50GbE and NG 100GbE Logic Baseline Proposal

Gary Nicholl - Cisco
Mark Gustlin - Xilinx
David Ofelt - Juniper

IEEE 802.3cd Task Force, July 25-28 2016, San Diego

Supporters

- Jonathan King - Finisar
- Tom Palkert - Molex
- Bharat Tailor - Semtech
- Hanan Leizerovich - Semtech
- Kent Lusted - Intel
- Pirooz Tooyserkani - Cisco
- Cedrik Begin - Cisco

Background

- At the 802.3cd meeting in Whistler it was agreed to adopt nicholl_3cd_01a_0516 as the basis for the 50GbE and 100GbE PCS and FEC architecture, with the exception of leaving the FEC lane count / distribution as TBD.
- http://www.ieee802.org/3/cd/public/May16/nicholl_3cd_01a_0516.pdf

Motion #4: 1:42 p.m.
Move to adopt nicholl_3cd_01a_0516 as the basis for the 50GbE and 100GbE PCS and FEC architecture, with the exception of leaving the FEC lane count / distribution as TBD

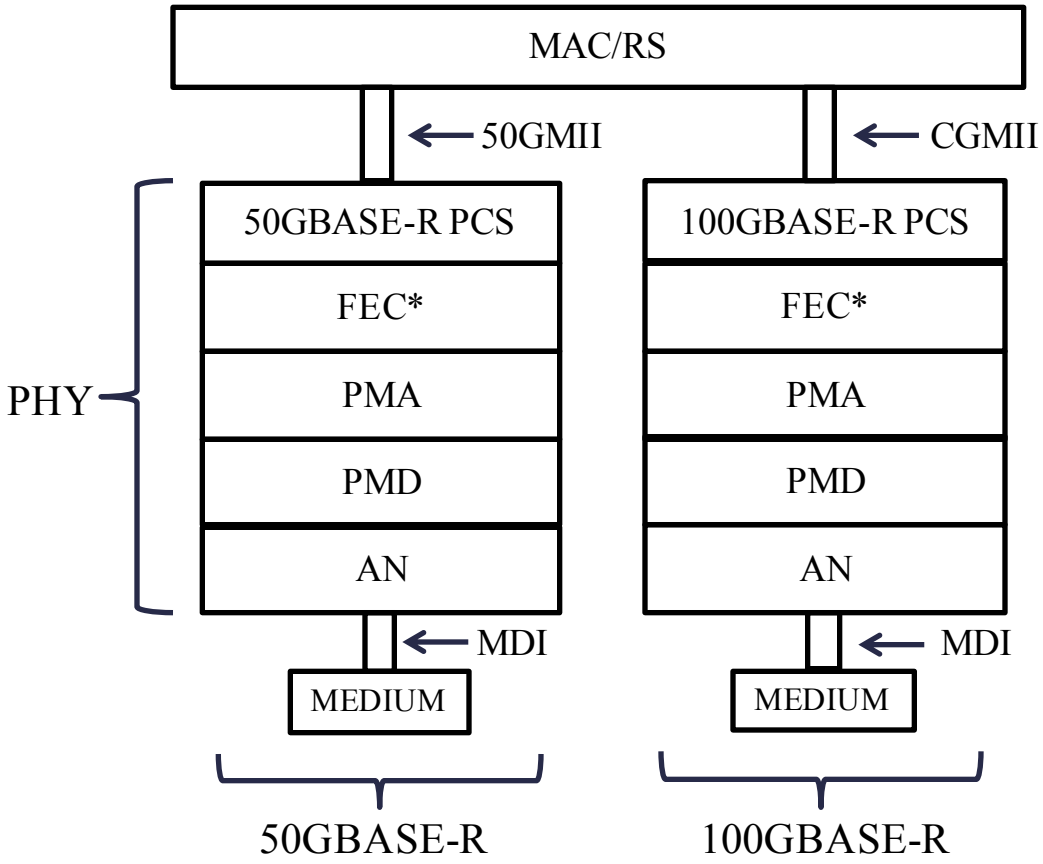
- M: Gary Nicholl
- S: Dave Ofelt
- Technical (>=75%),
- Y: 67 N: 1 A: 12
- Results: passes 1:52 p.m.

Introduction

- This presentation builds upon nicholl_3cd_01a_0516 and provides baseline proposals for the 50GbE and NG 100GbE logic layers
- The primary change for this presentation is related to the FEC lane count:
  - 50GbE: 2x26G FEC lanes; 100GbE 4x26G FEC lanes
  - A bit muxing PMA maps FEC lanes to 53Gb/s PAM4 lanes
  - This change was made to enable a broader range of implementation choices
- Additional changes include:
  - Update to AM to FEC lane mapping to reflect the adoption of 26G FEC lanes, including a proposal for the alignment marker (AM) bit patterns.
  - PAM4 precoding added to mitigate impact of burst errors (on links with dominant first DFE tap). Optional to enable
  - RS/MII baseline proposal
  - PMA baseline proposal

Architecture Overview

- Based on 802.3ba system architecture
- Separate PCS and FEC sub layers
- PCS and FEC can be separated by an optional AUI (not shown)
- FEC is mandatory for all 802.3cd PHY types
- FEC can be carried over a 25Gb/s or 50Gb/s per lane optional AUI (not shown)
- PMA is based on blind bit muxing



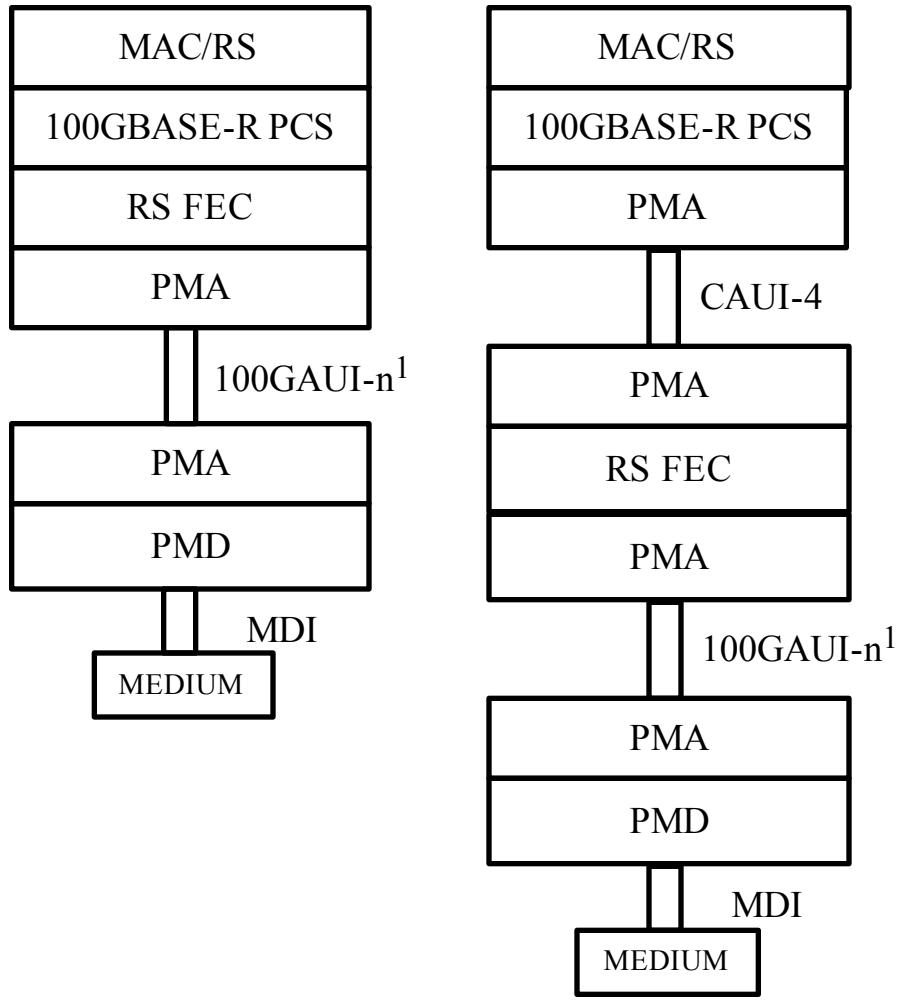
* FEC mandatory for all 802.3cd PHY types

