



# PCI Express

7 – July - 2020

Eng. # 3 SW Eng., James Hsiao

A photograph of a modern industrial factory floor. The scene is dominated by complex machinery, including large orange and silver components, yellow support structures, and a network of blue and yellow pipes. The floor is made of polished concrete. In the bottom right foreground, the words "INDUSTRIAL ONLY" are printed in large, white, sans-serif capital letters.

INDUSTRIAL  
ONLY

# Agenda

- Architecture
- Configuration
- Address Space & Transaction Routing
- TLP & DLLP Elements
- Flow Control
- ACK/NAK Protocol
- Physical Layer – Logical
- Link Initialization & Training
- Power Management
- Examples

# PCI Express

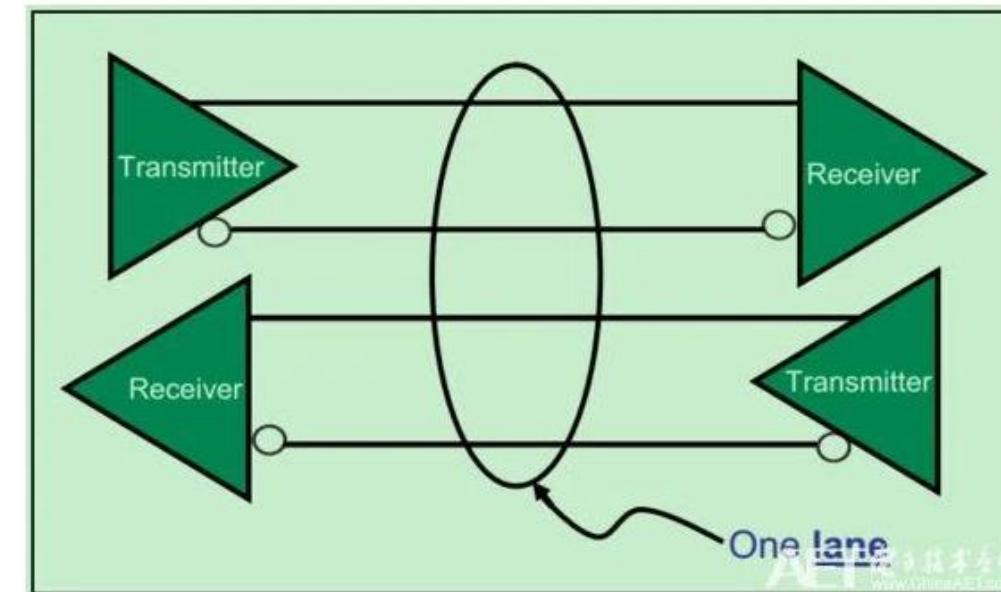
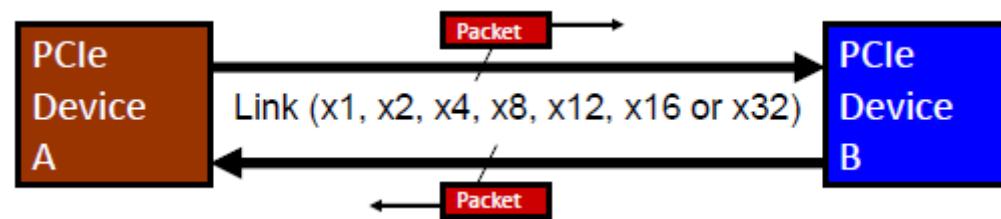
## Architecture



# Architecture

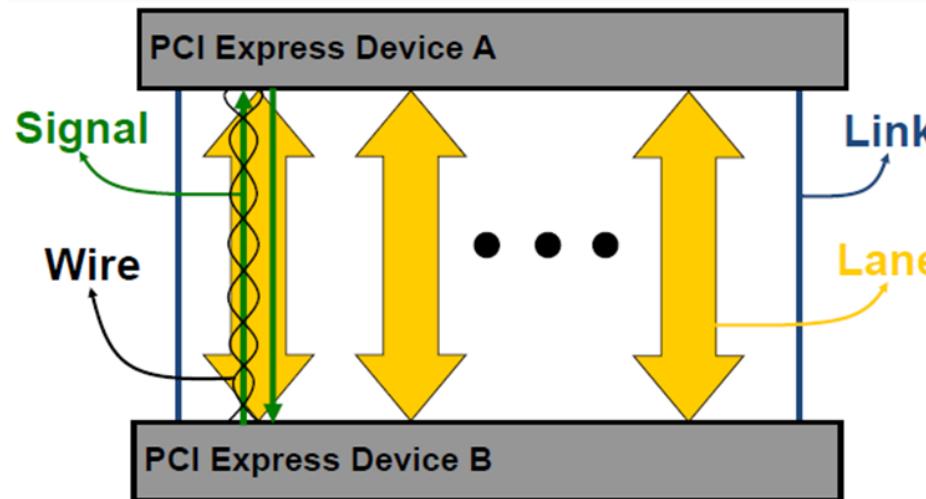
## ■ Serial Transport

- Links and Lanes
  - Dual Simplex point-to-point serial connection
    - Independent transmit and receive sides
  - Scalable Link Widths
    - x1, x2, x4, x8, x12, x16, x32
  - Scalable Link Speeds
    - 2.5, 5.0, 8.0, 16.0 GT/s, 32GT/s
  - Packet based transaction protocol



# Architecture

## ■ Bandwidth



Bandwidth (GB/s)	Link Width				
	x1	x2	x4	x8	x16
PCIe 1.x "2.5 GT/s"	0.25	0.5	1	2	4
PCIe 2.x "5 GT/s"	0.5	1	2	4	8
PCIe 3.x "8 GT/s"	~1	~2	~4	~8	~16
PCIe 4.0 "16GT/s"	~2	~4	~8	~16	~32
PCIe 5.0 "32GT/s"	~4	~8	~16	~32	~64

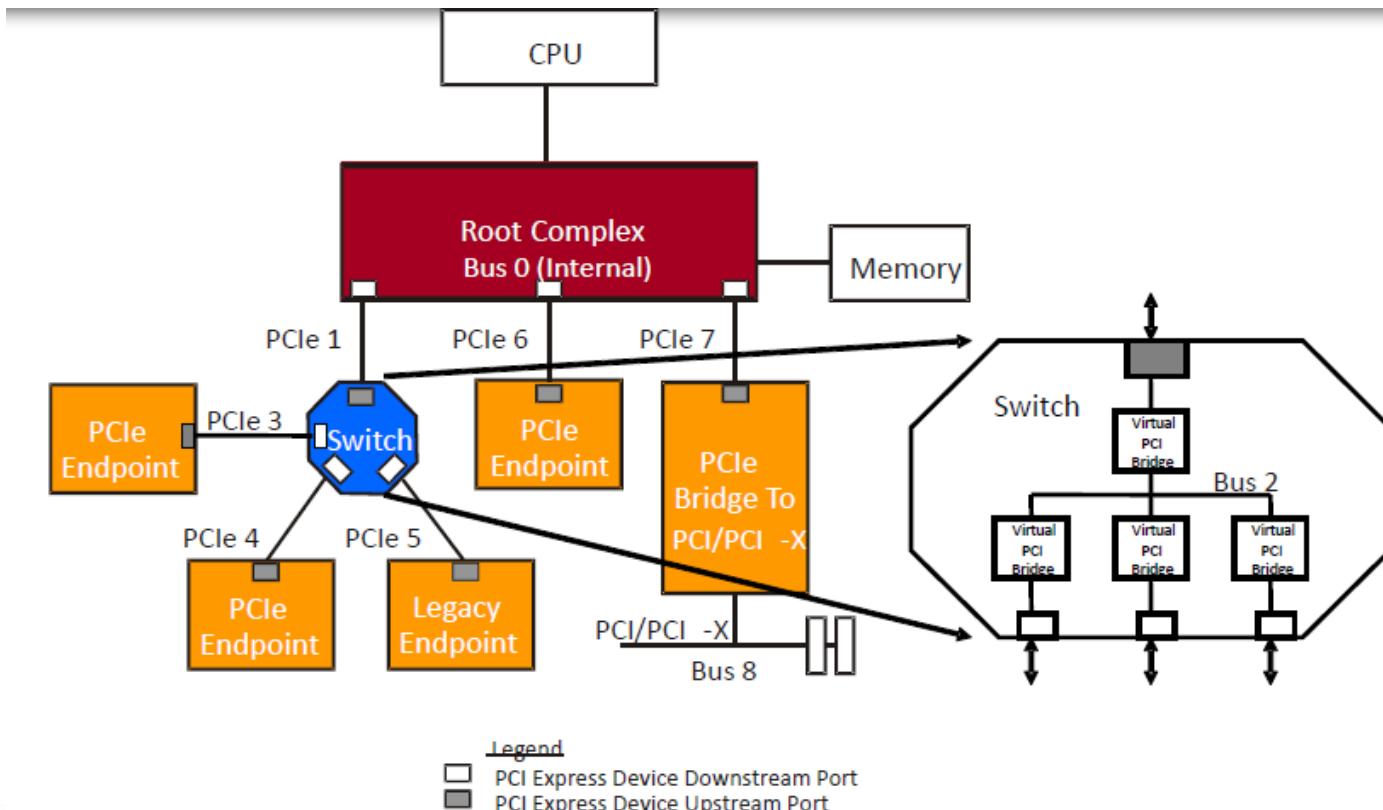
### Derivation of these numbers:

- 20% overhead due to 8b/10b encoding in 1.x and 2.x
- Note: ~1.5% overhead due to 128/130 encoding not reflected above in 3.x and beyond

# Architecture

## ■ Definitions

- ❑ Root Complex denotes the root of an I/O hierarchy that connects the CPU/memory subsystem to the I/O.
- ❑ A PCI Express to PCI/PCI-X Bridge provides a connection between a PCI Express fabric and a PCI/PCI-X hierarchy.
- ❑ A Switch is defined as a logical assembly of multiple virtual PCI-to-PCI Bridge devices.
- ❑ Endpoint refers to a type of Function that can be the Requester or Completer of a PCI Express transaction either on its own behalf or on behalf of a distinct non-PCI Express device (other than a PCI device or host CPU). Endpoints are classified as either legacy or PCI Express Endpoints.

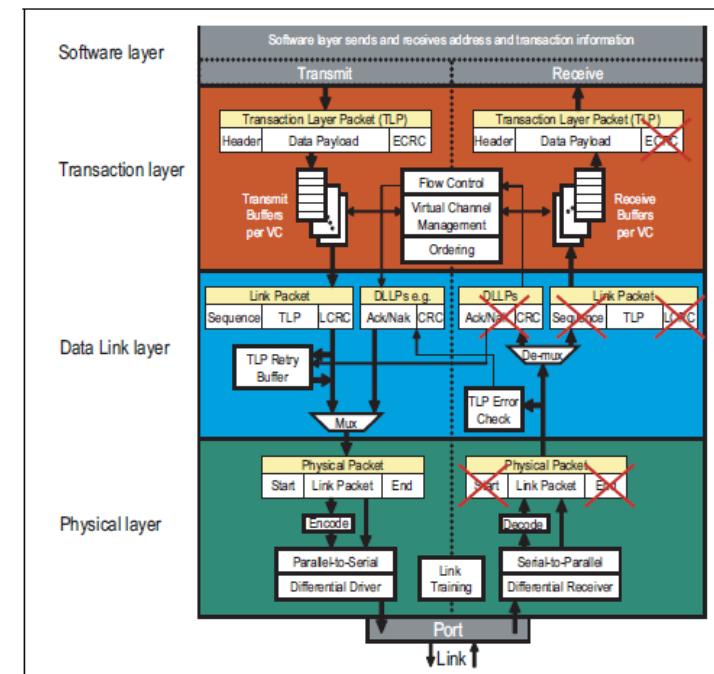


# Architecture

## ■ Device Layers

- The Transaction Layer's primary responsibility is the assembly and disassembly of TLPs. TLPs are used to communicate transactions, such as read and write, as well as certain types of events. The Transaction Layer is also responsible for managing credit-based flow control for TLPs.
- The Data Link Layer, serves as an intermediate stage between the Transaction Layer and the Physical Layer. The primary responsibilities of the Data Link Layer include Link management and data integrity, including error detection and error correction. The Data Link Layer also generates and consumes DLLPs that are used for Link management functions.
- The Physical Layer includes all circuitry for interface operation, including driver and input buffers, parallel-to-serial and serial-to-parallel conversion, PLL(s), and impedance matching circuitry. It also includes logical functions related to interface initialization and maintenance. This Layer is responsible for converting information received from the Data Link Layer into an appropriate serialized format and transmitting it across the PCI Express Link at a frequency and width compatible with the device connected to the other side of the Link.

Figure 11-1: PCIe Port Layers



# Architecture

## ■ Transaction

### □ Transaction Type

PCI Express transactions can be grouped into four categories:

1) memory, 2) IO, 3) configuration, and 4) message transactions. Memory, IO and configuration transactions are supported in PCI and PCI-X architectures, but the message transaction is new to PCI Express. Transactions are defined as a series of one or more packet transmissions required to complete an information transfer between a requester and a completer. Table 2-1 is a more detailed list of transactions. These transactions can be categorized into non-posted transactions and posted transactions.

*Table 2-1: PCI Express Non-Posted and Posted Transactions*

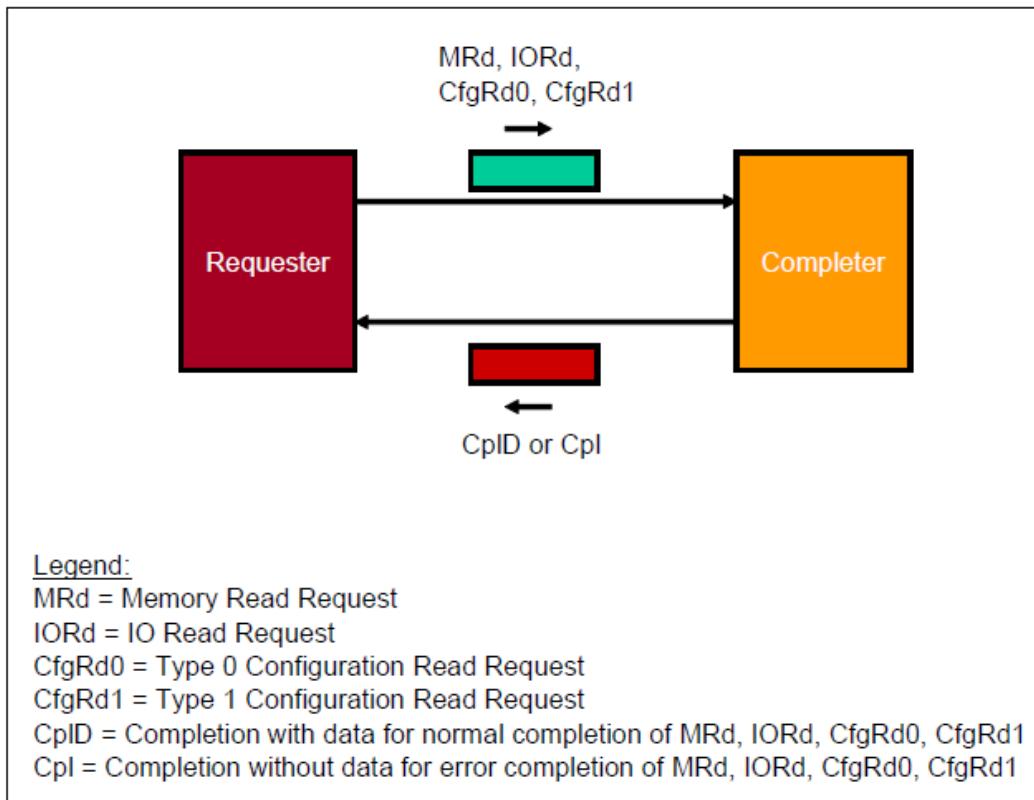
Transaction Type	Non-Posted or Posted
Memory Read	Non-Posted
Memory Write	Posted
Memory Read Lock	Non-Posted
IO Read	Non-Posted
IO Write	Non-Posted
Configuration Read (Type 0 and Type 1)	Non-Posted
Configuration Write (Type 0 and Type 1)	Non-Posted
Message	Posted

# Architecture

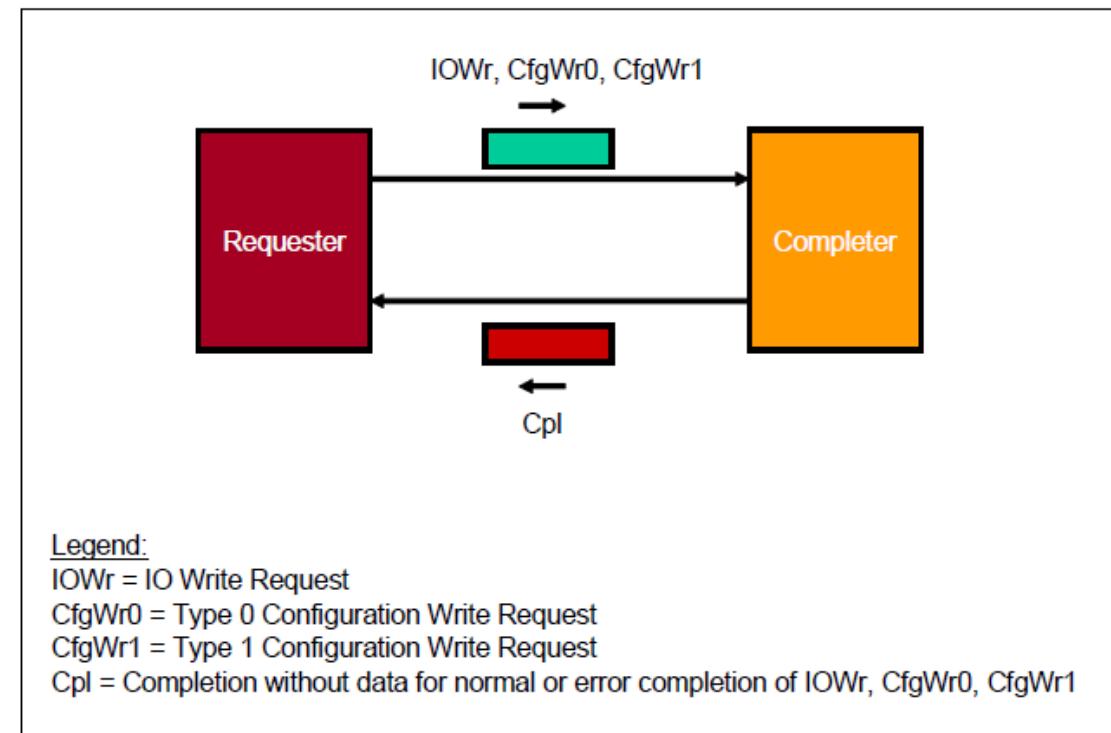
## ■ Transaction

### □ Non-Posted

*Figure 2-1: Non-Posted Read Transaction Protocol*



*Figure 2-3: Non-Posted Write Transaction Protocol*



# Architecture

## ■ Transaction

### □ Posted

Figure 2-4: Posted Memory Write Transaction Protocol

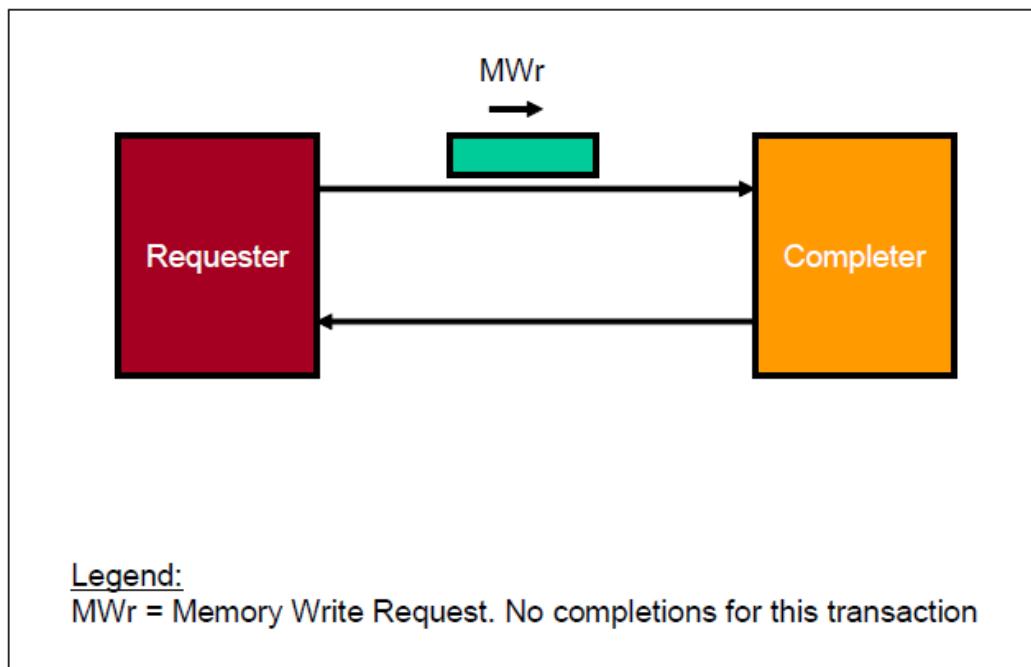
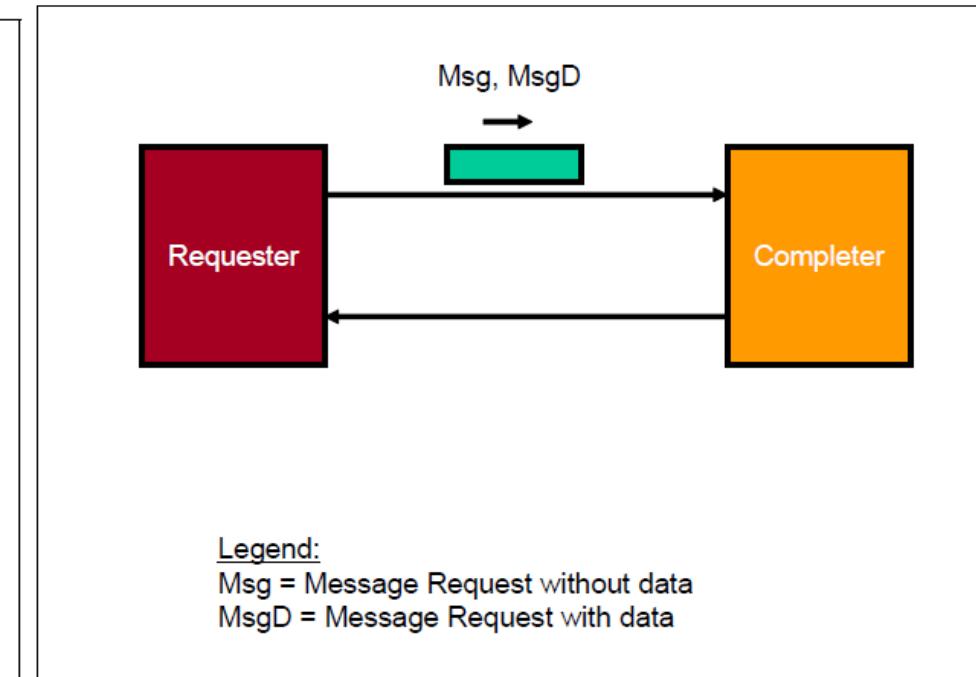


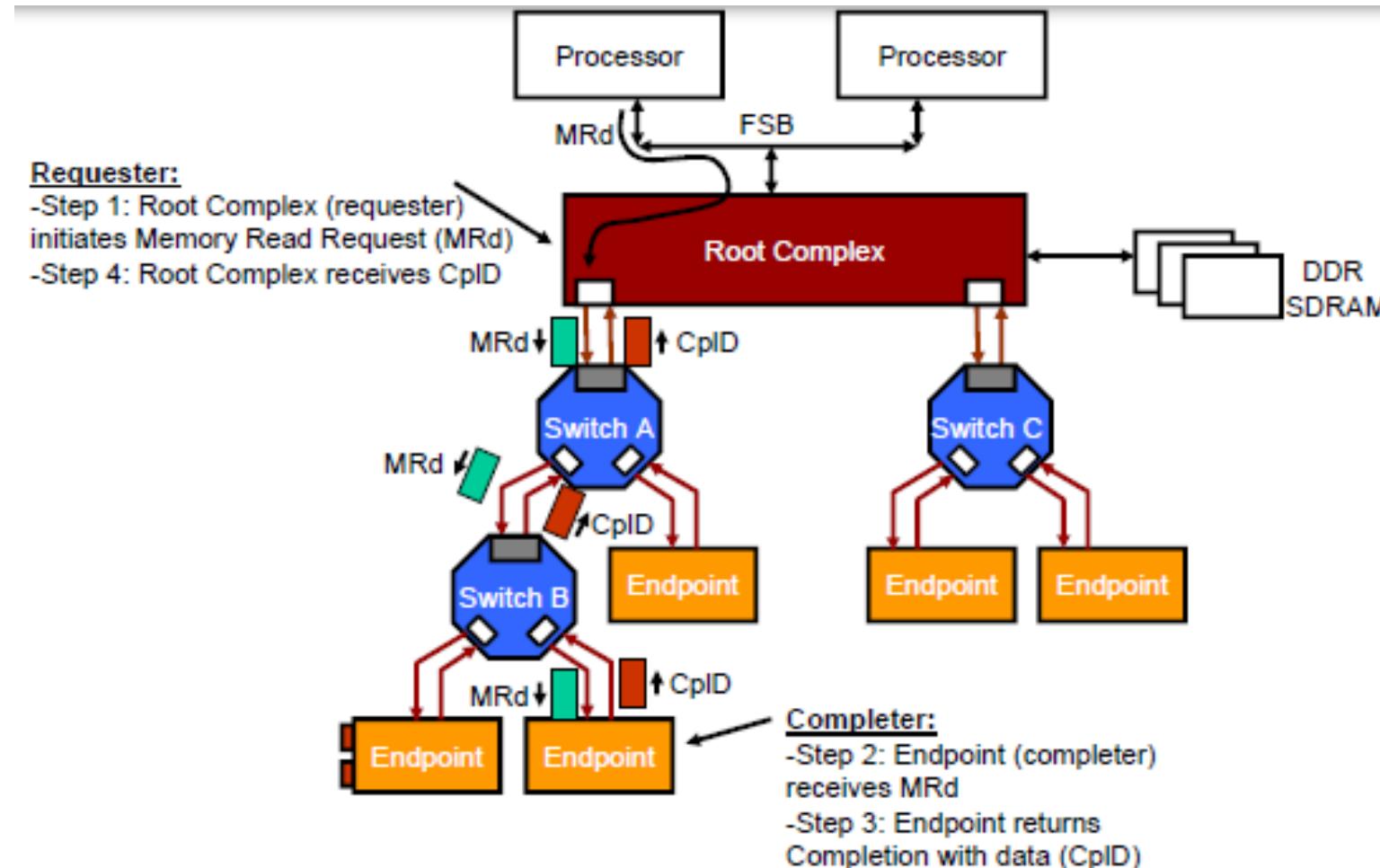
Figure 2-5: Posted Message Transaction Protocol



