The Scalable 400GbE PCS Architecture

Tongtong Wang, Wenbin Yang, Xinyuan Wang

Sep 2013

HUAWEI

Motivation

- To clarify that when FEC is integrated in PCS, the number of PCS lanes is not critical.
- We could use 16 PCS lanes and 80 AMs to provide the compatibility with 802.3bj.
- 16 PCS lanes with 80AMs scheme could reuse 802.3bj FEC.

Background

- Multi-rate support in 400GbE logic layer?
  - The key factors in 400GbE PHY/PCS to be compatible with 100/40GbE are the type of FEC and its implementation.
  - If 400GE standard needs to introduce new PMDs in 100/40GE PCS/PHYs, it would be out of scope of 400GbE study group.

  http://www.ieee802.org/3/400GSG/public/adhoc/logic/jun26_13/wangt_01_0613_logic.pdf
- Scalable 400 GbE Architecture
  - 80 VLs provide maximum flexibility by supporting existing PMD implementations at this early stage.
  - 400GbE based on 16 VL will not be compatible with MLG.
  - Reuse RS-FEC.

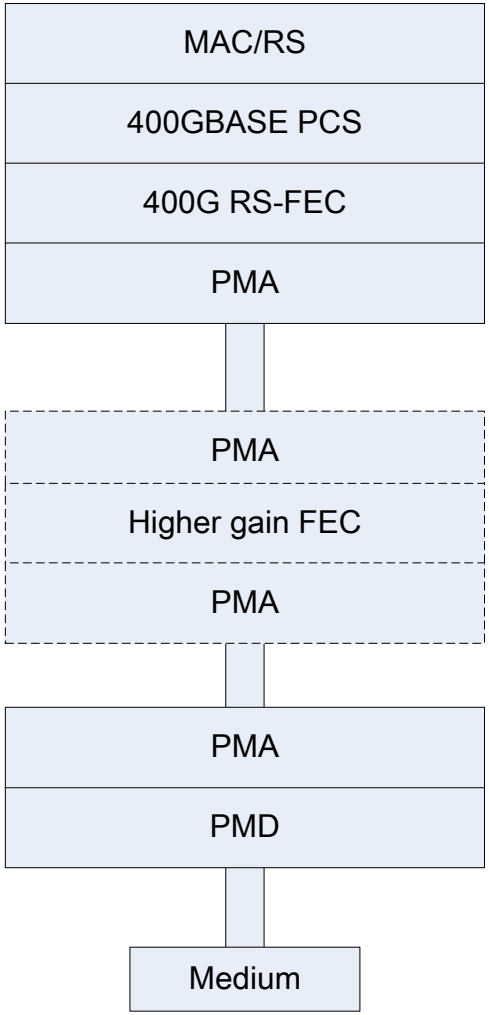
  http://www.ieee802.org/3/400GSG/public/13_07/ghiasi_400_01_0713.pdf
- 400GbE PCS Architectural Options
  - Possible Architecture #5: Based on a 80 Lane PCS with 256B/257B encoding. Not sure how 80 PCS lanes helps you when you only have 16 FEC lanes?

