Working Draft American National Standard

Project T10/BSR INCITS 550

Revision 0

Information technology - Zoned Block Commands - 2 (ZBC-2)

T10 Technical Editor:

Ralph O. Weber
ENDL Texas
18484 Preston Road
Suite 102, PMB 178
Dallas, TX 75252
USA

Telephone: (214) 912-1373
Email: roweber at ieee dot org

Reference number
ISO/IEC n/a
BSR INCITS 550-201x

Points of Contact

InterNational Committee for Information Technology Standards (INCITS) T10 Technical Committee

T10 Chair
Ralph O. Weber
ENDL Texas
18484 Preston Road
Suite 102, PMB 178
Dallas, TX 75252
USA

Telephone: (214) 912-1373
Email: roweber at ieee dot org

T10 Vice-Chair
William Martin
Samsung Semiconductor, Inc
7213 Marblethorpe Drive
Roseville, CA 95747
USA

Telephone: (916) 765-6875
Email: bill dot martin at ssi dot samsung dot com

T10 Web Site: http://www.t10.org

INCITS Secretariat
1101 K Street, NW
Suite 610
Washington, DC 20005
USA

Telephone: (202) 737-8888
Web site: http://www.incits.org
Email: incits@itic.org

Information Technology Industry Council
Web site: http://www.itic.org

Document Distribution
Purchase INCITS Standards
http://www.incits.org/standards-information/purchase-standards-or-download-dpans

American National Standard
for Information Technology

Draft Zoned Block Commands - 2 (ZBC-2)

Secretariat
Information Technology Industry Council

Approved mm.dd.yy

American National Standards Institute, Inc.

ABSTRACT

This standard incorporates the requirements for devices that require writing to occur at specific locations on their media. These devices allow random reading of logical block data that is already written.

Draft

Approval of an American National Standard requires verification by ANSI that the requirements for due process, consensus, and other criteria for approval have been met by the standards developer. Consensus is established when, in the judgment of the ANSI Board of Standards Review, substantial agreement has been reached by directly and materially affected interests. Substantial agreement means much more than a simple majority, but not necessarily unanimity. Consensus requires that all views and objections be considered, and that effort be made towards their resolution.

The use of American National Standards is completely voluntary; their existence does not in any respect preclude anyone, whether he has approved the standards or not, from manufacturing, marketing, purchasing, or using products, processes, or procedures not conforming to the standards.

American National Standard

The American National Standards Institute does not develop standards and will in no circumstances give interpretation on any American National Standard. Moreover, no person shall have the right or authority to issue an interpretation of an American National Standard in the name of the American National Standards Institute. Requests for interpretations should be addressed to the secretariat or sponsor whose name appears on the title page of this standard.

CAUTION NOTICE: This American National Standard may be revised or withdrawn at any time. The procedures of the American National Standards Institute require that action be taken periodically to reaffirm, revise, or withdraw this standard. Purchasers of American National Standards may receive current information on all standards by calling or writing the American National Standards Institute.

CAUTION: The developers of this standard have requested that holders of patents that may be required for the implementation of the standard, disclose such patents to the publisher. However, neither the developers nor the publisher have undertaken a patent search in order to identify which, if any, patents may apply to this standard. As of the date of publication of this standard, following calls for the identification of patents that may be required for the implementation of the standard, no such claims have been made. No further patent search is conducted by the developer or the publisher in respect to any standard it processes. No representation is made or implied that licenses are not required to avoid infringement in the use of this standard.

Published by
American National Standards Institute, Inc.
11 West 42nd Street, New York, New York 10036

I. Approved Documents Incorporated

Incorporated T10 Approved Documents (in document number order)

Doc	In Rev	Document title [a] (and notes)

[a] Document titles shown in blue text define significant new capabilities.

II. Revision History

II.1 Revision 00 (10 June 2016)

Revision 0 of ZBC-2 is substantially equal to revision 05 of ZBC. The only differences arise from changes made in ZBC by the ANSI Editor during the INCITS Public Review process.

Contents

Tables

Figures

FOREWORD

This foreword is not part of American National Standard BSR INCITS 536.

This purpose of this standard is to define the model and command set extensions to be used in conjunction with the SCSI Primary Command Set standard – 5 (SPC-5) and the SCSI Block Commands - 4 (SBC-4) to facilitate operation of zoned block devices.

Requests for interpretation, suggestions for improvement and addenda, or defect reports are welcome. They should be sent to the INCITS Secretariat, International Committee for Information Technology Standards, Information Technology Industry Council, Suite 610, 1101 K Street, NW, Washington, DC 20005.

This standard was processed and approved for submittal to ANSI by the International Committee for Information Technology Standards (INCITS). Committee approval of the standard does not necessarily imply that all committee members voted for approval. At the time it approved this standard, INCITS had the following members:

to be filled in at the time of publication

INCITS Technical Committee T10 on SCSI Storage Interfaces, which developed and reviewed this standard, had the following members:

Ralph O. Weber, Secretary
William Martin, Vice-Chair
John Geldman, Secretary

Organization Represented	*Name of Representative*
to be filled in at the time of publication	

INTRODUCTION

General

This standard is divided into the following clauses and annexes:

Clause 1 is the scope.
Clause 2 enumerates the normative references that apply to this standard.
Clause 3 describes the definitions, symbols, and abbreviations used in this standard.
Clause 4 describes the operational model.
Clause 5 describes commands for zoned block devices.
Clause 6 describes parameters for zoned block devices.
Annex A describes example methods to achieve optimal performance when accessing zoned block devices.
Bibliography is the bibliography for this standard.

SCSI standards family

Figure 0 shows the relationship of this standard to the other standards and related projects in the SCSI family of standards as of the publication of this standard.

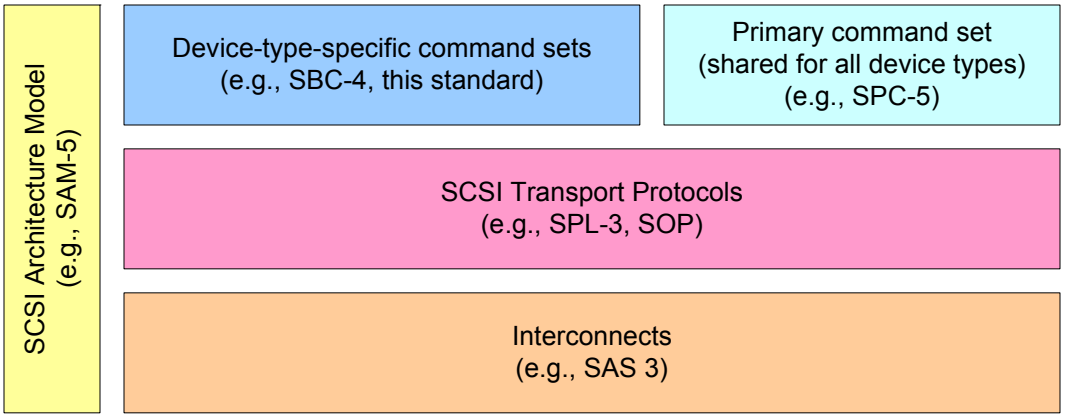

Figure 0 — SCSI document structure

Figure 0 is intended to show the general applicability of the documents to one another. Figure 0 is not intended to imply any hierarchy, protocol stack, or system architecture relationship.

The set of SCSI standards specifies the interfaces, functions, and operations necessary to ensure interoperability between conforming SCSI implementations. This standard is a functional description. Conforming implementations may employ any design technique that does not violate interoperability. See SAM-5 for more information about the relationships between the SCSI standards.

1 Scope

This standard defines the model and command set extensions to facilitate operation of zoned block devices.

The clauses in this standard, implemented in conjunction with the applicable clauses of SPC-5 and SBC-4, specify the standard command set for zoned block devices.

The objectives of this standard are to:

a) permit an application client to communicate over a SCSI service delivery subsystem (see SAM-5) with a logical unit that indicates support for:
   A) the zoned block device type in the PERIPHERAL DEVICE TYPE field of its standard INQUIRY data (see SPC-5); or
   B) the host aware zoned block device model by setting ZONED field to 01b in the Block Device Characteristics VPD page (see SBC-4);

   and

b) define commands and parameters unique to zoned block devices.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

T10/BSR INCITS 515, SCSI Architecture Model - 5 (SAM-5) (planned as ISO/IEC 14776-415)

T10/BSR INCITS 502, SCSI Primary Commands - 5 (SPC-5) (planned as ISO/IEC 14776-455)

T10/BSR INCITS 506, SCSI Block Commands - 4 (SBC-4) (planned as ISO/IEC 14776-323)

T10/BSR INCITS 491, SCSI / ATA Translation - 4 (SAT-4) (under national consideration).

3 Definitions, symbols, abbreviations, and conventions

3.1 Definitions

3.1.1 AND
boolean arithmetic function on two binary input values that results in an output value of one if both of the input values are one, or zero if either of the input values is zero

3.1.2 AND operation
performance of an AND (see 3.1.1) bitwise on two multiple-bit input values both having the same number of bits

Note 1 to entry: An example on the current content of a logical block and the content contained in the Data-Out Buffer having the same number of bytes.

3.1.3 application client
object that is the source of SCSI commands

Note 1 to entry: See SAM-5.

3.1.4 byte
8-bit construct

3.1.5 cache
temporary data storage area that is capable of containing a subset of the logical block data stored by the logical unit and is either volatile or non-volatile

Note 1 to entry: See 4.9.

3.1.6 close zone operation
device server process that makes zone resources non-volatile, releases open zone resources, and results in the Zone Condition becoming CLOSED or EMPTY

Note 1 to entry: See 4.4.3.2.3.

3.1.7 command
request describing a unit of work to be performed by a device server

Note 1 to entry: See SAM-5.

3.1.8 command descriptor block (CDB)
structure used to communicate commands from an application client to a device server

Note 1 to entry: See SPC-5.

3.1.9 conventional zone
zone that is not associated with a write pointer

Note 1 to entry: See 4.4.2.

3.1.10 device server
object within a logical unit (see 3.1.22) that processes SCSI commands according to the rules of command management

Note 1 to entry: See SAM-5.

3.1.11 ending LBA
highest numbered LBA of a command

3.1.12 field
group of one or more contiguous bits, a part of a larger structure

Note 1 to entry: An example of larger structures are a CDB (see 3.1.8) and sense data (see SPC-5).

3.1.13 finish zone operation
device server process that results in the Zone Condition becoming FULL

Note 1 to entry: See 4.4.3.2.4.

3.1.14 format operation
process by which a device server initializes the medium in a logical unit

Note 1 to entry: See 4.6.

3.1.15 hard reset
condition resulting from the events defined by SAM-5 during which the SCSI device performs the hard reset operations described in SAM-5, this standard, and other applicable command standards (see table 14)

3.1.16 host aware zoned block device
logical unit that implements the host aware zoned block device model

Note 1 to entry: See 4.2.2.

3.1.17 host managed zoned block device
logical unit that implements the host managed zoned block device model

Note 1 to entry: See 4.2.3.

3.1.18 I_T nexus
relationship between a SCSI initiator port and a SCSI target port

Note 1 to entry: See SAM-5.

3.1.19 logical block
set of data bytes accessed and referenced as a unit

Note 1 to entry: See SBC-4.

3.1.20 logical block address (LBA)
value used to reference a logical block

Note 1 to entry: See SBC-4.

3.1.21 logical block data
user data and protection information, if any

3.1.22 logical unit
externally addressable entity within a SCSI target device (see 3.1.41) that implements a SCSI device model

Note 1 to entry: See SAM-5.

3.1.23 logical unit reset
condition resulting from the events defined by SAM-5 in which the logical unit performs the logical unit reset operations described in SAM-5, this standard, and other applicable command standards

3.1.24 manage open zone resources operation
device server process that determines the availability of open zone resources

Note 1 to entry: See 4.4.3.2.6.

3.1.25 media

plural of medium

3.1.26 medium
material that is not cache on which data is stored (e.g., a magnetic disk)

3.1.27 non-sequential write operation
write operation performed on a write pointer zone for which the lowest LBA for that operation is not equal to the write pointer for that write pointer zone

3.1.28 non-volatile cache
cache that retains logical block data through any power cycle

3.1.29 open zone
zone with a Zone Condition of EXPLICITLY OPENED or IMPLICITLY OPENED

3.1.30 open zone operation
device server process that results in the Zone Condition becoming EXPLICITLY OPENED

3.1.31 open zone resources
zone resources that are available only to an open zone

Note 1 to entry: The availability of open zone resources is determined by the manage open zone resources operation (see 4.4.3.2.6).

Note 2 to entry: Open zone resources may be volatile (e.g., open zone resources may be cached copies of persistent zone resources).

3.1.32 persistent zone resources
zone resources that persist across all condition changes or SCSI events (see SAM-5) including a power on

Note 1 to entry: For a zone in any Zone Condition other than IMPLICITLY OPENED or EXPLICITLY OPENED, all zone resources are persistent zone resources.

3.1.33 physical block
set of data bytes accessed as a unit by the device server (see SBC-4)

3.1.34 power cycle
sequence of power being removed followed by power being applied to a SCSI device

3.1.35 power on
condition resulting from the events defined by SAM-5 during which a SCSI device performs the power on operations described in SAM-5, this standard, and other applicable command standards (see table 14)

3.1.36 protection information
group of fields at the end of each logical block or at specified intervals within each logical block that contain a logical block guard, an application tag, and a reference tag

Note 1 to entry: See SBC-4.

3.1.37 read command
command that requests read operations

3.1.38 read operation
read operations as described in SBC-4

Note 1 to entry: Examples of read operations include read cache operations (see SBC-4) and read medium operations (see SBC-4).

Note 2 to entry: This standard defines requirements for read operations in addition to those defined in SBC-4.

3.1.39 reset write pointer operation
device server process that results in the write pointer being set to the lowest LBA of the write pointer zone

Note 1 to entry: See 4.4.3.2.5.

3.1.40 sanitize operation
process by which a device server alters information on a logical unit such that recovery of previous logical block data from the cache and the medium is not possible

Note 1 to entry: See 4.7.

3.1.41 SCSI target device
SCSI device containing logical units and SCSI target ports that receives device service requests and task management requests for processing and sends device service responses and task management responses to SCSI initiator devices

Note 1 to entry: See SAM-5.

3.1.42 sense key
contents of the SENSE KEY field in the sense data

Note 1 to entry: See SPC-5.

3.1.43 sequential write preferred zone
write pointer zone in which the device server allows write operations that specify a lowest LBA that is within the zone but is not equal to the zone's write pointer

Note 1 to entry: See 4.4.3.3.

3.1.44 sequential write required zone
write pointer zone in which the device server requires that the lowest LBA for a write operation be the LBA designated by the zone's write pointer

Note 1 to entry: See 4.4.3.4.

3.1.45 starting LBA
lowest numbered LBA of a command

3.1.46 state machine variable
variable that exists within the context of a state machine

Note 1 to entry: A state machine variable may contain status from one state that is used in another state of the same state machine.

Note 2 to entry: The value contained in a state machine variable may affect subsequent state transitions or state machine outputs.

3.1.47 status
one byte of response information that contains a coded value defined in SAM-5, transferred from a device server to an application client upon completion of each command

Note 1 to entry: See SAM-5.

3.1.48 unit attention condition
state that a logical unit (see 3.1.22) maintains while the logical unit has asynchronous status information to report to the SCSI initiator ports associated with one or more I_T nexuses (see 3.1.18)

Note 1 to entry: See SAM-5.

3.1.49 user data
data contained in logical blocks that is accessible by an application client and is neither protection information nor other information that may not be accessible to the application client

3.1.50 verify operation
verify operations as described in SBC-4

Note 1 to entry: An example of verify operations include verify medium operations (see SBC-4).

3.1.51 volatile cache
cache that does not retain logical block data between power cycles

3.1.52 write command
command that requests write operations

3.1.53 write operation
write operations as described in SBC-4

Note 1 to entry: Examples of write operations include write cache operations (see SBC-4) and write medium operations (see SBC-4).

Note 2 to entry: This standard defines requirements for write operations in addition to those defined in SBC-4.

3.1.54 write pointer
pointer to a logical block in a write pointer zone where the next write operation in that zone should start

Note 1 to entry: See 4.4.3.1.

3.1.55 write pointer zone
zone that has an associated write pointer

Note 1 to entry: See 4.4.3.

3.1.56 zone
one of a set of disjoint contiguous ranges of LBAs that comprise a logical unit

Note 1 to entry: See 4.4.

3.1.57 zone attribute
named property associated with a zone

Note 1 to entry: This standard assumes that all zone attributes (see 4.3) are zone resources regardless of how the device sever maintains the defined information.

Note 2 to entry: In the absence of other conditions or actions (e.g., actions by the application client), zone attributes do not affect the operation of the Zone Condition state machine.

Note 3 to entry: See 4.3.

3.1.58 zone resources
device server resources that are used by a zone

Note 1 to entry: This standard defines open zone resources and persistent zone resources.

3.1.59 zoned block device
either a host aware zoned block device or a host managed zoned block device

Note 1 to entry: See 4.2.

3.2 Symbols and abbreviations

3.2.1 Abbreviations

Abbreviations used in this standard:

Abbreviation	Meaning
CDB	command descriptor block
LBA	logical block address
LSB	least significant bit
MSB	most significant bit
n/a	not applicable
SAM-5	SCSI Architecture Model - 5 (see clause 2)
SAT-4	SCSI / ATA Translation - 4 (see clause 2)
SBC-4	SCSI Block Commands-4 (see clause 2)
SCSI	Small Computer System Interface family of standards
SPC-5	SCSI Primary Commands-5 (see clause 2)
VPD	Vital product data

3.2.2 Units

No units are used in this standard.

