Comparison of PCS proposals with focus on burst immunity

Adee Ran, Cisco

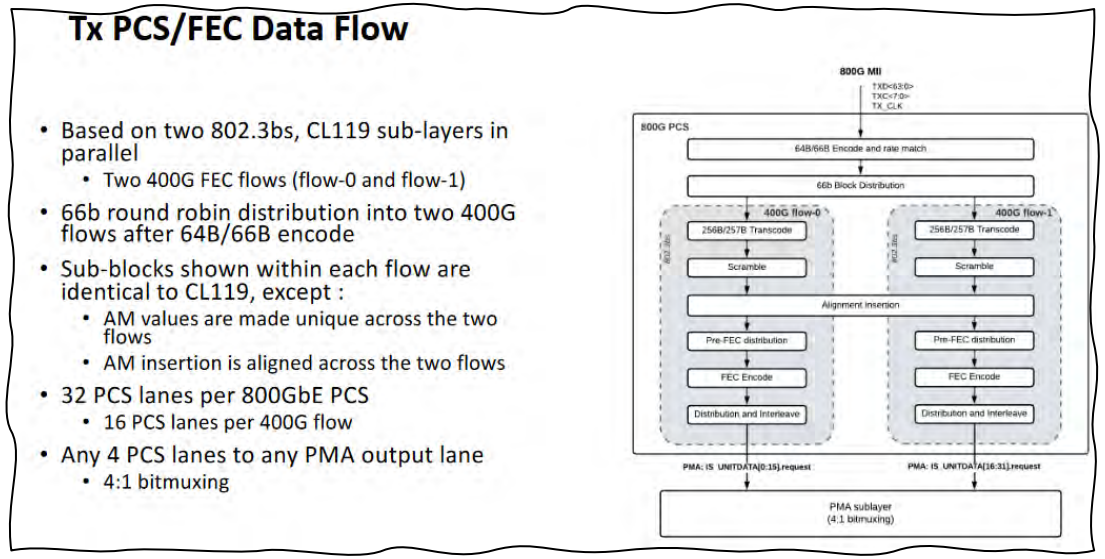
Outline

- Outline of proposals so far for 800GBASE-R PCS
- Why real-life tolerable BER is lower than the standard allows
- Considerations for choosing a PCS

Two proposals presented so far

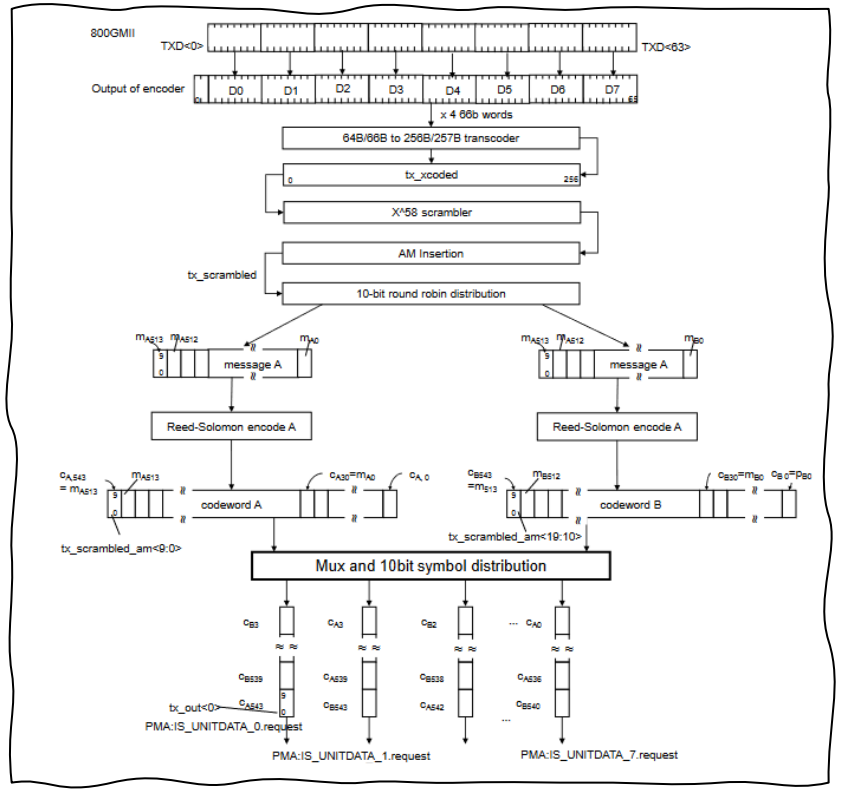
32×25G PCS lanes (2×400GBASE-R PCS, aka "2×400GBASE-R")

