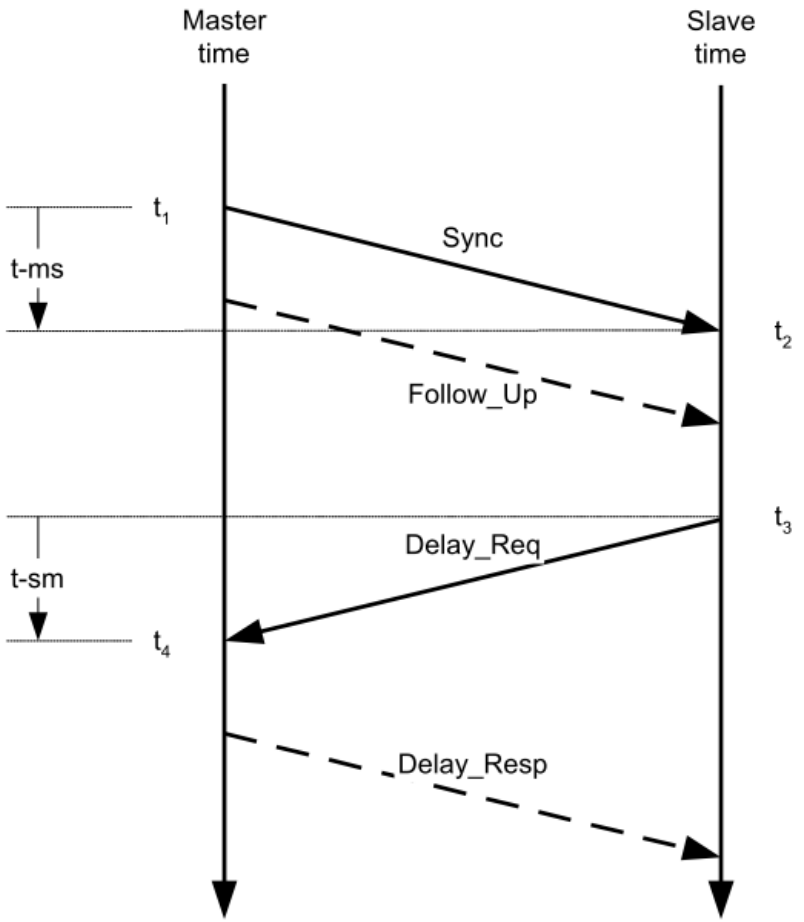
Timestamping across an 800GE MII Extender (supporting comment #2)

