Baseline proposal for 800GbE and 1.6TbE PCS, FEC and PMA using 100G PMD lanes

Yuchun LU, Yan ZHUANG, Huawei Technologies

IEEE P802.3df Task Force

June 23 2022

More competitive 800GbE&1.6TbE is required

Ethernet Rate	Signaling Rate	Electrical			Optical					
		AUI	Backplane	Copper Cable	MMF 50m	MMF 100m	SMF 500m	SMF 2km	SMF 10km	SMF 40km
200Gbps	200Gbps	Over 1 lane 200GAUI-1	TBD * 200GBASE-KR1	Over 1 pair 200GBASE-CR1			Over 1 pair 200GBASE-DR1	Over 1 pair 200GBASE-FR1		
400Gbps	100Gbps							Over 4 pairs 400GBASE-DR4-2		
	200Gbps	Over 2 lanes 400GAUI-2	TBD * 400GBASE-KR2	Over 2 pairs 400GBASE-CR2			Over 2 pairs 400GBASE-DR2			
800Gbps	100Gbps	Over 8 lanes 800GAUI-8	Over 8 lanes 800GBASE-KR8	Over 8 pairs 800GBASE-CR8	Over 8 pairs 800GBASE-VR8	Over 8 pairs 800GBASE-SR8	Over 8 pairs 800GBASE-DR8	Over 8 pairs 800GBASE-DR8-2		
	200Gbps	Over 4 lanes 800GAUI-4	TBD * 800GBASE-KR4	Over 4 pairs 800GBASE-CR4			Over 4 pairs 800GBASE-DR4	Over 4 pairs 800GBASE-DR4-2 Over 4 lambdas 800GBASE-FR4	TBD	
	TBD	The maximum number of AUI and PMD lanes is 8 for 800GbE and 16 for 1.6TbE.							Over single SMF in each direction ?	Over single SMF in each direction ?
?1.6Tbps	100Gbps	Over 16 lanes 1.6TAUI-16								
	200Gbps	Over 8 lanes 1.6TAUI-8		Over 8 pairs 1.6TGBASE-CR8			Over 8 pairs 1.6TBASE-DR8	Over 8 pairs 1.6TBASE-DR8-2		

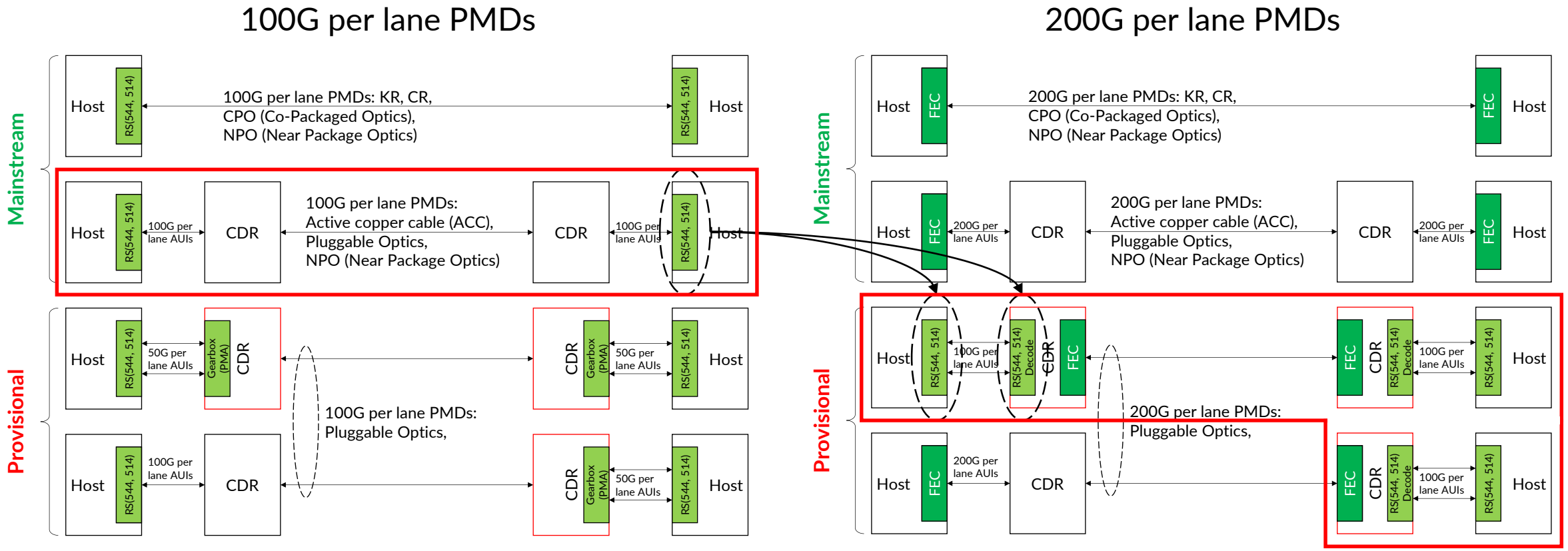
https://www.ieee802.org/3/df/proj_doc/objectives_P802d3df_220317.pdf

