Technology Guide

1

Hardware

TG1.1 Components of a Computer System

Computer hardware is composed of the following components: central processing unit (CPU), primary storage, secondary storage, input devices, output devices, and communication devices. Communication devices are covered in detail in Tech Guide 4.

The **input devices** accept data and instructions and convert them to a form that the computer can understand. The **output devices** present data in a form people can understand. The **CPU** manipulates the data and controls the tasks done by the other components. **Primary storage** (internal storage that is part of the CPU) temporarily stores data and program instructions during processing. **Secondary storage** (external storage such as flash drives) stores data and programs that have been saved for future use. **Communication devices** manage the flow of data from public networks (e.g., Internet, intranets) to the CPU and from the CPU to networks. A schematic view of a computer system is shown in Figure TG1.1.

REPRESENTING DATA, PICTURES, TIME, AND SIZE IN A COMPUTER

ASCII. Computers are based on integrated circuits (chips), each of which includes millions of sub-miniature transistors that are interconnected on a small (less than l-inch-square) chip area. Each transistor can be in either an "on" or an "off" position.

The on–off states of the transistors are used to establish a binary 1 or 0 for storing one binary digit, or bit. A fixed number of bits representing specific characters—letters, numbers, and special symbols—is known as a byte, usually 8 bits. Because a bit has only two states, 0 or 1, the bits comprising a byte can represent any of 2^8, or 256, unique characters.

The character that the byte represents depends upon on the coding scheme used. The two most commonly used coding schemes are:

1. **ASCII (American National Standard Code for Information Interchange),** pronounced *ask-ee.*
2. **EBCDIC (Extended Binary Coded Decimal Interchange Code),** pronounced *ebsa-dik.*

EBCDIC was developed by IBM and is used primarily on large mainframe computers. ASCII is the standard coding scheme for microcomputers. These coding schemes, and the characters they represent, are shown in Figure TG1.2. In addition to characters, it is possible to represent commonly agreed-upon symbols in a binary code. For example, the plus sign (+) is 00101011 in ASCII.

The 256 characters and symbols that are represented by ASCII and EBCDIC codes are sufficient for English and Western European languages but are not large enough for Asian and other languages that use different alphabets.

Unicode is a 16-bit code that has the capacity to represent more than 65,000 characters and symbols. The system employs the codes used by ASCII and also includes other alphabets (such as Cyrillic and Hebrew), special characters (including religious symbols), and some of the "word writing" symbols used in various Asian countries.

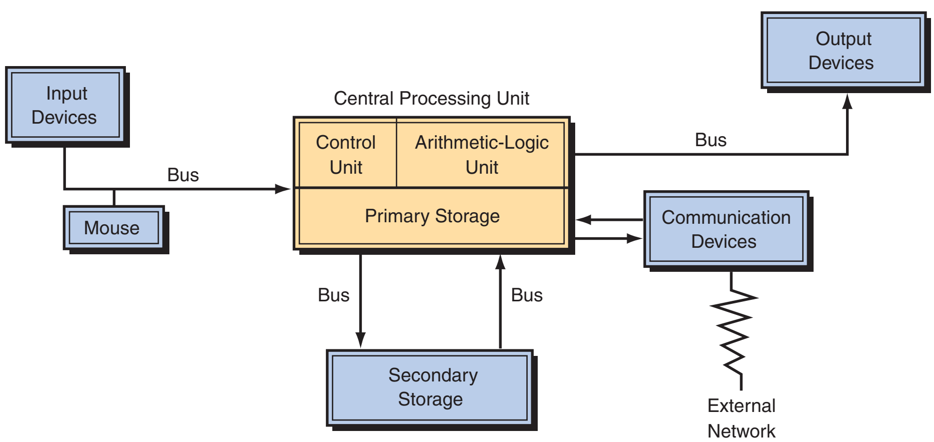

Figure TG1.1 Components of computer hardware. A bus is a connecting channel.

Character	EBCDIC Code	ASCII Code	Character	EBCDIC Code	ASCII Code
A	11000001	10100001	S	11100010	10110011
B	11000010	10100010	T	11100011	10110100
C	11000011	10100011	U	11100100	10110101
D	11000100	10100100	V	11100101	10110110
E	11000101	10100101	W	11100110	10110111
F	11000110	10100110	X	11100111	10111000
G	11000111	10100111	Y	11101000	10111001
H	11001000	10101000	Z	11101001	10111010
I	11001001	10101001	0	11110000	01010000
J	11010001	10101010	1	11110001	01010001
K	11010010	10101011	2	11110010	01010010
L	11010011	10101100	3	11110011	01010011
M	11010100	10101101	4	11110100	01010100
N	11010101	10101110	5	11110101	01010101
O	11010110	10101111	6	11110110	01010110
P	11010111	10110000	7	11110111	01010111
Q	11011000	10110001	8	11111000	01011000
R	11011001	10110010	9	11111001	01011001

Figure TG1.2 Internal computing coding schemes.

Representing Images. Images are represented by a grid overlay of the picture. The computer measures the color (or light level) of each cell of the grid. The unit measurement of this is called a **pixel.** Figure TG1.3 shows a pixel representation of the letter *A* and its conversion to an input code.

Time and Size of Bytes. Time is represented in fractions of a second, as follows:

- **Millisecond** = 1/1000 second
- **Microsecond** = 1/1,000,000 second
- **Nanosecond** = 1/1,000,000,000 second
- **Picosecond** = 1/1,000,000,000,000 second

Size of a file or storage space is measured in bytes. Measures of size are:

- **Kilobyte** = 1,000 bytes (actually 1,024)
- **Megabyte** = 1,000 kilobytes = 10^6 bytes
- **Gigabyte** = 10^9 bytes
- **Terabyte** = 10^{12} bytes
- **Petabyte** = 10^{15} bytes
- **Exabyte** = 10^{18} bytes
- **Zettabyte** = 10^{21} bytes

TG1.2 Evolution of Computer Hardware

Computer hardware has evolved through four stages, or generations, of technology. Each generation has provided increased processing power and storage capacity, while simultaneously exhibiting decreases in costs, as you see in Table TG1.1. The generations are distinguished by different technologies that perform the processing functions.

