Analysis of PMA muxing options for 200G/lane signaling

Adee Ran, Cisco

Supporters

- Will Bliss, Broadcom
- Vasu Parthasarathy, Broadcom
- Cathy Liu, Broadcom

Background

- P802.3df has adopted PCS and PMAs for 8×100G PHYs
  - PCS with 32 logical lanes and 4 FEC engines
  - 2 codewords symbol-muxed on each PCS lane (either A/B or C/D)
  - 4:1 bit muxing (32:8 lanes) in the PMA
  - Consideration for FEC performance: lane muxing restriction such that each physical lane has bits from all 4 codewords
- The next steps (possibly in P802.3dj) are
  - **200G/lane AUIs: 800GAUI-4, 1.6TAUI-8, as well as 400GAUI-2 and 200GAUI-1**
  - 1.6T PCS and 200G/lane PMDs
- We want to re-use as much as possible from the existing architecture
  - PCSs for 200G, 400G, and 800G already exist
  - Can we keep the bit-muxing PMAs?

Goals of this presentation

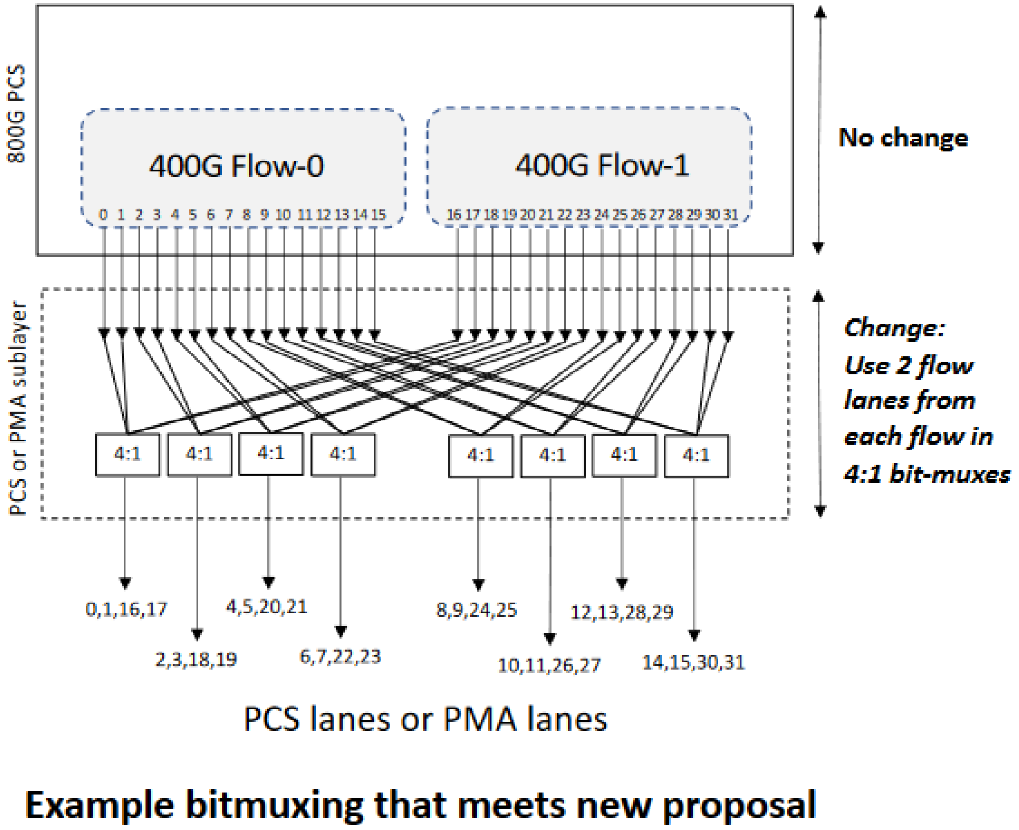
- Provide intuitive/graphical reasoning for the effect of error bursts
  - Compare 4:1 bit muxing (e.g., in 800GAUI-8) vs. 8:1 bit muxing
  - Compare 8:1 bit muxing vs. 8:1 symbol muxing
- Analyze performance (FLR vs. SNR plots) of RS FEC with correlated errors
  - Compare to results in wang_3df_01b_220928
- This is not a PMA proposal
  - A companion presentation describes a possible 8:1 symbol muxing specification in more detail
  - This presentation provides the motivation

Recap

(things we have discussed already)

