Computing basics

The first computers were used primarily for numerical calculations. However, as any information can be numerically encoded, people soon realized that computers are capable of general-purpose [information processing](https://www.britannica.com/technology/information-processing). Their capacity to handle large amounts of data has extended the range and accuracy of [weather forecasting](https://www.britannica.com/science/weather-forecasting). Their speed has allowed them to make decisions about routing telephone connections through a network and to [control](https://www.britannica.com/technology/control-system) mechanical systems such as automobiles, nuclear reactors, and robotic surgical tools. They are also cheap enough to be [embedded](https://www.britannica.com/dictionary/embedded) in everyday appliances and to make [clothes](https://www.britannica.com/topic/dress-clothing) dryers and rice cookers “smart.” Computers have allowed us to pose and answer questions that could not be pursued before. These questions might be about [DNA](https://www.britannica.com/science/DNA) sequences in genes, patterns of activity in a consumer market, or all the uses of a word in texts that have been stored in a [database](https://www.britannica.com/technology/database). Increasingly, computers can also learn and adapt as they operate.

Computers also have limitations, some of which are theoretical. For example, there are undecidable propositions whose truth cannot be determined within a given set of rules, such as the logical structure of a computer. Because no universal [algorithmic](https://www.britannica.com/science/algorithm) method can exist to identify such propositions, a computer asked to obtain the truth of such a proposition will (unless forcibly interrupted) continue indefinitely—a condition known as the “[halting problem](https://www.britannica.com/topic/halting-problem).” (See [Turing machine](https://www.britannica.com/technology/Turing-machine).) Other limitations reflect current [technology](https://www.britannica.com/technology/technology). Human minds are skilled at recognizing spatial patterns—easily distinguishing among human faces, for instance—but this is a difficult task for computers, which must process information sequentially, rather than grasping details overall at a glance. Another problematic area for computers involves natural language interactions. Because so much common knowledge and contextual information is assumed in ordinary human communication, researchers have yet to solve the problem of providing relevant information to general-purpose natural language programs.

## [Analog computers](https://www.britannica.com/technology/analog-computer)

[Analog](https://www.merriam-webster.com/dictionary/Analog) computers use continuous physical magnitudes to represent quantitative information. At first they represented quantities with mechanical components (see [differential analyzer](https://www.britannica.com/technology/differential-analyzer) and [integrator](https://www.britannica.com/technology/integrator)), but after [World War II](https://www.britannica.com/event/World-War-II) voltages were used; by the 1960s digital computers had largely replaced them. Nonetheless, analog computers, and some hybrid digital-analog systems, continued in use through the 1960s in tasks such as aircraft and spaceflight simulation.

One advantage of analog computation is that it may be relatively simple to design and build an analog computer to solve a single problem. Another advantage is that analog computers can frequently represent and solve a problem in “real time”; that is, the computation proceeds at the same rate as the system being modeled by it. Their main disadvantages are that analog representations are limited in precision—typically a few decimal places but fewer in complex mechanisms—and general-purpose devices are expensive and not easily programmed.

## [Digital computers](https://www.britannica.com/technology/digital-computer)

In contrast to analog computers, digital computers represent information in [discrete](https://www.britannica.com/dictionary/discrete) form, generally as sequences of 0s and 1s (binary digits, or bits). The modern era of digital computers began in the late 1930s and early 1940s in the [United States](https://www.britannica.com/place/United-States), Britain, and [Germany](https://www.britannica.com/place/Germany). The first devices used switches operated by [electromagnets](https://www.britannica.com/science/electromagnet) (relays). Their programs were stored on punched paper tape or cards, and they had limited internal data storage. For historical developments, see the section [Invention of the modern computer](https://www.britannica.com/technology/computer/Early-business-machines#ref216032).

## [Mainframe computer](https://www.britannica.com/technology/mainframe)

During the 1950s and ’60s, Unisys (maker of the [UNIVAC](https://www.britannica.com/technology/UNIVAC) computer), [International Business Machines Corporation](https://www.britannica.com/topic/International-Business-Machines-Corporation) (IBM), and other companies made large, expensive computers of increasing [power](https://www.britannica.com/science/power-physics). They were used by major corporations and government research laboratories, typically as the sole computer in the organization. In 1959 the [IBM 1401](https://www.britannica.com/technology/IBM-1401) computer rented for $8,000 per month (early IBM machines were almost always leased rather than sold), and in 1964 the largest IBM S/360 computer cost several million dollars.

