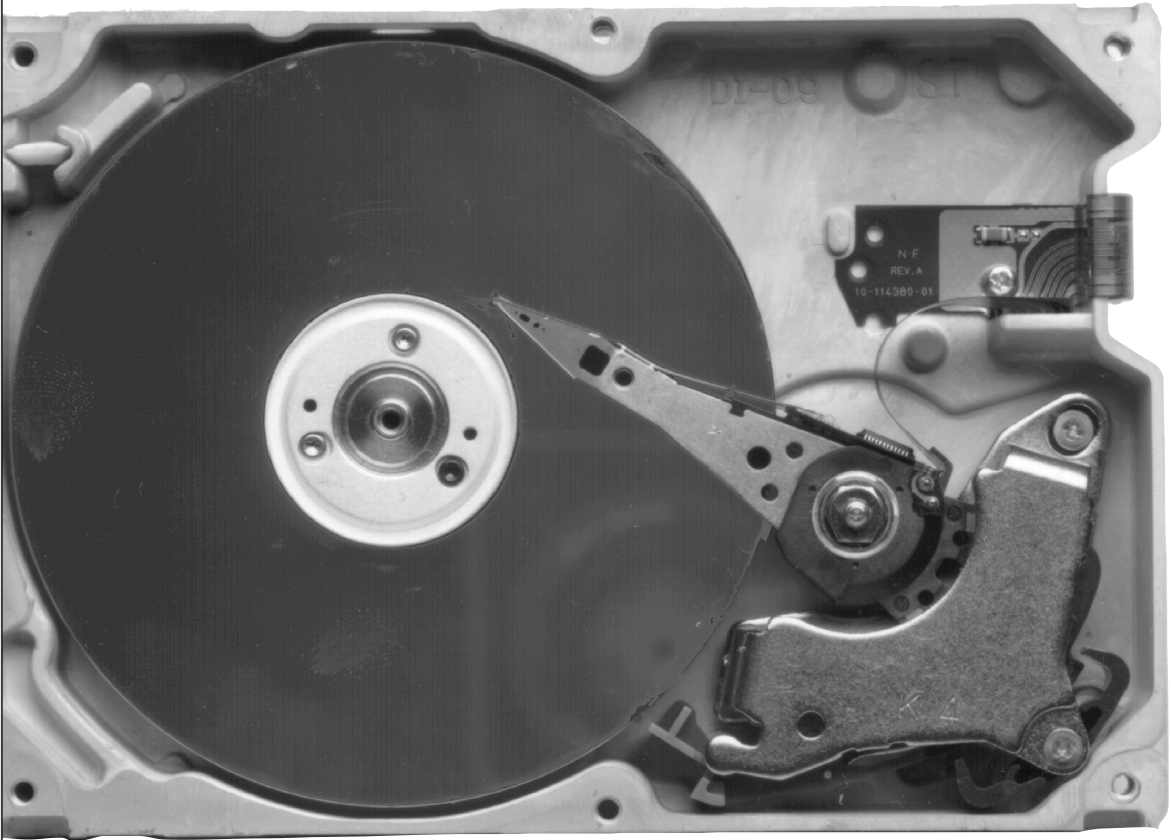


# CHAPTER 10

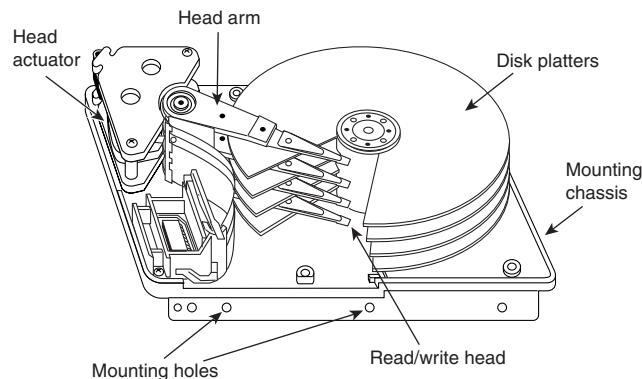
## Hard Disk Storage



## Definition of a Hard Disk

To many users, the hard disk drive is the most important and yet the most mysterious part of a computer system. A *hard disk drive* is a sealed unit that a PC uses for nonvolatile data storage. *Nonvolatile*, or semi-permanent, storage means that the storage device retains the data even when no power is supplied to the computer. Because the hard disk drive is expected to retain data until deliberately erased or overwritten, the hard drive is used to store crucial programming and data. As a result, when the hard disk fails, the consequences are usually very serious. To maintain, service, and upgrade a PC system properly, you must understand how the hard disk functions.

A hard disk drive contains rigid, disk-shaped platters, usually constructed of aluminum or glass (see Figure 10.1). Unlike floppy disks, the platters cannot bend or flex—hence the term *hard disk*. In most hard disk drives, you cannot remove the platters, which is why they are sometimes called *fixed* disk drives. Removable hard disk drives are also available. Sometimes this term refers to a device in which the entire drive unit (that is, the disk and the drive) is removable, but it is more commonly used to refer to cartridge drives, where the platters are contained in a removable cartridge.



**Figure 10.1** Hard disk heads and platters.

### Note

Hard disk drives are sometimes referred to as *Winchester drives*. This term dates back to 1973, when IBM introduced the model 3340 drive, which had 30MB of fixed platter and 30MB of removable platter storage on separate spindles. The drive was codenamed Winchester by project leader Ken Haughton, because the original capacity designation (30-30) sounded like the popular .30-30 (caliber-grains of charge) cartridge used by the Winchester 94 rifle introduced in 1895. The original 3340 “Winchester” drive was the first to use a sealed head/disk assembly, and the name has since been applied to all subsequent drives with similar technology.

## Hard Drive Advancements

In the almost 20 years that hard disks have commonly been used in PC systems, they have undergone tremendous changes. To give you an idea of how far hard drives have come in that time, I’ve outlined some of the more profound changes in PC hard disk storage:

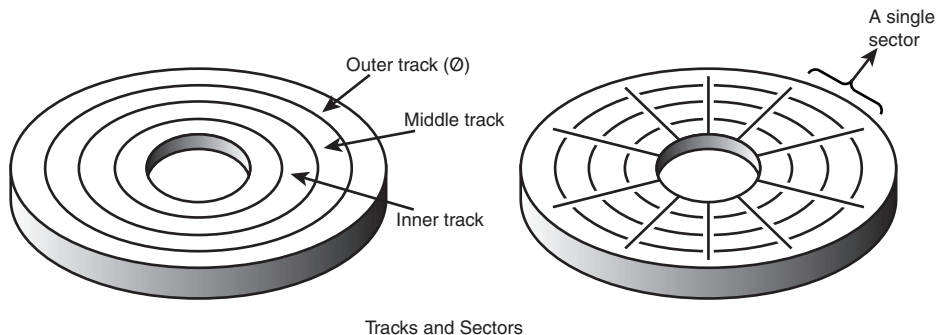
- Maximum storage capacities have increased from the 5MB and 10MB 5 1/4-inch full-height drives available in 1982 to 180GB or more for even smaller 3 1/2-inch half-height drives (Seagate Barracuda 180), and 32GB or more for notebook system 2 1/2-inch drives (IBM

Travelstar 32GH) that are 12.5mm (or less) in height. Hard drives smaller than 10GB are rare in today's desktop personal computers.

- Data transfer rates from the media (sustained transfer rates) have increased from 85KB to 102KB/sec for the original IBM XT in 1983 to an average of 51.15MB/sec or more for the fastest drives today (Seagate Cheetah 73LP).
- Average seek times (how long it takes to move the heads to a particular cylinder) have decreased from more than 85ms (milliseconds) for the 10MB XT hard disk in 1983 to 4.2ms or less for some of the fastest drives today (Seagate Cheetah X15).
- In 1982, a 10MB drive cost more than \$1,500 (\$150 per megabyte). Today, the cost of hard drives has dropped to one-half cent per megabyte or less!

## Hard Disk Drive Operation

The basic physical construction of a hard disk drive consists of spinning disks with heads that move over the disks and store data in tracks and sectors. The heads read and write data in concentric rings called *tracks*, which are divided into segments called *sectors*, which normally store 512 bytes each (see Figure 10.2).



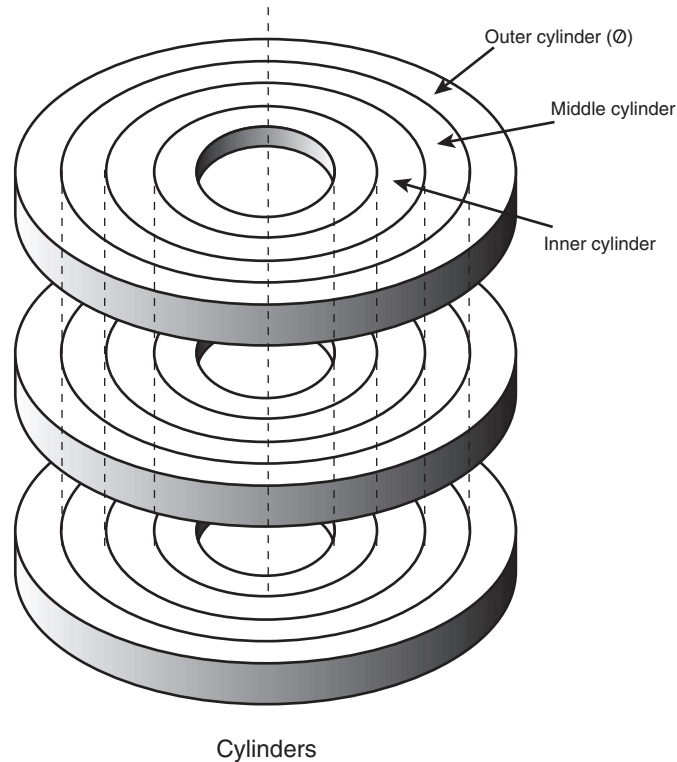
**Figure 10.2** The tracks and sectors on a disk.

Hard disk drives usually have multiple disks, called *platters*, that are stacked on top of each other and spin in unison, each with two sides on which the drive stores data. Most drives have two or three platters, resulting in four or six sides, but some PC hard disks have up to 12 platters and 24 sides with 24 heads to read them (Seagate Barracuda 180). The identically aligned tracks on each side of every platter together make up a cylinder (see Figure 10.3). A hard disk drive normally has one head per platter side, with all the heads mounted on a common carrier device or rack. The heads move radially across the disk in unison; they cannot move independently because they are mounted on the same carrier or rack, called an *actuator*.

Originally, most hard disks spun at 3,600rpm—approximately 10 times faster than a floppy disk drive. For many years, 3,600rpm was pretty much a constant among hard drives. Now, however, most drives spin the disks even faster. While speeds can vary, most modern drives spin the platters at 4,200; 5,400; 7,200; 10,000; or 15,000rpm. High rotational speeds combined with a fast head-positioning mechanism and more sectors per track are what make one hard disk faster than another.

The heads in most hard disk drives do not (and should not!) touch the platters during normal operation. When the heads are powered off, however, in most drives they land on the platters as they stop spinning. While the drive is running, a very thin cushion of air keeps each head suspended a short

distance above or below the platter. If the air cushion is disturbed by a particle of dust or a shock, the head can come into contact with the platter while it is spinning at full speed. When contact with the spinning platters is forceful enough to do damage, the event is called a *head crash*. The result of a head crash can be anything from a few lost bytes of data to a completely ruined drive. Most drives have special lubricants on the platters and hardened surfaces that can withstand the daily “takeoffs and landings” as well as more severe abuse.



**Figure 10.3** Hard disk cylinders.

Because the platter assemblies are sealed and nonremovable, the track densities on the disk can be very high. Hard drives today have up to 38,000 or more TPI (tracks per inch) recorded on the media (IBM Travelstar 30GT). Head Disk Assemblies (HDAs), which contain the platters, are assembled and sealed in clean rooms under absolutely sanitary conditions. Because few companies repair HDAs, repair or replacement of the parts inside a sealed HDA can be expensive. Every hard disk ever made eventually fails. The only questions are when the failure will occur and whether your data is backed up.

### **Caution**

It is strongly recommended that you do not even attempt to open a hard disk drive's HDA unless you have the equipment and the expertise to make repairs inside. Most manufacturers deliberately make the HDA difficult to open, to discourage the intrepid do-it-yourselfer. Opening the HDA almost certainly voids the drive's warranty.

Many PC users think that hard disks are fragile, and comparatively speaking, they are one of the more fragile components in your PC. In many of my PC Hardware and Troubleshooting or Data Recovery seminars, however, I have run various hard disks with the covers removed, and have even removed and installed the covers while the drives were operating! Those drives continue to store data perfectly to this day with their lids either on or off. Of course, I do not recommend that you try this with your own drives.

## The Ultimate Hard Disk Drive Analogy

There is an old analogy that compares the interaction of the heads and the medium in a typical hard disk drive as being similar in scale to a 747 flying a few feet off the ground at cruising speed (500+ mph). I have heard this analogy used over and over again for years, and I've even used it in my seminars many times without checking to see whether the analogy is technically accurate with respect to modern hard drives. It isn't.

One highly inaccurate aspect of the 747 analogy has always bothered me—the use of an airplane of any type to describe the head-and-platter interaction. This analogy implies that the heads fly very low over the surface of the disk, but technically, this is not true. The heads do not fly at all in the traditional aerodynamic sense; instead, they float on a cushion of air that is being dragged around by the platters.

A much better analogy would use a hovercraft instead of an airplane; the action of a hovercraft much more closely emulates the action of the heads in a hard disk drive. Like a hovercraft, the drive heads rely somewhat on the shape of the bottom of the head to capture and control the cushion of air that keeps them floating over the disk. By nature, the cushion of air on which the heads float forms only in very close proximity to the platter, and is often called an *air bearing* by the disk drive industry.

