



Transistor

A **transistor** is a semiconductor device used to amplify or switch electrical signals and power. It is one of the basic building blocks of modern electronics.^[1] It is composed of semiconductor material, usually with at least three terminals for connection to an electronic circuit. A voltage or current applied to one pair of the transistor's terminals controls the current through another pair of terminals. Because the controlled (output) power can be higher than the controlling (input) power, a transistor can amplify a signal. Some transistors are packaged individually, but many more in miniature form are found embedded in integrated circuits. Because transistors are the key active components in practically all modern electronics, many people consider them one of the 20th century's greatest inventions.^[2]

Physicist Julius Edgar Lilienfeld proposed the concept of a field-effect transistor (FET) in 1925,^[3] but it was not possible to construct a working device at that time.^[4] The first working device was a point-contact transistor invented in 1947 by physicists John Bardeen, Walter Brattain, and William Shockley at Bell Labs who shared the 1956 Nobel Prize in Physics for their achievement.^[5] The most widely used type of transistor, the metal–oxide–semiconductor field-effect transistor (MOSFET), was invented at Bell Labs between 1955 and 1960.^{[6][7][8][9][10][11]} Transistors revolutionized the field of electronics and paved the way for smaller and cheaper radios, calculators, computers, and other electronic devices.

Most transistors are made from very pure silicon, and some from germanium, but certain other semiconductor materials are sometimes used. A transistor may have only one kind of charge carrier in a field-effect transistor, or may have two kinds of charge carriers in bipolar junction transistor devices. Compared with the vacuum tube, transistors are generally smaller and require less power to operate. Certain vacuum tubes have advantages over transistors at very high operating frequencies or high operating voltages, such as traveling-wave tubes and gyrotrons. Many types of transistors are made to standardized specifications by multiple manufacturers.

Transistor



Component type Active

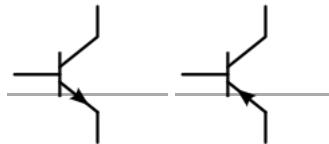
Inventor John Bardeen
Walter Brattain
William Shockley

Invention year 1947

First produced 1950s

Pin names Base, collector, and emitter

Electronic symbol



NPN and PNP symbols

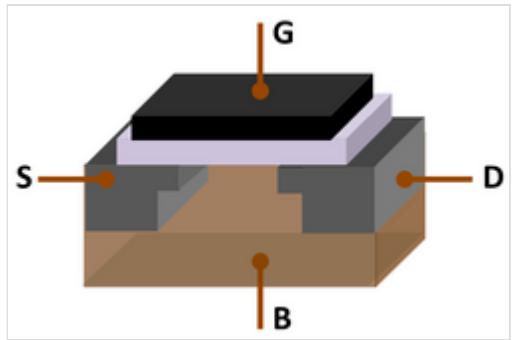
History

The thermionic triode, a vacuum tube invented in 1907, enabled amplified radio technology and long-distance telephony. The triode, however, was a fragile device that consumed a substantial amount of power. In 1909, physicist William Eccles discovered the crystal diode oscillator.^[12] Physicist Julius Edgar Lilienfeld filed a patent for a field-effect transistor (FET) in Canada in 1925,^[13] intended as a solid-state replacement for the triode.^{[14][15]} He filed identical patents in the United States in 1926^[16] and 1928.^{[17][18]} However, he did not publish any research articles about his devices nor did his patents cite any specific examples of a working prototype. Because the production of high-quality semiconductor materials was still decades away, Lilienfeld's solid-state amplifier ideas would not have found practical use in the 1920s and 1930s, even if such a device had been built.^[19] In 1934, inventor Oskar Heil patented a similar device in Europe.^[20]

Bipolar transistors

From November 17 to December 23, 1947, John Bardeen and Walter Brattain at AT&T's Bell Labs in Murray Hill, New Jersey, performed experiments and observed that when two gold point contacts were applied to a crystal of germanium, a signal was produced with the output power greater than the input.^[21] Solid State Physics Group leader William Shockley saw the potential in this, and over the next few months worked to greatly expand the knowledge of semiconductors. The term transistor was coined by John R. Pierce as a contraction of the term transresistance.^{[22][23][24]} According to Lillian Hoddeson and Vicki Daitch, Shockley proposed that Bell Labs' first patent for a transistor should be based on the field-effect and that he be named as the inventor. Having unearthed Lilienfeld's patents that went into obscurity years earlier, lawyers at Bell Labs advised against Shockley's proposal because the idea of a field-effect transistor that used an electric field as a *grid* was not new. Instead, what Bardeen, Brattain, and Shockley invented in 1947 was the first point-contact transistor.^[19] To acknowledge this accomplishment, Shockley, Bardeen and Brattain jointly received the 1956 Nobel Prize in Physics "for their researches on semiconductors and their discovery of the transistor effect".^{[25][26]}

Shockley's team initially attempted to build a field-effect transistor (FET) by trying to modulate the conductivity of a semiconductor, but was unsuccessful, mainly due to problems with the surface states, the dangling bond, and the germanium and copper compound materials. Trying to understand the mysterious reasons behind this failure led them instead to invent the bipolar point-contact and junction transistors.^{[27][28]}



Metal–oxide–semiconductor field-effect transistor (MOSFET), showing gate (G), body (B), source (S) and drain (D) terminals. The gate is separated from the body by an insulating layer (white).



Julius Edgar Lilienfeld proposed the concept of a field-effect transistor in 1925.

In 1948, the point-contact transistor was independently invented by physicists Herbert Mataré and Heinrich Welker while working at the Compagnie des Freins et Signaux Westinghouse, a Westinghouse subsidiary in Paris. Mataré had previous experience in developing crystal rectifiers from silicon and germanium in the German radar effort during World War II. With this knowledge, he began researching the phenomenon of interference in 1947. By June 1948, witnessing currents flowing through point-contacts, he produced consistent results using samples of germanium produced by Welker, similar to what Bardeen and Brattain had accomplished earlier in December 1947. Realizing that Bell Labs' scientists had already invented the transistor, the company rushed to get its transistor into production for amplified use in France's telephone network, filing his first transistor patent application on August 13, 1948.^{[29][30][31]}



John Bardeen, William Shockley, and Walter Brattain at Bell Labs in 1948; Bardeen and Brattain invented the point-contact transistor in 1947 and Shockley invented the bipolar junction transistor in 1948.

The first bipolar junction transistors were invented by Bell Labs' William Shockley, who applied for patent (2,569,347) on June 26, 1948. On April 12, 1950, Bell Labs chemists Gordon Teal and Morgan Sparks successfully produced a working bipolar NPN junction amplifying germanium transistor. Bell announced the discovery of this new sandwich transistor in a press release on July 4, 1951.^{[32][33]}

The first high-frequency transistor was the surface-barrier germanium transistor developed by Philco in 1953, capable of operating at frequencies up to 60 MHz.^[34] They were made by etching depressions into an n-type germanium base from both sides with jets of indium(III) sulfate until it was a few ten-thousandths of an inch thick. Indium electroplated into the depressions formed the collector and emitter.^{[35][36]}

AT&T first used transistors in telecommunications equipment in the No. 4A Toll Crossbar Switching System in 1953, for selecting trunk circuits from routing information encoded on translator cards.^[37] Its predecessor, the Western Electric No. 3A phototransistor, read the mechanical encoding from punched metal cards.

The first prototype pocket transistor radio was shown by INTERMETALL, a company founded by Herbert Mataré in 1952, at the Internationale Funkausstellung Düsseldorf from August 29 to September 6, 1953.^{[38][39]} The first production-model pocket transistor radio was the Regency TR-1, released in October 1954.^[26] Produced as a joint venture between the Regency Division of Industrial Development Engineering Associates, I.D.E.A. and Texas Instruments of Dallas, Texas, the TR-1 was manufactured in Indianapolis, Indiana. It was a near pocket-sized radio with four transistors and one germanium diode. The industrial design was outsourced to the Chicago firm of Painter, Teague and Petertil. It was initially released in one of six colours: black, ivory, mandarin red, cloud grey, mahogany and olive green. Other colours shortly followed.^{[40][41][42]}



A replica of the first working transistor, a point-contact transistor invented in 1947

The first production all-transistor car radio was developed by Chrysler and Philco corporations and was announced in the April 28, 1955, edition of *The Wall Street Journal*. Chrysler made the Mopar model 914HR available as an option starting in fall 1955 for its new line of 1956 Chrysler and Imperial cars, which reached dealership showrooms on October 21, 1955.^{[43][44]}

The Sony TR-63, released in 1957, was the first mass-produced transistor radio, leading to the widespread adoption of transistor radios.^[45] Seven million TR-63s were sold worldwide by the mid-1960s.^[46] Sony's success with transistor radios led to transistors replacing vacuum tubes as the dominant electronic technology in the late 1950s.^[47]

The first working silicon transistor was developed at Bell Labs on January 26, 1954, by Morris Tanenbaum. The first production commercial silicon transistor was announced by Texas Instruments in May 1954. This was the work of Gordon Teal, an expert in growing crystals of high purity, who had previously worked at Bell Labs.^{[48][49][50]}

Field-effect transistors

The basic principle of the field-effect transistor (FET) was first proposed by physicist Julius Edgar Lilienfeld when he filed a patent for a device similar to MESFET in 1926, and for an insulated-gate field-effect transistor in 1928.^{[15][51]} The FET concept was later also theorized by engineer Oskar Heil in the 1930s and by William Shockley in the 1940s.

In 1945, JFET was patented by Heinrich Welker.^[52] Following Shockley's theoretical treatment on JFET in 1952, a working practical JFET was made in 1953 by George C. Dacey and Ian M. Ross.^[53]

In 1948, Bardeen and Brattain patented the progenitor of MOSFET at Bell Labs, an insulated-gate FET (IGFET) with an inversion layer. Bardeen's patent, and the concept of an inversion layer, forms the basis of CMOS and DRAM technology today.^[54]

In the early years of the semiconductor industry, companies focused on the junction transistor, a relatively bulky device that was difficult to mass-produce, limiting it to several specialized applications. Field-effect transistors (FETs) were theorized as potential alternatives, but researchers could not get them to work properly, largely due to the surface state barrier that prevented the external electric field from penetrating the material.^[55]



Herbert Mataré (pictured in 1950)
independently invented a point-contact
transistor in June 1948.



A Philco surface-barrier transistor
developed and produced in 1953

MOSFET (MOS transistor)

In 1955, Carl Frosch and Lincoln Derick accidentally grew a layer of silicon dioxide over the silicon wafer, for which they observed surface passivation effects.^{[6][56]} By 1957 Frosch and Derick, using masking and predeposition, were able to manufacture silicon dioxide field effect transistors; the first planar transistors, in which drain and source were adjacent at the same surface.^[7] They showed that silicon dioxide insulated, protected silicon wafers and prevented dopants from diffusing into the wafer.^{[6][7]} After this, J.R. Ligenza and W.G. Spitzer studied the mechanism of thermally grown oxides, fabricated a high quality Si/SiO₂ stack and published their results in 1960.^{[57][58][59]}

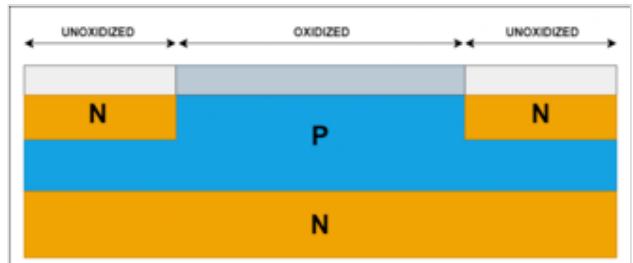


Diagram of one of the SiO₂ transistor devices made by Frosch and Derrick^[7]

Following this research, Mohamed Atalla and Dawon Kahng proposed a silicon MOS transistor in 1959^[60] and successfully demonstrated a working MOS device with their Bell Labs team in 1960.^{[61][62]} Their team included E. E. LaBate and E. I. Povilonis who fabricated the device; M. O. Thurston, L. A. D'Asaro, and J. R. Ligenza who developed the diffusion processes, and H. K. Gummel and R. Lindner who characterized the device.^{[8][9]} With its high scalability,^[63] much lower power consumption, and higher density than bipolar junction transistors,^[64] the MOSFET made it possible to build high-density integrated circuits,^[65] allowing the integration of more than 10,000 transistors in a single IC.^[66]

Bardeen and Brattain's 1948 inversion layer concept forms the basis of CMOS technology today.^[67] The CMOS (complementary MOS) was invented by Chih-Tang Sah and Frank Wanlass at Fairchild Semiconductor in 1963.^[68] The first report of a floating-gate MOSFET was made by Dawon Kahng and Simon Sze in 1967.^[69]

In 1967, Bell Labs researchers Robert Kerwin, Donald Klein and John Sarace developed the self-aligned gate (silicon-gate) MOS transistor, which Fairchild Semiconductor researchers Federico Faggin and Tom Klein used to develop the first silicon-gate MOS integrated circuit.^[70]

A double-gate MOSFET was first demonstrated in 1984 by Electrotechnical Laboratory researchers Toshihiro Sekigawa and Yutaka Hayashi.^{[71][72]} The FinFET (fin field-effect transistor), a type of 3D non-planar multi-gate MOSFET, originated from the research of Digh Hisamoto and his team at Hitachi Central Research Laboratory in 1989.^{[73][74]}

Importance

Because transistors are the key active components in practically all modern electronics, many people consider them one of the 20th century's greatest inventions.^[2]

The invention of the first transistor at Bell Labs was named an IEEE Milestone in 2009.^[75] Other Milestones include the inventions of the junction transistor in 1948 and the MOSFET in 1959.^[76]

The MOSFET is by far the most widely used transistor, in applications ranging from computers and electronics^[77] to communications technology such as smartphones.^[78] It has been considered the most important transistor,^[79] possibly the most important invention in electronics,^[80] and the device that enabled modern electronics.^[81] It has been the basis of modern digital electronics since the late 20th century, paving the way for the digital age.^[82] The US Patent and Trademark Office calls it a "groundbreaking invention that transformed life and culture around the world".^[78] Its ability to be mass-produced by a highly automated process (semiconductor device fabrication), from relatively basic materials, allows astonishingly low per-transistor costs. MOSFETs are the most numerously produced artificial objects in history, with more than 13 sextillion manufactured by 2018.^[83]

Although several companies each produce over a billion individually packaged (known as discrete) MOS transistors every year,^[84] the vast majority are produced in integrated circuits (also known as ICs, microchips, or simply chips), along with diodes, resistors, capacitors and other electronic components, to produce complete electronic circuits. A logic gate consists of up to about 20 transistors, whereas an advanced microprocessor, as of 2023, may contain as many as 134 billion transistors (and for exceptional chips, 2.6 trillion transistors, as of 2020).^[85] Transistors are often organized into logic gates in microprocessors to perform computation.^[86]

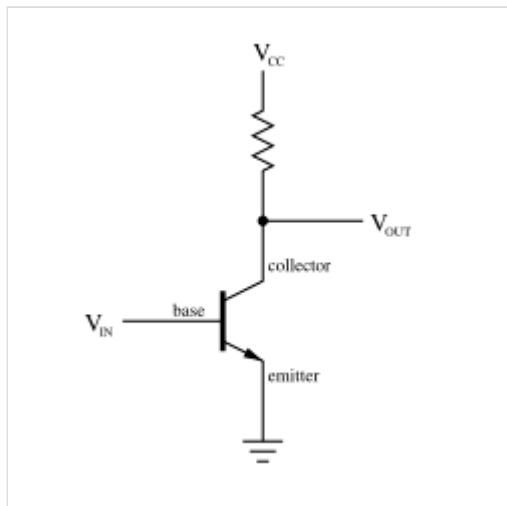
The transistor's low cost, flexibility and reliability have made it ubiquitous. Transistorized mechatronic circuits have replaced electromechanical devices in controlling appliances and machinery. It is often easier and cheaper to use a standard microcontroller and write a computer program to carry out a control function than to design an equivalent mechanical system.

Simplified operation

A transistor can use a small signal applied between one pair of its terminals to control a much larger signal at another pair of terminals, a property called gain. It can produce a stronger output signal, a voltage or current, proportional to a weaker input signal, acting as an amplifier. It can also be used as an electrically controlled switch, where the amount of current is determined by other circuit elements.^[87]

There are two types of transistors, with slight differences in how they are used:

- A bipolar junction transistor (BJT) has terminals labeled **base**, **collector** and **emitter**. A small current at the base terminal, flowing between the base and the emitter, can control or switch a much larger current between the collector and emitter.
- A field-effect transistor (FET) has terminals labeled **gate**, **source** and **drain**. A voltage at the gate can control a current between source and drain.^[88]



A simple circuit diagram showing the labels of an n–p–n bipolar transistor

The top image in this section represents a typical bipolar transistor in a circuit. A charge flows between emitter and collector terminals depending on the current in the base. Because the base and emitter connections behave like a semiconductor diode, a voltage drop develops between them. The amount of

this drop, determined by the transistor's material, is referred to as V_{BE} .^[88] (Base Emitter Voltage)

Transistor as a switch

Transistors are commonly used in digital circuits as electronic switches which can be either in an *on* or *off* state, both for high-power applications such as switched-mode power supplies and for low-power applications such as logic gates. Important parameters for this application include the current switched, the voltage handled, and the switching speed, characterized by the rise and fall times.^[88]

In a switching circuit, the goal is to simulate, as near as possible, the ideal switch having the properties of an open circuit when off, the short circuit when on, and an instantaneous transition between the two states. Parameters are chosen such that the *off* output is limited to leakage currents too small to affect connected circuitry, the resistance of the transistor in the *on* state is too small to affect circuitry, and the transition between the two states is fast enough not to have a detrimental effect.^[88]

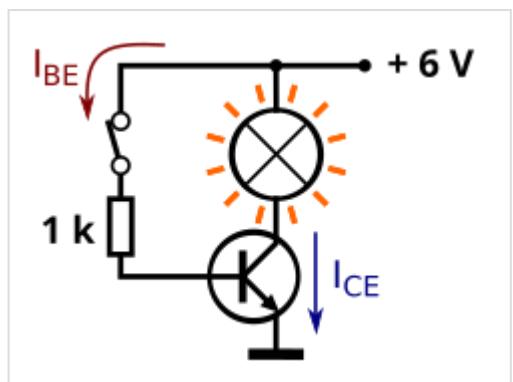
In a grounded-emitter transistor circuit, such as the light-switch circuit shown, as the base voltage rises, the emitter and collector currents rise exponentially. The collector voltage drops because of reduced resistance from the collector to the emitter. If the voltage difference between the collector and emitter were zero (or near zero), the collector current would be limited only by the load resistance (light bulb) and the supply voltage. This is called *saturation* because the current is flowing from collector to emitter freely. When saturated, the switch is said to be *on*.^[89]

The use of bipolar transistors for switching applications requires biasing the transistor so that it operates between its cut-off region in the off-state and the saturation region (*on*). This requires sufficient base drive current. As the transistor provides current gain, it facilitates the switching of a relatively large current in the collector by a much smaller current into the base terminal. The ratio of these currents varies depending on the type of transistor, and even for a particular type, varies depending on the collector current. In the example of a light-switch circuit, as shown, the resistor is chosen to provide enough base current to ensure the transistor is saturated.^[88] The base resistor value is calculated from the supply voltage, transistor C-E junction voltage drop, collector current, and amplification factor beta.^[90]

Transistor as an amplifier

The common-emitter amplifier is designed so that a small change in voltage (V_{in}) changes the small current through the base of the transistor whose current amplification combined with the properties of the circuit means that small swings in V_{in} produce large changes in V_{out} .^[88]

Various configurations of single transistor amplifiers are possible, with some providing current gain, some voltage gain, and some both.



BJT used as an electronic switch in grounded-emitter configuration

From mobile phones to televisions, vast numbers of products include amplifiers for sound reproduction, radio transmission, and signal processing. The first discrete-transistor audio amplifiers barely supplied a few hundred milliwatts, but power and audio fidelity gradually increased as better transistors became available and amplifier architecture evolved.^[88]

Modern transistor audio amplifiers of up to a few hundred watts are common and relatively inexpensive.

Comparison with vacuum tubes

Before transistors were developed, vacuum (electron) tubes (or in the UK *thermionic valves* or just *valves*) were the main active components in electronic equipment.

Advantages

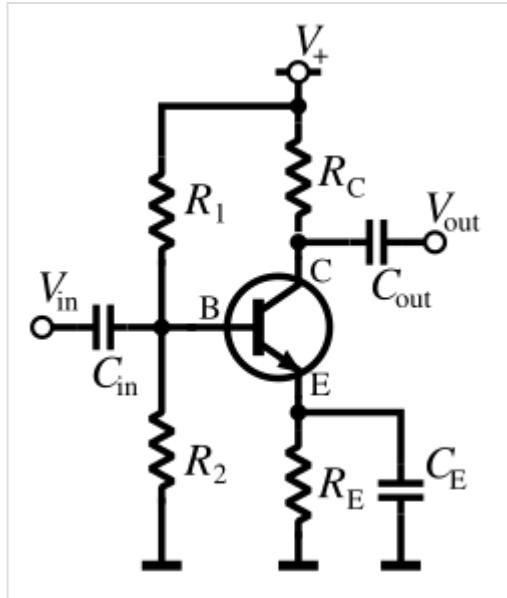
The key advantages that have allowed transistors to replace vacuum tubes in most applications are

- No cathode heater (which produces the characteristic orange glow of tubes), reducing power consumption, eliminating delay as tube heaters warm up, and immune from cathode poisoning and depletion.
- Very small size and weight, reducing equipment size.
- Large numbers of extremely small transistors can be manufactured as a single integrated circuit.
- Low operating voltages compatible with batteries of only a few cells.
- Circuits with greater energy efficiency are usually possible. For low-power applications (for example, voltage amplification) in particular, energy consumption can be very much less than for tubes.
- Complementary devices available, providing design flexibility including complementary-symmetry circuits, not possible with vacuum tubes.
- Very low sensitivity to mechanical shock and vibration, providing physical ruggedness and virtually eliminating shock-induced spurious signals (for example, microphonics in audio applications).
- Not susceptible to breakage of a glass envelope, leakage, outgassing, and other physical damage.

Limitations

Transistors may have the following limitations:

- They lack the higher electron mobility afforded by the vacuum of vacuum tubes, which is desirable for high-power, high-frequency operation – such as that used in some over-the-air television transmitters and in travelling-wave tubes used as amplifiers in some satellites
- Transistors and other solid-state devices are susceptible to damage from very brief electrical and thermal events, including electrostatic discharge in handling. Vacuum tubes are



An amplifier circuit, a common-emitter configuration with a voltage-divider bias circuit

electrically much more rugged.

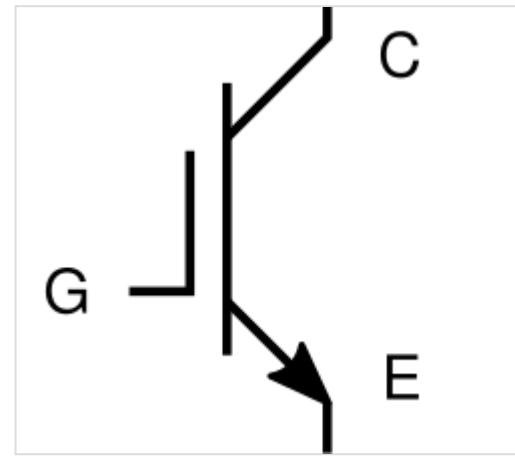
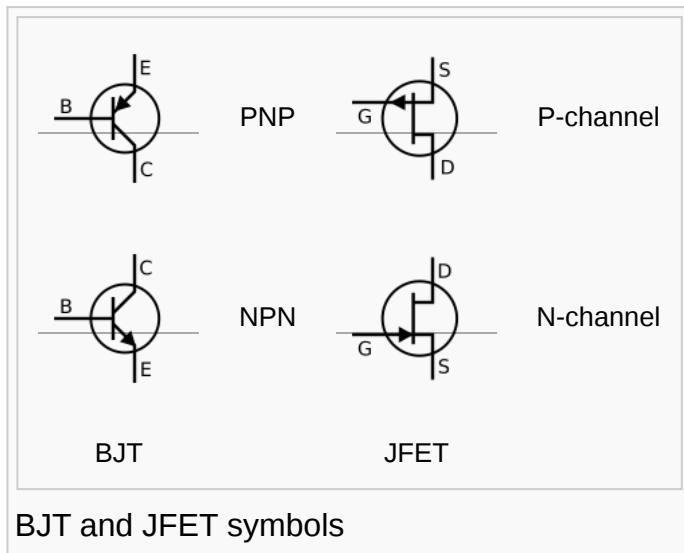
- They are sensitive to radiation and cosmic rays (special radiation-hardened chips are used for spacecraft devices).
- In audio applications, transistors lack the lower-harmonic distortion – the so-called tube sound – which is characteristic of vacuum tubes, and is preferred by some.^[91]

Types

Classification

Transistors are categorized by

- Structure: MOSFET (IGFET), BJT, JFET, insulated-gate bipolar transistor (IGBT), other type..
- Semiconductor material (dopants):
 - The metalloids; germanium (first used in 1947) and silicon (first used in 1954)—in amorphous, polycrystalline and monocrystalline form.
 - The compounds gallium arsenide (1966) and silicon carbide (1997).
 - The alloy silicon–germanium (1989)
 - The allotrope of carbon graphene (research ongoing since 2004), etc. (see Semiconductor material).
- Electrical polarity (positive and negative): NPN, PNP (BJTs), N-channel, P-channel (FETs).
- Maximum power rating: low, medium, high.
- Maximum operating frequency: low, medium, high, radio (RF), microwave frequency (the maximum effective frequency of a transistor in a common-emitter or common-source circuit is denoted by the term f_T , an abbreviation for transition frequency—the frequency at which the transistor yields unity voltage gain)
- Application: switch, general purpose, audio, high voltage, super-beta, matched pair.
- Physical packaging: through-hole metal, through-hole plastic, surface mount, ball grid array, power modules (see Packaging).
- Amplification factor h_{FE} , β_F (transistor beta)^[92] or g_m (transconductance).
- Working temperature: Extreme temperature transistors and traditional temperature transistors (−55 to 150 °C (−67 to 302 °F)). Extreme temperature transistors include high-temperature transistors (above 150 °C (302 °F)) and low-temperature transistors (below −55 °C (−67 °F)). The high-temperature transistors that operate thermally stable up to 250 °C (482 °F) can be developed by a general strategy of blending interpenetrating semi-crystalline conjugated polymers and high glass-transition temperature insulating polymers.^[93]

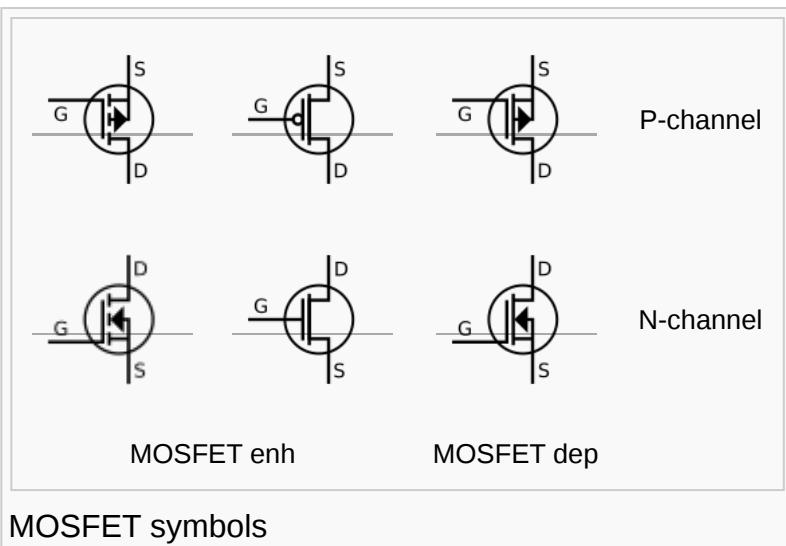


Insulated-gate bipolar transistor (IGBT)

Hence, a particular transistor may be described as *silicon*, *surface-mount*, *BJT*, *NPN*, *low-power*, *high-frequency switch*.

Mnemonics

Convenient mnemonic to remember the type of transistor (represented by an electrical symbol) involves the direction of the arrow. For the BJT, on an **n-p-n** transistor symbol, the arrow will "Not Point iN". On a **p-n-p** transistor symbol, the arrow "Points iN Proudly". However, this does not apply to MOSFET-based transistor symbols as the arrow is typically reversed (i.e. the arrow for the n-p-n points inside).



MOSFET symbols

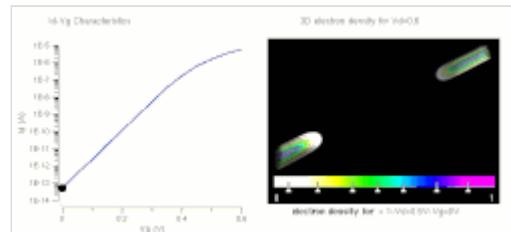
Field-effect transistor (FET)

The field-effect transistor, sometimes called a *unipolar transistor*, uses either electrons (in *n-channel FET*) or holes (in *p-channel FET*) for conduction. The four terminals of the FET are named *source*, *gate*, *drain*, and *body (substrate)*. On most FETs, the body is connected to the source inside the package, and this will be assumed for the following description.

In a FET, the drain-to-source current flows via a conducting channel that connects the *source* region to the *drain* region. The conductivity is varied by the electric field that is produced when a voltage is applied between the gate and source terminals, hence the current flowing between the drain and source is controlled by the voltage applied between the gate and source. As the gate–source voltage (V_{GS}) is increased, the drain–source current (I_{DS}) increases exponentially for V_{GS} below threshold, and then at a roughly quadratic rate: $(I_{DS} \propto (V_{GS} - V_T)^2)$, where V_T is the threshold voltage at which drain current begins)^[94] in the space-charge-limited region above threshold. A quadratic behavior is not observed in modern devices, for example, at the 65 nm technology node.^[95]

For low noise at narrow bandwidth, the higher input resistance of the FET is advantageous.

FETs are divided into two families: *junction FET (JFET)* and *insulated gate FET (IGFET)*. The IGFET is more commonly known as a *metal–oxide–semiconductor FET (MOSFET)*, reflecting its original construction from layers of metal (the gate), oxide (the insulation), and semiconductor. Unlike IGFETs, the JFET gate forms a p–n diode with the channel which lies between the source and drains. Functionally,



Operation of an FET and its I_d - V_g curve. At first, when no gate voltage is applied, there are no inversion electrons in the channel, so the device is turned off. As gate voltage increases, the inversion electron density in the channel increases, the current increases, and the device turns on.

this makes the n-channel JFET the solid-state equivalent of the vacuum tube triode which, similarly, forms a diode between its grid and cathode. Also, both devices operate in the *depletion-mode*, they both have a high input impedance, and they both conduct current under the control of an input voltage.

Metal–semiconductor FETs (MESFETs) are JFETs in which the reverse biased p–n junction is replaced by a metal–semiconductor junction. These, and the HEMTs (high-electron-mobility transistors, or HFETs), in which a two-dimensional electron gas with very high carrier mobility is used for charge transport, are especially suitable for use at very high frequencies (several GHz).

FETs are further divided into *depletion-mode* and *enhancement-mode* types, depending on whether the channel is turned on or off with zero gate-to-source voltage. For enhancement mode, the channel is off at zero bias, and a gate potential can *enhance* the conduction. For the depletion mode, the channel is on at zero bias, and a gate potential (of the opposite polarity) can *deplete* the channel, reducing conduction. For either mode, a more positive gate voltage corresponds to a higher current for n-channel devices and a lower current for p-channel devices. Nearly all JFETs are depletion-mode because the diode junctions would forward bias and conduct if they were enhancement-mode devices, while most IGFETs are enhancement-mode types.

Metal–oxide–semiconductor FET (MOSFET)

The metal–oxide–semiconductor field-effect transistor (MOSFET, MOS-FET, or MOS FET), also known as the metal–oxide–silicon transistor (MOS transistor, or MOS),^[65] is a type of field-effect transistor that is fabricated by the controlled oxidation of a semiconductor, typically silicon. It has an insulated gate, whose voltage determines the conductivity of the device. This ability to change conductivity with the amount of applied voltage can be used for amplifying or switching electronic signals. The MOSFET is by far the most common transistor, and the basic building block of most modern electronics.^[82] The MOSFET accounts for 99.9% of all transistors in the world.^[96]

Bipolar junction transistor (BJT)

Bipolar transistors are so named because they conduct by using both majority and minority carriers. The bipolar junction transistor, the first type of transistor to be mass-produced, is a combination of two junction diodes and is formed of either a thin layer of p-type semiconductor sandwiched between two n-type semiconductors (an n–p–n transistor), or a thin layer of n-type semiconductor sandwiched between two p-type semiconductors (a p–n–p transistor). This construction produces two p–n junctions: a base-emitter junction and a base-collector junction, separated by a thin region of semiconductor known as the base region. (Two junction diodes wired together without sharing an intervening semiconducting region will not make a transistor.)

BJTs have three terminals, corresponding to the three layers of semiconductor—an *emitter*, a *base*, and a *collector*. They are useful in amplifiers because the currents at the emitter and collector are controllable by a relatively small base current.^[97] In an n–p–n transistor operating in the active region, the emitter-base junction is forward-biased (electrons and holes recombine at the junction), and the base-collector junction is reverse-biased (electrons and holes are formed at, and move away from, the junction), and electrons are injected into the base region. Because the base is narrow, most of these electrons will diffuse into the reverse-biased base-collector junction and be swept into the collector; perhaps one-hundredth of the electrons will recombine in the base, which is the dominant mechanism in the base current. As well, as the base is lightly doped (in comparison to the emitter and collector regions), recombination rates are

low, permitting more carriers to diffuse across the base region. By controlling the number of electrons that can leave the base, the number of electrons entering the collector can be controlled.^[97] Collector current is approximately β (common-emitter current gain) times the base current. It is typically greater than 100 for small-signal transistors but can be smaller in transistors designed for high-power applications.

Unlike the field-effect transistor (see below), the BJT is a low-input-impedance device. Also, as the base-emitter voltage (V_{BE}) is increased the base-emitter current and hence the collector-emitter current (I_{CE}) increase exponentially according to the Shockley diode model and the Ebers-Moll model. Because of this exponential relationship, the BJT has a higher transconductance than the FET.

Bipolar transistors can be made to conduct by exposure to light because the absorption of photons in the base region generates a photocurrent that acts as a base current; the collector current is approximately β times the photocurrent. Devices designed for this purpose have a transparent window in the package and are called phototransistors.

Usage of MOSFETs and BJTs

The MOSFET is by far the most widely used transistor for both digital circuits as well as analog circuits,^[98] accounting for 99.9% of all transistors in the world.^[96] The bipolar junction transistor (BJT) was previously the most commonly used transistor during the 1950s to 1960s. Even after MOSFETs became widely available in the 1970s, the BJT remained the transistor of choice for many analog circuits such as amplifiers because of their greater linearity, up until MOSFET devices (such as power MOSFETs, LDMOS and RF CMOS) replaced them for most power electronic applications in the 1980s. In integrated circuits, the desirable properties of MOSFETs allowed them to capture nearly all market share for digital circuits in the 1970s. Discrete MOSFETs (typically power MOSFETs) can be applied in transistor applications, including analog circuits, voltage regulators, amplifiers, power transmitters, and motor drivers.



2N222A NPN Transistor.

Other transistor types

- Field-effect transistor (FET):
 - Metal–oxide–semiconductor field-effect transistor (MOSFET), where the gate is insulated by a shallow layer of insulator
 - p-type MOS (PMOS)
 - n-type MOS (NMOS)
 - Complementary MOS (CMOS)
 - RF CMOS, for radiofrequency amplification, reception
 - Multi-gate field-effect transistor (MuGFET)

- Fin field-effect transistor (FinFET), source/drain region shapes fins on the silicon surface
- GAAFET, Similar to FinFET but nanowires are used instead of fins, the nanowires are stacked vertically and are surrounded on 4 sides by the gate
- MBCFET, a variant of GAAFET that uses horizontal nanosheets instead of nanowires, made by Samsung. Also known as RibbonFET (made by Intel) and as horizontal nanosheet transistor.
- Thin-film transistor (TFT), used in LCD and OLED displays, types include amorphous silicon, LTPS, LTPO and IGZO transistors
- Floating-gate MOSFET (FGMOS), for non-volatile storage
- Power MOSFET, for power electronics
 - lateral diffused MOS (LDMOS)
- Carbon nanotube field-effect transistor (CNFET, CNTFET), where the channel material is replaced by a carbon nanotube
- Ferroelectric field-effect transistor (Fe FET), uses ferroelectric materials
- Junction gate field-effect transistor (JFET), where the gate is insulated by a reverse-biased p–n junction
- Metal–semiconductor field-effect transistor (MESFET), similar to JFET with a Schottky junction instead of a p–n junction
 - High-electron-mobility transistor (HEMT): GaN (gallium nitride), SiC (silicon carbide), Ga_2O_3 (gallium oxide), GaAs (gallium arsenide) transistors, MOSFETs, etc.
- Negative-capacitance FET (NC-FET)
- Inverted-T field-effect transistor (ITFET)
- Fast-reverse epitaxial diode field-effect transistor (FREDFET)
- Organic field-effect transistor (OFET), in which the semiconductor is an organic compound
- Ballistic transistor (disambiguation)
- FETs used to sense the environment
 - Ion-sensitive field-effect transistor (ISFET), to measure ion concentrations in solution,
 - Electrolyte–oxide–semiconductor field-effect transistor (EOSFET), neurochip,
 - Deoxyribonucleic acid field-effect transistor (DNAFET).
 - Field-effect transistor-based biosensor (Bio-FET)
- Bipolar junction transistor (BJT):
 - Heterojunction bipolar transistor, up to several hundred GHz, common in modern ultrafast and RF circuits
 - Schottky transistor
 - avalanche transistor
 - Darlington transistors are two BJTs connected together to provide a high current gain equal to the product of the current gains of the two transistors



A transistor symbol created on Portuguese pavement at the University of Aveiro

- Insulated-gate bipolar transistors (IGBTs) use a medium-power IGFET, similarly connected to a power BJT, to give a high input impedance. Power diodes are often connected between certain terminals depending on specific use. IGBTs are particularly suitable for heavy-duty industrial applications. The ASEA Brown Boveri (ABB) 5SNA2400E170100 ,^[99] intended for three-phase power supplies, houses three n-p-n IGBTs in a case measuring 38 by 140 by 190 mm and weighing 1.5 kg. Each IGBT is rated at 1,700 volts and can handle 2,400 amperes
- Phototransistor.
- Emitter-switched bipolar transistor (ESBT) is a monolithic configuration of a high-voltage bipolar transistor and a low-voltage power MOSFET in cascode topology. It was introduced by STMicroelectronics in the 2000s,^[100] and abandoned a few years later around 2012.^[101]
- Multiple-emitter transistor, used in transistor-transistor logic and integrated current mirrors
- Multiple-base transistor, used to amplify very-low-level signals in noisy environments such as the pickup of a record player or radio front ends. Effectively, it is a very large number of transistors in parallel where, at the output, the signal is added constructively, but random noise is added only stochastically.^[102]
- Tunnel field-effect transistor, where it switches by modulating quantum tunneling through a barrier.
- Diffusion transistor, formed by diffusing dopants into semiconductor substrate; can be both BJT and FET.
- Unijunction transistor, which can be used as a simple pulse generator. It comprises the main body of either p-type or n-type semiconductor with ohmic contacts at each end (terminals Base1 and Base2). A junction with the opposite semiconductor type is formed at a point along the length of the body for the third terminal (Emitter).
- Single-electron transistors (SET), consist of a gate island between two tunneling junctions. The tunneling current is controlled by a voltage applied to the gate through a capacitor.^[103]
- Nanofluidic transistor, controls the movement of ions through sub-microscopic, water-filled channels.^[104]
- Multigate devices:
 - Tetrode transistor
 - Pentode transistor
 - Trigate transistor (prototype by Intel)
 - Dual-gate field-effect transistors have a single channel with two gates in cascode, a configuration optimized for high-frequency amplifiers, mixers, and oscillators.
- Junctionless nanowire transistor (JNT), uses a simple nanowire of silicon surrounded by an electrically isolated wedding ring that acts to gate the flow of electrons through the wire.
- Nanoscale vacuum-channel transistor, when in 2012, NASA and the National Nanofab Center in South Korea were reported to have built a prototype vacuum-channel transistor in only 150 nanometers in size, can be manufactured cheaply using standard silicon



A Darlington transistor with the upper case removed so the transistor chip (the small square) can be seen. It is effectively two transistors on the same chip. One is much larger than the other, but both are large in comparison to transistors in large-scale integration because this particular example is intended for power applications.

semiconductor processing, can operate at high speeds even in hostile environments, and could consume just as much power as a standard transistor.^[105]

- Organic electrochemical transistor.
- Solaristor (from solar cell transistor), a two-terminal gate-less self-powered phototransistor.
- Germanium–tin transistor^[106]
- Wood transistor^{[107][108]}
- Paper transistor^[109]
- Carbon-doped silicon–germanium (Si–Ge:C) transistor
- Diamond transistor^[110]
- Aluminum nitride transistor^[111]
- Super-lattice castellated field effect transistors^[112]

Device identification

Three major identification standards are used for designating transistor devices. In each, the alphanumeric prefix provides clues to the type of the device.