3.2.3 Symbols

No symbols are used in this standard.

3.2.4 Mathematical operators

Mathematical operators used in this standard:

Mathematical Operators	Meaning
^ or XOR	exclusive logical OR
x	multiplication
/	division
±	plus or minus
x	multiply
+	add
–	subtract
AND	logical AND
< or LT	less than
≤ or LE	less than or equal to
= or EQ	equal
≠ or NE	not equal
> or GT	greater than
≥ or GE	greater than or equal to

3.3 Keywords

3.3.1 invalid
keyword used to describe an illegal or unsupported bit, byte, word, field, or code value

Note 1 to entry: Receipt by a device server of an invalid bit, byte, word, field, or code value shall be reported as error.

3.3.2 mandatory
keyword indicating an item that is required to be implemented as defined in this standard

3.3.3 may
keyword that indicates flexibility of choice with no implied preference

3.3.4 may not
keyword that indicates flexibility of choice with no implied preference

3.3.5 obsolete
keyword indicating that an item was defined in prior SCSI standards but has been removed from this standard

3.3.6 option, optional
keywords that describe features that are not required to be implemented by this standard

Note 1 to entry: If any optional feature defined by this standard is implemented, then it shall be implemented as defined in this standard.

3.3.7 prohibited
keyword used to describe a feature, function, or coded value that is defined in a non-SCSI standard (i.e., a standard that is not a member of the SCSI family of standards) to which this standard makes a normative reference where the use of said feature, function, or coded value is not allowed for implementations of this standard

3.3.8 reserved
keyword referring to bits, bytes, words, fields, and code values that are set aside for future standardization

Note 1 to entry: A reserved bit, byte, word, or field shall be set to zero, or in accordance with a future extension to this standard.

Note 2 to entry: Recipients are not required to check reserved bits, bytes, words, or fields for zero values.

Note 3 to entry: Receipt of reserved code values in defined fields shall be reported as an error.

3.3.9 restricted
keyword referring to bits, bytes, words, and fields that are set aside for other identified standardization purposes

Note 1 to entry: A restricted bit, byte, word, or field shall be treated as a reserved bit, byte, word, or field in the context where the restricted designation appears.

3.3.10 shall
keyword indicating a mandatory requirement; designers are required to implement all such mandatory requirements to ensure interoperability with other products that conform to this standard

3.3.11 should
keyword indicating flexibility of choice with a strongly preferred alternative

3.3.12 vendor specific

something (e.g., a bit, field, code value) that is not defined by this standard

Note 1 to entry: Specification of the referenced item is determined by the SCSI device vendor and may be used differently in various implementations.

3.4 Editorial conventions

Certain words and terms used in this standard have a specific meaning beyond the normal English meaning. These words and terms are defined either in the glossary or in the text where they first appear.

Upper case is used when referring to the name of a numeric value defined in this specification or a formal attribute possessed by an entity. When necessary for clarity, names of objects, procedure calls, arguments or discrete states are capitalized or set in bold type. Names of fields are identified using small capital letters (e.g., ZONED field).

Quantities having a defined numeric value are identified by large capital letters (e.g., CHECK CONDITION). Quantities having a discrete but unspecified value are identified using small capital letters (e.g., OFFLINE is a value of the Zone Condition zone attribute). Such quantities are associated with an event or indication whose observable behavior or value is specific to a given implementation standard.

Lists sequenced by lowercase or uppercase letters show no ordering relationship between the listed items.

EXAMPLE 1 - The following list shows no relationship between the named items:

a) red (i.e., one of the following colors):
   A) crimson; or
   B) amber;
b) blue; or
c) green.

Lists sequenced by numbers show an ordering relationship between the listed items.

EXAMPLE 2 -The following list shows an ordered relationship between the named items:

1) top;
2) middle; and
3) bottom.

Lists are associated with an introductory paragraph or phrase, and are numbered relative to that paragraph or phrase (i.e., all lists begin with an a) or 1) entry).

If a conflict arises between text, tables, or figures, the order of precedence to resolve the conflicts is text, then tables, and finally figures. Not all tables or figures are fully described in the text. Tables show data format and values.

Notes and examples do not constitute any requirements for implementors and notes are numbered consecutively throughout this standard.

3.5 Numeric and character conventions

3.5.1 Numeric conventions

A binary number is represented in this standard by any sequence of digits comprised of only the Arabic numerals 0 and 1 immediately followed by a lower-case b (e.g., 0101b). Underscores or spaces may be included in binary number representations to increase readability or delineate field boundaries (e.g., 0 0101 1010b or 0_0101_1010b).

A hexadecimal number is represented in this standard by any sequence of digits comprised of only the Arabic numerals 0 to 9 and/or the upper-case English letters A to F immediately followed by a lower-case h (e.g., FA23h). Underscores or spaces may be included in hexadecimal number representations to increase readability or delineate field boundaries (e.g., 3456FDCA 84BD5E7Ah, 3456FDCA_84BD5E7Ah, B FD8C FA23h, or B_FD8C_FA23h).

A decimal number is represented in this standard by any sequence of digits comprised of only the Arabic numerals 0 to 9 not immediately followed by a lower-case b or lower-case h (e.g., 25).

A range of numeric values is represented in this standard in the form “a to z”, where a is the first value included in the range, all values between a and z are included in the range, and z is the last value included in the range (e.g., the representation “0h to 3h” includes the values 0h, 1h, 2h, and 3h).

This standard uses the following conventions for representing decimal numbers:

a) the decimal separator (i.e., separating the integer and fractional portions of the number) is a period;
b) the thousands separator (i.e., separating groups of three digits in a portion of the number) is a space;
c) the thousands separator is used in both the integer portion and the fraction portion of a number; and
d) the decimal representation for a year is 1999 not 1 999.

Table 1 shows some examples of decimal numbers using various conventions.

Table 1 — Numbering conventions

French	English	This standard
0,6	0.6	0.6
3,141 592 65	3.14159265	3.141 592 65
1 000	1,000	1 000
1 323 462,95	1,323,462.95	1 323 462.95

A decimal number represented in this standard with an overline over one or more digits following the decimal point is a number where the overlined digits are infinitely repeating (e.g., $666.\overline{6}$ means 666.666 666... or 666 2/3, and $12.\overline{142\ 857}$ means 12.142 857 142 857... or 12 1/7).

3.5.2 Units of measure

This standard represents values using both decimal units of measure and binary units of measure. Values are represented by the following formats:

a) for values based on decimal units of measure:
   1) numerical value (e.g., 100);
   2) space; and
   3) prefix symbol and unit:
      1) decimal prefix symbol (e.g., M) (see table 2); and
      2) unit abbreviation (e.g., B);

   and

b) for values based on binary units of measure:
   1) numerical value (e.g., 1 024);
   2) space; and
   3) prefix symbol and unit:
      1) binary prefix symbol (e.g., Gi) (see table 2); and
      2) unit abbreviation (e.g., b).

Table 2 compares the prefix, symbols, and power of the binary and decimal units.

Table 2 — Comparison of decimal prefixes and binary prefixes

Decimal			Binary		
Prefix name	**Prefix symbol**	**Power (base-10)**	**Prefix name**	**Prefix symbol**	**Power (base-2)**
kilo	k	10^3	kibi	Ki	2^{10}
mega	M	10^6	mebi	Mi	2^{20}
giga	G	10^9	gibi	Gi	2^{30}
tera	T	10^{12}	tebi	Ti	2^{40}
peta	P	10^{15}	pebi	Pi	2^{50}
exa	E	10^{18}	exbi	Ei	2^{60}
zetta	Z	10^{21}	zebi	Zi	2^{70}
yotta	Y	10^{24}	yobi	Yi	2^{80}

3.6 Bit and byte ordering

In this standard, data structures may be defined by a table. A table defines a complete ordering of elements (i.e., bits, bytes, fields, and dwords) within the structure. The ordering of elements within a table does not in itself constrain the order of storage or transmission of the data structure, but in combination with other normative text in this standard, may constrain the order of storage or transmission of the structure.

In a table, any element that is presented in a row above another element in a lower row is more significant than the lower element, and any element presented to the left of another element in the same row is more significant than the element to the right.

If a table shows bit numbering (see table 3), the least significant bit (LSB) is numbered 0 and each more significant bit has the next greater number than the immediately less significant bit. If a table shows numbering of bytes or characters (see table 4), the most significant byte or character is represented at the lowest number and each less significant byte or character has the next greater number than the immediately more significant byte.

In a field in a table consisting of more than one bit that contains a single value (e.g., a number), the least significant bit (LSB) is shown on the right and the most significant bit (MSB) is shown on the left (e.g., in a byte, bit 7 is the MSB and is shown on the left, bit 0 is the LSB and is shown on the right). The MSB and LSB are not labeled if the field consists of eight or fewer bits. The MSB and LSB are labeled if the field consists of more than eight bits and has no internal structure defined.

In a field in a table consisting of more than one byte that contains multiple fields each with their own values (e.g., a descriptor), there is no MSB and LSB of the field itself and thus there are no MSB and LSB labels. Each individual field has an MSB and LSB, but they are not labeled.

In a field containing a text string (e.g., ASCII or UTF-8), only the MSB of the first character and the LSB of the last character are labeled.

Multiple byte fields are represented with only two rows, with the non-sequentially increasing byte number denoting the presence of additional bytes.

A data dword consists of 32 bits. Table 3 shows a data dword containing a single value, where the MSB is on the

upper left in bit 31 and the LSB is on the lower right in bit 0.

Table 3 — Example of ordering of bits and bytes within a data dword

Bit Byte	7	6	5	4	3	2	1	0
0	(MSB) Bit 31	Bit 30	Bit 29	Bit 28	Bit 27	Bit 26	Bit 25	Bit 24
1	Bit 23	Bit 22	Bit 21	Bit 20	Bit 19	Bit 18	Bit 17	Bit 16
2	Bit 15	Bit 14	Bit 13	Bit 12	Bit 11	Bit 10	Bit 9	Bit 8
3	Bit 7	Bit 6	Bit 5	Bit 4	Bit 3	Bit 2	Bit 1	Bit 0 (LSB)

Note - The Bit x labels in the individual table cells are for reference only and should not appear within tables that use this element format.

Table 4 shows a data dword containing four one-byte fields, where byte 0 (the first byte) is on the left and byte 3 (the fourth byte) is on the right. Each byte has an MSB on the left and an LSB on the right.

Table 4 — Example of ordering of bits and bytes within an element dword

Bit Byte	7	6	5	4	3	2	1	0
0	First byte							
	Bit 7 (MSB)	Bit 6	Bit 5	Bit 4	Bit 3	Bit 2	Bit 1	Bit 0 (LSB)
1	Second byte							
	Bit 7 (MSB)	Bit 6	Bit 5	Bit 4	Bit 3	Bit 2	Bit 1	Bit 0 (LSB)
2	Third byte							
	Bit 7 (MSB)	Bit 6	Bit 5	Bit 4	Bit 3	Bit 2	Bit 1	Bit 0 (LSB)
3	Fourth byte							
	Bit 7 (MSB)	Bit 6	Bit 5	Bit 4	Bit 3	Bit 2	Bit 1	Bit 0 (LSB)

Note - The Bit x labels in the individual table cells and the xx byte labels in the individual bytes are for reference only and should not appear within tables the use these element formats. In this example the MSB and LSB labels are for reference only, however, they may appear in multi-byte fields as described in this subclause.

3.7 Notation for state diagrams

All state diagrams use the notation shown in figure 1.

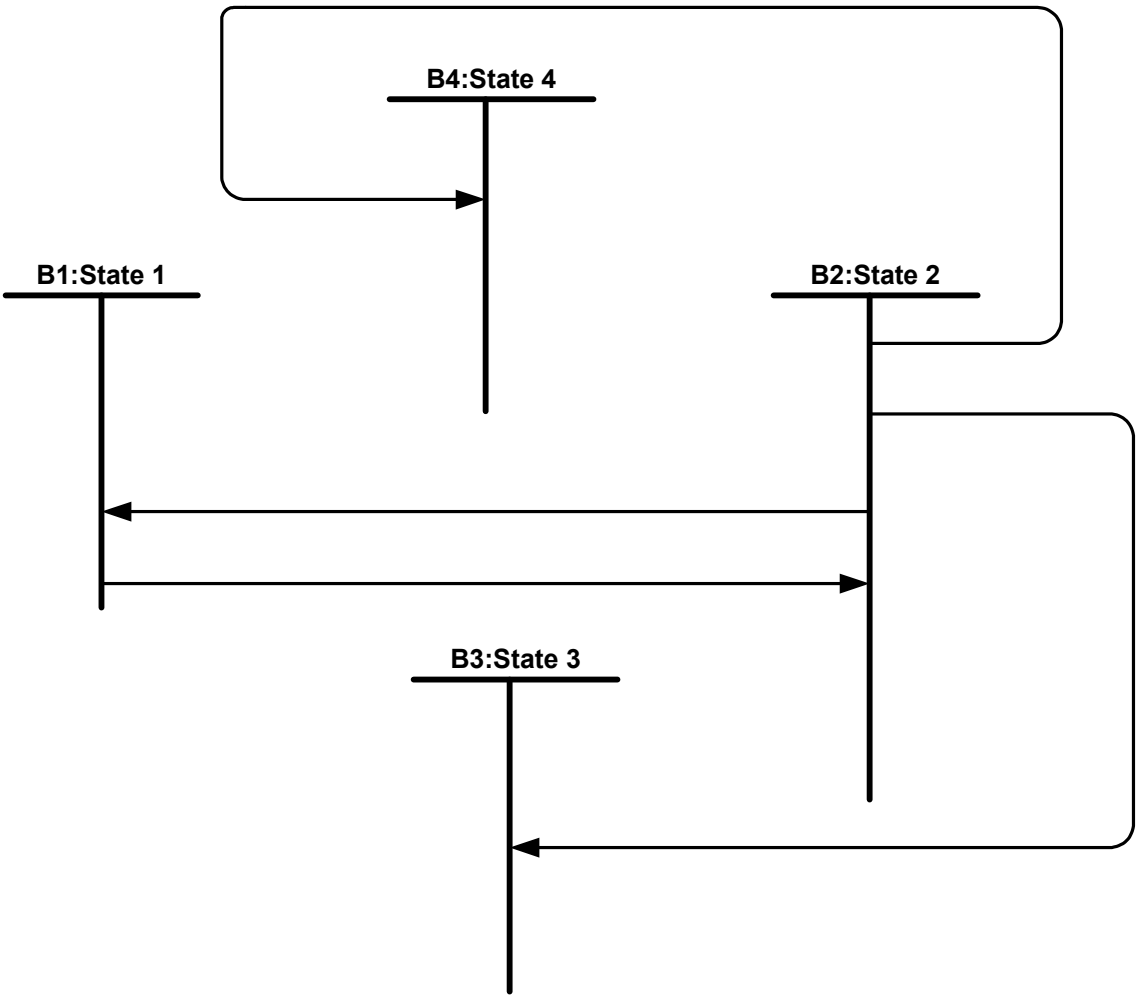

Figure 1 — Example state diagram

Each state is identified by a state designator and a state name. The state designator (e.g., B1) is unique among all state machines in this standard. The state name (e.g., State 1) is a brief description of the primary action taken during the state, and the same state name may be used by other state machines. Actions taken while in each state are described in the state description text.

Transitions between states are shown with solid lines, with an arrow pointing to the destination state.

The conditions and actions are described fully in the transition description text.

Transitions between states are instantaneous.

Upon entry into a state, all actions to be processed in that state are processed.

4 Zoned Block Device Model

4.1 Zoned Block Device model introduction

Table 5 lists topics that apply to zoned block devices.

Table 5 — Zoned block device model topics (part 1 of 2)

Topic [a]	Reference
Zoned block device models	4.2
Zone attributes	4.3
Zone type models	4.4
Error reporting	4.4.3 and SBC-4
Zoned block device extensions to block device model	4.5
Format operations	4.6 and SBC-4
Sanitize operations	4.7 and SBC-4
Reservations	4.8, SPC-5, and SBC-4
Caches	4.9
Background scan operations	SBC-4
Deferred microcode activation	SBC-4
Grouping function	SBC-4
Implicit head of queue command processing	SBC-4 [b]
Initialization	SBC-4
Media examples	SBC-4
Medium defects	SBC-4
Logical blocks	SBC-4
Physical blocks	SBC-4
Protection information model	SBC-4

[a] SBC-4 model subclauses not listed in this table should not be supported for host managed zoned block devices. Host aware zoned block devices may support any SBC-4 model topic.

[b] For a host managed zoned block device, the READ CAPACITY (16) command is found in this standard.

Table 5 — Zoned block device model topics (part 2 of 2)

Topic [a]	Reference
Ready state	SBC-4
START STOP UNIT and power conditions	SBC-4
Write and unmap failures	SBC-4
Write protection	SBC-4

[a] SBC-4 model subclauses not listed in this table should not be supported for host managed zoned block devices. Host aware zoned block devices may support any SBC-4 model topic.

[b] For a host managed zoned block device, the READ CAPACITY (16) command is found in this standard.

4.2 Zoned Block Device models

4.2.1 Zoned Block Device models overview

Zoned block device models described in this standard are as follows:

a) host aware zoned block device (see 4.2.2); and
b) host managed zoned block device (see 4.2.3).

Zoned block devices are accessed using LBAs. The LBAs are divided into ranges called zones (see 4.4). The entire capacity of a zoned block device is organized into a set of logically contiguous, non-overlapping zones. Figure 2 shows a zoned block device with n zones and m LBAs where LBA 0 is the lowest LBA of zone 0 and LBA m–1 is the highest LBA of zone n–1.

