



LEGAL NOTICE:

© **Copyright 2008 to 2024 NVM Express®, Inc. ALL RIGHTS RESERVED.**

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 2024 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, 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

Technical input submitted to the NVM Express® Workgroup is subject to the terms of the NVM Express® Participant’s agreement. Copyright © 2008 to 2024 NVM Express, Inc.

NVM Express® Technical Proposal (TP)

Technical Proposal ID	6036 CY2024 Reference and Capability Updates
Revision Date	2024.07.29
Builds on Specification(s)	NVM Express Base Specification 2.0d NVM Express Command Set Specification 1.0d NVM Express Management Interface Specification 1.2d NVM Express PCI Express Transport Specification 1.0d NVM Express TCP Transport Specification 1.0d NVM Express Boot 1.0
References	TP4055 Key Per I/O TP4152 Post Sanitize Media Verification TP8018 TLS Updates

Technical Proposal Author(s)

Name	Company
John Geldman	KIOXIA

Technical Proposal Overview

This proposal:

- updates the versions of many references used in the NVM Express family of specifications,
- unifies the form and sources of references across the NVM Express family of specifications,
- removes un-used references (so that there is at least one usage of each remaining reference), and
- adds a Bibliography section to the NVM Command Set specification for specifications that are informational.

The biggest technical change affecting NVM Express, was new capability descriptions from the PCI Express® Revision 6.2 reference in NVMe-MI.

A major improvement was made in the NVMe® over PCIe® Transport specification where many descriptions of the configuration registers were improved by removal of incomplete information normatively defined in the PCI Express Base specification.

Revision History

Revision Date	Change Description
2023.07.09	Initial draft
2023.10.10	
2024.04.01	Restarted from template and 2.0d releases (not complete, but a lot to review). TCP Transport, Boot specifications are not fully done, change descriptions are not in place. Presented 2024-04-04.
2024.04.04	Changes, comments from 2024-04-04 TWG discussion.
2024.04.09	Changes, comments from MI, Boot, and FMDS TG meetings. This includes notes that three specifications are in a publication process, and that if those specifications are released
2024.04.24	Changes based on comments from 2024-04-11 call. This adds a reference for RFC 8601 (the definition of UTC), determined that the IEEE 802.1q change is compatible, and that the iBFT version 1.02 is appropriate
2024.04.25	Included an update to MI for the name change to SMBus's type of DSA to DTA, and changed the requirement on PCIESN in non-Flit mode to a 'shall'. This version was approved for a Phase 2 exit vote in the 2024-04-25 call.
2024.05.20	Updated Technical Proposal Overview. Updated description of Changes box. Updated RFC 4622 to RFC 9562.
2024.05.20a	Removed old comments. Left comments on new changes and notes to editor.
2024.05.20b	Added the RFC 4122 to RFC 9562 to the change list (it was already in the proposal)
2024-05-21	Added in text for RFC 4122 to RFC 9562 changes in Base and Boot.
2024.06.30	<p>Member Comment Resolutions:</p> <p>Updated references of NVM Express NVMe-oF™ to 1.1a</p> <p>Newly released references (not in previous TP versions):</p> <ul style="list-style-type: none"> • PCI Express® M.2 Specification, Revision 5.1 • PCI Express® SFF-8639 Module Specification, Revision 5.0 • MCTP Base Specification (DSP0236), Version 1.3.3 • MCTP IDs and Codes (DSP0239), Version 1.11
2024.07.02	Integration Version, Editorial changes.
2024.07.18	<p>Preratification Version</p> <p>Fixed black text error in Base specification's Media Unit Status Descriptor</p> <p>Clean version, Note about pre-approved potential changes removed.</p>
2024.07.27	Editorial updates
2024.07.29	Updated hyperlinks for weblinks

Description for Changes Document for Multiple Specifications

New Features/Feature Enhancements:

- A Bibliography was added to the NVM Command Set specification for documents that were provided for examples or models, but are not required in NVM subsystems.
- The following list of revised versions of references are all compatible changes, with all changes are in the reference section of each listed NVM Express family specification:

New Reference	Old Reference	Affected NVM Express Specifications
ISO/IEC 27040:2024	ISO/IEC 27040:2015	NVM Express Base
JEDEC JESD218B-02: Solid State Drive (SSD) Requirements and Endurance Test Method standard.	JEDEC JESD218B-01: Solid State Drive (SSD) Requirements and Endurance Test Method standard.	NVM Express Base
PCI Express® Base Specification Revision 6.2	PCI Express® Base Specification Revision 5.0 Version 1.0	NVM Express Base, NVM Express Management Interface, NVMe over PCIe Transport, NVM Express Boot
PCI Express® Card Electromechanical Specification, Revision 5.1, Version 1.0	PCI Express® Card Electromechanical Specification, Revision 4.0, Version 1.0	NVM Express Management Interface
PCI Express® M.2 Specification, Revision 5.1	PCI Express® M.2 Specification, Revision 3.0, Version 1.2	NVM Express Management Interface
PCI Express® SFF-8639 Module Specification, Revision 5.0	PCI Express® SFF-8639 Module Specification, Revision 3.0, Version 1.0	NVM Express Management Interface
ACPI Version 6.5	ACPI Version 6.4	NVM Express Base, NVMe over PCIe Transport, NVM Express Boot
UEFI Specification Version 2.10	UEFI Specification Version 2.7A (2.9 in Boot)	NVM Express Base, NVM Express Boot
INCITS 555-2020 Information Technology – SCSI Enclosure Services – 4	INCITS 518-2017 Information Technology – SCSI Enclosure Services – 3	NVM Express Management Interface
MCTP Base Specification (DSP0236), Version 1.3.3	MCTP Base Specification (DSP0236), Version 1.3.1	NVM Express Management Interface
MCTP IDs and Codes (DSP0239), Version 1.11.0	MCTP IDs and Codes (DSP0239), Version 1.7.0	NVM Express Management Interface
MCTP PCIe VDM Transport Binding Specification (DSP0238), Version 1.2.1	MCTP PCIe VDM Transport Binding Specification (DSP0238), Version 1.2.0	NVM Express Management Interface
RFC 9562, “Universally Unique Identifiers”	RFC 4122 “A Universally Unique Identifier (UUID) URN Namespace”	NVM Base Specification, NVM Express Boot

SNIA Native NVMe-oF™ Drive Specification, Version 1.1	SNIA Native NVMe-oF™ Drive Specification, Version 1.0.1	NVM Express Management Interface
SNIA SFF-TA-1006 Enterprise and Datacenter 1U Short SSD Form Factor (E1.S) Specification, Revision 1.5	SNIA SFF-TA-1006 Enterprise and Datacenter 1U Short SSD Form Factor (E1.S) Specification, Revision 1.3a	NVM Express Management Interface
SNIA SFF-TA-1007 Enterprise and Datacenter 1U Long SSD Form Factor (E1.L) Specification, Revision 1.2	SNIA SFF-TA-1007 Enterprise and Datacenter 1U Long SSD Form Factor (E1.L) Specification, Revision 1.1	NVM Express Management Interface
SNIA SFF-TA-1008 Enterprise and Datacenter 3" SSD Form Factor Specification, Revision 2.1	SNIA SFF-TA-1008 Enterprise and Datacenter 3" SSD Form Factor Specification	NVM Express Management Interface
System Management Bus (SMBus) Specification, revision 3.2	System Management Bus (SMBus) Specification, revision 3.1	NVM Express Management Interface
IEEE 802.1q-2022: IEEE Standard for Local and Metropolitan Area Networks—Bridges and Bridged Networks	IEEE 802.1q-2018: IEEE Standard for Local and Metropolitan Area Networks—Bridges and Bridged Networks	NVM Express Boot
DMTF DSP0270, "Redfish Host Interface Specification", Version 1.3.1	DMTF DSP0270, "Redfish Host Interface Specification", Version 1.3	NVM Express Boot
DMTF DSP8010 "Redfish 2023.3 Schema Bundle", Version 2021.4	DMTF DSP8010 "Redfish 2021.4 Schema Bundle", Version 2021.4	NVM Express Boot

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.
Brown:	Moved text
Purple:	Introduced in TP4055 and modified in this TP
Purple Strikethrough:	Introduced in TP4055 and deleted in this TP
Orange	Introduced in TP4152
Light Blue:	Introduced in TP8018 and modified in this TP
Light Blue Strikethrough:	Introduced in TP8018 and deleted in this TP
<Green Bracketed>:	Notes to editor

Description of Specification Changes for NVM Express Base Specification 2.0d

1 Introduction

1.5 Definitions

1.5.53 sanitize operation

Process by which all user data in the NVM subsystem is altered such that recovery of the previous user data from any cache or the non-volatile media is infeasible for a given level of effort (refer to [IEEE 2883™-2022/ISO/IEC 27040](#)).

1.8 References

IEEE 2883™-2022, IEEE Standard for Sanitizing Storage. Available from <https://standards.ieee.org>.

INCITS 502-2019, Information technology – SCSI Primary Commands - 5 (SPC-5). Available from <https://webstore.ansi.org>.

~~INCITS 514-2014, Information technology – SCSI Block Commands – 3 (SBC-3). Available from <https://webstore.ansi.org>.~~

~~INCITS 529-2018, Information technology – ATA/ATAPI Command Set – 4 (ACS-4). Available from <http://webstore.ansi.org>.~~

INCITS 556-2020, Information technology – Non-Volatile Memory Express - 2 (FC-NVMe-2). Available from <https://webstore.ansi.org>.

ISO 8601, Data elements and interchange formats – Information interchange – Representations of dates and times. Available from <https://www.iso.org>.

ISO/IEC 27040:2024~~15~~ Information technology – Security techniques – Storage security. Available from <https://www.iso.org>.

JEDEC JESD218B-02~~4~~: Solid State Drive (SSD) Requirements and Endurance Test Method standard. Available from <https://www.jedec.org>.

NVM Express Management Interface Specification, Revision 1.2. Available from <https://www.nvmexpress.org>.

NVM Express NVM Command Set Specification, Revision 1.0. Available from <https://www.nvmexpress.org>.

NVM Express Zoned Namespace Command Set Specification, Revision 1.1. Available from <https://www.nvmexpress.org>.

NVM Express Key Value Command Set Specification, Revision 1.0. Available from <https://www.nvmexpress.org>.

NVM Express NVMe over PCIe Transport Specification, Revision 1.0. Available from <https://www.nvmexpress.org>.

NVM Express RDMA Transport Specification, Revision 1.0. Available from <https://www.nvmexpress.org>.

NVM Express TCP Transport Specification, Revision 1.0. Available from <https://www.nvmexpress.org>.

~~PCI Local Bus Specification, revision 3.0. Available from <https://www.pcisig.com>.~~

PCI-SIG PCI Express® Base Specification, Revision ~~6.24.0~~. Available from <https://www.pcisig.com>.

~~PCI Bus Power Management Interface Specification Revision 1.2. Available from <https://www.pcisig.com>.~~

~~PCI Single Root I/O Virtualization and Sharing Specification, revision 1.1. Available from https://www.pcisig.com/specifications/iov/single_root/.~~

Technical input submitted to the NVM Express® Workgroup is subject to the terms of the NVM Express® Participant's agreement. Copyright © 2008 to 2024 NVM Express, Inc.

~~PCI Firmware Specification Revision 3.2. Available from <https://www.pcisig.com>.~~

~~PCI Code and ID Assignment Specification Revision 1.11, 24 January, 2019. Available from <https://www.pcisig.com>.~~

RFC 1952, P. Deutsch, "GZIP file format specification version 4.3", May 1996. Available from ~~<https://www.ietf.org/rfc.html>~~<https://www.rfc-editor.org/info/rfc1952>.

