LEGAL NOTICE:

NVM Express Workgroup
c/o VTM, Inc.
3855 SW 153rd Drive
Beaverton, OR 97003
USA
info@nvmexpress.org

NVM Express Technical Proposal for New Feature

Technical Proposal ID	**4082 Multiple Controller Shutdown Enhancements**
Change Date	**2021-03-01**
Builds on Specification	**NVM Express 1.4b** **NVM Express Management Interface 1.1b**
References Specification	**TP 4009 Partitions and Domains** **TP 4029 Power Loss Signaling**

Technical Proposal Author(s)

Name	**Company**
Mike Allison, Jonathan Hughes, Nick Adams	Intel
David Black, Austin Bolen	Dell EMC

This proposal intends to update the process of shutting down a domain that contains multiple controllers.

Revision History

Revision Date	**Change Description**
2020-02-25	Initial version
2020-02-26	Used “main power” when for controller power. Added comments per Austin Bolen.
2020-02-27	Changes CPS field to be 2 bits. Updated descriptions in the CC.EN and CC.SHN fields.
2020-03-03	Added the requirement for secondary controllers to transition to Offline when its associated primary controller is shutting down. Added a reference to TP 4029 Power Loss Signaling as there is overlap on the Unsafe Shutdowns count.
2020-03-31	Jonathan Hughes and I determined that the value of CC.SHN is sufficient and that a transition of the value is not required. We determined that CSTS.RDY timing has to account for shutdown of the controller to coordinate when it is proper for a host to modify CC.EN.
2020-04-06	Modifications during Management Interface Task Group discussion. Updates the Reset Type definition.
2020-04-13	Feedback from 4/9/2020 Technical WG meeting was to have a bit in CSTS register to indicate when power is ready to be removed from controllers that share power.
2020-05-01	Updates the pattern to “Shdn”. Moved the NVMe-MI shutdown to a specific message instead of in the Reset command as the shutdown has to understand domains.
2020-06-02	Added the new event type for auto-clearing events not associated with a log page. Ties SMART Unsafe Shutdown to PRR when CAP.CPS not set to 00b (backwards compatible). Added new sections defining NVM subsystem and domain shutdown with NVM Subsystem Reset recovery.

2020-06-03	Changes the CSTS.PRR bit to a field so that the field can state that power can be removed the reset required to start command processing again without a power cycle. This gives the host indication of reset required to start executing commands without a power cycle. Removed new text from section 7.6.2 that was moved to the new 7.6.2.NEW_1 and 7.6.2.NEW_2 sections.
2020-06-17	Added a comment to discuss if a Controller Reset was allowed after an NVM Subsystem Shutdown completes to start command processing.
2020-06-25	Added definition that Controller Reset has no effect on NVM Subsystem Shutdown until NVM Controller Reset. Modified NVMe-MI Shutdown command to only be at NVM Subsystem Level to be consistent with Reset command being at NVM Subsystem Level. Use another TP to make Reset and Shutdown command support domains.
7/22/2020	Addressed comments submitted by Judy Brock that included editorial changes. The big items changed are: • If CAP.CPS is set to 00b then the PRR field is reserved. This is how the host know the existence of PRR field. • Renamed and changed the new category of asynchronous events. • Add the new event to the OAES and Asynchronous Event Configuration FID to allow the host to disable the event.
7/23/2020	PRR only needs to be 1 bit now.
7/27/2020	Accepted all changes as the number of edits was too large to manage. Removed CSTS.PRR and added CSTS.ST.
8/6/2020	Updates based on state machine presented in Technical WG meeting.
8/6/2020	Removed existing comments as have been addressed. Minor edits per Technical WG review.
8/10/2020	Updated CSTS.SHST to indicate that host software might not observe the transition from 00b to 01b is a controller shutdown is not aborted. Added same text to NVMe-MI for clarity. Updated the NVMe 1.4 changes to include the clarify added for PCI Conventional Resets and FLR.
9/24/2020	Added the text to CC.EN has no effect on an NVM Subsystem Shutdown. Editorial changes (bolding, capitalization, better language). NVMe-MI Shutdown command allows normal and abrupt so updated NVM Subsystem Shutdown sections to list the NVMe-MI capabilities in all cases (some were missing).
9/28/2020	Allowing Shutdown command during a sanitize operation. Made the Shutdown command optional in the in-band mechanism. Editorial changes to the Shutdown command. Clarified the optional/mandatory requirements for the Shutdown command shutdown types for in-band and out-of-band mechanisms.
10/1/2020	More editorial changes in getting ready for member review. Aligned to NVMe 1.4b.
10/7/2020	Added the update to section 7.1.2.
10/19/2020	Updated the wording to indicate the initiation of the shutdown type for an NVMe-Mi Shutdown command.
10/21/2020	Fixed all comments previous. Minor edits.
10/29/2020	Removed the use of PCIe level resets and used NVM Subsystem Reset and Controller Level Reset type text.
10/29/2020	Changed “a <acronym> to an <acronym>. Editorial changes for set/clears verbiage to be consistent.
11/2/2020	Remove addition of CSTS.ST bit in NVMe-MI as Management Interface Task Group said the value was not needed at this time.
11/5/2020	Removed all comments and accepted all changes for 30 day member review.
12/8/2020	Removed repeated sentence.
12/10/2020	Review in NVMe Technical WG approved integration. Accepted all changed and removed all comments.
01/03/2021	Integrated into the NVMe Base Specification and the NVMe Management Interface Specification, Revision 1.1. Shutdown command optional to align with TP 6028.

01/14/2021	Accepted all changes and removed all comments for member review. Adjusted filename to meet TP policy.
01/15/2021	Member review comment from Paul Suhler. Changed “writing of the value” to “writing the value”. Changed “abort shutdown” to “abrupt shutdown”.
02/03/2021	Fixing plural grammar.
02/04/2021	Fixing spelling “Rest” -> “Reset”
02/10/2021	Removed extra “associated” in a sentence.
02/13/2021	Added a note to the CSTS register that field values may change prior to setting the reset value.
02/14/2021	Updated the description of changes to match the new format and include a summary of the changes. Clarified text that detailed a normal shutdown and an abrupt shutdown. Removed “is shutdown” from a sentence as was not necessary.
02/15/2021	Made the statement about the host not seeing transition of the CSTS.SHST field to be plural.
2/18/2021	Accepted all changes and removed all comments for integration.
2/24/2021	Integrated into the NVMe Base Specification and the NVMe Management Interface Specification, Revision 1.1.
2/25/2021	Changed one more instance of “writing of the value” to “writing the value”.
3/1/2021	Accepted all changes, removed all comments, and converted all references/cross-references to text.

Description for NVMe Base Specification 1.4b Changes Document

Feature Enhancement:

This technical proposal adds the process of shutting down a domain or NVM subsystem that has multiple controllers. To support it the following has been changed:

- Added the new NSSD register that indicates to a controller that all controllers within the domain or NVM subsystem must be shut down (i.e., initiate a NVM Subsystem shutdown).
- Updated the shutdown process to include shutting down all controllers in a domain or NVM subsystem that included modifying the CAP, CC, and CSTS registers.
- Add the behavior of a shutdown occurring when the controller is disable (CC.EN is cleared to ‘0’).
- Add the new Immediate event type to notify hosts immediately of an event only if there is an outstanding Asynchronous Event Request command. Defined a new Normal NVM Subsystem Shutdown event to be issued if a normal NVM Subsystem shutdown is initiated.
- **New requirement / incompatible change**:
  - Changed the definition of the Unsafe Shutdowns field in the SMART / Health Information Log to increment when the controller is shutdown when the controller does not report it is safe to power down.
- **New requirement / incompatible change**
  - Added the requirement that a secondary controller transitions to the Offline state when the associated primary controller is shutdown.
- Clarified the effects on a controller shutdown due to PCIe Conventional reset and FLR.

