Hard Disk Performance, Quality and Reliability

- When considering the actual "real world" daily use of hard disks, and contemplating a hard disk purchase,
- PC users often ask **3 key questions:**
- 1. Is this hard disk fast?
- 2. Is this hard disk well manufactured?
- 3. Is this hard disk going to last?

- These questions are answered by considering carefully the three words that appear in the title of this page. Hard disk performance is important because hard disks are one of the slowest internal PC components, and therefore often limit the performance of the system as a whole. Quality and reliability are *critical* with hard disks, because they are where your data resides! No other PC component can lead so readily to disaster if it fails.

Hard Disk Performance, Quality and Reliability

- This section is quite large, and it has to be. Why? Because these matters are so crucial, and unfortunately, few of the "big players" in the industry do a good enough job of educating hard disk users about their products.
- Here I am talking especially about those who make hard disks, whole PCs and operating systems. Hard disks continue to be sold by flashing appetizing numbers on boxes, placed there to entice **consumers** who **can't easily understand** where they fit into the **big picture of overall performance.**
- **Benchmarks** are tossed around the Internet with little knowledge and even less scientific control.
- And **hard disk quality** and **reliability remain mysteries** to **many users**, who buy whatever seems fast and cheap, and simply hope for the best.

Hard Disk Performance

- There was a time when the performance of hard disks was one of the most underrated aspects of overall system performance. Hard disks were considered only "a place to keep stuff" and little heed given to how they affected the operation of the PC as a whole. Over the last few years this has changed dramatically, and hard disk performance issues are now getting the attention they deserve.
- The problem with hard **disk performance however is that it is not easy to understand. There** are **many different issues** in how **performance is assessed and measured**. There are interactions between components with which anyone seeking to really grasp hard disk performance must contend.
- And the **technology changes so rapidly** that what is the "fastest" today will probably be "second best" within a few months--or even weeks in many cases!

Hard Disk Performance

- A **performance *specification*** is a **figure** provided by a manufacturer that **attempts** to express an **element** of the drive's performance relative to other drives on the market.
- A **performance *factor*** is something about the drive (internal factor) or its environment (external factor) that **affects** the performance of the drive.
- Clearly, these are interrelated; factors affect specifications, and specifications indicate the impact of various factors.

- Before even delving into this section, it's important to remember that **performance always comes at a price**. As with most things, there are tradeoffs involved. If performance is your chief goal in a storage system you can have it, but you will either pay for it, or give up something else in exchange (capacity, flexibility, simplicity, etc.) There are no free lunches, although the general increase in performance of *all* hard disks means that even inexpensive drives today perform very well compared to their predecessors.

Hard Disk General Performance Issues

- Alright, so you want to know about hard disk performance. But what exactly does "performance" mean?

- This introductory section looks at some of the basic issues involved in understanding hard disk performance. This includes a discussion of the importance of hard disk performance, a look at various ways performance can be considered, and a discussion of the relative importance of **various performance specifications and factors.**

The Importance of Hard Disk Performance

- I keep insisting that the performance of the hard disk is important. But why? Surely the hard disk's performance level can't be as important as that of the CPU, or memory, or other core system components. Or can it?

- It is true that for many aspects of computing, the hard disk's performance level is not much of an issue. If you are recalculating a massive spreadsheet, or doing complex rendering of 3D objects, the amount of sheer processing power in the system is of paramount concern; the hard **disk will only come into play periodically**. However, these sorts of specialty operations are not indicative of how most of us use our PCs. Typical PC use **involves loading programs, and loading and saving data frequently.** All of these operations require access to the hard disk. And therefore, hard disk performance becomes an issue constantly as we use our machines.

- The importance of hard disk performance even goes beyond this however. After all, we also use all the *other* main components of our system constantly, so aren't they equally important to the performance equation? Well, yes and no. The importance of the CPU, motherboard and other core components *is* very important. But much as the strength of a chain is equal only to that of its weakest link, in many ways the performance of a system is only equal to that of its poorest-performing component. Compared to the solid state components in a PC, hard disks have by far the worst performance. And even as hard disks improve in speed, CPUs, video cards and motherboards improve in speed even faster, widening the gap. Thus, **hard disks continue** to **constrain the overall performance of many systems**.

The Importance of Hard Disk Performance

- In the amount of time it takes to perform one random hard disk access, one of today's CPUs can execute over a million instructions! Making the CPU fast enough to process *two* million instructions while it waits doesn't really gain you much unless it has something to do with that time. Only improving the hard disk's speed to reduce the wait time will get you where you want to go. Any time you see your hard disk's activity light flashing, the odds are that the rest of the system is twiddling its thumbs waiting for those slow mechanical components to "do their thing".

- The **applications** where hard disk performance issues are most important are obviously **those that do a lot of reading and writing to the hard disk**, instead of doing a lot of processing. Such tasks are said to be "I/O bound", where "I/O" stands for "input/output". These tasks are contrasted to those described earlier which use the CPU a great deal and are called (unsurprisingly) "CPU bound". **Multimedia editing applications**, especially those dealing with large audio and video files, are probably the ones most affected by the speed of the storage subsystem. Also up there are applications that process files, including compilers and many disk utilities. Initially starting your PC is also a very I/O-intensive application, as the operating system loads literally hundreds of files. Improving hard disk performance can shave time off the boot process in a very palpable way.

- The need for improved performance is a major driving factor behind the rise in popularity of RAID.

Internal vs. External Performance

- The hard disk's job is to **store data** from the system, **or get data** to the system, as fast as possible. When considering performance, it is this ability to move data into or out of the hard disk that we are looking to measure. There are **2 separate parts** of **this data movement job**. **For** a **write**, the data must be fetched from the system, and then written to the correct sector(s) on the disk. **For a read**, the process is reversed; data must be read from the disk, and then transmitted over the hard disk interface to the system.

- Clearly, the **hard disk itself is only responsible** for **some portion** of the **overall performance** we attribute to the hard disk subsystem. Some of the factors that affect performance are related to characteristics of the PC that are not specific to the particular hard drive being used.

- Performance characteristics that are largely a matter of how the hard disk itself is designed and implemented I call ***internal performance factors***;

- those that are mostly affected by the rest of the system, or how the hard disk is used (and hence are largely independent of the particular hard disk model) are ***external performance factors*****.**

Internal vs. External Performance

- The distinction between internal and external is a very important one! In any system, the **bottleneck** to **high performance can reside either within the disk**, or **within the rest of the system** (the interface, the system bus, CPU, drivers, file system, and so on.) It's usually not in both at the same time. If the main limiting factor in a particular system is, say, the system bus being used for the hard disk interface, putting a faster hard disk into that system will have very little impact on performance. Similarly, if the hard disk itself is slow, putting it on a faster interface will yield little improvement.

- In some cases, performance bottlenecks can change from being primarily affected by internal factors, to being more influenced by external factors, depending on what type of work is being done on the PC. In addition, making a change to a system can cause the bottleneck to "shift". Let's say you have a high-speed hard disk that is on a slow interface. The interface may be the bottleneck to high performance; if you move the disk onto a faster interface, the disk itself may become the bottleneck. In many PCs, external performance can be enhanced simply by making no-cost performance enhancements to the system.

Positioning vs. Transfer Performance

- As discussed here, the process of reading or writing to the hard disk really comprises two large parts: the work done by the drive itself (measured by internal performance factors) and the rest of the system (external factors).
- The **internal work** of the drive can itself be thought of as **2 functions**: **finding the correct location on the drive**, and **then reading or writing the data**. These are very different jobs, and two drives can be very similar in one regard but very different in another. They also depend on different design characteristics. I call these two different tasks positioning and transfer, and their performance positioning performance and transfer performance.
- Both of these are important to overall performance, although if you read the literature and the numbers that people talk about, positioning metrics are probably more commonly discussed than transfer measurements. You might be fooled by this into thinking they are more important, but often they are not--they are just simpler to explain in many cases, or people are used to using them to compare drives.

Positioning vs. Transfer Performance

- Which influences on performance are most important also depends on how you are using the device. If you are running a **file server**, the hard disk will be doing a **lot of random accesses** to files all over the disk, and **positioning performance** will be extremely important.
- If you are a single user doing **multimedia editing** where you need to read multi-gigabyte consecutive files as fast as possible, **data transfer** is far more important than positioning speed.

- Most of the performance specifications that hard disk manufacturers provide to describe their products can be broken down into categories by which aspect of performance they measure. I have designed the section on performance specifications with this in mind: there are sections discussing positioning performance specifications and transfer performance specifications. In addition, there are two key specifications that reflect aspects of both positioning and transfer. There are also some specifications that don't really fit into these categories.

Read vs. Write Performance

- Hard disks can of course both read and write data (which isn't true of all storage devices.) **The performance of a hard disk isn't exactly the same when it is doing a read as when it is doing a write, however.**
- For some performance measurements there is no difference in how the system performs when doing a read or a write; for example, the platters are always spinning at the same speed, and so the latency of the drive doesn't change depending on what the heads are doing. Other measurements though, such as seek time, are different for reads as opposed to writes.
- Almost all performance specifications given for hard disks are based upon how the hard disk performs while reading, not while writing. This is probably because hard disks spend more time reading than writing, and also because **hard disks** are generally *faster* when reading than when writing, so the numbers look better.

Read vs. Write Performance

- Some companies provide **explicit write specifications** in addition to their read specifications, while others do not. The most important specification that differs between reads and writes is seek time--a good rule of thumb is that the average seek time for writes on a hard disk is about 1 millisecond higher (slower) than the specification for reads.
- If a particular hard disk model doesn't mention the numbers you are interested in for writes, and if write performance is particularly important for your application, contact the manufacturer's technical support department. Someone there will know the answer, if you can get a hold of the right person. :^) It may be easier to try downloading the product manual for your model from the manufacturer's web site.

Component vs. System Performance

- As with every component of the PC, the hard disk is but one part of an integrated whole. It is not possible **to measure** the performance of the hard disk **in isolation**, since running any benchmark program involves using the processor, memory and other parts of the PC. The only way to isolate the hard disk would be if you were to use specialized test hardware, connected directly to the hard disk itself, and then you'd have a hard time being sure that the results really related to "real world" PC performance at all.

- Many **benchmarks** are designed to **try to isolate** the hard disk from the rest of the system to test "only" the hard disk. Some are more successful than others in doing this. Unfortunately, many of them don't take all the factors into account and end up (for example) testing the system's hard disk cache instead of the hard disk itself. They are getting smarter over time, however, but still, virtually every hard disk benchmark I have ever seen has allowed the rest of the system to impact the number. You take the hard disk out of a Pentium II 300 PC and put it into a Pentium III 600 system, run the benchmark again, and the score goes up.

Component vs. System Performance

- As with other components, the best way to compare two hard disks is still a comparative benchmark. Set up the system they are intended to be used in, test one with a benchmark that represents the intended use, and then replace it with the second drive, retest and compare. This eliminates much of the "background noise" that is associated with absolute benchmark numbers. Most better review sites do exactly this, maintaining constant "test benches" with hardware that does not change between tests of various drives.

- Another thing that the matter of "component vs. system" means is that there is no way to (legitimately) **compare directly two drives** that run under different interfaces or in different systems entirely. In particular, you can't directly compare the speed of an IDE/ATA drive to that of a SCSI drive and be sure that you are measuring *only the drive*. Some would rightly point out that the distinction isn't really that important however, since the "end result" of system performance is really most important.

Ranking Performance Specifications and Factors

- Amongst those who are concerned with hard disk performance, there has been considerable attention given lately to ranking the various performance specifications and factors by how important they are for a hard disk. The reason for this is usually to assist in choosing hardware: if you are comparing two hard disks that have similar specifications, how do you decide which one is likely to be better? Of course, reading reviews and benchmarks is one part of the answer. It's still useful at times to have a way to rank the various specifications in terms of which is most important to overall performance.