Joint Electron Device Engineering Council (JEDEC)

The JEDEC part numbering scheme evolved in the 1960s in the United States. The JEDEC EIA-370 transistor device numbers usually start with 2N, indicating a three-terminal device.^[113] Dual-gate field-effect transistors are four-terminal devices, and begin with 3N. The prefix is followed by a two-, three- or four-digit number with no significance as to device properties, although early devices with low numbers tend to be germanium devices. For example, 2N3055 is a silicon n–p–n power transistor, 2N1301 is a p–n–p germanium switching transistor. A letter suffix, such as A, is sometimes used to indicate a newer variant, but rarely gain groupings.

JEDEC prefix table

Prefix	Type and usage
1N	two-terminal device, such as diodes
2N	three-terminal device, such as transistors or single-gate <u>field-effect transistors</u>
3N	four-terminal device, such as dual-gate field-effect transistors

Japanese Industrial Standard (JIS)

In Japan, the JIS semiconductor designation (JIS-C-7012), labels transistor devices starting with 2S,^[114] e.g., 2SD965, but sometimes the 2S prefix is not marked on the package—a 2SD965 might only be marked D965 and a 2SC1815 might be listed by a supplier as simply C1815. This series sometimes has suffixes, such as R, O, BL, standing for red, orange, blue, etc., to denote variants, such as tighter h_{FE} (gain) groupings.

JIS transistor prefix table

Prefix	Type and usage
2SA	high-frequency p–n–p BJT
2SB	audio-frequency p–n–p BJT
2SC	high-frequency n–p–n BJT
2SD	audio-frequency n–p–n BJT
2SJ	P-channel FET (both JFET and MOSFET)
2SK	N-channel FET (both JFET and MOSFET)

European Electronic Component Manufacturers Association (EECA)

The European Electronic Component Manufacturers Association (EECA) uses a numbering scheme that was inherited from [Pro Electron](#) when it merged with EECA in 1983. This scheme begins with two letters: the first gives the semiconductor type (A for germanium, B for silicon, and C for materials like GaAs); the second letter denotes the intended use (A for diode, C for general-purpose transistor, etc.). A three-digit sequence number (or one letter and two digits, for industrial types) follows. With early devices this indicated the case type. Suffixes may be used, with a letter (e.g. C often means high h_{FE} , such as in: BC549C^[115]) or other codes may follow to show gain (e.g. BC327-25) or voltage rating (e.g. BUK854-800A^[116]). The more common prefixes are:

EECA transistor prefix table

Prefix	Type and usage	Example	Equivalent	Reference
AC	Germanium, small-signal <u>AF</u> transistor	AC126	NTE102A	
AD	Germanium, <u>AF</u> power transistor	AD133	NTE179	
AF	Germanium, small-signal <u>RF</u> transistor	AF117	NTE160	
AL	Germanium, <u>RF</u> power transistor	ALZ10	NTE100	
AS	Germanium, switching transistor	ASY28	NTE101	
AU	Germanium, power switching transistor	AU103	NTE127	
BC	Silicon, small-signal transistor ("general purpose")	BC548	<u>2N3904</u>	Datasheet (https://www.mccsemi.com/pdf/Products/2N3904(TO-92).pdf)
BD	Silicon, power transistor	BD139	NTE375	Datasheet (http://www.fairchildsemi.com/ds/BD/BD135.pdf)
BF	Silicon, <u>RF</u> (high frequency) <u>BJT</u> or <u>FET</u>	BF245	NTE133	Datasheet (http://www.onsemi.com/pub_lnk/Collateral/BF245A-D.PDF)
BS	Silicon, switching transistor (BJT or MOSFET)	<u>BS170</u>	<u>2N7000</u>	Datasheet (http://www.fairchildsemi.com/ds/BS/BS170.pdf)
BL	Silicon, high frequency, high power (for transmitters)	BLW60	NTE325	Datasheet (http://www.datasheetcatalog.org/datasheet/philips/BLW60.pdf)
BU	Silicon, high voltage (for <u>CRT</u> horizontal deflection circuits)	BU2520A	NTE2354	Datasheet (http://www.datasheetcatalog.org/datasheet/philips/BU2520A.pdf)
CF	Gallium arsenide, small-signal <u>microwave</u> transistor (<u>MESFET</u>)	CF739	—	Datasheet (https://web.archive.org/web/20150109012745/http://www.kesun.com/pdf/rf%20transistor/CF739.pdf)
CL	Gallium arsenide, <u>microwave</u> power transistor (<u>FET</u>)	CLY10	—	Datasheet (http://www.datasheetcatalog.org/datasheet/siemens/CLY10.pdf)

Proprietary

Manufacturers of devices may have their proprietary numbering system, for example CK722. Since devices are second-sourced, a manufacturer's prefix (like MPF in MPF102, which originally would denote a Motorola FET) now is an unreliable indicator of who made the device. Some proprietary naming schemes adopt parts of other naming schemes, for example, a PN2222A is a (possibly Fairchild Semiconductor) 2N2222A in a plastic case (but a PN108 is a plastic version of a BC108, not a 2N108, while the PN100 is unrelated to other xx100 devices).

Military part numbers sometimes are assigned their codes, such as the British Military CV Naming System.

Manufacturers buying large numbers of similar parts may have them supplied with *house numbers*, identifying a particular purchasing specification and not necessarily a device with a standardized registered number. For example, an HP part 1854,0053 is a (JEDEC) 2N2218 transistor^{[117][118]} which is also assigned the CV number: CV7763^[119]

Naming problems

With so many independent naming schemes, and the abbreviation of part numbers when printed on the devices, ambiguity sometimes occurs. For example, two different devices may be marked J176 (one the J176 low-power JFET, the other the higher-powered MOSFET 2SJ176).

As older through-hole transistors are given surface-mount packaged counterparts, they tend to be assigned many different part numbers because manufacturers have their systems to cope with the variety in pinout arrangements and options for dual or matched n-p-n + p-n-p devices in one pack. So even when the original device (such as a 2N3904) may have been assigned by a standards authority, and well known by engineers over the years, the new versions are far from standardized in their naming.

Construction

Semiconductor material

Semiconductor material characteristics

Semiconductor material	Junction forward voltage @ 25 °C, V	Electron mobility @ 25 °C, m ² /(V·s)	Hole mobility @ 25 °C, m ² /(V·s)	Max. junction temp., °C
Ge	0.27	0.39	0.19	70 to 100
Si	0.71	0.14	0.05	150 to 200
GaAs	1.03	0.85	0.05	150 to 200
Al-Si junction	0.3	—	—	150 to 200

The first BJTs were made from germanium (Ge). Silicon (Si) types currently predominate but certain advanced microwave and high-performance versions now employ the *compound semiconductor* material gallium arsenide (GaAs) and the *semiconductor alloy* silicon–germanium (SiGe). Single-element semiconductor material (Ge and Si) is described as *elemental*.

Rough parameters for the most common semiconductor materials used to make transistors are given in the adjacent table. These parameters will vary with an increase in temperature, electric field, impurity level, strain, and sundry other factors.

The *junction forward voltage* is the voltage applied to the emitter-base junction of a BJT to make the base conduct a specified current. The current increases exponentially as the junction forward voltage is increased. The values given in the table are typical for a current of 1 mA (the same values apply to semiconductor diodes). The lower the junction forward voltage the better, as this means that less power is required to drive the transistor. The junction forward voltage for a given current decreases with an increase in temperature. For a typical silicon junction, the change is $-2.1 \text{ mV/}^{\circ}\text{C}$.^[120] In some circuits special compensating elements (sensistors) must be used to compensate for such changes.

The density of mobile carriers in the channel of a MOSFET is a function of the electric field forming the channel and of various other phenomena such as the impurity level in the channel. Some impurities, called dopants, are introduced deliberately in making a MOSFET, to control the MOSFET electrical behavior.

The electron mobility and hole mobility columns show the average speed that electrons and holes diffuse through the semiconductor material with an electric field of 1 volt per meter applied across the material. In general, the higher the electron mobility the faster the transistor can operate. The table indicates that Ge is a better material than Si in this respect. However, Ge has four major shortcomings compared to silicon and gallium arsenide:

1. Its maximum temperature is limited.
2. It has relatively high leakage current.
3. It cannot withstand high voltages.
4. It is less suitable for fabricating integrated circuits.

Because the electron mobility is higher than the hole mobility for all semiconductor materials, a given bipolar n-p-n transistor tends to be swifter than an equivalent p-n-p transistor. GaAs has the highest electron mobility of the three semiconductors. It is for this reason that GaAs is used in high-frequency applications. A relatively recent FET development, the high-electron-mobility transistor (HEMT), has a heterostructure (junction between different semiconductor materials) of aluminium gallium arsenide (AlGaAs)-gallium arsenide (GaAs) which has twice the electron mobility of a GaAs-metal barrier junction. Because of their high speed and low noise, HEMTs are used in satellite receivers working at frequencies around 12 GHz. HEMTs based on gallium nitride and aluminum gallium nitride (AlGaN/GaN HEMTs) provide still higher electron mobility and are being developed for various applications.

Maximum junction temperature values represent a cross-section taken from various manufacturers' datasheets. This temperature should not be exceeded or the transistor may be damaged.

Al-Si junction refers to the high-speed (aluminum-silicon) metal-semiconductor barrier diode, commonly known as a Schottky diode. This is included in the table because some silicon power IGFETs have a parasitic reverse Schottky diode formed between the source and drain as part of the fabrication process. This diode can be a nuisance, but sometimes it is used in the circuit.

Packaging

Discrete transistors can be individually packaged transistors or unpackaged transistor chips.

Transistors come in many different semiconductor packages (see image). The two main categories are through-hole (or leaded), and surface-mount, also known as surface-mount device (SMD). The ball grid array (BGA) is the latest surface-mount package. It has solder balls on the underside in place of leads. Because they are smaller and have shorter interconnections, SMDs have better high-frequency characteristics but lower power ratings.



Assorted discrete transistors

Transistor packages are made of glass, metal, ceramic, or plastic. The package often dictates the power rating and frequency characteristics. Power transistors have larger packages that can be clamped to heat sinks for enhanced cooling. Additionally, most power transistors have the collector or drain physically connected to the metal enclosure. At the other extreme, some surface-mount *microwave* transistors are as small as grains of sand.



Soviet-manufactured KT315b transistors

Often a given transistor type is available in several packages.

Transistor packages are mainly standardized, but the assignment of a transistor's functions to the terminals is not: other transistor types can assign other functions to the package's terminals. Even for the same transistor type the terminal assignment can vary (normally indicated by a suffix letter to the part number, q.e. BC212L and BC212K).

Nowadays most transistors come in a wide range of SMT packages. In comparison, the list of available through-hole packages is relatively small. Here is a short list of the most common through-hole transistors packages in alphabetical order: ATV, E-line, MRT, HRT, SC-43, SC-72, TO-3, TO-18, TO-39, TO-92, TO-126, TO220, TO247, TO251, TO262, ZTX851.

Unpackaged transistor chips (die) may be assembled into hybrid devices.^[121] The IBM SLT module of the 1960s is one example of such a hybrid circuit module using glass passivated transistor (and diode) die. Other packaging techniques for discrete transistors as chips include *direct chip attach* (DCA) and *chip-on-board* (COB).^[121]

Flexible transistors

Researchers have made several kinds of flexible transistors, including organic field-effect transistors.^{[122][123][124]} Flexible transistors are useful in some kinds of flexible displays and other flexible electronics.

See also



- [Alpha cutoff frequency](#)
- [Band gap](#)
- [Digital electronics](#)
- [Diffused junction transistor](#)
- [Moore's law](#)
- [Optical transistor](#)
- [Magneto-Electric Spin-Orbit](#)

- [Nanoelectromechanical relay](#)
- [Semiconductor device modeling](#)
- [Transistor count](#)
- [Transistor model](#)
- [Transresistance](#)
- [Very Large Scale Integration](#)
- [Trancitor](#)

References

1. "Transistor" (<https://www.britannica.com/technology/transistor>). *Britannica*. Retrieved

January 12, 2021.

2. "A History of the Invention of the Transistor and Where It Will Lead Us" (<https://massless.info/images/A%20history%20of%20the%20invention%20of%20the%20transistor%20and%20where%20it%20will%20lead%20us.pdf>) (PDF). *IEEE JOURNAL OF SOLID-STATE CIRCUITS Vol 32 No 12*. December 1997.
3. "Patent 272437 Summary" (<https://brevets-patents.ic.gc.ca/opic-cipo/cpd/eng/patent/272437/summary.html>). *Canadian Patents Database*.
4. "1926 – Field Effect Semiconductor Device Concepts Patented" (<http://www.computerhistory.org/siliconengine/field-effect-semiconductor-device-concepts-patented/>). *Computer History Museum*. Archived (<https://web.archive.org/web/20160322023120/http://www.computerhistory.org/siliconengine/field-effect-semiconductor-device-concepts-patented/>) from the original on March 22, 2016. Retrieved March 25, 2016.
5. "The Nobel Prize in Physics 1956" (https://www.nobelprize.org/nobel_prizes/physics/laureates/1956/). *Nobelprize.org*. Nobel Media AB. Archived (https://web.archive.org/web/20141216204332/http://www.nobelprize.org/nobel_prizes/physics/laureates/1956/) from the original on December 16, 2014. Retrieved December 7, 2014.
6. Huff, Howard; Riordan, Michael (September 1, 2007). "Frosch and Derick: Fifty Years Later (Foreword)" (<https://iopscience.iop.org/article/10.1149/2.F02073IF>). *The Electrochemical Society Interface*. **16** (3): 29. doi:[10.1149/2.F02073IF](https://doi.org/10.1149/2.F02073IF) (<https://doi.org/10.1149/2.F02073IF>). ISSN 1064-8208 (<https://search.worldcat.org/issn/1064-8208>).
7. Frosch, C. J.; Derick, L (1957). "Surface Protection and Selective Masking during Diffusion in Silicon" (<https://iopscience.iop.org/article/10.1149/1.2428650>). *Journal of the Electrochemical Society*. **104** (9): 547. doi:[10.1149/1.2428650](https://doi.org/10.1149/1.2428650) (<https://doi.org/10.1149/1.2428650>).
8. KAHNG, D. (1961). "Silicon-Silicon Dioxide Surface Device" (https://doi.org/10.1142/9789814503464_0076). *Technical Memorandum of Bell Laboratories*: 583–596. doi:[10.1142/9789814503464_0076](https://doi.org/10.1142/9789814503464_0076) (https://doi.org/10.1142/9789814503464_0076). ISBN 978-981-02-0209-5.
9. Lojek, Bo (2007). *History of Semiconductor Engineering*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. p. 321. ISBN 978-3-540-34258-8.
10. Ligenza, J.R.; Spitzer, W.G. (1960). "The mechanisms for silicon oxidation in steam and oxygen" (<https://linkinghub.elsevier.com/retrieve/pii/0022369760902195>). *Journal of Physics and Chemistry of Solids*. **14**: 131–136. Bibcode:1960JPCS...14..131L (<https://ui.adsabs.harvard.edu/abs/1960JPCS...14..131L>). doi:[10.1016/0022-3697\(60\)90219-5](https://doi.org/10.1016/0022-3697(60)90219-5) ([https://doi.org/10.1016/0022-3697\(60\)90219-5](https://doi.org/10.1016/0022-3697(60)90219-5)).
11. Lojek, Bo (2007). *History of Semiconductor Engineering*. Springer Science & Business Media. p. 120. ISBN 9783540342588.
12. Moavenzadeh, Fred (1990). *Concise Encyclopedia of Building and Construction Materials* (<https://books.google.com/books?id=YiJaEAUj258C&q=Eccles+Oscillator+Galena&pg=PA430>). MIT Press. ISBN 9780262132480.
13. Lilienfeld, Julius Edgar (1927). *Specification of electric current control mechanism patent application* (https://worldwide.espacenet.com/publicationDetails/originalDocument?FT=D&date=19270719&DB=&CC=CA&NR=272437A&KC=A&locale=en_EP#).
14. Vardalas, John (May 2003) Twists and Turns in the Development of the Transistor (<http://www.todaysengineer.org/2003/May/history.asp>) Archived (<https://web.archive.org/web/20150108082709/http://www.todaysengineer.org/2003/May/history.asp>) January 8, 2015, at the Wayback Machine *IEEE-USA Today's Engineer*.
15. Lilienfeld, Julius Edgar, "Method and apparatus for controlling electric current" U.S. patent 1,745,175 (<https://patents.google.com/patent/US1745175>) January 28, 1930 (filed in Canada 1925-10-22, in US October 8, 1926).
16. "Method And Apparatus For Controlling Electric Currents" (<https://patents.google.com/patent/US1745175>). United States Patent and Trademark Office.

17. "Amplifier For Electric Currents" (<https://patents.google.com/patent/US1877140>). United States Patent and Trademark Office.
18. "Device For Controlling Electric Current" (<https://patents.google.com/patent/US1900018>). United States Patent and Trademark Office.
19. "Twists and Turns in the Development of the Transistor" (<https://web.archive.org/web/20150108082709/http://www.todaysengineer.org/2003/May/history.asp>). Institute of Electrical and Electronics Engineers, Inc. Archived from the original (<http://www.todaysengineer.org/2003/May/history.asp>) on January 8, 2015.
20. Heil, Oskar, "Improvements in or relating to electrical amplifiers and other control arrangements and devices" (<http://v3.espacenet.com/publicationDetails/biblio?CC=GB&NR=439457&KC=&FT=E>), Patent No. GB439457, European Patent Office, filed in Great Britain 1934-03-02, published December 6, 1935 (originally filed in Germany March 2, 1934).
21. "November 17 – December 23, 1947: Invention of the First Transistor" (<http://www.aps.org/publications/apsnews/200011/history.cfm>). American Physical Society. Archived (<https://web.archive.org/web/20130120065607/http://www.aps.org/publications/apsnews/200011/history.cfm>) from the original on January 20, 2013.
22. Millman, S., ed. (1983). *A History of Engineering and Science in the Bell System, Physical Science* (1925–1980). AT&T Bell Laboratories. p. 102.
23. Bodanis, David (2005). *Electric Universe*. Crown Publishers, New York. ISBN 978-0-7394-5670-5.
24. "transistor". *American Heritage Dictionary* (3rd ed.). Boston: Houghton Mifflin. 1992.
25. "The Nobel Prize in Physics 1956" (http://nobelprize.org/nobel_prizes/physics/laureates/1956/). nobelprize.org. Archived (https://web.archive.org/web/20070312091604/http://nobelprize.org/nobel_prizes/physics/laureates/1956/) from the original on March 12, 2007.
26. Guarnieri, M. (2017). "Seventy Years of Getting Transistorized". *IEEE Industrial Electronics Magazine*. **11** (4): 33–37. doi:[10.1109/MIE.2017.2757775](https://doi.org/10.1109/MIE.2017.2757775) (<https://doi.org/10.1109%2FMIE.2017.2757775>). hdl:[11577/3257397](https://hdl.handle.net/11577/3257397) (<https://hdl.handle.net/11577%2F3257397>). S2CID 38161381 (<https://api.semanticscholar.org/CorpusID:38161381>).
27. Lee, Thomas H. (2003). "The Design of CMOS Radio-Frequency Integrated Circuits" ([http://web.archive.org/web/20211021005313/https://www.semanticscholar.org/paper/The-Design-of-CMOS-Radio-Frequency-Integrated-Ellis/c0018d231b4960f7a6c4f581b086212d7f8b0d15?p2df](https://web.archive.org/web/20211021005313/https://www.semanticscholar.org/paper/The-Design-of-CMOS-Radio-Frequency-Integrated-Ellis/c0018d231b4960f7a6c4f581b086212d7f8b0d15?p2df)). *Soldering & Surface Mount Technology*. **16** (2). Cambridge University Press. doi:[10.1108/ssmt.2004.21916bae.002](https://doi.org/10.1108/ssmt.2004.21916bae.002) (<https://doi.org/10.1108%2Fssmt.2004.21916bae.002>). ISBN 9781139643771. S2CID 108955928 (<https://api.semanticscholar.org/CorpusID:108955928>). Archived from the original (<https://www.semanticscholar.org/paper/The-Design-of-CMOS-Radio-Frequency-Integrated-Ellis/c0018d231b4960f7a6c4f581b086212d7f8b0d15?p2df>) on October 21, 2021.
28. Puers, Robert; Baldi, Livio; Voorde, Marcel Van de; Nooten, Sebastiaan E. van (2017). *Nanoelectronics: Materials, Devices, Applications, 2 Volumes* (<https://books.google.com/books?id=JOqVDgAAQBAJ&pg=PA14>). John Wiley & Sons. p. 14. ISBN 9783527340538.
29. FR 1010427 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=FR1010427>) H. F. Mataré / H. Welker / Westinghouse: "Nouveau système cristallin à plusieurs électrodes réalisant des relais de effets électroniques" filed on August 13, 1948
30. US 2673948 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=US2673948>) H. F. Mataré / H. Welker / Westinghouse, "Crystal device for controlling electric currents by means of a solid semiconductor" French priority August 13, 1948
31. "1948, The European Transistor Invention" (<http://www.computerhistory.org/semiconductor/timeline/1948-European.html>). Computer History Museum. Archived (<https://web.archive.org/web/20120929202704/http://www.computerhistory.org/semiconductor/timeline/1948-European.html>) from the original on September 29, 2012.

32. "1951: First Grown-Junction Transistors Fabricated | The Silicon Engine | Computer History Museum" (<https://www.computerhistory.org/siliconengine/first-grown-junction-transistors-fabricated/>). www.computerhistory.org. Archived (<https://web.archive.org/web/20170404035446/http://www.computerhistory.org/siliconengine/first-grown-junction-transistors-fabricated/>) from the original on April 4, 2017.
33. "A Working Junction Transistor" (<https://www.pbs.org/transistor/science/info/junctw.html>). [PBS](http://www.pbs.org/transistor). Archived (<https://web.archive.org/web/20170703002246/http://www.pbs.org/transistor/science/info/junctw.html>) from the original on July 3, 2017. Retrieved September 17, 2017.
34. Bradley, W.E. (December 1953). "The Surface-Barrier Transistor: Part I-Principles of the Surface-Barrier Transistor". *Proceedings of the IRE*. **41** (12): 1702–1706. doi:10.1109/JRPROC.1953.274351 (<https://doi.org/10.1109%2FJRPROC.1953.274351>). S2CID 51652314 (<https://api.semanticscholar.org/CorpusID:51652314>).
35. *The Wall Street Journal*, December 4, 1953, page 4, Article "Philco Claims Its Transistor Outperforms Others Now In Use"
36. Electronics magazine, January 1954, Article "Electroplated Transistors Announced"
37. P. Mallory, *Transistors and Their Circuits in the 4A Toll Crossbar Switching System*, AIEE Transactions, September 1953, p.388
38. 1953 Foreign Commerce Weekly; Volume 49; pp.23
39. "Der deutsche Erfinder des Transistors – Nachrichten Welt Print – DIE WELT" (https://www.welt.de/welt_print/article2721871/Der-deutsche-Erfinder-des-Transistors.html). *Die Welt*. Welt.de. November 23, 2011. Archived (https://web.archive.org/web/20160515182422/http://www.welt.de/welt_print/article2721871/Der-deutsche-Erfinder-des-Transistors.html) from the original on May 15, 2016. Retrieved May 1, 2016.
40. "Regency TR-1 Transistor Radio History" (<http://www.regencytr1.com/>). Archived (<https://web.archive.org/web/20041021040145/http://www.regencytr1.com/>) from the original on October 21, 2004. Retrieved April 10, 2006.
41. "The Regency TR-1 Family" (<http://www.ericwrobbel.com/books/regency.htm>). Archived (<https://web.archive.org/web/20170427155821/http://www.ericwrobbel.com/books/regency.htm>) from the original on April 27, 2017. Retrieved April 10, 2017.
42. "Regency manufacturer in USA, radio technology from United St" (http://www.radiomuseum.org/dsp_hersteller_detail.cfm?company_id=3886). Archived (https://web.archive.org/web/20170410214244/http://www.radiomuseum.org/dsp_hersteller_detail.cfm?company_id=3886) from the original on April 10, 2017. Retrieved April 10, 2017.
43. Wall Street Journal, "Chrysler Promises Car Radio With Transistors Instead of Tubes in '56", April 28, 1955, page 1
44. "FCA North America - Historical Timeline 1950-1959" (<https://web.archive.org/web/20150402062327/http://www.fcanorthamerica.com/company/Heritage/Pages/Chrysler-Heritage-1950.aspx>). www.fcanorthamerica.com. Archived from the original (<http://www.fcanorthamerica.com/company/Heritage/Pages/Chrysler-Heritage-1950.aspx>) on April 2, 2015. Retrieved December 5, 2017.
45. Skrabec, Quentin R. Jr. (2012). *The 100 Most Significant Events in American Business: An Encyclopedia* (<https://books.google.com/books?id=2kc69qrid9oC&pg=PA195>). ABC-CLIO. pp. 195–7. ISBN 978-0313398636.
46. Snook, Chris J. (November 29, 2017). "The 7 Step Formula Sony Used to Get Back On Top After a Lost Decade" (<https://www.inc.com/chris-j-snook/sonys-7-step-formula-for-entrepreneurial-success-business-longevity.html>). *Inc*.
47. Kozinsky, Sieva (January 8, 2014). "Education and the Innovator's Dilemma" (<https://www.wired.com/insights/2014/01/education-innovators-dilemma/>). *Wired*. Retrieved October 14, 2019.

48. Riordan, Michael (May 2004). "The Lost History of the Transistor" (<https://web.archive.org/web/20150531113132/https://spectrum.ieee.org/biomedical/devices/the-lost-history-of-the-transistor>). *IEEE Spectrum*: 48–49. Archived from the original (<https://spectrum.ieee.org/biomedical/devices/the-lost-history-of-the-transistor>) on May 31, 2015.
49. Chelikowski, J. (2004) "Introduction: Silicon in all its Forms", p. 1 in *Silicon: evolution and future of a technology*. P. Siffert and E. F. Krimmel (eds.). Springer, ISBN 3-540-40546-1.
50. McFarland, Grant (2006) *Microprocessor design: a practical guide from design planning to manufacturing*. McGraw-Hill Professional. p. 10. ISBN 0-07-145951-0.
51. Lilienfeld, Julius Edgar, "Device for controlling electric current" U.S. patent 1,900,018 (<http://patents.google.com/patent/US1900018>) March 7, 1933 (filed in US March 28, 1928).
52. Grundmann, Marius (2010). *The Physics of Semiconductors*. Springer-Verlag. ISBN 978-3-642-13884-3.
53. Nishizawa, Jun-Ichi (1982). "Junction Field-Effect Devices" (https://link.springer.com/chapter/10.1007%2F978-1-4684-7263-9_11#page-1). *Semiconductor Devices for Power Conditioning*. pp. 241–272. doi:10.1007/978-1-4684-7263-9_11 (https://doi.org/10.1007%2F978-1-4684-7263-9_11). ISBN 978-1-4684-7265-3.
54. Howard R. Duff (2001). "John Bardeen and transistor physics". *AIP Conference Proceedings*. Vol. 550. pp. 3–32. doi:10.1063/1.1354371 (<https://doi.org/10.1063%2F1.1354371>).
55. Moskowitz, Sanford L. (2016). *Advanced Materials Innovation: Managing Global Technology in the 21st century* (<https://books.google.com/books?id=2STRDAAAQBAJ&pg=PA168>). John Wiley & Sons. p. 168. ISBN 9780470508923.
56. US2802760A (<https://patents.google.com/patent/US2802760A>), Lincoln, Derick & Frosch, Carl J., "Oxidation of semiconductive surfaces for controlled diffusion", issued August 13, 1957
57. Ligenza, J. R.; Spitzer, W. G. (July 1, 1960). "The mechanisms for silicon oxidation in steam and oxygen" (<https://linkinghub.elsevier.com/retrieve/pii/0022369760902195>). *Journal of Physics and Chemistry of Solids*. **14**: 131–136. Bibcode:1960JPCS...14..131L (<https://ui.adsabs.harvard.edu/abs/1960JPCS...14..131L>). doi:10.1016/0022-3697(60)90219-5 (<https://doi.org/10.1016%2F0022-3697%2860%2990219-5>). ISSN 0022-3697 (<https://search.worldcat.org/issn/0022-3697>).
58. Deal, Bruce E. (1998). "Highlights Of Silicon Thermal Oxidation Technology" (<https://books.google.com/books?id=cr8FPGkiRS0C&pg=PA183>). *Silicon materials science and technology*. The Electrochemical Society. p. 183. ISBN 978-1566771931.
59. Lojek, Bo (2007). *History of Semiconductor Engineering*. Springer Science & Business Media. p. 322. ISBN 978-3540342588.
60. Bassett, Ross Knox (2007). *To the Digital Age: Research Labs, Start-up Companies, and the Rise of MOS Technology* (<https://books.google.com/books?id=UUbB3d2UnaAC&pg=PA22>). Johns Hopkins University Press. pp. 22–23. ISBN 978-0-8018-8639-3.
61. Atalla, M.; Kahng, D. (1960). "Silicon-silicon dioxide field induced surface devices". *IRE-AIEE Solid State Device Research Conference*.
62. "1960 – Metal Oxide Semiconductor (MOS) Transistor Demonstrated" (<https://www.computerhistory.org/siliconengine/metal-oxide-semiconductor-mos-transistor-demonstrated/>). *The Silicon Engine*. Computer History Museum. Retrieved January 16, 2023.
63. Motoyoshi, M. (2009). "Through-Silicon Via (TSV)" (<https://web.archive.org/web/20190719120523/https://pdfs.semanticscholar.org/8a44/93b535463daa7d7317b08d8900a33b8cbaf4.pdf>) (PDF). *Proceedings of the IEEE*. **97** (1): 43–48. doi:10.1109/JPROC.2008.2007462 (<https://doi.org/10.1109%2FJPROC.2008.2007462>). ISSN 0018-9219 (<https://search.worldcat.org/issn/0018-9219>). S2CID 29105721 (<https://api.semanticscholar.org/CorpusID:29105721>). Archived from the original (<https://pdfs.semanticscholar.org/8a44/93b535463daa7d7317b08d8900a33b8cbaf4.pdf>) (PDF) on July 19, 2019.

64. "Transistors Keep Moore's Law Alive" (https://www.eetimes.com/author.asp?section_id=36&doc_id=1334068). *EETimes*. December 12, 2018. Retrieved July 18, 2019.
65. "Who Invented the Transistor?" (<https://www.computerhistory.org/atchm/who-invented-the-transistor/>). *Computer History Museum*. December 4, 2013. Retrieved July 20, 2019.
66. Hittinger, William C. (1973). "Metal-Oxide-Semiconductor Technology". *Scientific American*. **229** (2): 48–59. Bibcode:1973SciAm.229b..48H (<https://ui.adsabs.harvard.edu/abs/1973SciAm.229b..48H>). doi:10.1038/scientificamerican0873-48 (<https://doi.org/10.1038%2Fscientificamerican0873-48>). ISSN 0036-8733 (<https://search.worldcat.org/issn/0036-8733>). JSTOR 24923169 (<https://www.jstor.org/stable/24923169>).
67. Howard R. Duff (2001). "John Bardeen and transistor physics". *AIP Conference Proceedings*. Vol. 550. pp. 3–32. doi:10.1063/1.1354371 (<https://doi.org/10.1063%2F1.1354371>).
68. "1963: Complementary MOS Circuit Configuration is Invented" (<https://www.computerhistory.org/siliconengine/complementary-mos-circuit-configuration-is-invented/>). *Computer History Museum*. Retrieved July 6, 2019.
69. D. Kahng and S. M. Sze, "A floating gate and its application to memory devices", *The Bell System Technical Journal*, vol. 46, no. 4, 1967, pp. 1288–1295
70. "1968: Silicon Gate Technology Developed for ICs" (<https://www.computerhistory.org/siliconengine/silicon-gate-technology-developed-for-ics/>). *Computer History Museum*. Retrieved July 22, 2019.
71. Colinge, J.P. (2008). *FinFETs and Other Multi-Gate Transistors* (<https://books.google.com/books?id=t1ojkCdTGEEC&pg=PA11>). Springer Science & Business Media. p. 11. ISBN 9780387717517.
72. Sekigawa, Toshihiro; Hayashi, Yutaka (August 1, 1984). "Calculated threshold-voltage characteristics of an XMOS transistor having an additional bottom gate". *Solid-State Electronics*. **27** (8): 827–828. Bibcode:1984SSEle..27..827S (<https://ui.adsabs.harvard.edu/abs/1984SSEle..27..827S>). doi:10.1016/0038-1101(84)90036-4 (<https://doi.org/10.1016%2F0038-1101%2884%2990036-4>). ISSN 0038-1101 (<https://search.worldcat.org/issn/0038-1101>).
73. "IEEE Andrew S. Grove Award Recipients" (<https://web.archive.org/web/20180909112404/https://www.ieee.org/about/awards/bios/grove-recipients.html>). *IEEE Andrew S. Grove Award*. Institute of Electrical and Electronics Engineers. Archived from the original (<https://www.ieee.org/about/awards/bios/grove-recipients.html>) on September 9, 2018. Retrieved July 4, 2019.
74. "The Breakthrough Advantage for FPGAs with Tri-Gate Technology" (<https://www.intel.com/content/dam/www/programmable/us/en/pdfs/literature/wp/wp-01201-fpga-tri-gate-technology.pdf>) (PDF). Intel. 2014. Retrieved July 4, 2019.
75. "Milestones:Invention of the First Transistor at Bell Telephone Laboratories, Inc., 1947" (https://www.ieeeghn.org/wiki/index.php/Milestones:Invention_of_the_First_Transistor_at_Bell_Telephone_Laboratories,_Inc.,_1947). *IEEE Global History Network*. IEEE. Archived (https://web.archive.org/web/20111008193522/http://www.ieeeghn.org/wiki/index.php/Milestones:Invention_of_the_First_Transistor_at_Bell_Telephone_Laboratories,_Inc.,_1947) from the original on October 8, 2011. Retrieved August 3, 2011.
76. "List of IEEE Milestones" (<https://www.invent.org/inductees/dawon-kahng>). *National Inventors Hall of Fame*. Retrieved June 27, 2019.
78. "Remarks by Director Iancu at the 2019 International Intellectual Property Conference" (<https://www.uspto.gov/about-us/news-updates/remarks-director-iancu-2019-international-intellectual-property-conference>). *United States Patent and Trademark Office*. June 10, 2019. Retrieved July 20, 2019.

79. Ashley, Kenneth L. (2002). *Analog Electronics with LabVIEW* (<https://books.google.com/books?id=0qkc2f6EXnQC&pg=PA10>). Prentice Hall Professional. p. 10. ISBN 9780130470652.
80. Thompson, S. E.; Chau, R. S.; Ghani, T.; Mistry, K.; Tyagi, S.; Bohr, M. T. (2005). "In search of "Forever," continued transistor scaling one new material at a time". *IEEE Transactions on Semiconductor Manufacturing*. **18** (1): 26–36. doi:10.1109/TSM.2004.841816 (<https://doi.org/10.1109%2FTSM.2004.841816>). ISSN 0894-6507 (<https://search.worldcat.org/issn/0894-6507>). S2CID 25283342 (<https://api.semanticscholar.org/CorpusID:25283342>). "In the field of electronics, the planar Si metal–oxide–semiconductor field-effect transistor (MOSFET) is perhaps the most important invention."
81. Kubozono, Yoshihiro; He, Xuexia; Hamao, Shino; Uesugi, Eri; Shimo, Yuma; Mikami, Takahiro; Goto, Hidenori; Kambe, Takashi (2015). "Application of Organic Semiconductors toward Transistors" (<https://books.google.com/books?id=8wdCCwAAQBAJ&pg=PA355>). *Nanodevices for Photonics and Electronics: Advances and Applications*. CRC Press. p. 355. ISBN 9789814613750.
82. "Triumph of the MOS Transistor" (<https://www.youtube.com/watch?v=q6fBEjf9WPw>). YouTube. Computer History Museum. August 6, 2010. Archived (<https://ghostarchive.org/archive/youtube/20211211/q6fBEjf9WPw>) from the original on December 11, 2021. Retrieved July 21, 2019.
83. "The most manufactured human artifact in history" (<https://computerhistory.org/blog/13-sextillion-counting-the-long-winding-road-to-the-most-frequently-manufactured-human-artifact-in-history/?key=13-sextillion-counting-the-long-winding-road-to-the-most-frequently-manufactured-human-artifact-in-history>). *Computer History*. April 2, 2018. Retrieved January 21, 2021.
84. FETs/MOSFETs: Smaller apps push up surface-mount supply (<https://web.archive.org/web/20081206043949/http://www.globalsources.com/gsol/I/FET-MOSFET/a/900000085806.htm>). globalsources.com (April 18, 2007)
85. "Introducing M1 Pro and M1 Max: the most powerful chips Apple has ever built - Apple" (<https://www.apple.com/newsroom/2021/10/introducing-m1-pro-and-m1-max-the-most-powerful-chips-apple-has-ever-built/>). www.apple.com. Retrieved October 20, 2022.
86. Deschamps, Jean-Pierre; Valderrama, Elena; Terés, Lluís (October 12, 2016). *Digital Systems: From Logic Gates to Processors* (<https://books.google.com/books?id=GBVADQAQBAJ&q=processor+logic+gates>). Springer. ISBN 978-3-319-41198-9.
87. Roland, James (August 1, 2016). *How Transistors Work* (<https://books.google.com/books?id=g2qsDAAAQBAJ&q=how+do+transistors+work>). Lerner Publications. ISBN 978-1-5124-2146-0.
88. Pulfrey, David L. (January 28, 2010). *Understanding Modern Transistors and Diodes* (<https://books.google.com/books?id=y9dYENs2SVUC&q=how+do+transistors+work>). Cambridge University Press. ISBN 978-1-139-48467-1.
89. Kaplan, Daniel (2003). *Hands-On Electronics*. pp. 47–54, 60–61. Bibcode:2003hoe..book.....K (<https://ui.adsabs.harvard.edu/abs/2003hoe..book.....K>). ISBN 978-0-511-07668-8.
90. "Transistor Base Resistor Calculator" (<https://kaizerpowerelectronics.dk/calculators/transistor-base-resistor-calculator/>). January 27, 2012.
91. van der Veen, M. (2005). "Universal system and output transformer for valve amplifiers" (http://www.mennovanderveen.nl/nl/download/download_3.pdf) (PDF). *118th AES Convention, Barcelona, Spain*. Archived (https://web.archive.org/web/20091229114846/http://www.mennovanderveen.nl/nl/download/download_3.pdf) (PDF) from the original on December 29, 2009.
92. "Transistor Example" (<http://www.bcae1.com/transres.htm>). Archived (<https://web.archive.org/web/20080208150020/http://www.bcae1.com/transres.htm>) from the original on February 8, 2008. 071003 bcae1.com

93. Gamyusenge, Aristide; Tran, Dung T.; Luo, Xuyi; Pitch, Gregory M.; Zhao, Yan; Jenkins, Kaelon A.; Dunn, Tim J.; Ayzner, Alexander L.; Savoie, Brett M.; Mei, Jianguo (December 7, 2018). "Semiconducting polymer blends that exhibit stable charge transport at high temperatures" (<https://doi.org/10.1126%2Fscience.aa0759>). *Science*. **362** (6419): 1131–1134. Bibcode:2018Sci...362.1131G (<https://ui.adsabs.harvard.edu/abs/2018Sci...362.1131G>). doi:10.1126/science.aa0759 (<https://doi.org/10.1126%2Fscience.aa0759>). ISSN 0036-8075 (<https://search.worldcat.org/issn/0036-8075>). PMID 30523104 (<https://pubmed.ncbi.nlm.nih.gov/30523104>).
94. Horowitz, Paul; Winfield Hill (1989). *The Art of Electronics* (2nd ed.). Cambridge University Press. p. [115]. ISBN 978-0-521-37095-0.
95. Sansen, W. M. C. (2006). *Analog design essentials*. New York, Berlin: Springer. p. §0152, p. 28. ISBN 978-0-387-25746-4.
96. "13 Sextillion & Counting: The Long & Winding Road to the Most Frequently Manufactured Human Artifact in History" (<https://www.computerhistory.org/atchm/13-sextillion-counting-the-long-winding-road-to-the-most-frequently-manufactured-human-artifact-in-history/>). *Computer History Museum*. April 2, 2018. Retrieved July 28, 2019.
97. Streetman, Ben (1992). *Solid State Electronic Devices*. Englewood Cliffs, NJ: Prentice-Hall. pp. 301–305. ISBN 978-0-13-822023-5.
98. "MOSFET DIFFERENTIAL AMPLIFIER" (<http://sites.bu.edu/engcourses/files/2016/08/mosfet-differential-amplifier.pdf>) (PDF). *Boston University*. Retrieved August 10, 2019.
99. "IGBT Module 5SNA 2400E170100" ([https://web.archive.org/web/20120426020121/http://library.abb.com/GLOBAL/SCOT/scot256.nsf/VerityDisplay/E700072B04381DD9C12571FF002D2CFE/\\$File/5SNA%202400E170100_5SYA1555-03Oct%2006.pdf](https://web.archive.org/web/20120426020121/http://library.abb.com/GLOBAL/SCOT/scot256.nsf/VerityDisplay/E700072B04381DD9C12571FF002D2CFE/$File/5SNA%202400E170100_5SYA1555-03Oct%2006.pdf)) (PDF). Archived from the original ([http://library.abb.com/GLOBAL/SCOT/scot256.nsf/VerityDisplay/E700072B04381DD9C12571FF002D2CFE/\\$File/5SNA%202400E170100_5SYA1555-03Oct%2006.pdf](http://library.abb.com/GLOBAL/SCOT/scot256.nsf/VerityDisplay/E700072B04381DD9C12571FF002D2CFE/$File/5SNA%202400E170100_5SYA1555-03Oct%2006.pdf)) (PDF) on April 26, 2012. Retrieved June 30, 2012.
100. Buonomo, S.; Ronsivalle, C.; Scollo, R.; STMicroelectronics; Musumeci, S.; Pagano, R.; Raciti, A.; University of Catania Italy (October 16, 2003). IEEE (ed.). *A new monolithic emitter-switching bipolar transistor (ESBT) in high-voltage converter applications*. 38th IAS annual Meeting on Conference Record of the Industry Applications Conference. Vol. 3 of 3. Salt Lake City. pp. 1810–1817. doi:10.1109/IAS.2003.1257745 (<https://doi.org/10.1109%2FIAS.2003.1257745>).
101. STMicroelectronics. "ESBTs" (<https://www.st.com/en/power-transistors/esbts.html?querycriteria=productId=SC1775>). www.st.com. Retrieved February 17, 2019. "ST no longer offers these components, this web page is empty, and datasheets are obsoletes"
102. Zhong Yuan Chang, Willy M. C. Sansen, *Low-Noise Wide-Band Amplifiers in Bipolar and CMOS Technologies*, page 31, Springer, 1991 ISBN 0792390962.
103. "Single Electron Transistors" (<https://web.archive.org/web/20120426015942/http://snow.stanford.edu/~shimbo/set.html>). [Snow.stanford.edu](http://snow.stanford.edu). Archived from the original (<http://snow.stanford.edu/~shimbo/set.html>) on April 26, 2012. Retrieved June 30, 2012.
104. Sanders, Robert (June 28, 2005). "Nanofluidic transistor, the basis of future chemical processors" (http://www.berkeley.edu/news/media/releases/2005/06/28_transistor.shtml). [Berkeley.edu](http://www.berkeley.edu). Archived (https://web.archive.org/web/20120702182324/http://www.berkeley.edu/news/media/releases/2005/06/28_transistor.shtml) from the original on July 2, 2012. Retrieved June 30, 2012.
105. "The return of the vacuum tube?" (<http://www.gizmag.com/nasa-vacuum-channel-transistor/22626/>). [Gizmag.com](http://www.gizmag.com). May 28, 2012. Archived (<https://web.archive.org/web/20160414122940/http://www.gizmag.com/nasa-vacuum-channel-transistor/22626/>) from the original on April 14, 2016. Retrieved May 1, 2016.
106. "New Type of Transistor from a Germanium–Tin Alloy Developed" (https://www_azom_com/news.aspx?newsID=61206). April 28, 2023.

