PCI-SIG ENGINEERING CHANGE NOTICE

TITLE:	Optimized Buffer Flush/Fill
DATE:	Updated 30 April 2009, original request: 8 February 2008
AFFECTED DOCUMENTS:	PCI Express Base Specification, Revision 2.0 PCI Express Base Specification, Revision 2.1 PCI Express CEM Specification, Revision 2.0 PCI Express Mini CEM Specification, Revision 1.2
SPONSORS:	Intel Corporation

Part I

5   **1.   Summary of the Functional Changes**

Currently, asynchronous device activity prevents optimal power management of memory, CPU, and other Root Complex (RC) internals because device activity will tend to be misaligned with respect to other devices and with respect to the natural activity of the system, such as servicing the O/S timer tick.

This ECR proposes to add a new mechanism for platform central resource (RC) power state information
10   to be communicated to Devices. This mechanism enables Optimized Buffer Flush/Fill (OBFF) by allowing the platform to indicate optimal windows for device bus mastering & interrupt activity. Devices can use internal buffering to shape traffic to fit into these optimal windows, reducing platform power impact.

This OBFF indication is a hint - as a fallback devices are still permitted to initiate bus mastering & interrupt traffic at any time, although this will negatively impact the platform power and should be avoided as much
15   as possible. In such cases, the platform may use OBFF to signal other platform Devices so as to minimize the negative impact.

OBFF events are signaled using the WAKE# signal when this is supported by the platform topology because this prevents needless link reactivation for the (common) case where most devices have no need to perform bus master or interrupts. Also defined is a message for use when WAKE# not available,
20   and a firm/software configuration mechanism to determine which signaling mechanism is used.

This ECR impacts Endpoint devices, RCs and Switches that choose to implement the new optional feature.

2.   Benefits as a Result of the Changes

Using this mechanism, platform power management can be improved by enabling Endpoints to shape
25   their traffic to minimize impact to platform central resource power consumption.

3.   Assessment of the Impact

This is an optional normative capability.

Endpoints implementing the capability must support reception and processing of OBFF indications and may need to make implementation changes to optimize their behaviors.

30   Switches implementing the capability must support passing OBFF messages when configured to do so.

Root Complexes implementing the capability must provide capabilities for issuing OBFF indications at appropriate times.

4. Analysis of the Hardware Implications

OBFF requires new hardware and is an optional normative capability. Hardware that is not OBFF capable will continue to operate as it does today.

AC specifications for the WAKE# signal when used for OBFF are defined.

5 5. Analysis of the Software Implications

OBFF requires new firmware/software to enable this functionality and thus is optional normative. Software must not enable OBFF in an Endpoint unless the platform supports delivering OBFF indications to that Endpoint.

Part II

Detailed Description of the changes

Please note that changes are being requested to three separate specifications as annotated.

PCI Express Base Specification - add in Section 2.2.8:

This document defines the following groups of Messages:

INTx Interrupt Signaling

Power Management

Error Signaling

Locked Transaction Support

Slot Power Limit Support

Vendor-Defined Messages

OBFF Messages

PCI Express Base Specification - add Section 2.2.8.x:

2.2.8.x.   Optimized Buffer Flush/Fill (OBFF) Message

The OBFF Message is optionally used to report platform central resource states to Endpoints. This mechanism is described in detail in <Section 6.x below>.

The following rules apply to the formation of the OBFF Message:

❑   <ref table below> defines the OBFF Message.

❑   The OBFF Message does not include a data payload (TLP Type is Msg).

❑   The Length field is reserved.

❑   The Requester ID must be set to the Transmitting Port's ID

❑   The OBFF Message must use the default Traffic Class designator (TC0). Receivers that implement OBFF support must check for violations of this rule. If a Receiver determines that a TLP violates this rule, it must handle the TLP as a Malformed TLP.

