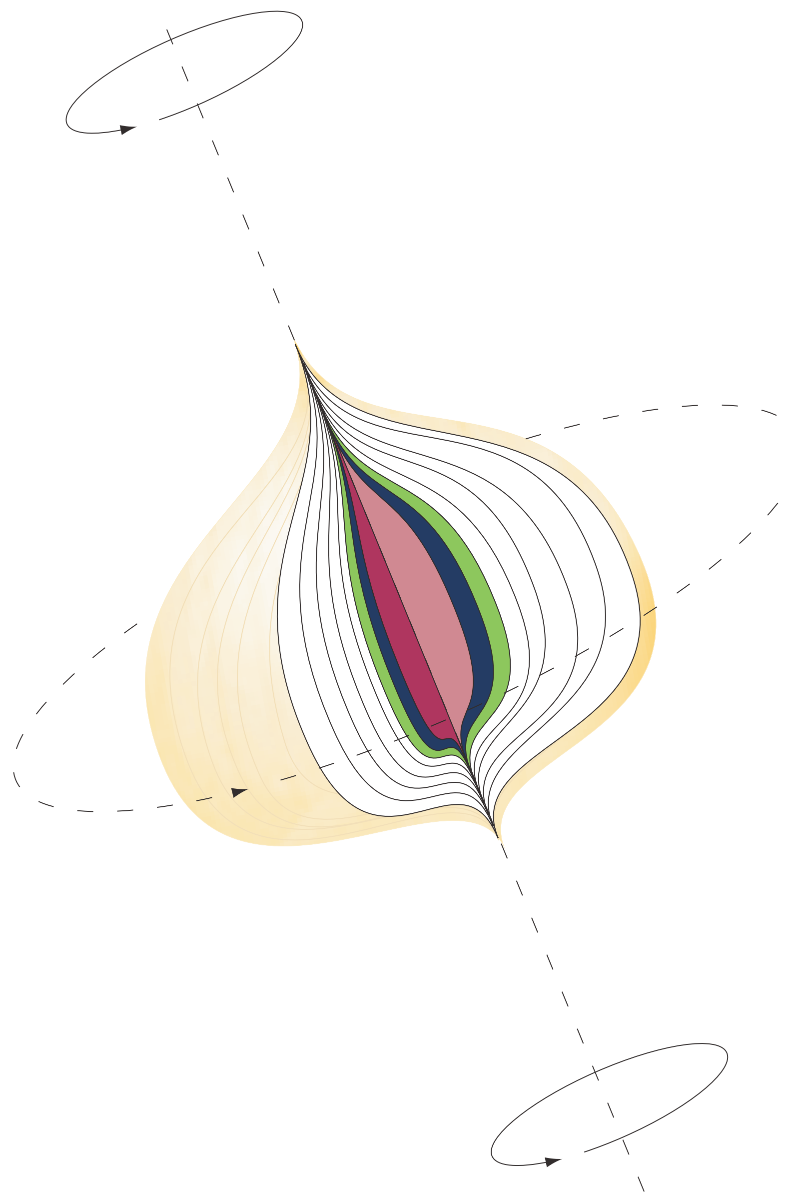
The Hardware Layer



5 Computing Components

Chapter 2 described the binary number system in which all information is represented on a computer. Chapter 4 described how we control electricity at a fundamental level to manage binary values. Now we can describe the primary components of a computer that capitalize on these technologies. These primary components are like Lego pieces; they can be combined to build a variety of different computers, just as Legos can form a variety of buildings.

Although these components, such as main memory and the central processing unit (CPU), are often thought of as the most fundamental parts of a computer, we know that they are abstractions of even more fundamental concepts.

Goals

After studying this chapter, you should be able to:

- read an ad for a computer and understand the jargon.
- list the components and their function in a von Neumann machine.
- describe the fetch-decode-execute cycle of the von Neumann machine.
- describe how computer memory is organized and accessed.
- name and describe the various auxiliary storage devices.
- define three alternative parallel computer configurations.
- explain the concept of embedded systems and give examples from your own home.

5.1 Individual Computer Components

Computing, more than most fields, has its own special jargon and acronyms. We begin this chapter by translating an ad for a desktop computer. We then examine the components of a computer as a logical whole before looking at each component in some detail.

Consider the following ad for a laptop computer.

Insatavialion 640 Laptop

Exceptional Performance and Portability

- Intel® Core™ 2 Duo (2.66GHz/1066Mhz FSB/6MB cache)
- 15.6″ High Definition (1080p) LED Backlit LCD Display (1366 x 768)
- 512MB ATI Mobility Radeon Graphics
- Built-in 2.0MP Web Camera
- 4GB Shared Dual Channel DDR2 at 800MHz
- 500GB SATA Hard Drive at 5400RPM
- 8X Slot Load DL DVD+/- RW Drive
- 802.11 a/g/n and Bluetooth 3.0
- 85 WHr Lithium Ion Battery
- (2) USB 2.0, HDMI, 15-pin VGA, Ethernet 10/100/1000, IEEE 1394 Firewire, Express Card, Audio line-in, line-out, mic-in
- 14.8″W X 1.2″H X 10.1″D, 5.6 lbs
- Microsoft® Windows 7® Professional
- Microsoft® Office Home and Student 2007
- 36-Month subscription to McAfee Security Center Anti-virus

There are two important and interesting things about this ad: The average person hasn't the foggiest idea what it all means, and by the time you are reading it, the machine that it describes will be obsolete. In this chapter, we try to interpret the acronyms; we can't do anything about the speed at which computer hardware and software change.

Before we go on to describe computer components in the abstract, let's go through this specification and decipher the acronyms. After this exercise, we go through all of the material again in more depth, so don't be concerned if the terms seem confusing. You'll see all of them defined again later.

The first line describes the central processor inside the laptop. Core™ 2 is a type of processor, and Duo refers to the presence of two of these processors (called cores) on a single chip. The 2.66GHz tells how fast the processors are. The G in GHz is the abbreviation for giga, a metric prefix indicating one billion. Hz stands for hertz, a unit of frequency that measures cycles per second, named after Heinrich R. Hertz. In a computer, a centrally generated series of electrical pulses, called the clock, is used to ensure that all of its actions are coordinated. You can think of the clock like an orchestra conductor's waving baton, which keeps all of the musicians playing together at a particular tempo. The clock in this processor pulses 2.66 billion times per second.

Following the clock speed number we read: 1066MHz FSB. Knowing that M in the metric system stands for million, we can guess that something called FSB is pulsing 1066 million (or just over a billion) times per second.

What is the FSB? A processor needs to access memory and input/output devices and does so through a set of wires called a bus. A computer has many different buses, but the one that makes the primary connection between the processor and the outside world is called the front side bus (FSB). Thus, these processors can communicate with the outside world 1066 million times per second. But if each of the processors is performing 2.66 billion operations per second, how can the FSB keep up at only one billion accesses per second?

The answer is related to the "6MB cache." MB stands for megabytes. A byte is a unit of memory, and a megabyte is 2^{20} (a little more than a million) bytes. So 6MB refers to six megabytes of cache memory. Cache is a small, fast memory that is usually built into the processor chip. Thus, the two processors have direct access to 6MB of memory without using the FSB. Many of the processor's attempts to access memory will find what they need within the cache. They only activate the FSB when they need something that is not in cache. Thus, the FSB can be slower than the processors and still not get in their way.

In general, a faster clock, faster FSB, and more cache would seem to yield a more powerful computer. But as in all areas of engineering, there are trade-offs. If the processor runs faster it consumes more power, which can cause the circuitry to overheat and shut down. A faster FSB requires faster devices in the outside world, which means their circuitry is more expensive. As cache gets bigger, access to its data becomes slower, which slows down the processors.

The next part of the ad describes the screen. The number 15.6″ refers to the diagonal measurement of the display area. High Definition (1080p) says it is compatible with the high definition television standard with 1080 horizontal lines of display elements. As we'll see, this isn't completely true. The screen is described as an LED backlit LCD. LED stands for light emitting diode, just like those found in some flashlights. A strip of these lights shine up from the bottom to illuminate the display. LEDs are replacing the use of a miniature fluorescent light bulb. The advantages are that LEDs last longer without growing dim and do not contain the toxic metal mercury. Lastly, the numbers 1366 × 768 refer to the screen's resolution in picture elements (pixels). This screen is 1366 pixels wide and 768 pixels high. Note that the number of vertical pixels is less than the 1080 claimed earlier. The computer compresses the 1080 lines from a high definition source, such as a movie, to fit the 768 lines in its screen. It takes an informed consumer to recognize marketing exaggerations such as this.

Next the ad lists the brand and model of graphics processor unit (GPU). We also see that it has 512MB of memory. The GPU is a separate computer that can be even more powerful than the main processors. Games and other graphics software send commands to the GPU that cause it to manipulate the image on the screen very quickly. It thus relieves the main processors of this task. The GPU keeps the data for the screen image in its own memory. The more it has, the better it is able to work with complex images, support external displays, and so on.

