400GbE PCS Architectural Requirements

IEEE 400 Gb/s Ethernet Study Group

TBD…

Mark Gustlin - Xilinx

Chris Cole – Finisar

Pete Anslow – Ciena

John D'Ambrosia - Dell

Introduction

- The following slides explore the architectural requirements for a 400GbE PCS

High Level Architectural Requirements

- Define an optional AUI
  - Chip-to-Module and Chip-to-Chip
- Support EEE
- Appropriate support for OTN
- Support FEC(s) either initially or in the future depending on the PMD or AUI requirements
  - Location of the FEC is TBD
- Ability to support (and change as necessary) TBD lane widths
  - Options detailed later in this document
- Define a common service interface for all sublayers
- Define an optional CDGMII
- Provide appropriate support for multi-rate implementations
  - For example common lane rates, common PCS technology etc.

AUI and MDI Evolution

- A big part of what the PCS architecture should support is a graceful evolution of the AUI and the MDI
- The next 8 slides look at the possible evolution of the AUIs and MDIs
- A possible evolution of the AUI is

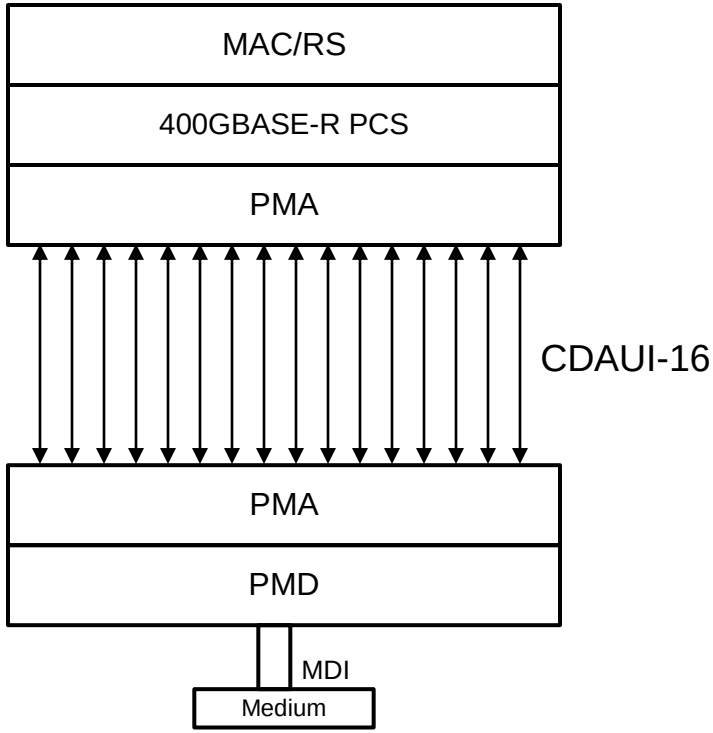
  CDAUI-16 -> CDAUI-8 -> CDAUI-4

  CDAUI-16 -> CDAUI-10 -> CDAUI-4

  It is also possible to define CDAUI-10 and CDAUI-8 at the same time
- The MDI likely won't evolve in a straight line due to the varying reach and medium needs, the PCS needs to be flexible enough to support the various lane requirements over time (a lane can be a lambda, fiber, copper cable or combination)

Electrical Interface Option #1

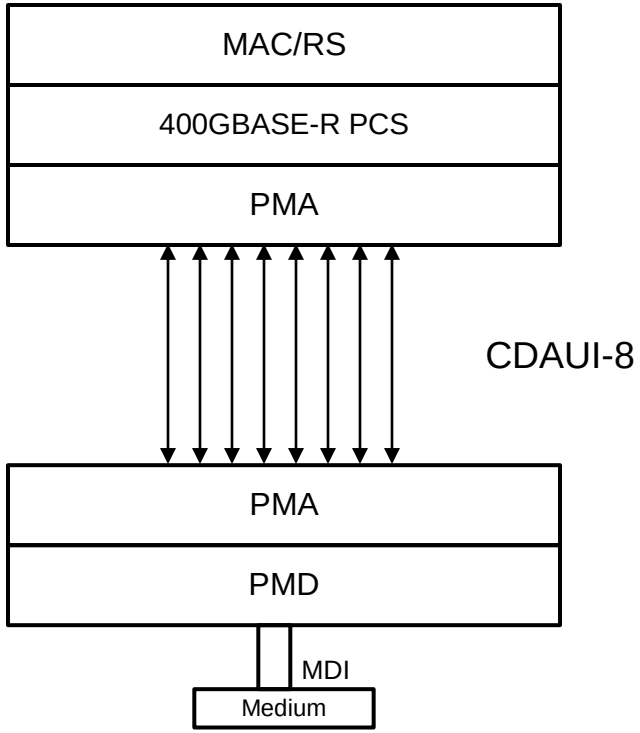
- Properties:
  - 16 lanes
  - 25.78125 Gb/s per lane
  - NRZ signaling
  - 10dB loss for chip-to-module
  - 15-20dB loss for chip-to-chip
  - No FEC required to close the interfaces*
- **Benefits**
  - Leverages existing CAUI-4 work



