Analysis of Potential Solutions for 100G CR/KR DFE Error Propagation

Yuchun Lu, Huawei
Yan Zhuang, Huawei

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

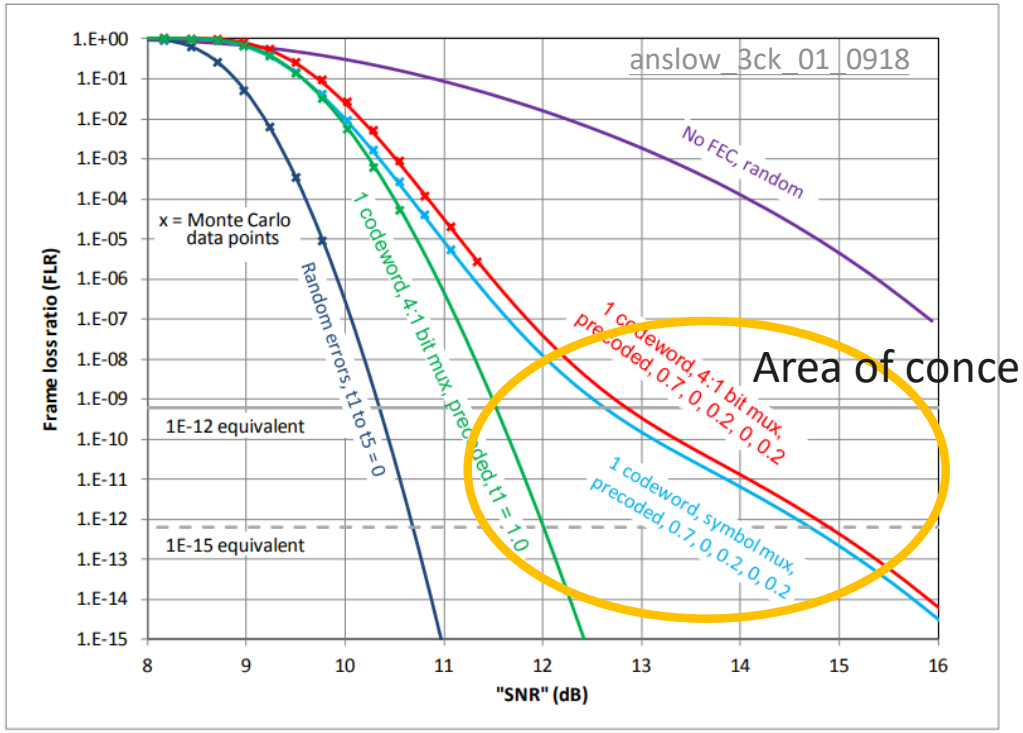
HUAWEI

Introduction

- **Performance concern of 100G CR/KR (4:1 bit mux PMA).**
- **Potential solutions to address the FEC performance concern.**
  1. PMD Sublayer solutions
     A. Use the mainstream n-tap FFE + 1-tap DFE receiver.
     B. Constrain the equivalent DFE weight to 0.7~0.85 for n-tap DFE receivers.
     C. Use enhanced precoding (v2.0) to eliminate the DFE error propagation (lu_3ck_01_0319).
  2. PMA Sublayer solution: **Bit mux PMA →Symbol mapping PMA**.
  3. FEC Sublayer solution: **2-way interleaved FEC**. See gustlin_3ck_01_0119 and lu_3ck_adhoc_01_022719.
- **Analysis on each solution (performance & cost)**
- **Summary & AUI Extender Sublayer supporting PMA remapping.**

Performance concern of 100G CR/KR (from 4:1 bit mux).

100G 5-tap DFE results (0.7, 0, 0.2, 0, 0.2) with precoding

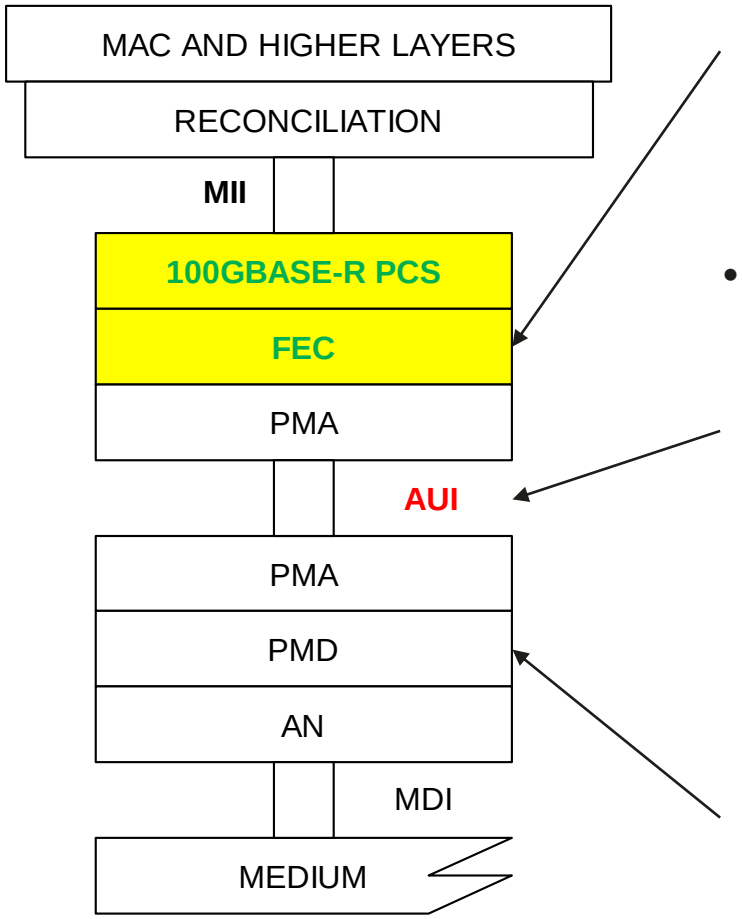


[0.7, 0, 0.2, 0, 0.2] is too pessimistic for DFE error propagation investigation.

1. 100G FEC performance concern was shown with multi-tap DFE burst errors, even with precoding and symbol mux. The concern comes from the **4:1 bit mux**.
2. Evaluated with DFE weight = [**0.7, 0, 0.2, 0, 0.2**], which is too pessimistic for DFE error propagation investigation.
3. The methodology of deriving worst case DFE weight for error propagation investigation was discussed, and worst case DFE weights were provided as below in (lu_3ck_adhoc_01a_010219).
   - **[0.78, 0.07, -0.01, 0.03, 0.02]**
   - **[0.87, 0.22, 0.09, 0.09, 0.06]**
4. Simulation results show that with DFE weights well constrained e.g. **[0.8 0.2 -0.05 0.05]**, the symbol-based solutions can address this FEC performance concern. (anslow_3ck_01_0119, also see page 6)

HUAWEI

Potential solutions



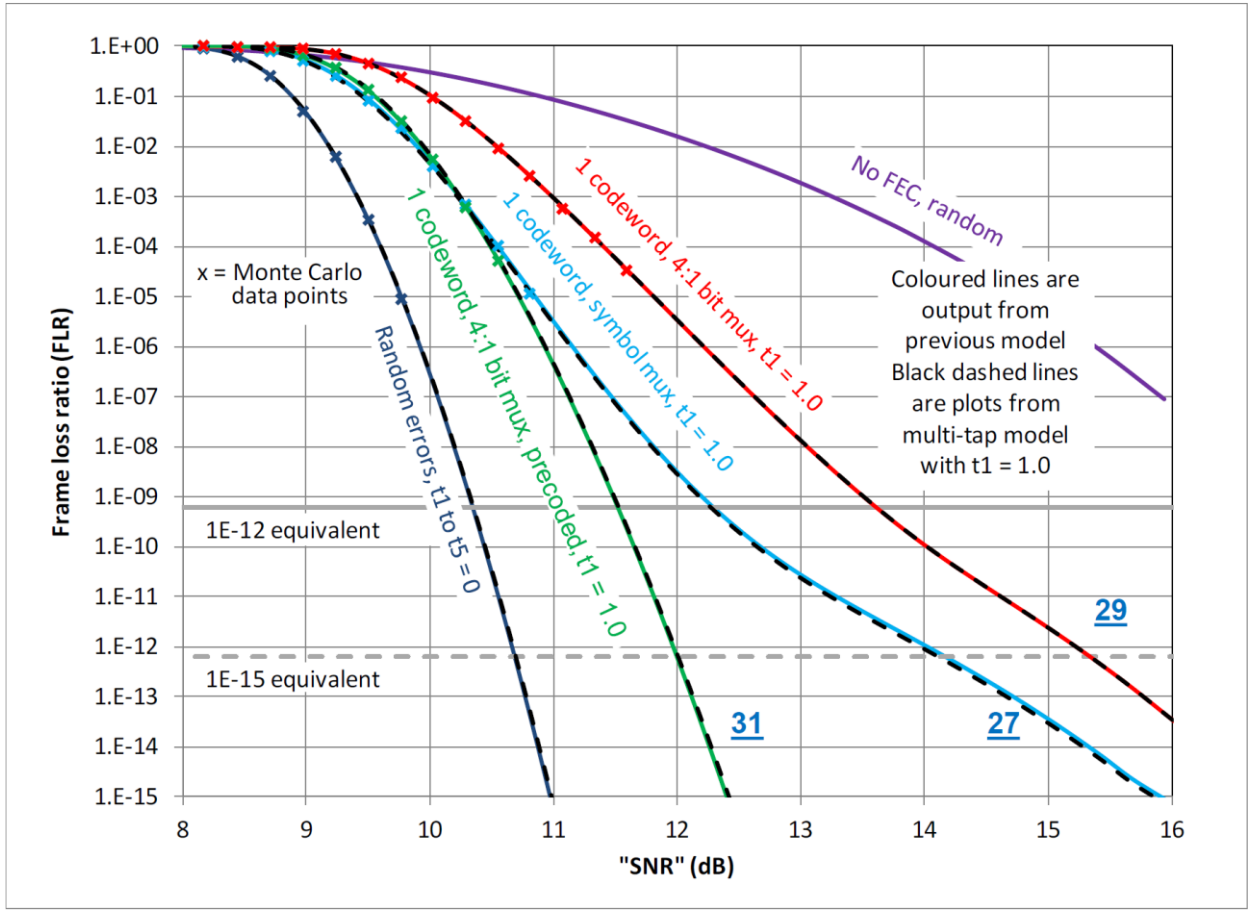
- Use new FEC (e.g. interleaved FEC) to mitigate potential burst error issues of multi-tap DFE receivers. ★
  - Analyzed in lu_3ck_adhoc_01_022719.

