NVM Express Technical Proposal for New Feature

Technical Proposal ID	4074 Defining Scope for Features
Change Date	2020-09-14
Builds on Specification	**NVM Express Base Specification 2.0** **NVM Express Command Set Specification 1.0** **NVM Express Key Value Command Set Specification 1.0**
References Specification	

Technical Proposal Author(s)

Name	Company
Andres Baez	Intel
Mike Allison	Intel

This proposal intends to do the following:
- Define a clear scope for each Feature;
- Remove ambiguity on Features affected by NVM subsystem with multiple controllers

Revision History

Revision Date	Change Description
2020-01-06	Initial version
2020-05-14	Added a table for the scoping of power and thermals. Added the features from other technical proposals. Updated section wording to clarify the scope.
2020-05-21	Changes scoping for Autonomous Power State Transition and Temperature Threshold.
2021-03-30	Aligned TP with NVMe 2.0
2021-04-16	Reverted Power features to controller scope per feedback. Added footnote to address unspecified behavior for a Subsystem with multiple controllers requesting different power scopes.
2021-05-13	• Revised wording providing clarifications • Reverted Power Management & Autonomous Power Transition scope wording as this is to be defined in TPAR 4121 • Moved FSP footnote from FID table into Base Spec Feature Identifiers Supported and Effects log page.
2021-05-27	• Updated wording in Feature Values section per feedback.

2021-09-01	• Integrated into the NVM Express Base Specification revision 2.0, the NVM Express Command Set Specification revision 1.0, and the NVM Express Key Value Command Set Specification revision 1.0
2021-09-14	• Capitalize "endurance group"

Description for NVMe 2.0 Changes Document

This technical proposal defines the scope of each Feature where the scoping is implied. In addition, Features that require the option of multiple scoping requirements are defined.

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 NVM Express Base Specification 2.0 (version dated 04/22/2021) as shown below:

Modify portions of section 4.2 as shown below:

4.2   Feature Values

Feature settings ~~may~~ apply to ~~may apply to:~~ the scope as defined in Figure 316.~~.~~:

> ~~a)   the controller (i.e., the feature is not namespace specific); or~~
> ~~b)   a namespace (i.e., the feature is namespace specific).~~

For feature values that apply to the controller scope:

- a) if the NSID field is cleared to 0h or set to FFFFFFFFh, then:
  - the Set Features command shall set the specified feature value for the controller; and
  - the Get Features command shall return the current setting of the requested feature value for the controller;

  and
- b) if the NSID field is set to a valid namespace identifier (refer to section 3.2.1.2), then:
  - the Set Features command shall abort with a status code of Feature Not Namespace Specific; and
  - the Get Features command shall return the current setting of the requested feature value for the controller.

For feature values that apply to ~~a~~the namespace scope:

- a) if the NSID field is set to an active namespace identifier (refer to section 3.2.1.4), then:
  - the Set Features command shall set the specified feature value of the specified namespace; and
  - the Get Features command shall return the current setting of the requested feature value for the specified namespace;
- b) if the NSID field is set to FFFFFFFFh, then:

  - for the Set Features command, the controller shall:
    - o if the MDS bit is set to '1' in the Identify Controller data structure, abort the command with Invalid Field in Command; or
    - o if the MDS bit is cleared to '0' in the Identify Controller data structure, unless otherwise specified, set the specified feature value for all namespaces attached to the controller processing the command;

    and
  - for the Get Features command, the controller shall, unless otherwise specified in section 5.27.1, abort the command with a status code of Invalid Namespace or Format;

    and
- c) if the NSID field is set to any other value, then the Set Features command and the Get Features command ~~shall~~is abort~~ed by the controller~~ as described ~~in the rules for namespace identifier usage~~ in Figure 87.

For feature values that apply to NVM subsystems, NVM Sets, or Endurance Groups scope:

- a) the NSID field should be cleared to 0h; and
- b) if the NSID field is set to a non-zero value, then the Set Features command and the Get Features command are aborted by the controller as described in Figure 87.

