



PCIe® 3.0 PHY Logical

Joe Cowan

Computer Systems Architect

Hewlett-Packard Company



Agenda

- Background
- New Encoding Scheme
- Transmitter Equalization and Training
- Testability Features
- Summary

Problem Statement

- PCIe® 3.0 data rate decision: 8 GT/s
 - ✓ High Volume Manufacturing channel for client/ servers
 - Same channels and length for backwards compatibility assuming worst-case
 - ✓ Low power and ease of design

- Requirement: **Double Bandwidth** from PCIe 2.0
 - ✓ PCIe 1.0 a data rate: 2.5 GT/s; doubled in PCIe 2.0 at 5GT/s
 - ✓ 8GT/s Data rate gives us a 60% boost in bandwidth
 - ✓ Rest will come from **Encoding**
 - Replace 8b/10b encoding with a scrambling-only mechanism at 8GT/s
 - ✓ Double B/W: Encoding efficiency improvement of 1.25X data rate improvement of 1.6 yields 2X improvement in bandwidth

- **Challenge:** 8b/10b encoded the 2^8 data patterns and 12 K-codes
 - ✓ K-codes used for PHY layer framing as well as Symbol boundary determination

- **Another Challenge** is related to operating the channel at 8GT/s
 - ✓ Need a Transmitter equalization mechanism

Encoding Considerations

- Basic Fault Model: Triple bit flip detection
- Eventual recovery with bit slip/add
- Framing Errors cause transition to Recovery
- Need to handle killer packets
 - ✓ Send a different bit stream on retry of a packet
- Bandwidth Inefficiency must be less than 2%
- Time Overhead through Recovery as well as L0s/L1 exit must be minimal
- Minimal changes beyond PHY layer

Agenda

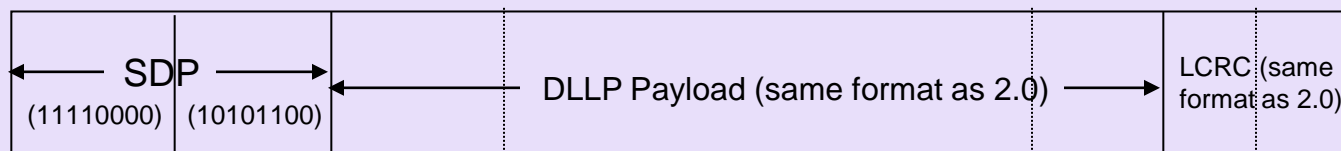
- Background
- New Encoding Scheme
- Transmitter Equalization and Training
- Testability Features
- Summary

128b/130b Encoding Scheme

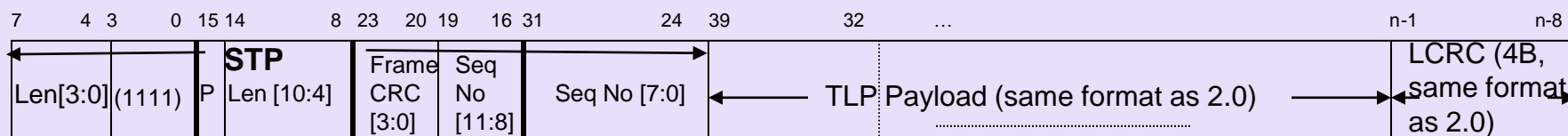
- Lane Level Encoding: 2 bit Sync header followed by 128 bit payload
 - ✓ Two types of Blocks:
 - Data Blocks: 10b Sync Header. Used for TLP, DLLP, IDL.
 - Ordered Set Blocks: 01b Sync Header. One OS per Block.
- Scrambling provides edge density
 - ✓ Sync header not scrambled
 - ✓ Payload in Data Blocks always scrambled
 - ✓ Ordered Set payload not scrambled except last 15 Symbols of TS1/ TS2
 - ✓ Degree 23 polynomial ($G(X) = X^{23} + X^{21} + X^{16} + X^8 + X^5 + X^2 + 1$)
 - Different taps for 8 adjacent lanes (or different seeds for same tap)
 - Minimizes cross talk as well as baseline wander
 - ✓ Electrical Idle Exit Ordered Set resets LFSR (Recovery/ Config)
- Electrical Idle Exit Ordered Set used for Block Alignment
 - ✓ Substitutes COM used for Symbol lock in 8b/10b
- Framing token defines the length of packets in Data Blocks
 - ✓ Multiple packets can exist in a Data Block and a packet can span across multiple Data Blocks

Data Block: Framing Tokens

- First Symbol of token indicates packet type:
 - ✓ 00000000: IDL; xxxx1111: STP; 11110000: SDP
 - ✓ Hamming distance 4 guarantees triple bit flip detect
- Token length is variable and indicates location of next token
- IDL Token is 1 Symbol. No payload (PAD merged with IDL).
- SDP Token is 2 Symbols
 - ✓ DLLP is 8 Symbols with no explicit End
- STP Token is 4 Symbols
 - ✓ Variable TLP Length; length field in STP protected by its own CRC/ parity