- Direct symbol mapping to the PMD lane will remove the impact of 4:1 bit-mux. Exactly the same as 2-lane 50GE; 25GE, 1-lane 50GE and 2-lane 100GE are even worse. ★★
  - Small impact on the standard and system.
  - Simplify CDR and can benefit more scenarios, such as link-segment OAM, FEC recovery in 100G/400G ZR.

- Use the mainstream n-tap FFE + 1-tap DFE receiver.
- Constrain the equivalent DFE weight to 0.7~0.8 for n-tap DFE receivers. ★★★
- Use precoding 2.0 to eliminate the DFE error propagation (lu_3ck_01_0319).
  - No impact on the standard and system.

HUAWEI

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.

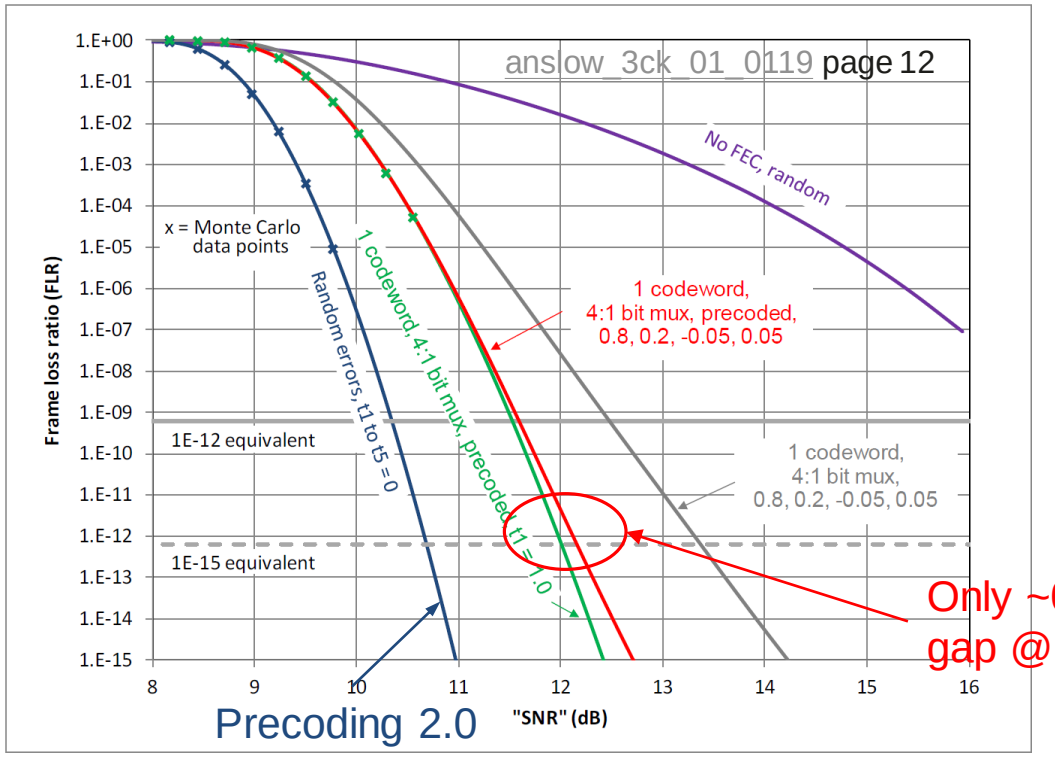
2. 1-tap DFE does not have FEC performance concern if pre-coding is applied, even with the worst case DFE weight t1=1.0 or error propagation probability a=0.75.

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)

HUAWEI

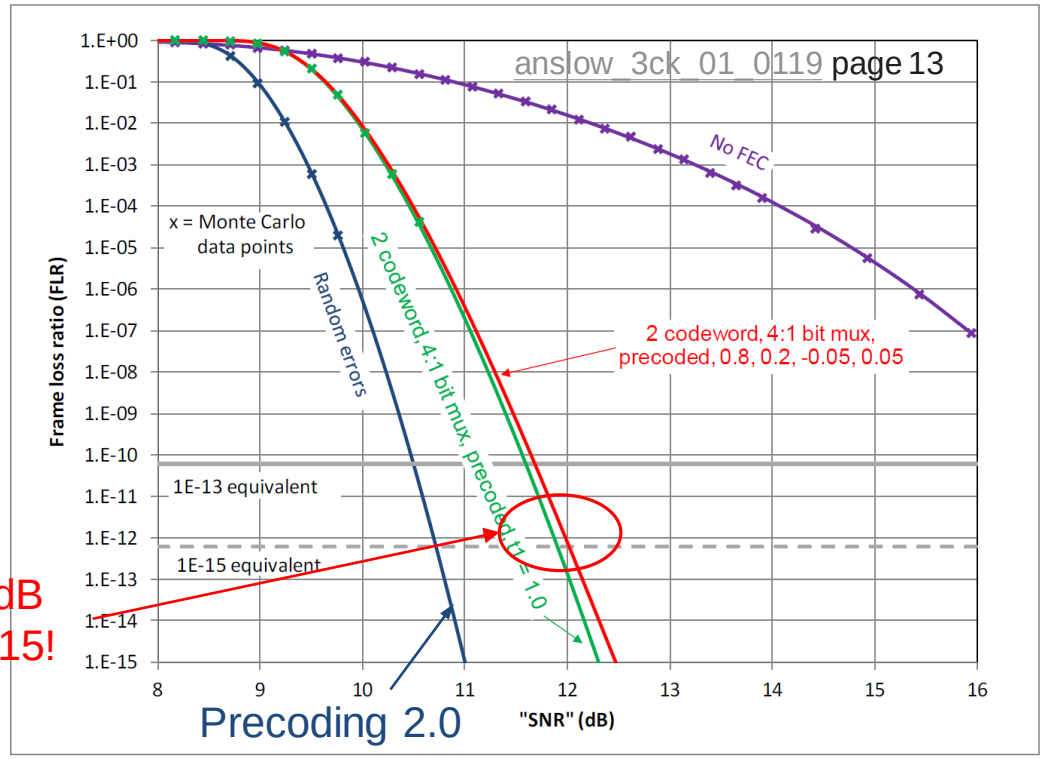
Constraining DFE weights can relief FEC performance concern

Non-interleaved FEC + 4:1 bit mux.



Only ~0.2dB gap @1e-15!

