Key Terms

access arm
access time
areal density
average access time
block
Blu-ray disc (BD)
CD digital audio (CD-DA)
CD read-only memory (CD-ROM)
coercivity
compact disc (CD)
cylinder
data transfer rate
direct access
disk defragmentation
disk drive
double data rate (DDR)
double inline memory module (DIMM)
drive array
DVD
DVD read-only memory (DVD-ROM)
dynamic RAM (DRAM)
firmware
flash RAM
fragmented
germanium, antimony, and tellurium (GST)
head-to-head (HTH) switching time
hybrid disk drive
magnetic decay
magnetic leakage
magnetic tape
magnetoresistive RAM (MRAM)
magneto-optical (MO) drive
multilevel cell (MLC)
nonvolatile
nonvolatile memory (NVM)
parallel access
phase-change memory (PCM)
platter
random access
random access memory (RAM)
read-only memory (ROM)
read/write head
read threshold
refresh cycle
rotational delay
sector
sequential access time
serial access
single inline memory module (SIMM)
single-level cell (SLC)
solid-state drive (SSD)
static RAM (SRAM)
storage medium
sustained data transfer rate
synchronous DRAM (SDRAM)
tape drive
track
track-to-track (TTT) seek time
vertical flash RAM
volatile

Vocabulary Exercises

1. Dynamic RAM requires frequent ________ to maintain its data content.
2. The ________ rate is the speed at which data can be moved to or from a storage device over a communication channel.
3. Three standard optical storage media that are written only during manufacture are called ________, ________, and ________.

Other features that server and data center customers look for include error detection and correction and predictable response to power loss. Read errors are more frequent with flash RAM than with magnetic drives, so all SSDs include some form of error detection and correction. For the server and data center market, manufacturers typically employ methods that are more reliable but also more expensive. In addition, many server and data center SSDs include a reserve power supply such as battery or high-powered capacitor to complete pending write operations if externally supplied power fails.

OPTICAL STORAGE

Optical storage came of age in the 1990s. Its primary advantages, compared with magnetic storage, are higher areal density and longer data life. Typical optical recording densities are at least 10 times higher than for magnetic storage devices. Higher density is achieved with tightly focused lasers that can access a very small storage medium area. Magnetic fields can't be focused on such a small area without overwriting surrounding bit positions.

Most optical storage media can retain data for decades because they aren't subject to magnetic decay and leakage. Optical storage is popular because of its standardized and inexpensive storage media. The CD and DVD formats used for musical and video recordings are also supported by most computer optical storage devices.

Optical storage devices store bit values as variations in light reflection (see Figure 5.13). The storage medium is a surface of highly reflective material, and the read mechanism consists of a low-power laser and a photoelectric cell. The laser is focused on 1-bit location of the storage medium at a specific angle, and the photoelectric cell is positioned at a complementary angle to intercept reflected laser light. A highly reflective surface spot causes the photoelectric cell to generate a detectable electrical current. A poorly reflective surface spot doesn't reflect enough light to cause the photoelectric cell to fire. The current, or lack of current, the photoelectric cell produces is interpreted as a binary 1 or 0.

There are multiple ways to implement bit areas with low reflectivity. In Figure 5.13, low reflection is a result of a concave dent, or pit, in the reflective layer. Other ways of achieving low reflectivity bit areas include burned areas, dyes, and substances that change from crystalline to amorphous states.

An optimal surface material could be changed rapidly and easily from highly to poorly reflective and then back again an unlimited number of times. The search for this material and a reliable and efficient method of altering its reflectivity has proved difficult. Some materials change quickly in one direction but not another, such as photosensitive sunglass materials. The capability of most materials to change reflectivity degrades over time and with repeated use. Several optical write technologies are currently in use or under development, but there's no clearly dominant technology.

Current optical storage devices use a disc storage medium. (Optical storage media are usually called *discs*, and magnetic storage media are called *disks*.) Because the recording surface is a disc, there are many similarities to magnetic disk storage, including a read/write mechanism mounted on an access arm, a spinning disc, and performance limitations caused by rotational delay and movement of the access arm.

head-to-head (HTH) switching time, is no more than a few nanoseconds. Switching occurs in sequence among read/write heads, and multiple switching operations might be required to activate a head. For example, if there are 10 read/write heads, switching from the first head to the last requires 9 separate switching operations (first to second, second to third, and so forth). HTH switching time is the least important component of access time.

The positioning servo that moves access arms and read/write heads is a bidirectional motor that can be started and stopped quickly to position read/write heads precisely over a track. The time needed to move from one track to another is called **track-to-track (TTT) seek time**, typically measured in milliseconds. It's an average number because of variations in the time required to start, operate, and stop the positioning servo.

After the read/write heads are positioned over the track, the controller waits for the sector to rotate beneath the heads. The controller keeps track of the current sector position by continually scanning sector synchronization data written on the disk during formatting. When the correct sector is sensed, the read or write operation is started. The time the disk controller must wait for the right sector to rotate beneath the heads is called **rotational delay**, which depends on platter rotation speed. Increasing platter spin rate reduces rotational delay.

Both TTT seek time and rotational delay vary from one read or write operation to another. For example, if two consecutive read requests access adjacent sectors on the same track and platter, rotational delay and TTT seek time for the second read operation are zero. If the second read access is to the sector preceding the first, rotational delay is high, almost a full platter rotation. If one read request references a sector on the outermost track and the other request references the innermost track, TTT seek time is high.

The amount of TTT seek time and rotational delay can't be known in advance because the location of the next sector to be accessed can't be known in advance. Access time for a disk drive is not a constant. Instead, it's usually stated by using a variety of performance numbers, including raw times for HTH switching and rotation speed. The most important performance numbers are average access time, sequential access time, and sustained data transfer rate.

Average access time is computed by assuming that two consecutive accesses are sent to random locations. Given a large number of random accesses, the expected value of HTH switching time corresponds to switching across half the number of recording surfaces. The expected value of TTT seek time corresponds to movement over half the tracks, and the expected value of rotational delay corresponds to one-half of a platter rotation. If HTH switching time is 5 ns, the number of read/write heads is 10, TTT seek time is 5 µs, each recording surface has 1000 tracks, and platters are rotated at 7500 rpm, the average access delay is computed as follows:

$$
\begin{aligned}
\text{average access delay} &= \text{HTH switch} + \text{TTT seek} + \text{rotational delay} = \\
&= \left(5\text{ ns} \times \frac{10}{2}\right) + \left(\frac{1000}{2} \times 5\ \mu\text{s}\right) + \left(\frac{60\text{ seconds}}{7500\text{ rpm}} \div 2\right) \\
&= .000000025 + (500 \times .000005) + (.008 \div 2)\text{ seconds} \\
&= .000000025 + .0025 + .004\text{ seconds} \\
&\approx 6.5\text{ ms}
\end{aligned}
$$

write performance is much slower than in DRAM. Also, each write operation is mildly destructive, resulting in storage cell destruction after 100,000 or more write operations. Because of its slower write speed and limited number of write cycles, flash RAM can't compete with DRAM as a primary storage technology.

Flash RAM is used in multiple forms as a secondary storage technology. Single chips installed within a computer or other device store firmware programs that aren't changed frequently and are loaded into memory when a computer powers up. Flash RAM is also used in portable secondary storage devices, such as compact flash cards in digital cameras and USB flash drives. Flash RAM is gradually supplanting magnetic disk drives as the dominant secondary storage technology in computer systems, as described later in this chapter.

Other NVM technologies under development could overcome some shortcomings of flash RAM. Two promising candidates are magnetoresistive RAM and phase-change memory. **Magnetoresistive RAM (MRAM)** stores bit values by using two magnetic elements, one with fixed polarity and the other with polarity that changes when a bit is written. The second magnetic element's polarity determines whether a current passing between the elements encounters low (a 0 bit) or high (a 1 bit) resistance. The latest MRAM generations have read and write speeds comparable with SRAM and densities comparable with DRAM, which make MRAM a potential replacement for both these RAM types. In addition, MRAM doesn't degrade with repeated writes, which gives it better longevity than conventional flash RAM.

Phase-change memory (PCM), also known as phase-change RAM (PRAM or PCRAM), uses a glasslike compound of **germanium, antimony, and tellurium (GST)**. When heated to the correct temperatures, GST can switch between amorphous and crystalline states. The amorphous state exhibits low reflectivity (useful in rewritable optical storage media) and high electrical resistance. The crystalline state exhibits high reflectivity and low electrical resistance. PCM has lower storage density and slower read times than conventional flash RAM, but its write time is much faster, and it doesn't wear out as quickly.

Memory Packaging

Memory packaging is similar to microprocessor packaging. Memory circuits are embedded in microchips, and groups of chips are packed on a small circuit board that can be installed or removed easily from a computer system. In the late 1980s, memory manufacturers adopted the **single inline memory module (SIMM)** as a standard RAM package. Each SIMM incorporates multiple memory chips on a tiny printed circuit board. The edge of the circuit board has a row of electrical contacts, and the entire package is designed to lock into a SIMM slot on a motherboard. The **double inline memory module (DIMM)**, a newer packaging standard, is essentially a SIMM with independent electrical contacts on both sides of the module, as shown in Figure 5.5.

Current microprocessors include a small amount of on-chip memory (described in Chapter 6). As fabrication techniques improve, the amount of memory that can be packaged with the CPU on a single chip grows. The logical extension of this trend is placing a CPU and all its primary storage on a single chip, which would minimize or eliminate the current gap between microprocessor clock rates and memory access speeds.

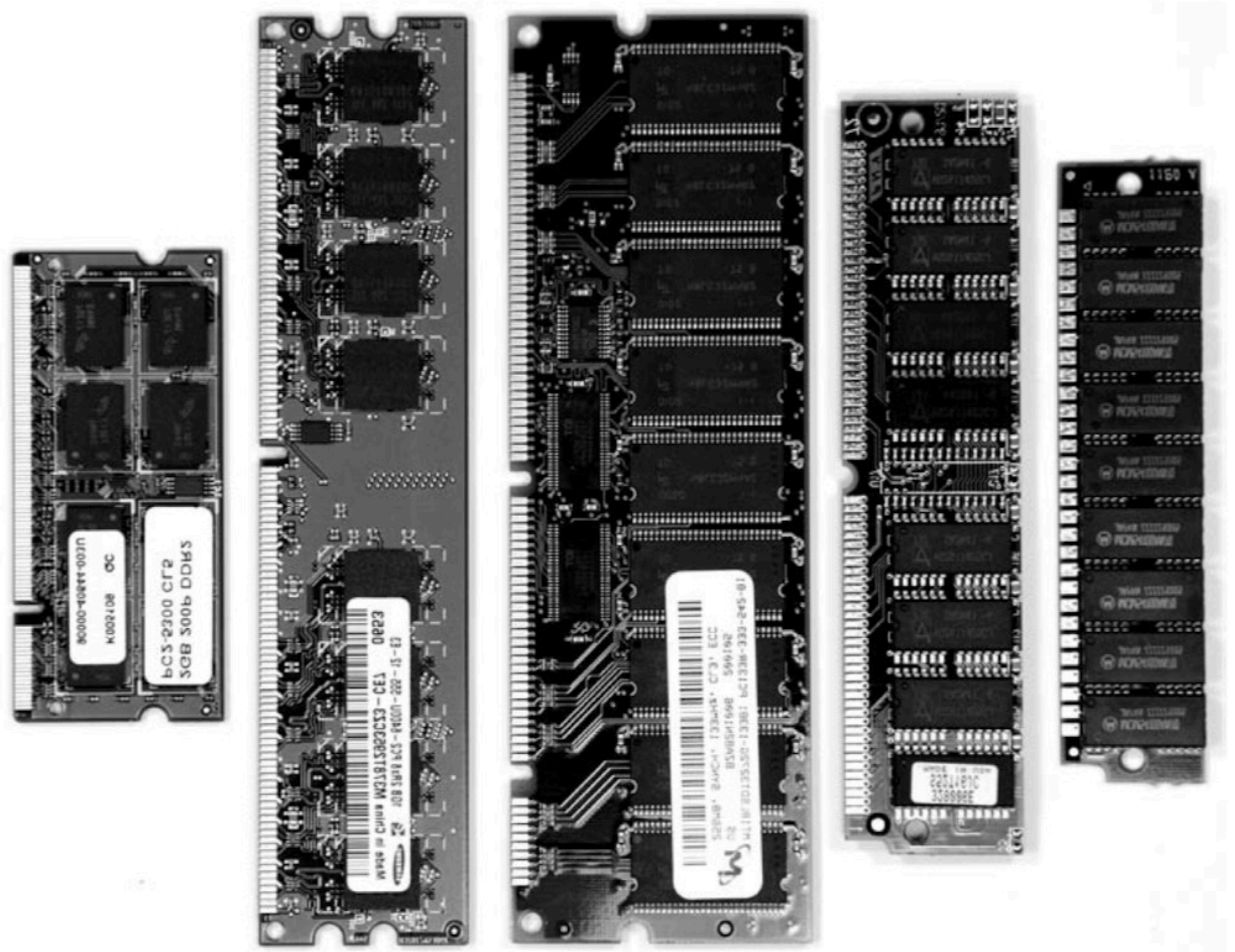

FIGURE 5.5 From left to right, 30-pin SIMM, 72-pin SIMM, DDR DIMM, DDR2 DIMM, and DDR2 DIMM for a laptop computer

MAGNETIC STORAGE

Magnetic storage devices exploit the duality of magnetism and electricity. That is, electrical current can generate a magnetic field, and a magnetic field can generate electricity. A magnetic storage device converts electrical signals into magnetic charges, captures the magnetic charge on a storage medium, and later regenerates an electrical current from the stored magnetic charge. The magnetic charge's polarity represents the bit values 0 and 1.

