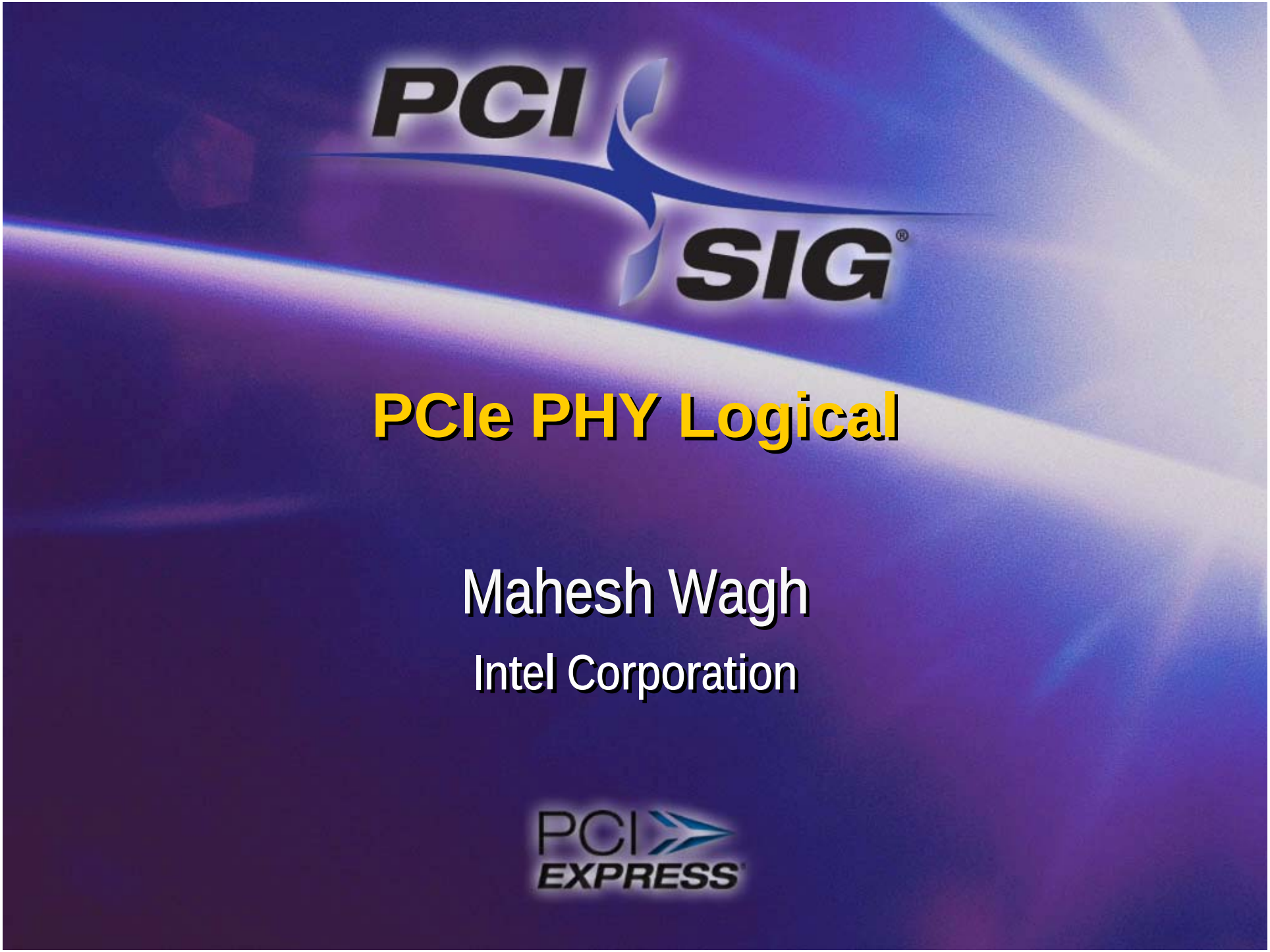

PCI
SIG
PCIe PHY Logical
Mahesh Wagh
Intel Corporation
PCI
EXPRESS

Disclaimer

- NOTE: The information in this presentation refers to a specification still in the development process. This presentation reflects the current thinking of the workgroup, but all material is subject to change before the specification is released.

Agenda

- Problem Statement
- Existing Usage of K-Codes
- Metrics considered for evaluation
- Current Direction on Encoding
- Summary & Call to Action

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
    - Avoid using complicated receiver equalization, etc.
- Requirement: Double Bandwidth from PCIe 2.0
  - ✓ PCIe 1.0a data rate: 2.5 GT/s
  - ✓ PCIe 2.0 data rate: 5 GT/s
    - Doubled the bandwidth from PCIe 1.x to PCIe 2.0 by doubling the data rate
  - ✓ Data rate gives us a 60% boost in bandwidth
  - ✓ Rest will come from Encoding
    - Replace 8b/10b encoding with a scrambling-only encoding scheme when operating at PCIe 3.0 data rate
- Double B/W: Encoding efficiency improvement of 1.25 X data rate improvement of 1.6 yields 2X improvement in bandwidth
- Challenge: 8b/10b encoded the 2^8 data patterns and 12 K-codes

Agenda

- Problem Statement
- Existing Usage of K-Codes
- Metrics used for evaluation
- Current Direction on Encoding
- Summary & Call to Action