- There's just one little problem with this: it's virtually impossible to do. Which specification is most critical depends entirely on how you use your hard disk and your PC as a whole. People disagree greatly on the matter of which specification or performance factor translates most directly into improved overall performance. There is no one "correct answer".

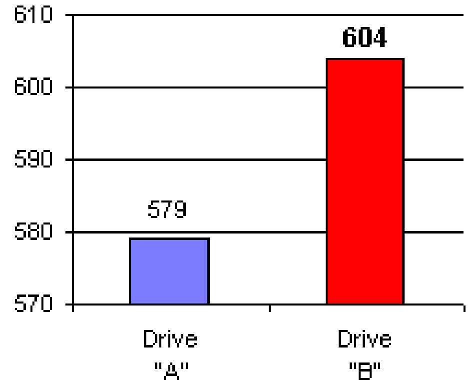
- The biggest argument seems to be over **which is more important**, **positioning** or **transfer**. The answer is: it depends. If you are doing a lot of work with large files, for example editing video or audio streams, transfer is typically more important. Few people fall into this category, however. If you are running a server or other application where many people are accessing smallish files on a regular basis, positioning is definitely more important. Even fewer people fall into this category. :^)

Hard Disk Performance Measurement

- There are many issues involved in measuring the performance of hard disk drives. For starters, there are different ways it can be measured, there are different ways to do it, and important considerations to keep in mind when making the measurements.

- The process of measuring the **performance** of a system or component is called ***benchmarking***. Its intention is to express the overall performance of a piece of hardware in numeric terms. Benchmarking can provide useful data if it is done properly, and kept in perspective.
- Unfortunately, **benchmarking is a process that is greatly prone** to **error** and to **exaggeration**. In order to use benchmarks wisely, one must have a proper understanding of the factors in measuring performance. It is also important to keep in mind that benchmarking is not the only way to look at the matter of performance.

Putting Performance Measurement In Perspective



- Here's an example of the type of benchmark chart you will run into all too frequently At first glance, drive "B" appears to blow the doors off of drive "A", right? In reality, the numbers differ by less than 4%--not even noticeable to most users. The scale has had its zero point "chopped off" to magnify the differences between the numbers. Note also the other marketing tricks: the larger number has its column in bright red to attract the eye; the smaller number is a plain light blue. The larger figure is also in bold and a larger font has been used.Also remember that there are other attributes besides performance that are important for selecting hardware. Lots of people try to get the very fastest hard disk but don't consider other equally important issues: quality, reliability, warranty, and data backup. People agonize over which hard disk is a teeny bit faster than another--and then never defragment their file system, or fill the hard disk up with junk so it runs less efficiently. Be sure to keep the big picture view.

- Finally, bear in mind that whatever is on the top of the hill in the hard disk world doesn't stay there for long. Sure, it's a good feeling to think you are getting *the* fastest disk around. But every few months, a new model comes out that is faster than anything that preceded it. If you really want to always have the best hard disk, you have to keep buying more hardware, which is an expensive proposition that few opt for.

Objective Performance Measurement (Benchmarking)

- There are many different programs used to ways to benchmark hard drives, and they generally fall into the following different categories:

- **1. High-Level (Application-Derived) Benchmarks:** These are programs that use code from popular applications software--usually office applications, web browsers and the like--to simulate the impact of hard disk performance on the use of those applications in the real world.
- The basic idea is to run a suite of tests that are comprised of typical actions that a user would take while using those applications, and time how much time elapses with the hardware in question. Then the hardware is changed and the test run again. This is generally a good concept for benchmarking, but only has relevance if you are actually using the types of applications around which the benchmarks is designed. If you are primarily a gamer for example, what do you care about the performance of spreadsheet software? Also, since the benchmark is running at a high level, there is a lot of room for interference from operating system and file system issues. One of the most common benchmarks of this type is the **ZDNet WinBench** series.

Objective Performance Measurement (Benchmarking

- **2. Low-Level (Synthetic) Benchmarks:** These programs attempt to test the hard disk directly, **isolating it** as much as possible from the rest of the system. They are often called "synthetic" because they don't try to reproduce the access patterns of real applications, instead using artificial patterns created by the programmers specifically for the benchmark, to test different types of hard disk use. They are often derided as being unrealistic because of their synthetic nature, and much of this criticism is in my opinion accurate. At the same time however, they provide much more control over the test process than application-derived benchmarks. This control lets you better "focus in" on one particular aspect of performance and more accurately compare different hardware units in a number of different areas. Common disk benchmarks of this variety include **Adaptec's Threadmark** and **Intel's IOMeter**.

- **3. "Real-World" Benchmarks:** These are not "formally" benchmarks, but are commonly used by hardware enthusiasts to compare real-world performance of hard disks. **The idea is simple**: **take something that you do often, measure how long it takes with one drive, and then how long it takes with another**. For example, if you have a system with an old hard disk that is very slow to boot up, measure how long it takes and then repeat the process with a newer disk to see how much things improve. In some ways these are the most realistic, and also the most *relevant* benchmarks. However, they are entirely system-dependent and therefore of no use whatsoever in communicating much in objective terms about the power of the hardware in question: the improvement you see between hard disk "A" and hard disk "B" on your system may be very different than the same hardware used on a friend's PC. Also, these measurements are usually fairly crude and can't be done on activities that take relatively little time, since the timing is often done with a regular clock or wristwatch.

Objective Performance Measurement (Benchmarking)

- Even leaving the matter of over-emphasizing benchmarks aside, there are some common "**benchmark traps" I** see all the time:
- **Poor Control Of Environmental Factors:** The only way to properly compare two pieces of hardware is to test them under *identical* conditions. Even seemingly irrelevant issues can influence the outcome. Most better hardware sites understand this, but many individual enthusiasts do not. The exact number you get from testing one drive on your system can be very different from the number someone else gets with the same drive, without this meaning anything is "wrong".
- **Small Sample Size:** All benchmarks have a tendency to produce **different numbers** if you run them more than once. To properly use a benchmark it must be run several times and the results averaged. It's even better to run at least five times and discard both the highest and lowest score for each piece of hardware.
- **Paying No Attention To Cost:** You will frequently see people talk about the "benchmark X" score of one drive versus another, but when's the last time you saw anyone take the ratio of two drives' respective benchmarks to their current market prices? I've seen people recommend "drive A" over "drive B" due to a difference in performance of well under 10% despite "drive A" costing 50% more than "drive B". That's rarely money well-spent.
- **Benchmark (In)Validity:** It's not uncommon to see a particular benchmark be used for a long time by many people... and then it is discovered that due to a flaw in how it is written, or the manner in which it interacts with the hardware, operating system or drivers, that its results were inaccurate or misleading. Another reason to use benchmarks only as guidelines.

Subjective Performance Measurement

- The opposite of objective performance measurement is ***subjective* performance measurement.** This technique dispenses with the use of benchmarks in favor of more intangible issues of assessing performance. In a nutshell, this method of measurement is based upon the system's usability or "feel" under different hardware conditions. It's not nearly as commonly discussed as objective measurement, for a couple of obvious reasons. It's much harder to quantify, to get a "handle" on than benchmarking. Subjective measurement also doesn't provide you with neat, easily-compared numbers. It's not really well-suited for testing hard disks in a general way that will be applicable to people reading an analysis or review of a piece of hardware.
- The guiding principle behind subjective performance measurement is this: "If you can't tell the difference, what does it matter?" As such, subjective evaluation is a very personal matter, best suited for an individual to use in assessing the performance of a particular piece of hardware. The best benchmark is always using your own system with your own software. If you change a piece of hardware and it doesn't make a difference big enough to really impact your use of the system, then the change is probably not worthwhile.
- The main problem with subjective measurement is that you can't always easily "test drive" hardware. If you have a slow hard disk and a friend is willing to let you borrow a faster one for a "trial run" then that's great--take advantage of it. Few of us have such a luxury, but fortunately, it's not strictly necessary. With subjective measures we are not dealing with specific numbers but rather "big picture" performance. As such, you can learn vicariously from others' experiences. Let's suppose a friend has a system that is similar to yours and has a similar "feel" when you use it. He upgrades to a new hard disk and suddenly the system feels much faster to you. If this is the case, the chances are good that you will experience a similar improvement in the usability of your own PC if you upgrade.

Hard Disk Performance Specifications

- Due to this "many-to-one and one-to-many" relationship between these two concepts, there is some overlap between the "specifications" and "factors" sections--there are headings and topics that are the same in both places. In this section, I am focused primarily on what the specifications are and what they represent, how to interpret them and what their importance is.

- I then relate each of these to the performance factors that influence them, and talk more about the direct impact on performance in the internal performance factors and external performance factors sections.

Positioning Plus Transfer Performance Specifications

- The two specifications discussed in this section are given the **"privilege"** of being discussed first--like they care, right? :^)--for an important reason. They are the only specs that illustrate aspects of the performance of the hard disk in both fundamental ways: **positioning** and **transfer**.
- As such, they are very important to look for and understand. They certainly are not the only important specifications, and there are some knowledgeable people who rank other metrics higher than these, but few would dispute that they are essential enough that you should always look for them on any hard disk you evaluate.

Spindle Speed

- The hard disk spindle is the shaft upon which the platters are mounted; it is driven by the spindle motor, one of the most important components in the hard disk. Obviously, the faster the motor turns, the faster the platters spin. The spindle speed of the hard disk is always given in RPM (revolutions per minute). Typical speeds of drives today range from **4,200 RPM to 15,000 RPM**, with 5,400 to 10,000 RPM being most common on desktop machines. See this operational discussion of spindle speed for a table of the most common speeds employed today and in the past, and a list of different applications that use them.
- Spindle speed has gone from being one of the least-discussed to one of the most-discussed hard disk specifications in only a few short years. The reason is the creation of increasingly fast spindles. For the first several years that hard disks were used in PCs, they all had the same spindle speed--3,600 RPM--so there was literally nothing to talk about in this regard! Over time, faster drives began to appear on the market, but slowly, and starting with high-end SCSI drives not used in most systems. Once the trend started, however, and the obvious advantages of higher spin speeds became apparent, the trend accelerated. Still, it is only since about 1998 that mainstream IDE/ATA drives have been readily available for the desktop in spindle speeds higher than 5,400 RPM. The most common speeds today are 5,400 and 7,200 RPM, and 10,000 RPM IDE/ATA drives are likely just around the corner (since they are now standard on SCSI with the SCSI high-end moving to 15,000 RPM!)

Areal Density

- *Areal density*, sometimes also (imprecisely) called *bit density* or even just *density*, refers to the amount of data that can be stored in a given amount of hard disk platter space. It is one of the most important indicators of overall hard disk performance, though one that outside the PC enthusiast community is sadly under-discussed.

- **Areal density is a two-dimensional measure** calculated by multiplying two linear measures: recording density (bit density) and track density. The result is measured in bits per square inch (BPSI). Since densities today are in the billions of bits per square inch, the most commonly seen unit is "Gbits/in2". Sometimes the two measures that comprise areal density, are specified separately; other data sheets don't show these components individually. It's much better to be able to evaluate the numbers separately, since they are very different in terms of how they reflect aspects of performance.

- **Areal density is strongly correlated** to the **transfer rate specifications** of a drive. The higher the drive's areal density, in general, the higher its transfer rates will be, however, most of the improvement in transfer rate is due to increases in *bit* density, not track density.

