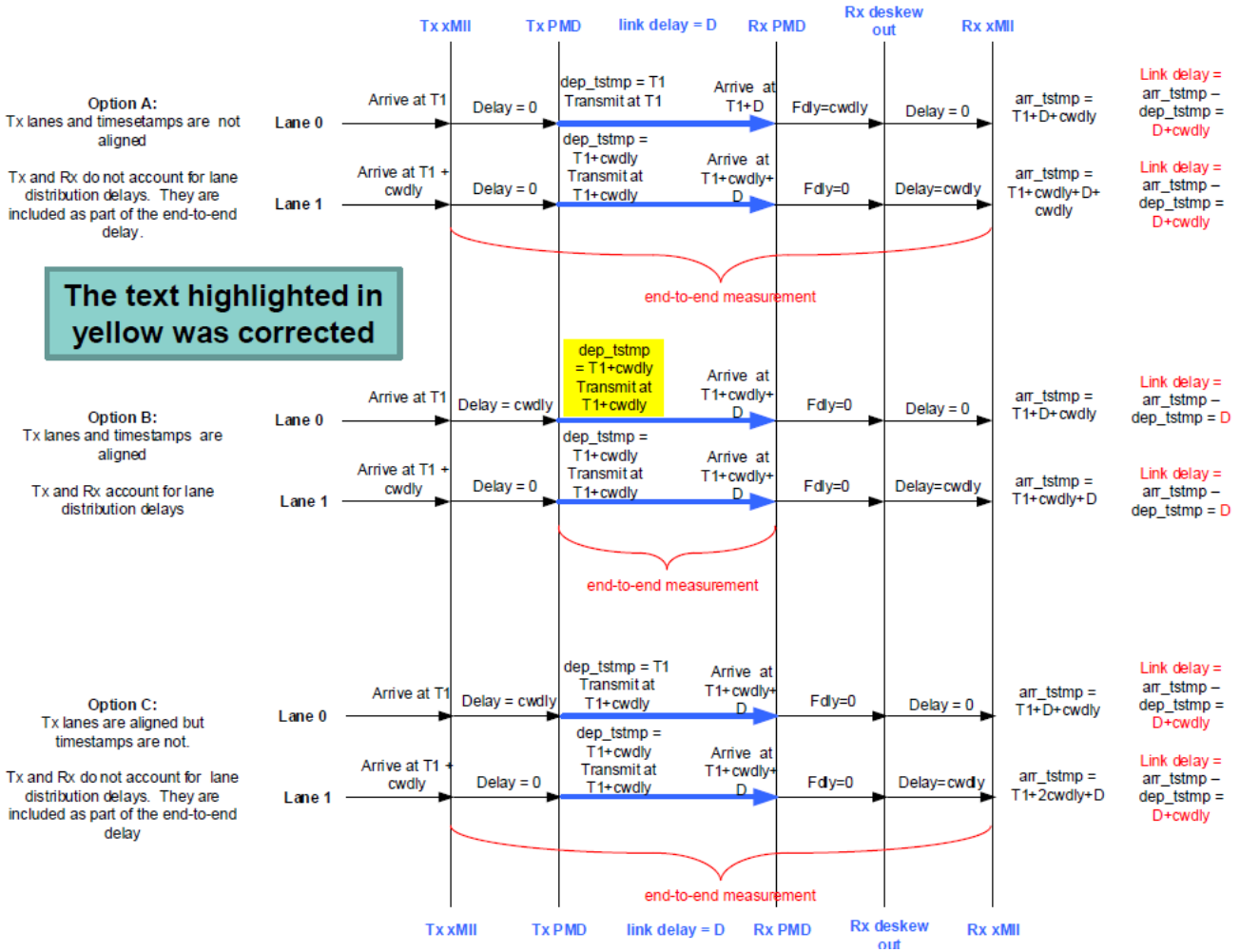
Timestamp Granularity for Multi-PCS Lane Interface

Xiang He (Huawei)
Jingfei Lv (Huawei)
Silvana Rodrigues (Huawei)

802.3cx ad hoc – 10/20/2020

Background – Multi-PCS lane distribution

- tse_3cx_02_0520 gave three possible solutions to compensate timestamp error caused by multi-PCS lane distribution.



Background – Proposed solutions

- tse_3cx_02_0520 lists three options to generate timestamps at Tx:
  - Option A: 66B blocks and timestamps are not aligned at NxPCS lane transmitter
  - Option B: 66B blocks and timestamps are aligned at NxPCS lane transmitter
  - Option C: 66B blocks are aligned but timestamps are not aligned at NxPCS lane transmitter
- And two methods for compensating multi-PCS lane distribution delay
  - Method 1: Account for the delay between the MII and the lane that carries the message timestamp point of the PTP message
  - Method 2: Use a constant delay regardless of which lane carries the message timestamp point, because the Tx+Rx lane distribution delay is a constant for every lane.
- Two different approaches were proposed in tse_3cx_02_0520 and he_3cx_01_0520.
  - Approach 1: Option C + Method 2, where Tx and Rx compensate a constant delay instead of the accurate delay, because the sum of the end-to-end intrinsic delay is constant.
  - Approach 2: Option B + Method 1, where Tx and Rx accurately compensate TS delay from xMII to MDI.
    - Fully interoperable with “Option A + Method 1”.

