



PCI-SIG ENGINEERING CHANGE NOTICE

Title of ECN:	Root Complex Topology Discovery
Date:	8.12.03
Affected Documents:	PCI Express Base Specification 1.0a
Sponsor:	Ramin Neshati (Intel)

1. Summary of Functional Changes:

This ECN describes changes necessary for supporting software discovery of Root Complex Register Blocks in PCI Express Root Complexes. This ECN does not impact any existing functionality or devices.

2. Benefits:

This change allows software to automatically discover Root Complex Register Blocks in a PCI Express Root Complex and establish topological relationships between Root Complex Register Blocks and other Root Complex elements such as Root Ports. This functionality is necessary for software to correctly program PCI Express Extended Capabilities present in Root Complex Register Blocks.

3. Assessment of the Impact:

There is no impact to systems or peripherals that conform to PCI Express Base Specification 1.0a.

4. Analysis of Hardware Implications:

This ECN allows Root Complexes to expose internal programming parameters correctly to software; this functionality is optional and not required for Root Complexes that do not expose internal programming parameters.

5. Analysis of Software Implications:

Software supporting the functionality described in this ECN will have to correctly interpret the new capability structures described; software not supporting this functionality is not impacted and operates as before

6. Additional Description and Rationale:

None

7. Details of Change:

Add term to Glossary:

Root Complex Component A logical aggregation of Root Ports, Root Complex Register Blocks and Root Complex Integrated Devices.

Pg 315, Section 7.2.3, Root Complex Register Block, remove first sentence of second paragraph in section:

~~System firmware communicates the base address of the RCRB for each Root Port or internal device in the Root Complex to the operating system.~~ Multiple Root Ports or internal devices may be associated with the same RCRB. The RCRB memory-mapped registers must not reside in the same address space as the memory-mapped configuration space.

Add following content:

6.

6.10 Root Complex Topology Discovery

A PCI Express Root Complex may present one of the following topologies to configuration software:

- A single opaque Root Complex such that software has no visibility with respect to internal operation of the Root Complex. All Root Ports are independent of each other from a software perspective; no mechanism exists to manage any arbitration among the various Root Ports for any differentiated services.
- A single Root Complex Component such that software has visibility and control with respect to internal operation of the Root Complex Component. As shown in Figure 6-10, software views the Root Ports as ingress ports for the component. The Root Complex internal port for traffic aggregation to a system egress port or an internal sink unit (such as memory) is represented by an RCRB structure. Controls for differentiated services are provided through a PCI Express Virtual Channel capability structure located in the RCRB.

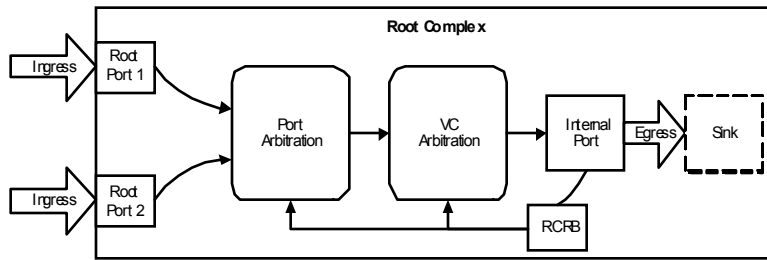


Figure 6-10 Root Complex represented as a single component

- Multiple Root Complex Components such that software not only has visibility and control with respect to internal operation of a given Root Complex Component but also has the ability to discover and control arbitration between different Root Complex Components. As shown in Figure 6-11, software views the Root Ports as ingress ports for a given component. An RCRB structure controls egress from the component to other Root Complex Components (RCRB C) or to an internal sink unit such as memory (RCRB A). In addition, an RCRB structure (RCRB B) may also be present in a given component to control traffic from other Root Complex Components. Controls for differentiated services are provided through PCI Express Virtual Channel capability structures located appropriately in the RCRBs respectively.