RFC 1994, W. Simpson, "PPP Challenge Handshake Authentication Protocol (CHAP)", August 1996. Available from ~~<https://www.ietf.org/rfc.html>~~<https://www.rfc-editor.org/info/rfc1994>.

RFC 2104, H. Krawczyk, M. Bellare, R. Canetti, "HMAC: Keyed-Hashing for Message Authentication", February 1997. Available from ~~<https://www.ietf.org/rfc.html>~~<https://www.rfc-editor.org/info/rfc2104>.

RFC 2631, E. Rescorla, "Diffie-Hellman Key Agreement Method", June 1999. Available from ~~<https://www.ietf.org/rfc.html>~~<https://www.rfc-editor.org/info/rfc2631>.

RFC 3629, F. Yergeau, "UTF-8, a transformation format of ISO 10646", November 2003. Available from ~~<https://www.ietf.org/rfc.html>~~<https://www.rfc-editor.org/info/rfc3629>.

RFC 3986, T. Berners-Lee, R. Fielding, L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", January 2005. Available from ~~<https://www.ietf.org/rfc.html>~~<https://www.rfc-editor.org/info/rfc3986>.

RFC 4086, D. Eastlake 3rd, J. Schiller, S. Crocker, "Randomness Requirements for Security", June 2005. Available from ~~<https://www.ietf.org/rfc.html>~~<https://www.rfc-editor.org/info/rfc4086>.

~~RFC 9562, K. Davis, B. Peabody, and P. Leach, "Universally Unique Identifiers, May 2024". 4122, P. Leach, M. Mealling, and R. Salz, "A Universally Unique Identifier (UUID) URN Namespace", July 2005. Available from <https://www.ietf.org/rfc.html><https://www.rfc-editor.org/info/rfc9562>.~~

RFC 4301, S. Kent, K. Seo, "Security Architecture for the Internet Protocol", December 2005. Available from ~~<https://www.ietf.org/rfc.html>~~<https://www.rfc-editor.org/info/rfc4301>.

RFC 4648, S. Josefsson, "The Base16, Base32, and Base64 Data Encodings", October 2006. Available from ~~<https://www.ietf.org/rfc.html>~~<https://www.rfc-editor.org/info/rfc4648>.

RFC 6234, D. Eastlake 3rd, and T. Hansen, "US Secure Hash Algorithms (SHA and SHA-based HMAC and HKDF)", May 2011. Available from ~~<https://www.ietf.org/rfc.html>~~<https://www.rfc-editor.org/info/rfc6234>.

RFC 7296, C. Kaufman, P. Hoffman, Y. Nir, P. Eronen, T. Kivinen, "Internet Key Exchange Protocol Version 2 (IKEv2)", October 2014. Available from ~~<https://www.ietf.org/rfc.html>~~<https://www.rfc-editor.org/info/rfc7296>.

RFC 7919, D. Gillmor, "Negotiated Finite Field Diffie-Hellman Ephemeral Parameters for Transport Layer Security (TLS)", August 2016. Available from ~~<https://www.ietf.org/rfc.html>~~<https://www.rfc-editor.org/info/rfc7919>.

RFC 8446, E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.3", August 2018. Available from ~~<https://www.ietf.org/rfc.html>~~<https://www.rfc-editor.org/info/rfc8446>.

UEFI Specification Version 2.7A10, ~~August 2022~~[September 2017](https://uefi.org). Available from <https://uefi.org>.

Advanced Configuration and Power Interface (ACPI) Specification, Version 6.45, ~~August 2022~~[January 2024](https://www.uefi.org). Available from <https://www.uefi.org>.

TCG Storage Architecture Core Specification, Version 2.01 Revision 1.00. Available from <https://www.trustedcomputinggroup.org>.

TCG Storage Interface Interactions Specification (SIIS), Version 1.1108 Revision 1.1800. Available from <https://www.trustedcomputinggroup.org>.

TCG Storage Security Subsystem Class: Key Per IO Version 1.00 Revision 1.41. Available from <https://trustedcomputinggroup.org>.

1.9 References Under Development

~~None. INCITS 506-201x, SCSI Block Commands – 4 (SBC-4). Available from <https://www.t10.org>.
TCG Storage Security Subsystem Class: Key Per I/O Specification~~

4.3.6 Universally Unique Identifier (UUID)

The Universally Unique Identifier is defined in RFC 9562~~4122~~ and contained in the Namespace Identification Descriptor (refer to Figure 340). Byte ordering requirements for a UUID are described in RFC 9562~~4122~~.

For historical reasons, UUID subtypes are called UUID versions. Multiple UUID versions are able to be used in the same implementation.

RFC 9562 defines UUID version 8 (i.e., UUIDv8) for experimental or vendor-specific use cases. Uniqueness of UUIDv8 values is implementation specific. NVM Express UUID use cases assume uniqueness within the set of hosts, NVM subsystems, and fabrics that are connected or accessible by a common instance of an administrative tool. For UUIDv8 values, that uniqueness is the responsibility of all of the implementations in that set. RFC 9562 Appendix B provides examples of UUIDv8 generation algorithms that produce unique UUIDs if all implementations in that set generate UUIDv8 values with the same algorithm.

4.5 NVMe Qualified Names

...

The second format may be used to create a unique identifier when there is not a naming authority or there is not a requirement for a human interpretable string. This format consists of:

- The string “nqn”;
- The string “.” (i.e., the ASCII period character);
- The string “2014-08.org.nvmexpress:uuid:”; and
- A 128-bit UUID based on the definition in RFC 9562~~4122~~ represented as a string formatted as “11111111-2222-3333-4444-555555555555”.

The following is an example of an NVMe Qualified Name using the UUID-based format:

- The string “nqn.2014-08.org.nvmexpress:uuid:f81d4fae-7dec-11d0-a765-00a0c91e6bf6”.

...

4.5.1.4 Namespace Unique Identifier

The Identify Namespace data structure (refer to the applicable I/O Command Set specification) contains the IEEE Extended Unique Identifier (EUI64) and the Namespace Globally Unique Identifier (NGUID) fields. The Namespace Identification Descriptor data structure (refer to Figure 340) contains the Namespace UUID. EUI64 is an 8-byte EUI-64 identifier (refer to section 4.3.4), NGUID is a 16-byte identifier based on EUI-64 (refer to section 4.3.5), and Namespace UUID is a 16-byte identifier described in RFC 9562~~4122~~ (refer to section 4.3.6).

When creating a namespace, the controller shall indicate a globally unique namespace identifier in one or more of the following:

- a) the EUI64 field;
- b) the NGUID field; or
- c) a Namespace Identification Descriptor with the Namespace Identifier Type field set to 3h (i.e., a UUID).

Refer to section 8.10.2 for additional globally unique namespace identifier requirements related to dispersed namespaces.

4.6 UTF-8 String Processing

...

Upon entry into the NVMe host (e.g., via a configuration interface at point 1 in Figure 139, described as “input” in RFC 95624122), NVMe host software may process a UTF-8 string (e.g., perform Unicode normalization). Upon entry into the NVM subsystem (e.g., via a configuration interface at point 3 in Figure 139, described as “input” in RFC 95624122), a controller may process a UTF-8 string (e.g., perform Unicode normalization). Upon receipt by the host (e.g., at point 2 in Figure 139) of a UTF-8 string from the controller, text processing (e.g., Unicode normalization) should not occur. Upon receipt by the controller (e.g., at point 4 in Figure 139) of a UTF-8 string from the host, text processing (e.g., Unicode normalization) should not occur.

...

5.23.2.16 UUID List (CNS 17h)

Figure 349: UUID List Entry

Description		
UUID Lists Entry Header:		
Bits	Description	
7:2	Reserved	
1:0	Identifier Association: This field indicates whether the UUID is associated with a vendor.	
	Value	Description
	00b	No association reported.
	01b	The UUID is associated with the vendor reported in the PCI Vendor ID field of the Identify Controller data structure (refer to Figure 337).
	10b	The UUID is associated with the vendor reported in the PCI Subsystem Vendor ID field of the Identify Controller data structure.
	11b	Reserved
Reserved		
UUID: This field contains a 128-bit Universally Unique Identifier (UUID) as specified in RFC 95624122. Refer to section 4.3.6.		

5 Admin Command Set

5.16 Get Log Page command

5.16.1 Log Specific Information

5.16.1.3 SMART / Health Information (Log Page Identifier 02h)

Figure 208: SMART / Health Information Log Page

Bytes	Description
05	<p>Percentage Used: Contains a vendor specific estimate of the percentage of NVM subsystem life used based on the actual usage and the manufacturer's prediction of NVM life. 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 shall be updated once per power-on hour (when the controller is not in a sleep state).</p> <p>Refer to the JEDEC JESD218B-024 standard for SSD device life and endurance measurement techniques.</p>

5.16.1.10 Endurance Group Information (Log Page Identifier 09h)

Figure 218: Endurance Group Information Log Page

Bytes	Description
05	<p>Percentage Used: Contains a vendor specific estimate of the percentage of NVM subsystem life used based on the actual usage and the manufacturer's prediction of NVM life. 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 shall be updated once per power-on hour (when the controller is not in a sleep state).</p> <p>Refer to the JEDEC JESD218B-024 standard for SSD device life and endurance measurement techniques.</p>

5.16.1.16 Media Unit Status (Log Page Identifier 10h)

Figure 250: Media Unit Status Descriptor

Bytes	Description
11	<p>Percentage Used: Contains a vendor specific estimate of the percentage of NVM subsystem life used based on the actual usage and the manufacturer's prediction of NVM life. 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 shall be updated once per power-on hour (when the controller is not in a sleep state).</p> <p>Refer to the JEDEC JESD218B-024 standard for SSD device life and endurance measurement techniques.</p> <p>The relationship between this value and the value in the Percentage Used field in the Endurance Group Information log page is outside the scope of this specification.</p>

5.27 Set Features command

5.27.1 Feature Specific Information

5.27.1.11 Timestamp (Feature Identifier 0Eh)

Figure 340: Timestamp – Data Structure for Set Features

Description
Timestamp: Number of milliseconds that have elapsed since midnight, 01-Jan-1970, UTC. Refer to ISO 8601 for format requirements.
Reserved

8 Extended Capabilities

8.26 Virtualization Enhancements

8.26.4 Single Root I/O Virtualization and Sharing (SR-IOV)

The PCI-SIG® [PCI Express Base specification defines](#) Single Root I/O Virtualization and Sharing (SR-IOV) ~~defines~~ extensions to PCI Express that allow multiple System Images (SIs), such as virtual machines running on a hypervisor, to share PCI hardware resources. The primary benefit of SR-IOV is that it eliminates the hypervisor from participating in I/O operations which may be a significant factor limiting storage performance in some virtualized environments and allows direct SI access to PCI hardware resources.

Description of Specification Changes for NVM Express NVM Command Set Specification 1.0d

1.6 References

<Editor's Note: no changes>

1.6.1 References Under Development

None

~~TCG Storage Security Subsystem Class: Key Per IO Specification~~

1.6.2 Bibliography

INCITS 506-2021 SCSI Block Commands - 4 (SBC-4). Available from <https://webstore.ansi.org/>.

INCITS 558-2021 ATA Command Set - 5 (ACS-5). Available from <https://webstore.ansi.org/>.

TCG Storage Security Subsystem Class: Key Per IO Version 1.00 Revision 1.41. Available from <https://trustedcomputinggroup.org/>

3.2.3.2.1 Deallocated or Unwritten Logical Blocks

...