* Should be adopted as long as the signaling & modulation & insertion loss objectives for CR/KR channels are determined.

Competitiveness is the key objective for new standard development, i.e. lower power consumption “pJ/bit”, lower cost “cost/bit”, lower latency and lower frame loss ratio (FLR).

“800GbE/1.6TbE” should be competitive over “2*400GbE/4*400GbE”.

800GbE&1.6TbE based on 100G PMDs have a long life cycle and impact the transaction to the next generation



800GbE & 1.6TbE PCS/PMA/PMD based on 100G/lane.
1.6TbE only has 1.6TAUI-16 objective for test and measurement perspective.
# of interleaved RS(544, 514)? # of FEC lanes? Bit-mux or symbol-mux PMA?
200GbE & 400GbE PCS/PMA/PMD was defined in previous IEEE task forces.
2-way interleaved RS(544, 514), 8 FEC lanes for 200GbE and 16 FEC lanes for 400GbE with bit-mux PMA.

200GbE & 400GbE & 800GbE & 1.6TbE PCS/PMA/PMD based on 200G/lane.
FEC architecture? FEC code selection? Bit-mux or symbol-mux PMA?
100G/lane based 800GbE PCS/PMA impacts the complexity of "gearbox" CDR.

General design rules: Simplify the CDR as much as possible and shift the necessary "complexity" to the host ASIC (lu_3df_01_220518).

