For Reference: Transmitter Training Proposal for 256GFC

Anil Mehta, Ramana Murty, John Johnson

BROADCOM®

Supporters

Cathy Liu - Broadcom

Jeff Slavick – Broadcom

Kent Lusted – Intel

Luz Osorio - Ciena

Purpose

- Link Training is a topic in industry for 200-256G/lane rates
- There has been interest in the transmitter training discussions occurring in INCITS FC
- This information is provided to the IEEE P802.3dj Task Force as potential reference material for Ethernet

Multimode Links

- The benefit of link training for multimode links is extended reach.

  Greater pre-emphasis can be applied without excess overshoot/undershoot in the optical waveform when it is determined that the fiber bandwidth falls below a set value. In practice, this can be implemented by have two (ore more) presets for the pre-emphasis setting that are cycled through at link bring up.

- Example of a 128GFC link with two presets T11-2022-00090-v000.pdf

  - Transceiver test: Capture one waveform with each preset

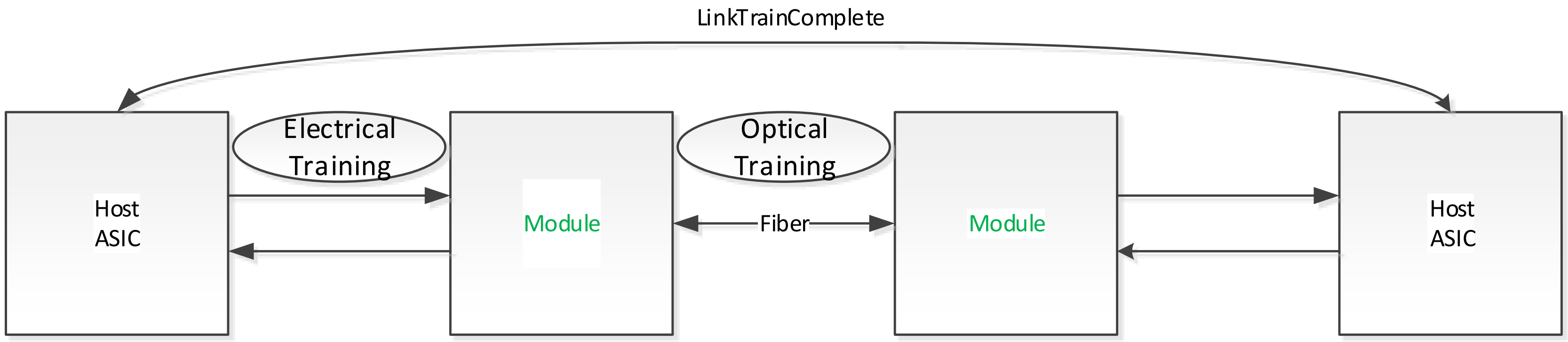
Preset 1		**Preset 2**	
OM4 100 m			OM4 70 m
Filter	Cascaded 28.05 and 18.6 GHz	Filter	Cascaded 28.05 and 26.6 GHz
TDECQ	max 4.4 dB		TDECQ max 4.4 dB
OM4 70 m		OM4 2 m	
Filter	Cascaded 28.05 and 26.6 GHz	Filter	28.05 GHz
Overshoot	max 29%	TECQ	max 4.4 dB
		Overshoot	max 29%

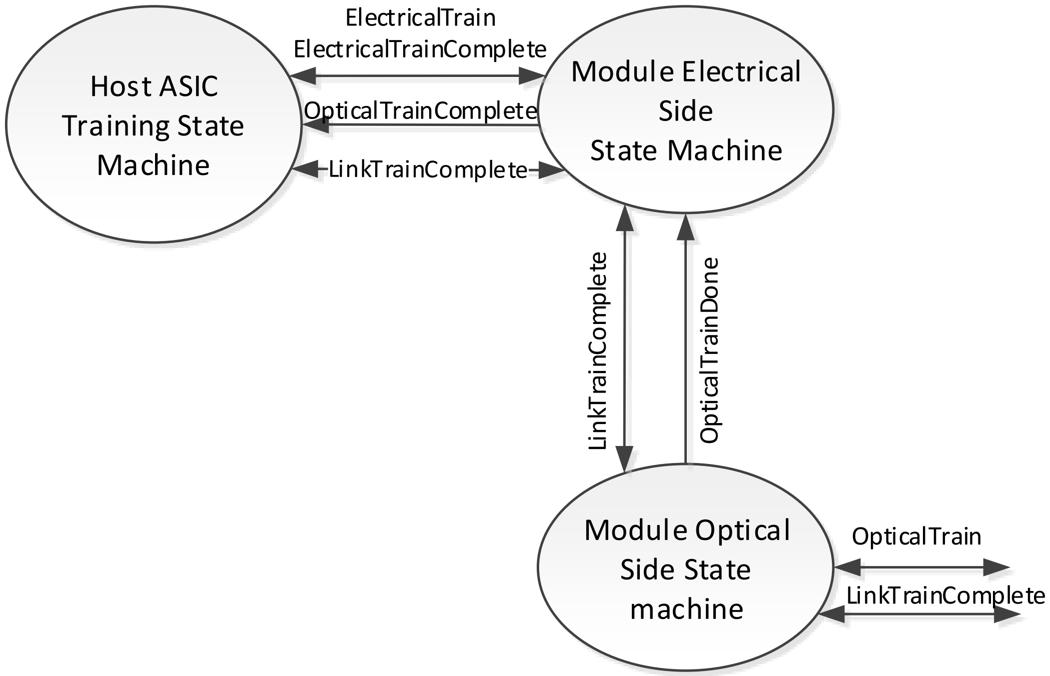
  - If implemented, most links will use preset 2 based on fiber link length statistics T11-2023-00122-v000.pdf