Figure 5.6 illustrates a simple magnetic storage device. A wire is coiled around a metallic **read/write head**. To perform a write operation, an electrical current is passed through the wire, which generates a magnetic field across the gap in the read/write head. The direction of current flow through the wire determines the field's polarity—that is, the position of the positive and negative poles of the magnetic field. Reversing the current's direction reverses the polarity.

A magnetic storage medium is placed next to the gap. The storage medium is coated with a metallic compound with high **coercivity**—the capability to accept and hold a magnetic charge. When the read/write head generates a magnetic field, the closest portion of the medium is permanently charged. The stored charge's polarity is identical to the polarity of the generated magnetic field. In addition, the strength of the stored charge is directly

Average access time is the sum of average access delay and the time required to read a single sector, which depends entirely on the disk's rotational speed and the number of sectors per track. Using the previous figures and assuming there are 250 sectors per track, the average access time is computed as follows:

$$\begin{aligned}\text{average access time} &= \text{average access delay} + \left(\frac{60 \text{ seconds}}{7500 \text{ rpm}} \div 250\right) \\ &= .0065 + .000032 \\ &\approx 6.53 \text{ ms}\end{aligned}$$

Sequential access time is the time required to read the second of two adjacent sectors on the same track and platter. HTH switching time, TTT seek time, and rotational delay are zero for this access. The only component of access time is the time required to read a single sector—that is, the second half of the preceding equation, or 32 µs in that example.

Fragmentation

Because sequential access time is so much faster than average access time, disk performance is improved dramatically if related data is stored in sequential sectors. For example, loading a program from disk to memory is fast if the entire program is stored in sequential locations on the same track. If the program won't fit in a single track, performance is maximized by storing the program in a single cylinder or group of adjacent cylinders.

If portions of a program or data file are scattered around the disk in random sectors, performance is reduced because of switching read/write heads, positioning the access arm, and waiting for sectors to rotate beneath the heads. A disk with many program and data files scattered on it is said to be **fragmented**.

NOTE

File storage in sequential sectors improves performance only if file contents are accessed sequentially, such as when loading a program into memory or playing a video. If file contents are accessed in random order, sequential storage might not improve performance substantially.

Over time, as files are created, modified, and deleted, file contents tend to become fragmented in many nonsequential sectors. Most OSs include a utility program for performing disk defragmentation. A **disk defragmentation** utility reorganizes disk content so that a file's contents are stored in sequential sectors, tracks, and platters (see Figure 5.11). The utility might be started by a user or might automatically run periodically in the background.

Data Transfer Rate

Beyond disk drive mechanics and geometry, disk drive performance also depends on the speed at which data can be moved through the disk controller circuitry to and from memory or the CPU. Communication channel capacity is rarely a restriction on a single disk drive's data transfer rate, but it can limit the effective throughput of a drive array.

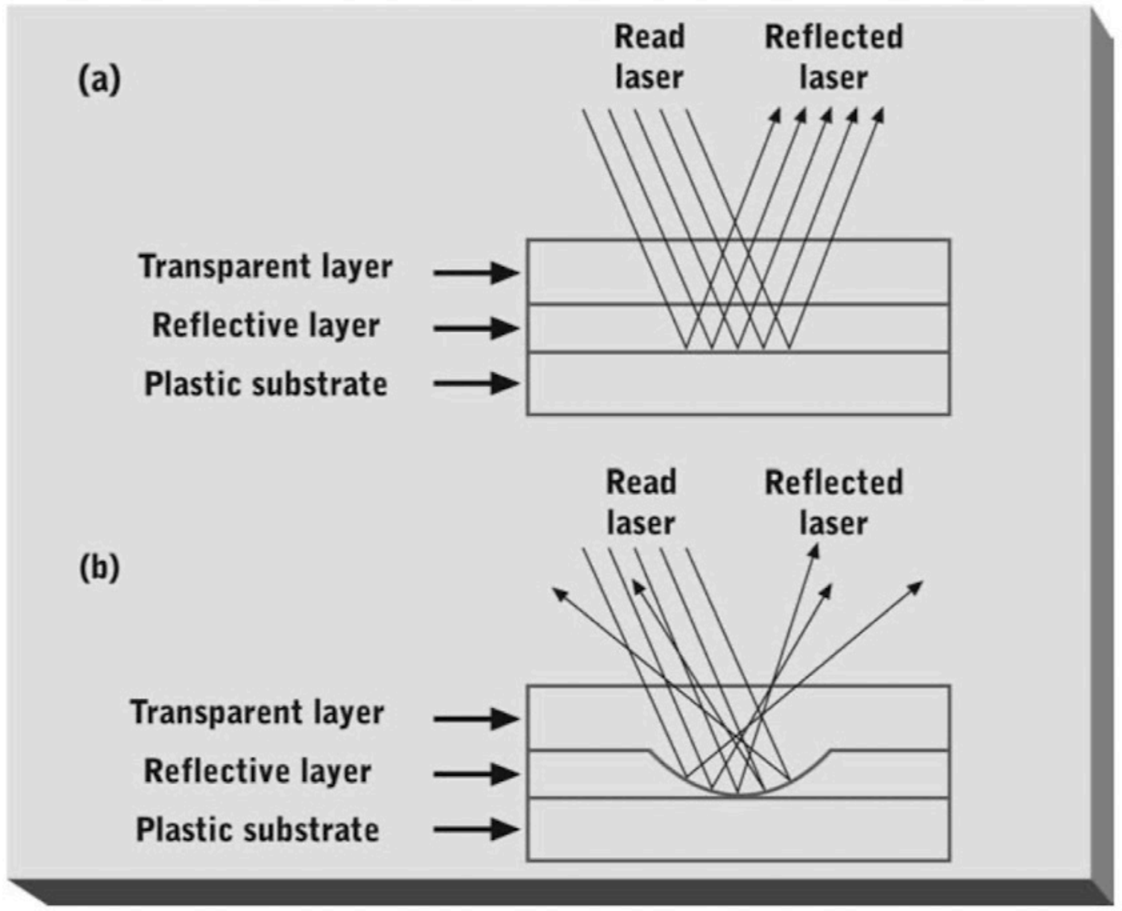

FIGURE 5.13 Optical disc read operations for a 1 bit (a) and a 0 bit (b)

Although both magnetic and optical devices use disk technology, their performance is quite different. Optical disc storage is slower for a number of reasons, including the use of removable media and the inherent complexity of an optical write operation. The read performance difference is typically 10:1 in favor of magnetic disks. Write access time is typically two or three times longer than read access time.

Optical storage doesn't directly compete with magnetic storage and SSDs because it has different cost/performance tradeoffs. SSDs or magnetic disks are usually the cost/performance leader for general-purpose online secondary storage depending on the specific use pattern. Optical storage is favored for read-only storage with low-performance requirements and when portability in a standardized format is needed. Optical storage devices are available in a wide variety of storage formats and write technologies, summarized in Table 5.4 and discussed in detail in the following sections. Standard development has slowed in recent years as music and video downloading has increased and in anticipation of the eventual replacement of both magnetic and optical disks with fixed and portable solid-state storage devices.

CD-ROM, DVD-ROM, and BD

Sony and Philips originally developed **compact disc (CD)** technology for storing and distributing music in the **CD digital audio (CD-DA)** format. **CD read-only memory (CD-ROM)** is compatible with CD-DA but includes additional formatting information to store directory and file information. Both CD-DA and CD-ROM are read-only storage technologies and formats because data is permanently embedded in the disc during manufacture.

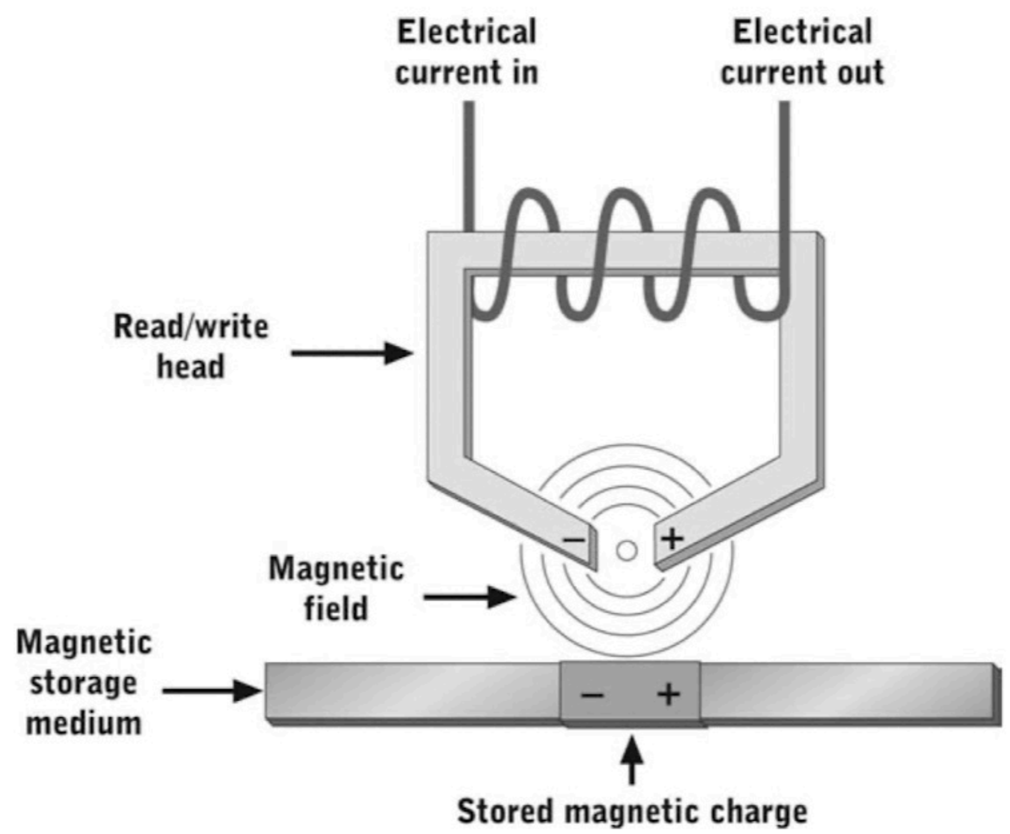

FIGURE 5.6 Principles of magnetic data storage

proportional to the magnetic field's strength, which is determined by several factors, including the number of coils in the wire, the strength of the electrical current, and the mass of the metallic read/write head.

