1.6TbE PCS/FEC Baseline Proposal

Mark Gustlin (Cisco), Kapil Shrikhande (Marvell), Eugene Opsasnick (Broadcom), Xinyuan Wang (Huawei), Xiang He (Huawei), Gary Nicholl (Cisco), David Ofelt (Juniper), Shawn Nicholl (AMD), Jeff Slavick (Broadcom), Adee Ran (Cisco), Kent Lusted (Intel), Jerry Pepper (Keysight), Matt Brown (Huawei)

February 6, 2023

IEEE P802.3dj Task Force interim meeting

Supporters

- Eric Maniloff, Ciena
- Tom Huber, Nokia
- Ben Jones, AMD
- Rick Rabinovich, Keysight
- Shimon Muller – Enfabrica
- Howard Heck – Intel
- Jeffery Maki – Juniper
- Arthur Marris – Cadence
- Ed Nakamoto – Spirent
- Dave Estes – Spirent
- Daniel Koehler - Synopsys
- Viet Tran, Keysight
- Rob Stone, Meta
- Alon Regev, Keysight
- Ted Sprague, Infinera
- Paul Brooks, Viavi
- Steve Gorshe, Microchip
- Mike Dudek, Marvell
- Ali Ghiasi, Ghiasi Quantum

Goals

- Enable efficient designs for 1.6TbE PCS/FEC/PMA
  - Maximizing common logic with other speeds is useful for designing port groups of many speeds
- Strong FEC performance
- Low latency
- Support both
  - 1.6TAUI-16
  - 1.6TAUI-8
- Focus on the FEC needs of the AUI(s)
  - RS(544,514,10) based
  - The FEC definition for the PMDs are not covered in this presentation
  - We do expect RS(544,514,10) to be part of the FEC structure for at least some of the PMDs

Architectural Direction

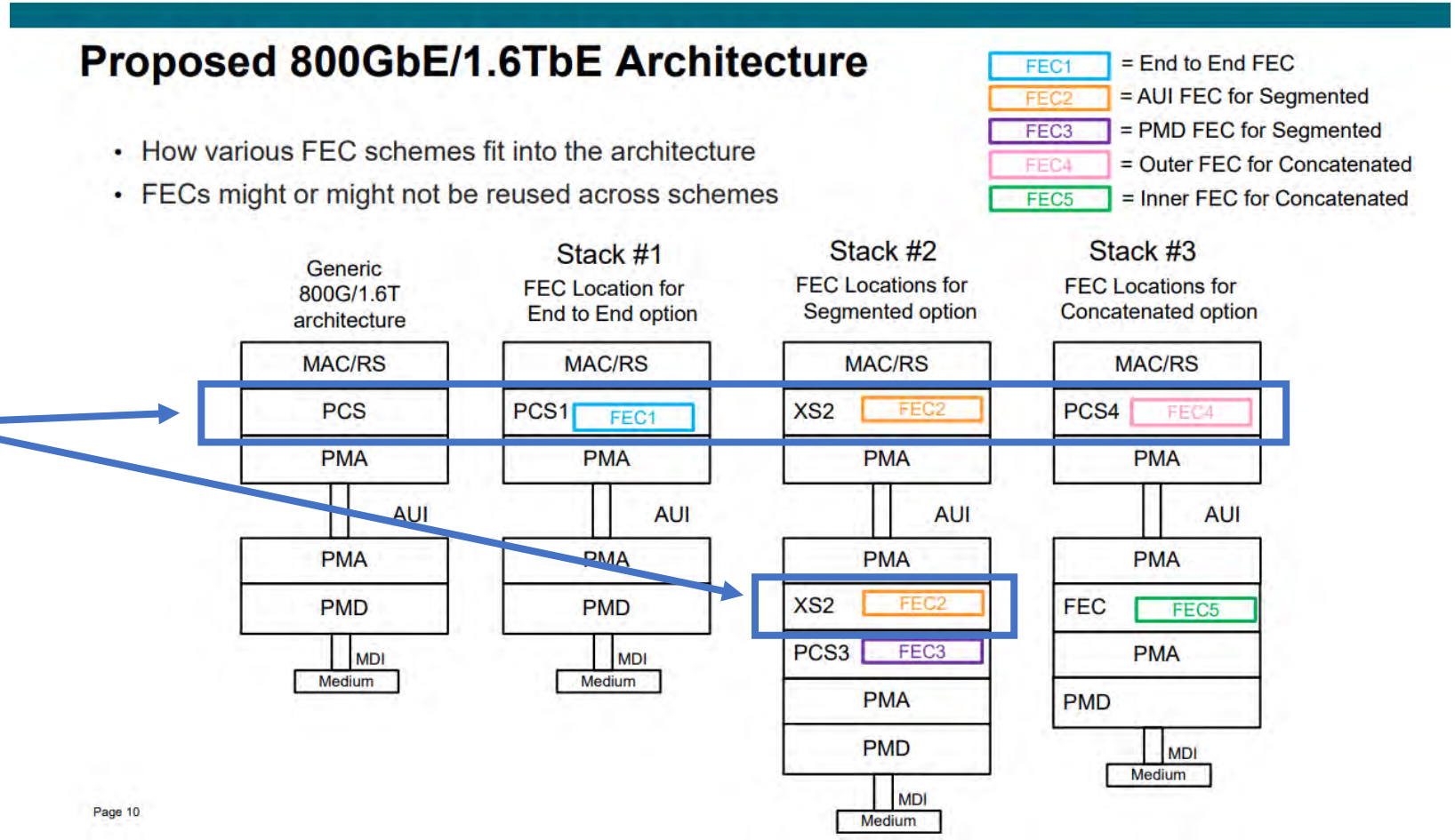
- Focus of this presentation is the data encoding/decoding, distribution, and FEC encoding/decoding
  - The method of FEC data distribution and interleave to the PMA output lanes requires further investigation