* If FEC is always sent, this interface can take advantage of that with a relaxed BER budget

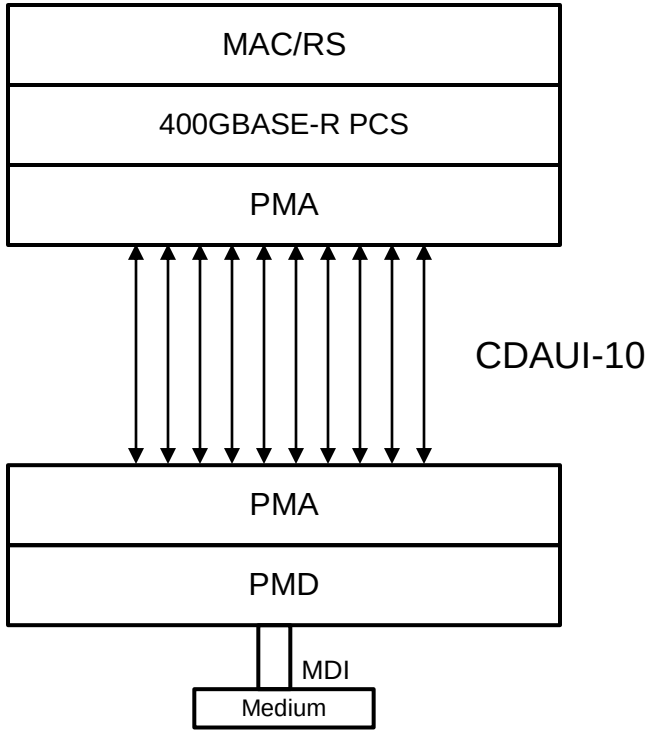
Electrical Interface Option #2

- Properties:
  - 8 lanes
  - 51.5625 Gb/s per lane
  - NRZ signaling?
  - 20dB? loss for chip-to-module
  - ?dB loss for chip-to-chip
    - Might not be NRZ for this one
  - No FEC required to close the chip-to-module interface?
  - Likely need FEC to close the chip-to-chip interface?



Electrical Interface Option #3

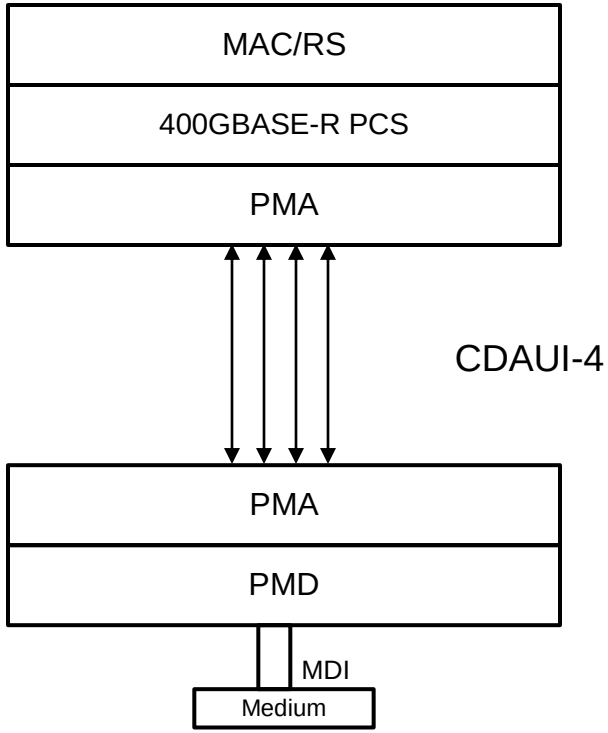
- Properties:
  - 10 lanes
  - 41.25 Gb/s per lane
  - NRZ? signaling
  - 20dB? loss for chip-to-module
  - ?dB loss for chip-to-chip
  - No FEC required to close the chip-to-module interface?
  - Likely need FEC to close the chip-to-chip interface?
  - Use case for supporting 40G VCSELs?
    - Question to answer: Is there is likely to be something that works for 400GBASE-?R10 but not 400GBASE-?R8 that is not a short term technology limit



