800GbE PCS and PMA Baseline Proposals for 100 Gb/s per lane PHYs

(Draft)

Xinyuan Wang, Huawei Technologies

Matt Brown, Huawei Technologies

Supporters:

- To be added.

Introduction:

- A baseline for the 800 Gb/s PCS/PMA for 100 Gb/s per lane PHYs has not yet been adopted.
- Two PCS/PMA proposals have been presented to the task force.
- This presentation proposes to adopt a PCS/PMA sublayer equivalent to the 200GBASE-R PCS/PMA sublayer defined in Clause 119/120.

Proposal Overview

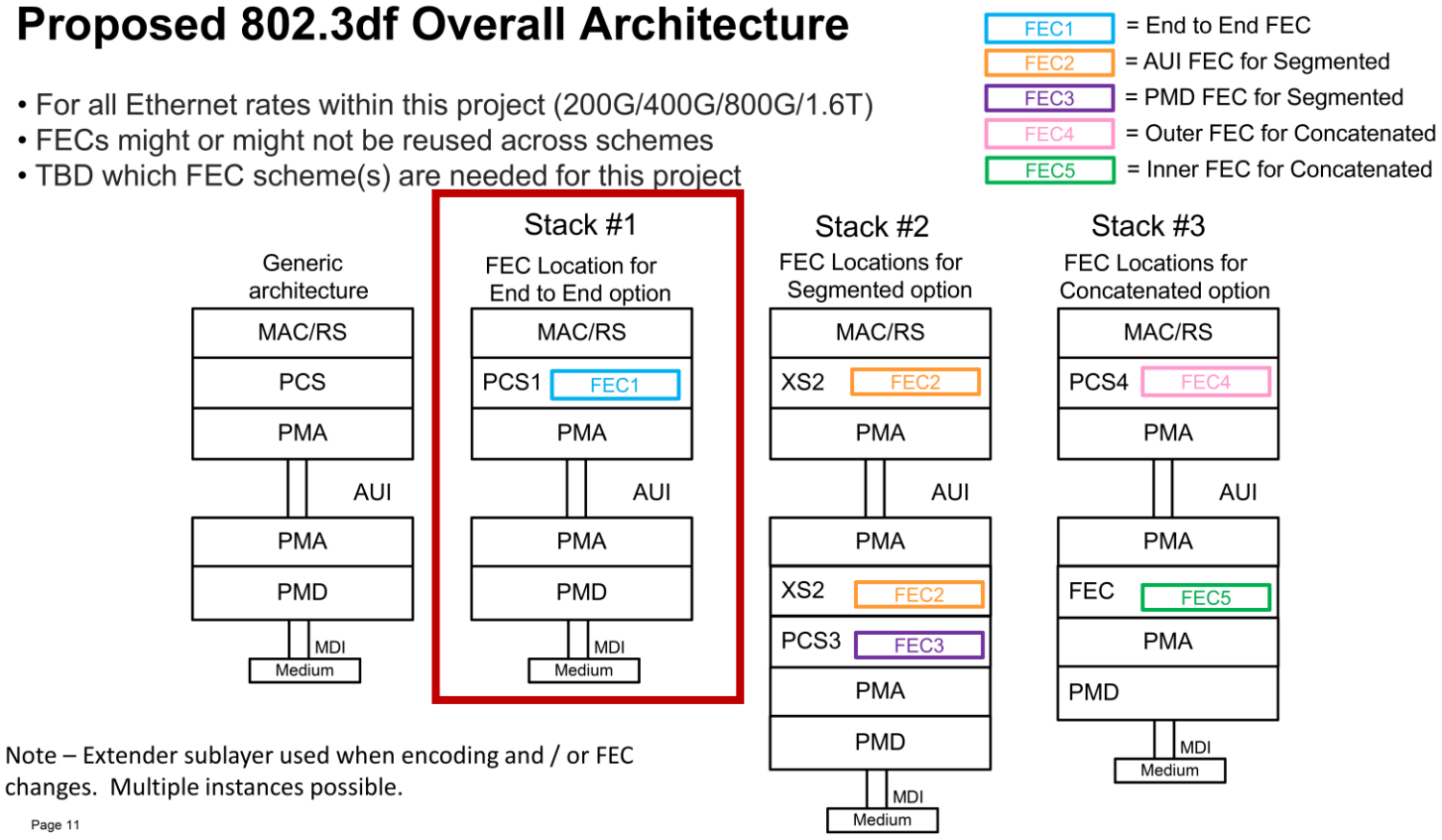
- The proposed PCS/FEC is summarized as follows:
  - PCS with FEC equivalent to the 200GBASE-R PCS defined in Clause 119.
  - Total data rate increased from 200 Gb/s to 800 Gb/s (increase 4X).
  - Eight PCS lanes at 106.25 Gb/s each (increase 4X).
  - Data 64B/66B encoded, 256B/257B encoded, scrambled.
  - Distributed to two RS(544,514) codewords and encoded.
  - Codewords interleaved then distributed to 8 PCS lanes.
  - No changes to alignment markers since rate will differentiate form 200GBASE-R lanes.
    - No analysis required.

