Gotchas in PCIe® Implementation Beyond 5GT/s

Suparna Behera

LSI Corp

Why Gotchas !

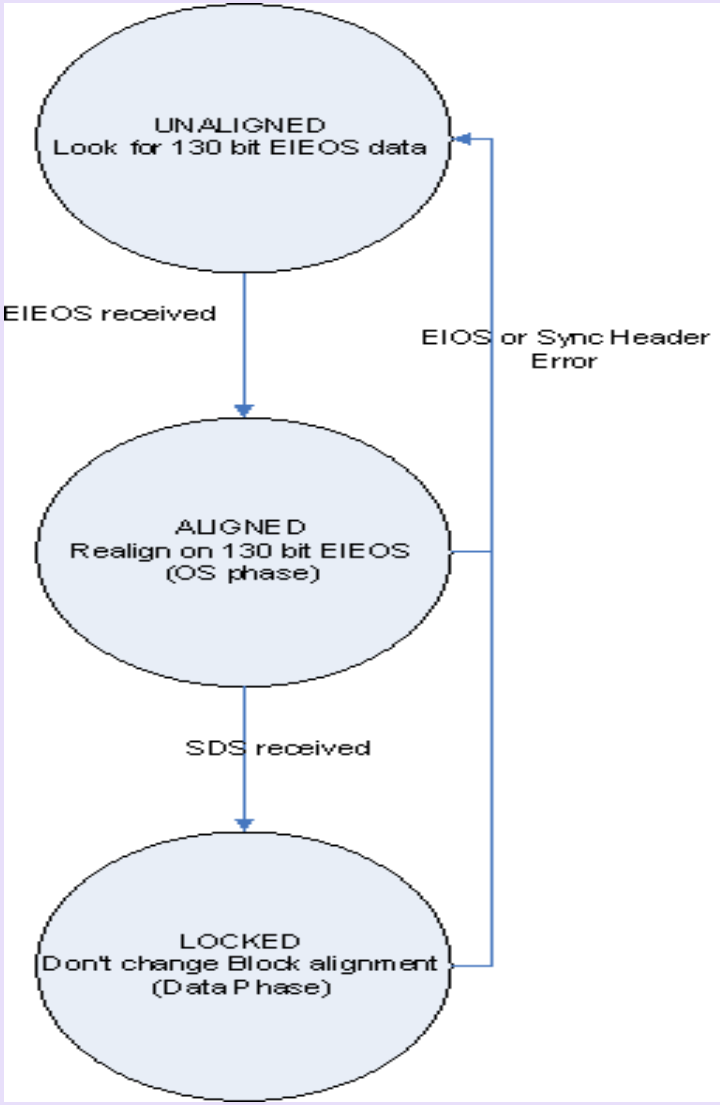
- The difference between PCIe 3.0 vs PCIe 2.0 is much greater than PCIe 2.0 vs PCIe 1.1 for Phy and Link layers.
- The assumptions used with 8b/10b encoding at 5.0 and 2.5 GT/s are no longer true while using 128/130 bit encoding at 8 GT/s.
- New Equalization substates have been added for 8 GT/s

Note: reference to PCIe 3.0 is with respect to rev 0.7 of PCIe 3.0 specs.

Block Alignment

- Block alignment using EIEOS at 8 GT/s is analogous to symbol alignment using K28.5 at 5.0 or 2.5 GT/s.
  - ✓ Data phase versus ordered set phase knowledge needed during block alignment for 128/130 bit encoding.
    - The following Ordered Set recognization is needed by the 8 GT/s Block alignment logic.
      - EIEOS (for block alignment itself).
      - SDS (to recognize that data phase is going to be entered, and entry into locked phase)
      - EIOS (to recognize exit of data phase, and entry into unaligned phase).
        - For case of EIOS recognition, if multiple EIOSs are being received at a time, instead of stopping after first 4 bytes of first EIOS, block alignment logic has to consider all incoming EIOS..