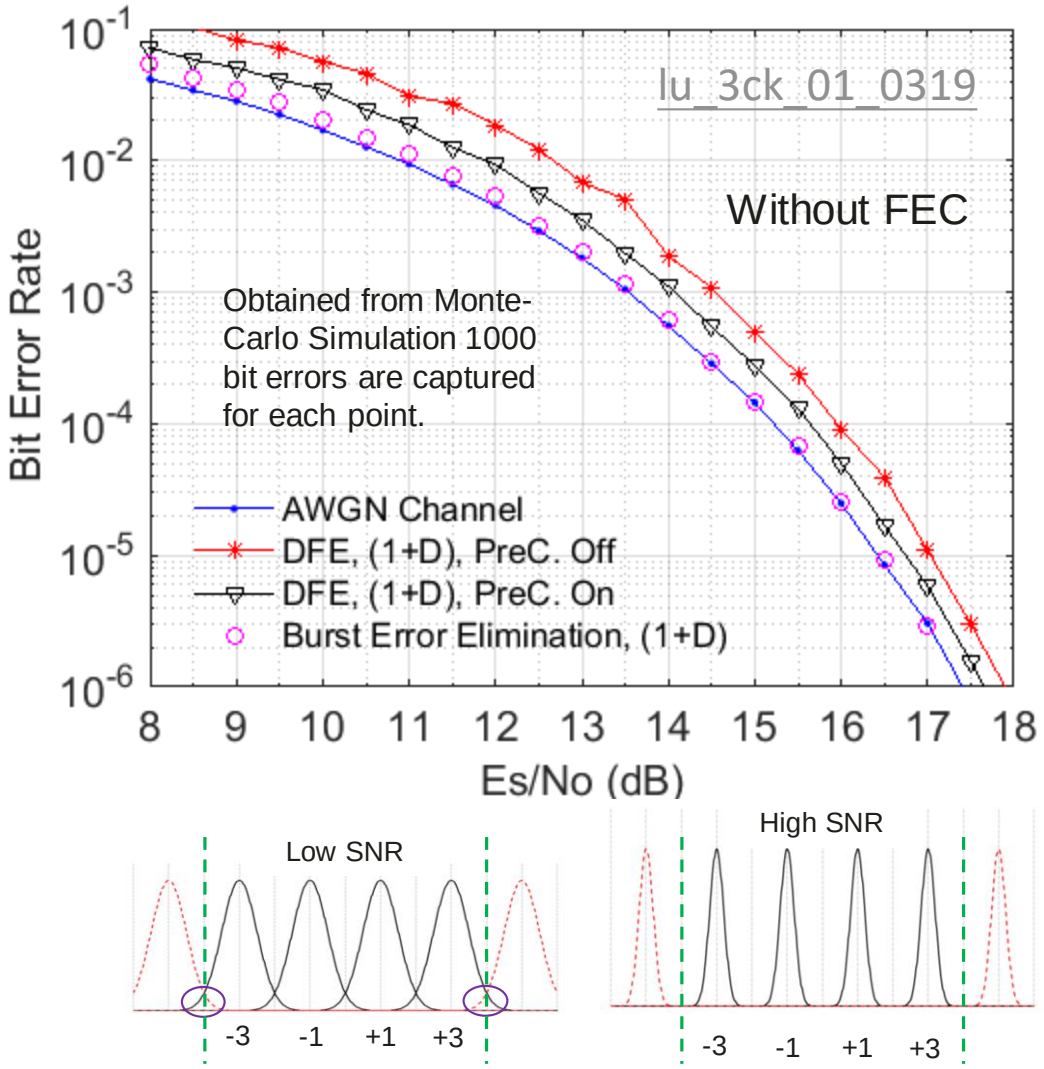
2-way interleaved FEC + 4:1 bit mux.

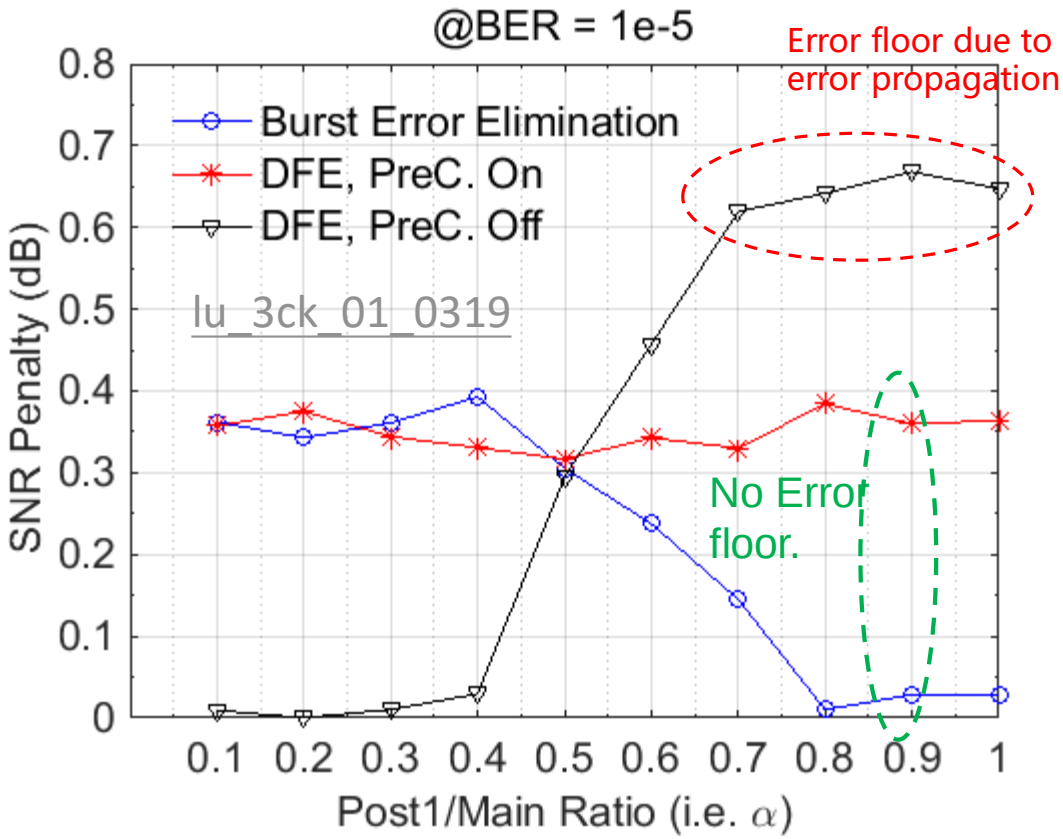


By constraining the DFE weight to **[0.8 0.2 -0.05 0.05]** the performance concern can be greatly relieved.
[0.8 0.2 -0.05 0.05] is a reasonable LR DFE weight setting which is close to the real worst case, bmax=[0.7, 0.2] was used in IEEE 802.3bj&cd, while relaxing "bmax" to [0.85, 0.35] is under discussion (wu_3ck_adhoc_01_022719).
If the 4:1 bit mux is replaced by symbol mapping, the FEC performance concern will be mitigated.

HUAWEI

Precoding 2.0 can remove DFE error propagation effect

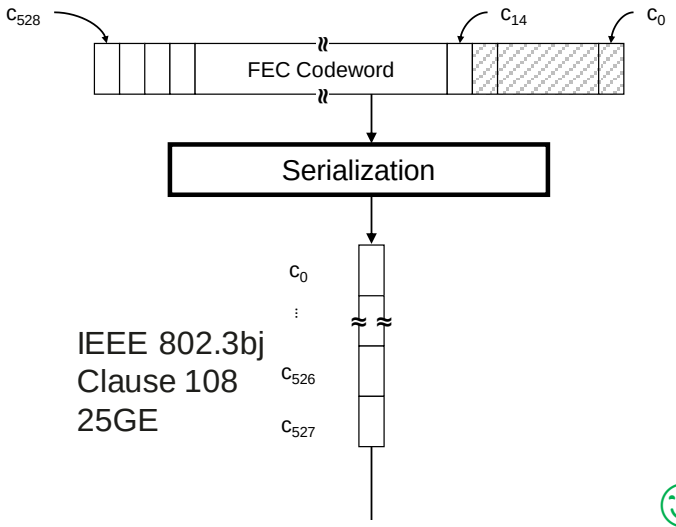




- The error propagation and post-FEC error floor can be eliminated by precoding 2.0.
- The penalty and complexity of precoding 2.0 is minor and negligible.