# Architecture

## ■ Transaction

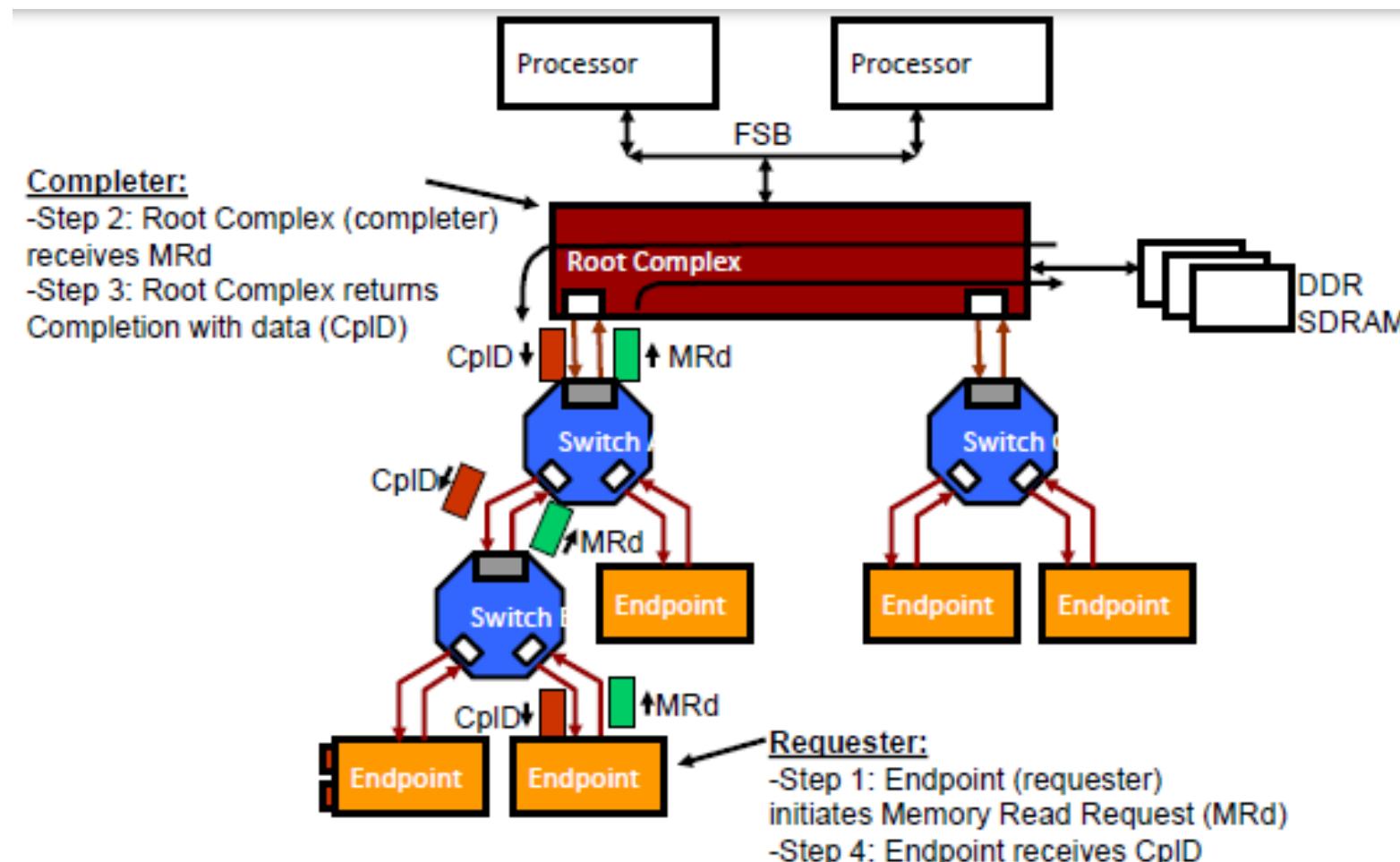
### □ Programmed I/O Transaction



# Architecture

## ■ Transaction

### DMA Transaction



# PCI Express

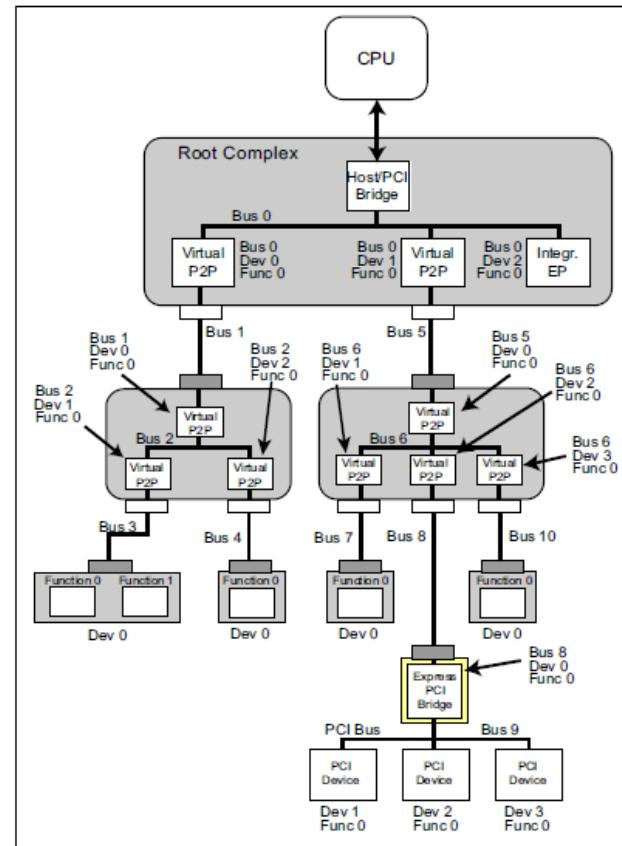
## Configuration



# Configuration

## ■ Bus, Device and Function

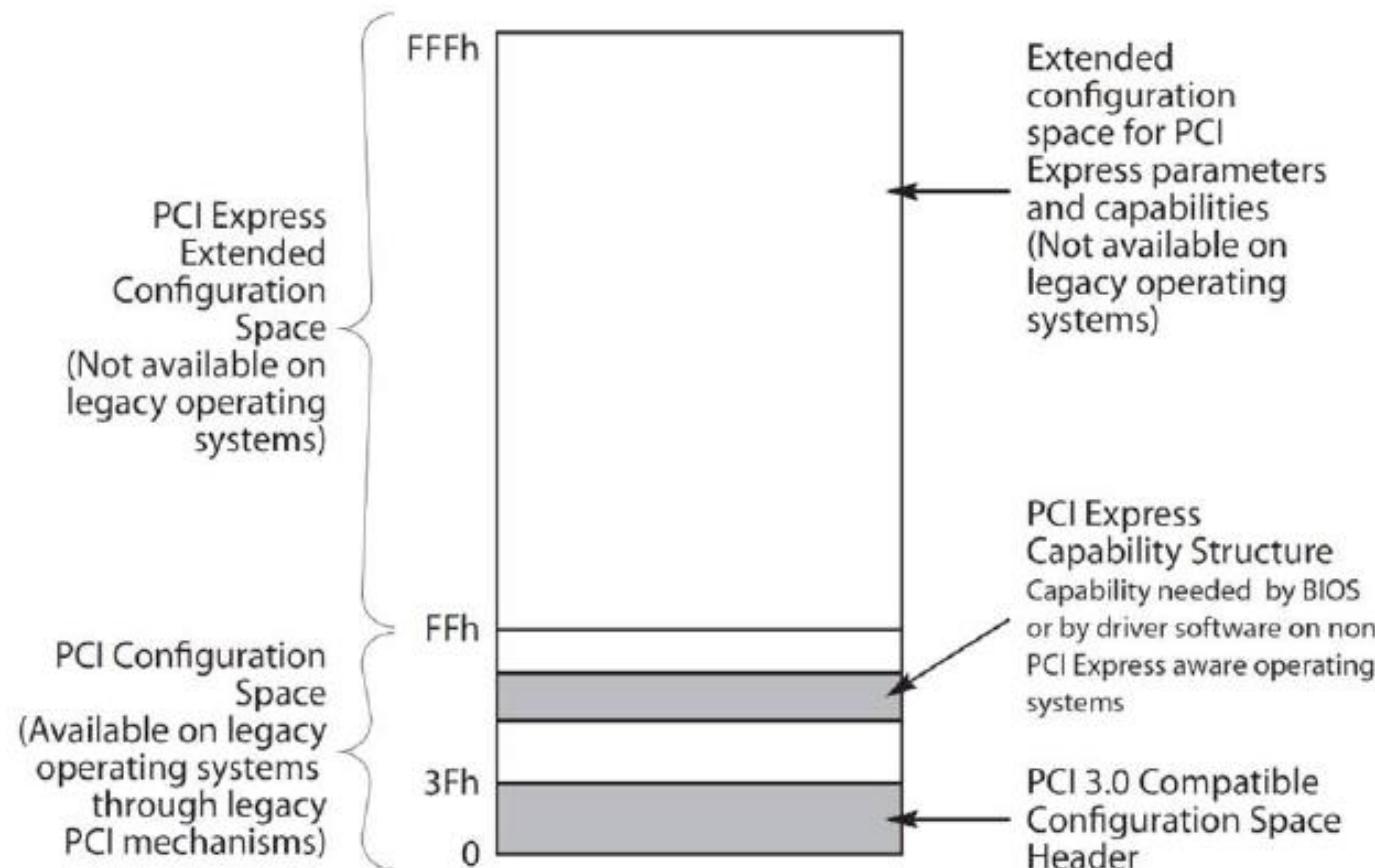
- Every PCIe Function is uniquely identified by the Device it resides within and the Bus to which the Device connects. This ID is commonly referred to as a 'BDF'. Up to 256 Bus Numbers can be assigned by configuration software. The initial Bus Number, Bus 0, is typically assigned by hardware to the Root Complex. PCIe permits up to 32 device attachments on a single PCI bus. Each Device must implement Function 0 and may contain a collection of up to eight Functions. When two or more Functions are implemented the Device is called a multi-function device. These Functions may include hard drive interfaces, display controllers, ethernet controllers, USB controllers, etc. Each Function has its own configuration address space that is used to setup the resources associated with the Function.



# Configuration

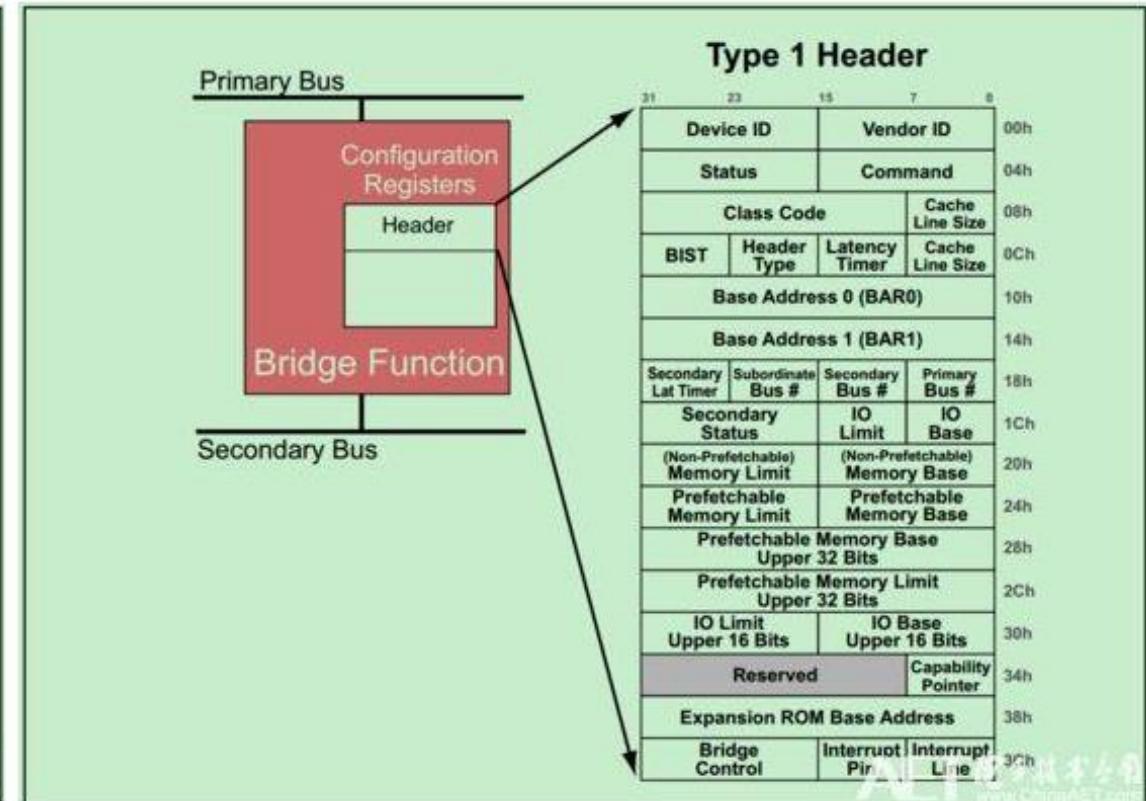
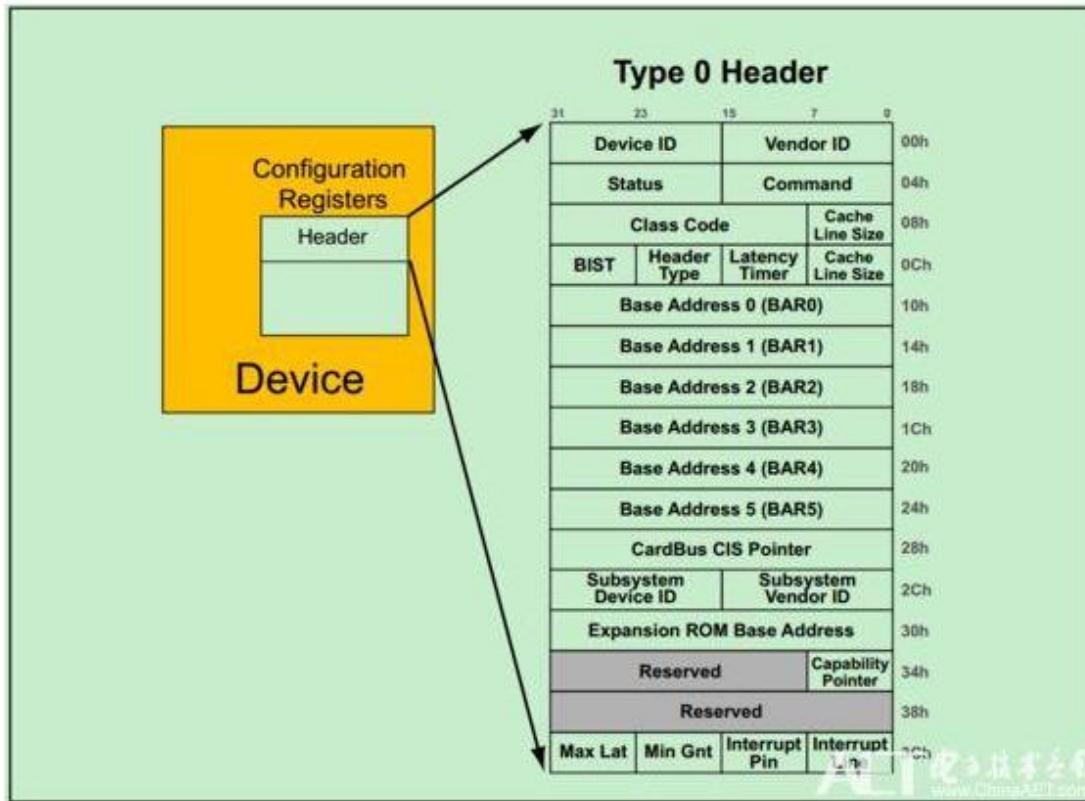
## ■ Configuration Space

- PCI Express extends the Configuration Space to 4096 bytes per Function as compared to 256 bytes allowed by PCI. PCI Express Configuration Space is divided into a PCI-compatible region, which consists of the first 256 bytes of a Function's Configuration Space, and a PCI Express Extended Configuration Space which consists of the remaining Configuration Space.



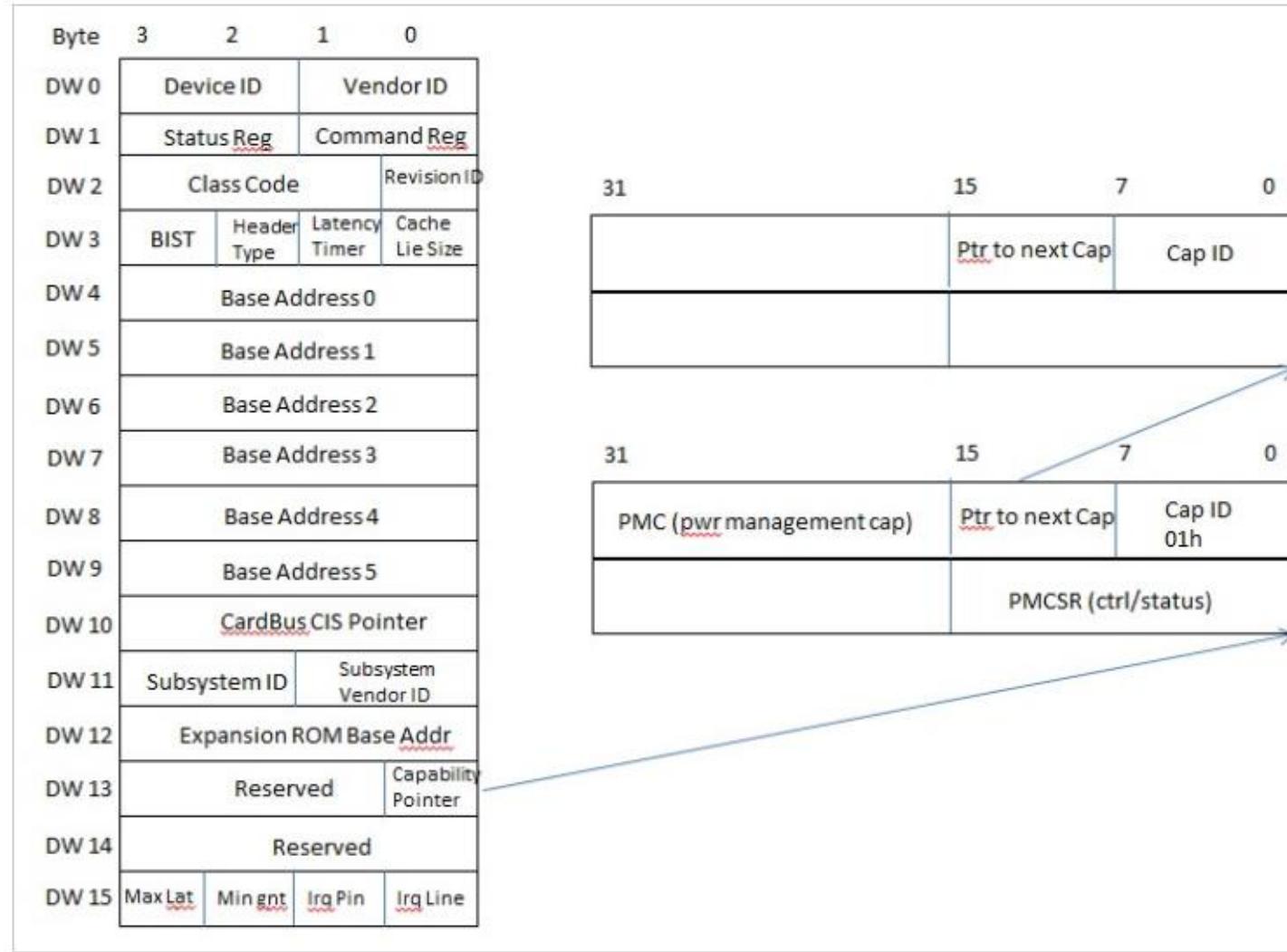
# Configuration

## ■ PCI-Compatible Configuration Space



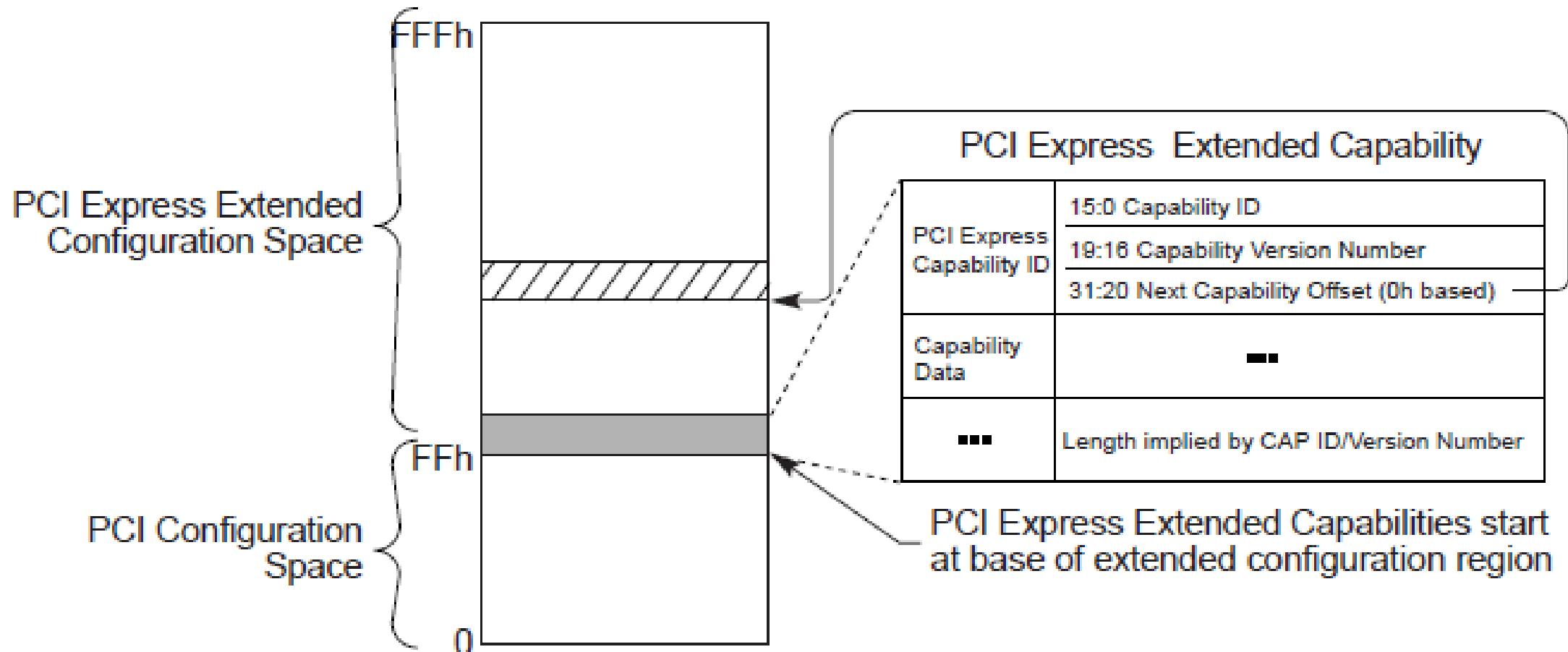
# Configuration

## ■ PCI-Compatible Configuration Space



# Configuration

## ■ Extended Configuration Space



# Configuration

## NVMe - System Bus (PCI Express) Registers

Figure 9: PCI Express Registers

Start	End	Name	Type
00h	3Fh	PCI Header	
PMCAP	PMCAP+7h	PCI Power Management Capability	PCI Capability
MSICAP	MSICAP+9h	Message Signaled Interrupt Capability	PCI Capability
MSIXCAP	MSIXCAP+Bh	MSI-X Capability	PCI Capability
PXCAP	PXCAP+29h	PCI Express Capability	PCI Capability
AERCAP	AERCAP+47h	Advanced Error Reporting Capability	PCI Express Extended Capability

Figure 10: PCI Header

Start	End	Symbol	Name
00h	03h	ID	Identifiers
04h	05h	CMD	Command Register
06h	07h	STS	Device Status
08h	08h	RID	Revision ID
09h	0Bh	CC	Class Codes
0Ch	0Ch	CLS	Cache Line Size
0Dh	0Dh	MLT	Master Latency Timer
0Eh	0Eh	HTYPE	Header Type
0Fh	0Fh	BIST	Built-In Self Test (Optional)
10h	13h	MLBAR (BAR0)	Memory Register Base Address, lower 32-bits <BAR0>
14h	17h	MUBAR (BAR1)	Memory Register Base Address, upper 32-bits <BAR1>
18h	1Bh	BAR2	Refer to section 2.1.12
1Ch	1Fh	BAR3	Vendor Specific
20h	23h	BAR4	Vendor Specific
24h	27h	BAR5	Vendor Specific
28h	2Bh	CCPTR	CardBus CIS Pointer
2Ch	2Fh	SS	Subsystem Identifiers
30h	33h	EROM	Expansion ROM Base Address (Optional)
34h	34h	CAP	Capabilities Pointer
35h	3Bh	R	Reserved
3Ch	3Dh	INTR	Interrupt Information
3Eh	3Eh	MGNT	Minimum Grant (Optional)
3Fh	3Fh	MLAT	Maximum Latency (Optional)

01h	07h	00h	Serial Attached SCSI (SAS) controller - vendor-specific interface
	01h		Obsolete
08h	00h		Non-volatile memory subsystem - vendor-specific interface
	01h		Non-volatile memory subsystem - NVMe interface (see note 8)
	02h		NVMe (NVMe) I/O controller (see Note 6)
	03h		NVMe (NVMe) administrative controller (see Note 6)

# Configuration

## NVMe - Controller Registers

Figure 68: Register Definition

Start	End	Symbol	Description
0h	7h	CAP	Controller Capabilities
8h	Bh	VS	Version
Ch	Fh	INTMS	Interrupt Mask Set
10h	13h	INTMC	Interrupt Mask Clear
14h	17h	CC	Controller Configuration
18h	1Bh	Reserved	Reserved
1Ch	1Fh	CTS	Controller Status
20h	23h	NSSR	NVM Subsystem Reset (Optional)
24h	27h	AQA	Admin Queue Attributes
28h	2Fh	ASQ	Admin Submission Queue Base Address
30h	37h	ACQ	Admin Completion Queue Base Address
38h	3Bh	CMBLOC	Controller Memory Buffer Location (Optional)
3Ch	3Fh	CMBSZ	Controller Memory Buffer Size (Optional)
40h	43h	BPINFO	Boot Partition Information (Optional)
44h	47h	BPRSEL	Boot Partition Read Select (Optional)
48h	4Fh	BPMBL	Boot Partition Memory Buffer Location (Optional)
50h	57h	CMBMSC	Controller Memory Buffer Memory Space Control (Optional)
58h	5Bh	CMBSTS	Controller Memory Buffer Status (Optional)
5Ch	DFFh	Reserved	Reserved
E00h	E03h	PMRCAP	Persistent Memory Capabilities (Optional)
E04h	E07h	PMRCTL	Persistent Memory Region Control (Optional)
E08h	E0Bh	PMRSTS	Persistent Memory Region Status (Optional)
E0Ch	E0Fh	PMREBS	Persistent Memory Region Elasticity Buffer Size
E10h	E13h	PMRSWTP	Persistent Memory Region Sustained Write Throughput
E14h	E1Bh	PMRMSC	Persistent Memory Region Controller Memory Space Control (Optional)
E1Ch	FFFh	Reserved	Command Set Specific
1000h	1003h	SQ0TDBL	Submission Queue 0 Tail Doorbell (Admin)
1000h + (1 * (4 << CAP.DSTRD))	1003h + (1 * (4 << CAP.DSTRD))	CQ0HDBL	Completion Queue 0 Head Doorbell (Admin)

Start	End	Symbol	Description
1000h + (2 * (4 << CAP.DSTRD))	1003h + (2 * (4 << CAP.DSTRD))	SQ1TDBL	Submission Queue 1 Tail Doorbell
1000h + (3 * (4 << CAP.DSTRD))	1003h + (3 * (4 << CAP.DSTRD))	CQ1HDBL	Completion Queue 1 Head Doorbell
1000h + (4 * (4 << CAP.DSTRD))	1003h + (4 * (4 << CAP.DSTRD))	SQ2TDBL	Submission Queue 2 Tail Doorbell
1000h + (5 * (4 << CAP.DSTRD))	1003h + (5 * (4 << CAP.DSTRD))	CQ2HDBL	Completion Queue 2 Head Doorbell
...	...	...	...
1000h + (2y * (4 << CAP.DSTRD))	1003h + (2y * (4 << CAP.DSTRD))	SQyTDBL	Submission Queue y Tail Doorbell
1000h + ((2y + 1) * (4 << CAP.DSTRD))	1003h + ((2y + 1) * (4 << CAP.DSTRD))	CQyHDBL	Completion Queue y Head Doorbell
			Vendor Specific (Optional)

# Configuration

## ■ Enhanced Configuration Access Mechanism

*Table 7-1 Enhanced Configuration Address Mapping*

Memory Address <sup>132</sup>	PCI Express Configuration Space
$A[(20+n-1):20]$	Bus Number $1 \leq n \leq 8$
$A[19:15]$	Device Number
$A[14:12]$	Function Number
$A[11:8]$	Extended Register Number
$A[7:2]$	Register Number
$A[1:0]$	Along with size of the access, used to generate Byte Enables

# Configuration

## ■ Configuration Requests

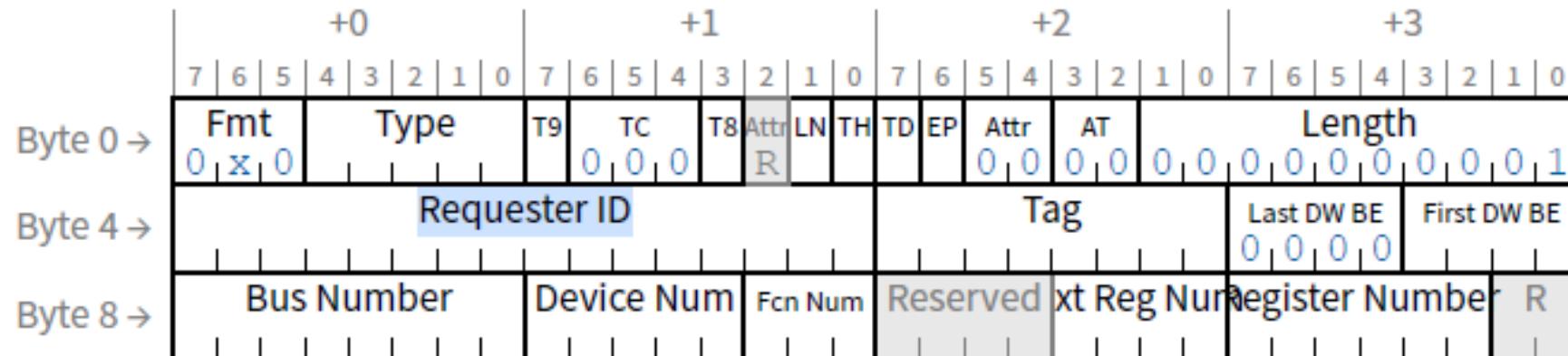


Figure 2-20 Request Header Format for Configuration Transactions

TLP Type	Fmt [2:0] (b)	Type [4:0] (b)	Description
CfgRd0	000	0 0 1 0 0	Configuration Read Type 0
CfgWr0	010	0 0 1 0 0	Configuration Write Type 0
CfgRd1	000	0 0 1 0 1	Configuration Read Type 1
CfgWr1	010	0 0 1 0 1	Configuration Write Type 1

# Configuration

## ■ Enumeration – Discovering the Topology

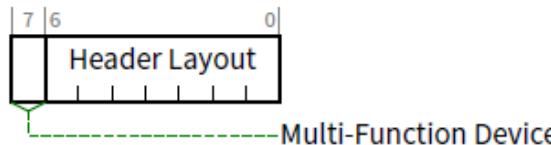
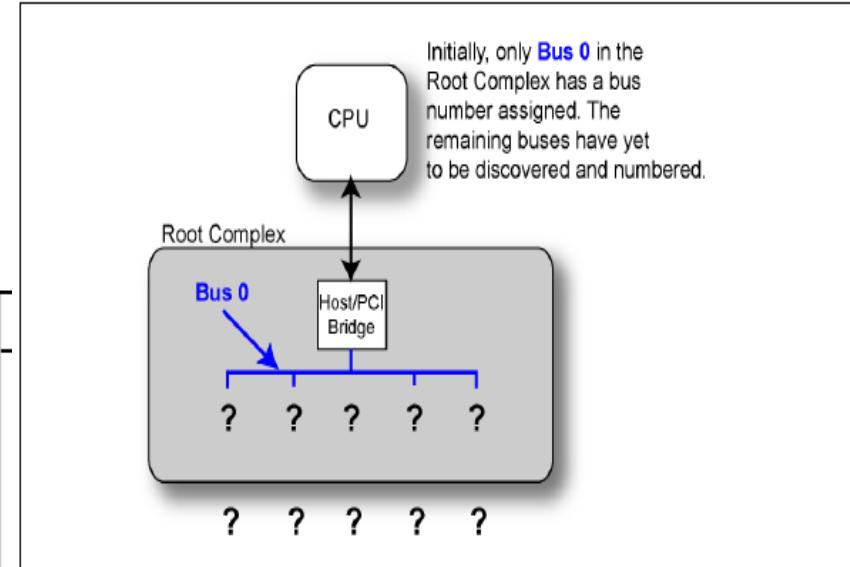