I thought it was time to come up with a new analogy that more correctly describes the dimensions and speeds at which a hard disk drive operates today. I looked up the specifications on a specific hard disk drive, and then magnified and rescaled all the dimensions involved to make the head floating height equal to one inch. For my example, I used an IBM Deskstar 75GXP drive, which is a 75GB (formatted capacity), 3 1/2-inch ATA (AT Attachment interface) drive. The head sliders (called pico sliders) in this drive are about 0.049 inches long, 0.039 inches wide, and 0.012 inches high. They float on a cushion of air about 15 nanometers (nm or billionths of a meter) over the surface of the disk while traveling at an average speed of 53.55 miles per hour (figuring an average track diameter of about 2.5 inches). These heads read and write individual bits spaced only 2.56 micro-inches (millionths of an inch) apart, along tracks separated by only 35.27 micro-inches. The heads can move from one track to another in 8.5 milliseconds during an average seek.

To create my analogy, I magnified the scale to make the head floating height equal to 5 millimeters (about 0.2 inches). Because 5 millimeters is about 333,333 times greater than 15 nanometers (nm), I scaled up everything else by the same amount.

Magnified to such a scale, the heads in this typical hard disk would be about 1,361 feet long, 1,083 feet wide, and 333 feet high (the length and height are about equal to the Sears Tower if it were toppled over sideways). These skyscraper-sized heads would float on a cushion of air which to scale would be only 5mm thick (about 0.2 inches) while travelling at a speed of 17.8 million miles per hour (4,958 miles per second), all while reading data bits spaced a mere 0.85 inches apart on tracks separated by only 0.98 feet!

The forward scale speed of this imaginary head is difficult to comprehend, so I'll elaborate. The diameter of the Earth at the equator is 7,926 miles, which means a circumference of about 24,900 miles. At 4,958 miles per second, this imaginary skyscraper-sized head would circle the earth once every five seconds (at only two-tenths of an inch over the surface)! It would also read 231.33MB in one lap around this equatorial track.

There is also sideways velocity to consider. Because the average seek time of 8.5 milliseconds is defined as the time it takes to move the heads over one-third of the total tracks (about 9,241 tracks in this case), the heads could move sideways within a scale distance of 1.71 miles in that short time. This results in a scale seek velocity of more than 726,321mph, or 202 miles per second!

This analogy should give you a new appreciation of the technological marvel that the modern hard disk drive actually represents. It makes the old 747 analogy look rather pathetic (not to mention inaccurate), doesn't it?

## Tracks and Sectors

A *track* is a single ring of data on one side of a disk. A disk track is too large to manage data effectively as a single storage unit. Many disk tracks can store 100,000 or more bytes of data, which would be very inefficient for storing small files. For that reason, tracks are divided into several numbered divisions known as sectors. These sectors represent arc-shaped pieces of the track.

Various types of disk drives split their disk tracks into different numbers of sectors, depending on the density of the tracks. For example, floppy disk formats use 8–36 sectors per track, although hard disks usually store data at a higher density and today can have 900 or more sectors per track physically. The sectors created by the standard formatting procedure on a PC system have a capacity of 512 bytes, which has been one constant throughout the history of the PC. One interesting phenomenon of the PC standard is that to be compatible with most older BIOS and drivers, drives will usually perform an internal translation to a logical 63 sectors per track.

The sectors on a track are numbered starting with 1, unlike the heads or cylinders that are numbered starting with 0. For example, a 1.44MB floppy disk contains 80 cylinders numbered 0–79 and two heads numbered 0 and 1, whereas each track on each cylinder has 18 sectors numbered 1–18.

When a disk is formatted, the formatting program creates ID areas before and after each sector's data that the disk controller uses for sector numbering and for identifying the start and end of each sector. These areas precede and follow each sector's data area and consume some of the disk's total storage capacity. This accounts for the difference between a disk's unformatted and formatted capacities. Note that most modern hard drives are sold preformatted and only advertise the formatted capacity. The unformatted capacity is usually not mentioned anymore. Another interesting development is that IBM and others now make drives with ID-less recording, which means that the sectors are recorded without ID marks before and after each sector. This means that more of the disk can be used for actual data.

Each sector on a disk normally has a prefix portion, or header, that identifies the start of the sector and contains the sector number, as well as a suffix portion, or trailer, that contains a checksum (which helps ensure the integrity of the data contents). Many newer drives omit this header and have what is called a No-ID recording, allowing more space for data.

Each sector contains 512 bytes of data. The low-level formatting process normally fills the data bytes with some specific value, such as F6h (hex), or some other repeating test pattern used by the drive manufacturer. Some patterns are more difficult for the drive to encode/decode, so these patterns normally are used when the manufacturer is testing the drive during initial formatting. A special test pattern might cause errors to surface that a normal data pattern would not show. This way, the manufacturer can more accurately identify marginal sectors during testing.

---

### Note

The type of disk formatting discussed here is a physical or low-level format, not the high-level format you perform when you use the Windows 9x/Me/2000 Explorer or the DOS FORMAT program on a disk. See the section "Disk Formatting" later in this chapter to learn about the difference between these two types of formatting.

---

The sector headers and trailers are independent of the operating system, the file system, and the files stored on the drive. In addition to the headers and trailers, gaps exist within the sectors, between the sectors on each track, and also between tracks, but none of these gaps contain usable data space. The gaps are created during the low-level format process when the recording is turned off momentarily. They serve the same function as having gaps of no sound between the songs recorded on a cassette tape. The prefix, suffix, and gaps account for the lost space between the unformatted capacity of a disk and the formatted capacity. For example, a 4MB (unformatted) floppy disk (3 1/2-inch) has a capacity of 2.88MB when it is formatted, a 2MB (unformatted) floppy has a formatted capacity of 1.44MB, and an older 38MB unformatted capacity (for instance, Seagate ST-4038) hard disk has a capacity of only 32MB when it is formatted. Because the ATA/IDE and SCSI hard drives you purchase today are low-level formatted at the factory, the manufacturers now advertise only the formatted capacity. Even so, nearly all drives use some reserved space for managing the data that will be stored on the drive.

Thus, while I stated earlier that each disk sector is 512 bytes in size, this statement is technically not true. Each sector does allow for the storage of 512 bytes of data, but the data area is only a portion of the sector. Each sector on a disk typically occupies 571 bytes of the disk, of which only 512 bytes are available for the storage of user data. The actual number of bytes required for the sector header and trailer can vary from drive to drive, but this figure is typical. As mentioned earlier, though, many modern drives now use a No-ID recording scheme that virtually eliminates the storage overhead of the sector header information. You might find it helpful to think of each disk sector as being a page in a book. In a book, each page contains text, but the entire page is not filled with text; rather, each page has top, bottom, left, and right margins. Information such as chapter titles (track and cylinder numbers) and page numbers (sector numbers) is placed in the margins. The “margin” areas of a sector are created and written to during the low-level formatting process. Formatting also fills the data area of each sector with dummy values. After you perform a high-level format on the disk, the PC’s file system can write to the data area of each sector, but the sector header and trailer information cannot be altered during normal write operations unless the disk is low-level formatted again.

Table 10.1 shows the format for each track and sector on a typical hard disk drive with 17 sectors per track.

**Table 10.1 Typical Disk Track/Sector Format Using ID Marks**

Bytes	Name	Description
16	POST INDEX GAP	All 4Eh, at the track beginning after the Index mark.
The following sector data (shown between the lines in this table) is repeated as many times as there are sectors on the track.		
13	ID VFO LOCK	All 00h; synchronizes the VFO for the sector ID.
1	SYNC BYTE	A1h; notifies the controller that data follows.
1	ADDRESS MARK	FEh; defines that ID field data follows.
2	CYLINDER NUMBER	A value that defines the head actuator position.
1	HEAD NUMBER	A value that defines the particular head selected.
1	SECTOR NUMBER	A value that defines the sector.
2	CRC	Cyclic Redundancy Check to verify ID data.
3	WRITE TURN-ON GAP	00h written by format to isolate the ID from DATA.
13	DATA SYNC VFO LOCK	All 00h; synchronizes the VFO for the DATA.
1	SYNC BYTE	A1h; notifies the controller that data follows.
1	ADDRESS MARK	F8h; defines that user DATA field follows.

**Table 10.1 Continued**

Bytes	Name	Description
512	DATA	The area for user DATA.
2	CRC	Cyclic Redundancy Check to verify DATA.
3	WRITE TURN-OFF GAP	00h; written by DATA update to isolate DATA.
15	INTER-RECORD GAP	All 00h; a buffer for spindle speed variation.
693	PRE-INDEX GAP	All 4Eh, at track end before Index mark.

512 = Total bytes per sector; 512 = Data (usable) bytes per sector

Note: "All XXh" indicates that field will be filled with XXh bytes.

As you can see, the usable space for data on each track is about 15% less than its total unformatted capacity. This is true for most disks, although the percentage can vary slightly, depending on how many sectors exist per track. The following paragraphs detail each piece of the sector data listed in Table 10.1.

The POST INDEX GAP provides a head-switching recovery period, so when switching from one track to another, the heads can read sequential sectors without waiting for an additional revolution of the disk. Because the disk is continuously spinning and the heads take some small amount of time to move radially from track to track, reading consecutive sectors on two different tracks, one right after the other, is not possible. By the time the head moves to the new track, the beginning of the second sector has already spun past it. Leaving a gap between sectors provides the heads with time to move to another track.

In some drives, this gap does not provide sufficient time for the heads to move. When this is the case, a drive can gain additional time by skewing the sectors on different tracks so the arrival of the first sector is delayed. In other words, the low-level formatting process offsets the sector numbering, so instead of the same numbered sectors on each track being adjacent to each other, Sector 9 on one track might be next to Sector 8 of the next track, which is next to Sector 7 on the next, and so forth. The optimum skew value is based on the rotational speed of the disk as compared to the lateral speed of the heads.

### Note

At one time, the head skew was a parameter you could set yourself while low-level formatting a drive. Today's ATA/IDE and SCSI drives are low-level formatted at the factory with the optimum skew values.

The Sector ID data consists of the Cylinder, Head, and Sector Number fields, as well as a CRC field used to verify the ID data. Most controllers use bit 7 of the Head Number field to mark a sector as bad during a low-level format or surface analysis. This convention is not absolute, however. Some controllers use other methods to mark a bad sector, but usually the mark involves one of the ID fields.

The WRITE TURN-ON GAP follows the ID field's CRC bytes and provides a pad to ensure a proper recording of the user data area that follows, as well as to enable full recovery of the ID CRC.

The user DATA field consists of all 512 bytes of data stored in the sector. This field is followed by a CRC field to verify the data. Although many controllers use two bytes of CRC here, the controller might implement a longer Error Correction Code (ECC) that requires more than two CRC bytes to store. The ECC data stored here provides the possibility of correcting errors in the DATA field as well



as detecting them. The correction/detection capabilities depend on the ECC code the drive uses and its implementation by the controller. The WRITE TURN-OFF GAP is a pad that enables the ECC (CRC) bytes to be fully recovered.

The INTER-RECORD GAP provides a means to accommodate variances in drive spindle speeds. A track might have been formatted while the disk was running slightly more slowly than normal and then written to while the disk was running slightly more quickly than normal. In such cases, this gap prevents the accidental overwriting of any information in the next sector. The actual size of this padding varies, depending on the speed of the DATA disk's rotation when the track was formatted and each time the DATA field is updated.

The PRE-INDEX GAP enables speed tolerance over the entire track. This gap varies in size, depending on the variances in disk rotation speed and write-frequency tolerance at the time of formatting.

This sector prefix information is extremely important because it contains the numbering information that defines the cylinder, head, and sector. So this information—except the DATA field, DATA CRC bytes, and WRITE TURN-OFF GAP—is written only during a low-level format.

## Disk Formatting

Two formatting procedures are required before you can write user data to a disk:

- Physical, or low-level formatting
- Logical, or high-level formatting

When you format a blank floppy disk, the Windows 9x/Me/2000 Explorer or the DOS `FORMAT` command performs both types of formats simultaneously. If the floppy was already formatted, DOS and Windows will default to doing only a high-level format.

A hard disk, however, requires two separate formatting operations. Moreover, a hard disk requires a third step, between the two formatting procedures, to write the partitioning information to the disk. Partitioning is required because a hard disk is designed to be used with more than one operating system. Using multiple operating systems on one hard drive is possible by separating the physical formatting in a procedure that is always the same, regardless of the operating system used and the high-level format (which is different for each operating system). Partitioning enables a single hard disk drive to run more than one type of operating system, or it can enable a single operating system to use the disk as several volumes or logical drives. A *volume* or *logical drive* is any section of the disk to which the operating system assigns a drive letter or name.

Consequently, preparing a hard disk drive for data storage involves three steps:

1. Low-level formatting (LLF)
2. Partitioning
3. High-level formatting (HLF)

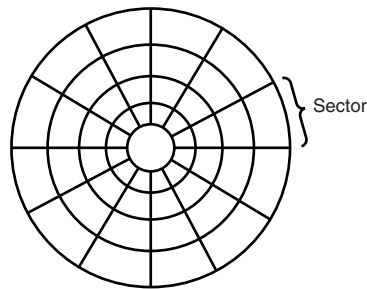
### Low-Level Formatting

During a low-level format, the formatting program divides the disk's tracks into a specific number of sectors, creating the intersector and intertrack gaps and recording the sector header and trailer information. The program also fills each sector's data area with a dummy byte value or a pattern of test values. For floppy disks, the number of sectors recorded on each track depends on the type of disk and drive. For hard disks, the number of sectors per track depends on the drive and the controller interface.

Originally, PC hard disk drives used a separate controller that took the form of an expansion card or was integrated into the motherboard. Because the controller could be used with various disk drives and might even have been made by a different manufacturer, some uniformity had to exist in the communications between the controller and the drive. For this reason, the number of sectors written to a track tended to be relatively consistent.

The original ST-506/412 MFM controllers always placed 17 sectors per track on a disk, although ST-506/412 controllers with RLL encoding increased the number of sectors to 25 or 26 per track; ESDI drives had 32 or more sectors per track. The ATA/IDE and SCSI drives found in PCs today can have anywhere from 17 to 900 or more sectors per track.

