Considerations on Beyond 400G

Leon Bruckman, Yuchun(Louis) Lu, Yan Zhuang, Huawei Technologies

Contributors

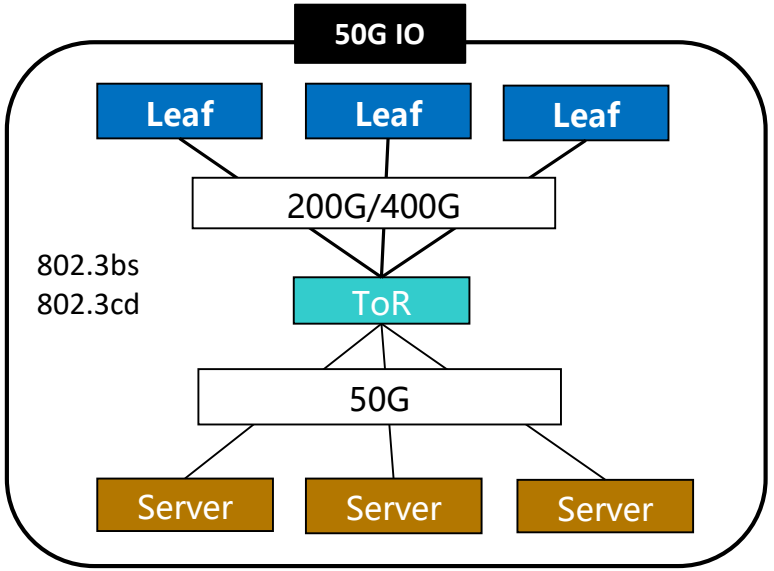
- Matt Brown, Huawei Technologies Canada
- Henry Wong, Huawei Technologies Canada
- Davide Tonietto, Huawei Technologies Canada

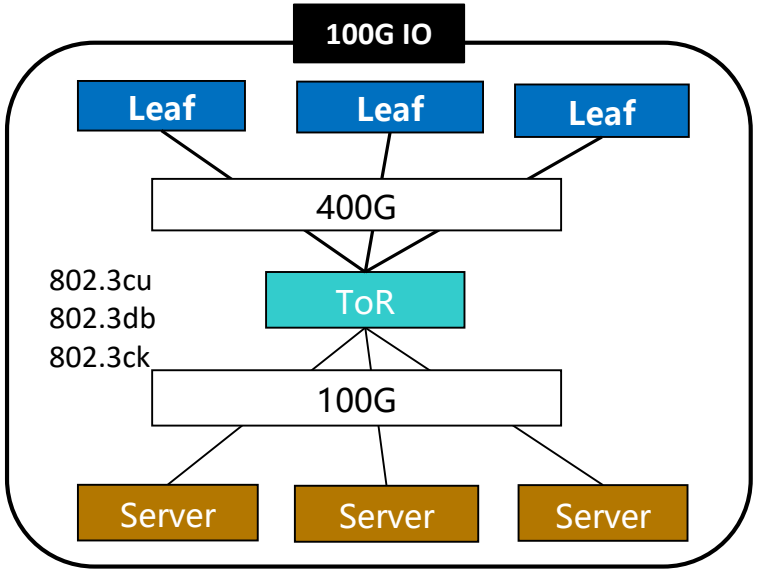
Many thanks to the contributors for reviewing and improving the slides.

What is beyond 400GbE?

- The impact of beyond 400GbE is not related only with standardizing an 800G or 1.6T port, but involves a whole system speed upgrade, the implications should be carefully analyzed and considered.
- So, we would like to start with presenting an example of a Datacenter interconnect network upgrading process, with the goal of opening and encouraging this group’s discussions.

