Larger Sector Sizes: A Drive Vendor's Perspective

By

Curtis E. Stevens

Seagate
We turn on ideas

WD Western Digital®

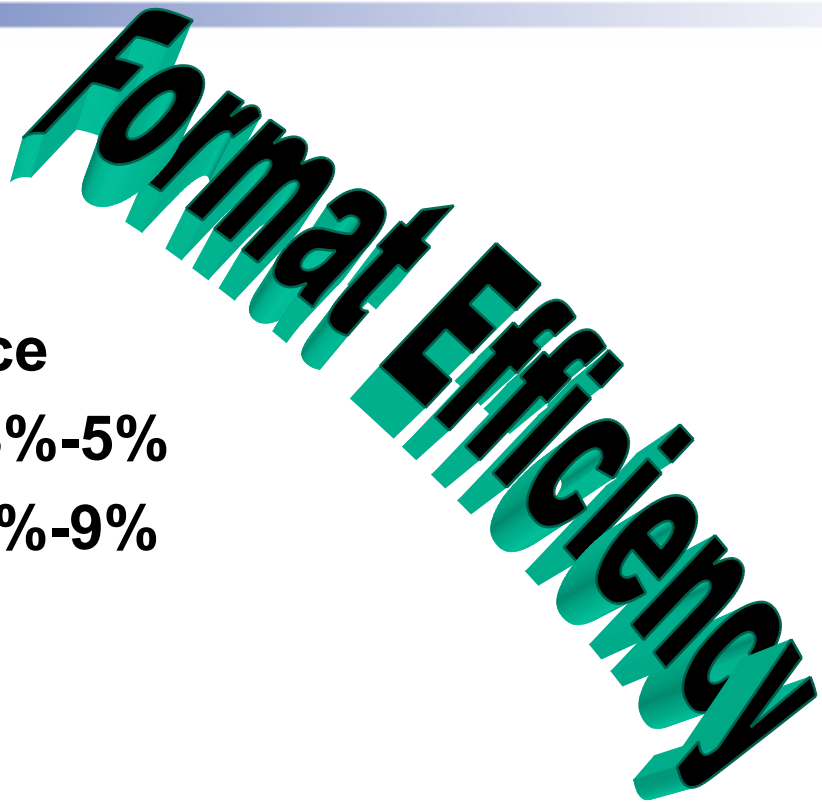
Maxtor®
What drives you.®

Agenda

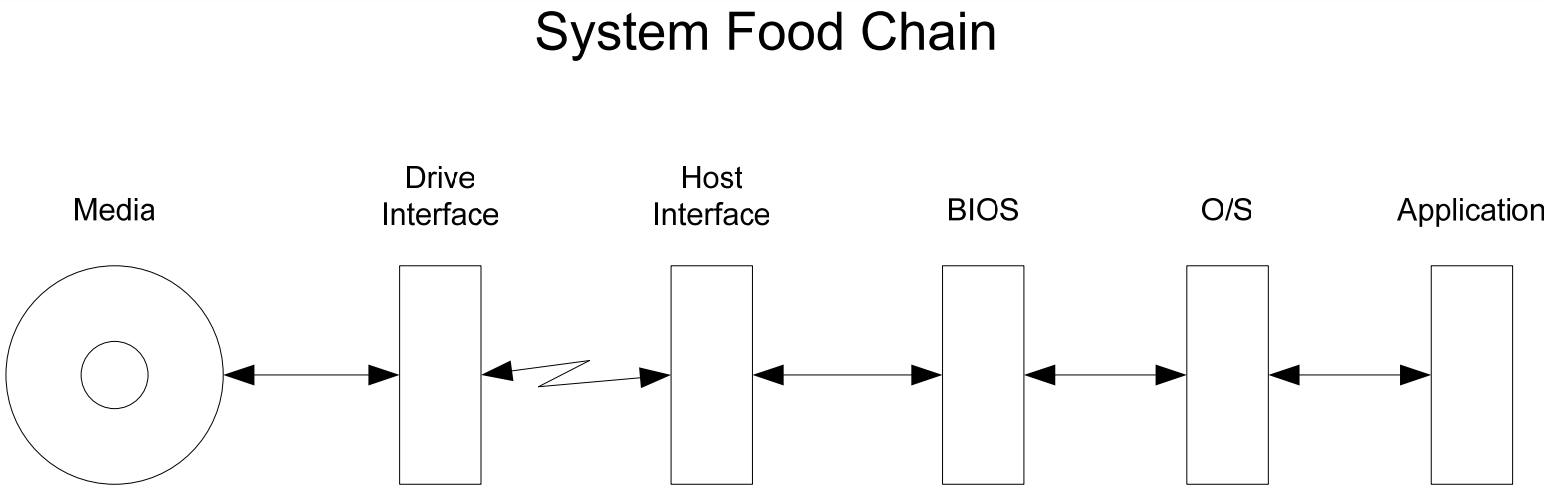
- Why larger sector sizes
- Issues with a technology transition to larger sectors
- Larger sector size implementation issues
- Conclusions

Why Larger Sector Sizes?

