

Chapter 9:
Peripheral Devices

By:

Derek Hildreth
Chad Davis

Brigham Young University - Idaho
CompE 324 Brother Fisher

Introduction

When discussing this chapter, it has been assumed that the reader has already gone through Chapter 8: Input and Output, where they learned about how the CPU interacts or interfaces with external peripheral devices. In this chapter, the reader will learn about the most common types of peripherals: hard disk drives, video displays, printers, input devices, and analog devices. Most of the emphasis is placed on the logical behavior of these devices, with some emphasis on their physical operation.

There is quite a bit of variation among these devices when discussing how they interact with the CPU and the bandwidth they require. Keyboard for instance, will only use a few bytes per second while the printer will use a few hundred bytes per second. Disk drives and video interfaces can really place a lot of demand on a CPU because some can operate in the tens of megabytes per second (mega=million). These devices generally use a special purpose processor that is built into the devices to off-load much of that demand. Peripheral bandwidth is something that the computer engineer might want to keep in mind when designing a computer.

Magnetic Disk Drives

Perhaps the most complicated and interesting peripheral of the three mentioned above is the modern hard drive. This section will discuss how the hard drive (HDD) is generally organized, then the different interfaces, followed by factors that influence its static (capacity) and dynamic (bandwidth and latency) attributes.

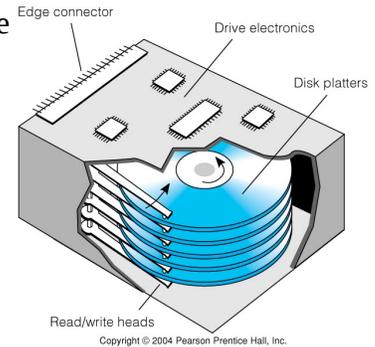
The organization of a HDD can be broken down in a series of steps starting from the outside, or physical dimensions, and then eventually breaking it down into individual bits.

Generally, a HDD comes in two flavors: one for desktop computers, and one for laptop (or notebook) computers. Desktop HDDs are 3.5" wide and notebook



computers are smaller at about 2.5” wide. And depending on how old the HDD is, 5.5” wide was another common size which is not in use anymore. Inside each of these different sized housings are platters, read/write heads, and drive electronics.

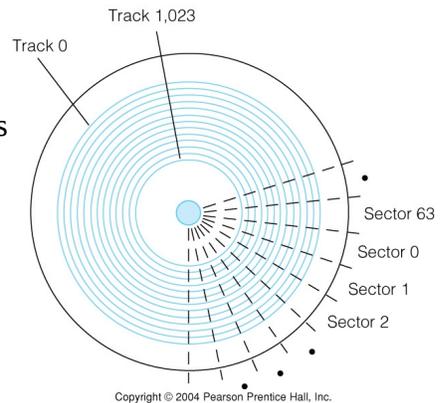
The platters are simply the disks that hold data by a magnetic charge. Typically, there are about four platters. These platters all spin together at a constant speed, called the rotational speed. Rotational speed is a dynamic characteristic of a HDD and is commonly equal to 7,500 RPMS, although there are some drives that can reach speeds of 15,000 RPMS. As the platters rotate, read/write heads have a stationary pivot and they “pick up” or “drop off” data onto the specific area of the HDD. There is a read/write head to each surface of the platters. Generally, only one head is active at a time and the data stream is always bit-serial to and from the surface. The platters can traverse the active area of the disk in around 17 ms.



