Improving PTP Timestamping Accuracy on Ethernet Interfaces Call For Interest Consensus Presentation

Steve Gorshe, Richard Tse

IEEE 802.3

Vienna, Austria

July 2019

CFI Objective

- To assess the support for the formation of the “Improving PTP Timestamping Accuracy on Ethernet Interfaces” Study Group in IEEE 802.3 to consider the development of a PAR and CSD to address high accuracy time transport for IEEE 802.3 Ethernet

- We will not:
  - Fully explore the problem
  - Debate strengths and weaknesses of solutions
  - Choose a solution
  - Create a PAR or 5 Criteria
  - Create a standard or a specification

- Anyone in the room may speak or vote

Agenda

- Ethernet for 5G Transport
- What's the Problem?
- Transport Timing
  - Legacy 4G RAN
  - New 5G C-RAN
  - Timing Requirements
  - Timing Consequences
- Why Can't High Accuracy Time Transport be Achieved Now?
  - PHY data delay variation
  - Basic PTP time distribution
  - Basic PTP time distribution with IEEE 802.3
  - Potential Areas of Improvement
  - Resulting Performance vs Target Performance
- What if IEEE 802.3 Doesn't Act?
- Contributors and Supporters
- Q&A
- Straw Polls

Ethernet for 5G Transport (1)

- Why Ethernet?
  - Packet-based transport supports load balancing on computing resources, which is a vital quality for the Centralized Radio Access Network (C-RAN)
  - Eco-system is mature
    - "…lowers cost by leveraging existing, mature packet-based solutions (e.g. Ethernet) for vital functions, such as QoS, synchronization, and data security" – IEEE P1914.1
  - Offers wide and sufficient range of capacities (10GE to 100GE, and eventually up to 400GE will be used)

Ethernet for 5G Transport (2)

- Ethernet has already been chosen for the 5G transport application and the market is huge!
  - "Researcher estimates that global **investments on C-RAN** architecture networks **will reach over $7 Billion by the end of 2016**. The market is further expected to grow at a **CAGR of nearly 20%** between 2016 and 2020. These investments will include spending on RRHs (Remote Radio Heads), BBUs (Baseband Units) and fronthaul transport networking gear." – Business Wire, Jan 2016
  - "Worldwide mobile **fronthaul equipment revenue totaled $787 million in 2016**, with the majority coming from Asia Pacific; the market is experiencing modest scale but has long-term potential as solutions **evolve from Common Public Radio Interface (CPRI) based to Ethernet based**. The global mobile fronthaul equipment market is forecast to grow at a compound annual growth rate **(CAGR) of 26.4 percent** from 2017 to 2021, when it **will reach $2.5 billion**" – IHS Markit, Nov 2017
  - "SNS Research estimates that global **investments in C-RAN** architecture networks **will reach nearly $9 Billion by the end of 2017**. The market is further expected to grow at a **CAGR of approximately 24%** between 2017 and 2020. These investments will include spending on RRHs (Remote Radio Heads), BBUs (Baseband Units) and fronthaul transport network equipment" – SNS Research, July 2017
  - "The **C-RAN is emerging as the critical network architecture for 5G**, it has innovative elastic and scalable network architectures which can provide the required capabilities to the incorporation of 5G network." – Grand View Research

Ethernet for 5G Transport (3)

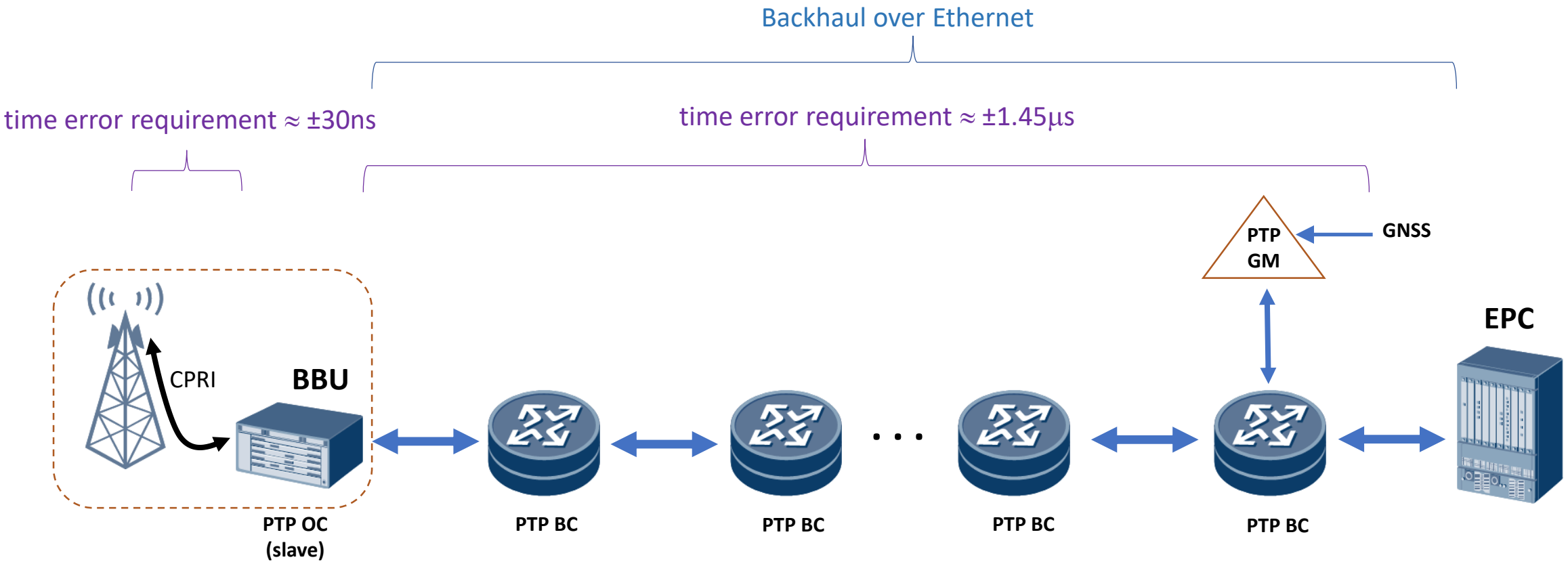
- New standards for fronthaul all use Ethernet as the transport layer and use Ethernet and IP-over-Ethernet encapsulated messages:
  - IEEE P1914.1: Draft Standard for Packet-based Fronthaul Transport Networks
  - IEEE 802.1CM: Time-Sensitive Networking for Fronthaul
  - IEEE 1914.3: Standard for Radio over Ethernet Encapsulations and Mappings
  - O-RAN Fronthaul Working Group: Control, User and Synchronization Plane Specification
  - CPRI: eCPRI Specification
  - IEEE 802.3cn: physical layers for 50/200/400Gb/s over 40km SMF
  - IEEE 802.3ct: physical layers for 100/400Gb/s over 80km DWDM
  - IEEE 802.3cp: bidirectional transmission at 10/25Gbps over 10/20/40km single fiber strand

Great! So What's the Problem?

- IEEE 1588-2008 (PTP, Precision Time Protocol) and associated ITU specifications on PTP and profiles of PTP (over Ethernet and over IP-over-Ethernet) are used for time synchronization in the 5G transport standards
- 5G's C-RAN based systems require high accuracy time synchronization for good radio performance

- But…
  - Current specifications in clause 90 (Ethernet support for time synchronization protocols) of IEEE 802.3 could reduce Ethernet's ability to support high accuracy time transport and, thus, limit its use in C-RAN applications

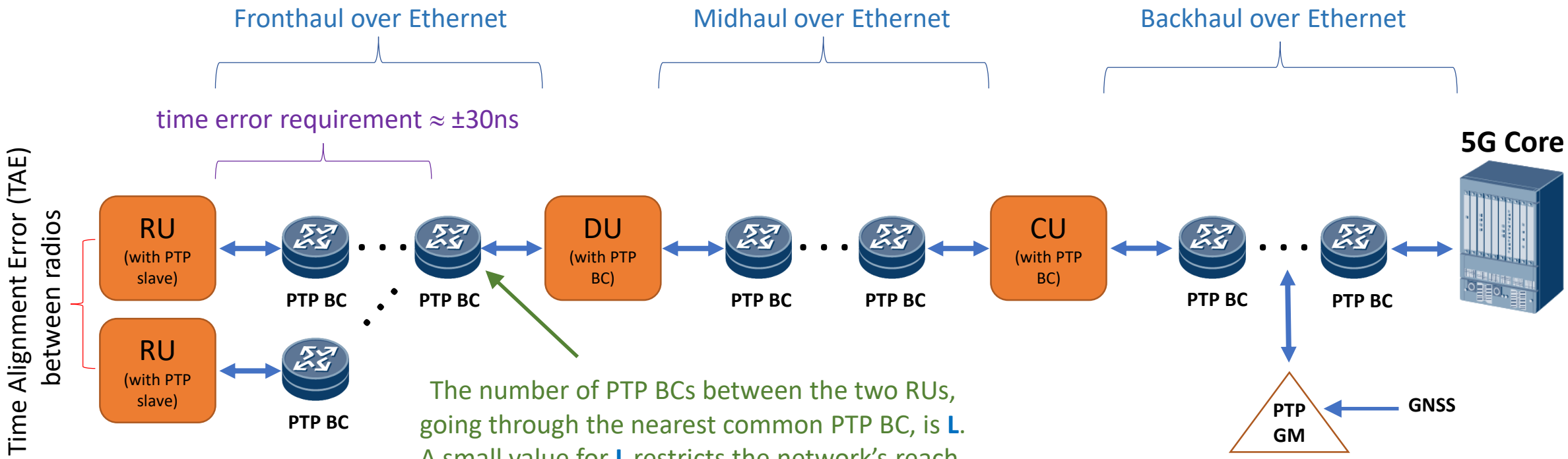
Transport Timing for 4G Radio Access Networks (RAN)



