FEC_I Sublayer Architecture Proposal for Type 2 PHYs

Xiang He, Hao Ren

Huawei Technologies

Contributors

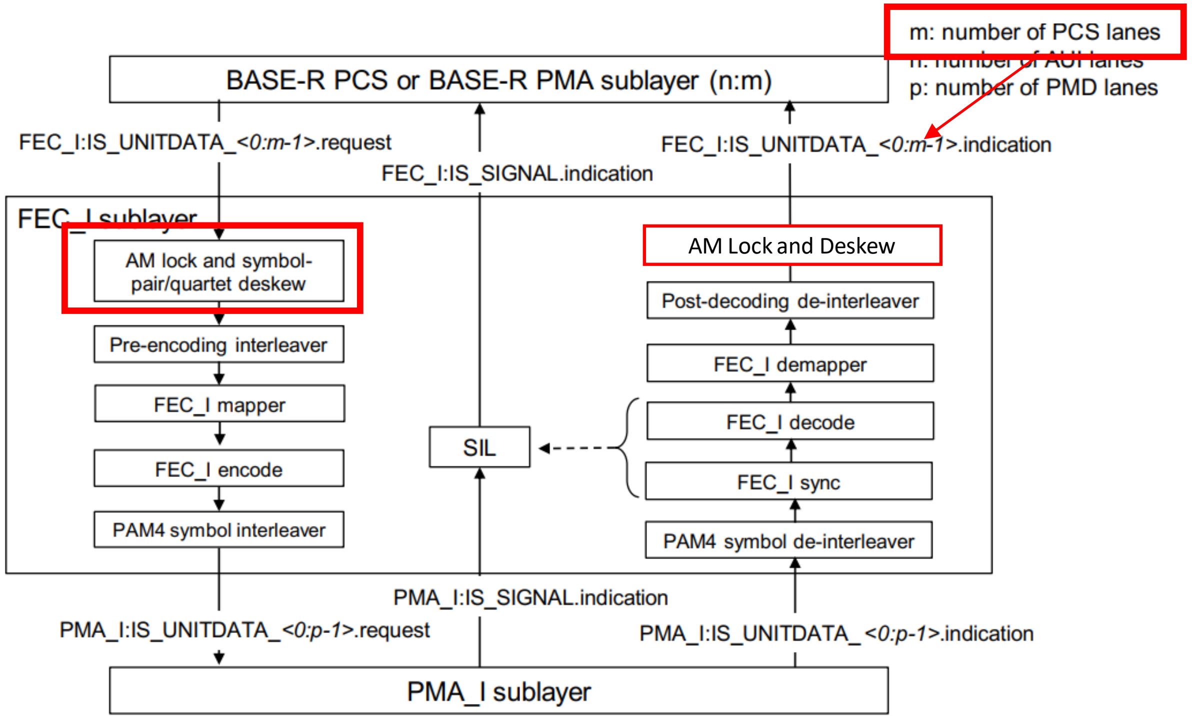
- Kechao Huang, Huawei
- Xiaoling Yang, Huawei

