



Flash Memory Summit

Challenges of Testing PCIe Gen. 4 SSDs *and Beyond*

Justin Treon
Advantest



Flash Memory Summit

Agenda

- MPT3000 SSD Tester
- Implementing PCIe Gen. 4 Early
- Specification Issues
- Testing Issues
- Test Compliance Issues
- PCIe Gen. 5 and 6
- SAS-4 and SAS-5



Flash Memory Summit

Our PCIe Gen. 4 Product

- The MPT3000 is a multi-protocol SSD tester
 - FPGA based implementation allows for the protocol switches





Flash Memory Summit

Risking Early Development

- Hardware developed for the 0.7 physical layer of the PCIe Gen. 4 specification
- By using the FPGA we were able to test before the Gen. 4 devices became available



Testing the Link

- No devices to test the link available
... so we made our own
 - To be an early adopter you need to build your own test equipment
- Test board has many debug features to test the link

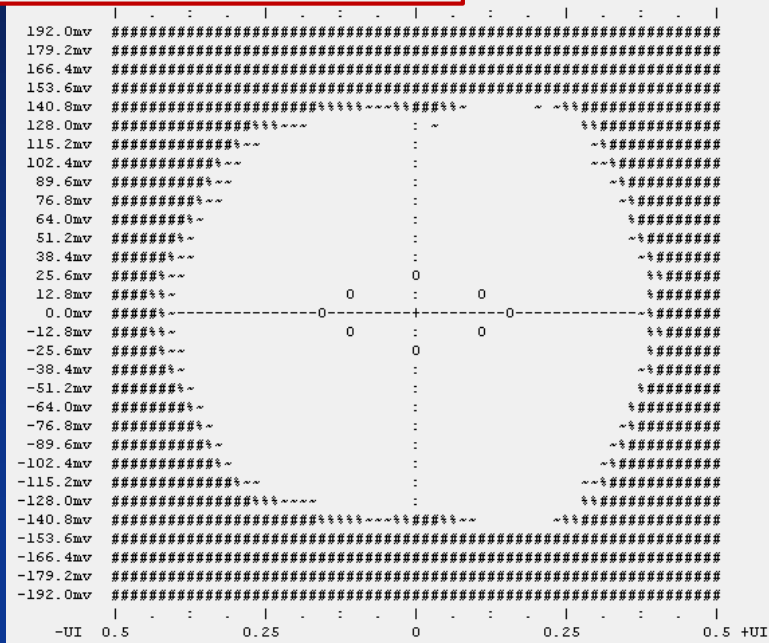




Eye Diagram

- Eye Diagram tool added to debug the link
 - Similar to Rx Margin (a.k.a. Lane Margining)
- Adding eye diagram to SSD controllers is suggested

Maximum Eye Width = 0.75 UI
 Maximum Eye Height = 256 mv





Changes to the Specification

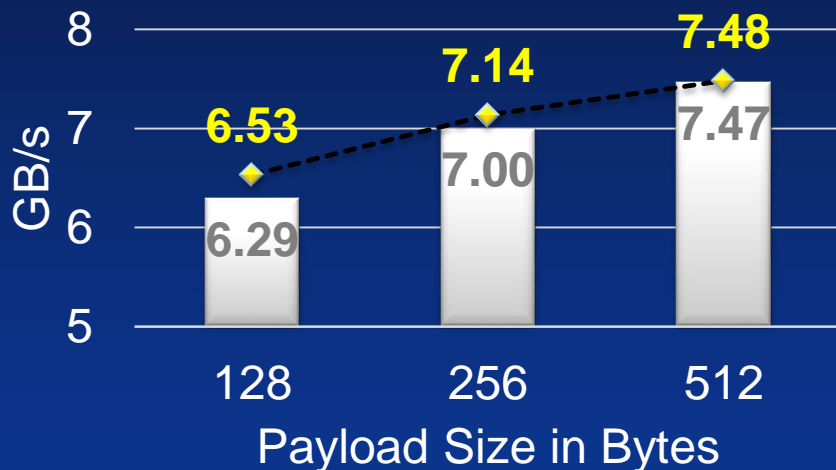
- The optional lane margining feature was made mandatory at the last moment ... *surprise*
- Causing implementation mismatch requiring rearchitecting
 - Budget time and resources for specification changes



Testing the Performance

- No SSD available to test with
... so we made our own
- DMA transfers simulated the NVMe command
 - Setting data engine speed
 - Determining performance with NVMe command overhead

Gen. 4 x4 Performance vs Payload Size



— Advantest Gen. 4 x4 with NVMe command overhead
-♦- Link Limit Gen. 4 x4



PCI SIG Compliance

- Test systems were not ready
 - False positives for pass and fail
 - Engineering resources used to correct tests
- Test specification as of last plug fest
 - Not possible to obtain PCI SIG certification



