PCS/FEC Baseline proposal for 50GbE and NG 100GbE

Gary Nicholl, Cisco Systems

IEEE 802.3 50GE & NGOATH Study Group ad-hoc, May 11, 2016

Supporters

- David Ofelt, Juniper
- TBA

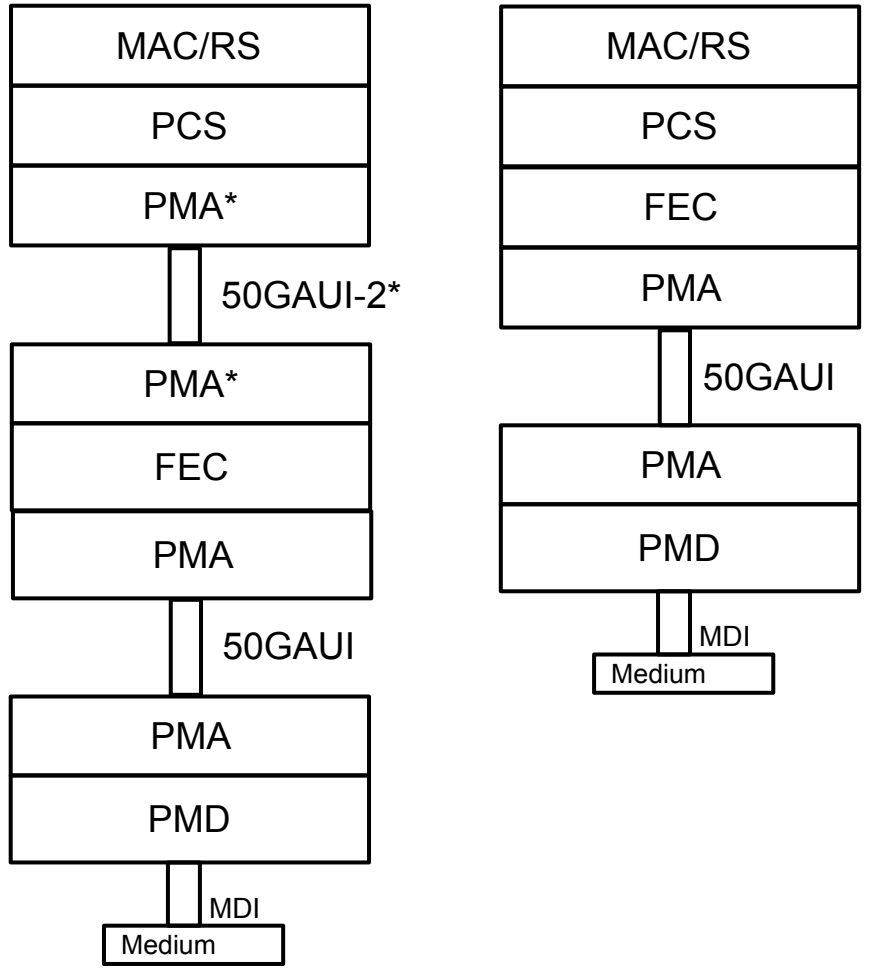
Introduction

- This presentation is developing a baseline proposal for the PCS and FEC for 50GbE and NG 100GbE. However there are still some open issues, e.g.
  - Distribution of FEC to physical lanes
  - Mapping of Alignment Markers (AMs) to FEC lanes
  - Definition of Alignment Marker bit patterns.
- Builds upon the following earlier presentations:
  - nicholl_042716_50GE_NGOATH_adhoc
  - nicholl_041316_50GE_NGOATH_adhoc-v2
- Proposals optimized for 50Gb/s AUI and PMD lane rates

50GbE

50GbE PCS Overview

- Separate PCS & FEC sub layers
  - similar to 100GbE architecture (802.3ba, 802.3bj)
  - PCS and FEC can be physically separated
- 4 lane PCS
  - based on overclocked 40GbE PCS (Clause 82)
  - AM spacing modified to support FEC sublayer
  - architecture supports optional AUI-2 /w no-FEC
- RS (544,514) FEC
  - based on 802.3bj (CL 91) but with single FEC lane
  - FEC can only operate over a single lane interface
  - optimized for 50Gb/s AUI and PMD lane rates
  - no FEC codeword interleaving (latency concerns)