•   This is a reported error associated with the Receiving Port (see Section 6.2)

Table 2.x: OBFF Message

Name	Code[7:0] (b)	Routing r[2:0] (b)	Support[1]				Req ID	Description/Comments
			RC	Ep	Sw	Br		
OBFF	0001 0010	100	t	r	tr		BD	Optimized Buffer Flush/Fill

Note 1: Support for OBFF is optional. Functions that support OBFF must implement the reporting and enable mechanisms described in Chapter 7.
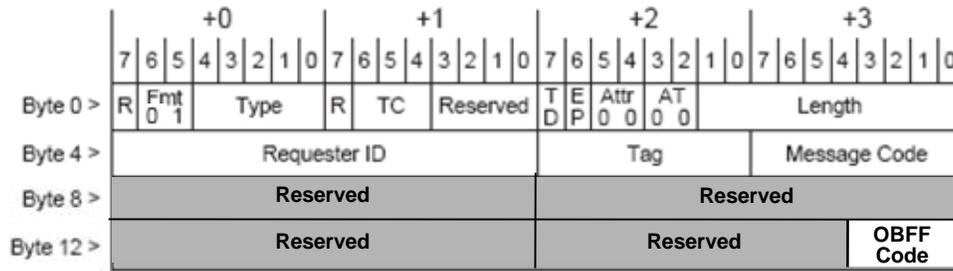
Figure 2.x: OBFF Message

PCI Express Base Specification – edit in Section 5.3.3.2:

…

WAKE# is not intended to be used as an input by any Endpoint, and tThe system is not required to route or buffer WAKE#it in such a way that an Endpoint is guaranteed to be able to detect that the signal has been asserted by another Function.

…

PCI Express Base Specification - add Section 6.x:

6.x   Optimized Buffer Flush/Fill (OBFF) Mechanism

The Optimized Buffer Flush/Fill (OBFF) Mechanism enables a Root Complex to report to Endpoints (throughout a hierarchy) time windows when the incremental platform power cost for Endpoint bus mastering and/or interrupt activity is relatively low. Typically this will correspond to time that the host CPU(s), memory, and other central resources associated with the Root Complex are active to service some other activity, for example the OS timer tick. The nature and determination of such windows is platform/implementation specific.

An OBFF indication is a hint - Functions are still permitted to initiate bus mastering and/or interrupt traffic whenever enabled to do so, although this will not be optimal for platform power and should be avoided as much as possible.

OBFF is indicated using either of the WAKE# signal or a message (see Section <ref 2.2.8.x>). The message is to be used exclusively on interconnects where the WAKE# signal is not available. WAKE# signaling of OBFF or CPU Active must only be initiated by a Root Port when the system is in an operational state, which in an ACPI compliant system corresponds to the S0 state. Functions that are in a non-D0 state must not respond to OBFF or CPU Active signaling.

The OBFF message routing is defined as 100b, for point-to-point, and is only permitted to be transmitted in the Downstream direction. There are multiple OBFF events distinguished. When using the OBFF Message, the OBFF Code field is used to distinguish between different OBFF cases:

1111   "CPU Active" - System fully active for all Device actions including bus mastering and interrupts

0001   "OBFF" – System memory path available for Device memory read/write bus master activities

0000   "Idle" – System in an idle, low power state

All other codes are Reserved.

These codes correspond to various assertion patterns of WAKE# when using WAKE# signaling, as shown in <ref figure 6-x below>. There is one negative-going transition when signaling OBFF and two negative going transitions each time CPU Active is signaled. The electrical parameters required when using WAKE# are defined in the WAKE# Signaling section of *PCI Express CEM Specification, Revision 2.0* (or later).