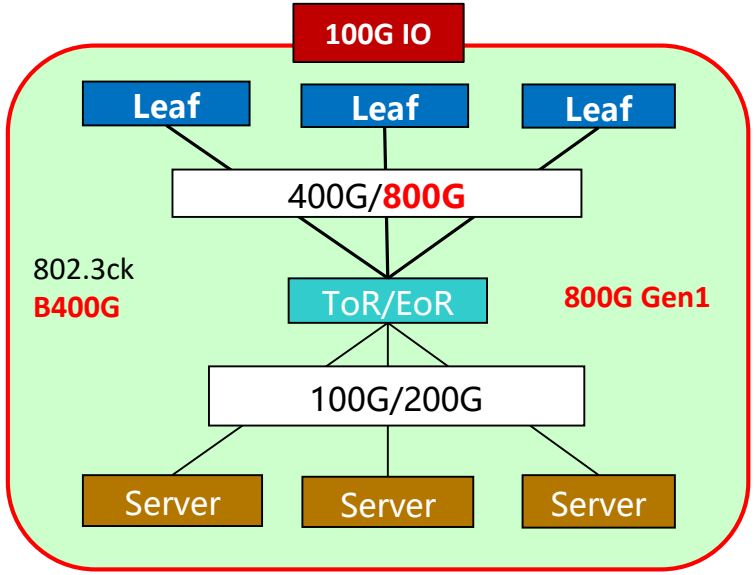
An example of Datacenter Speed Upgrade

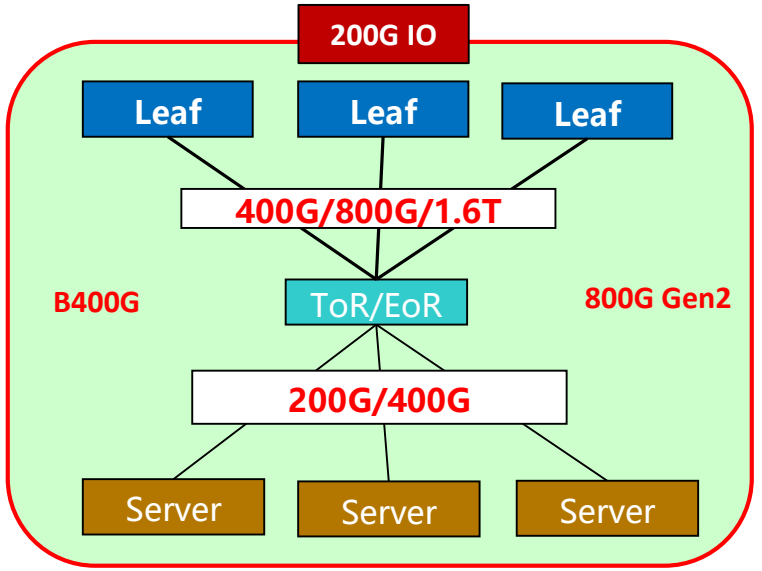




Nowadays

- IEEE 802.3 defines 50G IO and 100G IO-based ports in 802.3bs/cd/ck/db/cu to support up to 400Gbps.
- Products shipped to the market starting 2020 and deployed in the DCNs.
- Breakout is also supported on ToR ports to connect multiple lower speed servers.





B400G SG

- 100G IO might be considered for a Gen 1 of 800G which that can reuse the 100G/lane ecosystem
- But 200G IO should be the key for the B400G SG to define new 200G/lane techniques able to provide 800G Gen2 and 1.6T as well as existing rates with fewer lanes.
- Breakout is preferred to provide more flexible access (server) rates.

200Gbps/lane is the key for 800GbE/1600GbE (1)

- 200G/lane Serdes (key for 1.6T pluggable)
  - Objectives (TBD):
    - Reach (XSR, VSR/C2M, KR/C2C)
    - Performance (BER, power, cost, latency considerations are critical…)
  - Technologies:
    - Modulation, signaling and channel requirements
    - FEC
      - Architecture: end to end FEC, concatenated FEC, segmented FEC
      - Design: soft decision, hard decision
    - DSP: FFE, DFE, MLSE
      - DFE was used as an reference receiver [1+0.85D] channel
      - Is DFE still enough? Should we use MLSE?

200Gbps/lane is the key for 800GbE/1600GbE (2)

- 200G/lane Optics (key for 1.6T pluggable)
  - Objectives(TBD):
    - SR/DR/FR/LR/ER/ZR
    - Performance (BER, latency is very critical…)
  - Technologies:
    - Modulation: PAM4 is preferred (but we need to check its coverage)
    - FEC (Joint consideration with 200G Serdes is recommended)
      - Historically, E2E FEC demonstrated the best flexibility and competitiveness with minimal latency and power consumption.

The evolution of 800GbE and 1600GbE

	VSR/C2M	Optics	Specified
800GbE Gen1	8*100G	8*100G	802.3ck & 802.3cu
800GbE Gen1.5	8*100G	4*200G	802.3ck & B400G
800GbE Gen2	4*200G	4*200G	B400G