107. "Timber! The World's First Wooden Transistor – IEEE Spectrum" (<https://spectrum.ieee.org/wood-transistor>).
108. "Boffins claim to create the world's first wooden transistor" (https://www.theregister.com/2023/05/01/wooden_transistor_sweden/).
109. "Paper Transistor – IEEE Spectrum" (<https://spectrum.ieee.org/paper-transistor>). *IEEE*.
110. "This Diamond Transistor Is Still Raw, But Its Future Looks Bright – IEEE Spectrum" (<https://spectrum.ieee.org/this-diamond-transistor-is-still-raw-but-its-future-looks-bright>). *IEEE*.
111. "The New, New Transistor – IEEE Spectrum" (<https://spectrum.ieee.org/aluminum-nitride>). *IEEE*.
112. Staff, The SE (February 23, 2024). "Chip Industry Week In Review" (<https://semiengineering.com/chip-industry-week-in-review-23/>). *Semiconductor Engineering*.
113. "Type Registration, Data Sheets" (<https://www.jedec.org/standards-documents/type-registration-data-sheets>). JEDEC. Retrieved March 7, 2025.
114. "Transistor Data" (<http://www.clivetec.0catch.com/Transistors.htm#JIS>). Clivetec.0catch.com. Archived (<https://web.archive.org/web/20160426190413/http://www.clivetec.0catch.com/Transistors.htm#JIS>) from the original on April 26, 2016. Retrieved May 1, 2016.
115. "Datasheet for BC549, with A, B and C gain groupings" (<http://www.fairchildsemi.com/ds/BC/BC549.pdf>) (PDF). Fairchild Semiconductor. Archived (<https://web.archive.org/web/20120407001013/http://www.fairchildsemi.com/ds/BC/BC549.pdf>) (PDF) from the original on April 7, 2012. Retrieved June 30, 2012.
116. "Datasheet for BUK854-800A (800volt IGBT)" (<http://www.datasheetcatalog.org/datasheet/philips/BUK854-800A.pdf>) (PDF). Archived (<https://web.archive.org/web/20120415132635/http://www.datasheetcatalog.org/datasheet/philips/BUK854-800A.pdf>) (PDF) from the original on April 15, 2012. Retrieved June 30, 2012.
117. "Richard Freeman's HP Part numbers Crossreference" (<http://www.hpmuseum.org/cgi-sys/cgiwrap/hpmuseum/archv010.cgi?read=27258>). Hpmuseum.org. Archived (<https://web.archive.org/web/20120605183505/http://www.hpmuseum.org/cgi-sys/cgiwrap/hpmuseum/archv010.cgi?read=27258>) from the original on June 5, 2012. Retrieved June 30, 2012.
118. "Transistor–Diode Cross Reference – H.P. Part Numbers to JEDEC (pdf)" (<http://www.sphere.bc.ca/test/hp-parts/300-hpxref.pdf>) (PDF). Archived (<https://web.archive.org/web/20160508135527/http://www.sphere.bc.ca/test/hp-parts/300-hpxref.pdf>) (PDF) from the original on May 8, 2016. Retrieved May 1, 2016.
119. "CV Device Cross-reference by Andy Lake" (http://www.qsl.net/g8yoa/cv_table.html). Qsl.net. Archived (https://web.archive.org/web/20120121111531/http://www.qsl.net/g8yoa/cv_table.html) from the original on January 21, 2012. Retrieved June 30, 2012.
120. Sedra, A.S. & Smith, K.C. (2004). *Microelectronic circuits* (https://archive.org/details/microelectronicc00sedr_571) (Fifth ed.). New York: Oxford University Press. p. 397 (https://archive.org/details/microelectronicc00sedr_571/page/n426) and Figure 5.17. ISBN 978-0-19-514251-8.
121. Greig, William (April 24, 2007). *Integrated Circuit Packaging, Assembly and Interconnections* (<https://books.google.com/books?id=g29xoTAyAMgC&q=hybrid+&pg=PA63>). Springer. p. 63. ISBN 9780387339139. "A hybrid circuit is defined as an assembly containing both active semiconductor devices (packaged and unpackaged)"
122. Rojas, Jhonathan P.; Torres Sevilla, Galo A.; Hussain, Muhammad M. (2013). "Can We Build a Truly High Performance Computer Which is Flexible and Transparent?" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3767948>). *Scientific Reports*. **3**: 2609. Bibcode:2013NatSR...3.2609R (<https://ui.adsabs.harvard.edu/abs/2013NatSR...3.2609R>). doi:10.1038/srep02609 (<https://doi.org/10.1038%2Fsrep02609>). PMC 3767948 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3767948>). PMID 24018904 (<https://pubmed.ncbi.nlm.nih.gov/24018904>).

123. Zhang, Kan; Seo, Jung-Hun; Zhou, Weidong; Ma, Zhenqiang (2012). "Fast flexible electronics using transferrable [sic] silicon nanomembranes". *Journal of Physics D: Applied Physics*. **45** (14): 143001. Bibcode:2012JPhD...45n3001Z (<https://ui.adsabs.harvard.edu/abs/2012JPhD...45n3001Z>). doi:10.1088/0022-3727/45/14/143001 (<https://doi.org/10.1088%2F0022-3727%2F45%2F14%2F143001>). S2CID 109292175 (<https://api.semanticscholar.org/CorpusID:109292175>).
124. Sun, Dong-Ming; Timmermans, Marina Y.; Tian, Ying; Nasibulin, Albert G.; Kauppinen, Esko I.; Kishimoto, Shigeru; Mizutani, Takashi; Ohno, Yutaka (2011). "Flexible high-performance carbon nanotube integrated circuits". *Nature Nanotechnology*. **6** (3): 156–61. Bibcode:2011NatNa...6..156S (<https://ui.adsabs.harvard.edu/abs/2011NatNa...6..156S>). doi:10.1038/NNANO.2011.1 (<https://doi.org/10.1038%2FNNANO.2011.1>). PMID 21297625 (<https://pubmed.ncbi.nlm.nih.gov/21297625>). S2CID 205446925 (<https://api.semanticscholar.org/CorpusID:205446925>).

Further reading

Books

- Horowitz, Paul & Hill, Winfield (2015). *The Art of Electronics* (3 ed.). Cambridge University Press. ISBN 978-0521809269.
- Amos SW, James MR (1999). *Principles of Transistor Circuits*. Butterworth-Heinemann. ISBN 978-0-7506-4427-3.
- Riordan, Michael & Hoddeson, Lillian (1998). *Crystal Fire*. W.W Norton & Company Limited. ISBN 978-0-393-31851-7. The invention of the transistor & the birth of the information age
- Warnes, Lionel (1998). *Analogue and Digital Electronics*. Macmillan Press Ltd. ISBN 978-0-333-65820-8.
- *The Power Transistor - Temperature and Heat Transfer*; 1st Ed; John McWane, Dana Roberts, Malcom Smith; McGraw-Hill; 82 pages; 1975; ISBN 978-0-07-001729-0. (archive) (<https://archive.org/details/ThePowerTransistor/>)
- *Transistor Circuit Analysis - Theory and Solutions to 235 Problems*; 2nd Ed; Alfred Gronner; Simon and Schuster; 244 pages; 1970. (archive) (<https://archive.org/details/TransistorCircuitAnalysis/>)
- *Transistor Physics and Circuits*; R.L. Riddle and M.P. Ristenbatt; Prentice-Hall; 1957.

Periodicals

- Michael Riordan (2005). "How Europe Missed the Transistor" (<https://web.archive.org/web/20080214002109/http://www.spectrum.ieee.org/print/2155>). *IEEE Spectrum*. **42** (11): 52–57. doi:10.1109/MSPEC.2005.1526906 (<https://doi.org/10.1109%2FMSPEC.2005.1526906>). S2CID 34953819 (<https://api.semanticscholar.org/CorpusID:34953819>). Archived from the original (<http://spectrum.ieee.org/print/2155>) on February 14, 2008.
- "Herbert F. Mataré, An Inventor of the Transistor has his moment" (<https://web.archive.org/web/20090623050755/http://www.mindfully.org/Technology/2003/Transistor-Matare-Inventor24feb03.htm>). *The New York Times*. February 24, 2003. Archived from the original (<http://www.mindfully.org/Technology/2003/Transistor-Matare-Inventor24feb03.htm>) on June 23, 2009.
- Bacon, W. Stevenson (1968). "The Transistor's 20th Anniversary: How Germanium And A Bit of Wire Changed The World" (<https://books.google.com/books?id=mykDAAAAMBAJ>). *Popular Science*. **192** (6): 80–84. ISSN 0161-7370 (<https://search.worldcat.org/issn/0161-7370>).

Databooks

- Discrete Databook (https://archive.org/details/bitsavers_fairchildDiscreteDataBook_3512751); 1985; Fairchild (now ON Semiconductor)
- Small-Signal Semiconductors Databook (https://archive.org/details/bitsavers_motorolaSmallSignalSemiconductors_75896318/), 1987; Motorola (now ON semiconductor)
- Discrete Power Devices Databook (https://archive.org/details/bitsavers_sgsdataBooDevices5ed_46325378); 1982; SGS (now STMicroelectronics)
- Discrete Databook (<https://archive.org/details/NationalSemiconductor-DiscreteDatabook1978OCR>); 1978; National Semiconductor (now Texas Instruments)

External links

- BBC: Building the digital age (<http://news.bbc.co.uk/2/hi/technology/7091190.stm>) photo history of transistors
- The Bell Systems Memorial on Transistors (https://web.archive.org/web/20070928041118/https://www.porticus.org/bell/belllabs_transistor.html)
- IEEE Global History Network, *The Transistor and Portable Electronics* (https://www.ieeeghn.org/wiki/index.php/The_Transistor_and_Portable_Electronics). All about the history of transistors and integrated circuits.
- *This Month in Physics History: November 17 to December 23, 1947: Invention of the First Transistor* (<https://www.aps.org/publications/apsnews/200011/history.cfm>). From the American Physical Society
- Transistor | Definition & Uses | Britannica (<https://www.britannica.com/technology/transistor>) "Transistor" at *Encyclopædia Britannica*

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Transistor&oldid=1291367439>"

Resistor

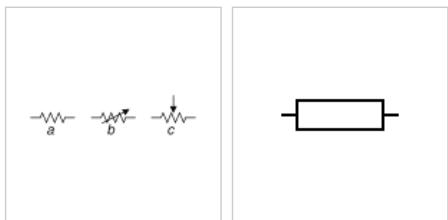
A **resistor** is a passive two-terminal electronic component that implements electrical resistance as a circuit element. In electronic circuits, resistors are used to reduce current flow, adjust signal levels, to divide voltages, bias active elements, and terminate transmission lines, among other uses. High-power resistors that can dissipate many watts of electrical power as heat may be used as part of motor controls, in power distribution systems, or as test loads for generators. Fixed resistors have resistances that only change slightly with temperature, time or operating voltage. Variable resistors can be used to adjust circuit elements (such as a volume control or a lamp dimmer), or as sensing devices for heat, light, humidity, force, or chemical activity.

Resistors are common elements of electrical networks and electronic circuits and are ubiquitous in electronic equipment. Practical resistors as discrete components can be composed of various compounds and forms. Resistors are also implemented within integrated circuits.

The electrical function of a resistor is specified by its resistance: common commercial resistors are manufactured over a range of more than nine orders of magnitude. The nominal value of the resistance falls within the manufacturing tolerance, indicated on the component.

Electronic symbols and notation

Two typical schematic diagram symbols are as follows:



ANSI-style:

- (a) resistor,
- (b) rheostat (variable resistor), and
- (c) potentiometer

IEC resistor symbol

Resistor

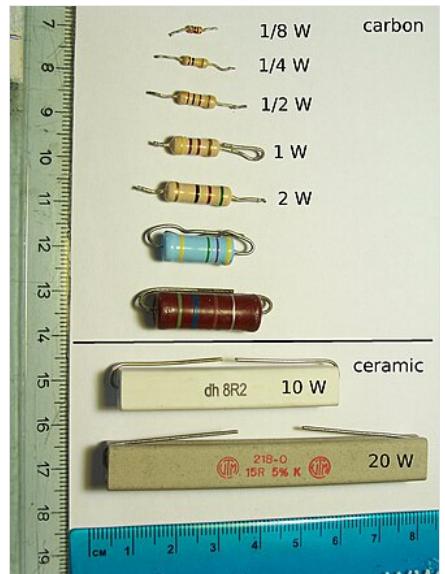
An array of axial-lead resistors

Component type	Passive
Working principle	Electrical resistance
Number of terminals	2
Electronic symbol	

ANSI and IEC symbols

The notation to state a resistor's value in a circuit diagram varies.

One common scheme is the RKM code following IEC 60062. Rather than using a decimal separator, this notation uses a letter loosely associated with SI prefixes corresponding with the part's resistance. For example, $8K2$ as part marking code, in a circuit diagram or in a bill of materials (BOM) indicates a resistor value of $8.2 \text{ k}\Omega$. Additional zeros imply a tighter tolerance, for example $15M0$ for three significant digits. When the value can be expressed without the need for a prefix (that is, multiplicator 1), an "R" is used instead of the decimal separator. For example, $1R2$ indicates 1.2Ω , and $18R$ indicates 18Ω .



Various resistor types of different shapes and sizes

Theory of operation

Ohm's law

An ideal resistor (i.e. a resistance without reactance) obeys Ohm's law:

$$V = I \cdot R.$$

Ohm's law states that the voltage (V) across a resistor is proportional to the current (I) passing through it, where the constant of proportionality is the resistance (R). For example, if a 300-ohm resistor is attached across the terminals of a 12-volt battery, then a current of $12 / 300 = 0.04$ amperes flows through that resistor.

The ohm (symbol: Ω) is the SI unit of electrical resistance, named after Georg Simon Ohm. An ohm is equivalent to a volt per ampere. Since resistors are specified and manufactured over a very large range of values, the derived units of milliohm ($1 \text{ m}\Omega = 10^{-3} \Omega$), kilohm ($1 \text{ k}\Omega = 10^3 \Omega$), and megohm ($1 \text{ M}\Omega = 10^6 \Omega$) are also in common usage.^{[2][3]:p.20}

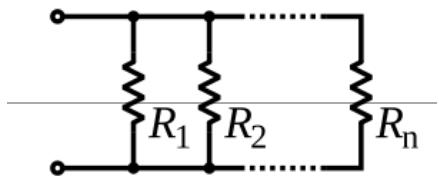
Series and parallel resistors

The total resistance of resistors connected in series is the sum of their individual resistance values.



$$R_{\text{eq}} = \sum_{i=1}^n R_i = R_1 + R_2 + \dots + R_n.$$

The total resistance of resistors connected in parallel is the reciprocal of the sum of the reciprocals of the individual resistors.^{[3]:p.20ff}



$$R_{\text{eq}} = \left(\sum_{i=1}^n \frac{1}{R_i} \right)^{-1} = \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \dots + \frac{1}{R_n} \right)^{-1}$$

For example, a 10 ohm resistor connected in parallel with a 5 ohm resistor and a 15 ohm resistor produces $\frac{1}{1/10 + 1/5 + 1/15}$ ohms of resistance, or $\frac{30}{11} = 2.727$ ohms.

A resistor network that is a combination of parallel and series connections can be broken up into smaller parts that are either one or the other. Some complex networks of resistors cannot be resolved in this manner, requiring more sophisticated circuit analysis. Generally, the Y-Δ transform, or matrix methods can be used to solve such problems.^{[4][5][6]}

Power dissipation

At any instant, the power P (watts) consumed by a resistor of resistance R (ohms) is calculated as:

$$P = IV = I^2 R = \frac{V^2}{R}$$

where V (volts) is the voltage across the resistor and I (amps) is the current flowing through it. Using Ohm's law, the two other forms can be derived. This power is converted into heat which must be dissipated by the resistor's package before its temperature rises excessively.^{[3]:p.22}

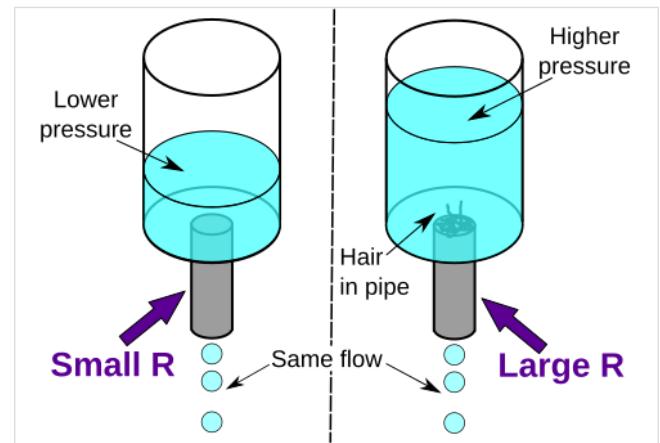
Resistors are rated according to their maximum power dissipation. Discrete resistors in solid-state electronic systems are typically rated as $\frac{1}{10}$, $\frac{1}{8}$, or $\frac{1}{4}$ watt. They usually absorb much less than a watt of electrical power and require little attention to their power rating.

Power resistors are required to dissipate substantial amounts of power and are typically used in power supplies, power conversion circuits, and power amplifiers; this designation is loosely applied to resistors with power ratings of 1 watt or greater. Power resistors are physically larger and may not use the preferred values, color codes, and external packages described below.

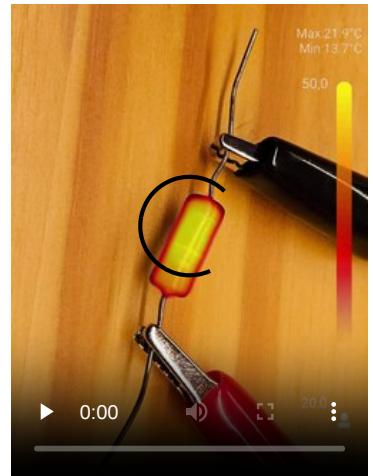
If the average power dissipated by a resistor is more than its power rating, damage to the resistor may occur, permanently altering its resistance; this is distinct from the reversible change in resistance due to its temperature coefficient when it warms. Excessive power dissipation may raise the temperature of the resistor to a point where it can burn the circuit board or adjacent components, or even cause a fire. There are flameproof resistors that will not produce flames with any overload of any duration.

Resistors may be specified with higher rated dissipation than is experienced in service to account for poor air circulation, high altitude, or high operating temperature.

All resistors have a maximum voltage rating; this may limit the power dissipation for higher resistance values.^[7] For instance, among $\frac{1}{4}$ watt resistors (a very common sort of leaded resistor) one is listed with a resistance of $100 \text{ M}\Omega$ ^[8] and a maximum rated voltage of 750 V. However even placing 750 V across a $100 \text{ M}\Omega$ resistor continuously would only result in a power dissipation of less than 6 mW, making the nominal $\frac{1}{4}$ watt rating



The hydraulic analogy compares electric current flowing through circuits to water flowing through pipes. When a pipe (left) is clogged with hair (right), it takes a larger pressure to achieve the same flow of water. Pushing electric current through a large resistance is like pushing water through a pipe clogged with hair: It requires a larger push (voltage) to drive the same flow (electric current).^[1]



Resistor warming caused by electrical current captured by thermal camera

meaningless.

Nonideal properties

Practical resistors have a series inductance and a small parallel capacitance; these specifications can be important in high-frequency applications. And while even an ideal resistor inherently has Johnson noise, some resistors have worse noise characteristics and so may be an issue for low-noise amplifiers or other sensitive electronics.

In some precision applications, the temperature coefficient of the resistance may also be of concern.

The unwanted inductance, excess noise, and temperature coefficient are mainly dependent on the technology used in manufacturing the resistor. They are not normally specified individually for a particular family of resistors manufactured using a particular technology.^[9] A family of discrete resistors may also be characterized according to its form factor, that is, the size of the device and the position of its leads (or terminals). This is relevant in the practical manufacturing of circuits that may use them.

Practical resistors are also specified as having a maximum power rating which must exceed the anticipated power dissipation of that resistor in a particular circuit: this is mainly of concern in power electronics applications. Resistors with higher power ratings are physically larger and may require heat sinks. In a high-voltage circuit, attention must sometimes be paid to the rated maximum working voltage of the resistor. While there is no minimum working voltage for a given resistor, failure to account for a resistor's maximum rating may cause the resistor to incinerate when current is run through it.

Fixed resistors

Lead arrangements

Through-hole components typically have "leads" (pronounced /li:dz/) leaving the body "axially", that is, on a line parallel with the part's longest axis. Others have leads coming off their body "radially" instead. Other components may be SMT (surface mount technology), while high power resistors may have one of their leads designed into the heat sink.

Carbon composition

Carbon composition resistors (CCR) consist of a solid cylindrical resistive element with embedded wire leads or metal end caps to which the lead wires are attached. The body of the resistor is protected with paint or plastic. Early 20th-century carbon composition resistors had uninsulated bodies; the lead wires were wrapped around the ends of the resistance element rod and soldered. The completed resistor was painted for color-coding of its value.

The resistive element in carbon composition resistors is made from a mixture of finely powdered carbon and an insulating material, usually ceramic. A resin holds the mixture together. The resistance is determined by the ratio of the fill material (the powdered ceramic) to the carbon. Higher concentrations of carbon, which is a good conductor, result in lower resistances. Carbon composition resistors were commonly used in the 1960s and earlier, but are not popular for general use now as other types have better specifications, such as tolerance, voltage dependence, and stress. Carbon composition resistors change value when stressed with over-voltages. Moreover, if internal moisture content, such as from exposure for some length of time to a humid environment, is significant, soldering heat creates a non-reversible change in resistance value. Carbon composition resistors have poor stability with time and were consequently factory sorted to, at best, only 5% tolerance.^[10] These resistors are non-inductive, which provides benefits when used in voltage pulse reduction and surge protection applications.^[11] Carbon composition resistors have higher capability to withstand overload relative to the component's size.^[12]

Carbon composition resistors are still available, but relatively expensive. Values ranged from fractions of an ohm to 22 megohms. Due to their high price, these resistors are no longer used in most applications. However, they are used in power supplies and welding controls.^[12] They are also in demand for repair of vintage electronic equipment where authenticity is a factor.



An aluminium-encased power resistor rated for dissipation of 50 W when mounted on a heat-sink



VZR power resistor 1.5 kΩ 12 W, manufactured in 1963 in the Soviet Union



A single in line (SIL) resistor package with 8 individual 47 ohm resistors. This package is also known as a SIP-9. One end of each resistor is connected to a separate pin and the other ends are all connected together to the remaining (common) pin – pin 1, at the end identified by the white dot.



Axial resistors with wire leads for through-hole mounting



Old style "dog bone" resistors with "body tip, dot" color code marking

Carbon pile

A carbon pile resistor is made of a stack of carbon disks compressed between two metal contact plates. Adjusting the clamping pressure changes the resistance between the plates. These resistors are used when an adjustable load is required, such as in testing automotive batteries or radio transmitters. A carbon pile resistor can also be used as a speed control for small motors in household appliances (sewing machines, hand-held mixers) with ratings up to a few hundred watts.^[13] A carbon pile resistor can be incorporated in automatic voltage regulators for generators, where the carbon pile controls the field current to maintain relatively constant voltage.^[14] This principle is also applied in the [carbon microphone](#).



Three carbon composition resistors in a 1960s [valve](#) (vacuum tube) radio

Carbon film

In manufacturing carbon film resistors, a carbon film is deposited on an insulating substrate, and a [helix](#) is cut in it to create a long, narrow resistive path. Varying shapes, coupled with the [resistivity of amorphous](#) carbon (ranging from 500 to 800 $\mu\Omega \text{ m}$), can provide a wide range of resistance values. Carbon film resistors feature lower noise compared to carbon composition resistors because of the precise distribution of the pure graphite without binding.^[15] Carbon film resistors feature a power rating range of 0.125 W to 5 W at 70 °C. Resistances available range from 1 ohm to 10 megaohm. The carbon film resistor has an [operating temperature](#) range of -55 °C to 155 °C. It has 200 to 600 volts maximum working voltage range. Special carbon film resistors are used in applications requiring high pulse stability.^[12]



Carbon film resistor with exposed carbon spiral (Tesla TR-212 1 k Ω)

Printed carbon resistors

Carbon composition resistors can be printed directly onto [printed circuit board](#) (PCB) substrates as part of the [PCB manufacturing](#) process. Although this technique is more common on hybrid PCB modules, it can also be used on standard fibreglass PCBs. Tolerances are typically quite large and can be in the order of 30%. A typical application would be non-critical [pull-up resistors](#).



Carbon resistors (black rectangles) printed directly onto the SMD pads on the PCB of a [Psion Organiser II](#) from 1989

Thick and thin film

Thick film resistors became popular during the 1970s, and most [SMD](#) (surface mount device) resistors today are of this type. The resistive element of thick films is 1000 times thicker than thin films,^[16] but the principal difference is how the film is applied to the cylinder (axial resistors) or the surface (SMD resistors).

Thin film resistors are made by [sputtering](#) (a method of [vacuum deposition](#)) the resistive material onto an insulating substrate. The film is then etched in a similar manner to the old (subtractive) process for making printed circuit boards; that is, the surface is coated with a [photo-sensitive material](#), covered by a pattern film, irradiated with [ultraviolet](#) light, and then the exposed photo-sensitive coating is developed, and underlying thin film is etched away.

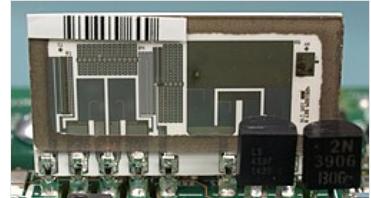
Thick film resistors are manufactured using screen and stencil printing processes.^[12]

Because the time during which the sputtering is performed can be controlled, the thickness of the thin film can be accurately controlled. The type of material also varies, consisting of one or more ceramic ([cermet](#)) conductors such as [tantalum nitride](#) (TaN), [ruthenium oxide](#) (RuO₂), [lead oxide](#) (PbO), [bismuth ruthenate](#) (Bi₂Ru₂O₇), [nickel chromium](#) (NiCr), or [bismuth iridate](#) (Bi₂Ir₂O₇).

The resistance of both thin and thick film resistors after manufacture is not highly accurate; they are usually trimmed to an accurate value by abrasive or [laser trimming](#). Thin film resistors are usually specified with tolerances of 1% and 5%, and with temperature coefficients of 5 to 50 ppm/K. They also have much lower [noise](#) levels, on the level of 10–100 times less than thick film resistors.^[17] Thick film resistors may use the same conductive ceramics, but they are mixed with [sintered](#) (powdered) glass and a carrier liquid so that the composite can be [screen-printed](#). This composite of glass and conductive ceramic (cermet) material is then fused (baked) in an oven at about 850 °C.

When first manufactured, thick film resistors had tolerances of 5%, but standard tolerances have improved to 2% or 1% in the last few decades. Temperature coefficients of thick film resistors are typically ± 200 or ± 250 ppm/K; a 40-[kelvin](#) (70 °F) temperature change can change the resistance by 1%.

Thin film resistors are usually far more expensive than thick film resistors. For example, SMD thin film resistors, with 0.5% tolerances and with 25 ppm/K temperature coefficients, when bought in full size reel quantities, are about twice the cost of 1%, 250 ppm/K thick film resistors.



Laser Trimmed Precision Thin Film Resistor Network from Fluke, used in the [Keithley DMM7510](#) multimeter. Ceramic backed with glass hermetic seal cover.

Metal film

A common type of axial-leaded resistor today is the metal-film resistor. Metal Electrode Leadless Face ([MELF](#)) resistors often use the same technology.

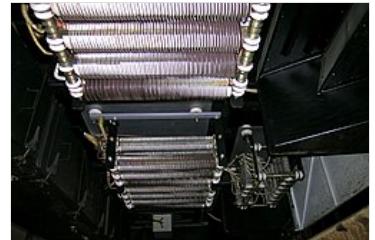
Metal film resistors are usually coated with nickel chromium (NiCr), but might be coated with any of the cermet materials listed above for thin film resistors. Unlike thin film resistors, the material may be applied using different techniques than sputtering (though this is one technique used). The resistance value is determined by cutting a helix through the coating rather than by etching, similar to the way carbon resistors are made. The result is a reasonable tolerance (0.5%, 1%, or 2%) and a temperature coefficient that is generally between 50 and 100 ppm/K.^[18] Metal film resistors possess good noise characteristics and low non-linearity due to a low voltage coefficient. They are also beneficial due to long-term stability.^[12]

Metal oxide film

Metal-oxide film resistors are made of metal oxides which results in a higher operating temperature and greater stability and reliability than metal film. They are used in applications with high endurance demands.

Wire wound

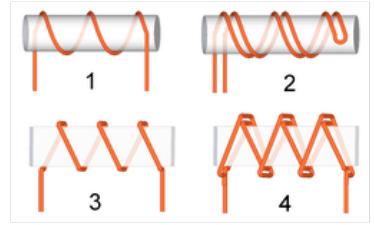
Wirewound resistors are commonly made by winding a metal wire, usually nichrome, around a ceramic, plastic, or fiberglass core. The ends of the wire are soldered or welded to two caps or rings, attached to the ends of the core. The assembly is protected with a layer of paint, molded plastic, or an enamel coating baked at high temperature. These resistors are designed to withstand unusually high temperatures of up to 450 °C.^[12] Wire leads in low power wirewound resistors are usually between 0.6 and 0.8 mm in diameter and tinned for ease of soldering. For higher power wirewound resistors, either a ceramic outer case or an aluminum outer case on top of an insulating layer is used. If the outer case is ceramic, such resistors are sometimes described as "cement" resistors, though they do not actually contain any traditional cement. The aluminum-cased types are designed to be attached to a heat sink to dissipate the heat; the rated power is dependent on being used with a suitable heat sink, e.g., a 50 W power rated resistor overheats at a fraction of the power dissipation if not used with a heat sink. Large wirewound resistors may be rated for 1,000 watts or more.



High-power wire wound resistors used for dynamic braking on an electric railway car. Such resistors may dissipate many kilowatts for an extended length of time.

Because wirewound resistors are coils they have more undesirable inductance than other types of resistor. However, winding the wire in sections with alternately reversed direction can minimize inductance. Other techniques employ bifilar winding, or a flat thin former (to reduce cross-section area of the coil). For the most demanding circuits, resistors with Ayrton–Perry winding are used.

Applications of wirewound resistors are similar to those of composition resistors with the exception of high frequency applications. The high frequency response of wirewound resistors is substantially worse than that of a composition resistor.^[12]



Types of windings in wire resistors:

1. common
2. bifilar
3. common on a thin former
4. Ayrton–Perry

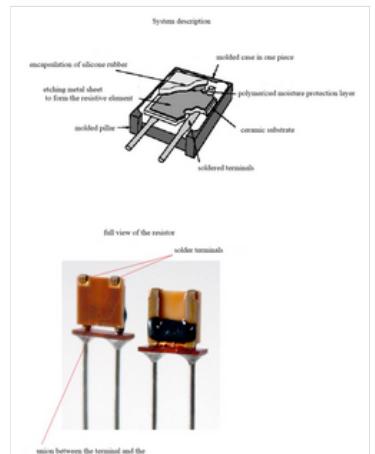
Metal foil resistor

In 1960, Felix Zandman and Sidney J. Stein^[19] presented a development of resistor film of very high stability.

The primary resistance element of a foil resistor is a chromium nickel alloy foil several micrometers thick. Chromium nickel alloys are characterized by having a large electrical resistance (about 58 times that of copper), a small temperature coefficient and high resistance to oxidation. Examples are Chromel A and Nichrome V, whose typical composition is 80 Ni and 20 Cr, with a melting point of 1420 °C. When iron is added, the chromium nickel alloy becomes more ductile. The Nichrome and Chromel C are examples of an alloy containing iron. The composition typical of Nichrome is 60 Ni, 12 Cr, 26 Fe, 2 Mn and Chromel C, 64 Ni, 11 Cr, Fe 25. The melting temperature of these alloys are 1350 °C and 1390 °C, respectively.^[20]

Since their introduction in the 1960s, foil resistors have had the best precision and stability of any resistor available. One of the important parameters of stability is the temperature coefficient of resistance (TCR). The TCR of foil resistors is extremely low, and has been further improved over the years. One range of ultra-precision foil resistors offers a TCR of 0.14 ppm/°C, tolerance ±0.005%, long-term stability (1 year) 25 ppm, (3 years) 50 ppm (further improved 5-fold by hermetic sealing), stability under load (2000 hours) 0.03%, thermal EMF 0.1 µV/°C, noise –42 dB, voltage coefficient 0.1 ppm/V, inductance 0.08 µH, capacitance 0.5 pF.^[21]

The thermal stability of this type of resistor also has to do with the opposing effects of the metal's electrical resistance increasing with temperature, and being reduced by thermal expansion leading to an increase in thickness of the foil, whose other dimensions are constrained by a ceramic substrate.



Metal foil resistor

Ammeter shunts

An ammeter shunt is a special type of current-sensing resistor, having four terminals and a value in milliohms or even micro-ohms. Current-measuring instruments, by themselves, can usually accept only limited currents. To measure high currents, the current passes through the shunt across which the voltage drop is measured and interpreted as current. A typical shunt consists of two solid metal blocks, sometimes brass, mounted on an insulating base. Between the blocks, and soldered or brazed to them, are one or more strips of low temperature coefficient of resistance (TCR)

manganin alloy. Large bolts threaded into the blocks make the current connections, while much smaller screws provide volt meter connections. Shunts are rated by full-scale current, and often have a voltage drop of 50 mV at rated current. Such meters are adapted to the shunt full current rating by using an appropriately marked dial face; no change need to be made to the other parts of the meter.

Grid resistor

In heavy-duty industrial high-current applications, a grid resistor is a large convection-cooled lattice of stamped metal alloy strips connected in rows between two electrodes. Such industrial grade resistors can be as large as a refrigerator; some designs can handle over 500 amperes of current, with a range of resistances extending lower than 0.04 ohms. They are used in applications such as dynamic braking and load banking for locomotives and trams, neutral grounding for industrial AC distribution, control loads for cranes and heavy equipment, load testing of generators and harmonic filtering for electric substations.^[22]

The term *grid resistor* is sometimes used to describe a resistor of any type connected to the control grid of a vacuum tube. This is not a resistor technology; it is an electronic circuit topology.

Special varieties

- Cermet
- Phenolic
- Tantalum
- Water resistor

Variable resistors

Adjustable resistors

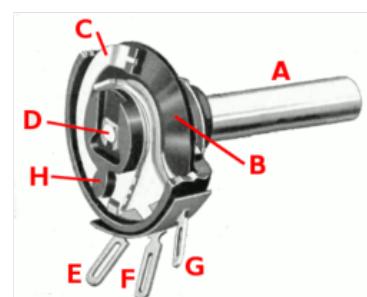
A resistor may have one or more fixed tapping points so that the resistance can be changed by moving the connecting wires to different terminals. Some wirewound power resistors have a tapping point that can slide along the resistance element, allowing a larger or smaller part of the resistance to be used.

Where continuous adjustment of the resistance value during operation of equipment is required, the sliding resistance tap can be connected to a knob accessible to an operator. Such a device is called a rheostat and has two terminals.

Potentiometers

A potentiometer (colloquially, *pot*) is a three-terminal resistor with a continuously adjustable tapping point controlled by rotation of a shaft or knob or by a linear slider.^[23] The name *potentiometer* comes from its function as an adjustable voltage divider to provide a variable potential at the terminal connected to the tapping point. Volume control in an audio device is a common application of a potentiometer. A typical low power potentiometer (see drawing) is constructed of a flat resistance element (**B**) of carbon composition, metal film, or conductive plastic, with a springy phosphor bronze wiper contact (**C**) which moves along the surface. An alternate construction is resistance wire wound on a form, with the wiper sliding axially along the coil.^[23] These have lower resolution, since as the wiper moves the resistance changes in steps equal to the resistance of a single turn.^[23]

High-resolution multturn potentiometers are used in precision applications. These have wire-wound resistance elements typically wound on a helical mandrel, with the wiper moving on a helical track as the control is turned, making continuous contact with the wire. Some include a conductive-plastic resistance coating over the wire to improve resolution. These typically offer ten turns of their shafts to cover their full range. They are usually set with dials that include a simple turns counter and a graduated dial, and can typically achieve three-digit resolution. Electronic analog computers used them in quantity for setting coefficients and delayed-sweep oscilloscopes of recent decades included one on their panels.



Potentiometer with case cut away, showing parts: (A) shaft, (B) stationary carbon composition resistance element, (C) phosphor bronze wiper, (D) shaft attached to wiper, (E, G) terminals connected to ends of resistance element, (F) terminal connected to wiper.



Typical panel mount potentiometer



An assortment of small through-hole potentiometers designed for mounting on printed circuit boards.

Resistance decade boxes

A resistance decade box or resistor substitution box is a unit containing resistors of many values, with one or more mechanical switches which allow any one of various discrete resistances offered by the box to be dialed in. Usually the resistance is accurate to high precision, ranging from laboratory/calibration grade accuracy of 20 parts per million, to field grade at 1%. Inexpensive boxes with lesser accuracy are also available. All types offer a convenient way of selecting and quickly changing a resistance in laboratory, experimental and development work without needing to attach resistors one by one, or even stock each value. The range of resistance provided, the maximum resolution, and the accuracy characterize the box. For example, one box offers resistances from 0 to 100 megohms, maximum resolution 0.1 ohm, accuracy 0.1%.^[24]



Resistance decade box

Special devices

There are various devices whose resistance changes with various quantities. The resistance of NTC thermistors exhibit a strong negative temperature coefficient, making them useful for measuring temperatures. Since their resistance can be large until they are allowed to heat up due to the passage of current, they are also commonly used to prevent excessive current surges when equipment is powered on. Similarly, the resistance of a humistor varies with humidity. One sort of photodetector, the photoresistor, has a resistance which varies with illumination.

