LEGAL NOTICE:

This NVM Express Management Interface revision 1.0a 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 NVM Express Management Interface revision 1.0a 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: **"© 2007 - 2018 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 service marks may be claimed as the property of their respective owners.

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

NVM Express Technical Proposal for New Feature

Technical Proposal ID	**6007 – Add Missing VPD and Basic features**
Change Date	**04/16/2018**
Builds on Specification	**NVM Express Management Interface 1.0a**

Technical Proposal Author(s)

Name	Company
Myron Loewen	Intel Corporation

Update to PCIe 4.0 specification and add new standard form factors. These changes will all go into existing reserved fields to avoid growing any structures larger or incur backwards compatibility issues. In addition, the Basic command from Appendix A is getting its first change to help fan speed controllers without MCTP do a better job managing temperature.

Revision History

Revision Date	Change Description
03/02/2018	First draft based on prior discussions for PCIe Gen 4 and Basic Temperatures
03/22/2018	Added BGA and EDSFF form factors, report power in Basic command 0 offset 6
03/26/2018	Improved the Power State text in the Basic command
04/16/2018	Fixed SMBus NACKs, added SFF-TA-1002, then Passed vote for approval

Description of Specification Changes

Modify Figure 64: PCIe Port Specific Data as shown below for PCIe Gen 4:

<table>
<tr><td>09</td><td>PCIe Supported Link Speeds Vector: This field indicates the Supported Link Speeds for the specified PCIe port.

<table>
<tr><th>Bit</th><th>Description</th></tr>
<tr><td>7:~~43~~</td><td>Reserved</td></tr>
<tr><td>3</td><td>This bit shall be set to ‘1’ if the link supports 16.0 GT/s</td></tr>
<tr><td>2</td><td>This bit shall be set to ‘1’ if the link supports 8.0 GT/s</td></tr>
<tr><td>1</td><td>This bit shall be set to ‘1’ if the link supports 5.0 GT/s</td></tr>
<tr><td>0</td><td>This bit shall be set to ‘1’ if the link supports 2.5 GT/s.</td></tr>
</table></td></tr>
<tr><td>10</td><td>PCIe Current Link Speed: The port’s PCIe negotiated link speed using the same encoding as the PCIe Supported Link Speed Vector field. A value of 0h in this field indicates the PCIe Link is not available.

<table>
<tr><th>Value</th><th>Definition</th></tr>
<tr><td>0h</td><td>Link not active</td></tr>
<tr><td>1h</td><td>The current link speed is the speed indicated in the supported link speed bit 0.</td></tr>
<tr><td>2h</td><td>The current link speed is the speed indicated in the supported link speed bit 1.</td></tr>
<tr><td>3h</td><td>The current link speed is the speed indicated in the supported link speed bit 2.</td></tr>
<tr><td>4h</td><td>The current link speed is the speed indicated in the supported link speed bit 3.</td></tr>
<tr><td>5h</td><td>The current link speed is the speed indicated in the supported link speed bit 4.</td></tr>
<tr><td>6h</td><td>The current link speed is the speed indicated in the supported link speed bit 5.</td></tr>
<tr><td>7h</td><td>The current link speed is the speed indicated in the supported link speed bit 6.</td></tr>
<tr><td>8h-FFh</td><td>Reserved</td></tr>
</table></td></tr>
</table>

Modify 9.2.3 NVMe MultiRecord Area for new Form Factors:

<table>
<tr><td>06</td><td>Impl Spec</td><td>Management Endpoint Form Factor (MEFF): This field indicates the form factor of the Management Endpoint.

<table>
<tr><th>Value</th><th>Definition</th></tr>
<tr><td>0</td><td>Other – unknown</td></tr>
<tr><td>1 – 15</td><td>Reserved</td></tr>
<tr><td>16</td><td>2.5” Form Factor – unknown</td></tr>
<tr><td>17</td><td>2.5” Form Factor – U.2 (SFF-8639) 15mm</td></tr>
</table></td></tr>
</table>