References

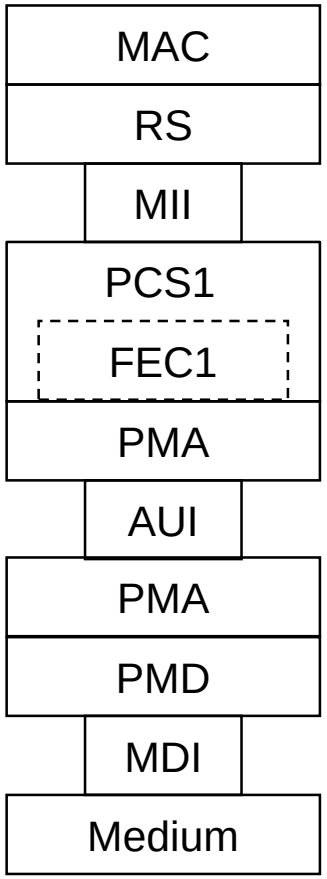
- Previous contributions in Task Force relating to the 800 Gb/s PCS/PMA baseline are listed here:
  - Option 1: 2X parallel Clause 119 (400GBASE-R PCS/PMA)
    - https://www.ieee802.org/3/df/public/22_03/shrikhande_3df_01_220329.pdf
    - https://www.ieee802.org/3/df/public/22_05/22_0517/shrikhande_3df_01a_220517.pdf
  - Option 2: Sped up Clause 119 (200GBASE-R PCS/PMA)
    - https://www.ieee802.org/3/df/public/22_03/wang_3df_01a_220308.pdf
    - https://www.ieee802.org/3/df/public/22_05/22_0517/he_3df_01_220517.pdf

References: (Cont'd)

- Adopted logic architecture baseline at May meeting
  - https://www.ieee802.org/3/df/public/22_05/motions_3df_2205_0524.pdf

