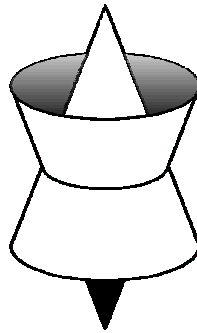


White Paper

**Western Digital Comments on
Sector Sizes Larger than 512 Bytes**

April 6, 2005
T13/e05122r0 Revision 0



Technical Editor:

Curtis E. Stevens
Western Digital
Phone: 949-672-7933
E-Mail: Curtis.Stevens@WDC.com

THIS WHITEPAPER IS MADE AVAILABLE WITHOUT CHARGE FOR USE IN DEVELOPING COMPUTER SYSTEMS AND DISK DRIVES. WESTERN DIGITAL MAKES NO REPRESENTATION OR WARRANTY REGARDING THIS WHITEPAPER OR ANY ITEM DEVELOPED BASED ON THIS WHITEPAPER, AND WESTERN DIGITAL DISCLAIMS ALL EXPRESS AND IMPLIED WARRANTIES, INCLUDING BUT NOT LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND FREEDOM FROM INFRINGEMENT. WITHOUT LIMITING THE GENERALITY OF THE FOREGOING, WESTERN DIGITAL MAKES NO WARRANTY OF ANY KIND THAT ANY ITEM DEVELOPED BASED ON THIS WHITEPAPER WILL NOT INFRINGE ANY COPYRIGHT, PATENT, TRADESECRET OR OTHER INTELLECTUAL PROPERTY RIGHT OF ANY PERSON OR ENTITY IN ANY COUNTRY. USE OF THIS WHITEPAPER FOR ANY PURPOSE IS AT THE RISK OF THE PERSON OR ENTITY USING IT. WESTERN DIGITAL RESERVES THE RIGHT TO REVISE THIS WHITEPAPER AND TO MAKE CHANGES FROM TIME TO TIME IN THE CONTENT WITHOUT OBLIGATION OF WESTERN DIGITAL TO NOTIFY ANY PERSON OF SUCH REVISIONS OR CHANGES.

Contributors	
Name	E-Mail

Revision History		
Date	Revision	Description
2-Mar-05	0	Initial Draft
		•
		•
		•
		•
		•

Table of Contents

1	INTRODUCTION	1
2	SCOPE	1
2.1	Revisioning.....	1
3	OVERVIEW	2
4	WD IMPLEMENTATION	4
4.1	1KB Sector Size Implementation	4
4.2	4KB Sector Size Implementation	4
4.3	Read-Modify-Write (RMW).....	5
5	IMPLEMENTATION ISSUES	6
5.1	Drive Partitioning.....	7
5.2	File System Formatting.....	7
5.3	Virtual Memory accessing	7
5.4	Bootting	7

LIST OF FIGURES

Figure 1 – System Food Chain..... 2
Figure 2 – Mapping Proposals 2
Figure 3 – Logical to Physical mapping 3
Figure 4 – Uncorrectable Error Handling 5
Figure 5 – Typical HDD Layout..... 6

1 Introduction

The disk drive industry has been standardized on a 512 byte sector size for over 25 years. In the continual pursuit for size and performance, larger sector sizes are being considered. Western Digital recommends that the 512 byte sector size be maintained in host communications while allowing the physical sector size of the media to change.

2 Scope

This white paper describes Western Digital's intentions for implementing a media format that incorporates sector sizes greater than 512 bytes. The target sector sizes are 1k followed by 4k. This paper does not make a case for using larger sector sizes. Instead, this paper assumes that the move to larger sector sizes will happen and addresses both system and industry implications.

The information provided in this paper enables sector sizes that are a binary multiple greater than 512 bytes. ATA/ATAPI-7 also specifies methods to report sector sizes that are not a binary multiple. Common sector sizes that are not binary multiples include 520, 524, 528 and 532 byte sectors. Non-binary multiples are beyond the scope of this paper.

