Low Latency Options for 800GBASE-DR4

Bill Simms and Vishnu Balan
NVIDIA

ACKNOWLEDGEMENTS

- Contributors
  - Tony Zortea - NVIDIA
  - Piers Dawe – NVIDIA
  - Vishnu Balan – NVIDIA

- Supporters
  - John Calvin - Keysight

- References
  - “Updates on Concatenated FEC Proposal for 200G/Lane PMD”, Lenin Patra – Marvell Semiconductor
    - https://www.ieee802.org/3/df/public/22_07/patra_3df_01a_2207.pdf
  - “FEC Architecture of B400GbE to Support BER Objective“, Xiang He, et al – Huawei Technologies
    - https://ieee802.org/3/B400G/public/21_05/he_b400g_01_210426.pdf
  - “DSP and FEC Considerations for 800GbE and 1.6TbE”, Yuchun Lu, et al – Huawei Technologies
    - https://www.ieee802.org/3/df/public/22_02/lu_3df_01b_220215.pdf
  - “Baseline Proposals for 800GBASE-DR-4, 800GBASE-DR4-2, and 800GBASE-FR4”, Brian Welch – Cisco Inc
    - https://ieee802.org/3/df/public/22_05/22_0602/welch_3df_01b_220602.pdf
  - MSA FFE 225G curve extracted from 800G Pluggable Multi-Source Agreement White Paper
    - https://static.s123-cdn-static-d.com/uploads/2598123/normal_60d5f55c54664.pdf

OBJECTIVES

- Propose a low-latency end-to-end FEC option for DR channels at 200G/lane
  - Leverage existing 100G/lane infrastructure and simply double the bandwidth
  - Saves the additional area, power, and latency of an SD-FEC in the module DSP
  - Saves any additional link up steps required by an SD-FEC

- Show preliminary simulations of EML based 200G/lane DR channel for pluggables
  - Channel models representative of the copper traces in the module PCB
    - PCB route from the line driver of the DSP output to the EML modulator
    - PCB route from the TIA output to the DSP receiver
  - Highlight achievable improvements in laser launch power, Rx sensitivity, laser RIN numbers, etc

- This presentation may not eliminate the need for SD-FEC in longer reach applications such as FR and LR
- This presentation is not an exhaustive evaluation of all DR links but shows what is feasible

BACKGROUND

- Enterprise AI and ML solutions are a growing segment of data center
- The majority of AI and ML servers will use <<500m optical links for interconnect making them equivalent to DR and less like FR and LR
- Some critical AI and ML applications require low-latency and low-power connections
- DR channels (200G/lane) with SMF-based optics under 500m should plan to provide low-latency connections for AI/ML applications
- Avoiding the need for an SD-FEC lowers latency, power, and area

BENEFITS OF AVOIDING SD-FEC FOR DR CHANNELS

- Lower power consumption and silicon area compared to the SD-FEC + Interleaver scheme
  - SD-FEC may use an additional 100mW per 200G lane (*1)
  - 0.17 to 1.36 relative area increase over a 2-way RS(544,514) HD-FEC [*2]
  - Bypassing SD-FEC rather than removal still adds silicon area and may use additional power in some implementations
- Data rate can be a simple doubling of existing 100G/lane solutions-> 212.5 Gbps
  - No new clock synthesis/gearbox required in the module
  - Maintaining the bit-mux features of the CDR for breakout applications may be made more cumbersome by SD-FEC
- Maintains DSP latency numbers comparable to previous generation
  - Additional SD-FEC latency estimated between 9.6 and 140 ns depending on the choice of coding and use of interleaver [*1, *2]
- Avoid complex SD-FEC related steps in module link-up
  - Symbol and frame alignment steps for DSP CDR with “high” raw BER (~5 x 10^{-3}) can be challenging
  - SD-FEC compatibility with DFE or MLSE based equalization is questionable and complex at best [*3]
    - DFE error propagation may corrupt SD-FEC decode
    - ‘There are exclusive relationships between FEC technologies (i.e. HD-FEC and SD-FEC) and DSP technologies and should be explored more deeply’ [*3]

*1: Updates on Concatenated FEC Proposal for 200G/Lane PMD [patra_3df_01a_2207.pdf]

*2: FEC Architecture of B400GbE to Support BER Objective [he_b400g_01_210426.pdf]

*3: DSP and FEC Considerations for 800GbE and 1.6TbE [lu_3df_01b_220215.pdf]