Figure 7-8 Header Type Register

Table 7-6 Header Type Register

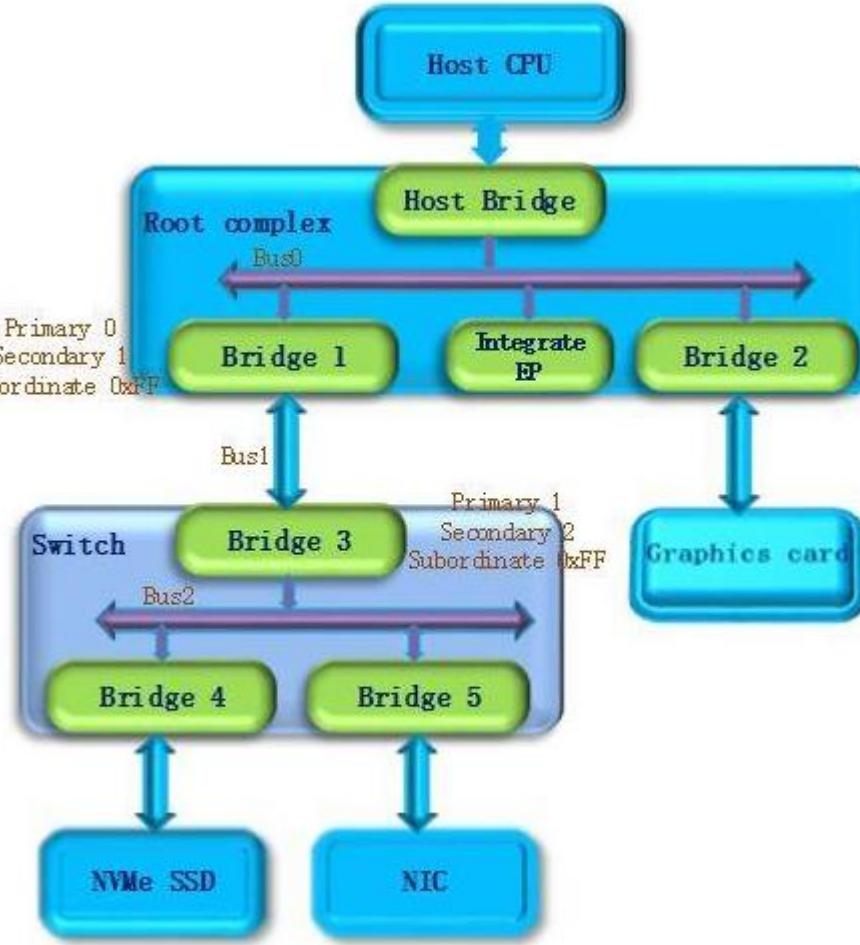
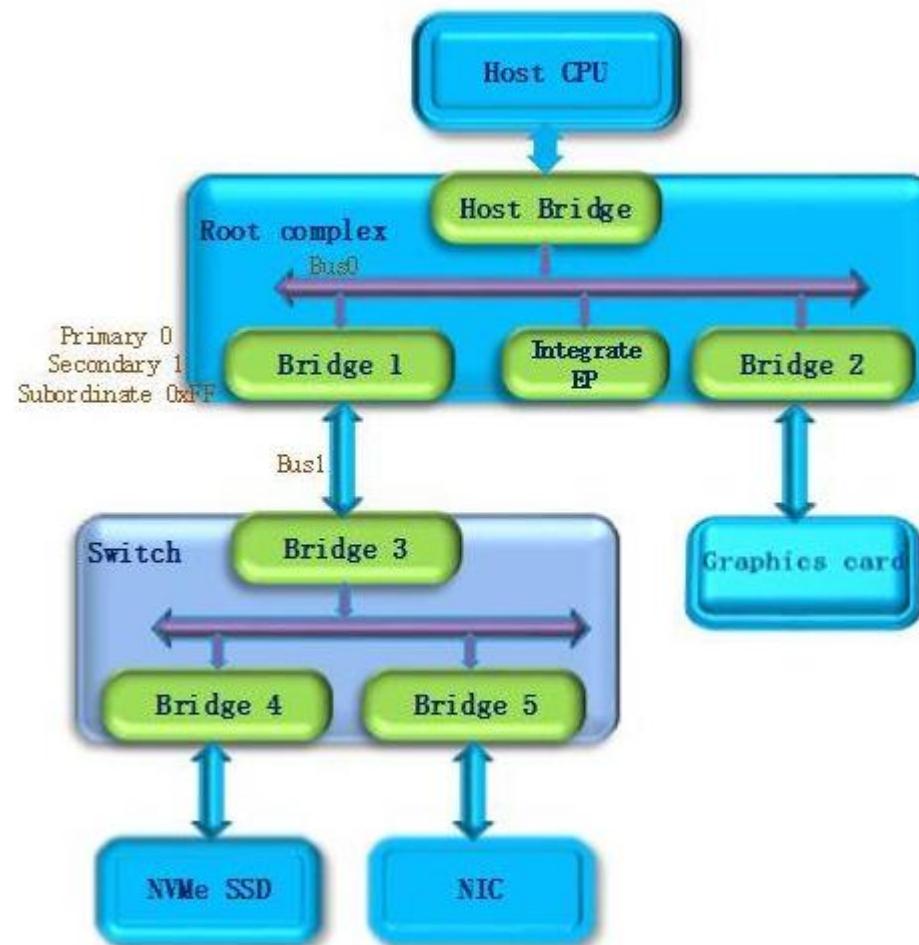
Bit Location	Register Description
6:0	<p><b>Header Layout</b> - This field identifies the layout of the second part of the predefined header.</p> <p>For Functions that implement a <a href="#">Type 0 Configuration Space Header</a> the encoding 000 0000b must be used.</p> <p>For Functions that implement a <a href="#">Type 1 Configuration Space Header</a> the encoding 000 0001b must be used.</p>

Figure 19-2: Topology View At Startup



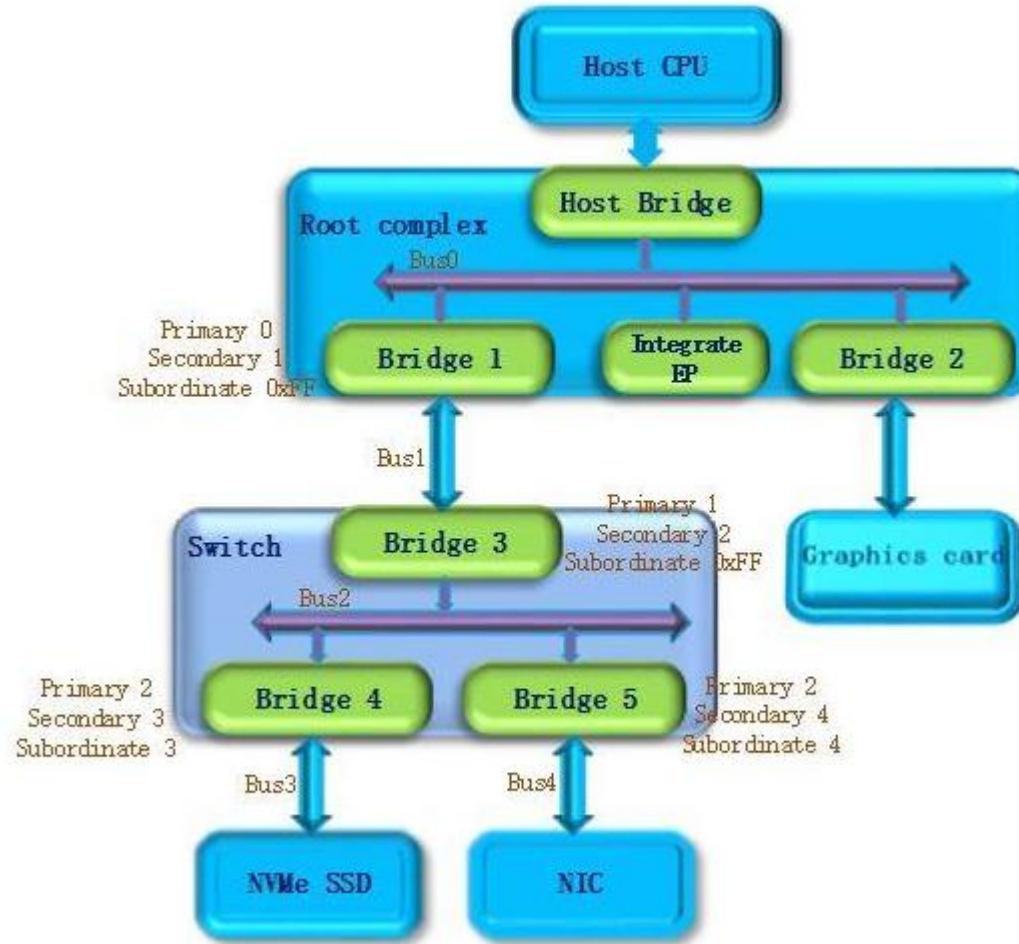
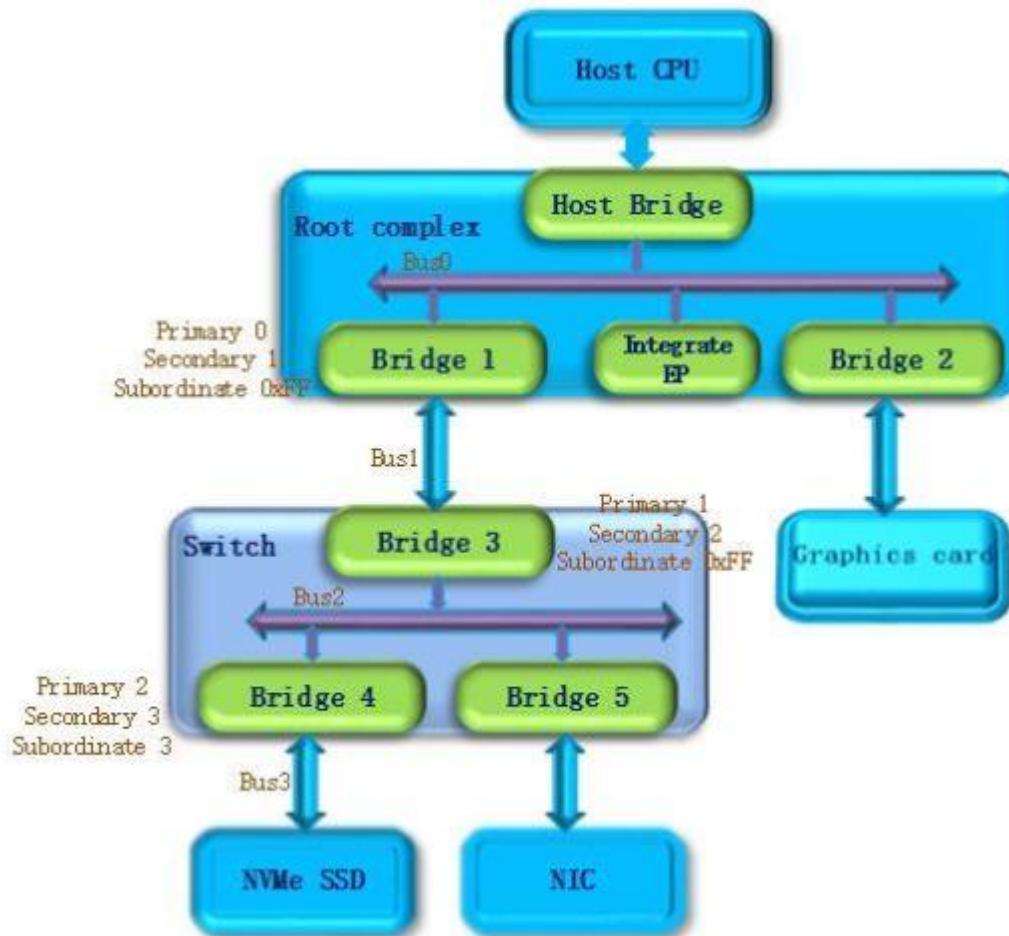
# Configuration

## ■ Enumeration Example



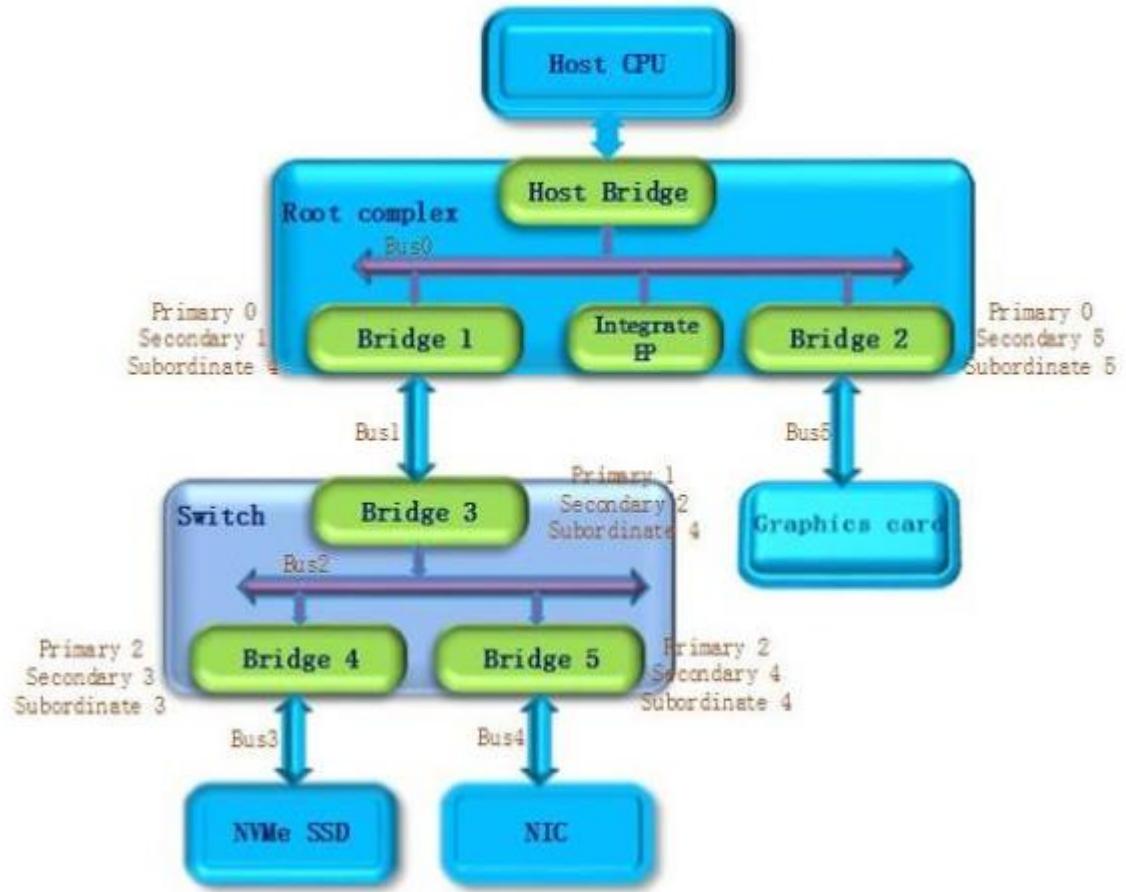
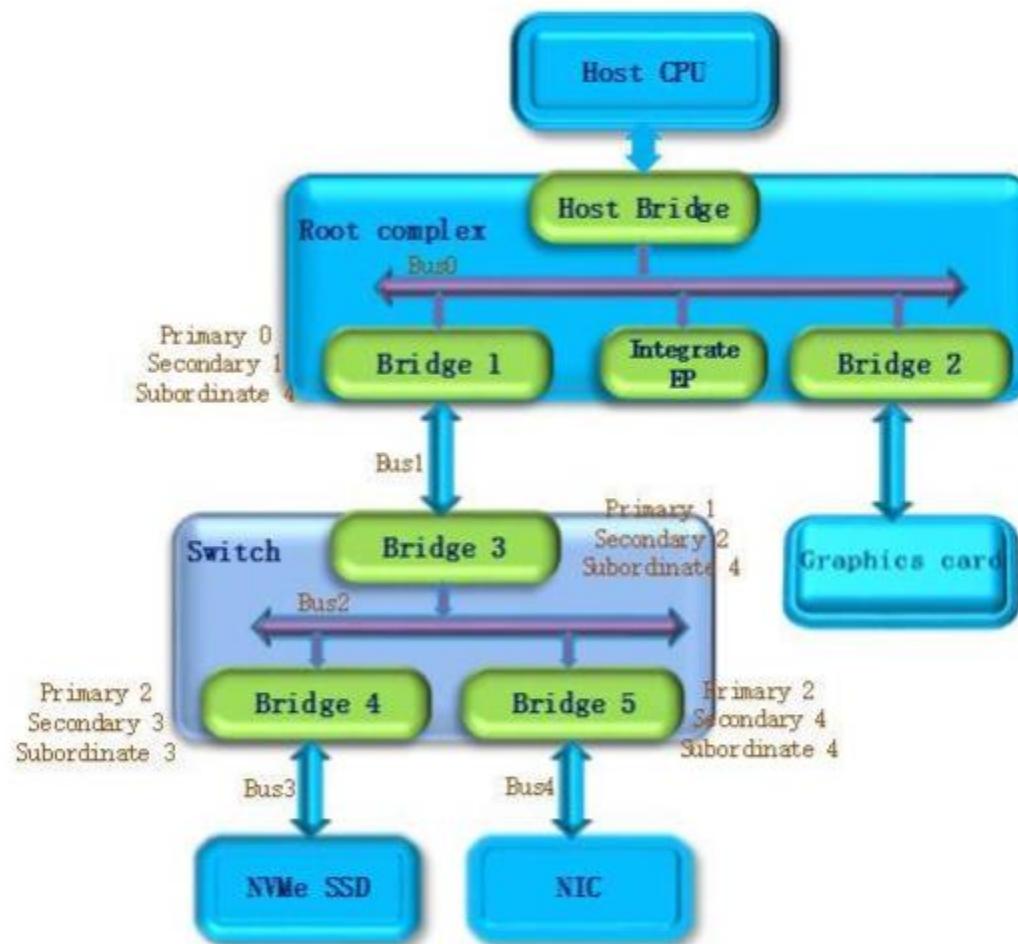
# Configuration

## ■ Enumeration Example



# Configuration

## ■ Enumeration Example



# PCI Express

Address Space & Transaction Routing



# Address Space & Transaction Routing

## ■ Base Address Registers (BARs)

31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0																									
Device ID															Vendor ID																																									
Status															Command																																									
Class Code															Revision ID																																									
BIST	Header Type		Latency Timer		Cache Line Size																																																			
Base Address Registers																																																								
Cardbus CIS Pointer																																																								
Subsystem ID															Subsystem Vendor ID																																									
Expansion ROM Base Address															Capabilities Pointer																																									
Reserved																																																								
Max_Lat	Min_Gnt		Interrupt Pin		Interrupt Line																																																			

Figure 7-10 Type 0 Configuration Space Header

31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0																									
Device ID															Vendor ID																																									
Status															Command																																									
Class Code															Revision ID																																									
BIST	Header Type		Latency Timer		Cache Line Size																																																			
Base Address Register 0																																																								
Base Address Register 1																																																								
Secondary Latency Timer	Subordinate Bus Number		Secondary Bus Number		Primary Bus Number										Secondary Status																																									
Memory Limit															Memory Base																																									
Prefetchable Memory Limit															Prefetchable Memory Base																																									
Prefetchable Memory Base Upper 32 Bits															Prefetchable Memory Limit Upper 32 Bits																																									
I/O Base Limit 16 Bits															U/O Base Upper 16 Bits																																									
Reserved															Capabilities Pointer																																									
Expansion ROM Base Address																																																								
Bridge Control															Interrupt Pin																																									

Figure 7-14 Type 1 Configuration Space Header

# Address Space & Transaction Routing

## ■ Base Address Registers (BARs)

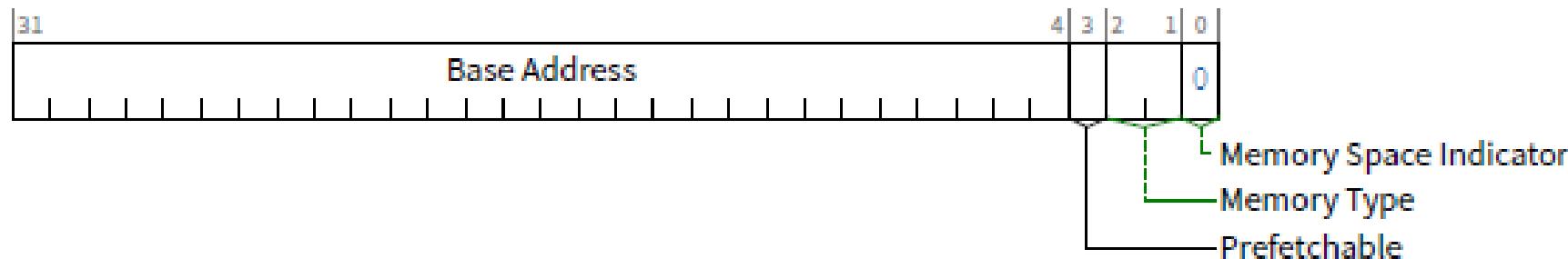


Figure 7-11 Base Address Register for Memory

Table 7-8 Memory Base Address Register Bits 2:1 Encoding

Bits 2:1(b)	Meaning
00	Base register is 32 bits wide and can be mapped anywhere in the 32 address bit Memory Space.
01	Reserved <sup>140</sup>
10	Base register is 64 bits wide and can be mapped anywhere in the 64 address bit Memory Space.
11	Reserved

# Address Space & Transaction Routing

## TLP Routing Basics

- Requests are translated to one of four transaction types by the Transaction Layer:
  - Memory Read or Memory Write: Used to transfer data from or to a memory mapped location.
    - The protocol also supports a locked memory read transaction variant
  - I/O Read or I/O Write: Used to transfer data from or to an I/O location.
    - These transactions are restricted to supporting legacy endpoint devices
  - Configuration Read or Configuration Write: Used to discover device capabilities, program features, and check status in the 4KB PCI Express configuration space.
  - Messages: Handled like posted writes. Used for event signaling and general purpose messaging.

*Table 2-1 Transaction Types for Different Address Spaces*

Address Space	Transaction Types	Basic Usage
Memory	Read Write	Transfer data to/from a memory-mapped location
I/O	Read Write	Transfer data to/from an I/O-mapped location
Configuration	Read Write	Device Function configuration/setup
Message	Baseline (including Vendor-Defined)	From event signaling mechanism to general purpose messaging

# Address Space & Transaction Routing

## ■ TLP Routing Basics

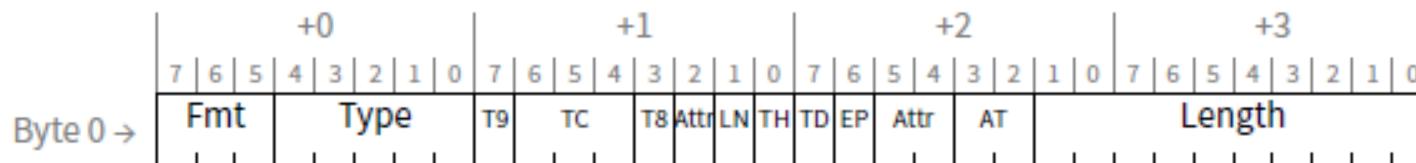


Figure 2-5 Fields Present in All TLP Headers

Table 2-2 Fmt[2:0] Field Values

Fmt[2:0]	Corresponding TLP Format
000b	3 DW header, no data
001b	4 DW header, no data
010b	3 DW header, with data
011b	4 DW header, with data
100b	TLP Prefix
	All encodings not shown above are Reserved (see Section 2.3).

Table 2-4 Length[9:0] Field Encoding

Length[9:0]	Corresponding TLP Data Payload Size
00 0000 0001b	1 DW
00 0000 0010b	2 DW
...	...
11 1111 1111b	1023 DW
00 0000 0000b	1024 DW

# Address Space & Transaction Routing

## TLP Routing Basics

Table 2-3 Fmt[2:0] and Type[4:0] Field Encodings

TLP Type	Fmt [2:0] <sup>3</sup> (b)	Type [4:0] (b)	Description
MRd	000 001	0 0000	Memory Read Request
MRdLk	000 001	0 0001	Memory Read Request-Locked
MWr	010 011	0 0000	Memory Write Request
IORd	000	0 0010	I/O Read Request
IOWr	010	0 0010	I/O Write Request
CfgRd0	000	0 0100	Configuration Read Type 0
CfgWr0	010	0 0100	Configuration Write Type 0
CfgRd1	000	0 0101	Configuration Read Type 1
CfgWr1	010	0 0101	Configuration Write Type 1

Table 2-3 Fmt[2:0] and Type[4:0] Field Encodings

TLP Type	Fmt [2:0] <sup>3</sup> (b)	Type [4:0] (b)	Description
Msg	001	1 0r2r1r0	Message Request - The sub-field r[2:0] specifies the Message routing mechanism (see <a href="#">Table 2-17</a> ).
MsgD	011	1 0r2r1r0	Message Request with data payload - The sub-field r[2:0] specifies the Message routing mechanism (see <a href="#">Table 2-17</a> ).
Cpl	000	0 1010	Completion without Data - Used for I/O and Configuration Write Completions with any Completion Status. Also used for AtomicOp Completions and Read Completions (I/O, Configuration, or Memory) with Completion Status other than Successful Completion.
CplD	010	0 1010	Completion with Data - Used for Memory, I/O, and Configuration Read Completions. Also used for AtomicOp Completions.
CplLk	000	0 1011	Completion for Locked Memory Read without Data - Used only in error case.
CplDLk	010	0 1011	Completion for Locked Memory Read - Otherwise like CplD.
FetchAdd	010 011	0 1100	Fetch and Add AtomicOp Request
Swap	010 011	0 1101	Unconditional Swap AtomicOp Request
CAS	010 011	0 1110	Compare and Swap AtomicOp Request
LPrfx	100	0 L <sub>3</sub> L <sub>2</sub> L <sub>1</sub> L <sub>0</sub>	Local TLP Prefix - The sub-field L[3:0] specifies the Local TLP Prefix type (see <a href="#">Table 2-36</a> ).
EPrfx	100	1 E <sub>3</sub> E <sub>2</sub> E <sub>1</sub> E <sub>0</sub>	End-End TLP Prefix - The sub-field E[3:0] specifies the End-End TLP Prefix type (see <a href="#">Table 2-37</a> ).

# Address Space & Transaction Routing

## ■ Applying Routing Mechanisms

- Each request or completion header is tagged as to its type, and each of the packet types is routed based on one of three schemes:
  - Address Routing
  - ID Routing
  - Implicit Routing
- Memory and IO requests use address routing
- Completions and Configuration cycles use ID routing
- Message requests have selectable routing based on a 3-bit code in the message routing sub-field of the header type field

# Address Space & Transaction Routing

## ■ Applying Routing Mechanisms

### □ Address Routing

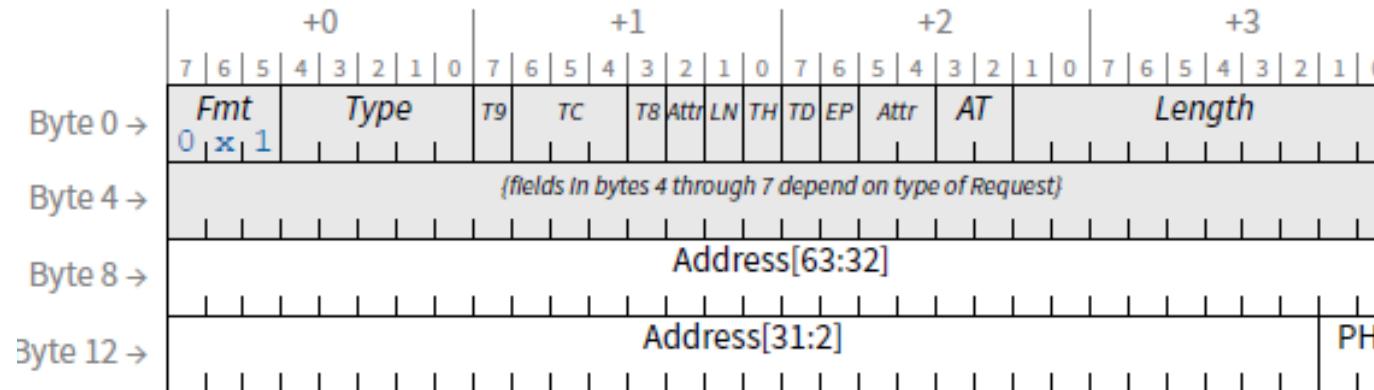


Figure 2-7 64-bit Address Routing

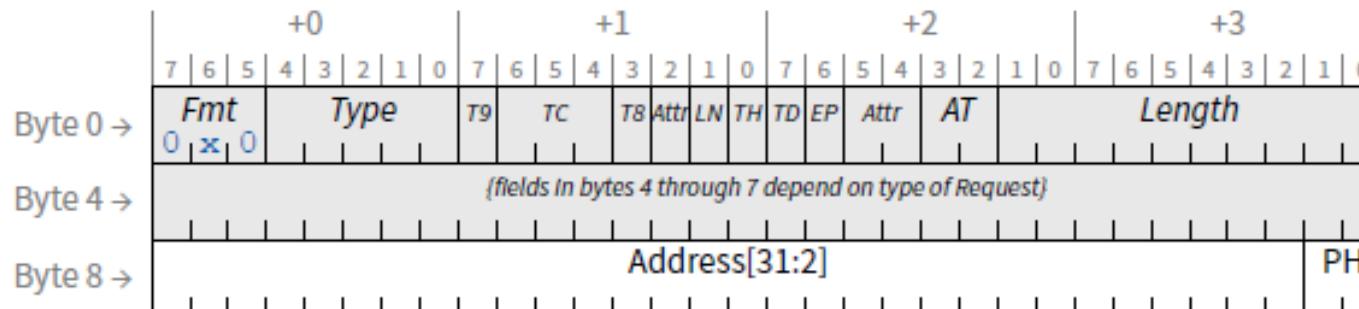


Figure 2-8 32-bit Address Routing

# Address Space & Transaction Routing

## ■ Applying Routing Mechanisms

### □ ID Routing

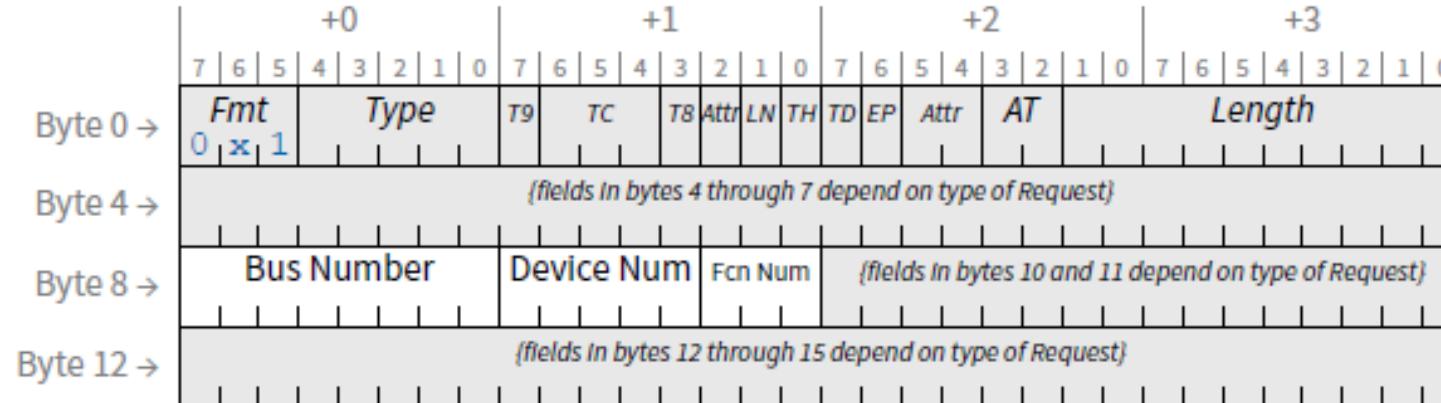


Figure 2-9 Non-ARI ID Routing with 4 DW Header

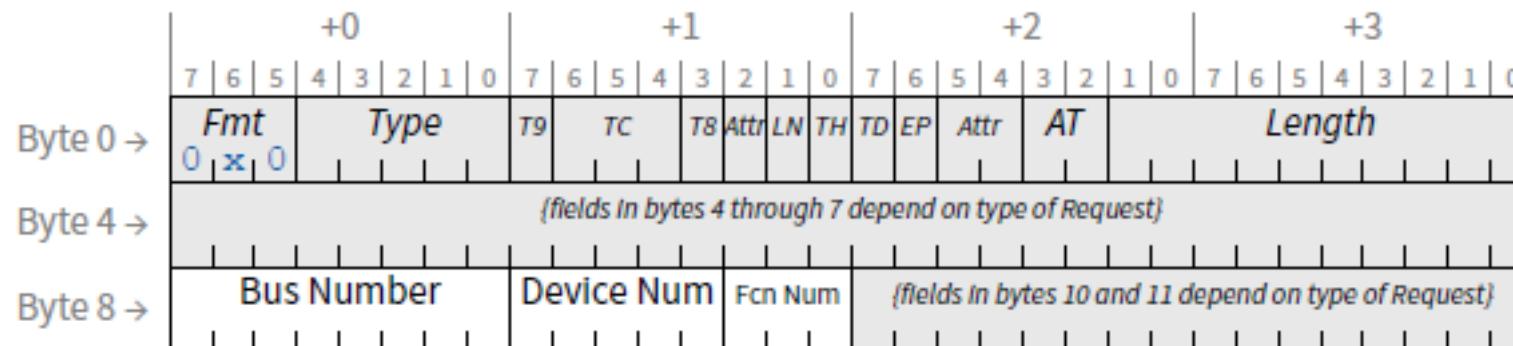


Figure 2-11 Non-ARI ID Routing with 3 DW Header

# Address Space & Transaction Routing

## ■ Applying Routing Mechanisms

### □ Implicit Routing

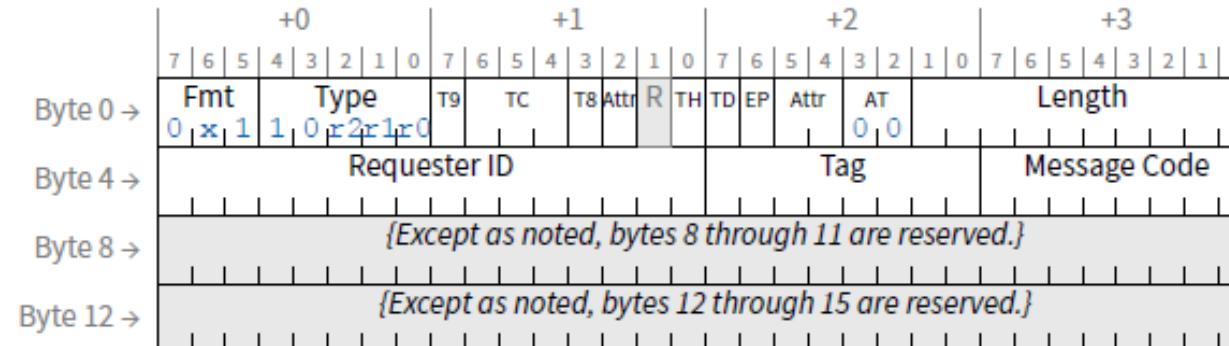


Figure 2-26 Message Request Header  
Table 2-17 Message Routing

r[2:0] (b)	Description	Bytes 8 to 15 <sup>15</sup>
000	Routed to Root Complex	Reserved
001	Routed by Address <sup>16</sup>	Address
010	Routed by ID	See <a href="#">Section 2.2.4.2</a>
011	Broadcast from Root Complex	Reserved
100	Local - Terminate at Receiver	Reserved
101	Gathered and routed to Root Complex <sup>17</sup>	Reserved
110 to 111	Reserved - Terminate at Receiver	Reserved