- BBU – baseband unit
- CPRI – Common Public Radio Interface
- EPC – enhanced packet core
- GNSS – Global Navigation Satellite System
- PTP BC – PTP boundary clock
- PTP GM – PTP grandmaster
- PTP OC – PTP ordinary clock

Transport Timing for 5G Centralized-RAN (C-RAN)

- C-RAN separates the BBU into “centralized” elements (Distributed Units (DUs) and Central Units (CUs)), allowing their resources to be efficiently shared between the Remote Units (RUs, radios)
- 5G mmWave NR (New Radio) has short reach (i.e. are densely packed) and high capacity
  - These qualities cause a need for a substantial fronthaul network (i.e. more timing hops) to connect RUs to their DUs



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	

Application Timing Consequences

- ITU Q13/SG15 WD13-25 shows why improved PTP performance is needed:
  - For radio time alignment error (TAE) of 260ns (see “TAE” in the figure on slide 9):
    - With all Class B Boundary Clocks everywhere, including in the RUs,
      L = 1 (only direct connect can satisfy requirements!)
    - With all Class C Boundary Clocks in network and class B Slave Clocks in the RUs,
      L = 5
    - With all Class C Boundary Clocks in network and “class C-like” Slave Clocks in the RUs,
      L = 7
    - If results were expanded to use class D Boundary Clocks in network and “class C-like” Slave Clocks in the RUs, L > 17

- To build a practical C-RAN network for 5G applications, PTP Clock performance should be Class C or better

Why Can't High Accuracy Time Transport be Achieved Now with IEEE 802.3?

- PTP timestamping is done at the MDI
- IEEE 802.3's timestamping is done at the xMII (per clause 90 of IEEE 802.3)
- PHY data delay must be known for the PTP message to move the timestamp from xMII to MDI
- Many newer 802.3 PHYs have fundamental dynamic variations in their data delay

- But
  - Data delay variations in the PHY are not inherently visible at the xMII

- Thus
  - IEEE 802.3's current timestamping mechanism does not inherently support high accuracy on PHYs with data delay variations
  - Specifications are needed on how to deal with each data delay variation

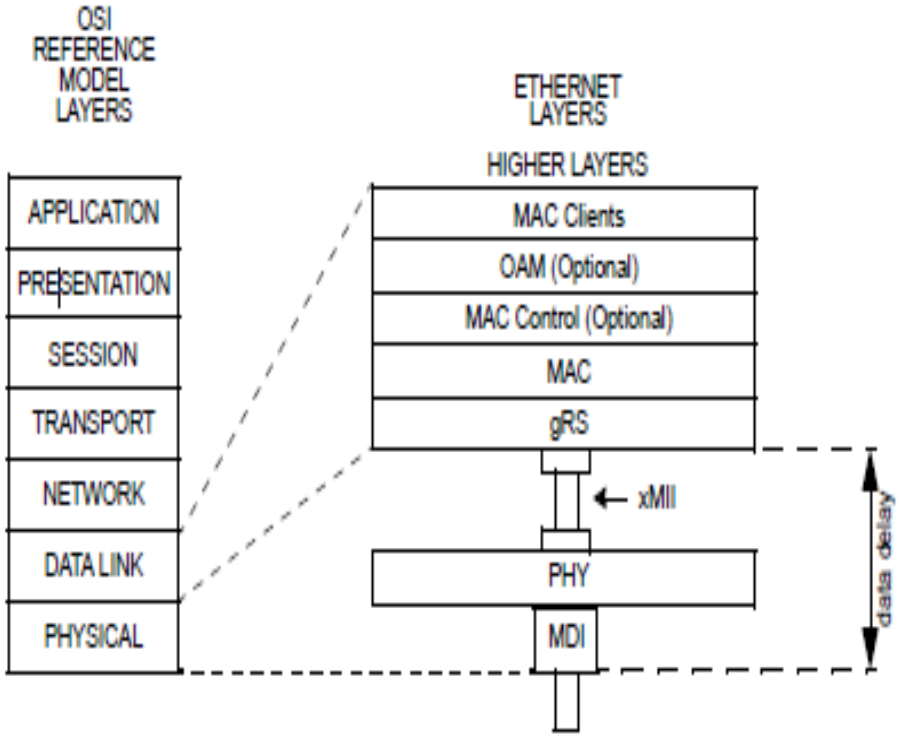


Figure 90–3—Data delay measurement

Data Delay Variations in 100GE PHY

Timestamps are captured at xMII

Block distribution to multi-PCS lanes, Alignment Marker insertion/removal (and their corresponding Idles), and FEC all inherently cause dynamic data delay variation

Timestamps should correspond to the time at MDI

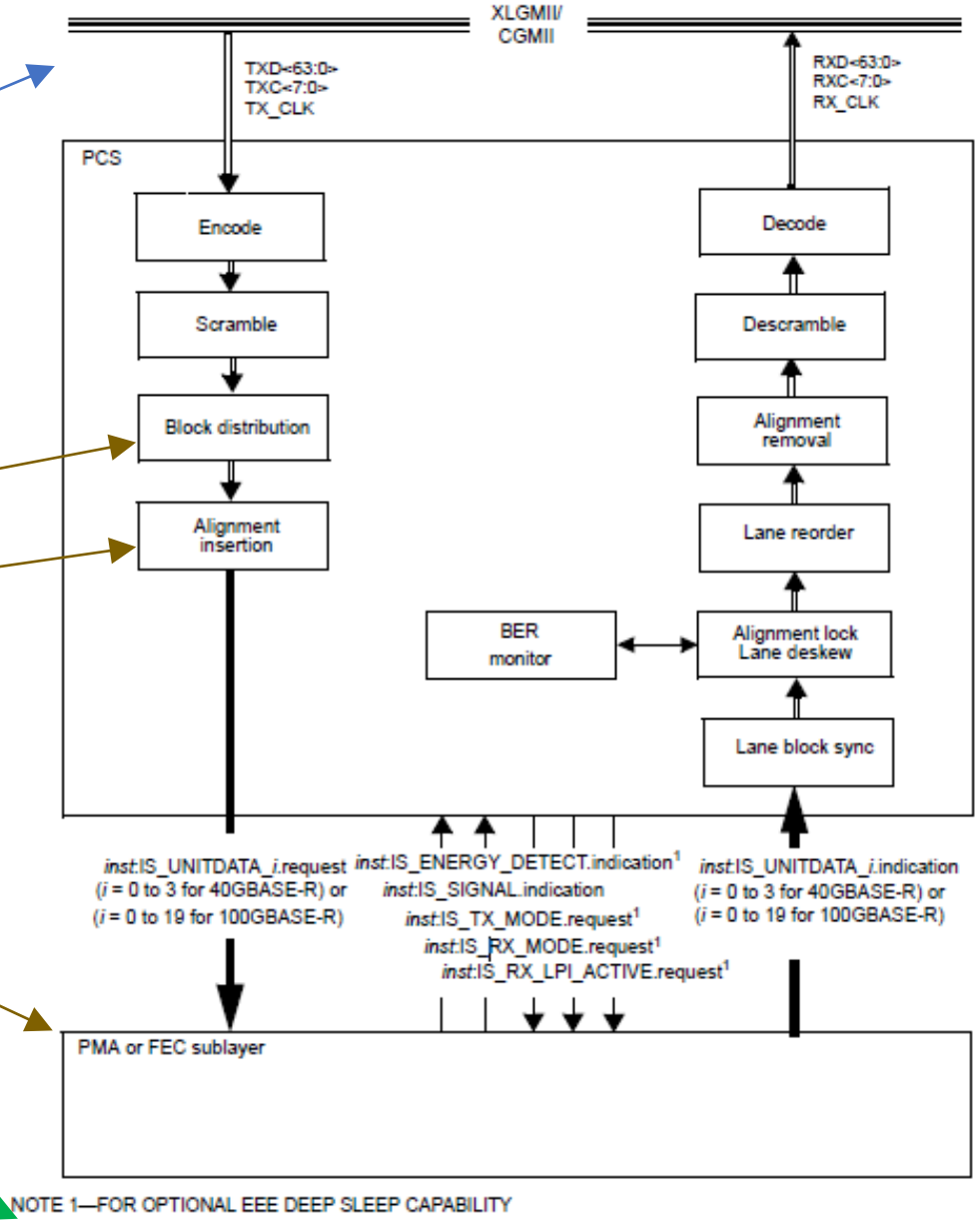
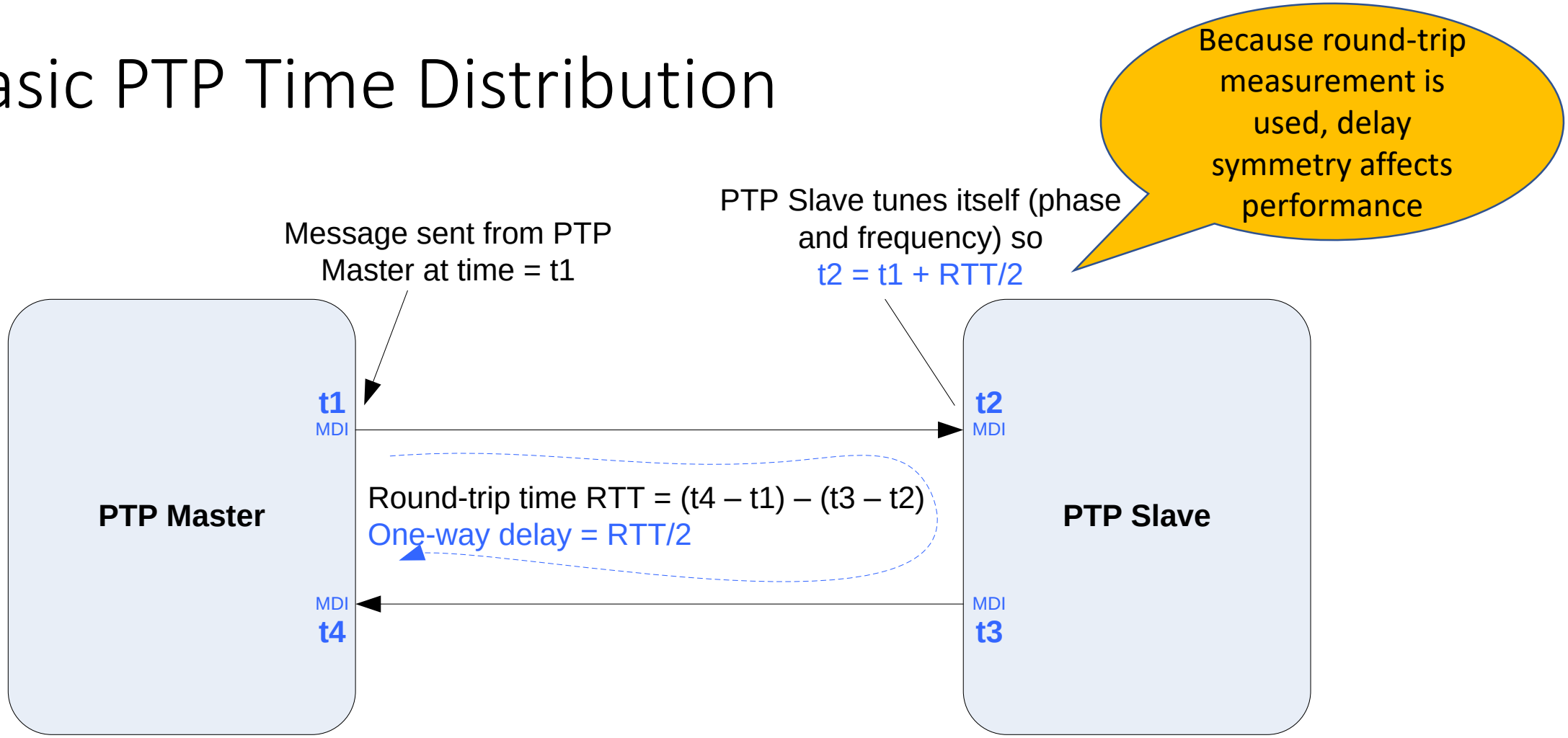