- References:
  - NVM Express Base Specification 1.1b sections 3.1, 3.1.1, 3.1.5, 3.1.6, 5.14.1.2, 5.15.2.2, 5.21.1.11, 6.4.2.1, 7.6.2, 8.5.3.

Description for NVMe Management Interface Specification 1.1b Changes Document

Feature Enhancement:

- Added the NVMe-MI Shutdown command to initiate an NVM Subsystem shutdown.
- References:
    - NVM Express Management Interface Specification 1.1b sections 5.3, 5.8, and 6.3

Description of Specification Changes

Markup Conventions:

Black:	Unchanged (however, hot links are removed)
~~Red Strikethrough~~:	Deleted
Blue:	New
Blue Highlighted:	TBD values, anchors, and links to be inserted in new text.
<Green Bracketed>:	Notes to editor

Modify portions of NVMe 1.4b as shown below:

Modify portions of section 3.1 as shown below:

3.1 Register Definition

Figure 68 describes the register map for the controller.

The Vendor Specific address range starts after the last doorbell supported by the controller and continues to the end of the BAR0/1 supported range. The start of the Vendor Specific address range starts at the same location and is not dependent on the number of allocated doorbells.

Figure 68: Register Definition

Start	End	Symbol	Description
…			
58h	5Bh	CMBSTS	Controller Memory Buffer Status (Optional)
64h	67h	NSSD	NVM Subsystem Shutdown (Optional)
68h~~5Ch~~	DFFh	Reserved	Reserved

Modify portions of section 3.1.1 as shown below:

3.1.1 Offset 0h: CAP – Controller Capabilities

This register indicates basic capabilities of the controller to host software.

Figure 69: Offset 0h: CAP – Controller Capabilities

Bits	Type	Reset	Description
...			
58	RO	Impl Spec	**NVM Subsystem Shutdown Supported (NSSS):** This bit indicates whether the controller supports the NVM Subsystem Shutdown feature defined in section 7.6.NEW. If the controller supports the NVM Subsystem Shutdown feature, then this bit is set to '1'. If the controller does not support the NVM Subsystem Shutdown feature, then this bit is cleared to '0'. If the NSSRS bit is cleared to '0', then this bit shall be cleared to '0'.

Figure 69: Offset 0h: CAP – Controller Capabilities

<table>
<tr><th>Bits</th><th>Type</th><th>Reset</th><th>Description</th></tr>
<tr><td colspan="4">…</td></tr>
<tr><td>51:48</td><td>RO</td><td>Impl Spec</td><td>Memory Page Size Minimum (MPSMIN): This field indicates the minimum host memory page size that the controller supports. The minimum memory page size is (2 ^ (12 + MPSMIN)). The host shall not configure a memory page size in CC.MPS that is smaller than this value.</td></tr>
<tr><td>47:46</td><td>RO</td><td>Impl Spec</td><td>Controller Power Scope (CPS): This field indicates scope of controlling the main power for this controller.
<table><tr><th>Value</th><th>Power Scope</th></tr><tr><td>00b</td><td>Not Reported</td></tr><tr><td>01b</td><td>Controller scope</td></tr><tr><td>10b</td><td>Domain scope (i.e, the NVM subsystem supports multiple domains (refer to TP 4009:7.NEW).</td></tr><tr><td>11b</td><td>NVM subsystem scope (i.e., the NVM subsystem does not support multiple domains).</td></tr></table>If the NSSS bit is set to ‘1’, then this field shall not be cleared to 00b.</td></tr>
</table>

Modify portions of section 3.1.5 as shown below:

3.1.5 Offset 14h: CC – Controller Configuration

This register modifies settings for the controller. Host software shall set the Arbitration Mechanism (CC.AMS), the Memory Page Size (CC.MPS), and the Command Set (CC.CSS) to valid values prior to enabling the controller by setting CC.EN to ‘1’. Attempting to create an I/O queue before initializing the I/O Completion Queue Entry Size (CC.IOCQES) and the I/O Submission Queue Entry Size (CC.IOSQES) should cause a controller to abort a Create I/O Completion Queue command or a Create I/O Submission Queue command with a status code of Invalid Queue Size.

Figure 78: Offset 14h: CC – Controller Configuration

<table>
<tr><th>Bits</th><th>Type</th><th>Reset</th><th>Description</th></tr>
<tr><td>…</td><td></td><td></td><td></td></tr>
<tr><td>15:14</td><td>RW</td><td>00b</td><td>Shutdown Notification (SHN): This field is used to initiate a controller shutdown ~~processing~~ when ~~a shutdown is occurring, (i.e., a~~ power down condition is expected~~)~~. For a normal controller shutdown ~~notification~~, it is expected that the controller is given time to process the controller shutdown ~~notification~~. For an abrupt shutdown ~~notification~~, the host may not wait for the controller shutdown ~~processing~~ to complete before power is lost.

The controller shutdown notification values are defined as:

<table><tr><th>Value</th><th>Definition</th></tr><tr><td>00b</td><td>No notification; no effect</td></tr><tr><td>01b</td><td>Normal shutdown notification</td></tr><tr><td>10b</td><td>Abrupt shutdown notification</td></tr><tr><td>11b</td><td>Reserved</td></tr></table>
This field should be written by host software prior to any power down condition and prior to any change of the PCI power management state. It is recommended that this field also be written prior to a warm reset (refer to the PCI Express Specification) ~~reboot~~. To determine when the controller shutdown processing is complete, refer to CSTS.ST and CSTS.SHST. Refer to section 7.6.2 for additional controller shutdown processing details.

Other fields in ~~the~~ this register (including the EN bit) may be modified as part of updating this field to 01b or 10b to initiate a controller shutdown. If the EN bit is cleared to ‘0’ such that the EN bit transitions from ‘1’ to ‘0’, then both a Controller Reset and a controller shutdown occur.

If an NVM Subsystem Shutdown is in progress or is being reported as completed (i.e., CSTS.ST is set to ‘1’, and CSTS.SHST is set to 01b or 10b), then writes to this field modify the field value but have no effect. Refer to section 7.6.New for details.</td></tr>
<tr><td>…</td><td></td><td></td><td></td></tr>
</table>

Figure 78: Offset 14h: CC – Controller Configuration

Bits	Type	Reset	Description
00	RW	0b	**Enable (EN):** When set to ‘1’, then the controller shall process commands based on Submission Queue Tail doorbell writes. When cleared to ‘0’, then the controller shall not process commands nor post completion queue entries to Completion Queues. When this bit transitions from ‘1’ to ‘0’, the controller is reset (i.e., a Controller Reset). That reset deletes all I/O Submission Queues and I/O Completion Queues, resets the Admin Submission Queue and Completion Queue, and brings the hardware to an idle state. That reset does not affect PCI Express registers (including MMIO MSI-X registers), nor the Admin Queue registers (AQA, ASQ, or ACQ). All other controller registers defined in this section and internal controller state (e.g., Feature values defined in section 5.21.1 that are not persistent across power states) are reset to their default values. The controller shall ensure that there is no data loss for commands that have had corresponding completion queue entries posted to an I/O Completion Queue prior to that Controller Reset. Refer to section 7.3. When this bit is cleared to ‘0’, the CSTS.RDY bit is cleared to ‘0’ by the controller once the controller is ready to be re-enabled. When this bit is set to ‘1’, the controller sets CSTS.RDY to ‘1’ when it is ready to process commands. CSTS.RDY may be set to ‘1’ before namespace(s) are ready to be accessed. Setting this bit from a ‘0’ to a ‘1’ when CSTS.RDY is a ‘1’ or clearing this bit from a '1' to a '0' when CSTS.RDY is cleared to '0' has undefined results. The Admin Queue registers (AQA, ASQ, and ACQ) shall only be modified when this bit ~~EN~~ is cleared to ‘0’. If an NVM Subsystem Shutdown is in progress or is completed (i.e., CSTS.ST is set to ‘1’, and CSTS.SHST is set to 01b or 10b), then writes to this field modify the field value but have no effect. Refer to section 7.6.New for details.