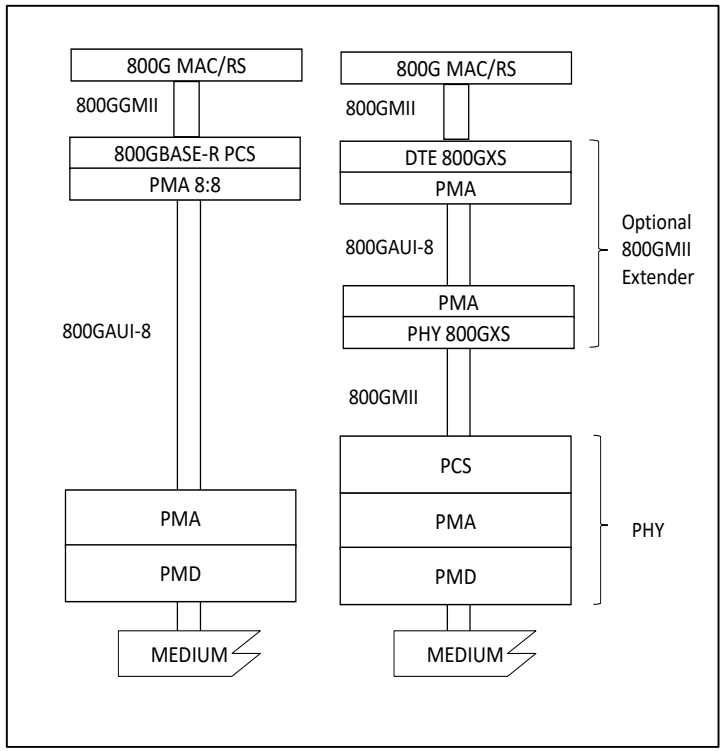
	VSR/C2M	Optics	Specified
1600GbE Gen1	~~16*100G~~	~~16*100G~~	802.3ck & 802.3cu
1600GbE Gen1.5	~~16*100G~~	8*200G	802.3ck & B400G
1600GbE Gen2	8*200G	8*200G	B400G

200G/lane Serdes is the key to 800GbE Gen2 and 1600GbE.

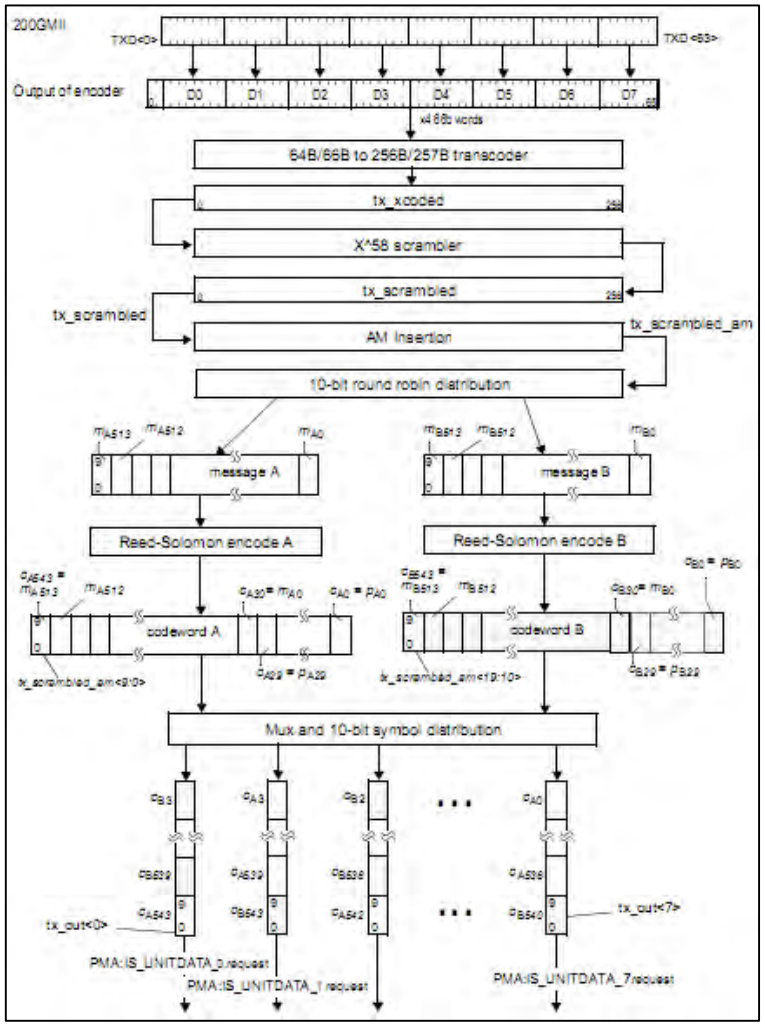
Note: 16*100G would not be used in practice

800GbE Gen1: 100G IO

- Reuse the 200GbE/400GbE architectures in 802.3bs for 800GbE and 1.6TbE.
- Increase lane rate to 100Gb/s.
- Define 800GMII.



