Physics Chapter 2 Notes

Computers and computing devices from different eras—left to right, top to bottom:

- Early vacuum tube computer (ENIAC)
- Mainframe computer (IBM System/360)
- Smartphone (LYF Water 2)
- Desktop computer (IBM ThinkCentre S50 with monitor)
- Video game console (Nintendo GameCube)
- Supercomputer (IBM Summit)

A **computer** is a machine that can be programmed to automatically carry out sequences of arithmetic or logical operations (*computation*). Modern digital electronic computers can perform generic sets of operations known as *programs*, which enable computers to perform a wide range of tasks. The term **computer system** may refer to a nominally complete computer that includes the hardware, operating system, software, and peripheral equipment needed and used for full operation; or to a group of computers that are linked and function together, such as a computer network or computer cluster.

A broad range of industrial and consumer products use computers as control systems, including simple special-purpose devices like microwave ovens and remote controls, and factory devices like industrial robots. Computers are at the core of general-purpose devices such as personal computers and mobile devices such as smartphones. Computers power the Internet, which links billions of computers and users.

Early computers were meant to be used only for calculations. Simple manual instruments like the abacus have aided people in doing calculations since ancient times. Early in the Industrial Revolution, some mechanical devices were built to automate long, tedious tasks, such as guiding patterns for looms. More sophisticated electrical machines did specialized analog calculations in the early 20th century. The first digital electronic calculating machines were developed during World War II, both electromechanical and using thermionic valves. The first semiconductor transistors in the late 1940s were followed by the silicon-based MOSFET (MOS transistor) and monolithic integrated circuit chip technologies in the late 1950s, leading to the microprocessor and the microcomputer revolution in the 1970s. The speed, power, and versatility of computers have been increasing dramatically ever since then, with transistor counts increasing at a rapid pace (Moore's law noted that counts doubled every two years), leading to the Digital Revolution during the late 20th and early 21st centuries.

Conventionally, a modern computer consists of at least one processing element, typically a central processing unit (CPU) in the form of a microprocessor, together with some type of computer memory, typically semiconductor memory chips. The processing element carries out arithmetic and logical operations, and a sequencing and control unit can change the order of operations in response to stored information. Peripheral devices include input devices (keyboards, mice, joysticks, etc.), output devices (monitors, printers, etc.), and input/output devices that perform both functions (e.g. touchscreens). Peripheral devices allow information to be retrieved from an external source, and they enable the results of operations to be saved and retrieved.

Etymology

A human computer, with microscope and calculator, 1952

It was not until the mid-20th century that the word acquired its modern definition; according to the *Oxford English Dictionary*, the first known use of the word *computer* was in a different sense, in a 1613 book called *The Yong Mans Gleanings* by the English writer Richard Brathwait: "I haue [*sic*] read the truest computer of Times, and the best Arithmetician that euer [*sic*] breathed, and he reduceth thy dayes into a short number." This usage of the term referred to a human computer, a person who carried out calculations or computations. The word continued to have the same meaning until the middle of the 20th century. During the latter part of this period, women were often hired as computers because they could be paid less than their male counterparts.[1] By 1943, most human computers were women.[2]

The *Online Etymology Dictionary* gives the first attested use of *computer* in the 1640s, meaning 'one who calculates'; this is an "agent noun from compute (v.)". The *Online Etymology Dictionary* states that the use of the term to mean "'calculating machine' (of any type) is from 1897." The *Online Etymology Dictionary* indicates that the "modern use" of the term, to mean 'programmable digital electronic computer' dates from "1945 under this name; [in a] theoretical [sense] from 1937, as *Turing machine*".[3] The name has remained, although modern computers are capable of many higher-level functions.

History

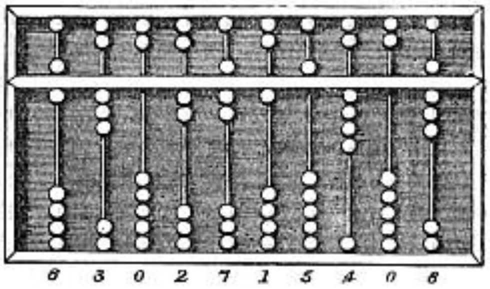
Main articles: History of computing and History of computing hardware

For a chronological guide, see Timeline of computing.

Pre-20th century

The Ishango bone, a bone tool dating back to prehistoric Africa

Devices have been used to aid computation for thousands of years, mostly using one-to-one correspondence with fingers. The earliest counting device was most likely a form of tally stick. Later record keeping aids throughout the Fertile Crescent included calculi (clay spheres, cones, etc.) which represented counts of items, likely livestock or grains, sealed in hollow unbaked clay containers.[a][4] The use of counting rods is one example.



The Chinese suanpan (算盘). The number represented on this abacus is 6,302,715,408.

The abacus was initially used for arithmetic tasks. The Roman abacus was developed from devices used in Babylonia as early as 2400 BCE. Since then, many other forms of reckoning boards or tables have been invented. In a medieval European counting house, a checkered cloth would be placed on a table, and markers moved around on it according to certain rules, as an aid to calculating sums of money.[5]

The Antikythera mechanism, dating back to ancient Greece circa 200–80 BCE, is an early analog computing device.

The Antikythera mechanism is believed to be the earliest known mechanical analog computer, according to Derek J. de Solla Price.[6] It was designed to calculate astronomical positions. It was discovered in 1901 in the Antikythera wreck off the Greek island of Antikythera, between Kythera and Crete, and has been dated to approximately c. 100 BCE. Devices of comparable complexity to the Antikythera mechanism would not reappear until the fourteenth century.[7]

Many mechanical aids to calculation and measurement were constructed for astronomical and navigation use. The planisphere was a star chart invented by Abū Rayhān al-Bīrūnī in the early 11th century.[8] The astrolabe was invented in the Hellenistic world in either the 1st or 2nd centuries BCE and is often attributed to Hipparchus. A combination of the planisphere and dioptra, the astrolabe was effectively an analog computer capable of working out several different kinds of problems in spherical astronomy. An astrolabe incorporating a mechanical calendar computer[9][10] and gear-wheels was invented by Abi Bakr of Isfahan, Persia in 1235.[11] Abū Rayhān al-Bīrūnī invented the first mechanical geared lunisolar calendar astrolabe,[12] an early fixed-wired knowledge processing machine[13] with a gear train and gear-wheels,[14] c. 1000 AD.

The sector, a calculating instrument used for solving problems in proportion, trigonometry, multiplication and division, and for various functions, such as squares and cube roots, was developed in the late 16th century and found application in gunnery, surveying and navigation.

The planimeter was a manual instrument to calculate the area of a closed figure by tracing over it with a mechanical linkage.

