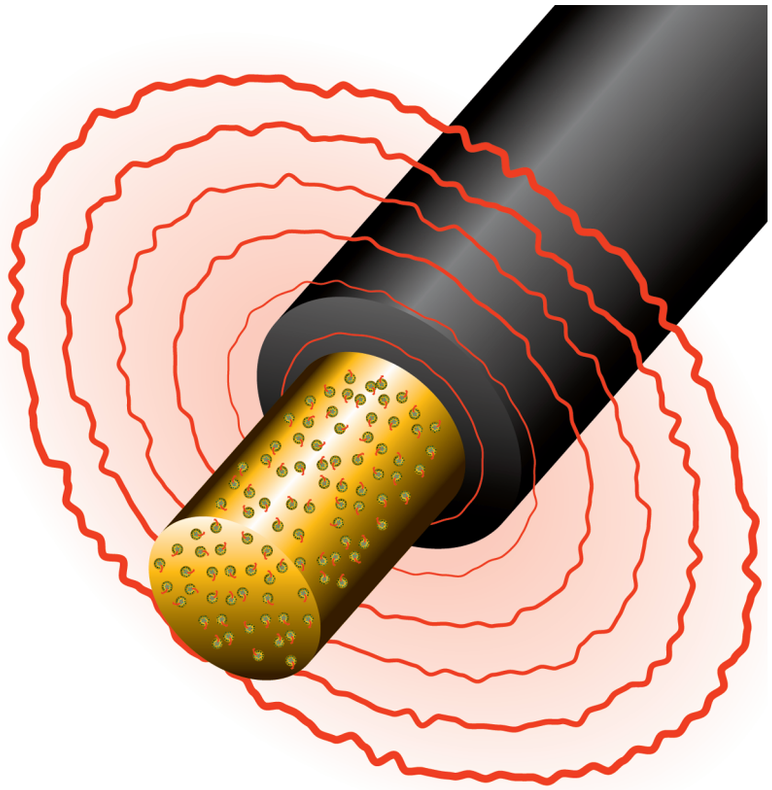
CHAPTER 10

How a Computer's Long-Term Memory Works



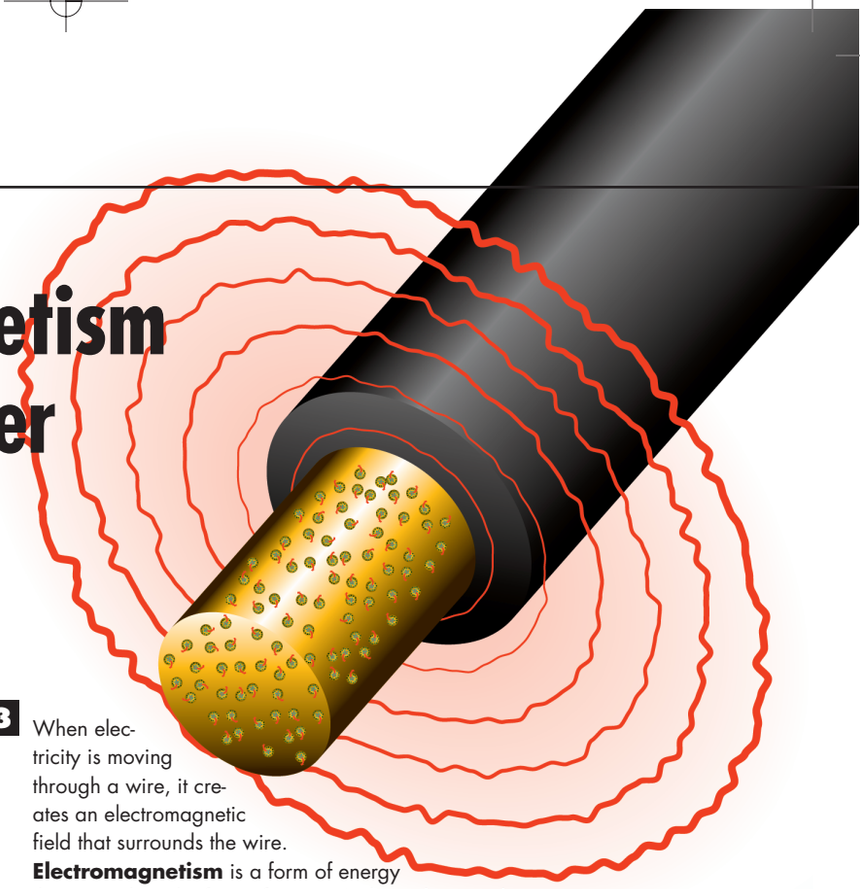
MAGNETIC disks are the most common form of permanent data storage. Their capacities can range from a few hundred kilobytes to scores of gigabytes, but they all have some elements in common. For one, the way that a drive's mechanism creates the ones and zeros that make up the binary language of computers might differ, but the goal is the same: to alter microscopically small areas of the disk surface so that some of the areas represent zeros and others represent ones. The new disk uses only those two numbers whether it records a great novel or this week's grocery list.

Another common element among magnetic drives is a scheme that determines how the data on the disk is organized. The computer's operating system, beginning with DOS and continuing with Windows and every other OS, determines the scheme. Many people forget just how long PC hard drives felt DOS's influence, even after Windows 95 took the world by storm. In Windows 95, Windows 98 and Windows Me, the older DOS is still there, it was just hidden beneath Windows's graphic interface. The operating system controls so many of a PC's operations that many PC users forget that DOS stands for **disk operating system** and that, originally, its primary function was to control disk drives.