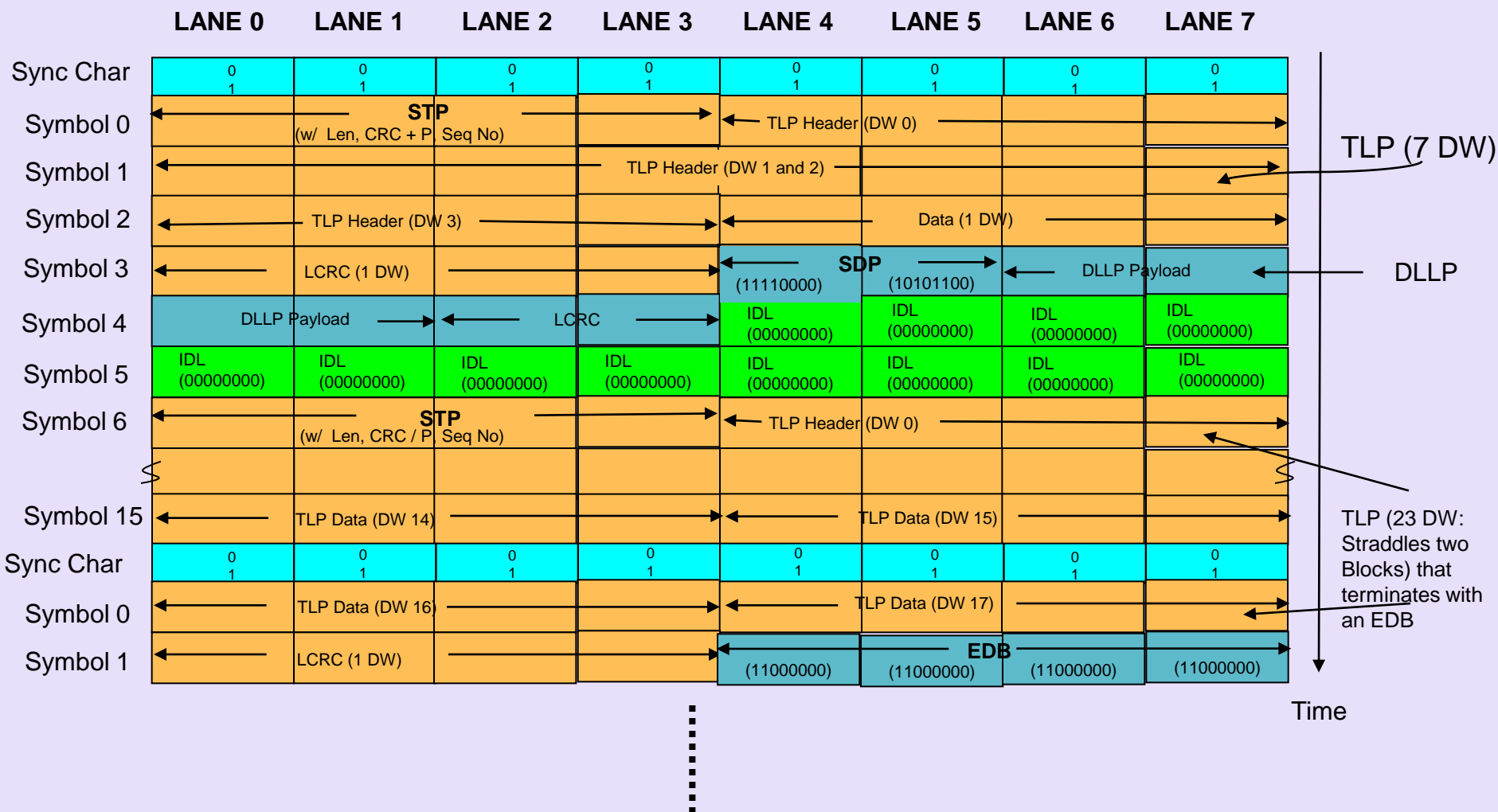
(DLLP Layout)



[Len[10:0]: length of the TLP in DWs, Frame CRC[4:0]: Check Bits covering Length[0:10], P: Frame Parity, No END]

