100GE AUI Extender Sublayer

Yuchun Lu, Huawei
Yan Zhuang, Huawei

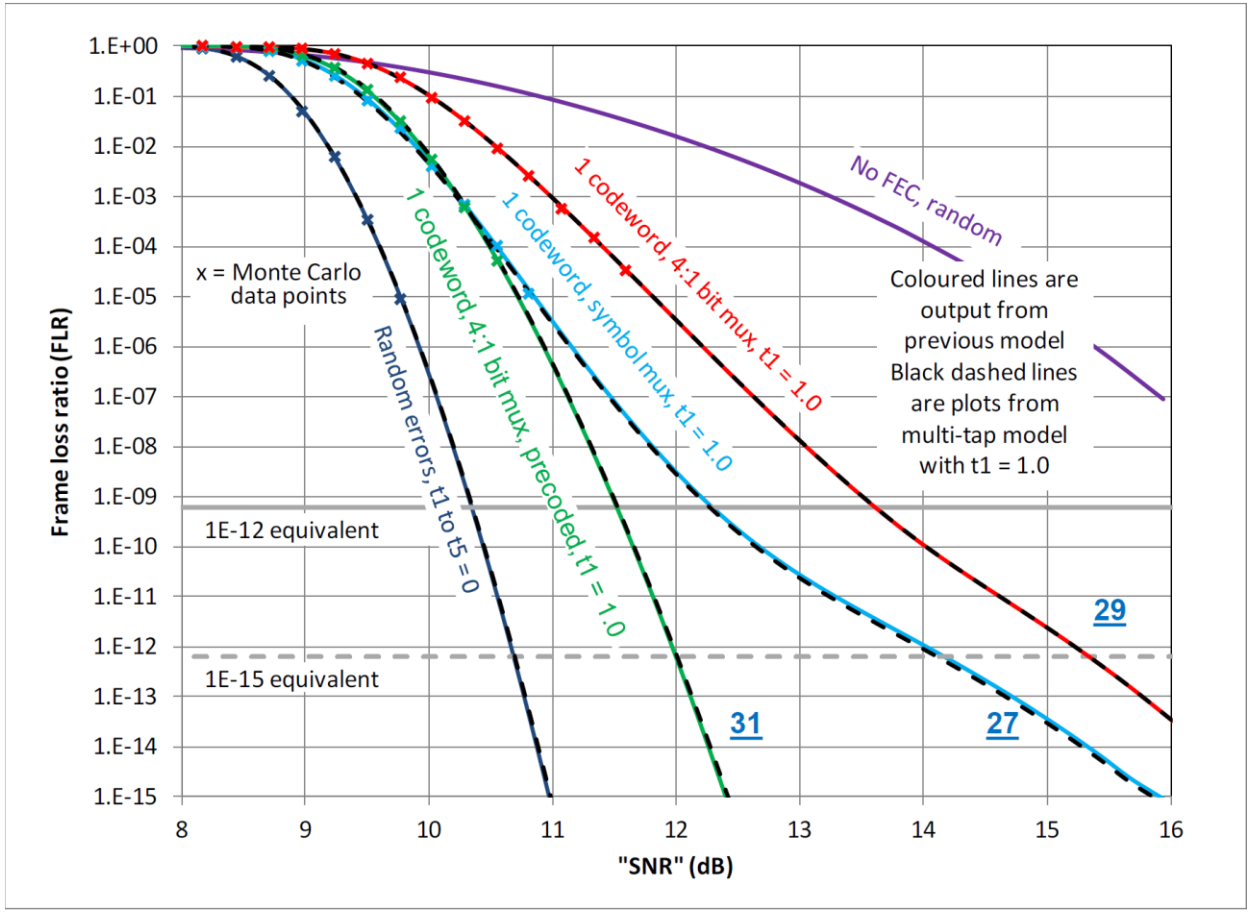
IEEE 802.3 100 Gb/s, 200 Gb/s, and 400 Gb/s Electrical Interfaces Task Force

HUAWEI

Introduction

- FEC performance concern for 100GE-CR1/KR1 multi-tap DFEs with 4:1 bitmux PMA was shown and interleaved FEC was proposed in gustlin_3ck_01_1118.
- Interleaved FEC will introduce more latency and complicated CDR is needed to address the interoperability and compatibility issues. Both the latency and the complicated CDR are not affordable in some applications. In-depth analysis was given lu_3ck_adhoc_01_022719.
- Further, analysis of potential solutions for 100G-CR1/KR1 multi-tap DFE error propagation was given in lu_3ck_02_0319 including PMD, PMA and FEC sublayer solutions.
- This presentation looks into the PMA solution, which introduces a new optional AUI Extender Sublayer to implement the PMA remapping to resolve the FEC performance concern of “multi-tap DFEs with 4:1 bitmux PMA” with negligible cost (both in latency and circuit) and negligible impact on the standard and system. It can simplify the CDR and support “protocol independent CDR”.

1-tap DFE does not have FEC performance concern



anslow_3ck_01_0119 page 5,
healey_100GEL_01_0318

1. Pre-coding should be used in 100 Gb/s per lane electrical PHYs as a tool to improve error correction performance. See healey_100GEL_01_0318, zhang_3ck_01a_0918, lu_3ck_01_0319.

2. 1-tap DFE does not have FEC performance concern if pre-coding is applied, even with the worst case DFE weight t1=1.0 or error propagation probability a=0.75.

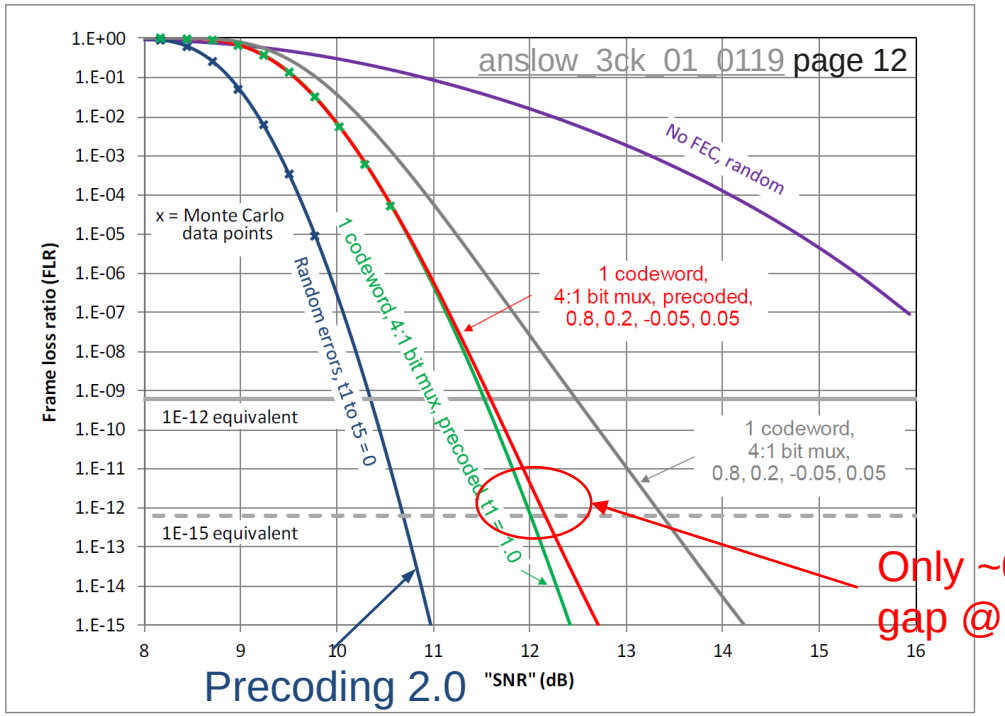
3. 1-tap DFE is a special case of n-tap DFE, and n-tap DFE can be viewed as an equivalent 1-tap DFE with time variant DFE weight. If DFE weights are well constrained, the FEC performance concern of n-tap DFE receiver can be mitigated. (More details see lu_3ck_adhoc_01a_010219)

The FEC concern only happens for extreme settings (maybe too pessimistic and un-realistic) of minority multi-tap DFE receiver.

HUAWEI

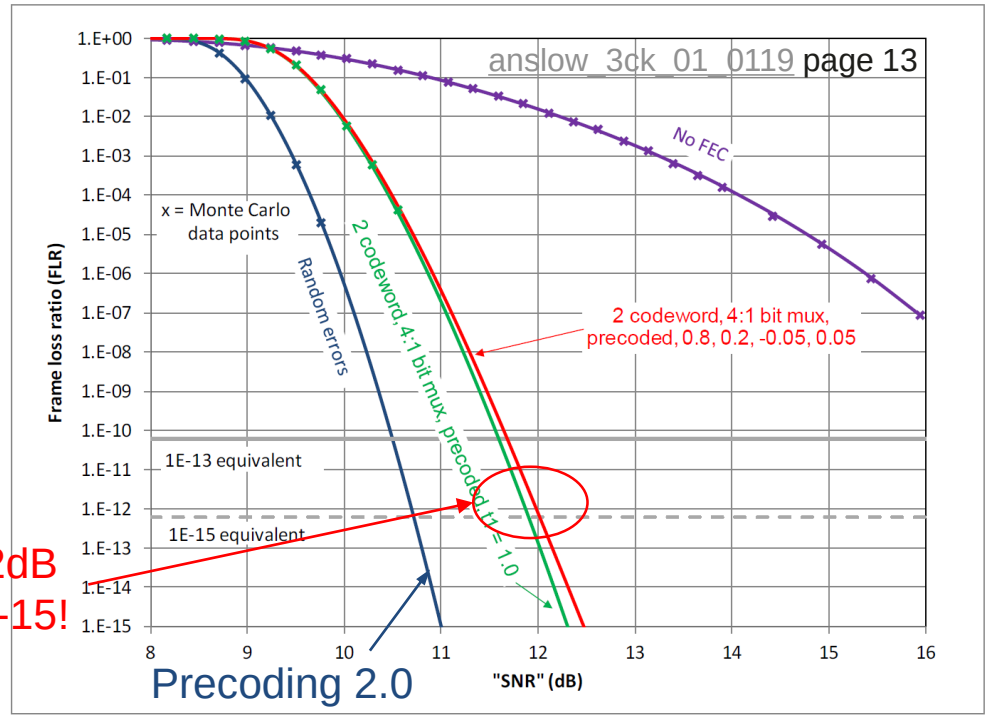
Constraining DFE weights can relief FEC performance concern

Non-interleaved FEC + 4:1 bit mux.