800GBASE-R PMA: 4:1 bit muxing (32:8)

- Each of the two flows contains two codewords
  - Flow 0: A and B
  - Flow 1: C and D
- PAM4 symbols merge the content of two lanes
  - Consecutive symbols alternate between A/B and C/D
  - “Checkerboard” pattern on PCS periodically swaps MSB and LSB
- A block consisting of one bit from each PCS lane is transmitted in a 2-UI cycle



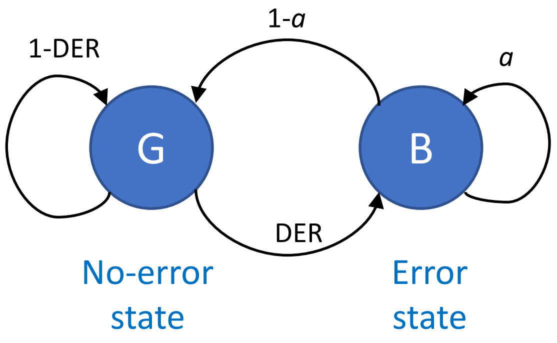
shrikhande_3df_01a_221004, slide #18

Why bit muxing affects burst sensitivity

- Bits from different PCS lanes are placed on the same physical lane
  - Each PCS lane transmits a different FEC symbol
  - **A burst of errors on one physical lane can affect bits from multiple PCS lanes ➔ multiple FEC symbols**
- As more PCS lanes are muxed on the same physical lane, burst sensitivity increases:
  - A burst of given length (L) can impact more FEC symbols
  - The probability of getting a number (n) of FEC symbol errors from a single error event increases
  - “Blast radius” increases
- Since each PCS lane carries 25 Gb/s:
  - 50G/lane signaling – mux ratio 2:1 (with PAM4, LSB from one lane and MSB from another lane)
  - 100G/lane signaling – mux ratio 4:1
  - 200G/lane signaling – mux ratio 8:1
- Interleaving multiple codewords mitigates the muxing effect, but only partly

PAM4 error model

- PAM4 symbols are formed by pairs of bits on the same PMA lane
- A detection error (with probability DER) **inverts one bit of the PAM4 symbol (either MSB or LSB)**
  - Due to Gray coding, two-bit errors are rare (<DER^2)
- We assume 1-tap DFE error propagation (Gilbert model)
  - Probability of a PAM4 detection error propagating to the next PAM4 symbol is denoted *a*
  - A random error event creates a burst of length L PAM4 symbols with probability $a^{L-1}(1-a)$
  - With PAM4, $0 \leq a \leq 0.75$
- Precoding converts a burst of length L into just two PAM4 symbol errors, in positions 1 and L+1
  - Effectively doubling the DER (and potentially the SER for RS-FEC)
  - It is only beneficial if long bursts are frequent.