Singlemode Links

- As the signal rate increases, operation over a wide dynamic range of link properties becomes ever more challenging.
- Link training can be a tool to adapt and optimize transceiver performance over a broad range of expected singlemode optical link properties.
- Potential applications of link training for singlemode optical PMDs
  - Optical insertion loss dynamic range
    - Adjust TX launch power to avoid RX overload or minimize TX power dissipation
    - Example of APD RX damage prevention in BiDi access links, 3dk_effenberger_2307_1
  - Chromatic dispersion dynamic range
    - Adjust TX chirp to best match actual CD. 200G EAM example, johnson_3df_01a_221011
    - Adjust TX EQ for CD pre-compensation and minimize noise enhancement from RX FFE
    - Bypass of inner FEC to match actual CD and latency objective, ghiasi_3dj_01a_2307
  - Four wave mixing mitigation
    - Adjust TX wavelength to prevent FWM phase-matching, kuschnerov_3dj_01a_230206
  - Interoperability between vendors and technologies
    - Adjust TX FIR to optimize signal for actual RX bandwidth, maximize BER margin
    - Bypass of pre-coding to match actual RX configuration, ghiasi_3dj_01a_2307

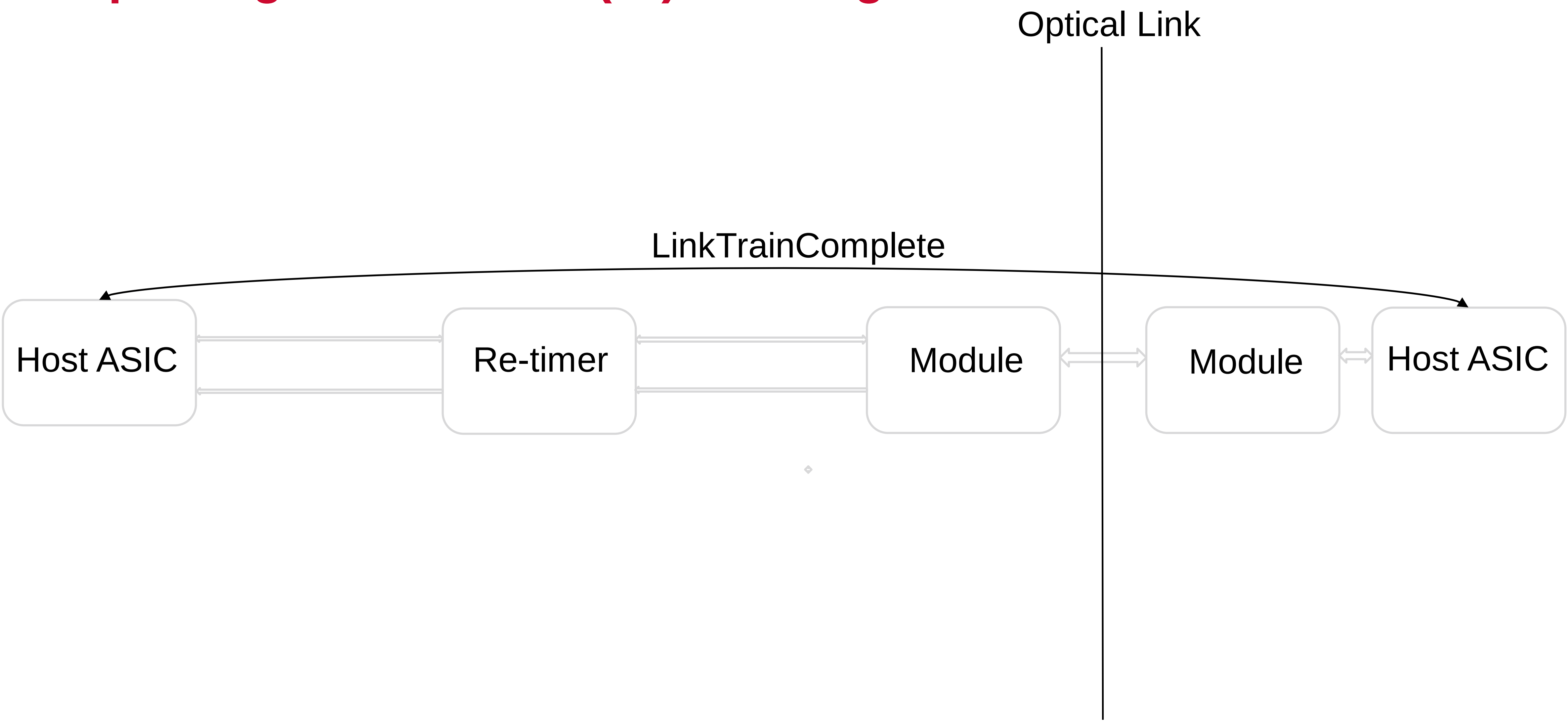
Link training for Electrical/Optical Interfaces





