PCI-SIG ENGINEERING CHANGE NOTICE

TITLE:	Readiness Notifications (RN)
DATE:	Last change: 27 August 2013 PWG approved for final release: 29 Aug 2013
AFFECTED DOCUMENT:	PCI Express Base Spec. Rev. 3.0, PCI Local Bus Specification Revision 3.0, PCI Bus Power Management Interface Specification Revision 1.2, Single Root I/O Virtualization and Sharing Revision 1.1, PCI Code and ID Assignment Specification Version 1.3
SPONSOR:	Dell, HP, IDT, Intel, NVIDIA, AMD

Part I

1. Summary of the Functional Changes

Defines mechanisms to reduce the time software needs to wait before issuing a Configuration Request to a PCIe Function or RC-integrated PCI Function following power on, reset, or power state transitions.

This ECN uses the PCI-SIG Defined VDM mechanism introduced in the Lightweight Notification (LN) ECN but is not otherwise related to LN.

2. Benefits as a Result of the Changes

Through the mechanisms defined by this ECR, we can avoid the long, architected, fixed delays following various forms of reset before software is permitted to perform its first Configuration Request. These delays are very large:

1 second if Configuration Retry Status (CRS) is not used

100ms for most cases if CRS is used

10ms "minimum recovery time" when a Function is programmed from $D3_{hot}$ to D0

In addition, we avoid the complexity of using the existing CRS mechanism, which potentially requires polling periodically up to 1 second following reset, by providing an explicit readiness indication.

Specific cases addressed by this ECR include the delays associated with:

Device becoming ready following DL_Down to DL_Up

- ✓ Exit from Cold Reset (initial power-up, hot-add, or D3cold)
- ✓ Exit from Warm Reset, Hot Reset, or Disabled, or Loopback
- ✓ Exit from L2/L3 Ready

Function becoming ready following

- ✓ $D3_{hot}$/D0 transition by the Function
- ✓ Completion of FLR by the Function
- ✓ Setting or Clearing of VF Enable in a PF (SR-IOV)

Readiness Notifications defines Messages to allow a Device or Function to send a Message in these cases when it's ready to respond to Configuration Requests with a Successful Completion, and Software Configuration discovery mechanisms which allow all Functions to indicate cases where they are guaranteed to be ready for configuration without requiring system software delay.

3. Assessment of the Impact

These are optional capabilities, for which new hardware and/or software elements are required.

4. Analysis of the Hardware Implications

Implementations that support RN will need to support new hardware mechanisms and capabilities as defined in this document. There is a Message-based link protocol, new configuration field(s) for existing registers and, for FRS Queuing, a new extended capability structure.

New read-only configuration bits are added to allow indication of Immediate Readiness for Configuration in certain cases.

5. Analysis of the Software Implications

No changes are required to existing software, and existing software will not benefit from this feature.

New firmware/software is required to achieve the benefits associated with RN. Note that it is possible for some firmware/software to be revised and gain benefit while older software continues to operate as-is, for example a new platform with a new system BIOS could make use of RN, but an older operating system installed on that platform would continue to operate according to the current configuration procedures once booted.

6. Analysis of the C&I Test Implications

Some previously reserved fields are now defined. New test cases would need to be created to test functionality of this feature.

Part II

Detailed Description of the change

Modify Terms and Acronyms as shown:

Terms and Acronyms

…

Configuration-Ready

A Function is "Configuration-Ready" when it is guaranteed that the Function will respond to a valid Configuration Request targeting the Function with a Completion indicating Successful Completion status.

…

Device Readiness Status, DRS

A mechanism for indicating that a Device is Configuration-Ready (see Section 6.x.1)

…

Function Readiness Status, FRS

A mechanism for indicating that a Function is Configuration-Ready (see Section 6.x.2)

Edit Section 2.2.8 as shown:

- OBFF Messages
- DRS Messages
- FRS Messages

Add Section 2.2.8.6.x as shown (Note to Reader: this references material added by the LN Protocol ECN defining PCI-SIG-Defined VDMs):

2.2.8.6.x. Device Readiness Status (DRS) Message

The Device Readiness Status (DRS) protocol (see Section 6.x.1) uses the PCI-SIG-Defined VDM mechanism (see Section 2.2.8.6.1). The DRS Message is a PCI-SIG-Defined VDM (Vendor-Defined Type 1 Message) with no payload.

Beyond the rules for other PCI-SIG-Defined VDMs, the following rules apply to the formation of DRS Messages:

- Table 2-x1 and Figure 2-x1 illustrate and define the DRS Message.
- The TLP Type must be Msg.
- The TC[2:0] field must be 000b.
- The Attr[2:0] field is Reserved.
- The Tag field is Reserved.
- The Message Routing field must be set to 100b – Local – Terminate at Receiver.

Receivers may optionally check for violations of these rules (but must not check reserved bits). These checks are independently optional (see Section 6.2.3.4). If a Receiver

implementing these checks determines that a TLP violates these rules, the TLP is a Malformed TLP.

- If checked, this is a reported error associated with the Receiving Port (see Section 6.2).

Table 2-x1: DRS Message

Name	Code[7:0] (b)	Routing r[2:0] (b)	Support				Description/Comments
			RC	Ep	Sw	Br	
DRS Message	0111 1111	100	r	t	tr		Device Readiness Status

The format the DRS Message is shown in Figure 2-x below.

+0								+1								+2								+3							
7	6	5	4	3	2	1	0	7	6	5	4	3	2	1	0	7	6	5	4	3	2	1	0	7	6	5	4	3	2	1	0
Fmt 0 0 1			Type 1 0 1 0 0					R	TC 0 0 0			R	Attr	LN	TH	TD	EP	Attr		AT 0 0		Length 00 0000 0000										
Requester ID																Tag								Message Code 0111 1111							
Reserved																Vendor ID 0000 0000 0000 0001															
Subtype 0000 1000								Reserved																							

Figure 2-x1: DRS Message

Add Section 2.2.8.6.y as shown (Note to Reader: this references material added by the LN Protocol ECN defining PCI-SIG-Defined VDMs):

2.2.8.6.y. Function Readiness Status (FRS) Message

The Function Readiness Status (FRS) protocol (see Section 6.x.2) uses the PCI-SIG-Defined VDM mechanism (see Section 2.2.8.6.1). The FRS message is a PCI-SIG-Defined VDM (Vendor-Defined Type 1 Message) with no payload.

Beyond the rules for other PCI-SIG-Defined VDMs, the following rules apply to the formation of FRS Messages:

- Table 2-x2 and Figure 2-x2 illustrate and define the FRS Message.
- The TLP Type must be Msg.
- The TC[2:0] field must be 000b.
- The Attr[2:0] field is Reserved.
- The Tag field is Reserved.
- The FRS Reason[3:0] field indicates why the FRS Message was generated:

- **0001b: DRS Message Received** - The Downstream Port indicated by the Message Requester ID received a DRS Message and has the DRS Signaling Control field in the Link Control register set to DRS to FRS Signaling Enabled
- **0010b: $D3_{hot}$ to D0 Transition Completed** - A $D3_{hot}$ to D0 transition has completed, and the Function indicated by the Message Requester ID is now Configuration-Ready and has returned to the $D0_{uninitialized}$ or $D0_{active}$ state depending on the setting of the No_Soft_Reset bit (see ***PCI Bus Power Management Interface Specification*** Section 3.2.4 and Section 5.4.1)
- **0011b: FLR Completed** - An FLR has completed, and the Function indicated by the Message Requester ID is now Configuration-Ready
- **1000b: VF Enabled** – The Message Requester ID indicates a PF - All VFs associated with that PF are now Configuration-Ready
- **1001b: VF Disabled** – The Message Requester ID indicates a PF - All VFs associated with that PF have been disabled and the SR-IOV data structures in that PF may now be accessed.
- All other values Reserved

❑ The Message Routing field must be set to 000b – Routed to Root Complex

Receivers may optionally check for violations of these rules (but must not check reserved bits). These checks are independently optional (see Section 6.2.3.4). If a Receiver implementing these checks determines that a TLP violates these rules, the TLP is a Malformed TLP.