A slide rule

The slide rule was invented around 1620–1630, by the English clergyman William Oughtred, shortly after the publication of the concept of the logarithm. It is a hand-operated analog computer for doing multiplication and division. As slide rule development progressed, added scales provided reciprocals, squares and square roots, cubes and cube roots, as well as transcendental functions such as logarithms and exponentials, circular and hyperbolic trigonometry and other functions. Slide rules with special scales are still used for quick performance of routine calculations, such as the E6B circular slide rule used for time and distance calculations on light aircraft.

In the 1770s, Pierre Jaquet-Droz, a Swiss watchmaker, built a mechanical doll (automaton) that could write holding a quill pen. By switching the number and order of its internal wheels different letters, and hence different messages, could be produced. In effect, it could be mechanically "programmed" to read instructions.

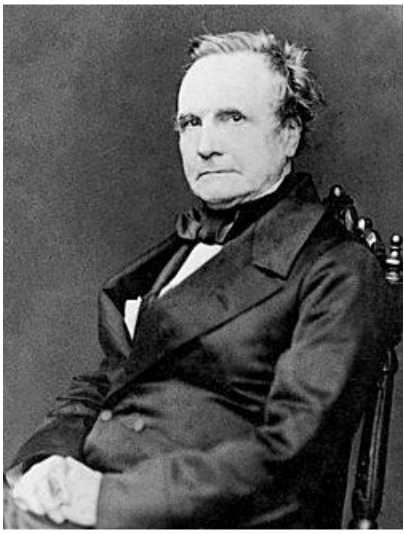
Along with two other complex machines, the doll is at the Musée d'Art et d'Histoire of Neuchâtel, Switzerland, and still operates.[15]

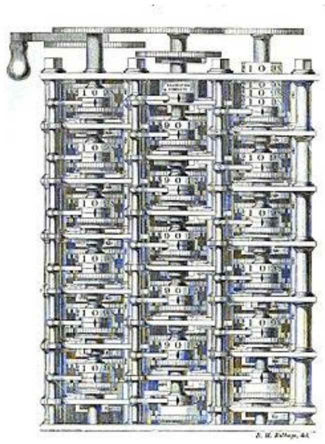
In 1831–1835, mathematician and engineer Giovanni Plana devised a Perpetual Calendar machine, which through a system of pulleys and cylinders could predict the perpetual calendar for every year from 0 CE (that is, 1 BCE) to 4000 CE, keeping track of leap years and varying day length. The tide-predicting machine invented by the Scottish scientist Sir William Thomson in 1872 was of great utility to navigation in shallow waters. It used a system of pulleys and wires to automatically calculate predicted tide levels for a set period at a particular location.

The differential analyser, a mechanical analog computer designed to solve differential equations by integration, used wheel-and-disc mechanisms to perform the integration. In 1876, Sir William Thomson had already discussed the possible construction of such calculators, but he had been stymied by the limited output torque of the ball-and-disk integrators.[16] In a differential analyzer, the output of one integrator drove the input of the next integrator, or a graphing output. The torque amplifier was the advance that allowed these machines to work. Starting in the 1920s, Vannevar Bush and others developed mechanical differential analyzers.

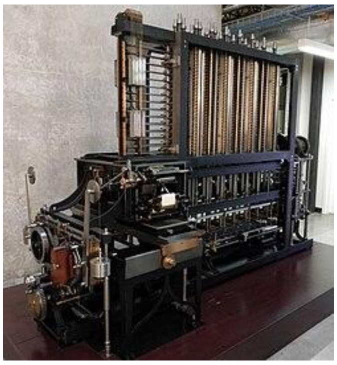
In the 1890s, the Spanish engineer Leonardo Torres Quevedo began to develop a series of advanced analog machines that could solve real and complex roots of polynomials,[17][18][19][20] which were published in 1901 by the Paris Academy of Sciences.[21]

First computer


Charles Babbage



A diagram of a portion of Babbage's Difference engine



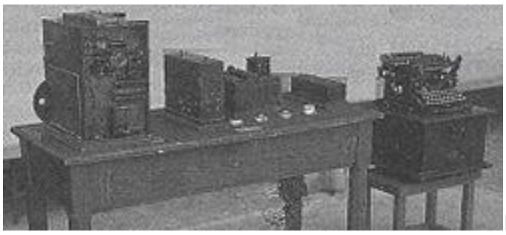
The Difference Engine Number 2 at the Intellectual Ventures laboratory in Seattle

Charles Babbage, an English mechanical engineer and polymath, originated the concept of a programmable computer. Considered the "father of the computer",[22] he conceptualized and invented the first mechanical computer in the early 19th century.

After working on his difference engine he announced his invention in 1822, in a paper to the Royal Astronomical Society, titled "Note on the application of machinery to the computation of astronomical and mathematical tables".[23] He also designed to aid in navigational calculations, in 1833 he realized that a much more general design, an analytical engine, was possible. The input of programs and data was to be provided to the machine via punched cards, a method being used at the time to direct mechanical looms such as the Jacquard loom. For output, the machine would have a printer, a curve plotter and a bell. The machine would also be able to punch numbers onto cards to be read in later. The engine would incorporate an arithmetic logic unit, control flow in the form of conditional branching and loops, and integrated memory, making it the first design for a general-purpose computer that could be described in modern terms as Turing-complete.[24][25]

The machine was about a century ahead of its time. All the parts for his machine had to be made by hand – this was a major problem for a device with thousands of parts. Eventually, the project was dissolved with the decision of the British Government to cease funding. Babbage's failure to complete the analytical engine can be chiefly attributed to political and financial difficulties as well as his desire to develop an increasingly sophisticated computer and to move ahead faster than anyone else could follow. Nevertheless, his son, Henry Babbage, completed a simplified version of the analytical engine's computing unit (the *mill*) in 1888. He gave a successful demonstration of its use in computing tables in 1906.

Electromechanical calculating machine



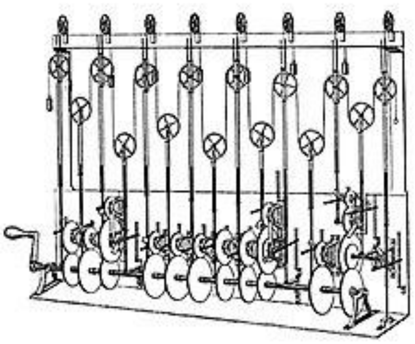
Electro-mechanical calculator (1920) by Leonardo Torres Quevedo.

