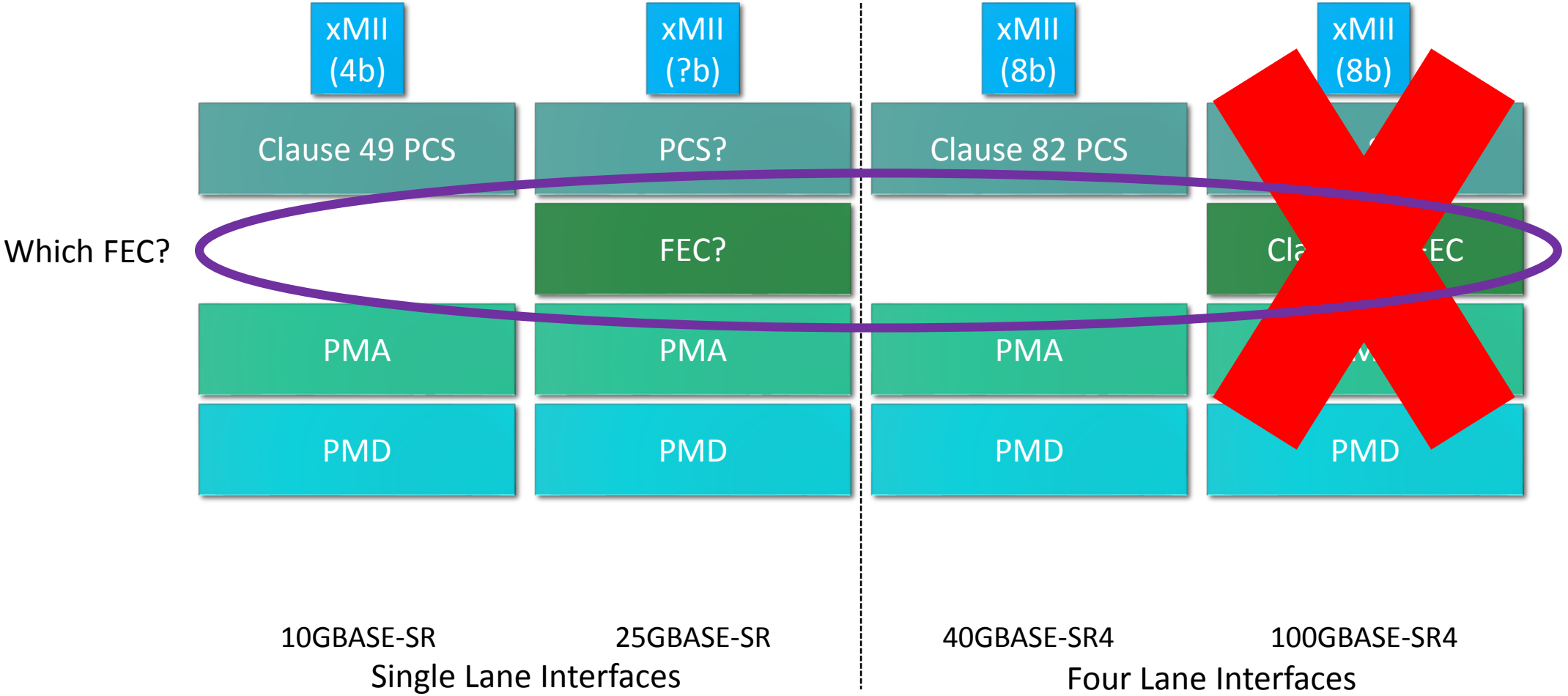
Architectural Overview – What FEC and PCS for 25G Optical?



