

Hardware Innovation for Windows

Platform and component design for Microsoft® Windows® Operating Systems

Hybrid Hard Disk Support in Windows “Longhorn”

Microsoft Confidential. © 2005 Microsoft Corporation. All rights reserved. These materials are confidential to and maintained as a trade secret by Microsoft Corporation. Information in these materials is restricted to Microsoft authorized recipients only. Any use, distribution or public discussion of, and any feedback to, these materials is subject to the terms of the attached license. By providing any feedback on these materials to Microsoft, you agree to the terms of that license.

Abstract

The next version of the Microsoft® Windows® operating system, codenamed Windows “Longhorn,” will support a new type of hard disk drive (HDD) called a hybrid hard disk. Hybrid hard disks contain a small amount of solid-state nonvolatile memory which, when used as a cache by the HDD, can realize significant benefits such as faster booting, extended battery life, and a more ruggedized PC. This paper describes how hybrid hard disks work, hybrid hard disk support for Windows Longhorn, and the opportunity for PC vendors and component suppliers to build and sell systems that leverage them.

Contents

Introduction	3
Technology Opportunity	4
Hybrid Hard Disk Hardware Introduction.....	5
Hybrid Hard Disk Market Opportunity.....	8
Market Potential.....	8
User Scenarios.....	10
Hybrid Hard Disk Technical Description.....	12
Boot/Resume Time	14
Power Control.....	14
Prototype ATA Interface.....	15
Windows Logo Program Issues	19
Call to Action.....	19

Microsoft Corporation Technical Documentation License Agreement (Standard)

READ THIS! THIS IS A LEGAL AGREEMENT BETWEEN MICROSOFT CORPORATION ("MICROSOFT") AND THE RECIPIENT OF THESE MATERIALS, WHETHER AN INDIVIDUAL OR AN ENTITY ("YOU"). IF YOU HAVE ACCESSED THIS AGREEMENT IN THE PROCESS OF DOWNLOADING MATERIALS ("MATERIALS") FROM A MICROSOFT WEB SITE, BY CLICKING "I ACCEPT", DOWNLOADING, USING OR PROVIDING FEEDBACK ON THE MATERIALS, YOU AGREE TO THESE TERMS. IF THIS AGREEMENT IS ATTACHED TO MATERIALS, BY ACCESSING, USING OR PROVIDING FEEDBACK ON THE ATTACHED MATERIALS, YOU AGREE TO THESE TERMS.

1. For good and valuable consideration, the receipt and sufficiency of which are acknowledged, You and Microsoft agree as follows:

(a) If You are an authorized representative of the corporation or other entity designated below ("**Company**"), and such Company has executed a Microsoft Corporation Non-Disclosure Agreement that is not limited to a specific subject matter or event ("**Microsoft NDA**"), You represent that You have authority to act on behalf of Company and agree that the Confidential Information, as defined in the Microsoft NDA, is subject to the terms and conditions of the Microsoft NDA and that Company will treat the Confidential Information accordingly;

(b) If You are an individual, and have executed a Microsoft NDA, You agree that the Confidential Information, as defined in the Microsoft NDA, is subject to the terms and conditions of the Microsoft NDA and that You will treat the Confidential Information accordingly; or

(c) If a Microsoft NDA has not been executed, You (if You are an individual), or Company (if You are an authorized representative of Company), as applicable, agrees: (a) to refrain from disclosing or distributing the Confidential Information to any third party for five (5) years from the date of disclosure of the Confidential Information by Microsoft to Company/You; (b) to refrain from reproducing or summarizing the Confidential Information; and (c) to take reasonable security precautions, at least as great as the precautions it takes to protect its own confidential information, but no less than reasonable care, to keep confidential the Confidential Information. You/Company, however, may disclose Confidential Information in accordance with a judicial or other governmental order, provided You/Company either (i) gives Microsoft reasonable notice prior to such disclosure and to allow Microsoft a reasonable opportunity to seek a protective order or equivalent, or (ii) obtains written assurance from the applicable judicial or governmental entity that it will afford the Confidential Information the highest level of protection afforded under applicable law or regulation. Confidential Information shall not include any information, however designated, that: (i) is or subsequently becomes publicly available without Your/Company's breach of any obligation owed to Microsoft; (ii) became known to You/Company prior to Microsoft's disclosure of such information to You/Company pursuant to the terms of this Agreement; (iii) became known to You/Company from a source other than Microsoft other than by the breach of an obligation of confidentiality owed to Microsoft; or (iv) is independently developed by You/Company. For purposes of this paragraph, "Confidential Information" means nonpublic information that Microsoft designates as being confidential or which, under the circumstances surrounding disclosure ought to be treated as confidential by Recipient. "Confidential Information" includes, without limitation, information in tangible or intangible form relating to and/or including released or unreleased Microsoft software or hardware products, the marketing or promotion of any Microsoft product, Microsoft's business policies or practices, and information received from others that Microsoft is obligated to treat as confidential.

2. You may review these Materials only (a) as a reference to assist You in planning and designing Your product, service or technology ("Product") to interface with a Microsoft Product as described in these Materials; and (b) to provide feedback on these Materials to Microsoft. All other rights are retained by Microsoft; this agreement does not give You rights under any Microsoft patents. You may not (i) duplicate any part of these Materials, (ii) remove this agreement or any notices from these Materials, or (iii) give any part of these Materials, or assign or otherwise provide Your rights under this agreement, to anyone else.

3. These Materials may contain preliminary information or inaccuracies, and may not correctly represent any associated Microsoft Product as commercially released. All Materials are provided entirely "AS IS." To the extent permitted by law, MICROSOFT MAKES NO WARRANTY OF ANY KIND, DISCLAIMS ALL EXPRESS, IMPLIED AND STATUTORY WARRANTIES, AND ASSUMES NO LIABILITY TO YOU FOR ANY DAMAGES OF ANY TYPE IN CONNECTION WITH THESE MATERIALS OR ANY INTELLECTUAL PROPERTY IN THEM.