Figure 82–2—Functional block diagram

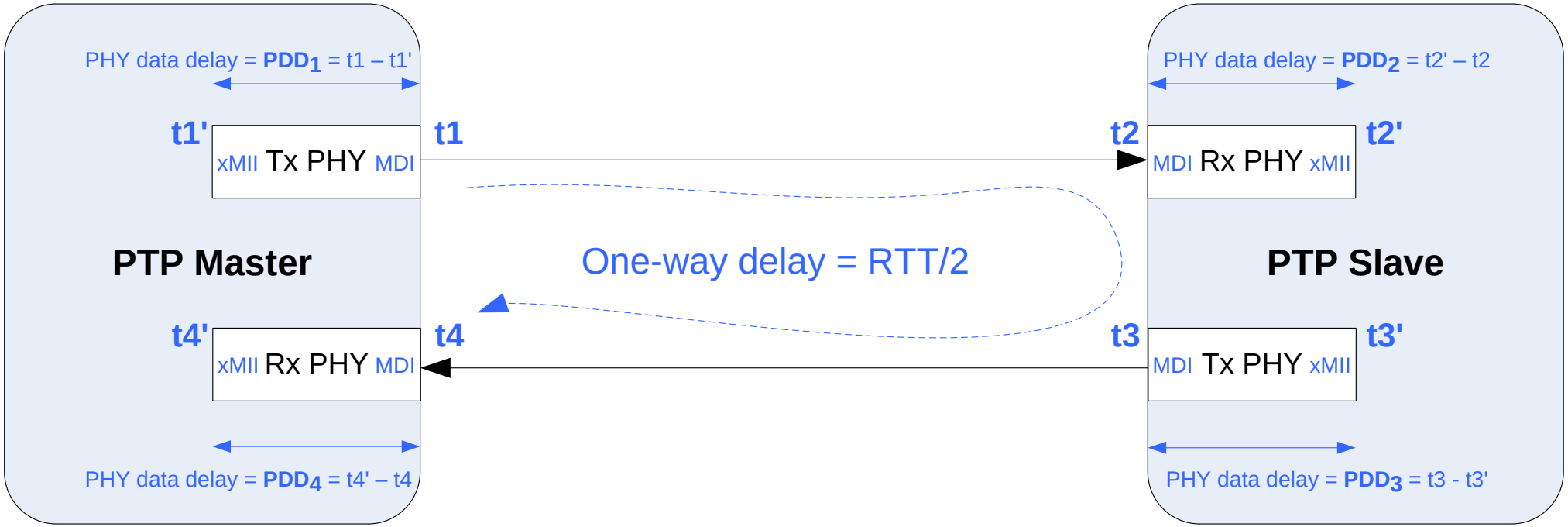
Basic PTP Time Distribution



-Timestamps **t1** and **t4** (corresponding to MDI) are captured at the PTP Master
-Timestamps **t2** and **t3** (corresponding to MDI) are captured at the PTP Slave
-All timestamps are given to the PTP Slave so it can:

- calculate RTT
- do adjustments to make t2 = t1 + RTT/2

PTP Time Distribution with IEEE 802.3



t1', **t2'**, **t3'**, and **t4'** are captured at the IEEE 802.3 xMII interfaces
t1, **t2**, **t3**, and **t4** are derived from **t1'**, **t2'**, **t3'**, and **t4'** using the corresponding **PHY data delay** (**PDD_x**)

Round-trip time RTT = (t4 – t1) – (t3 – t2)

$$= ((t4' - PDD_4) - (t1' + PDD_1)) - ((t3' + PDD_3) - (t2' - PDD_2))$$

To get an accurate RTT value, the following **PHY data delays must be known** for each PTP event message:

- All corresponding PDD_x or
- $(PDD_1 + PDD_2)$ or $(PDD_3 + PDD_4)$

Potential Areas for Improvement

1. Message Timestamp Point in IEEE 802.3 is different from IEEE 1588 and IEEE 802.1AS
   - If endpoints timestamp different events, RTT result will be wrong
2. Path Data Delay variance caused by Alignment Marker and Idle insertion/removal events needs to be accounted for in a standardized manner
   - PDD_x, (PDD_1 + PDD_2), and (PDD_3 + PDD_4) values might change because these events insert or extract data within a PHY
   - These changes must be detected and handled consistently in all PHYs so an accurate RTT can be measured
3. Path Data Delay variance from multi-PCS lane distribution function needs to be accounted for in a standardized manner
   - The characteristics of PDD_x, (PDD_1 + PDD_2), and (PDD_3 + PDD_4) must be specified to allow consistency between interworking PHYs so an accurate RTT can be measured
4. Others (for the potential Study Group to find and determine)

See Appendix 3 for details on items 1 - 3

Resulting Performance vs Target Performance

- Target Max|TE| = 30ns for class C Telecom Boundary Clock (see slide 10)
  - See Appendix 3 for details on these potential sources of errors in IEEE 802.3 timestamping
  - In a system, there are other sources of TE, in addition to those from timestamping, that use up the allowance

Ethernet Rate	Path Data Delay Variation per Tx/Rx Interface (ns)				Total TE per Tx or Rx Interface (ns)	**Path Data Delay Variation Contribution to Max|TE|, per PTP Boundary Clock (ns)**
	mismatched SFD timestamp point	Idle insert/remove (per Idle)	AM insert/remove	Lane Distribution		
GE	8	16	N/A	N/A	24	48
10GE	0.8	3.2	N/A	N/A	4	8
25GE	0.32	1.28	2.56	N/A	4.16	8.32
40GE	0.2	1.6	6.4	4.8	13	26
100GE	0.08	0.64	12.8	12.16	25.68	51.36
200GE	0.04	0.32	2.56	2.24	5.16	10.32
400GE	0.02	0.16	2.56	2.4	5.14	10.28

100GE is very important for C-RAN

What if IEEE 802.3 Doesn't Act?

- Ethernet has already been chosen for the 5G transport application
- Vendors are already releasing high accuracy timestamping solutions to get into this market
- Could result in development of incompatible implementations, which will not interoperate properly to meet performance goals
  - The industry might settle on one or more unofficial (and not clearly specified) but de facto standards, based on the popularity of certain solutions
  - Performance might always remain risky when interworking between different devices
- Conclusion: To enable a successful 5G transport network to be built with Ethernet, IEEE 802.3 should improve its PTP timestamping specifications

Contributors and Supporters

- Ali Ghiasi, Ghiasi Quantum
- Bill Powell, Nokia
- David Chalupsky, Intel
- Denny Wong, Xilinx
- Dino Pozzebon, Microchip Technology
- Gary Nicholl, Cisco
- Jeff Slavik, Broadcom
- Kapil Shrikhande, Innovium
- Marek Hajduczenia, Charter Communications
- Mark Bordogna, Intel
- Mark Gustlin, Cisco
- Matt Brown, Independent
- Nitzan Dror, Marvell
- Pete Anslow, Ciena
- Pirooz Tooyserkani, Cisco
- Richard Tse, Microchip Technology
- Shawn Nicholl, Xilinx
- Sriram Natarajan, Cisco
- Steve Carlson, High Speed Designs
- Steve Gorshe, Microchip Technology
- Steve Trowbridge, Nokia

Q & A

- Does anyone have any questions or comments?