  http://www.ieee802.org/3/400GSG/public/13_07/gustlin_400_02_0713.pdf

Key Factors of PCS Architecture

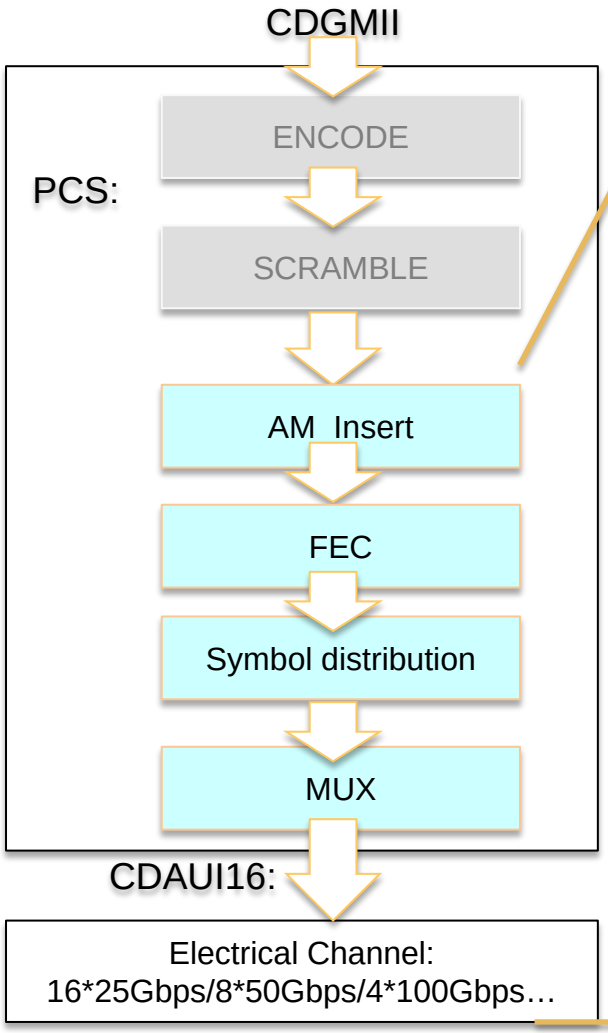
- FEC algorithm
  - If a generic FEC is mandatory in 400GE, making it integrated in PCS cost efficiently with acceptable net gain is straightforward and economical.
  - Base on 802.3 bm/bj and previous analysis, RS may be a proper algorithm in 400GE PHYs.
- PCS lanes/FEC symbol lanes
  - PCS lanes will be transformed into FEC symbol lanes, thus the number of PCS lanes has less impact on 400GbE/4x100GbE/10x40GbE interoperating. The less PCS lanes count, the less complexity and logic resource required.
  - FEC symbols but not PCS blocks are multiplexed on PMA/PMDs.
- Alignment Marker
  - 16 or 80 PCS lane design does not affect FEC mechanism by keeping AMG distance as a multiple of FEC block size;
- Scalable 400GE /4X100GE/ 10X40GE
  - To design a 400GbE ASIC with 4x100GbE/10x40GbE support is desirable. However, the compatibility requires 80 PCS lanes. Future MLG is possible to support 40GbE bundling in 400GbE.
  - If the cost and latency are acceptable, reusing off-the-shelf technology, e.g., RS-FEC in 802.3bj, could simplify 400GbE architecture design.