4. If You are an entity and (a) merge into another entity or (b) a controlling ownership interest in You changes, Your right to use these Materials automatically terminates and You must destroy them.

5. You have no obligation to give Microsoft any suggestions, comments or other feedback ("Feedback") relating to these Materials. However, any Feedback you voluntarily provide may be used in Microsoft Products and related specifications or other documentation (collectively, "Microsoft Offerings") which in turn may be relied upon by other third parties to develop their own Products. Accordingly, if You do give Microsoft Feedback on any version of these Materials or the Microsoft Offerings to which they apply, You agree: (a) Microsoft may freely use, reproduce, license, distribute, and otherwise commercialize Your Feedback in any Microsoft Offering; (b) You also grant third parties, without charge, only those patent rights necessary to enable other Products to use or interface with any specific parts of a Microsoft Product that incorporate Your Feedback; and (c) You will not give Microsoft any Feedback (i) that You have reason to believe is subject to any patent, copyright or other intellectual property claim or right of any third party; or (ii) subject to license terms which seek to require any Microsoft Offering incorporating or derived from such Feedback, or other Microsoft intellectual property, to be licensed to or otherwise shared with any third party.

6. Microsoft has no obligation to maintain confidentiality of any Microsoft Offering, but otherwise the confidentiality of Your Feedback, including Your identity as the source of such Feedback, is governed by Your NDA.

7. This agreement is governed by the laws of the State of Washington. Any dispute involving it must be brought in the federal or state superior courts located in King County, Washington, and You waive any defenses allowing the dispute to be litigated elsewhere. If there is litigation, the losing party must pay the other party's reasonable attorneys' fees, costs and other expenses. If any part of this agreement is unenforceable, it will be considered modified to the extent necessary to make it enforceable, and the remainder shall continue in effect. This agreement is the entire agreement between You and Microsoft concerning these Materials; it may be changed only by a written document signed by both You and Microsoft.

©2005 Microsoft Corporation. All rights reserved.

Microsoft, and Windows are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

The names of actual companies and products mentioned herein may be the trademarks of their respective owner



Introduction

The hard disk drive (HDD) is the storage media for both the operating system and most of the critical user data in a PC. The HDD typically represents a performance bottleneck and is also one of the major consumers of power in a mobile PC. In addition, the disk drive can be a significant and costly failure risk for the portable or mobile computer solution:

- Total loss of data can be catastrophic and expensive for the end user.
- Replacing a failed HDD under warranty is costly for the OEM or vendor.

The hybrid hard disk described in this whitepaper, when used with the release of Windows Longhorn, provides higher performance, greater reliability, and lower power consumption, particularly for mobile PC solutions. The hybrid hard disk has an integral solid-state, nonvolatile memory cache (NV Cache) that permits read and write operations to the drive while the spindle is stopped. The price of nonvolatile memory technologies such as NAND-based flash is low and predicted to drop further in the future, making it cost-effective to consider large cache sizes of 64–128 MB or greater.

The onboard disk controller for a hybrid hard disk supports an extension to the ATA interface, which allows Windows Longhorn to use this NV Cache. Windows Longhorn provides an NV Cache manager to programmatically control the logical block addressing (LBAs) that are mapped (or pinned) into the NV Cache, and to control the Power Management mode of the HDD. The onboard disk controller also has the flexibility to accept ad-hoc writes and to buffer this data when the drive spindle is stopped. This ad-hoc data is described as unpinned data, and is flushed out to magnetic media when all of the NV Cache is used, or the spindle is started.

The following new features of Windows Longhorn are improved by the addition of a hybrid hard disk:

- **Faster Boot and Resume** – The end-user experience of slow PC startup times (from S4 and S5) is due in part by the startup time and latencies of the disk drive. By including an NV Cache in the disk drive, data can be read and written to the disk subsystem prior to the disk spindle reaching the required speed to access data on the magnetic media. It may also be possible to sustain a faster data transfer from the NV Cache during Boot and Resume than is possible from disk. Boot and Resume times for a hybrid hard disk based system should be 2-5 seconds faster than for a conventional drive.
- **Longer battery life** – The end user can expect to see up to a 15% improvement in battery life with a hybrid hard disk, depending on system configuration and usage patterns. For a mobile PC providing up to 5 hours of operation, the hybrid hard disk may extend this by up to 30–45 minutes.
- **Higher data reliability** – By keeping the disk spindle spun down and the heads parked for long periods of time, the HDD is far less susceptible to failures that can occur when mobile PCs are moved, jostled, and dropped. In addition, the wear and tear on the drive is reduced if the spindle can be turned off for long periods, extending the lifetime of the drive.

Technology Opportunity

Microsoft's aim with the Windows Longhorn release is to provide increased end-user functionality and an enhanced Windows experience by improving all aspects of the operation of a PC. Windows Longhorn and the hybrid hard disk provide the opportunity for significant gains for the end user, PC manufacturer, and HDD manufacturer.

Table 1: Summary of Hybrid Hard Disk Benefits

Market Segment	Benefits
End user	<ul style="list-style-type: none"> • Faster Boot and Resume times • Longer battery life • Higher data reliability
PC Manufacturer	<ul style="list-style-type: none"> • Ability to differentiate and penetrate the Windows mobile PC market • Higher performance and feature-to-cost ratio • Reduced warranty and return costs
HDD Manufacturers	<ul style="list-style-type: none"> • Ability to differentiate and penetrate the growing 2.5 inch HDD market in the Windows mobile market • More reliable and rugged disk drives with longer mean time between failures (MTBF) • Reduced power consumption

Windows Support for Hybrid Hard Disks

Windows Longhorn will support the new hybrid hard disk technology to achieve faster Boot and Resume times and longer power off periods for the disk drive spindle. This will extend the battery life of mobile devices that use Windows Longhorn, and will improve the overall reliability of these systems.

The new Windows Longhorn features work together to permit the spindle on a hybrid hard disk to be powered down for long periods of time, while still permitting read and write requests to the drive. These new features include a new system for memory management and an NV Cache manager. When the spindle is powered down, the HDD is less prone to mechanical wear and failure, has reduced susceptibility to external shock, and has reduced power consumption.