- Contact Info:
  - Steve.Gorshe@microchip.com
  - Richard.Tse@microchip.com

Straw Polls

- I would support formation of the “Improving PTP Timestamping Accuracy on Ethernet Interfaces” Study Group in IEEE 802.3 to consider the development of a PAR and CSD to address high accuracy time transport for IEEE 802.3 Ethernet
  - Total individuals in room: _____
  - Total supporters in room: _____
- I would participate in the “Improving PTP Timestamping Accuracy on Ethernet Interfaces” Study Group in IEEE 802.3:
  - Total individuals in room: _____
  - Total supporters in room: _____
- I believe my affiliation would support participation in the “Improving PTP Timestamping Accuracy on Ethernet Interfaces” Study Group in IEEE 802.3:
  - Total individuals in room: _____
  - Total supporters in room: _____

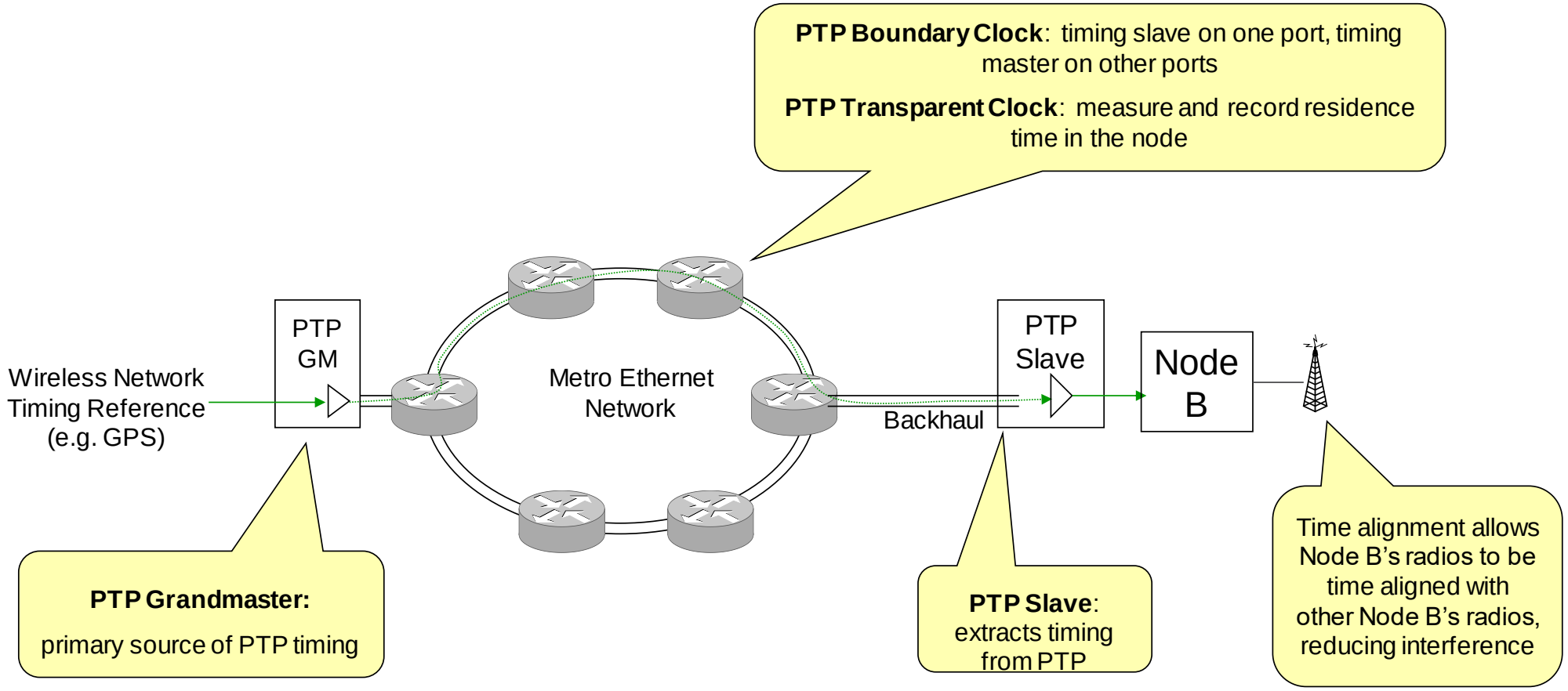
Outline of Appendices

1. PTP Fundamentals
   - PTP Application Example
   - Time Distribution Mechanism
   - Timestamp Generation Model
   - Time Error Measurement Model
2. Current State of Clause 90, IEEE 802.3
3. Potential Areas of Improvement in Support of High Accuracy Time Transport
   - History of Discussions and Contributions
   - Difference in Message Timestamp Point
   - AM and Idle Insertion/Removal
   - Multi-Lane PHY Ambiguities
   - Performance vs Target
4. More Details on PCS-Lane Distribution Delay

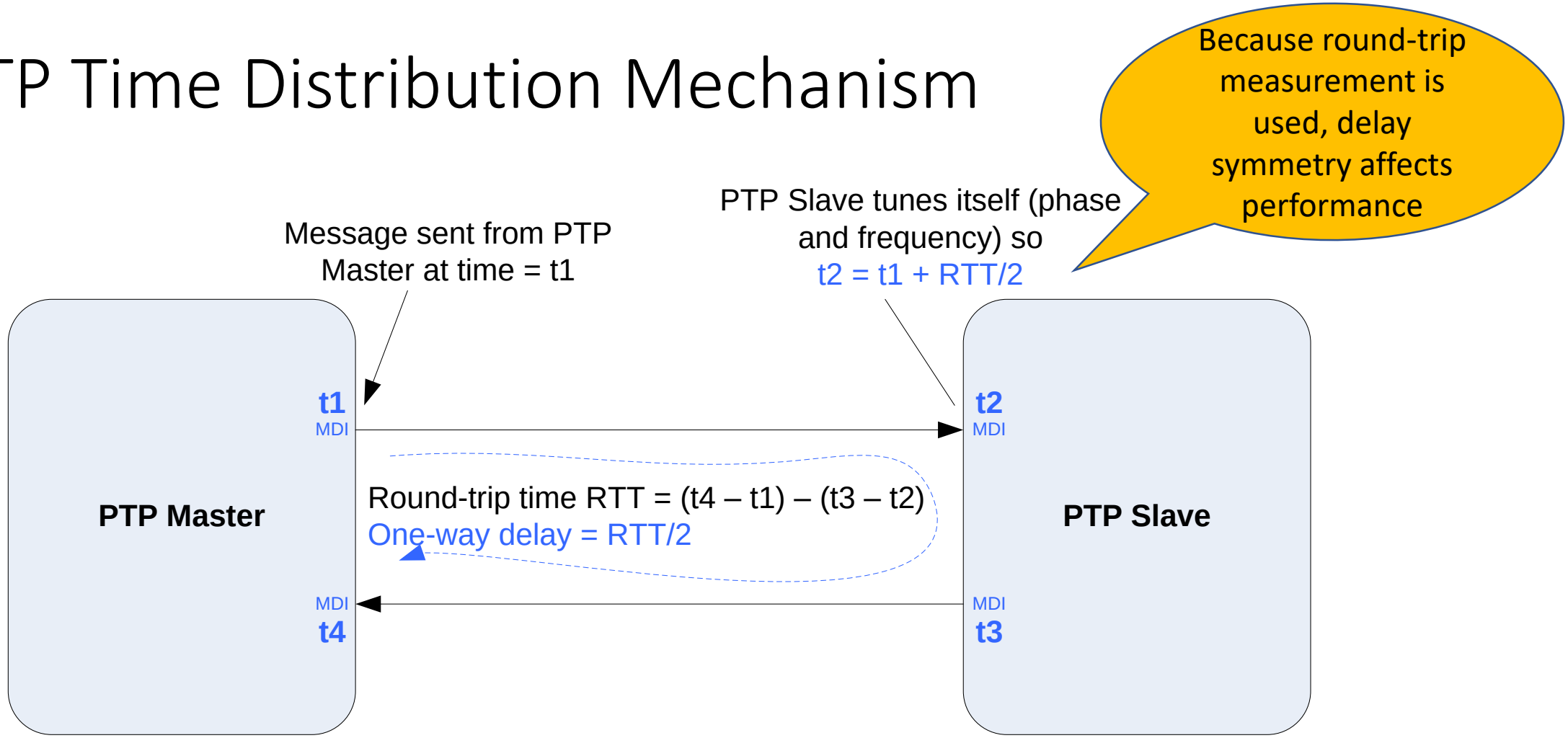
Appendix 1

PTP Fundamentals

PTP Application Example



PTP Time Distribution Mechanism

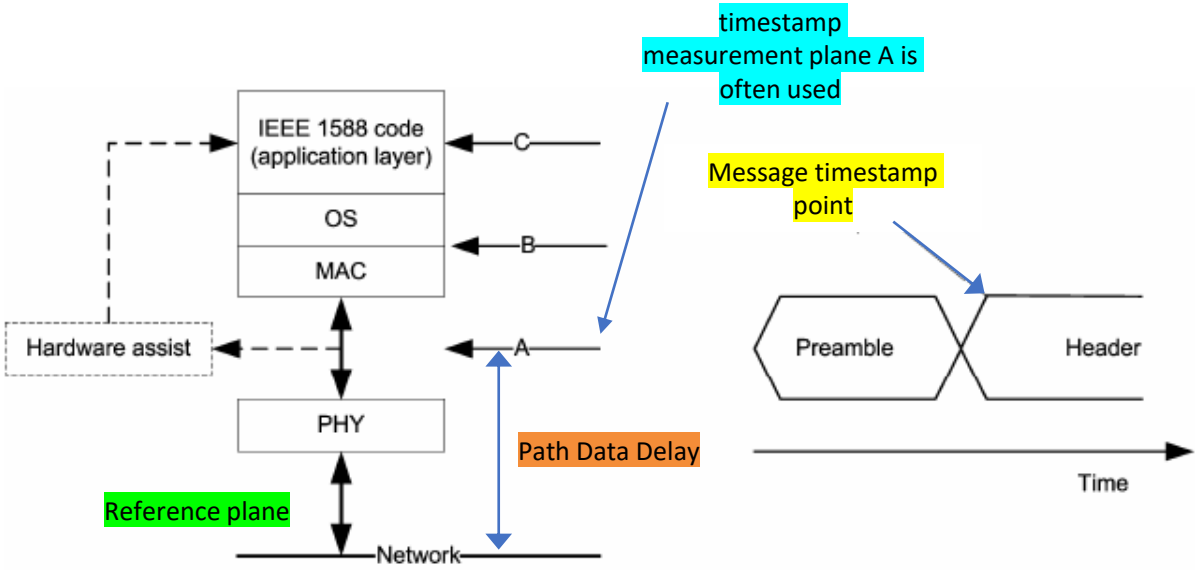


-Timestamps **t1** and **t4** (corresponding to MDI) are captured at the PTP Master
-Timestamps **t2** and **t3** (corresponding to MDI) are captured at the PTP Slave
-All timestamps are given to the PTP Slave so it can:

- calculate RTT
- do adjustments to make t2 = t1 + RTT/2

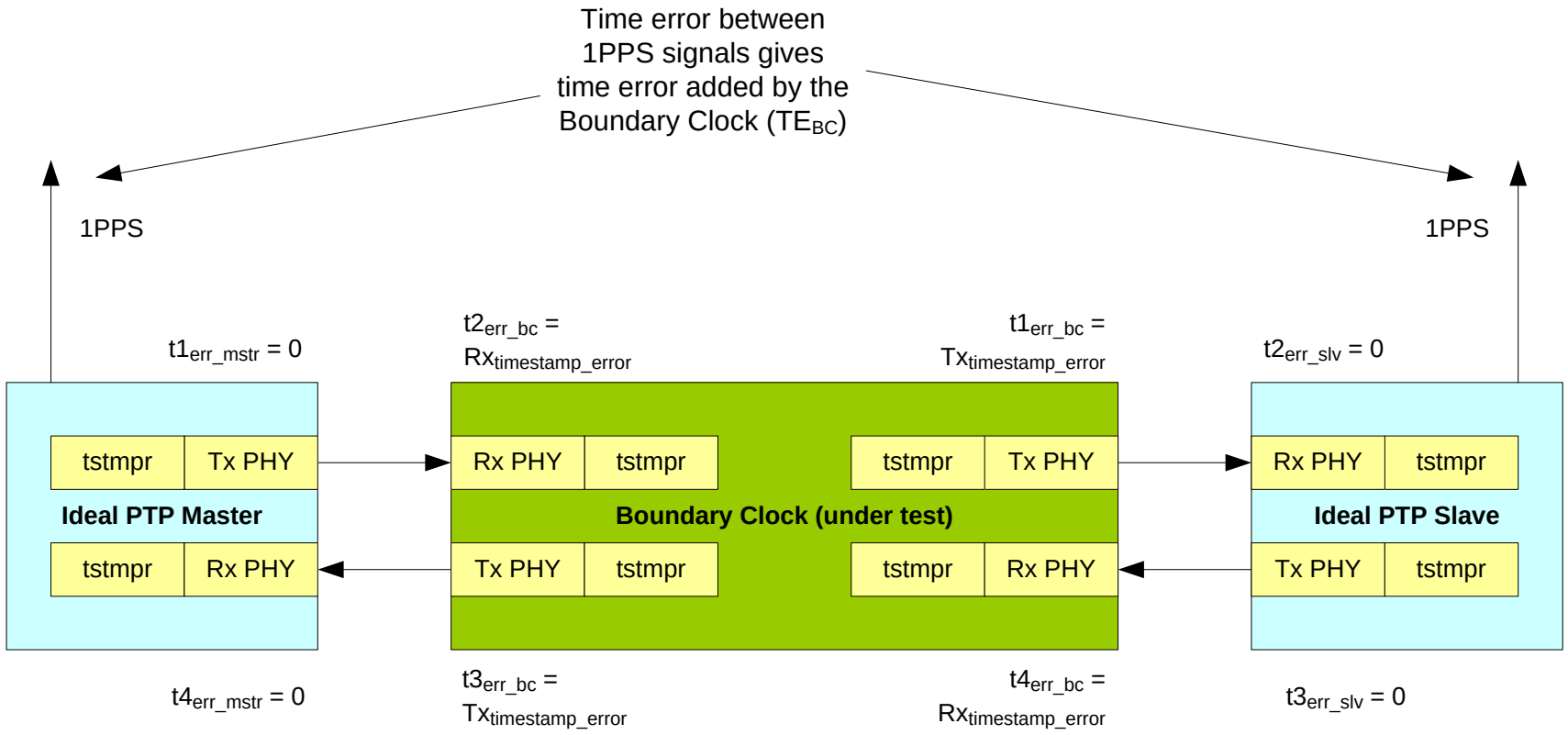
PTP Timestamp Generation Model

- A timestamp is generated at the time the "message timestamp point" crosses "reference plane", which is the intersection between the network (i.e. the medium) and the PHY
- Timestamp capture is implemented at the "timestamp measurement plane", which, in practice, occurs at point A and must be moved back to the reference plane
- *Good estimate of the PHY delay* ("path data delay", the time between the reference plane and the timestamp measurement plane) *is needed → varying delays should be compensated for*
- *Every endpoint needs to have the same understanding of the above concepts and how compensation is done*



Time Error Measurement Model (for Boundary Clock)

- PTP Master and PTP Slave are ideal (no timestamping errors, perfectly stable clocks)
- Boundary Clock's time error (TE) is affected by timestamping errors on messages to/from Master and to/from Slave
  - other sources of TE are ignored for this discussion
- $|TE_{BC}| = 0.5*(|t1_{err\_bc}| + |t2_{err\_bc}| + |t3_{err\_bc}| + |t4_{err\_bc}|) = (|Tx_{timestamp\_error}| + |Rx_{timestamp\_error}|)$



Appendix 2

Current State of Clause 90, Ethernet support for time synchronization protocols, IEEE 802.3

Current IEEE 802.3 Support for Time Synchronization (1)

- IEEE 802.3 Clause 90 provides support for a TimeSync Client
  - The optional Time Synchronization Service Interface (TSSI) supports protocols that require knowledge of packet egress and ingress time
  - Timestamping is done in the gRS, where the timestamp is captured when the message timestamp point crosses the xMII

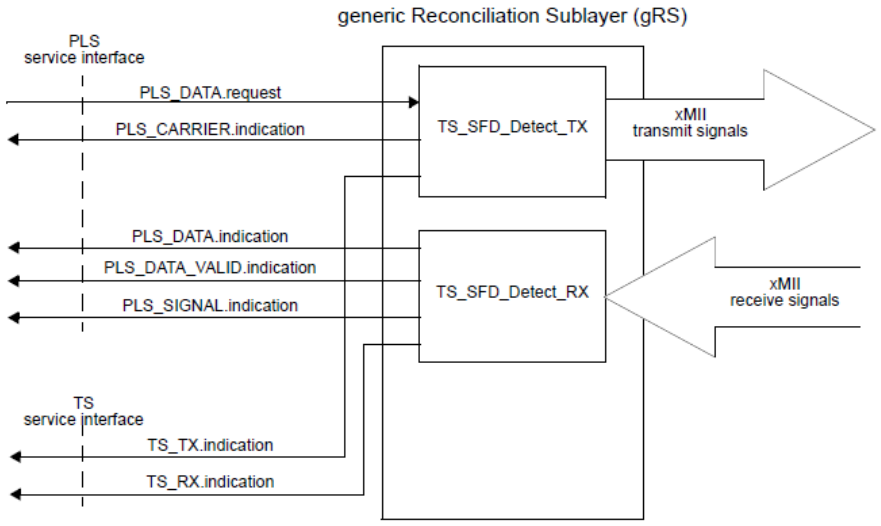
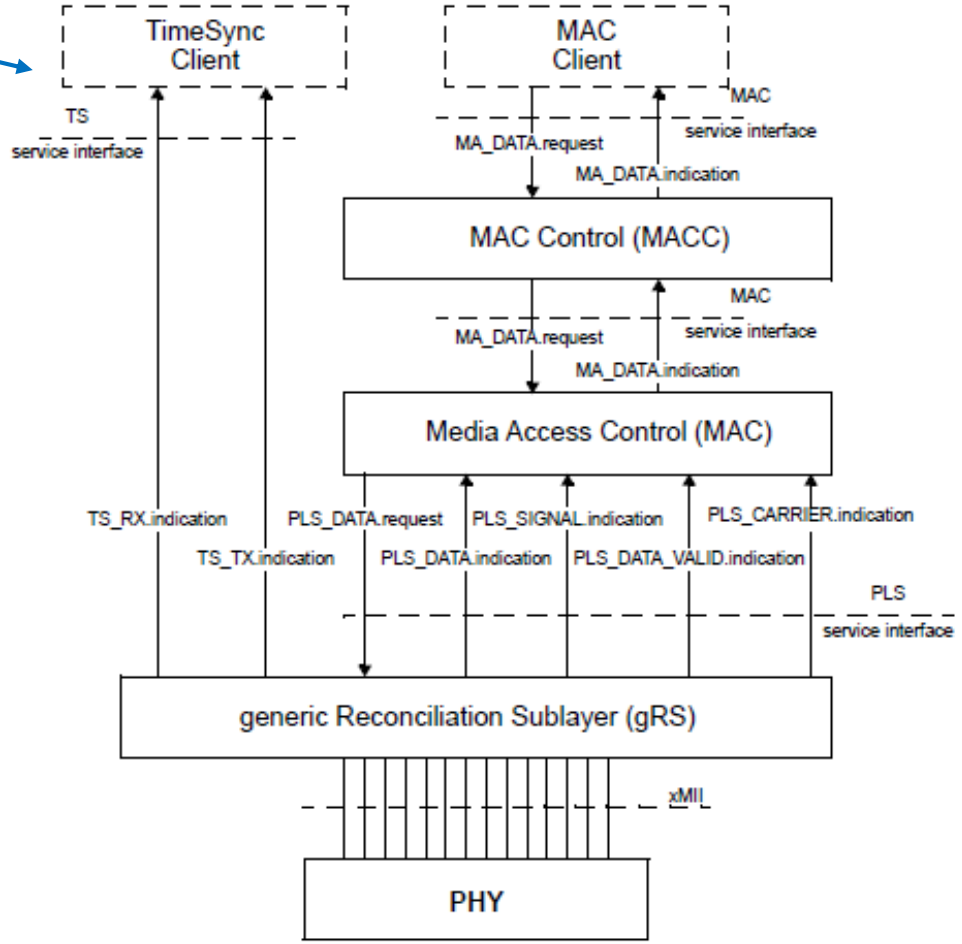


Figure 90–2—TS_SFD_Detect_TX and TS_SFD_Detect_RX functions within the generic Reconciliation Sublayer (gRS)



Current IEEE 802.3 Support for Time Synchronization (2)

- TSSI allows for “PHY” delay measurement to be done by TimeSync Client(s)
  - 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.