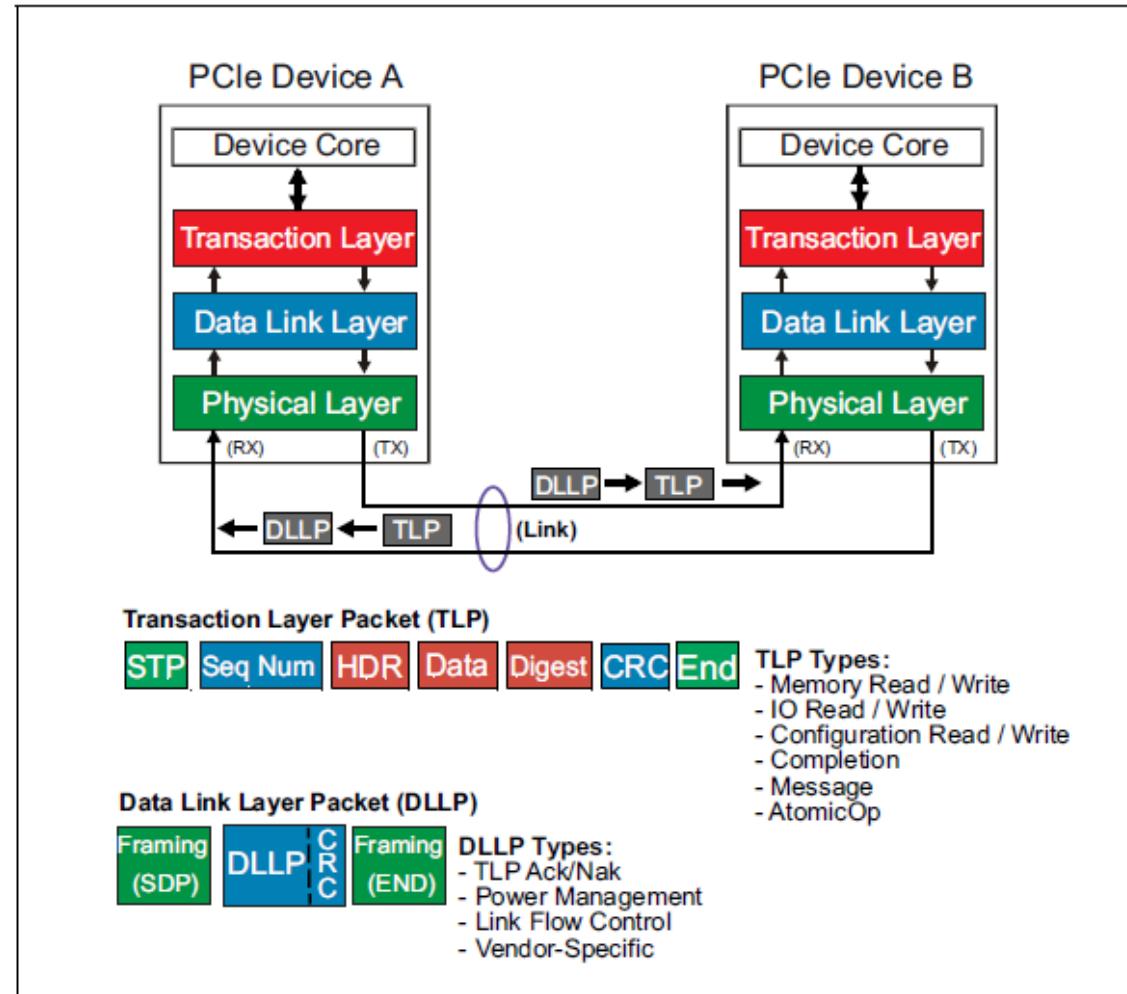
# PCI Express

## TLP & DLLP Elements

# TLP Elements

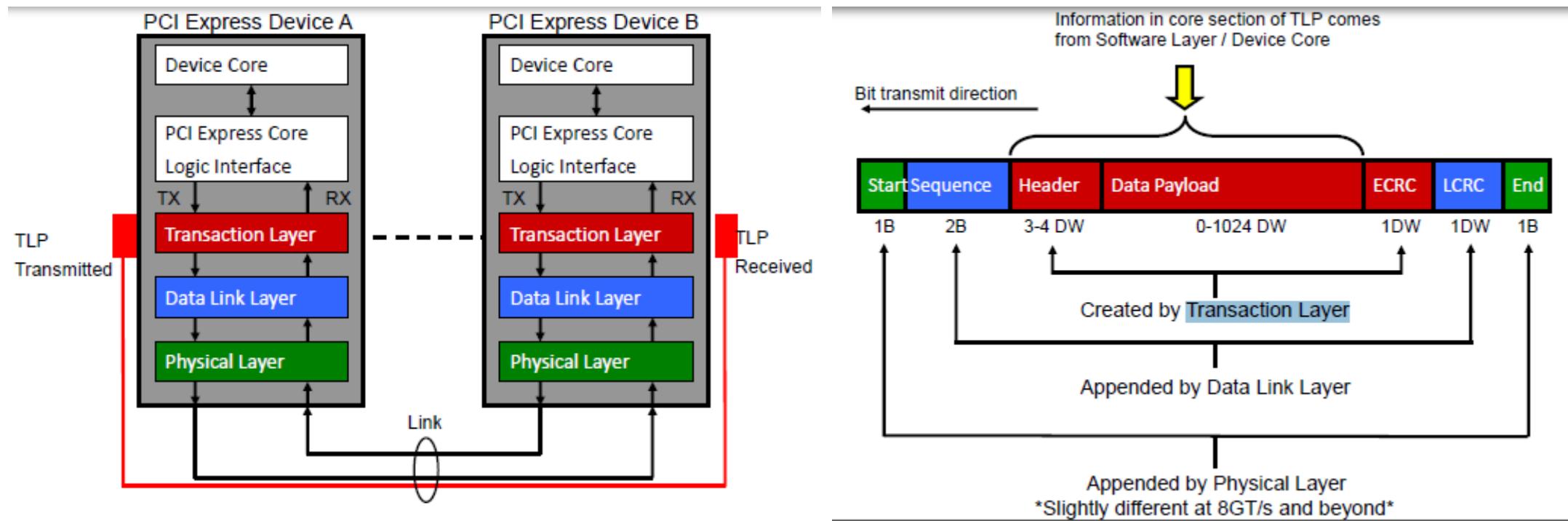
## ■ Packet-Based Protocol

Figure 5-1: TLP And DLLP Packets



# TLP Elements

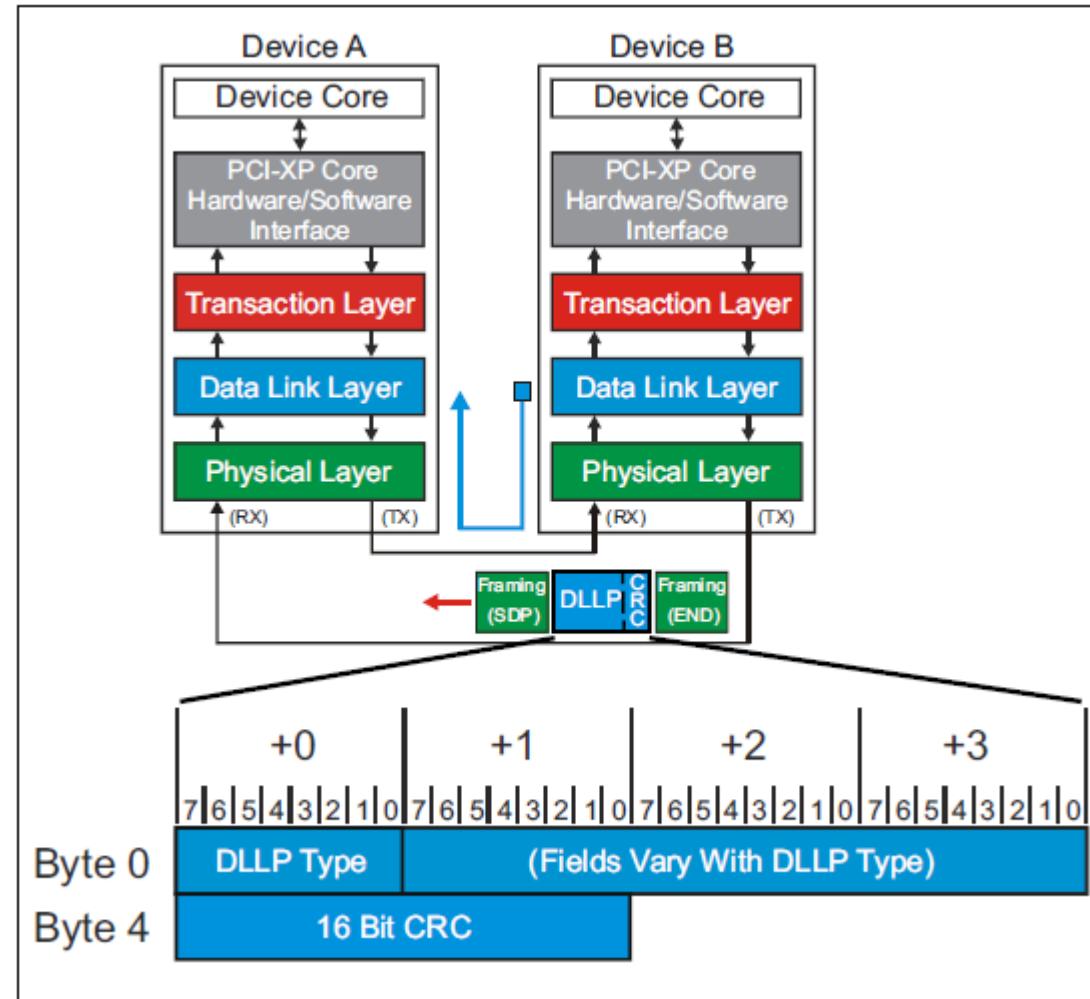
## ■ TLP Assembly And Disassembly



# DLLP Elements

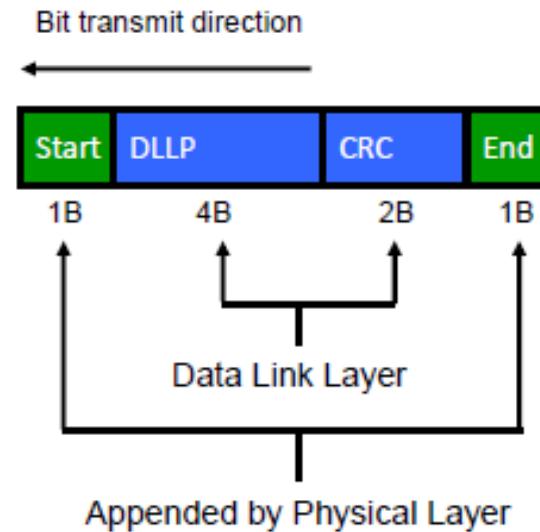
## Sending DLLPs

Figure 9-2: Generic Data Link Layer Packet Format



# DLLP Elements

## DLLP Packet Types



- **ACK / NAK Packets**
- **Flow Control Packets**
- **Power Management Packets**
- **Vendor Defined Packets**

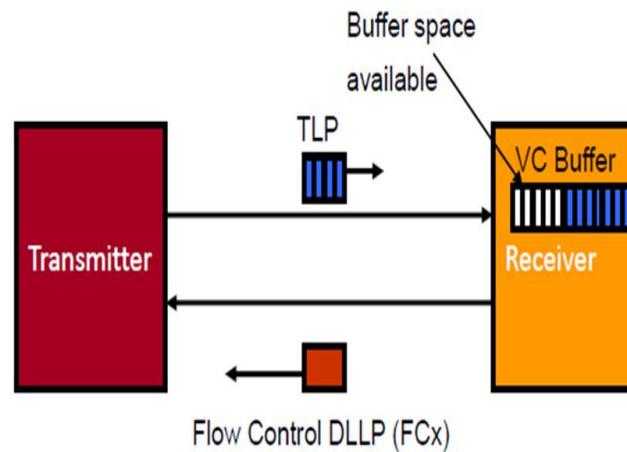
# PCI Express

## Flow Control



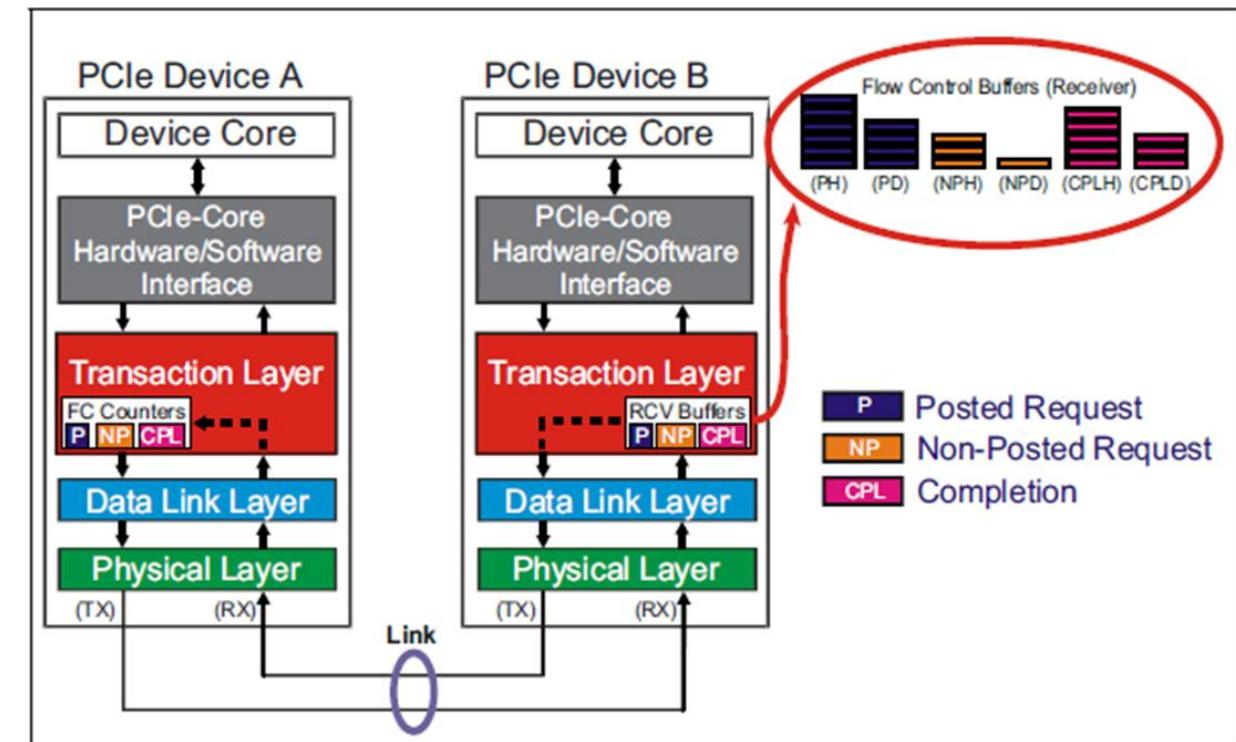
# Flow Control

## Buffers and Credits



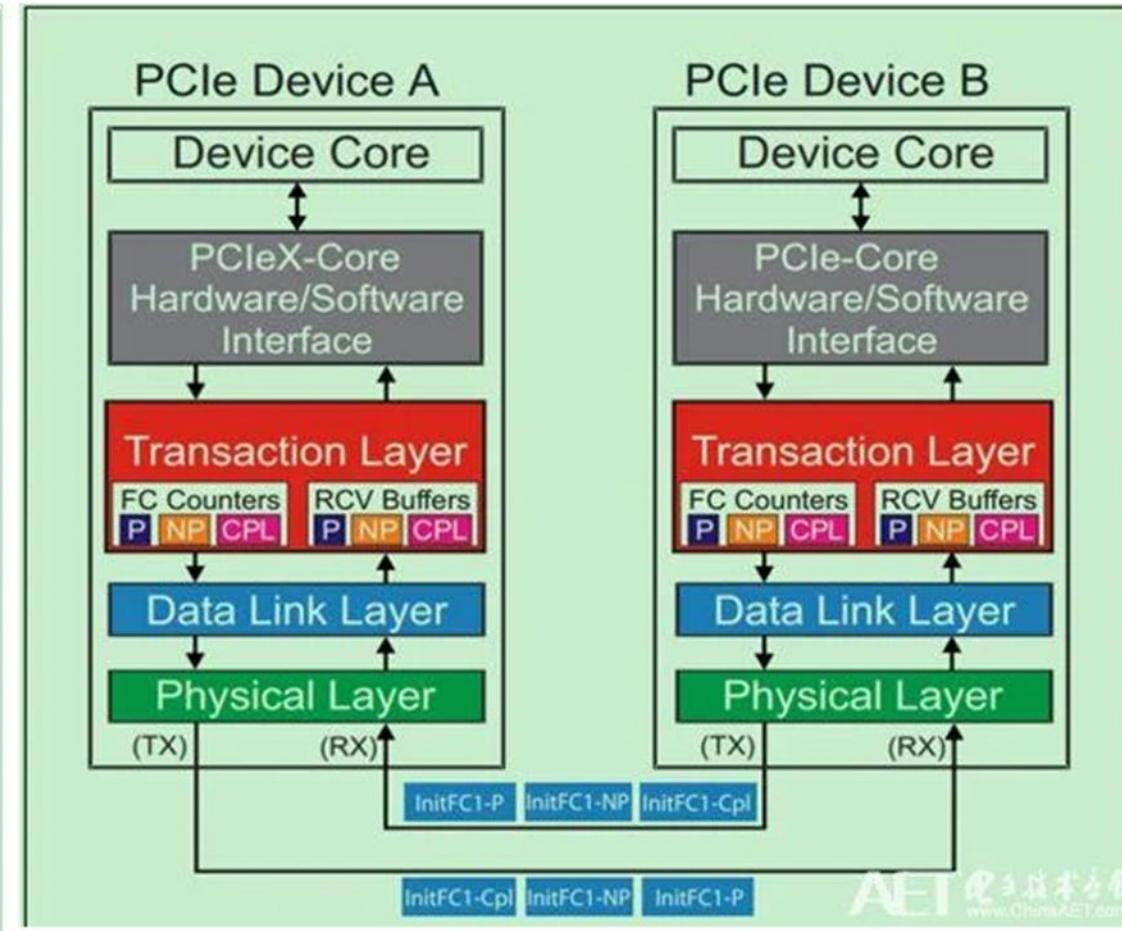
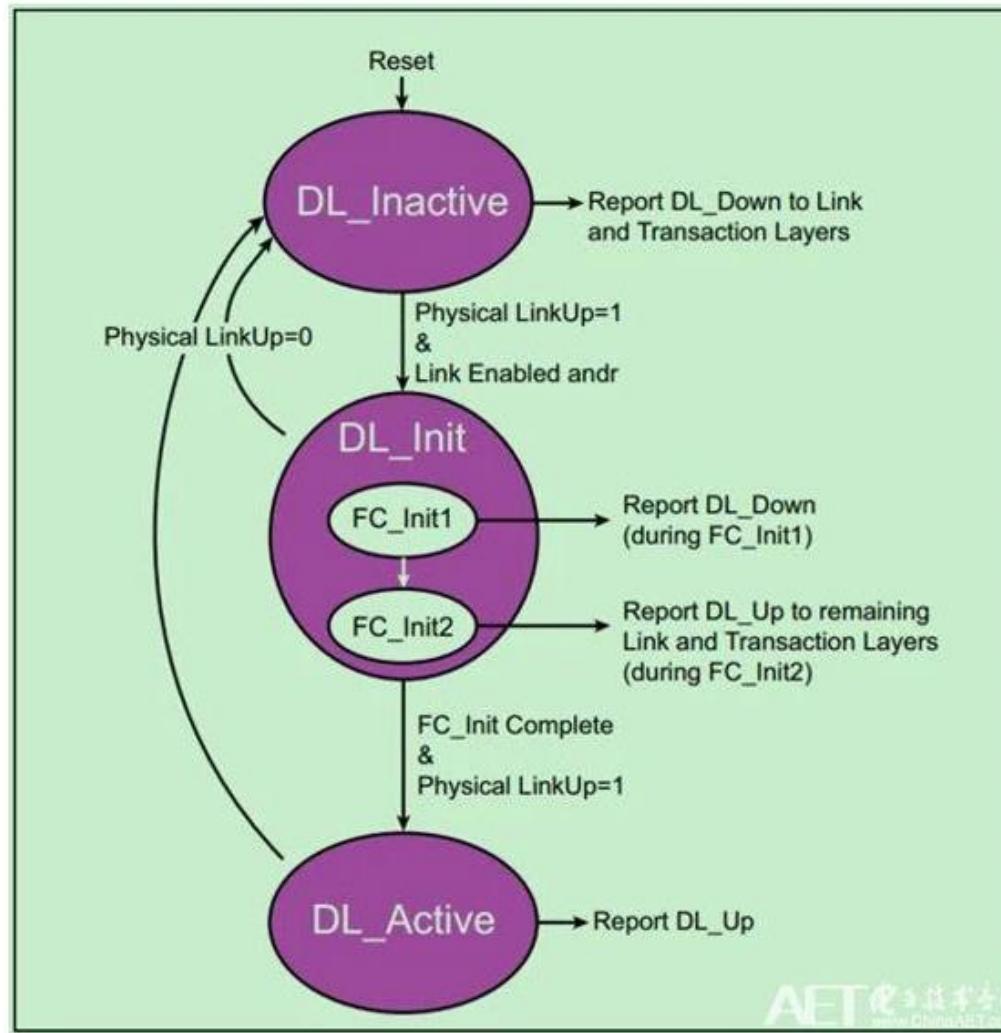
Receiver sends Flow Control Packets (FCP) which are a type of DLLP (Data Link Layer Packet) to provide the transmitter with credits so that it can transmit packets to the receiver

Figure 6-2: Flow Control Buffer Organization



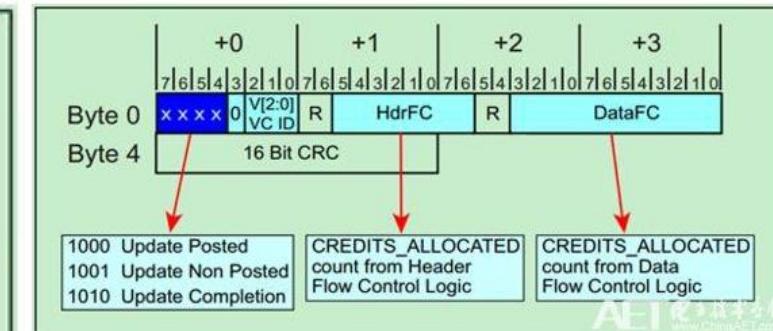
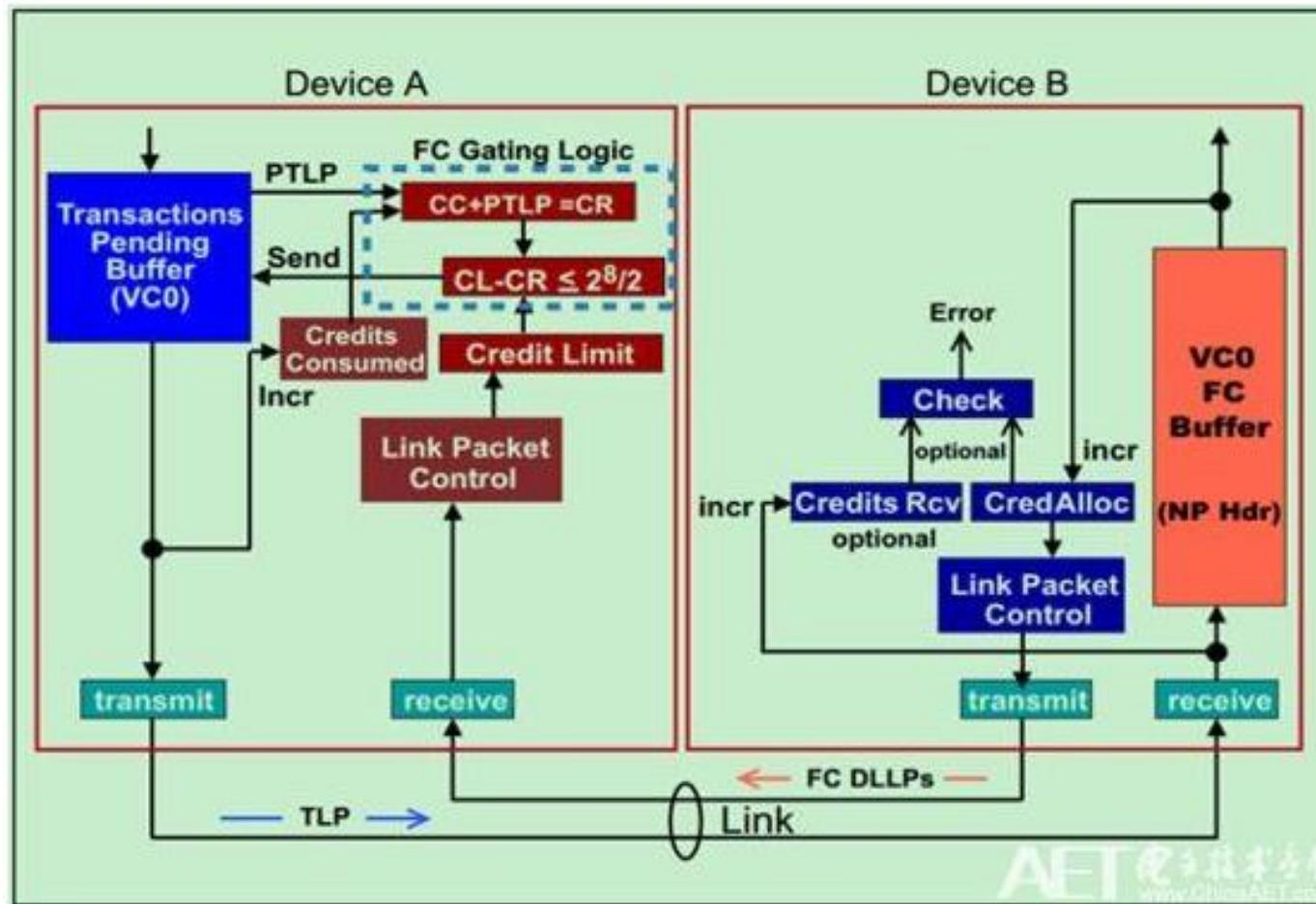
# Flow Control