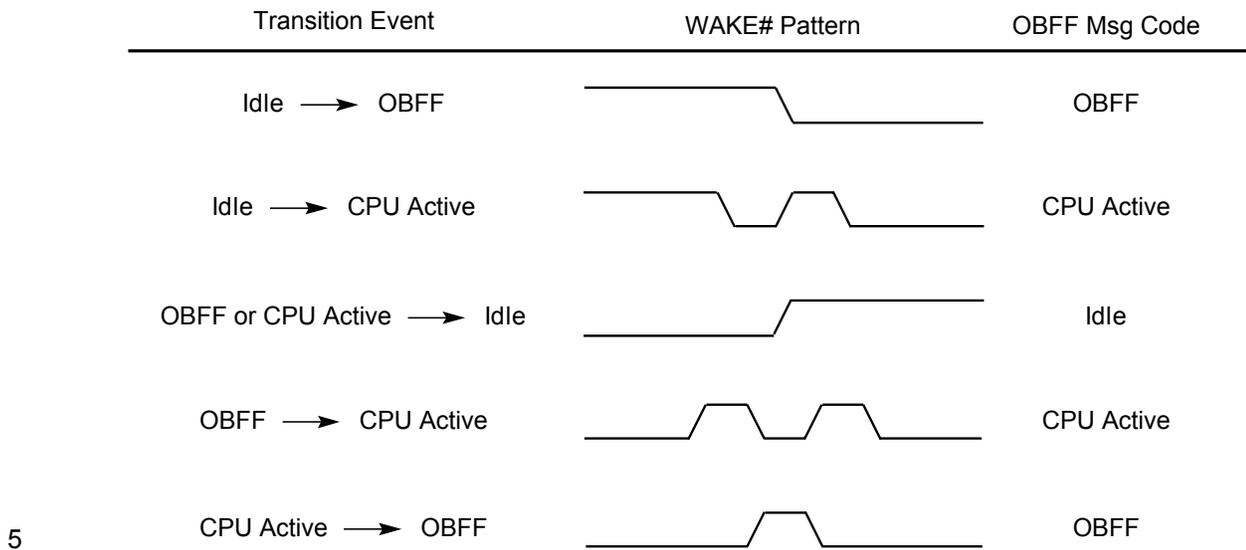| Transition Event	WAKE# Pattern	OBFF Msg Code
Idle → OBFF		OBFF
Idle → CPU Active		CPU Active
OBFF or CPU Active → Idle		Idle
OBFF → CPU Active		CPU Active
CPU Active → OBFF		OBFF

Figure 6-x: OBFF Codes and Equivalent WAKE# Patterns

When an OBFF Message is received that indicates a Reserved code, the Receiver, if OBFF is enabled, must treat the indication as a "CPU Active" indication.

An OBFF Message received at a Port that does not implement OBFF or when OBFF is not enabled must be handled as an Unsupported Request (UR).

OBFF indications reflect central resource power management state transitions, and are signaled using WAKE# when this is supported by the platform topology, or using a Message when WAKE# is not available. OBFF support is discovered and enabled through reporting and control registers described in Chapter 7. Software must not enable OBFF in an Endpoint unless the platform supports delivering OBFF indications to that Endpoint.

When the platform indicates the start of a CPU Active or OBFF window, it is recommended that the platform not return to the Idle state in less than 10µs. It is permitted to indicate a return to Idle in advance of actually entering platform idle, but it is strongly recommended that this only be done to prevent late Endpoint activity from causing an immediate exit from the idle state, and that the advance time be as short as possible.