- The obtained data delay measurement is reported in the form of a quartet of values as defined for the TimeSync managed object class.
    - maximum transmit data delay
    - minimum transmit data delay
    - maximum receive data delay
    - minimum receive data delay

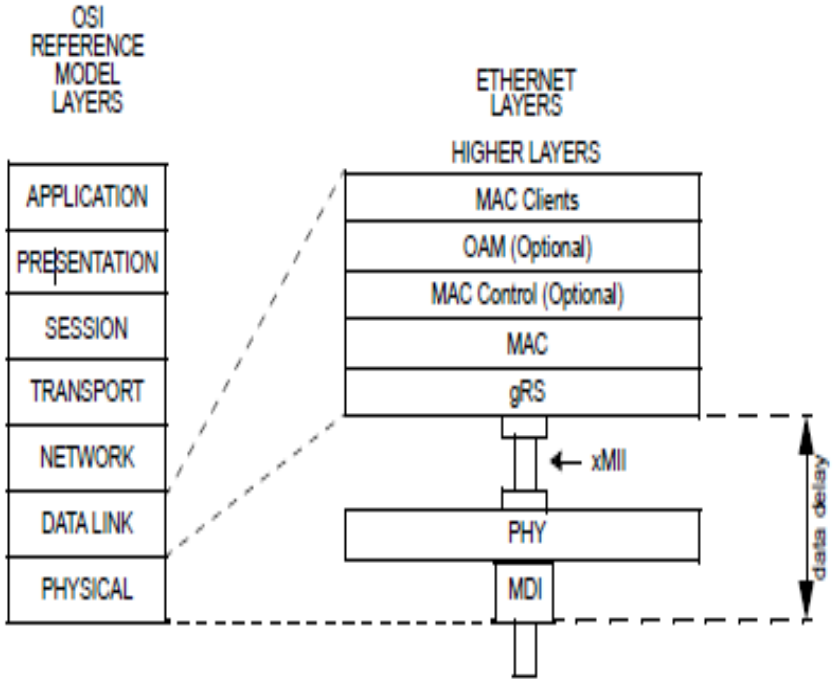


Figure 90–3—Data delay measurement

Current IEEE 802.3 Support for Time Synchronization (3)

- Multi-Lane – clause 90.7 (added in 2016):
  - “The receiver of a multi-lane PHY is expected to include a buffer to compensate for skew between the lanes. This buffer selectively delays each lane such that the lanes are aligned at the buffer output. The earliest arriving lane experiences the most delay through the buffer and the latest arriving lane experiences the least delay through the buffer. The receive path data delay for a multi-lane PHY is reported as if the beginning of the SFD arrived at the MDI input on the lane with the smallest buffer delay.”

- FEC – clause 90.7 (added in 2018):
  - “For a PHY that includes an FEC function, the transmit and receive path data delays may show significant variation depending upon the position of the SFD within the FEC block. However, since the variation due to this effect in the transmit path is expected to be compensated by the inverse variation in the receive path, it is recommended that the transmit and receive path data delays be reported as if the SFD is at the start of the FEC block.”

Appendix 3

Potential Areas of Improvement in Support of High Accuracy Time Transport

History of Discussions and Contributions (1)

- Liaison with ITU-T SG15
  - ITU_SG15-LS-72_to_IEEE_802d3.pdf (Oct 2017)
    - ITU requested advice on sources of timestamping error in PHYs with FEC, codeword markers, and/or alignment markers
  - IEEE_802d3_to_SG15_timing_0118.pdf (Jan 2018)
    - Indicated that Ethernet FEC streams are bit transparent through the FEC layer such that the delay variation in the Tx path is matched by a complementary variation in the Rx path
    - Indicated that some implementation introduce no timestamping inaccuracy due to markers
- IEEE 802.3 Maintenance Task Force:
  - gorshe_1_0718.pdf (Jul 2018)
    - Sought clarity in PTP timestamping in the presence of alignment markers
    - Highlighted differences for the message timestamp point between IEEE 802.3 clause 90 and IEEE 802.1AS and IEEE 1588-2018 (beginning of SFD vs beginning of symbol after SFD)
  - gorshe_1_0119.pdf (Jan 2019)
    - Reiterated above points
    - Highlighted new point about lane distribution/multiplexing delay variation and its impact on timestamping accuracy
  - Mar 2019
    - Moved this discussion into the IEEE 802.3 New Ethernet Applications (NEA) Task Force

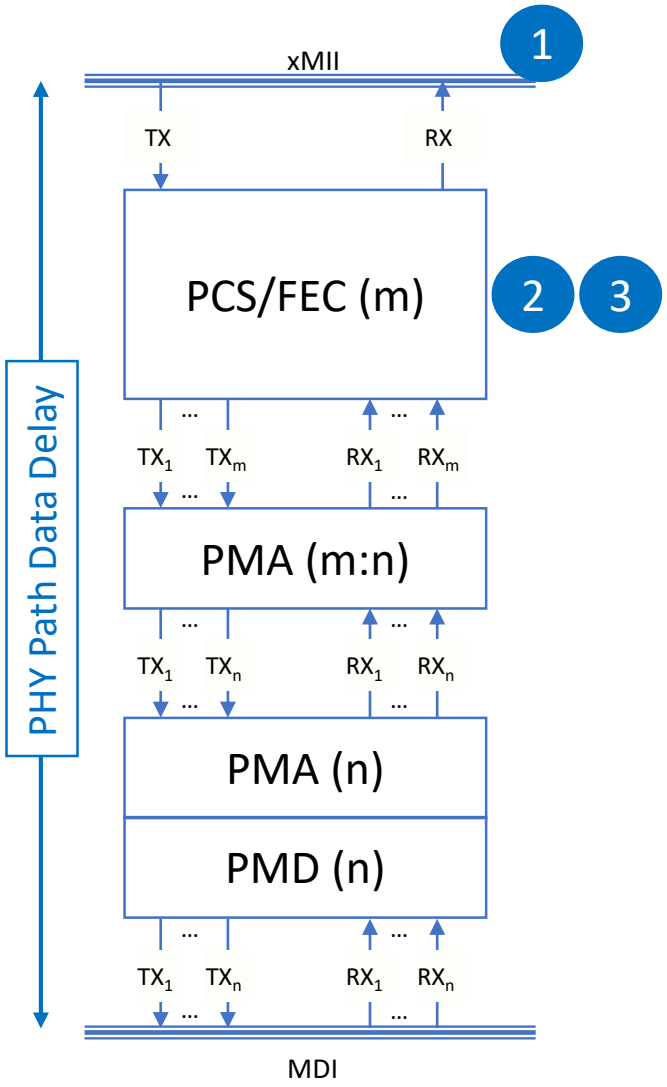
History of Discussions and Contributions (2)

- IEEE 802.3 NEA Task Force:
  - tse_nea_01_190416.pdf (Apr 2019)
    - Showed PTP fundamentals
    - Discussed requirements imposed by applications and other SDOs
    - Showed inaccuracies that can result from current clause 90 specifications
    - Discussed possible courses of action
  - nicholl_nea_01_190416.pdf (Apr 2019)
    - Provided historical background of timestamping discussions
    - Agreed with concepts described in the 7/2018 and 1/2019 Maintenance TF presentations
    - General agreement with proposed solutions
  - tse_nea_01a_0519.pdf (May 2019)
    - First draft of the CFI consensus building presentation

Potential Areas of Improvement in Support of High Accuracy Time Transport

Improvements to Clause 90 are needed to enable better PTP performance

1. Deal with Message Timestamp Point differences between IEEE 802.3 and IEEE 1588/IEEE 802.1AS and its effect on Tx/Rx Path Data Delay
2. Specify how delay variance from Alignment Marker (AM) and Idle insertion/removal events are accounted for
3. Multi-Lane
    - Clarify timestamping for multi-PCS lane PHYs
    - Specify how delay variance from multi-PCS lane distribution mechanism is accounted for



Message Timestamp Point

Subclause 90.7 of IEEE 802.3 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 follow:

- “the message timestamp point for an event message shall be the beginning of the first symbol after the Start of Frame (SOF) delimiter”
- “the message timestamp point for a PTP event message shall be the beginning of the first symbol following the start of frame delimiter”

Effect of Different Message Timestamp Points

- Link delay measurement is affected by the message timestamp point
  - A timestamp at the beginning of SFD is earlier than a timestamp at the beginning of the first symbol after SFD
  - Examples:
    - Master and slave both use symbol after SFD:
      - Measured link delay = X
    - Master and slave both use beginning of SFD:
      - Measured link delay = X
    - Master uses symbol after SFD and Slave uses beginning of SFD:
      - Measured link delay = X – T_{SFD}
        - T_{SFD} is the time occupied by a SFD symbol
        - creates a constant time error cTE = T_{SFD}
- Alignment marker could also separate the SFD and the symbol after the SFD, creating an even greater discrepancy between their corresponding timestamps

AM and IDLE Insertion/Removal

Alignment Marker (AM) and Idle insertion/removal affect the path data delay:

- Insertion of AM or Idle momentarily increases the path data delay by T_{AM} or T_{Idle}, respectively
- Removal of AM or Idle momentarily decreases the path data delay by T_{AM} or T_{Idle}, respectively
- Idle insertion/removal operate independently at Rx and Tx so delay changes do not have deterministic relationship
- AM removal at Rx deterministically undoes the delay change caused by AM insertion at Tx
  - However, AM events cause many additional Idle insertion/removal events

Multi-Lane PHY Ambiguities

Ambiguities in IEEE 802.3 can affect path data delay values.