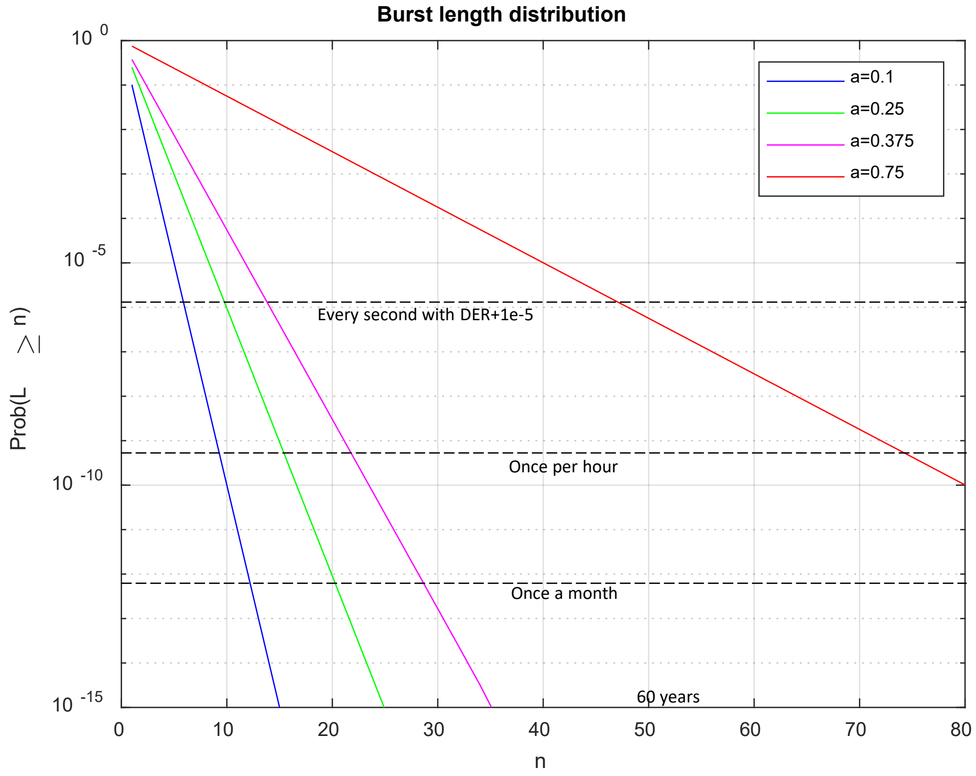


How often do bursts occur?

And how long can they get?

Expected burst lengths

- In annex 120G (100G/lane AUI-C2M) we have **DER<1e-5**, and a limited DFE assumption that results in *a*=0.25
  - With these values, typical bursts (expected to occur at least once per second) have L ≤ 10
  - Bursts with L≥25 occur once in ~60 years
- PMDs (CR, KR) can have stronger DFEs
  - Also, higher DER ➔ error events occur more often
- For a KR/CR receiver with DER=1e-4:
  - DFE tap value of 0.5 results in *a*=0.375; bursts with L≥15 occur every second, and bursts with L≥24 occur daily
  - Stronger DFE can reach *a*=0.75; this would cause bursts with L≥54 once per second(!) and L≥82 occurs daily



Burst effect on FEC

Should we care?

800GAUI-8 streams

- PCS output bits are allocated to the 8 PMA lanes in pairs as shown
- A burst usually affects up to one RS symbol per codeword
  - To affect more than one symbol, a burst has to cross a symbol-group boundary (once every 20 UI) and a specific MSB/LSB combination
  - This is shown in the highlighted case (either A9+A80 or B9+B80).
- A burst of errors can also “spill” into the other codewords

Bits A0-A9 are one RS symbol of codeword A
Bits A10-A19 are the next RS symbol of codeword A
Bits B0-B9 are one RS symbol of codeword B
Bits B10-B19 are the next RS symbol of codeword B
…

	UI\Lane	0	1	2	3	4	5	6	7
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Last	22	B81 A81	B91 A91	B101 A101	B111 A111	B121 A121	B131 A131	B141 A141	B151 A151
Tx/Rx order	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
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	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
First	0	A0 B0	A10 B10	A20 B20	A30 B30	A40 B40	A50 B50	A60 B60	A70 B70

Correlated errors in 800GAUI-8 C2M

- Error propagation in the reference C2M receiver is equivalent to a BER increase of at most 7.1%
  - Due to the limited DFE assumption
  - Detailed calculation →
- **The effect of correlated errors in 800GAUI-8 C2M is negligible!**

- Error spilling into other codewords increases the average BER
  - The effect is a factor of $1+\frac{1}{4}\left(a+\frac{a^2+a^3}{2}\right)$
  - For a=0.25, it is a 7% increase
- A 3-UI or longer burst can affect two RS symbols in the same codeword
  - With a=0.25, 1 of 16 error events creates a long enough burst
  - Combined with the required alignment, one of about 550 random error creates a 2-symbol event
  - Effectively increases the BER by ~0.1%
- To affect three RS symbols, a burst with L>21 is required
  - But this is extremely rare in AUI-C2M, and has negligible effect

Correlated errors in 800GBASE-CR8/KR8

- A receiver for a high-loss channel (e.g. 800GBASE-CR8 PMD) can stronger DFE and higher error correlation
- The requirement for a the PMD is stated as "a frame loss ratio lower than 9.2×10^{-13}"
  - Graphs of the effect of error correlation (for several values of *a*) on frame loss ratio vs. SNR are shown in a backup slide
- Using 4-codeword interleaving makes 800GBASE-CR8 more tolerant to bursts than 400GBASE-CR4 (with 2 codewords)
  - With *a*=0.375 (DFE limited to 0.5), the penalty is only 0.3 dB
  - *a*=0.75 may occur with the highest loss channels (larger DFE); if precoding is used, the penalty is 0.6 dB
- **The effect of correlated errors in 800GBASE-CR8/KR8 is tolerable!**

100G/lane ➔ 200G/lane

- Lane muxing ratio increases from 4:1 to 8:1
- For high-loss C2M channels, we expect stronger receiver equalization
  - Strong DFE and/or MLSE
  - Also expected for optical receivers at 200G/lane
  - DFEs are also expected in medium-loss C2M
- **Actual designs can differ – but we should expect much stronger error correlation than in 100G AUIs!**
- For AUIs with high DER, we assume the RS-FEC is terminated in the module
  - Therefore, the FLR is divided between the segments; assume 9.2×10^{-13} is allocated to each AUI
  - We will look at the FLR of the C2M segment as a function of its SNR and *a.*

8:1 muxing options

8:1 bit muxing for 800GAUI-4?