2.1 Revisioning

The revision numbers used in this document indicate the level of maturity and review level of the specification. The revisions are .6, .7, .8, .9, or 1.0, followed by a letter. The revision numbers can be interpreted as follows:

1. Rev 0.6 – Proposal, not reviewed by anybody except the author. A document advances from the .6 level to the .7 level when 1 or more people have read the document and provided input for a new revision.
2. Rev 0.7 – The specification in review by a closed group of people, usually known as a working group. The specification advances from .7 to .8 when the working group believes that the document is complete and is ready to be reviewed by a greater audience. Greater audiences include Customers, Suppliers, and other groups within WD.
3. Rev 0.8 – The specification is in review corporately and possibly by customers under NDA. At this level, the specification is content complete and some or all of the sections are usually in freeze. Sections that are not part of the critical path for implementation may still be under development. The specification advances to .9 or 1.0 (whichever is appropriate) when there are no changes required by the people reviewing the document.
4. Rev 0.9 – The specification is in review publicly. At this level, the document is content complete, and attempts are being made to freeze the document. Internal specifications do not usually take on a .9 revision. There is usually a set timeframe for public review, 30, 60, and 90 days are common. After the review period is over, comments are addressed and the document can advance to 1.0, or stay at .9 for another review cycle if there were substantial changes required.

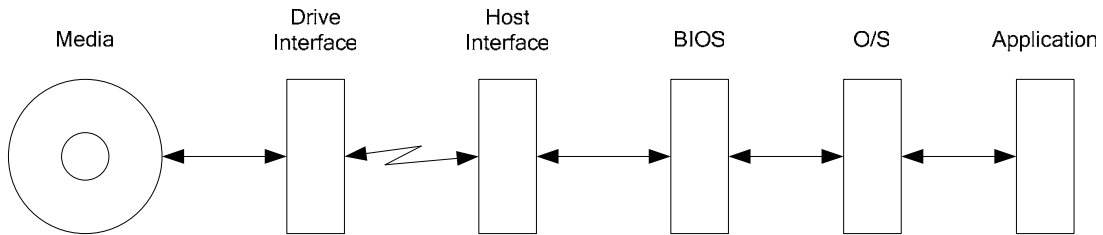
5. Rev 1.0 – Release. A revision goes to 1.0 usually when implementation is complete. At this point it is common to remove the revision history and to publish the specification.

The revision number has a letter as a suffix. This letter advances as an indication of how many drafts have been reviewed by the target audience. If more than 25 drafts are created at any revision, the suffix goes from a-z and then aa, ab, etc.

3 Overview

Western digital is considering implementing a drive with a media sector size larger than 512 bytes. The purpose of this change is to allow for greater format efficiency. Figure 1 shows major system components that are affected by a change in sector size.

Figure 1 – System Food Chain



There are two competing possibilities for expanding the sector size on the media. One proposal would expand the sector size seen at the drive interface; the other would keep the 512 byte sector size at the drive interface. Both possibilities have drawbacks. Figure 2 illustrates the possibilities. It is Western Digital’s belief that the Compatible mechanism should be implemented prior to the Future mechanism listed.

Figure 2 – Mapping Proposals

	Today	Compatible	Future	Compatible	Future
Interface Sector Size	512 Bytes	512 Bytes Requires RMW, is compatible with system food chain	1K Bytes Incompatible with food chain, does not require RMW	512 Bytes Requires RMW, is compatible with system food chain	4K Bytes Incompatible with food chain, does not require RMW
Media Sector Size	512 Bytes	1K Bytes	1K Bytes	4K Bytes	4K Bytes