Can MLG be used for this application instead of defining a CDAUI-10?

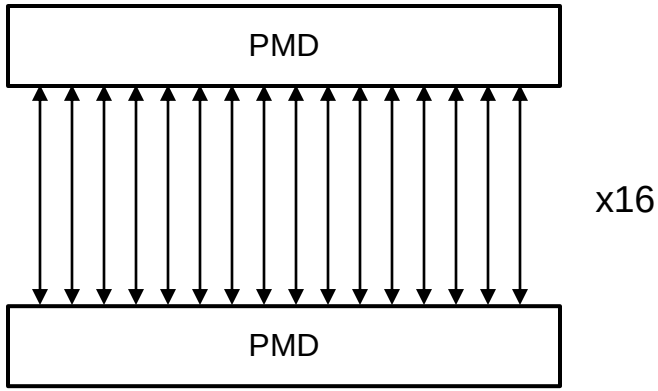
Electrical Interface Option #4

- Properties:
  - 4 lanes
  - 103.125 Gb/s per lane
  - PAMn? signaling
  - dB? loss for chip-to-module
  - ?dB loss for chip-to-chip
  - Likely need strong FEC to close the interfaces?



Possible Re-use Optical/Electrical PMD

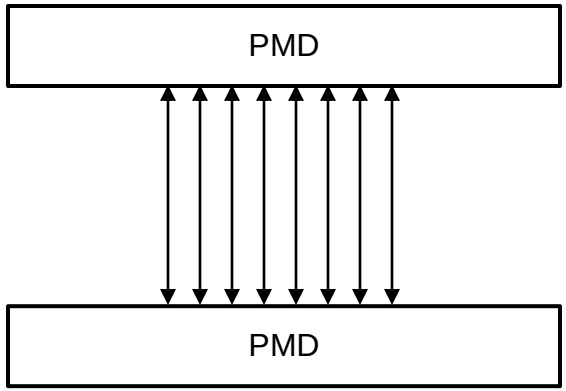
- Properties:
  - 16 lanes
  - 25.78125 Gb/s per lane
  - NRZ signaling
  - FEC required for some possible variations (SR16, CR16 for instance), not needed for others (LR16)



Note: number of lanes can be constituted from lambdas, fibers, twinax or some combination

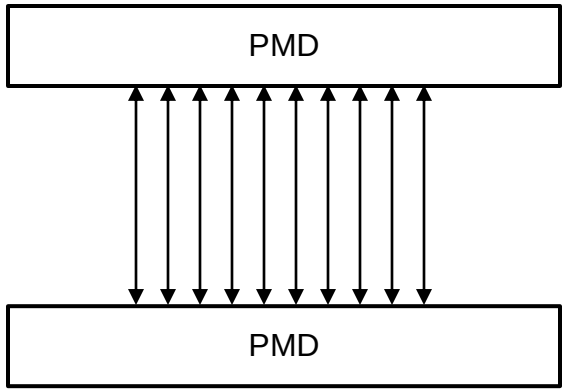
Possible Future Optical PMD

- Properties:
  - 8 lanes
  - 51.5625 Gb/s per lane
  - PAM4 signaling?
  - FEC required??



Possible Future Optical PMD?