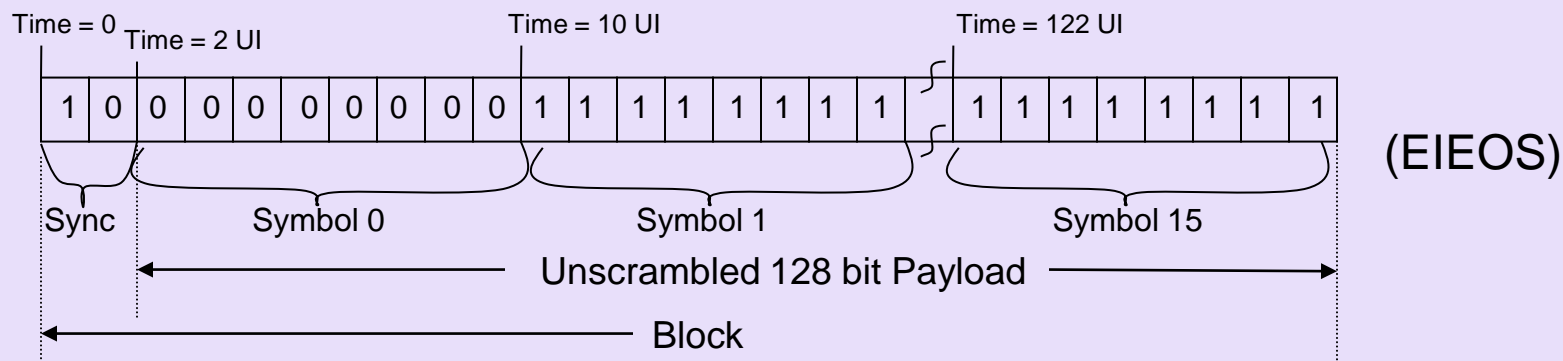
(TLP Layout)

Ex: TLP/ DLLP/ IDLs in x8



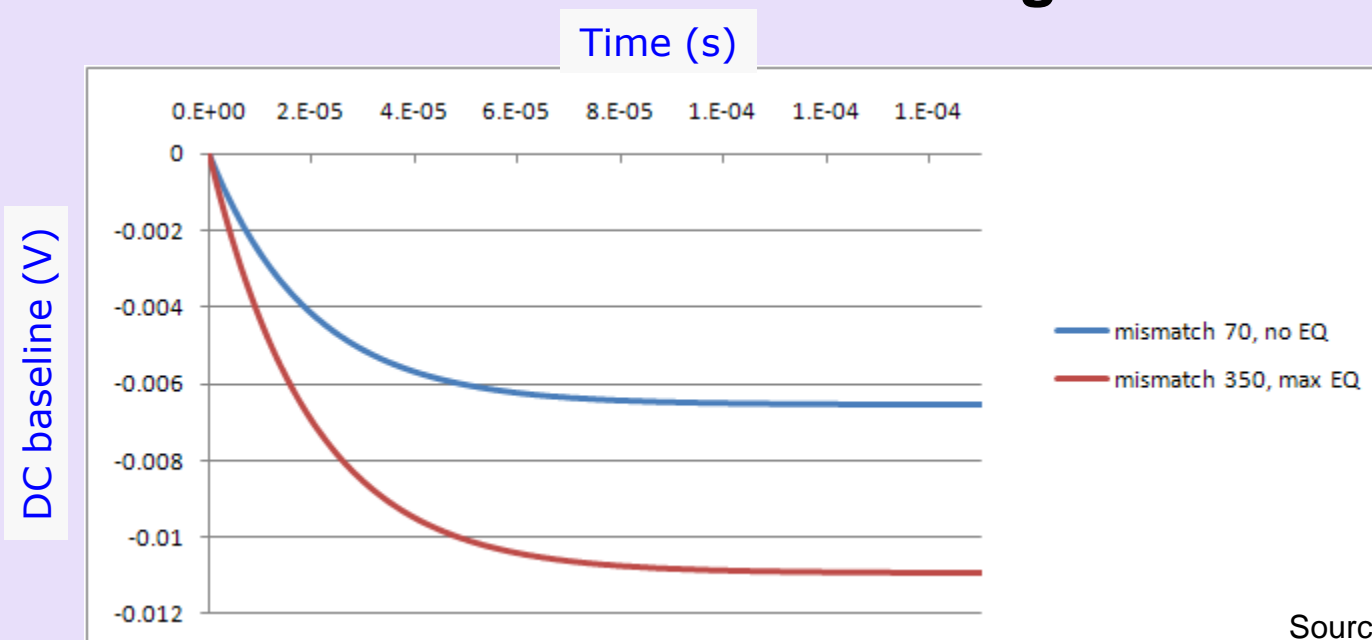
Ordered Sets

- First Symbol indicates the OS type
 - ✓ Not scrambled; DC balanced; at a Hamming distance of 4 from each other
 - ✓ TS1: 1E, TS2:2D, EIEOS: 00, SKP: AA, EIOS: 66, FTS: 55, SKP_END/ SDS: E1
- None of the Ordered Sets except TS1/TS2 are scrambled
- TS1/TS2:
 - ✓ Symbol 0 not scrambled; Symbols 1-13 are Scrambled
 - ✓ Symbols 14 and 15 may be used for DC balance
 - Will use unscrambled DC balancing patterns if the accumulated DC imbalance up to Symbol 11 is greater than 31 for Symbol 14 and 15 for Symbol 15; else will use scrambled value
- EIEOS: Low Frequency; Block Alignment, reset scrambler – Recovery/ Config
- EDS Token is sent in the last DW of last data block prior to switching to OS
 - ✓ Ensures a 2-bit error with the sync header does not alias to a TLP/DLLP
- SDS (Start Data Stream) OS sent prior to the first data block after a stream of OS
 - ✓ Ensures a 2-bit error with the sync header does not alias to a TLP/DLLP



