

To: T10 Committee
From: Gerry Houlder, Seagate Technology <gerry.houlder@seagate.com>
Subj: Background Medium Scan
Date: Nov. 4, 2004

This document proposes a standard method to control and retrieve status from background medium scan operations. Several drive vendors (including Seagate) have proprietary methods for controlling these operations today, but customers have asked that a standard method be defined so their processes can be simpler.

Many systems implement background medium scanning in their operating systems. Having the drive do the medium scanning instead has the advantage of reducing system overhead and removing an unproductive demand on interface bus bandwidth. During idle periods, the drive reads each sector on its medium to verify the data. If correctable errors are detected, the drive may rewrite and/or relocate the LBA to another sector depending on what has been allowed by the customer. Errors are logged to a list; if allowed by the customer, a future write to that LBA may result in the LBA being relocated before performing the write.

There is a related option called Pre-scan. This is a feature that starts a background medium scan right after power on. If the customer does a write operation to an area of the disk that has not been scanned yet, the drive turns the write operation into a write-and-verify operation. This has the advantage of checking (right after power on) that all data can be read back correctly. A disadvantage is that the drive's performance is reduced (because of the extra steps to do the verify operation) until the first medium scan is completed. Some customers choose to disable this option for that reason.

This proposal includes fields in a mode page that is used to control the scan and pre-scan operations. It also includes parameters in a log page that stores a list of LBAs that have or did have read errors. The system can read this list and may take action on LBAs that the drive was not able to handle on its own.

Editorial conventions – added text is underlined, deleted text is ~~struck through~~, changes to tables are not specifically marked.

Rev. 1: Added wording for model section; added wording change for EBF bit in IEC mode page; added description for using Extended Control mode page (page 0xA, subpage 1) for the BMS mode page controls; added description to use Self-Test Results log page for returning BMS log data.

1.0 Add background medium scan model in SPC

5.x Background medium read scan operations

Medium scanning is when logical blocks are read from the medium for the purpose of verifying that the blocks are readable. The blocks are not retained in cache after they are read. The intent is to identify blocks that are difficult to read or unreadable, log the problem, and possibly take action to make the block easily readable again. This action increases the reliability of storing data in the device. A system can do this by sending read commands over the SCSI interface to access all logical blocks periodically, but this has the disadvantage of using system resources (e.g., SCSI interface bandwidth and initiator processing power) for an unproductive purpose. This disadvantage is removed by having the device do the scanning itself, during time when it is otherwise idle, without using any bandwidth on the SCSI interface.

If a block is readable but requires extra actions (e.g., retries or application of correction algorithm) to be read, the device may resolve the problem by rewriting the original data to the same physical location or relocating the logical block to another physical location.

If a block is unreadable the device cannot resolve the problem itself, but it may add that logical block to an internal list so it gets relocated the next time that logical block is written by an initiator.

The initiator still has control over whether to allow the device to correct the problem or to address the problem itself. The ARRE bit in the Read Write Error Recovery mode page (see SBC) controls whether the device may automatically repair/ relocate recoverable read errors. The AWRE bit (also in the Read Write Error Recovery mode page) controls whether the device may relocate blocks during write operations. Blocks that have previously been logged as unrecoverable are reassigned at the start of the next write operation to that logical block. These bits control the device action during background medium scanning as well as during normal read and write operations.

5.x.1 Enabling medium pre-scan operation

The pre-scan feature is enabled by clearing, then setting the EN_PS bit in the Control Extension mode page (see xxx) and power cycling the device. After power is restored the device will begin scanning the medium starting at LBA zero and ending with the last LBA. During this time any commands from an initiator will cause the pre-scan operation to be suspended while the device services the commands. The pre-scan operation will resume where it left off when all initiator commands have been completed.

Write commands may get special handling by the device. If a write command accesses an LBA that have not yet been scanned during the pre-scan operation, then it is converted into a write and verify command. This verifies that the data just written was read back successfully. If a write command accesses an LBA that has already been scanned during the pre-scan operation then it is processed normally. All other commands are processed normally in either case.

5.x.2 Halting medium pre-scan operation

The pre-scan feature is halted when any of the following occurs:

- (a) the pre-scan operation completes scanning all LBAs on the device;
- (b) an initiator clears the EN_PS bit (see xxx [refer to Control Extension mode page]);
- (c) the Prescan Timeout Value expires (see xxx [refer to Control Extension mode page]);
- (d) or the device detects an unusual pattern of errors.

Once the pre-scan is halted, it can only be restarted from the beginning (see 5.x.1).

5.x.3 Enabling background medium scan operation

If pre-scan operation is enabled, it shall be completed or halted before background medium scan operation is started. The background medium scan feature is enabled by setting the EN_BMS bit in the Control Extension mode page (see xxx). After the BMS Interval Time (see xxx [refer to Control Extension mode page]) has been satisfied, the device will begin scanning the medium starting at LBA zero and ending with the last LBA. During this time any commands from an initiator will cause the background medium scan operation to be suspended while the device services the commands. The background medium scan operation will resume where it left off when all initiator commands have been completed.

