Microsemi
a Microchip company

PTP Timestamping clarifications_

Steve Gorshe, Richard Tse, Winston Mok

IEEE 802.3 2018

Introduction

- Clarifications are needed in clause 90 to unambiguously define how Alignment Marker functions affect PTP timestamping

1. Clarify Tx and Rx Path Data Delay
2. Clearly specify how AM and Idle insertion/deletion affect PTP timestamps
3. Clarify how to account for the lane distribution impact on the latency difference between the MII and the PHY of each lane

New Complicated PHYs

- Newer PHYs (200GBASE-R PHY is shown) are more complicated with respect to timestamping
  - Timestamp events are captured at the xMII
  - AM insertion/removal occurs between the xMII and MDI
  - The data shift resulting from AM insertion/removal creates new considerations for accurate timestamp transmission
  - There are different allowable methods for creating the bandwidth needed for AM insertion

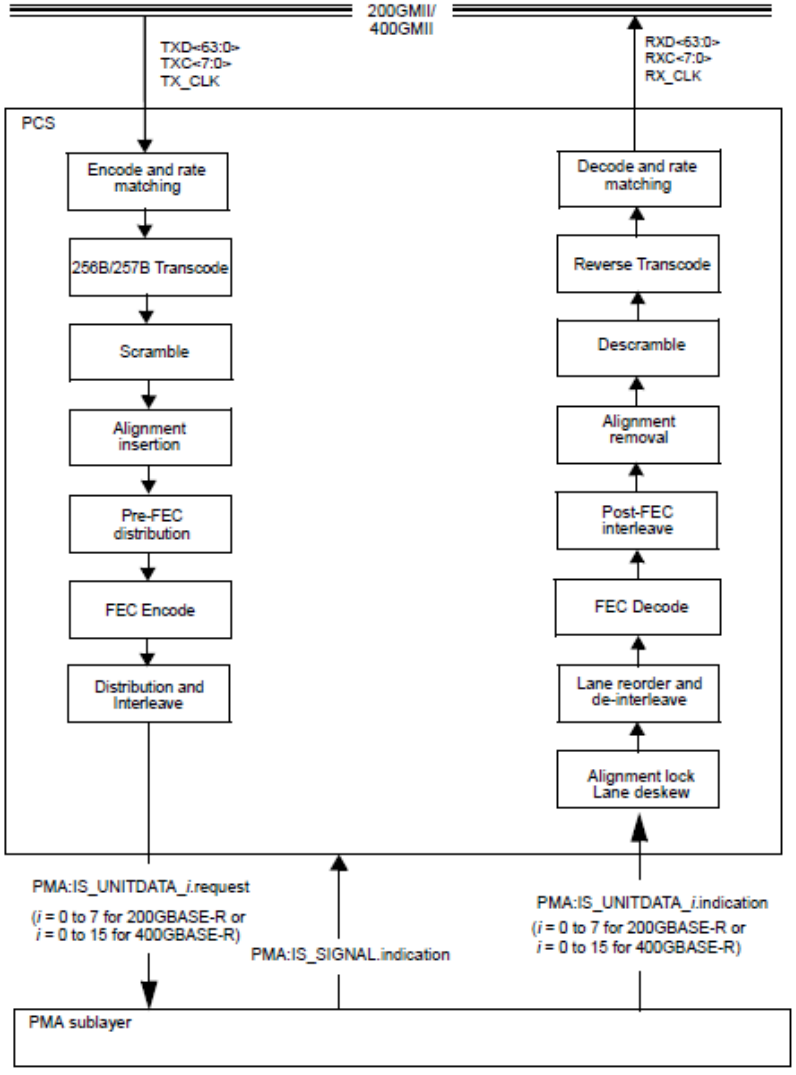


Figure 119–2—Functional block diagram

Tx and Rx Path Data Delay: Background

- Subclause 90.7 states
  - "The transmit path data delay is measured from the input of the beginning of the SFD at the xMII to its presentation by the PHY to the MDI. The receive path data delay is measured from the input of the beginning of the SFD at the MDI to its presentation by the PHY to the xMII."

however…

- Subclause 7.3.4.1 of IEEE 1588v2 and subclause 11.3.9 of IEEE 802.1AS define the message timestamp point as follows:
  - "the message timestamp point for an event message shall be the beginning of the first symbol after the Start of Frame (SOF) delimiter."

