200GbE Logic Baseline proposal

IEEE P802.3bs 200 Gb/s & 400 Gb/s Ethernet Task Force

May 2016 Whistler

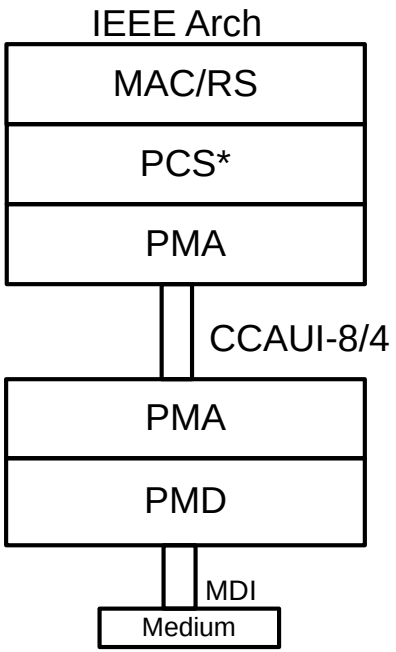
Mark Gustlin – Xilinx
Pete Anslow - Ciena
Eric Baden - Broadcom
Gary Nicholl – Cisco
Dave Ofelt – Juniper
Steve Trowbridge - Nokia
Tongtong Wang – Huawei
John D'Ambrosia- Futurewei, Subsidiary of Huawei

Introduction

- This is a baseline proposal for the 200GbE logic layers
- The following assumes reusing the 802.3bs architecture, and that FEC is always required
- Supports 8/4 lanes (25G and 50G)

PCS Architecture

- Based on the current 802.3bs system architecture
- End to end FEC is used, across up to 5 interfaces (located in the PCS sublayer)
- CCMII is an optional interface that is not shown in these figures, but may be present in a given implementation



*FEC is part of the PCS sublayer

CCMII Interface

- Why define it?
  - Electrically it won't be directly instantiated, but in the proposed 200GbE architecture it can be extended with an extender sublayer (CCXS) and interface (CCAUI-n)
  - Some will want it for RTL to RTL connections within devices
- Define it as a logical Interface only
  - Unless it is extended, then there is a physical instantiation via an extender sublayer

What is it?

- Base it directly on clause 81
- Same signal structure as shown below, just run faster, or in parallel
  - Same as 400GbE

81.1.6 XLGMII/CGMII structure

The XLGMII/CGMII is composed of independent transmit and receive paths. Each direction uses 64 data signals (TXD<63:0> and RXD<63:0>), 8 control signals (TXC<7:0> and RXC<7:0>), and a clock (TX_CLK and RX_CLK). Figure 81–2 depicts a schematic view of the RS inputs and outputs.

