4x RS Codeword Interleaving Proposal for 200 GbE and 400 GbE
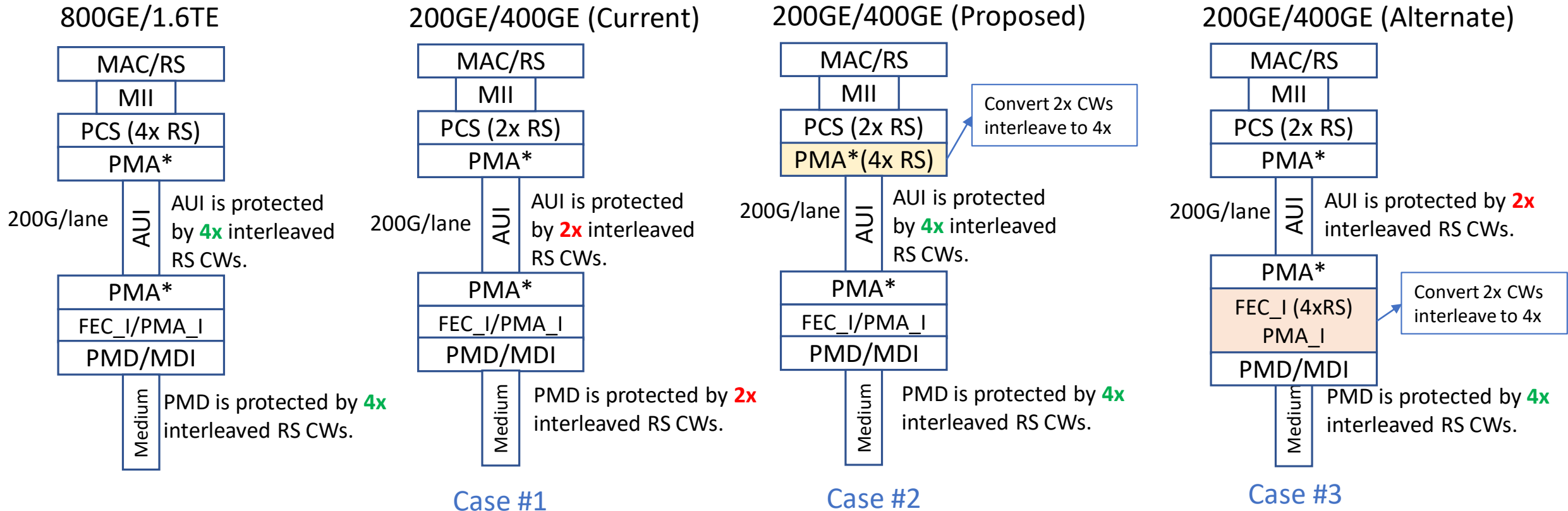
Xiang He, Hao Ren, Matt Brown (Huawei)

Adee Ran (Cisco)

Introduction

- P802.3dj covers various Ethernet rates for 200GE, 400GE, 800GE and 1.6TE.
    - Each rate needs to support 200Gb/s per lane AUIs and PMDs.
- All PCS designs for the above rates have been determined in the Task Force.
    - 200GE and 400GE both interleave 2x RS codewords in the PCS.
    - 800GE and 1.6TE both interleave 4x RS codewords in the PCS/PMA.
- Symbol-pair muxing PMA has been adopted for 200G/lane AUIs.
    - Please see ran_3dj_01a_2303.
- This presentation is a preliminary analysis on the performance difference due to different number of RS codewords for Type 2 PHY scheme.
    - This analysis is specific for Type 2 optical links with up to 2 AUIs at each end.
    - Type 1 PHY scheme needs to be addressed as well, but not covered in this presentation.
- Independent analysis in ran_3dj_elec_01_230420 also showed performance difference of 2/4 RS codeword interleaving.

Interleaving Depth vs AUI Burst Protection

800GE/1.6TE

| MAC/RS |
| MII |
| PCS (4x RS) |
| PMA* |

200G/lane — AUI

AUI is protected by **4x** interleaved RS CWs.

| PMA* |
| FEC_I/PMA_I |
| PMD/MDI |

Medium

PMD is protected by **4x** interleaved RS CWs.

