HOW THE CPU WORKS

As already discussed, a CPU consists of a variety of circuitry and components packaged together. The key element of the processor is the transistor—a device made of semiconductor material that controls the flow of electrons inside a chip. Today's CPUs contain hundreds of millions of transistors, and the number doubles approximately every 18 months. This phenomenon is known as Moore's Law

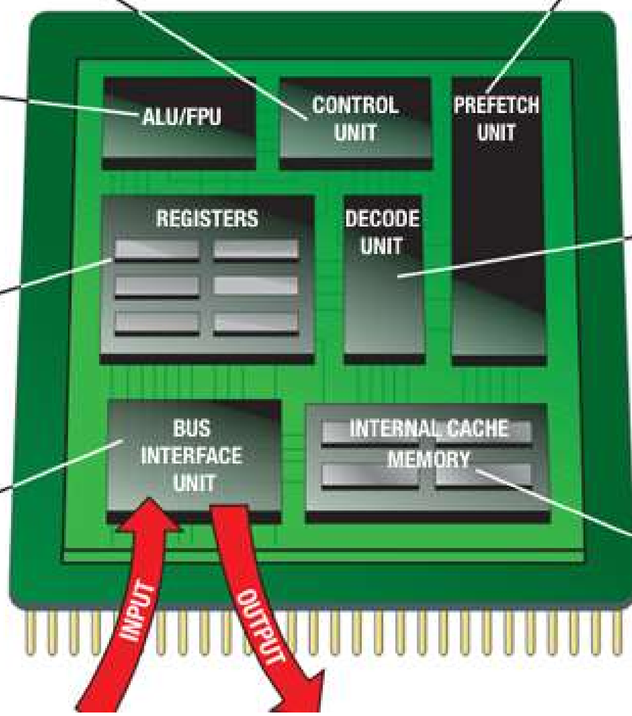
- Typical CPU components

To begin to understand how a CPU works, you need to know how the CPU is organized and what components it includes. This information will help you understand how electronic impulses move from one part of the CPU to another to process data. The architecture and components included in a CPU (referred to as microarchitecture) vary from processor to processor. A simplified example of the principal components that might be included in a single core of a typical CPU

CONTROL UNIT
Is in charge of the entire process, making sure everything happens at the right time. It instructs the ALU, FPU, and registers what to do, based on instructions from the decode unit.

PREFETCH UNIT
Requests instructions and data from cache or RAM and makes sure they are in the proper order for processing; it attempts to fetch instructions and data ahead of time so that the other components don't have to wait.

ARITHMETIC/LOGIC UNIT AND FLOATING POINT UNIT
Performs the arithmetic and logical operations, as directed by the control unit.



DECODE UNIT
Takes instructions from the prefetch unit and translates them into a form that the control unit can understand.

REGISTERS
Hold the results of processing.

BUS INTERFACE UNIT
The place where data and instructions enter or leave the core.

INTERNAL CACHE MEMORY
Stores data and instructions before and during processing.

- **Arithmetic/Logic unit (ALU) and Floating Point unit (FPU)**

The arithmetic/logic unit (ALU) is the section of a CPU core that performs arithmetic (addition, subtraction, multiplication, and division) involving integers and logical operations (such as comparing two pieces of data to see if they are equal or determining if a specific condition is true or false).

- **Control unit**

The control unit coordinates and controls the operations and activities taking place within a CPU core, such as retrieving data and instructions and passing them on to the ALU or FPU for execution. In other words, it directs the flow of electronic traffic within the core, much like a traffic cop controls the flow of vehicles on a roadway.

- **Prefetch unit**

The prefetch unit orders data and instructions from cache or RAM based on the cur rent task. The prefetch unit tries to predict what data and instructions will be needed and retrieves them ahead of time in order to help avoid delays in processing.

- **Decode unit**

The decode unit takes the instructions fetched by the prefetch unit and translates them into a form that can be understood by the control unit, ALU, and FPU. The decoded instructions go to the control unit for processing.

- **Registers and internal cache memory**

As mentioned earlier, registers and cache memory are both types of memory used by the CPU. Registers are groups of high-speed memory located within the CPU that are used during processing. The ALU and FPU use registers to store data, intermediary calculations, and the results of processing temporarily.

- **Bus interface unit**

The bus interface unit allows a core to communicate with other CPU components, such as the memory controller and other cores. As previously mentioned, the memory controller controls the flow of instructions and data going between the CPU cores and RAM.

- **The system clock and the machine cycle**

To synchronize all of a computer's operations, a system clock, which is located on the motherboard, is used. The system clock sends out a signal on a regular basis to all other computer components, similar to a musician's metronome or a person's heartbeat. Each signal is referred to as a cycle. The number of cycles per second is measured in hertz (Hz). One mega hertz (MHz) is equal to one million ticks of the system clock.