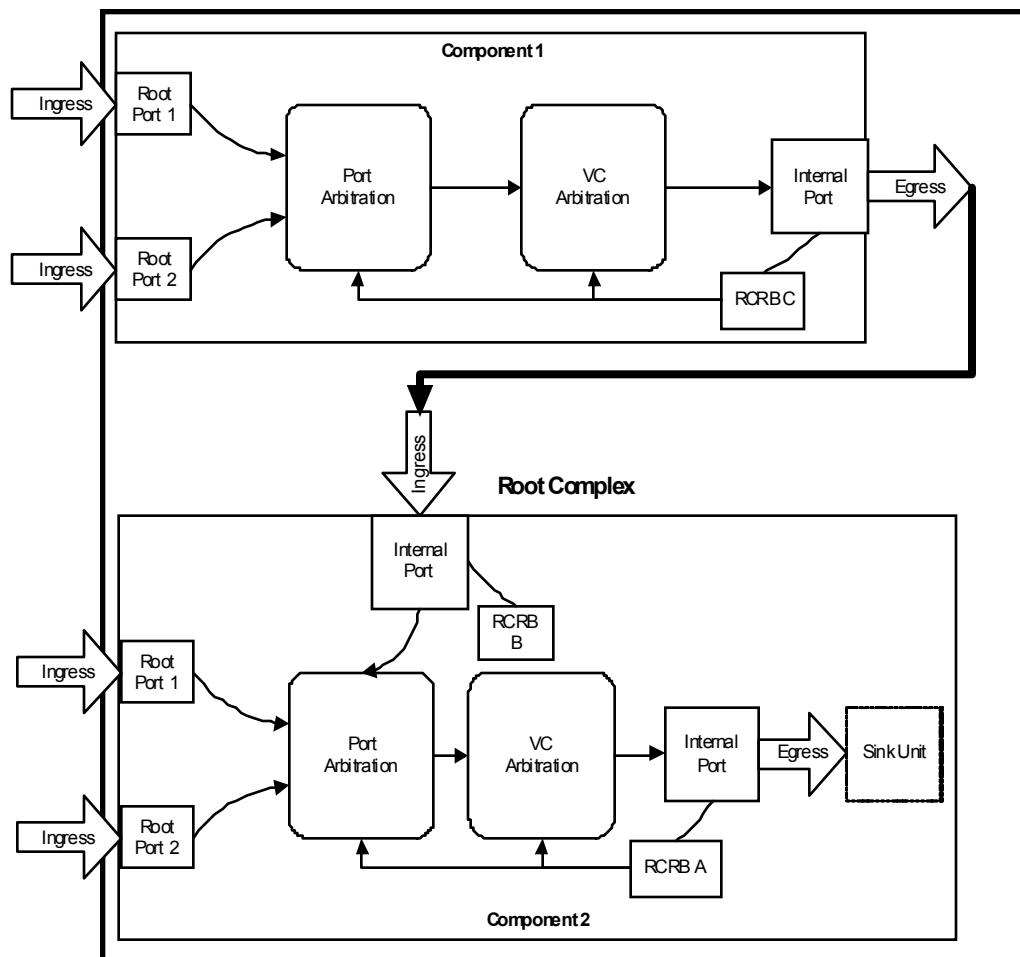


Figure 6-11 Root Complex represented as multiple components

More complex topologies are possible as well.

A Root Complex topology can be represented as a collection of logical Root Complex Components such that each logical component has:

- One or more ingress ports.
- An egress port.
- Optional associated virtual channel capabilities located either in the configuration space (for root ports) or in an RCRB (for internal ingress/egress ports) if the Root Complex supports virtual channels.
- Optional devices/functions integrated in the Root Complex.

In order for software to correctly program arbitration and other control parameters for PCI Express differentiated services, software must be able to discover a Root Complex's internal topology. Root Complex topology discovery is accomplished

by means of the PCI Express Root Complex Link Declaration Capability as described in Section 7.13.

7.

7.13 PCI Express Root Complex Link Declaration Capability

The PCI Express Root Complex Link Declaration Capability is an optional capability that may be implemented by Root Ports, Root Complex Integrated Devices or RCRBs to declare a Root Complex's internal topology.

A Root Complex consists of one or more following elements:

- PCI Express Root Port
- A default system egress port or an internal sink unit such as memory (represented by an RCRB)
- Internal Data Paths/Links (represented by an RCRB on either side of an internal link)
- Integrated devices/functions

A Root Complex Component is a logical aggregation of the above described Root Complex elements. No single element can be part of more than one Root Complex Component. Each Root Complex Component must have a unique Component ID.

A Root Complex is represented either as an opaque Root Complex or as a collection of one or more Root Complex Components.

The PCI Express Root Complex Link Declaration Capability may be present in a Root Complex element's configuration space or RCRB. It declares links from the respective element to other elements of the same Root Complex Component or to an element in another Root Complex Component. The links are required to be declared bidirectional such that each valid data path from one element to another has corresponding link entries in the configuration space (or RCRB) of both elements.