- Assuming the same 32-lane PCS, bits would be allocated to the 4 physical lanes as shown
- A 7-UI burst can affect up to four RS symbols in the same codeword
  - As shown in the highlighted case (either A9+A19+A80+A90 or B9+B19+B80+B90)
  - With *a*=0.75, 18% of errors create 7-UI or longer bursts
  - This should be multiplied with the probability of alignment and errors in specific bits
  - Overall, 4-symbol error events occur W.P. 6e-3
- **Any 3-UI burst** can affect two symbols in the same codeword
  - 2-symbol error events occur W.P. 28%
- Spilling into other codewords is severe
- **Overall, the FEC performance degradation is much worse than with 4:1 bit muxing**

Last

Tx/Rx order

First

UI\Lane	0	1	2	3
	⋮	⋮	⋮	⋮
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
⋮	⋮	⋮	⋮	⋮
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
4	A1 B1	A21 B21	A41 B41	A61 B61
3	C10 D10	C30 D30	C50 D50	C70 D70
2	A10 B10	A30 B30	A50 B50	A70 B70
1	C0 D0	C20 D20	C40 D40	C60 D60
0	A0 B0	A20 B20	A40 B40	A60 B60

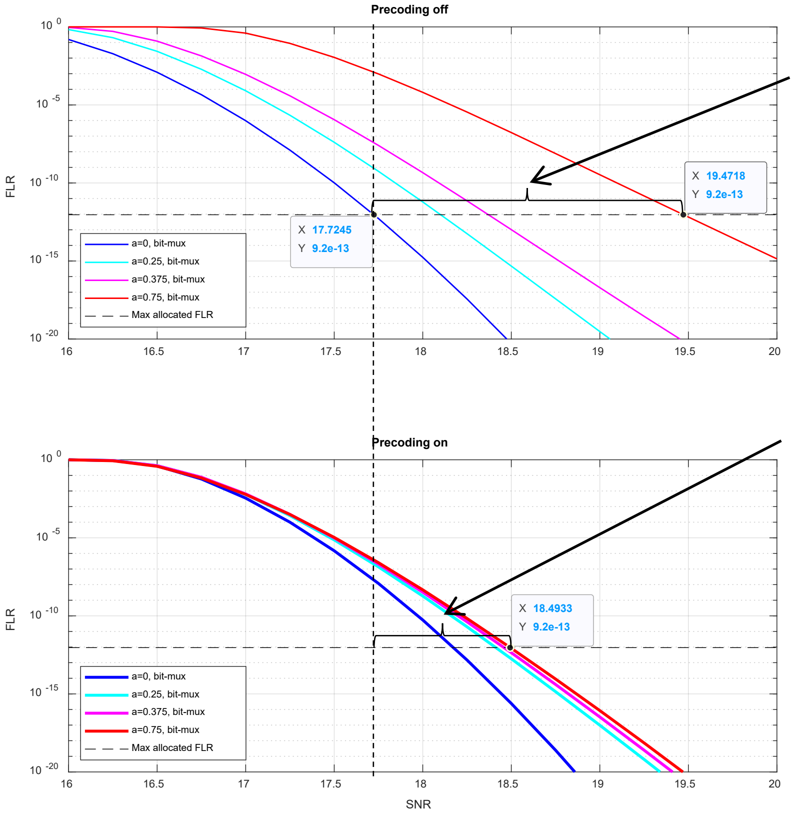
Can precoding save us?

- With precoding, a burst will affect two symbols in the same codeword **if the initial error and the termination error are 2 UI apart**, as shown on the right.
  - This happens if the number of propagation events is 1, 5, 9… or generally (4n+1)
  - With *a*=0.75, this happens W.P. 27%; it is almost as common with lower values.
- The impact is no more than two symbols…
  - But 2-symbol events happen much more often than with 4:1 bit muxing.
- Precoding has a penalty for all values of *a*
  - Even when there is no error propagation, the end-of-burst error would spill into another codeword
  - In past specifications, precoding was optional/negotiated; but this can't be done over optics.