200GE/400GE (Current)

| MAC/RS |
| MII |
| PCS (2x RS) |
| PMA* |

200G/lane — AUI

AUI is protected by **2x** interleaved RS CWs.

| PMA* |
| FEC_I/PMA_I |
| PMD/MDI |

Medium

PMD is protected by **2x** interleaved RS CWs.

Case #1

200GE/400GE (Proposed)

| MAC/RS |
| MII |
| PCS (2x RS) |
| PMA*(4x RS) |

Convert 2x CWs interleave to 4x

200G/lane — AUI

AUI is protected by **4x** interleaved RS CWs.

| PMA* |
| FEC_I/PMA_I |
| PMD/MDI |

Medium

PMD is protected by **4x** interleaved RS CWs.

Case #2

200GE/400GE (Alternate)

| MAC/RS |
| MII |
| PCS (2x RS) |
| PMA* |

200G/lane — AUI

AUI is protected by **2x** interleaved RS CWs.

| PMA* |
| FEC_I (4xRS) PMA_I |
| PMD/MDI |

Convert 2x CWs interleave to 4x

Medium

PMD is protected by **4x** interleaved RS CWs.

Case #3

*Symbol-pair muxing PMA as in ran_3dj_01a_2303.

- Assuming same AUI BER threshold for all Ethernet rates using 200G/lane AUIs, the current 200GE/400GE will have lower performance due to 2x RS CWs interleaving compared to 4x of 800GE/1.6TE.

# Two-part Link Simulation for 2x and 4x RS Codewords Interleave

- Case #1 (dotted lines): 2x RS

- Case #2 (solid lines): 4x RS(host)

- Increasing the interleaving to 4x RS codewords on the AUIs (Case # 2) is clearly beneficial.
  - The AUI electrical specs and/or the PMD's optical specs will be driven by the worst case.
  - It would be preferable if we can avoid burdening the AUI and PMD for the older rates.

- Latency will be increased but is reasonable as shown in brown_3dj_elec_01_230420.

- Simulation configuration:

AUI: 200G/lane, symbol-pair muxing.
Error propagation probability "a" = 0.75, pre-coding ON.
All BER values includes additional errors due to bursts and precoding.