When the last LBA has been scanned, the background medium scan operation becomes dormant until the BMS Interval Time is satisfied again. Then the background medium scan operation starts another scanning cycle starting at LBA zero.

5.x.4 Suspending background medium scan operation

The background medium scan feature is suspended when any of the following occurs:

- (a) the background medium scan operation completes scanning all LBAs on the device;
- (b) an initiator clears the EN_BMS bit (see xxx [refer to Control Extension mode page]);
- (c) or the device detects an unusual pattern of errors.

Once the background medium scan is suspended, it resumes from the spot where it left off when it is enabled again (see 5.x.3).

5.x.5 Interpreting the logged results

An initiator may poll the self-test results log page (see 7.2.10) to get information about pre-scan and background medium scan activity. The BMS status parameter has fields that indicate whether a pre-scan or background medium scan is active or suspended, the number of scans performed on the medium, and the next LBA to be scanned if a scan is active. This information can be used by an initiator to monitor the background scanning operations.

The medium scan parameters, if any, describe the physical and logical location of any suspected defects. The reassign status field indicates whether the defect was completely handled by the device or whether the initiator may have to take action (e.g., reassigning or re-writing an LBA) to fix a particular defect.

After an initiator analyzes the medium scan parameters and has completed actions, if any, to resolve the defects it may delete the list by issuing a LOG SELECT command with PCR bit set (see 6.5 [refer to Log Select command]). This eliminates the need to analyze the same entries at a later time.

2.0 Changes to Information Exception Control mode page in SPC

7.4.11 Information Exceptions Control mode page

[editors note: words describing everything but EBF bit are unchanged but not replicated here.]

If background functions are supported, an Enable Background Function (EBF) bit set to one indicates the device server shall enable background functions. An EBF bit set to zero indicates the device server shall disable the functions. Background functions with separate enable control bits are not controlled by this bit (e.g., EN_BMS bit in table 245).

For the purposes of the EBF bit, background functions are defined as idle time functions that may impact performance that are performed by a device server operating without errors but do not impact the reliability of the logical unit (e.g., read scan).

3.0 Changes to Control Extension mode page in SPC

7.4.7 Control Extension mode page

The Control Extension mode page (see table 245) is a subpage of the Control mode page (see 7.4.6) and provides controls over SCSI features that are applicable to all device types. The mode page policy (see 6.7) for this subpage shall be shared. If a field in this mode subpage is changed while there is a task already in the task set, it is vendor specific whether the old or new value of the field applies to that task.

Table 245 – Control Extension mode page

Bit	7	6	5	4	3	2	1	0
0	PS	SPF(1b)	PAGE CODE (0Ah)					
1	SUBPAGE CODE (01h)							
2	(MSB)	PAGE LENGTH (1Ch)						
3								(LSB)
4	RESERVED							IALUAE
5	RESERVED				INITIAL PRIORITY			
6	RESERVED							EN_PS
7	RESERVED							EN_BMS
8	(MSB)	BMS INTERVAL TIME						
9								(LSB)
10	(MSB)	PRESCAN TIMEOUT VALUE						
11								(LSB)
12	RESERVED							
13	RESERVED							
14	RESERVED							
31								

[editors note: words describing fields in bytes 4 and 5 are unchanged but not replicated here.]

An enable pre-scan (EN PS) bit set to zero specifies that pre-scan is disabled. If a pre-scan operation is in progress when EN PS is changed from a one to a zero then pre-scan is halted. An EN PS bit set to one specifies that a pre-scan operation is started after the next power on cycle. Once this pre-scan has completed, another pre-scan shall not occur unless the EN PS bit is set to zero, then set to one, and another power on cycle occurs.

An enable background medium scan (EN BMS) bit set to zero specifies that background medium scan is disabled. An EN BMS bit set to one specifies that background medium scan operations are allowed. If the EN PS bit is also set to one then a background medium scan operation will not start until after the pre-scan operation is halted or completed. If a background medium scan is in progress when the EN BMS bit is changed from one to zero, then the medium scan is suspended; when the EN BMS bit is subsequently set to one, the medium scan shall resume from the suspended location.

The BMS interval time field specifies the minimum time, in hours, between the end of one pre-scan or background medium scan operation and the start of the next background medium scan operation.

The pre-scan timeout field specifies the maximum time, in hours, for a pre-scan operation to complete. If the pre-scan operation doesn't complete within the specified time then it is halted.

4.0 Changes to Self-Test Results log page in SPC

7.2.10 Self-Test Results log page

The Self-Test Results log page (see table 210) provides the results from the twenty most recent self-tests (see 5.5). Results from the most recent self-test or the self-test currently in progress shall be reported in the first self-test log parameter; results from the second most recent self-test shall be reported in the second self-test log parameter; etc. If fewer than twenty self-tests have occurred, the unused self-test log parameter entries shall be zero filled.