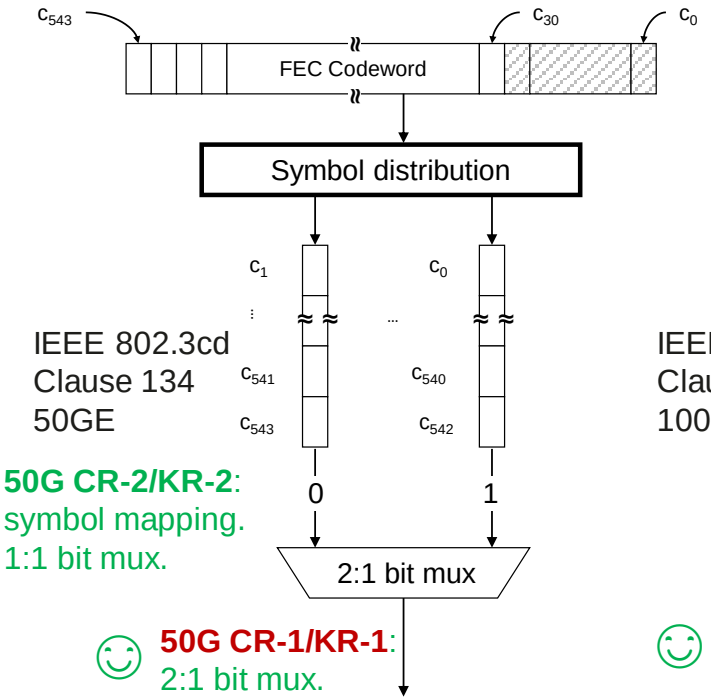
HUAWEI

Symbol mapping can relief FEC performance concern



IEEE 802.3bj
Clause 108
25GE

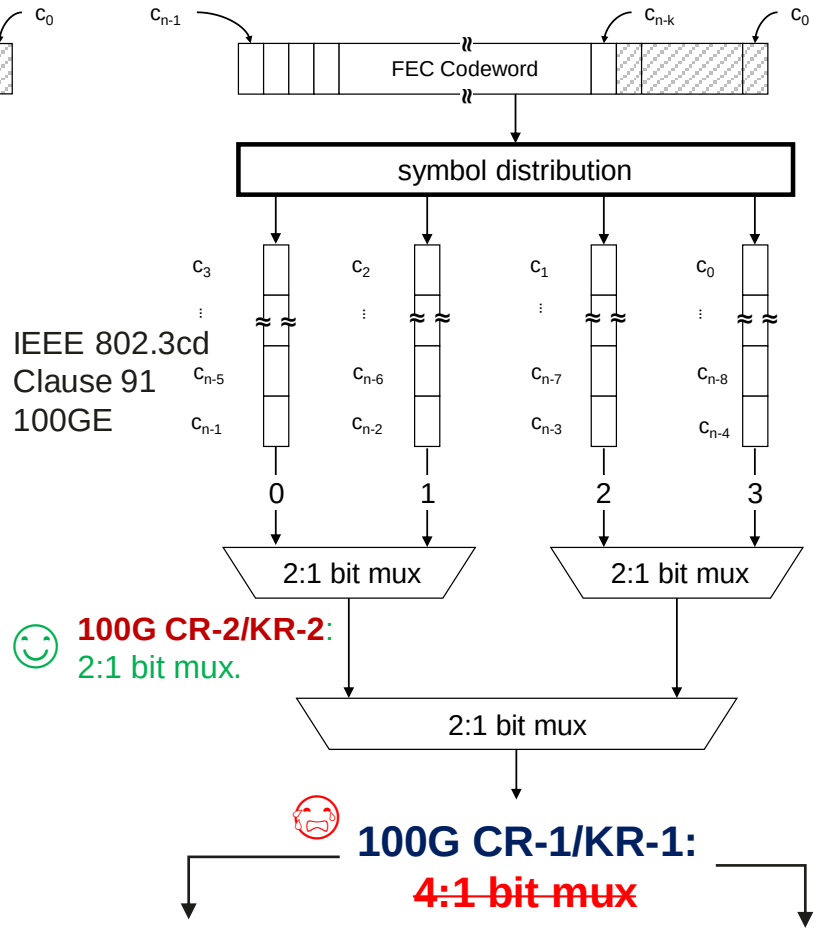
☺ **25G CR-1/KR-1**: direct codeword/symbol mapping. **RS(528, 514)** + 1:1 bit mux.



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. The FEC performance with bit mux PMA follows the order of "1:1 bit mux/symbol mapping" > "2:1 bit mux" > "**4:1 bit mux**".
2. "2-way interleaved FEC + 2:1 bit mux" (gustlin_3ck_01_1118) is equivalent to "1 FEC + 1:1 bit mux/symbol mapping".
3. **The error correction performance model does not scaled with the rate, thus the conclusions in IEEE 802.3bj, IEEE 802.3cd are still valid and should be consistent.**

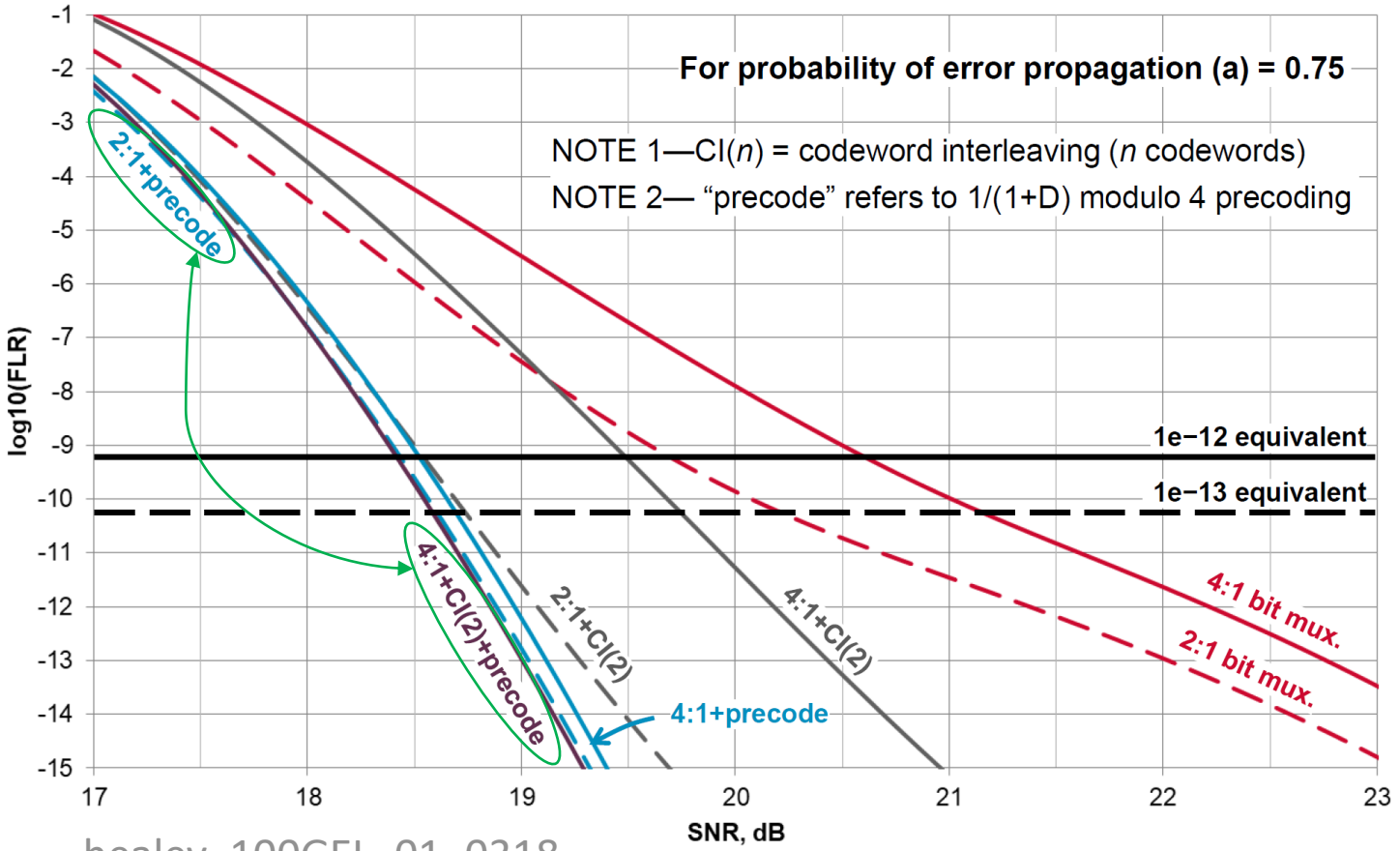
1 FEC + 1:1 bit mux (symbol mapping) ~= 2-way interleaved FEC + 2:1 bit mux.

gustlin_3ck_01_1118

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".

HUAWEI

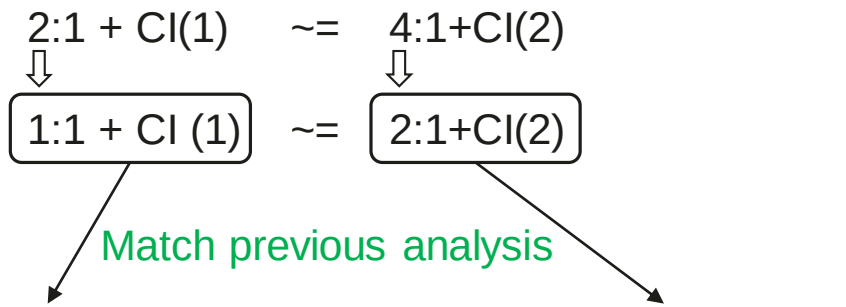
Symbol mapping can relief FEC performance concern



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:



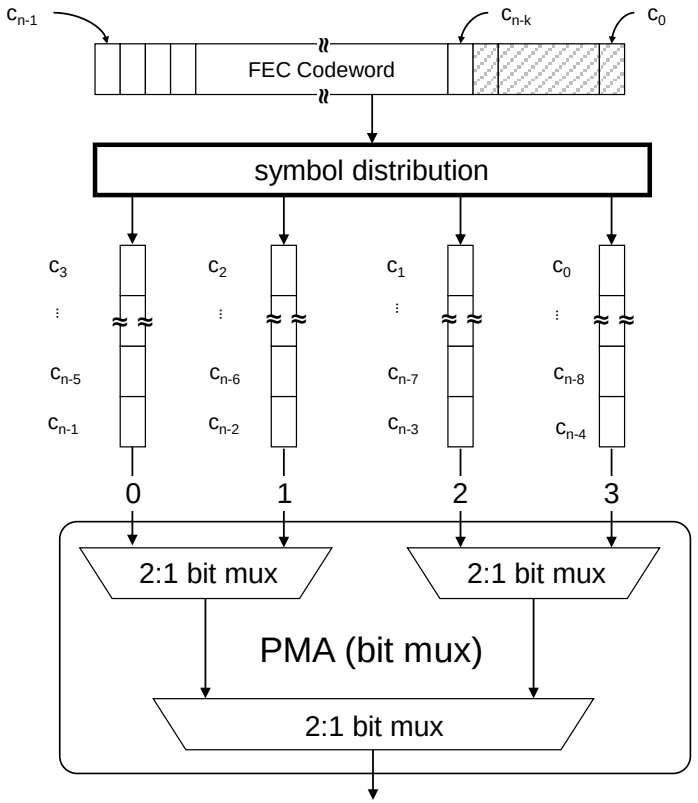
1 FEC + 1:1 bit mux (symbol mapping) ~= 2-way interleaved FEC + 2:1 bit mux.

