400GbE PCS Options

IEEE P802.3bs 400 Gb/s Ethernet Task Force

November 2014 San Antonio

Mark Gustlin – Xilinx
Gary Nicholl – Cisco
Dave Ofelt – Juniper
Steve Trowbridge - ALU

Introduction

- This looks at possible PCS architectures for a couple of PMD options that are being proposed
  - It does not exhaustively explore a PCS for each PMD type that has been proposed
  - But many would fit into one of the two categories, KP4 FEC that is relatively cheap to implement and the second category where a significantly stronger FEC needs to be used

References

- 400G PCS and FEC options:

http://www.ieee802.org/3/bs/public/14_09/anslow_3bs_02_0914.pdf
http://www.ieee802.org/3/bs/public/14_09/wang_z_3bs_01_0914.pdf
http://www.ieee802.org/3/bs/public/14_09/wang_t_3bs_01a_0914.pdf
http://www.ieee802.org/3/bs/public/14_07/wang_x_3bs_01_0714.pdf
http://www.ieee802.org/3/bs/public/14_07/trowbridge_3bs_01_0714.pdf
http://www.ieee802.org/3/bs/public/14_07/wang_t_3bs_01_0714.pdf
http://www.ieee802.org/3/bs/public/14_07/gustlin_3bs_04_0714.pdf
http://www.ieee802.org/3/bs/public/14_07/gustlin_3bs_02_0714.pdf
http://www.ieee802.org/3/bs/public/14_05/wang_x_3bs_01_0514.pdf
http://www.ieee802.org/3/bs/public/14_05/trowbridge_3bs_01_0514.pdf
http://www.ieee802.org/3/bs/public/14_05/barrass_3bs_01_0514.pdf
http://www.ieee802.org/3/400GSG/public/13_09/wang_400_01_0913.pdf
http://www.ieee802.org/3/400GSG/public/13_09/begin_400_01_0913.pdf
http://www.ieee802.org/3/400GSG/public/13_09/ghiasi_400_01_0913.pdf
http://www.ieee802.org/3/400GSG/public/13_09/song_400_01_0913.pdf
http://www.ieee802.org/3/400GSG/public/13_09/wang_z_400_01_0913.pdf
http://www.ieee802.org/3/400GSG/public/13_07/gustlin_400_02_0713.pdf
http://www.ieee802.org/3/400GSG/public/13_07/wang_400_01_0713.pdf
http://www.ieee802.org/3/400GSG/public/13_07/ghiasi_400_01_0713.pdf
http://www.ieee802.org/3/400GSG/public/13_05/ghiasi_400_01a_0513.pdf

Review Sublayer Functions

Sublayer	10GbE	100GbE	400GbE (proposed)
MAC	Framing, addressing, error detection	Framing, addressing, error detection	Framing, addressing, error detection
Extender	PCS + PMA	N/A	PCS + PMA + FEC
PCS	Coding (8B/10B, 64B/66B), lane distribution, EEE	Coding (64B/66B), lane distribution, EEE	Coding, lane distribution, EEE, FEC
FEC	FEC, transcoding	FEC, transcoding, align and deskew	N/A?
PMA	Serialization, clock and data recovery	Muxing, clock and data recovery, HOM	Muxing, clock and data recovery, HOM??
PMD	Physical interface driver	Physical interface driver	Physical interface driver

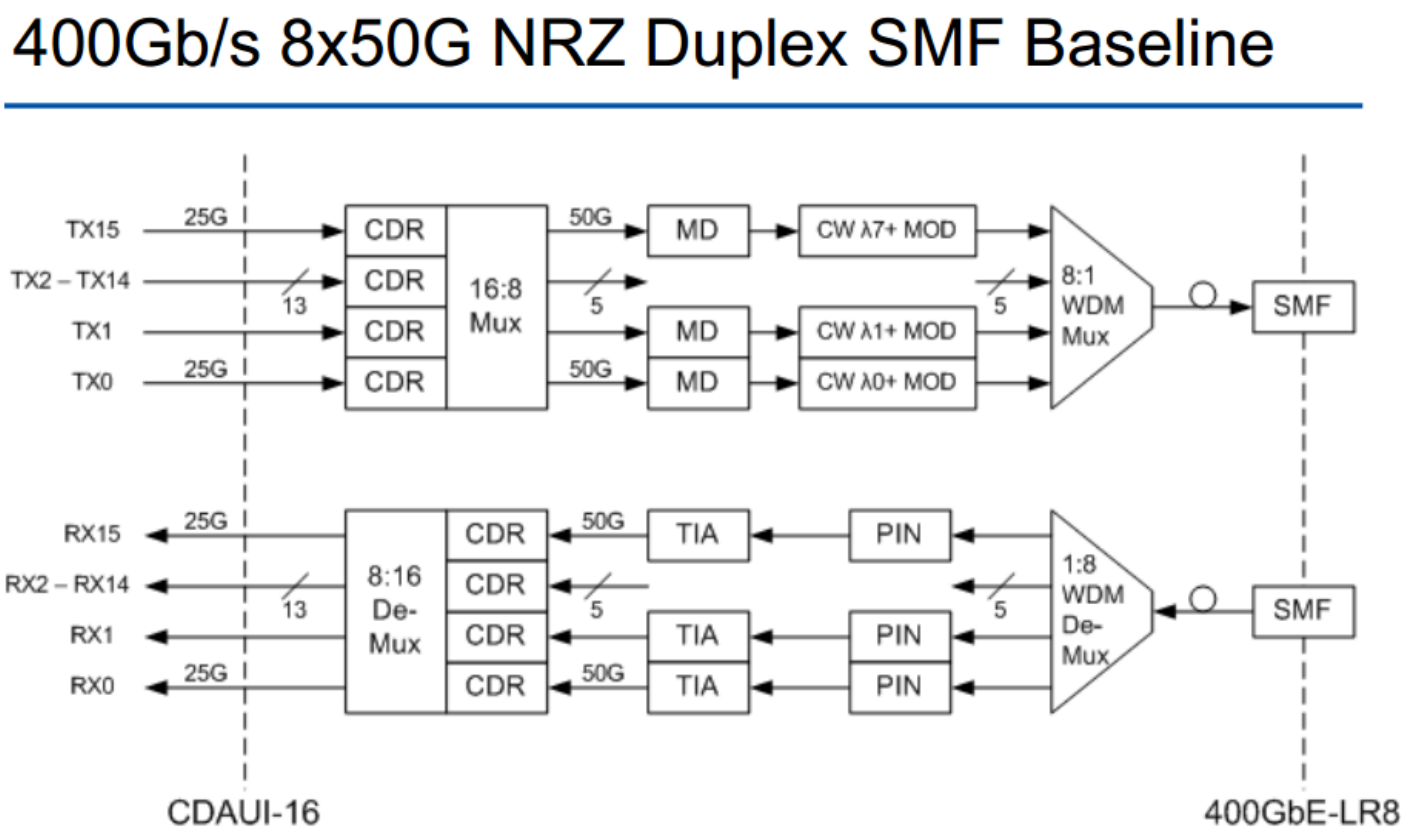
Note that there are variations with a single speed, not all are captured in this table

