PCI Firmware Specification

Revision 3.3

January 20, 2021

Revision	Revision History	Date
1.0	Original issue distributed by Intel Corporation.	9/28/1992
2.0	Updated to be in synch with PCI Bus Specification, Rev. 2.0.	7/20/1993
2.1	Added functions for PCI IRQ routing; clarifications.	8/26/1994
3.0	Updated revision evolving the specification to include all aspects of PCI and system firmware.	6/20/2005
3.1	Incorporated outstanding ECNs and made editorial changes.	12/13/2010
3.2	Incorporated the following ECNs: • ACPI additions for FW latency optimizations • UEFI related updates • PCIe hot plug	01/26/2015
3.3	Incorporated the following ECNs: • ECN_D3ColdTiming_Power_Firmware_Final • ECN_TPH-ST_Firmware_20191029 • Enabling Multiple Base Addresses per PCI Segment Group ECN • PCI FW 3.2 ECN - _DSM function 5 • PCI FW 3.2 ECN - HPX and Completion Timeout • PCI FW 3.2 ECN-DPC • PCI FW 3.2 ECN LED • PCIFW_3.2_ECN_DSM_Revisions_20200212 • _DSM additions for PCIe SSD Status LED Management ECN • _DSM additions for Runtime Device Power Management • System Firmware Intermediary (SFI) _OSC and DPC Updates	01/20/2021

Contact the PCI-SIG office to obtain the latest revision of the specification.

Questions regarding this specification or membership in PCI-SIG may be forwarded to:

Membership Services
www.pcisig.com
E-mail: administration@pcisig.com
Phone: 503-619-0569
Fax: 503-644-6708

Technical Support
techsupp@pcisig.com

Contents

Figures

Tables

1

1. Introduction

This document describes the hardware independent firmware interface for managing PCI, PCI-X, and PCI Express™ systems in a host computer.

1.1. Scope

This document is developed based on the *PCI BIOS Specification, Revision 2.1*. It continues to provide the PCI BIOS support on the PC-compatible systems.

In addition, this document also provides the descriptions or references of the following:

- Advanced Configuration and Power Interface (ACPI) services for supporting PCI, PCI-X, and PCI Express devices and systems.
- Extensible Firmware Interface (EFI) services for supporting PCI, PCI-X, and PCI Express devices and systems.
- Requirements and services for supporting PCI Expansion ROMs: The format, contents, and code entry points for PCI Expansion ROMs, along with system firmware services and execution environment, are described in this document. This information was documented in prior versions of the Conventional *PCI Local Bus Specification* (*Revision 2.3* and earlier), and is now maintained solely in this document.

1.2. Reference Documents

The following documents are a part of this specification to the extent specified herein:

Advanced Configuration and Power Interface, Revision 6.3 (ACPI 6.3), January 2019
https/uefi.org/specifications

Unified Extensible Firmware Interface Specification, Version 2.8, March 2019, https/uefi.org/specifications

Driver Writer's Guide for UEFI 2.3.1, Draft 1.10, April 2018, is https://edk2-docs.gitbooks.io/edk-ii-uefi-driver-writer-s-guide/

PCI Local Bus Specification, Revision 3.0 (PCI 3.0)

PCI Express Base Specification, Revision 5.0

PCI Express Card Electromechanical Specification, Revision 4.0

PCI Express to PCI/PCI-X Bridge Specification, Revision 1.0

PCI Express Mini Card Electromechanical Specification, Revision 4.0

PCI-X Addendum to the PCI Local Bus Specification, Revision 2.0

PCI Hot-Plug Specification, Revision 1.1

PCI Standard Hot-Plug Controller and Subsystem Specification, Revision 1.0 (SHPC 1.0)

DMTF Server Management Command Line Protocol (SM CLP) Specification, v1.0.2, Document Number DSP0214, March 7, 2007, http://www.dmtf.org/standards/published_documents/DSP0214.pdf

SMBIOS Reference Specification, v3.4.0a, Document Number DSP0134, December 7, 2019, https://www.dmtf.org/standards/smbios

1.3. Terms and Acronyms

Term	Definition
Active State Power Management (ASPM)	An autonomous hardware based active state mechanism, Active State Power Management defined in PCI Express.
Advanced Error Reporting (AER)	A capability for advanced error control and reporting defined in PCI Express.
Bus Number	A number in the range 0…255 that uniquely selects a PCI bus.
Configuration Space	A separate address space on PCI buses. Used for device identification and configuring devices into Memory and I/O spaces.
Device ID	A predefined field in configuration space that (along with Vendor ID) uniquely identifies the device.
Device Number	A number in the range 0…31 that uniquely selects a device on a PCI bus.
Function Number	A number in the range 0…7 that uniquely selects a function within a multi-function PCI device.
Message Signaled Interrupt (MSI)	An optional feature that enables a device to request service by writing a system-specified DW of data to a system-specified address using a Memory Write semantic Request.
Multi-function PCI device	A PCI device that contains multiple functions. For instance, a single device that provides both LAN and SCSI functions and has a separate configuration space for each function is a multi-function device.
Operating System Hot Plug (OSHP)	An ACPI control method to transfer control of Hot-Plug to the operating system. On some systems, this method is defined for Ports that are Hot-Plug capable and being controlled by ACPI firmware.
PCI	An acronym for Peripheral Component Interconnect.

Term	Definition
Segment Group Number	A number in the range 0…65535 that uniquely selects a PCI Segment Group. PCI Segment Group is purely a software concept managed by system firmware and used by the operating system. It is a logical collection of PCI buses (or bus segments). There is no tie to any physical entities. It is a way to logically group the PCI bus segments and PCI Express Hierarchies. PCI Segment Group concept enables support for more than 256 buses in a system by allowing the reuse of the PCI bus numbers. Within each PCI Segment Group, the bus numbers for the PCI buses must be unique. PCI buses in different PCI Segment Group are permitted to have the same bus number. A PCI Segment Group contains one or more PCI host bridges. There is at least one PCI Segment Group, PCI Segment Group 0, in a system.
Standard Hot-Plug Controller (SHPC)	A PCI Hot-Plug Controller compliant with SHPC 1.0.
Vendor ID	A predefined field in configuration space that (along with Device ID) uniquely identifies the device.

2

2.Traditional PCI BIOS

2.1. Functional Description

PCI BIOS functions provide a software interface to the hardware used to implement a PCI based system. Its primary usage is for generating operations in PCI specific address spaces (configuration space and Special Cycles).

PCI BIOS functions are specified to operate in the following modes of the X86 architecture. The modes are: real-mode, 16:16 protected mode (also known as 286 protected mode), 16:32 protected mode (introduced with the 386), and 0:32 protected mode (also known as "flat" mode, wherein all segments start at linear address 0 and span the entire 4 GiB address space).

Access to the PCI BIOS functions for 16-bit callers is provided through Interrupt 1Ah. 32-bit (i.e., protected mode) access is provided by calling through a 32-bit protected mode entry point. The PCI BIOS function code is B1h. Specific BIOS functions are invoked using a sub-function code. A user simply sets the host processors registers for the function and sub-function desired and calls the PCI BIOS software. Status is returned using the CARRY FLAG ([CF]) and registers specific to the function invoked.

2.2. Assumptions and Constraints

2.2.1. ROM BIOS Location

The PCI BIOS functions are intended to be located within an IBM-PC compatible ROM BIOS.

2.2.2. Calling Conventions

The PCI BIOS functions use the X86 CPU's registers to pass arguments and return status. The caller must use the appropriate sub-function code.

These routines preserve all registers and flags except those used for return parameters. The CARRY FLAG [CF] will be altered as shown to indicate completion status. The calling routine will be returned with the interrupt flag unmodified and interrupts will not be enabled during function execution. These routines are re-entrant. These routines require 1024 bytes of stack space and the stack segment must have the same size (i.e., 16-bit or 32-bit) as the code segment.

The PCI BIOS provides a 16-bit real and protected mode interface and a 32-bit protected mode interface. The 16-bit interface is provided through PC/AT Int 1Ah software interrupt. The PCI BIOS Int 1Ah interface operates in either real mode, virtual-86 mode, or 16:16 protected mode. The BIOS functions may also be accessed through the industry standard entry point for INT 1Ah (physical address 000F FE6Eh) by simulating an INT instruction.[3] The INT 1Ah entry point supports 16-bit code only. Protected mode callers of this interface must set the CS selector base to 0 F000h.

The protected mode interface supports 32-bit protected mode callers. The protected mode PCI BIOS interface is accessed by calling (not a simulated INT) through a protected mode entry point in the PCI BIOS. The entry point and information needed for building the segment descriptors are provided by the BIOS32 Service Directory (refer to Section 2.3). 32-bit callers invoke the PCI BIOS routines using CALL FAR.

The PCI BIOS routines (for both 16-bit and 32-bit callers) must be invoked with appropriate privilege so that interrupts can be enabled/disabled and the routines can access I/O space. Implementers of the PCI BIOS must assume that CS is execute-only and DS is read-only.

2.2.3. Interrupt Support

To support PC-compatible BIOS, the system must support the INTx routing. MSI or MSI-X may be supported for the operating system use after booted from BIOS.

2.3. BIOS32 Service Directory

Detecting the absence or presence of 32-bit BIOS services with 32-bit code can be problematic. Standard BIOS entry points cannot be called in 32-bit mode on all machines because the platform BIOS may not support 32-bit callers. This section describes a mechanism for detecting the presence of 32-bit BIOS services. While the mechanism supports the detection of the PCI BIOS, it is intended to be broader in scope to allow detection of any/all 32-bit BIOS services. The description of this mechanism, known as BIOS32 Service Directory, is provided in three parts; the first part specifies an algorithm for determining if the BIOS32 Service Directory exists on a platform, the second part specifies the calling interface to the BIOS32 Service Directory, and the third part describes how the BIOS32 Service Directory supports PCI BIOS detection.

2.3.1. Determining the Existence of BIOS32 Service Directory

A BIOS which implements the BIOS32 Service Directory must embed a specific, contiguous 16-byte data structure, beginning on a 16-byte boundary somewhere in the physical address range 0E 0000h - 0F FFFFh. A description of the fields in the data structure are given in Table 2-1.

[3] Note that accessing the BIOS functions through the industry standard entry point will bypass any code that may have “hooked” the INT 1Ah interrupt vector.

Table 2-1. Data Structure Fields for the BIOS32 Service Directory

Offset	Size	Description
0	4 bytes	Signature string in ASCII. The string is "_32_". This puts an “underscore” at offset 0, a “3” at offset 1, a “2” at offset 2, and another “underscore” at offset 3.
4	4 bytes	Entry point for the BIOS32 Service Directory. This is a 32-bit physical address.
8	1 byte	Revision level. This version has revision level 00h.
9	1 byte	Length. This field provides the length of this data structure in paragraph (i.e., 16-byte) units. This data structure is 16 bytes long so this field contains 01h.
0Ah	1 byte	Checksum. This field is a checksum of the complete data structure. The sum of all bytes must add up to 0.
0Bh	5 bytes	Reserved. Must be zero.

Clients of the BIOS32 Service Directory should determine its existence by scanning 0E 0000h to 0F FFF0h looking for the ASCII signature and a valid, checksummed data structure. If the data structure is found, the BIOS32 Service Directory can be accessed through the entry point provided in the data structure. If the data structure is not found, then the BIOS32 Service Directory (and also the PCI BIOS) is not supported by the platform.

2.3.2. Calling Interface for BIOS32 Service Directory

The BIOS32 Service Directory is accessed by doing a CALL FAR to the entry point provided in the Service data structure (see previous section). There are several requirements about the calling environment that must be met. The CS code segment selector and the DS data segment selector must be set up to encompass the physical page holding the entry point as well as the immediately following physical page. They must also have the same base. Platform BIOS writers must assume that CS is execute-only and DS is read-only. The SS stack segment selector must provide at least 1 KiB of stack space. The calling environment must also allow access to I/O space.

The BIOS32 Service Directory provides a single function to determine whether a particular 32-bit BIOS service is supported by the platform. All parameters to the function are passed in registers. Parameter descriptions are provided below. If a particular service is implemented in the platform BIOS, three values are returned. The first value is the base physical address of the BIOS service. The second value is the length of the BIOS service. These two values can be used to build the code segment selector and data segment selector for accessing the service. The third value provides the entry point to the BIOS service encoded as an offset from the base.

ENTRY:

[EAX] Service Identifier. This is a four character string used to specifically identify which 32-bit BIOS Service is being sought.

[EBX] The low order byte ([BL]) is the BIOS32 Service Directory function selector. Currently only one function is defined (with the encoding of zero) which returns the values provided below.

The upper three bytes of [EBX] are reserved and must be zero on entry.

EXIT:

[AL] Return Code:

00h = Service corresponding to Service Identifier is present.

80h = Service corresponding to Service Identifier is not present.

81h = Unimplemented function for BIOS Service Directory (i.e., BL has an unrecognized value).

[EBX] Physical address of the base of the BIOS service.

[ECX] Length of the BIOS service.

[EDX] Entry point into BIOS service. This is an offset from the base provided in EBX.

2.4. PCI BIOS 32-bit Service

The BIOS32 Service Directory may be used to detect the presence of the PCI BIOS. The Service Identifier for the PCI BIOS is "$PCI" (4943 5024h).

The 32-bit PCI BIOS functions must be accessed using CALL FAR. The CS and DS descriptors must be setup to encompass the physical addresses specified by the Base and Length parameters returned by the BIOS32 Service Directory. The CS and DS descriptors must have the same base. The calling environment must allow access to I/O space and provide at least 1 KiB of stack space. Platform BIOS writers must assume that CS is execute-only and DS is read-only.

2.5. Host Interface

2.5.1. Identifying PCI Resources

The following group of functions allow the caller to determine first, if the PCI BIOS support is installed, and second, if specific PCI devices are present in the system.

2.5.2. PCI BIOS Present

This function allows the caller to determine whether the PCI BIOS interface function set is present, and what the current interface version level is. It also provides information about what hardware mechanism for accessing configuration space is supported, and whether or not the hardware supports generation of PCI Special Cycles.

ENTRY:

```
[AH]          PCI_FUNCTION_ID

[AL]          PCI_BIOS_PRESENT
```

EXIT:

```
[EDX]         "PCI", "P" in [DL], "C" in [DH], etc.  There is a
              "space" character in the upper byte.

[AH]          Present Status, 00h = BIOS Present if and only if
              EDX set properly.

[AL]          Hardware mechanism.

[BH]          Interface Level Major Version.

[BL]          Interface Level Minor Version.

[CL]          Number of last PCI bus in the system.

[CF]          Present Status, set = No BIOS Present,
              reset = BIOS Present if and only if EDX set
              properly.
```

If the CARRY FLAG [CF] is cleared and AH is set to 00h, it is still necessary to examine the contents of [EDX] for the presence of the string "PCI" + (trailing space) to fully validate the presence of the PCI function set. [BX] will further indicate the version level, with enough granularity to allow for incremental changes in the code that do not affect the function interface. Version numbers are stored as Binary Coded Decimal (BCD) values. For example, Version 2.10 would be returned as a 02h in the [BH] registers and 10h in the [BL] registers.

A BIOS that complies with Version 3.1 of the specification will return with 0310h in the [BX] register.

The value returned in [AL] identifies what specific hardware characteristics the platform supports in relation to accessing configuration space and generating PCI Special Cycles (see Figure 2-1). The PCI Specification defines two hardware mechanisms for accessing configuration space. Bits 0 and 1 of the value returned in [AL] specify which mechanism is supported by this platform. Bit 0 will be Set (1) if Mechanism #1 is supported, and Clear (0) otherwise. Bit 1 will be Set (1) if Mechanism #2 is supported, and Clear (0) otherwise. Bits 2, 3, 6, and 7 are reserved and returned as zeros.

The PCI Specification also defines hardware mechanisms for generating Special Cycles. Bits 4 and 5 of the value return in [AL] specify which mechanism is supported (if any). Bit 4 will be Set (1) if the platform supports Special Cycle generation based on Config Mechanism #1, and Clear (0) otherwise. Bit 5 will be Set (1) if the platform supports Special Cycle generation based on Config Mechanism #2, and Clear (0) otherwise.

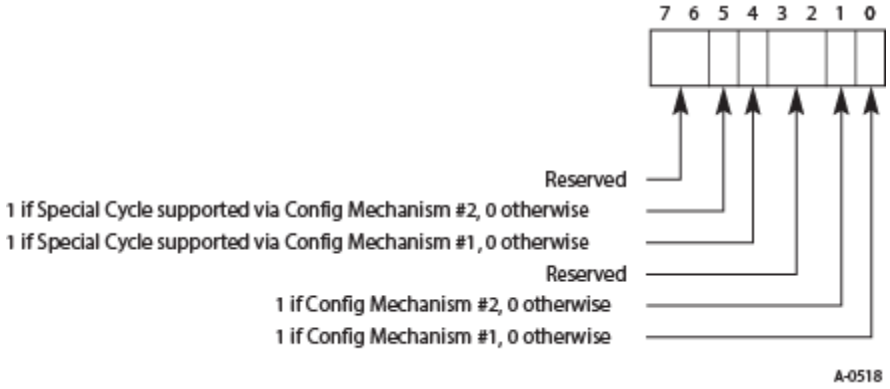


Figure 2-1: Layout of Value Returned in [AL]

The value returned in [CL] specifies the number of the last PCI bus in the system. PCI buses are numbered starting at zero and running up to the value specified in CL.

If the [BX] register indicates that the BIOS is compliant with version 3.0 (or later), then the [CH] register will have the following meaning:

CH = Level of BIOS support:
- bit 0 set = Functions 06h through 0Dh inclusively are implemented for register access below 256.
- bit 1 set = Functions 06h through 0Dh inclusively are implemented for register access in the range 256 - 4096 for POST only.
- bit 2 set = Function 0Eh has been implemented.
- bit 3 set = Function 0Fh has been implemented.
- bit 4 set = Function 02h has been implemented.
- bit 5 set = Function 03h has been implemented.
- bit 6 set = BIOS supports Option ROM Configuration Code execution.
- bit 7 set = BIOS supports calling the Option ROM with DMTF CLP style configuration data.

Note that version 3.0 and 3.1 of this specification requires that bit 0 and bit 1 be Set. This may not be the case in future versions of the specification.

2.5.3. Find PCI Device

This function returns the location of PCI devices that have a specific Device ID and Vendor ID. Given a Vendor ID, Device ID, and an Index (N), the function returns the Bus Number, Device Number, and Function Number of the Nth Device/Function whose Vendor ID and Device ID match the input parameters.

ENTRY:

[AH] PCI_FUNCTION_ID

[AL] FIND_PCI_DEVICE

[CX] Device ID (0...65535)

[DX] Vendor ID (0...65534)

[SI] Index (0...N)

EXIT:

[BH] Bus Number (0...255)

[BL] Device Number in upper 5 bits,
Function Number in bottom 3 bits.

[AH] Return Code:

SUCCESSFUL

DEVICE_NOT_FOUND

BAD_VENDOR_ID

[CF] Completion Status, set = error, cleared = success.

Calling software can find all devices having the same Vendor ID and Device ID by making successive calls to this function starting with Index set to zero, and incrementing it until the return code is "DEVICE_NOT_FOUND". A return code of BAD_VENDOR_ID indicates that the passed in Vendor ID value (in [DX]) had an illegal value of all 1's.

Values returned by this function upon successful completion must be the actual values used to access the PCI device if the INT 1Ah routines are bypassed in favor of the direct I/O mechanisms described in the PCI Specification.

2.5.4. Find PCI Class Code

This function returns the location of PCI devices that have a specific Class Code. Given a Class Code and an Index (N), the function returns the Bus Number, Device Number, and Function Number of the Nth Device/Function whose Class Code matches the input parameters.

ENTRY:

[AH]	PCI_FUNCTION_ID
[AL]	FIND_PCI_CLASS_CODE
[ECX]	Class Code (in lower 3 bytes)
[SI]	Index (0...N)

EXIT:

[BH]	Bus Number (0...255)
[BL]	Device Number in upper 5 bits, Function Number in bottom 3 bits.
[AH]	Return Code: SUCCESSFUL DEVICE_NOT_FOUND
[CF]	Completion Status, set = error, cleared = success.

Calling software can find all devices having the same Class Code by making successive calls to this function starting with Index set to zero, and incrementing it until the return code is "DEVICE_NOT_FOUND".

2.6. PCI Support Functions

The following functions provide support for several PCI specific operations.

2.6.1. Generate Special Cycle

This function allows for generation of PCI special cycles. The generated special cycle will be broadcast on a specific PCI bus in the system.

ENTRY:

[AH] PCI_FUNCTION_ID

[AL] GENERATE_SPECIAL_CYCLE

[BH] Bus Number (0...255)

[EDX] Special Cycle Data

EXIT:

[AH] Return Code:

SUCCESSFUL

FUNC_NOT_SUPPORTED

[CF] Completion Status, set = error, reset = success.

2.6.2. Get PCI Interrupt Routing Expansions

Description:

Logical input parameters:

RouteBuffer Pointer to buffer data structure

This routine returns the PCI interrupt routing Expansions available on the system motherboard and also the current state of what interrupts are currently exclusively assigned to PCI. Routing information is returned in a data buffer that contains an IRQ Routing for each PCI device or slot. The format of an entry in the IRQ routing table is shown in Table 2-2.

Table 2-2: Layout of IRQ Routing Table Entry

Offset	Size	Description
0	byte	PCI Bus number
1	byte	PCI Device number (in upper 5 bits)
2	byte	Link value for INTA#
3	word	IRQ bit-map for INTA#
5	byte	Link value for INTB#
6	word	IRQ bit-map for INTB#
8	byte	Link value for INTC#
9	word	IRQ bit-map for INTC#
11	byte	Link value for INTD#
12	word	IRQ bit-map for INTD#
14	byte	Slot Number
15	byte	Reserved (for OEM use)

Two values are provided for each PCI interrupt pin in every slot. One of these values is a bit-map that shows which of the standard AT IRQs this PCI interrupt can be routed to. This provides the routing Expansions for one particular PCI interrupt pin. In this bit-map, bit 0 corresponds to IRQ0, bit 1 to IRQ1, etc. A "1" bit in this bit-map indicates a routing is possible: a "0" bit indicates no routing is possible. The second value is a "link" value that provides a way of specifying which PCI interrupt pins are wire-OR'ed together on the motherboard. Interrupt pins that are wired together must have the same "link" value in their table entries. Values for the "link" field are arbitrary except that the value zero indicates that the PCI interrupt pin has no connection to the interrupt controller.[5]

The Slot Number value at the end of the structure is used to communicate whether the table entry is for a motherboard device or an add-in slot. For motherboard devices, Slot Number should be set to zero. For add-in slots, Slot Number should be set to a value that corresponds with the physical placement of the slot on the motherboard. This provides a way to correlate physical slots with PCI Device numbers. Values (with the exception of 00h) are OEM specific.[6] For end user ease-of-use, slots in the system should be clearly labeled (e.g., solder mask, back panel, etc.).

[5]This is typically used for motherboard devices that have only an IRQA# line and not IRQB#, IRQC#, or IRQD#.

[6]For example, a system with 4 ISA slots and 3 PCI slots arranged as 3-ISA, 3-PCI, and 1-ISA may choose to start numbering the slots at the 3-ISA end in which case the PCI slot numbers would be 4, 5, and 6. If slot numbering started at the 1-ISA end, PCI slot numbers would be 2, 3, and 4.

This routine requires one parameter, *RouteBuffer*, which is a far pointer to the data structure shown below.

```
typedef struct
    {
        WORD        BufferSize;
        BYTE  FAR * DataBuffer;
    } IRQRoutingExpansionsBuffer;
```

where

BufferSize: A word size value providing the size of the data buffer. If the buffer is too small, the routine will return with status of BUFFER_TOO_SMALL, and this field will be updated with the required size. To indicate that the running PCI system does not have any PCI devices, this function will update the *BufferSize* field to zero. On successful completion, this field is updated with the size (in bytes) of the data returned.

DataBuffer: Far pointer to the buffer containing PCI interrupt routing information for all motherboard devices and slots.

IMPLEMENTATION NOTE

Defining the DataBuffer in C

The code example above is syntactically correct for C language in all processor modes. But the storage size of the BYTE FAR * offset varies for each mode as follows:

- Real Mode: 2 WORD (Segment:Offset)
- PM16: 2 WORD (Selector:Offset)
- PM32: 3 WORD (Selector:32bitOffset)

This routine also returns information about which IRQs are currently dedicated for PCI usage. This information is returned as a bit map where a set bit indicates that the IRQ is dedicated to PCI and not available for use by devices on other buses. Note that if an IRQ is routed such that it can be used by PCI devices and other devices the corresponding bit in the bit map should not be set. The function returns this information in the [BX] register where bit 0 corresponds to IRQ0, bit 1 - IRQ1, etc. The caller must initialize [BX] to zero before calling this routine.

ENTRY:

[AH]	PCI_FUNCTION_ID
[AL]	GET_IRQ_ROUTING_EXPANSIONS
[BX]	Must be initialized to 0000h.
[DS]	Segment or Selector for BIOS data. For 16-bit code, the real-mode segment or PM selector must resolve to physical address 0F 0000h and have a limit of 64 KiB. For information on 32-bit code, refer to Section 2.4.
[ES]	Segment or Selector for *RouteBuffer* parameter.
[DI] for 16-bit code	Offset for *RouteBuffer* parameter.
[EDI] for 32-bit code	

EXIT:

[AH]	Return Code: SUCCESSFUL BUFFER_TOO_SMALL FUNC_NOT_SUPPORTED
[BX]	IRQ bitmap indicating which IRQs are exclusively dedicated to PCI devices.
[CF]	Completion Status, set = error, cleared = success.

2.6.3. Set PCI Hardware Interrupt

Description:

This function is intended to be used by a system-wide configuration utility or a PNP operating system. This function should never be called by device drivers or Expansion ROM code.

This function is optional in this version of the specification. The caller is responsible for checking the return code to determine if the call is supported.

Logical input parameters:

BusDev	Bus number and device number
IntPin	PCI Interrupt Pin (INTA .. INTD)
IRQNum	IRQ Number (0-15)

This routine causes the specified hardware interrupt (IRQ) to be connected to the specified interrupt pin of a PCI device. It makes the following assumptions:

1. The caller is responsible for all error checking to ensure no resource conflict exits between the specific hardware interrupt assigned to the PCI device and any other hardware interrupt resource in the system.

2. The caller is responsible for ensuring that the specified interrupt is configured properly (level triggered) in the interrupt controller. If the system contains hardware outside of the interrupt controller that controls interrupt triggering (edge/level), then the callee (i.e., the BIOS) is responsible for setting that hardware to level triggered for the specified interrupt.

3. The caller is responsible for updating PCI configuration space (i.e., Interrupt Line registers) for all effected devices.

4. The caller must be aware that changing IRQ routing for one device will also change the IRQ routing for other devices whose INTx# pins are WIRE-ORed together (i.e., they have the same link field in the Get Routing Expansions call).

If the requested interrupt cannot be assigned to the specified PCI device, then SET_FAILED status is returned. This routine immediately effects the interrupt routing and does nothing to remember the routing for the next system boot.