The improved memory management features in Windows Longhorn buffer both the operating system and user applications in dynamic random access memory (DRAM) to speed user access to data and to minimize random disk read activity. This read caching allows more aggressive power control for the HDD, and increases the time that the drive spindle can be powered off.

The NV Cache manager is used to control the power mode of the hybrid hard disk and to map (or pin) the required LBAs into the drive's NV Cache. The NV Cache manager maps the following types of information into the NV Cache:

- On Shutdown, sectors that will be read to accelerate cold boot
- On Sleep, sectors required for Resume functions when the PC is shut down or put into the Hibernate mode

Hybrid Hard Disk Hardware Introduction

A hybrid hard disk appears as a standard ATA HDD to the system with support for one additional new power mode (NV Cache mode) and a new ATA command for managing the NV Cache. The hybrid hard disk includes an NV Cache in addition to the DRAM cache. The NV Cache allows both read and write requests to the drive to be resolved while the drive spindle is powered down.

Figure 1 shows a block diagram of the hybrid hard disk in a PC that is running Windows Longhorn.



Figure 1: Illustration of Hybrid Hard Disk Interaction with Windows Longhorn

NV Cache Mode

The NV Cache mode permits the hybrid hard disk onboard disk controller to aggressively power down the disk drive spindle. The controller also uses the NV Cache to buffer write requests to the drive. The onboard controller maintains a map of LBAs currently in the NV Cache and maintains coherency between the NV Cache and magnetic media.

The NV Cache is logically separated into a pinned and unpinned data set, with Windows Longhorn responsible for defining the pinned set and the drive controller responsible for the unpinned set. The onboard disk controller decides where data is read from or written to, based on whether the LBAs addressed are pinned or unpinned in the NV Cache. The following definitions may be helpful in understanding the content of the NV Cache:

- **NV Cache Set** – The set of LBAs currently stored in the drive NV Cache. The onboard disk controller manages the NV Cache set based on the needs for programmatically pinned data and unpinned data generated by buffering of write operations to the disk.

- **NV Cache Pinned Set** – The set of LBAs specified by the host as needing to be present in the NV Cache. Windows will add LBAs to the NV Cache pinned set to ensure these specific LBAs can be accessed when the disk is spun down or still in the process of spinning up.
- **NV Cache Unpinned Set** – The set of LBAs that are represented in the NV Cache set but not represented in the NV Cache pinned set. The NV Cache pinned set and the unpinned set are mutually exclusive. The onboard disk controller creates unpinned set data while buffering write IO from the host and will flush this data set to magnetic media when no more unpinned cache is available, or opportunistically when the drive spindle is powered on.

Active Mode Operation

During normal (Active mode) disk operation, when the disk spindle is powered on, it is the purview of the onboard disk controller to determine the location of fresh data and whether read IO should be satisfied from the NV Cache if possible or from the magnetic media. This will likely depend on the relative performance of the nonvolatile memory being used for the NV Cache to that disk media accesses and throughput. The performance of a hybrid hard disk in this mode is expected to be as good as or better than that of a standard HDD.

Transitioning to NV Cache mode of the hybrid hard disk to save power would be initiated by Windows Longhorn when the PC was transitioned to being used on battery power. More specifically, Windows Longhorn will set the hybrid hard disk into NV Cache mode when entering a power conserving mode of Windows Power Management; this may be a user- or OEM-configured power management option. Upon transitioning back to AC power, or a Windows power management mode which does not conserve power, the hybrid hard disk would be returned to Active mode.

In a transition from NV Cache mode to Active mode, it is possible that data may still exist in the NV Cache. It is recommended that the onboard disk controller flush this data back to disk in a timely fashion. If the data is written too often and left mapped into the NV Cache, this may reduce the life expectancy of memory blocks within the NV Cache memory, particularly if no wear-leveling algorithms are implemented in the disk controller.

NV Cache Mode Operation

When the drive is put into NV Cache mode, the onboard disk controller will aggressively power down the drive spindle and any subsequent writes to the drive will be stored in the NV Cache. If the NV Cache fills with write data, or if the drive must resolve a read request from the magnetic media, the onboard disk controller will:

- Power up the spindle to complete the operation.
- Flush the NV Cache unpinned set to magnetic media.
- Perform any vendor specific operations or maintenance, if needed.
- Power down the spindle.

Note Depending on the size of the NV Cache, the onboard disk controller may not flush all data in the NV Cache unpinned set to magnetic media when the spindle is powered up, but this is implementation-specific and each HDD manufacturer may provide unique algorithms for cache management.

In this mode of operation, Windows Longhorn does not define the LBAs that the NV Cache contains (the unpinned data set). The hybrid hard disk’s onboard disk controller simply stores write data in the NV Cache and maintains a list of the LBAs that are currently in the NV Cache. This data is called the NV Cache unpinned set, and the

onboard disk controller is responsible for maintaining coherency of the NV Cache and magnetic media.

Figure 2 shows the duty cycle of the hybrid hard disk entering into NV Cache mode, where the spindle is powered up for only a short period of time to satisfy requests that could not be resolved from either Windows Longhorn read buffers or from the NV Cache in the drive.

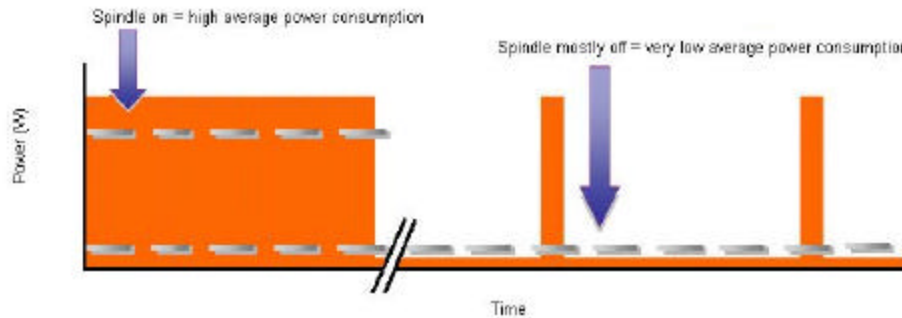


Figure 2: How the Disk Spindle Duty Cycle Reduces Power

NV Cache Data Management

The onboard disk controller is responsible for managing the NV Cache and for constantly maintaining coherency of the data written to the drive. The NV Cache is logically separated into a pinned (NV Cache pinned set) and unpinned (NV Cache unpinned set) data sets by the onboard disk controller, with Windows Longhorn responsible for defining the pinned set, and the controller responsible for the unpinned set. The onboard disk controller will read and write to either the NV Cache or to magnetic media, or to both, depending on the state of the spindle and the algorithms implemented in the firmware.