Machine cycle

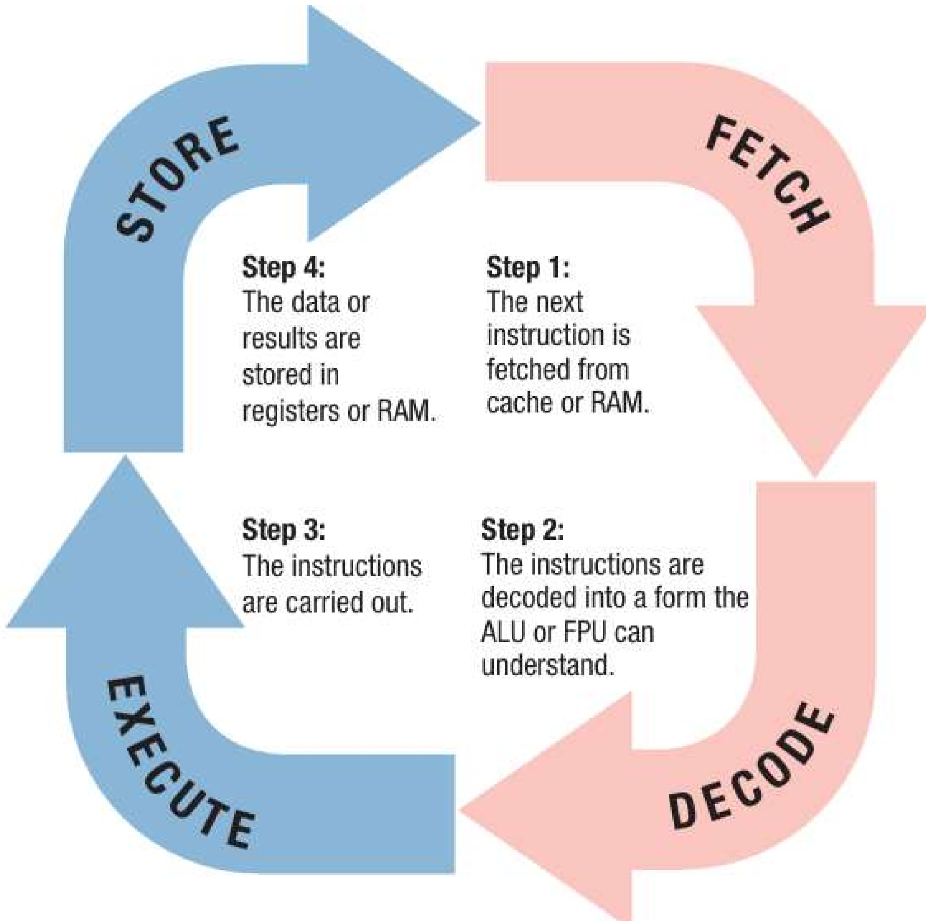
Whenever the CPU processes a single piece of micro code, it is referred to as a machine cycle. Each machine cycle consists of the four general operations illustrated in Figure and discussed next.

1. Fetch—the program instruction is fetched.

2. Decode—the instructions are decoded so the control unit, ALU, and FPU can understand them.

3. Execute—the instructions are carried out.

4. Store—the original data or the result from the ALU or FPU execution is stored in the CPU's registers.


STORE
FETCH
Step 4:
The data or results are stored in registers or RAM.
Step 1:
The next instruction is fetched from cache or RAM.
Step 3:
The instructions are carried out.
Step 2:
The instructions are decoded into a form the ALU or FPU can understand.
EXECUTE
DECODE

MAKING COMPUTERS FASTER AND BETTER NOW AND IN THE FUTURE

There are several possible remedies for a computer that is performing too slowly, including:

- adding more memory,
- performing system maintenance to clean up the computer's hard drive,
- buying a larger or additional hard drive, and
- upgrading the computer's Internet connection or video card, depending on the primary role of the computer and where the processing bottleneck appears to be.

To make computers work faster overall, computer designers have developed a number of strategies over the years, and researchers are continually working on new strategies. Some of the strategies already being implemented include improved architecture, pipelining, multiprocessing, parallel processing, and the use of new materials, such as graphene, and 3D chips.

One possibility for both current and future computers is **nanotechnology** research, which focuses on building computer components at the individual atomic and molecular levels. Some computer and consumer products (such as NRAM, solar cells, tennis rackets, and bikes) using carbon nanotubes or CNTs (tiny hollow tubes made of graphene) are currently on the market. Quantum computing and optical computers are other possibilities being researched, along with silicon photonics, terascale computing, and exascale computing

Storage

There are two parts to any storage system: the storage medium and the storage device.

