Technical Feasibility of Bit multiplexing in 400GbE PMA

Xinyuan Wang, Tongtong Wang, Wenbin Yang

HUAWEI TECHNOLOGIES CO., LTD.

IEEE 802.3 400 GbE Study Group

HUAWEI

Supporters

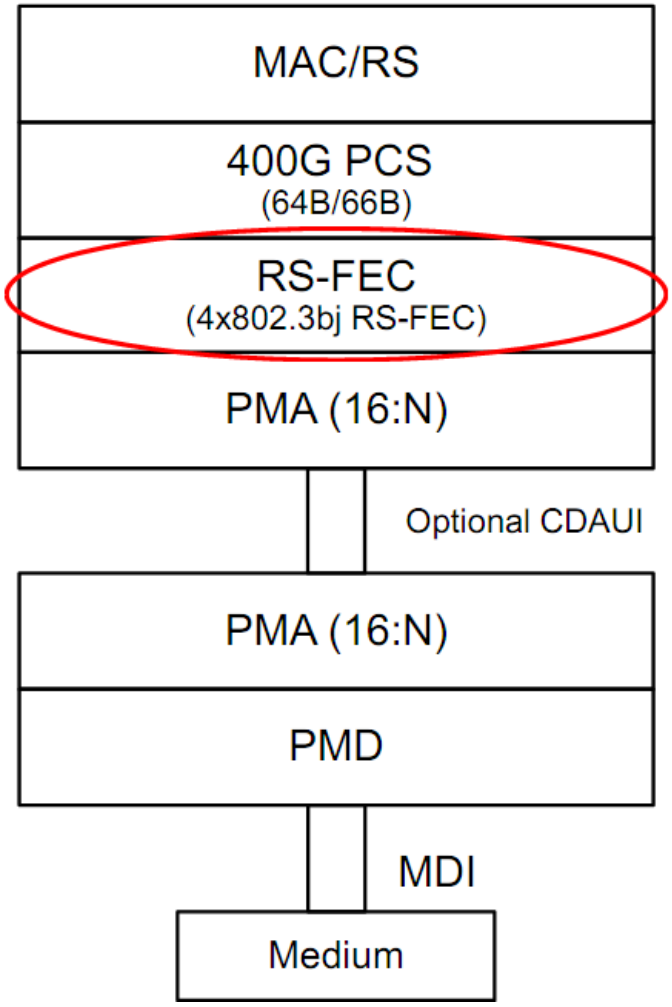
- Peter Stassar, Huawei
- Ali Ghiasi, Independent
- Chris Cole, Finisar
- Edward Nakamoto, Spirent

Motivation

- To present an improvement proposal to the 400GE PCS as presented in November 2013 in Dallas (in gustlin_400_01_1113), based on 10-bit blocks, probably precluding protocol-agnostic optical modules.
- An improved PCS/PMA architecture will be considered
  - On the basis of Bit-mux instead of muxing on the basis of 10-bit blocks
  - Supporting interoperation with CDAUI-n.
  - Enabling a protocol-agnostic optical module for usage in both Ethernet and OTN applications.
  - Increasing broad market potential
- Generic RS-FEC in PCS/PMA sub-layer
  - Correcting burst errors introduced by DFE in RX side of CDAUI or gearbox;
  - Keeping RS-FEC symbol integrity for MTTFPA when lane width change;
  - Fulfilling 1E-13 BER objective;

Background: Presentation “A 400GE PCS Option”, Dallas November 2013

- PCS is 64B/66B based
- Required RS-FEC sublayer
- Interface between the PCS and RS-FEC is not exposed (no concept of PCS lanes!)
- 16 FEC lanes below the RS-FEC sublayer