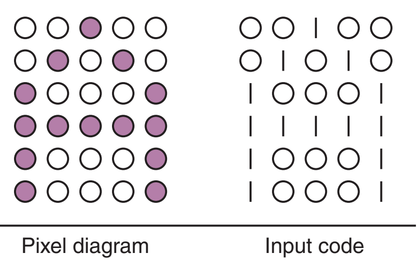

Figure TG1.3 Pixel representation of the letter A.

TABLE TG1.1 Hardware Generations

	Generations					
Feature	**1G**	**2G**	**3G**	**4G (early)**	**4G (1988)**	**4G (2001)**
Circuitry	Vacuum tubes	Transistors	Integrated circuits	LSI and VLSI	ULSI	GSI
Primary storage	2 KB	64 KB	4 MB	16 MB	64 MB	128 MB
Cycle times	100 millisecs	10 microsecs	500 nanosecs	800 picosecs	2,000 picosecs	333 MHz
Average cost	$2.5 million	$250 thousand	$25 thousand	$2.5 thousand	$2.0 thousand	$1.5 thousand

- 1G: The *first-generation* computers, from 1946 to about 1956, used *vacuum tubes* to store and process information. Vacuum tubes consumed large amounts of power, generated much heat, and were short-lived. Therefore, first-generation computers had limited memory and processing capability.
- 2G: The *second-generation* computers, 1957–1963, used **transistors** for storing and processing information. Transistors consumed less power than vacuum tubes, produced less heat, and were cheaper and more reliable. And 2G computers, with increased processing and storage capabilities, began to be more widely used for scientific and business purposes.
- 3G: *Third-generation* computers, 1964–1979, used **integrated circuits** for storing and processing information. Integrated circuits are made by printing numerous small transistors on silicon chips. These devices are called *semiconductors*. 3G computers employed software that could be used by nontechnical people, thus enlarging the computer's role in business.
- 4G: Early to middle *fourth-generation* computers, 1980–1995, used **very- large-scale integrated (VLSI) circuits** to store and process information. The VLSI technique allows the installation of hundreds of thousands of circuits (transistors and other components) on a small chip. With **ultra-large-scale integration (ULSI),** 100 million transistors could be placed on a chip. These computers are inexpensive and widely used in business and everyday life.
- Late 4G: Computers from 2001 to the present use **grand-scale integrated (GSI) circuits** to store and process information. With GSI, 1,000 million transistors can be placed on a chip.

The first 4Gs of computer hardware were based on the *Von Neumann architecture*, which processed information sequentially, one instruction at a time. The fifth generation (5G) of computers uses **massively parallel processing** to process multiple instructions simultaneously. Massively parallel computers use flexibly connected networks linking thousands of inexpensive, commonly used chips to address large computing problems, attaining supercomputer speeds. With enough chips networked together, massively parallel machines can perform more than a trillion floating point operations per second—a teraflop. A *floating point operation (flop)* is a basic computer arithmetic operation, such as addition or subtraction, on numbers that include a decimal point.

TG1.3 Types of Computers

Computers are distinguished on the basis of their processing capabilities. **Supercomputers** are the computers with the most processing power. The primary application of supercomputers has been in scientific and military work, but their use is growing rapidly in business as their prices decrease. Supercomputers are especially valuable for large simulation models of real-world phenomena, where complex mathematical representations and calculations are required, or for image creation and processing. Supercomputers are used to model the weather for better weather prediction, to test weapons nondestructively, to design aircraft (e.g., the Boeing 777) for more efficient and less costly production, and to make sequences in motion pictures (e.g., *Jurassic Park*).

Supercomputers use the technology of **parallel processing.** However, in contrast to neural computing, which uses massively parallel processing, supercomputers use noninterconnected CPUs. The difference is shown in Figure TG1.4. Parallel processing is also used in smaller computers where 2 to 64 processors are common.

MAINFRAMES

Mainframes are not as powerful and generally not as expensive as supercomputers. Large corporations, where data processing is centralized and large databases are maintained, often use mainframe computers. Applications that run on a mainframe can be large and complex, allowing for data and information to be shared throughout the organization.

MIDRANGE COMPUTERS

Midrange computers include *minicomputers* and *servers*.

Minicomputers. **Minicomputers** are smaller and less expensive than mainframe computers. Minicomputers are usually designed to accomplish specific tasks such as process control, scientific research, and engineering applications. Larger companies gain greater corporate flexibility by distributing data processing with minicomputers in organizational units instead of centralizing computing at one location. These minicomputers are connected to each other and often to a mainframe through telecommunication links.

Servers. **Servers** typically support computer networks, enabling users to share files, software, peripheral devices, and other network resources. Servers have large amounts of primary and secondary storage and powerful CPUs. Organizations with heavy e-commerce requirements and very large Web sites are running their Web and e-commerce applications on multiple servers in *server farms*. Server farms are large groups of servers maintained by an organization or by a commercial vendor and made available to customers. As companies pack greater numbers of servers in their server farms, they are using pizza-box-size servers called *rack servers* that can be stacked in racks. These computers run cooler and therefore can be packed more closely, requiring less space. To further increase density, companies are using a server design called a blade. A *blade* is a card about the size of a paperback book on which memory, processor, and hard drives are mounted.

A **Blade Server** is one component in a system. Blades can be individual servers that plug into a single cabinet or individual port cards that add connectivity to a switch. A blade is typically a hot swappable hardware device.

A blade server is a server architecture that houses multiple server modules (blades) in a single chassis. It is widely used in data centers to save space and improve system management. The chassis provides the power supply, and each blade has its own CPU, memory, and hard disk. Blade servers generally provide their own management systems and may include a network or storage switch.

With enterprise-class blade servers, disk storage is external, and the blades are diskless. This approach allows for more efficient failover because applications are not tied to specific hardware and a particular instance of the operating system. The blades are anonymous and interchangeable.

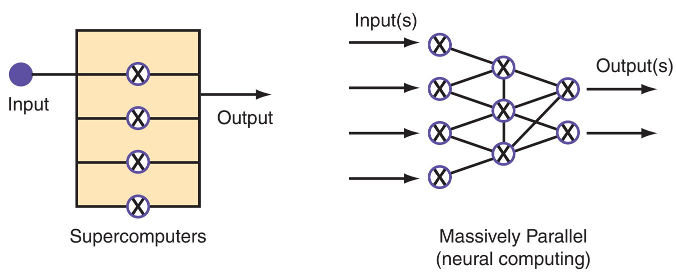

Figure TG1.4
Supercomputers vs. neural computing. (X is a CPU.)