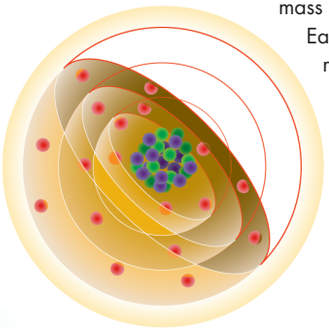
Before any information can be stored on a magnetic disk, the disk must first be **formatted**. Formatting creates a road map that allows the drive to store and find data in an orderly manner. The road map consists of magnetic markers embedded in the magnetic film on the surface of the disk. The codes divide the surfaces of the disk into sectors (pie slices) and tracks (concentric circles). These divisions organize the disk so that data can be recorded in a logical manner and accessed quickly by the read/write heads that move back and forth over the disk as it spins. The number of sectors and tracks that fit on a disk determines the disk's capacity to hold information.

After a disk is formatted, writing or reading even the simplest file is a complicated process. This process involves your software, operating system, the PC's **BIOS (basic input/output system)**, software drivers that tell the operating system how to use add-on hardware such as an external USB flash memory drive, and the mechanism of the disk drive itself.

How Electromagnetism Reacts with Matter

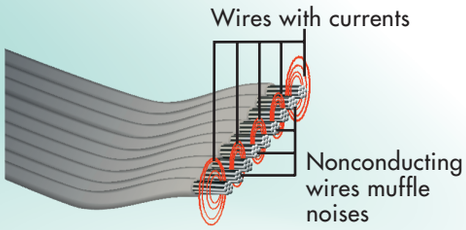


1 All atoms are made up of positively charged particles called **protons** at the center, or nucleus, of the atom, and the cloud of negatively charged **electrons** that surround the nucleus. (The nucleus also contains **neutrons**, particles that have mass but no electrical charge.) Each type of charged particle repels similarly charged particles and attracts particles with the opposite charge.



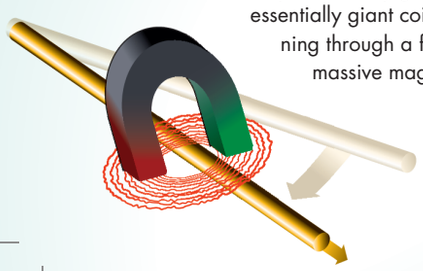
3 When electricity is moving through a wire, it creates an electromagnetic field that surrounds the wire. **Electromagnetism** is a form of energy that spreads in the form of waves. Light, radio signals, microwaves, television signals, and heat are examples of electromagnetism.

2 In some atoms, such as copper and aluminum, the attraction is weak between the protons and the electrons in the atoms' outermost layer. In such **conductive** materials, electrons jump freely from one atom to another under the right circumstances. This movement of electrons is **electricity**. In other materials, such as rubber and glass, electrons are more closely bound to their nuclei and do not easily move from one atom to another. These **nonconductive** materials are **insulators**. Still other materials, such as silicon, can act as either conductors or nonconductors under different conditions. These are **semiconductors**, an important component of microchips and transistors.

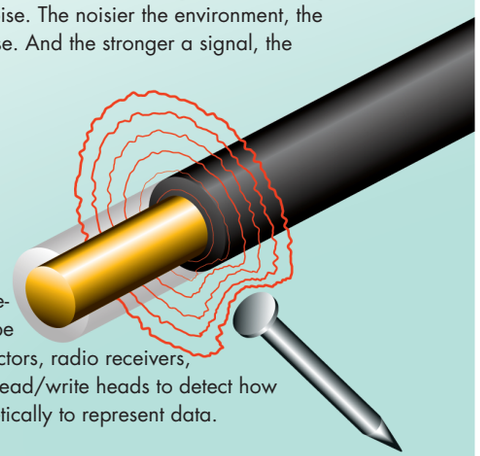


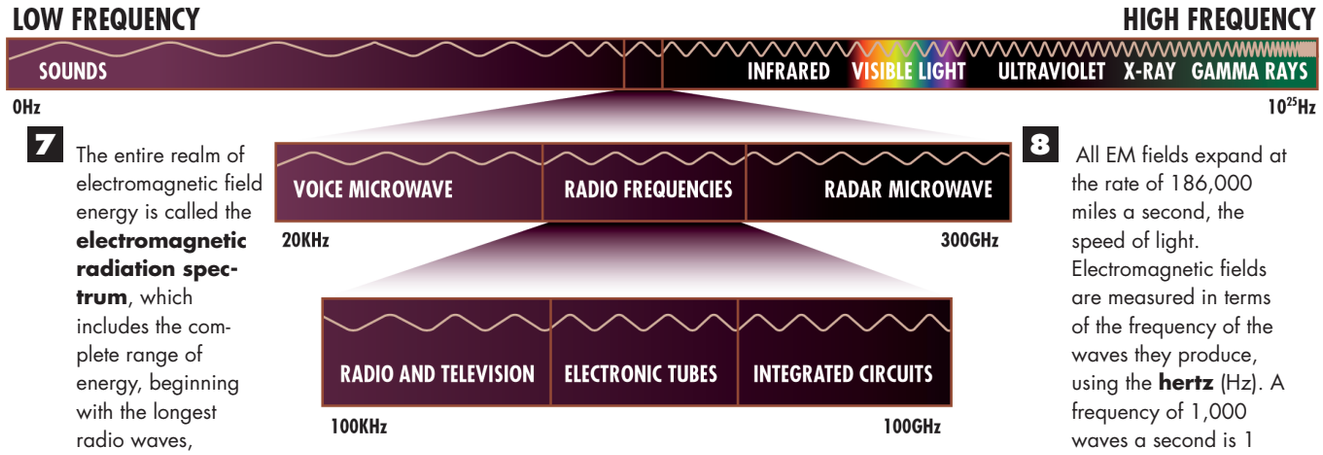
4 All electrical devices invariably create stray, unwanted electromagnetic fields called **noise**. Static on the radio and snow on a television screen are tangible examples of electrical noise. Much of the electromechanical design of computer components, such as adding nonfunctioning wires to cables, is aimed at reducing the interference of noise. The noisier the environment, the stronger a signal must be to make itself heard over the noise. And the stronger a signal, the more noise it creates that affect other components.