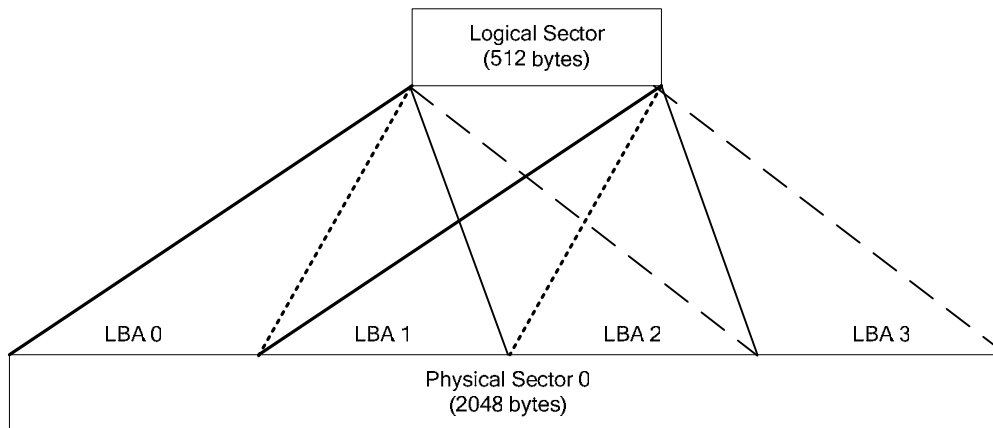
Using the Compatible mechanism, the Drive Interface, Host Interface, BIOS, OS, and Applications will still function. If the OS were modified to properly align the disk accesses then performance of the disk drive would be optimal. The Compatible mechanism also allows a drive manufacturer to ship a utility with the unit that can optimize performance. If the Future mechanism is employed, the Drive Interface, Host Interface, BIOS, OS, and Applications may stop functioning. The reason they may stop functioning is that many components in the System Food Chain are hardwired to 512

bytes. These hardwired elements include hardware, firmware and software. If you attach a drive with 1K interface sectors to a system today it will not be able to boot using Windows 2000/XP. If the host interface is able to transfer the data, it is highly likely that the system BIOS is hardwired to 512 bytes. If the BIOS were able to launch Windows, the user would find that Windows was hardwired to 512 bytes and the system would hang. In the case where the BIOS or host interface are hardwired to 512 bytes, no utility can reasonably be used to fix the problem.

The change in sector size can be compared to the transition from Cylinder-Head-Sector (CHS) addressing to Logical Block Addressing (LBA). The transition began in 1993. The issues were fairly well understood by 1994 or 1995. Systems started working well together in the 2000 timeframe. Today, even though CHS can not be used to access media above 137GB, we are still forced to support it. If WD is supported by the OS manufacturers in implementing the “Compatible” method where the disk drive implements Read-Modify-Write (RMW) for legacy compatibility and the OS implements alignment to mitigate the RMW performance impact, the transition to larger sectors can be less painful to the customer.

The ATA/ATAPI-7 standard provides a mechanism for describing media format and host LBA alignment requirements in the IDENTIFY DEVICE command and as a part of the Long Logical and Long Physical feature sets. Figure 3 illustrates an example of the capability documented by ATA/ATAPI-7.

Figure 3 – Logical to Physical mapping



In this example, the interface (logical) sector size is 512 bytes, and the media (physical) sector size is 2048 bytes. This mechanism allows an ATA device to both implement a larger physical sector and maintain compatibility with existing systems, interfaces, and software. One of the drawbacks of this method is that drive performance can suffer if the host writes data starting or ending on an LBA that is misaligned with respect to the physical sector boundaries. When misalignment occurs, the drive will be forced to perform a Read-Modify-Write (RMW) operation in order to satisfy the host request.

ATA also allows the Logical Sector size to be changed. This would allow a device to implement a 4KB sector on the media and require that the host transfer 4KB of data for each LBA requested. This type of implementation avoids the RMW issue noted above.

The main drawback of this implementation is that existing systems, interfaces, BIOS and system software (OS and otherwise) would have to change in order to accommodate the device.

Western Digital believes that any change in sector size should be compatible with existing systems.

4 Implementation

4.1 1KB Sector Size Implementation

The 1KB sector size allows for greater format efficiency, and a slight increase in performance. The change to 1KB sectors will cause some issues regarding access alignment. These issues will not be seen in an environment that has been optimized for 4KB accessing.

The device indicates the 1KB sector size to the host by returning 2001h in word 106 of IDENTIFY DEVICE. This indicates that the device has 2 512 byte logical sectors to compose a 1KB physical sector. The host can use this information to know that transfers should start on even LBAs and end on odd LBAs for best performance.