- 1.6TbE PCS/FEC leverages some aspects of the 800GbE PCS/FEC
  - Keeps two flows to simplify the scrambler implementation
  - Keeps 4 FEC engines/codewords for excellent burst error tolerance
  - RS(544,514) based FEC
- But changes the distribution to 1x257b (from 1x66b)
  - Provides a single OTN reference point

Fit Into the Adopted Architecture

- This proposal fits within the adopted 802.3dj logic architecture

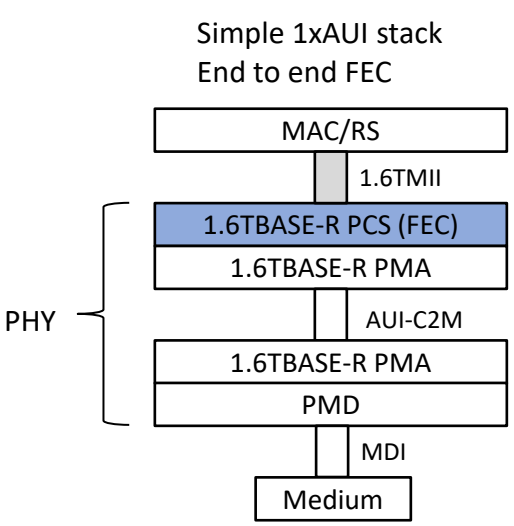
Focus of this presentation (for 1.6TbE)