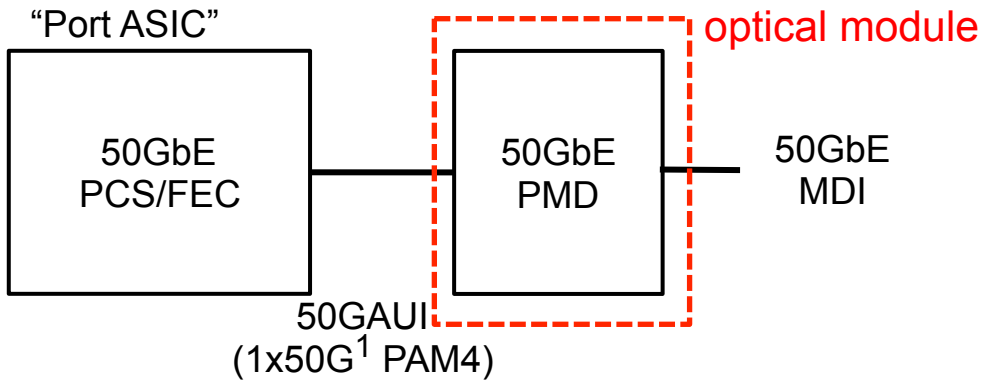


*Optional 50GAUI-2

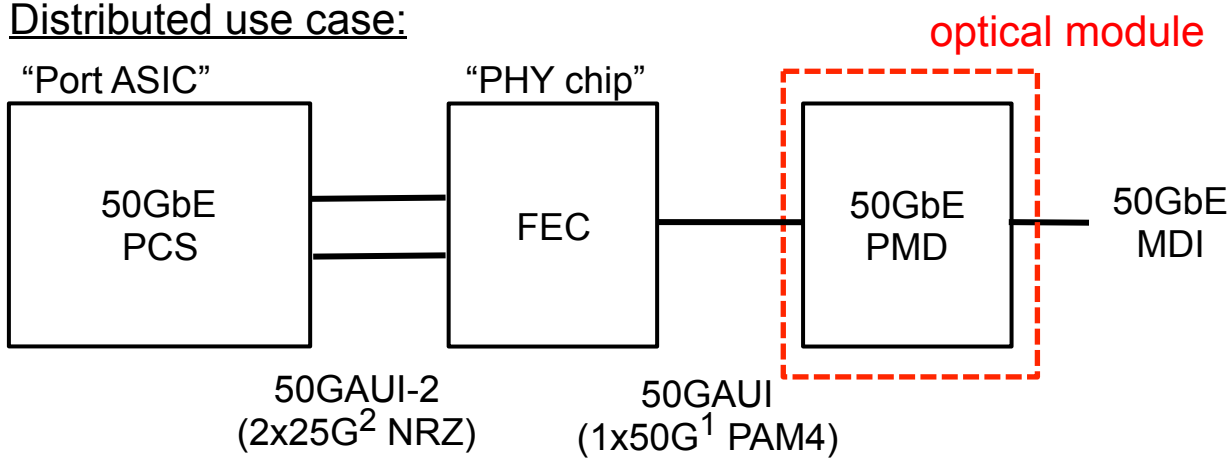
Note: PMA blocks not shown for clarity.

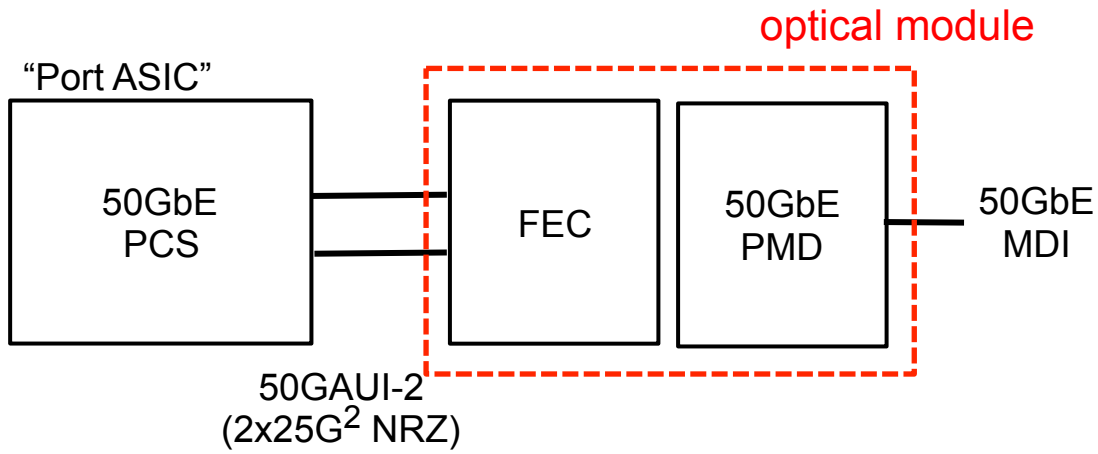
50GbE PCS Use Cases

Integrated use case (long term, single lane optimized):



Distributed use case:

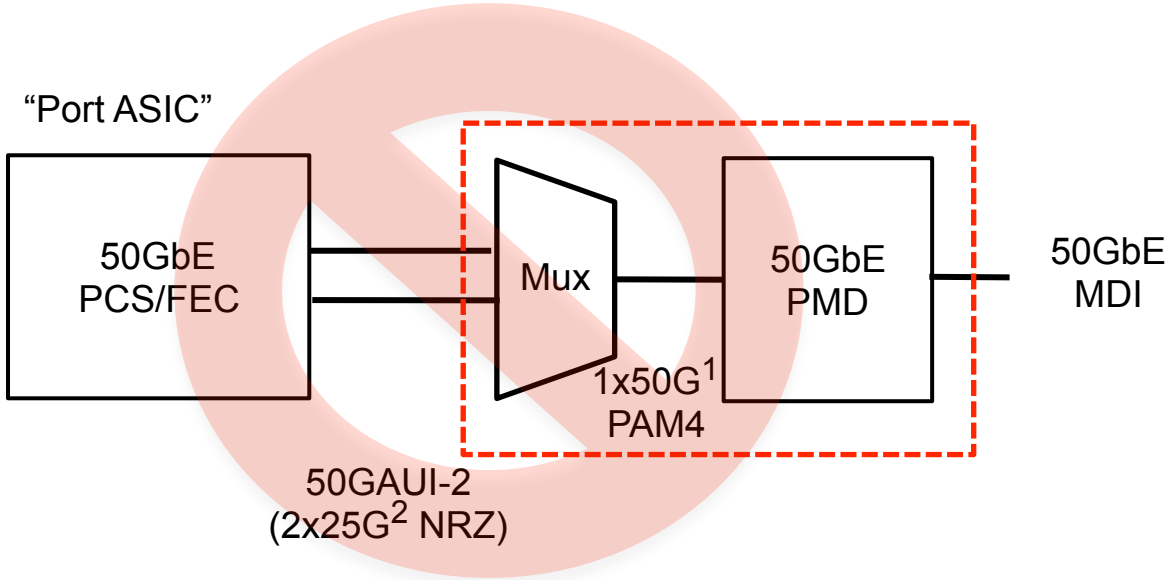




1 = 53.125 Gb/s, 2 = 25.78125Gb/s

Note: PMA blocks not shown for clarity.