Note: The operation of the Deallocate function is similar to the ATA DATA SET MANAGEMENT with Trim feature described in ACS-45 and SCSI UNMAP command described in SBC-34.

5.2 End-to-end Data Protection

...

The NVM Express interface supports the same end-to-end protection types defined in the SCSI Protection information model specified in SBC-34. The type of end-to-end data protection (i.e., Type 1, Type 2, or Type 3) is selected when a namespace is formatted and is reported in the Identify Namespace data structure (refer to Figure 97).

5.2.1.1 16b Guard Protection Information

If the Storage Tag Size (STS) field for the LBA Format is cleared to 0h, then the 16b Guard Protection Information is shown in Figure 114 and is contained in the metadata associated with each logical block. The Guard field contains a CRC-16 computed over the logical block data. The formula used to calculate the CRC-16 is defined in SBC-34. In addition to a CRC-16, DIX also specifies an optional IP checksum that is not supported by the NVM Express interface. The Application Tag is an opaque data field not interpreted by the controller and that may be used to disable checking of protection information. The Reference Tag associates logical block data with an address and protects against misdirected or out-of-order logical block transfer. Like the Application Tag, the Reference Tag may also be used to disable checking of protection information.

5.4 Key Per I/O

The Key Per I/O capability operates as defined by the NVM Express Base Specification with the exceptions specified by this section. [The Trusted Computing Group Security Subsystem Key Per I/O specification makes use of this NVM Express capability.](#)

Description of Specification Changes for NVM Express Management Interface

Specification 1.2d

1.6 NVMe Enclosure Architectural Model

...

SCSI Enclosure Services - 43 (SES-43) is a standard developed by the American National Standards Institute T10 committee for management of enclosures using the SCSI architecture. While the NVMe and SCSI architectures differ, the elements of an NVMe Enclosure and a SCSI enclosure are similar and the capabilities required to manage elements of an NVMe Enclosure and a SCSI enclosure are similar. Thus, this specification leverages SES for Enclosure Management. SES manages the elements of an enclosure using control and status diagnostic pages transferred using SCSI commands (refer to Enclosure Control and Enclosure Status diagnostic pages in SES-43). This specification uses these same control and status diagnostic pages but transfers them using the SES Send and SES Receive commands. this specification supports only the standalone Enclosure Services Process model as defined in SES.

A Requester manages an NVMe Enclosure using SES Send and SES Receive commands that are part of the Management Interface Command Set (refer to section 5). The SES Send command provides the functionality of the SES-43 SCSI SEND DIAGNOSTIC command and is used by a Requester to send SES control type diagnostic pages to modify the state of the NVMe Enclosure. The SES Receive command provides the functionality of the SES-43 SCSI RECEIVE DIAGNOSTIC RESULTS command and is used by a Requester to retrieve SES status type diagnostic pages that contain various status and warning information available from the NVMe Enclosure.

Refer to SES-43 for a list and description of SES control type diagnostic pages and SES status type diagnostic pages. The mapping of bytes in SES pages to NVMe MI Request and Response Data is one-to-one where byte x of the SES page maps to byte x in the NVMe MI Request or Response Data (e.g., byte zero of the SES control type diagnostic page corresponds to byte zero of NVMe MI Request Data). The NVMe firmware update process is used (i.e., Firmware Image Download and Firmware Commit commands) to update NVMe firmware. Download Microcode Control and Status diagnostic pages, if supported, shall only be supported on NVMe Enclosure elements.

An Enclosure Services Process, that is logically part of the NVMe Enclosure, is responsible for managing NVMe Enclosure elements and participates in servicing SES Send and SES Receive commands issued by a Requester. Unlike the SES-43 Enclosure Services Process model that maintains state for each I_T nexus (refer to SES-43), unless otherwise noted, this specification requires an NVMe Enclosure to maintain a single global state regardless of the Requester or path used to access that state.

An NVMe Enclosure may contain one or more Subenclosures (refer to SES-43). Each Subenclosure is identified by an SES-43 defined one-byte Subenclosure identifier. If multiple Subenclosures are present, then one of the Subenclosures is designated as the primary Subenclosure and the remaining Subenclosures are secondary Subenclosures. When an NVMe Enclosure consists of only a single Subenclosure, then that Subenclosure is the primary Subenclosure. The Enclosure Services Process associated with the primary Subenclosure is the one that provides access to NVMe Enclosure services information for all Subenclosures. Refer to SES-43 for more information.

Associated with each NVMe Enclosure slot is an SES element that may be used to manage the slot. Refer to SES-43 for more information.

1.8.8 Enclosure Services Process

A process that implements Enclosure services for an NVMe Enclosure that supports Enclosure Management. Refer to SCSI Enclosure Services - 43 (SES-43) for more information.

1.8.20 NVMe Subenclosure (Subenclosure)

A portion of an NVMe Enclosure accessed through a primary NVMe Enclosure's Enclosure Services Process. Refer to SCSI Enclosure Services - 43 (SES-43) for more information.

1.11 References

- I2C Bus specification, revision 6.0. Available from <https://www.i2c-bus.org>.
- IPMI Platform Management FRU Information Storage Definition 1.0, Version 1.3. Available from <https://www.intel.com>.
- INCITS 555-2020~~518-2017~~ Information Technology – SCSI Enclosure Services – ~~43~~ (SES-~~43~~). Available from <https://webstore.ansi.org/>.
- MCTP Base Specification (DSP0236), ~~V~~version 1.3.~~34~~. Available from <https://www.dmtf.org/pmci>.
- MCTP IDs and Codes (DSP0239), ~~V~~version 1.1~~7~~.0. Available from <https://www.dmtf.org/pmci>.
- ~~MCTP Overview White Paper (DSP2016), version 1.0.0. Available from https://www.dmtf.org.~~
- MCTP NVMe™ (NVM Express™) Management Messages over MCTP Binding specification (DSP0235), ~~Version~~revision 1.0.1. Available from <https://www.dmtf.org/pmci>.
- MCTP PCIe VDM Transport Binding Specification (DSP0238), ~~V~~version 1.2.1.~~0~~. Available from <https://www.dmtf.org/pmci>.
- MCTP SMBus/I2C Transport Binding Specification (DSP0237), ~~V~~version 1.2.0. Available from <https://www.dmtf.org/pmci>.
- NVM Express Base Specification, ~~R~~revision 2.0. Available from <https://www.nvmexpress.org>.
- NVM Express Key Value Command Set Specification, ~~R~~revision 1.0. Available from <https://www.nvmexpress.org>.
- NVM Express NVM Command Set Specification, ~~R~~revision 1.0. Available from <https://www.nvmexpress.org>.
- NVM Express NVMe over PCIe Transport Specification, ~~R~~revision 1.0. Available from <https://www.nvmexpress.org>.
- NVM Express Zoned Namespace Command Set Specification, ~~R~~revision 1.1. Available from <https://www.nvmexpress.org>.
- ~~NVMe™ (NVM Express™) Management Messages over MCTP Binding specification (DSP0235), revision 1.0.1. Available from https://www.dmtf.org.~~
- PCI-SIG PCI Express® Base Specification, revision ~~6.25-0~~. Available from <https://www.pcisig.com>.
- PCI-SIG PCI Express® Card Electromechanical Specification, Revision ~~5.14-0~~, Version 1.0. Available from <https://www.pcisig.com>.
- PCI-SIG PCI Express® M.2 Specification, Revision ~~53.10~~, ~~Version 1.2~~. Available from <https://www.pcisig.com>.
- PCI-SIG PCI Express® SFF-8639 Module Specification, Revision ~~53.0~~, ~~Version 1.0~~. Available from <https://www.pcisig.com>.
- RFC 3629, F. Yergeau, “UTF-8, a transformation format of ISO 10646”, November 2003. Available from <https://www.rfc-editor.org/info/rfc3629>~~https://www.ietf.org/rfc.html~~.
- SNIA Native NVMe-oF™ Drive Specification, Version 1.~~0~~.1. Available from <https://www.snia.org>.
- SNIA SFF-TA-1001 Universal x4 Link Definition for SFF-8639 Specification, Revision 1.1. Available from <https://www.snia.org>.
- SNIA SFF-TA-1006 Enterprise and Datacenter 1U Short SSD Form Factor (E1.S) Specification, Revision 1.~~53a~~. Available from <https://www.snia.org>.
- SNIA SFF-TA-1007 Enterprise and Datacenter 1U Long SSD Form Factor (E1.L) Specification, Revision 1.~~24~~. Available from <https://www.snia.org>.

SNIA SFF-TA-1008 Enterprise and Datacenter 3" SSD Form Factor Specification, [Revision 2.1](#). Available from <https://www.snia.org>.
System Management Bus (SMBus) Specification, revision 3.24. Available from <https://www.smbus.org>.

2.2 SMBus/I2C

...

SMBus/I2C elements that support ARP should be implemented as ~~DSA~~ DTA devices (refer to the SMBus Specification). These devices should not issue “Notify ARP ~~Controller~~Master” commands.

...

Figure 17: SMBus/I2C Element UDID

Bits	Field	Description													
79:64	Interface	This field defines the SMBus version and the Interface Protocols supported.													
		Bits	Description	15:08	Reserved	07	ZONE: This bit shall be cleared to '0'.	06	IPMI: This bit shall be cleared to '0'.	05	ASF: This bit shall be set to '1'. Refer to the MCTP SMBus/I2C Transport Binding Specification.	04	OEM: This bit shall be set to '1'.	03:00	SMBus Version: This field shall be set to 4h for SMBus Version 2.0, or to 5h for SMBus Version 3.0, 3.1 , and 3.24 .
		Bits	Description												
		15:08	Reserved												
		07	ZONE: This bit shall be cleared to '0'.												
		06	IPMI: This bit shall be cleared to '0'.												
		05	ASF: This bit shall be set to '1'. Refer to the MCTP SMBus/I2C Transport Binding Specification.												
		04	OEM: This bit shall be set to '1'.												
03:00	SMBus Version: This field shall be set to 4h for SMBus Version 2.0, or to 5h for SMBus Version 3.0, 3.1 , and 3.24 .														

5.7 Read NVMe MI Data Structure

Figure 97: PCIe Port Specific Data

Bytes	Description																
08	PCIe Maximum Payload Size: This field indicates the Max_Payload_Size setting for the specified PCIe port (refer to the PCI Express Base Specification). If the link is not active, this field should be cleared to 0h.																
	<table><tr><th>Value</th><th>Definition</th></tr><tr><td>0h</td><td>128 bytes</td></tr><tr><td>1h</td><td>256 bytes</td></tr><tr><td>2h</td><td>512 bytes</td></tr><tr><td>3h</td><td>1 KiB</td></tr><tr><td>4h</td><td>2 KiB</td></tr><tr><td>5h</td><td>4 KiB</td></tr><tr><td>6h to FFh</td><td>Reserved</td></tr></table>	Value	Definition	0h	128 bytes	1h	256 bytes	2h	512 bytes	3h	1 KiB	4h	2 KiB	5h	4 KiB	6h to FFh	Reserved
Value	Definition																
0h	128 bytes																
1h	256 bytes																
2h	512 bytes																
3h	1 KiB																
4h	2 KiB																
5h	4 KiB																
6h to FFh	Reserved																
	The value reported in this field by ARI Devices and Non-ARI Multi-Function Devices (refer to the PCI Express Base Specification) whose Max Payload Size settings are identical across all Functions is the setting in Function 0. The value reported in this field by non-ARI Multi-Function Devices whose Max Payload Size settings are not identical across all Functions is implementation specific.																