In his work *Essays on Automatics* published in 1914, Leonardo Torres Quevedo wrote a brief history of Babbage's efforts at constructing a mechanical Difference Engine and Analytical Engine. The paper contains a design of a machine capable to calculate formulas like , for a sequence of sets of values. The whole machine was to be controlled by a read-only program, which was complete with provisions for conditional branching. He also introduced the idea of floating-point arithmetic.[26][27][28] In 1920, to celebrate the 100th anniversary of the invention of the arithmometer, Torres presented in Paris the Electromechanical

Arithmometer, which allowed a user to input arithmetic problems through a keyboard, and computed and printed the results,[29][30][31][32] demonstrating the feasibility of an electromechanical analytical engine.[33]

Analog computers

Main article: Analog computer



Sir William Thomson's third tide-predicting machine design, 1879–81

During the first half of the 20th century, many scientific computing needs were met by increasingly sophisticated analog computers, which used a direct mechanical or electrical model of the problem as a basis for computation. However, these were not programmable and generally lacked the versatility and accuracy of modern digital computers.[34] The first modern analog computer was a tide-predicting machine, invented by Sir William Thomson (later to become Lord Kelvin) in 1872. The differential analyser, a mechanical analog computer designed to solve differential equations by integration using wheel-and-disc mechanisms, was conceptualized in 1876 by James Thomson, the elder brother of the more famous Sir William Thomson.[16]

The art of mechanical analog computing reached its zenith with the differential analyzer, completed in 1931 by Vannevar Bush at MIT.[35] By the 1950s, the success of digital electronic computers had spelled the end for most analog computing machines, but analog computers remained in use during the 1950s in some specialized applications such as education (slide rule) and aircraft (control systems).[*citation needed*]

Digital computers

Electromechanical

Claude Shannon's 1937 master's thesis laid the foundations of digital computing, with his insight of applying Boolean algebra to the analysis and synthesis of switching circuits being the basic concept which underlies all electronic digital computers.[36][37]

By 1938, the United States Navy had developed the Torpedo Data Computer, an electromechanical analog computer for submarines that used trigonometry to solve the problem of firing a torpedo at a moving target. During World War II, similar devices were developed in other countries.[38]

Replica of Konrad Zuse's Z3, the first fully automatic, digital (electromechanical) computer

Early digital computers were electromechanical; electric switches drove mechanical relays to perform the calculation. These devices had a low operating speed and were eventually superseded by much faster all-electric computers, originally using vacuum tubes. The Z2, created by German engineer Konrad Zuse in 1939 in Berlin, was one of the earliest examples of an electromechanical relay computer.[39]

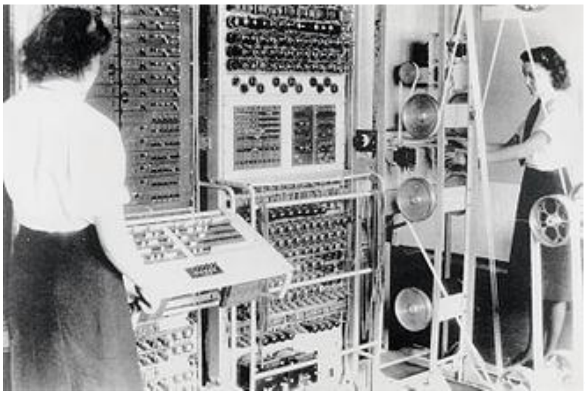
Konrad Zuse, inventor of the modern computer[40][41]

In 1941, Zuse followed his earlier machine up with the Z3, the world's first working electromechanical programmable, fully automatic digital computer.[42][43] The Z3 was built with 2000 relays, implementing a 22 bit word length that operated at a clock frequency of about 5–10 Hz.[44] Program code was supplied on punched film while data could be stored in 64 words of memory or supplied from the keyboard. It was quite similar to modern machines in some respects, pioneering numerous advances such as floating-point numbers. Rather than the harder-to-implement decimal system (used in Charles Babbage's earlier design), using a binary system meant that Zuse's machines were easier to build and potentially more reliable, given the technologies available at that time.[45] The Z3 was not itself a universal computer but could be extended to be Turing complete.[46][47]

Zuse's next computer, the Z4, became the world's first commercial computer; after initial delay due to the Second World War, it was completed in 1950 and delivered to the ETH Zurich.[48] The computer was manufactured by Zuse's own company, Zuse KG, which was founded in 1941 as the first company with the sole purpose of developing computers in Berlin.[48] The Z4 served as the inspiration for the construction of the ERMETH, the first Swiss computer and one of the first in Europe.[49]

Vacuum tubes and digital electronic circuits

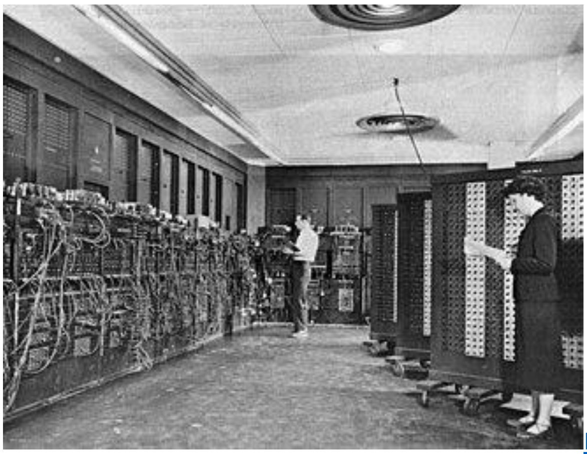
Purely electronic circuit elements soon replaced their mechanical and electromechanical equivalents, at the same time that digital calculation replaced analog. The engineer Tommy Flowers, working at the Post Office Research Station in London in the 1930s, began to explore the possible use of electronics for the telephone exchange. Experimental equipment that he built in 1934 went into operation five years later, converting a portion of the telephone exchange network into an electronic data processing system, using thousands of vacuum tubes.[34] In the US, John Vincent Atanasoff and Clifford E. Berry of Iowa State University developed and tested the Atanasoff–Berry Computer (ABC) in 1942,[50] the first "automatic electronic digital computer".[51] This design was also all-electronic and used about 300 vacuum tubes, with capacitors fixed in a mechanically rotating drum for memory.[52]



Colossus, the first electronic digital programmable computing device, was used to break German ciphers during World War II. It is seen here in use at Bletchley Park in 1943.

During World War II, the British code-breakers at Bletchley Park achieved a number of successes at breaking encrypted German military communications. The German encryption machine, Enigma, was first attacked with the help of the electro-mechanical bombes which were often run by women.[53][54] To crack the more sophisticated German Lorenz SZ 40/42 machine, used for high-level Army communications, Max Newman and his colleagues commissioned Flowers to build the Colossus.[52] He spent eleven months from early February 1943 designing and building the first Colossus.[55] After a functional test in December 1943, Colossus was shipped to Bletchley Park, where it was delivered on 18 January 1944[56] and attacked its first message on 5 February.[52]

Colossus was the world's first electronic digital programmable computer.[34] It used a large number of valves (vacuum tubes). It had paper-tape input and was capable of being configured to perform a variety of boolean logical operations on its data, but it was not Turing-complete. Nine Mk II Colossi were built (The Mk I was converted to a Mk II making ten machines in total). Colossus Mark I contained 1,500 thermionic valves (tubes), but Mark II with 2,400 valves, was both five times faster and simpler to operate than Mark I, greatly speeding the decoding process.[57][58]



ENIAC was the first electronic, Turing-complete device, and performed ballistics trajectory calculations for the United States Army.