It is recommended that Endpoints not assume CPU Active or OBFF windows will remain open for any particular length of time.
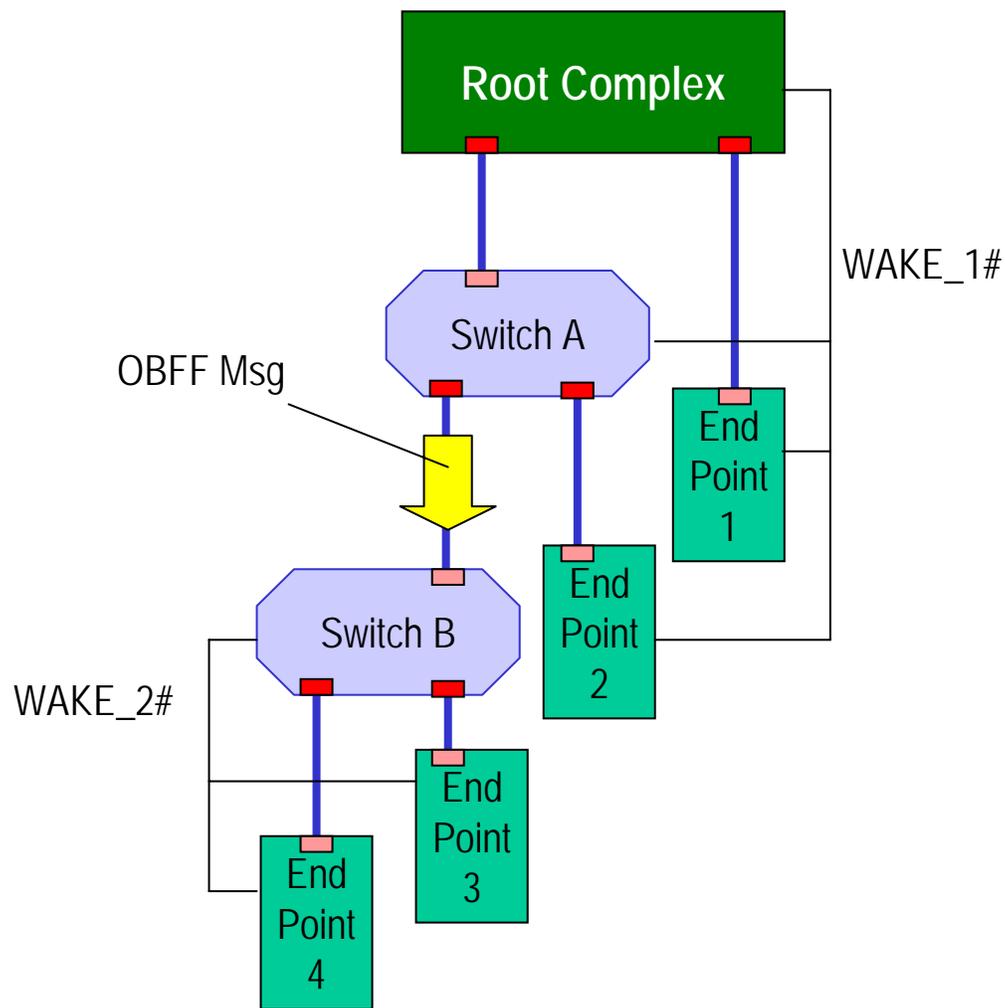
Figure 06-y: Example Platform Topology Showing a Link Where OBFF is carried by Messages

Figure <6-y> shows an example system where it is necessary for a Switch (A) to translate OBFF indications received using WAKE# into OBFF Messages, which in this case are received by another Switch (B) and translated back to using WAKE# signaling. A HwInit configuration mechanism (set by hardware or firmware) is used to identify cases such as shown in this example (where the link between Switch A and Switch B requires the use of OBFF Messages), and system firmware/software must configure OBFF accordingly.

When a Switch is configured to use OBFF Message signaling at its Upstream Port and WAKE# at one or more Downstream Ports, or vice-versa, when enabled for OBFF, the Switch is required to convert all OBFF indications received at the Upstream Port into the appropriate form at the Downstream Port(s).

When using WAKE#, the enable for any specific Root Port enables the global use of WAKE# unless there are multiple WAKE# signals, in which case only the associated WAKE# signals are affected. When using Message signaling for OBFF, the enable for a particular Root Port enables transmission of OBFF messages from that Root Port only. To ensure OBFF is fully enabled in a platform, all Root Ports indicating OBFF support must be enabled for OBFF. It is permitted for system firmware/software to selectively enable OBFF, but such enabling is beyond the scope of this specification.

To minimize power consumption, system firmware/software is strongly recommended to enable Message signaling of OBFF only when WAKE# signaling is not available for a given link.

OBFF signaling using WAKE# must only be reported as supported by all components connected to a Switch if it is a shared WAKE# signal. In these topologies it is permitted for software to enable OBFF for components connected to the Switch even if the Switch itself does not support OBFF.

It is permitted, although not encouraged, to indicate the same OBFF event more than once in succession.

5   When a Switch is propagating OBFF indications Downstream, it is strongly encouraged to propagate all OBFF indications. However, especially when using Messages, it may be necessary for the Switch to discard or collapse OBFF indications. It is permitted to discard and replace an earlier indication of a given type when an indication of the same or a different type is received.