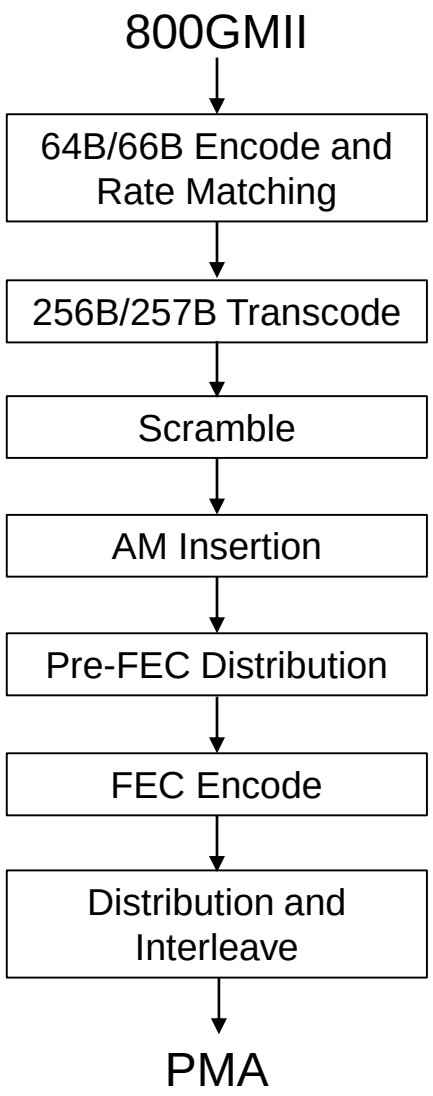


PCS and PMA in 800GbE Logic Architecture



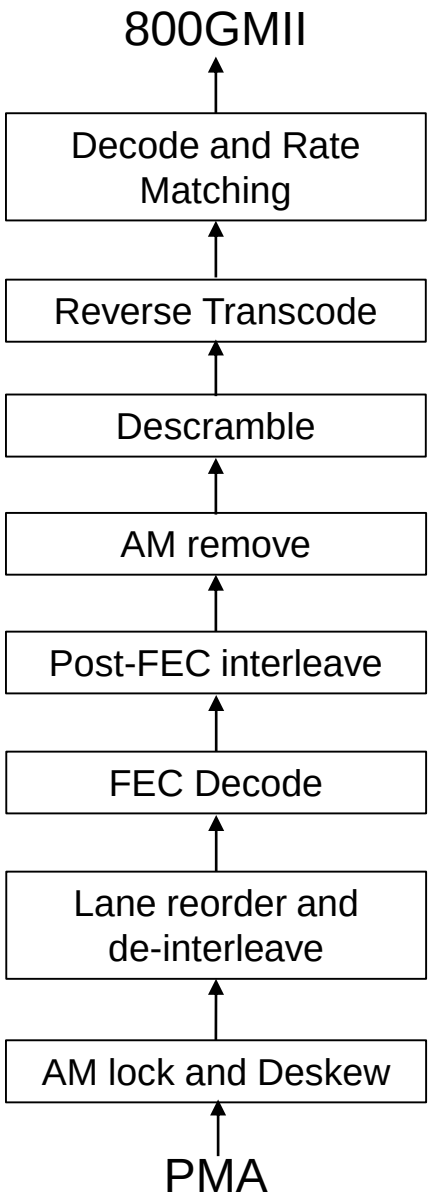
- The 800GBASE-R PCS are composed of Transmit and Receive processes, which shields the Reconciliation Sublayer (and MAC) from the specific nature of the underlying channel.
  - Communicating with the 800GMII: the PCS uses an eight octet-wide, synchronous data path, with frame delineation being provided by transmit control signals (TXC<n> = 1) and receive control signals (RXC<n> = 1).
- The PMA sublayer operates independently of block and frame boundaries.
  - The PCS provides the functions necessary to map frames between the 800GMII format and the PMA service interface format.

TX PCS Data Flow



- 64B/66B encode based on Clause 119.2.3/82.2.3.
- Transcode to 256B/257B based on Clause 119.2.4.2/91.5.2.5.
  - Allow direct encode from 64B/66B.
- Scramble based on Clause 119.2.4.3/82.2.5.
- FEC Encoder is RS(544,514,15,10) with 2-way interleave based on Clause 119.
  - All FEC processing is same as in Clause 119.2.4, including error correction and detection modes at RX.
- 8 PCS lanes @ 100 Gb/s.
- Support for any PCS lane on any physical lane.
- Compensation for any rate differences caused by the insertion or deletion of alignment markers or due to any rate difference between the 800GMII and PMA through the insertion or deletion of idle control characters.

RX PCS Data Flow



- Reverse of TX PCS data flow.
- Arbitrary PCS lanes order arrival from PMA.

Alignment Marker

- In order to support deskew and reordering of the individual PCS lanes at the RX PCS, alignment markers corresponding to PCS lanes are periodically inserted after being processed by the alignment marker mapping function.
- Refer to Clause 119.2.4.4.1, identical format as Figure 119-4 for 200GbE with 8 PCS lanes.

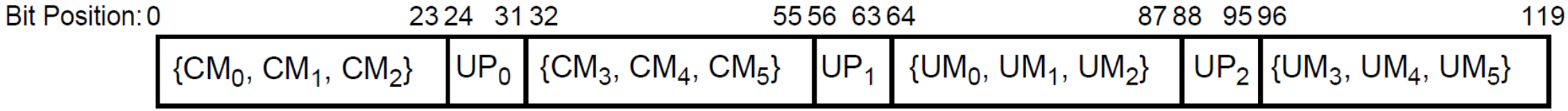


Figure 119–4—Alignment marker format

Alignment Marker Encoding

- Refer to Clause 119.2.4.4.1, identical encodings as Table 119-1 for 200GbE with 8 PCS lanes.