RS/MII Baseline

- RS adapts the bit serial protocols of the MAC to the parallel format of the PCS
- MII provides an optional logical interfaces between the MAC/RS sublayers and the Physical Layer (PHY).
  - not physically instantiated
  - defines a common logical interface for all PHY types
  - often used for connecting RTL blocks within ASICs/FPGAs
- 100G RS and MII are already defined in Clause 81
- 50G RS and MII to be based on Clause 81

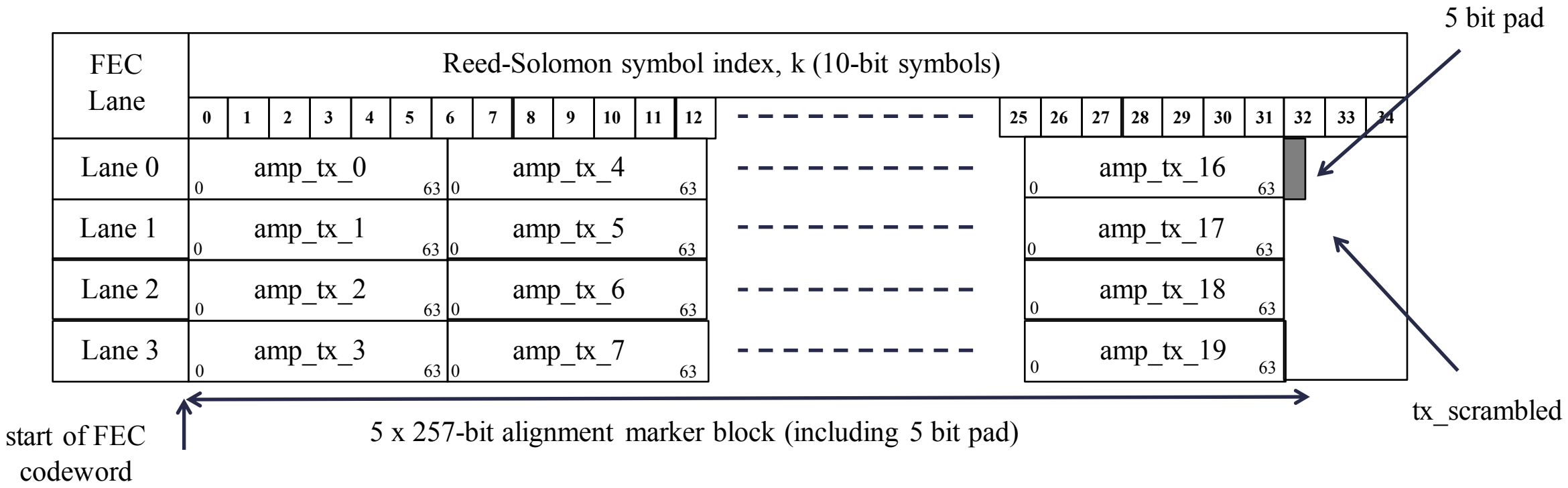
NG 100GbE PCS & FEC Overview

- PCS
  - Re-use existing 100GbE (Clause 82) PCS
  - No changes proposed
  - Supports optional CAUI-4
- FEC
  - Based on 802.3bj RS(544,514) FEC (Clause 91)
  - Need to modify AM mapping to enable bit muxing
  - Supports an optional 2 or 4 lane AUI interface



Note 1: n = 2 or 4 lanes

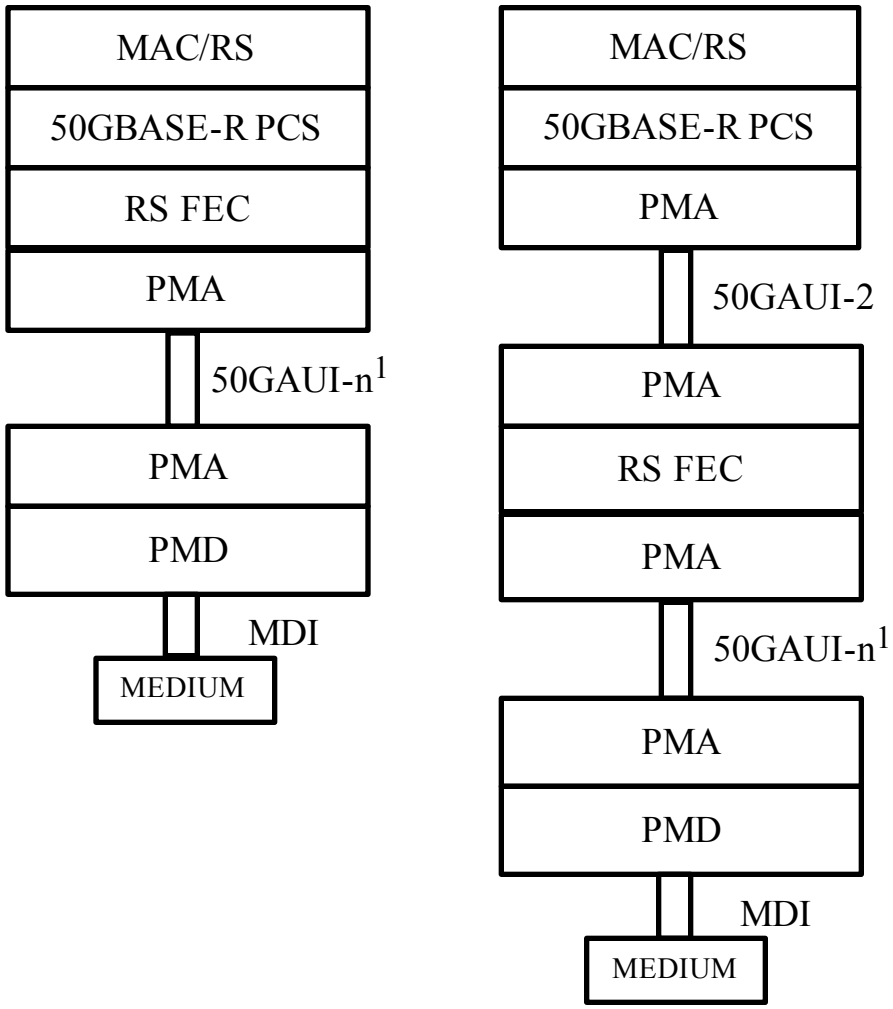
NG 100GbE - Alignment Marker mapping to FEC lanes



- Based on Clause 91. Exact AM mapping still TBD
- Initial analysis indicates that Clause 91 AM mapping needs to be modified to avoid clock content issues with repeating AM0 and AM16 patterns when bit muxing FEC lanes.