Downstream Ports can be configured to transmit OBFF Messages in two ways, which are referred to as
10 Variation A and Variation B. For Variation A, the Port must transmit the OBFF Message if the Link is in the L0 state, but discard the Message when the Link is in the Tx.L0s or L1 state. This variation is preferred when the Downstream Port leads to Devices that are expected to have communication requirements that are not time-critical, and where Devices are expected to signal a non-urgent need for attention by returning the Link state to L0. For Variation B, the Port must transmit the OBFF Message if the Link is in the L0 state, or, if
15 the Link is in the Tx.L0s or L1 state, it must direct the Link to the L0 state and then transmit the OBFF Message. This variation is preferred when the Downstream Port leads to devices that can benefit from timely notification of the platform state.

When initially configured for OBFF operation, the initial assumed indication must be the CPU Active state, regardless of the logical value of the WAKE# signal, until the first transition is observed.

20 When enabling Ports for OBFF, it is recommended that all Upstream Ports be enabled before Downstream Ports, and Root Ports must be enabled after all other Ports have been enabled. For hot pluggable Ports this sequence will not generally be possible, and it is permissible to enable OBFF using WAKE# to an unconnected hot pluggable Downstream Port. It is recommended that unconnected hot pluggable Downstream Ports not be enabled for OBFF message transmission.
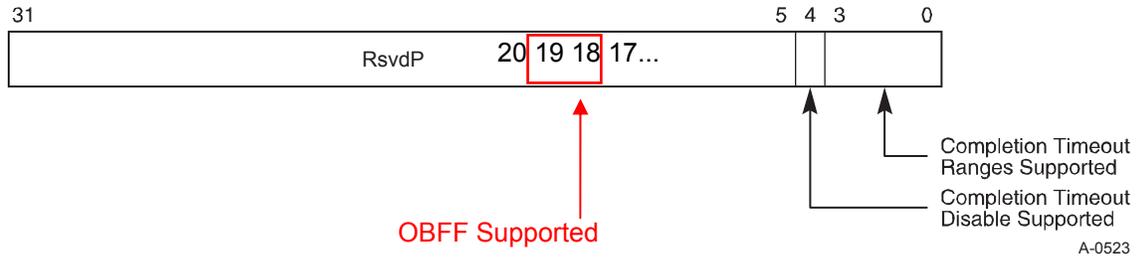

IMPLEMENTATION NOTE

OBFF Considerations for Endpoints

25 It is possible that during normal circumstances, events could legally occur that could cause an Endpoint to misinterpret transitions from an Idle window to a CPU Active window or OBFF window. For example, a non-OBFF Endpoint could assert WAKE# as a wakeup mechanism, masking the system's transitions of the signal. This could cause the Endpoint to behave in a manner that would be less than optimal for power or performance reasons, but should not be unrecoverable for the Endpoint or the host system.

30 In order to allow an Endpoint to maintain the most accurate possible view of the host state, it is recommended that the Endpoint place its internal state tracking logic in the CPU Active state when it receives a request that it determines to be host-initiated, and at any point where the Endpoint has a pending interrupt serviced by host software.

7.8.15. Device Capabilities 2 Register (Offset 24h)

OBFF Supported

Completion Timeout Ranges Supported

Completion Timeout Disable Supported

A-0523

5   ...

Bit Location	Register Description	Attributes
...		

19:18	**OBFF Supported** –
	
	00b – OBFF Not Supported
	01b – OBFF supported using Message signaling only
	10b – OBFF supported using WAKE# signaling only
	11b – OBFF supported using WAKE# and Message signaling
	
	The value reported in this field must indicate support for WAKE# signaling only if:
	  - for a Downstream Port, driving the WAKE# signal for OBFF is supported and the connector or component connected Downstream is known to receive that same WAKE# signal
	  - for an Upstream Port, receiving the WAKE# signal for OBFF is supported and, if the component is on an add-in-card, that the component is connected to the WAKE# signal on the connector.
	Root Ports, Switch Ports and Endpoints are permitted to implement this capability.
	For a multi-Function device associated with an Upstream Port, each Function must report the same value for this field.
	For Bridges and Ports that do not implement this capability, this field must be hardwired to 00b.
		HwInit

