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
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 (<DER²)
- 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.

![Diagram showing the error model with states No-error state and Error state, and transition probabilities 1-DER, 1-a, a.](image-url)
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 $L \leq 10$
  - Bursts with $L \geq 25$ occur once in ~60 years
- PMDs (CR, KR) can have stronger DFES
  - Also, higher DER $\Rightarrow$ 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 $L \geq 15$ occur every second, and bursts with $L \geq 24$ occur daily
  - Stronger DFE can reach $a=0.75$; this would cause bursts with $L \geq 54$ once per second(!) and $L \geq 82$ 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.
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 \(1 + \frac{1}{4}(a + \frac{a^2 + a^3}{2})\)
  - 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 \(\sim0.1\%\)

- 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 $9.2 \times 10^{-13}$”
  • 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 $9.2 \times 10^{-13}$ 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**
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 α=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 α
  - 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.
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 $L \geq 17$
  • With $a=0.75$, such bursts occur W.P. 0.7%
  • For three symbols – $L \geq 37$ (W.P. 2e-5)

• Spilling into other codewords is much less severe
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.

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

As a function of DER (2*BER)

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

FEC performance with an 8:1 symbol-muxing PMA

November 2022

IEEE P802.3df 21
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?
• 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

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

* 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 $a$)