The *BusDev* parameter specifies the PCI bus and device numbers for the PCI device/slot being modified. The high-order byte of *BusDev* contains the PCI bus number. The device number is provided in the top five bits of the low-order byte of *BusDev*. For example, to specify device 6 on PCI bus 2 the *BusDev* parameter would be 0230h.

The *IntPin* parameter specifies which interrupt pin (INTA#,..,INTD#) of the specified PCI device/slot is effected by this call. A value of 0Ah corresponds to INTA#, 0Bh to INTB#, etc.

The *IRQNum* parameter specifies which IRQ input is to be connected to the PCI interrupt pin. This parameter can have values of 0..15 specifying IRQ0 thru IRQ15 respectively.

ENTRY:

```
[AH]          PCI_FUNCTION_ID

[AL]          SET_PCI_HW_INT

[CL]          IntPin parameter.  Valid values 0Ah..0Dh.

[CH]          IRQNum parameter.  Valid values 0..0Fh.

[BX]          BusDev parameter.  [BH] holds bus number,
              [BL] holds Device (upper 5 bits), and
              Function (lower 3 bits) numbers.

[DS]          Segment or Selector for BIOS data.

              For 16-bit code, the real-mode segment or
              the Protected Mode selector must resolve to
              physical address 0F 0000h and have a limit
              of 64 KiB.

              For 32-bit code, see Section 2.4.
```

EXIT:

```
[AH]            Return Code:
                SUCCESSFUL
                SET_FAILED
                FUNC_NOT_SUPPORTED
[CF]            Completion Status, set = error, cleared
                = success.
```

2.7. Accessing Configuration Space

The 32-bit PCI BIOS functions must be accessed using CALL FAR. The CS and DS descriptors must be setup to encompass the physical addresses specified by the Base and Length parameters returned by the BIOS32 Service Directory. The CS and DS descriptors must have the same base. The ES descriptors must be setup to encompass the physical address reported by the PCI Express Base Address Register (BAR) region. The ES descriptor must have a base of zero with the limit of 4 GiB to encompass the 256-MiB PCI Express BAR region. The calling environment must allow access to I/O space and provide at least 1 KiB of stack space. Platform BIOS writers must assume that CS is execute-only, DS is read-only, and ES is read/write.

The underlying assumptions being made are the following:

- If PCI BIOS functions support access to extended configuration registers, the PCI Express memory mapped configuration base address must be programmed with an address region that exists below 4 GiB of memory.
- The caller has the responsibility to appropriately parse the data returned from accessing non-existing registers of a PCI Device. The caller has to ensure that Register accesses beyond 256 bytes be invoked only on devices that have the support for extended configuration space. For example: if the caller accesses DWORD Register 0FFCh (4092) of a regular PCI device, the data returned cannot be predicted.
- PCI BIOS functions do not comprehend the concept of PCI Segment Groups (for definition, refer to Section 4.3.2.3) and, hence, can only support access to the devices in the default PCI Segment Group, namely, PCI Segment Group 0.

2.7.1. Access Rules for PCI Express I/O and Memory Mapped Accesses

Firmware should use the PCI I/O Index/Data mechanism to access configuration space registers 0-255. Only accesses to configuration space registers 256 and beyond should use the PCI Express memory-mapped access mechanism.

2.7.2. INT1Ah Access Calls in Real Mode

The real mode invocations of INT 1Ah for reading the PCI Express Extended Configuration space are intended to be used for handling the internal BIOS/system firmware code calls during POST (if needed). After POST, it is recommended that the PCI Express Extended Configuration space be accessed directly without the use of INT 1Ah. The ACPI MCFG table describes the location of the PCI Express configuration space, and this table will be present in a firmware implementation compliant to this specification version 3.0 (or later). If INT 1Ah is invoked after POST for reading the PCI Express extended configuration space, the firmware may return FUNC_NOT_SUPPORTED.

2.7.3. Read Configuration Byte

This function allows reading individual bytes from the configuration space of a specific device.

ENTRY:

```
[AH]    PCI_FUNCTION_ID

[AL]    READ_CONFIG_BYTE

[BH]    Bus Number (0...255)

[BL]    Device Number in upper 5 bits,
        Function Number in lower 3 bits.

[DI]    Register Number (0...4095) [bits 11:0]
          To Read Register number greater than 255 [bit15=1]
          To Read Register number less than or equal to 255 [bit15=0]
```

EXIT:

```
[CL]    Byte Read

[AH]    Return Code:
          SUCCESSFUL
          BAD_REGISTER_NUMBER
          FUNC_NOT_SUPPORTED

[CF]    Completion Status, set = error, reset = success.
```

Input Register Requirements:

- If register DI has a register number less than or equal to 255 and does not have bit15=1, the BIOS will read the configuration space and return the value read with return code "SUCCESS".
- If register DI has a register number greater than 255 and does not have bit15=1, the BIOS will not try to read the configuration space and returns "BAD_REGISTER_NUMBER".
- If register DI has a register number greater than 255 with bit15=1 and the BIOS implements the extensions for accessing PCI Express Extended Configuration space, then the BIOS will read the configuration space and return the value read with return code "SUCCESS".
- If register DI has a register number greater than 255 with bit15=1 and the BIOS does not implement the extensions for accessing PCI Express Extended Configuration space, then the BIOS will return error code of "FUNC_NOT_SUPPORTED".

2.7.4. Read Configuration Word

This function allows reading individual words from the configuration space of a specific device. The Register Number parameter must be a multiple of two (i.e., bit 0 must be set to 0).

ENTRY:

```
[AH]    PCI_FUNCTION_ID

[AL]    READ_CONFIG_WORD

[BH]    Bus Number (0...255)

[BL]    Device Number in upper 5 bits,
        Function Number in lower 3 bits.

[DI]    Register Number (0, 2, 4,...4094) [bits 11:0]
          To Read Register number greater than 255 [bit15=1]
          To Read Register number less than or equal to 254 [bit15=0]
```

EXIT:

```
[CX]    Word Read

[AH]    Return Code:
          SUCCESSFUL
          BAD_REGISTER_NUMBER
          FUNC_NOT_SUPPORTED

[CF]    Completion Status, set = error, reset = success.
```

Input Register Requirements:

- If register DI has a register number less than or equal to 254 and does not have bit15=1, the BIOS will read the configuration space and return the value read with return code "SUCCESS".
- If register DI has a register number greater than 255 and does not have bit15=1, the BIOS will not try to read the configuration space and returns "BAD_REGISTER_NUMBER".
- If register DI has a register number greater than 255 with bit15=1 and the BIOS implements the extensions for accessing PCI Express Extended Configuration space, then the BIOS will read the configuration space and return the value read with return code "SUCCESS".
- If register DI has a register number greater than 255 with bit15=1 and the BIOS does not implement the extensions for accessing PCI Express Extended Configuration space, then the BIOS will return error code of "FUNC_NOT_SUPPORTED".

2.7.5. Read Configuration DWORD

This function allows reading individual DWORDs from the configuration space of a specific device. The Register Number parameter must be a multiple of four (i.e., bits 0 and 1 must be set to 0).

ENTRY:

```
[AH]       PCI_FUNCTION_ID

[AL]       READ_CONFIG_DWORD

[BH]       Bus Number (0...255)

[BL]       Device Number in upper 5 bits,
           Function Number in lower 3 bits.

[DI]       Register Number (0, 4, 8,...4092) [bits 11:0]

           To Read Register number greater than 255 [bit15=1]

           To Read Register number less than or equal to 252 [bit15=0]
```

EXIT:

```
[ECX]      DWORD Read

[AH]       Return Code:

               SUCCESSFUL

               BAD_REGISTER_NUMBER

               FUNC_NOT_SUPPORTED

[CF]       Completion Status, set = error, reset = success.
```

Input Register Requirements:

- If register DI has a register number less than or equal to 252 and does not have bit15=1, the BIOS will read the configuration space and return the value read with return code "SUCCESS".
- If register DI has a register number greater than 255 and does not have bit15=1, the BIOS will not try to read the configuration space and returns "BAD_REGISTER_NUMBER".
- If register DI has a register number greater than 255 with bit15=1 and BIOS implements the extensions for accessing PCI Express Extended Configuration space, then the BIOS will read the configuration space and return the value read with return code "SUCCESS".
- If register DI has a register number greater than 255 with bit15=1 and the BIOS does not implement the extensions for accessing PCI Express Extended Configuration space, then the BIOS will return error code of "FUNC_NOT_SUPPORTED".

2.7.6. Write Configuration Byte

This function allows writing individual bytes to the configuration space of a specific device.

ENTRY:

```
[AH]       PCI_FUNCTION_ID

[AL]       WRITE_CONFIG_BYTE

[BH]       Bus Number (0...255)

[BL]       Device Number in upper 5 bits,
           Function Number in lower 3 bits.

[DI]       Register Number (0...4095) [bits 11:0]

           To Write Register number greater than 255 [bit15=1]

           To Write Register number less than or equal to 255 [bit15=0]

[CL]       Byte Value to Write
```

EXIT:

```
[AH]       Return Code:

               SUCCESSFUL

               BAD_REGISTER_NUMBER

               FUNC_NOT_SUPPORTED

[CF]       Completion Status, set = error, reset = success.
```

Input Register Requirements:

- If register DI has a register number less than or equal to 255 and does not have bit15=1, the BIOS will read the configuration space and return the value read with return code "SUCCESS".
- If register DI has a register number greater than 255 and does not have bit15=1, the BIOS will not try to read the configuration space and returns "BAD_REGISTER_NUMBER".
- If register DI has a register number greater than 255 with bit15=1 and the BIOS implements the extensions for accessing PCI Express Extended Configuration space, then the BIOS will read the configuration space and return the value read with return code "SUCCESS".
- If register DI has a register number greater than 255 with bit15=1 and the BIOS does not implement the extensions for accessing PCI Express Extended Configuration space, then the BIOS will return error code of "FUNC_NOT_SUPPORTED".

2.7.7. Write Configuration Word

This function allows writing individual words from the configuration space of a specific device. The Register Number parameter must be a multiple of two (i.e., bit 0 must be set to 0).

ENTRY:

```
[AH]       PCI_FUNCTION_ID

[AL]       WRITE_CONFIG_WORD

[BH]       Bus Number (0...255)

[BL]       Device Number in upper 5 bits,
           Function Number in lower 3 bits.

[DI]       Register Number (0, 2, 4,...4094) [bits 11:0]

           To Write Register number greater than 255 [bit15=1]

           To Write Register number less than or equal to 254 [bit15=0]

[CX]       Word Value to Write
```

EXIT:

```
[AH]       Return Code:

               SUCCESSFUL

               BAD_REGISTER_NUMBER

               FUNC_NOT_SUPPORTED

[CF]       Completion Status, set = error, reset = success.
```

Input Register Requirements:

- If register DI has a register number less than or equal to 254 and does not have bit15=1, the BIOS will read the configuration space and return the value read with return code "SUCCESS".
- If register DI has a register number greater than 255 and does not have bit15=1, the BIOS will not try to read the configuration space and returns "BAD_REGISTER_NUMBER".
- If register DI has a register number greater than 255 with bit15=1 and the BIOS implements the extensions for accessing PCI Express Extended Configuration space, then the BIOS will read the configuration space and return the value read with return code "SUCCESS".
- If register DI has a register number greater than 255 with bit15=1 and the BIOS does not implement the extensions for accessing PCI Express Extended Configuration space, then the BIOS will return error code of "FUNC_NOT_SUPPORTED".

2.7.8. Write Configuration DWORD

This function allows writing individual DWORDs from the configuration space of a specific device. The Register Number parameter must be a multiple of four (i.e., bits 0 and 1 must be set to 0).

ENTRY:

```
[AH]    PCI_FUNCTION_ID

[AL]    WRITE_CONFIG_DWORD

[BH]    Bus Number (0...255)

[BL]    Device Number in upper 5 bits,
        Function Number in lower 3 bits.

[DI]    Register Number (0, 4, 8,...4092) [bits 11:0]
          To Write Register number greater than 255 [bit15=1]
          To Write Register number less than or equal to 252 [bit15=0]

[ECX]   DWORD Value to Write
```

EXIT:

```
[AH]    Return Code:
          SUCCESSFUL
          BAD_REGISTER_NUMBER
          FUNC_NOT_SUPPORTED

[CF]    Completion Status, set = error, reset = success.
```

Input Register Requirements:

- If register DI has a register number less than or equal to 252 and does not have bit15=1, the BIOS will read the configuration space and return the value read with return code "SUCCESS".
- If register DI has a register number greater than 255 and does not have bit15=1, the BIOS will not try to read the configuration space and returns "BAD_REGISTER_NUMBER".
- If register DI has a register number greater than 255 with bit15=1 and the BIOS implements the extensions for accessing PCI Express Extended Configuration space, then the BIOS will read the configuration space and return the value read with return code "SUCCESS".
- If register DI has a register number greater than 255 with bit15=1 and the BIOS does not implement the extensions for accessing PCI Express Extended Configuration space, then the BIOS will return error code of "FUNC_NOT_SUPPORTED".

2.8. Function List

Table 2-3: Function List

Function	AH	AL	Implementation	Notes
PCI_FUNCTION_ID	B1h			
PCI_BIOS_PRESENT		01h		
FIND_PCI_DEVICE		02h	Optional	1
FIND_PCI_CLASS_CODE		03h	Optional	1
GENERATE_SPECIAL_CYCLE		06h	Optional	1
READ_CONFIG_BYTE		08h	Optional	1
READ_CONFIG_WORD		09h	Optional	1
READ_CONFIG_DWORD		0Ah	Optional	1
WRITE_CONFIG_BYTE		0Bh	Optional	1
WRITE_CONFIG_WORD		0Ch	Optional	1
WRITE_CONFIG_DWORD		0Dh	Optional	1
GET_IRQ_ROUTING_EXPANSIONS		0Eh	Optional	2
SET_PCI_HW_INT		0Fh	Optional	2

Notes:

1. If the "PCI BIOS PRESENT (B101h)" function indicates "Presence", then all functions (06h-to-0Dh inclusively) must be implemented for register accesses below 256 and must be present during POST and at run-time (after POST completes). For register accesses above 255, these functions must only be present during POST. There is not a requirement to implement the above 255 register access functions after POST completes; however, if implemented, they must not change the processor mode when invoked after POST (option ROMs might be executing in v86 mode).
2. Implementation of these INT1Ah sub-functions is optional for compliance with this version of the specification.

2.9. Return Code List

Table 2-4: Return Code List

Return Codes	AH
SUCCESSFUL	00h
FUNC_NOT_SUPPORTED	81h
BAD_VENDOR_ID	83h
DEVICE_NOT_FOUND	86h
BAD_REGISTER_NUMBER	87h
SET_FAILED	88h
BUFFER_TOO_SMALL	89h

3

3.UEFI PCI Services

UEFI stands for Unified Extensible Firmware Interface. The *UEFI Specification, Version 2.4* or later (http://www.uefi.org) describes an interface between the operating system and the platform firmware. The interface is in the form of data tables that contain platform-related information and boot and run-time services calls that are available to the operating system and its loader. Together, these provide a standard environment for booting an operating system.

The following sections provide an overview of the UEFI Services relevant to PCI (including Conventional PCI, PCI-X, and PCI Express). For details, refer to the UEFI Specification. UEFI is processor-agnostic.

3.1. UEFI Driver Model

The UEFI Driver Model is designed to support the execution of drivers that run in the pre-boot environment present on systems that implement the UEFI firmware. These drivers may manage and control hardware buses and devices on the platform, or they may provide some software derived platform specific services.

The UEFI Driver Model is designed to extend the UEFI Specification in a way that supports device drivers and bus drivers. It contains information required by UEFI driver writers to design and implement any combination of bus drivers and device drivers that a platform may need to boot an UEFI-compliant operating system.

Applying the UEFI Driver Model to PCI, the UEFI Specification defines the PCI Root Bridge Protocol and the PCI Driver Model and describes how to write PCI bus drivers and PCI devices drivers in the UEFI environment. For details, refer to the UEFI Specification.

3.1.1. PCI Root Bridge Protocol

A PCI Root Bridge is represented in UEFI as a device handle that contains a Device Path Protocol instance and a PCI Root Bridge Protocol instance.

PCI Root Bridge Protocol provides an I/O abstraction for a PCI Root Bridge that the host bus can perform. This protocol is used by a PCI Bus Driver to perform PCI Memory, PCI I/O, and PCI Configuration cycles on a PCI Bus. It also provides services to perform different types of bus mastering DMA on a PCI bus.

PCI Root Bridge Protocol abstracts device specific code from the system memory map. This allows system designers to make changes to the system memory map without impacting platform independent code that is consuming basic system resources. An example of such system memory map changes is a system that provides non-identity memory mapped I/O (MMIO) mapping between the host processor view and the PCI device view.

3.1.2. PCI Driver Model

The PCI Driver Model is designed to extend the UEFI Driver Model in a way that supports PCI Bus Drivers and PCI Device Drivers. This applies to Conventional PCI, PCI-X, and PCI Express.

PCI Bus Drivers manage PCI buses present in a system. The PCI Bus Driver creates child device handles that must contain a Device Path Protocol instance and a PCI I/O Protocol instance. The PCI I/O Protocol is used by the PCI Device Driver to access memory and I/O on a PCI controller.

PCI Device Drivers manage PCI controllers present on PCI buses. The PCI Device Drivers produce an I/O abstraction that may be used to boot an UEFI compliant operating system.

3.2. PCI-X Mode 2 and PCI Express

The PCI-X Mode 2 and PCI Express provide a software programming model that is software compatible with the Conventional PCI.

UEFI PCI I/O Protocol supports up to 4 GiB of configuration space; therefore, it covers the PCI-X Mode 2 and PCI Express Extended Configuration space of 4 KiB in size.

UEFI uses a single timer interrupt in pre-boot, UEFI device drivers are polled so INTx, MSI, or MSI-X is not used by UEFI.

To identify a function (e.g., Conventional PCI, PCI-X vs. PCI Express), the UEFI driver uses Device ID, Vendor ID, and Capability Pointer in the compatibility configuration space.

3.3. EFI Byte Code

The UEFI Specification defines a virtual machine that provides a platform and CPU independent mechanism for loading and executing UEFI device drivers. The instruction set of the virtual machine is called EFI Byte Code, or EBC.

For details of the EBC Virtual Machine, refer to the UEFI Specification.

3.4. Graphics Output Protocol

Graphics Output Protocol (GOP) is defined in the UEFI Specification to remove the hardware requirement to support legacy VGA and INT 10h BIOS. GOP provides a software abstraction to draw on the video screen.

Note: A graphics adapter will still require a performance driver for high speed operation in the operating system.

Note: VGA hardware can support GOP.

3.5. Device State at Firmware/Operating System Handoff

System firmware is only required to configure the boot and console devices. This section specifies the state of the PCI subsystem at firmware handoff and provides guidance to the operating system on how to determine if a particular component was configured by firmware. PCI subsystem refers to components that are compliant to the PCI, PCI-X, or PCI-Express Specifications. In this section, "PCI Specifications" refers to the PCI, PCI-X, or PCI Express Specification.

Firmware owns the PCI subsystem prior to handing control off to the operating system. The handoff point is the return from UEFI ExitBootServices(). After the operating system loader calls ExitBootServices(), the operating system owns the PCI subsystem.

Firmware may provide pre-boot user interaction to allow the system operator to specify the desired boot and console devices.

Firmware must configure the entire path to the console (both input and output) and boot devices. This includes, the chipset, bridges, and multi-function devices. The device configuration is required to load the operating system loader, display boot up messages, and allow operator interaction with the boot process.

Optionally, firmware may configure all devices and bridges in the system. Firmware is not required to configure devices other than boot and console devices.

Since not all devices may be configured prior to the operating system handoff, the operating system needs to know whether a specific BAR register has been configured by firmware. The operating system makes the determination by checking the I/O Enable, and Memory Enable bits in the device's command register, and Expansion ROM BAR enable bits. If the enable bit is set, then the corresponding resource register has been configured.

Note: The operating system does not use the state of the Bus Master Enable bit to determine the validity of the BARs. If the BAR ranges are enabled, the device must respond to those addresses. The device may not be able to master a transaction, but enabled BARs must be configured correctly by firmware.

The operating system is required to configure PCI subsystems:

- During hotplug
- For devices that take too long to come out of reset
- PCI-to-PCI bridges that are at levels below what firmware is designed to configure

Firmware must configure all Host Bridges in the systems, even if they are not connected to a console or boot device. Firmware must configure Host Bridges in order to allow operating systems to use the devices below the Host Bridges. This is because the Host Bridges programming model is not defined by the PCI Specifications. "Configured" in this context means that:

- ❑ Memory and I/O resources are assigned and configured.
- ❑ Includes both the resources consumed by the Host Bridge and the resources passed through to the secondary bus.
- ❑ The bridge is enabled to receive and forward transactions.
- ❑ The bridge is operating in "safe" mode. Safe mode includes:
  - Enabling resources such as: I/O Port, Memory addresses, VGA routing, bus number, etc.
  - Enabling detection of parity and system errors.
  - Programming cacheline, latency timer, and other registers as required by the PCI Specifications.

Firmware must report Host Bridges in the ACPI namespace. Each Host Bridge object must contain the following objects:

- ❑ _HID and _CID
- ❑ _CRS to determine all resources consumed and produced (passed through to the secondary bus) by the host bridge. Firmware allocates resources (Memory Addresses, I/O Port, etc.) to Host Bridges. The _CRS descriptor informs the operating system of the resources it may use for configuring devices below the Host Bridge:
  - _TRA, _TTP, and _TRS translation offsets to inform the operating system of the mapping between the primary bus and the secondary bus.
- ❑ _PRT and the interrupt descriptor to determine interrupt routing.
- ❑ _BBN to obtain a bus number.
- ❑ _UID to match with UEFI device path.
- ❑ _SEG if it has a non-zero PCI Segment Group number.
- ❑ _STA if hot plug is supported.
- ❑ _MAT if hot plug is supported.

Firmware is required to configure all PCI-to-PCI Bridges in the hierarchy leading to boot and console devices. Firmware may optionally configure all PCI-to-PCI Bridges in the system. When configuring a PCI-to-PCI Bridge, Firmware must set it to safe mode. This includes:

- ❑ Programming and enabling resources such as: I/O Port, Memory addresses, VGA routing, bus number, etc.
- ❑ Enabling detection of parity and system errors.
- ❑ If applicable, program cache_line, latency timer, and other registers as required by the PCI Specifications.

- If applicable[7], disable Discard SERR# Enable. The Discard Timer SERR# Enable bit in the Bridge Control Register controls whether the timer waiting for the completion of a delayed transaction generates an SERR (value=1) or simply discards the transaction (value=0) on a time-out. The value of 1 is generally required when peer-to-peer transactions are allowed to and from cards under that bridge. Allowing peer-to-peer transactions is operating system policy and may not be supported on all platforms. Therefore, the bit should be set to 0 when firmware hands off to the operating system, and any changes to the setting to support peer-to-peer under the PCI-to-PCI Bridges should be made by the operating system.

The operating system may provide software to configure PCI-to-PCI bridges for optimum performance.

All slots with an open MRL must be disabled and their Power Indicators must be turned off. All occupied slots with MRL closed must be enabled and their Power Indicators must be turned on. The slot power state for unoccupied slots with MRL closed is implementation dependent and their Power Indicators must reflect the slot power state.

Firmware must deassert RST# for all occupied PCI slots below the Host Bridge. Firmware must observe required wait times such as Trhfa (RST# High to First configuration Access) after taking a bus out of reset. Firmware only needs to delay once for all PCI bus controllers before handing control to the operating system. This allows the operating system to successively walk PCI buses without having to successively delay (post reset quiesce period, 1 second) for each bus.

PCI-to-PCI Bridges have RST# asserted by default for the secondary bus. Firmware is not required to deassert RST# on secondary buses that are not used for boot and console devices.

UEFI drivers and applications must not change BAR assignments. The PCI BARs and the configuration of any PCI-to-PCI bridge controllers belong to the firmware component that configured the PCI Bus prior to the execution of the device driver.

The operating system must not assume that all devices have been configured. Per Section 2.5.6 of UEFI (Revision.2.8): "The presence of a UEFI driver in the system firmware or in an option ROM does not guarantee that the UEFI driver will be loaded, executed, or allowed to manage any devices in a platform." In addition, UEFI drivers are not involved during PCI hot plug.

Note: The operating system does not have to walk all buses during boot. The kernel can automatically configure devices on request; i.e., an event can cause a scan of I/O on demand.

The operating system can determine if the device's BARs have been configured by firmware by checking the I/O Enable, and Memory Enable bits in the device's command register, and Expansion ROM BAR enable bit. If the enable bit is set, then the corresponding resource has been configured. If the enable bit is not set, the operating system cannot assume that the associated BAR register contains valid information.

The address that a processor uses to access a device is not necessarily the same as the address stored in the device's BAR. The translation (_TRA, _TTP, and _TRS) information is not available to the operating system until after the operating system brings up the ACPI interpreter. The operating system must wait until after the ACPI interpreter is up to determine the address at the processor side associated with the BAR registers configured by firmware. Before re-enabling a resource, the operating system must reprogram the BAR register using a value that falls within the range reported

[7] For example, does not apply to PCI Express.

in the _CRS descriptor of the parent Host Bridge. The operating system must ensure that the address range is not used by any other device below that Host Bridge.

Operating systems must not configure devices with resources outside what is reported by the Host Bridge _CRS. Firmware reports the address ranges that are routed to that particular Host Bridge. There is no guarantee that devices under that bridge will respond to other address ranges.

The Expansion ROM BAR (at 30h) is normally not enabled at the firmware handoff to the operating system and the operating system must assume that the BAR content is invalid. For some devices, when the Expansion ROM BARs are enabled, the device's other BARs are disabled. PCI specifies that a device may share decoders between the Expansion ROM BAR and other BARs, and that device independent software must not access the other BARs when the Expansion ROM BAR is enabled. Firmware may leave the Expansion ROM BARs enabled if it happens to know that the device does not share address decoders. This could be firmware based on the Device ID or firmware that is shipped in the card itself. The device independent operating system software must disable the Expansion ROM when accessing the device via the other BARs. If the operating system wants to use the Expansion ROM, it must "take turns" enabling the Expansion ROM BAR, using the ROM, then disabling the BAR again before resuming access to the card via its other BARs. In other words, the operating system must not assume that the card has dual decoders. The operating system is not prohibited from accessing all the card's resources if it knows that the card has dual decoders and that the Expansion ROM BAR content is correct.

4

4. PCI Services in ACPI

The Advanced Configuration and Power Interface (ACPI) Specification describes a set of common firmware interfaces that enable robust operating system-directed platform device configuration and power management of both individual devices and the entire system. ACPI uses tables to describe system information, features, and methods for controlling those features.

The following sections provide an overview of ACPI interfaces that are relevant to platforms that support PCI hierarchies consisting of Conventional PCI, PCI-X, and PCI Express.

4.1. Enhanced Configuration Access Method Base Address

On PC-compatible systems, the enhanced configuration access mechanism allows PCI configuration space to be accessed using memory primitives rather than I/O-based primitives (CF8/CFC mechanism). For PCI Express and PCI-X Mode 2 hierarchies on these systems, the memory mapped configuration access mechanism is the only way to access the extended configuration space (offsets 256-4095).