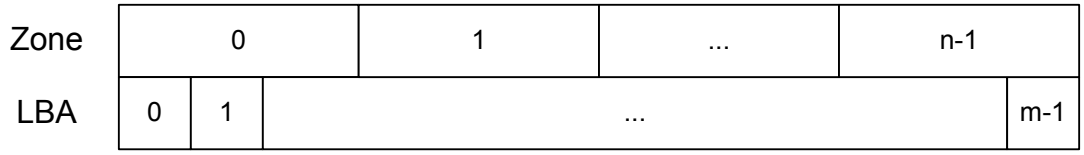

Figure 2 — Zones in a zoned block device

Each zone of a zoned block device has an associated zone type (see 4.4). Zoned block devices use the REPORT ZONES command (see 5.6) for reporting the zone type for each zone.

Table 6 defines the characteristics of zoned block devices defined in this standard.

Table 6 — Requirements of zoned block devices

Characteristics	Reference	Host aware zoned block device	Host managed zoned block device
Command support	n/a	See SPC-5 and SBC-4	See 5.1
PERIPHERAL DEVICE TYPE field value	SPC-5	00h	14h
ZONED field value in the Block Device Characteristics VPD page	SBC-4	01b	00b
Conventional zone	4.4.2	Optional	Optional
Sequential write preferred zone	4.4.3.3	Mandatory	Not supported
Sequential write required zone	4.4.3.4	Not supported	Mandatory
CLOSE ZONE command	5.2	Mandatory	Mandatory
FINISH ZONE command	5.3	Mandatory	Mandatory
OPEN ZONE command	5.4	Mandatory	Mandatory
REPORT ZONES command	5.6	Mandatory	Mandatory
RESET WRITE POINTER command	5.7	Mandatory	Mandatory

4.2.2 Host aware zoned block device model

A host aware zoned block device reports:

a) the PERIPHERAL DEVICE TYPE field set to 00h (i.e., direct access block device) (see SPC-5); and
b) the ZONED field set to 01b in the Block Device Characteristics VPD page (see SBC-4).

Host aware zoned block devices:

a) may support conventional zones (see 4.4.2);
b) shall support one or more sequential write preferred zones (see 4.4.3.3); and
c) shall not support sequential write required zones (see 4.4.3.4).

In addition to commands for direct access block devices (see SBC-4), host aware zoned block devices shall support the commands specified in table 6.

If a write command starts in one zone, and the transfer length extends into one or more additional zones, then the device server splits the command into separate write operations for each zone. The write pointer in each sequential write preferred zone (see 4.4.3.3), if any, is modified by the write operation within that write pointer zone.

4.2.3 Host managed zoned block device model

A host managed zoned block device reports:

a) the PERIPHERAL DEVICE TYPE field set to 14h (i.e., host managed zoned block device); and
b) the ZONED field set to 00h in the Block Device Characteristics VPD page (see SBC-4).

Host managed zoned block devices:

a) may support conventional zones (see 4.4.2);
b) shall not support sequential write preferred zones (see 4.4.3.3); and
c) shall support one or more sequential write required zones (see 4.4.3.4).

See 5.1 for a list of commands for host managed zoned block devices.

4.3 Zone attributes

4.3.1 Zone attributes summary

The zoned blocked device zone attributes are summarized in table 7.

Table 7 — Summary of zone attributes

Attribute	Description	Reference
Zone Type	Type of zone (e.g., CONVENTIONAL)	4.3.2
Zone Condition	Zone's operational characteristics	4.3.3
WPointer	Combination of the write pointer (see 4.4.3.1) and whether the write pointer is valid	4.3.4
RWP Recommended	Indication of whether the processing of a RESET WRITE POINTER command (see 5.7) for this zone is recommended	4.3.5
Non-Sequential Write Resources Active	Indication of whether zone resources are allocated for non-sequential writes	4.3.6

4.3.2 Zone Type zone attribute

Zone Type (see table 8) indicates the type of zone.

Table 8 — Zone Type zone attribute

Zone Type	Reference
CONVENTIONAL	4.4.2
SEQUENTIAL WRITE PREFERRED	4.4.3.3
SEQUENTIAL WRITE REQUIRED	4.4.3.4

The relationships between Zone Type and other zone attributes are shown in table 9.

Table 9 — Relationships between zone attributes

Zone type	Other zone attributes			
	Non-Sequential Write Resources Active	RWP Recommended	WPointer	Zone Condition
CONVENTIONAL	false [a]	false [a]	invalid [a]	see 4.3.3
SEQUENTIAL WRITE PREFERRED	see 4.3.6	see 4.3.5	see 4.4.3.1 and table 12	
SEQUENTIAL WRITE REQUIRED	false [b]			

[a] The zone attributes values for a Zone Type CONVENTIONAL zone are required to be compatible with the direct access block device type defined in SBC-4.

[b] Zone resources for non-sequential writes are not allocated for a zone in which all non-sequential write operations are terminated with an error.

4.3.3 Zone Condition zone attribute

Zone Condition (see table 10) is determined by the Zone Type (see 4.3.2) and the Zone Condition state machine state (see 4.4.3.5).

Table 10 — Zone Condition zone attribute

Zone Type	Zone Condition state machine	Reference	Resulting Zone Condition
CONVENTIONAL	not applicable	4.4.2	NOT WRITE POINTER
SEQUENTIAL WRITE PREFERRED or SEQUENTIAL WRITE REQUIRED	ZC1:Empty state	4.4.3.5.2	EMPTY
	ZC2:Implicit_Open state	4.4.3.5.3	IMPLICITLY OPENED
	ZC3:Explicit_Open state	4.4.3.5.4	EXPLICITLY OPENED
	ZC4:Closed state	4.4.3.5.5	CLOSED
	ZC5:Full state	4.4.3.5.6	FULL
	ZC6:Read_Only state	4.4.3.5.7	READ ONLY
	ZC7:Offline state	4.4.3.5.8	OFFLINE

4.3.4 WPointer zone attribute

For write pointer zones (see 4.4.3), WPointer is the combination of the write pointer (see 4.4.3.1) and whether the write pointer is valid (see table 12).

If Zone Type is CONVENTIONAL, then WPointer shall be invalid.

4.3.5 RWP Recommended zone attribute

For write pointer zones (see 4.4.3), if RWP Recommended is:

a) true, then the device server has determined (e.g., through the detection by the device server that the zone is using too many resources) that the processing of a RESET WRITE POINTER command (see 5.7) specifying this zone is recommended; or
b) false, then the device server has no recommendation for or against the processing of a RESET WRITE POINTER command specifying this zone.

RWP Recommended shall be set to false if:

a) the Zone Type is CONVENTIONAL; or
b) the Zone Condition is EMPTY, READ ONLY, or OFFLINE.

If RWP Recommended changes from false to true then the device server should establish a unit attention condition for the SCSI initiator port (see SAM-5) associated with each I_T nexus with additional sense code set to ZONE RESET WRITE POINTER RECOMMENDED (see SPC-5) and the INFORMATION field (see SPC-5) set to the lowest LBA of the zone for which RWP Recommended was set to true.

4.3.6 Non-Sequential Write Resources Active zone attribute

For sequential write preferred zones (see 4.4.3.3), if Non-Sequential Write Resources Active is:

a) true, then the device server has zone resources allocated for the processing of non-sequential writes in this zone; or

b) false, then the device server does not have zone resources allocated for the processing of non-sequential writes in this zone.

If a non-sequential write operation is performed in a zone, then the device server may set Non-Sequential Write Resources Active to true. The device server may set Non-Sequential Resources Active to false at any time if:

a) the write pointer does not indicate the lowest LBA in the zone; and
b) zone resources are no longer allocated for the processing of non-sequential writes in this zone.

If the write pointer indicates the lowest LBA in the zone and the device server has performed any non-sequential write operations in a zone since the last time the Zone Condition was EMPTY, then Non-Sequential Resources Active shall not be set to false until the write pointer does not indicate the lowest LBA in the zone.

If the device server has not performed a non-sequential write operation in a zone since the last time the Zone Condition was EMPTY, then Non-Sequential Resources Active shall be set to false.

Non-Sequential Write Resources Active shall be set to false if:

a) the Zone Type is CONVENTIONAL or SEQUENTIAL WRITE REQUIRED; or
b) the Zone Condition is EMPTY, READ ONLY, or OFFLINE.

4.4 Zone type models

4.4.1 Zone type models overview

In a zoned block device, each zone:

a) has the zone attributes defined in 4.3; and
b) is one of the following zone types:
   A) a conventional zone (see 4.4.2); or
   B) a write pointer zone (see 4.4.3) that is either:
      a) a sequential write preferred zone (see 4.4.3.3); or
      b) a sequential write required zone (see 4.4.3.4).

4.4.2 Conventional zone model

A conventional zone is a type of zone that is not associated with a write pointer and performs operations as described in SBC-4.

4.4.3 Write pointer zone models

4.4.3.1 Features common to all write pointer zones

4.4.3.1.1 Write pointer features

A write pointer zone is a type of zone that maintains a write pointer that indicates a location within that zone. Each write pointer zone has one write pointer. That write pointer indicates the LBA that the application client should specify in the LOGICAL BLOCK ADDRESS field of the next write command to that zone. The write pointer may be set to the lowest LBA of a zone with a reset write pointer operation (see 4.4.3.2.5) for that zone.

Figure 3 shows a write pointer zone where no LBAs have been written since the last reset write pointer operation for that zone.

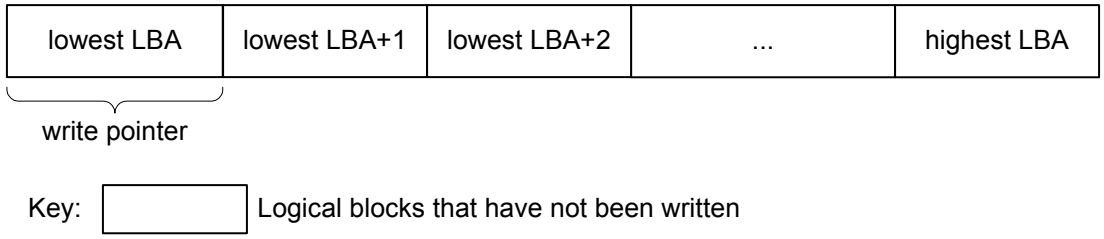

Figure 3 — Write pointer zone and write pointer after reset write pointer operation with no subsequent writes

Figure 4 shows a write pointer zone where LBAs have:

a) been written since the most recent time the zone's Zone Condition was EMPTY (i.e., LBAs less than the write pointer value); and
b) not been written since the most recent time the zone's Zone Condition was EMPTY (i.e., LBAs greater than or equal to the write pointer value).

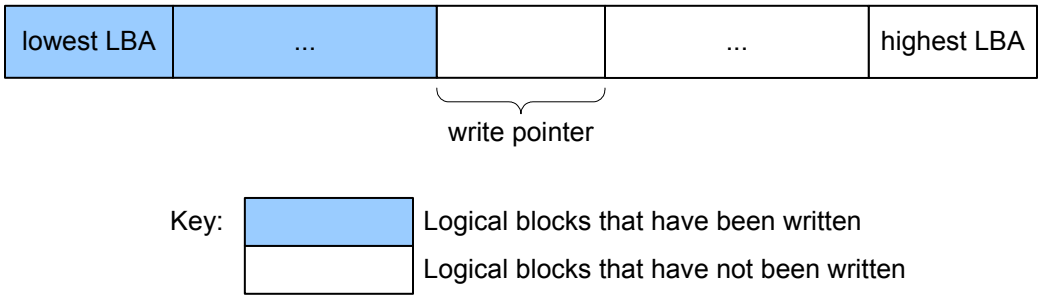

Figure 4 — Write pointer zone and write pointer

The example in figure 5 shows the effects of application client actions on the write pointer as follows:

1) at time 1, the Zone Condition is EMPTY (i.e., the write pointer is set to the lowest LBA of that zone);
2) at time 2:
   1) the Zone Condition becomes IMPLICITLY OPEN;
   2) a write operation writes logical block data starting at that zone's lowest LBA; and
   3) the write pointer is set to the highest LBA written by the write operation plus one;
3) at time 3, a write operation:
   1) writes logical block data that starts at the write pointer after time 2 and extends to the end of that zone; and
   2) as a result, the Zone Condition becomes FULL;

   and
4) at time 4, a reset write pointer operation sets the write pointer to the lowest LBA of that write pointer zone, and the Zone Condition returns to being EMPTY.

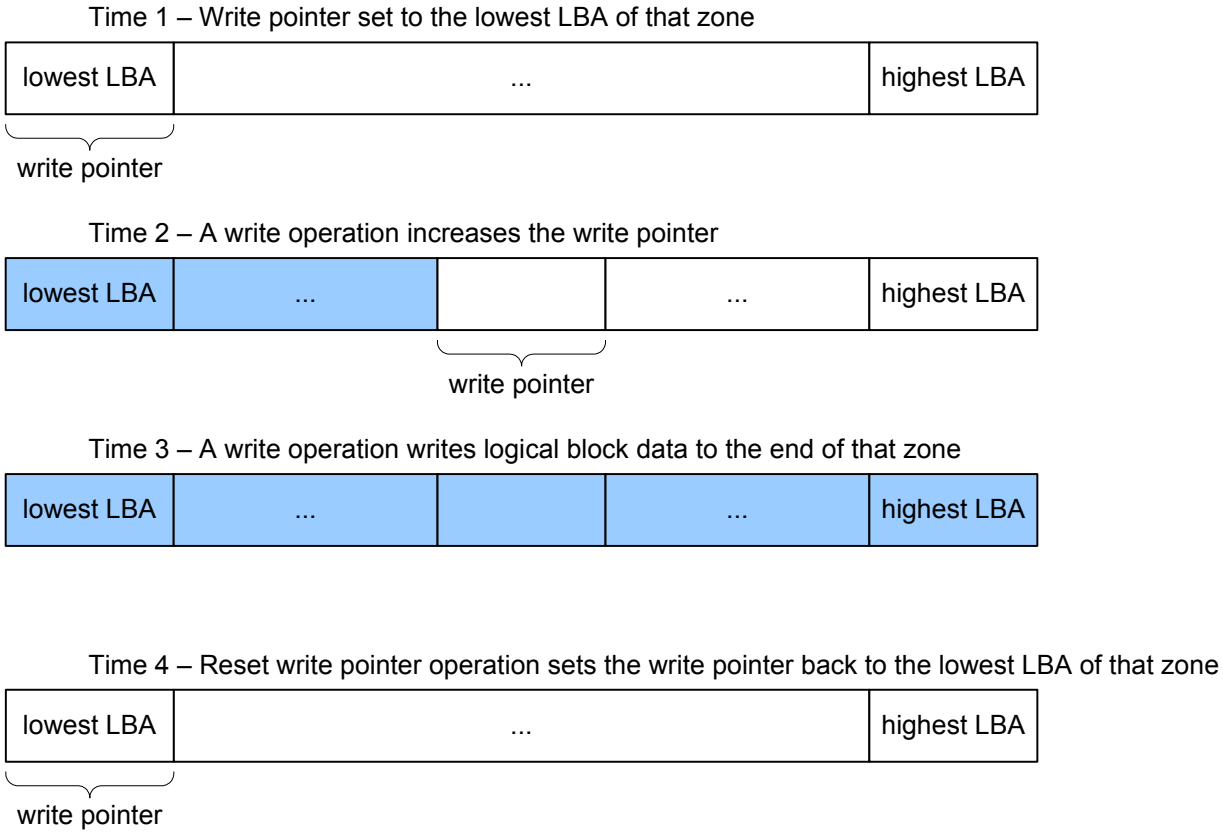

Figure 5 — Write pointer zone example operations

4.4.3.1.2 Resetting the write pointer

A reset write pointer operation (see 4.4.3.2.5) sets the write pointer in a write pointer zone to the lowest LBA of that zone and sets that zone's Zone Condition to EMPTY. A reset write pointer operation is requested by:

a) a RESET WRITE POINTER command (see 5.7);
b) a FORMAT UNIT command (see 4.6 and SBC-4); and

c) a SANITIZE command (see 4.7 and SBC-4).

4.4.3.1.3 Initialization pattern

The initialization pattern used by a zoned block device is the logical block data set:

a) as the default initialization pattern at the time of manufacture;
b) by the most recent FORMAT UNIT command (see SBC-4), if any; or
c) by the most recent SANITIZE command with the service action set to OVERWRITE (see SBC-4), if any.

4.4.3.2 Write pointer zone operations

4.4.3.2.1 Write pointer zone operations overview

The write pointer operations that are applicable to a device server for a zoned block device are summarized in table 11.

Table 11 — Summary of write pointer zone operations

Operation	Description	Reference
Operations defined in this standard for write pointer zones		
open zone operation	Results in the Zone Condition becoming EXPLICITLY OPENED [a]	4.4.3.2.2
close zone operation	Results in the Zone Condition becoming CLOSED	4.4.3.2.3
finish zone operation	Causes subsequent reads that specify logical sectors in the zone that have not been written since the last time the Zone Condition was EMPTY to return the initialization pattern (see 4.4.3.1.3) and results in the Zone Condition becoming FULL	4.4.3.2.4
reset write pointer operation	Results in the Zone Condition becoming EMPTY	4.4.3.2.5
manage open zone resources operation	Limits the number of open zones, if required, by closing a zone with a Zone Condition of IMPLICITLY OPENED	4.4.3.2.6
Operations defined in SBC-4 and modified by this standard for write pointer zones		
read operations	Reads logical block data stored in a write pointer zone or initialization pattern (see 4.4.3.1.3)	
write operations	Writes logical block data stored in a write pointer zone and may result in the Zone Condition being IMPLICITLY OPENED [a]	

[a] The Zone Condition state machine (see 4.4.3.5) requires a manage open zone resources operation to be performed before the Zone Condition is changed to IMPLICITLY OPENED or EXPLICITLY OPENED.