Table 119–1—200GBASE-R alignment marker encodings

PCS lane number	Encoding[a] $\{CM_0, CM_1, CM_2, UP_0, CM_3, CM_4, CM_5, UP_1, UM_0, UM_1, UM_2, UP_2, UM_3, UM_4, UM_5\}$
0	0x9A,0x4A,0x26,0x05,0x65,0xB5,0xD9,0xD6,0xB3,0xC0,0x8C,0x29,0x4C,0x3F,0x73
1	0x9A,0x4A,0x26,0x04,0x65,0xB5,0xD9,0x67,0x5A,0xDE,0x7E,0x98,0xA5,0x21,0x81
2	0x9A,0x4A,0x26,0x46,0x65,0xB5,0xD9,0xFE,0x3E,0xF3,0x56,0x01,0xC1,0x0C,0xA9
3	0x9A,0x4A,0x26,0x5A,0x65,0xB5,0xD9,0x84,0x86,0x80,0xD0,0x7B,0x79,0x7F,0x2F
4	0x9A,0x4A,0x26,0xE1,0x65,0xB5,0xD9,0x19,0x2A,0x51,0xF2,0xE6,0xD5,0xAE,0x0D
5	0x9A,0x4A,0x26,0xF2,0x65,0xB5,0xD9,0x4E,0x12,0x4F,0xD1,0xB1,0xED,0xB0,0x2E
6	0x9A,0x4A,0x26,0x3D,0x65,0xB5,0xD9,0xEE,0x42,0x9C,0xA1,0x11,0xBD,0x63,0x5E
7	0x9A,0x4A,0x26,0x22,0x65,0xB5,0xD9,0x32,0xD6,0x76,0x5B,0xCD,0x29,0x89,0xA4

[a] Each octet is transmitted LSB to MSB.

Alignment Marker Mapping into FEC Codewords and PCS Lanes

- Refer to Clause 119.2.4.4.1, identical mapping as Figure 119-5 for 200GbE with 8 PCS lanes.