This section defines an ACPI-based mechanisms to communicate the memory mapped configuration space base address(es) used for Enhanced Configuration Access Mechanism (defined in PCI-X and PCI Express Local Bus Specifications) to the operating system:

- MCFG: An ACPI table-based mechanism that is used to communicate the memory mapped configuration space base addresses corresponding to the (non-hot removable) PCI Segment Groups and/or base address ranges within a PCI Segment Group available to the system at boot. The memory mapped configuration space address ranges described exclusively through the table mechanism are considered to be non-relocatable and non-hot removable for the current boot.
- _CBA: An ACPI method that is used to report the Enhanced Configuration Access base address for any PCI Segment Groups and/or base address ranges within a PCI Segment group. This allows run-time update for the hot added PCI components.

The MCFG table is only used to communicate the base addresses corresponding to the non-hot removable PCI Segment Groups available to the system at boot. The _CBA method enables the system to describe the base address of the memory mapped configuration space for hot plug capable PCI Segment Groups.

4.1.1. Background

The PCI Express and PCI-X Specifications define the Enhanced Configuration Access Mechanism for the PC-compatible systems, which allows access to the configuration registers via memory mapped address space. The base address of this memory mapped configuration space is platform specific and is communicated to the operating system via system firmware.

In a hierarchy that supports the enhanced configuration access mechanism, the first 256 bytes (offsets 0-255) of PCI 2.3 compatible configuration space can be accessed by either the PCI 2.3 configuration mechanism (CF8/CFC) or using the enhanced configuration mechanism. The extended register configuration space (region from offset 256-4095) can be accessed only via the enhanced configuration mechanism for PCI Express devices and Mode 2 PCI-X devices on the PC-compatible systems.

Table 4-1: Memory Address PCI Express Configuration Space

Memory Address	PCI	PCI-X Mode 1	PCI-X Mode 2	PCI Express
A[(20+n):20] Bus[n:0], where n=0 to 7	Applies	Applies	Applies	Applies
A[19:15] Device[4:0]	Applies	Applies	Applies	Applies
A[14:12] Function[2:0]	Applies	Applies	Applies	Applies
A[11:8] Extended Register[3:0]	N/A	N/A	Applies	Applies
A[7:0] Register[7:0]	Applies	Applies	Applies	Applies

The 256-MiB window of memory mapped configuration space (assuming maximum addressable 4 KiB per function, eight functions per device, 32 devices per bus) defined by PCI-X and PCI Express is capable of describing the entire 256 bus PCI Segment Group as shown in Figure 4-1.

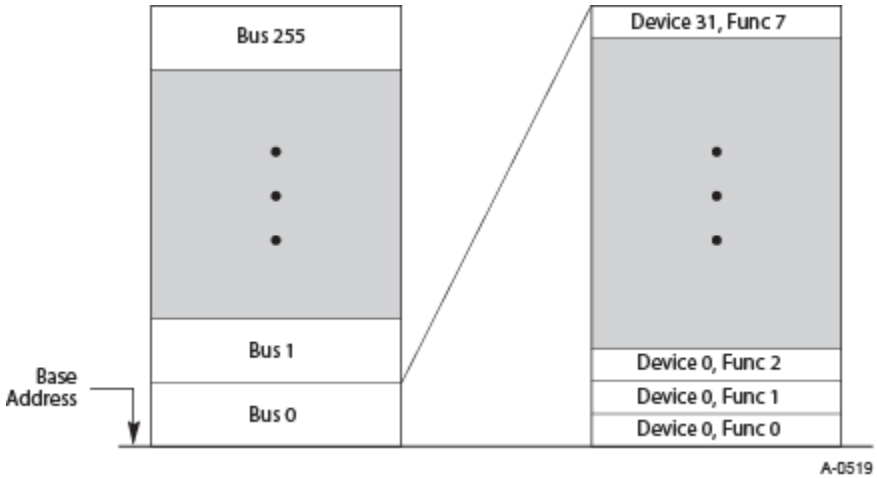


Figure 4-1: 256-MiB Region for Enhanced Configuration Space Access Mechanism

The Enhanced Configuration Access Mechanism can be applied to access heterogeneous hierarchies consisting of PCI/PCI-X/PCI Express. In this case, the extended configuration space up to a limit of 4 KiB per device-function is accessible for PCI Express devices and PCI-X Mode 2 devices.

For non-Mode 2 PCI-X and all PCI devices, only the first 256 bytes of configuration space are accessible.

Note that a given chipset implementation may choose to implement less than 256 MiB for the memory mapped configuration space. Further, implementations with multiple host bridges or mixed hierarchies (for example, a multi-chip implementation with PCI Express as well as PCI-X at the root level) are allowed to implement the memory mapped configuration space in a discontinuous fashion; that is, the 256 PCI buses could be distributed across multiple host bridges in a non-overlapping fashion.

Note that in a multiple host bridge hierarchy, there is no requirement for the host bridges to program the buses within a PCI Segment Group or Groups in a contiguous fashion.

4.1.2. MCFG Table Description

The MCFG table is an ACPI table that is used to communicate the base addresses corresponding to the non-hot removable PCI Segment Groups range within a PCI Segment Group available to the operating system at boot. This is required for the PC-compatible systems.

The MCFG table is only used to communicate the base addresses corresponding to the PCI Segment Groups available to the system at boot. This table directly refers to PCI Segment Groups defined in the system via the _SEG object in the ACPI namespace for the applicable host bridge device. For systems containing only a single PCI Segment Group, the default PCI Segment Group number, namely, PCI Segment Group 0, is implied. In such a case, the default PCI Segment Group need not be represented in the ACPI namespace (i.e., no _SEG method is required in such a hierarchy).

The size of the memory mapped configuration region is indicated by the start and end bus number fields in the Memory mapped Enhanced configuration space base address allocation structure as shown in Table 4-3. 0-255 is the range of allowed bus numbers supported for a given PCI Segment Group.

Table 4-2 provides a description of the MCFG table.

Table 4-2: MCFG Table to Support Enhanced Configuration Space Access

Field	Byte Length	Byte Offset	Description
Header			
Signature	4	0	"MCFG". Signature for the Memory mapped configuration space base address Description Table. (refer to Note 1)
Length	4	4	Length, in bytes, of the entire MCFG Description table including the memory mapped configuration space base address allocation structures.
Revision	1	8	1
Checksum	1	9	Entire table must sum to zero
OEMID	6	10	OEM ID
OEM Table ID	8	16	For the MCFG Description Table, the table ID is the manufacture model ID
OEM Revision	4	24	OEM revision of MCFG table for supplied OEM Table ID
Creator ID	4	28	Vendor ID of utility that created the table
Creator Revision	4	32	Revision of utility that created the table
Reserved	8	36	Reserved
Configuration space base address allocation structure [n]	---	44	A list of the memory mapped configuration base address allocation structures. This list will contain at least one entry corresponding to each PCI Segment Group present in the platform. The structure of this entry is defined in Table 4-3.

Notes:

1. A table signature "MCFG" is reserved for this purpose and the header for the table is shown in Table 4-2. Based on the signature and table revision, the operating system can then interpret the implementation-specific data within the table. The Table Revision for revision 1.0 of the MCFG table is set to 1.
2. If the operating system does not natively comprehend reserving the MMCFG region, the MMCFG region must be reserved by firmware. The address range reported in the MCFG table or by _CBA method (see Section 4.1.3) must be reserved by declaring a motherboard resource. For most systems, the motherboard resource would appear at the root of the ACPI namespace (under _SB) in a node with a _HID of EISAID (PNP0C02), and the resources in this case should not be claimed in the root PCI bus's _CRS. The resources can optionally be returned in Int15 E820h or EFIGetMemoryMap as reserved memory but must always be reported through ACPI as a motherboard resource.
3. This table must not include the memory mapped configuration base addresses for hot pluggable PCI Segment Groups. Such PCI Segment Groups must be described by using the _CBA method (see Section 4.1.3) in the corresponding ACPI namespace object.

The structure in Table 4-3 describes the association between the PCI Segment Group and the corresponding memory mapped configuration base address. This table describes the details of this structure.

Table 4-3: Memory Mapped Enhanced Configuration Space Base Address Allocation Structure

Field	Byte Length	Byte Offset	Description
Base Address	8	0	Processor-relative Base Address for the Enhanced Configuration Access Mechanism
PCI Segment Group Number	2	8	PCI Segment Group Number. Default is 0. For all other PCI Segment Groups, this field value should correspond to the value returned by _SEG object in ACPI namespace for the applicable host bridge device.
Start Bus Number	1	10	Start PCI Bus number decoded by the host bridge
End Bus Number	1	11	End PCI Bus number decoded by the host bridge
Reserved	4	12	Reserved

The MCFG table format allows for more than one memory mapped base address entry (instance of Table 4-3) provided each entry (memory mapped configuration space base address allocation structure) corresponds to a unique PCI Segment Group consisting of 256 PCI buses. Multiple entries corresponding to a single PCI Segment Group are also allowed provided the Field (PCI Segment Group Number, Start Bus Number, and End Bus Number) uniquely identifies each PCI Host Bridge and the bus number values do not overlap.

- The PCI Segment Group Number field denotes the PCI Segment Group corresponding to the base address field in the structure. For systems supporting multiple PCI Segment Groups, this field should correspond to the value returned by _SEG object in ACPI namespace for the applicable host bridge device. If the system only contains a single (default) PCI Segment Group, namely, PCI Segment Group 0, no corresponding _SEG object is required.
- The base address field provides the 64-bit physical address of the base of the memory mapped configuration space associated with the PCI Segment Group. It is the responsibility of the provider of the table to ensure that the base address reported is consistent with the requirements for the hardware implementation. For PCI-X and PCI Express platforms utilizing the enhanced configuration access method, the base address of the memory mapped configuration space always corresponds to bus number 0 (regardless of the start bus number decoded by the host bridge) and further must comply with alignment requirements of the corresponding local bus specification. The unsupported upper bits of the physical address must be set to 0.

4.1.3. The _CBA Method

Some systems may support hot plug of host bridges that introduce either a range of buses within an existing PCI Segment Group or introduce a new PCI Segment Group. For example, each I/O chip in a multi-chip PCI Express root complex implementation could start a new PCI Segment Group. The base address of the memory mapped configuration space for such a hot pluggable PCI Segment Group or a range of buses within a PCI Segment Group is described using an ACPI control method, _CBA, that is under the host bridge devices that are part of the PCI Segment Group. This applies to PC-compatible systems only.

The _CBA (Memory mapped Configuration Base Address) control method is an optional ACPI object that returns the 64-bit memory mapped configuration base address for the hot plug capable host bridge. The base address returned by _CBA is processor-relative address. The _CBA control method evaluates to an Integer.

This control method appears under a host bridge object. When the _CBA method appears under an active host bridge object, the operating system evaluates this structure to identify the memory mapped configuration base address corresponding to the bus number range specified in _CRS method. Within the same PCI Segment Group, different host bridges, each with its associated bus number range, may have a different configuration base address. An ACPI namespace object that contains the _CBA method must also contain a corresponding _SEG method.

For a host bridge that includes _CBA, the _CBA and _BBN control methods have to be executed first to enable PCI_Config_OpRegion access for devices below the bridge. As a result, the _CBA and _BBN methods must not include PCI_Config opregions that refer to devices below the host bridge.

A set of hot pluggable host bridges could have _CBA under each of the host bridge devices, where each host bridge device is typically described in the ACPI namespace with PNP0A08 for _HID and PNP0A03 for _CID. In this case, the memory mapped configuration base address (always corresponds to bus number 0 of a given bus number range) is provided by _CBA and the bus number range covered by the base address is indicated by the corresponding bus number range specified in _CRS.

If rebalancing of resources on a host bridge is supported via _PRS, _SRS, it is the responsibility of the operating system to reevaluate _CBA every time _CRS is evaluated.

Memory mapped configuration base addresses for non-hot pluggable host bridges must be described using MCFG table.

_CBA Control Method

Arguments:

None

Result Code:

Memory mapped configuration base address for the PCI-compatible host bridge returned as an integer

Note: Starting with ACPI 2.0, integers are 64-bit entities.

Example ASL for _CBA usage:

```
Scope(\_SB) {
            …
      Device(PCI1) {                        // Root PCI Bus
            Name(_HID, EISAID("PNP0A03")) // Need _HID for root device
            Name(_SEG, 1)                 // PCI Segment Group 1
            Method (_CRS, ResourceTemplate()
            {
                  …
            }
            Method(_CBA, 0) {
                  // Bits 63:0 of the base address
                  Return (0xE000000000000000)
            } // end of _CBA method
      } // end PCI1
}     // end scope SB
```

4.1.4. System Software Implication of MCFG and _CBA

The base address returned by MCFG table for a given bus number range is always with respect to bus 0 of that particular bus number range as specified in the *PCI Express Base Specification* (and the *PCI-X Specification*). It is the responsibility of system software to calculate the start and end of the supported memory mapped configuration address range based on the start and end bus numbers specified in the MCFG entry. System software must make no assumptions about the memory range corresponding to the base address up to the start of the memory mapped configuration space (as specified by start bus number).

The base address returned by _CBA for a given PCI Segment Group is always with respect to bus 0 as specified in the *PCI Express Base Specification* (and the *PCI-X Specification*). It is the responsibility of system software to calculate the start and end of the supported memory mapped configuration address range for the host bridge based on the bus range supported by the host bridge as identified by _CRS.

IMPLEMENTATION NOTE

Multiple Host Bridges

A platform may have multiple PCI Express or PCI-X host bridges. The base address for the MMCONFIG space for these host bridges may need to be allocated at different locations. Historically, in such cases, using MCFG table and _CBA method as defined in this section means that each of these host bridges must be in its own PCI Segment Group. This approach is referred to as the legacy interpretation of this specification. A newer interpretation of this specification allows "multiple host bridges with base addresses allocated at different locations" to exist within the same PCI Segment Group. Vendors who choose to implement to the newer interpretation bear the responsibility that supported operating systems can handle the newer interpretation.

Allocating base addresses to non-contiguous regions of memory has been achievable in some implementations with an address mapping feature to make non-contiguous regions of memory appear to be contiguous to the operating system. The newer interpretation removes the need for such an address mapping feature thereby simplifying implementations.

Allocating base addresses to non-contiguous regions of memory has also been possible using the legacy interpretation of the specification by employing different PCI Segment Group Numbers. However, the newer interpretation of the specification preserves the use of a single PCI Segment Group Number since implementations will not allow peer-to-peer communications across multiple PCI Segment Groups.

For example, the Base Address field value returned by MCFG of a bus number range starting at 0xE2000000, where the Start Bus Number is 0, would be 0xE2000000. The Base Address field value returned by MCFG of a bus number range starting at 0xE2000000, where the Start Bus Number is 0x40, would be (0xE20000000 – (0x40 << 20)) or 0xDE000000. Therefore, the lowest possible Base Address field value returned by MCFG is a function of Start Bus Number:

Base Address=(Start of memory mapped configuration address range)-((0x100000)*(Start Bus Number)).

If this platform also needs to support legacy operating systems or x86 BIOS Option ROMs, the CF8/CFCh access mechanism (which is PCI Segment Group unaware and only can support up to 256 bus numbers) must also be supported. There may be a number of implementation choices to make these two work together; for example, the bus numbers in each of the PCI Segment Group can be made to not overlap. This would make the CF8/CFCh access still appear to be Segment Group unaware and support up to 256 buses while using the MCFG/_CBA definition to describe host bridges MMCONFIG base addresses that are allocated at different locations.

Note that in this arrangement, even though the PCI Segment Group concept is used, the total number of PCI buses is still limited to 256 due to the CF8/CFCh limitation.

4.1.5. Plug-and-Play ID Defined for Enhanced Configuration Space Access Capable Devices

Currently, a Plug-and-Play ID (PNP ID) of PNP0A03 is used to indicate host-PCI bridge devices in ACPI namespace. This PNP ID is used to describe both PCI/PCI-X hierarchies.

A new PNP ID of PNP0A08 is defined to indicate PCI Express as well as PCI-X Mode 2 host bridges to the operating system. The unique ID for these host bridges will allow the host to distinguish a PCI Express, PCI-X Mode 2 hierarchy in a mixed host-bridge environment and also allow the operating system to tune the driver loading process to the capabilities of the underlying I/O hierarchy.

To retain compatibility with older operating systems that do not recognize the new PNP ID, when PNP0A08 is used to describe a device in the namespace, it is also required to include PNP0A03 as the compatible ID (_CID).

Example ASL for PNP0A08 usage:

```
Device(PCI1) {                          // Root PCI Bus
     Name(_HID, EISAID("PNP0A08"))    // Indicates PCI
     Express/PCI-X Mode 2 host hierarchy
     Name(_CID, EISAID("PNP0A03"))    // To support legacy OS
     that does
                                 // not understand the new HID
     }
```

Notes regarding use of PNPID and PNP0A08 to indicate PCI Express/PCI-X Mode 2 host bridge hierarchy:

1. PNP0A08 only indicates support for extended configuration space, that is, each device function supports 4 KiB of configuration space. The support for memory mapped configuration access is indicated by MCFG/_CBA.
2. For PCI-X Mode 2 host bridges, PNP ID of PNP0A08 only indicates that the host bridge is capable of supporting extended configuration space. OSPM must call _DSM to identify if the extended configuration capabilities are currently enabled.

4.2. Mechanism for Controlling System Wake From PCI Express

An errant PCI Express device may prevent the system from going to sleep by continuously asserting the PCI Express wake signal. The PCI Express wake registers, defined in ACPI FADT PM register, allows OSPM to disable the wake signal to allow the system to enter a sleep state.

Complete details are contained in the *ACPI Specification*.

4.3. PCI Root Bridge Description

4.3.1. Identification

PCI Host bridge devices in ACPI namespace are identified by PNP ID. PNP0A03 is used to indicate PCI/PCI-X host bridge hierarchies.

A PNP ID of PNP0A08 is used to represent PCI Express and PCI-X Mode 2 host bridge hierarchies to indicate support for extended configuration space.

For further details, see Section 4.1.4.

4.3.2. Resource Description

4.3.2.1. Resource Setting

Host bridges resources programming is communicated to the operating system using ACPI methods _CRS, _SRS, and _PRS. _CRS indicates the current resource setting for the host bridge. This includes I/O space, memory space, and bus range assigned to the bridge by platform firmware.

A non-configurable device only specifies _CRS. However, if they are configurable, devices include _PRS to indicate the possible resource setting and _SRS to allow OSPM to specify a new resource allocation for the device.

4.3.2.2. Boot Bus Number

In multiple host bridge systems, _BBN method is used to provide PCI_Config Operation Region access for a specific bus.

4.3.2.3. PCI Segment Group

PCI Segment Group concept enables support for more than 256 buses in a system by allowing the reuse of the PCI bus numbers. The _SEG method is used to uniquely identify PCI Segment Groups.

Complete details are located in the *ACPI Specification*.

4.4. PCI Interrupt Routing

The routing of PCI INTx interrupts to interrupt controllers cannot be deduced through the PCI register space. ACPI object, _PRT, is required under all PCI host bridges and is used to indicate the routing of the PCI INTx interrupts to the interrupt controllers. If the Root Ports are described in the ACPI namespace and have an associated _PRT, operating systems must evaluate and use routing as described in the _PRT.

MSI and MSI-X are standardized by the *PCI-X Specification* and the *PCI Express Base Specification*. ACPI is not involved in the use of MSI or MSI-X.

4.5. _OSC – A Mechanism for Exposing PCI Express Capabilities Supported by an Operating System

_OSC is an object that is used by OSPM to query the capabilities of a device (as defined in ACPI) and to communicate to the platform the feature support or capabilities provided by a device's driver. This object is a child object of the device in the ACPI namespace. Device specific objects are evaluated after _OSC invocation. For a complete description of _OSC method, refer to the *ACPI Specification*.

4.5.1. _OSC Interface for PCI Host Bridge Devices

The _OSC interface defined in this section applies only to "Host Bridge" ACPI devices that originate PCI, PCI-X, or PCI Express hierarchies. These ACPI devices must have a _HID of (or a _CID including either EISAID("PNP0A03") or EISAID("PNP0A08"). For a host bridge device that originates a PCI Express hierarchy, the _OSC interface defined in this section is required. For a host bridge device that originates a PCI/PCI-X bus hierarchy, inclusion of an _OSC object is optional.

The _OSC interface for a PCI/PCI-X/PCI Express hierarchy is identified by the Universal Unique IDentifier (UUID) 33DB4D5B-1FF7-401C-9657-7441C03DD766. A revision ID of 1 encompasses fields defined in this section of this revision of this specification comprised of three DWORDs, including the first DWORD described by the generic ACPI definition of _OSC.

The first DWORD in the _OSC Capabilities Buffer contains bits that are generic to _OSC. These include status and error information.

The second DWORD in the _OSC Capabilities Buffer is the Support Field. Bits defined in the Support Field provide information regarding operating system supported features. Contents in the Support Field are passed one way; the operating system will disregard any changes to this field when returned.

The third DWORD in the _OSC Capabilities Buffer is the Control Field. Bits defined in the Control Field are used to submit requests by the operating system for control/handling of the associated feature, typically (but not limited to) those features that utilize native interrupts or events handled by an operating system-level driver. If any bits in the Control Field are returned cleared (masked to zero) by the _OSC control method, the respective feature is designated unsupported by the platform and must not be enabled by the operating system. Some of these features may be controlled by platform firmware prior to operating system boot or during runtime for a legacy operating system, while others may be disabled/inoperative until native operating system support is available. System firmware must only mask a Control Field bit to zero if it has explicit knowledge that the feature will not work properly under native operating system control, due to platform errata or other incompatibilities.

Since _OSC applies to the entire hierarchy originated by a PCI Host Bridge, and system firmware cannot generally comprehend the features and capabilities of all devices that may be hot-plugged into a system, lack of explicit support of a feature in the system firmware is not a reason to mask a Control Field bit to zero. For example, unless system firmware has knowledge that the system contains hardware that does not work properly with Standard Hot-Plug Controller (SHPC) or PCI Express Native Hot Plug software, it must grant operating system requests for PCI Express Native Hot Plug control and SHPC Native Hot Plug control bits, even if the system does not explicitly support hot plug. It is the operating system's responsibility to determine whether a slot in the hierarchy is hot pluggable by examining the status of the slots based on the *PCI Express Base Specification* and the *PCI SHPC Specification*.

For the above reason, system firmware must always assume that PCI-X features are supported beneath a PCI Express hierarchy unless it has explicit knowledge to the contrary, and must not mask _OSC Control Field bits that apply only to PCI/PCI-X to zero in PCI Express hierarchies. However, if a bit in the _OSC Control Field applies to PCI Express, system firmware must mask this bit to zero in a PCI/PCI-X hierarchy.

If the _OSC control method is absent from the scope of a host bridge device, then the operating system must not enable or attempt to use any features defined in this section for the hierarchy originated by the host bridge. Doing so could contend with platform firmware operations or produce undesired results.

It is recommended that a machine with multiple host bridge devices should report the same capabilities for all host bridges of the same type and also negotiate control of the features described in the Control Field in the same way for all host bridges of the same type. That is, for a machine with multiple host bridge devices supporting both PCI/PCI-X and PCI Express hierarchies, this recommendation applies to each host bridge type separately. PCI/PCI-X host bridge devices can have one capabilities setting and PCI Express hot bridge devices can have another.

Table 4-4: Interpretation of the _OSC Support Field

Support Field Bit Offset	Interpretation
0	**Extended PCI Config operation regions supported** The operating system sets this bit to 1 if it supports ASL accesses through PCI Config operation regions to extended configuration space (offsets greater than FFh). Otherwise, the operating system sets this bit to 0.
1	**Active State Power Management supported** The operating system sets this bit to 1 if it natively supports configuration of Active State Power Management registers in PCI Express devices. Otherwise, the operating system sets this bit to 0.
2	**Clock Power Management supported** The operating system sets this bit to 1 if it supports Clock Power Management and will enable this feature during a native hot plug insertion event if supported by the newly added device. Otherwise, the operating system sets this bit to 0.

<table>
<tr><th>Support Field Bit Offset</th><th>Interpretation</th></tr>
<tr><td>3</td><td>PCI Segment Groups supported
The operating system sets this bit to 1 if it supports PCI Segment Groups as defined by the _SEG object and access to the configuration space of devices in PCI Segment Groups as described by this specification. Otherwise, the operating system sets this bit to 0.</td></tr>
<tr><td>4</td><td>MSI supported
The operating system sets this bit to 1 if it supports configuration of devices to generate message-signaled interrupts, either through the MSI Capability or the MSI-X Capability. Otherwise, the operating system sets this bit to 0.</td></tr>
<tr><td>5</td><td>Optimized Buffer Flush and Fill supported
The operating system sets this bit to 1 if it supports the OBFF feature and will enable this feature during a native hot plug insertion event if supported by the newly added device in a hierarchy where OBFF is supported. Otherwise, the operating system sets this bit to 0.</td></tr>
<tr><td>6</td><td>ASPM Optionality supported
The operating system sets this bit to 1 if it properly recognizes and manages ASPM support on PCI Express components which report support for ASPM L1 only in the ASPM Support field within the Link Capabilities Register. Otherwise, the operating system sets this bit to 0.</td></tr>
<tr><td>7</td><td>Error Disconnect Recover Supported
The operating system sets this bit to 1 if it supports Error Disconnect Recover notification on PCI Express Host Bridges, Root Ports and Switch Downstream Ports. Otherwise, the operating system sets this bit to 0. Refer to the ACPI Specification for more information.
In the context of PCIe, support for Error Disconnect Recover implies that the operating system will invalidate the software state associated with child devices of the port without attempting to access the child device hardware. If the operating system supports Downstream Port Containment (DPC), as indicated by the operating system setting bit 7 of _OSC control field, the operating system shall attempt to recover the child devices if the port implements the Downstream Port Containment Extended Capability. If the operating system continues operation, the operating system must inform the Firmware of the status of the recovery operation via the _OST method. The operating system shall invoke _OST on the same ACPI Device where the Error Disconnect Recover notification was received. The upper word of Arg1 argument of _OST carries the Bus, Device, and Function number of the port that experienced the containment event. The segment number of this port is equal to the segment number of the Error Disconnect Recover notification target. The layout of Arg1 is as follows:
Bit Position Definition
15:0 Status of the operation (Refer to the ACPI Specification).
18:16 Function number of the port that experienced the containment event
23:19 Device number of the port that experienced the containment event
31:24 Bus number of the port that experienced the containment event</td></tr>
</table>

Support Field Bit Offset	Interpretation
8	**_HPX PCI Express Descriptor Setting Record (Type 3) Supported** The operating system sets this bit to 1 if it supports _HPX PCI Express Descriptor Setting Record (Type 3). Otherwise, the operating system sets this bit to 0.
9:31	Reserved

IMPLEMENTATION NOTE

Recommended Use of ASPM Optionality Support Bit

The ASPM Optionality bit in the _OSC Support field is designed to be used by platform firmware for use with devices known a priori by platform firmware to have an ASPM Support field that is write-once or otherwise lockable, configurable by platform firmware, and support ASPM L1 but do not support L0s.

In the case where the operating system does not support ASPM Optionality, such devices would be configured by firmware to report support for both L0s and ASPM L1.