Flash Memory Summit

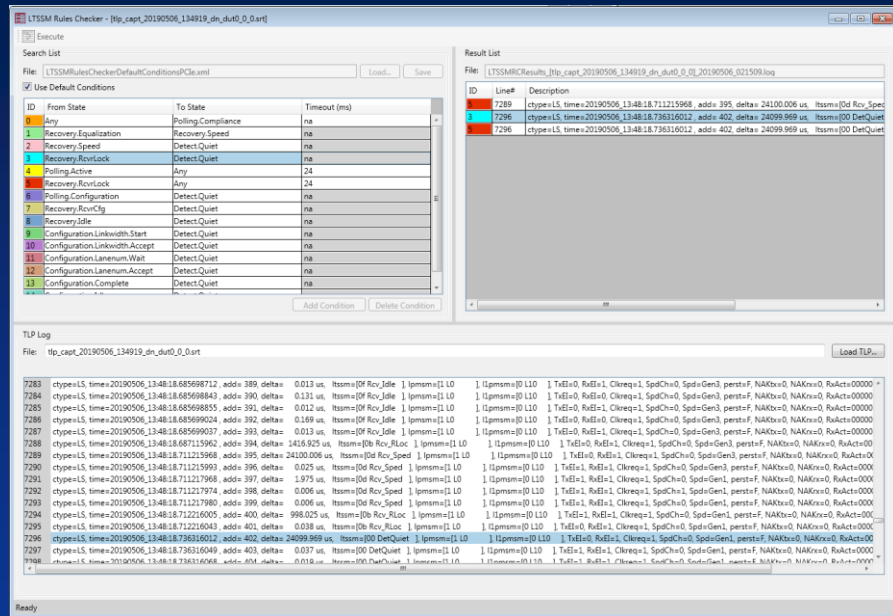
Hurry Up and Wait

- PCIe Gen. 4 market intercept pushed out from original expectations
- Extra time allowed for feature development and stability improvements



TCDT (Traffic Capture and Debug Tool)

- Traffic Capture and Debug Tool
- Suite of user friendly graphical tools for parsing and debugging captures
- LTSSM Rules Checker
- Tools allow users to parse and analyze logs in a quick and user friendly manner





TCDT (2)

- TLP Capture
 - Packet capture at the TLP layer and above
- PCIe Protocol Decoder
 - Protocol Decoder allows the user to graphically analyze the traffic

The screenshot shows two software windows. The left window is titled "PCIe Protocol Decoder - [tlp_capt_sample12capture.srt]". It features a "Filter" section with "Rules Violation" selected, and buttons for "EQ", "LS", "TLP", "<<LS", "LS>>", "Display Lock", and "Info...". Below this is a "TLP Log" section with a file path and a "Load TLP..." button. The main area contains a table with columns for "Time Stamp", "LS", "Delta Time", and "Type".

Time Stamp	LS	Delta Time	Type
2018081018:54:54.501811052	LS	2.06	Configuration.Linkwidth.Accept
2018081018:54:54.501811332	LS	0.78	Configuration.Lanenum.Wait
2018081018:54:54.501812420	LS	1.588	Configuration.Lanenum.Accept
2018081018:54:54.501812540	LS	0.62	Configuration.Complete
2018081018:54:54.501814420	LS	2.38	Configuration.Idle
2018081018:54:54.501814588	LS	0.668	Recovery.RcvrClg
2018081018:54:54.501814604	LS	0.516	Polling.Compliance
	LS		Type

The right window is titled "LTSSM Rules Checker - [tlp_capt_sample12capture.srt]". It has an "Execute" button and a "Search List" section with "File: UserDefinedConditio" and "Load..." and "Save" buttons. Below is a "Result List" table with columns for "ID", "Line#", and "Description".

ID	Line#	Description
0	51	ctype=LS, time=20180810_18:
0	52	ctype=LS, time=20180810_18:
0	53	ctype=LS, time=20180810_18:
1	75	ctype=LS, time=20180810_18:
1	8726	ctype=LS, time=20180810_18:
2	55	ctype=LS, time=20180810_18:
2	8727	ctype=LS, time=20180810_18:
3	7393	ctype=LS, time=20180810_18:
3	8716	ctype=LS, time=20180810_18:
3	27	ctype=LS, time=20180811_07:

Below the result list is a "TLP Log" section with a file path and a "Load TLP..." button. It shows log entries with details like "ctype=LS, time=20180810_18:54:54.501814420, add= 14, delta= 2.380, ltssm=[0a Cfg_Idle], TxEl=0" and "ctype=LS, time=20180810_18:54:54.501814588, add= 15, delta= 0.668, ltssm=[0e Cfg_Idle], TxEl=0".