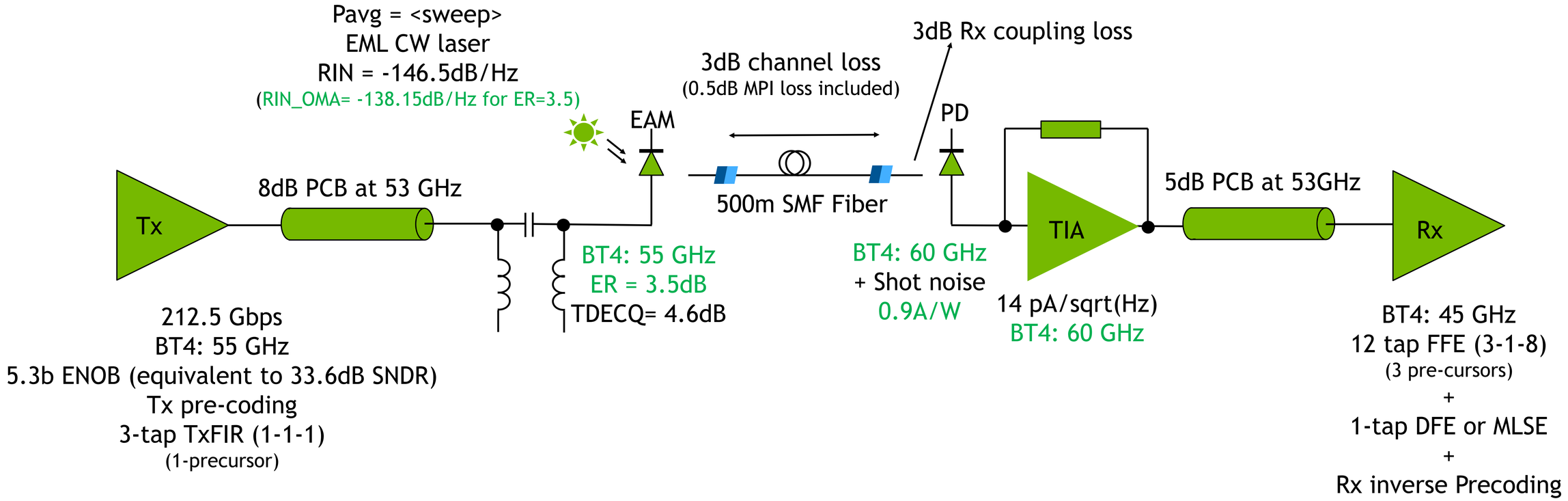
SIMULATION SETUP

- Construct a channel relevant to AI and ML applications
  - Simplified DR channel characteristics were developed to accelerate simulation time
- Sweep laser launch power, Rx sensitivity, laser RIN numbers, etc to optimize TDECQ
  - Based on prior studies [*4]
- Optimizations done for circuit parameters shown in diagram on following page
- Simulation parameters obtained from public domain 800G Pluggable Multi-Source Agreement

*4: Baseline Proposals for 800GBASE-DR-4, 800GBASE-DR4-2, and 800GBASE-FR4 [welch_3df_01b_220602.pdf]