8x50G NRZ PMD

- As an example, we will look at what a PCS to support the 8x50G NRZ PMD might look like
  - This would also apply to 8x50G PAM4 which at this point is targeted to use KP4 FEC

- From cole_3bs_02b_914
- 8 wavelengths WDM over a 2km single duplex fiber, LAN WDM
- Current FEC target is KP4 FEC
  - KP4 and KR4 FEC can be flexibly done with a single implementation, making sharing between 100GbE and 400GbE efficient (assuming 400G FEC is implemented at 4x100G)
- The next number of slides explore the architecture and details of a possible PCS etc. that are required for this link

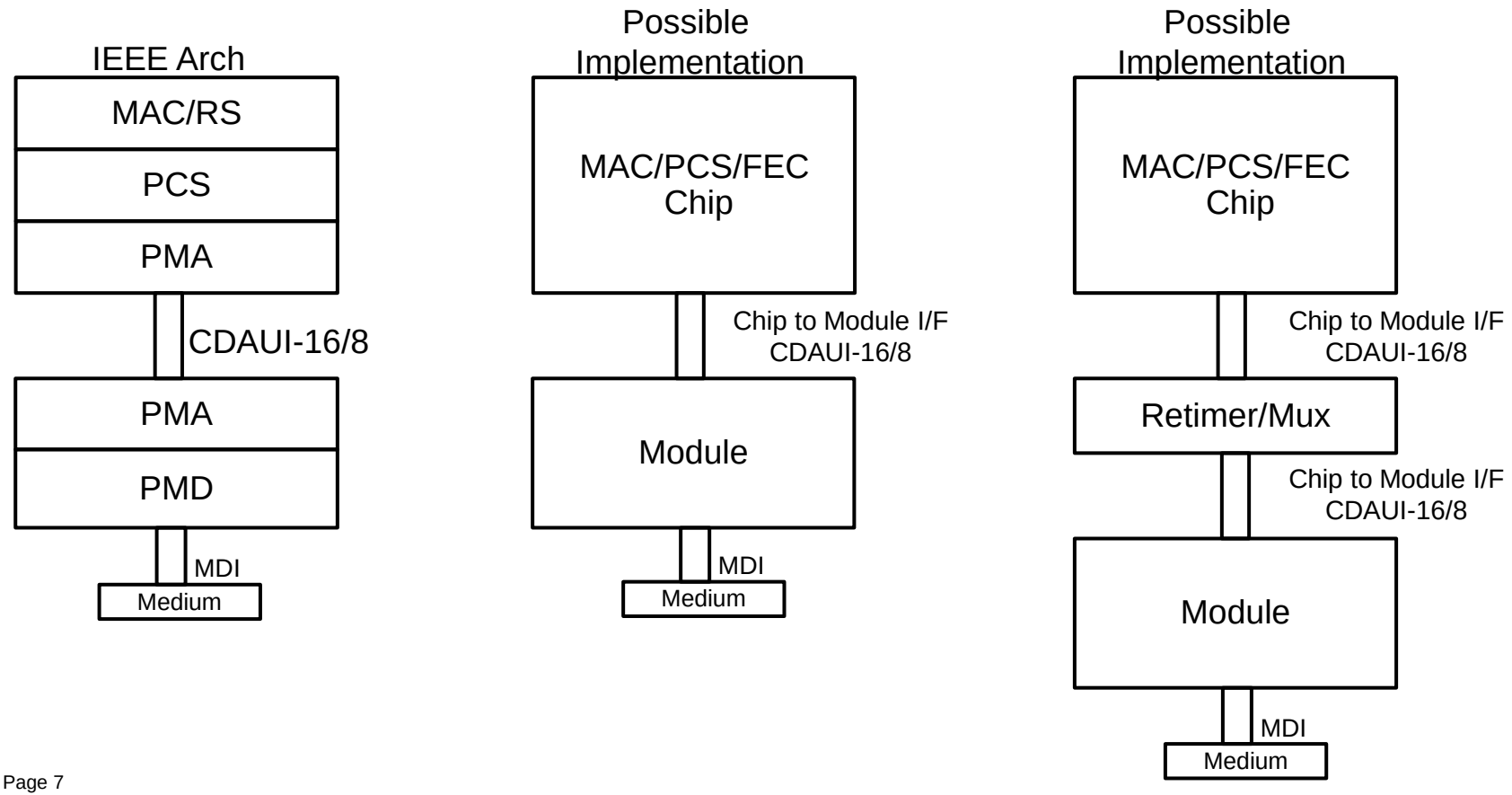
A Possible PMD Architecture



From cole_400_01a_1113

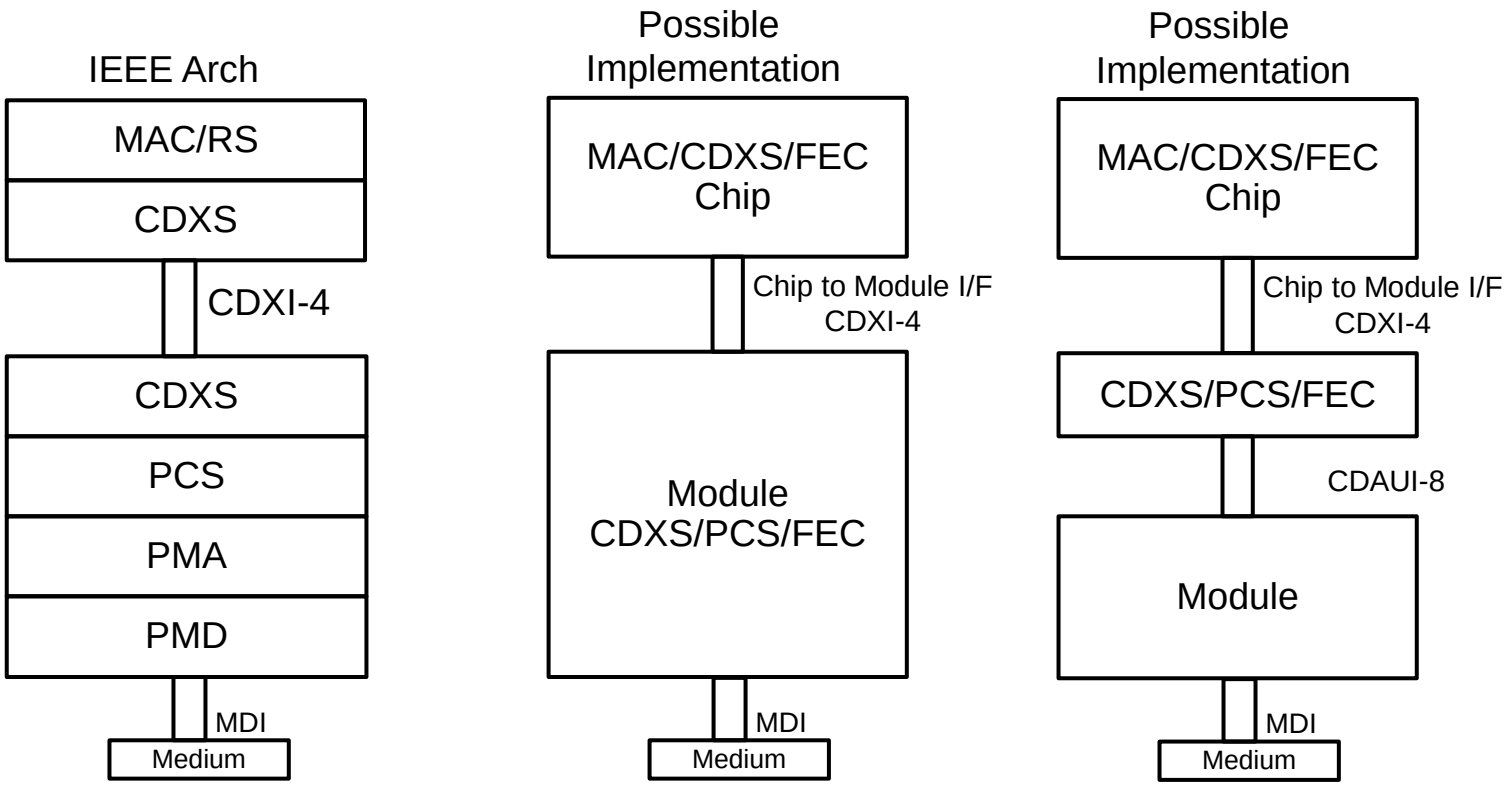
PCS Architecture for 8x50G NRZ PMD

- Likely implementation options, 16 lane CDAUI interfaces first and then 8 lane interfaces later
- You can mix the two, just PMA Muxing to go back and forth
- In this instance FEC is end to end, across up to 5 interfaces (in the PCS sublayer)
- Assuming a single FEC covers up to 5 interfaces