BROADCOM

Multiple Segment In Band(IB) Training



Re-Timer/Module/ASIC State Machines

- Host End(HE) – Side closest to Host ASIC
- Optical End(OE)– Side closest to Optical Fibre
- Separate State Diagrams for HE and OE State Machines
- Rx_TTS_Status[15] from the Host ASIC used to go into flow through mode on Re-Timers and Modules in the path

Currently Defined Training Frame Control Field with Proposed New Bits

Table 162–9—Control field structure

Bit(s)	Name	Description
15:14	Reserved	Transmit as 0, ignore on receipt
13:11	Initial condition request	13 12 11 1 1 1 = Reserved 1 0 1 = Reserved 0 1 1 = Preset 5 0 0 1 = Preset 4 1 1 0 = Preset 3 1 0 0 = Preset 2 0 1 0 = Preset 1 0 0 0 = Individual coefficient control
10	Reserved	Transmit as 0, ignore on receipt
9:8	Modulation and precoding request	9 8 1 1 = PAM4 with precoding 1 0 = PAM4 0 1 = Reserved 0 0 = PAM2
7:5	Reserved	Transmit as 0, ignore on receipt
4:2	Coefficient select	4 3 2 1 0 1 = $c(-3)$ 1 1 0 = $c(-2)$ 1 1 1 = $c(-1)$ 0 0 0 = $c(0)$ 0 0 1 = $c(1)$
1:0	Coefficient request	1 0 1 1 = No equalization 1 0 = Decrement 0 1 = Increment 0 0 = Hold

→ 6:HostEnd Train

→ 5:OpticalEnd Train

BROADCOM

Currently Defined Training Frame Status Field with proposed New Bits

Table 19 - 64GFC Training Frame Status field

Bits	Field name	Content
15	Receiver Ready	Set to one: training is complete and receiver is ready for data. Set to zero: request for Training to continue.
14	SN	Set to one: transmitter has not completed LSN. Set to zero: transmitter has completed LSN.
13	Reserved	Transmit as zero, ignore on receipt.
12	TF	Set to one: transmitter is operating with Fixed Coefficients. Set to zero: transmitter coefficients may be trained by the receiver.
11-10	Modulation and Precoding Status	Set to 11b: PAM4 with precoding. Set to 10b: PAM4. Set to 01b: reserved Set to 00b: PAM2
9	Receiver frame lock	Set to one: frame boundaries identified. Set to zero: frame boundaries not identified.
8	Initial Condi-tion Status	Set to one: updated. Set to zero: not updated.
7	Parity	Parity bit to provide DC balance.
6	Reserved	Transmit as zero, ignore on receipt.
5-3	Coefficient Select Echo	Set to 110b: c(-2) Set to 111b: c(-1) Set to 000b: c(0) Set to 001b: c(1)
2-0	Coefficient Status	Set to 111b: reserved Set to 110b: coefficient at limit and maximum voltage Set to 101b: reserved Set to 100b: maximum voltage Set to 011b: coefficient not supported Set to 010b: coefficient at limit Set to 001b: updated Set to 000b: not updated

→ Change name to LinkTrainComplete(Used to Signal Completion of Training on the entire Link)