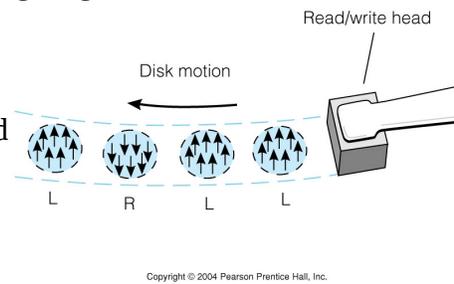
The drive electronics play a very important roll in interfacing the HDD with the CPU. To the CPU, they provide an interface to a standard serial bus such as SCSI (small computer systems interface), PATA (parallel advanced technology attachment), or SATA (serial advanced technology attachment). There are many more serial buses, but what was mentioned above is most commonly used. This allows the CPU to treat the HDD as another memory device. It's important to note that a HDD is *much* slower than memory found in the computer such as SDRAM. Following a read request, the drive must seek out and find the block requested, stream it off the surface, error-check and possibly error-correct it on the fly, assemble it into bytes, store it in an on-board buffer, and signal the processor that the task is complete. This is a lot of work, so a special purpose disk controller assists the drive controller. These components are attached to the drive, and I/O is by means of an edge connector.

When a block of data is requested, the HDD must know where it is located, so the surface of the HDD is broken up into what are called tracks, sectors, and cylinders. A track goes around the surfaces

circumference like you would expect to see on a high school runner track. These tracks are divided up again by means of a vertical line and the result is called a sector. The figure here shows a disk with 1,024 tracks and 63 sectors. The head can determine which track it is on by counting tracks from a known location, and sector identities are encoded in a header written on the disk at the front of each sector. The number of bytes per sector is fixed for a given disk drive, varying in size from 512 bytes to 2 KB. A cylinder is formed when you take all the tracks with the same number but on different surfaces. For the example shown here, with five platters, there would be 1,024 cylinders, each of which has 10 tracks.



Information is recorded to the drive 1 bit at a time by using magnetism. The write head will magnetize a small area on the track and that area is recognized as a bit as it passes under the read head. This is easily illustrated in the figure shown to the right. A brand new HDD from the factory has been initialized or formatted, meaning the drive has had the original track and sector information on the surfaces and has been made sure that it can be written and read from each sector.



The static property of a HDD is related to the drive capacity and unrelated to access speed. Drive capacity can be measured in several different ways including areal density (never realized in practice), and formatted density (actual storage capacity). Areal density is the bit density in bits per

unit area computed by $\frac{1}{(\textit{bit spacing} \times \textit{track spacing})}$. Formatted capacity is computed by

$(\frac{\textit{bytes}}{\textit{sector}}) * (\frac{\textit{sectors}}{\textit{track}}) * (\frac{\textit{track}}{\textit{surface}}) * (\textit{no. of surfaces})$. This is what manufactures use to sell HDDs to

consumers.

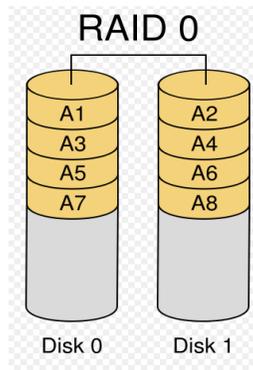
Dynamic properties include seek times, track-to-track time, rotational latency, access time (seek time plus rotational latency), burst rate, sustained rate, and head recalibration. The most important properties of all these, or the ones most realized in practice, are rotational latency and sustained rate. Rotational latency is the average time required for the needed sector to pass under the head once the head has been positioned at the correct track. Today, rotational latencies are about 4.6ms. Sustained rate is the rate at which data can be accessed over a sustained period of time. This is important when designing systems that must access large files in real time like video or audio data.

Improving Disk System Performance and Reliability

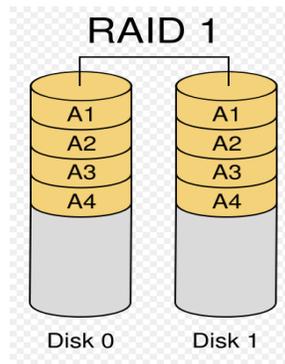
As the size of hard drives continues increasing, the problem of backing up important data becomes increasingly difficult. In the early days of hard drives, a single cdrom could easily hold more information than the entire drive, but with modern drives holding in the hundreds of gigabytes of information, even the densest dvd can only hold a small fraction of the hard drives capacity.

