Discussion on Annex 201A

William Lo

March 9, 2026

Agenda

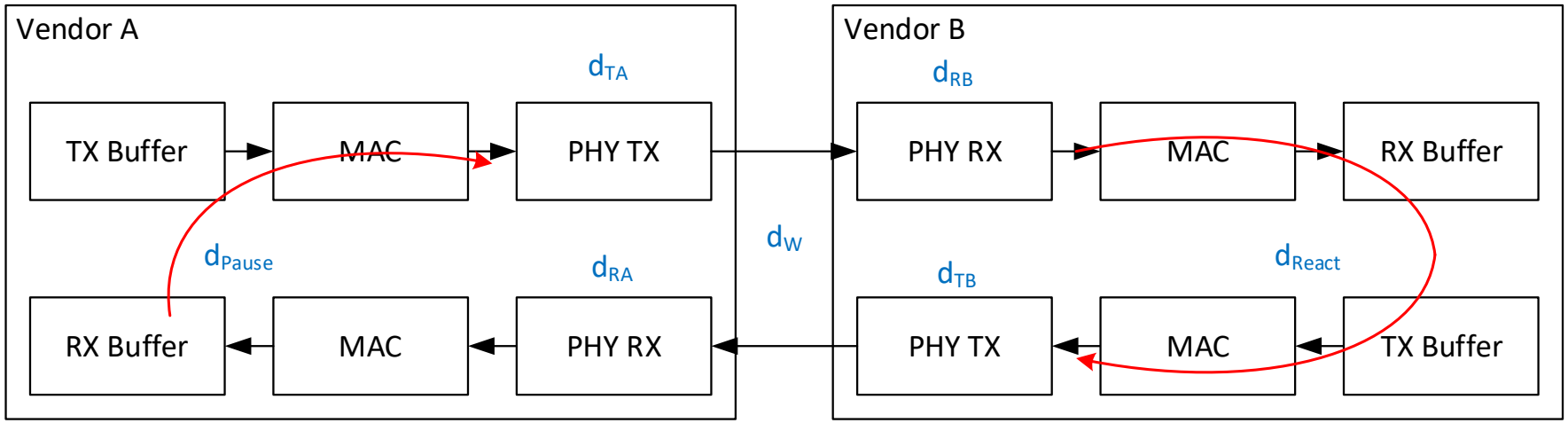
- Why PHY delay limits are needed for flow control calculations
- Show why asymmetrical PHY need finer breakdown of delay limits
- Justify some numbers proposed

Preliminaries – To avoid confusion

- Discussion here is about bit to bit delays through the PHY as it traverses from XGMII to MDI and MDI to XGMII
- It is not about timestamping the SFD of packets at the XGMII relative to some global PTP time.
- However, the isolated PHY delay numbers are very useful for PTP to compensate offsets (Not the subject of the current discussion).

Definitions

- d_{Pause} RX buffer high water mark to pause frame generated on XGMII
- d_{React} Pause frame received on XGMII to TX buffer stopping traffic
- d_{TA}, d_{TB} XGMII to MDI PHY delay for vendor A and B
- d_{RA}, d_{RB} MDI to XGMII PHY delay for vendor A and B
- d_{W} Wire propagation delay. Assumed to be the same both direction
- d_{MTU} Duration of maximum size packet



PHY Delays are Needed to Compute RX Buffer High Water Level

- Want link partner to stop transmitting to prevent RX buffer overflow
- Need enough buffering since reaction is not instantaneous
  - Detect high water and send pause frame d_{Pause}
  - Propagation time through PHYA to PHYB $d_{TA} + d_{W} + d_{RB}$
  - Reaction time to stop transmitting d_{React}
  - Worst case maximum length packet just started before stopping d_{MTU}
  - Propagation time through PHYB to PHYA $d_{TB} + d_{W} + d_{RA}$