Positioning Performance Specifications

- In this section I take a look at some of the more important hard disk specifications relevant to positioning performance. These include several very well-known and widely quoted specs, along with some specifications that are a bit more obscure but often just as important.
- Note that these specifications are not listed in any particular order. (OK, I *did* put seek time first since it is so important and interesting to talk about. :^))

- **Access Time** =
- **Command Overhead Time**
- **+**
- **Seek Time**
- **+**
- **Settle Time**
- **+**
- **Latency**

Seek Time

- The *seek time of a hard disk measures* the **amount of time required for the read/write heads to move between tracks over the surfaces of the platters**. Seek time is one of the most commonly discussed metrics for hard disks, and it *is* one of the most important positioning performance specifications. However, using this number to compare drives can be somewhat fraught with danger. Alright, that's a bit melodramatic; nobody's going to get hurt or anything. :^) Still, to use seek time properly, we must figure out exactly what it means.
- Switching between tracks requires the **head actuator** to move the head arms physically, which being a mechanical process, takes a specific amount of time. The amount of time required to switch between two tracks depends on the distance between the tracks. However, there is a certain amount of "overhead" involved in track switching, so the relationship is not linear. It does not take double the time to switch from track 1 to track 3 that it does to switch from track 1 to track 2, much as a trip to the drug store 2 miles away does not take double the time of a trip to the grocery store 1 mile away, when you include the overhead of getting into the car, starting it, etc.
- Seek time is **normally expressed in milliseconds** (commonly abbreviated "msec" or "ms"), with average seek times for most modern drives today in a rather tight range of 8 to 10 ms. Of course, in the modern PC, a millisecond is an *enormous* amount of time: your system memory has speed measured in nanoseconds, for example (one million times smaller). A 1 GHz processor can (theoretically) execute over one million instructions in a millisecond!

Seek Time

- At one point many years ago seek times were difficult to use because manufacturers wouldn't agree on a standardized way of reporting them. Today, this has largely been corrected. While seek time is usually given as a single number, in fact there are three different seek time specifications you should examine for a drive, as they represent the drive's performance when doing **different types of seeks**:
- **Average:** As discussed, this is meant to represent an average seek time from one random track (cylinder) to any other. This is the most common seek time metric, and is usually 8 to 10 ms, though older drives had much higher numbers, and top-of-the-line SCSI drives are now down to as low as 4 ms!
- **Track-to-Track:** This is the amount of time that is required to seek between adjacent tracks. This is similar in concept (but not exactly the same as) the **track switch time** and is usually around 1 ms. (Incidentally, getting this figure without at least two significant digits is pretty meaningless; don't accept "1 ms" for an answer, get the number after the decimal point! Otherwise every drive will probably round off to "1 ms".)
- **Full Stroke:** This number is the amount of time to seek the entire width of the disk, from the innermost track to the outermost. This is of course the largest number, typically being in the 15 to 20 ms range. In some ways, combining this number with the average seek time represents the way the drive will behave when it is close to being full.

Seek Time

- To really put seek time in proper context, it should be remembered that it is the largest component of access time, which is the composite metric that best represents positioning performance. However, it is only one component, and there is at one that is of at least equal importance (see the discussion of access time for more on seek time's role in overall positioning performance). Also, bear in mind that seek times are averages that make certain assumptions of how the disk will be used. For example, file system factors will always have an impact on seek performance in the real world.

- A couple of additional caveats on seek times. First, unless you see two numbers, one for read performance and one for write, **seek times always refer to reads**; see here for more details. Ask for the write numbers if you are interested, or you can approximate by **adding 1 ms to the average read numbers**. Second, watch out for "**less than X ms**" specifications. Rather bogus, and fortunately not seen as often as in the past, I interpret "less than X ms" as "X ms" and you generally should do so as well--if the true average were under "X-1", they'd say "less than X-1 ms" instead of "less than X ms".

- **Seek time is almost entirely** a **function of the design and characteristics of the hard disk's actuator assembly**. It is affected slightly by the read/write head design since the size of the heads affects movement speed.

Settle Time

- The ***settle time*** specification (sometimes called *settling time*) refers to the amount of time required, **after** the actuator has moved the head assembly during a **seek**, for the **heads to stabilize** sufficiently for the data to begin to be read.
- Since it is a component of access time and therefore part of the time required to position for reading a random file, I include it here for completeness. However, since settle time is usually so short (**typically less than 0.1 msec**) it is dwarfed in importance by seek time and rotational latency, and differences between drives in this regard are not really significant. Some manufacturers do not even bother to specify settle time, and some just lump it in with seek time.
- Settle time, like seek time, is a function of the drive's actuator characteristics.

Command Overhead Time

- *Command overhead* refers to the time that elapses from when a command is given to the hard disk until something actually starts happening to fulfill the command. In a way, it's sort of like a "**reaction time**" for the disk. Consider when you're driving a car and a streetlight suddenly turns red; your "command overhead" is the time that elapses from when the light changes, until your foot starts to move toward the brake pedal.

- Like settle time, command overhead is a component of access time and thus part of the overall equation of random positioning performance. Also like settle time, it is **generally very small** and not highly variable between drive designs; it is generally around 0.5 ms for pretty much all modern drives and therefore not something that requires a lot of attention.

Latency (rotational)

- The hard disk platters are spinning around at high speed, and the spin speed is not synchronized to the process that moves the read/write heads to the correct cylinder on a random access on the hard disk. Therefore, at the time that the heads arrive at the correct cylinder, the actual sector that is needed may be anywhere. After the actuator assembly has completed its seek to the correct track, the drive must wait for the correct sector to come around to where the read/write heads are located. This time is called *latency*. Latency is directly related to the spindle speed of the drive and such is influenced solely by the drive's spindle characteristics. This operation page discussing spindle speeds also contains information relevant to latency.

- Conceptually, latency is rather simple to understand; it is also easy to calculate. The faster the disk is spinning, the quicker the correct sector will rotate under the heads, and the lower latency will be. Sometimes the sector will be at just the right spot when the seek is completed, and the latency for that access will be close to zero. Sometimes the needed sector will have just passed the head and in this "worst case", a full rotation will be needed before the sector can be read. On average, latency will be half the time it takes for a full rotation of the disk. This table shows the latency for the most common hard disk spindle speeds:

Latency

- This table shows the latency for the most common hard disk spindle speeds:

Spindle Speed (RPM)	Worst-Case Latency (Full Rotation) (ms)	Average Latency (Half Rotation) (ms)
3,600	16.7	8.3
4,200	14.2	7.1
4,500	13.3	6.7
4,900	12.2	6.1
5,200	11.5	5.8
5,400	11.1	5.6
7,200	8.3	4.2
10,000	6.0	3.0
12,000	5.0	2.5
15,000	4.0	2.0

Latency

- The "average" value is almost always the one provided as a specification for the drive; sometimes the "worst case" number is also mentioned. Sometimes latency is not even mentioned specifically at all, but it can always be calculated using this formula:
- **(1 / (SpindleSpeed / 60)) * 0.5 * 1000**
- Which factors down to this much simpler formula:
- **30000 / SpindleSpeed**
- The result is a value in **milliseconds.**

- Again, as with seek times, latency is most relevant only to certain types of accesses.
- For multiple, frequent reads of random sectors on the disk, it is an important performance-limiting factor.
- For reading large continuous blocks of data, latency is a relatively minor factor because it will only happen while waiting to read the first sector of a file.

Access Time

- *Access time* is the metric that represents the composite of all the other specifications reflecting random performance positioning in the hard disk. As such, it is the best figure for assessing overall positioning performance, and you'd expect it to be the specification most used by hard disk manufacturers and enthusiasts alike. Depending on your level of cynicism then, you will either be very surprised, or not surprised much at all, to learn that it is rarely even discussed. :^) Ironically, in the world of CD-ROMs and other optical storage it *is* the figure that is universally used for comparing positioning speed. I am really not sure why this discrepancy exists.
- Perhaps the problem is that access time is really a derived figure, comprised of the other positioning performance specifications. The most common definition is:
- **Access Time** =
- **Command Overhead Time + Seek Time + Settle Time + Latency**
- Unfortunately, this definition is not universal, and is made complicated by the fact that manufacturers refuse to standardize on even what access time's subcomponents mean. Some companies incorporate settle time into seek time, some don't, for example. And to make matters worse, some companies use the term "access time" to mean "seek time"! They really are not the same thing at all.

Access Time

- Seek time and latency are a result of very different drive performance factors--seek time being primarily a matter of the actuator and latency the spindle motor--resulting in the possibility of some drives being better in one area and worse in another. In practice, high-end drives with faster spindles usually have better seek times as well since these drives are targeted to a performance-sensitive market that wouldn't buy a drive with slow seek time.
- Let's compare a high-end, mainstream IDE/ATA drive, the Maxtor DiamondMax Plus 40, to a high-end, mainstream SCSI drive, the IBM Ultrastar 72ZX. (When I say "high end" I mean that the drives are good performers, but neither drive is the fastest in its interface class at the time I write this.) The Maxtor is a 7200 RPM drive with a seek time spec of "< 9.0 ms", which to me means 9 ms. Its sum of its seek time and latency is about 13.2 ms. The IBM is a 10,000 RPM drive with a seek time spec of 5.3 ms. It's sum of seek time and latency is about 8.3 ms. This difference of 5 ms represents an enormous performance difference between these two drives, one that would be readily apparent to any serious user of the two drives.
- As you can see, the Cheetah beats the DiamondMax on both scores, seek time and latency. When comparing drives of a given class, say, IDE/ATA 7200 RPM drives, they will all have the same latency, which means, of course that the only number to differentiate them is seek time. Comparing the Maxtor above to say, the Seagate Barracuda ATA II with its 8.2 ms seek time shows a difference of 0.8 ms, or around 10%. But the proper comparison includes the other components of access time. So the theoretical access time of the Maxtor drive is about 13.7 ms (including 0.5 ms for command overhead) and that of the Seagate Barracuda drive 12.9. The difference now is about 6%. Is that significant? Only you can judge, but you also have to remember that even access time is only one portion of the overall performance picture.

Access Time

- Remember that access time is an average figure, comprised of other averages. In fact, access time on any particular read or write can vary greatly. For an illustration, let's consider the IBM 34GXP drive, look at its minimums and maximums, and see how they translate into access time minimums and maximums:

Attribute	Best-Case Figure (ms)	Worst-Case Figure (ms)
Command Overhead	0.5	0.5
Seek Time	2.2	15.5
Settle Time	<0.1	<0.1
Latency	0.0	8.3
Total	2.8	28.4

- As you can see, there's quite a range! In the real world these extremes will rarely occur, and over time will be "averaged out" anyway, which is the reason that average figures are used. However, it's important to remember that this wide range can occur on any given access, and random perturbations can affect benchmarks and other performance tests.

Transfer Performance Specifications

- Before we look at transfer specifications, we need to have a short word about terminology. :^) Transfer rates are confusing in part because of the phrase "transfer rate" can mean so many different things. Data transfer occurs in two main steps. For a read, data is first read from the disk platters by the heads and transferred to the drive's internal buffer; then it is moved from the buffer, over the interface, to the rest of the system. For a write, the process is reversed.
- The **rate that transfer occurs** within the disk is of course the ***internal* transfer rate;**
- the rate that transfer occurs over the interface is the ***external* transfer rate**. They are usually not the same, and in some cases can differ by an order of magnitude.

- **Internal transfer rates** are further broken down into the **media transfer rate** and the **sustained transfer rate**, and further complicating things is the fact that transfer rates are not constant over the surface of the drive. It sounds impossible to get a handle on, but it's not that bad once you place it all in the proper context and perspective, and that's exactly what we will do in this section.