Timestamp Accuracy vs Timestamp “Granularity”

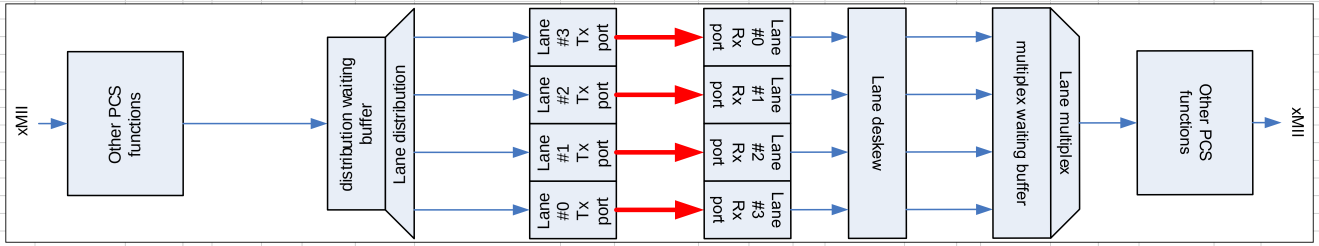
- Question was raised on timestamp granularity between PCS lanes.
  - All PCS lanes have the same timestamp when using Tx Option B.
  - Each PCS lane has different timestamps for Tx Option A and C.
- Granularity could be interpreted in two ways:
  1. Timestamp quantization steps between different PCS lanes
  2. Timestamp quantization steps between different PTP messages
- For higher timestamp accuracy, *“2”* is more important.
  - Both *1* & *2* could be maintained with proper definition, but there is no need to keep *1* to a finer value because non-PTP messages does not carry timestamps.
- The following slides give three examples explaining the different definitions of granularity.

Example 1: “Coarse granularity” on PCS lanes, “fine granularity” for PTP message timestamps

- “Option B + Method 1” as defined in tse_multilane_TE_analysis.

- Because blocks on the PCS lanes are aligned before sending, all lanes share the same “departure timestamp”.
- The timestamps on the PCS lanes may not be the actual timestamps transmitted for PTP messages because distribution latency will be compensated.
- The actual timestamp has a much finer granularity.

Block Time	640	ps	time of one block's worth of data (i.e. 64 bits of uncoded data)
Tx Option	B		Blocks on NxPCS lane Tx MDI are aligned, xMII to MDI delay is different for every Tx PCS lane, blocks on every Tx PCS lane have same timestamp because per-lane distribution delay is accounted for
Rx Option	1		Each Rx lane's MDI to xMII delay is different and is accounted for as such
Link delay	0	ps	assumes the same delay for all lanes
Number of lanes	20		number of PCS lanes (max = 20)
Resulting time error	**0**	ps	



Lane	actual arvl time	actual distrib dly	actual dept time	actual arrival time	actual deskew dly	actual merge dly	actual arrival time
0	0	12160	12160	12160	0	0	12160
1	640	11520	12160	12160	0	640	12800
2	1280	10880	12160	12160	0	1280	13440
3	1920	10240	12160	12160	0	1920	14080
4	2560	9600	12160	12160	0	2560	14720
5	3200	8960	12160	12160	0	3200	15360
6	3840	8320	12160	12160	0	3840	16000
7	4480	7680	12160	12160	0	4480	16640
8	5120	7040	12160	12160	0	5120	17280
9	5760	6400	12160	12160	0	5760	17920
10	6400	5760	12160	12160	0	6400	18560
11	7040	5120	12160	12160	0	7040	19200
12	7680	4480	12160	12160	0	7680	19840
13	8320	3840	12160	12160	0	8320	20480
14	8960	3200	12160	12160	0	8960	21120
15	9600	2560	12160	12160	0	9600	21760
16	10240	1920	12160	12160	0	10240	22400
17	10880	1280	12160	12160	0	10880	23040
18	11520	640	12160	12160	0	11520	23680
19	12160	0	12160	12160	0	12160	24320