shrikhande_3df_01a_220517 slide 10



8×100G PCS lanes (aka "Sped-up 200GBASE-R")

bruckman_3df_01_220308 slide 7



Detailed comparison was provided in he_3df_01_220517

What is the maximum BER?

- BER specifications of PMDs and AUIs are given based on the FEC performance analysis, assuming uncorrelated (“random”) errors
- With the maximum BER on all sub-links, assuming uncorrelated errors, the end-to-end link can be below the FLR target (6.2×10^{-11})
  - See analysis in opsasnick_3df_logic_220630a
- But…

124.1.1 Bit error ratio

The bit error ratio (BER) when processed according to Clause 120 shall be less than 2.4×10^{-4} provided that the error statistics are sufficiently random that this results in a frame loss ratio (see 1.4.344) of less than 1.7×10^{-12} for 64-octet frames with minimum interpacket gap when processed according to Clause 120 and then Clause 119. For a complete Physical Layer, the frame loss ratio may be degraded to 6.2×10^{-11} for 64-octet frames with minimum interpacket gap due to additional errors from the electrical interfaces.

120G.1.1 Bit error ratio

The bit error ratio (BER) when processed by the PMA according to Clause 135 for 100GAUI-1 C2M or Clause 120 for 200GAUI-2 or 400GAUI-4 C2M shall be less than 10^{-5}.

Tolerable BER in real life is lower

- It is widely known that considerable margin in BER is required
- Why?
  - Correlated errors can’t be ignored
  - The standard’s FLR target is not acceptable is many use cases