□ If checked, this is a reported error associated with the Receiving Port (see Section 6.2).

Table 2-x2: FRS Message

Name	Code[7:0] (b)	Routing r[2:0] (b)	Support				Description/Comments
			RC	Ep	Sw	Br	
FRS Message	0111 1111	000	r	t	tr		Function Readiness Status

The format the FRS Message is shown in Figure 2-x2 below.

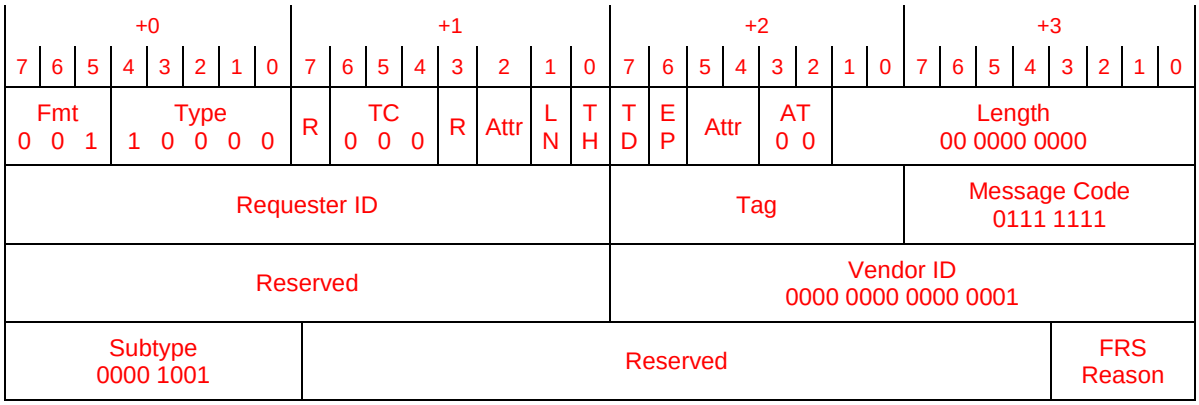


Figure 2-x2: FRS Message

Edit in Section 2.3.1 as shown:

2.3.1. Request Handling Rules

…

A device Function is explicitly not permitted to return CRS following a software-initiated reset (other than an FLR) of the device, e.g., by the device's software driver writing to a device-specific reset bit. A device Function is not permitted to return CRS after it has indicated that it is Configuration-Ready (see Section 6.x.) without an intervening valid reset (i.e., FLR or Conventional Reset) condition, or if the Immediate Readiness bit in the Function's Status register is Set. Additionally, a device Function is not permitted to return CRS after having previously returned a Successful Completion without an intervening valid reset (i.e., FLR or Conventional Reset) condition.

…

Edit within the Implementation Note "Configuration Request Retry Status" in Section 2.3.1. as shown:

…

Note that it is only legal to respond with a CRS Completion Status in response to a Configuration Request. Sending this Completion Status in response to any other Request type is illegal (see Section 2.3.2). Readiness Notifications (see Section 6.x.) and Immediate Readiness (see Conventional PCI Specification Revision 3.0 Section 6.2.3. and PCI-PM Specification Revision 1.2 Section 3.2.3) also forbid the use of CRS Completion Status in certain situations.

…

Edit within Section 5.3.1.4.. as shown:

5.3.1.4. D3 State

…

Unless the Immediate_Readiness_on_Return_to_D0 bit in the PCI-PM Power Management Capabilities register is Set, ~~S~~system software must allow a minimum recovery time following a $D3_{hot} \rightarrow D0$ transition of at least 10 ms, prior to accessing the Function.

…

Edit in Section 6.6 subsections as shown:

6.6.1. Conventional Reset

…

The second set of rules addresses requirements placed on the system:

- To allow components to perform internal initialization, system software must wait a specified minimum period following the end of a Conventional Reset of one or more devices before it is permitted to issue Configuration Requests to those devices, unless Readiness Notifications mechanisms are used (see Section 6.x.).

…

- Unless Readiness Notifications mechanisms are used (see Section 6.x.), ~~T~~the Root Complex and/or system software must allow at least 1.0 s after a Conventional Reset of a device, before it may determine that a device which fails to return a Successful Completion status for a valid Configuration Request is a broken device. This period is independent of how quickly Link training completes.

6.6.2. Function-Level Reset (FLR)

…

After an FLR has been initiated by writing a 1b to the Initiate Function Level Reset bit, the Function must complete the FLR within 100 ms. If software initiates an FLR when the Transactions Pending bit is 1b, then software must not initialize the Function until allowing adequate time for any associated Completions to arrive, or to achieve reasonable certainty that any remaining Completions will never arrive. For this purpose, it is recommended that software allow as much time as provided by the pre-FLR value for Completion Timeout on the device. If Completion Timeouts were disabled on the Function when FLR was issued, then the delay is system dependent but must be no less than 100 ms. If Function Readiness Status (FRS – see Section 6.x.2) is implemented, then system software is permitted to issue Configuration Requests to the Function immediately following receipt of an FRS Message indicating Configuration-Ready, however, this does not necessarily indicate that outstanding Requests initiated by the Function have completed.

…

Add new section in Chapter 6 as follows:

6.x Readiness Notifications (RN)

Readiness Notifications (RN) is intended to reduce the time software needs to wait before issuing Configuration Requests to a Device or Function following DRS Events or FRS Events. RN includes both the Device Readiness Status (DRS) and Function Readiness Status (FRS) mechanisms. These mechanisms provide a direct indication of Configuration-Readiness (see Terms and Acronyms entry for "Configuration-Ready"). When used, DRS and FRS allow an improved behaviour over the CRS mechanism, and eliminate its associated periodic polling time of up to 1 second following a reset.

It is permitted that system software/firmware provide mechanisms that supersede the FRS and/or DRS mechanisms, however such software/firmware mechanisms are outside the scope of this specification.

IMPLEMENTATION NOTE

Optimizing Configuration Readiness

It is strongly recommended that implementers of system firmware/software avoid unnecessary delays wherever possible. It is strongly recommended that hardware be designed to eliminate or minimize required delays, and to make full use of the mechanisms provided in this specification and related specifications to communicate what, if any, delays are required. Hardware implementers should appropriately document implementation behavior to enable system firmware/software to implement optimal behaviors.

Even with good documentation, some cases may at first appear problematic – for example, how can system firmware benefit from the Device Readiness Status (DRS) mechanism, when it is necessary to read from Root Port Configuration space to do so? In such cases, platform specific knowledge is required, i.e. that the Root Port supports Immediate Readiness.