Only ~0.2dB gap @1e-15!

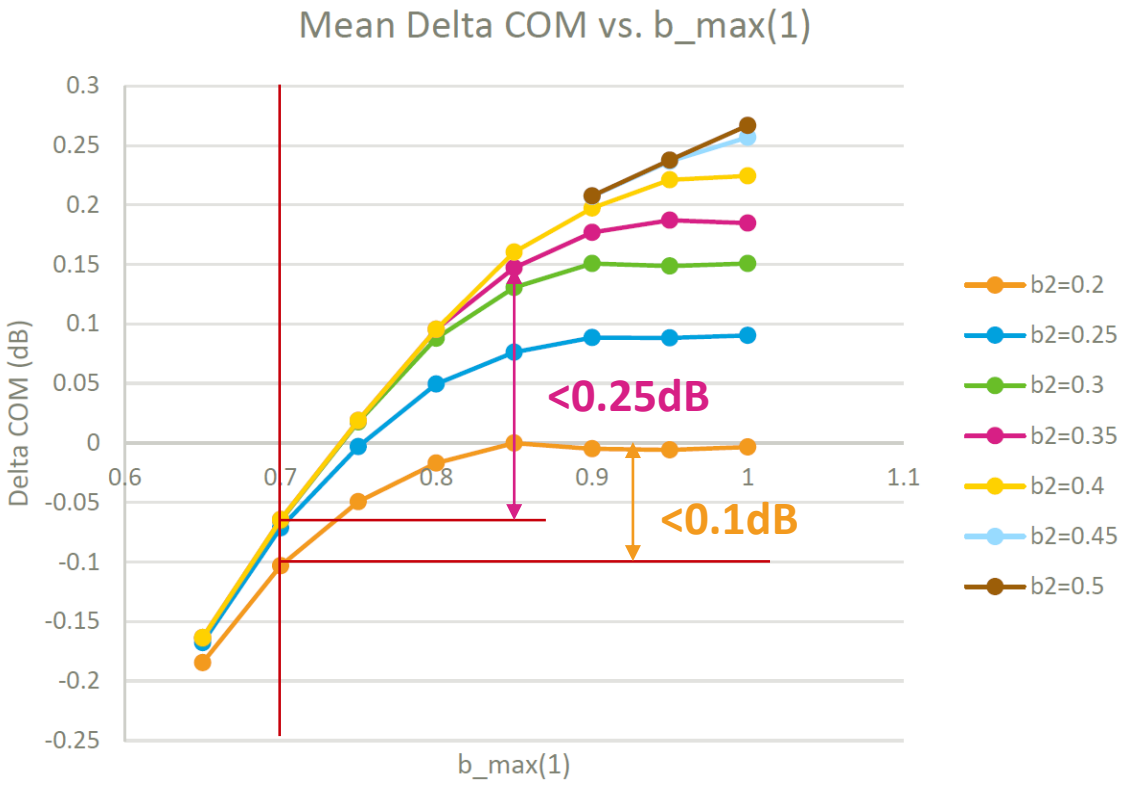
2-way interleaved FEC + 4:1 bit mux.



- **[0.8 0.2 -0.05 0.05]** is a reasonable DFE weight setting close to the real worst case, bmax=[0.7, 0.2] was used in IEEE 802.3bj&cd. No significant improvement **(<0.2dB**) was observed by "interleaved FEC" in this case.
- Relaxing "bmax" to [0.85, 0.35] is under discussion (wu_3ck_adhoc_01_022719), however the COM benefit by increasing bmax(1) from 0.7 to 0.85 is **<0.25dB** when bmax(2…Nb)=0.35; and **<0.1dB** when bmax(2…Nb)=0.2.
- "Increasing bmax(1) to 0.85" only makes sense when the FEC performance penalty is negligible.

HUAWEI

Constraining DFE weights make the whole work reasonable

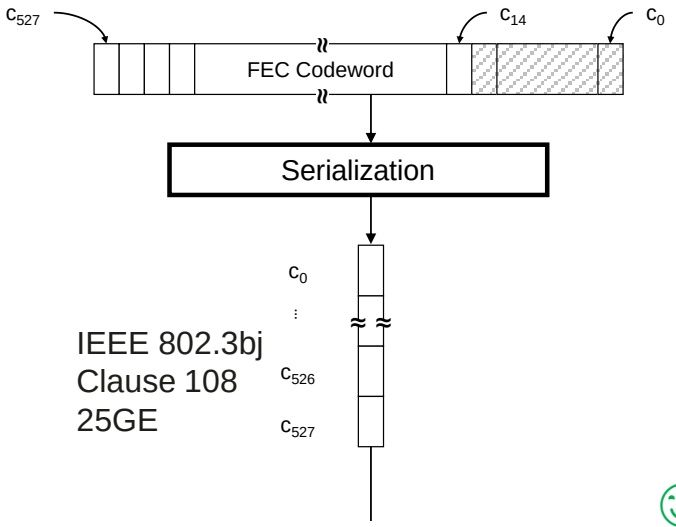


wu_3ck_01b_0319 Page 16.

- The benefits of increasing bmax(1) from 0.7 to 0.85 is minor.
  - <0.25dB for bmax(2) = 0.35; <0.1dB for bmax(2) = 0.2.
- If the penalty of increasing bmax(1) from 0.7 to 0.85 is not negligible. Bmax=[0.7 0.2] will be a better choice, because we did not have any FEC performance concern in IEEE 802.3bj&cd, i.e. “50G CR-2/KR-2”; “25G CR-1/KR-1”, “50G CR-1/KR-1” and “100G CR-2/KR-2”, in which “Symbol mapping” or “2:1 Bit mux PMA” was adopted.
- Constraining DFE weights is reasonable. There will be a concern if COM gain <0.25dB while FEC performance penalty >0.25dB due to the DFE error propagation.

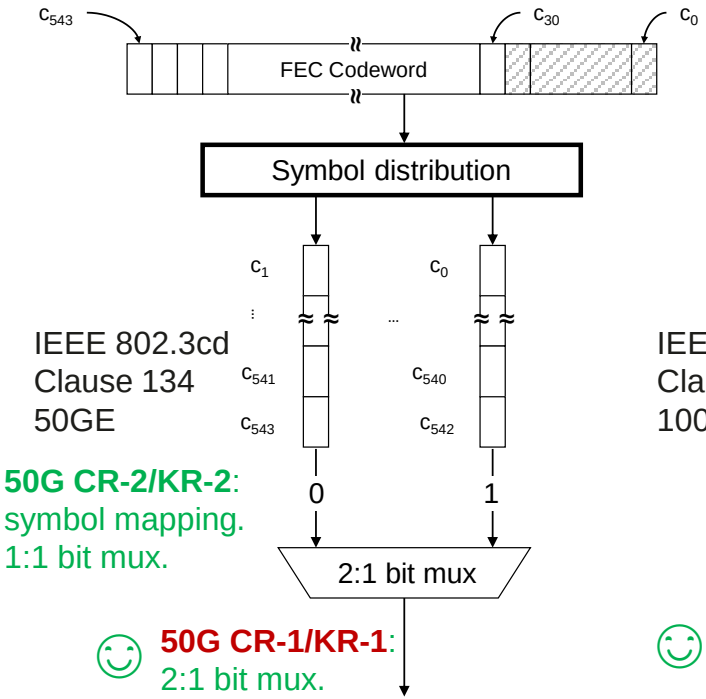
HUAWEI

Symbol mapping can address the FEC performance concern



IEEE 802.3bj
Clause 108
25GE

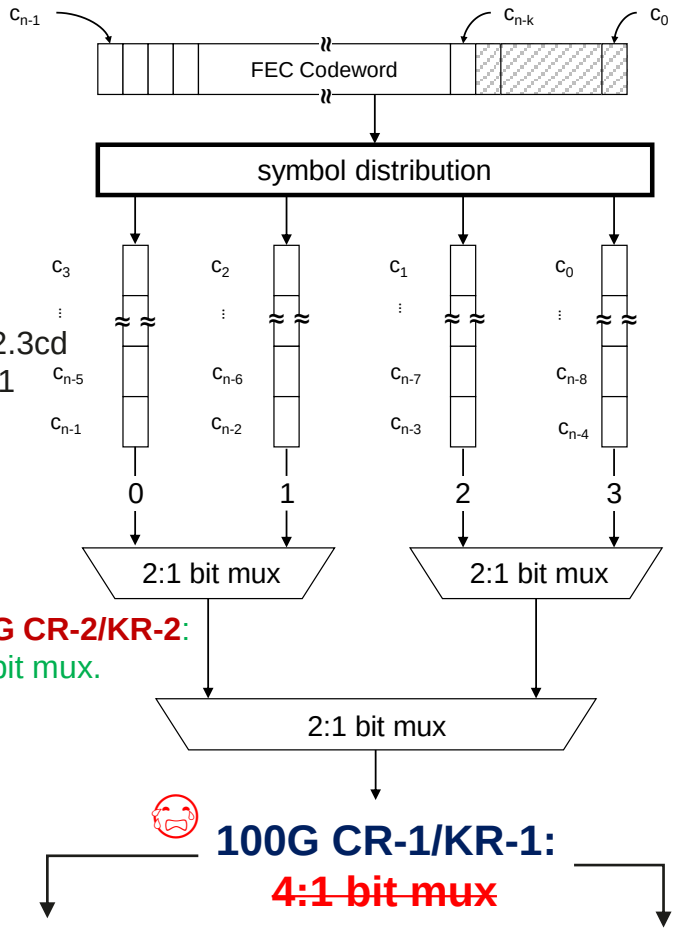
☺ **25G CR-1/KR-1**: direct codeword/symbol mapping. **RS(528, 514)** + 1:1 bit mux.