TABLE TG1.2	Characteristics of Netbooks	
Operating system	Most Netbooks Windows 7 starter	Other Options Linux or Windows XP
CPU	Single Core Intel Atom	AMD Athlon Neo and dual Core Intel Atom
Screen size	10.1 inches	7 to 12.1 inches
Memory	1GB	2 GB
Hard drive	160 GB or 250 GB SATA 5400 RPM	Solid-state drive (SSD) or larger SATA hard drives
Price	$300 to $400	$200 to $500

Virtual servers make it possible to place multiple applications on a single physical server yet run each within its own operating system environment, known as a virtual machine. So when one virtual server crashes or is rebooted, the others continue operating without interruption.

Workstations Computer vendors originally developed workstations to provide the high levels of performance demanded by technical users such as designers. Workstations are typically based on RISC (reduced instruction set computing) architecture and provide both very-high-speed calculations and high-resolution graphic displays. These computers have found widespread acceptance within the scientific community and, more recently, within the business community.

MICROCOMPUTERS

Microcomputers, or *personal computers (PCs)*, are the smallest and least expensive category of general-purpose computers. **Notebook computers** are small, easily transportable, lightweight microcomputers that fit easily into a briefcase.

Netbooks are smaller, more portable, less expensive, and less powerful than notebooks; they are primarily for connecting to the Internet. (See Table TG 1.2.)

Netbooks were inspired by OLPC XO. The first netbook designed for consumers was the Asus EEE PC. It shipped with an Intel Celeron CPU, 512 MB of RAM, a 2 GB or 4 GB solid-state hard drive, and a Linux OS. Now most netbooks have an Intel Atom CPU, 1 GB of RAM, 160 GB or 250 GB SATA 5400 RPM hard drive, and Windows 7 Starter.

MOBILE DEVICES

Platforms for computing and communications include such **mobile devices** as **personal digital assistants (PDAs)** or *handheld personal computers*. Another platform is mobile phone handsets with wireless and Internet access capabilities, often called smartphones. Such devices usually use a microversion of a desktop operating system, such as Android, iPhone OS, Windows Phone 7, or Web OS. They have the following characteristics:

- They cost much less than PCs.
- OSs are simpler than those on a desktop PC.
- They provide good performance at specific tasks but do not replace the full functions of a PC.
- They provide both computer and/or communications features.
- They offer a Web portal that is viewable on a screen.

A **personal digital assistant** is a palmtop computer that combines a processor with a multitasking operating system using a pen (stylus) for handwriting recognition rather than keyboard input. Some PDAs enable users to communicate via fax, electronic mail, and paging, or to access online services. A **smartphone** has telcom and computing capabilities. Comparisons of smartphones are shown in Table TG 1.3.

TABLE TG2.3 Comparison of Smartphone Operating Systems

Phones	iPhone 3GS	Nexus One	Kin 2	Nokia N8	BlackBerry Bold 9700	Palm Pre-Plus	HTC EVO
OS shipped with	iPhone OS 3.0	Android 2.1	Windows Phone OS for KIN	Symbian^3	BlackBerry OS 5.0.0.330	WebOS 1.3.5.1	Android 2.1
Wireless carrier in the United States	AT&T	T-Mobile, AT&T, unlocked	Verizon	Unknown; may only be available unlocked	T-Mobile, AT&T	Verizon, AT&T (original Palm Pre on Sprint)	Sprint
U.S. release date	06/17/2009	01/06/2010	05/13/2010	Unknown	10/2009	01/20/2010	06/04/10
Keyboard	Virtual	Virtual	Physical	Virtual	Physical	Physical	Virtual
Camera features	3 MP with no flash; First iPhone to record video	5 MP with flash	8 MP with flash; 720P video	12 MP shots; video in 720P	3.2 MP with flash	3 MP LED flash	8 MP and 1.3MP front-facing webcam
Notable facts	Will get full version of iPhone OS 4.0, unlike iPhone 3G or older models.	Official Google phone	Isn't running Windows Phone 7	First phone to use the open-source Symbian^3 OS	The BlackBerry Bold continues RIM focus on business users	Can be used as a Wi-Fi hotspot; costs extra monthly	First 4G phone in the United States and the world's first 4G Android phone
GPS	Apps available for real	Free real-time navigation	Doesn't come with real-time navigation	Free real-time navigation	Doesn't come with real-time navigation	Real-time navigation costs $9.99 per month	Free real-time navigation
Display	LCD	OLED	LCD	OLED	LCD (not touch screen)	LCD	LCD
Storage, internal and expansion	16 GB or 32 GB internal, no memory card support	512 MB internal, comes with 4 GB memory card with support for up to 32 GB	8 GB internal, no memory card support	16 GB internal, supports memory cards up to 32 GB	256 MB internal, comes with 2 GB memory card with support for up to 32 GB	16 GB internal, no memory card support	1 GB internal, comes with 8 GB memory cards with support for up to 32 GB
Application processor	600 MHz Samsung S5PC100	1 GHz Snapdragon Qualcomm QSD 8250	600 MHz	Unknown	624 MHz	600 MHz ARM Cortex A8	1 GHz Qualcomm Snapdragon QSD8650
RAM	256 MB	512 MB	256 MB	Unknown	256 MB	512 MB	512 MB

TABLE TG1.4 Comparison of e-readers

E-book reader	Sony Reader Daily Edition PRS-900BC	Amazon Kindle (global wireless)	Barnes & Noble Nook
Screen size	7.1 inches	6 inches	6 inches
Wireless	First Sony Reader to come with 3G service. 3G service is provided by AT&T	Global wires version uses AT&T 3G service Earlier versions came with Sprint 3G	3G service through AT&T. It can use Wi-Fi
User input	The whole screen is a touch screen	Physical keyboard below the screen along with other buttons on the sides	Small touch screen on the bottom
Memory	2G internal Supports both SD and Memory Sick Duo memory cards	2 GB internal No memory card support	2 GB internal Supports up to 16 GB memory cards
Other notable features	Free Google books and library books Comes with its own case	Offers text-to-speech when allowed by the author and publisher	Can browse full books for free inside a Barnes & Noble brick-and-mortar store

Tablet PCs/E-Readers. Tablet PC technology runs touch-sensitive displays that you can tap with your fingers or sometimes with a stylus, forgoing a mouse or touch pad. A tablet PC can put the full power of Windows 7 Professional in a laptop computer that's as simple as a pad and pen.

