PCI-SIG ENGINEERING CHANGE NOTICE

TITLE:	Dynamic Power Allocation
DATE:	May 24, 2008
AFFECTED DOCUMENT:	PCI Express Base Specification version 2.0
SPONSOR:	Intel Corporation, Hewlett-Packard, IBM

Part I

1. Summary of the Functional Changes

DPA (Dynamic Power Allocation) extends existing PCIe device power management to provide active (D0) device power management substates for appropriate devices, while comprehending existing PCIe PM Capabilities including PCI-PM and Power Budgeting.

2. Benefits as a Result of the Changes

DPA complements on-going industry efforts to optimize component and platform power management to meet new customer and regulatory operating requirements.   DPA provides a capability so that all Endpoint Functions can have a mechanism for the dynamic allocation of power. DPA is applicable to appropriate PCIe Endpoint Functions.

3. Assessment of the Impact

DPA is optional normative functionality, applicable to appropriate Endpoint Functions

4. Analysis of the Hardware Implications

Implementations that support this mechanism must implement the DPA Capability and associated ability to dynamically adjust Endpoint Function power utilization.

5. Analysis of the Software Implications

The DPA Capability enables a method for software to discover and actively manage Endpoint Function power usage. Power management policies used to determine which power state to operate at are outside the scope of the PCI-SIG.

Part II

Detailed Description of the change

Add Section 6.x

6.x Dynamic Power Allocation (DPA) Capability

A common approach to managing power consumption is through a negotiation between the device driver, operating system, and executing applications. Adding Dynamic Power Allocation for such devices is anticipated to be done as an extension of that negotiation, through software mechanisms that are outside of the scope of this specification. Some devices do not have a device specific driver to manage power efficiently. The DPA Capability provides a mechanism to allocate power dynamically for these types of devices. DPA is optional normative functionality applicable to Endpoint Functions that can benefit from the dynamic allocation of power and do not have an alternative mechanism.

The DPA Capability enables software to actively manage and optimize Function power usage when in the D0 state. DPA is not applicable to power states D1-D3 therefore the DPA Capability is independently managed from the PCI-PM Capability.

DPA defines a set of power substates, each of which with an associated power allocation. Up to 32 substates [0..31] can be defined per Function. Substate 0, the default substate, indicates the maximum power the Function is ever capable of consuming.

Substates must be contiguously numbered from 0 to Substate_Max, as defined in section 7.x.2. Each successive substate has a power allocation lower than or equal to that of the prior substate. For example, a Function with four substates could be defined as follows:

1. Substate 0 (the default) defines a power allocation of 25 Watts.
2. Substate 1 defines a power allocation of 20 Watts
3. Substate 2 defines a power allocation of 20 Watts
4. Substate 3 defines a power allocation of 10 Watts.

When the Function is initialized, it will operate within the power allocation associated with substate 0. Software is not required to progress through intermediate substates. Over time, software may dynamically configure the Function to operate at any of the substates in any sequence it chooses. Software is permitted to configure the Function to operate at any of the substates before the Function completes a previously initiated substate transition.

On the completion of the substate transition(s) the Function must compare its substate with the configured substate. If the Function substate does not match the configured substate, then the Function must begin transition to the configured substate. It is permitted for the Function to dynamically alter substate transitions on Configuration Requests instructing the Function to operate in a new substate.

In the prior example, software can configure the Function to transition to substate 4, followed by substate 1, followed by substate 3, and so forth. As a result, the Function must be able to transition between any substates when software configures the associated control field.

The Substate Control Enabled bit provides a mechanism that allows the DPA Capability to be used in conjunction with the software negotiation mechanism mentioned above. When Set, power allocation is controlled by the DPA Capability. When Clear, the DPA Capability is disabled, and the Function is not permitted to directly initiate substate transitions based on configuration of the Substate Control register field. At an appropriate point in time, software participating in the software negotiation mechanism mentioned above clears the bit, effectively taking over control of power allocation for the Function.

It is required that the Function respond to Configuration Space accesses while in any substate.

At any instant, the Function must never draw more power than it indicates through its Substate Status. When the Function is configured to transition from a higher power substate to a lower power substate, the Function's Substate Status must indicate the higher power substate during the transition, and must indicate the lower power substate after completing the transition. When the Function is configured to transition from a lower power substate to a higher power substate, the Function's Substate Status must indicate the higher power substate during the transition, as well as after completing the transition.