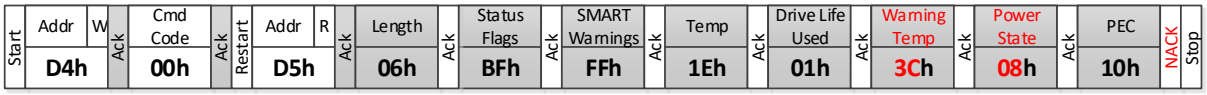
			18	2.5” Form Factor – U.2 (SFF-8639) 7mm
			19	2.5” Form Factor – (SFF-TA-1001) 15mm
			20	2.5” Form Factor – (SFF-TA-1001) 7mm
			21~~19~~ - 31	Reserved
			32	CEM add in card – unknown
			33	CEM add in card – Low Profile (HHHL)
			34	CEM add in card – Standard Height Half Length (FHHL)
			35	CEM add in card – Standard Height Full Length (FHFL)
			36-47	Reserved
			48	M.2 module – unknown
			49	M.2 module – 2230
			50	M.2 module – 2242
			51	M.2 module – 2260
			52	M.2 module – 2280
			53	M.2 module – 22110
			54-63	Reserved
			64	BGA SSD – unknown
			65	BGA SSD – 16 x 20mm (M.2 Type 1620)
			66	BGA SSD – 11.5 x 13mm (M.2 Type 1113)
			67-79	Reserved
			80	Enterprise & Datacenter SSD Form Factor – unknown
			81	1U Short Form Factor - (SFF-TA-1006) 5.9mm
			82	1U Short Form Factor - (SFF-TA-1006) 8mm
			83	1U Long Form Factor - (SFF-TA-1007) 9.5mm
			84	1U Long Form Factor - (SFF-TA-1007) 18mm
			85	3” Short Form Factor - (SFF-TA-1008) 7.5mm
			86	3” Short Form Factor - (SFF-TA-1008) 16.8mm
			87	3” Long Form Factor - (SFF-TA-1008) 7.5mm
			88	3” Long Form Factor - (SFF-TA-1008) 16.8mm
			89~~65~~-239	Reserved
			240-255	Vendor Specific

Modify 9.2.4 NVMe PCIe Port MultiRecord Area as shown below for PCIe Gen 4:

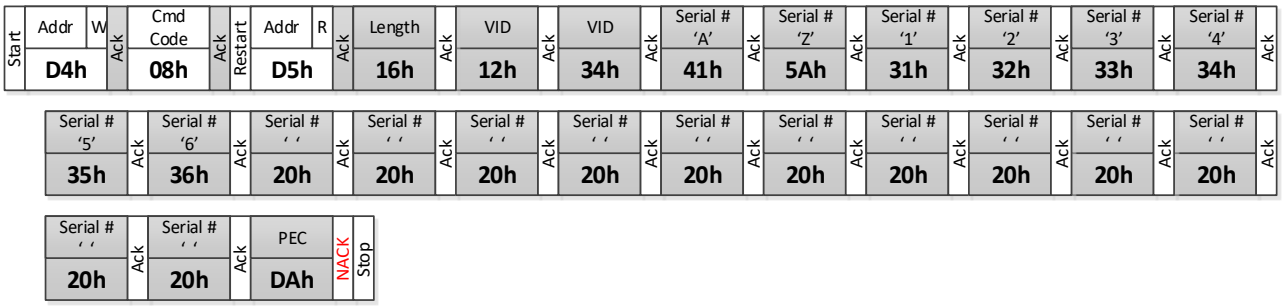
08	Impl Spec	**PCIe Link Speed:** This field indicates a bit vector of link speeds supported by the PCIe port.

Bit	Definition
7:~~43~~	Reserved
3	Set to ‘1’ if the PCIe link supports 16.0 GT/s. Otherwise cleared to ‘0’.
2	Set to ‘1’ if the PCIe link supports 8.0 GT/s. Otherwise cleared to ‘0’.
1	Set to ‘1’ if the PCIe link supports 5.0 GT/s. Otherwise cleared to ‘0’.
0	Set to ‘1’ if the PCIe link supports 2.5 GT/s. Otherwise cleared to ‘0’.