Figure 97: PCIe Port Specific Data

Bytes	Description																														
09	PCIe Supported Link Speeds Vector: This field indicates the Supported Link Speeds for the specified PCIe port.																														
	<table><tr><th>Bits</th><th>Description</th></tr><tr><td>7:x5</td><td>Reserved</td></tr><tr><td>TBD</td><td>64.0 GT/s Support (64GTS): Set to '1' if the PCIe link supports 64.0 GT/s, otherwise cleared to '0'.</td></tr><tr><td>4</td><td>Set to '1' if the PCIe link supports 32.0 GT/s, otherwise cleared to '0'.</td></tr><tr><td>3</td><td>Set to '1' if the PCIe link supports 16.0 GT/s, otherwise cleared to '0'.</td></tr><tr><td>2</td><td>Set to '1' if the PCIe link supports 8.0 GT/s, otherwise cleared to '0'.</td></tr><tr><td>1</td><td>Set to '1' if the PCIe link supports 5.0 GT/s, otherwise cleared to '0'.</td></tr><tr><td>0</td><td>Set to '1' if the PCIe link supports 2.5 GT/s, otherwise cleared to '0'.</td></tr></table>	Bits	Description	7:x5	Reserved	TBD	64.0 GT/s Support (64GTS): Set to '1' if the PCIe link supports 64.0 GT/s, otherwise cleared to '0'.	4	Set to '1' if the PCIe link supports 32.0 GT/s, otherwise cleared to '0'.	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'.														
	Bits	Description																													
	7:x5	Reserved																													
	TBD	64.0 GT/s Support (64GTS): Set to '1' if the PCIe link supports 64.0 GT/s, otherwise cleared to '0'.																													
	4	Set to '1' if the PCIe link supports 32.0 GT/s, otherwise cleared to '0'.																													
	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'.																														
10	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 to FFh</td><td>Reserved</td></tr></table>	Value	Definition	0h	Link not active	1h	The current link speed is the speed indicated in the supported link speed bit 0.	2h	The current link speed is the speed indicated in the supported link speed bit 1.	3h	The current link speed is the speed indicated in the supported link speed bit 2.	4h	The current link speed is the speed indicated in the supported link speed bit 3.	5h	The current link speed is the speed indicated in the supported link speed bit 4.	6h	The current link speed is the speed indicated in the supported link speed bit 5.	7h	The current link speed is the speed indicated in the supported link speed bit 6.	8h to FFh	Reserved										
	Value	Definition																													
	0h	Link not active																													
	1h	The current link speed is the speed indicated in the supported link speed bit 0.																													
	2h	The current link speed is the speed indicated in the supported link speed bit 1.																													
	3h	The current link speed is the speed indicated in the supported link speed bit 2.																													
	4h	The current link speed is the speed indicated in the supported link speed bit 3.																													
	5h	The current link speed is the speed indicated in the supported link speed bit 4.																													
6h	The current link speed is the speed indicated in the supported link speed bit 5.																														
7h	The current link speed is the speed indicated in the supported link speed bit 6.																														
8h to FFh	Reserved																														
11	PCIe Maximum Link Width: The maximum PCIe link width for this NVM Subsystem port. This is the expected negotiated link width that the port link trains to if the platform supports it. A Requester may compare this value with the PCIe Negotiated Link Width to determine if there has been a PCIe link training issue.																														
	<table><tr><th>Value</th><th>Definition</th></tr><tr><td>0</td><td>Reserved</td></tr><tr><td>1</td><td>PCIe x1</td></tr><tr><td>2</td><td>PCIe x2</td></tr><tr><td>3</td><td>Reserved</td></tr><tr><td>4</td><td>PCIe x4</td></tr><tr><td>5 to 7</td><td>Reserved</td></tr><tr><td>8</td><td>PCIe x8</td></tr><tr><td>9 to 11</td><td>Reserved</td></tr><tr><td>12</td><td>PCIe x12</td></tr><tr><td>13 to 15</td><td>Reserved</td></tr><tr><td>16</td><td>PCIe x16</td></tr><tr><td>17 to 31</td><td>Reserved</td></tr><tr><td>32</td><td>PCIe x32</td></tr><tr><td>33 to 255</td><td>Reserved</td></tr></table>	Value	Definition	0	Reserved	1	PCIe x1	2	PCIe x2	3	Reserved	4	PCIe x4	5 to 7	Reserved	8	PCIe x8	9 to 11	Reserved	12	PCIe x12	13 to 15	Reserved	16	PCIe x16	17 to 31	Reserved	32	PCIe x32	33 to 255	Reserved
	Value	Definition																													
	0	Reserved																													
	1	PCIe x1																													
	2	PCIe x2																													
	3	Reserved																													
	4	PCIe x4																													
	5 to 7	Reserved																													
	8	PCIe x8																													
	9 to 11	Reserved																													
	12	PCIe x12																													
	13 to 15	Reserved																													
	16	PCIe x16																													
	17 to 31	Reserved																													
	32	PCIe x32																													
	33 to 255	Reserved																													

Figure 97: PCIe Port Specific Data

Bytes	Description																														
12	<p>PCIe Negotiated Link Width: The negotiated PCIe link width for this port.</p> <table> <tr> <th>Value</th><th>Definition</th></tr> <tr><td>0</td><td>Link not active</td></tr> <tr><td>1</td><td>PCIe x1</td></tr> <tr><td>2</td><td>PCIe x2</td></tr> <tr><td>3</td><td>Reserved</td></tr> <tr><td>4</td><td>PCIe x4</td></tr> <tr><td>5 to 7</td><td>Reserved</td></tr> <tr><td>8</td><td>PCIe x8</td></tr> <tr><td>9 to 11</td><td>Reserved</td></tr> <tr><td>12</td><td>PCIe x12</td></tr> <tr><td>13 to 15</td><td>Reserved</td></tr> <tr><td>16</td><td>PCIe x16</td></tr> <tr><td>17 to 31</td><td>Reserved</td></tr> <tr><td>32</td><td>PCIe x32</td></tr> <tr><td>33 to 255</td><td>Reserved</td></tr> </table>	Value	Definition	0	Link not active	1	PCIe x1	2	PCIe x2	3	Reserved	4	PCIe x4	5 to 7	Reserved	8	PCIe x8	9 to 11	Reserved	12	PCIe x12	13 to 15	Reserved	16	PCIe x16	17 to 31	Reserved	32	PCIe x32	33 to 255	Reserved
Value	Definition																														
0	Link not active																														
1	PCIe x1																														
2	PCIe x2																														
3	Reserved																														
4	PCIe x4																														
5 to 7	Reserved																														
8	PCIe x8																														
9 to 11	Reserved																														
12	PCIe x12																														
13 to 15	Reserved																														
16	PCIe x16																														
17 to 31	Reserved																														
32	PCIe x32																														
33 to 255	Reserved																														
13	PCIe Port Number: This field contains the PCIe port number. This is the same value as that reported in the Port Number field in the PCIe Link Capabilities Register (refer to the NVMe over PCIe Transport Specification).																														
31:14	Reserved																														

Figure 99: Controller Information Data Structure

Bytes	Description								
00	Port Identifier (PORTID): This field specifies the PCIe Port Identifier with which the Controller is associated.								
04:01	Reserved								
05	<p>PCIe Routing ID Information (PRII): This field provides additional data about the PCI Express Routing ID (PRI) for the specified Controller.</p> <table> <tr> <th>Bits</th><th>Description</th></tr> <tr> <td>7:1</td><td>Reserved</td></tr> <tr> <td>0</td><td>PCIe Routing ID Valid: This bit is set to '1' if the device has captured a Bus Number and Device Number (Bus Number only for ARI devices). This bit is cleared to '0' if the device has not captured a Bus and Device number (Bus Number only for ARI devices).</td></tr> </table>	Bits	Description	7:1	Reserved	0	PCIe Routing ID Valid: This bit is set to '1' if the device has captured a Bus Number and Device Number (Bus Number only for ARI devices). This bit is cleared to '0' if the device has not captured a Bus and Device number (Bus Number only for ARI devices).		
Bits	Description								
7:1	Reserved								
0	PCIe Routing ID Valid: This bit is set to '1' if the device has captured a Bus Number and Device Number (Bus Number only for ARI devices). This bit is cleared to '0' if the device has not captured a Bus and Device number (Bus Number only for ARI devices).								
07:06	<p>PCIe Routing ID (PRI): This field contains the PCIe Routing ID for the specified Controller.</p> <table> <tr> <th>Bits</th><th>Description</th></tr> <tr> <td>15:08</td><td>PCI Bus Number: The Controller's PCI Bus Number.</td></tr> <tr> <td>07:03</td><td>PCI Device Number: The Controller's PCI Device Number.</td></tr> <tr> <td>02:00</td><td>PCI Function Number: The Controller's PCI Function Number.</td></tr> </table> <p>Note: For an ARI Device, bits 7:0 represents the (8-bit) Function Number, which replaces the (5-bit) Device Number and (3-bit) Function Number fields above.</p>	Bits	Description	15:08	PCI Bus Number: The Controller's PCI Bus Number.	07:03	PCI Device Number: The Controller's PCI Device Number.	02:00	PCI Function Number: The Controller's PCI Function Number.
Bits	Description								
15:08	PCI Bus Number: The Controller's PCI Bus Number.								
07:03	PCI Device Number: The Controller's PCI Device Number.								
02:00	PCI Function Number: The Controller's PCI Function Number.								
09:08	PCI Vendor ID: The PCI Vendor ID for the specified Controller.								
11:10	PCI Device ID: The PCI Device ID for the specified Controller.								
13:12	PCI Subsystem Vendor ID: The PCI Subsystem Vendor ID for the specified Controller.								
15:14	PCI Subsystem Device ID: The PCI Subsystem Device ID for the specified Controller.								
16	PCIe Segment Number (PCIESN): The Segment Number for the specified Controller when the PCI Express Link is in Flit mode. Refer to the PCI Express Base specification for more information. If the PCI Express interface is not in Flit mode, then this field shall be cleared to 0h.								
31:17	Reserved								

5.9 SES Receive

...

The Page Code (PCODE) field specifies the SES status type diagnostic page to be retrieved. Refer to SES-43 for a list and description of SES diagnostic pages. If the PCODE field specifies a reserved value, an unsupported value, or a value that only corresponds to an SES control type diagnostic page, then the Responder responds with an Invalid Parameter Error Response with the PEL field indicating the PCODE field.

The Allocation Length (ALENGTH) field specifies the maximum length of the Response Data field in the Response Message and is used to limit the maximum amount of SES diagnostic page data that may be returned. The length of the Response Data field shall be the total length of the SES diagnostic page specified by the PCODE field or the number of bytes specified by the ALENGTH field (i.e., the SES diagnostic page is truncated), whichever is less. When the SES diagnostic page is truncated, the value of fields within the SES diagnostic page are not altered to reflect the truncation.

All errors are detected and reported while servicing the SES Receive command and reported via an Error Response. If an invalid field is detected in an SES Receive command, then the Responder responds with an Invalid Parameter Error Response with the PEL field indicating the invalid field. If a condition occurs that in SES-43 results in a CHECK CONDITION, then the Responder responds with an Error Response. The mapping of Error Response Status values to SES-43 sense keys and additional sense codes is shown in Figure 13.

5.10 SES Send

...

Unlike the SES Receive command that specifies the page code of the SES status diagnostic page being retrieved, the SES Send command specifies the page code of the SES control type diagnostic page that is being transferred in the SES control type diagnostic page itself. Refer to SES-43 for a list and description of SES control type diagnostic pages. If the Page Code (PCODE) field in the SES control type diagnostic page specifies a reserved value, an unsupported value, or a value that only corresponds to an SES status diagnostic page, then the Responder responds with an Invalid Parameter Error Response with the PEL field indicating the PCODE field.

