Logic Architecture Baseline

IEEE P802.3df Task Force
IEEE 802.3 May 2022

Mark Gustlin – Cisco
John D’Ambrosia – Futurewei, U.S. Subsidiary of Huawei
Gary Nicholl – Cisco
Xinyuan Wang – Huawei
Matt Brown - Huawei
David Ofelt – Juniper
Jeff Slavick – Broadcom
Kapil Shrikhande – Marvell

Supporters

- Rob Stone – Meta
- Ali Ghiasi – Ghiasi Quantum
- Mike Dudek - Marvell
- Shawn Nicholl – AMD
- Phil Sun – Credo
- Weiqiang Cheng, China Mobile
- Haojie Wang, China Mobile
- Xiang He, Huawei
- Eric Maniloff – Ciena
- Jerry Pepper – Keysight
- Eugene Opsasnick – Broadcom
- Kent Lusted – Intel
- Arthur Marris – Cadence
- Mike Peng Li - Intel

Architecture Requirements

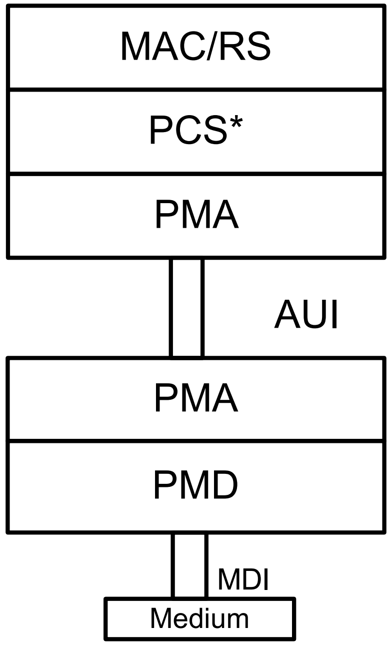
- Support Physical layer specifications for all Ethernet rates (200G, 400G, 800G, 1.6T)
  - A common architecture across all rates is a good to have
  - Commonality with previous 200GbE/400GbE architecture is a nice to have
- Support multiple FEC schemes
  - End to end, segmented, concatenated, or a mix; as determined by our PMD/AUI needs
- Support up to 2xAUIs per link side
  - AUIs are lane rate independent
    - As an example: one AUI at 200G/lane and another at 100G/lane for instance
    - Lane rates can vary on one side of the link, both sides etc.
- Support the FEC capabilities as needed for 100G/lane, 200G/lane and beyond
- Flexible enough to deal with the unknowns at this point in the project

Current 200GbE/400GbE Architecture

- PCS and FEC are integrated into a single sublayer
- We should reuse this architecture unless there is a good reason not to
- See clause 116 for details

802.3bs architecture

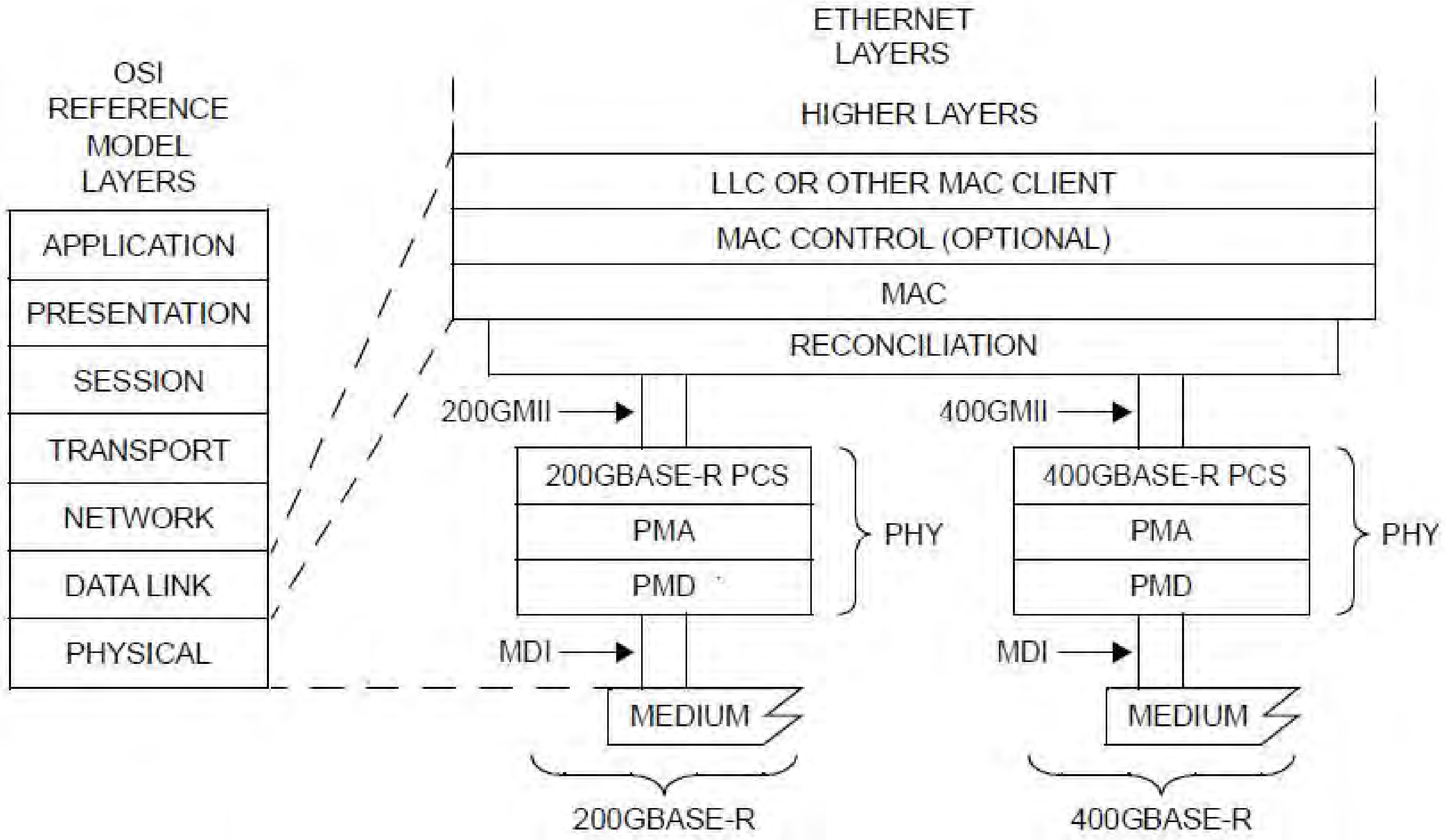
End to end FEC (for 100G/lane)
Corrects errors accumulated across all AUIs and the PMD