The ENIAC[59] (Electronic Numerical Integrator and Computer) was the first electronic programmable computer built in the U.S. Although the ENIAC was similar to the Colossus, it was much faster, more flexible, and it was Turing-complete. Like the Colossus, a "program" on the ENIAC was defined by the states of its patch cables and switches, a far cry from the stored program electronic machines that came later. Once a program was written, it had to be mechanically set into the machine with manual resetting of plugs and switches. The programmers of the ENIAC were six women, often known collectively as the "ENIAC girls".[60][61]

It combined the high speed of electronics with the ability to be programmed for many complex problems. It could add or subtract 5000 times a second, a thousand times faster than any other machine. It also had modules to multiply, divide, and square root. High speed memory was limited to 20 words (about 80 bytes). Built under the direction of John Mauchly and J. Presper Eckert at the University of Pennsylvania, ENIAC's development and construction lasted from 1943 to full operation at the end of 1945. The machine was huge, weighing 30 tons, using 200 kilowatts of electric power and contained over 18,000 vacuum tubes, 1,500 relays, and hundreds of thousands of resistors, capacitors, and inductors.[62]

Modern computers

Concept of modern computer

The principle of the modern computer was proposed by Alan Turing in his seminal 1936 paper,[63] *On Computable Numbers*. Turing proposed a simple device that he called "Universal Computing machine" and that is now known as a universal Turing machine. He proved that such a machine is capable of computing anything that is computable by executing instructions (program) stored on tape, allowing the machine to be programmable. The fundamental concept of Turing's design is the stored program, where all the instructions for computing are stored in memory. Von Neumann acknowledged that the central concept of the modern computer was due to this paper.[64] Turing machines are to this day a central object of study in theory of computation. Except for the limitations imposed by their finite memory stores, modern computers are said to be Turing-complete, which is to say, they have algorithm execution capability equivalent to a universal Turing machine.

Stored programs

Main article: Stored-program computer

A section of the reconstructed Manchester Baby, the first electronic stored-program computer

Early computing machines had fixed programs. Changing its function required the re-wiring and re-structuring of the machine.[52] With the proposal of the stored-program computer this changed. A stored-program computer includes by design an instruction set and can store in memory a set of instructions (a program) that details the computation. The theoretical basis for the stored-program computer was laid out by Alan Turing in his 1936 paper. In 1945, Turing joined the National Physical Laboratory and began work on developing an electronic stored-program digital computer. His 1945 report "Proposed Electronic Calculator" was the first specification for such a device. John von Neumann at the University of Pennsylvania also circulated his *First Draft of a Report on the EDVAC* in 1945.[34]

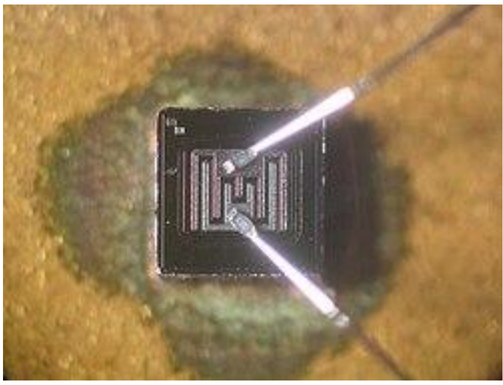
The Manchester Baby was the world's first stored-program computer. It was built at the University of Manchester in England by Frederic C. Williams, Tom Kilburn and Geoff Tootill, and ran its first program on 21 June 1948.[65] It was designed as a testbed for the Williams tube, the first random-access digital storage device.[66] Although the computer was described as "small and primitive" by a 1998 retrospective, it was the first working machine to contain all of the elements essential to a modern electronic computer.[67] As soon as the Baby had demonstrated the feasibility of its design, a project began at the university to develop it into a practically useful computer, the Manchester Mark 1.

The Mark 1 in turn quickly became the prototype for the Ferranti Mark 1, the world's first commercially available general-purpose computer.[68] Built by Ferranti, it was delivered to the University of Manchester in February 1951. At least seven of these later machines were delivered between 1953 and 1957, one of them to Shell labs in Amsterdam.[69] In October 1947 the directors of British catering company J. Lyons & Company decided to take an active role in promoting the commercial development of computers. Lyons's LEO I computer, modelled closely on the Cambridge EDSAC of 1949, became operational in April 1951[70] and ran the world's first routine office computer job.

Transistors

Main articles: Transistor and History of the transistor

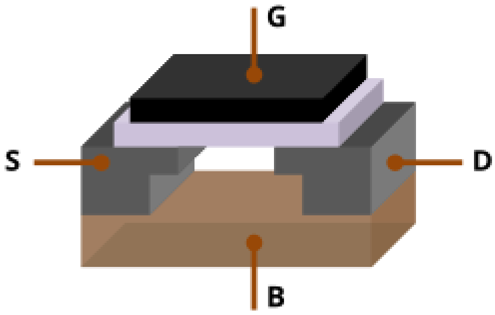
Further information: Transistor computer and MOSFET



Bipolar junction transistor (BJT)

The concept of a field-effect transistor was proposed by Julius Edgar Lilienfeld in 1925. John Bardeen and Walter Brattain, while working under William Shockley at Bell Labs, built the first working transistor, the point-contact transistor, in 1947, which was followed by Shockley's bipolar junction transistor in 1948.[71][72] From 1955 onwards, transistors replaced vacuum tubes in computer designs, giving rise to the "second generation" of computers. Compared to vacuum tubes, transistors have many advantages: they are smaller, and require less power than vacuum tubes, so give off less heat. Junction transistors were much more reliable than vacuum tubes and had longer, indefinite, service life. Transistorized computers could contain tens of thousands of binary logic circuits in a relatively compact space. However, early junction transistors were relatively bulky devices that were difficult to manufacture on a mass-production basis, which limited them to a number of specialized applications.[73]

At the University of Manchester, a team under the leadership of Tom Kilburn designed and built a machine using the newly developed transistors instead of valves.[74] Their first transistorized computer and the first in the world, was operational by 1953, and a second version was completed there in April 1955. However, the machine did make use of valves to generate its 125 kHz clock waveforms and in the circuitry to read and write on its magnetic drum memory, so it was not the first completely transistorized computer. That distinction goes to the Harwell CADET of 1955,[75] built by the electronics division of the Atomic Energy Research Establishment at Harwell.[75][76]



MOSFET (MOS transistor), showing gate (G), body (B), source (S) and drain (D) terminals. The gate is separated from the body by an insulating layer (pink).