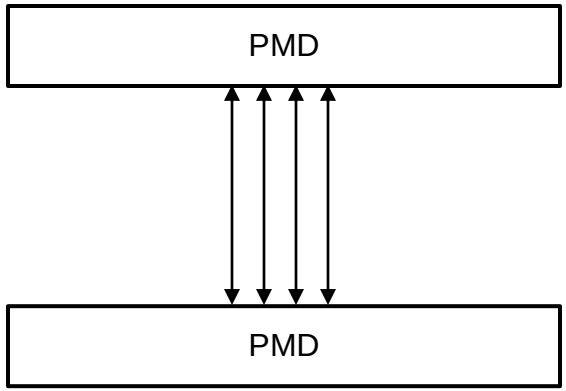
- Properties:
  - 10 lanes
  - 41.25 Gb/s per lane
  - NRZ signaling?
  - FEC required??
  - Re-use of 40GbE technology?
  - Use case for supporting 40G VCSELs?
    - Question to answer: Is there is likely to be something that works for 400GBASE-?R10 but not 400GBASE-?R8 that is not a short term technology limit



Can MLG be used for this application instead of defining a CDAUI-10?

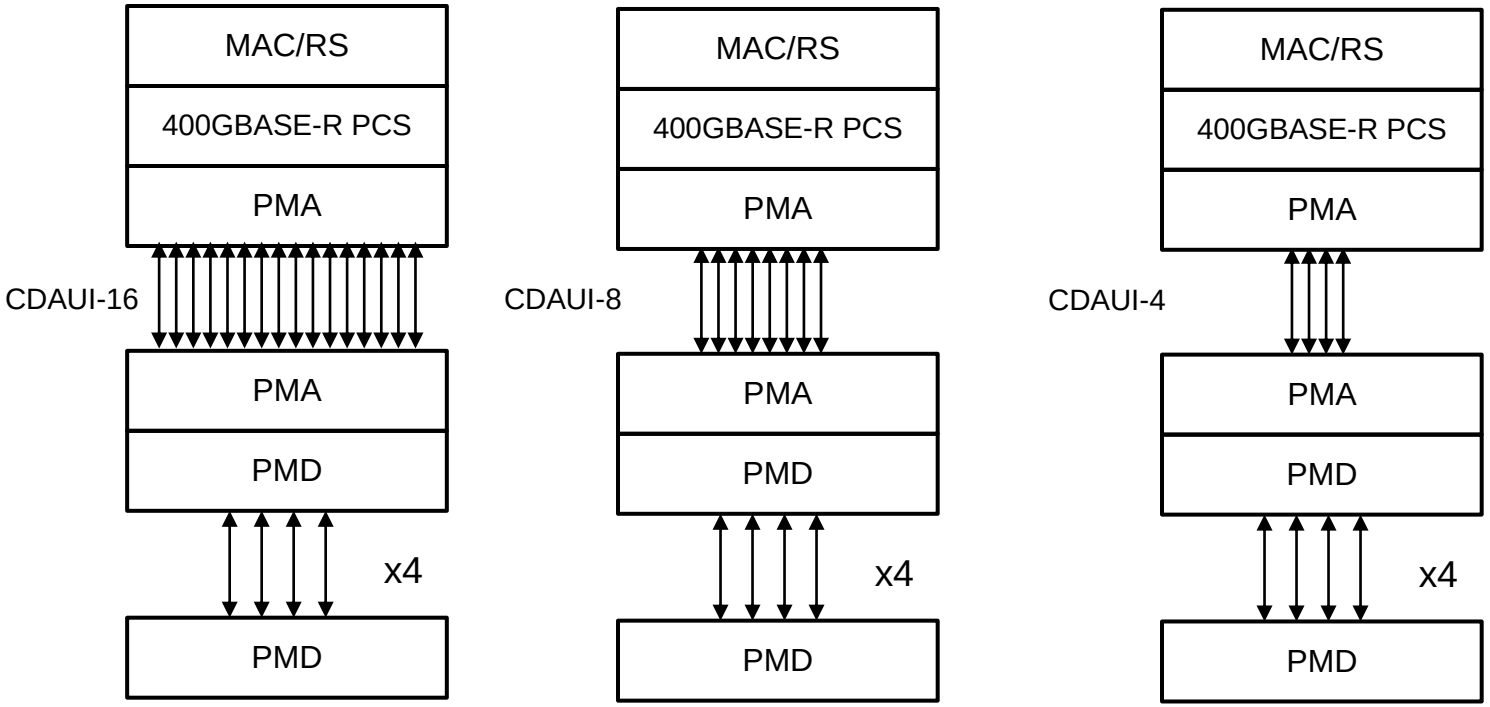
Possible Future Optical PMD

- Properties:
  - 4 lanes
  - 103.125 Gb/s per lane
  - PAM8, DMT etc..
  - Strong FEC required
  - Assuming that with higher order modulation alignment or other needed information can be added to the structure, does the PCS support for 4 lanes matter?