The fourth line in the ad describes a built-in digital camera that faces the user from just above the screen. This camera can be used for video conferencing over the Internet or recording still images and videos. 2.0MP indicates that the camera has a resolution of 2 million pixels, which is sufficient for these tasks.

Next the ad lists the computer's random access memory (RAM), also called main memory. Random access means that each byte of memory can be accessed directly, rather than having to begin at the beginning and access each byte in turn until you get to the one you want. 4GB means that there are 4×2^{30} bytes of storage (2^{30} is just over one billion). Shared means that both processors have access to this memory. Dual-channel DDR2 is the type of memory. It provides two access paths (called channels), and DDR2 stands for second generation, double-data rate. Through clever use of circuitry, memory designers doubled the rate at which a memory could operate, compared with earlier designs. Their achievement is acknowledged in this acronym.

This laptop contains a hard disk drive, which is the common name for the computer's secondary (also called auxiliary) storage device. It is listed as having 500GB (500×2^{30} bytes) of storage. The disk uses an interface called SATA, which stands for Serial ATA. Serial means that its data is transmitted to and from the computer as a stream of individual bits, rather than the older approach of sending 16 bits at once over 16 wires (known as Parallel ATA). The ATA acronym has a long history, referring to a means of attaching a hard drive to the IBM PC/AT — a computer that was introduced in 1984. Serial ATA is both faster and less costly to make, and it can transfer up to 300 MB per second, which is more than most hard disks can supply. The ad also mentions 5400 RPM (revolutions per minute), which is how fast the disk spins. Disks in laptops spin relatively slowly to conserve battery power. Disks are also available that spin at 7200 RPM and 15,000 RPM, enabling them to transfer data at a higher rate. Hard drives are gradually being replaced by all-electronic secondary storage, called solid-state disk (SSD). The technology of SSD is similar to RAM, except that data isn't lost when the power is turned off. Because it has no moving parts, it is faster and consumes less power than a hard drive. At this early stage in the transition, SSD is more expensive and has less storage capacity, but those factors can be expected to change as the technology advances.

A DVD drive comes with the machine. The ad describes it as being 8×, which means it can read data from a DVD as much as eight times faster than a DVD movie player. Slot load means that you insert a DVD into a narrow slit in the edge of the laptop, rather than pressing a button and having a drawer slide out to accept the disk. DL stands for dual layer, which means that the drive can work with second generation DVDs that store nearly twice as much data by using two layers of recording surface. Following the DVD acronym are the symbols +/-RW. The R indicates that the drive can record on special DVDs that are writeable. There are actually two standards for how these disks are made, called -R and +R, and the +/-

indicates that the drive is compatible with both standards. A DVD+/-R can have data written just once. After that, it can be read any number of times, but no more writing is allowed. Another type of DVD, called RW (for rewritable) can be written more than once. This laptop also supports RW disks. While DVD drives are still the most popular, laptops are starting to shift to the newer Blu-Ray format that has higher capacity and is being used to distribute high-definition movies.

The next line of the ad describes its wireless networking support. 802.11 is the number of a standard that has been defined by the Institute of Electrical and Electronics Engineers (IEEE), an engineering professional society. There are three accepted versions of the standard, a, g, and n. The original was 802.11a. The 802.11g version supports communication over longer distances, but at a slightly slower speed. With 802.11n, both greater speed and distance are achieved. This laptop is compatible with all three standards. Bluetooth is another form of wireless network, but it operates at much shorter ranges with a relatively weak signal. Typical uses for Bluetooth are to connect with a wireless keyboard, mouse, earphones, or for transferring data to and from a cell phone. There have been multiple versions of the Bluetooth standard, each adding various features, and version 3.0 was adopted in 2009.

Of course, laptops run on batteries. Even so, they still consume quite a bit of power. When a laptop is idle, with the screen turned off, it will use just a few watts. But in playing a game that makes heavy use of both processors and the GPU it can draw 50 watts. That's far more energy than normal rechargeable batteries can supply, so special technology, based on the metal lithium, provides high electrical storage capacity. This laptop's battery can store 85 watt-hours of energy, which means that it could supply 85 watts for one hour, or 42.5 watts for two hours, etc. More capacity means a longer time without recharging, but it also adds size and weight to the laptop.

Next the ad has a long list of external connections (often called ports). USB, or universal serial bus, uses a cable to transfer data. As its name suggests, it can connect to just about anything, including an external hard drive, a digital camera, a printer, a scanner, a music player, and so on. This laptop has two second-generation USB ports, which transfer data faster than USB 1.0. HDMI stands for high definition multimedia interface, which can send digital video and audio to, for example, a home theater system. A 15-pin VGA port is used to connect the laptop to an external analog monitor or projector. An Ethernet cable connects to a router or cable modem for wired network access. There are three versions of Ethernet that provide 10, 100, and 1000 million bits per second of data transfer capacity, and this laptop handles all three. IEEE 1394 is another communication standard, also called Firewire. This port provides very fast digital data transfer and is commonly used for connecting high definition camcorders and high performance disk drives. The express card slot allows the user to insert a small circuit board to provide extra functionality, such as a solid-state disk or wireless communication with a cellular phone network. Lastly, we see that we can connect analog audio inputs and outputs, such as electronic musical instruments and headphones, plus an external microphone.

Putting sizes in perspective

Admiral Grace Murray Hopper demonstrated the relative sizes of computer jargon by displaying a coil of wire nearly 1000 feet long, a short piece of wire about as long as your forearm, and a bag containing grains of pepper. She would point out that the wire coil was the distance traveled by an electron along the wire in the space of a microsecond. The short piece of wire was the distance traveled by an electron along the wire in the space of a nanosecond. The grains of pepper represented the distance traveled by an electron in a picosecond. She would admonish the members of her audience to remember their nanoseconds.

Physical size and weight are important parameters for a laptop that will be carried regularly. This is a mid-size, mid-weight model. At 5.6 pounds, it weighs over twice as much as this book. A lightweight laptop has roughly the same weight as this book, and heavier models, sometimes called desktop replacements, can weigh in at around 8 pounds. Generally, to reduce weight, the size shrinks and we give up features and battery life. However, it is also possible to reduce weight by replacing plastic in the case with aluminum, but for greater cost.

Lastly, the ad lists software that is preinstalled on the laptop. These include the operating system (Windows 7®), the Microsoft® Office suite of programs that includes a word processor, spreadsheet, and so on for performing common tasks, and a 3-year subscription to updates for a malware detection package. Malware is software that intends to do harm, and comes in many forms, such as viruses that can take over your computer when you open a downloaded file. Malware detection software constantly watches for such programs in files and web content to prevent them from running. But hackers are constantly creating new forms of malware, so it is necessary to regularly update the detection software to keep up with the latest threats.

Within this ad, multiple size measures have been used. Let's summarize the prefixes that are used frequently in computing.

Power of 10	Power of 2	Value of Power of 2	Prefix	Abbreviation	Derivation
10^{-12}			pico	p	Italian for little
10^{-9}			nano	n	Greek for dwarf
10^{-6}			micro	µ	Greek for small
10^{-3}			milli	m	Latin for thousandth
10^{3}	2^{10}	1024	kilo	K	Greek for thousand
10^{6}	2^{20}	1,048,576	mega	M	Greek for large
10^{9}	2^{30}	1,073,741,824	giga	G	Greek for giant
10^{12}	2^{40}	not enough room	tera	T	Greek for monster
10^{15}	2^{50}	not enough room	peta	P	Greek prefix for five

Did you notice that we used powers of 10 when referring to time and powers of 2 when referring to storage? Time is expressed in multiples of seconds in decimal notation. Storage capacity is expressed in multiples of bytes in binary notation. If you keep this distinction in mind, it is clear that K is 1000 when referring to speed and 1024 when referring to storage.

We now move from the specific to the general. In the next several sections we look at each of the pieces of hardware that make up a computer from the logical level, rather than from a specific computer configuration.

5.2 Stored-Program Concept

A major defining point in the history of computing was the realization in 1944–1945 that data and instructions to manipulate the data were logically the same and could be stored in the same place. The computer design built upon this principle, which became known as the *von Neumann architecture*, is still the basis for computers today. Although the name honors John von Neumann, a brilliant mathematician who worked on the construction of the atomic bomb, the idea probably originated with J. Presper Eckert and John Mauchly, two other early pioneers who worked on the ENIAC at the Moore School at the University of Pennsylvania during the same time period.

?

Does it matter who was the father of the modern computer?

All of the people involved in the research and development of electronic computing devices in the late 1930s and 1940s undoubtedly contributed to the computer as we know it. This list includes John Atanasoff, Clifford Berry, and Konrad Zuse, in addition to von Neumann, Eckert, and Mauchly.