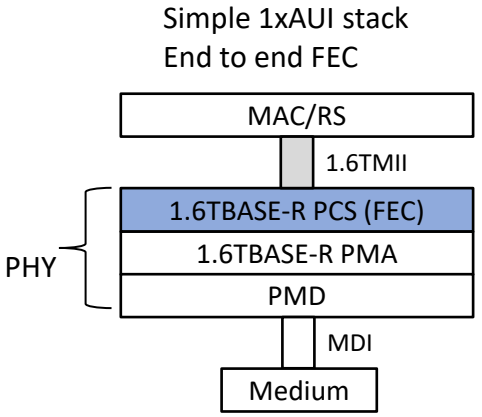


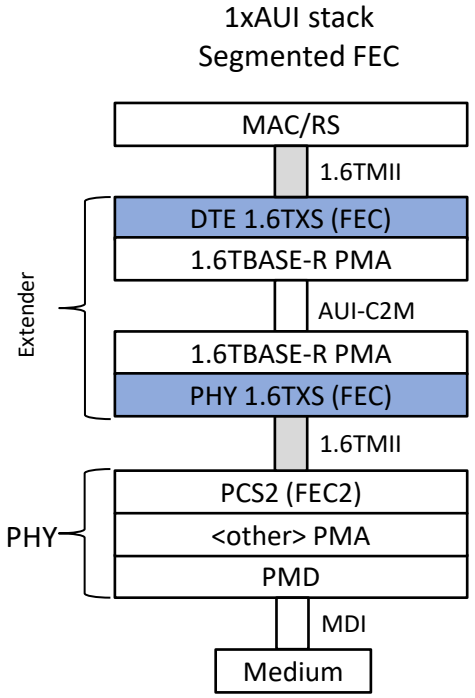
From: gustlin_3df_01a_220517

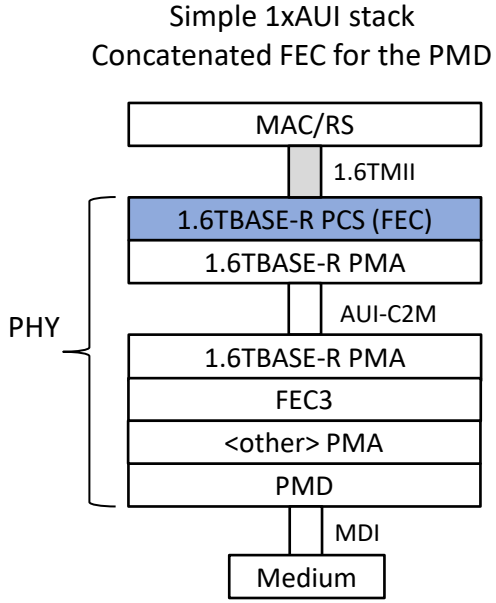
Details of the Fit Into the Adopted Architecture

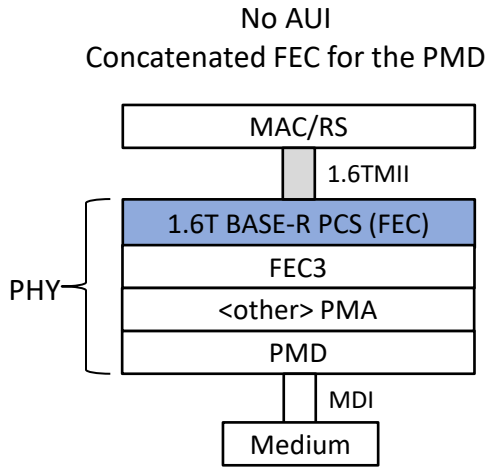
- The blue colored boxes are what is being defined in this presentation
  - Minus the method of FEC data distribution and interleave to the PMA output lanes



