LEGAL NOTICE:

This technical proposal is proprietary to the NVM Express, Inc. (also referred to as "Company") and/or its successors and assigns.

NOTICE TO USERS WHO ARE NVM EXPRESS, INC. MEMBERS: Members of NVM Express, Inc. have the right to use and implement this technical proposal subject, however, to the Member's continued compliance with the Company's Intellectual Property Policy and Bylaws and the Member's Participation Agreement.

NOTICE TO NON-MEMBERS OF NVM EXPRESS, INC.: If you are not a Member of NVM Express, Inc. and you have obtained a copy of this document, you only have a right to review this document or make reference to or cite this document. Any such references or citations to this document must acknowledge NVM Express, Inc. copyright ownership of this document. The proper copyright citation or reference is as follows: **"© 2008 to 2023 NVM Express, Inc. ALL RIGHTS RESERVED."** When making any such citations or references to this document you are not permitted to revise, alter, modify, make any derivatives of, or otherwise amend the referenced portion of this document in any way without the prior express written permission of NVM Express, Inc. Nothing contained in this document shall be deemed as granting you any kind of license to implement or use this document or the specification described therein, or any of its contents, either expressly or impliedly, or to any intellectual property owned or controlled by NVM Express, Inc., including, without limitation, any trademarks of NVM Express, Inc.

LEGAL DISCLAIMER:

THIS DOCUMENT AND THE INFORMATION CONTAINED HEREIN IS PROVIDED ON AN "**AS IS**" BASIS. TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, NVM EXPRESS, INC. (ALONG WITH THE CONTRIBUTORS TO THIS DOCUMENT) HEREBY DISCLAIM ALL REPRESENTATIONS, WARRANTIES AND/OR COVENANTS, EITHER EXPRESS OR IMPLIED, STATUTORY OR AT COMMON LAW, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE, VALIDITY, AND/OR NONINFRINGEMENT.

All product names, trademarks, registered trademarks, and/or servicemarks may be claimed as the property of their respective owners.

The NVM Express® design mark is a registered trademark of NVM Express, Inc.

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	**4104a HMB with Low Power Support**
Change Date	**2023-10-31**
Builds on Specification	**NVMe Base Specification, Revision 2.0a**
Refers to Ratified Technical Proposals	**ECN112**

Technical Proposal Author(s)

Name	**Company**
Yoni Shternhell	Western Digital, Inc.
Judy Brock. Mike Allison	Samsung
David Black	Dell EMC
Gerry Houlder	Seagate Technology

This proposal adds a mechanism to advise the controller about the transitioning of the host to non-operational low power state. This allows the controller to limit the access to HMB only on non-operational power state initiated by the host, and for the host to not change or deallocate the HMB already in use.

4104a adds changes to Annex TBD, which were originally proposed for ECN 112.

Revision History

Revision Date	Change Description
2021-02-09	Initial version
2021-03-16	Editorial changes in the feature definition
2021-04-06	Add a support bit in Controller Attributes
2021-04-28	- Several editorial changes - Remove the 'set explicitly by the host' text
2021-04-29	Editorial changes, changes of 'limit' text to 'prohibit' text
2021-05-12	Added clarifying text in section 8.4.1. (Non-Operational Power States)
2021-05-13	More editorial changes in 8.4.1
2021-05-27	- Re-writing text in section 5.21.1.13 and in the NAP field definition. - Edits in section 8.4.1
2021-08-24	- Added support for NACP field - Removed prohibition to access HMB if entered due to APST
2021-09-15	- Changes in the text on the behavior of the NAP field at non-operational state - Adding new field Non-Operational Access Currently Prohibited (NACP)
2021-09-20	- Aligned with NVMe Base specification 2.0a
2021-09-22	- Major rewrite, including use of restrict instead of prohibit in names and complete rework of CQE Dword 0 contents.
2021-09-30	- Updates from Mike Allison, Judy Brock and technical WG meeting
2021-10-07	- Minor editorial changes from the technical WG meeting - Added acronym for the HMB Restrict Non-Operational Power State Access bit
2021-10-26	- Editorial member review comments
2021-11-18	- Clean version for integration
2021-12-09	- Minor editorial edit at the TP description
2022-01-11	- Integration
2022-01-12	- Changed the overview to indicate non-operational power state.
2023-03-17	- Started 4104a, which adds Annex C to NVM Express 2.0a.
2023-03-23	- Aligned to NVMe 2023 template. Removed an "and" Aligned to changes requested at 3/23 Technical meeting.
2023-10-31	- Integrated