Modify portions of section 3.1.6 as shown below:

3.1.6 Offset 1Ch: CSTS – Controller Status

…

Figure 79: Offset 1Ch: CSTS – Controller Status

<table>
<tr><th>Bits</th><th>Type</th><th>Reset[1]</th><th>Description</th></tr>
<tr><td colspan="4">…</td></tr>
<tr><td>06</td><td>RO</td><td>Impl Spec</td><td>Shutdown Type (ST): When CSTS.SHST is set to a non-zero value, then this bit indicates the type of shutdown reported by CSTS.SHST. If this bit is set to ‘1’, then CSTS.SHST is reporting the state of an NVM Subsystem Shutdown. If this bit is cleared to ‘0’, then CSTS.SHST is reporting the state of a controller shutdown.

If CSTS.SHST is cleared to 00b, then this bit is ignored.</td></tr>
<tr><td colspan="4">…</td></tr>
<tr><td>03:02</td><td>RO</td><td>00b</td><td>Shutdown Status (SHST): This field indicates the status of shutdown processing that is initiated by the host setting the CC.SHN field, the host setting the NSSC register, or a Management Endpoint has processed an NVMe-MI Shutdown command (refer to the NVM Express Management Interface specification).

The shutdown status values are defined as:
<table><tr><th>Value</th><th>Definition</th></tr><tr><td>00b</td><td>Normal operation (no shutdown has been requested)</td></tr><tr><td>01b</td><td>Shutdown processing occurring</td></tr><tr><td>10b</td><td>Shutdown processing complete</td></tr><tr><td>11b</td><td>Reserved</td></tr></table>If this field is set to 01b, then:

• an NVM Subsystem Reset aborts a controller shutdown and an NVM Subsystem Shutdown; and
• any other type of Controller Level Reset:
o may or may not abort a controller shutdown; and
o shall not abort an NVM Subsystem Shutdown.

If this field is set to 01b when a Controller Level Reset is initiated and the shutdown is not aborted, then this field transitions to 00b on the reset and then to 01b to indicate the shutdown is still in progress and host software may or may not observe these transitions.

If CSTS.ST is cleared to ‘0’ and this field is set to 10b, then

• If CC.EN is set to ‘1’, t~~T~~o start executing commands on the controller ~~after a shutdown operation (CSTS.SHST set to 10b)~~, a Controller Level Reset (i.e., CC.EN is cleared to ‘0’) is required. If host software submits commands to the controller without issuing a Controller Level Reset, the behavior is undefined~~.~~; and

• If CC.EN is cleared to ‘0’, to start executing commands on the controller:
o a Controller Level Reset is required; or
o CC.EN is required to be set to ‘1’ and CC.SHN is required to be cleared to 00b with the same write to the CC register.

If CSTS.ST is set to ‘1’ and this field is set to 10b, then an NVM Subsystem Reset is required to start executing commands.

Refer to section 7.6.New on the reset behavior of this field when CAP.CPS is set to 10b or 11b.</td></tr>
<tr><td>01</td><td>RO</td><td>HwInit</td><td>Controller Fatal Status (CFS): This bit is set to ’1’ when a fatal controller error occurred that could not be communicated in the appropriate Completion Queue. This bit is cleared to ‘0’ when a fatal controller error has not occurred. Refer to section 10.5.

The reset value of this bit is set to '1' when a fatal controller error is detected during controller initialization.</td></tr>
</table>

Figure 79: Offset 1Ch: CSTS – Controller Status

Bits	Type	Reset[1]	Description
00	RO	0b	**Ready (RDY):** This bit is set to '1' when the controller is ready to accept Submission Queue Tail doorbell writes after CC.EN is set to '1'. This bit shall be cleared to '0' when CC.EN is cleared to '0' once the controller is ready to be re-enabled. Commands shall not be submitted to the controller until this bit is set to '1' after the CC.EN bit is set to '1'. Failure to follow this requirement produces undefined results. If shutdown processing is not occurring on the controller (CSTS.SHST is set to 00b), then ~~the H~~host software shall wait a minimum of CAP.TO seconds for this bit to be set to '1' after setting CC.EN to '1' from a previous value of '0'. If a controller shutdown processing is occurring on the controller (i.e., CSTS.ST is cleared to '0' and CSTS.SHST is set to 01b), then the host software shall wait a minimum of RTD3 Entry Latency microseconds (refer to the Identify Controller data structure) plus CAP.TO seconds for this bit to be set to '1' after setting CC.EN to '1' from a previous value of '0'. If an NVM Subsystem Shutdown has completed that affects this controller (i.e., CSTS.ST is set to '1' and CSTS.SHST is set to 10b), then an NVM Subsystem Reset is be required before this bit is allowed to be set to '1'. Refer to section 7.6.New.

NOTE:

1. During a Controller Level Reset, the field values may transition to values other than the reset value prior to indicating the reset value.

Add a new section 3.1.TBD as shown below:

3.1.NEW Offset 64h: NSSD – NVM Subsystem Shutdown

This optional register provides host software with the capability to initiate a normal or abrupt NVM subsystem shutdown.

Support for this register is indicated by the state of the NVM Subsystem Shutdown Supported (CAP.NSSS) field. If the register is not supported, then the address range occupied by the register is reserved.

Figure TBD_Figure1: Offset 64h: NSSD – NVM Subsystem Shutdown

Bits	Type	Reset	Description
31:00	RW	0h	**NVM Subsystem Shutdown Control (NSSC):** A write of the value 4E726D6Ch ("Nrml") to this field initiates a normal NVM Subsystem Shutdown on every controller: • in the domain associated with the controller when CAP.CPS is set to 10b as specified in section 7.6.2.NEW_1; or • in the NVM subsystem when CAP.CPS is set to 11b in the NVM Subsystem as specified in section 7.6.2.NEW_2. A write of the value 41627077h ("Abpt") to this field initiates an abrupt NVM subsystem shutdown on every controller: • in the domain associated with the controller when CAP.CPS is set to 10b as specified in section 7.6.2.NEW_1; or • in the NVM subsystem when CAP.CPS is set to 11b in the NVM Subsystem as specified in section 7.6.2.NEW_2. A write of any other value to this field has no functional effect on the operation of the NVM subsystem. This field shall return the value 0h when read.

Modify portions of section 5.2 as shown below:

5.2 Asynchronous Event Request command

Asynchronous events are used to notify host software of status, error, and health information as these events occur. To enable asynchronous events to be reported by the controller, host software needs to submit one or more Asynchronous Event Request commands to the controller. The controller specifies an event to the host by completing an Asynchronous Event Request command. Host software should expect that the controller may not execute the command immediately; the command should be completed when there is an event to be reported.

The Asynchronous Event Request command is submitted by host software to enable the reporting of asynchronous events from the controller. This command has no timeout. The controller posts a completion queue entry for this command when there is an asynchronous event to report to the host. If Asynchronous Event Request commands are outstanding when the controller is reset, then each of those commands is aborted and should not return a CQE.

All command specific fields are reserved.

Host software may submit multiple Asynchronous Event Request commands to reduce event reporting latency. The total number of simultaneously outstanding Asynchronous Event Request commands is limited by the Asynchronous Event Request Limit specified in the Identify Controller data structure in Figure 249.

Asynchronous events are grouped into event types. The event type information is indicated in Dword 0 of the completion queue entry for the Asynchronous Event Request command. When the controller posts a completion queue entry for an outstanding Asynchronous Event Request command and thus reports an asynchronous event, subsequent events of that event type are automatically masked by the controller until the host clears that event. Unless otherwise stated, ~~A~~an event is cleared by reading the log page associated with that event using the Get Log Page command (refer to section 5.14).