- Ambiguities for NxPCS lane Transmitter implementation
  A. 66B blocks and timestamps are not aligned at NxPCS lane transmitter outputs
  B. 66B blocks and timestamps are aligned at NxPCS lane transmitter outputs
  C. 66B blocks are aligned but timestamps are not aligned at NxPCS lane transmitter outputs
- Path data delays for the lane distribution function can be different for each PCS lane in Tx and Rx PHYs
  - Example: received lane 0 block goes to xMII first while received lane N goes to xMII last
  - No instructions are given on how to handle these deterministic but varying path data delays
- Interactions between implementations that interpret the specification differently will have additional time error
- See Appendix for details on the above items

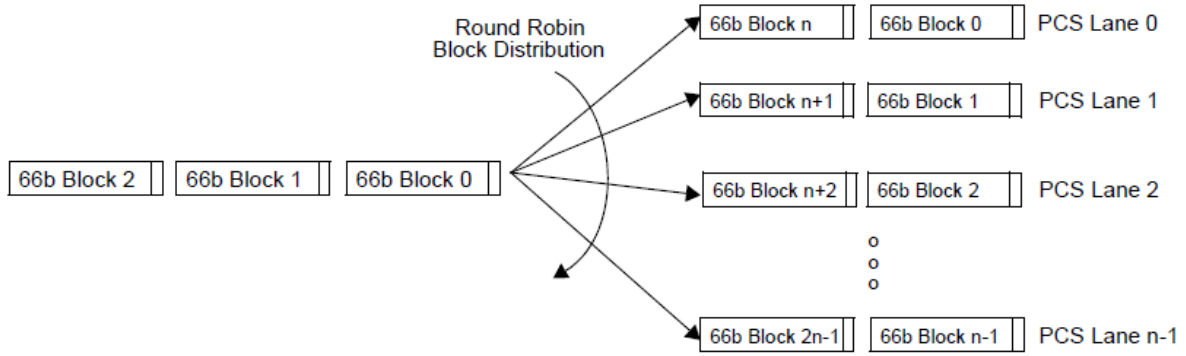


Figure 82–6—PCS Block distribution

Performance vs Target

- Max|TE| = 30ns for class C Telecom Boundary Clock (see slide 10)
  - There are other sources of TE in addition to those from timestamping

Ethernet Rate	Path Data Delay Variation per Tx/Rx Interface (ns)				Total TE per Tx or Rx Interface (ns)	Max|TE| contribution per PTP Boundary Clock (ns)
	mismatched SFD timestamp point	Idle insert/remove (per Idle)	AM insert/remove	Lane Distribution		
GE	8	16	N/A	N/A	24	48
10GE	0.8	3.2	N/A	N/A	4	8
25GE	0.32	1.28	2.56	N/A	4.16	8.32
40GE	0.2	1.6	6.4	4.8	13	26
100GE	0.08	0.64	12.8	12.16	25.68	51.36
200GE	0.04	0.32	2.56	2.24	5.16	10.32
400GE	0.02	0.16	2.56	2.4	5.14	10.28

Appendix 4

More Details on PCS-Lane Distribution Delay

PCS-Lane Distribution Interpretation Option Details (1)

Ambiguities in IEEE 802.3 affect path data delays.

No instructions are given in IEEE 802.3 on how to handle the following deterministic but varying delays

- NxPCS lane Transmitter Interpretation Options
  - A. 66B blocks and timestamps are not aligned at NxPCS lane transmitter
    - xMII to MDI has constant path data delay for every lane
      - Data for Lane 0 arrives first at xMII and is transmitted first at MDI
      - Data for Lane N arrives last at xMII and is transmitted last at MDI
    - 66B blocks on each lane have a different timestamp because they cross the reference plane at different times
      - Timestamper at Tx xMII uses the same xMII-to-MDI constant data path delay for every lane
    - Lane-to-lane skew of 66B blocks at the transmitter is removed by Rx deskew buffers
  - B. 66B blocks and timestamps are aligned at NxPCS lane transmitter
    - xMII to MDI path has different path data delay for each lane
      - Data for Lane 0 arrives first at xMII and is transmitted at the same time as lane N at MDI, causing largest path data delay
      - Data for Lane N arrives last at xMII and is transmitted at the same time as Lane 0 at MDI, causing smallest path data delay
    - 66B blocks on every lane have the same timestamp because they cross the reference plane at the same time
      - Timestamper at Tx xMII uses appropriate xMII-to-MDI path data delay for each lane
    - No lane-to-lane skew of 66B blocks

PCS-Lane Distribution Interpretation Option Details (2)

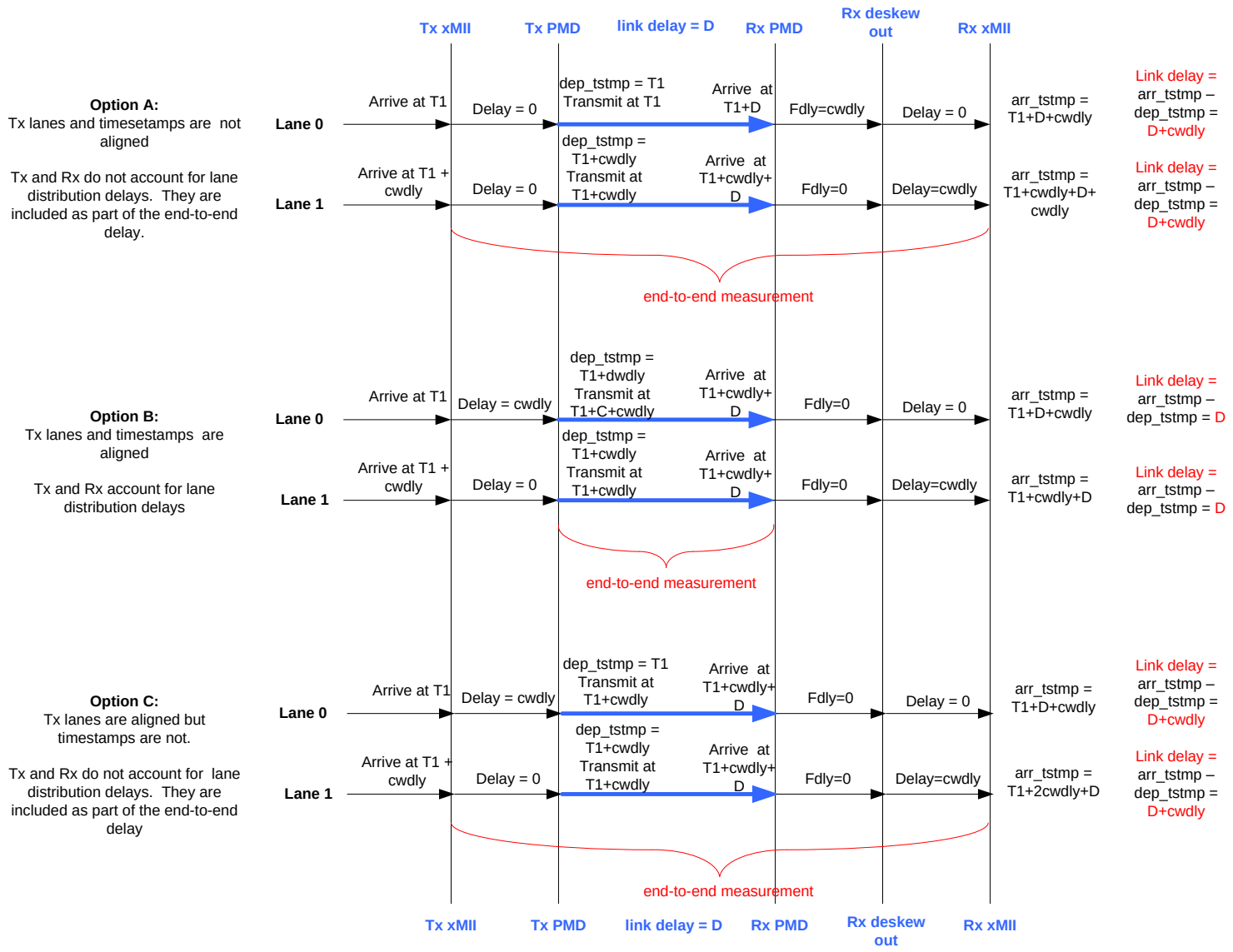
- NxPCS lane Transmitter Options (continued)
  - C. 66B blocks are aligned but timestamps are not aligned at NxPCS lane transmitter
    - xMII to MDI path has different path data delay for each lane
      - Data for Lane 0 arrives first at xMII and is transmitted at the same time as lane N at MDI, causing largest path data delay
      - Data for Lane N arrives last at xMII and is transmitted at the same time as Lane 0 at MDI, causing smallest path data delay
    - Timestamps assume a constant data path delay for all lanes
      - Timestamper at Tx xMII uses the same xMII-to-MDI constant path data delay for every lane
    - No lane-to-lane skew of 66B blocks

PCS-Lane Distribution Interpretation Option Details (3)

- NxPCS lane Receiver Options:
  - After deskew buffers, all lanes are aligned
    - For N-lane transmitter type “A”, intrinsic lane-to-lane skew of 66B blocks is “moved into the medium” by the deskew function
    - For N-lane transmitter types “B” and “C”, there is no skew of 66B blocks between lanes
  - MDI to xMII multiplexer causes varying path data delay
    - All lanes are deskewed and are ready to go to xMII
    - Data for Lane 0 goes to xMII first and has smallest path data delay
    - Data for Lane N goes to xMII last and has largest path data delay
  - How is this lane-to-lane varying delay handled?