The SES Send command does not use NVMe Management Dword 0 or the NVMe Management Response field. All of these are reserved.

All errors are detected and reported while processing the SES Send command and reported via an Error Response. If an invalid field is detected in the SES control type diagnostic page data transferred by an SES Send command, then the Responder responds with an Invalid Parameter Error Response with the PEL field indicating the invalid field. If a condition occurs that in SES-43 results in a CHECK CONDITION, then the Responder responds with an Error Response. The mapping of Response Message Status values to SES-43 sense keys and additional sense codes is shown in Figure 13.

8.2.4 NVMe PCIe Port MultiRecord Area

Figure 154: NVMe PCIe Port MultiRecord Area

Bytes	Factory Default	Description
00	0Ch	NVMe PCIe Port Record Type ID
01	02h or 82h	Record Format:

Figure 154: NVMe PCIe Port MultiRecord Area

Bytes	Factory Default	Description																														
		<table><tr><th>Bits</th><th>Definition</th></tr><tr><td>7</td><td>Set to '1' if last record in list.</td></tr><tr><td>6:0</td><td>Record format version shall be set to 2h.</td></tr></table>	Bits	Definition	7	Set to '1' if last record in list.	6:0	Record format version shall be set to 2h.																								
Bits	Definition																															
7	Set to '1' if last record in list.																															
6:0	Record format version shall be set to 2h.																															
02	08h or 0Bh	Record Length (RLEN): This field indicates the length of the MultiRecord Area in bytes without including the first 5 bytes that are common to all MultiRecords.																														
03	Impl Spec	Record Checksum: This field is used to give the record data a zero checksum (i.e., the modulo 256 sum of the record data bytes from byte offset 05 to the end of this record plus this checksum byte equals 0h).																														
04	Impl Spec	Header Checksum: This field is used to give the record header a zero checksum (i.e., the modulo 256 sum of the least-significant byte of the header through this checksum byte equals 0h).																														
05	1h	NVMe PCIe Port MultiRecord Area Version Number: This field indicates the version number of this NVMe PCIe Port MultiRecord. This field shall be set to 1h in this version of the specification.																														
06	Impl Spec	PCIe Port Number: This field contains the PCIe port number. This is the same value as that reported in the Port Number field in the PCIe Link Capabilities Register.																														
07	Impl Spec	Port Information: This field indicates information about the PCIe Ports in the device. <table><tr><th>Bits</th><th>Definition</th></tr><tr><td>7:1</td><td>Reserved</td></tr><tr><td>0</td><td>If this bit is set to '1', then all PCIe ports within the device have the same capabilities (i.e., the capabilities listed in this structure are consistent across each PCIe port). If this bit is cleared to '0', then all PCIe ports within the device do not have the same capabilities.</td></tr></table>	Bits	Definition	7:1	Reserved	0	If this bit is set to '1', then all PCIe ports within the device have the same capabilities (i.e., the capabilities listed in this structure are consistent across each PCIe port). If this bit is cleared to '0', then all PCIe ports within the device do not have the same capabilities.																								
Bits	Definition																															
7:1	Reserved																															
0	If this bit is set to '1', then all PCIe ports within the device have the same capabilities (i.e., the capabilities listed in this structure are consistent across each PCIe port). If this bit is cleared to '0', then all PCIe ports within the device do not have the same capabilities.																															
08	Impl Spec	PCIe Link Speed: This field indicates a bit vector of link speeds supported by the PCIe port. <table><tr><th>Bits</th><th>Definition</th></tr><tr><td>7:6</td><td>Reserved</td></tr><tr><td>5</td><td>64.0 GT/s Support (64GTS): Set to '1' if the PCIe link supports 64.0 GT/s, otherwise cleared to '0'.</td></tr><tr><td>4</td><td>Set to '1' if the PCIe link supports 32.0 GT/s, otherwise cleared to '0'.</td></tr><tr><td>3</td><td>Set to '1' if the PCIe link supports 16.0 GT/s, otherwise cleared to '0'.</td></tr><tr><td>2</td><td>Set to '1' if the PCIe link supports 8.0 GT/s, otherwise cleared to '0'.</td></tr><tr><td>1</td><td>Set to '1' if the PCIe link supports 5.0 GT/s, otherwise cleared to '0'.</td></tr><tr><td>0</td><td>Set to '1' if the PCIe link supports 2.5 GT/s, otherwise cleared to '0'.</td></tr></table>	Bits	Definition	7:6	Reserved	5	64.0 GT/s Support (64GTS): Set to '1' if the PCIe link supports 64.0 GT/s, otherwise cleared to '0'.	4	Set to '1' if the PCIe link supports 32.0 GT/s, otherwise cleared to '0'.	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'.														
Bits	Definition																															
7:6	Reserved																															
5	64.0 GT/s Support (64GTS): Set to '1' if the PCIe link supports 64.0 GT/s, otherwise cleared to '0'.																															
4	Set to '1' if the PCIe link supports 32.0 GT/s, otherwise cleared to '0'.																															
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'.																															
09	Impl Spec	PCIe Maximum Link Width: The maximum PCIe link width for this NVM Subsystem port. This is the expected negotiated link width that the port link trains to if the platform supports it. A Requester may compare this value with the PCIe Negotiated Link Width to determine if there has been a PCIe link training issue. <table><tr><th>Value</th><th>Definition</th></tr><tr><td>0</td><td>Reserved</td></tr><tr><td>1</td><td>PCIe x1</td></tr><tr><td>2</td><td>PCIe x2</td></tr><tr><td>3</td><td>Reserved</td></tr><tr><td>4</td><td>PCIe x4</td></tr><tr><td>5 to 7</td><td>Reserved</td></tr><tr><td>8</td><td>PCIe x8</td></tr><tr><td>9 to 11</td><td>Reserved</td></tr><tr><td>12</td><td>PCIe x12</td></tr><tr><td>13 to 15</td><td>Reserved</td></tr><tr><td>16</td><td>PCIe x16</td></tr><tr><td>17 to 31</td><td>Reserved</td></tr><tr><td>32</td><td>PCIe x32</td></tr><tr><td>33 to 255</td><td>Reserved</td></tr></table>	Value	Definition	0	Reserved	1	PCIe x1	2	PCIe x2	3	Reserved	4	PCIe x4	5 to 7	Reserved	8	PCIe x8	9 to 11	Reserved	12	PCIe x12	13 to 15	Reserved	16	PCIe x16	17 to 31	Reserved	32	PCIe x32	33 to 255	Reserved
Value	Definition																															
0	Reserved																															
1	PCIe x1																															
2	PCIe x2																															
3	Reserved																															
4	PCIe x4																															
5 to 7	Reserved																															
8	PCIe x8																															
9 to 11	Reserved																															
12	PCIe x12																															
13 to 15	Reserved																															
16	PCIe x16																															
17 to 31	Reserved																															
32	PCIe x32																															
33 to 255	Reserved																															

Figure 154: NVMe PCIe Port MultiRecord Area

Bytes	Factory Default	Description												
10	Impl Spec	MCTP Support: This field contains a bit vector that specifies the level of support for the NVMe Management Interface.												
		<table><tr><th>Bits</th><th>Definition</th></tr><tr><td>7:1</td><td>Reserved</td></tr><tr><td>0</td><td>If this bit is set to '1', then MCTP-based management commands are supported on the PCIe port. If this bit is cleared to '0', then MCTP-based management commands are not supported on the PCIe port.</td></tr></table>	Bits	Definition	7:1	Reserved	0	If this bit is set to '1', then MCTP-based management commands are supported on the PCIe port. If this bit is cleared to '0', then MCTP-based management commands are not supported on the PCIe port.						
		Bits	Definition											
7:1	Reserved													
0	If this bit is set to '1', then MCTP-based management commands are supported on the PCIe port. If this bit is cleared to '0', then MCTP-based management commands are not supported on the PCIe port.													
11	Impl Spec	Ref Clk Capability: This field contains a bit vector that specifies the PCIe clocking modes supported by the port.												
		<table><tr><th>Bits</th><th>Definition</th></tr><tr><td>7:4</td><td>Reserved</td></tr><tr><td>3</td><td>Set to '1' if the device automatically uses RefClk if provided and otherwise uses SRIS, otherwise cleared to '0'.</td></tr><tr><td>2</td><td>Set to '1' if the PCIe link supports Separate RefClk with SSC (SRIS), otherwise cleared to '0'.</td></tr><tr><td>1</td><td>Set to '1' if the PCIe link supports Separate RefClk with no SSC (SRNS), otherwise cleared to '0'.</td></tr><tr><td>0</td><td>Set to '1' if the PCIe link supports common RefClk, otherwise cleared to '0'.</td></tr></table>	Bits	Definition	7:4	Reserved	3	Set to '1' if the device automatically uses RefClk if provided and otherwise uses SRIS, otherwise cleared to '0'.	2	Set to '1' if the PCIe link supports Separate RefClk with SSC (SRIS), otherwise cleared to '0'.	1	Set to '1' if the PCIe link supports Separate RefClk with no SSC (SRNS), otherwise cleared to '0'.	0	Set to '1' if the PCIe link supports common RefClk, otherwise cleared to '0'.
		Bits	Definition											
		7:4	Reserved											
		3	Set to '1' if the device automatically uses RefClk if provided and otherwise uses SRIS, otherwise cleared to '0'.											
		2	Set to '1' if the PCIe link supports Separate RefClk with SSC (SRIS), otherwise cleared to '0'.											
1	Set to '1' if the PCIe link supports Separate RefClk with no SSC (SRNS), otherwise cleared to '0'.													
0	Set to '1' if the PCIe link supports common RefClk, otherwise cleared to '0'.													
12	Impl Spec	Port Identifier: This field contains the NVMe-MI Port Identifier.												
15:13	0h	If the RLEN field is set to 0Bh, then this field is reserved. If the RLEN field is set to 08h, then this field is not present.												

8.2.5.6 PCIe Switch Element Descriptor

Figure 171: PCIe Switch Port Descriptor

Bytes	Factory Default	Description															
00	00h	Type: This field indicates the type of PCIe Switch Port Descriptor. This field shall be cleared to 0h.															
01	Impl Spec	Length: This field indicates the length of the PCIe Switch Port Descriptor in bytes.															
02	Impl Spec	PCIe Link Speed: This field indicates a bit vector of link speeds supported by the PCIe port.															
		Bits	Description	7:x5	Reserved	TBD	64.0 GT/s Support (64GTS): Set to '1' if the PCIe link supports 64.0 GT/s, otherwise cleared to '0'.	4	Set to '1' if the PCIe link supports 32.0 GT/s, otherwise cleared to '0'.	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'.
		Bits	Description														
		7:x5	Reserved														
		TBD	64.0 GT/s Support (64GTS): Set to '1' if the PCIe link supports 64.0 GT/s, otherwise cleared to '0'.														
		4	Set to '1' if the PCIe link supports 32.0 GT/s, otherwise cleared to '0'.														
		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'.																