Long Lived Optical Interface

- Below shows an example of a possible long lived PMD with 3 electrical generations
- We should strive to define an architecture that is long lived and does not unduly burden the optics module
- The need or not for FEC on the electrical interface can complicate the module if that FEC must be terminated in the module
  - Even worse is if different FECs are used, Latency is multiplied for instance

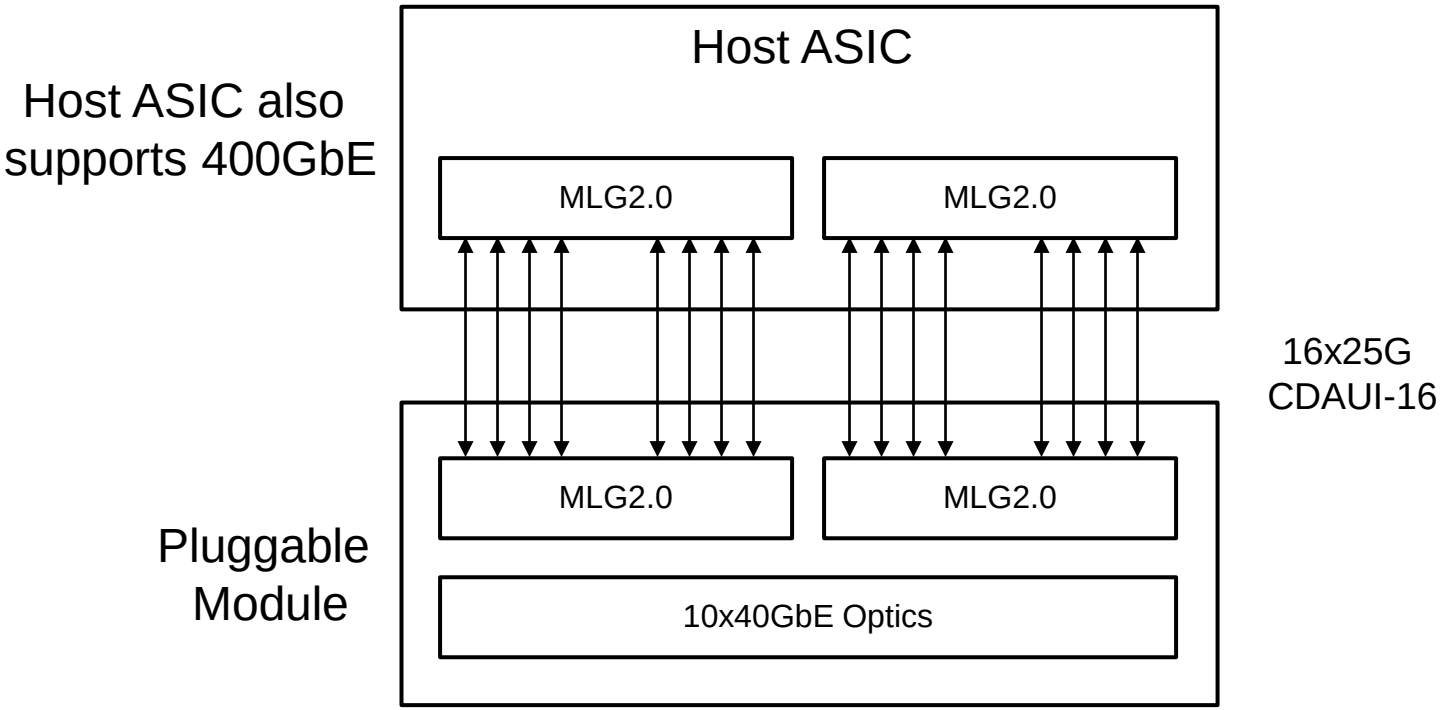


MLG and 400GbE

- The following slides show how MLG 1.0 and 2.0 can be used to carry nx40GbE and nx10GbE across CDAUI interfaces
- So instead of defining 40Gb/s lane modes, can MLG be used?
- MLG is based on 5G VLs, 400GbE can be based on 25G VLs and MLG can be used to carry 40GbE across the same 400GbE electrical interfaces
- Using 25G lanes precludes a CDAUI-10, but likely not needed given the MLG capability?