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".

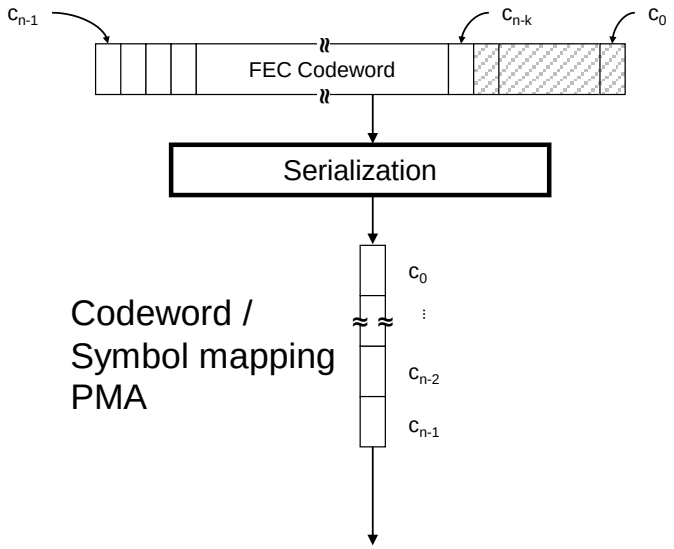
gustlin_3ck_01_1118

HUAWEI

Symbol mapping can be easily supported by host IC

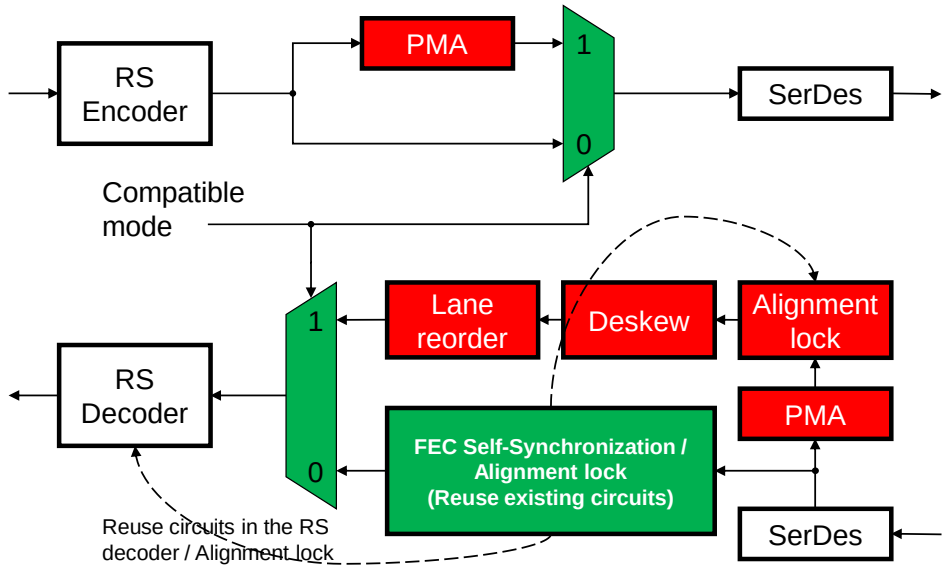


Compatible to legacy modules/host ICs. (Compatible mode)



New symbol mapping mode for new modules/host ICs. (Advanced mode)

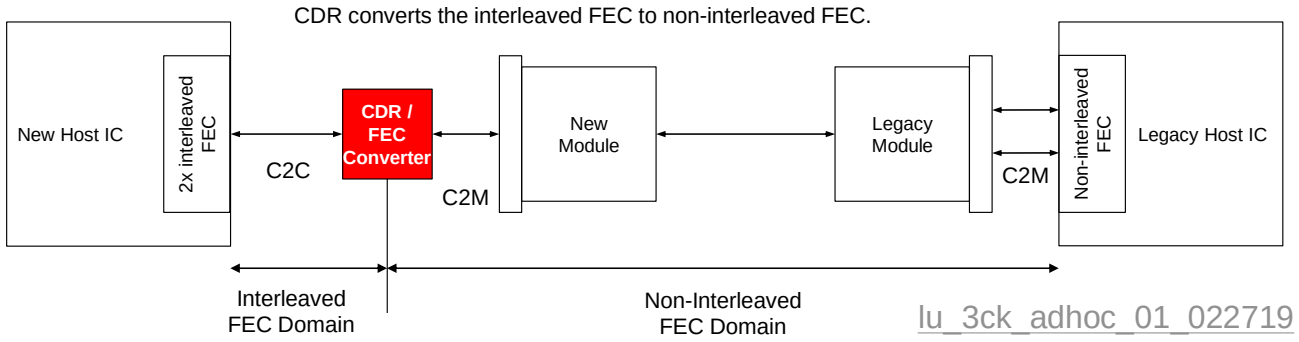
1. Relief the FEC performance concern due to 4:1 bit mux.
2. Simplify the CDR chip to support “PMA re-mapping”, “FEC recovery” & “segment-to-segment performance monitoring and fault location”.

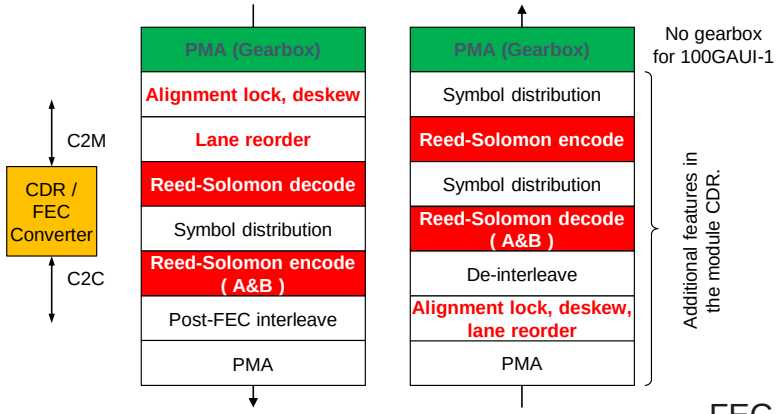


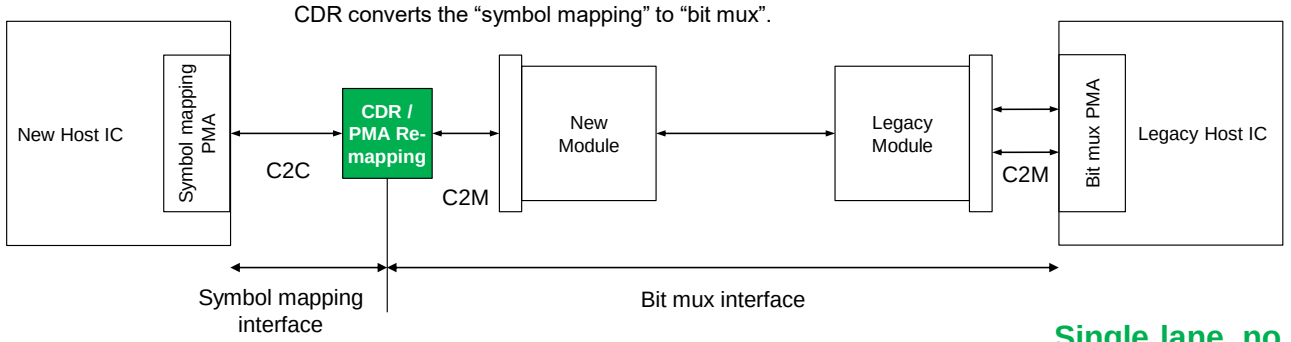
- The cost in chip area and latency of supporting a new PMA mapping mode in the host IC is minor.
- **Negligible chip area and latency cost in the data path.**
  1. Only “2:1 selectors” are needed. 160 ~ 320 “2:1 selectors” for 100G to achieve dual-mode design.
  2. No latency will be introduced.
- **Negligible chip area cost in the control path.**
  1. FEC synchronization can reuse the syndrome calculation circuits in the RS decoder.
  2. Remapping of the alignment markers (AMs) can make them reused. The “Alignment lock” circuits can be fully reused.