Figure 171: PCIe Switch Port Descriptor

Bytes	Factory Default	Description																														
03	Impl Spec	PCIe Maximum Link Width: The maximum PCIe link width for this port.																														
		<table><tr><th>Value</th><th>Definition</th></tr><tr><td>0</td><td>Reserved</td></tr><tr><td>1</td><td>PCIe x1</td></tr><tr><td>2</td><td>PCIe x2</td></tr><tr><td>3</td><td>Reserved</td></tr><tr><td>4</td><td>PCIe x4</td></tr><tr><td>5 to 7</td><td>Reserved</td></tr><tr><td>8</td><td>PCIe x8</td></tr><tr><td>9 to 11</td><td>Reserved</td></tr><tr><td>12</td><td>PCIe x12</td></tr><tr><td>13 to 15</td><td>Reserved</td></tr><tr><td>16</td><td>PCIe x16</td></tr><tr><td>17 to 31</td><td>Reserved</td></tr><tr><td>32</td><td>PCIe x32</td></tr><tr><td>33 to 255</td><td>Reserved</td></tr></table>	Value	Definition	0	Reserved	1	PCIe x1	2	PCIe x2	3	Reserved	4	PCIe x4	5 to 7	Reserved	8	PCIe x8	9 to 11	Reserved	12	PCIe x12	13 to 15	Reserved	16	PCIe x16	17 to 31	Reserved	32	PCIe x32	33 to 255	Reserved
		Value	Definition																													
		0	Reserved																													
		1	PCIe x1																													
		2	PCIe x2																													
		3	Reserved																													
		4	PCIe x4																													
		5 to 7	Reserved																													
		8	PCIe x8																													
		9 to 11	Reserved																													
		12	PCIe x12																													
		13 to 15	Reserved																													
		16	PCIe x16																													
		17 to 31	Reserved																													
		32	PCIe x32																													
33 to 255	Reserved																															
04	Impl Spec	RefClk Capability: This field contains a bit vector that specifies the PCIe clocking modes supported by the port.																														
		<table><tr><th>Bits</th><th>Description</th></tr><tr><td>7:4</td><td>Reserved</td></tr><tr><td>3</td><td>Set to '1' for upstream ports that automatically use RefClk if provided and otherwise uses SRIS, otherwise, cleared to '0'. Reserved for downstream ports.</td></tr><tr><td>2</td><td>Set to '1' if the PCIe port supports Separate RefClk with SSC (SRIS), otherwise cleared to '0'.</td></tr><tr><td>1</td><td>Set to '1' if the PCIe port supports Separate RefClk with no SSC (SRNS), otherwise cleared to '0'.</td></tr><tr><td>0</td><td>Set to '1' if the PCIe port supports common RefClk, otherwise cleared to '0'.</td></tr></table>	Bits	Description	7:4	Reserved	3	Set to '1' for upstream ports that automatically use RefClk if provided and otherwise uses SRIS, otherwise, cleared to '0'. Reserved for downstream ports.	2	Set to '1' if the PCIe port supports Separate RefClk with SSC (SRIS), otherwise cleared to '0'.	1	Set to '1' if the PCIe port supports Separate RefClk with no SSC (SRNS), otherwise cleared to '0'.	0	Set to '1' if the PCIe port supports common RefClk, otherwise cleared to '0'.																		
		Bits	Description																													
		7:4	Reserved																													
		3	Set to '1' for upstream ports that automatically use RefClk if provided and otherwise uses SRIS, otherwise, cleared to '0'. Reserved for downstream ports.																													
		2	Set to '1' if the PCIe port supports Separate RefClk with SSC (SRIS), otherwise cleared to '0'.																													
		1	Set to '1' if the PCIe port supports Separate RefClk with no SSC (SRNS), otherwise cleared to '0'.																													
0	Set to '1' if the PCIe port supports common RefClk, otherwise cleared to '0'.																															
05	Impl Spec	Port Number: This field indicates the PCIe Port Number, as defined by the PCI Express Base Specification, associated with this port.																														
06	Impl Spec	PCIe Pointer: In downstream ports this field contains the child index of the Element Descriptor that has a PCIe port connected to this PCIe port. In upstream ports this field is cleared to 0h.																														
07	Impl Spec	Destination Port: This field contains the index of the Port Descriptor in the child Element Descriptor. If the child Element Descriptor has one PCIe upstream port (i.e., a PCIe Switch Element Descriptor), this field shall be cleared to 0h.																														

8.2.5.7 NVM Subsystem Element Descriptor

Figure 173: NVM Subsystem Port Descriptor

Bytes	Factory Default	Description
00	00h	Type: This field indicates the type of an NVM Subsystem Port Descriptor. This field shall be cleared to 0h.
01	Impl Spec	Length: This field indicates the length of the NVM Subsystem Port Descriptor in bytes.

Figure 173: NVM Subsystem Port Descriptor

Bytes	Factory Default	Description																														
02	Impl Spec	<p>PCIe Link Speed: This field indicates a bit vector of link speeds supported by the PCIe port.</p> <table><tr><th>Bits</th><th>Description</th></tr><tr><td>7:65</td><td>Reserved</td></tr><tr><td>5</td><td>64.0 GT/s Support (64GTS): Set to '1' if the PCIe link supports 64.0 GT/s, otherwise cleared to '0'.</td></tr><tr><td>4</td><td>Set to '1' if the PCIe link supports 32.0 GT/s, otherwise cleared to '0'.</td></tr><tr><td>3</td><td>Set to '1' if the PCIe link supports 16.0 GT/s, otherwise cleared to '0'.</td></tr><tr><td>2</td><td>Set to '1' if the PCIe link supports 8.0 GT/s, otherwise cleared to '0'.</td></tr><tr><td>1</td><td>Set to '1' if the PCIe link supports 5.0 GT/s, otherwise cleared to '0'.</td></tr><tr><td>0</td><td>Set to '1' if the PCIe link supports 2.5 GT/s, otherwise cleared to '0'.</td></tr></table>	Bits	Description	7:65	Reserved	5	64.0 GT/s Support (64GTS): Set to '1' if the PCIe link supports 64.0 GT/s, otherwise cleared to '0'.	4	Set to '1' if the PCIe link supports 32.0 GT/s, otherwise cleared to '0'.	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'.														
Bits	Description																															
7:65	Reserved																															
5	64.0 GT/s Support (64GTS): Set to '1' if the PCIe link supports 64.0 GT/s, otherwise cleared to '0'.																															
4	Set to '1' if the PCIe link supports 32.0 GT/s, otherwise cleared to '0'.																															
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'.																															
03	Impl Spec	<p>PCIe Maximum Link Width: The maximum PCIe link width for this NVM Subsystem port.</p> <table><tr><th>Value</th><th>Description</th></tr><tr><td>0</td><td>Reserved</td></tr><tr><td>1</td><td>PCIe x1</td></tr><tr><td>2</td><td>PCIe x2</td></tr><tr><td>3</td><td>Reserved</td></tr><tr><td>4</td><td>PCIe x4</td></tr><tr><td>5 to 7</td><td>Reserved</td></tr><tr><td>8</td><td>PCIe x8</td></tr><tr><td>9 to 11</td><td>Reserved</td></tr><tr><td>12</td><td>PCIe x12</td></tr><tr><td>13 to 15</td><td>Reserved</td></tr><tr><td>16</td><td>PCIe x16</td></tr><tr><td>17 to 31</td><td>Reserved</td></tr><tr><td>32</td><td>PCIe x32</td></tr><tr><td>33 to 255</td><td>Reserved</td></tr></table>	Value	Description	0	Reserved	1	PCIe x1	2	PCIe x2	3	Reserved	4	PCIe x4	5 to 7	Reserved	8	PCIe x8	9 to 11	Reserved	12	PCIe x12	13 to 15	Reserved	16	PCIe x16	17 to 31	Reserved	32	PCIe x32	33 to 255	Reserved
Value	Description																															
0	Reserved																															
1	PCIe x1																															
2	PCIe x2																															
3	Reserved																															
4	PCIe x4																															
5 to 7	Reserved																															
8	PCIe x8																															
9 to 11	Reserved																															
12	PCIe x12																															
13 to 15	Reserved																															
16	PCIe x16																															
17 to 31	Reserved																															
32	PCIe x32																															
33 to 255	Reserved																															
04	Impl Spec	<p>RefClk Capability: This field contains a bit vector that specifies the PCIe clocking modes supported by the port.</p> <table><tr><th>Bits</th><th>Description</th></tr><tr><td>7:4</td><td>Reserved</td></tr><tr><td>3</td><td>Set to '1' if the device automatically uses RefClk if provided and otherwise uses SRIS, otherwise cleared to '0'.</td></tr><tr><td>2</td><td>Set to '1' if the PCIe link supports Separate RefClk with SSC (SRIS), otherwise cleared to '0'.</td></tr><tr><td>1</td><td>Set to '1' if the PCIe link supports Separate RefClk with no SSC (SRNS), otherwise cleared to '0'.</td></tr><tr><td>0</td><td>Set to '1' if the PCIe link supports common RefClk, otherwise cleared to '0'.</td></tr></table>	Bits	Description	7:4	Reserved	3	Set to '1' if the device automatically uses RefClk if provided and otherwise uses SRIS, otherwise cleared to '0'.	2	Set to '1' if the PCIe link supports Separate RefClk with SSC (SRIS), otherwise cleared to '0'.	1	Set to '1' if the PCIe link supports Separate RefClk with no SSC (SRNS), otherwise cleared to '0'.	0	Set to '1' if the PCIe link supports common RefClk, otherwise cleared to '0'.																		
Bits	Description																															
7:4	Reserved																															
3	Set to '1' if the device automatically uses RefClk if provided and otherwise uses SRIS, otherwise cleared to '0'.																															
2	Set to '1' if the PCIe link supports Separate RefClk with SSC (SRIS), otherwise cleared to '0'.																															
1	Set to '1' if the PCIe link supports Separate RefClk with no SSC (SRNS), otherwise cleared to '0'.																															
0	Set to '1' if the PCIe link supports common RefClk, otherwise cleared to '0'.																															
05	Impl Spec	<p>Port Identifier: This field contains the NVMe-MI Port Identifier associated with this port.</p>																														

Appendix A Technical Note: NVM Express Basic Management Command

Figure 176: Subsystem Management Data Structure

Command Code	Offset (byte)	Description
32+	32:255	Vendor Specific – These data structures shall not exceed the maximum read length of 255 specified in the SMBus version 3.x specifications. Preferably their lengths are not greater than 32 for compatibility with SMBus 2.0.

Description of Specification Changes for NVM Express PCI Express Transport Specification 1.0d

1.5 References

NVM Express® (NVMe®) Base Specification revision 2.0. Available from <https://www.nvmexpress.org>.

~~PCI Bus Power Management Interface Specification Revision 1.2. Available from <https://www.pcisig.com>.~~

~~PCI-SIG PCI Express® Base Specification, Revision 6.24.0. Available from <https://www.pcisig.com>.~~

~~PCI-to-PCI Bridge Architecture Specification, Revision 1.2. Available from <https://www.pcisig.com>.~~

Advanced Configuration and Power Interface (ACPI) Specification, Version 6.54, August 2022~~January 2024~~. Available from <https://www.uefi.org>.

3.8.1 PCI Express Type 0/1 Common Configuration Space

Refer to the PCIe Base Specification for the Type, Reset, and Descriptions of the Configuration Space entities, except for the definitions of Class Codes. Refer to the PCI Code and ID Assignment Specification for the definitions of Class Codes.~~where the PCI Code and ID Assignment Specification.~~

3.8.1.1 Offset 00h: ID – Identifiers

Figure 11: Offset 00h: ID - Identifiers

Bits	Type	Reset	NameDescription
31:16	RO	Impl Spec	Device ID (DID): Indicates the device number assigned by the vendor. Specific to each implementation.
15:00	RO	Impl Spec	Vendor ID (VID): Indicates the company vendor, assigned by the PCI-SIG.