Virtually all ATA and SCSI drives use a technique called *zoned-bit recording*, which writes a variable number of sectors per track. Without zoned-bit recording, the number of sectors, and therefore bits, on each track is a constant. This means the actual number of bits per inch will vary. More bits per inch will exist on the inner tracks, and fewer will exist on the outer. The data rate and rotational speed will remain constant, as will the number of bits per track. Figure 10.4 shows a drive recorded with the same number of sectors per track.



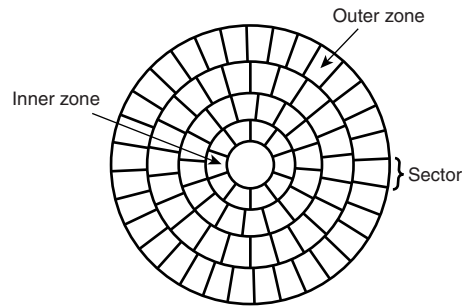
**Figure 10.4** Standard recording, where the same number of sectors comprise every track.

A standard recording wastes capacity on the outer tracks, because it is longer and yet holds the same amount of data (more loosely spaced) as the inner tracks. One way to increase the capacity of a hard drive during the low-level format is to create more sectors on the disks' outer cylinders than on the inner ones. Because they have a larger circumference, the outer cylinders can hold more data. Drives without zoned-bit recording store the same amount of data on every cylinder, even though the tracks of the outer cylinders might be twice as long as those of the inner cylinders. The result is wasted storage capacity because the disk medium must be capable of storing data reliably at the same density as on the inner cylinders. When the number of sectors per track is fixed, as in older controllers, the drive capacity is limited by the density of the innermost (shortest) track.

Drives that use zoned-bit recording split the cylinders into groups called *zones*, with each successive zone having more sectors per track as you move outward from the center of the disk. All the cylinders in a particular zone have the same number of sectors per track. The number of zones varies with specific drives, but most drives have 10 or more zones.

Figure 10.5 shows a drive with zoned-bit recording.

Another effect of zoned-bit recording is that transfer speeds vary depending on which zone the heads are in. A drive with zoned-bit recording still spins at a constant speed; because more sectors exist per track in the outer zones, however, data transfer is fastest there. Consequently, data transfer is slowest when reading or writing to the inner zones. That is why virtually all drives today report minimum and maximum sustained transfer rates, which depend on where on the drive you are reading from or writing to.



**Figure 10.5** Zoned-bit recording, where the number of sectors per track increases within each zone, moving out from the center.

As an example, see Table 10.2, which shows the zones defined for an IBM Travelstar 32GH 2 1/2-inch notebook drive, the sectors per track for each zone, and the resulting data transfer rate.

**Table 10.2 Zoned Bit Recording Information for the IBM Travelstar 32GH 32GB 2 1/2-inch Hard Disk Drive**

Zone	Sectors per Track	Data Transfer Rate (MB/sec)	Bytes per Track	Sectors per Zone
0	617	28.49	315,904	835,418
1	598	27.60	306,005	809,241
2	578	26.70	296,107	783,063
3	559	25.81	286,208	756,886
4	540	24.92	276,309	730,709
5	520	24.03	266,411	704,531
6	501	23.13	256,512	678,354
7	482	22.24	246,613	652,177
8	462	21.35	236,715	625,999
9	443	20.46	226,816	599,822
10	424	19.56	216,917	573,645
11	404	18.67	207,019	547,467
12	385	17.78	197,120	521,290
13	366	16.88	187,221	495,113
14	346	15.99	177,323	468,935
15	327	15.10	167,424	442,758

21,664 Total Tracks; 16 Zones; 1,354 Tracks per Zone

512 Bytes per Sector; 5,411 rpm; 10,225,408 Total Sectors per Side

This drive has a total of 21,664 tracks on each platter surface and, as you can see, the tracks are divided into 16 zones of 1,354 tracks each. It is not essential for all the zones to be the same size; this is simply how this drive is arranged. Zone 0 consists of the outermost 1,354 tracks, which are the longest and contain the most sectors: 617. Because each sector is 512 bytes, each track in this zone can therefore provide about 315,904 bytes of user data storage, although the 327 sector tracks in zone 15 can hold only 167,424 bytes.

Thus, with zoned-bit recording, each platter surface in this disk drive contains 10,225,408 sectors, for a storage capacity of 5,235MB per side. Without zoned-bit recording, the number of sectors per track would be limited to 327 over the entire surface of each platter, for a total of 7,084,128 sectors, storing 3,627MB. Zoned-bit recording, therefore, provides a 44% increase in the storage capacity of this particular drive.

Notice also the difference in the data transfer rates for each of the zones. The tracks in the outermost zone (0) yield a transfer rate of 28.49MB/sec, which is 89% higher than the 15.10MB/sec of the innermost zone (15). This is one reason you might notice huge discrepancies in the results produced by disk drive benchmark programs. A test that reads or writes files on the outer tracks of the disk naturally yields far better results than one conducted on the inner tracks. It might appear as though your drive is running more slowly, when the problem is actually that the test results you are comparing stem from disk activity on different zones.

Another thing to note is that this drive conforms to the ATA-5 specification and is capable of running in Ultra-ATA/66 mode (also called UDMA-66), which implies a transfer speed of 66MB/sec. As you can see, that is entirely theoretical because the true *media transfer speed* of this drive varies between about 15MB/sec and 28MB/sec, averaging about 21.8MB/sec overall. The interface transfer rate is just that: what the interface is capable of. It has little bearing on the actual capabilities of the drive.

Drives with separate controllers used in the past could not handle zoned-bit recording because no standard way existed to communicate information about the zones from the drive to the controller.

With SCSI and ATA disks, however, formatting individual tracks with different numbers of sectors became possible because these drives have the disk controller built in. The built-in controllers on these drives are fully aware of the zoning algorithm and can translate the physical Cylinder, Head, and Sector numbers to logical Cylinder, Head, and Sector numbers so the drive appears to have the same number of sectors on each track. Because the PC BIOS is designed to handle only a single number of sectors per track throughout the entire drive, a zoned drive must run by using a sector translation scheme.

The use of zoned-bit recording enables drive manufacturers to increase the capacity of their hard drives by 20%–50% compared with a fixed-sector-per-track arrangement. All modern ATA (IDE) and SCSI drives today use zoned-bit recording.

## Partitioning

Creating a partition on a hard disk drive enables it to support separate file systems, each in its own partition.

Each file system can then use its own method to allocate file space in logical units called *clusters* or *allocation units*. Every hard disk drive must have at least one partition on it and can have up to four partitions, each of which can support the same or different type file systems. Three common file systems are used by PC operating systems today:

- **FAT (File Allocation Table).** The standard file system supported by DOS, Windows 9x/Me/2000, and Windows NT. FAT partitions support filenames of 11 characters maximum (8 characters+3-character extension) under DOS, and 255 characters under Windows 9x/Me/2000 or Windows NT 4.0 (or later). The standard FAT file system uses 12- or 16-bit numbers to identify clusters, resulting in a maximum volume size of 2GB.

Using FDISK, you can create only two physical FAT partitions on a hard disk drive—primary and extended—but you can subdivide the extended partition into as many as 25 logical volumes. Alternative partitioning programs, such as Partition Magic, can create up to four primary partitions or three primary and one extended.

- **FAT32 (File Allocation Table, 32-bit).** An optional file system supported by Windows 95 OSR2 (OEM Service Release 2), Windows 98, Windows Me, and Windows 2000.

FAT32 uses 32-bit numbers to identify clusters, resulting in a maximum single volume size of 2TB or 2,048GB.

- **NTFS (Windows NT File System).** The native file system for Windows NT/2000 that supports file-names up to 256 characters long and partitions up to (a theoretical) 16 exabytes. NTFS also provides extended attributes and file system security features that do not exist in the FAT file system. Of these three file systems, the FAT file system still is by far the most popular and is accessible by nearly every operating system, which makes it the most compatible as well. FAT32 and NTFS provide additional features but are not universally accessible by other operating systems.

►► See "FAT Disk Structures," p. 1316 and "FAT32," p. 1334.

Partitioning normally is accomplished by running the FDISK program that comes with your operating system. FDISK enables you to select the amount of space on the drive to use for a partition, from a single megabyte or 1% of the drive up to the entire capacity of the drive, or as much as the particular file system will allow. Normally, it is recommended to have as few partitions as possible, and many people (myself included) try to stick with only one or two at the most. This was more difficult before FAT32 because the maximum partition size for a FAT16 partition was only 2GB. With FAT32, the maximum partition size can be up to 2,048GB.

### Caution

FDISK cannot be used to change the size of a partition; all it can do is remove or create partitions. The act of removing or creating a partition destroys any data that was contained in the partition or was on that part of the disk. To manipulate partitions without destroying data, you can use third-party utility programs, such as Partition Magic from PowerQuest or Partition Commander from V Communications.

After a drive is partitioned, each partition must then be high-level formatted by the operating system that will use it.

### High-Level Formatting

During the high-level format, the operating system (such as Windows 9x/Me/2000, Windows NT, or DOS) writes the structures necessary for managing files and data on the disk. FAT partitions have a Volume Boot Sector (VBS), two copies of a file allocation table (FAT), and a root directory on each formatted logical drive. These data structures enable the operating system to manage the space on the disk, keep track of files, and even manage defective areas so they do not cause problems.

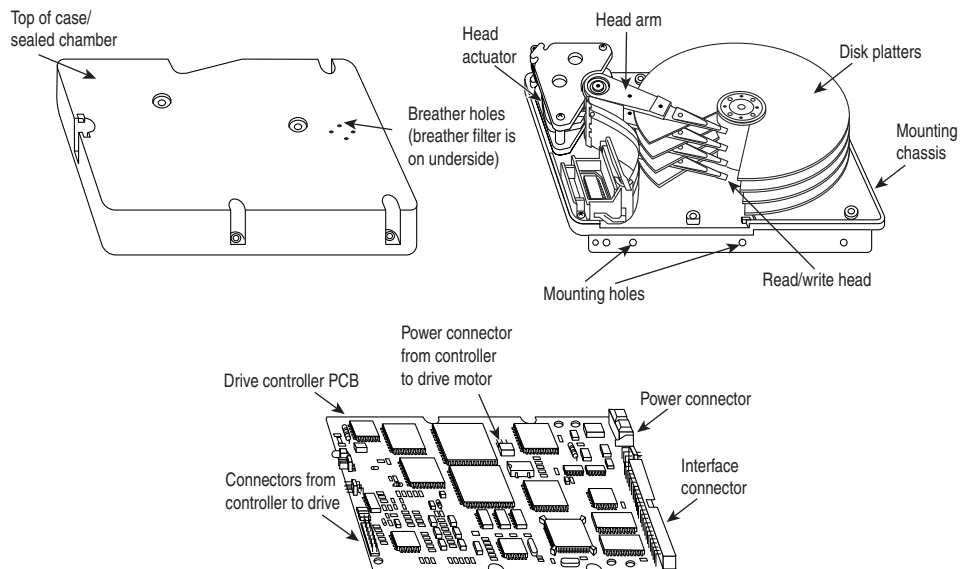
High-level formatting is not really a physical formatting of the drive, but rather the creation of a table of contents for the disk. In low-level formatting, which is the real physical formatting process, tracks and sectors are written on the disk. As mentioned, the DOS `FORMAT` command can perform both low-level and high-level format operations on a floppy disk, but it performs only the high-level format for a hard disk. Low-level formats of ATA and SCSI hard disk drives are performed by the manufacturer and should almost never be performed by the end user. The only time I low-level format ATA or SCSI drives is when I am attempting to repair a format that has become damaged (parts of the disk become unreadable), or in some cases when I want to wipe away all data on the drive.

## Basic Hard Disk Drive Components

Many types of hard disk drives are on the market, but nearly all share the same basic physical components. Some differences might exist in the implementation of these components (and in the quality

of the materials used to make them), but the operational characteristics of most drives are similar. The basic components of a typical hard disk drive are as follows (see Figure 10.6):

- Disk platters
- Read/write heads
- Head actuator mechanism
- Spindle motor (inside platter hub)
- Logic board (controller or Printed Circuit Board)
- Cables and connectors
- Configuration items (such as jumpers or switches)



**Figure 10.6** Typical hard disk drive components.

The platters, spindle motor, heads, and head actuator mechanisms usually are contained in a sealed chamber called the Head Disk Assembly (HDA). The HDA is usually treated as a single component; it is rarely opened. Other parts external to the drive's HDA, such as the logic boards, bezel, and other configuration or mounting hardware, can be disassembled from the drive.

## Hard Disk Platters (Disks)

A hard disk drive has one or more platters, or disks. Hard disks for PC systems have been available in a number of form factors over the years. Normally, the physical size of a drive is expressed as the size of the platters. Following are the platter sizes that have been associated with PC hard disk drives:

- 5 1/4-inch (actually 130mm, or 5.12 inches)
- 3 1/2-inch (actually 95mm, or 3.74 inches)
- 2 1/2-inch (actually 65mm, or 2.56 inches)
- 1-inch (actually 34mm, or 1.33 inches)

Larger hard disk drives that have 8-inch, 14-inch, or even larger platters are available, but these drives are not used with PC systems. Currently, the 3 1/2-inch drives are the most popular for desktop and some portable systems, whereas the 2 1/2-inch and smaller drives are very popular in portable or notebook systems.

During 1998, IBM introduced a drive called the MicroDrive, which currently can store up to 1GB or more on a single platter about the size of a quarter! This type of drive can be used in information appliances, digital cameras, and anywhere else flash memory cards have been used.

Most hard disk drives have two or more platters, although some of the smaller drives used in portable systems have only one. The number of platters a drive can have is limited by the drive's vertical physical size. The maximum number of platters I have seen in any 3 1/2-inch drive is 12; however, most drives have 6 or fewer.