50GbE PCS & FEC Overview

- PCS
  - Based on overclocked 40GbE PCS (Clause 82)
  - 4 x PCS lanes running at 12.890625 Gb/s
  - AM spacing changed from 16k to 20k* to better support FEC sublayer
  - Supports an optional 50GAUI-2
- FEC
  - Based on 802.3bj RS(544,514) FEC (Clause 91)
  - FEC symbols distributed to 2 x 26G FEC lanes
  - Modified AM mapping (4 PCS lanes, 2 FEC lanes, and to enable bit muxing)
  - Supports an optional 1 or 2 lane AUI interface

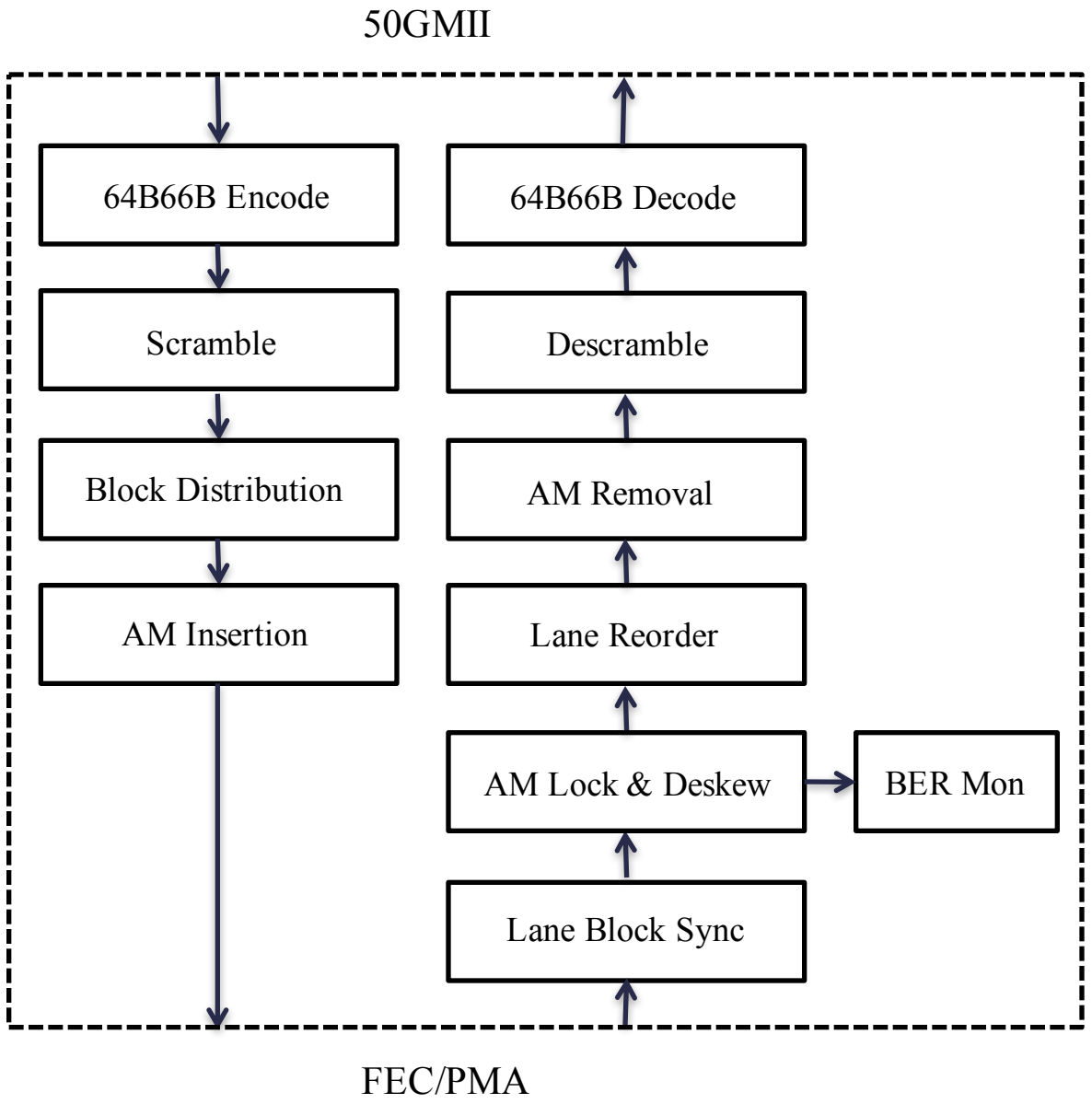


Note 1: n = 1 or 2 lanes

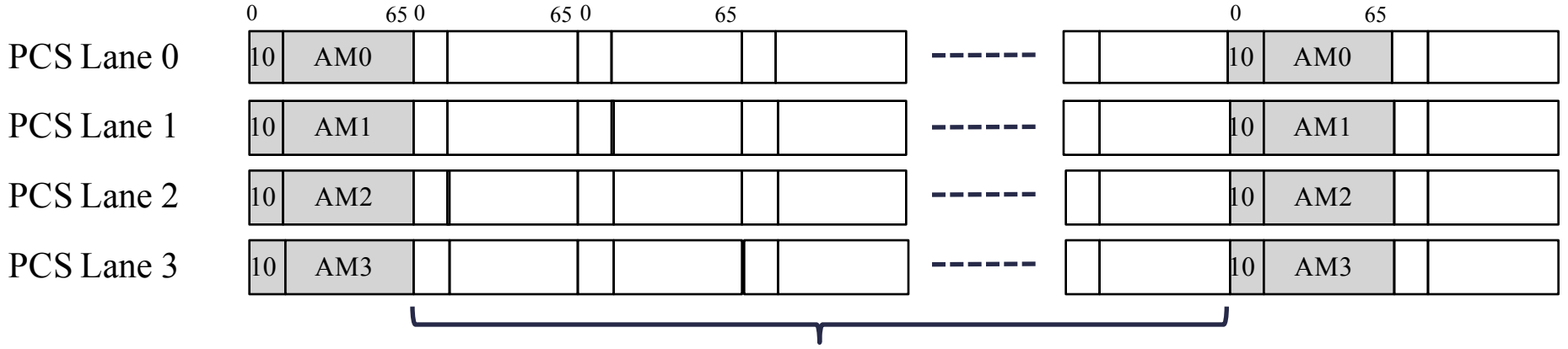
* 20k spacing is consistent with other 50G FEC implementations.