	UI\Lane	0	1	2	3
		⋮	⋮	⋮	⋮
Last	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
	⋮	⋮	⋮	⋮	⋮
	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
Termination	4	A1 B1	A21 B21	A41 B41	A61 B61
1 propagation	3	C10 D10	C30 D30	C50 D50	C70 D70
Initial error	2	A10 B10	A30 B30	A50 B50	A70 B70
	1	C0 D0	C20 D20	C40 D40	C60 D60
First	0	A0 B0	A20 B20	A40 B40	A60 B60

FEC performance with an 8:1 bit-muxing PMA with 4-way interleaved FEC

As a function of SNR



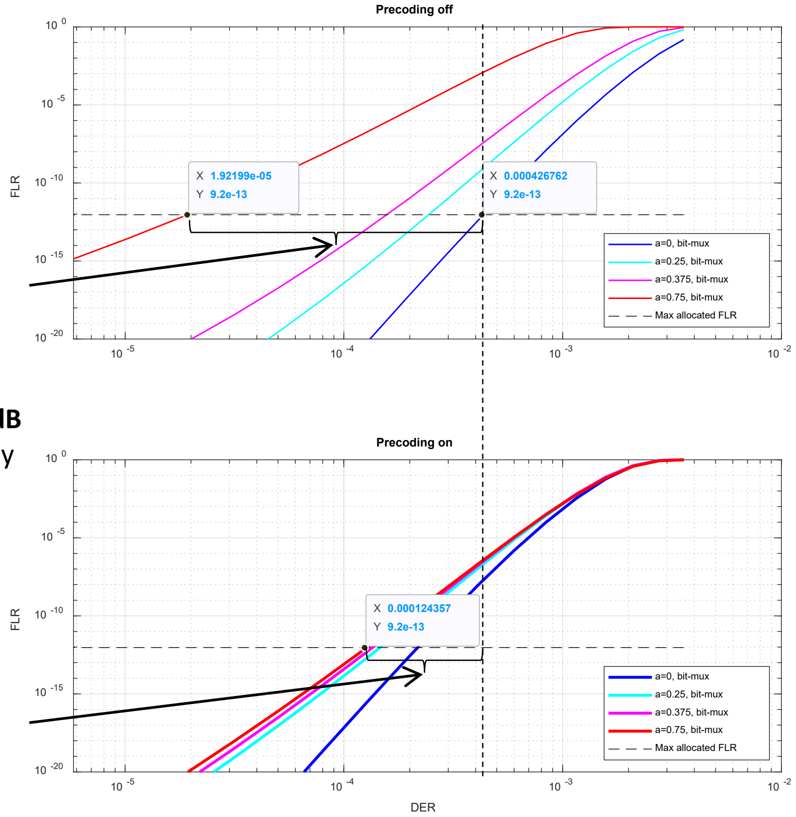
Without precoding, the worst-case penalty at 8:1 muxing (with a=0.75) is **1.75 dB** (much worse than 4:1)

A factor of 20 in BER!

Precoding reduces the worst-case penalty to **0.8 dB** but has a minimum penalty (again, worse than 4:1)

A factor of 4 in BER in almost all cases

As a function of DER (2*BER)



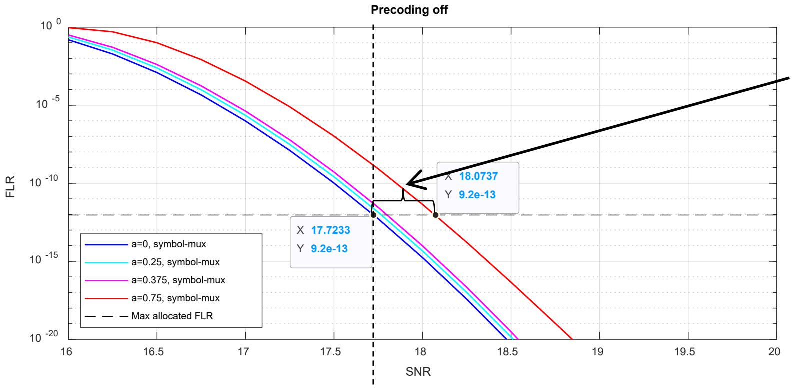
The solution: 8:1 symbol muxing in the PMA