In the case where the operating system supports ASPM Optionality, such devices would be configured to report support for only ASPM L1.

Table 4-5: Interpretation of the _OSC Control Field, Passed in via Arg3

Control Field Bit Offset	Interpretation
0	**PCI Express Native Hot Plug control** The operating system sets this bit to 1 to request control over PCI Express native hot plug. If the operating system sets this bit, it must also set bit 4 to request control of the PCI Express Capability Structure.
1	**SHPC Native Hot Plug control** The operating system sets this bit to 1 to request control over PCI/PCI-X SHPC hot plug. If the operating system sets this bit, it must also set bit 4, PCI Express Capability Structure control, to request control of the PCI Express Capability Structure.
2	**PCI Express Native Power Management Events control** The operating system sets this bit to 1 to request control over PCI Express native power management event interrupts (PMEs). If the operating system sets this bit, it must also set bit 4, PCI Express Capability Structure control, to request control of the PCI Express Capability Structure.
3	**PCI Express Advanced Error Reporting control** The operating system sets this bit to 1 to request control over the PCI Express Advanced Error Reporting Extended Capability structure, if supported, and the other error enable/status bits defined in the *PCI Express Base Specification*. If the operating system sets this bit, it must also set bit 4, PCI Express Capability Structure control, to request control of the PCI Express Capability Structure.

Control Field Bit Offset	Interpretation
4	**PCI Express Capability Structure control** The operating system sets this bit to 1 to request control over the PCI Express Capability structures (standard and extended) defined in the *PCI Express Base Specification, Revision 1.1.* These capability structures are the PCI Express Capability, the Virtual Channel Extended Capability (Capability ID 002, **not** ID 009), the Power Budgeting Extended Capability, and the Device Serial Number Extended Capability structures.
5	**Latency Tolerance Reporting control** The operating system sets this bit to 1 to request control over PCI Express Latency Tolerance Reporting. If the operating system sets this bit, it must also set bit 4, PCI Express Capability Structure control, to request control of the PCI Express Capability tructure.
6	The operating system sets this bit to 1 to indicate that all certified IO drivers will never interpret as valid data an all 1's response to a read from a Function's Memory Mapped IO space (MMIO BAR memory), IO space, or Configuration space. An all 1's read response is returned by hardware to indicate certain error conditions including surprise removal of a PCIe device, and reliable detection of those error cases requires that an all 1's response must never be interpreted as valid by an IO driver.
7	**PCI Express Downstream Port Containment Configuration Control** The operating system sets this bit to 1 to request control over PCI Express Downstream Port Containment configuration. If the operating system sets this bit, it must also set bit 7 of the Support field (indicating support for Error Disconnect Recover notifications) and bits 3 and 4 of the Control field (requesting control of PCI Express Advanced Error Reporting and the PCI Express Capability Structure).
8	**PCI Express Completion Timeout Control** The operating system sets this bit to 1 to request control over the PCI Express Completion Timeout fields. If the operating system sets this bit, it must also set bit 4, PCI Express Capability Structure control, to request control of the PCI Express Capability Structure.
9	**PCI Express System Firmware Intermediary Configuration Control** The operating system sets this bit to 1 to request control over configuration of PCI Express System Firmware Intermediary (SFI). SFI is architected primarily for use by system firmware. In a few cases like embedded systems, it may make sense for the operating system to own SFI instead. It is recommended that general-purpose operating systems never request ownership, and that general-purpose platforms never grant ownership to an operating system if requested.
10 - 31	Reserved

Table 4-6: Interpretation of the _OSC Control Field, Returned Value

Control Field Bit Offset	Interpretation
0	**PCI Express Native Hot Plug control** The firmware sets this bit to 1 to grant control over PCI Express native hot plug interrupts. If firmware allows the operating system control of this feature, it means that the firmware has made the necessary configurations to ensure that all hot plug events are routed to device interrupts as described in the *PCI Express Base Specification* (e.g., switched from SCI/GPE interrupt mechanism used in the legacy hot plug mechanisms), regardless of whether there are hot pluggable slots down the hierarchy. It is the operating system's responsibility to determine whether a slot is hot pluggable by examining the status of the slots based on the *PCI Express Base Specification*. Additionally, after control is transferred to the operating system, firmware must not update the state of hot plug slots, including the state of the indicators and power controller. If control of this feature was requested and denied or was not requested, firmware returns this bit set to 0.
1	**SHPC Native Hot Plug control** The firmware sets this bit to 1 to grant control over PCI/PCI-X SHPC hot plug. If firmware allows the operating system control of this feature, it means that the firmware has made the necessary configurations to ensure that all hot plug events are routed to device interrupts as described in the PCI SHPC Specification 1.0 (e.g., switched from SCI/GPE interrupt mechanism used in the legacy hot plug mechanisms), regardless of whether there are hot pluggable slots down the hierarchy. If the operating system successfully receives control of this feature, it must track and update the status of hot plug slots and handle hot plug events as described in the SHPC Specification. It is the operating system's responsibility to determine whether a slot is hot pluggable by examining the status of the slots based on the PCI SHPC Specification. Additionally, after control is transferred to the operating system, firmware must not update the state of hot plug slots, including the state of the indicators and power controller. If control of this feature was requested and denied or was not requested, firmware returns this bit set to 0. System firmware must always assume that PCI-X features are supported beneath a PCI Express hierarchy unless it has explicit knowledge to the contrary, and must not mask this bit to zero in PCI Express hierarchy.

Control Field Bit Offset	Interpretation
2	**PCI Express Native Power Management Events control** The firmware sets this bit to 1 to grant control over control over PCI Express native power management event interrupts (PMEs). If firmware allows the operating system control of this feature, then in the context of the _OSC method, it must ensure that all PMEs are routed to Root Port interrupts as described in the *PCI Express Base Specification*. Additionally, after control is transferred to the operating system, firmware must not update the PME Status field in the Root Status register or the PME Interrupt Enable field in the Root Control register. If the operating system successfully receives control of this feature, it must handle power management events as described in the *PCI Express Base Specification*. If control of this feature was requested and denied or was not requested, firmware returns this bit set to 0.
3	**PCI Express Advanced Error Reporting control** The firmware sets this bit to 1 to grant control over the PCI Express Advanced Error Reporting Capability structure, if supported, and the other error enable/status bits defined in the *PCI Express Base Specification*: • Correctable Error Reporting Enable • Non-Fatal Error Reporting Enable • Fatal Error Reporting Enable • Unsupported Request Reporting Enable • Correctable Error Detected • Non-Fatal Error Detected • Fatal Error Detected • Unsupported Request Detected If firmware allows the operating system control of this feature, then in the context of the _OSC method, it must ensure that error messages are routed to device interrupts as described in the *PCI Express Base Specification*. Additionally, after control is transferred to the operating system, firmware must not modify the Advanced Error Reporting Capability and the other error enable/status bits listed above. If firmware grants control of Advanced Error Reporting, it must also grant control of the PCI Express Capability structure (bit 4). If firmware retains ownership of Downstream Port containment, it must also retain ownership of Advanced Error Reporting. Additionally, the operating system retains control of AER and the other error enable/status bits across power transitions to and from S1, S2, S3 system power states, and is responsible for saving and restoring the AER and error enable configuration across those transitions when register context may have been lost. If control of this feature was requested and denied or was not requested, firmware returns this bit set to 0, and the operating system must not modify the Advanced Error Reporting Capability or the other error enable/status bits listed above.

Control Field Bit Offset	Interpretation
4	**PCI Express Capability Structure control** The firmware sets this bit to 1 to grant control over the PCI Express Capability, the Virtual Channel Extended Capability, the Power Budgeting Extended Capability, the Advanced Error Reporting Extended Capability, and the Serial Number Extended Capability. If the firmware does not grant control of this feature, firmware must handle configuration of these Capabilities. If firmware grants the operating system control of this feature, any firmware configuration of these Capabilities may be overwritten by an operating system configuration, depending on operating system policy. If the operating system successfully receives control of this feature, it is responsible for configuring the registers in these Capabilities in a manner that complies with the *PCI Express Base Specification*. Additionally, the operating system is responsible for saving and restoring all settings within these Capabilities registers listed above across power transitions to and from S1, S2, S3 system power states when register context may have been lost. If control of this feature was requested and denied or was not requested, firmware returns this bit set to 0, and the operating system must not modify these Capabilities.
5	**Latency Tolerance Reporting control** The firmware sets this bit to 1 to grant control over control over PCI Express Latency Tolerance Reporting. If firmware allows the operating system control of this feature, then in the context of the _OSC method, the firmware must ensure that the LTR Extended Capabilities Structure is initialized properly in all devices that support LTR as described in the *PCI Express Base Specification*. If control of this feature was requested and denied or was not requested, firmware returns this bit set to 0. If the operating system successfully receives control of this feature, it must manage configuration of the LTR Extended Capability structure and LTR Mechanism Enable fields as described in the *PCI Express Base Specification*. Additionally, the operating system is responsible for saving and restoring LTR configuration across power transitions to and from S1, S2, S3 system power states when register context may have been lost. If control of this feature was requested and denied or was not requested, firmware returns this bit set to 0, and the operating system must not modify LTR Capabilities.
6	The firmware sets this bit to 1 to indicate that it will suppress error notification caused by surprise hot remove (Completion Time Out errors).

7	**PCI Express Downstream Port Containment configuration control** Firmware sets this bit to 1 to grant the operating system control over PCI Express Downstream Port Containment configuration. If firmware grants control of Downstream Port Containment, it must also grant control of the Advanced Error Reporting and PCI Express Capability Structures. If firmware retains ownership of Downstream Port containment, it must also retain ownership of Advanced Error Reporting. If firmware allows the operating system control of this feature, then, in the context of the _OSC method firmware must clear the DPC ERR_COR Enable bit in the DPC Control Register (refer to the *PCI Express Base Specification*) to 0. Additionally, after control is transferred to the operating system, firmware must not modify any registers in the Downstream Port Containment Extended Capability Structure. Any firmware configuration of the Downstream Port Containment Capability may be overwritten by the operating system, depending on operating system policy. If control of this feature was requested and denied, or was not requested, the firmware sets this bit to 0, and the operating system must not modify Downstream Port Containment Capabilities except as provided below. If control of this feature was requested and denied, or was not requested, firmware is responsible for initializing Downstream Port Containment Extended Capability Structures per firmware policy. Further, the operating system is permitted to write the following: • Device Status Register • Uncorrectable Error Status Register • Correctable Error Status Register • Root Error Status Register • RP PIO Status Register in the Port that triggered DPC while processing an Error Disconnect Recover notification from firmware. The Error Disconnect Recover notification processing window begins with the Error Disconnect Recover Notify from firmware and ends when the operating system invokes the _OST method to inform firmware of the status of the Error Disconnect Recover notification. During this window, firmware is not permitted to write to these registers. If control of this feature was requested and denied, or was not requested, then the operating system is permitted to write to the DPC Trigger Status bit in the DPC Status Register in the Port that triggered containment and firmware is prohibited from writing to the DPC Status Register: • during the Error Disconnect Recover notification processing window • while processing the subsequent hot-remove of a device under a port with the DPC Trigger Status bit set to 1.

Control Field Bit Offset	Interpretation
	If control of this feature was requested and denied, or was not requested, then the operating system is permitted to write to the following: • DPC Status bit in the DPC Status Register • Device Status Register • Uncorrectable Error Status Register • Correctable Error Status Register in any PCI Function below the port that triggered DPC after firmware returns from the _OST corresponding to the Error Disconnect Recover notification but before the operating system finishes re-initializing the Function or disables the Function. During that time, firmware is prohibited from writing to the DPC Status Register, Device Status Register, Uncorrectable Error Status Register, and Correctable Error Status Register in any PCI Function below the port that triggered DPC.
8	**PCI Express Completion Timeout Control** The firmware sets this bit to 1 to grant control over the PCI Express Completion Timeout fields. If the firmware does not grant control of this feature, firmware must handle configuration of the PCI Express Completion Timeout fields. If firmware grants the operating system control of this feature, any firmware configuration of the PCI Express Completion Timeout fields may be overwritten by an operating system configuration, depending on operating system policy. These fields include Completion Timeout Value (Device Control 2 Register bits 3:0 in the PCI Express Capability Structure) and Completion Timeout Disable (Device Control 2 Register bit 4 in the PCI Express Capability Structure). If the operating system successfully receives control of this feature, it is responsible for configuring the PCI Express Completion Timeout fields in a manner that complies with the *PCI Express Base Specification*. Additionally, the operating system is responsible for saving and restoring all PCI Express Completion Timeout field settings across power transitions to and from S1, S2, S3 system power states when register context may have been lost. If control of this feature was requested and denied, or was not requested, the firmware returns this bit set to 0, and the operating system must not modify PCI Express Completion Timeout fields. If the firmware does not grant control of this feature, firmware must handle configuration of the PCI Express Completion Timeout fields.

Control Field Bit Offset	Interpretation
9	**PCI Express System Firmware Intermediary Configuration Control** Firmware sets this bit to 1 to grant the operating system control over PCI Express System Firmware Intermediary configuration. If firmware grants the operating system control of this feature, then the operating system is responsible for initializing System Firmware Intermediary Extended Capability Structure; further, the operating system retains control of SFI across power transitions for S1, S2, S3 system power states. If this bit is set by the operating system, this indicates that it supports both native operating system control and firmware ownership models of SFI. If control of this feature was requested and denied, or was not requested, the firmware returns this bit set to 0, and the operating system must not modify System Firmware Intermediary registers. If control of this feature was requested and denied, firmware is responsible for initializing System Firmware Intermediary Extended Capability Structures per firmware policy.
10 – 31	Reserved

4.5.2. Rules for Evaluating _OSC

This section defines when and how the operating system must evaluate _OSC as well as restrictions on firmware implementations.

4.5.2.1. Query Flag

If the Query Support Flag (Capabilities DWORD 1, bit 0) is set by the operating system when evaluating _OSC, no hardware settings are permitted to be changed by firmware in the context of the _OSC call. It is strongly recommended that the operating system evaluate _OSC with the Query Support Flag set until _OSC returns the Capabilities Masked bit clear, to negotiate the set of features to be granted to the operating system for native support. A platform may require a specific combination of features to be supported natively by an operating system before granting native control of a given feature.

4.5.2.2. Evaluation Conditions

The operating system must evaluate _OSC under the following conditions:

- During initialization of any driver that provides native support for features described in the section above. These features may be supported by one or many drivers, but should only be evaluated by the main bus driver for that hierarchy. Secondary drivers must coordinate with the bus driver to install support for these features. Drivers may not relinquish control of features previously obtained; i.e., bits set in Capabilities DWORD3 after the negotiation process must be set on all subsequent negotiation attempts.
- When a Notify(<device>, 8) is delivered to the PCI Host Bridge device.
- Upon resume from S4. Platform firmware will handle context restoration when resuming from S1-S3.

4.5.2.3. Sequence of _OSC Calls

The following rules govern sequences of calls to _OSC that are issued to the same host bridge and occur within the same boot:

- The operating system is permitted to evaluate _OSC an arbitrary number of times.
- If the operating system declares support of a feature in the Status Field in one call to _OSC, then it must preserve the set state of that bit (declaring support for that feature) in all subsequent calls.
- If the operating system is granted control of a feature in the Control Field in one call to _OSC, then it must preserve the set state of that bit (requesting that feature) in all subsequent calls.
- Firmware may not reject control of any feature it has previously granted control to.

There is no mechanism for the operating system to relinquish control of a feature previously requested and granted.

4.5.2.4. Dependencies Between _OSC Control Bits

Because handling of hot-plug events, power management events, advanced error reporting, downstream port containment, and completion timeout all require the modification of PCI Express Capability registers, the operating system is required to claim control over the PCI Express Capability (bit 4 of the Control field) in conjunction with claiming control over PCI Express Native Hot Plug, PCI Express Native Power Management Events, PCI Express Advanced Error Reporting, Latency Tolerance Reporting, Downstream Port Containment, or PCI Express Completion Timeout (bits 0, 2, 3, 5, 7 and 8 of the Control field). If the operating system attempts to claim control of any of these features without also claiming control over the PCI Express Capability, the firmware is required to refuse control of the feature being illegally claimed and mask the corresponding bit.

Because handling of Downstream Port Containment has a dependency on Advanced Error Reporting, the operating system is required to request control over Advanced Error Reporting (bit 3 of the Control field) while requesting control over Downstream Port Containment Configuration (bit 7 of the Control field). If the operating system attempts to claim control of Downstream Port Containment Configuration without also claiming control over Advanced Error Reporting, firmware is required to refuse control of the feature being illegally claimed and mask the corresponding bit. Firmware is required to maintain ownership of Advanced Error Reporting if it retains ownership of Downstream Port Containment Configuration.

If the operating system sets bit 7 of the Control field, it must set bit 7 of the Support field, indicating support for the Error Disconnect Recover event.

Operating systems must comprehend that platforms may not grant control of the PCI Express Advanced Error Reporting feature and PCI Express Downstream Port Containment Configuration feature.

If firmware does not support _DSM for Downstream Port Containment Enable and Hot-Plug Surprise Control, then Operating Systems, drivers, and software determine that slots in the platform support surprise removal by consulting the Hot-Plug Surprise bit in the Slot Capabilities register of corresponding root port or switch downstream port.

If firmware does support _DSM for Downstream Port Containment Enable and Hot-Plug Surprise Control, then the Hot-Plug Surprise bit may be cleared by the platform and therefore, it is recommended that firmware advertise whether or not slots in the platform support surprise removal (also referred to as 'async removal') via bit 4 of Slot Characteristics 2 in the SMBIOS Type 9 structure (System Slots). Refer to the *SMBIOS Reference Specification*, Version 3.4.0a or newer for details on bit 4 of Slot Characteristics 2 in the SMBIOS Type 9 structure.

4.5.3. ASL Example

A sample _OSC implementation for a mobile system incorporating a PCI Express hierarchy is shown below:

```
Device(PCI0)  // Root PCI bus
{
   Name(_HID,EISAID("PNP0A08"))  // PCI Express Root Bridge
   Name(_CID,EISAID("PNP0A03"))  // Compatible PCI Root Bridge

   Name(SUPP,0)  // PCI _OSC Support Field value
   Name(CTRL,0)  // PCI _OSC Control Field value

   Method(_OSC,4)
   {  // Check for proper UUID
      If(LEqual(Arg0,ToUUID("33DB4D5B-1FF7-401C-9657-7441C03DD766")))
      {
         // Create DWORD-addressable fields from the Capabilities Buffer
         CreateDWordField(Arg3,0,CDW1)
         CreateDWordField(Arg3,4,CDW2)
         CreateDWordField(Arg3,8,CDW3)

         // Save Capabilities DWORD2 & 3
         Store(CDW2,SUPP)
         Store(CDW3,CTRL)

         // Only allow native hot plug control if the OS supports:
         // * ASPM
         // * Clock PM
         // * MSI/MSI-X
         If(LNotEqual(And(SUPP, 0x16), 0x16))
         {
            And(CTRL,0x1E,CTRL) // Mask bit 0 (and undefined bits)
         }

         // Always allow native PME, AER (no dependencies)

         // Never allow SHPC (no SHPC controller in this system)
         And(CTRL,0x1D,CTRL)

         If(Not(And(CDW1,1)))    // Query flag clear?
         {  // Disable GPEs for features granted native control.
            If(And(CTRL,0x01))  // Hot plug control granted?
            {
               Store(0,HPCE) // clear the hot plug SCI enable bit
               Store(1,HPCS) // clear the hot plug SCI status bit
            }
            If(And(CTRL,0x04))  // PME control granted?
            {
               Store(0,PMCE) // clear the PME SCI enable bit
               Store(1,PMCS) // clear the PME SCI status bit
            }
            If(And(CTRL,0x10))  // OS restoring PCI Express cap structure?
            {  // Set status to not restore PCI Express cap structure
               // upon resume from S3
               Store(1,S3CR)
            }
         }

         If(LNotEqual(Arg1,One))
         {  // Unknown revision
            Or(CDW1,0x08,CDW1)
```

```
        }

        If(LNotEqual(CDW3,CTRL))
        {   // Capabilities bits were masked
            Or(CDW1,0x10,CDW1)
        }
        // Update DWORD3 in the buffer
        Store(CTRL,CDW3)
        Return(Arg3)
    } Else {
        Or(CDW1,4,CDW1)      // Unrecognized UUID
        Return(Arg3)
    }
}   // End _OSC

  // End PCI0
```

4.6. _DSM Definitions for PCI

Device Specific Method (_DSM) is defined in the *ACPI Specification*. This object is a control method that enables devices to provide device specific control functions that are consumed by the device driver. Table 4-7 below lists the UUID, Revision, and Function definitions.

The valid values of the Revision ID argument for an individual _DSM Function range from the Initial Revision ID value (lowest value) as shown in Table 4-7 to the Current Revision ID value (highest value).

The Current Revision ID value for this revision of this specification is 6. For example: a valid Revision ID value for Function 8 is 3 through the Current Revision ID value inclusive; a value of 0, 1 or 2 is not valid for Function 8. Any value passed in as the Revision ID argument greater than the Current Revision ID is invalid.

The Current Revision ID value for all Functions of this _DSM is incremented whenever a new Function is added, or when the definition of an existing Function definition is modified in a way that affects its operation.

Note: It is possible that the functionality of a given _DSM Function may change between Revision ID values, in a backwards-compatible manner. System firmware may wish to support all valid values of the Revision ID, but in doing so must process the _DSM Function invocation in a manner expected by the Revision ID value passed in by OSPM. If system firmware chooses to support only a subset of the valid Revision IDs, it uses the return value of _DSM Function 0 to convey the supported Functions based on the Revision ID value to OSPM.

System firmware must support the same Revision ID value for all implemented functions. If OSPM invokes a Function with an invalid Revision ID, system firmware must not take any action on inputs, and any returned data may be invalid.

OSPM must invoke all Functions other than Function 0 with the same Revision ID value. The Revision ID value is determined by OSPM's invocation of Function 0 with the highest known (Current Revision ID) value for the Revision ID. If system firmware returns an empty Buffer or a Buffer containing a single entry with a value of 0 (no bits set), the UUID is unknown and OSPM should not attempt to invoke any Functions with that UUID. If system firmware returns a Buffer containing a single entry with a value of 1 (bit 0 set), the UUID is known and supported by system firmware, but the Revision ID is unknown. If this occurs, OSPM should successively invoke Function 0 with decremented values of Revision ID until system firmware returns a value indicating support for more than Function 0.

Table 4-7: _DSM Definitions for PCI

UUID	Initial Revision ID	Function	Description
E5C937D0-3553-4D7A-9117-EA4D19C3434D	1	1	PCI Express Slot Information
	1	2	PCI Express Slot Number
	1	3	Vendor-specific Token ID
	1	4	PCI Bus Capabilities
	2	5	Preserve PCI Boot Configuration
	2	6	LTR Maximum Latency
	2	7	Naming a PCI or PCI Express device under OS
	3	8	Avoid power-on Reset Delay Duplication on Sx Resume
	3	9	Device Readiness Durations
	4	0Ah	Request D3cold Aux Power Limit
	4	0Bh	Add PERST# Assertion Delay
	5	0Ch	Downstream Port Containment and Hot-Plug Surprise Control
	5	0Dh	Locate the port that experienced the containment event
	6	0Eh	Query Platform Vendor Specific TPH Features

4.6.1. _DSM for PCI Express Slot Information

This section describes how the PCI Express slot information is exposed through the _DSM ACPI method. In the future, this information may be reported through hardware mechanisms that the operating system has direct access to; if and when that is available, the information provided through the hardware mechanism overrides the information provided through the mechanism defined in this section. The operating system should not use the mechanism defined in this section.

Note: The SMBIOS Specification defines the Type 9 entry structure and enumerations that allows reporting slot information for PCI/PCI-X/PCI Express slots. However, the SMBIOS access mechanism does not comprehend platforms with dynamic characteristics that would render statically reported SMBIOS data invalid. _DSM mechanism is used on such platforms to report the equivalent of the SMBIOS Type 9 slot information. The UUID in _DSM in this context is {E5C937D0-3553-4D7A-9117-EA4D19C3434D}, the revision is 1, and the function is 1.

Note: Function 0 is a generic Query function that is supported by _DSMs with any UUID and Revision ID. The definition of function 0 is generic to _DSM.

Location:

This object will be placed under the virtual PCI-to-PCI bridge object representing the PCI Express root port or switch port that generates the slot on the motherboard.

Arguments:

Arg0: UUID: E5C937D0-3553-4d7a-9117-EA4D19C3434D

Arg1: Revision ID: lowest valid Revision ID value: 1

Arg2: Function Index: 1

Return:

ACPI Buffer type; the definition of the return is a package of two items and the description is as follows:

Package item 1:

Type: Integer

Purpose: status of operation

Description:

0: Failure

1: Success

Package item 2:

Type: Package

Purpose: PCI Express Slot Information

Description: A package of three integers:

Integer 1:

Bit	Position
0	Supports x1
1	Supports x2
2	Supports x4
3	Supports x8
4	Supports x12
5	Supports x16
Others	Reserved

This integer indicates the link width capabilities of the slot. More than one bit can be set in this field to indicate the link widths supported by the slot. For example, a x8 slot that is capable of supporting x4, x2, and x1 would be indicated by setting bits 0, 1, and 2 in this field. This field indicates the downshift capabilities to management software. For maximum and negotiated link width, refer to the *PCI Express Base Specification*.

Integer 2:

Value	Meaning
0h	Unknown
1h	PCI Express Card Slot
2h	PCI Express Server I/O Module Slot
3h	PCI Express ExpressCard* Slto
4h	PCI Express Mini Card Slot
5h	PCI Express Wireless Form Factor Slot
Others	Reserved

This field indicates the type of the slot.

Integer 3:

Bit	Position
0	SMBus signal
1	WAKE# signal
Others	Reserved

This field indicates the supported signal(s) of the slot. More than one bit can be set in this field to indicate the signals supported by the slot.

4.6.2. _DSM for PCI Express Slot Number

This section describes how PCI Express slot number information is exposed through the _DSM ACPI method. It provides a method to break the monolithic slot numbers that are already provided both in ACPI firmware (through the _SUN method) and in PCI Express hardware (through the PCI-to-PCI bridge Chassis ID and the PCI Express Slot Number) into tokens that each represent a discrete hardware element. This allows the slot numbers to be displayed to the user in a more meaningful way.

This method only indicates the system-specific mechanism for interpreting the sub-fields of slot numbers. It does not mandate the exact format in which this data must be displayed to the user.

This method may exist in any device in the _SB scope in the ACPI namespace. The information returned from this method applies to any PCI device underneath the device containing the method, unless it is over-ridden by another instance of this method further down in the namespace tree.

The UUID in _DSM in this context is {E5C937D0-3553-4D7A-9117-EA4D19C3434D}, the revision is 1, and the function is 2.

Note: Function 0 is a generic Query function that is supported by _DSMs with any UUID and Revision ID. The definition of function 0 is generic to _DSM.

Arguments:

Arg0: UUID: E5C937D0-3553-4d7a-9117-EA4D19C3434D

Arg1: Revision ID: lowest valid Revision ID value: 1

Arg2: Function Index: 2

Return:

A Package of token packages. The returned package must contain at lease one token package.