The iPad is a tablet PC and e-reader from Apple. It runs a modified version of the iPhone OS and is designed for all user input to be done through the touch screen. There are two different versions of the iPad, the 3G and the Wi-Fi-only versions. The 3G iPad can use AT&T's 3G network in the U.S. for data but not for voice. It is important to note that the iPad is not a phone. The 3G model includes all of the features of the Wi-Fi only model plus a better 3G micro SIM card slot, 3G antenna, and GPS.

The iPad has three size options for internal storage: 16 GB, 32 GB, and 64 GB. The amount of memory you will need is important to consider because there is no way to add more. It has no memory card slots or USB ports to use USB flash drives. However, there are apps that work with Web-based storage.

The iPad can run the iPhone's 200,000+ apps, which can be stretched to fit the larger screen. The iPad also has some apps that have been made specifically for the iPad that cannot be used on the iPhone.

E-readers are devices used to read digital books, newspapers, and so forth. Most e-readers come with 3G, with no monthly charge which is used to connect to bookstores and to download some books. Most e-readers do not have a back light, which makes them easier on the eyes than computer monitors. However, this means an external light source will be needed just like with a normal book. Comparisons of e-readers are shown in Table TG1.4.

Wi-Fi. The spread of wireless fidelity, or Wi-Fi, has had a huge impact on the ability to connect to the Internet via laptops and mobiles. Wi-Fi is the common name for the wireless networking standard 802.11b (now 802.11n) that is a standard feature for most laptops and PDAs.

Wi-Fi provides the convenience of finding a hotspot for Internet connectivity. HP's iPAQ 5450 is the first handheld that has both wireless local area network (WLAN) and Bluetooth connectivity. It also has a built-in fingerprint security scanner—a small bar just beneath the navigation button over which the user swipes his or her finger to be identified. IEEE 802.11n is the wireless standard that was finalized in 2009.

TG1.4 Microprocessor and Primary Storage

MICROPROCESSORS

The **central processing unit (CPU)** performs all processing. The CPU is where all processing is controlled, data is manipulated, arithmetic computations are performed, and logical comparisons are made. The CPU consists of the control unit, the arithmetic-logic unit (ALU), and the primary storage (or main memory). Because of its small size, the CPU is also referred to as a *microprocessor*.

How a Microprocessor Works. The CPU operates like a tiny factory. Inputs come in and are stored until needed, at which point they are retrieved and processed and the output is stored and then delivered somewhere. Figure TG1.5 illustrates this process, which works as follows:

- The inputs are data and brief instructions about what to do with the data. These instructions come from software in other parts of the computer. Data might be entered by the user through the keyboard, for example, or read from a data file in another part of the computer. The inputs are stored in registers until they are sent to the next step in the processing.
- Data and instructions travel in the chip via electrical pathways called *buses*. The size of the bus—analogous to the width of a highway—determines how much information can flow at any time.
- The control unit directs the flow of data and instruction within the chip.
- The ALU receives the data and instructions from the registers and makes the desired computation. These data and instructions have been translated into *binary form*, that is, only 0s and 1s. The CPU can process only binary data.
- The data in its original form and the instructions are sent to storage registers and are then sent back to a storage place outside the chip, such as the computer's hard drive (discussed below). Meanwhile, the transformed data goes to another register and then on to other parts of the computer (to the monitor for display, or to be stored, for example).
- This cycle processing, known as a **machine instruction cycle,** occurs millions of times or more per second. The speed of a chip, which is an important benchmark, depends on four things: the clock speed, the word length, the data bus width, and the design of the chip.

1. The **clock.** This is located within the control unit and is the component that provides the timing for all processor operations. The beat frequency of the clock

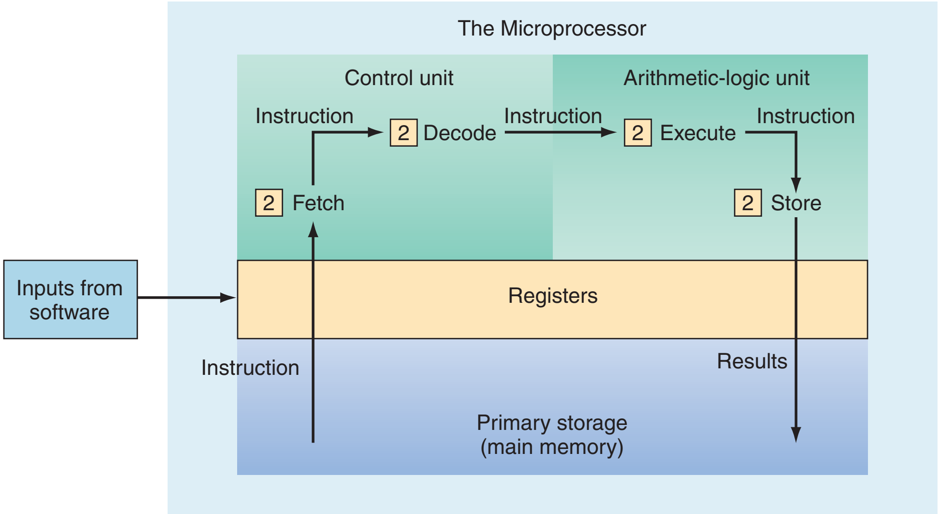

Figure TG1.5 How the CPU works.

(measured in megahertz [MHz], or millions of cycles per second) determines how many times per second the processor performs operations.

2. The **word length.** This is the number of bits (0s and 1s) that can be processed by the CPU at any one time.

3. The **bus width.** The wider the *bus* (the physical paths down which the data and instructions travel as electrical impulses), the more data can be moved and the faster the processing. A process's *bus bandwidth* is the product of the width of its bus (measured in bits) times the frequency at which the bus transfers data (measured in megahertz).