Internal Media Transfer Rate

- The ***internal media transfer rate*** of a drive (often just called the *media transfer rate* or the *media rate*) **refers to the actual speed that the drive can read bits from the surface of the platter, or write bits to the surface of the platter**. It is normally quoted in units of megabits per second, abbreviated Mbit/sec or Mb/s. Typical values for today's drives are in the hundreds of Mb/s, with a maximum media rate of about **500 Mb/s being high-end at the time of this writing.**

- **Media transfer rate** can be confusing to understand even for the serious hard disk enthusiast; it's equally difficult to describe. :^) For starters, let's explain what it is *not*. It is only related to what is going on inside the hard disk, and therefore has nothing directly to do with the interface transfer rate. It differs from the sustained transfer rate in that it refers ***only* to the speed of reading or writing bits to a *single* track of one surface of the disk. Nothing else is included--no positioning, no track or head switching.** A track holds a relatively small amount of data--under 0.25 MB with current technology. This means that almost no real-world reads or writes occur on a single track except for very short files, and the performance when reading those is primarily limited by positioning, not transfer. The end result of this is that the media transfer rate does not have much relevance to real-world use of a drive. It is primarily a "theoretical" specification that illustrates the state of the drive's technology. It is used almost exclusively for comparing drives against each other. It is also the basis for the calculation of the sustained transfer rate specification.

Internal Media Transfer Rate

- **Media transfer rates are not constant across the entire surface of a platter.** Let's recall for a moment the fact that modern disk drives use zoned bit recording. This is done because the length of the inner tracks on the disk is much shorter than that of the outer tracks. **ZBR allows the outer tracks to have more sectors per track than the inner tracks**. However, since every track is spinning at the same speed, this means that when reading the outer tracks, the disk is transferring more data per second when when reading the inner tracks. For this reason, the media transfer rate decreases as you move from the outer tracks of the disk to the inner ones.
- The explanation above is the reason that there is no single "media transfer rate" figure for a modern hard disk. They are typically stated as a range, from minimum to maximum (with the maximum figure given alone, of course, if only one number is provided). For example, the IBM Deskstar 34GXP (model DPTA-373420) has a media transfer rate of between approximately 171 Mb/s and 284 Mb/s depending where on the disk you are reading: that drive has 12 different zones. This drive has 272 sectors in its innermost zone, and 452 sectors on its outside tracks.
- Another important thing to remember about the media transfer rate (and another reason why it is a theoretical measure only) is that it includes *all* bits read or written to the disk, not just user data. **As discussed in detail here, some of the data storage space in a sector is reserved for overhead**. This means that you cannot assume that the media rate represents the rate at which user data can be read from the disk. Taking the IBM drive above again as an example, its maximum media transfer rate is 284 Mb/s, but the maximum rate that the drive can read user data is about 222 Mb/s in the outside zone.

Internal Media Transfer Rate

- **User Data Transfer Rate =**
- **(Spindle Speed / 60 * Sectors Per Track * 512 * 8) / 1,000,00**
- In Mb/s
- This formula shows the derivation of the 222 Mb/s figure above: use 7200 for the 34GXP's spindle speed, and 452 sectors on its outside tracks. Note that you need the true physical geometry here; the logical BIOS setup parameters will give incorrect results. (If the geometry you are using says the disk has 63 sectors per track and 16 heads, it's almost certain that you are looking at the logical BIOS geometry!) And again, remember that this is not the same as the media transfer rate; to get that figure you'd have to replace the "512" above with the total number of bits, including overhead, contained in each sectors of the disk.
- The media transfer rate of the drive is primarily affected by all of the various data recording and encoding factors, as well as the size of the platters, and the drive's spindle speed. In addition, the drive's controller must be fast enough to be able to handle the fastest rate that the disk can read or write, but manufacturers ensure that this is never an issue by beefing up their controllers where necessary.

Head Switch Time

- **Switching between heads** within a cylinder still requires a certain amount of time, called the head switch time. This is usually less than the track switch time, and is usually on the order of 1 to 2 milliseconds. (Seems kind of slow for an electronic process, doesn't it? The reason is that this time includes all of the overhead of the switch as well; it is all of the time that passes between when the read stops on one head and when it actually starts again on the next one.)
- Head switch time is not commonly discussed, but it is an important component of sustained transfer rate, since STR measures transfer rate over larger reads or writes that encompass more than one track. See the discussion of sustained transfer rate for more details. You may also want to read about head and cylinder skew here.
- Head switch time is primarily influenced by the characteristics of the hard disk's controller. It does not vary greatly from drive model to model or between manufacturers.

Cylinder Switch Time

- Similar in concept to head switch time, cylinder switch time is the time that elapses when the drive finishes reading (or writing) all the data on a given cylinder and **needs to switch to the next one**. This normally only occurs during fairly long reads or writes, since the drive will read all the tracks in a cylinder before switching cylinders. Cylinder switch time is slower than head switch time because it involves a mechanical process: moving the actuator assembly. It is usually somewhere around 2 to 3 milliseconds.
- Note: **You might think** that cylinder switch time would be the same as track-to-track seek time, after all, it's the same thing, isn't it? They aren't the same however, because cylinder switch time includes all of the overhead time that passes from the time the read stops on one track until it starts again on the next one. This is why cylinder switch times are typically double those of track-to-track seeks.

- Cylinder switch time is another specification that is fairly obscure and not commonly discussed, but is an important component of sustained transfer rate, since STR measures transfer rate over larger reads or writes that can encompass more than one cylinder. See the discussion of sustained transfer rate for more details. You may also want to read about head and cylinder skew here.
- Cylinder switch time is influenced by the characteristics of the hard disk's controller as well as its actuator mechanics. It does not vary greatly from drive model to model or between manufacturers.

Internal Sustained Transfer Rate (STR)

- The **media transfer rate** is the **maximum rate** that any particular track of the hard disk can have bits written to it or read from it. However, most transfers from the hard disk involve more than a single track (and the performance of accesses short enough to fit in a single track is typically dominated by positioning concerns more than transfer issues anyway). For real-world transfers of average files, what we are concerned with is the rate at which the drive can transfer data sequentially from multiple tracks and cylinders on the disk. This specification is the drive's sustained transfer rate (sometimes the sequential transfer rate), abbreviated STR.

- **Sustained transfer rates** are most relevant for reflecting the drive's performance when dealing with **largish files**. It is based upon the drive's media transfer rate, but includes the overheads required for head switch time and cylinder switch time. Also, STR is normally measured in bytes, not bits like the media transfer rate, and includes only data, not the overhead portions of each sector or track. An example: let's say we want to read a 4 MB file from a hard disk that has 300 sectors per track in the zone where the file is located; that's about 0.15 MB per track. If the drive has three platters and six surfaces, this means that if this file is stored sequentially, it will on average occupy 26 tracks over some portion of 5 cylinders. Reading this file in its entirety would require (at least) 25 head switches and 4 cylinder switches.

Internal Sustained Transfer Rate (STR)

- STR can be calculated from various characteristics of a disk, but this isn't nearly as conceptually simple as calculating a media transfer rate on a single track. Rather than just provide a lengthy formula, I'll try to explain how the calculation is done. A transfer rate is of course data transferred per unit of time. So our equation will be a ratio of data transferred to the time taken to transfer it. Now, to represent a sustained transfer we need to cover an entire cylinder, so we include all the head switches while reading the cylinder, and one cylinder switch time as well (to get us to the next cylinder). The data that is transferred for an entire cylinder read is as follows:
- **Data transferred per cylinder = Number of surfaces * Sectors per track * 512 bytes**
- where "number of surfaces" is identical to the number of tracks per cylinder, of course. Now, how much time is taken? First, we of course have to wait for the disk to make one complete revolution for each track read, as the data is read. Then we need to add a number of head switches equal to the number of surfaces *less one*, and finally, one cylinder switch. So the time taken to transfer an entire cylinder is as follows:
- **Time per cylinder transfer =**
- **Number of surfaces * Platter revolution time**
- **+ (Number of surfaces - 1) * Head Switch Time**
- **+ Cylinder Switch Time**

Internal Sustained Transfer Rate (STR)

- The easiest way to calculate platter revolution is to double the disk's latency specification. The final equation then looks like this:
- **STR = (Number of surfaces * Sectors per track * 512) /**
- **(2 * Number of surfaces * Latency + (Number of surfaces - 1) * Head Switch Time + Cylinder Switch Time) in MB/s**

- The result is in bytes per second. Simple, right? ;^) Let's use the same IBM Deskstar 34GXP model that we discussed in the media transfer rate section. This drive has 452 sectors in its outermost zone, and a 7200 RPM spin speed (for latency of 4.17 ms). This family's head switch time is 1.5 ms and cylinder switch time is 2.0 ms. We'll consider the flagship drive that has five platters and hence ten surfaces:
- STR = (10 * 452 * 512) / (2 * 10 * 0.00417 + (10 - 1) * 0.0015 + 0.002) = 23,399,798 bytes per second
- The specification for maximum STR for this drive is in fact 23.4 MB/s. Out of curiosity, let's do the same calculation for the 13.6 GB version of this drive, which has only two platters:
- STR = (4 * 452 * 512) / (2 * 4 * 0.00417 + (4 - 1) * 0.0015 + 0.002) = 23,223,683 bytes per second

Internal Sustained Transfer Rate (STR)

- The change is due entirely to the difference between head switch time and cylinder switch time: if they were identical the STRs would be as well. Since the drive with more platters performs a higher ratio of (faster) head switches compared to (slower) cylinder switches, its STR is a bit higher. Still, it's only a difference of less than 1% between the biggest and smallest members of the family.

- An important question to consider is how meaningful the STR numbers really are: if you have the drive above, will it really let you read at a rate of about 23 MB/second? I'm sure you won't be shocked if I say "no". There are a number of issues involved. First, since STR is derived directly from the media transfer rate, its value also depends on what part of the disk is being read; larger outer cylinders have the highest STR, smaller inner cylinders have the lowest. Second, there's the matter of whether the access is *really* sequential. There is a big difference between a 10 MB file that is laid out contiguously on the disk, and one that is fragmented into a dozen pieces. Once you fragment the file, you aren't doing a consecutive data transfer any more. Each fragment of the file introduces the need for an additional positioning step to the location where the next piece starts, which slows the transfer and introduces other factors into the performance measurement. Finally, real-world transfers incur all sorts of penalties due to operating system overhead and other considerations. A good rule of thumb in the computer world is that you never get the theoretical maximum of *anything*. :^)

Internal Sustained Transfer Rate (STR)

- STR has in the last few years started to get more attention than it traditionally has-- some would say too much. :^) It is important to those who do a lot of work with large files, but not as critical to those who work with a large number of smaller files, which includes many, if not most, Windows users. It is probably best to value it roughly equally with key positioning specifications such as access time.

- Sustained transfer rate is affected by just about every internal performance factor you can name. :^) The **number of platters** influences **it by changing the mix of head and cylinder switches; actuator design and controller circuitry affect the switch times**; media issues and spindle speed influence the all-important underlying media transfer rates. There's probably no other performance specification that is affected by so many different design factors.

- A final point about internal sustained transfer rates vs. external (interface) transfer rates. In order to get the most from the hard disk, the interface must be fast enough to be able to handle the maximum STR of the drive. This is usually not a problem because most disk interfaces have sufficient "headroom" to handle drives that run on them. However, many interfaces are also backward-compatible; for example, you can put the drive discussed above on an older IDE/ATA interface running at 16.6 MB/s. It will work, but clearly you will not get STR of 23 MB/s over that interface. The converse is that putting a drive on an interface much faster than it won't improve performance much; but that's a topic for another section.