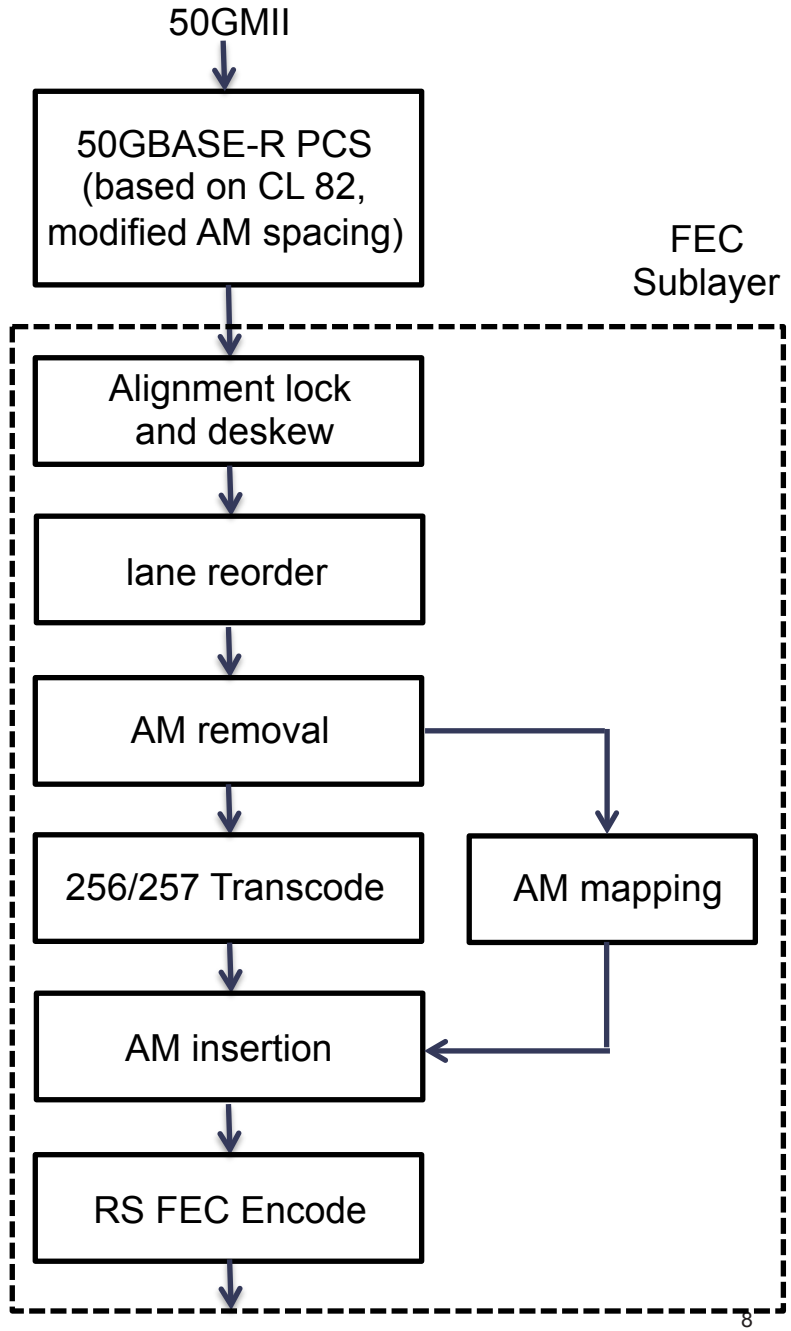
50GbE PCS Use Cases (not supported)