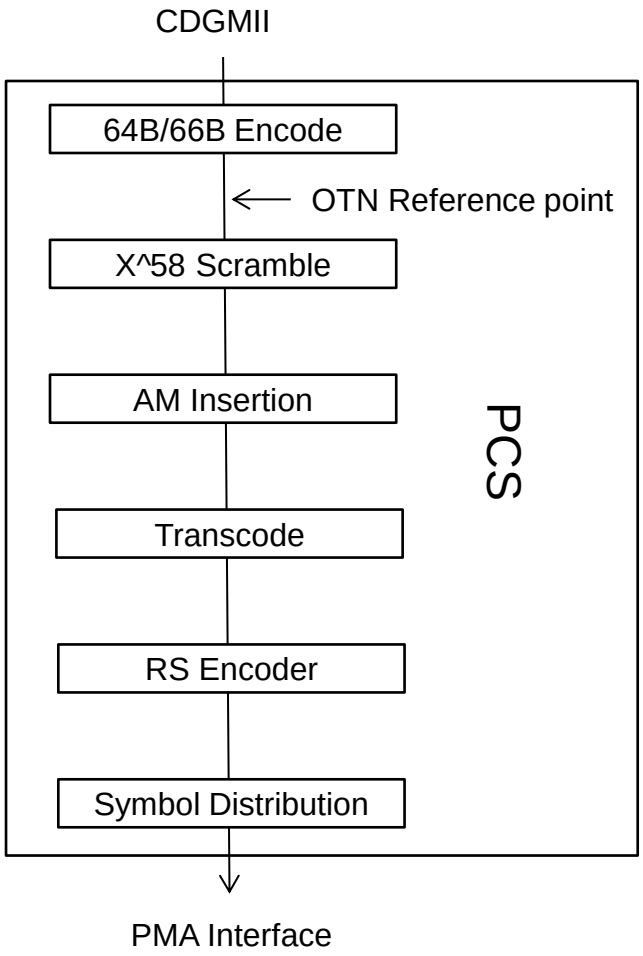
PCS Architecture for 8x50G NRZ PMD – Future?

- Possible future implementation, though might be unlikely?
- Module has shrunk, now has a 4 lane interface that requires a new high performance FEC, FEC is no longer end to end
- Might be more likely that the optical interface will move to 100G wavelengths and require a stronger FEC than the electrical interface



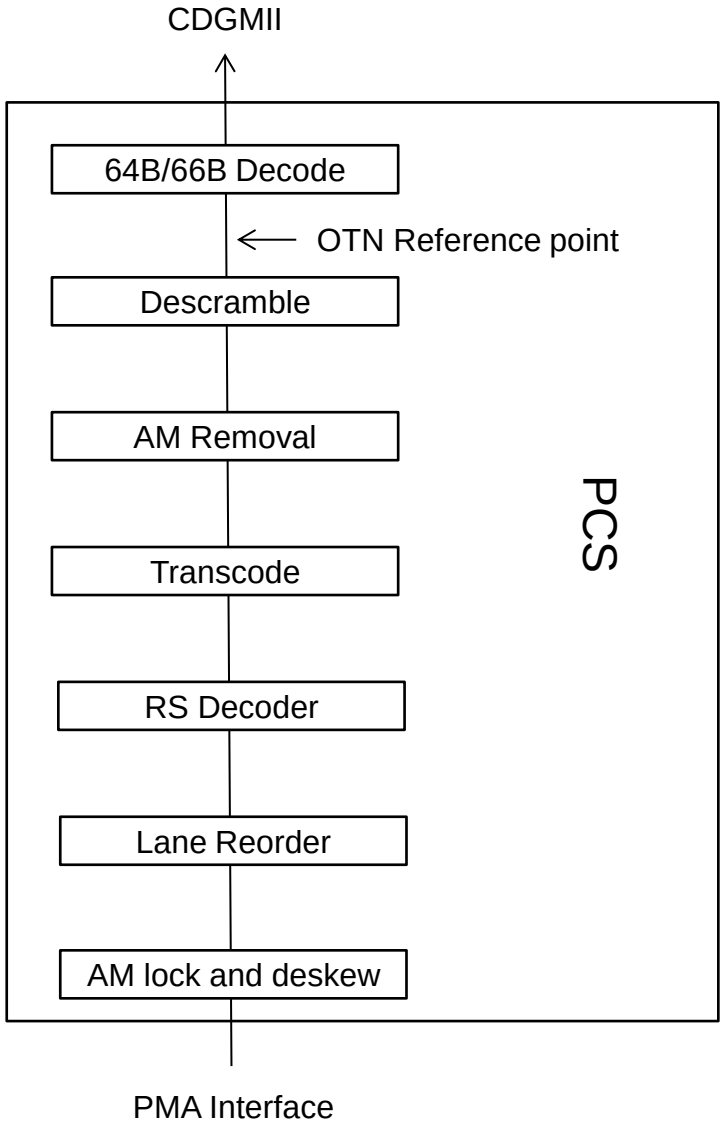
TX Data Flow

- 64B/66B encode based on clause 82
- Can look at direct 256B/257B encode as an option
- RS Encoder would support KP4 FEC RS(544,514,10)
- Open question if we should have a single 400G FEC or 4x100G
  - 4x100G has obvious re-use for 100GbE
  - Would the KP4 FEC be useful for mainstream 100G interfaces?
  - KP4 and KR4 FEC can share and encoder/decoder
- 16 PMA lanes (similar to PCS/FEC lanes)