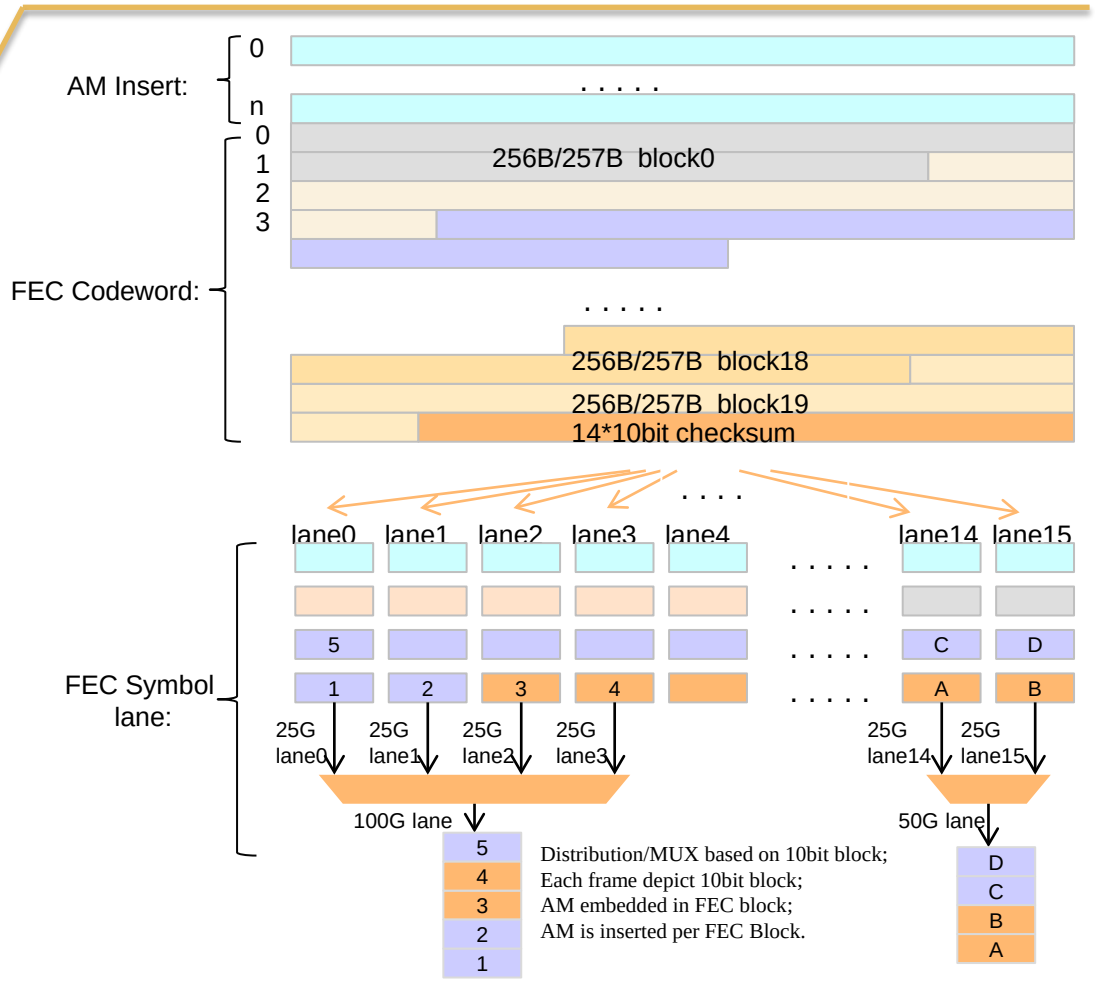
Requirement for 400GbE Architecture with FEC



- One unified PCS architecture with generic FEC is feasible to cover most PMAs/PMDs with 25/50/100Gbps lanes per lambda, fiber, copper cable, or SerDes. 40Gbps electrical/optical PMDs is infeasible in current technology.
- 400GbE should have one set of logic architecture to serve most types of PMDs. E.g., RS-FEC may be used for multiple solutions, such as Backplane/Cable, MMF, and some SMF scenarios.
- Optional higher gain FEC would be used in specific scenario with higher complexity and cost.

A Possible 400GbE PCS Architecture with FEC





**Refer to wangt_01_0613_logic.pdf, gustlin_400_02_0713.pdf*

802.3bj Compatible with 16/80 PCS Lanes and 80 AMs

- To be compatible with 802.3bj, 80 PCS lanes are suggested to provide flexibility, while 16 FEC lanes with 80 AMs could do the same with less resource.