MLG Across CDAUI-16

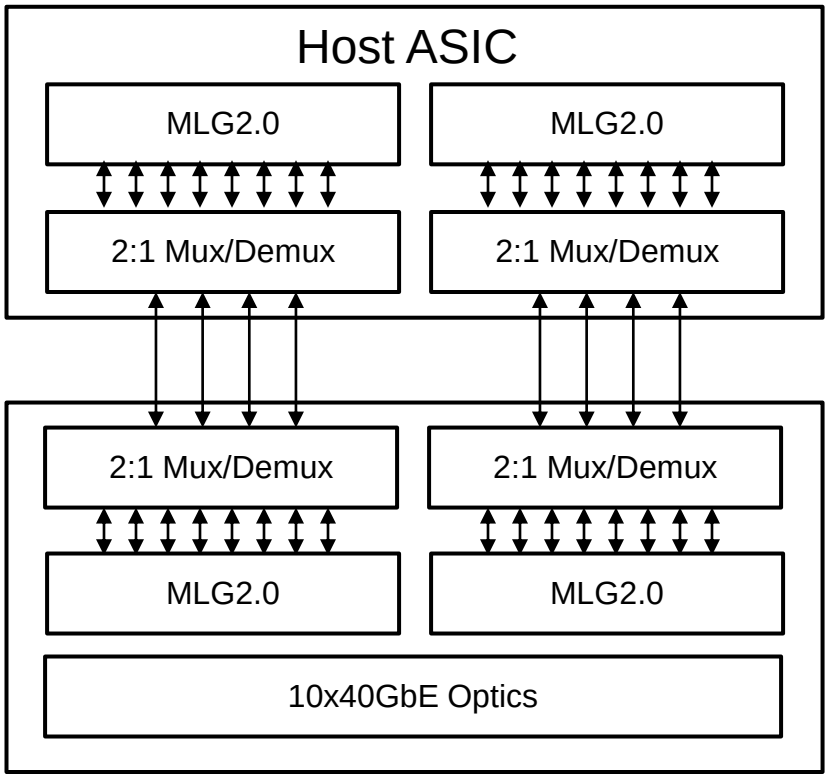
- How to carry 10x40G across a 16 lane 400GbE electrical interface?
- Very simple, MLG 2.0 is defined to support 8x40GbE across 8 lanes of 25G SerDes, so run two instances of MLG 2.0 across the 16x25G interface



MLG Across CDAUI-8

- How to carry 10x40G across an 8 lane 400GbE electrical interface?
- Very simple, MLG 2.0 is defined to support 8x40GbE across 8 lanes of 25G SerDes, so run two instances of MLG 2.0 across the 16x25G interface and then mux one more time to 50G