In part because of this problem, a great deal of effort has gone into finding methods to improve the reliability and performance of the hard drive itself. Two powerful methods have come out of this effort. The first is called Redundant Array of Inexpensive Disk, otherwise referred to as RAID. The second is known as Self Monitoring and Reporting Technology, also called SMART.

RAID is a very powerful way of protecting information through the use of multiple hard drives. There are several different setups that can be used, we will discuss the 5 most common here. The common theme between all RAID setups is that they spread data across multiple drives.



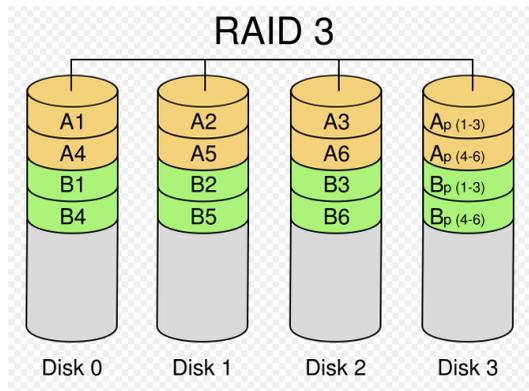
RAID 0 is simply tying multiple drives together without any extra error checking involved. The only advantage of doing this is that you gain a lot of extra speed in read and write time, but if any of the drives fails, you lose all of your data.



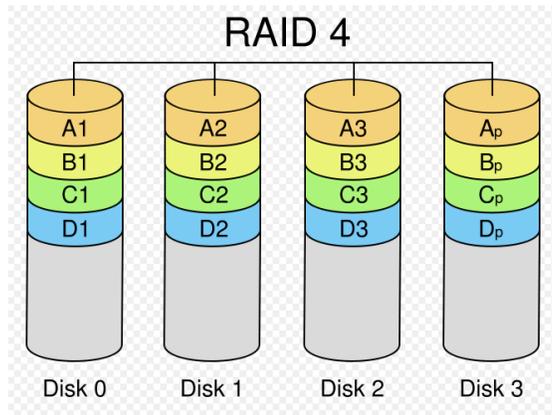
RAID 1 uses exactly 2 drives, but instead of spreading the same data across each, anytime data is written to the hard drive, it writes the exact same data to both drives. This effectively provides an easy backup drive should either drive fail.

RAID 2

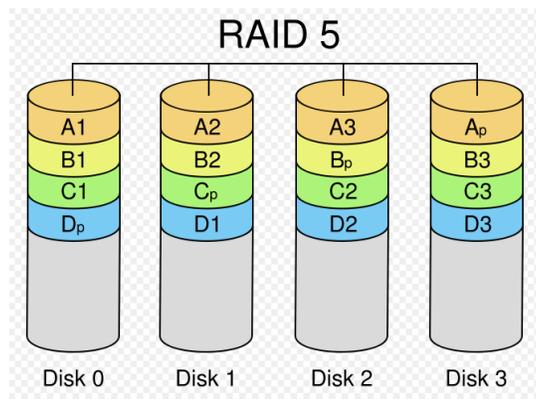
RAID 2 uses Hamming encoding to encode the parity bits, and dedicates some of the drives to store just the error checking and correction (ECC) bits. It implements the striping (spreading of information across multiple disks) at the bit level.



RAID 3 stripes the data across multiple disks similar to RAID 0, except at the byte level. It also dedicates one drive to store the parity bits on. This results in the same high read rates as RAID 0, but slower write speeds because of the need to write the parity bits.



RAID 4 is the same as RAID 3 in all aspects except one: its hard drives are set up to act independently of each other. This allows the drives to handle multiple different I/O requests at the same time, and will perform very well where that functionality is needed.



RAID 5 is again similar to RAID 4, except that instead of using a single drive for parity, parity bits are spread across all of the drives. The parity bits are also always written to a drive other than the one containing the information to which they pertain.

SMART is actually an idea that is incorporated into the hardware of the disk drive. It records error rates as the drive reads and writes information for many different categories related to the drives "health", and is designed to issue a warning if any of those error rates go above some threshold value. This allows the drive to warn users when the drive may be nearing failure, but the information is still easily retrievable.

