Benefits of Supporting RTS Across Optical Links

Ali Ghiasi – Ghiasi Quantum/Marvell
Mike Dudek – Marvell
Whay Lee – Marvell

IEEE 802.3dj Task Force
Adhoc Meeting

April 18, 2024

Overview

- Why optical link up is challenging
- Adopted multi-segment LT with RTS
- Improving optical link up with RTS
- Propagating RTS across optical links
- Defining optical PMDs as duplex group
- Summary.

Contributors: Adee Ran and Gary Nicholl

Optical Link Up Process Outdated

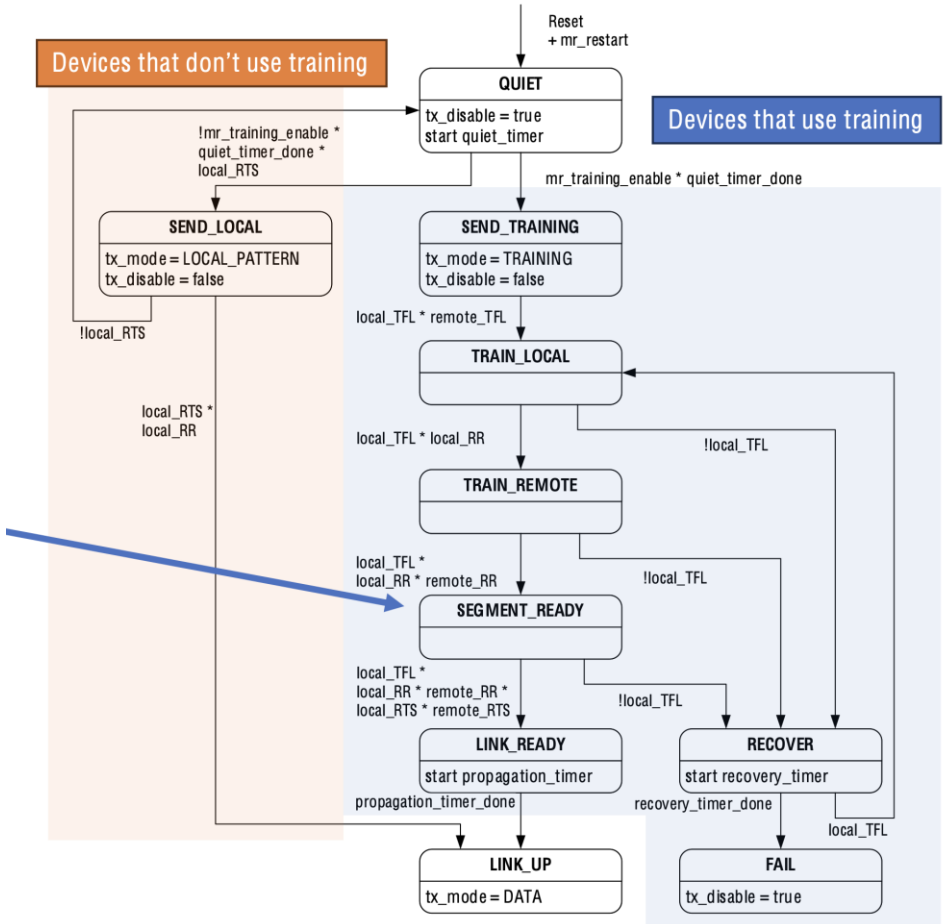
- **Current IMDD link up goes back to early days of FC**
  - Early optics were unretimed without any equalization on AUIs/PPI or optics with PCS relying on SD (Signal Detect)
- **What is making the current optical link up cumbersome and complex**
  - CMIS initializes module and data paths
  - CMIS adjust AUI SI parameters, then module enables it's optical TX
  - SD and CDR locks are used as indicator of good optical signal
  - As the module CDR/DSP is calibrating its ADC, recovering, equalizing, and propagating data to host there is no guarantee good data is being transmitted to the optical receiver
  - Downstream electrical link may simultaneously be trying to adapt/equalize to "not guaranteed good data" and may end up adapting to invalid data
  - Optical links are multi-segments and complex, and inherently unpredictable to host-module and PMD-PMD interactions
  - By relying on SD/CDR lock to unsquelch module out may result sending invalid data to the host that require host keep resetting the module Egress DSP and hoping for the best!