The following event types are defined:

a) **Error event:** Indicates a general error that is not associated with a specific command (refer to Figure 146). To clear this event, host software reads the Error Information log (refer to section 5.14.1.1) using the Get Log Page command with the Retain Asynchronous Event bit cleared to '0';
b) **SMART / Health Status event:** Indicates a SMART or health status event (refer to Figure 147). To clear this event, host software reads the SMART / Health Information log (refer to section 5.14.1.2) using the Get Log Page command with the Retain Asynchronous Event bit cleared to '0'. The SMART / Health conditions that trigger asynchronous events may be configured in the Asynchronous Event Configuration feature using the Set Features command (refer to section 5.21);
c) **Notice event:** Indicates a general event (refer to Figure 148). To clear this event, host software reads the appropriate log page as described in Figure 148. The conditions that trigger asynchronous events may be configured in the Asynchronous Event Configuration feature using the Set Features command (see section 5.21.1.11). These notice events include:
    A. Namespace Attribute Changed;
    B. Firmware Activation Starting;
    C. Telemetry Log Changed;
    D. Asymmetric Namespace Access Change;
    E. Predictable Latency Event Aggregate Log Change;
    F. LBA Status Information Alert; and
    G. Endurance Group Event Aggregate Log Page Change;
d) **NVM Command Set Specific events:** Events that are defined by an I/O command set:
    A. **Reservation Log Page Available event:** Indicates that one or more Reservation Notification log pages (refer to section 5.14.1.16.1) are available. To clear this event, host software reads the Reservation Notification log page using the Get Log Page command with the Retain Asynchronous Event bit cleared to '0';
    B. **Sanitize Operation Completed event:** Indicates that a sanitize operation has completed (including any associated additional media modification, refer to the No-Deallocate Modifies Media After Sanitize field in Figure 249) without unexpected deallocation of all logical blocks

(refer to section 5.21.1.23) and status is available in the Sanitize Status log page (refer to section 5.14.1.16.2). To clear this event, host software reads the Sanitize Status log page using the Get Log Page command with the Retain Asynchronous Event bit cleared to '0'; and

C. **Sanitize Operation Completed With Unexpected Deallocation event:** Indicates that a sanitize operation has completed with unexpected deallocation of all LBAs (refer to section 5.21.1.23) and status is available in the Sanitize Status log page (refer to section 5.14.1.16.2). To clear this event, host software reads the Sanitize Status log page using the Get Log Page command with the Retain Asynchronous Event bit cleared to '0';

e) **Immediate events:** Events that are only reported when an outstanding Asynchronous Event Request command exists at the time the event occurs. If the event occurs and there is no outstanding Asynchronous Event Request command, then the event shall not be reported. No log page is associated with these events. These events include:

A. Normal NVM Subsystem Shutdown event;

and

f) **Vendor Specific event:** Indicates a vendor specific event. To clear this event, host software reads the indicated vendor specific log page using the Get Log Page command with the Retain Asynchronous Event bit cleared to '0'.

The Sanitize Operation Completed With Unexpected Deallocation asynchronous event shall be supported if the controller supports the Sanitize Config feature (refer to section 5.21.1.23).

Asynchronous events are reported due to a new entry being added to a log page (e.g., Error Information log) or a status update (e.g., status in the SMART / Health log). A status change may be permanent (e.g., the media has become read only) or transient (e.g., the temperature reached or exceeded a threshold for a period of time). Host software should modify the event threshold or mask the event for transient and permanent status changes before issuing another Asynchronous Event Request command to avoid repeated reporting of asynchronous events.

If an event occurs for which reporting is enabled and there are no Asynchronous Event Request commands outstanding, the controller should retain the event information for that Asynchronous Event Type and use that information as a response to the next Asynchronous Event Request command that is received. If a Get Log Page command clears the event prior to receiving the Asynchronous Event Request command or if a power off condition occurs, then a notification is not sent. If multiple events of the same type occur that have identical responses to the Asynchronous Event Request command, then those events may be reported as a single response to an Asynchronous Event Request command. If multiple events occur that are of different types, then the controller should retain a queue of those events for reporting in responses to subsequent Asynchronous Event Request commands.

5.2.1 Command Completion

A completion queue entry is posted to the Admin Completion Queue if there is an asynchronous event to report to the host. Command specific status values associated with Asynchronous Event Request are defined in Figure 145.

Figure 145: Status Code – Command Specific Status Values

Value	Description
05h	**Asynchronous Event Request Limit Exceeded:** The number of concurrently outstanding Asynchronous Event Request commands has been exceeded.

Dword 0 of the completion queue entry contains information about the asynchronous event. The definition of Dword 0 of the completion queue entry is in Figure 146.

Figure 146: Asynchronous Event Request – Completion Queue Entry Dword 0

Bits	Description
31:24	Reserved
23:16	**Log Page Identifier:** Indicates the log page associated with the asynchronous event. This log page needs to be read by the host to clear the event.

Figure 146: Asynchronous Event Request – Completion Queue Entry Dword 0

Bits	Description
15:08	**Asynchronous Event Information:** Refer to Figure 146, Figure 147, Figure 148, and Figure 149 for detailed information regarding the asynchronous event.
07:03	Reserved
02:00	**Asynchronous Event Type:** Indicates the type of the asynchronous event. More specific information on the event is provided in the Asynchronous Event Information field.

Value	Definition	Reference
000b	Error status	Figure 146
001b	SMART / Health status	Figure 147
010b	Notice	Figure 148
011b	Immediate	Figure TBD_EVENT
100b~~011b~~ to 101b	Reserved	-
110b	I/O Command Set specific status	Figure 149
111b	Vendor specific	-

The information in either Figure 147, Figure 148, Figure 149, or Figure 150 is returned in the Asynchronous Event Information field, depending on the Asynchronous Event Type.

Figure 147: Asynchronous Event Information – Error Status

Value	Description
00h	**Write to Invalid Doorbell Register:** Host software wrote the doorbell of a queue that was not created.
01h	**Invalid Doorbell Write Value:** Host software attempted to write an invalid doorbell value. Some possible causes of this error are: • the value written was out of range of the corresponding queue's base address and size; • the value written is the same as the previously written doorbell value; • the number of commands that would be added as part of a doorbell write would exceed the number of available entries; • host software attempts to add a command to a full Submission Queue; and • host software attempts to remove a completion queue entry from an empty Completion Queue.
02h	**Diagnostic Failure:** A diagnostic failure was detected. This may include a self test operation.
03h	**Persistent Internal Error:** A failure occurred that is persistent and the controller is unable to isolate to a specific set of commands. If this error is indicated, then the CSTS.CFS bit may be set to '1' and the host should perform a reset as described in section 7.3.
04h	**Transient Internal Error:** A transient error occurred that is specific to a particular set of commands; controller operation may continue without a reset.
05h	**Firmware Image Load Error:** The firmware image could not be loaded. The controller reverted to the previously active firmware image or a baseline read-only firmware image.
06h to FFh	Reserved

Figure 148: Asynchronous Event Information – SMART / Health Status

Value	Description
00h	**NVM subsystem Reliability:** NVM subsystem reliability has been compromised. This may be due to significant media errors, an internal error, the media being placed in read only mode, or a volatile memory backup device failing. This status value shall not be used if the read-only condition on the media is due to a change in the write protection state of a namespace (refer to section 8.19.1).
01h	**Temperature Threshold:** A temperature is greater than or equal to an over temperature threshold or less than or equal to an under temperature threshold (refer to section 5.21.1.4).
02h	**Spare Below Threshold:** Available spare capacity has fallen below the threshold.
03h to FFh	Reserved