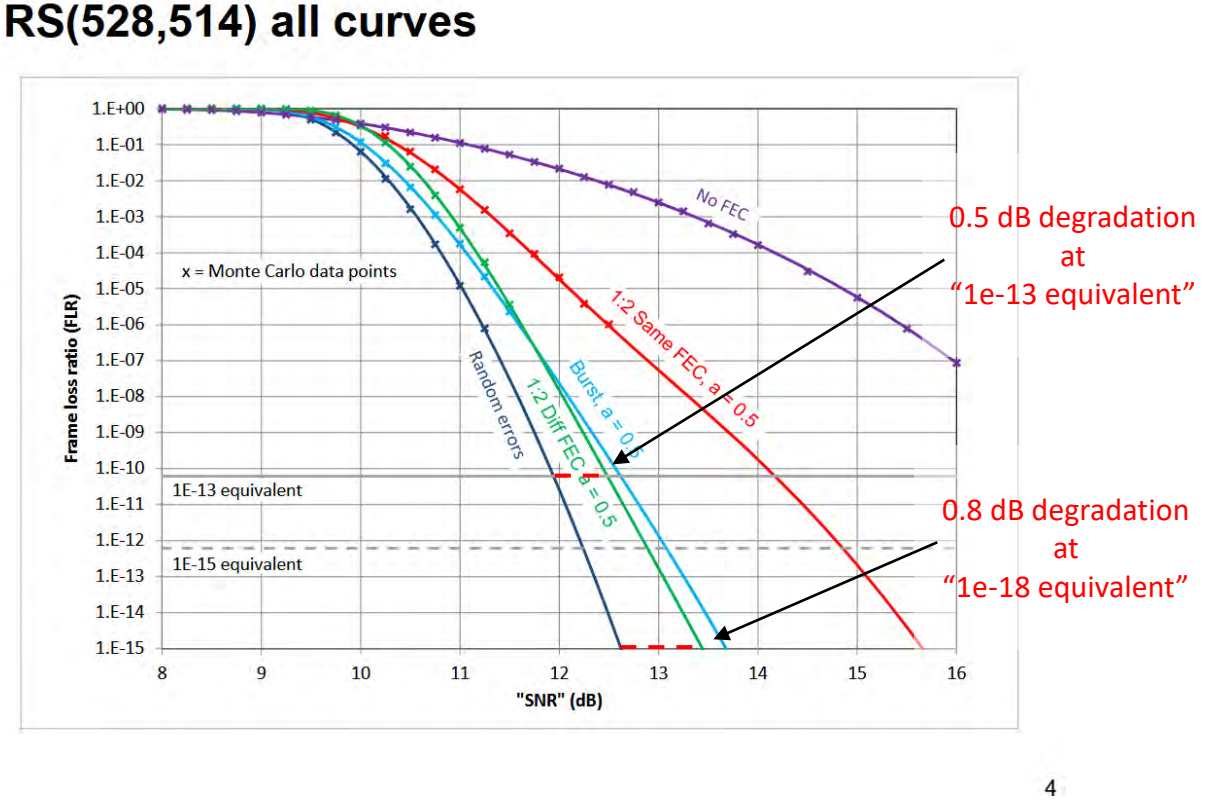
Correlated errors

- Many receiver designs include a DFE
  - Almost universal with CR/KR links
  - Part of the reference receiver in 100G/lane AUIs
  - DFE is used in some optical receivers (although we never accounted for that)
  - Even small DFE coefficients create error bursts occasionally
  - Precoding can mitigate DFE bursts, but is not defined for optical links
- There are other sources of correlated errors
  - Often not traceable to a single issue in one component
  - Very difficult to debug
- Even mild cases of correlated errors change the slope of the waterfall curve considerably
  - Lower BER/DER (or higher SNR) is required to achieve the same FLR
  - Coding gain is degraded
  - By how much?

How much margin is enough

- At 800 Gb/s, “equivalent of BER=1e-13” means an error every 12 seconds!
  - Unacceptable in many cases
- ~1e-18 is a common expectation
  - Error every two weeks
- The degradation in coding gain is larger when the target performance is better!
  - The waterfall may even start “flattening” going to the right
- FEC scheme should consider coding gain loss at better expected performance

Typical example
(Source: anslow_3bs_02_1114)



4

My parameters for comparing proposals

- Performance with uncorrelated errors
  - pre-FEC DER and AWGN SNR required to meet FLR objective
- Performance with correlated errors (strong DFE)
  - With 100 Gb/s per lane (8-lane AUI/PMD)
  - With 200 Gb/s per lane (4-lane AUI/PMD)
- Ease of implementation
  - Considering backward compatibility
  - Considering breakout, possibly in a module
  - Considering existing pre-standard implementations…
- Latency

Burst immunity at 100G/lane (illustrated)

