Thoughts for 400GbE FEC Architecture with Pre-interleave

Xinyuan Wang, Tongtong Wang

HUAWEI TECHNOLOGIES CO., LTD.

IEEE 802.3bs 400 GbE Task Force

HUAWEI

Background and Introduction

- In Hawaii meeting, how to interleave from multi-Codewords in KP4 FEC is one of key item to be investigated for moving 400GbE standard forward

  Method of forming PCS lanes from FEC codewords is TBD and dependent on further error analysis

  gustlin_3bs_02c_0715

- In this contribution, following work are investigated to address FEC performance & architecture in 400GbE

  - Update FEC performance from "wang_x_3bs_01_0715"
  - Observation of FEC architecture of 400GbE

Assumption for KP4 FEC Performance Analysis

- Summary from previous discussion or consensus
  - End-to-End FEC to cover both of optical and electrical link
  - Two-part link model
  - Bit-multiplexing only in PMA
  - Up to 5 interfaces from optical and electrical link
  - Random error only from optical link
  - Random error only from C-M electrical link
  - Burst error by DFE error propagation only from C-C electrical link
  - Single signal level transition error in PAM symbol as only Gaussian noise included as in “wang_t_3bs_01a_0315”
  - DFE Error propagation probability for PAM4 signaling:
    - a=0.75 as in “wang_t_3bs_01_0515”
  - 0.2dB KP4 FEC coding gain to cover electrical links

Options of Forming PCS lanes from FEC Codewords

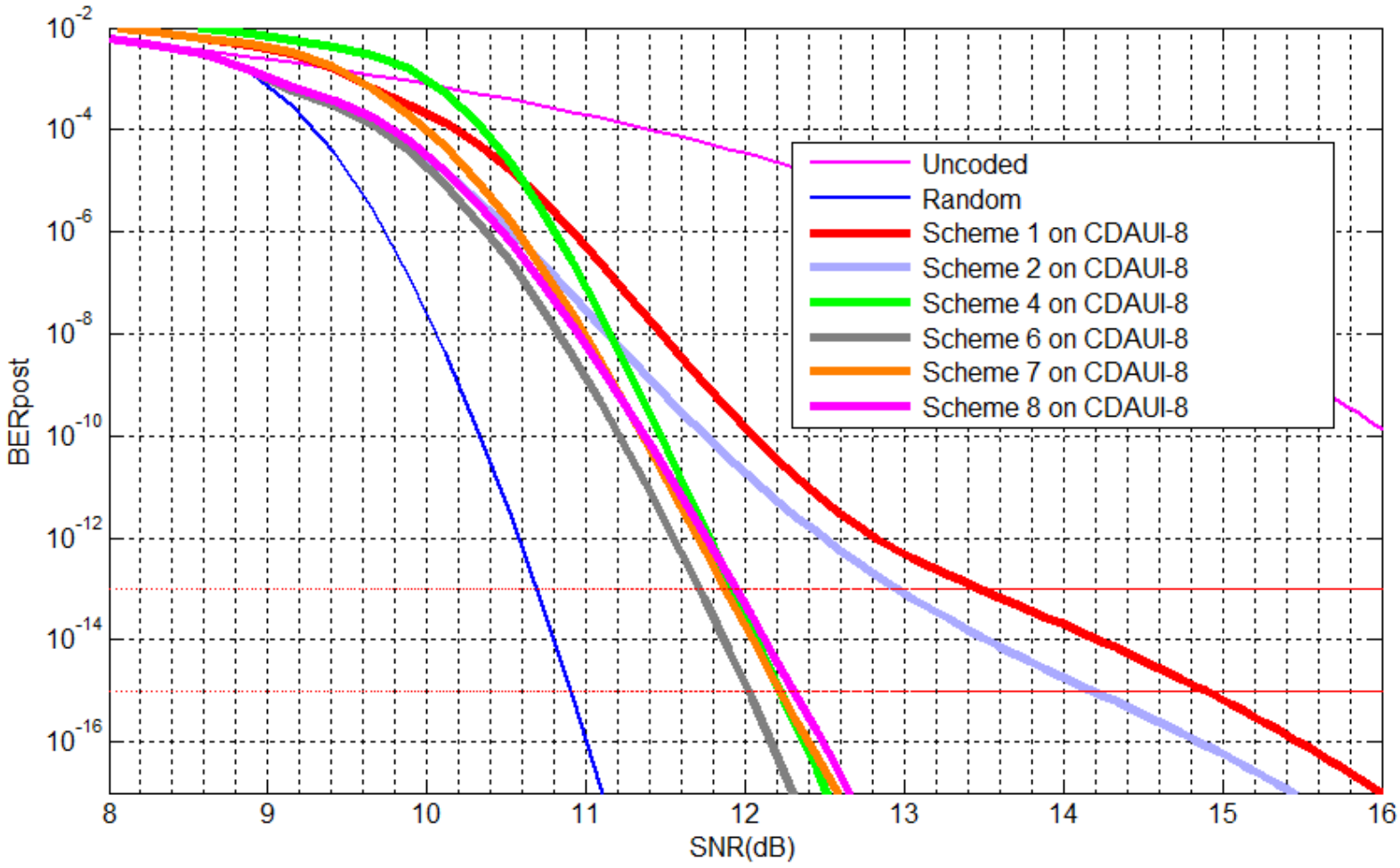
- Scheme 1: One FEC instances with bit multiplex in PMA. No Pre-interleave
- Scheme 4: Pre-interleaved by bit granularity as "wang_t_3bs_01a_0115"
- Scheme 6: Alternative 4-way Pre-interleaved by RS FEC symbol granularity
  - Scheme 6a: Worst case w/ bit mux between RS FEC symbols from same FEC codeword;
  - Scheme 6b: Best case w/ bit mux between RS FEC symbol from different FEC codewords;
- Scheme 7: FOM bit mux
  - Scheme 7a: 2:1 mux for 50G per lane
  - Scheme 7b: 4:1 mux for 100G per lane
- Scheme 8: alternative 2-way Pre-interleaved by RS FEC symbol granularity in worst case