4.4.3.2.2 Open zone operation

Prior to performing an open zone operation, the Zone Condition state machine (see 4.4.3.5) requires the processing of a manage open zone resources operation (see 4.4.3.2.6).

The result of a successful open zone operation is the Zone Condition becoming EXPLICITLY OPENED.

An open zone operation may be performed on a zone with a Zone Condition of EXPLICITLY OPENED. Such an operation has no effect and shall not be considered to be an error.

4.4.3.2.3 Close zone operation

A successful close zone operation annuls any preparations made for writing an open zone (e.g., by deallocating any open zone resources associated with the write pointer zone).

As a result of a close zone operation, the following become non-volatile:

a) any logical block data stored in the specified write pointer zone; and
b) the zone attributes.

The results of a successful close zone operation are:

a) the Zone Condition becoming CLOSED; and
b) all zone resources becoming persistent zone resources.

A close zone operation may be performed on a zone with a Zone Condition of CLOSED. Such an operation has no effect and shall not be considered to be an error.

4.4.3.2.4 Finish zone operation

A finish zone operation results in:

a) all logical blocks in the zone containing non-volatile data;
b) all LBAs in the zone becoming available for reading; and
c) the Zone Condition becoming FULL.

The Zone Condition state machine (see 4.4.3.5) requires the specified zone to have a Zone Condition of EXPLICITLY OPENED or IMPLICITLY OPENED before a finish zone operation is performed. If a zone with a Zone Condition of EMPTY or CLOSED is specified for a finish zone operation, prior to processing the finish zone operation, then the Zone Condition state machine requires that:

a) a manage open zone resources operation (see 4.4.3.2.6) be performed; and
b) the Zone Condition becomes IMPLICITLY OPENED.

The result of a successful finish zone operation is the Zone Condition becoming FULL.

A finish zone operation may be performed on a zone with a Zone Condition of FULL. Such an operation has no effect and shall not be considered to be an error.

4.4.3.2.5 Reset write pointer operation

A reset write pointer operation results in the Zone Condition becoming EMPTY (e.g., the write pointer is set to the lowest LBA in the zone).

A reset write pointer operation may be performed on a zone with a Zone Condition of EMPTY. Such an operation has no effect and shall not be considered to be an error.

A reset write pointer operation is the only method for causing an invalid write pointer to become valid.

4.4.3.2.6 Manage open zone resources operation

4.4.3.2.6.1 Manage open zone resources operation overview

The manage open zone resources operation provides a way for the Zone Condition state machine to ensure that a zone is allowed to become an open zone while maintaining the limits on the number of open zones that the device server indicates in:

a) the OPTIMAL NUMBER OF OPEN SEQUENTIAL WRITE PREFERRED ZONES field (see 6.4.2); and
b) the MAXIMUM NUMBER OF OPEN SEQUENTIAL WRITE REQUIRED ZONES field (see 6.4.2).

The manage open zone resources operation is performed only within the Zone Condition state machine. The manage open zone resources operation shall set the OZR Available state machine variable (see 4.4.3.5.1) to:

a) SUCCESS to indicate that sufficient open zone resources are available to allow a Zone Condition to become IMPLICITLY OPENED or EXPLICITLY OPENED; or
b) FAILURE to indicate that insufficient open zone resources are available to allow a Zone Condition to become IMPLICITLY OPENED or EXPLICITLY OPENED.

The manage open zone resources operation:

a) determines whether open zone resources are available to allow a zone to become an open zone; and
b) may close a zone with a Zone Condition of IMPLICITLY OPENED in order to make open zone resources available.

As detailed in 4.4.3.2.6, an instance of the manage open zone resources operation sets the OZR Available state machine variable to SUCCESS if:

a) closing a zone is not required in order to limit the number of open zones; or
b) a zone is closed in order to limit the number of open zones,

otherwise, the manage open zone resources operation sets the OZR Available state machine variable to FAILURE.

4.4.3.2.6.2 Select a sequential write preferred zone

The device server:

a) may choose a zone with a Zone Condition of IMPLICITLY OPENED and perform a close zone operation (see 4.4.3.2.3) on that zone; and
b) shall set the OZR Available state machine variable to SUCCESS (see 4.4.3.2.6.1),

if the following formula evaluates to true:

$$(k \leq (x + y)) \text{ AND } (y > 0)$$

where:

x is the number of sequential write preferred zones with a Zone Condition of EXPLICITLY OPEN;
y is the number of sequential write preferred zones with a Zone Condition of IMPLICITLY OPEN; and
k is the contents of the OPTIMAL NUMBER OF OPEN SEQUENTIAL WRITE PREFERRED ZONES field (see 6.4.2).

All manage open zone resources operations on sequential write preferred zones set the OZR Available state machine variable to SUCCESS.

4.4.3.2.6.3 Select a sequential write required zone

The device server:

a) should not select a zone to be closed; and
b) shall set the OZR Available state machine variable to SUCCESS (see 4.4.3.2.6.1),

if the following formula evaluates to true:

$$z > (x + y)$$

where:

x is the number of sequential write required zones with a Zone Condition of EXPLICITLY OPEN;
y is the number of sequential write required zones with a Zone Condition of IMPLICITLY OPEN; and
z is the contents of the MAXIMUM NUMBER OF OPEN SEQUENTIAL WRITE REQUIRED ZONES field (see 6.4.2).

The device server shall:

a) set the OZR Available state machine variable to FAILURE (see 4.4.3.2.6.1); and
b) terminate the command that initiated the manage open zone resources operation with CHECK CONDITION status, with sense key set to DATA PROTECT and the additional sense code set to INSUFFICIENT ZONE RESOURCES,

if the following formula evaluates to true:

$$x \geq z$$

where:

x is the number of sequential write required zones with a Zone Condition of EXPLICIT OPEN; and
z is the contents of the MAXIMUM NUMBER OF OPEN SEQUENTIAL WRITE REQUIRED ZONES (see 6.4.2).

The device server shall:

a) choose a zone with a Zone Condition of IMPLICITLY OPENED and perform a close zone operation (see 4.4.3.2.3) on that zone; and
b) set the OZR Available state machine variable to SUCCESS (see 4.4.3.2.6.1),

if the following formula evaluates to true:

$$(z \leq (x + y)) \text{ AND } (y > 0)$$

where:

x is the number of sequential write required zones with a Zone Condition of EXPLICITLY OPEN;
y is the number of sequential write required zones with a Zone Condition of IMPLICITLY OPEN; and
z is the contents of the MAXIMUM NUMBER OF OPEN SEQUENTIAL WRITE REQUIRED ZONES field (see 6.4.2).

4.4.3.2.7 Zoned block device modifications to SBC-4 write operations and read operations

If the Zone Type is:

a) CONVENTIONAL, then zoned block devices perform write operations and read operations as described in SBC-4; or
b) SEQUENTIAL WRITE PREFERRED or SEQUENTIAL WRITE REQUIRED, then zoned block devices perform read operations and write operations:
   A) with the limitation that one read operation or one write operation affects only one zone (e.g., a command that affects multiple zones is split into multiple operations);
   B) as described in the Zone Condition state machine (see 4.4.3.5) based on the Zone Condition of the specified zone;
   C) as described in this subclause; and
   D) as described in SBC-4.

The Zone Condition state machine (see 4.4.3.5) requires the specified zone to have a Zone Condition of EXPLICITLY OPENED or IMPLICITLY OPENED before a write operation is performed. If a zone with a Zone Condition of EMPTY or CLOSED is specified for a write operation, prior to processing the write operation, then the Zone Condition state machine requires that:

a) a manage open zone resources operation (see 4.4.3.2.6) be performed; and
b) the Zone Condition become IMPLICITLY OPENED.

The result of a successful write operation on a zone with a Zone Condition of EMPTY or CLOSED is a zone with a Zone Condition of IMPLICITLY OPENED.

The Zone Condition state machine may or may not require a read operation on a zone with a Zone Condition of EMPTY or CLOSED to be processed without an error. The result of a successful read operation on a zone with a Zone Condition of EMPTY or CLOSED does not change the Zone Condition of that zone.

4.4.3.3 Sequential write preferred zone model

4.4.3.3.1 Writing in sequential write preferred zones

A sequential write preferred zone is a write pointer zone (see 4.4.3.1) in which the device server allows write commands that specify a starting LBA that is not equal to the associated zone's write pointer. The device server may allocate non-sequential write zone resources for processing these writes, see 4.3.6.

A write command that starts at the write pointer of a sequential write preferred zone and ends at the highest LBA of that sequential write preferred zone results in the Zone Condition changing to FULL.

Figure 6 shows an example of a write command that starts at the write pointer of a sequential write preferred zone.

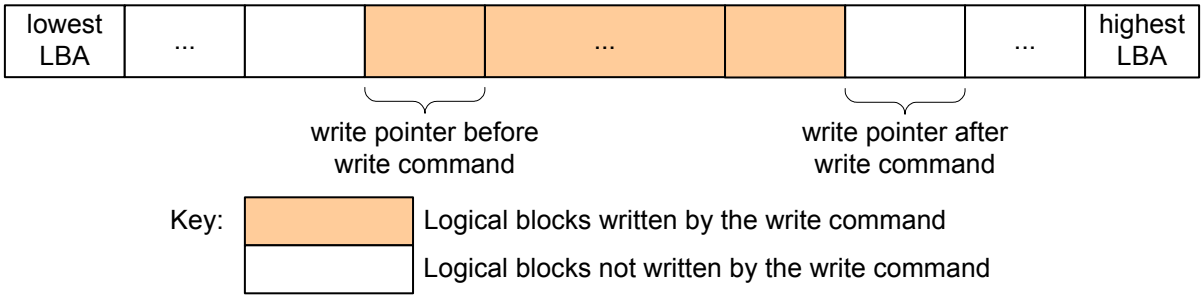

Figure 6 — Example write command that starts at the write pointer

Figure 7 shows examples of write commands that do not start at the write pointer (i.e., a non-sequential write operation).

A non-sequential write operation shall result in:

a) the associated zone's write pointer being set to an LBA that is:
   A) greater than or equal to the value of that write pointer before that operation; and
   B) less than or equal to the minimum of:
      a) one plus the highest LBA written in the zone since the last time the Zone Condition was EMPTY; and
      b) the highest LBA in the zone;

   or

b) the Zone Condition becoming FULL (e.g., the write pointer becoming invalid).

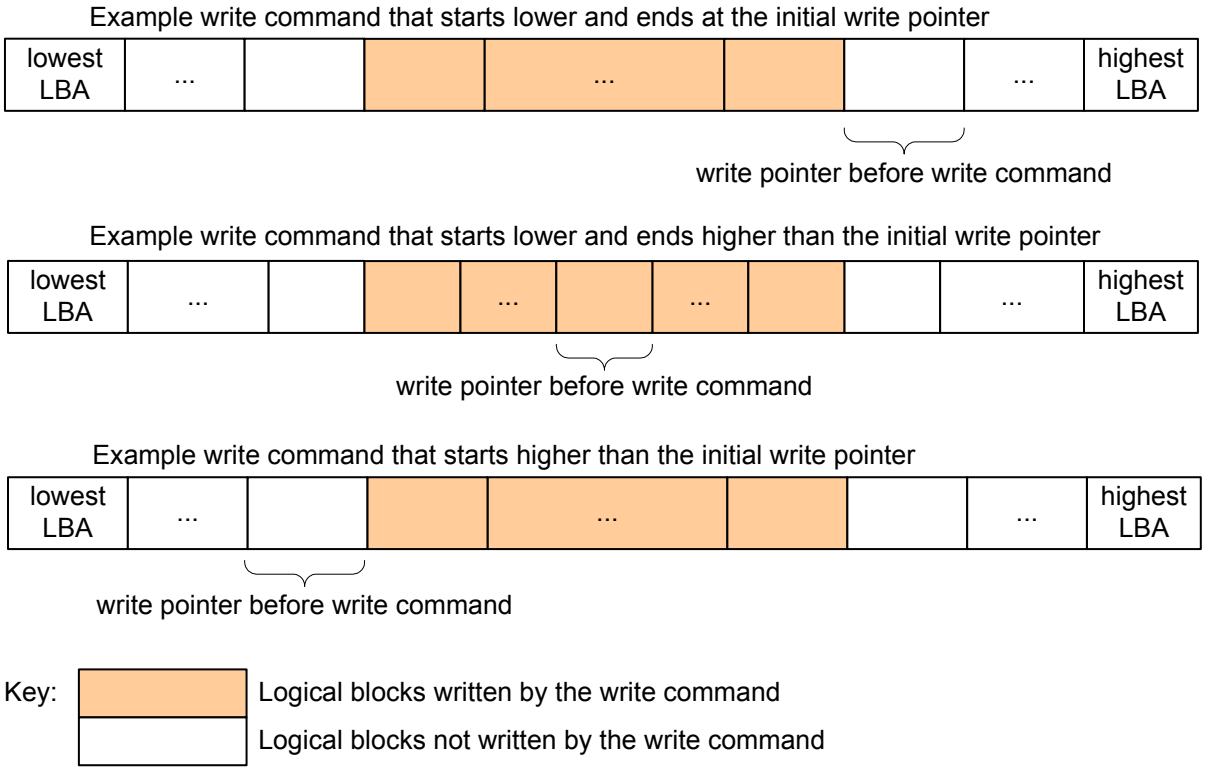

Figure 7 — Examples of write commands that do not start at the write pointer

As described in the Zone Condition state machine (see 4.4.3.5), a non-sequential write operation to a sequential write preferred zone may result in the Zone Condition becoming FULL. If a non-sequential write command completes without error and causes the Zone Condition to become FULL, then the device server shall complete that command or a subsequent command accessing that zone with GOOD status with the sense key set to COMPLETED and the additional sense code set to ZONE TRANSITION TO FULL.

The device server may incur delays during the processing of subsequent commands if the device server processes a command that results in:

a) the number of open sequential write preferred zones exceeding the value in the OPTIMAL NUMBER OF OPEN SEQUENTIAL WRITE PREFERRED ZONES field (see 6.4.2); or
b) the number of non-sequentially written sequential write preferred zones exceeding the value in the OPTIMAL NUMBER OF NON-SEQUENTIALLY WRITTEN SEQUENTIAL WRITE PREFERRED ZONES field (see 6.4.2).

Regardless of the starting LBA in a write command, the resulting value of the write pointer is outside the scope of this standard if:

a) at least one logical block has been written;
b) not all the logical blocks specified by the command have been written; and
c) write caching:
   A) is not enabled and the write command is terminated with an error; or
   B) is enabled and an error occurs after the command completes with success.

4.4.3.3.2 Reading in sequential write preferred zones

A read operation on an LBA that has been written since the last reset write pointer operation in that sequential write preferred zone shall return the last written logical block data. A read operation on an LBA that has not been written since the last reset write pointer operation in that sequential write preferred zone shall:

a) return logical block data set to the last initialization pattern that was set as described in 4.4.3.1.3; and

b) not write any data for a read command that has the FUA bit set to one (see SBC-4).

A verify operation on an LBA that has been written since the last reset write pointer operation in that sequential write preferred zone shall use the last written data for that verify operation. A verify operation on an LBA that has not been written since the last reset write pointer operation in that sequential write preferred zone shall use logical block data set to the last initialization pattern that was set as described in 4.4.3.1.3.

4.4.3.4 Sequential write required zone model

4.4.3.4.1 Writing in sequential write required zones

A sequential write required zone is a write pointer zone (see 4.4.3.1) in which the device server requires that write commands specify a starting LBA that is equal to that zone's write pointer.

If the device server processes a write command without error, then the write pointer is modified to indicate the LBA where a subsequent write operation within the zone shall be performed.

If the ending LBA of a write command that completed without error is equal to the highest LBA of a sequential write required zone, then the Zone Condition becomes FULL and the write pointer becomes invalid.

If the device server processes a write command that is terminated with CHECK CONDITION status, then the value of the write pointer may be unknown to the application client. The application client should use the REPORT ZONES command (see 5.6) to determine the current write pointer and other attributes of the zone.

If the value in the MAXIMUM NUMBER OF OPEN SEQUENTIAL WRITE REQUIRED ZONES field is non-zero (see 6.4.2) and the number of zones with a Zone Condition of EXPLICITLY OPENED is equal to the value in the MAXIMUM NUMBER OF OPEN SEQUENTIAL WRITE REQUIRED ZONES field, then a command that writes to or attempts to open a zone with a zone condition of EMPTY or CLOSED is terminated with CHECK CONDITION status with sense key set to DATA PROTECT and the additional sense code set to INSUFFICIENT ZONE RESOURCES as described in 4.4.3.2.6.

If the device server processes a write command with:

a) the starting LBA equal to the write pointer;
b) the ending LBA in the same sequential write required zone; and
c) an ending LBA that is equal to the last logical block within a physical block,

then the device server shall perform write operations for that command.

If the device server processes a write command with:

a) the starting LBA in a sequential write required zone set to a value that is in the write required zone but that is not equal to the write pointer; or
b) an ending LBA that is not equal to the last logical block within a physical block (see SBC-4),

then the device server shall not write any data and shall terminate the command with CHECK CONDITION status, with the sense key set to ILLEGAL REQUEST, and the additional sense code set to UNALIGNED WRITE COMMAND.

If the device server processes a write command with an ending LBA that is not in the same sequential write required zone as the starting LBA, then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to ILLEGAL REQUEST, and the additional sense code set to WRITE BOUNDARY VIOLATION.

If the device server processes a write command with the starting LBA that is not in a sequential write required zone and specifies an LBA that is in a sequential write required zone, then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to ILLEGAL REQUEST, and the additional sense code set to WRITE BOUNDARY VIOLATION.

