



LEGAL NOTICE:

© Copyright 2007 to 2021 NVM Express™, Inc. ALL RIGHTS RESERVED.

This NVM Express revision 1.4 technical proposal is proprietary to the NVM Express, Inc. (also referred to as "Company") and/or its successors and assigns.

NOTICE TO USERS WHO ARE NVM EXPRESS, INC. MEMBERS: Members of NVM Express, Inc. have the right to use and implement this NVM Express revision 1.4 technical proposal subject, however, to the Member's continued compliance with the Company's Intellectual Property Policy and Bylaws and the Member's Participation Agreement.

NOTICE TO NON-MEMBERS OF NVM EXPRESS, INC.: If you are not a Member of NVM Express, Inc. and you have obtained a copy of this document, you only have a right to review this document or make reference to or cite this document. Any such references or citations to this document must acknowledge NVM Express, Inc. copyright ownership of this document. The proper copyright citation or reference is as follows: "© 2007 to 2021 NVM Express, Inc. ALL RIGHTS RESERVED." When making any such citations or references to this document you are not permitted to revise, alter, modify, make any derivatives of, or otherwise amend the referenced portion of this document in any way without the prior express written permission of NVM Express, Inc. Nothing contained in this document shall be deemed as granting you any kind of license to implement or use this document or the specification described therein, or any of its contents, either expressly or impliedly, or to any intellectual property owned or controlled by NVM Express, Inc., including, without limitation, any trademarks of NVM Express, Inc.

LEGAL DISCLAIMER:

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

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

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

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

NVM Express Technical Proposal for New Feature

Technical Proposal ID	4064
Change Date	2020-02-25
Builds on Specification	NVM Express 1.4

Technical Proposal Author(s)

Name	Company
Jim Hatfield	Seagate
Lee Prewitt	Microsoft

Define a backward-compatible optional threshold for any single command for the number of SGL Data Block Descriptors and SGL Bit Bucket Descriptors that have a non-zero length.

Revision History

Revision Date	Change Description
2019-10-02	Initial version
2019-10-11	Incorporated comments from Yoni Shternhell, Fred Knight, and Lee Prewitt: adding the word 'optimal' in several places, make 0='not reported', adding 'if reduced performance is not desired'
2019-10-23	Incorporated comments from Judy Brock: move the bulk of the text to a new subsection, change the name of the field, and make the text fit the style of the current document better.
2019-10-31	Collapsed the requirements to apply to all commands. Moved the text requirements to the SGL section of the main spec.
2019-12-19	Fine tuning of the wording based on comments from David Black
2020-01-16	Ready for member review by updating to 2020. Editorial change to SDT field. In 4.4, changed 'shall be equal to' to 'shall be less than or equal to'.
2021-02-24	Integrated into the NVMe Base Specification.
2021-02-25	Accepted all changes, removed comments, and converted all references/cross-references to text.

Description for NVMe 1.4 Changes Document

Define a backward-compatible optional threshold for any single command for the number of SGL Data Block descriptors and SGL Bit Bucket descriptors that have a non-zero length.

Description of Specification Changes

Black text is unchanged from NVMe 1.4

~~Red strikethru~~ is to be deleted

Blue underscore is new text to be added

<Green text is an editor's note, not to be included in the specification>

<ed note: add text to section 4.4 >

4.4 Scatter Gather List (SGL)

A Scatter Gather List (SGL) is a data structure in memory address space used to describe a data buffer. The controller indicates the SGL types that the controller supports in the Identify Controller data structure. A data buffer is either a source buffer or a destination buffer. An SGL contains one or more SGL segments. The total length of the Data Block and Bit Bucket descriptors in an SGL shall be equal to or exceed the amount of data required by the number of logical blocks transferred.

An SGL segment is a qword aligned data structure in a contiguous region of physical memory describing all, part of, or none of a data buffer and the next SGL segment, if any. An SGL segment consists of an array of one or more SGL descriptors. Only the last descriptor in an SGL segment may be an SGL Segment descriptor or an SGL Last Segment descriptor.