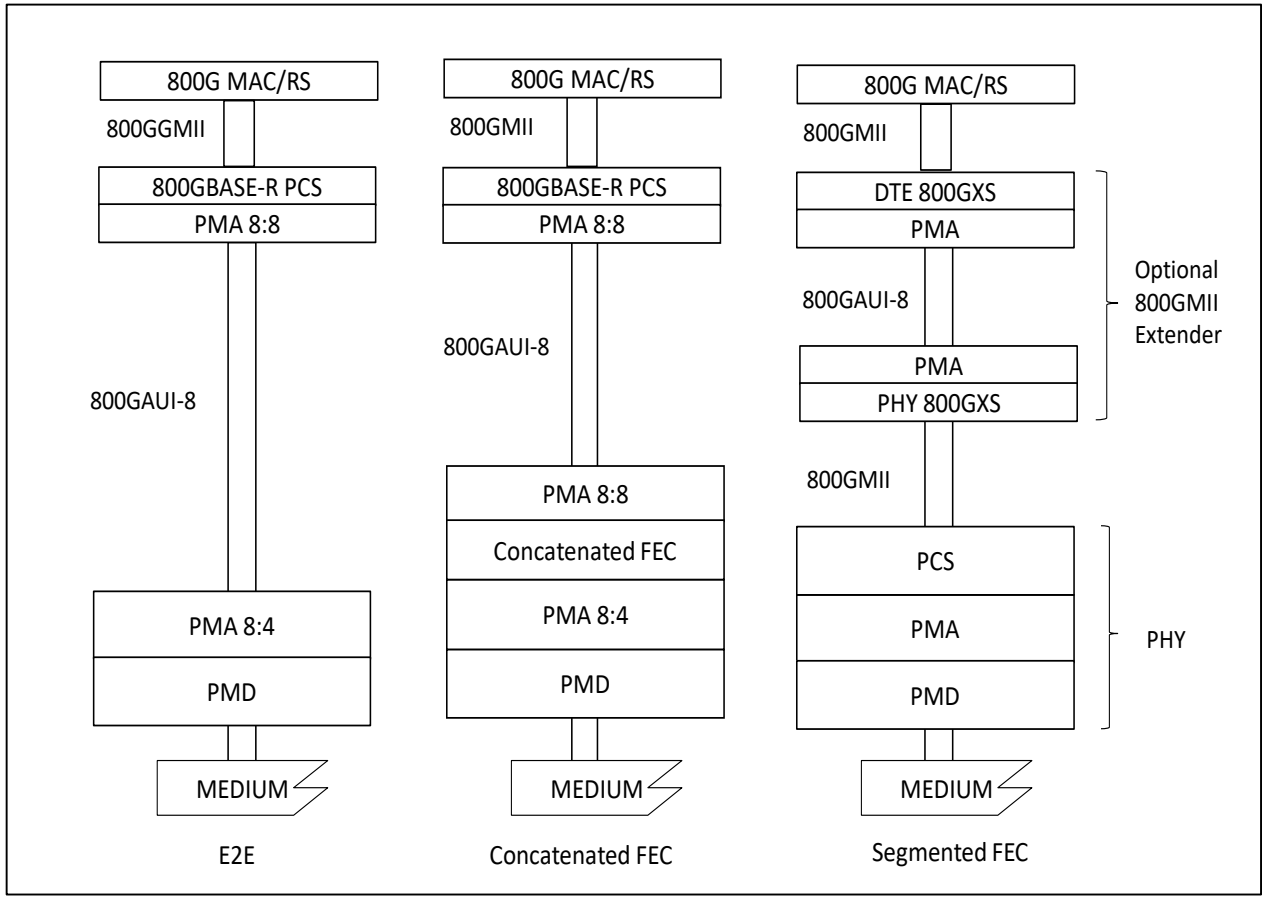
- No 16 lanes pluggable module.
- No gearbox due to same rates on optics and electrics.
- No performance concerns (8:8 PMA).