The strain gauge, invented by Edward E. Simmons and Arthur C. Ruge in 1938, is a type of resistor that changes value with applied strain. A single resistor may be used, or a pair (half bridge), or four resistors connected in a Wheatstone bridge configuration. The strain resistor is bonded with adhesive to an object that is subjected to mechanical strain. With the strain gauge and a filter, amplifier, and analog/digital converter, the strain on an object can be measured.

A related but more recent invention uses a Quantum Tunnelling Composite to sense mechanical stress. It passes a current whose magnitude can vary by a factor of 10^{12} in response to changes in applied pressure.

Measurement

The value of a resistor can be measured with an ohmmeter, which may be one function of a multimeter. Usually, probes on the ends of test leads connect to the resistor. A simple ohmmeter may apply a voltage from a battery across the unknown resistor (with an internal resistor of a known value in series) producing a current which drives a meter movement. The current, in accordance with Ohm's law, is inversely proportional to the sum of the internal resistance and the resistor being tested, resulting in an analog meter scale which is very non-linear, calibrated from infinity to 0 ohms. A digital multimeter, using active electronics, may instead pass a specified current through the test resistance. The voltage generated across the test resistance in that case is linearly proportional to its resistance, which is measured and displayed. In either case the low-resistance ranges of the meter pass much more current through the test leads than do high-resistance ranges. This allows for the voltages present to be at reasonable levels (generally below 10 volts) but still measurable.

Measuring low-value resistors, such as fractional-ohm resistors, with acceptable accuracy requires four-terminal connections. One pair of terminals applies a known, calibrated current to the resistor, while the other pair senses the voltage drop across the resistor. Some laboratory quality ohmmeters, milliohmmeters, and even some of the better digital multimeters sense using four input terminals for this purpose, which may be used with special test leads called Kelvin clips. Each of the two clips has a pair of jaws insulated from each other. One side of each clip applies the measuring current, while the other connections are only to sense the voltage drop. The resistance is again calculated using Ohm's Law as the measured voltage divided by the applied current.

Standards

Production resistors

Resistor characteristics are quantified and reported using various national standards. In the US, MIL-STD-202^[25] contains the relevant test methods to which other standards refer.

There are various standards specifying properties of resistors for use in equipment:

- IEC 60062 (IEC 62) / DIN 40825 / BS 1852 / IS 8186 / JIS C 5062 etc. (Resistor color code, RKM code, date code)
- EIA RS-279 / DIN 41429 (Resistor color code)
- IEC 60063 (IEC 63) / JIS C 5063 (Standard E series values)
- MIL-PRF-26
- MIL-PRF-39007 (Fixed power, established reliability)
- MIL-PRF-55342 (Surface-mount thick and thin film)
- MIL-PRF-914
- MIL-R-11 Standard Canceled
- MIL-R-39017 (Fixed, General Purpose, Established Reliability)
- MIL-PRF-32159 (zero ohm jumpers)

- UL 1412 (fusing and temperature limited resistors)^[26]

There are other United States military procurement MIL-R- standards.

Resistance standards

The primary standard for resistance, the "mercury ohm" was initially defined in 1884 in as a column of mercury 106.3 cm long and 1 square millimeter in cross-section, at 0 degrees Celsius. Difficulties in precisely measuring the physical constants to replicate this standard result in variations of as much as 30 ppm. From 1900 the mercury ohm was replaced with a precision machined plate of manganin.^[27] Since 1990 the international resistance standard has been based on the quantized Hall effect discovered by Klaus von Klitzing, for which he won the Nobel Prize in Physics in 1985.^[28]

Resistors of extremely high precision are manufactured for calibration and laboratory use. They may have four terminals, using one pair to carry an operating current and the other pair to measure the voltage drop; this eliminates errors caused by voltage drops across the lead resistances, because no charge flows through voltage sensing leads. It is important in small value resistors (100–0.0001 ohm) where lead resistance is significant or even comparable with respect to resistance standard value.^[29]

Resistor marking

Axial resistor cases are usually tan, brown, blue, or green (though other colors are occasionally found as well, such as dark red or dark gray), and display three to six colored stripes that indicate resistance (and by extension tolerance), and may include bands to indicate the temperature coefficient and reliability class. In four-striped resistors, the first two stripes represent the first two digits of the resistance in ohms, the third represents a multiplier, and the fourth the tolerance (which if absent, denotes ±20%). For five- and six- striped resistors the third band is the third digit, the fourth is the multiplier and the fifth is the tolerance; a sixth stripe represents the temperature coefficient. The power rating of the resistor is usually not marked and is deduced from its size.

Surface-mount resistors are marked numerically.

Early 20th century resistors, essentially uninsulated, were dipped in paint to cover their entire body for color-coding. This base color represented the first digit. A second color of paint was applied to one end of the element to represent a second digit, and a color dot (or band) in the middle provided the third digit. The rule was "body, tip, dot", providing two significant digits for value and the decimal multiplier, in that sequence. Default tolerance was ±20%. Closer-tolerance resistors had silver (±10%) or gold-colored (±5%) paint on the other end.



Wheel-based RMA Resistor Color Code guide. Circa 1945–1950.

Preferred values

Early resistors were made in more or less arbitrary round numbers; a series might have 100, 125, 150, 200, 300, etc.^[30] Early power wirewound resistors, such as brown vitreous-enameled types, were made with a system of preferred values like some of those mentioned here. Resistors as manufactured are subject to a certain percentage tolerance, and it makes sense to manufacture values that correlate with the tolerance, so that the actual value of a resistor overlaps slightly with its neighbors. Wider spacing leaves gaps; narrower spacing increases manufacturing and inventory costs to provide resistors that are more or less interchangeable.

A logical scheme is to produce resistors in a range of values which increase in a geometric progression, so that each value is greater than its predecessor by a fixed multiplier or percentage, chosen to match the tolerance of the range. For example, for a tolerance of ±20% it makes sense to have each resistor about 1.5 times its predecessor, covering a decade in 6 values. More precisely, the factor used is $1.4678 \approx 10^{1/6}$, giving values of 1.47, 2.15, 3.16, 4.64, 6.81, 10 for the 1–10-decade (a decade is a range increasing by a factor of 10; 0.1–1 and 10–100 are other examples); these are rounded in practice to 1.5, 2.2, 3.3, 4.7, 6.8, 10; followed by 15, 22, 33, ... and preceded by ... 0.47, 0.68, 1. This scheme has been adopted as the E6 series of the IEC 60063 preferred number values. There are also E12, E24, E48, E96 and E192 series for components of progressively finer resolution, with 12, 24, 48, 96, and 192 different values within each decade. The actual values used are in the IEC 60063 lists of preferred numbers.

A resistor of 100 ohms ±20% would be expected to have a value between 80 and 120 ohms; its E6 neighbors are 68 (54–82) and 150 (120–180) ohms. A sensible spacing, E6 is used for ±20% components; E12 for ±10%; E24 for ±5%; E48 for ±2%, E96 for ±1%; E192 for ±0.5% or better. Resistors are manufactured in values from a few milliohms to about a gigaohm in IEC60063 ranges appropriate for their tolerance. Manufacturers may sort resistors into tolerance-classes based on measurement. Accordingly, a selection of 100 ohms resistors with a tolerance of ±10%, might not lie just around 100 ohm (but no more than 10% off) as one would expect (a bell-curve), but rather be in two groups – either between 5 and 10% too high or 5 to 10% too low (but not closer to 100 ohm than that) because any resistors the factory had measured as being less than 5% off would have been marked and sold as resistors with only ±5% tolerance or better. When designing a circuit, this may become a consideration. This process of sorting parts based on post-production measurement is known as "binning", and can be applied to other components than resistors (such as speed grades for CPUs).

SMT resistors

Surface mounted resistors of larger sizes (metric 1608 and above) are printed with numerical values in a code related to that used on axial resistors. Standard-tolerance surface-mount technology (SMT) resistors are marked with a three-digit code, in which the first two digits are the first two significant digits of the value and the third digit is the power of ten (the number of zeroes). For example:

- $334 = 33 \times 10^4 \Omega = 330 \text{ k}\Omega$
- $222 = 22 \times 10^2 \Omega = 2.2 \text{ k}\Omega$
- $473 = 47 \times 10^3 \Omega = 47 \text{ k}\Omega$
- $105 = 10 \times 10^5 \Omega = 1 \text{ M}\Omega$

Resistances less than 100Ω are written: 100, 220, 470. The final zero represents ten to the power zero, which is 1. For example:

- $100 = 10 \times 10^0 \Omega = 10 \Omega$
- $220 = 22 \times 10^0 \Omega = 22 \Omega$

Sometimes these values are marked as 10 or 22 to prevent a mistake.

Resistances less than 10Ω have 'R' to indicate the position of the decimal point (radix point). For example:

- $4R7 = 4.7 \Omega$
- $R300 = 0.30 \Omega$
- $0R22 = 0.22 \Omega$
- $0R01 = 0.01 \Omega$

000 and 0000 sometimes appear as values on surface-mount zero-ohm links, since these have (approximately) zero resistance.

More recent surface-mount resistors are too small, physically, to permit practical markings to be applied.

Precision resistor markings

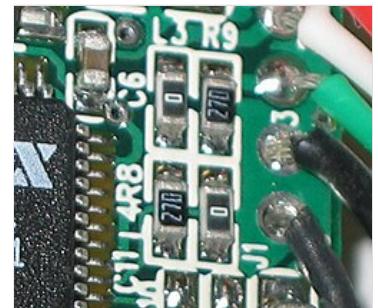
Many precision resistors, including surface mount and axial-lead types, are marked with a four-digit code. The first three digits are the significant figures and the fourth is the power of ten. For example:

- $1001 = 100 \times 10^1 \Omega = 1.00 \text{ k}\Omega$
- $4992 = 499 \times 10^2 \Omega = 49.9 \text{ k}\Omega$
- $1000 = 100 \times 10^0 \Omega = 100 \Omega$

Axial-lead precision resistors often use color code bands to represent this four-digit code.

EIA-96 marking

The former EIA-96 marking system now included in [IEC 60062:2016](#) is a more compact marking system intended for physically small high-precision resistors. It uses a two-digit code plus a letter (a total of three alphanumeric characters) to indicate 1% resistance values to three significant digits.^[31] The two digits (from "01" to "96") are a code that indicates one of the 96 "positions" in the standard E96 series of 1% resistor values. The uppercase letter is a code that indicates a power of ten multiplier. For example, the marking "01C" represents 10 kOhm; "10C" represents 12.4 kOhm; "96C" represents 97.6 kOhm.^{[32][33][34][35][36]}



This image shows four surface-mount resistors (the component at the upper left is a capacitor) including two zero-ohm resistors. Zero-ohm links are often used instead of wire links, so that they can be inserted by a resistor-inserting machine. Their resistance is negligible.

Code	Series	Letter						
Digits	E96	Y / S	X / R	A	B / H	C	D	E
01	1.00	1R00	10R0	100R	1K00	10K0	100K	1M00
02	1.02	1R02	10R2	102R	1K02	10K2	102K	1M02
03	1.05	1R05	10R5	105R	1K05	10K5	105K	1M05
04	1.07	1R07	10R7	107R	1K07	10K7	107K	1M07
05	1.10	1R10	11R0	110R	1K10	11K0	110K	1M10
06	1.13	1R13	11R3	113R	1K13	11K3	113K	1M13
07	1.15	1R15	11R5	115R	1K15	11K5	115K	1M15
08	1.18	1R18	11R8	118R	1K18	11K8	118K	1M18
09	1.21	1R21	12R1	121R	1K21	12K1	121K	1M21
10	1.24	1R24	12R4	124R	1K24	12K4	124K	1M24
11	1.27	1R27	12R7	127R	1K27	12K7	127K	1M27
12	1.30	1R30	13R0	130R	1K30	13K0	130K	1M30
13	1.33	1R33	13R3	133R	1K33	13K3	133K	1M33
14	1.37	1R37	13R7	137R	1K37	13K7	137K	1M37
15	1.40	1R40	14R0	140R	1K40	14K0	140K	1M40
16	1.43	1R43	14R3	143R	1K43	14K3	143K	1M43
17	1.47	1R47	14R7	147R	1K47	14K7	147K	1M47
18	1.50	1R50	15R0	150R	1K50	15K0	150K	1M50
19	1.54	1R54	15R4	154R	1K54	15K4	154K	1M54
20	1.58	1R58	15R8	158R	1K58	15K8	158K	1M58
21	1.62	1R62	16R2	162R	1K62	16K2	162K	1M62
22	1.65	1R65	16R5	165R	1K65	16K5	165K	1M65
23	1.69	1R69	16R9	169R	1K69	16K9	169K	1M69
24	1.74	1R74	17R4	174R	1K74	17K4	174K	1M74
25	1.78	1R78	17R8	178R	1K78	17K8	178K	1M78
26	1.82	1R82	18R2	182R	1K82	18K2	182K	1M82
27	1.87	1R87	18R7	187R	1K87	18K7	187K	1M87
28	1.91	1R91	19R1	191R	1K91	19K1	191K	1M91
29	1.96	1R96	19R6	196R	1K96	19K6	196K	1M96
30	2.00	2R00	20R0	200R	2K00	20K0	200K	2M00
31	2.05	2R05	20R5	205R	2K05	20K5	205K	2M05
32	2.10	2R10	21R0	210R	2K10	21K0	210K	2M10
33	2.15	2R15	21R5	215R	2K15	21K5	215K	2M15
34	2.21	2R21	22R1	221R	2K21	22K1	221K	2M21
35	2.26	2R26	22R6	226R	2K26	22K6	226K	2M26
36	2.32	2R32	23R2	232R	2K32	23K2	232K	2M32
37	2.37	2R37	23R7	237R	2K37	23K7	237K	2M37
38	2.43	2R43	24R3	243R	2K43	24K3	243K	2M43
39	2.49	2R49	24R9	249R	2K49	24K9	249K	2M49
40	2.55	2R55	25R5	255R	2K55	25K5	255K	2M55
41	2.61	2R61	26R1	261R	2K61	26K1	261K	2M61
42	2.67	2R67	26R7	267R	2K67	26K7	267K	2M67
43	2.74	2R74	27R4	274R	2K74	27K4	274K	2M74
44	2.80	2R80	28R0	280R	2K80	28K0	280K	2M80
45	2.87	2R87	28R7	287R	2K87	28K7	287K	2M87
46	2.94	2R94	29R4	294R	2K94	29K4	294K	2M94
47	3.01	3R01	30R1	301R	3K01	30K1	301K	3M01
48	3.09	3R09	30R9	309R	3K09	30K9	309K	3M09

Code	Series	Letter						
Digits	E96	Y / S	X / R	A	B / H	C	D	E
49	3.16	3R16	31R6	316R	3K16	31K6	316K	3M16
50	3.24	3R24	32R4	324R	3K24	32K4	324K	3M24
51	3.32	3R32	33R2	332R	3K32	33K2	332K	3M32
52	3.40	3R40	34R0	340R	3K40	34K0	340K	3M40
53	3.48	3R48	34R8	348R	3K48	34K8	348K	3M48
54	3.57	3R57	35R7	357R	3K57	35K7	357K	3M57
55	3.65	3R65	36R5	365R	3K65	36K5	365K	3M65
56	3.74	3R74	37R4	374R	3K74	37K4	374K	3M74
57	3.83	3R83	38R3	383R	3K83	38K3	383K	3M83
58	3.92	3R92	39R2	392R	3K92	39K2	392K	3M92
59	4.02	4R02	40R2	402R	4K02	40K2	402K	4M02
60	4.12	4R12	41R2	412R	4K12	41K2	412K	4M12
61	4.22	4R22	42R2	422R	4K22	42K2	422K	4M22
62	4.32	4R32	43R2	432R	4K32	43K2	432K	4M32
63	4.42	4R42	44R2	442R	4K42	44K2	442K	4M42
64	4.53	4R53	45R3	453R	4K53	45K3	453K	4M53
65	4.64	4R64	46R4	464R	4K64	46K4	464K	4M64
66	4.75	4R75	47R5	475R	4K75	47K5	475K	4M75
67	4.87	4R87	48R7	487R	4K87	48K7	487K	4M87
68	4.99	4R99	49R9	499R	4K99	49K9	499K	4M99
69	5.11	5R11	51R1	511R	5K11	51K1	511K	5M11
70	5.23	5R23	52R3	523R	5K23	52K3	523K	5M23
71	5.36	5R36	53R6	536R	5K36	53K6	536K	5M36
72	5.49	5R49	54R9	549R	5K49	54K9	549K	5M49
73	5.62	5R62	56R2	562R	5K62	56K2	562K	5M62
74	5.76	5R76	57R6	576R	5K76	57K6	576K	5M76
75	5.90	5R90	59R0	590R	5K90	59K0	590K	5M90
76	6.04	6R04	60R4	604R	6K04	60K4	604K	6M04
77	6.19	6R19	61R9	619R	6K19	61K9	619K	6M19
78	6.34	6R34	63R4	634R	6K34	63K4	634K	6M34
79	6.49	6R49	64R9	649R	6K49	64K9	649K	6M49
80	6.65	6R65	66R5	665R	6K65	66K5	665K	6M65
81	6.81	6R81	68R1	681R	6K81	68K1	681K	6M81
82	6.98	6R98	69R8	698R	6K98	69K8	698K	6M98
83	7.15	7R15	71R5	715R	7K15	71K5	715K	7M15
84	7.32	7R32	73R2	732R	7K32	73K2	732K	7M32
85	7.50	7R50	75R0	750R	7K50	75K0	750K	7M50
86	7.68	7R68	76R8	768R	7K68	76K8	768K	7M68
87	7.87	7R87	78R7	787R	7K87	78K7	787K	7M87
88	8.06	8R06	80R6	806R	8K06	80K6	806K	8M06
89	8.25	8R25	82R5	825R	8K25	82K5	825K	8M25
90	8.45	8R45	84R5	845R	8K45	84K5	845K	8M45
91	8.66	8R66	86R6	866R	8K66	86K6	866K	8M66
92	8.87	8R87	88R7	887R	8K87	88K7	887K	8M87
93	9.09	9R09	90R9	909R	9K09	90K9	909K	9M09
94	9.31	9R31	93R1	931R	9K31	93K1	931K	9M31
95	9.53	9R53	95R3	953R	9K53	95K3	953K	9M53
96	9.76	9R76	97R6	976R	9K76	97K6	976K	9M76

Industrial type designation

Power Rating at 70 °C				Tolerance code		
Type no.	Power rating (watts)	MIL-R-11 style	MIL-R-39008 style	Industrial type designation	Tolerance	MIL designation
BB	1/8	RC05	RCR05	5	±5%	J
CB	1/4	RC07	RCR07	2	±20%	M
EB	1/2	RC20	RCR20	1	±10%	K
GB	1	RC32	RCR32	-	±2%	G
HB	2	RC42	RCR42	-	±1%	F
GM	3	-	-	-	±0.5%	D
HM	4	-	-	-	±0.25%	C
				-	±0.1%	B

Steps to find out the resistance or capacitance values:^[37]

1. First two letters gives the power dissipation capacity.
2. Next three digits gives the resistance value.
 1. First two digits are the significant values
 2. Third digit is the multiplier.
 3. Final digit gives the tolerance.

If a resistor is coded:

- EB1041: power dissipation capacity = 1/2 watts, resistance value = $10 \times 10^4 \pm 10\% =$ between 9×10^4 ohms and 11×10^4 ohms.
- CB3932: power dissipation capacity = 1/4 watts, resistance value = $39 \times 10^3 \pm 20\% =$ between 31.2×10^3 and 46.8×10^3 ohms.

Common usage patterns

There are several common usage patterns that resistors are commonly configured in.^[38]

Current limiting

Resistors are commonly used to limit the amount of current flowing through a circuit. Many circuit components (such as LEDs) require the current flowing through them to be limited, but do not themselves limit the amount of current. Therefore, oftentimes resistors will be added to prevent overcurrent situations. Additionally, oftentimes circuits do not need the amount of current that would be otherwise flowing through them, so resistors can be added to limit the power consumption of such circuits.

Voltage divider

Oftentimes circuits need to provide various reference voltages for other circuits (such as voltage comparators). A fixed voltage can be obtained by taking two resistors in series between two other fixed voltages (such as the source voltage and ground). The terminal between the two resistors will be at a voltage that is between the two voltages, at a linear distance based on the relative resistances of the two resistors. For instance, if a 200 ohm resistor and a 400 ohm resistor are placed in series between 6 V and 0 V, the terminal between them will be at 4 V.

Pull-down and pull-up resistors

When a circuit is not connected to power, the voltage of that circuit is not zero but undefined (it can be influenced by previous voltages or the environment). A pull-up or pull-down resistor provides a voltage for a circuit when it is otherwise disconnected (such as when a button is not pushed down or a transistor is not active). A pull-up resistor connects the circuit to a high positive voltage (if the circuit requires a high positive default voltage) and a pull-down resistor connects the circuit to a low voltage or ground (if the circuit requires a low default voltage). The resistor value must be high enough so that, when the circuit is active, the voltage source it is attached to does not over influence the function of the circuit, but low enough so that it "pulls" quickly enough when the circuit is deactivated, and does not significantly alter the voltage from the source value.

Electrical and thermal noise

In amplifying faint signals, it is often necessary to minimize electronic noise, particularly in the first stage of amplification. As a dissipative element, even an ideal resistor naturally produces a randomly fluctuating voltage, or noise, across its terminals. This Johnson–Nyquist noise is a fundamental noise source which depends only upon the temperature and resistance of the resistor, and is predicted by the fluctuation–dissipation theorem. Using a larger value of resistance produces a larger voltage noise, whereas a smaller value of resistance generates more current noise, at a given temperature.

The thermal noise of a practical resistor may also be larger than the theoretical prediction and that increase is typically frequency-dependent. Excess noise of a practical resistor is observed only when current flows through it. This is specified in unit of $\mu\text{V/V}/\text{decade}$ – μV of noise per volt applied across the resistor per decade of frequency. The $\mu\text{V/V}/\text{decade}$ value is frequently given in dB so that a resistor with a noise index of 0 dB exhibits 1 μV (rms) of excess noise for each volt across the resistor in each frequency decade. Excess noise is thus an example of $1/\text{f}$ noise. Thick-film and carbon composition resistors generate more excess noise than other types at low frequencies. Wire-wound and thin-film resistors are often used for their better noise characteristics. Carbon composition resistors can exhibit a noise index of 0 dB while bulk metal foil resistors may have a noise index of -40 dB, usually making the excess noise of metal foil resistors insignificant.^[39] Thin film surface mount resistors typically have lower noise and better thermal stability than thick film surface mount resistors. Excess noise is also size-dependent: in general, excess noise is reduced as the physical size of a resistor is increased (or multiple resistors are used in parallel), as the independently fluctuating resistances of smaller components tend to average out.

While not an example of "noise" per se, a resistor may act as a thermocouple, producing a small DC voltage differential across it due to the thermoelectric effect if its ends are at different temperatures. This induced DC voltage can degrade the precision of instrumentation amplifiers in particular. Such voltages appear in the junctions of the resistor leads with the circuit board and with the resistor body. Common metal film resistors show such an effect at a magnitude of about 20 $\mu\text{V}/^\circ\text{C}$. Some carbon composition resistors can exhibit thermoelectric offsets as high as 400 $\mu\text{V}/^\circ\text{C}$, whereas specially constructed resistors can reduce this number to 0.05 $\mu\text{V}/^\circ\text{C}$. In applications where the thermoelectric effect may become important, care has to be taken to mount the resistors horizontally to avoid temperature gradients and to mind the air flow over the board.^[40]

Failure modes

The failure rate of resistors in a properly designed circuit is low compared to other electronic components such as semiconductors and electrolytic capacitors. Damage to resistors most often occurs due to overheating when the average power delivered to it greatly exceeds its ability to dissipate heat (specified by the resistor's *power rating*). This may be due to a fault external to the circuit but is frequently caused by the failure of another component (such as a transistor that shorts out) in the circuit connected to the resistor. Operating a resistor too close to its power rating can limit the resistor's lifespan or cause a significant change in its resistance. A safe design generally uses overrated resistors in power applications to avoid this danger.

Low-power thin-film resistors can be damaged by long-term high-voltage stress, even below maximum specified voltage and below maximum power rating. This is often the case for the startup resistors feeding a switched-mode power supply integrated circuit.

When overheated, carbon-film resistors may decrease or increase in resistance.^[41] Carbon film and composition resistors can fail (open circuit) if running close to their maximum dissipation. This is also possible but less likely with metal film and wirewound resistors.

There can also be failure of resistors due to mechanical stress and adverse environmental factors including humidity. If not enclosed, wirewound resistors can corrode.

Surface mount resistors have been known to fail due to the ingress of sulfur into the internal makeup of the resistor. This sulfur chemically reacts with the silver layer to produce non-conductive silver sulfide. The resistor's impedance goes to infinity. Sulfur resistant and anti-corrosive resistors are sold into automotive, industrial, and military applications. ASTM B809 is an industry standard that tests a part's susceptibility to sulfur.

An alternative failure mode can be encountered where large value resistors are used (hundreds of kilohms and higher). Resistors are not only specified with a maximum power dissipation, but also for a maximum voltage drop. Exceeding this voltage causes the resistor to degrade slowly reducing in resistance. The voltage dropped across large value resistors can be exceeded before the power dissipation reaches its limiting value. Since the maximum voltage specified for commonly encountered resistors is a few hundred volts, this is a problem only in applications where these voltages are encountered.

Variable resistors can also degrade in a different manner, typically involving poor contact between the wiper and the body of the resistance. This may be due to dirt or corrosion and is typically perceived as "crackling" as the contact resistance fluctuates; this is especially noticed as the device is adjusted. This is similar to crackling caused by poor contact in switches, and like switches, potentiometers are to some extent self-cleaning: running the wiper across the resistance may improve the contact. Potentiometers which are seldom adjusted, especially in dirty or harsh environments, are most likely to develop this problem. When self-cleaning of the contact is insufficient, improvement can usually be obtained through the use of contact cleaner (also known as "tuner cleaner") spray. The crackling noise associated with turning the shaft of a dirty potentiometer in an audio circuit (such as the volume control) is greatly accentuated when an undesired DC voltage is present, often indicating the failure of a DC blocking capacitor in the circuit.

See also



- [Circuit design](#)
- [Dummy load](#)
- [Electrical impedance](#)
- [High value resistors \(electronics\)](#)
- [Iron-hydrogen resistor](#)
- [Piezoresistive effect](#)

- Shot noise
- Thermistor
- Trimmer (electronics)

References

1. Harder, Douglas Wilhelm. "Resistors: A Motor with a Constant Force (Force Source)" (https://ece.uwaterloo.ca/~dwharder/Analogy_Resistors/). Department of Electrical and Computer Engineering, University of Waterloo. Retrieved 9 November 2014.
2. American Radio Relay League (ARRL) (2021). "Fundamental Theory—Circuits and Components". *ARRL Handbook for Radio Communications* (98 ed.). American Radio Relay League. ISBN 978-1-62595-139-7.
3. Doug DeMaw, ed. (1968). "Electrical Laws and Circuits —Resistance". *Radio Amateurs Handbook* (https://archive.org/details/arrl_1968_handbook) (45 ed.). American Radio Relay League.
4. Farago, P.S. (1961) *An Introduction to Linear Network Analysis*, pp. 18–21, The English Universities Press Ltd.
5. Wu, F. Y. (2004). "Theory of resistor networks: The two-point resistance". *Journal of Physics A: Mathematical and General*. **37** (26): 6653–6673. arXiv:math-ph/0402038 (<https://arxiv.org/abs/math-ph/0402038>). Bibcode:2004JPhA...37.6653W (<https://ui.adsabs.harvard.edu/abs/2004JPhA...37.6653W>). doi:10.1088/0305-4470/37/26/004 (<https://doi.org/10.1088%2F0305-4470%2F37%2F26%2F004>). S2CID 119611570 (<https://api.semanticscholar.org/CorpusID:119611570>).
6. Wu, Fa Yueh; Yang, Chen Ning (2009). *Exactly Solved Models: A Journey in Statistical Mechanics : Selected Papers with Commentaries (1963–2008)* (<https://books.google.com/books?id=H-k8dhB7ImwC&pg=PA489>). World Scientific. pp. 489–. ISBN 978-981-281-388-6.
7. "Specifications and How to Interpret Them" (https://seiselect.com/news/20170821_-_Resistor_Data_Sheet_Information.pdf) (PDF). Stackpole Electronics. Retrieved July 6, 2021.
8. "Through Hole Resistor, 0.1 Gohm, RGP Series, 250 mW, ± 5%, Axial Leaded, 750 V" (<https://web.archive.org/web/20210709190647/https://nl.farnell.com/te-connectivity/rgp0207chj100m/res-100m-5-250mw-axial-thick-film/dp/2805251>). nl.farnell.com. Archived from the original (<https://nl.farnell.com/te-connectivity/rgp0207chj100m/res-100m-5-250mw-axial-thick-film/dp/2805251>) on 2021-07-09. Retrieved 2023-10-07.
9. A family of resistors may also be characterized according to its *critical resistance*. Applying a constant voltage across resistors in that family below the critical resistance will exceed the maximum power rating first; resistances larger than the critical resistance fail first from exceeding the maximum voltage rating. See Middleton, Wendy; Van Valkenburg, Mac E. (2002). *Reference data for engineers: radio, electronics, computer, and communications* (9 ed.). Newnes. pp. 5–10. ISBN 0-7506-7291-9.
10. Harter, James H. and Lin, Paul Y. (1982) *Essentials of electric circuits*. Reston Publishing Company. pp. 96–97. ISBN 0-8359-1767-3.
11. HVR International (ed.). "SR Series: Surge Resistors for PCB Mounting" (<http://www.hvrint.com/SRSeries.pdf>). (PDF; 252 kB), 26 May 2005, retrieved 24 January 2017.
12. Beyschlag, Vishay (2008). "Basics of Linear Fixed Resistors Application Note" (<https://www.vishay.com/docs/28771/basics.pdf>), Document Number 28771.
13. Morris, C. G. (ed.) (1992). *Academic Press Dictionary of Science and Technology*. Gulf Professional Publishing. p. 360. ISBN 0122004000.
14. *Principles of automotive vehicles*. United States Department of the Army (1985). p. 13
15. "Carbon Film Resistor" (<http://www.resistorguide.com/carbon-film-resistor/>). The Resistorguide. Retrieved 10 March 2013.
16. "Thick Film and Thin Film" (https://web.archive.org/web/20110927122203/http://www.digikey.com/Web%20Export/Supplier%20Content/Stackpole_738/PDF/Stackpole_ThickFilmXThinFilm.pdf) (PDF). Digi-Key (SEI). Archived from the original (http://www.digikey.com/Web%20Export/Supplier%20Content/Stackpole_738/PDF/Stackpole_ThickFilmXThinFilm.pdf) (PDF) on 27 September 2011. Retrieved 23 July 2011.
17. "Thin and Thick film" (<http://www.resistorguide.com/thin-and-thick-film/>). resistorguide.com. resistor guide. Retrieved 3 December 2017.
18. Kuhn, Kenneth A. "Measuring the Temperature Coefficient of a Resistor" (<https://web.archive.org/web/20160304081455/http://www.kennethkuhn.com/students/ee431/text/ee431lab3.pdf>) (PDF). Archived from the original (<http://www.kennethkuhn.com/students/ee431/text/ee431lab3.pdf>) on 2016-03-04. Retrieved 2010-03-18.
19. Zandman, F.; Stein, S. (1964). "A New Precision Film Resistor Exhibiting Bulk Properties". *IEEE Transactions on Component Parts*. **11** (2): 107–119. doi:10.1109/TCP.1964.1135008 (<https://doi.org/10.1109%2FTCP.1964.1135008>).
20. *Procedures in Experimental Physics*, John Strong, p. 546.
21. "Alpha Electronics Corp. Metal Foil Resistors" (http://www.alpha-elec.co.jp/e_machine.html). Alpha-elec.co.jp. Retrieved 2008-09-22.
22. "Grid Resistors: High Power/High Current" (<http://www.milwaukeeeresistor.com/gridresistors.html>). Milwaukee Resistor Corporation. Retrieved 14 May 2012.
23. Mazda, F. F. (1981). *Discrete Electronic Components* (<https://books.google.com/books?id=3qk8AAAAIAAJ&pg=PA57>). CUP Archive. pp. 57–61. ISBN 0521234700.
24. "Decade Box – Resistance Decade Boxes" (<http://www.ietlabs.com/decaderes.html>). ietlabs.com. Retrieved 2008-09-22.
25. "Test method standard: electronic and electrical component parts" (<https://web.archive.org/web/20150209025640/http://www.dscc.dla.mil/downloads/milspec/docs/mil-std-202/std202.pdf>) (PDF). Department of Defense. Archived from the original (<http://www.dscc.dla.mil/Downloads/MilSpec/Docs/MIL-STD-202/std202.pdf>) (PDF) on 2015-02-09.
26. Fusing Resistors and Temperature-Limited Resistors for Radio- and Television- Type Appliances UL 1412 (<https://web.archive.org/web/20080210020811/http://ulstandardsinfonet.ul.com/scopes/scopes.asp?fn=1412.html>). ulstandardsinfonet.ul.com
27. Stability of Double-Walled Manganin Resistors (<http://nvl.nist.gov/pub/nistpubs/sp958-lide/063-065.pdf>) Archived (<https://web.archive.org/web/20061006125243/http://nvl.nist.gov/pub/nistpubs/sp958-lide/063-065.pdf>) 2006-10-06 at the Wayback Machine. NIST.gov
28. Klaus von Klitzing The Quantized Hall Effect (http://nobelprize.org/nobel_prizes/physics/laureates/1985/klitzing-lecture.pdf). Nobel lecture, December 9, 1985. nobelprize.org

29. "Standard Resistance Unit Type 4737B" (<https://web.archive.org/web/20080821154924/http://www.tinsley.co.uk/products/standard-resistors/4737b.htm>). Tinsley.co.uk. Archived from the original (<http://www.tinsley.co.uk/products/standard-resistors/4737b.htm>) on 2008-08-21. Retrieved 2008-09-22.
30. "1940 Catalog – page 60 – Resistors" (<https://web.archive.org/web/20170711160604/http://www.radioshackcatalogs.com/html/1940-hr060.html>). RadioShack. Archived from the original (<http://www.radioshackcatalogs.com/html/1940-hr060.html>) on 11 July 2017. Retrieved 11 July 2017.
31. "Chapter 2 - Resistor standards and codes" (<https://eepower.com/resistor-guide/resistor-standards-and-codes/resistor-smd-code/>).
32. "CRP0603 Series - Precision Chip Resistors" (<https://www.bourns.com/pdfs/CRP0603.pdf>). p. 3.
33. "Online calculator - EIA-96 SMD resistor" (<http://kiloohm.info/eia96-resistor/96C>).
34. "SMD Resistor Codes: How to Find the Value of SMD Resistors" (<https://www.electricaltechnology.org/2013/07/how-to-calculate-or-find-value-of-smd.html>).
35. "Marking Codes used on Welwyn Chip Resistors" (<https://www.ttelectronics.com/TTElectronics/media/ProductFiles/ApplicationNotes/TN004-Methods-for-Marking-Values-on-Resistors.pdf>). p. 2.
36. "Surface Mount Resistor: codes & markings" (https://www.electronics-notes.com/articles/electronic_components/resistors/smt-smd-resistor-codes-markings.php).
37. Maini, A. K. (2008), *Electronics and Communications Simplified*, 9th ed., Khanna Publications. ISBN 817409217X
38. Bartlett, Jonathan (2020). "Basic Resistor Circuit Patterns" (https://link.springer.com/chapter/10.1007/978-1-4842-5979-5_9). *Electronics for Beginners*. pp. 129–144. doi:10.1007/978-1-4842-5979-5_9 (https://doi.org/10.1007%2F978-1-4842-5979-5_9). ISBN 978-1-4842-5978-8. S2CID 226539488 (<https://api.semanticscholar.org/CorpusID:226539488>).
39. *Audio Noise Reduction Through the Use of Bulk Metal Foil Resistors – "Hear the Difference"* (<https://web.archive.org/web/20130119120843/http://www.c-c-i.com/sites/default/files/vse-an00.pdf>) (PDF). Archived from the original (<http://www.c-c-i.com/sites/default/files/vse-an00.pdf>) on 2013-01-19. Retrieved 2009-08-03., Application note AN0003, Vishay Intertechnology Inc, 12 July 2005.
40. Jung, Walt (2005). "Chapter 7 – Hardware and Housekeeping Techniques" (http://www.analog.com/library/analogDialogue/archives/39-05/Web_Ch7_final_J.pdf) (PDF). *Op Amp Applications Handbook* (http://www.analog.com/library/analogDialogue/archives/39-05/op_amp_applications_handbook.html). Newnes. p. 7.11. ISBN 0-7506-7844-5.
41. "Electronic components – resistors" (https://web.archive.org/web/20080403111045/https://www.fda.gov/ora/Inspect_ref/itg/itg31.html). *Inspector's Technical Guide*. US Food and Drug Administration. 1978-01-16. Archived from the original (https://www.fda.gov/ora/Inspect_ref/itg/itg31.html) on 2008-04-03. Retrieved 2008-06-11.

External links

- Color Coded Resistance Calculator (<https://web.archive.org/web/20110401175312/http://www.ese.upenn.edu/rca/calcs.html>) - University of Pennsylvania
 - Resistor Types – Does It Matter? (<https://web.archive.org/web/20130407011352/http://www.aikenamps.com/ResistorNoise.htm>) - Aiken Amps
 - Difference between types of resistors (http://www.analog.com/library/analogDialogue/archives/31-1/Ask_Engineer.html) - Analog Devices
 - Basics of Linear Fixed Resistors (<https://www.vishay.com/docs/28771/basics.pdf>) - Vishay
 - 4-terminal resistors – How ultra-precise resistors work (<https://web.archive.org/web/20110715110844/http://www.powerstandards.com/4terminal.htm>) - PSL
 - Beginners' Guide to Potentiometers (<https://www.sound-au.com/pots.htm>) - ESP
-

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Resistor&oldid=1282499534>"



Inductor

An **inductor**, also called a **coil**, **choke**, or **reactor**, is a passive two-terminal electrical component that stores energy in a magnetic field when an electric current flows through it.^[1] An inductor typically consists of an insulated wire wound into a coil.

When the current flowing through the coil changes, the time-varying magnetic field induces an electromotive force (emf) (voltage) in the conductor, described by Faraday's law of induction. According to Lenz's law, the induced voltage has a polarity (direction) which opposes the change in current that created it. As a result, inductors oppose any changes in current through them.

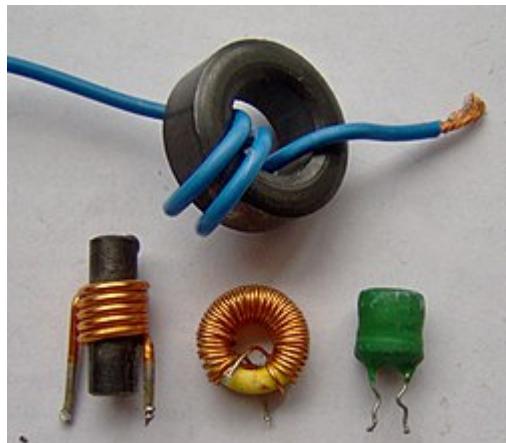
An inductor is characterized by its inductance, which is the ratio of the voltage to the rate of change of current. In the International System of Units (SI), the unit of inductance is the henry (H) named for 19th century American scientist Joseph Henry. In the measurement of magnetic circuits, it is equivalent to weber ampere. Inductors have values that typically range from $1 \mu\text{H}$ (10^{-6} H) to 20 H. Many inductors have a magnetic core made of iron or ferrite inside the coil, which serves to increase the magnetic field and thus the inductance. Along with capacitors and resistors, inductors are one of the three passive linear circuit elements that make up electronic circuits. Inductors are widely used in alternating current (AC) electronic equipment, particularly in radio equipment. They are used to block AC while allowing DC to pass; inductors designed for this purpose are called chokes. They are also used in electronic filters to separate signals of different frequencies, and in combination with capacitors to make tuned circuits, used to tune radio and TV receivers.