In 1951, Sperry Rand bought the patent for the ENIAC and its underlying concepts and began charging royalties to other computer manufacturers. Not wanting to pay royalties, Honeywell researched the history of modern computers and presented evidence that the work of John Atanasoff at Iowa State College had directly influenced Mauchly and Eckert. Because of this evidence, the patent for the ENIAC was invalidated in 1973.

von Neumann Architecture

Another major characteristic of the von Neumann architecture is that the units that process information are separate from the units that store information. This characteristic leads to the following five components of the von Neumann architecture, shown in Figure 5.1:

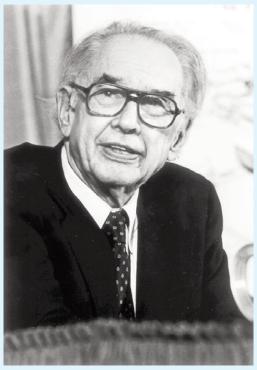
- The memory unit that holds both data and instructions
- The arithmetic/logic unit that is capable of performing arithmetic and logic operations on data
- The input unit that moves data from the outside world into the computer
- The output unit that moves results from inside the computer to the outside world
- The control unit that acts as the stage manager to ensure that all the other components act in concert

Memory

Recall from the discussion of number systems that each storage unit, called a bit, is capable of holding a 1 or a 0; these bits are grouped together into bytes (8 bits), and these bytes are in turn grouped together into words. Memory is a collection of cells, each with a unique physical address. We use the generic word *cell* here rather than byte or word, because the number of bits in each addressable location, called the memory's **addressability**, varies from one machine to another. Today, most computers are byte addressable.

Addressability The number of bits stored in each addressable location in memory

John Vincent Atanasoff


Courtesy of ISU Photo Service

John Vincent Atanasoff was born in Hamilton, New York, on October 4, 1903, one of nine children. When he was about ten, his father bought a new slide rule. After reading the instructions, John Vincent became more interested in the mathematics involved than in the slide rule itself. His mother picked up on his interest and helped him study his father's old college algebra book. He continued his interest in mathematics and science and graduated from high school in two years. His family moved to Old Chicara, Florida, where John Vincent graduated from the University of Florida in 1925 with a degree in electrical engineering because the university didn't offer a degree in theoretical physics. A year later, he received a master's degree in mathematics from Iowa State College. In 1930, after receiving his PhD in theoretical physics, he returned to Iowa State College as an assistant professor in mathematics and physics.

Dr. Atanasoff became interested in finding a machine that could do the complex mathematical work he and his graduate students were doing. He examined computational devices in existence at that time, including the Monroe calculator and the IBM tabulator. Upon concluding that these machines were too slow and inaccurate, he became obsessed with finding a solution. He said that at night in a tavern after a drink of bourbon he began generating ideas of how to build this computing device. It would be electronically operated and would compute by direct logical action rather than enumeration, as in analog devices. It would use binary numbers rather than decimal numbers, condensers for memory, and a regenerative process to avoid lapses due to leakage of power.

In 1939, with a $650 grant from the school and a new graduate assistant named Clifford Berry, Dr. Atanasoff began work on the first prototype of the Atanasoff Berry Computer (ABC) in the basement of the physics building. The first working prototype was demonstrated that year.

In 1941, John Mauchly, a physicist at Ursinus College whom Dr. Atanasoff had met at a conference, came to Iowa State to visit the Atanasoffs and see a demonstration of the ABC machine. After extensive discussions, Mauchly left with papers describing its design. Mauchly and J. Presper Eckert continued their work on a computation device at the Moore School of Electrical Engineering at the University of Pennsylvania. Their machine, the ENIAC, completed in 1945, became known as the first computer.

Dr. Atanasoff went to Washington in 1942 to become director of the Underwater Acoustics Program at the Naval Ordnance Laboratory, leaving the patent application for the ABC computer in the hands of the Iowa State attorneys. The patent application was never filed and the ABC was eventually dismantled without either Atanasoff or Berry being notified. After the war, Dr. Atanasoff was chief scientist for the Army Field Forces and director of the Navy Fuse program at the Naval Ordnance Laboratory.

In 1952, Dr. Atanasoff established The Ordnance Engineering Corporation, a research and engineering firm, which was later sold to

» continued

John Vincent Atanasoff, continued

Aerojet General Corporation. He continued to work for Aerojet until he retired in 1961.

Meanwhile, in 1947 Mauchly and Eckert applied for the patent on their ENIAC computer. Sperry Rand brought suit. The subsequent trial lasted 135 working days and filled more than 20,000 pages of transcript from the testimony of 77 witnesses, including Dr. Atanasoff. Judge Larson found that Mauchly and Eckert "did not themselves first invent the automatic electronic digital computer, but instead derived that subject matter from one Dr. John Vincent Atanasoff."

In 1990, President George Bush acknowledged Dr. Atanasoff's pioneering work by awarding him the National Medal of Technology. Dr. Atanasoff died on June 15, 1995.

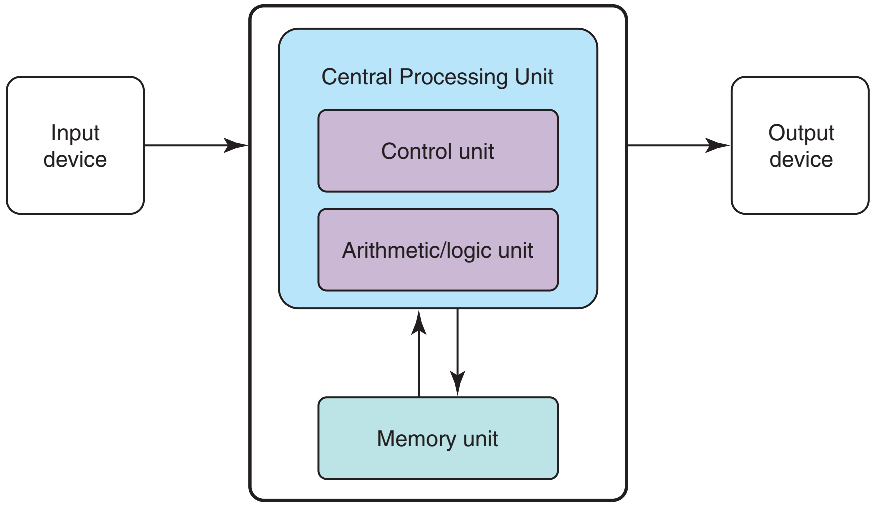


FIGURE 5.1 The von Neumann architecture

The ad in the previous section describes a memory of 4×2^{30} bytes. This means that each of the 4GB is uniquely addressable. Therefore, the addressability of the machine is 8 bits. The cells in memory are numbered consecutively beginning with 0. For example, if the addressability is 8, and there are 256 cells of memory, the cells would be addressed as follows:

Address	Contents
00000000	11100011
00000001	10101001
⋮	⋮
11111100	00000000
11111101	11111111
11111110	10101010
11111111	00110011

What are the contents of address 11111110? The bit pattern stored at that location is 10101010. What does it mean? We can't answer that question in the abstract. Does location 11111110 contain an instruction? An integer with a sign? A two's complement value? Part of an image? Without knowing what the contents represent, we cannot determine what it means: It is just a bit pattern. We must apply an interpretation on any bit pattern to determine the information it represents.

When referring to the bits in a byte or word, the bits are numbered from right to left beginning with zero. The bits in address 11111110 are numbered as follows:

7	6	5	4	3	2	1	0	← Bit position
1	0	1	0	1	0	1	0	← Contents

Arithmetic/Logic Unit

Arithmetic/logic unit (ALU) The computer component that performs arithmetic operations (addition, subtraction, multiplication, division) and logical operations (comparison of two values)

Register A small storage area in the CPU used to store intermediate values or special data

The **arithmetic/logic unit** (**ALU**) is capable of performing basic arithmetic operations such as adding, subtracting, multiplying, and dividing two numbers. This unit is also capable of performing logical operations such as AND, OR, and NOT. The ALU operates on words, a natural unit of data associated with a particular computer design. Historically the word length of a computer has been the number of bits processed at once by the ALU. However, the current Intel line of processors has blurred this definition by defining the word length to be 16 bits. The processor can work on words (16 bits), double words (32 bits), and quadwords (64 bits). In the rest of this discussion we continue to use "word" in its historical sense.

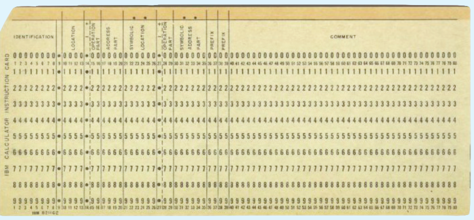
Most modern ALUs have a small number of special storage units called **registers**. These registers contain one word and are used to store information that is needed again immediately. For example, in the calculation of

One * (Two + Three)

Two is first added to Three and the result is then multiplied by One. Rather than storing the result of adding Two and Three in memory and then retrieving it to multiply it by One, the result is left in a register and

Who Was Herman Hollerith?