IEEE 802.3cd
Clause 134
50GE

☺ **50G CR-2/KR-2**: symbol mapping. 1:1 bit mux.

☺ **50G CR-1/KR-1**: 2:1 bit mux.



IEEE 802.3cd
Clause 91
100GE

☺ **100G CR-2/KR-2**: 2:1 bit mux.

☹ **100G CR-1/KR-1:**
~~**4:1 bit mux**~~

1 FEC + 1:1 bit mux (symbol mapping) ≈ 2-way interleaved FEC + 2:1 bit mux.

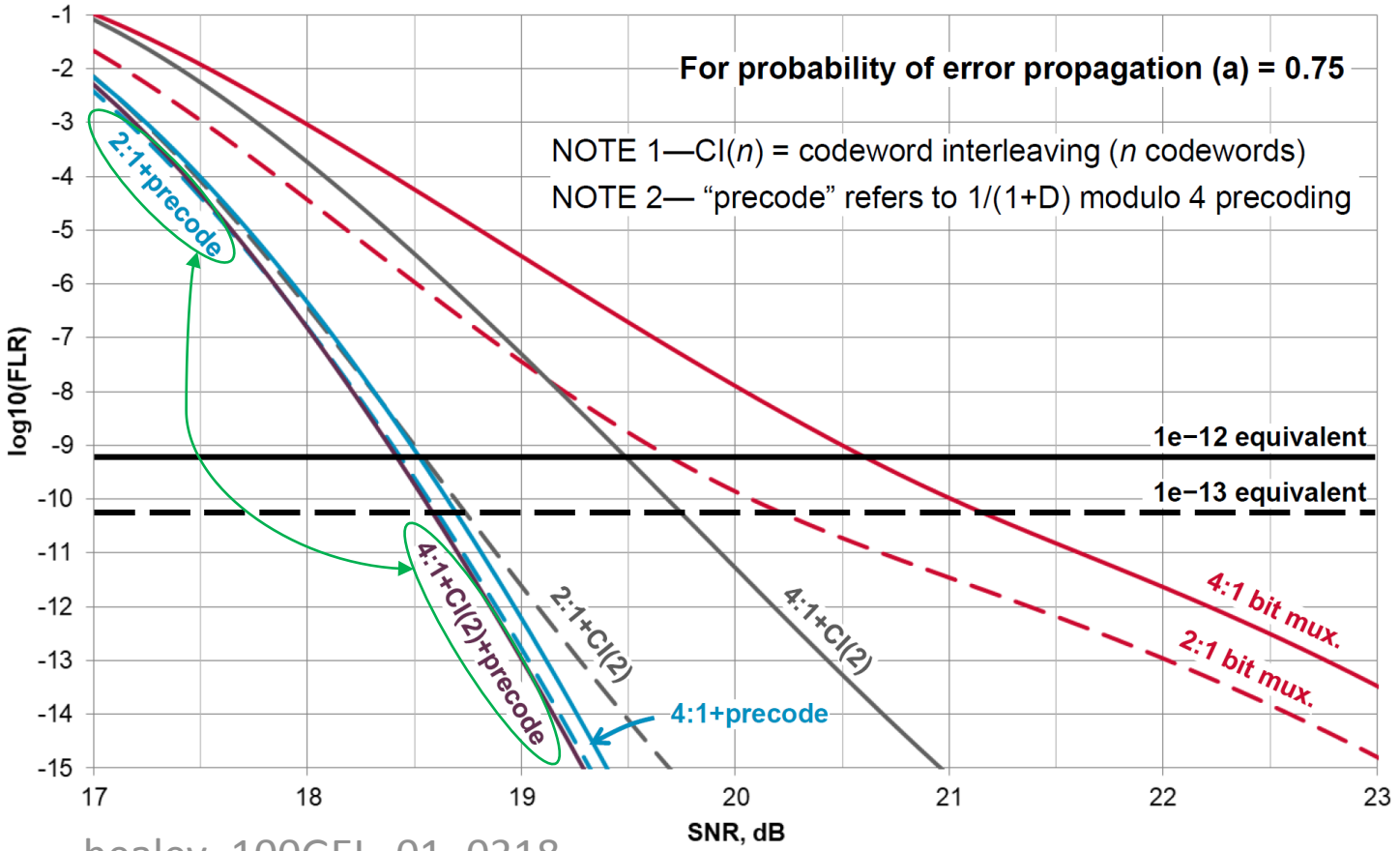
Same as “50G CR-2/KR-2”; Better than “25G CR-1/KR-1”, “50G CR-1/KR-1” and “100G CR-2/KR-2”.

1. “Non-Interleaved FEC + Symbol Mapping PMA” has the same performance as 50G CR-2/KR-2 and is better than “25G CR-1/KR-1”, “50G CR-1/KR-1”, and “100G CR-2/KR-2”.
2. “Symbol mapping PMA” can definitely solve the FEC performance concern for 100G CR-1/KR-1 which was approved by previous projects (802.3bj&802.3cd).

gustlin_3ck_01_1118

HUAWEI

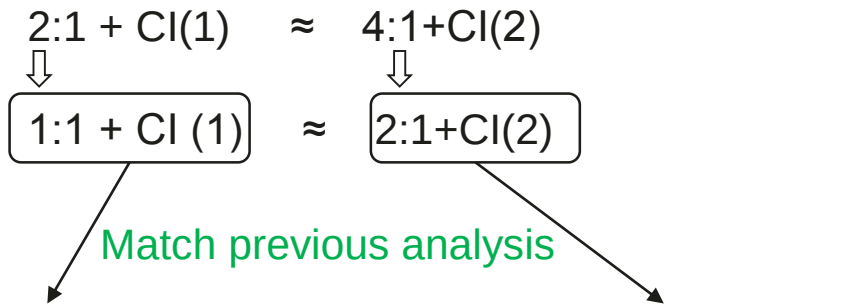
Symbol mapping can address the FEC performance concern



healey_100GEL_01_0318

May need some reasonable constrains on DFE weights for multi-tap DFE receivers.

- Relationship between reducing 'n' in 'n:1 bit mux' and increasing 'm' in 'm-way interleaved FEC' can be observed.
- "1 FEC + 2:1 bit mux (2:1 precode)" has identical performance compared with "2-way interleaved FEC + 4:1 bit mux (4:1+CI(2)+precode)".
- With the n-tap DFE weight well controlled, and precoding turned on, we can find the following equivalent configurations:



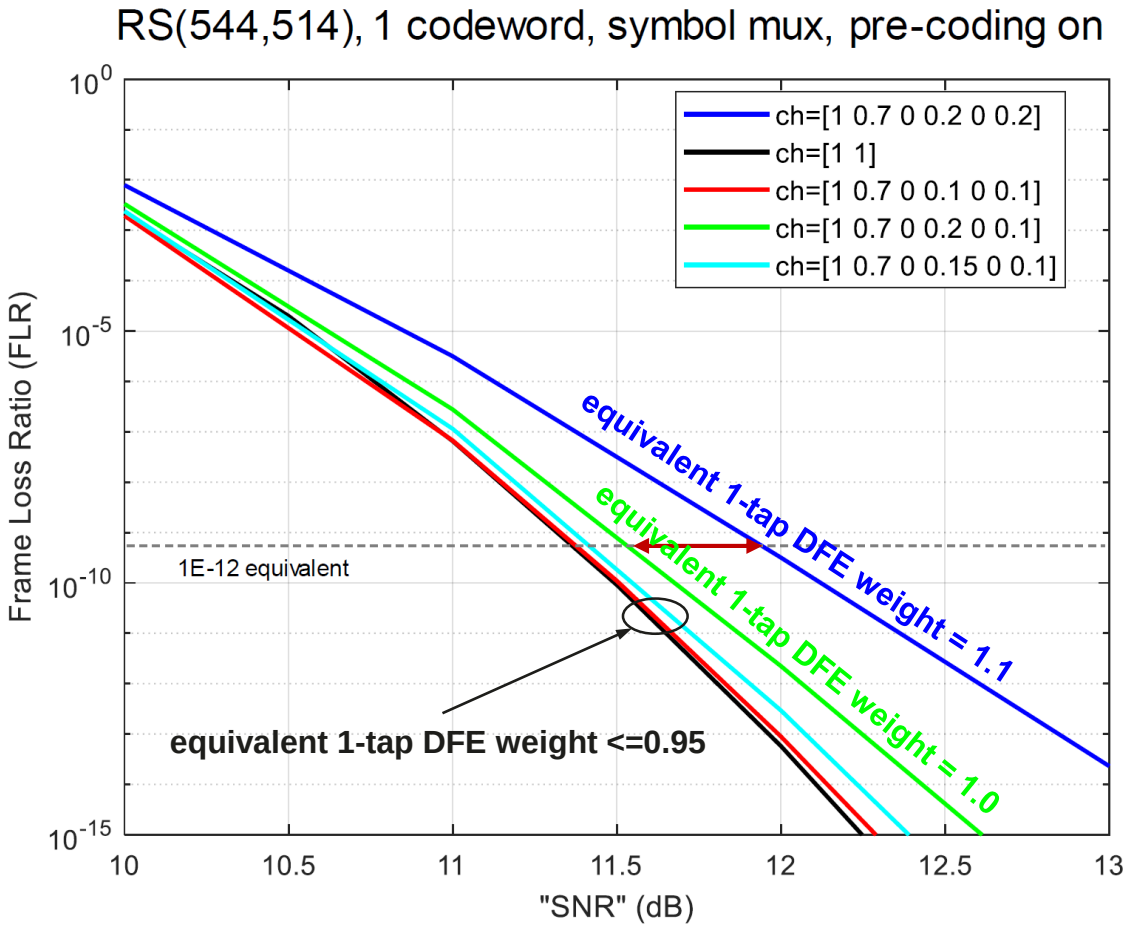
1 FEC + 1:1 bit mux (symbol mapping) ≈ 2-way interleaved FEC + 2:1 bit mux.