The term inductor seems to come from Heinrich Daniel Ruhmkorff, who called the induction coil he invented in 1851 an inductorium.^[2]

Description

An electric current flowing through a conductor generates a magnetic field surrounding it. The magnetic flux linkage Φ_B generated by a given current I depends on the geometric shape of the circuit. Their ratio defines the inductance L .^{[3][4][5][6]} Thus

Inductor



A selection of low-value inductors

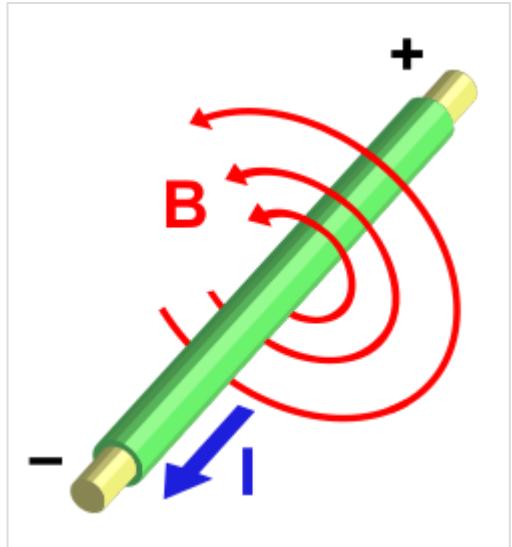
<u>Component type</u>	Passive
<u>Working principle</u>	<u>Electromagnetic induction</u>
<u>Inventor</u>	<u>Michael Faraday</u>
<u>Invention year</u>	1831
<u>Number of terminals</u>	2

Electronic symbol



$$L := \frac{\Phi_B}{I}.$$

The inductance of a circuit depends on the geometry of the current path as well as the magnetic permeability of nearby materials. An inductor is a component consisting of a wire or other conductor shaped to increase the magnetic flux through the circuit, usually in the shape of a coil or helix, with two terminals. Winding the wire into a coil increases the number of times the magnetic flux lines link the circuit, increasing the field and thus the inductance. The more turns, the higher the inductance. The inductance also depends on the shape of the coil, separation of the turns, and many other factors. By adding a "magnetic core" made of a ferromagnetic material like iron inside the coil, the magnetizing field from the coil will induce magnetization in the material, increasing the magnetic flux. The high permeability of a ferromagnetic core can increase the inductance of a coil by a factor of several thousand over what it would be without it.



An electric current I creates a magnetic field B around it

Constitutive equation

Any change in the current through an inductor creates a changing flux, inducing a voltage across the inductor. By Faraday's law of induction, the voltage \mathcal{E} induced by any change in magnetic flux through the circuit is given by^[6]

$$\mathcal{E} = -\frac{d\Phi_B}{dt}.$$

Reformulating the definition of L above, we obtain^[6]

$$\Phi_B = LI.$$

It follows that

$$\mathcal{E} = -\frac{d\Phi_B}{dt} = -\frac{d}{dt}(LI)$$

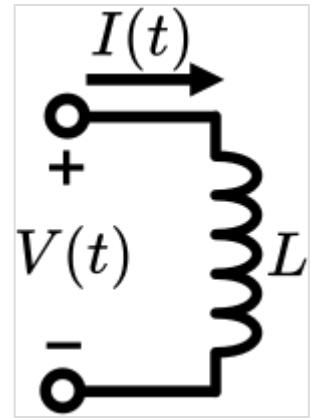
$$\boxed{\mathcal{E} = -L \frac{dI}{dt}}$$

if L is independent of time, current and magnetic flux linkage. Thus, inductance is also a measure of the amount of electromotive force (voltage) generated for a given rate of change of current. This is usually taken to be the constitutive relation (defining equation) of the inductor.

Because the induced voltage is positive at the current's entrance terminal, the inductor's current–voltage relationship is often expressed without a negative sign by using the current's exit terminal as the reference point for the voltage $V(t)$ at the current's entrance terminal (as labeled in the schematic). The current–voltage relationship is then:

$$V(t) = L \frac{dI(t)}{dt} .$$

The dual of the inductor is the capacitor, which stores energy in an electric field rather than a magnetic field. Its current–voltage relation replaces L with the capacitance C and has current and voltage swapped from these equations.



Schematic using current's exit terminal as reference for voltage

Lenz's law

The polarity (direction) of the induced voltage is given by Lenz's law, which states that the induced voltage will be such as to oppose the change in current.^[7] For example, if the current through an inductor is increasing, the induced potential difference will be positive at the current's entrance point and negative at the exit point, tending to oppose the additional current.^{[8][9][10]} The energy from the external circuit necessary to overcome this potential "hill" is being stored in the magnetic field of the inductor. If the current is decreasing, the induced voltage will be negative at the current's entrance point and positive at the exit point, tending to maintain the current. In this case energy from the magnetic field is being returned to the circuit.

Energy stored in an inductor

One intuitive explanation as to why a potential difference is induced on a change of current in an inductor goes as follows:

When there is a change in current through an inductor there is a change in the strength of the magnetic field. For example, if the current is increased, the magnetic field increases. This, however, does not come without a price. The magnetic field contains potential energy, and increasing the field strength requires more energy to be stored in the field. This energy comes from the electric current through the inductor. The increase in the magnetic potential energy of the field is provided by a corresponding drop in the electric potential energy of the charges flowing through the windings. This appears as a voltage drop across the windings as long as the current increases. Once the current is no longer increased and is held constant, the energy in the magnetic field is constant and no additional energy must be supplied, so the voltage drop across the windings disappears.

Similarly, if the current through the inductor decreases, the magnetic field strength decreases, and the energy in the magnetic field decreases. This energy is returned to the circuit in the form of an increase in the electrical potential energy of the moving charges, causing a voltage rise across the windings.

Derivation

The work done per unit charge on the charges passing through the inductor is $-\mathcal{E}$. The negative sign indicates that the work is done *against* the emf, and is not done *by* the emf. The current I is the charge per unit time passing through the inductor. Therefore, the rate of work W done by the charges against the emf, that is the rate of change of energy of the current, is given by

$$\frac{dW}{dt} = -\mathcal{E}I$$

From the constitutive equation for the inductor, $-\mathcal{E} = L \frac{dI}{dt}$ so

$$\frac{dW}{dt} = L \frac{dI}{dt} \cdot I = LI \cdot \frac{dI}{dt}$$

$$dW = LI \cdot dI$$

In a ferromagnetic core inductor, when the magnetic field approaches the level at which the core saturates, the inductance will begin to change, it will be a function of the current $L(I)$. Neglecting losses, the energy W stored by an inductor with a current I_0 passing through it is equal to the amount of work required to establish the current through the inductor.

This is given by: $W = \int_0^{I_0} L_d(I) I dI$, where $L_d(I)$ is the so-called "differential inductance" and is defined as: $L_d = \frac{d\Phi_B}{dI}$. In an air core inductor or a ferromagnetic core inductor below saturation, the inductance is constant (and equal to the differential inductance), so the stored energy is

$$W = L \int_0^{I_0} I dI$$

$$W = \frac{1}{2} LI_0^2$$

For inductors with magnetic cores, the above equation is only valid for linear regions of the magnetic flux, at currents below the saturation level of the inductor, where the inductance is approximately constant. Where this is not the case, the integral form must be used with L_d variable.

Voltage step response

When a voltage step is applied to an inductor:

- In the short-time limit, since the current cannot change instantaneously, the initial current is zero. The equivalent circuit of an inductor immediately after the step is applied is an open circuit.
- As time passes, the current increases at a constant rate with time until the inductor starts to saturate.
- In the long-time limit, the transient response of the inductor will die out, the magnetic flux through the inductor will become constant, so no voltage would be induced between the terminals of the inductor. Therefore, assuming the resistance of the windings is negligible, the equivalent circuit of an inductor a long time after the step is applied is a short circuit.

Ideal and real inductors

The constitutive equation describes the behavior of an *ideal inductor* with inductance L , and without resistance, capacitance, or energy dissipation. In practice, inductors do not follow this theoretical model; *real inductors* have a measurable resistance due to the resistance of the wire and energy losses in the core, and parasitic capacitance between turns of the wire.^{[11][12]}

A real inductor's capacitive reactance rises with frequency, and at a certain frequency, the inductor will behave as a resonant circuit. Above this self-resonant frequency, the capacitive reactance is the dominant part of the inductor's impedance. At higher frequencies, resistive losses in the windings increase due to the skin effect and proximity effect.

Inductors with ferromagnetic cores experience additional energy losses due to hysteresis and eddy currents in the core, which increase with frequency. At high currents, magnetic core inductors also show sudden departure from ideal behavior due to nonlinearity caused by magnetic saturation of the core.

Inductors radiate electromagnetic energy into surrounding space and may absorb electromagnetic emissions from other circuits, resulting in potential electromagnetic interference.

An early solid-state electrical switching and amplifying device called a saturable reactor exploits saturation of the core as a means of stopping the inductive transfer of current via the core.

Q factor

The winding resistance appears as a resistance in series with the inductor; it is referred to as DCR (DC resistance). This resistance dissipates some of the reactive energy. The quality factor (or *Q*) of an inductor is the ratio of its inductive reactance to its resistance at a given frequency, and is a measure of its efficiency. The higher the *Q* factor of the inductor, the closer it approaches the behavior of an ideal inductor. High *Q* inductors are used with capacitors to make resonant circuits in radio transmitters and receivers. The higher the *Q* is, the narrower the bandwidth of the resonant circuit.

The *Q* factor of an inductor is defined as

$$Q = \frac{\omega L}{R}$$

where *L* is the inductance, *R* is the DC resistance, and the product ωL is the inductive reactance

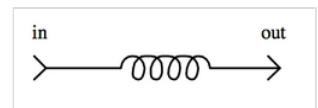
Q increases linearly with frequency if *L* and *R* are constant. Although they are constant at low frequencies, the parameters vary with frequency. For example, skin effect, proximity effect, and core losses increase *R* with frequency; winding capacitance and variations in permeability with frequency affect *L*.

At low frequencies and within limits, increasing the number of turns *N* improves *Q* because *L* varies as N^2 while *R* varies linearly with *N*. Similarly increasing the radius *r* of an inductor improves (or increases) *Q* because *L* varies with r^2 while *R* varies linearly with *r*. So high *Q* air core inductors often have large diameters and many turns. Both of those examples assume the diameter of the wire stays the same, so both examples use proportionally more wire. If the total mass of wire is held constant, then there would be no advantage to increasing the number of turns or the radius of the turns because the wire would have to be proportionally thinner.

Using a high permeability ferromagnetic core can greatly increase the inductance for the same amount of copper, so the core can also increase the *Q*. Cores however also introduce losses that increase with frequency. The core material is chosen for best results for the frequency band. High *Q* inductors must avoid saturation; one way is by using a (physically larger) air core inductor. At VHF or higher frequencies an air core is likely to be used. A well designed air core inductor may have a *Q* of several hundred.

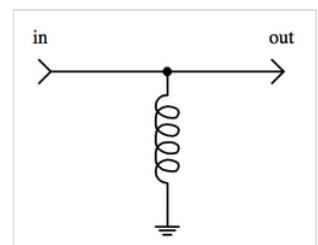
Applications

Inductors are used extensively in analog circuits and signal processing. Applications range from the use of large inductors in power supplies, which in conjunction with filter capacitors remove ripple which is a multiple of the mains frequency (or the switching frequency for switched-mode power supplies) from the direct current output, to the small inductance of the ferrite bead or torus installed around a cable to prevent radio frequency interference from being transmitted down the wire.



Example of signal filtering. In this configuration, the inductor blocks AC current, while allowing DC current to pass.

Inductors are used as the energy storage device in many switched-mode power supplies to produce DC current. The inductor supplies energy to the circuit to keep current flowing during the "off" switching periods and enables topographies where the output voltage is higher than the input voltage.



Example of signal filtering. In this configuration, the inductor decouples DC current, while allowing AC current to pass.

A tuned circuit, consisting of an inductor connected to a capacitor, acts as a resonator for oscillating current. Tuned circuits are widely used in radio frequency equipment such as radio transmitters and receivers, as narrow bandpass filters to select a single frequency from a composite signal, and in electronic oscillators to generate sinusoidal signals.

Two (or more) inductors in proximity that have coupled magnetic flux (mutual inductance) form a transformer, which is a fundamental component of every electric utility power grid. The efficiency of a transformer may decrease as the frequency increases due to eddy currents in the core material and skin effect on the windings. The size of the core can be decreased at higher frequencies. For this reason, aircraft use 400 hertz alternating current rather than the usual 50 or 60 hertz, allowing a great saving in weight from the use of smaller transformers.^[13] Transformers enable switched-mode power supplies that galvanically isolate the output from the input.

Transformers enable switched-mode power supplies that galvanically isolate the output from the input.

Inductors are also employed in electrical transmission systems, where they are used to limit switching currents and fault currents. In this field, they are more commonly referred to as reactors.

Inductors have parasitic effects which cause them to depart from ideal behavior. They create and suffer from electromagnetic interference (EMI). Their physical size prevents them from being integrated on semiconductor chips. So the use of inductors is declining in modern electronic devices, particularly compact portable devices. Real inductors are increasingly being replaced by active circuits such as the gyrator which can synthesize inductance using capacitors.

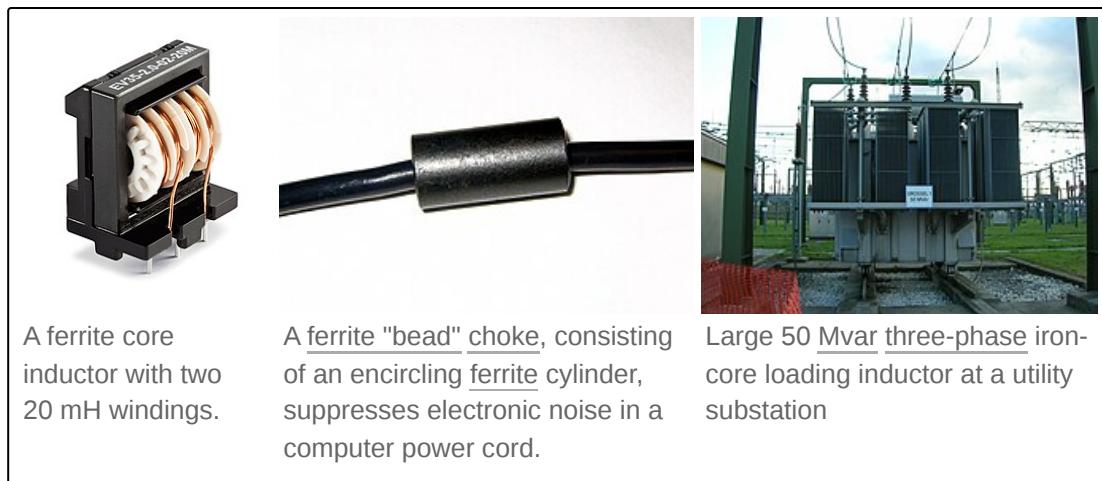
Inductor construction

An inductor usually consists of a coil of conducting material, typically insulated copper wire, wrapped around a core either of plastic (to create an air-core inductor) or of a ferromagnetic (or ferrimagnetic) material; the latter is called an "iron core" inductor. The high permeability of the ferromagnetic core increases the magnetic field and confines it closely to the inductor, thereby increasing the inductance. Low frequency inductors are constructed like transformers, with cores of electrical steel laminated to

prevent eddy currents. 'Soft' ferrites are widely used for cores above audio frequencies, since they do not cause the large energy losses at high frequencies that ordinary iron

alloys do. Inductors come in many shapes. Some inductors have an adjustable core, which enables changing of the inductance. Inductors used to block very high frequencies are sometimes made by stringing a ferrite bead on a wire.

Small inductors can be etched directly onto a printed circuit board by laying out the trace in a spiral pattern. Some such planar inductors use a planar core. Small value inductors can also be built on integrated circuits using the same processes that are used to make interconnects. Aluminium interconnect is typically used, laid out in a spiral coil pattern. However, the small dimensions limit the inductance, and it is far more common to use a circuit called a gyrator that uses a capacitor and active components to behave similarly to an inductor. Regardless of the design, because of the low inductances and low power dissipation on-die inductors allow, they are currently only commercially used for high frequency RF circuits.



Shielded inductors

Inductors used in power regulation systems, lighting, and other systems that require low-noise operating conditions, are often partially or fully shielded.^{[14][15]} In telecommunication circuits employing induction coils and repeating transformers shielding of inductors in close proximity reduces circuit cross-talk.

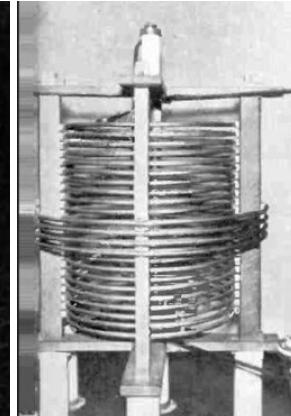
Types

Air-core inductor

The term air core coil describes an inductor that does not use a magnetic core made of a ferromagnetic material. The term refers to coils wound on plastic, ceramic, or other nonmagnetic forms, as well as those that have only air inside the windings. Air core coils have lower inductance than ferromagnetic core coils, but are often used at high frequencies because they are free from energy losses called core losses that occur in ferromagnetic cores, which increase with frequency. A side effect that can occur in air core coils in which the winding is not rigidly supported on a form is 'microphony': mechanical vibration of the windings can cause variations in the inductance.



High Q tank coil in tuned circuit of radio transmitter

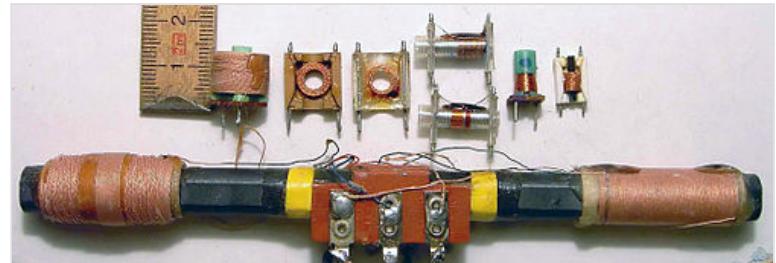


An antenna tuning coil at an AM radio station.

These coils illustrate high power high Q construction: single layer winding with turns spaced apart to reduce proximity effect losses, made of silver-plated wire or tubing to reduce skin effect losses, supported by narrow insulating strips to reduce dielectric losses

Radio-frequency inductor

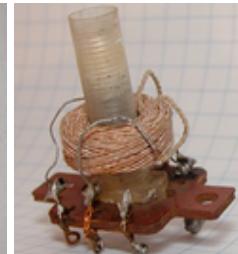
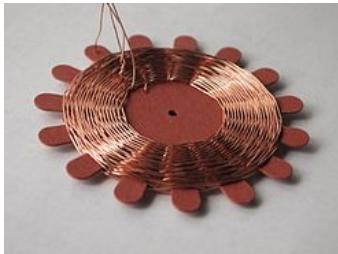
At high frequencies, particularly radio frequencies (RF), inductors have higher resistance and other losses. In addition to causing power loss, in resonant circuits this can reduce the Q factor of the circuit, broadening the bandwidth. In RF inductors specialized construction techniques are used to minimize these losses. The losses are due to these effects:



Collection of RF inductors, showing techniques to reduce losses. The three top left and the ferrite loopstick or rod antenna, [16][17][18][19] bottom, have basket windings.

- **Skin effect:** The resistance of a wire to high frequency current is higher than its resistance to direct current because of skin effect.^{[20][21]:p.141} Due to induced eddy currents, radio frequency alternating current does not penetrate far into the body of a conductor but travels along its surface. For example, at 6 MHz the skin depth of copper wire is about 0.001 inches (25 μm); most of the current is within this depth of the surface. Therefore, in a solid wire, the interior portion of the wire may carry little current, effectively increasing its resistance.
- **Proximity effect:** Another similar effect that also increases the resistance of the wire at high frequencies is proximity effect, which occurs in parallel wires that lie close to each other.^{[22][21]:p.98} The individual magnetic field of adjacent turns induces eddy currents in the wire of the coil, which causes the current density in the conductor to be displaced away from the adjacent surfaces. Like skin effect, this reduces the effective cross-sectional area of the wire conducting current, increasing its resistance.
- **Dielectric losses:** The high frequency electric field near the conductors in a tank coil can cause the motion of polar molecules in nearby insulating materials, dissipating energy as heat. For this reason, coils used for tuned circuits may be suspended in air, supported by narrow plastic or ceramic strips rather than being wound on coil forms.
- **Parasitic capacitance:** The capacitance between individual wire turns of the coil, called parasitic capacitance, does not cause energy losses but can change the behavior of the coil. Each turn of the coil is at a slightly different potential, so the electric field between neighboring turns stores charge on the wire, so the coil acts as if it has a capacitor in

parallel with it. At a high enough frequency this capacitance can resonate with the inductance of the coil forming a tuned circuit, causing the coil to become self-resonant.



(left) Spiderweb coil (right) Adjustable ferrite slug-tuned RF coil with basketweave winding and litz wire

To reduce parasitic capacitance and proximity effect, high Q RF coils are constructed to avoid having many turns lying close together, parallel to one another. The windings of RF coils are often limited to a single layer, and the turns are spaced apart. To reduce resistance due to skin effect, in high-power inductors such as those used in transmitters the windings are sometimes made of a metal strip or tubing which has a larger surface area, and the surface is silver-plated.

Basket-weave coils

To reduce proximity effect and parasitic capacitance, multilayer RF coils are wound in patterns in which successive turns are not parallel but crisscrossed at an angle; these are often called *honeycomb* or basket-weave coils. These are occasionally wound on a vertical insulating supports with dowels or slots, with the wire weaving in and out through the slots.

Spiderweb coils

Another construction technique with similar advantages is flat spiral coils. These are often wound on a flat insulating support with radial spokes or slots, with the wire weaving in and out through the slots; these are called *spiderweb* coils. The form has an odd number of slots, so successive turns of the spiral lie on opposite sides of the form, increasing separation.

Litz wire

To reduce skin effect losses, some coils are wound with a special type of radio frequency wire called litz wire. Instead of a single solid conductor, litz wire consists of a number of smaller wire strands that carry the current. Unlike ordinary stranded wire, the strands are insulated from each other, to prevent skin effect from forcing the current to the surface, and are twisted or braided together. The twist pattern ensures that each wire strand spends the same amount of its length on the outside of the wire bundle, so skin effect distributes the current equally between the strands, resulting in a larger cross-sectional conduction area than an equivalent single wire.

Axial Inductor

Small inductors for low current and low power are made in molded cases resembling resistors. These may be either plain (phenolic) core or ferrite core. An ohmmeter readily distinguishes them from similar-sized resistors by showing the low resistance of the inductor.

Ferromagnetic-core inductor

Ferromagnetic-core or iron-core inductors use a magnetic core made of a ferromagnetic or ferrimagnetic material such as iron or ferrite to increase the inductance. A magnetic core can increase the inductance of a coil by a factor of several thousand, by increasing the magnetic field due to its higher magnetic

permeability. However the magnetic properties of the core material cause several side effects which alter the behavior of the inductor and require special construction:



Core losses

A time-varying current in a ferromagnetic inductor, which causes a time-varying magnetic field in its core, causes energy losses in the core material that are dissipated as heat, due to two processes:

Eddy currents

From Faraday's law of induction, the changing magnetic field can induce circulating loops of electric current in the conductive metal core. The energy in these currents is dissipated as heat in the resistance of the core material. The amount of energy lost increases with the area inside the loop of current.



A variety of types of ferrite core inductors and transformers

Hysteresis

Changing or reversing the magnetic field in the core also causes losses due to the motion of the tiny magnetic domains it is composed of. The energy loss is proportional to the area of the hysteresis loop in the BH graph of the core material. Materials with low coercivity have narrow hysteresis loops and so low hysteresis losses.

Core loss is non-linear with respect to both frequency of magnetic fluctuation and magnetic flux density. Frequency of magnetic fluctuation is the frequency of AC current in the electric circuit; magnetic flux density corresponds to current in the electric circuit. Magnetic fluctuation gives rise to hysteresis, and magnetic flux density causes eddy currents in the core. These nonlinearities are distinguished from the threshold nonlinearity of saturation. Core loss can be approximately modeled with Steinmetz's equation. At low frequencies and over limited frequency spans (maybe a factor of 10), core loss may be treated as a linear function of frequency with minimal error. However, even in the audio range, nonlinear effects of magnetic core inductors are noticeable and of concern.

Saturation

If the current through a magnetic core coil is high enough that the core saturates, the inductance will fall and current will rise dramatically. This is a nonlinear threshold phenomenon and results in distortion of the signal. For example, audio signals can suffer intermodulation distortion in saturated inductors. To prevent this, in linear circuits the current through iron core inductors must be limited below the saturation level. Some laminated cores have a narrow air gap in them for this purpose, and powdered iron cores have a distributed air gap. This allows higher levels of magnetic flux and thus higher currents through the inductor before it saturates.^[23]

Curie point demagnetization

If the temperature of a ferromagnetic or ferrimagnetic core rises to a specified level, the magnetic domains dissociate, and the material becomes paramagnetic, no longer able to support magnetic flux. The inductance falls and current rises dramatically, similarly to what happens during saturation. The effect is reversible: When the temperature falls below the Curie point, magnetic flux resulting from current in the electric circuit will realign the magnetic domains of the core and its magnetic flux will be restored. The Curie point of ferromagnetic materials (iron alloys) is quite high; iron is highest at 770 °C. However, for

some ferrimagnetic materials (ceramic iron compounds – ferrites) the Curie point can be close to ambient temperatures (below 100 °C).

Laminated-core inductor

Low-frequency inductors are often made with laminated cores to prevent eddy currents, using construction similar to transformers. The core is made of stacks of thin steel sheets or laminations oriented parallel to the field, with an insulating coating on the surface. The insulation prevents eddy currents between the sheets, so any remaining currents must be within the cross sectional area of the individual laminations, reducing the area of the loop and thus reducing the energy losses greatly. The laminations are made of low-conductivity silicon steel to further reduce eddy current losses.



Laminated iron core ballast inductor for a metal halide lamp

Ferrite-core inductor

For higher frequencies, inductors are made with cores of ferrite. Ferrite is a ceramic ferrimagnetic material that is nonconductive, so eddy currents cannot flow within it. The formulation of ferrite is $xx\text{Fe}_2\text{O}_4$ where xx represents various metals. For inductor cores soft ferrites are used, which have low coercivity and thus low hysteresis losses.

Powdered-iron-core inductor

Another material is powdered iron cemented with a binder. Medium frequency equipment almost exclusively uses powdered iron cores, and inductors and transformers built for the lower shortwaves are made using either cemented powdered iron or ferrites.

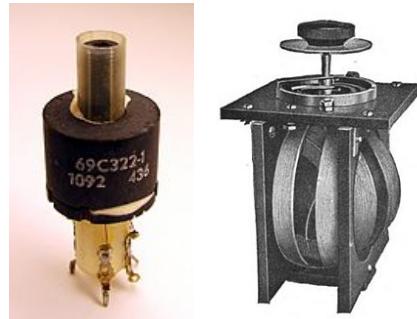
Toroidal-core inductor

In an inductor wound on a straight rod-shaped core, the magnetic field lines emerging from one end of the core must pass through the air to re-enter the core at the other end. This reduces the field, because much of the magnetic field path is in air rather than the higher permeability core material and is a source of electromagnetic interference. A higher magnetic field and inductance can be achieved by forming the core in a closed magnetic circuit. The magnetic field lines form closed loops within the core without leaving the core material. The shape often used is a toroidal or doughnut-shaped ferrite core. Because of their symmetry, toroidal cores allow a minimum of the magnetic flux to escape outside the core (called leakage flux), so they radiate less electromagnetic interference than other shapes. Toroidal core coils are manufactured of various materials, primarily ferrite, powdered iron and laminated cores.^[24]



Toroidal inductor in the power supply of a wireless router

Variable inductor



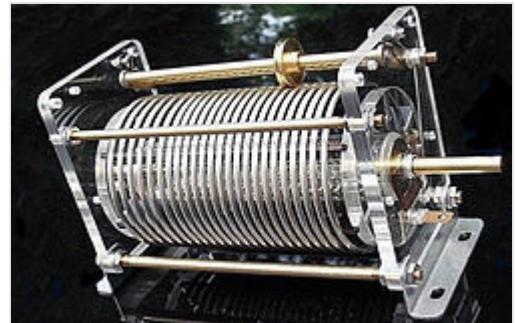
(left) Inductor with a threaded ferrite slug (visible at top) that can be turned to move it into or out of the coil, 4.2 cm high. (right) A variometer used in radio receivers in the 1920s

Probably the most common type of variable inductor today is one with a moveable ferrite magnetic core, which can be slid or screwed in or out of the coil. Moving the core farther into the coil increases the permeability, increasing the magnetic field and the inductance. Many inductors used in radio applications (usually less than 100 MHz) use adjustable cores in order to tune such inductors to their desired value, since manufacturing processes have certain tolerances (inaccuracy). Sometimes such cores for frequencies above 100 MHz are made from highly conductive non-magnetic material such as aluminum.^[25] They decrease the inductance because the magnetic field must bypass them.

Air core inductors can use sliding contacts or multiple taps to increase or decrease the number of turns included in the circuit, to change the inductance. A type much used in the past but mostly obsolete today has a spring contact that can slide along the bare surface of the windings. The disadvantage of this type is that the contact usually short-circuits one or more turns. These turns act like a single-turn short-circuited transformer secondary winding; the large currents induced in them cause power losses.

A type of continuously variable air core inductor is the *variometer*. This consists of two coils with the same number of turns connected in series, one inside the other. The inner coil is mounted on a shaft so its axis can be turned with respect to the outer coil. When the two coils' axes are collinear, with the magnetic fields pointing in the same direction, the fields add and the inductance is maximum. When the inner coil is turned so its axis is at an angle with the outer, the mutual inductance between them is smaller so the total inductance is less. When the inner coil is turned 180° so the coils are collinear with their magnetic fields opposing, the two fields cancel each other and the inductance is very small. This type has the advantage that it is continuously variable over a wide range. It is used in antenna tuners and matching circuits to match low frequency transmitters to their antennas.

Another method to control the inductance without any moving parts requires an additional DC current bias winding which controls the permeability of an easily saturable core material. See Magnetic amplifier.



A "roller coil", an adjustable air-core RF inductor used in the tuned circuits of radio transmitters. One of the contacts to the coil is made by the small grooved wheel, which rides on the wire. Turning the shaft rotates the coil, moving the contact wheel up or down the coil, allowing more or fewer turns of the coil into the circuit, to change the inductance.

Choke

A choke is an inductor designed specifically for blocking high-frequency alternating current (AC) in an electrical circuit, while allowing DC or low-frequency signals to pass. Because the inductor restricts or "chokes" the changes in current, this type of inductor is called a choke. It usually consists of a coil of insulated wire wound on a magnetic core, although some consist of a donut-shaped "bead" of ferrite material strung on a wire. Like other inductors, chokes resist changes in current passing through them increasingly with frequency. The difference between chokes and other inductors is that chokes do not require the high Q factor construction techniques that are used to reduce the resistance in inductors used in tuned circuits.



An MF or HF radio choke for tenths of an ampere, and a ferrite bead VHF choke for several amperes.

Circuit analysis

The effect of an inductor in a circuit is to oppose changes in current through it by developing a voltage across it proportional to the rate of change of the current. An ideal inductor would offer no resistance to a constant direct current; however, only superconducting inductors have truly zero electrical resistance.

The relationship between the time-varying voltage $v(t)$ across an inductor with inductance L and the time-varying current $i(t)$ passing through it is described by the differential equation:

$$v(t) = L \frac{di(t)}{dt}$$

When there is a sinusoidal alternating current (AC) through an inductor, a sinusoidal voltage is induced. The amplitude of the voltage is proportional to the product of the amplitude (I_P) of the current and the angular frequency (ω) of the current.

$$\begin{aligned} i(t) &= I_P \sin(\omega t) \\ \frac{di(t)}{dt} &= I_P \omega \cos(\omega t) \\ v(t) &= L I_P \omega \cos(\omega t) \end{aligned}$$

In this situation, the phase of the current lags that of the voltage by $\pi/2$ (90°). For sinusoids, as the voltage across the inductor goes to its maximum value, the current goes to zero, and as the voltage across the inductor goes to zero, the current through it goes to its maximum value.

If an inductor is connected to a direct current source with value I via a resistance R (at least the DCR of the inductor), and then the current source is short-circuited, the differential relationship above shows that the current through the inductor will discharge with an exponential decay:

$$i(t) = I e^{-\frac{R}{L}t}$$

Reactance

The ratio of the peak voltage to the peak current in an inductor energised from an AC source is called the reactance and is denoted X_L .

$$X_L = \frac{V_P}{I_P} = \frac{\omega L I_P}{I_P}$$

Thus,

$$X_L = \omega L$$

where ω is the angular frequency.

Reactance is measured in ohms but referred to as *impedance* rather than resistance; energy is stored in the magnetic field as current rises and discharged as current falls. Inductive reactance is proportional to frequency. At low frequency the reactance falls; at DC, the inductor behaves as a short circuit. As frequency increases the reactance increases and at a sufficiently high frequency the reactance approaches that of an open circuit.

Corner frequency

In filtering applications, with respect to a particular load impedance, an inductor has a corner frequency defined as:

$$f_{3 \text{ dB}} = \frac{R}{2\pi L}$$

Laplace circuit analysis (s-domain)

When using the Laplace transform in circuit analysis, the impedance of an ideal inductor with no initial current is represented in the *s* domain by:

$$Z(s) = Ls$$

where

L is the inductance, and
 s is the complex frequency.

If the inductor does have initial current, it can be represented by:

- adding a voltage source in series with the inductor, having the value:
 LI_0

where

L is the inductance, and
 I_0 is the initial current in the inductor.

(The source should have a polarity that is aligned with the initial current.)

- or by adding a current source in parallel with the inductor, having the value:

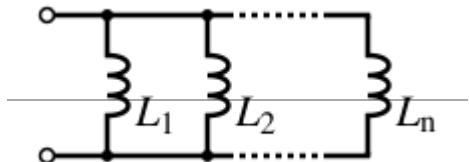
$$\frac{I_0}{s}$$

where

- I_0 is the initial current in the inductor.
- s is the complex frequency.

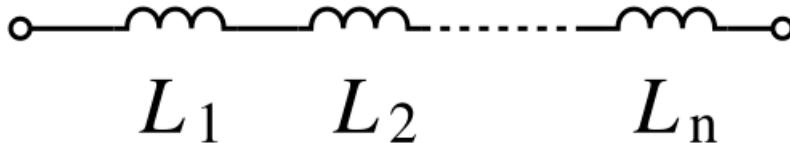
Inductor networks

Inductors in a parallel configuration each have the same potential difference (voltage). To find their total equivalent inductance (L_{eq}):



$$L_{\text{eq}} = \left(\sum_{i=1}^n \frac{1}{L_i} \right)^{-1} = \left(\frac{1}{L_1} + \frac{1}{L_2} + \dots + \frac{1}{L_n} \right)^{-1}.$$

The current through inductors in series stays the same, but the voltage across each inductor can be different. The sum of the potential differences (voltage) is equal to the total voltage. To find their total inductance:



$$L_{\text{eq}} = \sum_{i=1}^n L_i = L_1 + L_2 + \dots + L_n.$$

These simple relationships hold true only when there is no mutual coupling of magnetic fields between individual inductors.

Mutual inductance

Mutual inductance occurs when the magnetic field of an inductor induces a magnetic field in an adjacent inductor. Mutual induction is the basis of transformer construction.

$$M = \sqrt{L_1 L_2}$$

where M is the maximum mutual inductance possible between 2 inductors and L_1 and L_2 are the two inductors. In general

$$M \leq \sqrt{L_1 L_2}$$

as only a fraction of self flux is linked with the other. This fraction is called "Coefficient of flux linkage (K)" or "Coefficient of coupling".

$$M = K \sqrt{L_1 L_2}$$

Inductance formulas

The table below lists some common simplified formulas for calculating the approximate inductance of several inductor constructions.

Construction	Formula	Notes
Cylindrical air-core coil ^[26]	$L = \mu_0 K N^2 \frac{A}{\ell}$ <ul style="list-style-type: none"> ▪ L = inductance in <u>henries</u> (H) ▪ μ_0 = <u>permeability of free space</u> = $4\pi \times 10^{-7}$ H/m ▪ K = Nagaoka coefficient^{[26][a]} ▪ N = number of turns ▪ A = area of cross-section of the coil in square metres (m^2) ▪ ℓ = length of coil in metres (m) 	$K \approx 1$ Calculation of Nagaoka's coefficient (K) is complicated; normally it must be looked up from a table. ^[27]
Straight wire conductor ^[28]	$L = \frac{\mu_0}{2\pi} \ell (A - B) + C,$ <p>where:</p> $A = \ln\left(\frac{\ell}{r} + \sqrt{\left(\frac{\ell}{r}\right)^2 + 1}\right)$ $B = \frac{1}{\frac{r}{\ell} + \sqrt{1 + \left(\frac{r}{\ell}\right)^2}}$ $C = \text{Im} \left(\frac{n\rho J_0(nr)}{2\pi\omega\mu r J_1(nr)} \right)$ <ul style="list-style-type: none"> ▪ L = inductance ▪ ℓ = cylinder length ▪ r = cylinder radius ▪ μ_0 = permeability of free space = $4\pi \times 10^{-7}$ H/m ▪ μ = conductor permeability ▪ ρ = resistivity ▪ ω = angular frequency ▪ $n = \sqrt{-i\frac{\omega\mu}{\rho}} = (-1+i)\sqrt{\frac{\omega\mu}{2\rho}}$ ▪ J_0, J_1 are <u>Bessel functions</u>. ▪ $\frac{\mu_0}{2\pi} = 0.2 \mu\text{H}/\text{m}$, exactly. 	<p>The term C gives the <i>internal</i> inductance of the wire with skin-effect correction (the imaginary part of the internal impedance of the wire). If $\omega = 0$ (DC) then $C = \frac{\mu}{8\pi}$, and as ω approaches ∞, C approaches 0.^[29]</p> <p>The term B subtracts rather than adds.</p>
Small loop or very short coil ^[33]	$L = \frac{\mu_0}{2\pi} \ell \left[\ln\left(\frac{4\ell}{d}\right) - 1 \right] \quad (\text{when } d^2 f \gg 1 \text{ mm}^2 \text{ MHz})$ $L = \frac{\mu_0}{2\pi} \ell \left[\ln\left(\frac{4\ell}{d}\right) - \frac{3}{4} \right] \quad (\text{when } d^2 f \ll 1 \text{ mm}^2 \text{ MHz})$ <ul style="list-style-type: none"> ▪ L = inductance (nH)^{[30][31]} ▪ ℓ = length of conductor (mm) ▪ d = diameter of conductor (mm) ▪ f = frequency ▪ $\frac{\mu_0}{2\pi} = 0.2 \mu\text{H}/\text{m}$, exactly. 	<p>Requires $\ell > 100 d$^[32]</p> <p>For relative permeability $\mu_r = 1$ (e.g., <u>Cu</u> or <u>Al</u>).</p>

	<ul style="list-style-type: none"> ▪ L = inductance in the same units as μ_0. ▪ D = Diameter of the coil (conductor center-to-center) ▪ d = diameter of the conductor ▪ N = number of turns ▪ f = operating frequency (regular f, not ω) ▪ σ = specific conductivity of the coil conductor ▪ μ_r = relative permeability of the conductor ▪ Total conductor length $\ell_c \approx N\pi D$ should be roughly $\frac{1}{10}$ wavelength or smaller.^[34] ▪ Proximity effects are not included: edge-to-edge gap between turns should be $2 \times d$ or larger. ▪ $\frac{\mu_0}{2\pi} = 0.2 \text{ } \mu\text{H/m}$, exactly. 	than a magnetic or paramagnetic metal.
Medium or long air-core cylindrical coil ^{[35][36]}	$L = \frac{r^2 N^2}{23r + 25\ell}$ <ul style="list-style-type: none"> ▪ L = inductance (μH) ▪ r = outer radius of coil (cm) ▪ ℓ = length of coil (cm) ▪ N = number of turns 	Requires cylinder length $\ell > 0.4 r$: Length must be at least $\frac{1}{5}$ of the diameter. Not applicable to single-loop antennas or very short, stubby coils.
Multilayer air-core coil ^[37]	$L = \frac{r^2 N^2}{19r + 29\ell + 32d}$ <ul style="list-style-type: none"> ▪ L = inductance (μH) ▪ r = mean radius of coil (cm) ▪ ℓ = physical length of coil winding (cm) ▪ N = number of turns ▪ d = depth of coil (outer radius minus inner radius) (cm) 	
Flat spiral air-core coil ^{[38][39][40]}	$L = \frac{r^2 N^2}{20r + 28d}$ <ul style="list-style-type: none"> ▪ L = inductance (μH) ▪ r = mean radius of coil (cm) ▪ N = number of turns ▪ d = depth of coil (outer radius minus inner radius) (cm) $L = \frac{r^2 N^2}{8r + 11d}$ <ul style="list-style-type: none"> ▪ L = inductance (μH) ▪ r = mean radius of coil (in) ▪ N = number of turns ▪ d = depth of coil (outer radius minus inner radius) (in) 	Accurate to within 5 percent for $d > 0.2 r$. ^[41]
Toroidal air-core (circular cross-section) ^[42]	$L = 2\pi N^2 \left(D - \sqrt{D^2 - d^2} \right)$ <ul style="list-style-type: none"> ▪ L = inductance (nH) ▪ d = diameter of coil winding (cm) ▪ N = number of turns ▪ D = $2 \times$ radius of revolution (cm) $L \approx \pi \frac{d^2 N^2}{D}$ <ul style="list-style-type: none"> ▪ L = inductance (nH) 	Approximation when $d < 0.1 D$

	<ul style="list-style-type: none"> ▪ d = diameter of coil winding (cm) ▪ N = number of turns ▪ $D = 2 * \text{radius of revolution (cm)}$ 	
Toroidal air-core (rectangular cross-section) ^[41]	$L = 2N^2 h \ln\left(\frac{d_2}{d_1}\right)$ <ul style="list-style-type: none"> ▪ L = inductance (nH) ▪ d_1 = inside diameter of toroid (cm) ▪ d_2 = outside diameter of toroid (cm) ▪ N = number of turns ▪ h = height of toroid (cm) 	

See also

- [Bellini–Tosi direction finder](#) (radio goniometer)
- [Hanna curve](#)
- [Induction coil](#)
- [Induction cooking](#)
- [Induction loop](#)
- [LC circuit](#)
- [RLC circuit](#)
- [Saturable reactor](#) – a type of adjustable inductor
- [Accumulator \(energy\)](#)

Notes

- a. Nagaoka's coefficient (K) is approximately 1 for a coil which is much longer than its diameter and is tightly wound using small gauge wire (so that it approximates a current sheet).