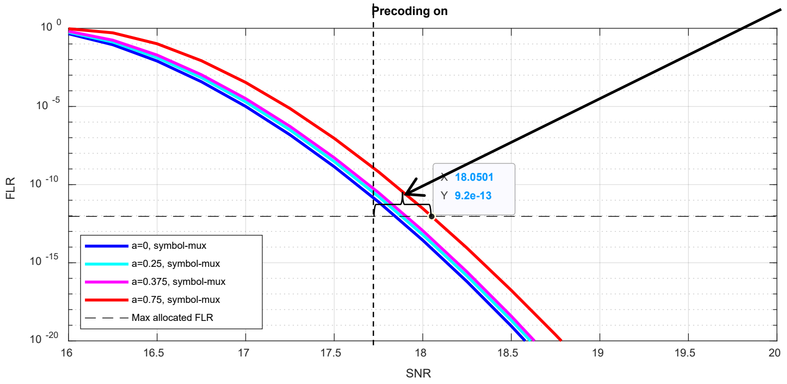
- Instead of taking one bit from each PCS lane, the PMA takes a full FEC symbol (10 bits)
  - Each PAM4 symbol contains two bits from the same FEC symbol
- Bits are allocated to the 4 lanes as shown on the right.
- Short error bursts affect up to 1 symbol per codeword
  - **Affecting two symbols in the same codeword requires a burst with L≥17**
  - With a=0.75, such bursts occur W.P. 0.7%
  - For three symbols – L≥37 (W.P. 2e-5)
- **Spilling into other codewords is much less severe**

	UI\Lane	0	1	2	3
		⋮	⋮	⋮	⋮
Last	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
First	0	A0 A1	A10 A11	A20 A21	A30 A31

FEC performance with an 8:1 symbol-muxing PMA

As a function of SNR



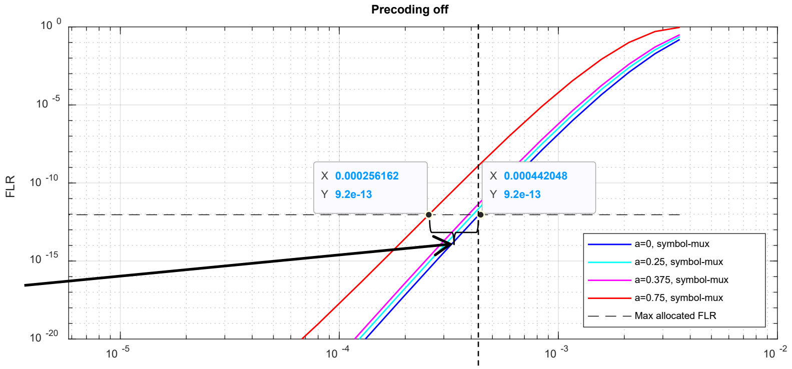


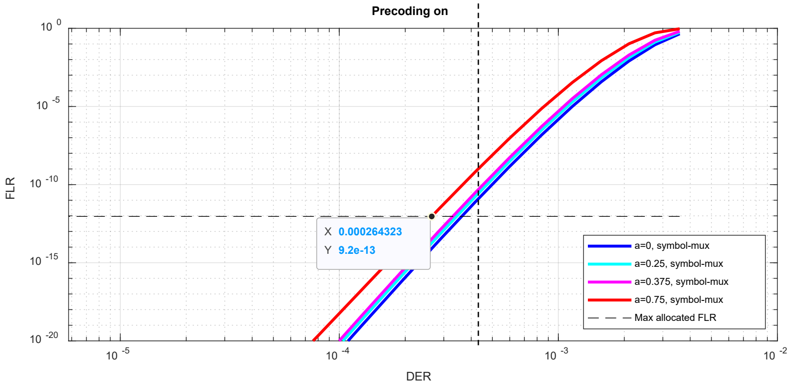
Without precoding, the worst-case penalty at 8:1 muxing (with *a*=0.75) is **0.35 dB**
(much better than 4:1)
Lower values of *a* create negligible penalty