Existing Usage of K-Codes

- Two flavors for K-code use
  - ✓ Packet Stream (independent of link width)
  - ✓ Lane Stream (per-lane)
- Packet Stream relates to Packet Framing (Link-Wide)
  - ✓ STP - Start of TLP
  - ✓ END - End (Good) of TLP
  - ✓ EDB - End Bad of TLP
  - ✓ SDP - Start of DLLP
- Lane Stream relates to Ordered Sets:
  - ✓ Training Set #1 & #2
    - – Link training and negotiation
  - ✓ SKP Ordered Sets
    - – Periodic link clock compensation
    - – Recovery from bit slip/add
  - ✓ Electrical Idle Start/ Exit sequence
    - – Power management
- New encoding scheme needs to accommodate these existing usages

Agenda

- Problem Statement
- Existing Usage of K-Codes
- Metrics used for evaluation
- Current Direction on Encoding
- Summary & Call to Action

Error Detection Ability

- Robustness against bit errors considered
  - ✓ Bit flip, bit slip/add
- **Basic Fault Model:**
  - ✓ Guaranteed error detection against random bit flips in any TLP or DLLP or IDL or Ordered Set
    - – Must not alias to a TLP or a DLLP with up to three bit flips
      - Can cause data corruption or flow-control problems
- No guaranteed detection of error with bit slip/add
  - ✓ Same as 2.0 ability
- No self healing for physical layer detected errors
  - ✓ Errors may cause transition to Recovery
- Eventual guaranteed recovery in the presence of multiple errors above including bit slip/add
- Need to handle killer packets
  - ✓ Send a different bit stream on retry of a packet

Other Metrics

- Bandwidth Inefficiency must be low enough
  - ✓ 8b/10b had a 20% inefficiency
  - ✓ New scheme must be in the 1-2% range for inefficiency
    - Would result in close to 2X the bandwidth from PCIe 2.0
- Time Overhead through Recovery as well as L0s/L1 exit must be minimal
  - ✓ Enables better power management without performance penalty
- Bytes continue to be the unit of transmission
  - ✓ Enables single-wide/double-wide type of parallel implementation
    - E.g., no end TLP in bit 3 and a new TLP starts in bit 4 within a byte
  - ✓ Preservation of framing rules and length of TLP/DLLP
- Switch to new encoding after speed change from electrical idle in Recovery.Speed
- Minimal changes beyond PHY layer
  - ✓ Ease of implementation

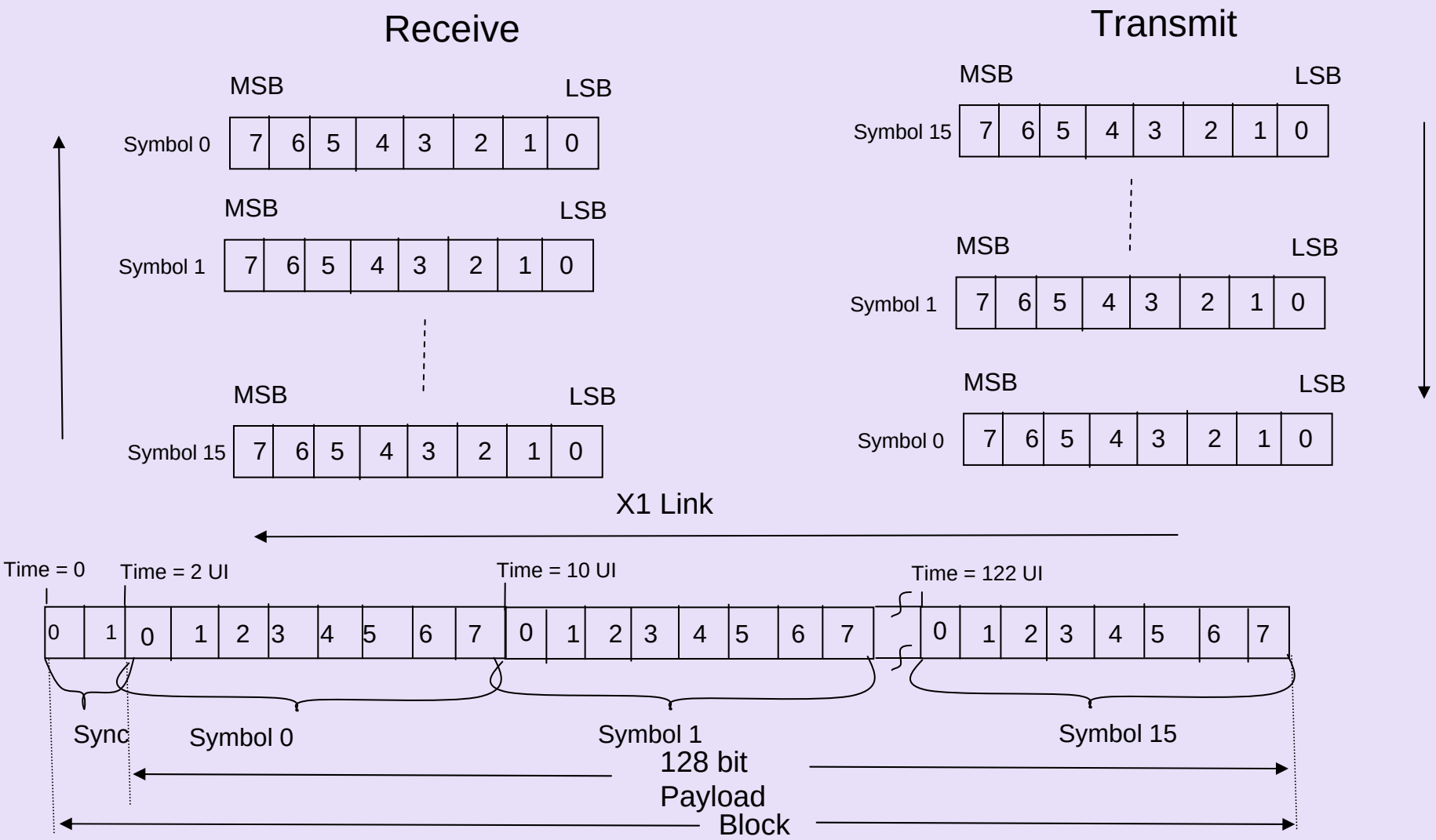
Agenda

- Problem Statement
- Existing Usage of K-Codes
- Metrics used for evaluation
- Current Direction on Encoding
- Summary & Call to Action

