XOR Command Study Group meeting minutes - Document number X3T10/95-161r0
Meeting place: Hyatt Newporter, Newport Beach
Date: March 6, 1995

Attendees:

Gerry Houlder	Seagate
Jay Elrod	Seagate
Stephen Fuld	StorageTek
Mike Chenery	Fujitsu
George Penokie	IBM
Lansing Sloan	Lawrence Livermore Nat. Lab.
Paul Boulay	Hitachi
Bill Hutchison	Hewlett Packard
Mark Hamel	EMC
Edward Fong	Amdahl
Doug Prins	QLogic
Rod DeKoning	Symbios Logic
Jim Whitworth	Conner Peripherals
Bob Snively	Sun Microsystems
Edward Gardner	Quantum
Eric Griffith	Western Digital
Robert Liu	Fujitsu
Paul Hodges	IBM
Charles Monia	Digital Equipment Corp.
Jeff Stai	Western Digital Corp.

Gerry Houlder acted as chairman for the meeting, conducting a page by page walkthrough of the XOR document (X3T10/94-111r6). Below is a summary of the major issues discussed. Note: several minor editorial changes were discussed which are not included in this list. The next revision should be referred to if details of these changes are needed.

1) It was recommended that the Model section of the document take more of a "just the facts" approach, rather than including phrases such as "An advantage of this technique is...". It was emphasized that the intent of a standard is to explain an idea, not to sell it.

2) It was recommended that the word "Host" be replaced with "Controller" in the title of section 1.1, "Host Supervised XOR Operations".

3) Terminology compatibility between the XOR document and the SCC document was discussed. For example, the XOR document uses terms such as "Data" and "Parity", whereas the SCC document uses "Protected space" and "Check data", respectively, to mean the same thing. It was suggested that perhaps the terminology in the XOR document should more closely resemble that of the SCC document to avoid confusion.

4) It was pointed out that the term "domain" could most likely be used in the context of SCSI devices capable of peer-to-peer communication. For example, the phrase "devices on the same bus or loop" could most

likely be replaced with "devices in the same domain".

5) It was suggested that a step be added in section 1.2.1 which describes the fact that an XOR operation takes place in the check data device.

6) It was recommended that paragraph c of section 1.4.2 be modified to be more correct with regard to the term "inconsistent", since the first sentence of the paragraph could be interpretted as a partial redefinition of the term (with respect to the preceding paragraphs). One suggestion was to modify the last sentence of this section to read something like "Consistency should be maintained during regenerates and rebuilds.".

7) It was recommended that example diagrams describing different possible RAID system configurations, similar to those which existed in the last revision (but were removed), be put back into the document. It was mentioned that the model section should contain any generic drawings, and that drawings which describe specific examples should be contained in the appendix.

8) It was recommended that the XDWRITE(16) command should specify that the LUN of the secondary target shall be zero. (Text to this effect was removed from an earlier revision.)

9) It was recommended that a description of the Transfer Length field be added to the XDWRITE(16) command. (This was removed from an earlier revision.)

10) There was discussion about removing the requirement for the implied exclusive access extent reservation during the XDWRITE(16) command.

11) It was mentioned that the final version of the XOR document will need to be in Frame format (the current format is Microsoft Word).

12) It was decided that the document should contain information specifying what events will relieve a target of having to retain xor'd data awaiting an XDREAD command. It was decided that, in addition to the appropriate XDREAD command, a Bus Device Reset, Reset, power cycle, Clear Queue, Abort, and Abort Tag would accomplish this.

13) Since there is nothing is the XOR document which disallows simultaneous writes to medium on both a data and parity device in Third Party mode (an XDWRITE command may be writing to the data disk at the same time as the associated XPWRITE command is writing to the parity disk), there is the possibility that two devices in the same stripe could become corrupted in the case of, for example, a power outage during disk writes. This would cause the stripe to be non-rebuildable. The idea of a special bit somewhere to prevent such simultaneous writes was mentioned. It was also mentioned that perhaps this issue should not be addressed at the spec level, but rather between vendors and customers if there is a concern.

14) It was mentioned that the recently added "buffer full" status may have the same meaning, in certain cases, as "queue full".

15) Order of appearance of commands in the XOR document was discussed.

It was pointed out that the order should be alphabetical, and should be patterned after other ANSI documents when two or more commands exist with the same name but different cdb lengths. For example, XDWRITE(10) should appear before XDWRITE(16), etc.

16) Paul Hodges (IBM) presented an idea in which the XOR task during a rebuild or regenerate operation would be distributed among all of the involved source devices rather than handled completely by the REBUILD/REGENERATE target, in order to offload some of the work from that target. No specific technique was presented since Paul was only looking for feedback as to whether the idea should be further investigated. Since the response was positive, the idea will be pursued by Paul and handled on the SCSI reflector.

17) It was mentioned that there should be an error reporting scheme for the REBUILD and REGENERATE commands since, without such a scheme, the controller has no information regarding which source device may have failed during such an operation. A sense index byte was recommended for this purpose. The byte would point to the source descriptor field (of the REBUILD/REGENERATE parameter data) which contained the address of the failing device. It was also suggested that certain error codes from the SCC document could be used, such as "Rebuild Failure".

18) REBUILD and REGENERATE parameter data:

- It was requested that provision be made for a "pad" between the last source descriptor and the intermediate data in the REGENERATE and REBUILD parameter data. The reason for such a pad would be to facilitate those controllers which are unable to send contigious parameter data (i.e. the first byte of intermediate data immediately follows the last byte of the last source descriptor). A "Source Descriptor Length" value was requested for bytes 2 and 3 of the parameter data, which would specify in bytes the sum of the source descriptor lengths and the pad length.

- There was a request for an 8 byte LUN field and 4 reserved bytes in the Source Physical Address field within the source descriptor.

- It was pointed out that the byte numbering scheme for the parameter data needs to be such that all byte numbers are relative to the first byte of parameter data, rather than to the first byte of a particular section within the parameter data. For example, the first byte of the first source descriptor should have 4, not 0, as its byte number.

19) It was recommended that the Log Mode bit be removed from the XOR Control mode page until the meaning of "Logging Device" is specified. It was also recommended that 2 reserved bytes be added to this page, increasing the total page length to 24 bytes, for byte alignment purposes.