Each sub-package represents a slot number token, which is a bit range within one of the monolithic slot numbers that has a distinct meaning. For example, bits 4-7 of the PCI Express Slot Number could represent the I/O Cabinet the slot sits in, while bits 0-3 would be the slot number within the cabinet. Each of these fields would be represented by a different token.

Each token is a package of four items:

Source ID – This field indicates the source data of the bit range that this token represents.

Token ID – This field indicates the type of element the token represents. Token IDs 0-7Fh are reserved by this specification. Token IDs 80h - FFhare vendor-specific.

Start bit – This field indicates the first bit number of the token in the source data.

End bit – This field indicates the last bit number of the token in the source data.

Token IDs:

0 – Chassis
1 – Cabinet
2 – I/O Tray
3 – Module
4 – Slot
5-7Fh – Reserved
80h-FFh – Vendor-specific

Source IDs:

0 – The _SUN method in the ACPI namespace

1 – The data in the Chassis ID register of the PCI-to-PCI bridge Slot Numbering Capability

2 – The data in the Physical Slot Number field of the Slot Capabilities register in the PCI Express Capability

Example:

```
Method (_DSM, 3) {           // Assume GUID and revision match
  Return (Package(3){        // PCI Express Slot Parsing
     Package(4){
        1,1,2,7              // bit 7:2 in the PPB Chassis ID
     },                      // Token represents a "Cabinet"
     Package(4){
        1,2,0,1              // bit 1:0 in the PPB Chassis ID
     },                      // Token represents an "I/O Tray"
     Package(4){
        2,4,0,7              // bit 7:0 in the PCI Express Physical
     }                       // Slot Number
  })
}
```

If the PPB Chassis ID of a device in the hierarchy under the location of this method contained the value 0100 1110b and the PCI Express Physical Slot Number contained 17Ah, the device's slot number could be displayed as "Cabinet 13h, I/O Tray 2, Slot 7Ah".

4.6.3. _DSM for Vendor-specific Token ID Strings

For vendor-specific token IDs, the tokens themselves do not provide enough information to display slot number information to the user. This section provides a method for the BIOS to return a human readable description in multiple languages of the hardware element represented by a vendor-specific token ID. In the example usage of the _DSM for PCI Express Slot Number in Section 4.6.2, if a vendor-specific token ID is used, the resulting string displayed to the user would contain a string returned from this method instead of a standard string like "Cabinet" or "Slot".

This method returns a package of packages. Each sub-package consists of a Language identifier and corresponding Unicode string for a given locale. Specifying a language identifier allows OSPM to easily determine if support for displaying the Unicode string is available. OSPM can use this information to determine whether or not to display the device string, or which string is appropriate for a user's preferred locale. It is assumed that OSPM will always support the primary English locale to accommodate English embedded in a non-English string, such as a brand name. If OSPM does not support the specific sub-language ID, it may choose to use the primary language ID for displaying device text.

The language IDs returned by this method indicate that the corresponding string follows the format specified in RFC 3066. Strings are returned in Unicode (UTF-16).

In addition to supporting the existing strings in RFC 3066, the following aliases are also supported:

RFC String	Supported Alias String
zh-Hans	zh-chs
zh-Hant	zh-cht

Location:

This method must exist in the same context as the _DSM for PCI Express Slot Number.

.

Arguments:

Arg0: UUID: E5C937D0-3553-4d7a-9117-EA4D19C3434D

Arg1: Revision ID: lowest valid Revision ID value: 1

Arg2: Function Index: 3

Arg3: A package containing one element. The element is the Token ID to return the string for. The Token ID must be an integer in the range 80h-FFh.

Return:

A package of packages.

Each sub-package contains two elements:

The language ID of the string

The Unicode string representing the token ID

4.6.4. _DSM for PCI Bus Capabilities

This section defines a mechanism for a PCI root bus to report its capabilities and operating mode to software. PCI buses that are created by PCI to PCI bridges already report this information through hardware registers in the configuration space header, PCI-X capability and/or SHPC capability. Since root buses are exposed through ACPI, no such hardware registers exist, and this information must be described in a different manner.

The _DSM method, defined in the *ACPI Specification*, is used to provide the information about the capabilities of a root bus. The UUID used for the bus capabilities information and the structure of the buffer returned is described below.

Note: SMBIOS has type 9 entry reporting slot information, but the interface is static with no dynamic update capability. The entry has not been updated to reflect the most current technologies. The _DSM mechanism replaces the SMBIOS interface.

4.6.4.1. Bus Capabilities Structure

This section defines the concept of Bus Capabilities Structure. This structure describes the different attributes of a root bus. The Bus Capabilities Structure is bus-type specific and is returned by the _DSM defined here. A root bus can only have a single Bus Capability Structure.

The Bus Capabilities Structure must be aligned to a 4-byte boundary to ensure that future structure definitions do not cause any alignment issues for the consumer of these structures.

4.6.4.1.1. Bus Types

Table 4-8 describes the different bus types that can be used in the type field of the Bus Capabilities Structures. The type field is 2 bytes in size.

Table 4-8: Bus Types

Type	Description
0	Reserved
1	PCI
2+	Reserved

4.6.4.1.2. Bus Capabilities Structure Definitions

The Bus Capabilities Structures, for different bus types, are discussed in the following table.

Table 4-9: PCI Bus Capability Structure

Field	Byte Length	Byte Offset	Description
Type	2	0	PCI (Type 1)
Length	2	2	The length of the structure, in bytes, including the type, starting from offset 0. This field is used to record the size of the entire structure. This field must return 16.
Attributes	1	4	The bit corresponding to each attribute of the bus should be set: 01h: 64-bit Device – This bit should be set if the secondary interface of the bus is 64 bits wide. 02h: PCI-X Mode 1 ECC Capable – This bit should be set if the bus is capable of supporting ECC in PCI-X Mode 1. 04h: Device ID Messaging Capable – This bit should be set if the bridge is capable of forwarding Device ID Message transactions. 08h: OBFF Capable – This bit should be set if the OBFF/WAKE# signal is provided to devices beneath this bridge, but the bridge does not report support for OBFF in its Device Capabilities 2 register.
Current Speed/Mode	1	5	Defined encodings are as follows: 0000 0000b – 33 MHz Conventional Mode 0000 0001b – 66 MHz Conventional Mode 0000 0010b – 66 MHz PCI-X with Parity (Mode 1) 0000 0011b – 100 MHz PCI-X with Parity (Mode 1) 0000 0100b – 133 MHz PCI-X with Parity (Mode 1) 0000 0101b – 66 MHz PCI-X with ECC (Mode 1) 0000 0110b – 100 MHz PCI-X with ECC (Mode 1) 0000 0111b – 133 MHz PCI-X with ECC (Mode 1) 0000 1000b – 66 MHz PCI-X 266 (Mode 2) 0000 1001b – 100 MHz PCI-X 266 (Mode 2) 0000 1010b – 133 MHz PCI-X 266 (Mode 2) 0000 1011b – 66 MHz PCI-X 533 (Mode 2) 0000 1100b – 100 MHz PCI-X 533 (Mode 2) 0000 1101b – 133 MHz PCI-X 533 (Mode 2) 0000 1110b to 1111 1110b – Reserved

Field	Byte Length	Byte Offset	Description
Supported Speeds/Modes	2	6	The bit corresponding to each supported bus speed/mode should be set: Bit[5:0] 00 0001b: Conventional PCI 33 MHz 00 0010b: Conventional PCI 66 MHz 00 0100b: PCI-X 66 MHz 00 1000b: PCI-X 133 MHz 01 0000b: PCI-X 266 MHz 10 0000b: PCI-X 533 MHz Bit[15:6] Reserved
Voltage	1	8	0: 3.3 V 1: 5 V 2+: Reserved
Reserved	7	9	0

4.6.4.1.3. _DSM for Bus Capabilities

This section describes how the Bus Capabilities structure is exposed through the _DSM ACPI method.

The _DSM method must appear in the context of a PCI root bus, identified by a _HID or _CID of PNP0A03. The UUID for reporting a Bus Capabilities structure in a _DSM in this context is {E5C937D0-3553-4D7A-9117-EA4D19C3434D}, the revision is 1, and the function is 4.

Note: Function 0 is a generic Query function that is supported by _DSMs with any UUID and Revision ID.

Location:
This object will be placed under the object representing the PCI or PCI-X root bus in the ACPI namespace. This object should have a _HID or _CID of PNP0A03.

Arguments:
Arg0: UUID: E5C937D0-3553-4d7a-9117-EA4D19C3434D
Arg1: Revision ID: lowest valid Revision ID value: 1
Arg2: Function Index: 4
Arg3: Empty Package

Return:
ACPI Buffer type; the definition of the return a package of two items and the description is as follows.

Package item 1:
Type: Integer
Purpose: status of operation
Description: 0: Failure
1: Success

Package item 2:
Type: Buffer
Purpose: Capabilities structure
Description: The buffer layout will be as follows:

Field	Byte Length	Byte Offset	Value	Description
Version	2	0	1	Version of the return buffer
Status	2	2	Zero or Ones	Status return of this method: Zero indicates success, Ones indicate failure.
Length	4	4	Computed	The length of the entire buffer, in bytes, including the version, starting from offset 0. This field is used to record the size of the entire buffer.
Bus Capabilities Structure	Dependent on capability type and size of structure	8	Computed	The bus capabilities as defined in the Bus Capabilities Structure Definitions

4.6.5. _DSM for Preserving PCI Boot Configurations

This section describes how system firmware can indicate to an operating system that it should preserve assignments of PCI bus numbers, I/O ports, and memory address space done by firmware, or whether it can reassign those resources. The indication is exposed through Function 5 of the _DSM ACPI method.

This _DSM function is optional.

Location:
This _DSM function can be placed under any object representing:

- A PCI Express host bridge (PNP0A08)
- A PCI Express Root Port, Switch Port, or Endpoint
- A PCI Express Root Port Complex Integrated Endpoint
- A Conventional PCI host bridge (PNP0A03)
- Conventional PCI/PCI-X bridge
- A Conventional PCI/PCI-X device

The function that applies to a device is either:

- A _DSM Function 5 implemented by an ACPI object that represents the device itself (if present), or
- _DSM Function 5 in the nearest ACPI namespace scope that encloses the device.

If Function 5 applies to a bridge, it applies to the bridge itself (BARs, windows, and secondary/subordinate bus numbers), as well as to the hierarchy produced by the bridge, unless it is overridden by another Function 5 lower in the hierarchy.

Note: The resources consumed by a PCI host bridge, including those it makes available on the PCI root bus below it, are described by the bridge's _CRS method. If an operating system needs to change the root bus resources, it must do so by evaluating the host bridge _PRS and _SRS methods. A host bridge that does not supply _PRS/_SRS is not configurable by the operating system.

Arguments:

Arg0: UUID: E5C937D0-3553-4d7a-9117-EA4D19C3434D

Arg1: Revision ID: lowest valid Revision ID value: 2

Arg2: Function Index: 5

Arg3: Empty Package

Return:

An integer whose description is as follows:
Type: Integer
Purpose: Preserve PCI device configuration
Description:

0: Yes (The operating system must preserve PCI resource assignments made by firmware at boot time. Except while sizing BARs, the operating system must not modify device BARs,

bridge windows, or bridge secondary and subordinate bus numbers of this device even though this restriction applies to the hierarchy below a bridge (unless another _DSM in that hierarchy also implements Function 5), the operating system is free to assign resources to devices in this hierarchy that have invalid BAR, window, or bus number settings. There may be a reduced level of hot plug capability support in this hierarchy due to resource constraints.)

1: No (The operating system need not preserve any PCI resource assignments made by firmware at boot time and it may reconfigure/rebalance BARs, bridge windows and bus numbers in the hierarchy.)

4.6.6. _DSM Definitions for Latency Tolerance Reporting

This section describes how PCI Express firmware describes Latency Tolerance Reporting information to the operating system.

For Root Complexes that support Latency Tolerance Reporting, there is a set of maximum latency values that could never be exceeded in normal operation. Firmware must convey the maximum values for each Downstream Port embedded within the platform. The operating system is responsible for calculating latencies along the path between each Downstream Port and any Upstream Port (Switch Upstream Port or Endpoint) supporting LTR beneath the port, and programming the sum into the Upstream Port's Latency Tolerance Reporting Extended Capability Structure.

Arguments:

Arg0: UUID: E5C937D0-3553-4d7a-9117-EA4D19C3434D

Arg1: Revision ID: lowest valid Revision ID value: 2

Arg2: Function Index: 6

Arg3: Empty Package

Return:

A Package of four integers corresponding with the LTR encoding defined in the *PCI Express Base Specification*, as follows:

Integer 0: Maximum Snoop Latency Scale
Integer 1: Maximum Snoop Latency Value
Integer 2: Maximum No-Snoop Latency Scale
Integer 3: Maximum No-Snoop Latency Value

These values correspond directly to the LTR Extended Capability Structure fields described in the *PCI Express Base Specification*.

4.6.7. _DSM for Naming a PCI or PCI Express Device Under Operating Systems

This interface will return a package with the following information relative to the device or slot:

- First entry of the package is the instance number.

- Instance number must be unique under _SB scope. This instance number does not have to be sequential in a given system configuration.
- This entry is mandatory when this _DSM is implemented.
- In typical usage model, in a configuration with multiple devices of the same type, device with lower instance number would be assigned lower device label by the operating system compared to another device with higher instance number.

❑ Second entry of the package is a string name:

- This string is optional even when this _DSM is implemented; when not implemented, this entry must return a null string.
- When populated, this string object should match the label/name on the chassis. String should be specified in Unicode.
- When implemented, this string must be unique for a given platform under _SB.

This interface is defined for devices that are defined in ACPI namespace for a given platform by the BIOS. For expansion slots, typically, platform BIOS will include entry for the PCI or PCI Express slot in its namespace and each of these expansion slots would be expected to include their respective _DSM entries. Assignment of specific device names to multi-function devices installed in expansion slots, and/or PCI or PCI Express devices that are hot-added to expansion slots in operating system-environment would be handled in operating system-specific manner, and is not specified via this specification.

Location:
This object will be placed under the ACPI object representing the embedded PCI or PCI Express device/function or under the ACPI PCI Slot description (as defined in Section 4.7) representing the add-in PCI or PCI Express Slot.

Arguments:

Arg0: UUID: E5C937D0-3553-4d7a-9117-EA4D19C3434D

Arg1: Revision ID: lowest valid Revision ID value: 2

Arg2: Function Index: 7

Arg3: None

Return:
A package of two elements.

```
Method (_DSM, 3) { // Assume GUID and revision match

  If (LAnd( Lequal(Arg1, 2), Lequal(Arg2,7))) {
    Return ( Package(2){ // PCI Express Slot Parsing
               1,        // Instance of the enumeration
               String1   //String name which matches the label on the chassis
    }
  } // end of If (LAnd...
} // end of method
```

IMPLEMENTATION NOTE

When implementing support for this method, the following should be considered: System firmware should include this method for all applicable PCI or PCI Express devices, both embedded and in PCI or PCI Express slots. System firmware should assign appropriate instance numbers to each device to influence the device labeling under operating system accordingly for the given platform design. For an example system configuration that includes two networking devices, Table 4-10 shows a) the location for each device (Bus#, Device#, Function#), b) description of device, c) _DSM (function 7) for each device, and d) the resultant sample device label as assigned by an operating system that takes advantage of this method. In case of this sample configuration, operating system would sort the different networking devices in the system using the instance numbers provided via _DSM, and assign lower NIC-label (NIC1) to device which has lower instance number, as shown in Table 4-10.

Table 4-10: Example Usage of Device/Slot Enumeration Ordering Hints

PCI/PCI Express Device Location	System Device	_DSM Fn 7 Package	Operating System Device Label
Bus 2, Device 4, Function 0	Networking device in PCI Express Slot 1	(**3**, “PCIe Slot 1”)	NIC 2
Bus 4, Device 8, Function 0	Embedded networking device	(2, “Embedded NIC1”)	NIC 1

4.6.8. _DSM for Avoiding Power-On Reset Delay Duplication on Sx Resume

This section describes how system firmware can inform operating system that no additional delay is needed before issuing the first Configuration Access to a device in the PCI subsystem underneath the host bus. System hardware/firmware assume the responsibility of providing the post power-on delay (Conventional Reset delay and Data Link Up (DL_Up) Time delay).

This function is optional. If not present, the operating system should adhere to the post Conventional Reset delays as described in the PCIe Base Specification.

If system firmware assumes the responsibility of post Conventional Reset delay (and informs the Operating System via this _DSM function) on Sx Resume (such as boot from ACPI S5, or resume from ACPI S4 or S3 states), the Operating System may assume sufficient time has elapsed since the end of reset, and devices within the PCI subsystem are ready for Configuration Access.

It is the system integrator's responsibility to ensure that the pre-OS power-on delay sequence covers all devices in the PCI subsystem. If the system firmware supports runtime power gating on any of the device within PCI subsystem covered by this _DSM function, the system firmware is responsible for covering the necessary post power-on reset delay.

If an Operating System (after it has control of the PCI subsystem) has the ability to apply Conventional Reset to devices without system firmware involvement, then it will need to adhere to delays after such reset.

This _DSM function is applicable whether reduction in device readiness timing, via Readiness Notification or _DSM function 9, are available or not.

The returned value of this _DSM function also applies to PCI devices integrated into the Root Complex within the PCI hierarchy.

Location:
This object can only be placed within the scope of a PCI host bus. This _DSM function is intended to cover the PCI subsystem underneath the host bus. If a _DSM Method implementing this function is found within the scope of any other device, any values returned by this function should be ignored.

Arguments:

Arg0: UUID: E5C937D0-3553-4d7a-9117-EA4D19C3434D

Arg1: Revision ID: lowest valid Revision ID value: 3

Arg2: Function Index: 8

Arg3: Empty Package

Return:
An integer whose description is as follows:

Type: Integer

Purpose: Allow operating system to avoid duplicating post power-on delay on Sx resume flow

Description:
0: No (There is no hardware or firmware sequencing to provide post power-on reset delay to PCI subsystem. This situation is the same as the legacy situation where this _DSM is not provided.)

1: Yes (System hardware or firmware assumes responsibility to provide post power-on reset delay to the PCI subsystem. Operating System may assume all devices in the PCI subsystems have completed power-on reset.)

4.6.9. _DSM for Specifying Device Readiness Durations

This section describes how system firmware can indicate to an operating system that it can lower the PCI device initialization delays to less than the duration specified in the *PCI Express Base Specification*. This provides an ACPI interface to provide information equivalent to the Readiness Time Reporting extended capability structure. There are several delays covered in this _DSM:

1. For devices that support faster self-initialization after Conventional Reset (section 6.6.1 in *PCI Express Base Specification* Rev 3.0).
2. For device that support faster recovery time following D3hot-to-D0 transition (section 5.3.1 Device Power Management States in *PCI Express Base Specification*).
3. For a function that has a shorter Functional Level Reset (FLR) latency (section 6.6.2 Function-Level Reset in *PCI Express Base Specification*).
4. For a device that has a shorter Data Link Up (DL_Up) latency (section 6.6.1 in *PCI Express Base Specification*).
5. For a device (PF) supporting one or more VFs with a shorter VF enable latency (section 3.3.3.1 in *PCI Express Base Specification*).

This function is optional. If the platform does not provide it, the operating system must adhere to all timing requirements as described in the *PCI Express Base Specification* and/or applicable form factor specification, including values contained in a Readiness Time Reporting capability structure. Operating systems that take advantage of this function can reduce PCI device initialization delay. Values provided in this function can only be used to lower (reduce) the latency required by specification or values discovered from device.

Delay values reported via this Function which are greater than the minimum delays required by the *PCI Express Base Specification* should be ignored by operating system software.

Delay values reported by firmware must be interpreted as overriding any Configuration Ready indicator from hardware, whether increasing or decreasing required delays. This includes ignoring FRS and DRS notifications where overridden by this _DSM function, as well as ignoring values specified in the Readiness Time Reporting extended capability structure, if present.

IMPLEMENTATION NOTE

It is recommended that the ASL constant of *Ones* be used for any element in the returned package where overriding the default value is not desired.

For systems where FRS and DRS messages are supported and enabled, it is expected that FRS and DRS Events will only be overridden by this _DSM function in cases where these Events are determined to be problematic or sub-optimal by the system vendor.

Location:
This object can be placed within the scope of any PCI device object, including RCIE scope, PCI devices integrated into the Root Complex, and all other PCI Express functions. The delay values returned by this function are only applicable to the device object. Downstream devices (if any) will retain the default delay requirement according to *PCI Express Base Specification* and/or applicable form factor specification.

Arguments:

Arg0: UUID: E5C937D0-3553-4d7a-9117-EA4D19C3434D

Arg1: Revision ID: lowest valid Revision ID value: 3

Arg2: Function Index: 9

Arg3: Empty Package

Return:
A Package of five integers covering the various device related delay in the PCIe Base Specification:

Integer 0: **FW Reset Time**-- Number of microseconds that the operating system should wait after a Conventional Reset, before issuing the first Configuration Access. For downstream ports supporting the Data Link Layer Active Reporting capability, this entry should not be used. This value may also be specified for PCI devices integrated into the Root Complex.

Integer 1: **FW DL_Up Time**-- Number of microseconds that the operating system should wait after the Downstream Port above this device reports Data Link Layer Link Active, before issuing the first Configuration Access. This entry should only be used for downstream ports supporting the Data Link Layer Active Reporting Capability.

Integer 2: **FW FLR Reset Time**-- Number of microseconds that the operating system should wait after an FLR, before issuing the first Configuration Access.

Integer 3: **FW D3hot to D0 Time**-- Number of microseconds that the operating system should wait on D3hot-to-D0 transition, before issuing the first request (Configuration Access or any other Request). This value may also be specified for PCI devices integrated into the Root Complex.

Integer 4: **FW VF Enable Time**-- Number of microseconds that the operating system should wait after setting the VF Enable bit, before issuing the first request (Configuration Access or any other Request) to respective VFs.

4.6.10. _DSM for Requesting $D3_{cold}$ Aux Power Limit

The Request $D3_{cold}$ Aux Power Limit function describes how a device driver conveys its auxiliary power requirements to the host platform when the device is in $D3_{cold}$. The system firmware responds with a value indicating whether the request can be supported. The power request is specific to the Auxiliary power supply; main power[8] may be removed while in $D3_{cold}$. A device must not draw any more power than what has been negotiated via this mechanism after entering $D3_{cold}$.

A device driver may invoke this function multiple times, either to determine the maximum available power, to retry a request that was temporarily rejected, or to modify (raise or lower) the amount of required power.

For a Multi-Function Device, the driver for Function 0 is required to report an aggregate power requirement covering all functions contained within the device.

Earlier power limits are superseded when this function grants a request. In all other cases, this function has no effect on earlier power limits. This includes grants from earlier calls of this function as well as requirements for the D3 PM state in the form factor and *PCI Express Base Specification*.

Location:

Support for the Request $D3_{cold}$ Aux Power Limit function is only applicable for a _DSM Control Method within the scope of a PCI Express Downstream Port.

For bus hierarchies where multiple Functions reside beneath the PCI Express Downstream Port supporting this _DSM function, system software is responsible for tracking and aggregating requests from child devices and requesting the sum of the requested power limits.

Arguments:

Arg0: UUID: E5C937D0-3553-4d7a-9117-EA4D19C3434D

Arg1: Revision ID: lowest valid Revision ID value: 4

Arg2: Function Index: 0Ah

Arg3: Integer. The value of the integer is the amount of power requested, in mW, for the auxiliary power supply. The minimum allowed value is 0; the maximum allowed value varies by form factor. A value of 8000 0000h signifies that the hierarchy connected via the slot cannot support main power removal when in D3cold while the system is in S0.

[8] This term “main power” matches the usage in the *PCI Express Card Electromechanical Specification*, for other form factors, a mapping to proper terminology may be needed.

Return:
Integer

0h - Denied. Indicates that the platform cannot support the power requested. Software may retry the request with either a lower power request or require no power removal. This is also the value returned to signify any error with the request.

1h - Granted. Indicates that the device is permitted to draw the requested auxiliary power.

2h - No main power removal. Indicates that the platform will not remove main power from the slot while the system is in S0. This value may be returned even though it was not requested that main power remain on, based on requirements of other devices within the platform or other platform configuration variations. This is the only valid return value when Arg3 is 8000 0000h.

11h to 1Fh - Retry, with interval. Indicates that the platform cannot support the power requested at this time, but that it may be able to in the future. Software should retry the request in the number of seconds corresponding with the lower 4 bits of the return value (1 to 15). Firmware is not permitted to return a value in this range more than once for each _DSM instance (located within the ACPI namespace of a single downstream port DeviceObject), unless there is a subsequent invocation of this function before the previously returned retry interval has expired.

All other values are Reserved.

IMPLEMENTATION NOTE

Platform Firmware Budgeting of Aux Power Availability

Platform firmware must not grant more power than what is available within the system. For example, a board may be designed with four CEM slots (one x16 slot, one x4 slot, and two x1 slots). The board may implement a power delivery circuit capable of supplying 2 W of power for the 3.3 Vaux rail supplying all 4 slots. The 3.3 Vaux pins on each CEM slot can supply 1 W each. Platform firmware may use the retry mechanism to prioritize requests from devices in preferred slots in the following manner:

- Requests from a device in the highest priority slot (e.g., x16) are granted immediately, if available.
- Requests from devices in lower priority slots (e.g., x2, x1) are initially rejected, with a retry interval inversely proportional to the slot priority. For instance, if the x2 slot is higher priority than the x1 slots, so the retry interval for the x2 slot may be 4 seconds, and the x1 slots may be 8 and 10 seconds.
- As requests are granted, the granted values are subtracted from the available budget.
- Retried requests are granted based on the remaining power budget or denied if insufficient power budget is available. Another retry is not permitted.
- When there is insufficient power budget for a request, firmware may choose to keep main power on and return no power removal (2h).

4.6.11. _DSM for Adding PERST# Assertion Delay

The Add PERST# Assertion Delay function is used to convey the requirement for a fixed delay in timing between the time the PME_TO_Ack message is received at the PCI Express Downstream Port that originated the PME_Turn_Off message, and the time the platform asserts PERST# to the slot during the corresponding Endpoint's or PCI Express Upstream Port's transition to $D3_{cold}$ while the system is in an ACPI operational state. This delay is not guaranteed to be applied during the transition to a system sleeping state. Assertion of PERST# may be an indicator that power to the device is about to be removed. There is no guaranteed delay between assertion of PERST# and power removal. If any Function on the device is armed for wake, auxiliary power rails required to supply wake logic will not be turned off.

Host platforms implementing this feature must ensure that the delay is observed by the device that is being transitioned to $D3_{cold}$. Once set, this delay is met on every applicable $D3_{cold}$ transition of the device. This function may be invoked multiple times, allowing the delay value to be changed at any time, so that the new value can be applied for the next applicable $D3_{cold}$ transition of the device.

Location:
Support for the Add PERST# Assertion Delay function is only applicable for a _DSM Control Method within the scope of a PCI Express Downstream Port terminated by any sort of add-in slot/connector.

For bus hierarchies where multiple Functions reside beneath the PCI Express Downstream Port supporting this _DSM function, system software is responsible for tracking and aggregating requests from child devices and requesting the maximum of the requested delay values.