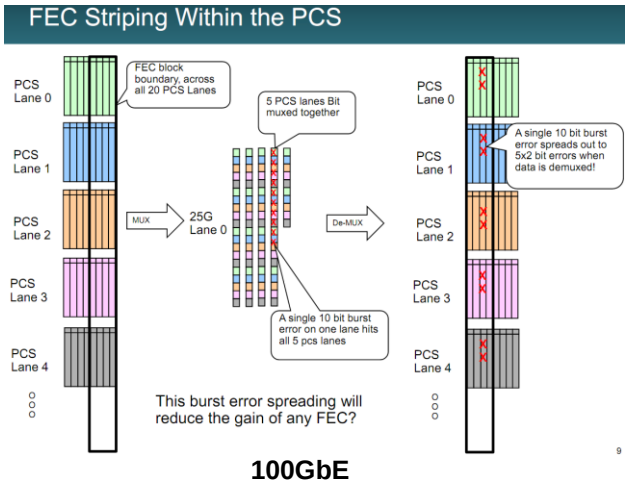


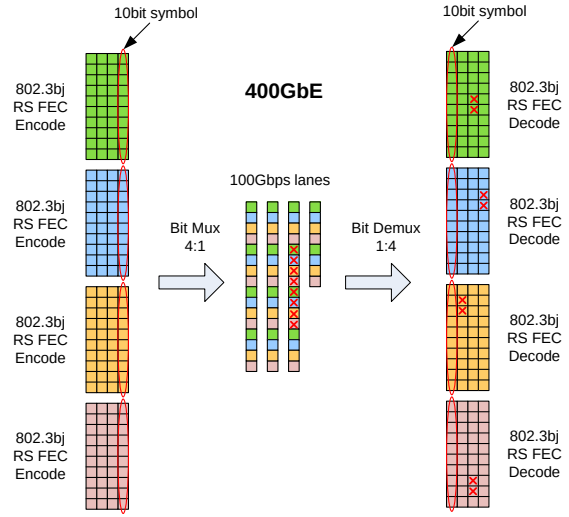
- Is it possible to define an alternative option for 400GE, using bit-mux/demux in the PMA and satisfying the requirement for RS-FEC symbol integrity?

http://www.ieee802.org/3/400GSG/public/13_11/gustlin_400_02a_1113.pdf

How to Address Burst Error Spreading into Multiple errors?

- In 802.3 bj concerns were raised that a single burst error spreads into multiple error symbols by bit-mux/demux in single RS-FEC.
  - Refer to “gustlin_02_0911”