Requirements for 100G/lane 800GbE and 1.6TbE

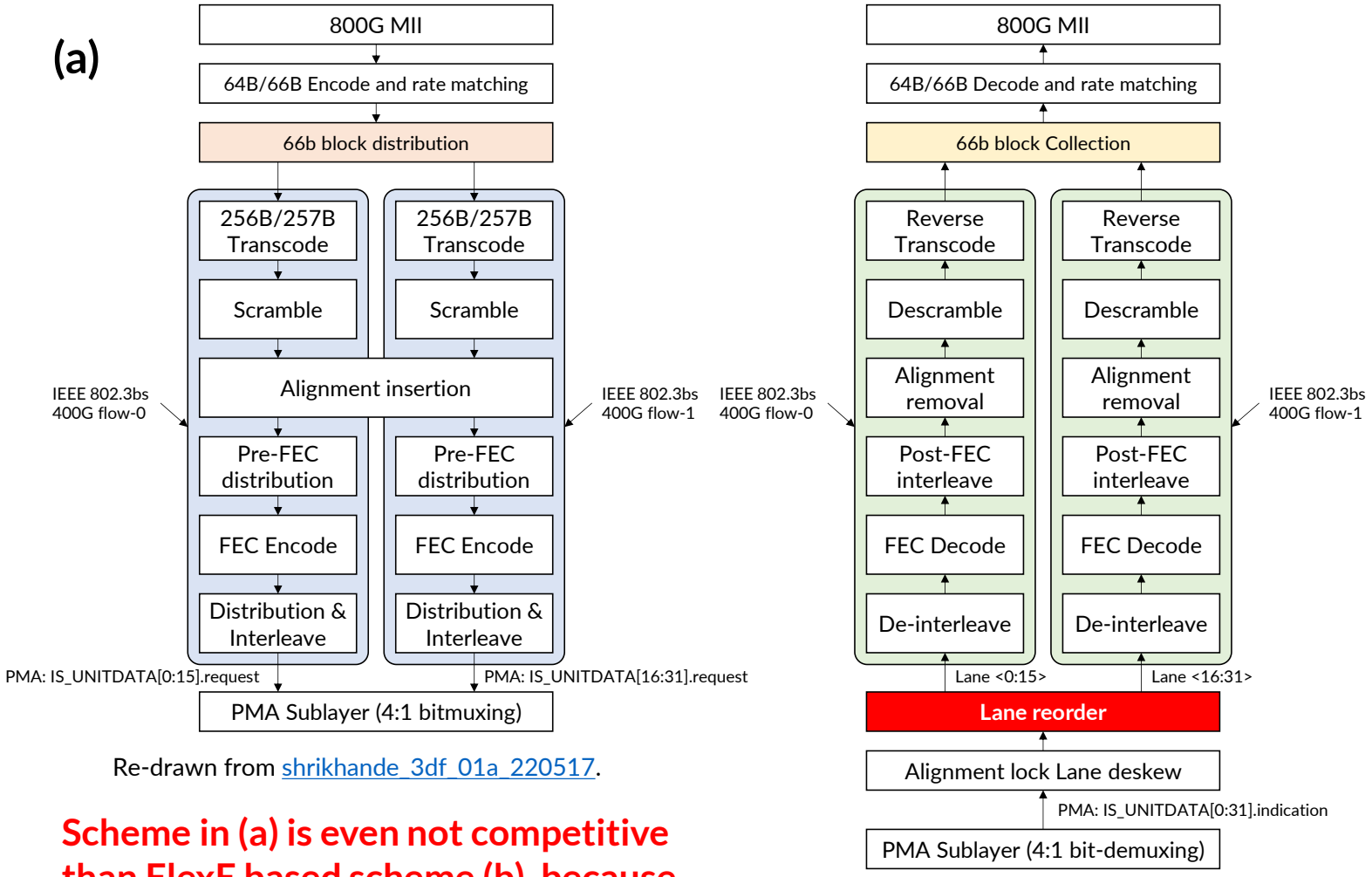
- 800GbE (1.6TbE) should be competitive compared with 2*400GbE (4*400GbE)
  - lower power consumption (“pJ/bit”), lower area (“cost/bit”), lower latency (“ns”)
  - lower frame loss ratio (FLR) or higher margin.
- Achieve a low cost transit from 100G/lane to 200G/lane & ZR “800GbE&1.6TbE”.
  - Simplify the CDR chip (Extender Sublayer /Inverse FEC) as much as possible
    - Use as small number of FEC lanes as possible.
- Implementation and editorial consideration, i.e. reuse of logic blocks and clauses.
  - “200GbE&400GbE” can re-use and benefit from the new design of “800GbE&1.6TbE” logic blocks but not vice versa, because native “800GbE&1.6TbE” design is expected to be much better than “200GbE&400GbE” in all aspects.
  - Reuse the IEEE 802.3bs “200GbE&400GbE” clauses as much as possible.
  - 800GbE can reuse 1.6TbE logic blocks.
- **“2*400GbE bonding” is much less competitive and not recommended**, it does not offer any improvements and is uncompetitive in almost all aspects. It deviates from the original intention of a new Ethernet standard development with higher rate. It is not a native Ethernet speed upgrade technology.

Goals for 100G/lane 800GbE and 1.6TbE

- Build competitive 100G/lane based “800GbE&1.6TbE” over 400GbE.
  - lower power consumption (“pJ/bit”)
  - lower area (“cost/bit”)
  - lower latency (“ns”)
  - lower frame loss ratio (FLR) or higher margin.
- Achieve a low cost transit from 100G/lane to 200G/lane “800GbE&1.6TbE”.
  - Simplify the CDR chip (Extender Sublayer /Inverse FEC) as much as possible.
    - Use as small number of FEC lanes as possible.
- Fast time to an 100G/lane based 800GbE&1.6TbE PCS/FEC/PMA specification.
  - **Fully re-use the “200GbE&400GbE” clauses without modification.**
- Leverage existing industry investment in “200GbE & 400GbE” technology.
  - **No change to the architecture and the clauses**.
  - **Fully re-use the design and validation efforts of “200GbE & 400GbE”.**
  - **In some specific designs, even RTL code can be re-used by “speed-up” with advanced processes.**

“2*400GbE bonding” is not even as competitive as FlexE

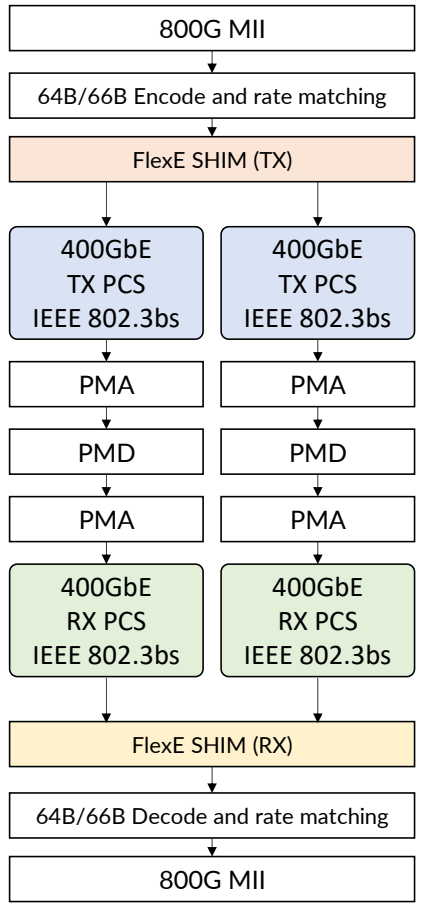
(a)