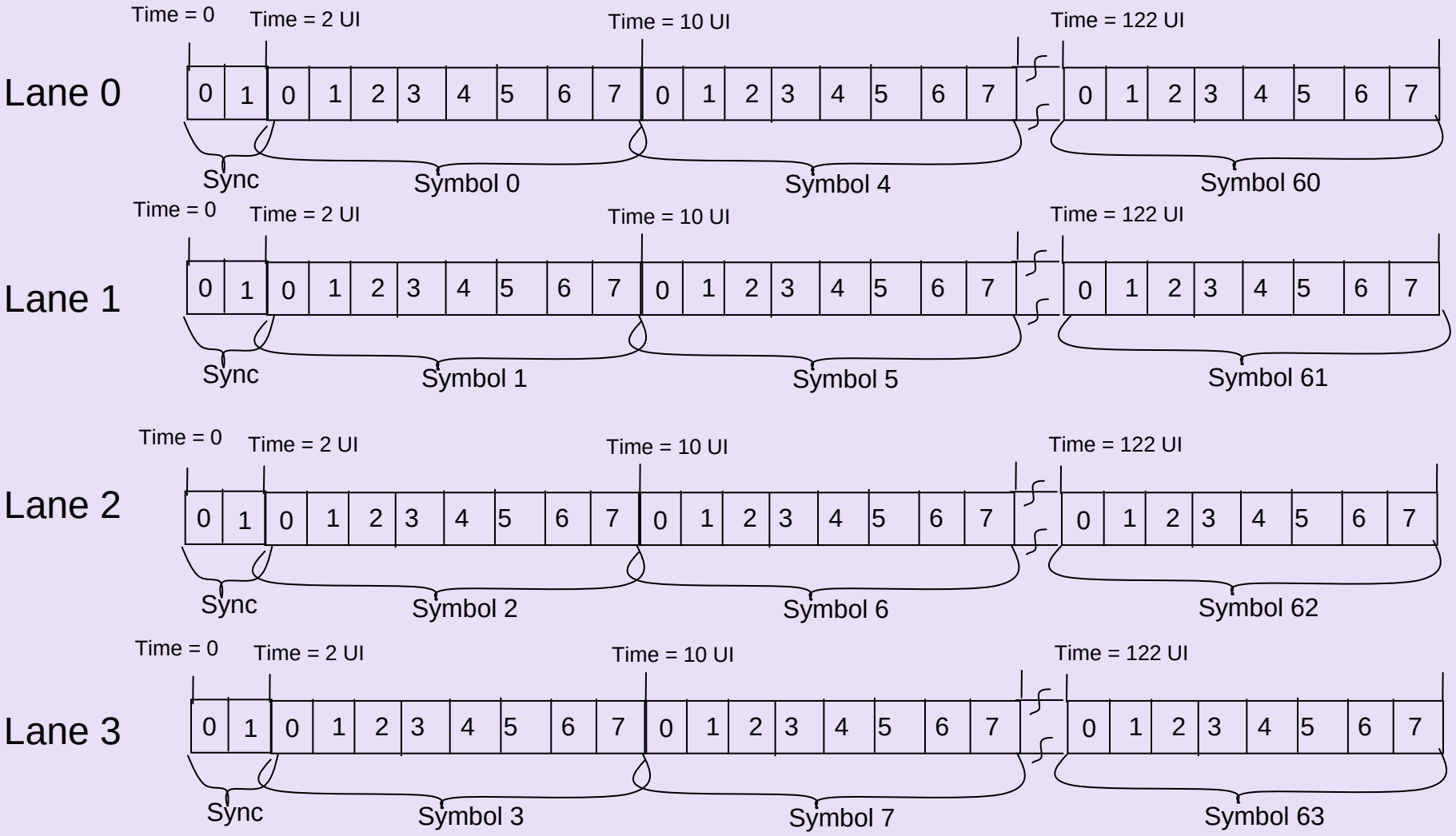
Overall Scheme

- Current direction is to use two levels of encapsulation
  - ✓ Lane level encoding (e.g., 128/130 code) on individual lanes
  - ✓ Physical layer packetization to identify “packet” boundaries
- Lane Level Encoding
  - ✓ Mostly 128/130 bit code. The per-Lane code is called a Block
    - Two types of Blocks: Data Block (TLP, DLLP, LIDL) and Ordered Set Blocks
    - SKP OS blocks Variable in length (may not be 130 bits)
  - ✓ 2 bit Sync Header followed by payload (mostly 128 bit)
  - ✓ Sync header not scrambled
    - 10b (0b followed by 1b in the wire) used for Data Blocks
    - 01b (1b followed by 0b in the wire) used for Ordered Set Blocks
  - ✓ Block lock
    - EIEOS Ordered Set substitutes COM used for Symbol lock in 8b/10b
- Phy Layer packetization to identify packet boundaries. Packet types:
  - ✓ Link Level (TLP or DLLP or LIDL)
  - ✓ Lane Level (Ordered Sets)
- Scrambling only (no 8b/10b) to provide edge density
  - ✓ Additive scrambling on a per-lane basis
  - ✓ Degree 23 polynomial for LFSR with different taps for 8 adjacent lanes (or different seeds for same tap)
  - ✓ Electrical Idle Exit Ordered Set resets scrambler (Recovery/ Config)