Platters have traditionally been made from an aluminum/magnesium alloy, which provides both strength and light weight. However, manufacturers' desire for higher and higher densities and smaller drives has led to the use of platters made of glass (or, more technically, a glass-ceramic composite). One such material, produced by the Dow Corning Corporation, is called MemCor. MemCor is composed of glass with ceramic implants, enabling it to resist cracking better than pure glass. Glass platters offer greater rigidity than metal (because metal can be bent and glass cannot) and can therefore be machined to one-half the thickness of conventional aluminum disks—sometimes less. Glass platters are also much more thermally stable than aluminum platters, which means they do not expand or contract very much with changes in temperature. Several hard disk drives made by companies such as IBM, Seagate, Toshiba, Areal Technology, and Maxtor currently use glass or glass-ceramic platters. For most manufacturers, glass disks will probably replace the standard aluminum/magnesium substrate over the next few years, especially in high-performance 2 1/2- and 3 1/2-inch drives.

## Recording Media

No matter which substrate is used, the platters are covered with a thin layer of a magnetically retentive substance, called the *medium*, on which magnetic information is stored. Two popular types of magnetic media are used on hard disk platters:

- Oxide medium
- Thin-film medium

The oxide medium is made of various compounds, containing iron oxide as the active ingredient. The magnetic layer is created on the disk by coating the aluminum platter with a syrup containing iron-oxide particles. This syrup is spread across the disk by spinning the platters at high speed; centrifugal force causes the material to flow from the center of the platter to the outside, creating an even coating of the material on the platter. The surface is then cured and polished. Finally, a layer of material that protects and lubricates the surface is added and burnished smooth. The oxide coating is normally about 30 millionths of an inch thick. If you could peer into a drive with oxide-coated platters, you would see that the platters are brownish or amber.

As drive density increases, the magnetic medium needs to be thinner and more perfectly formed. The capabilities of oxide coatings have been exceeded by most higher-capacity drives. Because the oxide medium is very soft, disks that use it are subject to head-crash damage if the drive is jolted during operation. Most older drives, especially those sold as low-end models, use oxide media on the drive platters. Oxide media, which have been used since 1955, remained popular because of their relatively low cost and ease of application. Today, however, very few drives use oxide media.

The thin-film medium is thinner, harder, and more perfectly formed than oxide medium. Thin film was developed as a high-performance medium that enabled a new generation of drives to have lower

head-floating heights, which in turn made increases in drive density possible. Originally, thin-film media were used only in higher-capacity or higher-quality drive systems, but today, virtually all drives use thin-film media.

The thin-film medium is aptly named. The coating is much thinner than can be achieved by the oxide-coating method. Thin-film media are also known as plated, or sputtered, media because of the various processes used to deposit the thin film on the platters.

Thin-film plated media are manufactured by depositing the magnetic medium on the disk with an electroplating mechanism, in much the same way that chrome plating is deposited on the bumper of a car. The aluminum/magnesium or glass platter is immersed in a series of chemical baths that coat the platter with several layers of metallic film. The magnetic medium layer itself is a cobalt alloy about 1  $\mu$ -inch thick.

Thin-film sputtered media are created by first coating the aluminum platters with a layer of nickel phosphorus and then applying the cobalt-alloy magnetic material in a continuous vacuum-deposition process called *sputtering*. This process deposits magnetic layers as thin as 1  $\mu$ -inch or less on the disk, in a fashion similar to the way that silicon wafers are coated with metallic films in the semiconductor industry. The same sputtering technique is then used again to lay down an extremely hard, 1  $\mu$ -inch protective carbon coating. The need for a near-perfect vacuum makes sputtering the most expensive of the processes described here.

The surface of a sputtered platter contains magnetic layers as thin as 1  $\mu$ -inch. Because this surface also is very smooth, the head can float more closely to the disk surface than was possible previously. Floating heights as small as 15nm (nanometers, or about 0.6  $\mu$ -inch) above the surface are possible. When the head is closer to the platter, the density of the magnetic flux transitions can be increased to provide greater storage capacity. Additionally, the increased intensity of the magnetic field during a closer-proximity read provides the higher signal amplitudes needed for good signal-to-noise performance.

Both the sputtering and plating processes result in a very thin, hard film of magnetic medium on the platters. Because the thin-film medium is so hard, it has a better chance of surviving contact with the heads at high speed. In fact, modern thin-film media are virtually uncrashable. If you could open a drive to peek at the platters, you would see that platters coated with the thin-film medium look like mirrors.

## Read/Write Heads

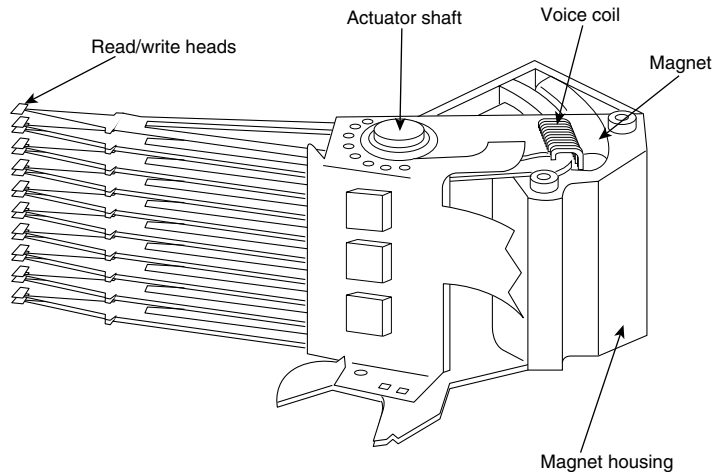
A hard disk drive usually has one read/write head for each platter surface (meaning that each platter has two sets of read/write heads—one for the top side and one for the bottom side). These heads are connected, or ganged, on a single movement mechanism. The heads, therefore, move across the platters in unison.

Mechanically, read/write heads are simple. Each head is on an actuator arm that is spring-loaded to force the head into contact with a platter. Few people realize that each platter actually is “squeezed” by the heads above and below it. If you could open a drive safely and lift the top head with your finger, the head would snap back down into the platter when you released it. If you could pull down on one of the heads below a platter, the spring tension would cause it to snap back up into the platter when you released it.

Figure 10.7 shows a typical hard disk head-actuator assembly from a voice coil drive.

When the drive is at rest, the heads are forced into direct contact with the platters by spring tension, but when the drive is spinning at full speed, air pressure develops below the heads and lifts them off the surface of the platter. On a drive spinning at full speed, the distance between the heads and the platter can be anywhere from 0.5 to 5  $\mu$ -inch or more in a modern drive.





**Figure 10.7** Read/write heads and rotary voice coil actuator assembly.

In the early 1960s, hard disk drive recording heads operated at floating heights as large as 200–300  $\mu$ -inch; today's drive heads are designed to float as low as 10nm (nanometers) or 0.4  $\mu$ -inch above the surface of the disk. To support higher densities in future drives, the physical separation between the head and disk is expected to drop even further, such that on some drives there will even be contact with the platter surface. New media and head designs will make full or partial contact recording possible.

### Caution

The small size of the gap between the platters and the heads is why you should never open the disk drive's HDA except in a clean-room environment. Any particle of dust or dirt that gets into this mechanism could cause the heads to read improperly or possibly even to strike the platters while the drive is running at full speed. The latter event could scratch the platter or the head.

To ensure the cleanliness of the interior of the drive, the HDA is assembled in a class-100 or better clean room. This specification means that a cubic foot of air cannot contain more than 100 particles that measure up to 0.5 microns (19.7  $\mu$ -inch). A single person breathing while standing motionless spews out 500 such particles in a single minute! These rooms contain special air-filtration systems that continuously evacuate and refresh the air. A drive's HDA should not be opened unless it is inside such a room.

Although maintaining a clean-room environment might seem to be expensive, many companies manufacture tabletop or bench-size clean rooms that sell for only a few thousand dollars. Some of these devices operate like a glove box; the operator first inserts the drive and any tools required, and then closes the box and turns on the filtration system. Inside the box, a clean-room environment is maintained, and a technician can use the built-in gloves to work on the drive.

In other clean-room variations, the operator stands at a bench where a forced-air curtain maintains a clean environment on the bench top. The technician can walk in and out of the clean-room field by walking through the air curtain. This air curtain is very similar to the curtain of air used in some stores and warehouses to prevent heat from escaping in the winter while leaving a passage wide open.

Because the clean environment is expensive to produce, few companies, except those that manufacture the drives, are properly equipped to service hard disk drives.

### *Read/Write Head Designs*

As disk drive technology has evolved, so has the design of the read/write head. The earliest heads were simple iron cores with coil windings (electromagnets). By today's standards, the original head designs were enormous in physical size and operated at very low recording densities. Over the years, head designs have evolved from the first simple Ferrite Core designs into the Magneto-Resistive and Giant Magneto-Resistive types available today.

For more information on the various head designs, see Chapter 9, "Magnetic Storage Principles."

## **Head Actuator Mechanisms**

Possibly more important than the heads themselves is the mechanical system that moves them: the head actuator. This mechanism moves the heads across the disk and positions them accurately above the desired cylinder. Many variations on head actuator mechanisms are in use, but all fall into one of two basic categories:

- Stepper motor actuators
- Voice coil actuators

The use of one or the other type of actuator has profound effects on a drive's performance and reliability. The effects are not limited to speed; they also include accuracy, sensitivity to temperature, position, vibration, and overall reliability. The head actuator is the single most important specification in the drive, and the type of head actuator mechanism in a drive tells you a great deal about the drive's performance and reliability characteristics. Table 10.3 shows the two types of hard disk drive head actuators and the affected performance characteristics.

**Table 10.3 Characteristics of Stepper Motor Versus Voice Coil Drives**

Characteristic	Stepper Motor	Voice Coil
Relative access speed	Slow	Fast
Temperature sensitive	Yes (very)	No
Positionally sensitive	Yes	No
Automatic head parking	Not usually	Yes
Preventive maintenance	Periodic reformat	None required
Relative reliability	Poor	Excellent

Generally, a stepper motor drive has a slower average access rating, is temperature sensitive during read and write operations, is sensitive to physical orientation during read and write operations, does not automatically park its heads above a save zone during power-down, and usually requires annual or biannual reformats to realign the sector data with the sector header information due to mistracking. To put it bluntly, a drive equipped with a stepper motor actuator is much less reliable (by a large margin) than a drive equipped with a voice coil actuator.

Floppy disk drives position their heads by using a stepper motor actuator. The accuracy of the stepper mechanism is suited to a floppy disk drive because the track densities usually are nowhere near those of a hard disk. The track density of a 1.44MB floppy disk is 135 tracks per inch, whereas hard disk drives have densities of more than 5,000 tracks per inch. Virtually all the hard disk drives being manufactured today use voice coil actuators because stepper motors cannot achieve the degree of accuracy necessary.

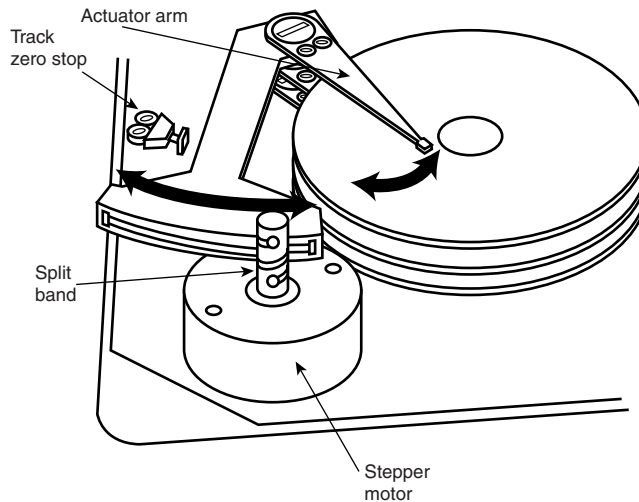
## Stepper Motor Actuators

A stepper motor is an electrical motor that can “step,” or move from position to position, with mechanical detents or click-stop positions. If you were to grip the spindle of one of these motors and spin it manually, you would hear a clicking or buzzing sound as the motor passed each detent position with a soft click.

Stepper motors cannot position themselves between step positions; they can stop only at the predetermined detent positions. The motors are small (between 1 and 3 inches) and can be square, cylindrical, or flat. Stepper motors are outside the sealed HDA, although the spindle of the motor penetrates the HDA through a sealed hole.

Stepper motor mechanisms are affected by a variety of problems; the greatest problem is temperature. As the drive platters heat and cool, they expand and contract, and the tracks on the platters move in relation to a predetermined track position. The stepper mechanism cannot move in increments of less than a single track to correct for these temperature-induced errors. The drive positions the heads to a particular cylinder according to a predetermined number of steps from the stepper motor, with no room for nuance.

Figure 10.8 shows a common stepper motor design, where a split metal band is used to transfer the movement from the rotating motor shaft to the head actuator itself.



**Figure 10.8** A stepper motor actuator.

## Voice Coil Actuators

The voice coil actuators used in virtually all hard disk drives made today—unlike stepper motor actuators—use a feedback signal from the drive to accurately determine the head positions and adjust them, if necessary. This arrangement provides significantly greater performance, accuracy, and reliability than traditional stepper motor actuator designs.

A voice coil actuator works by pure electromagnetic force. The construction of the mechanism is similar to that of a typical audio speaker, from which the term *voice coil* is derived. An audio speaker uses a stationary magnet surrounded by a voice coil, which is connected to the speaker's paper cone. Energizing the coil causes it to move relative to the stationary magnet, which produces sound from

the cone. In a typical hard disk drive's voice coil system, the electromagnetic coil is attached to the end of the head rack and placed near a stationary magnet. No physical contact occurs between the coil and the magnet; instead, the coil moves by pure magnetic force. As the electromagnetic coils are energized, they attract or repulse the stationary magnet and move the head rack. Systems like these are extremely quick and efficient and usually much quieter than systems driven by stepper motors.

Unlike a stepper motor, a voice coil actuator has no click-stops, or detent positions; rather, a special guidance system stops the head rack above a particular cylinder. Because it has no detents, the voice coil actuator can slide the heads in and out smoothly to any position desired. Voice coil actuators use a guidance mechanism called a *servo* to tell the actuator where the heads are in relation to the cylinders and to place the heads accurately at the desired positions. This positioning system often is called a *closed loop feedback mechanism*. It works by sending the index (or servo) signal to the positioning electronics, which return a feedback signal that is used to position the heads accurately. The system also is called *servo-controlled*, which refers to the index or servo information that is used to dictate or control head-positioning accuracy.