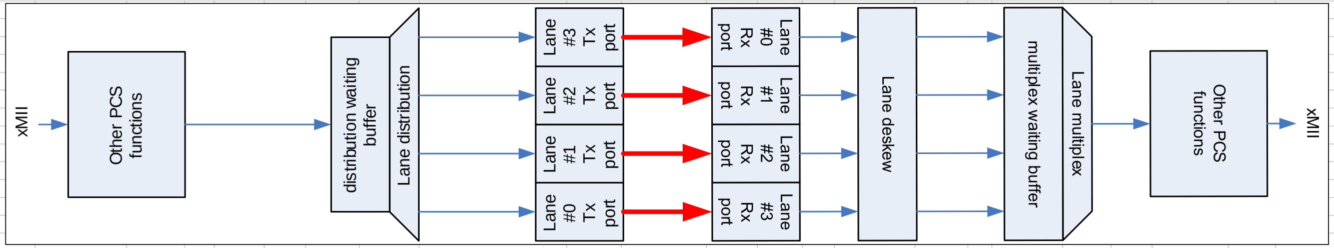
Lane	modelled PHY dly	dept timestamp	modelled PHY dly	arrival timestamp	measured link dly
0	12160	12160	0	12160	
1	11520	12160	640	12160	
2	10880	12160	1280	12160	
3	10240	12160	1920	12160	
4	9600	12160	2560	12160	
5	8960	12160	3200	12160	
6	8320	12160	3840	12160	
7	7680	12160	4480	12160	
8	7040	12160	5120	12160	
9	6400	12160	5760	12160	
10	5760	12160	6400	12160	
11	5120	12160	7040	12160	
12	4480	12160	7680	12160	
13	3840	12160	8320	12160	
14	3200	12160	8960	12160	
15	2560	12160	9600	12160	
16	1920	12160	10240	12160	
17	1280	12160	10880	12160	
18	640	12160	11520	12160	
19	0	12160	12160	12160	

Example 2: “Fine granularity” on PCS lanes and “fine granularity” for PTP message timestamps

- This is equivalent to “Option A + Method 1”.

- Because Rx uses the same method as Example 1, it is fully interoperable with Option B.
- The timestamps on the PCS lanes are different because Tx does not require alignment at MDI.

Block Time	640	ps	time of one block's worth of data (i.e. 64 bits of uncoded data)
Tx Option	A		Blocks on NxPCS lane Tx MDI are not aligned, xMII to MDI delay is constant for every Tx PCS lane, blocks on every Tx PCS lane have different timestamps, differing by the block arrival delay on xMII
Rx Option	1		Each Rx lane's MDI to xMII delay is different and is accounted for as such
Link delay	0	ps	assumes the same delay for all lanes
Number of lanes	20		number of PCS lanes (max = 20)
Resulting time error	**0**	ps	



Lane		actual arvl time	actual distrib dly	actual dept time	actual arrival time	actual deskew dly	actual merge dly	actual arrival time
	0	0	0	0	0	12160	0	12160
	1	640	0	640	640	11520	640	12800
	2	1280	0	1280	1280	10880	1280	13440
	3	1920	0	1920	1920	10240	1920	14080
	4	2560	0	2560	2560	9600	2560	14720
	5	3200	0	3200	3200	8960	3200	15360
	6	3840	0	3840	3840	8320	3840	16000
	7	4480	0	4480	4480	7680	4480	16640
	8	5120	0	5120	5120	7040	5120	17280
	9	5760	0	5760	5760	6400	5760	17920
	10	6400	0	6400	6400	5760	6400	18560
	11	7040	0	7040	7040	5120	7040	19200
	12	7680	0	7680	7680	4480	7680	19840
	13	8320	0	8320	8320	3840	8320	20480
	14	8960	0	8960	8960	3200	8960	21120
	15	9600	0	9600	9600	2560	9600	21760
	16	10240	0	10240	10240	1920	10240	22400
	17	10880	0	10880	10880	1280	10880	23040
	18	11520	0	11520	11520	640	11520	23680
	19	12160	0	12160	12160	0	12160	24320

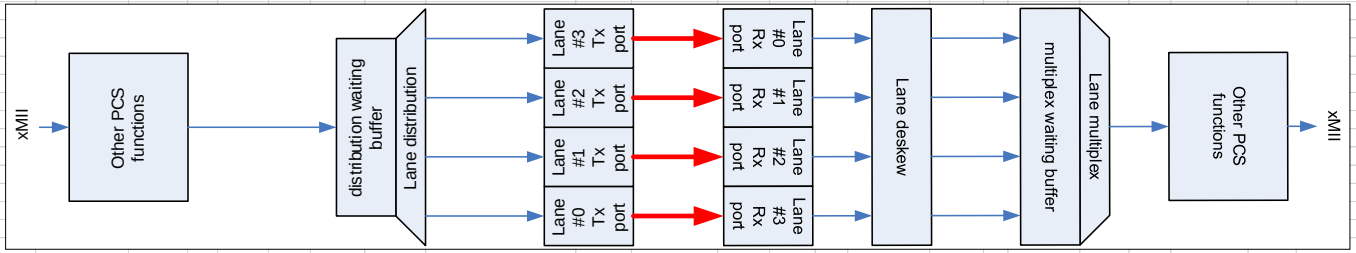
Lane		modelled PHY dly	dept timestamp	modelled PHY dly	arrival timestamp	measured link dly
	0	0	0	12160	0	
	1	0	640	12160	640	
	2	0	1280	12160	1280	
	3	0	1920	12160	1920	
	4	0	2560	12160	2560	
	5	0	3200	12160	3200	
	6	0	3840	12160	3840	
	7	0	4480	12160	4480	
	8	0	5120	12160	5120	
	9	0	5760	12160	5760	
	10	0	6400	12160	6400	
	11	0	7040	12160	7040	
	12	0	7680	12160	7680	
	13	0	8320	12160	8320	
	14	0	8960	12160	8960	
	15	0	9600	12160	9600	
	16	0	10240	12160	10240	
	17	0	10880	12160	10880	
	18	0	11520	12160	11520	