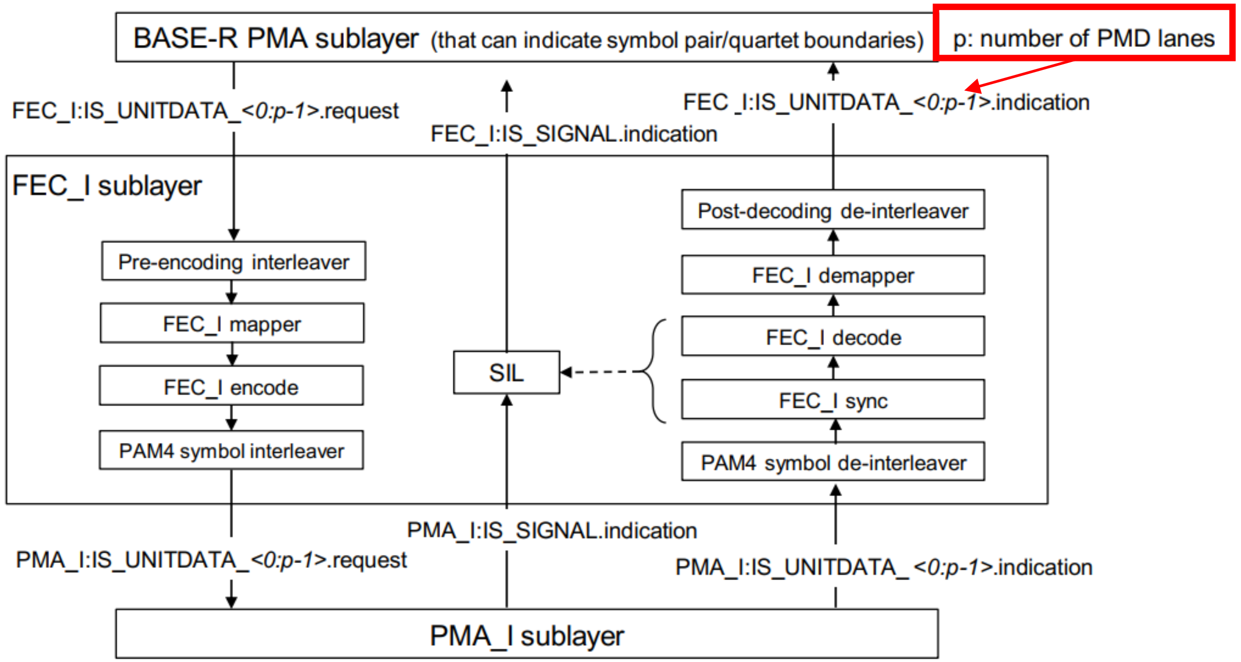
Introduction

- Type 2 PHY/FEC scheme has been adopted with Hamming(128,120) as the inner FEC.
  - See parthasarathy_3dj_02a_2303, and motions_3dfdj_2303.
  - Detailed design regarding convolutional interleaver and FEC lane rates are TBD.
- Symbol-pair muxing has been adopted for 200G/lane AUIs.
  - See ran_3dj_01a_2303, and motions_3dfdj_2303.
- Convolutional interleaver has been proposed to randomize errors from inner FEC.
  - See patra_3dj_01b_2303.pdf, huang_3df_01a_2211 and he_3dj_01a_230206.pdf.
  - Three different lane rates were proposed: 25G/lane, 100G/lane and 200G/lane.
  - The convolutional interleaver should be avoided for shorter PMDs due to high latency.
    - See he_3dj_02a_230206, dawe_3dj_01a_2303.
    - Latency impact has been analyzed in brown_3dj_optx_01b_230413 and brown_3dj_elec_01_230420.
- This presentation focuses on FEC_I lane rates, and recommend to use 200G/lane design.
  - Convolutional interleaver in the following slides could be excluded if not needed.

Architecture Overview

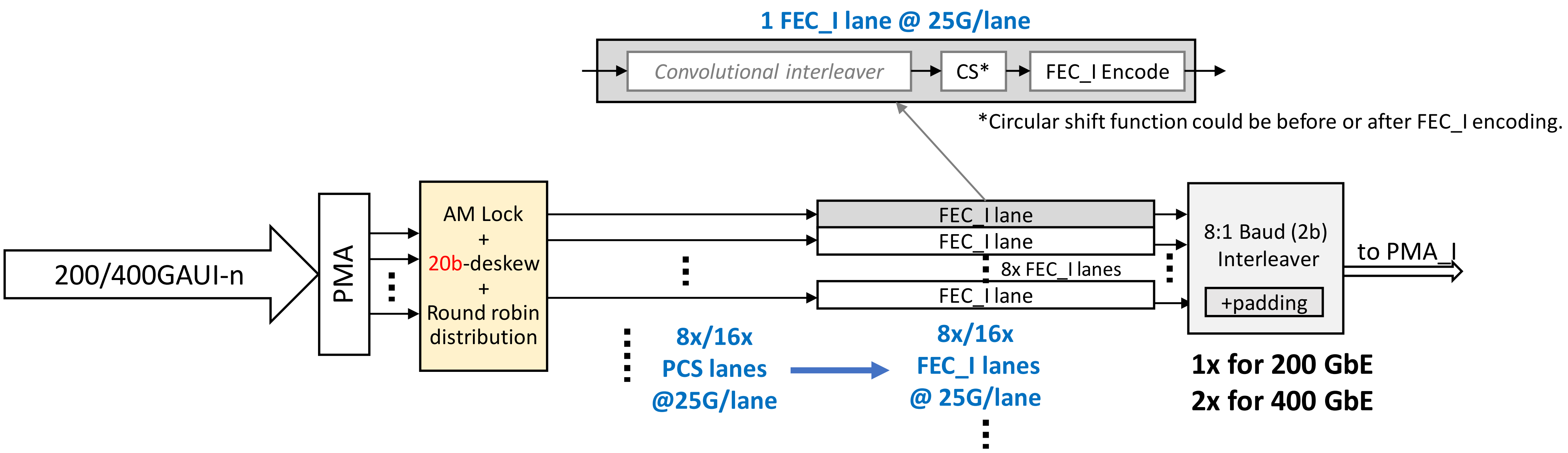
- The three FEC lane designs can be viewed as two main options, one has PCS lane based design, and the other has PMA/PMD lane based design.
  - Both options have exactly the same performance in terms of FEC gain.
  - Both options have ~~the same~~ similar number of bits storage for convolutional interleaver if used.
    - Using 25G/lane + 100G/lane design will result in 5.5% more bit storage than unified 200G/lane design.