800GbE Gen1.5: 100G IO

800GbE Gen1.5	8*100G (electrical)	4*200G (optics)	802.3ck & B400G

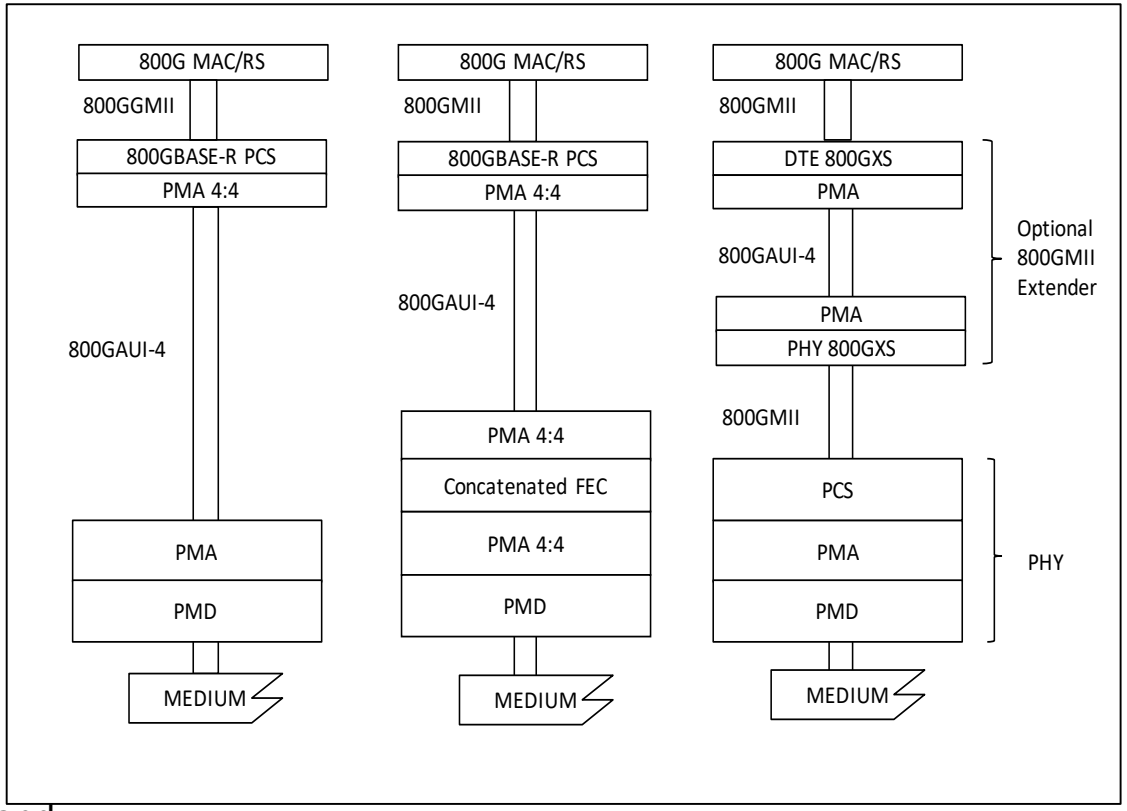
- Reuse 100G/lane electrical defined in 802.3ck.
- Potential FEC architectures
  - End to End FEC: CL119
    - ✓ With better optic and DSP design (good to have, simple)
    - ✓ Whether the coding gain is enough to cover 200G/lane optics
  - Segmented FEC: KP4 + new FEC
    - ✓ Most flexibility, same as 100G-ZR
  - Concatenated FEC: KP4 + inner FEC (new choice, need to be investigated)
    - ✓ Troubleshooting
    - ✓ Coding gain and performance should be investigated especially with burst errors.



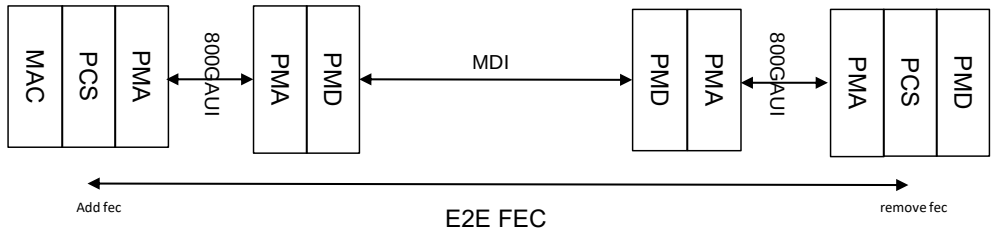
800GbE Gen2: 200G IO

800GbE Gen2	4*200G	4*200G	B400G

- 200G/lane Serdes is critical:
  - Objectives:
    - Reach (XSR, VSR, KR, DR, SR, FR)
    - Performance (BER, power, cost, latency…)
  - Technologies:
    - Modulation: PAM4, PAM6, PAM8, …
    - Potential FEC architectures:
      - E2E FEC
        - RS code (e.g. RS 576,514)
        - …
      - Concatenated FEC
        - (e.g. KP4 on host + inner FEC on module)
      - Segmented FEC
    - FEC Design
      - Hard decision: fits all DSP designs
      - Soft decision: DSP design related (costly for MLSE DSP and difficult for DFE).
    - DSP: FFE, DFE, MLSE