TLP Capture Example (1)

- PCIe link framing
- Test Log

Example of link framing failure

```

Minimize voltage and disconnect!
*****
PCIe Link Status :
LinkUp L0      = true
LTSSM         = 0x10, LTSSM_L0
LinkWidth     = x4
LinkSpeed     = 8.0G
ActiveLanes   = 00001111 (0x0F)
ValidLanes    = 00001111 (0x0F)
LinkUpCount   = 1
LinkRetrainCnt = 22
*****
Sun Jan 27 23:54:47 2019
Error in connecting and
Power up PCIe levels! ERROR in setting the PCIe Power Level.

```

link retrain count 22 > 20.

- Log from TLP Capture

```

delta=473746.780 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.005 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.005 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.005 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.240 us, ltssm=[0f Rcv_Idle ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.340 us, ltssm=[10 --L0-- ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta=122430.750 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.005 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.005 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.005 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.940 us, ltssm=[0e Rcv_RCfg ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.245 us, ltssm=[0f Rcv_Idle ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.335 us, ltssm=[10 --L0-- ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 38072.200 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.005 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.005 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.005 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.970 us, ltssm=[0e Rcv_RCfg ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.240 us, ltssm=[0f Rcv_Idle ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.340 us, ltssm=[10 --L0-- ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 38072.180 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.005 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.005 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111
delta= 0.005 us, ltssm=[0b Rcv_RLOC ], lpmssm=[1 L0 ], l1pmsm=[0 L10 ], TxEI=0, RxEI=0, Clkreq=0, SpdCh=0, Gen3, perst=F, NAKtx=0, NAKrx=0, RkAct=1111

```

l0recovery=[framing]
l0recovery=[directed]

l0recovery=[framing]
l0recovery=[directed]

l0recovery=[framing]
l0recovery=[directed]

l0recovery=[framing]
l0recovery=[directed]



TLP Capture Example (2)

- No block device not ready
 - Test Log

Example of device ready failure

```

Minimize Voltage and disconnect!
*****
PCIe Link Status 1:
LinkUp LO      = true
LTSSM         = 0x10, LTSSM_LO
LinkWidth      = x4
LinkSpeed      = 8.00
ActiveLanes    = 00001111 (0x0f)
ValidLanes     = 00001111 (0x0f)
LinkUpCount    = 1
LinkRetrainCnt = 3
Sun Mar 24 01:55:04 2019 ERROR : Block Device is not present.
find_protocol_fpga_virtual_ep: ---> bus (number=14 primary=12 s
nvme_wait_ready: Error: device not ready after 20500
nvme_probe: nvme_configure_admin_queue failed, result
  
```

Block Device is not present.

- Log from TLP

```

20190324_01:54:12.277536572 [ 142]: -->DN1[ Mem32 RdReq ] 'h00000001 'h0000000f 'hca01001c ..... (3:11,0,1.12) LEN=4 CPU
20190324_01:54:12.277537080 [ 138]: <--DN1[ Cpl w/ data ] 'h4a000001 'h0e000004 'h0000001c 0x00000000 ..... (4:11,0,0.02) LEN=4 CPU

[... many lines reporting the same status deleted ...]

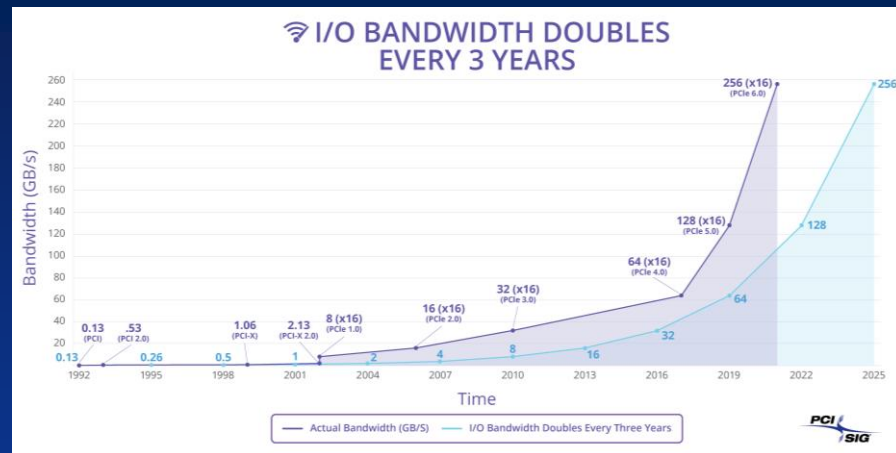
20190324_01:54:32.585139616 [ 337]: -->DN1[ Mem32 RdReq ] 'h00000001 'h0000000f 'hca01001c ..... (3:11,0,1.12) LEN=4 CPU
20190324_01:54:32.585140124 [ 333]: <--DN1[ Cpl w/ data ] 'h4a000001 'h0e000004 'h0000001c 0x00000000 ..... (4:11,0,0.02) LEN=4 CPU
20190324_01:54:32.689161116 [ 338]: -->DN1[ Mem32 RdReq ] 'h00000001 'h0000000f 'hca01001c ..... (3:11,0,1.12) LEN=4 CPU
20190324_01:54:32.689161616 [ 334]: <--DN1[ Cpl w/ data ] 'h4a000001 'h0e000004 'h0000001c 0x00000000 ..... (4:11,0,0.02) LEN=4 CPU
  