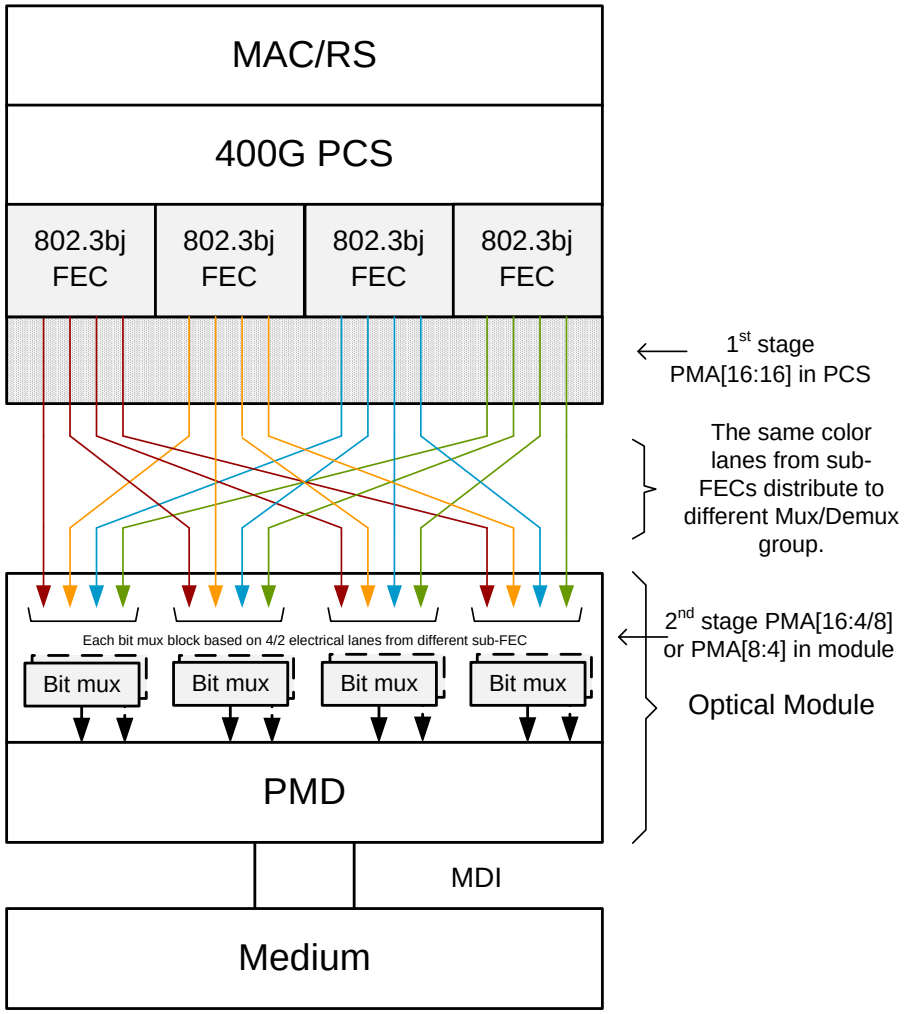




Potential solution of 400GE bit-mux in optical module:

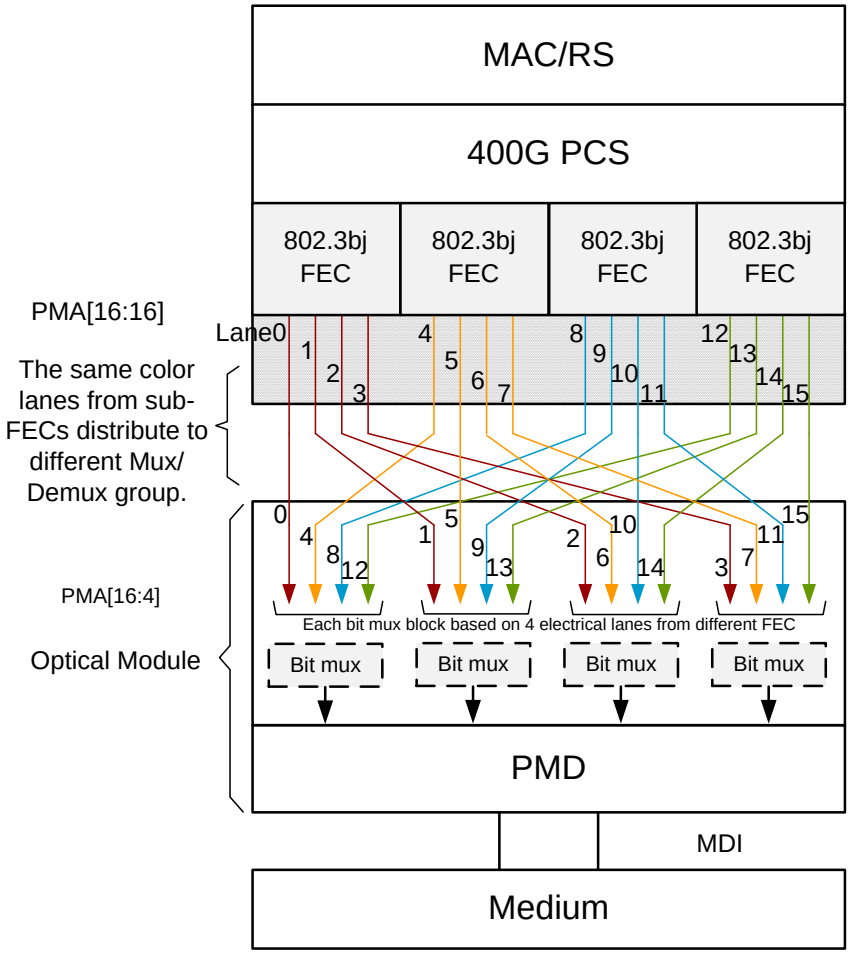
- By interleaving four parallel RS-FEC one burst error can be broken into multiple short burst errors to different RS-FECs.
- Each RS-FEC corrects short burst errors as in 802.3bj.
- The advantages of this method will be quantified and compared with symbol mux scheme.