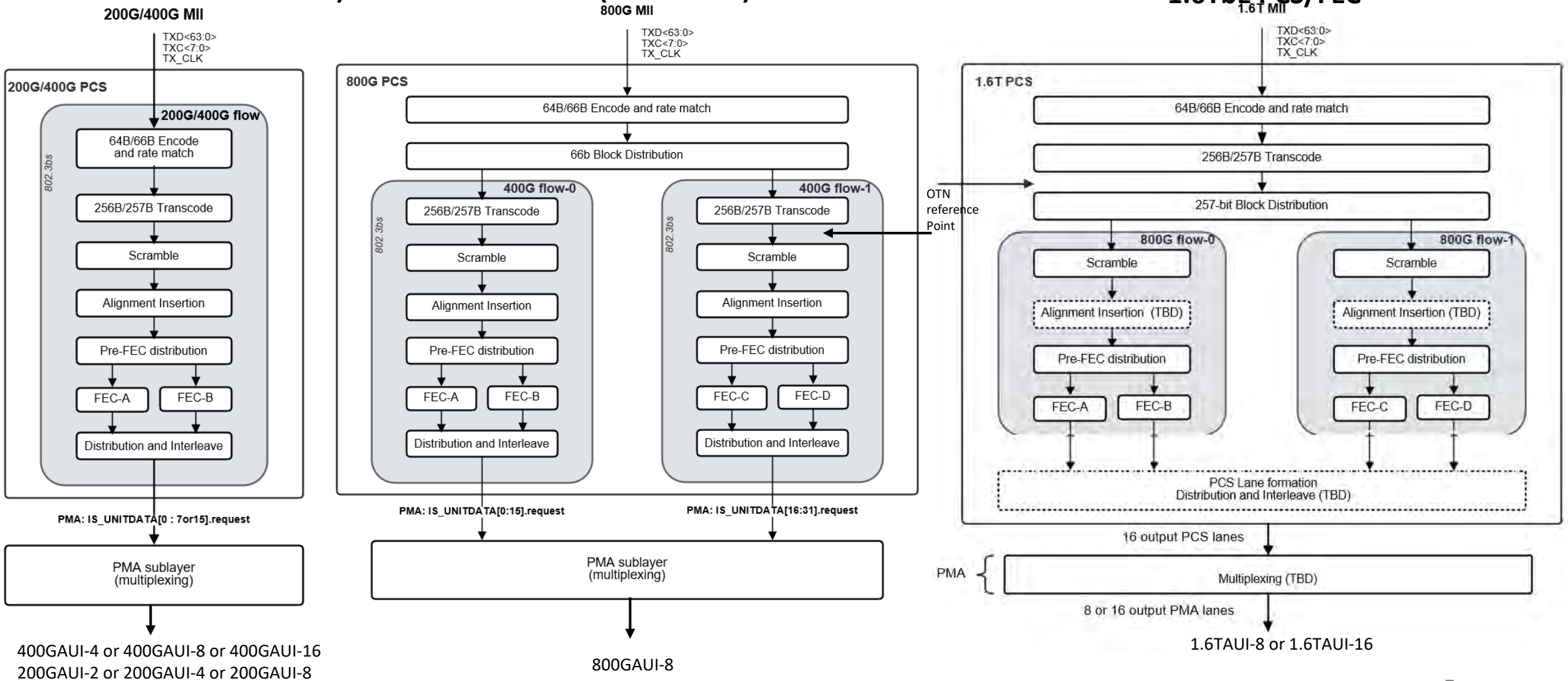






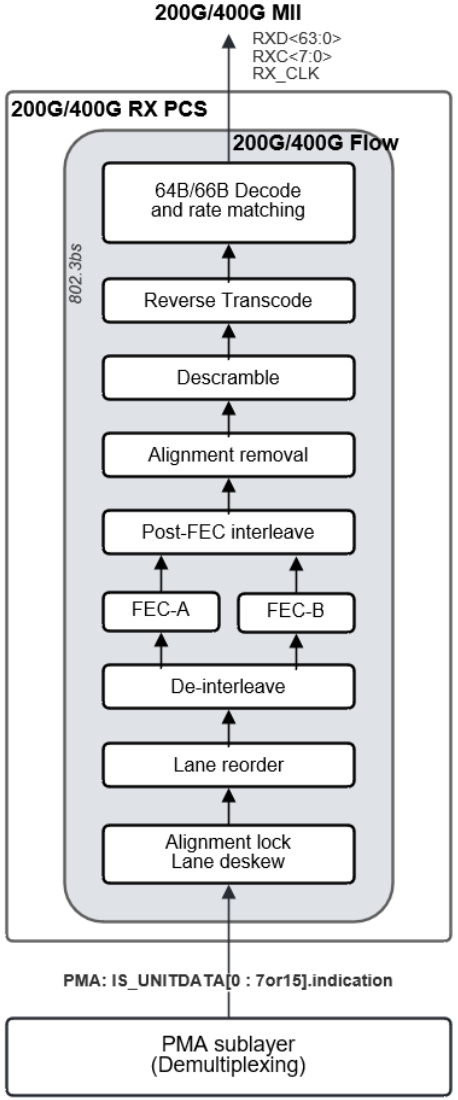


1.6TbE PCS/FEC Architecture Evolution : Tx Flow