## ■ FC\_Init1 & FC\_Init2



# Flow Control

## ■ UpdateFC



# PCI Express

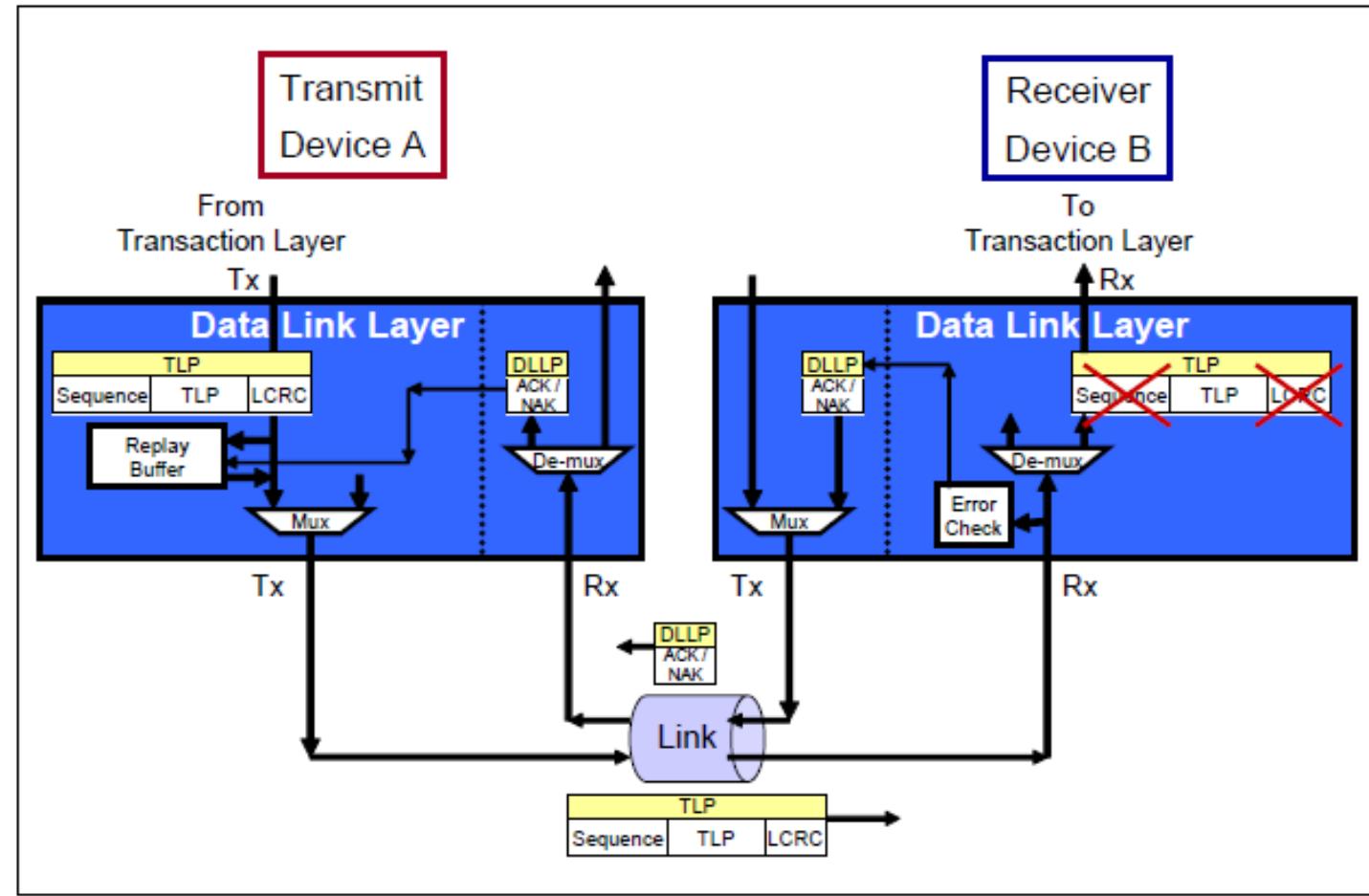
ACK/NAK Protocol



# ACK/NAK Protocol

## ■ Overview

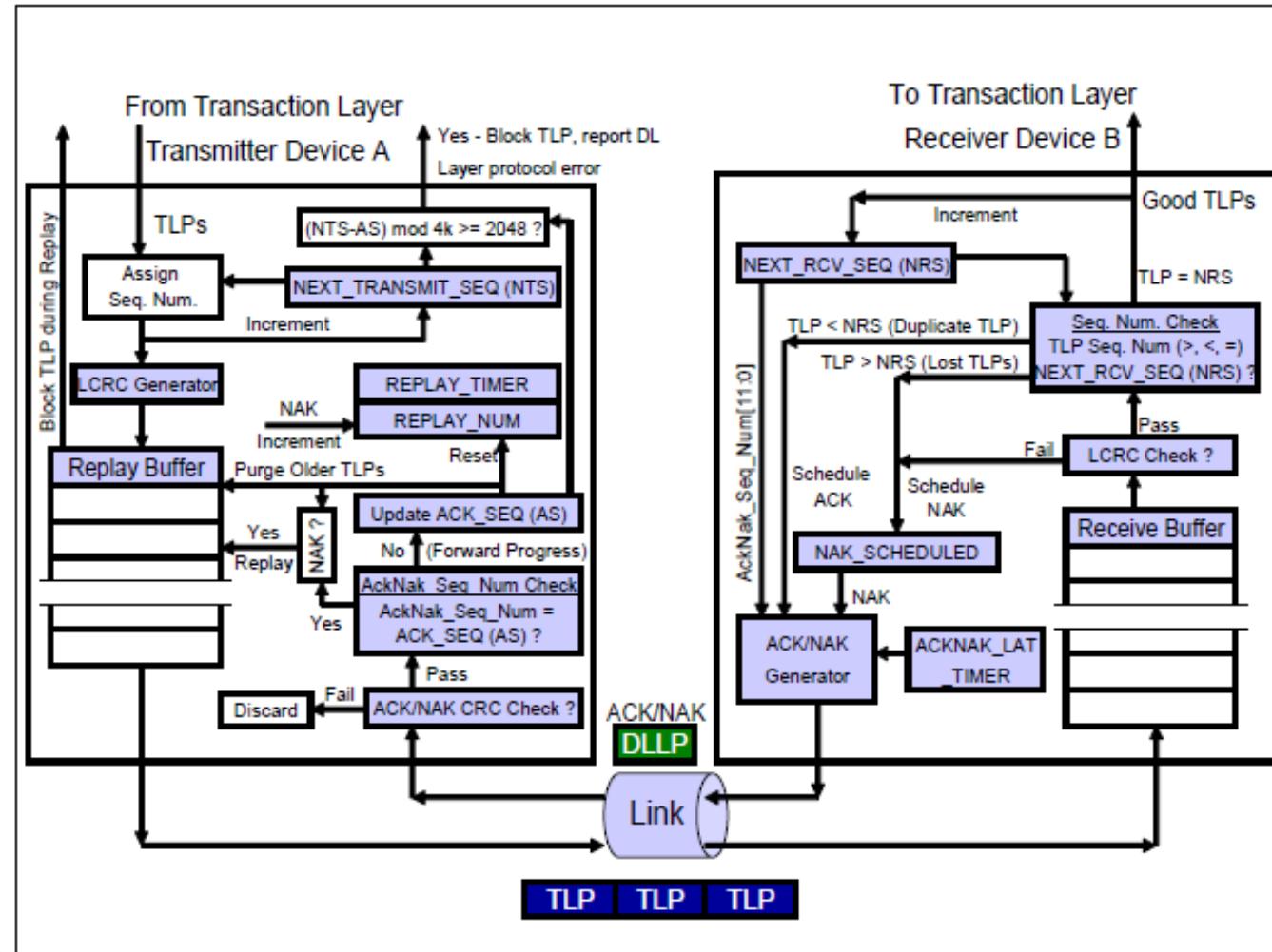
Figure 5-2: Overview of the ACK/NAK Protocol



# ACK/NAK Protocol

## Elements

Figure 5-3: Elements of the ACK/NAK Protocol



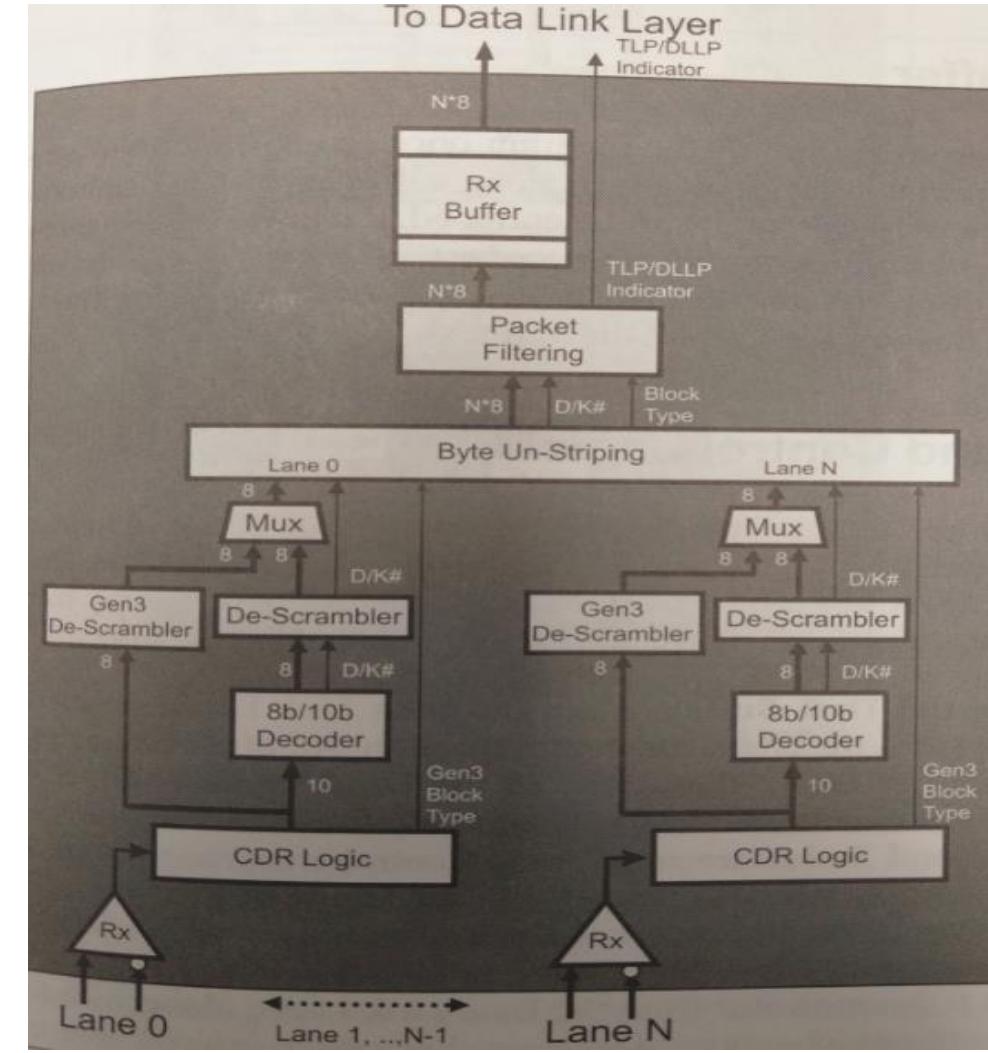
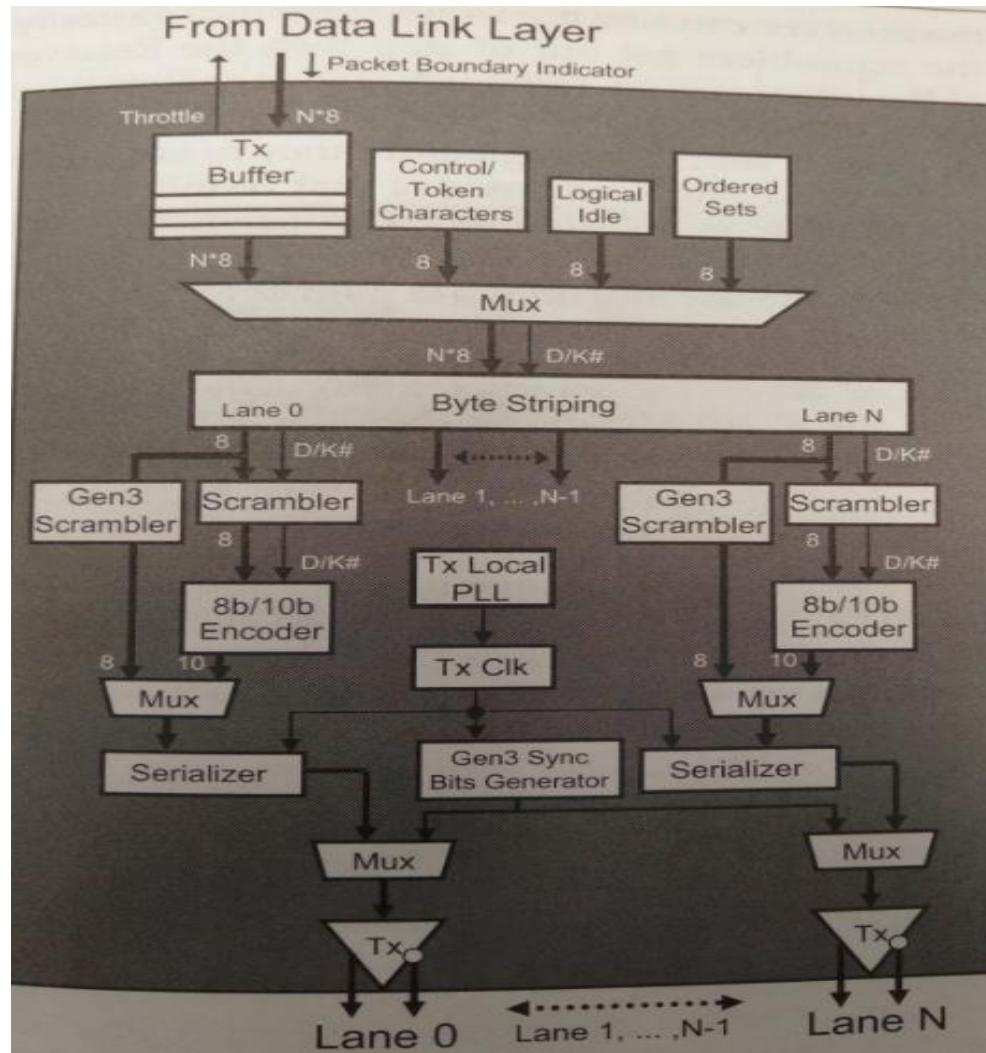
# PCI Express

Physical Layer - Logical



# Physical Layer - Logical

## ■ Transmit / Receive Logic



# Physical Layer - Logical

## ■ 8b/10b & 128b/130b Encoding

Figure 12-1: 8b/10b Lane Encoding

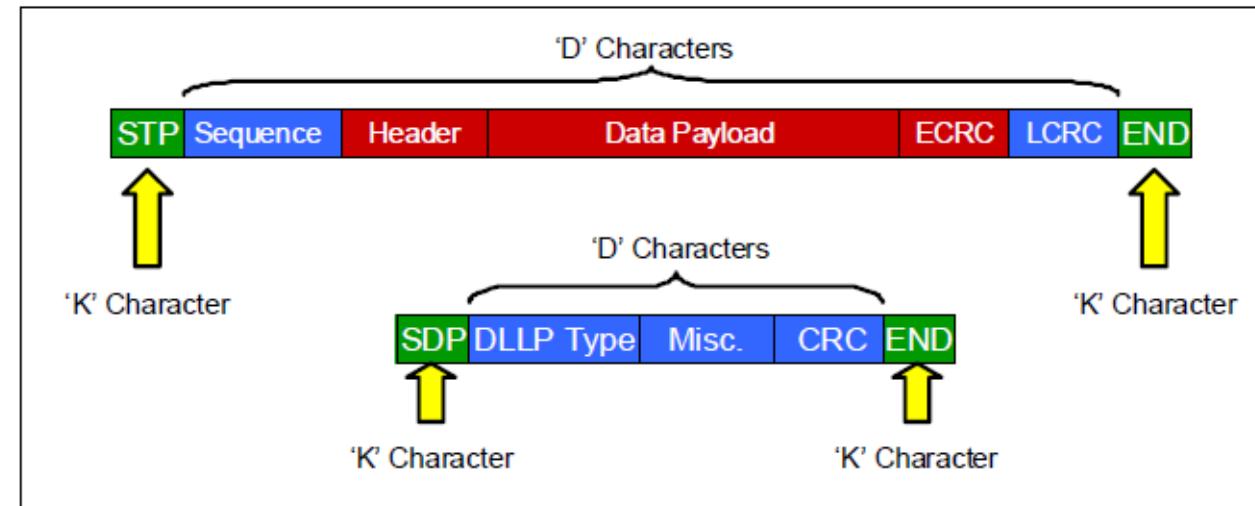
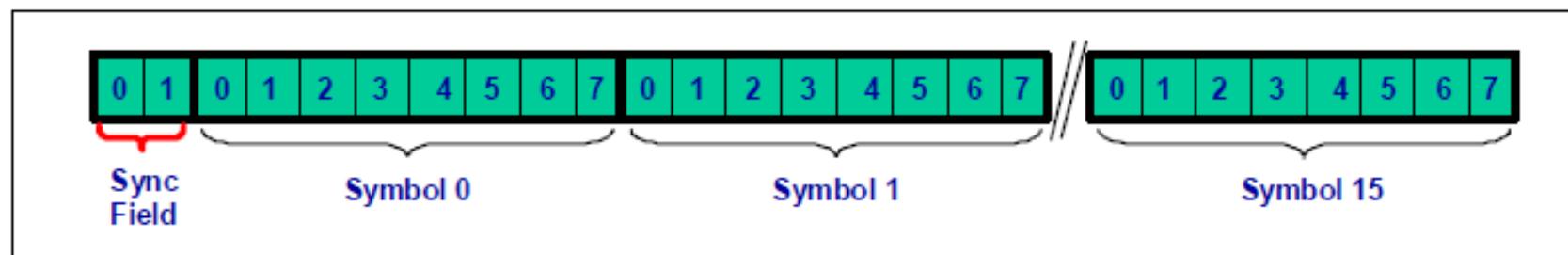


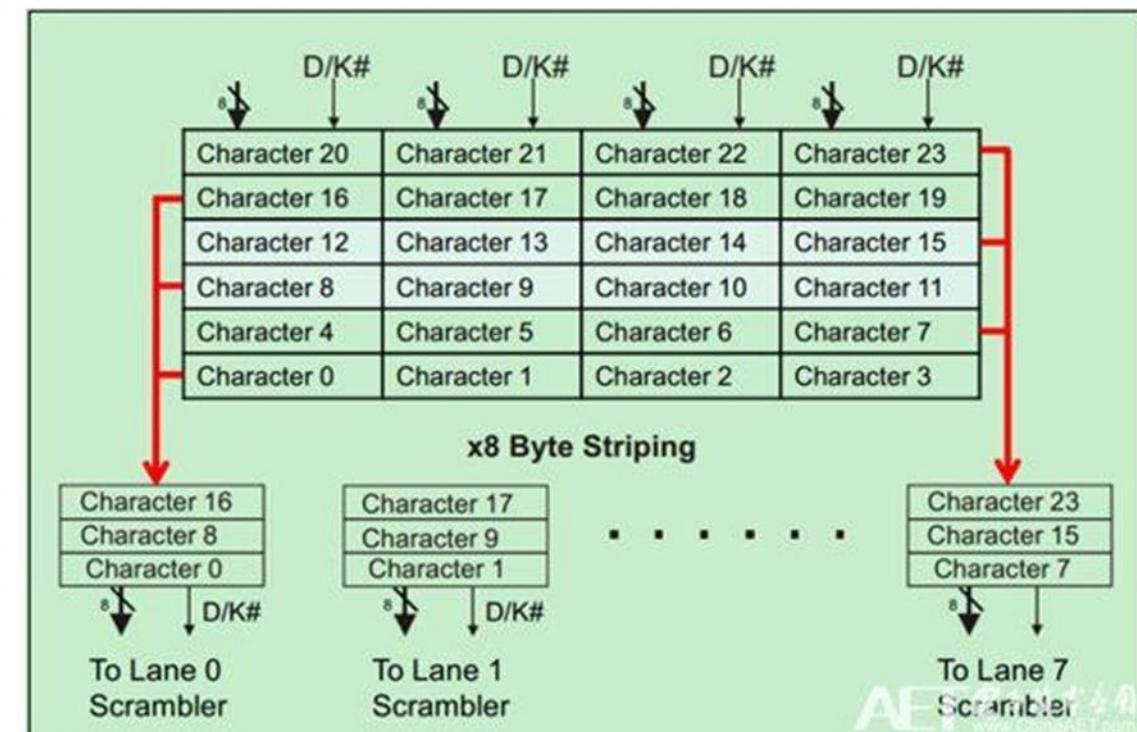
Figure 12-2: 128b/130b Block Encoding



# Physical Layer - Logical

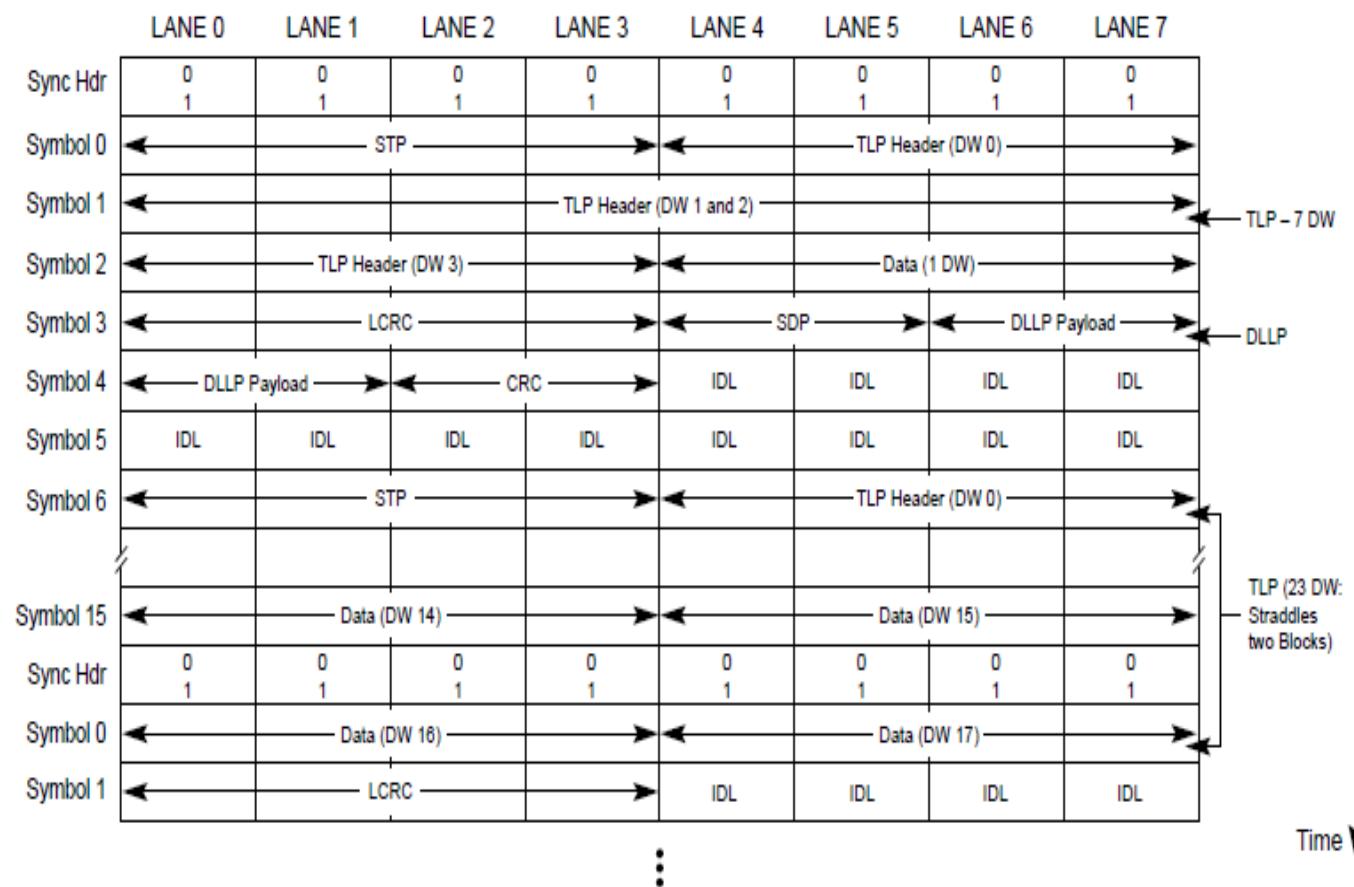
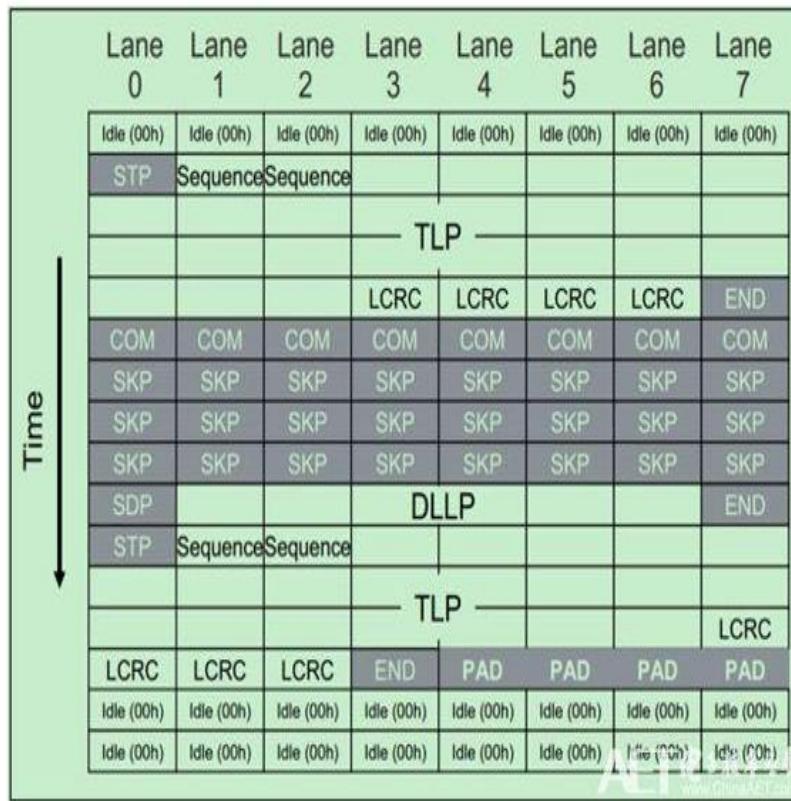
## ■ Byte Striping

Character Name	8b/10b Name	Description
COM	K28.5	First character in any ordered set. Also used by Rx to achieve Symbol lock during training.
PAD	K23.7	Packet filler
SKP	K28.0	Used in SKIP ordered set for Clock Tolerance Compensation
STP	K27.7	Start of a TLP
SDP	K28.2	Start of a DLLP
END	K29.7	End of Good Packet
EDB	K30.7	End of a bad or 'nullified' TLP.
FTS	K28.1	Used to exit from L0s low power state to L0
IDL	K28.3	Used to place Link into Electrical Idle state
EIE	K28.7	Part of the Electrical Idle Exit Ordered Set sent prior to bringing the Link back to full power for speeds higher than 2.5 GT/s



# Physical Layer - Logical

## Packet Format

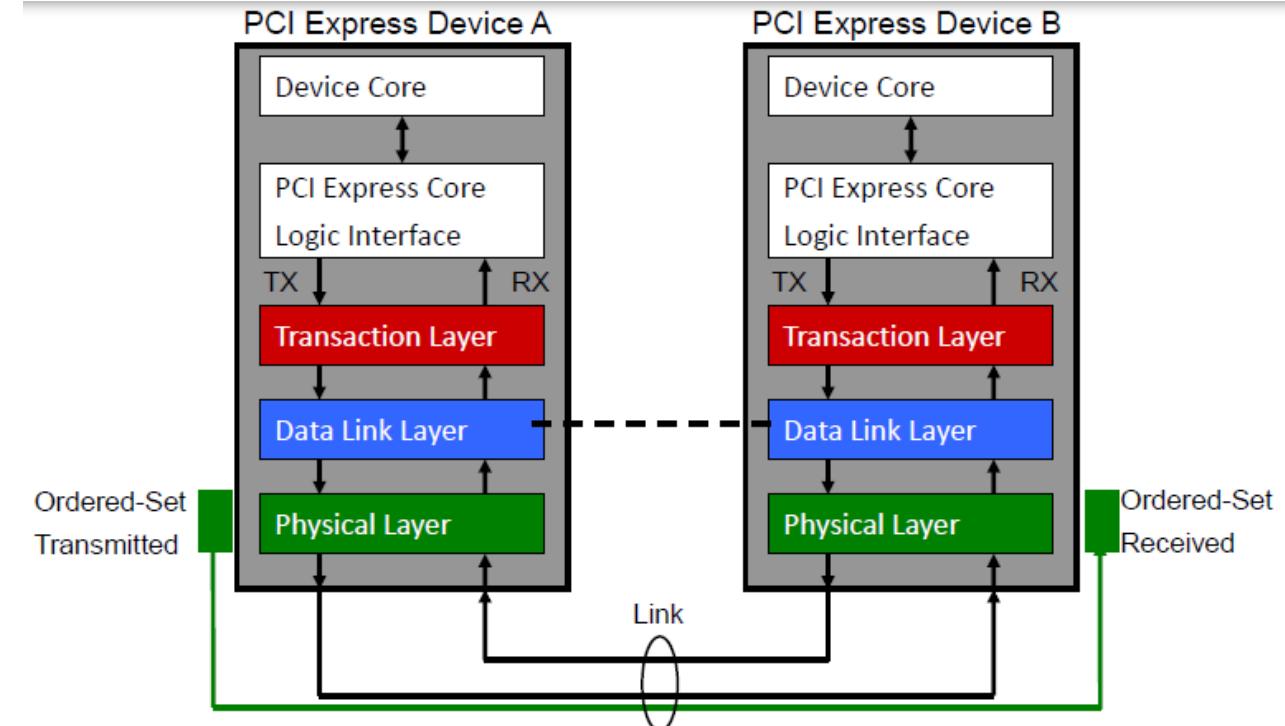


# Physical Layer - Logical

## ■ Ordered-Set

COM Identifier Identifier • • • Identifier

- **Training Sequence One (TS1)**
  - 16 character set: 1 COM, 15 TS1 data characters
- **Training Sequence Two (TS2)**
  - 16 character set: 1 COM, 15 TS2 data characters
- **SKIP**
  - 4 character set: 1 COM followed by 3 SKP identifiers
- **Fast Training Sequence (FTS)**
  - 4 characters: 1 COM followed by 3 FTS identifiers
- **Electrical Idle (IDLE)**
  - 4 characters: 1 COM followed by 3 IDL identifiers
- **Electrical Idle Exit (EIEOS) (new to 2.0 spec)**
  - 16 characters