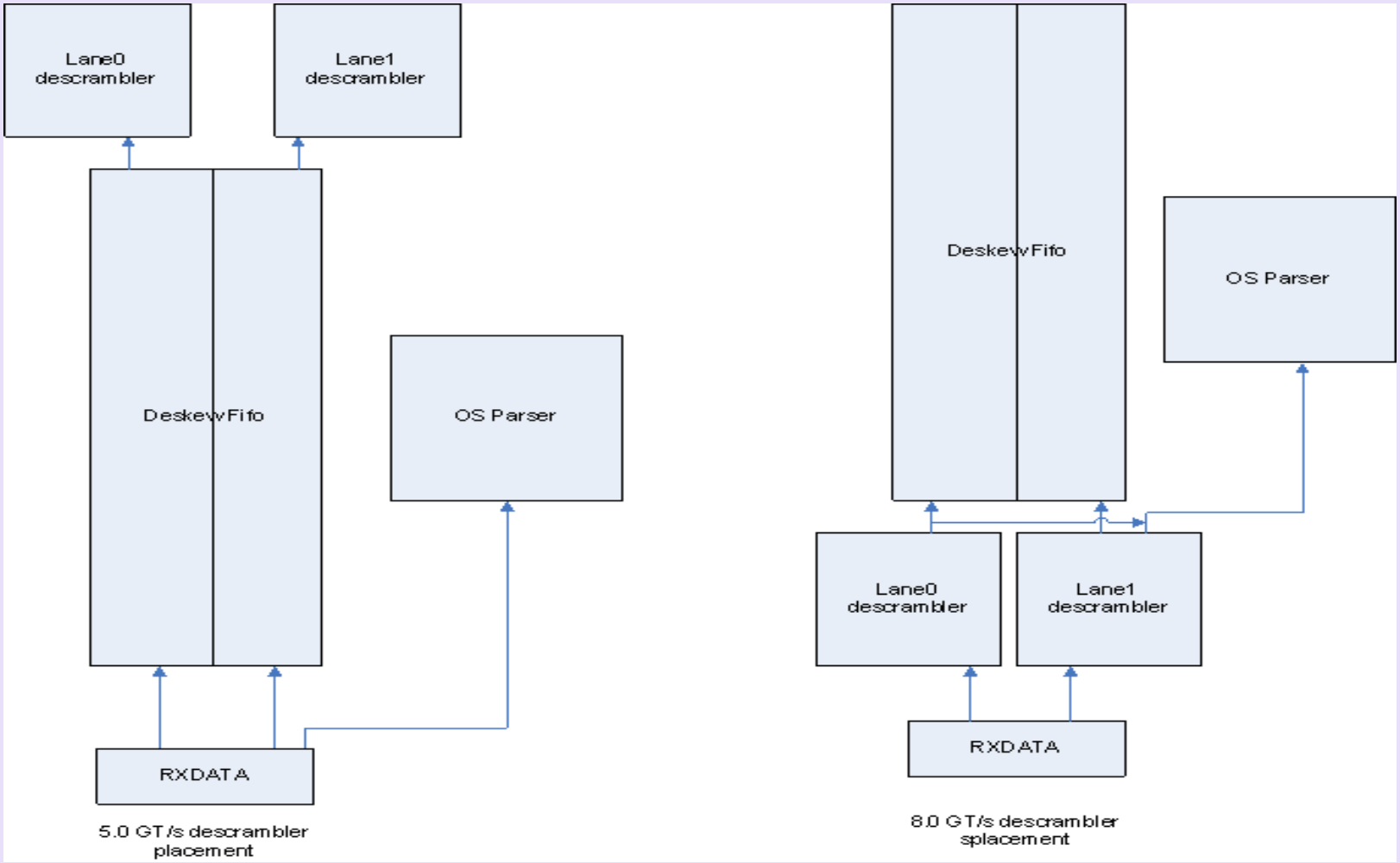
Example Block Alignment State Diagram



Scrambling/descrambling

- Some Ordered Sets are also scrambled at 8 GT/s, while only logical Idle and Data packets are scrambled at 5.0 and 2.5 GT/s.
  - ✓ Placement of scrambler with respect to Lane Deskew Fifo is important at 8 GT/s. if Ordered set detection is done on undeskewed data, then it is preferable to place the descrambler before the deskew Fifo so that TS1/TS2 can be descrambled and fed to the Order set detection logic.
- Scrambling can't be turned off at 8.0 GT/s. The disable scrambling bit in Training Control field of TS1/TS2 doesn't apply at 8.0 GT/s. The disable scrambling information received at 2.5 GT/s has to be stored for re-entry into 2.5/5.0 GT/s.