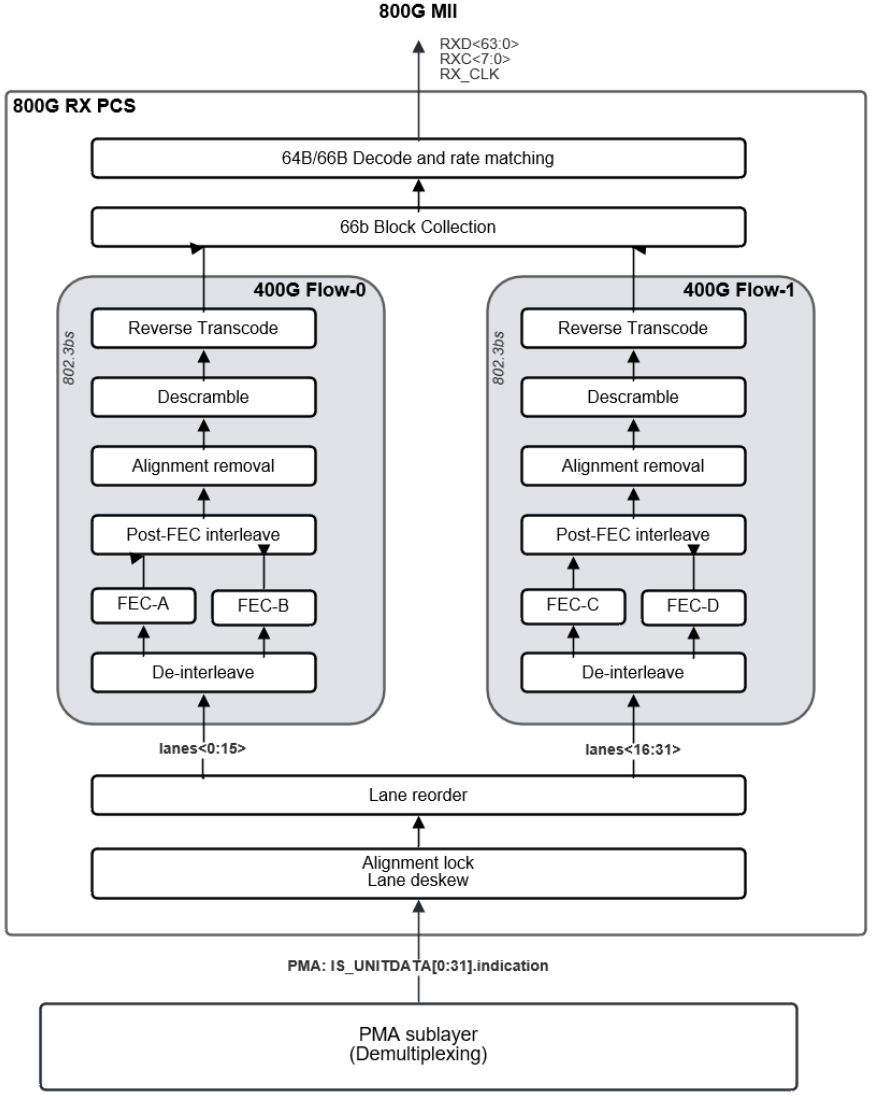


1.6TbE PCS/FEC Architecture Evolution : Rx Flow

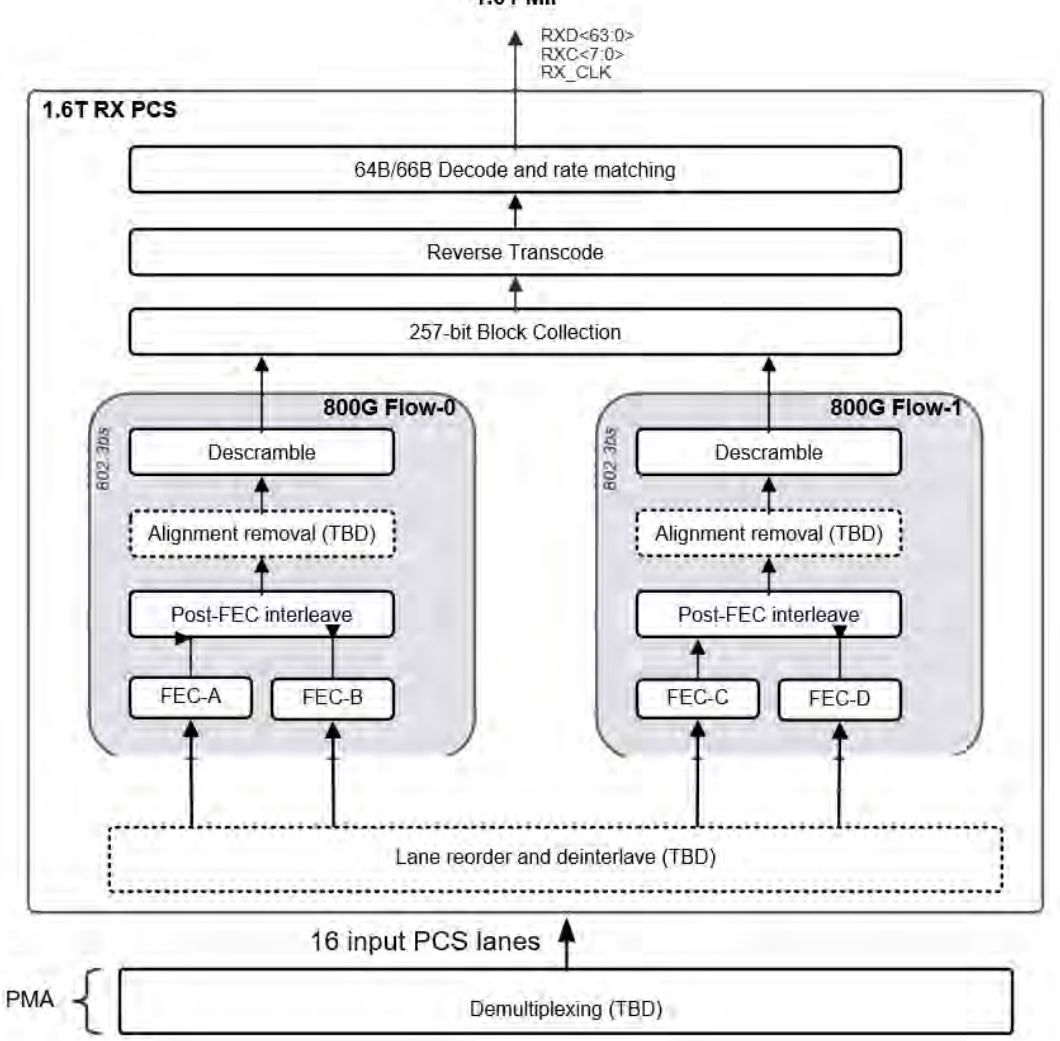
200/400GbE (CL 91 PCS)