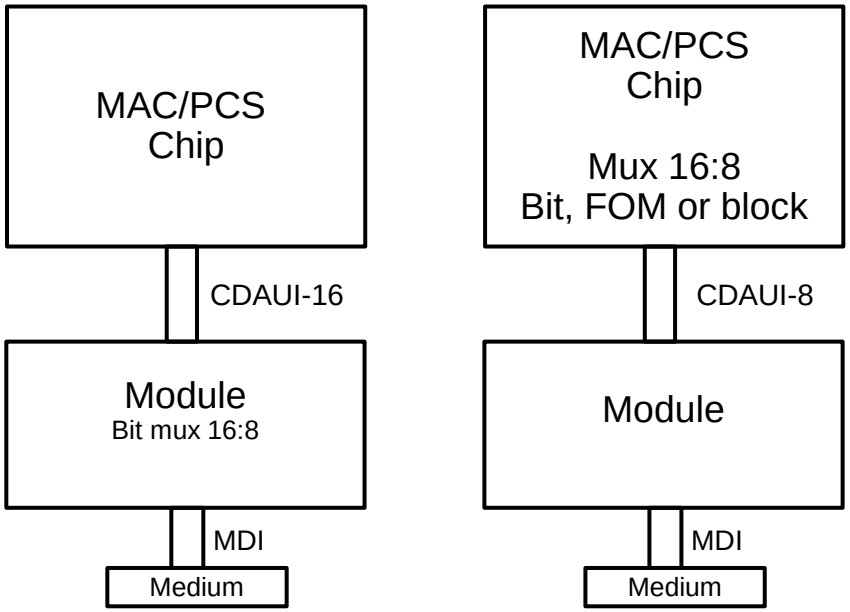
RX Data Flow

- Reverse of TX
- Allowing for arbitrary lane arrival



PMA Multiplexing

- With 16 PMA lanes, you can multiplex down to 8 lanes
- If muxing in the module, and if there are no correlated errors, you can bit mux without concern of the FEC block boundaries
- If muxing before an 8 lane (or less) electrical interface, or if the optical lanes have correlate errors, we need to understand the error models to see if we can do bit muxing, or if we need to do FOM or block level muxing