25G/lane Design – 200 GbE and 400 GbE ~~and 800 GbE~~

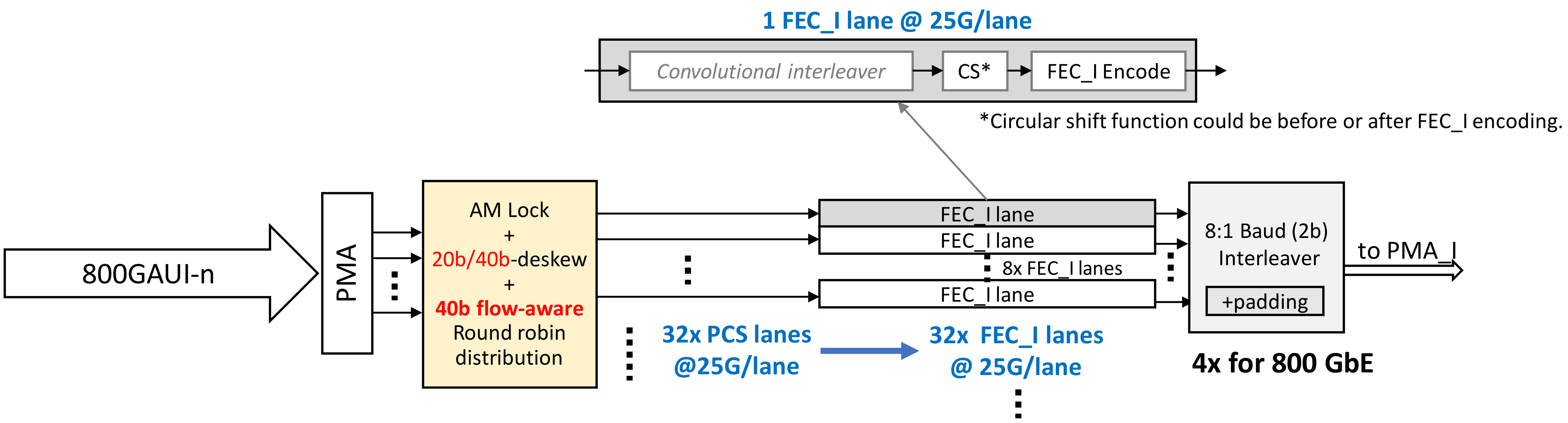
- Data from AUI is de-muxed to PCS lanes first, and deskewed to RS symbol-pair (20b) boundary.
- FEC_I sublayer is based on 25G/lane, as proposed in patra_3dj_02_2305.
  - For 200 GbE and 400 GbE, PCS lane rate is 25G/lane, and each has its own FEC_I lane.
  - 200 GbE and 400 GbE has 2x RS codewords, so deskew and round-robin distribution to each FEC_I lane is based on 20b.
- 8-lane per 200G PMD lane naturally supports 8:1 baud interleaver.



*Highlighted boxes are rate-specific functions.