5 Conversely, when a wire moves through a magnetic field, the interaction creates an electrical current in the wire. **AC** electrical current is produced by generators that are essentially giant coils of wire spinning through a field created by massive magnets.



6 Similarly, electromagnetic fields are influenced by the presence, proximity, shape, composition, and mass of objects moving through the field. Those changes to the frequency or amplitude of an EM field can be detected and are the basis for metal detectors, radio receivers, and the capability of a computer drive's read/write heads to detect how particles on the disk are arranged magnetically to represent data.

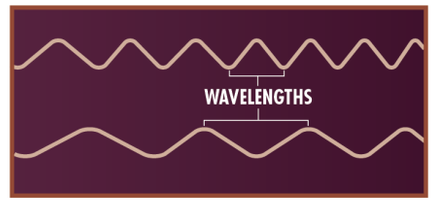


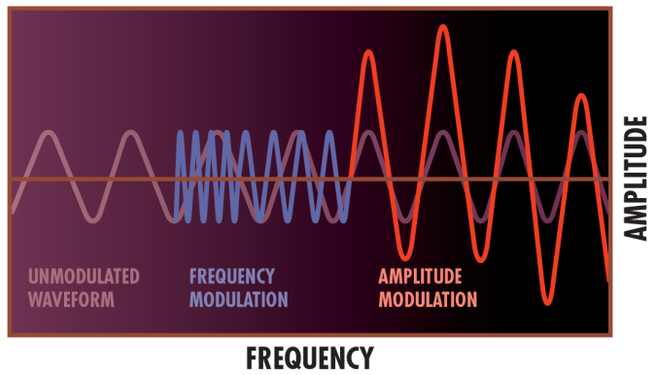


7 The entire realm of electromagnetic field energy is called the **electromagnetic radiation spectrum**, which includes the complete range of energy, beginning with the longest radio waves, through visible light—a very small part of the spectrum—to the extremely short gamma rays produced by radioactive atoms and major astronomical events, such as novas.

8 All EM fields expand at the rate of 186,000 miles a second, the speed of light. Electromagnetic fields are measured in terms of the frequency of the waves they produce, using the **hertz** (Hz). A frequency of 1,000 waves a second is 1 kilohertz (kHz). In the case of infrared, visible light, ultraviolet, and gamma radiation, the wavelength is more often specified in **nanometers** (units of 10^{-9} meter) or **Angstrom units** (units of 10^{-10} meter).

9 The frequency of an EM field is inversely related to its **wavelength**, the distance between identical points in adjacent waves. The higher the frequency of the signal, the shorter the wavelength. A signal at 100MHz—in the middle of the FM radio broadcast band—has a wavelength of about 10 feet. A signal at 30 gigahertz (GHz)—in the range of radar and microwaves—has a wavelength of a little less than half an inch.