PMD: Pure AWGN.
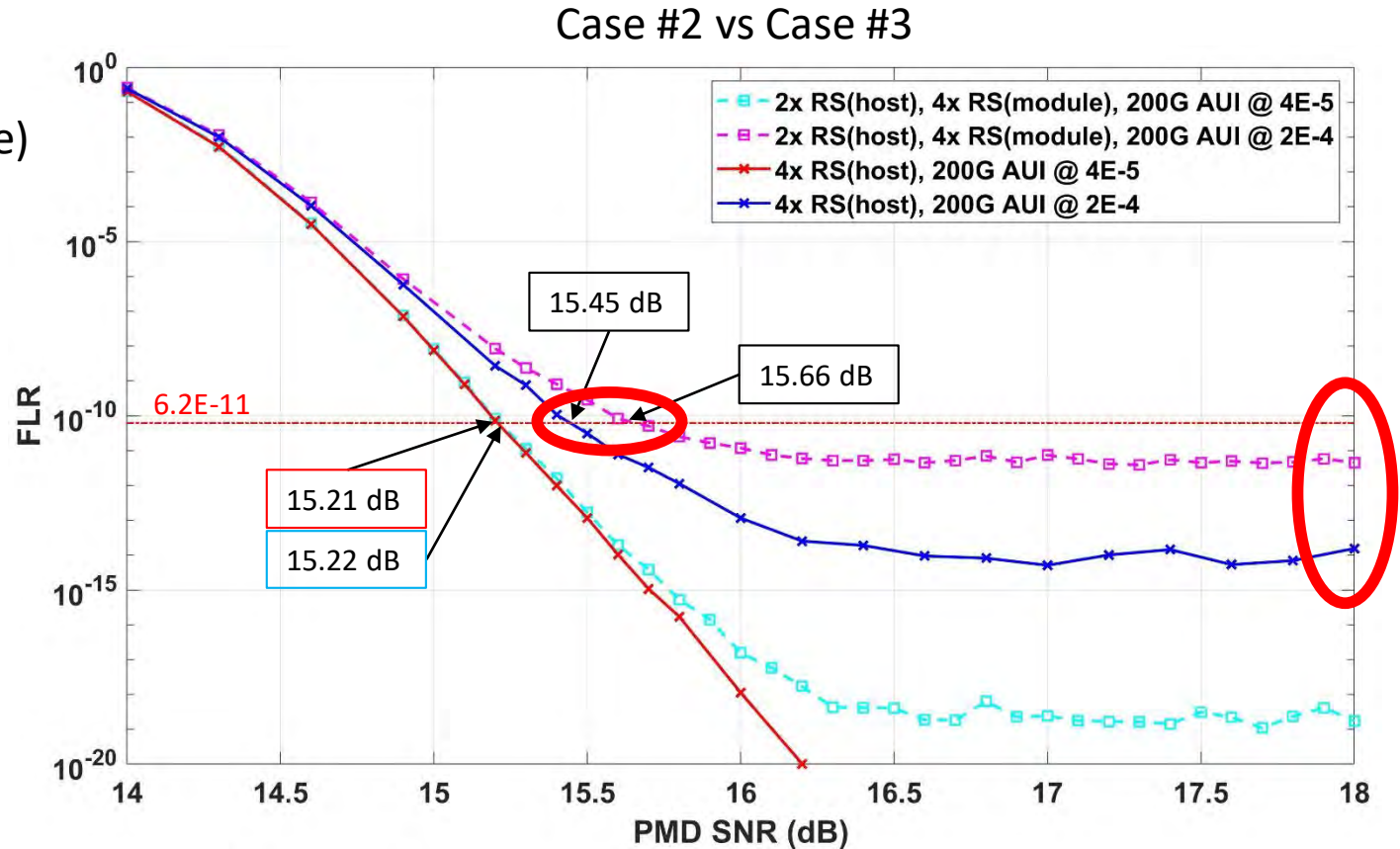
FEC_I: Hamming(128,120), w/o convolutional interleaver.

Case #1 vs Case #2