Description for Changes Document for NVM Express® Base Specification, Revision 2.0a

New Features/Feature Enhancements/Required Changes:

- Identify Controller Data Structure
  - Description of change.
    - New support field in the Controller Attribute
- Host Memory Buffer
    - New text to specify the HMB access restrictions in operational power states.
    - New enable bit in the Host Memory Buffer – Command Dword 11
    - 2 new bits in the Host Memory Buffer – Completion Queue Entry Dword 0
  - References
    - Technical Proposal 4104

Markup Conventions:

Black:	Unchanged (however, hot links are removed)
~~Red Strikethrough~~:	Deleted
Blue:	New
Highlighted:	TBD values, anchors, and links to be inserted
Green Bracketed:	Notes to editor
Orange:	Changes from another ECN or TP

Modify a portions of Figure 275 (Identify – Identify Controller Data Structure) as shown below:

5.17.2.1 Identify Controller data structure (CNS 01h)

…

Figure 275: Identify – Identify Controller Data Structure

<table>
<tr><th>Bytes</th><th>O/M [1]</th><th>Description</th></tr>
<tr><td colspan="3">…</td></tr>
<tr><td colspan="3">Controller Capabilities and Features</td></tr>
<tr><td colspan="3">…</td></tr>
<tr><td>99:96</td><td>M</td><td>Controller Attributes (CTRATT): This field indicates attributes of the controller.
<table>
<tr><th>Bits</th><th>Description</th></tr>
<tr><td>31:18</td><td>Reserved</td></tr>
<tr><td>17</td><td>HMB Restrict Non-Operational Power State Access (HMBR): If set to ‘1’, then the controller supports restricting HMB access in non-operational power states as defined in section 5.27.1.10. If cleared to ‘0’, then the controller does not support restricting HMB access in non-operational power states as defined by section 5.27.1.10.</td></tr>
</table></td></tr>
</table>

Figure 275: Identify – Identify Controller Data Structure

Bytes	O/M [1]	Description	
		15	**Extended LBA Formats Supported (ELBAS):** If set to '1' indicates that the controller supports the I/O command set specific extended protection information formats (refer to the Protection Information Formats section of the applicable I/O command set specification). If cleared to '0' indicates that the controller does not support the I/O command set specific extended protection information formats (refer to the Protection Information Formats section of the NVM Command Set Specification). Refer to the LBA Format Extension Enable (LBAFEE) field in the Host Behavior Support feature (refer to section 5.27.1.18) for details for host software to enable the controller to operate on namespaces using the protection information formats. NOTE: This bit field applies to all I/O Command Sets. The original name has been retained for historical continuity.
		14	**Delete NVM Set:** If set to '1', then the controller supports the Delete NVM Set operation (refer to section 8.3.3). If cleared to '0', then the controller does not support the Delete NVM Set operation.
		13	**Delete Endurance Group:** If set to '1', then the controller supports the Delete Endurance Group operation (refer to section 8.3.3). If cleared to '0', then the controller does not support the Delete Endurance Group operation.
		12	**Variable Capacity Management:** If set to '1', then the controller supports Variable Capacity Management (refer to section 8.3.3). If cleared to '0', then the controller does not support Variable Capacity Management.
		11	**Fixed Capacity Management**: If set to '1', then the controller supports Fixed Capacity Management (refer to section 8.3.2). If cleared to '0', then the controller does not support Fixed Capacity Management.
		10	**Multi-Domain Subsystem (MDS):** If set to '1', then the NVM subsystem supports the multiple domains (refer to section 3.2.4). If cleared to '0', then the NVM subsystem does not support the reporting of multiple domains and the NVM subsystem consists of a single domain.
		9	**UUID List:** If set to '1', then the controller supports reporting of a UUID List (refer to Figure 284). If cleared to '0', then the controller does not support reporting of a UUID List (refer to section 8.25).
		8	**SQ Associations:** If set to '1', then the controller supports SQ Associations (refer to section 8.22). If cleared to '0', then the controller does not support SQ Associations.
		7	**Namespace Granularity:** If set to '1', then the controller supports reporting of Namespace Granularity (refer to section 5.17.2.15). If cleared to '0', the controller does not support reporting of Namespace Granularity. If the Namespace Management capability (refer to section 8.11) is not supported, then this bit shall be cleared to '0'.
		6	**Traffic Based Keep Alive Support (TBKAS):** If set to '1', then the controller supports restarting the Keep Alive Timer if an Admin command or an I/O command is processed during the Keep Alive Timeout Interval (refer to section 3.9.2). If cleared to '0', then the controller supports restarting the Keep Alive Timer only if a Keep Alive command is processed during the Keep Alive Timeout Interval (refer to section 3.9.1).
		5	**Predictable Latency Mode:** If set to '1', then the controller supports Predictable Latency Mode (refer to section 8.16). If cleared to '0', then the controller does not support Predictable Latency Mode.
		4	**Endurance Groups:** If set to '1', then the controller supports Endurance Groups (refer to section 3.2.3). If cleared to '0', then the controller does not support Endurance Groups.