Andras de Koos – Microchip Technology

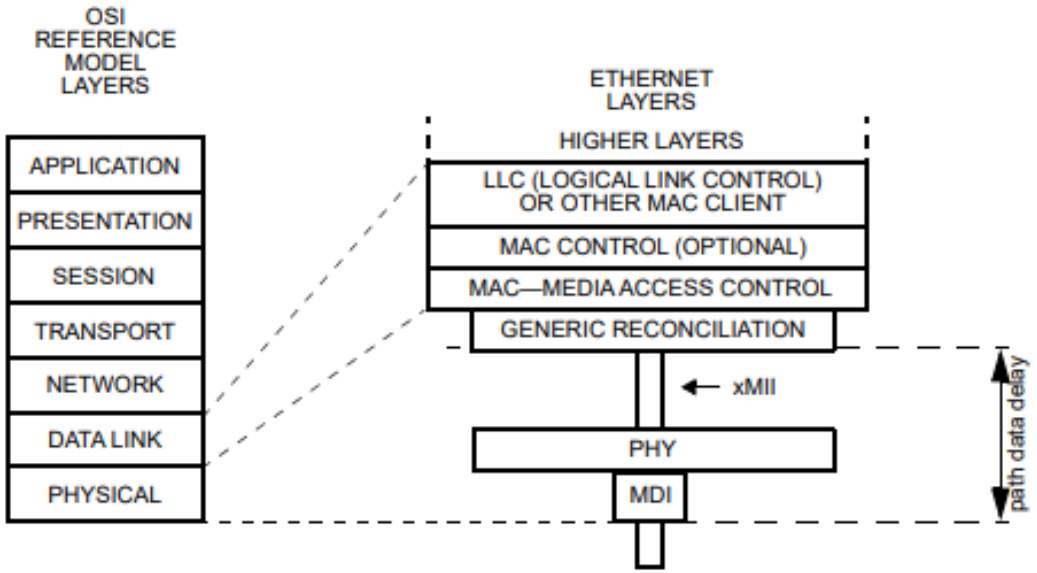
IEEE 802.3df Task Force

PTP/Timestamping Background

- Synchronize Time-of-Day across a network by exchanging messages and their measured arrival/departure times.
- With the round-trip delay, the time difference between the TimeTransmitter and TimeReceiver can be calculated.
- Requirements for time synchronization accuracy are defined in ITU-T Recommendation G.8273.2
  - “Class D” targets end-to-end accuracy to within +/- 5ns!
- The more precise the timestamps, the smaller the end-to-end TimeSyncError.



Timestamping Model for Ethernet

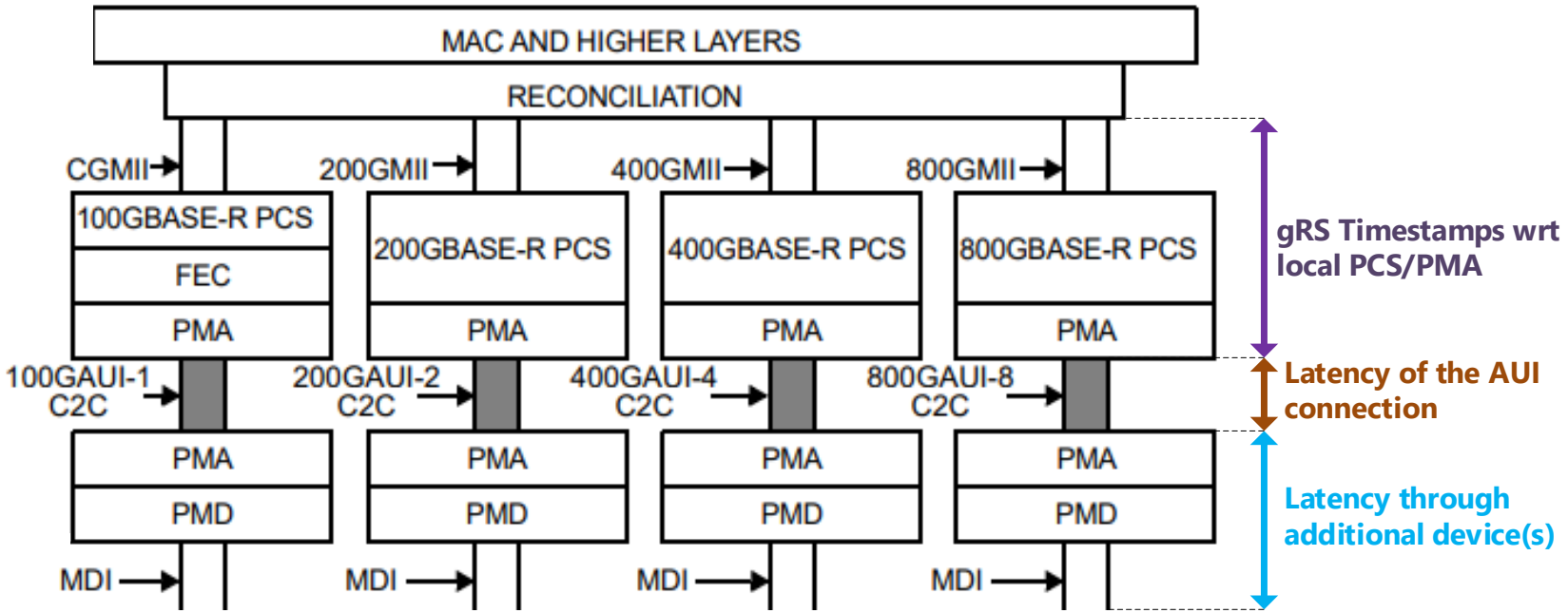


- A timestamp is simply a packet's arrival or departure time
- 802.3 specifies that the timestamp is calculated as the time at the Generic Reconciliation Sublayer (gRS), and then to adjusted to account for the delay through the PHY (path data delay).
- 802.3cx explains how to account for the cyclic delay variations through the PHY (e.g. FEC parity bits) and use a constant value for the path data delay.