The metal–oxide–silicon field-effect transistor (MOSFET), also known as the MOS transistor, was invented at Bell Labs between 1955 and 1960[77][78][79][80][81][82] and was the first truly compact transistor that could be miniaturized and mass-produced for a wide range of uses.[73] With its high scalability,[83] and much lower power consumption and higher density than bipolar junction transistors,[84] the MOSFET made it possible to build high-density integrated circuits.[85][86] In addition to data processing, it also enabled the practical use of MOS transistors as memory cell storage elements, leading to the development of MOS semiconductor memory, which replaced earlier magnetic-core memory in computers. The MOSFET led to the microcomputer revolution,[87] and became the driving force behind the computer revolution.[88][89] The MOSFET is the most widely used transistor in computers,[90][91] and is the fundamental building block of digital electronics.[92]

Integrated circuits

Main articles: Integrated circuit and Invention of the integrated circuit

Further information: Planar process and Microprocessor

Integrated circuits are typically packaged in plastic, metal, or ceramic cases to protect the IC from damage and for ease of assembly.

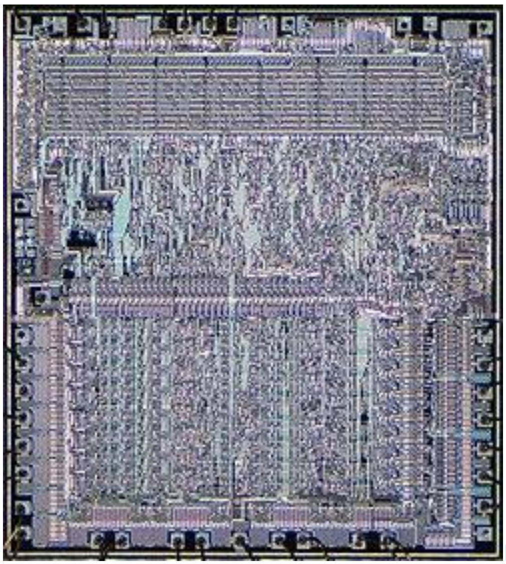
The next great advance in computing power came with the advent of the integrated circuit (IC). The idea of the integrated circuit was first conceived by a radar scientist working for the Royal Radar Establishment of the Ministry of Defence, Geoffrey W.A. Dummer. Dummer presented the first public description of an integrated circuit at the Symposium on Progress in Quality Electronic Components in Washington, D.C., on 7 May 1952.[93]

The first working ICs were invented by Jack Kilby at Texas Instruments and Robert Noyce at Fairchild Semiconductor.[94] Kilby recorded his initial ideas concerning the integrated circuit in July 1958, successfully demonstrating the first working integrated example on 12 September 1958.[95] In his patent application of 6 February 1959, Kilby described his new device as "a body of semiconductor material ... wherein all the components of the electronic circuit are completely integrated".[96][97] However, Kilby's invention was a hybrid integrated circuit (hybrid IC), rather than a monolithic integrated circuit (IC) chip.[98] Kilby's IC had external wire connections, which made it difficult to mass-produce.[99]

Noyce also came up with his own idea of an integrated circuit half a year later than Kilby.[100] Noyce's invention was the first true monolithic IC chip.[101][99] His chip solved many practical problems that Kilby's had not.

Produced at Fairchild Semiconductor, it was made of silicon, whereas Kilby's chip was made of germanium. Noyce's monolithic IC was fabricated using the planar process, developed by his colleague Jean Hoerni in early 1959. In turn, the planar process was based on Carl Frosch and Lincoln Derick work on semiconductor surface passivation by silicon dioxide.[102][103][104][105][106][107]

Modern monolithic ICs are predominantly MOS (metal–oxide–semiconductor) integrated circuits, built from MOSFETs (MOS transistors).[108] The earliest experimental MOS IC to be fabricated was a 16-transistor chip built by Fred Heiman and Steven Hofstein at RCA in 1962.[109] General Microelectronics later introduced the first commercial MOS IC in 1964,[110] developed by Robert Norman.[109] Following the development of the self-aligned gate (silicon-gate) MOS transistor by Robert Kerwin, Donald Klein and John Sarace at Bell Labs in 1967, the first silicon-gate MOS IC with self-aligned gates was developed by Federico Faggin at Fairchild Semiconductor in 1968.[111] The MOSFET has since become the most critical device component in modern ICs.[108]


Die photograph of a MOS 6502, an early 1970s microprocessor integrating 3500 transistors on a single chip

The development of the MOS integrated circuit led to the invention of the microprocessor,[112][113] and heralded an explosion in the commercial and personal use of computers. While the subject of exactly which device was the first microprocessor is contentious, partly due to lack of agreement on the exact definition of the term "microprocessor", it is largely undisputed that the first single-chip microprocessor was the Intel 4004,[114] designed and realized by Federico Faggin with his silicon-gate MOS IC technology,[112] along with Ted Hoff, Masatoshi Shima and Stanley Mazor at Intel.[b][116] In the early 1970s, MOS IC technology enabled the integration of more than 10,000 transistors on a single chip.[86]

System on a Chip (SoCs) are complete computers on a microchip (or chip) the size of a coin.[117] They may or may not have integrated RAM and flash memory. If not integrated, the RAM is usually placed directly above (known as Package on package) or below (on the opposite side of the circuit board) the SoC, and the flash memory is usually placed right next to the SoC. This is done to improve data transfer speeds, as the data signals do not have to travel long distances. Since ENIAC in 1945, computers have advanced enormously, with modern SoCs (such as the Snapdragon 865) being the size of a coin while also being hundreds of thousands of times more powerful than ENIAC, integrating billions of transistors, and consuming only a few watts of power.

Mobile computers

The first mobile computers were heavy and ran from mains power. The 50 lb (23 kg) IBM 5100 was an early example. Later portables such as the Osborne 1 and Compaq Portable were considerably lighter but still needed to be plugged in. The first laptops, such as the Grid Compass, removed this requirement by incorporating batteries – and with the continued miniaturization of computing resources and advancements in portable battery life, portable computers grew in popularity in the 2000s.[118] The same developments allowed manufacturers to integrate computing resources into cellular mobile phones by the early 2000s.

These smartphones and tablets run on a variety of operating systems and recently became the dominant computing device on the market.[119] These are powered by System on a Chip (SoCs), which are complete computers on a microchip the size of a coin.[117]

Types

See also: Classes of computers

Computers can be classified in a number of different ways, including:

By architecture

- Analog computer
- Digital computer
- Hybrid computer
- Harvard architecture
- Von Neumann architecture
- Complex instruction set computer
- Reduced instruction set computer

By size, form-factor and purpose

See also: List of computer size categories

- Supercomputer
- Mainframe computer
- Minicomputer (term no longer used),[120] Midrange computer
- Server
  - Rackmount server
  - Blade server
  - Tower server
- Personal computer
  - Workstation
  - Microcomputer (term no longer used)[121]
    - Home computer (term fallen into disuse)[122]
  - Desktop computer
    - Tower desktop
    - Slimline desktop
      - Multimedia computer (non-linear editing system computers, video editing PCs and the like, this term is no longer used)[123]