25G/lane Design – 800 GbE

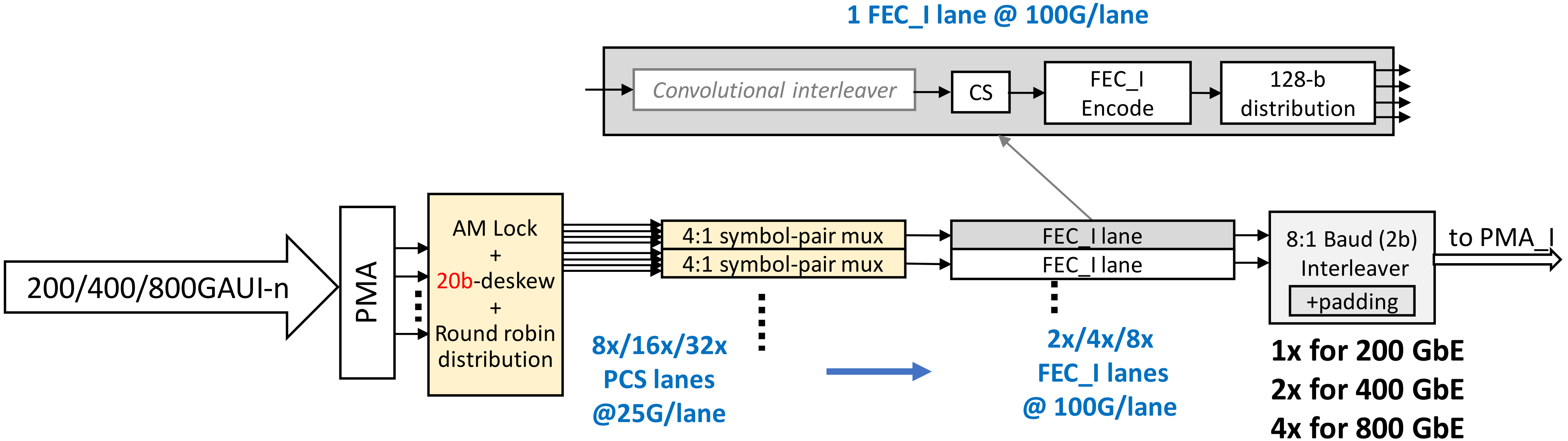
- Data from AUI is de-muxed to PCS lanes first, and deskewed to RS symbol-pair (20b) boundary.
- FEC_I sublayer is based on 25G/lane, as proposed in patra_3dj_02_2305.
    - For 800 GbE, PCS lane rate is 25G/lane, but FEC_I lane is not designed based on each PCS lane.
    - Round-robin distribution to each FEC_I lane is based on **40b** blocks, and has to maintain the 2-flow restrictions.
- 8-lane per 200G PMD lane naturally supports 8:1 baud interleaver.



*Highlighted boxes are rate-specific functions.