DC Imbalance in TS1/TS2

- **Substantial DC imbalance accumulated in Recovery**
- **350 bits imbalance in 4290 UI**
 - ✓ DC baseline wander of 11mV with max TX EQ (9.5dB boost)
 - ✓ Will cause bad set-up during Equalization
- **Need DC correction on exceeding certain threshold**



Source: Intel Corporation

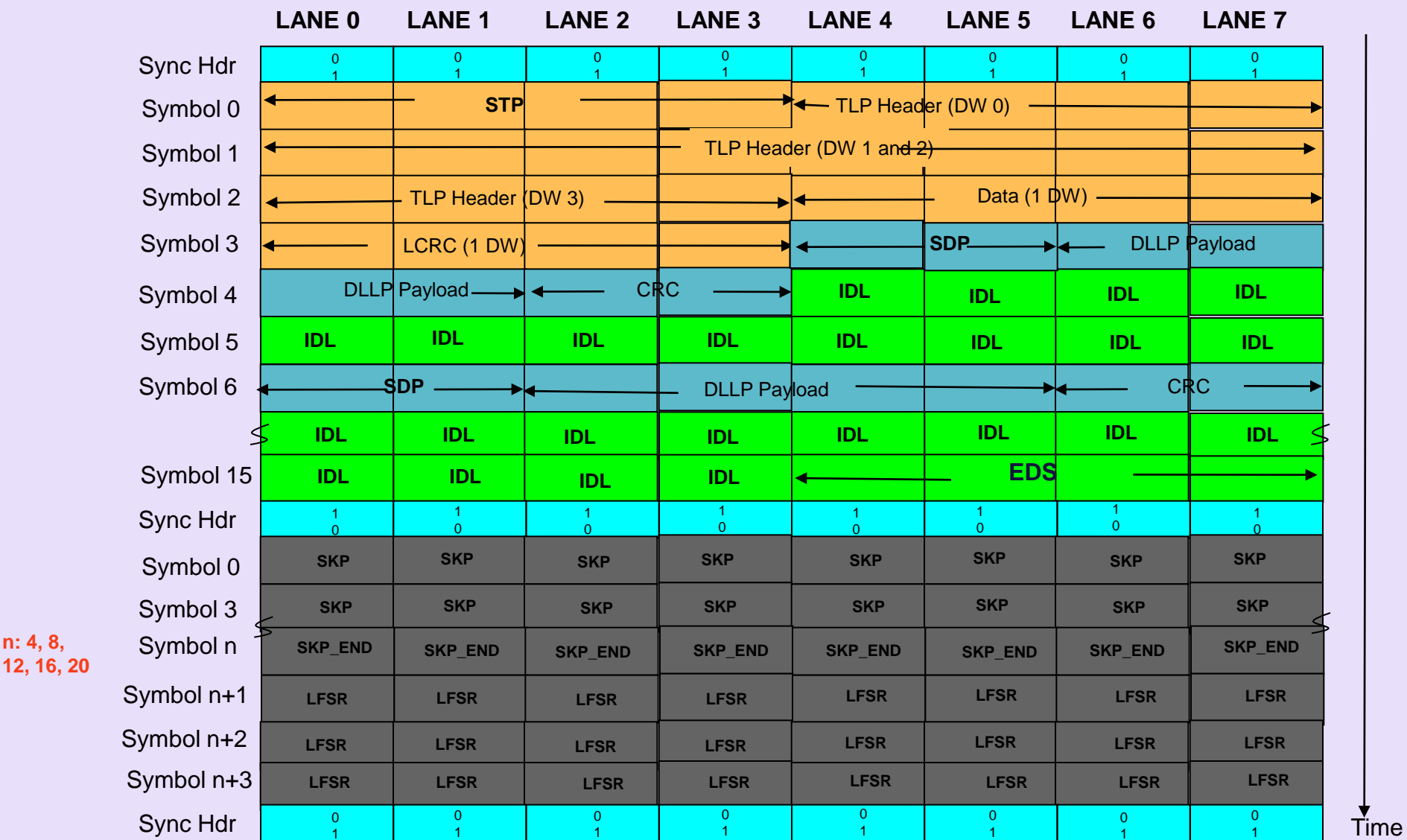
TS1 Ordered Set

| Symbol No | Value | Scrambled? | Scrambler Advances? | Description |
|-----------|-------------------|------------|---------------------|---|
| Sync Hdr | 01b | No | 01b | |
| 0 | 1Eh | No | Yes | TS1 OS Identifier |
| 1 | 0-31, PAD | Yes | Yes | Link No |
| 2 | 0-31, PAD | Yes | Yes | Lane No |
| 3 | 0-255 | Yes | Yes | N_FTS |
| 4 | - | Yes | Yes | Data Rate Identifier |
| 5 | - | Yes | Yes | Training Control |
| 6 | - | Yes | Yes | {Use Preset[7], Tx Preset[6:3], Reset EIEOS Interval[2], Equalization Control[1:0]} |
| 7 | - | Yes | Yes | {Reserved[7:6], FS or Precursor Coefficient[5:0]} |
| 8 | - | Yes | Yes | {Reserved[7:6], LF or Cursor Coefficient[5:0]} |
| 9 | - | Yes | Yes | {Parity[7], Reject Coefficient Values[6], Post Cursor Coefficients[5:0]} |
| 10-13 | 4Ah | Yes | Yes | TS1 Identifier |
| 14-15 | 4Ah or DC balance | Yes / No | Yes | DC balance: Symbol 14: DFh or 20h, Symbol 15: F7h or 08h |

SKP Ordered Set

- Usage: Logic Analyzer, Clock Compensation
- Can be of variable length at Receiver
 - ✓ Payload can be 8, 12, 16, 20, or 24 Symbols
 - ✓ Add or delete 4 SKP symbols in each receiver
 - ✓ Block boundary needs to be adjusted at end of the Block
- Transmitter always sends 16 Symbols
- Scrambler not advanced