Arguments:

Arg0: UUID: E5C937D0-3553-4d7a-9117-EA4D19C3434D

Arg1: Revision ID: lowest valid Revision ID value: 4

Arg2: Function Index: 0Bh

Arg3: Integer. The value, in microseconds, of the delay needed between the PME_TO_Ack message receipt at the PCI Express Downstream Port that originated the PME_Turn_Off message, and the subsequent assertion of PERST# on the corresponding slot. This delay is injected during a transition to $D3_{cold}$ during the ACPI operational state (S0). It does not apply during transitions from the ACPI operational state to an ACPI sleeping state (S3/S4/S5). The maximum permitted requested delay is 10 ms.

Return:

Integer: If the request succeeded, it returns the value specified in Arg3. If the request failed (e.g., if the request in Arg3 is greater than 10ms), it returns the value specified in Arg3 in the most recent successful request, or 0 if there was no such request.

4.6.12. _DSM for Downstream Port Containment and Hot-Plug Surprise Control

If the System Firmware Intermediary (SFI) capability structure is implemented, then this section assumes that system firmware has ownership of SFI. Use cases where system firmware does not own SFI are outside the scope of this specification.

When firmware controls Downstream Port Containment (DPC), this section describes how the operating system can request the firmware to enable DPC and suppress Hot-Plug Surprise for a given DPC capable root port or a switch port. When the operating system controls DPC, this section describes how the operating system can request the firmware to suppress Hot-Plug Surprise for a given DPC capable root port or a switch port. The request is communicated through the _DSM ACPI method. This _DSM function is mandatory for firmware to implement on a port if the firmware supports the Error Disconnect Recovery mechanism (defined in the *ACPI Specification*) on that port. This _DSM function must be evaluated by the operating system if the operating system supports the Error Disconnect Recovery notification regardless of whether the operating system or firmware owns DPC. This _DSM function, if present, must be evaluated by the Error Disconnect Recovery aware OS at least once after the _OSC handshake. As a result of plug and play operations, this _DSM function may be evaluated multiple times during the OS operation with differing inputs. The OS is not permitted to evaluate _DSM while the port is in DPC triggered state. It is recommended that the OS quiesce all traffic below the port while evaluating this _DSM function.

Note that *PCI Express Base Specification* recommends suppressing the Hot-Plug Surprise bit in the Slot Capabilities register when DPC is enabled and so it is recommended that firmware suppress or enable Hot-Plug Surprise when it enables or disables DPC, respectively. Note that if HPS is suppressed while DPC is disabled, then there is window where HPS is not suppressed and DPC is disabled. A Surprise Down error that happens during this window may be reported as an uncorrectable error even though DPC has not yet been enabled to handle it and therefore the system may crash. It is therefore recommended that firmware never have Hot-Plug Surprise suppressed while DPC is disabled.

Also note that when the firmware hands control to the operating system, it may have already enabled DPC and may have already cleared the Hot-Plug Surprise bit. This model is to handle legacy hardware that does not provide a way to clear the Hot-Plug Surprise bit at runtime. The details of what the firmware does are outside the scope of this specification. However, the operating system behavior remains same. The operating system must still evaluate this _DSM function to enable or disable DPC and follow the flow in the implementation note below titled "DPC Event Handling".

The operating system must evaluate this _DSM function when enabling or disabling DPC regardless of whether the operating system or system firmware owns DPC. If the operating system owns DPC then evaluating this _DSM function lets the system firmware know when the operating system is ready to handle DPC events and gives the system firmware an opportunity to clear the Hot-Plug Surprise bit, if applicable.

Location:
This object can be placed under any object representing a DPC capable PCI Express Root Port or Switch Downstream Port. Firmware is permitted to apply the settings of this _DSM function to any ports downstream from the port where this _DSM function is located that do not include an instance of this _DSM function.

Arguments:

Arg0: UUID: E5C937D0-3553-4d7a-9117-EA4D19C3434D

Arg1: Revision ID: lowest valid Revision ID value: 5 (Refer to the *Downstream Port Containment Related Enhancements ECN* for details on Revision 5 of this function. The current definition of this function was introduced with a Revision ID of 6.)

Arg2: Function Index: 0Ch

Arg3: Downstream Port Containment and Hot-Plug Surprise Control
Type: Package
Description: A package of one integer
Integer 1:

0: Disable DPC. If this _DSM function is implemented, then support for disabling DPC is optional. If disabling DPC is not supported, then the return value from this _DSM function will be set to 1 to indicate DPC is enabled. If disabling DPC is supported, then the following steps are required:

1. First, firmware should set the Hot-Plug Surprise bit if the port is Hot-Plug Surprise capable and the port provides a way to set to the Hot-Plug Surprise bit. If SFI HPS Suppress bit is implemented, then firmware should clear it to 0. Proprietary mechanisms are permitted to be used to set the Hot-Plug Surprise to 1 if the SFI HPS Suppress bit is not implemented. This step is skipped if the port does not provide a way to set the Hot-Plug Surprise bit.
2. If firmware owns DPC:
   - Firmware disables DPC
   Else if the operating system owns DPC:
   - The operating system disables DPC after evaluating this _DSM function.

It is recommended that disabling DPC be performed as soon as possible after enabling Hot-Plug Surprise in order to minimize the window where DPC is enabled but HPS is not suppressed.

1: Enable DPC. If this _DSM function is implemented, then support for enabling DPC is mandatory; however, firmware may choose not to enable DPC in certain cases. If firmware chooses not to enable DPC, then the return value from this _DSM function will be cleared to 0 to indicate DPC is disabled. The following steps are required if firmware chooses to enable DPC:

1. If firmware owns DPC:
   - Firmware enables DPC
   Else if the operating system owns DPC:
   - The operating system enables DPC prior to evaluating this _DSM function.
2. Second, firmware should clear the Hot-Plug Surprise bit if the port is Hot-Plug Surprise capable and the port provides a way to clear to the Hot-Plug Surprise bit. If SFI HPS Suppress bit is implemented, then firmware should set it to 1. Proprietary mechanisms are permitted to be used to clear the Hot-Plug Surprise to 0 if the SFI HPS Suppress bit is not implemented.

It is recommended that suppressing Hot-Plug Surprise be performed as soon as possible after enabling DPC in order to minimize the window where DPC is enabled but HPS is not suppressed.

Return: Downstream Port Containment Status from firmware

Type: Integer

Description:

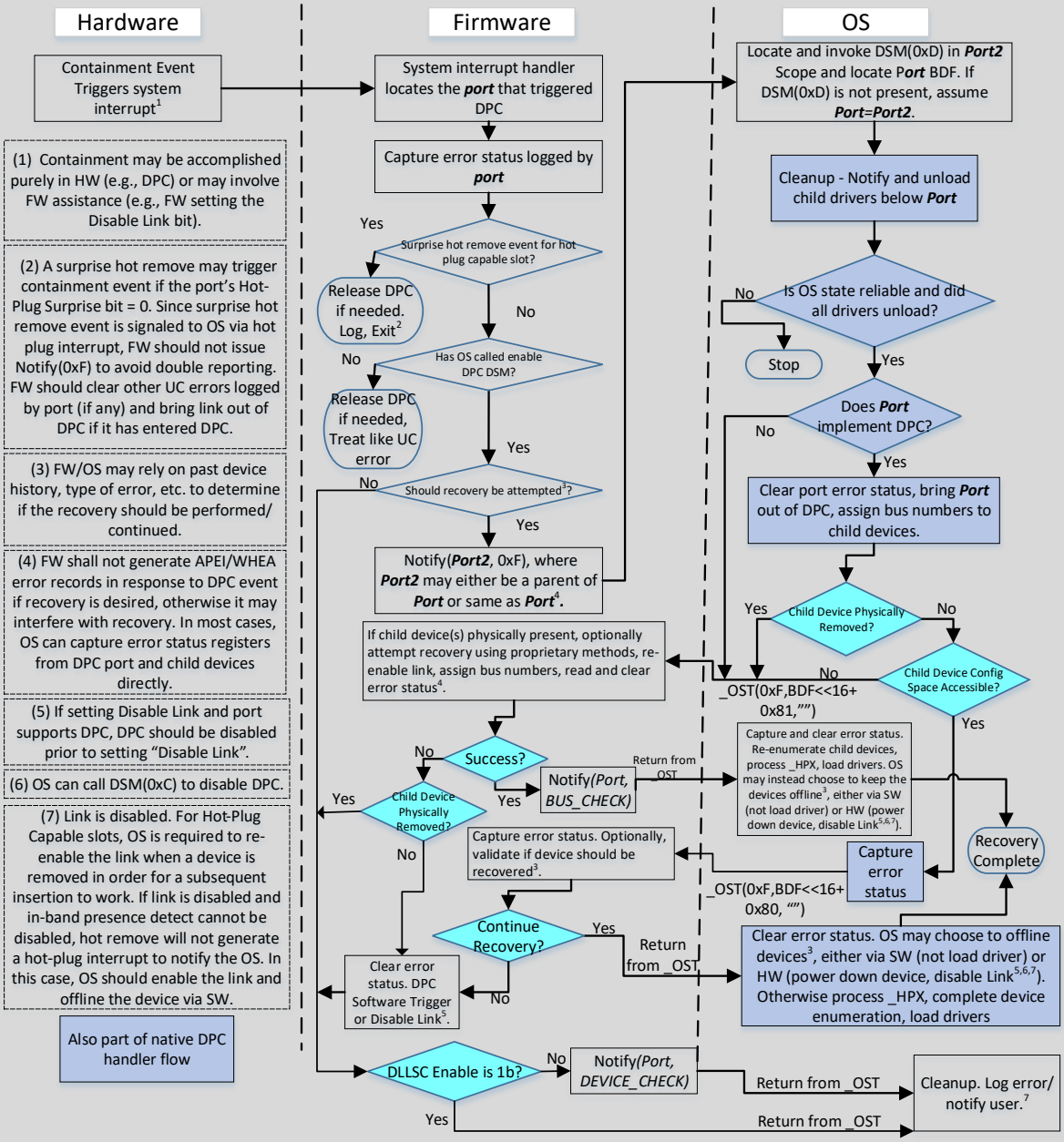
0: The status of the operation is success. DPC is disabled and the Hot-Plug Surprise bit may be set to 1 for hot-plug surprise capable ports.

1: The status of the operation is success. DPC is enabled and the Hot-Plug Surprise bit may be cleared to 0 for hot-plug surprise capable ports.

2: The status of the operation is failure. DPC and Hot-Plug Surprise state are not specified.

IMPLEMENTATION NOTE

DPC Event Handling: The following flow chart documents the behavior when firmware maintains control of AER and DPC and grants control of PCIe Hot-Plug to the operating system.



For devices with persistent errors, a port may be kept in the DPC triggered state (disabled) to keep those devices from continuing to generate errors. For hot-plug slots, the errant device may be removed and replaced with a new device. If the DPC trigger state is not cleared, then the port above the newly inserted device will still be disabled and will be non-operational. Therefore, operating systems may need to modify their hot-plug interrupt handling code to clear DPC Trigger Status when a device is removed so that a subsequent insertion will succeed. Likewise, if the Disable Link bit or Secondary Bus Reset bit is set then the operating system is required to clear it when a device is removed in order for a subsequent device insertion to work. Refer to note 7 in the flow chart.

Also note that in this flow chart, certain events (such as DPC trigger) may generate an interrupt to the firmware (e.g., an ERR_COR message may be sent when DPC is triggered which may be used to generate a SMI to invoke firmware) and other events may generate interrupts to the operating system (e.g., MSI or MSI-X for Data Link Layer State Changed and Presence Detect State Changed hot-plug interrupts).

The *PCI Express Base Specification* requires that the events be signaled in a specific order. For example, on a DPC trigger event the ERR_COR for the DPC event is required to be generated prior to the MSI or MSI-X for the Data Link Layer State Changed event. However, many processors cannot guarantee the order of execution of these events. In some cases, the interrupt handler for firmware may be executed first and in other cases the operating system interrupt handler may be executed first. In other cases, the operating system interrupt handler may partially run, then get interrupted by the firmware which runs to completion, and then returns to the operating system interrupt handler to finish executing.

Sample ASL Code

```
Scope(\_SB){
Device(PCI0) {   //
   Name (_HID, EISAID("PNP0A03"))
   Name (_UID, 0)
   //..
   Method(_OSC,4, Serialized) {
    // Extended to support DPC specific extensions
   }

   Device(RP0) { // Root port 0, supports DPC
     Name (_ADR , 0)  // Device 0, function 0

     OperationRegion (DPC0, PCI_Config, 0x160, 0x20)
     Field (DPC0, AnyAcc, NoLock, Preserve)
     {
       Hdr, 32 // Start of DPC register block
       // define rest of the DPC registers so that FW can access these from ASL
     }

   Method (_DSM, 4, Serialized)
   {
     // check for GUID and revision match
     Switch(Arg2) {
       Case(0xC) {
         // Enable/disable DPC request, replace x with the spec assigned value
         If (LEqual(Arg3, 0)) {
         // Enable HPS if supported/applicable, disable DPC by modifying
         //  registers in DPC0 op region
         }
         If (LEqual(Arg3, 1)) {
         // Enable DPC by modifying registers in DPC0 op region
         }
       } // End Arg2=0xC
       Case (0xD) {
         // return BDF of port that experienced containment
       } // End Arg2=0xD
     } // End Switch
   } // End DSM

   Method(_OST, 3, Serialized) {
     // OSPM calls this method after processing ErrorDisconnectRecover
     //  notification from firmware
     Switch(And(Arg0,0xFF)) // Mask to retain low byte
     {
       Case(0x0F) { // Error Disconnect Recover request
```

```
          Switch(AND(Arg1, 0xFF))  {
            Case (0x80) {// Success
            // Extract BDF of the port with containment event from upper
            //  word of Arg1.
            // Read AER status register from the endpoint below this port.
            } // End Case(0x80)

            Default {
            // IO recovery failed or unrecognized status code
            // Optionally, attempt FW specific recovery.  OK to do nothing
            } // End Default
          } // End Switch
        } // End Case(0xF)
      } // End Switch
    } // End _OST
  } // End RP0
} // End PCI0
} // End \_SB
```

4.6.13. _DSM for Locating the Port that Experienced the Containment Event

Firmware may wish to issue Error Disconnect Recover notification on a port that is parent of the port that experienced the containment event. This can be facilitated by implementing this DSM under the parent port scope. This method, if present, must be evaluated by the DPC aware OS after Error Disconnect Recover notification. This method returns the bus, device and function number of the child port where the containment event occurred. If the parent is a root port, the return value may represent the bus, device and function number of a switch downstream port that entered containment mode. The segment number of the child port is assumed to be identical to the parent port.

If this DSM is not implemented under the port that is the target of Error Disconnect Recover, the OS assumes that the target of the notify event is the port that experienced the containment event.

Location:
This object can be placed under any object representing a DPC capable PCI Express Root Port or Switch Downstream Port. If a port implements this DSM, its child devices cannot instantiate this DSM function.

Arguments:

Arg0: UUID: E5C937D0-3553-4d7a-9117-EA4D19C3434D

Arg1: Revision ID: lowest valid Revision ID value: 5

Arg2: Function Index: 0Dh

Arg3: Empty Package

Return: Location of the port that experienced containment event

Type: Integer

Description:

Port Bus, Device, and Function number encoded as a 16-bit quantity.
(Bits 2:0 = Function, Bits 7:3 = Device, Bits 15:8 = Bus)

Status of Operation as a 1-bit quantity:

Bit 31 = 0: Success

Bit 31 = 1: Failure

All other bits are reserved.

4.6.14. _DSM for Quering Platform Vendor Specific TPH Features

The _DSM for Quering Platform Vendor Specific TPH Features function describes how PCI Express Root Complex support for Vendor Specific decode of Steering Tags is conveyed to the PCI bus driver. PCIe transactions may be recognized and handled specially by the root complex based on the Steering Tag field within a TLP Header for upstream memory TLPs.

Operating System Software uses information returned from this Function to convey to a driver managing an Endpoint to set up the Endpoint hardware to use the correct Steering Tag value as needed. Endpoints describe their ability to use vendor-specific ST values by setting the Device Specific Mode Supported bit in their TPH Requester Capability Register in the TPH Extended Capability Structure.

Location:
Support for this function is only applicable for a _DSM Control Method within the scope of a PCI Express Root Port or a Root Complex Integrated Endpoint.

Arguments:

Arg0: UUID: E5C937D0-3553-4d7a-9117-EA4D19C3434D

Arg1: Revision ID: lowest valid Revision ID value: 6

Arg2: Function Index: 0x0E

Arg3 (bits 15:0): Integer: Vendor ID

Arg3 (bits 31:16): Feature ID: Vendor specific feature being inquired

Return:
Integer
Steering Tag value to be used for the Feature ID inquired by Arg3. If the feature is implemented by the platform, the value must not be Zero.

4.7. _DSM Definitions for PCIe SSD Status LED

_DSM (Device Specific Method) is defined in the *ACPI Specification*. The _DSM method defined in this section is required to be under the ACPI object representing the embedded PCIe device/function or under the ACPI PCIe description representing the add-in PCI Express slot. The following Table 4-11 outlines the required UUID, revision, Arg1, Arg2 parameters, and **Error! Reference source not found.** describes function definitions.

OSPM should use this _DSM when available. If this _DSM is not available, OSPM should use Native PCIe Enclosure Management (NPEM) or SCSI Enclosure Services (SES) instead, if available.

Table 4-11: _DSM Method Input Parameters

Argument	Parameter
Arg0	UUID (set to 5D524D9D-FFF9-4D4B-8CB7-747ED51E194D).
Arg1	Revision ID (set to 1).
Arg2	Function Index as described in **Error! Reference source not found.**.
Arg3	A package containing parameters for the function specified by the UUID, Revision ID, and Function Index. The layout of the package for each Function Index along with the corresponding output is illustrated in the following tables. The input and output packages are a list of bytes (Buffer).

Table 4-12: Function Index

Index	_DSM Function Name
0	Query supported functions (as per 9.1.1 of the *ACPI 6.3 Specification*).
1	Get PCIe SSD Supported LED States.
2	Get PCIe SSD Status LED State.
3	Set PCIe SSD Status LED State.
4 - 0xFFFF	Reserved

4.7.1. _DSM Method Output

Functions 1, 2, and 3 shall return a buffer of length greater than or equal to 4. The first four bytes of the output buffer represent status are structured in Table 4-13.

Table 4-13: Status Field

Field	Byte Length	Byte Offset	Description
General Status Code	2	0	0 – Success 1 – Not Supported 2 – Invalid Input Parameters 3 – Communication Error 4 – Function-Specific Error Code 5 – Vendor-Specific Error Code 6 - 0xFFFF – Reserved
Function-specific Error Code	1	2	This field contains an error code that is specific to the function. This field only contains valid information if General Status Code = Function-Specific Error Code.
Vendor-specific Error Code	1	3	This field contains vendor-specific status codes. It only contains valid information if General Status Code = Vendor-Specific Error Code.
Note: Any non-zero General Status Code indicates that the function failed.			

4.7.2. Query Supported Functions (Function Index 0)

Function 0 of every _DSM is a query function that returns the set of supported function indexes and is always required. This function returns the functions supported by this interface version

Input (Arg3)
None

Return
This function returns an ACPI Buffer containing the byte values {0x0F}

4.7.3. Get PCIe SSD Supported LED States

This function returns the supported LED states that OSPM can control. See Table 4-14 for Output Buffer descriptions.

Input (Arg3)
None

Table 4-14: Output Buffer

Field	Byte Length	Byte Offset	Description
Status	4	0	See Table 4-13
PCIe SSD Status LED State	4	4	Table 4-15 has a detail description of the different LED states: Bit [0] – Reserved Bit [1] – Reserved Bit [2] – OK Bit [3] – Locate Bit [4] – Fail Bit [5] – Rebuild Bit [6] – PFA Bit [7] – Hot Spare Bit [8] – In Critical Array Status Bit [9] – In Failed Array Status Bit [10] – Invalid Device Bit [11] – Disabled Bit [12-31] – Reserved

Bit [2] thru Bit [11] indicate support for different LED states. Table 4-15 provides a detailed description of each state. If the bit is set, then OSPM controls the state.

Table 4-15: PCIe SSD Device Status Bitmap Description

Bit	State	Description
Bit [2]	OK	OK state may mean the drive is functioning normally. This state may implicitly mean that a PCIe SSD is present, powered on, and working normally as seen by the software. A more granular indication of the drive not physically present or present but not powered up are both outside the scope of this specification.
Bit [3]	Locate	Locate state may mean the specific drive is being identified by an OSPM.
Bit [4]	Fail	Fail state may mean the drive is not functioning properly.
Bit [5]	Rebuild	Rebuild state may mean this drive is part of a multi-drive storage volume/array that is rebuilding or reconstructing data from redundancy on to this specific drive.
Bit [6]	PFA	PFA stands for Predicted Failure Analysis. This state may mean the drive is still functioning normally but predicted to fail soon.
Bit [7]	Hot Spare	Hot Spare state may mean this drive is marked to be automatically used as a replacement for a failed drive and contents of the failed drive may be rebuilt on this drive.
Bit [8]	In a Critial Array	In A Critical Array state may mean the drive is part of a multidrive storage array and that array is degraded.
Bit [9]	In a Failed Array	In A Failed Array state may mean the drive is part of a multidrive storage array and that array is failed.
Bit [10]	Invalid Device Type	Invalid Device Type state may mean the drive is not the right type for the connector (e.g., An enclosure supports SAS and NVMe drives and this drive state indicates that a SAS drive is plugged into an NVMe Slot).
Bit [11]	Disabled	Disabled state mean slot is disabled. The power from this slot may be removed.

4.7.4. Get PCIe SSD Status LED States (Function Index 2)

This function returns the PCIe SSD Status LED State. See Table 4-16 for Output Buffer for Function Index 2 descriptions.

Input (Arg3)

None

Table 4-16: Output Buffer for Function Index 2

Field	Byte Length	Byte Offset	Description
Status	4	0	See Table 4-13
PCIe SSD Status LED State	4	4	Table 4-15 has a detail description of each bit: Bit [0] – Reserved Bit [1] – Reserved Bit [2] – OK Bit [3] – Locate Bit [4] – Fail Bit [5] – Rebuild Bit [6] – PFA Bit [7] – Hot Spare Bit [8] – In Critical Array Status Bit [9] – In Failed Array Status Bit [10] – Invalid Device Bit [11] – Disabled Bit [12-31] – Reserved

4.7.5. Set PCIe SSD Status LED States (Function Index 3)

This function returns the PCIe SSD Status LED State. OSPM should call this method after it calls the query method of this _DSM. This function must return success if it is called with no bit set. Table 4-17 lists the Input (Arg3) field description. Table 4-18 lists the Output Buffer for Function Index 3. Table 4-19 lists the Function-specific Error Codes for Function Index 3.

Table 4-17: Input (Arg3)

Field	Byte Length	Byte Offset	Description
PCIe SSD Status LED State	4	0	Table 4-15 has a detail description of each bit: Bit [0] – Reserved Bit [1] – Reserved Bit [2] – OK Bit [3] – Locate Bit [4] – Fail Bit [5] – Rebuild Bit [6] – PFA Bit [7] - Hot Spare Bit [8] –Critical Array Control Bit [9] –Failed Array Control Bit [10] – Invalid Device Bit [11] – Disabled Bit [12-31] – Reserved

Table 4-18: Output Buffer for Function Index 3

Field	Byte Length	Byte Offset	Description
Status	4	0	See Table 4-13
PCIe SSD Status LED State	4	4	Table 4-15 has a detail description of each bit: Bit [0] – Reserved Bit [1] – Reserved Bit [2] – OK Bit [3] – Locate Bit [4] – Fail Bit [5] – Rebuild Bit [6] – PFA Bit [7] - Hot Spare Bit [8] –Critical Array Status Bit [9] –Failed Array Status Bit [10] – Invalid Device Bit [11] – Disabled Bit [12-31] – Reserved

Table 4-19: Function-specific Error Code for Function Index 3

Field	Byte Length	Byte Offset	Description
Function-specific Error Code	1	0	Bit [0] – Not all bits are set. If this bit is set, the platform disregarded some or all of the request state changes. OSPM should check the resulting PCIe SSD Status LED States to see what, if anything, has changed. Bit [1-7] – Reserved

4.8. Generic ACPI PCI Slot Description

To describe a PCI slot, either for manageability purposes or for supporting the ACPI-based PCI hot plug, the ACPI namespace must list eight device objects corresponding to the possible eight functions on the device in the slot. For each of the device object, _ADR is used to identify the PCI address (device number and function number) for each of the functions. Note that each slot can only support one device.

The following is an example name space for a slot:

```
// Device definitions for Slot 1
Device (S1F0) {    // Slot 1, Func#0
    Name(_ADR,0x00020000)
    Name(_SUN,0x00000001)
}
Device (S1F1) {    // Slot 1, Func#1
    Name(_ADR,0x00020001)
    Name(_SUN,0x00000001)
}
Device (S1F2) {    // Slot 1, Func#2
    Name(_ADR,0x00020002)
    Name(_SUN,0x00000001)
}
Device (S1F3) {    // Slot 1, Func#3
    Name(_ADR,0x00020003)
    Name(_SUN,0x00000001)
}
Device (S1F4) {    // Slot 1, Func#4
    Name(_ADR,0x00020004)
    Name(_SUN,0x00000001)
}
Device (S1F5) {    // Slot 1, Func#5
    Name(_ADR,0x00020005)
    Name(_SUN,0x00000001)
}
Device (S1F6) {    // Slot 1, Func#6
    Name(_ADR,0x00020006)
    Name(_SUN,0x00000001)
}
Device (S1F7) {    // Slot 1, Func#7
    Name(_ADR,0x00020007)
    Name(_SUN,0x00000001)
}
```

4.9. The OSHP Control Method

Some systems that include SHPC that are released before ACPI-compliant operating systems with native SHPC support are available, use ACPI firmware to operate the SHPC. Firmware control of the SHPC must be disabled if an operating system with native support is used. Platforms that provide ACPI firmware to operate the SHPC must also provide a control method to transfer control to the operating system. This method is called OSHP (Operating System Hot Plug) and is provided for each SHPC that is controlled by ACPI firmware. Operating systems with native SHPC support must execute the OSHP method, if present, for each SHPC before accessing the SHPC's registers and when returning from a hibernated state. If an SHPC's OSHP method is executed multiple times, and the switch to operating system control has already been achieved, the method must return successfully without doing anything. After the OSHP method is executed, the firmware must not access the SHPC registers. If any signals such as the System Interrupt or PME# have been redirected for servicing by the firmware, they must be restored appropriately for operating system control.