```

- NVMe controller status register CSTS.RDY reports not ready



PCIe Gen. 5 and Beyond

- PCIe Gen. 5 is almost the same as Gen. 4
 - Early adopters will need to make their own tools
- Gen. 6 specification in 2021
 - Significant changes
 - 64 GT/s
 - Pulse Amplitude Modulation
 - Forward Error Correction
- PCI SIG on three year cycle, Gen. 7 in 2024





SAS-4 and SAS-5

- There may be a SAS-4 market ... *maybe*
 - SAS-4 demand is limited to companies that do not want to update/upgrade databases
 - SAS optimized databases do not make good use of potential performance gains
- Will there even be a SAS-5?



Closing

- Making your own tools is crucial as an early implementor
- Starting early allows the product to mature, but is costly ... *and frustrating*
- Beware supply chain issues
 - High speed components may be in short supply
 - Resin for high speed PCBs will be costly and is in short supply



Flash Memory Summit

Questions?





Flash Memory Summit

Backup



TLP Capture Example (3)

Flash Memory Summit

- PCIe link rcvd_ts
- Test Log

Example of link training failure

```
=====Test: Display Link Status, Wed Mar
ActiveLanes = 00001111 (0x0F)
ValidLanes = 00001111 (0x0F)
TxElectIdle = 11110000 (0xF0)
TxDetectRx = 00001111 (0x0F)
RXElectIdle = 11110000 (0xF0)
RXPolarity = 00000000 (0x00)
LSSM = 10000 (0x10) L0
Link Width = x4 Changed
Link Speed = 5.0G Changed
Link Up = 1
Retrain = 0
Linkup_Count = 1
```

Retrain Count = 65535

Time since last Retrain: 1.1e-05 seconds

- Log from TLP Capture

```
delta= 0.000 us, tssm=10 --10-- , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=000
delta= 0.000 us, tssm=10 --10-- , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=000
delta= 0.016 us, tssm=10 --10-- , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=000
delta= 0.016 us, tssm=0e Rcv_Rcfc , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=000
delta= 1.332 us, tssm=0e Rcv_Rcfc , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=000
delta= 0.448 us, tssm=0f Rcv_Idle , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111
delta= 0.504 us, tssm=10 --10-- , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111
delta= 14.004 us, tssm=10 --10-- , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111, 10recovery=[ rcvd_ts ]
delta= 0.016 us, tssm=10 --10-- , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111
delta= 0.016 us, tssm=0b Rcv_RLoc , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111
delta= 1.340 us, tssm=0e Rcv_Rcfc , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111
delta= 0.448 us, tssm=0f Rcv_Idle , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111
delta= 0.496 us, tssm=10 --10-- , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111
delta= 14.012 us, tssm=10 --10-- , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111, 10recovery=[ rcvd_ts ]
delta= 0.016 us, tssm=10 --10-- , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111
delta= 0.016 us, tssm=0b Rcv_RLoc , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111
delta= 1.332 us, tssm=0e Rcv_Rcfc , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111
delta= 0.448 us, tssm=0f Rcv_Idle , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111
delta= 0.512 us, tssm=10 --10-- , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111
delta= 13.996 us, tssm=10 --10-- , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111, 10recovery=[ rcvd_ts ]
delta= 0.016 us, tssm=10 --10-- , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111
delta= 0.016 us, tssm=0b Rcv_RLoc , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111
delta= 1.332 us, tssm=0e Rcv_Rcfc , lpsm=1 L0 , lpsm=10 L10 , TXE=0, RxE=0, Clkreq=0, SpdCh=0, Spd=Gen2, perat=F, NAKtx=0, NAKrx=0, RxAAct=00001111
```

10recovery=[rcvd_ts]
10recovery=[rcvd_ts]



PCIe Gen. 4 Changes

1. **Speed change to 16GT/s**
2. **Equalization updates for 4.0 (8 GT/s to 16 GT/s)**
3. TSx OS changes
4. 16 GT/s EIEOS
5. SKP OS changes (CTRL SKP)
6. Polling.Compliance update
7. **10-Bit Tag**
 - Allowing for greater token count
8. Data Link Feature Exchange
9. Flow Control Scaling
10. **Rx Margining (a.k.a. Lane Margining)**
 - **Host cabling feature**
 - Pushed through as a requirement by Intel
11. Retimer
12. **Configuration space register updates**