A voice coil actuator with servo control is not affected by temperature changes, as a stepper motor is. When temperature changes cause the disk platters to expand or contract, the voice coil system compensates automatically because it never positions the heads in predetermined track positions. Rather, the voice coil system searches for the specific track, guided by the prewritten servo information, and then positions the head rack precisely above the desired track, wherever it happens to be. Because of the continuous feedback of servo information, the heads adjust to the current position of the track at all times. For example, as a drive warms up and the platters expand, the servo information enables the heads to "follow" the track. As a result, a voice coil actuator is sometimes called a *track following system*.

The two main types of voice-coil positioner mechanisms are

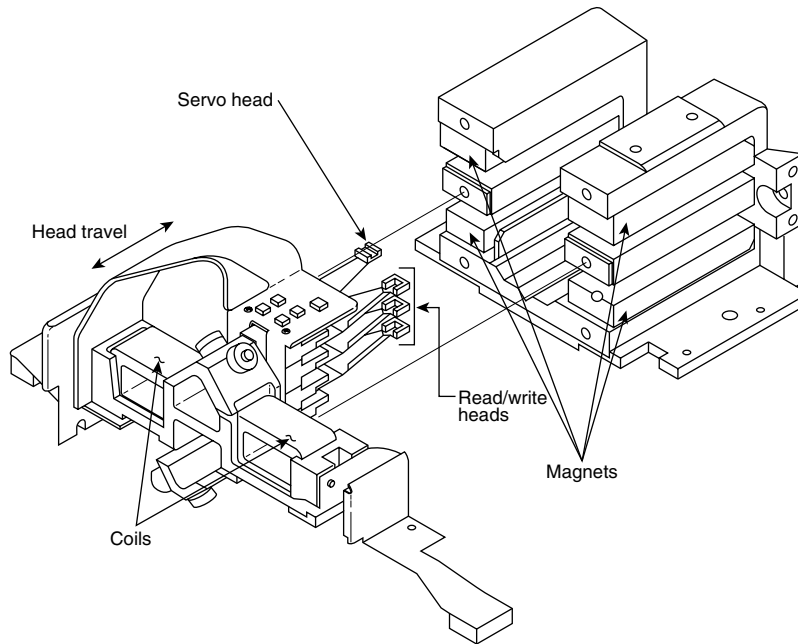
- Linear voice-coil actuators
- Rotary voice-coil actuators

The two types differ only in the physical arrangement of the magnets and coils.

A linear actuator (see Figure 10.9) moves the heads in and out over the platters in a straight line. The coil moves in and out on a track surrounded by the stationary magnets. The primary advantage of the linear design is that it eliminates the head *azimuth* variations that occur with rotary positioning systems. (Azimuth refers to the angular measurement of the head position relative to the tangent of a given cylinder.) A linear actuator does not rotate the head as it moves from one cylinder to another, thus eliminating this problem.

Although the linear actuator seems to be a good design, it has one fatal flaw: The devices are much too heavy. As drive performance has increased, the desire for lightweight actuator mechanisms has become very important. The lighter the mechanism, the faster it can accelerate and decelerate from one cylinder to another. Because they are much heavier than rotary actuators, linear actuators were popular only for a short time; they are virtually nonexistent in drives manufactured today.

Rotary actuators also use stationary magnets and a movable coil, but the coil is attached to the end of an actuator arm. As the coil moves relative to the stationary magnet, it swings the head arms in and out over the surface of the disk. The primary advantage of this mechanism is its light weight, which means that the heads can accelerate and decelerate very quickly, resulting in very fast average seek times. Because of the lever effect on the head arm, the heads move faster than the actuator, which also helps to improve access times. (Refer to Figure 10.7, which shows a rotary voice coil actuator.)



**Figure 10.9** A linear voice coil actuator.

The disadvantage of a rotary system is that as the heads move from the outer to the inner cylinders, they rotate slightly with respect to the tangent of the cylinders. This rotation results in an azimuth error and is one reason why the area of the platter in which the cylinders are located is somewhat limited. By limiting the total motion of the actuator, the azimuth error is contained to within reasonable specifications. Virtually all voice coil drives today use rotary actuator systems.

## Servo Mechanisms

Three servo mechanism designs have been used to control voice coil positioners over the years:

- Wedge servo
- Embedded servo
- Dedicated servo

The three designs are slightly different, but they accomplish the same basic task: They enable the head positioner to adjust continuously so it is precisely positioned above a given cylinder on the disk. The main difference between these servo designs is where the gray code information is actually written on the drive.

All servo mechanisms rely on special information that is written to the disk when it is manufactured. This information is usually in the form of a special code called a *gray code*. A gray code is a special binary notational system in which any two adjacent numbers are represented by a code that differs in only one bit place or column position. This system makes it easy for the head to read the information and quickly determine its precise position.

At the time of manufacture, a special machine called a servowriter writes the servo gray code on the disk. The servowriter is basically a jig that mechanically moves the heads to a given reference position

and then writes the servo information at that position. Many servowriters are themselves guided by a laser-beam reference that calculates its own position by calculating distances in wavelengths of light. Because the servowriter must be capable of moving the heads mechanically, the process requires either that the lid of the drive be removed or that access be available through special access ports in the HDA. After the servowriting is complete, these ports are usually covered with sealing tape. You often see these tape-covered holes on the HDA, usually accompanied by warnings that you will void the warranty if you remove the tape. Because servowriting exposes the interior of the HDA, it requires a clean-room environment.

A servowriter is an expensive piece of machinery, costing up to \$50,000 or more, and often must be custom-made for a particular make or model of drive. Some drive-repair companies have servowriting capability, which means they can rewrite the servo information on a drive if it becomes damaged. If a servowriter is not available, a drive with servo-code damage must be sent back to the drive manufacturer for the servo information to be rewritten.

Fortunately, damaging the servo information through disk read and write processes is impossible. Drives are designed so the heads cannot overwrite the servo information, even during a low-level format. One myth that has been circulating (especially with respect to ATA drives) is that you can damage the servo information by improper low-level formatting. This is not true. An improper low-level format can compromise the performance of the drive, but the servo information is totally protected and cannot be overwritten. Even so, the servo information on some drives can be damaged by a strong adjacent magnetic field or by jarring the drive while it is writing, causing the heads to move off track.

The track-following capabilities of a servo-controlled voice coil actuator eliminate the positioning errors that occur over time with stepper motor drives. Voice coil drives are not affected by conditions such as thermal expansion and contraction of the platters. In fact, many voice coil drives today perform a special thermal-recalibration procedure at predetermined intervals while they run. This procedure usually involves seeking the heads from cylinder 0 to some other cylinder one time for every head on the drive. As this sequence occurs, the control circuitry in the drive monitors how much the track positions have moved since the last time the sequence was performed, and a thermal-recalibration adjustment is calculated and stored in the drive's memory. This information is then used every time the drive positions the heads to ensure the most accurate positioning possible.

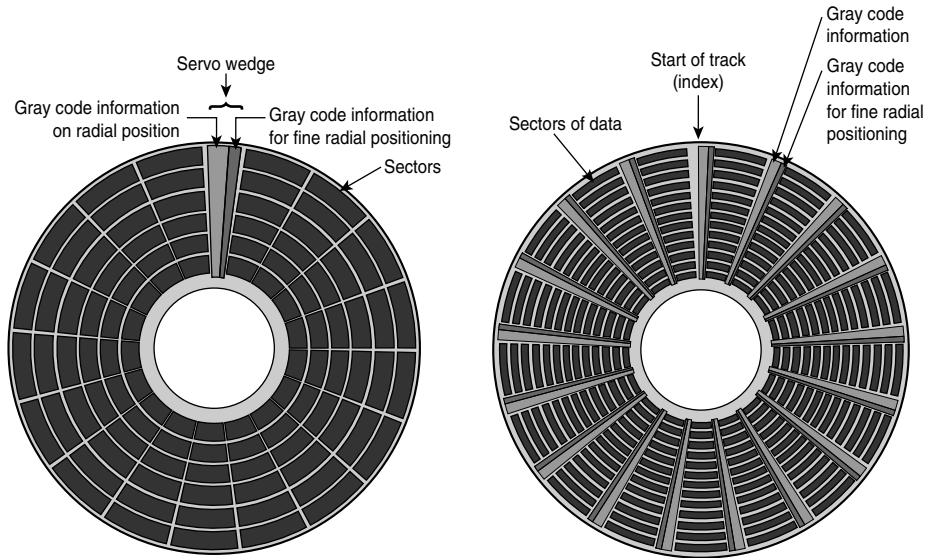
Most drives perform the thermal-recalibration sequence every 5 minutes for the first 30 minutes that the drive is powered on and then once every 25 minutes after that. With some drives, this thermal-recalibration sequence is very noticeable; the drive essentially stops what it is doing, and you hear rapid ticking for a second or so. Some people think this is an indication that their drive is having a problem reading something and perhaps is conducting a read retry, but this is not true. Most drives today (ATA and SCSI) employ this thermal-recalibration procedure to maintain positioning accuracy.

As multimedia applications grew in popularity, thermal recalibration became a problem with some manufacturers' drives. The thermal-recalibration sequence sometimes interrupted the transfer of a large data file, such as an audio or a video file, which resulted in audio or video playback jitter. Some companies released special A/V (audio visual) drives that hide the thermal-recalibration sequences so they never interrupt a file transfer. Most of the newer ATA and SCSI drives are A/V capable, which means that the thermal-recalibration sequences will not interrupt a transfer such as a video playback.

While we are on the subject of automatic drive functions, most of the drives that perform thermal-recalibration sequences also automatically perform a function called a *disk sweep*. Also called *wear leveling* by some manufacturers, this procedure is an automatic head seek that occurs after the drive has been idle for a period of time. The disk-sweep function moves the heads to a cylinder in the outer portion of the platters, which is where the head float-height is highest (because the head-to-platter

velocity is highest). Then, if the drive continues to remain idle for another period, the heads move to another cylinder in this area, and the process continues indefinitely as long as the drive is powered on.

The disk-sweep function is designed to prevent the head from remaining stationary above one cylinder in the drive for too long, where friction between the head and platter eventually would dig a trench in the medium. Although the heads are not in direct contact with the medium, they are so close that the constant air pressure from the head floating above a single cylinder could cause friction and excessive wear. Figure 10.10 shows both a wedge and an embedded servo.



**Figure 10.10** A wedge and an embedded servo.

### *Wedge Servo*

Early servo-controlled drives used a technique called a *wedge servo*. In these drives, the gray-code guidance information is contained in a “wedge” slice of the drive in each cylinder immediately preceding the index mark. The index mark indicates the beginning of each track, so the wedge-servo information was written in the PRE-INDEX GAP, which is at the end of each track. This area is provided for speed tolerance and normally is not used by the controller.

Some controllers had to be notified that the drive was using a wedge servo so they could shorten the sector timing to allow for the wedge-servo area. If they were not correctly configured, these controllers would not work properly with the drive.

Another problem was that the servo information appears only one time every revolution, which means that the drive often needed several revolutions before it could accurately determine and adjust the head position. Because of these problems, the wedge servo never was a popular design; it no longer is used in drives.

### *Embedded Servo*

An embedded servo is an enhancement of the wedge servo. Instead of placing the servo code before the beginning of each cylinder, an embedded servo design writes the servo information before the start of each sector. This arrangement enables the positioner circuits to receive feedback many times

in a single revolution, making the head positioning much faster and more precise. Another advantage is that every track on the drive has its own positioning information, so each head can quickly and efficiently adjust position to compensate for any changes in the platter or head dimensions, especially for changes due to thermal expansion or physical stress.

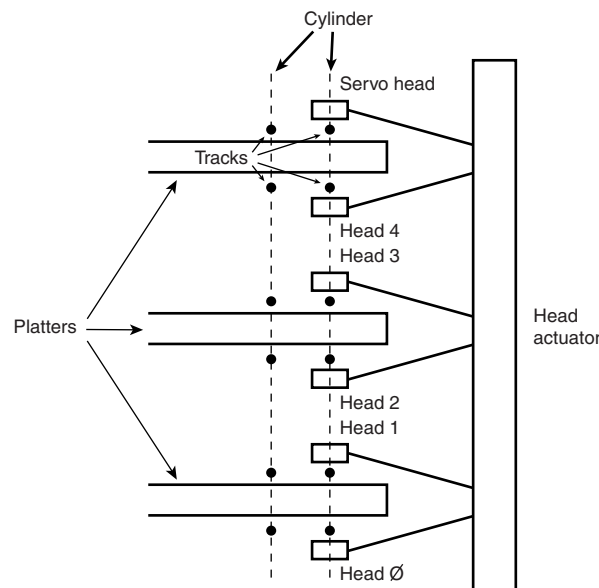
Most drives today use an embedded servo to control the positioning system. As in the wedge servo design, the embedded servo information is protected by the drive circuits, and any write operations are blocked whenever the heads are above the servo information. Thus, it is impossible to overwrite the servo information with a low-level format, as some people incorrectly believe.

Although the embedded servo works much better than the wedge servo because the servo feedback information is made available several times in a single disk revolution, a system that offered continuous servo feedback information would be better.

#### *Dedicated Servo*

A dedicated servo is a design in which the servo information is written continuously throughout the entire track, rather than just once per track or at the beginning of each sector. Unfortunately, if this procedure were used on the entire drive, no room would be left for data. For this reason, a *dedicated servo* uses one side of one of the platters exclusively for the servo-positioning information. The term “dedicated” comes from the fact that this platter side is completely dedicated to the servo information and cannot contain any data.

When building a dedicated servo drive, the manufacturer deducts one side of one platter from normal read/write usage and records a special set of gray-code data there that indicates the proper track positions. Because the head that rests above this surface cannot be used for normal reading and writing, the gray code can never be erased, and the servo information is protected—as in the other servo designs. No low-level format or other procedure can possibly overwrite the servo information. Figure 10.11 shows a dedicated servo mechanism. Normally, the head on top or one in the center is dedicated for servo use.