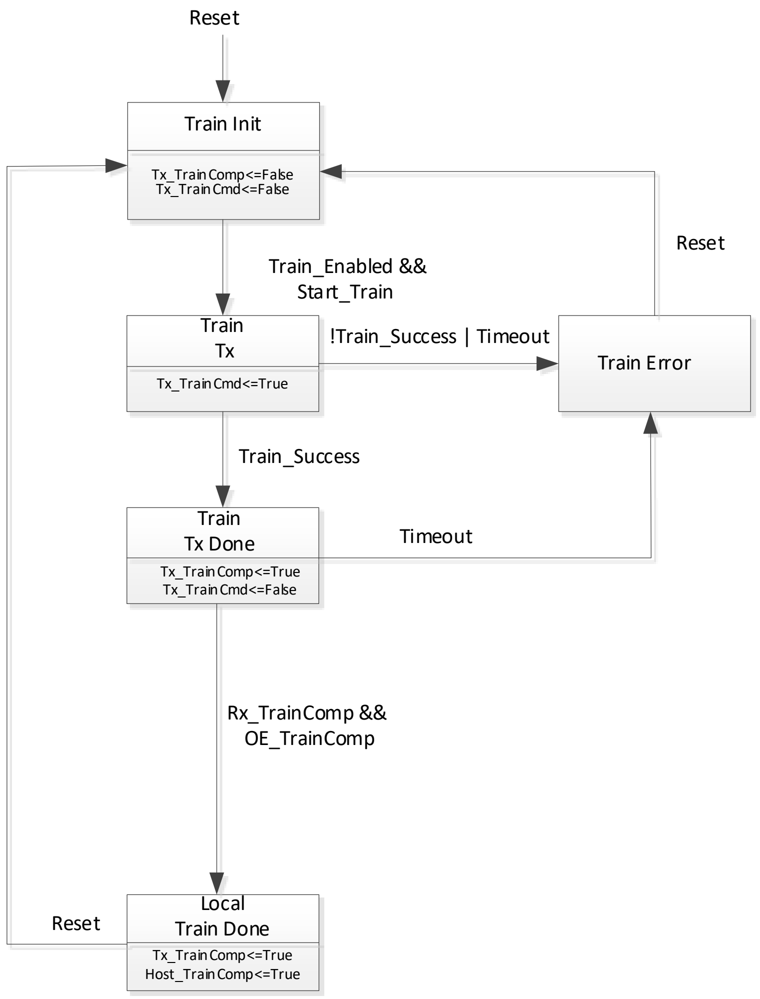
→ HostEndTrainComplete

→ OpticalEndTrainComplete

→ Set to 101 for c(-3)