50GbE PCS Data Flow

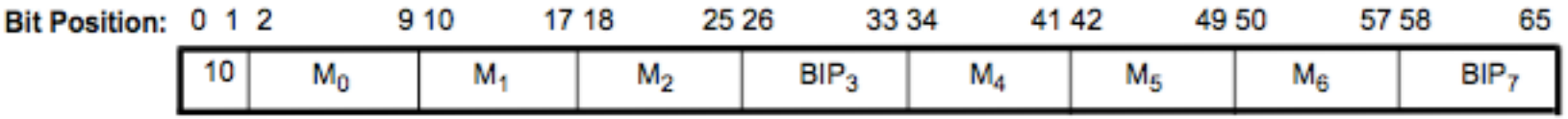
- Identical data flow to 40GbE PCS in Clause 82
- 4 x PCS lanes running at 12.890625 Gb/s (overclocked rate)
- 4 x 66-bit alignment markers (AM), one per PCS lane, inserted periodically
- AM spacing (start of one AM to the start of next AM) modified to 20480 66-bit blocks, to better align with FEC codeword boundaries



50GbE PCS AM Details



50GBASE-R Alignment marker spacing



50GBASE-R Alignment marker format

50GbE PCS AM Details

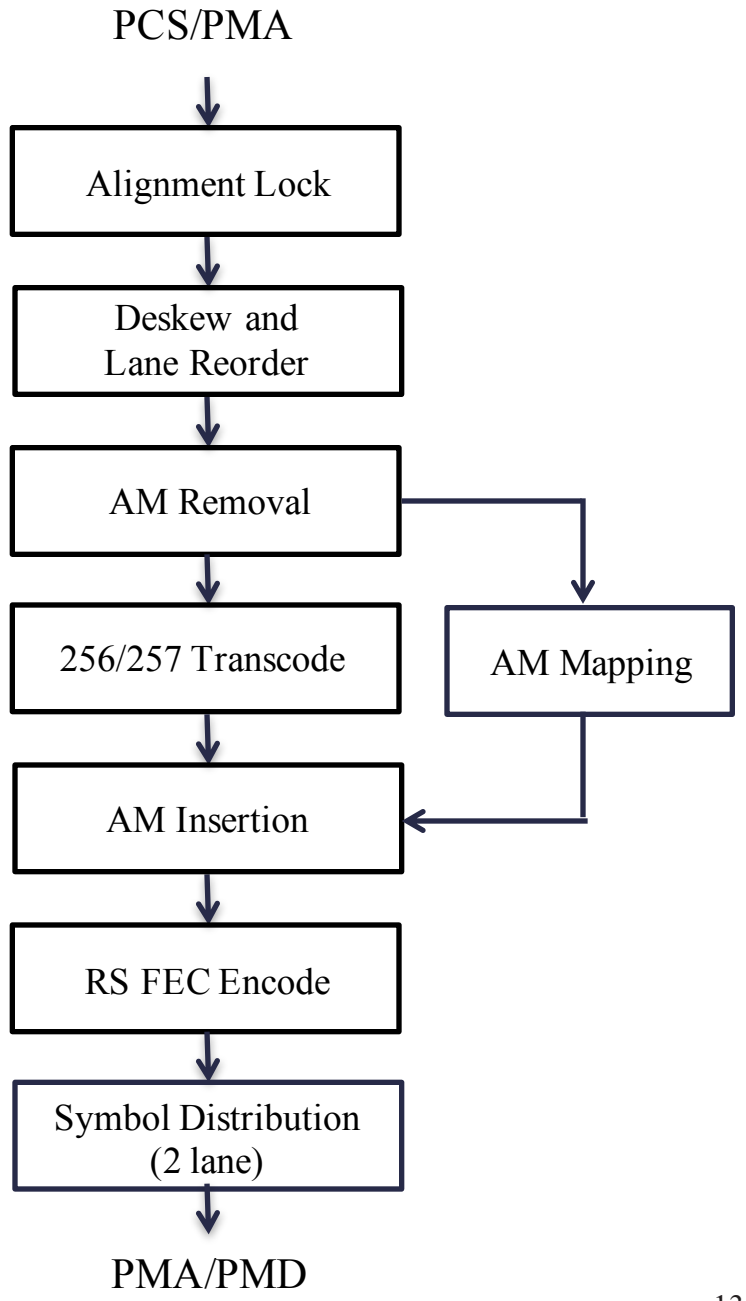
PCS Lane Number	Encoding {M0,M1,M2,BIP3,M4,M5,M6,BIP7}
0	0x90, 0x76, 0x47, BIP3, 0x6F, 0x89, 0xB8, BIP7
1	0xF0, 0xC4, 0xE6, BIP3, 0x0F, 0x3B, 0x19, BIP7
2	0xC5, 0x65, 0x9B, BIP3, 0x3A, 0x9A, 0x64, BIP7
3	0xA2, 0x79, 0x3D, BIP3, 0x5D, 0x86, 0xC2, BIP7

Note: Each octet is transmitted LSB to MSB

50GBASE-R Alignment marker encodings
(identical to 40GBASE-R encodings)