References

1. Alexander, Charles K.; Sadiku, Matthew N. O. (2013). *Fundamentals of Electric Circuits* (5 ed.). McGraw-Hill. p. 226. ISBN 978-0-07-338057-5.
2. Urbanitzky, Alfred Ritter von (1886). *Electricity in the Service of Man*. Macmillan and Company. p. 195 (<https://archive.org/details/electricityinse01wormgoog/page/194/mode/2up>).
3. Singh, Yaduvir (2011). *Electro Magnetic Field Theory* (<https://books.google.com/books?id=0-PfbT49tJMC&q=inductance&pg=PA65>). Pearson India. p. 65. ISBN 978-8131760611.
4. Wadhwa, C. L. (2005). *Electrical Power Systems* (<https://books.google.com/books?id=Su3-0UhVF28C&q=inductance&pg=PA18>). New Age International. p. 18. ISBN 978-8122417227.
5. Pelcovits, Robert A.; Josh Farkas (2007). *Barron's AP Physics C* (<https://books.google.com/books?id=yON684oSjbEC&q=inductance&pg=PA646>). Barron's Educational Series. p. 646. ISBN 978-0764137105.
6. Purcell, Edward M.; David J. Morin (2013). *Electricity and Magnetism* (<https://books.google.com/books?id=A2rS5vISFq0C&pg=PA364>). Cambridge Univ. Press. p. 364. ISBN 978-1107014022.

7. Shamos, Morris H. (2012-10-16). *Great Experiments in Physics: Firsthand Accounts from Galileo to Einstein* (<https://books.google.com/books?id=J0fCAgAAQBAJ&q=1834+Lenz%E2%80%99s+Law&pg=PT238>). Courier Corporation. ISBN 9780486139623.
8. Schmitt, Ron (2002). *Electromagnetics Explained: A Handbook for Wireless/ RF, EMC, and High-Speed Electronics* (<https://books.google.com/books?id=7gJ4RocvEskC&q=%22lenz%27s+law%22+energy&pg=PA75>). Elsevier. pp. 75–77. ISBN 978-0080505237.
9. Jaffe, Robert L.; Taylor, Washington (2018). *The Physics of Energy* (<https://books.google.com/books?id=RMhJDwAAQBAJ&q=%22lenz%27s+law%22+energy+stored+inductor+current&pg=PA51>). Cambridge Univ. Press. p. 51. ISBN 978-1108547895.
10. Lerner, Lawrence S. (1997). *Physics for Scientists and Engineers, Vol. 2* (<https://books.google.com/books?id=Nv5GAyAdijoC&q=inductor+energy+%22magnetic+field%22+current&pg=PA856>). Jones and Bartlet Learning. p. 856. ISBN 978-0763704605.
11. Bowick, Christopher (2011). *RF Circuit Design, 2nd Ed* (<https://books.google.com/books?id=zpTnMsiUkmwC&q=inductor+%22parasitic+capacitance%22&pg=PA7>). Newnes. pp. 7–8. ISBN 978-0080553429.
12. Kaiser, Kenneth L. (2004). *Electromagnetic Compatibility Handbook* (<https://books.google.com/books?id=nZzOAsroBIEC&q=inductor+%22parasitic+capacitance%22>). CRC Press. pp. 6.4 – 6.5. ISBN 978-0849320873.
13. "What is an inductor? What is the definition of an inductor? Overview of inductor characteristics, parameters, and types" (<https://www.gotrend.com.tw/ec99/rwd1606/news.asp?newsno=55>). GOTREND Technology. Retrieved 2010-09-24.
14. Ott, Henry W. (2011). *Electromagnetic Compatibility Engineering* (<https://books.google.com/books?id=2-4WJKxzzigC&q=shielding+transformer+inductor&pg=PA203>). John Wiley and Sons. p. 203. ISBN 978-1118210659.
15. Violette, Norman (2013). *Electromagnetic Compatibility Handbook* (https://books.google.com/books?id=K7_nCAAQBAJ&q=shielding+transformer+inductor&pg=PA516). Springer. pp. 515–516. ISBN 978-9401771443.
16. "An Unassuming Antenna – The Ferrite Loopstick" (<http://radio-timetraveller.blogspot.com/2011/01/unassuming-antenna-ferrite-loopstick.html>). Radio Time Traveller. January 23, 2011. Retrieved March 5, 2014.
17. Frost, Phil (December 23, 2013). "What's an appropriate core material for a loopstick antenna?" (<http://ham.stackexchange.com/questions/1156/whats-an-appropriate-core-material-for-a-loopstick-antenna>). Amateur Radio beta. Stack Exchange, Inc. Retrieved March 5, 2014.
18. Poisel, Richard (2011). *Antenna Systems and Electronic Warfare Applications* (<https://books.google.com/books?id=1YA1NZuo6u0C&q=%22ferrite+rod+loop+antenna&pg=PA280>). Artech House. p. 280. ISBN 978-1608074846.
19. Yadava, R. L. (2011). *Antenna and Wave Propagation* (<https://books.google.com/books?id=MMjtYYrE2r8C&q=%22ferrite+loop+antenna&pg=PA261>). PHI Learning Pvt. Ltd. p. 261. ISBN 978-8120342910.
20. Zurek, Stan (2023). "Skin effect" (http://www.e-magnetica.pl/doku.php/proximity_effect). Encyclopedia Magnetica website. Retrieved 21 May 2024.
21. Kazimierczuk, Marian K. (2011). *High-Frequency Magnetic Components* (<https://books.google.com/books?id=t2TgU-uuNQ0C&pg=PA141>). John Wiley and Sons. ISBN 978-1-119-96491-9.
22. Zurek, Stan (2023). "Proximity effect" (http://www.e-magnetica.pl/doku.php/proximity_effect). Encyclopedia Magnetica website. Retrieved 21 May 2024.
23. "Inductors 101" (<http://www.newark.com/pdfs/techarticles/vishay/Inductors101.pdf>) (PDF). vishay. Retrieved 2010-09-24.
24. "Inductor and Magnetic Product Terminology" (<http://www.vishay.com/docs/34053/definit.pdf>) (PDF). Vishay Dale. Retrieved 2012-09-24.

25. "page with aluminum cores" (<http://www.coilcraft.com/pdfs/uni5.pdf>) (PDF). Coilcraft catalog. Retrieved 10 July 2015.
26. Nagaoka, Hantaro (1909-05-06). "The Inductance Coefficients of Solenoids" (<http://www.g3ynh.info/zdocs/refs/Nagaoka1909.pdf>) (PDF). *Journal of the College of Science, Imperial University, Tokyo, Japan*. 27: 18. Retrieved 2011-11-10.
27. Kenneth L. Kaiser, *Electromagnetic Compatibility Handbook*, p. 30.64, CRC Press, 2004 ISBN 0849320879.
28. Rosa, Edward B. (1908). "The Self and Mutual Inductances of Linear Conductors" (<http://www.g3ynh.info/zdocs/refs/NBS/Rosa1908.pdf>) (PDF). *Bulletin of the Bureau of Standards*. 4 (2): 301–344. doi:10.6028/bulletin.088 (<https://doi.org/10.6028%2Fbulletin.088>).
29. Skilling, Hugh Hildreth (1951). *Electric transmission lines : distributed constants, theory, and application*. McGraw-Hill. pp. 153–159.
30. Rosa 1908, equation (11a), subst. radius $\rho = d/2$ and cgs units
31. Terman 1943, pp. 48–49, convert to natural logarithms and inches to mm.
32. Terman (1943, p. 48) states for $\ell < 100 d$, include $d/2\ell$ within the parentheses.
33. Burger, O. & Dvorský, M. (2015). *Magnetic Loop Antenna*. Ostrava, Czech Republic: EDUCA TV o.p.s.
34. Values of πD up to $\frac{1}{3}$ wavelength are feasible antennas, but for windings that long, this formula will be inaccurate.
35. ARRL Handbook, 66th Ed. American Radio Relay League (1989).
36. "Helical coil calculator" (<https://kaizerpowerelectronics.dk/calculators/helical-coil-calculator/>). Kaizer Power Electronics. 2014-07-09. Retrieved 2020-12-29.
37. Wheeler, H.A. (October 1928). "Simple Inductance Formulas for Radio Coils". *Proceedings of the Institute of Radio Engineers*. 16 (10): 1398. doi:10.1109/JRPROC.1928.221309 (<https://doi.org/10.1109%2FJRPROC.1928.221309>). S2CID 51638679 (<https://api.semanticscholar.org/CorpusID:51638679>).
38. For the second formula, Terman (1943, p. 58) which cites to Wheeler 1928.
39. "A Magnetic Elevator for Neutral Atoms into a 2D State-dependent Optical Lattice Experiment" (<https://web.archive.org/web/20180523063438/http://quantum-technologies.iap.uni-bonn.de/de/diplom-theses.html?task=download&file=302&token=fff191dcc4193aae12b9b5b0e9e199c9>). Uni-Bonn. Archived from the original (<http://quantum-technologies.iap.uni-bonn.de/de/diplom-theses.html?task=download&file=302&token=fff191dcc4193aae12b9b5b0e9e199c9>) on 2018-05-23. Retrieved 2017-08-15.
40. "Spiral coil calculator" (<https://kaizerpowerelectronics.dk/calculators/spiral-coil-calculator/>). Kaizer Power Electronics. 2014-07-10. Retrieved 2020-12-29.
41. Terman 1943, p. 58
42. Terman 1943, p. 57

Source

- Terman, Frederick (1943). *Radio Engineers' Handbook*. McGraw-Hill.

External links

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Inductor&oldid=1291516231>"



Capacitor

In electrical engineering, a **capacitor** is a device that stores electrical energy by accumulating electric charges on two closely spaced surfaces that are insulated from each other. The capacitor was originally known as the **condenser**,^[1] a term still encountered in a few compound names, such as the *condenser microphone*. It is a passive electronic component with two terminals.

The utility of a capacitor depends on its capacitance. While some capacitance exists between any two electrical conductors in proximity in a circuit, a capacitor is a component designed specifically to add capacitance to some part of the circuit.

The physical form and construction of practical capacitors vary widely and many types of capacitor are in common use. Most capacitors contain at least two electrical conductors, often in the form of metallic plates or surfaces separated by a dielectric medium. A conductor may be a foil, thin film, sintered bead of metal, or an electrolyte. The nonconducting dielectric acts to increase the capacitor's charge capacity. Materials commonly used as dielectrics include glass, ceramic, plastic film, paper, mica, air, and oxide layers. When an electric potential difference (a voltage) is applied across the terminals of a capacitor, for example when a capacitor is connected across a battery, an electric field develops across the dielectric, causing a net positive charge to collect on one plate and net negative charge to collect on the other plate. No current actually flows through a perfect dielectric. However, there is a flow of charge through the source circuit. If the condition is maintained sufficiently long, the current through the source circuit ceases. If a time-varying voltage is applied across the leads of the capacitor, the source experiences an ongoing current due to the charging and discharging cycles of the capacitor.

Capacitors are widely used as parts of electrical circuits in many common electrical devices. Unlike a resistor, an ideal capacitor does not dissipate energy, although real-life capacitors do dissipate a small amount ().

The earliest forms of capacitors were created in the 1740s, when European experimenters discovered that electric charge could be stored in water-filled glass jars that came to be known as Leyden jars. Today, capacitors are widely used in electronic circuits for blocking direct current while allowing alternating current to pass. In analog filter networks, they smooth the output of power supplies. In resonant circuits they tune radios to particular frequencies. In electric power transmission systems, they stabilize voltage and power flow.^[2] The property of energy storage in capacitors was exploited as dynamic memory in early digital computers,^[3] and still is in modern DRAM.

Capacitor



<u>Component type</u>	Passive
<u>Working principle</u>	Capacitance
<u>Inventor</u>	Ewald Georg von Kleist (1745) Pieter van Musschenbroek (1746)
<u>Number of terminals</u>	2

Electronic symbol



History

Natural capacitors have existed since prehistoric times. The most common example of natural capacitance are the static charges accumulated between clouds in the sky and the surface of the Earth, where the air between them serves as the dielectric. This results in bolts of lightning when the breakdown voltage of the air is exceeded.^[4]



Battery of four Leyden jars in Museum Boerhaave, Leiden, the Netherlands

In October 1745, Ewald Georg von Kleist of Pomerania, Germany, found that charge could be stored by connecting a high-voltage electrostatic generator by a wire to a volume of water in a hand-held glass jar.^[5] Von Kleist's hand and the water acted as conductors and the jar as a dielectric (although details of the mechanism were incorrectly identified at the time). Von Kleist found that touching the wire resulted in a powerful spark, much more painful than that obtained from an electrostatic machine. The following year, the Dutch physicist Pieter van Musschenbroek invented a similar capacitor, which was named the Leyden jar, after the University of Leiden where he worked.^[6] He also was impressed by the power of the shock he received, writing, "I would not take a second shock for the kingdom of France."^[7]

Daniel Gralath was the first to combine several jars in parallel to increase the charge storage capacity.^[8] Benjamin Franklin investigated the Leyden jar and came to the conclusion that the charge was stored on the glass, not in the water as others had assumed. He also adopted the term "battery",^{[9][10]} (denoting the increase of power with a row of similar units as in a battery of cannon), subsequently applied to clusters of electrochemical cells.^[11] In 1747, Leyden jars were made by coating the inside and outside of jars with metal foil, leaving a space at the mouth to prevent arcing between the foils.^[12] The earliest unit of capacitance was the jar, equivalent to about 1.11 nanofarads.^[13]

Leyden jars or more powerful devices employing flat glass plates alternating with foil conductors were used exclusively up until about 1900, when the invention of wireless (radio) created a demand for standard capacitors, and the steady move to higher frequencies required capacitors with lower inductance. More compact construction methods began to be used, such as a flexible dielectric sheet (like oiled paper) sandwiched between sheets of metal foil, rolled or folded into a small package.

Early capacitors were known as condensers, a term that is still occasionally used today, particularly in high power applications, such as automotive systems. The term condensatore was used by Alessandro Volta in 1780 to refer to a device, similar to his electrophorus, he developed to measure electricity, and translated in 1782 as condenser,^[14] where the name referred to the device's ability to store a higher density of electric charge than was possible with an isolated conductor.^{[15][1]} The term became deprecated because of the ambiguous meaning of steam condenser, with capacitor becoming the recommended term in the UK from 1926,^[16] while the change occurred considerably later in the United States.

Since the beginning of the study of electricity, non-conductive materials like glass, porcelain, paper and mica have been used as insulators. Decades later, these materials were also well-suited for use as the dielectric for the first capacitors. Paper capacitors, made by sandwiching a strip of impregnated paper

between strips of metal and rolling the result into a cylinder, were commonly used in the late 19th century; their manufacture started in 1876,^[17] and they were used from the early 20th century as decoupling capacitors in telephony.

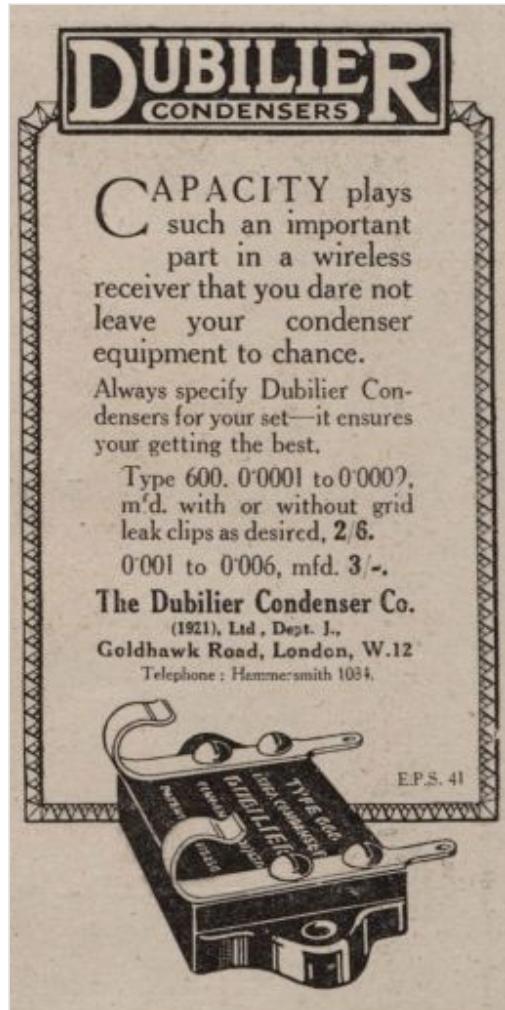
Porcelain was used in the first ceramic capacitors. In the early years of Marconi's wireless transmitting apparatus, porcelain capacitors were used for high voltage and high frequency application in the transmitters. On the receiver side, smaller mica capacitors were used for resonant circuits. Mica capacitors were invented in 1909 by William Dubilier. Prior to World War II, mica was the most common dielectric for capacitors in the United States.^[17]

Charles Pollak (born Karol Pollak), the inventor of the first electrolytic capacitors, found out that the oxide layer on an aluminum anode remained stable in a neutral or alkaline electrolyte, even when the power was switched off. In 1896 he was granted U.S. Patent No. 672,913 for an "Electric liquid capacitor with aluminum electrodes". Solid electrolyte tantalum capacitors were invented by Bell Laboratories in the early 1950s as a miniaturized and more reliable low-voltage support capacitor to complement their newly invented transistor.

With the development of plastic materials by organic chemists during the Second World War, the capacitor industry began to replace paper with thinner polymer films. One very early development in film capacitors was described in British Patent 587,953 in 1944.^[17]

Electric double-layer capacitors (now supercapacitors) were invented in 1957 when H. Becker developed a "Low voltage electrolytic capacitor with porous carbon electrodes".^{[17][18][19]} He believed that the energy was stored as a charge in the carbon pores used in his capacitor as in the pores of the etched foils of electrolytic capacitors. Because the double layer mechanism was not known by him at the time, he wrote in the patent: "It is not known exactly what is taking place in the component if it is used for energy storage, but it leads to an extremely high capacity."

The MOS capacitor was later widely adopted as a storage capacitor in memory chips, and as the basic building block of the charge-coupled device (CCD) in image sensor technology.^[20] In 1966, Dr. Robert Dennard invented modern DRAM architecture, combining a single MOS transistor per capacitor.^{[21][22]}



Advert from the 28 December 1923 edition of The Radio Times for Dubilier condensers, for use in wireless receiving sets

Theory of operation

Overview

A capacitor consists of two conductors separated by a non-conductive region.^[23] The non-conductive region can either be a vacuum or an electrical insulator material known as a dielectric. Examples of dielectric media are glass, air, paper, plastic, ceramic, and even a semiconductor depletion region chemically identical to the conductors. From Coulomb's law a charge on one conductor will exert a force on the charge carriers within the other conductor, attracting opposite polarity charge and repelling like polarity charges, thus an opposite polarity charge will be induced on the surface of the other conductor. The conductors thus hold equal and opposite charges on their facing surfaces,^[24] and the dielectric develops an electric field.

An ideal capacitor is characterized by a constant capacitance C , in farads in the SI system of units, defined as the ratio of the positive or negative charge Q on each conductor to the voltage V between them:^[23]

$$C = \frac{Q}{V}$$

A capacitance of one farad (F) means that one coulomb of charge on each conductor causes a voltage of one volt across the device.^[25] Because the conductors (or plates) are close together, the opposite charges on the conductors attract one another due to their electric fields, allowing the capacitor to store more charge for a given voltage than when the conductors are separated, yielding a larger capacitance.

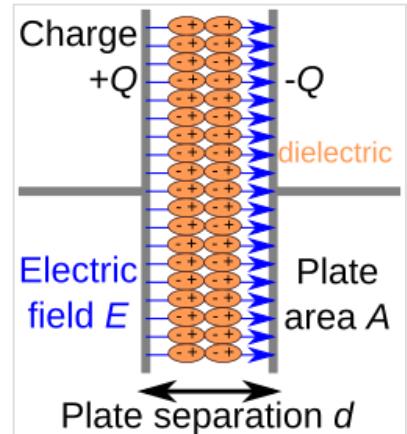
In practical devices, charge build-up sometimes affects the capacitor mechanically, causing its capacitance to vary. In this case, capacitance is defined in terms of incremental changes:

$$C = \frac{dQ}{dV}$$

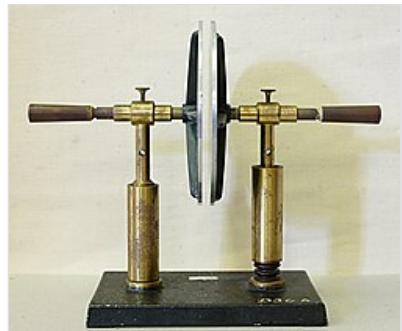
Hydraulic analogy

In the hydraulic analogy, voltage is analogous to water pressure and electrical current through a wire is analogous to water flow through a pipe. A capacitor is like an elastic diaphragm within the pipe. Although water cannot pass through the diaphragm, it moves as the diaphragm stretches or un-stretches.

- Capacitance is analogous to diaphragm elasticity. In the same way that the ratio of charge differential to voltage would be greater for a larger capacitance value



Charge separation in a parallel-plate capacitor causes an internal electric field. A dielectric (orange) reduces the field and increases the capacitance.



A simple demonstration capacitor made of two parallel metal plates, using an air gap as the dielectric



In the hydraulic analogy, a capacitor is analogous to an elastic diaphragm within a pipe. This animation shows a diaphragm being stretched and un-stretched, which is analogous to a capacitor being charged and discharged.

$(C = Q/V)$, the ratio of water displacement to pressure would be greater for a diaphragm that flexes more readily.

- In an AC circuit, a capacitor behaves like a diaphragm in a pipe, allowing the charge to move on both sides of the dielectric while no electrons actually pass through. For DC circuits, a capacitor is analogous to a hydraulic accumulator, storing the energy until pressure is released. Similarly, they can be used to smooth the flow of electricity in rectified DC circuits in the same way an accumulator damps surges from a hydraulic pump.
- Charged capacitors and stretched diaphragms both store potential energy. The more a capacitor is charged, the higher the voltage across the plates ($V = Q/C$). Likewise, the greater the displaced water volume, the greater the elastic potential energy.
- Electrical current affects the charge differential across a capacitor just as the flow of water affects the volume differential across a diaphragm.
- Just as capacitors experience dielectric breakdown when subjected to high voltages, diaphragms burst under extreme pressures.
- Just as capacitors block DC while passing AC, diaphragms displace no water unless there is a change in pressure.

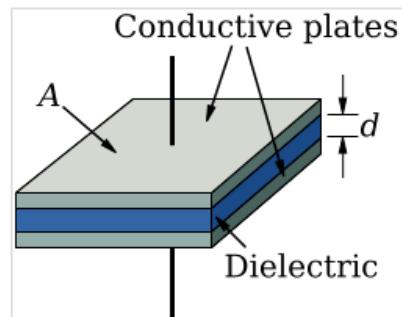
Circuit equivalence at short-time limit and long-time limit

In a circuit, a capacitor can behave differently at different time instants. However, it is usually easy to think about the short-time limit and long-time limit:

- In the long-time limit, after the charging/discharging current has saturated the capacitor, no current would come into (or get out of) either side of the capacitor; Therefore, the long-time equivalence of capacitor is an open circuit.
- In the short-time limit, if the capacitor starts with a certain voltage V , since the voltage drop on the capacitor is known at this instant, we can replace it with an ideal voltage source of voltage V . Specifically, if $V=0$ (capacitor is uncharged), the short-time equivalence of a capacitor is a short circuit.

Parallel-plate capacitor

The simplest model of a capacitor consists of two thin parallel conductive plates each with an area of A separated by a uniform gap of thickness d filled with a dielectric of permittivity ϵ . It is assumed the gap d is much smaller than the dimensions of the plates. This model applies well to many practical capacitors which are constructed of metal sheets separated by a thin layer of insulating dielectric, since manufacturers try to keep the dielectric very uniform in thickness to avoid thin spots which can cause failure of the capacitor.



Parallel plate capacitor model consists of two conducting plates, each of area A , separated by a gap of thickness d containing a dielectric.

Since the separation between the plates is uniform over the plate area, the electric field between the plates E is constant, and directed perpendicularly to the plate surface, except for an area near the edges of the plates where the field decreases because the electric field lines "bulge" out of the sides of the capacitor. This "fringing field" area is approximately the same width as the plate separation, d , and assuming d is small compared to the plate dimensions, it is small enough to be ignored. Therefore, if a charge of $+Q$ is placed on one plate and $-Q$ on the other plate (the situation for unevenly charged plates is discussed below), the charge on each plate will be spread evenly in a surface charge layer of constant charge density $\sigma = \pm Q/A$ coulombs per square

meter, on the inside surface of each plate. From Gauss's law the magnitude of the electric field between the plates is $E = \sigma/\epsilon$. The voltage(difference) V between the plates is defined as the line integral of the electric field over a line (in the z-direction) from one plate to another

$$V = \int_0^d E(z) dz = Ed = \frac{\sigma}{\epsilon} d = \frac{Qd}{\epsilon A}$$

The capacitance is defined as $C = Q/V$. Substituting V above into this equation

$$C = \frac{\epsilon A}{d}$$

Therefore, in a capacitor the highest capacitance is achieved with a high permittivity dielectric material, large plate area, and small separation between the plates.

Since the area A of the plates increases with the square of the linear dimensions and the separation d increases linearly, the capacitance scales with the linear dimension of a capacitor ($C \propto L$), or as the cube root of the volume.

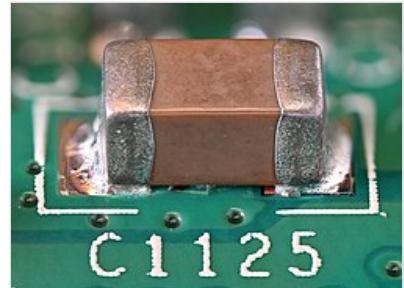
A parallel plate capacitor can only store a finite amount of energy before dielectric breakdown occurs. The capacitor's dielectric material has a dielectric strength U_d which sets the capacitor's breakdown voltage at $V = V_{bd} = U_d d$. The maximum energy that the capacitor can store is therefore

$$E = \frac{1}{2} CV^2 = \frac{1}{2} \frac{\epsilon A}{d} (U_d d)^2 = \frac{1}{2} \epsilon A d U_d^2$$

The maximum energy is a function of dielectric volume, permittivity, and dielectric strength. Changing the plate area and the separation between the plates while maintaining the same volume causes no change of the maximum amount of energy that the capacitor can store, so long as the distance between plates remains much smaller than both the length and width of the plates. In addition, these equations assume that the electric field is entirely concentrated in the dielectric between the plates. In reality there are fringing fields outside the dielectric, for example between the sides of the capacitor plates, which increase the effective capacitance of the capacitor. This is sometimes called parasitic capacitance. For some simple capacitor geometries this additional capacitance term can be calculated analytically.^[26] It becomes negligibly small when the ratios of plate width to separation and length to separation are large.

For unevenly charged plates:

- If one plate is charged with Q_1 while the other is charged with Q_2 , and if both plates are separated from other materials in the environment, then the inner surface of the first plate will have $\frac{Q_1 - Q_2}{2}$, and the inner surface of the second plated will have $-\frac{Q_1 - Q_2}{2}$ charge. Therefore, the voltage V between the plates is $V = \frac{Q_1 - Q_2}{2C}$. Note that the outer surface of both plates will have $\frac{Q_1 + Q_2}{2}$, but those charges do not affect the voltage between the plates.
- If one plate is charged with Q_1 while the other is charged with Q_2 , and if the second plate is connected to ground, then the inner surface of the first plate will have Q_1 , and the inner



A surface-mount capacitor. The plates, not visible, are layered horizontally between ceramic dielectric layers, and connect alternately to either end-cap, which are visible.

surface of the second plated will have $-Q_1$. Therefore, the voltage V between the plates is $V = \frac{Q_1}{C}$. Note that the outer surface of both plates will have zero charge.

Interleaved capacitor

For n number of plates in a capacitor, the total capacitance would be

$$C = \epsilon_0 \frac{A}{d} (n - 1)$$

where $C = \epsilon_0 A/d$ is the capacitance for a single plate and n is the number of interleaved plates.

As shown to the figure on the right, the interleaved plates can be seen as parallel plates connected to each other. Every pair of adjacent plates acts as a separate capacitor; the number of pairs is always one less than the number of plates, hence the $(n - 1)$ multiplier.

Energy stored in a capacitor

To increase the charge and voltage on a capacitor, work must be done by an external power source to move charge from the negative to the positive plate against the opposing force of the electric field.^{[27][28]} If the voltage on the capacitor is V , the work dW required to move a small increment of charge dq from the negative to the positive plate is $dW = Vdq$. The energy is stored in the increased electric field between the plates. The total energy W stored in a capacitor (expressed in joules) is equal to the total work done in establishing the electric field from an uncharged state.^{[29][28][27]}

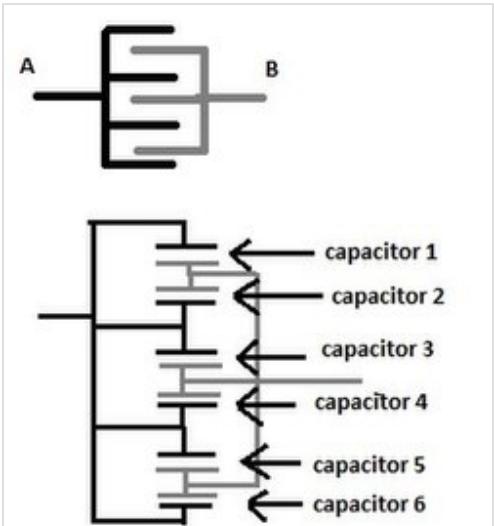
$$W = \int_0^Q V(q) dq = \int_0^Q \frac{q}{C} dq = \frac{1}{2} \frac{Q^2}{C} = \frac{1}{2} VQ = \frac{1}{2} CV^2$$

where Q is the charge stored in the capacitor, V is the voltage across the capacitor, and C is the capacitance. This potential energy will remain in the capacitor until the charge is removed. If charge is allowed to move back from the positive to the negative plate, for example by connecting a circuit with resistance between the plates, the charge moving under the influence of the electric field will do work on the external circuit.

If the gap between the capacitor plates d is constant, as in the parallel plate model above, the electric field between the plates will be uniform (neglecting fringing fields) and will have a constant value $E = V/d$. In this case the stored energy can be calculated from the electric field strength

$$W = \frac{1}{2} CV^2 = \frac{1}{2} \frac{\epsilon A}{d} (Ed)^2 = \frac{1}{2} \epsilon AdE^2 = \frac{1}{2} \epsilon E^2 (\text{volume of electric field})$$

The last formula above is equal to the energy density per unit volume in the electric field multiplied by the volume of field between the plates, confirming that the energy in the capacitor is stored in its electric field.



The interleaved capacitor can be seen as combination of several parallel connected capacitors.

The interleaved capacitor can be seen as a combination of several parallel connected capacitors.

Current–voltage relation

The current $I(t)$ through any component in an electric circuit is defined as the rate of flow of a charge $Q(t)$ passing through it. Actual charges – electrons – cannot pass through the dielectric of an *ideal* capacitor.^[note 1] Rather, one electron accumulates on the negative plate for each one that leaves the positive plate, resulting in an electron depletion and consequent positive charge on one electrode that is equal and opposite to the accumulated negative charge on the other. Thus the charge on the electrodes is equal to the integral of the current as well as proportional to the voltage, as discussed above. As with any antiderivative, a constant of integration is added to represent the initial voltage $V(t_0)$. This is the integral form of the capacitor equation:^[30]

$$V(t) = \frac{Q(t)}{C} = V(t_0) + \frac{1}{C} \int_{t_0}^t I(\tau) d\tau$$

Taking the derivative of this and multiplying by C yields the derivative form:^[31]

$$I(t) = \frac{dQ(t)}{dt} = C \frac{dV(t)}{dt}$$

for C independent of time, voltage and electric charge.

The dual of the capacitor is the inductor, which stores energy in a magnetic field rather than an electric field. Its current–voltage relation is obtained by exchanging current and voltage in the capacitor equations and replacing C with the inductance L .

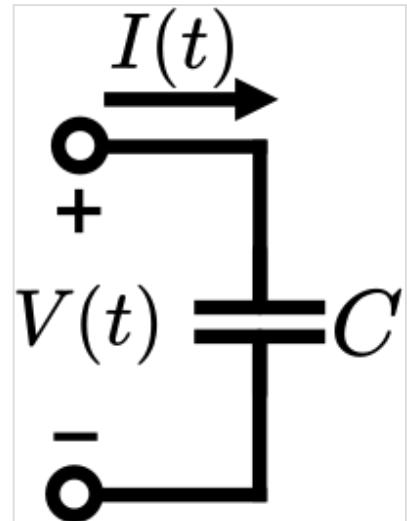
RC circuits

A series circuit containing only a resistor, a capacitor, a switch and a constant DC source of voltage V_0 is known as a *charging circuit*.^[32] If the capacitor is initially uncharged while the switch is open, and the switch is closed at $t = 0$, it follows from Kirchhoff's voltage law that

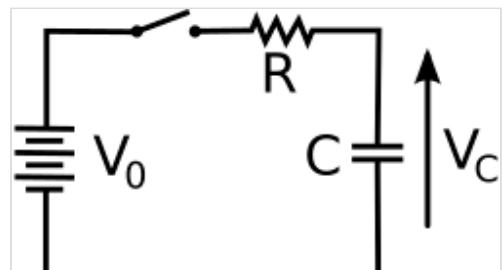
$$V_0 = v_{\text{resistor}}(t) + v_{\text{capacitor}}(t) = i(t)R + \frac{1}{C} \int_{t_0}^t i(\tau) d\tau$$

Taking the derivative and multiplying by C , gives a first-order differential equation:

$$RC \frac{di(t)}{dt} + i(t) = 0$$



Schematic showing polarity of voltage and direction of current for this current–voltage relation



A simple resistor–capacitor circuit demonstrates charging of a capacitor.

At $t = 0$, the voltage across the capacitor is zero and the voltage across the resistor is V_0 . The initial current is then $I(0) = V_0/R$. With this assumption, solving the differential equation yields

$$I(t) = \frac{V_0}{R} e^{-t/\tau_0}$$

$$V(t) = V_0 \left(1 - e^{-t/\tau_0}\right)$$

$$Q(t) = CV_0 \left(1 - e^{-t/\tau_0}\right)$$

where $\tau_0 = RC$ is the time constant of the system. As the capacitor reaches equilibrium with the source voltage, the voltages across the resistor and the current through the entire circuit decay exponentially. In the case of a *discharging* capacitor, the capacitor's initial voltage (V_{Ci}) replaces V_0 . The equations become

$$I(t) = \frac{V_{Ci}}{R} e^{-t/\tau_0}$$

$$V(t) = V_{Ci} e^{-t/\tau_0}$$

$$Q(t) = CV_{Ci} e^{-t/\tau_0}$$

AC circuits

Impedance, the vector sum of reactance and resistance, describes the phase difference and the ratio of amplitudes between sinusoidally varying voltage and sinusoidally varying current at a given frequency. Fourier analysis allows any signal to be constructed from a spectrum of frequencies, whence the circuit's reaction to the various frequencies may be found. The reactance and impedance of a capacitor are respectively

$$X = -\frac{1}{\omega C} = -\frac{1}{2\pi f C}$$

$$Z = \frac{1}{j\omega C} = -\frac{j}{\omega C} = -\frac{j}{2\pi f C}$$

where j is the imaginary unit and ω is the angular frequency of the sinusoidal signal. The $-j$ phase indicates that the AC voltage $V = ZI$ lags the AC current by 90° : the positive current phase corresponds to increasing voltage as the capacitor charges; zero current corresponds to instantaneous constant voltage, etc.

Impedance decreases with increasing capacitance and increasing frequency.^[33] This implies that a higher-frequency signal or a larger capacitor results in a lower voltage amplitude per current amplitude – an AC "short circuit" or AC coupling. Conversely, for very low frequencies, the reactance is high, so that a capacitor is nearly an open circuit in AC analysis – those frequencies have been "filtered out".

Capacitors are different from resistors and inductors in that the impedance is *inversely* proportional to the defining characteristic; i.e., capacitance.

A capacitor connected to an alternating voltage source has a displacement current flowing through it. In the case that the voltage source is $V_0 \cos(\omega t)$, the displacement current can be expressed as:

$$I = C \frac{dV}{dt} = -\omega CV_0 \sin(\omega t)$$

At $\sin(\omega t) = -1$, the capacitor has a maximum (or peak) current whereby $I_0 = \omega CV_0$. The ratio of peak voltage to peak current is due to capacitive reactance (denoted X_C).

$$X_C = \frac{V_0}{I_0} = \frac{V_0}{\omega CV_0} = \frac{1}{\omega C}$$

X_C approaches zero as ω approaches infinity. If X_C approaches 0, the capacitor resembles a short wire that strongly passes current at high frequencies. X_C approaches infinity as ω approaches zero. If X_C approaches infinity, the capacitor resembles an open circuit that poorly passes low frequencies.

The current of the capacitor may be expressed in the form of cosines to better compare with the voltage of the source:

$$I = -I_0 \sin(\omega t) = I_0 \cos(\omega t + 90^\circ)$$

In this situation, the current is out of phase with the voltage by $+\pi/2$ radians or $+90$ degrees, i.e. the current leads the voltage by 90° .

Laplace circuit analysis (s-domain)

When using the Laplace transform in circuit analysis, the impedance of an ideal capacitor with no initial charge is represented in the s domain by:

$$Z(s) = \frac{1}{sC}$$

where

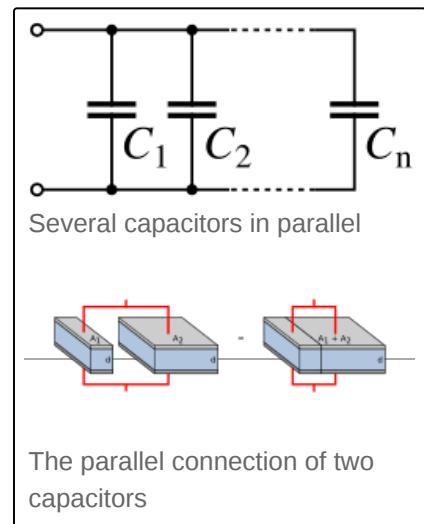
- C is the capacitance, and
- s is the complex frequency.

Circuit analysis

Capacitors in parallel

Capacitors in a parallel configuration each have the same applied voltage. Their capacitance values add up. Charge is apportioned among them by capacitance value. Using the schematic diagram to visualize parallel plates, it is apparent that each capacitor contributes to the total surface area.