BROADCOM

Host ASIC State Machine

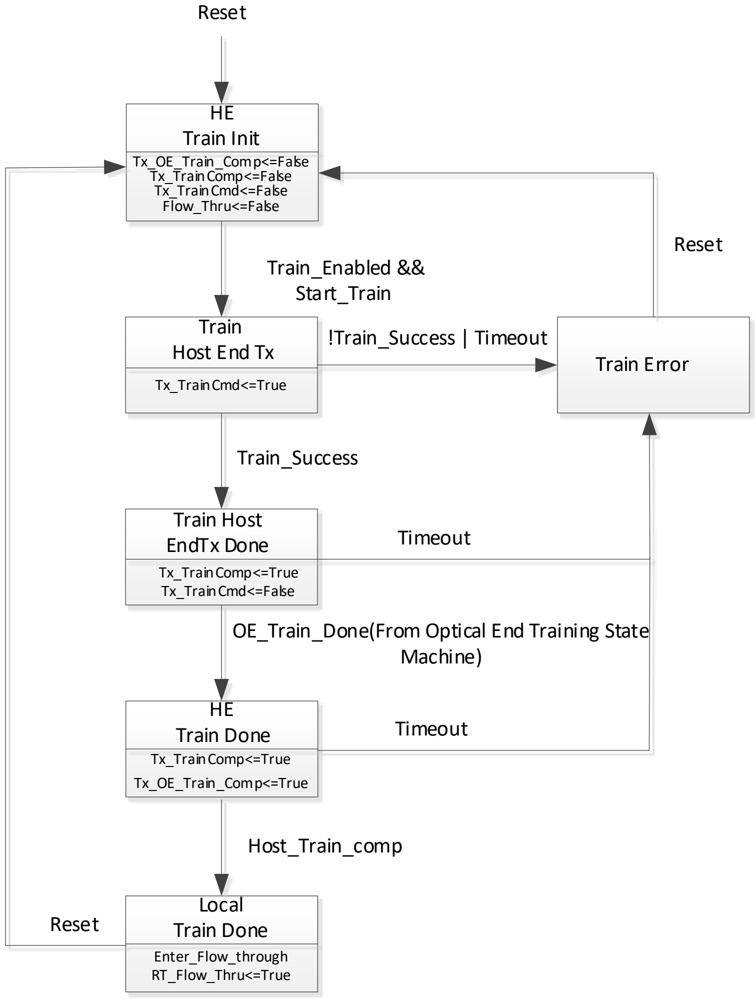


Tx_TrainCmd = Tx_TTS_Control[5]
Tx_TrainComp = Tx_TTS_Status[6]
OE_Train_comp= Rx_TTS_Status[6]
Rx_TrainComp <= Rx_TTS_Status[13]
Host_TrainComp = Tx_TTS_Status[15]

Host ASIC Training State Machine

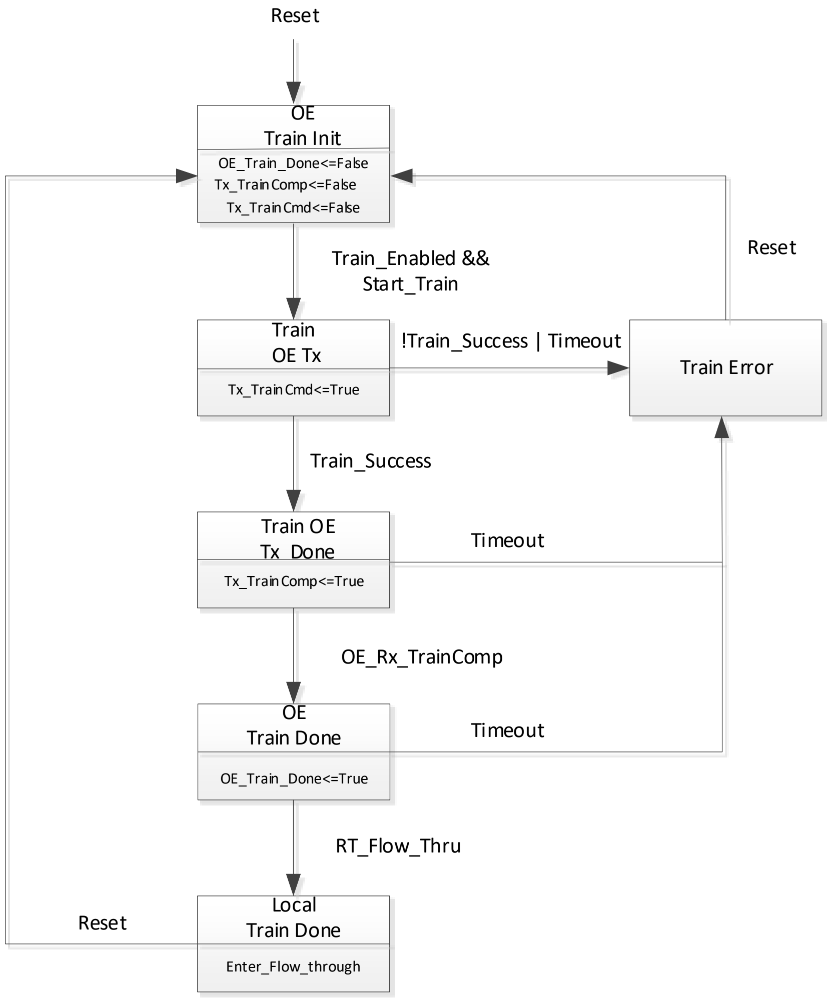
BROADCOM

Module State Diagrams



Tx_TrainCmd = Tx_TTS_Control[6]
Tx_TrainComp = Tx_TTS_Status[13]
Tx_OE_Train_comp= Tx_TTS_Status[6]
Host_Train_comp = Rx_TTS_Status[15]