2×400GBASE-R

UI\Lane	0	1	2	3	4	5	6	7
22	B81 A81	B91 A91	B101 A101	B111 A111	B121 A121	B131 A131	B141 A141	B151 A151
21	D80 C80	D90 C90	D100 C100	D110 C110	D120 C120	D130 C130	D140 C140	D150 C150
20	B80 A80	B90 A90	B100 A100	B110 A110	B120 A120	B130 A130	B140 A140	B150 A150
19	C9 D9	C19 D19	C29 D29	C39 D39	C49 D49	C59 D59	C69 D69	C79 D79
18	A9 B9	A19 B19	A29 B29	A39 B39	A49 B49	A59 B59	A69 B69	A79 B79
17	C8 D8	C18 D18	C28 D28	C38 D38	C48 D48	C58 D58	C68 D68	C78 D78
16	A8 B8	A18 B18	A28 B28	A38 B38	A48 B48	A58 B58	A68 B68	A78 B78
15	C7 D7	C17 D17	C27 D27	C37 D37	C47 D47	C57 D57	C67 D67	C77 D77
14	A7 B7	A17 B17	A27 B27	A37 B37	A47 B47	A57 B57	A67 B67	A77 B77
13	C6 D6	C16 D16	C26 D26	C36 D36	C46 D46	C56 D56	C66 D66	C76 D76
12	A6 B6	A16 B16	A26 B26	A36 B36	A46 B46	A56 B56	A66 B66	A76 B76
11	C5 D5	C15 D15	C25 D25	C35 D35	C45 D45	C55 D55	C65 D65	C75 D75
10	A5 B5	A15 B15	A25 B25	A35 B35	A45 B45	A55 B55	A65 B65	A75 B75
9	C4 D4	C14 D14	C24 D24	C34 D34	C44 D44	C54 D54	C64 D64	C74 D74
8	A4 B4	A14 B14	A24 B24	A34 B34	A44 B44	A54 B54	A64 B64	A74 B74
7	C3 D3	C13 D13	C23 D23	C33 D33	C43 D43	C53 D53	C63 D63	C73 D73
6	A3 B3	A13 B13	A23 B23	A33 B33	A43 B43	A53 B53	A63 B63	A73 B73
5	C2 D2	C12 D12	C22 D22	C32 D32	C42 D42	C52 D52	C62 D62	C72 D72
4	A2 B2	A12 B12	A22 B22	A32 B32	A42 B42	A52 B52	A62 B62	A72 B72
3	C1 D1	C11 D11	C21 D21	C31 D31	C41 D41	C51 D51	C61 D61	C71 D71
2	A1 B1	A11 B11	A21 B21	A31 B31	A41 B41	A51 B51	A61 B61	A71 B71
1	C0 D0	C10 D10	C20 D20	C30 D30	C40 D40	C50 D50	C60 D60	C70 D70
0	A0 B0	A10 B10	A20 B20	A30 B30	A40 B40	A50 B50	A60 B60	A70 B70

Sped-up 200GBASE-R

UI\Lane	0	1	2	3	4	5	6	7
14	A88 A89	B88 B89	A98 A99	B98 B99	A108 A109	B108 B109	A118 A119	B118 B119
13	A86 A87	B86 B87	A96 A97	B96 B97	A106 A107	B106 B107	A116 A117	B116 B117
12	A84 A85	B84 B85	A94 A95	B94 B95	A104 A105	B104 B105	A114 A115	B114 B115
11	A82 A83	B82 B83	A92 A93	B92 B93	A102 A103	B102 B103	A112 A113	B112 B113
10	A80 A81	B80 B81	A90 A91	B90 B91	A100 A101	B100 B101	A110 A111	B110 B111
9	B48 B49	A48 A49	B58 B59	A58 A59	B68 B69	A68 A69	B78 B79	A78 A79
8	B46 B47	A46 A47	B56 B57	A56 A57	B66 B67	A66 A67	B76 B77	A76 A77
7	B44 B45	A44 A45	B54 B55	A54 A55	B64 B65	A64 A65	B74 B75	A74 A75
6	B42 B43	A42 A43	B52 B53	A52 A53	B62 B63	A62 A63	B72 B73	A72 A73
5	B40 B41	A40 A41	B50 B51	A50 A51	B60 B61	A60 A61	B70 B71	A70 A71
4	A8 A9	B8 B9	A18 A19	B18 B19	A28 A29	B28 B29	A38 A39	B38 B39
3	A6 A7	B6 B7	A16 A17	B16 B17	A26 A27	B26 B27	A36 A37	B36 B37
2	A4 A5	B4 B5	A14 A15	B14 B15	A24 A25	B24 B25	A34 A35	B34 B35
1	A2 A3	B2 B3	A12 A13	B12 B13	A22 A23	B22 B23	A32 A33	B32 B33
0	A0 A1	B0 B1	A10 A11	B10 B11	A20 A21	B20 B21	A30 A31	B30 B31

2 symbols per CW:
Possible for $3 \leq L \leq 22$

2 symbols per CW:
Guaranteed for $7 \leq L \leq 16$

Minimum burst that might impacts 2 symbols: shorter in “2×400GBASE-R”
Minimum burst that is guaranteed to impact 2 symbols: shorter in “sped-up 200GBASE-R”

We should not ignore 200G/lane

- If we defer the analysis of the PCS FEC with 200G/lane (e.g. a future 800GAUI-4), we may face unexpected problems in the next project
  - Will we redefine the PCS?
- The FEC defined in 802.3df should be a foundation for re-use