Other Mass Storage Devices

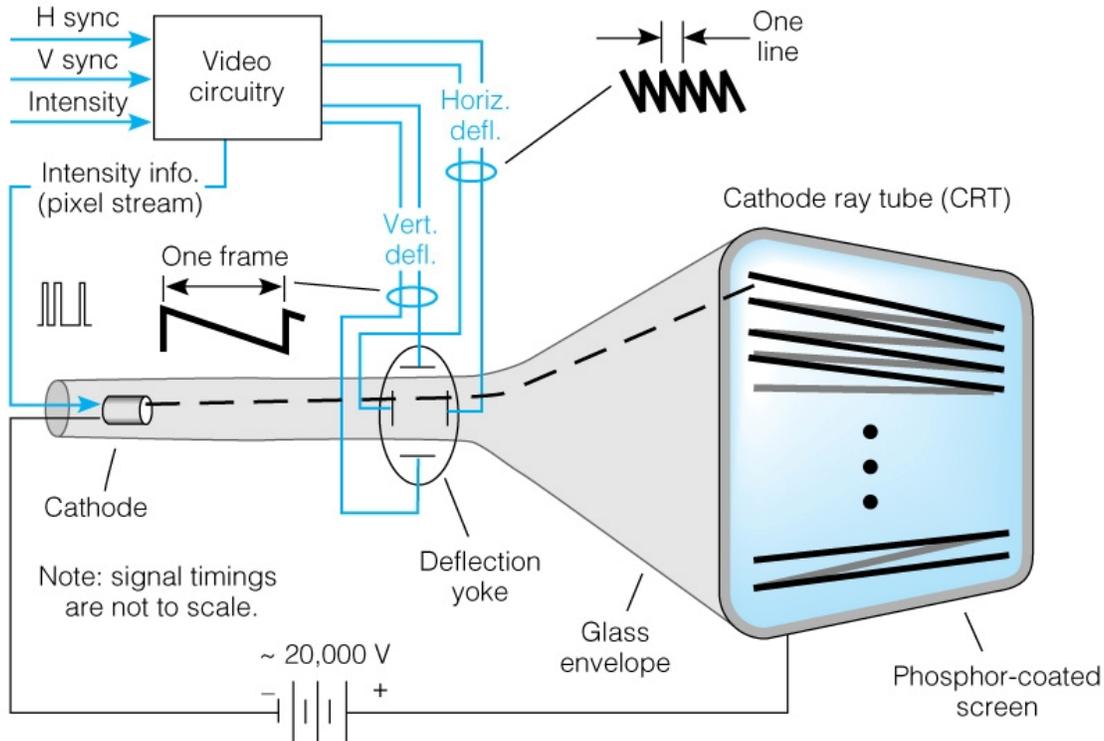
Hard drives aren't the only form of data storage. There are many types of storage devices ranging from CDs to Tape Drives, each having the same basic principle of a HDD (that is, the disk spins and a stationary head read or writes the data). Magnetic tape gives the user plenty of backup space (80GB), but due to its nature, it's incredibly slow to retrieve data from it. The CD and CD ROM are much faster than a magnetic tape, but slower than a HDD, and they can fit about 700MB worth of data. They differ from HDDs because instead of using magnetic areas on a surface, they actually use pits that have been 'burned' into the disk. Other forms of mass storage devices include DVDs and Flash Memory (which doesn't use any moving parts).

Display Devices

Video monitors are the most common display device used in computers today, and are what we will be discussing here. They actually originated from TV monitors, and the only significant difference between the two even today is that a video monitor can display much higher resolutions than a TV monitor.