In 1889 the United States Census Bureau realized that unless it found a better way to count the 1890 census, the results might not be tabulated before the next required census in 1900. Herman Hollerith had designed a method of counting based on cards with holes punched in them. This method was used for tabulating the census and the cards became known as Hollerith cards. Hollerith's electrical tabulating system led to the founding of the company known today as IBM.



Courtesy of Douglas W. Jones at the University of Iowa

the contents of the register are multiplied by One. Access to registers is much faster than access to memory locations.

Input/Output Units

All of the computing power in the world wouldn't be useful if we couldn't input values into the calculations from outside or report to the outside the results of those calculations. Input and output units are the channels through which the computer communicates with the outside world.

An **input unit** is a device through which data and programs from the outside world are entered into the computer. The first input units interpreted holes punched on paper tape or cards. Modern-day input devices include the keyboard, the mouse, and the scanning devices used at supermarkets.

An **output unit** is a device through which results stored in the computer memory are made available to the outside world. The most common output devices are printers and displays.

Control Unit

The **control unit** is the organizing force in the computer, for it is in charge of the fetch–execute cycle, discussed in the next section. There are two special registers in the control unit. The **instruction register** (**IR**) contains the instruction that is being executed, and the **program counter** (**PC**) contains the address of the next instruction to be executed. Because the ALU and the control unit work so closely together, they are often thought of as one unit called the central processing unit, or **CPU**.

Figure 5.2 shows a simplified view of the flow of information through the parts of a von Neumann machine. The parts are connected to one another by a collection of wires called a bus, through which data travels in

Input unit A device that accepts data to be stored in memory

Output unit A device that prints or otherwise displays data stored in memory or makes a permanent copy of information stored in memory or another device

Control unit The computer component that controls the actions of the other components so as to execute instructions in sequence

Instruction register (IR) The register that contains the instruction currently being executed

Program counter (PC) The register that contains the address of the next instruction to be executed

CPU The central processing unit, a combination of the arithmetic/logic unit and the control unit; the "brain" of a computer that interprets and executes instructions

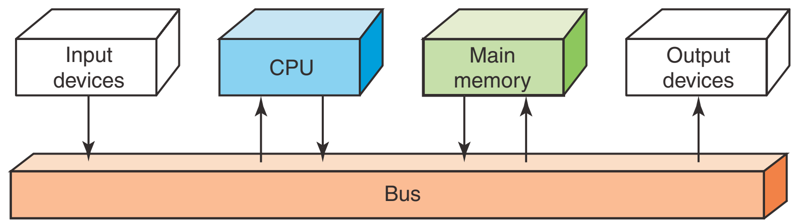


FIGURE 5.2 Data flow through a von Neumann machine

Bus width The number of bits that can be transferred in parallel over the bus

Cache memory A type of small, high-speed memory used to hold frequently used data

Pipelining A technique that breaks an instruction into smaller steps that can be overlapped

Motherboard The main circuit board of a personal computer

the computer. Each bus carries three kinds of information: address, data, and control. An address is used to select the memory location or device to which data will go, or from which it will be taken. Data then flows over the bus between the CPU, memory, and I/O devices. The control information is used to manage the flow of addresses and data. For example, a control signal will typically be used to determine the direction in which the data is flowing, either to or from the CPU. The **bus width** is the number of bits that it can transfer simultaneously. The wider the bus, the more address or data bits it can move at once.

Because memory accesses are very time consuming relative to the speed of the processor, many architectures provide **cache memory**. Cache memory is a small amount of fast-access memory into which copies of frequently used data are stored. Before a main memory access is made, the CPU checks whether the data is stored in the cache memory. **Pipelining** is another technique used to speed up the fetch–execute cycle. This technique splits an instruction into smaller steps that can be overlapped.

In a personal computer, the components in a von Neumann machine reside physically in a printed circuit board called the **motherboard**. The motherboard also has connections for attaching other devices to the bus, such as a mouse, a keyboard, or additional storage devices. (See the section on secondary storage devices later in this chapter.)

So just what does it mean to say that a machine is an *n*-bit processor? The variable *n* usually refers to the number of bits in the CPU general registers: Two *n*-bit numbers can be added with a single instruction. It also can refer to the width of the address bus, which is the size of the addressable memory—but not always. In addition, *n* can refer to the width of the data bus—but not always.

The Fetch–Execute Cycle

Before looking at *how* a computer does what it does, let's look at *what* it can do. The definition of a computer outlines its capabilities: A computer is a device that can store, retrieve, and process data. Therefore, all of the instructions that we give to the computer relate to storing, retrieving, and processing data. In Chapters 6 and 9, we look at various languages that we can use to give instructions to the computer. For our examples here, we use simple English-like instructions.

Recall the underlying principle of the von Neumann machine: Data and instructions are stored in memory and treated alike. This means that instructions and data are both addressable. Instructions are stored in contiguous memory locations; data to be manipulated are stored together in a different part of memory. To start the fetch–execute cycle, the address of the first instruction is loaded into the program counter.

The processing cycle includes four steps:

- Fetch the next instruction.
- Decode the instruction.
- Get data if needed.
- Execute the instruction.

Let's look at each of these steps in more detail. The process starts with the address in memory of the first instruction being stored in the program counter.

Fetch the Next Instruction

The program counter contains the address of the next instruction to be executed, so the control unit goes to the address in memory specified in the PC, makes a copy of the contents, and places the copy in the instruction register. At this point the IR contains the instruction to be executed. Before going on to the next step in the cycle, the PC must be updated to hold the address of the next instruction to be executed when the current instruction has been completed. Because the instructions are stored contiguously in memory, adding the number of bytes in the current instruction to the program counter should put the address of the next instruction into the PC. Thus the control unit increments the PC. It is possible that the PC may be changed later by the instruction being executed.

In the case of an instruction that must get additional data from memory, the ALU sends an address to the memory bus, and the memory responds by returning the value at that location. In some computers, data retrieved from memory may immediately participate in an arithmetic or logical operation. Other computers simply save the data returned by the memory into a register for processing by a subsequent instruction. At the end of execution, any result from the instruction may be saved either in registers or in memory.

Decode the Instruction

To execute the instruction in the instruction register, the control unit has to determine what instruction it is. It might be an instruction to access data from an input device, to send data to an output device, or to perform some operation on a data value. At this phase, the instruction is decoded into control signals. That is, the logic of the circuitry in the CPU determines which operation is to be executed. This step shows why a computer can execute only instructions that are expressed in its own machine language. The instructions themselves are literally built into the circuits.

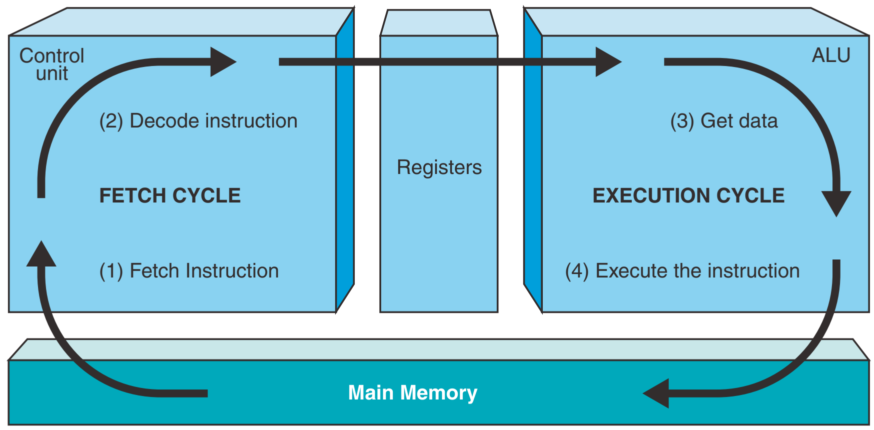


FIGURE 5.3 The fetch-execute cycle

Get Data if Needed

The instruction to be executed may potentially require additional memory accesses to complete its task. For example, if the instruction says to add the contents of a memory location to a register, the control unit must get the contents of the memory location.

Execute the Instruction

Once an instruction has been decoded and any operands (data) fetched, the control unit is ready to execute the instruction. Execution involves sending signals to the arithmetic/logic unit to carry out the processing. In the case of adding a number to a register, the operand is sent to the ALU and added to the contents of the register.

When the execution is complete, the cycle begins again. If the last instruction was to add a value to the contents of a register, the next instruction probably says to store the results into a place in memory. However, the next instruction might be a control instruction—that is, an instruction that asks a question about the result of the last instruction and perhaps changes the contents of the program counter.

Figure 5.3 summarizes the fetch–execute cycle.

Hardware has changed dramatically in the last half-century, yet the von Neumann machine remains the basis of most computers today. As Alan Perlis, a well-known computer scientist, said in 1981, “Sometimes I think the only universal in the computing field is the fetch–execute cycle.”[1] This statement is still true today, nearly three decades later.