A factor of 2 in BER, only in the worst case

Precoding has a minor effect (worst-case 0.33 dB)
Minimum penalty is small

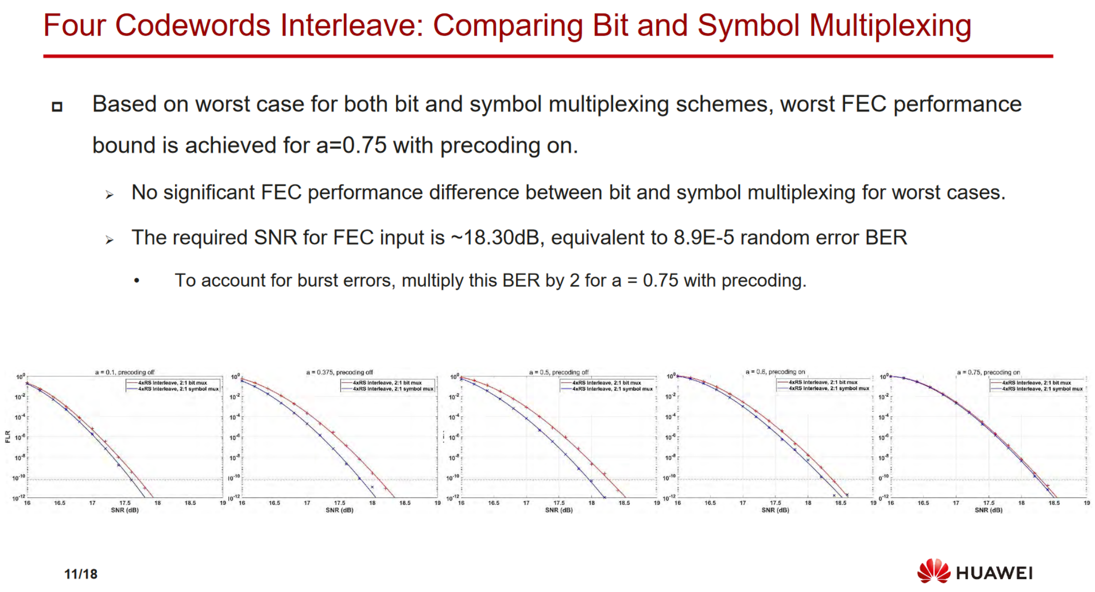
As a function of DER (2*BER)



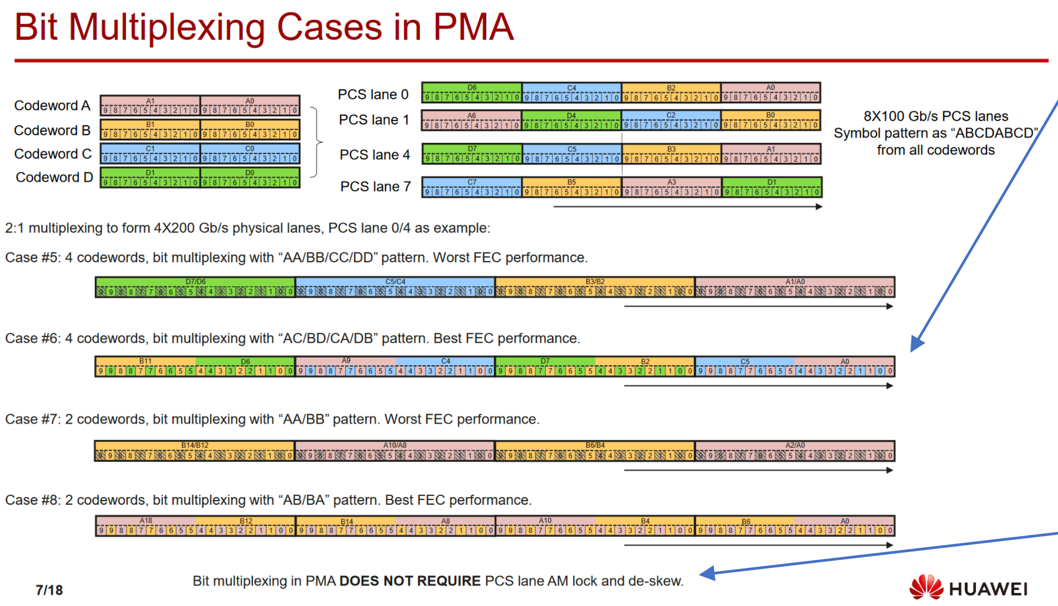


Is there really a difference?

- In wang_3df_01b_220928 it has been stated that bit and symbol muxing have no significant FEC performance difference
- Why is there a difference in my analysis?