Modify portions of section 5.16.1.18 as shown below:

5.16.1.18 Feature Identifiers Supported and Effects (Log Identifier 12h)

An NVM subsystem may support several interfaces for submitting a Get Log Page command such as an Admin Submission Queue, PCIe VDM Management Endpoint, or SMBus/I2C Management Endpoint (refer the NVM Express Management Interface Specification for details on Management Endpoints) and may

have zero or more instances of each of those interfaces. The feature identifiers (FIDs) supported on each instance of each interface may be different. This log page describes the FIDs that are supported on the interface to which the Get Log Page command was submitted and the effects of those features on the state of the NVM subsystem. The log page is defined in Figure 255. Each Feature Identifier's effects are described in a FID Supported and Effects data structure defined in Figure 256.

If the Feature Identifiers Supported and Effects log page is supported and the Feature is supported, then the scope, as defined in Figure 316, shall be indicated in the FID Scope field (FSP) for that Feature (refer to Figure 256).

The features that the controller supports are dependent on the I/O Command Set that is based on:
- the I/O Command Set selected in CC.CSS, if CC.CSS is not set to 110b; and
- the Command Set Identifier (CSI) field in CDW 14, if CC.CSS is set to 110b.

If CC.CSS is set to 110b, I/O Command Sets that have not been enabled by the I/O Command Set Profile (FID 19h) (refer to section 5.27.1.21) are treated as unsupported.

Figure 255: Get Log Page – Feature Identifiers Effects Log

Bytes	Description
03:00	**Feature Identifier Supported 0:** Contains the FID Supported and Effects data structure (refer to Figure 256) for FID 0h.
07:04	**Feature Identifier Supported 1:** Contains the FID Supported and Effects data structure (refer to Figure 256) for FID 1h.
…	…
1019:1016	**Feature Identifier Supported 254:** Contains the FID Supported and Effects data structure (refer to Figure 256) for FID FEh.
1023:1020	**Feature Identifier Supported 255:** Contains the FID Supported and Effects data structure (refer to Figure 256) for FID FFh.

The FID Supported and Effects data structure describes the effect of a Set Features command for the FID, including any optional features of the FID.

If multiple hosts are connected to the NVM subsystem, then those hosts should coordinate their commands to meet the Command Submission and Execution requirements (refer to Figure 256). The details of this coordination are outside the scope of this specification.

Figure 256: Get Log Page – FID Supported and Effects Data Structure

Bits	Description
31:20	**FID Scope (FSP):** This field defines the scope for the associated feature identifier. If the value of this field is 0h, then no scope is reported. If this field is non-zero, then only one bit shall be set to '1'. <table><tr><th>Bits</th><th>Description</th></tr><tr><td>11:6</td><td>Reserved</td></tr><tr><td>5</td><td>**NVM Subsystem Scope**: If set to '1', then modifying the value of the FID may impact the whole NVM subsystem. If cleared to '0' and the FSP field is non-zero, then modifying the value of the FID does not impact the whole NVM subsystem.</td></tr><tr><td>4</td><td>**Domain Scope**: If set to '1', then modifying the value of the FID may impact a single Domain. If cleared to '0' and the FSP field is non-zero, then modifying the value of the FID does not impact a single Domain.</td></tr><tr><td>3</td><td>**Endurance Group Scope**: If set to '1', then modifying the value of the FID may impact Endurance Groups. If cleared to '0' and the FSP field is non-zero, then modifying the value of the FID does not impact Endurance Groups.</td></tr><tr><td>2</td><td>**NVM Set Scope**: If set to '1', then modifying the value of the FID may impact NVM Sets. If cleared to '0' and the FSP field is non-zero, then modifying the value of the FID does not impact NVM Sets.</td></tr><tr><td>1</td><td>**Controller Scope**: If set to '1', then modifying the value of the FID may impact the controller. If cleared to '0' and the FSP field is non-zero, then the FID does not have controller scope.</td></tr><tr><td>0</td><td>**Namespace Scope**: If set to '1', then modifying the value of the FID may impact namespaces. If cleared to '0' and the FSP field is non-zero, then modifying the value of the FID does not impact namespaces.</td></tr></table>

Modify portions of section 5.27 as shown below:

5.27.1 Feature Specific Information

Figure 316: Set Features – Feature Identifiers

Feature Identifier	Current Setting Persists Across Power Cycle and Reset [2]	Uses Memory Buffer for Attributes	Feature Name	Scope[6]
00h	Reserved			
01h	No	No	Arbitration	Controller
02h	No	No	Power Management	Controller[7]
04h	No	No	Temperature Threshold	Controller
06h	No	No	Volatile Write Cache	Controller
07h	No	No	Number of Queues	Controller
08h	No	No	Interrupt Coalescing	Controller
09h	No	No	Interrupt Vector Configuration	Controller
0Bh	No	No	Asynchronous Event Configuration	Controller
0Ch	No	Yes	Autonomous Power State Transition	Controller[7]
0Dh	No [3]	No [4]	Host Memory Buffer	Controller
0Eh	No	Yes	Timestamp	Controller
0Fh	No	No	Keep Alive Timer	Controller
10h	Yes	No	Host Controlled Thermal Management	Controller
11h	No	No	Non-Operational Power State Config	Controller
12h	Yes	No	Read Recovery Level Config	NVM Set / NVM subsystem
13h	No	Yes	Predictable Latency Mode Config	NVM Set
14h	No	No	Predictable Latency Mode Window	NVM Set
16h	No	Yes	Host Behavior Support	Controller
17h	Yes	No	Sanitize Config	NVM subsystem
18h	No	No	Endurance Group Event Configuration	Endurance Group
19	Yes	No	I/O Command Set Profile	Controller
1Ah	Yes	No	Spinup Control	NVM subsystem
1Bh to 1Fh	Reserved			
20h			Command Set Specific	
21h to 77h	Reserved			
78h to 7Fh	Reserved for Management Features			
7Dh	No	Yes	Enhanced Controller Metadata	Controller
7Eh	No	Yes	Controller Metadata	Controller
7Fh	No	Yes	Namespace Metadata	Namespace per controller
80h	Yes	No	Software Progress Marker	Controller
81h	No	Yes	Host Identifier	Controller
82h	No	No	Reservation Notification Mask	Namespace
83h	Yes	No	Reservation Persistence	Namespace
84h	No	No	Namespace Write Protection Config	Namespace
85h to BFh			Reserved	

Figure 316: Set Features – Feature Identifiers

Feature Identifier	Current Setting Persists Across Power Cycle and Reset [2]	Uses Memory Buffer for Attributes	Feature Name	Scope[6]
C0h to FFh			Vendor Specific[1, 5]	

NOTES:
1. The behavior of a controller in response to an inactive namespace ID to a vendor specific Feature Identifier is vendor specific.
2. This column is only valid if the feature is not saveable (refer to section 4.2). If the feature is saveable, then this column is not used.
3. The controller does not save settings for the Host Memory Buffer feature across power states and reset events, however, host software may restore the previous values. Refer to section 8.9.
4. The feature does not use a memory buffer for Set Features commands and does use a memory buffer for Get Features commands. Refer to section 8.9.
5. Selection of a UUID may be supported. Refer to section 8.25.
6. Refer to Feature Identifiers Supported and Effects log page in section 5.16.1.18 for how scope is reported to the host.
7. For NVM Subsystems with multiple controllers in the same domain, specifying different power states results in an unspecified power state for that domain.

5.27.1.1 Arbitration (Feature Identifier 01h)

This Feature controls command arbitration within the controller. Refer to section 3.4.4 for command arbitration details. The attributes are specified in Command Dword 11.

5.27.1.2 Power Management (Feature Identifier 02h)

This Feature allows the host to configure the controller power state. The attributes are specified in Command Dword 11 (refer to Figure 318).

Upon successful completion of a Set Features command for this fFeature, the controller shall be in the Power State specified. For a transition to a non-operational power state, the device may exceed the power indicated for that non-operational power state as defined in section 8.15.1 (e.g., while completing this command). If enabled, autonomous power state transitions continue to occur from the new state.

If a Get Features command is submitted for this Feature, the attributes described in Figure 319 are returned in Dword 0 of the completion queue entry for that command.

Figure 318: Power Management – Command Dword 11

Bits	Description
31:08	Reserved
07:05	**Workload Hint (WH):** This field indicates the type of workload expected. This hint may be used by the NVM subsystem to optimize performance. Refer to section 8.15.3 for more details.
04:00	**Power State (PS):** This field indicates the new power state into which the controller is requested to transition. This power state shall be one supported by the controller as indicated in the Number of Power States Supported (NPSS) field in the Identify Controller data structure. If the power state specified is not supported, the controller shall abort the command and should return an error of Invalid Field in Command.