If the device server terminates a command with the additional sense code set to UNALIGNED WRITE COMMAND or WRITE BOUNDARY VIOLATION, then the device server shall return the write pointer associated

with the sequential write required zone specified by the LOGICAL BLOCK ADDRESS field of that command in the INFORMATION field of the sense data (see SPC-5).

4.4.3.4.2 Reading in sequential write required zones

If the device server processes a read command with:

a) the starting LBA in a sequential write required zone;
b) the ending LBA in the same sequential write required zone; and
c) the ending LBA is less than the write pointer,

then the device server shall perform read operations for that command.

The URSWRZ bit (see 6.4.2) affects the behavior of read commands in sequential write required zones.

If the device server processes a read command:

a) with the ending LBA greater than or equal to the write pointer in the same zone as the starting LBA; and
b) the URSWRZ bit is set to:
   A) one, then the device server shall process the command as described in 4.4.3.4.3; or
   B) zero, then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to ILLEGAL REQUEST, and the additional sense code set to ATTEMPT TO READ INVALID DATA.

If the device server processes a read command with the ending LBA in a different sequential write required zone than the starting LBA, then the device server shall:

a) process the command as described in 4.4.3.4.3 without regard to sequential write required zone boundary, if the URSWRZ bit is set to one and all LBAs indicated by the command are in a sequential write required zone;
b) terminate the command with CHECK CONDITION status, with the sense key set to ILLEGAL REQUEST, and the additional sense code set to READ BOUNDARY VIOLATION, if the URSWRZ bit is set to one and one or more LBAs indicated by the command are in a zone other than a sequential write required zone;

   and

c) terminate the command with CHECK CONDITION status, with the sense key set to ILLEGAL REQUEST, and the additional sense code set to READ BOUNDARY VIOLATION, if the URSWRZ bit is set to zero.

If the device server processes a read command with any LBA that is not in a sequential write required zone and specifies at least one LBA that is in a sequential write required zone, then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to ILLEGAL REQUEST, and the additional sense code set to ATTEMPT TO READ INVALID DATA.

If the device server terminates a command with the additional sense code set to ATTEMPT TO READ INVALID DATA or READ BOUNDARY VIOLATION, then the device server shall return the write pointer associated with the sequential write required zone specified by the LOGICAL BLOCK ADDRESS field of that command in the INFORMATION field of the sense data (see SPC-5).

4.4.3.4.3 Processing unrestricted reads in sequential write required zones

If the URSWRZ bit is set to one (see 6.4.2), then a read operation an LBA that has not been written since the last reset write pointer operation in that LBA's sequential write required zone, shall:

a) return logical block data set to the last initialization pattern that was set as described in 4.4.3.1.3; and
b) not write any data for a read command that has the FUA bit set to one (see SBC-4).

4.4.3.5 Zone condition state machine

4.4.3.5.1 Zone condition state machine overview

There is one zone condition state machine for each write pointer zone. The zone condition state machine (see figure 8) controls the operation of each write pointer zone. The state machine consists of the following states:

a) ZC1:Empty (see 4.4.3.5.2);
b) ZC2:Implicit_Open (see 4.4.3.5.3);
c) ZC3:Explicit_Open (see 4.4.3.5.4);
d) ZC4:Closed (see 4.4.3.5.5);
e) ZC5:Full (see 4.4.3.5.6);
f) ZC6:Read_Only (see 4.4.3.5.7); and
g) ZC7:Offline (see 4.4.3.5.8).

Each zone condition state machine is an integral part of the device server that participates in the processing of specified commands and in the performing of specified operations. The processing of a single command (e.g., a FINISH ZONE command (see 5.3)) may include zero or more transitions in the zone condition state machine. Information about device server progress in the processing of such commands affects transitions in the zone condition state machine.

Unless otherwise specified in the zone condition state machine, the device server shall maintain zone condition state after any condition changes or SCSI events (see SAM-5) except for a power on.

The initial state for a write pointer zone (see 4.4.3.1) after power on is:

a) the ZC1:Empty state if:
   A) the write pointer is valid and indicates the lowest LBA in the zone; and
   B) Non-Sequential Write Resources Active is false;
b) the ZC4:Closed state, if:
   A) the write pointer is valid and indicates an LBA that is not the lowest LBA in the zone; or
   B) Non-Sequential Write Resources Active is true;
c) the ZC5:Full state, if the most recent Zone Condition was FULL;
d) the ZC6:Read_Only state, if the most recent Zone Condition was READ ONLY; and
e) the ZC7:Offline state, if the most recent Zone Condition was OFFLINE.

This state machine shall maintain the OZR Available state machine variable to indicate the success or failure of the most recent manage open zone resources operation (see 4.4.3.2.6).

The characteristics associated with the state of a zone are summarized in table 12.

Table 12 — Characteristics associated with zone state

State	Zone characteristics			
			Accessible for	
	Write pointer valid [a]	Open zone resources in use	Reads	Writes
ZC1:Empty state	Yes	No	see [b]	see [c]
ZC2:Implicit_Open state	Yes	Yes	Yes	Yes
ZC3:Explicit_Open state	Yes	Yes	Yes	Yes
ZC4:Closed state	Yes	No	Yes	see [c]
ZC5:Full state	No	No	Yes	see [d]
ZC6:Read_Only state	No	No	Yes	No
ZC7:Offline state	No	No	No	No

[a] A valid write pointer (i.e., Yes) indicates a specific LBA in the zone as described in 4.4.3.1. An invalid write pointer (i.e., No) provides no information.

[b] This zone is accessible for reads if:
a) the zone type is sequential write preferred (see 4.4.3.3.2); or
b) the zone type is sequential write required and the URSWRZ bit is set to one (see 4.4.3.4.2).

[c] This zone is accessible for writes if the manage open zone resources operation (see 4.4.3.2.6) is able to complete with success and the zone transitions to the ZC2:Implicit_Open state.

[d] This zone is accessible for writes if the zone type is sequential write preferred (see 4.4.3.3).

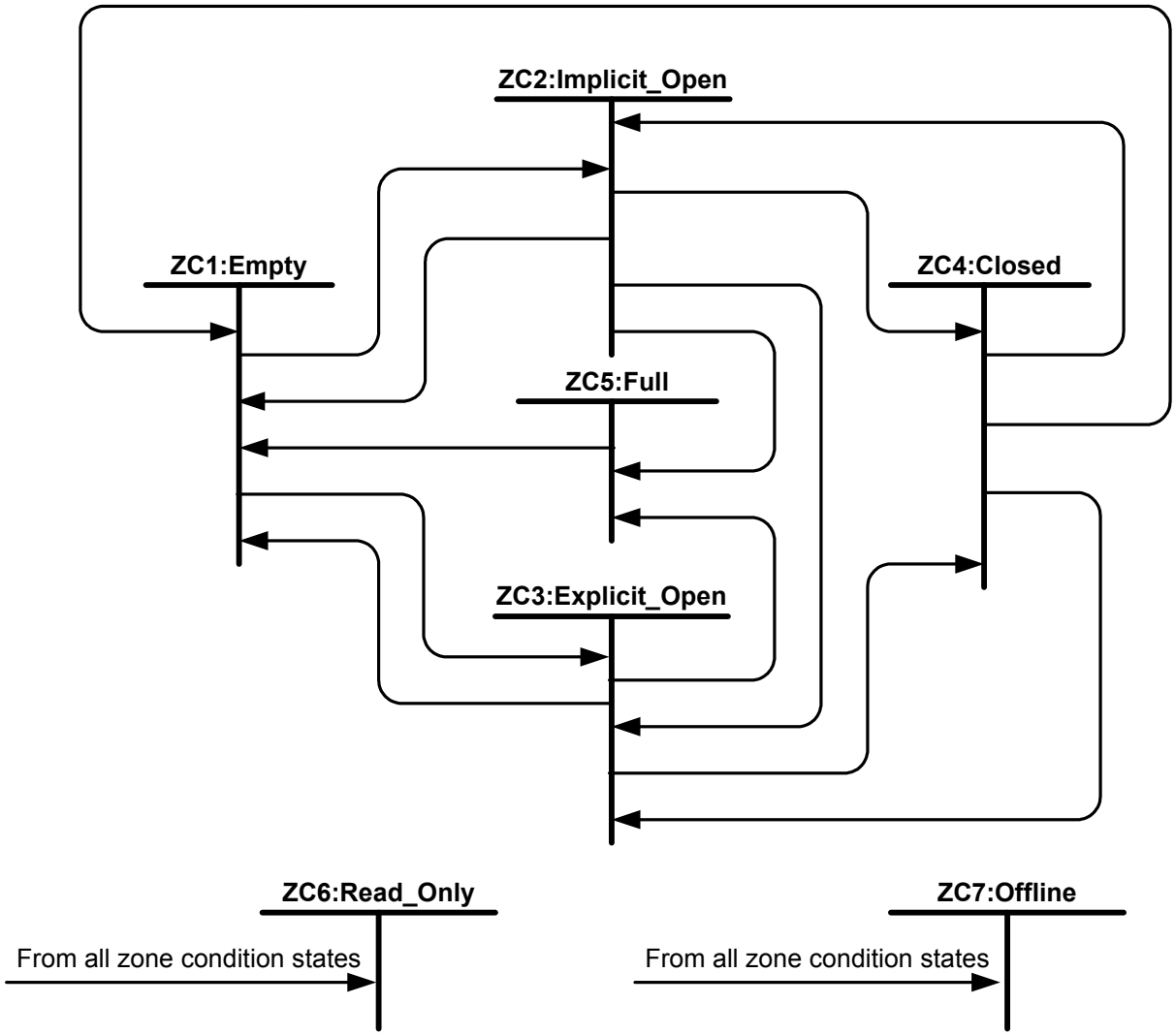

Figure 8 — Zone condition state machine

4.4.3.5.2 ZC1:Empty state

4.4.3.5.2.1 ZC1:Empty state overview

While in this state:

a) Zone Condition shall be set to EMPTY;
b) the write pointer is valid and indicates the lowest LBA in the zone;
c) all zone resources (e.g., the write pointer) shall be persistent zone resources (i.e., open zone resources shall not be used);
d) RWP Recommended is set as defined in 4.3.5;
e) Non-Sequential Write Resources Active is set as defined in 4.3.6;
f) if the device server begins to perform:
   A) a write operation;
   B) a finish zone operation; or
   C) an open zone operation (i.e., as part of processing an OPEN ZONE command),

   then, the device server shall:

   1) perform a manage open zone resources operation (see 4.4.3.2.6), including the setting of the OZR Available state machine variable; and
   2) if the OZR Available state machine variable is set to FAILURE, then the device server shall terminate the command that initiated the manage open zone resources operation with CHECK CONDITION status, with sense key set to DATA PROTECT and the additional sense code set to INSUFFICIENT ZONE RESOURCES;

   and

g) if the device server performs a read operation, then the device server shall perform the read operation and process the read command that resulted in the read operation as described in:
   A) 4.4.3.3.2 for sequential write preferred zones; or
   B) 4.4.3.4.2 for sequential write required zones.

4.4.3.5.2.2 Transition ZC1:Empty to ZC2:Implicit_Open

In order to continue performing the write operation or the finish zone operation, this transition shall occur if:

1) as described in 4.4.3.5.2.1, the device server begins to perform:
   A) a write operation; or
   B) a finish zone operation;

   and

2) the OZR Available state machine variable (see 4.4.3.2.6) is set to SUCCESS.

4.4.3.5.2.3 Transition ZC1:Empty to ZC3:Explicit_Open

This transition shall occur if:

1) as described in 4.4.3.5.2.1, the device server performs an open zone operation (i.e., as part of processing an OPEN ZONE command); and
2) the OZR Available state machine variable (see 4.4.3.2.6) is set to SUCCESS.

4.4.3.5.2.4 Transition ZC1:Empty to ZC6:Read_Only

This transition may occur for reasons outside the scope of this standard.

4.4.3.5.2.5 Transition ZC1:Empty to ZC7:Offline

This transition may occur:

a) as a result of media failure (see 4.5.3); or
b) for reasons outside the scope of this standard.

4.4.3.5.3 ZC2:Implicit_Open state

4.4.3.5.3.1 ZC2:Implicit_Open state overview

While in this state:

a) Zone Condition shall be set to IMPLICITLY OPENED;
b) the write pointer (see 4.4.3.1) is valid;
c) all zone resources (e.g., the write pointer) shall be open zone resources, persistent zone resources, or both (i.e., this zone uses the open zone resources allocated to it);
d) the RWP Recommended zone attribute shall be maintained as defined in 4.3.5;
e) the Non-Sequential Write Resources Active zone attribute shall be maintained as defined in 4.3.6;
f) if the device server performs a write operation, then the device server shall perform the write operation and process the write command that resulted in the write operation as described in:
   A) 4.4.3.3.1 for sequential write preferred zones; or
   B) 4.4.3.4.1 for sequential write required zones;

   and

g) if the device server performs a read operation, then the device server shall perform the read operation and process the read command that resulted in the read operation as described in:
   A) 4.4.3.3.2 for sequential write preferred zones; or
   B) 4.4.3.4.2 for sequential write required zones.

4.4.3.5.3.2 Transition ZC2:Implicit_Open to ZC1:Empty

This transition shall occur:

a) after successful completion of a reset write pointer operation; or
b) if the write pointer is at the lowest LBA of the zone and Non-Sequential Write Resources Active is false, then after successful completion of:
   A) a close zone operation for this zone; or
   B) a manage open zone resources operation for a different zone that selects this zone to be closed.

4.4.3.5.3.3 Transition ZC2:Implicit_Open to ZC3:Explicit_Open

This transition shall occur after successful completion of an open zone operation (i.e., as part of processing an OPEN ZONE command).

4.4.3.5.3.4 Transition ZC2:Implicit_Open to ZC4:Closed

If the write pointer is not at the lowest LBA of the zone or Non-Sequential Write Resources Active is true, then this transition shall occur after successful completion of:

a) a close zone operation for this zone; or
b) a manage open zone resources operation for a different zone that selects this zone to be closed.

4.4.3.5.3.5 Transition ZC2:Implicit_Open to ZC5:Full

For a sequential write required zone, this transition shall occur after successful completion of:

a) a finish zone operation in this zone; or
b) a write operation that writes from the write pointer to the highest LBA in the zone.

For a sequential write preferred zone, this transition:

a) may occur after successful completion of a non-sequential write operation (see 4.4.3.3.1);
b) shall occur after successful completion of a write operation that writes from the write pointer to the highest LBA in this zone; and
c) shall occur after a successful completion of a finish zone operation in this zone.

4.4.3.5.3.6 Transition ZC2:Implicit_Open to ZC6:Read_Only

This transition may occur for reasons outside the scope of this standard (e.g., an error occurs while performing a write operation).

4.4.3.5.3.7 Transition ZC2:Implicit_Open to ZC7:Offline

This transition may occur:

a) as a result of media failure (see 4.5.3); or
b) for reasons outside the scope of this standard.

4.4.3.5.4 ZC3:Explicit_Open state

4.4.3.5.4.1 ZC3:Explicit_Open state overview

While in this state:

a) Zone Condition shall be set to EXPLICITLY OPENED;
b) the write pointer (see 4.4.3.1) is valid;
c) all zone resources (e.g., the write pointer) shall be open zone resources, persistent zone resources, or both (i.e., this zone uses the open zone resources allocated to it);
d) RWP Recommended shall be maintained as defined in 4.3.5;
e) Non-Sequential Write Resources Active shall be maintained as defined in 4.3.6;
f) if the device server performs a write operation, then the device server shall perform the write operation and process the write command that resulted in the write operation as described in:
   A) 4.4.3.3.1 for sequential write preferred zones; or
   B) 4.4.3.4.1 for sequential write required zones;

   and
g) if the device server performs a read operation, then the device server shall perform the read operation and process the read command that resulted in the read operation as described in:
   A) 4.4.3.3.2 for sequential write preferred zones; or
   B) 4.4.3.4.2 for sequential write required zones.

4.4.3.5.4.2 Transition ZC3:Explicit_Open to ZC1:Empty

This transition shall occur after successful completion of:

a) a reset write pointer operation; or
b) a close zone operation, if the write pointer is at the lowest LBA of the zone and Non-Sequential Write Resources Active is false.

4.4.3.5.4.3 Transition ZC3:Explicit_Open to ZC4:Closed

This transition shall occur after successful completion of a close zone operation, if the write pointer is not at the lowest LBA of the zone or Non-Sequential Write Resources Active is true.

4.4.3.5.4.4 Transition ZC3:Explicit_Open to ZC5:Full

For a sequential write required zone, this transition shall occur after successful completion of:

a) a finish zone operation in this zone; or
b) a write operation that writes from the write pointer to the highest LBA in the zone.

For a sequential write preferred zone, this transition:

a) may occur after successful completion of a non-sequential write operation (see 4.4.3.3.1);
b) shall occur after successful completion of a write operation that writes from the write pointer to the highest LBA in this zone; and
c) shall occur after a successful completion of a finish zone operation in this zone.

4.4.3.5.4.5 Transition ZC3:Explicit_Open to ZC6:Read_Only

This transition may occur for reasons outside the scope of this standard (e.g., an error occurs while performing a write operation).

4.4.3.5.4.6 Transition ZC3:Explicit_Open to ZC7:Offline

This transition may occur:

a) as a result of media failure (see 4.5.3); or
b) for reasons outside the scope of this standard.