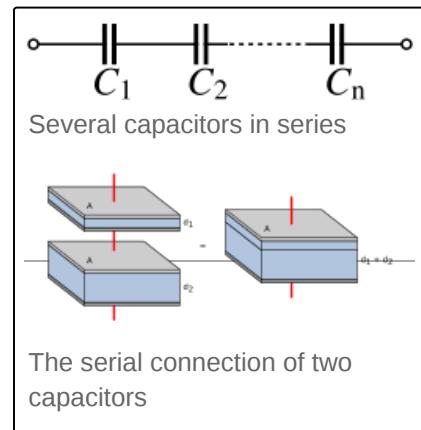
$$C_{eq} = \sum_{i=1}^n C_i = C_1 + C_2 + \dots + C_n$$



For capacitors in series

Connected in series, the schematic diagram reveals that the separation distance, not the plate area, adds up. The capacitors each store instantaneous charge build-up equal to that of every other capacitor in the series. The total voltage difference from end to end is

apportioned to each capacitor according to the inverse of its capacitance. The entire series acts as a capacitor smaller than any of its components.



$$C_{\text{eq}} = \left(\sum_{i=1}^n \frac{1}{C_i} \right)^{-1} = \left(\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots + \frac{1}{C_n} \right)^{-1}$$

Capacitors are combined in series to achieve a higher working voltage, for example for smoothing a high voltage power supply. The voltage ratings, which are based on plate separation, add up, if capacitance and leakage currents for each capacitor are identical. In such an application, on occasion, series strings are connected in parallel, forming a matrix. The goal is to maximize the energy storage of the network without overloading any capacitor. For high-energy storage with capacitors in series, some safety considerations must be applied to ensure one capacitor failing and leaking current does not apply too much voltage to the other series capacitors.

Series connection is also sometimes used to adapt polarized electrolytic capacitors for bipolar AC use.

Voltage distribution in parallel-to-series networks.

To model the distribution of voltages from a single charged capacitor (**A**) connected in parallel to a chain of capacitors in series (**B_n**):

$$\begin{aligned} (\text{volts}) A_{\text{eq}} &= A \left(1 - \frac{1}{n+1} \right) \\ (\text{volts}) B_{1..n} &= \frac{A}{n} \left(1 - \frac{1}{n+1} \right) \\ A - B &= 0 \end{aligned}$$

Note: This is only correct if all capacitance values are equal.

The power transferred in this arrangement is:

$$P = \frac{1}{R} \cdot \frac{1}{n+1} A_{\text{volts}} (A_{\text{farads}} + B_{\text{farads}})$$

Non-ideal behavior

In practice, capacitors deviate from the ideal capacitor equation in several aspects. Some of these, such as leakage current and parasitic effects are linear, or can be analyzed as nearly linear, and can be accounted for by adding virtual components to form an equivalent circuit. The usual methods of network analysis can then be applied.^[34] In other cases, such as with breakdown voltage, the effect is non-linear and ordinary (normal, e.g., linear) network analysis cannot be used, the effect must be considered separately. Yet another group of artifacts may exist, including temperature dependence, that may be linear but invalidates the assumption in the analysis that capacitance is a constant. Finally, combined parasitic effects such as inherent inductance, resistance, or dielectric losses can exhibit non-uniform behavior at varying frequencies of operation.

Breakdown voltage

Above a particular electric field strength, known as the dielectric strength E_{ds} , the dielectric in a capacitor becomes conductive. The voltage at which this occurs is called the breakdown voltage of the device, and is given by the product of the dielectric strength and the separation between the conductors,^[35]

$$V_{bd} = E_{ds}d$$

The maximum energy that can be stored safely in a capacitor is limited by the breakdown voltage. Exceeding this voltage can result in a short circuit between the plates, which can often cause permanent damage to the dielectric, plates, or both. Due to the scaling of capacitance and breakdown voltage with dielectric thickness, all capacitors made with a particular dielectric have approximately equal maximum energy density, to the extent that the dielectric dominates their volume.^[36]

For air dielectric capacitors the breakdown field strength is of the order 2–5 MV/m (or kV/mm); for mica the breakdown is 100–300 MV/m; for oil, 15–25 MV/m; it can be much less when other materials are used for the dielectric.^[37] The dielectric is used in very thin layers and so absolute breakdown voltage of capacitors is limited. Typical ratings for capacitors used for general electronics applications range from a few volts to 1 kV. As the voltage increases, the dielectric must be thicker, making high-voltage capacitors larger per capacitance than those rated for lower voltages.

The breakdown voltage is critically affected by factors such as the geometry of the capacitor conductive parts; sharp edges or points increase the electric field strength at that point and can lead to a local breakdown. Once this starts to happen, the breakdown quickly tracks through the dielectric until it reaches the opposite plate, leaving carbon behind and causing a short (or relatively low resistance) circuit. The results can be explosive, as the short in the capacitor draws current from the surrounding circuitry and dissipates the energy.^[38] However, in capacitors with particular dielectrics^{[39][40]} and thin metal electrodes, shorts are not formed after breakdown. It happens because a metal melts or evaporates in a breakdown vicinity, isolating it from the rest of the capacitor.^{[41][42]}

The usual breakdown route is that the field strength becomes large enough to pull electrons in the dielectric from their atoms thus causing conduction. Other scenarios are possible, such as impurities in the dielectric, and, if the dielectric is of a crystalline nature, imperfections in the crystal structure can result in an avalanche breakdown as seen in semi-conductor devices. Breakdown voltage is also affected by pressure, humidity and temperature.^[43]

Equivalent circuit

An ideal capacitor only stores and releases electrical energy, without dissipation. In practice, capacitors have imperfections within the capacitor's materials that result in the following parasitic components.^[44]

- **ESL**, the *equivalent series inductance*, due to the leads. This is usually significant only at relatively high frequencies.
- Two resistances that add a *real-valued* component to the total impedance, which wastes power:
 - R_{lead} , a small series resistance in the leads. Becomes more relevant as frequency increases.
 - $G_{dielectric}$, a small *conductance* (or reciprocally, a large resistance) in parallel with the capacitance, to account for imperfect dielectric material. This causes a small leakage current across the dielectric (see § *Leakage*)^[45] that slowly discharges the capacitor over

time. This conductance dominates the total resistance at very low frequencies. Its value varies greatly depending on the capacitor material and quality.

Simplified RLC series model

As frequency increases, the capacitive impedance (a negative reactance) reduces, so the dielectric's conductance becomes less important and the series components become more significant. Thus, a simplified RLC series model valid for a large frequency range simply treats the capacitor as being in series with an equivalent series inductance **ESL** and a frequency-dependent equivalent series resistance **ESR**, which varies little with frequency. Unlike the previous model, this model is not valid at DC and very low frequencies where **G_{dielectric}** is relevant.

Inductive reactance increases with frequency. Because its sign is positive, it counteracts the capacitance.

At the RLC circuit's natural frequency $\omega_0 = \frac{1}{\sqrt{\text{ESL} \cdot C}}$, the inductance perfectly cancels the capacitance, so total reactance is zero. Since the total impedance at ω_0 is just the real-value of **ESR**, average power dissipation reaches its maximum of $\frac{V_{\text{RMS}}^2}{\text{ESR}}$, where V_{RMS} is the root mean square (RMS) voltage across the capacitor.

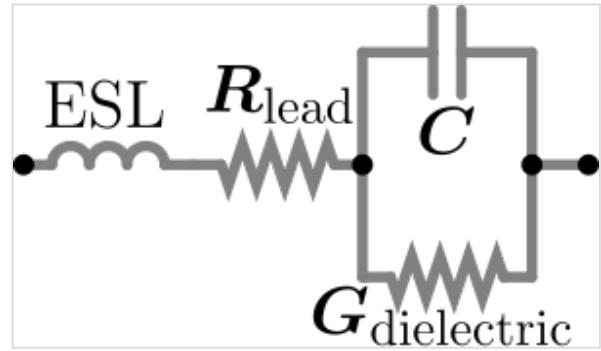
At even higher frequencies, the inductive impedance dominates, so the capacitor undesirably behaves instead like an inductor. High-frequency engineering involves accounting for the inductance of all connections and components.

Q factor

For a simplified model of a capacitor as an ideal capacitor in series with an equivalent series resistance **ESR**, the capacitor's quality factor (or **Q**) is the ratio of the magnitude of its capacitive reactance **X_C** to its resistance at a given frequency **ω** :

$$Q(\omega) = \frac{|X_C(\omega)|}{\text{ESR}} = \frac{1}{\omega C \cdot \text{ESR}}.$$

The Q factor is a measure of its efficiency: the higher the Q factor of the capacitor, the closer it approaches the behavior of an ideal capacitor. Dissipation factor is its reciprocal.



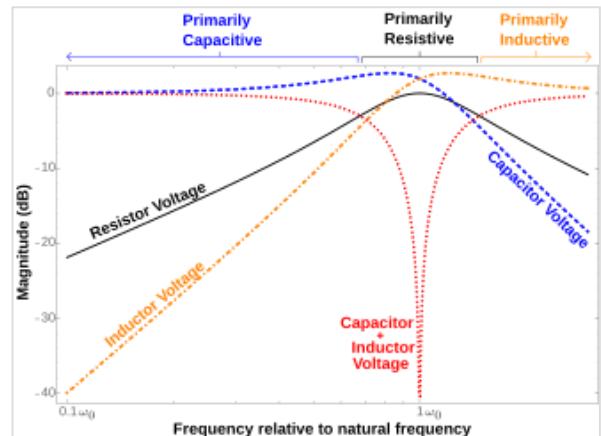
Real capacitor model that adds an inductance and resistance in series and a conductance in parallel to its capacitance. Its total impedance is:

$$Z_\Sigma = Z_{\text{ESL}} + R_{\text{lead}} + (Z_C \parallel G_{\text{dielectric}}) = j\omega \cdot \text{ESL} + R_{\text{lead}} + \frac{1}{j\omega \cdot C + G_{\text{dielectric}}}.$$



Simplified RLC series capacitor model. Its total equivalent impedance is:

$$j\omega \cdot \text{ESL} + \text{ESR} - \frac{j}{\omega \cdot C}.$$



Bode magnitude plot of voltages in an RLC circuit.

Frequency is relative to the natural frequency ω_0 . (Its damping ratio ζ and ω_0 would depend on the particular capacitor.) Lower frequencies are more capacitive. Around ω_0 , the total impedance and voltage drop is primarily resistive. Higher frequencies are more inductive.

Ripple current

Ripple current is the AC component of an applied source (often a switched-mode power supply) whose frequency may be constant or varying. Ripple current causes heat to be generated within the capacitor due to the dielectric losses caused by the changing field strength together with the current flow across the slightly resistive supply lines or the electrolyte in the capacitor. The equivalent series resistance (ESR) is the amount of internal series resistance one would add to a perfect capacitor to model this.

Some types of capacitors, primarily tantalum and aluminum electrolytic capacitors, as well as some film capacitors have a specified rating value for maximum ripple current.

- Tantalum electrolytic capacitors with solid manganese dioxide electrolyte are limited by ripple current and generally have the highest ESR ratings in the capacitor family. Exceeding their ripple limits can lead to shorts and burning parts.
- Aluminum electrolytic capacitors, the most common type of electrolytic, suffer a shortening of life expectancy at higher ripple currents. If ripple current exceeds the rated value of the capacitor, it tends to result in explosive failure.
- Ceramic capacitors generally have no ripple current limitation and have some of the lowest ESR ratings.
- Film capacitors have very low ESR ratings but exceeding rated ripple current may cause degradation failures.

Capacitance instability

The capacitance of certain capacitors decreases as the component ages. In ceramic capacitors, this is caused by degradation of the dielectric. The type of dielectric, ambient operating and storage temperatures are the most significant aging factors, while the operating voltage usually has a smaller effect, i.e., usual capacitor design is to minimize voltage coefficient. The aging process may be reversed by heating the component above the Curie point. Aging is fastest near the beginning of life of the component, and the device stabilizes over time.^[46] Electrolytic capacitors age as the electrolyte evaporates. In contrast with ceramic capacitors, this occurs towards the end of life of the component.

Temperature dependence of capacitance is usually expressed in parts per million (ppm) per °C. It can usually be taken as a broadly linear function but can be noticeably non-linear at the temperature extremes. The temperature coefficient may be positive or negative, depending mostly on the dielectric material. Some, designated C0G/NP0, but called **NPO**, have a somewhat negative coefficient at one temperature, positive at another, and zero in between. Such components may be specified for temperature-critical circuits.^[47]

Capacitors, especially ceramic capacitors, and older designs such as paper capacitors, can absorb sound waves resulting in a microphonic effect. Vibration moves the plates, causing the capacitance to vary, in turn inducing AC current. Some dielectrics also generate piezoelectricity. The resulting interference is especially problematic in audio applications, potentially causing feedback or unintended recording. In the reverse microphonic effect, the varying electric field between the capacitor plates exerts a physical force, moving them as a speaker. This can generate audible sound, but drains energy and stresses the dielectric and the electrolyte, if any.

Current and voltage reversal

Current reversal occurs when the current changes direction. Voltage reversal is the change of polarity in a circuit. Reversal is generally described as the percentage of the maximum rated voltage that reverses polarity. In DC circuits, this is usually less than 100%, often in the range of 0 to 90%, whereas AC circuits experience 100% reversal.

In DC circuits and pulsed circuits, current and voltage reversal are affected by the damping of the system. Voltage reversal is encountered in RLC circuits that are underdamped. The current and voltage reverse direction, forming a harmonic oscillator between the inductance and capacitance. The current and voltage tends to oscillate and may reverse direction several times, with each peak being lower than the previous, until the system reaches an equilibrium. This is often referred to as ringing. In comparison, critically damped or overdamped systems usually do not experience a voltage reversal. Reversal is also encountered in AC circuits, where the peak current is equal in each direction.

For maximum life, capacitors usually need to be able to handle the maximum amount of reversal that a system may experience. An AC circuit experiences 100% voltage reversal, while underdamped DC circuits experience less than 100%. Reversal creates excess electric fields in the dielectric, causes excess heating of both the dielectric and the conductors, and can dramatically shorten the life expectancy of the capacitor. Reversal ratings often affect the design considerations for the capacitor, from the choice of dielectric materials and voltage ratings to the types of internal connections used.^[48]

Dielectric absorption

Capacitors made with any type of dielectric material show some level of "dielectric absorption" or "soakage". On discharging a capacitor and disconnecting it, after a short time it may develop a voltage due to hysteresis in the dielectric. This effect is objectionable in applications such as precision sample and hold circuits or timing circuits. The level of absorption depends on many factors, from design considerations to charging time, since the absorption is a time-dependent process. However, the primary factor is the type of dielectric material. Capacitors such as tantalum electrolytic or polysulfone film exhibit relatively high absorption, while polystyrene or Teflon allow very small levels of absorption.^[49] In some capacitors where dangerous voltages and energies exist, such as in flashtubes, television sets, microwave ovens and defibrillators, the dielectric absorption can recharge the capacitor to hazardous voltages after it has been shorted or discharged. Any capacitor containing over 10 joules of energy is generally considered hazardous, while 50 joules or higher is potentially lethal. A capacitor may regain anywhere from 0.01 to 20% of its original charge over a period of several minutes, allowing a seemingly safe capacitor to become surprisingly dangerous.^{[50][51][52][53][54]}

Leakage

No material is a perfect insulator, thus all dielectrics allow some small level of current to leak through, which can be measured with a megohmmeter.^[55] Leakage is equivalent to a resistor in parallel with the capacitor. Constant exposure to factors such as heat, mechanical stress, or humidity can cause the dielectric to deteriorate resulting in excessive leakage, a problem often seen in older vacuum tube circuits, particularly where oiled paper and foil capacitors were used. In many vacuum tube circuits, interstage coupling capacitors are used to conduct a varying signal from the plate of one tube to the grid circuit of the next stage. A leaky capacitor can cause the grid circuit voltage to be raised from its normal bias setting, causing excessive current or signal distortion in the downstream tube. In power amplifiers this can cause the plates

to glow red, or current limiting resistors to overheat, even fail. Similar considerations apply to component fabricated solid-state (transistor) amplifiers, but, owing to lower heat production and the use of modern polyester dielectric-barriers, this once-common problem has become relatively rare.

Electrolytic failure from disuse

Aluminum electrolytic capacitors are *conditioned* when manufactured by applying a voltage sufficient to initiate the proper internal chemical state. This state is maintained by regular use of the equipment. If a system using electrolytic capacitors is unused for a long period of time it can lose its conditioning. Sometimes they fail with a short circuit when next operated.

Lifespan

All capacitors have varying lifespans, depending upon their construction, operational conditions, and environmental conditions. Solid-state ceramic capacitors generally have very long lives under normal use, which has little dependency on factors such as vibration or ambient temperature, but factors like humidity, mechanical stress, and fatigue play a primary role in their failure. Failure modes may differ. Some capacitors may experience a gradual loss of capacitance, increased leakage or an increase in equivalent series resistance (ESR), while others may fail suddenly or even catastrophically. For example, metal-film capacitors are more prone to damage from stress and humidity, but will self-heal when a breakdown in the dielectric occurs. The formation of a glow discharge at the point of failure prevents arcing by vaporizing the metallic film in that spot, neutralizing any short circuit with minimal loss in capacitance. When enough pinholes accumulate in the film, a total failure occurs in a metal-film capacitor, generally happening suddenly without warning.

Electrolytic capacitors generally have the shortest lifespans. Electrolytic capacitors are affected very little by vibration or humidity, but factors such as ambient and operational temperatures play a large role in their failure, which gradually occur as an increase in ESR (up to 300%) and as much as a 20% decrease in capacitance. The capacitors contain electrolytes which will eventually diffuse through the seals and evaporate. An increase in temperature also increases internal pressure, and increases the reaction rate of the chemicals. Thus, the life of an electrolytic capacitor is generally defined by a modification of the Arrhenius equation, which is used to determine chemical-reaction rates:

$$L = Be^{\frac{e_A}{kT_0}}$$

Manufacturers often use this equation to supply an expected lifespan, in hours, for electrolytic capacitors when used at their designed operating temperature, which is affected by both ambient temperature, ESR, and ripple current. However, these ideal conditions may not exist in every use. The rule of thumb for predicting lifespan under different conditions of use is determined by:

$$L_a = L_0 2^{\frac{T_0 - T_a}{10}}$$

This says that the capacitor's life decreases by half for every 10 degrees Celsius that the temperature is increased,^[56] where:

- L_0 is the rated life under rated conditions, e.g. 2000 hours
- T_0 is the rated max/min operational temperature
- T_a is the average operational temperature
- L_a is the expected lifespan under given conditions

Capacitor types

Practical capacitors are available commercially in many different forms. The type of internal dielectric, the structure of the plates and the device packaging all strongly affect the characteristics of the capacitor, and its applications.

Values available range from very low (picofarad range; while arbitrarily low values are in principle possible, stray (parasitic) capacitance in any circuit is the limiting factor) to about 5 kF supercapacitors.

Above approximately 1 microfarad electrolytic capacitors are usually used because of their small size and low cost compared with other types, unless their relatively poor stability, life and polarised nature make them unsuitable. Very high capacity supercapacitors use a porous carbon-based electrode material.

Dielectric materials

Most capacitors have a dielectric spacer, which increases their capacitance compared to air or a vacuum. In order to maximise the charge that a capacitor can hold, the dielectric material needs to have as high a permittivity as possible, while also having as high a breakdown voltage as possible. The dielectric also needs to have as low a loss with frequency as possible.

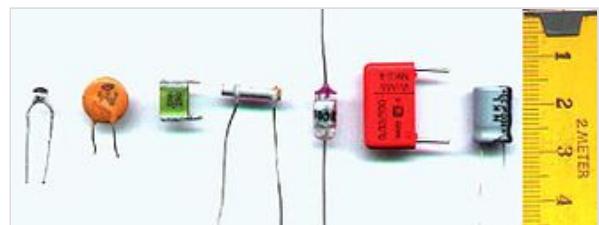
However, low value capacitors are available with a high vacuum between their plates to allow extremely high voltage operation and low losses. Variable capacitors with their plates open to the atmosphere were commonly used in radio tuning circuits. Later designs use polymer foil dielectric between the moving and stationary plates, with no significant air space between the plates.

Several solid dielectrics are available, including paper, plastic, glass, mica and ceramic.^[17]

Paper was used extensively in older capacitors and offers relatively high voltage performance. However, paper absorbs moisture, and has been largely replaced by plastic film capacitors.

Most of the plastic films now used offer better stability and ageing performance than such older dielectrics such as oiled paper, which makes them useful in timer circuits, although they may be limited to relatively low operating temperatures and frequencies, because of the limitations of the plastic film being used. Large plastic film capacitors are used extensively in suppression circuits, motor start circuits, and power-factor correction circuits.

Ceramic capacitors are generally small, cheap and useful for high frequency applications, although their capacitance varies strongly with voltage and temperature and they age poorly. They can also suffer from the piezoelectric effect. Ceramic capacitors are broadly categorized as class 1 dielectrics, which have predictable variation of capacitance with temperature or class 2 dielectrics, which can operate at higher voltage. Modern multilayer ceramics are usually quite small, but some types have inherently wide value tolerances, microphonic issues, and are usually physically brittle.



An assortment of capacitor types. From left: multilayer ceramic, ceramic disc, multilayer polyester film, tubular ceramic, polystyrene, metalized polyester film, aluminum electrolytic. Major scale divisions are in centimetres.

Glass and mica capacitors are extremely reliable, stable and tolerant to high temperatures and voltages, but are too expensive for most mainstream applications.

Electrolytic capacitors and supercapacitors are used to store small and larger amounts of energy, respectively, ceramic capacitors are often used in resonators, and parasitic capacitance occurs in circuits wherever the simple conductor-insulator-conductor structure is formed unintentionally by the configuration of the circuit layout.

Electrolytic capacitors use an aluminum or tantalum plate with an oxide dielectric layer. The second electrode is a liquid electrolyte, connected to the circuit by another foil plate. Electrolytic capacitors offer very high capacitance but suffer from poor tolerances, high instability, gradual loss of capacitance especially when subjected to heat, and high leakage current. Poor quality capacitors may leak electrolyte, which is harmful to printed circuit boards. The conductivity of the electrolyte drops at low temperatures, which increases equivalent series resistance. While widely used for power-supply conditioning, poor high-frequency characteristics make them unsuitable for many applications. Electrolytic capacitors suffer from self-degradation if unused for a period (around a year), and when full power is applied may short circuit, permanently damaging the capacitor and usually blowing a fuse or causing failure of rectifier diodes. For example, in older equipment, this may cause arcing in rectifier tubes. They can be restored before use by gradually applying the operating voltage, often performed on antique vacuum tube equipment over a period of thirty minutes by using a variable transformer to supply AC power. The use of this technique may be less satisfactory for some solid state equipment, which may be damaged by operation below its normal power range, requiring that the power supply first be isolated from the consuming circuits. Such remedies may not be applicable to modern high-frequency power supplies as these produce full output voltage even with reduced input.

Tantalum capacitors offer better frequency and temperature characteristics than aluminum, but higher dielectric absorption and leakage.^[57]

Polymer capacitors (OS-CON, OC-CON, KO, AO) use solid conductive polymer (or polymerized organic semiconductor) as electrolyte and offer longer life and lower ESR at higher cost than standard electrolytic capacitors.

A feedthrough capacitor is a component that, while not serving as its main use, has capacitance and is used to conduct signals through a conductive sheet.

Several other types of capacitor are available for specialist applications. Supercapacitors store large amounts of energy. Supercapacitors made from carbon aerogel, carbon nanotubes, or highly porous electrode materials, offer extremely high capacitance (up to 5 kF as of 2010) and can be used in some applications



Three aluminum electrolytic capacitors of varying capacity



3D model of a capacitor

instead of rechargeable batteries. Alternating current capacitors are specifically designed to work on line (mains) voltage AC power circuits. They are commonly used in electric motor circuits and are often designed to handle large currents, so they tend to be physically large. They are usually ruggedly packaged, often in metal cases that can be easily grounded/earthed. They also are designed with direct current breakdown voltages of at least five times the maximum AC voltage.

Voltage-dependent capacitors

The dielectric constant for a number of very useful dielectrics changes as a function of the applied electrical field, for example ferroelectric materials, so the capacitance for these devices is more complex. For example, in charging such a capacitor the differential increase in voltage with charge is governed by:

$$dQ = C(V) dV$$

where the voltage dependence of capacitance, $C(V)$, suggests that the capacitance is a function of the electric field strength, which in a large area parallel plate device is given by $\mathcal{E} = V/d$. This field polarizes the dielectric, which polarization, in the case of a ferroelectric, is a nonlinear S-shaped function of the electric field, which, in the case of a large area parallel plate device, translates into a capacitance that is a nonlinear function of the voltage.^{[58][59]}

Corresponding to the voltage-dependent capacitance, to charge the capacitor to voltage V an integral relation is found:

$$Q = \int_0^V C(V) dV$$

which agrees with $Q = CV$ only when C does not depend on voltage V .

By the same token, the energy stored in the capacitor now is given by

$$dW = Q dV = \left[\int_0^V dV' C(V') \right] dV.$$

Integrating:

$$W = \int_0^V dV \int_0^V dV' C(V') = \int_0^V dV' \int_{V'}^V dV C(V') = \int_0^V dV' (V - V') C(V'),$$

where interchange of the order of integration is used.

The nonlinear capacitance of a microscope probe scanned along a ferroelectric surface is used to study the domain structure of ferroelectric materials.^[60]

Another example of voltage dependent capacitance occurs in semiconductor devices such as semiconductor diodes, where the voltage dependence stems not from a change in dielectric constant but in a voltage dependence of the spacing between the charges on the two sides of the capacitor.^[61] This effect is intentionally exploited in diode-like devices known as varicaps.

Frequency-dependent capacitors

If a capacitor is driven with a time-varying voltage that changes rapidly enough, at some frequency the polarization of the dielectric cannot follow the voltage. As an example of the origin of this mechanism, the internal microscopic dipoles contributing to the dielectric constant cannot move instantly, and so as

frequency of an applied alternating voltage increases, the dipole response is limited and the dielectric constant diminishes. A changing dielectric constant with frequency is referred to as dielectric dispersion, and is governed by dielectric relaxation processes, such as Debye relaxation. Under transient conditions, the displacement field can be expressed as (see electric susceptibility):

$$\mathbf{D}(\mathbf{t}) = \epsilon_0 \int_{-\infty}^t \epsilon_r(t-t') \mathbf{E}(t') dt',$$

indicating the lag in response by the time dependence of ϵ_r , calculated in principle from an underlying microscopic analysis, for example, of the dipole behavior in the dielectric. See, for example, linear response function.^{[62][63]} The integral extends over the entire past history up to the present time. A Fourier transform in time then results in:

$$\mathbf{D}(\omega) = \epsilon_0 \epsilon_r(\omega) \mathbf{E}(\omega),$$

where $\epsilon_r(\omega)$ is now a complex function, with an imaginary part related to absorption of energy from the field by the medium. See permittivity. The capacitance, being proportional to the dielectric constant, also exhibits this frequency behavior. Fourier transforming Gauss's law with this form for displacement field:

$$\begin{aligned} I(\omega) &= j\omega Q(\omega) = j\omega \oint_{\Sigma} \mathbf{D}(\mathbf{r}, \omega) \cdot d\mathbf{\Sigma} \\ &= [G(\omega) + j\omega C(\omega)] V(\omega) = \frac{V(\omega)}{Z(\omega)}, \end{aligned}$$

where j is the imaginary unit, $V(\omega)$ is the voltage component at angular frequency ω , $G(\omega)$ is the *real* part of the current, called the conductance, and $C(\omega)$ determines the *imaginary* part of the current and is the capacitance. $Z(\omega)$ is the complex impedance.

When a parallel-plate capacitor is filled with a dielectric, the measurement of dielectric properties of the medium is based upon the relation:

$$\epsilon_r(\omega) = \epsilon'_r(\omega) - j\epsilon''_r(\omega) = \frac{1}{j\omega Z(\omega) C_0} = \frac{C_{\text{cmplx}}(\omega)}{C_0},$$

where a single *prime* denotes the real part and a double *prime* the imaginary part, $Z(\omega)$ is the complex impedance with the dielectric present, $C_{\text{cmplx}}(\omega)$ is the so-called *complex* capacitance with the dielectric present, and C_0 is the capacitance without the dielectric.^{[64][65]} (Measurement "without the dielectric" in principle means measurement in free space, an unattainable goal inasmuch as even the quantum vacuum is predicted to exhibit nonideal behavior, such as dichroism. For practical purposes, when measurement errors are taken into account, often a measurement in terrestrial vacuum, or simply a calculation of C_0 , is sufficiently accurate.^[66])

Using this measurement method, the dielectric constant may exhibit a resonance at certain frequencies corresponding to characteristic response frequencies (excitation energies) of contributors to the dielectric constant. These resonances are the basis for a number of experimental techniques for detecting defects. The conductance method measures absorption as a function of frequency.^[67] Alternatively, the time response of the capacitance can be used directly, as in deep-level transient spectroscopy.^[68]

Another example of frequency dependent capacitance occurs with MOS capacitors, where the slow generation of minority carriers means that at high frequencies the capacitance measures only the majority carrier response, while at low frequencies both types of carrier respond.^{[61][69]}

At optical frequencies, in semiconductors the dielectric constant exhibits structure related to the band structure of the solid. Sophisticated modulation spectroscopy measurement methods based upon modulating the crystal structure by pressure or by other stresses and observing the related changes in absorption or reflection of light have advanced our knowledge of these materials.^[70]

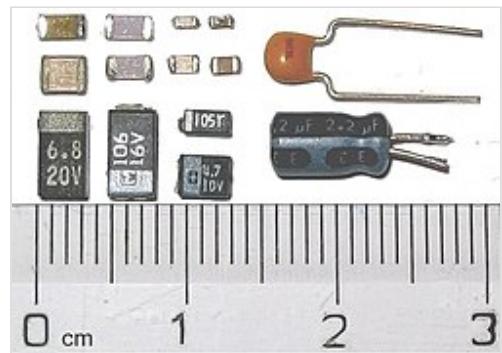
Styles

The arrangement of plates and dielectric has many variations in different styles depending on the desired ratings of the capacitor. For small values of capacitance (microfarads and less), ceramic disks use metallic coatings, with wire leads bonded to the coating. Larger values can be made by multiple stacks of plates and disks. Larger value capacitors usually use a metal foil or metal film layer deposited on the surface of a dielectric film to make the plates, and a dielectric film of impregnated paper or plastic – these are rolled up to save space. To reduce the series resistance and inductance for long plates, the plates and dielectric are staggered so that connection is made at the common edge of the rolled-up plates, not at the ends of the foil or metallized film strips that comprise the plates.

The assembly is encased to prevent moisture entering the dielectric – early radio equipment used a cardboard tube sealed with wax. Modern paper or film dielectric capacitors are dipped in a hard thermoplastic. Large capacitors for high-voltage use may have the roll form compressed to fit into a rectangular metal case, with bolted terminals and bushings for connections. The dielectric in larger capacitors is often impregnated with a liquid to improve its properties.

Capacitors may have their connecting leads arranged in many configurations, for example axially or radially. "Axial" means that the leads are on a common axis, typically the axis of the capacitor's cylindrical body – the leads extend from opposite ends. Radial leads are rarely aligned along radii of the body's circle, so the term is conventional. The leads (until bent) are usually in planes parallel to that of the flat body of the capacitor, and extend in the same direction; they are often parallel as manufactured.

Small, cheap discoidal ceramic capacitors have existed from the 1930s onward, and remain in widespread use. After the 1980s, surface mount packages for capacitors have been widely used. These packages are extremely small and lack connecting leads, allowing them to be soldered directly onto the surface of printed circuit boards. Surface mount components avoid undesirable high-frequency effects due to the leads and simplify automated assembly, although manual handling is made difficult due to their small size.



Capacitor packages: SMD ceramic at top left; SMD tantalum electrolytic at bottom left; through-hole ceramic at top right; through-hole aluminium electrolytic at bottom right. Major scale divisions are cm.



Several axial-lead electrolytic capacitors

Mechanically controlled variable capacitors allow the plate spacing to be adjusted, for example by rotating or sliding a set of movable plates into alignment with a set of stationary plates. Low cost variable capacitors squeeze together alternating layers of aluminum and plastic with a screw. Electrical control of capacitance is achievable with varactors (or varicaps), which are reverse-biased semiconductor diodes whose depletion region width varies with applied voltage. They are used in phase-locked loops, amongst other applications.

Capacitor markings

Marking codes for larger parts

Most capacitors have designations printed on their bodies to indicate their electrical characteristics. Larger capacitors, such as electrolytic types usually display the capacitance as value with explicit unit, for example, $220 \mu F$.

For typographical reasons, some manufacturers print *MF* on capacitors to indicate microfarads (μF).^[71]

Three-/four-character marking code for small capacitors

Smaller capacitors, such as ceramic types, often use a shorthand-notation consisting of three digits and an optional letter, where the digits (XYZ) denote the capacitance in picofarad (pF), calculated as $XY \times 10^Z$, and the letter indicating the tolerance. Common tolerances are $\pm 5\%$, $\pm 10\%$, and $\pm 20\%$, denoted as J, K, and M, respectively.

A capacitor may also be labeled with its working voltage, temperature, and other relevant characteristics.

Example: A capacitor labeled or designated as *473K 330V* has a capacitance of $47 \times 10^3 \text{ pF} = 47 \text{ nF}$ ($\pm 10\%$) with a maximum working voltage of 330 V. The working voltage of a capacitor is nominally the highest voltage that may be applied across it without undue risk of breaking down the dielectric layer.

Two-character marking code for small capacitors

For capacitances following the E3, E6, E12 or E24 series of preferred values, the former ANSI/EIA-198-D:1991, ANSI/EIA-198-1-E:1998 and ANSI/EIA-198-1-F:2002 as well as the amendment IEC 60062:2016/AMD1:2019 to IEC 60062 define a *special two-character marking code for capacitors* for very small parts which leave no room to print the above-mentioned three-/four-character code onto them. The code consists of an uppercase letter denoting the two significant digits of the value followed by a digit indicating the multiplier. The EIA standard also defines a number of lowercase letters to specify a number of values not found in E24.^[72]

Code	Series	Digit									
Letter ^[nb 1]	E24	9	0	1	2	3	4	5	6	7	8
A	1.0	0.10 pF	1.0 pF	10 pF	100 pF	1.0 nF	10 nF	100 nF	1.0 µF	10 µF	100 µF
B	1.1	0.11 pF	1.1 pF	11 pF	110 pF	1.1 nF	11 nF	110 nF	1.1 µF	11 µF	110 µF
C	1.2	0.12 pF	1.2 pF	12 pF	120 pF	1.2 nF	12 nF	120 nF	1.2 µF	12 µF	120 µF
D	1.3	0.13 pF	1.3 pF	13 pF	130 pF	1.3 nF	13 nF	130 nF	1.3 µF	13 µF	130 µF
E	1.5	0.15 pF	1.5 pF	15 pF	150 pF	1.5 nF	15 nF	150 nF	1.5 µF	15 µF	150 µF
F	1.6	0.16 pF	1.6 pF	16 pF	160 pF	1.6 nF	16 nF	160 nF	1.6 µF	16 µF	160 µF
G	1.8	0.18 pF	1.8 pF	18 pF	180 pF	1.8 nF	18 nF	180 nF	1.8 µF	18 µF	180 µF
H	2.0	0.20 pF	2.0 pF	20 pF	200 pF	2.0 nF	20 nF	200 nF	2.0 µF	20 µF	200 µF
J	2.2	0.22 pF	2.2 pF	22 pF	220 pF	2.2 nF	22 nF	220 nF	2.2 µF	22 µF	220 µF
K	2.4	0.24 pF	2.4 pF	24 pF	240 pF	2.4 nF	24 nF	240 nF	2.4 µF	24 µF	240 µF
L	2.7	0.27 pF	2.7 pF	27 pF	270 pF	2.7 nF	27 nF	270 nF	2.7 µF	27 µF	270 µF
M	3.0	0.30 pF	3.0 pF	30 pF	300 pF	3.0 nF	30 nF	300 nF	3.0 µF	30 µF	300 µF
N	3.3	0.33 pF	3.3 pF	33 pF	330 pF	3.3 nF	33 nF	330 nF	3.3 µF	33 µF	330 µF
P	3.6	0.36 pF	3.6 pF	36 pF	360 pF	3.6 nF	36 nF	360 nF	3.6 µF	36 µF	360 µF
Q	3.9	0.39 pF	3.9 pF	39 pF	390 pF	3.9 nF	39 nF	390 nF	3.9 µF	39 µF	390 µF
R	4.3	0.43 pF	4.3 pF	43 pF	430 pF	4.3 nF	43 nF	430 nF	4.3 µF	43 µF	430 µF
S	4.7	0.47 pF	4.7 pF	47 pF	470 pF	4.7 nF	47 nF	470 nF	4.7 µF	47 µF	470 µF
T	5.1	0.51 pF	5.1 pF	51 pF	510 pF	5.1 nF	51 nF	510 nF	5.1 µF	51 µF	510 µF
U	5.6	0.56 pF	5.6 pF	56 pF	560 pF	5.6 nF	56 nF	560 nF	5.6 µF	56 µF	560 µF
V	6.2	0.62 pF	6.2 pF	62 pF	620 pF	6.2 nF	62 nF	620 nF	6.2 µF	62 µF	620 µF
W	6.8	0.68 pF	6.8 pF	68 pF	680 pF	6.8 nF	68 nF	680 nF	6.8 µF	68 µF	680 µF
X	7.5	0.75 pF	7.5 pF	75 pF	750 pF	7.5 nF	75 nF	750 nF	7.5 µF	75 µF	750 µF
Y	8.2	0.82 pF	8.2 pF	82 pF	820 pF	8.2 nF	82 nF	820 nF	8.2 µF	82 µF	820 µF
Z	9.1	0.91 pF	9.1 pF	91 pF	910 pF	9.1 nF	91 nF	910 nF	9.1 µF	91 µF	910 µF

Code	Series	Digit									
		9	0	1	2	3	4	5	6	7	8
a	2.5	0.25 pF	2.5 pF	25 pF	250 pF	2.5 nF	25 nF	250 nF	2.5 µF	25 µF	250 µF
b? ^[73]	3.0? ^[73]	0.30 pF	3.0 pF	30 pF	300 pF	3.0 nF	30 nF	300 nF	3.0 µF	30 µF	300 µF
b? ^[72] /c? ^[73]	3.5	0.35 pF	3.5 pF	35 pF	350 pF	3.5 nF	35 nF	350 nF	3.5 µF	35 µF	350 µF
d	4.0	0.40 pF	4.0 pF	40 pF	400 pF	4.0 nF	40 nF	400 nF	4.0 µF	40 µF	400 µF
e	4.5	0.45 pF	4.5 pF	45 pF	450 pF	4.5 nF	45 nF	450 nF	4.5 µF	45 µF	450 µF
f	5.0	0.50 pF	5.0 pF	50 pF	500 pF	5.0 nF	50 nF	500 nF	5.0 µF	50 µF	500 µF
m	6.0	0.60 pF	6.0 pF	60 pF	600 pF	6.0 nF	60 nF	600 nF	6.0 µF	60 µF	600 µF
n	7.0	0.70 pF	7.0 pF	70 pF	700 pF	7.0 nF	70 nF	700 nF	7.0 µF	70 µF	700 µF
t	8.0	0.80 pF	8.0 pF	80 pF	800 pF	8.0 nF	80 nF	800 nF	8.0 µF	80 µF	800 µF
g	9.0	0.90 pF	9.0 pF	90 pF	900 pF	9.0 nF	90 nF	900 nF	9.0 µF	90 µF	900 µF

RKM code

The [RKM code](#) following [IEC 60062](#) and [BS 1852](#) is a notation to state a capacitor's value in a circuit diagram. It avoids using a [decimal separator](#) and replaces the decimal separator with the SI prefix symbol for the particular value (and the letter F for weight 1). The code is also used for part markings. Example: 4n7 for 4.7 nF or 2F2 for 2.2 F.

Historical

In texts prior to the 1960s and on some capacitor packages until more recently,^[17] obsolete capacitance units were utilized in electronic books,^[74] magazines, and electronics catalogs.^[75] The old units "mfd" and "mf" meant *microfarad* (μF); and the old units "mmfd", "mmf", "uuf", " $\mu\mu\text{f}$ ", "pfd" meant *pico farad* (pF); but they are rarely used any more.^[76] Also, "Micromicrofarad" or "micro-microfarad" are obsolete units that are found in some older texts that is equivalent to *pico farad* (pF).^[74]

Summary of obsolete capacitance units: (upper/lower case variations are not shown)

- μF (*microfarad*) = mf, mfd
- pF (*pico farad*) = mmf, mmfd, pfd, $\mu\mu\text{f}$

Applications

Energy storage

A capacitor can store electric energy when disconnected from its charging circuit, so it can be used like a temporary [battery](#), or like other types of [rechargeable energy storage system](#).^[77] Capacitors are commonly used in electronic devices to maintain power supply while batteries are being changed. (This prevents loss of information in volatile memory.)

A capacitor can facilitate conversion of kinetic energy of charged particles into electric energy and store it.^[78]

There are tradeoffs between capacitors and batteries as storage devices. Without external resistors or inductors, capacitors can generally release their stored energy in a very short time compared to batteries. Conversely, batteries can hold a far greater charge per their size. Conventional capacitors provide less than 360 joules per kilogram of specific energy, whereas a conventional alkaline battery has a density of 590 kJ/kg. There is an intermediate solution: supercapacitors, which can accept and deliver charge much faster than batteries, and tolerate many more charge and discharge cycles than rechargeable batteries. They are, however, 10 times larger than conventional batteries for a given charge. On the other hand, it has been shown that the amount of charge stored in the dielectric layer of the thin film capacitor can be equal to, or can even exceed, the amount of charge stored on its plates.^[79]



A capacitor discharging its stored energy through a flashtube. The mylar-film capacitor has very low inductance and low resistance, producing a 3.5 microsecond pulse with 24 million watts of power, to operate a dye laser.