4.2 4KB Sector Size Implementation

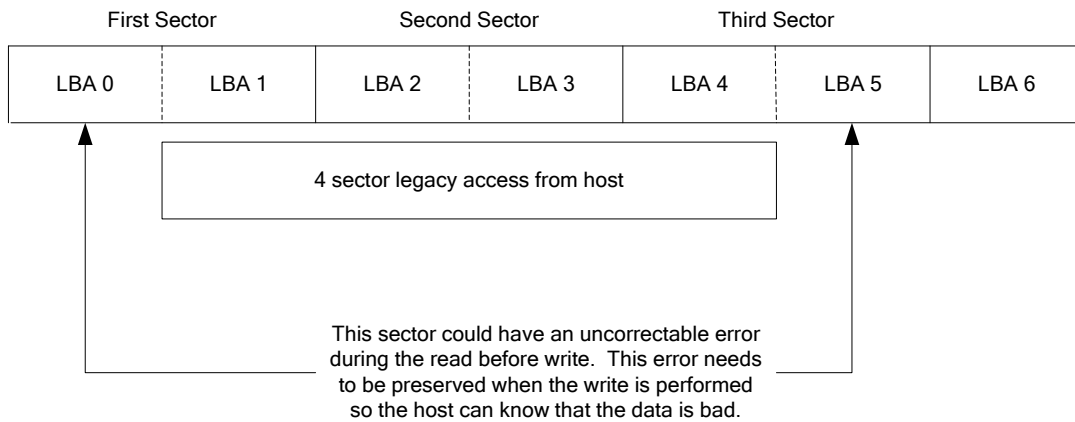
The 4KB sector size allows for greater format efficiency than the 1KB sector size; as well as a slight increase in performance. The change to 4KB sectors will cause additional issues regarding access alignment.

The device indicates the 4KB sector size to the host by returning 2003h in word 106 of IDENTIFY DEVICE. This indicates that the device has 8 512 byte logical sectors to compose a 4KB physical sector. The host can use this information to know that transfers should start with an LBA where the low order 3 bits are zero and the transfer ends on an LBA where the low order 3 bits are 1.

4.3 Read-Modify-Write (RMW)

If the recommendation for keeping the logical sector size at 512 bytes is implemented, the drive will be forced to perform RMW when it receives an unaligned transfer. The ATA/ATAPI-7 WRITE commands do not provide a way to return an error other than an ABORT or a DEVICE FAULT. If there is an uncorrectable error encountered during the initial read operation, the WRITE command has no way to report the issue. Further, this error may affect sectors not accessed by the WRITE command. Western Digital believes this issue is not a problem for ATA drives, nor does the issue require a modification to the ATA standard. There are several possible solutions for drive vendors to choose from in providing the information to the host. Figure 4 illustrates the issue.