5.27.1.3 Temperature Threshold (Feature Identifier 04h)

A primary controller may report up to nine temperature values in the SMART / Health Information log (i.e., the Composite Temperature and Temperature Sensor 1 through Temperature Sensor 8; refer to Figure 207). Associated with each implemented temperature sensor is an over temperature threshold and an under temperature threshold. When a temperature is greater than or equal to its corresponding over temperature

threshold or less than or equal to its corresponding under temperature threshold, then bit one of the Critical Warning field in the SMART / Health Information Log (refer to section 5.16.1.3) is set to one. This may trigger an asynchronous event.

5.27.1.4 Volatile Write Cache (Feature Identifier 06h), (Optional)

This Feature controls the volatile write cache, if present, on the controller. If a volatile write cache is present (refer to the VWC field in Figure 275), then this fFeature shall be supported. The attributes are specified in Command Dword 11.

Note: If the controller is able to guarantee that data present in a write cache is written to non-volatile media on loss of power, then that write cache is considered non-volatile and this fFeature does not apply to that write cache.

5.27.1.5 Number of Queues (Feature Identifier 07h)

This Feature indicates the number of queues that the host requests for the controller processing the command. This fFeature shall only be issued during initialization prior to creation of any I/O Submission and/or Completion Queues. If a Set Features command is issued for this fFeature after creation of any I/O Submission and/or I/O Completion Queues, then the Set Features command shall abort with status code of Command Sequence Error. The controller shall not change the value allocated between resets. For a Set Features command, the attributes are specified in Command Dword 11 (refer to Figure 322). For a Get Features command, Dword 11 is ignored.

5.27.1.6 Interrupt Coalescing (Feature Identifier 08h)

This Feature configures controller interrupt coalescing settings. The controller should signal an interrupt when either the Aggregation Time or the Aggregation Threshold conditions are met. If either the Aggregation Time or the Aggregation Threshold fields are cleared to 0h, then an interrupt may be generated (i.e., interrupt coalescing is implicitly disabled). This Feature applies only to the I/O Queues. It is recommended that interrupts for commands that complete in error are not coalesced. The settings are specified in Command Dword 11.

5.27.1.7 Interrupt Vector Configuration (Feature Identifier 09h)

This Feature configures settings specific to a particular controller interrupt vector. The settings are specified in Command Dword 11.

5.27.1.8 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 by the controller 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 specified in Command Dword 11.

5.27.1.9 Autonomous Power State Transition (Feature Identifier 0Ch), (Optional)

This fFeature configures the settings for autonomous controller power state transitions, refer to section 8.15.2.

5.27.1.10 Host Memory Buffer (Feature Identifier 0Dh), (Optional)

This Feature controls use of the Host Memory Buffer by the controller. The attributes are specified in Command Dword 11, Command Dword 12, Command Dword 13, Command Dword 14, and Command Dword 15.

5.27.1.11 Timestamp (Feature Identifier 0Eh), (Optional)

This Feature~~Timestamp feature~~ enables the host to set a timestamp value in the controller. A controller indicates support for the Timestamp feature through the Optional NVM Command Support (ONCS) field in the Identify Controller data structure. The Timestamp field value (refer to Figure 339) in a Set Features command sets a timestamp value in the controller. After the current value for this fFeature is set, the controller updates that value as time passes. A Get Features command that requests the current value reports the timestamp value in the controller at the time the Get Features command is processed (e.g., the value set with a Set Features command for the current value plus the elapsed time since being set).

5.27.1.12 Keep Alive Timer (Feature Identifier 0Fh)

This Feature ~~controls~~configures the controller Keep Alive Timer. Refer to section 3.9 for Keep Alive details. The attributes are specified in Command Dword 11.

5.27.1.13 Host Controlled Thermal Management (Feature Identifier 10h), (Optional)

This fFeature configures the controller settings for the host controlled thermal management feature, refer to section 8.15.5. The host controlled thermal management feature uses Command Dword 11 with the attributes shown in Figure 342.

If a Get Features command is submitted for this fFeature, then the attributes shown in Figure 342 are returned in Dword 0 of the completion queue entry for that command.

This fFeature is not namespace specific.