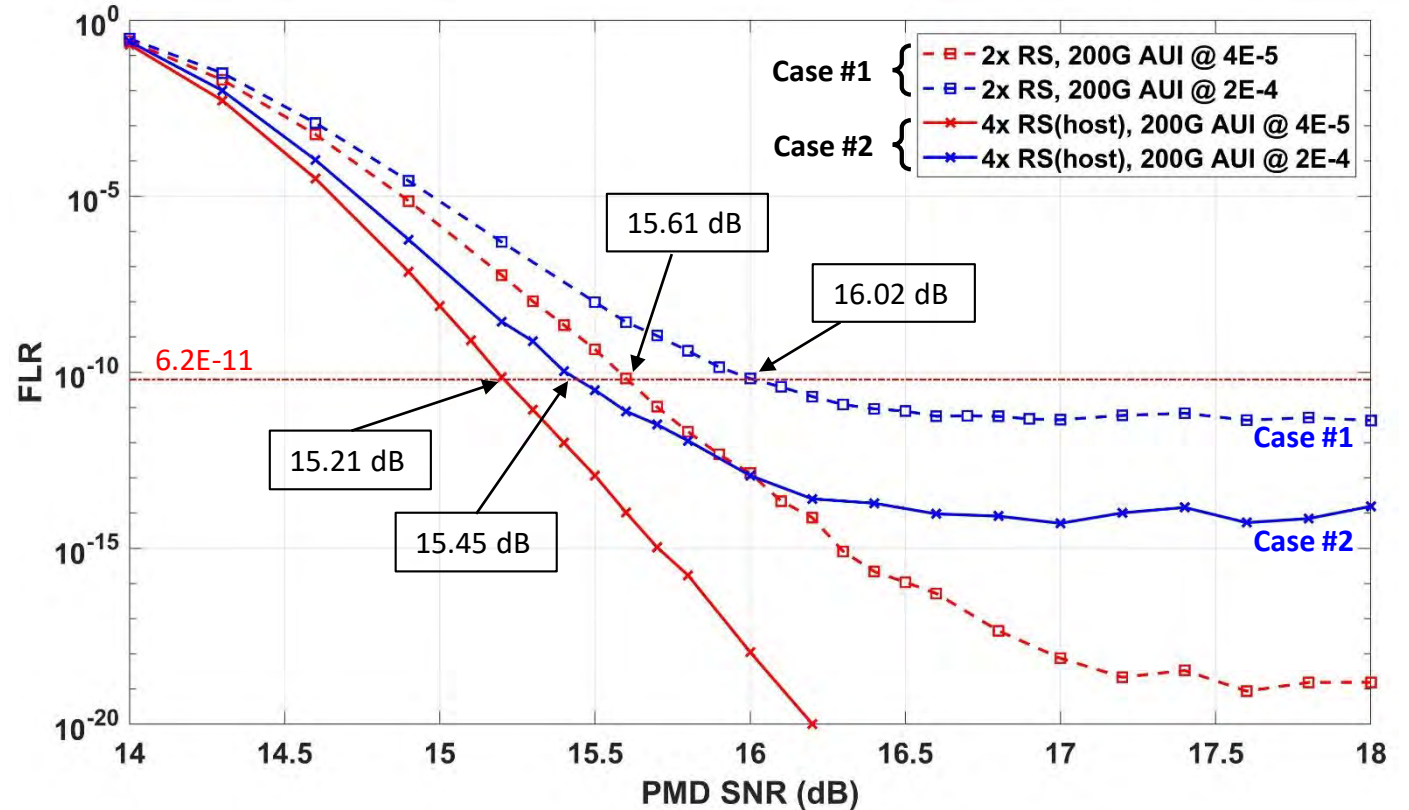
AUI BER (per PHY)	Total AUI BER	PMD BER Threshold	
		Case #1	Case #2
2E-5	4E-5	2.62E-3	3.74E-3
1E-4	2E-4	1.76E-3	3.03E-3