The following is an example of a namespace entry for an SHPC that is managed by firmware:

```
Device(PPB1){
...
Method(OSHP, 0) {
// Disable firmware access to SHPC and restore
// the normal System Interrupt and Wakeup Signal
// connection. (See the Implementation Note below.)
}
...
}
```

IMPLEMENTATION NOTE

Controlling the SHPC Using ACPI

- When using ACPI to control the SHPC, the following should be considered: Firmware should redirect the System Interrupt and the Wakeup signal to a GPE so that ACPI can service the interrupts instead of the operating system. An appropriate _Lxx GPE handler should be provided. When an operating system with native SHPC support executes the OSHP method, the firmware restores the normal System Interrupt and Wakeup signal connection so the interrupts can be serviced by the operating system. In a PCI-to-PCI bridge implementation, access to the SHPC registers is recommended to be done through the DWORD Select/DWORD Data pair in Configuration Space and not the memory Base Address register. This is because versions of ACPI prior to 2.0 do not allow memory access through a Base Address register. In a Host Bridge implementation, the Host Bridge Register Block can be accessed directly by declaring a memory space operation region to encompass it.
- A platform may implement both OSHP and _OSC. However, an operating system that comprehends _OSC will preferentially call _OSC over OSHP.

4.10. Hot Plug Parameters

4.10.1. _HPP

This optional object evaluates to the cache-line size, latency timer, SERR enable, and PERR enable values to be used when configuring a PCI device inserted into a hot-plug slot or for performing

configuration of PCI devices not configured by the BIOS at system boot. The object is placed under a PCI bus where this behavior is desired, such as a bus with hot-plug slots. _HPP provided settings apply to all child buses until another _HPP object is encountered.

4.10.2. _HPX

This optional object provides settings that apply to all child buses until another such object is encountered. The _HPX method supersedes the _HPP method; operating systems will preferentially call _HPX over _HPP if both objects exist.

4.10.3. Device State During Hot Plug

PCI Hot Plug support is done through ACPI mechanism or native support. In the case of using ACPI, _PS3 method brings the device to its D3 state. It is device dependent as to whether this is $D3_{hot}$ or $D3_{cold}$. The *PCI Express Base Specification* defines the context preserved by hardware upon resume from the D3 state, either $D3_{hot}$ or $D3_{cold}$. After the operating system has brought the device out of the D3 state by using the _PS0 method, firmware with knowledge of the PCI Bus Power Management support in the device can determine what context was preserved upon resume from D3.

4.10.4. Slot Power State After Device Removal

PCI Hot Plug support is done through ACPI mechanism or native support. In the case using ACPI, _EJ0 is used to online remove a PCI device. At _EJ0 completion, the PCI device must be isolated for physical removal. _EJ0 is recommended to remove the power from the slot if the platform supports.

5

5.PCI Expansion ROMs

The conventional *PCI Local Bus Specification* provides a mechanism where devices can provide Expansion ROM code that can be executed for device-specific initialization and possibly a system boot function.

This section describes the format, contents, and code entry points for these Expansion ROMs. In addition, this section also describes the services and execution environment provided to those Expansion ROMs by system firmware compliant to this specification, version 3.0 or later.

The information in the Expansion ROMs is laid out to be compatible with existing Intel x86 Expansion ROM headers for ISA add-in cards, but it will also support other machine architectures.

PCI Expansion ROM code is never executed in place. It is always copied from the ROM device to RAM and executed (initialized) from RAM. After initialization the code is moved, by the Expansion ROM, to its final execution location in RAM, assuming the Expansion ROM code is compliant to this specification, version 3.0 or later. This enables dynamic sizing of the code (for initialization and run time) and provides speed improvements when executing run-time code.

5.1. PCI Expansion ROM Contents

PCI device Expansion ROMs may contain code (executable or interpretive) for multiple processor architectures. This may be implemented in a single physical ROM which can contain as many code images as desired for different system and processor architectures (see Figure 5-1). Each image must start on a 512-byte boundary and must contain the PCI Expansion ROM header. The starting point of each image depends on the size of previous images. The last image in a ROM has a special encoding in the header to identify it as the last image.

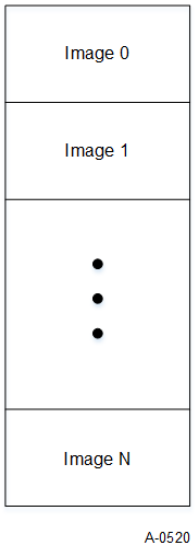


Figure 5-1: PCI Expansion ROM Structure

5.1.1. PCI Expansion ROM Header Format

The information required in each ROM image is split into two different areas. One area, the ROM header, is required to be located at the beginning of the ROM image. The second area, the PCI Data Structure, must be located in the first 64 KiB of the image. The format for the PCI Expansion ROM Header is given below. The offset is a hexadecimal number from the beginning of the image, and the length of each field is given in bytes.

Offset	Length	Value	Description
0h	1	55h	ROM Signature, byte 1
1h	1	AAh	ROM Signature, byte 2
2h-17h	16h	xx	Reserved (processor architecture unique data)
18h-19h	2	xx	Pointer to PCI Data Structure

ROM Signature
The ROM Signature is a two-byte field containing a 55h in the first byte and AAh in the second byte. This signature must be the first two bytes of the ROM address space for each image of the ROM.

Pointer to PCI Data Structure
The Pointer to the PCI Data Structure is a two-byte pointer in little-endian format that points to the PCI Data Structure. The reference point for this pointer is the beginning of the ROM image.

5.1.2. PCI Data Structure Format

The PCI Data Structure must be located within the first 64 KiB of the ROM image and must be DWORD aligned. The PCI Data Structure contains the following information:

Offset	Length	Description
0	4	Signature, the string “PCIR"
4h	2	Vendor Identification
6h	2	Device Identification
8h	2	Device List Pointer
Ah	2	PCI Data Structure Length
Ch	1	PCI Data Structure Revision
Dh	3	Class Code
10h	2	Image Length
12h	2	Revision Level of the Vendor’s ROM
14h	1	Code Type
15h	1	Last Image Indicator
16h	2	Maximum Run-time Image Length
18h	2	Pointer to Configuration Utility Code Header
1Ah	2	Pointer to DMTF CLP Entry Point

Signature
These four bytes provide a unique signature for the PCI Data Structure. The string “PCIR” is the signature with “P” being at offset 0, “C” at offset 1, etc.

Vendor Identification
The Vendor Identification field is a 16-bit field with the same definition as the Vendor Identification field in the Configuration Space for this device.

Device Identification
The Device Identification field is a 16-bit field with the same definition as the Device Identification field in the Configuration Space for this device.

Device List Pointer
The Device List Pointer is a two-byte pointer in little-endian format that points to the list of Device IDs supported by this ROM. The beginning reference point (“offset zero”) for this pointer is the beginning of the PCI Data structure (the first byte of the Signature field). This field is only present in Revision 3.0 (and greater) PCI Data structures. A value of 0000h will be present in this field if the Expansion ROM does not support Device List. A value of 0000h must be present in this field if the Code Type field contains a value of 3 (UEFI).

PCI Data Structure Length
The PCI Data Structure Length is a 16-bit field that defines the length of the data structure from the start of the data structure (the first byte of the Signature field). This field is in little-endian format and is in units of bytes.

PCI Data Structure Revision
The PCI Data Structure Revision field is an eight-bit field that identifies the data structure revision level. The revision level is 3 for the current specification.

Class Code
The Class Code field is a 24-bit field with the same fields and definition as the class code field in the Configuration Space for this device.

Image Length
The Image Length field is a two-byte field that represents the length of the image. This field is in little-endian format, and the value is in units of 512 bytes.

Revision Level
The Revision Level field is a two-byte field that contains the revision level of the Vendor's code in the ROM image.

Code Type
The Code Type field is a one-byte field that identifies the type of code contained in this section of the ROM. The code may be executable binary for a specific processor and system architecture or interpretive code. The following code types are assigned:

Type	Description
0	Intel x86, PC-AT compatible
1	Open Firmware standard for PCI[9]
2	Hewlett-Packard PA RISC
3	Unified Extensible Firmware Interface (UEFI)
4-FFh	Reserved

Last Image Indicator
Bit 7 in this field tells whether or not this is the last image in the ROM. A value of 1 indicates "last image;" a value of 0 indicates that another image follows. Bits 0-6 are reserved.

Maximum Run-Time Image Length
The Image Length field is a two-byte field that represents the maximum length of the image after the initialization code has been executed. This field is in little-endian format, and the value is in units of 512 bytes. This field will be used to determine if the run-time image size is small enough to fit in the memory remaining in the system. This field is only present in Revision 3.0 and later of the PCI Data structure.

---

[9] Open Firmware is a processor architecture and system architecture independent standard for dealing with device specific option ROM code. Documentation for Open Firmware is available in the *IEEE 1275-1994 Standard for Boot (Initialization, Configuration) Firmware Core Requirements and Practices.* A related document, *PCI Bus Binding to IEEE 1275-1994*, specifies the application of Open Firmware to the PCI local bus, including PCI-specific requirements and practices. This document may be obtained using anonymous FTP to the machine *playground.sun.com* with the filename /pub/p1275/bindings/postscript/PCI.ps.

Pointer to Configuration Utility Code Header

This pointer is a two-byte pointer in little-endian format that points to the Expansion ROM's Configuration Utility Code Header table at the beginning of the configuration code block described in Section 5.2.1.24. The beginning reference point ("offset zero") for this pointer is the beginning of the Expansion ROM image. This field is only present in Revision 3.0 (and greater) PCI Data structures. A value of 0000h will be present in this field if the Expansion ROM does not support a Configuration Utility Code Header. A value of 0000h must be present in this field if the Code Type field contains a value of 3 (UEFI).

Pointer to DMTF CLP Entry Point

This pointer is a two-byte pointer in little-endian format that points to the execution entry point for the DMTP CLP code supported by this ROM. The beginning reference point ("offset zero") for this pointer is the beginning of the Expansion ROM image. This field is only present in Revision 3.0 (and greater) PCI Data structures. A value of 0000h will be present in this field if the Expansion ROM does not support a DMTF CLP code entry point as described in Section 5.2.1.25. A value of 0000h must be present in this field if the Code Type field contains a value of 3 (UEFI).

5.1.3. Device List Format

Revision 3.0 (or later) defines a Device List Pointer that points to the list of Device IDs supported by the Expansion ROM image whose Code Type is not 3 (UEFI).

The beginning reference point ("offset zero") for this pointer is the beginning of the PCI Data structure (the first byte of the Signature field). If this field does not exist (i.e., its contents are zero), then the Expansion ROM only supports the one specific Device ID listed in the Device ID field in the PCI Data structure. However, if this field is non-zero then it must point to a Device List Table. The format of this table is shown in Table 5-1.

Table 5-1: Device List Table

Offset	Length	Description
0	2	Device ID 0 supported
2	2	Device ID 1 supported
4	2	Device ID N supported
N	2	Value of 0000h terminates list

5.2. Firmware Power-on Self Test (POST) Firmware

The description of the POST firmware (BIOS) in this section is only pertinent to the management of PCI Expansion ROMs. The POST Firmware goes through several steps to configure a PCI Expansion ROM. These steps are described below:

1. The POST Firmware examines the PCI bus topology to individually examine each PCI device. If the device has implemented an Expansion ROM Base Address register in Configuration Space, then the POST firmware proceeds to the next step.

   Note that individual functions on a multi-function PCI Device may each have an Expansion ROM Base register implementation.

   Note that the order in which PCI Devices are examined and initialized is not defined.

2. The POST firmware enables the Expansion ROM at an unused memory address.
3. The POST firmware checks the first two bytes in the Expansion ROM for the AA55h signature. If that signature is found, then there is a ROM present. Otherwise, no ROM is attached to the device, and the POST firmware proceeds to step 10.
4. If a ROM is attached, the POST firmware must search the ROM for an image that has the proper Code Type. The following steps can be used to search multiple images.
   - If the pointer is valid (non-zero), examine the target location of the pointer. If the pointer is invalid, no further images can be located.
   - The target location must contain the signature string "PCIR". If a valid signature is not present, no further images can be located.
   - If the signature string is valid, examine the Code Type field to determine if it is correct for the expected type of execution environment. (Refer to the Code Type field entry for more information).
   - If the Code Type field is not appropriate, then the BIOS needs to continue searching for Images as described in the next step. If the Code Type field is appropriate, the BIOS proceeds to step 5 below.
   - The BIOS should then examine the "Last Image" field to determine if more images are present. If no further images are present, the BIOS should proceed to step 9.
   - If more images are present, the Image Length field should be examined to determine the starting location of the next image. The Image Length is added to the starting address of the current image (not Image 0). Note that Image Length is in units of 512 bytes.
   - The BIOS proceeds to step 3 above.
5. After the correct image has been selected, POST Firmware will verify that the Vendor ID and Device ID fields match the corresponding fields in the device.

6. If Vendor ID matches but the Device ID does not, the POST Firmware will examine the Device List Pointer. Assuming the Device List Pointer is not zero, the POST firmware will examine the Device List to find a match to the Device ID in the device. A value of 0000h indicates the end of the Device List.

   Note that Expansion ROMs of Code Type 3 (UEFI) do not support the Device List Pointer. For other Code Types, only Expansion ROMs compliant to this specification version 3.0 or later support the Device List Pointer. POST Firmware should not examine the Device List Pointer field until it has confirmed that the PCI Data Structure Revision Level is 3 or greater.

7. After finding the proper image, the POST Firmware determines if there is sufficient space for the image to be copied into RAM in preparation for initialization. This is accomplished by examining the Maximum Run-Time Image Length field.

   Note that only Expansion ROMs compliant to this specification version 3.0 or later support the Maximum Run-Time Image Length field. POST Firmware should not examine the Maximum Run-Time Image Length field until it has confirmed that the PCI Data Structure Revision Level is 3.

   Note that if the image is a PCI 2.1 compliant Expansion ROM, the POST firmware must continue examining the images to determine if an Expansion ROM compliant to this specification version 3.0 or later exists for the same Vendor ID and Device ID. Expansion ROMs compliant to this specification version 2.1 or earlier must only be used by the POST firmware when an Expansion ROM compliant to this specification version 3.0 or later cannot be found.

8. The System firmware must confirm that the Option ROM image has a valid checksum before copying it into RAM. This checksum is calculated using Current Image Size as described in Section 5.2.1.4.

9. The POST firmware then copies the appropriate amount of data into RAM. The amount of data to copy is determined by examining the Image Length field in the PCI Data structure in the header.

10. The POST firmware then disables the Expansion ROM Base Address register.

Subsequent steps will vary depending on the value of the Code Type field.

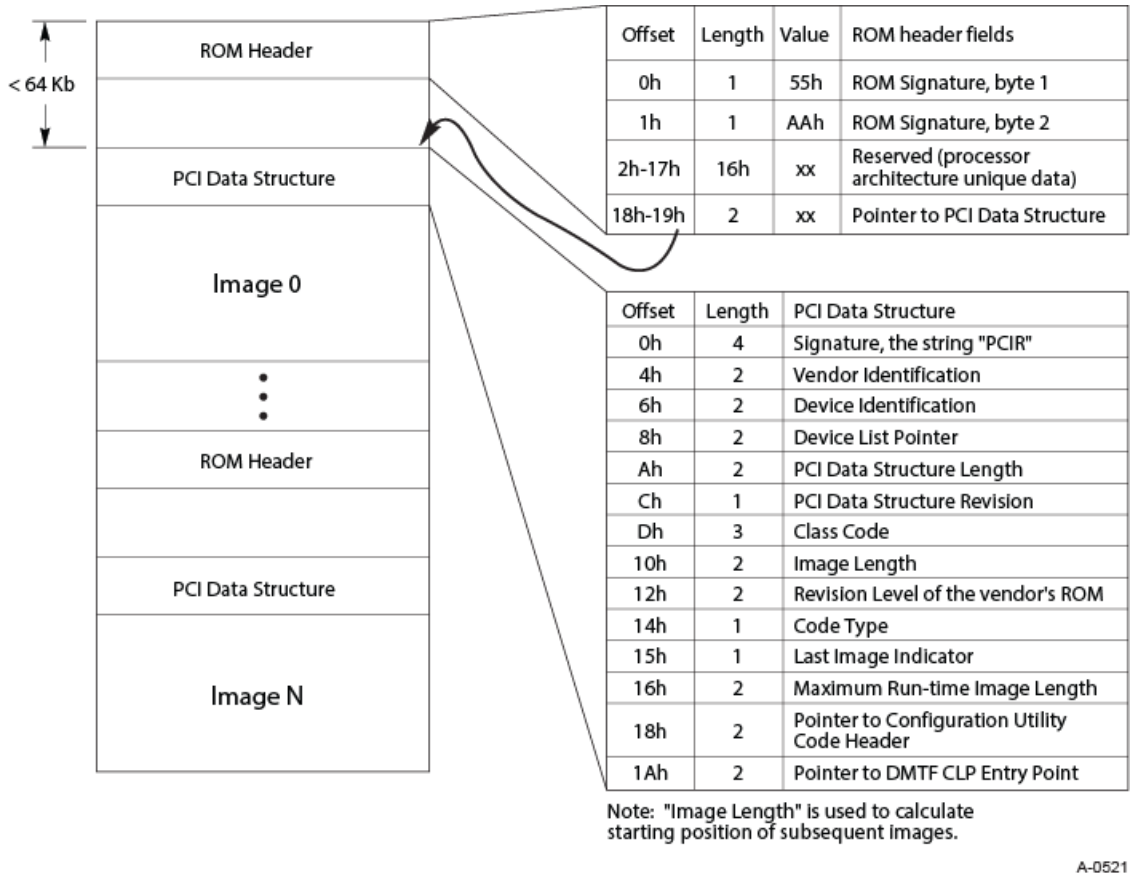


Figure 5-2: Image and Header Organization

5.2.1. PC-compatible Expansion ROMs (Code Type 0)

This section describes the additional steps performed by the POST Firmware when handling of ROM images that have a value of 0 in the Code Type field, indicating the image is Intel x86, PC-compatible.

In addition, this section also describes the new field for the Expansion ROM header for Code Type 0.

5.2.1.1. Expansion ROM Header Extensions

The standard header for PCI Expansion ROM images is expanded slightly for PC-compatibility (Code Type 0). Code Type 0 headers use offset 03h as the entry point for the Expansion ROM INIT function.

Offset	Length	Value	Description
0h	1	55h	ROM Signature byte 1
1h	1	AAh	ROM Signature byte 2
2h	1	xx	Current Image Size in units of 512 bytes
3h	3	xx	Entry point for INIT function. POST does a FAR CALL to this location.
6h-17h	12h	xx	Reserved (application unique data)
18h-19h	2	xx	Pointer to PCI Data Structure

5.2.1.2. POST Firmware Extensions

The POST firmware copies the entire Expansion ROM image into a RAM address, as described above. The RAM address can vary depending on whether the Expansion ROM identifies itself as being compliant to this specification version 3.0 or later. (A value of 03h in the PCI Data Structure Revision field indicates compliance.)

If the Expansion ROM is compliant, then the following steps will be followed by the POST Firmware:

1. POST Firmware will place the Expansion ROM in RAM at an address that may not be the final run-time execution address. This address will be below the 1 MiB address boundary.
2. The POST firmware will then call the INIT function entry point and supply the six arguments described in Table 5-2.
3. As part of the INIT function, the Expansion ROM code must place the final run-time Expansion ROM image into the run-time address provided by the POST firmware.

   Note that the Expansion ROM code must ensure that the device is quiescent until the interrupt service routine is in place at the final run-time location. Caution must be taken in chaining the interrupt service routine. It is possible that an interrupt may arrive while the interrupt service routine is being installed.
4. The POST Firmware will not write-protect the RAM containing the run-time Expansion ROM code at this step. In a system compliant to this specification version 3.0 or later, the write-protection step will be delayed until the Setup portion of the Expansion ROM code, if any, has been executed.
5. The POST firmware will erase the old image from the temporary RAM location.

Table 5-2: Arguments for an Expansion ROM Compliant to this Specification Version 3.0 or Later

Argument Number	Register	Meaning
1	[AH]	Bus number
2	[AL]	Upper 5 bits are the Device number
3	[AL]	The lower 3 bits are the Function number
4	[BX]	Run-time address. This is the final run-time address where the ROM image will be located after Initialization. This address is in units of 16-bytes.
5	[CX]	Segment of the PMM Services Entry Point
6	[DX]	Offset of the PMM Services Entry Point

Table 5-3: Arguments for a Legacy Expansion ROM

Argument Number	Register	Meaning
1	[AH]	Bus number
2	[AL]	Upper 5 bits are the Device number
3	[AL]	The lower 3 bits are the Function number

If the Expansion ROM is not compliant with this specification version 3.0 or later, then the following steps will be followed by the POST Firmware:

1. POST Firmware will place the Expansion ROM in RAM within the legacy compatibility address (usually 0C 0000h to 0E 0000h).
2. The POST firmware will then call the INIT function entry point and supply the three arguments described in Table 5-3.
3. The POST Firmware will then write-protect the RAM containing the run-time Expansion ROM code. The size of the Expansion ROM and amount of RAM to write-protect is determined by examining the Current Image Size field at offset 2 in the ROM header.

The POST firmware should be prepared to handle any mixture of Revision 3.0 (or later) compliant and non-complaint Expansion ROMs in the system.

5.2.1.3. Resizing of Expansion ROMs During INIT

When the Expansion ROM INIT function is called, it is possible for the Expansion ROM code to reduce the amount of space needed for the run-time code and run-time data. This permits the Expansion ROM to occupy the minimal amount of RAM at run-time.

For example, a device Expansion ROM may require 24 KiB for its initialization and run-time code but only 8 KiB for the run-time code. The image in the ROM will show a size of 24 KiB (in the Image Length field in the PCI Data structure), so that the POST firmware copies the whole thing into RAM. Then when the INIT function is running, it can adjust the size down to 8 KiB. This is accomplished by updating the Current Image Size field (at offset 2h) in the ROM Header.

When the INIT function completes, the POST firmware sees that the run-time size is 8 KiB in this example and can copy the next Expansion firmware to the optimum location.

If the INIT function wants to completely remove itself from the Expansion ROM area, it does so by writing a zero to the Initialization Size field (the byte at offset 02h). In this case, no checksum has to be generated (since there is no length to checksum across).

5.2.1.3.1. Calculating a New Checksum at the End of INIT

The INIT function is responsible for guaranteeing that the checksum across the size of the image is correct. If the INIT function modifies the RAM area in any way, a new checksum must be calculated and stored in the image.

The checksum is validated by performing a byte-sum of the Expansion ROM across the entire address. The address range is calculated from the Current Image Size field in the ROM Header. The resulting byte-sum must be zero.

5.2.1.4. Image Structure and Length

A PC-compatible image has three lengths associated with it: a run-time length, an initialization length, and an image length. The image length is the total length of the image, and it must be greater than or equal to the initialization length. The image length is contained in the Image Length field in the PCI Data structure.

Once the image has been copied into RAM by the POST firmware, the Image Length field is no longer referenced by POST firmware and is no longer relevant. It does not need to be adjusted when the Expansion ROM size changes.

The initialization length specifies the amount of the image that contains both the initialization and run-time code. This is the amount of data that POST firmware will copy into RAM before executing the initialization routine. This length is calculated from the Current Image Size at offset 2h in the ROM header. Initialization length must be greater than or equal to run-time length.

The run-time length specifies the amount of the image that contains the run-time code. This is the amount of data the POST firmware will leave in RAM while the system is operating. Again this is calculated from the Current Image Size at offset 2h in the ROM header. The Current Image Size must be updated if the size of the Expansion ROM is changed during the INIT phase.

The PCI Data structure must be contained within the run-time portion of the image (if there is any), otherwise, it must be contained within the initialization portion.

It is critical that the checksum be maintained at all times. When the POST firmware is examining the images in the Expansion ROM the checksum for the Image Length must be zero or the image will not be used. Later when the Expansion ROM is initialized and the Current Image Size is updated, the checksum must again be updated to maintain a valid image.

5.2.1.5. Memory Usage

In *PCI Firmware Specifications* version 2.1 and earlier, the Expansion ROM code had no method for reliably sharing the system resources with the BIOS, particularly memory. PCI Expansion ROMs attempted to allocate memory from the Extended BIOS Data Area (EBDA) or search for erased areas of low memory for temporary use. Many PCI Expansion ROMs fail to implement robust algorithms for sharing memory, and unpredictable behavior may occur.

The Expansion ROM code must make a call to the POST Memory Manager (PMM) to reserve memory, either for temporary usage or for permanent usage. This BIOS function call may only be called during the Expansion ROM initialization phase (INIT phase).

IMPLEMENTATION NOTE

PMM Failure

The Expansion ROM code must be prepared to handle several PMM failure scenarios such as:

- The absence of PMM support in the BIOS
- The absence of PMM "Get Permanent Memory" support in the BIOS
- The PMM denying the request for the memory

In all cases, the 3.0 (or later) compliant Expansion ROM code must be prepared to gracefully handle these conditions.

5.2.1.6. Verification of BIOS Support

The Expansion ROM code must first verify that the system BIOS is compliant with version 3.0 (or later) of this specification. If the BIOS is compliant, the Expansion ROM code can call the PMM function REQUEST_SYSTEM_MEMORY to obtain memory for its own use. The Expansion ROM code also provides an argument to the PMM function to request either permanent memory or temporary memory from the BIOS. Refer to Section 5.2.1.20 for more details on the PMM functions.

Note that a 3.0 (or later) compliant Expansion ROM is not required to have backward compatibility and operate successfully with a BIOS compliant with this specification version 2.1 in a system. It is a design choice whether the 3.0 (or later) Expansion ROM includes support for 3.0 BIOSes only, or both the 3.0 and 2.1 BIOSes.

5.2.1.7. Permanent Memory

Permanent memory will continue to be assigned to the Expansion ROM even after the operating system has booted. This is accomplished by the BIOS updating the data for the Interrupt 15h, E820h function ("Get Memory Map"), to show that this memory allocated to the Expansion ROM is reserved, private, and unmovable.

In addition, the BIOS must support the ACPI memory tables that describe reserved regions of memory. Interrupt 15h, E820h function ("Get Memory Map"), and the ACPI memory tables must both contain entries for the memory reserved by Expansion ROMs.

The operating system must make the "Get Memory Map" call to the BIOS to understand the usage of this reserved area of memory. If the operating system fails to understand that this memory is reserved, then unpredictable behavior may occur.

5.2.1.8. Temporary Memory

If the Expansion ROM initialization code requests temporary memory via PMM, the BIOS will allocate memory that is usable only during the time of Expansion ROM code initialization. This is useful as a buffer for the Expansion ROM code to decompress the ROM for example. When the Expansion ROM initialization code finishes execution, and returns control to the BIOS, the memory will no longer be usable by the Expansion ROM. Any data left in this region may be erased or over-written.

5.2.1.9. Memory Locations

When the Expansion ROM initialization code requests memory, it can specify whether the memory should be located above or below the 1 MiB (F FFFFh) boundary. Memory below the 1 MiB boundary is very limited and Expansion ROMs should be frugal with this valuable system resource.

When the Expansion ROM initialization code requests permanent memory above 1 MiB, it is assumed that the code is capable of accessing this space at a later time when the operating system may be running in protected mode.

Note that the Expansion ROM initialization code will be called from the BIOS in "big real mode,"[10] and the code will have data access to the full 32-bit address range of memory. This will permit access to any memory buffers allocated above the 1 MiB address boundary.

[10] In "Big Real Mode" (aka 32-bit Flat model) is an x86 CPU state where normal Real Mode has been enhanced to have all the segment limits set to 4 GiB. The BIOS will accomplish this mode by entering Protected Mode, setting the segment limits to 4 GiB, and returning to Real Mode (now "Big Real Mode"). In this mode, the processor will honor the segment register limits and they are unaffected by segment register loads. Refer to http://www.phoenix.com/resources/specs-pmm101.pdf for more information.