4.4.3.5.5 ZC4:Closed state

4.4.3.5.5.1 ZC4:Closed state overview

While in this state:

a) Zone Condition shall be set to CLOSED;
b) the write pointer (see 4.4.3.1) is valid;
c) all zone resources (e.g., the write pointer) shall be persistent zone resources (i.e., open zone resources shall not be used);
d) RWP Recommended shall be maintained as defined in 4.3.5;
e) Non-Sequential Write Resources Active shall be maintained as defined in 4.3.6;
f) if the device server begins to perform:
   A) a write operation;
   B) a finish zone operation; or
   C) an open zone operation (i.e., as part of processing an OPEN ZONE command),

   then, the device server shall:

   1) perform a manage open zone resources operation (see 4.4.3.2.6), including the setting of the OZR Available state machine variable; and
   2) if the OZR Available state machine variable is set to FAILURE, then the device server shall terminate the command that initiated the manage open zone resources operation with CHECK CONDITION status, with sense key set to DATA PROTECT and the additional sense code set to INSUFFICIENT ZONE RESOURCES;

   and

g) if the device server performs a read operation, then the device server shall perform the read operation and process the read command that resulted in the read operation as described in:
   A) 4.4.3.3.2 for sequential write preferred zones; or
   B) 4.4.3.4.2 for sequential write required zones.

4.4.3.5.5.2 Transition ZC4:Closed to ZC1:Empty

This transition shall occur after successful completion of a reset write pointer operation.

4.4.3.5.5.3 Transition ZC4:Closed to ZC2:Implicit_Open

In order to continue performing the write operation or the finish zone operation, this transition shall occur if:

1) as described in 4.4.3.5.2.1, the device server begins to perform:
   A) a write operation; or
   B) a finish zone operation;

   and

2) the OZR Available state machine variable (see 4.4.3.2.6) is set to SUCCESS.

4.4.3.5.5.4 Transition ZC4:Closed to ZC3:Explicit_Open

This transition shall occur if:

1) as described in 4.4.3.5.2.1, the device server performs an open zone operation (i.e., as part of processing an OPEN ZONE command); and
2) the OZR Available state machine variable (see 4.4.3.2.6) is set to SUCCESS.

4.4.3.5.5.5 Transition ZC4:Closed to ZC6:Read_Only

This transition may occur for reasons outside the scope of this standard.

4.4.3.5.5.6 Transition ZC4:Closed to ZC7:Offline

This transition may occur:

a) as a result of media failure (see 4.5.3); or
b) for reasons outside the scope of this standard.

4.4.3.5.6 ZC5:Full state

4.4.3.5.6.1 ZC5:Full state overview

While in this state:

a) Zone Condition shall be set to FULL;
b) the write pointer is invalid;
c) all LBAs in the zone are able to return logical block data in response to a read command;
d) all zone resources (e.g., the write pointer) shall be persistent zone resources (i.e., open zone resources shall not be used);
e) RWP Recommended shall be maintained as defined in 4.3.5;
f) Non-Sequential Write Resources Active shall be maintained as defined in 4.3.6;
g) if the device server performs a write operation, then:
   A) for sequential write preferred zones, the device server shall perform the write operation and process the write command that resulted in the write operation as described in 4.4.3.3.1; or
   B) for sequential write required zones, the device server shall terminate the write command with CHECK CONDITION status, the sense key set to ILLEGAL REQUEST and the additional sense code set to INVALID FIELD IN CDB;

   and

h) if the device server performs a read operation, then the device server shall perform the read operation and process the read command that resulted in the read operation as described in:
   A) 4.4.3.3.2 for sequential write preferred zones; or
   B) 4.4.3.4.2 for sequential write required zones.

4.4.3.5.6.2 Transition ZC5:Full to ZC1:Empty

This transition shall occur after successful completion of a reset write pointer operation.

4.4.3.5.6.3 Transition ZC5:Full to ZC6:Read_Only

This transition may occur for reasons outside the scope of this standard.

4.4.3.5.6.4 Transition ZC5:Full to ZC7:Offline

This transition may occur:

a) as a result of media failure (see 4.5.3); or
b) for reasons outside the scope of this standard.

4.4.3.5.7 ZC6:Read_Only state

4.4.3.5.7.1 ZC6:Read_Only state overview

While in this state:

a) Zone Condition shall be set to READ ONLY;
b) the write pointer is invalid;
c) all zone resources (e.g., the write pointer) shall be persistent zone resources (i.e., open zone resources shall not be used);
d) RWP Recommended is set as defined in 4.3.5;
e) Non-Sequential Write Resources Active is set as defined in 4.3.6;
f) if the device server performs a write operation or a finish zone operation, then the device server shall terminate the command that resulted in the operation with CHECK CONDITION status, with the sense key set to DATA PROTECT, and additional sense code set to ZONE IS READ ONLY; and
g) if the device server performs a read operation, then the device server shall perform the read operation and process the read command that resulted in the read operation as described in:
   A) 4.4.3.3.2 for sequential write preferred zones; or
   B) 4.4.3.4.2 for sequential write required zones.

With the exception of transitions to the ZC7:Offline state, transitions out of this state are outside the scope of this standard.

4.4.3.5.7.2 Transition ZC6:Read_Only to ZC7:Offline

This transition may occur:

a) as a result of media failure (see 4.5.3);
b) as the result of performing a sanitize operation (see 4.7); or
c) for reasons outside the scope of this standard.

4.4.3.5.8 ZC7:Offline state

While in this state:

a) Zone Condition shall be set to OFFLINE;
b) the write pointer is invalid;
c) all zone resources (e.g., the write pointer) shall be persistent zone resources (i.e., open zone resources shall not be used);
d) RWP Recommended is set as defined in 4.3.5;
e) Non-Sequential Write Resources Active is set as defined in 4.3.6; and
f) if the device server performs a read operation, a write operation, or a finish zone operation, then the device server shall terminate the command that resulted in the operation with CHECK CONDITION status, with the sense key set to MEDIUM ERROR, and additional sense code set to ZONE IS OFFLINE.

All transitions out of this state are outside the scope of this standard.

4.5 Zoned block device extensions to block device model

4.5.1 Zoned block device internal resource management

A zoned block device requires internal resources (e.g., persistent zone resources) to maintain each zone. Insufficient resources may result in degraded functionality (e.g., reduced performance, increased power consumption, or increased reporting of write errors).

A RESET bit set to one in a zone descriptor (see 5.6.2) indicates that an insufficient resources condition has been detected for that zone.

An application client that detects the RESET bit being set to one may respond by sending a RESET WRITE POINTER command (see 5.7) that specifies that zone. Before sending the RESET WRITE POINTER command, the application client may copy the logical block data in the affected zone to another location.

4.5.2 Unexpected power removal

If power is removed from a zoned block device prior to the device server completing the processing of a power loss expected event (see SAM-5) and:

a) there are partially completed write operations;
b) there is logical block data in volatile write cache; or
c) zone attributes for completed write operations are not stored in persistent zone resources,

then an unexpected power removal condition has occurred.

An unexpected power removal condition may result in zone attributes for partially completed write commands to be updated to reflect the partial logical block data written to the medium (e.g., a write pointer may indicate the LBA plus one of the last logical block that the device server attempted to write before the unexpected power removal).

If:

1) write caching is disabled (see SBC-4);
2) a write command completes without error; and
3) an unexpected power removal occurs,

then, after power is restored, a read command that completes without error for an LBA included in that write command shall return the logical block data that was written by that write command before the unexpected power removal.

If:

1) a write command with the FUA bit set to one (see SBC-4) completes without error; and
2) an unexpected power removal occurs,

then, after power is restored, a read command that completes without error for an LBA included in that write command shall return the logical block data that was written by that write command before the unexpected power removal.

If:

1) a write command completes without error;
2) a SYNCHRONIZE CACHE command (see SBC-4) completes without error; and
3) an unexpected power removal occurs,

then, after power is restored, a read command that completes without error for an LBA included in that write command shall return the logical block data that was written by that write command before the unexpected power removal.

If:

1) a verify command (see SBC-4) completes without error; and
2) an unexpected power removal occurs,

then, after power is restored, a read command that completes without error for an LBA included in that verify command shall return the logical block data that was accessed by that verify command before the unexpected power removal.

For sequential write required zones, if an unexpected power loss results in unwritten logical blocks in LBAs lower than the write pointer for a zone and read operations access the unwritten logical blocks for that zone, then the device server shall terminate the operation with CHECK CONDITION status with sense key set to MEDIUM ERROR with additional sense code set to UNWRITTEN DATA IN ZONE.

After an unexpected power removal condition, the parameter data returned by a REPORT ZONES command (see 5.6) may contain write pointers that do not correspond to the ending LBA plus one of the last command to report completion prior to the unexpected power removal condition.

4.5.3 Media failure

If a host managed zoned block device has a media failure that affects one or more zones, then the device server may indicate that those zones are unavailable for reading or writing by changing the Zone Condition to OFFLINE.

If the device server terminates a write command to a write pointer zone with the additional sense code set to:

a) WRITE ERROR - RECOVERY NEEDED;
b) WRITE ERROR - RECOVERY SCAN NEEDED; or
c) WRITE ERROR - INSUFFICIENT ZONE RESOURCES,

then the device server shall return the lowest numbered LBA where a media failure may have occurred during processing of that command in the INFORMATION field of the sense data (see SPC-5). If the additional sense code is set to WRITE ERROR - RECOVERY SCAN NEEDED, then the application client should read all the logical blocks from the LBA indicated by the INFORMATION field to the LBA indicated by the write pointer minus one and take recovery actions for LBAs that return errors.

4.6 Format operations

All of the functions defined for the FORMAT UNIT command (see SBC-4) are available for zoned block devices. A FORMAT UNIT command affects all zones, however some characteristics are dependent on the zone type (see 4.4) present in the zoned block device. A zoned block device may support multiple types of zones.

For each conventional zone (see 4.4.2), a format operation is performed as specified in SBC-4.

For each write pointer zone (see 4.4.3), a format operation is performed as specified in SBC-4 and a reset write pointer operation (see 4.4.3.1) is performed.

4.7 Sanitize operations

See SBC-4 for the definition of the SANITIZE command. A SANITIZE operation affects all zones, however some characteristics are dependent on the zone types (see 4.4) present in the zoned block device. The ZNR bit in the SANITIZE command controls whether a reset write pointer zone operation is performed on each write pointer zone as part of the sanitize operation.

4.8 Reservations

Reservation restrictions are placed on commands as a result of access qualifiers associated with the type of reservation. See SPC-5 for a description of reservations for commands described in that standard. See SBC-4 for a description of reservations for commands described in that standard. See table 13 for a description of reservations for commands described in this standard.

Commands from I_T nexuses holding a reservation should complete normally. Table 13 specifies the behavior of commands from registered I_T nexuses when a registrants only or all registrants type persistent reservation is present.

For each command in table 13, this standard or SPC-5 defines the conditions that result in the device server completing the command with RESERVATION CONFLICT status.

Table 13 — ZBC commands that are allowed in the presence of various reservations

Command	Addressed logical unit has this type of persistent reservation held by another I_T nexus				
	From any I_T nexus		From registered I_T nexus (RR all types)	From I_T nexus not registered	
	Write Exclusive	Exclusive Access		Write Exclusive - RR	Exclusive Access - RR
CLOSE ZONE	Conflict	Conflict	Allowed	Conflict	Conflict
FINISH ZONE	Conflict	Conflict	Allowed	Conflict	Conflict
OPEN ZONE	Conflict	Conflict	Allowed	Conflict	Conflict
REPORT ZONES	Allowed	Allowed	Allowed	Allowed	Allowed
RESET WRITE POINTER	Conflict	Conflict	Allowed	Conflict	Conflict

Key:
RR = Registrants Only or All Registrants
Allowed = Commands received from I_T nexuses not holding the reservation or from I_T nexuses not registered when a registrants only or all registrants type persistent reservation is present should complete normally.
Conflict = Commands received from I_T nexuses not holding the reservation or from I_T nexuses not registered when a registrants only or all registrants type persistent reservation is present shall not be performed, and the device server shall complete the command with RESERVATION CONFLICT status.

4.9 Caches

4.9.1 Caches overview

Zoned block devices may implement caches. A cache is an area of temporary storage in the zoned block device (e.g., to enhance performance) separate from the medium that is not directly accessible by the application client.

A cache stores logical block data.

A cache may be volatile or non-volatile. A volatile cache does not retain data through power cycles. A non-volatile cache retains data through power cycles. There may be a limit on the amount of time a non-volatile cache is able to retain data without power.

The cache model description for zone block devices is as specified in SBC-4 with additional requirements as specified in 4.9.

4.9.2 Write caching

While processing write commands, as a result of using write-back caching there is a period of time during which the logical block data may be lost if:

a) an unexpected power removal occurs (see 4.5.2); or
b) a hardware failure occurs.

If an error occurs during a write medium operation in a sequential write required zone and that error is reported as a deferred error, then the device server may invalidate cached logical block data for LBAs in that zone that are larger than the LBA reported with the deferred error. This invalidation may occur for data cached in both volatile and non-volatile caches.

4.9.3 Command interactions with caches

If the FUA bit is set to one in a read command or a write command that specifies LBAs in a sequential write required zone and any logical blocks in that zone are cached in a volatile cache, then the device server shall write all such logical blocks to the medium or non-volatile cache.

4.9.4 Write operation and write medium operation interactions with caches

When the device server performs a write cache operation that updates cached logical block data for a write pointer zone, the device server shall update the write pointer to reflect the completed write operation. If an unexpected power removal occurs (see 4.5.2), then after power is restored, the value of every write pointer shall reflect the state of the medium and non-volatile cache, if any (e.g., a write pointer value may become smaller as a result of data lost from volatile cache).

4.9.5 Close zone and finish zone operation interactions with cache

As part of performing a close zone operation or the zone becomes full (e.g., the device completes a finish zone operation), if any logical blocks in that zone are cached in a volatile cache, the device server shall write those cached logical blocks to the medium or to non-volatile cache.

5 Commands for zoned block devices

5.1 Commands for zoned block devices overview

The commands for host aware zoned block devices are listed in table 6 and SBC-4.

The commands for host managed zoned block devices are listed in table 14.

Table 14 — Commands for host managed zoned block devices (part 1 of 2)

Command	Operation code	Type	LBACT	Reference
ATA PASS-THROUGH (12)	A1h	O	n/a	SAT-4
ATA PASS-THROUGH (16)	85h	O	n/a	SAT-4
CLOSE ZONE	94h/01h	M	Z	5.2
FINISH ZONE	94h/02h	M	Z	5.3
FORMAT UNIT	04h	O	Z	SBC-4
INQUIRY	12h	M	n/a	SPC-5
LOG SELECT	4Ch	O	n/a	SPC-5
LOG SENSE	4Dh	M	n/a	SPC-5
MODE SELECT (10)	55h	M	n/a	SPC-5
MODE SENSE (10)	5Ah	M	n/a	SPC-5
OPEN ZONE	94h/03h	M	Z	5.4
PERSISTENT RESERVE IN	5Eh	O	n/a	SPC-5
PERSISTENT RESERVE OUT	5Fh	O	n/a	SPC-5
READ (16)	88h	M	R	SBC-4
READ BUFFER (10)	3C	O	n/a	SPC-5
READ BUFFER (16)	9Bh	O	n/a	SPC-5
READ CAPACITY (16)	9Eh/10h	M	n/a	5.5
READ DEFECT DATA (12)	B7h	O	n/a	SBC-4
REPORT LUNS	A0h	M	n/a	SPC-5

Key:
O = optional
M = mandatory
R = read command
U = unmap command
V = verify command
W = write command
Z = other command
LBACT = logical block access command type (see SBC-4)

Table 14 — Commands for host managed zoned block devices (part 2 of 2)

Command	Operation code	Type	LBACT	Reference
REPORT SUPPORTED OPERATION CODES	A3h/0Ch	M	n/a	SPC-5
REPORT SUPPORTED TASK MANAGEMENT FUNCTIONS	A3h/0Dh	M	n/a	SPC-5
REPORT TIMESTAMP	A3/0Fh	O	n/a	SPC-5
REPORT ZONES	95h/00h	M	Z	5.6
REQUEST SENSE	03h	M	n/a	SPC-5
RESET WRITE POINTER	94h/04h	M	Z	5.7
SET TIMESTAMP	A4h/0Fh	O	n/a	SPC-5
SANITIZE	48h	O	Z	SBC-4
SECURITY PROTOCOL IN	A2h	O	n/a	SPC-5
SECURITY PROTOCOL OUT	B5h	O	n/a	SPC-5
SEND DIAGNOSTIC	1Dh	O	n/a	SPC-5
START STOP UNIT	1Bh	M	n/a	SBC-4
SYNCHRONIZE CACHE (16)	91h	M	W	SBC-4
TEST UNIT READY	00h	M	n/a	SPC-5
VERIFY (16)	8Fh	O	V,W	SBC-4
WRITE (16)	8Ah	M	W	SBC-4
WRITE BUFFER	3Bh	O	n/a	SPC-5
WRITE SAME (16)	93h	M	U, W	SBC-4

Key:
O = optional
M = mandatory
R = read command
U = unmap command
V = verify command
W = write command
Z = other command
LBACT = logical block access command type (see SBC-4)

5.2 CLOSE ZONE command

The CLOSE ZONE command (see table 15) requests the device server to perform close zone operations (see 4.4.3.2.3).

Table 15 — CLOSE ZONE command

Bit / Byte	7	6	5	4	3	2	1	0
0	OPERATION CODE (94h)							
1	Reserved			SERVICE ACTION (01h)				
2	(MSB)							
•••	ZONE ID							
9								(LSB)
10	Reserved							
•••								
13								
14	Reserved							ALL
15	CONTROL							

The OPERATION CODE field and the SERVICE ACTION field are defined in SPC-5 and shall be set to the values shown in table 15 for the CLOSE ZONE command.