In either Active or NV Cache mode, Windows Longhorn can programmatically define LBAs that should be pinned in the NV Cache. In defining this pinned set, Windows Longhorn can either:

- Request that this data be read from the magnetic media into the NV Cache.
- Request that the data not be read from magnetic media and wait until the data at those LBAs is written to the disk instead.

Requesting that the data not be read from magnetic media can save time, but it is then the responsibility of Windows Longhorn to write this data to the drive. This is useful when Windows Longhorn is about to write information, such as the hibernation state files that will be used for a resume operation.

The onboard disk controller is responsible for managing bad physical blocks within the NV Cache. Because flash memory may degrade (such as when blocks no longer support data integrity), the controller may reduce the available NV Cache memory size returned to Windows Longhorn over time. Windows Longhorn must be able at runtime to dynamically adjust the use of NV Cache if the size drops. In the worst case scenario, where there is a total failure of the physical memory in the NV Cache before failure of the drive itself, the HDD will revert to a conventional mode of operation.

Hybrid Hard Disk Market Opportunity

When the spindle is powered down the hybrid hard disk provides a low power, low latency write capability with write bandwidth similar to the nonvolatile memory-based solid state storage devices (SSDs). While not an SSD, the hybrid hard disk provides many of the benefits with much reduced overall cost. Future cost reduction and density improvements in nonvolatile memory will apply to the hybrid hard disk, while also benefiting from the increasing capacity of magnetic media.

SSDs are often seen as ideal HDD replacements because of their high shock resistance and lack of rotational and seek delays. Regardless of their underlying technology (DRAM or nonvolatile memory), they are not the ideal solution for mobile PCs. DRAM-based SSDs are typically faster than an HDD, but are high power, high cost, and volatile devices suitable mainly for the server market. Nonvolatile memory based SSDs have been traditionally slower than an HDD for data transfers, although they have no rotational or seek delays. These nonvolatile memory-based SSDs are available in the mobile market, but to become mainstream options, they are dependant on significant future cost reduction and density improvements in nonvolatile memory. Based on current density and cost predictions, it is unlikely that fully solid state solutions of viable capacity will be available in the same time frame as the hybrid hard disk, and SSDs are therefore unlikely to become a mainstream solution in the near-term.

Considering the end user and OEM/ODM benefits of Windows Longhorn support for hybrid hard disks, they should be available as a selectable option sometime in 2006, and as a mainstream option for mobile solutions by 2008. Any cost, density or performance advances achieved by the nonvolatile memory industry can be quickly incorporated into a hybrid hard disk with immediate potential end user gain.

Market Potential

Future mobile HDD shipments will continue to shift toward higher capacities as aerial densities increase. Currently, most 2.5-inch HDDs ship with storage capacities of 20–49.9 GB, and are expected to move to higher capacity 50–99.9 GB drives in the Windows Longhorn introduction time frame. The total worldwide market size for 2.5-inch drives is projected to exceed 60M units in 2006.

In the future, the mobile PC HDD market will begin to adopt smaller 1.x drives as the cost per gigabyte comes down, but for the near to mid term 2.5-inch drives are the most cost effective, and therefore the largest market potential for the hybrid hard disk.

Targeted geographies include markets such as U.S., Western Europe, and Japan, which account for 77% of the total mobile PC market and 26% of worldwide PC sales by 2006. Table 2 lists past and projected worldwide PC and notebook sales. (Source: IDC, Worldwide Quarterly PC Tracker, Figure 3, February 2005.)

Table 2: Past and Projected Notebook Sales by Region

Portable PC Shipments By Region & Share Of Total Client PCs				
Region	2003 Units	2004 Units	2005 Units	2006 Units
USA	13,807,701	16,623,579	20,230,305	24,852,720
Western Europe	11,628,199	15,266,475	18,393,556	21,088,332
Japan	6,256,300	6,544,589	6,707,890	6,861,100
Subtotal Tier One	31,692,200	38,434,643	45,331,751	52,802,152
Tier One Share - WW Portable PCs	81%	78%	78%	77%
Tier Two Share - WW Client PCs	21%	22%	24%	26%
Subtotal Tier Two	7,664,092	10,594,873	13,065,359	15,514,945
Tier Two Share - WW Portable PCs	19%	22%	22%	23%
Total	39,356,292	49,029,516	58,397,111	68,317,096
Total Portables Share - Client PCs	26%	28%	31%	33%
Total Client PC's	150,096,960	172,392,593	189,151,613	205,411,847

Note: Unit Data does not include x86 Server Shipments

Source: IDC Worldwide Quarterly PC Tracker, February 2005

Today's Market Horizon

The overall demand in 2003 for mobile-class HDDs increased 41% (year over year) with nearly 51 million drives shipped. The revenue for 2003 (\$5.1 billion) set a new record high against a backdrop of increasing ASPs (\$100, a \$4 increase from 2002). IDC expects mobile-class HDD shipments to increase again in 2004 by approximately 28% from 2003, to almost 67 million units. This represents a revenue base of over \$6 billion (a 25% increase over 2003). A large part of this demand has been generated by the introduction of HDD storage into consumer electronics devices, such as digital video recorders, portable music players, and other portable devices. As additional suppliers launch mobile-class products, aggressive pricing tactics are expected to impact the forward-looking revenue opportunities of new and existing HDD manufacturers.