Mapping of bits on a x1 Link



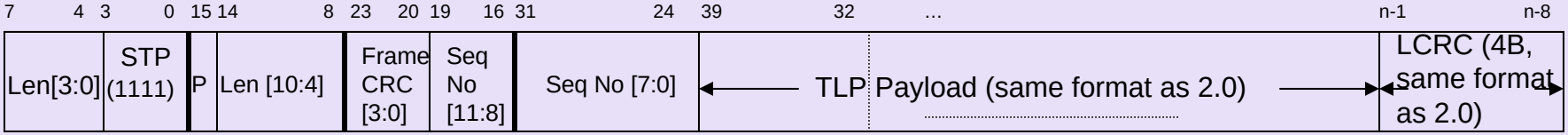
Mapping of bits on a x4 Link



Physical Layer Encapsulation

- First Symbol (scrambled) indicates packet type:
  - ✓ 00000000 is Logical IDL
    - All subsequent lanes in same Symbol-time should be LIDL
    - Receivers check for all 0s (after descrambling) in LIDL
    - PAD functionality merged with LIDL
  - ✓ 1111xxxx is STP
    - Subsequent 11 bits (link wide) define the length
  - ✓ 00001111 is SDP
    - 2nd Symbol also gets a fixed encoding
  - ✓ 00000011 is EDB
    - EDB packet is 4 Symbols; each with the same value 00000011

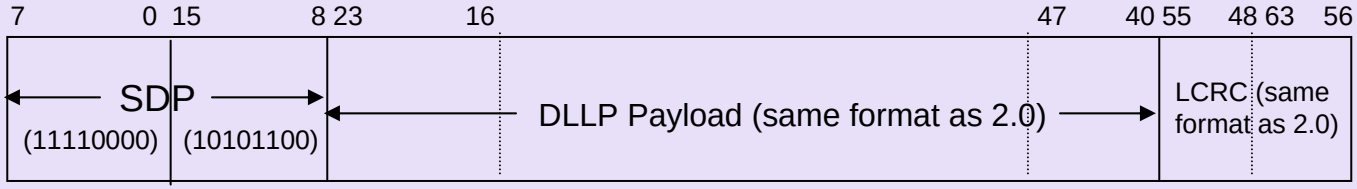
P-Layer Encapsulation: TLP



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

- Length known from the first 3 Symbols
  - ✓ First 4 bits are 1111 (bit[0:3] = 4'b1111)
  - ✓ Bits 4:14 has the length of the TLP (valid values: 5 to 1031)
  - ✓ Bits 15 and 20:23 is check bits to cover the TLP Length field
    - Primitive Polynomial (X4 + X + 1) protects 15 bit field
      - Provides double bit flip detection guarantee (length 11 bits + CRC 4 bits)
    - Odd parity covers the 15 bits (length 11 bits + CRC 4 bits)
      - Guaranteed detection of triple bit errors (over 16 bits)
- Sequence Number occupies bits 16:19 and 24:31
- TLP payload is from the 4th Symbol position (same as 2.0)
- No explicit END. Need to check first Symbol after TLP for implicit END vs an explicit EDB => Ensures triple bit flip detection
- All Symbols are (de)scrambled