*: All of Scheme # is inline with "anslow_01_0815_logic"

Updated Assumption after July Meeting

- Maximum Burst Length:
  - The initial proposal of Scheme 4 in "wang_t_3bs_01a_0115"
    - All PAM4 error patterns for 2-17 symbols long burst errors
    - Good candidate for facing short burst as in previous analysis
    - Average BER between multi physical lanes in "wang_x_3bs_01_0715"
  - From "FEC Group Weekly Meeting" after Hawaii meeting, up to 75 PAM4 symbol burst error per FEC codeword is considered
  - The maximum burst length will significantly influence FEC performance and final selection for FEC architecture

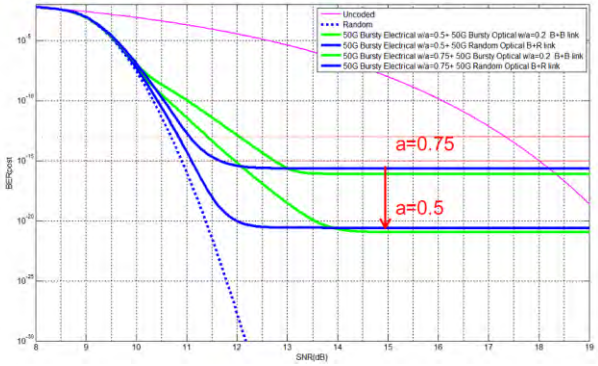
Updated KP4 FEC Performance for PAM4 Links

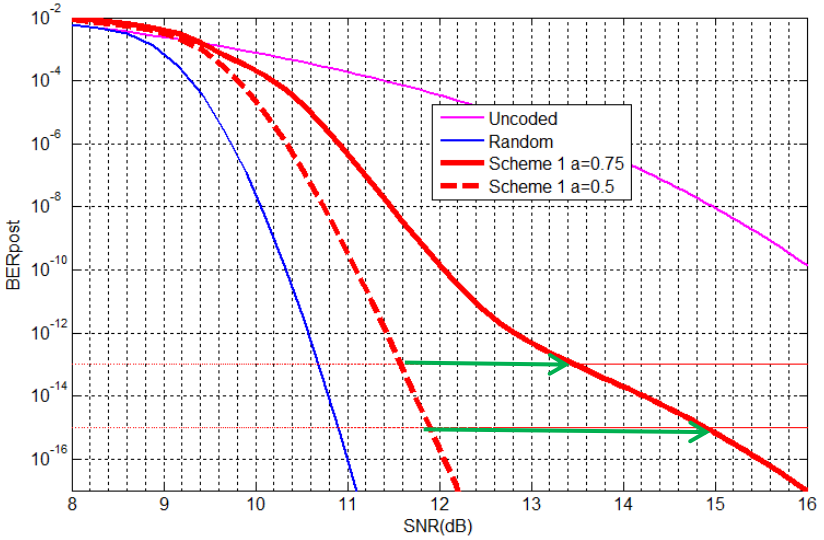


Observation 1: Risk of Scheme 1 with limit “a”

- In “wang_x_01_3bs_0715”, the following figure shows the error floor impacted by “a” parameter from 0.5 to 0.75
- FEC performance degraded by “a” parameter from 0.5 to 0.75, especially for Post-BER>1E-15

- Error propagation parameter “a” of DFE will significantly shift error floor of Non-FOM Bit Mux even from 0.75 to 0.5.
- Considering burst error from optical physical link, FEC performance by Non-FOM Bit Mux will be further degraded.
- Architecture with Non-FOM Bit Mux can’t effective benefit from burst error correct capability of RS FEC.

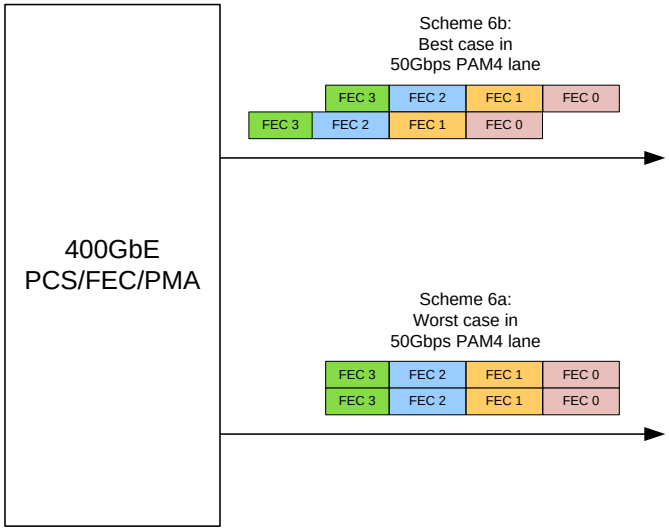


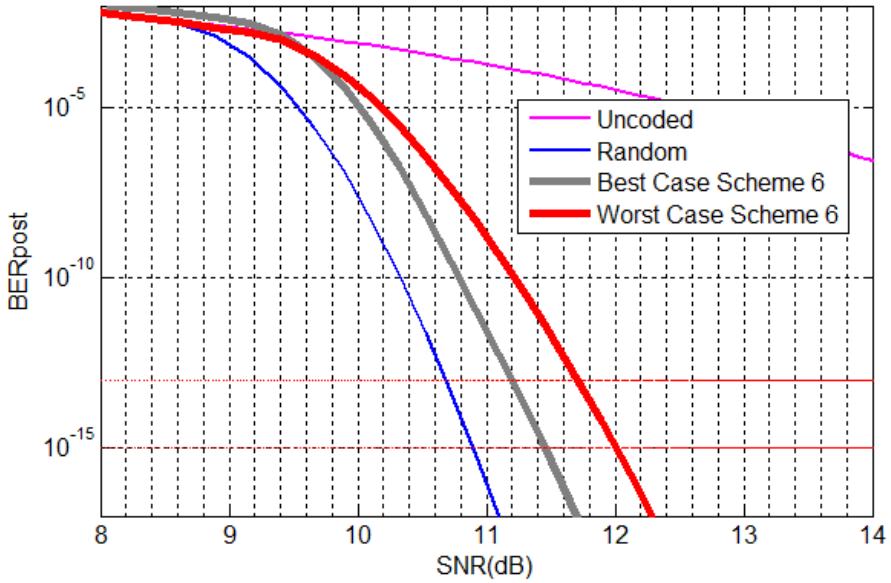


- Limiting implementation of DFE architecture and tap coefficient

Observation 2: Deliverable of Scheme 6b

- For Host ASIC with 8X50G PAM4 C-M electrical interface, it is feasible to multiplex LSB/MSB of PAM4 symbols from different FEC codeword and thus improve FEC performance.