- Symbols 0 through $4N-1$ is AAh
- Symbol $4N$ is E1 (denotes SKP_END)
- Symbol $4N+1$ through $4N+3$ carry different information depending on LTSSM state:
 - ✓ Polling.Compliance: {AA, Err_Cnt[7:0], ~Err_Cnt[7:0]}
 - ✓ Else:
 - LFSR[22:0] value for helping trace tools achieve block alignment in the midst of data stream
 - If Data Stream also carries a Parity bit to help isolate a faulty Lane (else carries ~LFSR[22])

Use of EDS Token



L0s Entry and Exit

- Entry to L0s: Data Block with EDS -> EIOS
Block -> EI
- Exit from L0s: EIEOS -> (N) FTS -> EIEOS ->
SDS -> Data Block
 - ✓ Block lock either with the last EIEOS (optionally on FTS)
 - ✓ Lane to lane deskew on SDS or EIEOS or SKP OS (on extended sync)
 - ✓ On Extended Sync, the (N) FTS number will have SKP OS inserted periodically
 - ✓ EIEOS after every 32 FTS
 - ✓ Receiver exit condition on each Lane is SDS followed by a Data Block. Must go to Recovery if this is not met.

Error Detection and Recovery

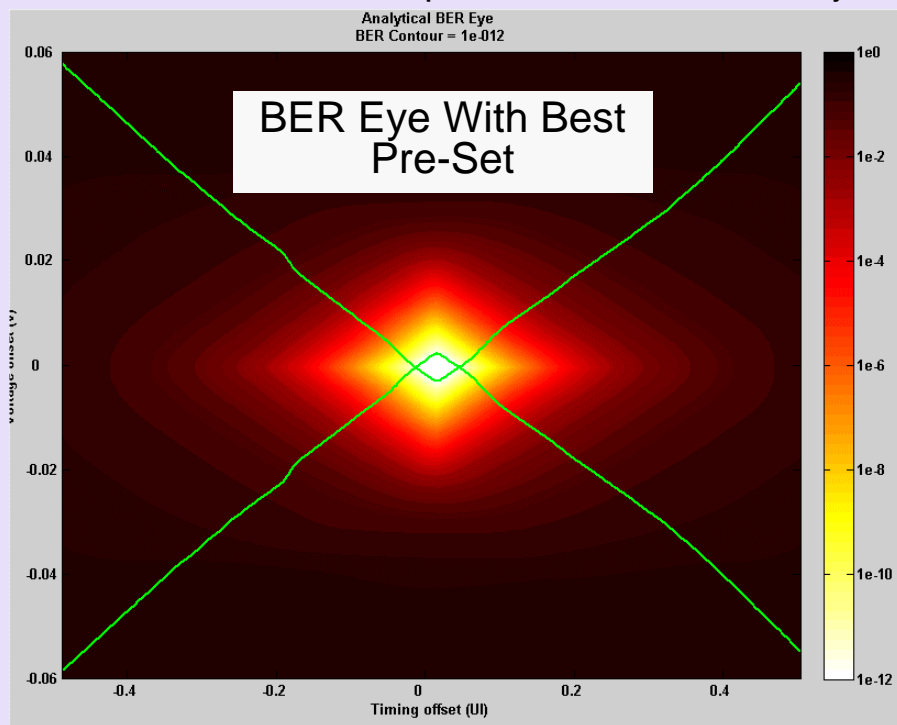
- Framing error detected by the physical layer based on some rules. Some examples:
 - ✓ Token does not match defined types (first byte not SDP, STP, IDL, ...)
 - ✓ Sync header is 00b or 11b
 - ✓ Same sync character not present in all lanes after deskew
 - ✓ CRC / parity error in the length field of an STP token
 - ✓ No EDS in the Data Block prior to the first OS
- Any framing error requires directing LTSSM to Recovery
 - ✓ Stop processing any received TLP/ DLLP
 - ✓ Block lock acquired with EIEOS
 - ✓ Scrambler reset with each EIEOS
- Error Detection Guarantees
 - ✓ Triple bit flip detection within each TLP/ DLLP/ IDL/ OS