Same as "50G CR-2/KR-2"; Better than "25G CR-1/KR-1", "50G CR-1/KR-1" and "100G CR-2/KR-2".

gustlin_3ck_01_1118

HUAWEI

Symbol mapping can address the FEC performance concern



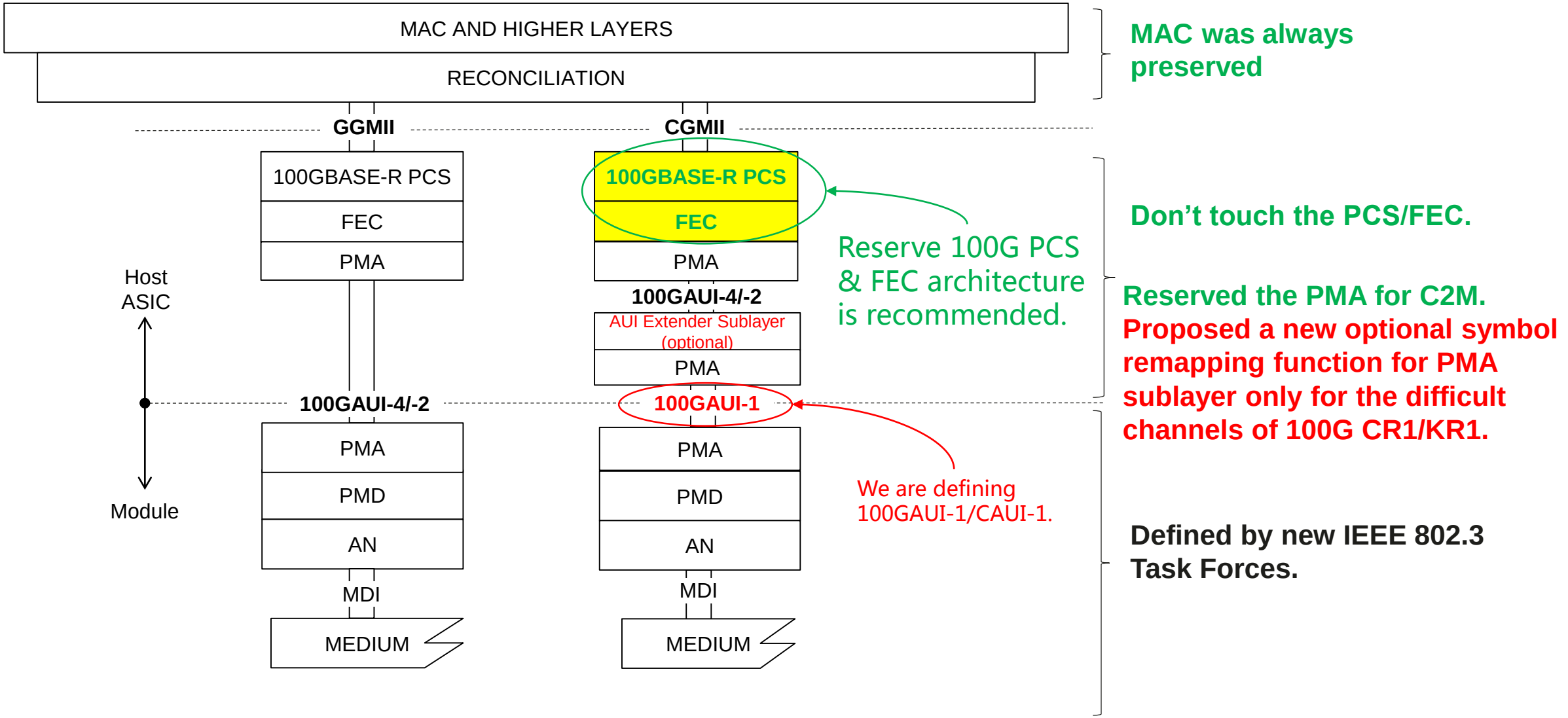
lyubomirsky_3ck_01a_0319 page 9.

Channel	Equivalent 1-tap DFE weight	"SNR" penalty (dB)
[1 1] (Reference)	**1**	**0**
[1 0.7 0 0.20 0 0.2]	**1.1**	**0.65**
[1 0.7 0 0.20 0 0.1]	**1.0**	**<0.2**
[1 0.7 0 0.15 0 0.1]	**0.95**	**<0.1**
[1 0.7 0 0.10 0 0.1]	**0.9**	**<0.05**

- Changing DFE weights from
  [1 0.7 0 0.2 0 **0.1**] (**equivalent DFE weight = 1.0**) to
  [1 0.7 0 0.2 0 **0.2**] (**equivalent DFE weight = 1.1**)
  contributes >**0.4dB** of the SNR penalty.
- As long as **equivalent DFE weight is constrained to be <=1.0 (DFE error propagation probability <0.75)**, the SNR penalty for "symbol mux" is minor.
- The Equivalent 1-tap DFE weight is <**0.8**, even with bmax=[0.85, 0.35] constrain (wu_3ck_01b_0319 page 21).
- Equivalent 1-tap DFE weight for multi-tap DFE was defined in lu_3ck_adhoc_01a_010219.

HUAWEI

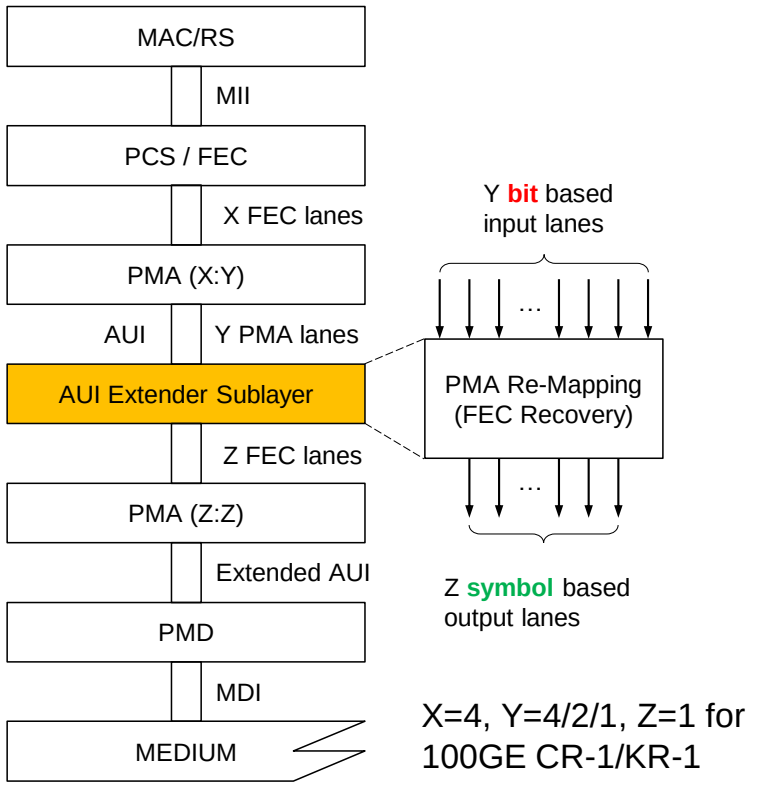
How to introduce “Symbol Mapping”