Example 3: “Fine granularity” on PCS lanes but “coarse granularity” for PTP message timestamps

- This is equivalent to “Option C + Method 2”.

- Each PCS lane has its own timestamp.
- The timestamp transmitted did not account for PCS lane distribution delay.
- Actual PTP message timestamps do not benefit from the 640 ps timestamp steps between the PCS lanes.
- Actual PTP message timestamps granularity is 12160 ps because they are not compensated based on the lane number.

Block Time	640	ps	time of one block's worth of data (i.e. 64 bits of uncoded data)		
Tx Option	C		Blocks on NxPCS lane Tx MDI are aligned, xMII to MDI delay is different for every Tx PCS lane, blocks on Tx PCS lanes have different timestamps because a predetermined constant lane distribution delay is used for		
Rx Option	2		Each Rx lane's MDI to xMII delay is different, but is accounted for as a predetermined constant delay because it can be balanced by a mirror delay on the Tx side		
Link delay	0	ps	assumes the same delay for all lanes		
Number of lanes	20		number of PCS lanes (max = 20)		
Resulting time error	**0**	ps			



Lane	actual arvl time	actual distrib dly	actual dept time	actual arrival time	actual deskew dly	actual merge dly	actual arrival time
0	0	12160	12160	12160	0	0	12160
1	640	11520	12160	12160	0	640	12800
2	1280	10880	12160	12160	0	1280	13440
3	1920	10240	12160	12160	0	1920	14080
4	2560	9600	12160	12160	0	2560	14720
5	3200	8960	12160	12160	0	3200	15360
6	3840	8320	12160	12160	0	3840	16000
7	4480	7680	12160	12160	0	4480	16640
8	5120	7040	12160	12160	0	5120	17280
9	5760	6400	12160	12160	0	5760	17920
10	6400	5760	12160	12160	0	6400	18560
11	7040	5120	12160	12160	0	7040	19200
12	7680	4480	12160	12160	0	7680	19840
13	8320	3840	12160	12160	0	8320	20480
14	8960	3200	12160	12160	0	8960	21120
15	9600	2560	12160	12160	0	9600	21760
16	10240	1920	12160	12160	0	10240	22400
17	10880	1280	12160	12160	0	10880	23040
18	11520	640	12160	12160	0	11520	23680
19	12160	0	12160	12160	0	12160	24320

Lane	modelled PHY dly	dept timestamp	modelled PHY dly	arrival timestamp	measured link dly
0	6080	6080	6080	6080	
1	6080	6720	6080	6720	
2	6080	7360	6080	7360	
3	6080	8000	6080	8000	
4	6080	8640	6080	8640	
5	6080	9280	6080	9280	
6	6080	9920	6080	9920	
7	6080	10560	6080	10560	
8	6080	11200	6080	11200	
9	6080	11840	6080	11840	
10	6080	12480	6080	12480	
11	6080	13120	6080	13120	
12	6080	13760	6080	13760	
13	6080	14400	6080	14400	
14	6080	15040	6080	15040	
15	6080	15680	6080	15680	
16	6080	16320	6080	16320	
17	6080	16960	6080	16960	
18	6080	17600	6080	17600	

Summary

- Option B + Method 1
  - All PCS lanes share the same timestamp (granularity_PCS = 12160 ps)
  - Each PTP message has accurate timestamp (granularity_PTP = 1 bit time).

- Option A + Method 1
  - PCS lanes have different timestamps (granularity_PCS = 640 ps)
  - Each PTP message has accurate timestamp (granularity_PTP = 1 bit time).

- Option C + Method 2
  - PCS lanes have different timestamps (granularity_PCS = 640 ps)
  - Each PTP message has inaccurate timestamp (granularity_PTP = 12160 ps).

THANK YOU!