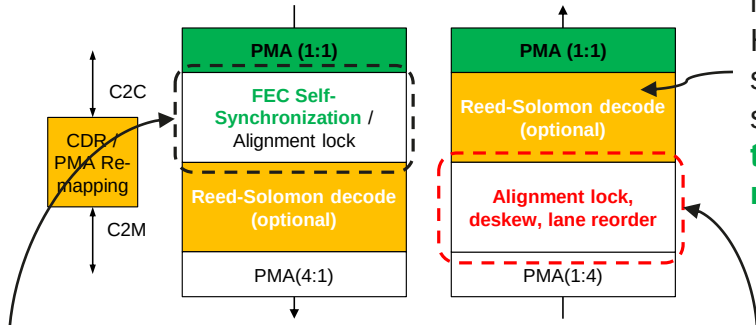
HUAWEI

Symbol mapping can have CDR greatly simplified









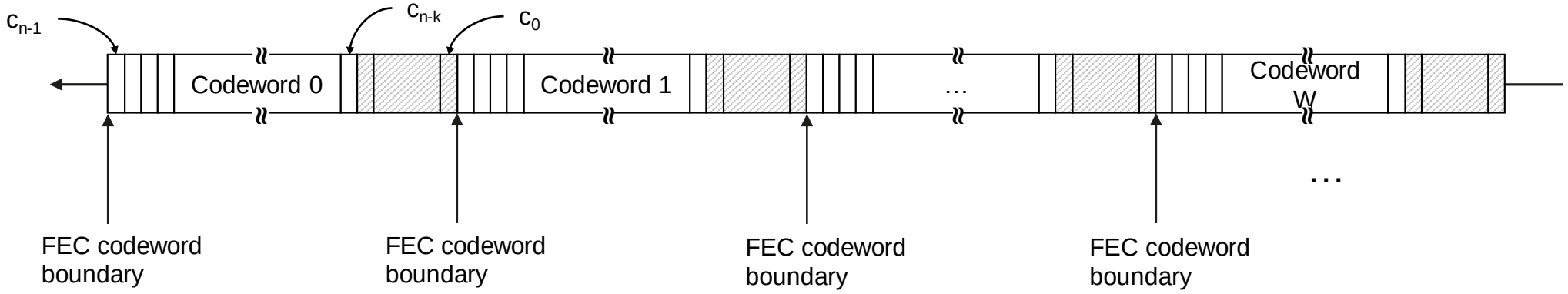
FEC decoders have already been integrated inside CDR chips. However the Ethernet standard does not support **protocol-transparency “FEC recovery”** feature.

Single lane, no skew, minor cost for FEC synchronization. The syndrome calculation circuits in the RS decoder can be reused.

The complexity mainly comes from the skew ambiguity @the legacy module side, due to the bit mux PMA. It can be shared between both directions.

HUAWEI

Symbol mapping can have CDR greatly simplified (Cont'd)

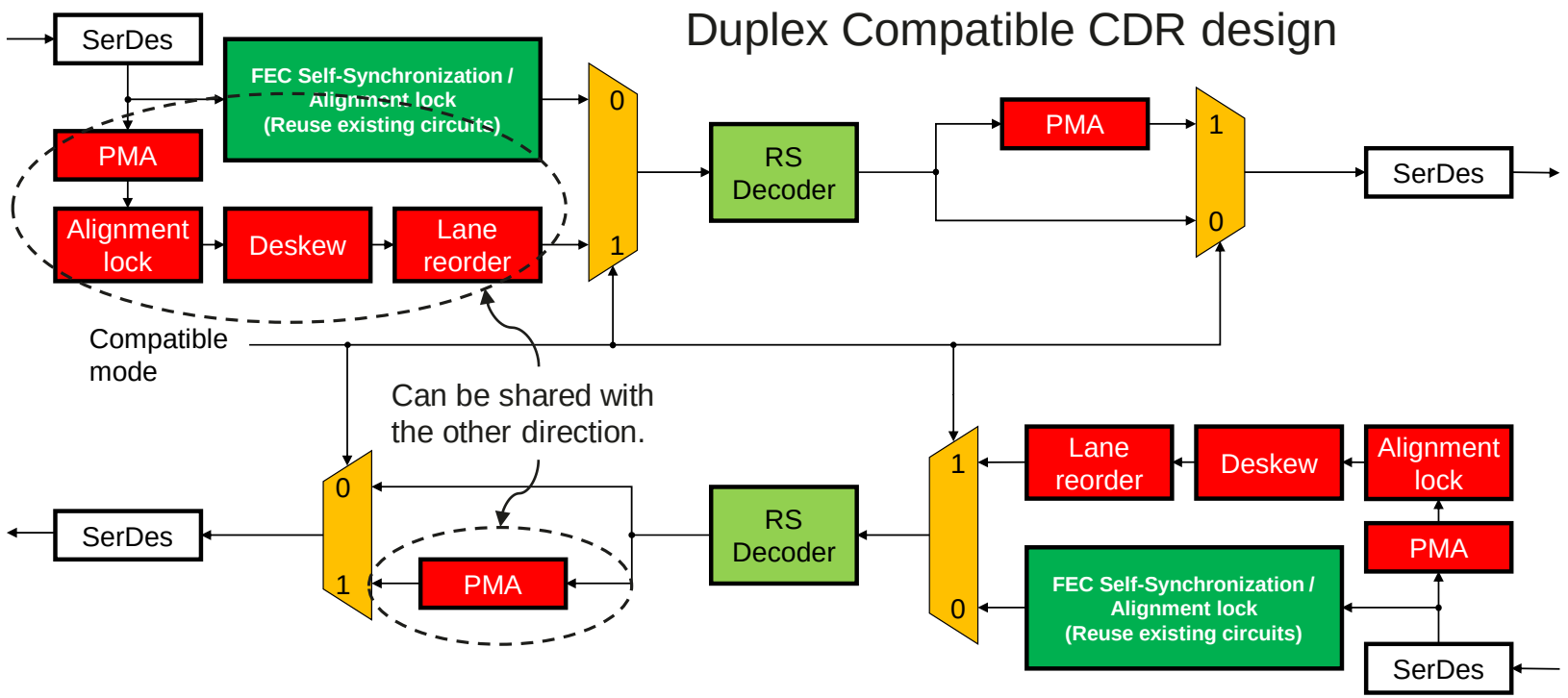


FEC can be self-synchronized! FEC codeword boundary is aligned with the alignment markers. As long as FEC codeword is aligned, the FECLs can be easily recovered. There is no need to detect the protocol dependent alignment markers. **Protocol independent CDR** is applicable.

A general procedure for FEC self-synchronization:

1. Start from an arbitrary bit position P.
2. Receive 1 FEC frame data, starts from bit position P.
3. Verify whether the received 1 FEC frame size data is a FEC codeword or not.
4. If the received 1 FEC frame data is not a FEC codeword, update bit position P to a new position and repeat step 2 and 3 until the FEC codeword is found.

Symbol mapping can have CDR greatly simplified (Cont'd)



Function blocks of CDR that support duplex backward compatibility.

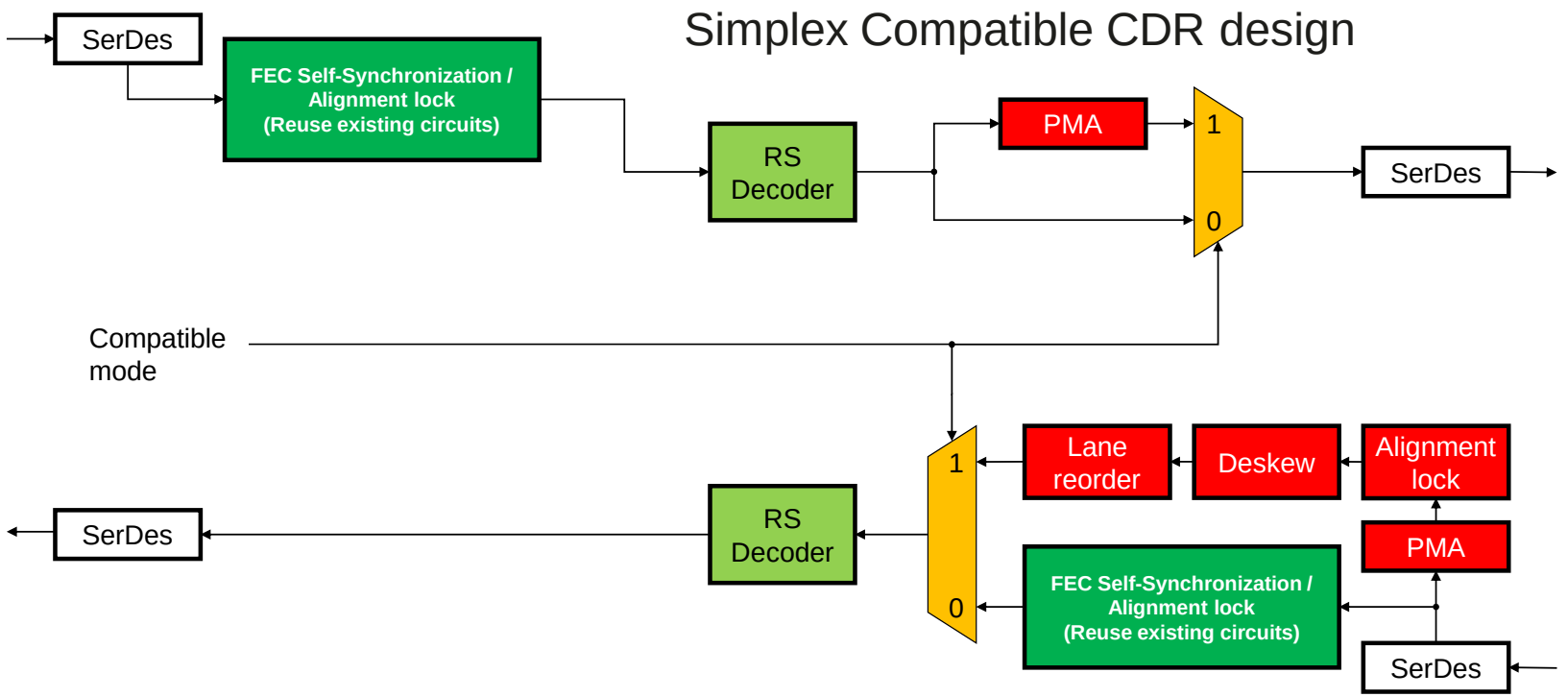
Both side supports bit-mux and symbol mapping PMA.