*FEC above the AUI is part of the PCS sublayer

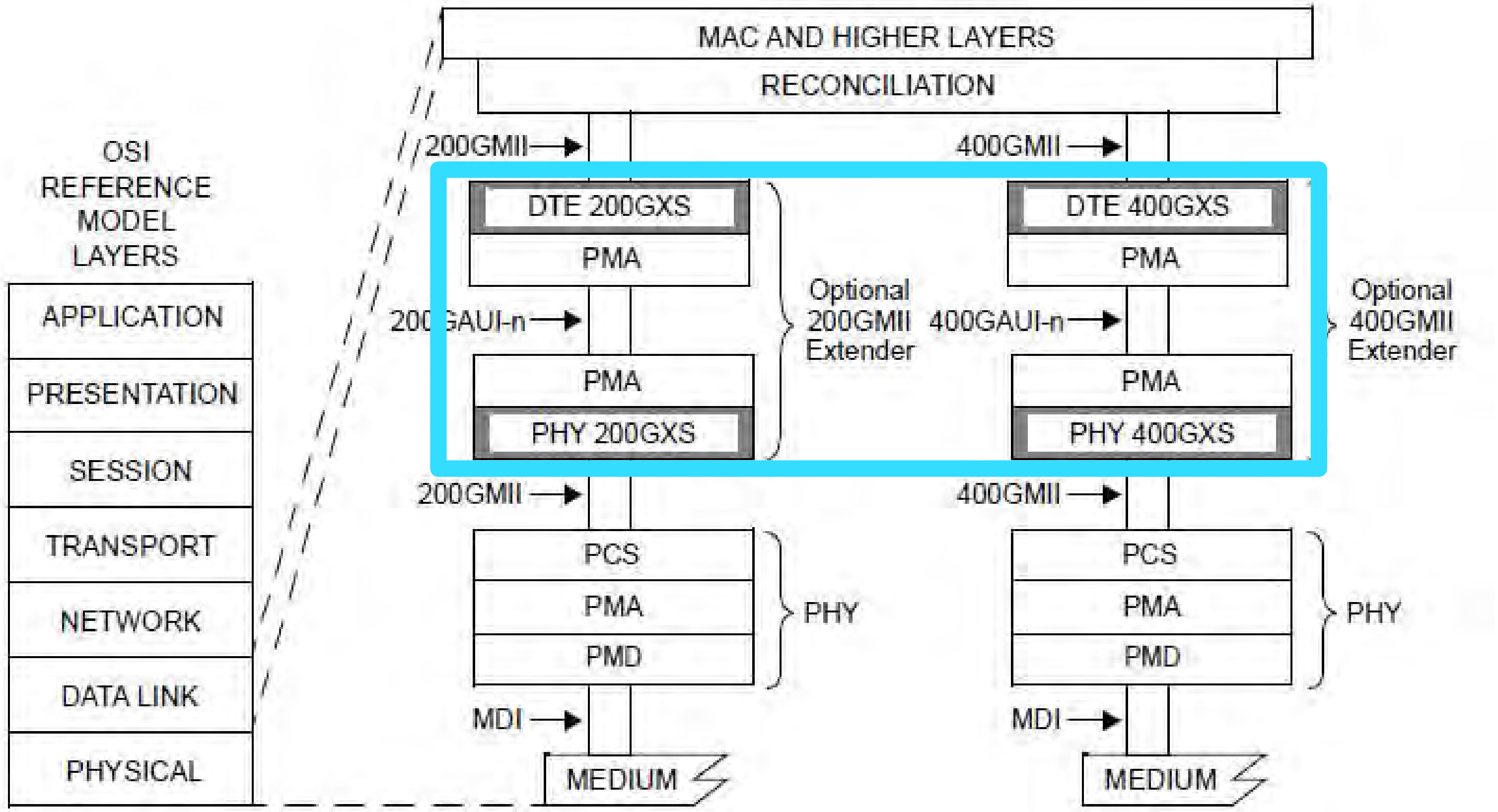
Full 200GbE/400GbE Architecture

- Architecture in the larger OSI context (from the base standard)



200GbE/400GbE MII Extender

- An extender sublayer is defined to allow us extend the MII across an AUI
- This is used for a 400GBASE-ZR interface, for instance
- Implementations don't necessarily go back to the MII interface
- See clause 118 for details



Note: Physical layer ends at MDI

Proposed 200GbE/400GbE Architecture

FEC1 = End to End FEC

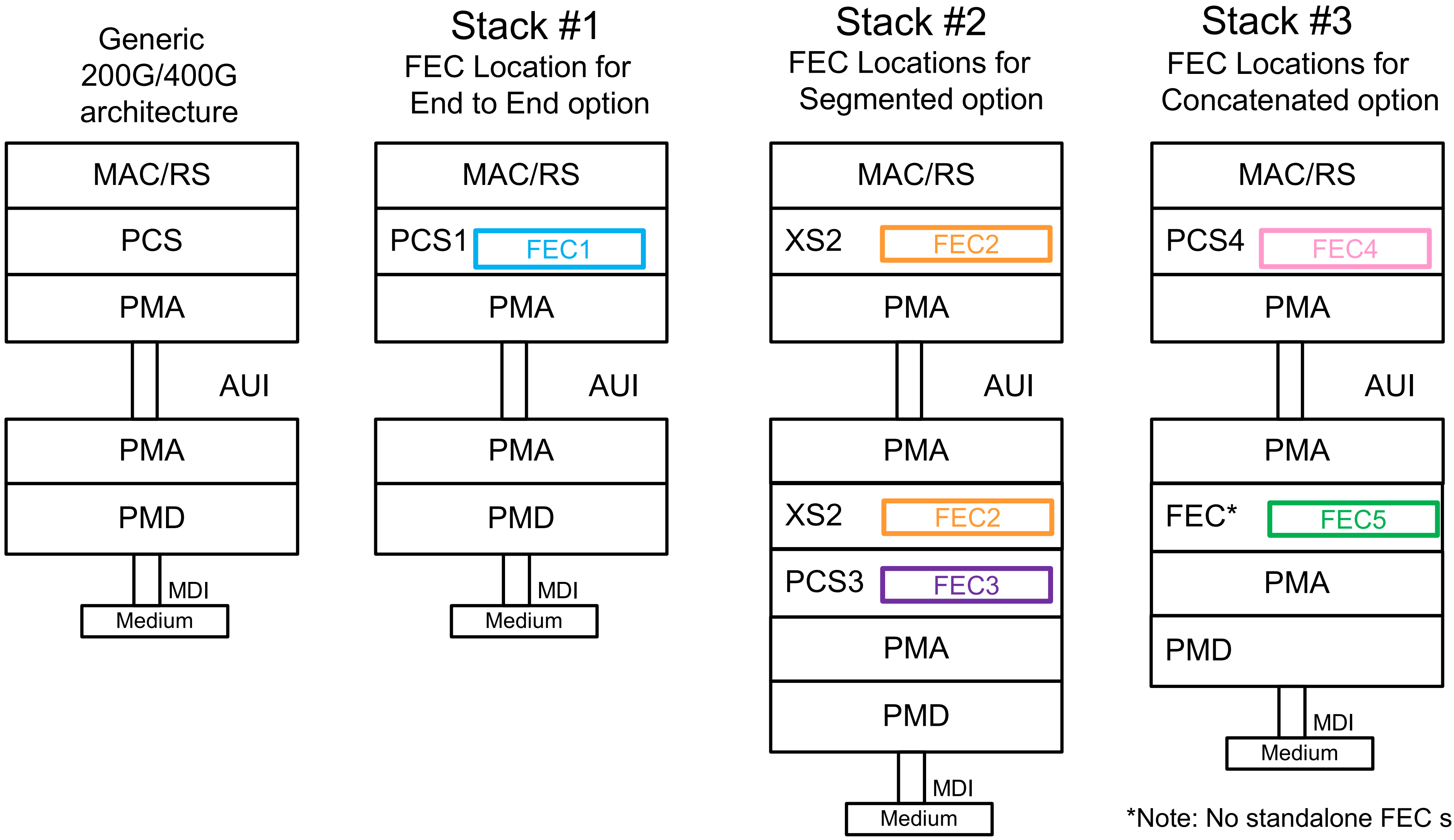
FEC2 = AUI FEC for Segmented

FEC3 = PMD FEC for Segmented

FEC4 = Outer FEC for Concatenated

FEC5 = Inner FEC for Concatenated

- How various FEC schemes fit into the architecture
- FECs might or might not be reused across schemes