            - Gaming computer
        - All-in-one PC
        - Nettop (Small form factor PCs, Mini PCs)
        - Home theater PC
        - Keyboard computer
        - Portable computer
        - Thin client
        - Internet appliance
    - Laptop computer
        - Desktop replacement computer
        - Gaming laptop
        - Rugged laptop
        - 2-in-1 PC
        - Ultrabook
        - Chromebook
        - Subnotebook
        - Smartbook
        - Netbook
    - Mobile computer
        - Tablet computer
        - Smartphone
        - Ultra-mobile PC
        - Pocket PC
        - Palmtop PC
        - Handheld PC
        - Pocket computer
    - Wearable computer
        - Smartwatch
        - Smartglasses
- Single-board computer
- Plug computer
- Stick PC

- Programmable logic controller
- Computer-on-module
- System on module
- System in a package
- System-on-chip (Also known as an Application Processor or AP if it lacks circuitry such as radio circuitry)
- Microcontroller

Unconventional computers

Main article: Human computer

See also: Harvard Computers

A computer does not need to be electronic, nor even have a processor, nor RAM, nor even a hard disk. While popular usage of the word "computer" is synonymous with a personal electronic computer,[c] a typical modern definition of a computer is: "*A device that computes*, especially a programmable [usually] electronic machine that performs high-speed mathematical or logical operations or that assembles, stores, correlates, or otherwise processes information."[124] According to this definition, any device that *processes information* qualifies as a computer.

Hardware

Main articles: Computer hardware, Personal computer hardware, Central processing unit, and Microprocessor

Duration: 1 minute and 55 seconds.1:55 Video demonstrating the standard components of a "slimline" computer

The term *hardware* covers all of those parts of a computer that are tangible physical objects. Circuits, computer chips, graphic cards, sound cards, memory (RAM), motherboard, displays, power supplies, cables, keyboards, printers and "mice" input devices are all hardware.

History of computing hardware

Main article: History of computing hardware

First generation (mechanical/electromechanical)	Calculators	Pascal's calculator, Arithmometer, Difference engine, Quevedo's analytical machines
	Programmable devices	Jacquard loom, Analytical engine, IBM ASCC/Harvard Mark I, Harvard Mark II, IBM SSEC, Z1, Z2, Z3
Second generation (vacuum tubes)	Calculators	Atanasoff–Berry Computer, IBM 604, UNIVAC 60, UNIVAC 120
	Programmable devices	Colossus, ENIAC, Manchester Baby, EDSAC, Manchester Mark 1, Ferranti Pegasus, Ferranti Mercury, CSIRAC, EDVAC, UNIVAC I, IBM 701, IBM 702, IBM 650, Z22

Third generation (discrete transistors and SSI, MSI, LSI integrated circuits)	Mainframes	IBM 7090, IBM 7080, IBM System/360, BUNCH
	Minicomputer	HP 2116A, IBM System/32, IBM System/36, LINC, PDP-8, PDP-11
	Desktop Computer	HP 9100
Fourth generation (VLSI integrated circuits)	Minicomputer	VAX, IBM AS/400
	4-bit microcomputer	Intel 4004, Intel 4040
	8-bit microcomputer	Intel 8008, Intel 8080, Motorola 6800, Motorola 6809, MOS Technology 6502, Zilog Z80
	16-bit microcomputer	Intel 8088, Zilog Z8000, WDC 65816/65802
	32-bit microcomputer	Intel 80386, Pentium, Motorola 68000, ARM
	64-bit microcomputer[d]	Alpha, MIPS, PA-RISC, PowerPC, SPARC, x86-64, ARMv8-A
	Embedded computer	Intel 8048, Intel 8051
	Personal computer	Desktop computer, Home computer, Laptop computer, Personal digital assistant (PDA), Portable computer, Tablet PC, Wearable computer
Theoretical/experimental	Quantum computer	IBM Q System One
	Chemical computer	
	DNA computing	
	Optical computer	
	Spintronics-based computer	
	Wetware/Organic computer	

Other hardware topics

Peripheral device (input/output)	Input	Mouse, keyboard, joystick, image scanner, webcam, graphics tablet, microphone
	Output	Monitor, printer, loudspeaker

	Both	Floppy disk drive, hard disk drive, optical disc drive, teleprinter
Computer buses	Short range	RS-232, SCSI, PCI, USB
	Long range (computer networking)	Ethernet, ATM, FDDI

A general-purpose computer has four main components: the arithmetic logic unit (ALU), the control unit, the memory, and the input and output devices (collectively termed I/O). These parts are interconnected by buses, often made of groups of wires. Inside each of these parts are thousands to trillions of small electrical circuits which can be turned off or on by means of an electronic switch. Each circuit represents a bit (binary digit) of information so that when the circuit is on it represents a "1", and when off it represents a "0" (in positive logic representation). The circuits are arranged in logic gates so that one or more of the circuits may control the state of one or more of the other circuits.

Input devices

Input devices are the means by which the operations of a computer are controlled and it is provided with data. Examples include:

- Computer keyboard
- Digital camera
- Graphics tablet
- Image scanner
- Joystick
- Microphone
- Mouse
- Overlay keyboard
- Real-time clock
- Trackball
- Touchscreen
- Light pen

Output devices

Output devices are the means by which a computer provides the results of its calculations in a human-accessible form. Examples include:

- Computer monitor
- Printer
- PC speaker
- Projector
- Sound card

- Graphics card

Control unit

Main articles: CPU design and Control unit

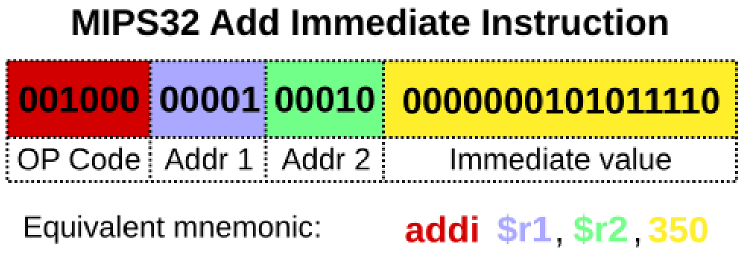


Diagram showing how a particular MIPS architecture instruction would be decoded by the control system

The control unit (often called a control system or central controller) manages the computer's various components; it reads and interprets (decodes) the program instructions, transforming them into control signals that activate other parts of the computer.[e] Control systems in advanced computers may change the order of execution of some instructions to improve performance.

A key component common to all CPUs is the program counter, a special memory cell (a register) that keeps track of which location in memory the next instruction is to be read from.[f]

The control system's function is as follows— this is a simplified description, and some of these steps may be performed concurrently or in a different order depending on the type of CPU:

1. Read the code for the next instruction from the cell indicated by the program counter.
2. Decode the numerical code for the instruction into a set of commands or signals for each of the other systems.
3. Increment the program counter so it points to the next instruction.
4. Read whatever data the instruction requires from cells in memory (or perhaps from an input device). The location of this required data is typically stored within the instruction code.
5. Provide the necessary data to an ALU or register.
6. If the instruction requires an ALU or specialized hardware to complete, instruct the hardware to perform the requested operation.
7. Write the result from the ALU back to a memory location or to a register or perhaps an output device.
8. Jump back to step (1).