6.x.1 Device Readiness Status (DRS)

When implemented, DRS must be used to indicate when a Device is Configuration-Ready following any of the following Device-level occurrences, which are subsequently referred to as "DRS Events":

- Exit from Cold Reset
- Exit from Warm Reset, Hot Reset, Loopback, or Disabled
- Exit from L2/L3 Ready
- Any other scenario where the Port transitions from DL_Down to DL_Up status.

The DRS Message protocol requirements include the following:

- There is no enable or disable mechanism for DRS. For Downstream Ports that support DRS, the DRS Supported bit in the Link Capabilities 2 register must be Set. For Upstream Ports that support DRS, it is strongly recommended that the DRS Supported bit in the Link Capabilities 2 register be Set. It is expressly permitted for Upstream Ports to send DRS Messages even when the DRS Supported bit is Clear.
- A DRS Message must be transmitted by a DRS-capable Upstream Port following every DL_Down to DL_Up transition when all non-VF Functions on the Logical Bus associated with that Upstream Port become ready.
    - A Type 0 Function is ready when it is Configuration-Ready.
    - A Type 1 Function that is a Switch Upstream Port is ready when it is Configuration-Ready and all Functions on its secondary bus are Configuration-Ready.
    - A Type 1 Function that is not a Switch Upstream Port is ready when the Function itself is Configuration-Ready.

- ❑ After a Device transmits a DRS Message, non-VF Functions indicated as Configuration-Ready by that DRS Message must not return Completions with CRS unless a subsequent DRS Event occurs.

Additional requirements relating to Switches implementing DRS include:

- ❑ Must support DRS functionality in all Ports
- ❑ Implementation at each Downstream Port of the DRS Signaling Control field.
- ❑ For any physically-integrated Device that appears beneath a Switch Downstream Port, the DRS sent by the Switch does not indicate Configuration Readiness for that Device
    - o For such a Device, implementation and use of DRS is independent of the Switch

Additional requirements for Root Ports and Switch Downstream Ports include:

- ❑ Implementation of the DRS Message Received bit, which indicates receipt of a DRS Message

6.x.2 Function Readiness Status (FRS)

When implemented, FRS must be used to indicate a specific Function as being Configuration-Ready following any of the following Function-level occurrences, which are subsequently referred to as "FRS Events":

- ❑ Function Level Reset (FLR)
- ❑ Completion of $D3_{hot}$ to D0 transition
- ❑ Setting or Clearing of VF Enable in a PF (SR-IOV)

The FRS Message protocol requirements include the following:

- ❑ The Requester ID of the FRS Message must indicate the Function that has changed readiness status (see section 2.2.8.6.y)
- ❑ The FRS Reason field in the FRS Message must indicate why that Function changed readiness status.
- ❑ After a Function transmits an FRS Message, the indicated Function(s) must not return Completions with CRS unless a subsequent DRS Event or FRS Event occurs

Additional requirements for Switches implementing FRS include:

- ❑ Must support FRS functionality in the Upstream Port and all Downstream Ports
- ❑ The ability to transmit FRS Messages Upstream when required by the FRS protocol

Additional requirements for Physical Functions (PFs) include:

- ❑ The ability to transmit FRS Message Upstream when the VF Enable or VF Disable process completes

Additional requirements for Root Ports and Root Complex Event Collectors implementing FRS include:

- ❑ Must implement the FRS Queuing Extended Capability (see Section 7.x)

6.x.3 FRS Queuing

Root Ports and Root Complex Event Collectors that support FRS must implement the FRS Queuing Extended Capability (see Section 7.x).

For a Root Port, the FRS Message Queue contains FRS Messages received by the Root Port or generated by the Root Port.

For a Root Complex Event Collector, the FRS Message Queue contains FRS Messages generated by Root Complex Integrated Endpoints associated with the Root Complex Event Collector (see Section 7.17) or generated by the Root Complex Event Collector.

The FRS Message Queue must satisfy the following requirements:

- ❑ The FRS Message Queue must be empty following Reset.
- ❑ For a Root Port, the FRS Message Queue must be emptied when the Link goes to DL_Down.
- ❑ FRS Messages must be queued in the order received.
- ❑ If the FRS Message Queue is not full at the time an FRS Message is received or is internally generated, that FRS Message must be entered in the queue and the FRS Message Received bit must set to 1b.
- ❑ If the FRS Message Queue is full at the time an FRS Message is received or is internally generated, that FRS Message must be discarded and the FRS Message Overflow bit must be set to 1b. The pre-existing FRS Message Queue must be preserved.
- ❑ The oldest FRS Message must be visible in the FRS Message Queue register (see Section 7.x.4).
- ❑ Writing the FRS Message Queue register must remove the oldest element from the queue.
- ❑ When either FRS Message Received or FRS Message Overflow transitions from 0b to 1b, an interrupt must be generated if enabled.

In Section 7.5.3.6. edit as shown:

7.5.3.6. Bridge Control Register (Offset 3Eh)

…

Bit Location	Register Description	Attributes
…	…	…
6	**Secondary Bus Reset** – Setting this bit triggers a hot reset on the corresponding PCI Express Port. Software must ensure a minimum reset duration (T_{rst}) as defined in the PCI Local Bus Specification. Software and systems must honor first-access following-reset timing requirements defined in Section 6.6, unless the Readiness Notifications mechanism (see Section 6.x) is used or if the Immediate Readiness bit in the relevant Function's Status register is Set. Port configuration registers must not be changed, except as required to update Port status. Default value of this bit is 0b.	RW
…	…	…

…

Modify Section 7.8.7 as shown:

7.8.7. Link Control Register (Offset 10h)

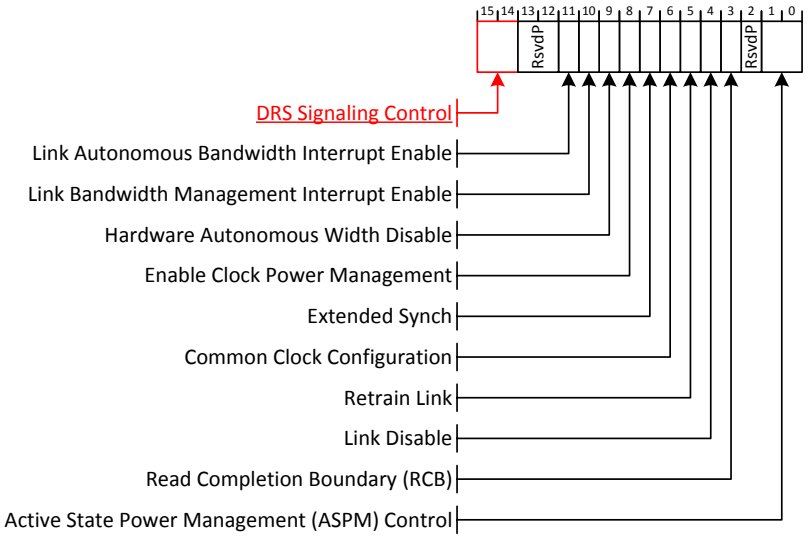


Figure 7-17: Link Control Register

Table 7-16: Link Control Register

Bit Location	Register Description	Attributes
…	…	…
15:14	DRS Signaling Control – Indicates the mechanism used to report reception of a DRS Message. Must be implemented for Downstream Ports with the DRS Supported bit Set in the Link Capabilities 2 register. Encodings are: 00b DRS not Reported: If DRS Supported is Set, receiving a DRS Message will set DRS Message Received in the Link Status 2 register but will otherwise have no effect 01b DRS Interrupt Enabled: If the DRS Message Received bit in the Link Status 2 register transitions from 0 to 1, and either MSI or MSI-X is enabled, an MSI or MSI-X interrupt is generated using the vector in Interrupt Message Number (section 7.8.2) 10b DRS to FRS Signaling Enabled: If the DRS Message Received bit in the Link Status 2 register transitions from 0 to 1, the Port must send an FRS Message Upstream with the FRS Reason field set to DRS Message Received Behavior is undefined if this field is set to 10b and the FRS Supported bit in the Device Capabilities 2 register is Clear. Behavior is undefined if this field is set to 11b. Downstream Ports with the DRS Supported bit Clear in the Link Capabilities 2 register must hardwire this field to 00b. This field is Reserved for Upstream Ports. Default value of this field is 00b.	RW/RsvdP