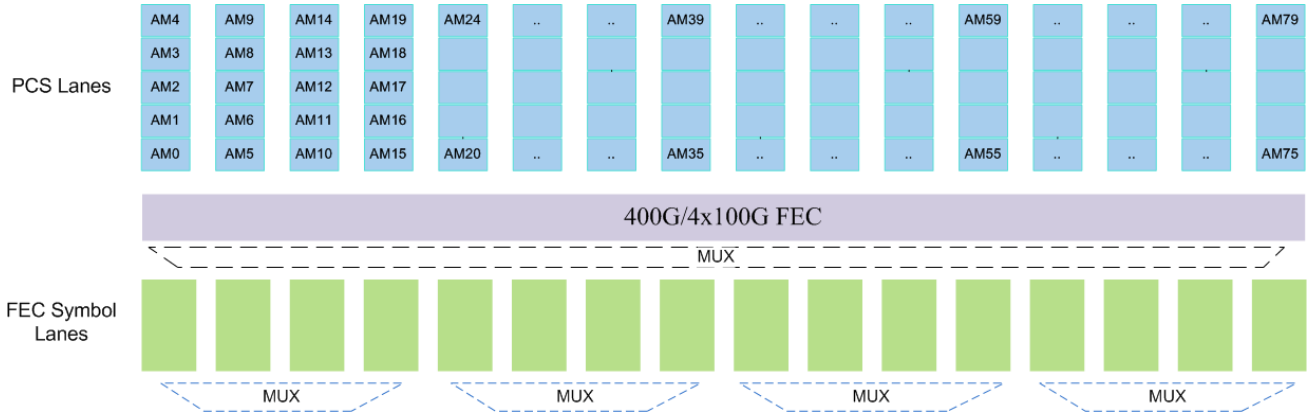


Figure 1: 16 FEC lanes with 80 AMs

- A simpler 16 PCS lanes + 16 AMs scheme can limit its application, but it is possible to change AM format/period to smaller size and be more efficient.

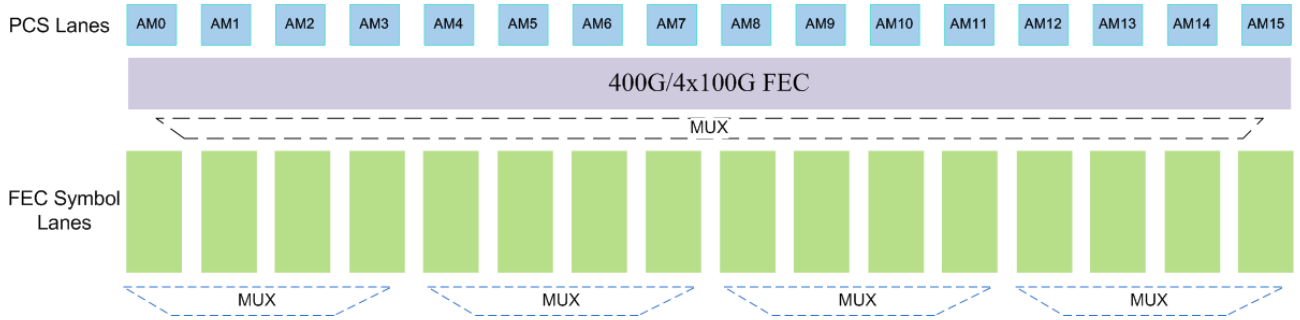


Figure 2: 16 FEC lanes with 16 AMs

Alignment Marker Details

- Using 80 AM headers in PCS lane, AM period could remain 4*4096=16384 as 802.3bj.

PCS Lane number	AM number	Compatible with 802.3bj?	AM period of PCS lane (64/66b)	AM period of FEC layer (5280b)	FEC lane AM Period changed?
5	5	no	16384	1024	Yes
	5*16	yes	16384*16	16384	No
10	10	no	16384	2048	Yes
	10*8	yes	16384*8	16384	No
15	15	no	16384	3072	Yes
16	16	no	variable	variable	Yes
	16*5	yes	16384*5	16384	No
20	20	no	16384	4096	Yes
	20*4	yes	16384*4	16384	No
80	80	yes	16384	16384	No