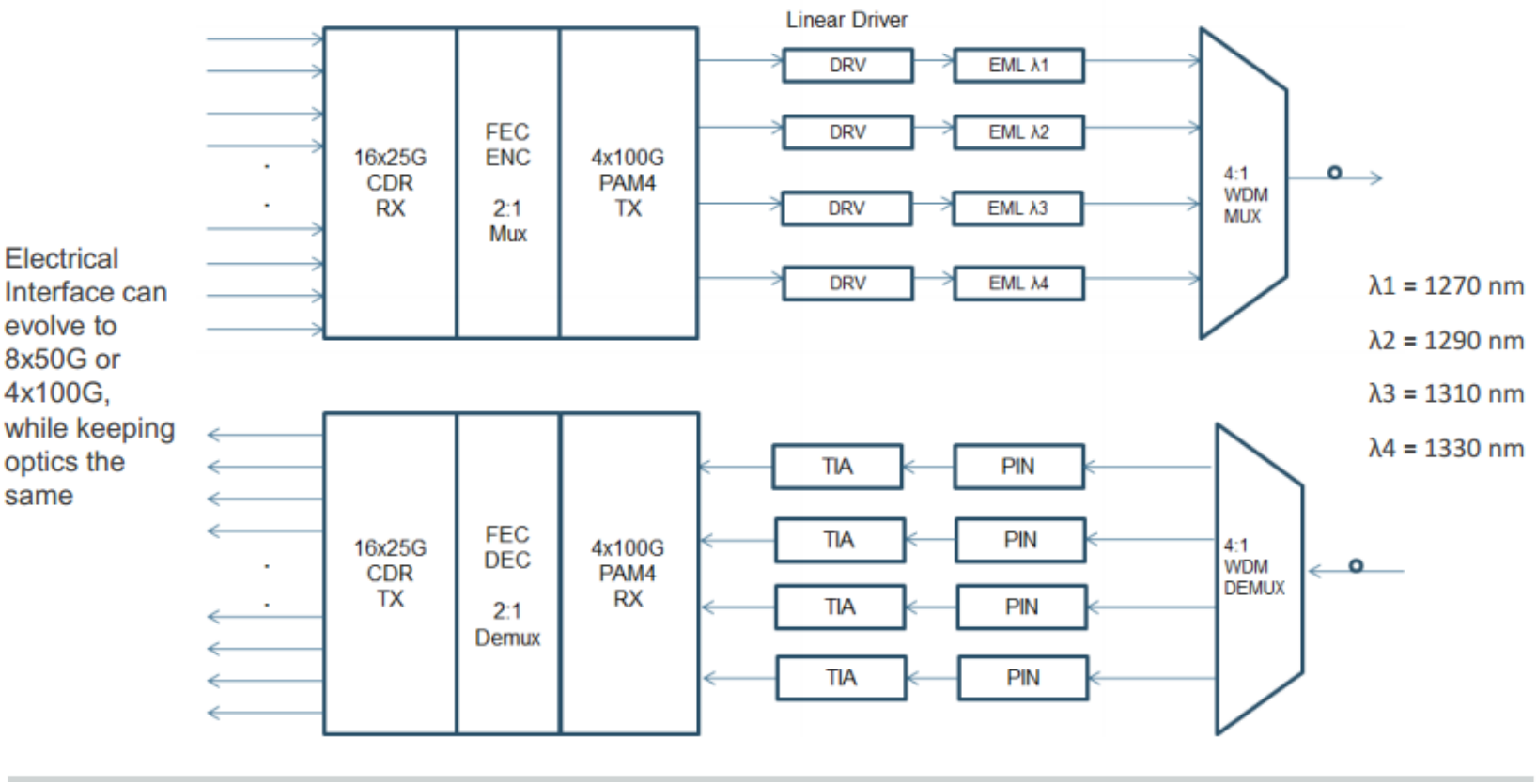
4x100G PAM-4 PMD

- As an example, we will look at what a PCS to support the 4x100G PAM4 PMD might look like
  - This analysis can also apply to other PMDs that need very strong FEC

- From bhatt_3bs_01a_0714
- 4 wavelengths WDM over a 2km single duplex fiber, PAM-4 Encoding
- Current FEC target was KP4 FEC. But have heard that stronger FEC is likely needed?
  - If KP4 is all that is needed, then use the architecture as shown for the NRZ 8x50G PMD
  - The following slides explore the need and impact of a stronger FEC, assuming high gain FEC (BCH or equivalent) is needed
  - There is a tradeoff to be made between a higher gain FEC vs. low SNR components
- The next number of slides explore the architecture and details of a possible PCS etc. that are required for this link

Possible PMD Architecture

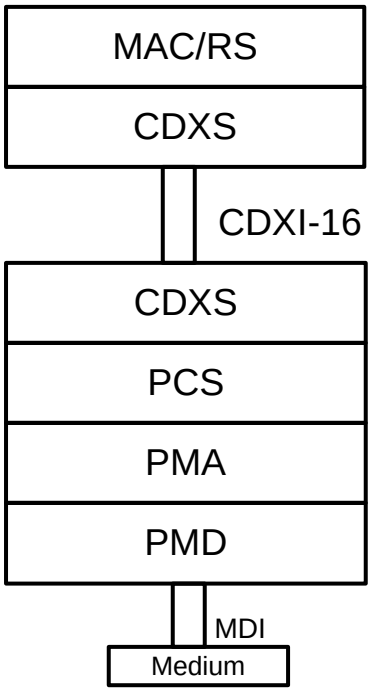
Proposed PMD for 2 km Objective

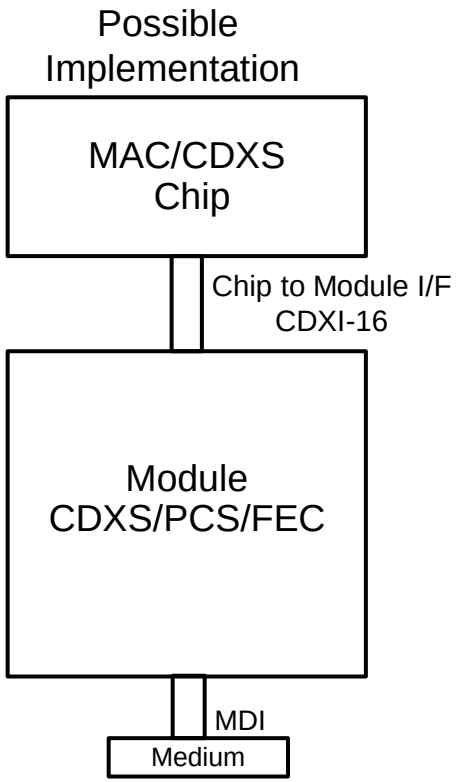


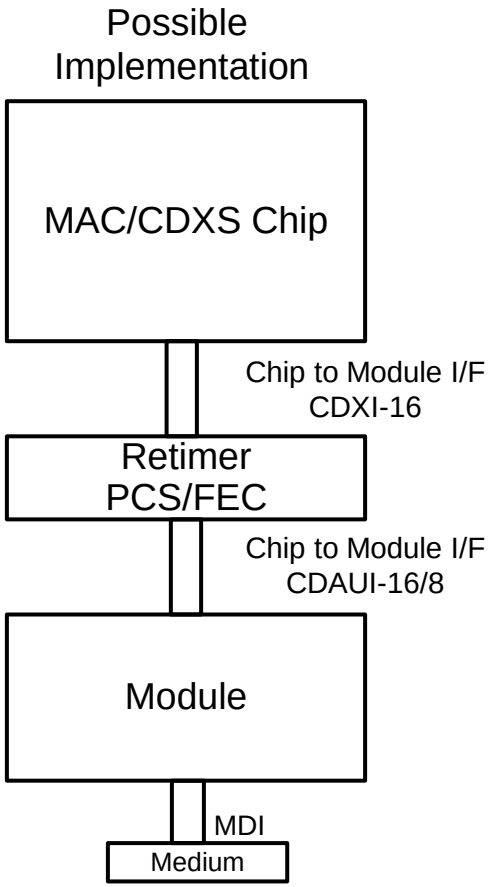
From bhatt_3bs_01a_0714

4x100G PAM4 Architecture – Option 1

- Likely implementation options, 16 lane CDAUI interfaces first
  - No FEC required, but you can have FEC if desired
- PCS is in the module or in a ‘retimer’ chip
- Strong FEC is from PCS to PCS, does not cover the 16 lane electrical interface
- Assuming a single FEC covers up to 5 interfaces