Modify Section 7.8.15 as shown:

7.8.15. Device Capabilities 2 Register (Offset 24h)

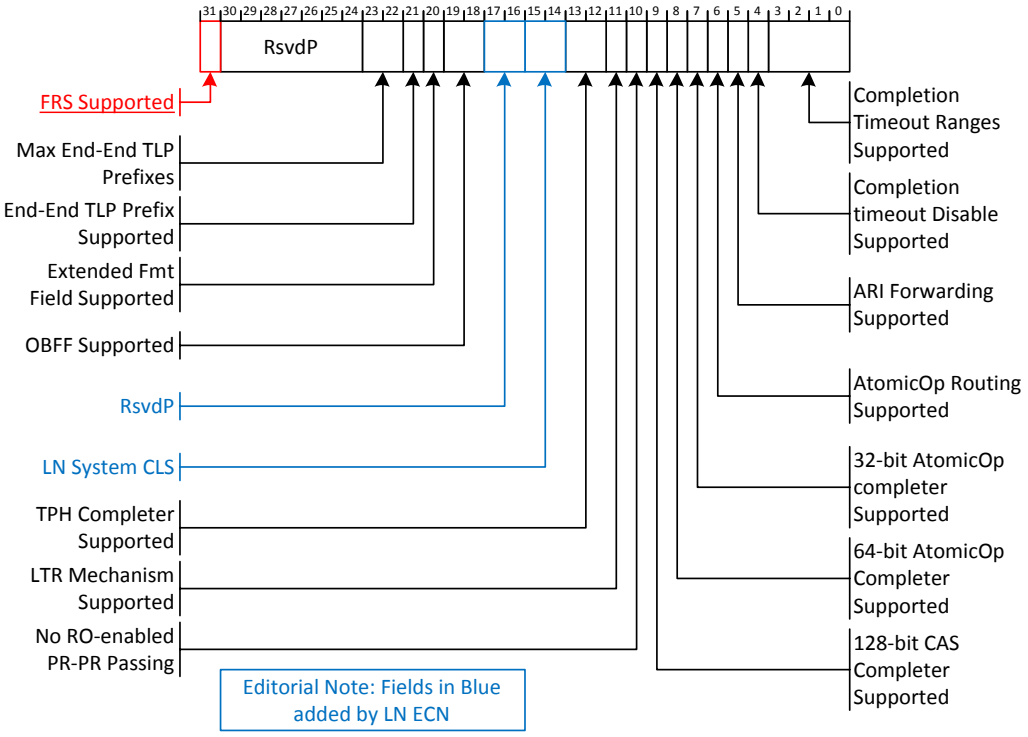


Figure 7-25: Device Capabilities 2 Register

5

Table 7-24: Device Capabilities 2 Register

Bit Location	Register Description	Attributes
…	…	…
31	**FRS Supported** – When Set, indicates support for the optional Function Readiness Status (FRS) capability. Must be Set for all Functions that support generation or receipt capabilities of FRS Messages. Must not be Set by Switch Functions that do not generate FRS Messages on their own behalf.	HwInit

Modify Section 7.8.18 as shown:

7.8.18. Link Capabilities 2 Register (Offset 2Ch)

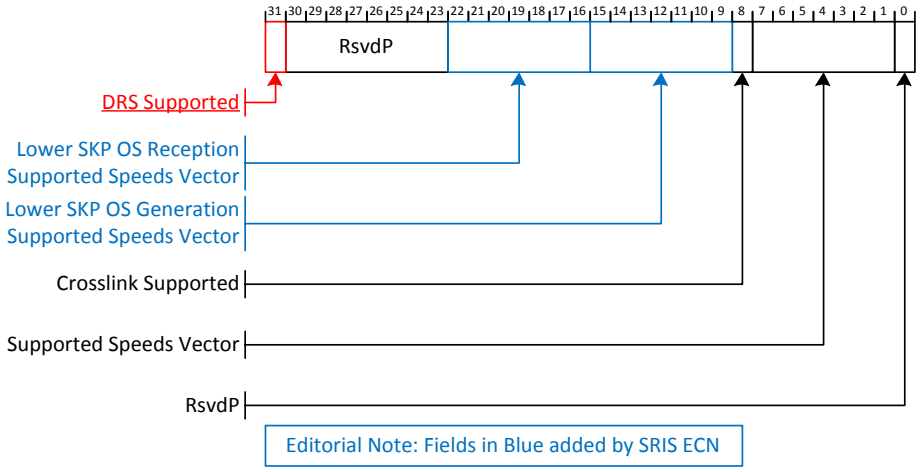


Figure 7-27: Link Capabilities 2 Register

5

Table 7-26: Link Capabilities 2 Register

Bit Location	Register Description	Attributes
…	…	…
31	DRS Supported – When Set, indicates support for the optional Device Readiness Status (DRS) capability. Must be Set in Downstream Ports that support DRS. Must be Set in Downstream Ports that support FRS. For Upstream Ports that support DRS, it is strongly recommended that this bit be Set in Function 0. For all other Functions associated with an Upstream Port, this bit must be Clear.[1] Must be Clear in Functions that are not associated with a Port. RsvdP in all other Functions.	HwInit/RsvdP

[1] It is expressly permitted for Upstream Ports to send DRS Messages even when the DRS Supported bit is Clear.