IMPLEMENTATION NOTE

Topologies to Avoid

Topologies that create more than one data path between any two Root Complex elements (either directly or through other Root Complex elements) may not be able to support bandwidth allocation in a standard manner. The description of

how traffic is routed through such a topology is implementation specific, meaning that general purpose operating systems may not have enough information about such a topology to correctly support bandwidth allocation. In order to circumvent this problem, these operating systems may require that a single RCRB element (of type Internal Link) not declare more than one link to a Root Complex Component other than the one containing the RCRB element itself.

The PCI Express Root Complex Link Declaration Capability, as shown in Figure 7-56, consists of the PCI Express Enhanced Capability Header and Root Complex Element Self Description followed by one or more Root Complex Link Entries.

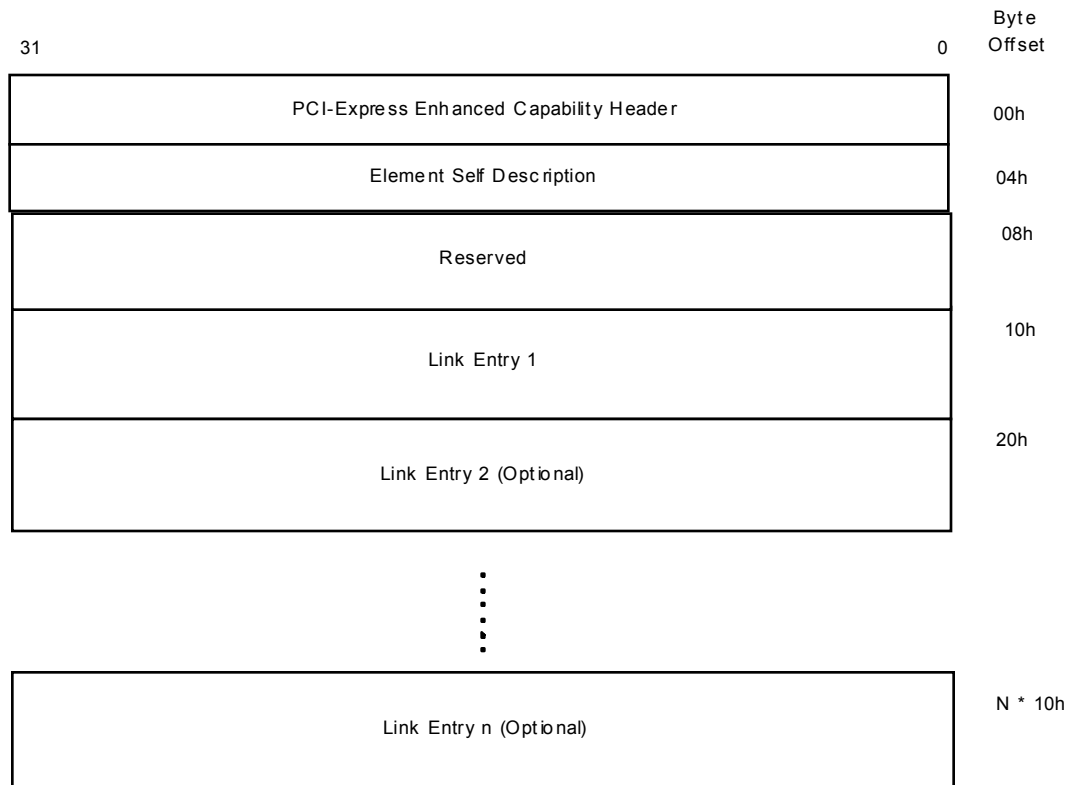


Figure 7-56 PCI Express Root Complex Link Declaration Capability

7.13.1 Root Complex Link Declaration Enhanced Capability Header

The Extended Capability ID for the Root Complex Link Declaration Capability is 0005h.