Scrambling/descrambling



Lane Deskew

- At 8GT/s there is no COMMA character present to perform lane deskew, instead specific Ordered Sets have to be used, i.e.
  - ✓ SDS OS – unambiguous deskew
  - ✓ EIEOS - unambiguous deskew
  - ✓ SKP OS – unambiguous deskew
  - ✓ Other OS (e.g. FTS/TS1/TS2). – can be used for deskew check.

Note: It is best to use all 3 mechanisms of unambiguous deskew, such that if the Port receives back-to-back OS of type1, then an OS of type2 would provide unambiguous deskew mechanism.

Ordered Sets

- At 8 GT/s
  - ✓ TS1/TS2 are scrambled while other ordered sets aren't. Additional Equalization fields have been added in symbols 6,7,8,9 of TS1 and in symbol 6 of TS2.
  - ✓ All Ordered Sets are transmitted with a block length of 16 bytes, even FTS OS and SKP OS.
  - ✓ New Ordered Sets added for 128/130 bit encoding. (SDS).
- At 5.0 or 2.5 GT/s, additional field added in TS1/TS2 symbol 6 to support Equalization.
  - ✓ The Symbol 6 check for TS1/TS2 identifier has to be modified to account for the new equalization field addition.

SKIP Ordered Set

- SKIP Ordered set has fields in last 3 bytes of block which are used differently in special LTSSM states at 8 GT/s.
  - ✓ In Data phase, the field is used for scrambler's LFSR value, and data parity. Parity is calculated in data phase over:
    - SDS to SKP OS
    - SKP OS to next SKP OS
    - SKP OS to EIEOS or EIOS (this value is not used, but parity is calculated, since the EIEOS or EIOS can't be predicted beforehand).
  - ✓ In OS phase, the field is used for scrambler's LFSR value only.
  - ✓ In Polling compliance state, it has special values, and Error status if modified Polling Compliance transmission is in progress.

SKIP Ordered Set

Last 4 bytes of SKP Ordered Set

Symbol number	Value	Description
Last-3	E1	SKP_END
Last-2	0-FF	(i) If LTSSM state is Polling.Compliance: AAh (ii) Else if prior block was a Data Block: Bit[7] = Data Parity; Bit[6:0] = LFSR[22:16] (iii) Else: Bit[7] = ~LFSR[22]; Bit[6:0] = LFSR[22:16]
Last-1	0-FF	(i) If the LTSSM state is Polling.Compliance: Error_Status[7:0] (ii) Else LFSR[15:8]
Last	0-FF	(i) If the LTSSM state is Polling.Compliance: ~Error_Status[7:0] (ii) Else: LFSR[7:0]

SKIP Ordered Set

- SKIP Ordered set can be a minimum of 8 bytes and a maximum of 24 bytes coming into the Port receive side.
  - ✓ The deskew Fifo should be prepared to handle a SKIP OS from 4 bytes upto 28 bytes, if the clock compensation scheme in the elastic buffer uses addition and deletion of 4 bytes in SKIP OS.
  - ✓ If clock compensation mechanism doesn't use addition and deletion of bytes in SKIP OS; instead it makes use of data_valid assertion/deassertion of received data stripped off it's sync header bits, then deskew Fifo can get a SKIP OS ranging from 8 bytes to 24 bytes.