Modify Section 7.8.20 as shown:

7.8.20. Link Status 2 Register (Offset 32h)

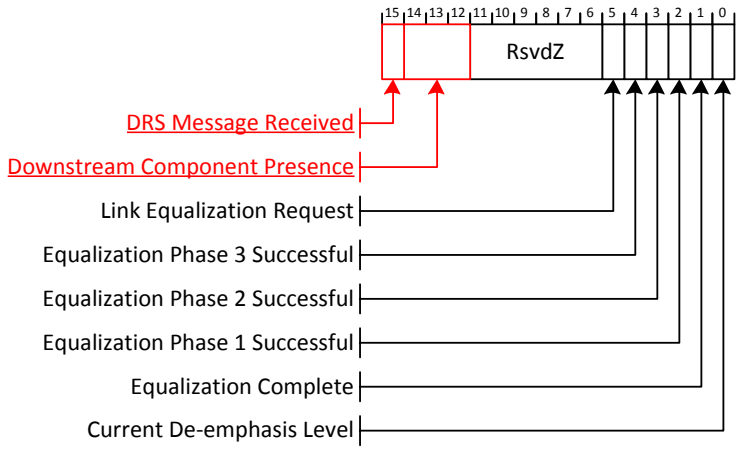


Figure 7-29: Link Status 2 Register

Table 7-28: Link Status 2 Register

Bit Location	Register Description	Attributes
…	…	…
14:12	**Downstream Component Presence** – This field indicates the presence and DRS status for the Downstream Component, if any, connected to the Link; defined values are: 000b Link Down – Presence Not Determined 001b Link Down – Component Not Present indicates the Downstream Port (DP) has determined that a Downstream Component is not present 010b Link Down – Component Present indicates the DP has determined that a Downstream Component is present, but the Data Link Layer is not active 011b Reserved 100b Link Up – Component Present indicates the DP has determined that a Downstream Component is present, but no DRS Message has been received since the Data Link Layer became active 101b Link Up – Component Present and DRS Received indicates the DP has received a DRS Message since the Data Link Layer became active 110b Reserved 111b Reserved Component Presence state must be determined by the logical "OR" of the Physical Layer in-band presence detect mechanism and, if present, any out-of-band presence detect mechanism implemented for the Link. If no out-of-band presence detect mechanism is implemented, then Component Presence state must be determined solely by the Physical Layer in-band presence detect mechanism. This field must be implemented in any Downstream Port where the DRS Supported bit is Set in the Link Capabilities 2 register. This field is RsvdZ for all other Functions. Default value of this field is 000b.	RO/RsvdZ
15	**DRS Message Received** – This bit must be Set whenever the Port receives a DRS Message. This bit must be Cleared in DL_Down. This bit must be implemented in any Downstream Port where the DRS Supported bit is Set in the Link Capabilities 2 register. This bit is RsvdZ for all other Functions. Default value of this bit is 0b.	RW1C / RsvdZ

Add the following sections to Chapter 7:

7.x. Function Readiness Status (FRS) Queuing Extended Capability

The Function Readiness Status (FRS) Queuing Extended Capability is required for Root Ports and Root Complex Event Collectors that support the optional normative FRS Queuing capability. See Section 6.x. This extended capability is only permitted in Root Ports and Root Complex Event Collectors.

If this capability is present in a Function, that Function must also implement either MSI, MSI-X, or both.

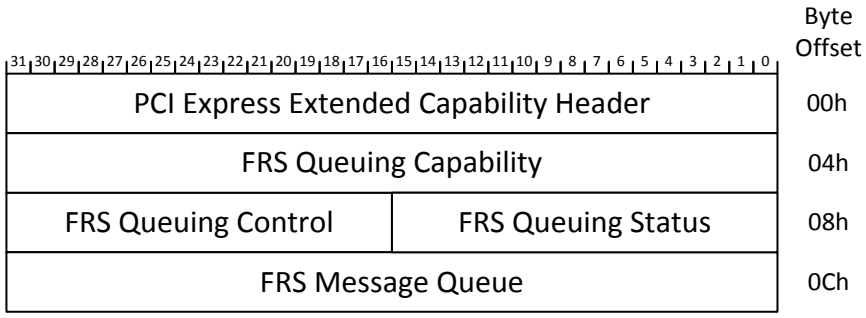


Figure 7-x1 FRS Extended Capability

7.x.1. Function Readiness Status (FRS) Queuing Extended Capability Header (Offset 00h)

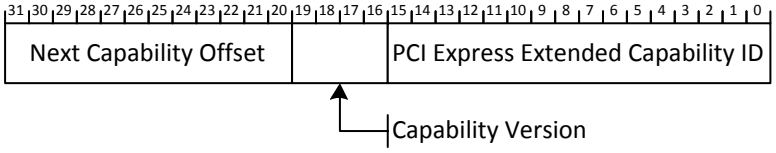


Figure 7-x2 FRS Queuing Extended Capability Header

Table 7-x1 FRS Queuing Extended Capability Header

Bit Location	Register Description	Attributes
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. PCI Express Extended Capability ID for the FRS Queuing Extended Capability is 0021h.	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 Extended Capability structure or 000h if no other items exist in the linked list of capabilities.	RO

7.x.2. FRS Queuing Capability Register (Offset 04h)

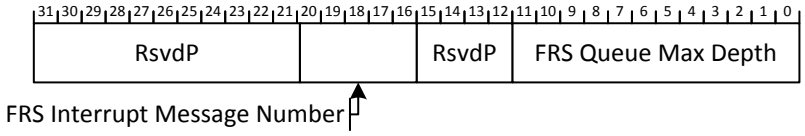
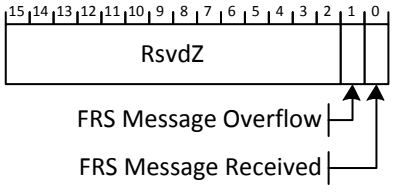


Figure 7-x3 FRS Queuing Capability Register

Table 7-x2 FRS Queuing Capability Register

Bit Location	Register Description	Attributes
11:0	**FRS Queue Max Depth** – Indicates the implemented queue depth, with valid values ranging from 001h (queue depth of 1) to FFFh (queue depth of 4095). The value of FRS Message Queue Depth must not exceed this value. The value 000h is Reserved.	HwInit
20:16	**FRS Interrupt Message Number** – This register indicates which MSI/MSI-X vector is used for the interrupt message generated in association with FRS Message Received or FRS Message Overflow. For MSI, the value in this register indicates the offset between the base Message Data and the interrupt message that is generated. Hardware is required to update this field so that it is correct if the number of MSI Messages assigned to the Function changes when software writes to the Multiple Message Enable field in the MSI Message Control register. For MSI-X, the value in this register indicates which MSI-X Table entry is used to generate the interrupt message. The entry must be one of the first 32 entries even if the Function implements more than 32 entries. For a given MSI-X implementation, the entry must remain constant. If both MSI and MSI-X are implemented, they are permitted to use different vectors, though software is permitted to enable only one mechanism at a time. If MSI-X is enabled, the value in this register must indicate the vector for MSI-X. If MSI is enabled or neither is enabled, the value in this register must indicate the vector for MSI. If software enables both MSI and MSI-X at the same time, the value in this register is undefined.	RO

7.x.3. FRS Queuing Status Register (Offset 08h)



5 **Figure 7-x4 FRS Queueing Status Register**

Table 7-x3 FRS Queuing Status Register

Bit Location	Register Description	Attributes
0	**FRS Message Received** – This bit is Set when a new FRS Message is received or generated by this Root Port or Root Complex Event Collector. Root Ports must Clear this bit when the Link is DL_Down. Default value of this bit is 0b.	RW1C
1	**FRS Message Overflow** – This bit is Set if the FRS Message queue is full and a new FRS Message is received or generated by this Root Port or Root Complex Event Collector. Root Ports must Clear this bit when the Link is DL_Down. Default value of this bit is 0b.	RW1C
15:2	**Reserved**	RsvdZ

7.x.4. FRS Queuing Control Register (Offset 0Ah)

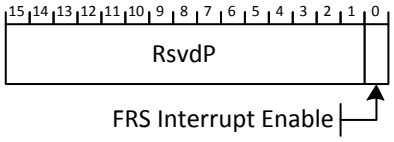


Figure 7-x5 FRS Queueing Control Register

Table 7-x4 FRS Queuing Control Register

Bit Location	Register Description	Attributes
0	**FRS Interrupt Enable** – When Set and MSI or MSI-X is enabled, the Port must issue an MSI/MSI-X interrupt to indicate the 0b to 1b transition of either the FRS Message Received or the FRS Message Overflow bits. Default value of this bit is 0b.	RW
15:1	**Reserved**	RsvdP

7.x.5. FRS Message Queue Register (Offset 0Ch)

The FRS Message Queue Register contains fields from the oldest FRS message in the queue. It also indicates the number of FRS messages in the queue.