PCI Express Base Specification - change Figure 7-25 and Table 7-24 in Section 7.8.16 as follows:
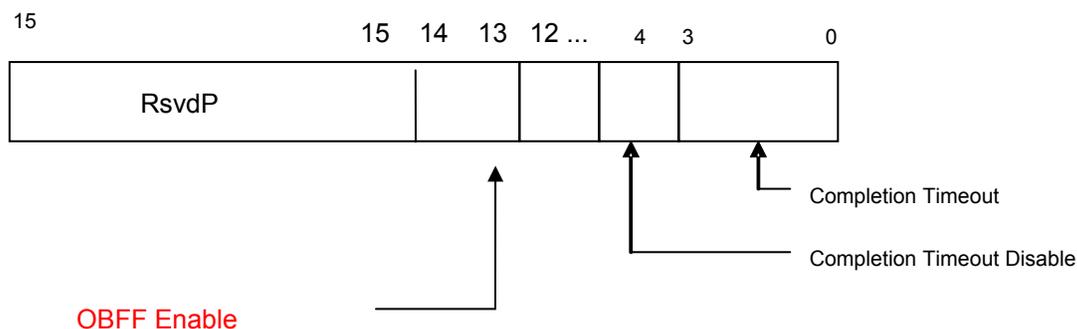
7.8.16.   Device Control 2 Register (Offset 28h)

...

5

Bit Location	Register Description	Attributes
...		
14:13	**OBFF Enable** – 00b – Disabled 01b – Enabled using Message signaling [Variation A] 10b – Enabled using Message signaling [Variation B] 11b – Enabled using WAKE# signaling See Section <6.x!!!> for an explanation of the above encodings. This field is required for all Ports that support the OBFF Capability. For a Multi-Function Device associated with an Upstream Port of a Device that implements OBFF, the field in Function 0 is of type RW, and only Function 0 controls the Component's behavior. In all other Functions of that Device, this field is of type RsvdP. Ports that do not implement OBFF are permitted to hardwire this field to 00b. Default value of this field is 00b.	RW /RsvdP (see description)

PCI Express Base Specification - Add to the Message table in Appendix F as shown:

Message Code	Routing r[2:0]	Type	Description
...			
0001 0010	100	Msg	Optimized Buffer Flush/Fill (OBFF) Message, see Section 2.2.8.x

PCI Express CEM Specification - revise a portion of Section 2 as follows:

1.5.　　Electrical Overview

.
.
.

5

❑ Wake (WAKE#), required only if the device/system supports wakeup and/or OBFF mechanism

PCI Express CEM Specification - revise a portion of Section 2 as follows:

10 # 2.　　Auxiliary Signals

.
.
.

15　❑ WAKE#: an open-drain, active low signal that is driven low by a PCI Express function to re-activate the PCI Express Link hierarchy's main power rails and reference clocks. It is required on any add-in card or system board that supports wakeup functionality compliant with this specification. WAKE# is also used by the system to signal to the PCI Express function in conjunction with the Optimized Buffer Flush/Fill (OBFF) mechanism.

20

2.3.　　WAKE# Signal

The WAKE# signal is an open drain, active low signal that is driven low by a PCI Express
5　component to reactivate the PCI Express slot's main power rails and reference clocks. The
WAKE# signal is also used by Downstream Ports to signal to functions on the add-in card in
conjunction with the OBFF mechanism. Only add-in cards that support either the wake process or
the OBFF mechanism connect to this pin. If the add-in card has wakeup capabilities, it must
support the WAKE# function. Likewise, only systems that support the wakeup function or the
10　OBFF mechanism need to connect to this pin, but if they do, they must fully support the WAKE#
function. Such systems are not required to support Beacon as a wakeup mechanism, but are
encouraged to support it. If the wakeup process is used, the +3.3Vaux supply must be present and
used for this function. The assertion and de-assertion of WAKE# are asynchronous to any system
clock. (See Chapter 5 of the *PCI Express Base Specification, Revision 2.0* for more details on PCI-
15　compatible power management.)