Total Delay in Loop

- $d_{Total} = d_{Pause} + (d_{TA} + d_{W} + d_{RB}) + d_{React} + d_{MTU} + (d_{TB} + d_{W} + d_{RA})$
- $d_{Total} = (d_{TA} + d_{RA}) + (d_{TB} + d_{RB}) + (d_{Pause} + d_{React} + d_{MTU} + 2d_{W})$
  - Regroup transmit and receive of each path to be from same device
- $d_{Total} = d_{A} + d_{B} + d_{O}$
- Where:
  - d_{A} and d_{B} are PHY Delay limit specified by IEEE (i.e. Clause 149.10) reported by vendor
    - This is measureable by taking 2 instances of the same vendor's PHY and measuring XGMII to XGMII
  - d_{O} is everything else outside the PHYs
- Implication – vendors can independently assign how much delay to allocate between transmit and receive

Delay Limits in Asymmetrical

- $d_{HighSpeed} = d_{TA} + d_{RB}$
- $d_{LowSpeed} = d_{TB} + d_{RA}$

- Cannot do ($d_{TA} + d_{RA}$) + ($d_{TB} + d_{RB}$) since vendor A transmitter cannot talk to vendor A receiver and similarly for vendor B.

- Implication – Vendor A has to make assumptions on the delay limits of the complementary portions of vendor B.
  - Vendor A + Vendor B may exceed the delay limits
  - Design really aggressively to give more margin to unknown vendor

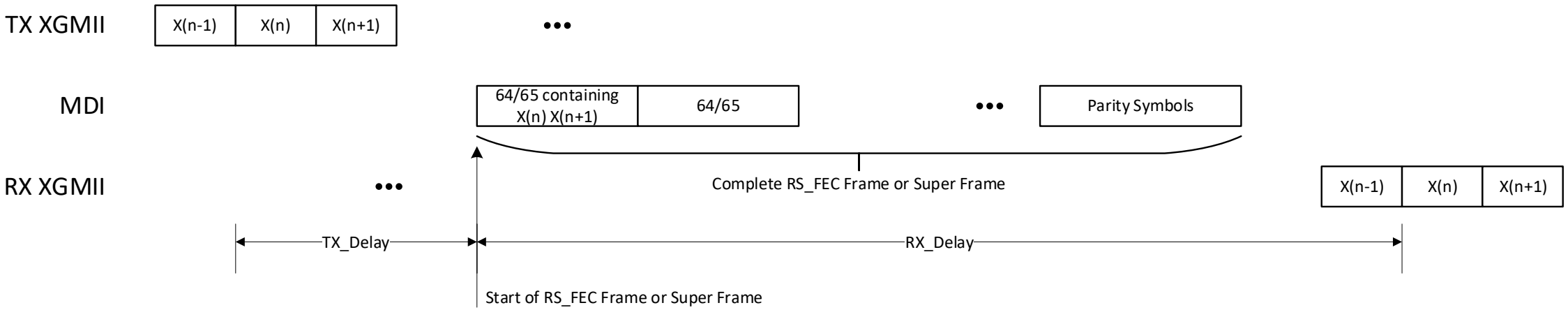
- XGMII to XGMII measurable between each half of vendor A and B
- XGMII to MDI or MDI to XGMII not practically measurable

Annex 201A Purpose

- Since delay measurements to the MDI is not practical it is not realistic to require d_{TA}, d_{RA}, d_{TB}, d_{RB} to be normative
- Measurements to the MDI point can be done in simulations
- Assumes vendors good faith in reporting the delay breakdown

Annex 201A Purpose

- Assign a common point at the MDI so all vendor's reported numbers are consistent
- XGMII to first 64/65 in RS-Frame delay is longer than XGMII to final 64/65 in RS-Frame
- But delay difference on MDI to XGMII on the receive side is reversed