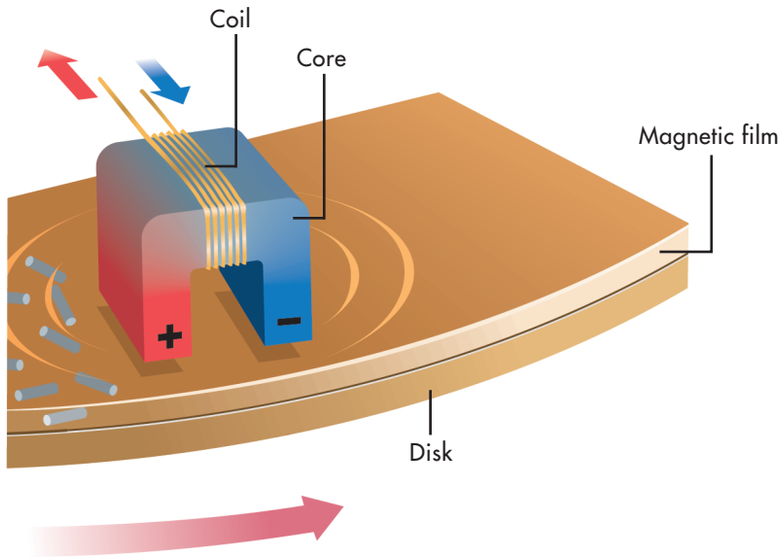


10 Electromagnetic fields are used as signals to carry data by creating variations in the **waveform**.

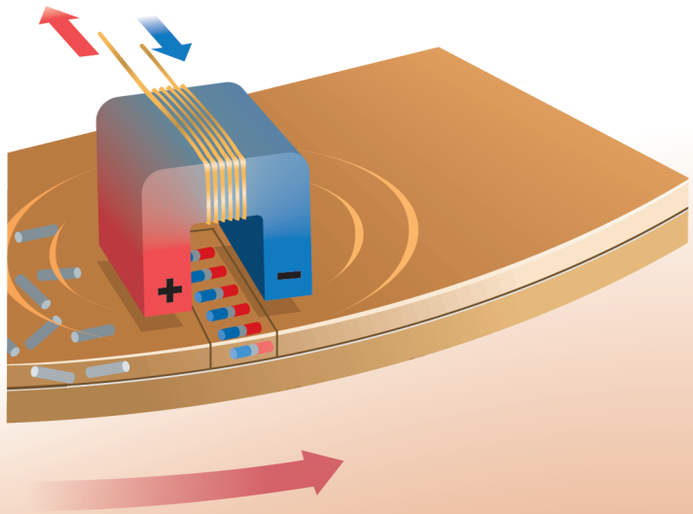
Frequency modulation (FM) carries data by varying the frequency of a fundamental waveform.

Amplitude modulation (AM) varies the strength, or amplitude, of a basic waveform. The amount of data that a signal can carry increases with the frequency of the electromagnetic field creating them. Because there are more variations in one second of higher frequency waves, there are more opportunities to modulate the wave so that it carries data.

How a Drive Writes and Reads Bits on a Disk

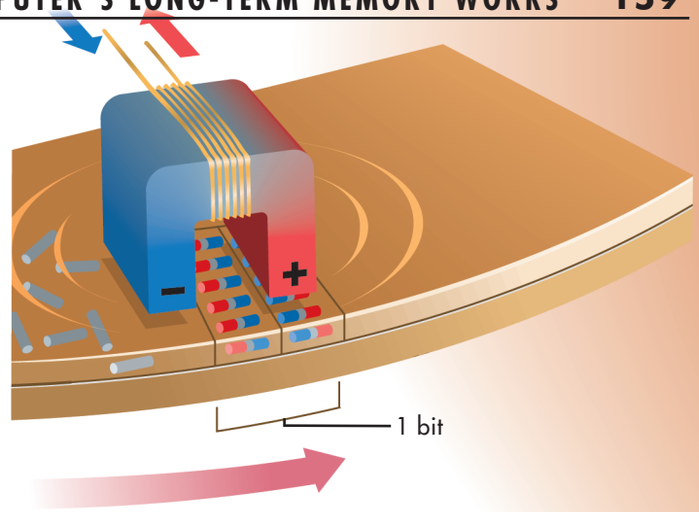


1 Before a PC writes any data to a magnetic drive, iron particles are scattered in a random pattern within a magnetic film that coats the surface of the disk. The film is similar to the surface of audio and video tapes. To organize the particles into data, electricity pulses through a coil of wire wrapped around an iron core in the drive mechanism's read/write head, which is suspended over the disk's surface. The electricity turns the core into an electromagnet that can magnetize the particles in the coating, much as a child uses a magnet to play with iron filings.

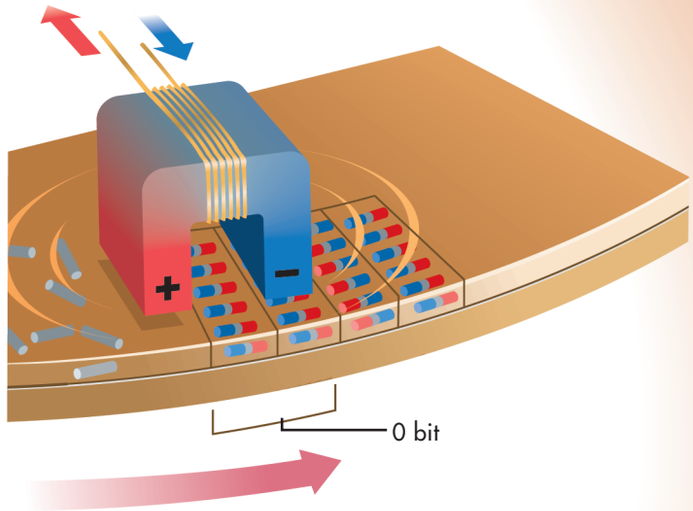


2 The coil induces a magnetic field in the core as it passes over the disk. The field, in turn, magnetizes the iron particles in the disk coating so their positive poles (red) point toward the negative pole of the read/write head, and their negative poles (blue) point to the head's positive pole.

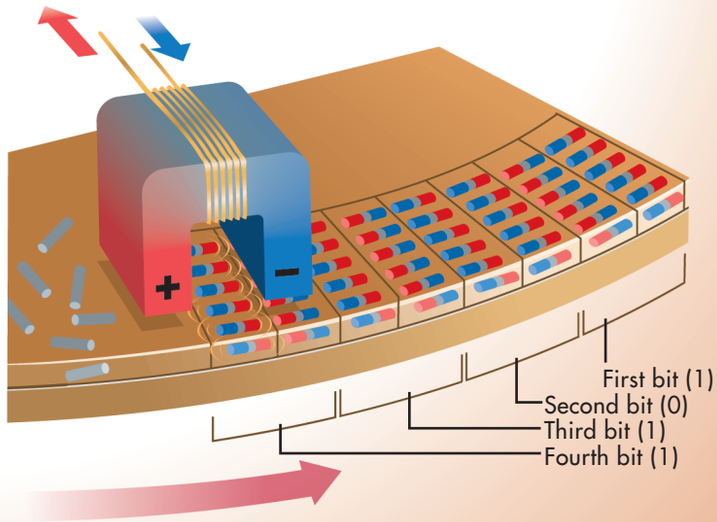
3 After the head creates one band of aligned, magnetized particles on the revolving disk, a second band is created next to it. Together, the two bands represent the smallest discrete unit of data that a computer can handle—a *bit*. If the bit is to represent a binary 1, after creating the first band, the current in the coil reverses so that the magnetic poles of the core are swapped and the particles in the second band are magnetized in the opposite direction. If the bit is a binary 0, the particles in both bands are aligned in the same direction.