- Higher capacity
- Better data integrity
- Faster drive performance
- 1K sector size gains ≈ 3%-5%
- 4k sector size gains ≈ 5%-9%



What Are the Issues?



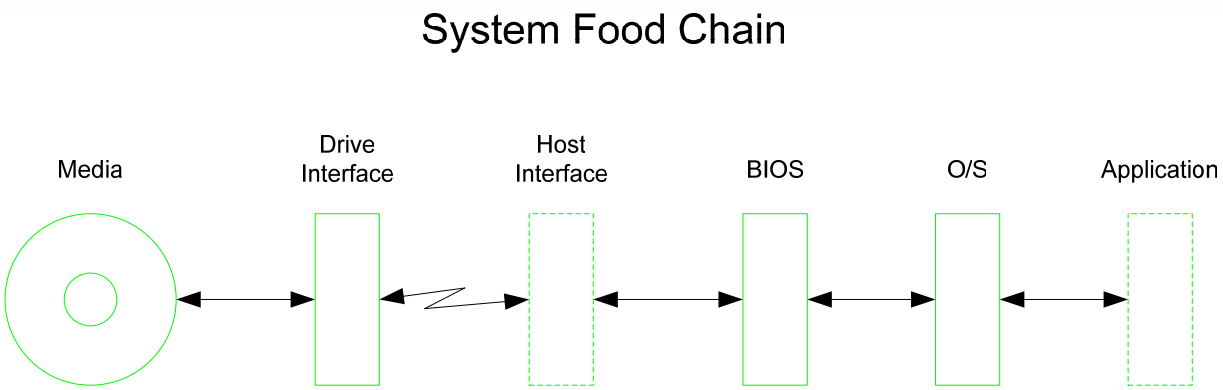
- Many points in the system are hardwired for 512 bytes
- Emulation mechanisms impact system performance
- Larger sector sizes at the drive interface impact compatibility

Our Direction

	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

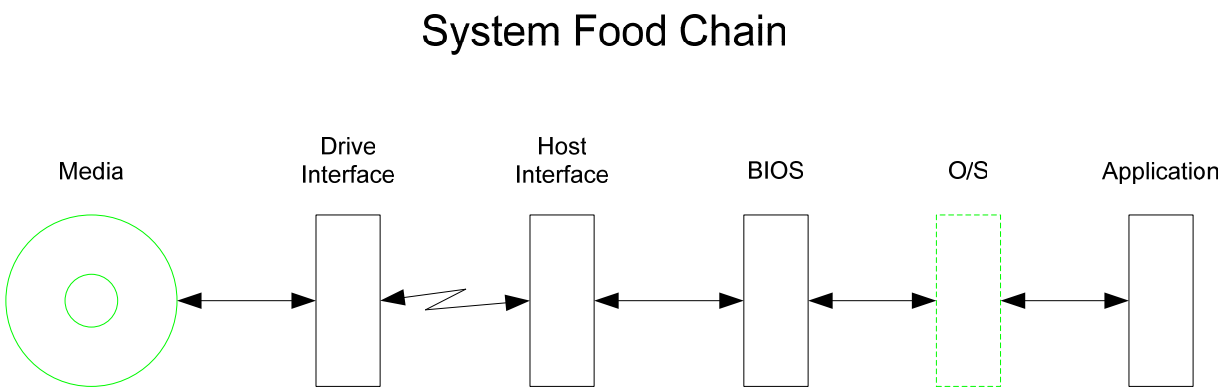
- **The technologically pure solution would be to require the food chain to deal with the issue of larger sector size**
- **Initial solution is to provide a 512 byte sector at the interface and a larger sector on the media**
- **Over time - can bridge to larger sector sizes at the interface**
- **At some point - will need to provide the host with a way to "turn on" larger sectors at the interface**
  - This is important for many server applications that would prefer the larger sector sizes
  - This would need to be a sticky setting preserved across power cycles

Food Chain Impacts – Future Implementation



- **Today, very few, or possibly no x86 systems will boot from an ATA device that does not return 512 byte sectors.**
- **Many host interfaces will handle the larger sectors, but**
  - Some will suffer efficiency issues
  - Others will not work at all
- **No windows OS currently accepts a larger sector size for ATA devices**
- **Some applications do not use the OS filesystem and will stop functioning**