Burst immunity at 200G/lane (bit muxing)

2×400GBASE-R

2 symbols per CW: $3 \leq L \leq 4$

3 symbols per CW: $5 \leq L \leq 6$

4 symbols per CW: $7 \leq L \leq 42$

5 symbols per CW: $43 \leq L \leq 62$

…

UI\Lane	0	1	2	3
44	B81 A81	B101 A101	B121 A121	B141 A141
43	D90 C90	D110 C110	D130 C130	D150 C150
42	B90 A90	B110 A110	B130 A130	B150 A150
41	D80 C80	D100 C100	D120 C120	D140 C140
40	B80 A80	B100 A100	B120 A120	B140 A140
39	C19 D19	C39 D39	C59 D59	C79 D79
38	A19 B19	A39 B39	A59 B59	A79 B79
37	C9 D9	C29 D29	C49 D49	C69 D69
36	A9 B9	A29 B29	A49 B49	A69 B69
35	C18 D18	C38 D38	C58 D58	C78 D78
34	A18 B18	A38 B38	A58 B58	A78 B78
33	C8 D8	C28 D28	C48 D48	C68 D68
32	A8 B8	A28 B28	A48 B48	A68 B68
31	C17 D17	C37 D37	C57 D57	C77 D77
30	A17 B17	A37 B37	A57 B57	A77 B77
29	C7 D7	C27 D27	C47 D47	C67 D67
28	A7 B7	A27 B27	A47 B47	A67 B67
27	C16 D16	C36 D36	C56 D56	C76 D76
26	A16 B16	A36 B36	A56 B56	A76 B76
25	C6 D6	C26 D26	C46 D46	C66 D66
24	A6 B6	A26 B26	A46 B46	A66 B66
23	C15 D15	C35 D35	C55 D55	C75 D75
22	A15 B15	A35 B35	A55 B55	A75 B75
21	C5 D5	C25 D25	C45 D45	C65 D65
20	A5 B5	A25 B25	A45 B45	A65 B65
19	C14 D14	C34 D34	C54 D54	C74 D74
18	A14 B14	A34 B34	A54 B54	A74 B74
17	C4 D4	C24 D24	C44 D44	C64 D64
16	A4 B4	A24 B24	A44 B44	A64 B64
15	C13 D13	C33 D33	C53 D53	C73 D73
14	A13 B13	A33 B33	A53 B53	A73 B73
13	C3 D3	C23 D23	C43 D43	C63 D63
12	A3 B3	A23 B23	A43 B43	A63 B63
11	C12 D12	C32 D32	C52 D52	C72 D72
10	A12 B12	A32 B32	A52 B52	A72 B72
9	C2 D2	C22 D22	C42 D42	C62 D62
8	A2 B2	A22 B22	A42 B42	A62 B62
7	C11 D11	C31 D31	C51 D51	C71 D71
6	A11 B11	A31 B31	A51 B51	A71 B71
5	C1 D1	C21 D21	C41 D41	C61 D61

Sped-up 200GBASE-R

2 symbols per CW: $2 \leq L \leq 12$

3 symbols per CW: $13 \leq L \leq 22$

4 symbols per CW: $23 \leq L \leq 32$
5 symbols per CW: $33 \leq L \leq 42$

…

UI\Lane	0	1	2	3
23	B82 B83	B92 B93	B102 B103	B112 B113
22	A82 A83	A92 A93	A102 A103	A112 A113
21	B80 B81	B90 B91	B100 B101	B110 B111
20	A80 A81	A90 A91	A100 A101	A110 A111
19	A48 A49	A58 A59	A68 A69	A78 A79
18	B48 B49	B58 B59	B68 B69	B78 B79
17	A46 A47	A56 A57	A66 A67	A76 A77
16	B46 B47	B56 B57	B66 B67	B76 B77
15	A44 A45	A54 A55	A64 A65	A74 A75
14	B44 B45	B54 B55	B64 B65	B74 B75
13	A42 A43	A52 A53	A62 A63	A72 A73
12	B42 B43	B52 B53	B62 B63	B72 B73
11	A40 A41	A50 A51	A60 A61	A70 A71
10	B40 B41	B50 B51	B60 B61	B70 B71
9	B8 B9	B18 B19	B28 B29	B38 B39
8	A8 A9	A18 A19	A28 A29	A38 A39