100G/lane Design – 200 GbE, 400 GbE and 800 GbE

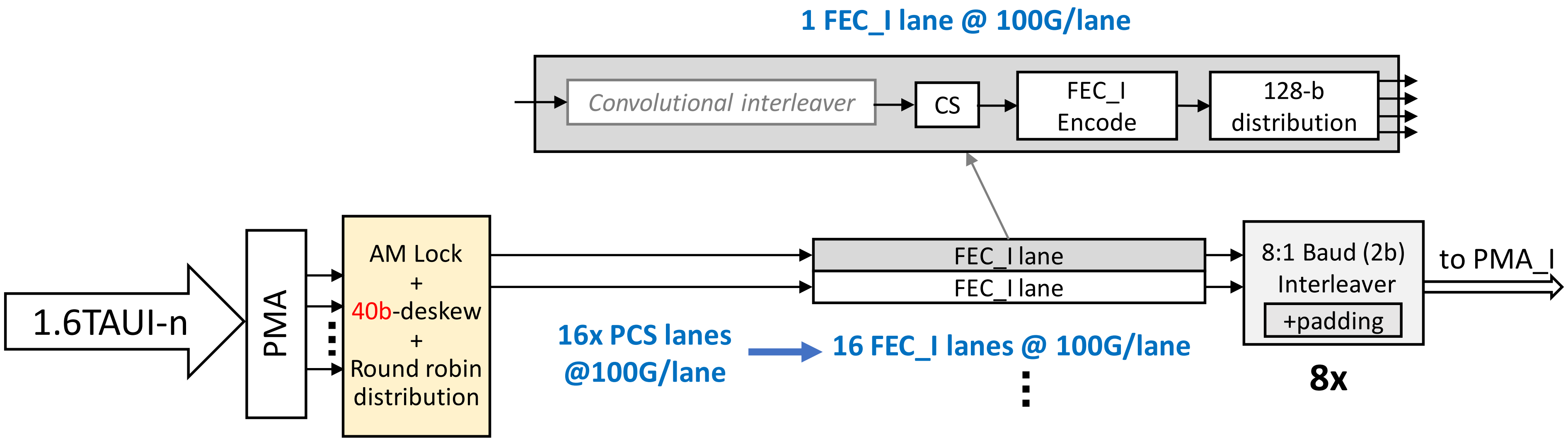
- Data from AUI is de-muxed to PCS lanes first, and deskewed to RS symbol-pair (20b) boundary.
- FEC_I sublayer is based on 100G/lane.
  - For 200/400/800 GbE, a 4:1 symbol-pair mux is needed for each FEC_I lane, and it need to be "flow aware".
- 100G/lane design could also support 8:1 (or higher ratio) post-encoding baud interleaver.
  - Circular shift function could be before or after FEC_I encoding.
  - Using single lane "CS + FEC_I encode", aligned with patra_3dj_optx_01_230427.
  - "CS" logic is different between the two FEC_I lanes within a 8:1 baud-interleaved group.



*Highlighted boxes are rate-specific functions.

100G/lane Design – 1.6 TbE

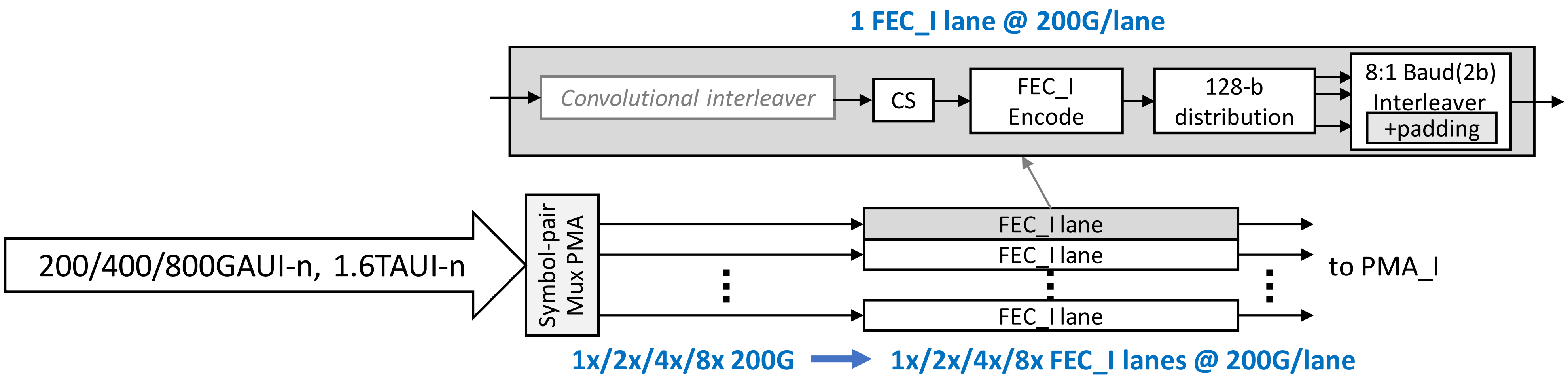
- Data from AUI is de-muxed to PCS lanes first, and deskewed to RS symbol-quartet (40b) boundary.
- FEC_I sublayer is based on 100G/lane.
  - For 1.6 TbE, each FEC_I lane takes the 100G/lane PCS input directly.
- 100G/lane design could also support 8:1 (or higher ratio) post-encoding baud interleaver.
  - Circular shift function could be before or after FEC_I encoding.
  - Using single lane "CS + FEC_I encode", aligned with patra_3dj_optx_01_230427.
  - "CS" logic is different between the two FEC_I lanes within a 8:1 baud-interleaved group.



*Highlighted boxes are rate-specific functions.