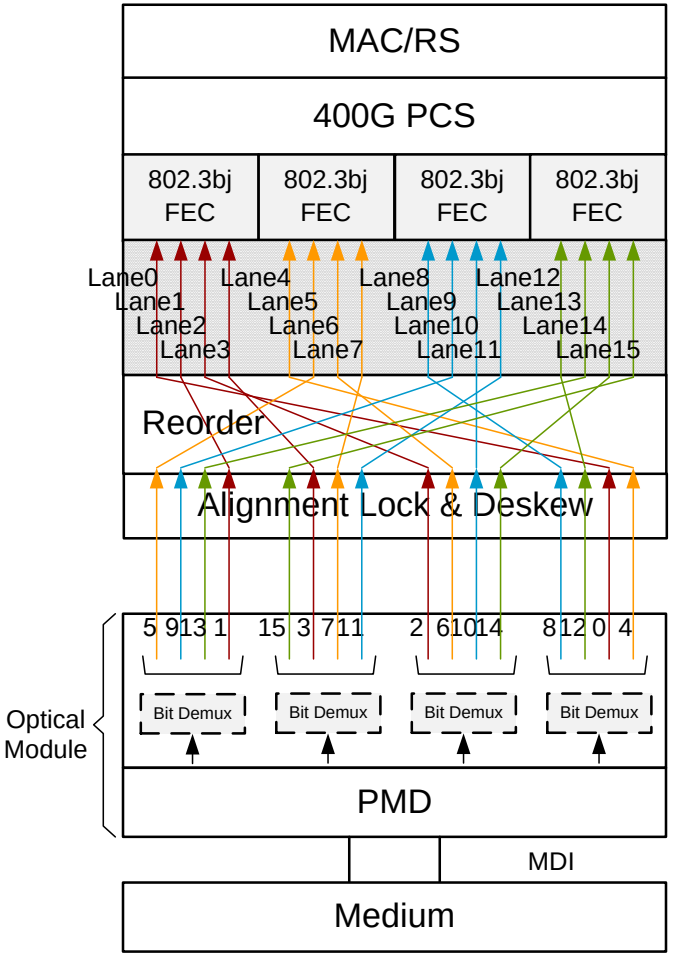
Improved PCS/PMA Architecture with Bit Mux



- Use different muxing in PMAs. PMA in PCS is 16:16 or 16:8 with symbol-mux/demux and PMA [16:4], PMA[16:8] or PMA[8:4] in module is bit-mux/demux.
- Grouping FEC lanes from 4 different RS-FECs and do bit-mux/demux in optical module.
- Gives loose constraint in layout design of CDAUI interface .
- No specific lane sequence and skew constraint in each bit-mux/demux group.
- If burst error happens in each group, it breaks to shorter errors, then distributing to different sub RS-FEC codewords.