Re-drawn from shrikhande_3df_01a_220517.

Scheme in (a) is even not competitive than FlexE based scheme (b), because (a) needs reorder over 32 FEC lanes.

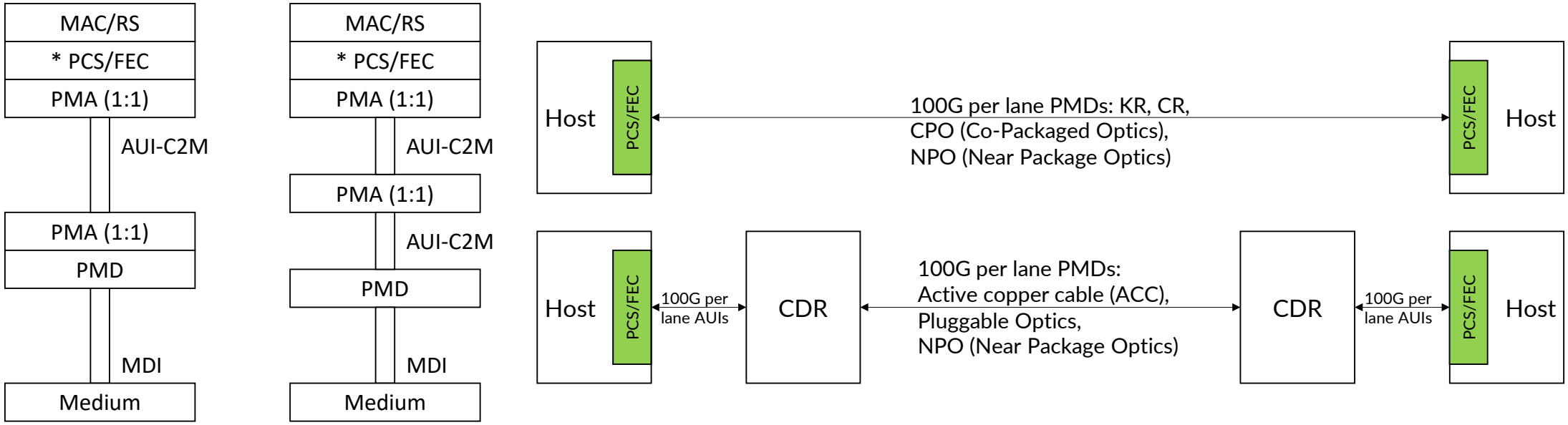
(b)



Bonding of Multiple Links by FlexE.

Re-drawn from figure 7 of https://www.oiforum.com/wp-content/uploads/2019/01/OIF_FlexE_White_Paper.pdf