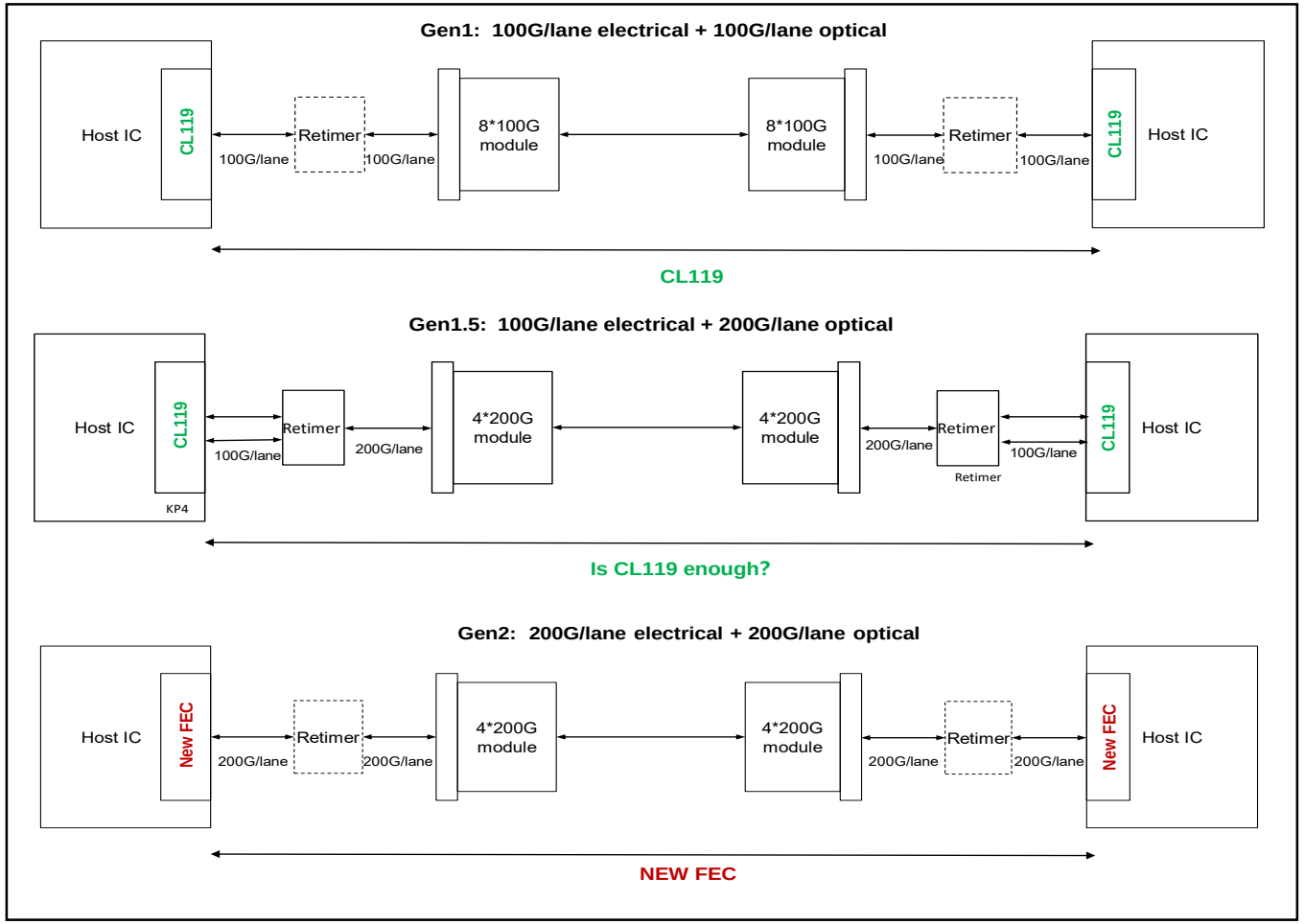
Potential FEC architecture: End to End FEC



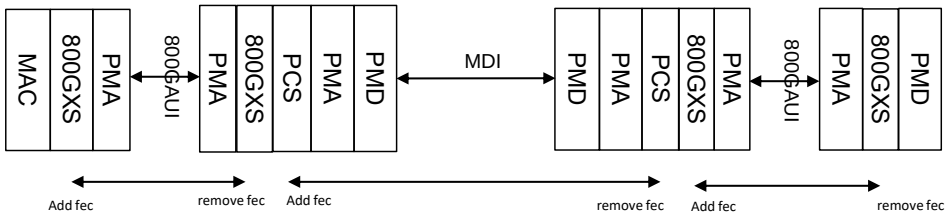
reference architecture

- Considerations:
  - Simple, lowest complexity, latency (latency should not increase too much compared with 400GbE) and power
  - Whether KP4 coding gain is enough for 200G/lane optics and its coverage (if not, better optic and DSP design? Or use a new FEC?).
  - How to allocate FEC error budget across interfaces?