PCS-Lane Distribution Interpretation Options Details (4)

- Figure shows examples of the 3 Options
- Arrival times at each stage are shown (Arrive at, Transmit at)
- The delays through each functional stage are shown (Delay, Fdly, link delay)
  - Constant delays are assumed to be 0 where the actual values don't matter
- The departure timestamps at Tx (dep_tstmp) and arrival timestamps at Rx (arr_tstmp) are shown
- The calculated link delay (Link_delay) is shown for the span (end-to-end measurement)



PCS-Lane Distribution Delays – Constant vs per-Lane

- There are two inherent approaches for determining the xMII-to-MDI delay on multi-PCS lane PHYs
  1. Method 1 – Account for the delay between the MII and the lane that carries the message timestamp point of the PTP message.
  2. Method 2 – Because the Tx + Rx lane distribution delay is a constant for every lane, use this constant delay regardless of which lane carries the message timestamp point.
     - This is like how IEEE 802.3 handles FEC delays

PCS-Lane Distribution Delays: 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 MDI, 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 MDI, it would seem this is a valid conclusion.

PCS-Lane Distribution Delays: Method 1 (continued)

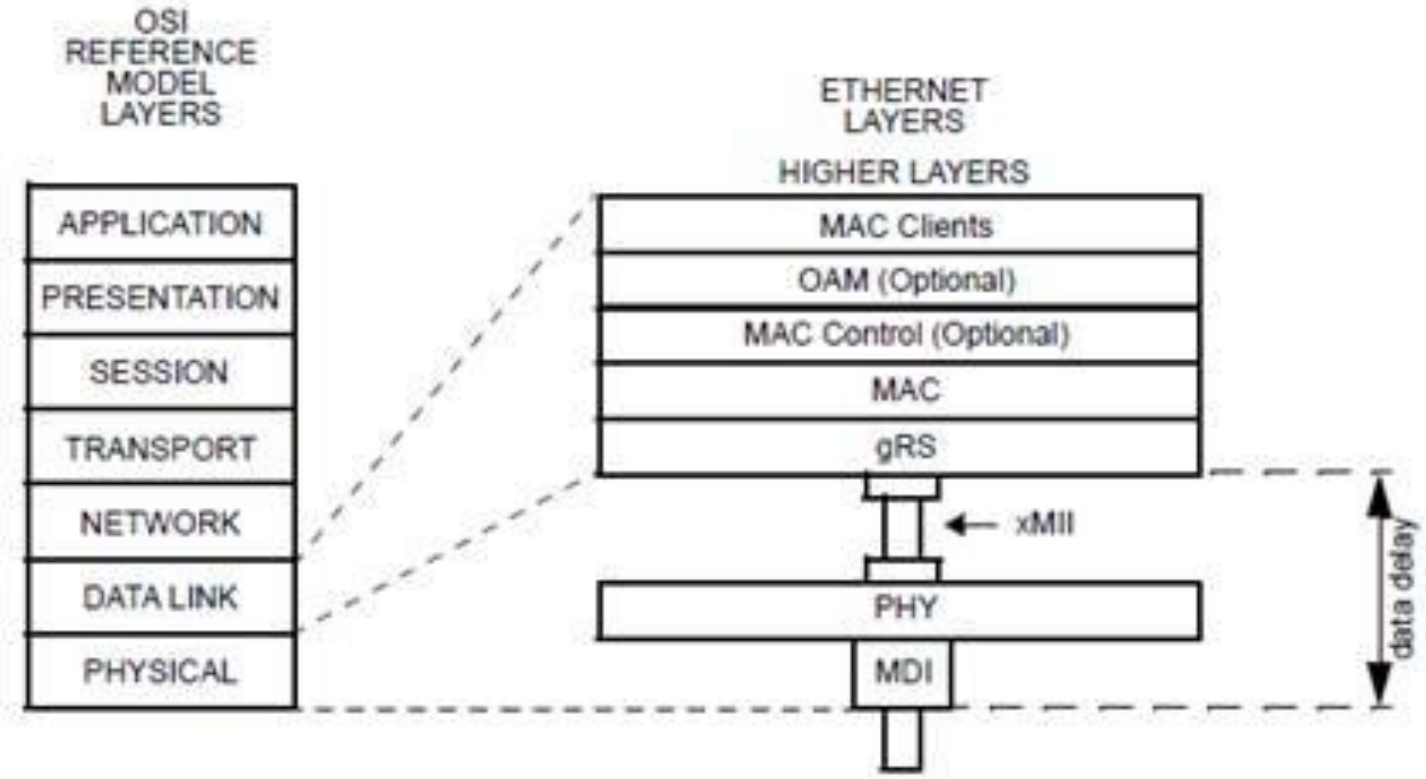
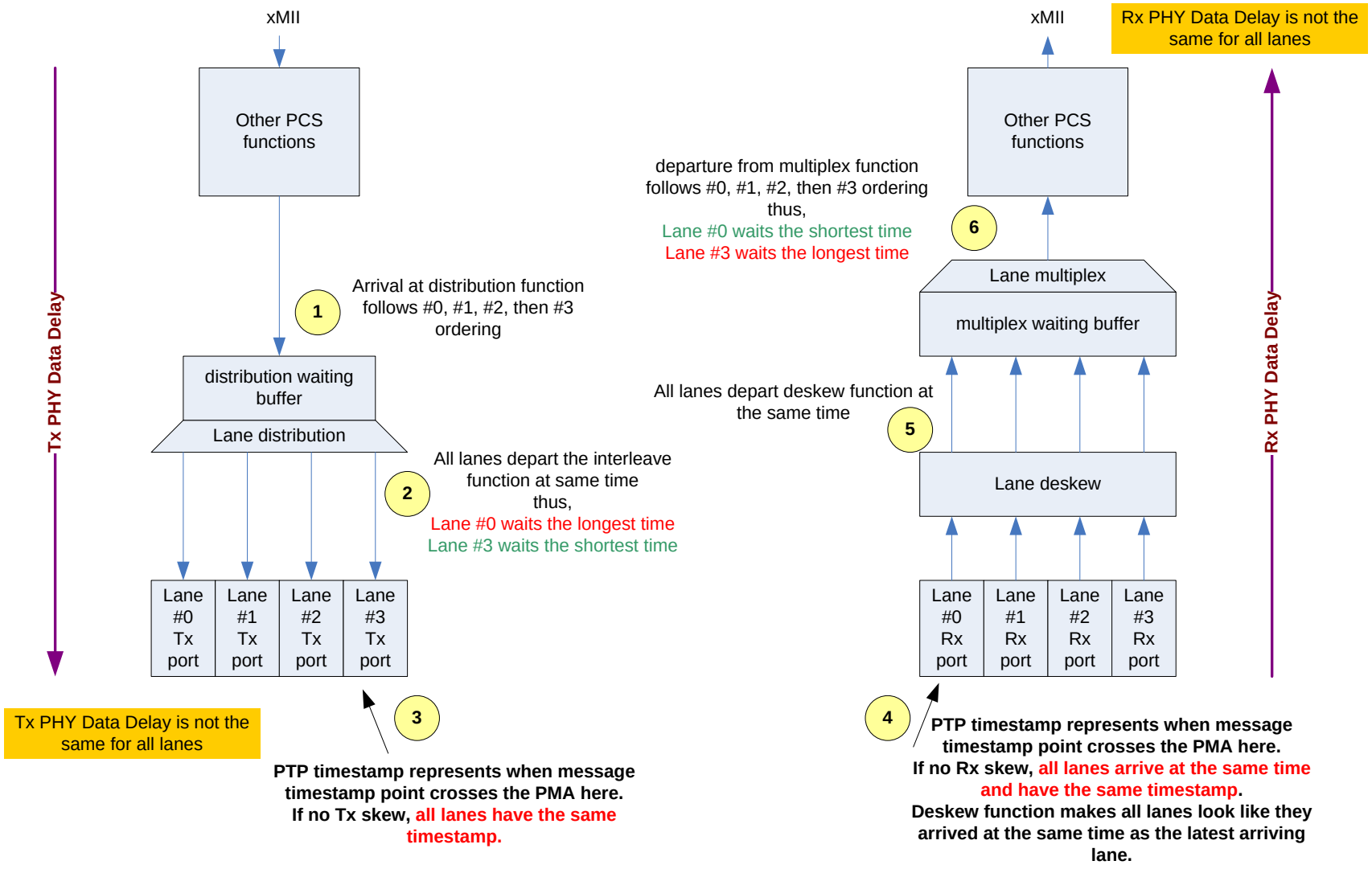


Figure 90–3—Data delay measurement

PCS-Lane Distribution Delays: Method 1 (continued)

- However, this means that PHY data delay (between xMII and MDI, as per Figure 90-3 above) is not the same for every lane because the MDI-to-xMII multiplexing delay (for Rx) and xMII-to-MDI demultiplexing delay (for Tx) is different for each lane (as shown in Figures 82-3 and 82-4 below). In the Tx direction, 66B blocks going to lane 0 have the most delay and 66B blocks 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 IEEE 802.3 model), the lane-based intrinsic delay must be included as part of the PHY data delay.
  - Was this the intent?

PCS-Lane Distribution Delays: Method 1 (continued)



PCS-Lane Distribution Delays: Method 2

- These multi-PCS lane 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?

PCS-Lane Distribution Delays: Method 2 (continued)