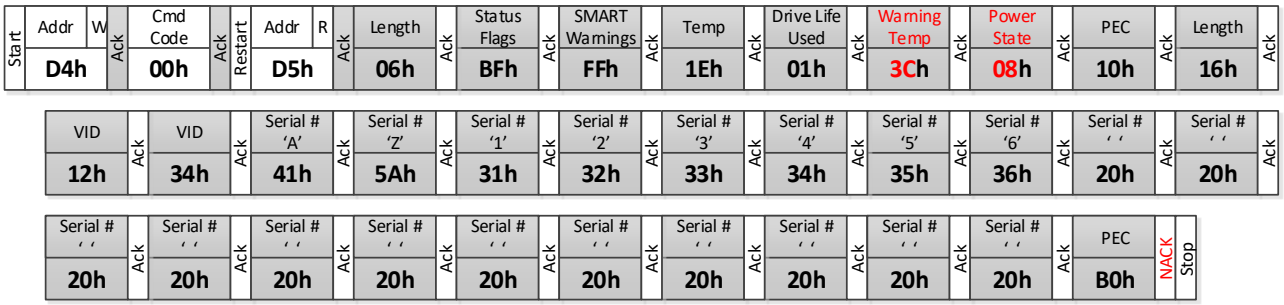
Modify Appendix A Example 1 as shown below for Basic temperature thresholds and NACK field:



Modify Appendix A Example 2 as shown below for the NACK field:



Modify Appendix A Example 4 as shown below for Basic temperature thresholds and NACK:



Modify Appendix A Figure 112 as shown below for Basic temperature thresholds:

Figure 1: Subsystem Management Data Structure

Command Code	Offset (byte)	Description
0	00	**Length of Status:** Indicates number of additional bytes to read before encountering PEC. This value should always be 6 (06h) in implementations of this version of the spec.

<table>
<tr><th>Command Code</th><th>Offset (byte)</th><th>Description</th></tr>
<tr><td></td><td>01</td><td>Status Flags (SFLGS): This field indicates the status of the NVM Subsystem.

SMBus Arbitration – Bit 7 is set ‘1’ after an SMBus block read is completed all the way to the stop bit without bus contention and cleared to ‘0’ if an SMBus Send Byte FFh is received on this SMBus slave address.

Drive Not Ready – Bit 6 is set to ‘1’ when the subsystem is not capable of processing NVMe management commands, and the rest of the transmission may be invalid. If cleared to ‘0’, then the NVM Subsystem is fully powered and ready to respond to management commands. This logic level intentionally identifies and prioritizes powered up and ready drives over their powered off neighbors on the same SMBus segment.

Drive Functional – Bit 5 is set to ‘1’ to indicate an NVM Subsystem is functional. If cleared to ‘0’, then there is an unrecoverable failure in the NVM Subsystem and the rest of the transmission may be invalid. Note that this bit may default to ‘0’ after reset and transition to ‘1’ after the NVM Subsystem has completed initialization and this case should not be considered an error.

Reset Not Required - Bit 4 is set to ‘1’ to indicate the NVM Subsystem does not need a reset to resume normal operation. If cleared to ‘0’ then the NVM Subsystem has experienced an error that prevents continued normal operation. A Controller Level Reset is required to resume normal operation.

Port 0 PCIe Link Active - Bit 3 is set to ‘1’ to indicate the first port’s PCIe link is up (i.e., the Data Link Control and Management State Machine is in the DL_Active state). If cleared to ‘0’, then the PCIe link is down.

Port 1 PCIe Link Active - Bit 2 is set to ‘1’ to indicate the second port’s PCIe link is up. If cleared to ‘0’, then the second port’s PCIe link is down or not present.

Bits 1-0 shall be set to ‘1’.</td></tr>
<tr><td></td><td>02</td><td>SMART Warnings: This field shall contain the Critical Warning field (byte 0) of the NVMe SMART / Health Information log. Each bit in this field shall be inverted from the NVMe definition (i.e., the management interface shall indicate a ‘0’ value while the corresponding bit is ‘1’ in the log page). Refer to the NVMe specification for bit definitions.