Time error caused by different compensation methods

- The spreadsheet also shows time error could be as much as ~6ns if Tx and Rx use different options.
  - Data delay measurement as in current standard requires inclusion of all latency caused by the protocol stack between the point timestamp is generated and the reference plane (MDI).
  - All ports in compliance with the current standard will suffer the 6ns time error if the other side chooses to ignore the PCS lane distribution delay and relying on error cancellation.

Block Time	640	ps	Block Time	640	ps
Tx Option	A		Tx Option	A	
Rx Option	1		Rx Option	2	
Link delay	0	ps	Link delay	0	ps
Number of lanes	20		Number of lanes	20	
Resulting time error	**0**	ps	**Resulting time error**	**6080**	ps
Block Time	640	ps	Block Time	640	ps
Tx Option	B		Tx Option	B	
Rx Option	1		Rx Option	2	
Link delay	0	ps	Link delay	0	ps
Number of lanes	20		Number of lanes	20	
Resulting time error	**0**	ps	**Resulting time error**	**6080**	ps
Block Time	640	ps	Block Time	640	ps
Tx Option	C		Tx Option	C	
Rx Option	1		Rx Option	2	
Link delay	0	ps	Link delay	0	ps
Number of lanes	20		Number of lanes	20	
Resulting time error	**6080**	ps	**Resulting time error**	**0**	ps

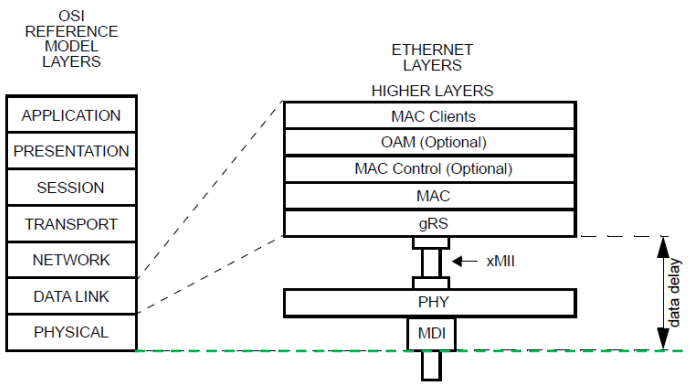


Figure 90–3—Data delay measurement

IEEE Std 1588-2008
IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems

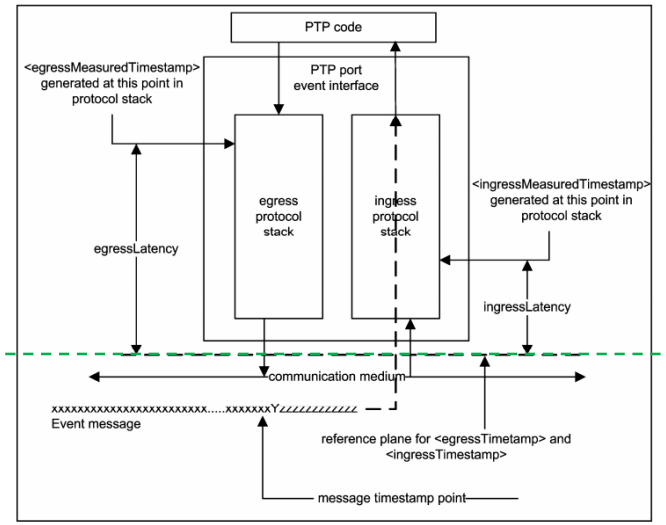


Figure 19—Definition of latency constants

Reference plane defined in IEEE802.3 and IEEE1588

Inaccurate compensation of multilane distribution delay

- Option C + method 2 proposed in tse_3cx_02_0520 recommends to use the constant sum of Tx+Rx to cancel out multi-lane distribution delay.
  - tse_3cx_03_0520 extends this method to all other intrinsic delay introduced by the protocol stack.
- The equivalent reference plane is “floating”.
  - Time error is expected when two sides are following different methods.