If the ALL bit is set to zero, then the ZONE ID field specifies the lowest LBA of the write pointer zone on which the device server shall perform a close zone operation. If the ALL bit is set to one, then the device server shall ignore the ZONE ID field.

If the ALL bit is set to zero and the ZONE ID field:

a) does not specify the lowest LBA of a write pointer zone (see 4.4.3), then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to ILLEGAL REQUEST and the additional sense code set to INVALID FIELD IN CDB; or
b) specifies the lowest LBA of a write pointer zone (see 4.4.3) and for the zone specified by the ZONE ID field, the Zone Condition is:
   A) READ ONLY, then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to DATA PROTECT and additional sense code set to ZONE IS READ ONLY;
   B) OFFLINE, then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to DATA PROTECT and additional sense code set to ZONE IS OFFLINE;
   C) EMPTY, CLOSED, or FULL, then the device server shall make no changes in the Zone Condition and not return an error; or
   D) IMPLICITLY OPENED or EXPLICITLY OPENED, then the device server shall perform a close zone operation (see 4.4.3.2.3) on the zone specified by the ZONE ID field.

An ALL bit set to one specifies that the device server shall perform a close zone operation (see 4.4.3.2.3) on each zone with a Zone Condition of EXPLICIT OPEN or IMPLICIT OPEN.

The CONTROL byte is defined in SAM-5.

5.3 FINISH ZONE command

The FINISH ZONE command (see table 16) requests the device server to perform finish zone operations (see 4.4.3.2.4).

The device server returns the initialization pattern for all unwritten LBAs in this zone (see 4.4.3.3 and 4.4.3.4.3) in response to a read operation. The device server may write the initialization pattern to the media for unwritten LBAs.

Table 16 — FINISH ZONE command

Bit / Byte	7	6	5	4	3	2	1	0
0	OPERATION CODE (94h)							
1	Reserved			SERVICE ACTION (02h)				
2	(MSB)							
•••	ZONE ID							
9								(LSB)
10	Reserved							
•••								
13								
14	Reserved							ALL
15	CONTROL							

The OPERATION CODE field and the SERVICE ACTION field are defined in SPC-5 and shall be set to the values shown in table 16 for the FINISH ZONE command.

If the ALL bit is set to zero, then the ZONE ID field specifies the lowest LBA of the write pointer zone on which the device server shall perform a finish zone operation. If the ALL bit is set to one, then the device server shall ignore the ZONE ID field.

If the ALL bit is set to zero and the ZONE ID field:

a) does not specify the lowest LBA of a write pointer zone (see 4.4.3), then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to ILLEGAL REQUEST and the additional sense code set to INVALID FIELD IN CDB; or
b) specifies the lowest LBA of a write pointer zone (see 4.4.3) and for the zone specified by the ZONE ID field, the Zone Condition is:
   A) READ ONLY, then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to DATA PROTECT and additional sense code set to ZONE IS READ ONLY;
   B) OFFLINE, then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to DATA PROTECT and additional sense code set to ZONE IS OFFLINE;
   C) FULL, then the device server shall make no changes in the Zone Condition and not return an error; or
   D) IMPLICITLY OPENED, EXPLICITLY OPENED, CLOSED, or EMPTY, then the device server shall perform a finish zone operation (see 4.4.3.2.4) on the zone specified by the ZONE ID field.

An ALL bit set to one specifies that the device server shall perform a finish zone operation (see 4.4.3.2.4) on each zone with a Zone Condition of EXPLICIT OPEN, IMPLICIT OPEN, or CLOSED.

The CONTROL byte is defined in SAM-5.

5.4 OPEN ZONE command

The OPEN ZONE command (see table 17) requests the device server to perform open zone operations (see 4.4.3.2.2).

Table 17 — OPEN ZONE command

Bit / Byte	7	6	5	4	3	2	1	0
0	OPERATION CODE (94h)							
1	Reserved			SERVICE ACTION (03h)				
2	(MSB)							
•••	ZONE ID							
9								(LSB)
10	Reserved							
•••								
13								
14	Reserved							ALL
15	CONTROL							

The OPERATION CODE field and the SERVICE ACTION field are defined in SPC-5 and shall be set to the values shown in table 17 for the OPEN ZONE command.

If the ALL bit is set to zero, then the ZONE ID field specifies the lowest LBA of the write pointer zone on which the device server shall perform an open zone operation. If the ALL bit is set to one, then the device server shall ignore the ZONE ID field.

If the ALL bit is set to zero and the ZONE ID field:

a) does not specify the lowest LBA of a write pointer zone (see 4.4.3), then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to ILLEGAL REQUEST and the additional sense code set to INVALID FIELD IN CDB; or
b) specifies the lowest LBA of a write pointer zone (see 4.4.3) and for the zone specified by the ZONE ID field, the Zone Condition is:
   A) READ ONLY, then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to DATA PROTECT and additional sense code set to ZONE IS READ ONLY;
   B) OFFLINE, then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to DATA PROTECT and additional sense code set to ZONE IS OFFLINE;
   C) EXPLICITLY OPENED or FULL, then the device server shall make no changes in the Zone Condition and not return an error; or
   D) IMPLICITLY OPENED, CLOSED, or EMPTY, then the device server shall perform an open zone operation (see 4.4.3.2.2) on the zone specified by the ZONE ID field.

An ALL bit set to one specifies that the device server shall perform an open zone operation (see 4.4.3.2.2) on each zone with a Zone Condition of CLOSED. If the ALL bit is set to one, then the device server shall ignore the ZONE ID field.

For a host managed zoned block device, if the ALL bit is set to one and the number of zones with a Zone Condition of EXPLICIT OPEN plus the number of zones with a Zone Condition of CLOSED is greater than the maximum number of open sequential write required zones, then the device server shall terminate the command

with CHECK CONDITION status, with sense key set to DATA PROTECT and the additional sense code set to INSUFFICIENT ZONE RESOURCES.

The CONTROL byte is defined in SAM-5.

5.5 READ CAPACITY (16) command

5.5.1 READ CAPACITY (16) command overview

The READ CAPACITY (16) command is defined in SBC-4.

For a zoned block device, the READ CAPACITY (16) parameter data is defined in 5.5.2.

5.5.2 READ CAPACITY (16) parameter data

The READ CAPACITY (16) parameter data is defined in table 18. Any time the READ CAPACITY (16) parameter data changes, the device server should establish a unit attention condition (see SBC-4).

Table 18 — READ CAPACITY (16) parameter data

Bit / Byte	7	6	5	4	3	2	1	0
0	(MSB)							
•••	RETURNED LOGICAL BLOCK ADDRESS							
7								(LSB)
8	(MSB)							
•••	LOGICAL BLOCK LENGTH IN BYTES							
11								(LSB)
12	Reserved			RC BASIS		P_TYPE		PROT_EN
13	P_I_EXPONENT				LOGICAL BLOCKS PER PHYSICAL BLOCK EXPONENT			
14	LBPME	LBPRZ	(MSB)	LOWEST ALIGNED LOGICAL BLOCK ADDRESS				
15								(LSB)
16	Reserved							
•••								
31								

The LOGICAL BLOCK LENGTH IN BYTES field, P_TYPE field, PROT_EN bit, P_I_EXPONENT field, LOGICAL BLOCKS PER PHYSICAL BLOCK EXPONENT field, LBPME bit, LBPRZ bit, and LOWEST ALIGNED LOGICAL BLOCK ADDRESS field are defined in SBC-4.

The RC BASIS field indicates the meaning of the value returned by the RETURNED LOGICAL BLOCK ADDRESS field in the READ CAPACITY (16) parameter data and is described in table 19.

Table 19 — RC BASIS field

Code	Description
00b	The RETURNED LOGICAL BLOCK ADDRESS field indicates the highest LBA of a contiguous range of zones that are not sequential write required zones starting with the first zone.
01b	The RETURNED LOGICAL BLOCK ADDRESS field indicates the LBA of the last logical block on the logical unit.
10b	Reserved
11b	Reserved

5.6 REPORT ZONES command

5.6.1 REPORT ZONES command introduction

The REPORT ZONES command (see table 20) requests that the device server transfer parameter data describing the zone structure of the zoned block device.

Table 20 — REPORT ZONES command

Bit / Byte	7	6	5	4	3	2	1	0
0	OPERATION CODE (95h)							
1	Reserved			SERVICE ACTION (00h)				
2	(MSB)							
•••	ZONE START LBA							
9								(LSB)
10	(MSB)							
•••	ALLOCATION LENGTH							
13								(LSB)
14	PARTIAL	Reserved		REPORTING OPTIONS				
15	CONTROL							

The OPERATION CODE field and the SERVICE ACTION field are defined in SPC-5 and shall be set to the values shown in table 20 for the REPORT ZONES command.

The ZONE START LBA field specifies an LBA in the first zone to be reported. If the ZONE START LBA field does not specify the lowest LBA of a zone, then the device server uses the lowest LBA of the zone that contains the specified LBA to specify the first zone to be reported. If the ZONE START LBA field specifies an LBA that is greater than the value in the MAXIMUM LBA field (see table 24), then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to ILLEGAL REQUEST, and the additional sense code set to LOGICAL BLOCK ADDRESS OUT OF RANGE.

The ALLOCATION LENGTH field is defined in SPC-5.

The PARTIAL bit modifies the definition of the ZONE LIST LENGTH field as described in 5.6.2.

The REPORTING OPTIONS field (see table 21) specifies the information to be returned in the parameter data.

Table 21 — REPORT ZONES REPORTING OPTIONS field

Code	Description
00h	List all of the zones in the zoned block device.
01h	List the zones with a Zone Condition of EMPTY.
02h	List the zones with a Zone Condition of IMPLICITLY OPENED.
03h	List the zones with a Zone Condition of EXPLICITLY OPENED.
04h	List the zones with a Zone Condition of CLOSED.
05h	List the zones with a Zone Condition of FULL.
06h	List the zones with a Zone Condition of READ ONLY.
07h	List the zones with a Zone Condition of OFFLINE.
08h to 0Fh	Reserved
10h	List of the zones with RWP Recommended set to true.
11h	List of the zones with Non-Sequential Write Resources Active set to true.
12h to 3Eh	Reserved
3Fh	List of the zones with a Zone Condition of NOT WRITE POINTER.

The CONTROL byte is defined in SAM-5.

5.6.2 REPORT ZONES parameter data

The REPORT ZONES parameter data is defined in table 22.

Table 22 — REPORT ZONES parameter data

Bit / Byte	7	6	5	4	3	2	1	0
0	(MSB)							
•••	ZONE LIST LENGTH (n–63)							
3								(LSB)
4	Reserved				SAME			
5	Reserved							
•••								
7								
8	(MSB)							
•••	MAXIMUM LBA							
15								(LSB)
16	Reserved							
•••								
63								
Zone descriptors list								
64	Zone descriptor [first]							
•••								
127								
	⋮							
n–63	Zone descriptor [last]							
•••								
n								

The ZONE LIST LENGTH field shall contain the length in bytes of the zone descriptors list. The zone descriptors list is the list of zones that:

a) meet the requirements of the REPORTING OPTIONS field; and
b) include the LBA specified by the ZONE START LBA field or have a lowest LBA that is greater than the LBA specified by the ZONE START LBA field.

If the PARTIAL bit is set to zero, then the content of the ZONE LIST LENGTH field is not altered based on the allocation length (see SPC-5). If the PARTIAL bit is set to one then the ZONE LIST LENGTH field shall be set to the lesser of:

a) the allocation length minus 64 if the allocation length is greater than 64;

b) zero if the allocation length is less than or equal to 64; or
c) the length of the zone descriptors list.

The SAME field is defined in table 23. If the ZONE LIST LENGTH field is zero then the SAME field is invalid and should be ignored by the application client.

Table 23 — SAME field description

Code	Description
0h	The zone type and zone length in each zone descriptor may be different.
1h	The zone type and zone length in each zone descriptor are equal to the zone type and zone length indicated in the first zone descriptor in the zone descriptor list.
2h	The zone type in each zone descriptor is equal to the zone type indicated in the first zone descriptor in the zone descriptor list. The zone length of each zone except the last zone is equal to the zone length of the first zone descriptor in the zone descriptor list. The zone length of the last zone descriptor is different than the zone length of the first descriptor in the zone descriptor list.
3h	The zone type in each descriptor may be different. The zone length in each zone descriptor is equal to the zone length indicated in the first zone descriptor in the zone descriptor list.
4h to Fh	Reserved

The MAXIMUM LBA field contains the LBA of the last logical block on the logical unit.

The zone descriptors list contains zone descriptors that shall be sorted in ascending order based on the ZONE START LBA field of each zone descriptor.

Each zone descriptor (see table 24) contains the description of a single zone.

Table 24 — Zone descriptor format

Bit Byte	7	6	5	4	3	2	1	0
0	Reserved				ZONE TYPE			
1	ZONE CONDITION				Reserved		NON_SEQ	RESET
2	Reserved							
•••								
7								
8	(MSB)							
•••	ZONE LENGTH							
15								(LSB)
16	(MSB)							
•••	ZONE START LBA							
23								(LSB)
24	(MSB)							
•••	WRITE POINTER LBA							
31								(LSB)
32	Reserved							
•••								
63								

The ZONE TYPE field indicates the Zone Type (see 4.3.2) of zone as described in table 25.

Table 25 — Zone descriptor ZONE TYPE field

Code	Zone Type
0h	Reserved
1h	CONVENTIONAL
2h	SEQUENTIAL WRITE REQUIRED
3h	SEQUENTIAL WRITE PREFERRED
4h to Fh	Reserved

The ZONE CONDITION field indicates the Zone Condition (see 4.3.3) of the zone as described in table 26.

Table 26 — Zone descriptor ZONE CONDITION field

Code	Description	The content of the WRITE POINTER LBA field is invalid
0h	NOT WRITE POINTER	yes
1h	EMPTY	no
2h	IMPLICITLY OPENED	no
3h	EXPLICITLY OPENED	no
4h	CLOSED	no
5h to Ch	Reserved	
Dh	READ ONLY	yes
Eh	FULL	yes
Fh	OFFLINE	yes

The value of the non-sequential (NON_SEQ) bit is based on the Non-Sequential Write Resources Active zone attribute (see 4.3.6). If Non-Sequential Write Resources Active is:

a) false, then the NON_SEQ bit shall be set to zero; or
b) true, then the NON_SEQ bit shall be set to one.

The value of the RESET bit is based on the RWP Recommended zone attribute (see 4.3.5). If RWP Recommended is:

a) false, then the RESET bit shall be set to zero; or
b) true, then the RESET bit shall be set to one.

The ZONE LENGTH field indicates the number of logical blocks in this zone.

The ZONE START LBA field indicates the lowest LBA in this zone.

The WRITE POINTER LBA field indicates the starting LBA that the application client should specify in the next write command associated with this zone (i.e., the write pointer). The content of the WRITE POINTER LBA field is invalid if the content of the ZONE CONDITION field (see table 26) indicates that the WRITE POINTER LBA field is invalid.

5.7 RESET WRITE POINTER command

The RESET WRITE POINTER command (see table 27) requests the device server to perform reset write pointer operations (see 4.4.3.2.5).

Table 27 — RESET WRITE POINTER command

Bit / Byte	7	6	5	4	3	2	1	0
0	OPERATION CODE (94h)							
1	Reserved			SERVICE ACTION (04h)				
2	(MSB)							
•••	ZONE ID							
9								(LSB)
10	Reserved							
•••								
13								
14	Reserved							ALL
15	CONTROL							

The OPERATION CODE field and the SERVICE ACTION field are defined in SPC-5 and shall be set to the values shown in table 27 for the RESET WRITE POINTER command.

If the ALL bit is set to zero, then the ZONE ID field specifies the lowest LBA of the write pointer zone on which the device server shall perform a reset write pointer operation. If the ALL bit is set to one, then the device server shall ignore the ZONE ID field.

If the ALL bit is set to zero and the ZONE ID field:

a) does not specify the lowest LBA of a write pointer zone (see 4.4.3), then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to ILLEGAL REQUEST and the additional sense code set to INVALID FIELD IN CDB; or
b) specifies the lowest LBA of a write pointer zone (see 4.4.3) and for the zone specified by the ZONE ID field, the Zone Condition is:
   A) READ ONLY, then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to DATA PROTECT and additional sense code set to ZONE IS READ ONLY;
   B) OFFLINE, then the device server shall terminate the command with CHECK CONDITION status, with the sense key set to DATA PROTECT and additional sense code set to ZONE IS OFFLINE;
   C) EMPTY, then the device server shall make no changes in the Zone Condition and not return an error; or
   D) IMPLICITLY OPENED, EXPLICITLY OPENED, FULL, or CLOSED, then the device server shall perform a reset write pointer operation (see 4.4.3.2.5) on the zone specified by the ZONE ID field.

An ALL bit set to one specifies that the device server shall perform a reset write pointer operation (see 4.4.3.2.5) on each zone with a Zone Condition of IMPLICITLY OPENED, EXPLICITLY OPENED, FULL, or CLOSED.

The CONTROL byte is defined in SAM-5.

6 Parameters for zoned block devices

6.1 Diagnostic parameters

The diagnostic pages and their corresponding page codes for host aware zoned block devices are defined in SPC-5 and SBC-4.

The diagnostic pages and their corresponding page codes for host managed zoned block devices are defined in table 28.

Table 28 — Diagnostic page codes for host managed zoned block devices

Diagnostic page name	Page code	Reference
Diagnostic pages assigned by SPC-5	30h to 3Fh	SPC-5
Direct access device diagnostic pages	40h to 7Fh	SBC-4
SCSI enclosure services diagnostic pages	01h to 2Fh	SES-4
Vendor specific diagnostic pages	80h to FFh	

6.2 Log parameters

The log pages and their corresponding page codes and subpage codes for host aware zoned block devices are defined in SPC-5 and SBC-4.

