MTTFPA Concerns

IEEE P802.3bj

July 2012 San Diego

Roy Cideciyan – IBM

Mark Gustlin – Xilinx

Contributors and Supporters

Pete Anslow – Ciena

Stephen Bates – PMC-Sierra

Andre Szczepanek – Inphi

Pravin Patel – IBM

Mounir Meghelli – IBM

Mike Dudek – Qlogic

Barry Barnett – IBM

Arthur Marris – Cadence

MTTFPA Summary

- **Draft 1.0 allows you to send 64b/66b encoded data if FEC is not needed (loss < 30dB) for the NRZ PHY (backplane and copper cable)**
  - This reduces the latency for those channels/applications that don't need FEC
- **Roy's presentation shows (in cideciyan_01_0512) that sending 64b/66b data at a 10^{-12} BER has an MTTFPA of ~10^{4} years**
  - Mainly due to the high probability of an error burst that extends to 4 bits due to the DFE, and how that error burst is spread in the packet due to the PCS lane bit multiplexing
- **Roy also shows that the MTTFPA of FEC transcoded data has an MTTFPA of ~10^{3} years at a 10^{-12} BER if FEC is not used for correction or error detection**
- **Assuming that people agree that these MTTFPAs are not sufficient, what do we do?**
- **We held 4 meetings to build consensus on this issue**

Possible Solutions

1. **The MTTFPA is good enough at BERs where people really run their systems, so don't do anything**
2. **Add pre-coding**
   - First thoughts are that this can help, but is not a complete solution
3. **Perform block interleaving with the 64b/66b blocks instead of bit interleaving**
4. **Don't allow 64b/66b to be sent, always send FEC encoded data, receiver can correct, detect only or do something else (trailing error detection etc).**
5. **Terminate the 100G PCS, create a 4 PCS lane PCS at 100G.**

Possible Solution #2

2. Add pre-coding

The current estimate is that adding in precoding will improve the MTTFPA by a couple orders of magnitude

This is not sufficient to solve the 64B/66B bit interleaved concern without doing something else also

Possible Solution #3

3. Perform block interleaving with the 64b/66b blocks instead of bit interleaving

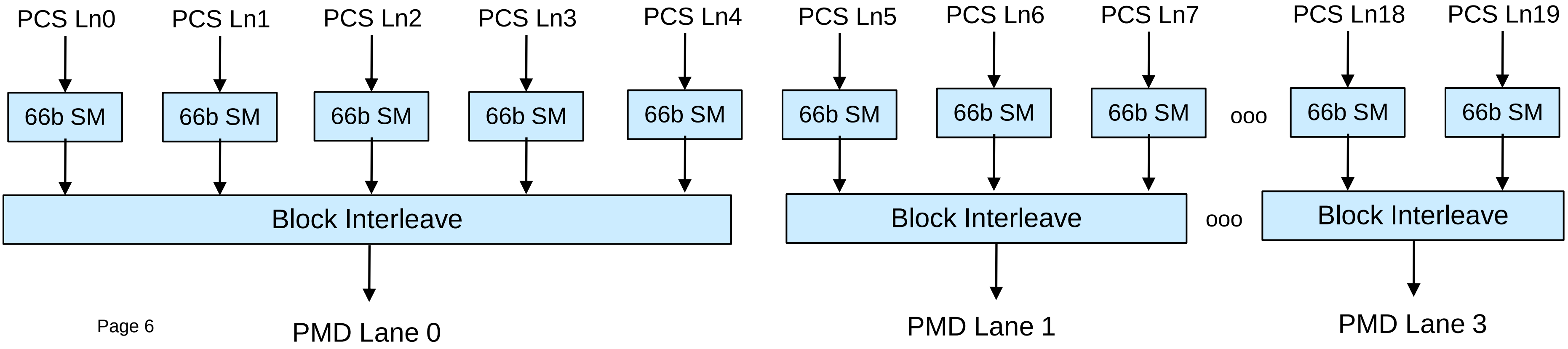
In clause 83 add a new PMA for 100GBASE-KR4 PMDs

You take 5 PCS lanes, find 66b lock, then interleave a block at a time on that PMD lane in a round robin fashion

No need to align or re-order the PCS lanes, any PCS lane can appear on any PMD lane, though once you start sending a PCS lane on a given PMD lane it is always sent in order on that PMD lane

Any burst error will be broken up into at most two separate errors, and the CRC32 can detect up to two burst errors of 9 bits each in 802.3 ethernet frames, and two burst errors of 8 bits each in jumbo frames as discussed in cideciyan_01_0712.pdf

This will add ~13ns (5 x 64 x 40ps = 12.8 ns) of latency (to gather the 66b blocks on each PCS lane), when a non co-located interface



Possible Solution #4

4. **Don’t allow 64b/66b to be sent, always send FEC encoded data, receiver can correct, detect only or do something else (trailing error detection etc.)**

When correcting FEC errors, the MTTFPA is acceptable (> LOU), added latency ~100ns

When detecting FEC errors and not correcting, MTTFPA is acceptable (> LOU), added latency is ~50ns

If an application is very latency sensitive, then a proprietary trailing FEC check can be implemented to reduce the latency further, details left up to the implementer (this assumes a channel where FEC correction is not needed), added latency is ~5ns

Proposal: Require that FEC is always transmitted in the standard for NRZ backplane and copper cable. The receiver behavior is not defined except how to decode/correct if the receiver chooses to.