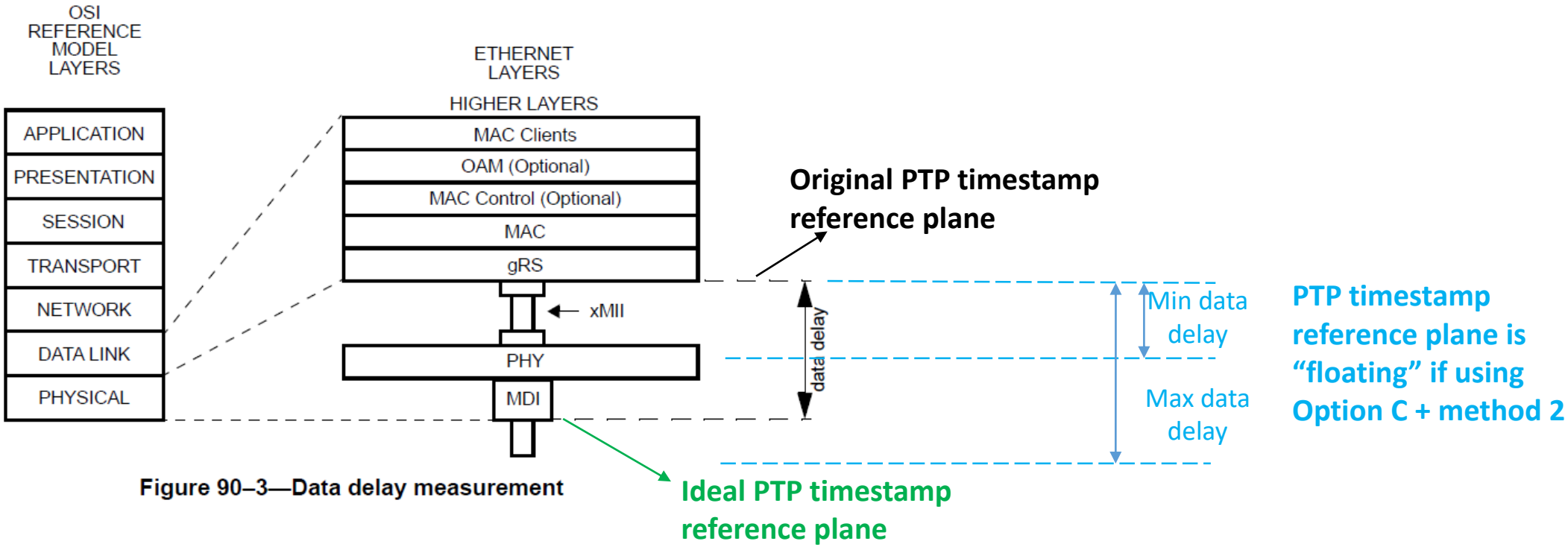


Figure 90–3—Data delay measurement

Accurate compensation of multilane distribution delay

- Option B + Method 1 proposed by he_3cx_01_0520 requires accurate compensation on both sides.
- Extra work might be needed in the logic layer, and a method was proposed in this contribution.
- After the accurate compensation of multilane distribution delay, the PTP timestamp reference plane aligns with the ideal plane (MDI).