Figure 275: Identify – Identify Controller Data Structure

Bytes	O/M [1]	Description
		3 — **Read Recovery Levels:** If set to '1', then the controller supports Read Recovery Levels (refer to section 8.17). If cleared to '0', then the controller does not support Read Recovery Levels. 2 — **NVM Sets:** If set to '1', then the controller supports NVM Sets (refer to section 3.2.2). If cleared to '0', then the controller does not support NVM Sets. 1 — **Non-Operational Power State Permissive Mode:** If set to '1', then the controller supports host control of whether the controller may temporarily exceed the power of a non-operational power state for the purpose of executing controller initiated background operations in a non-operational power state (i.e., Non-Operational Power State Permissive Mode supported). If cleared to '0', then the controller does not support host control of whether the controller may exceed the power of a non-operational state for the purpose of executing controller initiated background operations in a non-operational state (i.e., Non-Operational Power State Permissive Mode not supported). Refer to section 5.27.1.14. 0 — **Host Identifier Support:** If set to '1', then the controller supports a 128-bit Host Identifier. Bit 0 if cleared to '0', then the controller does not support a 128-bit Host Identifier.
...		

Make the following addition to section 5.27.1.10 (Host Memory Buffer) *as shown below:*

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

...

After a successful completion of a Set Features command that disables the host memory buffer, the controller shall not access any data in the host memory buffer until the host memory buffer has been enabled. The controller should retrieve any necessary data from the host memory buffer in use before posting the completion queue entry for the Set Features command that disables the host memory buffer. Posting of the completion queue entry for the Set Features command that disables the host memory buffer acknowledges that it is safe for the host software to modify the host memory buffer contents. Refer to section 8.9.

A host is able to restrict access to the host memory buffer (HMB) while the controller is in a non-operational power state that was configured by the host (refer to section 5.27.1.2). If this HMB non-operational power state access restriction is enabled by the host (refer to Figure 330) and the host configures a non-operational power state, then the controller does not access the HMB until the controller transitions to an operational power state except for HMB access required to process Admin commands and background operations initiated by Admin commands. Enabling or disabling Non-Operational Power State Permissive Mode (refer to section 5.27.1.14) shall have no effect on HMB non-operational power state access restriction.

Enabling or disabling HMB non-operational power state access restriction should not affect the Entry Latency (ENLAT) for non-operational power states (refer to section 8.15) that are reported in the power state descriptors in Identify Controller data structure (e.g., if HMB non-operational power state access restriction is enabled, the controller may consume additional time beyond the applicable Entry Latency value in order to retrieve necessary data from the HMB before the controller transitions to a non-operational power state).

If HMB non-operational power state access restriction is enabled and the controller autonomously transitions from an operational power state to a non-operational power state, then HMB access by the controller is not restricted and that access should be minimized (e.g., access ceases as soon as possible after that transition and does not resume until after the controller transitions to an operational power state).

If HMB non-operational power state access restriction is enabled and the host configures a non-operational power state while the controller is in a non-operational power state, then HMB access by the controller is restricted.

If a Get Features command is issued for this Feature, then the completion queue entry indicates whether HMB non-operational power state access restriction is enabled and whether HMB non-operational power state access restriction is currently restricting controller access to the HMB (refer to Figure 337):

