FEC baseline proposal for 200Gb/s per Lane IM-DD Optical PMDs

Co-Authors:

Lenin Patra, Marvell
Arash Farhood, Marvell
Rajesh Radhamohan, Cisco
Will Bliss, Broadcom
Sridhar Ramesh, Maxlinear
Dave Cassan, Alphawave

Supporters list:

- Ali A. (Reza) Eftekhar, Amazon
- Alex Levin, Microsoft
- Rob Stone, Meta
- Gary Nicholl, Cisco
- Brian Welch, Cisco
- Scott Schube, Intel
- Maxim Kuschnerov, Huawei
- Masoud Barakatain, Huawei
- Kechao Huang, Huawei
- Dirk Lutz, Eoptolink
- K.T. Tsai, Molex
- Vipul Bhatt, Coherent
- Roberto Rodes, Coherent
- Frank Chang, Source Photonics
- Marek Tlalka, Macom
- Kohichi Tamura, Cig
- Chris Cole, Quintessent Inc.
- Atul Srivastava, NEL
- Chris Doerr, Aloe Semiconductor
- Ali Ghiasi, Ghiasi Quantum
- Vasu Parthasarathy, Broadcom
- Tony Chan Carusone, Alphawave
- Jamal Riani, Marvell
- Chungjue Chen, Broadcom
- Henry Wong, Alphawave
- Drew Guckenberger, Maxlinear
- Kishore Kota, Marvell
- Kapil Shrikhande, Marvell
- Rich Melitz, Samtec
- Samuel Kocsis, Amphenol
- Mark Kimber, Semtech

Introduction

This presentation consists of a proposal for the FEC scheme for the following PMDs:

FEC Approach	PMDs
RS(544,514) based	800GBASE-DR4, 800GBASE-DR4-2 400GBASE-DR2, 400GBASE-DR2-2* 200GBASE-DR1, 200GBASE-FR1
RS(544,514) + Hamming (128,120) based	800GBASE-FR4

Proposal is consistent with the adopted P802.3dj architecture: gustlin_3df_01a_220517.pdf

* If objective adopted

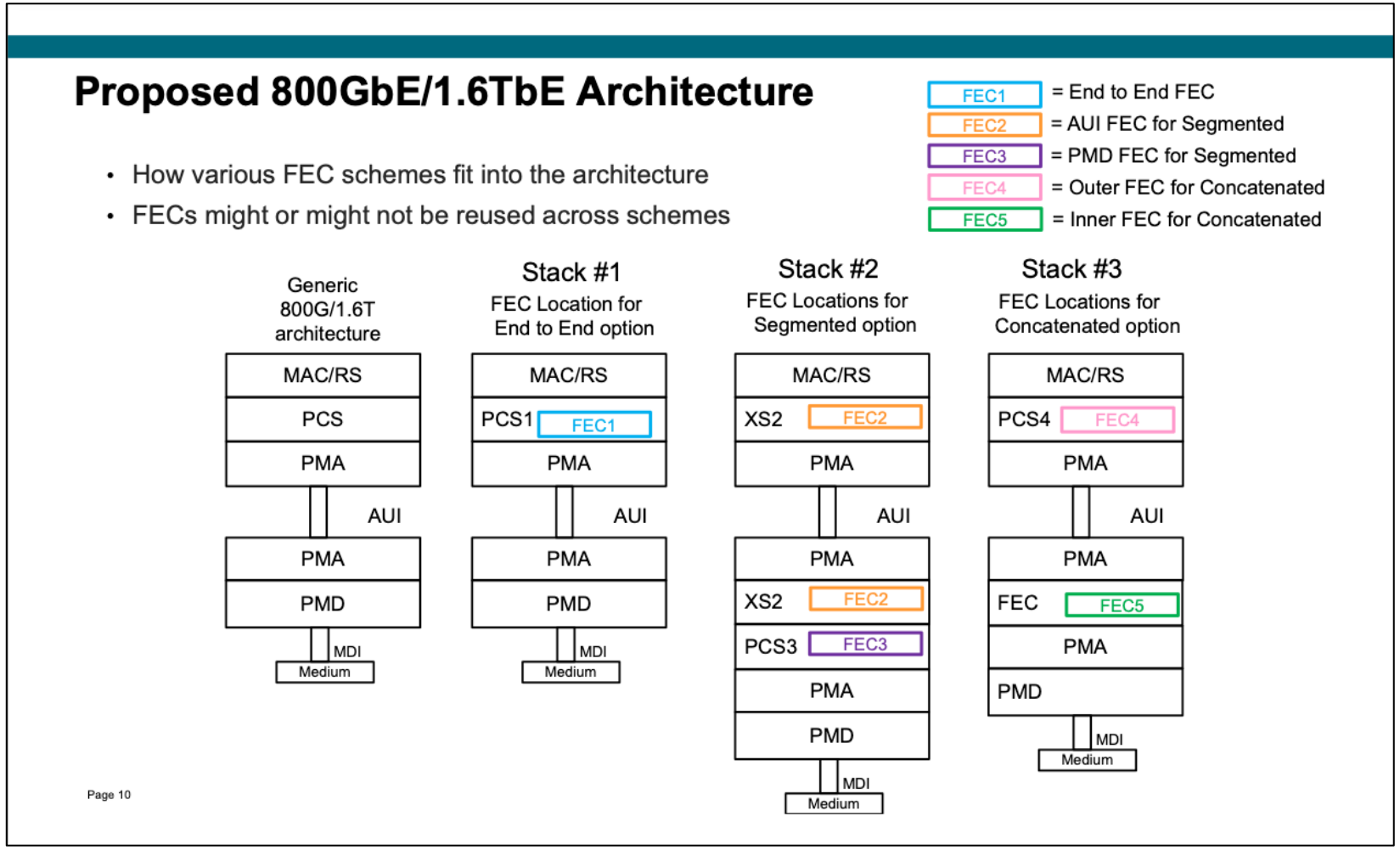
Culmination of past contributions for DR & FR link segments

- January 2023 baseline proposal for Inner FEC for FR4: https://www.ieee802.org/3/dj/public/23_01/23_0206/farhood_3dj_01a_230206.pdf
- AUI BER trade off for Inner code FEC study: https://www.ieee802.org/3/df/public/22_11/farhood_3df_01_2211.pdf
- January Proposal for DR4 mode: https://www.ieee802.org/3/dj/public/23_01/23_0206/welch_3dj_01a_230206.pdf
- A case for Concatenated code :https://www.ieee802.org/3/dj/public/23_03/parthasarathy_3dj_01_2303.pdf

Other supporting presentation:

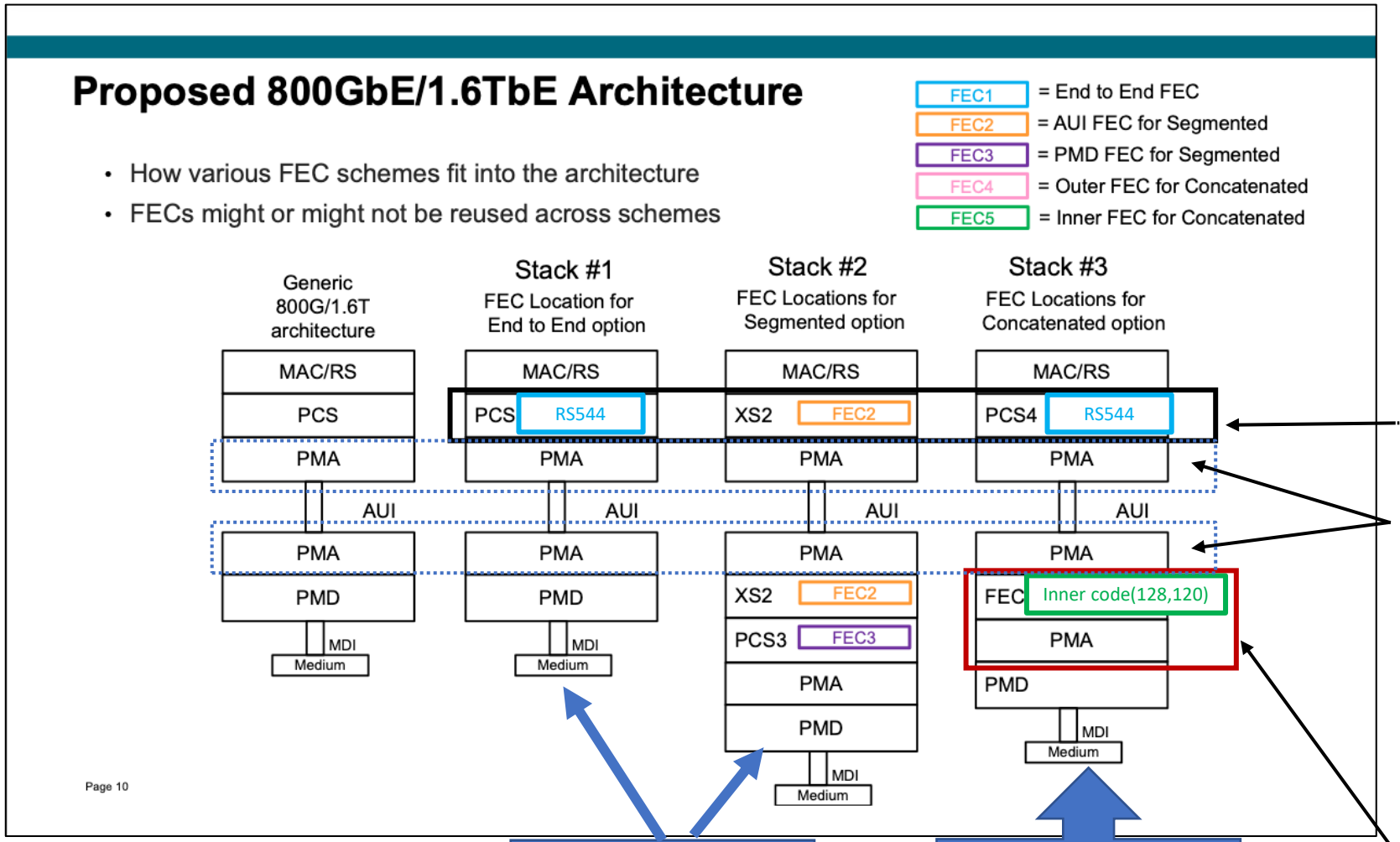
- https://www.ieee802.org/3/df/public/22_07/patra_3df_01a_2207.pdf
- https://www.ieee802.org/3/df/public/22_05/22_0517/bliss_3df_01a_220517.pdf
- https://www.ieee802.org/3/df/public/22_02/welch_3df_02a_220222.pdf
- https://www.ieee802.org/3/df/public/22_03/welch_3df_01a_220315.pdf
- https://www.ieee802.org/3/df/public/22_05/22_0602/welch_3df_01b_220602.pdf
- https://www.ieee802.org/3/df/public/22_07/bernier_3df_01_2207.pdf
- https://www.ieee802.org/3/df/public/22_10/22_1005/simms_3df_01_221005.pdf
- https://www.ieee802.org/3/df/public/22_10/22_1011/welch_3df_01a_221011.pdf
- https://www.ieee802.org/3/df/public/22_10/22_1011/ingham_3df_01_221011.pdf
- https://www.ieee802.org/3/df/public/22_11/lu_3df_01a_2211.pdf
- https://www.ieee802.org/3/dj/public/23_01/23_0206/li_3dj_01a_230206.pdf
- https://www.ieee802.org/3/dj/public/23_01/23_0206/li_3dj_01a_23020 6.pdf
- https://www.ieee802.org/3/df/public/22_10/22_1005/bliss_3df_01_220929.pdf
- https://www.ieee802.org/3/df/public/22_11/huang_3df_01a_2211.pdf
- https://www.ieee802.org/3/df/public/22_11/he_3df_01_2211.pdf