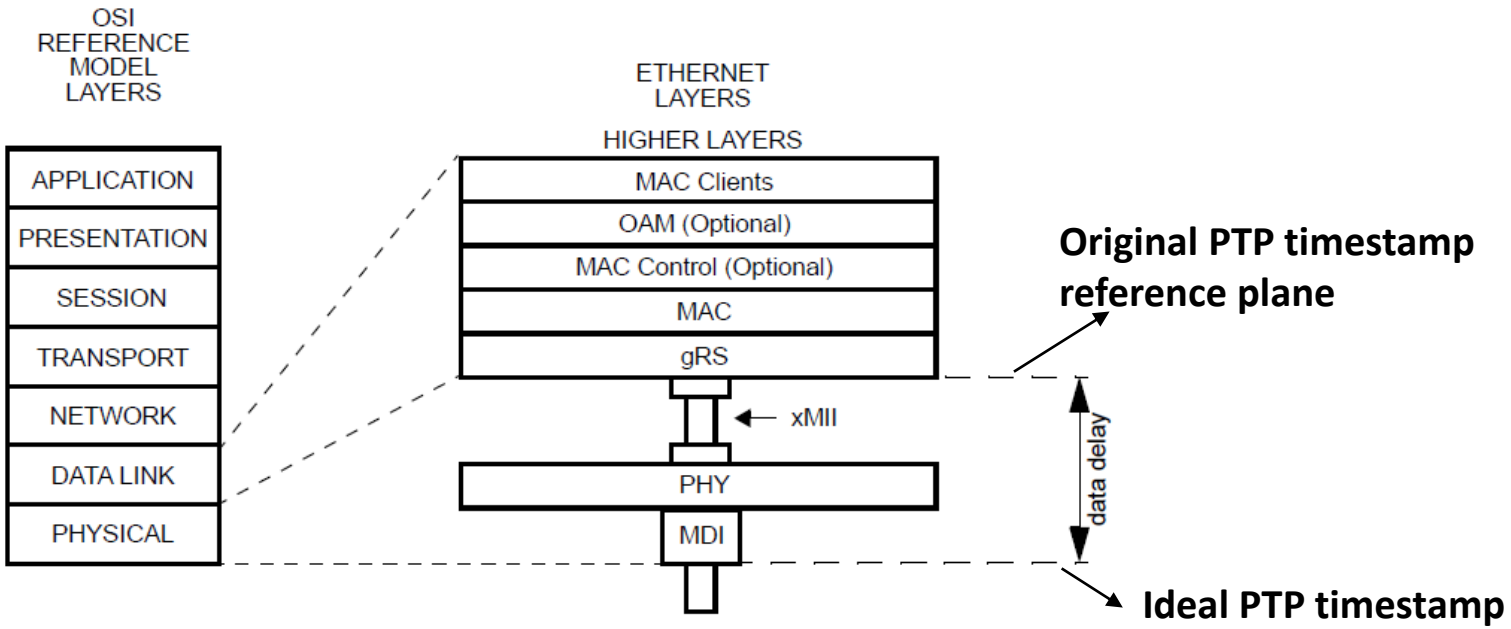


Figure 90–3—Data delay measurement

Accurate compensation of multilane distribution delay

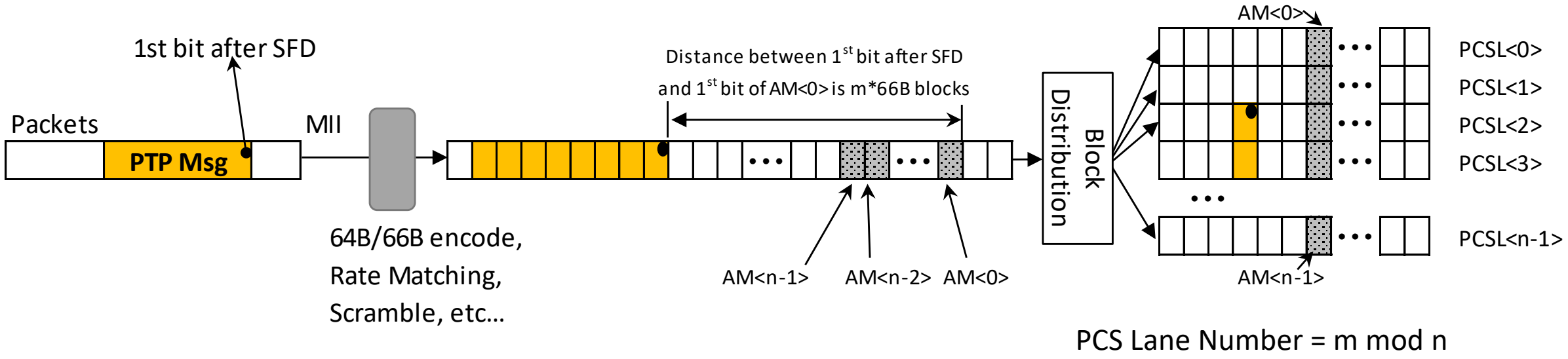
- This slide provides a method how to do the accurate compensation.
  - If the PCS lane number of PTP reference point can be known, the multilane distribution delay can be calculated.
  - For example, if the PTP reference point is carried by lane #3, the multilane Tx distribution delay would be (19 – 3) * 640ps = 10240ps, and the multilane Rx distribution delay would be 3*640 = 1920ps.

Lane	0	actual arvl time	0	actual distrib dly	12160	actual dept time	12160	actual arrival time	12160	actual deskew dly	0	actual merge dly	0	actual arrival time	12160
	1		640		11520		12160		12160		0		640		12800
	2		1280		10880		12160		12160		0		1280		13440
	3		1920		10240		12160		12160		0		1920		14080
	4		2560		9600		12160		12160		0		2560		14720
	5		3200		8960		12160		12160		0		3200		15360
	6		3840		8320		12160		12160		0		3840		16000
	7		4480		7680		12160		12160		0		4480		16640
	8		5120		7040		12160		12160		0		5120		17280
	9		5760		6400		12160		12160		0		5760		17920
	10		6400		5760		12160		12160		0		6400		18560
	11		7040		5120		12160		12160		0		7040		19200
	12		7680		4480		12160		12160		0		7680		19840
	13		8320		3840		12160		12160		0		8320		20480
	14		8960		3200		12160		12160		0		8960		21120
	15		9600		2560		12160		12160		0		9600		21760
	16		10240		1920		12160		12160		0		10240		22400
	17		10880		1280		12160		12160		0		10880		23040
	18		11520		640		12160		12160		0		11520		23680
	19		12160		0		12160		12160		0		12160		24320

  - Then the MAC can compensate the accurate delay into the PTP timestamp generated at xMII.
- Can we predict the lane number where the PTP reference point is?

Lane number prediction based on AM

- AM<0> is ALWAYS on lane 0.
- If the distance between the PTP TS reference point and AM<0> is known, then the PCS lane number where the PTP TS reference point is transmitted on can be accurately predicted.
  - An accurate compensation value thus can be calculated.
  - Similarly, knowing the distance between the PTP reference point and any AM block will also work.