The log pages and their corresponding page codes and subpage codes for host managed zoned block devices are defined in table 29. See SPC-5 for a detailed description of logging operations.

Table 29 — Log page codes and subpage codes for host managed zoned block devices

Log page name	Page code [a]	Subpage code [a]	Reference
Application Client	0Fh	00h	SPC-5
Background Scan Results	15h	00h	SBC-4
Environmental Limits	0Dh	02h	SPC-5
Environmental Reporting	0Dh	01h	SPC-5
Informational Exceptions	2Fh	00h	SPC-5
Power Condition Transitions	1Ah	00h	SPC-5
Protocol-Specific Port	18h	00h to FEh	SPC-5
Read Error Counters	03h	00h	SPC-5
Self-Test Results	10h	00h	SPC-5
Start-Stop Cycle Counter	0Eh	00h	SPC-5
Supported Log Pages	00h	00h	SPC-5
Supported Log Pages and Subpages	00h	FFh	SPC-5
Supported Subpages	01h to 3Fh	FFh	SPC-5
Temperature	0Dh	00h	SPC-5
Verify Error Counters	05h	00h	SPC-5
Write Error Counters	02h	00h	SPC-5
Utilization	0Eh	01h	SBC-4
Vendor specific	30h to 3Eh	00h to FEh	n/a

[a] All page code and subpage code combinations not shown in this table are reserved for host managed zoned block devices.

6.3 Mode parameters

The mode pages and their corresponding page codes and subpage codes for host aware zoned block devices are defined in SPC-5 and SBC-4.

The mode pages and their corresponding page codes and subpage codes for host managed zoned block devices are defined in table 30. See SBC-4 for mode parameter block descriptors used by host managed zoned block devices.

Table 30 — Mode page codes and subpage codes for host managed zoned block devices

Mode page name	Page code	Subpage code	Reference
Application Tag	0Ah	02h	SBC-4
Background Control	1Ch	01h	SBC-4
Caching	08h	00h	SBC-4
Control	0Ah	00h	SPC-5
Control Extension	0Ah	01h	SPC-5
Disconnect-Reconnect	02h	00h	SPC-5
Informational Exceptions Control	1Ch	00h	SBC-4
Power Condition	1Ah	00h	SPC-5
Protocol-Specific Logical Unit	18h	00h	SPC-5
Protocol-Specific Port	19h	00h	SPC-5
Read-Write Error Recovery	01h	00h	SBC-4
Return all mode pages and subpages [a]	3Fh	FFh	SPC-5
Return all mode pages only (i.e., not including subpages) [a]	3Fh	00h	SPC-5
Return all subpages for the specified mode page code [a]	00h to 3Eh	FFh	SPC-5
Verify Error Recovery	07h	00h	SBC-4
Vendor specific [b]			
Reserved	all other combinations		

[a] Valid only for the MODE SENSE command.
[b] These vendor specific mode page code and subpage code combinations do not require a page format: mode page 00h with subpage code 00h and mode page codes 20h to 3Eh with all subpage codes.

6.4 Vital product data (VPD) parameters

6.4.1 VPD parameters overview

The VPD pages and their corresponding page codes for host aware zoned block devices are defined in table 31, SPC-5, and SBC-4.

The VPD pages and their corresponding page codes for host managed zoned block devices are defined in table 31 and in SPC-5.

Table 31 — VPD page codes for zoned block devices

VPD page name	Page code [a]	Reference	Host managed zoned block device support requirements [b]	Host aware zoned block device support requirements [b]
ATA Information	89h	SAT-4	See SAT-4	See SAT-4
Block Device Characteristics	B1h	SBC-4	mandatory	mandatory
Block Device Characteristics Extension	B5h	SBC-4	optional	optional
Block Limits	B0h	SBC-4	mandatory	mandatory
Supported Block Lengths And Protection Types	B4h	SBC-4	optional	optional
Zoned Block Device Characteristics	B6h	6.4.2	mandatory	mandatory

[a] All page codes not shown in this table or SPC-5 are reserved for host managed zoned block devices.
[b] Support requirements defined in this table take precedence over support requirements defined in SBC-4.

6.4.2 Zoned Block Device Characteristics VPD page

The Zoned Block Device Characteristics VPD page (see table 32) provides a list of parameters that are specific to zoned block devices.

Table 32 — Zoned Block Device Characteristics VPD page

Bit / Byte	7	6	5	4	3	2	1	0
0	PERIPHERAL QUALIFIER			PERIPHERAL DEVICE TYPE				
1	PAGE CODE (B6h)							
2	PAGE LENGTH (003Ch)							
3								
4	Reserved							URSWRZ
5	Reserved							
•••								
7								
8	(MSB)							
•••	OPTIMAL NUMBER OF OPEN SEQUENTIAL WRITE PREFERRED ZONES							
11								(LSB)
12	(MSB)							
•••	OPTIMAL NUMBER OF NON-SEQUENTIALLY WRITTEN SEQUENTIAL WRITE PREFERRED ZONES							
15								(LSB)
16	(MSB)							
•••	MAXIMUM NUMBER OF OPEN SEQUENTIAL WRITE REQUIRED ZONES							
19								(LSB)
20	Reserved							
•••								
63								

The PERIPHERAL QUALIFIER field and PERIPHERAL DEVICE TYPE field are defined in SPC-5.

The PAGE CODE field and PAGE LENGTH field are defined in SPC-5 and shall be set to the values shown in table 32 for the Zoned Block Device Characteristics VPD page.

An unrestricted read in sequential write required zone (URSWRZ) bit set to zero indicates that the device server does not support reading unwritten logical blocks or reads across zone boundaries in sequential write required zones. A URSWRZ bit set to one indicates that the device server supports reading unwritten logical blocks and reads across zone boundaries in sequential write required zones. The processing of reads in sequential write required zones is described in 4.4.3.4.3.

The OPTIMAL NUMBER OF OPEN SEQUENTIAL WRITE PREFERRED ZONES field indicates the recommended upper limit for the number of sequential write preferred zones that are open zones. A value of FFFF_FFFFh indicates that the optimal number of sequential write preferred zones that are open zones is not reported.

The OPTIMAL NUMBER OF NON-SEQUENTIALLY WRITTEN SEQUENTIAL WRITE PREFERRED ZONES field indicates the recommended upper limit for the number of sequential write preferred zones in which at least one non-sequential write operation has occurred (see 4.4.3.3). A value of FFFF_FFFFh indicates that the optimal number of non-sequentially written sequential write preferred zones is not reported.

The MAXIMUM NUMBER OF OPEN SEQUENTIAL WRITE REQUIRED ZONES field indicates the maximum number of sequential write required zones that are allowed to be open zones (see 4.4.3.4). A value of FFFF_FFFFh indicates that there is no limit for the number of sequential write required zones having a Zone Condition of IMPLICIT OPEN and EXPLICIT OPEN.

Annex A
(informative)

Application Client Considerations for Zoned Devices

A.1 Application client considerations for zoned devices overview

Zoned block devices add host management requirements to the SBC-4 requirements associated with the direct access block device type. One perspective on the added zoned block device requirements is found in the article *Shingled Magnetic Recording: Areal Density Increase Requires New Data Management* (see Bibliography) that discusses:

a) Cooperatively Managed SMR concepts that are related to host aware zoned block device model (see 4.2.2);
b) Host-Managed SMR concepts that are related to host managed zoned block device model (see 4.2.3); and
c) Drive-Managed SMR concepts that are outside the scope of this standard.

To facilitate better interactions between application clients and device servers, this annex considers descriptions of ZBC requirements from an application client perspective.

A.2 Writing to write pointer zones

In a write pointer zone (see 4.4.3), the preferred place to write is at a valid write pointer.

While a zone is an open zone, the write pointer may be volatile.

Although a zone with a Zone Condition of FULL does not have a valid write pointer, the device server is able to perform write operations in sequential write preferred zones (see 4.4.3.3) with a Zone Condition of FULL.

A.3 Open zone considerations

A.3.1 Open zone considerations overview

A zoned block device is able to process write commands only on LBAs:

a) in open zones; or
b) in sequential write preferred zones (see 4.4.3.3) with a Zone Condition of FULL.

Open zones have access to open zone resources that are not available to zones that are not open. Limitations on the availability of open zone resources and the number of zones that are allowed to be open are indicated by:

a) the OPTIMAL NUMBER OF OPEN SEQUENTIAL WRITE PREFERRED ZONES field (see A.4.1); and
b) the MAXIMUM NUMBER OF OPEN SEQUENTIAL WRITE REQUIRED ZONES field (see A.4.2).

Zones with a Zone Condition of EMPTY or CLOSED are opened as a result of performing:

a) a write operation;
b) a finish zone operation (see 4.4.3.2.4); or
c) an open zone operation (see 4.4.3.2.2).

A write operation results in an error if performed on any zone with a Zone Condition of READ ONLY, OFFLINE, or a sequential write required zone with a Zone Condition of FULL. If a write operation is performed on LBAs in a zone with a Zone Condition of EMPTY or CLOSED, then the device server:

1) performs a manage open zone resources operation (see 4.4.3.2.6);
2) opens the zone, if open zone resources are available; and

3) performs the write operation.

An open zone remains open until:

a) a write operation is performed for this zone that writes from the write pointer to the last LBA in the zone;
b) this zone has a Zone Condition of IMPLICITLY OPENED and a manage open zone resources operation (see 4.4.3.2.6) that is performed for a different zone selects this zone to be closed;
c) a finish zone operation (see 4.4.3.2.4) is performed for this zone;
d) a reset write pointer operation (see 4.4.3.2.5) is performed for this zone;
e) a close zone operation (see 4.4.3.2.3) is performed for this zone;
f) the zone's Zone Condition becomes READ ONLY or OFFLINE; or
g) a power on occurs.

Until a zone is closed, the following may be volatile:

a) logical block data written in that zone; and
b) the write pointer, if any.

A.3.2 Explicitly opened zones and implicitly opened zones

An open zone is a zone with a Zone Condition of IMPLICITLY OPENED or EXPLICITLY OPENED. A device server opens a zone with a Zone Condition of:

a) EXPLICITLY OPENED as a result of performing an open zone operation (see 4.4.3.2.2); or
b) IMPLICITLY OPENED as a result of performing:
   A) a write operation; or
   B) a finish zone operation (see 4.4.3.2.4).

If a device server performs an open zone operation on a zone with a Zone Condition of IMPLICITLY OPENED, the result is a zone with a Zone Condition of EXPLICITLY OPENED.

The manage open zone resources operation (see 4.4.3.2.6) is:

a) allowed to close a zone with a Zone Condition of IMPLICITLY OPENED; and
b) not allowed to close a zone with a Zone Condition of EXPLICITLY OPENED.

After a power on reset, no zones have a Zone Condition of IMPLICITLY OPENED or EXPLICITLY OPENED. For any zone that was an open zone before the power on, after the power on the Zone Condition is:

a) EMPTY, if the write pointer indicates the lowest LBA in the zone and Non-Sequential Resources Active is false; or
b) CLOSED, if the write pointer does not indicate the lowest LBA in the zone or Non-Sequential Resources Active is true.

A.3.3 Opening and closing zones

An application client uses one of the following techniques to open a zone for writing and later to close that zone:

a) an OPEN ZONE command (see 5.4) to open a zone and a CLOSE ZONE command (see 5.2) to close that zone; or
b) a write command to open a zone and:
   A) a CLOSE ZONE command to close that zone; or
   B) allowing the manage open zone resources operation (see 4.4.3.2.6) to close that zone.

An application client may use different techniques to manage the opening and closing of different zones.

An application client may use CLOSE ZONE commands to close more zones than the number necessary to satisfy limitations on the number of zones that are allowed to be open (see A.3.1).

A.3.4 Finish zone operation considerations

A zone may be opened (e.g., with an OPEN ZONE command (see 5.4)) prior to processing a FINISH ZONE command (see 5.3).

Regardless of how a zone is opened before performing a finish zone operation (see 4.4.3.2.4), completion of that Finish zone operation changes the Zone Condition to FULL, with the result that open zone resources become available to open a different zone.

A.4 Open zone resources considerations based on zone type

A.4.1 Sequential write preferred zones

An application client:

1) reads the OPTIMAL NUMBER OF OPEN SEQUENTIAL WRITE PREFERRED ZONES field (see 6.4.2) to determine the recommended number of zones open at the same time; and
2) sends:
   A) a REPORT ZONES command (see 5.6) with:
      a) the REPORTING OPTIONS field (see 5.6.1) set to 2h; and
      b) the PARTIAL bit (see 5.6.1) set to zero,
      to return the list of zones with a Zone Condition of IMPLICITLY OPENED;

   and

   B) a REPORT ZONES command with:
      a) the REPORTING OPTIONS field set to 3h; and
      b) the PARTIAL bit set to zero,
      to return the list of zones with a Zone Condition of EXPLICITLY OPENED,

   to obtain the number of open zones (i.e., zones with a Zone Condition of IMPLICITLY OPENED plus zones with a Zone Condition of EXPLICITLY OPENED).

If the number of open zones is equal to or greater than the optimal number of open sequential write preferred zones, then:

a) the preferred host behavior is to write to a zone with Zone Condition of EXPLICITLY OPENED or IMPLICITLY OPENED instead of writing to a zone with Zone Condition of EMPTY or CLOSED; or
b) before writing to a zone with Zone Condition of EMPTY or CLOSED, the application client may send CLOSE ZONE command (see 5.2) to specify the zone to be closed instead of allowing the device server to select the zone to be closed when the write occurs.

An application client:

1) reads the OPTIMAL NUMBER OF NON-SEQUENTIALLY WRITTEN SEQUENTIAL WRITE PREFERRED ZONES field (see 6.4.2) to determine the optimum number of zones that are allowed to be non-sequentially written at the same time; and
2) sends a REPORT ZONES command (see 5.6) with:
   a) the REPORTING OPTIONS field (see 5.6.1) set to 11h; and
   b) the PARTIAL bit (see 5.6.1) set to zero,
   to return the list of zones with Non-Sequential Write Resources Active set to true.

The preferred application client behavior is to manage writes to a zoned block device to ensure that the number of sequential write preferred zones with Non-Sequential Write Resources Active set to true is less than the optimal number of non-sequentially written sequential write preferred zones.

A.4.2 Sequential write required zones

Unlike Sequential Write Preferred zones (see A.4.1), the value in the MAXIMUM NUMBER OF OPEN SEQUENTIAL WRITE REQUIRED ZONES field (see 6.4.2) indicates the allowed maximum number of zones with a Zone Condition of EXPLICITLY OPENED. After all zones with a Zone Condition of IMPLICITLY OPENED have been closed by the manage open zone resources operation (see 4.4.3.2.6), the device server is required to complete any command with an error if that command requires a zone to be opened (e.g., a write command, an OPEN ZONE command (see 5.4)).

An application client:

1) reads the MAXIMUM NUMBER OF OPEN SEQUENTIAL WRITE REQUIRED ZONES field (see 6.4.2) to determine the recommended number of zones open at the same time; and
2) sends:
   A) a REPORT ZONES command (see 5.6) with:
      a) the REPORTING OPTIONS field (see 5.6.1) set to 2h; and
      b) the PARTIAL bit (see 5.6.1) set to zero,
      to return the list of zones with a Zone Condition of IMPLICITLY OPENED;

   or

   B) a REPORT ZONES command (see 5.6) with:
      a) the REPORTING OPTIONS field (see 5.6.1) set to 3h; and
      b) the PARTIAL bit (see 5.6.1) set to zero,
      to return the list of zones with a Zone Condition of EXPLICITLY OPENED,

   to obtain the number of open zones (i.e., zones with a Zone Condition of IMPLICITLY OPENED plus zones with a Zone Condition of EXPLICITLY OPENED).

If the number of open zones is equal to or greater than the maximum number of open sequential write required zones, then:

a) the preferred application client behavior is to write to a zone with Zone Condition of EXPLICITLY OPENED or IMPLICITLY OPENED instead of writing to a zone with Zone Condition of EMPTY or CLOSED; or
b) to avoid commands that are completed with an error when writing to a zone with Zone Condition of EMPTY or CLOSED, the application client sends CLOSE ZONE command (see 5.2) to specify a zone to be closed.

A.5 Partial failures

A.5.1 Partial failures overview

The Zone Conditions READ ONLY and OFFLINE provide the ability for application clients to continue using a zone block device after parts of its storage capacity have stopped operating (e.g., the use of the OFFLINE Zone Condition as a response to a media failures (see 4.5.3)).

Many of the conditions that result in the Zone Condition becoming READ ONLY and OFFLINE are outside the scope of this standard.

A.5.2 Sanitize considerations

Zoned block devices may change the Zone Condition to OFFLINE in any zone that has a Zone Condition of READ ONLY prior to performing a sanitize operation (see SBC-4). Zoned block devices may avoid failing a sanitize operation by changing the Zone Condition to OFFLINE in any zone where SBC-4 requirements for media failures (see 4.5.3) prevent successful completion of a sanitize operation.

Bibliography

ISO/IEC 14776-154, *Serial Attached SCSI - 3 (SAS-3)* (under consideration)

ISO/IEC 14776-263, *SAS Protocol Layer (SPL-3)* (under consideration)

INCITS 489-2014, *SCSI over PCIe® architecture (SOP)*

USENIX ;login:, v 38, n 3, *Shingled Magnetic Recording: Areal Density Increase Requires New Data* Management, Tim Feldman, Garth Gibson, June 2013 [1] [2]

1.See https://www.usenix.org/publications/login.

2.Reprints available from, http://www.cs.cmu.edu/~./garth/#pubs.