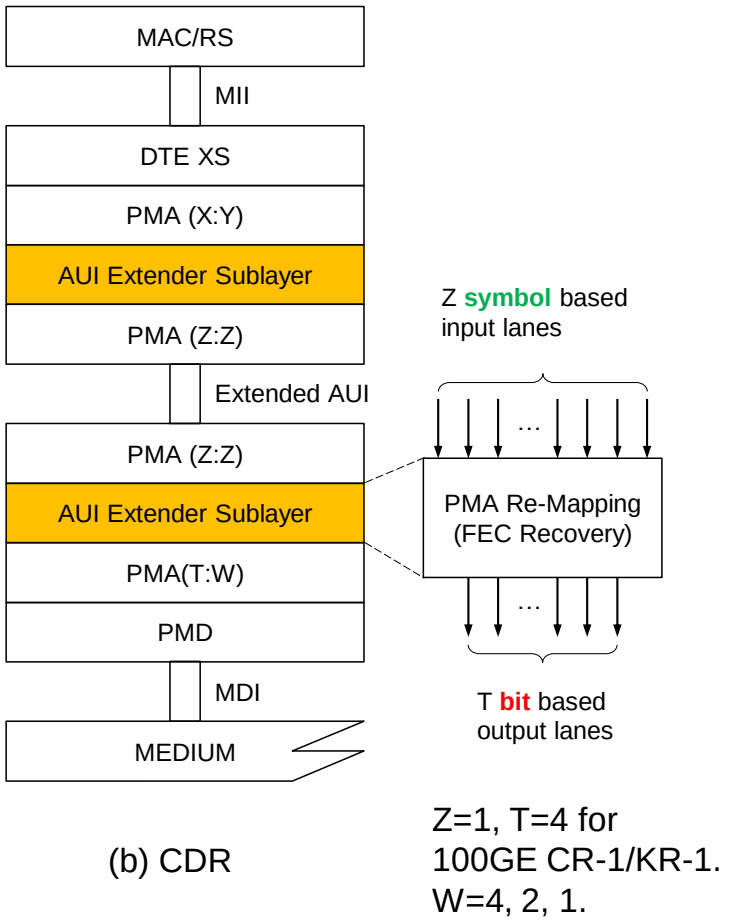


MII: Medium Independent Interface
AUI: Attachment Unit Interface

HUAWEI

Optional 100GE AUI Extender Sublayer

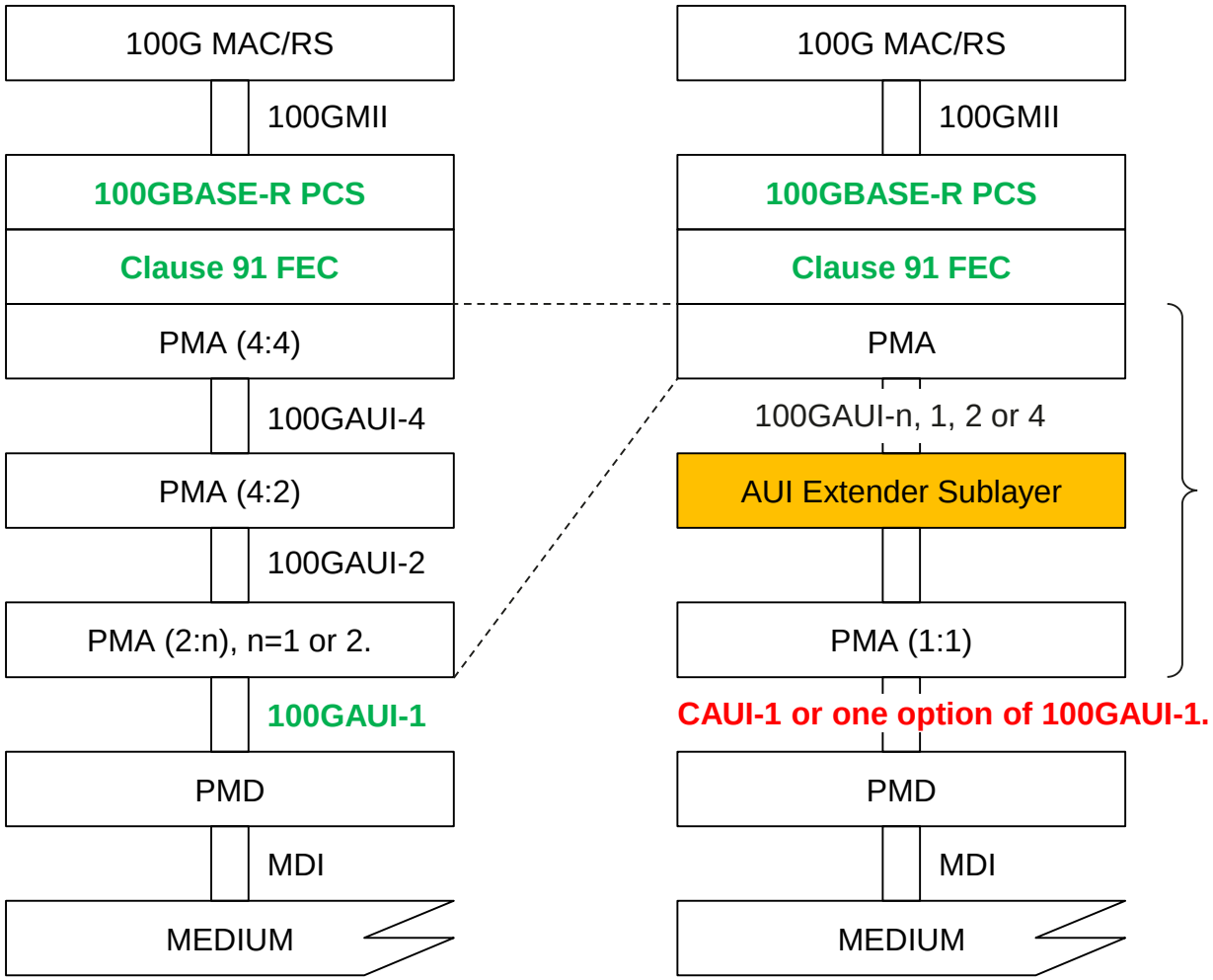




- Introduce a new optional AUI Extender Sublayer (XS) to extend the definition of AUI interface.
- AUI XS achieves PMA remapping function when necessary.
- The FEC performance concern can be resolved by reversing 4:1 bitmux and using symbol mapping for distribution.
- No latency or complicated CDR will be introduced.
- It can also support “FEC recovery” if needed.

HUAWEI

Proposed 100GE AUI Extender Sublayer architectural view



100GBASE-DR/KR/CR, C2M/C2C Interface.

Reserve the 100GE PCS FEC. No interoperability and compatibility issue.

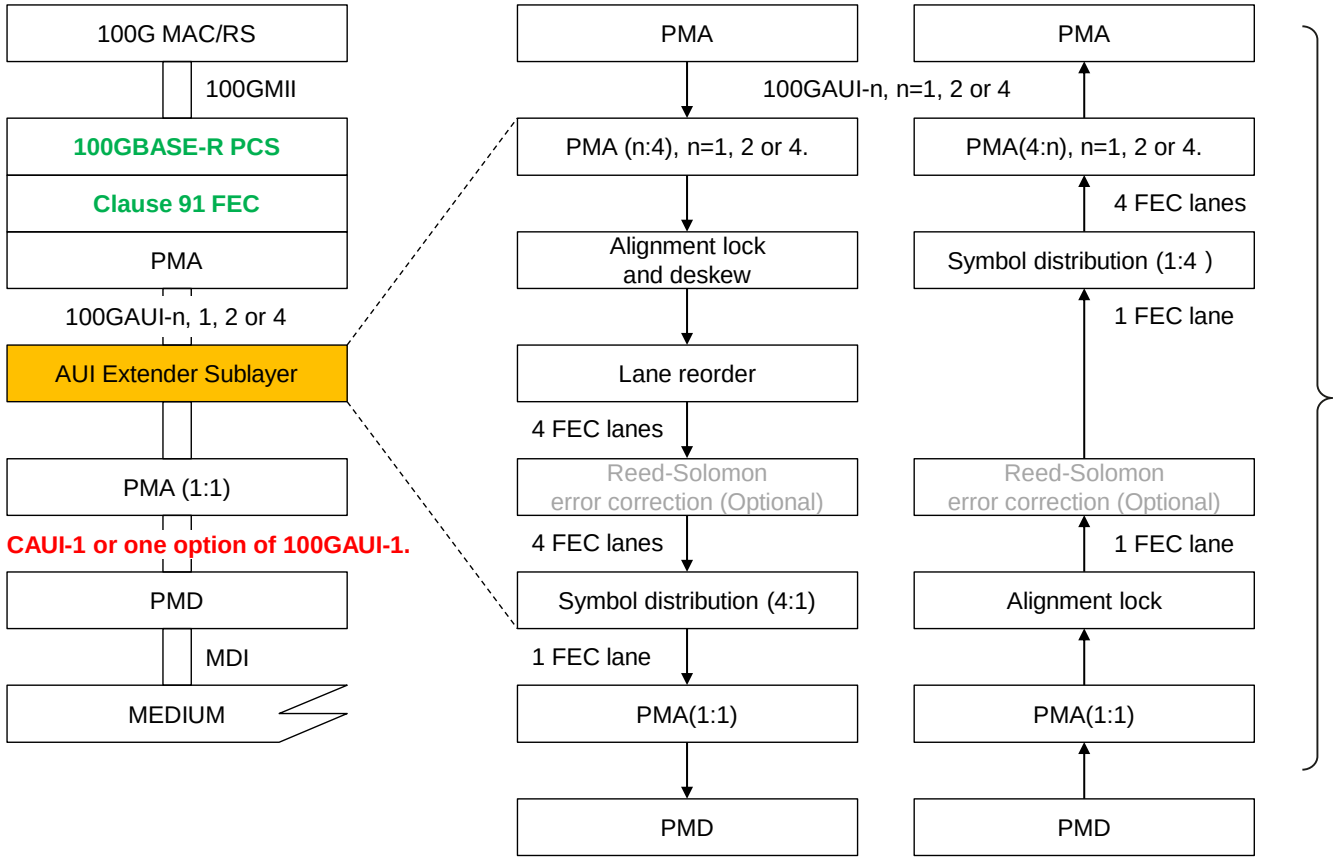
Bitmux PMA architecture is also reserved.
AUI XS is a shim layer within PMA sublayer.
Map 4 FEC lanes to 1 FEC lane.