Figure 149: Asynchronous Event Information – Notice

Value	Description
00h	**Namespace Attribute Changed:** The Identify Namespace data structure (refer to Figure 247) for one or more namespaces, as well as the Namespace List returned when the Identify command is issued with the CNS field set to 02h, have changed. Host software may use this event as an indication to read the Identify Namespace data structures for each namespace to determine what has changed. Alternatively, host software may request the Changed Namespace List (Log Identifier 04h) (refer to section 5.14.1.4) to determine which namespaces in this controller have changed information in the Identify Namespace data structure since the last time the log page was read. A controller shall not send this event if: a) Namespace Utilization (refer to Figure 247) has changed, as this is a frequent event that does not require action by the host; b) the ANAGRPID field (refer to Figure 247) has changed; or c) capacity information (i.e., the NUSE field and the NVMCAP field) returned in the Identify Namespace data structure (refer to Figure 247) changed as a result of an ANA state change. A controller shall only send this event for changes to the Format Progress Indicator field when bits 6:0 of that field transition from a non-zero value to 0h, or from 0h to a non-zero value.
01h	**Firmware Activation Starting:** The controller is starting a firmware activation process during which command processing is paused. Host software may use CSTS.PP to determine when command processing has resumed. To clear this event, host software reads the Firmware Slot Information log page.
02h	**Telemetry Log Changed:** The controller has saved the controller internal state in the Telemetry Controller-Initiated log page and set the Telemetry Controller-Initiated Data Available field to 1h in that log page. To clear this event, the host issues a Get Log Page command with Retain Asynchronous Event bit cleared to '0' for the Telemetry Controller-Initiated log.
03h	**Asymmetric Namespace Access Change:** The Asymmetric Namespace Access information (refer to section 5.14.1.12) related to an ANA Group that contains namespaces attached to this controller has changed (e.g., an ANA state has changed, an ANAGRPID has changed). The current Asymmetric Namespace Access information for attached namespaces is indicated in the Asymmetric Namespace Access log page (refer to section 5.14.1.12). To clear this event, the host issues a Get Log Page with the Retain Asynchronous Event bit cleared to '0' for the Asymmetric Namespace Access log. A controller shall not send this event if: a) the change is due to the creation of a namespace (refer to section 5.20); or b) the change is due to the deletion of a namespace (refer to section 5.20), as the Namespace Attribute Changed event is sent for these changes.
04h	**Predictable Latency Event Aggregate Log Change:** Indicates that event pending entries for one or more NVM Sets (refer to section 5.14.1.11) have been added to the Predictable Latency Event Aggregate log.
05h	**LBA Status Information Alert:** The criteria for generating an LBA Status Information Alert Notice event have been met (refer to section 8.22). Information about Potentially Unrecoverable LBAs is available in the LBA Status Information log page (refer to section 5.14.1.14). To clear this event, the host issues a Get Log Page command with Retain Asynchronous Event bit cleared to '0' for the LBA Status Information log.
06h	**Endurance Group Event Aggregate Log Page Change:** Indicates that event entries for one or more Endurance Groups (refer to section 5.14.1.9) have been added to the Predictable Latency Event Aggregate log. To clear this event, the host issues a Get Log Page command with the Retain Asynchronous Event bit cleared to '0' for the Endurance Group Event Aggregate log.
07h to EFh	Reserved
F0h to FFh	Refer to the NVMe over Fabrics specification

Figure 150: Asynchronous Event Information – NVM Command Set Specific Status

Value	Description
00h	**Reservation Log Page Available:** Indicates that one or more Reservation Notification log pages (refer to section 5.14.1.16.1) have been added to the Reservation Notification log.
01h	**Sanitize Operation Completed:** Indicates that a sanitize operation has completed (including any associated additional media modification, refer to the No-Deallocate Modifies Media After Sanitize field in Figure 249) without unexpected deallocation of all logical blocks (refer to section 5.21.1.23) and status is available in the Sanitize Status log page (refer to section 5.14.1.16.2).
02h	**Sanitize Operation Completed With Unexpected Deallocation:** Indicates that a sanitize operation for which No-Deallocate After Sanitize (refer to Figure 332) was requested has completed with the unexpected deallocation of all logical blocks (refer to section 5.21.1.23) and status is available in the Sanitize Status log page (refer to section 5.14.1.16.2).
03h to FFh	Reserved

Figure TBD_EVENT: Asynchronous Event Information – Immediate

Value	Description
00h	**NVM Subsystem Normal Shutdown:** This controller has started performing a normal NVM Subsystem Shutdown that is due to: • the value 4E726D6Ch ("Nrml") has been written to an NSS.NCCR register within the NVM subsystem or Domain; or • an NVMe-MI Shutdown command (refer to the NVM Express Management Interface Specification) being processed. Refer to section 7.6.NEW.
01h to FFh	Reserved

Modify figure 198 in section 5.14.1.2 as shown below:

5.14.1.2 SMART / Health Information (Log Identifier 02h)

….

Figure 198: Get Log Page – SMART / Health Information Log

Bytes	Description
...	
159:144	**Unsafe Shutdowns:** Contains the number of unsafe shutdowns. This count is incremented when ~~a Shutdown Notification (CC.SHN) is not received prior to loss of power~~ the controller does not report it is safe to power down prior to loss of main power. If CAP.CPS is set to 00b or 01b, it is safe to power down the controller when a controller shutdown processing is complete (i.e., CSTS.ST is cleared to '0' and CSTS.SHST is set to 10b). If CAP.CPS is set to 10b, it is safe to power down the domain when NVM Subsystem Shutdown processing is complete (i.e., CSTS.ST is set to '1' and CSTS.SHST is set to 10b). If CAP.CPS is set to 11b, it is safe to power down the NVM Subsystem when NVM Subsystem Shutdown processing is complete (i.e., CSTS.ST is set to '1' and CSTS.SHST is set to 10b).

<Note: This is proposed in TP 4029 as well. This definition affects both TP's. Author would like the first TP to get to 30 day member review to include it.>

Modify figure 251 in section 5.15.2.2 as shown below:

5.15.2.2 Identify Controller data structure (CNS 01h)

...

Figure 251: Identify – Identify Controller Data Structure

Bytes	O/M [1]	Description
...		
95:92	M	**Optional Asynchronous Events Supported (OAES):** This field indicates the optional asynchronous events supported by the controller. A controller shall not send optional asynchronous events before they are enabled by host software. Bits 31:~~15~~16 are reserved. BIT 15 is set to '1' if the controller supports the Normal NVM Subsystem Shutdown event. If cleared to '0', then the controller does not support the Normal NVM Subsystem Shutdown event. Bit 14 is set to '1' if the controller supports the Endurance Group Event Aggregate Log Page Change Notices event. If cleared to '0', then the controller does not support the Endurance Group Event Aggregate Log Page Change Notices event. Bit 13 is set to '1' if the controller supports the LBA Status Information Notices event. If cleared to '0', then the controller does not support the LBA Status Information Notices event. Bit 12 is set to '1' if the controller supports the Predictable Latency Event Aggregate Log Change Notices event. If cleared to '0', then the controller does not support the Predictable Latency Event Aggregate Log Change Notices event. Bit 11 is set to '1' if the controller supports sending Asymmetric Namespace Access Change Notices. If cleared to '0', then the controller does not support the Asymmetric Namespace Access Change Notices event. Bit 10 is reserved. Bit 9 is set to '1' if the controller supports the Firmware Activation Notices event. If cleared to '0', then the controller does not support the Firmware Activation Notices event. Bit 8 is set to '1' if the controller supports the Namespace Attribute Notices event and the associated Changed Namespace List log page. If cleared to '0', then the controller does not support the Namespace Attribute Notices event nor the associated Changed Namespace List log page. Bits 7:0 are reserved.