# PCI Express

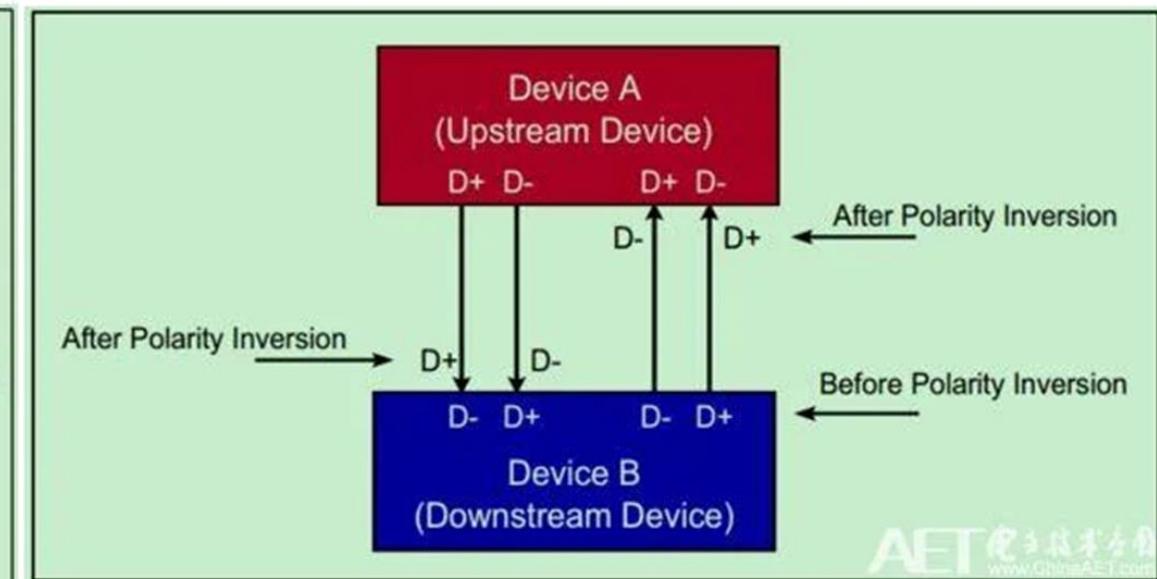
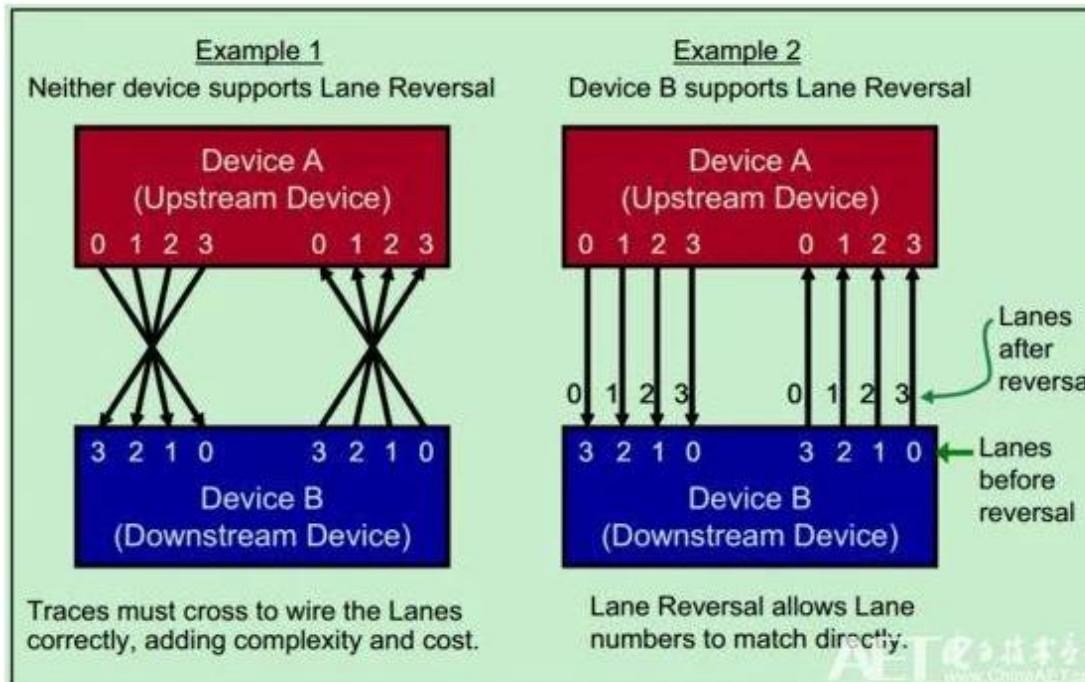
Link Initialization & Training



# Link Initialization & Training

## ■ Overview

- The following are discovered and determined during the training process:
  - Link width (Link width negotiation)
  - Link data rate (Link data rate negotiation)
  - Lane reversal (Lane ordering within a Link)
  - Lane polarity



# Link Initialization & Training

## ■ Training Sequences

- TS1 and TS2 Ordered Sets

*Figure 14-4: TS1 and TS2 Ordered Sets When In Gen1 or Gen2 Mode*

0	COM	K28.5
1	Link #	0 - 255 = D0.0 - D31.7, PAD = K23.7
2	Lane #	0 - 31 = D0.0 - D17.1, PAD = K23.7
3	# FTS	# of FTSs required by Receiver for L0s recovery
4	Rate ID	Bit 1 must be set, indicates 2.5 GT/s support
5	Train Ctl	
6	TS ID or EQ Info	Equalization info when changing to 8.0 GT/s, else TS1 or TS2 Identifier
9		
10	TS ID	TS1 Identifier = D10.2 TS2 Identifier = D5.2
15		

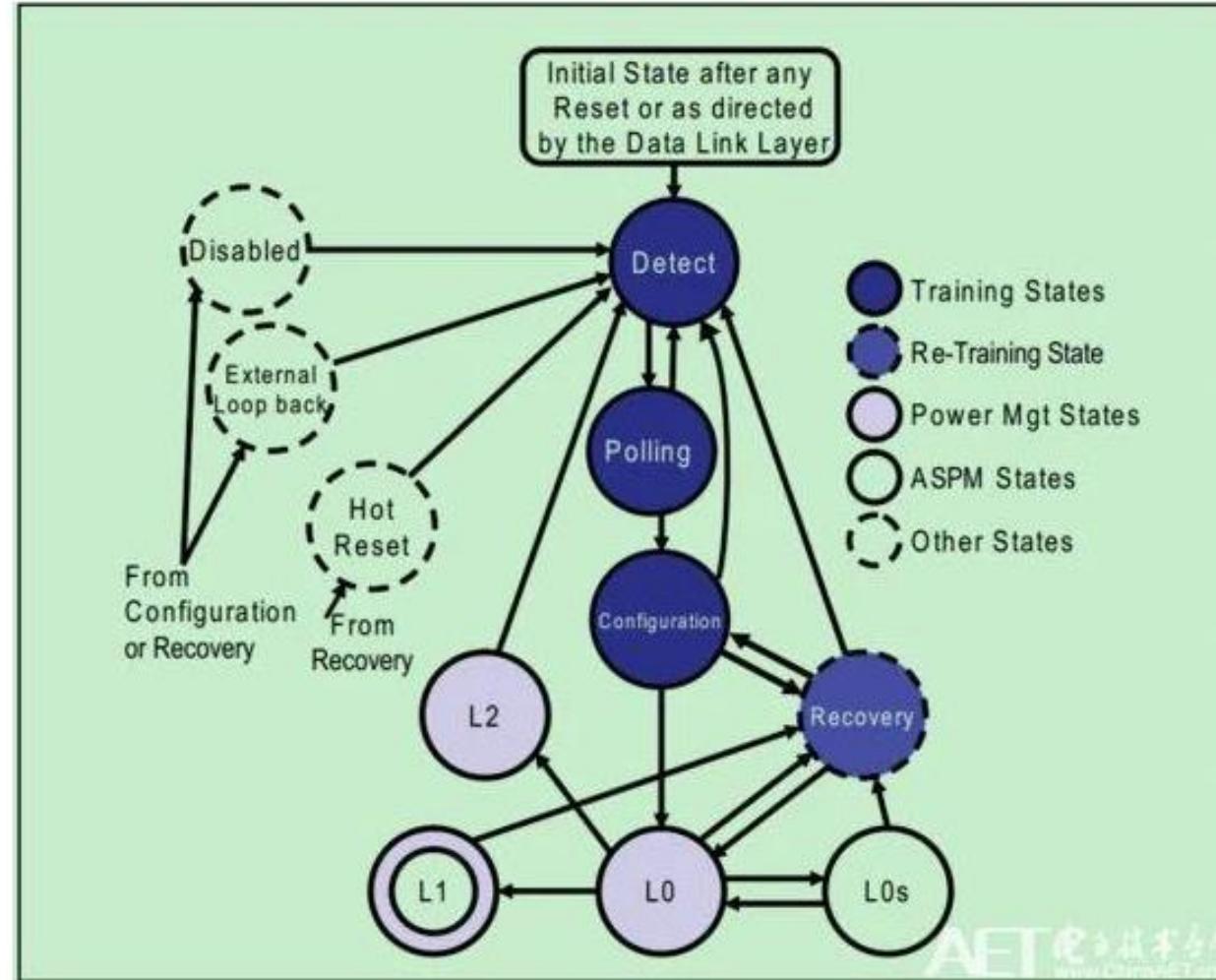
*Figure 14-5: TS1 and TS2 Ordered Set Block When In Gen3 Mode of Operation*

0		TS1 = 1Eh, TS2 = 2Dh
1	Link #	0 - 31, PAD = F7h
2	Lane #	0 - 31, PAD = F7h
3	# FTS	# of FTSs required by Receiver for L0s recovery
4	Rate ID	Bit 3 indicates 8.0 GT/s support
5	Train Ctl	
6	EQ Info	Equalization presets and coefficients or TS2
9		
10	TS ID	TS1 Identifier = 4Ah TS2 Identifier = 45h
13		
14	TS ID	TS1, TS2, or DC Balance Symbols
15		

# Link Initialization & Training

## ■ Link Training and Status State Machine (LTSSM)

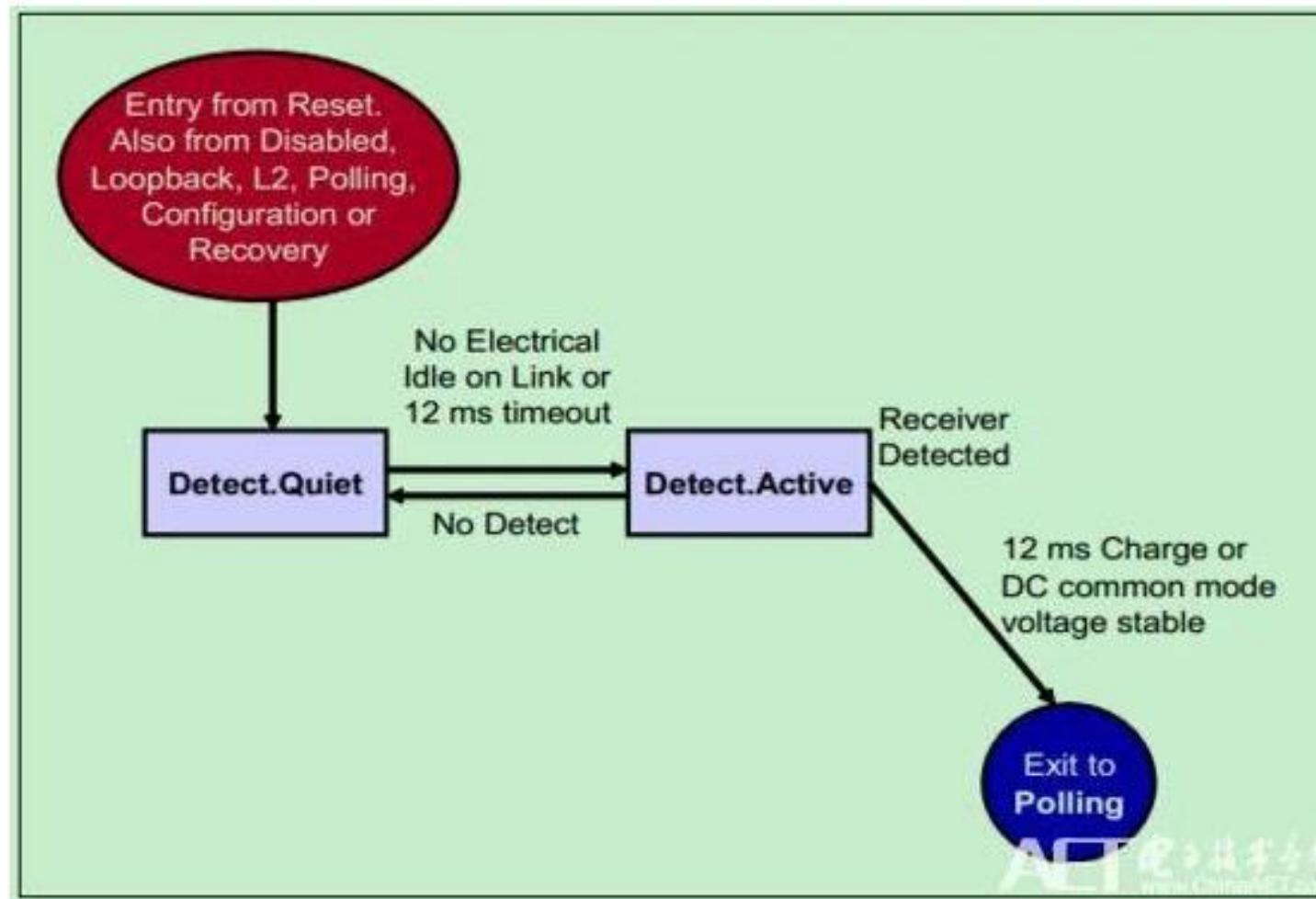
### □ Overview



# Link Initialization & Training

## ■ Link Training and Status State Machine (LTSSM)

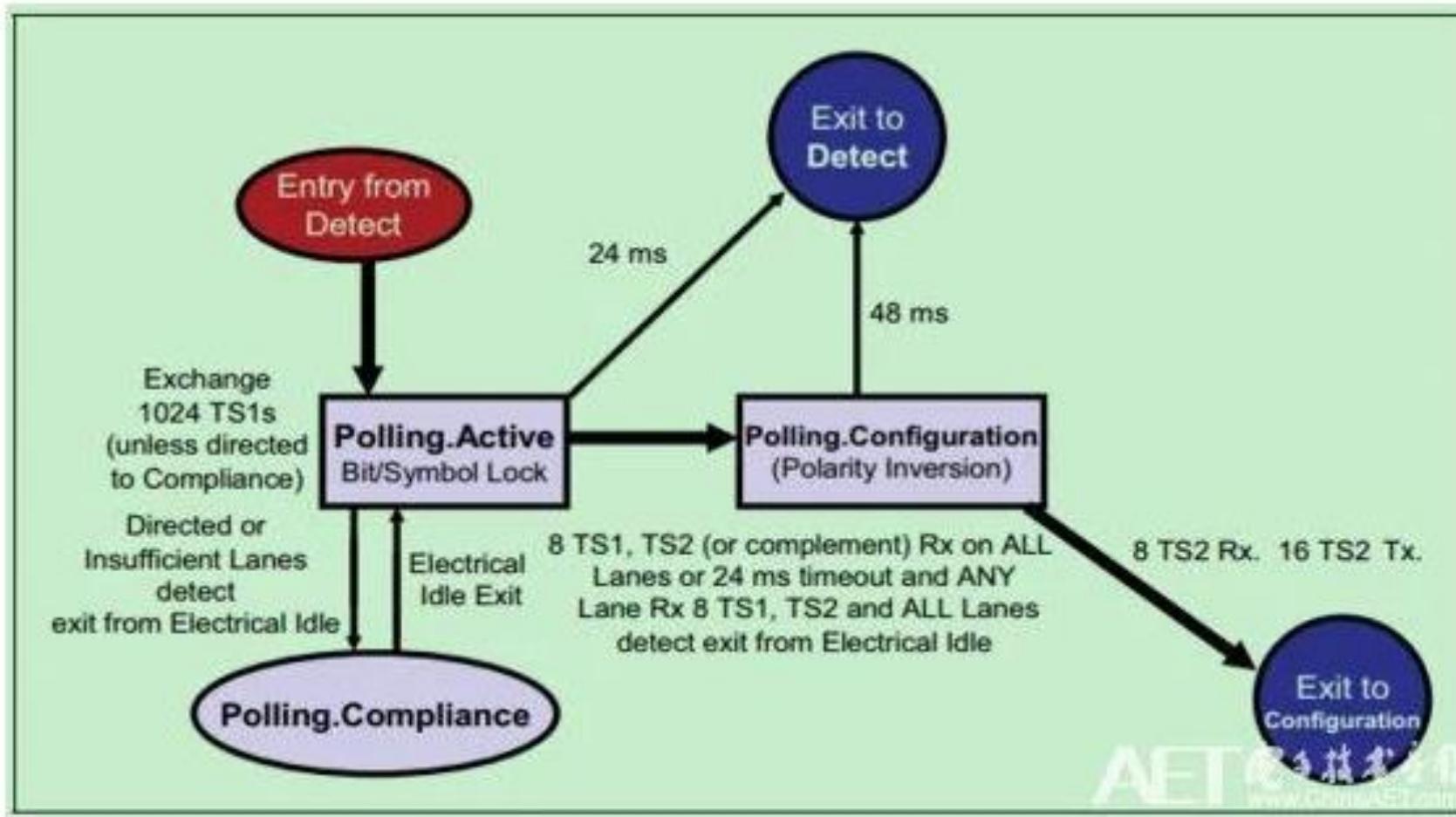
□ Detect



# Link Initialization & Training

## ■ Link Training and Status State Machine (LTSSM)

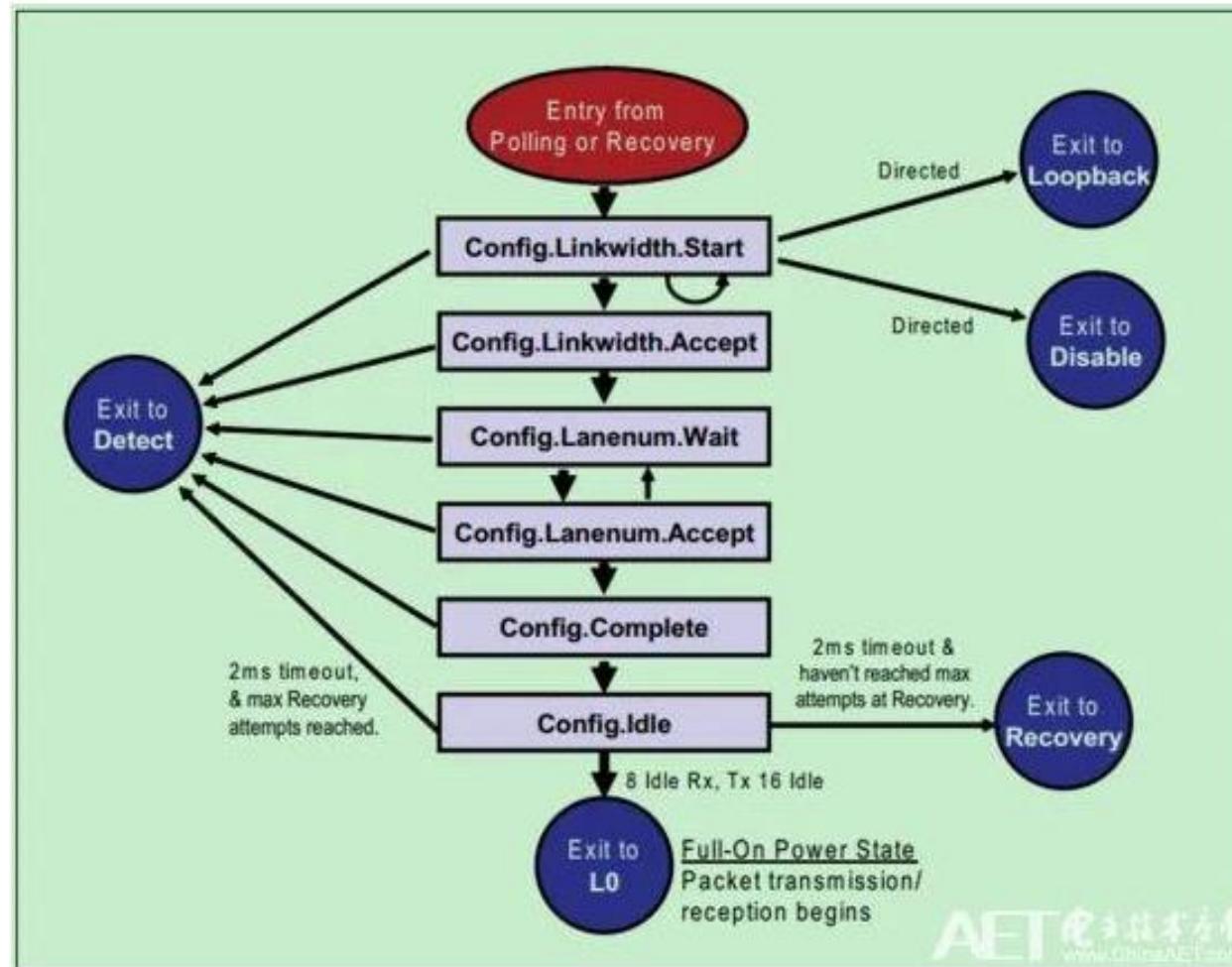
□ Polling



# Link Initialization & Training

## ■ Link Training and Status State Machine (LTSSM)

- Configuration



# Link Initialization & Training

## ■ Link Training and Status State Machine (LTSSM)

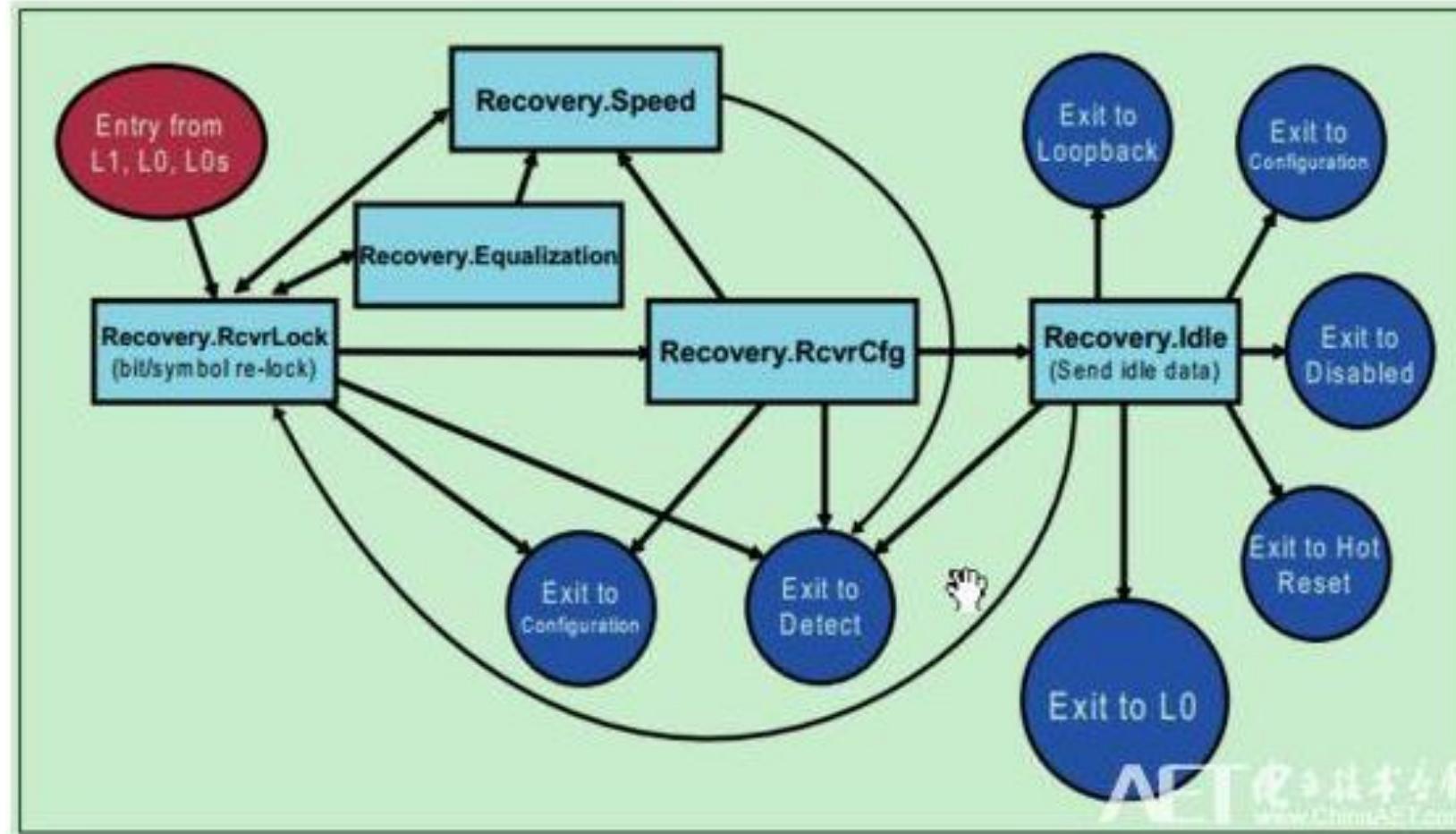
- L0

L0 is the normal operational state where data and control packets can be transmitted and received. All power management states are entered from this state.

# Link Initialization & Training

## ■ Link Training and Status State Machine (LTSSM)

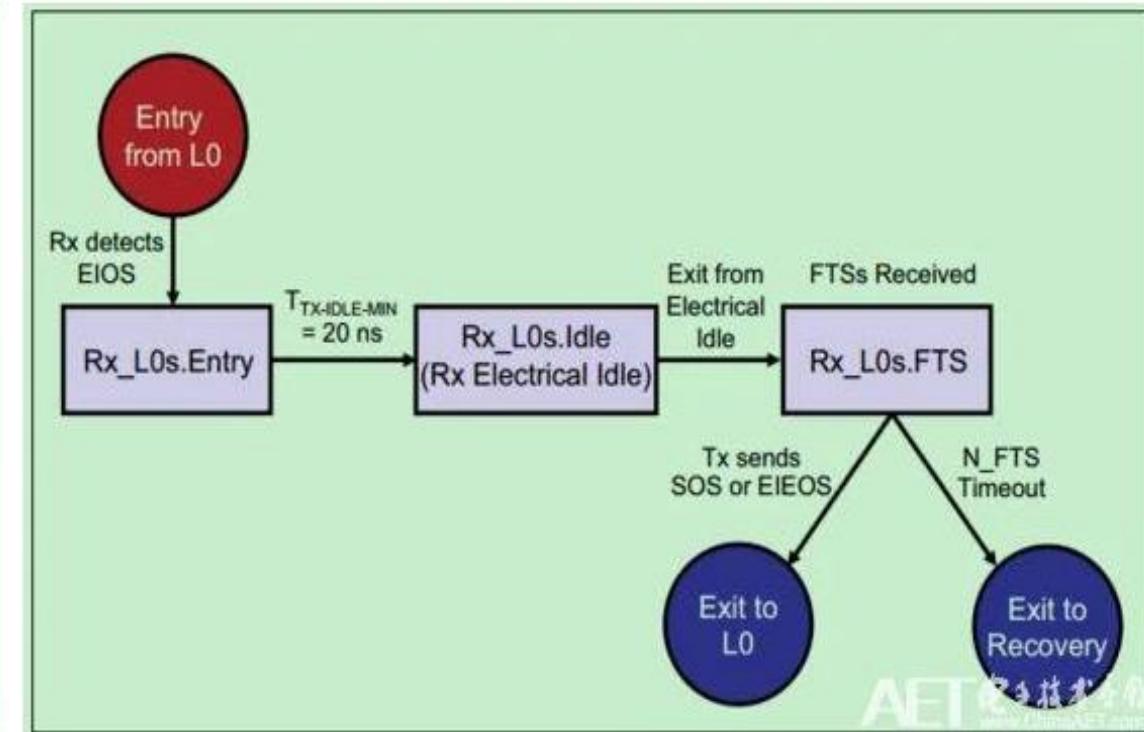
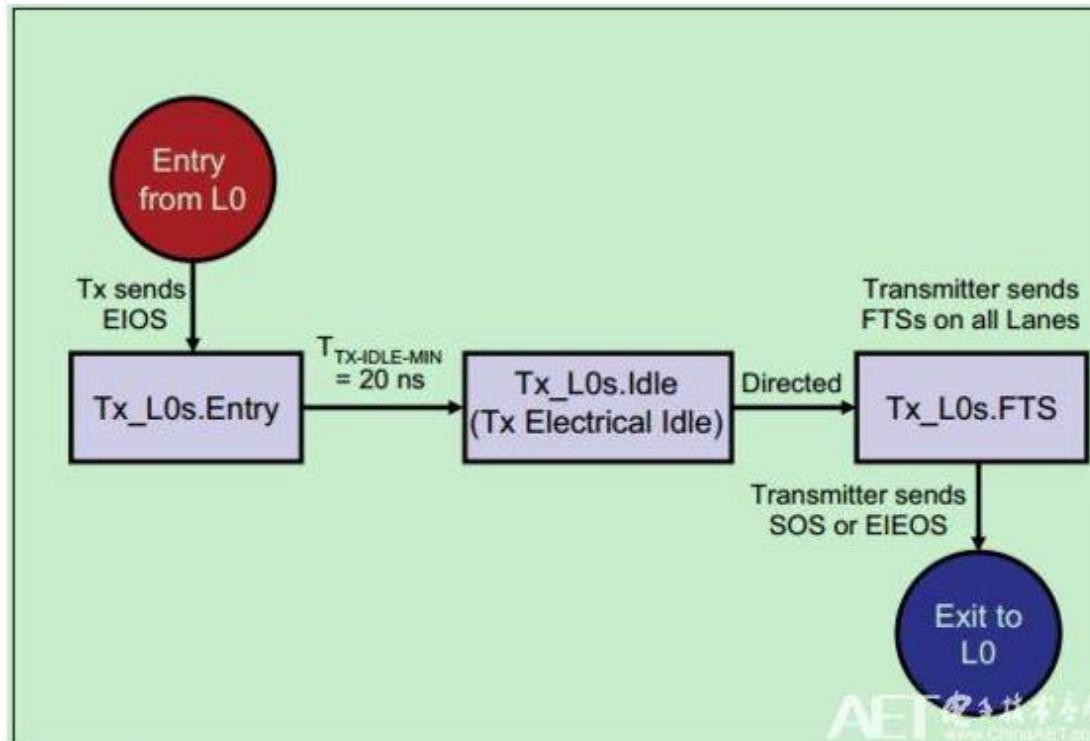
□ Recovery