4 When a second bit is stored, the polarity of its first band is always the opposite of the band preceding it to indicate that it's beginning a new bit. Even the slowest drive takes only a fraction of a second to create each band. The stored bits in the illustration represent the binary numeral 1011, which is 11 in decimal numbers.



5 To read the data, no current is sent to the read/write head as it passes over the disk. Instead, the magnetic opposite of the writing process happens. The banks of polarized particles in the disk's coating are themselves tiny magnets that create a magnetic field through which the read/write head passes. The movement of the head through the magnetic field generates an electrical current that travels in one direction or the other through the wires leading from the head. The direction the current flows depends on the polarities of the bands. By sensing the changes in direction of the current, the computer can tell whether the read/write head is passing over a 1 or a 0.

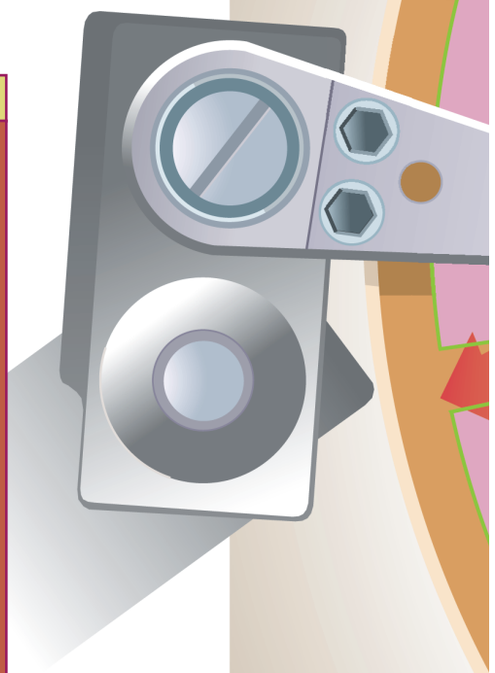


How a Drive Maps a Disk's Surface

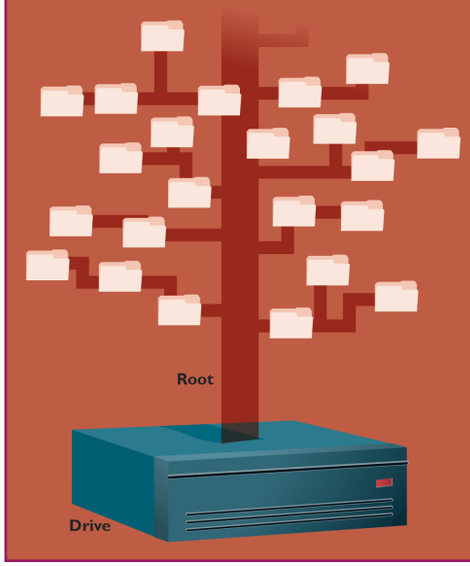
1 The first task a magnetic drive must accomplish is to format any disk that is used with it so that there is a way to organize and find files saved to the disk. It does this by writing a pattern of ones and zeros onto the surface of the disk—like magnetic signposts. The pattern divides the disk radially into sectors and into concentric circles called **tracks**. As the read/write head moves back and forth over the spinning disks, it reads these magnetic signposts to determine where it is in relation to the data on the disk's surface.

2 Two or more sectors on a single track make up a **cluster** or **block**. The number of bytes in a cluster varies according to the disk's size and the version of the operating system used to format the disk. A cluster is the minimum unit the operating system uses to store information. Even if a file has a size of only 1 byte, a cluster as large as 32 kilobytes (KB) might be used to contain the file on large drives. The number of sectors and tracks and, therefore, the number of clusters that a drive can create on a disk's surface, determine the capacity of the disk.