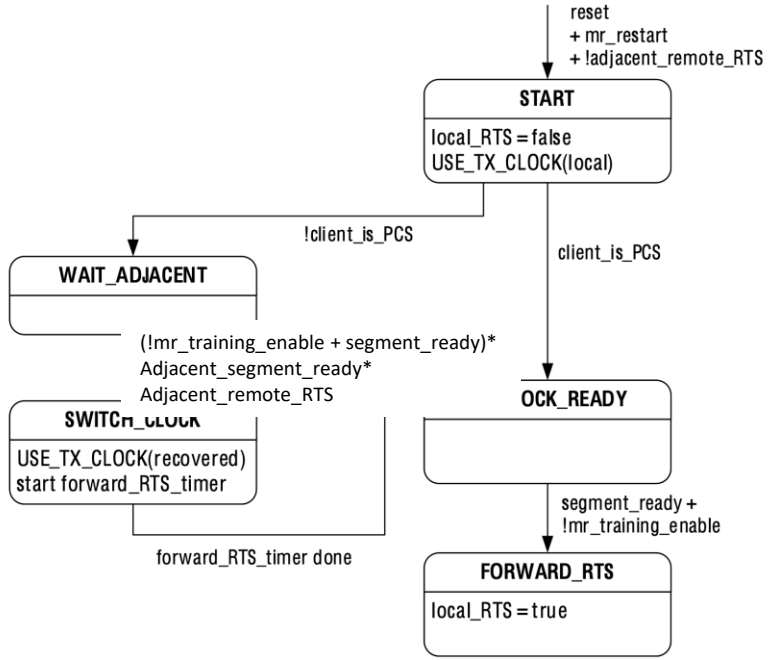
Improving Optical Link Up Process

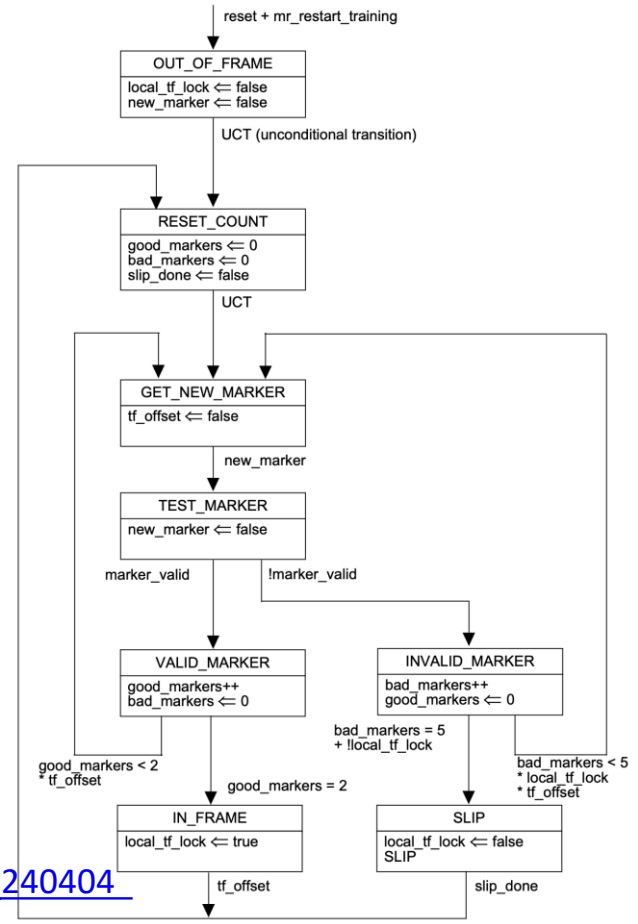
- **OLT (Optical Link Training) start up with known NRZ PRBS then switching to PAM4 improve optical receiver/DSP calibration and adaptation**
- **OLT passing RTS, ran_3dj_elec_01_240229, will improve reliable link up on both AUIs segments and optical link**
  - Optical SD/CDR lock are not always a reliable indicator of signal goodness and may falsely send RTS onto local AUI
  - The real benefit of RTS is not realized for optical links unless RTS propagates end-end from PCS-PMA to PMA-PCS
- **Reliable and predictive optical link up is required for 200G optics considering 100G optics pain points**
  - OLT with RTS will improve 200G optics to have more reliable and consistent link up in a timely manner!