2.5-inch HDDs

In 2003, shipments of HDDs were up 37% from 2002 to 47.3 million units. Revenue also increased from \$3.3 billion (in 2002) to around \$4.7 billion (in 2003); this represents a 43% increase. All of the 2.5-inch vendors released 40GB/platter products during 2003. Shipments of 5,400 rpm disk drives with the same capacity points occurred in 2003, although volumes and yields were lower. During the year, 2.5-inch 5,400 rpm drives represented about 17% of all 2.5-inch mobile-class HDD shipments. Seagate, which entered the market in the second half of 2003, shipped around 28% of the industry's 5,400 rpm 2.5-inch drives. Hitachi, determined to maintain its 2.5-inch mobile-class HDD leadership, launched a 7,200 rpm configuration leveraging new femto head technology that reduces power consumption to that of a 5,400 rpm drive and improves non-operating shock (by reducing mass at the end of the suspension). Given its price premium, the product saw shipments that were quite limited (less than 500,000). This limited demand trend will likely remain unchanged until additional competition gives rise to more aggressive price erosion. (Source: IDC report #31208 "WW Hard Disk Drive 2005-2008 Forecast and Analysis".)

Additional Market Potential

While the mobile PC market is the initial target market for hybrid hard disk technology, there also is potential application for this technology in desktop and enterprise server systems, as described in this section.

Desktop Systems

If a user leaves applications open in a typical desktop system, the system may never power down the HDD. Particularly for corporate desktops, this means that the desktop

system may always be powered on, or, at best, only turned off during weekends or holidays.

Windows Longhorn and the hybrid hard disk can, under this continuously powered scenario, permit the hard disk to power down for extensive periods. These extended power-down periods should dramatically reduce the number of hard disk failures and power consumption in large client-based environments.

In addition, desktop and workstation systems typically have higher-performance drives, which spin at a higher speed (up to 15K rpm), and take longer to power up than lower-performance drives. The hybrid hard disk can significantly shorten the Boot and Resume time for these systems.

Server Systems

In a typical server or data center environment, there may be many configurations using direct attached storage. Disk drives in these configurations are typically selected for the highest performance, with little regard to power dissipation.

With the impact of serial ATA (SATA) on the storage marketplace, the hybrid hard disk that uses a SATA interface has the potential to significantly reduce the power budget and disk replacement costs for server-based storage arrays.

Typical HDDs in these enterprise environments consume 8-12W, and with storage capacities requiring hundreds of drives in a typical storage area network (SAN) system, hybrid hard disks could result in significant power consumption reductions.

User Scenarios

The primary scenario is a user who buys a new mobile PC that is running Windows Longhorn. The user will use their PC to perform any number of typical computing tasks and will experience an 8–12% longer battery life and a fast resume from their Sleep state. The user is also less likely to experience catastrophic disk failure and loss of data due to rough handling of the mobile PC. The user experiences no degradation in performance when plugged into AC power, and only minimal delays when opening rarely used files and applications while operating on battery power with the hybrid hard disk spindle powered off.

Table 3: Summary of Hybrid Hard Disk User Scenarios

Scenario	Description
Faster Boot	When starting a laptop which has been turned completely off, users see their PC boot quickly and consistently to the logon screen or to the desktop. The hybrid hard disk is able to instantly provide data that has been pinned in the NV Cache when power is applied; this eliminates the delay waiting for the disk to become ready.
Resume from Sleep	When users suspend their laptop by using the Windows Longhorn Sleep mode, Resume time is less than 2 seconds if they are in S3. If they are in S4, the Resume time may be slightly longer due to the reading of the hibernate file from the NV Cache. Windows Longhorn will resume very rapidly from S3 because the contents of DRAM are maintained and are simply re-initialized without reading from the disk drive. In S4, Windows Longhorn pins the LBAs associated with the files for the resume process before entering S4, and is able to load them faster after power applied to the HDD.

Scenario	Description
Running on AC Power	<p>When AC powered, the hybrid hard disk will behave by default as a standard HDD, and the disk spindle will not be powered down to save power.</p> <p>Windows Longhorn will still place sectors needed for Boot or Resume in the NV Cache during Shutdown or Suspend to ensure the faster Boot or Resume experience.</p>
Running on Battery	<p>When running their PC on battery power only, the hybrid hard disk will be put by default into NV Cache mode to save power. Typically, users will not perceive a degradation of their user experience except when they use a file or application they haven't used in some time. In this scenario, a short delay occurs while the disk is spun up to access the file. Occasionally, users may be able to hear the HDD spin-up and write data to disk. After a short period of time (10–20 seconds), the HDD will turn off and be completely silent once again.</p> <p>If users perform a large number of writes while running on battery, such as writing a DVD or otherwise copying or streaming a large amount of data to the disk, they may detect a slow start to the write process, followed by the sound of their HDD spinning up, and higher performance of the write data rate as the disk comes up to full speed.</p>

Windows User Interface for Hybrid Hard Disk

The Windows Longhorn user interface for the hybrid hard disk will present the user with options to enable or disable the use of the NV Cache. If the NV Cache is disabled, the drive will function as a conventional HDD.

The properties of the drive will indicate if the drive is a hybrid and will also provide the NV Cache size built into the hybrid hard disk. The cache size returned by the drive will be dynamic. Over time, wear failures in the NV Cache may result in a reduction of the cache size reported to Windows Longhorn.

Hybrid Hard Disk Technical Description

The hybrid hard disk will connect to a mobile PC by using either a standard parallel ATA (PATA) or serial ATA (SATA) interface. The onboard disk controller will implement the ATA extensions necessary to enable NV Cache operation and functionality.

Operational Scenarios

There are six hybrid hard disk operational scenarios defined by the current PC operation mode. Windows Longhorn will turn off the hybrid hard disk NV Cache mode when the PC is connected to AC power, and turn on the NV Cache mode when running on battery power. The functions provided by the hybrid hard disk vary depending on the PC operational mode. Table 4 lists each scenario and the corresponding PC operations.