*Note: No standalone FEC sublayer in 802.3bs

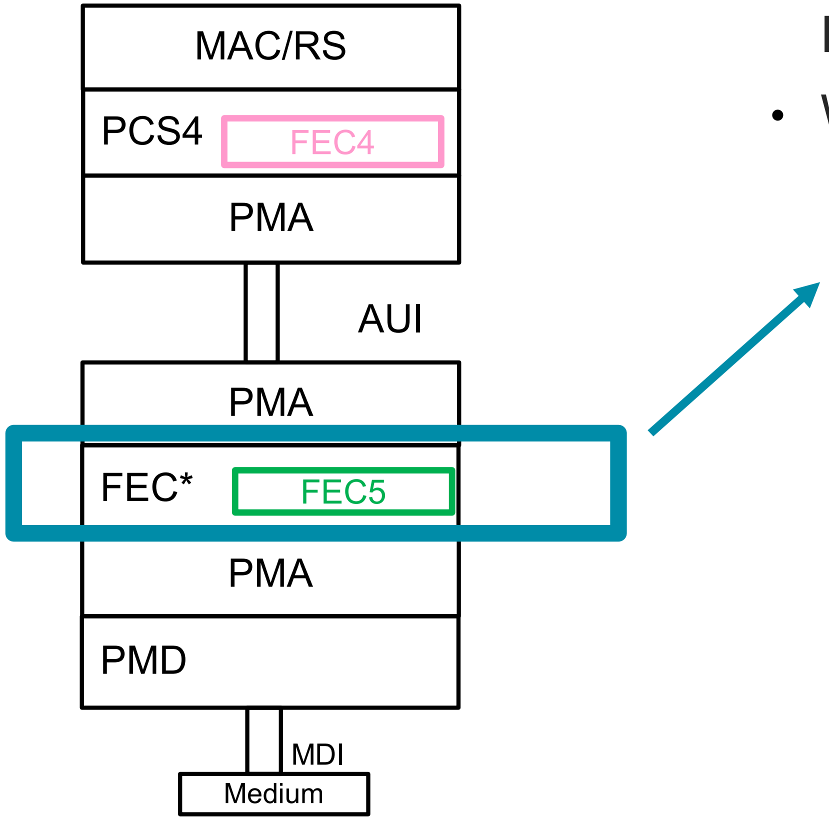
More on the FEC Sublayer

FEC4 = Outer FEC for Concatenated

FEC5 = Inner FEC for Concatenated

Stack #3

FEC Locations for Concatenated option


MAC/RS

PCS4 FEC4

PMA

AUI

PMA

FEC* FEC5

PMA

PMD

MDI

Medium

- When/if we do concatenation, we have this new sublayer labeled FEC
- What is it?
  - We would need to fully define the functionality; it was not in the original 802.3bs architecture
  - On Tx, it would take the data stream and wrap the inner FEC around it
  - And do whatever else is needed, permute the data stream as needed
  - On the RX side it would remove the FEC, correct, un-permute the data etc.

Proposed Direction for 800GbE/1.6TbE

- Keep the 800GbE/1.6TbE architecture consistent with the 802.3bs architecture
- No reason to change it, we must support 200G/lanes with the current 802.3bs architecture anyhow
- There are some benefits to keeping the high-level logic architecture consistent between the 4 Ethernet rates within this project (200G/400G/800G/1.6T)

Proposed 800GbE/1.6TbE Architecture

FEC1 = End to End FEC

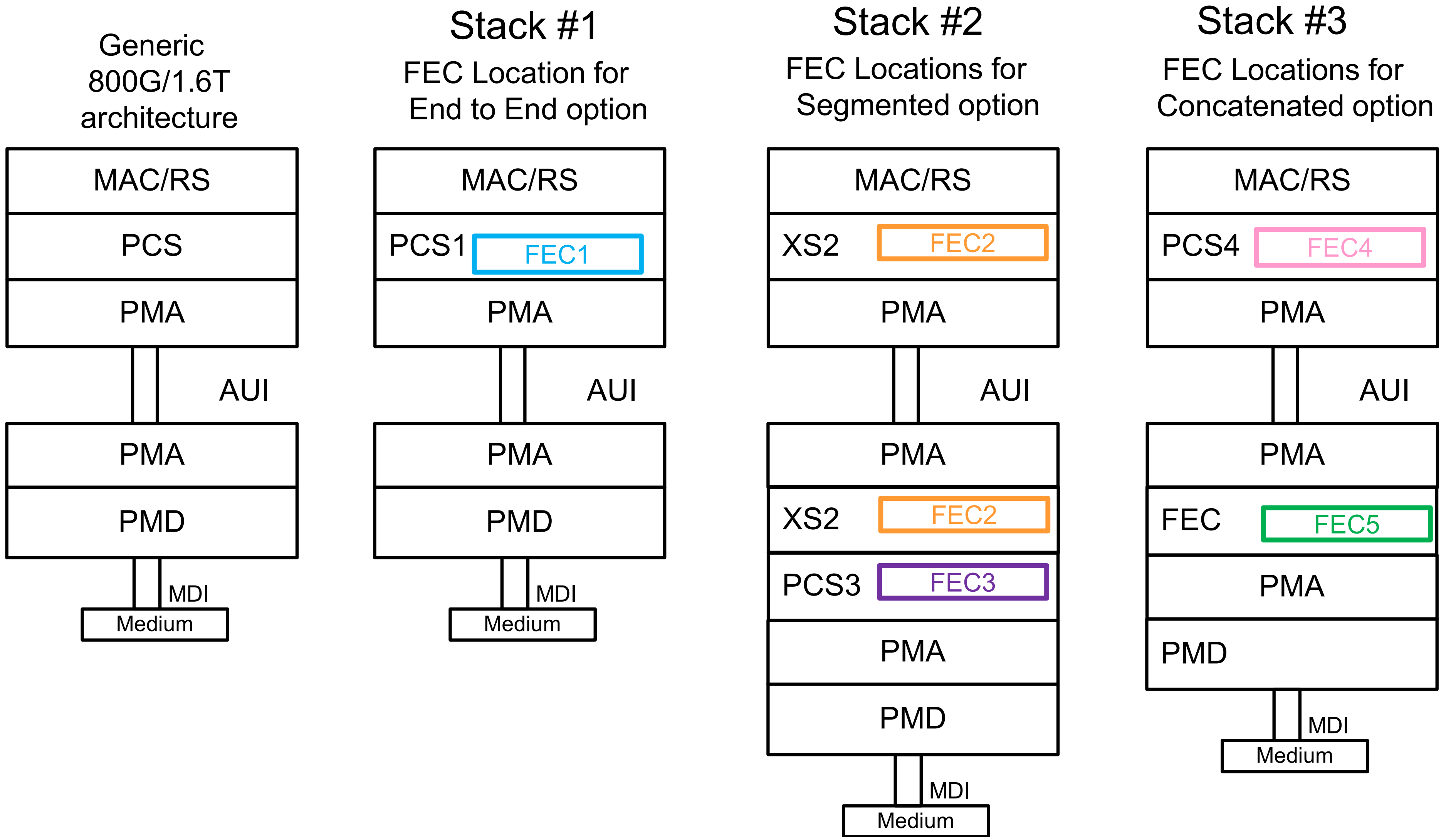
FEC2 = AUI FEC for Segmented

FEC3 = PMD FEC for Segmented

FEC4 = Outer FEC for Concatenated

FEC5 = Inner FEC for Concatenated

- How various FEC schemes fit into the architecture
- FECs might or might not be reused across schemes