# Link Initialization & Training

## ■ Link Training and Status State Machine (LTSSM)

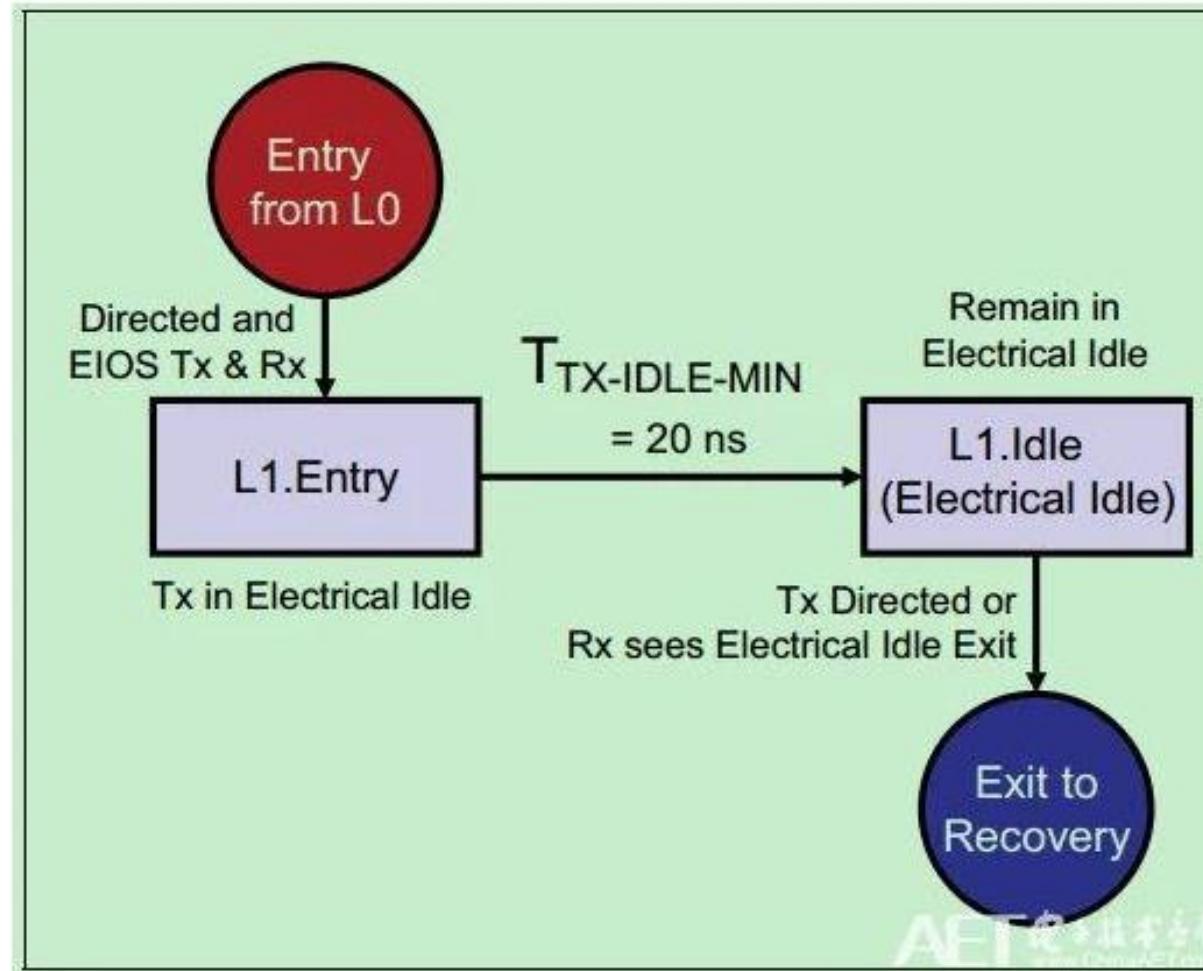
□ L0s



# Link Initialization & Training

## ■ Link Training and Status State Machine (LTSSM)

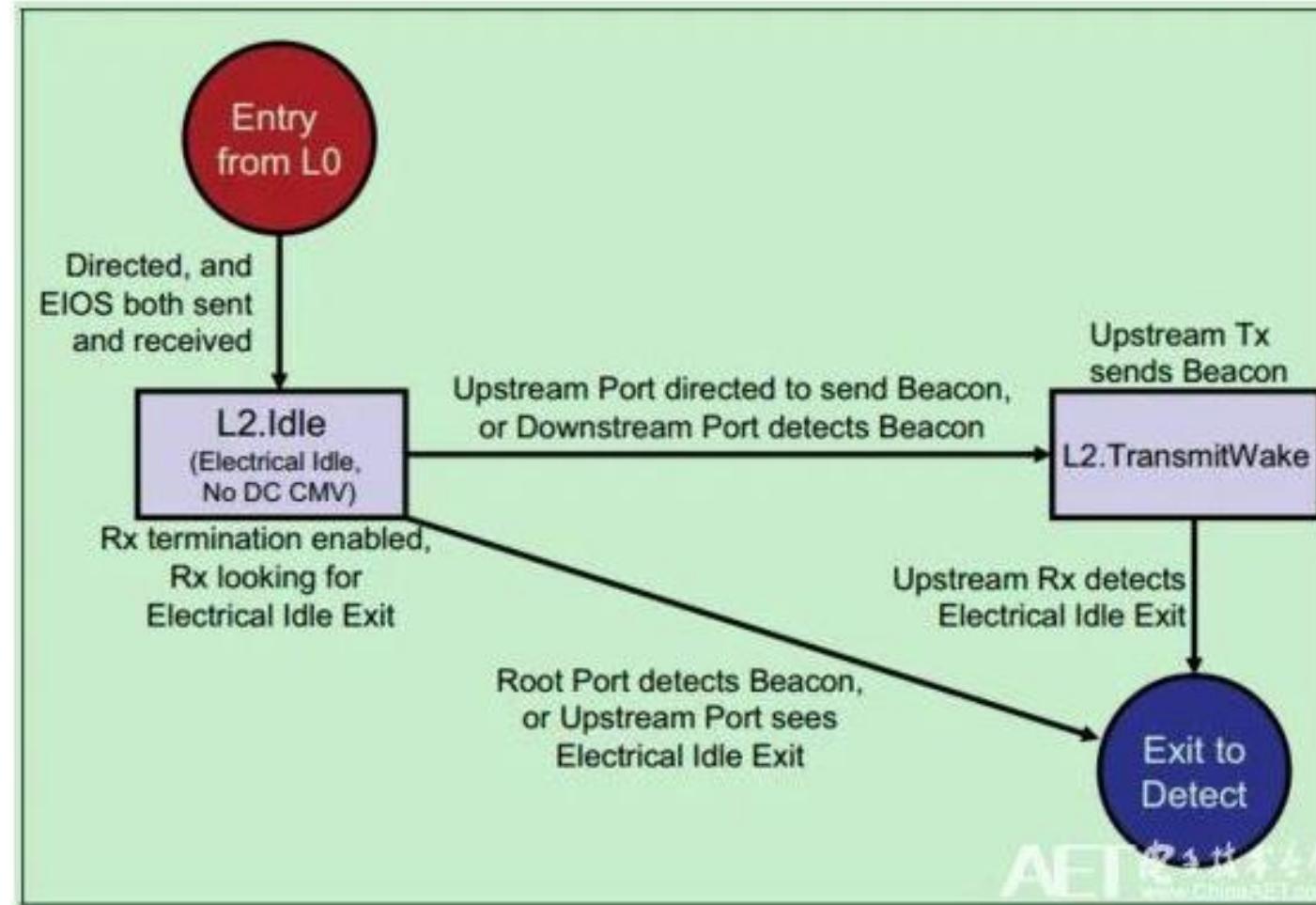
□ L1



# Link Initialization & Training

## ■ Link Training and Status State Machine (LTSSM)

□ L2



# PCI Express

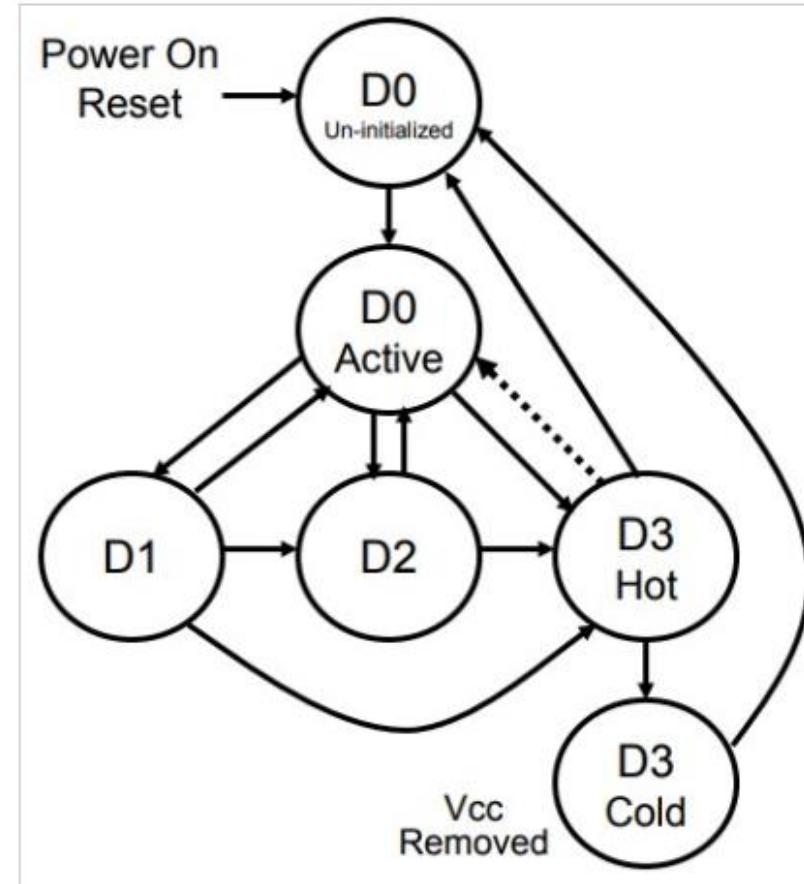
## Power Management



# Power Management

## ■ Function Power Management

### □ Device PM States



# Power Management

## ■ Function Power Management

### □ PCI-PM Registers

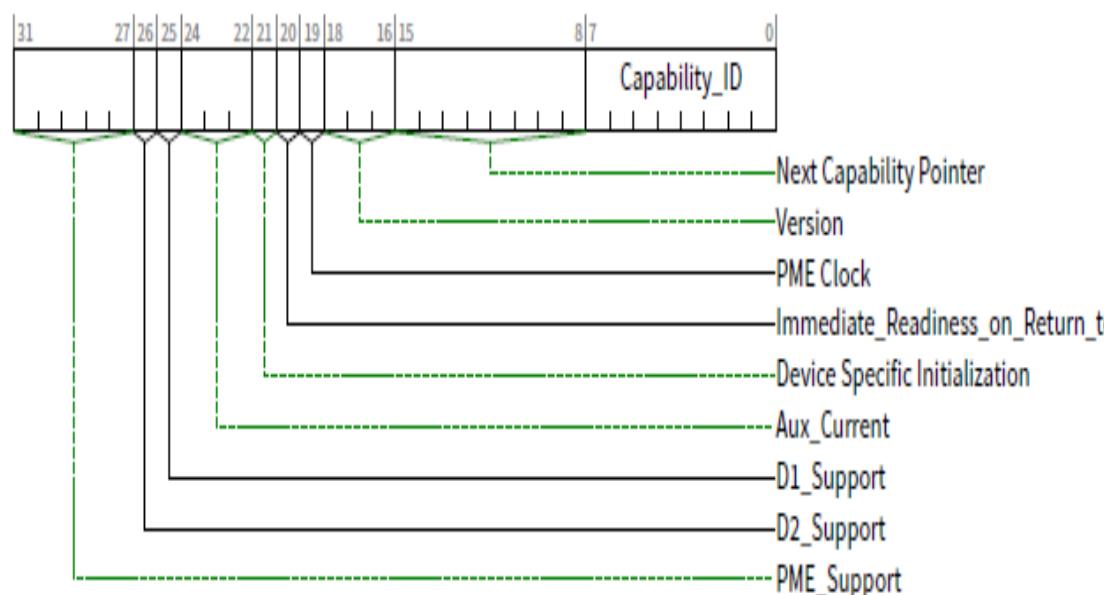


Figure 7-18 Power Management Capabilities Register

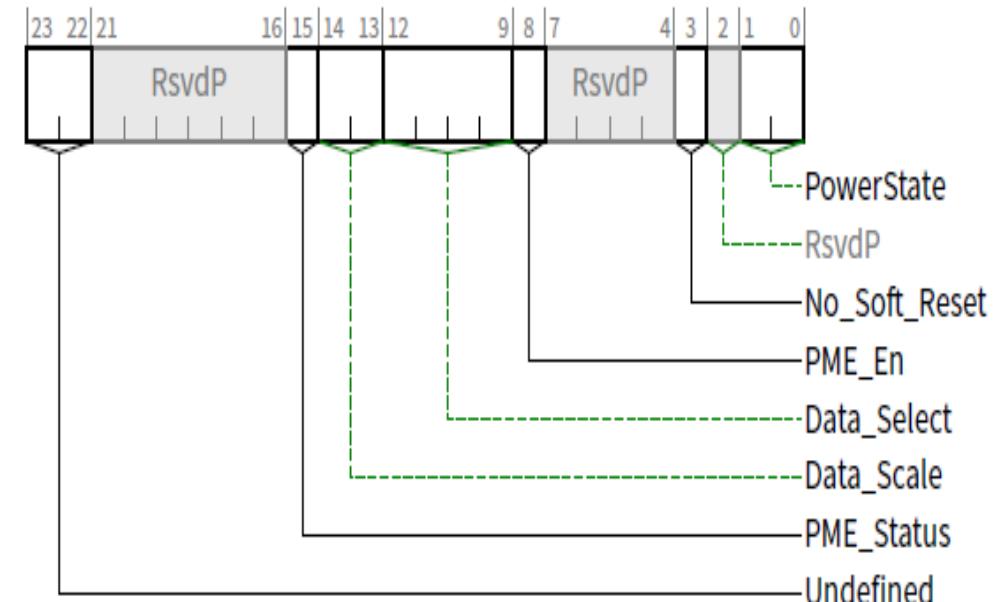
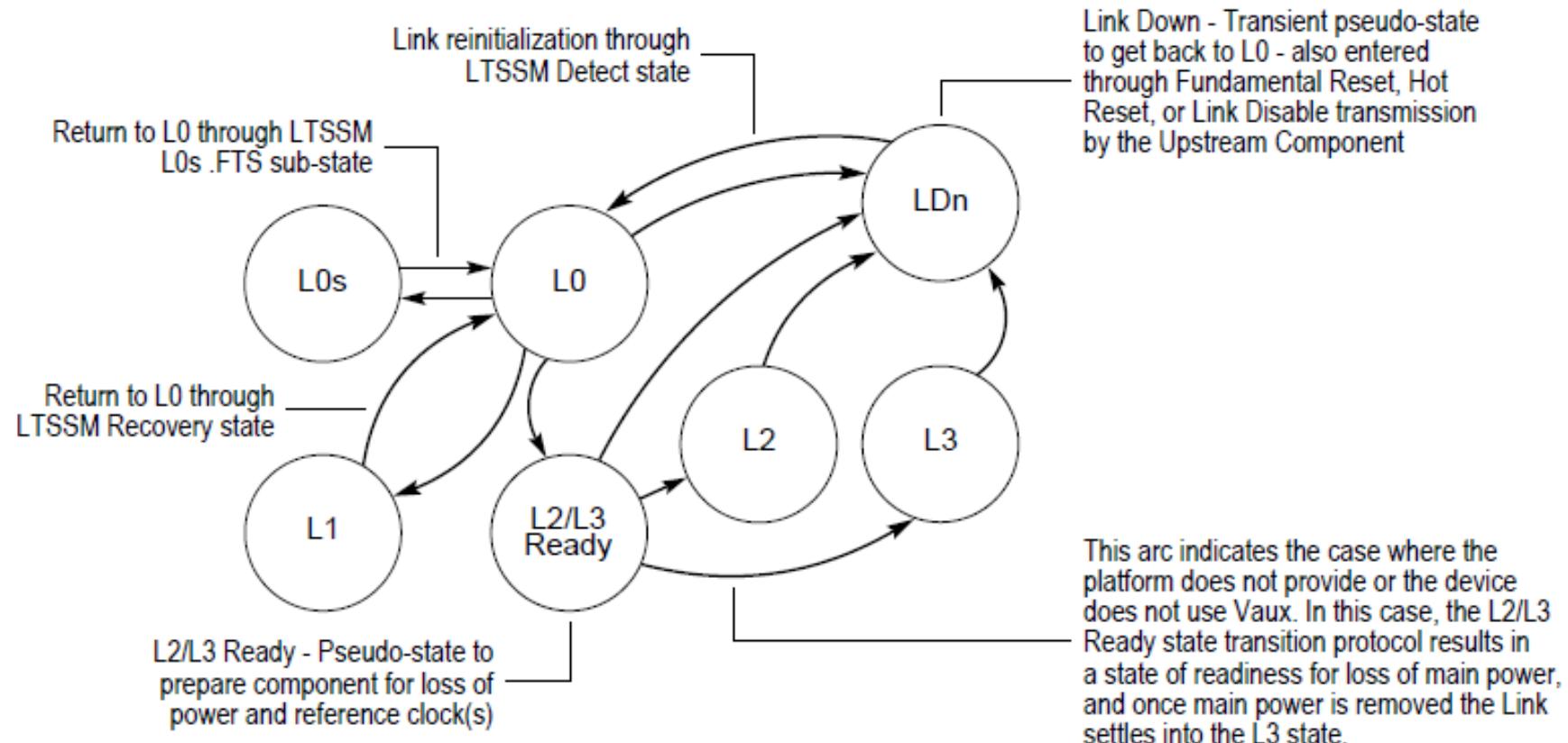


Figure 7-19 Power Management Control/Status Register

# Power Management

## ■ Link Power Management

### □ State Flow Diagram



OM13819B

Figure 5-1 Link Power Management State Flow Diagram

# Power Management

## ■ Link Power Management

### □ State Characteristics

*Table 5-1 Summary of PCI Express Link Power Management States*

	L-State Description	Used by S/W Directed PM	Used by ASPM	Platform Reference Clocks	Platform Main Power	Component Internal PLL	Platform Vaux
<u>L0</u>	Fully active Link	Yes ( <u>D0</u> )	Yes ( <u>D0</u> )	On	On	On	On/Off
<u>L0s</u>	Standby state	No	Yes <sup>1</sup> (opt., <u>D0</u> )	On	On	On	On/Off
<u>L1</u>	Lower power standby	Yes ( <u>D1-D3 Hot</u> )	Yes (opt., <u>D0</u> )	On/Off <sup>6</sup>	On	On/Off <sup>2</sup>	On/Off
<u>L2/L3 Ready</u> (pseudo-state)	Staging point for power removal	Yes <sup>3</sup>	No	On/Off <sup>6</sup>	On	On/Off	On/Off
<u>L2</u>	Low power sleep state (all clocks, main power off)	Yes <sup>4</sup>	No	Off	Off	Off	On <sup>5</sup>
<u>L3</u>	Off (zero power)	n/a	n/a	Off	Off	Off	Off
<u>LDn</u> (pseudo-state)	Transitional state preceding L0	Yes	N/A	On	On	On/Off	On/Off

# Power Management

## ■ Link Power Management

- Relation Between Power Management States of Link and Components

*Table 5-2 Relation Between Power Management States of Link and Components*

Downstream Component D-State	Permissible Upstream Component D-State	Permissible Interconnect State
D0	D0	L0, L0s, L1 <sup>(1)</sup> , L2/L3 Ready
D1	D0-D1	L1, L2/L3 Ready
D2	D0-D2	L1, L2/L3 Ready
D3Hot	D0- D3Hot	L1, L2/L3 Ready
D3Cold	D0- D3Cold	L2 <sup>(2)</sup> , L3

Notes:

1. Requirements for ASPM L0s and ASPM L1 support are form factor specific.
2. If Vaux is provided by the platform, the Link sleeps in L2. In the absence of Vaux, the L-state is L3.

# Power Management

## ■ Active State Power Management (ASPM)

### □ L0s State

L0s, also called L0 standby. L0s applied to a single direction of pcie link. So if device initiates L0s, device can put its tx into L0s while its rx is still in L0.

*Table 7-22 Link Capabilities Register*

Bit Location	Register Description	Attributes
11:10	<p><b>ASPM Support / Active State Power Management Support</b> - This field indicates the level of ASPM supported on the given PCI Express Link. See <a href="#">Section 5.4.1</a> for ASPM support requirements.</p> <p>Defined encodings are:</p> <ul style="list-style-type: none"> <li><b>00b</b> No ASPM Support</li> <li><b>01b</b> L0s Supported</li> <li><b>10b</b> L1 Supported</li> <li><b>11b</b> L0s and L1 Supported</li> </ul> <p><u>Multi-Function Devices</u> associated with an Upstream Port must report the same value in this field for all Functions.</p>	RO

*Table 7-23 Link Control Register*

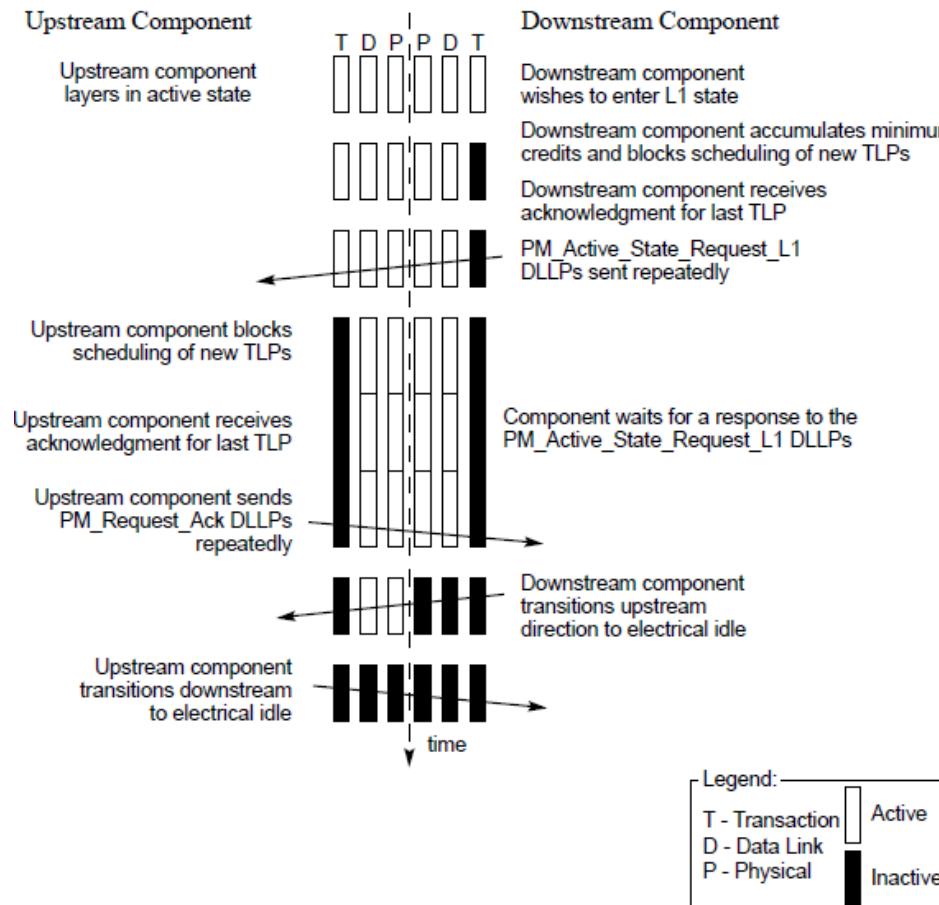
Bit Location	Register Description	Attributes
1:0	<p><b>ASPM Control / Active State Power Management Control</b> - This field controls the level of ASPM enabled on the given PCI Express Link. See <a href="#">Section 5.4.1.3</a> for requirements on when and how to enable ASPM.</p> <p>Defined encodings are:</p> <ul style="list-style-type: none"> <li><b>00b</b> Disabled</li> <li><b>01b</b> L0s Entry Enabled</li> <li><b>10b</b> L1 Entry Enabled</li> <li><b>11b</b> L0s and L1 Entry Enabled</li> </ul> <p>Note: "L0s Entry Enabled" enables the Transmitter to enter L0s. If L0s is supported, the Receiver must be capable of entering L0s even when the Transmitter is disabled from entering L0s (00b or 10b).</p>	RW

When either system or device side pcie phy hardware detects idle time on pcie link, it can put its tx into L0s state. It doesn't need to tell higher layer software to block outbound TLP transactions. L0s exit is also initiated by the corresponding phy hw.

# Power Management

## ■ Active State Power Management (ASPM)

### □ L1 ASPM State



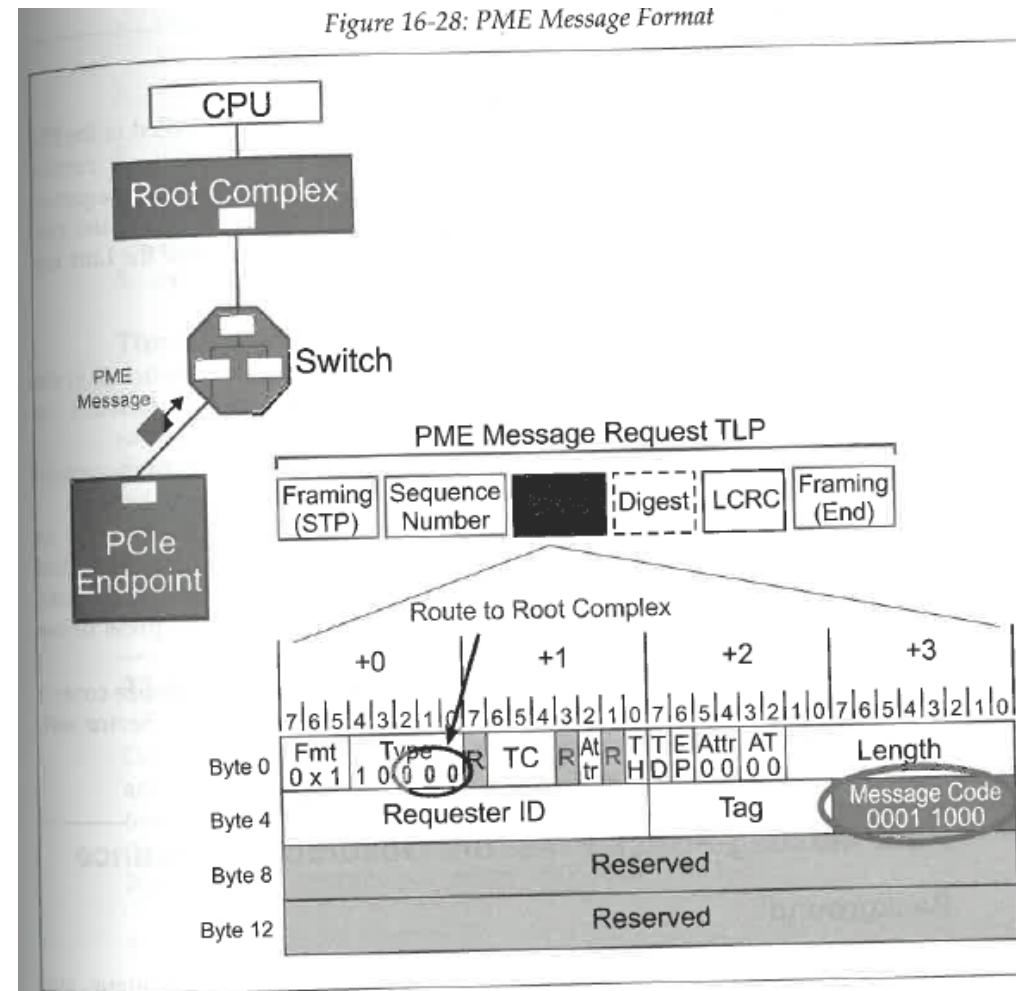
OM13824B

Figure 5-7 L1 Successful Transition Sequence

# Power Management

## ■ Power Management Event Mechanisms

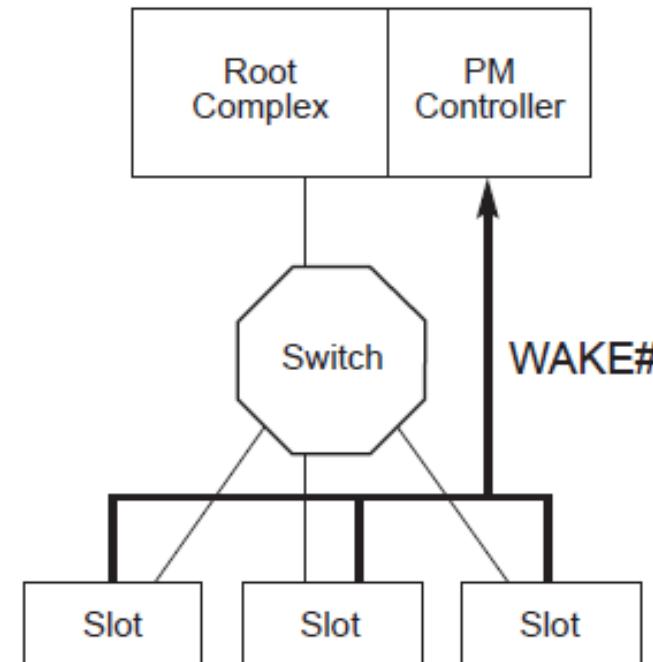
- PME Message



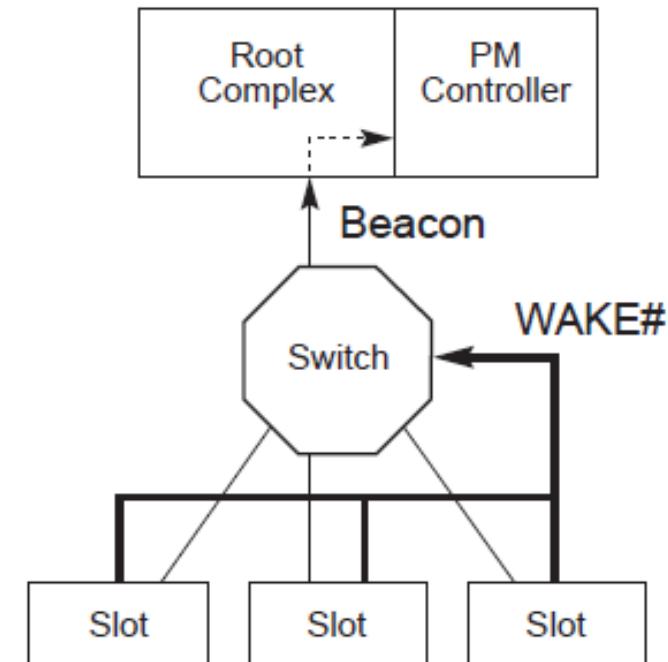
# Power Management

## ■ Power Management Event Mechanisms

### □ Link Wake



Case 1: WAKE# routed directly to system Power Management Controller



Case 2: WAKE# routed to Switch;  
Switch in turn generates Beacon

# Power Management

## ■ L1 PM Substates

- State Diagram

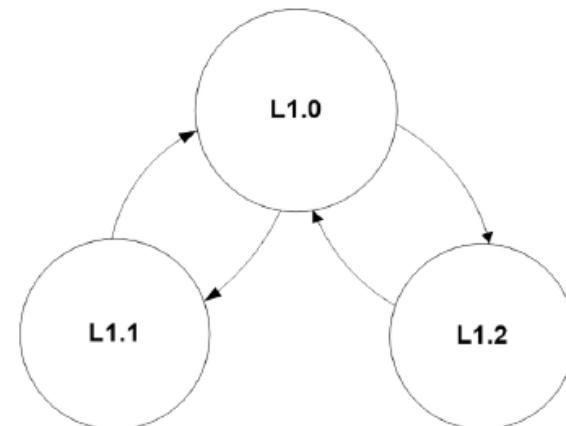


Figure 5-9 State Diagram for L1 PM Substates

Power Sub-states		Status (On/Off)			Targets	
Link	PHY/PIPE	PLL	Rx/Tx	Common Mode Keepers	1-Lane Power	Exit Latency
L1	P1	On/Off	Off/idle	On	20's of mW 10's of mW	< 5 µs (retrain) < 20 µs (PLL Off)
L1.1	P1.1	Off	Off	On	< 500 µW	< 20 µs
L1.2	P1.2	Off	Off	Off	< 10 µW	< 70 µs

Table 1: Link and PHY/PIPE Power Sub-states in Idle States

# Power Management

## L1 PM Substates

### Entry conditions

The following rules define how the L1.1 and L1.2 substates are entered:

- When in PCI-PM L1.0 and the PCI-PM L1.2 Enable bit is Set, the L1.2 substate must be entered when CLKREQ# is deasserted.
- When in PCI-PM L1.0 and the PCI-PM L1.1 Enable bit is Set, the L1.1 substate must be entered when CLKREQ# is deasserted and the PCI-PM L1.2 Enable bit is Clear.
- When in ASPM L1.0 and the ASPM L1.2 Enable bit is Set, the L1.2 substate must be entered when CLKREQ# is deasserted and the reported LTR value last sent or received by this Port is greater than or equal to the value set by the LTR\_L1.2\_THRESHOLD Value and Scale fields.
- When in ASPM L1.0 and the ASPM L1.1 Enable bit is Set, the L1.1 substate must be entered when CLKREQ# is deasserted and the conditions for entering the L1.2 substate are not satisfied.