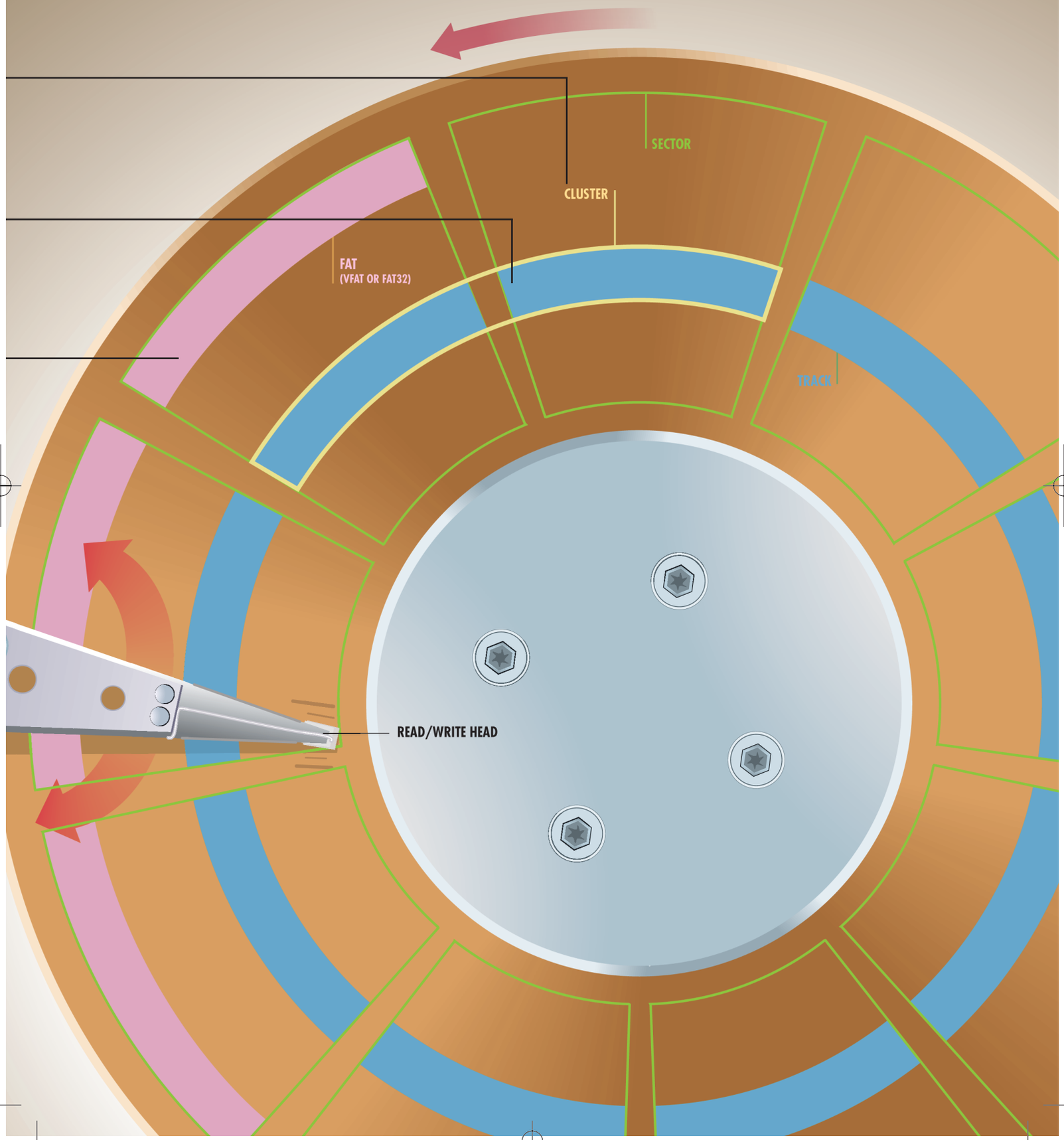
3 The drive creates a special file located in the disk's sector 0. (In the computer world, numbering often begins with 0 instead of 1.) This file is called the **file allocation table**, or **FAT**, in DOS, and the **VFAT (virtual FAT)** in Windows 95/98. VFAT is faster because it allows the computer to read files 32 bits at a time, compared to the 16-bit reads of the older FAT. VFAT also allows the use of filenames up to 255 characters long, compared to the 11 used by DOS. The FATs are where the operating systems store the information about the disk's directory, or folder structure, and which clusters are used to store which files. They also permit clusters of 4KB regardless of disk size. With Windows 98 came FAT32, which allows hard drives larger than 2 gigabytes to be formatted as a single disk. An identical copy of the FAT is kept in another location in case the data in the first version becomes corrupted. Ordinarily, you will never see the contents of the FAT, VFAT, or FAT32. In the NT file system (NTFS), which is used in Windows 2000 and XP, the FAT disappears entirely. Information about a file's clusters is instead stored in each one of those clusters.



The Computer Filing Cabinet

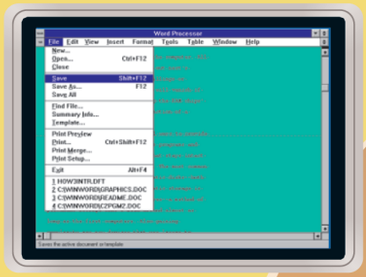


Think of a disk as being a filing cabinet in which you keep all your documents. Each drawer in the cabinet is the equivalent of one of your drives—floppy, hard disk, or optical. On each drive, the first level of organization, called the *root*, contains **directories**, or **folders**—the digital equivalent of a file cabinet's cardboard file folders. Each directory contains the individual files—documents, spreadsheets, graphics, programs—the same way that a drawer's file folders contain individual letters, reports, and other hard copy. One important difference is that it's easy for drive folders to contain other folders, which can contain still more folders, and so on, indefinitely. This directory/folder structure is called a **tree** because a diagram of how it's organized looks like the branching structure of a tree.



How a PC Saves a File to Disk

1 When you click your mouse to save a file, the program you're using sends a command to Windows, asking the operating system to carry out the steps needed to save the file from RAM, where it's being held temporarily, to disk for permanent storage. For this example, we'll assume you're using a word processor to save a file named Letter to Mom.doc.



2 Windows modifies the record of the folder (directory) structure stored in the **virtual file allocation table**, or **VFAT** (simply **FAT** in DOS or **FAT32** in Windows 98), to indicate that a file named Letter to Mom.doc will be stored in the current folder, or in another folder if you provide a different directory path.
There is no FAT in Windows NT, 2000, and XP. Location information for each cluster in a file is saved in every cluster that helps make up the file.