Benefit is no need to communicate turning on/off FEC.

Observable standard behavior is meeting the BER, not important the details of how to get there.

There is concern over being able to accurately detect 10^{-12} BER and deciding to turn FEC correction on, but this is there for all other options also

Possible Solution #5

5. Terminate the 100G PCS, create a 4 PCS lane PCS at 100G

- Add a new PCS clause with 4 PCS lanes
- This will have identical performance as option #3
- The complexity is higher though, at least for non collocated sublayers

Background on 40GBASE-KR4 MTTFPA

- The question has come up, what was the MTTFPA for 40GBASE-KR4, and how come it does not have an issue
- 40GBASE-KR4 stripes data across 4 PCS lanes, 1 PCS lane per physical lane, so it should be similar to the MTTFPA of option 3 or 5
- In gustlin_02_0308.pdf we used error propagation probabilities from liu_01_1105, 11b error was the maximum size, probability of propagation is 0.1
- An 11b error is 100% detectable, so next worst case is two errors that are propagated to 2 and 4 bits respectively
  - This calculates to an MTTFPA of ~10^{17} years
  - This was for errors contained within the payload, the analysis did not include control block corruption impacts, and so this analysis might be optimistic

Comparison of the Options

9018B jumbo frame, Prob. of staying in burst state=0.5

Option	Draft 1.0	Option 3	Option 4	Option 5
Description	20-lane PCS 64b/66b Bit-Muxed PMA	20-lane PCS 64b/66b Block-Muxed PMA	20-lane PCS Mandatory TC and FEC Encoding	4-lane PCS 64b/66b PMA(4:4)
Added Latency[1]	~0 ns	~13 ns	~5ns[2] or 50ns[3]	~13 ns
MTTFPA for BER=10^{-12}	3×10^{4} years	1×10^{9} years	~10^{86} years	1×10^{9} years
MTTFPA for BER=10^{-6}	12 days	1×10^{3} years	~10^{71} years	1×10^{3} years
MTTFPA for BER=10^{-4}	1 hour	50 days	~10^{56} years	50 days
Complexity	Simple bypass of FEC, autoneg issues	Moderate, add in block interleaving logic, autoneg issues	Simple[4], reuse of FEC coding, requires FEC be implemented	Most complicated, add in new PCS mode, autoneg issues

[1]: This is an estimate of the added latency for non collocated sublayers

[2]: No error correction, just error detection in trailing mode *after* FEC decoder outputs codeword payload

[3]: No error correction, just error detection *before* FEC decoder outputs codeword payload

[4]: The trailing error detection can add complexity, that is implementation dependent

Comparison of the Options

9018B jumbo frame, Prob. of staying in burst state=0.1

Option	Draft 1.0	Option 3	Option 4	Option 5
Description	20-lane PCS 64b/66b Bit-Muxed PMA	20-lane PCS 64b/66b Block-Muxed PMA	20-lane PCS Mandatory TC and FEC Encoding	4-lane PCS 64b/66b PMA(4:4)
Added Latency[1]	~0 ns	~13 ns	~5ns[2] or 50ns[3]	~13 ns
MTTFPA for BER=10^{-12}	3×10^{6} years	4×10^{16} years	$>10^{86}$ years	4×10^{16} years
MTTFPA for BER=10^{-6}	2 years	1×10^{10} years	$>10^{71}$ years	1×10^{10} years
MTTFPA for BER=10^{-4}	1 hour	186 days	$>10^{56}$ years	186 days
Complexity	Simple bypass of FEC, autoneg issues	Moderate, add in block interleaving logic, autoneg issues	Simple[4], reuse of FEC coding, requires FEC be implemented	Most complicated, add in new PCS mode, autoneg issues

[1]: This is an estimate of the added latency for non collocated sublayers

[2]: No error correction, just error detection in trailing mode *after* FEC decoder outputs codeword payload

[3]: No error correction, just error detection *before* FEC decoder outputs codeword payload

[4]: The trailing error detection can add complexity, that is implementation dependent

Recommendation

- Given the MTTFPA issues with sending bit multiplexed 64B/66B encoded data even on a low loss backplane or copper cable channel, the majority recommendation from the participants in the MTTFPA discussion is option 4: Require that FEC encoded data is always sent by the TX
- The receiver has the option to always correct, only detect errors for low loss channels, or do some proprietary trailing error detection if absolute lowest latency is needed
- This is the lowest complexity solution that also provides a robust MTTFPA
- This solution simplifies overall operation, no need for auto-negotiation between the TX and RX, the RX can decide what to do based on its knowledge of the channel

- Use comment #76 and change it to say that FEC is required to always be sent for both clause 92 and 93.

Thanks!