Adopted Multi-Segment LT with RTS

- **Leverage DJ proposed electrical link training [lusted_3dj_02_2401] as well as RTS from [ran_3dj_elec_01_240229] with goal to stay align with the AUI segment by segment training.**







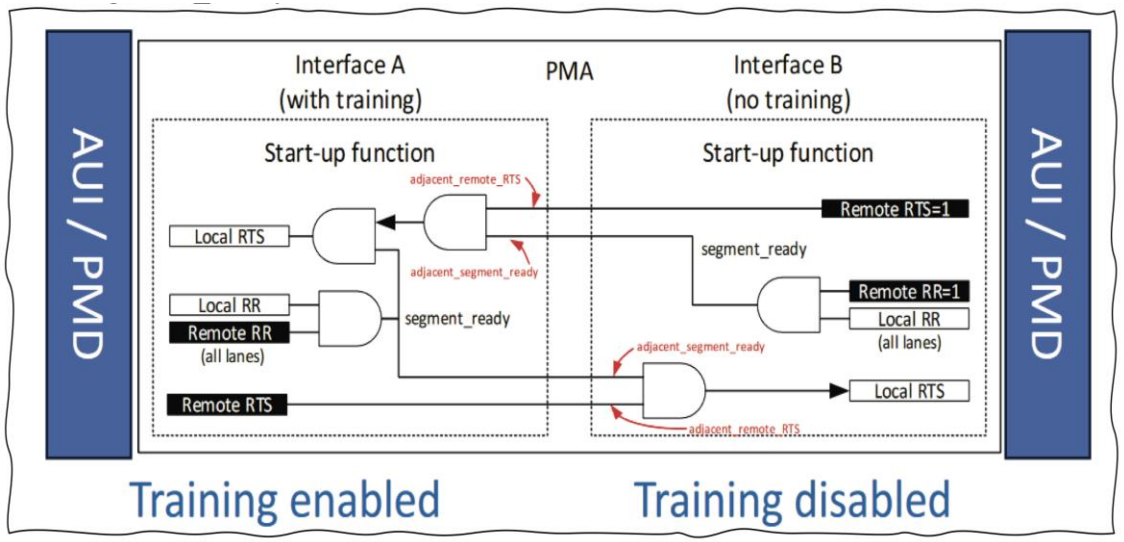
*RTS State Diagram Per Updated [Ran_3dj_elec_01_240404]

Figure 136–8—Training frame lock state diagram

PMD control flow Fig 136-7 replaced with Ran control diagram with RTS

Segment Ready and RTS Propagation

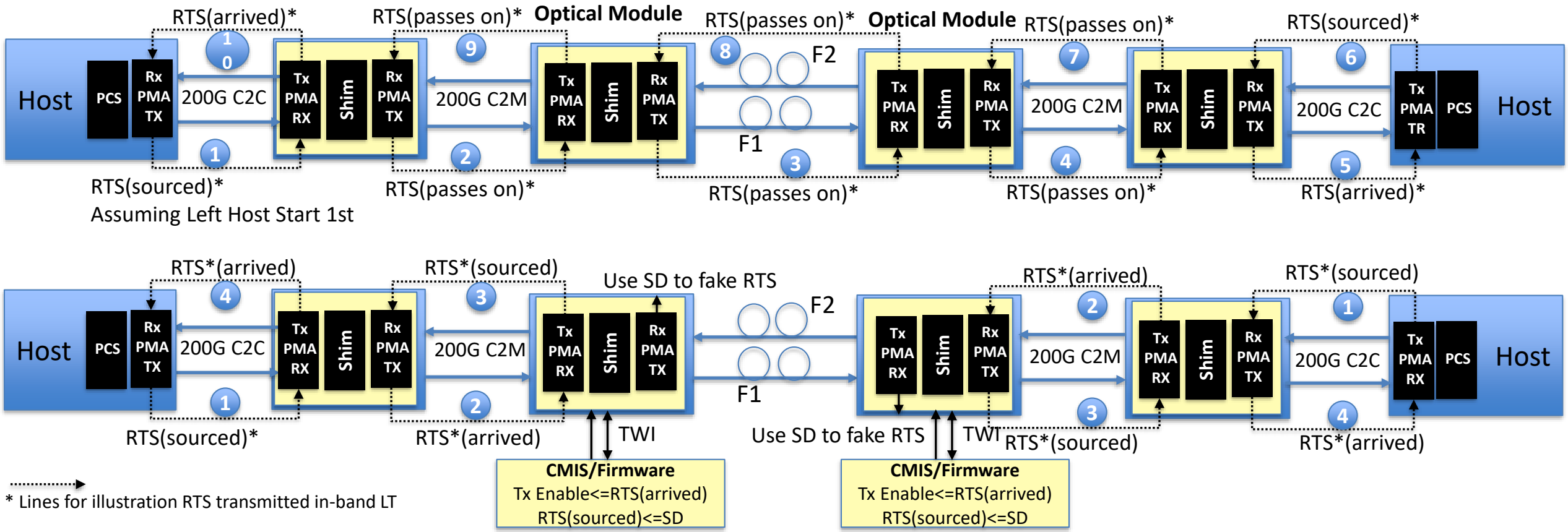
- **If training is disabled or next segment (optical PMD) doesn't support training then Remote RTS=1 and RR=1 permanently with local AUI not taking advantage of propagated RTS**
  - Local AUI instead of relaying on remote RTS instead relies on PMA unsquelching (Local RR) driven by the optical PMD OMA level or CDR lock without knowledge of Remote RR
    - OMA/CDR lock indicator are not a reliable link up indicator
  - Even if OMA/CDR lock were reliable link up indicator there is no guarantee when the remote PMD/PMA will unsquelch
  - Multi-segment link with optical segment are most problematic regarding reliable and predictable link up and will benefit the most from RTS.



