

Computer Basics

Computers manipulate digital data. Every sort of information a computer deals with — text, sound, graphics, to name a few — is represented within the computer as a stream of numbers.

Computers represent these numbers using the binary system. While modern human number systems are based on 10 (we have 10 digits, 0 through 9) computers use a base 2 system. In base 2, there are only 2 digits: 1 and 0. These binary digits are called bits. Everything a computer does involves moving, manipulating, calculating and storing numbers using the base 2 system. While computers may translate these base 2 numbers into other sorts of information that is easier for humans to work with, a computer's only true ability is to work with base 2 numbers.

Computers are built of a collection of components or subsystems, each performing a different task. Three critical ones are processing, short term storage and long term storage. Every modern computer, from desktop machines to smartphones, uses this same basic arrangement of functions.

Most processing is handled by the computer's CPU, its Central Processing Unit. The CPU can be likened to the engine of the computer. This is where the binary data is manipulated. A CPU can only perform a few basic arithmetic operations, such as adding, subtracting and shifting groups of numbers around. However, since modern computers can perform billions of these basic operations per second, large and complex tasks can be performed on computers by breaking the tasks down into billions of very simple arithmetical operations, the kind of operation a computer's CPU is designed to do.

Most computers also have specialized processors for manipulating graphics and other specialized data. These are called Graphic Processing Units (GPUs) and are important for specialized functions like 3D graphics and artificial intelligence features.

The timing of these operations is synchronized by a component of the CPU called the clock. The clock generates a metronome-like signal at regular intervals. The clock speed of modern computers is measured in billions of clock signals per second, (gigahertz, abbreviated GHz).

While the CPU manipulates data, the data also needs to be stored somewhere so that the CPU can access the raw data and store the results of its calculations. Most computer systems have two types of storage: long-term storage and short-term storage.

The CPU's ability to deal with billions of basic arithmetical operations per second means it needs to be fed data at a very high rate. The data that supplies the CPU's appetite is held in short-term storage. This takes the form of Random Access Memory chips, or RAM. The data the CPU needs to access is held in RAM. It is then moved to the processor along a path called a bus (think of a bus as a many-laned highway for binary digits).

The CPU makes the necessary calculation on a small number of bits, and the result of the calculation is shuttled along the bus back to the RAM memory, where it is held in storage. The cycle is then repeated.

RAM memory allows for very fast access to data, but it has two inherent disadvantages. One is that it is relatively expensive. A greater disadvantage is that most types of RAM memory can only store data as long as the memory chips are supplied with electrical power. Cut off the electricity, and the data vanishes instantly. These two drawbacks are the reasons we have a second sort of storage, long-term storage.

In most computer systems, long-term storage once took the form of hard disk drives (HDD), but now is usually a special kind of memory chip called a solid-state drive (SSD). Disk storage is much slower than RAM, but it is also much cheaper. SSDs are generally faster than disks, but still much slower than RAM. More importantly, and unlike RAM, information stored on a disk or SSD is preserved when the supply of electricity is shut off. While not fast enough to directly feed the CPU's ravenous appetite for data, long-term storage is used to store information not immediately needed by the computer's CPU, and to store all information when electrical power is shut off.

Computers generally move information from the long-term storage to RAM as the information is needed, shuttling it back and forth between RAM and CPU as the information is manipulated, and finally storing the information back in long-term storage (a hard drive or similar system) when the processed information needs to be saved. The rate at which data can be shuttled from long-term storage to RAM to CPU and back is often a more important factor when gauging the overall speed of a computer than the speed of the CPU alone. The computer's Graphic Processing Unit (GPU) makes a large contribution to the speed of many graphical operations, especially 3D graphics.

Other components, such as USB ports, and network interfaces like ethernet and wifi communicate with external devices such as displays and peripherals such as printers.