External (Interface) Transfer Rate

- The internal transfer rate of the drive represents the speed with which bits can be moved to (from) the hard disk platters from (to) the hard disk's integrated controller. The external or interface transfer rate represents the speed which which those bits are moved between the hard disk and the rest of the PC. This is usually faster than the internal rate because it is a purely electronic operation, which is typically much faster than the mechanical operations involved in accessing the physical disk platters themselves. This is in fact a major reason why modern disks have an internal buffer.

- The external transfer rate is unique among hard disk specifications in that it has almost nothing to do with the hard disk itself. The integrated controller must have the right chip to support the interface, but that's about it. The external transfer rate is dictated primarily by the type of interface used, and the mode that the interface operates in. Support for a given mode has two requirements: the drive itself must support it, and the system--usually meaning the system BIOS and chipset, or add-in controller card--must support it as well. Only one or the other does absolutely no good. External transfer rate is affected by a variety of interface issues, discussed in much more detail in the section on external interface performance factors.

- External transfer rate is a perennial candidate for "most overrated hard disk specification". The reason is that external transfer rate specs are usually very high and impressive; manufacturers print them in big bold letters on their retail boxes, and system makers highlight them on their spec sheets. Unfortunately, they usually have very little to do with real-world performance, because the drive's internal characteristics limit transfer performance.

External (Interface) Transfer Rate

- As I've mentioned before, transfer consists of two steps, internal and external. For typical transfers, the net speed of the transfer cannot be any higher than the slower of these two components. Since the external transfer rate of a drive is usually much higher than its internal sustained transfer rate, that means that the STR will be the bottleneck, and thus the factor that limits performance; the high transfer rate of the interface is mostly wasted. As an analogy, suppose you have a 1/2" garden hose connected to a 3/4" pipe. The 1/2" segment will be what limits the flow of water; increasing the 3/4" pipe to 1" or even higher won't make a difference in how much water you get at the end of the pipe.

- There is one occasion where the external transfer rate does come into play: if the data requested by the system is already in the disk's internal cache or buffer. In that case, the data can be sent from the buffer to the rest of the system at the full speed of the interface, whatever that happens to be. Unfortunately, these situations represent such a small percentage of total requests that the net effect of the higher interface speed on overall performance is small. Today's IDE/ATA hard disks are designed to operate with an interface speed of **100 MB/s**, but their sustained transfer rates are barely pushing 40 MB/s. This means the 100 MB/s speed only applies for the occasional transfer that does not require actual access to the hard disk platters.

- There is one area where the interface speed is very important to pay attention to: you do not want it to be too low or performance will suffer. If you take the 3/4" pipe mentioned above and reduce its diameter to 1/4", suddenly it becomes the bottleneck, not the 1/2" diameter hose. If the interface does not have enough speed to allow the hard disk to run at its full STR, then performance can be substantially degraded. Since interfaces are relatively inexpensive this is a situation you generally want to avoid: instead, upgrade the interface. This issue occurs only when putting a new, fast drive into a rather old, slow system.

External (Interface) Transfer Rate

- Hard disk manufacturers always provide lots of "head room" by upping the interface standards in anticipation of advances in sustained transfer rates.
- In 2000 they moved from Ultra ATA/66, which was already sufficiently fast for modern drives, to the 100 MB/s Ultra ATA/100 interface. This despite there being no IDE/ATA drive available that has an STR of even half that figure. It's good to plan for the future; certainly a motherboard supporting a 100 MB/s interface will give you more "room for expansion".
- Just don't think it will be noticeably faster than one that "only" supports 66 MB/s, with today's drives. And also don't forget that by the time drives need that throughput, you may be using a different motherboard or PC altogether.

Other Performance Specifications

- The important matters of positioning and transfer performance are of course the ones that get most of the attention when considering hard disk performance specifications--and rightly so. However, they are not the only specifications that exist for hard disks. There are a few other specs that are routinely found in hard disk data sheets that are indicative of various performance characteristics of the drive, even if they aren't strictly related to the drive's ability to do random accesses or sequential transfers. I take a look at these in this section.

Internal Cache (Buffer) Size

- All modern hard disks have an internal buffer, or cache, that is used as an intermediate repository for data being transferred between the hard disk and the PC. It is described in detail in this operation section. The size of this buffer is usually given as a standard specification on modern drives.
- Having some cache in a drive is somewhat important to overall performance; the drive will use it to buffer recent requests and to "pre-fetch" data likely to be requested by the system in the future. If this data in the cache is in fact needed, it will be transferred to the system at the drive's external transfer rate--much faster than would be possible if there were no cache. However, the number of requests that fit into this category is relatively small. Increasing the size of the cache even by a substantial percentage doesn't change this very much, because no matter how large the cache, it will always be a very small percentage of the total capacity of the drive. Caches today, despite significant increases in size, are still far less than 0.1% of the size of the disk drives they serve.
- As memory prices have fallen into the "dirt cheap" range, drive manufacturers have realized that they can increase the size of their buffers at very little cost. Certainly nothing is lost in doing this; extra cache won't hurt performance; but neither does it greatly improve it. As a result, if interface transfer rate is the "reigning champion" of overrated performance specifications, then cache size is probably the "prime contender". :^) Some people seem to think a 2 MiB buffer makes a drive four times as fast as one with a 512 kiB buffer! In fact, you'd be hard pressed to find even a 4% difference between them in most cases, all else being equal. Not surprisingly, both external transfer rate and cache size are overrated for the same reason: they apply to only a small percentage of transfers.
- The cache size specification is of course a function of the drive's cache characteristics. Unfortunately, manufacturers rarely talk about any characteristics other than the cache's size.

Power Consumption

- There are several reasons why power consumption is an area of concern for PC users. The **first** is that the amount of power needed by the hard disks must be provided for when specifying the power supply (although modern systems with one hard disk don't generally need to worry about this). The **second** is that the start-up power requirements of hard disks exceed their normal requirements and must be given special consideration in systems with multiple storage drives. The **third** is that more power consumption, all else being equal, equates to more heat dissipated by the drive. The final one is environmental: the trend is towards systems that use less power just for the sake of using less power!
- The power consumption specifications provided for a drive vary from manufacturer to manufacturer. Some provide only a "typical" rating for the drive during average conditions, a start-up peak value for the +12 V voltage, and that's it. Others provide a comprehensive look at the drive's use of both +5 V and +12 V power under various conditions. For example, the table below contains the power consumption specifications for the IBM Deskstar 75GXP, four-and five platter models. Note that unlike most hard disk specifications, power consumption normally is higher for drives with more platters even within the same family--since they have more mass to move, more power is required to turn the platters.

Power Consumption

- Many manufacturers just quote an average for the whole family, but IBM generally doesn't:

Operating Condition		+5 V draw (Amps, RMS)	+12 V draw (Amps, RMS)	Power Consumption (W)
Start-Up	Peak	0.81	1.81	--
Random R/W Operation	Peak	1.02	2.23	--
	Average	0.41	0.78	11.5
Seek	Peak	0.47	2.23	--
	Average	0.27	0.84	11.4
Idle	Average	0.24	0.57	8.1
Standby	Average	0.26	0.015	1.5
Sleep	Average	0.17	0.015	1.0

- Examining these numbers reveals a number of facts about how the drive uses power. First, notice that when operating (platters spinning and actuator moving), the +12 V draw is about 0.8 A; when idle (platters spinning but actuator stationary), it is about 0.6 A; and when in standby (platters stationary), +12 V is about zero. This tells you that roughly 3/4 of the +12 V power is taken by the spindle motor and roughly 1/4 by the actuator assembly. +5 V is primarily used to drive the controller's components, which is why even in standby mode a fair percentage of the +5 V power required during operation.is needed. This is typical of modern drives. "Real-world" power consumption will generally be close to what the manufacturer specifies, but bear in mind that actual consumption will depend on a number of factors, most especially the manner in which the drive is used.

Hard Disk Internal Performance Factors

- There are a number of design factors and issues that affect the performance--and hence the performance specifications--of the hard disk. Of these, I refer to performance considerations that relate only or primarily to the capabilities of the hard disk drive itself as internal performance factors. In theory, these are not directly related to the interface or other parts of the system external to the hard disk, which means they should be reasonably consistent and even "portable" from one system to another. These are really the basis of hard disk performance, since they dictate the theoretical maximums; external factors can only further constrain the limits imposed by the design of the hard disk itself.

- This section takes a look at the most important internal performance factors of the modern hard disk. They are divided into three sections, reflecting the different major areas of concern regarding internal performance considerations. First, I take a look at three major design factors related to the mechanics of the drive, which are probably the most important influence on performance. Then, I'll discuss issues related to how data is recorded on the platters. Finally, I'll describe some factors that relate to the drive's integrated controller. For each factor, I will provide a reference to the performance specifications it most directly impacts.

Mechanical Design Factors

- The defining performance-limiting characteristic of hard disks compared to the other main "performance" components is the fact that they operate mechanically.
- Consider the other components of the PC that have a fundamental impact on overall system performance: processors, motherboards, system memory, video cards. They are all solid-state--no moving parts--and as such much faster than the mechanical parts that make up hard disks. That's why **memory access times** are in nanoseconds, for example, while **hard disk access times** are in milliseconds--a **million times slower**!
- Since mechanical parts are largely what limit hard disk performance, that makes them the most important factors affecting performance. This section takes a look at three of these

Size and Number of Platter Surfaces

- The data in the hard disk is stored on the platter surfaces. (The operation and characteristics of the platters and media are described in detail here, including a lot of performance-relevant detail.) The number of platters and the size of the platters themselves vary between different hard disk designs, and have an important impact on performance in several ways.
- First, let's look at **platter size**. As discussed in much detail here, the trend is towards smaller and smaller platter sizes for a number of reasons; two of them being particularly important to performance. The first one is that smaller platters allow the data on the drive to be located physically closer together, so there is less distance for the hard disk actuator to have to move when doing random reads or writes on the disk. This directly improves positioning performance on random accesses. The second is that smaller platters have lower mass and higher rigidity, which enables them to be spun at higher speeds for a given power of spindle motor (or conversely, to use a lower-powered spindle motor for the same spin speed). The main cost of using smaller platters is reduced capacity, but with areal density constantly increasing--thus doubling capacity per square inch every year or two anyway--this is a trade-off more people than ever are willing to make.
- The **number of platters** has a more subtle influence on performance; this is why you will sometimes see small differences in the specifications of drives of different capacity in the same model family. The first impact is a relatively simple: more platters means more weight and thus more for the spindle motor to turn. This generally means that the spin-up speed and power consumption of a drive with four platters will be a little higher than those figures for the same drive with two platters.

Size and Number of Platter Surfaces

- The other impact of the number of platters is a bit more controversial: not everyone agrees on the extent to which these effects exist. All else being equal, a drive with more platters will have slightly better positioning performance and a slightly higher sustained transfer rate than one with fewer platters. If you double the number of data storage surfaces, you can store the same amount of data in (roughly) half as many cylinders; this keeps the data "closer together" physically on the drive, reducing the extent to which the actuator must move when doing seeks. You also replace many cylinder switches with head switches when you have more platters; a **one-platter drive will have a 1:1 ratio of head switches to cylinder switches on a sustained read; a four-platter drive will have a 7:1 ratio**. Head switches are faster than cylinder switches, so this slightly improves STR, though it's certainly not a large effect. I show the difference between drives of the same family in the discussion of the sustained transfer rate specification.
- The size and number of platter surfaces on the drive have an impact on seek time (and hence access time), media transfer rate and sustained transfer rate, spin-up speed, and power consumption. Of course, the basic design of the drive also matches the platter size and number to the power of the spindle motor.