See [24_0404/ran_3dj_elec_01_240404]

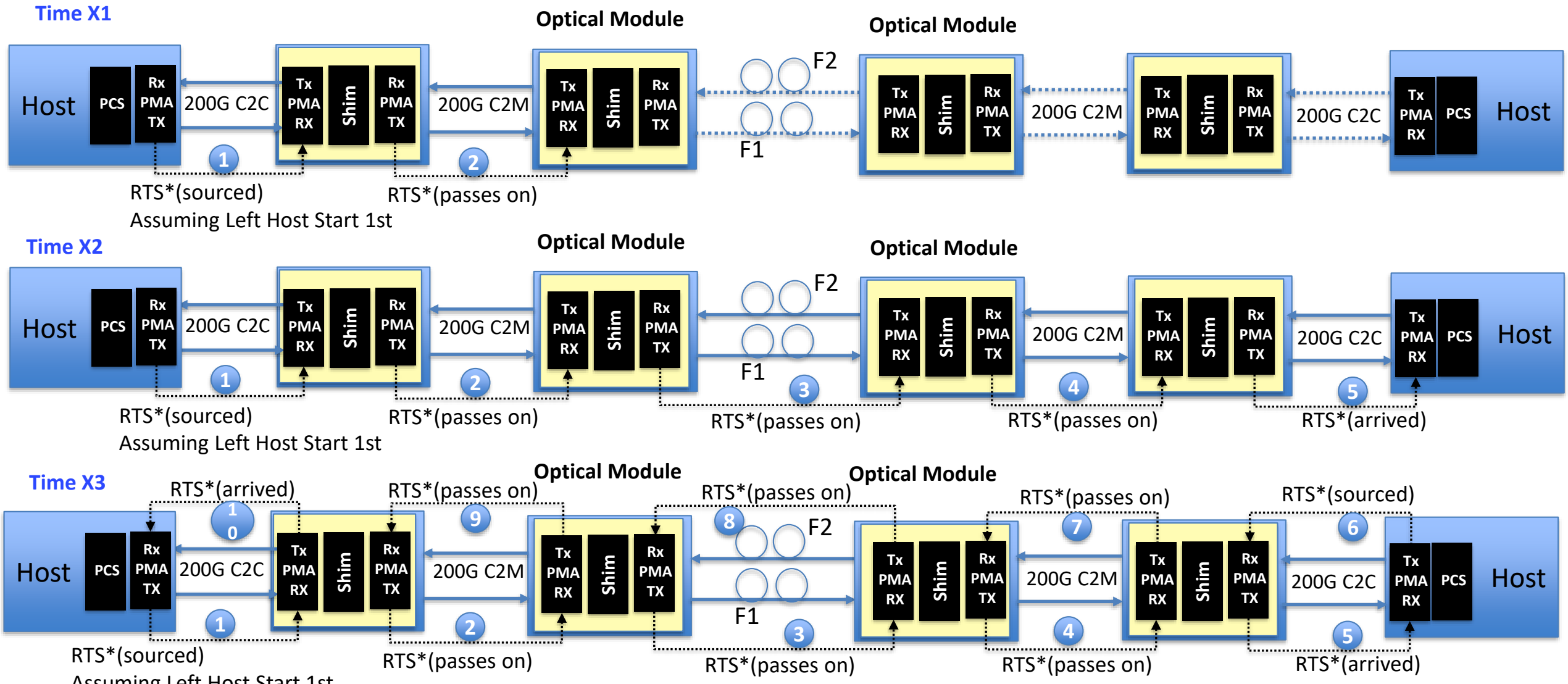
RTS Propagation

- **RTS propagated across optical link with OLT offers reliable-consistent link up in timely manner**
  - RTS terminated in the optical module doesn't address optical link up challenges
  - Generating RTS based on not always reliable SD/CDR lock can exasperate optical link up!