Table 4: Hybrid Hard Disk Functions

PC Operational Mode	Operational Functions
Running on AC Power	<ul style="list-style-type: none"> • Hybrid hard disk keeps its platter spinning according to Windows Longhorn power management policy for all HDDs . This enables highest performance, but with some sacrifice of MTBF and shock protection. • NV Cache unpinned set is flushed to the magnetic media . • Read IO to the disk may be satisfied from NV Cache (better performance) or from magnetic media. • Windows Longhorn can specify that sectors can be transferred from magnetic media to NV Cache by using an enhanced ATA command specification.
Running on Battery (Hybrid Hard Disk Enabled)	<ul style="list-style-type: none"> • The hybrid hard disk spindle spins up when NV Cache is full, a read occurs which cannot be satisfied from NV Cache, or an ATA command to change to Active mode is received. The onboard disk controller then: <ul style="list-style-type: none"> ○ Performs any required read IOs. ○ Flushes valid NV Cache unpinned set data to magnetic media. ○ Has the option of writing data directly to the magnetic media until spin-down is initiated. • Onboard disk controller spins down the platter.
S4 Hibernate/Power Off	<ul style="list-style-type: none"> • Windows Longhorn sets the hybrid hard disk's mode to Active resulting in the disk spinning up. • Windows Longhorn flushes the NV Cache to magnetic media. • Windows Longhorn copies all files needed for Boot (hibernate file, NTLDR, metadata files , etc.) from disk to NV Cache pinned set by issuing pin commands .

PC Operational Mode	Operational Functions
S5 Full Power On	<ul style="list-style-type: none"> • BIOS runs and initializes HDD (which starts spindle spin-up) and reads the loader file from NV Cache. • The loader file reads needed files from the disk using standard ATA Read commands. As long as they are present, the onboard disk controller supplies these sectors from NV Cache. • Additional necessary sectors are read from disk after it spins up. • Once Boot has concluded, some or all of the pinned set will be removed to make space available in the NV Cache for unpinned set. • Windows Longhorn memory management will continue to fill DRAM memory cache. • Windows Longhorn spins down hybrid hard disk according to current power management policy.
S3 Resume	<ul style="list-style-type: none"> • Memory is still valid. • Windows Longhorn performs disk spin-up. • Windows Longhorn flushes NV Cache to magnetic media. • Windows Longhorn performs any desired reads or writes. • Windows Longhorn spins down hybrid hard disk according to power management policy.
S4 Resume	<ul style="list-style-type: none"> • BIOS runs and initializes HDD. • The loader file sectors are supplied from NV Cache when read. • Hibernate file sectors are supplied from NV Cache when read. • Any remaining sectors should be supplied from whichever media is faster once the rotating media has spun up. • Windows Longhorn removes some or all of the pinned set to make space for unpinned set • Windows Longhorn makes any needed reads/writes to the disk while it remains spinning • Windows Longhorn spins down hybrid HDD according to power management policy.

NV Cache Size

Initial performance predictions based on the current builds of Windows Longhorn indicate that the recommended cache size should be at least 128 MB. The larger the cache size, the more flexibility Windows Longhorn has in allocating disk sectors to NV Cache, but benefits may be realized for smaller cache sizes such as 64 MB. The typical requirements to map disk data required for cold Boot and Resume are in the 100–120 MB range, indicating that 128 MB or more would be the most desirable NV Cache size.

Future price drops in the cost of nonvolatile memory may eventually make larger cache sizes such as 512 MB–1 GB practical.

The other factor in NV Cache size is the capacity needed to keep unpinned set used for buffering write data during power savings. Figure 3 shows the amount of data written in a 10-minute period by active users of PCs (both desktop and mobile).

- Field data from Windows XP-based PCs
 - >9,000,000 10min "active use" intervals
 - 90% of 10min active use intervals < 64MB unique data written

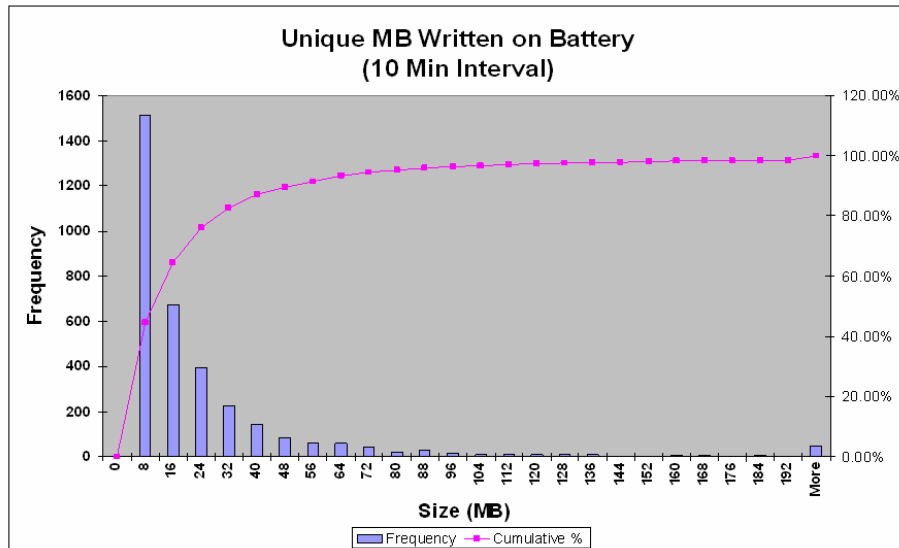


Figure 3: Establishing Cache Size

Boot/Resume Time

The time a Windows XP-based mobile PC takes to transition from S4 (Hibernate) or S5 (Off) to a usable S0 (On) state is inconsistent across platforms and can take between 8 and 60 seconds. Windows Longhorn client team has set aggressive targets to make Longhorn the fastest booting and resuming client to date with a consistent experience for the end user.