Since the program counter is (conceptually) just another set of memory cells, it can be changed by calculations done in the ALU. Adding 100 to the program counter would cause the next instruction to be read from a place 100 locations further down the program. Instructions that modify the program counter are often known as "jumps" and allow for loops (instructions that are repeated by the computer) and often conditional instruction execution (both examples of control flow).

The sequence of operations that the control unit goes through to process an instruction is in itself like a short computer program, and indeed, in some more complex CPU designs, there is another yet smaller computer called a microsequencer, which runs a microcode program that causes all of these events to happen.

Central processing unit (CPU)

Main articles: Central processing unit and Microprocessor

The control unit, ALU, and registers are collectively known as a central processing unit (CPU). Early CPUs were composed of many separate components. Since the 1970s, CPUs have typically been constructed on a single MOS integrated circuit chip called a *microprocessor*.

Arithmetic logic unit (ALU)

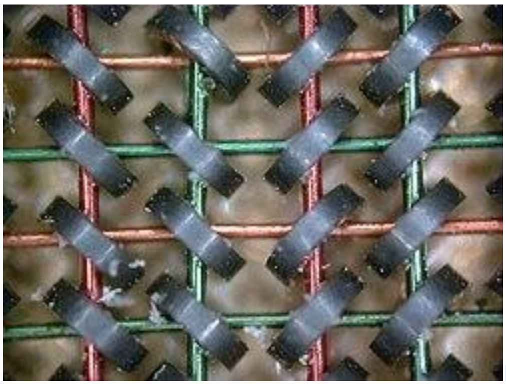
Main article: Arithmetic logic unit

The ALU is capable of performing two classes of operations: arithmetic and logic.[125] The set of arithmetic operations that a particular ALU supports may be limited to addition and subtraction, or might include multiplication, division, trigonometry functions such as sine, cosine, etc., and square roots. Some can operate only on whole numbers (integers) while others use floating point to represent real numbers, albeit with limited precision. However, any computer that is capable of performing just the simplest operations can be programmed to break down the more complex operations into simple steps that it can perform. Therefore, any computer can be programmed to perform any arithmetic operation—although it will take more time to do so if its ALU does not directly support the operation. An ALU may also compare numbers and return Boolean truth values (true or false) depending on whether one is equal to, greater than or less than the other ("is 64 greater than 65?"). Logic operations involve Boolean logic: AND, OR, XOR, and NOT. These can be useful for creating complicated conditional statements and processing Boolean logic.

Superscalar computers may contain multiple ALUs, allowing them to process several instructions simultaneously.[126] Graphics processors and computers with SIMD and MIMD features often contain ALUs that can perform arithmetic on vectors and matrices.

Memory

Main articles: Computer memory and Computer data storage



Magnetic-core memory (using magnetic cores) was the computer memory of choice in the 1960s, until it was replaced by semiconductor memory (using MOS memory cells).

A computer's memory can be viewed as a list of cells into which numbers can be placed or read. Each cell has a numbered "address" and can store a single number. The computer can be instructed to "put the number 123 into the cell numbered 1357" or to "add the number that is in cell 1357 to the number that is in cell 2468 and put the answer into cell 1595." The information stored in memory may represent practically anything. Letters, numbers, even computer instructions can be placed into memory with equal ease. Since the CPU does not differentiate between different types of information, it is the software's responsibility to give significance to what the memory sees as nothing but a series of numbers.

In almost all modern computers, each memory cell is set up to store binary numbers in groups of eight bits (called a byte). Each byte is able to represent 256 different numbers ($2^8 = 256$); either from 0 to 255 or −128 to +127. To store larger numbers, several consecutive bytes may be used (typically, two, four or eight). When negative numbers are required, they are usually stored in two's complement notation. Other arrangements are possible, but are usually not seen outside of specialized applications or historical contexts. A computer can store any kind of information in memory if it can be represented numerically. Modern computers have billions or even trillions of bytes of memory.

The CPU contains a special set of memory cells called registers that can be read and written to much more rapidly than the main memory area. There are typically between two and one hundred registers depending on the type of CPU. Registers are used for the most frequently needed data items to avoid having to access main memory every time data is needed. As data is constantly being worked on, reducing the need to access main memory (which is often slow compared to the ALU and control units) greatly increases the computer's speed.

Computer main memory comes in two principal varieties:

- random-access memory or RAM
- read-only memory or ROM

RAM can be read and written to anytime the CPU commands it, but ROM is preloaded with data and software that never changes, therefore the CPU can only read from it. ROM is typically used to store the computer's initial start-up instructions. In general, the contents of RAM are erased when the power to the computer is turned off, but ROM retains its data indefinitely. In a PC, the ROM contains a specialized program called the BIOS that orchestrates loading the computer's operating system from the hard disk drive into RAM whenever the computer is turned on or reset. In embedded computers, which frequently do not have disk drives, all of the required software may be stored in ROM. Software stored in ROM is often called firmware, because it is notionally more like hardware than software. Flash memory blurs the distinction between ROM and RAM, as it retains its data when turned off but is also rewritable. It is typically much slower than conventional ROM and RAM however, so its use is restricted to applications where high speed is unnecessary.[g]

In more sophisticated computers there may be one or more RAM cache memories, which are slower than registers but faster than main memory. Generally computers with this sort of cache are designed to move frequently needed data into the cache automatically, often without the need for any intervention on the programmer's part.

Input/output (I/O)

Main article: Input/output

Hard disk drives are common storage devices used with computers.

I/O is the means by which a computer exchanges information with the outside world.[128] Devices that provide input or output to the computer are called peripherals.[129] On a typical personal computer, peripherals include input devices like the keyboard and mouse, and output devices such as the display and printer. Hard disk drives, floppy disk drives and optical disc drives serve as both input and output devices. Computer networking is another form of I/O. I/O devices are often complex computers in their own right, with their own CPU and memory. A graphics processing unit might contain fifty or more tiny computers that perform the calculations necessary to display 3D graphics.[citation needed] Modern desktop computers contain many smaller computers that assist the main CPU in performing I/O. A 2016-era flat screen display contains its own computer circuitry.

Multitasking

Main article: Computer multitasking

While a computer may be viewed as running one gigantic program stored in its main memory, in some systems it is necessary to give the appearance of running several programs simultaneously. This is achieved by multitasking, i.e. having the computer switch rapidly between running each program in turn.[130] One means by which this is done is with a special signal called an interrupt, which can periodically cause the computer to stop executing instructions where it was and do something else instead. By remembering where it was executing prior to the interrupt, the computer can return to that task later. If several programs are running "at the same time". Then the interrupt generator might be causing several hundred interrupts per second, causing a program switch each time. Since modern computers typically execute instructions several orders of magnitude faster than human perception, it may appear that many programs are running at the same time, even though only one is ever executing in any given instant. This method of multitasking is sometimes termed "time-sharing" since each program is allocated a "slice" of time in turn.[131]

Before the era of inexpensive computers, the principal use for multitasking was to allow many people to share the same computer. Seemingly, multitasking would cause a computer that is switching between several programs to run more slowly, in direct proportion to the number of programs it is running, but most programs spend much of their time waiting for slow input/output devices to complete their tasks. If a program is waiting for the user to click on the mouse or press a key on the keyboard, then it will not take a "time slice" until the event it is waiting for has occurred. This frees up time for other programs to execute so that many programs may be run simultaneously without unacceptable speed loss.