A write of any value that includes byte 0 to this register removes the oldest FRS Message from the queue and updates these fields. A write to this register when the queue is empty has no effect.

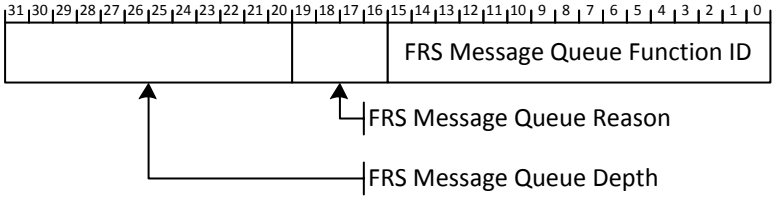


Figure 7-x6 FRS Message Queue Register

Table 7-x5 FRS Message Queue Register

Bit Location	Register Description	Attributes
15:0	**FRS Message Queue Function ID** – Recorded from the Requester ID of the oldest FRS Message received or generated by this Root Port or Root Complex Event Collector and still in the queue. Undefined if FRS Message Queue Depth is 000h.	RO
19:16	**FRS Message Queue Reason** – Recorded from the FRS Reason of the oldest FRS Message received or generated by this Root Port or Root Complex Event Collector and still in the queue. Undefined if FRS Message Queue Depth is 000h.	RO
31:20	**FRS Message Queue Depth** – indicates the current number of FRS Messages in the queue. The value of 000h indicates an empty queue. Default value of this field is 000h.	RO

5

Insert new section 7.y as follows:

7.y Readiness Time Reporting Extended Capability

The Readiness Time Reporting Extended Capability provides an optional mechanism for describing the time required for a Device or Function to become Configuration-Ready. In the indicated situations, software is permitted to issue Requests to the Device or Function after waiting for the time advertised in this capability and need not wait for the (longer) times required elsewhere.

Software is permitted to issue Requests upon the earliest of:

- Receiving a Readiness Notifications message (see Section 6.x).
- Waiting the appropriate time as specified in this document or in applicable specifications including the PCI Local Bus Specification, PCI Bus Power Management Interface Specification.
- Waiting the time indicated in the associated field of this capability.
- Waiting the time defined by system software or firmware[2].

Software is permitted to cache values from this capability and to use those cached values when the device topology has not changed.

This capability is permitted to be implemented in all Functions.

A Function must be Configuration-Ready if:

- The Immediate Readiness bit is Clear and at least **Reset Time** has elapsed after the completion of Conventional Reset
    - If the Immediate Readiness bit is Set, **Reset Time** does not apply, and is Reserved
- The Function is associated with an Upstream Port and at least **DL_Up Time** has elapsed after the Downstream Port above that Function reported Data Link Layer Link Active (see Section 7.8.8)
- The Function supports Function Level Reset and at least **FLR Time** has elapsed after that Function was issued a Function Level Reset
- Immediate_Readiness_on_Return_to_D0 is Clear and at least **$D3_{hot}$ to D0 Time** has elapsed after that Function was directed to the D0 state from $D3_{hot}$.
    - If the Immediate_Readiness_on_Return_to_D0 bit is Set, **$D3_{hot}$ to D0 Time** does not apply, and is Reserved

When Immediate_Readiness_on_Return_to_D0 is Clear, a Function must be Configuration-Ready when at least **$D3_{hot}$ to D0 Time** has elapsed after the Function was directed to the D0 state from $D3_{hot}$. In addition, the Function must be in either the $D0_{uninitialized}$ or $D0_{active}$ state, depending on the value of the No_Soft_Reset bit.

For VFs additional behavior is defined in the *Single Root I/O Virtualization and Sharing Specification*.

---

[2] For example, using ACPI tables to provide the equivalent of this capability.

If the above conditions do not apply, Function behavior is not determined by the Readiness Time Reporting Extended Capability, and the Function must respond as defined elsewhere (including, for example, no response or a response with Configuration Retry Status).

The time values reported are determined by implementation-specific mechanisms. A valid bit is defined in this capability to permit a device to defer reporting time values, for example to allow hardware initialization through driver-based mechanisms. If the Valid bit remains Clear and 1 minute has elapsed after device driver(s) have started, software is permitted to assume that no values will be reported.

Registers and fields in the Readiness Time Reporting Extended Capability are shown in Figure 7-y1. Time values are encoded in floating point as shown in Figure 7-y2. The actual time value is $Value \times Multiplier[Scale]$. For example, the value A1Eh represents about 1 second (actually 1.006 sec) and the value 40Ah represents about 10 ms (actually 10.240 ms).

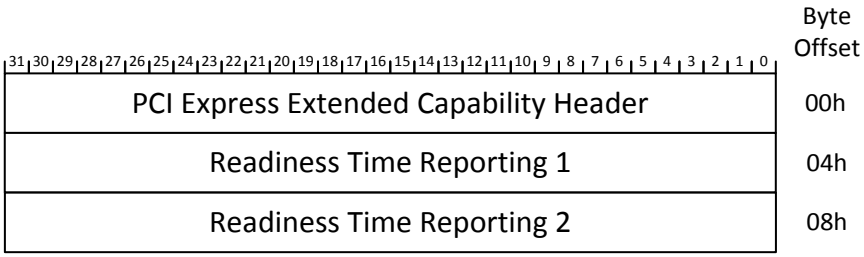
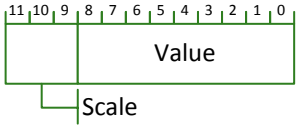


Figure 7-y1: Readiness Time Reporting Extended Capability



Scale	Multiplier
0	1 ns
1	32 ns
2	1,024 ns
3	32,768 ns
4	1,048,576 ns
5	33,554,432 ns
6	Reserved
7	Reserved

Multiplier = 32^{scale} ns

Figure 7-y2: Readiness Time Encoding

7.y.1 Readiness Time Reporting Extended Capability Header (Offset 00h)

Figure 7-y3 and Table 7-y1 detail allocation of fields in the Extended Capability header.

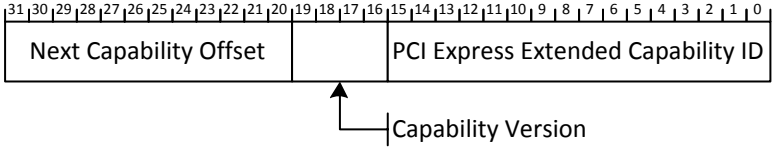


Figure 7-y3: Readiness Time Reporting Extended Capability Header

Table 7-y1: Readiness Time Reporting Extended Capability Header

Bit Location	Register Description	Attributes
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 Readiness Time Reporting Extended Capability is 0022h.	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 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.	RO

7.y.2 Readiness Time Reporting 1 (Offset 04h)

Figure 7-yx4 and Table 7-y2 detail allocation of fields in the Readiness Time Reporting 1 Register.