3.8.1.2 Offset 04h: CMD – Command

Figure 12: Offset 04h: CMD - Command

Bits	Type	Reset	NameDescription
15:11	RO	0h	Reserved by PCI-SIG®
10	RW	0b	Interrupt Disable (ID): Disables the controller from generating pin-based INTx# interrupts. This bit does not have any effect on MSI or MSI-X operation.
09	RO	0b	Fast Back-to-Back Enable (FBE): Not supported by the NVM Express interface.
08	RW/RO	0b	SERR# Enable (SEE): Controls error reporting.
07	RO	0b	IDSEL Stepping/Wait Cycle Control Reserved by PCI-SIG
06	RW/RO	0b	Parity Error Response Enable (PEE): When set to '1', the controller shall generate PERR# when a data parity error is detected. If parity is not supported, then this bit is read-only '0'.
05	RO	0b	VGA Palette Snooping Enable (VGA): Shall be cleared to zero for NVM Express use.
04	RO	0b	Memory Write and Invalidate Enable (MWIE): Shall be cleared to zero for NVM Express use.
03	RO	0b	Special Cycle Enable (SCE): Shall be cleared to zero for NVM Express use.
02	RW	0b	Bus Master Enable (BME): Enables the controller to act as a master for data transfers. When set to '1', bus master activity is allowed. When cleared to '0', the controller is not allowed to issue any Memory or I/O Requests.
01	RW	0b	Memory Space Enable (MSE): Controls access to the controller's register memory space.
00	RW	0b	I/O Space Enable (IOSE): Controls access to the controller's target I/O space.

3.8.1.3 Offset 06h: STS - Device Status

Figure 13: Offset 06h: STS – Device Status

Bits	Type	Reset	NameDescription
15	RWC	0b	Detected Parity Error (DPE): Set to '1' by hardware when the controller detects a parity error on its interface.
14	RWC/RO	0b	Signaled System Error (SSE): Refer to the PCI-SIG specifications.
13	RWC	0b	Received Master-Abort (RMA): Set to '1' by hardware when the controller receives a master abort to a cycle the controller generated.
12	RWC	0b	Received Target Abort (RTA): Set to '1' by hardware when the controller receives a target abort to a cycle the controller generated.
11	RO	0b	Signaled Target-Abort (STA): Not supported by the NVM Express interface.
10:09	RO	Impl Spec	DEVSEL# Timing (DEVT): Controls the device select time for the controller's PCI interface. This field is not applicable to PCI Express implementations.
08	RWC	0b	Master Data Parity Error Detected (DPD): Set to '1' by hardware when the controller, as a master, either detects a parity error or sees the parity error line asserted, and the Parity Error Response Enable bit (CMD.PEE) is set to '1'.
07	RO	Impl Spec	Fast Back-to-Back Capable (FBC): Shall be cleared to zero for NVM Express use.
06	RO	0b	Reserved by PCI-SIG
05	RO	Impl Spec	66 MHz Capable (C66): Shall be cleared to zero for NVM Express use.
04	RO	1b	Capabilities List (CL): Indicates the presence of a capabilities list. The controller shall support the PCI Power Management capability as a minimum.
03	RO	0	Interrupt Status (IS): Indicates the interrupt status of the device ('1' = asserted).
02:01	RO	000b	Reserved by PCI-SIG
00			Immediate Readiness (IR)

3.8.1.4 Offset 08h: RID - Revision ID

Figure 14: Offset 08h: RID - Revision ID

Bits	Type	Reset	Description
07:00	RO	Impl Spec	Revision ID (RID): Indicates stepping of the controller hardware.

3.8.1.5 Offset 09h: CC - Class Code

Figure 15: Offset 09h: CC - Class Code

Bits	Type	Reset	Description
23:16	RO	04h	Base Class Code (BCC): Indicates the base class code as a mass storage controller.
15:08	RO	08h	Sub Class Code (SCC): Indicates the sub class code as a Non-Volatile Memory controller.
07:00	RO	02h or 03h	Programming Interface (PI): This field specifies that the controller uses the NVM Express programming interface. I/O Controllers shall report 02h and Administrative controllers shall report 03h as defined by the PCI Code and ID Assignment Specification.

3.8.1.6 Offset 0Ch: CLS – Cache Line Size

Figure 16: Offset 0Ch: CLS – Cache Line Size

Bits	Type	Reset	Description
07:00	RW	00h	Cache Line Size (CLS): Cache Line Size register is set by the system firmware or operating system to the system cache size.

3.8.1.7 Offset 0Dh: MLT – Master Latency Timer

Figure 17: Offset 0Dh: MLT – Master Latency Timer

Bits	Type	Reset	Description
07:00	RO	00h	Master Latency Timer (MLT): Indicates the number of clocks the controller is allowed to act as a master on PCI. For a PCI Express device, this register does not apply and shall be hardwired to '0'.

3.8.1.8 Offset 0Eh: HTYPE – Header Type

Figure 18: Offset 0Eh: HTYPE – Header Type

Bits	Type	Reset	Description
07	RO	Impl Spec	Multi-Function Device (MFD): Indicates whether the controller is part of a multi-function device.
06:00	RO	00h	Header Layout (HL): Indicates that the controller uses a target device layout.

3.8.1.9 Offset 0Fh: BIST – Built-In Self Test (Optional)

Figure 19: Offset 0Fh: BIST – Built-In Self Test (Optional)

Bits	Type	Reset	Description
07	RO	Impl Spec	BIST Capable (BC): Indicates whether the controller has a BIST function.
06	RW	0b	Start BIST (SB): Host software sets this bit to '1' to invoke BIST. The controller clears this bit to '0' when BIST is complete.
05:04	RO	00b	Reserved by PCI-SIG
03:00	RO	0h	Completion Code (CC): Indicates the completion code status of BIST. A non-zero value indicates a failure.

3.8.1.11 Offset 14h: MUBAR (BAR1) – Memory Register Base Address, upper 32-bits

Note: NVM Express implementations that reside behind PCI compliant bridges, such as PCI Express Endpoints, are restricted to having 32-bit assigned base address registers due to limitations on the maximum address that may be specified in the bridge ~~for non-prefetchable memory~~. Refer to the [PCI Express Base Specification](#) ~~PCI-to-PCI Bridge Architecture Specification 1.2~~ for more information on this restriction.

Description of Specification Changes for NVM Express TCP Transport Specification 1.0d

1 Introduction

1.4 References

CNSSP 15CNSA 1.0, “USE OF PUBLIC STANDARDS FOR SECURE INFORMATION SHARING”, CNSSP 15 ANNEX B “NSA-APPROVED COMMERCIAL NATIONAL SECURITY ALGORITHM (CNSA) SUITE”, 20 October 2016. Available from <https://www.cnss.gov/CNSS/issuances/Policies.cfm>.

NVM Express® Base Specification, Revision 2.0. Available from <https://www.nvmexpress.org>.

NVM Express over Fabrics Specification, Revision 1.1a. Available from. <https://www.nvmexpress.org>.

Note: the obsoleted NVM Express over Fabrics Specification is included only for its requirements for TLS version 1.2, refer to [section 3.6.1.1](#).

RFC 1952, P. Deutsch, “GZIP file format specification version 4.3”, May 1996. Available from <https://www.rfc-editor.org/rfc/rfc1952><https://www.ietf.org/rfc.html>.

RFC 4648, S. Josefsson, “The Base16, Base32, and Base64 Data Encodings”, October 2006. Available from <https://www.rfc-editor.org/rfc/rfc4648><https://www.ietf.org/rfc.html>.

RFC 5869, H. Krawczyk, P. Eronen, “HMAC-based Extract-and-Expand Key Derivation Functions (HKDF)”, May 2010. Available from <https://www.rfc-editor.org/rfc/rfc5869><https://www.ietf.org/rfc.html>.

RFC 7296, C. Kaufman, P. Hoffman, Y. Nir, P. Eronen, T. Kivinen, “Internet Key Exchange Protocol Version 2 (IKEv2)”, October 2014. Available from <https://www.rfc-editor.org/rfc/rfc7296><https://www.ietf.org/rfc.html>.

RFC 7301, S. Friedl, A. Popov, A. Langley, E. Stephan, “Transport Layer Security (TLS) Application-Layer Protocol Negotiation Extension”, July 2014. Available from <https://www.ietf.org/rfc.html><https://www.rfc-editor.org/rfc/rfc7301>.

RFC 8446, E. Rescorla, “The Transport Layer Security (TLS) Protocol Version 1.3”, August 2018. Available from <https://www.rfc-editor.org/rfc/rfc8446><https://www.ietf.org/rfc.html>.

RFC 8996, K. Moriarty, S. Farrell, “Deprecating TLS 1.0 and TLS 1.1”, March 2021. Available from <https://www.rfc-editor.org/rfc/rfc8996><https://www.ietf.org/rfc.html>.

3.1.2 Transport Service Identifier

TCP port 4420 has been assigned for use by NVM Express over Fabrics and TCP port 8009 has been assigned by IANA (refer to <https://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xhtml>) for use by NVM Express over Fabrics discovery. TCP port 8009 is the default TCP port for NVMe/TCP Discovery controllers. There is no default TCP port for NVMe/TCP I/O controllers. The Transport Service Identifier (TRSVCID) field in the Discovery Log Entry indicates the TCP port to use.

3.3 Data Transfer Model

Figure 10: NVMe/TCP PDU Types

PDU Name	Opcode by field		Combined Opcode ²	Section	PDU Description
	Function (07:01)	PDU Direction (00) ¹			
...					

Figure 10: NVMe/TCP PDU Types

PDU Name	Opcode by field		Combined Opcode ²	Section	PDU Description
	Function (07:01)	PDU Direction (00) ¹			
CapsuleCmd	0000010b	0b	04h	3.6.2.6	Command Capsule: A PDU sent from a host to a controller to transfer aan-NVMe-over Fabrics Command Capsule
CapsuleResp	0000010b	1b	05h	3.6.2.7	Response Capsule: A PDU sent from a controller to a host to transfer aan-NVMe-over Fabrics Response Capsule
...					
Notes:					
1. Indicates the opcode encoded direction of the PDU. All PDUs shall follow this convention:					
a. 0b = Host to Controller (H2C); and					
b. 1b = Controller to Host (C2H).					
2. Opcodes not listed are reserved.					

3.6.1.1 Transport Specific Address Subtype: TLS

Figure 17: Transport Specific Address Subtype Definition for NVMe/TCP Transport

Bytes	Description										
00	Security Type (SECTYPE): Specifies the type of security used by the NVMe/TCP port. If SECTYPE is a value of 0h (No Security), then the host shall set up a normal TCP connection.										
	<table><tr><th>Value</th><th>Definition</th></tr><tr><td>00</td><td>No Security</td></tr><tr><td>01</td><td>Transport Layer Security (TLS) version 1.2 (refer to thethe obsolete NVMe over Fabrics Specification describes requirements for TLS version 1.2). TLS version 1.2 should not be used with NVMe/TCP.</td></tr><tr><td>02</td><td>Transport Layer Security (TLS) version 1.3 (refer to RFC 8446) or a subsequent version. The TLS protocol negotiates the version and cipher suite for each TCP connection.</td></tr><tr><td>255:03</td><td>Reserved</td></tr></table>	Value	Definition	00	No Security	01	Transport Layer Security (TLS) version 1.2 (refer to the the obsolete NVMe over Fabrics Specification describes requirements for TLS version 1.2). TLS version 1.2 should not be used with NVMe/TCP.	02	Transport Layer Security (TLS) version 1.3 (refer to RFC 8446) or a subsequent version. The TLS protocol negotiates the version and cipher suite for each TCP connection.	255:03	Reserved
	Value	Definition									
	00	No Security									
	01	Transport Layer Security (TLS) version 1.2 (refer to the the obsolete NVMe over Fabrics Specification describes requirements for TLS version 1.2). TLS version 1.2 should not be used with NVMe/TCP.									
02	Transport Layer Security (TLS) version 1.3 (refer to RFC 8446) or a subsequent version. The TLS protocol negotiates the version and cipher suite for each TCP connection.										
255:03	Reserved										
255:01	Reserved										

3.6.1.2 Mandatory and Recommended Cipher Suites

TLS for NVMe/TCP is based on pre-shared key (PSK) authentication. NVMe/TCP implementations that support TLS 1.3 shall support the TLS_AES_128_GCM_SHA256 {13h, 01h} cipher suite and should support the TLS_AES_256_GCM_SHA384 {13h, 02h} cipher suite. Implementation and use of the TLS_AES_256_GCM_SHA384 cipher suite may be necessary to meet requirements of security policies that are not defined by NVM Express (e.g., CNSA 1.0 as specified in CNSSP 15, Annex B (the NSA-Approved Commercial National Security Algorithm (CNSA) Suite)).