4. The physical design of the chip. The distance between transistors is known as **line width.** Historically, line width has been expressed in microns (millionths of a meter), but as technology has advanced, it has become more convenient to express line width in nanometers (billionths of a meter).

Running a Program on a Computer. To see how a program is run on a computer, look at Figure TG1.6. A computer program can be stored on a disk or on the hard drive (drive 'C'). To run this program, the operating system will retrieve the program from its location (step 1 in the figure) and place it into the RAM (step 2). Then the control unit "fetches" the first instruction in the program from the RAM (step 3) and acts upon it (e.g., sends a message to the user, via an output device, to enter a number, or say "yes" or "no"; step 4). Once the message is answered (step 5) (e.g., via an input device), it is stored in the RAM. This concludes the first instruction. Then the control unit "fetches" the second instruction (step 6), and the process continues on and on.

If one of the instructions calls for some computation, the control unit sends it, together with any relevant data stored in the RAM, to the arithmetic-logic unit (ALU) (step 7). The ALU executes the processing and returns the results to the RAM (step 8). The control unit then "fetches" one more instruction (step 9), which tells what to do with the result—for example, display it (step 10) or store it on the hard drive (step 11).

When instructions are "fetched," they are decoded. The computer can process large numbers of instructions per second, usually millions. Therefore, we measure the speed of computers by "millions of instructions per minute," or MIPS.

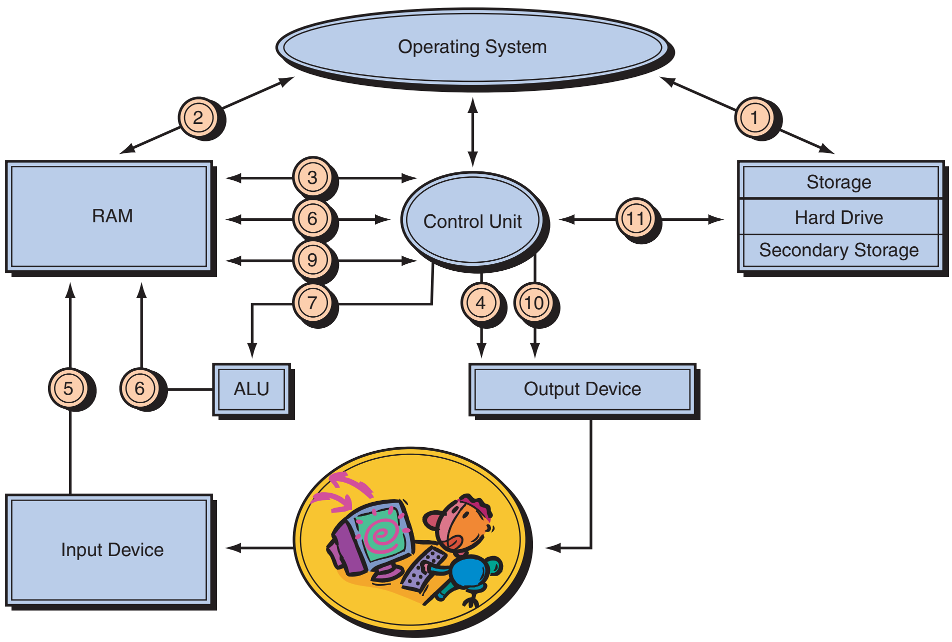

Figure TG1.6 Running a program on a computer.

Parallel Processing. A computer system with two or more processors is referred to as a **parallel processing system.** Today, some PCs have 2 to 6 processors while workstations have 20 or more. Processing data in parallel speeds up processing. Larger computers may have a hundred processors.

Computer Architecture. The arrangement of the components and their interactions is called computer *architecture*. Computer architecture includes the instruction set and the number of processors, the structure of the internal buses, the use of caches, and the types and arrangements of input/output (I/O) device interfaces.

Every processor comes with a unique set of operational codes or commands that represent the computer's instruction set. An **instruction set** is the set of machine instructions that a processor recognizes and can execute. Today, two instruction set strategies, **complex instruction set computer (CISC)** and **reduced instruction set computer (RISC),** dominate the processor instruction sets of computer architectures. These two strategies differ by the number of operations available and how and when instructions are moved into memory.

A *CISC processor* contains more than 200 unique coded commands, one for virtually every type of operation. The CISC design goal is for its instruction set to look like a sophisticated programming language. Inexpensive hardware can then be used to replace expensive software, thereby reducing the cost of developing software. The penalty for this ease of programming is that CISC processor–based computers have increased architectural complexity and decreased overall system performance. In spite of these drawbacks, most computers still use CISC processors.

The other approach is *RISC processors*, which eliminate many of the little-used codes found in the complex instruction set. Underlying RISC design is the claim that a very small subset of instructions accounts for a very large percentage of all instructions executed. The instruction set, therefore, should be designed around a few simple "hardwired" instructions that can be executed very quickly. The rest of the needed instructions can be created in software.

Arithmetic-Logic Unit. The **arithmetic-logic unit (ALU)** performs required arithmetic and comparisons, or logic, operations. The ALU adds, subtracts, multiplies, divides, compares, and determines whether a number is positive, negative, or zero. All computer applications are achieved through these six operations. The ALU operations are performed sequentially, based on instructions from the control unit. For these operations to be performed, the data must first be moved from the storage to the arithmetic registers in the ALU. **Registers** are specialized, high-speed memory areas for storing temporary results of ALU operations as well as for storing certain control information.

PRIMARY STORAGE

Primary storage, or **main memory,** stores data and program statements for the CPU. It has four basic purposes:

1. To store data that has been input until it is transferred to the ALU for processing
2. To store data and results during intermediate stages of processing
3. To hold data after processing until it is transferred to an output device
4. To hold program statements or instructions received from input devices and from secondary storage

Primary storage utilizes **integrated circuits.** These circuits are interconnected layers of etched semiconductor materials forming electrical transistor memory units with on–off positions that direct the electrical current passing through them. The on–off states of the transistors are used to establish a binary 1 or 0 for storing one binary digit, or bit.