200G/lane Design – 200 GbE, 400 GbE, 800 GbE and 1.6TbE

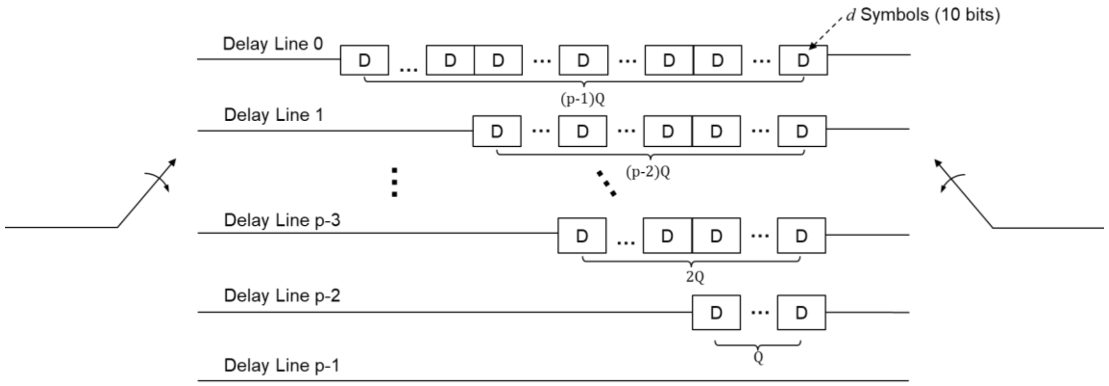
- Data from AUI is NOT required to be de-muxed to PCS lanes first.
  - Relying on the symbol-pair muxing PMA functions to establish FEC_I lane mapping.
  - The PMA cannot be an n:n retiming PMA because convolutional interleavers require to know the RS symbol boundaries.
- Maximizing the common functional blocks across different rates.
- 200G/lane design could also support 8:1 (or higher ratio) post-encoding baud interleaver.
  - Circular shift function could be before or after FEC_I encoding.



***No rate-specific functions.**

200G/lane Convolutional Interleaver Design

- 200G/lane FEC_I design could utilize the same design blocks across different rates.
- 200G/lane convolutional interleaver design has lower latency for most cases.
- For latency sensitive applications, convolutional interleavers should not be used.



Proposal	PCS	d (RS symbol)	P	Q	Depth	Latency ns	FEC_I Lane Rate
patra_3dj_optx_01_230427	1.6TE	4	3	11	12x RS	24.85	100G/lane
	800GE	4	3	6	12x RS	54.21	25G/lane
	400GE	2	6	6	12x RS	135.53	
	200GE	2	6	12	12x RS	271.06	
he_3dj_optx_01_230427	1.6TE	4	3	23	12x RS	25.98	200G/lane
	800GE	4	3	45	12x RS	50.82	
	400GE	2	6	46	12x RS	129.88	
	200GE	2	6	91	12x RS	256.94	

Hardware Resource Comparison

Proposal	Lane Rate		Convo. Interleaver Area	PMA Demux + Distribution Area	Hamming Encoding + Decoding Area
	200/400/800GE	1.6TE			
patra_3dj_02_2305	25G/lane	100G/lane	**1.5xA**	~2xB	**C**
barakatain_3dj_01a_2303 & he_3dj_01_2303	100G/lane		**A**	B	**C**
he_3dj_01_2305	200G/lane		**A**	None	**C**

Notes:

- A is tightly related to FEC_I lane rate selection, and will increase if more flavors if interleaver is used.
- C is algorithm dependent, and is slightly smaller than A for typical implementation like Chase-II.
- B is negligible compared to A or C.

- Overall patra_3dj will **consume 28% more chip area** (gate counts) than he_3dj for the whole inner FEC Tx + Rx.

Summary

- The FEC_I lane rate does not affect the FEC performance.
  - However it needs to be defined clearly to ensure interop.
- If using different FEC_I lane rate for different rates, the module cannot support rate agnostic breakout.
- FEC_I sublayer designing based on 200G/lane PMA lanes enables unified design across all rates that supports 200G/lane optical PMDs.
  - With a single FEC_I sublayer defined, it could be used to define all Ethernet rates using 200G/lane optics.
  - When the convolutional interleaver is bypassed, it could support rate agnostic breakout.

Thank you!