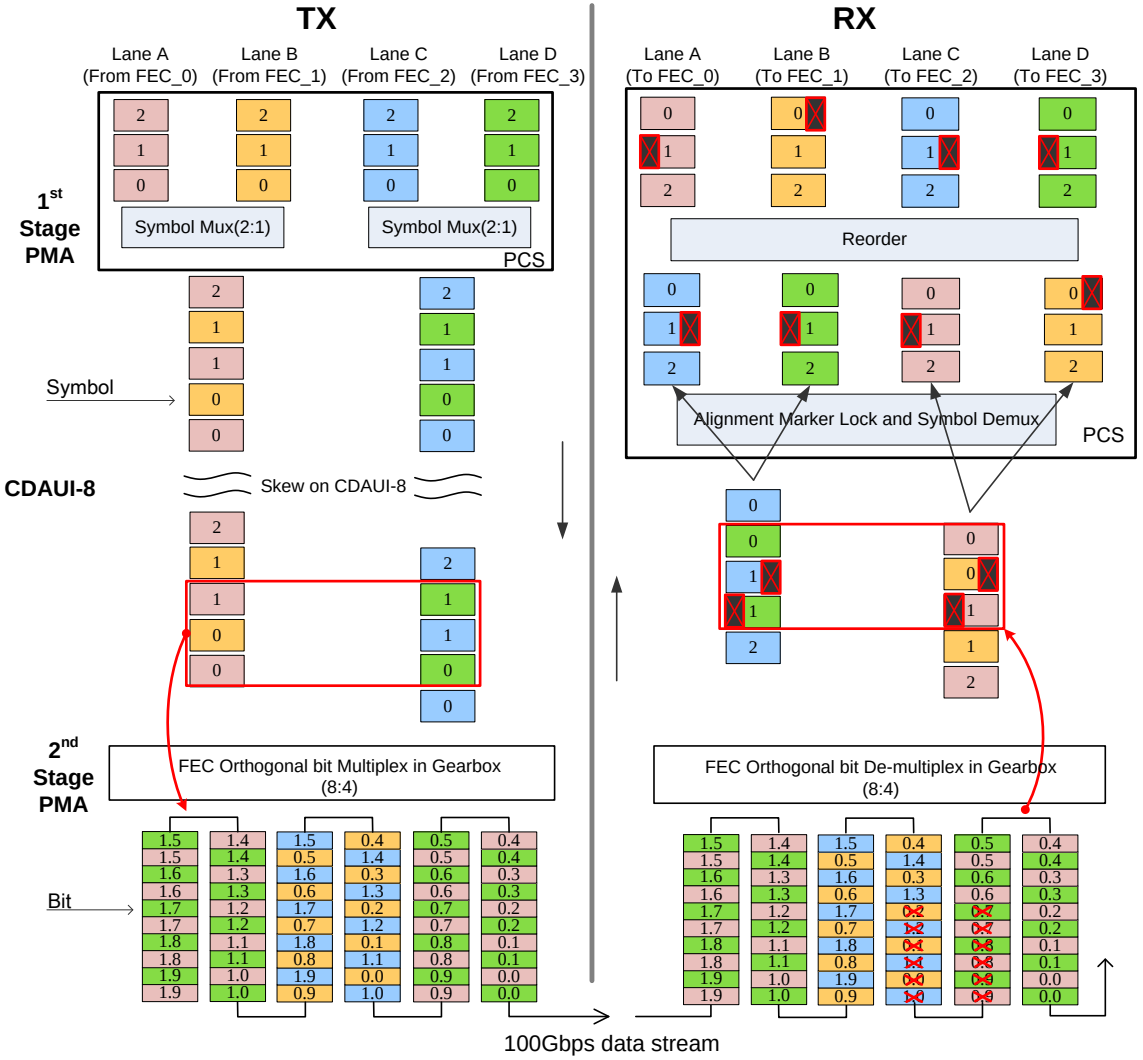
FEC Lanes Alignment and Reorder





- No lanes ordering constraint for bit-demuxing in each multiplexed group.
- Use alignment and reorder mechanism in RX side across all lanes as in 100GE.

Analysis of Error Spreading Beyond Symbols



- This is an example of cross lane multiplexing on CDAUI-8 and 4X100Gbps PMD. Same mechanism is useful in CDAUI-16 with 8X50Gbps PMD.
- Use symbol mux in 1st Stage PMA for better performance against burst error on CDAUI interface.
- Burst error from gearbox distributes to different sub-FEC blocks by 2nd Stage PMA, which keeps system MTTFPA as in 802.3bj.