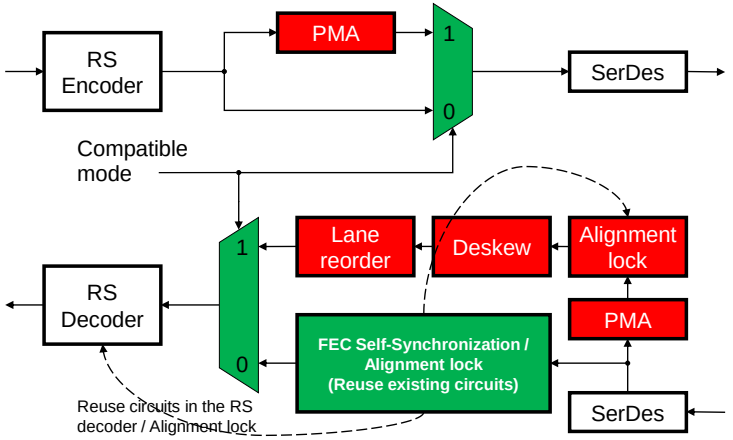
1. Can be stacked in Host ASIC.
2. Should be optional and only work for the most difficult channels at 100GE-CR1/KR1. Just like “Precoding”.

Bitmux PMA is not necessary for 1 PMD lane cases, because there are no “Skew” or “Out-of-order” issues.

HUAWEI

Proposed 100GE AUI Extender Sublayer detailed view

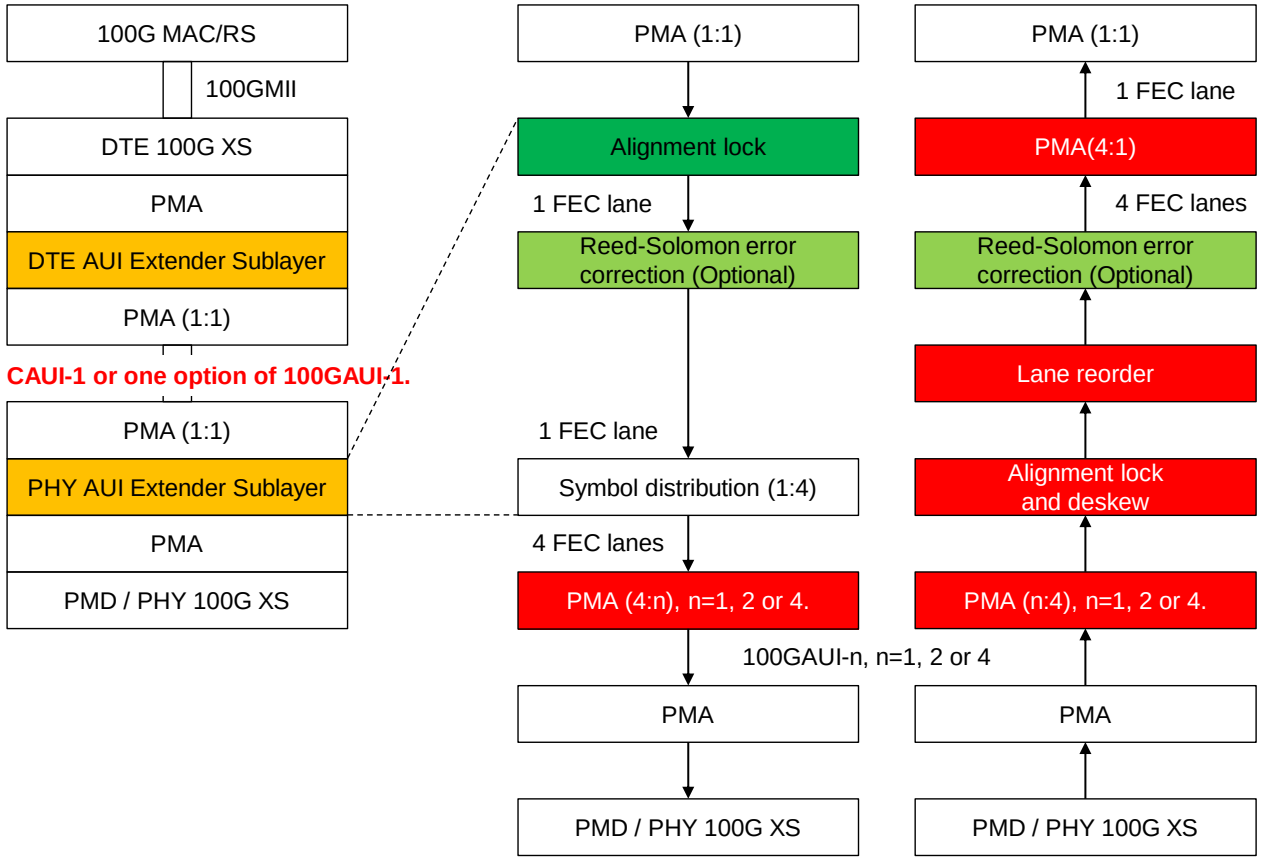


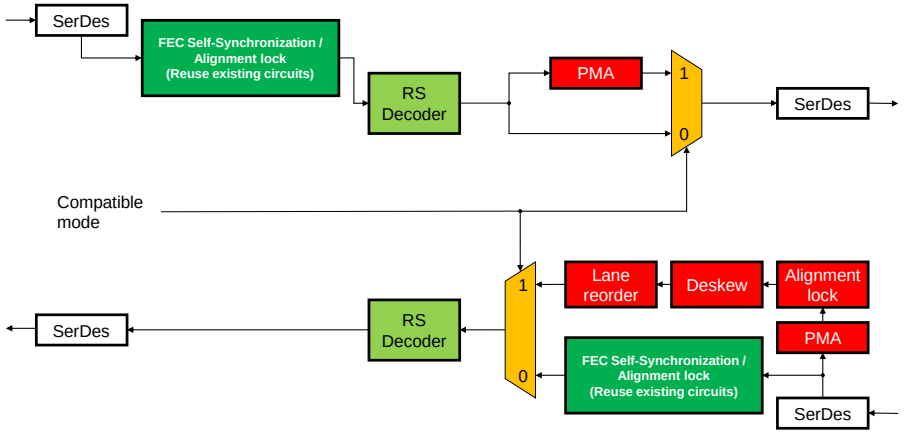


1. No need to implement in Host ASIC. Collapses with PCS and PMA for collocated Sublayers.
2. The "**red blocks**" are for the "Bit mux PMA", not needed for "Symbol mux PMA".
3. Only 2:1 selectors are needed to support dual model. No latency will be introduced.

HUAWEI

Proposed 100GE AUI Extender Sublayer detailed view





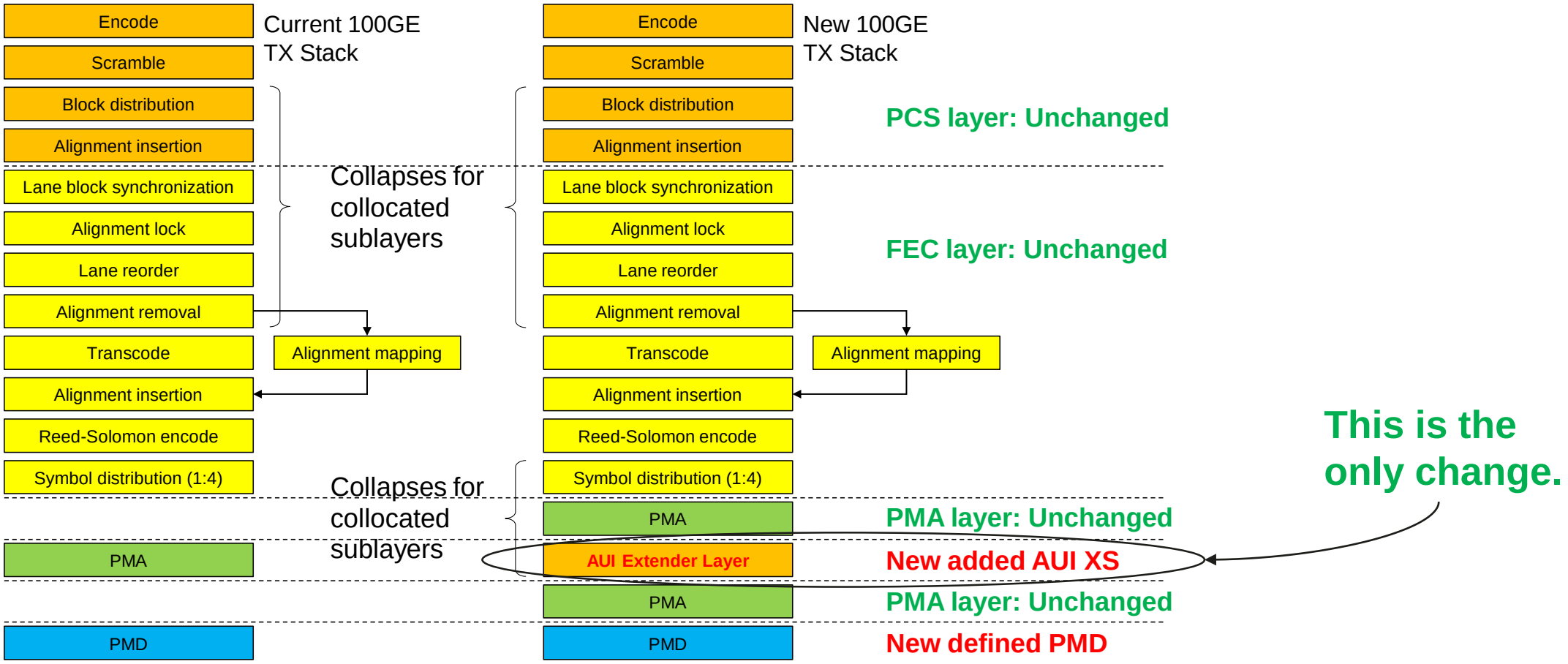
1. Small cost of implementation in module. Negligible cost if compared with interleaved FEC scheme. No latency will be introduced if “FEC recovery” is turned off.