1 = 53.125 Gb/s, 2 = 26.5625Gb/s

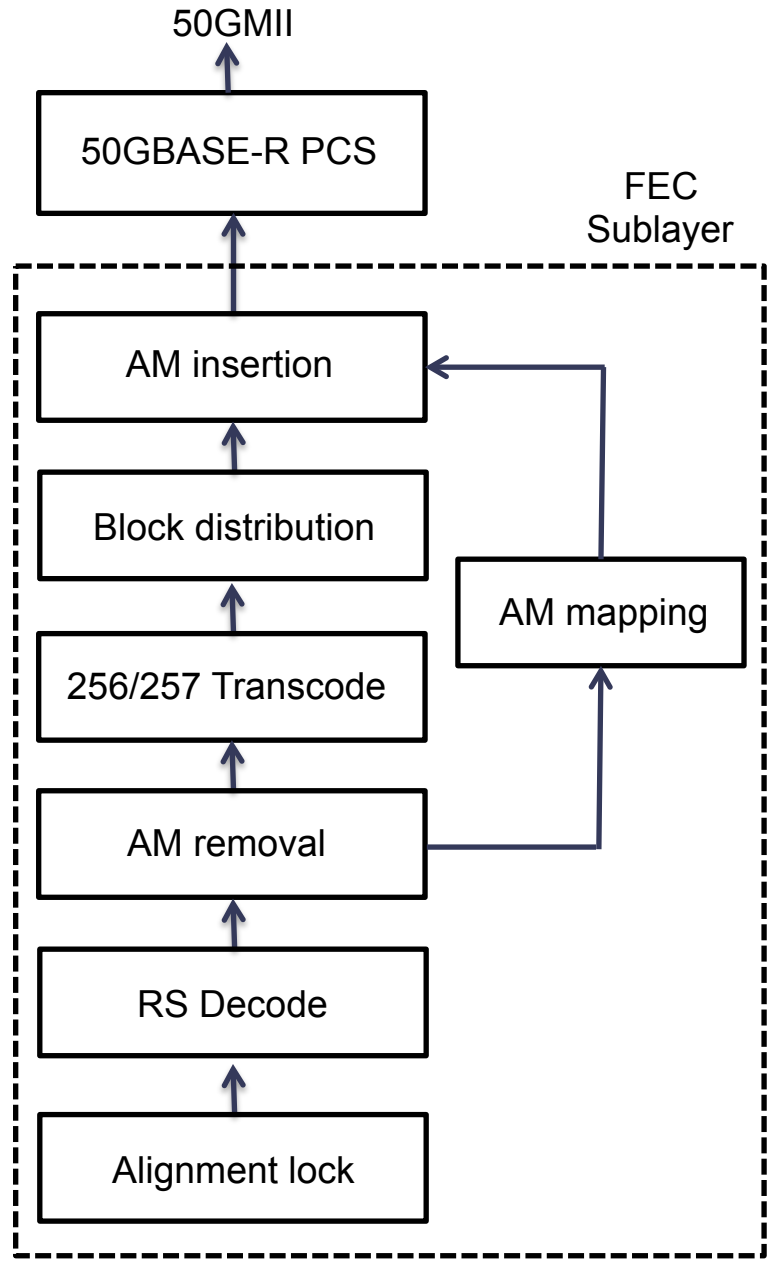
50GbE Tx Data Flow

- PCS based on 802.3ba Clause 82
  - overclocked 40GbE PCS
  - 4 x PCS lanes running at 12.890625 Gb/s
  - 4 x 66-bit alignment markers (AM), one per PCS lane
  - AM spacing (start of one AM to the start of next AM) modified to 20480 66-bit blocks, to align with FEC codeword boundaries
  - standard PMA bit muxing to map the 4 PCS lanes onto 50GAUI-2
- FEC sublayer leverages 802.3bj Clause 91
  - RS (544, 514)
  - Single FEC lane (serialized RS symbols on output)
  - 257-bit alignment marker inserted at beginning of every 1024 FEC codewords.
  - Fairly simply AM mapping (4 x 64-bit PCS AMs + 1 bit pad)

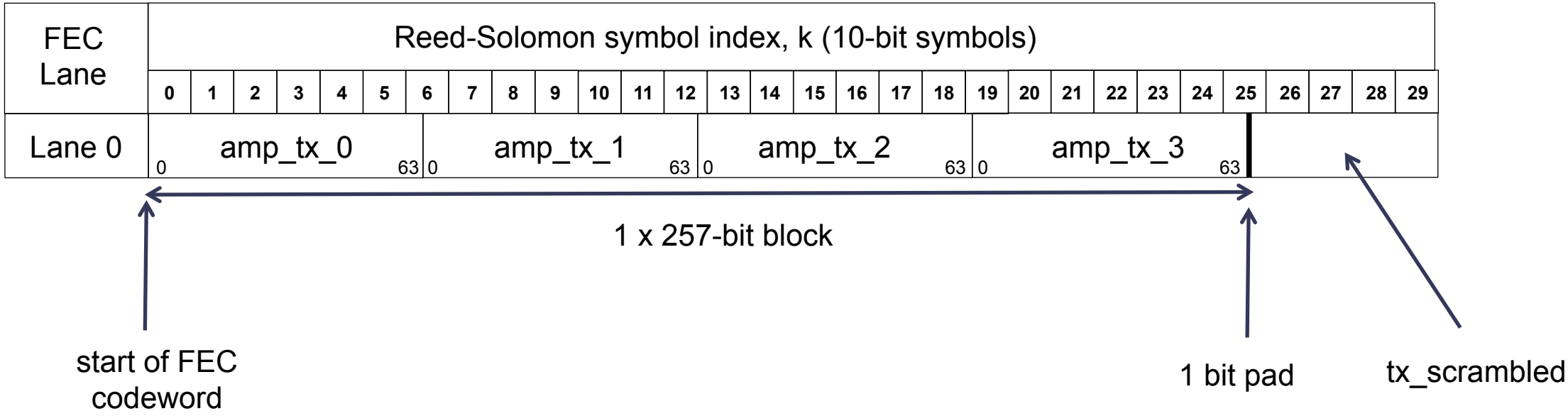


50GbE Rx Data Flow

- Reverse of Tx



50GbE - Alignment marker mapping to FEC lane



50GbE - Open Items

- Definition of PCS Alignment Marker bit patterns
  - how different to 40GbE AMs ?
- FEC distribution over single 1x50G lane (current proposal) or over 2x25G lanes ?