A read operation is the inverse of a write operation. The portion of the storage medium with the data being read is placed near the gap of the read/write head, and the stored charge radiates a small magnetic field. When the magnetic field comes in contact with the read/write head, a small electrical current is induced in the read/write head and the wire wrapped around it. The current direction depends on the stored charge's polarity. Electrical switches at either end of the coiled wire detect the direction of current flow and sense a 0- or 1-bit value. The switches require a minimum current level for reliable operation. Because induced current power is proportional to the magnetic charge's strength, a successful read operation requires a magnetic charge above a certain threshold, called the **read threshold**.

The surface area allocated to store a bit is called the **areal density**, recording density, or bit density. Areal density is typically expressed in bits per inch or centimeter for tapes and bits per square inch or centimeter for disk platters. Areal density can be increased by reducing the surface area used to store a single bit value, thus increasing the total number of bits that can be stored on the medium. For example, the areal density of a two-dimensional medium such as a disk platter can be quadrupled by halving the length and width of each bit area (see Figure 5.7).

Increasing areal density also reduces coercible material per bit and risks having the stored charge fall below the read threshold. Therefore, areal density increases must be accompanied by more coercible medium coatings or more sensitive read/write electronics that can operate reliably with a lower read threshold. Increased areal density also requires

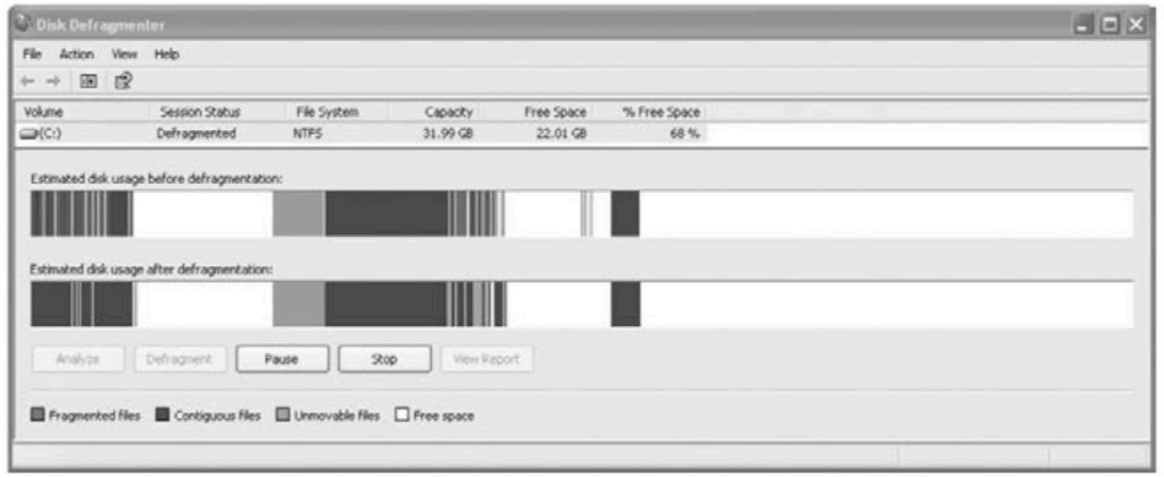

FIGURE 5.11 Display of fragmentation before and after disk defragmentation

A disk drive's data transfer rate is a summary performance number combining the physical aspects of data access with the electronic aspects of data transfer to the disk controller or computer system. As with access times, generally two different numbers are stated: maximum data transfer rate and sustained data transfer rate. The maximum data transfer rate is the fastest rate the drive can support and assumes sequential access to sectors. Because mechanical delays are minimal with sequential accesses, this number is quite high, typically several megabytes per second. The maximum data transfer rate for the previously described performance figures, assuming no controller delays, is as follows:

$$\frac{1}{0.000032 \text{ seconds}} \times 512 \text{ bytes per sector} \approx 15.3 \text{ MB per second}$$

The **sustained data transfer rate** is calculated with much less optimistic assumptions about data location. The most straightforward method is to use average access time. For the previous example, this is the sustained data transfer rate:

$$\frac{1}{0.00653} \times 512 \text{ bytes per sector} \approx 76.6 \text{ KB per second}$$

Because OSs typically allocate disk space to files sequentially, the sustained data transfer rate experienced during normal use is generally much higher than the preceding calculation implies. Table 5.3 lists performance statistics for three hard disk drive models from Seagate Technology, Inc.

TABLE 5.3 Hard disk drive performance statistics

Model	Use	Rotation speed (rpm)	Average access time (ms)	Maximum sustained DTR (MBps)
ST500LT012	Laptop	5400	12.5	118.5
ST1000DM003	Desktop	7200	8.5	156
ST600MP0005	Server	15,000	2.5	202

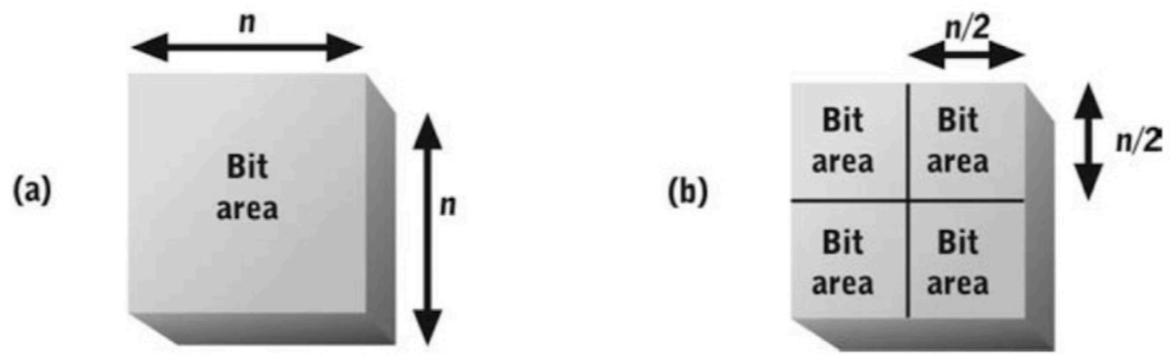

FIGURE 5.7 Areal density is a function of a bit area's length and width (a); density can be quadrupled by halving the length and width of bit areas (b)

more precise positioning of the read/write head over the storage medium. Note that these issues also impact optical disc storage drives and media.

Magnetic decay is the tendency of magnetically charged particles to lose their charge over time. Over enough time, stored charge decays below the read threshold and the data content of the storage medium is effectively lost. To minimize potential data loss, data bits must be written with a charge high enough to compensate for decay. Magnetic storage devices write data at a substantially higher charge than the read threshold, therefore ensuring long-term readability. However, there are limits to how much charge a read/write head can generate and how much the storage medium can hold.

The strength of a bit charge can also decrease because of **magnetic leakage** from adjacent bits. Any charged area continuously generates a magnetic field that affects nearby bit areas. If the polarity of adjacent bits is opposite, their magnetic fields tend to cancel out the charge of both areas, and the strength of both charges falls below the read threshold. Magnetic leakage is counteracted by lowering areal density (therefore increasing the charge per bit) or by placing uncharged buffer areas between bit areas. Unfortunately, both methods reduce storage capacity.

Magnetic storage device manufacturers carefully balance all of these factors to achieve cost-effective and reliable storage (see Table 5.2). Storage devices intended for different

TABLE 5.2 Factors that must be balanced in magnetic storage devices

Factor	Description
Magnetic decay	Natural charge decay over time; data must be written at a higher power than the read threshold to avoid data loss.
Magnetic leakage	Cancellation of adjacent charges of opposite polarity and migration of charge to nearby areas; data must be written at a higher power than the read threshold to avoid data loss.
Areal density	The coercible material per bit decreases as the areal density increases; higher areal density makes stored data more susceptible to loss caused by decay and leakage if all other factors are equal
Media integrity	Stability of coercible material and its attachment to the substrate; physical stress and extremes of temperature and humidity must be avoided to prevent loss of coercible material.
Device electronics and mechanics	Speed or capacity increases must be coupled with improvements in components that position the read/write heads and the storage medium.

TABLE 5.4 Technologies and storage formats for optical and magneto-optical storage

Technology/format	Writable?	Description
CD-ROM	No	Adaptation of musical CD technology; 650 or 700 MB capacity.
CD-R	One time only	CD-ROM format with a dye reflective layer that can be written by a low-power laser.
CD-RW	Yes	CD-ROM format with phase-change reflective layer; can be written up to 1000 times.
DVD-ROM	No	Adaptation of DVD video technology; similar to CD-ROM but more advanced; 4.7 GB (single layer) or 8.5 GB (dual layer) capacity.
DVD+/-R	One time only	DVD-ROM single- and dual-layer formats; similar to CD-R with improved performance and capacity. DVD-R and DVD+R are slightly different formats.
DVD+/-RW	Yes	DVD-ROM single- and dual-layer formats with phase-change reflective layer. DVD-RW and DVD+RW are slightly different formats.
BD	No	Trade name is Blu-ray disc. Improved DVD technology using higher wavelength lasers; 25 GB (single layer) or 50 GB (dual layer) capacity. Higher capacity discs with more than two layers are under development.
BD-R	One time only	Blu-ray version of DVD+/-R.
BD-RE	Yes	Blu-ray version of DVD+/-RW.
Magneto-optical	Yes	Outdated combination of magnetic and optical technologies; capacity up to 30 GB per disk.
HVD	Yes	Stores bits holographically in the optical recording layer. A 2007 multivendor standard specifies a 500 GB disc, but no drives are currently in production.

A CD is 120 millimeters (approximately 4.75 inches) in diameter and made largely of durable polycarbonate. Bit values are represented as flat areas called *lands* and concave dents called *pits* in the reflective layer. Data is recorded in a single continuous track that spirals outward from the disc's center. Error correction data written along the track at regular intervals helps the CD drive's read mechanism recover from errors introduced by dust, dirt, and scratches.

DVD is an acronym for both *digital video disc* and *digital versatile disc*. (The two terms are equivalent.) A consortium of companies developed DVD as a standard format for distributing movies and other audiovisual content. Like CD-ROM, **DVD read-only memory (DVD-ROM)** is an adaptation of audiovisual DVD in which DVD discs can store computer data. Most DVD-ROM drives can also read CD, CD-R, and CD-RW discs.

Blu-ray disc (BD), an important update to DVD-ROM, was originally designed for high-definition video discs. As with CD and DVD, the technology has been adapted to computer data storage. DVDs and BDs are the same size as CDs, and their read and write technologies are similar enough that storage and most drives are backward compatible with all earlier standards.

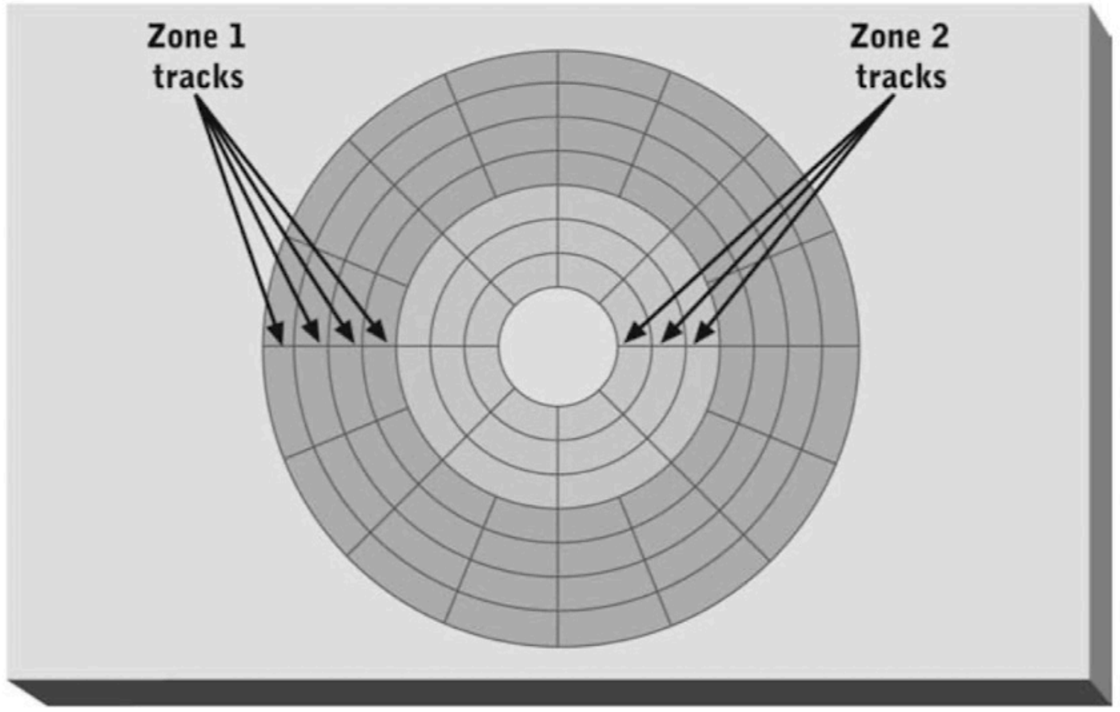

FIGURE 5.12 A platter divided into two zones with more sectors per track in Zone 1

Another factor that complicates performance calculations is that most manufacturers pack more sectors in the outer tracks of a platter. Look closely at Figure 5.10 again, and note that the surface area of each track and sector increases as you move from the spindle toward the outer platter edge. Coercible material per sector is much greater for the outermost track than for the innermost track. To increase capacity per platter, disk manufacturers divide tracks into two or more zones and vary the sectors per track in each zone (see Figure 5.12).