A last SGL segment is an SGL segment that does not contain an SGL Segment descriptor, or an SGL Last Segment descriptor.

A controller may support byte or dword alignment and granularity of Data Blocks. If a controller supports only dword alignment and granularity as indicated in the SGL Support field of the Identify Controller data structure (refer to Figure 247), then the values in the Address and Length fields of all Data Block descriptors shall have their lower two bits cleared to 00b. This requirement applies to Data Block descriptors that indicate data and/or metadata memory regions.

The SGL Descriptor Threshold (SDT) field in the Identify Controller data structure (refer to Figure 247) indicates the recommended maximum number of SGL descriptors for a command. If the SDT field is set to a non-zero value and a command is submitted for which the sum of:

- a) the number of SGL Bit Bucket descriptors with non-zero Length field contents; and
 - b) the number of SGL Data Block descriptors with non-zero Length field contents,
- exceeds the value of the SDT field, then the performance of the controller may be reduced.

The value of the SDT field shall be less than or equal to the value of the Maximum SGL Data Block Descriptors field (MSDBD) field in the Identify Controller data structure (refer to the NVMe over Fabrics specification for the definition of the MSDBD field).

<ed note: "Maximum SGL Data Block Descriptors (MSDBD): This field indicates the maximum number of SGL Data Block or Keyed SGL Data Block descriptors that a host is allowed to place in a capsule." >

A Keyed SGL Data Block descriptor is a Data Block descriptor that includes a key that is used as part of the host memory access. The maximum length that may be specified in a Keyed SGL Data Block descriptor is (16 MiB – 1).

A Transport SGL Data Block descriptor is a Data Block descriptor that specifies a data block that is transferred by the NVMe Transport using a transfer mechanism and data buffers that are specific to the NVMe Transport.

The SGL Identifier Descriptor Sub Type field may indicate additional information about a descriptor. As an example, the Sub Type may indicate that the Address field is an offset rather than an absolute address. The Sub Type may also indicate NVMe Transport specific information.

The controller shall abort a command if:

- an SGL segment contains an SGL Segment descriptor or an SGL Last Segment descriptor in other than the last descriptor in the segment;
- a last SGL segment contains an SGL Segment descriptor, or an SGL Last Segment descriptor;
- an SGL descriptor has an unsupported format; or
- an SGL Data Block descriptor contains Address or Length fields with either of the two lower bits set to 1b and the controller supports only dword alignment and granularity as indicated in the SGL Support field of the Identify Controller data structure. Refer to Figure 247.

Figure 110 defines the SGL segment.

....

Make the following change to section 5.15.2.2 Identify Controller data structure (CNS 01h)

Figure 247 - Identify - Identify Controller Data Structure

Figure 2-17. Secondary Controller Data Structure

Bytes	O/M ¹	Description								
Controller Capabilities and Features										
539:536	O	<p>SGL Support (SGLS): This field indicates if SGLs are supported for the NVM command Set and the particular SGL types supported. Refer to section 4.4.</p> <table><tr><th>Bits</th><th>Description</th></tr><tr><td>...</td><td>...</td></tr><tr><td><u>15:08</u></td><td><u>SGL Descriptor Threshold (SDT): This field indicates the recommended maximum number of SGL descriptors in a command (refer to section 4.4). If this field is cleared to 0h, then no recommended maximum number of SGL descriptors is reported.</u></td></tr><tr><td>...</td><td>...</td></tr></table>	Bits	Description	<u>15:08</u>	<u>SGL Descriptor Threshold (SDT): This field indicates the recommended maximum number of SGL descriptors in a command (refer to section 4.4). If this field is cleared to 0h, then no recommended maximum number of SGL descriptors is reported.</u>
Bits	Description									
...	...									
<u>15:08</u>	<u>SGL Descriptor Threshold (SDT): This field indicates the recommended maximum number of SGL descriptors in a command (refer to section 4.4). If this field is cleared to 0h, then no recommended maximum number of SGL descriptors is reported.</u>									
...	...									