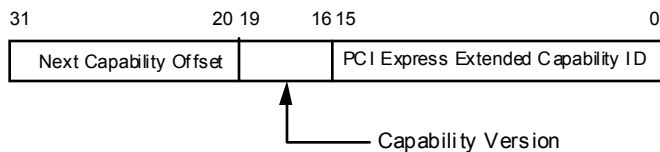


Figure 7-57 Root Complex Link Declaration Enhanced Capability Header

Table 7-50 Root Complex Link Declaration Enhanced Capability Header

Bit Location	Description	Register Attribute
15:0	PCI Express Extended Capability ID – This field is a PCI-SIG defined ID number that indicates the nature and format of the extended capability. Extended Capability ID for the Link Declaration Capability is 0005h.	RO
19:16	Capability Version – This field is a PCI-SIG defined version number that indicates the version of the capability structure present. Must be 1h for this version of the specification.	RO
31:20	Next Capability Offset – This field contains the offset to the next PCI Express capability structure or 000h if no other items exist in the linked list of capabilities. For Extended Capabilities implemented in device configuration space, this offset is relative to the beginning of PCI compatible configuration space and thus must always be either 000h (for terminating list of capabilities) or greater than 0FFh. The bottom two bits of this offset are reserved and must be implemented as 00b although software must mask them to allow for future uses of these bits.	RO

7.13.2 Element Self Description

The Element Self Description register provides information about the Root Complex element containing the Link Declaration Capability

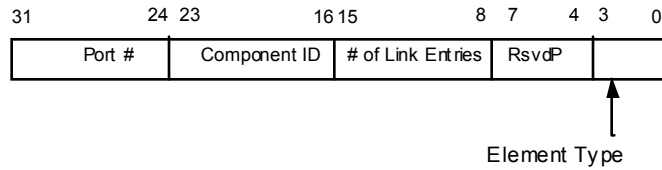


Figure 7-58 Element Self Description Register

Table 7-51 Element Self Description Register

Bit Location	Description	Register Attribute
3:0	<p>Element Type – This field indicates the type of the Root Complex Element. Defined encodings are:</p> <p>0h Configuration Space Element 1h System egress port or internal sink (memory) 2h Internal Root Complex Link</p>	RO
15:8	<p>Number of Link Entries – This field indicates the number of link entries following the Element Self Description. This field must report a value of 1 or higher.</p>	HwInit
23:16	<p>Component ID – This field identifies the Root Complex Component that contains this Root Complex Element. A value of 0 is reserved; Component IDs start at 1.</p>	HwInit
31:24	<p>Port Number – This field specifies the port number associated with this element with respect to the Root Complex Component that contains this element.</p> <p>An element with a port number of 0 indicates the default egress port to configuration software.</p>	HwInit

7.13.3 Link Entries

Link Entries start at offset 10h of the PCI Express Root Complex Link Declaration Capabilities structure. Each Link Entry consists of a link description followed by a 64-bit link address at offset 08h from the start of link entry identifying the target element for the declared link. A Link Entry declares an internal link to another Root Complex Element.

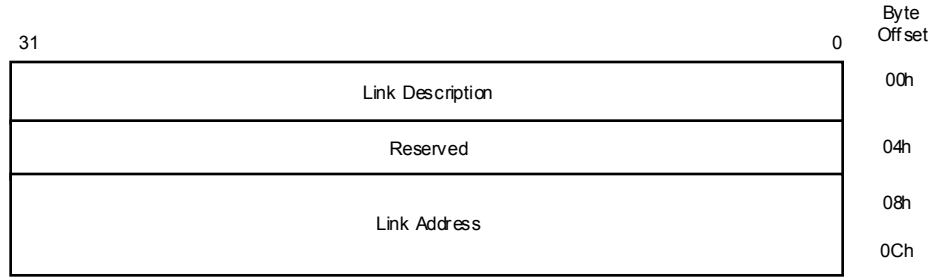


Figure 7-59 Link Entry

7.13.3.1 Link Description

The Link Description is located at offset 00h from the start of a Link Entry and is defined as follows:

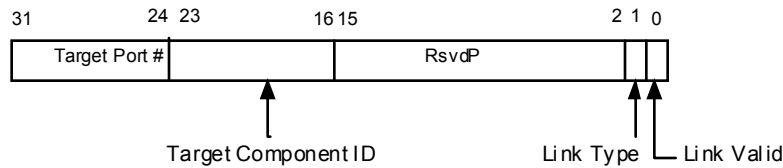


Figure 7-60 Link Description Register

Table 7-52 Link Description Register

Bit Location	Description	Register Attribute
0	Link Valid – This field when set to 1 indicates that the Link Entry specifies a valid link. Link entries that do not have this bit set are ignored by software.	Hwlnit
1	Link Type – This field indicates the target type of the link and defines the format of the link address field. Defined encodings are: 0 – Link points to memory-mapped space (for RCRB). The link address specifies the 64-bit base address of the target RCRB. 1 – Link points to configuration space (for a Root Port or Root Complex Integrated Device). The link address specifies the configuration address (segment, bus, device, function) of the target element.	Hwlnit
23:16	Target Component ID – This field identifies the Root Complex Component that is targeted by this link entry. A value of 0 is reserved; Component IDs start at 1.	Hwlnit