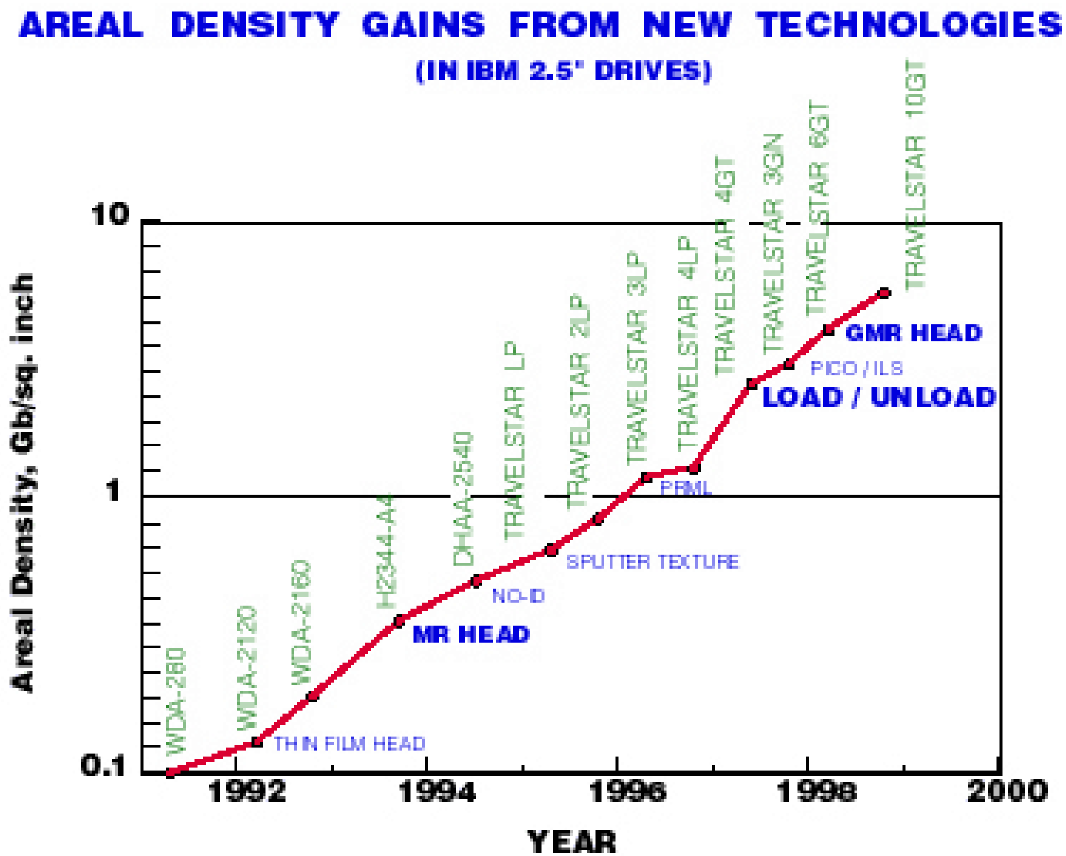
Actuator Characteristics

- The hard disk's actuator--or more correctly, its actuator assembly, comprising the actuator, head arms, head sliders and read/write heads--is one of the most important performance-limiting components in the hard disk. It's also one of the least-discussed and least-understood; I discuss it in detail in this operation section.
- The primary impact that the actuator assembly has is on positioning performance. Since random accesses require the heads to move over the surface of the disk to the correct location, and the actuator controls this process, the actuator assembly is the primary influence on the drive's seek time. Seek time in turn is the largest component of access time. In order to improve performance, manufacturers are constantly striving to reduce seek times by improving the speed of the actuator assembly. The first step taken was to move to voice-coil actuators from stepper-motor designs. Today, improvement is evolutionary, done by increasing the strength of the actuator itself, reducing the weight and size of the actuator arms and sliders, and tweaking other design parameters.
- The characteristics of the actuator also have an impact, albeit relatively minor, on transfer performance. The reason is that the actuator has an impact on cylinder switch time, which is a component of sustained transfer rate. Again, the impact is relatively small compared to the impact of the other factors that influence STR.

Spindle Motor Speed and Power

- The spindle motor is one of the most important components in the hard disk, because its quality and power have a direct impact on many key performance and reliability concerns. It is discussed in detail in this section.
- The drive's spindle speed affects both positioning and transfer performance and is thus one of the most important directly-quoted performance specifications unto itself; it is described in its own specification section. It affects positioning performance because it directly correlates to the latency of the drive (and in fact, is the only factor that affects the latency of regular hard disk drives). Latency, in turn, is an important component of access time, the specification that best correlates to overall positioning performance. Spindle speed affects transfer position because it is related to the drive's media transfer rate, which is the prime transfer performance specification.
- Note: While spindle speed affects the media transfer rate, it is not proportional to it. The reason is that it is more difficult to read and write at very high linear densities when running at very high spindle speeds. This means that in some cases a drive running at 5400 RPM will have a higher areal density than a similar drive running at 7200 RPM. The transfer rate is still normally higher for the 7200 RPM drive, because the spindle speed is 33% higher and the linear areal density is usually only smaller by a factor of 10% or less (though this could change at any time; who knows what those engineers are up to! :^))
- The power of the spindle motor has an impact on the drive's spin-up speed, for obvious reasons. Also, since the spindle motor is the primary consumer of power in the hard disk, its design has the biggest impact on overall power consumption. In some ways, slower drives have an advantage here; it takes longer to spin anything up to 10,000 RPM than to 5400 RPM (unless you use a correspondingly larger motor in the faster drive.)

Data Recording and Encoding Factors



- This chart shows the progress of areal density over the last several years, as refleced in IBM's Travelstar series of 2.5" form factor hard disk drives. Increased areal density leads directly to larger capacity and better performance. Note that the density scale is logarithmic!

Read/Write Head Technology

- The read/write heads actually write and read the data to and from the hard disk, so you'd think they'd have at least some impact on the drive's overall performance! And in fact this is true: they do. Improvements to read/write head technologies, and related components such as the head sliders, are key to allowing increases in linear density, which in turn affects areal density and hence both positioning and transfer performance. They also allow further miniaturization of the head sliders and related components, which indirectly allows faster and more accurate positioning performance.
- However, you rarely hear read/write heads mentioned in discussions of hard disk performance. The reason it isn't often discussed is not because it isn't important, but rather because it doesn't change very frequently. There have been only five general read/write head technologies used in the last 30 years! So in some ways, nobody talks about the heads because there isn't a lot to talk about. :^) Most manufacturers make the change to new head technologies at roughly the same time. Further, many of the manufacturers license head technology from each manufacturers. Today virtually every disk drive sold uses GMR heads; therefore, this is basically assumed and not really a topic for discussion. This equality in basic design generally leaves only minor tweaks to differentiate head designs between models.

Encoding Method

- The encoding method of the disk refers primarily to the algorithm(s) used to change user data into a pattern of bits that can be stored on the platters. This isn't nearly as simple as it sounds; therefore, several different but related methods have been developed over time to facilitate this conversion (and of course, subsequent conversion in the other direction when the data is read back from the disk.)
- Improved encoding methods have one main impact on hard disk performance: they increase recording density, and thus areal density. This improves positioning and transfer performance as described in those sections. Encoding methods are discussed in detail in this section; it contains a relevant comparison between the different methods used over the years, showing how data density has been increased as they have been improved.
- Like hard disk read/write heads, encoding method is rarely discussed in hard disk performance circles because it doesn't change very often, and also because it doesn't vary greatly from one drive to the next. Mostly, its effects are bound up in the areal density specification, as well as the transfer rate specifications.

Track and Sector Layout

- There are several effects on performance that come about as a result of how the data on the surface of the platter is organized. In order to make sense of a platter surface that can store 10 GB of data or more, there has to be a way of organizing it into smaller, more manageable pieces. To accomplish this end, each surface is split into tracks, and then each track is further split into sectors, each sector holding 512 bytes of user data (normally). The track and sector layout of the hard disk, and some of the issues in how this organizing is done, are discussed here.
- The most important impact of the track and sector layout is on sustained transfer rate. The various techniques used in improving the way data is organized on the hard disk all are primarily oriented around increasing the amount of data that can be stored in a given amount of space, and this mainly improves sustained transfer rate. Some advances also improve positioning speed. Here's a quick list of the performance enhancements that fall into this general category:
- Optimal Interleaving: All modern drives use optimal (1:1) interleaving, which cannot be changed. Thus, this factor is not really relevant for distinguishing modern drives. However, you should understand that proper interleaving does ensure that the maximum possible transfer rate is realized on a drive. Older drives that were sometimes set to the wrong interleave factor would have greatly reduced transfer rates compared to their potential maximums.
- Zoned Bit Recording: The use of zoned bit recording has allowed the larger outer tracks of a hard disk to be used to their full potential. It is also the reason that the media transfer rate of a disk depends on what part of the disk is being accessed; since outer tracks have more sectors, they have a higher transfer rate.
- Cylinder and Head Skew: Optimal cylinder and head skew factors built into the drive controller are necessary for high sustained transfer rates; they optimize cylinder switch time and head switch time respectively.
- Sector Format: Improved sector formats and higher sector format efficiency allow a larger percentage of a track to contain data, increasing capacity and transfer rates. Also, the "no ID" sector format improves random positioning by saving time during random seeks, and avoiding "detours" around remapped sectors.

Controller and Cache Factors

- There are a few performance factors that are related to the hard disk's integrated controller, and its internal buffer. Since these parts of the hard disk aren't typically the most important in terms of their influence on overall performance, this discussion is relatively brief. However, these components do have an impact on performance. I also want to point out that some of the aspects of how the cache functions are more important than just its *size*, which it sometimes seems is all that most people focus on today...

Controller Circuitry and Algorithms

- All modern drives include an integrated controller, also called its logic board, that functions as its "nerve center"; it runs the drive, controlling everything that happens within it. It also interfaces the drive to the rest of the PC system. The controller is discussed in this operation section.
- The controller has an important but basically non-measurable impact on the overall performance of the hard disk. In some ways, I liken the controller of the hard disk to the umpires of a baseball game. One way of assessing how good a job the umpires are doing is as follows: if you don't notice that they are there, they are doing a good job; if you do notice them then maybe not. Similarly, the controller's job is to make sure everything happens as it should without drawing undue attention to itself.
- The controller's speed in some ways affects the performance of everything in the drive simply because it manages everything in the drive. However, since it is much faster than the drive's mechanical components, its speed is shadowed by that of the slower components. The controller's design doesn't usually improve performance much, but if it is not designed to meet the needs of the other parts of the system, it can in theory harm performance. For example, when you are doing a sustained read from the drive, the controller cannot make the platters spin faster or affect the disk's recording density. It has no way to "speed up" the transfer, but if it is not designed to have enough capacity to handle the disk's maximum throughput, it can limit performance. Obviously, hard disk manufacturers make sure this doesn't happen.
- Special algorithms within the controller, such as those used to manage multiple commands, can have a direct impact on performance under some types of usage.

Cache Size and Type

- All modern hard disks have an internal cache, that is used as a buffer between the fast PC system and the slow mechanics of the hard disk. It is discussed in some detail here.
- The size of the cache is a commonly-quoted (and often overhyped) performance specification. It of course has some impact on overall performance, but the exact size of the cache is not nearly as important as many people have been led to believe. The impact of the cache is on burst transfers of data over the external interface.
- The type of the cache is rarely mentioned in specification sheets, but its speed has some small impact on performance as well. Since reads and writes depending on the cache are themselves small, differences in cache technologies don't make much real-world performance difference at all. Most disks use for cache whatever current memory technology is mainstream and inexpensive; today, that's SDRAM.