Computing the average access time is more complex when sectors are more densely packed on the platter's outer portions because the assumption that an average access requires moving the read/write head over half the tracks is no longer valid. For this reason, most average access time computations assume movement over a quarter to a third of the tracks. Also, more data is read per revolution when accessing the outer zones than the inner zones. Therefore, manufacturers sometimes quote sustained data transfer rates using a phrase such as *outer to inner diameter* with a corresponding numeric range.

Magnetic disk drive performance can be improved significantly by a number of caching techniques that are fully described in Chapter 6. Sustained data transfer rate improvements of 50 to 200% over the limits implied by drive mechanics and geometry are common for accesses to files as large as 16 MB. Magnetic drives with multi-gigabyte flash memory caches can provide even larger performance improvements, as described later in this chapter.

SOLID-STATE DRIVES

Declining cost and improved performance and capacity of flash RAM have created a new class of storage devices that are gradually supplanting magnetic disk as the dominant secondary storage technology. A **solid-state drive (SSD)** is a storage device that mimics the behavior of a magnetic disk drive but uses flash RAM or other NVM devices as the

DVD and BD technologies improve on CD technology in several ways:

- Increased areal density achieved with smaller wavelength lasers and more precise mechanical control
- Higher rotational speeds
- Improved error correction
- Multiple recording sides and layers

CD-ROMs and DVD-ROMs are popular media for distributing software and large data sets, such as maps, census data, phone directories, and other large databases. Standardized formats, high density, and cheap manufacturing cost make them well suited for these purposes. Their primary drawback is that discs can't be rewritten.

Recordable Discs

Recordable versions of the major optical disc formats (CD-R, DVD+R, DVD-R, DVD+R DL, DVD-R DL, and BD-R) use a laser that can be switched between high and low power and a laser-sensitive dye embedded in the disc. The dye is stable when scanned at low power during a read operation but changes its reflective properties when scanned at higher power during a write operation. The write operation is destructive, so recordable disc formats can be written only one time. Because the dyes are sensitive to heat and light, recordable discs should be stored in a dark location at room temperature. Also, because the dyes degrade over time, data might be lost after a few decades.

Phase-Change Optical Discs

Phase-change optical technology enables nondestructive writing to optical storage media. The technology is based on the same GST material used in MRAM, which can change state easily from amorphous to crystalline and then back again. The reflective characteristics of this material are quite different in the amorphous and crystalline states. The difference is less than with manufactured or dye-based discs but enough to be detected by newer optical scanning technologies.

GST changes from an amorphous state to a crystalline state when heated to a precise temperature. Heating the material to its melting point changes it back to an amorphous state. The melting point is low, so high-power lasers aren't required. However, multiple passes are usually required to generate enough heat, so write times are substantially longer than read times. The reflective layer loses its capability to change state with repeated heating and cooling. Current rewritable media wear out after about 1000 write operations.

CD-Rewritable (CD-RW) is a standard for phase-change optical discs that use the CD-ROM format. Rewritable DVDs use one of four different standards: DVD+RW, DVD-RW, DVD+RW DL, and DVD-RW DL. BD-RE is the rewritable version of the Blu-ray standard.

Magneto-Optical

A **magneto-optical (MO) drive** uses a laser and reflected light to sense magnetically recorded bit values. Reading is based on the polarity of the reflected laser light, which is determined by the polarity of the magnetic charge. This laser polarity shift is known as the Kerr effect.

uses can balance these factors differently to best match a specific need or market. For example, one device might sacrifice storage capacity (areal density) to gain longer storage life. Another might balance factors differently to gain speed at the cost of capacity or longevity.

Magnetic Tape

A **magnetic tape** is a ribbon of plastic with a coercible (usually metallic oxide) coating. Tapes are mounted in a **tape drive**, which is a slow serial access device, for reading and writing. A tape drive contains motors that wind and unwind the tape, one or more read/write heads, and related circuitry. Tapes are used primarily to make backup copies of data stored on faster secondary storage devices and to physically transport large data sets.

Tapes and tape drives come in a variety of shapes and sizes. Early tape drives used tapes up to 1 inch wide wound on large open reels up to 10 inches in diameter. Modern tape drives use much narrower tapes mounted permanently in plastic cassettes or cartridges (see Figure 5.8). When a cassette is inserted into a tape drive, an access door opens automatically, and the read/write head is positioned around or next to the exposed tape.

Compared to magnetic disk, tapes are much more susceptible to the problems described in the previous section. Magnetic leakage can occur from adjacent bit positions on the same area of the tape as well as from the layer of tape wound above or below on the reel. Also, as a tape is wound and unwound, it tends to stretch. If enough stretching occurs, the distance between bit positions can be altered, making bits difficult or impossible to read. Stretching can combine with temperature and humidity variation to reduce the long-term integrity of the storage medium.

The role of magnetic tape is rapidly shrinking in modern computer systems. Significant improvements in tape technology and standards have slowed due to declining market share. High-speed networking, rapid decreases in the relative cost of magnetic disk storage,

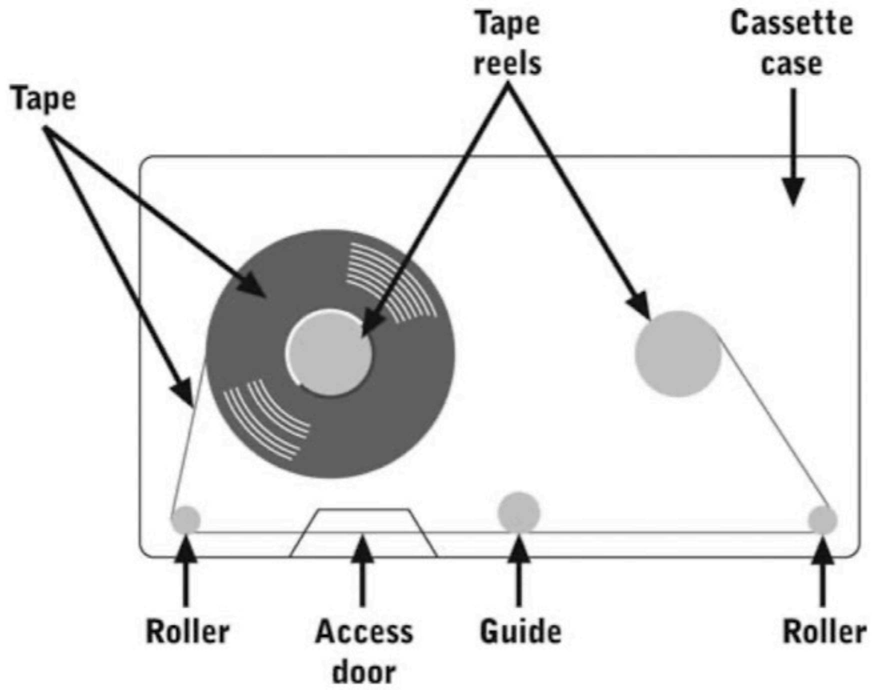

FIGURE 5.8 Components of a typical cassette or cartridge tape

storage medium and read/write mechanism. For backward compatibility, most SSDs have the same physical size and interfaces as magnetic drives, which enables them to compete directly and be used interchangeably. For higher performance in servers and workstations, SSDs can also be purchased as add-in cards that connect directly to the system bus.

Current SSDs use flash RAM as the storage medium, although newer NVM types, such as MRAM and PCM, might supplant it in the future. At present, the advantages and disadvantages of SSDs compared to other secondary storage technologies, especially magnetic disks, are byproducts of the current state of flash RAM technology. However, the comparisons are expected to change rapidly under the assumption that magnetic disks are a more mature technology than NVM. The rate of improvement in SSD characteristics is expected to be faster than for magnetic disks, resulting in magnetic disks being gradually replaced with SSDs.

Speed

By avoiding the mechanics of spinning platters and access arms, SSDs can achieve significantly higher random access times than magnetic disks. Fast access times translate into high data transfer rates. For SSDs in desktop and portable computers, random read data transfer rates of 250 to 350 MBps are typical. Write data transfer rates are typically 80% of read data transfer rates, and sequential data transfer rates are typically 50 to 75% higher than random rates. Those rates are roughly two to three times higher than typical desktop magnetic hard drives for sequential accesses and 10 to 100 times higher for random accesses.

Volatility

A typical magnetic hard drive has a lifespan of 5 to 10 years and can perform an unlimited number of read and write operations during that lifetime. In read-only mode, an SSD can last for decades. But the destructive nature of write operations to flash RAM limits general-purpose lifetimes. The actual lifetime depends on the underlying flash RAM technology and the quantity and distribution of write operations.

Individual flash RAM memory cells wear out after approximately 100,000 writes. By itself, that would imply an SSD drive lifetime of a few months to a year or two. But SSDs employ a technique called wear-leveling to extend that lifetime. An SSD tracks the write frequency to each memory region and *levels out* the impact of write operations by distributing them evenly throughout the device. An SSD also reserves some memory regions to replace regions that wear out prematurely. This technique extends the life of a typical SSD to approximately five years in general-purpose use, though that can vary widely depending on use characteristics.

The destruction of flash RAM cells resulting from write operations doesn't occur all at once. Some cells fail well before 100,000 writes and others last longer. In addition, read operations slow down or fail as a memory cell approaches the end of its life. Therefore, SSDs need to use error correction codes (covered in Chapter 8) to enable recovery of partially corrupt data blocks and movement of their content to spare storage locations. The net impact of these characteristics and techniques is a gradual decrease in performance over an SSD's lifetime. In contrast, the performance of a magnetic drive is virtually unchanged from day 1 to the end of its life.

The disk surface is a material that accepts a magnetic charge when heated to approximately 150 degrees Celsius. To bring the material to a chargeable temperature, a high-power laser is focused on a bit position. A magnetic write head positioned below the media surface then generates the appropriate charge. Because the laser beam focuses on a single bit position, only that bit position is raised to a sufficient temperature. Surrounding bits are unaffected by the magnetic field from the write head because they're cooler.

Magneto-optical technology peaked in the mid-1990s. Newer optical formats such as CD, DVD, and BD drives provide cheaper alternatives though not necessarily with significantly higher performance. However, there's still a market for MO drives because many organizations created data archives on MO disks and still need access to these archives.

CLOUD-BASED STORAGE

Though this chapter focuses on storage hardware, no discussion of modern storage would be complete without at least mentioning cloud-based storage. Cloud-based storage services come in many forms, but they have common features and underlying principles:

- A service provider manages large amounts of storage hardware in remote locations shared by many users, therefore realizing significant economies of scale.
- Data is stored in multiple remote locations to protect against loss and ensure availability when and where needed.
- High-speed network connections among local computer systems and storage service providers enable rapid movement of data.
- Data stored on local storage devices is frequently or continuously synchronized with copies stored in remote locations.
- Software on both local and remote computer systems keeps track of all data copies and coordinates data movement among local and remote storage locations as needed.
- Users aren't aware of (and don't care about) the work that software does behind the scenes to ensure data protection and availability.