RAM and ROM

As mentioned, RAM stands for random-access memory. RAM is memory in which each cell (usually a byte) can be directly accessed. Inherent in the idea of being able to access each location is the ability to *change* the contents of each location. That is, storing something else into that place can change the bit pattern in each cell.

In addition to RAM, most computers contain a second kind of memory, called ROM. ROM stands for read-only memory. The contents in locations in ROM cannot be changed. Their contents are permanent and cannot be altered by a stored operation. Placing the bit pattern in ROM is called *burning*. The bit pattern is burned either at the time the ROM is manufactured or at the time the computer parts are assembled.

RAM and ROM are differentiated by a very basic property: RAM is volatile; ROM is not. This means that RAM does not retain its bit configuration when the power is turned off, but ROM does. The bit patterns within ROM are permanent. Because ROM is stable and cannot be changed, it is used to store the instructions that the computer needs to start itself. Frequently used software is also stored in ROM so that the system does not have to read the software in each time the machine is turned on. Main memory usually contains some ROM along with the general-purpose RAM.

Secondary Storage Devices

As mentioned earlier, an input device is the means by which data and programs are entered into the computer and stored into memory. An output device is the means by which results are sent back to the user. Because most of main memory is volatile and limited, it is essential that there be other types of storage devices where programs and data can be stored when they are no longer being processed or when the machine is not turned on. These other types of storage devices (other than main memory) are called *secondary* or *auxiliary* storage devices. Because data must be read from them and written to them, each secondary storage device is also an input and an output device.

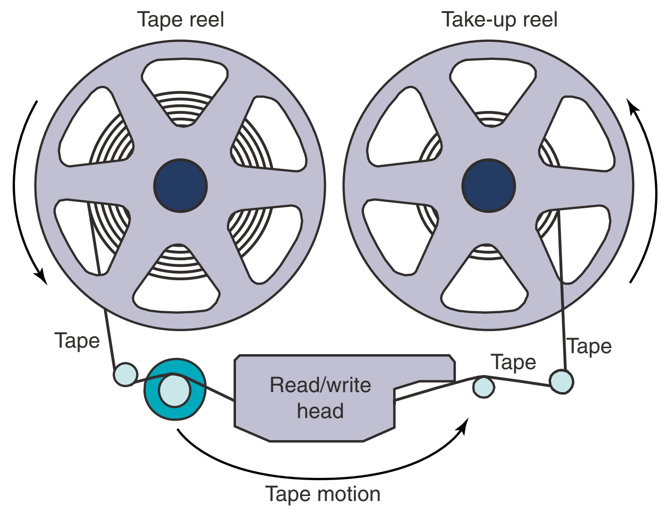
Secondary storage devices can be installed within the computer box at the factory or added later as needed. Because these devices can store large quantities of data, they are also known as mass storage devices. For example, the hard disk drive that comes with the laptop specified in the ad can store 500×2^{30} bytes as opposed to 4×2^{30} bytes in main memory.

The next sections describe some secondary storage devices.

Magnetic Tape

Card readers and card punches were among the first input/output devices. Paper tape readers and punches were the next input/output devices. Although paper tapes, like cards, are permanent, they cannot hold much data. The first truly mass auxiliary storage device was the *magnetic tape drive*. A magnetic tape drive is like a tape recorder and is most often used to back up (make a copy of) the data on a disk in case the disk is later damaged. Tapes come in several varieties, from small streaming-tape cartridges to large reel-to-reel models.

FIGURE 5.4 A magnetic tape



Tape drives have one serious drawback: To access data in the middle of the tape, all the data before the piece you want must be accessed and discarded. Although modern streaming-tape systems have the capability of skipping over segments of tape, the tape must physically move through the read/write heads. Any physical movement of this type is time-consuming. See Figure 5.4.

Magnetic Disks

A *disk drive* is a cross between a compact disk player and a tape recorder. A read/write head (similar to the record/playback head in a tape recorder) travels across a spinning magnetic disk, retrieving or recording data. As on a compact disk, the heads travel directly to the information desired; as on a tape, the information is stored magnetically.

Disks come in several varieties, but all of them consist of a thin disk made out of magnetic material. The surface of each disk is logically organized into **tracks** and **sectors**. Tracks are concentric circles around the surface of the disk. Each track is divided into sectors. Each sector holds a **block** of information as a continuous sequence of bits. [See Figure 5.5(a).] The figure depicts the original layout of data on a disk, in which each track has the same number of sectors, and each sector holds the same number of bits. The blocks of data nearer the center were more densely packed. On modern disks, there are fewer sectors near the middle and more toward the outside. The actual number of tracks per surface and the number of sectors per track vary, but 512 bytes or 1024 bytes is common. (The power of 2 strikes again.) The locations of the tracks and sectors are marked magnetically when a disk is formatted; they are not physically part of the disk.

> **Track** A concentric circle on the surface of a disk
>
> **Sector** A section of a track
>
> **Block** The information stored in a sector

The read/write head in a disk drive is positioned on an arm that moves from one track to another. [See Figure 5.5(b).] An input/output instruction

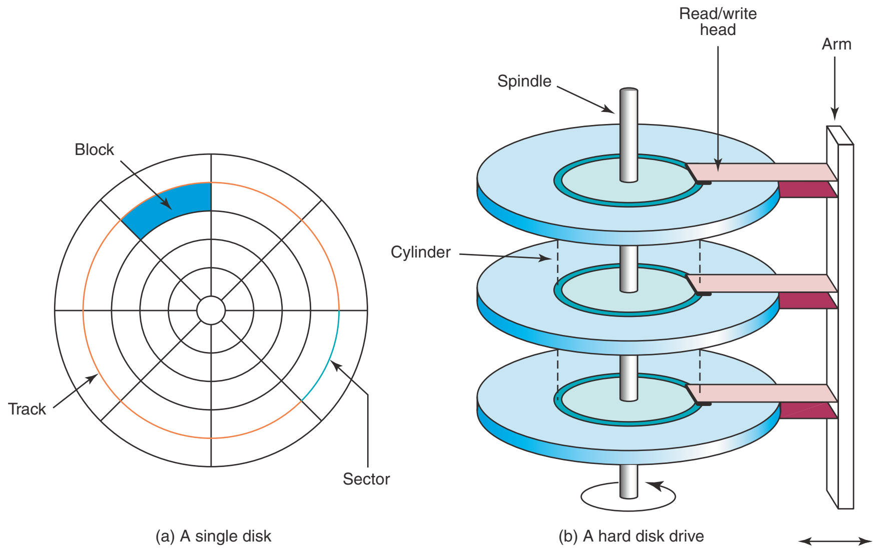


FIGURE 5.5 The organization of a magnetic disk

specifies the track and sector. When the read/write head is over the proper track, it waits until the appropriate sector is beneath the head; it then accesses the block of information in that sector. This process gives rise to four measures of a disk drive's efficiency: **seek time**, **latency**, **access time**, and **transfer rate**. Seek time is the time it takes for the read/write head to get into position over the specified track. Latency is the time it takes for the specified sector to spin to the read/write head. The average latency is one-half the time for a full rotation of the disk. For this reason, latency is also called rotation delay. Access time is the sum of seek time and latency. Transfer rate is the rate at which data is transferred from the disk to memory.

Now let's look at some of the varieties of disks. One classification of disk is hard versus floppy. These terms refer to the flexibility of the disk itself. The original floppy disk, introduced in the 1970s, was 8″ in diameter and even its case was floppy. By the time of the rise in personal computers in the late 1970s, the floppy disk had been reduced in size to 5 1/2″ in diameter. Today's generic "floppy" disks are 3 1/2″ in diameter, encased in a hard plastic cover, and capable of storing 1.44MB of data. Newer machines do not automatically have built-in drives for these disks as they did a few years ago, but drives for them can be added.

Hard disks actually consist of several disks—this sounds strange, so let's explain. Let's call the individual disks platters. Hard disks consist of several platters attached to a spindle that rotates. Each platter has its own read/write head. All of the tracks that line up under one another are called a **cylinder** [see Figure 5.5(b)]. An address in a hard drive consists of the cylinder number, the surface number, and the sector. Hard drives rotate at much higher speeds

Seek time The time it takes for the read/write head to get positioned over the specified track

Latency The time it takes for the specified sector to be in position under the read/write head

Access time The time it takes for a block to start being read; the sum of seek time and latency.

Transfer rate The rate at which data moves from the disk to memory

Cylinder The set of concentric tracks on all surfaces than floppy drives do, and the read/write heads don't actually touch the surface of the platters but rather float above them. A typical hard disk drive rotates at 7200 revolutions per minute. Laptop hard disks usually spin at 5400 RPM, conserving battery power. The disks in high performance servers may run at 15,000 RPM, providing lower latency and a higher transfer rate.

CDs and DVDs

The world of compact discs and their drivers looks like acronym soup. The ad we examined used the acronym: DVD +/–/RW. In addition, we have to decipher CD-DA, CD-RW, and DVD.