Role of Buses. Instructions and data move between computer subsystems and the processor via communications channels called buses. A bus is a channel through which data is passed in electronic form. Three types of buses link the CPU, primary storage, and the other devices in the computer system. The data bus moves data to and from primary storage. The address bus transmits signals for locating a given address in primary storage. The control bus transmits signals specifying whether to "read" or "write" data to or from a given primary storage address, input device, or output device. The capacity of a bus, called **bus width,** is defined by the number of bits it carries at one time.

Control Unit. The **control unit** reads instructions and directs the other components of the computer system to perform the functions required by the program. It interprets and carries out instructions contained in computer programs, selecting program statements from the primary storage, moving them to the instruction registers in the control unit, and then carrying them out. It controls input and output devices and data-transfer processes from and to memory. The control unit does not actually change or create data; it merely directs the data flow within the CPU. The control unit can process only one instruction at a time, but it can execute instructions so quickly (millions per second) that it can appear to do many different things simultaneously.

The series of operations required to process a single machine instruction is called a **machine cycle.** Each machine cycle consists of the *instruction cycle*, which sets up circuitry to perform a required operation, and the *execution cycle*, during which the operation is actually carried out.

Categories of Memory. There are two categories of memory: the *register*, which is part of the CPU and is very fast, and the **internal memory chips,** which reside outside the CPU and are slower. A register is circuitry in the CPU that allows for the fast storage and retrieval of data and instructions during the processing. The control unit, the CPU, and the primary storage all have registers. Small amounts of data reside in the register for very short periods prior to their use.

The **internal memory** is used to store data just before it is processed by the CPU. Immediately after the processing, it comprises two types of storage space: RAM and ROM.

Random-access memory (RAM) is the place in which the CPU stores the instructions and data it is processing. The larger the memory area, the larger the programs that can be stored and executed.

More than one program may be operating at a time, each occupying a portion of RAM.

The advantage of RAM is that it is very fast in storing and retrieving any type of data, whether textual, graphical, sound, or animation-based. Its disadvantages are that it is relatively expensive and volatile. This volatility means that all data and programs stored in RAM are lost when the power is turned off. To lessen this potential loss of data, many of the newer application programs perform periodic automatic "saves" of the data.

Many software programs are larger than the internal, primary storage (RAM) available to store them. To get around this limitation, some programs are divided into smaller blocks, with each block loaded into RAM only when necessary. However, depending on the program, continuously loading and unloading **blocks** can slow down performance considerably, especially since secondary storage is so much slower than RAM. As a compromise, some architectures use high-speed **cache memory** as a temporary storage for the most frequently used blocks. Then the RAM is used to store the next most frequently used blocks, and secondary storage (described later) for the least used blocks.

There are two types of cache memory in the majority of computer systems—Level 1 (Ll) cache is located in the processor, and Level 2 (L2) cache is located on the motherboard but not actually in the processor. L1 cache is smaller and faster than L2 cache. Chip manufacturers are now designing chips with L1 cache and L2 cache in the processor and Level 3 (L3) cache on the motherboard.

Since cache memory operates at a much higher speed than conventional memory (i.e., RAM), this technique greatly increases the speed of processing because it reduces the number of times the program has to fetch instructions and data from RAM and secondary storage.

Dynamic random-access memories (DRAMs) are the most widely used RAM chips. These are known to be volatile since they need to be recharged and refreshed hundreds of times per second in order to retain the information stored in them.

Read-only memory (ROM) is that portion of primary storage that cannot be changed or erased. ROM is *nonvolatile*; that is, the program instructions are continually retained within the ROM, whether power is supplied to the computer or not. ROM is necessary to users who need to be able to restore a program or data after the computer has been turned off or, as a safeguard, to prevent a program or data from being changed. For example, the instructions needed to start, or "boot," a computer must not be lost when it is turned off.

Programmable read-only memory (PROM) is a memory chip on which a program can be stored. But once the PROM has been used, you cannot wipe it clean and use it to store something else. Like ROMs, PROMs are nonvolatile.

TG1.5 Input/Output Devices

The input/output (I/O) devices of a computer are not part of the CPU but are channels for communicating between the external environment and the CPU. Data and instructions are entered into the computer through **input devices (I/O),** and processing results are provided through **output devices.** Widely used I/O devices are the cathode-ray tube (CRT) or visual display unit (VDU), magnetic storage media, printers, keyboards, "mice," and image-scanning devices.

I/O devices are controlled directly by the CPU or indirectly through special processors dedicated to input and output processing. Generally speaking, I/O devices are subclassified into *secondary storage devices* (primarily disk and tape drives) and *peripheral devices* (any input/output device that is attached to the computer).

SECONDARY STORAGE

Secondary storage is separate from primary storage and the CPU but directly connected to it. It stores the data in a format that is compatible with data stored in primary storage, but secondary storage provides the computer with vastly increased space for storing and processing large quantities of software and data. Primary storage is volatile, contained in memory chips, and very fast in storing and retrieving data. In contrast, secondary storage is nonvolatile, uses many different forms of media that are less expensive than primary storage, and is relatively slower than primary storage.

Magnetic tape is kept on a large open reel or in a small cartridge or cassette. Today, cartridges and cassettes are replacing reels because they are easier to use and access. The principal advantages of magnetic tape are that it is inexpensive, relatively stable, long lasting, and can store very large volumes of data. Magnetic tape is excellent for backup or archival storage of data and can be reused. The main disadvantage of magnetic tape is that it must be searched from the beginning to find the desired data. This process is called *sequential access*. The magnetic tape itself is fragile and must be handled with care. Magnetic tape is also labor intensive to mount and dismount in a mainframe computer. Magnetic tape storage is often used for information that an organization must maintain but uses rarely or does not need immediate access to. Industries with huge numbers of files (e.g., insurance companies) use magnetic tape systems. Modern versions of magnetic tape systems use cartridges and often a robotic system that selects and loads the appropriate cartridge automatically.

Magnetic disks, also called **hard disks,** alleviate some of the problems associated with magnetic tape by assigning specific address locations for data, so that users can go directly to the address without having to go through intervening locations looking for the right data to retrieve. This process is called *direct access*.