Tx and Rx Path Data Delay: The Problem

- When AMs are part of the PCS, the “beginning of the SFD” and the “beginning of the first symbol after SFD” could have path data delays that are different by a non-constant value, depending on whether or not AMs separate these two symbols

Tx and Rx Path Data Delay: Example

- Figure 82-5 defines the format of the control block that contains the implied /S/ start of packet delimiter
- As described in clause 81.2.2, the SFD occurs in the last 8 bits of this control block.
- Consequently, the “beginning of the SFD” and the “beginning of the first symbol after the Start of Frame (SOF) delimiter” are in different 66B blocks.
- The 66B block containing the /S/ and SFD can appear in any of the four different locations within the 256B/257B block code.
- Per clause 119.2.4.4, “The alignment marker group … interrupts any data transfer that is already in progress.”
- Consequently, if the AM insertion occurs immediately following a 257B block in which the SFD is carried at the end of that block, the AM adds ≈5ns of latency (from the 4×257 AM block bits for 200GbE and 8×257 for 400GbE) between the “beginning of the SFD at the xMII to its presentation by the PHY to the MDI.”
- ”

Tx and Rx Path Data Delay: Considerations

- Our understanding from 802.3 participants who were involved at the time is that 802.3 Clause 90.7 uses the reference point requested by 802.1AS.
  - Did 802.1AS request using a different reference bit position for the MII-MDI delay measurement than 802.1AS/1588 uses for the timestamp?
  - The editor of 802.1AS said they did not intend to request for any difference in 802.3
- If 802.3 moved the MII-MDI delay reference point to being the same as the 802.1AS timestamp reference point, would it cause any problems?
  - For PHYs without AMs, there should be no change to the PTP timestamping performance.
  - Can we change “beginning of the SFD” to “beginning of the first symbol after SFD” in clause 90.7?

Effect of AM and Idle insertion/removal on PTP Timestamping: The Problem

- Insertion/deletion of AMs and Idles can affect the time at which the PTP event message's timestamping point crosses the timestamping reference plane and consequently affect its timestamp
- A noteworthy number of legacy complex PHY implementations (with AM insertion/removal) are already in use
  - Not desirable to degrade the performance of these legacy PHYs

Effect of AM and Idle insertion/removal on PTP Timestamping: Considerations

- Figure 119-2 (see slide 3) shows a Tx functional flow of the gRS (where timestamps are inserted) followed by rate matching followed by AM insertion
  - Accommodating AM insertion is one of the potential reasons for rate matching
  - Rate matching is typically performed by Idle insertion/removal
- Strict adherence to the functional flow of Figure 119-2 implies that the timestamp event captured at the xMII will be shifted by the AM and Idle insertion/removal process, thus creating an inaccuracy for which (unlike FEC) there is no guarantee that the Rx processes will cancel it out in the same location of the received bit stream
- However, 802.3 has a long established principle of recognizing implementations as being compliant as long as the device output bit stream is the same as if the process described in the standard had been applied

Effect of AM and Idle insertion/removal on PTP Timestamping: Considerations (continued)

- The current standard allows both
  - Making room for AMs by deleting idles
  - Making room for AMs by other means without deleting idles
- To ensure accurate timestamps on a message-by-message basis
  - Departure (Tx) timestamp: some legacy implementations perform AM insertion and Idle insertion/removal before capturing the timestamp
  - Arrival (Rx) timestamp: some legacy implementations perform AM removal and Idle insertion/removal after capturing the timestamp
- CONCLUSION: Achieving accurate timestamp transfer in a manner that is compatible with all implementations requires using a message-by-message approach in which the timestamp at the xMII is modified to account for path data delay changes resulting from AM and Idle insertions/removals

Effect of AM and Idle insertion/removal on PTP Timestamping: Considerations (continued)

- Can the handling of AMs be clarified to ensure a consistent interpretation?
  - For example:
    “If the insertion or removal of AMs and/or Idles in these PCSs affects the transmit or receive data path delay, this effect must be accounted for in the timestamp. In this way, the timestamp operation is performed as if alignment markers are present at the xMII (i.e., as if AM insertion and Idle insertion/removal is performed ahead of the Tx xMII and AM deletion and Idle insertion/removal is performed after the Rx xMII).”