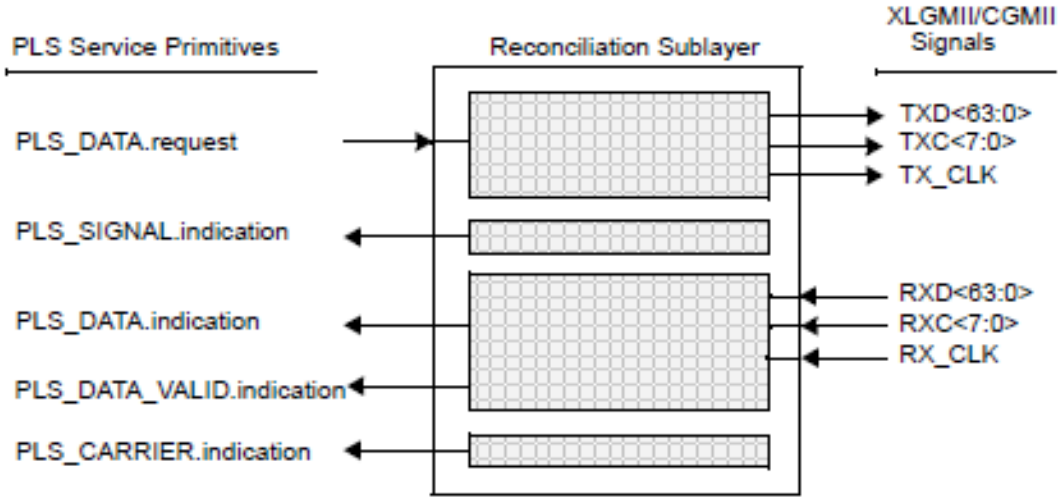
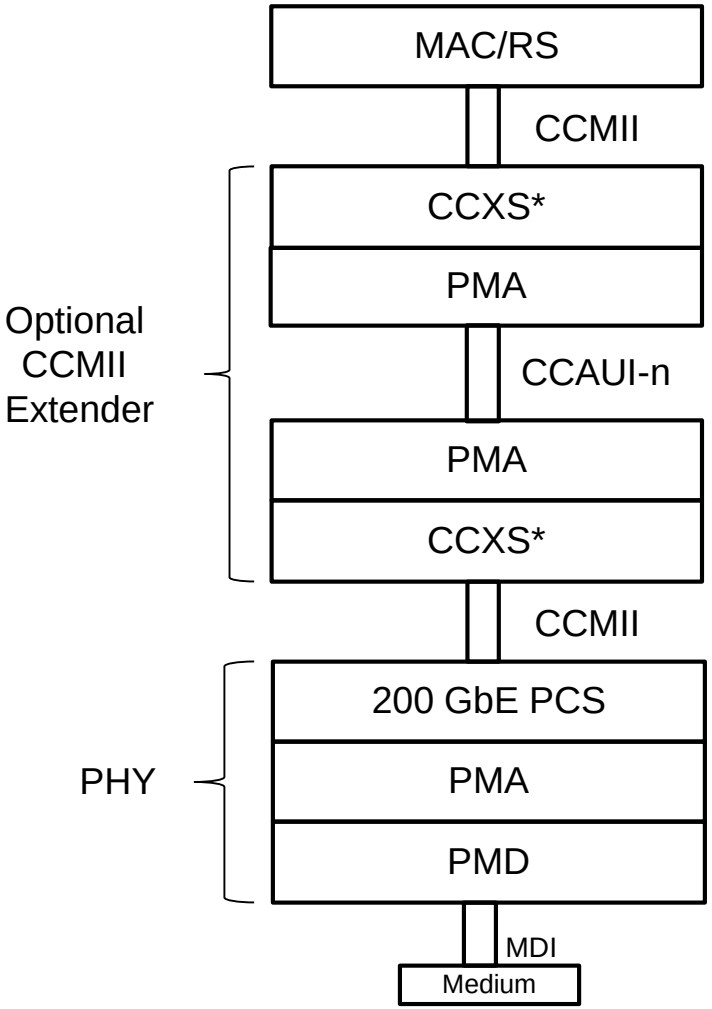


Figure 81–2—Reconciliation Sublayer (RS) inputs and outputs

Extender Sublayer (CCXS)

- The CCXS is the proposed extender sublayer to extend the CCMII
- It is optional, only used if the PCS does not cover both the electrical and optical interface needs
- The CCXS contains PCS and FEC, functionality

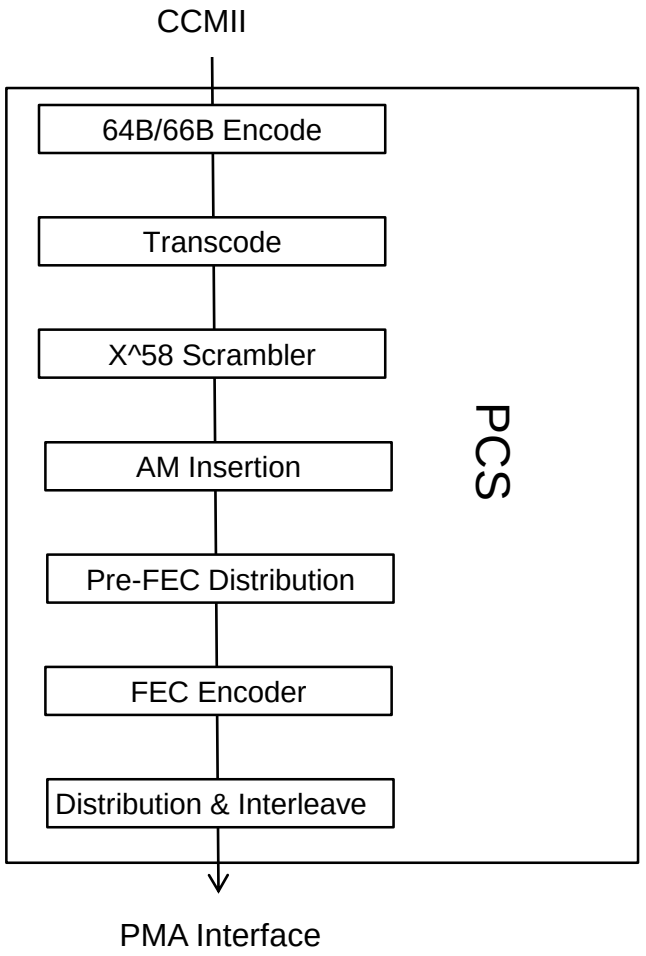
CCMII Extender (Optional)