Although computing, storage, and network hardware are the enablers of cloud-based storage, its implementation is primarily driven by software in general and operating systems in particular. As such, we'll delay a more detailed discussion of this technology until Chapter 12, after we've covered networks and operating systems in detail.

and cloud-based storage services have all contributed to this decline. Tape's role as a backup storage medium is increasingly filled by redundant copies stored on both on- and off-premises disks.

Magnetic Disk

Magnetic disk media are flat, circular **platters** with metallic coatings that are rotated beneath read/write heads (see Figure 5.9). Data is normally recorded on both sides of a platter, and multiple platters are mounted and rotated on a common spindle. A **track** is one concentric circle of a platter, or the surface area that passes under a read/write head when its position is fixed. A **cylinder** consists of all tracks at an equivalent distance from the edge or spindle on all platter surfaces.

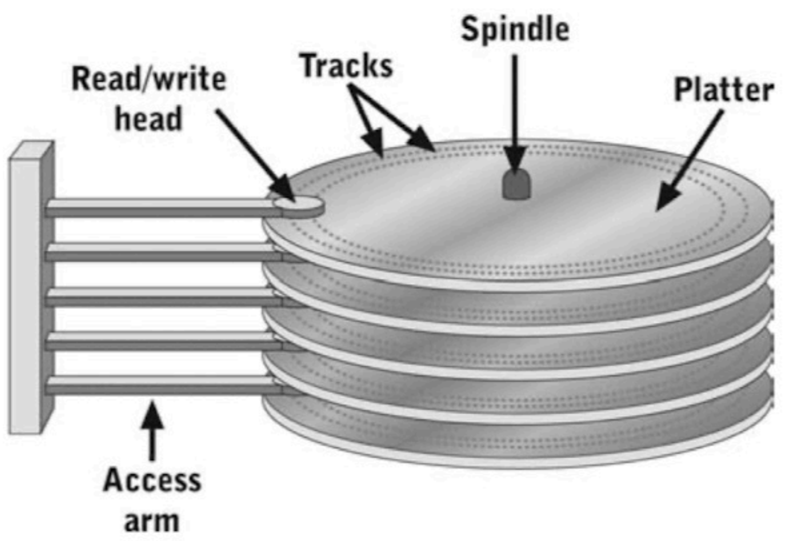

FIGURE 5.9 Primary components of a typical magnetic disk drive

A read/write head is mounted on an **access arm** for each side of each platter. Access arms of equal length are attached to a positioning servo. This servo moves the access arms so that read/write heads can be positioned anywhere between the outermost track, closest to the platter's outer edge, and the innermost track, closest to the spindle. Read/write heads float above the platter on a thin cushion of air as the disk spins.

A sector is a fractional portion of a track (see Figure 5.10). Until recently, a single sector usually held 512 bytes, which was also the normal data transfer unit to or from the drive. Many disks manufactured since 2010 have a 4096-byte sector size.

A magnetic **disk drive** is a single device with rigid platters, access arms and read/write heads, a motor to spin the platters, a servo to move the access arms, and all of the electronics needed to control the drive, read and write sectors, and communicate with the system bus. Currently, platter diameter is between 1 and 3 inches. Platters are spun at high speeds, up to 15,000 revolutions per minute (rpm). Drive capacity depends on the number of platters, platter size, and areal density. At this writing (early 2015), hard disk drive capacities range up to 6 TB, but maximum capacity typically doubles every two years. Multiple disk drives can be enclosed in a single storage cabinet; this arrangement is referred to as a **drive array**.

Portability

Lacking mechanical components, SSDs are much more tolerant to shock and other negative environmental factors, and they generate far less vibration. Higher shock tolerance makes them much more reliable in portable devices such as cell phones, tablets, and laptop computers. Magnetic disk drives used in portable applications typically have much lower performance than in fixed applications to compensate for environmental factors. Reduced vibration is an advantage in some server applications that use densely packed storage arrays. The vibrations of dozens of magnetic drives in close proximity can create interference and shorten drive lifetimes.

Power Consumption

The lack of motors and servos in SSDs significantly reduces their power consumption compared to magnetic drives of similar capacity. As flash RAM fabrication scales to smaller sizes, power requirements drop. Lower power requirements are a significant advantage in battery-powered applications and to a lesser extent in densely packed storage arrays.

Capacity and Cost

Other than volatility, the primary comparative disadvantage for SSDs is cost. As of this writing (early 2015), the cost of magnetic disk storage ranges from 10¢ per GB for desktop computers to $1.50 per GB for high-performance servers. Capacity per drive ranges up to 6 TB. In contrast, SSD cost ranges from $1.50 per GB for low-performance portable computers to $4 per GB for higher performance high-capacity server drives. Capacity per drive ranges up to 2 TB.

Hybrid Magnetic/Solid-State Drives

Whenever two storage technologies have significant cost, capacity, and performance differences, information system managers and computer and device manufacturers attempt to combine the technologies in a way that provides the advantages of each at a reasonable cost. Managers of large servers and server farms often do this today by combining terabytes to exabytes of magnetic disk capacity with gigabytes to terabytes of SSD capacity. The SSDs hold data that requires the fastest possible access, whereas magnetic disks hold data that isn't accessed as frequently or isn't needed as quickly. Of course, system managers must determine which data will reside on which devices and must periodically reallocate data as usage patterns and access needs change.

Storage device manufacturers have adopted this same technique on a smaller scale. A **hybrid disk drive** combines flash RAM and magnetic disk storage in a single device. Flash RAM capacity typically ranges from a few to a few dozen gigabytes, and magnetic disk capacity typically ranges from 500 MB to 4 TB. The drive contains a dedicated processor and sophisticated firmware that tracks data reads and writes. The device allocates data to flash RAM or magnetic disk according to its access pattern and frequency. Frequently read data that's seldom changed is held in flash RAM for faster access. Examples of such content include operating system device drivers, management utilities, and user interface routines. Other examples include frequently used programs such as Web browsers and frequently accessed documents and other data.

BUSINESS FOCUS

Which Storage Technology?

The Ackerman, Holt, Sanchez, and Trujillo (AHST) law firm has thousands of clients, dozens of partners and associates, and a costly document storage and retention problem. Before 1990, AHST archived legal documents and billing records on paper. As in-office file cabinets filled, the paper was boxed, labeled, and transported to a warehouse for storage. AHST retains all records for at least 20 years and longer in many cases.

In the late 1980s, AHST adopted automated procedures for archiving documents. Most documents were stored as computer files in their original format (WordPerfect, at that time). Documents with signatures were scanned and stored as JPEG files though paper copies were also saved because many courts didn't (and still don't) recognize that validity of copied signatures. Several possible storage media were considered, including different forms of magnetic tape and removable MO disks. The firm chose Sony MO drives and disks based on the following selection criteria, listed roughly in the order of importance:

- Permanence (25-year minimum usable lifetime)
- Cost
- High storage capacity
- Ease of use
- Stability and market presence of the technology and selected vendor(s)

AHST has used several different MO media, including 600 MB, 2.3 GB, and 9.1 GB 5.25-inch disks, and several different drives. Document formats have also changed to include Microsoft Word and Adobe Acrobat files. In 2009, AHST began recording depositions with a digital video camera and archiving them as MPEG files stored on 9.1 GB MO disks. AHST estimates its current annual storage requirements for document and video deposition archives at 10 to 25 TB, approximately 90% of which is video.

Electronic document and secondary storage technologies have evolved considerably since AHST chose MO storage. Although MO technology evolved through 2010, there have been no new products or technology updates since that time. Drives and media are becoming harder to find, and costs are high compared with more modern alternatives. For these reasons, the firm is reassessing the choice of technology for permanent document storage. Many new formats have been developed and cloud-based storage has become commonplace. To date, AHST has resisted cloud-based storage because of concerns over reliability and privacy.

Questions:

- Describe the pros and cons of the removable secondary storage technologies described in this chapter with respect to the firm's selection criteria.
- Which technologies should AHST adopt for future document and video storage? Why?
- Should older archived content be migrated to the new technologies? Why or why not?
- How and when might the technologies you recommended become obsolete?

Portability

Storage device portability is typically implemented in one of two ways:

- The entire storage device—storage medium, read/write mechanism, and possibly controller—can be transported between computer systems (for example, a USB flash drive).
- The storage medium can be removed from the storage device and transported to a compatible storage medium on another computer (for example, a DVD).

Portable secondary storage devices and devices with removable storage media typically have slower access speed than permanently installed devices and those with nonremovable media. High-speed access requires tight control of environmental factors and high-speed communication channels to connect the device to the system bus. For example, high-speed access for magnetic disks is achieved, in part, by sealed enclosures for the storage media, therefore minimizing or eliminating dust and air density variations. Removable storage media come in contact with a wider variety of damaging environmental influences than isolated media do. Also, an internal magnetic disk drive is usually connected to the system bus by a high-speed channel, whereas a portable magnetic disk drive typically uses a slower channel, such as FireWire or USB.

Cost and Capacity

Each storage device attribute described so far is related to device cost per unit of storage (see Table 5.1). For example, improving speed, volatility, or portability increases cost per unit if all other factors are held constant. Cost per unit also increases as an access method moves from serial to random to parallel. Primary storage is generally expensive compared with secondary storage because of its high speed and combination of parallel and random access methods.

Secondary storage devices are usually thought of as having much higher capacity than primary storage devices. In fact, the capacity difference between primary and secondary storage in most computers results from a compromise between cost per unit and other

TABLE 5.1 Storage device characteristics and their relationship to cost

Characteristic	Description	Cost
Speed	Time required to read or write a bit, byte, or larger unit of data	Cost increases as speed increases.
Volatility	Capability to hold data indefinitely, particularly in the absence of external power	For devices of a similar type, cost decreases as volatility increases.
Access method	Can be serial, random, or parallel; parallel devices are also serial or random access	Serial is the least expensive; random is more expensive than serial; parallel is more expensive than nonparallel.
Portability	Capability to easily remove and reinstall storage media from the device or the device from the computer	For devices of a similar type, portability increases cost, if all other characteristics are held constant.
Capacity	Maximum data quantity the device or storage medium holds	Cost usually increases in direct proportion to capacity.

device characteristics. For example, if cost weren't a factor, most users would opt for solid-state drives rather than magnetic disk drives. However, solid-state drives with terabytes of capacity cost more than many users can afford. So users sacrifice speed and parallel access to gain the capacity they need at an acceptable cost.

Memory-Storage Hierarchy

A typical computer system has a variety of primary and secondary storage devices. The CPU and a small amount of high-speed RAM usually occupy the same chip. Slower RAM on separate chips composes the bulk of primary storage. One or more magnetic disk or solid-state drives are usually complemented by an optical disk drive and at least one form of removable magnetic storage.

The range of storage devices in a single computer system forms a memory-storage hierarchy, as shown in Figure 5.3. Cost and access speed generally decrease as you move down the hierarchy. Because of lower cost, capacity tends to increase as you move down the hierarchy. A computer designer or purchaser attempts to find an optimal mix of cost and performance for a particular purpose.