These computers came to be called mainframes, though the term did not become common until smaller computers were built. Mainframe computers were characterized by having (for their time) large storage capabilities, fast components, and powerful computational abilities. They were highly reliable, and, because they frequently served vital needs in an organization, they were sometimes designed with [redundant](https://www.merriam-webster.com/dictionary/redundant) components that let them survive partial failures. Because they were complex systems, they were operated by a staff of systems programmers, who alone had access to the computer. Other users submitted “batch jobs” to be run one at a time on the mainframe.

Such systems remain important today, though they are no longer the sole, or even primary, central computing resource of an organization, which will typically have hundreds or thousands of [personal computers](https://www.britannica.com/technology/personal-computer) (PCs). Mainframes now provide high-capacity data storage for [Internet](https://www.britannica.com/technology/Internet) servers, or, through [time-sharing](https://www.britannica.com/technology/time-sharing) techniques, they allow hundreds or thousands of users to run programs simultaneously. Because of their current roles, these computers are now called [servers](https://www.britannica.com/technology/server) rather than mainframes.

# [Supercomputer](https://www.britannica.com/technology/supercomputer)

The most powerful computers of the day have typically been called [supercomputers](https://www.britannica.com/technology/supercomputer). They have historically been very expensive and their use limited to high-priority computations for government-sponsored research, such as nuclear [simulations](https://www.britannica.com/dictionary/simulations) and weather modeling. Today many of the computational techniques of early supercomputers are in common use in PCs. On the other hand, the design of costly, special-purpose processors for supercomputers has been supplanted by the use of large arrays of commodity processors (from several dozen to over 8,000) operating in parallel over a high-speed communications network.

## [Minicomputer](https://www.britannica.com/technology/minicomputer)

Although [minicomputers](https://www.britannica.com/technology/minicomputer) date to the early 1950s, the term was introduced in the mid-1960s. Relatively small and inexpensive, minicomputers were typically used in a single department of an organization and often dedicated to one task or shared by a small group. Minicomputers generally had limited computational [power](https://www.britannica.com/science/power-physics), but they had excellent compatibility with various laboratory and industrial devices for collecting and inputting [data](https://www.britannica.com/dictionary/data).

One of the most important manufacturers of minicomputers was [Digital Equipment Corporation](https://www.britannica.com/topic/Digital-Equipment-Corporation) (DEC) with its [Programmed Data Processor](https://www.britannica.com/technology/Programmed-Data-Processor) (PDP). In 1960 DEC’s PDP-1 sold for $120,000. Five years later its PDP-8 cost $18,000 and became the first widely used minicomputer, with more than 50,000 sold. The DEC PDP-11, introduced in 1970, came in a variety of models, small and cheap enough to [control](https://www.britannica.com/technology/control-system) a single [manufacturing](https://www.britannica.com/technology/manufacturing) process and large enough for shared use in university computer centres; more than 650,000 were sold. However, the [microcomputer](https://www.britannica.com/technology/microcomputer) overtook this market in the 1980s.

[Microcomputer](https://www.britannica.com/technology/microcomputer)

[personal computer and peripherals](https://cdn.britannica.com/61/74061-050-47B62C0E/computer-peripherals-images-laser-printer-inkjet-layout.jpg)

A microcomputer is a small computer built around a [microprocessor](https://www.britannica.com/technology/microprocessor) [integrated circuit](https://www.britannica.com/technology/integrated-circuit), or [chip](https://www.britannica.com/technology/computer-chip). Whereas the early minicomputers replaced [vacuum tubes](https://www.britannica.com/technology/electron-tube) with discrete [transistors](https://www.britannica.com/technology/transistor), microcomputers (and later minicomputers as well) used microprocessors that [integrated](https://www.merriam-webster.com/dictionary/integrated) thousands or millions of transistors on a single chip. In 1971 the [Intel Corporation](https://www.britannica.com/topic/Intel) produced the first microprocessor, the [Intel 4004](https://www.britannica.com/technology/Intel-4004), which was powerful enough to [function](https://www.britannica.com/science/function-mathematics) as a computer although it was produced for use in a Japanese-made [calculator](https://www.britannica.com/technology/calculator). In 1975 the first

[personal computer](https://www.britannica.com/technology/personal-computer), the Altair, used a successor chip, the Intel 8080 microprocessor. Like minicomputers, early microcomputers had relatively limited storage and data-handling capabilities, but these have grown as storage [technology](https://www.britannica.com/technology/technology) has improved alongside processing power.