Agenda

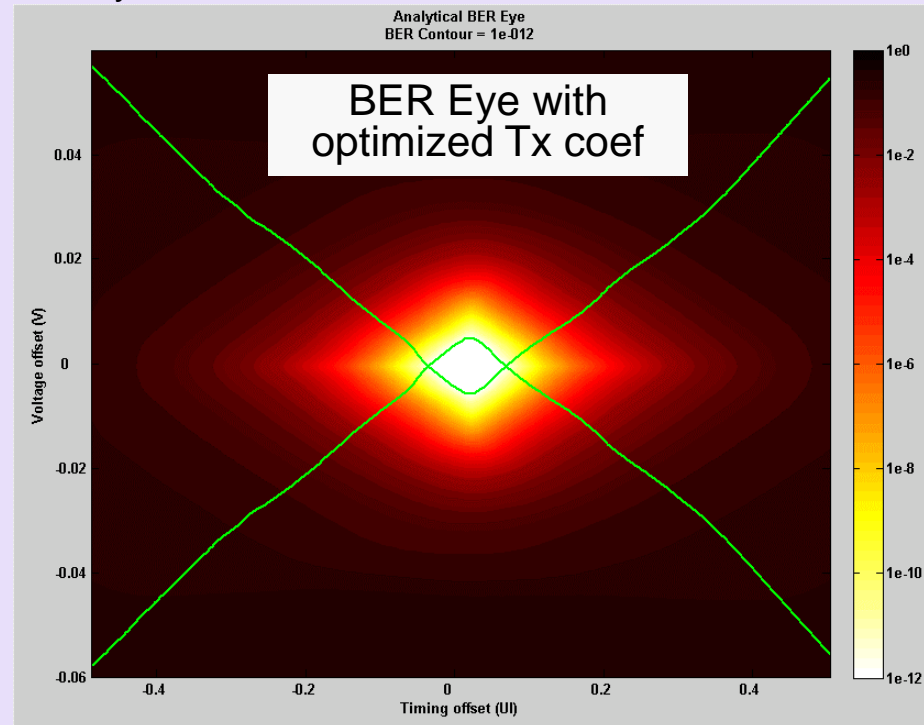
- Background
- New Encoding Scheme
- Transmitter Equalization and Training
- Testability Features
- Summary

Transmitter Equalization

- 2.5GT/s and 5GT/s: Fixed de-emphasis for Link
- 8 GT/s: Analysis demonstrates need for per Tx-Rx EQ
 - ✓ Variations in receiver design, channel, PVT
 - ✓ Adjust each Tx by its Rx individually
 - ✓ Start with a preset value and then adjust dynamically



Results from an 18" 2C channel



Source: Intel Corporation

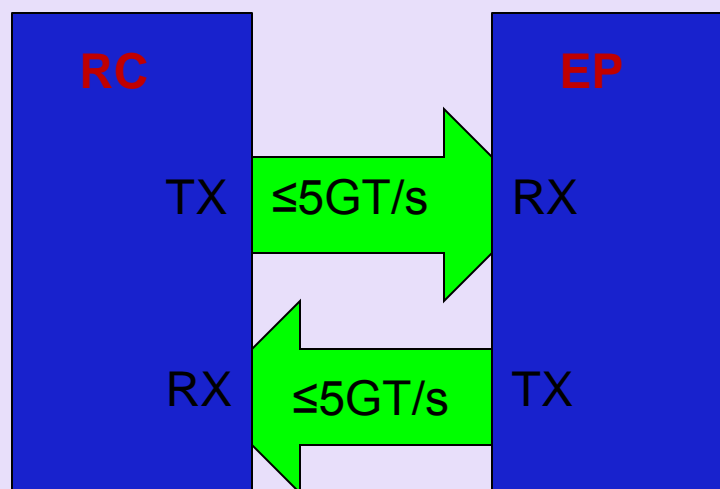
Co-efficient based Tx EQ provides better margin