Alternative Location to Convert 2x Codewords Interleave to 4x

- Case #2 (solid lines): 4x RS(host)

- Case #3 (dotted lines): 2x RS(host), 4x RS(module)

- Case #3 has very little margin below the FLR target when the total AUI BER is 2E-4.
  - This leaves little room for PMD BER.
  - Case #2 improves this margin by ~3 orders.
  - Case #2 has 0.2dB higher coding gain than Case #3.

- Simulation configuration:

AUI: 200G/lane, symbol-pair muxing.
Error propagation probability "a" = 0.75, pre-coding ON.
All BER values includes additional errors due to bursts and precoding.

PMD: Pure AWGN.

FEC_I: Hamming(128,120), w/o convolutional interleaver.

Case #2 vs Case #3

| AUI BER (per PHY)	Total AUI BER	PMD BER Threshold	
		Case #2	Case #3
2E-5	4E-5	3.74E-3	3.71E-3
1E-4	2E-4	3.03E-3	2.50E-3

Option 1 – Delay Half of PCS Lanes

- On Tx side, PCS lanes with odd index ("*p*" as shown below) are delayed by 2 codewords.
  - Each odd index PCS lane is delayed by 10880/8 = 1360 bits.
  - PMA performs symbol-pair muxing between even and odd lanes to form 4x RS CWs interleaving pattern (ABA'B' as shown below).
- On Rx side, deskew is performed on all lanes in the PCS and no special handling is required.
- Total delay penalty is 2 codewords for Tx + Rx.
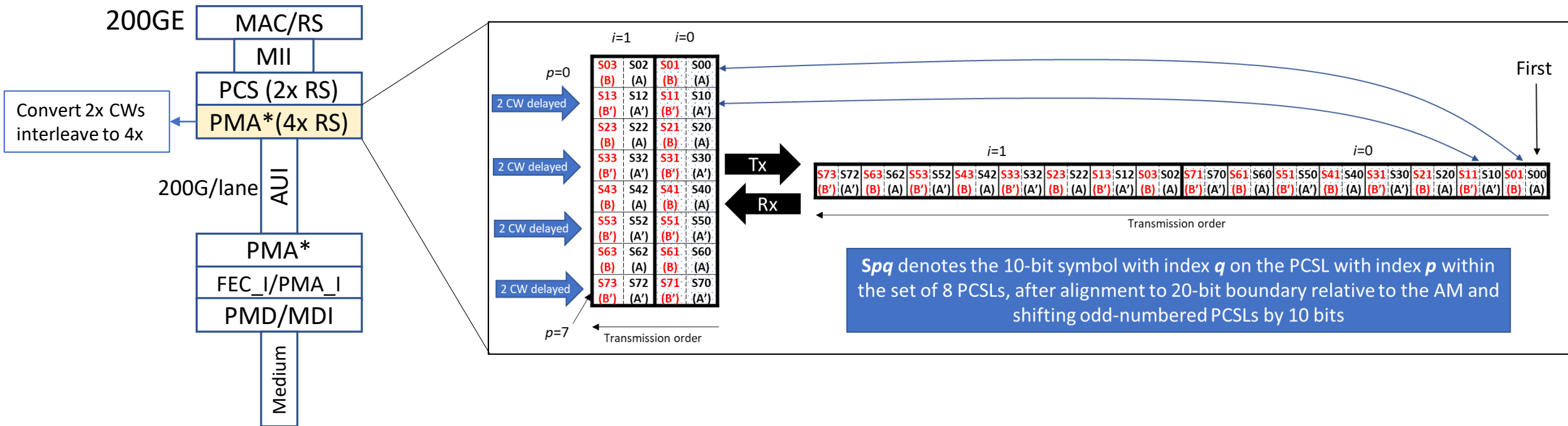
Option 1 – Delay Half of PCS Lanes, continued

- When interoperating with 100G/lane AUIs, the 2x to 4x RS CWs conversion can be done after PMA symbol-pair muxing function.
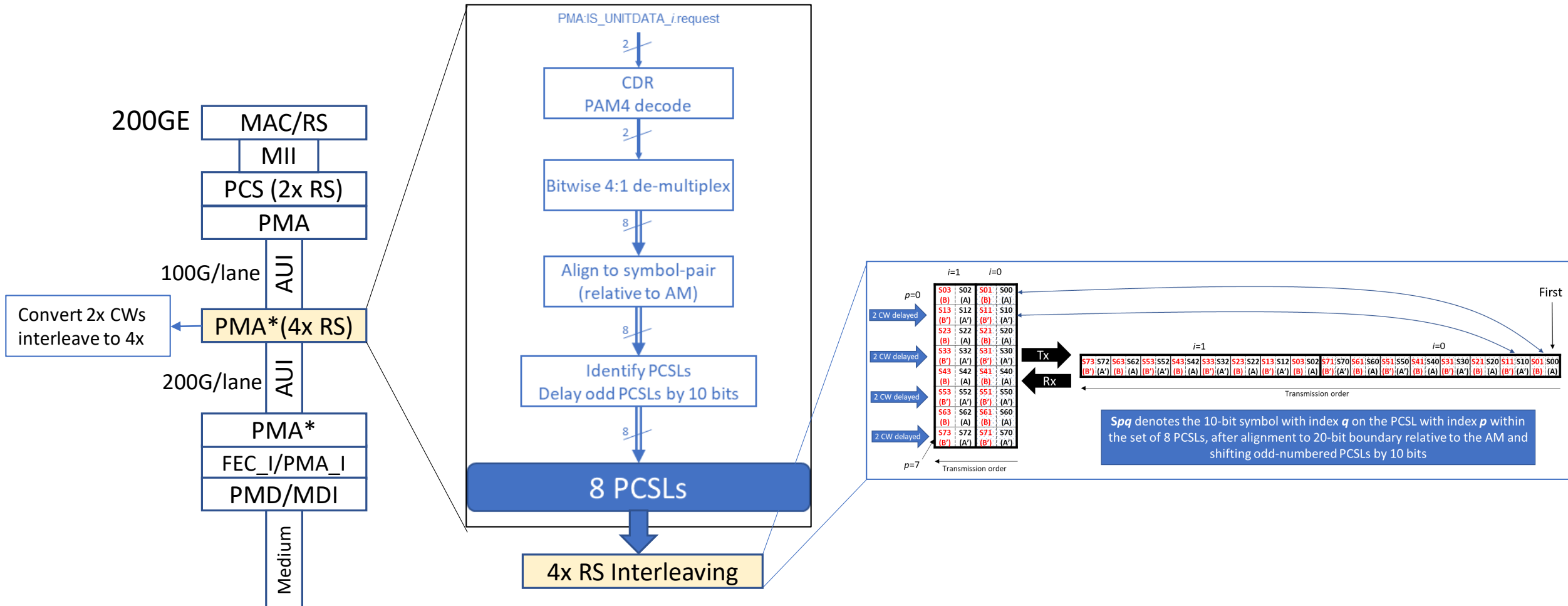
Option 2 – Convolutional Interleaver on 200G PMA Lanes