By keeping the Windows Longhorn loader file and other required files in the NV Cache, there can be a significant reduction in Boot and Resume times. The amount of time saved will depend largely on:

- The system having a fast POST BIOS which transfers to the loader file as quickly as possible.
- The speed at which data can be read from the NV Cache.
- All other standard system considerations for Boot and Resume time including CPU speeds, HDD RPM, DRAM memory speeds and good BIOS INT13 implementation.

Other aspects of Boot time acceleration in Windows Longhorn are not covered in this document.

Power Control

This section describes the hybrid hard disk power signature.

The HDD in a mobile PC can be in one of several power states:

- **Drive powered off** – This is where internal logic removes the power from both the onboard disk controller electronics and the drive system for the HDD. This is the lowest power state and is typically represented by only status logic and interface leakage. This state is typically set in the PC BIOS.
- **Controller powered on, spindle stopped** – This is a lower power state, where the disk drive is powered up to process read or write requests. This state may be managed by Windows Longhorn or third party disk control utilities.
- **Controller powered on and spindle powered** – This is the idle-active state, represented when no track-to-track searches are required. There is a higher transitory power level where the disk is actively seeking or reading/writing data to a cylinder.

There is also an important transitory power state that occurs in normal operation of a disk drive:

- **Starting the disk drive spindle** – This is typically the largest power draw for the disk drive, but occurs only for the spin up time of the spindle.

To take advantage of the hybrid hard disk, the spindle needs to be powered off long enough to gain a power advantage over either the idle-active state or the spindle power cycling state of a conventional HDD.

Figure 4 shows the data acquired on an instrumented PC using prototype hybrid hard disk architecture to test power savings.

- **80% average reduction in power (from 1.2W to 0.24W)**
- **Assumptions**
 - $P_{avg\ active} = 1.2W$ (measured)
 - $P_{avg\ with\ Flash\ write\ buffer\ and\ Longhorn\ kernel} = 0.18W$ (calculated)
 - $T_{off} = 600s @ .18W$
 - $T_{on} = 15s @ 2.5W$
 - $T_{on} = \text{spin-up time (2s)} + \text{Flash buffer flush time (13s)}$
 - » Flash buffer size = 128MB
 - » Transfer rate = 10MB/s
 - » $P_{avg} = (600 * 0.18 + 15 * 2.5) / 615 = 0.24W$

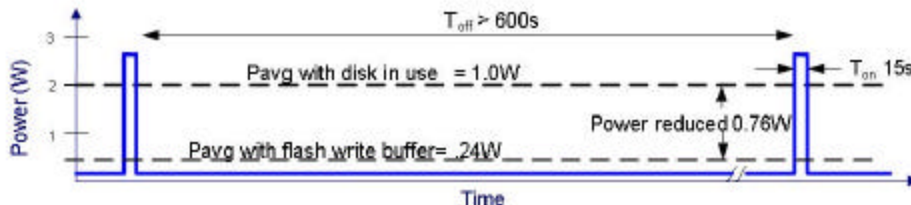


Figure 4: Power Savings Comparison for a 2.5-inch Hybrid Hard Disk

Prototype ATA Interface

This interface description is used in conjunction with the current ATA7 specification, and it contains extended ATA commands designed to identify and manage an NV Cache contained within a hybrid hard disk.

The modifications and additions to the ATA7 specification detailed in this document are the minimum set of functionality that a prototype must support for development testing. Based on prototype functional testing, a proposal to add commands to the future ATA8-ACS draft specification will be made.

Note The ATA7 command extensions used for construction of prototype devices may not represent the command set finally ratified by the T13 committee.

Exceptions and Changes to ATA7 for Hybrid Hard Disk Technology

This interface description is to be used in conjunction with the ATA7 specification and contains extended ATA commands designed to identify and manage an NV Cache contained within a hybrid hard disk.

Note The ATA7 command extensions used for construction of prototype devices may not represent the command set finally ratified by the T13 committee.

The following ATA7 Volume 1 sections are modified to support the NV Cache:

- The Identify Device command (ATA7 Volume 1 Section 6.17) is modified to be able to identify device support for the new commands and power state.
- The Set Features command (ATA7 Volume 1 Section 6.49) is modified to be able to identify host support for the new commands and power states.
- The Check Power Mode command (ATA7 Volume 1 Section 6.8) is modified to reflect the new NV Cache power modes.

The following feature set is added to the Command Descriptions (ATA7 Volume1 Section 4):

- The NV Cache power management feature set.

The following commands are added to the Command Descriptions (ATA7 Volume1 Section 6):

- Add LBAs To NV Cache Set
- Remove LBAs From NV Cache Set
- Set NV Cache Power Mode
- Return From NV Cache Power Mode
- Synchronize NV Cache
- Query NV Cache Pinned Set
- Query NV Cache Misses

Hybrid Hard Disk Power Management

Because the power management feature set is mandatory for all ATA devices, a new power mode that exists outside of the power management feature set is needed for hybrid hard disk. Figure 5 illustrates how adding this power mode changes the Power Management State Diagram. (Source: ATA7 Volume 1 Figure 4.)

When a hybrid hard disk is in the NV Cache power mode, the device attempts to remove power from its rotational media and satisfy all reads and writes from the NV Cache. If a hybrid hard disk is not able to resolve a read or write from its NV Cache, it shall satisfy the read or write by powering up the spindle and reading from the magnetic media. Each manufacturer can define the length of time that the device powers up its spindle.

If the device uses a SATA interface and has powered its rotational media, incurring a spin-up, it may send an Asynchronous Notification to the host. (Source: Serial ATA 1.0a Extensions revision 1.2.)

When the device is in the NV Cache power mode, all writes to LBAs that are in the devices NV Cache pinned set shall update the NV Cache as a write-through operation.

The NV Cache power mode and the power modes defined in the power management feature set (Source: ATA7 Volume 1 Section 4.5.3) are mutually exclusive. When a device is in the NV Cache power mode, it may only transition to the active power mode when commanded to do so by a received RETURN FROM NV CACHE MODE command.