Food Chain Impacts – Compatible Implementation



- **The format on the media is changed, the sector size at the drive interface remains the same**
- **The entire food chain remains functional**
- **The cost of this method is a possible decrease in performance**
  - The solution is alignment – If the OS aligns the filesystem on natural drive boundaries there is no decrease in drive performance, in fact there may be a slight increase in performance.
  - Applications that do not use the OS filesystem may also suffer in performance
- **All performance issues can be handled with software tools**
  - This will cause all of us some heartburn, however, we will be able to support our customers.

512 byte sectors and performance

- **PIO transfers have an overhead between each sector or group of sectors transferred**
  - Many systems that use PIO use interrupts to tell when the next sector is ready for transfer
  - READ/WRITE MULTIPLE cuts down on the number of interrupts
  - SATA implementations such as AHCI need not have any overhead since PIO transfers are accomplished using DMA
- **DMA operation has no sector based overhead**
  - A 64K transfer takes the same amount of time regardless of the transfer unit size – 512 byte, 1K, or 4K

Technology Transition Timeline

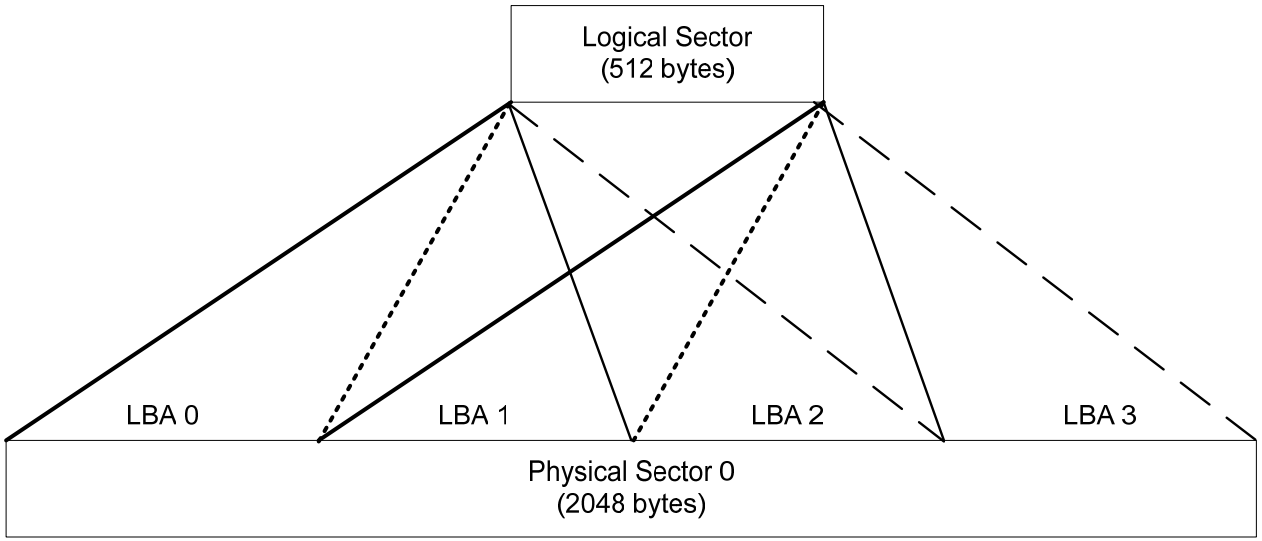
- **Disk Drives begin production in 2006**
- **Windows XP and its predecessors are the target operating systems**
  - These operating systems require 512 byte emulation to function
- **Longhorn release is projected for 2006**
  - It generally takes 2-3 years for a new MS operating system to mainstream This means that Longhorn will become the OS of choice in 2008-2009
- **BIOS development also takes time to reach the end user**
  - If there was a new implementation today, it would also achieve high levels of acceptance in 2008-2009
  - This will only succeed if motherboard vendors require the feature. If it seems like an add-on that will not be used immediately the transition will take longer

The Bottom Line

- **There will be approximately 50 million hard disk drives with 512 byte emulation distributed by 2009**
- **The end user will have a less than stellar experience because software tools are required to get the best system performance**
- **If Longhorn implements the alignment required for the 512 byte compatible solution, users will naturally be drawn to upgrade**
  - There will be no more tools required
  - OS will work more efficiently with the current technology