The Self-Test Results log page also returns the BMS Status parameter and zero or more Medium Scan parameters when the background medium scan feature is implemented. The BMS Status parameter provides information about pre-scan and background medium scan operations. Each medium error log entry corresponds to a site where an error was detected. If the Medium Scan log area is filled up, a new medium scan log entry will overwrite the oldest entry. When a LOG SELECT command with PCR bit set to one is processed all Medium Scan parameters are deleted, but the values in the BMS Status Parameter are not affected.

Table 210 – Self-Test Results log page

Byte	Bit	7	6	5	4	3	2	1	0								
0		PAGE CODE (10h)															
1		RESERVED															
2	(MSB)	PAGE LENGTH (N-3)															
3		(LSB)															
		Self-Test Results parameters															
4		FIRST SELF-TEST RESULTS PARAMETER															
23																	
										:							
										:							
384										TWENTIETH SELF-TEST RESULTS PARAMETER							
403																	
										Background Medium Scan parameters							
404										BMS STATUS PARAMETER							
423																	
424										FIRST MEDIUM SCAN PARAMETER							
447																	
										:							
										:							
423 + 24N										LAST MEDIUM SCAN PARAMETER							

[editors note: for N=2048 medium scan parameters, total number of bytes will be 49,576. This is comfortably within the maximum page length of 65,535.]

[editors note: descriptions of self-test results parameters are unchanged but not replicated here.]

7.2.10.1 BMS Status parameter

The background medium scan (BMS) status parameter (see table new1) contains status information about the background medium scan feature.

Table new1 – BMS Status parameter format

Bit	7	6	5	4	3	2	1	0
Byte 0	(MSB) PARAMETER CODE (1000h)							
1	(LSB)							
2	DU	DS	TSD	ETC	TMC		LBIN	LP
3	PARAMETER LENGTH (10h)							
4	(MSB)							
5	TIME STAMP							
6								
7	(LSB)							
8	RESERVED							
9	BMS STATUS							
10	(MSB) NUMBER OF SCANS PERFORMED							
11	(LSB)							
12	(MSB)							
13	CURRENT LBA FOR SCAN							
14								
15								
16								
17								
18								
19	(LSB)							

The TIME STAMP indicates the drive power on minutes at the time the log page is requested.

Table new2 specifies the meaning of the BMS STATUS field.

Table new2 – BMS Status field

Value	Meaning of definition parameter
0h	No scans active.
1h	Background medium scan is active.
2h	Pre-scan is active.
3h	Scan suspended due to fatal error.
4h	Scan suspended due to unusual pattern of errors.
5h	Scan suspended due to medium formatted without P-List.
6h	Scan suspended – vendor specific cause.
7h	Scan suspended due to temperature out of allowed range.
8h	Scan suspended until BMS Interval Time (see 7.4.7 [refer to Control Extension mode page]) expires.
9h - FFh	Reserved

The NUMBER OF SCANS PERFORMED field indicates the number of background scans that have been performed since the drive was originally shipped by the drive vendor.

The CURRENT LBA FOR SCAN field indicates the next LBA that will be scanned when a pre-scan or background medium scan is active. When no scans are active, the value is set to zero.

7.2.10.2 Medium Scan parameter

A Medium Scan parameter (see table new3) describes a suspected defect location on the medium.

Table new3 – Medium Scan parameter format

Bit	7	6	5	4	3	2	1	0
0	(MSB) PARAMETER CODE (1001h through 1800h)							
1	(LSB)							
2	DU	DS	TSD	ETC	TMC		LBIN	LP
3	PARAMETER LENGTH (14h)							
4	(MSB)							
5	TIME STAMP							
6								
7	(LSB)							
8	REASSIGN STATUS				SENSE KEY			
9	ADDITIONAL SENSE CODE							
10	ADDITIONAL SENSE CODE QUALIFIER							
11	HEAD NUMBER				(MSB) CYLINDER NUMBER			
14	CYLINDER NUMBER							
15	(LSB)							
14	(MSB) SECTOR NUMBER							
15	(LSB)							
16	(MSB)							
17								
18	CURRENT LBA FOR SCAN							
19								
20								
21								
22								
23	(LSB)							

The TIME STAMP indicates the drive power on minutes at the time the error was detected.

Table new4 specifies the meaning of the REASSIGN STATUS field.

Table new4 – Reassign Status field

Value	Meaning of definition parameter
0h	No reassignment needed.
1h	Reassignment pending receipt of Reassign command or write command (if auto write reallocation is allowed) from the initiator.
2h	LBA successfully reassigned by drive.
3h	Reserved
4h	Reassignment failed.
5h	LBA recovered via re-write only.
6h - Fh	Reserved

The Sense Key, Additional Sense Code, and Additional Sense Code Qualifier fields are the same as the sense bytes (see 4.5 [refer to sense data section of SPC]). They indicate the sense bytes that would be reported for this site if the medium error were detected by a normal read or write. These sense bytes are not reported by any other method because the error was detected during pre-scan or background scan.

The head number field indicates the head on which the medium error was detected.

The cylinder number field indicates the cylinder number associated with the medium error.

The sector number field indicates the sector number associated with the medium error.

The current LBA for scan field indicates the next LBA to be accessed during pre-scan or background medium scan activity. If pre-scan and background scan are not active, this value is set to zero.