A storage medium is the hardware where data is actually stored (for example, a DVD or a flash memory card); a storage medium is inserted into its corresponding storage device (such as a DVD drive or a flash memory card reader) in order to be read from or written to. Often the storage device and storage medium are two separate pieces of hardware (that is, the storage medium is removable), although with some systems—such as a hard drive or most USB flash drives—the two parts are permanently sealed together to form one piece of hardware.

Storage devices can be internal (located inside the system unit), external (plugged into an external port on the system unit), or remote (located on another computer, such as a net work server or Web server). Internal devices have the advantage of requiring no additional desk space and are usually faster than their external counterparts. External devices, however, can be eas ily transported from one location to another (such as to share data with others, to transfer data between a work computer and a home computer, or to take digital photos to a photo store). They can also be removed from the computer and stored in a secure area (such as for backup purposes or to protect sensitive data). Remote devices are accessed over a network. Some remote storage devices, such as those accessed via the Internet, have the additional advantage of being accessible from any computer with an Internet connection.

Volatility

Conventional RAM is volatile so programs and documents held in RAM are erased when they are no longer needed by the computer or when the power to the computer is turned off. Storage media, however, are nonvolatile, so the data remains on the media even when the power to the computer or storage device is off. Consequently, storage media are used for anything that needs to be saved for future use.

Random vs. Sequential access

When the computer receives an instruction that requires data located on a storage medium, it must go to the designated location on the appropriate storage medium and retrieve the requested data. This procedure is referred to as access. Two basic access methods are available: random and sequential.

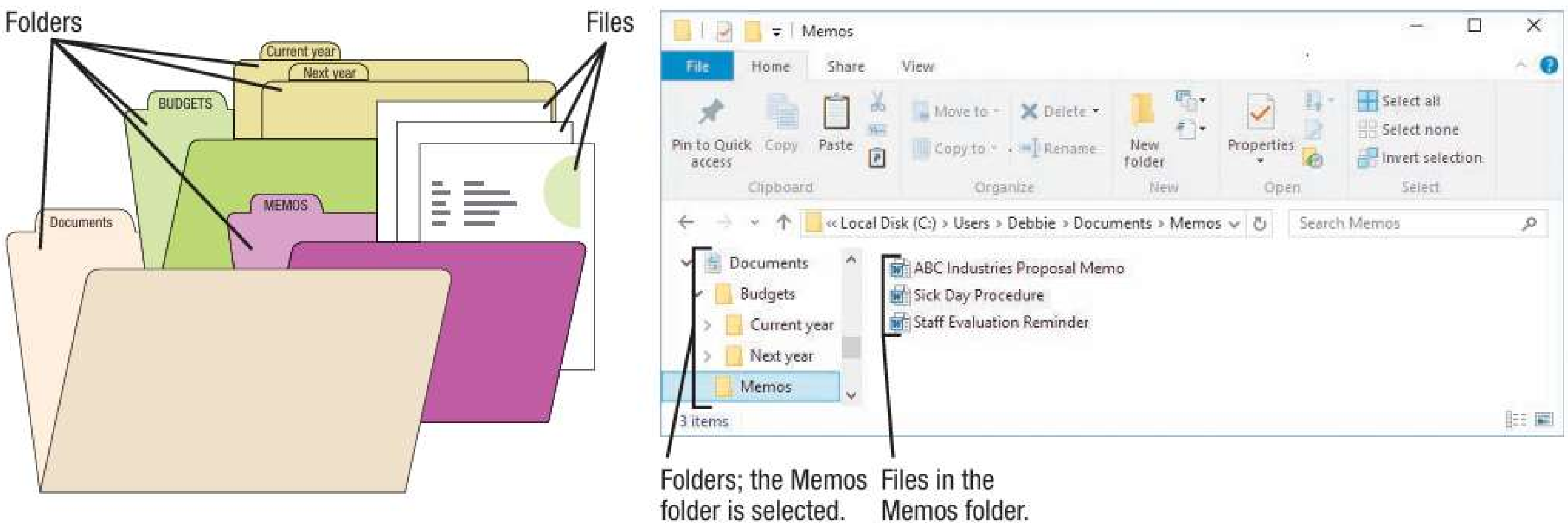
Random access

Also called direct access, means that data can be retrieved directly from any location on the storage medium, in any order. A random access device works in a manner similar to a DVD player used to play music or movies; that is, it can jump directly to a particular location on the medium when data located at that location is needed. Virtually all storage devices used with computers today for day-to-day storage—including hard drives, DVD drives, and USB flash drives—are random access devices. Media that allow random access are sometimes referred to as addressable media. This means that the storage system can locate each piece of stored data at a unique address, which is determined by the computer system.