Let's look for a moment at the acronym CD. CD, of course, stands for compact disk—you probably have a collection of them with recorded music. A CD drive uses a laser to read information that is stored optically on a plastic disk. Rather than having concentric tracks, a CD has one track that spirals from the inside out. As on magnetic disks, this track is broken into sectors. A CD has the data evenly packed over the whole disk, so more information is stored in the track on the outer edges and read in a single revolution. To make the transfer rate consistent throughout the disk, the rotation speed varies depending on the position of the laser beam.

The other letters attached to CD refer to various properties of the disk, such as formatting and whether the information on the disk can be changed. CD-DA is the format used in audio recordings; CD-DA stands for compact disk–digital audio. Certain fields in this format are used for timing information. A sector in a CD-DA contains 1/75 of a second of music.

CD-ROM is the same as CD-DA but the disk is formatted differently. Data is stored in the sectors reserved for timing information in CD-DA. ROM stands for read-only memory. As we said earlier, read-only memory means that the data is permanent and cannot be changed. A sector on a CD-ROM contains 2KB of data. CD-ROM capacity is in the neighborhood of 600MB.

CD-R stands for recordable, allowing data to be written after it is manufactured. The contents of a CD-R cannot be changed after data is recorded on it. A CD-RW is rewritable, meaning that it can have data recorded on it multiple times.

> **What is FiOS?**
> FiOS ("fiber optic service") is Verizon's fiber optic cable (superceding copper wire cable) that provides Internet, telephone, and television service to homes. The service is not available in all towns or areas. According to some, FiOS is an expensive gamble—it will cost $23 billion to run fiber to 18 million homes—that may never pay off. FiOS throughput speeds are variable, depending on a variety of factors, such as the physical location, competing broadband providers, and customer budget considerations. As of June 2008, Verizon offered bandwidth tiers (Mbits download and upload) of 10/2, 20/5, 20/20, and 50/20.

The most common format for distributing movies is now a DVD, which stands for digital versatile disk (although the acronym generally stands on its own these days). Because of its large storage capacity, a DVD is well suited to hold multimedia presentations that combine audio and video.

DVDs come in multiple forms: DVD+R, DVD-R, DVD+RW, and DVD-RW, and each of these may be preceded by DL. As we noted in describing the ad, the + and - refer to two competing formats. As with CD, R means recordable and RW means rewritable. DL stands for dual layer, which nearly doubles the capacity of a DVD. DVD-R has a capacity of 4.7GB while DL DVD-R can hold 8.5GB. More recently, Blu-Ray disks with 25GB capacity and DL 50GB capacity have been introduced. Writable versions are also available. The name Blu-Ray refers to its use of a blue laser instead of the red laser in CD and DVD drives.

Note that the × used in rating CD and DVD speeds indicates the relative speed of access compared with a standard CD or DVD player. When evaluating these devices, be aware that the higher speeds listed represent maximums that are usually attainable only when retrieving data from certain parts of the disk. They are not averages. Therefore, faster may not be better in terms of the added cost.

Flash Drives

IBM introduced the flash drive in 1998 as an alternative to floppy disks. Figure 5.6 shows a flash drive (or thumb drive), which uses flash memory, a nonvolatile computer memory that can be erased and rewritten. The drive is integrated with a USB (universal serial bus). Most computers today do not come with floppy disks, but they do come with USB ports. In 2010, this little (thumb-sized) 8GB storage device could be bought for less than $20.

FIGURE 5.6 A flash drive

Flash memory is also being used to build solid state disks (SSD) that can directly replace a hard disk. Because SSD is all electronic and has no moving parts, it is faster and consumes less power than a hard disk. Even so, its storage elements can eventually wear out, meaning that it can suffer failures just as a hard disk can.

Touch Screens

We've seen how secondary memory devices provide locations in which to store programs and data used by the CPU. Other input/output (I/O) devices allow the human user to interact with an executing program. Many of these are commonplace—we often provide information through a keyboard and mouse, and we usually view information displayed on a monitor screen. Other input devices include bar code readers and image scanners; other output devices include printers and plotters.

Let's examine one particular type of I/O device in some detail. A *touch screen* displays text and graphics like a regular monitor, but it can also detect and respond to the user touching the screen with a finger or stylus. Usually, an I/O device serves either as an input device or an output device. A touch screen serves as both.

You've probably seen touch screens used in a variety of situations such as information kiosks, restaurants, and museums. Figure 5.7 shows someone using a touch screen. These devices are most helpful in situations in which complex input is not needed, and they have the added benefit of

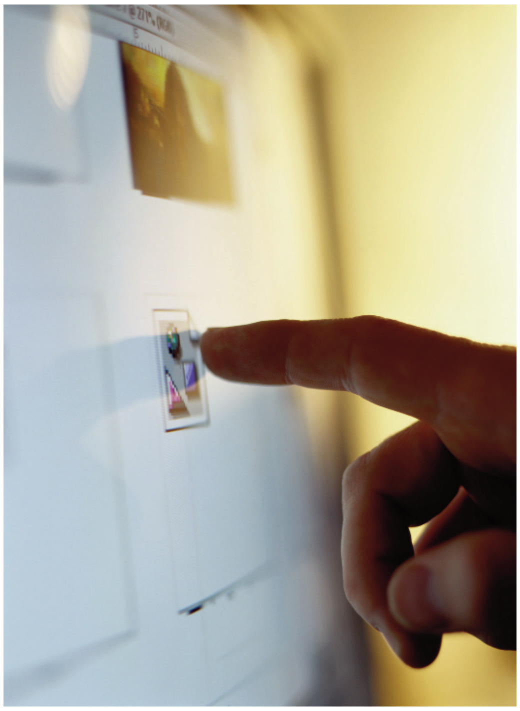


FIGURE 5.7 A touch screen
© Randy Allbritton/Photodisc/Getty Images

being fairly well protected. It's far better for a waiter at a restaurant to make a few choices using a touch screen than to have to deal with a keyboard, which has more keys than necessary (for the task) and may easily get damaged from food and drink.

A touch screen not only detects the touch, but also knows where on the screen it is being touched. Choices are often presented using graphical buttons that the user selects by touching the screen where the button is positioned. In this sense, using a touch screen is not much different from using a mouse. The mouse position is tracked as the mouse is moved; when the mouse button is clicked, the position of the mouse pointer determines which graphical button is pushed. In a touch screen, the location at which the screen is touched determines which button is pushed.

So how does a touch screen detect that it is being touched? Furthermore, how does it know where on the screen it is being touched? Several technologies are used today to implement touch screens. Let's briefly explore them.

A *resistive* touch screen is made up of two layers—one with vertical lines and one with horizontal lines of electrically conductive material. The two layers are separated by a very small amount of space. When the top layer is pressed, it comes in contact with the second layer, which allows

?

Evolution of the BlackBerry®

The first BlackBerry device, known as the RIM Inter@ctive Pager 850, was introduced to the public in 1999 by Research in Motion, Ltd. The device appeared on the hit NBC television show *ER*, and is known as the first mobile email machine. Each year brought newer and better models. For example, 2006 saw the introduction of the BlackBerry Pearl, which, with its smaller size, digital camera, enhanced software, and media player, appealed to the mass public in a way that the previous models hadn't. In 2008, the BlackBerry Storm was introduced as a sleek, touch screen handset built to compete with Apple's iPhone. How popular is the BlackBerry? Even Barack Obama refused to give up his BlackBerry when he became president.

electrical current to flow. The specific vertical and horizontal lines that make contact dictate the location on the screen that was touched.

A *capacitive* touch screen has a laminate applied over a glass screen. The laminate conducts electricity in all directions, and a very small current is applied equally on the four corners. When the screen is touched, current flows to the finger or stylus. The current is so low that the user doesn't even feel it. The location of the touch on the screen is determined by comparing the strength of the flow of electricity from each corner.

An *infrared* touch screen projects crisscrossing horizontal and vertical beams of infrared light just over the surface of the screen. Sensors on opposite sides of the screen detect the beams. When the user breaks the beams by touching the screen, the location of the break can be determined.

A *surface acoustic wave* (SAW) touch screen is similar to an infrared touch screen except that it projects high-frequency sound waves across the horizontal and vertical axes. When a finger touches the surface, the corresponding sensors detect the interruption and determine the location of the touch.

Note that a gloved hand could be used in resistive, infrared, and SAW touch screens, but cannot be used with capacitive screens, which rely on current flowing to the touch point.

5.3 Embedded Systems

Embedded systems are computers that are designed to perform a narrow range of functions as part of a larger system. Typically, an embedded system is housed on a single microprocessor chip with the programs stored in ROM. Virtually all appliances that have a digital interface—watches, microwaves, VCRs, cars—utilize embedded systems. In fact, embedded systems are everywhere: From consumer electronics, to kitchen appliances, to automobiles, to networking equipment, to industrial control systems, you find embedded systems lurking in the device. Some embedded systems include an operating system, but many are so specialized that the entire logic can be implemented as a single program.[2]

Early embedded systems were stand-alone 8-bit microprocessors with their own homegrown operating system. Today, they range from 8-bit controllers to 32-bit digital signal processors (DSPs) to 64-bit RISC (Reduced Instruction Set) chips. More and more embedded systems are based on networks of distributed microprocessors that communicate through wired and wireless buses, remotely monitored and controlled by regular network management communications protocols.