Duplex backward compatibility is not needed, because the newly developed chips can talk with 100GAUI-1 based on "symbol mapping PMA".

1. The complexity of CDR chip mainly comes from the skew ambiguity @the legacy optical module side, due to the "bit mux PMA". The blocks marked in **red**.
2. The FEC can be self-synchronized by reusing circuits in the RS decoder.
3. The RS decoders can also be switched off to save power and reduce latency.

HUAWEI

Symbol mapping can have CDR greatly simplified (Cont'd)



Function blocks of CDR that support simplex backward compatibility.

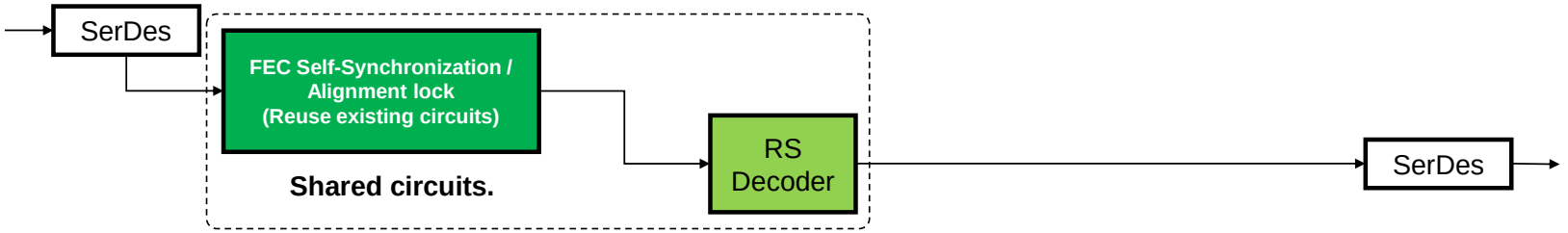
One side supports bit-mux and symbol mapping PMA; the other side supports symbol mapping PMA only.

Newly developed chips can talk with 100GAUI-1 based on symbol mapping PMA; and new developed chip talk to legacy chips with bit mux PMA.

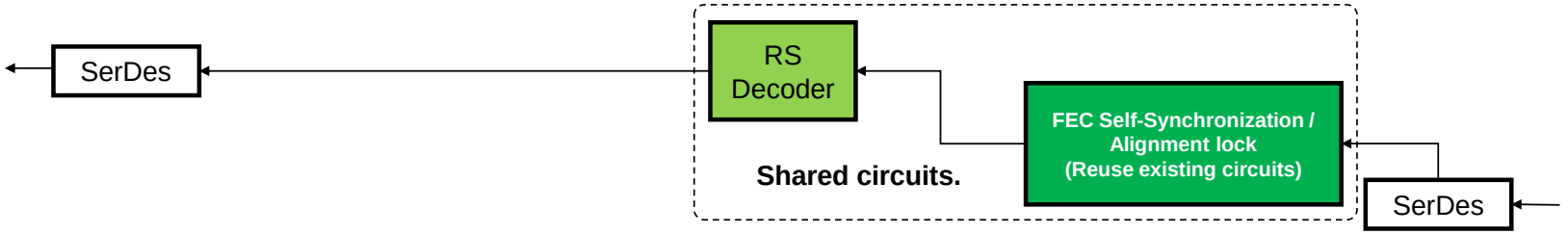
Though the "host side SerDes" and "Line side SerDes" are different, the "PMA gearbox" , "Alignment lock", "Deskew" and "Lane reoder" can be shared.

The complexity of CDR chip is mainly due to backward compatibility to the bit mux PMA.

Symbol mapping can have CDR greatly simplified (Cont'd)



Protocol independent CDR.



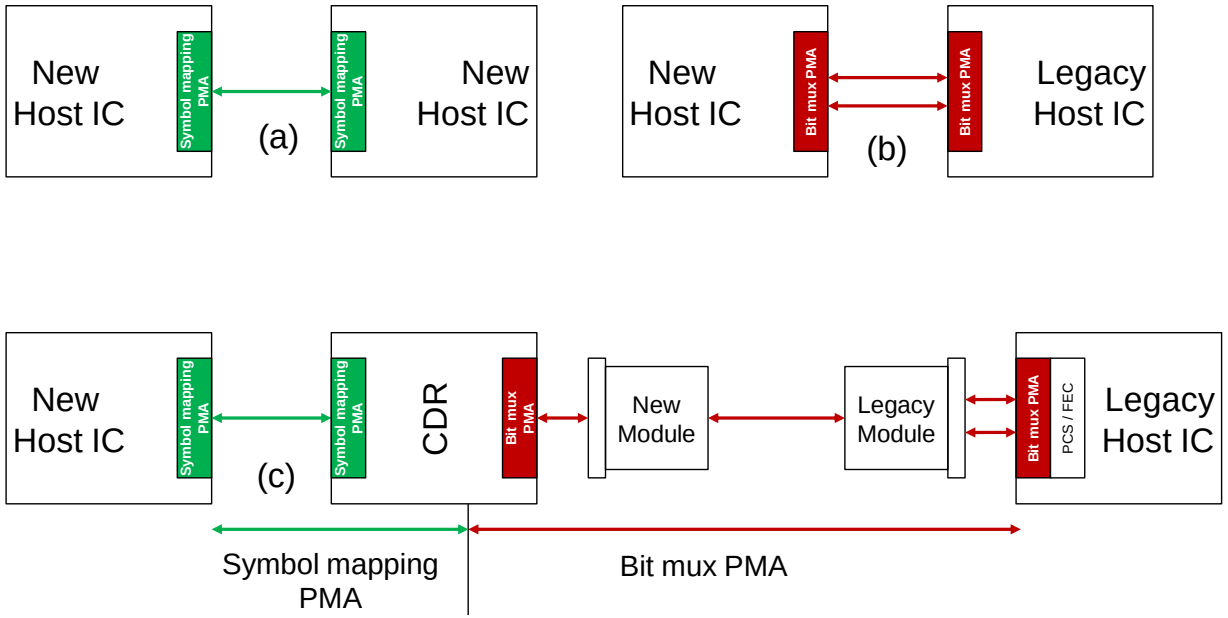
Function blocks of CDR that supporting protocol-transparency "FEC recovery".

Both sides supports "symbol mapping PMA only".

This can be used for newly defined 100GE ports.

1. The FEC can be self-synchronized so the CDR chip does not need to identify protocol dependent alignment markers.
2. RS(544, 514) or its compatible variants have already been widely used and integrated into CDR chips. Integration of RS(544, 514) compatible FEC decoder is becoming popular in real CDR chip implementations.
3. No additional complexity in chip area, power and latency will be introduced if "symbol mapping PMA" is applied. Ethernet can use protocol independent CDR chip to achieve protocol-transparency "FEC Recovery".

Scenarios of “symbol mapping PMA”