If the WAKE# signal is supported by a slot, the signal is connected to the platform's power
management (PM) controller. WAKE# may be bused to multipleall PCI Express add-in card
connectors, forming a single input connection at the PM controller or individual connectors can
have individual connections to the PM controller. Hot-Plug requires that WAKE# be isolated
20　between connectors and driven inactive during the Hot-Plug/Hot Removal events. Refer to Section
<> for the connector pin assignment for the WAKE# signal.

Auxiliary power (+3.3Vaux) must be used by the asserting and receiving ends of WAKE# in order
to revive the hierarchy. The system vendor must also provide a pull-up on WAKE# with its bias
voltage reference being supplied by the auxiliary power source in support of Link reactivation. Note
25　that the voltage that the system board uses to terminate the WAKE# signal can be lower than the
auxiliary supply voltage to be compatible with lower voltage processes of the system PM controller.
However, all potential drivers of the WAKE# signal must be 3.3 V tolerant.

WAKE# must only be asserted by the add-in card when all of its functions are in the D3 state and
at least one of its functions is enabled for wakeup generation using the PME Enable bit in the
30　PMCSR.

Note: WAKE# is not PME# and should not be attached to the PCI-PME# interrupt signals.
WAKE# causes power to be restored but must not directly cause an interrupt.

If the PCI Express add-in card supports the OBFF mechanism defined in the *PCI Express Base
Specification*, then the WAKE# signal may be used as an input to the add-in card. See Chapter 6.x of
35　the *PCI Express Base Specification* for specifics on the OBFF mechanism.

WAKE# will only be asserted by End Points in the L2 state or by Root Complexes (for OBFF or CPU
Active signaling) in the L1, L0 or L0S states.

2.6.1. DC Specifications

Table 2-3: Auxiliary Signal DC Specifications - PERST#, WAKE#, and SMBus

Symbol	Parameter	Conditions	Min	Max	Unit	Notes
V_{IL1}	Input Low Voltage		-0.5	0.8	V	2, 6
V_{IH1}	Input High Voltage		2.0	V_{cc3_3} + 0.5	V	2, 6
V_{IL2}	Input Low Voltage		-0.5	0.8	V	4
V_{IH2}	Input High Voltage		2.1	V_{ccSus3_3} + 0.5	V	4
V_{OL1}	Output Low Voltage	4.0 mA		0.2	V	1, 3
V_{HMAX}	Max High Voltage			V_{cc3_3} + 0.5	V	3
V_{OL2}	Output Low Voltage	4.0 mA		0.4	V	1, 4
I_{in}	Input Leakage Current	0 to 3.3 V	-10	+10	μA	2, 4
I_{lkg}	Output Leakage Current	0 to 3.3 V	-50	+50	μA	3, 5
C_{in}	Input Pin Capacitance			7	pF	2
C_{out}	Output (I/O) Pin Capacitance			30	pF	3,4

Notes:

1. Open-drain output a pull-up is required on the system board. There is no V_{OH} specification for these signals. The number given is the maximum voltage that can be applied to this pin.
2. Applies to PERST#.
3. Applies to WAKE#.
4. Applies to SMBus signals SMBDATA and SMBCLK.
5. Leakage at the pin when the output is not active (high impedance).
6. Applies to WAKE# issued by Switch Downstream Ports and Root Complex for signaling of OBFF indications as received at the input of the Endpoint(s).

5

**PCI Express CEM Specification - change Table 2-4 and Figure 2-14 in Section 2.6.2 as follows:**

2.6.2.   AC Specifications

Table <>: Power Sequencing and Reset Signal Timings

Symbol	Parameter	Min	Max	Units	Notes	Figure
T$_{PVPERL}$	Power stable to PERST# inactive	100		ms	1	<>
T$_{PERST-CLK}$	REFCLK stable before PERST# inactive	100		µs	2	<>
T$_{PERST}$	PERST# active time	100		µs		<>
T$_{FAIL}$	Power level invalid to PERST# active		500	ns	3	<>
T$_{WKRF}$	WAKE# rise – fall time		100	ns	4	<>
T$_{WAKE-TX-MIN-PULSE}$	Minimum WAKE# pulse width; applies to both active-inactive-active and inactive-active-inactive cases	300		ns	5	
T$_{WAKE-FALL-FALL-CPU-ACTIVE}$	Time between two falling WAKE# edges when signaling CPU Active	700	1000	ns	5	