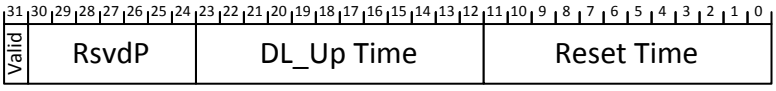


Figure 7-y4: Readiness Time Reporting 1 Register

Table 7-y2: Readiness Time Reporting 1 Register

Bit Location	Register Description	Attributes
11:0	**Reset Time** – is the time the Function requires to become Configuration-Ready after the completion of Conventional Reset. This field is RsvdP if the Immediate Readiness bit is Set. This field is undefined when the Valid bit is Clear. This field must be less than or equal to the encoded value A1Eh.	HwInit/RsvdP
23:12	**DL_Up Time** – is the time the Function requires to become Configuration-Ready after the Downstream Port above the Function reports Data Link Layer Link Active. This field is RsvdP in Functions that are not associated with an Upstream Port. This field is undefined when the Valid bit is Clear. This field must be less than or equal to the encoded value A1Eh.	HwInit/RsvdP
30:24	**Reserved**	RsvdP
31	**Valid** – If Set, indicates that all time values in this capability are valid. If Clear, indicates that the time values in this capability are not yet available. Time values may depend on device configuration. Device specific mechanisms, possibly involving the device driver(s), could be involved in determining time values. If this bit remains Clear and 1 minute has elapsed after all associated device driver(s) have started, software is permitted to assume that this bit will never be set.	HwInit

7.y.3 Readiness Time Reporting 2 (Offset 08h)

Figure 7-y5 and Table 7-y3 detail allocation of fields in the Readiness Time Reporting 2 Register.

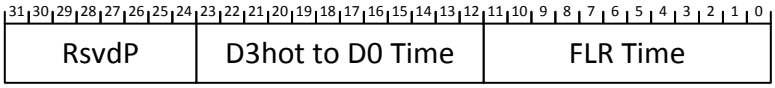


Figure 7-y5: Readiness Time Reporting 2 Register

Table 7-y3: Readiness Time Reporting 2 Register

Bit Location	Register Description	Attributes
11:0	**FLR Time** – is the time that the Function requires to become Configuration-Ready after it was issued an FLR. This field is RsvdP when the Function Level Reset Capability bit is Clear (see Section 7.8.3). This field is undefined when the Valid bit is Clear. This field must be less than or equal to the encoded value A1Eh.	HwInit/RsvdP
23:12	**$D3_{hot}$ to D0 Time** – If Immediate_Readiness_on_Return_to_D0 is Clear, $D3_{hot}$ to D0 Time is the time that the Function requires after it is directed from $D3_{hot}$ to D0 before it is Configuration-Ready and has returned to either $D0_{unitialized}$ or $D0_{active}$ state (see the *PCI Bus Power Management Interface Specification*). This field is RsvdP if the Immediate_Readiness_on_Return_to_D0 bit is Set. This field is undefined when the Valid bit is Clear. This field must be less than or equal to the encoded value 40Ah.	HwInit/RsvdP
31:24	**Reserved**	RsvdP

5

In the Conventional PCI Specification, Section 6.2.3. Device Status, change as shown:

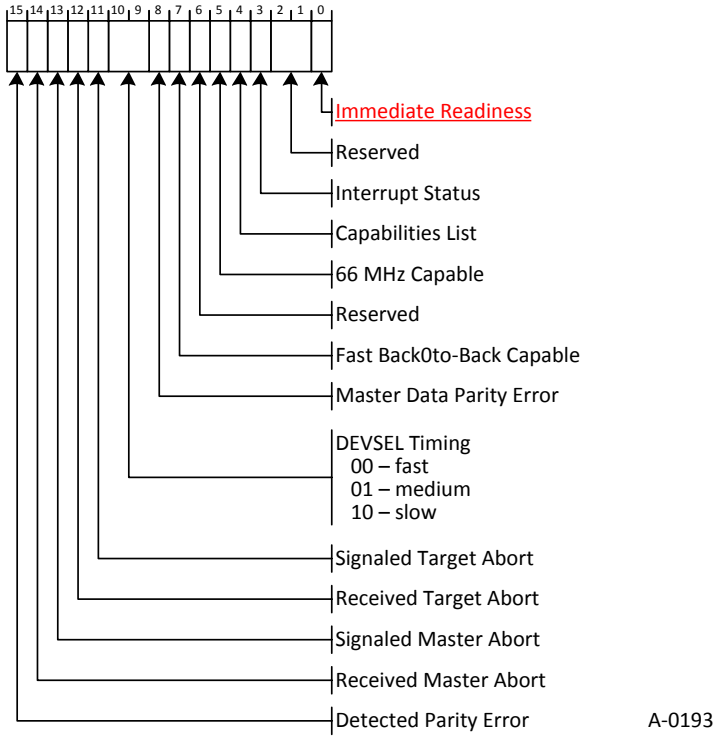


Figure 6-3: Status Register Layout

Table 6-2: Status Register Bits

Bit Location	Description
0	**Immediate Readiness** – This optional read-only bit, when Set, indicates the Function is guaranteed to be ready to successfully complete valid configuration accesses at any time following any reset that the host is capable of issuing Configuration Requests to this Function. When this bit is set, for accesses to this Function, software is exempt from all requirements to delay configuration accesses following any type of reset, including but not limited to the timing requirements defined in Section 4.3.2. How this guarantee is established is beyond the scope of this document. It is permitted that system software/firmware provide mechanisms that supersede the indication provided by this bit, however such software/firmware mechanisms are outside the scope of this specification.
…	…

In the PCI-PM Specification (1.2), Section 3.2.3. PMC - Power Management Capabilities (Offset = 2), change as shown:

Bits	Default Value	Read/ Write	Description
…	…	…	…
04	~~0b~~ Device Specific	Read Only	~~**Reserved**~~ **Immediate Readiness on Return to D0** - If this bit is a "1", this Function is guaranteed to be ready to successfully complete valid accesses immediately after being set to D0. These accesses include Configuration cycles, and if the Function returns to $D0_{initialized}$, they also include Memory and I/O Cycles. When this bit is "1", for accesses to this Function, software is exempt from all requirements to delay accesses following a transition to D0, including but not limited to the 10 ms delay defined in Section 5.4, and the delays described in Section 5.6. How this guarantee is established is beyond the scope of this document. It is permitted that system software/firmware provide mechanisms that supersede the indication provided by this bit, however such software/firmware mechanisms are outside the scope of this specification.
…	…	…	…

5

In the Single Root I/O Virtualization and Sharing Specification, Section 3.3.3.1. VF Enable, change as shown:

3.3.3.1. VF Enable

VF Enable manages the assignment of VFs to the associated PF. If VF Enable is Set, the VFs associated with the PF are accessible in the PCI Express fabric. When Set, VFs respond to and may issue PCI Express transactions following the rules for PCI Express Endpoint Functions.

If VF Enable is Clear, VFs are disabled and not visible in the PCI Express fabric; Requests to these VFs shall receive UR and these VFs shall not issue PCI Express transactions.