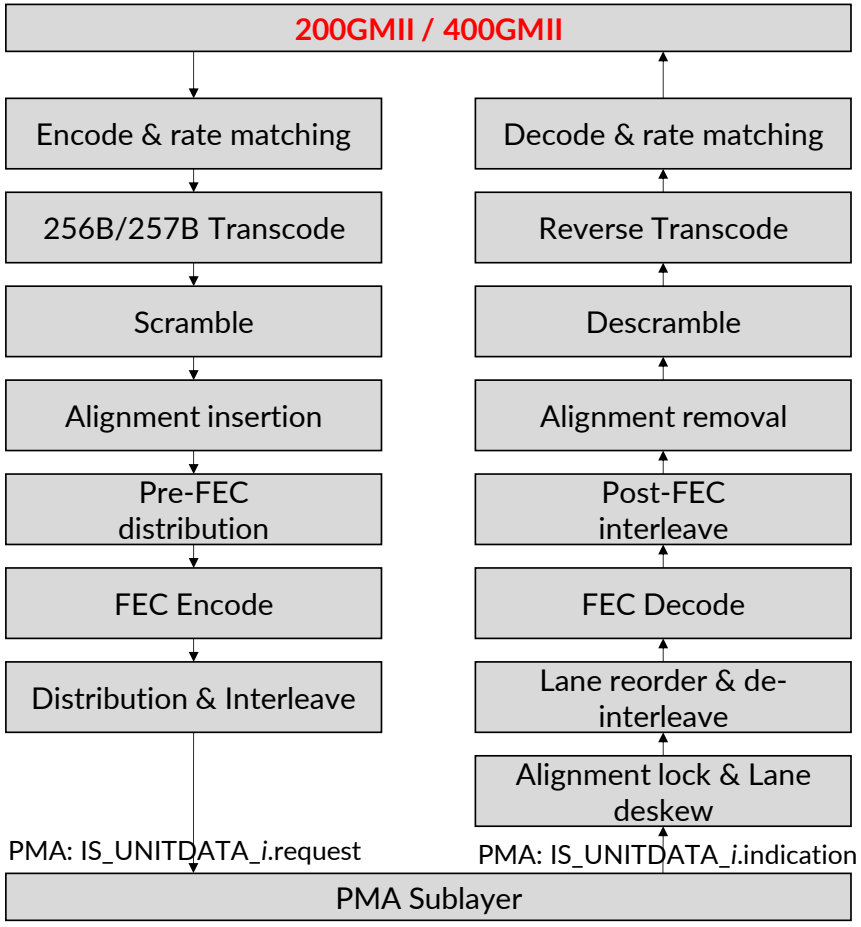
Architecture



* PCS and FEC functions proposed to be co-located in the PCS sub-layer (same as CL119).

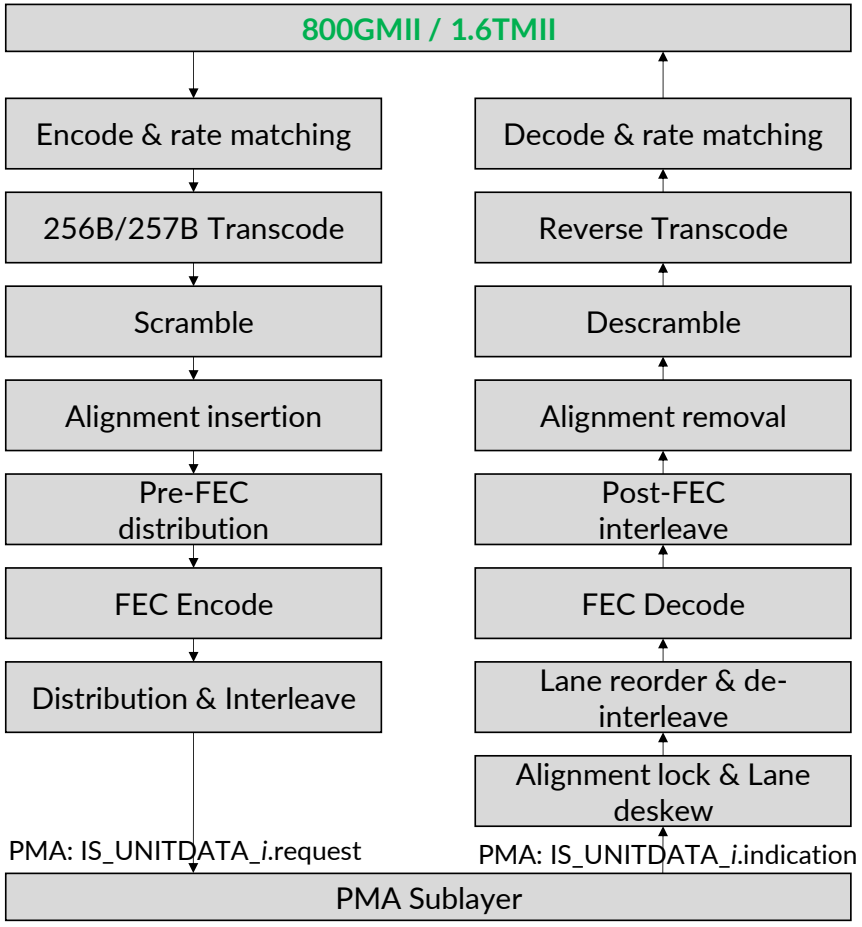
- End-to-end FEC architecture to cover both the AUIs and PMDs.
- 8 FEC lanes for 800GbE and 16 FEC lanes for 1.6TbE which covers all the scenario of the IEEE 802.3df objective.
- It was discussion in bruckman_3df_01_220308.