Sequential access

The data can only be retrieved in the order in which it is physically stored on the medium. One type of storage device that is sometimes used with computers for backup purposes and that uses sequential access is a magnetic tape drive. Computer magnetic tapes work like audio cassette tapes or videotapes—to get to a specific location on the tape, you must play or fast forward through all of the tape that comes before the location you want to access.

on a storage medium.



>**File.** Something stored on a storage medium, such as a program, a document, or an image. >**Filename.** A name given to a file by the user; it is used to retrieve the file at a later time. >**Folder.** A named place on a storage medium into which the user can place files in order to keep the files stored on that medium organized.

TYPE OF STORAGE TECHNOLOGY USED

Data is stored magnetically or optically on many types of storage media.

MAGNETIC STORAGE SYSTEMS

With magnetic storage systems, such as conventional hard drives, data is stored magnetically on the storage medium, which means the data (0s and 1s) is represented using different magnetic alignments. The storage device can change the magnetic alignment when needed, so data can be written to the medium, deleted from the medium, or rewritten to the medium.

OPTICAL STORAGE MEDIA

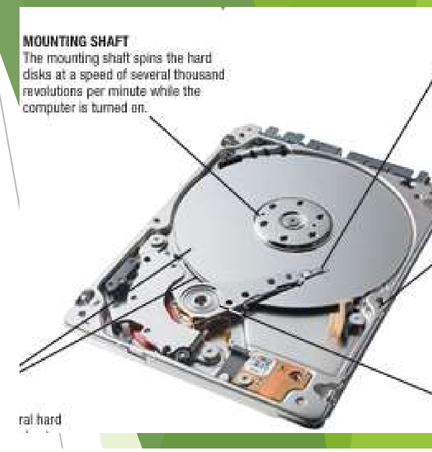
Optical storage media (such as DVDs) store data optically using laser beams. On some optical media, the laser burns permanent marks to represent 0s and 1s into the surface of the medium so the data cannot be erased or rewritten. With rewritable optical media, the laser changes the reflectivity of the medium to represent 0s and 1s but it does not permanently alter the disc surface so the reflectivity of the medium can be changed back again as needed. Consequently, the data stored on a rewritable optical disc can be changed.

ome storage systems use a combination of magnetic and optical technology. Others use a different technology altogether, such as flash memory storage systems that represent data using electrons inside flash memory cells to represent 0s and 1s.

Hard Drives

With the exception of computers designed to use only network storage devices (such as thin clients and some Internet appliances), virtually all personal computers come with a hard drive that is used to store most programs and data. Internal hard drives (those located inside the system unit) are not designed to be removed, unless they need to be repaired or replaced. External hard drives typically connect to a computer via an external port (such as a USB or Thunderbolt port) or a wireless connection and are frequently used for additional storage (such as for digital photos, videos, and other large multimedia files—particularly for users of tablets and other devices with limited internal storage.

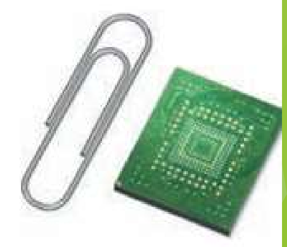
- Magnetic Hard Drives
- Solid-State Drives (SSDs)
- Solid-State Hybrid Drives (SSHDs)



FLASH MEMORY STORAGE SYSTEMS

Embedded Flash Memory

Embedded flash memory refers to flash memory chips embedded into products. Because flash memory media are physically very small, they are increasingly being embedded directly into a variety of consumer products—such as smartphones, tablets, smart watches, and even sunglasses and wristwatches—to provide built-in data storage. In fact, flash memory is usually the primary storage for mobile devices, such as tablets and smartphones.



EMBEDDED FLASH MEMORY CHIP

Flash memory Cards and readers

One of the most common types of flash memory media is the flash memory card—a small card containing one or more flash memory chips, a controller chip, other electrical components, and metal contacts to connect the card to the device or reader being used. Flash memory cards are available in a variety of formats, such as CompactFlash (CF), Secure Digital (SD), Multimedia Card (MMC), xD Picture Card (xD), XQD, and Memory Stick (MS) (see Figure 3-16). These formats are not interchangeable, so the type of flash memory card used with a device is determined by the type and size of flash media card that device can accept. Flash memory cards are the most common type of storage media for digital cameras, smartphones, and other portable devices. Flash memory cards can also be used to store data for a personal computer, as needed, as well as to transfer data from a portable device (digital camera or tablet, for instance) to a computer. Consequently, most personal computers and many mobile devices today have a built-in flash memory card reader capable of reading flash memory cards

FLASH MEMORY CARD READERS AND ADAPTERS

FLASH MEMORY CARDS