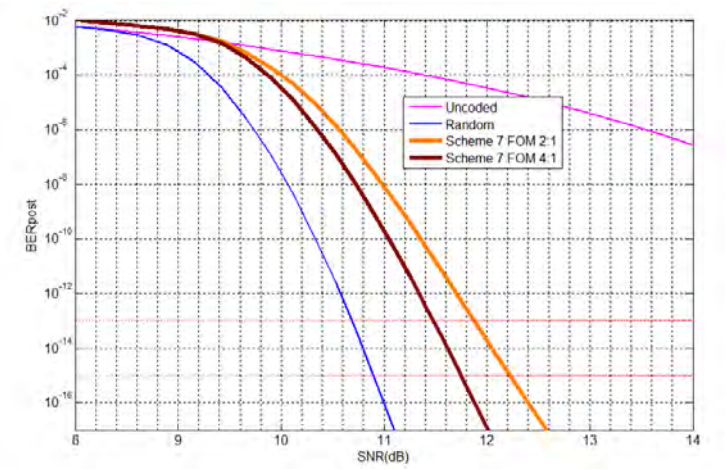




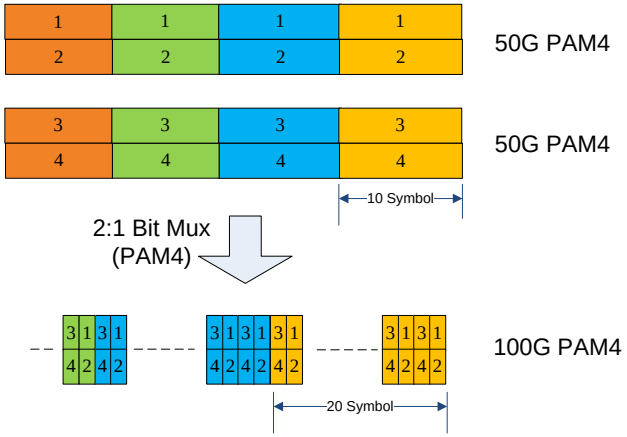
- If no 16X25G physical lanes required, scheme 6b can be implemented and provide ~0.4dB additional coding gain than scheme 6a

Observation 3: 100Gbps per Optical Lanes

- If assuming ONLY random error from optical link, FEC performance for 100Gbps and 50G optical lane will be same
- However, considering burst errors:
  - With Scheme 7a vs 7b (aka FOM), FEC performance will be improved for 100Gbps lane



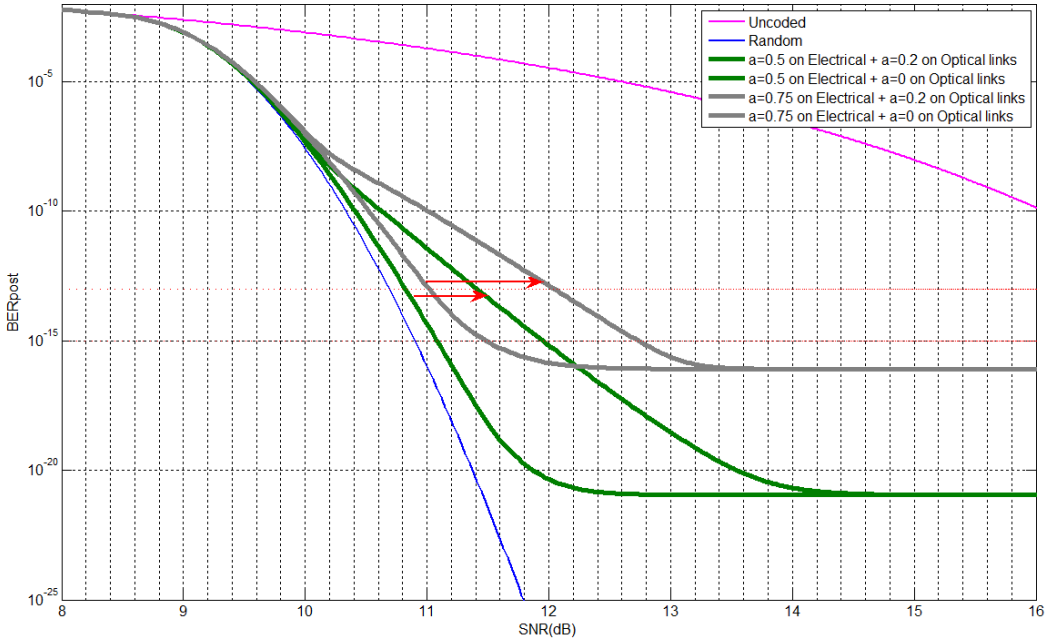
  - With Scheme 6 and 8, FEC performance will degrade for 100Gbps lane in worst case



- Foresee FEC performance degrade for Scheme 6&8 in 100Gbps per lane, and scheme 6 gives more coding gain

Observation 4: Burst Error Impact from Optical Link

- In "wang_x_01_3bs_0715", even assuming a=0.2 error propagation probability for burst error in optical link, performance of Scheme 1 will degrade ~0.6/1.0dB with a=0.5/0.75 in electrical link
- With bitmux between symbols in same FEC codeword, Scheme 1 weaken the burst error correct capability of RS FEC



Potential Source of Burst Error from "wang_x_01_3bs_0715"

- Infrequent pattern dependent event.
- DC blocking caps: low frequency cut off coupled with data wonder.
- Long data transition charging effect coupled with non-linear response of O/E devices.
- VCSEL slow turn off (see page 9) has similar error floor as seen commonly in 802.3bs SMF contributions as result of top 3 effects.
  - http://www.ieee802.org/3/100GNGOPTX/public/mar12/plenary/ghiasi_02_0312_NG100GOPTX.pdf
- Implementing long DFE with PAM4 signaling is complex but we shouldn't rule out a short 2-3 taps DFE and an MLSE and assume optical links have absolutely no burst error!

