Addressing PTP timestamp accuracy for 800GBASE-ER1 with an 800GMII Extender

Gary Nicholl, Mike Sluyski - Cisco
Ulf Parkholm - Ericsson
Xiang He - Huawei
Dave Ofelt - Juniper
Andras Dekoos - Microchip
Tom Huber - Nokia

IEEE P802.3dj Task Force Joint Logic & Optics Ad hoc, 11 April 2024

Supporters

- Jerry Pepper – Keysight
- Peter Sinn – Alphawave
- Eugene Opasnick - Broadcom
- TBA (your name could be here!)

References

Recommendation ITU-T G.8273/Y.1368 - Framework of phase and time clocks (06/2023). https://www.itu.int/rec/T-REC-G.8273/en

IEEE 1588-2019 – Standard for a Precision Clock Synchronization Protocol for Network Measurement and Control Systems. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9120376

IEEE 802.3cx; Improving PTP Timestamping Accuracy on Ethernet Interfaces: https://www.ieee802.org/3/ad_hoc/ngrates/public/calls/19_0702/tse_nea_01_190702.pdf

IEEE P802.3df: Incoming: MOPA: Time synchronization error in PTP networks. https://www.ieee802.org/3/minutes/nov23/incoming/MOPA_to_IEEE_802p3_231102_Redacted.pdf

Updated logic baseline for an 800GbE coherent PHY based on oFEC/C-band. https://grouper.ieee.org/groups/802/3/dj/public/23_07/nicholl_3dj_02a_2307.pdf

Consideration of timestamp accuracy with MII-extender in coherent 800GBASE-ER1

https://www.ieee802.org/3/dj/public/23_11/parkholm_3dj_01_2311.pdf

Introduction

- PTP accuracy is becoming more important in a broader set of networks
  - Reference liaison letter from MOPA: https://www.ieee802.org/3/minutes/nov23/incoming/MOPA_to_IEEE_802p3_231102_Redacted.pdf
- Ethernet PHYs, including those being developed within 802.3dj, can address timestamping accuracy by following Clause 90. The exception to this is the currently adopted 800GBASE-ER1 baseline with an 800GMII Extender (at either end).
- This contribution proposes an update to the current 800GBASE-ER1 baseline to address this limitation.

Problem statement

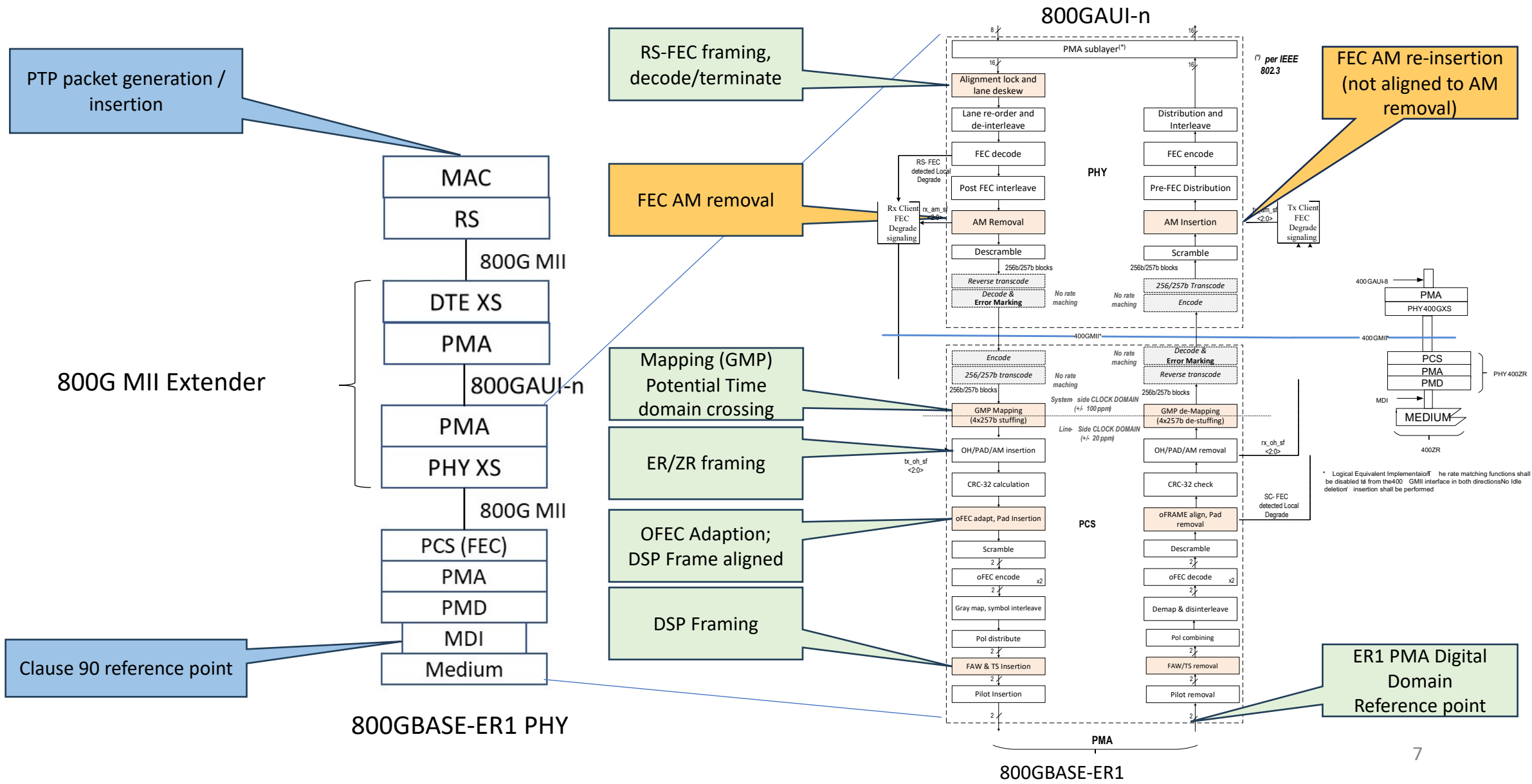
- In addition to the usual need for very careful design to achieve high PTP accuracy, 800GBASE-ER1 when used with an 800GMII Extender has an architectural constraint that limits the achievable accuracy (even after using Clause 90 compensation):
  - In the transmit path AMs are removed before mapping the Ethernet symbols into the GMP frame
  - In the receive path, after Ethernet symbols are de-mapped from the GMP frame, AMs are then re-inserted at some random location
  - This changes the relationship between packets and introduces timestamping errors