VIRTUAL FILE ALLOCATION TABLE

VIRTUAL FILE ALLOCATION TABLE

FILE	CLUSTER
Letter to Mom.doc	3
New Budget.xls	4
EMPTY	5

7 Finally, Windows or DOS changes the information contained in the VFAT to mark which clusters contain Letter to Mom.doc, so that later the operating system will know the clusters are already in use and won't overwrite Mom's letter.

CLUSTER ADDRESS

CLUSTER	TRACK	SECTORS
12	3	6,7,8,9

What Happens When You Delete a File?

When you delete a file, the data that makes up the file is not actually changed on the disk. Instead, the operating system changes the information in the VFAT to indicate that the clusters that had been used by that file are now available for reuse by other files. Because the data remains on disk until the clusters are reused, you often can restore—or *undelete*—a file that you've accidentally erased.

VIRTUAL FILE ALLOCATION TABLE

Available

FILE	CLUSTER
Letter to Mom.doc	3
New Budget.xls	4
EMPTY	5
Old Budget.xls	6

3 The operating system also checks the VFAT for the number of a cluster where the VFAT says Windows can save the file without overwriting any other data that's already been saved. In this example, the VFAT tells Windows that Cluster 3 is available to record data.

VIRTUAL FILE ALLOCATION TABLE

FILE	1ST CLUSTER
Expenses.xls	1
Annual Report.doc	2
AVAILABLE	3
New Budget.xls	4

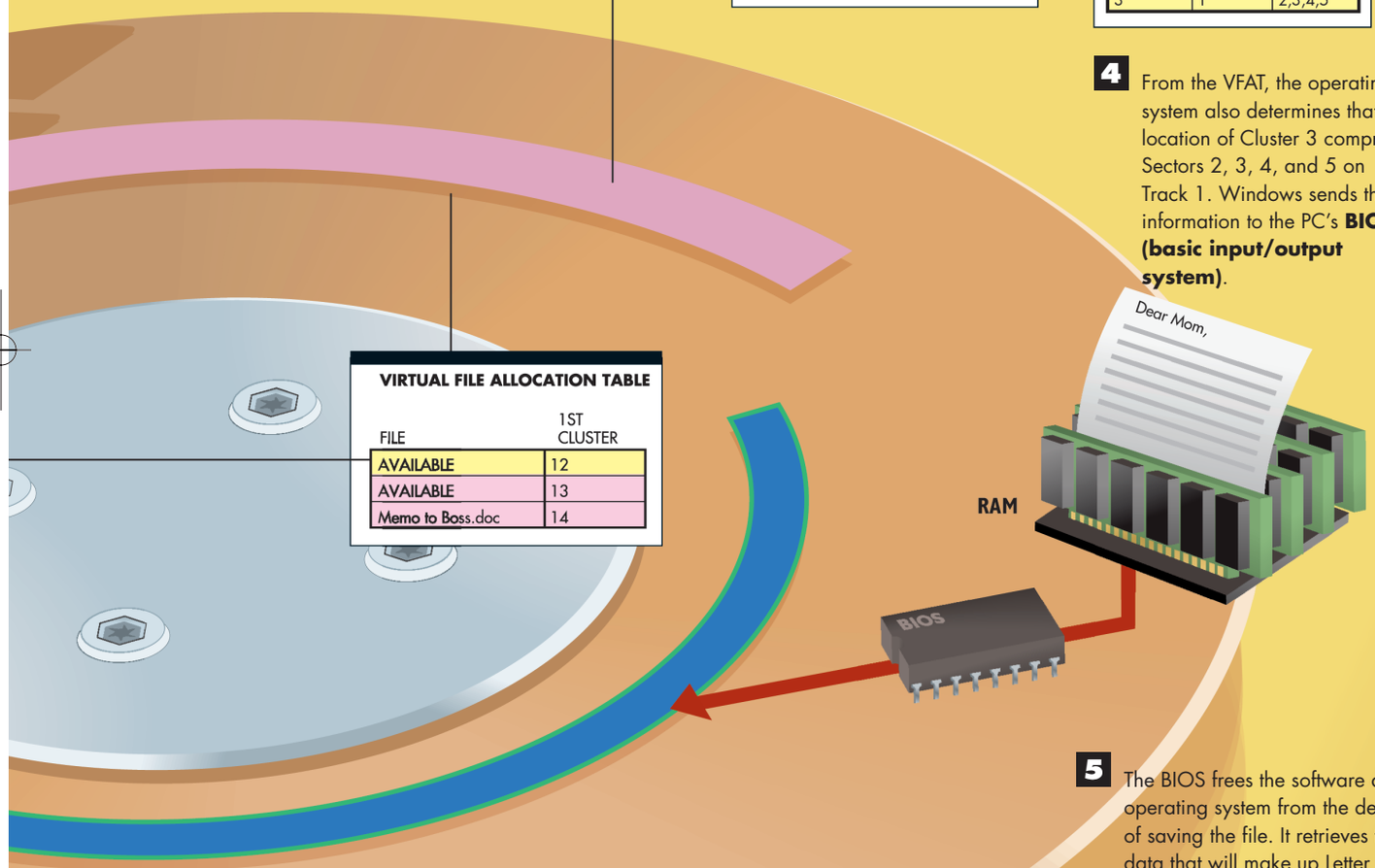
CLUSTER ADDRESS

CLUSTER	TRACK	SECTORS
3	1	2,3,4,5

4 From the VFAT, the operating system also determines that the location of Cluster 3 comprises Sectors 2, 3, 4, and 5 on Track 1. Windows sends this information to the PC's **BIOS (basic input/output system)**.