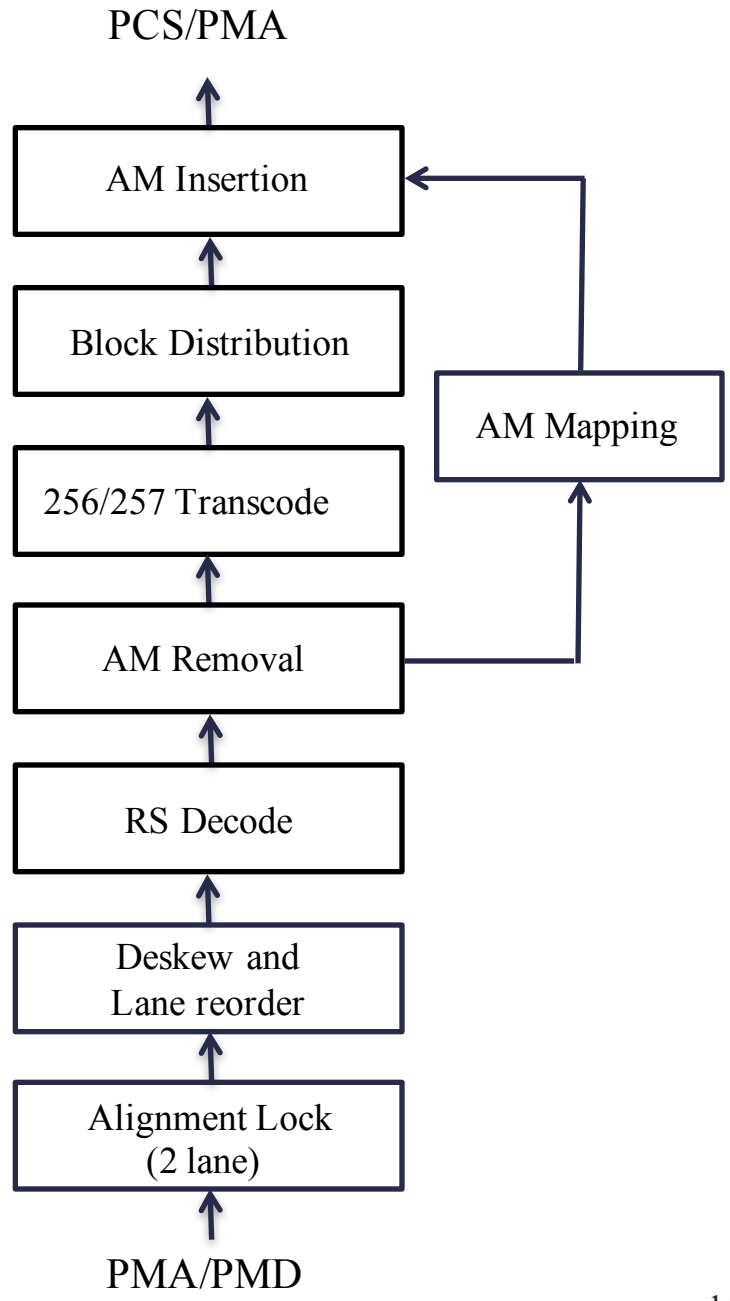
50GbE Tx FEC Data Flow

- Data flow is essentially identical to 802.3bj Clause 91
- FEC encoder is RS(544,514) running in a 1x50G configuration
- FEC encoder output is distributed to 2 FEC lanes on a symbol by symbol basis
- A single 257-bit alignment marker (AM) is inserted into the first 257 message bits to be transmitted from every 1024th FEC codeword

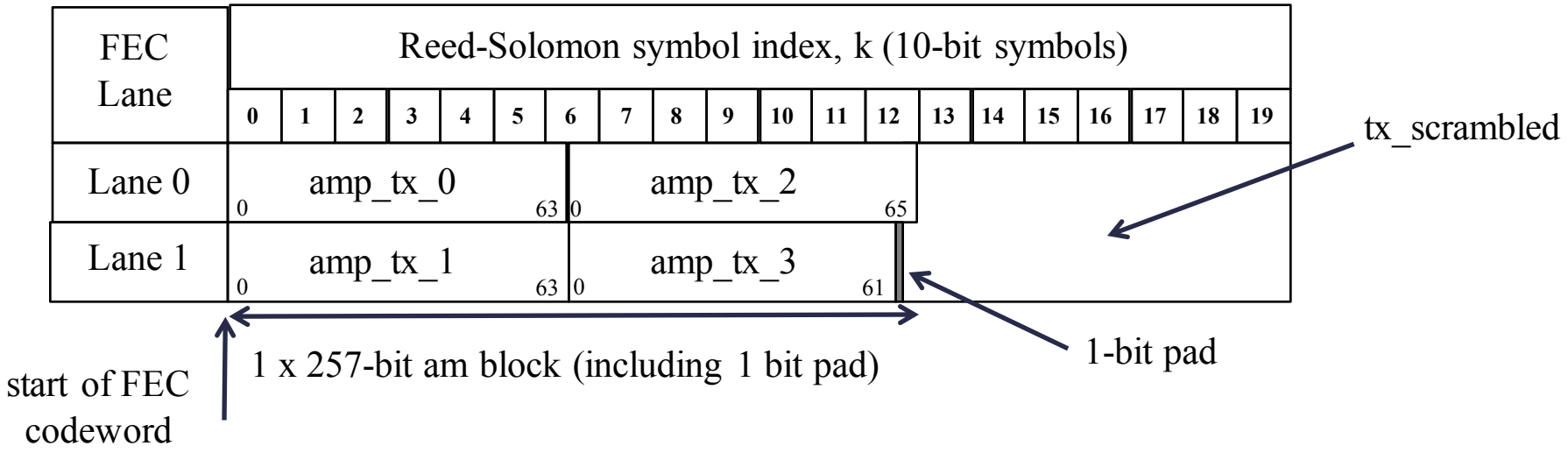


50GbE Rx FEC Data Flow

- Reverse of Tx



50GbE - Alignment Marker mapping to FEC lane



- Based on Clause 91 mapping, but modified to support 4 PCS lanes, 2 FEC lanes and to enable bit muxing of FEC lanes
- Exact AM mapping still TBD
- Note: amp_tx_0, amp_tx_1 =64 bits, amp_tx_2=66 bits, amp_tx_3=62 bit

PMA Baseline

- Identical PMA functions as described in Clause 120
  - Support for bit muxing
  - Support for Gray coding and PAM4 coding on 50G interfaces
  - Mapping between logical and physical lanes is not defined nor constrained
- PAM4 Precoding
  - Mandatory to implement (Tx & Rx) on PMAs driving backplane, copper and C2C.
  - Optional to enable (on links with dominant first DFE tap)
  - Precoding is implemented after Gray coding
- With no FEC the per lane signaling rate is 25.78125Gb/s
  - 2x25.78125Gb/s NRZ for 50GbE or 4x25.78125Gb/s NRZ for 100GbE (defined in Clause 83)
- With RS 544 FEC the per lane signaling rate is 26.5625Gb/s or 53.125 Gb/s
  - 2x26.5625Gb/s NRZ for 50GbE or 4x26.5625Gb/s NRZ for 100GbE
  - 1x53.125 Gb/s PAM4 for 50GbE or 2x53.125 Gb/s PAM4 for 100GbE

EEE

- The EEE baseline proposal is being addressed in a separate presentation

Conclusion

- This presentation provides 50GbE and NG 100GbE baseline proposals for:
  - RS/MII
  - PCS
  - FEC
  - PMA

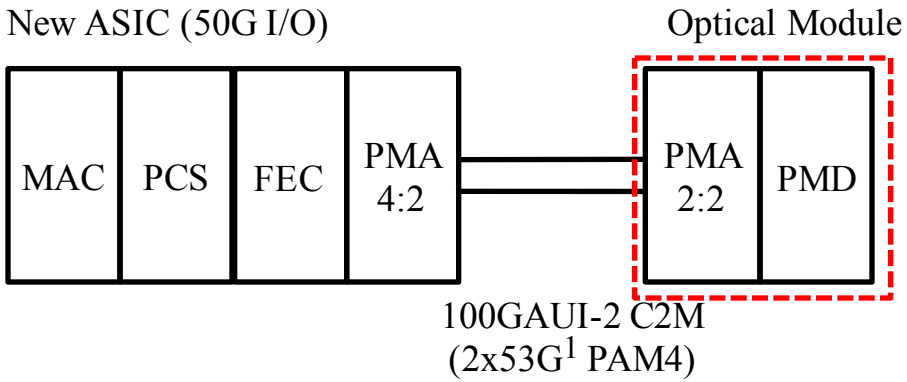
Thanks !!