Modify figure 291 in section 5.21.1.11 as shown below:

5.21.1.11 Asynchronous Event Configuration (Feature Identifier 0Bh)

This Feature controls the events that trigger an asynchronous event notification to the host. This Feature may be used to disable reporting events in the case of a persistent condition (refer to section 5.2). If the condition for an event is true when the corresponding notice is enabled, then an event is sent to the host. The attributes are indicated in Command Dword 11.

If a Get Features command is submitted for this Feature, the attributes specified in Figure 291 are returned in Dword 0 of the completion queue entry for that command.

Figure 291: Asynchronous Event Configuration – Command Dword 11

Bits	Description
…	
15	**Normal NVM Subsystem Shutdown:** This bit determines whether an asynchronous event notification is sent to the host when the NVM subsystem has started performing a normal shutdown due to an NVM Subsystem Shutdown (refer to Figure TBD_EVENT). If this bit is set to '1', then the Normal NVM Subsystem Shutdown event is sent to the host if an outstanding Asynchronous Event Request command exists at the time this condition occurs. If this bit is cleared to '0', then the controller shall not send the Normal NVM Subsystem Shutdown event to the host.
14	**Endurance Group Event Aggregate Log Change Notices:** This bit determines whether an asynchronous event notification is sent to the host when an event entry for an Endurance Group (refer to section 8.17) has been added to the Endurance Group Event Aggregate log (refer to section 5.14.1.15). If this bit is set to '1', then the Endurance Group Event Aggregate Log Change event is sent to the host when this condition occurs. If this bit is cleared to '0', then the controller shall not send the Endurance Group Event Aggregate Log Change event to the host. If Endurance Groups are not supported and this bit is set to '1', then the Set Features command shall be aborted with a status of Invalid Field in Command.

Modify section 6.4.2.1 as shown below:

6.4.2.1 AWUPF/NAWUPF Example (Informative)

…

After a write command has completed, reads for that location which are subsequently submitted shall return the data from that write command and not an older version of the data from previous write commands with the following exception;

If all of the following conditions are met:

a) the controller supports a volatile write cache;
b) the volatile write cache is enabled;
c) the FUA bit for the write is not set;
d) no flush commands, associated with the same namespace as the write, successfully completed before the controller reports shutdown complete (CSTS.SHST set to 10b); and
e) main power loss occurs on a controller ~~shutdown occurs~~ without completing the normal or abrupt shutdown procedure outlined in section 7.6.2,

then subsequent reads for locations written to the volatile write cache that were not written to non-volatile storage may return older data.

Modify section 7.1.2 as shown below:

Administrative Controller

An administrative controller is a controller whose intended purpose is to provide NVM subsystem management capabilities. While an I/O controller may support these same management capabilities, an administrative controller has fewer mandatory capabilities. Unlike an I/O controller, an administrative controller does not support commands that provide access to logical block data and metadata stored on an NVM subsystem's non-volatile storage medium. This prevents a host managing an NVM subsystem using an administrative controller from accessing user data. Finally, an administrative controller has a dedicated PCI programming interface value (refer to CC.PI field) allowing a dedicated NVMe management driver to be loaded instead of a generic NVMe driver.

Examples of management capabilities that may be supported by an administrative controller include the following.

- Ability to efficiently poll NVM subsystem health status via NVMe-MI using the NVMe-MI Send and NVMe-MI Receive commands;
- Ability to manage an NVMe enclosure via NVMe-MI using the NVMe-MI Send and NVMe-MI Receive commands;
- Ability to manage NVM subsystem namespaces using the Namespace Attachment and Namespace commands;
- Ability to perform virtualization management using the Virtualization Management command; ~~and~~
- Ability to reset an entire NVM subsystem using the NVM Subsystem Reset (NSSR) register; and
- Ability to shutdown an entire NVM subsystem using the NVM Subsystem Shutdown (NSSD) register.

Modify section 7.6.2 as shown below:

7.6.2 Controller Shutdown

It is recommended that the host perform an orderly shutdown of the controller by following the procedure in this section when a power-off or shutdown condition is imminent.

The host should perform the following actions in sequence for a normal controller shutdown:

1. If the controller is enabled (i.e., CC.EN is set to '1'):
    a. Stop submitting any new I/O commands to the controller and allow any outstanding commands to complete;
    b. If the controller implements I/O queues, then the host should delete all I/O Submission Queues, using the Delete I/O Submission Queue command. A result of the successful completion of the Delete I/O Submission Queue command is that any remaining commands outstanding are aborted;
    c. If the controller implements I/O queues, then the host should delete all I/O Completion Queues, using the Delete I/O Completion Queue command; and
2. The host should set the Shutdown Notification (CC.SHN) field to 01b to indicate a normal controller shutdown operation. The controller indicates when shutdown processing is completed by updating the Shutdown Status (CSTS.SHST) field to 10b and the Shutdown Type (CSTS.ST) field is cleared to '0'.

~~For entry to the D3 power state, the shutdown steps outlined for a normal shutdown should be followed.~~

The host should perform the following actions in sequence for an abrupt controller shutdown:

1. If the controller is enabled (i.e., CC.EN is set to '1'), then ~~S~~stop submitting any new I/O commands to the controller; and
2. The host should set the Shutdown Notification (CC.SHN) field to 10b to indicate an abrupt shutdown operation. The controller indicates when shutdown processing is completed by updating the Shutdown Status (CSTS.SHST) field to 10b and CSTS.ST is cleared to '0'.

For entry to the D3 power state, the shutdown steps outlined for a normal controller shutdown should be followed.

It is recommended that the host wait a minimum of the RTD3 Entry Latency reported in the Identify Controller data structure for the shutdown operations to complete; if the value reported in RTD3 Entry Latency is 0h, then the host should wait for a minimum of one second. It is not recommended to disable the controller via the CC.EN field. This causes a Controller Reset which may impact the time required to complete shutdown processing.

I~~I~~t is safe to power off the controller when CSTS.ST is cleared to '0', and CSTS.SHST indicates controller shutdown processing is complete (regardless of the value of CC.EN). It remains safe to power off the controller until CC.EN transitions from '0' to '1'.

To start executing commands on the controller after that controller reports ~~a~~ controller shutdown processing complete ~~operation~~ (i.e., CSTS.ST is cleared to '0' and CSTS.SHST is set to 10b) utilizing CC.EN:~~,~~

- If CC.EN is set to '1', then a Controller Reset (i.e., CC.EN cleared from '1' to '0') is required on that controller; or
- If CC.EN is cleared to '0', then the controller is required to be enabled (i.e., CC.EN is set to '1' from '0').

The initialization sequence should then be executed on that controller.

It is an implementation choice whether the host aborts all outstanding commands to the Admin Queue prior to the controller shutdown. The only commands that should be outstanding to the Admin Queue ~~at~~ when the controller reports shutdown processing complete are Asynchronous Event Request commands.

Add section 7.6.NEW as shown below:

7.6.NEW NVM Subsystem Shutdown

An NVM Subsystem Shutdown initiates a shutdown of all controllers in a domain or NVM subsystem from a single controller.

7.6.2.NEW_1 NVM Subsystem Shutdown in a Single Domain NVM Subsystem

A normal shutdown on all controllers within the NVM subsystem (i.e., normal NVM Subsystem Shutdown) is initiated by:

- a host writing the value 4E726D6Ch ("Nrml") to NSAS.NSSC when CAP.CPS is set to 11b; or
- issuing an NVMe-MI Shutdown command to a Management Endpoint (refer to the NVM Express Management Interface Specification) specifying a normal shutdown.

For each controller in the NVM subsystem for this normal NVM Subsystem Shutdown, if:

- CSTS.SHST is set to 00b; and
- An outstanding Asynchronous Event Request command exists,

then the controller shall issue a Normal NVM Subsystem Shutdown event prior to shutting down the controller.