31:24	Target Port Number – This field specifies the port number associated with the element targeted by this link entry; the target port number is with respect to the Root Complex Component (identified by the Target Component ID) that contains the target element.	Hwlnit
-------	--	--------

7.13.3.2 Link Address

The link address is a Hwlnit field located at offset 08h from the start of a Link Entry that identifies the target element for the link entry. For a link of Link Type 0 in its Link Description, the link address specifies the memory-mapped base address of RCRB. For a link of Link Type 1 in its Link Description, the link address specifies the configuration address of a PCI Express Root Port or an Root Complex Integrated Device.

7.13.3.2.1 Link Address for Link Type 0

For a link pointing to a memory-mapped RCRB (Link Type = 0), the first DWORD specifies the lower 32-bits of the RCRB base address of the target element as shown below; bits 11:0 are hardwired to 0 and reserved for future use. The second DWORD specifies the high order 32-bits (63:32) of the base address of the target element.

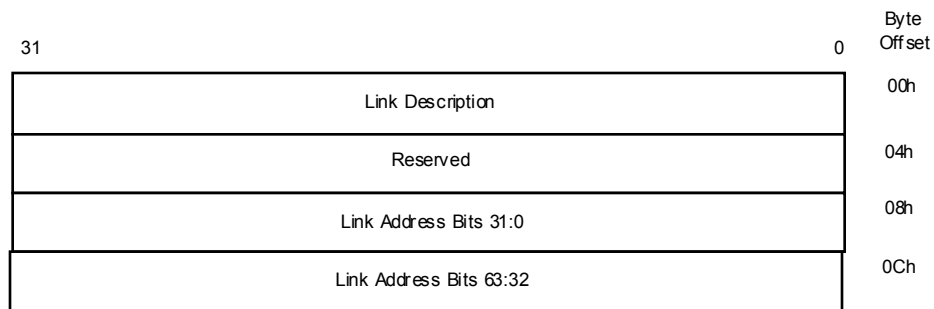


Figure 7-61 Link Address for Link Type 0

7.13.3.2.2 Link Address for Link Type 1

For a link pointing to the configuration space of a Root Complex element (Link Type = 1), Bits 27:12 of the first DWORD specify the bus, device and function number of the configuration space of the target element as shown in the table

below; bits 11:0 are reserved and hardwired to 0. Bits 31:28 of the first DWORD together with the second DWORD optionally identify the target element's hierarchy (for multi-hierarchy systems) by specifying bits 63:28 of the configuration space base address of the PCI Express hierarchy associated with the targeted element; single hierarchy systems that do not implement more than one memory mapped configuration space are allowed to report a value of 0 to indicate default configuration space.

A configuration space base address [63:28] equal to zero indicates that the configuration space address defined by bits [27:12] (bus, device number, and function number) exists in the default configuration space segment; any non-zero value indicates a separate configuration space base address.

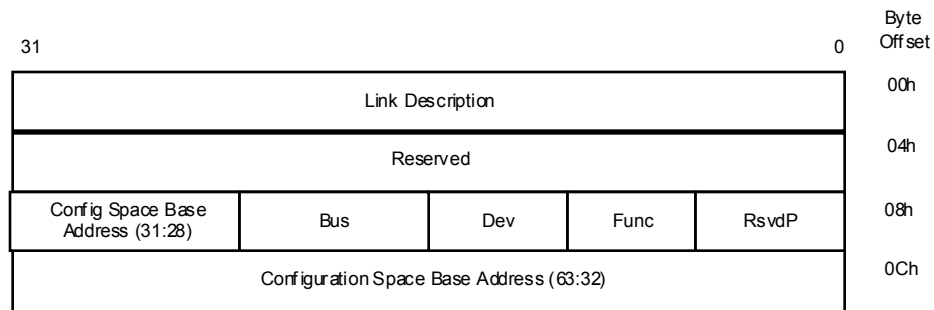


Figure 7-62 Link Address for Link Type 1

Table 7-53 Link Address for Link Type 1

Bit Location	Description	Register Attribute
14:12	Function Number	HwInit
19:15	Device Number	HwInit
27:20	Bus Number	HwInit
63:28	PCI Express Configuration Space Base Address Note: A Root Complex that does not implement multiple configuration spaces is allowed to report this field as 0.	HwInit

7.14 PCI Express Root Complex Internal Link Control Capability

The PCI Express Root Complex Internal Link Control Capability is an optional capability that controls an internal Root Complex link between two distinct Root Complex Components. This capability is valid for RCRBs that declare element type as Internal Link in the Element Self-Description register of the Root Complex Link Declaration Capability structure.