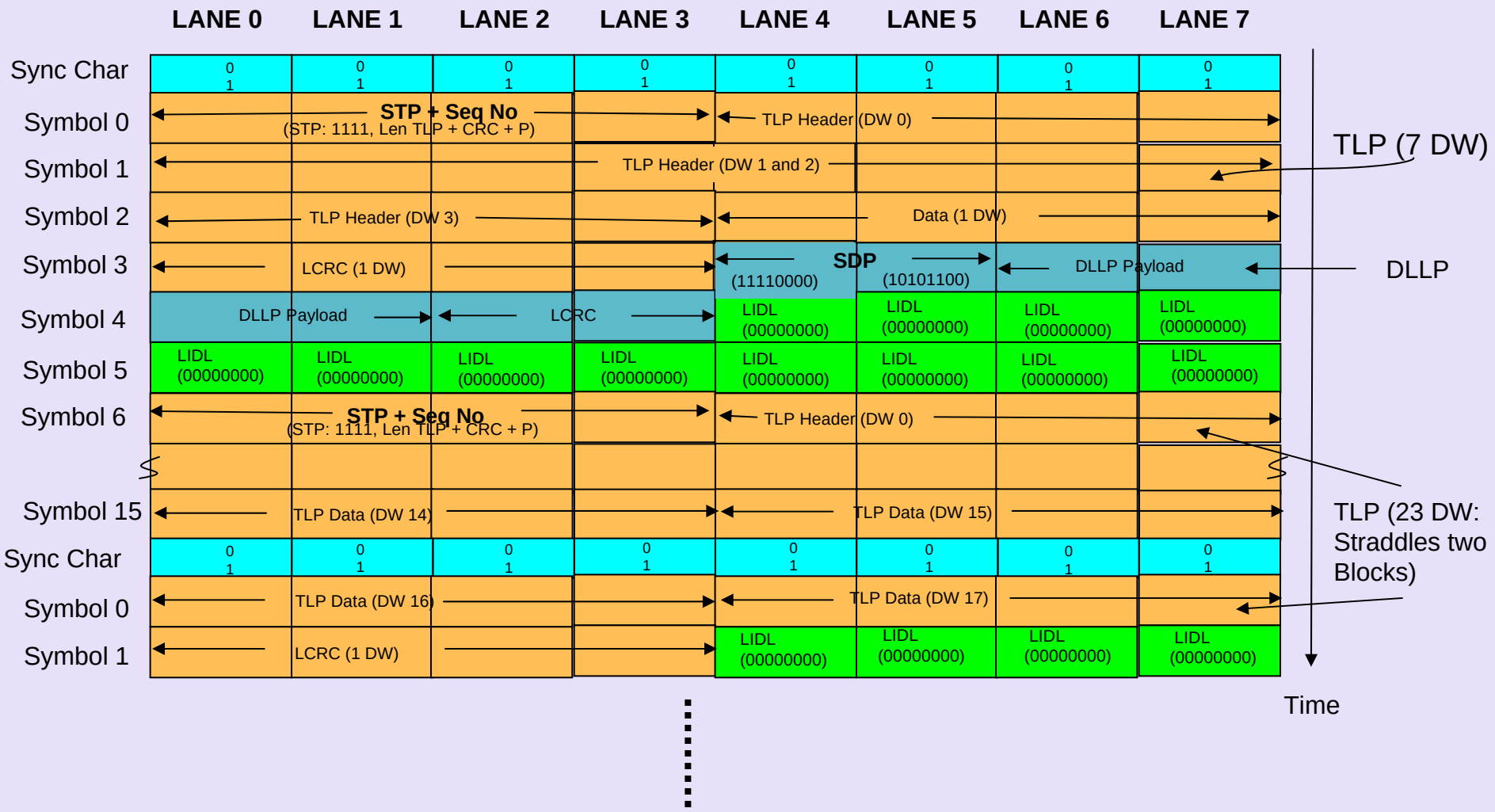
P-Layer Encapsulation: DLLP



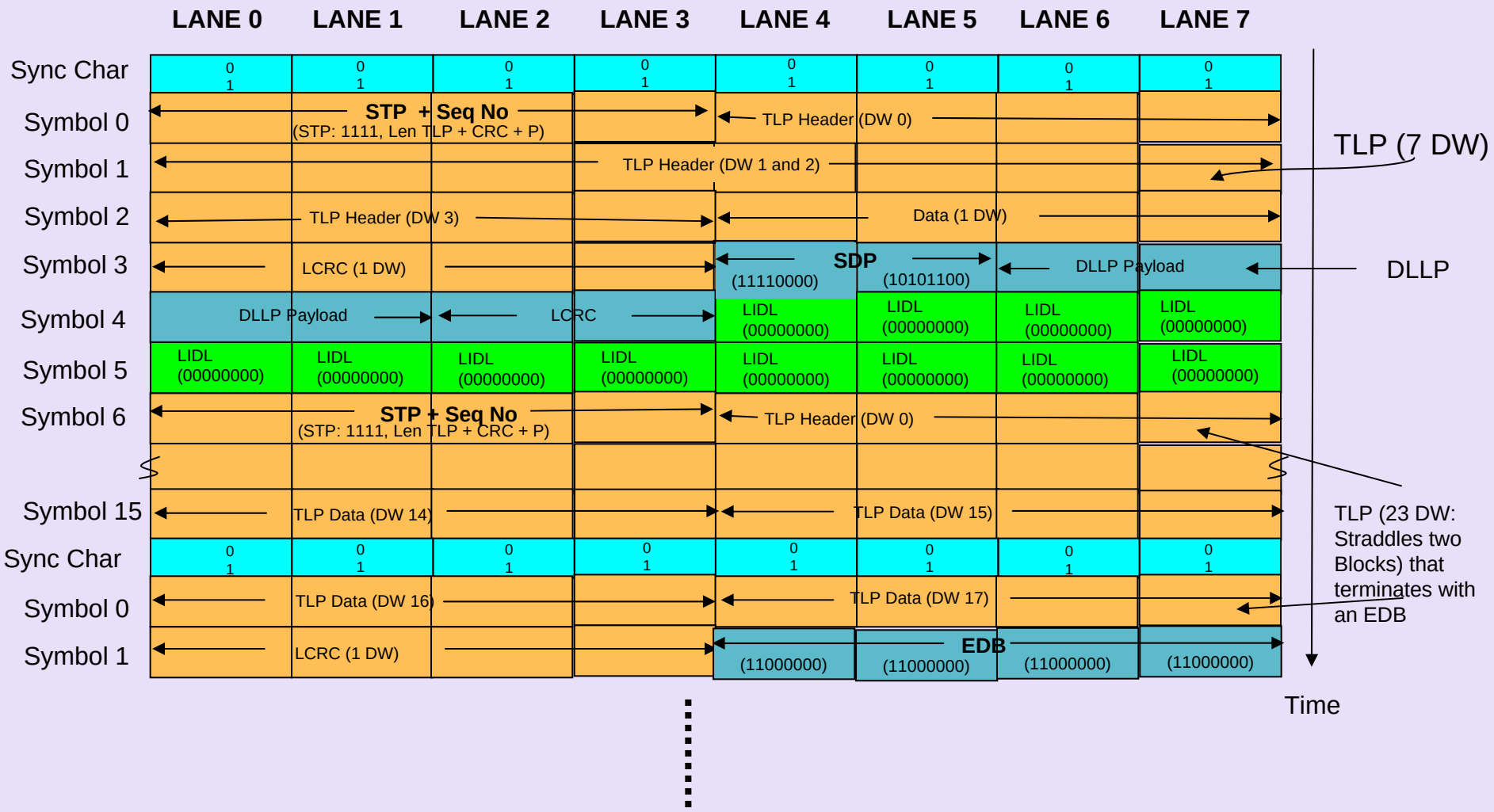
(DLLP Layout)