Generic 800G/1.6T architecture

MAC/RS
PCS
PMA
AUI
PMA
PMD
MDI
Medium

Stack #1

FEC Location for End to End option

MAC/RS
PCS1 FEC1
PMA
AUI
PMA
PMD
MDI
Medium

Stack #2

FEC Locations for Segmented option

MAC/RS
XS2 FEC2
PMA
AUI
PMA
XS2 FEC2
PCS3 FEC3
PMA
PMD
MDI
Medium

Stack #3

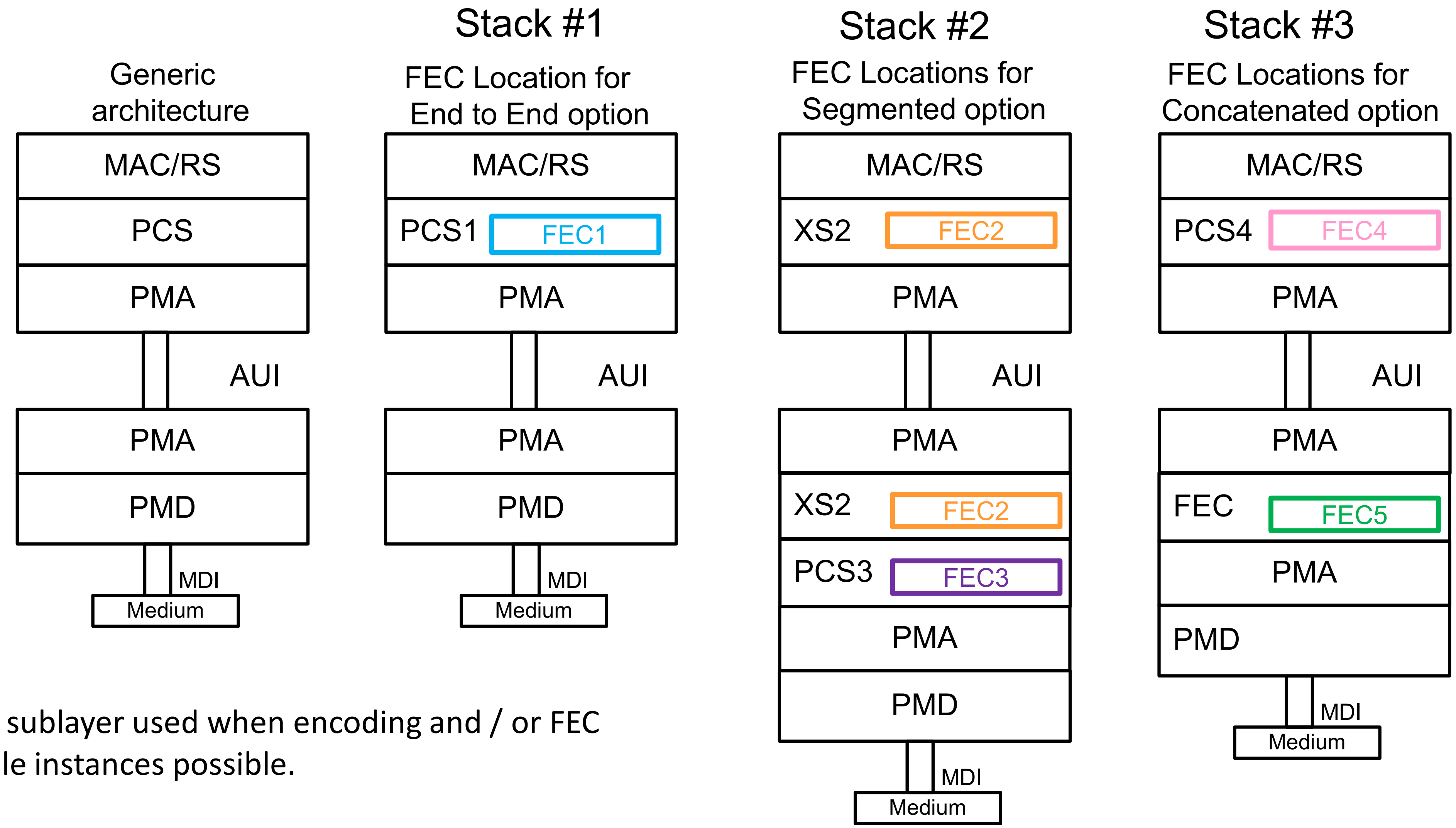
FEC Locations for Concatenated option

MAC/RS
PCS4 FEC4
PMA
AUI
PMA
FEC FEC5
PMA
PMD
MDI
Medium

Proposed 802.3df Overall Architecture

- For all Ethernet rates within this project (200G/400G/800G/1.6T)
- FECs might or might not be reused across schemes
- TBD which FEC scheme(s) are needed for this project

FEC1 = End to End FEC
FEC2 = AUI FEC for Segmented
FEC3 = PMD FEC for Segmented
FEC4 = Outer FEC for Concatenated
FEC5 = Inner FEC for Concatenated



Generic architecture
MAC/RS
PCS
PMA
AUI
PMA
PMD
MDI
Medium

Stack #1
FEC Location for End to End option
MAC/RS
PCS1 FEC1
PMA
AUI
PMA
PMD
MDI
Medium

Stack #2
FEC Locations for Segmented option
MAC/RS
XS2 FEC2
PMA
AUI
PMA
XS2 FEC2
PCS3 FEC3
PMA
PMD
MDI
Medium

Stack #3
FEC Locations for Concatenated option
MAC/RS
PCS4 FEC4
PMA
AUI
PMA
FEC FEC5
PMA
PMD
MDI
Medium

Note – Extender sublayer used when encoding and / or FEC changes. Multiple instances possible.

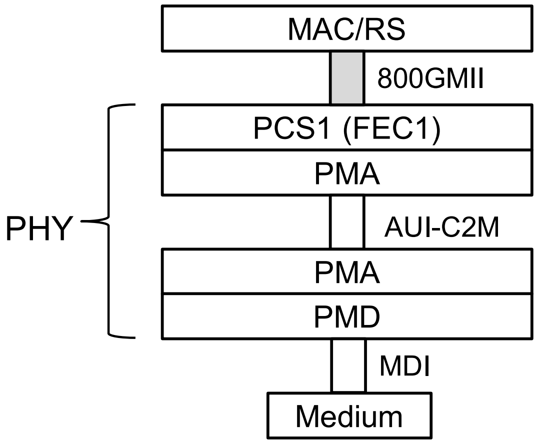
A Note about the Following Slides

- The following slide 13-18 are examples to illustrate the flexibility of the architecture
- 800GbE is used in the examples, but architectural aspects would apply to any of our rates
- The intent of the following slides is not to recommend one FEC scheme over another or to propose a FEC baseline