The Root Complex Internal Link Control Capability Structure is defined as shown in the following figure:

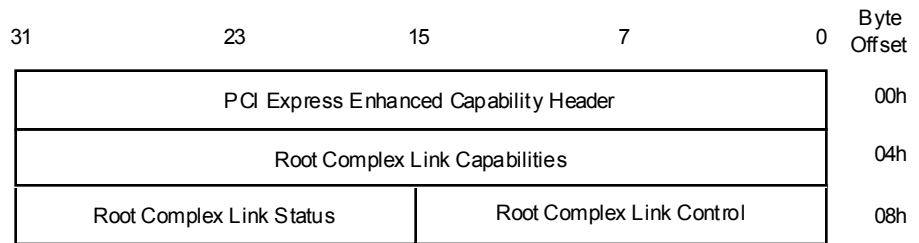


Figure 7-63 Root Complex Internal Link Control Capability

7.14.1 Root Complex Internal Link Control Enhanced Capability Header

The Extended Capability ID for the Root Complex Internal Link Control Capability is 0006h.

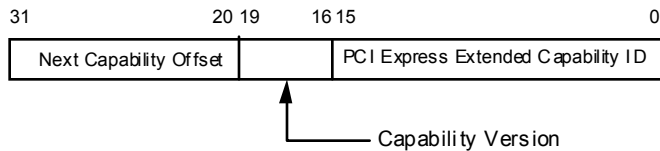


Figure 7-64 Root Complex Internal Link Control Enhanced Capability Header

Table 7-54 Root Complex Internal Link Control Enhanced Capability Header

Bit Location	Description	Register Attribute
15:0	<p>PCI Express Extended Capability ID – This field is a PCI-SIG defined ID number that indicates the nature and format of the extended capability.</p> <p>Extended Capability ID for the Internal Link Control Capability is 0006h.</p>	RO
19:16	<p>Capability Version – This field is a PCI-SIG defined version number that indicates the version of the capability structure present.</p>	RO

	Must be 1h for this version of the specification.	
31:20	<p>Next Capability Offset – This field contains the offset to the next PCI Express capability structure or 000h if no other items exist in the linked list of capabilities.</p> <p>For Extended Capabilities implemented in device configuration space, this offset is relative to the beginning of PCI compatible configuration space and thus must always be either 000h (for terminating list of capabilities) or greater than 0FFh.</p> <p>The bottom two bits of this offset are reserved and must be implemented as 00b although software must mask them to allow for future uses of these bits.</p>	RO

7.14.2 Root Complex Link Capabilities Register

The Root Complex Link Capabilities register identifies capabilities for this link.