5.2.1.10. Permanent Memory Size Limits

An Expansion ROM needs to limit the request for permanent allocation to 64 KiB (total) for above 1 MiB memory requests. The limit is 40 KiB (total) for below 1 MiB permanent memory requests. This total upper bound for permanent memory allocation can be tracked by the Firmware (System BIOS). The total amount of permanent memory available has to be shared between multiple Expansion ROMs. Before requesting memory, an Expansion ROM can query using PMM function 0 (pmmAllocate) call with the length "0" and appropriate flags to obtain the maximum permanent memory size available for allocation.

5.2.1.11. Multiple Requests for Memory

Expansion ROM initialization code may make multiple calls to the BIOS to request memory. Expansion ROM initialization code may request temporary memory or permanent memory any number of times. However, the total amount of permanent memory requested must not exceed the limits described in Section 5.2.1.10.

The Expansion ROM code must always check the status code being returned to determine if the requested memory was actually allocated to the Expansion ROM.

5.2.1.12. Protected Mode

If the Expansion ROM initialization code switches into Protected Mode, it must return control to the BIOS in Big Real Mode (a 32-bit flat model). In addition many BIOS services are not available in Protected Mode during Expansion ROM initialization. Refer to the 32-bit BIOS Service Directory (Section 2.3) for further details.

5.2.1.13. Run-Time Expansion ROM Size

The conventional *PCI Local Bus Specification* requires that the Expansion ROM code resize itself after initialization to reduce its use of the memory space. However it is not possible for the BIOS to know, prior to running the Expansion ROM initialization code, if the Expansion ROM will fit into the space remaining for Expansion ROM placement. An Expansion ROM compliant to this specification version 3.0 or later will have a new field in the PCI Data Structure header (Maximum Run-Time Image Length) to indicate the run-time space occupied by the Expansion ROM.

It may not be possible for the Expansion ROM code to know the actual run-time size (which is affected by the presence or absence of other PCI devices in the system). This run-time size field describes the maximum run-time size that the Expansion ROM could occupy after the initialization code has been run.

The BIOS firmware will examine this field prior to calling the Expansion ROM initialization code. If there is insufficient space remaining to place the run-time Expansion ROM code, the BIOS may elect to not initialize the Expansion ROM.

5.2.1.14. Relocation of Expansion ROM Run-time Code

Prior revisions of the *PCI Specification* described the three arguments that are passed to the Expansion ROM initialization code. An Expansion ROM compliant to this specification version 3.0 or later will accept a fourth argument which describes the final run-time address of the Expansion ROM code.

Expansion ROM code compliant to this specification version 3.0 or later must call the BIOS Function (Section 2.5.2) to determine if the BIOS is compliant to this specification version 3.0 or later and will provide this fourth argument.

This new argument permits the BIOS firmware to execute the Expansion ROM initialization code in a memory space other than its final run-time location. This is useful when there is insufficient space remaining in the compatibility region for the BIOS to copy the Expansion ROM initialization code, but there is sufficient space remaining for the run-time code to be placed correctly. The BIOS may initialize the Expansion ROM code in a region other than where the final run-time code will reside. In all cases, this region will be below the 1 MiB address boundary.

A 3.0 (or later) compliant BIOS will not provide a run-time address that partially overlaps the temporary INIT address. The two addresses (if different) will be entirely separate ranges of memory. However a 3.0 (or later) compliant BIOS may provide the same address for run-time as for the INIT address, similar to how a 2.1 compliant BIOS functions now. In either case, the Expansion ROM can examine the CS register and the fourth argument (run-time address) to determine if they are the same region.

Prior to returning control to the BIOS, the Expansion ROM code moves the run-time code, as described in the earlier section of the document. If the Expansion ROM code updates interrupt vectors to point to the run-time Expansion ROM code, those interrupt vectors must point to the final run-time location, not the temporary location where the BIOS may have temporarily placed the Expansion ROM code for initialization.

For example, if there are 16 KiB of available space remaining to place Expansion ROMs, and a 64 KiB Expansion ROM indicates (via the run-time size field) that its run-time size is 16 KiB then there is sufficient space for the run-time code to fit. However, the 64 KiB Expansion ROM initialization code must be executed first, but the 64 KiB of code cannot be placed in memory (there is only 16 KiB remaining), so it will fail to be initialized. However an Expansion ROM compliant to this specification version 3.0 or later can be initialized in another location and the run-time code then moved by the Expansion ROM code into its final address.

5.2.1.15. Expansion ROM Placement Address

Prior versions of the *PCI Specifications* have described the address where Expansion ROMs will be placed as typically being from 0C 0000h up to 0E 0000h. This version of the specification now expands that region to be from A 0000h to F FFFFh inclusively. System firmware compliant to this specification version 3.0 or later will place the Expansion ROM code that is compliant to this specification version 3.0 or later at any aligned address within this expanded range.

If an Expansion ROM not identified as being 3.0 (or later) compliant, the system firmware will place the Expansion ROM within the legacy address range (0C 0000h up to 0E 0000h).

The system firmware must write-protect the region where the PCI Expansion ROM run-time code resides in the new range. System firmware must have a clear understanding of the systems capability to write-protect the various regions of RAM. The system firmware must write-protect the region where the PCI Expansion ROM run-time code resides in the new range. It is possible that a 3.0 (or later) compliant Expansion ROM cannot be placed in the A 0000h address region; for example, because the system does not support write-protecting this region.

The region from A 0000h-B FFFFh is also the legacy frame buffer location for VGA device. The system firmware should exclude this range from the available Expansion ROMs space as long as VGA device is supported. For the system without supporting VGA device, the system firmware is responsible for determining if there are unacceptable risks in placing the Expansion ROMs in the region from A 0000h-B FFFFh.

5.2.1.16. VGA Expansion ROM

Prior versions of this specification have stated that VGA Expansion ROMs must be initialized into the 0C 0000h address space. That requirement is lifted in this version of the specification. A 3.0 (or later) compliant VGA Expansion ROM can be initialized at any location in the Expansion ROM placement address range.

Diagnostic firmware and POST firmware should not assume that if a valid PCI Expansion ROM exists at the 0C 0000h address that it is a VGA Expansion ROM. Firmware should look at the Class Code field in the PCI Data structure to determine if the Expansion ROM belongs to a PCI VGA device.

The 3.0 (or later) compliant BIOS may elect to always place the VGA Expansion ROM at the legacy 0C 0000h address. The BIOS is responsible for determining if there are unacceptable risks to placing the VGA Expansion ROM at non-legacy addresses.

5.2.1.17. Expansion ROM Placement Alignment

Traditional ISA legacy Expansion ROMs were placed in memory at addresses aligned on a 2-KiB boundary. This tradition has carried through into PCI and it results in an average 1-KiB gap between Expansion ROMs.

3.0 (or later) compliant Expansion ROMs may be placed at addresses evenly aligned on a 512-byte boundary. However, the BIOS must be aware of legacy operating systems and legacy applications that will continue to search for Expansion ROMs on the traditional 2-KiB boundaries. The BIOS is responsible for determining if there are unacceptable risks in placing the Option ROMs at non-legacy addresses.

5.2.1.18. BIOS Boot Specification

All 3.0 (or later) compliant Expansion ROMs must support the *BIOS Boot Specification* found at http://www.acpica.org/sites/acpica/files/specsbbs101.pdf. However this specification supersedes the *BIOS Boot Specification* for the definitions of the Expansion ROM placement address, the VGA Expansion ROM placement, and the Expansion ROM placement alignment.

5.2.1.19. Extended BIOS Data Area (EBDA) Usage

The Extended BIOS Data Area (EBDA) is a memory buffer allocated in the below 1 MiB address region of memory. Historically, this region has been used by Expansion ROM code to create a permanent memory buffer for its own use. It is important that the EBDA memory buffer be used consistently and carefully as it is a shared system resource.

If an Expansion ROM requires memory below the 1 MiB boundary, it should use the PMM functions to obtain this memory. The use of EBDA is discouraged due to the complexity of managing and accessing the EBDA memory.

If an Expansion ROM determines that the PMM services are not available, then it is permissible to use EBDA with the following guidelines in mind:

1. First the Expansion ROM must determine if the EBDA is present.
   Is 40:0Eh non-zero? Yes, then it is present. Go to step 2.
   Allocate an EBDA.
   Read 40:13h size and subtract 1 KiB.
   Convert new size to segment and save in 40:0Eh.
   Set 40:0E offset 0 to 1 (1 KiB).
2. Make sure EBDA can handle the amount of memory needed.
   Obtain current EBDA size in 1 KiB.
   Add needed amount of memory (rounded up to nearest KiB).
   If the total memory is greater than 64 KiB, then an error has occurred. The memory request is too large.
3. Move existing EBDA down in memory.
   Disable Interrupts when moving EBDA
   Save current EBDA size as size of memory to move.
   Move current EBDA down to lower memory.
   Zero out old "new" region in upper memory.
   Adjust the 40:0Eh pointer to the new EBDA.
   Set 40:0Eh offset 0 to new size.
   Subtract the size allocated (in KiB) from the value in 40:13h.

After following the above steps, the Expansion ROM may use the new memory (at the highest offset) in the EBDA buffer.

All code references to the EBDA buffer must be indirect references through the pointer at 40:0Eh. The Expansion ROM code must always read the pointer at 40:0Eh to determine the beginning segment of the EBDA buffer. This value will change as the EBDA buffer is expanded, while the offset within the buffer will remain constant.

The 8-bit location at 40:13h is the amount of memory installed below 1 MiB. It is the highest usable memory address below 1 MiB. Many BIOSes install the EBDA immediately below the highest usable memory address but not in all cases. The variable at 40:13h is not a reliable indicator of whether an EBDA has been allocated (use 40:0Eh instead), nor is it a reliable indicator of the size of the EBDA (use 40:0Eh offset 0 instead). Expansion ROM code should make no inferences about the EBDA from the value in 40:13h.

5.2.1.20. POST Memory Manager (PMM) Functions

This description of the new PMM functions is based on the existing *PMM 1.01 Specification*.[11]

Section 3.3.3 of that specification describes the input arguments for the various PMM calls. The flags field is now expanded to include a definition for bit 3 as shown in Table 5-4.

Table 5-4: Fields and Values for PMM Functions

Bits	Field	Value	Description
1:0	*Memory Type*	1..3	0 = Invalid 1 = Requesting convention memory block (0 MiB to 1 MiB) 2 = Requesting extended memory block (1 MiB to 4 GiB) 3 = Requesting either conventional or extended memory
2	*Alignment*	0..1	0 = No alignment 1 = Use alignment from length parameter
3	*Attribute*	0...1	0 = Temporary memory use during POST 1 = Permanent memory use
4:15	*Reserved*	0	Reserved for future use

5.2.1.21. Backward Compatibility of Option ROMs

It is intended that 3.0 (or later) definitions of Option ROMs be compatible with the earlier 2.1 definitions. It is possible for a single 3.0 (or later) Option ROM image to function with 3.0 (or later) compliant System Firmware and with 2.1 compliant System Firmware. The 3.0 fields and functions in the Option ROM will be ignored by 2.1 compliant System Firmware. Any 2.1 compliant System Firmware should have no difficulty locating the 3.0 (or later) Option ROM and initializing it in a manner consistent with this specification version 2.1.

It is also possible to have two separate ROM images for the same PCI device: one for 2.1-compliant System Firmware and one for 3.0 (or later) compliant System Firmware. In this case, the 2.1 Option ROM image must appear first in the sequence of images. 3.0 (or later) compliant System Firmware will first search for a 3.0 (or later) Option ROM image and only use the 2.1 Option ROM image if no 3.0 (or later) Option ROM image is found.

[11] The *POST Memory Manager (PMM) Specification, Version 1.01*, can be found at http://www.pcisig.com/specifications/conventional/pci_firmware/.

5.2.1.22. Option ROM and IRET Handling

The Option ROMs must be capable of handling the level triggered PCI interrupt model where an interrupt can be shared between multiple devices. The Option ROMs must implement appropriate interrupt chaining mechanism and pass control to the next ISR (if present). There should be a default System BIOS interrupt handler for all hardware interrupts that executes the IRET and EOI (acknowledge). Typically, most of the System BIOSes already have a default interrupt handler installed for handling spurious interrupts.

Note: The Option ROM always copies the current address (original Interrupt Vector Address) into its local buffer and proceeds to install its own Interrupt Vector address. On an ISR invocation, if an interrupt vector address has been registered by the Option ROM, it is supposed to invoke the original Interrupt Vector address. The Option ROM that is the first to install in the interrupt ISR chain will be the last Option ROM invoked during an ISR execution prior to the control being passed to the default System BIOS handler.

The interrupt can be left masked in the "Interrupt controller" prior to an Option ROM hooking an interrupt. Option ROM can unmask the interrupt once the ISR is installed. The default System BIOS interrupt handler must not mask the interrupt once it has been unmasked by the Option ROM. Option ROMs in the ISR chain should not mask the level triggered shared interrupt in the "Interrupt controller". System BIOS default interrupt handler should be capable of handling regular and spurious interrupts by appropriately issuing an EOI to Master/Slave Interrupt controller if IRQ is greater than or equal to 8.

In order to maintain backward compatibility with current system ROM implementations, the PCI option ROMs should check for 3.0 (or later) compliant system BIOS via the PCI BIOS Present function. If a 3.0 (or later) compliant BIOS is detected, then the option ROM's interrupt handler should chain to the next interrupt handler as stated in this section and not perform an EOI. If a 3.0 (or later) compatible system BIOS is not present, then the option ROM should just EOI and IRET.

5.2.1.23. Stack Size Requirement by Expansion ROM

The system firmware will provide a minimum 4-KiB size of stack to POST Expansion ROM. It is recommended that system firmware, operating system, and applications provide a minimum 4-KiB stack when calling to Expansion ROM at run time.

5.2.1.24. Configuration Code for Expansion ROMs

Certain Expansion ROMs contain configuration code or utility code that may be executed by the end user during Expansion ROM initialization. Since prior versions of this specification did not provide any mechanism for executing this code, the Expansion ROMs code writers implemented a wide variety of methods to invoke this configuration code.

In many cases the Expansion ROM code would place a prompt on the screen, such as "Press F2 to enter the XXXX Configuration utility", and snoop the keystrokes looking for an invocation. This caused issues to arise in many cases. There was no guarantee that the invocation keystroke was not already assigned to another purpose by the system firmware.

Assigning multiple functionality to a single keystroke may cause a portion of the system to become inaccessible during Expansion ROM initialization. In addition, the Expansion ROM must wait for some arbitrary period of time to be certain that the user had not pressed the keystroke. This adds an unnecessary delay to the Expansion ROM initialization code and to the overall system POST time, particularly in the cases where no change in configuration is needed by the end user.

Beginning with this version of the specification, the Expansion ROM vendors can isolate the configuration code and the system firmware will execute the configuration code only when specifically requested by the user. The Expansion ROM header (defined in Section 5.1.2) now contains an optional pointer to the header for the configuration code. During POST when the system firmware is preparing the Expansion ROM for the INIT stage, the system firmware will make a copy of the configuration utility code (if present in the ROM image) and leave it in temporary memory for later execution.

At the completion of the INIT phase, the Expansion ROM will shrink itself to the final run-time size (without the configuration utility code), and the configuration code will remain in temporary memory. Note that the BIOS will not write-protect the run-time Expansion ROM image until the very end of POST. This permits the Expansion ROM Configuration Utility code to update the run-time Expansion ROM code with configuration parameters or any other data collected by the Configuration Utility code. The Expansion ROM run-time size must include the region where configuration parameters are stored, and the size must not increase when the Configuration Utility code updates the run-time code with the configuration parameters. The Configuration Utility code must also update the checksum for the run-time Expansion ROM code to keep the run-time image valid.

System firmware will be responsible for managing the location and execution of the configuration utilities for all the Expansion ROMs. However, the Expansion ROM can control some of the ways that the System firmware treats the configuration utility code. There are three possible scenarios.

- **Legacy:** In this scenario, the Expansion ROM image does not contain a pointer to the configuration utility code header. The Expansion ROM either has no configuration code, or it prefers to run the configuration code during the INIT phase as 2.1 compliant Expansion ROMs do today. The system firmware does not manage the configuration utility code in this scenario.
- **Hybrid:** In this scenario, the Expansion ROM has a valid configuration utility code header. However the Expansion ROM runs the configuration utility code during the INIT phase, as described in the legacy style above. In addition the system firmware has preserved a copy of the configuration utility code and may execute it at a later point in POST. The configuration utility code must be prepared to be executed twice in this scenario.
- **Delayed Execution:** In this scenario, the Expansion ROM has a valid configuration utility code header. Unlike the previous two scenarios, the configuration utility is not executed (by the Expansion ROM) during the INIT phase. The configuration utility code is only executed by the system firmware at a later point in POST.

The preferred method is the "delayed execution" style described above. This permits the system firmware to manage the execution of the Configuration Utility code in a consistent and organized manner.

Support for Configuration Utility Code management is optional for the system firmware. The Expansion ROM must call the "PCI BIOS Present" function, as described in Section 2.5.2, to determine whether the system firmware has implemented the support for Expansion ROM Configuration Code.

5.2.1.24.1. Executing the Expansion ROM Configuration Code

The configuration code remains in temporary memory so long as it is possible that an end user may need to use the configuration code. The system firmware will make a far call into the configuration utility code, and the configuration utility will perform a far return to pass control back to system firmware when it has completed. The configuration code need not preserve any registers, and is expected to update the BH register as defined below. The system firmware will call the configuration code in "Big Real Mode", as defined in the footnote to Section 5.2.1.9. In addition, the Bus/Device/Function number and the far pointer to the PMM entry point will be provided as argument as shown below.

Arg.	Register	Meaning
1	[AH]	Bus number
2	[AL]	Upper 5 bits are the Device number
3	[AL]	Lower 3 bits are the Function number
4	[BL]	System Firmware state: Bit 0 set = BIOS is performing Console Redirection. Bit 1:7 = Reserved.
5	[BH]	Configuration utility status. Cleared to zero by system firmware prior to call.
6	[CX]	Segment of the PMM Services Entry Point.
7	[DX]	Offset of the PMM Services Entry Point.
8	[SI]	Segment of the run-time Expansion ROM code.
9	[DI]	Offset of the run-time Expansion ROM code.

Prior to returning control to the system firmware via a far return, the configuration code will update the BH register as follows:

Bit 0 set = Configuration utility is requesting a system reset
Bits 1 to 7 = Reserved

The system firmware provides the far pointer to the PMM Services Entry point. The PMM Services Entry pointer is described in Section 3.1.1 "PMM Structure", (offset 7) of the *POST Memory Manager Specification, Version 1.01*. By providing the PMM Services Entry pointer, it is no longer necessary for the Expansion ROM code to search for the PMM Structure, validate the PMM signature, and validate the PMM checksum.

The configuration utility code can assume that all basic devices have been initialized and the system firmware is providing input and output services. The configuration utility code must use the system firmware Interrupt services for input and output functions.

System firmware will have a list of all the configuration code from the various Expansion ROMs. System firmware will be responsible for creating a method for the end user to selectively execute the configuration code. This may be done within a specific page of the ROM based Setup program.

However, the exact method for selecting and executing the individual Option ROM configuration code is not defined by this specification.

5.2.1.24.2. Configuration Utility Behavior Under Console Redirection

When the system firmware is running Console Redirection, it will set bit 0 of the BL register prior when calling the configuration utility code. This places additional restrictions on the configuration utility code, as described below:

- The configuration code must not change the video mode at any point during its execution. Attempts to enter any graphical video mode, in particular, are destructive to Console Redirection session. The BIOS will already be in the preferred text mode when it calls the configuration utility code, and the it must not be changed.
- The only permitted input device is keyboard via Int 16h. Mouse and other types of input devices that cannot be easily redirected over the Console Redirection will not function.
- The configuration code must perform all output via Int 10h services. Writing output directly to a video buffer may cause data to not be redirected correctly in the Console Redirection session. Preferrably, the configuration code would use the Int 10h Write-TTY function for screen output and would minimize the use of cursor repositioning, color dependencies, and "screen-scraping" (screen clearing) features, which perform poorly under Console Redirection.

5.2.1.24.3. Configuration Utility Code Header

The configuration utility code must be separated from the standard Expansion ROM code and be in a contiguous, self-contained block of code that can be independently executed by system firmware. The configuration utility code header resides at the very beginning of this code block and has the following format:

Offset	Length	Description
0	1	Configuration Code Length
1	40	Configuration Code Identifier string
41	1	Revision of this header structure

Configuration Code Length — This byte indicates size of the Configuration code in units of 512 bytes.

Configuration Code Identifier string — The Configuration Code Identifier string is a null-terminated ASCII text string giving the name of the vendor's configuration utility. For example, it may contain "ABC SCSI configuration utility". This text will be displayed to the end user by the System Firmware. Note that the system firmware will display the text until a null is encountered or until 40 bytes have been displayed.

Header Revision — This field is the version of the structure of the Configuration Code Header. This version of the header have a revision field value of 1.

The Configuration Code header and associated code is included in the Image Length field for the PCI Data Structure ROM size.

5.2.1.25. DMTF Server Management Command Line Protocol (SM CLP) Support

The Option ROM may optionally provide an entry point that will support device configuration via the DMTF SM CLP standard.[12] This interface will follow an API defined in the *DMTF SM CLP Specification* 1.0.2 Final Standard. This interface is accessed with a FAR CALL in Big Real Mode where the system ROM will pass in a DMTF SM CLP compatible configuration message that will target the device or a child of that device. This interface may be called multiple times in order to completely configure a device during pre-boot. It is the responsibility of the System ROM to perform any discovery, enumeration, and subsequent translation of the SM CLP UFiTs for any given Container in an implementation. The option ROM code should assume that the system firmware will call the entry point in Big Real Mode and with a minimum of 4 KiB for the stack and a least 128 KiB of extended memory available via PMM.

Input parameters are described below:

Table 5-5: Input Arguments for SM CLP Entry Point

Argument Number	Register	Meaning
1	[AH]	Bus number
2	[AL]	Upper 5 bits are the Device number
3	[AL]	The lower 3 bits are the Function number
4	[ES:EDI]	Pointer to NULL-terminated SM CLP Command Line string buffer
5	[DS:ESI]	Pointer to SM CLP Command Response string buffer

AX contains the Bus/Device/Function of the source of the PCI option ROM, same as the PCI option ROM INIT and Configuration entry points described earlier in this specification. This provides an indicator of which PCI device is being targeted to the SM CLP code. If a PCI option ROM's SM CLP entry point supports more than one device as an SM CLP target, the Command Line Protocol string buffer must also be examined for target information, according to the *SM ME Addressing Specification*. ES:EDI points to a Command Line string as defined in the *SM CLP Specification* (see Section 5.1.9, "Input Data"). DS:ESI points to a buffer of at least 4 KiB in size. Strings are always NULL-terminated.

Note: The scope of the target in the SM CLP Command Line string buffer is limited to the Address Space of the PCI device and is not expected to be scoped to the overall containing Computer System. SM CLP implementations in PCI option ROMs are expected to adhere to the *SM ME Addressing Specification* with the exception that Collections, Logical Devices, Computer Systems, and other Managed Elements are allowed to be in the root of the Address Space.

[12] This document references v1.0.2 of the *DMTF SM CLP Specification* which is available from the DMTF group at http://www.dmtf.org/sites/default/files/standards/documents/DSP0214.pdf.

The output parameters are described in :

Table 5-6: Output Arguments for SM CLP Entry Point

Argument Number	Register	Meaning
1	[AH]	If [AL] = 2 (COMMAND_PROCESSING_FAILED) the contents of [AH] are derived from the SM CLP Processing Error Value (refer to *SM CLP Specification* – Table 6: Processing Error Values and Tags). If [AL] = 3 (COMMAND_EXECUTION_FAILED) the contents of [AH] are derived from the SM CLP CIM Status Code Values (refer to *SM CLP Specification* – Table 9: CIM Status Code Values and Descriptions)
2	[AL]	SM CLP Command Status (refer to *SM CLP Specification* – Table 4: Command Status Values and Tags)
3	[EAX]	Bit 31: OEM Code Flag 0 = Execution Code is a~~n~~ SM CLP Probable Cause Value (refer to *SM CLP Specification* Table 11 – Probable Cause Values and Descriptions) 1 = Execution Code is an OEM Specific value
4	[EAX]	Bits 30-16: Execution Code
5	[ES:EDI]	Pointer to NULL-terminated SM CLP Command Line string buffer
6	[DS:ESI]	Pointer to NULL-terminated SM CLP Command Response string buffer

AL contains 0 for a successful command completion status. If AL is returned with a value of 2, a command processing error is signified and AH will contain a code signifying the cause of the processing error. If AL is returned with a value of 3, an execution error has occurred and the upper 16-bits of EAX will contain an execution error code. Bit 31 of EAX specifies whether the Execution Code is an SM CLP defined code or an OEM-specific code. The string buffer pointers (ES:EDI and DS:ESI) and the contents of the SM CLP Command Line string buffer are preserved. The SM CLP Command Response string buffer is filled in by the option ROM in the "keyword=value" format described in the *SM CLP Specification* (refer to Section 5.1.10, "Output Data"), unless otherwise requested via the SM CLP –output option in the Command Line string buffer. Option ROM support for this default "keyword=value" output format is required, while support for other SM CLP output formats is optional (the SM CLP option ROM should return an OPTION NOT SUPPORTED error in AH if the SM CLP – output option requested by the caller is not supported by the option ROM).

All other x86 registers are preserved.

Note that not all 3.0 (or later) system firmware will support calling the DMTF SM CLP entry point. The support for this function can be determined by examining the results of the "PCI BIOS Present" call described in Section 2.5.2. This is useful for diagnostics and validation in determining if the DMTF SM CLP interface will be used. In addition an Expansion ROM may change its configuration behavior if it determines that the 3.0 (or later) system firmware will not call the DMTF SM CLP API.

Similarly an Expansion ROM is not required to have a DMTF SM CLP entry point. If the entry point is not present, the "Pointer to the DMTF SM CLP entry point" field in the PCI Data Structure (see Section 5.2.1) should be null.

The normal power-up sequence concerning SM CLP is as follows:

1. An SM CLP configuration session is started via a remote console.
2. The system ROM facilitates the configuration of the PCI device(s) by passing SM CLP commands via the SM CLP Entry Point interface.
3. The Expansion ROM SM CLP code stores the configuration settings in a non-volatile location associated with the targeted PCI device so that the configuration information is available to the PCI device's option ROM Init code on current and subsequent boots.
4. The system ROM initializes the PCI device(s) by calling the PCI Expansion ROM Init entry point.

Note: SM CLP configuration sessions that occur after the PCI Expansion ROM Init code has executed may require the system to be rebooted so that the new configuration changes can take effect.

5.2.2. UEFI Expansion ROM (Type 3)

Section 13.4.2 "PCI Option ROMs" of the *Unified Extensible Firmware Interface Specification, Version* 2.4 (http://www.uefi.org) describes ways to store UEFI images (e.g., UEFI drivers) in PCI Expansion ROMs. The description in Section 5.2.1 of this document does not apply to UEFI Expansion ROMs.