Equalization – Phase 0

Phase 0 is used to exchange spec determined TxEq presets while operating at 2.5GT/s or 5GT/s speed before jumping to 8GT/s

| Preset |
|--------|
| 0 |
| 1 |
| 2 |
| 3 |
| 4 |
| 5 |
| 6 |
| 7 |
| 8 |
| 9 |
| 10 |

- Preset is exchanged in link training ordered set
 - ✓ Preset is sent from Upstream to Downstream
 - ✓ Downstream component implicitly accepts for its Tx
 - ✓ Upstream may use a different preset in its Tx

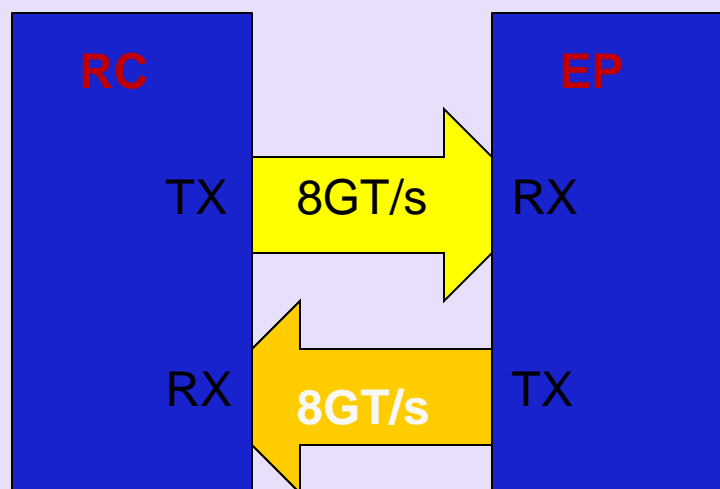


Source: Intel Corporation

Equalization – Phase 1

Link jumps to 8GT/s in phase 1. Both RC and EP use TxEq values that comply w/ the presets (defined by the spec).

Expectation is that link will operate “good” enough to allow progression at 8GT/s ($BER \leq 10^{-4}$); else link will go to a lower Data Rate



Source: Intel Corporation

Back Channel – Phase 2/ 3

Ph 2: Intended for EP Receiver to achieve $BER \leq 10^{-12}$. Starts at the preset.
Coefficients/ presets are exchanged in sub-loops until this is accomplished
May decide not to make any new requests

- Receiver full swing (FS) defines granularity of coeff
 - ✓ Table at bottom-right is for illustrative purposes
 - ✓ X-axis is pre-cursor, y-axis post-cursor, diagonal defines the boostline
 - ✓ Each tile represents a coeff (e.g. p7=4/6, p8=5/5, etc)
 - ✓ Numbers in tiles represent presets; black tiles are illegal coeff space

Example: start from preset 7 (coef=4/6)

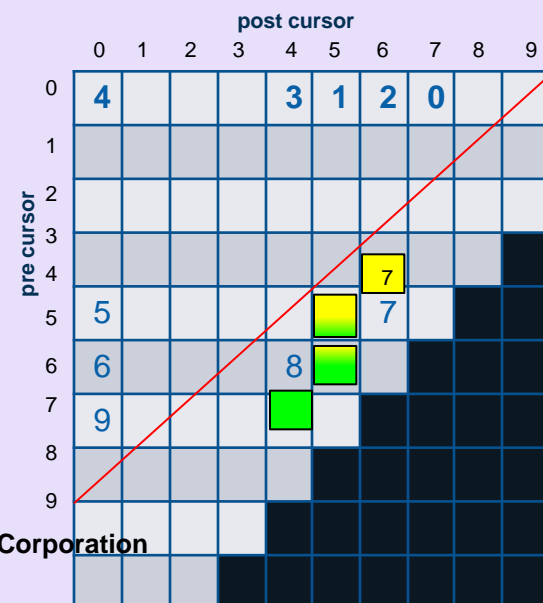
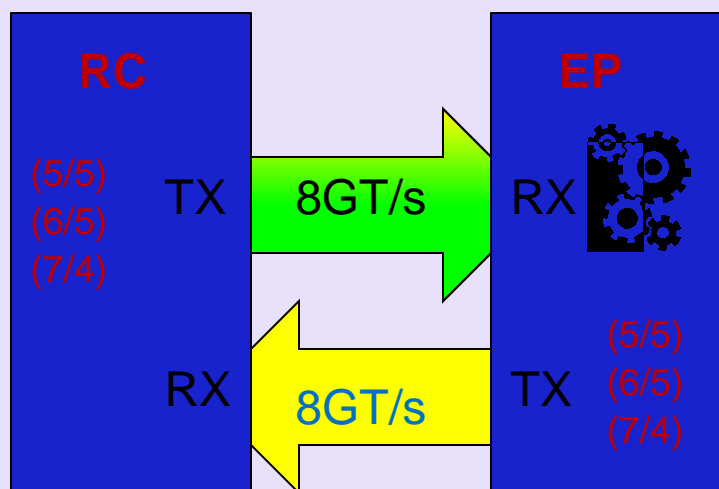
1st sub-loop