50GbE - Proposed Mapping to IEEE Documentation

- New 50GbE PCS Clause
  - highly reference Clause 82
  - similar to approach taken in 802.3by for 25GbE
- New 50GbE FEC Clause
  - highly reference Clause 91
  - similar to approach taken in 802.3by for 25GbE

50GbE Baseline Summary

Pros:

- supports an optimized single lane architecture (with PCS & FEC in port ASIC)
- supports both 50GAUI (1x50G) and optional 50GAUI-2 (2x25G NRZ) interfaces
  - but for 50GAUI-2 the FEC must be external
- enables easy transition from 25Gb/s to 50Gb/s port ASIC IO
- supports ‘bump in the wire’ applications for server ports

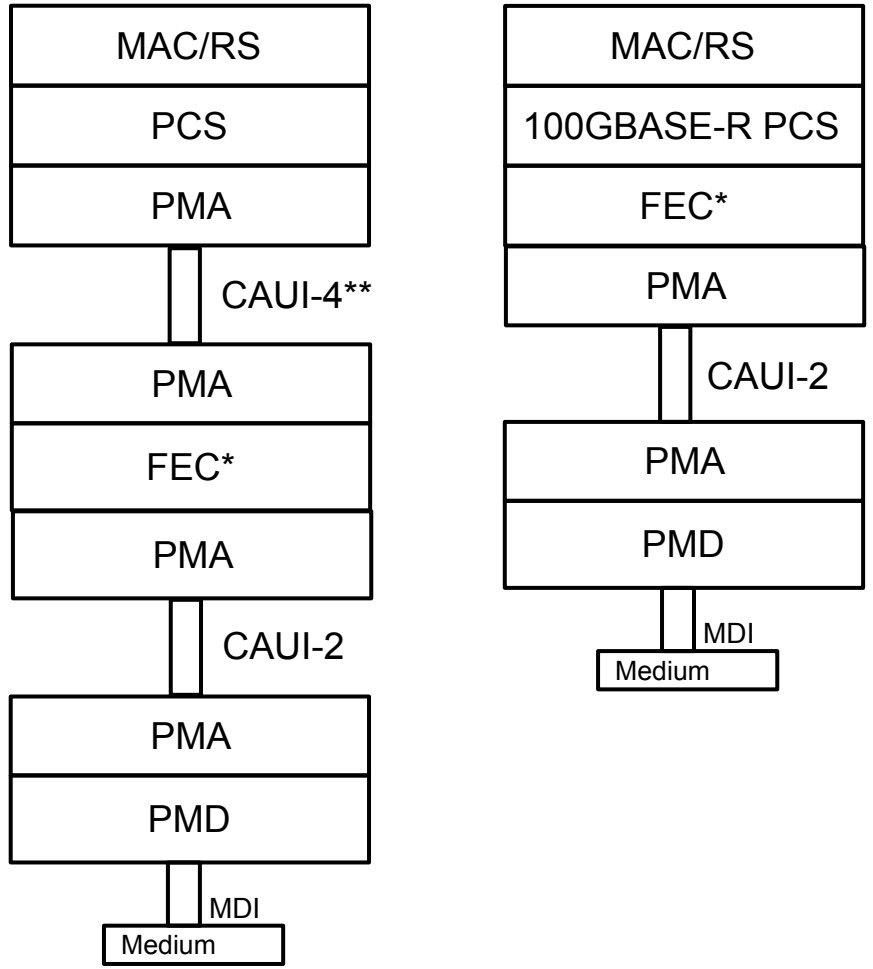
Cons:

- long term the 4 lane MLD functionality in the PCS is redundant
  - trivial impact

100GbE

NG 100GbE PCS Overview

- Separate PCS & FEC sub-layers
  - same as 802.3bj
  - same as proposed 50GbE architecture
- Existing 100GbE (CL82) PCS
  - no changes required
  - supports optional CAUI-4 /w no-FEC
- RS (544,514) FEC
  - based on 802.3bj (CL 91) but distributed over 2 FEC lanes
  - optimized for 50Gb/s AUI and PMD lane rates
  - no FEC codeword interleaving (latency concerns)