- Note that the gRS has awareness of the Alignment Marker position inserted (Tx) or deleted (Rx) by the PCS
  - In 802.3cx, this is done by "passing notes" about the AM location from the PCS
  - Passing such notes isn't possible if the PCS is not in the same physical device as the gRS

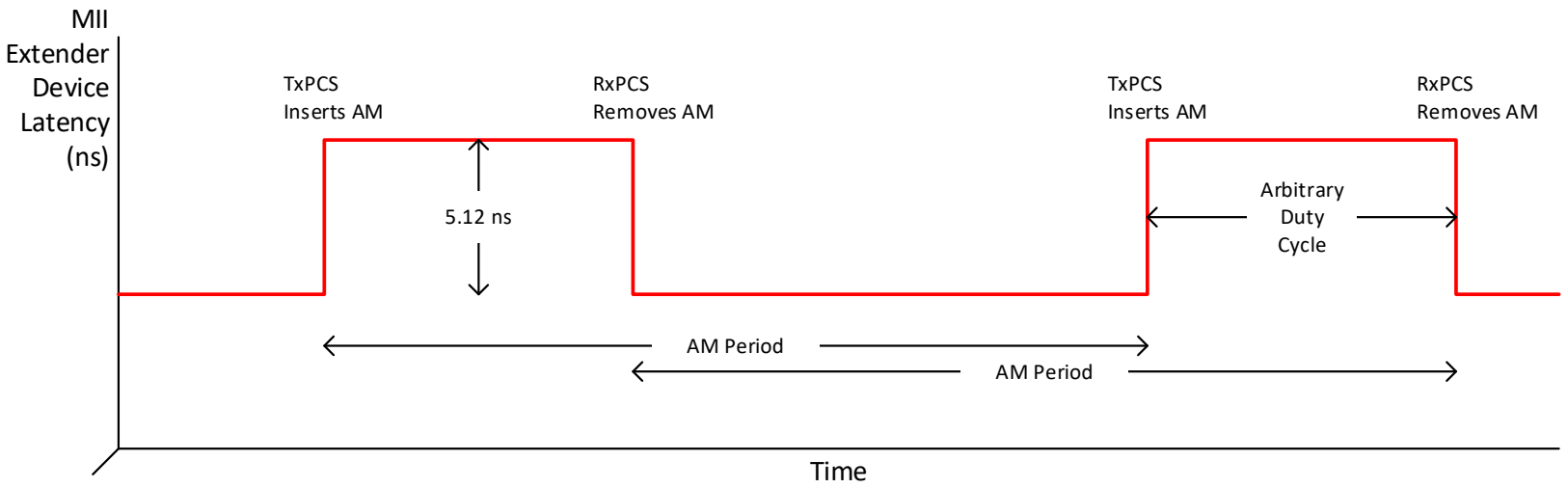
Example with a C2C/C2M AUI connection

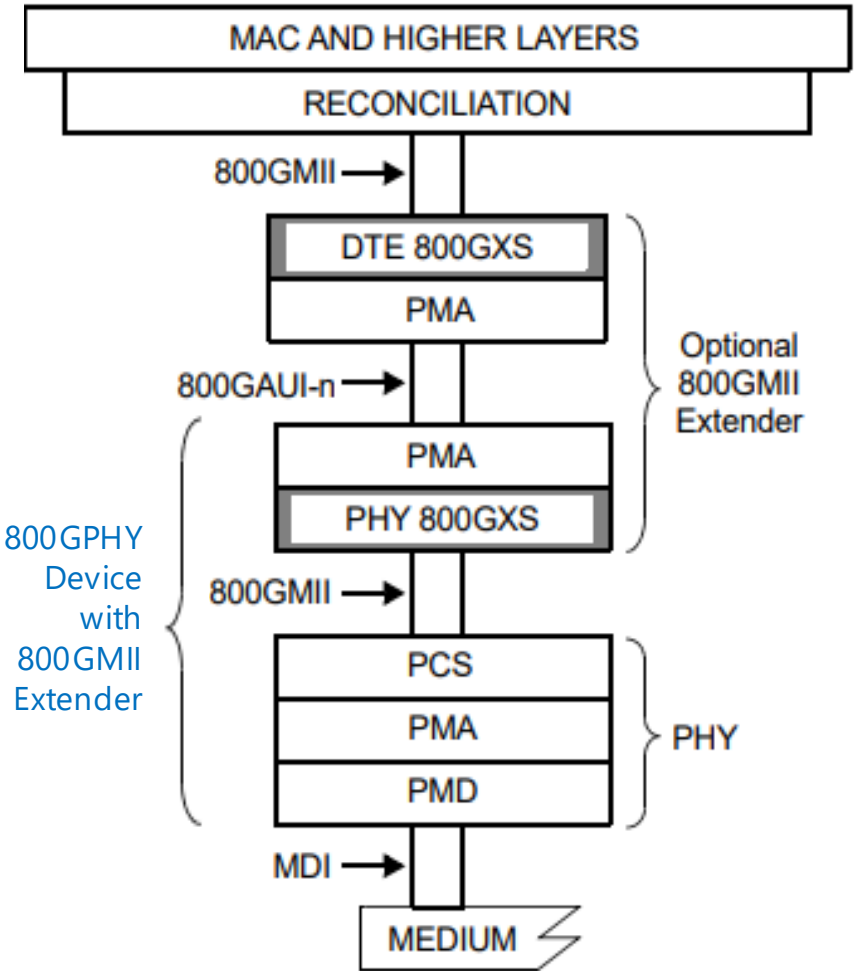
- As long as the device in the path does not alter the MII sequence, the MII <-> MDI delay can be calculated piece-meal.
  - Use the calculated Path Data Delay with respect to the local device.
  - Add the AUI Latency