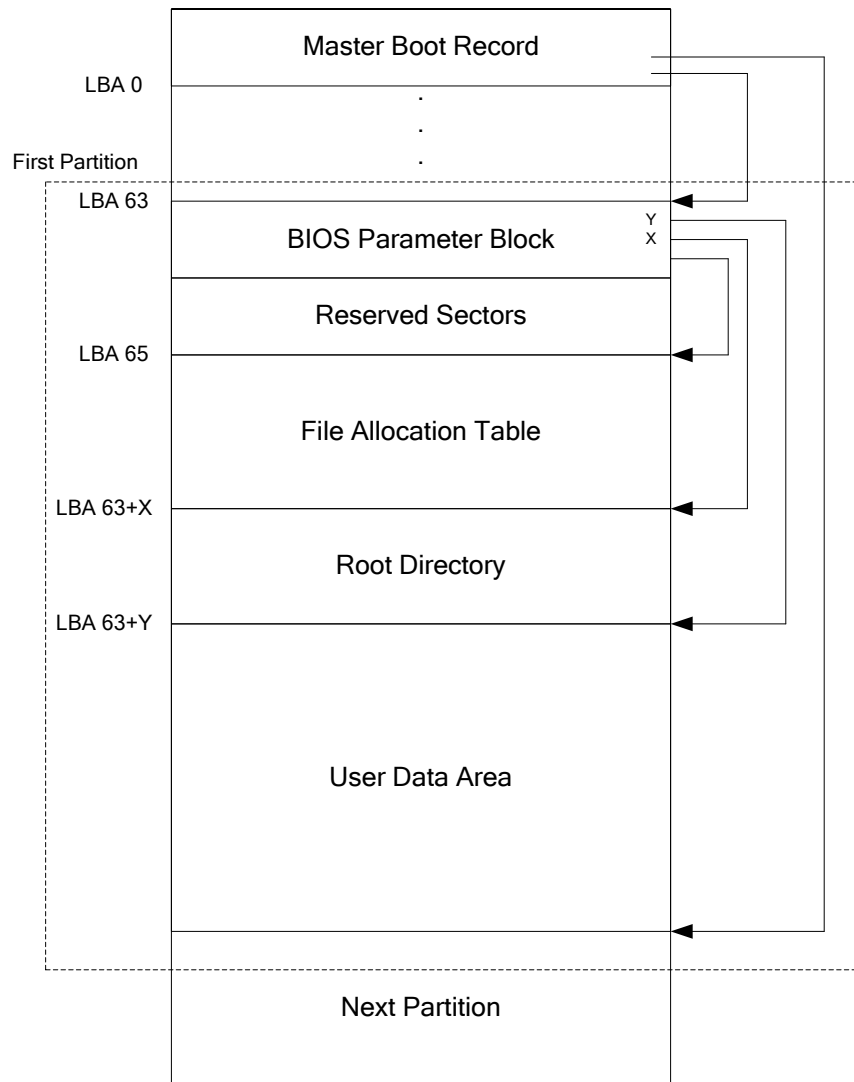
Figure 4 – Uncorrectable Error Handling



5 Implementation Issues

Although the implementation described here allows a drive to function in a legacy system without modification, there are some issues that are critical in allowing the drive to perform at peak efficiency. Figure 5 describes a typical device media layout showing the positions of the Master Boot Record (MBR), BIOS Parameter Block (BPB), and the remainder of a FAT based file system. This layout varies based on the type of FAT file system used, but all the elements described here are generally present. The sector numbers on the left hand side of the figures show typical and/or legacy locations for the various data structures on the media. The following sections describe alignment issues associated with current media layout.

Figure 5 – Typical HDD Layout



5.1 Drive Partitioning

In 1993 when the HDD industry was still dealing in cylinders heads and sectors, an important milestone was reached which caused drive manufacturers to standardize on 63 sectors per track. The norm for disk partitioning software was to place the Master Boot Record (MBR) at Cylinder 0, Head 0, sector 1 (or LBA 0). The MBR contains a pointer to the first partition. The common practice was to place the first partition at Cylinder 0 Head 1, sector 1. This meant that the LBA value of the first sector in the first partition could vary. Once the sectors per track standardized on 63, the LBA value of the first sector in the first partition standardized on LBA 63. Today, there are some applications that check to make sure that partitions start on a track boundary, even though there is no meaning for cylinders heads and sectors.

As we move forward and create larger sectors, partition alignment becomes an important issue. In the case of a 1KB sector device, the partitions should start on an even numbered sector and end on an odd numbered sector. If the drive implements a 4KB sector on the media, then the partition should start on an LBA where the low order 3 bits are zero.

Western Digital recommends that all partitions start on a LBA that is aligned with the start of a physical sector on the media.. We know that this effects some applications that check to make sure the first partition starts on sector 63, but a change is required to implement larger sectors on the media.

5.2 File System Formatting

There are many file systems that cluster sectors together to create an allocation unit larger than a single 512 byte sector. These file systems generally implement a table to associate clusters with files, commonly called a File Allocation Table (FAT). A typical cluster size is 4KB or 8 512 byte sectors. Even if the Partition is properly aligned, there is an issue where the size of the FAT can cause the individual clusters in the user data area to be unaligned relative to the physical sectors on the media. This would also result in performance degradation.

If the clusters in the file system are properly aligned, file accesses will naturally be aligned in many cases and performance will not be degraded.

5.3 Virtual Memory accessing

Once the clusters in the file system are aligned, the OS memory manager needs to be modified to prevent unaligned accesses. When a drive has alignment requirements, disk performance tests may show acceptable performance, but if the virtual memory activity is not aligned, CPU performance tests may provide unacceptable results.

5.4 Booting

The drives with alignment requirements should not show significant performance degradation on unaligned reads. Since booting is mainly a reading process, we do not expect to see an impact on system boot times in an unaligned environment.