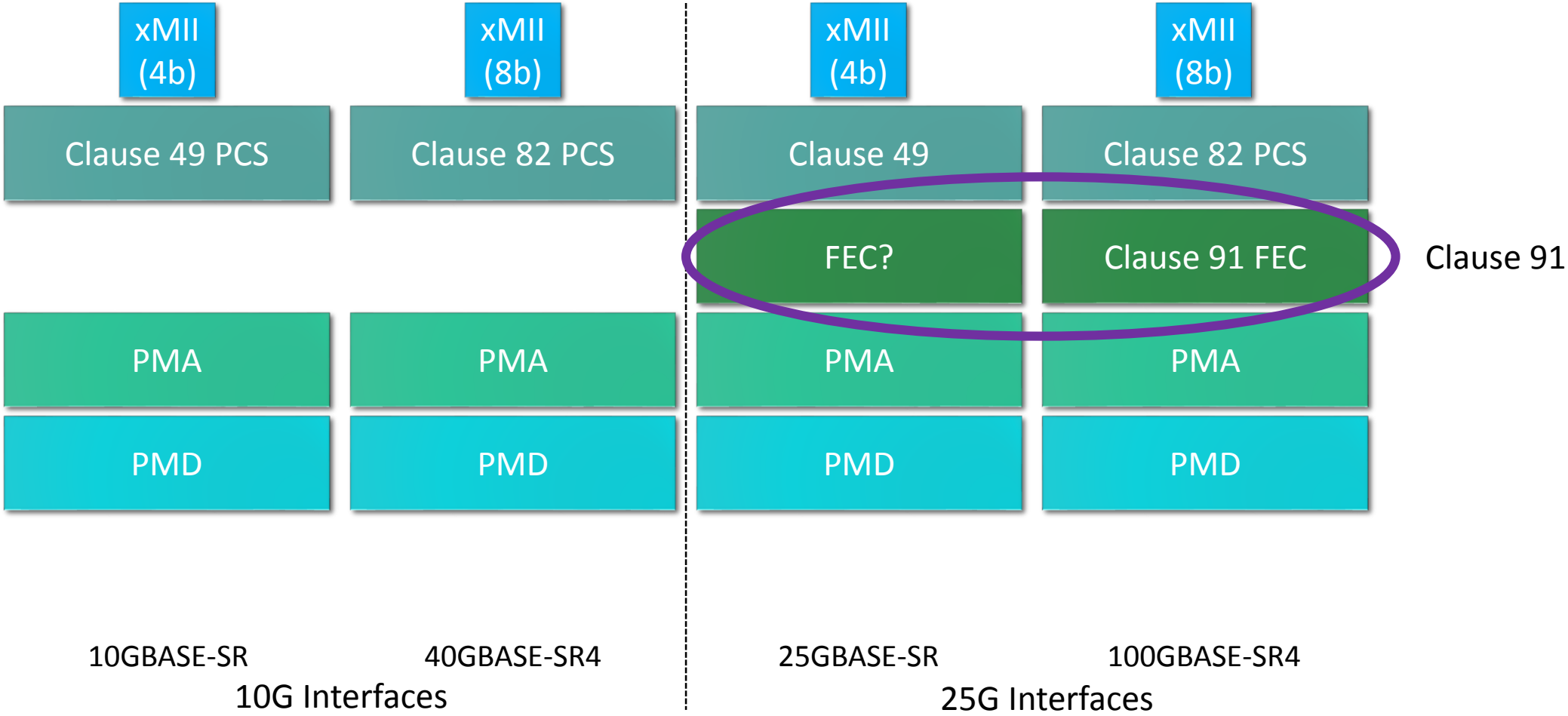
10G and 25G Network Interface Card



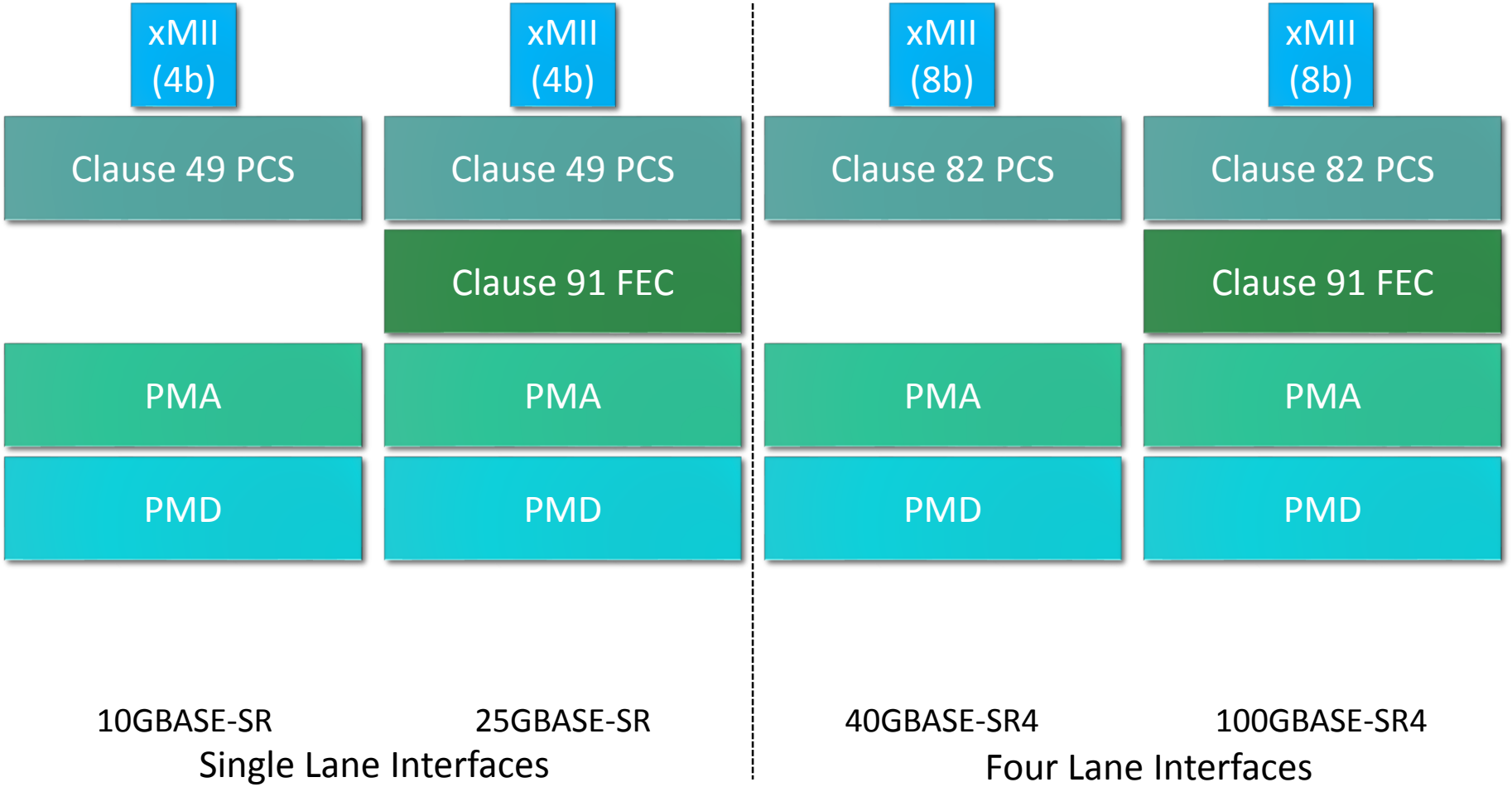
10G/25G/40G Network Interface Card



Change the View – Interface Rate



Resulting Block Diagram – Optical



Optical Observations

No auto-negotiation capability

10G/25G NIC

- Support of Clause 49 PCS would be the simplest from an implementation and silicon cost point of view
- Channel capabilities are different though
- MAC/RS/xMII would be closest between 10G and 25G

10G/25G/40G NIC

- Both Clause 49 and Clause 82 PCS would need to be supported
- No FEC exists for 10G or 40G optical
- How likely is 10G/25G/40G to be implemented due to difference in connector

Clause 91 FEC would offer compatibility to 100GBASE-SR4

- And to 100GBASE-KR4 and CR4!

OTN Thoughts

OTN not used between server and switch; only in switch to switch connections

MMF is used in the switch to OTN connection

- Copper is not used today, but cannot guarantee that moving forward
- Compatibility between copper and MMF bit stream is helpful

Link into and out of OTN needs to operate in same manner as though the OTN doesn't exist

- Treat the OTN as only a media translator

Assume that any connection to OTN must be based upon Clause 49 PCS and Clause 91 FEC

General Observations

PCS selection

- Clause 49 as the single-lane
- Clause 82 as the multi-lane
- Easiest to support Clause 49 in a 10G/25G NIC implementation

FEC selection

- Analyzing from a SerDes interface point of view
- FEC choice would best match that of the similar data rate
- RS-FEC may be a requirement for the 5m channel or MMF
- Clause 74 for 10G SerDes and RS-FEC for 25G SerDes
- Auto-negotiation would permit selection of the FEC to be done prior to link establishment

Recommendations

Select Clause 49 as the PCS for 25G

- Single-lane 10G/25G NICs are likely to be lion share
- Clause 82 is a great multi-lane PCS, but has greatest impact to re-use of 10G implementations

Use auto-negotiation to select the FEC for twinax and backplane

- For copper NICs, FEC could be:
  - Off
  - Clause 74 FEC (optional)
  - Clause 91 FEC (mandatory?)

For optical implementations, FEC would be based on Clause 91

For OTN connectivity, FEC should be based on Clause 91

- Per above, this would work with both MMF and copper interconnect

Thank You!