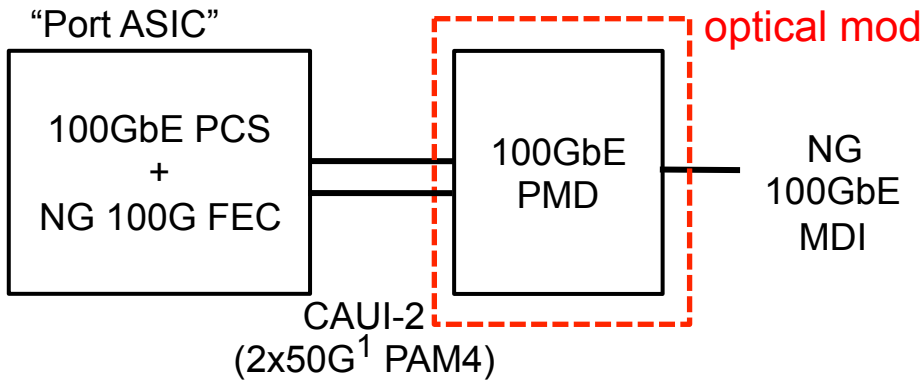
*FEC is a separate sublayer

**optional CAUI-4

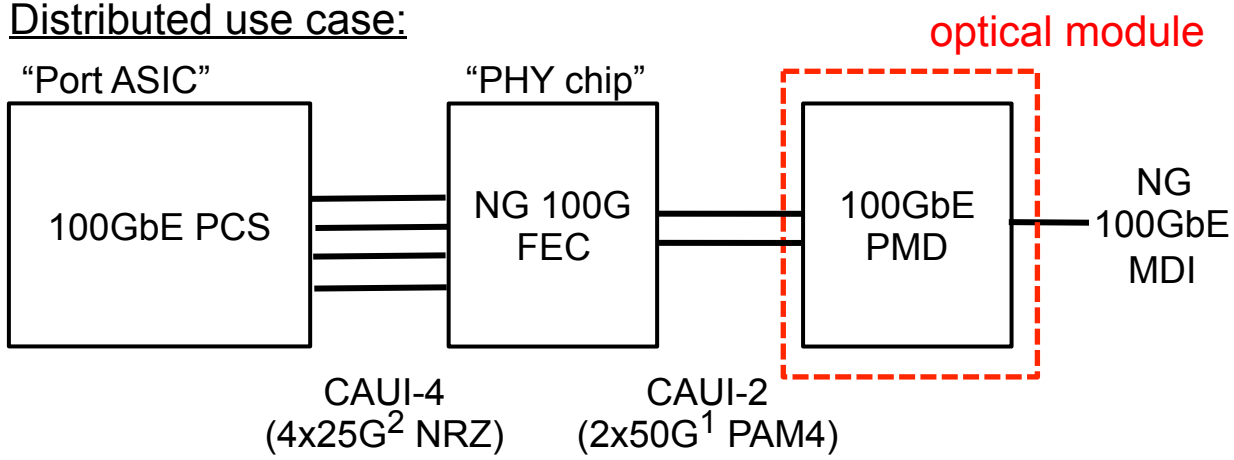
Note: PMA blocks not shown for clarity.

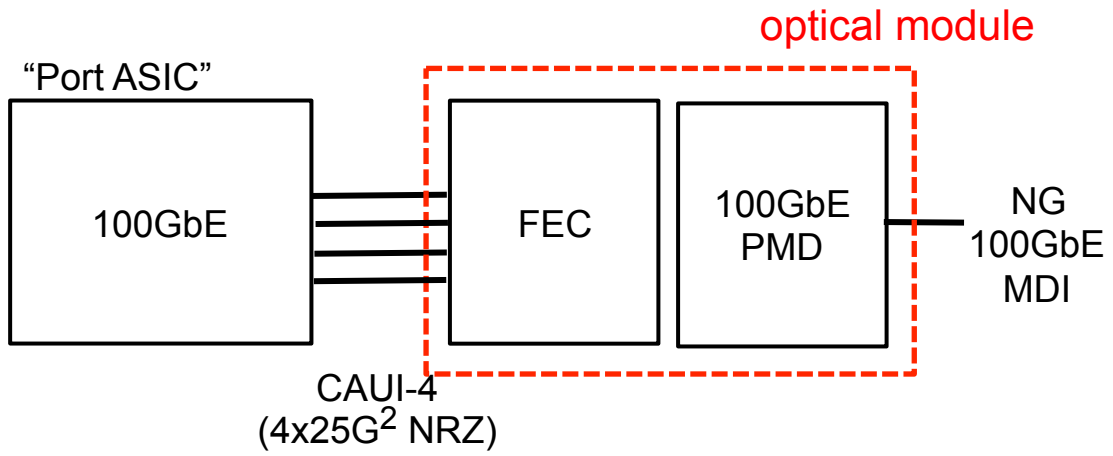
NG 100GbE PCS Use Cases

Integrated use case (long term, 2x50G lane optimized):



Distributed use case:

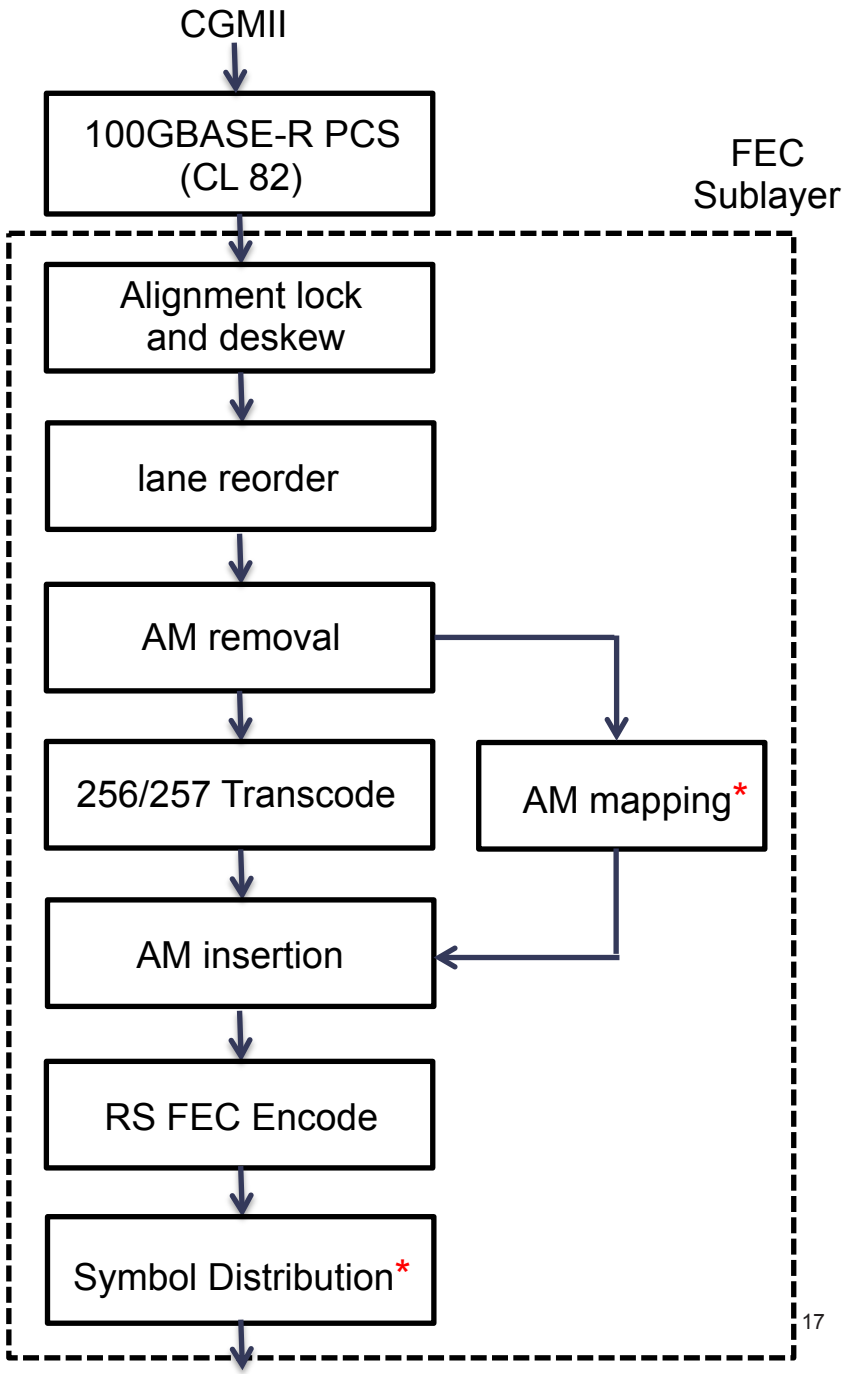




1 = 53.125 Gb/s, 2 = 25.78125Gb/s

NG 100GbE Tx Data Flow

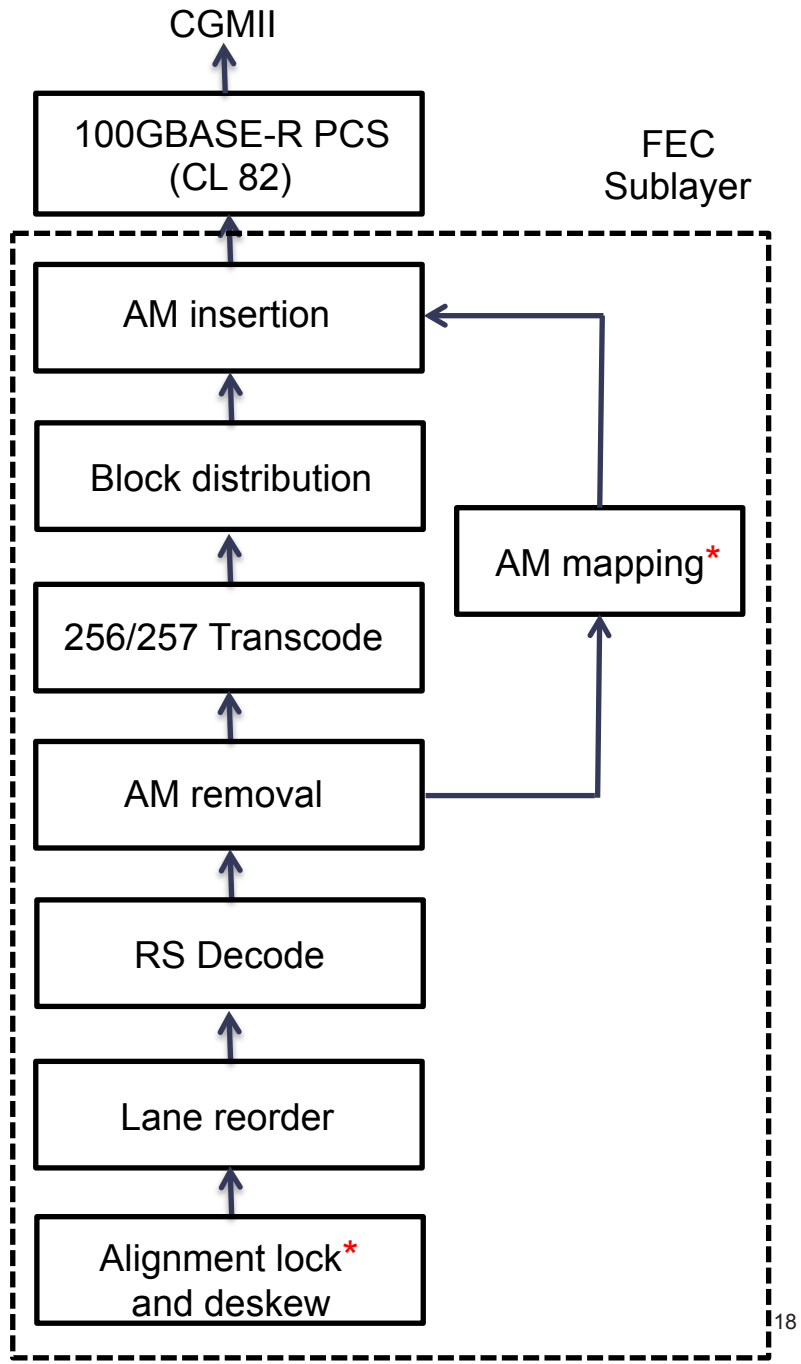
- PCS identical to 802.3ba Clause 82
  - no changes required
- FEC sublayer data flow identical to 802.3bj Clause 91, with following exceptions:
  - FEC symbols distributed over 2 rather than 4 lanes
  - Minor change to AM mapping to accommodate fact that distributing over 2 rather than 4 FEC lanes