A hard disk is like a phonograph containing a stack of metal-coated platters (usually permanently mounted) that rotate rapidly. Magnetic read/write heads, attached to arms, hover over the platters. To locate an address for storing or retrieving data, the head moves inward or outward to the correct position, then waits for the correct location to spin underneath.

The speed of access to data on hard-disk drives is a function of the rotational speed of the disk and the speed of the read/write heads. The read/write heads must position themselves, and the disk pack must rotate until the proper information is located. Advanced disk drives have access speeds of 8 to 12 milliseconds.

Magnetic disks provide storage for large amounts of data and instructions that can be rapidly accessed. Another advantage of disks over reel is that a robot can change them. This can drastically reduce the expenses of a data center. Storage Technology is the major vendor of such robots. The disks' disadvantages are that they are more expensive than magnetic tape and they are susceptible to "disk crashes."

In contrast to large, fixed disk drives, one approach is to combine a large number of small disks drives, each with 10- to 40-gigabyte capacity, developed originally for microcomputers. These devices are called **redundant arrays of inexpensive disks (RAID).** Because data is stored redundantly across many drives, the overall impact on system performance is lessened when one drive malfunctions. Also, multiple drives provide multiple data paths, improving performance. Finally, because of manufacturing efficiencies of small drives, the cost of RAID devices is significantly lower than the cost of large disk drives of the same capacity.

To take advantage of faster technologies, *disk-drive interfaces* must also be faster. Most PCs and workstations use one of two high-performance disk-interface standards: **Enhanced Integrated Drive Electronics (EIDE)** or **Small Computer Systems Interface (SCSI).** EIDE offers good performance, is inexpensive, and supports up to four disks, tapes, or CD-ROM drives. The latest version is called Serial ATA (SATA). For details, refer to *serialata.org*. SCSI drives are more expensive than EIDE drives, but they offer a faster interface and support more devices. SCSI interfaces are therefore used for graphics workstations, server-based storage, and large databases.

Optical storage devices have extremely high storage density. Typically, much more information can be stored on a standard 5.25-inch optical disk than on a comparably sized floppy (about 400 times more). Since a highly focused laser beam is used to read/write information encoded on an optical disk, the information can be highly condensed. In addition, the amount of physical disk space needed to record an optical bit is much smaller than that usually required by magnetic media.

Another advantage of optical storage is that the medium itself is less susceptible to contamination or deterioration. First, the recording surfaces (on both sides of the disk) are protected by two plastic plates, which keep dust and dirt from contaminating the surface. Second, only a laser beam of light, not a flying head, comes in contact with the recording surface; the head of an optical disk drive comes no closer than 1 mm from the disk surface. Optical drives are also less fragile, and the disks themselves may easily be loaded and removed. In addition, optical disks can store much more information, both on a routine basis and also when combined into storage systems.

Compact disk read-only memory (CD-ROM) disks have high capacity, low cost, and high durability. CD-ROM technology is very effective and efficient for mass-producing many copies of large amounts of information that do not need to be changed; for example, encyclopedias, directories, and online databases. However, because it is a read-only medium, the CD-ROM can only be read, not written on. **Compact disk, rewritable (CD-RW)** adds rewritability to the recordable compact disk market.

Digital video disks (DVDs) offer higher quality and denser storage capabilities.

Blu-ray disks offer higher quality and denser storage than DVDs. Also, Blu-ray disks use hard coating technology to make them more scratch resistant than CDs or DVDs.

TABLE TG1.5 Comparison of Secondary Storage

Type	Advantages	Disadvantages	Application
Magnetic Storage Devices			
Magnetic tape	Lowest cost per unit stored.	Sequential access means slow retrieval speeds.	Corporate data archiving.
Hard drive	Relatively high capacity and fast retrieval speed.	Fragile; high cost per unit stored.	Personal computers through mainframes.
RAID	High capacity; designed for fault tolerance and reduced risk of data loss; low cost per unit stored.	Expensive, semipermanent installation.	Corporate data storage that requires frequent, rapid access.
SAN	High capacity; designed for large amounts of enterprise data.	Expensive.	Corporate data storage that requires frequent, rapid access.
NAS	High capacity; designed for large amounts of enterprise data.	Expensive.	Corporate data storage that requires frequent, rapid access.
Magnetic diskettes	Low cost per diskette, portability.	Low capacity; very high cost per unit stored; fragile.	Personal computers.
Memory cards	Portable; easy to use; less failure-prone than hard drives.	Expensive.	Personal and laptop computers.
Memory sticks	Extremely portable and easy to use.	Expensive.	Consumer electronic devices; moving files from portable devices to desktop computers.
USB flash drives	Portable; easy to use; fast; require only a USB port, not a special drive.		Personal computers.
SSD hard drives	Faster than HDD.	High cost per unit stored.	Personal computers through corporate data storage.
Expandable storage	Portable; high capacity.	More expensive than hard drives.	Backup of internal hard drive.
Optical Storage Devices			
CD-ROM	High capacity; moderate cost per unit stored; high durability.	Slower retrieval speeds than hard drives; only certain types can be rewritten.	Personal computers through corporate data storage.
DVD	High capacity; moderate cost per unit stored.	Slower retrieval speeds than hard drives.	Personal computers through corporate data storage.
FMD-ROM	Very high capacity; moderate cost per unit stored.	Faster retrieval speeds than DVD or CD-ROM; slower retrieval speeds than hard drives.	Personal computers through corporate data storage.
Blu-ray	Higher capacity than DVDs, up to 50 GB currently.	Expensive.	Personal computers through corporate data storage.

Memory PC Cards (also known as *memory sticks*) expand the amount of available memory. They have been widely used, particularly in portable devices such as PDAs and smartphones.

Table TG1.5 summarizes the major secondary storage devices and their advantages, limitations, and applications.