Notes:

1. Any supplied power is stable when it meets the requirements specified for that power supply.
2. A supplied reference clock is stable when it meets the requirements specified for the reference clock. The PERST# signal is asserted and de-asserted asynchronously with respect to the supplied reference clock.
3. The PERST# signal must be asserted within T$_{FAIL}$ of any supplied power going out of specification.
4. Measured from WAKE# assertion/de-assertion to valid input level at the system PM controller. Since WAKE# is an open-drain signal, the rise time is dependent on the total capacitance on the platform and the system board pull-up resistor. It is the responsibility of the system designer to meet the rise time specification.
5. Refers to timing requirement for indicating an active window.
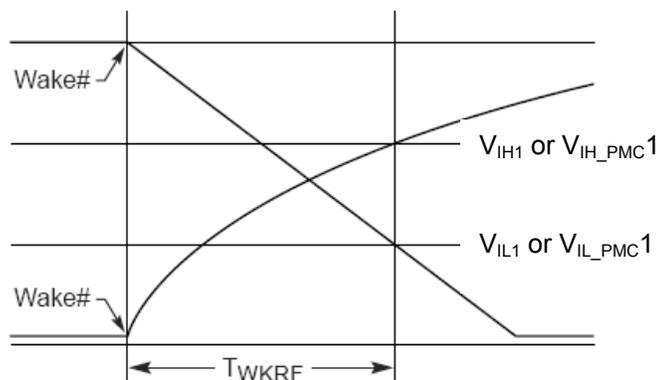
Note 1: Power Management Controller input switching levels are platform dependent and are not set by this specification.

A-0338

Figure 2-14: WAKE# Rise and Fall Time Measurement Points

5

PCI Express Mini CEM Specification - revise the affected portion of Table 3-1 as follows:

Table 3-1: PCI Express Mini Card System Interface Signals

Signal Group	Signal	Direction	Description
Auxiliary Signals (3.3V Compliant)	PERST#	Input	Functional reset to the card
	CLKREQ#	Output	Reference clock request signal
	WAKE#	Input/Output	Open Drain active Low signal. When the add-in card supports wakeup, Tthis signal is used by the add-in card to request that the system return from a sleep/suspended state to service a function initiated wake event. When the add-in card supports the OBFF mechanism, this signal is used by the system to indicate OBFF or CPU Active state transitions.
	SMB_DATA	Input/Output	SMBus data signal compliant to the SMBus 2.0 specification
	SMB_CLK	Input	SMBus clock signal compliant to the SMBus 2.0 specification

PCI Express Mini CEM Specification - revise Section 3.2.4.4 as follows:

3.2.4.4.   WAKE# Signal

PCI Express Mini Cards must implement WAKE# if the card supports either the wakeup function or the OBFF mechanism. See the *PCI Express Card Electromechanical Specification* for details on the functional requirements for the WAKE# signal.

5

3.4.1. Logic Signal Requirements

The 3.3V card logic levels for single-ended digital signals (WAKE#, CLKREQ#, PERST#, and W_DISABLE#) are given in Table 3-7.

Table 3-7: DC Specification for 3.3V Logic Signaling

Symbol	Parameter	Conditions	Min	Max	Units	Notes
+3.3Vaux	Supply Voltage		3.3 – 9%	3.3 + 9%	V	3
V_{IH}	Input High Voltage		2.0	3.6	V	1
V_{IL}	Input Low Voltage		-0.5	0.8	V	1
I_{OL}	Output Low Current for open-drain signals	0.4 V	4		mA	2
I_{IN}	Input Leakage Current	0 V to 3.3 V	-10	+10	µA	1
I_{LKG}	Output Leakage Current	0 V to 3.3 V	-50	+50	µA	1
C_{IN}	Input Pin Capacitance			7	pF	1
C_{OUT}	Output Pin Capacitance			30	pF	2

Notes:
1. Applies to PERST# ~~and,~~ W_DISABLE# and WAKE# (when used for OBFF signaling).
2. Applies to CLKREQ# and WAKE#.
3. As measured at the card connector pad.