FEC update

Piers Dawe
Avago Technologies

Supporters

- John Petrilla — Avago Technologies
- Jonathan King — Finisar
- Petar Pepeljugoski — IBM
- Richard Mellitz — Intel
- Andre Szczepanek — Texas Instruments

Contents

- Introduction
- FEC and striping
- FEC and non-FEC coexistence options
- Conclusions
- References

Introduction

- With every generation, FEC becomes more attractive and ubiquitous
- Backplane and copper-cable PMDs need protection against burst errors
  - 10GBASE-R FEC protects against 11-UI errors and provides ~2 dBe gain
- Optical PMDs can benefit from FEC gain
- ~1 dBo gain unless optically amplified (100GBASE-ER4) when 1.5 dBo gain
- Common PCS implies common FEC, irrespective of number of physical lanes
- 10GBASE-R FEC (Clause 74) exists, does not affect throughput or line rate, will pass through existing CDRs and optical modules such as XFP, SFP
- FEC adds latency (e.g. for 10GBASE-R, ~410 ns minimum at 10G or 34% of a 1500-bit frame or 82 m of cable, max. per 74.6 is 614.4 ns sum of transmit and receive delay)
- FEC adds some silicon area (16 kgates + RAM for both directions, one end) and power (<50 mW per 10G per end)
  - Not sure that all will accept these costs
- 10GBASE-R FEC is auto-negotiated
  - Would like to avoid this, at least for optical PMDs
- How to extend 10GBASE-R FEC for multi-lane use, both copper and optical

FEC and striping

Because number of physical lanes may vary and number of FEC lanes should be fixed, consider these possible scenarios:

1. FEC lane(s) is/are wider than physical lanes
   - e.g. FEC acts on the whole stream
2. FEC lanes are same as physical lanes
   - 4 lanes in 40G
3. FEC lanes are narrower than physical lanes
   - e.g. one FEC machine per virtual lane, 2 or 5 (or 10 or 20) per physical lane
4. FEC lanes are different to physical lanes
   - e.g. in 100G, FEC acts on pairs of VLs which are muddled up before transmission

1 If FEC lanes are wider than physical lanes

- One FEC machine per physical lane
- Latency is same number of bits but 1/10 or 1/4 the time of 10GBASE-KR because link is running faster
- Burst protection (11 bits on the line in KR) is divided by number of physical lanes per FEC lane
  - Not desirable for MTTFPA for backplane, electrical-cable
- 2048 MAC bits "lost" (marked as bad) per uncorrected error event (17% of a 1500-byte frame)
- For short-enough error bursts, FEC gain is as for 10GBASE-KR
  - For longer bursts, less good
- A single FEC machine would have to be fast
  - Because FEC blocks are independent, can implement in multiple parallel machines
    - At a small cost in latency

2 If FEC lanes are same as physical lanes

- E.g. one FEC machine "above the MLD"
- Latency is e.g. 4x number of bits, same time as 10GBASE-KR, because link is running faster
- Burst protection is same as KR
  - Good for MTTFPA for backplane, electrical-cable
  - Probably overkill for optical
- E.g. 4 x 2048 = 8192 MAC bits "lost" (marked as bad) per uncorrected error event (68% of a 1500-byte frame)
- FEC gain is as for 10GBASE-KR
- FEC machine can run at 10G
  - Because FEC blocks are independent, can implement in a larger number of parallel machines or a smaller number of faster machines
    - At a small cost in latency

3 If FEC lanes are narrower than physical lanes

- E.g. a FEC machine per virtual lane
- Latency is e.g. 4 (8) or 20 times as many bits, same or double the time of 10GBASE-KR, because link is running faster
- Burst protection (11 bits on the line in KR) is multiplied by number of FEC lanes per physical lane (e.g. x2)
  - Very good for MTTFPA for backplane, electrical-cable, overkill for optical
- 40960 MAC bits (3½ 1500-byte frames) "lost" (marked as bad) per uncorrected error event if 20 FEC lanes for 100G, 16384 MAC bits (1.4 1500-byte frames) if 8 FEC lanes for 40G
- FEC gain is a very little lower than 10GBASE-KR (~0.2 dB See Valliappan)
- Can have many 5G FEC machines
  - Because each block is independent, can share fewer machines if wished (at small extra cost in latency?)

4 If FEC lanes are different to physical lanes

- E.g. 10 FEC machines, one per two VLs, 10 physical lanes, but mismatched
- Latency is e.g. 10 times as many bits, same or time of 10GBASE-KR because link is running faster
- Burst protection (11 bits on the line in KR) is doubled
  - Unless there are coincident error bursts on the two physical lanes that one FEC instance half-protects
  - Very good for MTTFPA for backplane, electrical-cable, overkill for optical
- 4096 MAC bits (34% of a 1500-byte frame) "lost" (marked as bad) per uncorrected error event
- FEC gain is a very little lower than 10GBASE-KR, as in case 3
- FEC machine can run at 10G
  - Because FEC blocks are independent, can implement in a larger or smaller number of parallel machines
    - At a small cost in latency

FEC and non-FEC coexistence options

- Transmitter to encode or not?
- Receiver to ignore FEC, check or correct?

Detect or correct?