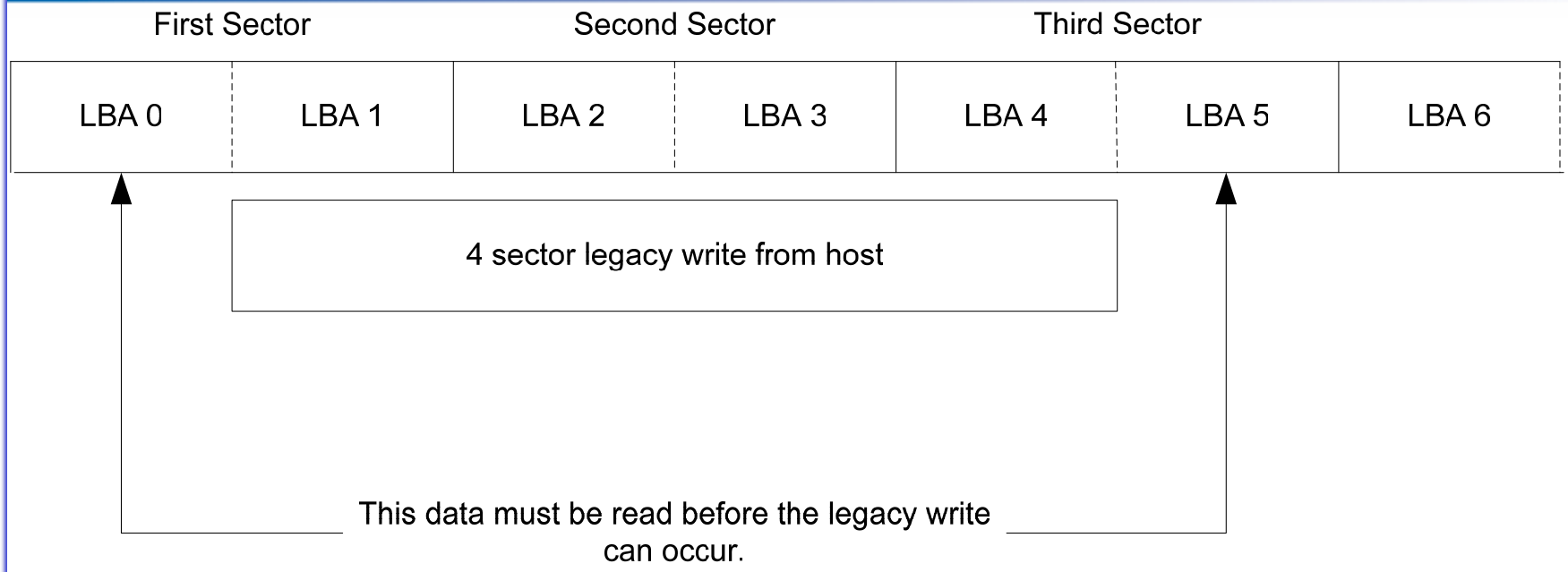
Implementation Issues

If we preserve the 512 byte sector for the host



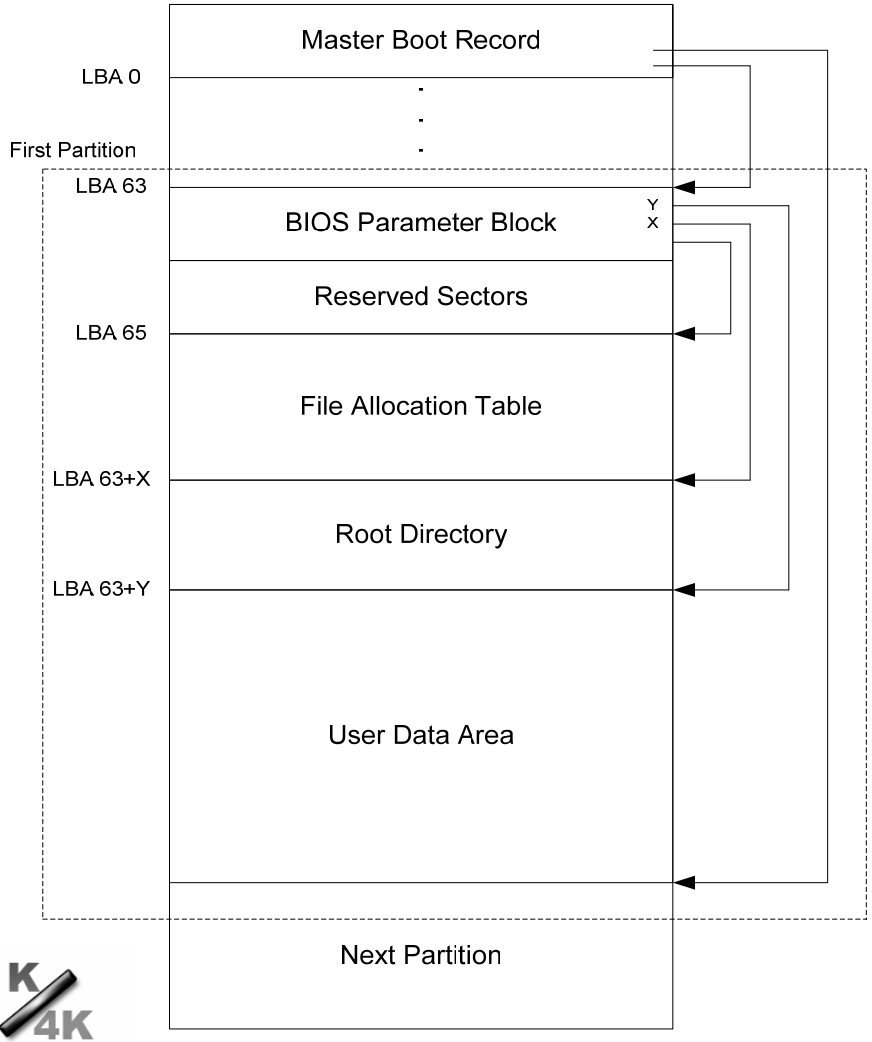
- **Un-enlightened hosts will have poor performance for a variety of reasons**
  - Unaligned transfers can require the drive to read the data before writing it.
- **Enlightened hosts could start transfers at the beginning of a physical sector and end them at the end of a physical sector.**