PERIPHERAL INPUT DEVICES

Users can command the computer and communicate with it by using one or more **input devices.** Each input device accepts a specific form of data. For example, keyboards transmit typed characters, and handwriting recognizers "read" handwritten

TABLE TG1.6 Input Devices

Categories	Examples
Keying devices	• Punched card reader • Keyboard • Point-of-sale (POS) terminal
Pointing devices (devices that point to objects on the computer screen)	• Mouse (including rollerballs and trackballs) • Touch screen • Touchpad (or trackpad) • Light pen • Joy stick
Optical character recognition (devices that scan characters)	• Barcode scanner (e.g., at POS) • Optical character reader • Wand reader • Cordless reader • Optical mark reader
Handwriting recognizers	• Pen
Voice recognizers (data entered by voice)	• Microphone
Other devices	• Magnetic ink character readers • Digital cameras • Automated teller machines (ATMs) • Smart cards • Digitizers (for maps, graphs, etc.) • RFID

characters. Users want communication with computers to be simple, fast, and error-free. Therefore, a variety of input devices fit the needs of different individuals and applications (Table TG 1.6).

Keyboards. The most common input device is the keyboard. The keyboard is designed like a typewriter but with many additional special keys. Most computer users utilize keyboards regularly. Unfortunately, a number of computer users have developed repetitive stress injury, which they allege comes from excessive use of poorly designed keyboards. As a result, new keyboards have been developed that are ergonomically designed. For example, some keyboards are now "split" in half, loosely approximating the natural angle of the arms and wrists.

Mice and Trackballs. The computer *mouse* is a handheld device used to point a cursor at a desired place on the screen, such as an icon, a cell in a table, an item in a menu, or any other object. Once the arrow is placed on an object, the user clicks a button on the mouse, instructing the computer to take some action. The use of the mouse reduces the need to type in information or use the slower arrow keys.

A variant of the mouse is the **trackball,** which is often used in graphic design. The user holds an object much like a mouse, but rather than moving the entire device to move the cursor (as with a mouse), he or she rotates a ball that is built into the top of the device. Portable computers have some other mouse-like technologies, such as the glide-and-tap pad, used in lieu of a mouse. Many portables also allow a conventional mouse to be plugged in when desired.

Another variant of the mouse, the **optical mouse,** replaces the ball, rollers, and wheels of the mechanical mouse with a light, lens, and camera chip. It replicates the action of a ball and rollers by taking photographs of the surface it passes over and comparing each successive image to determine where it is going.

Touch Screens. An alternative to the mouse or other screen-related device is a *touch screen*. **Touch screens** are a technology that divides a computer screen into different areas. Users simply touch the desired area (often buttons or squares) to trigger an action.

Stylus. A **stylus** is a pen-style device that allows the user either to touch parts of the predetermined menu of options or to handwrite information into the computer (as with some PDAs). The technology may respond to pressure of the stylus, or the stylus can be a type of light pen that emits light that is sensed by the computer.

Joysticks. *Joysticks* are used primarily at workstations that can display dynamic graphics. They are also used in playing video games. The joystick moves and positions the cursor at the desired object on the screen.

Electronic Forms. **Electronic forms** provide a standardized format whose headings serve as prompts for the input. In **form interaction,** the user enters data or commands into designated spaces (fields) in forms. The computer may produce some output after input is made, and the user may be requested to continue the form interaction process. Electronic forms can alleviate many of the resource-intensive steps of processing forms, making traditional typesetting and printing unnecessary. Finally, processing centers do not need to rekey data from paper-based forms, since the data remains in electronic format throughout the process.

Whiteboard. A **whiteboard** is an area on a display screen that multiple users can write or draw on. Whiteboards are a principal component of teleconferencing applications because they enable visual as well as audio communication.

Source Data Automation. Source data automation captures data in computer readable form at the moment the data is created. Point-of-sale systems, optical barcodes and code scanners, other optical character recognition devices, handwriting recognizers, voice recognizers, digitizers, and cameras are examples of source data automation. Source data automation devices eliminate errors arising from humans keyboarding data and allow for data to be captured directly and immediately, with built-in error correction. The major devices are described below.

Point-of-Sale Terminals. Many retail organizations utilize **point-of-sale (POS) terminals.** The POS terminal has a specialized keyboard. For example, the POS terminals at fast-food restaurants include all of the items on the menu, sometimes labeled with the picture of the item. POS terminals in a retail store are equipped with a barcode scanner that reads the barcoded sales tag. POS devices increase the speed of data entry and reduce the chance of errors. POS terminals may include many features such as scanner, printer, voice synthesis (which pronounces the price by voice), and accounting software.

Barcode Scanner. **Barcode scanners** scan the one-dimensional (1-D) black-and-white bars written in the *Universal Product Code* (UPC). This code specifies the name of the product and its manufacturer (product ID). Then a computer finds in the database the price equivalent to the product's ID. Barcodes are especially valuable in high-volume processing, where keyboard energy is too slow and/or inaccurate. Applications include supermarket checkout, airline baggage stickers, and transport companies' packages (Federal Express, United Parcel Service, and the U.S. Postal Service). The **wand reader** is a special handheld barcode reader that can read codes that are also readable by people.

Radio Frequency Identification (RFID) Tag. *Radio frequency identification* (RFID) is a system of technologies that use radio waves to automatically identify people or objects. The unique information (usually a serial number) is stored on a microchip (tag) that is attached to an antenna, which can transmit to a nearby reader. The reader then converts the radio waves from the RFID tag into digital information for the computer to use.

Optical Mark Reader. An **optical mark reader** is a special scanner for detecting the presence of pencil marks on a predetermined grid, such as multiple-choice test answer sheets.

Sensors. **Sensors** are extremely common technologies embedded in other technologies. They collect data directly from the environment and input it into a computer system. Examples might include your car's airbag activation sensor or fuel mixture/pollution control sensor, inventory control sensors in retail stores, and the myriad types of sensors built into a modern aircraft.

Universal Serial Bus (USB). This is a low-cost interfacing port for computer peripherals. USB 1.1 has a maximum transfer rate of 12 Mbps, which cannot fulfill some speedy peripherals, such as external hard drives. USB 2.0 has a maximum transfer rate of 480 Mbps, which is 40 times faster than USB 1.1. It is faster than its competitor IEEE 1394, which has a maximum transfer rate of 480 Mbps. The newest USB standard is USB 3.0. It is about 10 times faster than USB 2.0, with a maximum transfer speed of 5 Gbps.

Monitors. The data entered into a computer can be visible on the computer **monitor,** which is basically a video screen that displays both input and output. Monitors come in different sizes, ranging from inches to several feet. The major benefit is the interactive nature of the device.