  - Add the Latency of the additional device(s)
  - The final path data delay is the sum of the segments' latencies.
  - Can be applied as an offset in the gRS timestamp.



The latency across a PHY device with an MII-Extender is NOT Constant

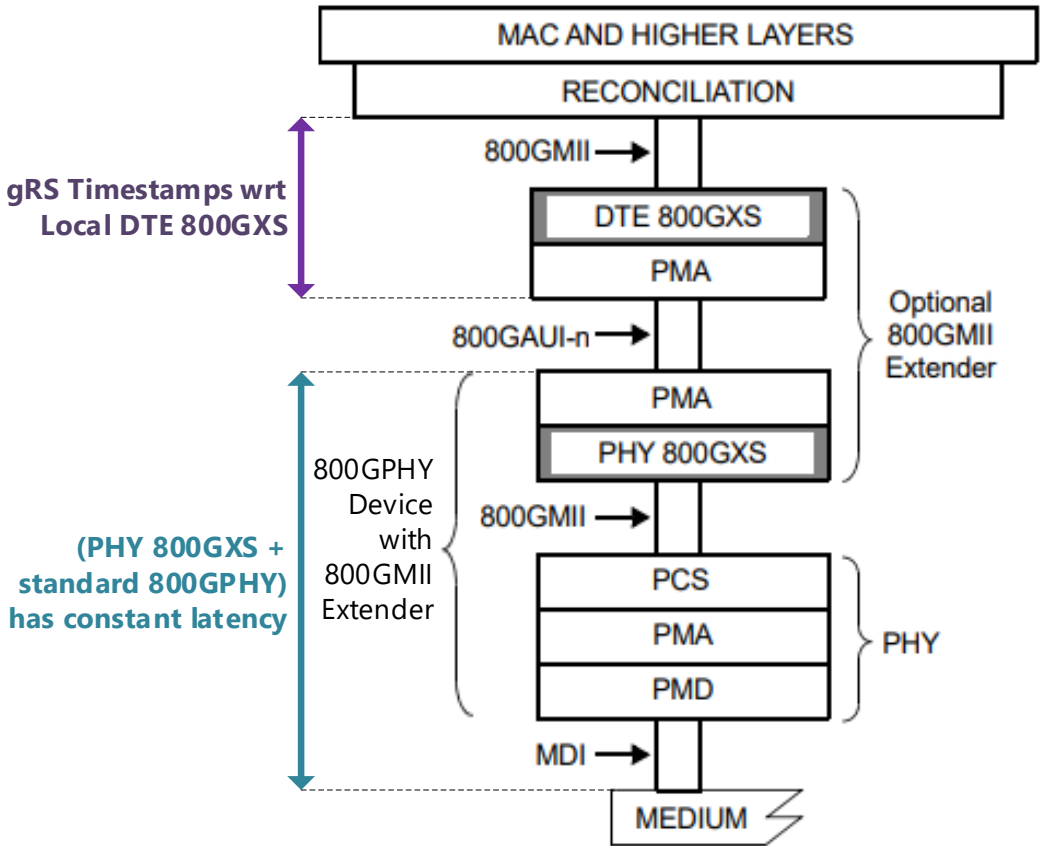
- An "PHY Device with an MII-Extender" comprises
  - standard PHY (MDI-facing)
  - PHYXS+PMA (AUI-facing)
- Device removes (Rx) and adds (Tx) alignment markers, so its effective latency varies over time
- Across an 800G PHY device with an 800GXS, the latency varies by 5.12ns
- The relative phase of the AMs across the device is completely arbitrary