Write Caching Policy

- There's a complication involved in caching write requests to the disk that doesn't exist when doing reads: should the write be cached (put into the disk's buffer) or forced to go through directly to the disk platters? If you don't cache the write you effectively remove the buffering that is in place to isolate the system from the slow mechanics of the drive. If you do cache the write you improve performance, but what happens to the write if the power is cut off to the disk before it can be put out to the platters? This is not a simple question, and doesn't have a simple answer; see here for a full discussion on the subject.
- Most people won't worry about the potential risks of write caching, especially when they find out that it improves performance. :^) That's fine; but you should try to at least find your drive manufacturer's "write policy policy" when shopping for a drive; it's a good thing to know. A lot of drives use write caching without explicitly saying so.
- It should be noted that write caching improves performance pretty much only in a random write environment: writing small blocks to the disk. The cache size is small relative to the size of the disk, so write caching won't improve performance much on a long sequential write. (On a long sustained write the buffer will fill up, and thus force writes to the platters to occur in order to provide space in the cache for later blocks in the sequence.)

Thermal Recalibration

- Thermal recalibration is a procedure that was at one time commonly employed to correct for shifts in the positions of tracks on drive surfaces as they heated up. It is discussed in this section.
- On some older drives, thermal recalibration caused performance problems due to unexpected interruptions for the recalibration to be performed. These primarily affected transfer performance, particularly for users manipulating large files. To combat this problem, manufacturers created special (expensive) drives that did not use thermal recalibration. Fortunately today, this is no longer an issue, since recalibration is not required with today's drives the way it once was.
- The most important thing to remember about thermal recalibration today in terms of performance is: don't pay extra for a drive with a big label on it that says "no thermal recalibration!" :^)

Hard Disk External Performance Factors

- However, there are also ***external performance factors*** to be considered, which relate to how the hard disk relates to other components in the system. Everything from the file system used on the disk, to the disk's interface speed, to the performance of the CPU and other key components, can affect disk performance.

- This section discusses external factors that affect the performance of a hard disk in the "real world" of a system. Note that many of these factors are so external that they don't relate directly to the drive itself, but rather the rest of your system. As such, they often don't correspond directly to any of the hard disk performance specifications. Rather, they influence the real performance you will see when using the drive on a daily basis.

- Since external factors can be different for the same drive in two different systems, this is a good reason to be careful about assuming that a given performance level on one PC will translate directly to another.

Disk Interface Factors

- While overrated in many ways, the speed of the interface is important in the grand scheme of hard disk performance. The interface can form a bottleneck to overall performance if it is too low for the hard disk's maximum sustained transfer rate, and can also have other effects on real-world performance. The issues discussed in this section of course primarily affect the disk's interface speed performance specification.
- The **two most popular hard disk interfaces** used today, by far, are **IDE/ATA and SCSI** (and enhancements of each), so those are the ones that I focus on in this discussion. There is an entire large section devoted discussing interface issues in general, so I won't delve into too much on that subject here. I will focus primarily on how the interface affects performance, but you will still find more information on interface performance effects in the interface section. Also see the comparison of SCSI and IDE for more issues of relevance to interface performance.

Interface Type, Mode and Speed

- The nature of the interface between the hard disk and the rest of the PC system plays a role in its overall performance. The type of interface is the sole determinant of the interface transfer performance specification, and the way the interface is actually implemented has an impact on the real-world performance of the storage subsystem.
- The interface type of almost all drives today is either IDE/ATA or SCSI. Clearly, the decision of which to use involves a much "bigger picture" analysis of your needs and overall system design, since you can't interchange the drives. (I compare the two interfaces here.) However, even within each of the two types there are various modes that control how fast the interface runs. Choosing the best mode for your drive helps to ensure optimal performance. I discuss IDE/ATA modes here, and SCSI modes in this section.
- While new hard disks are generally designed to be able to run at the fastest interface speeds possible for their interface, actually obtaining this interface speed requires an appropriate controller, hardware and/or drivers on the system side. Drives will generally "fall back" to slower interface speeds if required. For example, all new IDE/ATA drives are designed to run in Ultra DMA mode 5, allowing a theoretical maximum transfer rate of 100 MB/s. These drives will function on older interfaces at 66 MB/s, 33 MB/s or 16.7 MB/s if required. Performance will be negatively affected, but only significantly eroded if the speed of the interface falls below the maximum sustained transfer rate specification of the drive.
- High interface speeds also require appropriate support from the system bus upon which the interface runs.

CPU Utilization

- Whenever a hard disk is transferring data over the interface to the rest of the system, it uses some of the system's resources. One of the more critical of these resources is how much CPU time is required for the transfer. This is called the CPU utilization of the transfer. CPU utilization is important because the higher the percentage of the CPU used by the data transfer, the less power the CPU can devote to other tasks. When multitasking, too high a CPU utilization can cause slowdowns in other tasks when doing large data transfers. Of course, if you are only doing a large file copy or similar disk access, then CPU utilization is less important.
- CPU utilization is usually highest when running an IDE/ATA hard disk using a programmed I/O mode, and lower when using a DMA mode. Most newer systems use DMA modes for hard disk data transfer, provided that the appropriate hardware and drivers are installed. SCSI drives use a dedicated controller (host adapter) and typically also have low CPU utilization.
- CPU utilization is one of those performance factors that is both grossly underrated and overrated at the same time. :^) Most people have never even heard of it; it often seems though that a big percentage of those who do understand its role worry about it way too much. :^) Like most performance issues, sweating small differences in numbers is usually pointless; it doesn't matter much if your CPU utilization is 5% or 10%; but if it is 80% or 90% then you are going to see an impact on the usability of the system if you multitask.
- Another key issue is that faster drives transfer more data, and more data--all else being equal--requires more processing time. It's totally invalid to compare the CPU utilization of drives of different generations without correcting for this very important consideration.
- One hard disk utility commonly employed for testing CPU utilization is HD Tach. (Note that this should be considered as information, not an endorsement!)

Command Overhead and Multiple Device Considerations

- As discussed in this section, a certain amount of overhead is required to process any command to the hard disk. However, that's only one type of overhead, the kind within the hard disk involved in doing a random access to the platters. There are other overhead considerations as well that exist within the system itself. These include the time for the system to process the command at a high level, operating system overhead, and so on. Every "piece" of this overhead reduces overall performance by a small amount.
- In comparing the SCSI and IDE/ATA interfaces, command overhead is an important consideration. SCSI is a much more intelligent and capable interface, but it is also more complex, which means more work must be done to set up a transfer. This means that SCSI can be slower than IDE/ATA in a single-user, single-tasking environment, even though it can be much faster and more capable in a machine that is supporting multiple users or multiple devices on the same bus. SCSI shines when you need to use multiple devices on a single bus, where IDE/ATA starts to become cumbersome..
- There is also another consideration: the number of devices that are sharing the interface. This is particularly a concern with SCSI, which allows for many devices on a bus (IDE/ATA and enhancements allow just two per channel). If you are using four hard disks on a SCSI bus in a server that is handling many simultaneous requests, and each drive has an internal sustained transfer rate of 18 MB/s, that 80 MB/s for Ultra2 Wide SCSI will probably, at many points in time, be in full use. On an IDE/ATA machine only one device can use any given channel at a time, so you only need to compare the speed of the interface to the speed of each drive that will use it, not the sum of their transfer rates.

PC System Factors

- As I probably reiterate in too many places on this site, the components in a PC system are interrelated, and affect each other's performance in many ways. This makes it difficult to measure the performance of any component in isolation. Some tests are better able than others to isolate the component being tested, but it also depends on the component. Hard disks are virtually impossible to completely isolate from the rest of the system, because every access of the hard disk involves a transfer through the main processing subsystems, and thus involves almost all of the main components of the PC.
- Various parts of the PC not only affect the way a hard disk benchmarks; they also affect the real-world usability of the storage subsystem as a whole. In this section I take a look at some key issues related to the PC system as a whole and how they influence hard disk performance.

CPU Speed

- Since the system processor (CPU) is involved in everything that happens in the PC, it of course has an impact on hard disk performance. However, this impact isn't nearly as great as you might think. The reason is that the CPU is so much faster than anything related to the hard disk, particularly today, that it normally spends a great deal of time waiting for the hard disk. There's virtually no way for the CPU to really affect the performance of the storage subsystem all that much if the hard disk and system are properly set up. On older systems though, the performance of the CPU was sometimes a real factor in hard disk performance. And even today, if an interface mode is used that results in high CPU utilization by the hard disk interface, overall performance can become somewhat dependent on the performance level of the CPU itself.
- The real issues with CPUs and hard disks are related to benchmarks. Every benchmark that you run on your hard disk involves instructions that are processed on the main system CPU. A faster CPU will run these benchmarks faster than a slower one, and I have found it to be quite consistent that testing the same drive on a machine with a much faster CPU, will result in higher scores. This is accentuated when using one of the transfer modes that requires intervention from the processor, such as programmed I/O.

Interface Bus Speed

- Every hard disk read or write involves a sequence of data transfers. Looking at a read: first the data is retrieved from the hard disk platters. It is transferred to the drive's internal cache by its controller. Then the data is sent over the interface cable to the interface controller on the PC system. That controller resides on a system bus, and uses the system bus to communicate the data to the CPU and the rest of the PC. System buses are discussed in detail in their own section.
- Normally the speed of the bus used for the hard disk interface is not something that you really need to be concerned with. Virtually all systems today use the PCI bus for interfacing to their hard disks and other storage devices, which is fast enough to handle even the high interface transfer rates of modern drives. (Even if your IDE cables plug into the motherboard directly, they are still going to an IDE controller chip that logically "resides" on the PCI bus.) However, as the interface transfer rate of IDE/ATA drives (maximum 100 MB/s) now approaches the limits of the PCI bus (about 127 MB/s), at some point this will become an issue even on new systems; probably within two or three years. Hard disks continue to get faster and faster.
- On older systems interface bus speed limits can become a real issue. There are still systems around that use ISA bus hard disk controllers, for example. Even if you could get one of these older cards to work with a large, modern drive, the slow speed of the ISA bus would drag it down. ISA is limited to a maximum bus bandwidth of under 16 MB/s, easily exceeded even for sustained transfers by most any modern drive, not to mention burst transfers!
- Even on new systems, alternative means of interfacing hard disks can have a major impact on performance. The widespread adoption of the universal serial bus (USB) standard has been a boon for portability of devices and easy interchanging of hardware. Some companies are now even offering USB-based hard disks. These are convenient, but the slow speed of the USB interface--which was designed for slow items like scanners and keyboards, not hard disks--effectively cripples these drives, limiting them to a maximum transfer rate of about 1 MB/s. That may be OK for moving data between PCs, archiving seldom-needed data or doing smallish backups, but it's very slow for just about anything else!

System BIOS Issues

- The system BIOS is the set of core routines that provides the primary interface between the main hardware of the system and the software that runs upon it. It plays a critical role in the functioning of any PC; see here for a full section covering it.
- The BIOS affects hard disk performance in two distinct ways. The first is that the BIOS itself was traditionally used for access to hard disks, and thus the BIOS's routines had an impact on overall performance. Most of today's operating systems now "bypass" the BIOS to access the hard disks directly, reducing this influence greatly. See this section for a discussion of how the BIOS relates to the hard disk in a general way.
- The second is related to the way most systems are designed. In a typical "regular" PC, the motherboard contains an integrated IDE/ATA system controller. Since it is part of the motherboard, it is configured and controlled using code in the system BIOS. This means that the BIOS must provide support for increased capacity when larger drives come out--to avoid BIOS **capacity barrier problems-**-and also support for performance-enhancing features like higher-speed transfer modes, block mode, etc. If your system is a few years old, its BIOS may need to be updated or you could find the performance of newer hard drives restricted.