Possible implementation examples:



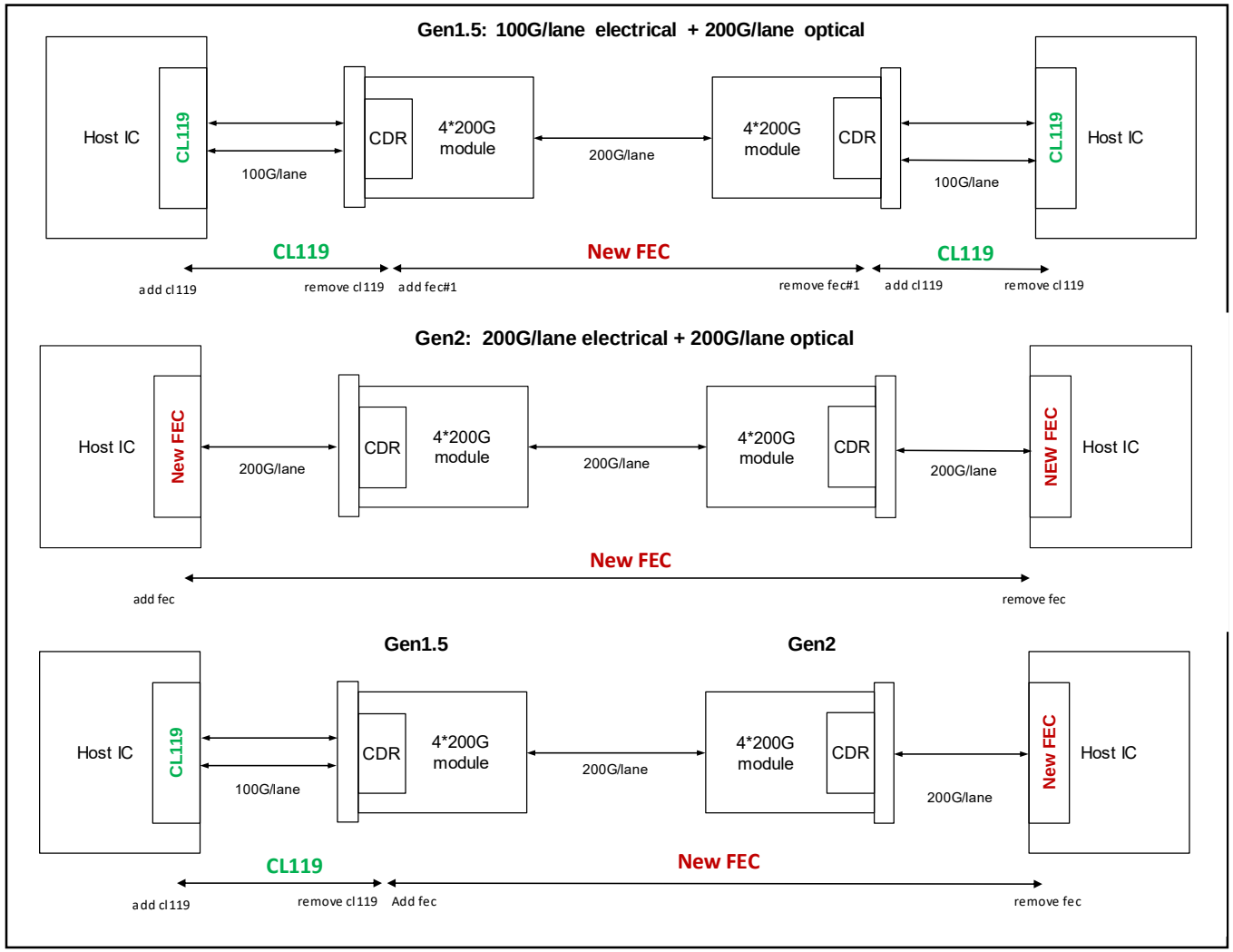
Potential FEC architecture: Segmented FEC



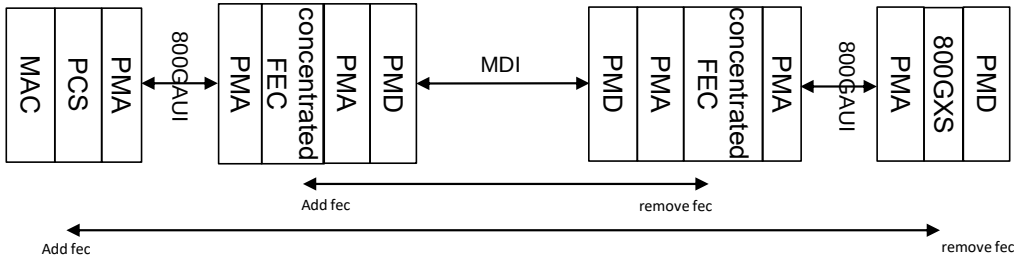
reference architecture

- Considerations:
  - Decoupled electrical and optical channels and FEC can be optimized for each segment.
  - Easy to cope with multiple PMDs.
  - Gen1.5 and Gen2 can interoperate properly.
  - Highest complexity, power, latency etc.