Adopted logic architecture for reference



gustlin_3df_01a_220517.pdf

This proposal:



gustlin_3df_01a_220517.pdf

RS(544,514) based

RS(544,514) + Inner code (128,120) based

Per:

brown_3dj_optx_adhoc_01a_230222.pdf

gustlin_3dj_optx_adhoc_01a_230222.

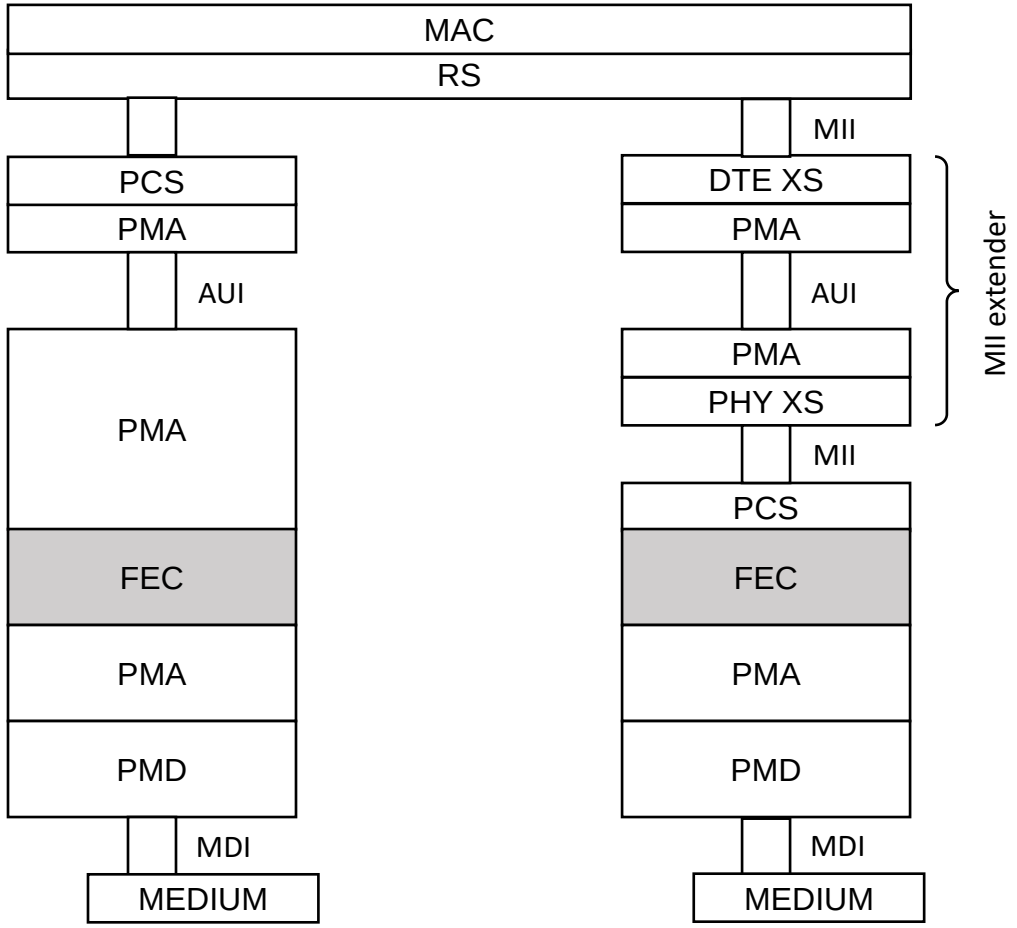
Both approaches support segmentation of the KP4 FEC using extender sublayer, if needed.

Existing Clause 172

200G/lane: Based on New symbol mux PMA as proposed by : https://www.ieee802.org/3/dj/public/23_03/ran_3dj_01_2303.pdf

This presentation provides a baseline proposal for these two new sublayers

Adopted architecture for inner code FEC

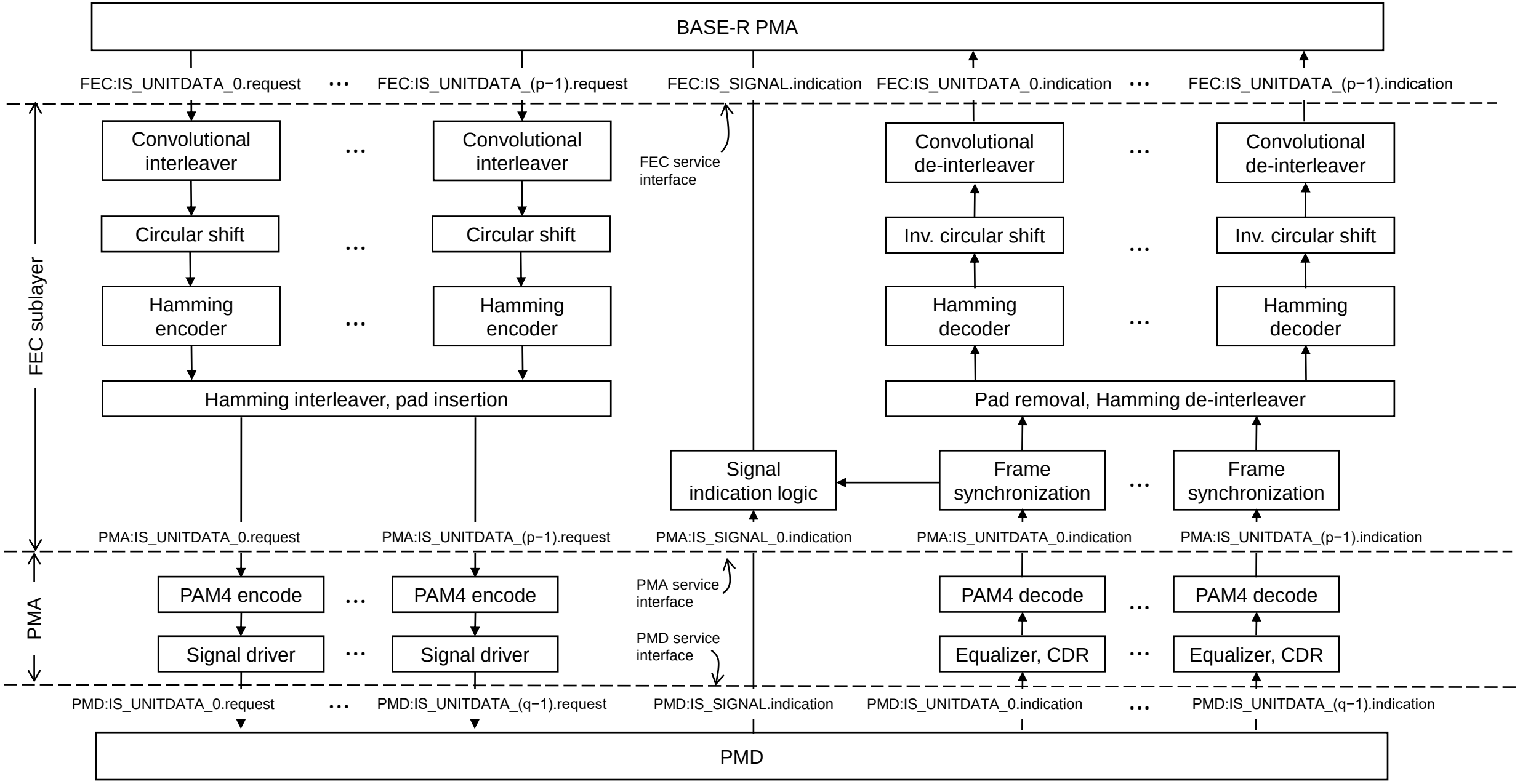


Type 2 scheme

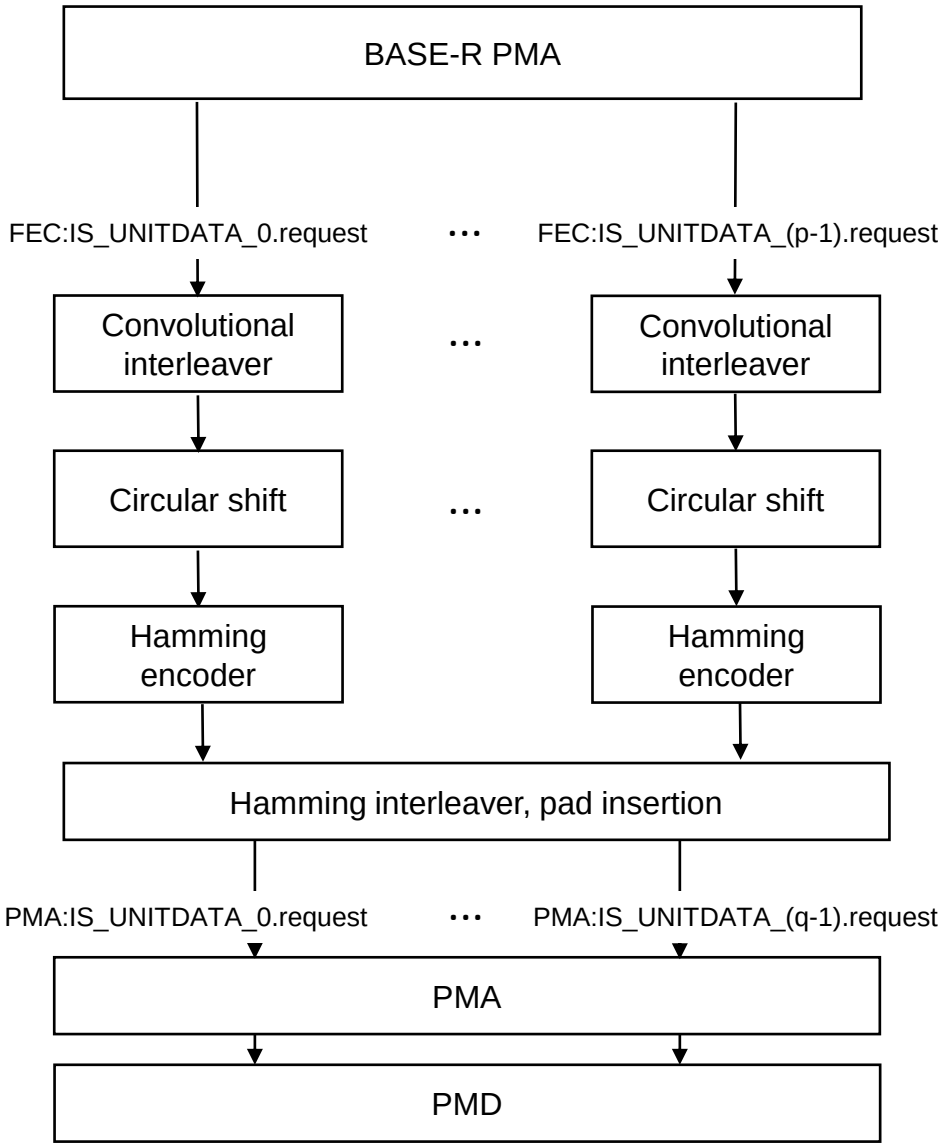
Type 2 + MII extender

- The following presentation describes a new FEC sublayer to be used in conjunction with 800GBASE-FR4 PMDs
- It is compatible with the Type 2 FEC schemes as described in [brown_3dj_optx_adhoc_01a_230222]
- It is compatible with the PCS and PMA defined for 800 Gb/s Ethernet
- Although, the diagram shows only one AUI per direction, it is understood there may be up to two AUIs per direction.

Functionality of FEC sublayer block diagram



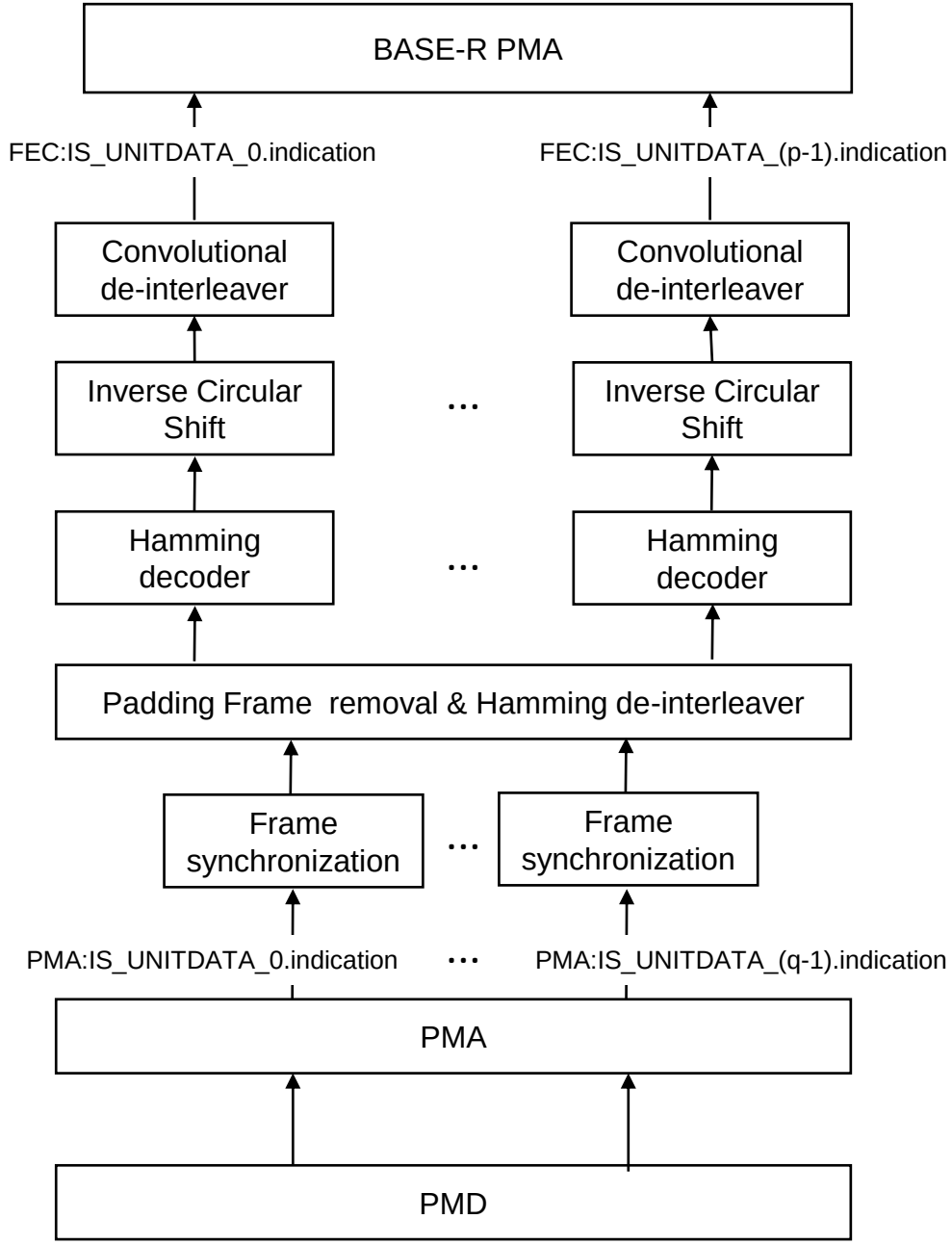
Transmit path



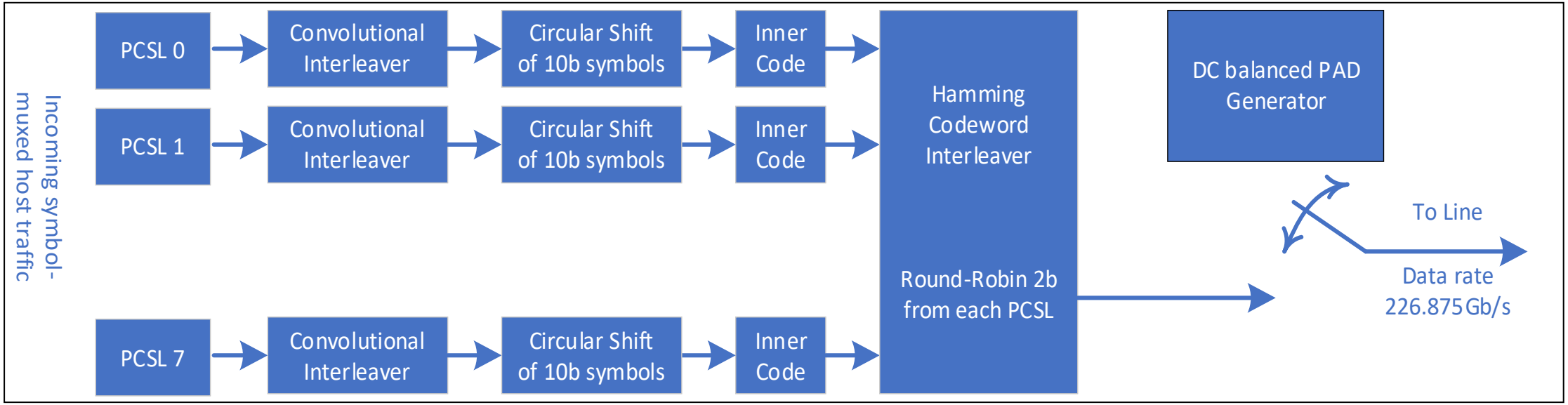
- Encode *p* PCS lanes into *q* physical lanes
- All the encoding is done per PCS lane
- Circular Shift block maximizes the distance in Bauds between transmitted PAM4 symbols from two different RS symbols in the same RS(544,514) FEC
- Circular shift, Hamming Encoder, Hamming Interleaver and padding functionality are provided in details in the subsequent slides

Receive path

- Decode *q* physical lanes into *p* PCS lanes
- All the decoding is done per PCS lane
- Frame synchronization can be done based on the padding frames as described in the padding sections – in Appendix A



Detailed representation of Transmit Datapath with Inner code (128,120) and padding



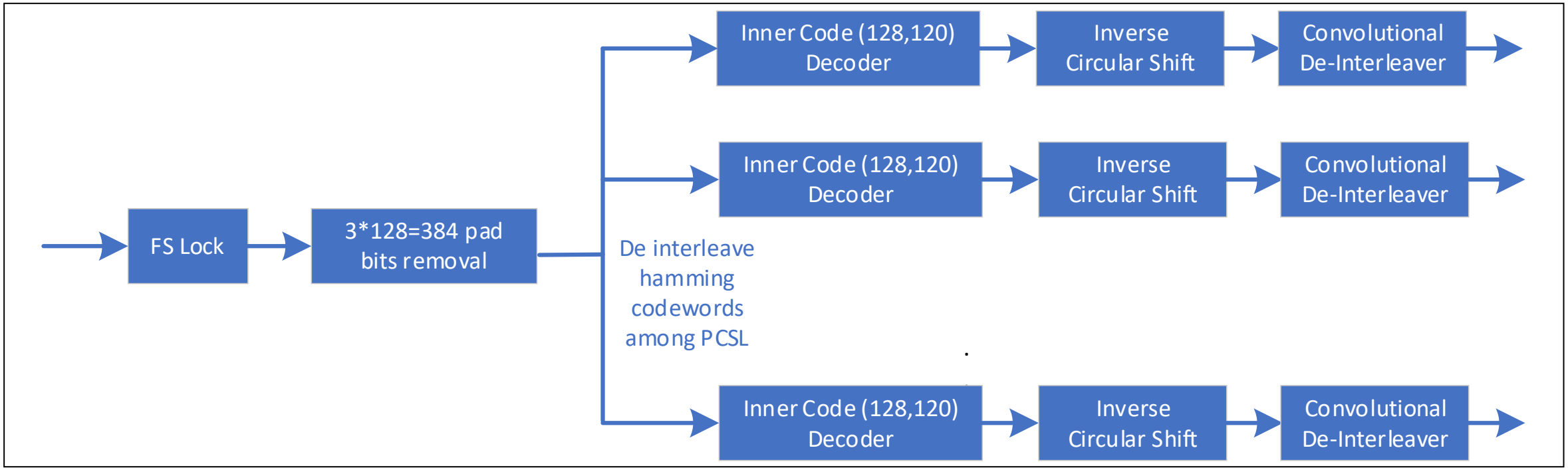
Convolutional Interleaving (CI): To form Hamming payload as 12x10b KP4 symbols
Circular shift: To enable a simple 2-bit MUXing for the Hamming Interleaver
Hamming Encoder: Appends 8b parity to 120b payload
Hamming Codeword Interleaver: 8-ways Hamming Interleaving is proposed

- The 8 ENC outputs are aligned with respect to Hamming codeword boundaries
- The 8 ENC@25G codewords are round-robin inter-leaved, in units of 2b per FECL

Padding bits:

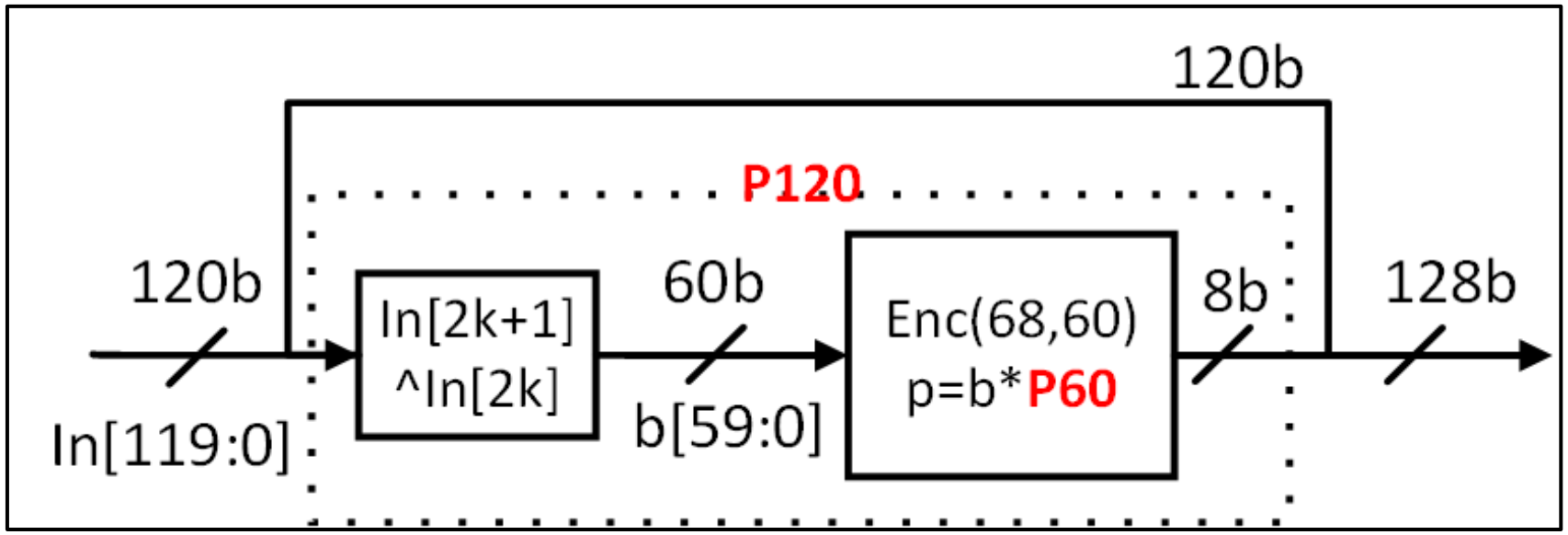
- Padding Symbol: 384 bits (3x128bits) being inserted after every 3264 hamming codewords on TX
- On RX: Padding bits needs to be removed before any processing happens

Detailed representation of Receiver Datapath with Inner code (128,120) and padding



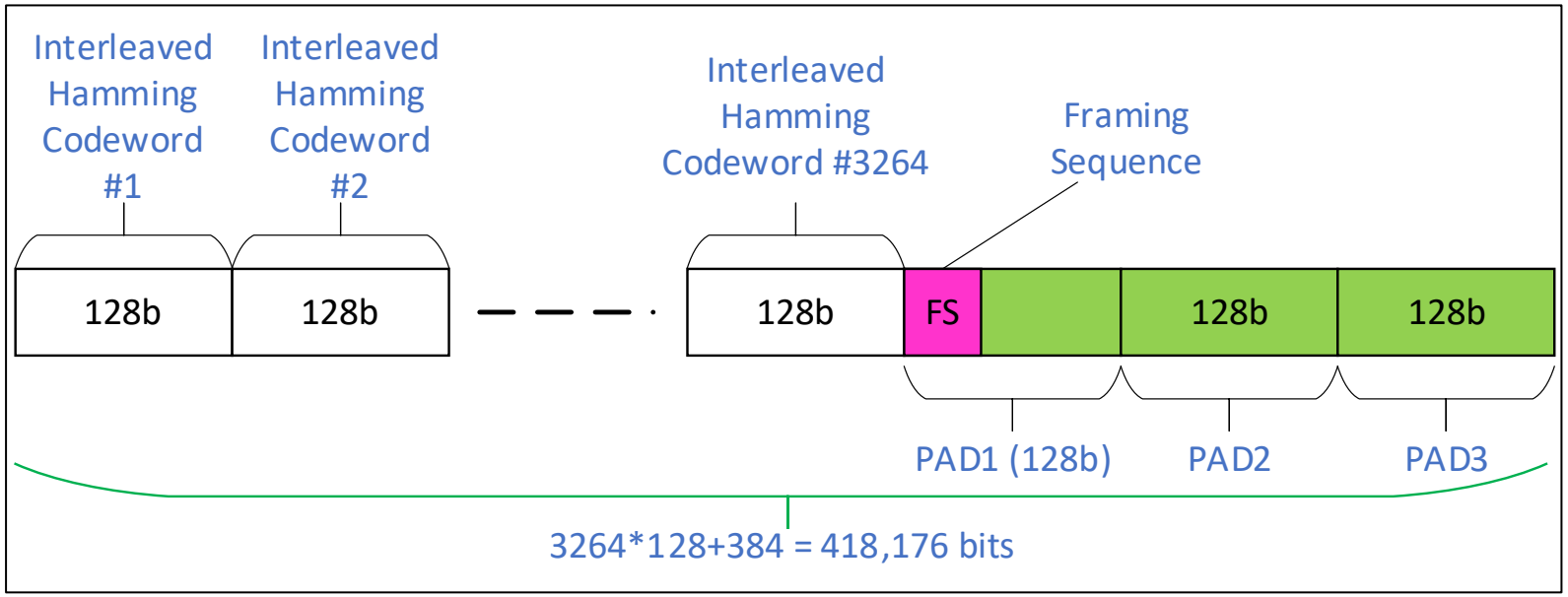
- FS Lock: 6 bytes out of 384 pad bits can be used for frame synchronization (i.e. similar to 400G PCS AM scheme which offers DC balance & hardware reusability)
- See Appendix A for an example

Inner code (128,120) based on Hamming (68,60)



- 60b Hamming payload is formed by XOR of bits in 2bx60 input
  - As an example refer to scheme : https://www.ieee802.org/3/df/public/22_10/22_1005/bliss_3df_01_220929.pdf
- Same rate as extended hamming code (128/120) and block length of 128b
- Input is aligned with incoming 12 x 10b RS symbols from Host
- 1b per payload PAM4 UI, 2b per parity PAM4 UI
  - Benefits of smaller area due to reduction in logic for syndrome/parity calculation

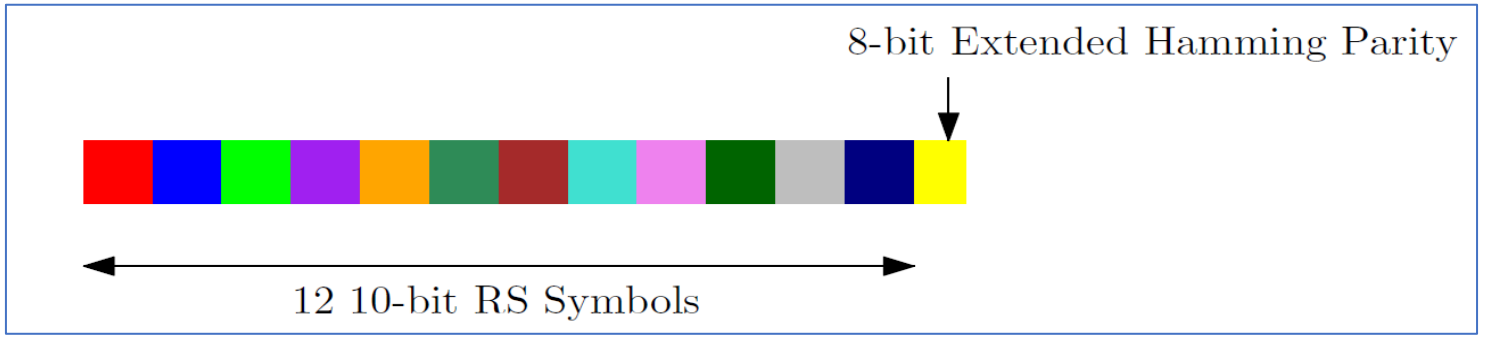
Insertion of padding bits to make the line rate a multiple of 156.25MHz reference clock frequency



- Padding Symbol: 384 bits (3x128bits) being inserted after every 3264 hamming codewords on TX
  - Inner Code (128,120) baud rate without padding → 113.3333GBaud
  - Inner code (128,120) baud rate with padding → 113.4375GBaud – i.e., an integer multiple of 156.25M*__726__
- On RX: Padding bits needs to be removed before any processing happens
- The DC balanced pad bits includes Framing Sequence (FS) to help identifying the location of the pad bits as well as the boundary and the order of each inner code (128,120). For the rest of the contents of the 3*128=384 pad bits, please refer to appendix A of this presentation

RS(544,514) +Inner Code (128,120): Convolutional Interleaver

- Convolutional Interleaver (CI) implementation guarantees that the 12x10 bit payload of the Hamming encoder comes from 12 distinct RS codewords
  - It also helps with randomly breaking up burst errors and make the concatenated code operates closer to AWGN limit

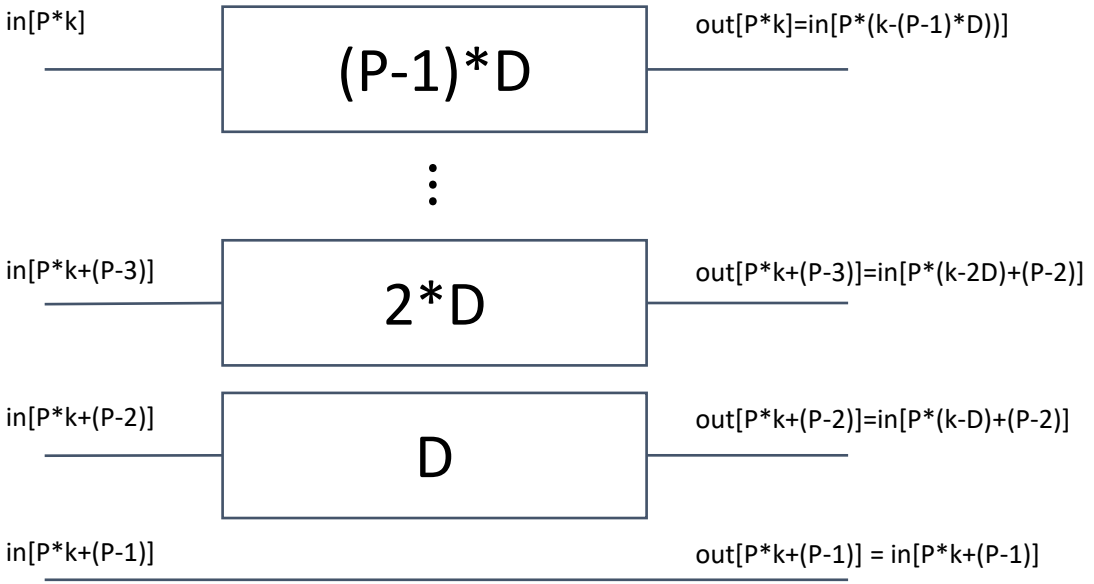


*8 parity bits are computed over **12** (10b) RS Symbols, each RS symbol from distinct codewords*

Parameterized view of per-lane Convolutional Interleaver

- Convolutional Interleaver is defined per FECL lane
- Parameters for the per-lane Convolutional Interleaver
  - W: Number of KP4 RS codewords in each “word”
  - P: Number of sub-lanes of interleaver
  - D: Number of “word” delays
  - k : Time index
  - in[k]: Input “word” at time index k
  - out[k]: Output “word” at time index k

W-symbol words at interleaver input are round-robin distributed to P sub-lanes

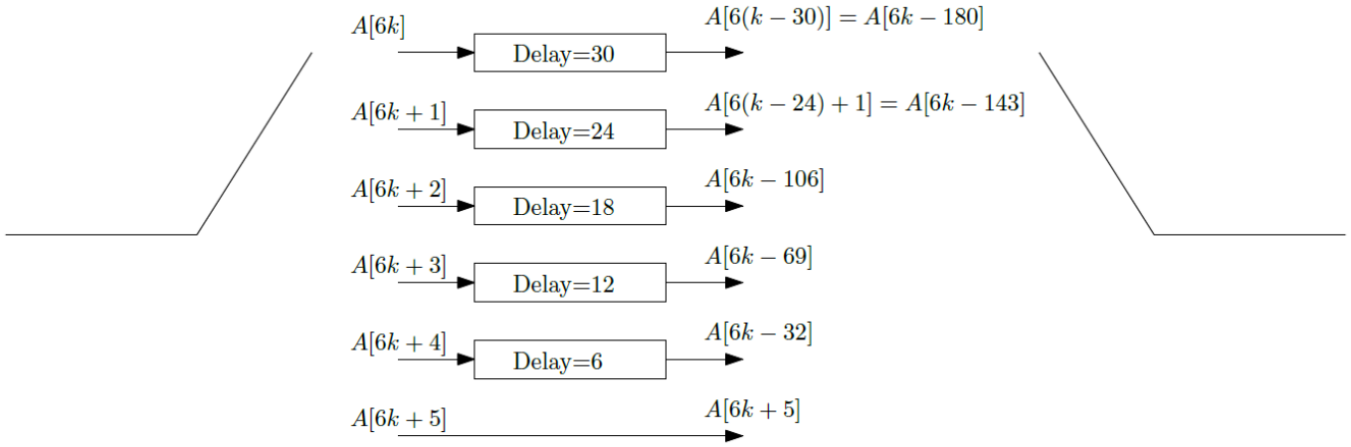


W-symbol words from P sub-lanes are round-robin multiplexed to interleaver output

Illustration : 800G/400G mode (2 way interleaved) Convolutional Interleaver

- **20b** (FEC_A,FEC_B) symbols represented by **A[m]**
  - Delay lines operate on 20b symbols
- 544/16=34
  - A[m] and A[m-n] are guaranteed to come from distinct RS codewords if n>=34
- 6 “branches” of CI due to 6-way interleaving of (FEC_A,FEC_B) symbols in Hamming payload
  - Hamming payload is **(A[6k-180],A[6k-143],A[6k-106],A[6k-69],A[6k-32],A[6k+5])**

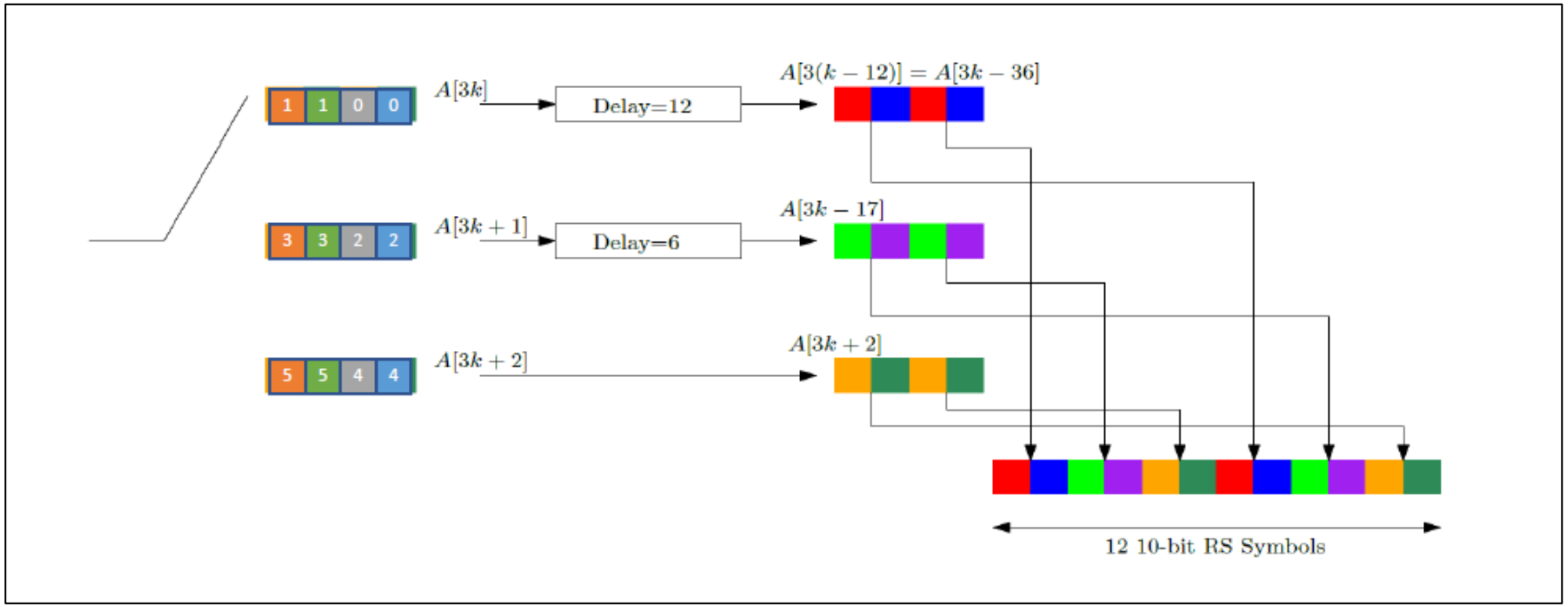
- Input/output switches are always in sync, but **no requirement** to sync relative to AM position
  - At TX, switches move from top to bottom
- Synchronization at RX is **implied** by Hamming codeword boundaries (i.e., Hamming sync implies CI sync)



Translating this to Parametrized CI will result in W = 2, P = 6, D = 6

Illustration: 800G mode (4 way interleaved) Convolutional Interleaver

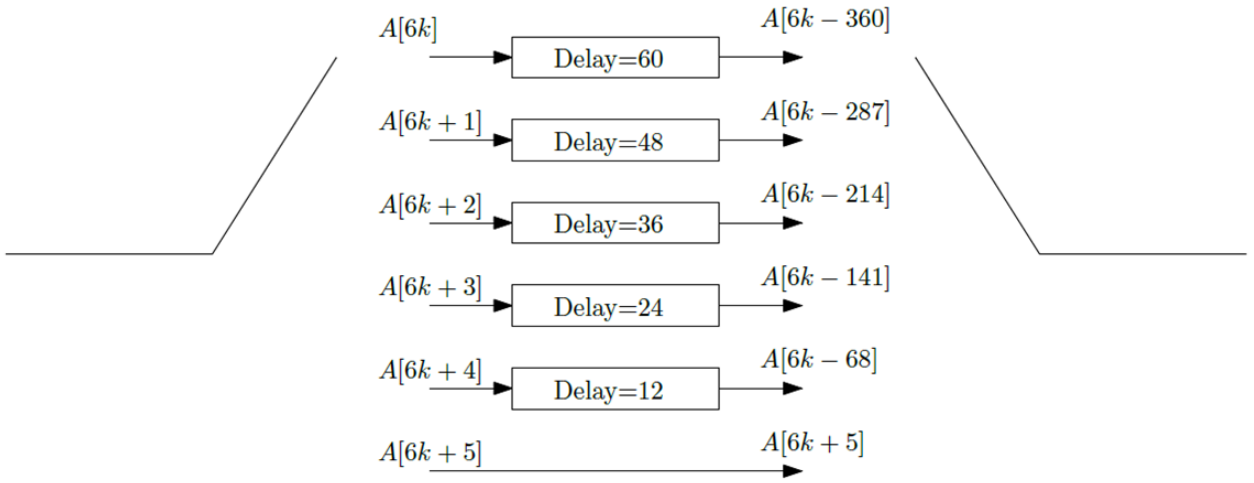
- **40b** (FEC_A,FEC_B, FEC_A,FEC_B) symbols represented by **A[m]**
  - Delay lines operate on 40b symbols
- **Total Latency (CI+CDI): 36x40=1440b @ 25G**



Translating this to Parametrized CI will result in W = 4, P = 3, D = 6

Illustration: 200G (2 way interleaved) Convolutional Interleaver

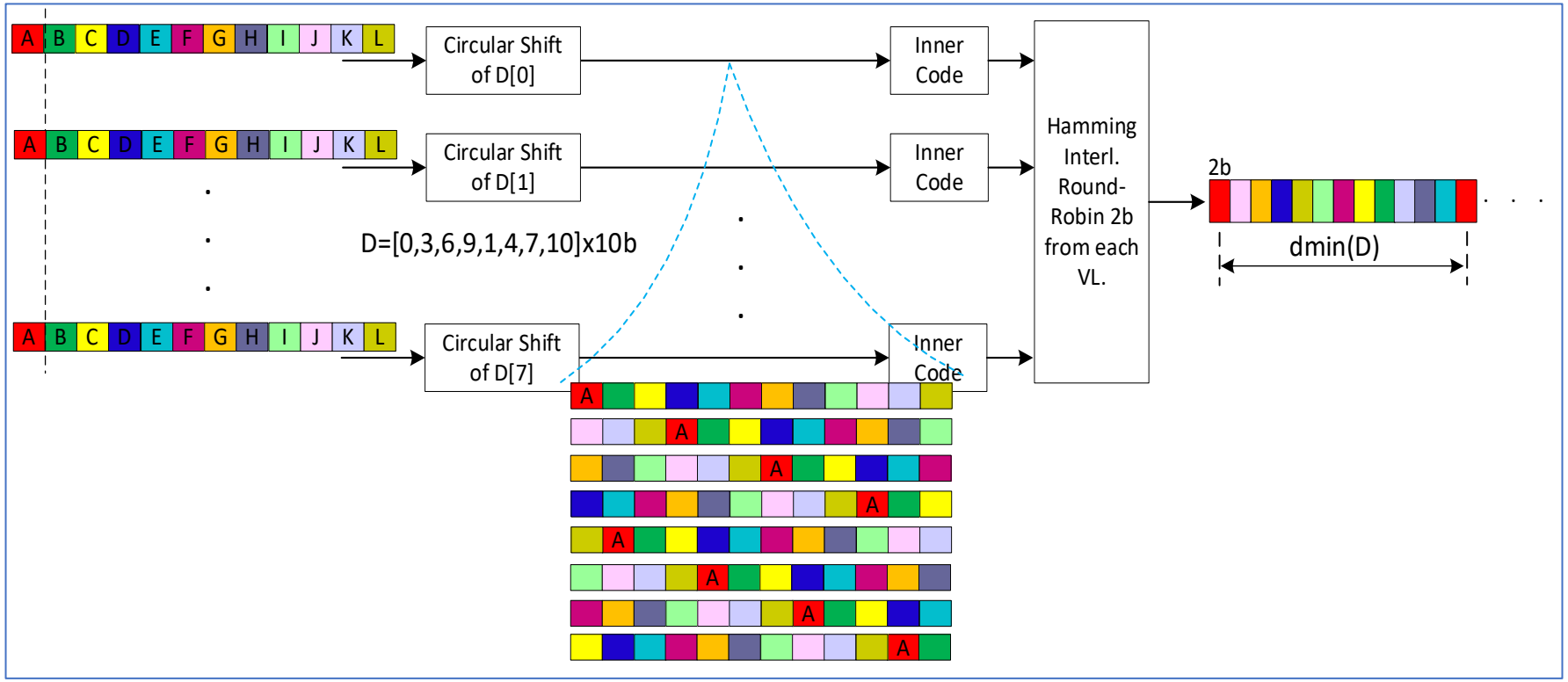
- **20b** (FEC_A,FEC_B) symbols represented by **A[m]**
  - **Delay line operates in 20b symbols**
- 544/8=68
  - A[m] and A[m-n] are guaranteed to come from distinct RS codewords if **n>=68**
- 6 “branches” of CI due to 6-way interleaving of (FEC_A,FEC_B) symbols in Hamming payload
  - Hamming payload is **(A[6k-360],A[6k-287],A[6k-214],A[6k-141],A[6k-68],A[6k+5])**



Translating this to Parametrized CI will result in W = 2, P = 6, D = 12

Functionality of the circular shift block

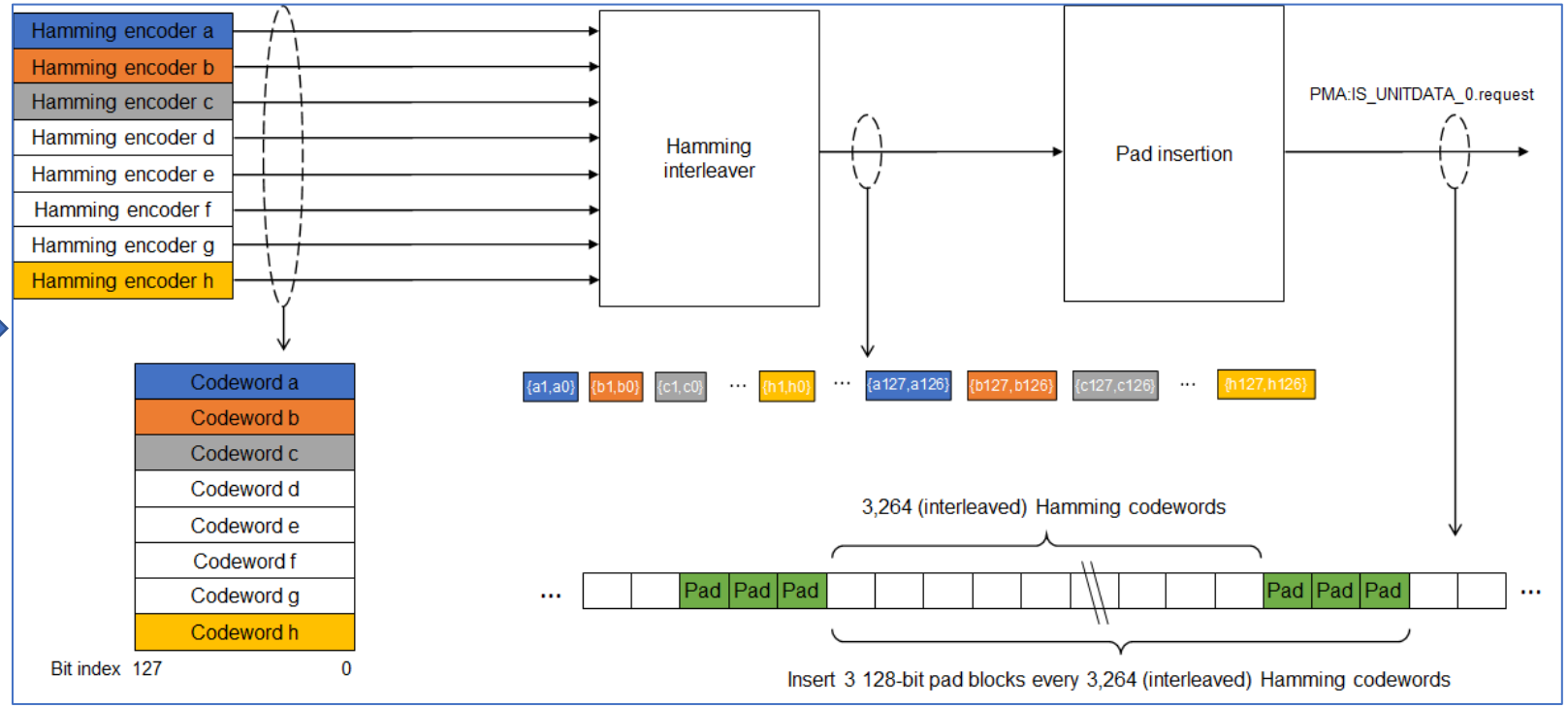
- Circular Shift block can be simply visualized as a simple rewiring of the 10b RS symbols (120b input bus) to the Inner code (128,120) block
- It maximizes the distance in Bauds between transmitted PAM4 symbols from two different RS symbols in the same KP FEC codeword
- Concept is very similar to the 200G AUI symbol muxing scheme to improve the burst error tolerance



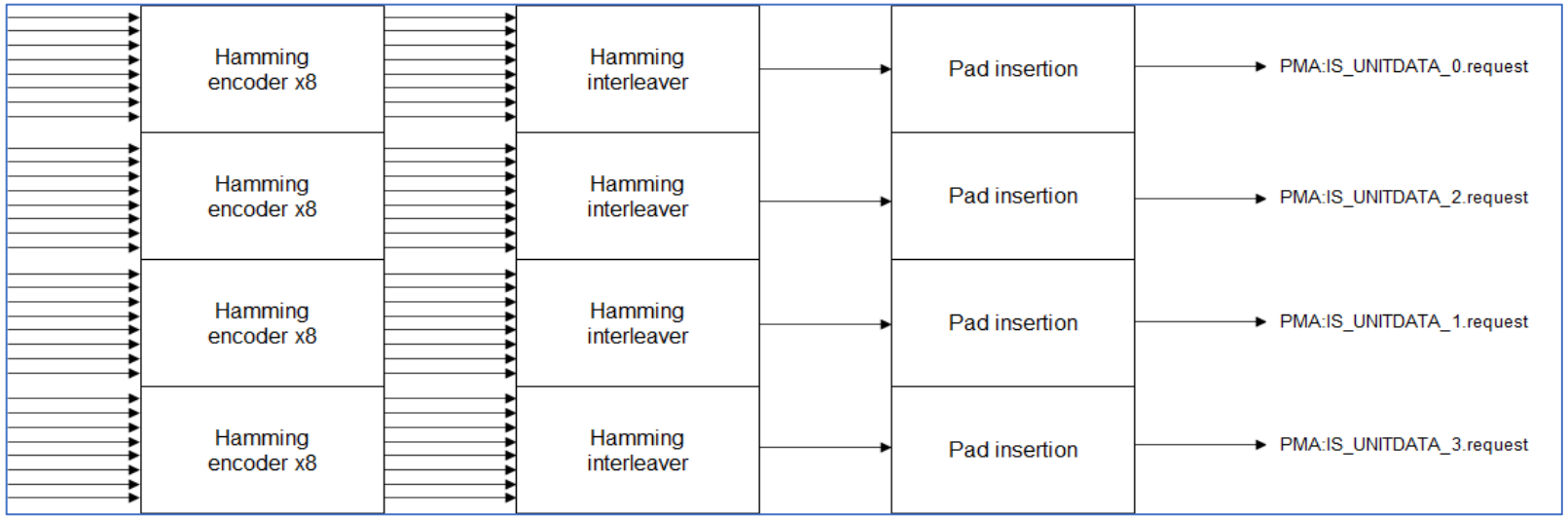
See the Burst Error tolerance presentation : https://www.ieee802.org/3/dj/public/23_03/riani_3dj_01_2303.pdf

Illustration of bit-flows with 8 ways Hamming Interleaver and padding block towards line side

Visualization of bit-flows from Hamming Encoder → Hamming Interleaver → padding blocks in 200G mode



Corresponding bits flows in 800G mode towards line side



Hamming encoder generation matrix

1	0	0	1	0	1	0	0
1	0	0	1	0	1	0	0
0	1	0	0	1	0	1	0
0	1	0	0	1	0	1	0
0	0	1	0	0	1	0	1
0	0	1	0	0	1	0	1
1	1	0	0	1	0	1	1
1	1	0	0	1	0	1	1
1	0	1	1	1	1	0	0
1	0	1	1	1	1	0	0
0	1	0	1	1	1	1	0
0	1	0	1	1	1	1	0
0	0	1	0	1	1	1	1
0	0	1	0	1	1	1	1
1	1	0	0	1	1	1	0
1	1	0	0	1	1	1	0
0	1	1	0	0	1	1	1
0	1	1	0	0	1	1	1
1	1	1	0	1	0	1	0
1	1	1	0	1	0	1	0
0	1	1	1	0	1	0	1
0	1	1	1	0	1	0	1
1	1	1	0	0	0	1	1
1	1	1	0	0	0	1	1
1	0	1	0	1	0	0	0
1	0	1	0	1	0	0	0
0	1	0	1	0	1	0	0
0	1	0	1	0	1	0	0
0	0	1	0	1	0	1	0
0	0	1	0	1	0	1	0
0	0	0	1	0	1	0	1
0	0	0	1	0	1	0	1
1	1	0	1	0	0	1	1
1	1	0	1	0	0	1	1
1	0	1	1	0	0	0	0
1	0	1	1	0	0	0	0
0	1	0	1	1	0	0	0
0	1	0	1	1	0	0	0

- SFEC(128,120): P120 matrix
- This p120 matrix is simply twice the replication of P60 matrix

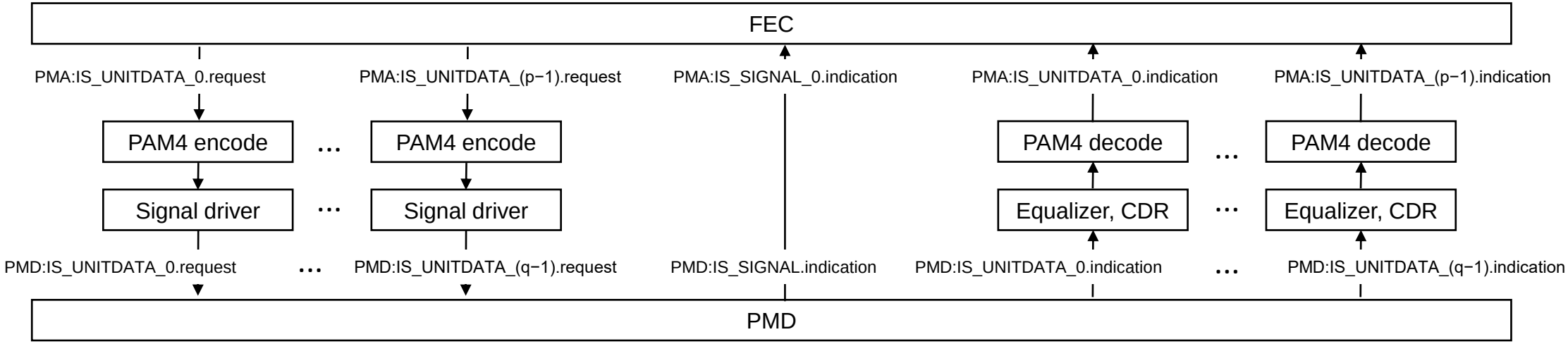
The 60x8 matrix P is given by:

P=

1	0	0	1	0	1	0	0
0	1	0	0	1	0	1	0
0	0	1	0	0	1	0	1
1	1	0	0	1	0	1	1
1	0	1	1	1	1	0	0
0	1	0	1	1	1	1	0
0	0	1	0	1	1	1	1
1	1	0	0	1	1	1	0
0	1	1	0	0	1	1	1
1	1	1	0	1	0	1	0
0	1	1	1	0	1	0	1
1	1	1	0	0	0	1	1
1	0	1	0	1	0	0	0
0	1	0	1	0	1	0	0
0	0	1	0	1	0	1	0
0	0	0	1	0	1	0	1
1	1	0	1	0	0	1	1
1	0	1	1	0	0	0	0
0	1	0	1	1	0	0	0
0	0	1	0	1	1	0	0
0	0	0	1	0	1	1	0
0	0	0	0	1	0	1	1
1	1	0	1	1	1	0	0
0	1	1	0	1	1	1	0
0	0	1	1	0	1	1	1
1	1	0	0	0	0	1	0
0	1	1	0	0	0	0	1
1	1	1	0	1	0	0	1
1	0	1	0	1	1	0	1
1	0	0	0	1	1	1	1
1	0	0	1	1	1	1	0
0	1	0	0	1	1	1	1
1	1	1	1	1	1	1	0
0	1	1	1	1	1	1	1
1	1	1	0	0	1	1	0
0	1	1	1	0	0	1	1
1	1	1	0	0	0	0	0
0	1	1	1	0	0	0	0
0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0
0	0	0	0	1	1	1	0
0	0	0	0	0	1	1	1
1	1	0	1	1	0	1	0
0	1	1	0	1	1	0	1
1	1	1	0	1	1	1	1
1	0	1	0	1	1	1	0

- Hamming (68,60)
- P60 matrix

PMA functions



- **PAM4 encode**
  - Maps pairs of bits to Gray-coded PAM4 symbols as defined in 120.5.7.1.
- **Equalizer, CDR**
  - Generates an estimate of the Gray-coded PAM4 symbol received from the PMD service interface and a measure of the quality (or reliability) of the estimate.
  - The quality of the estimate is used by the Hamming decoder in the FEC sublayer
- **PAM4 decode**
  - Maps estimate of the Gray-coded PAM4 symbol to a pair of bits as defined in 120.5.7.1.
  - The bits, and the corresponding quality estimate, are communicated to the FEC sublayer via the PMA:IS_UNITDATA_i.indication primitive for physical lane *i*

Refresh of Convolutional Interleaver + Inner Code (128,120) latency

Client Type	Parameters for Interleaver	FEC	Decoder Input BER	Latency
800GBASE-R (2 way interleaved)	W=2,P=6,D=6	RS(544,514) + Inner code (128,120)	4.85E-3	~140ns
800GBASE-R (4 way interleaved)	W=4,P=3,D=6			~56ns
400GBASE-R (2 way interleaved)	W =2,P =6,D =6			~140ns

- Proposed 800G Latency is highlighted in dotted box

- BER thresholds are simulated using AWGN model. The CI helps with randomly distributing equalizer burst errors which results in AWGN model BER estimate to be accurate. See:
  - https://www.ieee802.org/3/df/public/22_05/22_0517/bliss_3df_01a_220517.pdf
  - For channels with correlated errors, see:
  - https://www.ieee802.org/3/df/public/22_10/22_1005/bliss_3df_01_220929.pdf

Summary

- We presented a FEC baseline proposal that enables the 200G/Lambda Optical IM-DD links for DR and FR reaches.

- **The Presented proposal is a result of consensus building by combining multiple past 802.3df/dj presentations as well as recent work since the last task force meeting.**

- The proposed inner code is a low complexity solution that works in conjunction with the existing RS(544,514) FEC to act as a booster to the overall coding gain for FR4 Link

- Leveraging the existing RS (544,514) FEC for 200G AUI will benefit the industry and will ease the backward compatibility issues.

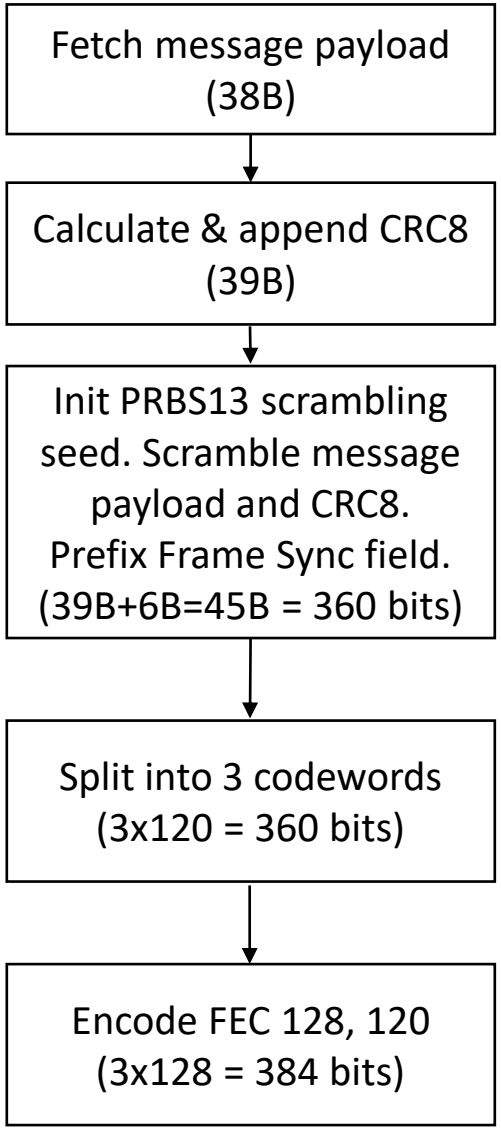
Appendix A: Padding bits

Padding Specification

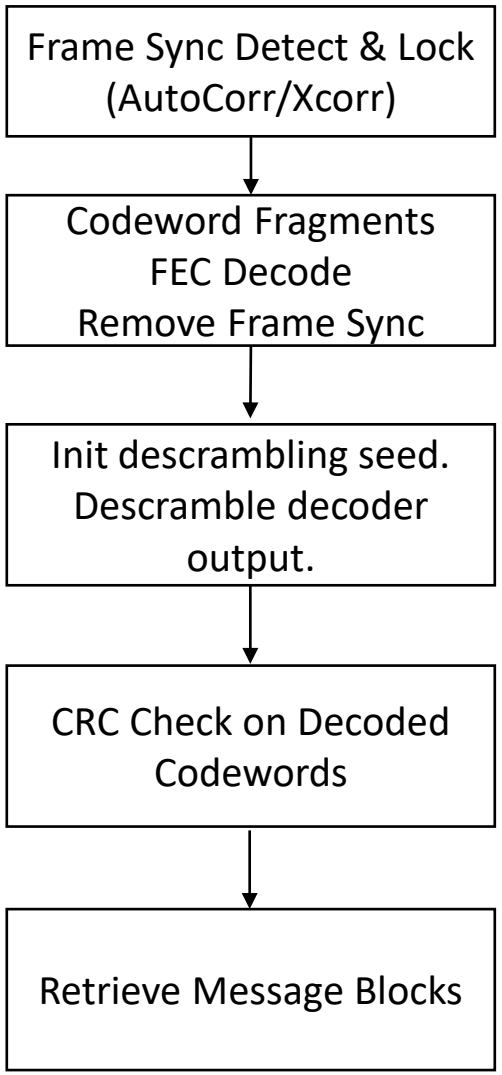
- **384 bits = 3 CW using 128, 120 code**
  - Payload bits = 360 (=45 B), parity = 24 bits

- **45 data bytes composed as follows**
  - 6 byte frame sync field (same as 200G/400G PCS AM, offers DC balance & hardware reuse):
    - 0x9A4A2665B5D9
  - Remaining 312 bits are scrambled with PRBS13, using generator polynomial $X^{13} + X^{12} + X^2 + X + 1$, seed reset to 0xCCC for each pad fragment):
    - 38 byte Message field – Start of scrambling with PRBS
      - 8 bit message index (8 bit counter 0 to 255)
      - 8 bit message type (see slides 4 & 5)
      - 36 bytes message content
    - 1 byte CRC8 (calculated on previous 38 bytes) – polynomial is $X^8+X^5+X^4+1$
- **The 38-bytes message field (details to be specified) can be used to convey link and signal-related information, such as receiver state, channel pulse response, FEC stats, etc**

Padding Field Construction – Reference Implementation



Padding Field Consumption – Reference Implementation (Informative)



Thanks !