Consideration of FEC Architecture and Performance

- As in “dambrosia_3bs_01_1114” for “FEC Architecture Discussion”:

  Architecture needs to be flexible for future evolution

- Consideration for FEC architecture, it should be
  - Robust enough for all of physical links and independent to specific implementation, for example RX Equalizer
  - Trade off on latency, performance and hardware complexity as in “wang_x_3bs_01a_0115”
- Steep FEC performance curve will help to improve system stability
- Provide good margin of FEC performance that
  - Not at edge working point and has severe error floor;
  - Tolerant for other error except random only, even it is hard to quantify;
  - Permit to further lower post-BER to several order of magnitude

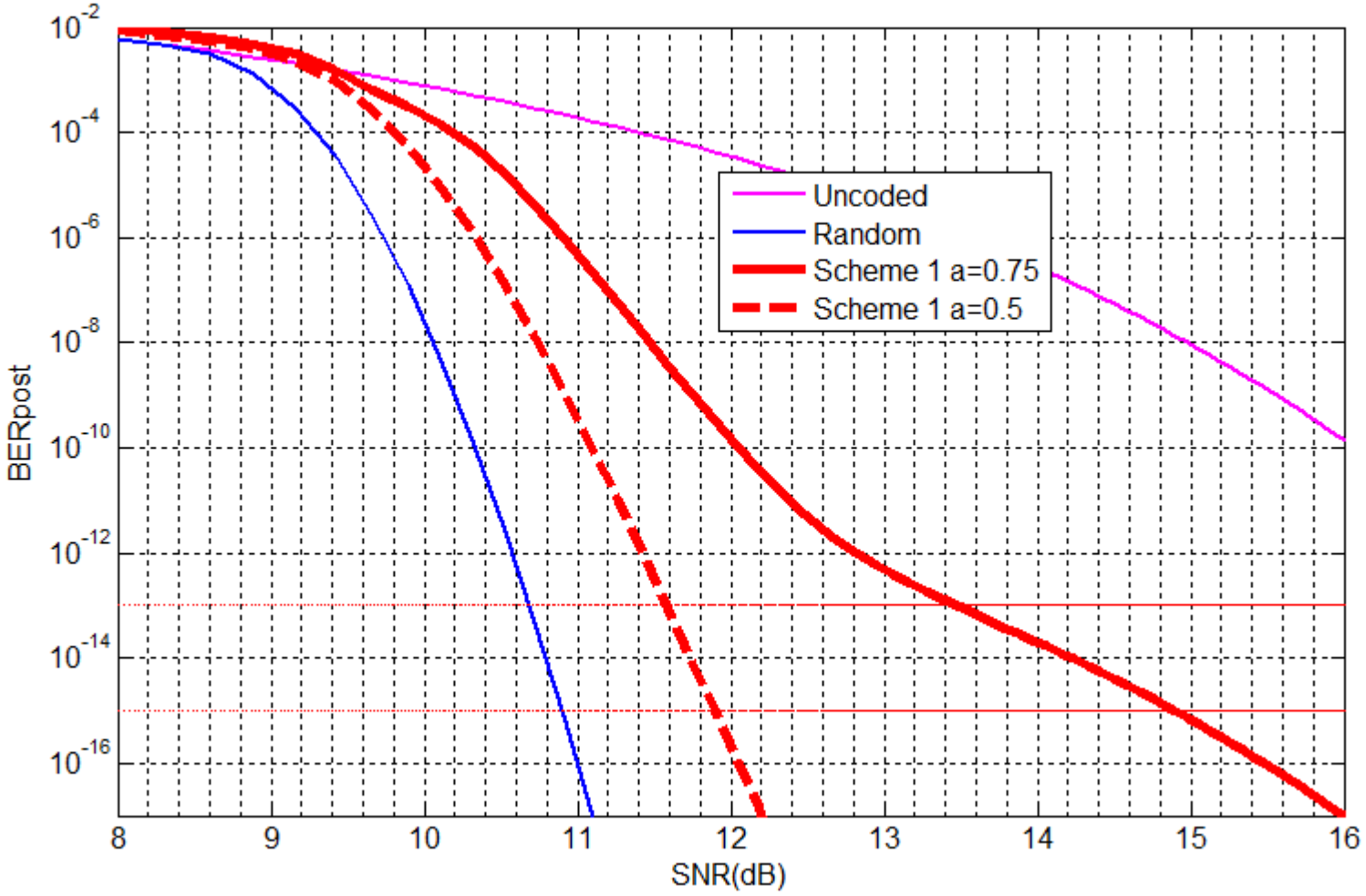
Conclusion

- Scheme 1 has poor FEC performance or limiting to specific implementation.
- Scheme 6 &8 is much better on FEC performance
  - From latency perspective, Scheme 8 is preferred as saving ~26ns
  - From performance perspective, Scheme 6 is preferred as improving ~0.3dB
- 400GbE FEC architecture from implementation perspective:
  - 4X100G architecture is best for Scheme 6 with same latency as 802.3bj
  - 2X200G architecture is best for Scheme 8 with further lower latency
  - 4X100G architecture is nature to support Breakout and Flexible Ethernet