- Performed on symbol-pair muxed PMA lanes.

- 2-branch convolutional interleaver, with 1 branch delayed by 5440 bits on Tx side.
  - Rx side deskew is performed on all lanes in the PCS and no special handling is required.
  - Total delay penalty is 2 codewords for Tx + Rx.

- For 100G AUIs, performing the symbol-pair mux first.

- Can be interoperable with option 1.
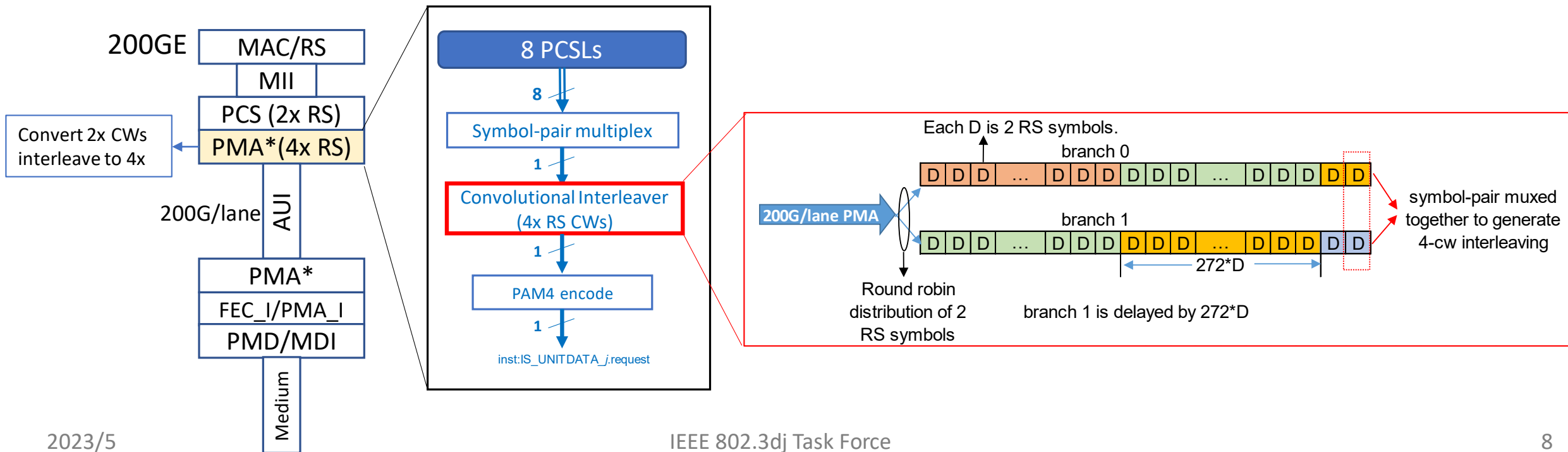
Summary

- Increasing interleaving depth from 2x codewords to 4x for 200GE and 400GE could improve overall coding gain by 0.57dB, if total AUI BER is 1E-4 on each side.

  - 2x codewords interleaving on AUI leads to error floor around 1E-11 FLR.

  - Increasing the interleaving depth in the host rather than in the module is preferred.

  - Latency impact is 51.2ns for 200GE, and 25.6ns for 400GE.

- If AUI BER is at 1E-5 per segment, increasing interleaving depth from 2x to 4x for 200GE and 400GE could improve overall coding gain by 0.4dB.

  - Increasing the interleaving depth in the host or module are both OK.

- The two options proposed could provide 4x codewords interleaving in the PMA, with a total (Tx+Rx) added latency of 2 codewords.

Thank you!