To allow components to perform internal initialization, ~~system software must wait for at least 100 ms~~ after changing the VF Enable bit from a 0 to a 1, the system is not permitted to issue Requests to the VFs which are enabled by that VF Enable bit until one of the following is true:

- At least 100 ms has passed
- An FRS Message has been received from the PF with a Reason Code of VF Enabled
- At least VF Enable Time has passed. VF Enable Time is either (1) the Reset Time value in the Readiness Time Reporting capability associated with the VF, or (2) a value determined by system software / firmware[3].

The Root Complex and/or system software must allow at least 1.0 s after Setting the VF Enable bit, before it may determine that a VF which fails to return a Successful Completion status for a valid Configuration Request is broken. After Setting the VF Enable bit, the VFs enabled by that VF Enable bit are permitted to return a CRS status to Configuration Requests up to the 1.0 s limit, if they are not ready to provide a Successful Completion status for a valid Configuration Request. After a PF transmits an FRS Message with a Reason Code of VF Enabled, no VF associated with that PF is permitted to return CRS without an intervening VF disable or other valid reset condition. After returning a Successful Completion to any Request, no ~~Additionally, a~~ VF is ~~not~~ permitted to return CRS ~~after having previously returned a Successful Completion~~ without an intervening VF disable or other valid reset condition.

Since VFs do not have an MSE bit (MSE in VFs is controlled by the VF MSE bit in the SR-IOV capability in the PF), it's possible for software to issue a Memory Request before the VF is ready to handle it. Therefore, Memory Requests must not be issued to a VF until at least one of the following conditions has been met:

- ~~At least 1.0 s has passed since either issuing an FLR to the VF or setting VF Enable.~~
- The VF has responded successfully (without returning CRS) to a Configuration Request.
- After issuing an FLR to the VF, at least one of the following is true:
  - At least 1.0 s has passed since the FLR was issued.

[3] For example, ACPI tables.

- The VF supports FRS and, after the FLR was issued, an FRS Message has been received from the VF with a Reason Code of FLR Completed.
- At least FLR Time has passed since the FLR was issued. FLR Time is either (1) the FLR Time value in the Readiness Time Reporting capability associated with the VF or (2) a value determined by system software / firmware[4].

❑ After Setting VF Enable in a PF, at least one of the following is true:

- At least 1.0 s has passed since VF Enable was Set.
- The PF supports FRS and, after VF Enable was Set, an FRS Message has been received from the PF with a Reason Code of VF Enabled.
- At least VF Enable Time has passed since VF Enable was Set. VF Enable Time is either (1) the Reset Time value in the Readiness Time Reporting capability associated with the VF or (2) a value determined by system software / firmware[5].

The VF is permitted to silently drop Memory Requests after an FLR has been issued to the VF or VF Enable has been Set in the associated PF's SR-IOV capability until the VF responds successfully (without returning CRS) to any Request ~~the first of the two conditions listed above have occurred~~.

Clearing VF Enable effectively destroys the VFs. Setting VF Enable effectively creates VFs. Setting VF Enable after it has previously been Cleared shall result in a new set of VFs. If the PF is in the D0 power state, the new VFs are in the $D0_{uninitialized}$ state. If the PF is in a lower power state behavior is undefined (see Sections 6.1 and 6.2).

When Clearing VF Enable, a PF that supports FRS shall send an FRS Message with FRS Reason VF Disabled to indicate when this operation is complete. The PF is not permitted to send this Message if there are outstanding Non-Posted Requests issued by the PF or any of the VFs associated with the PF. The FRS Message may only be sent after these Requests have completed (or timed out).

~~If software Clears VF Enable, software must allow 1.0 s a~~After VF Enable is Cleared, ~~before reading any~~ no field in the SR-IOV Extended Capability or the VF Migration State Array (see Section 3.3.15.1) may be accessed until either:

❑ At least 1.0 s has elapsed after VF Enable was Cleared.

❑ The PF supports FRS and after VF Enable was Cleared, an FRS Message has been received from the PF with a Reason Code of VF Disabled.

Section 3.3.7 NumVFs, Section 3.3.5 InitialVFs, Section 3.3.6 TotalVFs, Section 3.3.9 First VF Offset, Section 3.3.13 System Page Size, and Section 3.3.14 VF BARx describe additional semantics associated with this field.

[4] For example, ACPI tables.

[5] For example ACPI tables.

In the Single Root I/O Virtualization and Sharing Specification, Section 3.7, Table 3-22 add the following rows:

3.7. PCI Express Extended Capabilities

Extended Capability ID	Description	PF Attributes	VF Attributes
0021h	FRS Queuing	n/a	n/a
0022h	Readiness Time Reporting	Base	See Section 3.7.6

In the Single Root I/O Virtualization and Sharing Specification, add new Section 3.75:

3.7.6. Readiness Time Reporting Extended Capability

The **Reset Time** field contains the time required following setting of VF Enable (see Section 6.1).

The **DL_Up Time** field is RsvdP.

All VFs associated with a PF shall report the same time values.

In the Single Root I/O Virtualization and Sharing Specification, Section 6.1. VF Device Power Management States, change as shown:

6.1. VF Device Power Management States

If a VF does not implement the Power Management Capability, then the VF behaves as if it had been programmed into the equivalent power state of its associated PF.

If a VF implements the Power Management Capability, the functionality is defined in the *PCI Express Base Specification* except as noted in Section 6.4.

If a VF implements the Power Management Capability, the Device behavior is undefined if the PF is placed in a lower power state than the VF. Software should avoid this situation by placing all VFs in lower power state before lowering their associated PF's power state.

A VF in the D0 state is in the $D0_{active}$ state when the VF has completed its internal initialization and either the VF's Bus Master Enable bit is Set (see Section 3.4.1.3) or the VF MSE bit in the SR-IOV Control (see Section 3.3.3) Extended Capability is Set. The VF's internal initialization must have completed when ~~either~~ any of the following conditions have occurred:

- The VF has responded successfully (without returning CRS) to a Configuration Request.
- After issuing an FLR to the VF, one of the following is true:
  - At least 1.0 s has passed ~~since either issuing an FLR to the VF or setting VF Enable~~ since the FLR was issued.
  - The VF supports Function Readiness Status and, after the FLR was issued, an FRS Message from the VF with Reason Code FLR Completed has been received.

- At least FLR time has passed since the FLR was issued. FLR Time is either (1) the FLR Time value in the Readiness Time Reporting capability associated with the VF or (2) a value determined by system software / firmware[6].

- After Setting VF Enable in a PF, at least one of the following is true:
  - At least 1.0 s has passed since VF Enable was Set.
  - The PF supports Function Readiness Status and, after VF Enable was Set, an FRS Message from the PF with Reason Code VF Enabled has been received.

- After transitioning a VF from $D3_{hot}$ to D0, at least one of the following is true:
  - At least 10 ms has passed since the request to enter D0 was issued.
  - The VF supports Function Readiness Status and, after the request to enter D0 was issued, an FRS Message from the VF with Reason Code $D3_{hot}$ to D0 Transition Completed has been received.
  - At least $D3_{hot}$ to D0 Time has passed since the request to enter D0 was issued. $D3_{hot}$ to D0 Time is either (1) the $D3_{hot}$ to D0 Time in the Readiness Time Reporting capability associated with the VF or (2) a value determined by system software / firmware[7].

For PCI Code and ID Assignment Specification modify Section 3 as shown:

Table 3-1: Extended Capability IDs

0021h	FRS Queuing
0022h	Readiness Time Reporting

[6] For example, ACPI tables.

[7] For example ACPI tables.