800GbE (CL 172 PCS)

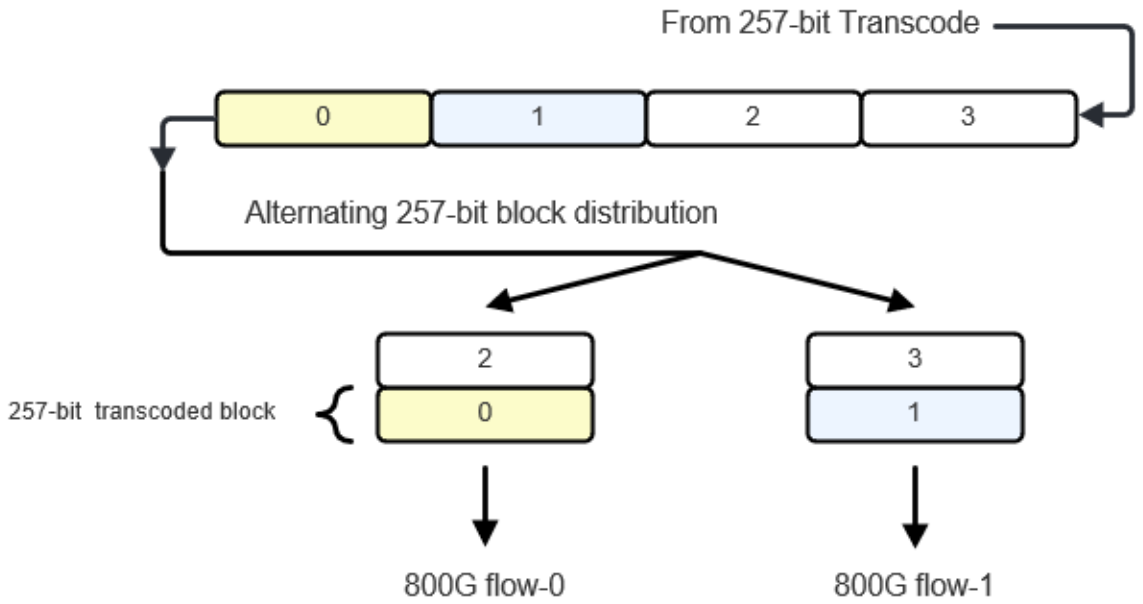


1.6TbE PCS/FEC

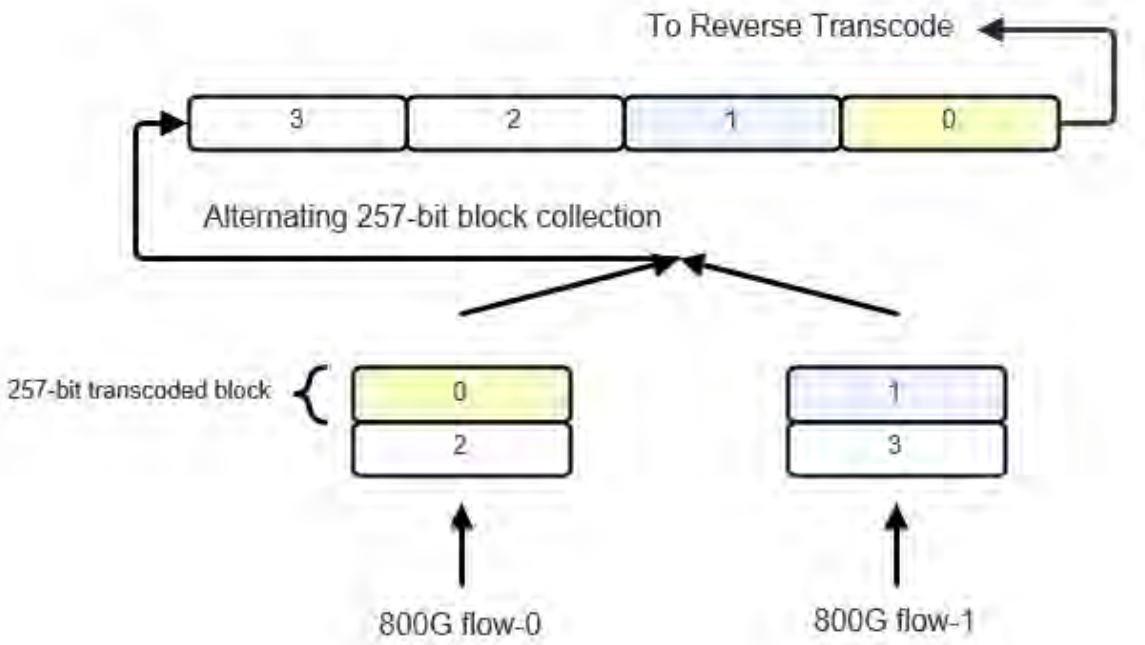


Tx 257b Block Distribution and Rx Block Collection

- Tx 257-bit Block Distribution



- Rx 257-bit Block Collection



Alignment Marker Insertion

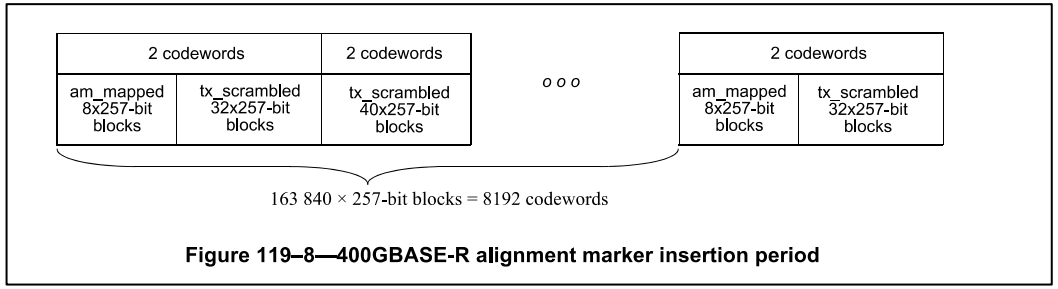