Probability of burst impacting 2 or more symbols is too high in both proposals

Burst immunity at 200G/lane (symbol muxing)

2×400GBASE-R

2 symbols per CW: $17 \leq L \leq 36$

3 symbols per CW: $37 \leq L \leq 56$

UI\Lane	0	1	2	3
39	D48 D49	D58 D59	D68 D69	D78 D79
38	D46 D47	D56 D57	D66 D67	D76 D77
37	D44 D45	D54 D55	D64 D65	D74 D75
36	D42 D43	D52 D53	D62 D63	D72 D73
35	D40 D41	D50 D51	D60 D61	D70 D71
34	C48 C49	C58 C59	C68 C69	C78 C79
33	C46 C47	C56 C57	C66 C67	C76 C77
32	C44 C45	C54 C55	C64 C65	C74 C75
31	C42 C43	C52 C53	C62 C63	C72 C73
30	C40 C41	C50 C51	C60 C61	C70 C71
29	B48 B49	B58 B59	B68 B69	B78 B79
28	B46 B47	B56 B57	B66 B67	B76 B77
27	B44 B45	B54 B55	B64 B65	B74 B75
26	B42 B43	B52 B53	B62 B63	B72 B73
25	B40 B41	B50 B51	B60 B61	B70 B71
24	A48 A49	A58 A59	A68 A69	A78 A79
23	A46 A47	A56 A57	A66 A67	A76 A77
22	A44 A45	A54 A55	A64 A65	A74 A75
21	A42 A43	A52 A53	A62 A63	A72 A73
20	A40 A41	A50 A51	A60 A61	A70 A71
19	D8 D9	D18 D19	D28 D29	D38 D39
18	D6 D7	D16 D17	D26 D27	D36 D37
17	D4 D5	D14 D15	D24 D25	D34 D35
16	D2 D3	D12 D13	D22 D23	D32 D33
15	D0 D1	D10 D11	D20 D21	D30 D31
14	C8 C9	C18 C19	C28 C29	C38 C39
13	C6 C7	C16 C17	C26 C27	C36 C37
12	C4 C5	C14 C15	C24 C25	C34 C35
11	C2 C3	C12 C13	C22 C23	C32 C33
10	C0 C1	C10 C11	C20 C21	C30 C31
9	B8 B9	B18 B19	B28 B29	B38 B39
8	B6 B7	B16 B17	B26 B27	B36 B37
7	B4 B5	B14 B15	B24 B25	B34 B35
6	B2 B3	B12 B13	B22 B23	B32 B33
5	B0 B1	B10 B11	B20 B21	B30 B31
4	A8 A9	A18 A19	A28 A29	A38 A39
3	A6 A7	A16 A17	A26 A27	A36 A37
2	A4 A5	A14 A15	A24 A25	A34 A35
1	A2 A3	A12 A13	A22 A23	A32 A33
0	A0 A1	A10 A11	A20 A21	A30 A31

Sped-up 200GBASE-R

2 symbols per CW: $7 \leq L \leq 16$

3 symbols per CW: $17 \leq L \leq 36$

UI\Lane	0	1	2	3
15	B60 B61	A60 A61	B70 B71	A70 A71
14	A48 A49	B48 B49	A58 A59	B58 B59
13	A46 A47	B46 B47	A56 A57	B56 B57
12	A44 A45	B44 B45	A54 A55	B54 B55
11	A42 A43	B42 B43	A52 A53	B52 B53
10	A40 A41	B40 B41	A50 A51	B50 B51
9	B28 B29	A28 A29	B38 B39	A38 A39
8	B26 B27	A26 A27	B36 B37	A36 A37
7	B24 B25	A24 A25	B34 B35	A34 A35
6	B22 B23	A22 A23	B32 B33	A32 A33
5	B20 B21	A20 A21	B30 B31	A30 A31
4	A8 A9	B8 B9	A18 A19	B18 B19
3	A6 A7	B6 B7	A16 A17	B16 B17
2	A4 A5	B4 B5	A14 A15	B14 B15
1	A2 A3	B2 B3	A12 A13	B12 B13
0	A0 A1	B0 B1	A10 A11	B10 B11

Impact of bursts is significantly lower with 4-way interleaving!