In the 1980s it was common to distinguish between microprocessor-based scientific [workstations](https://www.britannica.com/technology/workstation) and personal computers. The former used the most powerful microprocessors available and had high-performance colour graphics [capabilities](https://www.britannica.com/dictionary/capabilities) costing thousands of dollars. They were used by scientists for computation and data visualization and by engineers for [computer-aided engineering](https://www.britannica.com/technology/computer-aided-engineering). Today the distinction between [workstation](https://www.britannica.com/technology/workstation) and PC has virtually vanished, with PCs having the power and display capability of workstations.

## [Embedded processors](https://www.britannica.com/technology/embedded-processor)

Another class of computer is the embedded processor. These are small computers that use simple microprocessors to control electrical and mechanical functions. They generally do not have to do elaborate computations or be extremely fast, nor do they have to have great “input-output” capability, and so they can be inexpensive. Embedded processors help to control aircraft and industrial [automation](https://www.britannica.com/technology/automation), and they are common in automobiles and in both large and small household appliances. One particular type, the [digital signal processor](https://www.britannica.com/technology/digital-signal-processor) (DSP), has become as prevalent as the microprocessor. DSPs are used in wireless telephones, digital telephone and cable modems, and some stereo equipment.

## [Computer hardware](https://www.britannica.com/technology/hardware-computing)

The physical elements of a computer, its hardware, are generally divided into the central processing unit (CPU), main memory (or [random-access memory](https://www.britannica.com/technology/RAM-computing), RAM), and [peripherals](https://www.britannica.com/technology/input-output-device). The last class [encompasses](https://www.merriam-webster.com/dictionary/encompasses) all sorts of input and output (I/O) devices: keyboard, display monitor, printer, disk drives, network connections, scanners, and more.

The CPU and RAM are [integrated circuits](https://www.britannica.com/technology/integrated-circuit) (ICs)—small [silicon](https://www.britannica.com/science/silicon) wafers, or chips, that contain thousands or millions of transistors that function as electrical switches. In 1965 [Gordon Moore](https://www.britannica.com/biography/Gordon-Moore), one of the founders of Intel, stated what has become known as [Moore’s law](https://www.britannica.com/technology/Moores-law): the number of transistors on a chip doubles about every 18 months. Moore suggested that financial constraints would soon cause his law to break down, but it has been remarkably accurate for far longer than he first [envisioned](https://www.merriam-webster.com/dictionary/envisioned). It now appears that technical constraints may finally invalidate Moore’s law, since sometime between 2010 and 2020 transistors would have to consist of only a few atoms each, at which point the laws of [quantum](https://www.merriam-webster.com/dictionary/quantum) physics imply that they would cease to function reliably.

## [Central processing unit](https://www.britannica.com/technology/central-processing-unit)

The CPU provides the circuits that [implement](https://www.merriam-webster.com/dictionary/implement) the computer’s instruction set—its [machine language](https://www.britannica.com/technology/machine-language). It is composed of an [arithmetic-logic unit](https://www.britannica.com/technology/arithmetic-logic-unit) (ALU) and control circuits. The ALU carries out basic [arithmetic](https://www.britannica.com/science/arithmetic) and logic operations, and the control section determines the sequence of operations, including branch instructions that transfer control from one part of a program to another. Although the main memory was once considered part of the CPU, today it is regarded as separate. The boundaries shift, however, and CPU chips now also contain some high-speed [cache memory](https://www.britannica.com/technology/cache-memory) where data and instructions are temporarily stored for fast access.

The ALU has circuits that add, subtract, multiply, and divide two arithmetic [values](https://www.britannica.com/dictionary/values), as well as circuits for logic operations such as AND and OR (where a 1 is interpreted as true and a 0 as false, so that, for instance, 1 AND 0 = 0; see [Boolean algebra](https://www.britannica.com/topic/Boolean-algebra)). The ALU has several to more than a hundred registers that temporarily hold results of its computations for further arithmetic operations or for transfer to main memory.