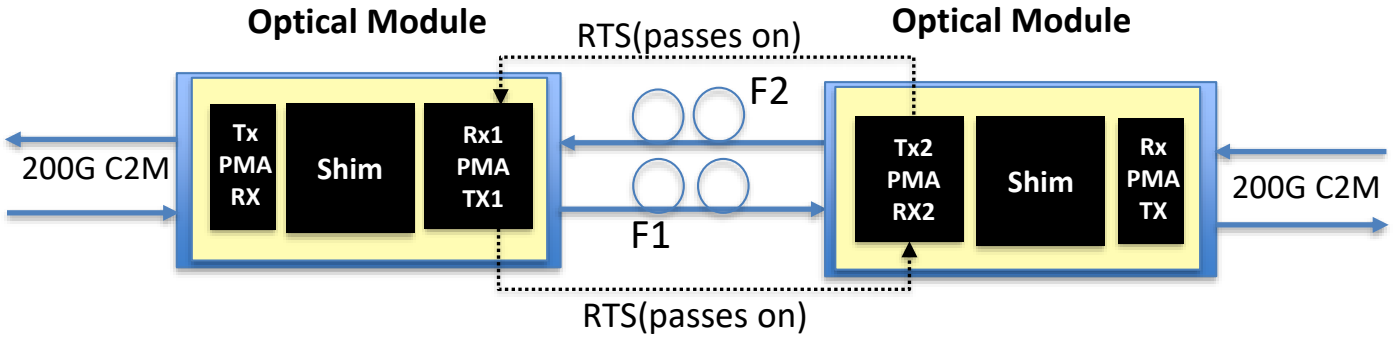
* Lines for illustration RTS transmitted in-band LT

Example of RTS Propagation



Optical OLT is Needed to Pass RTS

- **Propagate the RTS from the 200G AUI link across the 200G optical links**
  - In case of 200G optical PMDs operating with 100G AUI RTS not utilized
  - In case of 200G optical PMDs with 100G AUI on one end and 200G AUI on the other end, RTS is sourced and terminated in the 200G module PMA attached to 100G AUI
- **Illustration F1/F2 is for one duplex link, OLT can be supported across any 1 to N lanes PMDs as long as Tx/Rx pairing are preserved.**



Optical OLT PMD Requirements

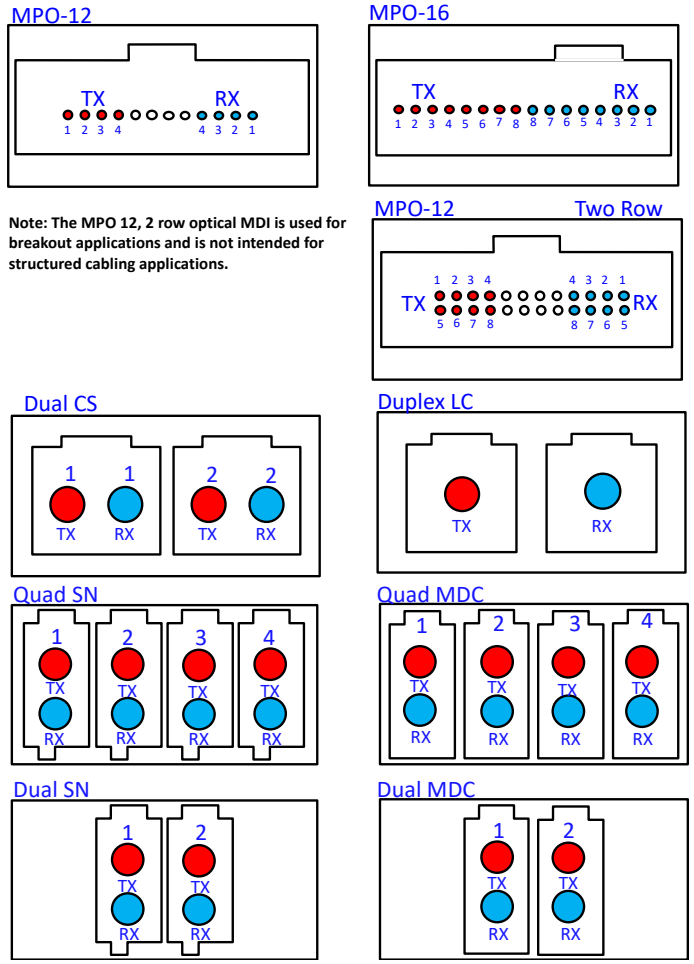
- **Transmit and receive must be grouped in duplex pairs PMDs to support optical loop back and breakout applications**
  - Current definition of optical PMDs
    - TX TX TX TX - RX RX RX RX
    - Was defined for connivence of routing any TX to RX lanes
  - We would need to define the optical PMDs as following
    - TX1 TX2 TX3 TX4 - RX4 RX3 RX2 RX1
    - Support breakout and optical loopback
- **In IEEE we have not grouped the TX/RX into duplex pairs but actual products are based on duplex pair grouping**
  - QSFP-DD MSA and OSFP MSA specify TX/RX lane # on the MDI
  - CMIS associate PMA lanes to optical PMD lanes – otherwise you wouldn't know which transmitter was turned on
- **Structure cable plants compliant to ANSI/TIA-568.3-E preserve duplex pairs and lane grouping.**