In fact, the term *embedded system* is nebulous because it encompasses just about everything except desktop PCs. The term originated because the first such computers were physically embedded within a product or device and could not be accessed. Now the term refers to any computer that is preprogrammed to perform a dedicated or narrow range of functions as part of a larger system. The implication is that there is only minimal end-user or operator intervention, if any.

Because the average person encounters an embedded system only in his or her kitchen, entertainment room, or car, we tend to equate these systems with hardware. In reality, programs must be written and burned into the read-only memory that comes with the system to make it accomplish its assigned function. Given that programs cannot be developed and tested on the embedded processor itself, how are they implemented? Programs are written on a PC and compiled for the target system, where the executable code is generated for the processor in the embedded system.

In early embedded systems, the size of the code and the speed at which it executed were very important. Because assembly-language programs provided the best opportunity to streamline and speed up the code, they were used almost exclusively for embedded systems. Even when the C language became popular and cross-compilers for C to embedded systems became available, many programmers continued to use assembly language for this purpose. C programs are approximately 25% larger and slower, but are easier to write than assembly-language programs. Even today, the size of the ROM may dictate that the code be as small as possible, leading to an assembly-language program.[3]

5.4 Parallel Architectures[4]

If a problem can be solved in n time units on a computer with one processor (von Neumann machine), can it be solved in $n/2$ time units on a computer with two processors, or $n/3$ on a computer with three processors? This question has led to the rise of parallel computing architectures.

Parallel Computing

There are four general forms of parallel computing: bit level, instruction level, data level, and task level.

Bit-level parallelism is based on increasing the word size of a computer. In an 8-bit processor, an operation on a 16-bit data value would require two operations: one for the upper 8 bits and one for the lower 8 bits. A 16-bit processor could do the operation in one instruction. Thus *increasing* the word size *reduces* the number of operations on data values larger than the word size. The current trend is to use 64-bit processors.

Instruction-level parallelism is based on the idea that some instructions in a program can be carried out independently in parallel. For example, if a program requires operations on unrelated data, these operations can be done at the same time. A superscalar is a processor that can recognize this situation and take advantage of it by sending instructions to different functional units of the processor. Note that a superscalar machine does not have multiple processors but does have multiple execution resources. For example, it might contain separate ALUs for working on integer and real numbers, enabling it to simultaneously compute the sum of two integers and the product of two real numbers. Such resources are called execution units.

Data-level parallelism is based on the idea that a single set of instructions can be run on different data sets at the same time. This type of parallelism is called SIMD (single instructions, multiple data) and relies on a control unit directing multiple ALUs to carry out the same operation, such as addition, on different sets of operands. This approach, which is also called synchronous processing, is effective when the same process needs to be applied to many data sets. For example, increasing the brightness of an image involves adding a value to every one of several million pixels. These additions can all be done in parallel. See Figure 5.8.

Synchronous processing Multiple processors apply the same program in lockstep to multiple data sets

Task-level parallelism is based on the idea that different processors can execute different tasks on the same or different data sets. If the different processors are operating on the same data set, then it is analogous to pipelining in a von Neumann machine. When this organization is applied to data, the first processor does the first task. Then the second processor starts working on the output from the first processor, while the first processor applies its computation to the next data set. Eventually, each processor is working on one phase of the job, each getting material or data from the previous stage of processing, and each in turn handing over its work to the next stage. See Figure 5.9.

In a data-level environment, each processor is doing the same thing to a different data set. For example, each processor might be computing the

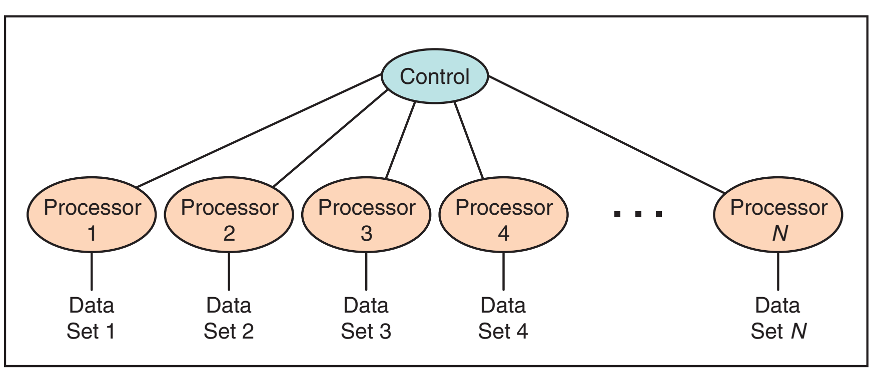


FIGURE 5.8 Processors in a synchronous computing environment

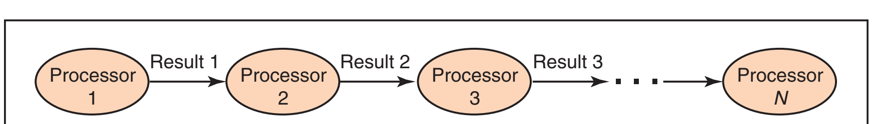


FIGURE 5.9 Processors in a pipeline

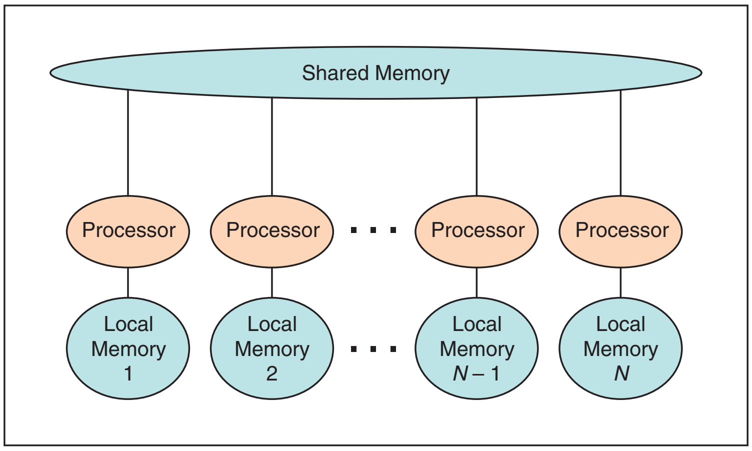


FIGURE 5.10 A shared-memory parallel processor

grades for a different class. In the pipelining task-level example, each processor is contributing to the grade for the same class. Another approach to task-level parallelism is to have different processors doing different things with different data. This configuration allows processors to work independently much of the time, but introduces problems of coordination among the processors. This leads to a configuration where each of the processors have both a local memory and a shared memory. The processors use the shared memory for communication, so the configuration is called a **shared memory parallel processor**. See Figure 5.10.

> **Shared memory parallel processor** The situation in which multiple processors share a global memory

Classes of Parallel Hardware

The classes of parallel hardware reflect the various types of parallel computing. Multicore processors have multiple independent cores, usually CPUs. Whereas a superscalar processor can issue multiple instructions to execution units, each multicore processor can issue multiple instructions to multiple execution units. That is, each independent core can have multiple execution units attached to it.

Symmetric multiprocessors (SMPs) have multiple identical cores. They share memory, and a bus connects them. The number of cores in an SMP is usually limited to 32 processors. A distributed computer is one in which multiple memory units are connected through a network. A cluster is a group of stand-alone machines connected through an off-the-shelf network. A massively parallel processor is a computer with many networked processors connected through a specialized network. This kind of device usually has more than 1000 processors.

The distinctions between the classes of parallel hardware are being blurred by modern systems. A typical processor chip today contains two to eight cores that operate as an SMP. These are then connected via a network to form a cluster. Thus, it is common to find a mix of shared and distributed memory in parallel processing. In addition, graphics processors that support general-purpose data-parallel processing may be connected to each of the multicore processors. Given that each of the cores is also applying instruction-level parallelism, you can see that modern parallel computers no longer fall into one or another specific classification. Instead, they typically embody all of the classes at once. They are distinguished by the particular balance that they strike among the different classes of parallel processing they support. A parallel computer that is used for science may emphasize data parallelism, whereas one that is running an Internet search engine may emphasize task-level parallelism.

Summary

The components that make up a computer cover a wide range of devices. Each component has characteristics that dictate how fast, large, and efficient it is. Furthermore, each component plays an integral role in the overall processing of the machine.

The world of computing is filled with jargon and acronyms. The speed of a processor is specified in GHz (gigahertz), the amount of memory is specified in MB (megabytes) and GB (gigabytes), and a display screen is specified in pixels.