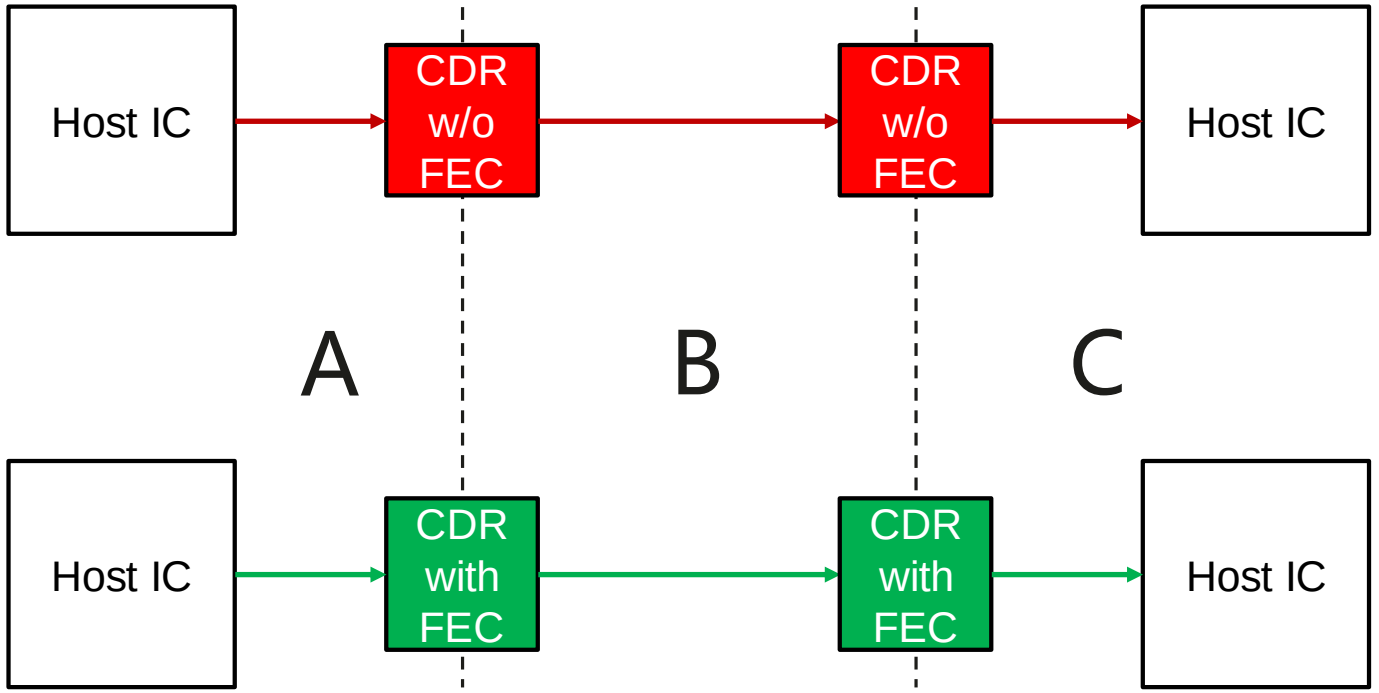


1. New chip can talk to new chip with “symbol mapping PMA” and “bit mux PMA”.
2. New chip can talk to legacy chip with “bit mux PMA”.
3. Host IC can support dual modes, i.e. “symbol mapping PMA” and “bit mux PMA” with negligible cost.
4. CDR chips can achieve the “Symbol mapping PMA” and “Bit mux PMA” 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

System benefits when integrating FEC decoder

FEC on the host IC cover the end-to-end link.



FEC is recovered segment by segment.

Only one direction is shown here.

Besides "PMA remapping", advanced features can be achievable in CDR:

1. "FEC recovery" to guarantee performance.
2. "segment-to-segment performance monitoring and fault location".

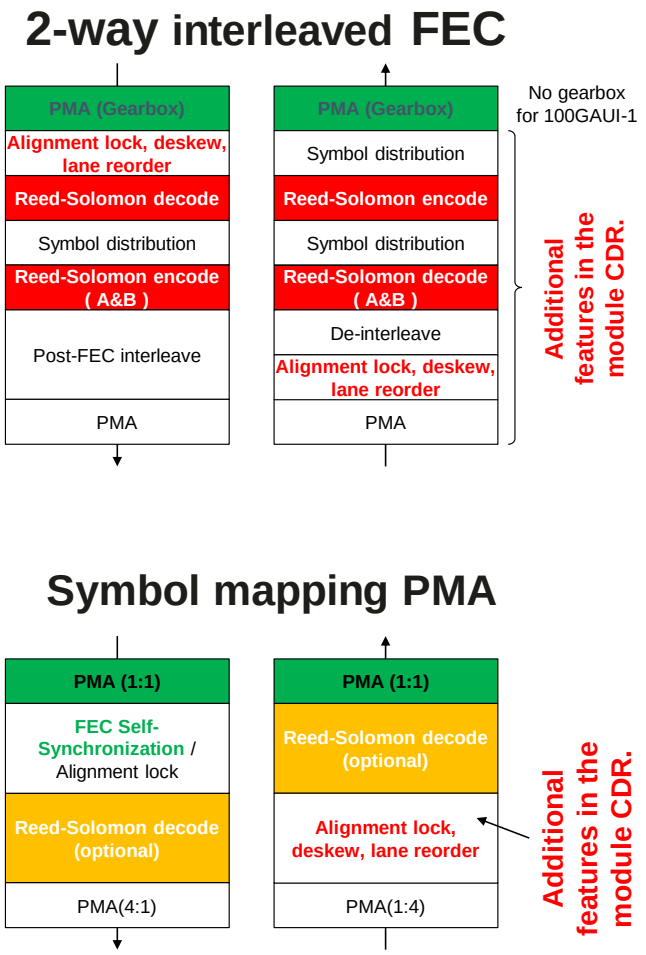
If there is no "FEC recovery" in CDR:

1. Errors in link A, B, C, will merge together, it is hard to allocate the error correction capability of FEC locates in the host IC.
   - Use different constrain in the BER? E.g. $BER_A = BER_C$ = 1e-6, and BER_B = 1e-4?
2. How to evaluate the end-to-end margin?
3. How to achieve OAM functions such as segment-by-segment performance monitoring and fault location?

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”.**

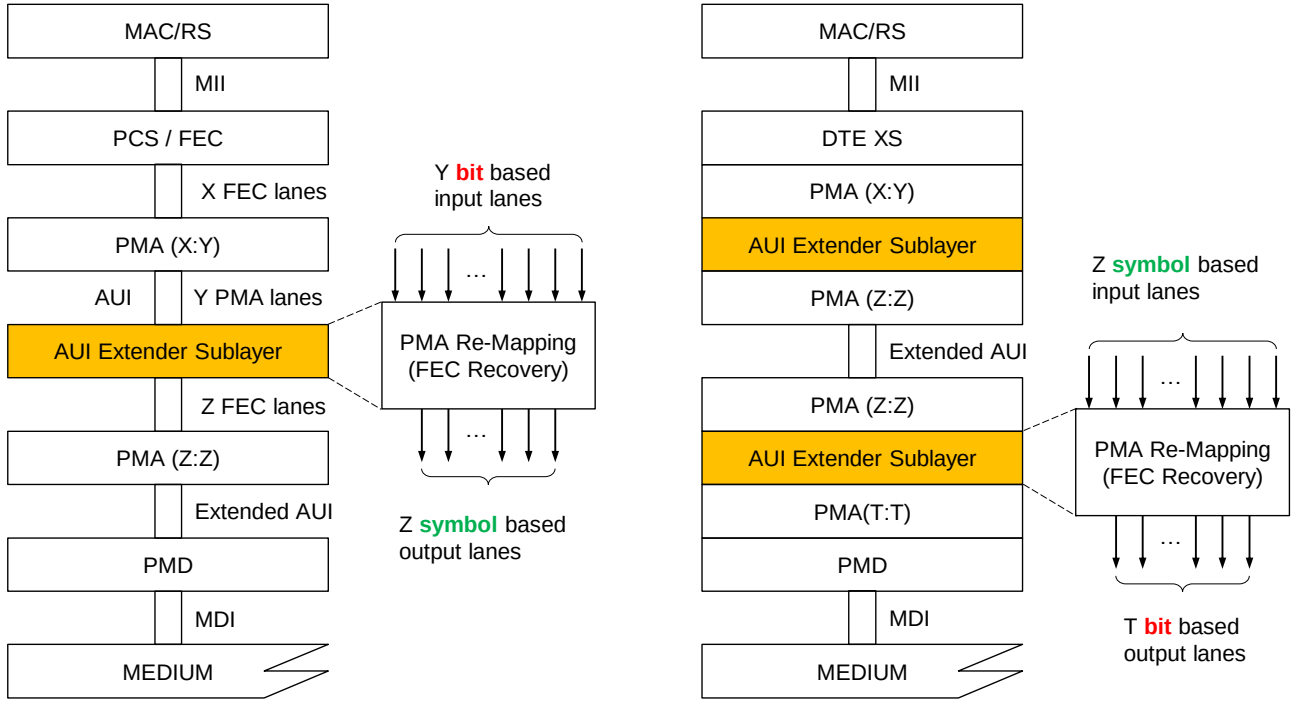


HUAWEI

Summary

- 1-tap DFE receiver does not have FEC performance concern if pre-coding is applied. No new feature should be added for this architecture.
- There may exist FEC performance concern for n-tap DFE receivers. They may have lower performance and need to relax the DFE weight constrains to get better COM, e.g. bmax=[0.85, 0.35]. Some optional solutions are considered, including PMD, PMA sublayer and FEC sublayer:
    1. PMD Sublayer solutions, i.e. constraining DFE weights may address this concern without impact on the standard and system, and precoding 2.0 may even provide better performance.
    2. PMA Syblayer solution, i.e. “symbol-mapping PMA”, can address this concern with negligible cost and impact on the standard and system. It can simplify the CDR and support “protocol independent CDR”.
    3. FEC Sublayer solution, i.e. “2-way interleaved FEC + 2:1 bit mux” can address this FEC performance concern, but may bring some system issues, such as latency and complicated CDR implementation.

AUI Extender Sublayer if supporting symbol remapping PMA



(a) Host

(b) CDR

Inherit the bit mux PMA architecture, just define symbol-mapping option when necessary.
Bit mux PMA(m:n) is equivalent to symbol-mapping when “m=n”, actually nothing new will be introduced.

An AUI Extender Sublayer (AUI XS) can be considered to:

1. Support “PMA Re-mapping” to mitigate the FEC performance concern of multi-tap DFE for electrical links as well as optical links.
2. Support “FEC Recovery”; Support protocol independent CDR.
3. Support “Segment-to-segment performance monitoring and fault location”.

Some Criteria:

1. Inherit bit mux PMA architecture, compatible with bit mux PMA. PMA remapping can be view as an optional function, just like precoding.
2. Reuse alignment markers (AM).
3. Keep the module as simple as possible.

Detailed protocol stacks and AM mapping rules need more study.

HUAWEI

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.