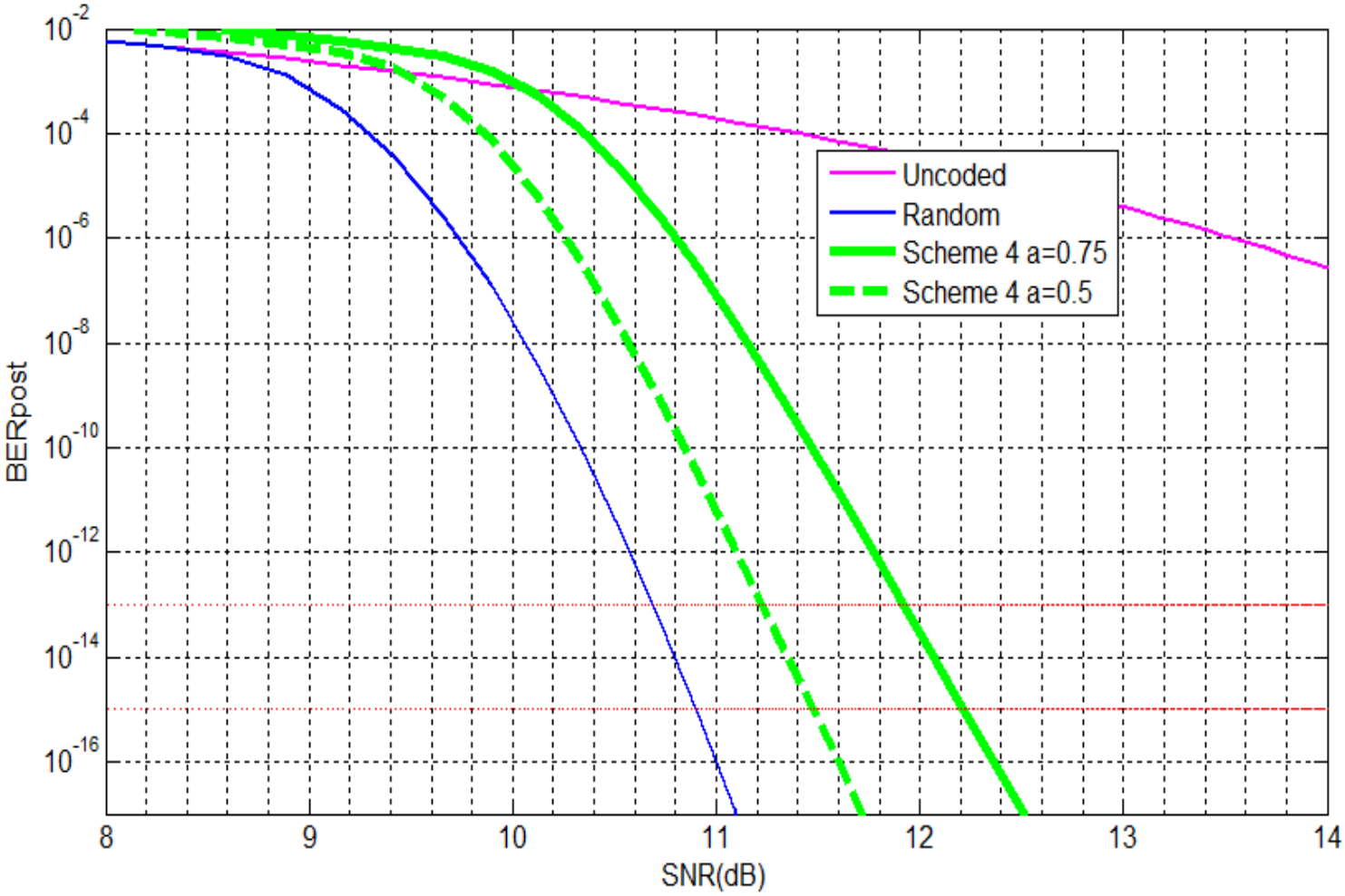
Thank You

BACKUP

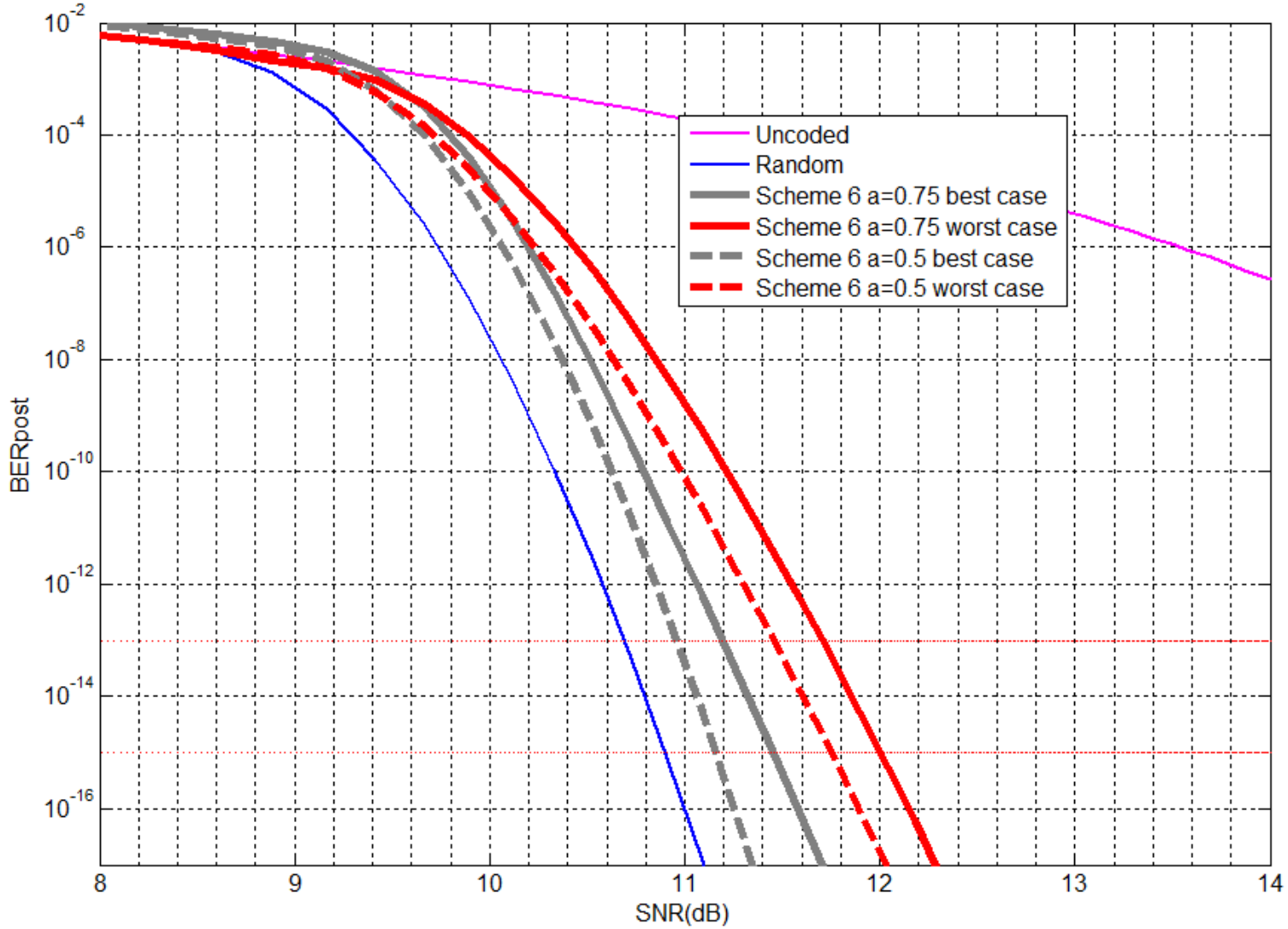
KP4 FEC Performance of Scheme 1



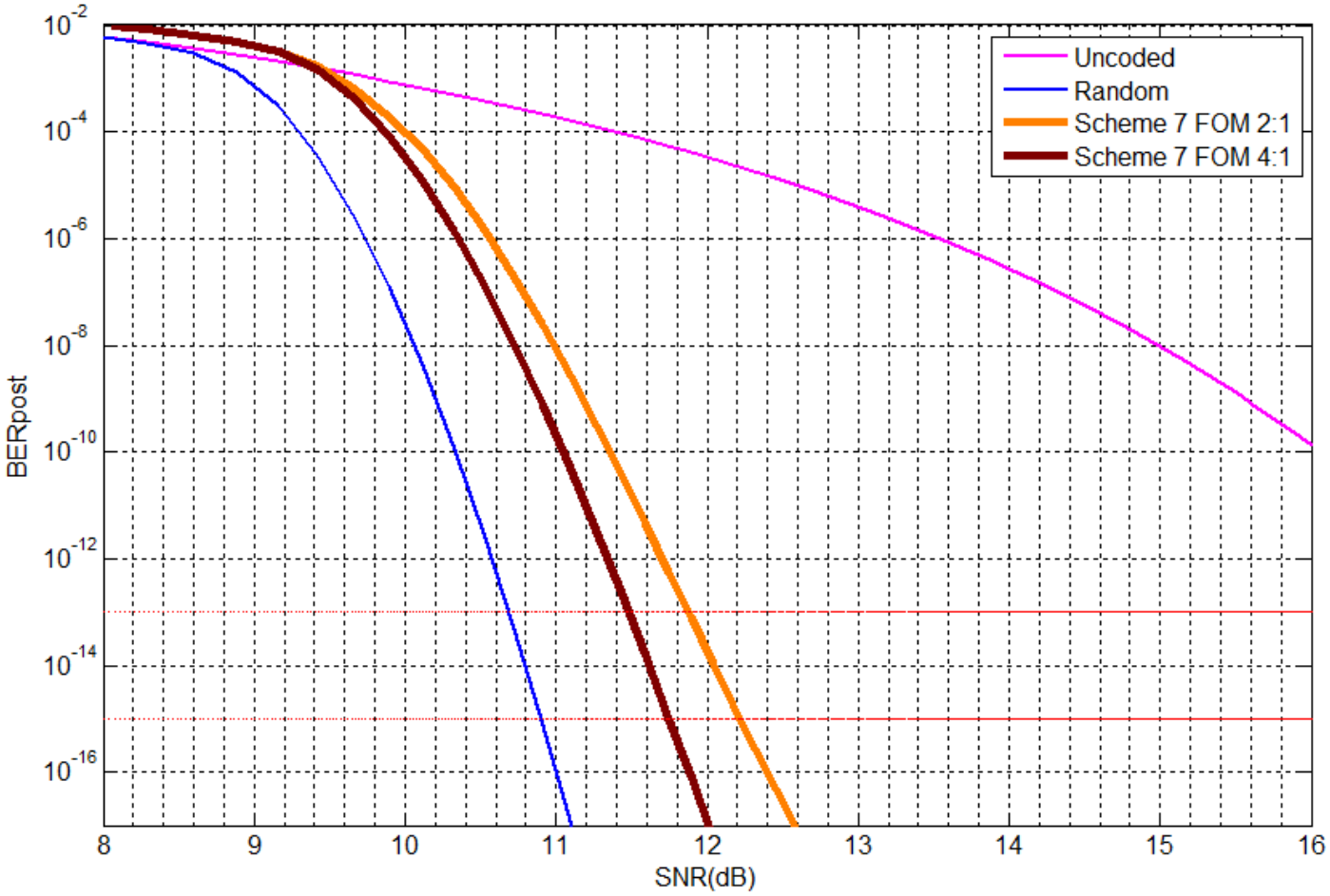
KP4 FEC Performance of Scheme 4



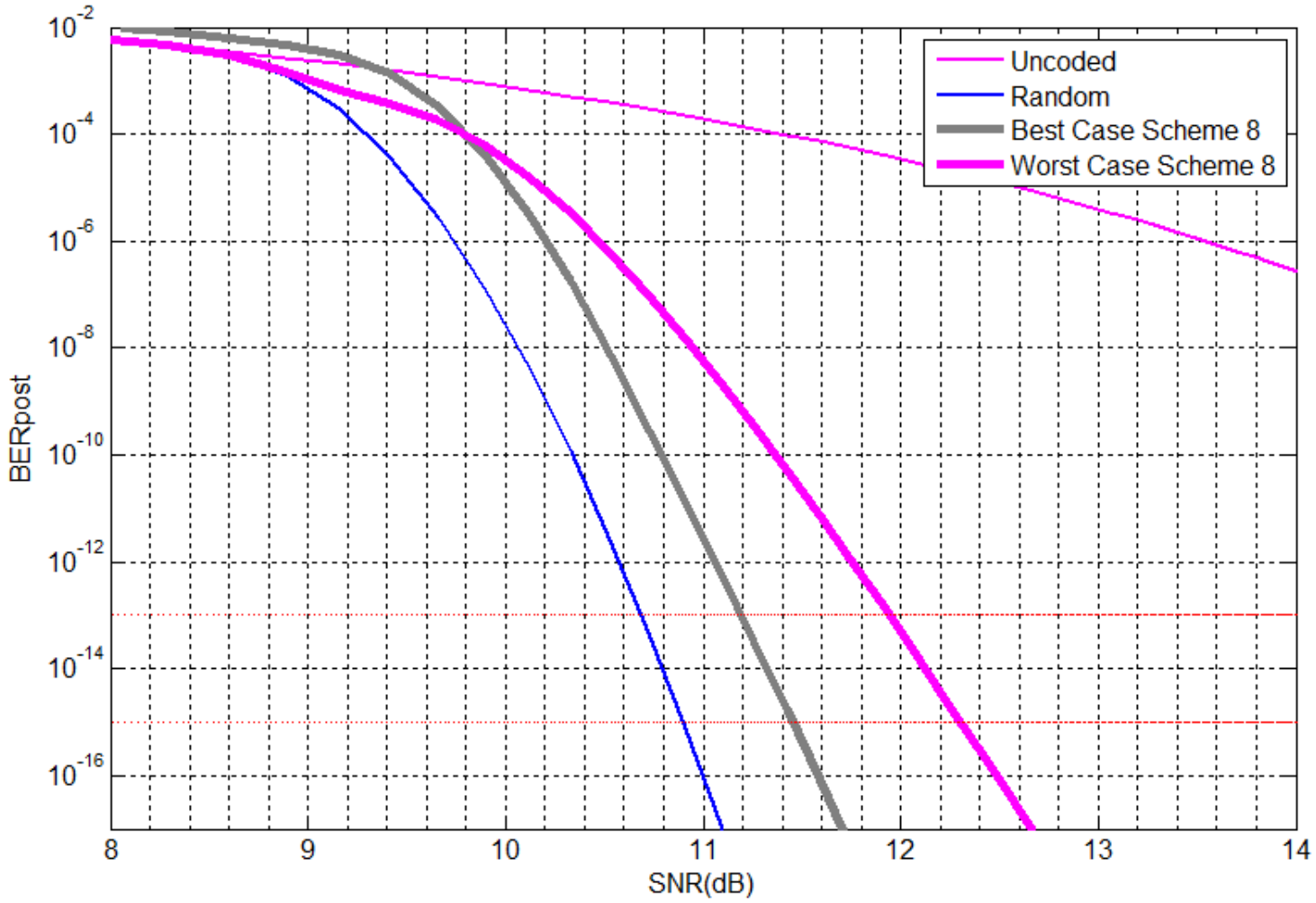
KP4 FEC Performance of Scheme 6



KP4 FEC Performance of Scheme 7



KP4 FEC Performance of Scheme 8



Precoding in "wang_t_3bs_01_1114"

Use Precoding in Bursty Links

- **Precoding Characteristics**
  - Help reduce ISI by coding at source
  - "Pre-coding [17] reduces the effect of DFE error propagation for a 1-tap DFE by breaking up DFE error bursts so that each error burst turns into two single errors after decoding. Pre-coding is less effective in breaking error bursts generated by multi-tap DFEs"*

    -- *R. Cideciyan et al. 《Next Generation Backplane and Copper Cable Challenges》, Dec 2013*
- **FOM Bitmux**
  - "Divide and conquer" errors
  - No dependence on DFE Tap numbers, comparing to precoding.