- Preserve DLLP layout of 2.0 spec
- First Symbol is F0h
- Second Symbol is ACh
- Next 4 Symbols (2 through 5) are the DLLP layout
- Next 2 Symbols (6 and 7): LCRC (identical to 2.0)
- No explicit END
- All Symbols are (de)scrambled

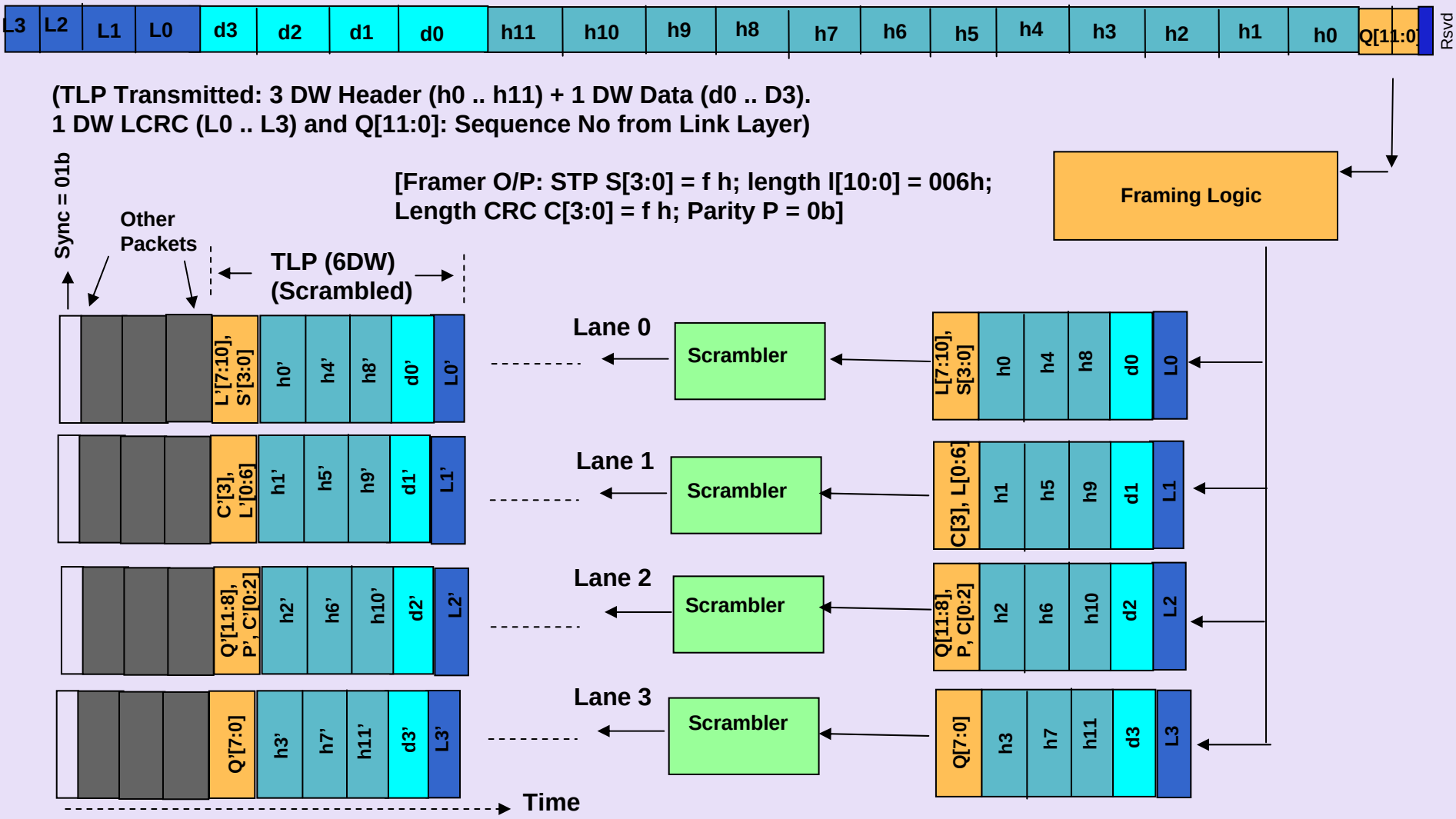
Ex: TLP/ DLLP/ IDLs in x8



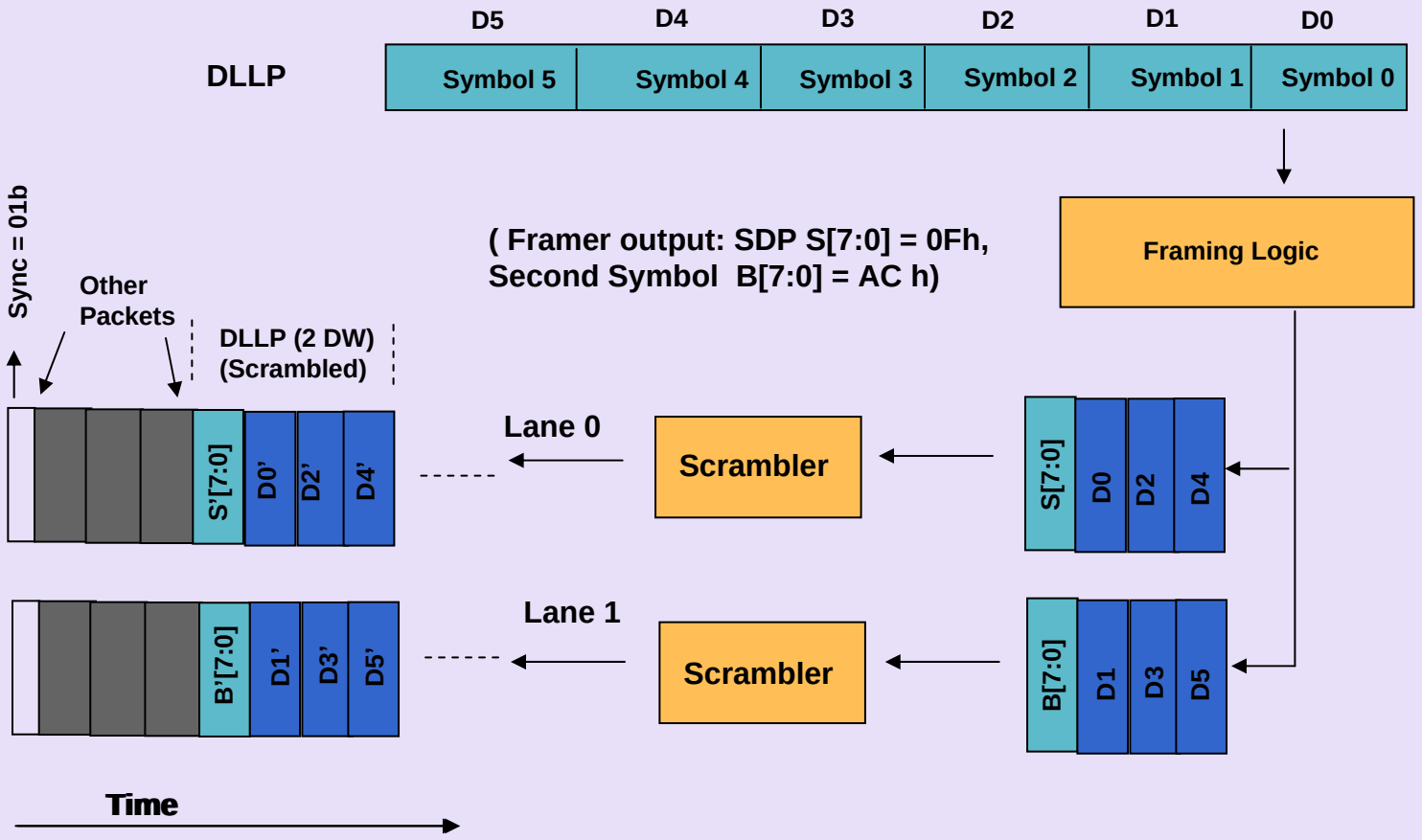
Ex: TLP/ DLLP/ IDLs in x8



TLP Transmission in a X4 Link

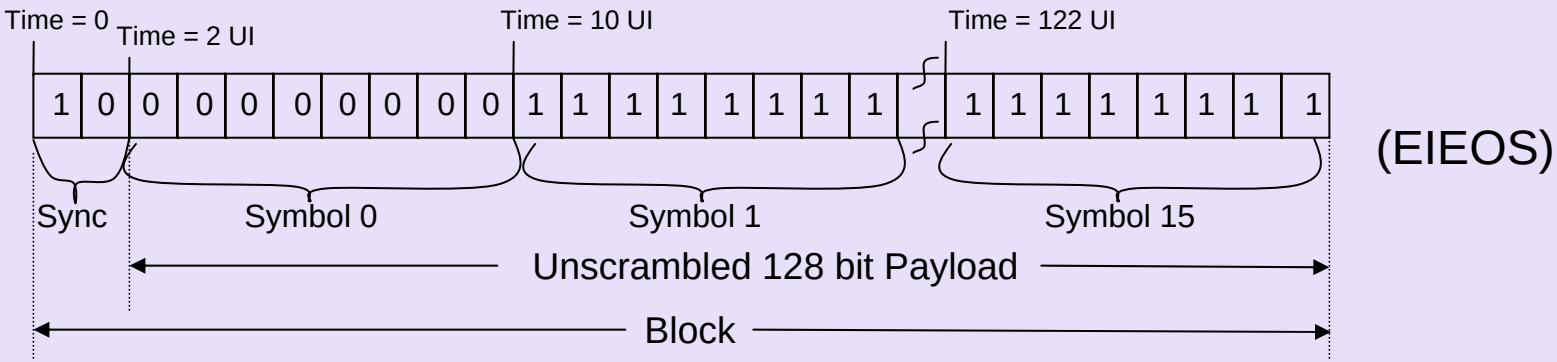


DLLP Transmission in a x2 Link

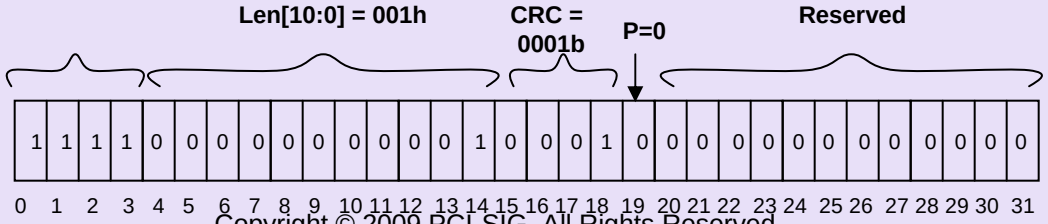


Ordered Sets

- Sync character 01b with the first byte not scrambled
  - ✓ First byte is DC balanced and at a hamming distance 4 from each other
    - TS1: 0001_1110, TS2: 0010_1101, EIEOS: 0000_0000, SKP: 0101_0101, EIOS: 0110_0110, FTS: 1010_1010, SKP_END/ Marker: 1110_0001
  - ✓ EIEOS, EIOS, FTS, Marker OS: 130 bits: not scrambled
  - ✓ EIEOS has low frequency component and used for Block lock in Recovery/ Config
  - ✓ TS1/TS2: 130 bits: last 15 bytes payload are scrambled
  - ✓ SKP OS: 66 bits to 194 bits: not scrambled
- 1DW marker packet 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
- A special marker 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