Figure 330: Host Memory Buffer – Command Dword 11

Bits	Description
31:04~~02~~	Reserved
03	This bit shall be cleared to '0'.
02	**Host Memory Non-operational Access Restriction Enable (HMNARE):** If the HMBR bit is set to '1' in the Controller Attributes (CTRATT) field in the Identify Controller data structure and: • this bit is set to '1', then HMB non-operational power state access restriction shall be enabled (i.e., the controller shall not access the HMB after a non-operational power state is configured by the host (refer to section 5.27.1.2) until the controller is in an operational power state except for access required to process Admin commands and background operations initiated by Admin commands); and • this bit is cleared to '0', then HMB non-operational power state access restriction shall be disabled (i.e., the controller may access the HMB while the controller is in any non-operational power state). If this bit set to '1' and the HMBR bit is cleared to '0', then the controller shall abort the command with a status code of Invalid Field in Command. If this bit is cleared to '0' and the HMBR bit is cleared to '0', then this bit has no effect.
01	**Memory Return (MR):** If set to '1', then the host is returning memory previously allocated to the controller for use as the host memory buffer (HMB). That memory may have been in use for the HMB prior to a reset or entering the Runtime D3 state (e.g., prior to the HMB being disabled). A returned host memory buffer shall have the exact same size, descriptor list address, descriptor list contents, and host memory buffer contents as last seen by the controller before the host memory buffer was disabled (i.e., a Set Features command with the EHM bit cleared to '0' was processed). If cleared to '0', then the host is allocating host memory resources with undefined content.
00	**Enable Host Memory (EHM):** If set to '1', then the host memory buffer shall be enabled and the controller may use the host memory buffer. If cleared to '0', then the host memory buffer shall be disabled, and the controller shall not use the host memory buffer. If a Set Features command is processed with this bit cleared to '0', then the controller shall ignore Command Dword 12, Command Dword 13, Command Dword 14, and Command Dword 15.

Figure 337: Host Memory Buffer – Completion Queue Entry Dword 0

Bits	Description
31:~~01~~ 04	Reserved
03	**Host Memory Non-operational Access Restricted (HMNAR):** If set to '1', then HMB non-operational power state access restriction is currently restricting the controller from accessing the HMB. If cleared to '0', then HMB non-operational power state access restriction is not currently restricting the controller from accessing the HMB (e.g., the HMB is not enabled, the controller is currently in an operational power state, HMB non-operational power state access restriction is not supported, or HMB non-operational power state access restriction is not enabled).

Figure 337: Host Memory Buffer – Completion Queue Entry Dword 0

Bits	Description
02	**Host Memory Non-operational Access Restriction Enable (HMNARE):** If set to '1', then HMB non-operational power state access restriction is enabled (refer to the HMNARE bit description in Figure 330). If cleared to '0', then HMB non-operational power state access restriction is not enabled.
01	This bit is not used for a Get Feature command and shall be cleared to '0'.
00	**Enable Host Memory (EHM):** If set to '1', then the host memory buffer is enabled and the controller may use the host memory buffer. If cleared to '0', then the host memory buffer is disabled, and the controller is not using the host memory buffer.

8.15.1 Non-Operational Power States

…

For all of the cases in the preceding paragraph, the controller shall:

- logically remain in the current non-operational power state unless an I/O command is received or if an explicit transition is requested by a Set Features command with the Power Management identifier; and
- not exceed the maximum power advertised for the most recent operational power state.

HMB non-operational power state access restriction (refer to section 5.27.1.10) does not prohibit the controller from accessing the HMB while processing Admin commands and while performing host-initiated background operations initiated by Admin commands. HMB non-operational power state access restriction does prohibit the controller from accessing the HMB in order to perform controller-initiated activity (i.e. activity not directly associated with a command).

…

<Reader's Note: The following is from ECN112>

Annex TBD. Power Management and Consumption (Informative)

…

Controller thermal management may cause a transition to a lower power state, interacting with these Features:

a) Temperature Threshold (refer to section 5.27.1.3); ~~and~~
b) Host Controlled Thermal Management (see section 5.21.1.3 and section 8.15.5)~~.~~; and
c) access to host memory buffer (refer to section 5.27.1.10) may be prohibited in non-operational power state.

Reporting mechanisms for NVM Express power management include:

a) this property:
   - Controller Power Scope (CAP.CPS) (refer to Figure 36);
b) these fields in the Identify Controller data structure (refer to Figure 275):
   - RTD3 Resume Latency (RTD3R);
   - RTD3 Entry Latency (RTD3E);
   - Non-Operational Power State Permissive Mode;
   - Number of Power States Support (NPSS);
   - Autonomous Power State Transition Attributes (APSTA); and
   - Power State 0 Descriptor (PSD0) through Power State 31 Descriptor (PSD31) (refer to Figure 276);
c) Feature Identifiers:

- Power Management (refer to section 5.27.1.2);
- Temperature Threshold (refer to section 5.27.1.3);
- Autonomous Power State Transition (refer to section 5.27.1.9 and section 8.15.4);
- Non-Operational Power State Configuration (refer to section 5.27.1.14 and section 8.15.3);
- Host Controlled Thermal Management (see section 5.21.1.3 and section 8.15.5); ~~and~~
- Host Memory Buffer (refer to section 5.27.1.10); and
- Spinup Control (refer to section 5.27.1.22);

and

d) log pages:

…