Both CMIS and Optical Modules Enforce Tx/Rx Pairs

- **CMIS data-path pairs electrical Tx/Rx lanes to optical Tx/Rx lanes**
  - Table and figure below from QSFP-DD MSA Rev. 7.0 illustrates the concept.

Electrical data input/output	Optical port mapping (see Figure 15)				
	Duplex LC, CS, SN, or MDC	MPO-12, Dual (CS, SN, MDC, Duplex LC, or MPO-12)	MPO-12, Quad (SN or MDC)	MPO-12 (two row), MPO-16, or Dual MPO-12	MPO-12, SN, MDC (BiDi)
	1 TX fiber 1 RX fiber [1]	2 TX fibers 2 RX fibers [1]	4 TX fibers 4 RX fibers [1]	8 TX fibers 8 RX fibers [1,3]	8 Tx (Rx) fibers [2,3]
Tx1	TX-1	TX-1	TX-1	TX-1	TR1
Tx2				TX-2	RT1
Tx3			TX-2	TX-3	TR2
Tx4				TX-4	RT2
Tx5		TX-2	TX-3	TX-5	TR3
Tx6				TX-6	RT3
Tx7			TX-4	TX-7	TR4
Tx8				TX-8	RT4
Rx1	RX-1	RX-1	RX-1	RX-1	RT1
Rx2				RX-2	TR1
Rx3			RX-2	RX-3	RT2
Rx4				RX-4	TR2
Rx5		RX-2	RX-3	RX-5	RT3
Rx6				RX-6	TR3
Rx7			RX-4	RX-7	RT4
Rx8				RX-8	TR4

Notes:
1. TX-n or RX-n where n is the optical port number as defined Figure 15.
2. TRn or RTn where n is the optical port number as defined Figure 15.
3. Some QSFP-DD/QSFP-DD800/QSFP-DD1600 modules may require fewer CS, SN, or MDC connectors. In such cases, Port #1 is always the left-most port. Successive ports then follow sequentially from left-to-right as shown in Figure 15.



Summary

- **The OLT training facility starting with NRZ then transition to PAM4 provides robust receiver/ADC/DSP start up**
- **Multi-segment link training with RTS is a very important feature that has been adopted in the 802.3dj AUI links**
  - Unless we carry RTS across optical link from PCS-PCS multi-segments AUI LT with RTS benefits will not be realized
- **RTS transmitted across optical links provides graceful mechanisam to transition from training to PCS data in predictable and timely manner**
- **The task force need to adopt the basic OLT facility even if the analog aspect of optical training isn't enabled just to support multi-segment training with RTS**
  - Basic OLT facility will address many of the 100G optical link/DSP start up pain-point issues
- **RTS propagated across optical link with OLT will provide for 1st time reliable and predictable optical/AUI link up process**
- **Maintaining optical lane assignments within a PMDs already done to support break-out and remote/local loop-backs**
  - ANS/TIA-568.3-E Optical Fiber Cabling and Components Standard also maintain duplex pairs and lane grouping
  - Just need to simply define MDI lane definition with lane grouping such as TX1TX2TX3TX4-RX4RX3RX2RX1 instead of TXTXTXTX-RXRXRXRX.