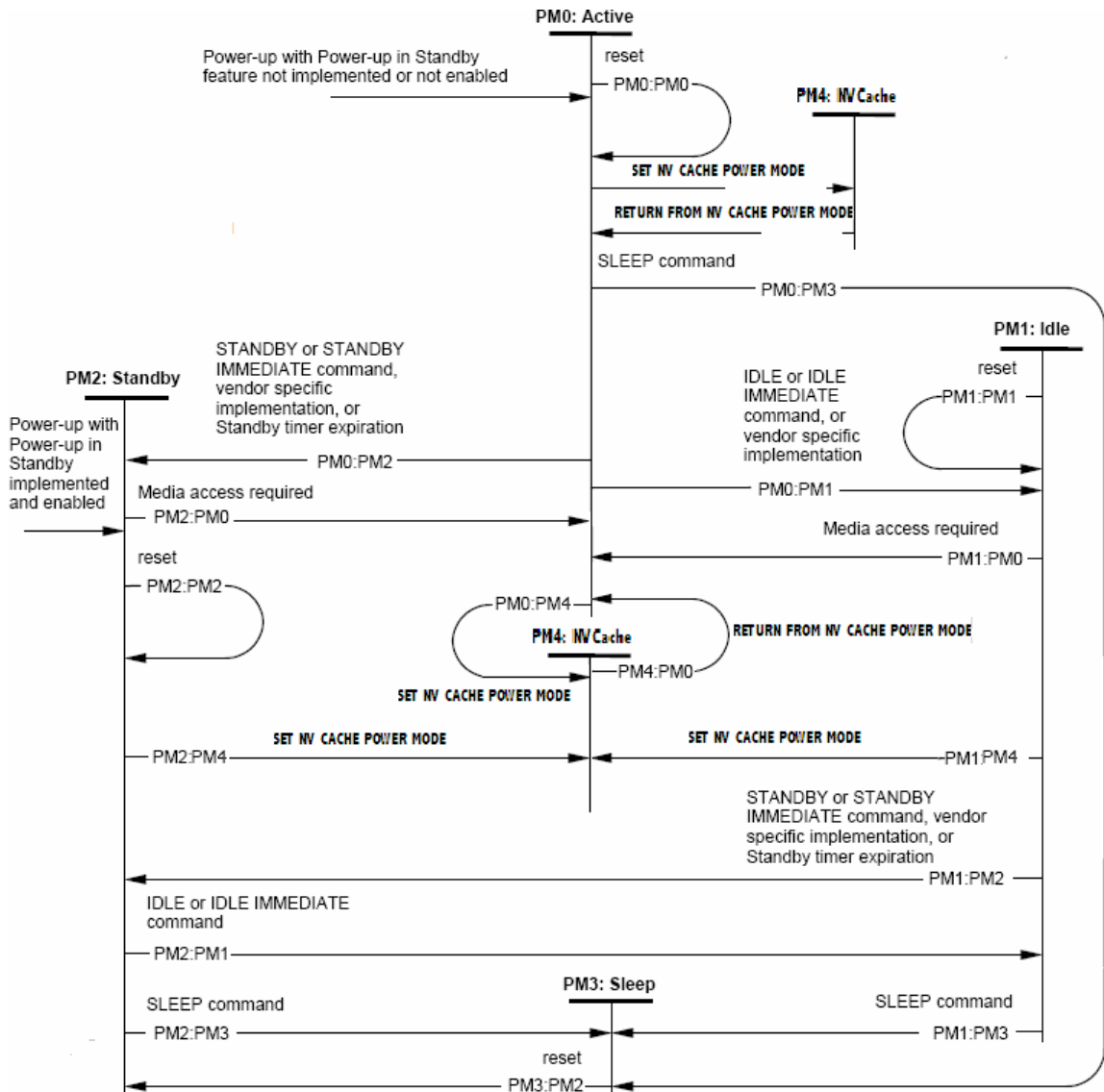


Figure 5: Changes to the Power Management State

NV Cache Power Mode

The NV Cache power mode, shown in Figure 5 as PM4, is entered when a device that supports the NV Cache power mode receives a SET NV CACHE MODE command.

In this mode, the device attempts to satisfy read and write commands from its NV Cache while keeping the spindle powered down.

Table 5: Hybrid Hard Disk Transition States

Transition States	Definition
Transition PM0:PM4	When a SET NV CACHE MODE command is received, the device will make the transition to PM4:NV Cache mode.
Transition PM1:PM4	When a SET NV CACHE MODE command is received, the device will make the transition to PM4:NV Cache mode.
Transition PM2:PM4	When a SET NV CACHE MODE command is received, the device will make the transition to PM4:NV Cache mode.
Transition PM4:PM0	When a RETURN FROM NV CACHE MODE command is received, the device will make the transition to PM0:Active mode. This transition is also implied if the host issues a SET FEATURES command that disables the NV Cache power mode support in the device while the device is in the PM4:NV Cache mode.

Note The ATA7 command extensions are detailed in a specification released under NDA. Contact your Microsoft representative for more information about the specification and architecture documentation.

Windows Logo Program Issues

Windows Logo Program requirements for this technology are still under development. Future versions of this white paper will include all applicable Logo Program requirements and recommendations.

Call to Action

Call to Action:

Support for hybrid hard disks as described in this paper is planned for Windows Longhorn. Microsoft expects significant market demand for this premium HDD offering, and encourages PC and HDD manufacturers to build and support hybrid hard disk functionality.

Additional information and a technical session on hybrid hard disks are currently planned for WinHEC 2005. For more information, see

<http://www.microsoft.com/whdc/winhec/default.aspx>

- For PC manufacturers – Review the information presented in this paper, provide feedback, and determine how to incorporate support for hybrid hard disks in your systems. Include these systems as part of your product offerings for the Windows Longhorn time frame.
- For HDD manufacturers – Begin planning to build hybrid hard disks that take advantage of the support planned for Windows Longhorn. Contact your Microsoft representative or the Feedback program listed below to receive the most up-to-date information about hybrid hard disk support planned for Windows Longhorn.

Feedback:

- To provide feedback or ask questions about the hybrid hard disk technology, please send e-mail to hybriddisk@microsoft.com