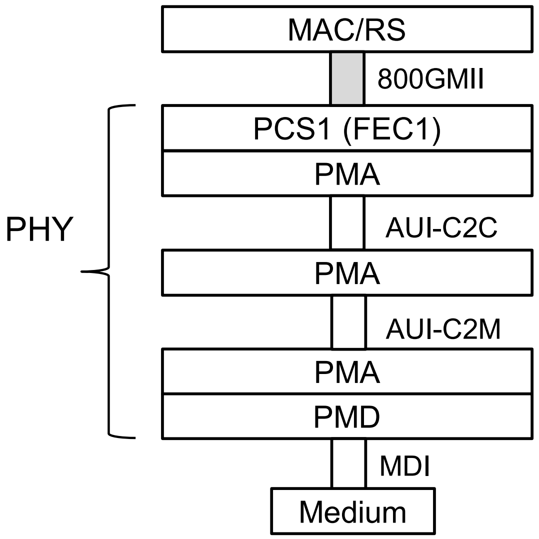
Architecture Examples - End to End FEC

- Examples show end to end FEC schemes

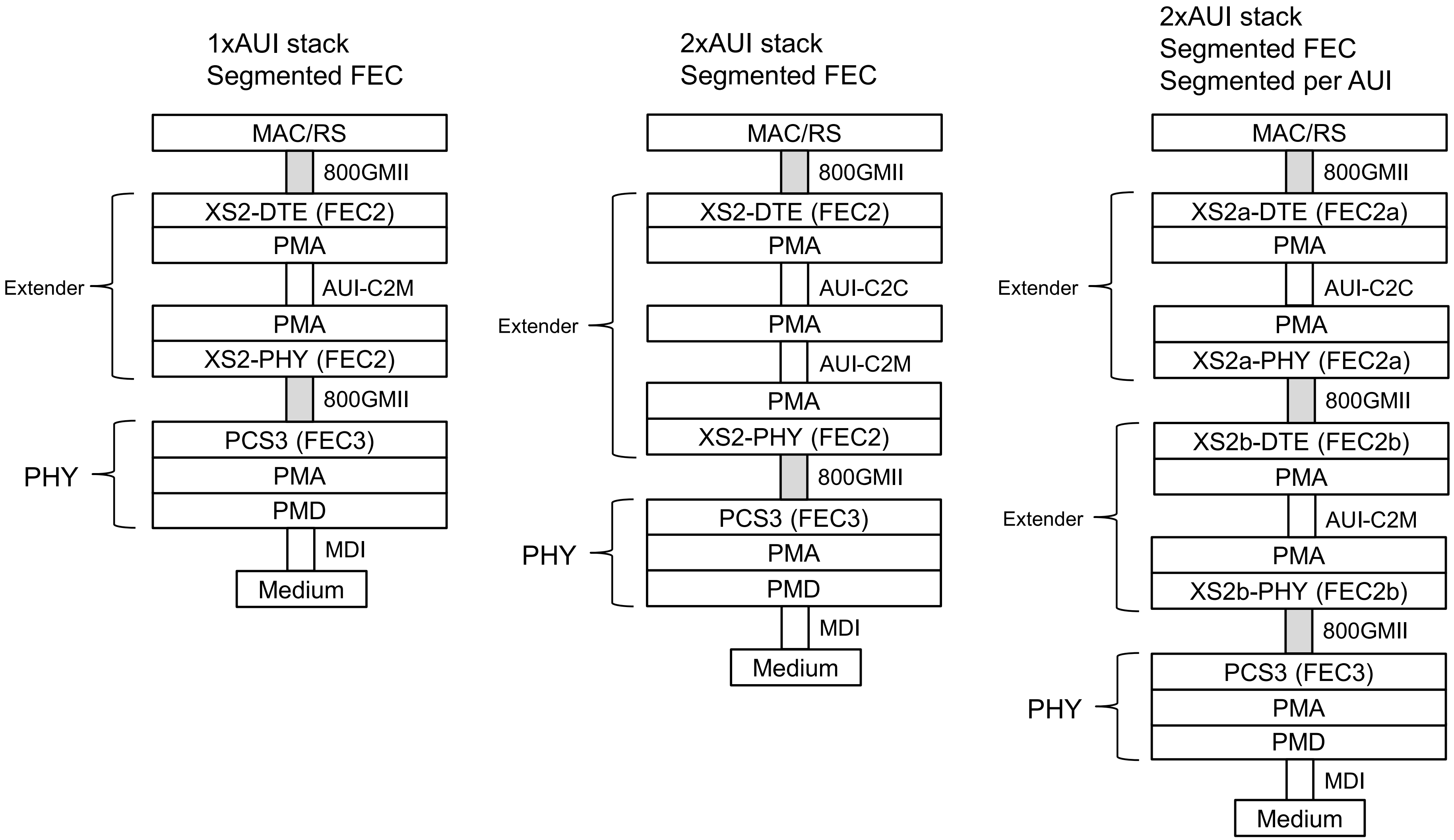
Simple 1xAUI stack



Simple 2xAUI stack



Architecture Examples – Segmented FEC



1xAUI stack
Segmented FEC

MAC/RS
800GMII
Extender
XS2-DTE (FEC2)
PMA
AUI-C2M
PMA
XS2-PHY (FEC2)
800GMII
PHY
PCS3 (FEC3)
PMA
PMD
MDI
Medium

2xAUI stack
Segmented FEC

MAC/RS
800GMII
Extender
XS2-DTE (FEC2)
PMA
AUI-C2C
PMA
AUI-C2M
PMA
XS2-PHY (FEC2)
800GMII
PHY
PCS3 (FEC3)
PMA
PMD
MDI
Medium

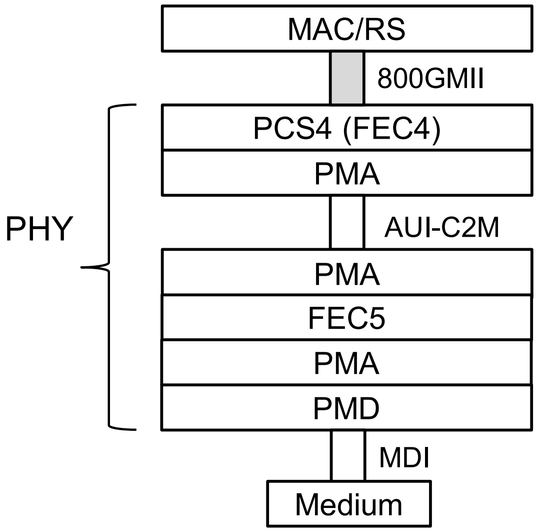
2xAUI stack
Segmented FEC
Segmented per AUI

MAC/RS
800GMII
Extender
XS2a-DTE (FEC2a)
PMA
AUI-C2C
PMA
XS2a-PHY (FEC2a)
800GMII
Extender
XS2b-DTE (FEC2b)
PMA
AUI-C2M
PMA
XS2b-PHY (FEC2b)
800GMII
PHY
PCS3 (FEC3)
PMA
PMD
MDI
Medium

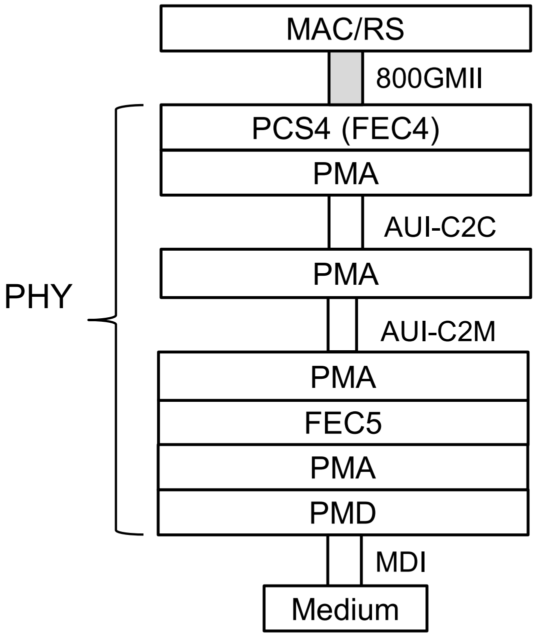
Note: No current examples of two XS in 802.3bs