Operating System and Controller Disk Caching

- The process of caching describes the use of buffers to separate operations that differ significantly in speed, so the fast one is not held up by the slower one (or at least, not as much). See here for more general details on caching and how it works. In a system there are many levels of caching that are used to allow different-speed components to run unimpeded; in the disk subsystem there are usually two levels.
- The disk drive's logic board contains an integral cache. This cache is used to separate the internal mechanical read/write operations from transfers over the bus, and to hold recently accessed data. A larger cache will result in improved performance by cutting down on the required number of physical seeks and transfers on the platters themselves. Smarter caching algorithms can have the same effect.
- In addition to this hardware caching, most operating systems use software disk caching. Since the system memory is many orders of magnitude faster than the hard disk, a small area of system memory (usually a few megabytes) is set aside to buffer requests to the hard disk. When the disk is read, the data are stored in this cache in case they are needed again in the near future (which they often are). If they are needed again, they can be supplied from the cache memory instead of requiring another read of the hard disk.
- As with the disk's internal buffer, increasing the size of the cache improves performance--to a point. If you increase it too much, your operating system will run out of usable memory for programs and data, and the system will be forced to rely on much slower virtual memory. In this case your use of memory as virtual disk is causing the system to also need to use your disk as virtual memory, defeating your original intent!
- In addition to the two typical caches in the hard disk subsystem, some SCSI host adapters add a third level of cache on the controller itself; these are sometimes called caching controllers. This cache can be several megabytes in size and logically sits between any system disk cache and any buffer on the hard disk. Again, this improves performance by reducing the number of accesses required to the disk. In this case when the system tries to read data from the hard disk, the controller will intercept the request and if it is in the cache, satisfy it from there instead of going to the hard disk. This both improves system speed greatly and also cuts down on traffic on the SCSI bus.

Redundant Arrays of Inexpensive Disks (RAID)

- Many higher-end systems, especially servers, now employ a technology called *redundant arrays of inexpensive disks*, or *RAID*. This concept allows for great improvements in both reliability and performance. The idea is to store data on multiple disk drives running in parallel. The primary motivation in many cases is reliability. From a performance standpoint, most RAID levels improve performance by allowing multiple accesses to happen simultaneously, and also by using algorithms that reduce seek time and latency by taking advantage of having multiple drives at their disposal. The exact performance impact depends entirely on the level of RAID used; some improve read performance at the expense of write performance, for example.
- I have written an entire section on RAID that covers its issues, levels and implementation in some detail, including a discussion of its impact on storage subsystem performance. Once used almost exclusively in a corporate setting for large, expensive machines, a new crop of inexpensive RAID controllers and hard disks is bringing RAID into the "mainstream", and small RAID arrays are now commonly seen in the machines of "power users".

File System Factors

- There are several factors related to how the disk is logically structured and the file system set up and maintained, that can have a tangible effect on performance.
- These are basically independent of the hard disk and will have similar impacts on any hard disk. They serve almost solely to influence the real-world performance of the hard disk.
- Better decisions about how you manage your hard disk's file system can translate directly into better performance on *any* PC.

File System Type

- The file system refers to the structures that are used to organize data at a high level on the disk; the file system is used by the operating system to store files, directories and other relevant information on the disk for later retrieval. As such, the file system is highly operating-system-dependent. In most cases you don't generally have a "choice" between different file system types. However, in some operating systems you do, and there can be a performance impact from the choice.
- Some file systems store files in packages as small as 512 bytes, while others store files in larger chunks called *allocation units* or *clusters*. Some are very simple file systems with few features and little overhead (such as the FAT file system used in DOS and Windows 9x), and others have many features but comparatively higher overhead (NTFS used in NT). Windows NT and 2000 typically give you your choice of file system; Windows 2000 supports FAT16, FAT32 and NTFS. See here for more on the different file systems used in the PC.
- Which file system you use can have an effect on overall performance, but it is relatively small: typically a few percentage points. It's also difficult to predict exactly what the effect will be for a given system when selecting from one file system to another. Since the file system affects so many other usability factors of the PC, performance is usually *not* one of the primary factors for deciding between them. As an example, consider the "FAT vs. NTFS" decision, which is probably the most common "file system decision" in the PC world today. These two file systems are so different in so many ways that most people choose one or the other for reasons particular to their use, not performance. If you need the high security and advanced management features of NTFS, you are probably going to use NTFS even if FAT is a few percentage points "faster". Similarly, if you need the compatibility and simplicity of FAT, changing to NTFS for a few ticks on a benchmark is probably unwise.
- One file system choice that *is* commonly made in part for performance reasons is "FAT16 vs. FAT32"; this is really a "sub-file-system" choice, since FAT16 and FAT32 are really two flavors of the same file system. The primary performance impact of changing between these has nothing to do with anything inherently different between FAT16 or FAT32, but rather the difference in cluster size that results from the choice. See here for more details on this.

Partitioning and Volume Position

- Partitioning is the process of dividing the hard disk into subsections, called *volumes*. It is an important initial step in preparing a hard disk for useThe choice of how the hard disk is partitioned can have a *tangible* impact on real-world performance. This is due to several different but related effects that you should keep in mind when deciding how to partition your drive:
- **Cluster Size:** The way that the hard disk is partitioned in most cases determines the cluster size of the partition, which has a performance impact.
- **Zone Effects:** Modern hard disks use zoned bit recording to allow more data to be stored on the outer tracks of the hard disk than the inner ones. This directly impacts the media transfer rate of the disk when reading one zone of the disk as opposed to another; see here for details. Hard disks fill their space starting from the outer tracks and working inward. This means that if you split a hard disk into three partitions of equal size, the first partition will have the highest transfer rate, the second will be lower, and the third lower still. Therefore, you can put the more important files on the faster partitions if transfer performance is important to you.
- **Seek Confinement:** Seek times are roughly proportional to the linear distance across the face of the platter surfaces that the actuator must move the read/write heads. Using platters of smaller diameter improves seek time, all else being equal, and partitioning can have the same net effect. If you split a drive into multiple partitions, you restrict the read/write heads to a subsection of the physical disk when seeking, as long as you stay within the same partition. The tradeoff is that if you do a lot of moving data *between* partitions, or accessing multiple partitions simultaneously, you'll force the heads to "jump" back and forth between two completely different areas of the disk, reducing performance. Some who truly desire performance over all else will buy a hard disk with double the capacity that they need, partition it in two pieces and use only the first half! Or use the second half only for archiving infrequently-used data.
- **Defragmentation Time:** Larger partitions tend to become full of, well, more data, obviously. :^) A larger partition can take much longer to defragment than a smaller one. Since fragmentation reduces performance, some people prefer to partition their drives to reduce defragmentation time, enabling them to do it more frequently.

Cluster Size

- The FAT file system used by DOS and Windows divides all of the file data on the hard disk into clusters comprised of multiple sectors. A cluster is normally between 2 kiB and 32 kiB in size, in powers of two, containing between 4 and 64 sectors of user data. This is done to make managing the location of data easier. Clusters and related file system structures are discussed here.
- The choice of cluster size has an impact on real-world performance, though for most people it is not all *that* significant. In a nutshell, **larger clusters waste more space due to slack** but generally provide for slightly better performance because there will be less fragmentation and more of the file will be in consecutive blocks. This occurs because when clusters are larger, fewer of them are needed than when they are small. A 10,000 byte file would require three 4 kiB clusters but only one 16 kiB cluster. This means this file will always be in a contiguous block if stored in a 16 kiB cluster, but could be fragmented if stored in a 4 kiB cluster size partition. (The slack tradeoff is a waste of 4 kiB more storage in the case of the 16 kiB clusters.) Small cluster sizes also have a negative effect on partition because they require larger file allocation tables, to manage their much larger numbers of clusters. These tradeoffs are discussed in detail here.
- Traditionally, most people have tried to use cluster sizes as small as possible in order to reduce slack and make more efficient use of disk space. This is of course a valid goal, but it has become increasingly irrelevant today as hard disks approach truly gargantuan sizes and the price per GB of storage drops to amazingly low levels. Today, the large file allocation tables resulting from enormous FAT32 partitions means that balancing slack reduction with performance effects is also important, unless you are on a very tight budget. I certainly can't recommend forcing Windows to use 4 kiB clusters on a 30 GB partition "to reduce slack" as some people do, because I personally wouldn't want to take the performance hit of having 30 MiB file allocation tables--and I wouldn't want to have to wait for that puppy to defragment either! :^)

Volume Free Space

- A relevant performance consideration that most people don't pay attention to is how full their hard disk is. The amount of free space on a hard disk affects the performance of the drive, for most of the same reasons that partitioning affects it:
- The more data on the volume, the more the data is spread out over the disk, reducing positioning performance;
- A disk that is more full forces new files to the inner tracks of the disk where transfer performance is reduced; and
- Drives that are full both take longer to defragment and tend to be more fragmented in the first place.
- The "Peter Principle" of hard disks is that the amount of junk put on a hard disk expands to fill the available space, regardless of the size of the hard disk. Imagine what PC users 10 years ago would have thought about people with 6 GB hard disks needing an upgrade because their "disk is too full!" I had the personal experience the other day of surprisingly discovering that a 15 GB hard disk volume I had just installed was down to 2 GB free! Most people don't clean out their disks until they have to. :^)
- The bottom line though is clear: the more "stuff" you put on that humongous hard disk you just bought, the more you will slow it down. :^) Don't fill your drive with clutter just because you have the space. Regularly go through your hard disk to get rid of files you don't need; if you think you will need them "at some point" then archive them to a tape, CD-R disk or other removable medium.
- Another impact of this is that you cannot reliably compare performance benchmarks even on the same disk in the same system if you change the amount of data on the drive between runs. All drives will tend to show a reduction in performance as you fill them up, so benchmarking should always be done on empty drives to eliminate this variable.

Fragmentation

- ***Fragmentation*** refers to the tendency of files stored in the FAT file system to become broken into pieces that can end up in very different places within a hard disk volume. How fragmentation occurs is explained in detail here.
- A fragmented file system leads to performance degradation. Instead of a file being in one continuous "chunk" on the disk, it is split into many pieces, which can be located anywhere on the disk. Doing this introduces additional positioning tasks into what should be a sequential read operation, often greatly reducing speed. For example, consider a 100,000 byte file on a volume using 8,192 byte clusters; this file would require 13 clusters. If these clusters are contiguous then to read this file requires one positioning task and one sequential read of 100,000 bytes. If the 13 clusters are broken into four fragments, then three additional accesses are required to read the file, which could easily double the amount of time taken to get to all the data.
- **Defragmenting** a very fragmented hard disk will often result in tangible improvements in the "feel" of the disk. To avoid excessive fragmentation, defragment on a regular basis; usually once every week or two is sufficient. See the system care guide for more.

Disk Compression

- Compression is a technique that is used to reduce the amount of space required by files on a hard disk volume. Very popular during the late 1980s and early 1990s, compression is rarely used today on new systems. (I am speaking of *volume* compression here, not file-by-file compression such as ZIP files, which are as popular as ever!) Compression is discussed in detail in this section.
- Most people have been told (repeatedly) to avoid compression because it greatly reduces performance. This is true to some extent, but the picture isn't nearly as simple as this blanket statement might suggest. Compression adds overhead to every disk access but reduces the amount data that needs to be retrieved from the platters for a given file. With modern high-speed CPUs, the overhead isn't nearly the issue it once was, while hard disks haven't improved in performance by nearly the same percentage. Therefore, there is a performance tradeoff at work here.
- The proliferation of huge, cheap hard disks has made volume disk compression largely irrelevant today. I don't recommend it except for use on older systems that cannot be upgraded to a larger disk for one reason or another. The reason for this is not primarily due to performance effects, but rather because compression complicates disk usage and simply isn't needed in an era where hard disk space costs half a penny per megabyte.