Host End Training State Machine

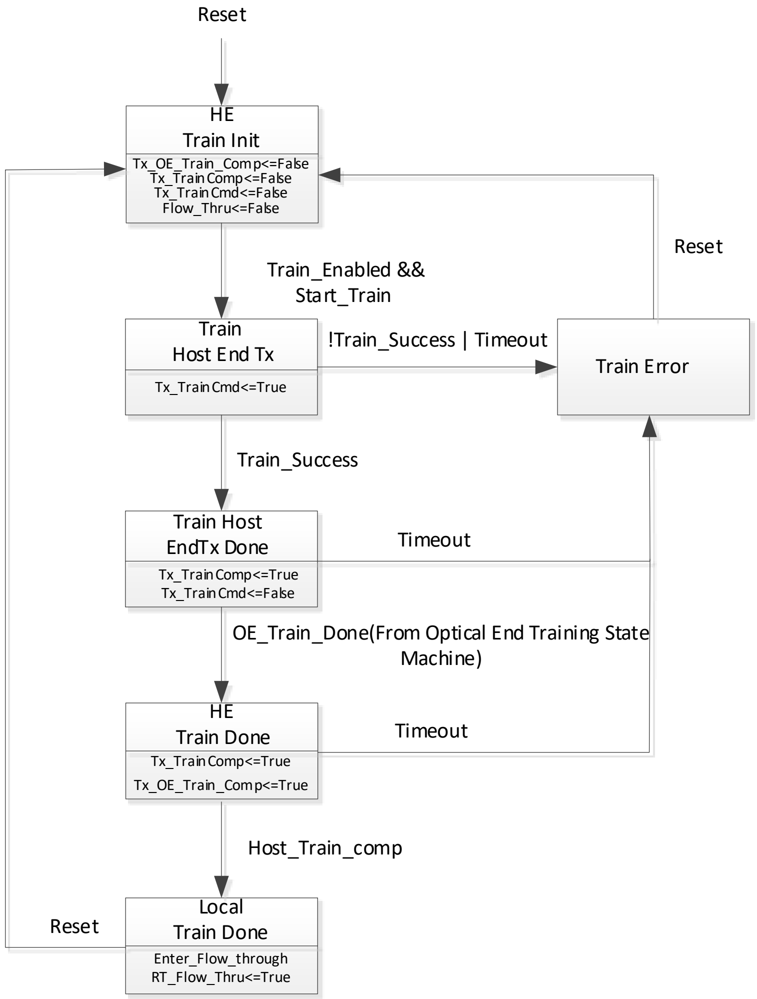


Tx_TrainComp = Tx_TTS_Status[6]
Tx_Train_cmd= Tx_TTS_Control[5]
OE_TrainComp = Rx_TTS_Status[6]

Optical End Training State Machine

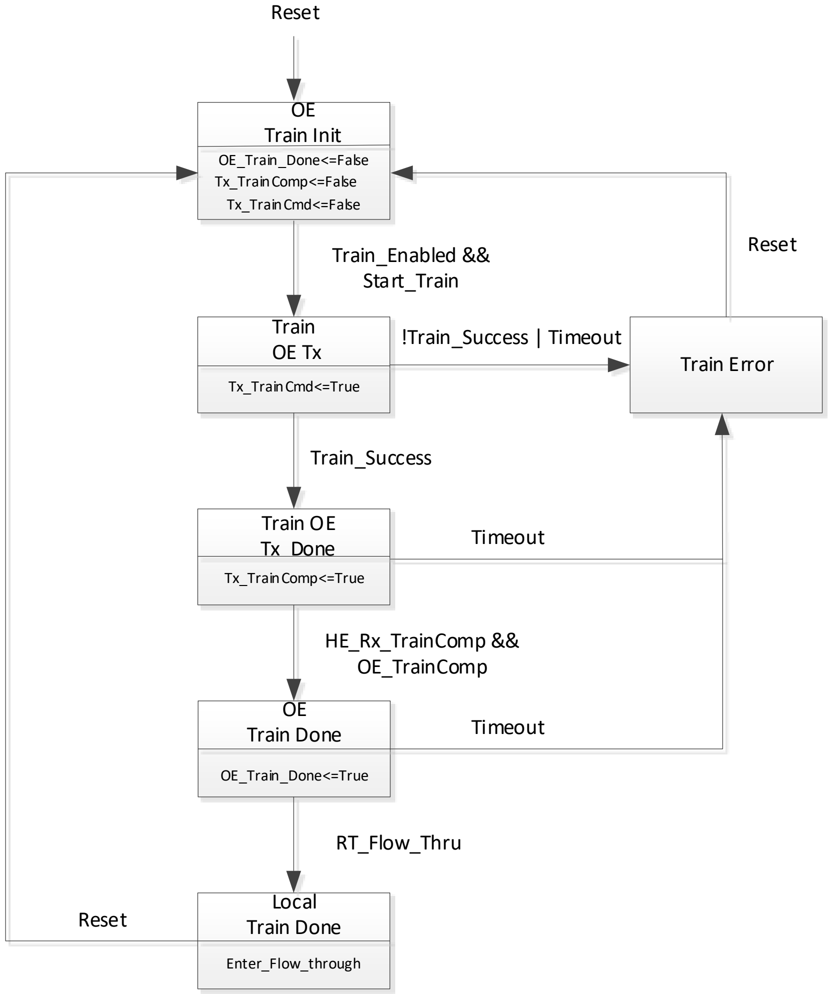
BROADCOM

Re-timer State Diagrams



Tx_TrainCmd = Tx_TTS_Control[6]
Tx_TrainComp = Tx_TTS_Status[13]
Tx_OE_Train_comp= Tx_TTS_Status[6]
Host_Train_comp = Rx_TTS_Status[15]