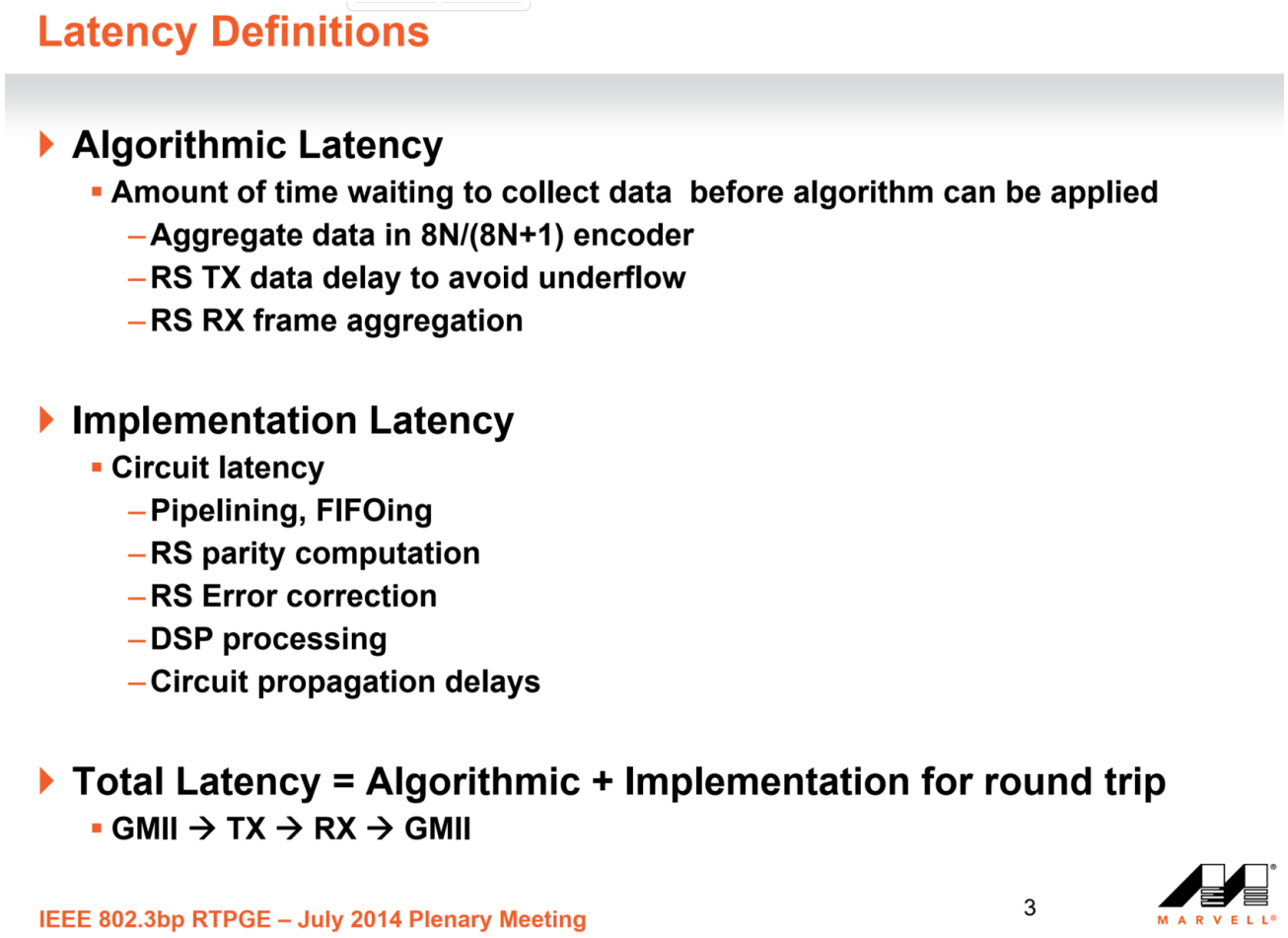
Annex 201A Delay Limit Allocation (Informative)

- The HS_TX TX_Delay is allocated 10% of the delay budget.
- The HS_RX RX_Delay is allocated 90% of the delay budget.
- The LS_TX TX_Delay is allocated 25% of the delay budget.
- The LS_RX RX_Delay is allocated 75% of the delay budget.