The von Neumann architecture is the underlying architecture of most of today's computers. It has five main parts: memory, the arithmetic/logic (ALU) unit, input devices, output devices, and the control unit. The fetch–execute cycle, under the direction of the control unit, is the heart of the processing. In this cycle, instructions are fetched from memory, decoded, and executed.

RAM and ROM are acronyms for two types of computer memory. RAM stands for random-access memory; ROM stands for read-only memory. The values stored in RAM can be changed; those in ROM cannot.

Secondary storage devices are essential to a computer system. These devices save data when the computer is not running. Magnetic tape, magnetic disk, and flash drives are three common types of secondary storage.

Touch screens are peripheral devices that serve both input and output functions and are appropriate in specific situations such as restaurants and information kiosks. They respond to a human touching the screen with a finger or stylus, and can determine the location on the screen where the touch occurred. Several touch screen technologies exist, including resistive, capacitive, infrared, and surface acoustic wave (SAW) touch screens. They have varying characteristics that make them appropriate in particular situations.

Although von Neumann machines are by far the most common, other computer architectures have emerged. For example, there are machines with more than one processor so that calculations can be done in parallel, thereby speeding up the processing.

ETHICAL ISSUES ▸ Computer Hoaxes and Scams

As long as humans have known that other humans could be taken advantage of, there have been scammers, con artists, and hoaxers. The principal difference between a hoax and a scam is the financial purpose of the latter. The motives of a hoaxster are sometimes difficult to discern and may be as simple as the adolescent's impulse "to leave a mark" or "just for kicks." Hoaxes are annoying and time consuming. The ultimate motive of the con artist and scammer, however, is to trick the naive and unwary out of their money or possessions.

Before computers, these predators led difficult lives. They had to spend their own time and money to find individual victims. The amount taken from a victim could range from a few dollars—as would be the case in a typical "shell game"—to huge sums of money—as in "Brooklyn Bridge" schemes wherein victims thought they had purchased property that did not exist or that was not for sale. The perpetrators of these crimes were limited in the number of potential victims they could reach at any one time.

Then came the Internet. With a few clicks of a mouse, a scammer can now reach thousands of potential victims through email. The gathering of email addresses can be automated, which creates an enormous population of potential victims. Websites can act as virtual spider webs, entrapping those who innocently wander in.

There was a time when the most common complaint of Internet users was the annoyance of commercial spam. Today, good email services provide filters that catch most commercial spam before it reaches the individual. According to the Federal Trade Commission (FTC), the most common complaints of computer users are now the following: Internet auctions, Internet access services, credit card fraud, international model dialing, Web cramming, multilevel marketing plans/pyramids, travel and vacation scams, bogus business opportunities/investments, healthcare products and services, and phishing scams.

Most serious are those crimes that steal financial information and passwords from Web surfers. Websites may be used to lull people into believing that they are responding to surveys or providing credit card information merely to prove they are 18. By stealing passwords, criminals can gain access to their victims' entire financial records. Identity theft is devastating to the victims and can take years to recover from. Perhaps the greatest threat comes from those who really want to wreak havoc. Today, airlines, banks, and municipal infrastructures are all tied into computer networks. The damage a determined cybercriminal can cause is boundless.

The challenge of policing these schemes cannot be overstated. Perpetrators can disguise not only their identities, but also their geographical locations. For now, the best protection users have is skepticism. Refusing to give out credit card or other personal information to any request is mandatory. As computer use becomes even more widespread, chances are that the scammers, hoaxers, and con artists will keep pace. Until there is a viable way to stop their activities, surfers beware.

?

Scam email received by Nell Dale

IT Department Service,

You have exceeded the limit of your mailbox set by your IT Department service. And you will be having problems in sending and receiving new emails. To prevent this, you will have to contact the IT Department Service by email with your:

Current username:{ } and
Password:{ } to help increase your storage limit.

IT Department Service
E-mail:
<mailto:it.dept@administrativos.com>it.dept @administrativos.com

Failure to do this, will result in limited access to your mailbox.

Regards,
IT Department Service

Would you have answered? What would have happened if you did?

Key Terms

Access time
Addressability
Arithmetic/logic unit (ALU)
Block
Bus width
Cache memory
Control unit
CPU
Cylinder
Input unit
Instruction register (IR)
Latency
Motherboard
Output unit
Pipelining
Program counter (PC)
Register
Sector
Seek time
Shared memory parallel processor
Synchronous processing
Track
Transfer rate

Exercises

For Exercises 1–16, match the power of 10 to its name or use.

- **A.** 10^{-12}
- **B.** 10^{-9}
- **C.** 10^{-6}
- **D.** 10^{-3}
- **E.** 10^{3}
- **F.** 10^{6}
- **G.** 10^{9}
- **H.** 10^{12}
- **I.** 10^{15}

1. Nano
2. Pico
3. Micro
4. Milli
5. Tera
6. Giga
7. Kilo
8. Mega
9. Often used to describe processor speed
10. Often used to describe size of memory
11. Used in relation to Internet speeds
12. Latin for "thousandth"

13. Italian for "little"
14. Peta
15. Roughly equivalent to 2^{10}
16. Roughly equivalent to 2^{50}

For Exercises 17–23, match the acronym with its most accurate definition.

- A. CD-ROM
- B. CD-DA
- C. CD-R
- D. DVD
- E. CD-RW
- F. DL DVD
- G. Blu-Ray

17. Format using two layers
18. Data is stored in the sectors reserved for timing information in another variant
19. Can be read many times, but written after its manufacture only once
20. Can be both read from and written to any number of times
21. Format used in audio recordings
22. A new technology storing up to 50 GB
23. The most popular format for distributing movies

Exercises 24–66 are problems or short-answer exercises.

24. Define the following terms:
    - a. Core 2 processor
    - b. Hertz
    - c. Random access memory
25. What does FSB stand for?
26. What does it mean to say that a processor is 1.4 GHz?
27. What does it mean to say that memory is 133 MHz?
28. How many bytes of memory are there in the following machines?
    - a. 512MB machine
    - b. 2GB machine
29. Define RPM and discuss what it means in terms of speed of access to a disk.
30. What is the stored-program concept, and why is it important?
31. What does "units that process information are separate from the units that store information" mean in terms of computer architecture?

32. Name the components of a von Neumann machine.
33. What is the addressability of an 8-bit machine?
34. What is the function of the ALU?
35. Which component in the von Neumann architecture would you say acts as the stage manager? Explain.
36. Punched cards and paper tape were two early input/output media. Discuss their advantages and disadvantages.
37. What is an instruction register, and what is its function?
38. What is a program counter, and what is its function?
39. List the steps in the fetch–execute cycle.
40. Explain what is meant by "fetch an instruction."
41. Explain what is meant by "decode an instruction."
42. Explain what is meant by "execute an instruction."
43. Compare and contrast RAM and ROM.
44. What is a secondary storage device, and why are such devices important?
45. Discuss the pros and cons of using magnetic tape as a storage medium.
46. What are the four measures of a disk drive's efficiency?
47. Define what is meant by a block of data.
48. What is a cylinder?
49. Define the steps that a hard disk drive goes through to transfer a block of data from the disk to memory.
50. Distinguish between a compact disk and a magnetic disk.
51. Describe a parallel architecture that uses synchronous processing.
52. Describe a parallel architecture that uses pipeline processing.
53. How does a shared-memory parallel configuration work?
54. How many different memory locations can a 16-bit processor access?
55. Why is a faster clock not always better?
56. Why is a larger cache not necessarily better?
57. In the ad, why is the 1080p specification for the screen not entirely true?
58. Keep a diary for a week of how many times the terms *hardware* and *software* appear in television commercials.
59. Take a current ad for a laptop computer and compare that ad with the one shown at the beginning of this chapter.
60. What is the common name for the disk that is a secondary storage device?
61. To what does the expression *pixels* refer?

62. What is a GPU?
63. If a battery in a laptop is rated for 80 WHr, and the laptop draws 20 watts, how long will it run?
64. What is the difference between 1K of memory and a 1K transfer rate?
65. Compare and contrast a DVD-ROM and a flash drive.
66. "Giga" can mean both 10^9 and 2^{30}. Explain to which each refers. Can this cause confusion when reading a computer advertisement?

??? Thought Questions

1. Would octal or hexadecimal be a better way to refer to the addresses in a 16-bit processor? Justify your answer.
2. Relate the concept of a program to the fetch–execute cycle of the von Neumann machine.
3. Personal computers originally came equipped with one, then two floppy drives. After that, floppy drives became optional as CD drives became standard equipment. Now USB flash drives are the current medium for storage of data. What are the advantages of flash drives over other forms of disk storage? Do you think they will replace other media?
4. Why don't we just use powers of 10 when referring to storage? Aren't powers of 10 and powers of 2 close enough?
5. Walk through your kitchen and list the number of items that include embedded computers.
6. Have you ever been taken in by a hoax? Were you angry or just annoyed?
7. Have you or anyone you know been the victim of a scam artist?