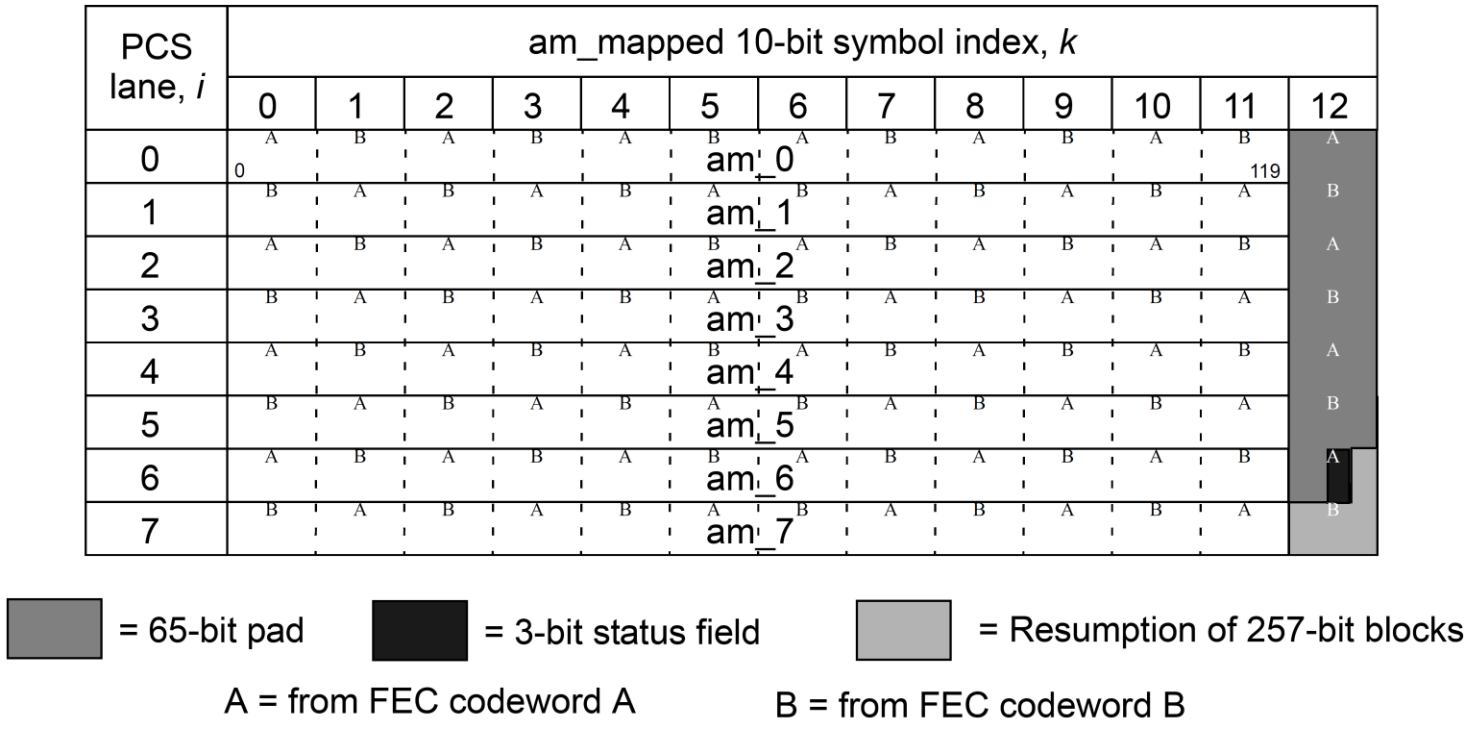
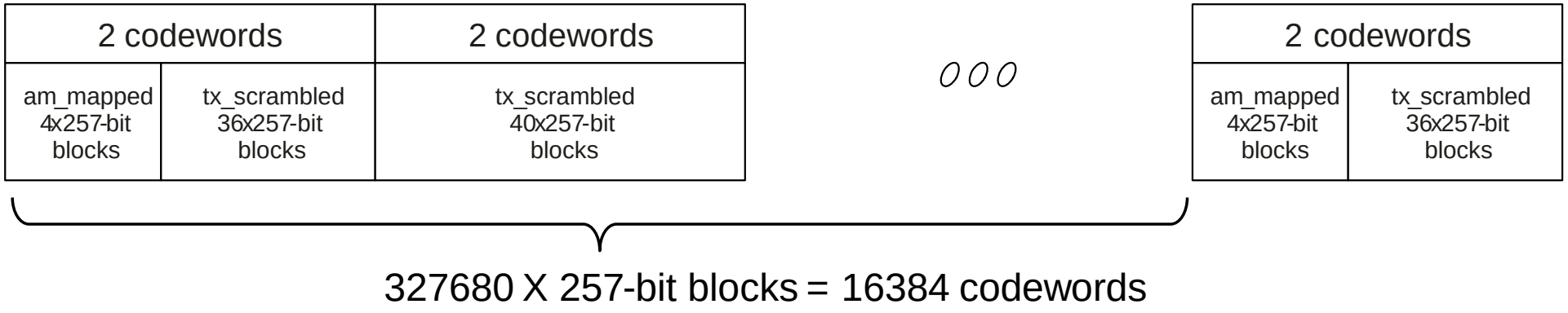


Figure 119–5—200GBASE-R alignment marker mapping to PCS lanes

Alignment Marker Insertion Period

- Refer to Clause 119.2.4.4.1/2, AMs are always aligned to the beginning of a RS FEC codeword, repetition distance is 16384 FEC codewords comparing to 8192 for 400GbE and 4096 for 200GbE in IEEE 802.3bs.



RX Process Function

- Refer to Clause 119.2.5
- Alignment lock and deskew:
  - The RX PCS forms 8 separate bit streams from PMA and obtains lock to the alignment markers as specified by the alignment marker lock state diagram shown in Figure 119–12.
  - After alignment marker lock is achieved on each of the 8 PCS lanes (bit streams), all inter-lane skew is removed as specified by the PCS synchronization state diagram shown in Figure 119–13.
- Reorder and De-interleave:
  - The RX PCS shall order the PCS lanes according to the PCS lane number. The PCS lane number is defined by the unique portion (UM0 to UM5) of the alignment marker that is mapped to each PCS lane.
  - After all PCS lanes are aligned, deskewed, and reordered, the two FEC codewords are de-interleaved to reconstruct the original stream of two FEC codewords.