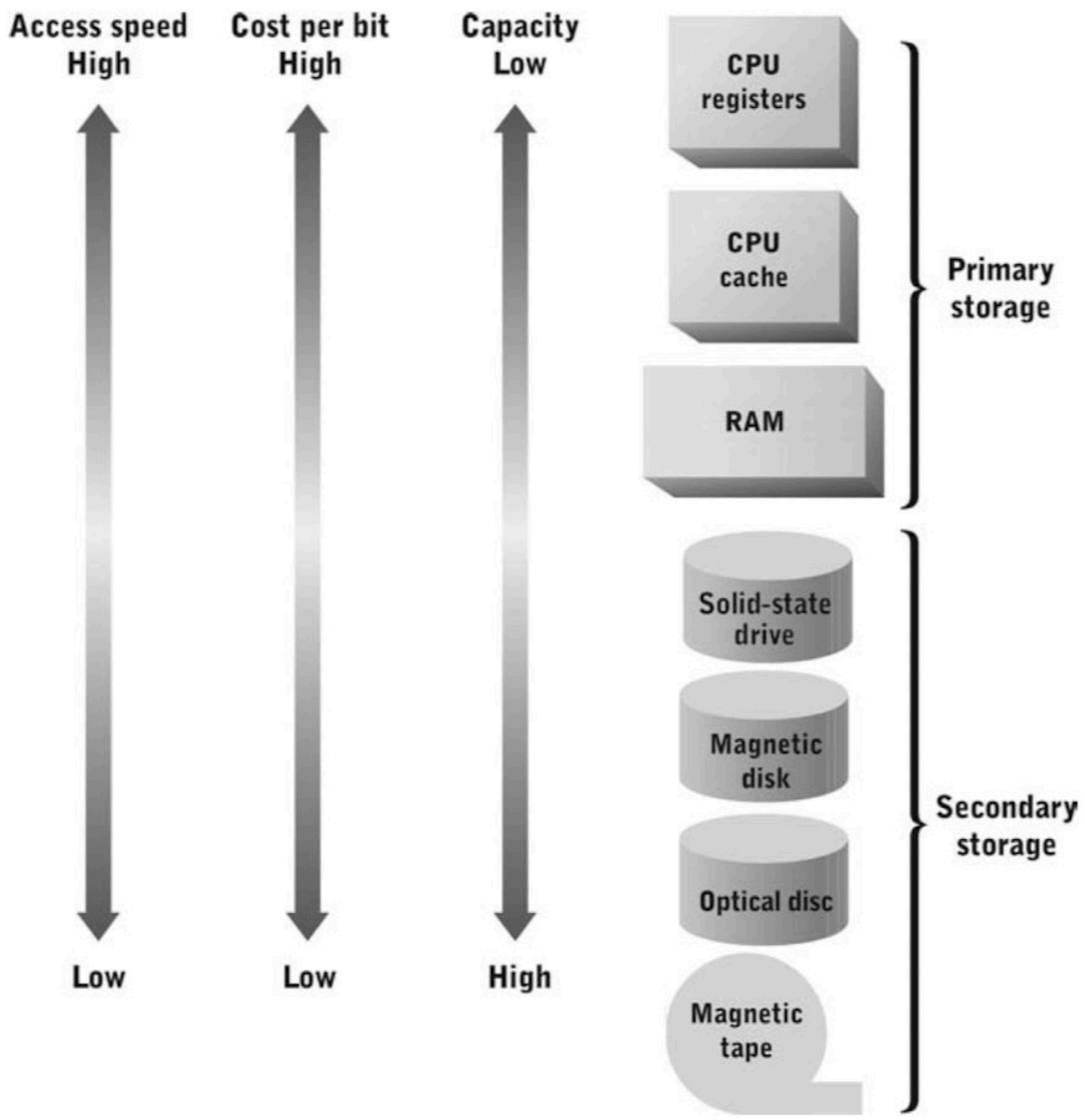

FIGURE 5.3 Comparison of storage devices in terms of cost and access speed

PRIMARY STORAGE DEVICES

As discussed, the critical performance characteristics of primary storage devices are access speed and data transfer unit size. Primary storage devices must closely match CPU speed and word size to avoid wait states. CPU and primary storage technologies have evolved in tandem—in other words, CPU technology improvements are applied to the construction of primary storage devices.

Storing Electrical Signals

Data is represented in the CPU as digital electrical signals, which are also the basis of data transmission for all devices attached to the system bus. Any storage device or controller must accept electrical signals as input and generate electrical signals as output.

Electrical power can be stored directly by various devices, including batteries and capacitors. Unfortunately, there's a tradeoff between access speed and volatility. Batteries are quite slow to accept and regenerate electrical current. With repeated use, they also lose their capability to accept a charge. Capacitors can charge and discharge much faster than batteries. However, small capacitors lose their charge fairly rapidly and require a fresh injection of electrical current at regular intervals (hundreds or thousands of times per second).

An electrical signal can be stored indirectly by using its energy to alter the state of a device, such as a mechanical switch, or a substance, such as a metal. An inverse process regenerates an equivalent electrical signal. For example, electrical current can generate a magnetic field. The magnetic field's strength can induce a permanent magnetic charge in a nearby metallic compound, therefore *writing* the bit value to the metallic compound. To read the stored value, the stored magnetic charge is used to generate an electrical signal equivalent to the one used to create the original magnetic charge. Magnetic polarity, which is positive or negative, can represent the values 0 and 1.

Modern computers use memory implemented with semiconductors. Basic types of semiconductor memory include random access memory and nonvolatile memory. There are many variations of each memory type, described in the following sections.

Random Access Memory

Random access memory (RAM) is a generic term describing primary storage devices with the following characteristics:

- Microchip implementation with semiconductors
- Capability to read and write with equal speed
- Random access to stored bytes, words, or larger data units

RAM is fabricated in the same manner as microprocessors. You might assume that microprocessor clock rates are well matched to RAM access speeds. Unfortunately, this isn't the case for many reasons, including the following:

- Reading and writing many bits in parallel requires additional circuitry.
- When RAM and microprocessors are on separate chips, there are delays when moving data from one chip to another.

There are two basic RAM types and several variations of each type. **Static RAM (SRAM)** is implemented entirely with transistors. The basic storage unit is a flip-flop circuit (see Figure 5.4). Each flip-flop circuit uses two transistors to store 1 bit. Additional transistors (typically two or four) perform read and write operations. A flip-flop circuit is an electrical switch that remembers its last position; one position represents 0 and the other represents 1. These circuits require a continuous supply of electrical power to maintain their positions. Therefore, SRAM is volatile unless a continuous supply of power can be guaranteed.

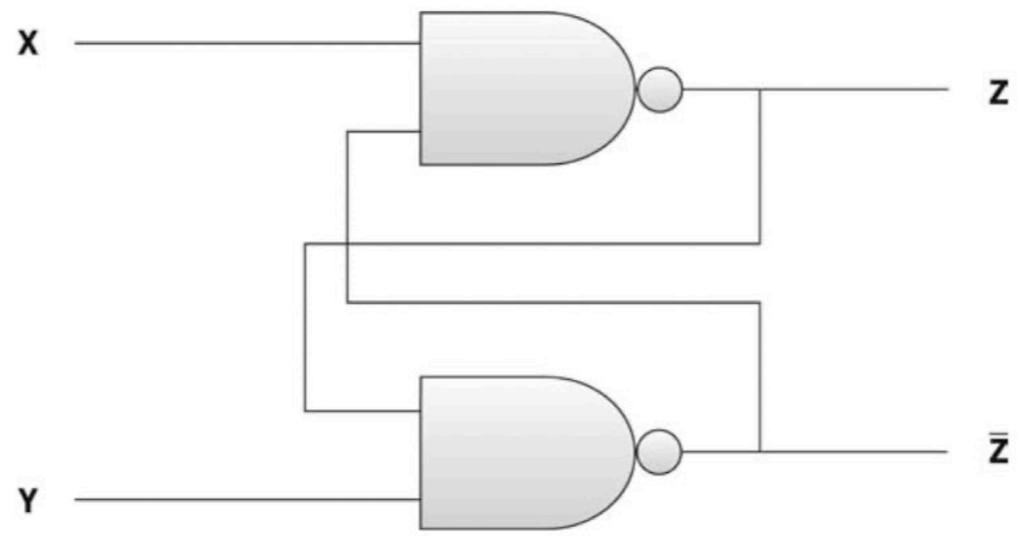

FIGURE 5.4 A flip-flop circuit composed of two NAND gates: the basic component of SRAM and CPU registers

Dynamic RAM (DRAM) stores each bit by using a single transistor and capacitor. Capacitors are the dynamic element. However, because they lose their charge quickly, they require a fresh infusion of power thousands of times per second, so DRAM chips include circuitry that performs refresh operations automatically. Each refresh operation is called a **refresh cycle**. Unfortunately, a DRAM chip can't perform a refresh operation at the same time it performs a read or write operation.

Because DRAM circuitry is simpler, more memory cells can be packed into each chip. (In other words, DRAM has higher density.) In contrast, fewer SRAM bits can be implemented in each chip, effectively making SRAM more expensive than DRAM. Despite its simpler circuitry, DRAM is slower than SRAM because it requires refresh cycles and because it uses less efficient circuitry for accessing bits. With current fabrication technology, typical access times are 10 to 20 ns for DRAM and 1 to 5 ns for SRAM. Improvements in fabrication technology can decrease both access times but can't change the performance difference.

Note that neither RAM type can match current microprocessor clock rates, which range up to 5 GHz at this writing. For zero wait states in memory accesses, a CPU with a 5GHz clock rate requires the following memory access speed:

$$\frac{1}{5\text{ GHz}} = \frac{1}{5{,}000{,}000{,}000} = 0.0000000002 = 0.2\text{ ns}$$

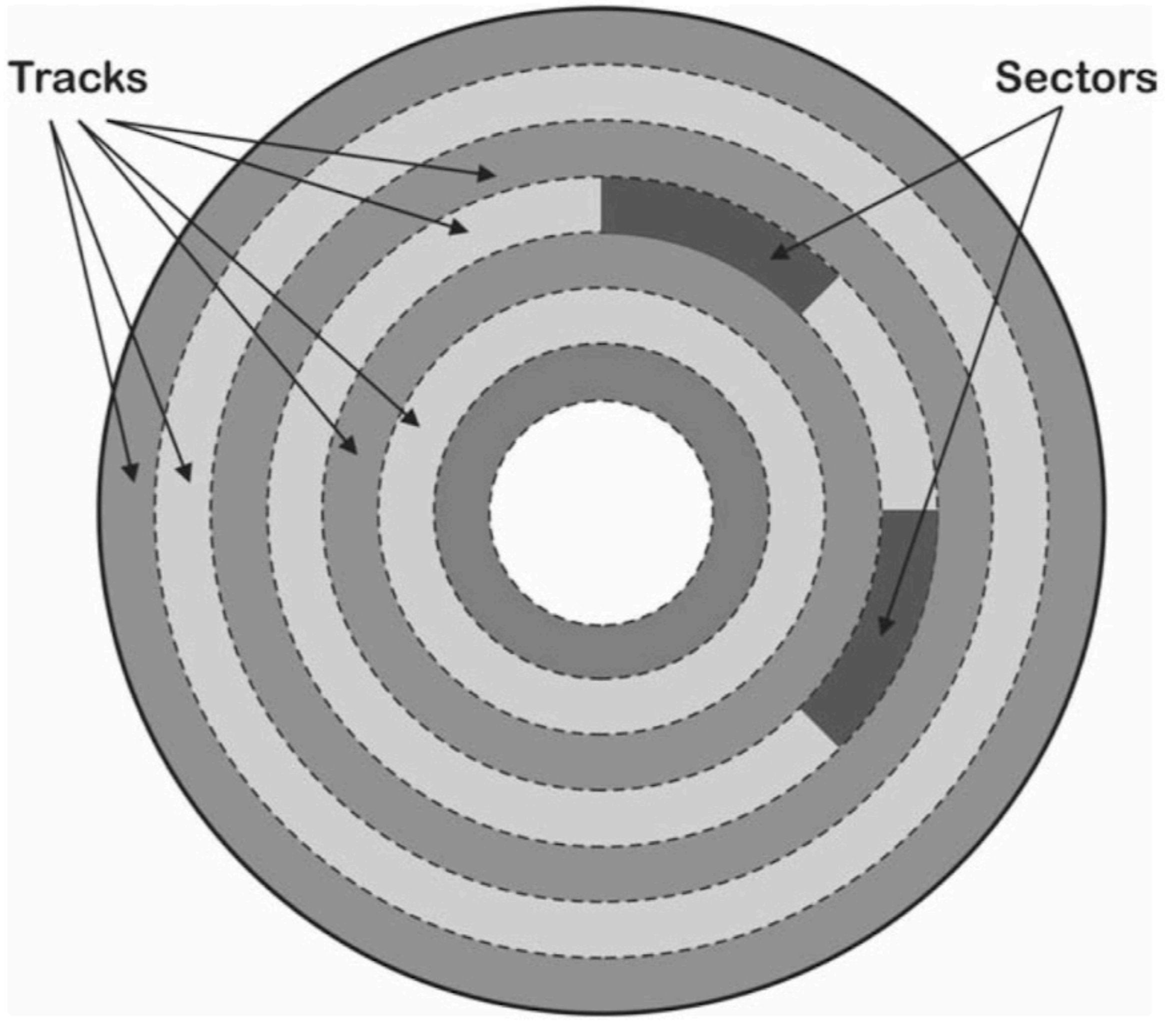

FIGURE 5.10 Organization of tracks and sectors on one surface of a disk platter

A read or write command sent to the disk drive identifies a specific storage location (sector), and the drive's electronics determine the corresponding platter, track in the platter, and sector in the track. Controller circuitry switches activate the read/write head for the platter, and the positioning servo moves the access arms over the track. The controller circuitry then waits for the correct sector to rotate beneath the read/write head; at this point, the read or write operation is started. The operation is completed within microseconds as the last portion of the sector passes under the read/write head. The access arms are typically left in their last position until the next read or write command is received.