Architecture Examples – Concatenated FEC

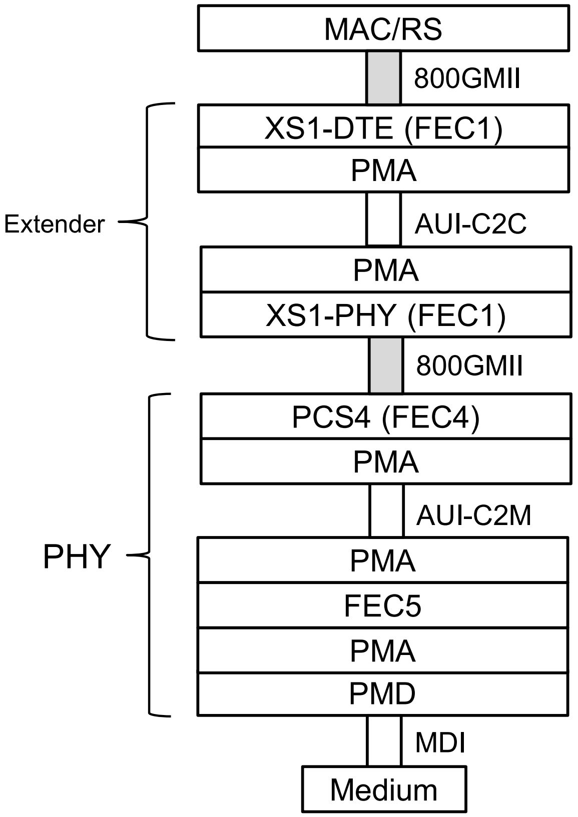
Simple 1xAUI stack



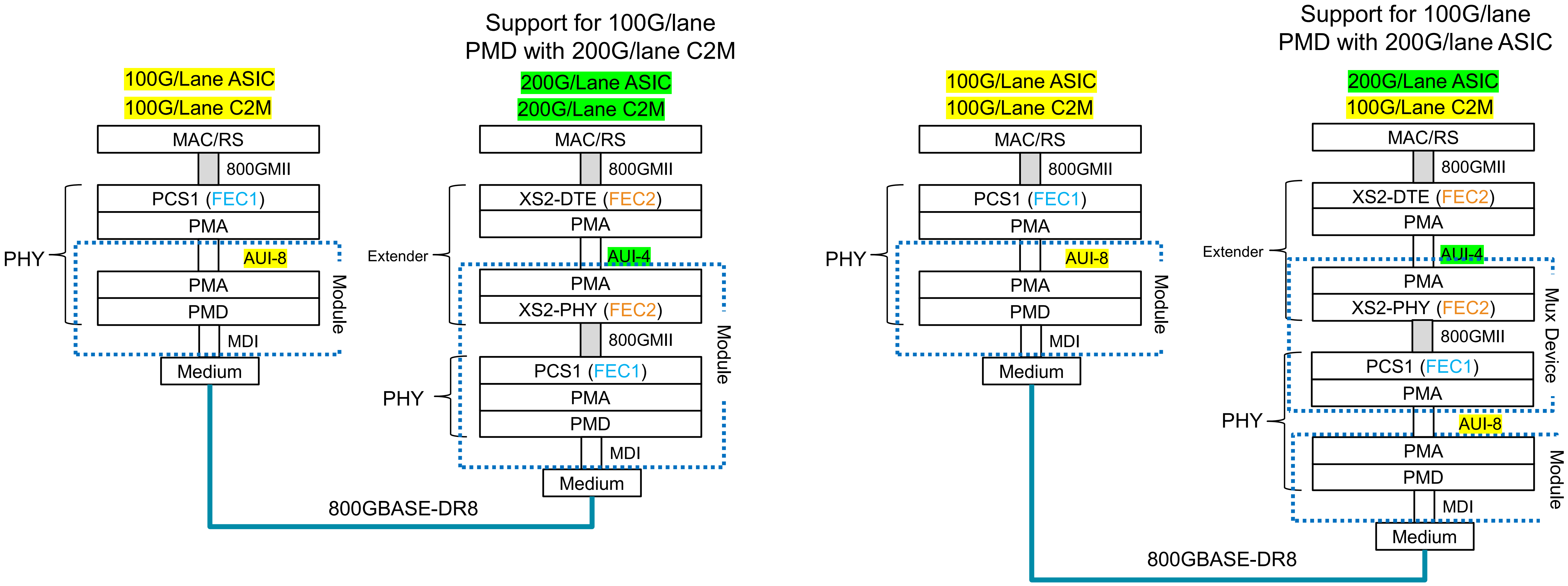
Simple 2xAUI stack



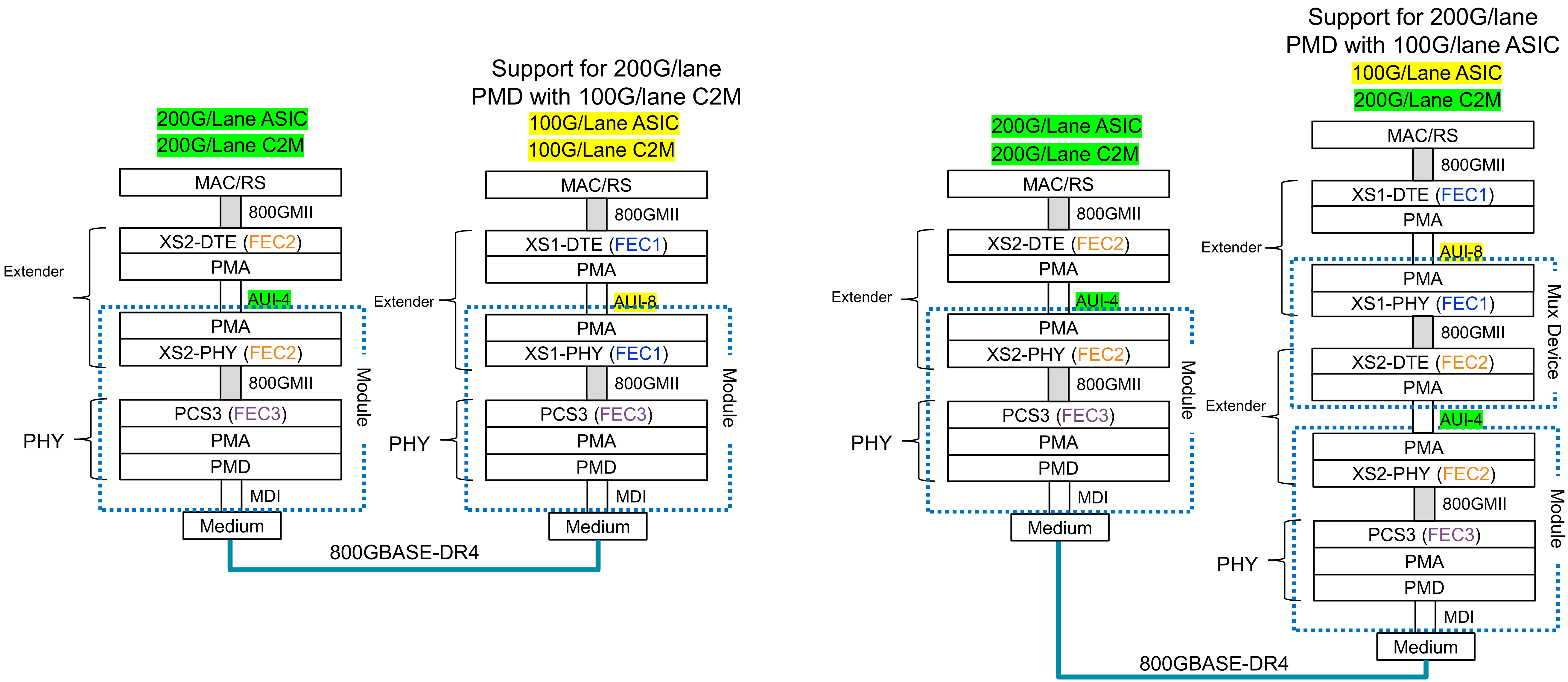
2xAUI stack
Segmented + concatenation of FECs
Unique FEC per AUI



Connectivity Examples – End to End FEC across the PHY



Connectivity Examples – Segmented FEC



200G/Lane ASIC
200G/Lane C2M
MAC/RS
800GMII
Extender
XS2-DTE (FEC2)
PMA
AUI-4
PMA
XS2-PHY (FEC2)
800GMII
Module
PHY
PCS3 (FEC3)
PMA
PMD
MDI
Medium

Support for 200G/lane PMD with 100G/lane C2M
100G/Lane ASIC
100G/Lane C2M
MAC/RS
800GMII
Extender
XS1-DTE (FEC1)
PMA
AUI-8
PMA
XS1-PHY (FEC1)
800GMII
Module
PHY
PCS3 (FEC3)
PMA
PMD
MDI
Medium