VIRTUAL FILE ALLOCATION TABLE

FILE	1ST CLUSTER
AVAILABLE	12
AVAILABLE	13
Memo to Boss.doc	14

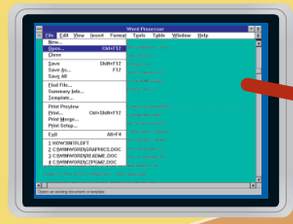


5 The BIOS frees the software and operating system from the details of saving the file. It retrieves the data that will make up Letter to Mom.doc from where the word processor is using it in RAM. At the same time, it issues the instructions to the disk drive controller to save the data that the BIOS is sending it, beginning at Sectors 2 through 5 on Track 1.

6 If the file is larger than the number of bytes contained in a single cluster, the operating system asks the VFAT for the location of another cluster in which it can continue saving the file. The clusters need not be adjacent to each other on the disk. The VFAT maintains a record of the chain of clusters over which the file is spread. The process of moving data from RAM to disks repeats itself until the operating system encounters a special code called an **end-of-file marker**.

How a PC Retrieves a File from a Disk

1 When you use the File menu command to open a file—for example, Letter to Mom.doc—the first thing your word processor does is call on its API and, in Windows', DLLs (see Chapter 8 ,"How Windows Works"). These tools build a File Open dialog box, retrieve the list of files from the current default folder, display them in the box, and wait on your selection.



2 When you click on your letter to Mom, the operating system and its associated programs take over. In earlier operating systems, the operation was pretty much handled by the operating system and a few DLLs. In Windows XP and Vista, there are more layers to the file system to put protection between you and the operating system and to handle today's larger number of devices for saving files. To handle all this, Windows calls on the **Installable File System (IFS) Manager.** It's the job of the IFS to pass control, whether it be from DOS, a 16-bit application, or a 32-bit application, to the appropriate **file system driver (FSD)** from a choice of four or more that work with different storage systems—NTFS, VFAT, CDFS for optical drives, and Network.

3 The FSD gets the disk location of the first cluster of the letter for Mom from VFAT or, in Windows XP, from the **MFT (master file table)**, which is the VFAT on steroids, containing more details about a file than its location. In fact, if a file is smaller than 2KB, the file is stored entirely in the MFT itself. Copies of the MFT are stored at various places on the disk as a precaution against damage to any single copy.

API

COMMDLG.DLL

MASTER FILE TABLE OR VIRTUAL FILE ALLOCATION TABLE

FILE SYSTEM DRIVERS

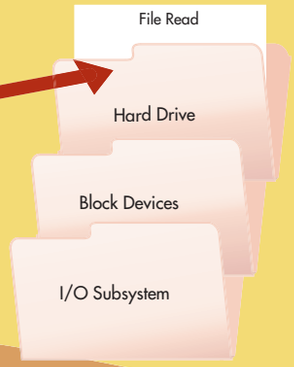
VFAT

NTFS

CDFS

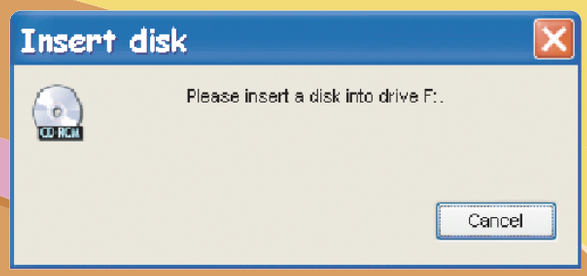
NET

INSTALLABLE FILE SYSTEM MANAGER

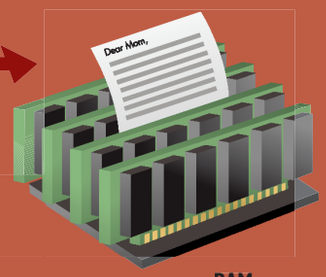


4 The FSD passes the read command and the info about the file to the next stage, the **IO Subsystem (IOS)**. This is generally an assistant to the FSDs, carrying out such chores as routing messages back and forth between the FSDs and lower, device-specific drivers.

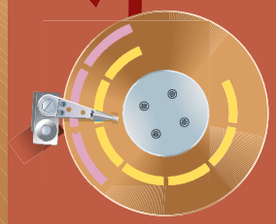
5 The **volume tracking driver (VTD)** might get into the action if the file you want to read is on a floppy, CD, DVD, or other removable drive. The VTD's only job is to make sure that the correct disc, or disk, is in the correct drive. If you leave a file open on a floppy and remove the floppy disk before you try to save the file, the VTD pops up to tell you to reinsert the right floppy.



6 Now the operation passes off to a **type-specific driver (TSD)**. There are individual TSDs for hard drives, floppy drives, network drivers, and any other class of hardware on your PC. A TSD might get help from drivers that the hardware's seller provides and by the **port driver** in Windows to let a message get from the main bus—the motherboard–to the drive's adapter.



RAM



7 Finally, the adapter takes over, moving the read/write heads to the correct series of disk clusters to retrieve the file, which is copied to memory, where you and your computer can work with it.