Implementation Workaround

- If, across the PHY device, the alignment marker position relative to the MII is the same for incoming/outgoing interfaces, then the effective latency of the MII-extender device is *constant*
  - Only applicable when the MII rate is unchanged through the PHY device
  - Means that idle addition/removal is not needed
- Piecemeal approach to Path Data Delay (MII <-> MDI delay) is thus valid.
  - the gRS can calculate the path data delay with respect to the local DTE 800GXS/PMA
  - The gRS timestamps take the AM position at the local DTE 800GXS into account
  - Additional segments have constant latency
- Overall, the workaround results in a completely compliant signal at the MDI.



There is no explicit provision to allow such a workaround

- Only hinted at in 802.3cx D3.3 90.7.2:

NOTE 5—When TX_NUM_BIT_CHANGE and RX_NUM_BIT_CHANGE are not available (e.g., over physical interfaces such as instantiated xMII or AUI), it is recommended to avoid insertion and removal of Idles, alignment markers, and codeword markers in the sublayers below the xMII/AUI, when possible, to reduce timestamping accuracy impairments (see Annex 90A).

- What is missing is a note in Clause 171 to point out that:
  - Removing and adding alignment markers across a PHY device with an MII extender can be problematic for timestamping
  - Maintaining the AM position across such a device (and foregoing idle compensation) is desirable, while being fully compliant to 802.3.

Conclusion

- Avoiding the latency variation across an 800G PHY + XS due to alignment marker removal and insertion is straightforward and compliant with the current 802.3df draft.
- But the need/ability to do so may not be obvious to most 802.3df readers.
- A note relating specifically to 800GMII-Extenders in Clause 171 could clarify this.

Thanks!

Appendix (From 802.3cx CFI): Application Timing Requirements

Classes C and D were added in 2018 for 5G transport applications

- From ITU-T Recommendation G.8273.2, Timing characteristics of telecom boundary clocks and telecom slave clocks
  - Specifies the max timing errors that can be added by a telecom boundary clock
  - cTE: constant time error
  - dTE_L: low-passed dynamic time error
    - MTIE: Maximum Time Interval Error
    - TDEV: Time Deviation
  - TE_L: constant time error + low-passed dynamic time error
  - TE: constant time error + unfiltered dynamic time error

Time Error Type	Class	Requirement (ns)
max|TE|	A	100
	B	70
	C	30
	D	for further study
max|TE_L|	A, B, C	not defined
	D	5

Class	cTE Requirement (ns)
A	±50
B	±20
C	±10
D	for further study

Time Error Type	Class	Requirement (ns)	Observation interval τ (s)
dTE_L	A and B	MTIE = 40	$m < \tau \leq 1000$ (for constant temp)
	A and B	MTIE = 40	$m < \tau \leq 10000$ (for variable temp)
	C	MTIE = 10	$m < \tau \leq 1000$ (for constant temp)
	D	MTIE = for further study	
	A and B	TDEV = 4	$m < \tau \leq 1000$ (for constant temp)
	C	TDEV = 2	
	D	TDEV = for further study	