If there are multiple Controllers in the NVM Subsystem, the management endpoint shall combine the Critical Warning field from every Controller such that a bit in this field is:
Cleared to ‘0’ if any Controller in the subsystem indicates a critical warning for that corresponding bit.
Set to ‘1’ if all Controllers in the NVM Subsystem do not indicate a critical warning for the corresponding bit.</td></tr>
<tr><td></td><td>03</td><td>Composite Temperature (CTemp): This field indicates the current temperature in degrees Celsius. If a temperature value is reported, it should be the same temperature as the Composite Temperature from the SMART log of hottest Controller in the NVM Subsystem. The reported temperature range is vendor specific, and shall not exceed the range -60 to +127°C. The 8 bit format of the data is shown below.

This field should not report a stale temperature, which means that it was sampled more ~~when that is older~~ than 5 seconds prior. If recent data is not available, the Management Endpoint should indicate a value of 80h for this field.

<table>
<tr><th>Value</th><th>Description</th></tr>
<tr><td>00h-7Eh</td><td>Temperature is measured in degrees Celsius (0 to 126C)</td></tr>
<tr><td>7Fh</td><td>127C or higher</td></tr>
<tr><td>80h</td><td>No temperature data or temperature data is more the 5 seconds old.</td></tr>
<tr><td>81h</td><td>Temperature sensor failure</td></tr>
<tr><td>82h-C3h</td><td>Reserved</td></tr>
</table></td></tr>
</table>

<table>
<tr><th>Command Code</th><th>Offset (byte)</th><th>Description</th></tr>
<tr><td></td><td></td><td>
<table>
<tr><td>C4</td><td>Temperature is -60C or lower</td></tr>
<tr><td>C5-FFh</td><td>Temperature measured in degrees Celsius is represented in two’s complement (-1 to -59C)</td></tr>
</table>
</td></tr>
<tr><td></td><td>04</td><td>Percentage Drive Life Used (PDLU): Contains a vendor specific estimate of the percentage of NVM Subsystem NVM life used based on the actual usage and the manufacturer’s prediction of NVM life. If an NVM Subsystem has multiple Controllers the highest value is returned. A value of 100 indicates that the estimated endurance of the NVM in the NVM Subsystem has been consumed, but may not indicate an NVM Subsystem failure. The value is allowed to exceed 100. Percentages greater than 254 shall be represented as 255. This value should be updated once per power-on hour and equal the Percentage Used value in the NVMe SMART Health Log Page.</td></tr>
<tr><td></td><td>05:~~06~~</td><td>~~Reserved~~ Current Over Temperature Warning Threshold (Optional): This field indicates the composite temperature over temperature warning threshold in degrees Celsius. This is intended to initially match the temperature reported in the WCTEMP field in the NVMe Identify Controller data structure. If the Over Temperature threshold for Composite Temperature is modified with set features, then the most recent value should be reported. The data format should match the same single byte format as the CTemp field with a range from -60 to 127 degrees Celsius. A value of zero means that this field is not reported or that the threshold is set to 0 degrees C.</td></tr>
<tr><td></td><td>06</td><td>Current Power (Optional): This field reports the curent NVM subsystem power consumption. If both bit mapped fields are zero it means that this field is not reported.
<table>
<tr><th>Bit</th><th>Definition</th></tr>
<tr><td>7</td><td>NVM Subsystem Idle (NVMSI): This bit is set to ‘1’ when the NVM subsystem is idle and has been idle for at least 5 seconds. Refer to the NVMe Idle Power (IDLP) definition.</td></tr>
<tr><td>6:0</td><td>NVM Subsystem Power (NVMSP): This field reports the ceiling function of the power consumed by the NVM subsystem in Watts. If this value is greater than 127 Watts, then 127 Watts is reported.

Power reported by the NVM subsystem is determined in the following manner. If NVMSI bit is set to ‘1’, then the value returned is equal to that reported in the Idle Power (IDLP) field in the Power State Descriptor Data Structure for the corresponding NVMe power state. If NVMSI bit is cleared to ‘0’, then the value returned is equal to that reported in the Active Power Workload (APW) field in the Power State Descriptor Structure for the corresponding NVMe power state. The Maximum Power (MP) field value is substituted for IDLP or APW if these are not for reported in the Power State Descriptor Structure for the current NVMe power state.</td></tr>
</table>
</td></tr>
<tr><td></td><td>07</td><td>PEC: An 8 bit CRC calculated over the slave address, command code, second slave address and returned data. The algorithm is defined in the SMBus specification.</td></tr>
</table>