Multiprocessing

Main article: Multiprocessing

Cray designed many supercomputers that used multiprocessing heavily.

Some computers are designed to distribute their work across several CPUs in a multiprocessing configuration, a technique once employed in only large and powerful machines such as supercomputers, mainframe computers and servers. Multiprocessor and multi-core (multiple CPUs on a single integrated circuit) personal and laptop computers are now widely available, and are being increasingly used in lower-end markets as a result.

Supercomputers in particular often have highly unique architectures that differ significantly from the basic stored-program architecture and from general-purpose computers.[h] They often feature thousands of CPUs, customized high-speed interconnects, and specialized computing hardware. Such designs tend to be useful for only specialized tasks due to the large scale of program organization required to use most of the available resources at once. Supercomputers usually see usage in large-scale simulation, graphics rendering, and cryptography applications, as well as with other so-called "embarrassingly parallel" tasks.

Software

Main article: Software

Software is the part of a computer system that consists of the encoded information that determines the computer's operation, such as data or instructions on how to process the data. In contrast to the physical hardware from which the system is built, software is immaterial. Software includes computer programs, libraries and related non-executable data, such as online documentation or digital media. It is often divided into system software and application software. Computer hardware and software require each other and neither is useful on its own. When software is stored in hardware that cannot easily be modified, such as with BIOS ROM in an IBM PC compatible computer, it is sometimes called "firmware".

Operating system /system software	Unix and BSD	UNIX System V, IBM AIX, HP-UX, Solaris (SunOS), IRIX, List of BSD operating systems
	Linux	List of Linux distributions, Comparison of Linux distributions
	Microsoft Windows	Windows 95, Windows 98, Windows NT, Windows 2000, Windows ME, Windows XP, Windows Vista, Windows 7, Windows 8, Windows 8.1, Windows 10, Windows 11
	DOS	86-DOS (QDOS), IBM PC DOS, MS-DOS, DR-DOS, FreeDOS
	Macintosh operating systems	Classic Mac OS, macOS (previously OS X and Mac OS X)
	Embedded and real-time	List of embedded operating systems
	Experimental	Amoeba, Oberon–AOS, Bluebottle, A2, Plan 9 from Bell Labs
Library	Multimedia	DirectX, OpenGL, OpenAL, Vulkan (API)
	Programming library	C standard library, Standard Template Library
Data	Protocol	TCP/IP, Kermit, FTP, HTTP, SMTP
	File format	HTML, XML, JPEG, MPEG, PNG
User interface	Graphical user interface (WIMP)	Microsoft Windows, GNOME, KDE, QNX Photon, CDE, GEM, Aqua
	Text-based user interface	Command-line interface, Text user interface
Application software	Office suite	Word processing, Desktop publishing, Presentation program, Database management system, Scheduling & Time management, Spreadsheet, Accounting software
	Internet Access	Browser, Email client, Web server, Mail transfer agent, Instant messaging

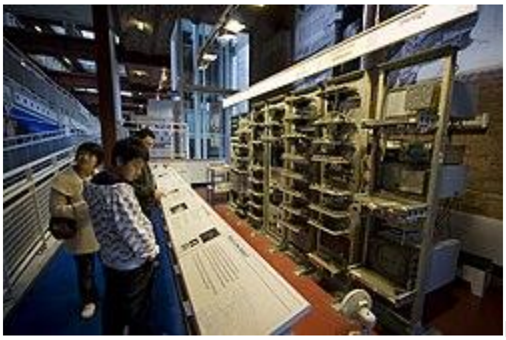
	Design and manufacturing	Computer-aided design, Computer-aided manufacturing, Plant management, Robotic manufacturing, Supply chain management
	Graphics	Raster graphics editor, Vector graphics editor, 3D modeler, Animation editor, 3D computer graphics, Video editing, Image processing
	Audio	Digital audio editor, Audio playback, Mixing, Audio synthesis, Computer music
	Software engineering	Compiler, Assembler, Interpreter, Debugger, Text editor, Integrated development environment, Software performance analysis, Revision control, Software configuration management
	Educational	Edutainment, Educational game, Serious game, Flight simulator
	Games	Strategy, Arcade, Puzzle, Simulation, First-person shooter, Platform, Massively multiplayer, Interactive fiction
	Misc	Artificial intelligence, Antivirus software, Malware scanner, Installer/Package management systems, File manager

Programs

The defining feature of modern computers which distinguishes them from all other machines is that they can be programmed. That is to say that some type of instructions (the program) can be given to the computer, and it will process them. Modern computers based on the von Neumann architecture often have machine code in the form of an imperative programming language. In practical terms, a computer program may be just a few instructions or extend to many millions of instructions, as do the programs for word processors and web browsers for example. A typical modern computer can execute billions of instructions per second (gigaflops) and rarely makes a mistake over many years of operation. Large computer programs consisting of several million instructions may take teams of programmers years to write, and due to the complexity of the task almost certainly contain errors.

Stored program architecture

Main articles: Computer program and Computer programming



Replica of the Manchester Baby, the world's first electronic stored-program computer, at the Museum of Science and Industry in Manchester, England

This section applies to most common RAM machine–based computers.

In most cases, computer instructions are simple: add one number to another, move some data from one location to another, send a message to some external device, etc. These instructions are read from the computer's memory and are generally carried out (executed) in the order they were given. However, there are usually specialized instructions to tell the computer to jump ahead or backwards to some other place in the program and to carry on executing from there. These are called "jump" instructions (or branches). Furthermore, jump instructions may be made to happen conditionally so that different sequences of instructions may be used depending on the result of some previous calculation or some external event. Many computers directly support subroutines by providing a type of jump that "remembers" the location it jumped from and another instruction to return to the instruction following that jump instruction.

Program execution might be likened to reading a book. While a person will normally read each word and line in sequence, they may at times jump back to an earlier place in the text or skip sections that are not of interest. Similarly, a computer may sometimes go back and repeat the instructions in some section of the program over and over again until some internal condition is met. This is called the flow of control within the program and it is what allows the computer to perform tasks repeatedly without human intervention.

Comparatively, a person using a pocket calculator can perform a basic arithmetic operation such as adding two numbers with just a few button presses. But to add together all of the numbers from 1 to 1,000 would take thousands of button presses and a lot of time, with a near certainty of making a mistake. On the other hand, a computer may be programmed to do this with just a few simple instructions. The following example is written in the MIPS assembly language: