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

Co-Authors:

Arash Farhood, Marvell
Will Bliss, Broadcom
Sridhar Ramesh, Maxlinear
Dave Cassan, Alphawave

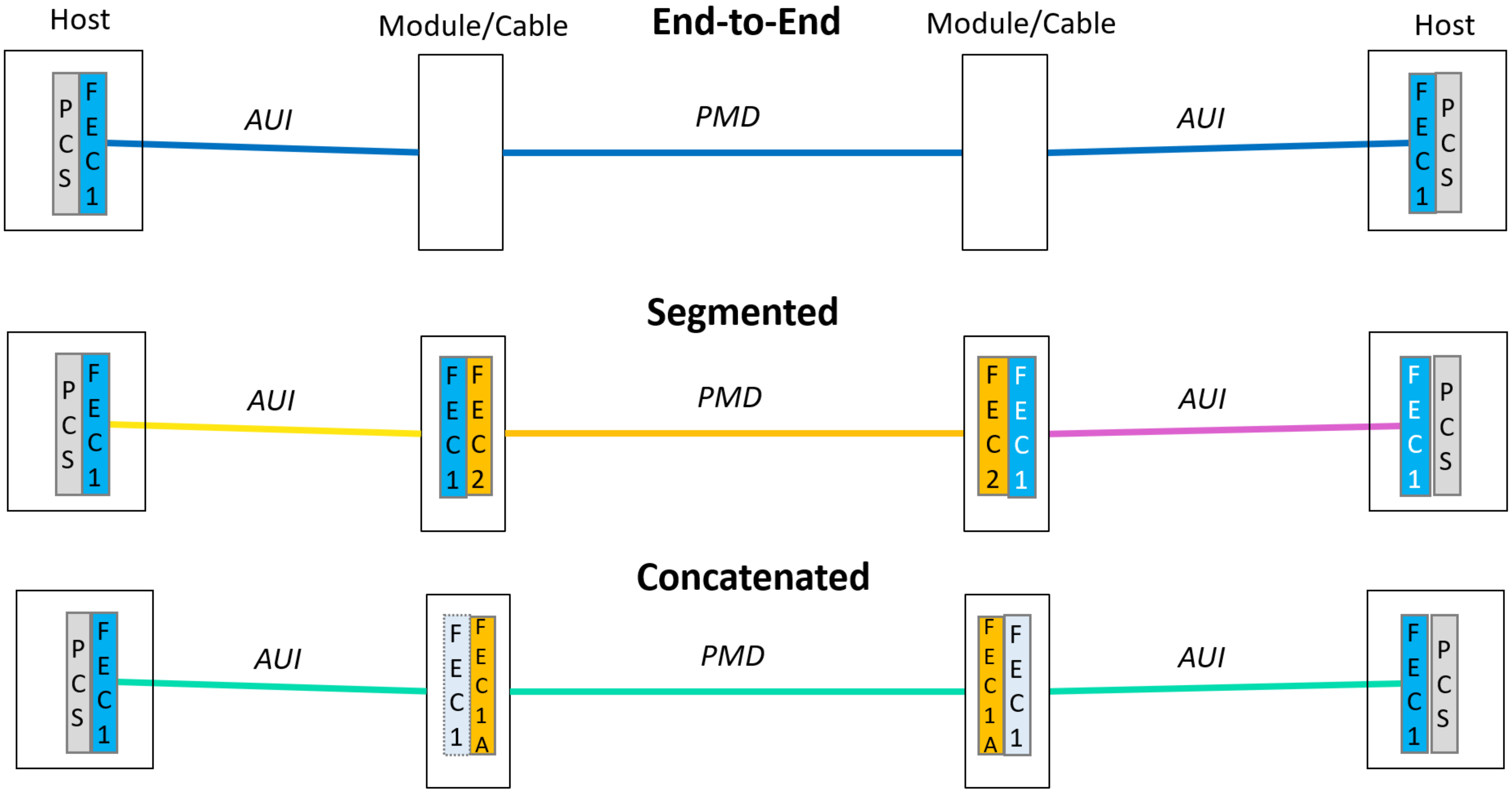
Supporters list:

- Xiang Zhou, Google
- Shuang Yin, Google
- Ali A. (Reza) Eftekhar, Amazon
- Vishnu Balan, Nvidia
- Scott Schube, Intel
- Rajesh Radhamohan, Cisco
- Hacene Chaouch, Arista
- Jeffery Maki, Juniper Networks
- Maxim Kuschnerov, Huawei
- Xiang Liu, Huawei
- Masoud Barakatain, Huawei
- Kechao Huang, Huawei
- Xiang Liu, Huawei
- Dirk Lutz, Eoptolink
- K.T. Tsai, Molex
- Vipul Bhatt, Coherent
- Roberto Rodes, Coherent
- Osa Mok, Innolight
- Frank Chang, Source Photonics
- Jeff Hutchins, Ranovus Inc.
- Chris Cole, Quintessent Inc.
- Rangchen Yu, Sifotonics
- David Piehler, Dell
- Atul Srivastava, NEL
- Rich Mellitz, Samtec
- Marek Tlalka, Macom
- Ryan Latchman, Macom
- Chris Doerr, Aloe Semiconductor
- Ali Ghiasi, Ghiasi Quantum
- David Chen, Independent
- Lenin Patra, Marvell
- Vasu Parthasarathy, Broadcom
- Tony Chan Carusone, Alphawave
- Jamal Riani, Marvell
- Chungjue Chen, Broadcom
- Henry Wong, Alphawave
- Drew Guckenberger, Maxlinear
- Kishore Kota, Marvell

A 200Gbps IMDD Inner FEC Consensus Proposal *overview*

- A large group of contributors
  - A new proposal that includes many components of prior proposals
- Support DR, FR, and potentially LR links with low complexity
  - Over 2.5dB soft coding gain
- Support current and future AUI interfaces with no burden placed on them
  - Can support both concatenated as well as segmented+concatenated FEC schemes
  - All forms of bit or symbol muxing in AUI are supported, with no performance impact on the optical link
- Key Format Features
  - A (128,120) extended Hamming Inner mother code
    - A special shortening technique that reduces both encoding and decoding complexity
  - An extra data (pad) field
    - Baud Rate = 156.25MHz * 726
    - Minor 0.098..% increase in Baud rate
    - Creates a back channel that allows future enhancements

Refresh of FEC Architecture already discussed in this forum: End-to-End, Segmented, Concatenated scheme



End-to-End

FEC 1: is end-to-end (AUI+PMD+AUI)

Segmented

FEC 1: AUI only

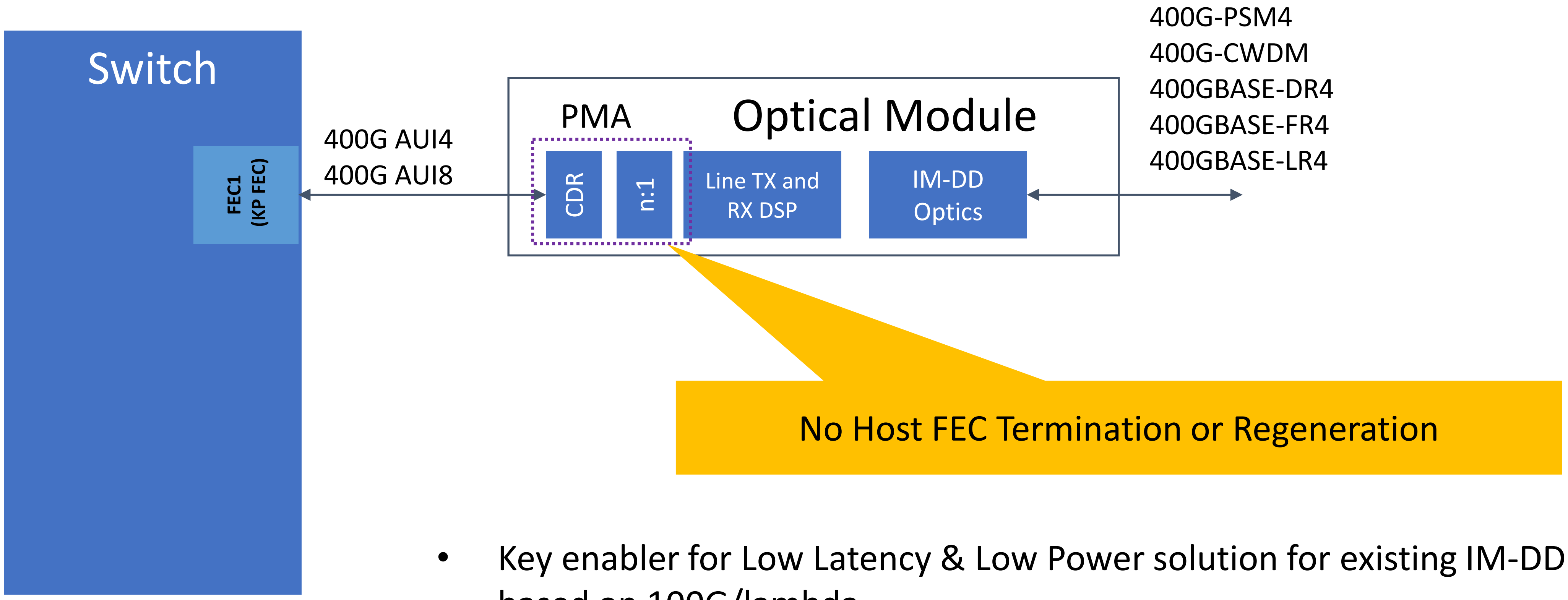
FEC 2: PMD only

Concatenated

FEC 1 (outer) FEC: (AUI + PMD + AUI**)**
Optional PMD FEC1
FEC 1A (inner) FEC: (PMD only)

What is inside the Data Center Optical modules today?

- “Re-timers” and “gearboxes” represent the bulk of DSP deployed inside the IM-DD optics today.

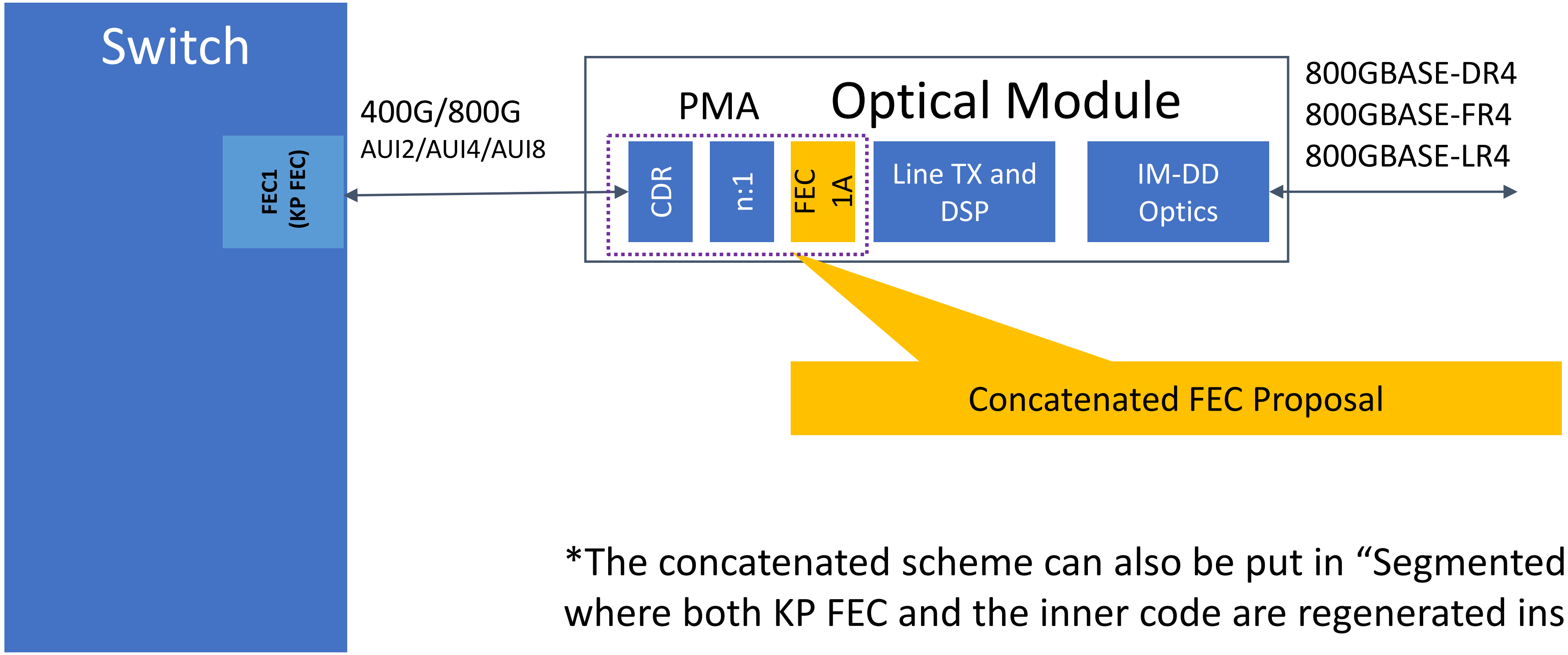


No Host FEC Termination or Regeneration

- Key enabler for Low Latency & Low Power solution for existing IM-DD optics based on 100G/lambda

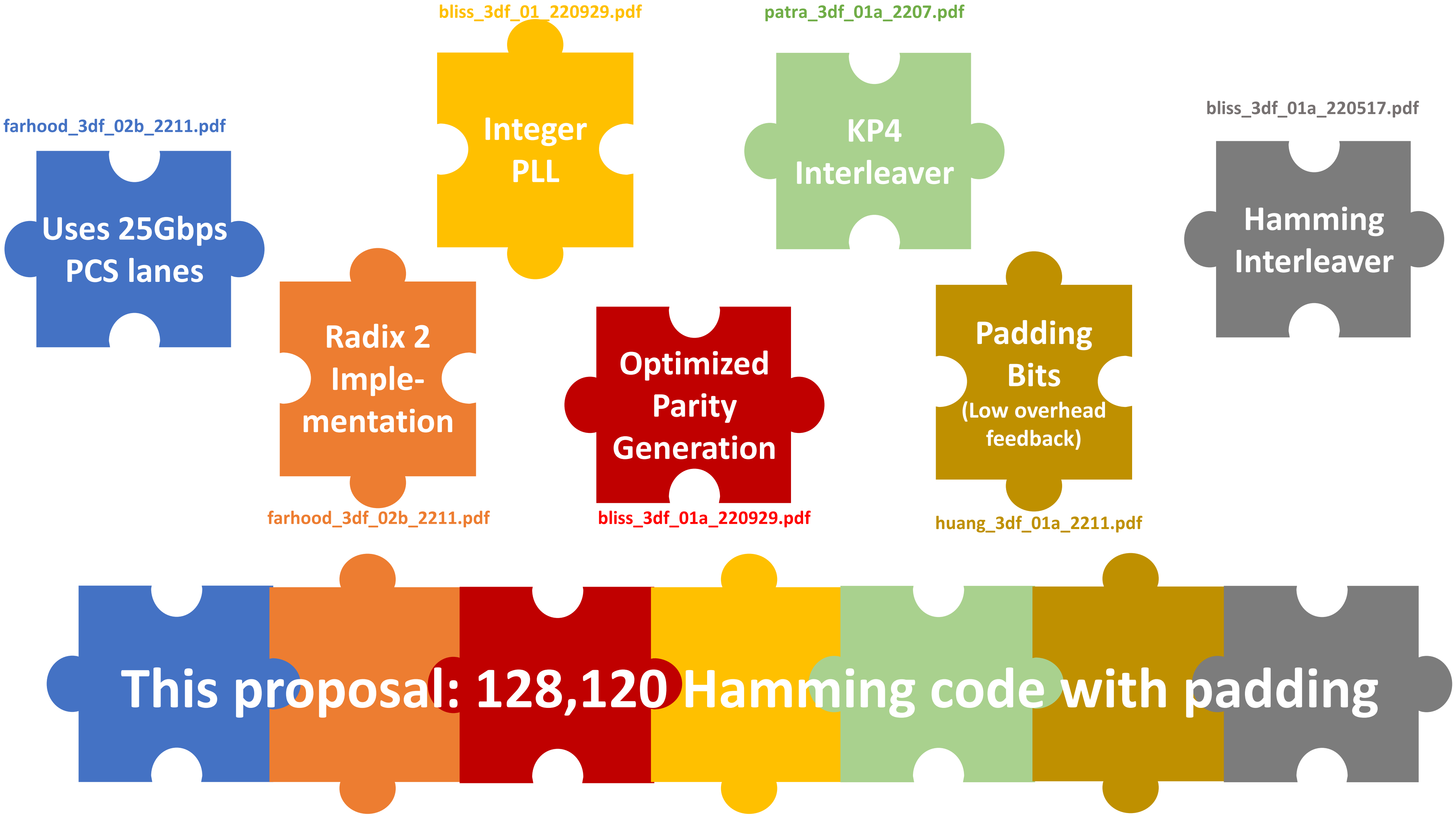
Concatenated FEC extends the Low complexity for NextGen 200G/lambda

- n:1 "gearbox" generalized to a simple convolutional interleaver
- Inner FEC code concatenated with the interleaved bit stream



*The concatenated scheme can also be put in "Segmented+Concatenated" mode where both KP FEC and the inner code are regenerated inside the optical module

Path to Convergence: Diverse thoughts driving a new approach



farhood_3df_02b_2211.pdf

Uses 25Gbps PCS lanes

bliss_3df_01_220929.pdf

Integer PLL

patra_3df_01a_2207.pdf

KP4 Interleaver

bliss_3df_01a_220517.pdf

Hamming Interleaver

Radix 2 Imple-mentation

farhood_3df_02b_2211.pdf

Optimized Parity Generation

bliss_3df_01a_220929.pdf

Padding Bits

(Low overhead feedback)

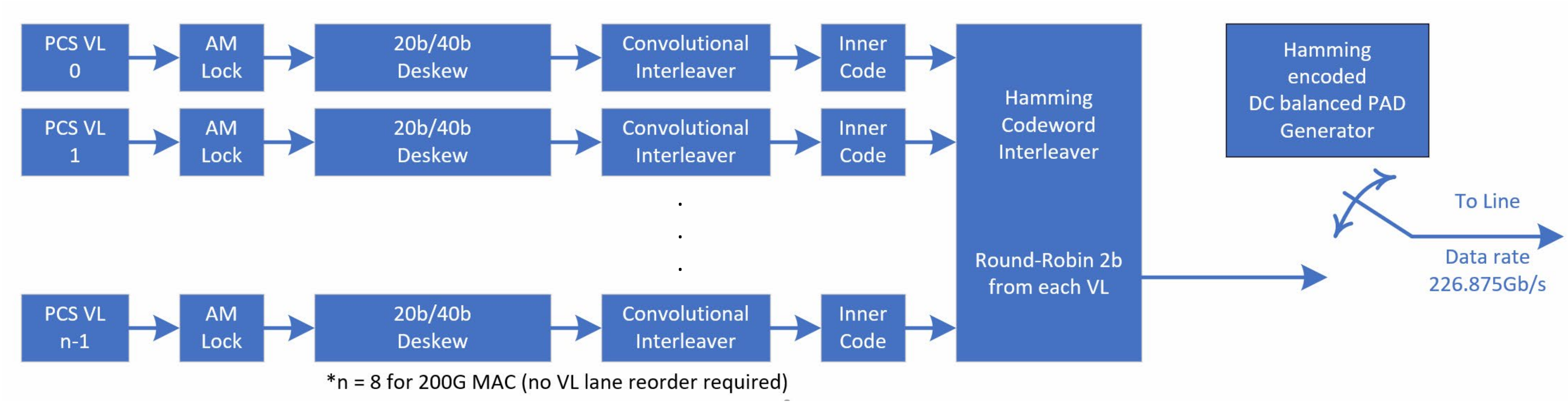
huang_3df_01a_2211.pdf

This proposal: 128,120 Hamming code with padding

Inner Code Data Path

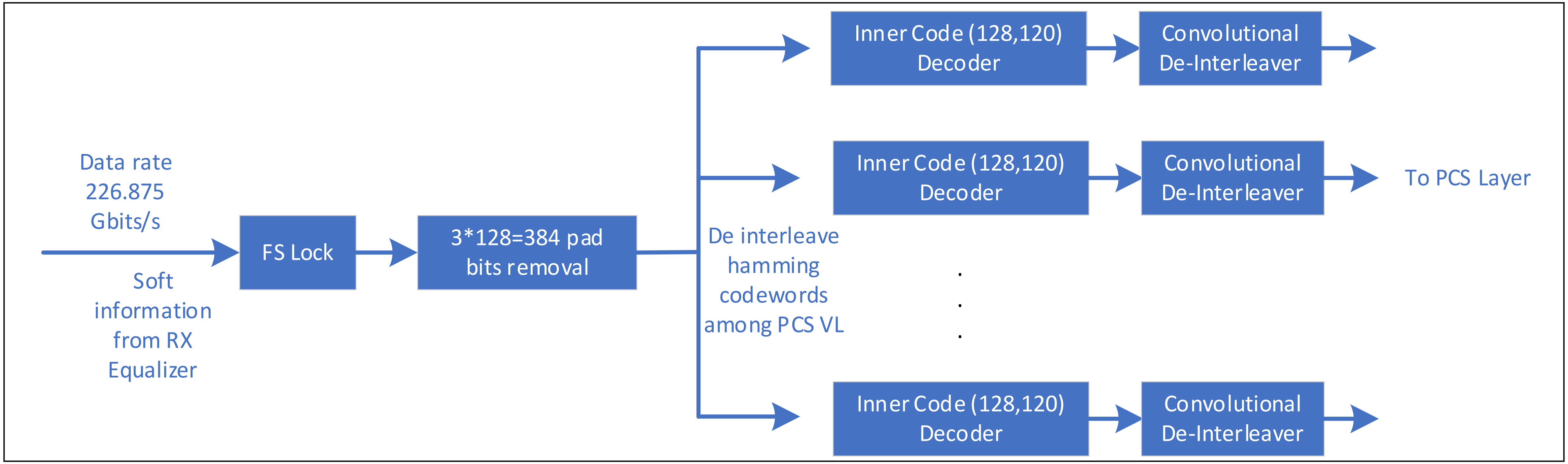
- Details of Inner Code mode Processing:
    - **Per FECL** TX processing
        - **AM lock:** To determine KP4 symbol boundaries
        - **20b/40b deskew:** Enable high performance when connected to 100G AUI, 50G AUI
        - **Convolutional Interleaving (CI):** To form Hamming payload as 12x10b KP4 symbols
        - **Hamming Encoder:** Appends 8b parity to 120b payload
    - **Hamming Codeword Interleaver**
        - **8-way Hamming Interleaving:**
            - 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
            - E.g., 2b from ENC0, 2b from ENC1, 2b from ENC2, …, 2b from ENC7, 2b from ENC0, etc.
    - **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

Overall Representation of TX Encoding Datapath with Inner Code(128,120) and padding

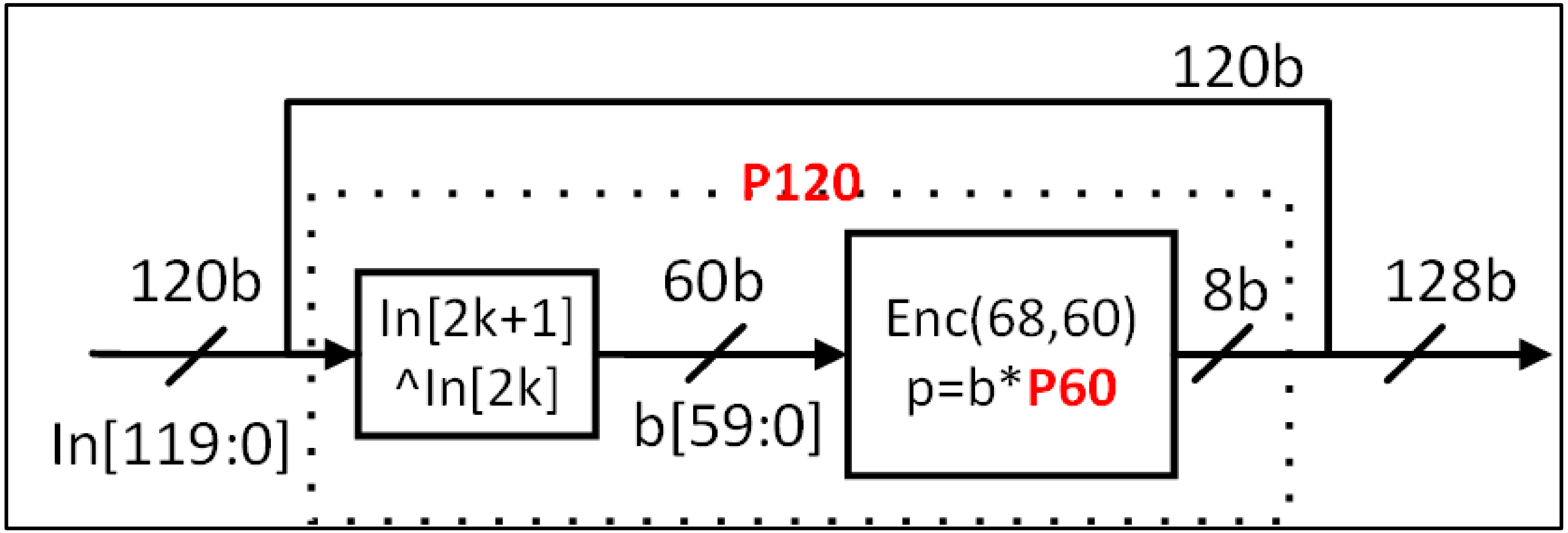

PCS VL 0
AM Lock
20b/40b Deskew
Convolutional Interleaver
Inner Code
PCS VL 1
AM Lock
20b/40b Deskew
Convolutional Interleaver
Inner Code
.
.
.
PCS VL n-1
AM Lock
20b/40b Deskew
Convolutional Interleaver
Inner Code
Hamming Codeword Interleaver
Round-Robin 2b from each VL
Hamming encoded DC balanced PAD Generator
To Line
Data rate 226.875Gb/s

*n = 8 for 200G MAC (no VL lane reorder required)

Overall Representation of RX Decoding Datapath with Inner Code(128,120) and padding

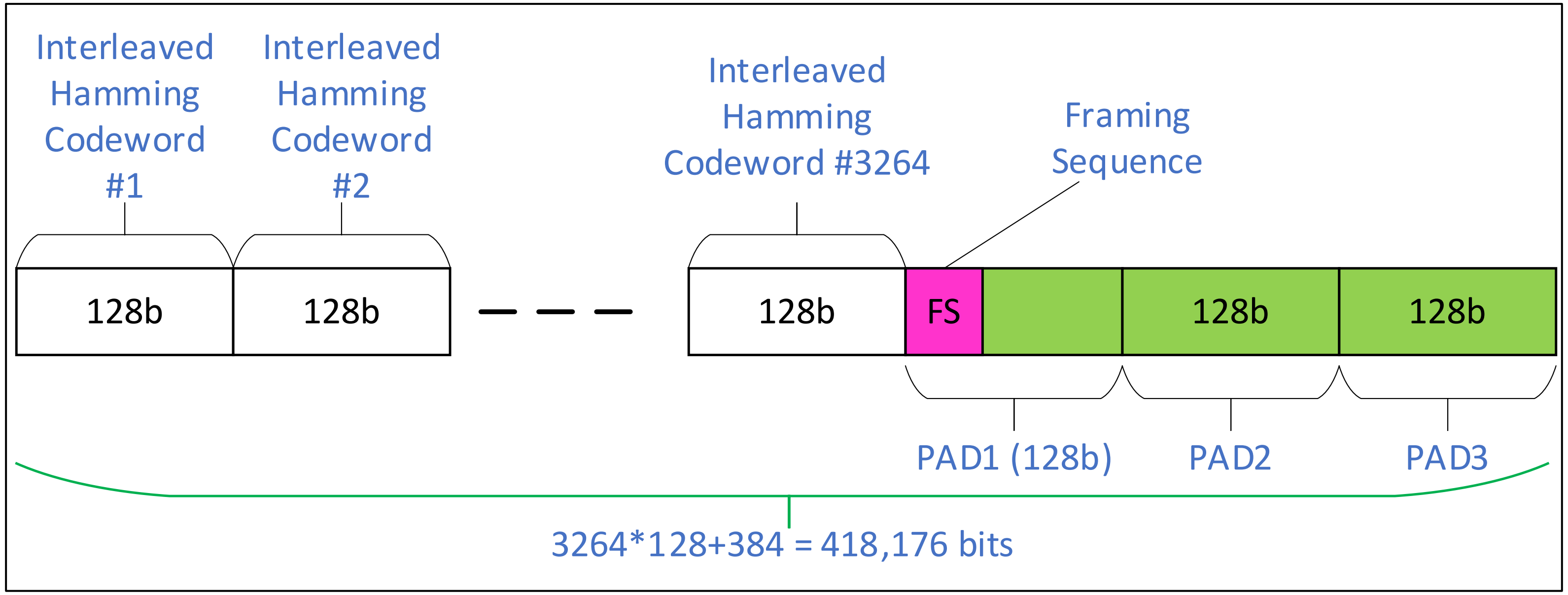


Inner code (128,120) based on Hamming(68,60)



- 60b Hamming payload is formed by XOR of bits in 2bx60 input
  - As an example scheme see: 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

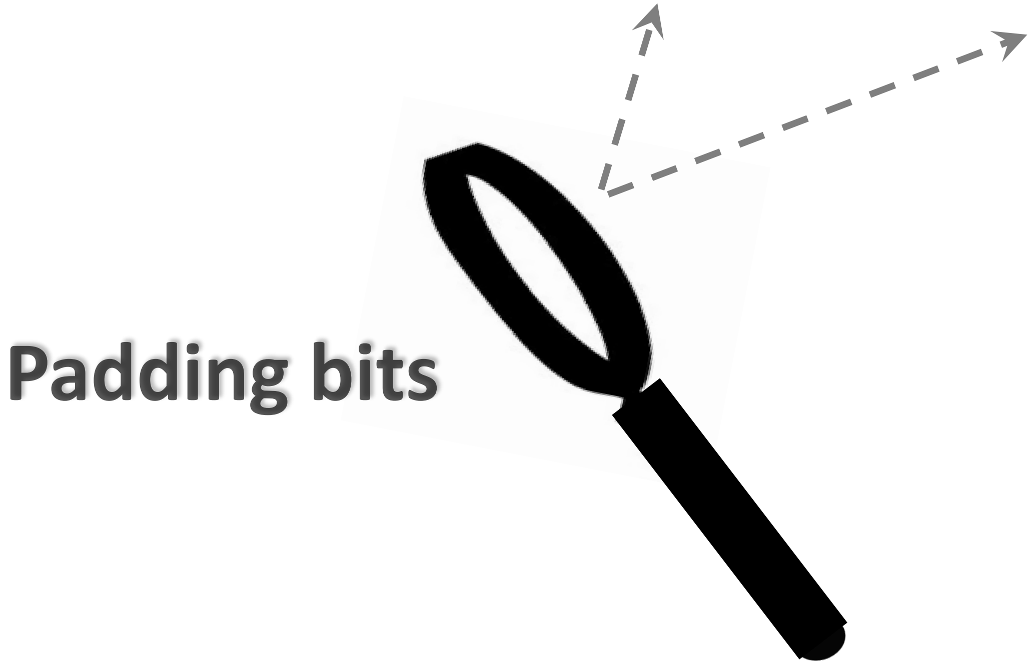

Interleaved Hamming Codeword #1
Interleaved Hamming Codeword #2
Interleaved Hamming Codeword #3264
Framing Sequence
128b
128b
128b
FS
128b
128b
PAD1 (128b)
PAD2
PAD3
3264*128+384 = 418,176 bits

- Padding Symbol : 384 bits (3x128bits) being inserted after every 3264 hamming codewords on TX
  - Inner Code (128,120) baud rate <u>without</u> padding → 113.3333GBaud
  - Inner code (128,120) baud rate <u>with</u> 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 include Framing Sequence (FS) to help with 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 see the appendix A of this presentation
- Padding proposal concept raised before in 802.3df: www.ieee802.org/3/df/public/22_11/huang_3df_01a_2211.pdf

Can the padding bits be defined to create value?

Information Theory states that feedback increases the capacity of noisy channels

- *“The Capacity of Channels With Feedback”*, S.Tatikonda and S.Mitter, IEEE Transactions on Information Theory, Vol. 55, No. 1, Jan 2009
- “*Two-way communication channels*” C. Shannon, in Proc. 4th Berkeley Symp. Math. Statist. Probab., J. Neyman, Ed., Berkeley, CA, 1961, vol. 1, pp. 611–644
- “*Arbitrarily varying channels with general alphabets and states*” I. Csiszár, IEEE Trans. Inf. Theory, vol. 38, no. 6, pp. 1725–1742, Nov. 1992



Padding bits

SUMMARY

□ **A number of uses could potentially be defined:**

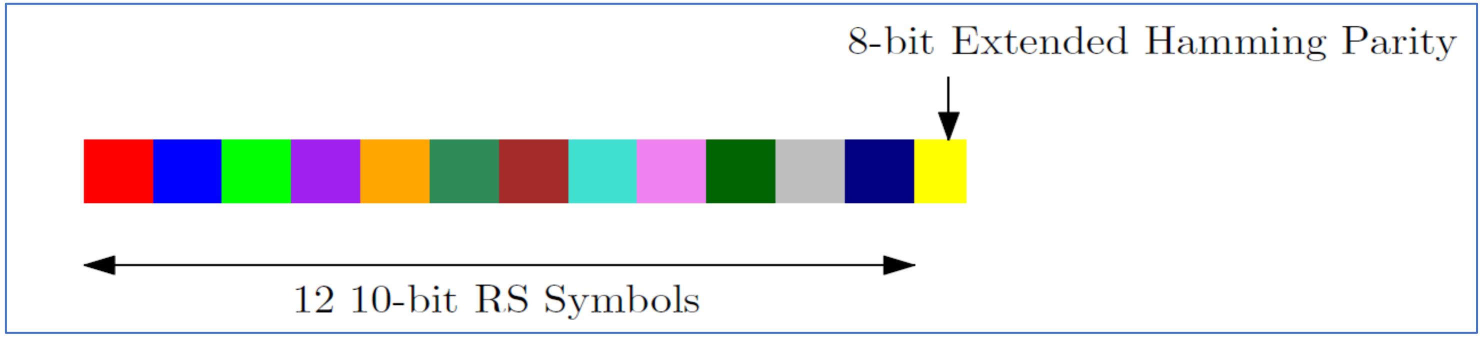
- Link training?
- Speed up signal acquisition (pilot patterns)?
- Auto-negotiation?
- Receiver link quality feedback?
- Optimize TDECQ?

□ **Definition will need to ensure appropriate characteristics:**

- DC Balance
- Scrambled to prevent tones
- Completely specified to ensure Interop (non-proprietary interface)

KP4+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 operate closer to AWGN limit



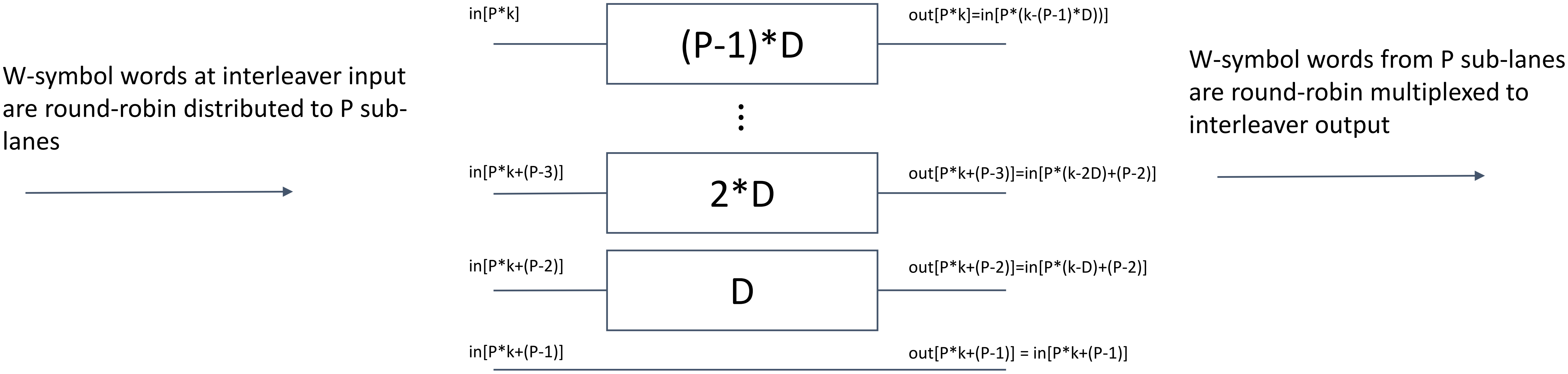
*8 parity bits are computed <u>over **12** (10b) RS Symbols</u>, each RS symbol from distinct codewords*

- Inner Code (128,120) data path will provide few distinct Convolutional Interleaver options for standardization:

1) 802.3bs 400G or 800G Ethernet Consortium or 800G IEEE
2) 802.3bs 200G

Parametrized 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

in[P*k]
(P-1)*D
out[P*k]=in[P*(k-(P-1)*D))]
in[P*k+(P-3)]
2*D
out[P*k+(P-3)]=in[P*(k-2D)+(P-2)]
in[P*k+(P-2)]
D
out[P*k+(P-2)]=in[P*(k-D)+(P-2)]
in[P*k+(P-1)]
out[P*k+(P-1)] = in[P*k+(P-1)]

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

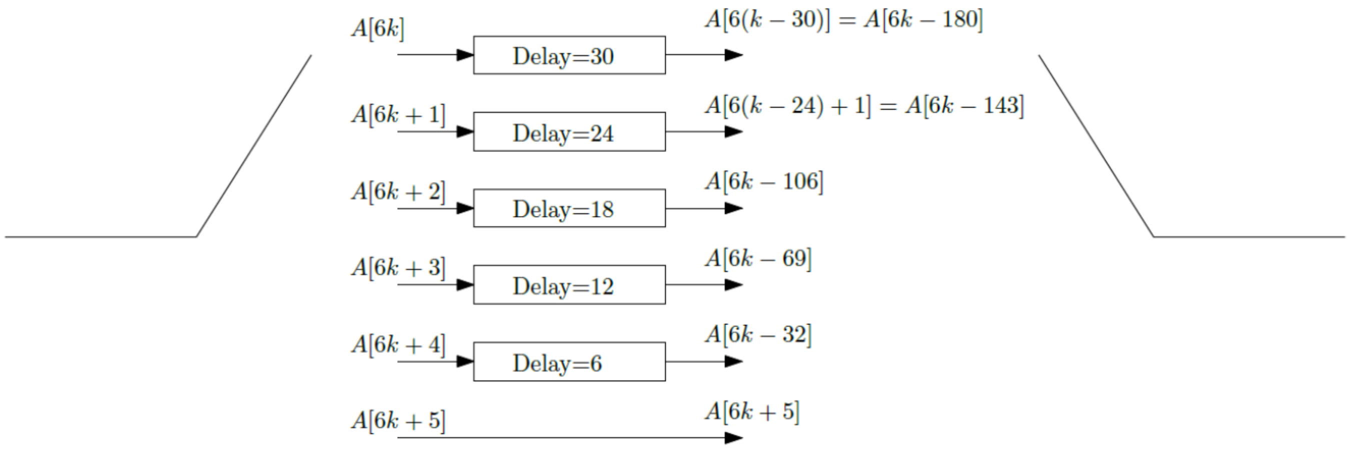
Illustration : 400G/800G mode (2 way interleaved) Convolutional Interleaver

- **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

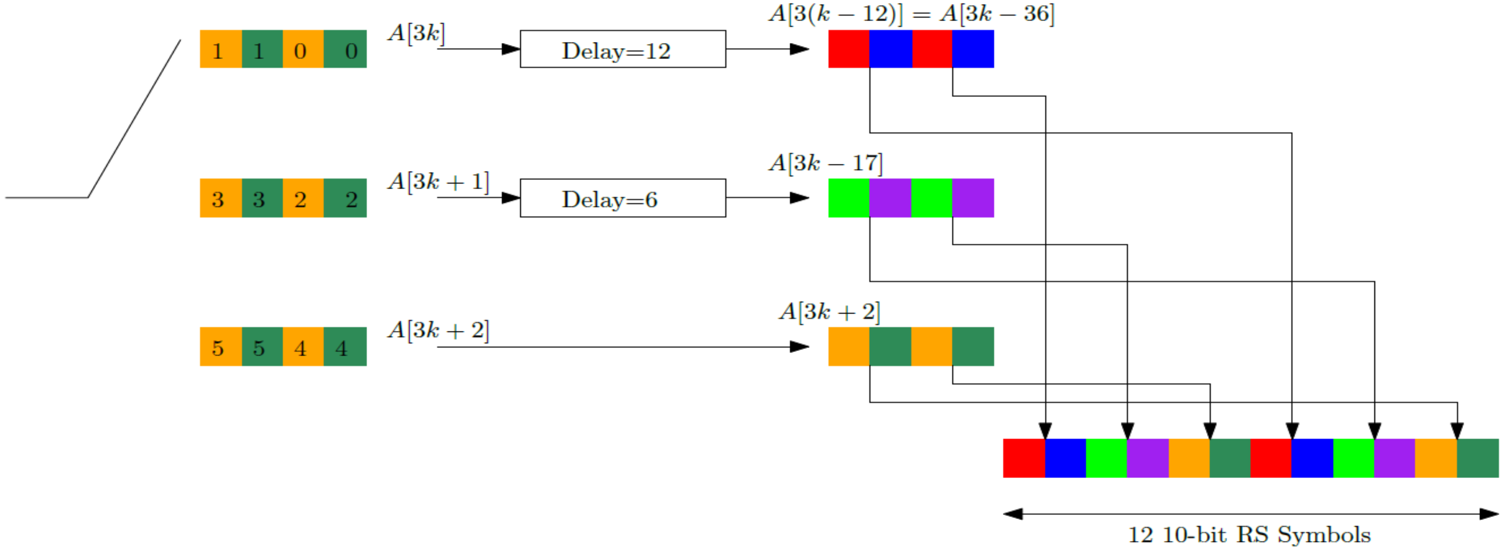
— Synchronization at RX is **implied** by Hamming codeword boundaries (i.e., Hamming sync implies CI sync)



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

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

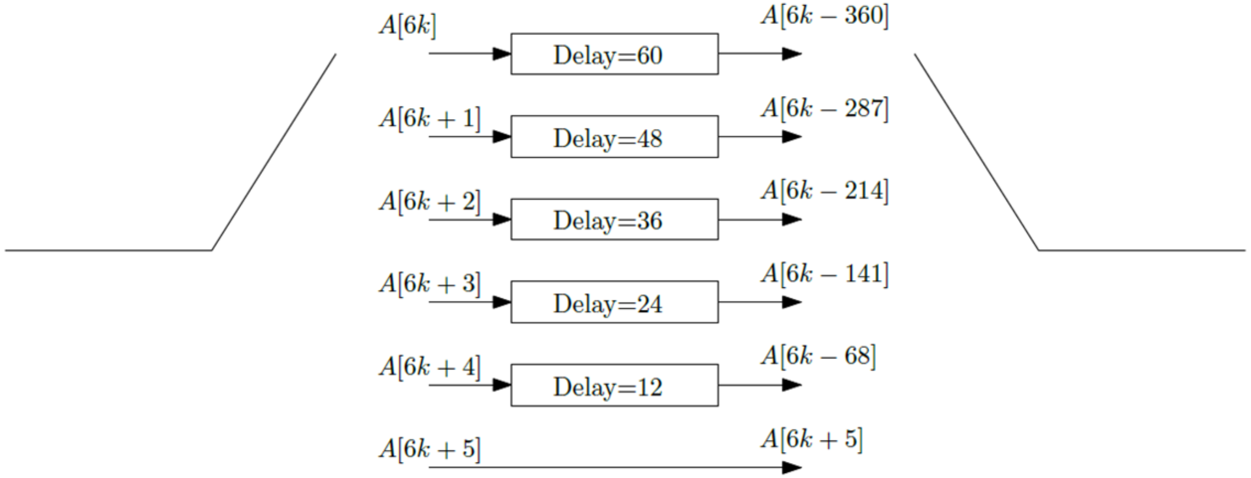
- **40b** (FEC_A,FEC_B, FEC_A,FEC_B) symbols represented by **A[m]**
  - Delay lines operate on 40b symbols
- **Total Latency (CI+CDI): 36x40=1440b @ 25G**



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])**



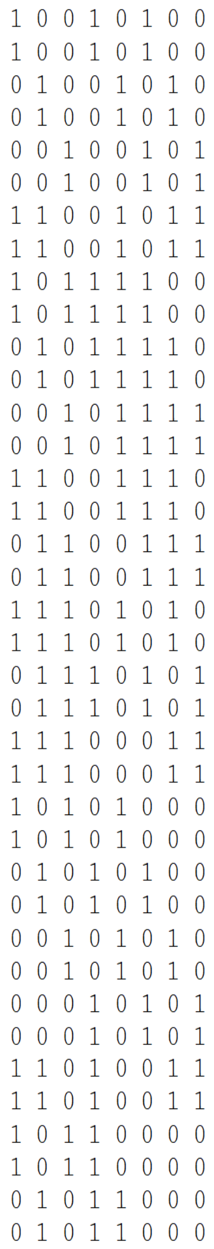
Translating this to Parametrized CI will result in W = 2, P = 6, D = 12

Convolutional Interleaver + Inner Code (128,120) Latency for 200G per Lane

Client Type	Parameters for Interleaver	FEC	Decoder Input BER	Latency
800GBASE-R (2 way interleaved)	W=2,P=6,D=6	KP4 + 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
200GBASE-R (2 way interleaved)	W =2,P =6,D =12			~280ns

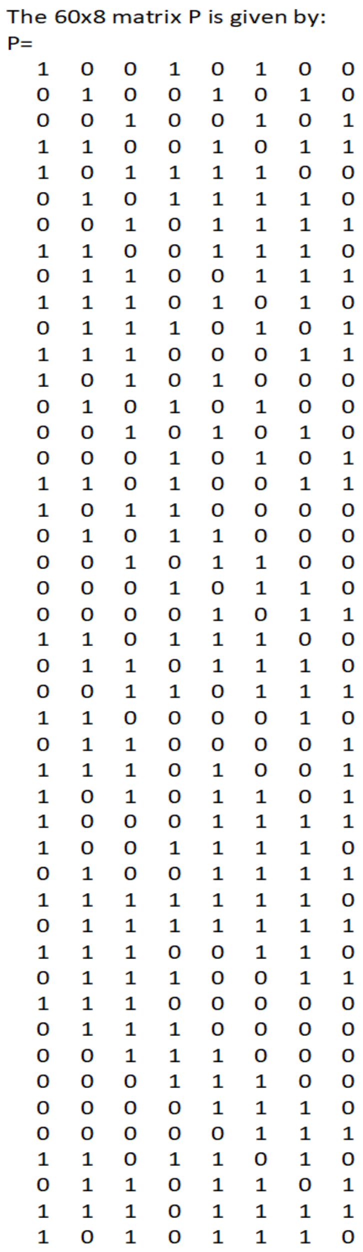
- BER thresholds are simulated using AWGN model. The CI helps with randomly distributing equalizer burst errors which results in AWGN model BER estimate to be accurate. See: https://www.ieee802.org/3/df/public/22_05/22_0517/bliss_3df_01a_220517.pdf
  - For channels with correlated errors, see: https://www.ieee802.org/3/df/public/22_10/22_1005/bliss_3df_01_220929.pdf
- The AUI symbol Muxing scheme will have no effect on the performance of the concatenated scheme
- The 200G latency can be reduced in half in exchange for 0.3dB/4.0E-3 AWGN coding gain penalty

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

Summary

- ❑ We presented a complete inner code scheme that enables 200G/Lambda Optical IM-DD links.

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

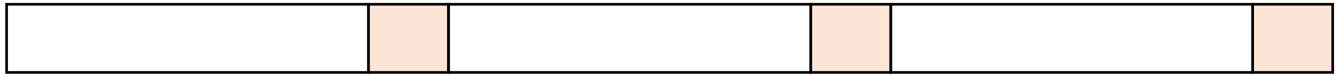
- ❑ The proposed inner code is a low complexity solution that works in conjunction with the existing KP FEC to act as a booster to the overall coding gain of the end-to-end link.

- ❑ Leveraging the existing KP4 FEC for 200G AUI will benefit the industry and will ease the backward compatibility issues.

Appendix A: Padding bits

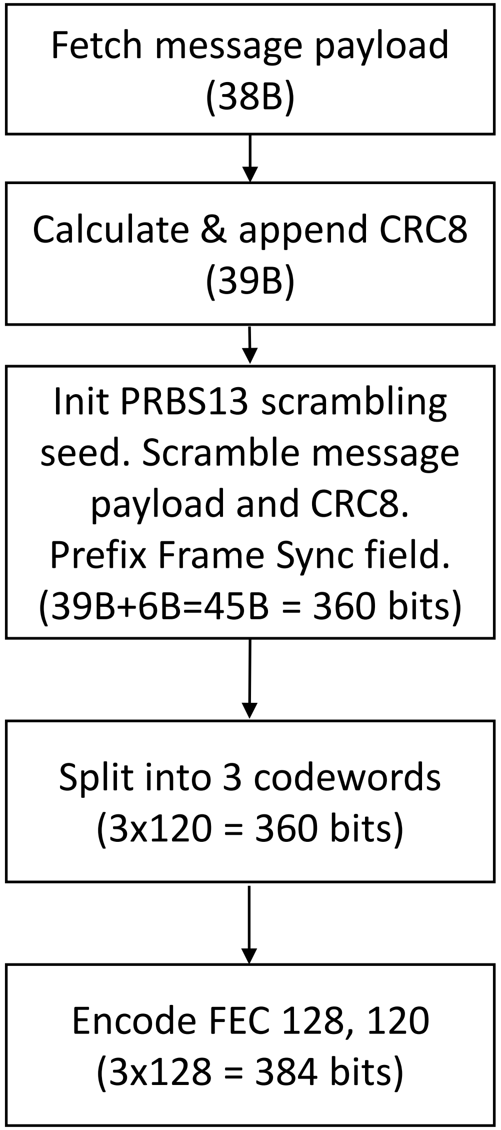
Padding Specification

- **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



Fetch message payload (38B)

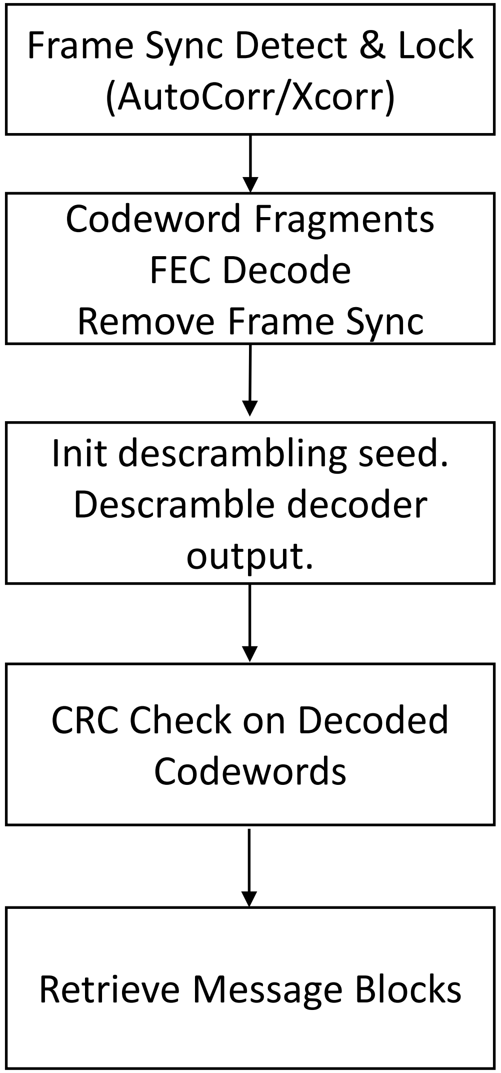
Calculate & append CRC8 (39B)

Init PRBS13 scrambling seed. Scramble message payload and CRC8. Prefix Frame Sync field. (39B+6B=45B = 360 bits)

Split into 3 codewords (3x120 = 360 bits)

Encode FEC 128, 120 (3x128 = 384 bits)

Padding Field Consumption – Reference Implementation (Informative)



Frame Sync Detect & Lock (AutoCorr/Xcorr)

Codeword Fragments
FEC Decode
Remove Frame Sync

Init descrambling seed. Descramble decoder output.

CRC Check on Decoded Codewords

Retrieve Message Blocks

Thanks !