The first monitors were Cathode Ray Tube (CRT) monitors, and worked by sending a stream of

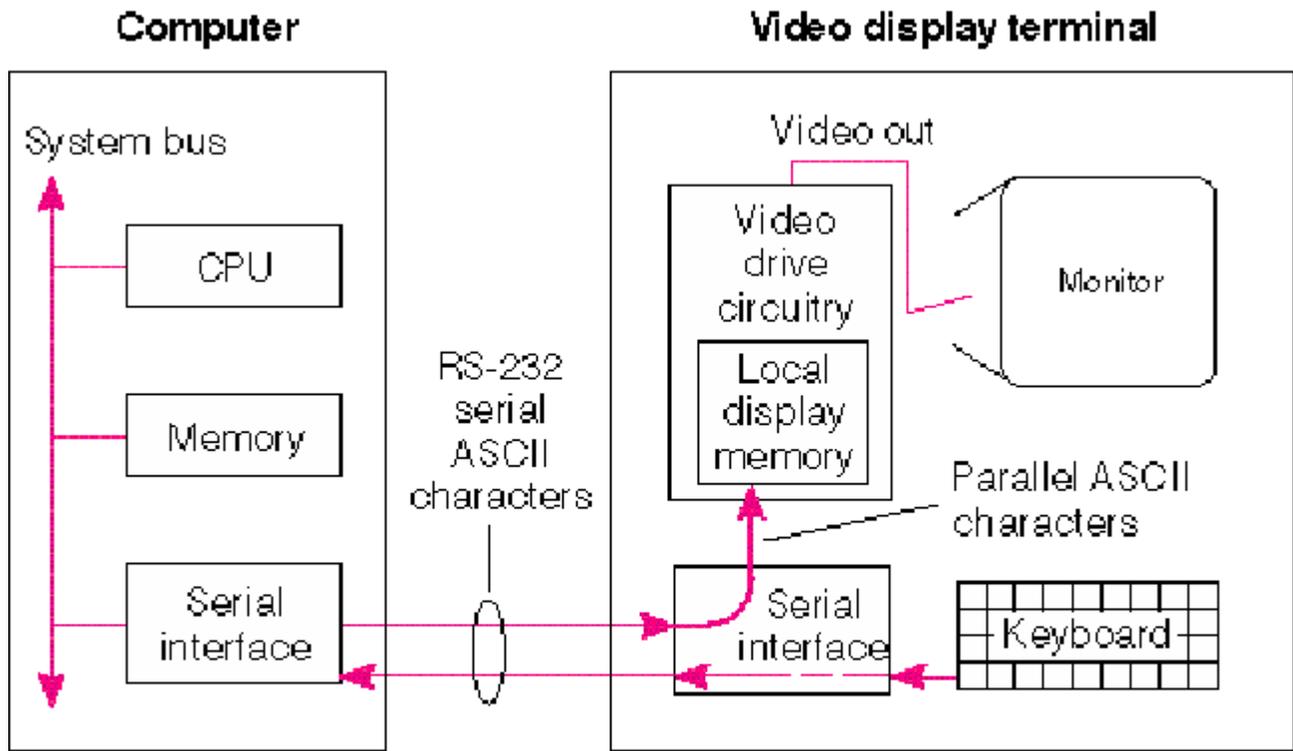
electrons at the inside of the display surface, which was coated with phosphorous material. When the electrons struck that material, they would cause it to light up. The stream of electrons is controlled using inductors, some for horizontal direction and some for vertical, as shown in the figure below.



Copyright © 2004 Pearson Prentice Hall, Inc.