An abrupt shutdown on all controllers within the NVM subsystem (i.e., abrupt NVM Subsystem Shutdown) is initiated by:

- a host writing the value 41627077h ("Abpt") to NSAS.NSSC when CAP.CPS is set to 11b; or
- issuing an NVMe-MI Shutdown command to a Management Endpoint (refer to the NVM Express Management Interface Specification) specifying an abrupt shutdown.

For either a normal shutdown or an abrupt shutdown, it is safe to power off the NVM subsystem when CSTS.ST is set to '1' and CSTS.SHST indicates shutdown processing complete (i.e., CSTS.SHST is set to 10b) on any controller in the NVM subsystem. It remains safe to power off the NVM subsystem until an NVM Subsystem Reset occurs.

If a normal or abrupt NVM Subsystem Shutdown is being processed or completed within the NVM subsystem (i.e., CSTS.ST is set to '1' and CSTS.SHST is set to 01b or 10b on all controllers in the NVM Subsystem), then:

- an NVM Subsystem Reset clears CSTS.SHST to 00b in all controllers in the NVM Subsystem; and
- any other type of Controller Level Reset has no effect on the processing of that shutdown.

7.6.2.NEW_2 Domain Shutdown in a Multiple Domain NVM Subsystem

A normal shutdown on this controller and all controllers within the associated domain is initiated by:

- a host writing the value 4E726D6Ch ("Nrml") to NSAS.NSSC when CAP.CPS is set to 10b; or
- issuing an NVMe-MI Shutdown command to a Management Endpoint (refer to the NVM Express Management Interface Specification) specifying a normal shutdown.

For each controller in the domain for this normal NVM subsystem shutdown, if:

- CSTS.SHST is cleared to 00b; and
- An outstanding Asynchronous Event Request command exists,

then the controller shall issue a Normal NVM Subsystem Shutdown event prior to shutting down the controller.

An abrupt shutdown to this controller and all controllers within the associated domain is initiated by:

- a host writing the value 41627077h ("Abpt") to NSAS.NSSC when CAP.CPS is set to 10b; or
- issuing an NVMe-MI Shutdown command to a Management Endpoint (refer to the NVM Express Management Interface Specification) specifying an abrupt shutdown.

For either a normal shutdown or an abrupt shutdown on the domain, it is safe to power off the domain when CSTS.ST is set to '1' and CSTS.SHST indicates shutdown processing complete (i.e., CSTS.SHST is set to 10b) on any controller in the domain. It remains safe to power off the domain until an NVM Subsystem Reset occurs on that domain.

If a normal or abrupt Domain Shutdown is being processed or completed within a domain (i.e., CSTS.ST is set to '1' and CSTS.SHST is set to 01b or 10b on all controllers in the domain), then:

- an NVM Subsystem Reset clears CSTS.SHST to 00b in all controllers in the Domain; and
- any other type of Controller Level Reset has no effect on the processing of that shutdown.

Modify section 8.5.3 as shown below:

8.5.3 Secondary Controller States and Resource Configuration

...

A primary controller or secondary controller is enabled when CC.EN and CSTS.RDY are both set to '1' for that controller. A secondary controller is able to be enabled only when in the Online state. If the primary controller associated with a secondary controller is disabled or undergoes a Controller Level Reset, then the secondary controller shall implicitly transition to the Offline state. A secondary controller shall transition to the Offline state when a shutdown occurs (refer to section 3.1.5 and section 3.1.NEW) on the primary controller associated with that secondary controller.

Modify portions of NVMe-MI 1.1b as shown below:

Modify a portion of Figure 55, Figure 56, and Figure 57 in section 5 as shown below:

5 Management Interface Command Set

…

Figure 55: Opcodes for Management Interface Command Set

Opcode	Command
00h	Read NVMe-MI Data Structure
01h	NVM Subsystem Health Status Poll
02h	Controller Health Status Poll
03h	Configuration Set
04h	Configuration Get
05h	VPD Read
06h	VPD Write
07h	Reset
08h	SES Receive
09h	SES Send
0Ah	Management Endpoint Buffer Read
0Bh	Management Endpoint Buffer Write
0Ch	Shutdown
0Dh~~0C~~h to BFh	Reserved
C0h to FFh	Vendor specific

…

Figure 56: Management Interface Command Set Support using an Out-of-Band Mechanism

NVMe Storage Device O/M/P[1]	NVMe Enclosure O/M/P[1]	Command
M	M	Read NVMe-MI Data Structure
M	O[3]	NVM Subsystem Health Status Poll
M	O[3]	Controller Health Status Poll
M	M[2]	Configuration Set
M	M[2]	Configuration Get
M	O[3]	VPD Read
M	O[3]	VPD Write
M	O[3]	Reset
P	M	SES Receive
P	M	SES Send
O	O[3]	Shutdown
O	M	Management Endpoint Buffer Read
O	M	Management Endpoint Buffer Write
O	O	Vendor specific

NOTES:
1. O/M/P definition: O = Optional, M = Mandatory, P = Prohibited from being supported. An NVMe Enclosure that is also an NVMe Storage Device (i.e., implements namespaces) shall implement mandatory commands required by either an NVMe Storage Device or an NVMe Enclosure and may implement optional commands allowed by either an NVMe Storage Device or an NVMe Enclosure.
2. This command was architected for an NVMe Storage Device. The mapping of Health Status Change Configuration Identifier to an NVMe Enclosure is outside the scope of this specification.
3. This command was architected for an NVMe Storage Device. The mapping of this command to an NVMe Enclosure is outside the scope of this specification.

…

Figure 57: Management Interface Command Set Support using In-Band Tunneling Mechanism

NVMe Storage Device		NVMe Enclosure		Command
O/M/P[1]	NVMe-MI Send/Receive Mapping[3]	O/M/P[1]	NVMe-MI Send/Receive Mapping[3]	
M	NVMe-MI Receive	O[2]	NVMe-MI Receive	Read NVMe-MI Data Structure
M	NVMe-MI Receive	O[2]	NVMe-MI Receive	NVM Subsystem Health Status Poll
M	NVMe-MI Receive	O[2]	NVMe-MI Receive	Controller Health Status Poll
M	NVMe-MI Send	O[2]	NVMe-MI Send	Configuration Set
M	NVMe-MI Receive	O[2]	NVMe-MI Receive	Configuration Get
M	NVMe-MI Receive	O[2]	NVMe-MI Receive	VPD Read
M	NVMe-MI Send	O[2]	NVMe-MI Send	VPD Write
M	NVMe-MI Send	O[2]	NVMe-MI Send	Reset
P	n/a	M	NVMe-MI Receive	SES Receive
P	n/a	M	NVMe-MI Send	SES Send
P	n/a	P	n/a	Management Endpoint Buffer Read
P	n/a	P	n/a	Management Endpoint Buffer Write
O	NVMe-MI Send	O[2]	NVMe-MI Send	Shutdown
O	Vendor Specific	O	Vendor Specific	Vendor specific

NOTES:
1. O/M/P definition: O = Optional, M = Mandatory, P = Prohibited from being supported. An NVMe Enclosure that is also an NVMe Storage Device (i.e., implements namespaces) shall implement mandatory commands required by either an NVMe Storage Device or an NVMe Enclosure and may implement optional commands allowed by either an NVMe Storage Device or an NVMe Enclosure.
2. This command was architected for an NVMe Storage Device. The mapping of this command to an NVMe Enclosure is outside the scope of this specification.
3. This column indicates whether the NVMe-MI Command is tunneled in-band using the NVMe-MI Send or NVMe-MI Receive command.

<Editor – when updating this figure please use a font size 14 on all note numbers in the figure like shown for the NVMe-MI Shutdown command. NVMe has adopted this as the size is too small. This may cause cells to have to be centered.>

Modify figure 77 in section 5.3 as shown below:

5.3 Controller Health Status Poll

…

Figure 77: Controller Health Data Structure (CHDS)

Bytes	Description
01:00	**Controller Identifier (CTLID):** This field specifies the Controller Identifier with which the data contained in this data structure is associated.