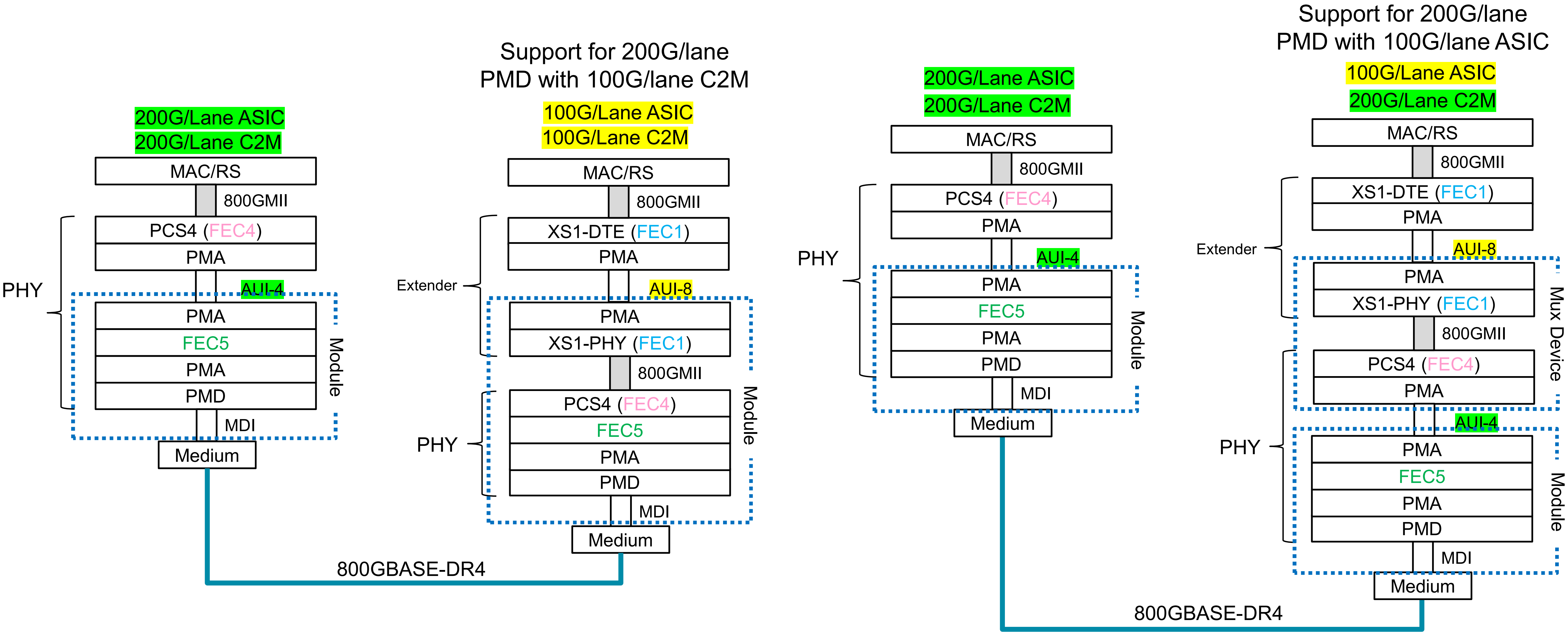
800GBASE-DR4

200G/Lane ASIC
200G/Lane C2M
MAC/RS
800GMII
Extender
XS2-DTE (FEC2)
PMA
AUI-4
PMA
XS2-PHY (FEC2)
800GMII
Module
PHY
PCS3 (FEC3)
PMA
PMD
MDI
Medium

Support for 200G/lane PMD with 100G/lane ASIC
100G/Lane ASIC
200G/Lane C2M
MAC/RS
800GMII
Extender
XS1-DTE (FEC1)
PMA
AUI-8
PMA
XS1-PHY (FEC1)
800GMII
Mux Device
Extender
XS2-DTE (FEC2)
PMA
AUI-4
PMA
XS2-PHY (FEC2)
800GMII
Module
PHY
PCS3 (FEC3)
PMA
PMD
MDI
Medium

800GBASE-DR4

Connectivity Examples – Concatenated FEC across the PHY



200G/Lane ASIC
200G/Lane C2M
MAC/RS
800GMII
PCS4 (FEC4)
PMA
AUI-4
PMA
FEC5
PMA
PMD
MDI
Medium
PHY
Module

Support for 200G/lane PMD with 100G/lane C2M
100G/Lane ASIC
100G/Lane C2M
MAC/RS
800GMII
XS1-DTE (FEC1)
PMA
AUI-8
PMA
XS1-PHY (FEC1)
800GMII
PCS4 (FEC4)
FEC5
PMA
PMD
MDI
Medium
Extender
PHY
Module

800GBASE-DR4

200G/Lane ASIC
200G/Lane C2M
MAC/RS
800GMII
PCS4 (FEC4)
PMA
AUI-4
PMA
FEC5
PMA
PMD
MDI
Medium
PHY
Module

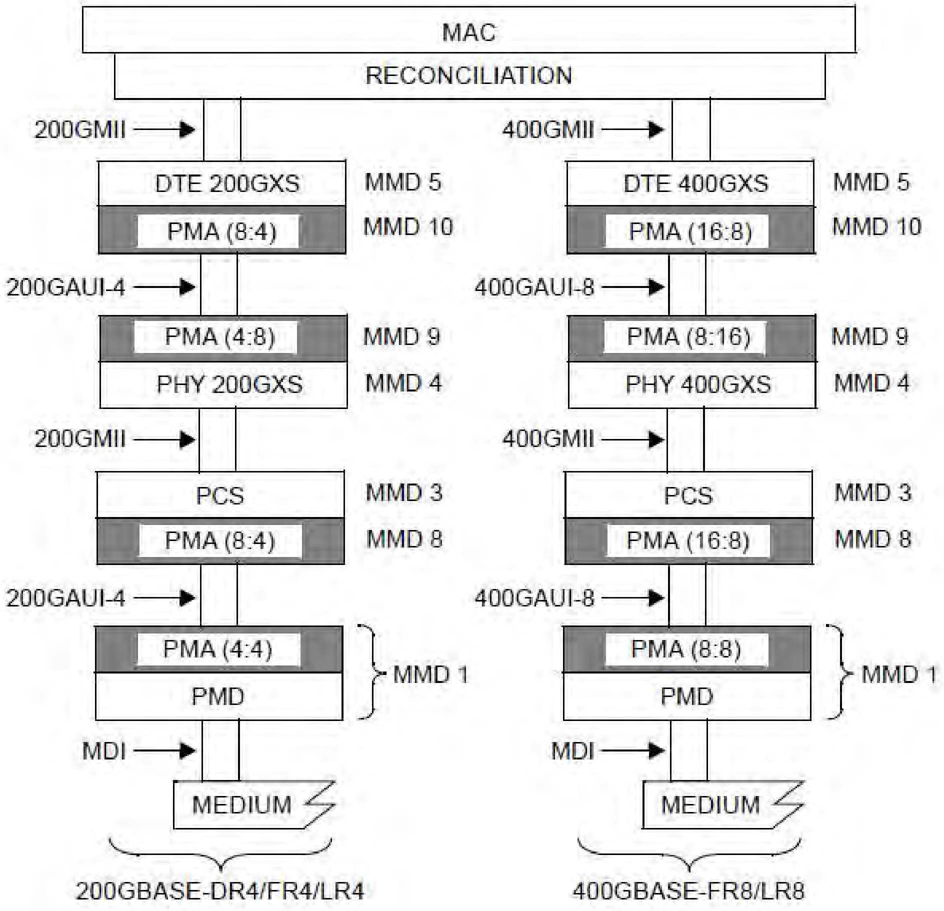
Support for 200G/lane PMD with 100G/lane ASIC
100G/Lane ASIC
200G/Lane C2M
MAC/RS
800GMII
XS1-DTE (FEC1)
PMA
AUI-8
PMA
XS1-PHY (FEC1)
800GMII
PCS4 (FEC4)
PMA
AUI-4
PMA
FEC5
PMA
PMD
MDI
Medium
Extender
PHY
Mux Device
Module