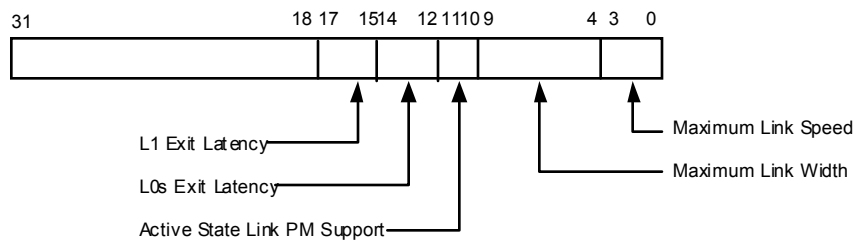


Figure 7-65 Root Complex Link Capabilities Register

Table 7-55 Root Complex Link Capabilities Register

Bit Location	Register Description	Attributes														
3:0	<p>Maximum Link Speed - This field indicates the maximum Link speed of the given Link. Defined encodings are:</p> <table border="1" style="margin-left: 20px;"> <tr> <td>0001b</td> <td>2.5 Gb/s Link</td> </tr> </table> <p>All other encodings are reserved. A Root Complex that does not support this feature reports a 0 in this field.</p>	0001b	2.5 Gb/s Link	RO												
0001b	2.5 Gb/s Link															
9:4	<p>Maximum Link Width - This field indicates the maximum width of the given Link. Defined encodings are:</p> <table border="1" style="margin-left: 20px;"> <tr> <td>000001b</td> <td>x1</td> </tr> <tr> <td>000010b</td> <td>x2</td> </tr> <tr> <td>000100b</td> <td>x4</td> </tr> <tr> <td>001000b</td> <td>x8</td> </tr> <tr> <td>001100b</td> <td>x12</td> </tr> <tr> <td>010000b</td> <td>x16</td> </tr> <tr> <td>100000b</td> <td>x32</td> </tr> </table>	000001b	x1	000010b	x2	000100b	x4	001000b	x8	001100b	x12	010000b	x16	100000b	x32	RO
000001b	x1															
000010b	x2															
000100b	x4															
001000b	x8															
001100b	x12															
010000b	x16															
100000b	x32															

	<table border="1"> <tr> <td>100000b</td> <td>x32</td> </tr> </table> <p>All other encodings are reserved. A Root Complex that does not support this feature reports a 0 in this field.</p>	100000b	x32															
100000b	x32																	
11:10	<p>Active State Link PM Support – This field indicates the level of active state power management supported on the given Link. Defined encodings are:</p> <table border="1"> <tr> <td>00b</td> <td>No Active State PM Support</td> </tr> <tr> <td>01b</td> <td>L0s Entry Supported</td> </tr> <tr> <td>10b</td> <td>L1 Entry Supported</td> </tr> <tr> <td>11b</td> <td>L0s and L1 Supported</td> </tr> </table>	00b	No Active State PM Support	01b	L0s Entry Supported	10b	L1 Entry Supported	11b	L0s and L1 Supported	RO								
00b	No Active State PM Support																	
01b	L0s Entry Supported																	
10b	L1 Entry Supported																	
11b	L0s and L1 Supported																	
14:12	<p>L0s Exit Latency - This field indicates the L0s exit latency for the given Link. The value reported indicates the length of time this Port requires to complete transition from L0s to L0. Defined encodings are:</p> <table border="1"> <tr> <td>000b</td> <td>Less than 64 ns</td> </tr> <tr> <td>001b</td> <td>64 ns to less than 128 ns</td> </tr> <tr> <td>010b</td> <td>128 ns to less than 256 ns</td> </tr> <tr> <td>011b</td> <td>256 ns to less than 512 ns</td> </tr> <tr> <td>100b</td> <td>512 ns to less than 1 μs</td> </tr> <tr> <td>101b</td> <td>1 μs to less than 2 μs</td> </tr> <tr> <td>110b</td> <td>2 μs-4 μs</td> </tr> <tr> <td>111b</td> <td>L0s transition not supported</td> </tr> </table>	000b	Less than 64 ns	001b	64 ns to less than 128 ns	010b	128 ns to less than 256 ns	011b	256 ns to less than 512 ns	100b	512 ns to less than 1 μ s	101b	1 μ s to less than 2 μ s	110b	2 μ s-4 μ s	111b	L0s transition not supported	RO
000b	Less than 64 ns																	
001b	64 ns to less than 128 ns																	
010b	128 ns to less than 256 ns																	
011b	256 ns to less than 512 ns																	
100b	512 ns to less than 1 μ s																	
101b	1 μ s to less than 2 μ s																	
110b	2 μ s-4 μ s																	
111b	L0s transition not supported																	
17:15	<p>L1 Exit Latency - This field indicates the L1 exit latency for the given Link. The value reported indicates the length of time this Port requires to complete transition from L1 to L0. Defined encodings are:</p> <table border="1"> <tr> <td>000b</td> <td>Less than 1 μs</td> </tr> <tr> <td>001b</td> <td>1 μs to less than 2 μs</td> </tr> <tr> <td>010b</td> <td>2 μs to less than 4 μs</td> </tr> <tr> <td>011b</td> <td>4 μs to less than 8 μs</td> </tr> <tr> <td>100b</td> <td>8 μs to less than 16 μs</td> </tr> <tr> <td>101b</td> <td>16 μs to less than 32 μs</td> </tr> <tr> <td>110b</td> <td>32 μs to less than 64 μs</td> </tr> <tr> <td>111b</td> <td>L1 transition not supported</td> </tr> </table>	000b	Less than 1 μ s	001b	1 μ s to less than 2 μ s	010b	2 μ s to less than 4 μ s	011b	4 μ s to less than 8 μ s	100b	8 μ s to less than 16 μ s	101b	16 μ s to less than 32 μ s	110b	32 μ s to less than 64 μ s	111b	L1 transition not supported	RO
000b	Less than 1 μ s																	
001b	1 μ s to less than 2 μ s																	
010b	2 μ s to less than 4 μ s																	
011b	4 μ s to less than 8 μ s																	
100b	8 μ s to less than 16 μ s																	
101b	16 μ s to less than 32 μ s																	
110b	32 μ s to less than 64 μ s																	
111b	L1 transition not supported																	

7.14.3 Root Complex Link Control Register

The Link Control register controls parameters for this internal link.