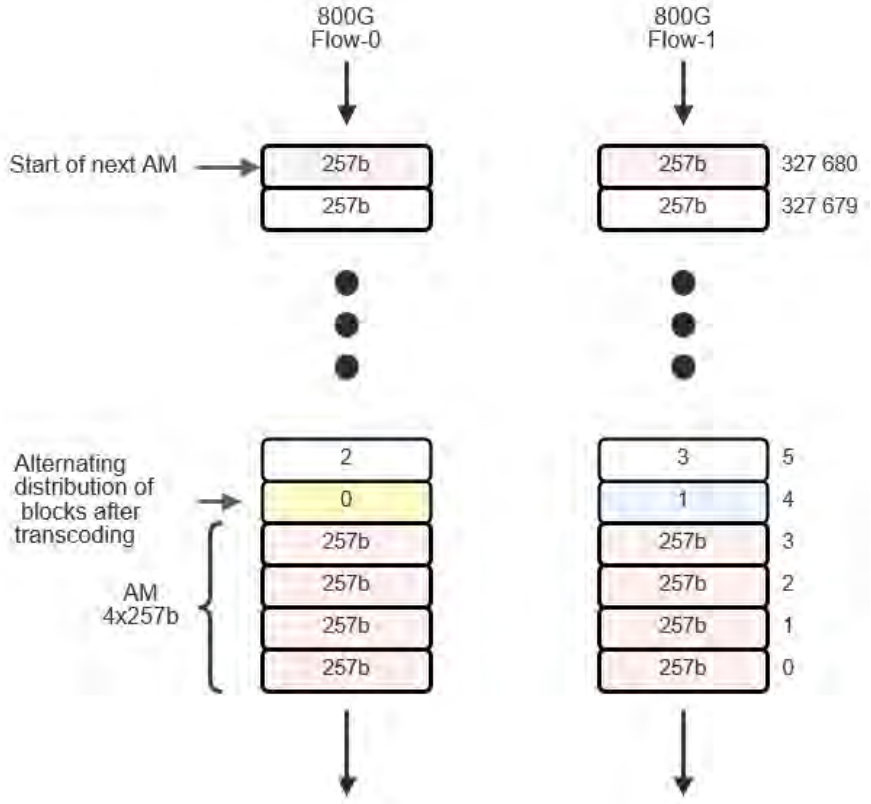


Figure 119–8—400GBASE-R alignment marker insertion period

Source: IEEE Std 802.3-2018

- Markers inserted at consecutive 257b blocks across both 800G flows
- Uses 16 PCSL
  - Total size of markers is same as 400GbE
  - Increased spacing between markers to maintain frequency of arrival.