Backup: Use cases

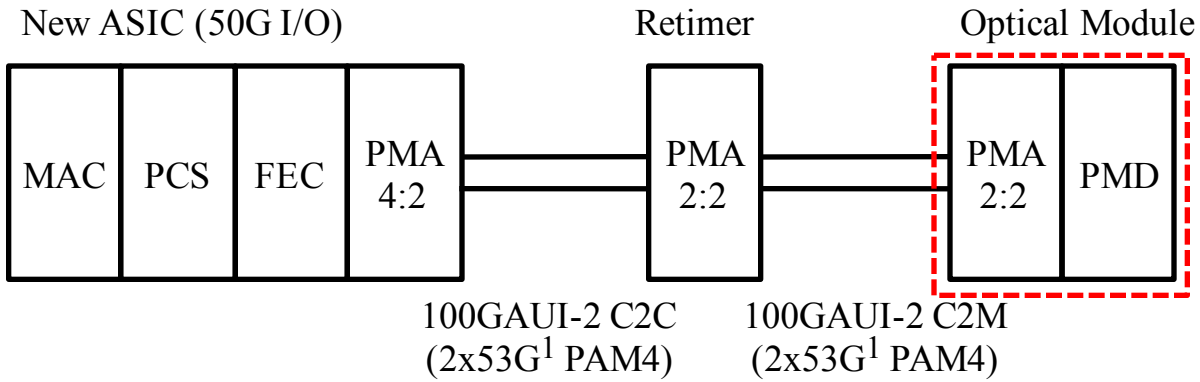
- The following slides provide examples of use cases that can be supported with the proposed 50GbE and NG 100GbE architecture.

NG 100GbE Use Cases (New Port ASIC)

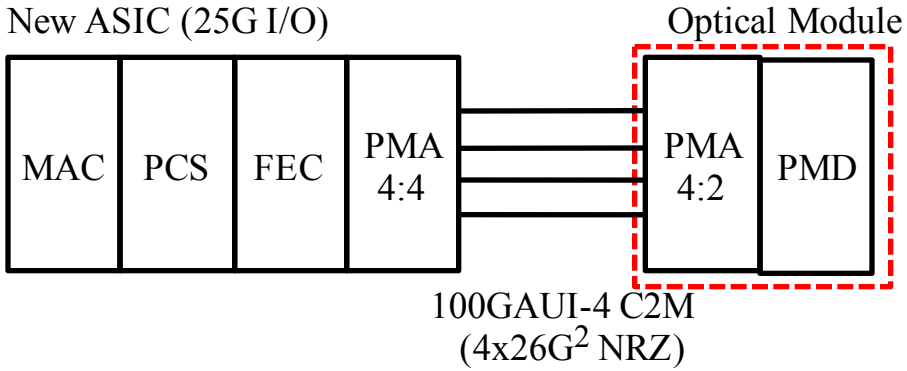
USE CASE #1



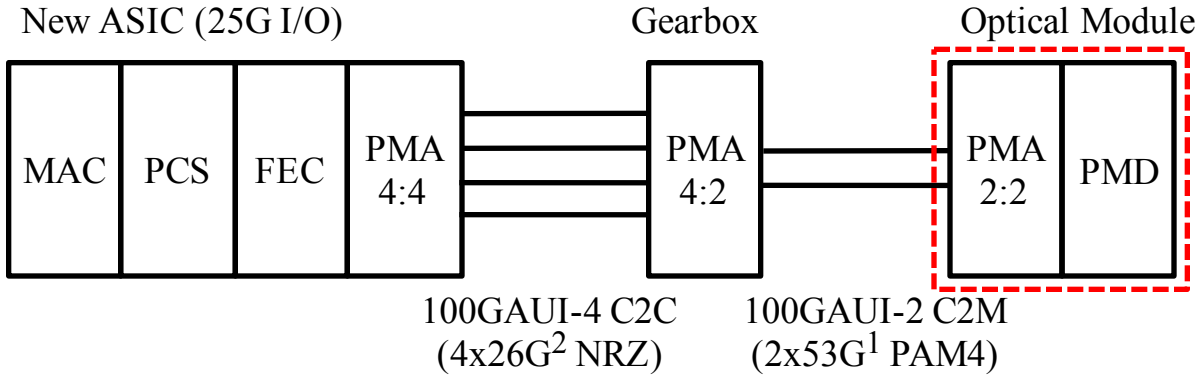
USE CASE #2



USE CASE #3



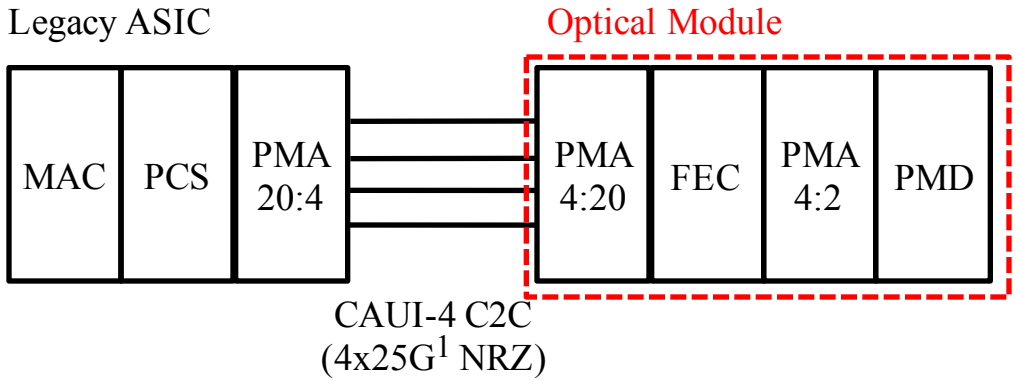
USE CASE #4



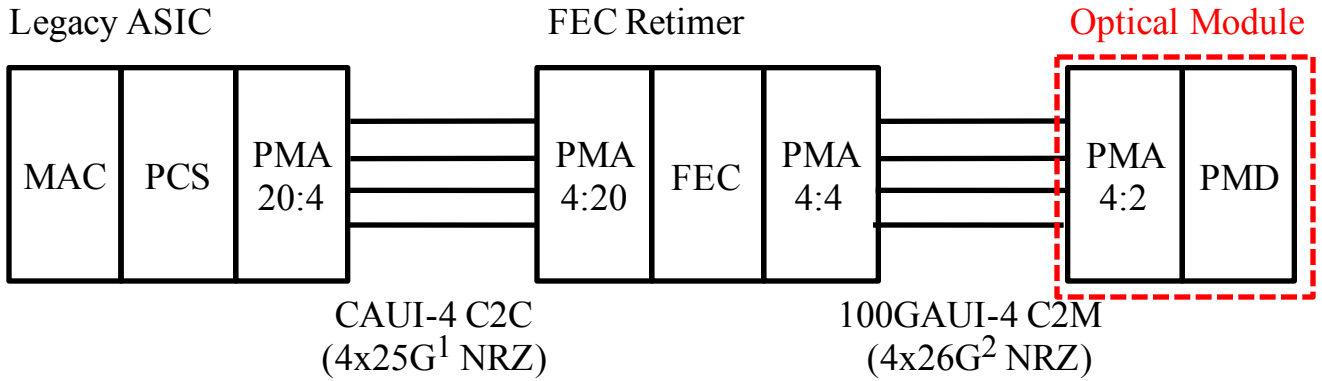
1 = 53.125 Gb/s, 2 = 26.5625Gb/s

NG 100GbE Use Cases (Legacy Port ASIC)