2. The complexity comes from the support of “Bit mux PMA”, i.e. the “**red blocks**”.

3. No need to implement the “**red blocks**” for symbol mapping PMA or “protocol independent CDR”.

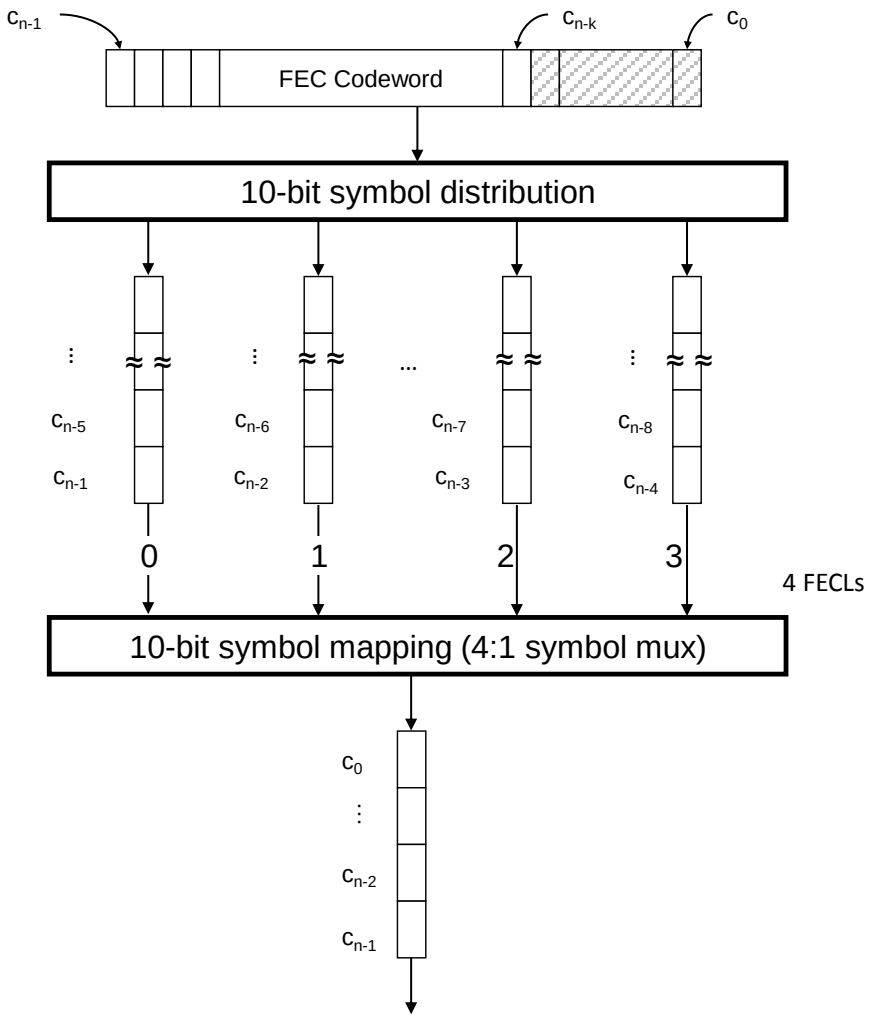
HUAWEI

Protocol stack comparison with current 100GE

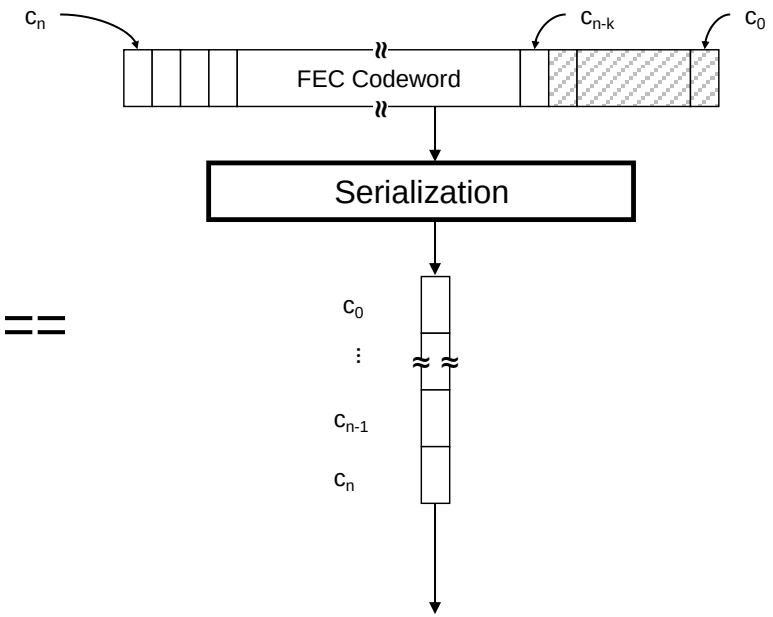


HUAWEI

Symbol mapping relationships



==

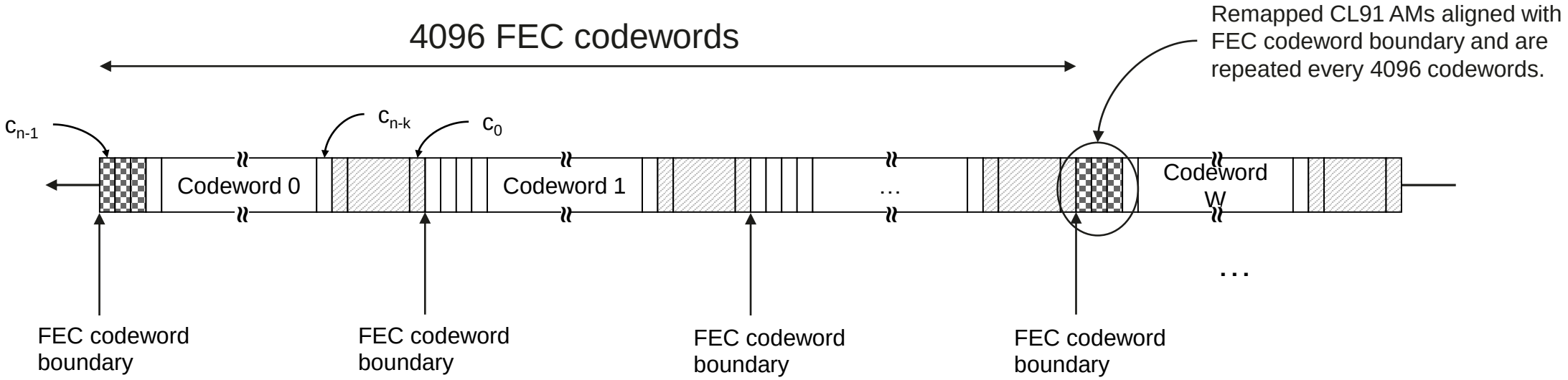


"Keep all the FECLs in order and use 4:1 symbol mux." is equivalent to direct "symbol mapping" from the FEC codeword to the 1 PMD lane.

HUAWEI

FEC self-synchronization and Alignment Markers

- No new Alignment Markers (AM) are needed to be defined. Two ways to achieve alignment
  1. The RS(544, 514) can be self-synchronized.
  2. Reuse the remapped alignment markers of Clause 91.



The CL91 Alignment Markers are aligned with RS(544, 514) FEC boundary. As long as the FEC boundary is founded by the self-synchronization algorithm, the FECLs can be easily recovered.

FEC self-synchronization and Alignment Markers

- No new Alignment Markers (AM) are needed to be defined. Two ways to achieve alignment
  1. The RS(544, 514) can be self-synchronized.
  2. Reuse the remapped alignment markers of Clause 91.

amp_tx_x={M_0, M_1, M_2, BIP_3, M_4, M_5, M_6, BIP_7}

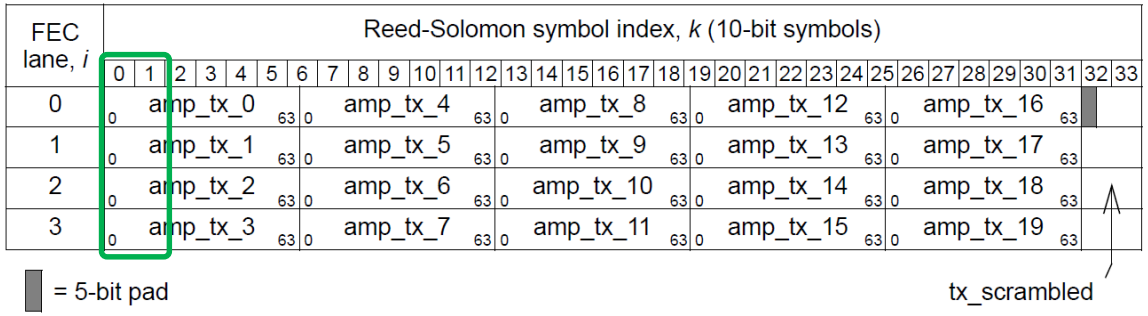


Figure 91–4—Alignment marker mapping to FEC lanes

An example of 80bits remapped AM is
amp_tx_0{ 9: 0}, amp_tx_1{ 9: 0}, amp_tx_2{ 9: 0}, amp_tx_3{ 9: 0}, amp_tx_0{19:10}, amp_tx_1{19:10}, amp_tx_2{19:10}, amp_tx_3{19:10}.
These bits are aligned with RS(544, 514) FEC codeword and repeat every 4096 FEC code words.

More bits are available for alignment except for the BIP bits.