800GBASE-DR4

MMD Numbering

- Need to provide for MMD numbering for all possible implementations

120A.4 Partitioning example using 200GXS and 400GXS

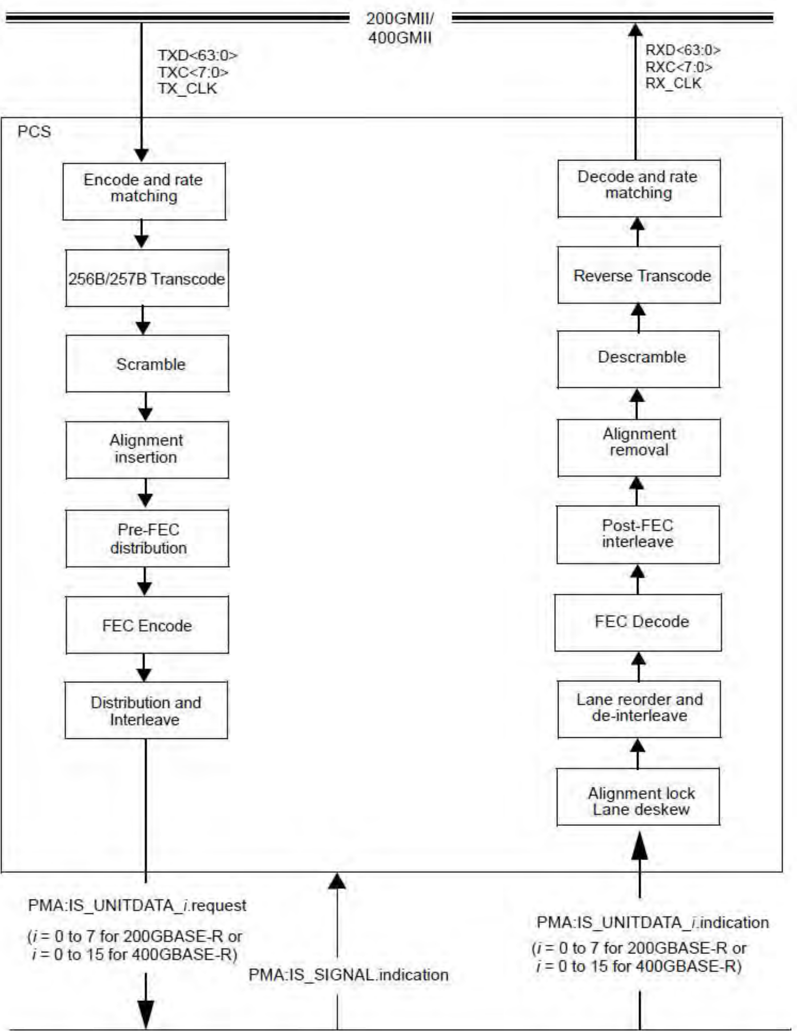


This MII Extender can be Confusing…What is it Really?

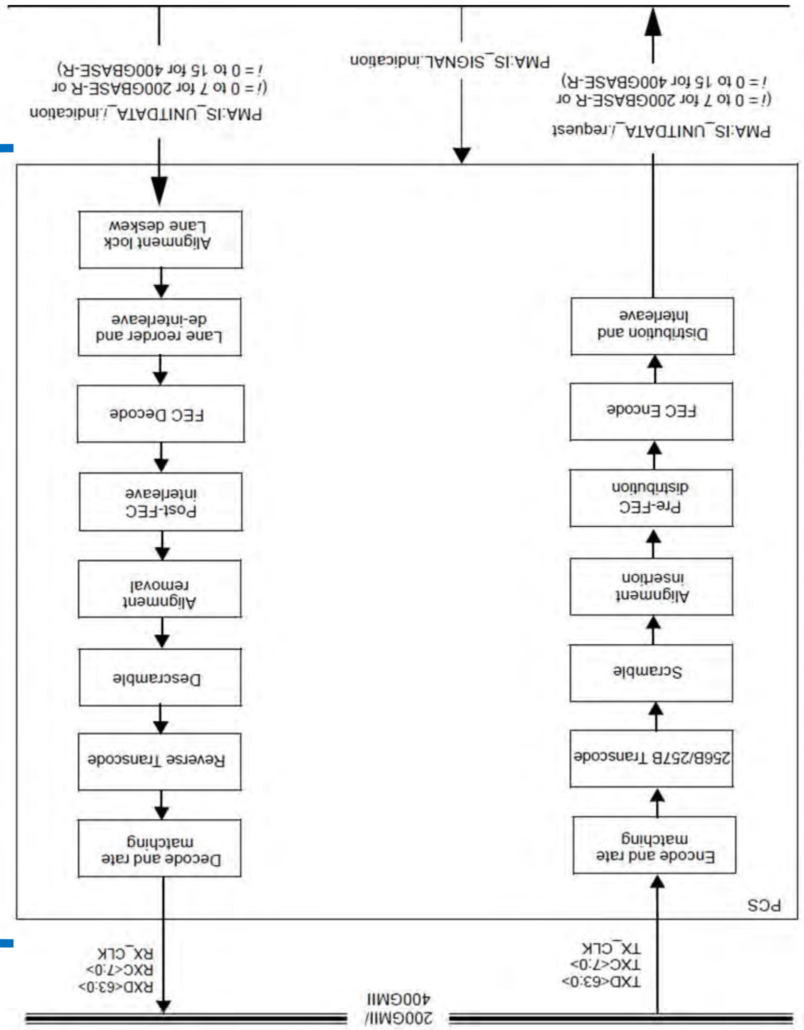
- It is a way to get back to the MII, by removing the PCS/FEC coding, so you can add different coding
  - The DTE XS is a forward PCS, the PHY XS is a reverse (opposite direction) PCS
    - In the proposed architecture the PCS includes a FEC
  - The XS clause mostly refers to the PCS clause, and adds MMD numbering etc. (see Clause 118)
- In an implementation, you don’t need to go back to the MII, you can optimize logic as makes sense
  - As an example, if an implementation only requires the FEC to be terminated, then you terminate the FEC and leave the transcoded blocks for re-encoding to the new FEC

XS Example (Based on Clause 119/118)

XS-DTE



XS-PHY



Things to add or think more about

- Assume that the OTN reference point would be the same as in 802.3bs (a stream of 64b/66b blocks with local/remote degrade)?
  - More discussion needed

- How do we ensure that we can maintain accurate timestamping
- If we adopt a segmented FEC scheme, we can think about specifying the XS required blocks more completely
  - Full PCS termination down to the MII provides an opportunity for timing adaptation – resulting in timing uncertainty
  - The XS could be defined as a full termination, leaving solutions open to individual implementations, or a partial termination could be specified with timing transparency

Summary

- Adopt the same high-level architecture for all 4 Ethernet rates in this project
  - Supports multiple FEC schemes
    - End to end, segmented, concatenated or a mix
    - Allows us to address all objectives, regardless of FEC scheme, and allow system designers to minimize the potential for stranded ports
- All based on the 802.3bs architecture
  - A FEC is collocated in the PCS sublayer
  - Other FECs can be added as need for a given requirement
- The specified FEC schemes will be dependent on the PHY baseline proposals
  - Different PHYs can use different FEC schemes

Proposed Straw Poll

- I would support adopting slide 11 as the logic architecture for this project
  - Y
  - N
  - Need more information

Thanks!