Proposal for Multi-PCS Lane Distribution Path Delay Variance

A Leading Provider of Smart, Connected and Secure Embedded Control Solutions

Richard Tse

IEEE 802.3cx Teleconference Sept 23, 2020

Supporters:

- Andras de Koos, Microchip Technology
- Dino Pozzebon, Microchip Technology
- Denny Wong, Xilinx
- Marek Hajduczenia, Charter Communications
- Mark Bordogna, Intel
- Richard Tse, Microchip Technology
- Sriram Natarajan, Cisco
- Ulf Parkholm, Ericsson (for implementation option 2)

Proposal

- Agree on “Soln #3” to deal with data delay variations due to multi-PCS lane distribution

Summary: Multi-PCS Lane Distribution Soln #3

#	Proposed Solution	Pros	Cons	Comments
3	"Method 2" and "Option C" from http://www.ieee802.org/3/cx/public/april20/tse_3cx_02a_0420.pdf Define Tx lane distribution so all lanes transmit their blocks at the same time Define Tx lane distribution time as a constant value, M Define Rx lane multiplexing time as a constant value, N M + N = intrinsic constant delay of both the Tx and Rx multi-PCS lane operations	Conceptually congruent with IEEE 1588 timestamping rules Consistent with how FEC delays are dealt with in 802.3 Because the delays are treated as constants, 802.3 delay registers can be used to record their values (M and N)	Literally incongruent with IEEE 1588 timestamping rules	This generic solution works for all variable delay functions that have mirrored Tx and Rx delays that sum to a constant value. It simplifies the estimation of the delay for a single PHY function or for multiple cascaded PHY functions. Should be compatible with split PHYs as long as each Rx PHY produces an output that is identical to its corresponding Tx PHY's input.

copied from tse_3cx_01_0720.pdf

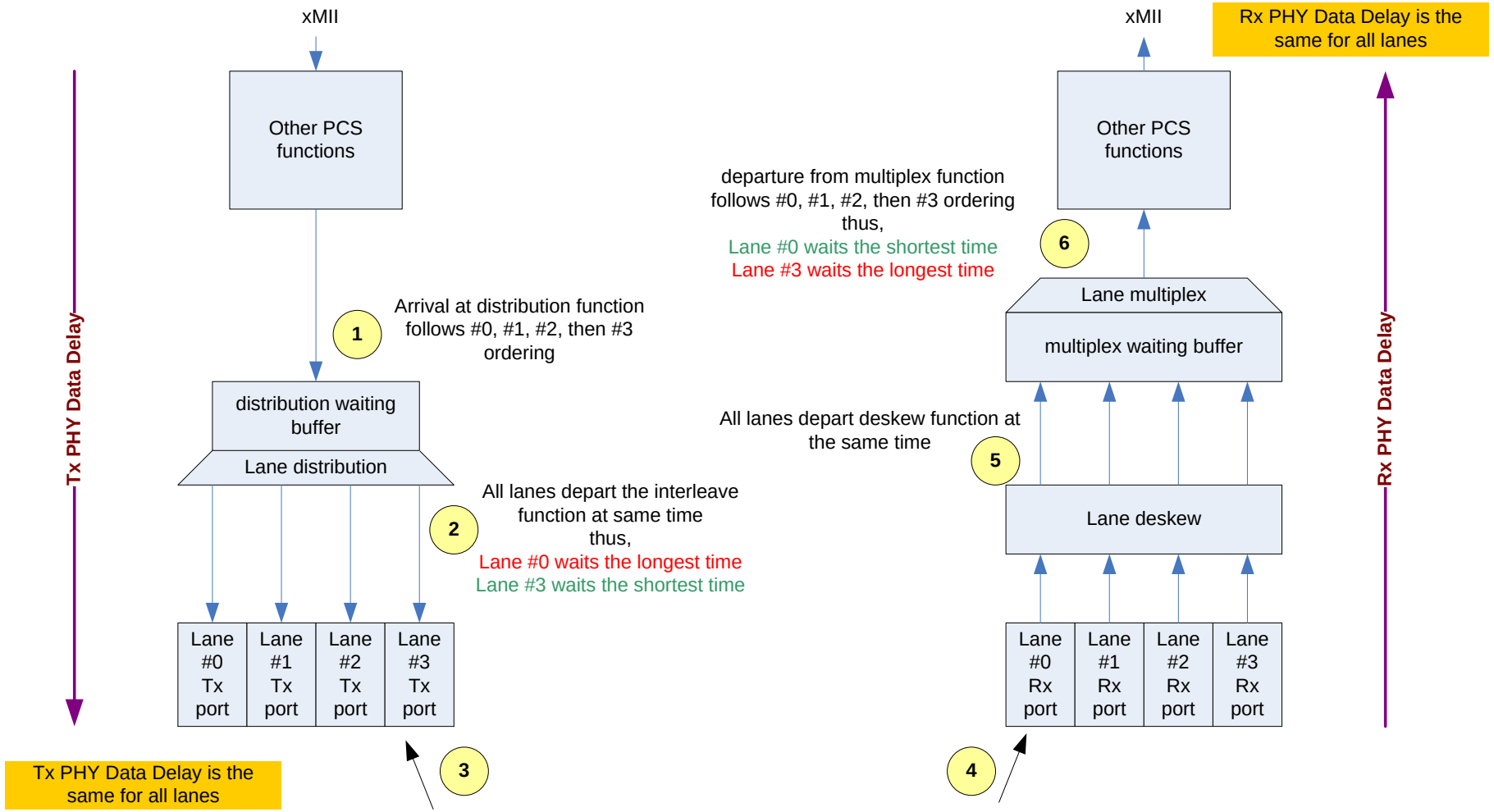
“Soln #3” described again (1/4)