**Figure 10.11** A dedicated servo, showing one entire head/side used for servo reading.



When the drive moves the heads to a specific cylinder, the internal drive electronics use the signals received by the servo head to determine the position of the read/write heads. As the heads move, the track counters are read from the dedicated servo surface. When the servo head detects the requested track, the actuator stops. The servo electronics then fine-tune the position so the heads are positioned precisely above the desired cylinder before any writing is permitted. Although only one head is used for servo tracking, the other heads are attached to the same rack so if one head is above the desired cylinder, all the others will be as well.

One way of telling whether a drive uses a dedicated servo platter is if it has an odd number of heads. For example, the Toshiba MK-538FB 1.2GB drive that I used to have in one of my systems had eight platters, but only 15 read/write heads. That drive uses a dedicated servo positioning system, and the 16th head is the servo head. The advantage of the dedicated servo concept is that the servo information is continuously available to the drive, making the head positioning process faster and more precise.

The drawback to a dedicated servo is that dedicating an entire platter surface for servo information is wasteful. Virtually all drives today use a variation on the embedded servo technique instead. Some drives combined a dedicated servo with an embedded servo, but this type of hybrid design is rare. Regardless of whether the servo mechanism is dedicated or embedded, it is far more accurate than the stepper motor mechanisms of the past.

Of course, as mentioned earlier, today's ATA and SCSI drives have head, track, and sector-per-track parameters that are translated from the actual physical numbers. Therefore, you usually can't tell from the published numbers exactly how many heads or platters are contained within a drive.

### *Automatic Head Parking*

When you power off a hard disk drive, the spring tension in each head arm pulls the heads into contact with the platters. The drive is designed to sustain thousands of takeoffs and landings, but it is wise to ensure that the landing occurs at a spot on the platter that contains no data. Older drives required manual head parking; you had to run a program that positioned the drive heads to a landing zone, usually the innermost cylinder, before turning the system off. Modern drives automatically park the heads, so park programs are no longer necessary.

Some amount of abrasion occurs during the landing and takeoff process, removing just a "micro puff" from the magnetic medium—but if the drive is jarred during the landing or takeoff process, real damage can occur.

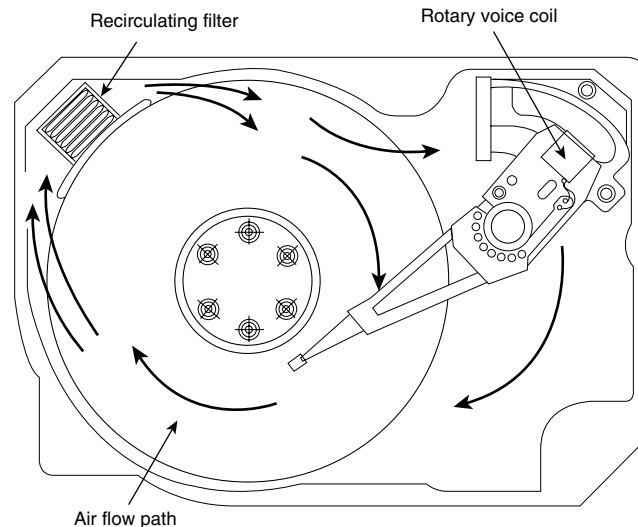
One benefit of using a voice coil actuator is automatic head parking. In a drive that has a voice coil actuator, the heads are positioned and held by magnetic force. When the power to the drive is removed, the magnetic field that holds the heads stationary over a particular cylinder dissipates, enabling the head rack to skitter across the drive surface and potentially cause damage. In the voice coil design, the head rack is attached to a weak spring at one end and a head stop at the other end. When the system is powered on, the spring is overcome by the magnetic force of the positioner. When the drive is powered off, however, the spring gently drags the head rack to a park-and-lock position before the drive slows down and the heads land. On some drives, you can actually hear the "ting...ting...ting...ting" sound as the heads literally bounce-park themselves, driven by this spring.

On a drive with a voice coil actuator, you activate the parking mechanism by turning off the computer; you do not need to run a program to park or retract the heads. In the event of a power outage, the heads park themselves automatically. (The drives unpark automatically when the system is powered on.)

## Air Filters

Nearly all hard disk drives have two air filters. One is called the recirculating filter, and the other is called either a barometric or breather filter. These filters are permanently sealed inside the drive and are designed never to be changed for the life of the drive, unlike many older mainframe hard disks that had changeable filters.

A hard disk on a PC system does not circulate air from inside to outside the HDA or vice versa. The recirculating filter permanently installed inside the HDA is designed to filter only the small particles scraped off the platters during head takeoffs and landings (and possibly any other small particles dislodged inside the drive). Because PC hard disk drives are permanently sealed and do not circulate outside air, they can run in extremely dirty environments (see Figure 10.12).



**Figure 10.12** Air circulation in a hard disk.

The HDA in a hard disk drive is sealed but not airtight. The HDA is vented through a barometric or breather filter element that enables pressure equalization (breathing) between the inside and outside of the drive. For this reason, most hard drives are rated by the drive's manufacturer to run in a specific range of altitudes, usually from 1,000 feet below to 10,000 feet above sea level. In fact, some hard drives are not rated to exceed 7,000 feet while operating because the air pressure would be too low inside the drive to float the heads properly. As the environmental air pressure changes, air bleeds into or out of the drive so internal and external pressures are identical. Although air does bleed through a vent, contamination usually is not a concern because the barometric filter on this vent is designed to filter out all particles larger than 0.3 microns (about 12  $\mu$ -inch) to meet the specifications for cleanliness inside the drive. You can see the vent holes on most drives, which are covered internally by this breather filter. Some drives use even finer grade filter elements to keep out even smaller particles.

I conducted a seminar in Hawaii several years ago, and several of the students were from one of the astronomical observatories atop Mauna Kea. They indicated that virtually all the hard disk drives they had tried to use at the observatory site had failed very quickly, if they worked at all. This was no surprise because the observatories are at the 13,796-foot peak of the mountain, and at that altitude, even people don't function very well! At the time, they had to resort to solid-state (RAM) disks, tape drives, or even floppy disk drives as their primary storage medium. Since then, IBM's Adstar division (which

produces all IBM hard drives) has introduced a line of rugged 3 1/2-inch drives that are hermetically sealed (airtight), although they do have air inside the HDA. Because they carry their own internal air under pressure, these drives can operate at any altitude and can also withstand extremes of shock and temperature. The drives are designed for military and industrial applications, such as systems used aboard aircraft and in extremely harsh environments. They are, of course, more expensive than typical hard drives that operate under ambient pressures.

## Hard Disk Temperature Acclimation

Because hard drives have a filtered port to bleed air into or out of the HDA, moisture can enter the drive, and after some period of time, it must be assumed that the humidity inside any hard disk is similar to that outside the drive. Humidity can become a serious problem if it is allowed to condense—and especially if you power up the drive while this condensation is present. Most hard disk manufacturers have specified procedures for acclimating a hard drive to a new environment with different temperature and humidity ranges, and especially for bringing a drive into a warmer environment in which condensation can form. This situation should be of special concern to users of laptop or portable systems. If you leave a portable system in an automobile trunk during the winter, for example, it could be catastrophic to bring the machine inside and power it up without allowing it to acclimate to the temperature indoors.

The following text and Table 10.4 are taken from the factory packaging that Control Data Corporation (later Imprimis and eventually Seagate) used to ship with its hard drives:

If you have just received or removed this unit from a climate with temperatures at or below 50°F (10°C) do not open this container until the following conditions are met, otherwise condensation could occur and damage to the device and/or media may result. Place this package in the operating environment for the time duration according to the temperature chart.

**Table 10.4 Hard Disk Drive Environmental Acclimation Table**

Previous Climate Temperature	Acclimation Time	Previous Climate Temperature	Acclimation Time
+40°F (+4°C)	13 hours	-10°F (-23°C)	20 hours
+30°F (-1°C)	15 hours	-20°F (-29°C)	22 hours
+20°F (-7°C)	16 hours	-30°F (-34°C) or less	27 hours
+10°F (-12°C)	17 hours		
0°F (-18°C)	18 hours		

As you can see from this table, you must place a hard disk drive that has been stored in a colder-than-normal environment into its normal operating environment for a specified amount of time to allow it to acclimate before you power it on.

## Spindle Motors

The motor that spins the platters is called the spindle motor because it is connected to the spindle around which the platters revolve. Spindle motors in hard disk drives are always connected directly; no belts or gears are involved. The motor must be free of noise and vibration; otherwise, it can transmit a rumble to the platters, which can disrupt reading and writing operations.

The spindle motor also must be precisely controlled for speed. The platters in hard disk drives revolve at speeds ranging from 3,600rpm to 15,000rpm (60–250 revolutions per second) or more, and the motor has a control circuit with a feedback loop to monitor and control this speed precisely. Because

the speed control must be automatic, hard drives do not have a motor-speed adjustment. Some diagnostics programs claim to measure hard drive rotation speed, but all these programs do is estimate the rotational speed by the timing at which sectors pass under the heads.

There is actually no way for a program to measure the hard disk drive's rotational speed; this measurement can be made only with sophisticated test equipment. Don't be alarmed if some diagnostics program tells you that your drive is spinning at an incorrect speed; most likely, the program is wrong, not the drive. Platter rotation and timing information is not provided through the hard disk controller interface. In the past, software could give approximate rotational speed estimates by performing multiple sector read requests and timing them, but this was valid only when all drives had the same number of sectors per track and spun at the same speed. Zoned-bit recording—combined with the many various rotational speeds used by modern drives, not to mention built-in buffers and caches—means that these calculation estimates cannot be performed accurately by software.

On most drives, the spindle motor is on the bottom of the drive, just below the sealed HDA. Many drives today, however, have the spindle motor built directly into the platter hub inside the HDA. By using an internal hub spindle motor, the manufacturer can stack more platters in the drive because the spindle motor takes up no vertical space.

---

**Note**

Spindle motors, particularly on the larger form-factor drives, can consume a great deal of 12-volt power. Most drives require two to three times the normal operating power when the motor first spins the platters. This heavy draw lasts only a few seconds or until the drive platters reach operating speed. If you have more than one drive, you should try to sequence the start of the spindle motors so the power supply does not have to provide such a large load to all the drives at the same time. Most SCSI and some ATA drives have a delayed spindle-motor start feature.

---

## Logic Boards

All hard disk drives have one or more logic boards mounted on them. The logic boards contain the electronics that control the drive's spindle and head actuator systems and present data to the controller in some agreed-upon form. On ATA drives, the boards include the controller itself, whereas SCSI drives include the controller and the SCSI bus adapter circuit.

Many disk drive failures occur in the logic board, not in the mechanical assembly. (This statement does not seem logical, but it is true.) Therefore, you sometimes can repair a failed drive by replacing the logic board rather than the entire drive. Replacing the logic board, moreover, enables you to regain access to the data on the drive—something that replacing the entire drive does not provide.

In many cases, logic boards plug into the drive and are easily replaceable. These boards are usually mounted with standard screw hardware. If a drive is failing and you have a spare, you might be able to verify a logic-board failure by taking the board off the known good drive and mounting it on the bad one. If your suspicions are confirmed, you can order a new logic board from the drive manufacturer, but unless you have data on the drive you need to recover, it might make more sense to buy a new drive, considering today's low disk drive costs.

To reduce costs further, many third-party vendors can also supply replacement logic-board assemblies. These companies often charge much less than the drive manufacturers for the same components. (See the Vendor List on the CD for vendors of drive components, including logic boards.)

## Cables and Connectors

Hard disk drives typically have several connectors for interfacing to the computer, receiving power, and sometimes grounding to the system chassis. Most drives have at least these three types of connectors:

- Interface connector(s)
- Power connector
- Optional ground connector (tab)

Of these, the interface connectors are the most important because they carry the data and command signals between the system and the drive. In most cases, the drive interface cables can be connected in a daisy-chain or bus-type configuration. Most interfaces support at least two devices, and SCSI (Small Computer System Interface) can support up to seven (Wide SCSI can support up to 15) devices in the chain, in addition to the host adapter. Older interfaces, such as ST-506/412 or ESDI (Enhanced Small Device Interface), used separate cables for data and control signals, but today's SCSI and ATA (AT Attachment) drives have a single connector.

◀◀ See "ATA I/O Connector," p. 486, and "SCSI Cables and Connectors," p. 526.

The power connector is usually the same four-pin type that is used in floppy disk drives, and the same power-supply connector plugs into it. Most hard disk drives use both 5- and 12-volt power, although some of the smaller drives designed for portable applications use only 5-volt power. In most cases, the 12-volt power runs the spindle motor and head actuator, and the 5-volt power runs the circuitry. Make sure your power supply can supply adequate power for the hard disk drives installed in your system.

The 12-volt power consumption of a drive usually varies with the physical size of the unit. The larger the drive is, the faster it spins. In addition, the more platters there are to spin, the more power it requires. For example, most of the 3 1/2-inch drives on the market today use roughly one-half to one-fourth the power (in watts) of the older 5 1/4-inch drives. Some of the very small (2 1/2- or 1.8-inch) hard disks barely sip electrical power and actually use 1 watt or less!

A grounding tab provides an optional ground connection between the drive and the system's chassis. In most computers, the hard disk drive is mounted directly to the chassis using screws, or the drive is grounded via the ground wires in the power connector, so an extra ground wire is unnecessary.