(EIEOS)



(Marker Packet)

SKP Ordered Set

- SKP Ordered Set
  - ✓ Usage: Retiming Repeaters, Logic Analyzer, Clock Compensation
  - ✓ Not Scrambled
  - ✓ Variable Block Length to allow repeaters to insert/delete SKPs
    - – Four SKPs added/ deleted
  - ✓ Distinct SKP_END to denote the end of SKP OS
  - ✓ Always starts on block boundary and ends the block boundary with SKP_END
  - ✓ LFSR value after SKP OS sent
  - ✓ LFSR not advanced during transmit/receive of SKP OS
  - ✓ Prior block can not carry a TLP or DLLP onto next block
    - – Can not insert SKP OS by breaking a TLP/DLLP

SKP OS Layout

Bit/ Symbol #	Value(s)	Scrambled	Description
Sync Hdr	10b	No	Sync Header: 1b followed by 0b in the wire
0 (SKP)	55h	No	
1 (SKP)	55h	No	
2 (SKP)	55h	No	
3 (SKP)	55h	No	
4 (SKP)	55h	No	
5 (SKP)	55h	No	
6 (SKP)	55h	No	
7 (SKP)	55h	No	
8 (SKP)	55h	No	
9 (SKP)	55h	No	
10 (SKP)	55h	No	
11 (SKP)	55h	No	
12 (SKP_END)	E1h	No	End of SKP OS
13 {LFSR[22], LFSR[22:16]}	00h-FFh	No	LFSR value (bits 22..16) for next block
14 LFSR[15:8]	00h-FFh	No	LFSR value (bits 15..8) for next block
15 LFSR[7:0]	00h-FFh	No	LFSR value (bits 7..0) for next block

Ex: SKP OS in x8

TS1/ TS2 Ordered Set

Bit/ Symbol #	Value(s)	Scrambled	Scrambler Advances?	Description
Sync Hdr	10b	No	No	Sync Header: bit 0 = 1, bit 1 = 0
0	1Eh (2Dh)	No	Yes	TS1 sends 1Eh and TS2 sends 2Dh
1	00-1Fh and FFh	Yes	Yes	Link Number (FFh denotes PAD)
2	00h-1Fh and FFh	Yes	Yes	Lane Number (FFh denotes PAD)
3	00h-FFh	Yes	Yes	N_FTS
4	00h-FFh	Yes	Yes	Data Rate Identifier
5	00h-FFh	Yes	Yes	Training Identifier
6 - 15	4Ah (45h)	Yes	Yes	TS1 sends 1Eh and TS2 sends 2Dh

- Most of TS1/ TS2 scrambled to get a good frequency spectrum along with good transition density
- Effectively the PRBS pattern; reset after 32 X 16 bits due to EIEOS

L0s Entry and Exit

- Entry to L0s: Data Block with EODS Marker -> EIOS Block -> EI
- Exit from L0s: EIEOS -> (N) FTS -> EIEOS -> Marker OS -> Data Block
  - ✓ Block lock either with the last EIEOS (optionally on FTS)
  - ✓ Lane to lane deskew on Marker OS or EIEOS or SKP OS (on extended sync)
  - ✓ On Extended Sync, the (N) FTS number will have SKP OS inserted periodically
  - ✓ TBD: Send EIEOS after every 32 FTS??
  - ✓ Rcvr must check for the marker OS followed by data block and the first data block preceded by marker OS; else direct to Recovery
  - ✓ Can not have a SKP OS preceding the first Data Block even in Extended Sync mode

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 symbol alignment once it switches over looping back
  - ✓ 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 marker packet as slave is just looping back
  - ✓ Master must expect variable number of SKPs in the SKP OS when it does the comparison
- Slave only adjusts the SKP (add, delete, or keep in tact) 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.

Error Detection and Recovery

- Framing error is detected by the physical layer
  - ✓ The first byte of a packet is not one of the allowed sets (e.g., TLP, DLLP, LIDL)
  - ✓ Sync character is not 01 or 10
  - ✓ Same sync character not present in all lanes after deskew
  - ✓ CRC error in the length field of a TLP
  - ✓ Ordered set not one of the allowed encodings or not all lanes sending the same ordered set after deskew (if applicable)
  - ✓ 10 sync header received after 01 sync header without a marker packet in the 01 sync header OR received a marker packet in the 01 sync header and the subsequent sync header in any lane not 10
- Any framing error requires directing LTSSM to Recovery
  - ✓ Stop processing any received TLP/ DLLP after error until we get through Recovery
  - ✓ Block lock acquired with EIEOS
  - ✓ Scrambler reset with each EIEOS
- Error Detection Guarantees
  - ✓ Triple bit flip detection within each TLP/ DLLP/ IDL/ OS

Agenda

- Problem Statement
- Existing Usage of K-Codes
- Metrics used for evaluation
- Current Direction on Encoding
- Summary & Call to Action

Summary & Call to Action

- Encoding scheme decided and development in progress
- Offers advantage of 25% bandwidth for 8GT/s (and above) data rate over 8b/10b encoding
- PCIe 3.0 Base Spec Rev 0.5 complete, awaiting 0.7 draft release.
- Track the spec development and plan for products accordingly

Thank you for attending the
PCI-SIG Developers Conference 2009

For more information please go to
www.pcisig.com