*** Note - Same as 200 GbE PCS**

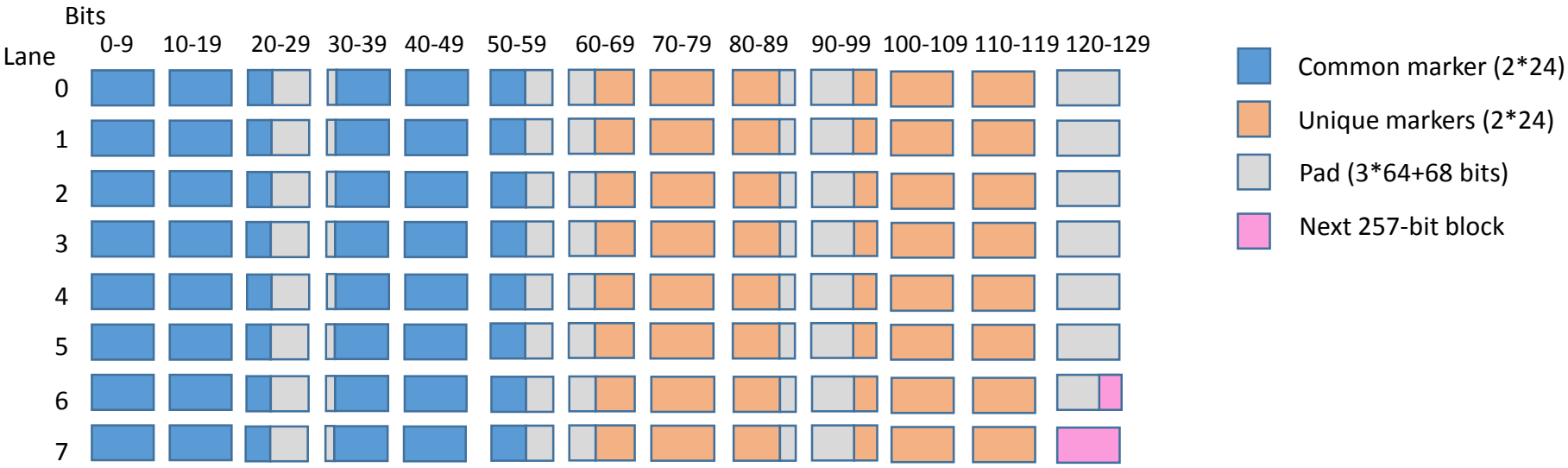
Proposed TX PCS Data Flow

- 64B/66B encode based on clause 82/119
- Transcode to 256B/257B based on clause 119
- Scrambler is located after the Transcoding to simplify the flow, standard X^58 scrambler
- Alignment Markers are the same format as clause 119
- FEC Encoder is RS(544,514,10)
  - Proposed that all FEC processing is as in clause 119, including data distribution and interleaving
- Support for any logical lane on any physical lane