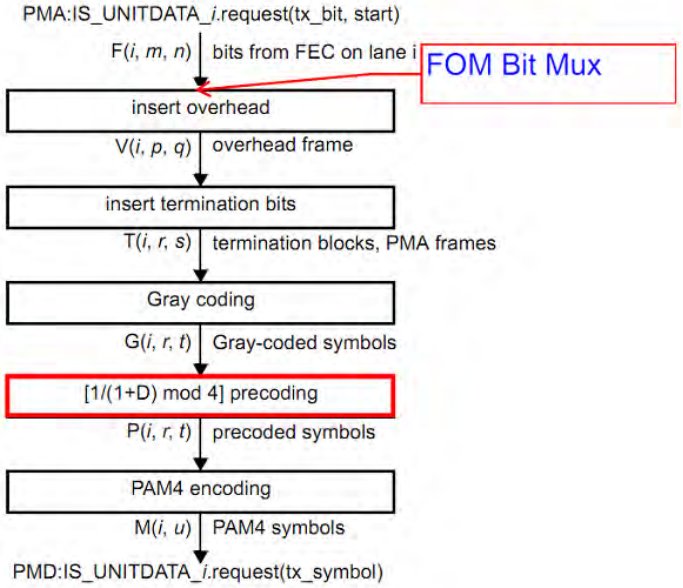
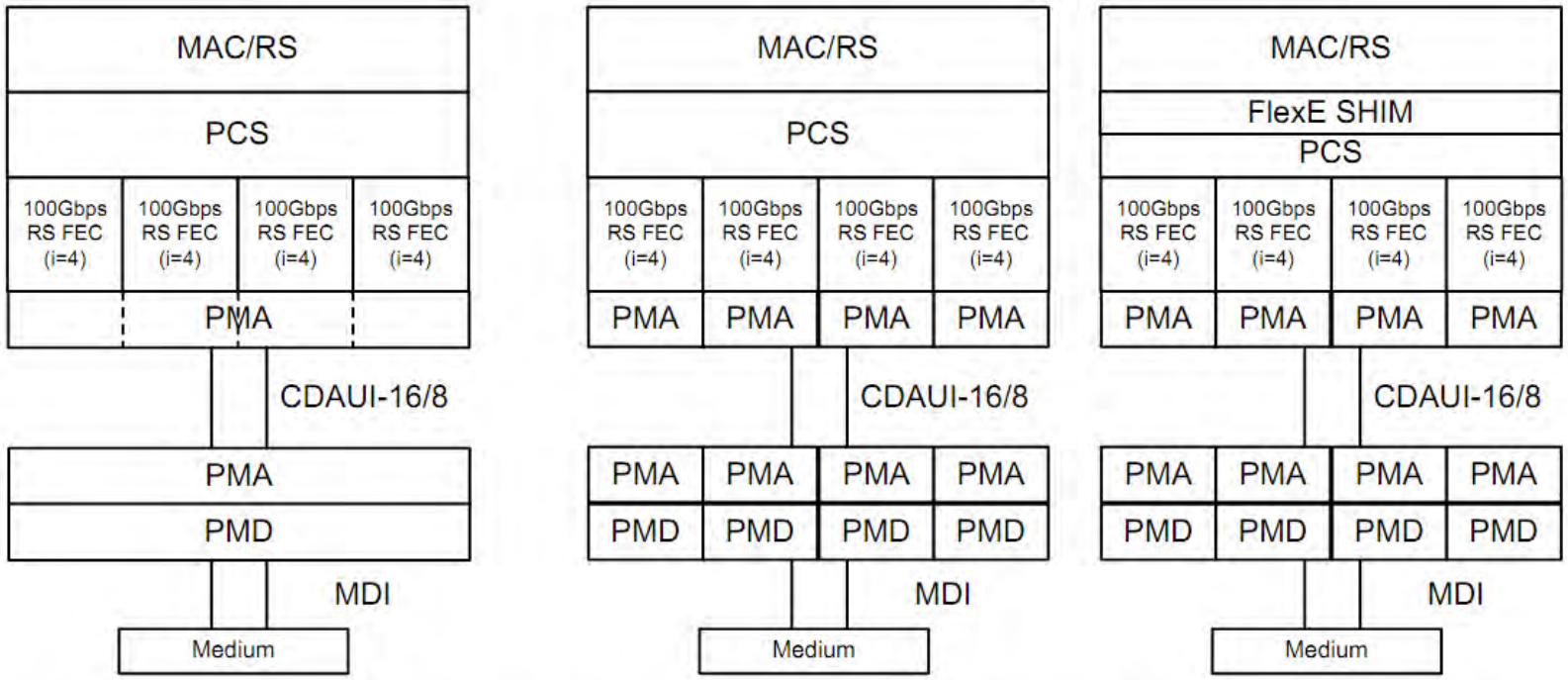


Figure 94–2—Transmit adaptation process diagram

- **Precoding can be enabled with FOM bitmux method to improve FEC performance.**

ONE FEC Architecture in Ethernet in "wangx_01_0615_logic"

- 400GbE with 4 pipeline 4X100G FEC
- Breakout into 4X100GbE
- 400Gbps FlexE with 4 PHY bonding



- 4X100Gbps FEC proposal is an excellent option to unify FEC architecture in Ethernet, even in ITU B100G is one of potential candidate yet.

FEC Discussion in 802.3bj

- From performance perspective, concern on ONE FEC architecture with simple bit mux is already raised.

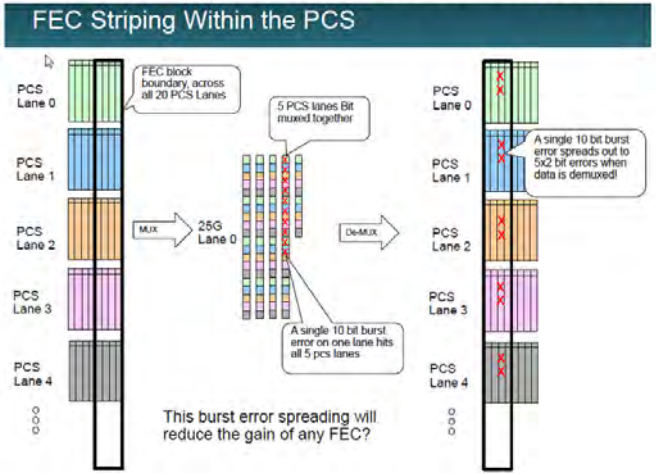
Thoughts?

- Striping of a FEC block across multiple physical lanes causes burst error spreading (bursts are split up)
- This is fatal for a block FEC like the KR Firecode
  - KR FEC can correct a single burst error up to 11 bits, if that burst is split up then it cannot correct the errors
- This can greatly impact the correction capability of a Reed Solomon code, depending on how the FEC blocks are striped
- Creating FEC blocks within the PCS, and distributing the FEC blocks to all 20 PCS lanes, and then allowing the normal 100 Gb/s bit multiplexing seems fatal to the FEC burst error correction capability
- What will work well with a FEC code is distributed to multiple lanes?
  - It looks like Reed Solomon codes are good candidates for this, as long as striping is done multiples of the symbol size (m)

gustlin_02_0911

FEC within the PCS Layer

- Why not put FEC into the PCS layer and allow the normal PMA muxing that we do for 802.3ba?
- See gustlin_02_0911, doing this would cause any burst errors to be split up into multiple individual errors and weaken most FEC codes



gustlin_01a_1111