Figure 77: Controller Health Data Structure (CHDS)

<table>
<tr><th>Bytes</th><th>Description</th></tr>
<tr><td>03:02</td><td>Controller Status (CSTS): This field reports the Controller status.
<table>
<tr><th>Bit</th><th>Reset</th><th>Description</th></tr>
<tr><td>15:08</td><td>0</td><td>Reserved</td></tr>
<tr><td>07</td><td>HwInit</td><td>Firmware Activated (FA): This bit is set to ‘1’ when a new firmware image is activated. Firmware activation is described in the NVM Express specification.

The reset value of this bit is set to ‘1’ if a reset caused a new firmware image to be activated.</td></tr>
<tr><td>06</td><td>0</td><td>Namespace Attribute Changed (NAC): This bit is set to ‘1’ under the same conditions that causes the Namespace Attribute Changed asynchronous event to be sent if Namespace Attribute Notices are enabled as specified in the NVM Express specification. This bit may be set to ‘1’ regardless of whether Namespace Attribute Notices are enabled or not.</td></tr>
<tr><td>05</td><td>0</td><td>Controller Enable Change Occurred (CECO): This bit is set to ‘1’ when the Enable bit (refer to CC.EN in the NVM Express specification) changes state.</td></tr>
<tr><td>04</td><td>HwInit</td><td>NVM Subsystem Reset Occurred (NSSRO): This bit corresponds to the value of the NVM Subsystem Reset Occurred (refer to CSTS.NSSRO in the NVM Express specification) bit.</td></tr>
<tr><td>03:02</td><td>00b</td><td>Shutdown Status (SHST): This field corresponds to the value of the Shutdown Status (refer to CSTS.SHST in the NVM Express specification) field.</td></tr>
<tr><td>01</td><td>HwInit</td><td>Controller Fatal Status (CFS): This bit corresponds to the value of the Controller Fatal Status (refer to CSTS.CFS in the NVM Express specification) bit.</td></tr>
<tr><td>00</td><td>0</td><td>Ready (RDY): This bit corresponds to the value of the Ready (refer to CSTS.RDY in the NVM Express specification) bit.</td></tr>
</table></td></tr>
<tr><td colspan="2">…</td></tr>
</table>

Modify portions of section 5.8 as shown below:

5.8 Reset

The Reset command may be used to initiate a reset.

The Reset command uses NVMe Management Dword 0. The format of NVMe Management Dword 0 is shown in Figure 99. All other command specific fields in the Request Message and Response Message are reserved.

Figure 99: Reset - NVMe Management Dword 0

<table>
<tr><th>Bit</th><th>Description</th></tr>
<tr><td>31:24</td><td>1. Reset Type: This field specifies the type of reset to be performed.
<table>
<tr><th>Value</th><th>O/M[1]</th><th>Description</th></tr>
<tr><td>00h</td><td>~~O~~/M[2]</td><td>Reset NVM Subsystem</td></tr>
<tr><td>01h to FFh</td><td>-</td><td>Reserved</td></tr>
</table></td></tr>
<tr><td>23:00</td><td>Reserved</td></tr>
<tr><td colspan="2">NOTES:
1. O/M definition: O = Optional, M = Mandatory</td></tr>
</table>

Figure 99: Reset - NVMe Management Dword 0

Bit	Description
2.	~~The Reset Type for Reset NVM Subsystem is r~~Required if the NVM Subsystem Reset feature is supported in-band as defined in the NVM Express specification; else, it is optional.

When a Reset command that specifies a Reset NVM Subsystem in the Reset Type field ~~is~~ completed~~s~~ successfully, the NVM Subsystem Reset is immediately initiated (refer to section 9.3). No Success Response is transmitted.

A Management Controller should shutdown all NVMe Controllers in an NVM Subsystem prior to resetting the NVM Subsystem. Refer to the Shutdown command in section 5.TBD.

Add a new section 5.TBD as shown below:

5.TBD Shutdown

The Shutdown command sent to one Management Endpoint initiates a shutdown on all Controllers in the NVM Subsystem.

The Shutdown command uses NVMe Management Dword 0. The format of NVMe Management Dword 0 is shown in Figure FIG_TBD. All other command specific fields in the Request Message and Response Message are reserved.

Figure FIG_TBD: Shutdown - NVMe Management Dword 0

<table>
<tr><th>Bit</th><th>Description</th></tr>
<tr><td>31:24</td><td>Shutdown Type: This field specifies the type of shutdown to be performed.
<table>
<tr><th>Value</th><th>O/M[1]</th><th>Description</th></tr>
<tr><td>00h</td><td>O[2]</td><td>Normal NVM Subsystem Shutdown</td></tr>
<tr><td>01h</td><td>O[2]</td><td>Abrupt NVM Subsystem Shutdown</td></tr>
<tr><td>02h to FFh</td><td>-</td><td>Reserved</td></tr>
</table></td></tr>
<tr><td>23:00</td><td>Reserved</td></tr>
<tr><td colspan="2">NOTES:
1. O/M definition: O = Optional, M = Mandatory
2. Mandatory for the out-of-band mechanism if the NVM Subsystem Shutdown feature is supported on all NVMe Controllers in the NVM Subsystem (refer to the NVM Express Base Specification).</td></tr>
</table>

Upon receipt of a Shutdown command specifying a Normal NVM Subsystem Shutdown, then: for each Controller in the NVM Subsystem:

- if:
  - CSTS.SHST is cleared to 00b on that Controller; and
  - An outstanding Asynchronous Event Request command exists on that Controller (refer to the NVM Express Base Specification),

  then the Controller shall issue a Normal NVM Subsystem Shutdown event prior to shutting down the Controller (refer to the NVM Express Base Specification);
- a normal shutdown is initiated on the Controller as specified by the NVM Express Base Specification.

Upon receipt of a Shutdown command specifying an Abrupt NVM Subsystem Shutdown, then for each Controller in the NVM subsystem an abrupt shutdown is initiated as specified by the NVM Express Base Specification.

The Shutdown command completes successfully when all NVMe controllers in the NVM Subsystem report shutdown process complete (i.e., CSTS.SHST is set to 10b). Refer to the NVMe Express Base Specification on the condition when it is safe to power down the NVM Subsystem.

Update figure 116 in section 6.3 as shown below:

6.3 Sanitize Operation

Figure 116 specifies the Command Messages allowed during a sanitize operation. Refer to the NVM Express specification for the definition of a sanitize operation.

Figure 116: Command Messages Allowed During Sanitize Operation

Command Set	Command Message	Allowed During Sanitize Operation [1]
Management Interface Command Set	Configuration Get	Yes
	Configuration Set	
	Controller Health Status Poll	
	Management Endpoint Buffer Read	
	Management Endpoint Buffer Write	
	NVM Subsystem Health Status Poll	
	Read NVMe-MI Data Structure	
	Reset	
	SES Receive	
	SES Send	
	Shutdown	
	VPD Read	
	VPD Write	
NVMe Admin Command Set [2]	Device Self-test	Same restrictions as defined by the NVM Express specification
	Firmware Activate/Commit	
	Firmware Image Download	
	Format NVM	
	Get Features	
	Get Log Page	
	Identify	
	Namespace Attachment	
	Namespace Management	
	Sanitize	
	Security Receive/Send	
	Security Send	
	Set Features	
	Vendor Specific	
	Virtualization Management	
PCIe Command Set	PCIe Configuration Read	Yes
	PCIe Configuration Write	
	PCIe I/O Read	
	PCIe Memory Read	
	PCIe Memory Write	

NOTES:
1. Refer to the NVM Express specification for the definition of a sanitize operation.
2. NVMe Admin Commands that are prohibited via the out-of-band mechanism (refer to Figure 110) are not listed since they are always prohibited including during a sanitize operation.