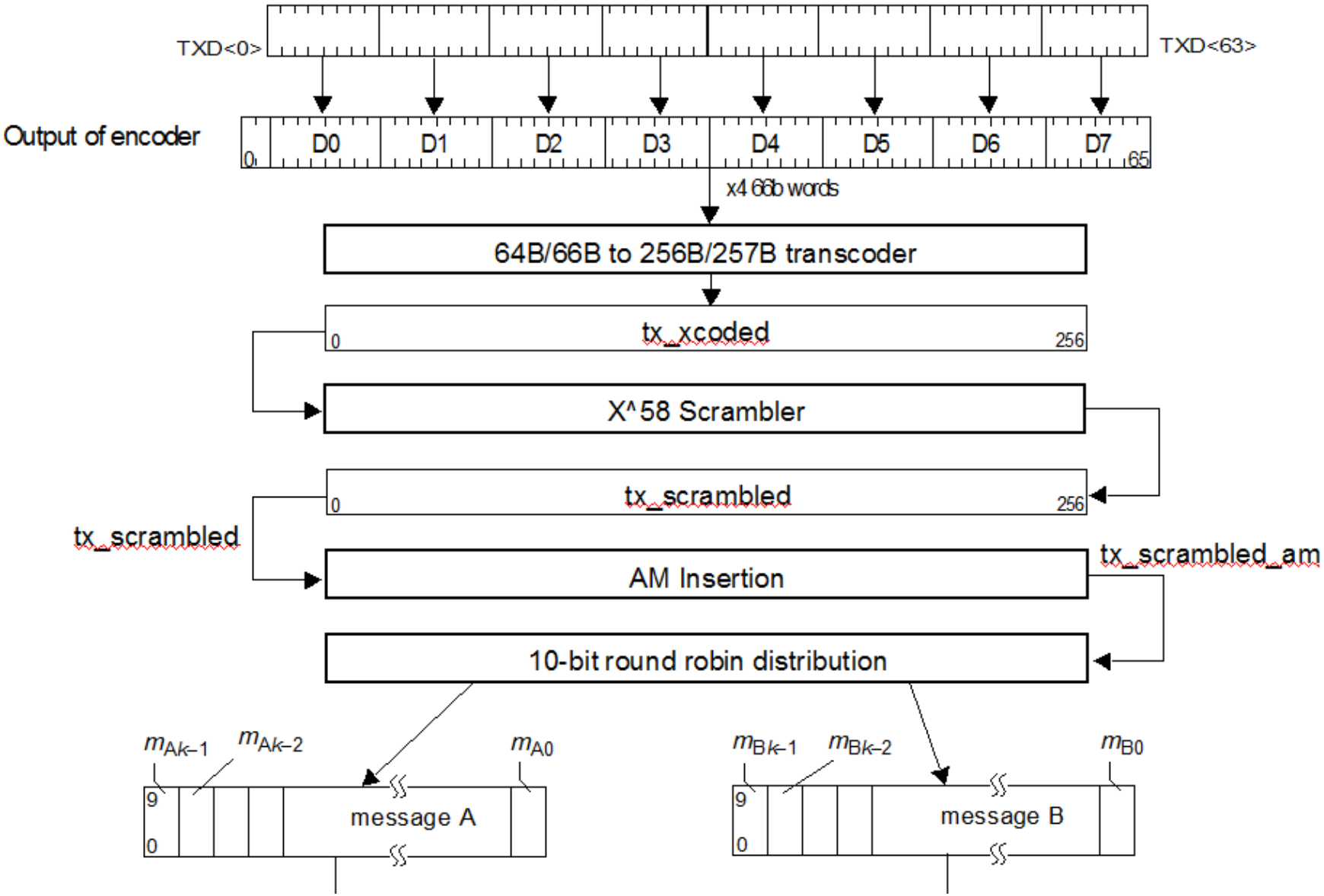


AM Details

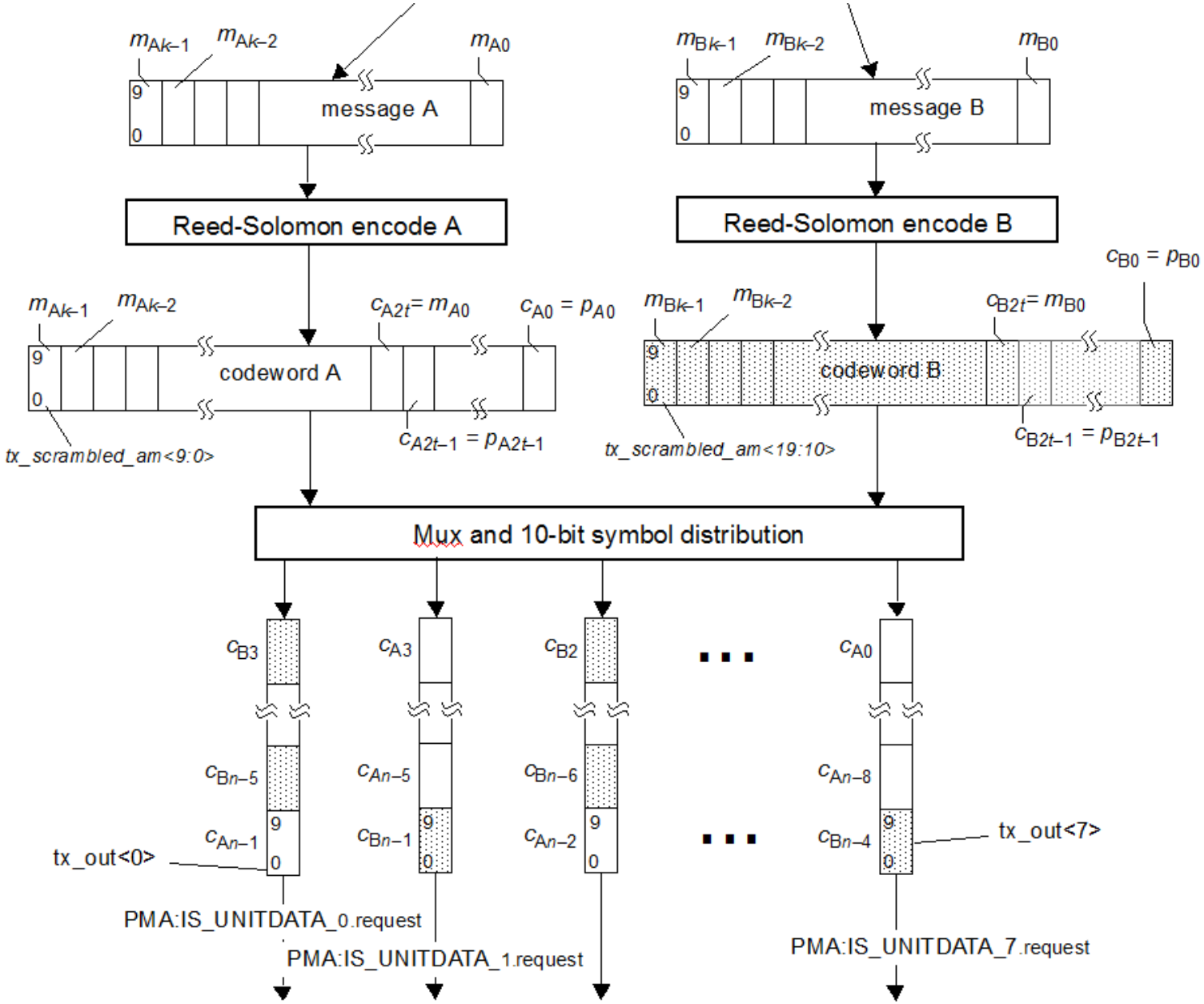
- Similar format to 400GbE
  - 120b AM field per lane with CM0 to CM5 and UM0 to UM5
  - For now use the first 8 AMs from 400GbE
- Distance is ½ that of 400GbE
  - 81920 × 257-bit blocks between AM insertions



200GbE Data Distribution Cont



Distribution Cont



PMA Functions

- Identical PMA functions as described in Clause 120
- Support for bit muxing and any logical lane to any physical lane
- With KP4 FEC the per lane signaling rate is:
  - 544/514*257/256*25G = 26.5625G
  - When running 8 lanes
  - When running 4 lanes it is 53.125G per lane

Misc Stuff

- All skew and delay budgets are identical to 400GbE
  - Delay is the same in time units (ns), number of bits is ½

EEE

- Support fast wake only, same as 400GbE

What should be done for 802.3bs?

- Adopt Fast Wake mode of operation for the 802.3bs PHY types.
- Add these PHY types to "Table 78-1 Clauses associated with each PHY or interface type" and indicate that they do not support deep sleep with the "b" suffix.
- The CDMII will need to be able to signal LPI and the RS will need to include a transmit LPI state machine to defer transmission for the wake time period after de-assertion of LPI.
- The PCS will need to be able to encode and decode LPI.

From: marris_3bs_01_0115.pdf

Conclusion

- This presentation looks at a baseline PCS/PMA for the 200GbE architecture
- This architecture is feasible, it follows 802.3bs architecture which has been shown to be technically feasible
- Achievable latency is ~110ns with similar performance/gain as 400GbE

Thanks!