Impact of Lane Distribution: The Problem

- There are two inherent approaches for determining the MII-to-MDI delay on multi-lane PHYs
  1. Method 1 – Take into account the delay between the timestamp insertion at the MII and when it is transmitted at the PHY for the lane that will carry it.
  2. Method 2 – Use a constant delay regardless of which lane will carry the timestamp

Impact of Lane Distribution: Method 1

90.7 Data delay measurement

The TimeSync capability requires measurement of data delay in the transmit and receive paths, as shown in Figure 90–3. The transmit path data delay is measured from the beginning of the SFD at the xMII input to the beginning of the SFD at the MDI output. The receive path data delay is measured from the beginning of the SFD at the MDI input to the beginning of the SFD at the xMII output.

- For a multilane PHY, after deskew delays are accounted for appropriately and since timestamping is at the PMD, would the timestamps be the same regardless of which lane the message's timestamp reference point is transmitted on (or received on)?
  - Since all lanes are transmitted at the same time and received at the same time (after deskew) at the PMD, it would seem this is a valid conclusion.

Impact of Lane Distribution: Method 1 (continued)

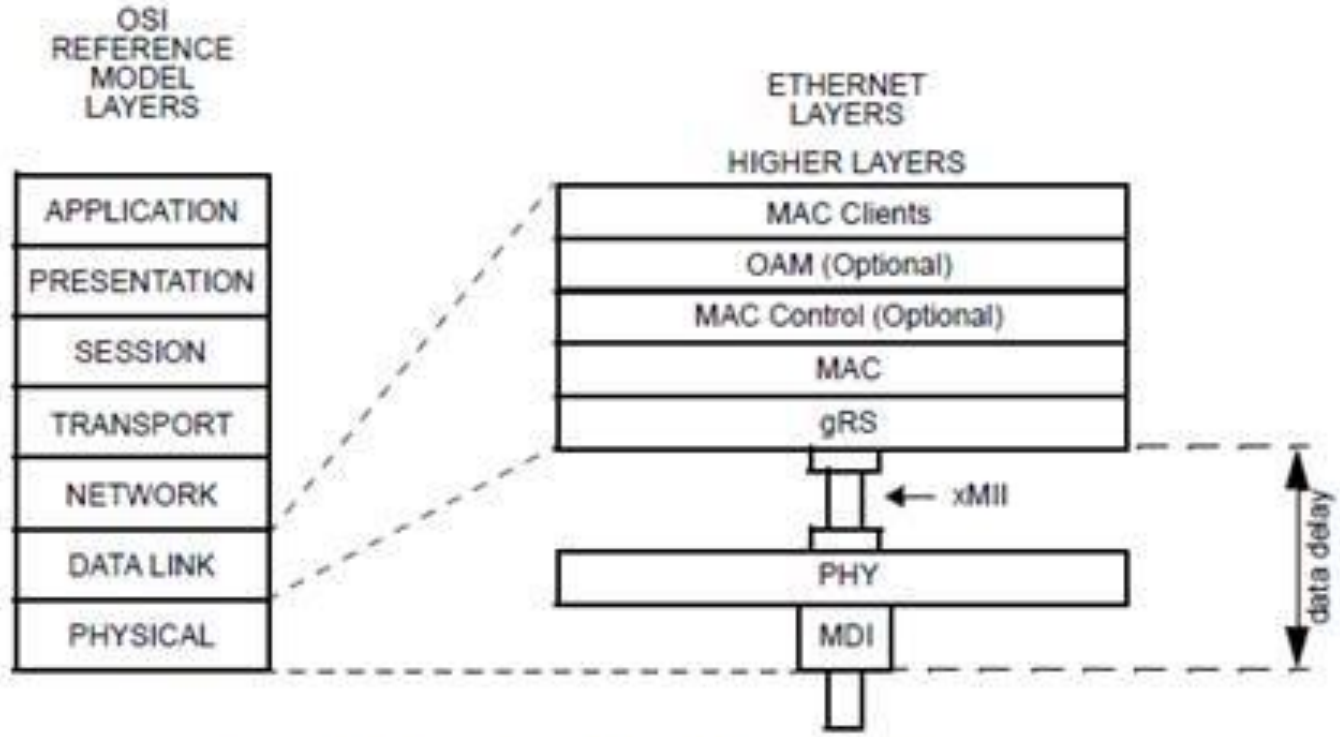
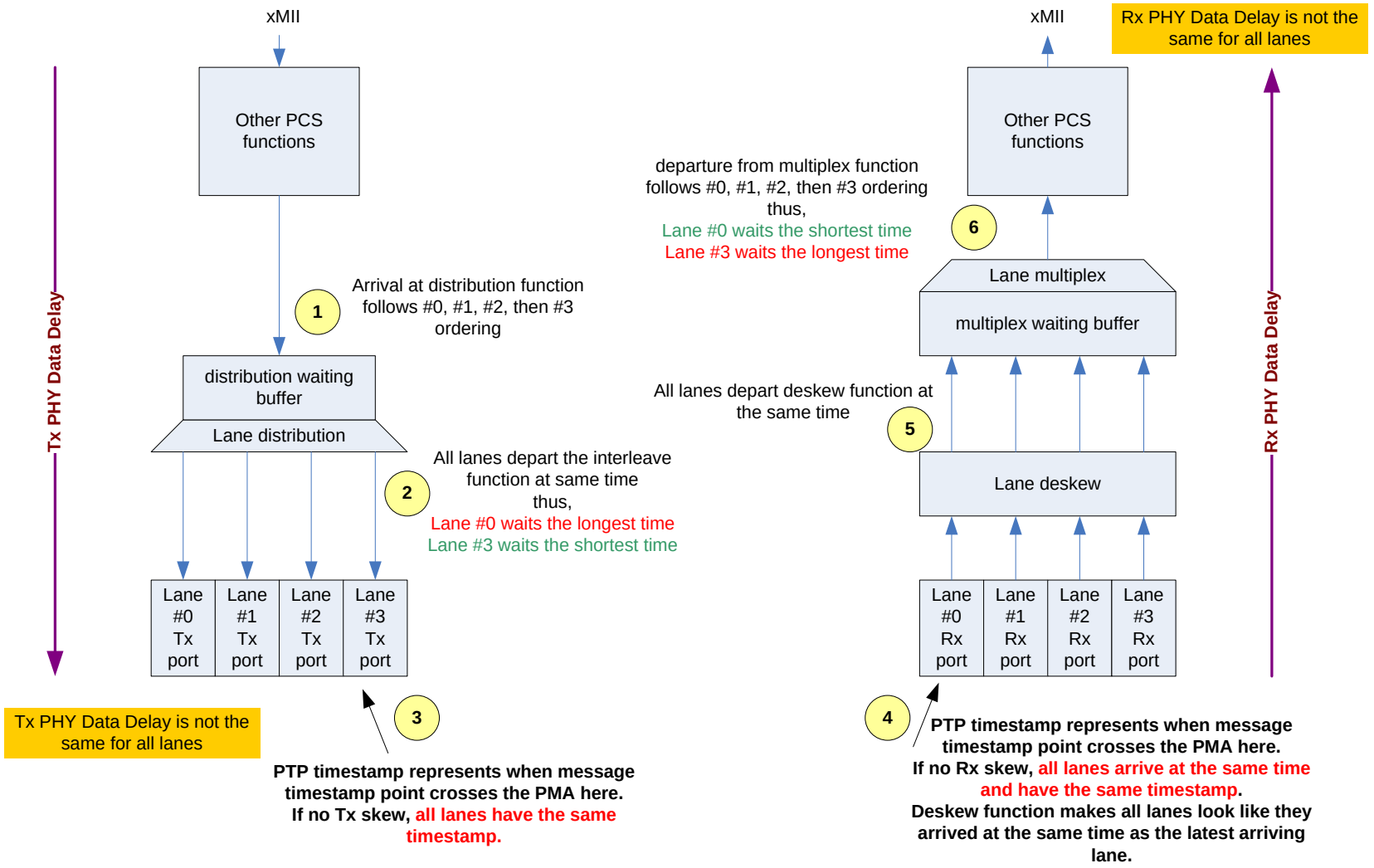