Host ASIC also supports 400GbE



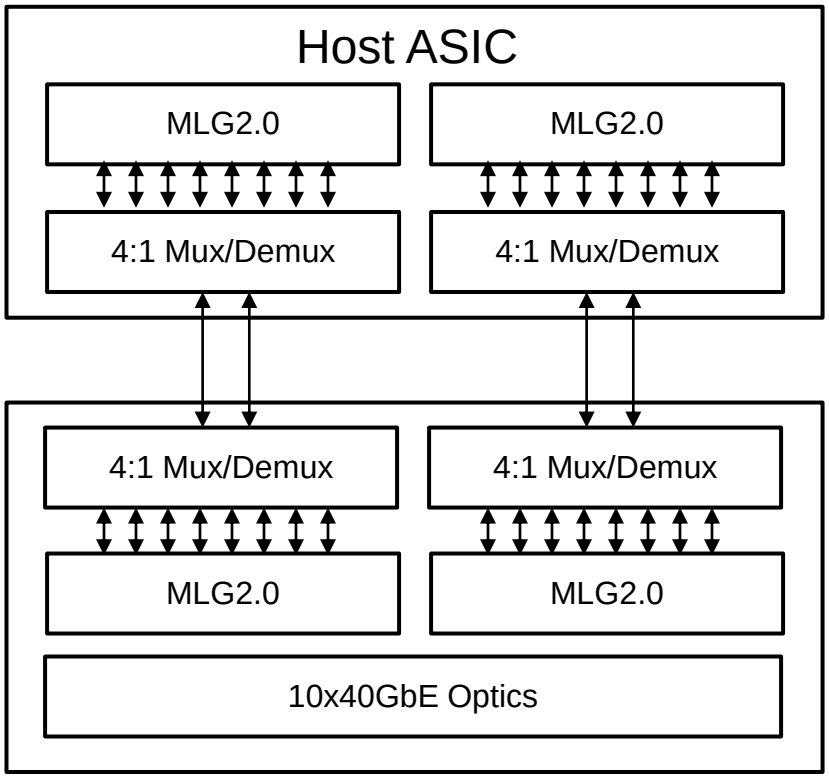
8x50G CDAUI-8

Pluggable Module

MLG Across CDAUI-4

- How to carry 10x40G across a 4 lane 400GbE electrical interface?
- Very simple, MLG 2.0 is defined to support 8x40GbE across 8 lanes of 25G SerDes, so run two instances of MLG 2.0 across the 16x25G interface and then mux one more time to 50G

Host ASIC also supports 400GbE



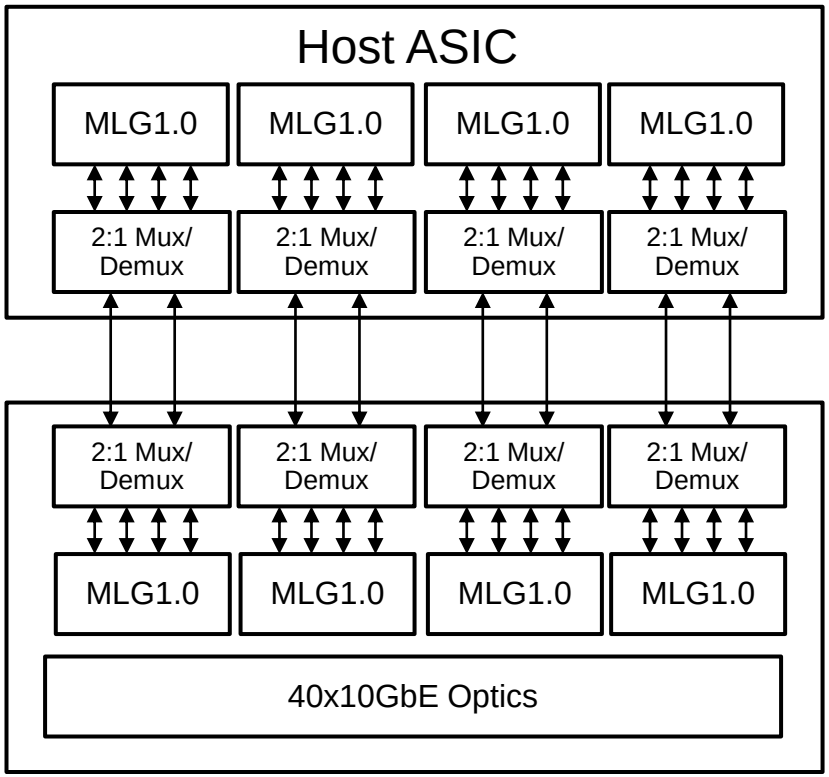
4x100G CDAUI-4

Pluggable Module

MLG Across CDAUI-8

- How to carry 40x10GbE across an 8 lane 400GbE electrical interface?
- Very simple, MLG 1.0 is defined to support 10x10GbE across 4 lanes of 25G SerDes, so run four instances of MLG 1.0 across the 16x25G interface and then mux one more time to 50G

Host ASIC also supports 400GbE



8x50G CDAUI-8

Pluggable Module

Summary

- Chip-to-module lane widths: 16, 10, 8, 4
- PMD lane widths: 16, 10, 8, 4
- If we do MLD, then the LCM of these is 80!
  - With 10 being the sticky number that causes a jump in number of PCS lanes
  - Question to answer: Is there is likely to be something that works for 400GBASE-?R10 but not 400GBASE-?R8 that is not a short term technology limit
  - Would the IEEE really standardize two electrical interfaces (CDAUI-8 and CDAUI-10) that are so similar, and can lead to plug and play issues?
- MLG can remain unchanged and stay with 5G VLs and support 10G, 40G etc. across CDAUI-16, -8 etc…
- A lot of questions remain around FEC

Interface Type	Lane Widths the Architecture should support over time?					
Chip-to-module I/F	16	10	8	4	2	1
PMD Lanes	16	10	8	4	2	1

Thanks!