Function block diagram



(*i*=0 to 7 for 200GBASE-R or *i*=0 to 15 for 400GBASE-R)

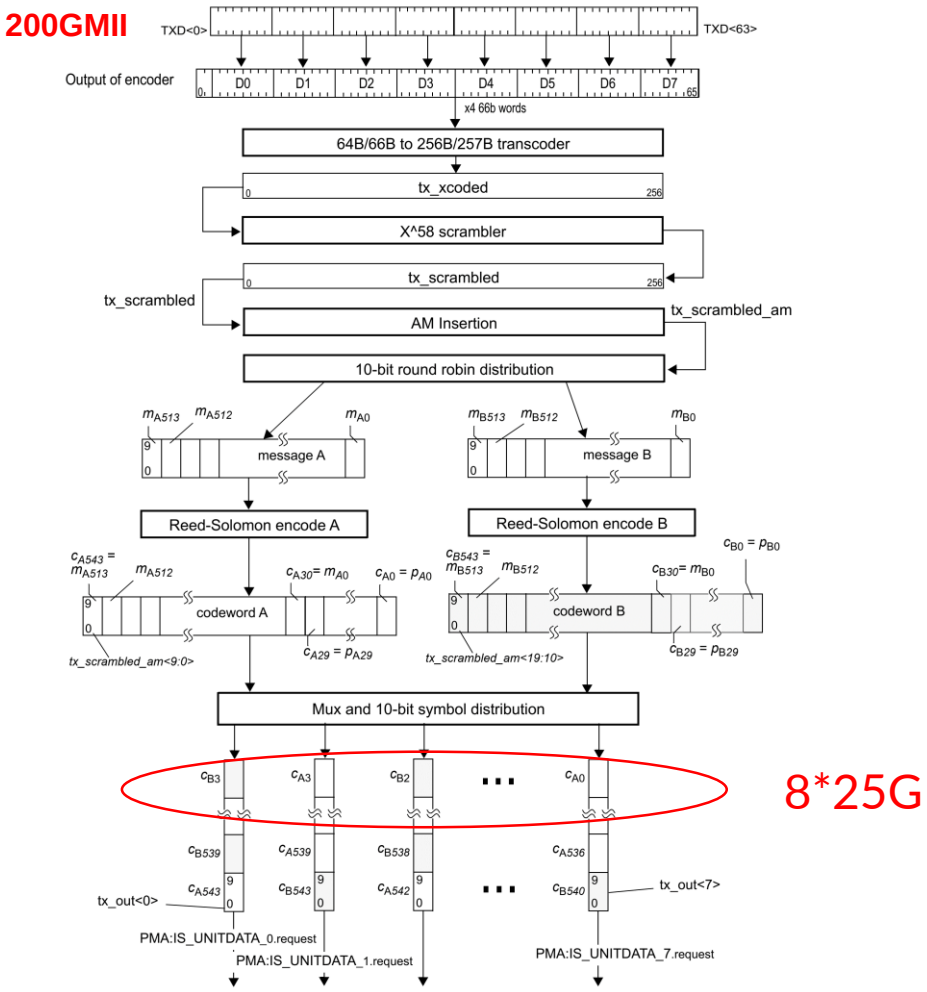
Clause 119



(*i*=0 to 7 for 800GBASE-R or *i*=0 to 15 for 1.6TBASE-R)