Access Time

Disk access time depends on several factors, including the time required to do the following:

- Switch between read/write heads
- Position read/write heads over a track
- Wait for the correct sector to rotate beneath read/write heads

Disk drives share one set of read/write circuits among all read/write heads. The read/write circuitry must be electronically attached, or switched, to the correct read/write head before a sector can be accessed. The time needed to perform this switching, referred to as

This approach to allocating data across two storage pools of differing speed, cost, and performance is called caching. Caching is described in depth in Chapter 6.

TECHNOLOGY FOCUS

Personal vs. Data Center SSDs

The usage patterns of disk drives vary widely by computer class. Most personal and desktop computing devices are idle much of the time. Even when they're in use, their disk drives are still idle while users perform tasks that require little secondary storage access. In contrast, disk drives installed in servers and data centers are usually busy around the clock.

With both magnetic disk drives and SSDs, price per GB varies considerably among devices intended for the personal/desktop market and the server/data center market. For SSDs, which are already much more expensive than magnetic disk drives, the price difference is 3:1 or more. So what SSD features does the extra money buy and what technology improvements provide those features?

Server and data center managers are usually willing to pay more money for better performance if the intended application demands it. Many applications such as e-commerce Web sites and data analytics require access to large amounts of data stored on secondary storage devices. The slower the devices, the more time server CPUs sit idle waiting for read operations to complete, and the more time end users stare at screens that haven't yet updated.

Because SSDs are based on semiconductors, their speed depends directly on their size. Shrinking the size of flash RAM cells and traces speed up data access. Flash RAM is a commodity produced by multiple manufacturers on a variety of semiconductor process sizes. For example, at this writing (2015) flash RAM chips are produced with fabrication processes ranging from 16 to 45 nanometers, representing three generations of fabrication technology. Smaller fabrication processes are more expensive, but SSD manufacturers that buy flash RAM chips produced on those processes can provide devices with improved access time and data transfer rates at higher per-GB cost.

Another factor that impacts SSD performance, capacity, and cost is the number of bits stored per flash RAM cell. **Single-level cells (SLCs)** store a single bit value. **Multilevel cells (MLCs)** store two bits per cell. Both types are currently mass produced. Cells that store three or more bits have also been produced though they aren't widely used at present. For the same number of cells, an SSD using MLCs stores twice as much data at roughly the same cost. However, MLC performance and reliability are lower than for SLCs. Read and write times are typically twice as long, MLCs typically fail after 10,000 writes, and they generate read errors more frequently before they fail. Given the higher performance and reliability requirements of servers and data centers, SSDs for that market are much more likely to use SLCs even though they have lower capacity for the same cost.

Another flash RAM technology that impacts performance and capacity is three-dimensional (3D), stacked, or **vertical flash RAM**. Compared to conventional 2D or planar flash RAM, vertical flash RAM adds capacity by using many more layers on a chip to stack cells atop another. The technique can be used with both SLCs and MLCs. Some flash RAM manufacturers have already switched to this type of memory to improve both capacity and speed. But the fabrication process is much more complex and expensive. Again, server and data center customers are often willing to pay the extra price for this new technology as long as it delivers improved performance and capacity.

(continued)

Summary

- A typical computer system has multiple storage devices filling a variety of needs. Primary storage devices support immediate execution of programs, and secondary storage devices provide long-term storage of programs and data. The characteristics of speed, volatility, access method, portability, cost, and capacity vary among storage devices, forming a memory-storage hierarchy. Primary storage tends to be at the top of the hierarchy, with faster access speeds and higher costs per bit of storage, and secondary storage tends to be in the lower portion of the hierarchy, with slower access speeds and lower costs.
- The critical performance characteristics of primary storage devices are their access speed and the number of bits that can be accessed in a single read or write operation. Modern computers use memory implemented with semiconductors. Basic types of memory built from semiconductor microchips include random access memory (RAM) and nonvolatile memory (NVM). The primary RAM types are static and dynamic. Static RAM (SRAM) is implemented entirely with transistors; dynamic RAM (DRAM) is implemented with transistors and capacitors. SRAM's more complex circuitry is more expensive but provides faster access times. Neither type of RAM can match current microprocessor clock rates. NVM is usually relegated to secondary storage roles because of its slower write speeds and limited number of rewrites.
- Magnetic secondary storage devices store data bits as magnetic charges. Magnetic disks are platters coated with coercible material. Platters are rotated in a disk drive, and read/write heads access data at various locations on the platters. Magnetic disk drives are random access devices because the read/write head can be moved to any location on a disk platter. They face increasing competition from solid-state drives based on flash RAM and other forms of nonvolatile memory. Hybrid drives combine both storage types in a single device to cost-effectively achieve the best characteristics of each.
- Optical discs store data bits as variations in light reflection. An optical disc drive reads data bits by shining a laser beam onto a small disc location. High and low reflections of the laser are interpreted as 1s and 0s. The cost per bit of optical storage is less than magnetic storage, at the expense of slower access speed. Types of optical discs (and drives) include different forms of CDs and DVDs. CD-ROM, DVD-ROM, and BD are written during manufacture. CD-R, DVD+/-R, and BD-R are manufactured blank and can be written to once. CD-RW and rewritable DVD and Blu-ray formats use phase-change technology to write and rewrite media.

Now that CPUs, primary storage, and secondary storage have been covered in some detail, it's time to turn your attention to how these devices are integrated in a computer system. Chapter 6 describes how processing, I/O, and storage devices interact and discusses methods of improving overall computer system performance. Chapter 7 describes types of I/O devices.

The fastest DRAM is one to two orders of magnitude slower than modern microprocessors. SRAM is up to one order of magnitude slower. The following technologies are used to bridge the performance gap between memory and microprocessors:

- Read-ahead memory access
- Synchronous read operations

Computer memory contains more than just SRAM or DRAM circuits. Memory circuits are grouped in memory modules containing tens or hundreds of megabytes. Each module contains additional circuitry to implement read and write operations to random locations—circuitry that must be activated before data can be accessed. Activating the read/write circuitry consumes time, which lengthens the time required to access a memory location.

Programs usually access instructions and data items sequentially. Read-ahead memory access takes advantage of this tendency by activating the read/write circuitry for location $n+1$ during or immediately after an access to memory location n. If the CPU subsequently issues an access command for location $n+1$, the memory module has already performed part of the work needed to complete the access.

Synchronous DRAM (SDRAM) is a read-ahead RAM that uses the same clock pulse as the system bus. Read and write operations are broken into a series of simple steps, each of which can be completed in one bus clock cycle (similar to pipelined processing, described in Chapter 4). Several clock cycles are required to complete a single random access. However, if memory is being written or read sequentially, accesses occur once per system bus clock tick after the first access has completed.

SDRAM has improved steadily with a series of technologies that doubled the data transfer rate of the previous technology. These technologies are called **double data rate (DDR)** and have been named DDR, DDR2, DDR3, and DDR4. DDR4 SDRAM supports data transfer rates up to 6 GBps.

Nonvolatile Memory

Memory manufacturers have worked for decades to develop semiconductor and other forms of RAM with long-term or permanent data retention. The generic terms for these memory devices are **nonvolatile memory (NVM)** and **read-only memory (ROM)**. Of course, manufacturers and consumers would like NVM to cost the same or less than conventional SRAM and DRAM and have similar or faster read/write access times. So far, those goals have proved elusive, but some current and emerging memory technologies show considerable promise.

Even though NVM currently lacks the access speed of RAM, it has many applications in computing and other industries, including storage of programs and data in portable devices, such as tablet computers and cell phones; portable secondary storage in devices such as digital cameras; and permanent program storage in motherboards and peripheral devices. One of the earliest uses in computers is storing programs such as boot subroutines—the system BIOS, for example. These instructions can be loaded at high speed into main memory from NVM. Software stored in NVM is called **firmware**.

The most common NVM in use today is **flash RAM** (also called *flash memory*). It's competitive with DRAM in storage density (capacity) and read performance. Unfortunately, CHAPTER 5

DATA STORAGE TECHNOLOGY

LEARNING OBJECTIVES

After studying this chapter, you should be able to:

- Describe the distinguishing characteristics of primary and secondary storage
- Describe the devices used to implement primary storage
- Compare secondary storage alternatives
- Describe factors that determine storage device performance
- Choose appropriate secondary storage technologies and devices

OVERVIEW

In Chapter 2, you were briefly introduced to the topic of storage, including the role of storage in a computer system and the differences between primary and secondary storage. In this chapter, you explore storage devices and their underlying technologies in depth (see Figure 5.1). This chapter also outlines characteristics common to all storage devices and compares primary and secondary storage technologies.

STORAGE DEVICE CHARACTERISTICS

A storage device consists of a read/write mechanism and a **storage medium**. The storage medium is the device or substance that actually holds data. The read/write mechanism is the device used to read or write data to and from the storage medium. A device controller provides the interface between the storage device and the system bus.

NOTE

Device controllers are discussed in detail in Chapter 6.

In some storage devices, the read/write mechanism and storage medium are a single unit using the same technology. For example, most types of primary storage use electrical circuits implemented with semiconductors for both the read/write mechanism and the storage medium. In other storage devices, the read/write mechanism and the storage