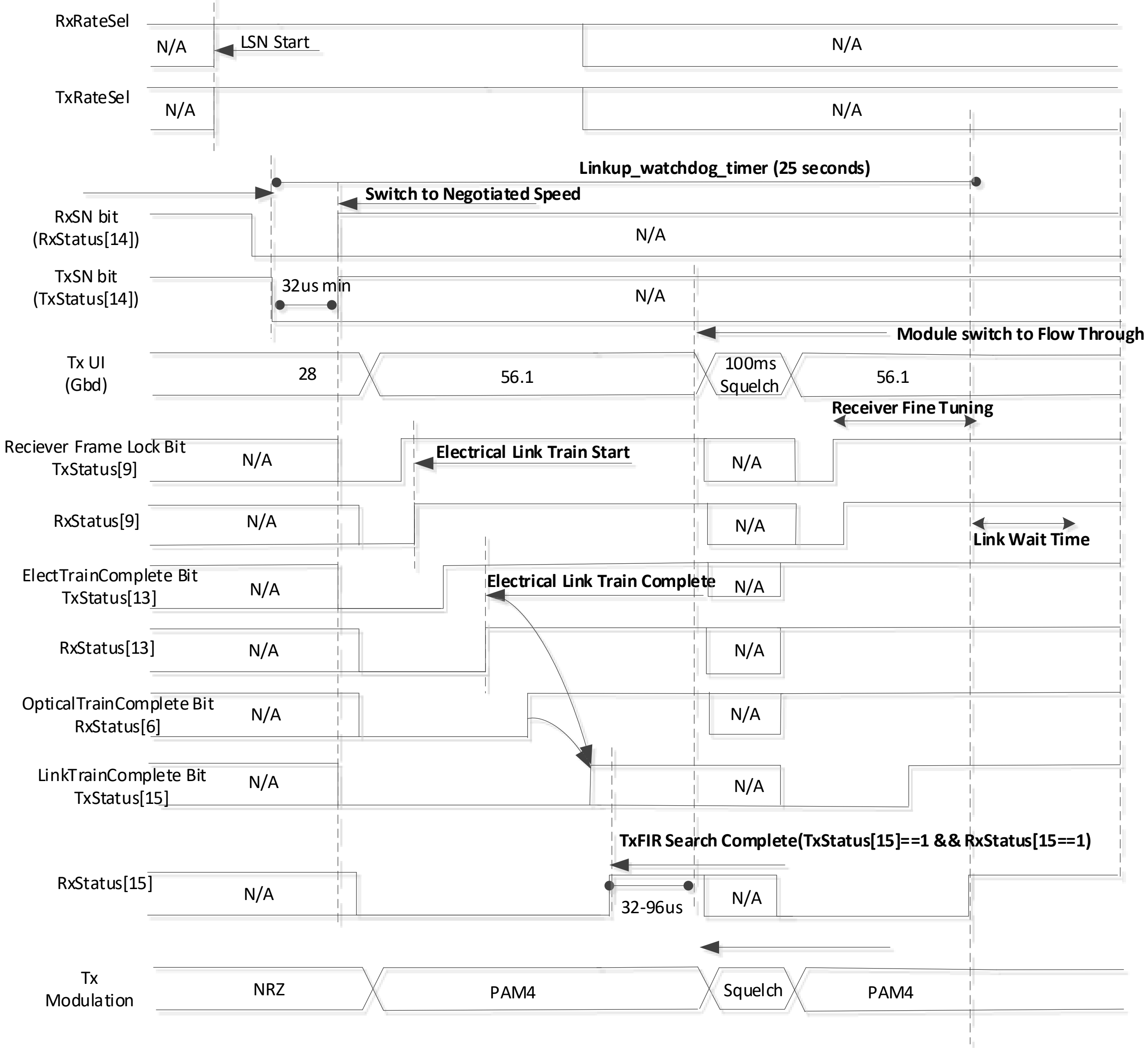
Host End Training State Machine



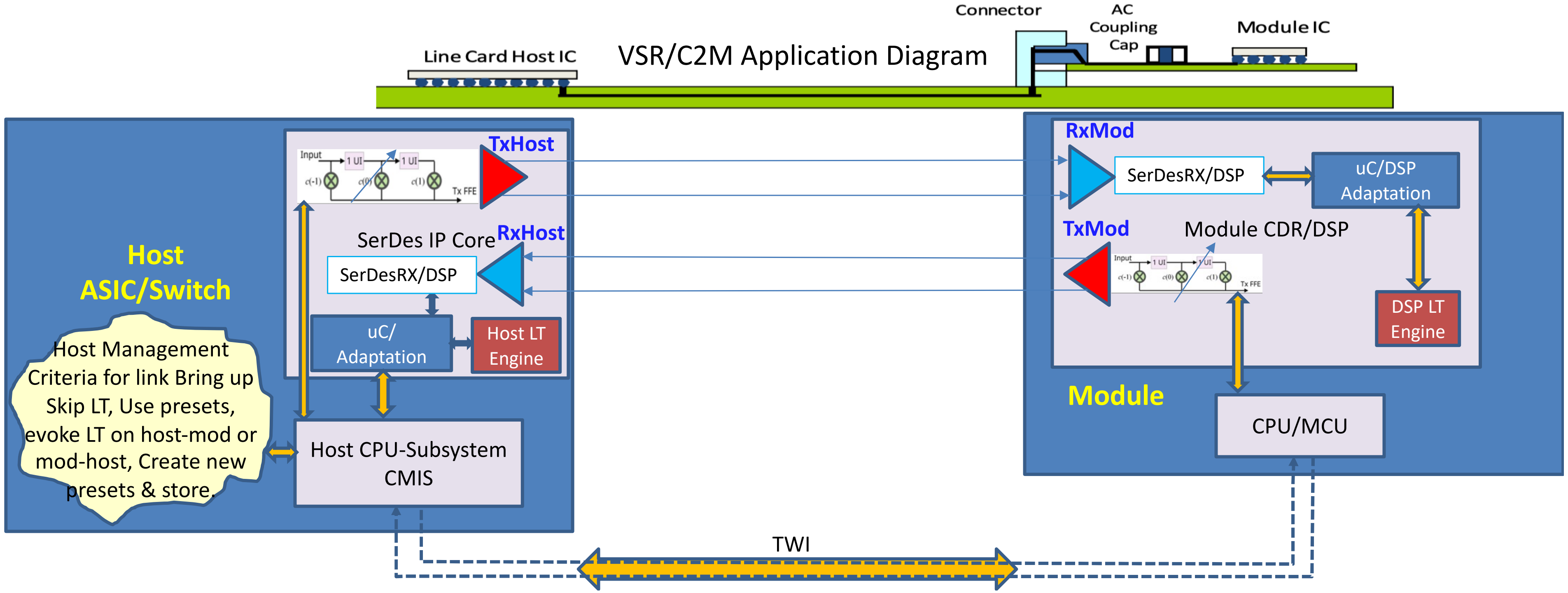
Tx_TrainComp = Tx_TTS_Status[6]
HE_Rx_TrainComp= Rx_TTS_Status[13]
Tx_Train_cmd= Tx_TTS_Control[5]
OE_TrainComp = Rx_TTS_Status[6]

Optical End Training State Machine

Host ASIC Link Bring Up Flow with In Band TxFIR Training



OB Link Training Concept aka CMIS LT(oif2022.430.01)



BROADCOM

OB vs IB Link training Comparison

	IB Link training	OB Link Training
Rx Serdes Tuning of FFE on Tx Serdes	Yes	Yes
Message Passing	In Band (High Bandwidth)	Out of Band Using Two Wire Link e.g. I2C(Lower Bandwidth)
Training of Electrical Link	Yes	Yes
Multiple Electrical Segment Links	Yes	Yes
Optical Links	Yes	No
System Management CPU Involvement	No	Yes
Incremental Training After Link Up	No	Yes(Non-Disruptive?)
Scalable to High Port Count Switches	Yes	?

Further Study Areas

- Is RX Adaptation good enough for the Host ASIC/Re-timer Serdes after switching to a recovered clock instead of an internally generated clock?
  - How to force this RX adaptation – Squelch or some other mechanism
- What is one of the entities in the link segments does not support IB LT or is not currently in a state to complete it successfully?
- Time allotment
  - Every link segment
  - Overall Link Bring Up
- New Common Frame formats for IEEE/FC that allow optimization of all link parameters

Summary

- Propose using IB Link training with TTS frames as the training mechanism for all Link Segments
  - Support for Single/Multiple Electrical Segment Links and Optical Links
  - Training State Machines already exist in several Serdes/DSP Architecture. Training frame generators/checkers/convergence algorithms don’t need to be added in most cases
  - No need to move this already existing IB function to an OB Management CPU as
    - Management CPU Bandwidth is limited especially in large port count switches
    - Management CPU cannot handle OB Optical training

BROADCOM®

Thanks