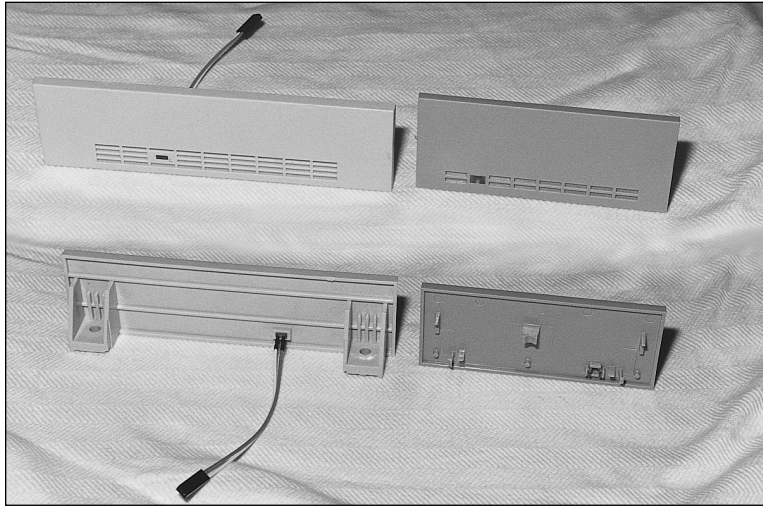
## Configuration Items

To configure a hard disk drive for installation in a system, you usually must set several jumpers (and, possibly, terminating resistors) properly. These items vary from interface to interface and often from drive to drive, as well.

## The Faceplate or Bezel

Many hard disk drives offer as an option a front faceplate, or bezel (see Figure 10.13). A bezel usually is supplied as an option for the drive rather than as a standard item. In most cases today, the bezel is a part of the case and not the drive itself.

Older systems had the drive installed so it was visible outside the system case. To cover the hole in the case, you would use an optional bezel or faceplate. Bezels often come in several sizes and colors to match various PC systems. Many faceplate configurations for 3 1/2-inch drives are available, including bezels that fit 3 1/2-inch drive bays as well as 5 1/4-inch drive bays. You even have a choice of colors (usually black, cream, or white).



**Figure 10.13** Typical 5 1/4- and 3 1/2-inch hard drives bezel shown from the front (as seen on the outside of the PC case) (top) and from the back (bottom—the inside mounting and LED wiring).

Some bezels feature a light-emitting diode (LED) that flickers when your hard disk is in use. The LED is mounted in the bezel; the wire hanging off the back of the LED plugs into the drive. In some drives, the LED is permanently mounted on the drive, and the bezel has a clear or colored window so you can see the LED flicker while the drive is being accessed.

In systems in which the hard disk is hidden by the unit's cover, a bezel is not needed. In fact, using a bezel can prevent the cover from resting on the chassis properly, in which case the bezel will have to be removed. If you are installing a drive that does not have a proper bezel, frame, or rails to attach to the system, check the Vendor List on the CD; several listed vendors offer these accessories for a variety of drives.

## Hard Disk Features

To make the best decision in purchasing a hard disk for your system or to understand what distinguishes one brand of hard disk from another, you must consider many features. This section examines some of the issues you should consider when you evaluate drives:

- Reliability
- Performance
- Cost

### Reliability

When you shop for a drive, you might notice a statistic called the Mean Time Between Failures (MTBF) described in the drive specifications. MTBF figures usually range from 300,000 to 1,000,000 hours or more. I usually ignore these figures because they are derived theoretically.

In understanding the MTBF claims, you must understand how the manufacturers arrive at them and what they mean. Most manufacturers have a long history of building drives, and their drives have seen millions of hours of cumulative use. They can look at the failure rate for previous drive models

with the same components and calculate a failure rate for a new drive based on the components used to build the drive assembly. For the electronic circuit board, they also can use industry standard techniques for predicting the failure of the integrated electronics. This enables them to calculate the predicted failure rate for the entire drive unit.

To understand what these numbers mean, you must know that the MTBF claims apply to a population of drives, not an individual drive. This means that if a drive claims to have an MTBF of 500,000 hours, you can expect a failure in that population of drives in 500,000 hours of total running time. If 1,000,000 drives of this model are in service and all 1,000,000 are running at once, you can expect one failure out of this entire population every half-hour. MTBF statistics are not useful for predicting the failure of any individual drive or a small sample of drives.

You also need to understand the meaning of the word *failure*. In this sense, a failure is a fault that requires the drive to be returned to the manufacturer for repair, not an occasional failure to read or write a file correctly.

Finally, as some drive manufacturers point out, this measure of MTBF should really be called mean time to first failure. "Between failures" implies that the drive fails, is returned for repair, and then at some point fails again. The interval between repair and the second failure here would be the MTBF. Because in most cases, a failed hard drive that would need manufacturer repair is replaced rather than repaired, the whole MTBF concept is misnamed.

The bottom line is that I do not really place much emphasis on MTBF figures. For an individual drive, they are not accurate predictors of reliability. However, if you are an information systems manager considering the purchase of thousands of PCs or drives per year, or a system vendor building and supporting thousands of systems, it is worth your while to examine these numbers and study the methods used to calculate them by each vendor. If you can understand the vendor's calculations and compare the actual reliability of a large sample of drives, you can purchase more reliable drives and save time and money in service and support.

### S.M.A.R.T.

S.M.A.R.T. (Self-Monitoring, Analysis, and Reporting Technology) is an industry standard providing failure prediction for disk drives. When S.M.A.R.T. is enabled for a given drive, the drive monitors predetermined attributes that are susceptible to or indicative of drive degradation. Based on changes in the monitored attributes, a failure prediction can be made. If a failure is deemed likely to occur, S.M.A.R.T. makes a status report available so the system BIOS or driver software can notify the user of the impending problems, perhaps enabling the user to back up the data on the drive before any real problems occur.

Predictable failures are the types of failures S.M.A.R.T. attempts to detect. These failures result from the gradual degradation of the drive's performance. According to Seagate, 60% of drive failures are mechanical, which is exactly the type of failures S.M.A.R.T. is designed to predict.

Of course, not all failures are predictable, and S.M.A.R.T. cannot help with unpredictable failures that occur without any advance warning. These can be caused by static electricity; improper handling or sudden shock; or circuit failure, such as thermal-related solder problems or component failure.

S.M.A.R.T. was originally created by IBM in 1992. That year IBM began shipping 3 1/2-inch hard disk drives equipped with Predictive Failure Analysis (PFA), an IBM-developed technology that periodically measures selected drive attributes and sends a warning message when a predefined threshold is exceeded. IBM turned this technology over to the ANSI organization, and it subsequently became the ANSI-standard S.M.A.R.T. protocol for SCSI drives, as defined in the ANSI-SCSI Informational Exception Control (IEC) document X3T10/94-190.

Interest in extending this technology to IDE/ATA drives led to the creation of the S.M.A.R.T. Working Group in 1995. Besides IBM, other companies represented in the original group were Seagate Technology, Conner Peripherals (now a part of Seagate), Fujitsu, Hewlett-Packard, Maxtor, Quantum, and Western Digital. The S.M.A.R.T. specification produced by this group and placed in the public domain covers both IDE/ATA and SCSI hard disk drives and can be found in most of the more recently produced drives on the market.

The S.M.A.R.T. design of attributes and thresholds is similar in IDE/ATA and SCSI environments, but the reporting of information differs.

In an IDE/ATA environment, driver software on the system interprets the alarm signal from the drive generated by the S.M.A.R.T. “report status” command. The driver polls the drive on a regular basis to check the status of this command and, if it signals imminent failure, sends an alarm to the operating system where it will be passed on via an error message to the end user. This structure also enables future enhancements, which might allow reporting of information other than drive failure conditions. The system can read and evaluate the attributes and alarms reported in addition to the basic “report status” command.

SCSI drives with S.M.A.R.T. communicate a reliability condition only as either good or failing. In a SCSI environment, the failure decision occurs at the disk drive, and the host notifies the user for action. The SCSI specification provides for a sense bit to be flagged if the drive determines that a reliability issue exists. The system then alerts the end user via a message.

The basic requirements for S.M.A.R.T. to function in a system are simple. All you need are a S.M.A.R.T.-capable hard disk drive and a S.M.A.R.T.-aware BIOS or hard disk driver for your particular operating system. If your BIOS does not support S.M.A.R.T., utility programs are available that can support S.M.A.R.T. on a given system. These include Norton Disk Doctor from Symantec, EZ Drive from StorageSoft, and Data Advisor from Ontrack Data International.

Note that traditional disk diagnostics, such as Scandisk and Norton Disk Doctor, work only on the data sectors of the disk surface and do not monitor all the drive functions that are monitored by S.M.A.R.T. Most modern disk drives keep spare sectors available to use as substitutes for sectors that have errors. When one of these spares is reallocated, the drive reports the activity to the S.M.A.R.T. counter but still looks completely defect-free to a surface analysis utility, such as Scandisk.

Drives with S.M.A.R.T. monitor a variety of attributes that vary from one manufacturer to another. Attributes are selected by the device manufacturer based on their capability to contribute to the prediction of degrading or fault conditions for that particular drive. Most drive manufacturers consider the specific set of attributes being used and the identity of those attributes as vendor specific and proprietary.

Some drives monitor the floating height of the head above the magnetic media. If this height changes from a nominal figure, the drive could fail. Other drives can monitor different attributes, such as ECC (error correcting code) circuitry that indicates whether soft errors are occurring when reading or writing data. Some of the attributes monitored on various drives include the following:

- Head floating height
- Data throughput performance
- Spin-up time
- Reallocated (spared) sector count
- Seek error rate
- Seek time performance



- Drive spin-up retry count
- Drive calibration retry count

Each attribute has a threshold limit that is used to determine the existence of a degrading or fault condition. These thresholds are set by the drive manufacturer and cannot be changed.

When sufficient changes occur in the monitored attributes to trigger a S.M.A.R.T. alert, the drive sends an alert message via an IDE/ATA or a SCSI command (depending on the type of hard disk drive you have) to the hard disk driver in the system BIOS, which then forwards the message to the operating system. The operating system then displays a warning message as follows:

Immediately back up your data and replace your hard disk drive. A failure may be imminent.

The message might contain additional information, such as which physical device initiated the alert; a list of the logical drives (partitions) that correspond to the physical device; and even the type, manufacturer, and serial number of the device.

The first thing to do when you receive such an alert is to heed the warning and back up all the data on the drive. It also is wise to back up to new media and not overwrite any previous good backups you might have, just in case the drive fails before the backup is complete.

After backing up your data, what should you do? S.M.A.R.T. warnings can be caused by an external source and might not actually indicate that the drive itself is going to fail. For example, environmental changes, such as high or low ambient temperatures, can trigger a S.M.A.R.T. alert, as can excessive vibration in the drive caused by an external source. Additionally, electrical interference from motors or other devices on the same circuit as your PC can induce these alerts.

If the alert was not caused by an external source, a drive replacement might be indicated. If the drive is under warranty, contact the vendor and ask them whether they will replace it. If no further alerts occur, the problem might have been an anomaly, and you might not need to replace the drive. If you receive further alerts, replacing the drive is recommended. If you can connect both the new and existing (failing) drive to the same system, you might be able to copy the entire contents of the existing drive to the new one, saving you from having to install or reload all the applications and data from your backup.

## Performance

When you select a hard disk drive, one of the important features you should consider is the performance (speed) of the drive. Hard drives can have a wide range of performance capabilities. As is true of many things, one of the best indicators of a drive's relative performance is its price. An old saying from the automobile-racing industry is appropriate here: "Speed costs money. How fast do you want to go?"

You can measure the speed of a disk drive in two ways:

- Average seek time
- Transfer rate

Average seek time, normally measured in milliseconds (ms), is the average amount of time it takes to move the heads from one cylinder to another a random distance away. One way to measure this specification is to run many random track-seek operations and then divide the timed results by the number of seeks performed. This method provides an average time for a single seek.

The standard method used by many drive manufacturers to measure the average seek time involves measuring the time it takes the heads to move across one-third of the total cylinders. Average seek

time depends only on the drive itself; the type of interface or controller has little effect on this specification. The rating is a gauge of the capabilities of the head actuator.

### Note

Be wary of benchmarks that claim to measure drive seek performance. Most ATA/IDE and SCSI drives use a scheme called sector translation, so any commands the drive receives to move the heads to a specific cylinder might not actually result in the intended physical movement. This situation renders some benchmarks meaningless for those types of drives. SCSI drives also require an additional step because the commands first must be sent to the drive over the SCSI bus. These drives might seem to have the fastest access times because the command overhead is not factored in by most benchmarks. However, when this overhead is factored in by benchmark programs, these drives receive poor performance figures.

### Average Access Time

A slightly different measurement, called average access time, involves another element called *latency*. Latency is the average time (in milliseconds) it takes for a sector to be available after the heads have reached a track. On average, this figure is half the time it takes for the disk to rotate once. A drive that spins twice as fast would have half the latency. A measurement of a drive's average access time is the sum of its average seek time and latency. This number provides the average amount of time required before the drive can access a randomly requested sector.

### Latency

Latency is a factor in disk read and write performance. Decreasing the latency increases the speed of access to data or files and is accomplished only by spinning the drive platters more quickly. Latency figures for most popular drive rotational speeds are shown in Table 10.5.

**Table 10.5 Hard Disk Rotation Speeds and Their Latencies**

Revs/Minute	Revs/Second	Latency
3,600	60	8.33
4,200	70	7.14
5,400	90	5.56
7,200	120	4.17
10,000	167	3.00
15,000	250	2.00

Many drives today spin at 7,200rpm, resulting in a latency time of only 4.17ms, whereas others spin at 10,000rpm or even 15,000rpm, resulting in incredible 3.00ms or 2.00ms latency figures. In addition to increasing performance where real-world access to data is concerned, spinning the platters more quickly also increases the data-transfer rate after the heads arrive at the desired sectors.

### Transfer Rate

The transfer rate is probably more important to overall system performance than any other statistic, but it is also one of the most misunderstood specifications. The problem stems from the fact that now several transfer rates can be specified for a given drive.

Most drive manufacturers now report up to five transfer rates. One is the interface transfer rate, which for most newer ATA drives is 100MB/sec. The other (FAR more important) transfer rate specifications are the media transfer rates, which can be expressed as a raw maximum, a raw minimum, a formatted maximum, or a formatted minimum. Few report the average, but that can be easily calculated.

The media transfer rate is far more important than the interface transfer rate because the media transfer rate is the true rate at which data can be read from the disk, which is how fast data can be read from the drive platters (media). It is the maximum rate that any sustained transfer can hope to achieve.