GbE	#PCSL	AM group size (x 257b)	Spacing (in CWs)	#PCS flows	AM Spacing (in 257b per flow)
200	8	4	4k	1	81,920
400	16	8	8k	1	163,840
800	32	16	16k	2	163,840
1600	16	8	32k	2	327,680



Alignment Marker Encoding

- With 16 PCSL
- CM0-CM5 and UP0-UP2 are unchanged from 400GbE CL119
- UM0-UM5 are inverted from 400GbE
- Resulting UMs differ from 400GbE and 800GbE
- Clock Content and Baseline Wander Analysis – TBD
- UP and UM values can be adjusted if necessary
- Open issue: How to form the AMs in a coherent way so they appear correctly on physical lanes

PCS Lane #	Encoding														
	CM0	CM1	CM2	UP0	CM3	CM4	CM5	UP1	UM0	UM1	UM2	UP2	UM3	UM4	UM5
0	0x9A	0x4A	0x26	0xB6	0x65	0xB5	0xD9	0xD9	**0xFE**	**0x8E**	**0x0C**	0x26	**0x01**	**0x71**	**0xF3**
1	0x9A	0x4A	0x26	0x04	0x65	0xB5	0xD9	0x67	**0xA5**	**0x21**	**0x81**	0x98	**0x5A**	**0xDE**	**0x7E**
2	0x9A	0x4A	0x26	0x46	0x65	0xB5	0xD9	0xFE	**0xC1**	**0x0C**	**0xA9**	0x01	**0x3E**	**0xF3**	**0x56**
3	0x9A	0x4A	0x26	0x5A	0x65	0xB5	0xD9	0x84	**0x79**	**0x7F**	**0x2F**	0x7B	**0x86**	**0x80**	**0xD0**
4	0x9A	0x4A	0x26	0xE1	0x65	0xB5	0xD9	0x19	**0xD5**	**0xAE**	**0x0D**	0xE6	**0x2A**	**0x51**	**0xF2**
5	0x9A	0x4A	0x26	0xF2	0x65	0xB5	0xD9	0x4E	**0xED**	**0xB0**	**0x2E**	0xB1	**0x12**	**0x4F**	**0xD1**
6	0x9A	0x4A	0x26	0x3D	0x65	0xB5	0xD9	0xEE	**0xBD**	**0x63**	**0x5E**	0x11	**0x42**	**0x9C**	**0xA1**
7	0x9A	0x4A	0x26	0x22	0x65	0xB5	0xD9	0x32	**0x29**	**0x89**	**0xA4**	0xCD	**0xD6**	**0x76**	**0x5B**
8	0x9A	0x4A	0x26	0x60	0x65	0xB5	0xD9	0x9F	**0x1E**	**0x8C**	**0x8A**	0x60	**0xE1**	**0x73**	**0x75**
9	0x9A	0x4A	0x26	0x6B	0x65	0xB5	0xD9	0xA2	**0x8E**	**0x3B**	**0xC3**	0x5D	**0x71**	**0xC4**	**0x3C**
10	0x9A	0x4A	0x26	0xFA	0x65	0xB5	0xD9	0x04	**0x6A**	**0x14**	**0x27**	0xFB	**0x95**	**0xEB**	**0xD8**
11	0x9A	0x4A	0x26	0x6C	0x65	0xB5	0xD9	0x71	**0xDD**	**0x99**	**0xC7**	0x8E	**0x22**	**0x66**	**0x38**
12	0x9A	0x4A	0x26	0x18	0x65	0xB5	0xD9	0x5B	**0x5D**	**0x09**	**0x6A**	0xA4	**0xA2**	**0xF6**	**0x95**
13	0x9A	0x4A	0x26	0x14	0x65	0xB5	0xD9	0xCC	**0xCE**	**0x68**	**0x3C**	0x33	**0x31**	**0x97**	**0xC3**
14	0x9A	0x4A	0x26	0xD0	0x65	0xB5	0xD9	0xB1	**0x35**	**0x04**	**0x59**	0x4E	**0xCA**	**0xFB**	**0xA6**
15	0x9A	0x4A	0x26	0xB4	0x65	0xB5	0xD9	0x56	**0x59**	**0x45**	**0x86**	0xA9	**0xA6**	**0xBA**	**0x79**

Note: in table above, bolded text indicates inverted values from CL 119 AM values

Other PCS Aspects

- Leverages detailed state machines and processes from 800GbE
  - 64B/66B encoding/decoding is stateless or stateful
  - 256B/257B transcode and reverse transcode is identical
  - Scrambler/descrambler follows 800GbE
    - Diverse seeding of the two scramblers is required
  - Alignment insertion/removal details TBD
  - Pre FEC distribution/interleave follows 800GbE
  - FEC encode/decode is the same as 800GbE

Summary

- Suggested course of action
  - Adopt the top portion of the PCS for 1.6Tb/s as described in this presentation

- The method of forming the PCS lanes requires further investigation
  - How the AMs are formed is also open

Proposed Straw Poll

- I would support gustlin_3df_01_230206, slides 7-12, as the baseline for the 1.6TbE PCS/FEC
  - With the noted details (PCS lane forming and AM construction) to be determined later

  Y: N: A:

Thanks