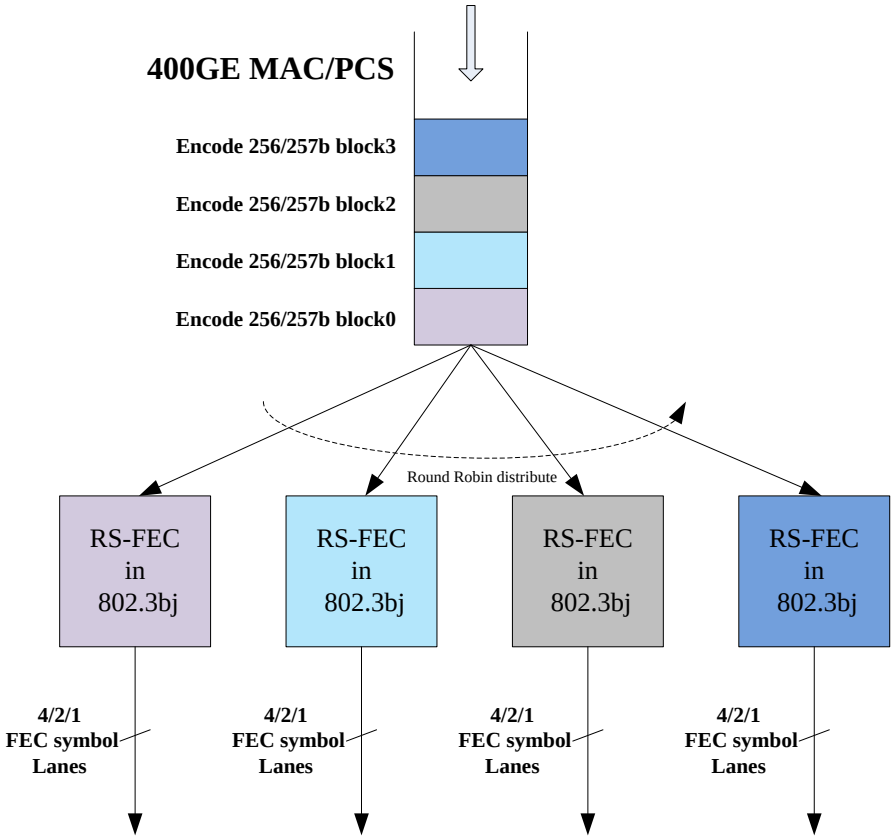
- Changing AM format/period is possible if no need for compatibility requirement;

Cost and Feature Comparison

	16 FEC Lanes + 80 AM	1x16 + 4x20 PCS Lanes	80 PCS Lanes	4x20 PCS Lanes
Capacity	400GbE/ 4x100GbE	400GbE/ 4x100GbE	400GbE/4x100 GbE/10x40GbE	4x100GbE
with FEC	Yes	No	No	No
Reorder MUX count	1x(16 mux16) x64bit @390MHz	4x(20Mux20) x 16bit @312MHz; 1X(16Mux16) x 64bit @390MHz	1x(80mux80) x16bit @312MHz	4x(20Mux20) x16bit @312MHz
Reorder resource ratio	1	2.25	5	1.25
Total resource ratio	1	2.25	3~4	-
Support 40GE MLG	Yes	No	Yes	-
Support 25Gbps PMA	Yes	Yes	Yes	-

- With integral FEC, PCS lane reordering could be simplified, after FEC lane alignment and deskew.
- Using 80 AMs on 16 FEC symbol lanes is flexible and cost-efficient.
- With 16 PCS lanes/16 AMs scheme, we should reconsider the MLG compatibility .

Reuse RS-FEC from 802.3bj



- Take egress side as example for 802.3bj RS-FEC reuse.
- Encoded 256/257b data blocks are distributed in round robin fashion to 4 parallel 100Gbps RS-FEC channels.
- Use round robin mechanism on ingress side.
- Pros:
  - Off the shelf technology in 802.3bj;
  - Lower complexity comparing to 1X400GbE RS-FEC;
- Cons:
  - ~2X latency and about ~100ns.

Summary

- Using 80 AMs on 16 FEC symbol lanes is a suitable architecture for 400GbE/4x100GbE, It could reuse 802.3bj FEC and support 16*25Gbps/8*50Gbps/4*100Gbps connections.
- Future MLG is possible to support 40GbE bundling in 400GbE.