- Options are

A. 64B/66B transmit and receive, as in 802.3ae

B. 64B/65B transmit with 32-bit check word as Clause 74, receiver ignores the check word

C. 64B/65B transmit, check the 32-bit check word but don't attempt to correct

D. 64B/65B transmit, check the 32-bit check word and (attempt to) correct, as in Clause 74 FEC

A 64B/66B transmit and receive

- As in 802.3ae
- Second sync bit improves detectability of some errors?
- PCS throws itself out of lock if BER is poor
  - This protects the MAC from too many false frames to challenge its CRC
  - PCS recovers lock very rapidly if this happens because of a transient event e.g. lightning
- Uses pairs of sync bits to determine BER?
  - This is what nicholl_?_0?08 proposes can be counted for BER monitor
- No special protection against burst errors (apart from the MAC CRC's good but not perfect probability of detecting them)

B 64B/65B transmit with check word which receiver ignores

- Check word encoding adds negligible latency
- Can we live without second sync bit to improve detectability of some errors?
- Use lane markers to gain sync
  - Original 10GBASE-KR couldn't do this
- Need to mark the link as bad if BER is poor
  - To protect the MAC from too many false frames to challenge its CRC
  - What BER should the threshold be at?
    - Is it significantly affected by multi-lane considerations?
  - PCS sync-up without frequent 2-bit headers would be slow
    - Rather than throw itself out of lock, PCS should carry on in lock but set its link OK primitive to bad for a while
      - Trivial change to Clause 74 spec
  - Use what to determine BER?
    - Lane markers (presumably not FEC coded) are very dilute
    - Idle (visible after 64B/66B decoding) also fairly dilute at worst
- No special protection against burst errors (apart from the MAC CRC's good but not perfect probability of detecting them)

C 64B/65B transmit with check block which receiver checks

- Just checking adds a small latency (~one FEC block if whole block to be marked: depends on implementation)
- Can we live without second sync bit to improve detectability of some errors? Presumably so if 802.3ap did
- Need to mark the link as bad if BER is poor
  - Not just the bad block
  - To protect the MAC from too many false frames to challenge its CRC
  - What BER should the threshold be at?
    - Is it significantly affected by multi-lane considerations?
  - PCS sync-up without frequent 2-bit headers would be slow
    - Rather than throw itself out of lock, PCS should carry on in lock but set its link OK primitive to bad for a while
      - Trivial change to Clause 74 spec
  - Use 32-bit check blocks to determine BER
- Burst errors are detected but not corrected
- Seems affordable and robust. Compatible with 4. No auto-neg.

D 64B/65B transmit with check block which receiver corrects

- Correcting adds another small latency (more than one FEC block and depends on implementation?)
- Consider marking the link as bad if BER is poor
  - Not just the bad block
  - To protect the MAC from too many false frames to challenge its CRC
  - What BER should the threshold be at?
    - Is it significantly affected by multi-lane considerations?
  - PCS sync-up without frequent 2-bit headers would be slow
    - Rather than throw itself out of lock, PCS should carry on in lock but set its link OK primitive to bad for a while
      - Trivial change to Clause 74 spec
  - Use 32-bit check blocks to determine BER
- Burst errors are detected and corrected
  - How many uncorrectable bursts are detectable?
- Seems good for those who can afford the latency. Compatible with 3. No auto-neg.

Conclusions

- Should still revisit every PMD to see if a stronger FEC is needed
- If not, in 802.3ba always encode for FEC
- Use lane markers for PCS/FEC sync
- Receiver must always check the FEC word
  - Use FEC word checks for BER monitor counter
  - Use FEC words to blank bad links for excellent MTTFPA even in presence of burst errors
- Receiver can optionally(?) correct using FEC word
- No auto-negotiation required
  - Receiver can even change its mind (probably outside the standard)
- Receiver chooses low latency or high robustness

References

- High Speed Ethernet
- http://ieee802.org/3/hssg/public/nov07/trowbridge_03_1107.pdf
- http://ieee802.org/3/ba/public/may08/dawe_02_0508.pdf
- 10GBASE-KR FEC Tutorial
  http://ieee802.org/802_tutorials/july06/10GBASE-KR_FEC_Tutorial_1407.pdf
  See IEEE Std 802.3, 74
- 802.3ah EFM
- Reed-Solomon code (255, 239) operating on 8-bit symbols
  See IEEE Std 802.3, 65.2 and e.g.
  http://ieee802.org/3/efm/public/sep01/rennie_1_0901.pdf
  http://ieee802.org/3/efm/public/jul02/p2mp/khermosh_general_1_0702.pdf
  http://ieee802.org/3/efm/public/sep02/p2mp/rennie_p2mp_1_0902.pdf
- 802.3av
  RS(255, 223)
- See P802.3av 92.3 and e.g.
  http://ieee802.org/3/av/public/2007_07/3av_0707_daido_1.pdf
- http://ieee802.org/3/av/public/2007_01/3av_0701_effenberger_1.pdf
- http://ieee802.org/3/av/public/2007_03/3av_0703_kramer_1.pdf
- http://ieee802.org/3/av/public/2007_05/3av_0705_lynskey_1.pdf
- http://ieee802.org/3/av/public/2007_05/3av_0705_effenberger_4.pdf