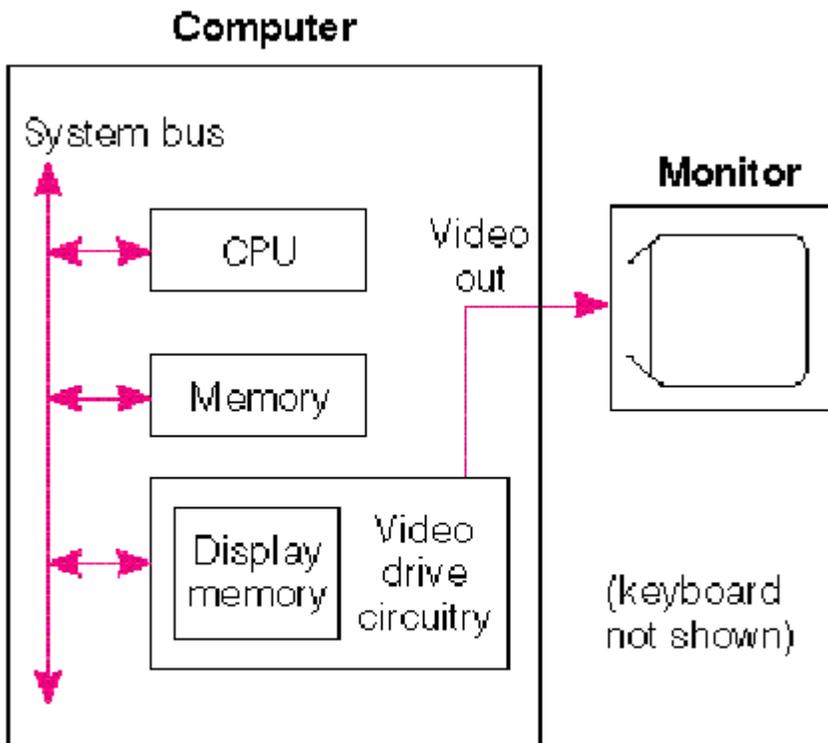
Although this type of monitor is still fairly common today, they are quickly being replaced by LCD monitors which are able to display information without the bulky CRT through the use of liquid crystals which change their color and light transmittance in response to applied electric fields.

There are also two common setups for video displays: memory-mapped video and video display terminals. The video display terminal is designed to be used when the display is located too far away from the computer to easily stretch the thick wiring needed to get a good image to it. It is generally only capable of displaying simple ASCII text. An example of such a setup is shown below.



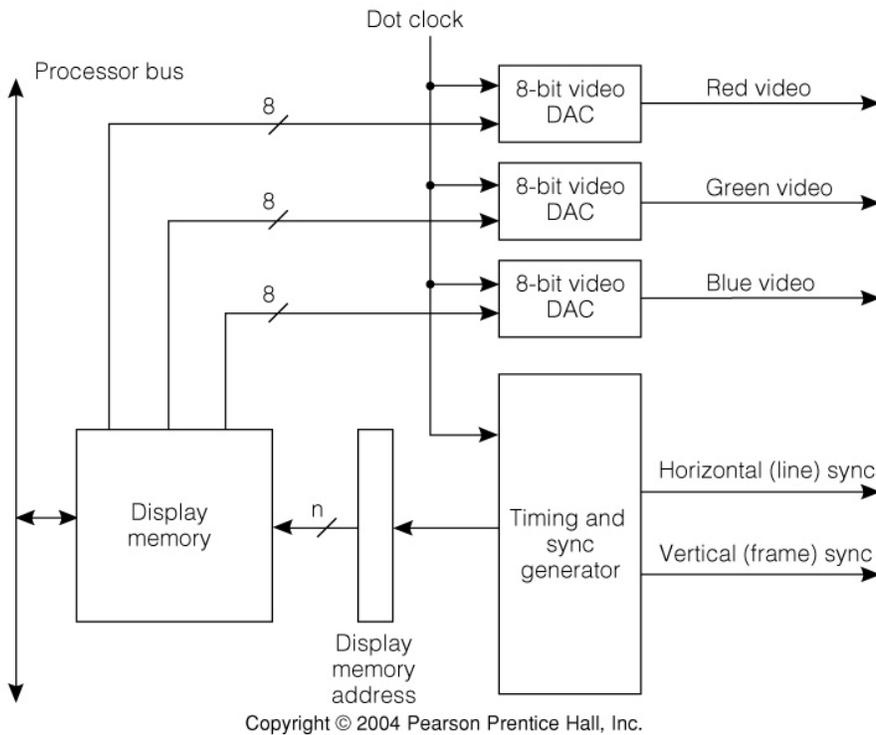
(a) Video display terminal

Memory mapped video is a much more common setup partially because it allows us to display complex images and other material to the screen, and also because it is easy to use when the computer is nearby. In this setup, the display memory is located within the computer, and the processor is able to transfer the graphic information to the screen at the full bandwidth of the internal bus, which more easily allows things like graphics and real time video to be seen. An example of such a setup is shown below.