RS FEC Decode Process

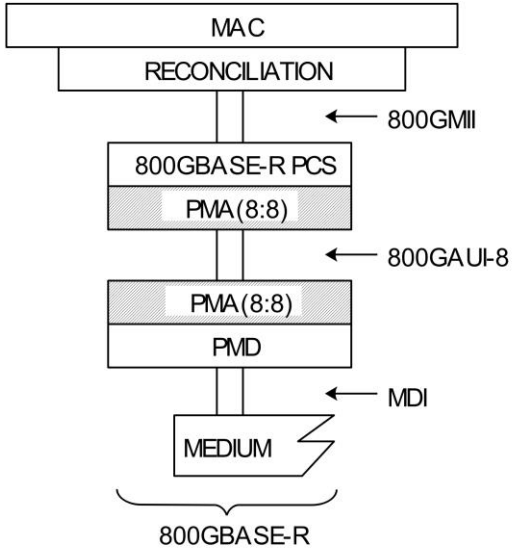
- RS FEC Decode:
  - Extracts the message symbols from the codeword, corrects them as necessary, and discards the parity symbols.
  - Capable of indicating when an errored codeword was not corrected.
  - When decoder determines that a codeword contains errors that were not corrected, it shall cause the RX PCS to set every 66-bit block within the two associated codewords to an error block (set to EBLOCK_R).
- Post-FEC Distribution:
  - After decode, data is interleaved on a 10-bit basis into rx_scrambled_am from two codewords corresponding to 40 transcoded blocks for RS(544,514) in order to recreate the transmitted data stream.
  - The first 1028 bits of rx_scrambled_am blocks is the vector am_rx<1027:0> where bit 0 is the first bit received.
  - The vector am_rx shall be removed from rx_scrambled_am to create rx_scrambled prior to descrambling.

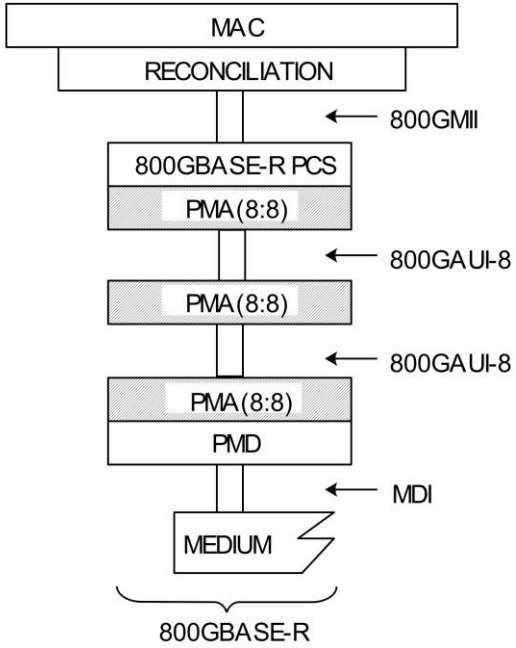
PMA Function

- Refer to Clause 120, both the transmit and receive directions:
  - Adapt the PCSL formatted signal to the appropriate number of abstract or physical lanes.
  - Provide per input-lane clock and data recovery.
  - Provide bit-level multiplexing.
  - Provide clock generation.
  - Provide signal drivers.
  - Optionally provide local loopback to/from the PMA service interface.
  - Optionally provide remote loopback to/from the PMD service interface.
  - Optionally provide test-pattern generation and detection.
  - Tolerate skew variation.
  - Perform PAM4 encoding and decoding for 800GBASE-R PMAs where the number of physical lanes is 8.

PMA Sublayer Positioning

- An implementation may use one or more PMA sublayers to adapt the number and rate of the PCS lanes to the number and rate of the PMD lanes.
- Potential forward looking compatible to Segmented, Concatenated FEC schemes.





PMA Multiplexing

- The PMA will support bit multiplexing, without regard to skew or PMA lane identity.
  - All skew is only handled in the RX PCS process as in Slide #14.
  - Refer to Clause 120.5.3 for skew variant
- The maximum cumulative delay contributed by up to four PMA stages in a PHY (sum of transmit and receive delays at one end of the link) shall meet the values specified in Table 120-1 for 92.16ns.

Standard Specification:

- New subclauses can explicitly refer to the existing subclauses in Clause 119/120.

Conclusion:

- This baseline proposes 800GbE PCS and PMA based on sped up of Clause 119 to support 8X100 Gb/s per lane AUIs, electrical and optical PMDs.

Thanks!