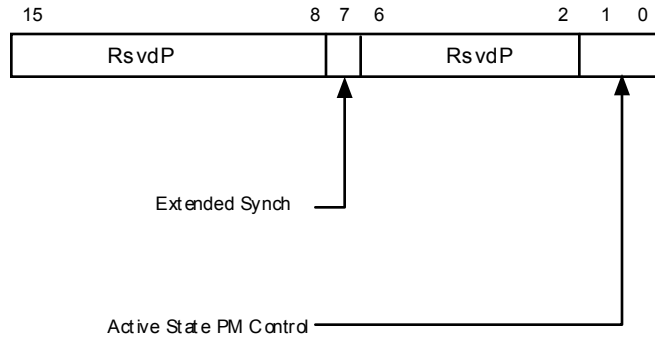


Figure 7-66 Root Complex Link Control Register

Table 7-56 Root Complex Link Control Register

Bit Location	Register Description	Attributes								
1:0	<p>Active State Link PM Control - This field controls the level of active state PM supported on the given Link. Defined encodings are:</p> <table border="1"> <tr> <td>00b</td> <td>Disabled</td> </tr> <tr> <td>01b</td> <td>L0s Entry Enabled</td> </tr> <tr> <td>10b</td> <td>L1 Entry Enabled</td> </tr> <tr> <td>11b</td> <td>L0s and L1 Entry Enabled</td> </tr> </table> <p>A Root Complex that does not support this feature for the given internal link hardwires this field to 0.</p>	00b	Disabled	01b	L0s Entry Enabled	10b	L1 Entry Enabled	11b	L0s and L1 Entry Enabled	RW
00b	Disabled									
01b	L0s Entry Enabled									
10b	L1 Entry Enabled									
11b	L0s and L1 Entry Enabled									
7	<p>Extended Synch – This bit when set forces extended transmission of FTS ordered sets in FTS and extra TS2 at exit from L1 prior to entering L0. This mode provides external devices monitoring the link time to achieve bit and symbol lock before the link enters L0 state and resumes communication. Default value for this bit is 0</p> <p>A Root Complex that does not support this feature for the given internal link hardwires this field to 0.</p>	RW								

7.14.4 Root Complex Link Status Register

The Link Status register provides information about Link specific parameters.

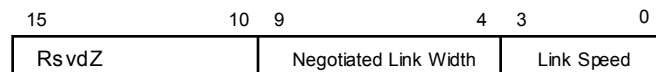


Figure 7-67 Root Complex Link Status Register

Table 7-57 Root Complex Link Status Register

Bit Location	Register Description	Attributes														
3:0	<p>Link Speed - This field indicates the negotiated Link speed of the given Link. Defined encodings are:</p> <table border="1" data-bbox="431 415 662 449"> <tr> <td>0001b</td> <td>2.5 Gb/s</td> </tr> </table> <p>All other encodings are reserved. A Root Complex that does not support this feature reports a 0 in this field.</p>	0001b	2.5 Gb/s	RO												
0001b	2.5 Gb/s															
9:4	<p>Negotiated Link Width - This field indicates the negotiated width of the given Link. Defined encodings are:</p> <table border="1" data-bbox="431 663 638 888"> <tr> <td>000001b</td> <td>X1</td> </tr> <tr> <td>000010b</td> <td>X2</td> </tr> <tr> <td>000100b</td> <td>X4</td> </tr> <tr> <td>001000b</td> <td>X8</td> </tr> <tr> <td>001100b</td> <td>X12</td> </tr> <tr> <td>010000b</td> <td>X16</td> </tr> <tr> <td>100000b</td> <td>X32</td> </tr> </table> <p>All other encodings are reserved. A Root Complex that does not support this feature reports a 0 in this field.</p>	000001b	X1	000010b	X2	000100b	X4	001000b	X8	001100b	X12	010000b	X16	100000b	X32	RO
000001b	X1															
000010b	X2															
000100b	X4															
001000b	X8															
001100b	X12															
010000b	X16															
100000b	X32															