Due to the variety of applications and the wide range of maximum power required for a given Function, the transition time required between any substates is implementation specific. To enable software to construct power management policies (outside the scope of this specification), the Function defines two Transition Latency Values. Each of the Function substates associates its maximum Transition Latency with one of the Transition Latency Values, where the maximum Transition Latency is the time it takes for the Function to enter the configured substate from any other substate. A Function is permitted to complete the substate transition faster than the maximum Transition Latency for the substate.

6.x.1 DPA Capability with Multi-Function Devices

It is permitted for some or all Functions of a multi-Function device to implement a DPA Capability. The power allocation for the multi-Function device is the sum of power allocations set by the DPA Capability for each Function. It is permitted for the DPA Capability of a Function to include the power allocation for the Function itself as well as account for power allocation for other Functions that do not implement a DPA Capability. The association between multiple Functions for DPA is implementation specific and beyond the scope of this specification.

Add Section 7.x

7.x Dynamic Power Allocation (DPA) Capability

The DPA Capability structure is shown in Figure 7-xx below

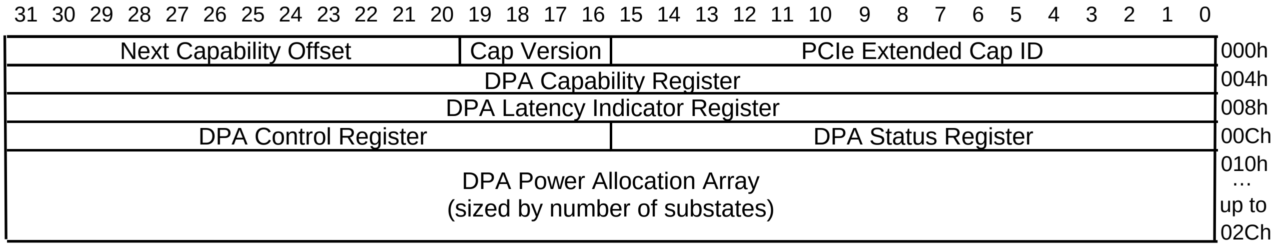

Figure 7-xx Dynamic Power Allocation Capability Structure

7.x.1. DPA Extended Capability Header (Offset 00h)

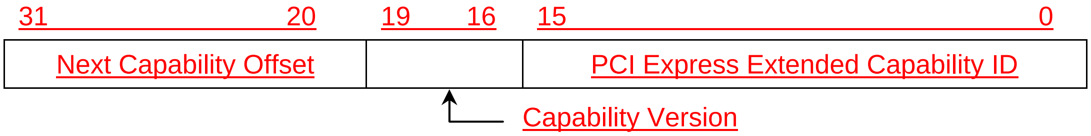

Figure 7-xx DPA Extended Capability Header

Table 7-xx DPA Extended Capability Header

Bit Location	**Register Description**	**Attributes**
15:0	**PCI Express Extended Capability ID** – This field is a PCI-SIG defined ID number that indicates the nature and format of the Extended Capability. PCI Express Extended Capability ID for the DPA Extended Capability is 0016h.	RO
19:16	**Capability Version** – This field is a PCI-SIG defined version number that indicates the version of the Capability structure present. Must be 1h for this version of the specification.	RO
31:20	**Next Capability Offset** – This field contains the offset to the next PCI Express Extended Capability structure or 000h if no other items exist in the linked list of capabilities.	RO

7.x.2. DPA Capability Register (Offset 04h)

31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Xlcy1								Xlcy0								RsvdZ		PAS		RsvdZ		Tlunit		RsvdZ			Substate_Max				

Figure 7-xx DPA Capability Register

Table 7-xx DPA Capability Register

Bit Location	Register Description	Attributes
4:0	**Substate_Max:** Value indicates the maximum substate number, which is the total number of supported substates minus one. A value of 00000b indicates support for one substate.	RO
9:8	**Transition Latency Unit (Tlunit):** A substate's Transition Latency Value is multiplied by the Transition Latency Unit to determine the maximum Transition Latency for the substate. Defined encodings are 00b – 1ms 01b – 10ms 10b – 100 ms 11b – Reserved	RO
13:12	**Power Allocation Scale (PAS):** The encodings provide the scale to determine power allocation per substate in Watts. The value corresponding to the substate in the Substate Power Allocation Register is multiplied by this field to determine the power allocation for the substate. Defined encodings are 00b – 10.0x 01b – 1.0x 10b – 0.1x 11b – 0.01x	RO
23:16	**Transition Latency Value 0 (Xlcy0):** This value is multiplied by the Transition Latency Unit to determine the maximum Transition Latency for the substate	RO
31:24	**Transition Latency Value 1 (Xlcy1):** This value is multiplied by the Transition Latency Unit to determine the maximum Transition Latency for the substate	RO

