NVM Express Technical Errata

Errata ID	**003**
Affected Spec Ver.	**NVM Express 1.0**
Corrected Spec Ver.	

Submission info

Name	Company	Date
Mike Shapiro	DSSD	3/3/2011
Matthew Wilcox	Intel	3/8/2011
Amber Huffman	Intel	3/3/2011

This erratum enables future implementations to support a dynamic doorbell size. This capability enables more efficient software emulation of an NVM Express controller. For example, a software thread monitoring doorbell notifications may observe a cacheline accessed by a spin loop or utilize the monitor/mwait CPU instructions. This capability is enabled in a way that does not impact any hardware implementations.

Description of the specification technical flaw

Modify section 3.1 as shown:

The following table describes the register map for the controller.

Start	End	Symbol	Description
00h	07h	CAP	Controller Capabilities
08h	0Bh	VS	Version
0Ch	0Fh	INTMS	Interrupt Mask Set
10h	13h	INTMC	Interrupt Mask Clear
14h	1Bh	CC	Controller Configuration
1Ch	1Fh	CSTS	Controller Status
20h	23h	Reserved	Reserved
24h	27h	AQA	Admin Queue Attributes
28h	2Fh	ASQ	Admin Submission Queue Base Address
30h	37h	ACQ	Admin Completion Queue Base Address
38h	EFFh	Reserved	Reserved
F00h	FFFh	Reserved	Command Set Specific
1000h	1003h	SQ0TDBL	Submission Queue 0 Tail Doorbell (Admin)
~~1004h~~ 1000h + (1 * (4 << CAP.DSTRD))	~~1007h~~ 1003h + (1 * (4 << CAP.DSTRD))	CQ0HDBL	Completion Queue 0 Head Doorbell (Admin)
~~1008h~~ 1000h + (2 * (4 << CAP.DSTRD))	~~100Bh~~ 1003h + (2 * (4 << CAP.DSTRD))	SQ1TDBL	Submission Queue 1 Tail Doorbell
~~100Ch~~ 1000h + (3 * (4 << CAP.DSTRD))	~~100Fh~~ 1003h + (3 * (4 << CAP.DSTRD))	CQ1HDBL	Completion Queue 1 Head Doorbell
~~1010h~~ 1000h + (4 * (4 << CAP.DSTRD))	~~1013h~~ 1003h + (4 * (4 << CAP.DSTRD))	SQ2TDBL	Submission Queue 2 Tail Doorbell
~~1014h~~ 1000h + (5 * (4 << CAP.DSTRD))	~~1017h~~ 1003h + (5 * (4 << CAP.DSTRD))	CQ2HDBL	Completion Queue 2 Head Doorbell
…	…	…	…
1000h + (2y * (4 << CAP.DSTRD)) ~~(8*y)~~	1003h + (2y * (4 << CAP.DSTRD)) ~~(8*y)~~	SQyTDBL	Submission Queue y Tail Doorbell
~~1004h + (8*y)~~ 1000h + ((2y + 1) * (4 << CAP.DSTRD))	~~1007h + (8*y)~~ 1003h + ((2y + 1) * (4 << CAP.DSTRD))	CQyHDBL	Completion Queue y Head Doorbell
			Vendor Specific (Optional)

Modify the definition of the Controller Capabilities register in section 3.1.1 as shown:

36~~:32~~	RO	0~~h~~	Reserved
35:32	RO	Impl Spec	**Doorbell Stride (DSTRD):** Each Submission Queue and Completion Queue Doorbell register is 32-bits in size. This register indicates the stride between doorbell registers. The stride is specified as (2 ^ (2 + DSTRD)) in bytes. A value of 0h indicates a stride of 4 bytes, where the doorbell registers are packed without reserved space between each register. Refer to section 8.6.

Modify the heading of section 3.1.10 as shown:

3.1.10 Offset (1000h + ((2y) * (4 << CAP.DSTRD)) ~~8*y~~): SQyTDBL – Submission Queue y Tail Doorbell

Modify the heading of section 3.1.11 as shown:

3.1.11 Offset (1000h + ((2y + 1) * (4 << CAP.DSTRD)) ~~1004h + 8*y~~): CQyHDBL – Completion Queue y Head Doorbell

Add section 8.6 to the specification, as shown:

8.6 Doorbell Stride for Software Emulation

The doorbell stride, specified in CAP.DSTRD, may be used to separate doorbells by a number of bytes in memory space. The doorbell stride is a number of bytes equal to (2 ^ (2 + CAP.DSTRD)). This is useful in software emulation of an NVM Express controller. In this case, a software thread is monitoring doorbell notifications. The software thread may be made more efficient by monitoring one doorbell per discrete cacheline or utilize the monitor/mwait CPU instructions. For hardware implementations of NVM Express, the expected doorbell stride value is 0h.

Disposition log

3/3/2011	Erratum captured.
3/8/2011	Updated formulas in erratum, added introductory section.
3/13/2011	Updated formulas in erratum.
3/17/2011	Added sentence to section 8.6.
3/21/2011	Updates to formulas based on reflector feedback.
4/26/2011	Erratum ratified.