Figure 90–3—Data delay measurement

Impact of Lane Distribution: Method 1 (continued)

- However, this means that PHY data delay (between xMII and PMD, as per Figure 90-3 above) is not the same for every lane because the PMD-to-xMII multiplexing delay (for Rx) and xMII-to-PMD demultiplexing delay (for Tx) is different for each lane (as shown in Figures 82-3 and 82-4 below). In the Tx direction, codewords going to lane 0 have the most delay and codewords going to lane 3 have the least delay. In the Rx direction, the opposite is true. To capture an accurate timestamp at the xMII (as per the 802.3 model), the lane-based intrinsic delay must be included as part of the PHY data delay.
  - Was this the intent?

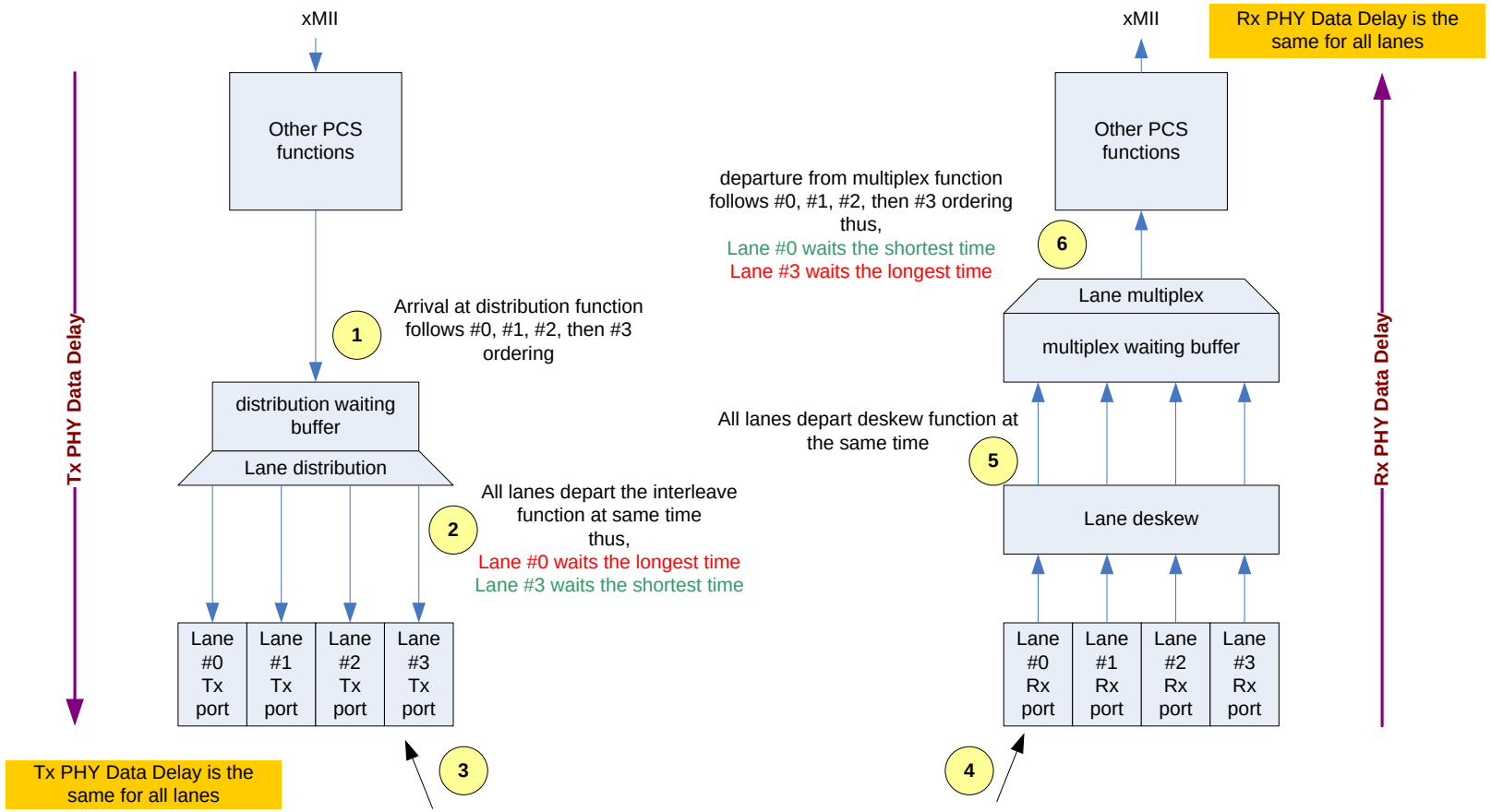
Impact of Lane Distribution: Method 1 (continued)



Impact of Lane Distribution: Method 2

- These multilane PHY data delays could also be designated to be a constant value for all lanes if the principle that is used for FEC's varying intrinsic delays is applied for multilane's multiplexing/demultiplexing varying intrinsic delays.
  - i.e., the Tx intrinsic demultiplexing delay is balanced by the Rx multiplexing intrinsic delay, making the aggregated demux/mux delay a constant.
  - Was this principle on anyone's mind when the multiplane PHY function was defined?

Impact of Lane Distribution: Method 2 (continued)



Impact of Lane Distribution: Considerations

- Which of these two methods, or other method, was intended?
- Is method #1 is the most-commonly used method?
- Can we clarify this in 802.3?