”Sped-up 200GBASE-R” proposal

- Performance with uncorrelated errors
  - 200GBASE-R has been analyzed in 802.3bs; standard FLR target requires **DER<6.1e-4, SNR>16.965 dB**
- Performance with correlated errors (strong DFE, a=0.75)
  - 8 physical lanes (analyzed in opsasnick_3df_logic_220630a): DER<2.5e-4, ΔSNR=0.66 dB (with precoding)
    - Better immunity to bursts than 400GBASE-R due to symbol-by-symbol lane distribution
  - 4 physical lanes (separate analysis): DER<1.1e-4, ΔSNR=1.21 dB (with precoding)
    - Large coding gain degradation compared to 400GBASE-R
    - Restored to 100G (ΔSNR=0.66 dB) with symbol muxing
- Implementation
  - Same architecture as 200GBASE-R, but higher data rate (no simple re-use)
  - Breakout to 2×400GBASE-R is not simple
  - Not compatible with existing implementations; supporting them will require some duplication
- Latency
  - Buffering time for two codewords at 800 Gb/s: 12.8 ns

” 2×400GBASE-R” proposal

- Performance with uncorrelated errors
  - 400GBASE-R has been analyzed in 802.3bs
  - 4-way codeword interleaving requires a factor of ½ in “post-FEC BER” (codeword error ratio) to get the same FLR as 2-way – but the impact on pre-FEC DER is negligible (5.8E-04 vs. 6.1E-04); ΔSNR=0.04 dB
- Performance with correlated errors (strong DFE, a=0.75)
  - 8 physical lanes (analyzed in opsasnick_3df_logic_220630a): DER<2.6e-4, ΔSNR= 0.63 dB (with precoding)
    - Better burst immunity than 400GBASE-R (0.07 dB) due to 4-way codeword interleaving
  - 4 physical lanes (separate analysis): DER<1.9e-4, ΔSNR= 0.85 dB (with precoding)
    - Significant coding gain degradation compared to 400GBASE-R
    - Improved compared to 100G (ΔSNR≈0 dB) with symbol muxing
- Implementation
  - Same logic and clock frequency for 1×800GBASE-R and 2×400GBASE-R (simple re-use)
  - Easy breakout
  - Compatible with existing implementations
- Latency
  - Buffering time for four codewords at 800 Gb/s: 25.6 ns

Comparison table

Parameter		“Sped-up 200GBASE-R”	“2×400GBASE-R”
Performance with uncorrelated errors		Good	Good (ΔSNR=0.04 dB)
Performance with correlated errors	8 physical lanes	Degraded, ΔSNR=0.66 dB	Degraded, ΔSNR= 0.63 dB
	4 physical lanes	Degraded, ΔSNR=1.21 dB	Degraded, ΔSNR= 0.85 dB
	4 physical lanes, symbol muxing	Degraded, ΔSNR=0.66 dB	Close to uncorrelated
Re-use		200GBASE-R architecture	400GBASE-R design
400G compatibility (breakout)		No	Yes
Codeword buffering latency		12.8 ns	25.6 ns

Summary

- Performance with error bursts should be considered
- With 8 physical lanes, both PCS proposals are roughly equivalent (and are similar to 400GBASE-R PCS)
- With 4 physical lanes:
  - With bit-muxing, both PCS proposals seem too vulnerable for bursts
  - If we consider symbol-muxing:
    - “sped-up 200GBASE-R” can be restored back to 8-lane degradation
    - “2×400GBASE-R” can become almost immune to bursts (no degradation)
  - “2×400GBASE-R” has a significant advantage
- “2×400GBASE-R” is favorable from most other aspects too

Thank you

Questions?