Difference explained



- The "bit multiplexing" in wang_3df_01b_220928 is performed between 2 RS symbols at a time
- This method is suitable for an 8-lane PCS (2:1 muxing ratio)
- But the 800GBASE-R PCS has 32 lanes, and generates 32 symbols in parallel
  - If a host ASIC uses the 800GBASE-R PCS, the "bit muxing" shown on this slide would require an external 8-lane PCS (XS)
- The claimed benefits of bit muxing would only be achieved if a host implements an 8-lane PCS internally
  - Having two different PCS implementations in an ASIC is a pain

Summary

Compare muxing options for 800G: SNR [dB] and DER for meeting FLR=9.2e-13

Scenario	8-lane AUI/PMD 4:1 bit muxing	4-lane AUI/PMD 8:1 bit muxing	4-lane AUI/PMD 8:1 symbol muxing
Uncorrelated errors	17.7 (reference) 4.3e-4		
Limited DFE, *a*=0.375	18.05 (Δ=0.35 dB) 2.7e-4	18.4 (Δ=0.6 dB) 1.6e-4	17.8 (Δ=0.1 dB) 3.9e-4
Unlimited DFE, *a*=0.75	18.7 (Δ=1 dB) 8.9e-5	19.5 (Δ=1.75 dB) 1.9e-5	18.07 (Δ=0.35 dB) 2.6e-4
Unlimited DFE, *a*=0.75 + precoding	18.3 (Δ=0.6 dB) 1.8e-4	18.5 (Δ=0.75 dB) 1.2e-4	18.05 (Δ=0.33 dB) 2.6e-4
Overall	Acceptable for PMD where precoding can be negotiated AUI and optics assumed not to have *a*=0.75	Not acceptable unless precoding is negotiated	Minimal degradation in all cases Precoding not required*

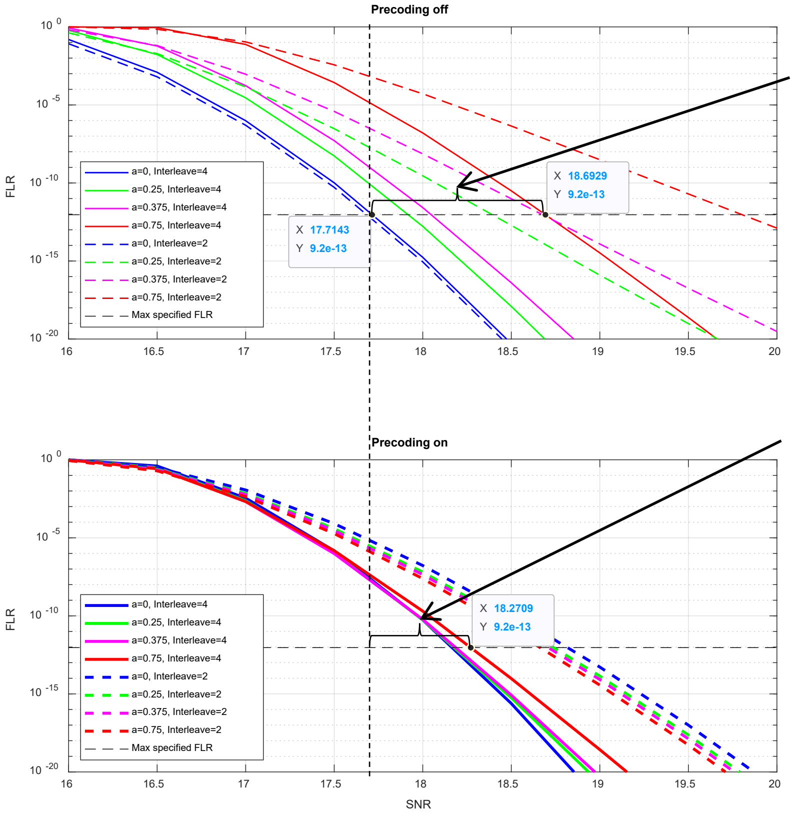
* Precoding may be needed for 400G and 200G with only 2-way codeword interleaving

Backup

FEC performance in 800GBASE-CR8

(dashed lines: 400GBASE-CR4, 2-way codeword interleaving)
(color denotes value of *a*)

As a function of SNR



Without precoding, the worst-case penalty at 800G (with a=0.75) is 1 dB (better than 400G)

Precoding reduces the worst-case penalty to 0.6 dB but has a minimum penalty (again, better than 400G)

As a function of DER (2*BER)