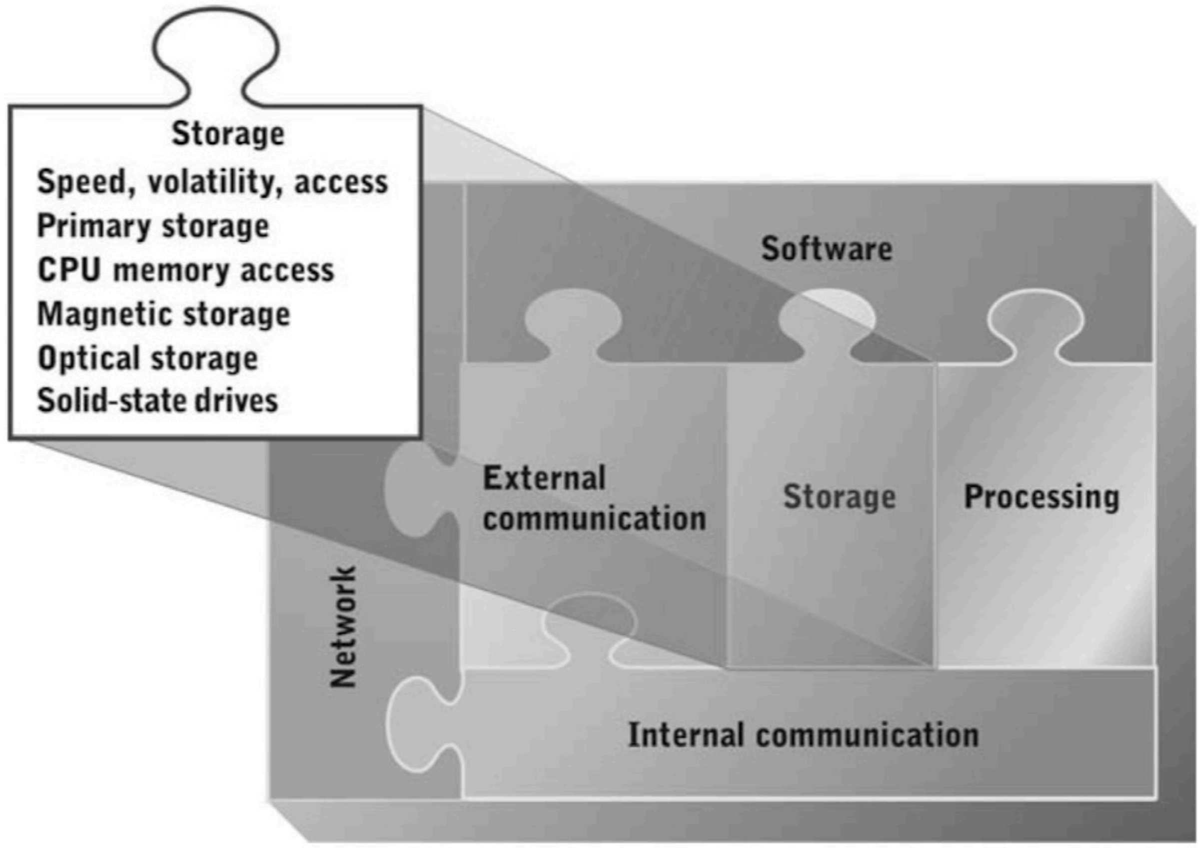

FIGURE 5.1 Topics covered in this chapter

medium use fundamentally different technologies. For example, magnetic disk drives use an electromechanical device for the read/write mechanism and a magnetic storage medium composed of metallic alloys.

A typical computer system has many storage devices (see Figure 5.2). Storage devices and technologies vary in several important characteristics, which include the following:

- Speed
- Volatility
- Access method
- Portability
- Cost and capacity

No single device or technology is optimal in all characteristics, so any storage device optimizes some characteristics at the expense of others. A computer system has a variety of storage devices, each offering a cost-effective solution to a particular storage requirement.

Speed

Speed is the most important characteristic differentiating primary and secondary storage. It's essential because the CPU must be continuously supplied with instructions and data to keep it busy. For example, a CPU with a 1 GHz clock rate needs a new instruction and supporting data every nanosecond. As you learned in Chapter 4, a wait state is a CPU cycle spent waiting for access to an instruction or data. Wait states reduce CPU and computer system performance.

As discussed in Chapter 2, registers in the CPU are storage locations for instructions and data. Their location enables zero wait states for access, but CPUs have a limited

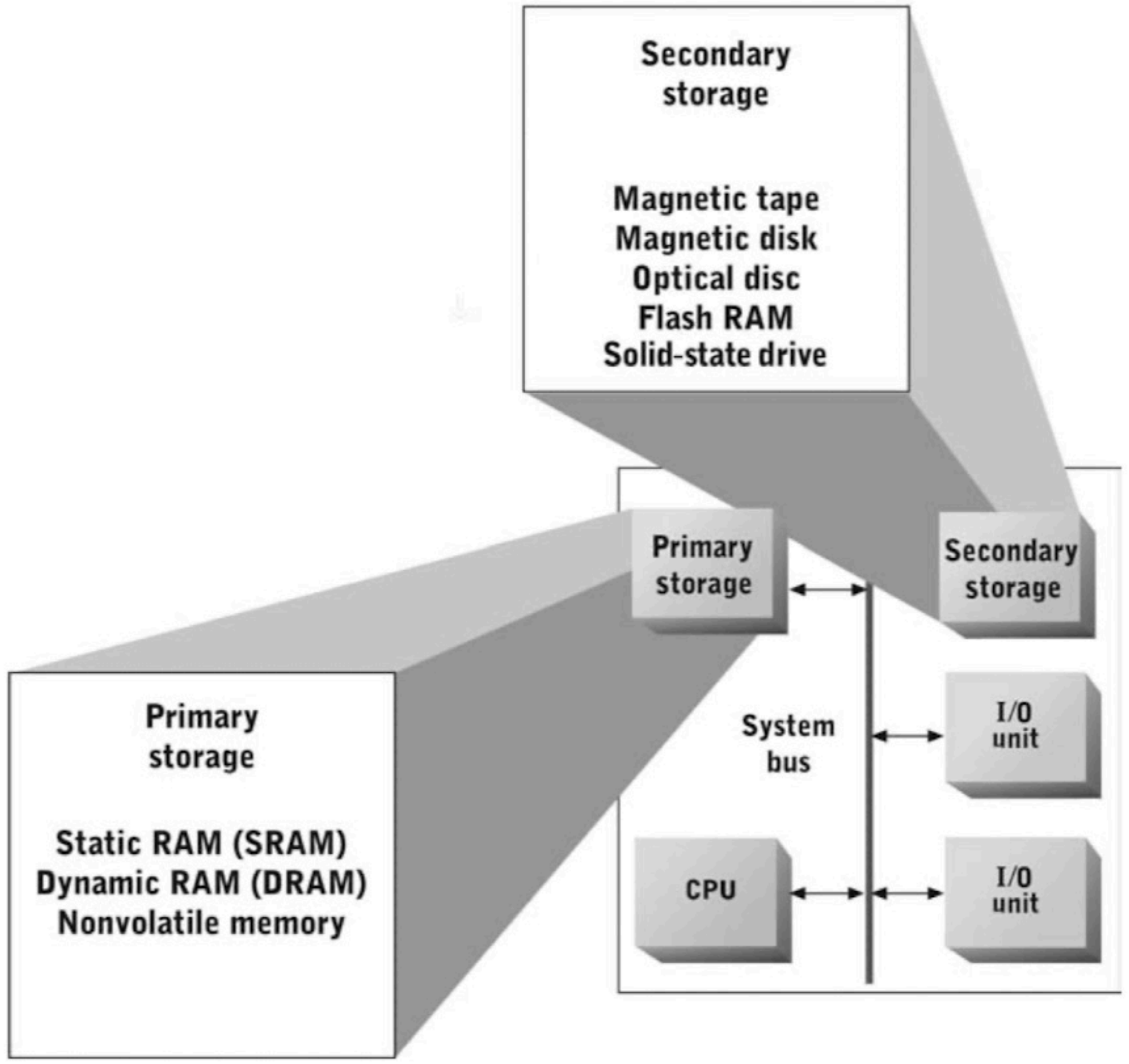

FIGURE 5.2 Primary and secondary storage and their component devices

number of registers—far fewer than are needed to hold typical programs and their data. Primary storage extends the limited capacity of CPU registers. The CPU moves data and instructions continually between registers and primary storage. To ensure that this movement incurs few or no wait states, all or part of primary storage is implemented with the fastest available storage devices. With current technology, primary storage speed is typically faster than secondary storage speed by a factor of 10^5 or more.

Speed is also an important issue for secondary storage. Many information system applications need access to large databases to support ongoing processing. Program response time in these systems depends on secondary storage access speed, which also affects overall computer performance in other ways. Before a program can be executed, its executable code is copied from secondary to primary storage. The delay between a user request for program execution and the first prompt for user input depends on the speed of both primary and secondary storage.

Storage device speed is called access time. **Access time** is the time required to perform one complete read or write operation. It's assumed to be the same for both reading and writing unless otherwise stated. For some storage devices, such as random access memory (RAM), access time is the same regardless of which storage location is accessed.

For other storage devices, such as disks, access time varies with storage location and is typically expressed as an average of access times for all storage locations, called **average access time**. Access times of primary storage devices are generally expressed in nanoseconds (ns, billionths of a second). Access times for secondary storage devices are typically expressed in milliseconds (ms, thousandths of a second) or microseconds (μs, millionths of a second).

By itself, access time is an incomplete measure of storage device speed. A complete measure consists of access time and the unit of data transfer to or from the storage device. Data transfer unit size varies from one storage device to another. For primary storage devices, it's usually a word. Depending on the CPU, a word can represent 2, 4, 8, or more bytes, with 4 and 8 bytes being the most common word sizes. Data transfer unit size is sometimes used as an adjective, as in *64-bit memory*.

Secondary storage devices read or write data in units larger than words. **Block** is a generic term for describing secondary storage data transfer units. Block size is normally stated in bytes and can vary widely between storage devices. 512-byte and 4096-byte block sizes are typical for magnetic and solid-state disk drives. The term **sector** describes the data transfer unit for magnetic disk and optical disc drives. As with blocks, sector size is generally stated in bytes and can vary from one device to another.

A storage device's **data transfer rate** is computed by dividing 1 by the access time (expressed in seconds) and multiplying the result by the unit of data transfer (expressed in bytes). For example, the data transfer rate for a primary storage device with 15 ns access time and a 64-bit word data transfer unit can be calculated as follows:

$$\frac{1 \text{ seconds}}{15 \text{ nanoseconds}} \times 64 \text{ bits} = \frac{1}{0.000000015} \times 8 \text{ bytes} \approx 533{,}333{,}333 \text{ bytes per second}$$

Using average access time, this formula can be used to calculate an average data transfer rate for devices with variable access times.

Volatility

A storage device or medium is **nonvolatile** if it holds data without loss over long periods and is **volatile** if it can't reliably hold data for long periods. Primary storage devices are generally volatile, and secondary storage devices are generally nonvolatile.

Volatility is actually a matter of degree and conditions. For example, RAM is nonvolatile as long as external power is supplied continuously. However, it's generally considered a volatile storage device because continuous power can't be guaranteed under normal operating conditions (for example, during a system restart after installing an OS update). Magnetic tape and disk are considered nonvolatile storage media, but data on magnetic media is often lost after a few years because of natural magnetic decay and other factors. Data stored on nonvolatile media might also be lost because of compatible read/write devices becoming obsolete, which has happened with many older diskette and tape formats.

Access Method

The physical structure of a storage device's read/write mechanism and storage medium determines the ways in which data can be accessed. Three broad classes of access are recognized:

- Serial access
- Random access
- Parallel access

A single device can use multiple access methods.

Serial Access

A **serial access** storage device stores and retrieves data items in a linear, or sequential, order. Magnetic tape is the only widely used form of serial access storage. Data is written to a linear storage medium in a specific order and can be read back only in that same order. For example, viewing the past few minutes recorded on a video cassette tape requires playing or fast-forwarding past all the minutes preceding it. Similarly, if n digital data items are stored on a magnetic tape, the nth data item can be accessed only after the preceding $n - 1$ data items have been accessed.

Serial access time depends on the current position of the read/write mechanism and the position of the target data item in the storage medium. If both positions are known, access time can be computed as the difference between the current and target positions multiplied by the time required to move from one position to the next.

Because of their inefficient access method, serial access devices aren't used for frequently accessed data. Even when data items are accessed in the same order as they're written, serial access devices are much slower than other forms of storage. Serial access storage is used mainly to hold backup copies of data stored on other storage devices—for example, in a weekly tape backup of user files from magnetic disk.

Random Access

A **random access** device, also called a **direct access** device, isn't restricted to any specific order when accessing data. Rather, it can access any storage location directly. All primary storage devices and disk storage devices are random access devices.

Access time for a random access storage device might or might not be a constant. It's a constant for most primary storage devices but not for disk storage because of the physical, or spatial, relationships of the read/write mechanism, storage media, and storage locations on the media. This issue is discussed in more detail later in the section "Magnetic Storage."

Parallel Access

A **parallel access** device can simultaneously access multiple storage locations. Although RAM is considered a random access storage device, it's also a parallel access device. This confusion comes from differences in defining the unit of data access. If you consider the unit of data access to be a bit, access is parallel. That is, random access memory circuitry can access all the bits in a byte or word at the same time.

Parallel access can also be achieved by subdividing data items and storing the component pieces on multiple storage devices. For example, some OSs can store file content on several disk drives. Different parts of the same file can be read at the same time by issuing parallel commands to each disk drive.