7.x.3. DPA Latency Indicator Register (Offset 08h)

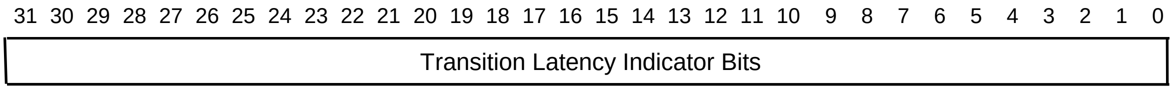

Figure 7-xx DPA Latency Indicator Register

Table 7-xx DPA Latency Indicator Register

Bit Location	Register Description	Attributes
Substate_Max:0	**Transition Latency Indicator Bits:** Each bit indicates which Transition Latency Value is associated with the corresponding substate. A value of 0b indicates Transition Latency Value 0; a value of 1b indicates Transition Latency Value 1.	RO
All other bits	Reserved	RsvdP

7.x.4. DPA Status Register (Offset 0Ch)

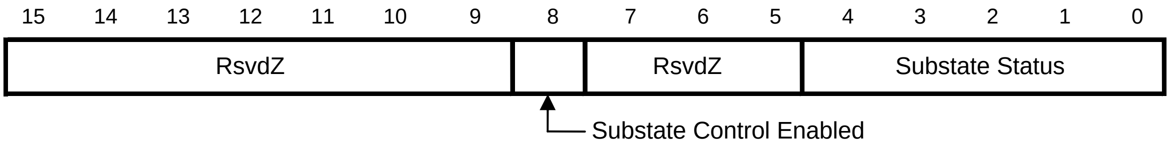

Figure 7-xx DPA Status Register

Table 7-xx DPA Status Register

Bit Location	Register Description	Attributes
4:0	**Substate Status:** Indicates current substate for this Function Default is 00000b	RO
8	**Substate Control Enabled:** Used by software to disable the Substate Control field in the DPA Control register. Hardware sets this bit following a Conventional Reset or FLR. Software clears this bit by writing a 1b to it. Software is unable to set this bit directly. When this bit is Set, the Substate Control field determines the current substate. When this bit is Clear, the Substate Control field has no effect on the current substate. Default value is 1b.	RW1C

7.x.5. DPA Control Register (Offset 0Eh)

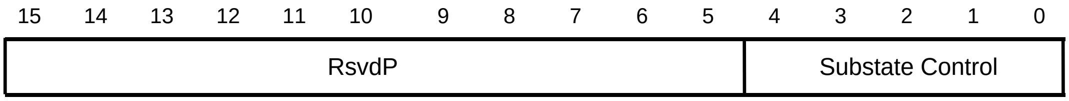

Figure 7-xx DPA Control Register

Table 7-xx DPA Control Register

Bit Location	**Register Description**	**Attributes**
4:0	**Substate Control:** Used by software to configure the Function substate. Software writes the substate value in this field to initiate a substate transition. When the Substate Control Enabled bit in the DPA Status register is Set, this field determines the Function substate. When the Substate Control Enabled bit in the DPA Status register is Clear, this field has no effect the Function substate. Default value is 00000b.	RW

7.x.6. DPA Power Allocation Array

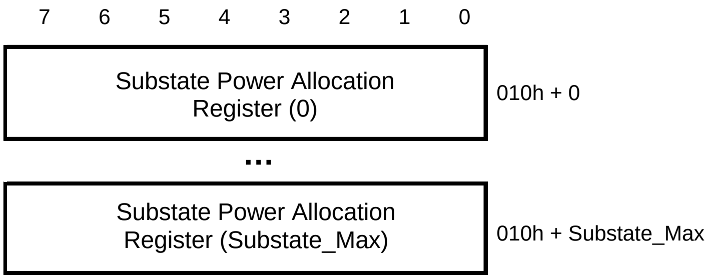
7
6
5
4
3
2
1
0
Substate Power Allocation Register (0)
010h + 0
...
Substate Power Allocation Register (Substate_Max)
010h + Substate_Max

Figure 7-xx DPA Power Allocation Array

Each Substate Power Allocation Register indicates the power allocation value for its associated substate. The number of Substate Power Allocation Registers implemented must be equal to the number of substates supported by Function, which is Substate_Max plus one.

Table 7-xx Substate Power Allocation Register (0 to Substate_Max)

Bit Location	**Register Description**	**Attributes**
7:0	**Substate Power Allocation:** The value in this field is multiplied by the Power Allocation Scale to determine power allocation in Watts for the associated substate.	RO