(b) Memory mapped video

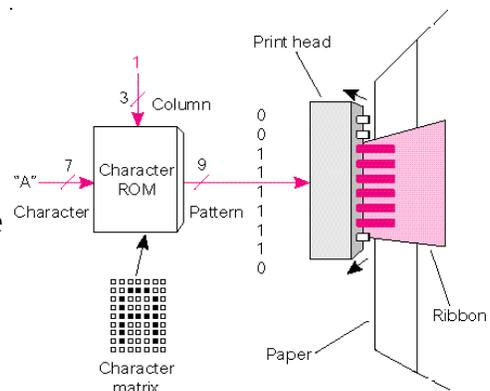
This setup has so much potential for speed that it has become very common to have video processors that are dedicated to processing the graphics being sent to the screen. This allows the central processor to concentrate on other tasks and further speeds up the graphics processing abilities. The video processor cycles through the video RAM once each frame period and translates that into the RGB values for each individual pixel on the screen, then sends the RGB values to the monitor. A basic idea of how this setup works is shown below.



Printers

So much work goes into putting data inside the computer, it would be a shame if the user couldn't retrieve it and physically bring it with them somewhere where the computer isn't. Printers allow this to happen by literally printing out what is on the computer screen. There are many forms of printers including dot-matrix, ink jet, and laser printers. The printer interfaces are varied as well from RS-232 (serial), parallel, to USB connections. RS-232 is the slowest of the group and is typically used with dot-matrix printers because they use very little bandwidth.

More complicated printers such as ink jet need a faster connection like USB to allow for more data to flow. The “backstage” action of a dot-matrix printer is shown in the figure to the right. Information comes in, and the information is physically recorded onto a sheet of paper for the user to take anywhere with them.



Input Devices

Input devices is far too large a subject to go into great detail here, but the most common such devices are the mouse and keyboard. These devices are what actually allow people to use computers, and are generally easy to handle in the computer because the input usually cannot come in faster than a human can think about it. In the case of the mouse and keyboard, they cannot come in faster than you can move your hands, which for a computer is extremely slow. Essentially input devices allow users to give instructions to the computer or to send information to it.

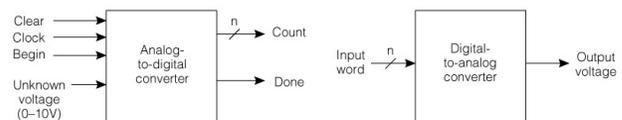
Interfacing to the Analog World

Quite simply put, the computer is digital and the world is analog. Interfacing the two takes a conversion process known as analog to digital conversion (ADC) or digital to analog conversion (DAC). Working with audio is a good example

because it requires both ADC and DAC converters.

Say, you want to record a song and store it to the hard drive. You'll have to take the analog signal of your voice, your musical instrument, or whatever else and

turn it into something that the HDD can recognize. You reverse the process when you want to playback the recording you just made from the HDD onto your computer speakers (DAC). The conversion itself is quite complicated and would require a few more pages to explain. The most important part is that conversion is possible using AC and DC circuits.



Copyright © 2004 Pearson Prentice Hall, Inc.

Conclusion

It is important to understand that, regardless of all the incredible technology and capability contained within the modern computer, of all the power in today's microprocessors, or of all the speed on current motherboards, without peripheral devices computers would be completely useless (unless you happen to need something to put things on, like a table or perhaps a nightstand).

Processors cannot process information unless it can be given to them, and that cannot be accomplished without some sort of input devices. Even then, once the information is processed it is no good unless we can see and evaluate the results, which is impossible without some form of output device. That all then brings up the need to store that information for future needs, otherwise what was the point in doing the work in the first place? Thus came about tools like monitors, keyboards, hard drives, and printers that allow us to give input and receive output from our computers in the various mediums we need.