Drive makers report a minimum and maximum media transfer rate because drives today have a zoned recording with various numbers of sectors per track. Typically, a drive is divided into 16 zones, and the inner zone has about half the sectors (and therefore about half the transfer rate) of the outer zone. Because the drive spins at a constant rate, data is read more quickly from the outer cylinders (with more sectors per track) than from the inner cylinders.

Another issue is the raw transfer rate versus the formatted transfer rate. The *raw* rate refers to how fast bits can be read off the media. Because not all bits represent data (some are intersector or ID bits), and because some time is lost when the heads have to move from track to track, the *formatted* transfer rate represents the true rate at which data can be read from or written to the drive.

Note that some manufacturers report only raw transfer rates, but you usually can figure that the formatted rates are about two-thirds of the raw rates. Likewise, some manufacturers report only maximum transfer rates (either raw, formatted, or both); in that case you can generally assume the minimum transfer rate is one-half of the maximum and that the average transfer rate is three-fourths of the maximum.

Let's look at a specific drive as an example. The IBM Deskstar 60GXP is one of the fastest ATA/IDE drives on the market, spins at 7,200rpm, and fully supports the ATA/100 interface transfer rate (100MB/sec from the controller to the motherboard). As with all current drives, the actual transfer rate is much less.

Table 10.6 shows the specifications for the 7,200rpm Ultra-ATA/100 IBM Deskstar 60GXP drive.

**Table 10.6 Transfer Rate Specifications for the IBM Deskstar 60GXP Drive**

Transfer Rates	Megabits/Sec	Megabytes/Sec
Interface Transfer Rate	800Mb/sec	100.0MB/sec
Raw Media Transfer Rate (Min)	253Mb/sec	31.6MB/sec
Raw Media Transfer Rate (Max)	494Mb/sec	61.8MB/sec
Formatted Media Transfer Rate (Min)	167Mb/sec	20.9MB/sec
Formatted Media Transfer Rate (Max)	326Mb/sec	40.8MB/sec
Formatted Media Transfer Rate (Avg)	247Mb/sec	30.8MB/sec

MB/sec = Million bytes per second; Mb/sec = Million bits per second

As you can see, the TRUE media transfer rate for this drive is between 20.9MB/sec and 40.8MB/sec, or an average of about 30.8MB/sec, or less than one-third of the interface transfer rate. Of course, if this were your drive you shouldn't be disappointed because 30.8MB/sec is excellent performance—in fact, this is one of the fastest ATA drives on the market. Many, if not most other, ATA/IDE drives would have equal or slower performance.

A common question I get is about upgrading the ATA interface in a system. Many people are using older motherboards that support only ATA/33 or ATA/66 modes and not the faster ATA/100 specification. After studying the true formatted media transfer rates of most drives, you can see why I generally do not recommend installing a separate ATA/100 controller for those systems. Those who perform such an upgrade will see little if any increase in performance. This is because in almost all cases, the drives they are using will be on average slower than even ATA/33—and often significantly slower than the ATA/66 or ATA/100 interface speeds.

Two primary factors contribute to transfer rate performance: rotational speed and the linear recording density or sector-per-track figures. When comparing two drives with the same number of sectors per track, the drive that spins more quickly will transfer data more quickly. Likewise, when comparing two drives with identical rotational speeds, the drive with the higher recording density (more sectors per track) will be faster. A higher-density drive can be faster than one that spins faster—both factors have to be taken into account.

Let's look at another example. One of the fastest rotating drives today is the Seagate Cheetah X15, which spins at 15,000rpm.

Table 10.7 shows the specifications for the 15,000rpm Ultra3-SCSI/160 Seagate Cheetah X15 drive.

**Table 10.7 Transfer Rate Specifications for the Seagate Cheetah X15 Drive**

Transfer Rates	Megabits/Sec	Megabytes/Sec
Interface Transfer Rate	1,280Mb/sec	160.0MB/sec
Raw Media Transfer Rate (Min)	385Mb/sec	48.1MB/sec
Raw Media Transfer Rate (Max)	508Mb/sec	63.5MB/sec
Formatted Media Transfer Rate (Min)	299Mb/sec	37.4MB/sec
Formatted Media Transfer Rate (Max)	391Mb/sec	48.9MB/sec
Formatted Media Transfer Rate (Avg)	345Mb/sec	43.2MB/sec

As a comparison, Table 10.8 contains the specifications for the 10,000rpm Ultra3-SCSI/160 Seagate Cheetah 73LP drive.

**Table 10.8 Transfer Rates for the Seagate Cheetah 73LP Drive**

Transfer Rates	Megabits/Sec	Megabytes/Sec
Interface Transfer Rate	1,280Mb/sec	160.0MB/sec
Raw Media Transfer Rate (Min)	399Mb/sec	49.9MB/sec
Raw Media Transfer Rate (Max)	671Mb/sec	83.9MB/sec
Formatted Media Transfer Rate (Min)	307Mb/sec	38.4MB/sec
Formatted Media Transfer Rate (Max)	511Mb/sec	63.9MB/sec
Formatted Media Transfer Rate (Avg)	409Mb/sec	51.2MB/sec

As you can see, in this case the 10,000rpm Cheetah 73LP is significantly faster than the 15,000rpm Cheetah X15. Neither of these drives, however, is close to the Ultra3 SCSI (160MB/sec) bandwidth that the interface would allow. With a comparison such as this, you can see that you need to be careful. Don't just compare one specification, such as rotational speed, because in some cases (like this one), the drive that spins twice as quickly is actually slower in transfer performance. This is a good

reminder why you need to be careful with simplistic comparisons. With hard drives, the bottom line is that the media transfer rate is probably the most important specification you can know about a drive, and faster is better.

With older drives the manufacturers often reported minimum and maximum sector per track specifications, which—combined with the rotational speed—can be used to calculate true formatted media performance. Also, be aware that many drives (especially zoned-bit recording drives) are configured with sector translation, so the number of sectors per track reported by the BIOS has little to do with the actual physical characteristics of the drive. You must know the drive's true physical parameters, rather than the values the BIOS uses.

When you know these figures, you can use the following formula to determine the true media data transfer rate in millions of bytes per second (MBps):

$$\text{Media Transfer Rate (MBps)} = \text{SPT} \times 512 \text{ bytes} \times \text{rpm} / 60 \text{ seconds} / 1,000,000 \text{ bytes}$$

For example, the IBM Travelstar 32GH 32GB 2 1/2-inch drive spins at 5,411rpm and has an average of 472 sectors per track. The average media transfer rate for this drive is figured as follows:

$$472 \times 512 \times (5,411 / 60) / 1,000,000 = 21.8 \text{ MBps}$$

Using this formula, you can calculate the media transfer rate of any drive, if you know the rotational speed and average sectors per track.

Some drives report their transfer speeds as a maximum sustainable formatted rate. For example, the IBM Deskstar 60GXP drive detailed earlier has minimum and maximum formatted transfer rates of 20.9MB/sec and 40.8MB/sec, respectively (to or from the media), and spins at 7,200rpm, which works out to be:

$$20.9 \times 1,000,000 / (7200 / 60) / 512 = 340 \text{ sectors per track maximum}$$

$$40.8 \times 1,000,000 / (7200 / 60) / 512 = 664 \text{ sectors per track maximum}$$

Averaging these figures results in a 30.8MB/sec average transfer rate and 502 sectors per track average.

Note that these are the true numbers of sectors per track (with differing values across 18 zones in this case) even though the drive would report only 63 sectors per track if queried with the ATA Identify command, such as in a BIOS Setup with autodetect. Don't be confused because the numbers reported by the drive are different from what is actually inside. For example, this drive has 6 actual heads even though it would report 16, and it has about 34,000 actual cylinders even though it would report 16,384 to the BIOS setup.

## *Cache Programs and Caching Controllers*

At the software level, disk cache programs, such as SMARTDRV (in DOS) and VCCACHE (in Windows 9x, Windows NT, and Windows 2000), can have a major effect on disk drive performance. These cache programs hook into the BIOS hard drive interrupt and intercept the read and write calls to the disk BIOS from application programs and device drivers.

When an application program wants to read data from a hard drive, the cache program intercepts the read request, passes the read request to the hard drive controller in the usual way, saves the data read from the disk in its cache memory buffer, and then passes the data back to the application program. Depending on the size of the cache buffer, data from numerous sectors can be read into and saved in the buffer.

When the application wants to read more data, the cache program again intercepts the request and examines its buffers to see whether the requested data is still in the cache. If so, the program passes

the data back from the cache to the application immediately, without another hard drive operation. Because the cached data is stored in memory, this method speeds access tremendously and can greatly affect disk drive performance measurements.

Most controllers now have some form of built-in hardware buffer or cache that doesn't intercept or use any BIOS interrupts. Instead, the drive caches data at the hardware level, which is invisible to normal performance-measurement software. Manufacturers originally included track read-ahead buffers in controllers to permit 1:1 interleave performance. Some manufacturers now increase the size of these read-ahead buffers in the controller, whereas others add intelligence by using a cache instead of a simple buffer.

Many ATA and SCSI drives have cache memory built directly into the drive's onboard controller. Most newer ATA drives have between 512KB and 2MB of built-in cache, whereas many SCSI drives have up to 16MB. I remember when 640KB was a lot of memory for an entire system. Now, tiny 3 1/2-inch hard disk drives can have 16MB built right in! These integrated caches are part of the reason many ATA (IDE) and SCSI drives perform so well.

Although software and hardware caches can make a drive faster for routine transfer operations, a cache will not affect the true maximum transfer rate the drive can sustain.

### *Interleave Selection*

In a discussion of disk performance, the issue of interleave often comes up. Although traditionally this was more a controller performance issue than a drive issue, modern ATA/IDE and SCSI hard disk drives with built-in controllers are fully capable of processing the data as fast as the drive can send it. In other words, all modern ATA and SCSI drives are formatted with no interleave (sometimes expressed as a 1:1 interleave ratio). On older hard drive types, such as MFM and ESDI, you could modify the interleave during a low-level format to optimize the drive's performance. Today, drives are low-level formatted at the factory and interleave adjustments are a moot topic.

### **Note**

---

For more information on interleaving and cylinder skewing as used on older drives, see Chapter 10 of *Upgrading and Repairing PCs*, 12th Edition, included in its entirety on the CD included with this book.

---

## **Cost**

The cost of hard disk storage is continually falling. You can now purchase a 30GB ATA drive for around \$150, which is about half a cent per megabyte.

A drive I bought in 1983 had a maximum capacity of 10MB and cost \$1,800. At current pricing (0.5 cents per megabyte), that drive is worth about 5 cents!

Of course, the cost of drives continues to fall, and eventually, even half a cent per megabyte will seem expensive. Because of the low costs of disk storage today, not many 3-1/2 inch drives with capacities of less than 10GB are even being manufactured.

## **Capacity**

Four figures are commonly used in advertising drive capacity:

- Unformatted capacity, in millions of bytes
- Formatted capacity, in millions of bytes
- Unformatted capacity, in megabytes
- Formatted capacity, in megabytes

The term *formatted*, in these figures, refers to the low-level (or physical) formatting of the drive. Most manufacturers of ATA/IDE and SCSI drives now report only the formatted capacities because these drives are delivered preformatted. Usually, advertisements and specifications refer to the unformatted or formatted capacity in millions of bytes because these figures are larger than the same capacity expressed in megabytes. This situation generates a great deal of confusion when the user runs FDISK (which reports total drive capacity in megabytes) and wonders where the missing space is. This question can seem troubling. Fortunately, the answer is easy; it only involves a little math to figure it out.

Perhaps the most common questions I get are concerning “missing” drive capacity. Consider the following example: I just installed a new Seagate ST330630A drive in a system, which is advertised as having 30.6GB capacity. After entering the drive parameters in the BIOS Setup (I used the autodetect feature to automate the process), when I went to partition the drive using FDISK, the capacity was reported by FDISK as only 203MB! “What happened to the other 29.8GB?”

The answer is only a few calculations away. By multiplying the drive specification parameters, you get this result:

Total sectors:	59,777,640
Bytes per sector:	512
Total bytes (in decimal megabytes):	30,606
Total bytes (in decimal gigabytes):	30.6
Total bytes (in binary megabytes):	29,188
Total bytes (in binary gigabytes):	28.5
As reported by FDISK:	29,188

All the numbers in the table above are correct. Drive manufacturers usually report drive capacity in decimal megabytes (millions of bytes) because they result in larger, more impressive sounding numbers, although your BIOS and especially the FDISK drive partitioning software report the capacity in binary megabytes. One decimal megabyte equals one million bytes, whereas one binary megabyte equals 1,048,576 bytes (or 1,024KB, in which each kilobyte is 1,024 bytes). So the bottom line is that, because the same abbreviations are often used for both millions of bytes and megabytes, this 30.6GB drive could also be called a 28.5GB drive, depending on how you look at it!

## Specific Recommendations

If you are going to add a hard disk to a system today, I can give you a few recommendations. For the drive interface, there really are only two types to consider:

- ATA (AT Attachment, also called IDE)
- SCSI (Small Computer System Interface)

These interfaces are covered more completely in Chapters 7, “The IDE Interface,” and 8, “The SCSI Interface.” SCSI offers great expandability, cross-platform compatibility, high capacity, performance, and flexibility. ATA is less expensive than SCSI and also offers a very high-performance solution, but expansion, compatibility, capacity, and flexibility are more limited when compared with SCSI. I usually recommend ATA for most people for two reasons. First, the interface is already integrated into virtually every motherboard and BIOS sold today. Second, most users will not need more capacity than the four devices supported by the standard primary and secondary ATA interfaces found in most systems. SCSI offers additional performance potential with a multithreaded operating system, such as Windows NT/2000, as well as support for more devices, but it also requires the purchase of a separate host adapter card, which is in addition to the higher cost for the drive itself.

**Note**

---

Note that the current fastest ATA standard is UltraDMA/100 or UltraATA/100, which runs at 100MB/sec raw interface transfer rate. Serial ATA is coming out with a 150MB/sec transfer rate. The current fastest SCSI standard is Ultra4 (also called Ultra/320)-SCSI, which runs a raw interface transfer rate of 320MB/sec. Both interfaces have improvements slated that should double their respective transfer rates in the future.

---