What is Read-Modify Write (RMW)



- **We do not see a measurable penalty for read operations**
- **We do not see any penalty for writes that begin and end on physical sector boundaries**
- **Writes that begin or end in the middle of a physical sector, as shown, will incur a performance penalty**
  - Up to 1 rev for writes within a track
  - Up to 2 revs for writes that span a track.

Typical Alignment Stumbling Blocks



- Partition starts on odd block number
- FAT starts on odd block number
- Root directory start varies based on length of FAT
- Cluster start locations vary based on structures listed above
- Some file-systems place data in-between the clusters.

The Need To Report Alignment

- **Drive vendors will customize their products to fit the target market**
- **Windows 3.1/95/98/me/NT/2000/XP naturally format on odd sector boundaries as shown earlier**
  - First Partition
- **Other operating systems naturally format on even sector boundaries**
- **This creates the need to report alignment requirements**
  - Using odd alignment target legacy systems, such as Windows XP, will function at near optimal performance
  - Newer systems can read device alignment requirements and format the media to work at peak efficiency
    - Addresses future legacy compatibility issues

Logical/Physical Alignment

Natural Alignment

Physical 0		Physical 1		Physical 2		Physical 3		Physical 4	
LBA 0	LBA 1	LBA 2	LBA 3	LBA 4	LBA 5	LBA 6	LBA 7	LBA 8	LBA 9

Odd Alignment

Physical 0		Physical 1		Physical 2		Physical 3		Physical 4	
------	LBA 0	LBA 1	LBA 2	LBA 3	LBA 4	LBA 5	LBA 6	LBA 7	LBA 8

Practical Issues 1K/4K At The Host Interface

- **Manufacturing**
  - Software tools are currently 512 byte oriented
  - Manufacturing OSes are limited to 512 bytes
  - Manufacturing hardware will need to be changed
- **OEM Logistics**
  - Impacts JIT, Forecasting, MPS, etc.
- **Retail – The drive won't function on many systems**
  - The System Food Chain does not allow a software solution on many systems
  - If the System Food Chain did permit a software solution
    - Need driver stack to talk to larger logical sector at the interface.
    - Need a tools package to optimize performance with existing OSes
      - Need special FORMAT and FDISK tools

The Industry Today

- **Need a utility package to support legacy OSes**
  - Windows 98/ME/2000/XP
    - FDISK and FORMAT will do it for most systems
    - The addition of a port driver would be a stronger solution
    - Filesystem adjustment could be used to optimize an existing installation
- **Linux community is already evolving**
  - Includes Apple

Call To Action

- We need to provide Microsoft with devices that support 512 byte emulation ASAP
- We need 512 byte emulation and alignment fully implemented in Longhorn

Contact Information

- **Maxtor**
  - Mark Evans
  - Mark_Evans@Maxtor.com
- **Seagate**
  - Marc Noblitt
  - Marc.A.Noblitt@seagate.com
- **Western Digital**
  - Curtis E. Stevens
  - Curtis.Stevens@wdc.com