Possible implementation examples:



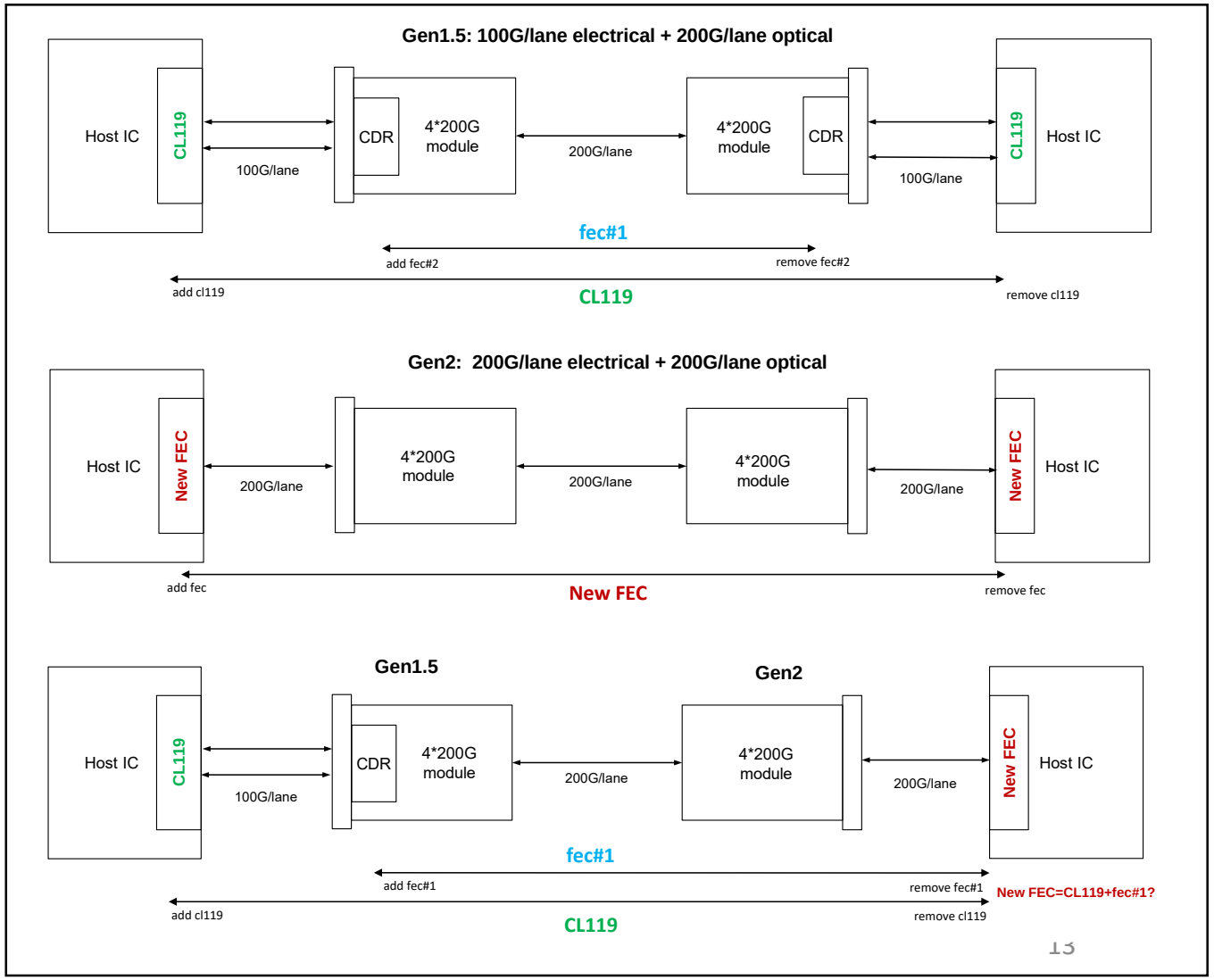
Potential FEC architecture: Concatenated FEC



reference architecture

- Considerations:
  - Flexible FEC choice for each application
  - Should consider the efficiency of the coding gain.
  - Moderate complexity, latency and power
  - Bit rate on different interfaces

Possible implementation examples:



Summary

- Look at the large picture: Start from system considerations.
- Each of the listed generations should be considered by the group to provide flexible upgrade routine for the industry.
  - For 800GbE Gen1 (100G/lane), it can reuse the architecture and PCS sublayer defined in 802.3bs. Fast path to 800G early adopters.
  - For 800GbE Gen2 (200G/lane), which should be the key for this group to provide more cost-efficient solutions.
  - For 800GbE Gen1.5 (200G/lane optics, 100G/lane electric), it can use 100GAUI and 200G/lane modules. The technical path should be considered.
- With this, potential FEC architectures are discussed to cope with the 3 generations.

Thanks!