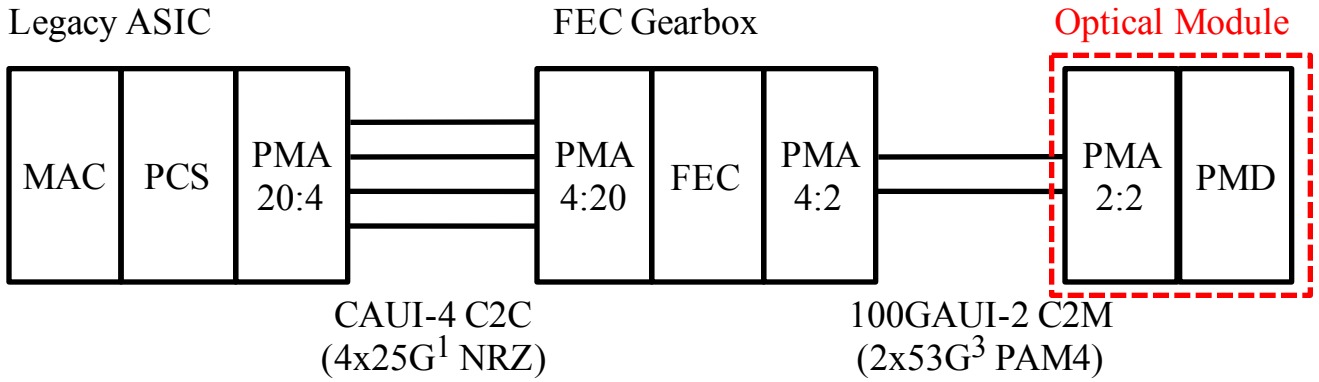
USE CASE #5



USE CASE #6



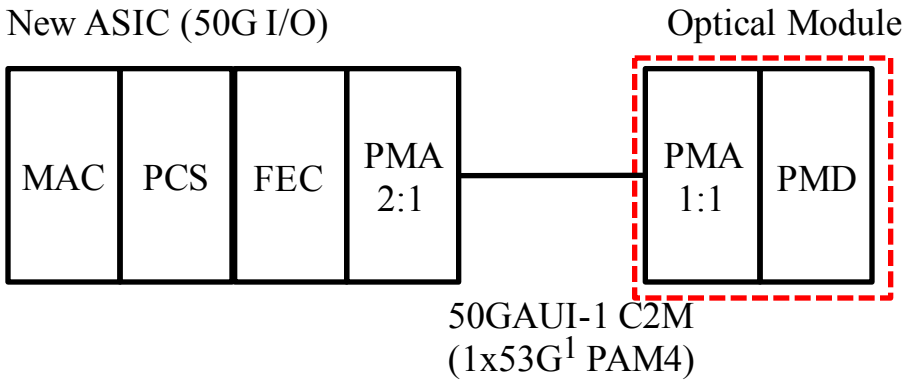
USE CASE #7



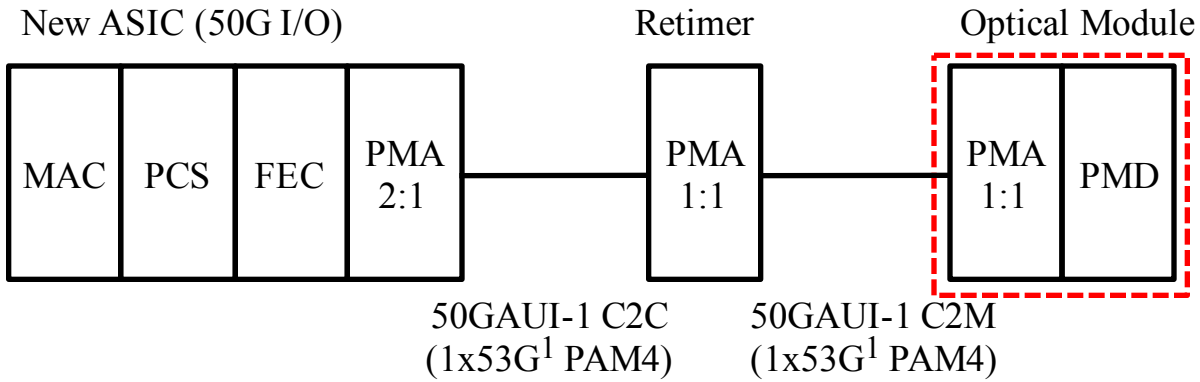
1 = 25.78125Gb/s , 2 = 26.5625Gb/s, 3 = 53.125 Gb/s

50GbE Use Cases (New Port ASIC)