Description of Specification Changes for NVM Express Boot Specification 1.0d

1 Introduction

1.4 References

- NVM Express® Base Specification, Revision 2.0e. Available from <https://www.nvmexpress.org>.
- NVM Express® Command Set Specification, Revision 1.0b. Available from <https://www.nvmexpress.org>.
- NVM Express® TCP Transport Specification, Revision 1.0b. Available from <https://www.nvmexpress.org>.
- Win 32 API: NL_PREFIX_ORIGIN enumeration, ~~Accessed April 19, 2022.~~ Available from https://docs.microsoft.com/en-us/windows/win32/api/nldef/ne-nldef-nl_prefix_origin.
- RFC 7143, M. Chadalapaka, J. Satran, K. Meth, D. Black, "Internet Small Computer System Interface (iSCSI) Protocol (Consolidated)", April 2014. Available from <https://www.rfc-editor.org/info/rfc7143>~~<https://datatracker.ietf.org/doc/html/rfc7143>~~.
- ~~RFC 3720, J. Satran, K. Meth, C. Sapuntzakis, M. Chadalapaka, E. Zeidner, "Internet Small Computer Systems Interface (iSCSI)", April 2004. Available from <https://datatracker.ietf.org/doc/html/rfc3720>.~~
- RFC 4173, P. Sarkar, D. Missimer, C. Sapuntzakis, "Bootstrapping Clients using the Internet Small Computer System Interface (iSCSI) Protocol", April 2005. Available from <https://www.rfc-editor.org/info/rfc4173>~~<https://datatracker.ietf.org/doc/html/rfc4173>~~.
- iSCSI Boot Firmware Table (iBFT) as Defined in ACPI 3.0b Specification, Version 1.023. Available from <https://docs.microsoft.com/en-us/windows-hardware/drivers/bringup/acpi-system-description-tables#iscsi-boot-firmware-table-ibft>~~<https://download.microsoft.com/download/1/e/5/1e5a2287-366c-431a-8e05-b958540230b1/iBFT.docx>~~.
- IEEE Guidelines for Use of Extended Unique Identifiers, August 2017. Available from <https://standards.ieee.org/wp-content/uploads/import/documents/tutorials/eui.pdf>.
- IEEE 802.1q-202248: IEEE Standard for Local and Metropolitan Area Networks—Bridges and Bridged Networks, ~~December 2022~~~~August 2018~~. Available from <https://ieeexplore.ieee.org/document/10004498>~~<https://standards.ieee.org/ieee/802.1Q/6844/>~~.
- ~~IETF~~ RFC 2132, S. Alexander, R. Droms, "DHCP Options and BOOTP Vendor Extensions", March 1997. Available from <https://rfc-editor.org/rfc/rfc2132>~~<https://rfc-editor.org/rfc/rfc2132.txt>~~.
- RFC 2631, E. Rescorla, "Diffie-Hellman Key Agreement Method", June 1999. Available from <https://www.rfc-editor.org/info/rfc2631>~~<https://www.ietf.org/rfc.html>~~.
- ~~RFC~~ 2732, R. Hinden, B. Carpenter, L. Masinter, "Format for Literal IPv6 Addresses in URL's", December 1999. Available from <https://rfc-editor.org/rfc/rfc2732>~~<https://rfc-editor.org/rfc/rfc2732.txt>~~.
- RFC 3986, T. Berners-Lee, R. Fielding, L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", January 2005. Available from <https://rfc-editor.org/rfc/rfc3986>~~<https://www.ietf.org/rfc.html>~~.
- ~~IETF~~ RFC 4038, M-K. Shin, Ed., Y-G. Hong, J. Hagino, P. Savola, E. M. Castro, "Application Aspects of IPv6 Transition", March 2005. Available from <https://rfc-editor.org/rfc/rfc4038>~~<https://rfc-editor.org/rfc/rfc4038.txt>~~.
- ~~IETF~~ RFC 4249, B. Haberman, "IP Forwarding Table MIB", April 2006. Available from <https://rfc-editor.org/rfc/rfc4249>~~<https://www.ietf.org/rfc/rfc4292.txt>~~.
- ~~IETF~~ RFC 4291, R. Hinden, S. Deering, "IP Version 6 Addressing Architecture", February 2006. Available from <https://rfc-editor.org/rfc/rfc4291>~~<https://rfc-editor.org/rfc/rfc4291.txt>~~.
- DMTF DSP0270, "Redfish Host Interface Specification", ~~Version~~~~Revision~~ 1.3.1. Available from https://www.dmtf.org/sites/default/files/standards/documents/DSP0270_1.3.1.pdf~~https://www.dmtf.org/sites/default/files/standards/documents/DSP0270_1.3.0.pdf~~.

DMTF DSP8010 “Redfish 2023.3 1.4 Schema Bundle”, Version 2023.3, January 2024~~Version 2021.4, December 2021.~~ Available from https://www.dmtf.org/sites/default/files/standards/documents/DSP8010_2023.3.zip~~https://www.dmtf.org/sites/default/files/standards/documents/DSP8010_2021.4.zip~~

Tianocore, “EDK2 Build Specification”, Version 1.0.0. Available from <https://edk2-docs.gitbook.io/edk-ii-build-specification/>.

~~ACPI~~ “Advanced Configuration and Power Interface (ACPI) Specification”, Version 6.54. Available from <https://uefi.org/specifications>~~https://uefi.org/specs/ACPI/6.4/~~.

UEFI, “Unified Extensible Firmware Interface (UEFI) Specification”, Version 2.109. Available from <https://uefi.org/specifications>.

IANA TLS Parameters Registry, TLS Cipher Suites, ~~accessed April 26th, 2022.~~ Available from <https://www.iana.org/assignments/tls-parameters/tls-parameters.xhtml#tls-parameters-4>.

IANA TLS Parameters Registry, TLS Supported Groups, ~~accessed April 26th, 2022.~~ Available from <https://www.iana.org/assignments/tls-parameters/tls-parameters.xhtml#tls-parameters-8>.

NIST Interagency Report 7695, Common Platform Enumeration: Naming Specification Version 2.3~~NISTIR 7695 from ISO/IEC 19770-2.~~ Available from <https://csrc.nist.gov/pubs/ir/7695/final>~~https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir7695.pdf~~

PCI-SIG PCI Express Base Specification, Revision 6.20. Available from <https://members.pcisig.com/wg/PCI-SIG/document>~~https://members.pcisig.com/wg/PCI-SIG/document/16609~~

PCI-SIG PCI Firmware Specification, Revision 3.3. Available from <https://members.pcisig.com/wg/PCI-SIG/document>~~https://members.pcisig.com/wg/PCI-SIG/document/15350~~

1.5.9.3 NVMe UUID String Format

The UUID textual representation shall follow the ~~textual format defined in~~ [IETF RFC 9562.4122 textual format](#).

The following is an example of such an NVMe UUID formatted string:

“urn:uuid:4eff7f8e-d353-4e9b-a4ec-deea8eab84d7”.

2.1.3 Boot from NVMe-oF: TCP Transport

Boot from NVMe-oF storage on TCP, behaves similarly to NVMe PCIe-based storage and NVMe-oF FC-based storage. Once an NVMe-oF association is created and a controller exists, expansion ROM software or boot environment software enumerates namespaces on the NVMe subsystem and convert the basic block storage device interface (select/read/write/deselect) used by the boot environment to the NVM Express protocol commands.

Similar to historical requirements in iSCSI storage (See further detail in Annex A.2.4 [and in RFC 7341](#)), NVMe-oF TCP associations are not self-discovering. The host adapter, or NVMe-oF stack utilizing a network adapter, requires information to locate and/or create the NVMe-oF TCP associations. Information includes IP addresses and port numbers for the initial TCP connections, Host NQN and Host ID values, and Subsystem NQN values to be connected to. To aid in the amount of information required to be known and stored, IP services such as DHCP may be used to query for the NVMe-oF TCP boot device session information. Even in those cases, address information and service type information may have to be specified.

3.2.2.4.1.1 HFI Transport Info Descriptor – NVMe/TCP

Figure 14: HFI Transport Info Descriptor – NVMe/TCP

Bytes	O/M ¹	Description													
...															
10:07	M	<p>PCI Express Routing ID for the HFI Transport Function: This field indicates the PCI Express Routing ID as specified in the PCI Express Base Specification. Note: If the platform supports PCIe ARI, then the Device bits and the Function bits may be concatenated for that purpose.</p> <table><tr><th>Bits</th><th>Description</th><th>Refer to</th></tr><tr><td>31:16</td><td>Segment Group Number</td><td>PCI Firmware Specification</td></tr><tr><td>15:08</td><td>Bus</td><td rowspan="3">PCI Express Base Specification</td></tr><tr><td>07:03</td><td>Device</td></tr><tr><td>02:00</td><td>Function</td></tr></table>	Bits	Description	Refer to	31:16	Segment Group Number	PCI Firmware Specification	15:08	Bus	PCI Express Base Specification	07:03	Device	02:00	Function
Bits	Description	Refer to													
31:16	Segment Group Number	PCI Firmware Specification													
15:08	Bus	PCI Express Base Specification													
07:03	Device														
02:00	Function														
...															
116	O	<p>PCIe Segment Number (PCIESN): The Segment Number for the specified Controller when the PCI Express Link is in Flit mode. Refer to the PCI Express Base specification for more information. If the PCI Express interface is not in Flit mode, or if the pre-OS driver does not support Flit mode, then this field should be cleared to 0h.</p>													
127:117		Reserved													
Notes:															
1. O/M definition: O = Optional. M = Mandatory.															

A.2.4 Boot from SAN: iSCSI Storage

Boot from iSCSI storage is another example of booting from SCSI-based block storage. [Information on iSCSI is found in RFC 7143](#). Most references to Boot from SAN refer to the administrative settings made per host adapter to specify iSCSI session information for the iSCSI target(s) and SCSI LUN number for the block storage devices to enumerate with the boot environment. The settings may also include IP addresses for discovery services to help locate the iSCSI target device.

In contrast to FC storage, iSCSI storage is not self-discovering. The host adapter, or iSCSI stack utilizing a network adapter, requires information to locate and/or create the iSCSI session. Information includes IP addresses and port numbers for the initial TCP connection, iSCSI names to use for the host initiator, and iSCSI names for the iSCSI target to connect to. To aid in the amount of information that is required, IP services such as DHCP, iSNS, or SLP may be used to query for the iSCSI boot device session information. Even in those cases, address information and service type information may have to be specified. Information on such services is found in ~~IEEE~~ RFC 4173.