Feasibility of Implementing CDAUI Layout

- Require less than 4 signal layers for partial SerDes crisscrossing interconnection in CDAUI interface between chip and optical module;
- Usually PCS/PMA chip and connector are placed near each other for shorter trace of high speed SerDes, one layer signal routing is impractical in most designs, even with no routing constraints.
- Mature hardware design examples of multi-lanes SerDes interface with specific ordering:
  - XAUI

XGMII

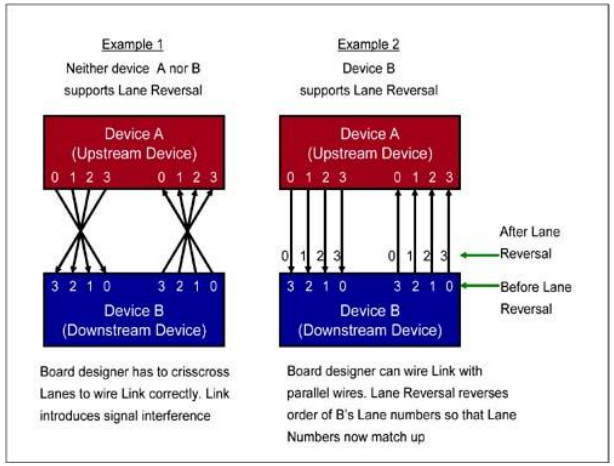
T/RXD<7:0>	I	I	S	Dp	D	D	D	---	D	D	D	D	I	I	I	I	I	I
T/RXD<15:8>	I	I	Dp	Dp	D	D	D	---	D	D	D	T	I	I	I	I	I	I
T/RXD<23:16>	I	I	Dp	Dp	D	D	D	---	D	D	D	I	I	I	I	I	I	I
T/RXD<31:24>	I	I	Dp	Ds	D	D	D	---	D	D	D	I	I	I	I	I	I	I

PCS

LANE 0	K	R	S	Dp	D	D	D	---	D	D	D	D	A	R	R	K	K	R
LANE 1	K	R	Dp	Dp	D	D	D	---	D	D	D	T	A	R	R	K	K	R
LANE 2	K	R	Dp	Dp	D	D	D	---	D	D	D	K	A	R	R	K	K	R
LANE 3	K	R	Dp	Ds	D	D	D	---	D	D	D	K	A	R	R	K	K	R

Legend:
Dp represents a data character containing the preamble pattern
Ds represents a data character containing the SFD pattern

  - X4/8/16/32 PCI Express



Courtesy: PCI Express System Architecture

Generic Rules for Effect of Burst Errors on Multi-FECs

*m is symbol size.

	Burst Error Length (bits)	Maximum Error Symbols in one sub-FEC
No Bit Mux/Demux	1	1
	2 to m+1	2
	m+2 to 2*m+1	3
Bit Mux/Demux 2:1	1 to 2	1
	3 to 2*m+2	2
	2*m+3 to 4*m+2	3
Bit Mux/Demux 4:1	1 to 4	1
	5 to 4*m+4	2
	4*m+5 to 8*m+4	3

- The table shows the number of error symbols caused by burst error in worst case.
- Less error symbols occur if burst error happens within symbol boundary.

Summary

- A PCS/PMA architecture on the basis of FOM(FEC Orthogonal Multiplexing) is proposed for 400GE.
  - Taking advantage of multiple parallel FEC blocks;
  - Supporting protocol agnostic optical module with less complexities and shorter verification & test time;
  - Supporting both Ethernet and OTN applications for broad market potential;
- Routing requirements from hardware route group for CDAUI is a loose constraint and easy to implement, simpler than XAUI.
- Both CDAUI and gearbox may have burst errors:
  - By using bit demux in gearbox burst errors are spread to different sub FECs;
  - RS-FEC has good performance against burst errors on CDAUI SerDes interface;
  - This solution can keep system MTTFPA performance as in 802.3bj.

Further Work

- Investigate Burst Error model of various PMDs in 400GE;
  - Different DFE architecture in PAMn/NRZ gearbox and CDAUI interface may influence the error probability and pattern.
- CDAUI-n interoperation based on FOM;
- MTTFPA analysis for different random/burst error model:
  - Bit Mux 4/2/1 with CDAUI-n options;
  - Evaluation based on BER objective 1E-13;

Thank you

www.huawei.com