- EP Rx eval reveals need for less post, more pre
- EP sends (5/5) to RC
- RC applies (5/5) to TX
- RC echo's (5/5) to EP

2nd sub-loop

- EP Rx eval needs more pre, post ok
- d. repeat with (6/5)

3rd sub-loop finds good result with (7/4) so moves to phase 3



Source: Intel Corporation

Phase 3 is same as phase 2 in opposite direction

Upstream component may skip Ph 2/ 3 if presets are good enough for the Link

Equalization Procedure

- Expected to be done once immediately following Link Training to L0 (Autonomous)
 - ✓ No DLLP/TLP exchange till equalization completes
 - Ensures no TLP can timeout as this can take more than 50 ms
 - ✓ Software should look at DL_Active prior to accessing downstream component
- Equalization can be initiated by Software by accessing CSRs in the Upstream component
 - ✓ Must ensure that no side-effects (e.g., ensure no timeout)
- A device may choose to withhold advertising 8GT/s (and hence equalization) if its associated software can guarantee no side effects by doing equalization when it advertises 8GT/s Data Rate
- Error during equalization or later
 - ✓ Not expected to use this other than error condition
 - ✓ Upstream component can redo EQ
 - ✓ Downstream component must report in its register and request
 - Upstream has two choices (i) redo equalization (ii) log and report

Agenda

- Background
- New Encoding Scheme
- Transmitter Equalization and Training
- **Testability Features**
- Summary

Compliance Pattern

- Same entry/ exit as 2.0
 - ✓ CSR based, TS Ordered Set, as well as CLB/CBB
- Preset used during compliance
 - ✓ CSR based: Link Control 2 Register[15:12]
 - SKP Ordered Set CSR has no impact in 8GT/s
 - ✓ TS Ordered set: Symbol 6 of received TS1 OS
 - ✓ CLB/ CBB: Cycle through 2.5 GT/s, 5 GT/s at -3.5 dB de-emphasis, 5 GT/s at -6 dB de-emphasis, and then the 11 presets at 8 GT/s
 - Moves by detecting exit from electrical idle
- Pattern: 36 Blocks
 - ✓ Sync Hdr: 01b. Payload: 64 1's followed by 64 0's
 - ✓ 2 blocks with Sync Hdr: 01b, Payload: different values in different Lanes to achieve DC balance across 36 blocks
 - ✓ EIEOS
 - ✓ 32 Data Blocks: Payload scrambled 16 IDL Symbols
 - Different in 8 adjacent Lanes of a Link
 - Scrambler reset at start of 32 blocks due to EIEOS transmission

Modified Compliance Pattern

- 65792 Blocks:
 - ✓ One EIEOS
 - Resets scrambler – different taps/ seed values in 8 Lanes
 - ✓ 256 Data Blocks
 - IDL scrambled in payload
 - ✓ 255 Sets of
 - One SKP Ordered Set
 - 256 Data Blocks
 - IDL scrambled in payload
- Presets chosen using the same mechanism as in compliance pattern (CSR or TS1 based)

Loopback

- Must use the 128/130 code throughout
 - ✓ Either 01b or 10b sync header must be used
 - OS with first byte same as SKP OS or EIEOS must be used for that OS only
 - ✓ Periodically send SKP OS for Slave to adjust for ppm differences at same frequency as it normally goes out (01b Sync header)
- Slave must not readjust the Block alignment once it switches over looping back except on the SKP OS boundary
 - ✓ Rationale: Valid data can alias to an EIEOS in an unaligned location
- Loopback master expected to send the SKP OS periodically
 - ✓ No need to send SDS as slave is just looping back
 - ✓ Master must expect variable number of SKPs in the SKP OS

Loopback (Continued)

- Slave only adjusts the SKP (add, delete, or keep intact) as per the normal SKP adjustment rules – does not generate any SKP OS once it loops back
- Same LTSSM transitions for LB as in PCIe 1.x / PCIe 2.x with following changes
 - ✓ Need to send EIEOS on LB Entry – once every 32 blocks
 - ✓ Slave terminates its OS whenever it decides to loopback
 - ✓ Master constantly adjusts its block alignment while in LB.Entry. That will help the LB Master to align to the new boundary that the slave will move to when it starts looping back the contents.

Agenda

- Background
- New Encoding Scheme
- Transmitter Equalization and Training
- Testability Features
- Summary

Summary

- Encoding scheme developed and stable
 - ✓ Offers advantage of 25% bandwidth for 8GT/s (and above) data rate over 8b/10b encoding
- Equalization mechanism defined
- Robust testability features
- PCIe 3.0 Base Spec released
- Plan for products

Thank you for attending the
PCI-SIG Developers Conference 2012

For more information please go to
www.pcisig.com