The circuits in the CPU control section provide branch instructions, which make elementary decisions about what instruction to execute next. For example, a branch instruction might be “If the result of the last ALU operation is negative, jump to location A in the program; otherwise, continue with the following instruction.” Such instructions allow “if-then-else” decisions in a program and execution of a sequence of instructions, such as a “while-loop” that repeatedly does some set of instructions while some condition is met. A related instruction is the subroutine call, which transfers execution to a subprogram and then, after the subprogram finishes, returns to the main program where it left off.

In a [stored-program computer](https://www.britannica.com/technology/stored-program-concept), programs and data in memory are indistinguishable. Both are [bit](https://www.britannica.com/technology/bit-communications) patterns—strings of 0s and 1s—that may be interpreted either as data or as program instructions, and both are fetched from memory by the CPU. The CPU has a program counter that holds the memory address (location) of the next instruction to be executed. The basic operation of the CPU is the “fetch-decode-execute” cycle:

* Fetch the instruction from the address held in the program counter, and store it in a register.
* Decode the instruction. Parts of it specify the operation to be done, and parts specify the data on which it is to operate. These may be in CPU registers or in memory locations. If it is a branch instruction, part of it will contain the memory address of the next instruction to execute once the branch condition is satisfied.
* Fetch the operands, if any.
* Execute the operation if it is an ALU operation.
* Store the result (in a register or in memory), if there is one.
* Update the program counter to hold the next instruction location, which is either the next memory location or the address specified by a branch instruction.

At the end of these steps the cycle is ready to repeat, and it continues until a special halt instruction stops execution.

Steps of this cycle and all internal CPU operations are regulated by a [clock](https://www.britannica.com/technology/clock) that [oscillates](https://www.britannica.com/dictionary/oscillates) at a high frequency (now typically measured in gigahertz, or billions of cycles per second). Another factor that affects performance is the “word” size—the number of bits that are fetched at once from memory and on which CPU instructions operate. Digital words now consist of 32 or 64 bits, though sizes from 8 to 128 bits are seen.

Processing instructions one at a time, or serially, often creates a bottleneck because many program instructions may be ready and waiting for execution. Since the early 1980s, CPU design has followed a style originally called reduced-instruction-set computing ([RISC](https://www.britannica.com/technology/RISC)). This design minimizes the transfer of data between memory and CPU (all ALU operations are done only on data in CPU registers) and calls for simple instructions that can execute very quickly. As the number of transistors on a chip has grown, the RISC design requires a relatively small portion of the CPU chip to be devoted to the basic instruction set. The remainder of the chip can then be used to speed CPU operations by providing circuits that let several instructions execute simultaneously, or in parallel.

There are two major kinds of instruction-level parallelism (ILP) in the CPU, both first used in early supercomputers. One is the pipeline, which allows the fetch-decode-execute cycle to have several instructions under way at once. While one instruction is being executed, another can obtain its operands, a third can be [decoded](https://www.britannica.com/dictionary/decoded), and a fourth can be fetched from memory. If each of these operations requires the same time, a new instruction can enter the pipeline at each phase and (for example) five instructions can be completed in the time that it would take to complete one without a pipeline. The other sort of ILP is to have multiple execution units in the CPU—duplicate arithmetic circuits, in particular, as well as specialized circuits for graphics instructions or for [floating-point calculations](https://www.britannica.com/science/floating-point-calculation) (arithmetic operations involving noninteger numbers, such as 3.27). With this “superscalar” design, several instructions can execute at once.

Both forms of ILP face complications. A branch instruction might render preloaded instructions in the pipeline useless if they entered it before the branch jumped to a new part of the program. Also, superscalar execution must determine whether an arithmetic operation depends on the result of another operation, since they cannot be executed simultaneously. CPUs now have additional [circuits](https://www.britannica.com/dictionary/circuits) to predict whether a branch will be taken and to analyze instructional dependencies. These have become highly sophisticated and can frequently rearrange instructions to execute more of them in parallel.

References: https://www.britannica.com/technology/computer

Computer once meant a person who did computations, but now the term almost universally refers to automated electronic [machinery](https://www.britannica.com/technology/machine). The first section of this article focuses on modern digital electronic computers and their design, [constituent](https://www.merriam-webster.com/dictionary/constituent) parts, and applications. The second section covers the history of computing. For details on [computer architecture](https://www.britannica.com/technology/computer-architecture), [software](https://www.britannica.com/technology/software), and theory, see [computer science](https://www.britannica.com/science/computer-science).