* = minor difference to Clause 91

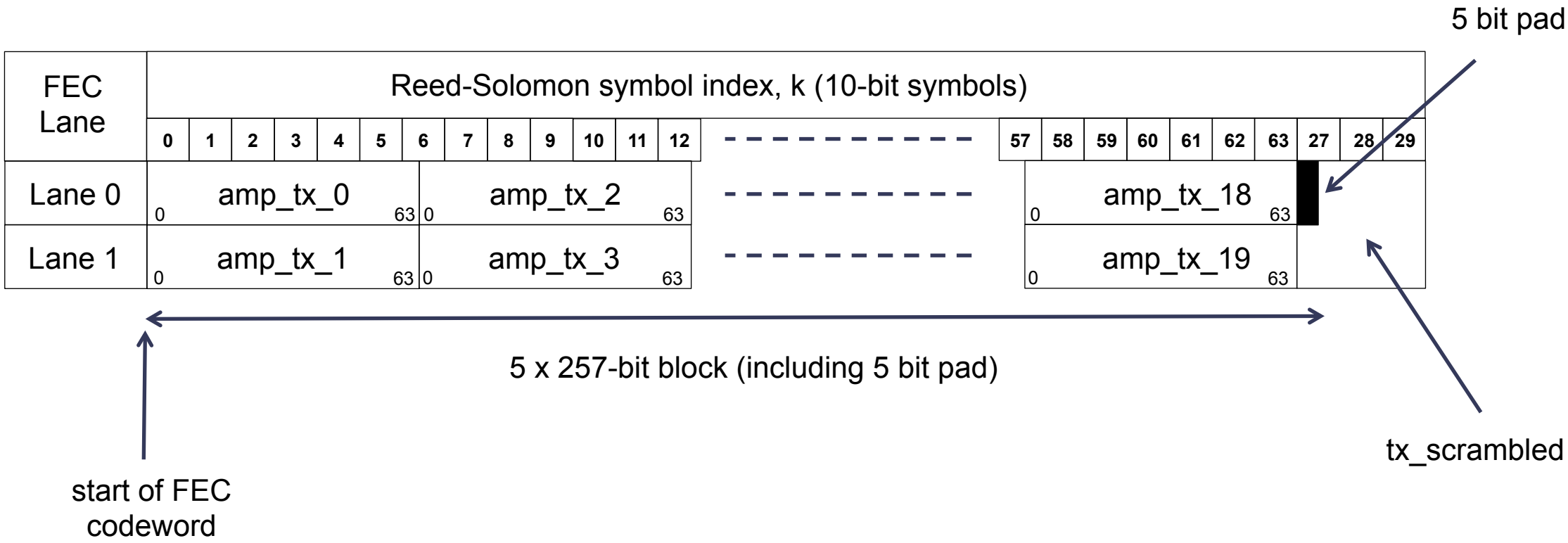
NG 100GbE Rx Data Flow

- Reverse of Tx



* = minor difference to Clause 91

NG 100GbE - Alignment marker mapping to FEC lanes



NG 100GbE - Open Items

- Details of changes to AM mapping to accommodate 2 lane distribution
  - Do we also need to change the AM mapping to put out unique AMs on both FEC lanes to support future bit muxing to a single 100G/Lane PMD ? Current 802.3bj proposal duplicates AM0 and AM16 on all four FEC lanes.
- Details of changes to Alignment locker state machine in Rx to accommodate the different AM pattern per FEC lane (compared to the case with 4 FEC lanes)
- FEC distribution over 2x50G lanes (current proposal) or over 4x25G lanes

NG 100GbE - Proposed Mapping to IEEE Documentation

- PCS - Existing Clause 82
  - no changes required
- FEC - Existing Clause 91
  - minor changes required to a small number of sub-clauses to accommodate the distribution over 2 rather than 4 FEC lanes
  - sub-clauses include AM mapping, Symbol distribution, Alignment lock
  - we could either edit the existing sub-clauses and add a ‘2 lane mode’, or add new sub-clauses to capture the new 2 lane requirements (maybe with reference to the current ‘4 lane’ sub-clauses).
- Potentially requires very minimal changes to the existing 100GbE PCS and FEC Clauses.

Backup

Recap - NG 100GbE Backwards Compatibility

- Proposal supports backwards compatibility with legacy hosts:
  - use downspeed serdes (run in 4x25G NRZ mode)
  - reduced bandwidth on new line card (but no different to 1G/10G and 40G/100G transition)
  - requires absolutely no new standards and/or product development
- Proposal supports backwards compatibility with legacy hosts at full bandwidth:
  - new module development with RS544 FEC sublayer installed in legacy host (Rob's Brown Field B)
- Proposal supports backwards compatibility with legacy silicon:
  - new line card with legacy silicon + new (4:2) PHY chip with RS544 FEC
  - this is identical to how RS528 FEC was introduced in transition from 802.3ba to 802.3bj/bm