- There is no requirement to meet these numbers to claim compliance
- This is just a guideline that if all vendors follow there is no chance that some combination of vendors exceeds the maximum delay limit in a link.

Discussion on Delay Limit Allocation

- The following is based on latency discussions in https://www.ieee802.org/3/bp/public/jul14/Lo_3bp_01a_0714.pdf
- Algorithmic Latency is the minimum theoretical latency
- Implementation latency assumed to be 0 in current discussion as this is vendor dependent



Latency Definitions

- **Algorithmic Latency**
  - Amount of time waiting to collect data before algorithm can be applied
    - Aggregate data in 8N/(8N+1) encoder
    - RS TX data delay to avoid underflow
    - RS RX frame aggregation
- **Implementation Latency**
  - Circuit latency
    - Pipelining, FIFOing
    - RS parity computation
    - RS Error correction
    - DSP processing
    - Circuit propagation delays
- **Total Latency = Algorithmic + Implementation for round trip**
  - GMII → TX → RX → GMII

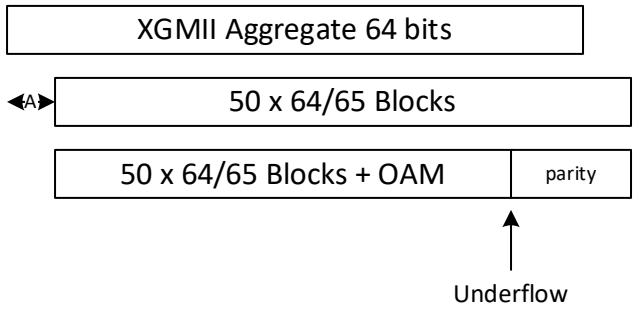
IEEE 802.3bp RTPGE – July 2014 Plenary Meeting