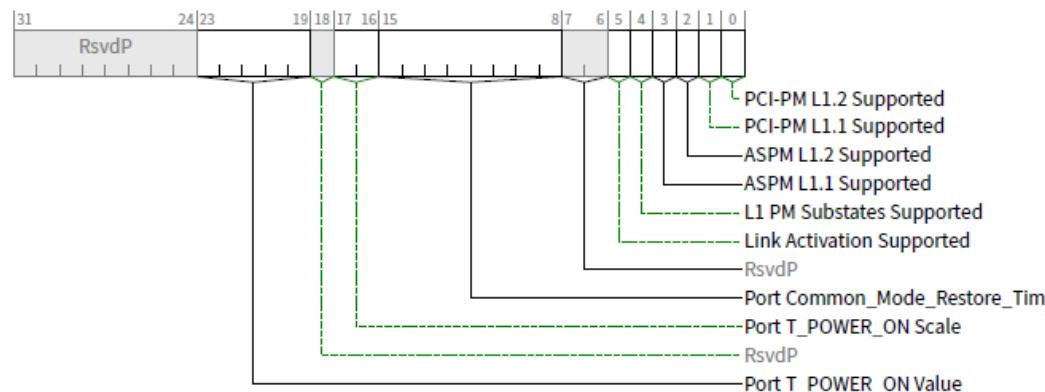


Figure 7-118 L1 PM Substates Capabilities Register

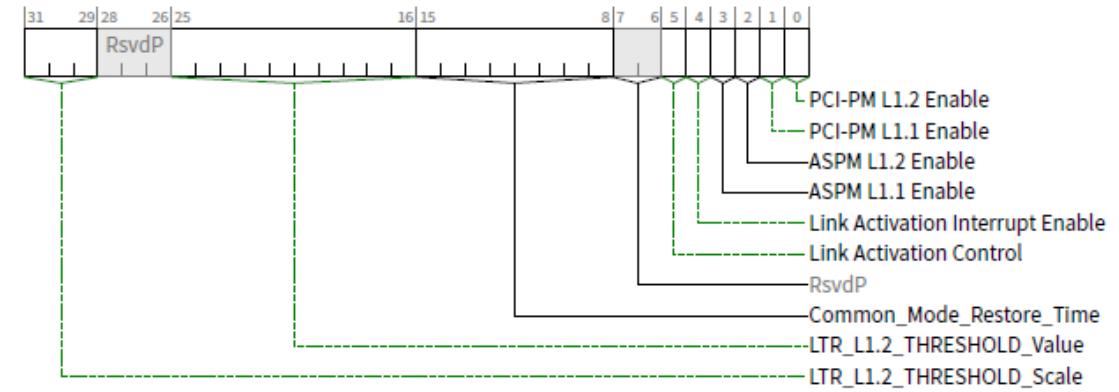


Figure 7-119 L1 PM Substates Control 1 Register

# PCI Express

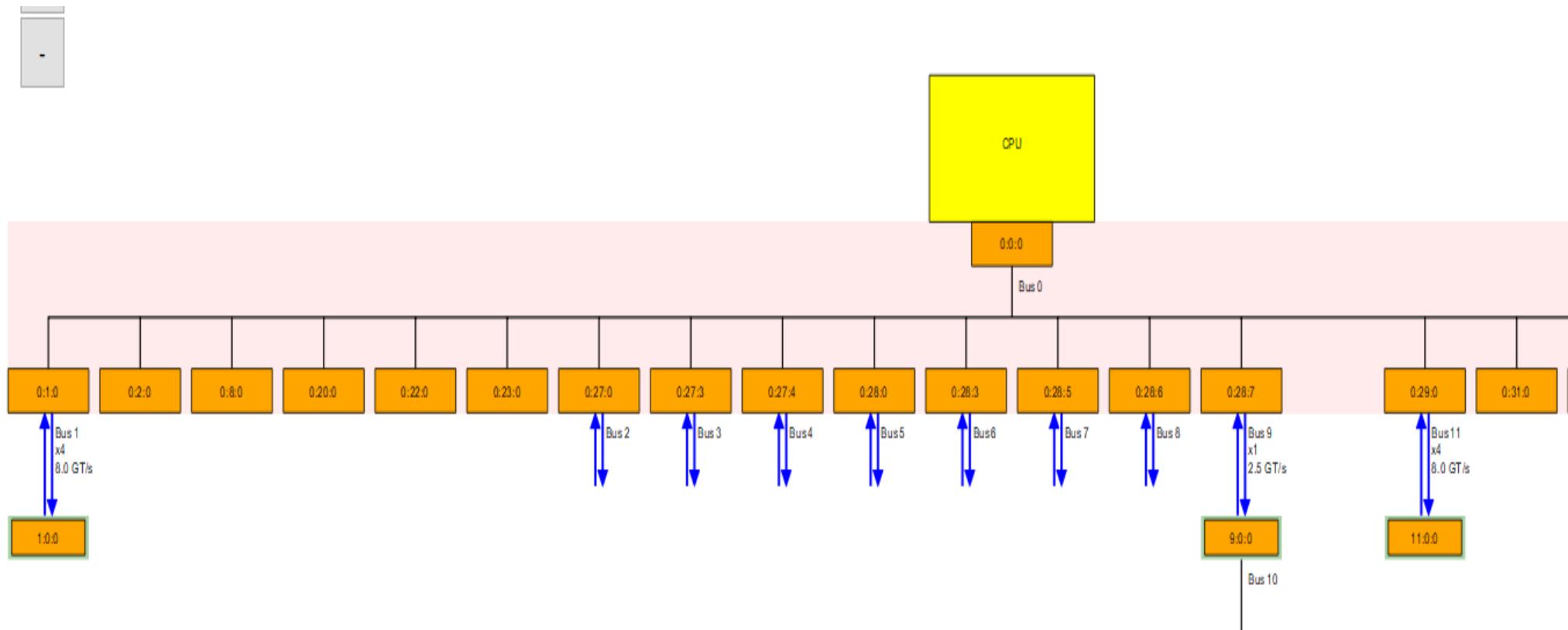
Examples



# Examples

## ■ Enumeration

### □ PCI Map



# Examples

## PCI Config

### Header0

Scan Options PCI Map PCI Config Memory / MMIO NVMe AHCI xHCI IO MSR CPUID Errors

Linear Tree **PCI View - Scanned:**

B:D:F	Class	Device Description	Device Type	Capabilities
0:1:0 (0,1,1)	PCI/PCI bridge	Intel Corporation Xeon E3-1200 v5/E3-1500 v5/6th Gen Core Processor PCIe Controller (x16)	Root Port	Header1, PwrMgmt (01h), SubVID (0Dh), MSI (05h), PCIe (10h), VC (0002h), RCLnkD (0005h), SecPCle (0019h),
1:0:0	NVM Express	Unknown Unknown	PCIe Endpoint	Header0, PwrMgmt (01h), MSI (05h), PCIe (10h), MSI-X (11h), AER (0001h), SecPCle (0019h), LTR (0018h), L1PMSub (001Eh).

decode raw | **1:0:0 - Unknown Unknown** submit feedback

**Type 0 Header**

Device ID 23 02	Vendor ID 1D B2	00h		
Status 00 10	Command 04 06	04h		
Class Code 01 08 02		08h		
BIST 00	HdrType 00	LatTimer 00	\$LineSze 10	0Ch
BAR0 F7 10 00 04				10h
BAR1 00 00 00 00				14h
BAR2 00 00 00 00				18h
BAR3 00 00 00 00				1Ch

**Base Address Register 0 & Base Address Register 1 (Bar0) & (Bar1)**

31	14	4 3 2 1 0
1 1 1 1 0 1 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0	BAR0	

HEX EDIT BINARY EDIT REFRESH

31	0
0 0	BAR1

HEX EDIT BINARY EDIT REFRESH

Address Range: F710\_0000h - F710\_3FFFh

Type: non prefetchable MMIO, 64-bit

Size: 16 KB (bit 14 is first writable bit; sizing value: FFFF\_FFFF\_FFFF\_C004h)

The BARs are used to determine what type of address space(s) each function may need and how much. A BAR that is hardcoded to all 0's is not implemented and means no address space is being requested by that BAR. If a function supports decoding addresses larger than 32 bits, two consecutive BARs will be used together and treated as a single 64-bit BAR.

# Examples

## PCI Config

### PCIe Capability

PCI View - Scanned:

B:D:F	Class	Device Description	Device Type	Capabilities
1:0:0	NVMe Express	Unknown Unknown	PCIe Endpoint	<a href="#">Header0</a> , <a href="#">PwrMgmt (01h)</a> , <a href="#">MSI (05h)</a> , <a href="#">PCIe (10h)</a> , <a href="#">MSI-X (11h)</a> , <a href="#">AER (0001h)</a> , <a href="#">SecPCIe (0019h)</a> , <a href="#">LTR (0018h)</a> , <a href="#">L1PMSub (001Eh)</a> .

decode raw 1:0:0 - Unknown Unknown submit feedback

**PCIe - PCI Express Capability Structure (10h)**

**Version 2**

PCIe Capabilities 00 02	NxtCapPtr B0	CapID 10	70h
Device Capabilities 11 2C 8F C1			74h
Device Status 00 19	Device Control 20 30	78h	
Link Capabilities 00 45 C8 43			7Ch
Link Status 10 43	Link Control 00 40	80h	
Slot Capabilities *			84h
Slot Status * 88h	Slot Control * 88h	88h	
Root Capabilities * 8Ch	Root Control * 8Ch	8Ch	

**Link Status (LinkStatus)**

15 14 13 12 11 10 9      4 3      0

0	0	1	0	0	0	0	1	0	0	0	1
---	---	---	---	---	---	---	---	---	---	---	---

[HEX EDIT](#) [BINARY EDIT](#) [REFRESH](#)

**3:0 - Current Link Speed (RO)**  
 Indicates the negotiated link speed.  

- 0001b = 2.5 GT/s
- 0010b = 5.0 GT/s
- 0011b = 8.0 GT/s
- 0100b = 16.0 GT/s

(Revision 3.0 of the base spec, made this field a binary encoding of the bit location in the Supported Link Speeds Vector in Link Capabilities 2 register that corresponds to the current link speed.)

**9:4 - Negotiated Link Width (RO)**  
 Indicates the negotiated link width (number of lanes).  

- 000001b = x1
- 000010b = x2
- 000100b = x4

# Examples

## ■ NVMe

### □ Controller Registers

Scan Options | PCI Map | PCI Config | Memory / MMIO | **NVMe** | AHCI | xHCI | IO | MSR | CPUID | Errors

**System (NVMe View)**

Name	Description	Decode File	Decode Path
PCI (1:0:0)	NVMe: MMIO: Controller Registers	internal	internal
PCI (1:0:0)	NVMe: Memory: Admin Submission Queue	internal	internal
PCI (1:0:0)	NVMe: Memory: Admin Completion Queue	internal	internal

**PCI (1:0:0) - NVMe: MMIO: Controller Registers** submit feedback

**base: F710\_0000h**

Controller Capabilities Low - CAPLo F0 01 3F FF	00h
Controller Capabilities High - CAPHi 00 10 00 30	04h
<b>Version - VS 00 01 03 00</b>	08h
Interrupt Mask Set - INTMS 00 00 00 00	0Ch
Interrupt Mask Clear - INTMC 00 00 00 00	10h
Controller Configuration - CC 00 46 00 01	14h
Controller Status - CSTS 00 00 00 01	1Ch
NVM Subsystem Reset - NSSR 00 00 00 00	20h

**Version - VS**

31 16:15 8:7 0

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	0 0 0 0 0 0 1 1	0 0 0 0 0 0 0 0
---------------------------------	-----------------	-----------------

[HEX EDIT](#) [BINARY EDIT](#) [REFRESH](#)

**7:0 - Tertiary Version Number - TER (RO)**  
- 0h = Tertiary Version: 0

**15:8 - Minor Version Number - MNR (RO)**  
- 3h = Minor Version: 3

**31:16 - Major Version Number - MJR (RO)**  
- 1h = Major Version: 1

# Examples

## ■ Enumeration

### □ PCI Map

```
jameshsiao@jameshsiao-B365M-D3H:~$ lspci -v -t
-[0000:00]-+00.0 Intel Corporation 8th Gen Core Processor Host Bridge/DRAM Registers
    +-01.0-[01]----00.0 Device 1db2:2302
        +-02.0 Intel Corporation Device 3e98
        +-08.0 Intel Corporation Xeon E3-1200 v5/v6 / E3-1500 v5 / 6th/7th Gen Core Processor Gaussian Mixture Model
        +-14.0 Intel Corporation 200 Series/Z370 Chipset Family USB 3.0 xHCI Controller
        +-16.0 Intel Corporation 200 Series PCH CSME HECI #1
        +-17.0 Intel Corporation 200 Series PCH SATA controller [AHCI mode]
        +-1b.0-[02]--
        +-1b.3-[03]--
        +-1b.4-[04]--
        +-1c.0-[05]--
        +-1c.3-[06]--
        +-1c.5-[07]--
        +-1c.6-[08]--
        +-1c.7-[09-0a]----00.0-[0a]--
            +-1d.0-[0b]----00.0 Intel Corporation Device f1a8
            +-1f.0 Intel Corporation Device a2cc
            +-1f.2 Intel Corporation 200 Series/Z370 Chipset Family Power Management Controller
            +-1f.3 Intel Corporation 200 Series PCH HD Audio
            +-1f.4 Intel Corporation 200 Series/Z370 Chipset Family SMBus Controller
            \-1f.6 Intel Corporation Ethernet Connection (2) I219-V
jameshsiao@jameshsiao-B365M-D3H:~$
```

# Examples

## ■ PCI Config

### □ Header0

```
root@jameshsiao-B365M-D3H:/home/jameshsiao# lspci -xxx -s 01:00.0
01:00.0 Non-Volatile memory controller: Device 1db2:2302 (rev 03)
    00: b2 1d 02 23 06 04 10 00 03 02 08 01 10 00 00 00
    10: 04 00 10 f7 00 00 00 00 00 00 00 00 00 00 00 00
    20: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 b2 1d 02 23
    30: 00 00 00 00 40 00 00 00 00 00 00 00 00 00 0b 01 00 00
    40: 01 50 03 00 08 00 00 00 00 00 00 00 00 00 00 00 00 00
    50: 05 70 86 01 00 00 00 00 00 00 00 00 00 00 00 00 00 00
    60: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
    70: 10 b0 02 00 c1 8f 2c 11 3f 20 19 00 43 c8 45 00
    80: 40 00 43 10 00 00 00 00 00 00 00 00 00 00 00 00 00 00
    90: 00 00 00 00 1f 08 00 00 00 04 00 00 0e 00 00 00 00 00
    a0: 03 00 1e 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
    b0: 11 00 0f 80 00 20 00 00 00 21 00 00 00 00 00 00 00 00
    c0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
    d0: 03 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
    e0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
    f0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00

root@jameshsiao-B365M-D3H:/home/jameshsiao#
```

# Examples

## ■ PCI Config

### □ PCIe Capability

```
root@jameshsiao-B365M-D3H:/home/jameshsiao# lspci -vvv -s 01:00.0
01:00.0 Non-Volatile memory controller: Device 1db2:2302 (rev 03) (prog-if 02 [NVM Express])
    Subsystem: Device 1db2:2302
    Control: I/O- Mem+ BusMaster+ SpecCycle- MemWINV- VGASnoop- ParErr- Stepping- SERR- FastB2B- DisINTx+
    Status: Cap+ 66MHz- UDF- FastB2B- ParErr- DEVSEL=fast >Abort- <Abort- <MAbort- >SERR- <PERR- INTx-
    Latency: 0, Cache Line Size: 64 bytes
    Interrupt: pin A routed to IRQ 16
    Region 0: Memory at f7100000 (64-bit, non-prefetchable) [size=16K]
    Capabilities: [40] Power Management version 3
        Flags: PMEClk- DSI- D1- D2- AuxCurrent=0mA PME(D0-,D1-,D2-,D3hot-,D3cold-)
        Status: D0 NoSoftRst+ PME-Enable- DSel=0 DScale=0 PME-
    Capabilities: [50] MSI: Enable- Count=1/8 Maskable+ 64bit+
        Address: 0000000000000000 Data: 0000
        Masking: 00000000 Pending: 00000000
    Capabilities: [70] Express (v2) Endpoint, MSI 00
        DevCap: MaxPayload 256 bytes, PhantFunc 0, Latency L0s unlimited, L1 unlimited
            ExtTag- AttnBtn- AttnInd- PwrInd- RBE+ FLReset+
        DevCtl: Report errors: Correctable+ Non-Fatal+ Fatal+ Unsupported+
            RxldOrd+ ExtTag- PhantFunc- AuxPwr- NoSnoop- FLReset-
            MaxPayload 256 bytes, MaxReadReq 512 bytes
        DevSta: CorrErr+ UncorrErr- FatalErr- UnsuppReq+ AuxPwr+ TransPend-
        LnkCap: Port #0, Speed 8GT/s, Width x4, ASPM L1, Exit Latency L0s <1us, L1 <8us
            ClockPM+ Surprise- LLActRep- BwNot- ASPMOptComp+
        LnkCtl: ASPM Disabled; RCB 64 bytes Disabled- CommClk+
            ExtSynch- ClockPM- AutWidDis- BWInt- AutBWInt-
        LnkSta: Speed 8GT/s, Width x4 TrErr- Train- SlotClk+ DLActive- BWMgrt- ABWMgmt-
        DevCap2: Completion Timeout: Range ABCD, TimeoutDis+, LTR+, OBFF Not Supported
        DevCtl2: Completion Timeout: 50us to 50ms, TimeoutDis-, LTR+, OBFF Disabled
        LnkCtl2: Target Link Speed: 8GT/s, EnterCompliance- SpeedDis-
            Transmit Margin: Normal Operating Range, EnterModifiedCompliance- ComplianceSOS-
            Compliance De-emphasis: -6dB
        LnkSta2: Current De-emphasis Level: -6dB, EqualizationComplete+, EqualizationPhase1+
            EqualizationPhase2+, EqualizationPhase3+, LinkEqualizationRequest-
    Capabilities: [b0] MSI-X: Enable+ Count=16 Masked-
        Vector table: BAR=0 offset=00002000
        PBA: BAR=0 offset=00002100
```

# Examples

## NVMe

### Set Features

NVMe Cmd		OPC	SQID	CQID	CID	PRP1	PRP2	FID	SV	NSQR	NCQR	NCQA	NSQA	ST	SCT	SC	Device ID	MN	Explicit SQ						
0	H	Set Features	0x0000	0x0000	0x0000	0x00000000:00000000	0x00000000:00000000	Number of Queues	0	0x0001	0x0001	0x000F	0x000F	ST	Generic Command Status	Successful Completion	001:00:0	ATP NVMe M.2 2280 SSD	NVMe #						
Host writes updated Submission Queue tail pointer to doorbell	NVMe 12 H	Device ID 001:00:0	QID 0x0000	SQyTDBL Admin SQT 0x0001	MN ATP NVMe M.2 2280 SSD	Metrics # Link & Split Trans 1	Time Delta 344.000 ns	Time Stamp 0007 . 966 493 466 500 s	Link Tra 8.0 R-> x4 2523 TLP Mem MWr(32) Length 010:00000 RequesterID 000:00:0 Tag 0 Address F7101000 1st BE 1111 Last BE 0000 Data 1 dword VC ID 0 Explicit ACK Packet #2714362 Metrics # Packets 2 Time Delta 344.000 ns Time Stamp 0007 . 966 493 466 500 s	NVMe 14 H	Device ID 001:00:0	QID 0x0000	CID 0x00000000:C991E000	ASQ Set Features Normal operation	FUSE PRP0x0000 NSID 0x00000000:MPTR 0x00000000:00000000	Address PRP1 0x00000000:00000000 Address PRP2 0x00000000:00000000	Address PRP1 0x00000000:00000000 Address PRP2 0x00000000:00000000	FID 0 Number of Queues 0 SV 0x0001 NCQR 0x0001 AT							
Controller fetches command	NVMe 1257 H	Device ID 001:00:0	QID 0x0000	MRd(32) 000:00000	RequesterID 001:00:0 Tag 0 TC 0 VC ID 0 Address C991E000 Status SC Data 16 dwords Metrics # LinkTrans 2 Time Delta 158.750 ns Time Stamp 0007 . 966 493 810 500 s	Link Tra 8.0 R-> x4 2525 TLP Mem MRd(32) Length 000:00000 RequesterID 001:00:0 Tag 0 Address 1st BE 1111 Last BE 1111 VC ID 0 Explicit ACK Packet #2714365 Metrics # Packets 2 Time Delta 158.750 ns Time Stamp 0007 . 966 493 810 500 s	Link Tra 8.0 R-> x4 2527 TLP Cpl CplID 010:01010 Length 16 RequesterID 001:00:0 Tag 0 CompleterID 000:00:0 Status SC BCM Byte Cnt 0x00 Lwr Addr 0x00 Data 16 dwords VC ID 0 Explicit ACK Packet #2714369 Metrics # Packets 2 Time Delta 528.720 us Time Stamp 0007 . 966 493 969 250 s	NVMe 16 D	Device ID 001:00:0	QID 0x0000	CID 0x00000000:C9918000	ACQ 0x0001 SQHD 0x0000 SQID 0x0000 CID 1 P 1 DWO 0x000FF00F Command Specific ST Generic Command Status	SCT Successful Completion M 0 DNR 0 MN ATP NVMe M.2 2280 SSD	# Link & Split Trans 1 Time Delta 3.663 us Time Stamp 0007 . 967 022 689 250 s											
Controller processes command	Link Tra 8.0 R-> x4 2606 TLP Mem MWr(32) Length 010:00000 RequesterID 001:00:0 Tag 0 Address C9918000 1st BE 1111 Last BE 1111 Data 4 dwords VC ID 0 Explicit ACK Packet #2714528 Metrics # Packets 2 Time Delta 3.663 us Time Stamp 0007 . 967 022 689 250 s	NVMe 17 H	Device ID 001:00:0	QID 0x0000	CQyHDBL Admin CQH 0x0001	MN ATP NVMe M.2 2280 SSD	Metrics # Link & Split Trans 1	Time Delta 4.702 us	Time Stamp 0007 . 967 026 352 750 s	Link Tra 8.0 R-> x4 2608 TLP Mem MWr(32) Length 010:00000 RequesterID 000:00:0 Tag 0 Address F7101004 1st BE 1111 Last BE 0000 Data 1 dword VC ID 0 Explicit ACK Packet #2714531 Metrics # Packets 2 Time Delta 4.702 us Time Stamp 0007 . 967 026 352 750 s															
Host writes updated Completion Queue head pointer to doorbell																									

# Examples

## NVMe

### Identify Controller

NVMe Cmd		OPC		1 warning(s)		SQID	CQID	CID	Data	PRP1	PRP2	CNS	CNTID	Controller Capabilities	VID	SSVID	SN	MN			
1	D	Identify Controller		NVMe Cmd Warnings (3)		0x0000	0x0000	0x0001	1024 dwords	0x00000000:C4E12018	0x00000000:C4E14000	Identify Controller	0x0000	Controller Capabilities	0x1DB2	0x1DB2	5511561	ATP NVMe M.2 2280 SSD			
Host writes updated Submission Queue tail pointer to doorbell	NVMe	H	Device ID	QID	CID	SQyTDBL	Admin SQT	MN	Metrics	# Link & Split Trans	Time Delta	Time Stamp									
	18	001:00:0	0x0000	0x0002	ATP NVMe M.2 2280 SSD				1	350.000 ns	0007 . 967 031 055 250 s										
Controller fetches command	NVMe	H	Device ID	QID	CID	Address	ASQ	OPC	FUSE	PSDT	CID	NSID	MPTR	Address	PRP1	Address	PRP2	Address	CNS CNTID MN		
	20	001:00:0	0x0000	0x0001	00000000:C991E040	Identify	Normal operation	PRP	0x0001	0x00000000	0x00000000:00000000	0x00000000:00000000	0x00000000:00000000	0x00000000:C4E12018	0x00000000:C4E14000	Controller	0x0000	ATP NVMe M.2 2280 SSD			
Controller processes command	NVMe	D	Device ID	QID	CID	Address	identify Controller	Controller Capabilities	VID	SSVID	SN	MN	FR	RAB	IEEE	CMIC	Bit 0	Bit 1	Bit 2	MDTS CNTID VER RTD3R	
	22	001:00:0	0x0000	0x0001	00000000:C4E12018			0x1DB2	0x1DB2	5511561	ATP NVMe M.2 2280 SSD	42A4S6B	0x06	0x141357	0	0	0	0x06	0x0001 0x00010300 0x0007A12		
	Link Tra	R← 8.0	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp			
	2913	x4	1453		010:00000	58	001:00:0	0	C4E12018	1111	1111	58 dwords	0	Packet #2715142		2	63.750 ns	0007 . 968 910 478 500 s			
	Link Tra	R← 8.0	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp			
	2914	x4	1454		010:00000	64	001:00:0	0	C4E12100	1111	1111	64 dwords	0	Packet #2715145		2	70.250 ns	0007 . 968 910 542 250 s			
Controller writes completion to Completion Queue	Link Tra	R← 8.0	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp			
	2935	x4	1475		010:00000	64	001:00:0	0	C4E12F00	1111	1111	64 dwords	0	Packet #2715187		2	70.000 ns	0007 . 968 911 591 500 s			
	Link Tra	R← 8.0	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp			
	2936	x4	1476		010:00000	6	001:00:0	0	C4E13000	1111	1111	6 dwords	0	Packet #2715189		2	7.250 ns	0007 . 968 911 661 500 s			
Host writes updated Completion Queue head pointer to doorbell	NVMe	D	Device ID	QID	CID	Address	ACQ	SQHD	SQID	CID	P	DWO	Command Specific	SCT	SC	M DNR	MN	# Link & Split Trans	Time Delta	Time S	
	23	001:00:0	0x0000	0x0001	00000000:C9918010	0x0002	0x0000	0x0001	1	0x00000000	0x0000	0x00000000	ST	Generic Command Status	Successful Completion	0 0	ATP NVMe M.2 2280 SSD	Metrics	1	5.970 us	0007 . 968 917 638 750 s
	Link Tra	R← 8.0	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp			
	24	001:00:0	0x0000	CQyHDBL	Admin CQH	MN	Metrics	# Link & Split Trans	Time Delta	Time Stamp											
	Link Tra	R← 8.0	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp			
	2937	x4	1477		010:00000	64	001:00:0	0	C4E13F00	1111	1111	64 dwords	0	Packet #2715191		2	70.000 ns	0007 . 968 911 591 500 s			
	Link Tra	R← 8.0	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp			
	2938	x4	1478		010:00000	6	001:00:0	0	C4E13000	1111	1111	6 dwords	0	Packet #2715193		2	7.250 ns	0007 . 968 911 661 500 s			
Host writes updated Completion Queue head pointer to doorbell	NVMe	H	Device ID	QID	CID	Address	ACQ	SQHD	SQID	CID	P	DWO	Command Specific	SCT	SC	M DNR	MN	# Link & Split Trans	Time Delta	Time S	
	24	001:00:0	0x0000	CQyHDBL	Admin CQH	MN	Metrics	# Link & Split Trans	Time Delta	Time Stamp											

# Examples

## NVMe

### Read command

NVMe Cmd		OPC	SQID	CQID	CID	PRP1	QID	QSIZE	PC	IEN	IV	ST	SCT	SC	Device ID	MN	Explicit SQyTDBL	Explicit ASQ	Explicit ACQ	Explicit CQyHDB		
3	H	Create I/O CQ	0x0000	0x0000	0x0003	0x00000000:C9914000	0x0001	0x00FF	1	0	0x0000	ST	Generic Command Status	Successful Completion	001:00:0	ATP NVMe M.2 2280 SSD	NVMe #33	NVMe #35	NVMe #37	NVMe #38		
4	H	Create I/O SQ	0x0000	0x0000	0x0004	0x00000000:C990E000	0x0001	0x00FF	1	Urgent	0x0001	ST	Generic Command Status	Successful Completion	001:00:0	ATP NVMe M.2 2280 SSD	NVMe #39	NVMe #41	NVMe #43	NVMe #4		
5	D	Identify Namespace	1 warning(s)		SQID	CQID	CID	Data	PRP1	PRP2	CNS	CNTID	NSZE	NCAP	NUSE	Bit 0 Bit 1	NSFEAT	0	0			
6	D	Read	0x0001	0x0001	0x0000	128 dwords	0x00000000:00000000	0x00000000:C4E0F898	PRP1	PRP2	SLBA	NLB	PRINFO	FUA	LR	DSM	ACCF	ACCL	SEQR	INCOM	EILBR	
52	H	Device ID	QID	IO SQT	MN	Metrics	# Link & Split Trans	Time Delta	Time Stamp	1	343.750 ns	0007 . 969 282 125 500 s										
Host writes updated Submission Queue tail pointer to doorbell		Link Tra	R← 8.0	TLP	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp			
3065	x4	1497	3065	010:00000	1	000:00:0	0	F7101008	1111	0000	1 dword	0	Packet #2715446			2	343.750 ns	0007 . 969 282 125 500 s				
54	H	Device ID	QID	CID	Address	IOSC	OPC	FUSE	PSDT	CID	NSID	MPTR	Address	PRP1	Address	PRP2	Address	SLBA	NLB	PRINFO	PRCHK	
1477	x4	MRd(32)	Mem	000:00000	001:00:0	000:00:0	0	0	0	C990E000	0x00000000:00000001	0x00000000:C4E0F898	0x00000000:00000000	0x00000000:00000000	0x00000000:00000000	0x00000000:00000000	0x00000000:00000000	0x00000000:00000000	0x00000000:00000000	0x00000000:00000000	0x00000000:00000000	0x00000000:00000000
Controller fetches command		Link Tra	R← 8.0	TLP	Mem	MRd(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp			
3067	x4	1568	3067	000:00000	16	001:00:0	0	C990E000	1111	1111	0	Packet #2715449				2	177.250 ns	0007 . 969 282 469 250 s				
3069	x4	1499	Cpl	Cpl	Length	RequesterID	Tag	CompleterID	Status	BCM	Byte Cnt	Lwr Addr	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp			
56	D	Device ID	QID	CID	Address	PRP Data	Data Len	Data	MN	Metrics	# Link & Split Trans	Time Delta	Time Stamp	1	53.000 ns	0007 . 969 630 753 500 s						
3128	x4	1599	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp					
3129	x4	1600	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp					
3130	x4	1601	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp					
57	D	Device ID	QID	CID	Address	IOQO	SQHD	SOID	CID	P	DW0	Command Specific	ST	SCT	SC	M	DNR	MN	Metrics	# Link & Split Trans	Time Delta	Time
3131	x4	1602	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp	1	2.291 us	0007 . 969 630 920 250 s		
58	H	Device ID	QID	IO CQH	MN	Metrics	# Link & Split Trans	Time Delta	Time Stamp	1	119.728 us	0007 . 969 633 210 750 s										
3132	x4	1529	Mem	MWr(32)	Length	RequesterID	Tag	Address	1st BE	Last BE	Data	VC ID	Explicit ACK	Metrics	# Packets	Time Delta	Time Stamp	2	119.728 us	0007 . 969 633 210 750 s		

# PCI Express

Q & A





[www.atpinc.com](http://www.atpinc.com)

A composite background image featuring three distinct scenes: a blurred view from inside a car showing the dashboard and a road ahead; a large industrial facility with complex steel structures and pipes; and a satellite in space against a backdrop of Earth's horizon and clouds.

# INDUSTRIAL ONLY