EIOS handling at 8.0 GT/s

- Transmitter is allowed to truncate upto the last 2 bytes of the last EIOS, when Electrical Idle is to be entered after the last EIOS.
- Receivers indicate that EIOS has been received when the first 4 symbols of the Ordered Set match the first 4 symbols of EIOS.
  - ✓ Receivers shouldn't be checking for all 16 bytes of EIOS since last 2 bytes of EIOS can be truncated by Transmitter.

TS handling at 8.0 GT/s

- The Link numbers used at 8.0 GT/s can't be more than 31 decimal. If a Root Complex can support 8.0 GT/s, then it should limit the Link number being used to within 31 decimal, else there will be issue in maintaining the Link number at 8.0 GT/s.
  - ✓ Value of F7 is used to indicate PAD in Link and Lane number fields at 8.0 GT/s.
- Symbol 6,7,8,9 of TS1 and Symbol 6 of TS2 shouldn't be checked to match with TS identifier, since these are being used for equalization fields.

TS1 handling during Equalization sub-state

- The TS1's have to be checked for Equalization request and request response in Phase2/Phase3 equalization sub states at 8 GT/s speeds.
  - ✓ If request preset, then the recipient of the request, should change the transmitter settings to the requested preset, and reflect back the requested preset and corresponding coefficients being used.
  - ✓ If request coefficient, then recipient shouldn't change it's transmitted preset value, but reflect back the coefficient received even if the coefficient is rejected.
    - Note: Special monitors had to be added in verification environment to differentiate response to request preset versus request coefficient.

TLP length

- The TLP framing includes a length field in framing for 128/130 bit encoding.
- If Nullified TLP is indicated, the Nullified TLP has to be transmitted for the length field indicated in the framing for 128/130 bit encoding. This is an additional requirement at 8GT/s.
- EDB indication is provided as information to Link after detecting first byte of EDB, but if the received EDB token is not 4 bytes long, then a framing error is triggered.

Framing Error

- Framing error checking is “required” for 8 GT/s, with some checks as optional.
- Framing error detection in data phase, results in entry into Recovery, and ends data block processing at 8GT/s.
  - ✓ If there is framing Error detected, the OS parser and packet processor should not parse any OS or packets received after the framing error, else there could be multiple errors reported which is not allowed by the spec. The first OS processed after the framing error will be received EIEOS followed by TS1 or TS2.

Misc LTSSM state transition

- Changes in Recovery state specifically for 8 GT/s operation:
  - ✓ The Upstream component initiates speed change to 8 GT/s if rate field of 8 GT/s is both transmitted and received the first time through configuration.
  - ✓ Entry into Recovery.Equalization sub-states at 8 GT/s, after first time in Recovery.Rcvrlock at 8 GT/s.
  - ✓ If framing error was received at 8 GT/s in Recovery.Rcvrcfg, transition to Recovery.Idle happens first and then transition to Recovery.Rcvrlock depending on idle_to_rlock variable.

L0s Exit at 8 GT/s

- The sequence of Ordered Set Transmission and Reception during L0s exit is as follows:
  - ✓ In TXL0s State, the transmitter transmits EIEOS first, followed by required number of FTS, with an EIEOS transmitted every 32 FTS; a final EIEOS following the last FTS, and an SDS after that to indicate Start of Data Stream.(Up to 2 EIEOS can be transmitted at the end of final FTS).
  - ✓ In RXL0s State, the transition to L0 state happens after receiving SDS

Special consideration for Width Change

- During width change EDS token will be received with old width, so checking of EDS token on receive side has to be done keeping the old width in mind.
- If width upconfiguration is initiated, then the port responding to the initiator, will transmit TS1 in Config.Linkwidth.Start state, without transmitting EIEOS first on the lanes which are being upconfigured. Verification environment has to be aware of this detail.

Summary

- Special attention has to be given to Phy and Link layer implementations in order to make changes for PCIe 3.0 .
- The presentation provides some examples of the areas to be careful about, while making changes for PCIe 3.0 .
- PCIe 3.0 is still an evolving specs, so the list of examples is in no way complete.

Thank you for attending the
PCI-SIG Developers Conference 2010.

For more information please go to
www.pcisig.com