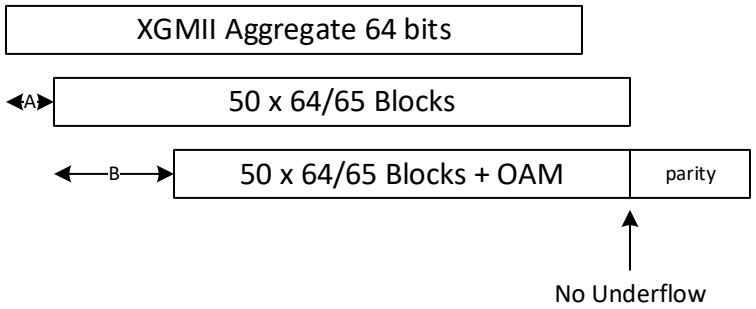
3

MARVELL

Algorithm Latency Transmit

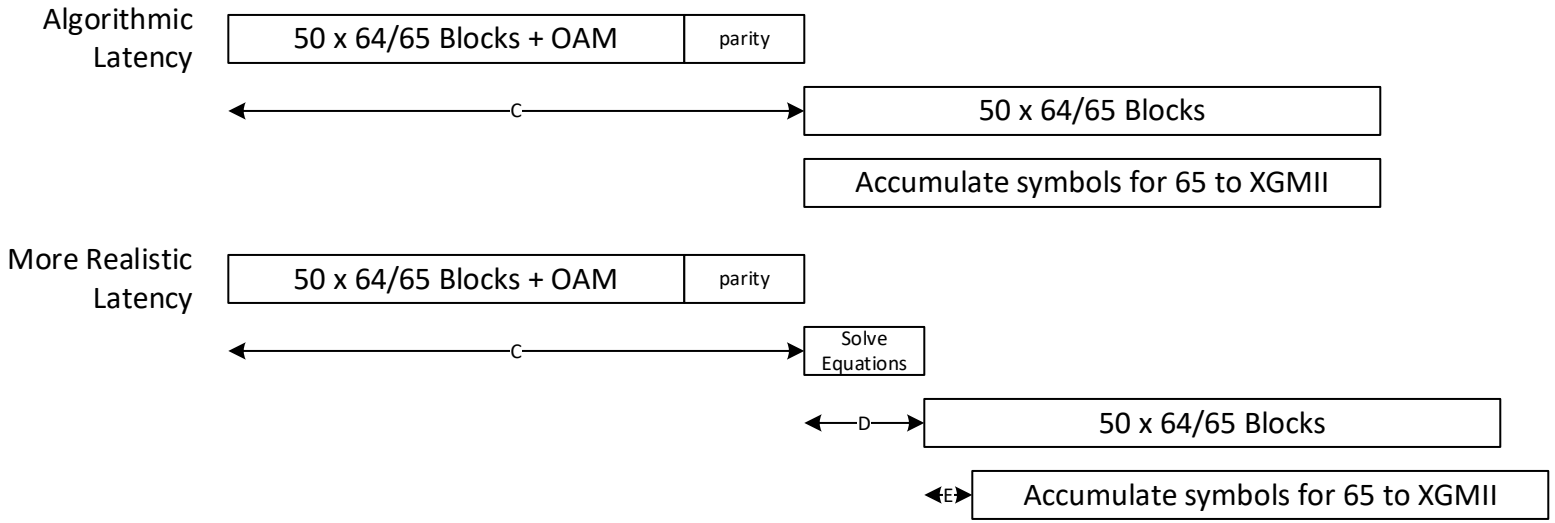
- A = 64 bit aggregation
- B = RS-Encode Underflow
- Everything else assumed to be 0 delay
- 1X interleave in diagram below
- 2X, 4X interleave same concept – parity 2X or 4X longer
- Low speed same concept except 4 x 64/65 + OAM + Reserve





Algorithm Latency Receive

- C = RS-Frame Aggregation and Syndrome Calculation
- Difficult to do the following in zero time
- D = Solve equations for correction
- E = Retrieve and correct symbols and decode to XGMII
- In slow receiver – can decrease D and E with faster clock



Rough Delays – Justification for TX and RX Budget

- All estimates are best case
- Sufficient margin left over for implementation delay.
- Lot more margin given to receiver given complexity of RS decoder and DSP.
- D assumed to be duration of parity symbols
- 100M D & E assume running 4X faster clock

Mode	Interleave	Bit times	Pause Quanta	Delay (ns)	TX Budget	RX Budget	TX Budget (ns)	RX Budget (ns)
2.5G	1x	10240	20	4096.0	10%	90%	409.6	3686.4
5G	1x	10240	20	2048.0	10%	90%	204.8	1843.2
5G	2x	13824	27	2764.8	10%	90%	276.5	2488.3
10G	1x	10240	20	1024.0	10%	90%	102.4	921.6
10G	2x	13824	27	1382.4	10%	90%	138.2	1244.2
10G	4x	20480	40	2048.0	10%	90%	204.8	1843.2
100M	--	512	1	5120.0	25%	75%	1280.0	3840.0

Mode	Interleave	A (ns)	B	C	D	E	Tx Delay	RX Delay	RX Realistic
2.5G	1x	25.6	124.4	1280.0	124.4	25.6	150.0	1280.0	1430.0
5G	1x	12.8	62.2	640.0	62.2	12.8	75.0	640.0	715.0
5G	2x	12.8	124.4	1280.0	124.4	12.8	137.2	1280.0	1417.2
10G	1x	6.4	31.1	320.0	31.1	6.4	37.5	320.0	357.5
10G	2x	6.4	62.2	640.0	62.2	6.4	68.6	640.0	708.6
10G	4x	6.4	124.4	1280.0	124.4	6.4	130.8	1280.0	1410.8
100M	--	640.0	341.3	2560.0	85.3	160.0	981.3	2560.0	2805.3

THANK YOU