



PCI-SIG ENGINEERING CHANGE REQUEST

TITLE:	ACPI additions for FW latency optimizations
DATE:	January 28, 2014
AFFECTED DOCUMENT:	PCI Firmware Spec 3.1
SPONSOR:	Anton Cheng, Rob Gough, Mahesh Wagh, Intel

Part I

1. Summary of the Functional Changes

In the PCIe base specification, there are several specified delays:

1. Initialization delay after Conventional Reset (section 6.6.1 in PCI Express Base Specification Rev 3.0), 100 ms.
2. Recovery time following D3_{hot}-to-D0 transition (section 5.3.1.4 Device Power Management States in PCI Express Base Specification Rev 3.0), 10 ms.
3. Delay after Function-Level Reset (FLR) (section 6.6.2 in PCI Express Base Specification 3.0), 100 ms.
4. Data Link Up (DL_Up) delay for downstream ports supporting the Data Link Layer Active Reporting Capability (section 6.6.1 in PCI Express Base Specification 3.0), 100 ms.
5. The Reset Time described in the Readiness Time Reporting extended capability structure is also used as the VF Enable time for PFs supporting VFs (section 3.3.3.1, <with RN ECR>).

This ECR provides two additional ACPI DSM functions to inform OS about the possible time reduction opportunities:

1. Hardware or system firmware has already met the post Conventional Reset delay requirement, there is no need for OS to provide any reset delay unless a new Conventional Reset is issued while OS is in control of the PCIe subsystem.
2. A PCIe device within a closed system has lower delays than required by the PCIe spec. The PCIe specification-provides similar latency reduction (Readiness Notification) information directly via in band protocol and Extended Capability Structure, while this ECR can be implemented on existing systems without hardware impact.

2. Benefits as a Result of the Changes

With an OS PCIe stack that support this ECR, closed system integrators with this ECR will be able to take advantage of the timing reduction in device and silicon and translate them to faster overall device startup and resume for end-user, allowing PCIe devices to be competitive against competitions (devices with different bus interface) on the instant-on and fast boot latency metrics.

3. Assessment of the Impact

System software- both platform firmware and OS will need to incorporate support for these changes in order for the features referenced above to function optimally.

Request Request Request Request Request Request Request Request

Analysis of the Hardware Implications

No hardware implications. Closed system integrators might want to request device vendors to specify lower post Conventional Reset and D3_{hot}-to-D0 transition recovery latencies.

4. Analysis of the Software Implications

Each of the proposed changes will require modified ACPI BIOS firmware to support the corresponding features, if implemented on the platform. OS-level bus driver and OSPM modifications are needed in order to take advantage of reduction in latency.

5. Analysis of the C&I Test Implications

If the ACPI DSM functions are declared incorrectly, it can lead to device initialization and recovery failure. The shortened delays are visible to debugger via logic analyzer as well as OS PCIe stack instrumentations. It is up to the system integrator to ensure that values reported in _DSM function 9 are correct.

Request Request Request Request Request Request Request Request

Part II

Detailed Description of the change

Adding Section 4.6.8 and 4.6.9:

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 devices 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: 3

Arg2: Function Index: 8

Arg3: Empty Package

Return:

Request Request Request Request Request Request Request Request

An integer whose description is as follows:

Type: Integer

Purpose: Allow OS 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 D3_{hot}-to-D0 transition (section 5.3.1 Device Power Management States).
3. For a function that has a shorter FLR (Functional Level Reset) latency (section 6.6.2 Function-Level Reset).
4. For a device that has a shorter DL_Up (Data Link Up) latency (section 6.6.1).
5. For a device (PF) supporting one or more VFs with a shorter VF enable latency (section 3.3.3.1).

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 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 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.

Request Request Request Request Request Request Request Request

<**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.>

<**Implementation note:** 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: 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 (Functional Level Reset), before issuing the first Configuration Access.

Integer 3: **FW D3_{hot} to D0 Time**-- Number of microseconds that the operating system should wait on D3_{hot}-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.