5.27.1.15 Read Recovery Level Config (Feature Identifier 12h)

This Feature is used to configure the Read Recovery Level (refer to section 8.17). The attributes are specified in Command Dword 11 and Command Dword 12. Modifying the Read Recovery Level has no effect on the data contained in any associated namespace.

The scope of this Feature is:

- NVM Set if NVM Sets are supported; or
- NVM Subsystem if NVM Sets are not supported.

5.27.1.20 Endurance Group Event Configuration (Feature Identifier 18h), (Optional)

This Feature controls the events that trigger adding an Endurance Group Event Aggregate Log Change Notices event to the Endurance Group Event Aggregate log for the specified Endurance Group. 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 specified in Command Dword 11.

5.27.1.22 Spinup Control (Feature Identifier 1Ah)

This Feature allows the host to configure the method for initial spinup for Endurance Groups that store data on rotational media (refer to section 8.20). ~~The NVM Subsystem is the scope of this feature~~.

5.27.1.23.1 Enhanced Controller Metadata (Feature Identifier 7Dh)

This fFeature is used to store enchanced controller metadata about the host platform in an NVM subsystem for later retrieval.

The metadata element types defined in Figure 369 are used by this fFeature.

5.27.1.23.2 Controller Metadata (Feature Identifier 7Eh)

This fFeature is used to store controller metadata about the host platform in an NVM subsystem for later retrieval.

…

The metadata element types defined in Figure 369 are used by this fFeature.

…

5.27.1.23.3 Namespace Metadata (Feature Identifier 7Fh)

This fFeature is used to store metadata about a namespace associated with a controller in the NVM subsystem for later retrieval. This fFeature is namespace specific and controller specific. The Add Entry Multiple action is prohibited for this fFeature.

5.27.1.24 Software Progress Marker (Feature Identifier 80h), (Optional)

This Feature is a controller software progress marker. The software progress marker is persistent across power states. This information may be used to indicate to an OS software driver whether there have been issues with the OS successfully loading. The attributes are specified in Command Dword 11.

5.27.1.25 Host Identifier (Feature Identifier 81h), (Optional[1])

This fFeature allows the host to register a Host Identifier with the controller. The Host Identifier is used by the controller to determine whether other controllers in the NVM subsystem are associated with the same host. The Host Identifier may be used to designate host elements that access an NVM subsystem independently of each other or for reservations.

---

[1] Mandatory if reservations are supported by the controller as indicated in the ONCS field in the Identify Controller data structure.

Modify portions of the NVM Express Command Set Specification 1.0 (version dated 2021.04.22) as shown below

Modify portions of section 4.1.3

4.1.3   Get Features & Set Features commands

Figure 78 defines the Features support requirements for I/O Controllers supporting the NVM Command Set.

Figure 78: Feature Identifiers – NVM Command Set

Feature Identifier	Persistent Across Power Cycle and Reset[1]	Uses Memory Buffer for Attributes	Description	Scope
03h	Yes	Yes	LBA Range Type	Namespace
05h	No	No	Error Recovery	Namespace
0Ah	No	No	Write Atomicity Normal	Controller
15h	No	No	LBA Status Information Report Interval	Controller

NOTES:
1.   This column is only valid if the feature is not saveable (refer to the NVME Base specification). If the feature is saveable, then this column is not used and any feature may be configured to be saved across power cycles and reset.

4.1.3.5   Write Atomicity Normal (Feature Identifier 0Ah)
This Feature ~~controls~~configures the controller operation of the AWUN and NAWUN parameters (refer to section 2.1.4.1). The attributes are specified in Command Dword 11.

Modify portions of the NVM Express Key Value Command Set Specification 1.0 (version dated 2021.03.11) as shown below

Modify portions of section 4.1.3

4.1.3 Get Features & Set Features commands

Figure 32 defines the Features support requirements for I/O Controllers supporting the Key Value Command Set.

Figure 32: Feature Identifiers – Key Value Command Set

Feature Identifier	Persistent Across Power Cycle and Reset[1]	Uses Memory Buffer for Attributes	Description	Scope
20h	Yes	No	Key Value Configuration	Namespace
NOTES: 1. This column is only valid if the feature is not saveable (refer to the NVME Base specification). If the feature is saveable, then this column is not used and any feature may be configured to be saved across power cycles and reset.				