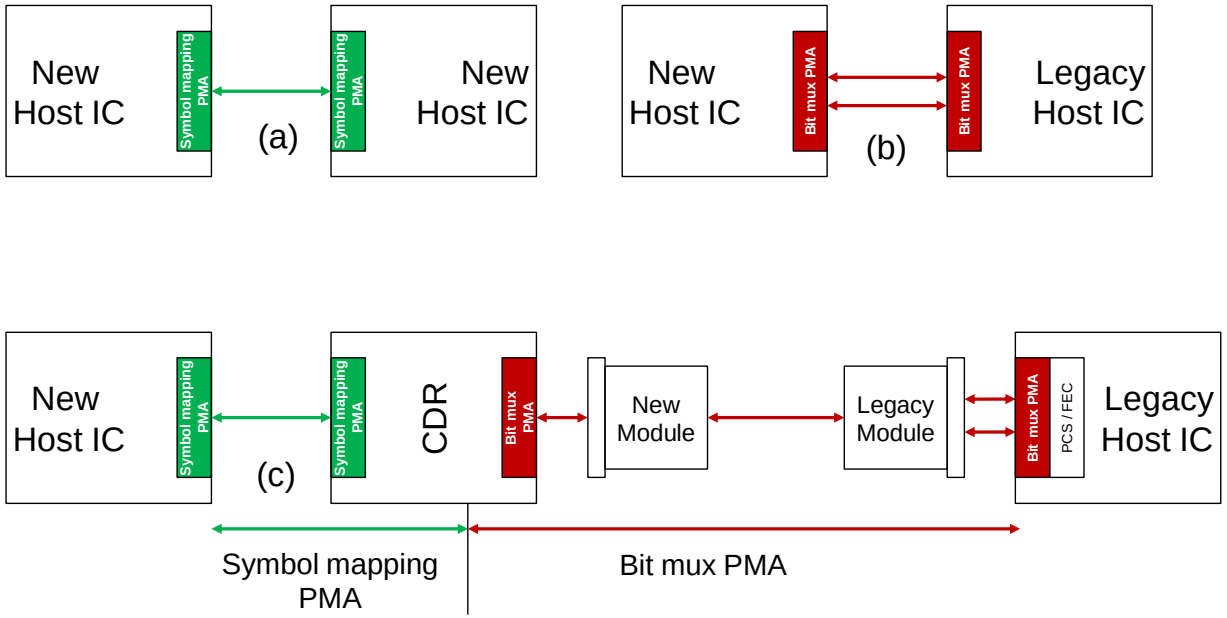
Table 82–2—100GBASE-R Alignment marker encodings

PCS lane number	Encoding[a] {M_0, M_1, M_2, BIP_3, M_4, M_5, M_6, BIP_7}	PCS lane number	Encoding[a] {M_0, M_1, M_2, BIP_3, M_4, M_5, M_6, BIP_7}
0	0xC1, 0x68, 0x21, BIP_3, 0x3E, 0x97, 0x1E, BIP_7	10	0xFD, 0x6C, 0x99, BIP_3 0x02, 0x93, 0x66, BIP_7
1	0x9D, 0x71, 0x8E, BIP_3, 0x62, 0x8E, 0x71, BIP_7	11	0xB9, 0x91, 0x55, BIP_3, 0x46, 0x6E, 0xAA, BIP_7
2	0x59, 0x4B, 0xE8, BIP_3, 0xA6, 0xB4, 0x17, BIP_7	12	0x5C, 0x B9, 0xB2, BIP_3, 0xA3, 0x46, 0x4D, BIP_7
3	0x4D, 0x95, 0x7B, BIP_3, 0xB2, 0x6A, 0x84, BIP_7	13	0x1A, 0xF8, 0xBD, BIP_3, 0xE5, 0x07, 0x42, BIP_7
4	0xF5, 0x07, 0x09, BIP_3, 0x0A, 0xF8, 0xF6, BIP_7	14	0x83, 0xC7, 0xCA, BIP_3, 0x7C, 0x38, 0x35, BIP_7
5	0xDD, 0x14, 0xC2, BIP_3, 0x22, 0xEB, 0x3D, BIP_7	15	0x35, 0x36, 0xCD, BIP_3, 0xCA, 0xC9, 0x32, BIP_7
6	0x9A, 0x4A, 0x26, BIP_3, 0x65, 0xB5, 0xD9, BIP_7	16	0xC4, 0x31, 0x4C, BIP_3, 0x3B, 0xCE, 0xB3, BIP_7
7	0x7B, 0x45, 0x66, BIP_3, 0x84, 0xBA, 0x99, BIP_7	17	0xAD, 0xD6, 0xB7, BIP_3, 0x52, 0x29, 0x48, BIP_7
8	0xA0, 0x24, 0x76, BIP_3, 0x5F, 0xDB, 0x89, BIP_7	18	0x5F, 0x66, 0x2A, BIP_3, 0xA0, 0x99, 0xD5, BIP_7
9	0x68, 0xC9, 0xFB, BIP_3, 0x97, 0x36, 0x04, BIP_7	19	0xC0, 0xF0, 0xE5, BIP_3, 0x3F, 0x0F, 0x1A, BIP_7

[a]Each octet is transmitted LSB to MSB.

HUAWEI

Scenarios of “Symbol Mapping PMA”

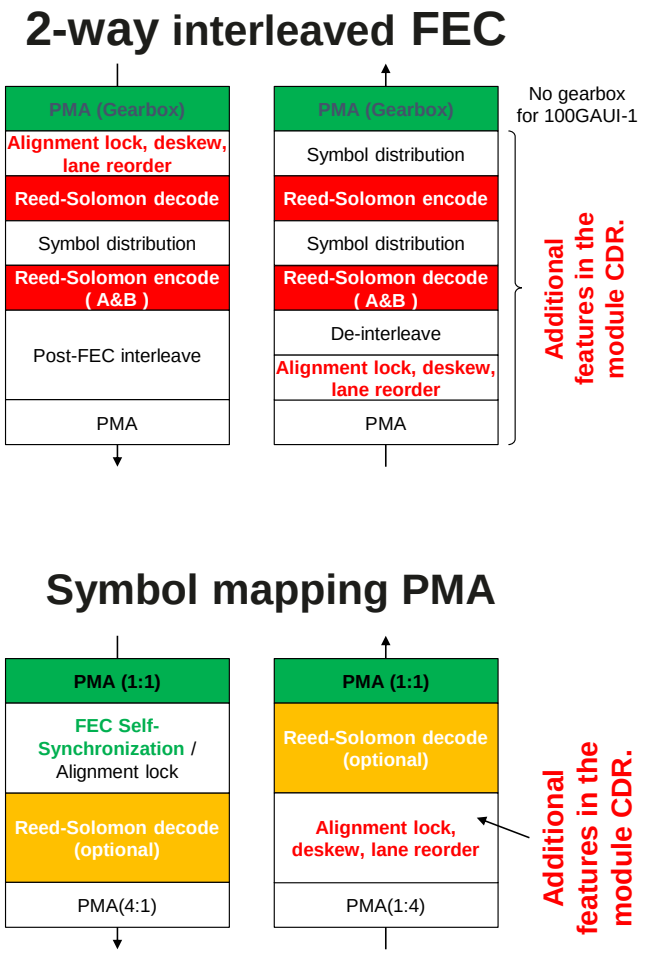


1. New chip can talk to new chip with “symbol mapping PMA” and “bit mux PMA”.
2. New chip can talk to legacy chip with “bit mux PMA”.
3. Host IC can support dual modes, i.e. “symbol mapping PMA” and “bit mux PMA” with negligible cost.
4. CDR chips can achieve the “Symbol mapping PMA” and “Bit mux PMA” conversion, i.e PMA-remapping with affordable cost. The complexity mainly comes from “Bit mux PMA”.
5. Only simplex PMA-remapping is required. Even though the “host side” and “line side” Serdes are different, circuits can be shared.

HUAWEI

“Symbol mapping” vs. “2-way interleaved FEC+2:1 bit mux”

#		“Symbol mapping PMA” (similar to 25GE)	“2-way interleaved FEC + 2:1 bit mux” (gustlin_3ck_01_0119)
Performance		**Same** (“Symbol mapping PMA” may need some reasonable constrain in DFE weights).	
Complexity	**Host IC**	160~320 “2:1 selector”	160~320 “2:1 selector” **2x 50G RS(544, 514) Encoder/Decoder.**
	CDR	1x FECL processing (PMA Gearbox, Alignment lock, deskew, lane reorder), **not needed for “symbol mapping PMA”**. 2x 100G RS(544, 514) Decoder, **not needed w/o “FEC recovery” support**. **All the above functions are optional and can be by-passed.**	2x FECL processing (PMA Gearbox, Alignment lock, deskew, lane reorder) 2x 50G RS(544, 514) Encoder/Decoder. 1x 100G RS(544, 514) Encoder/Decoder. **All the above functions are mandatory.**
Latency Increase	**Host IC**	**0**	**>50ns**
	CDR	**0ns w/o “FEC recovery” support;** ~100ns w/ FEC recovery support.	**>150ns 1 CDR; >250ns 2 CDR.**
Protocol independent CDR support		**Yes, FEC can be self-synchronized, no need to identify the AMs.**	**No, need to process the PCS to support “FEC Recovery”.**
Newly defined Alignment Markers		**No**	**Yes**



HUAWEI

Summary

This presentation shows a possible PMA Sublayer solution to mitigate the FEC performance concern of difficult 100GE-CR1/KR1 channels.

No mandatory new feature is needed. No latency and complicated CDR will be introduced. “Protocol independent CDR” with “FEC recovery” is supported.

Thank you !

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home, and organization for a fully connected, intelligent world.