- Analysis has shown that the insertion/removal of the Alignment Mechanism (AM) fields can affect c|TE| in a measurable way (2.4 nS@ 800G).

Solution overview

- A solution is proposed that ensures AMs are re-inserted in the same positions by the receive 800GMII Extender, relative to whether they were removed by the transmit 800GMII Extender.
- This preserves the Host-to-Host timing relationship of the underlying PTP packet across the 800GBASE-ER1 link.

800GBASE-ER1 Architecture with 800GMII Extender



Proposal – Mark AM position using GMP extender (JC9-JC7) OH bytes.

Frame	MFAS	Bytes 1	2	3	4	5	6	7	8	9	10	11	12	13	…	26	27	28	29	…	40
1	xxxxx000	MFAS	STAT	GID		GID		GID	RES	IID		MAP			CRC			FCC1		OSMC/RES			
2	xxxxx001	MFAS	STAT	AVAIL		RES		JC4		JC1		MAP			CRC			FCC1		RES			
3	xxxxx010	MFAS	STAT	JC7		JC8		JC5		JC2		MAP			CRC			FCC1		RES			
4	xxxxx011	MFAS	STAT			JC9		JC6		JC3		MAP			CRC			FCC1		RES			
5	xxxxx100	MFAS	STAT					MSI[x]		PT		MAP			CRC			FCC1		RES			
6	xxxxx101	MFAS	STAT	RES				JC4		JC1		MAP			CRC			FCC1		RES			
7	xxxxx110	MFAS	STAT	JC7		JC8		JC5		JC2		MAP			CRC			FCC1		RES			
8	xxxxx111	MFAS	STAT	CSTAT		JC9		JC6		JC3		MAP			CRC			FCC1		RES			

1	2	3	4	5	6	7	8
RPF	MNT			RES			

Link status overhead

1	2	3	4	5	6	7	8
CSF	MNT			RES		RD	LD

Client status overhead

1	2	3	4	5	6	7	8
A	O	RES				PORT	

MSI overhead

	1	2	3	4	5	6	7	8
JC1	C1	C2	C3	C4	C5	C6	C7	C8
JC2	C9	C10	C11	C12	C13	C14	II	DI
JC3	CRC8							
JC4	RES					D1	D2	D3
JC5	RES					D4	D5	RES
JC6	RES					CRC3		

Proposed GMP extensions for use in AM synchronization

GMP and Tributary OH fields

FlexO-x information Structure OH defined by ITU-T G.709.1 to carry various mapped payloads

Overhead Field	800G BASE-ER1
MFAS	Yes
STAT-RPF	Yes
STAT-MNT	Yes
STAT-LD	Yes
STAT-RD	Yes
STAT-SPARE	Yes
STAT-RES	Yes
GID	Yes
RES – row 0, byte 5	No
OSMC/RES	No
RES	No
CRC	Yes
MAP	Yes
CSTAT	Yes
IID	Yes