- NxPCS lane Transmitter
  - 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 intrinsic lane-to-lane skew of 66B blocks at MDI

“Soln #3” described again (2/4)

- NxPCS lane Receiver
  - 66B blocks are aligned by deskew buffer
    - No intrinsic lane-to-lane skew of 66B blocks at MDI
    - MDI to xMII path has different path data delay for each lane
      - Data for Lane 0 is released from deskew buffer first and arrives first at xMII, causing smallest path data delay
      - Data for Lane N is released last from deskew buffer and arrives last at xMII, causing largest path data delay
    - Timestamps assume a constant data path delay for all lanes
      - Timestamper at Rx xMII uses the same xMII-to-MDI constant path data delay for every lane

“Soln #3” described again (3/4)



“Soln #3” described again (4/4)

- Other notes
  - The allocation of the constant intrinsic demux/mux delay between the transmit PHY and the receive PHY must be standardized to ensure compatibility of all implementations (e.g., 50% to Tx PHY and 50% to Rx PHY)
  - All implementation-dependent delays (which, in principal, are a constant value) are dealt with independently
  - The use of this constant delay value allows the TimeSync PCS transmit and receive path data delay registers to be used for multi-lane PCS

Proposed Text – implementation option 1 (1/2)

- **Add the following text to Clause 90.7**

  Block distribution in a multi-lane PCS causes variance in the path data delay. Because the data stream crossing the transmit xMII is the same as the data stream crossing the receive xMII, the sum of the transmit block distribution functional delay and the receive block distribution functional delay is the same for every PCS lane.

  For a transmit PHY that performs block distribution from the xMII to multiple PCS lanes (e.g., the 100GBASE-R PCS in clause 82), the path data delay variance experienced by blocks transiting from the xMII to different PCS lanes is treated as a constant value. The constant value that represents the block distribution function's delay is equal to half of the difference between the shortest distribution time from the xMII to a PCS lane (e.g., for lane N of an N-lane PCS) and the largest distribution time from the xMII to a PCS lane (e.g., for lane 0).

Proposed Text – implementation option 1 (2/2)

- **Add the following text to Clause 90.7, continued…**

  For a receive PHY that performs block distribution from multiple PCS lanes to the xMII (e.g., the 100GBASE-R PCS in clause 82), the path data delay variance experienced by blocks transiting from the per-lane outputs of the deskew buffer to the xMII is treated as a constant value. The constant value that represents the block distribution function's delay is equal to half of the difference between the shortest distribution time from the output of a deskew buffer lane to the xMII (e.g., for lane 0) and the largest distribution time from the output of a deskew buffer lane to the xMII (e.g., for lane N of an N-lane PCS).

  The constant value for the receive PHY is equal to the constant value for the transmit PHY. This constant value can be used to represent the multi-lane block distribution function's portion of the PCS delay when using the TimeSync PCS transmit path data delay and the TimeSync receive path data delay.

Proposed Text – implementation option 2

- **Enhance existing text in 90.7 on FEC so it also deals with multi-lane PCS.**
- **Replace "SFD" with "message timestamp point" throughout 90.7 (not all are shown below)**
- **Insertions are highlighted in blue and deletions are highlighted in ~~red~~.**

For a PHY that includes an FEC and/or multilane distribution functions, the transmit and receive path data delays may show significant variation depending upon the position of the ~~SFD~~message timestamp point within the FEC block and in the multilane distribution sequence. 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~~message timestamp point is at the start of the FEC block and multilane distribution sequence. For PHYs with both FEC and multilane distribution, the start of the FEC block is guaranteed to coincide with the start of a multilane distribution sequence.

Questions?

Thanks!