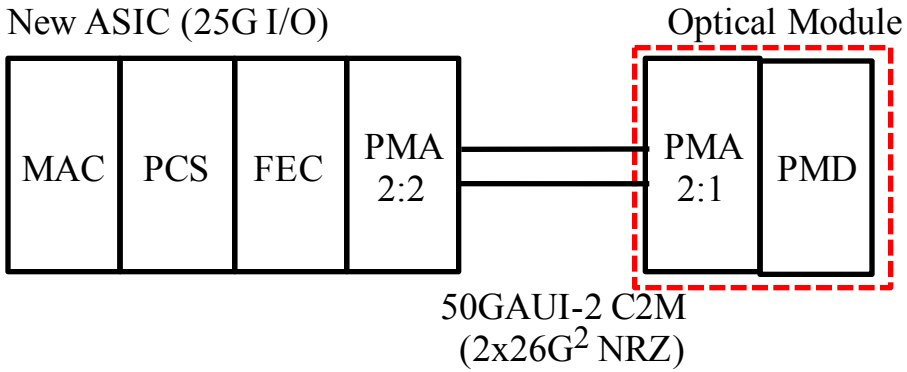
USE CASE #1



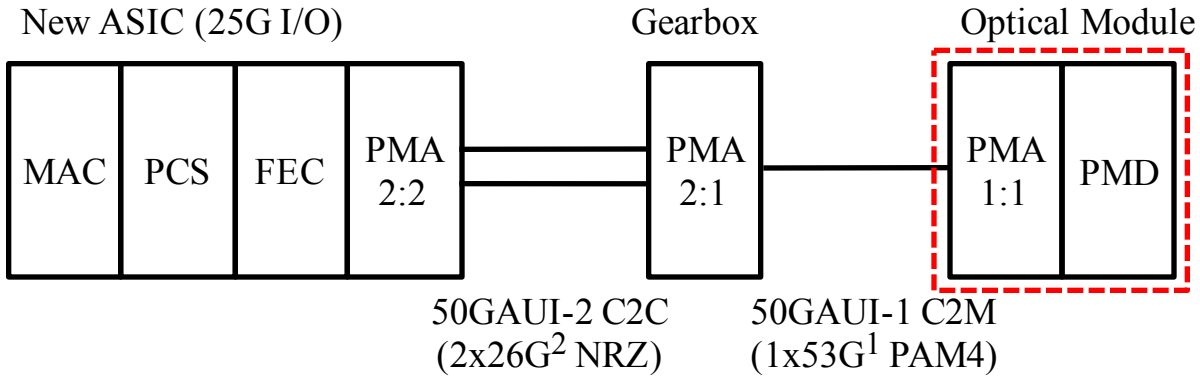
USE CASE #2



USE CASE #3



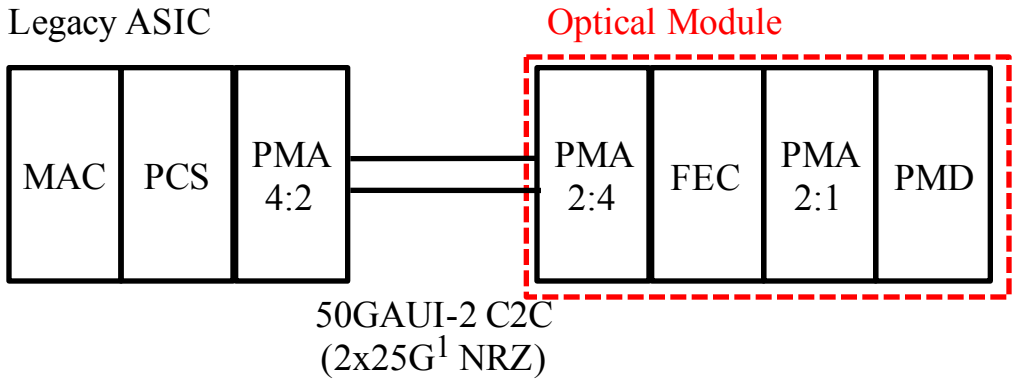
USE CASE #4



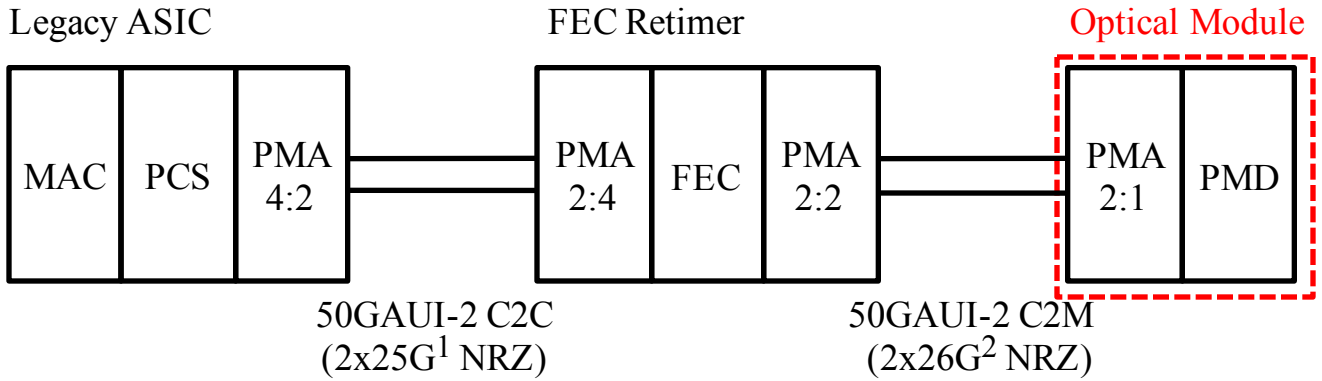
1 = 53.125 Gb/s, 2 = 26.5625Gb/s

50GbE Use Cases ("Legacy" Port ASIC - PCS only)

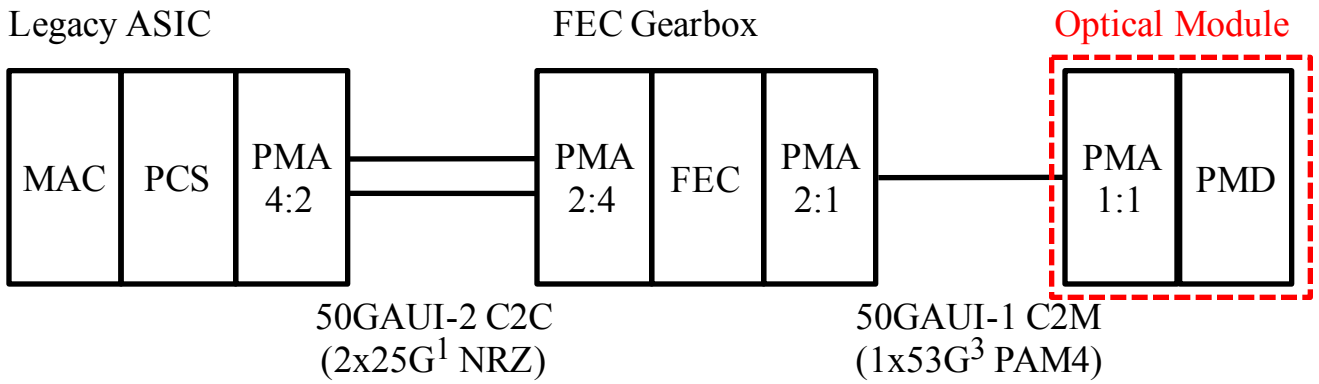
USE CASE #5



USE CASE #6



USE CASE #7



1 = 25.78125Gb/s , 2 = 26.5625Gb/s, 3 = 53.125 Gb/s

Backup: PMA Examples

50GbE PMA Examples:

PMA (2:2): 2x25G NRZ retimer (for connecting the PCS with the FEC sublayer)
PMA (2:2): 2x26G NRZ retimer (for connecting RS 544 FEC sublayer to a PMA 2:1)
PMA (2:1): 2x26G NRZ to 1x53G PAM4 (to connect RS 544 FEC sublayer to a PMA 1:1 or PMD)
PMA (1:1): 1x53G PAM4 retimer (to go between a PMA 1:1 and a PMD)

100GbE PMA Examples:

PMA (4:4): 4x25G NRZ retimer (for connecting the PCS with the FEC sublayer)
PMA (4:4): 4x26G NRZ retimer (for connecting RS 544 FEC to a PMA 4:2 sublayer)
PMA (4:2): 4x26G NRZ to 2x53G PAM4 (to connect RS 544 FEC sublayer to a PMA 2:2 or a PMD)
PMA (2:2): 2x53G PAM4 retimer (to go between a PMA 2:2 and a PMA 2:2 or a PMD)