Timestamp reference point tracing over MAC-PHY interconnection

- High speed MAC-PHY connection runs over AUI (e.g. CAUI for 100 GbE), which requires PCS functions in the MAC chip.
  - IEEE 802.3 Figure 83D-1 gives and example of 100 GbE chip-to-chip interface.
- PCS lane distribution is done in 100GBASE-R PCS.
- The number of PCS lane that transmits the PTP TS reference can be either inferred or calculated.
- The PCS lane number is kept unchanged over the whole interface.

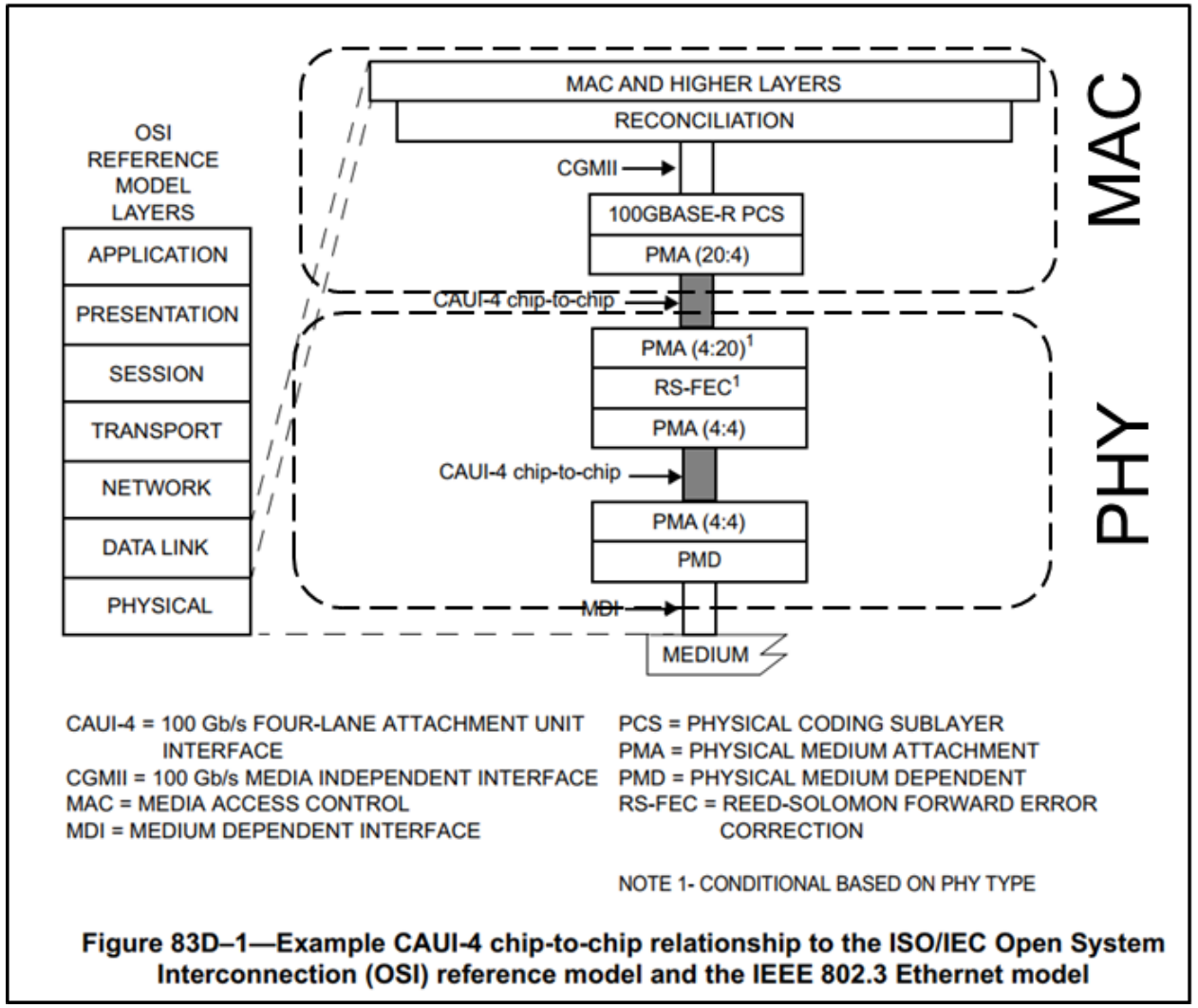


Figure 83D–1—Example CAUI-4 chip-to-chip relationship to the ISO/IEC Open System Interconnection (OSI) reference model and the IEEE 802.3 Ethernet model

Summary

- The multilane distribution delay can be accurately calculated and compensated.
  - Location (# of PCSL) of PTP timestamp reference point can be predicted using the distance between PTP timestamp reference point and the AM.
- It is recommended to compensate the multilane distribution delay so that the timestamp is aligned with the MDI reference plane.
  - To avoid the possible ~6ns time error between the old and new standards.