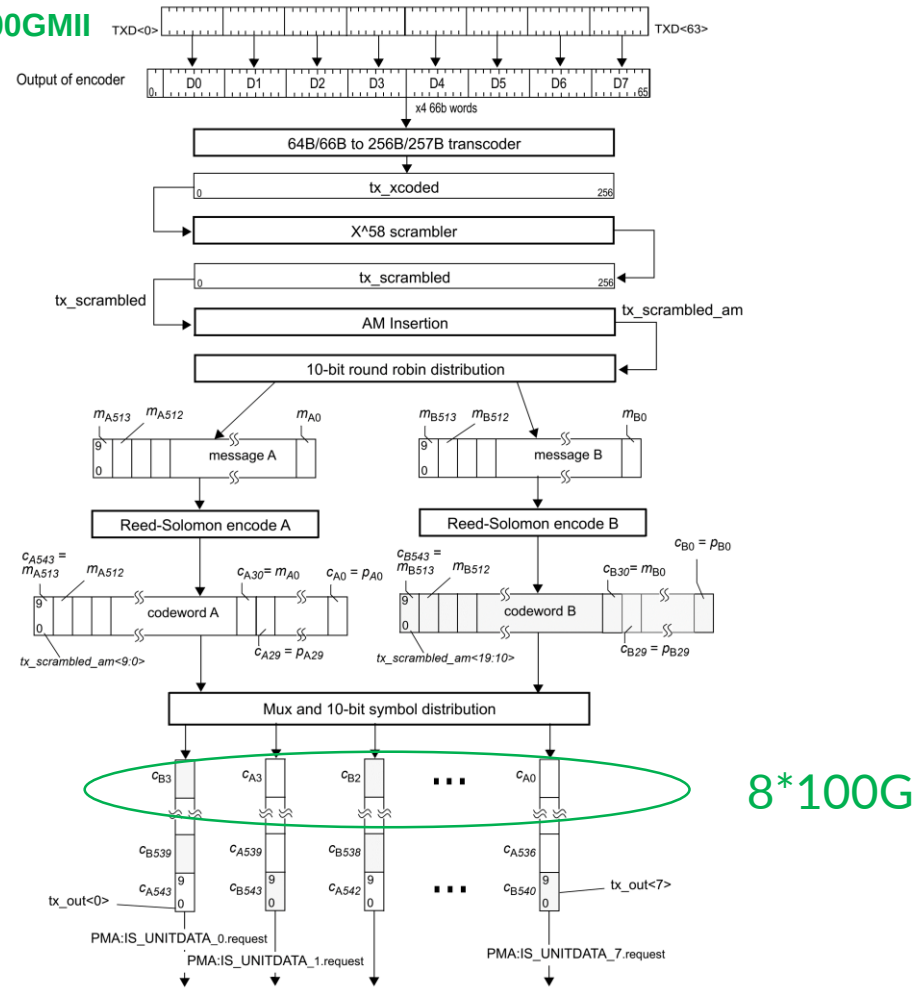
Proposed scheme

Transmit bit ordering and distribution for 800GbE



8*25G

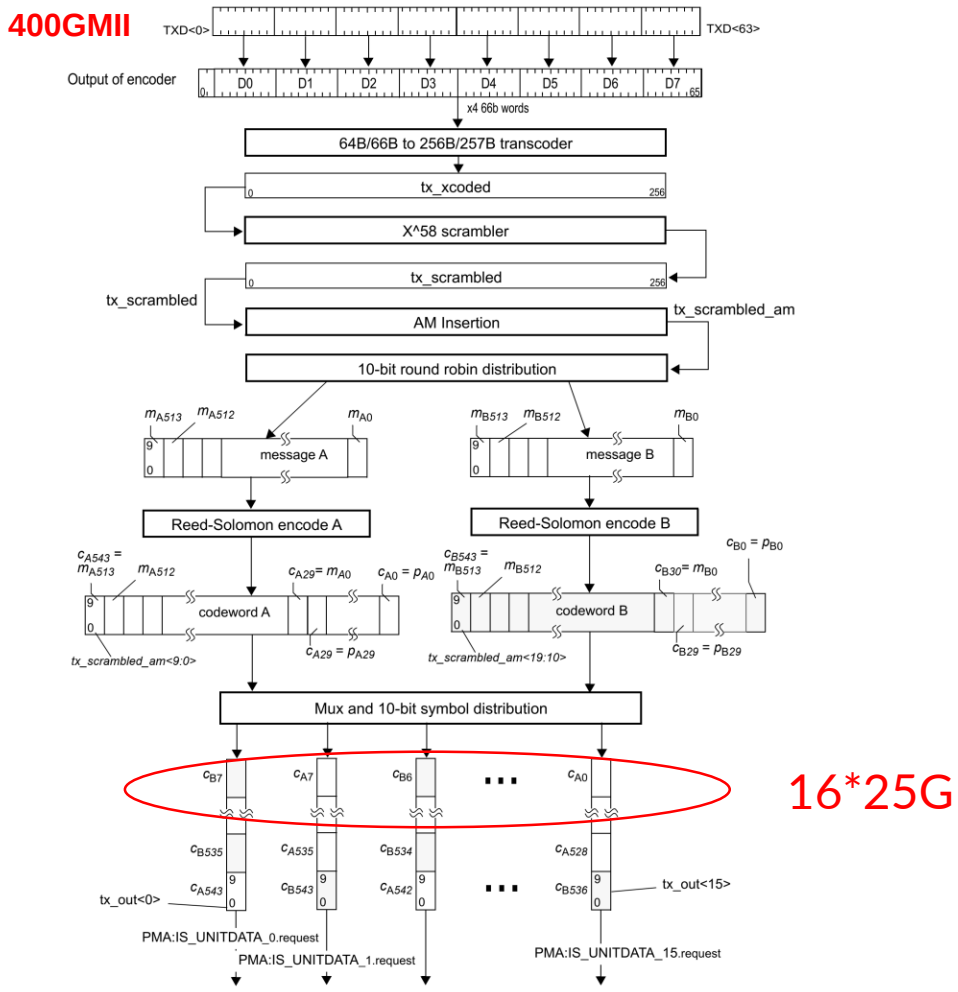
Clause 119



8*100G

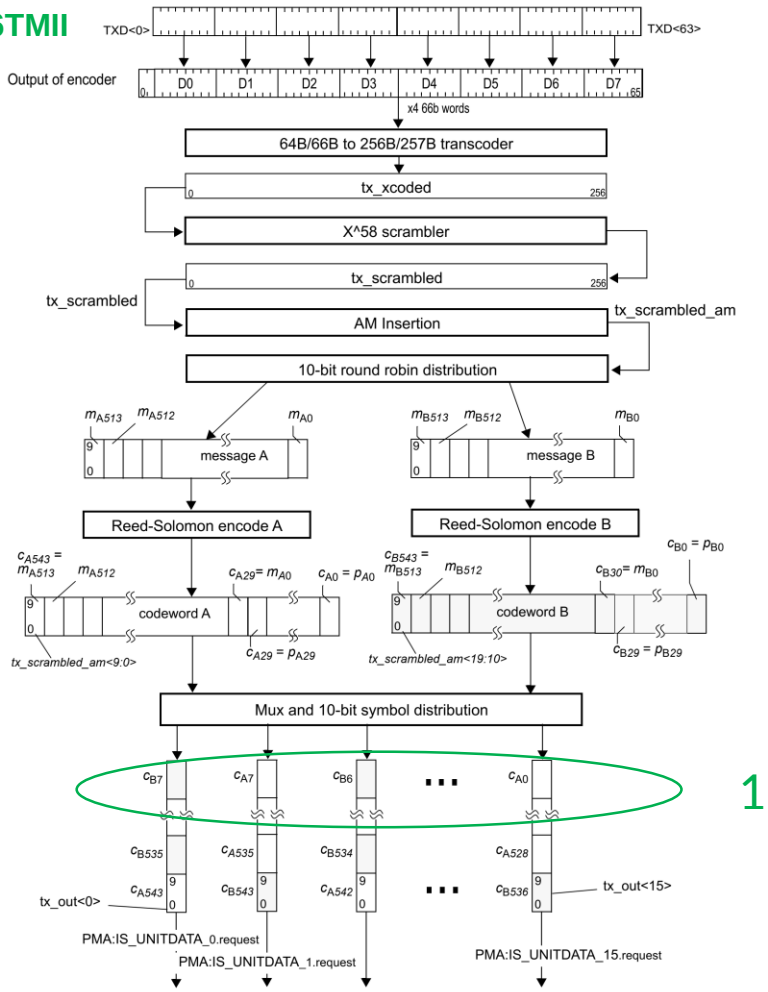
Proposed scheme

Transmit bit ordering and distribution for 1.6TbE



16*25G

Clause 119



16*100G

Proposed scheme

Summary

- A PCS, FEC and PMA baseline is proposed for 800GbE and 1.6TbE using 100G PMD lanes.
  - This baseline proposal is superior in all the aspects in terms of power consumption (“pJ/bit”), area (“cost/bit”), latency (“ns”) frame loss ratio (FLR) compared with “PHY bonding” solutions.
- Supports all adopted 802.3df copper and optical PMDs baselines of 100G/lane.
- Highly leverages existing IEEE802.3bs specifications.
  - 200GbE & 400GbE clause 119 without modifications, only “speed-up” is required.
- “200GbE & 400GbE” can highly re-use the optimized “800GbE and 1.6TbE” and gain benefits.
  - 200GbE & 400GbE can be implemented with “800GbE and 1.6TbE” logic by using time division multiplexing (TDM).
  - Architectural benefits such as low latency, low power consumption and low cost are achievable for combo IPs.
  - Fully re-use the design and validation efforts of “200GbE & 400GbE”.
  - In some specific designs, even RTL code can be re-used by “speed-up” with advanced processes.
- Simplify the extender sublayer as much as possible to better fit into the CDR chips and support schemes using 200G/lane AUIs and PMDs and/or Coherent/ZR PMDs.
- 800GbE & 1.6TbE PCS/FEC can fully share logic, and also the clause.

Q & A