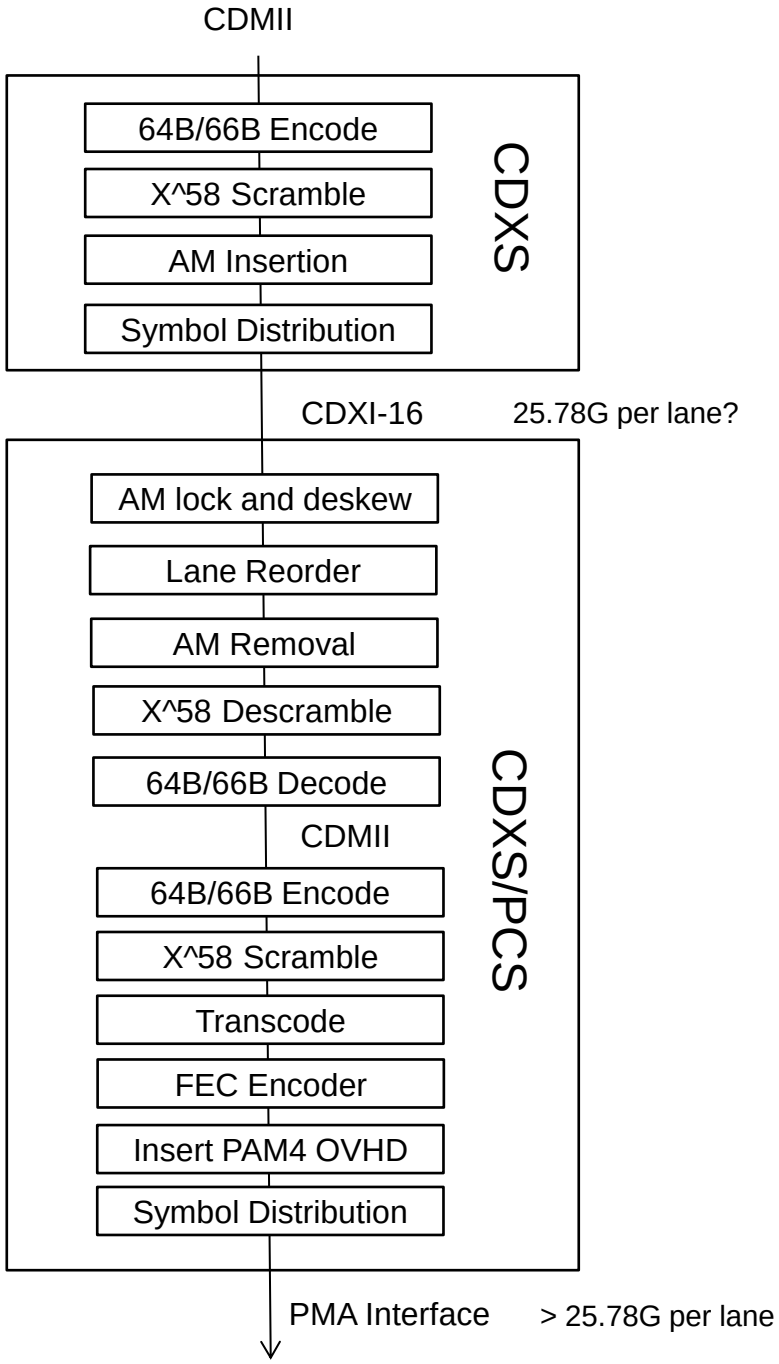






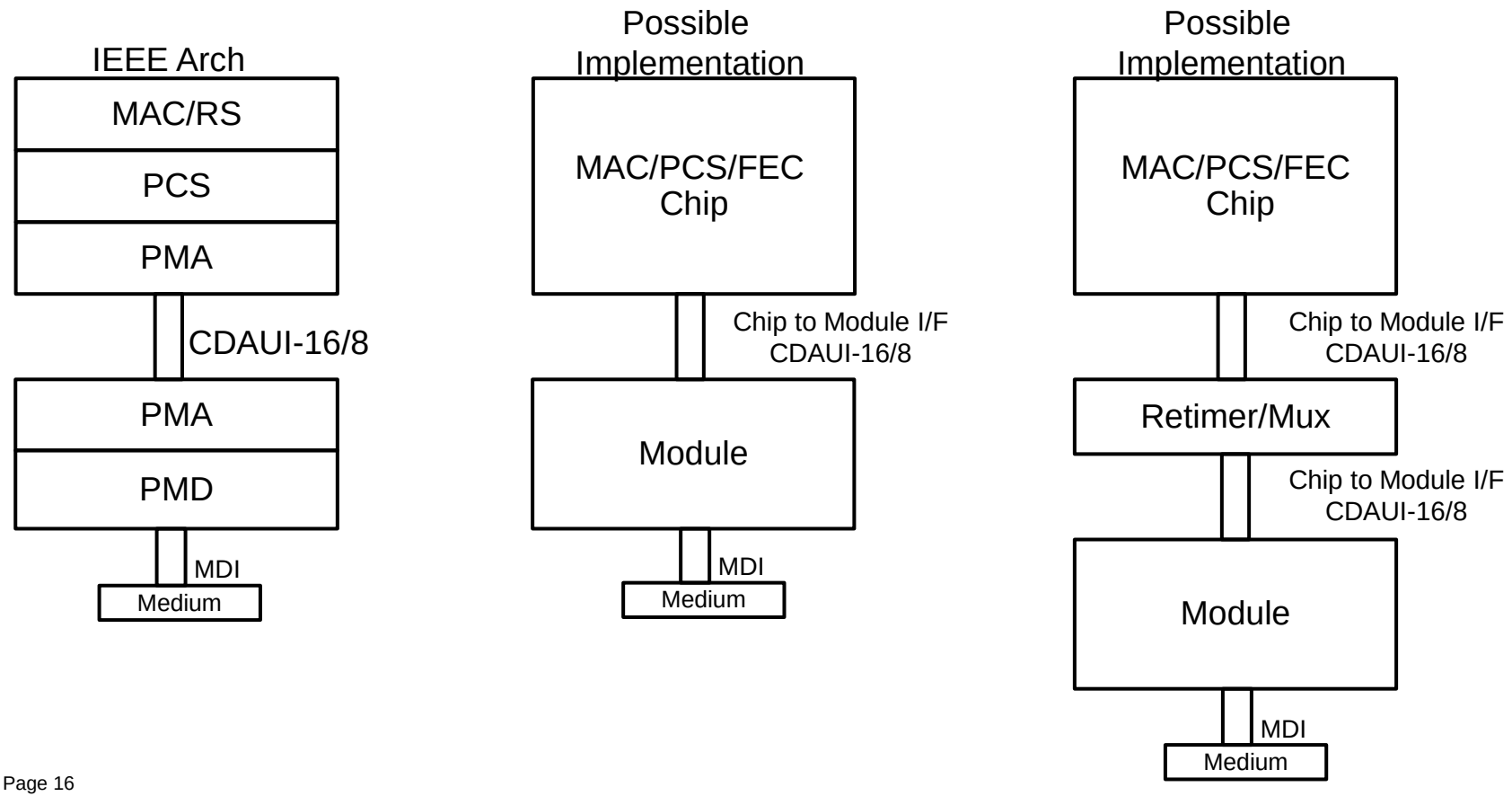
TX Data Flow Option 1

- 64B/66B encode based on clause 82
- Can look at direct 256B/257B encode as an option
- The CDXI might have FEC across it?
- FEC type is TBD, but high gain
- The diagram shows the standards defined functions, in the combo CDXS/PCS chip, the actual implementation can be simplified greatly, descramble/scramble and decode/encode can be eliminated for example



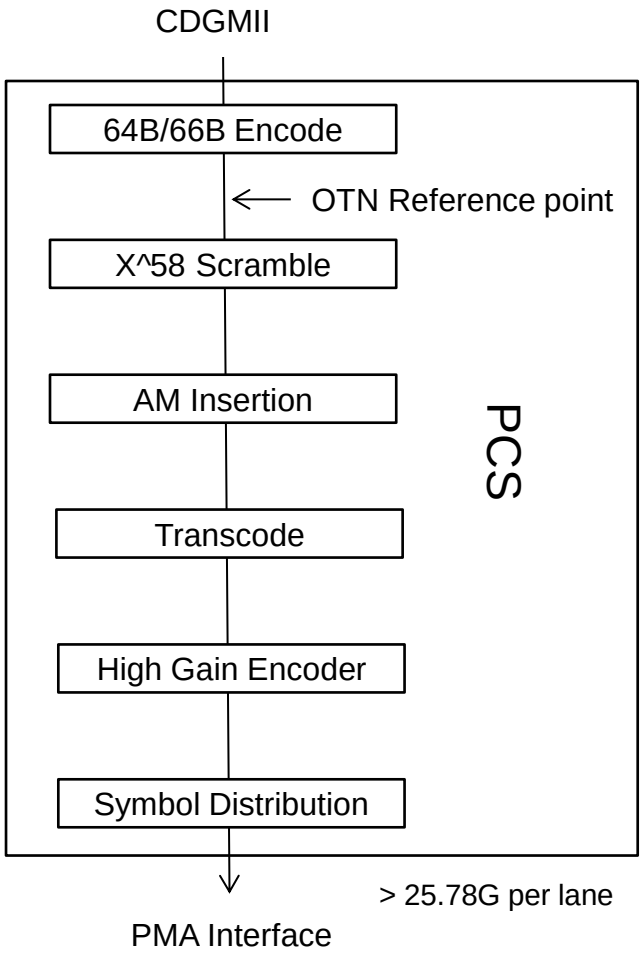
4x100G PAM4 Architecture – Option 2

- Likely implementation options, 16 lane CDAUI interfaces first
- PCS is in the big chip, with high gain FEC
- FEC is end to end
- Assuming a single FEC covers up to 5 interfaces



TX Data Flow - Option 2

- 64B/66B encode based on clause 82
- Can look at direct 256B/257B encode as an option
- High Gain FEC encoder, gain TBD
- 16 PMA lanes (similar to PCS/FEC lanes)
  - Muxed in some form for the PMD
  - PMD might need to add extra overhead



PMA Multiplexing

- What muxing should be done is dependent on the error models, FEC code chosen etc.

EEE

- It is assumed that the basis for the EEE implementation is based on 802.3bm, fast wake only

Conclusion

- Having end to end FEC will greatly simplify systems, but the right tradeoff between FEC gain and complexity/power needs to be made so that we can possibly include a single FEC in the large ASIC/FPGA/ASSP
- Other presentations at this meeting further explore FEC sizing concerns

Thanks!