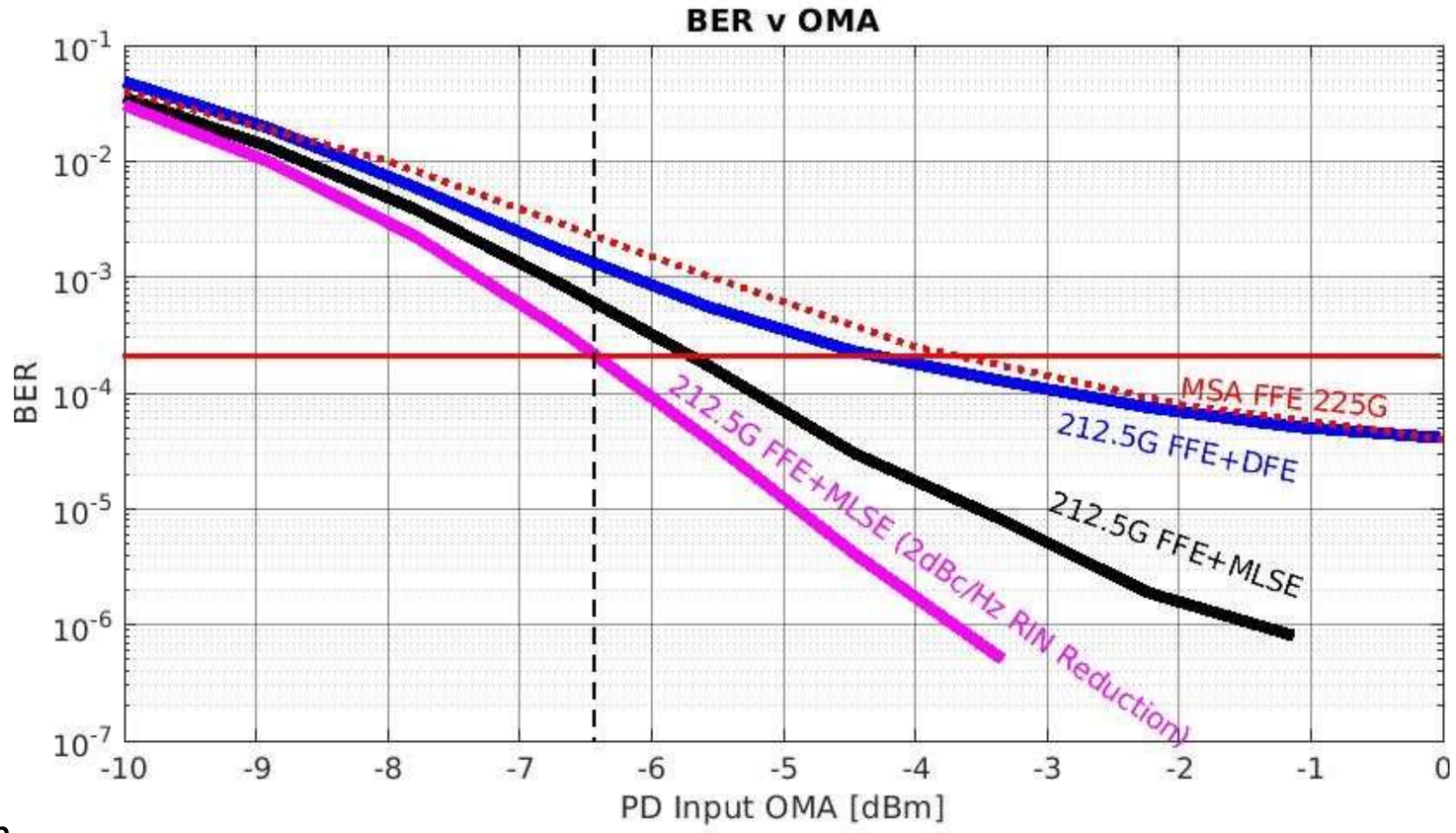
SIMULATED DR CHANNEL

- A simplified DR channel model that illustrates the feasibility is shown
  - Simplified PCB model
  - No Xtalk
- The model assumes certain channel parameters as indicated in the figure
  - Green highlighted numbers aligned with MSA choices



SIMULATION RESULTS

- A plot of BER vs. OMA (at the Rx photo-diode) is shown for 212.5 Gbps:
  - Baseline EQ : FFE + 1-tap DFE (Blue)
  - Advanced EQ : FFE + MLSE (Black)
  - Advanced EQ : FFE + MLSE + Reduced RIN (Magenta)
  - 800G MSA shown for reference (Red dotted)
- BER target of 2 x 10^{-4} (Red) can be achieved under given conditions
- We can rely on end-to-end KP4 RS(544,514) FEC in the host under the usual assumptions that the AUI links can meet BER of 10^{-5} or better
- With this example the link budget is closed with -6.4dBm



*5: MSA FFE 225G curve extracted from 800G Pluggable Multi-Source Agreement White Paper 200g-per-lane-for-future-800g-and-16t-modules

CONCLUSION

- Simulation results for a DR (500m, SMF) channel which closes the link budget without requiring an SD-FEC in the module for pluggables

- Achievable improvements compared to 100G/lane may be needed to make this a goal for DR channels
  - Laser launch power, Rx sensitivity, and laser RIN numbers, etc
  - Preliminary data from component vendors (Lasers, Laser Drivers, TIA, PD …) indicate this is possible
  - Optical component vendors expect to improve performance in the 200G /lane timeline
  - Prior presentations have demonstrated system parameter scaling requirements to achieve link budget closure [*4]

- Significant advantage in terms of latency, power, area, and complexity by skipping SD-FEC for DR channels
  - Some critical AI and ML systems are particularly sensitive to latency and power concerns
  - Other DR applications will benefit as well
  - The longer reach (2km or 10km) SMF channels (FR, LR) may still justify the use of SD-FEC

- Leverage existing 100G/lane ecosystem
  - Doubling of 106.25Gb/s data rate and no need for additional SD-FEC BW overhead

- We will continue to share further detailed simulations to justify the goal of not requiring SD-FEC for DR channels

THANK YOU