AM Synchronization Mechanisms

- JC9-JC7 (MSB-LSB) bytes can be used to provide a mechanism (16-counter) to transmit phase information of the Alignment Marker field (AM) across the 800GBASE-ER1 link. The counter tracks the number of valid (non-stuff) 4 x 257b data blocks transferred between AMs.
  - Purpose of the counter is to convey the “Start/End” of the 800GE Reed Solomon Block transfer.
  - For 800G there are 8192 (163840 x 257b blocks) 514b code words between AM marked RS coded blocks.
  - GMP stuffing is done in 4 x 257b blocks.
  - Counter increments for each valid (non-stuff) 4 x 257b block mapped into transmit container. The value in J7-J9 corresponds to the value of the counter on the first 4 x 257-bit payload block of the ZR/FlexO-8e multiframe that the J7-J9 bytes are located in.
  - Count value is initialized to zero by the TX GMP mapper, aligned to the removed AM fields. It increments every 4 x 257b valid data block. The Max Modulo is the expected end of the 4 x 257b valid data for the RS AM marked interface. RX de-mapper keeps equivalent RX counter and synchronizes to JC9-JC7 signals.
  - AM are inserted between the payload block where the counter equals 40957 and the payload block where the counter equals 0.

Host Rate	Code Words/AM Block	257-bit blocks/AM	800G Valid Data Counter (MAX modulo)
800G	8192	163840	40960

AM Synchronization

- The purpose of the Counter is to convey the TX_START/END (Point when AM removed at TX mapper) to RX_START/END (Point where AM insertion is required in RX de-mapper)

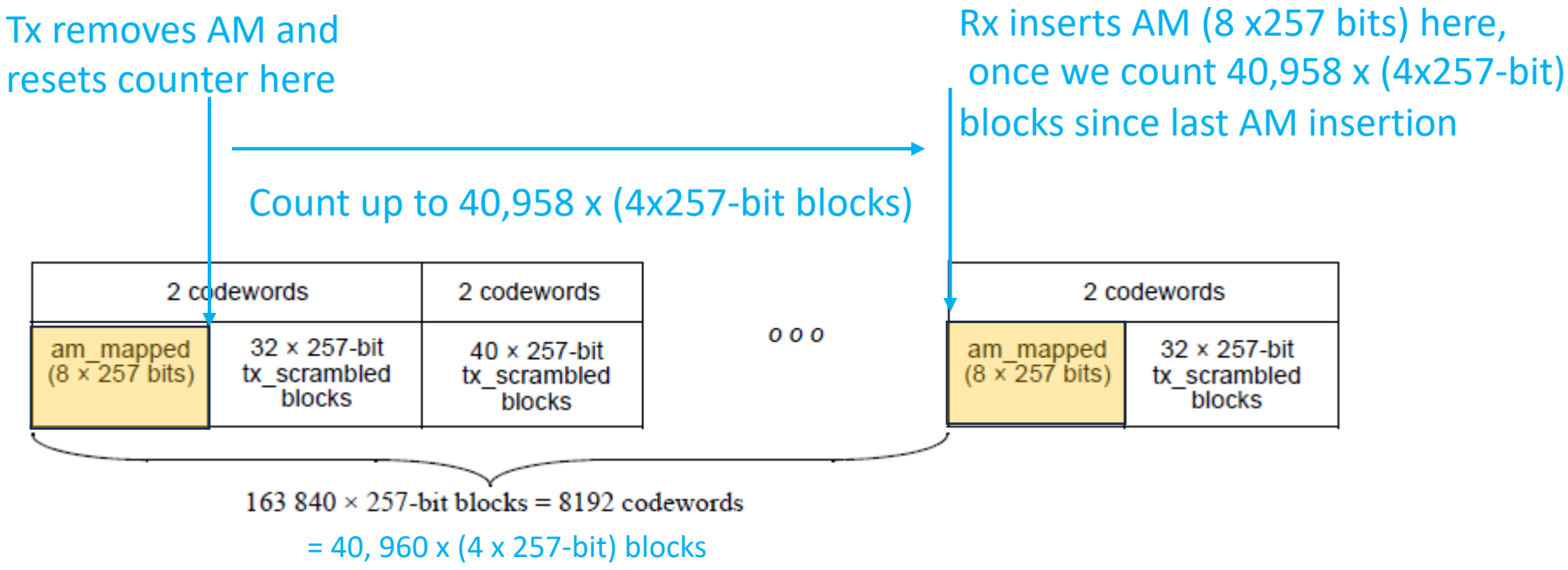


Figure 119–8—400GBASE-R alignment marker insertion period

Summary

- A potential issue was identified related to timestamping accuracy when using an 800GBASE-ER1 PHY with an 800GMII Extender, which is not adequately compensated for by Clause 90.
- This contribution proposes a method to mark the position where the Alignment Markers (AM) are removed in the transmit 800GMII Extender and carry this information over the 800BASE-ER1 link such that the receive 800GMII Extender can re-insert the AMs in the same location from which they were removed.
- This optional mechanism would allow an implementation choice for 800GBASE-ER1 that could comply with the intent of the “Note” in 90.7.2 (referenced in our response to the MPOA liaison letter) and essentially brings 800GBASE-ER1 into line with the other 802.3dj PHYs in this regard.
- Note: the same solution is applicable to 800GBASE-ER1-20

Thanks !