In car audio systems, large capacitors store energy for the amplifier to use on demand. Also, for a flash tube, a capacitor is used to hold the high voltage.

Digital memory

In the 1930s, John Atanasoff applied the principle of energy storage in capacitors to construct dynamic digital memories for the first binary computers that used electron tubes for logic.^[80]

Pulsed power and weapons

Pulsed power is used in many applications to increase the power intensity (watts) of a volume of energy (joules) by releasing that volume within a very short time. Pulses in the nanosecond range and powers in the gigawatts are achievable. Short pulses often require specially constructed, low-inductance, high-voltage capacitors that are often used in large groups (*capacitor banks*) to supply huge pulses of current for many pulsed power applications. These include electromagnetic forming, Marx generators, pulsed lasers (especially TEA lasers), pulse forming networks, radar, fusion research, and particle accelerators.^[81]

Large capacitor banks (reservoir) are used as energy sources for the exploding-bridgewire detonators or slapper detonators in nuclear weapons and other specialty weapons. Experimental work is under way using banks of capacitors as power sources for electromagnetic armour and electromagnetic railguns and coilguns.

Power conditioning

Reservoir capacitors are used in power supplies where they smooth the output of a full or half wave rectifier. They can also be used in charge pump circuits as the energy storage element in the generation of higher voltages than the input voltage.

Capacitors are connected in parallel with the power circuits of most electronic devices and larger systems (such as factories) to shunt away and conceal current fluctuations from the primary power source to provide a "clean" power supply for signal or control circuits. Audio equipment, for example, uses several capacitors in this way, to shunt away power line hum before it gets into the signal circuitry. The capacitors act as a

local reserve for the DC power source, and bypass AC currents from the power supply. This is used in car audio applications, when a stiffening capacitor compensates for the inductance and resistance of the leads to the lead-acid car battery.

Power-factor correction

In electric power distribution, capacitors are used for power-factor correction. Such capacitors often come as three capacitors connected as a three phase load. Usually, the values of these capacitors are not given in farads but rather as a reactive power in volt-amperes reactive (var). The purpose is to counteract inductive loading from devices like electric motors and transmission lines to make the load appear to be mostly resistive. Individual motor or lamp loads may have capacitors for power-factor correction, or larger sets of capacitors (usually with automatic switching devices) may be installed at a load center within a building or in a large utility substation.

Suppression and coupling

Signal coupling

Because capacitors pass AC but block DC signals (when charged up to the applied DC voltage), they are often used to separate the AC and DC components of a signal. This method is known as *AC coupling* or "capacitive coupling". Here, a large value of capacitance, whose value need not be accurately controlled, but whose reactance is small at the signal frequency, is employed.

Decoupling

A decoupling capacitor is a capacitor used to protect one part of a circuit from the effect of another, for instance to suppress noise or transients. Noise caused by other circuit elements is shunted through the capacitor, reducing the effect they have on the rest of the circuit. It is most commonly used between the power supply and ground. An alternative name is bypass capacitor as it is used to bypass the power supply or other high impedance component of a circuit.

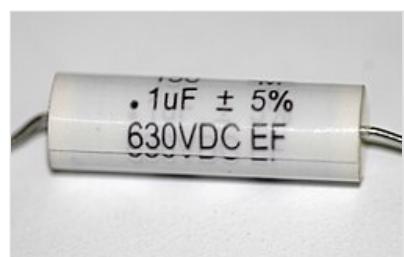
Decoupling capacitors need not always be discrete components. Capacitors used in these applications may be built into a printed circuit board, between the various layers. These are often referred to as embedded capacitors.^[82] The layers in the board contributing to the capacitive properties also function as power and ground planes, and have a dielectric in between them, enabling them to operate as a parallel plate capacitor.



A 10,000 microfarad capacitor in an amplifier power supply



A high-voltage capacitor bank used for power-factor correction on a power transmission system



Polyester film capacitors are frequently used as coupling capacitors.

High-pass and low-pass filters

Noise suppression, spikes, and snubbers

When an inductive circuit is opened, the current through the inductance collapses quickly, creating a large voltage across the open circuit of the switch or relay. If the inductance is large enough, the energy may generate a spark, causing the contact points to oxidize, deteriorate, or sometimes weld together, or destroying a solid-state switch. A snubber capacitor across the newly opened circuit creates a path for this impulse to bypass the contact points, thereby preserving their life; these were commonly found in contact breaker ignition systems, for instance. Similarly, in smaller scale circuits, the spark may not be enough to damage the switch but may still radiate undesirable radio frequency interference (RFI), which a filter capacitor absorbs. Snubber capacitors are usually employed with a low-value resistor in series, to dissipate energy and minimize RFI. Such resistor-capacitor combinations are available in a single package.

Capacitors are also used in parallel with interrupting units of a high-voltage circuit breaker to equally distribute the voltage between these units. These are called "grading capacitors".

In schematic diagrams, a capacitor used primarily for DC charge storage is often drawn vertically in circuit diagrams with the lower, more negative, plate drawn as an arc. The straight plate indicates the positive terminal of the device, if it is polarized (see electrolytic capacitor).

Motor starters

In single phase squirrel cage motors, the primary winding within the motor housing is not capable of starting a rotational motion on the rotor, but is capable of sustaining one. To start the motor, a secondary "start" winding has a series non-polarized starting capacitor to introduce a lead in the sinusoidal current. When the secondary (start) winding is placed at an angle with respect to the primary (run) winding, a rotating electric field is created. The force of the rotational field is not constant, but is sufficient to start the rotor spinning. When the rotor comes close to operating speed, a centrifugal switch (or current-sensitive relay in series with the main winding) disconnects the capacitor. The start capacitor is typically mounted to the side of the motor housing. These are called capacitor-start motors, that have relatively high starting torque. Typically they can have up-to four times as much starting torque as a split-phase motor and are used on applications such as compressors, pressure washers and any small device requiring high starting torques.

Capacitor-run induction motors have a permanently connected phase-shifting capacitor in series with a second winding. The motor is much like a two-phase induction motor.

Motor-starting capacitors are typically non-polarized electrolytic types, while running capacitors are conventional paper or plastic film dielectric types.

Signal processing

The energy stored in a capacitor can be used to represent information, either in binary form, as in DRAMs, or in analogue form, as in analog sampled filters and CCDs. Capacitors can be used in analog circuits as components of integrators or more complex filters and in negative feedback loop stabilization. Signal processing circuits also use capacitors to integrate a current signal.

Tuned circuits

Capacitors and inductors are applied together in tuned circuits to select information in particular frequency bands. For example, radio receivers rely on variable capacitors to tune the station frequency. Speakers use passive analog crossovers, and analog equalizers use capacitors to select different audio bands.

The resonant frequency f of a tuned circuit is a function of the inductance (L) and capacitance (C) in series, and is given by:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

where L is in henries and C is in farads.

Sensing

Most capacitors are designed to maintain a fixed physical structure. However, various factors can change the structure of the capacitor, and the resulting change in capacitance can be used to sense those factors.

Changing the dielectric

The effects of varying the characteristics of the **dielectric** can be used for sensing purposes. Capacitors with an exposed and porous dielectric can be used to measure humidity in air. Capacitors are used to accurately measure the fuel level in airplanes; as the fuel covers more of a pair of plates, the circuit capacitance increases. Squeezing the dielectric can change a capacitor at a few tens of bar pressure sufficiently that it can be used as a pressure sensor.^[83] A selected, but otherwise standard, polymer dielectric capacitor, when immersed in a compatible gas or liquid, can work usefully as a very low cost pressure sensor up to many hundreds of bar.

Changing the distance between the plates

Capacitors with a flexible plate can be used to measure strain or pressure. Industrial pressure transmitters used for process control use pressure-sensing diaphragms, which form a capacitor plate of an oscillator circuit. Capacitors are used as the sensor in condenser microphones, where one plate is moved by air pressure, relative to the fixed position of the other plate. Some accelerometers use MEMS capacitors etched on a chip to measure the magnitude and direction of the acceleration vector. They are used to detect changes in acceleration, in tilt sensors, or to detect free fall, as sensors triggering airbag deployment, and in many other applications. Some fingerprint sensors use capacitors. Additionally, a user can adjust the pitch of a theremin musical instrument by moving their hand since this changes the effective capacitance between the user's hand and the antenna.

Changing the effective area of the plates

Capacitive touch switches are now used on many consumer electronic products.

Oscillators

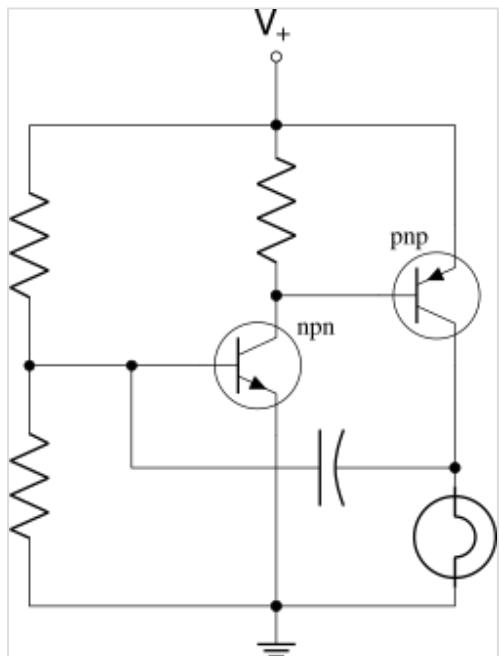
A capacitor can possess spring-like qualities in an oscillator circuit. In the image example, a capacitor acts to influence the biasing voltage at the npn transistor's base. The resistance values of the voltage-divider resistors and the capacitance value of the capacitor together control the oscillatory frequency.

Producing light

A light-emitting capacitor is made from a dielectric that uses phosphorescence to produce light. If one of the conductive plates is made with a transparent material, the light is visible. Light-emitting capacitors are used in the construction of electroluminescent panels, for applications such as backlighting for laptop computers. In this case, the entire panel is a capacitor used for the purpose of generating light.

Hazards and safety

The hazards posed by a capacitor are usually determined, foremost, by the amount of energy stored, which is the cause of things like electrical burns or heart fibrillation. Factors such as voltage and chassis material are of secondary consideration, which are more related to how easily a shock can be initiated rather than how much damage can occur.^[54] Under certain conditions, including conductivity of the surfaces, preexisting medical conditions, the humidity of the air, or the pathways it takes through the body (i.e.: shocks that travel across the core of the body and, especially, the heart are more dangerous than those limited to the extremities), shocks as low as one joule have been reported to cause death, although in most instances they may not even leave a burn. Shocks over ten joules will generally damage skin, and are usually considered hazardous. Any capacitor that can store 50 joules or more should be considered potentially lethal.^{[84][54]}



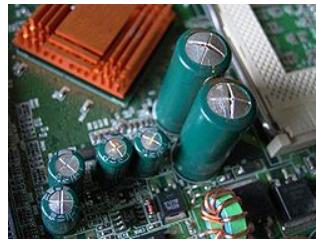
Example of a simple oscillator incorporating a capacitor

Capacitors may retain a charge long after power is removed from a circuit; this charge can cause dangerous or even potentially fatal shocks or damage connected equipment. For example, even a seemingly innocuous device such as the flash of a disposable camera, has a photoflash capacitor which may contain over 15 joules of energy and be charged to over 300 volts. This is easily capable of delivering a shock. Service procedures for electronic devices usually include instructions to discharge large or high-voltage capacitors, for instance using a Brinkley stick. Larger capacitors, such as those used in microwave ovens, HVAC units and medical defibrillators may also have built-in discharge resistors to dissipate stored energy to a safe level within a few seconds after power is removed. High-voltage capacitors are stored with the terminals shorted, as protection from potentially dangerous voltages due to dielectric absorption or from transient voltages the capacitor may pick up from static charges or passing weather events.^[54]

Some old, large oil-filled paper or plastic film capacitors contain polychlorinated biphenyls (PCBs). It is known that waste PCBs can leak into groundwater under landfills. Capacitors containing PCBs were labelled as containing "Askarel" and several other trade names. PCB-filled paper capacitors are found in very old (pre-1975) fluorescent lamp ballasts, and other applications.

Capacitors may catastrophically fail when subjected to voltages or currents beyond their rating, or in case of polarized capacitors, applied in a reverse polarity. Failures may create arcing that heats and vaporizes the dielectric fluid, causing a build up of pressurized gas that may result in swelling, rupture, or an explosion. Larger capacitors may have vents or similar mechanism to allow the release of such pressures in the event of failure. Capacitors used in RF or sustained high-current applications can overheat, especially in the center of the capacitor rolls. Capacitors used within high-energy capacitor banks can violently explode when a short in one capacitor causes sudden dumping of energy stored in the rest of the bank into the failing unit. High voltage vacuum capacitors can generate soft X-rays even during normal operation. Proper containment, fusing, and preventive maintenance can help to minimize these hazards.

High-voltage capacitors may benefit from a pre-charge to limit in-rush currents at power-up of high voltage direct current (HVDC) circuits. This extends the life of the component and may mitigate high-voltage hazards.



Swollen electrolytic capacitors. The vent on the tops allows the release of pressurized gas build-up in the event of failure, preventing it from exploding.



This high-energy capacitor from a defibrillator has a resistor connected between the terminals for safety, to dissipate stored energy.



An exploded electrolytic capacitor, showing fragments of paper and metallic foil

See also



Electronics portal

- [Capacitance meter](#)
- [Capacitor plague](#)
- [Electric displacement field](#)
- [Electroluminescence](#)
- [List of capacitor manufacturers](#)

Notes

1. Most real capacitors may have a small dielectric leakage current that passes through the resistive dielectric layer in between the plates.
1. In order to reduce the risk for read errors, the letters I and O are not used as their glyphs look similar to other letters and digits.

References

1. Duff, Wilmer (1916) [1908]. *A Text-Book of Physics* (<https://archive.org/stream/atextbookphysic00carmgoog#page/n378/mode/2up>) (4th ed.). Philadelphia: P. Blakiston's Son & Co. p. 361. Retrieved 2016-12-01.
2. Bird, John (2010). *Electrical and Electronic Principles and Technology* (https://books.google.com/books?id=A1tAHm_5sl0C). Routledge. pp. 63–76. ISBN 978-0-08089056-2. Retrieved 2013-03-17.
3. Floyd, Thomas (2005) [1984]. *Electronic Devices* (7th ed.). Upper Saddle River, New Jersey, USA: Pearson Education. p. 10. ISBN 0-13-127827-4.

4. "Molecular Expressions: Electricity and Magnetism - Interactive Java Tutorials: Lightning: A Natural Capacitor" (<https://micro.magnet.fsu.edu/electromag/java/lightning/index.html#:~:text=An+Example+Of+A+Natural,ionizing+radiation,+and+each+other.>). *micro.magnet.fsu.edu*.
5. Williams, Henry Smith. "A History of Science Volume II, Part VI: The Leyden Jar Discovered" (<https://web.archive.org/web/20071024034729/http://www.worldwideschool.org/library/books/sci/history/AHistoryofScienceVolumell/chap49.html>). Archived from the original (<http://www.worldwideschool.org/library/books/sci/history/AHistoryofScienceVolumell/chap49.html>) on 2007-10-24. Retrieved 2013-03-17.
6. Keithley, Joseph F. (1999). *The Story of Electrical and Magnetic Measurements: From 500 BC to the 1940s* (<https://books.google.com/books?id=uwgNAtqSHuQC&pg=PA23>). John Wiley & Sons. p. 23. ISBN 978-0780311930. Retrieved 2013-03-17.
7. Houston, Edwin J. (1905). *Electricity in Every-day Life* (<https://books.google.com/books?id=ko9BAAAAIAAJ&pg=PA71>). P. F. Collier & Son. p. 71. Retrieved 2013-03-17.
8. Benjamin, Park (1895). *A History of Electricity: (The Intellectual Rise in Electricity) from Antiquity to the Days of Benjamin Franklin* (<https://books.google.com/books?id=hkMPAAAAMA AJ&pg=PA523>). J. Wiley & Sons. pp. 522–524.
9. Isaacson, Walter (2003). *Benjamin Franklin: An American Life* (<https://books.google.com/books?id=oIW915dDMBwC&pg=PA135>). Simon and Schuster. p. 136. ISBN 978-0-74326084-8. Retrieved 2013-03-17.
10. Franklin, Benjamin (1749-04-29). "Experiments & Observations on Electricity: Letter IV to Peter Collinson" (<http://www.chemteam.info/Chem-History/Franklin-1749/Franklin-1749-all.pdf>) (PDF). p. 28. Retrieved 2009-08-09.
11. Morse, Robert A. (September 2004). "Franklin and Electrostatics – Ben Franklin as my Lab Partner" (<http://www.compadre.org/Repository/document/ServeFile.cfm?ID=3430&DocID=2402&DocFID=3925&Attachment=1>) (PDF). *Wright Center for Science Education*. Tufts University. p. 23. Retrieved 2009-08-10. "After Volta's discovery of the electrochemical cell in 1800, the term was then applied to a group of electrochemical cells"
12. Wolf, A; McKie, D. (1962). *A history of science, technology and philosophy in the 18th century* (<https://archive.org/stream/in.ernet.dli.2015.460063/2015.460063.A-History#page/n261/mode/2up>) (2nd ed.). London: George Allen & Unwin. p. 224.
13. "eFunda: Glossary: Units: Electric Capacitance: Jar" (http://www.efunda.com/glossary/units/units--electric_capacitance--jar.cfm). eFunda. Retrieved 2013-03-17.
14. Pancaldi, G. (2003). *Volta: Science and culture in the Age of Enlightenment*. Princeton: Princeton University Press. pp. 112–126. ISBN 0691096856.
15. "Sketch of Alessandro Volta" (<https://books.google.com/books?id=eCADAAAAMBAJ&pg=PA117>). *The Popular Science Monthly*. New York: Bonnier Corporation: 118–119. May 1892. ISSN 0161-7370 (<https://search.worldcat.org/issn/0161-7370>).
16. British Engineering Standards Association, *British Standard Glossary of Terms in Electrical Engineering*, C. Lockwood & Son, 1926
17. Ho, Janet; Jow, T. Richard; Boggs, Steven (January 2010). "Historical Introduction to Capacitor Technology" (<https://zenodo.org/record/1232215>). *IEEE Electrical Insulation Magazine*. **26** (1): 20–25. doi:10.1109/mei.2010.5383924 (<https://doi.org/10.1109%2Fmei.2010.5383924>). S2CID 23077215 (<https://api.semanticscholar.org/CorpusID:23077215>).
18. US 2800616 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=US2800616>), Becker, H. I., "Low voltage electrolytic capacitor", issued 1957-07-23
19. A brief history of supercapacitors AUTUMN 2007 Batteries & Energy Storage Technology (<http://www.cantecsystems.com/ccrdocs/brief-history-of-supercapacitors.pdf>) Archived (<https://web.archive.org/web/20140106032318/http://www.cantecsystems.com/ccrdocs/brief-history-of-supercapacitors.pdf>) 2014-01-06 at the Wayback Machine
20. Sze, Simon Min; Lee, Ming-Kwei (May 2012). "MOS Capacitor and MOSFET". *Semiconductor Devices: Physics and Technology* (https://www.oreilly.com/library/view/semiconductor-devices-physics/9780470537947/13_chap05.html). John Wiley & Sons. ISBN 978-0-47053794-7. Retrieved 2019-10-06.

21. "DRAM" (<https://www.ibm.com/ibm/history/ibm100/us/en/icons/dram/>). *IBM100*. IBM. 2017-08-09. Retrieved 2019-09-20.
22. Sze, Simon M. (2002). *Semiconductor Devices: Physics and Technology* (https://web.archive.org/web/20230123184804/http://www.fulviofrisone.com/attachments/article/453/Semiconductor.Devices_Physics.Technology_Sze.2ndEd_Wiley_2002.pdf) (PDF) (2nd ed.). Wiley. p. 214. ISBN 0-471-33372-7. Archived from the original (http://www.fulviofrisone.com/attachments/article/453/Semiconductor.Devices_Physics.Technology_Sze.2ndEd_Wiley_2002.pdf) (PDF) on 2023-01-23.
23. Ulaby 1999, p. 168.
24. Ulaby 1999, p. 157.
25. Ulaby 1999, p. 69.
26. Pillai, K. P. P. (1970). "Fringing field of finite parallel-plate capacitors". *Proceedings of the Institution of Electrical Engineers*. **117** (6): 1201–1204. doi:10.1049/piee.1970.0232 (<https://doi.org/10.1049%2Fpiee.1970.0232>).
27. Purcell, Edward (2011). *Electricity and Magnetism, 2nd Ed* (<https://books.google.com/books?id=Z3bkNh6h4WEC&pg=PA110>). Cambridge University Press. pp. 110–111. ISBN 978-1-13950355-6.
28. Serway, Raymond A.; Vuille, Chris (2014). *College Physics, 10th Ed* (<https://books.google.com/books?id=xETAAgAAQBAJ&pg=PA582>). Cengage Learning. p. 582. ISBN 978-1-30514282-4.
29. Hammond, P. (2013). *Electromagnetism for Engineers: An Introductory Course* (<https://books.google.com/books?id=-8QgBQAAQBAJ&pg=PA44>). Elsevier Science. pp. 44–45. ISBN 978-1-48314978-3.
30. Dorf & Svoboda 2001, p. 263.
31. Dorf & Svoboda 2001, p. 260.
32. "Capacitor charging and discharging" (http://www.allaboutcircuits.com/vol_6/chpt_3/17.html). *All About Circuits*. Retrieved 2009-02-19.
33. Current percolation through resistors and capacitors (https://www.researchgate.net/publication/313799141_Universality_of_the_emergent_scaling_in_finite_random_binary_percolation_networks). PLoS one 2017
34. "Universality of the emergent scaling in finite random binary percolation networks" (<https://www.researchgate.net/publication/313799141>).
35. Ulaby 1999, p. 170.
36. Pai, S. T.; Zhang, Qi (1995). *Introduction to High Power Pulse Technology* (https://books.google.com/books?id=spZ_H4nwlN0C&pg=PA47). Advanced Series in Electrical and Computer Engineering. Vol. 10. World Scientific. ISBN 978-9810217143. Retrieved 2013-03-17.
37. Dyer, Stephen A. (2004). *Wiley Survey of Instrumentation and Measurement* (<https://books.google.com/books?id=Wr6l42rEizUC&pg=PA397>). John Wiley & Sons. p. 397. ISBN 978-0-47122165-4. Retrieved 2013-03-17.
38. Scherz, Paul (2006). *Practical Electronics for Inventors* (<https://books.google.com/books?id=C9pL3iL6eSMC&pg=PA100>) (2nd ed.). McGraw Hill Professional. p. 100. ISBN 978-0-07177644-8. Retrieved 2013-03-17.
39. Inuishi, Y.; Powers, D. A. (1957). "Electric breakdown and conduction through Mylar films". *Journal of Applied Physics*. **28** (9): 1017–1022. Bibcode:1957JAP....28.1017I (<https://ui.adsabs.harvard.edu/abs/1957JAP....28.1017I>). doi:10.1063/1.1722899 (<https://doi.org/10.1063%2F1.1722899>).
40. Reed, C. W.; Cichanowski, S. W. (1994). "The fundamentals of aging in HV polymer-film capacitors". *IEEE Transactions on Dielectrics and Electrical Insulation*. **1** (5): 904–922. doi:10.1109/94.326658 (<https://doi.org/10.1109%2F94.326658>).

41. Klein, N.; Gafni, H. (1966). "The maximum dielectric strength of thin silicon oxide films". *IEEE Transactions on Electron Devices*. **13** (2): 281–289. Bibcode:1966ITED...13..281K (<https://ui.adsabs.harvard.edu/abs/1966ITED...13..281K>). doi:10.1109/T-ED.1966.15681 (<https://doi.org/10.1109%2FT-ED.1966.15681>).
42. Belkin, A.; et al. (2017). "Recovery of alumina nanocapacitors after high voltage breakdown" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5430567>). *Scientific Reports*. **7** (1): 932. Bibcode:2017NatSR...7..932B (<https://ui.adsabs.harvard.edu/abs/2017NatSR...7..932B>). doi:10.1038/s41598-017-01007-9 (<https://doi.org/10.1038%2Fs41598-017-01007-9>). PMC 5430567 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5430567>). PMID 28428625 (<https://pubmed.ncbi.nlm.nih.gov/28428625>).
43. Bird, John (2007). *Electrical Circuit Theory and Technology* (<https://books.google.com/books?id=Q9zpWdgQeM4C&pg=PA501>). Routledge. p. 501. ISBN 978-0-75068139-1. Retrieved 2013-03-17.
44. Bisquert, J.; Garcia-Belmonte, G.; Fabregat-Santiago, F. (2000). "The role of instrumentation in the process of modeling real capacitors" (<https://ieeexplore.ieee.org/document/883355>). *IEEE Transactions on Education*. **43** (4): 439–442. Bibcode:2000ITEdu..43..439F (<https://ui.adsabs.harvard.edu/abs/2000ITEdu..43..439F>). doi:10.1109/13.883355 (<https://doi.org/10.1109%2F13.883355>). ISSN 1557-9638 (<https://search.worldcat.org/issn/1557-9638>).
45. Ulaby 1999, p. 169.
46. "Ceramic Capacitor Aging Made Simple" (<https://web.archive.org/web/20121226160520/http://www.johansondielectrics.com/technical-notes/general/ceramic-capacitor-aging-made-simple.html>). Johanson Dielectrics. 2012-05-21. Archived from the original (<http://www.johansondielectrics.com/technical-notes/general/ceramic-capacitor-aging-made-simple.html>) on 2012-12-26. Retrieved 2013-03-17.
47. "Concise Guide to Capacitor Types" (<https://www.allaboutcircuits.com/technical-articles/x7r-x5-c0g---a-concise-guide-to-ceramic-capacitor-types/>). EETech Media LLC. Retrieved 2023-09-07.
48. "The Effect of Reversal on Capacitor Life" (<https://web.archive.org/web/20140714134127/http://www.ga.com/websites/ga/images/products/ep/tech-bulletins/voltage-reversal.pdf>) (PDF). *Engineering Bulletin 96-004*. Sorrento Electronics. November 2003. Archived from the original (<http://www.ga.com/websites/ga/images/products/ep/tech-bulletins/voltage-reversal.pdf>) (PDF) on 2014-07-14. Retrieved 2013-03-17.
49. Kaiser, Cletus J. (2012-12-06). *The Capacitor Handbook* (<https://books.google.com/books?id=FOatBAAAQBAJ>). Springer Science & Business Media. ISBN 978-9-40118090-0.
50. *Electronics*. McGraw-Hill 1960 p. 90
51. Xenon Strobe and Flash Safety Hints (<http://donklipstein.com/xesafe.html>). donklipstein.com. 2006-05-29
52. Prutchi, David (2012). *Exploring Quantum Physics through Hands-on Projects* (<https://books.google.com/books?id=LLxsnx74KEYC&pg=PA10>). John Wiley & Sons. p. 10. ISBN 978-1-11817070-0.
53. Dixit, J. B.; Yadav, Amit (2010). *Electrical Power Quality* (<https://books.google.com/books?id=LWYe5sGMsC&pg=PA63>). Laxmi Publications, Ltd. p. 63. ISBN 978-9-38038674-4.
54. Winburn (1989). *Practical Laser Safety, Second Edition* (<https://books.google.com/books?id=x1I5eQybE8C&pg=PA189>). CRC Press. p. 189. ISBN 978-0-82478240-5.
55. *Robinson's Manual of Radio Telegraphy and Telephony by S.S. Robinson -- US Naval Institute 1924 Pg. 170*
56. Gupta, Anunay; Yadav, Om Prakash; DeVoto, Douglas; Major, Joshua (October 2018). "A Review of Degradation Behavior and Modeling of Capacitors" (<https://www.nrel.gov/docs/fy19osti/71386.pdf>) (PDF). National Renewable Energy Laboratory. Archived (<https://web.archive.org/web/20200605220416/https://www.nrel.gov/docs/fy19osti/71386.pdf>) (PDF) from the original on 2020-06-05. Retrieved 2021-07-23.
57. Quinta, Steve. "Ask The Applications Engineer – 21" (<http://www.analog.com/library/analogDialogue/Anniversary/21.html>). Analog Devices. Retrieved 2013-03-17.

58. de Araujo, Carlos Paz; Ramesh, Ramamoorthy; Taylor, George W., eds. (2001). *Science and Technology of Integrated Ferroelectrics: Selected Papers from Eleven Years of the Proceedings of the International Symposium on Integrated Ferroelectrics* (<https://books.google.com/books?id=QMI0KeJ4qN4C&pg=PA504>). CRC Press. Figure 2, p. 504. ISBN 90-5699-704-1.
59. Musikant, Solomon (1991). *What Every Engineer Should Know about Ceramics* (<https://books.google.com/books?id=Jc8xRdgdh38C&pg=PA44>). CRC Press. Figure 3.9, p. 43. ISBN 0-8247-8498-7.
60. Yasuo Cho (2005). *Scanning Nonlinear Dielectric Microscope* (<https://books.google.com/books?id=wQ09DhMBJrcC&pg=PA304>) (in *Polar Oxides*; R. Waser, U. Böttger & S. Tiedke, editors ed.). Wiley-VCH. Chapter 16. ISBN 3-527-40532-1.
61. Sze & Ng 2006, p. 217.
62. Giuliani, Gabriele; Vignale, Giovanni (2005). *Quantum Theory of the Electron Liquid* (<https://books.google.com/books?id=kFkIKRfgUpsC&pg=PA538>). Cambridge University Press. p. 111. ISBN 0-521-82112-6.
63. Rammer, Jørgen (2007). *Quantum Field Theory of Non-equilibrium States* (<https://books.google.com/books?id=A7TbrAm5Wq0C&pg=PA6>). Cambridge University Press. p. 158. ISBN 978-0-52187499-1.
64. Czichos, Horst; Saito, Tetsuya; Smith, Leslie (2006). *Springer Handbook of Materials Measurement Methods* (<https://books.google.com/books?id=8IANaR-Pqi4C&pg=PA1>). Springer. p. 475. ISBN 3-540-20785-6.
65. Coffey, William; Kalmykov, Yu. P. (2006). *Fractals, diffusion and relaxation in disordered complex systems. Part A* (<https://books.google.com/books?id=mgtQslaXBc4C&pg=PA18>). Wiley. p. 17. ISBN 0-470-04607-4.
66. 2005 IEEE Instrumentation and Measurement Technology Conference (IMTC): Ottawa, Ontario Canada, 16–19 May 2005 (<https://books.google.com/books?id=85I5nQAACAAJ&pg=PA1350>). IEEE. 2005. pp. 1350–1353. doi:10.1109/IMTC.2005.1604368 (<https://doi.org/10.1109%2FIMTC.2005.1604368>). ISBN 978-0-78038879-6. S2CID 37739028 (<https://api.semanticscience.org/CorpusID:37739028>).
67. Schroder 2006, p. 347.
68. Schroder 2006, p. 305.
69. Kasap, Safa O.; Capper, Peter (2006). *Springer Handbook of Electronic and Photonic Materials* (<https://books.google.com/books?id=rVvW22pnzhoc&pg=PA425>). Springer. Figure 20.22, p. 425.
70. PY Yu; Cardona, Manuel (2001). *Fundamentals of Semiconductors* (<https://books.google.com/books?id=W9pdJZoAeyEC&pg=PA244>) (3rd ed.). Springer. §6.6 "Modulation Spectroscopy". ISBN 3-540-25470-6.
71. Kaplan, Daniel M.; White, Christopher G. (2003). *Hands-On Electronics: A Practical Introduction to Analog and Digital Circuits* (<https://books.google.com/books?id=kNbGGpuJyjmC&pg=PA19>). Cambridge University Press. p. 19. ISBN 978-0-52189351-0.
72. "Annex B: Special two-character code system for capacitors". *SLOVENSKI STANDARD SIST EN 60062:2016/A1:2019* (<https://cdn.standards.iteh.ai/samples/67713/7c42aed17a454cdaaff49b69445f0ee8/SIST-EN-60062-2016-A1-2019.pdf>) (PDF) (preview). 2019-12-01. pp. 3–4. Archived (<https://web.archive.org/web/20220617174351/https://cdn.standards.iteh.ai/samples/67713/7c42aed17a454cdaaff49b69445f0ee8/SIST-EN-60062-2016-A1-2019.pdf>) (PDF) from the original on 2022-06-17. Retrieved 2022-06-17.
73. Zabkar, Franc (2011-09-15). "EIA-198-D capacitance codes" (<https://www.electronicspoint.com/forums/threads/eia-198-d-capacitance-codes.238921/>). Archived (<https://web.archive.org/web/20220617225054/https://www.electronicspoint.com/forums/threads/eia-198-d-capacitance-codes.238921/>) from the original on 2022-06-17. Retrieved 2022-06-18.
74. "Fundamentals of Electronics, Volume 1b: Basic Electricity, Alternating Current, NAVPERS 93400A-1b" (<http://archive.org/details/FundamentalsOfElectronics93400A1b>). 1965-04-12 – via Internet Archive.

75. "1930 Catalog – Capacitors (Condensers)" (https://web.archive.org/web/20170711160331/http://www.alliedcatalogs.com/html/catalogs_additional/1930/hr139.html). *Allied Radio*. p. 139. Archived from the original (http://www.alliedcatalogs.com/html/catalogs_additional/1930/hr139.html) on 2017-07-11. Retrieved 2017-07-11.
76. "Capacitor MF - MMFD Conversion Chart" (<https://www.justradios.com/MFMMFD.html>). www.justradios.com.
77. Miller, Charles (2011). *Illustrated Guide to the National Electrical Code* (<https://books.google.com/books?id=RSsJAAAAQBAJ&pg=PA445>). Cengage Learning. p. 445.
78. Shinn, Eric; et al. (2012). "Nuclear energy conversion with stacks of graphene nanocapacitors". *Complexity*. **18** (3): 24–27. Bibcode:2013Cmplx..18c..24S (<https://ui.adsabs.harvard.edu/abs/2013Cmplx..18c..24S>). doi:10.1002/cplx.21427 (<https://doi.org/10.1002%2Fcplx.21427>).
79. Bezryadin, A.; Belkin, A.; et al. (2017). "Large energy storage efficiency of the dielectric layer of graphene nanocapacitors". *Nanotechnology*. **28** (49): 495401. arXiv:2011.11867 (<https://arxiv.org/abs/2011.11867>). Bibcode:2017Nanot..28W5401B (<https://ui.adsabs.harvard.edu/abs/2017Nanot..28W5401B>). doi:10.1088/1361-6528/aa935c (<https://doi.org/10.1088%2F1361-6528%2Fa935c>). PMID 29027908 (<https://pubmed.ncbi.nlm.nih.gov/29027908>). S2CID 44693636 (<https://api.semanticscholar.org/CorpusID:44693636>).
80. Floyd, Thomas L. (2017). *Electronic Devices* (<https://books.google.com/books?id=vg41vgAACAAJ&pg=PA10>). Pearson. p. 10. ISBN 978-0-13441444-7.
81. *Pulsed Power* by Gennady A. Mesyats -- Springer 2005 Page 1--5
82. Alam, Mohammed; Azarian, Michael H.; Osterman, Michael; Pecht, Michael (2010). "Effectiveness of embedded capacitors in reducing the number of surface mount capacitors for decoupling applications". *Circuit World*. **36** (1): 22. doi:10.1108/03056121011015068 (<https://doi.org/10.1108%2F03056121011015068>).
83. Downie, Neil A and Mathilde Pradier, 'Method and apparatus for monitoring fluid pressure", US Patent 7526961 (2009)
84. "Some Xenon Strobe and Flash Safety Hints" (<http://donklipstein.com/xesafe.html>). donklipstein.com.

Bibliography

- Dorf, Richard C.; Svoboda, James A. (2001). *Introduction to Electric Circuits* (<https://books.google.com/books?id=l-weAQAAIAAJ>) (5th ed.). New York: John Wiley & Sons. ISBN 978-0-47138689-6.
- Philosophical Transactions of the Royal Society LXXII, Appendix 8, 1782 (Volta coins the word *condenser*)
- Ulaby, Fawwaz Tayssir (1999). *Fundamentals of Applied Electromagnetics* (https://books.google.com/books?id=a_C8QgAACAAJ) (2nd ed.). Upper Saddle River, New Jersey, USA: Prentice Hall. ISBN 978-0-13011554-6.
- Schroder, Dieter K. (2006). *Semiconductor Material and Device Characterization* (<https://books.google.com/books?id=OX2cHKJWCKgC&pg=PA305>) (3rd ed.). Wiley. p. 270 ff. ISBN 978-0-47173906-7.
- Sze, Simon M.; Ng, Kwok K. (2006). *Physics of Semiconductor Devices* (<https://books.google.com/books?id=o4unkmHBHb8C>) (3rd ed.). Wiley. ISBN 978-0-47006830-4.

Further reading

- *Tantalum and Niobium-Based Capacitors – Science, Technology, and Applications*; 1st Ed; Yuri Freeman; Springer; 120 pages; 2018; ISBN 978-3-31967869-6.
- *Capacitors*; 1st Ed; R. P. Deshpande; McGraw-Hill; 342 pages; 2014; ISBN 978-0-07184856-5.

- *The Capacitor Handbook*; 1st Ed; Cletus Kaiser; Van Nostrand Reinhold; 124 pages; 1993; ISBN 978-9-40118092-4.
- *Understanding Capacitors and their Uses*; 1st Ed; William Mullin; Sams Publishing; 96 pages; 1964. (archive) (<https://worldradiohistory.com/BOOKSHELF-ARH/Sams-Books/Sams-Understanding-Capacitors-And-Their-Uses-1964-Mullin.pdf>)
- *Fixed and Variable Capacitors*; 1st Ed; G. W. A. Dummer and Harold Nordenberg; Maple Press; 288 pages; 1960. (archive) (<https://archive.org/details/FixedAndVariableCapacitors/>)
- *The Electrolytic Capacitor*; 1st Ed; Alexander Georgiev; Murray Hill Books; 191 pages; 1945. (archive) (<https://archive.org/details/TheElectrolyticCapacitor/>)

External links

- The First Condenser – A Beer Glass (<http://www.sparkmuseum.com/BOOKLEYDEN.HTM>) – SparkMuseum
- How Capacitors Work (<http://electronics.howstuffworks.com/capacitor.htm/printable>) – Howstuffworks
- Capacitor Tutorial (<http://www.robotplatform.com/electronics/capacitor/capacitor.html>)

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Capacitor&oldid=1292099951>"



Printed electronics

Printed electronics is a set of printing methods used to create electrical devices on various substrates. Printing typically uses common printing equipment suitable for defining patterns on material, such as screen printing, flexography, gravure, offset lithography, and inkjet. By electronic-industry standards, these are low-cost processes. Electrically functional electronic or optical inks are deposited on the substrate, creating active or passive devices, such as thin film transistors, capacitors, coils, and resistors. Some researchers expect printed electronics to facilitate widespread, very low-cost, low-performance electronics for applications such as flexible displays, smart labels, decorative and animated posters, and active clothing that do not require high performance.^[1]



Gravure printing of electronic structures on paper

The term *printed electronics* is often related to organic electronics or plastic electronics, in which one or more inks are composed of carbon-based compounds.^[2] These other terms refer to the ink material, which can be deposited by solution-based, vacuum-based, or other processes. Printed electronics, in contrast, specifies the process, and, subject to the specific requirements of the printing process selected, can utilize any solution-based material. This includes organic semiconductors, inorganic semiconductors, metallic conductors, nanoparticles, and nanotubes. The solution usually consist of filler materials dispersed in a suitable solvent. The most commonly used solvents include ethanol, xylene, Dimethylformamide (DMF), Dimethyl sulfoxide (DMSO), toluene and water, whereas, the most common conductive fillers include silver nanoparticles, silver flakes, carbon black, graphene, carbon nanotubes, conductive polymers (such as polyaniline and polypyrrole), and metal powders (such as copper or nickel). Considering the environmental impacts of the organic solvents, researchers are now focused on developing printable inks using water.^{[3][4][5]}

For the preparation of printed electronics nearly all industrial printing methods are employed. Similar to conventional printing, printed electronics applies ink layers one atop another.^[6] So the coherent development of printing methods and ink materials are the field's essential tasks.^[7]

The most important benefit of printing is low-cost volume fabrication. The lower cost enables use in more applications.^[8] An example is RFID-systems, which enable contactless identification in trade and transport. In some domains, such as light-emitting diodes printing does not impact performance.^[6] Printing on flexible substrates allows electronics to be placed on curved surfaces, for example: printing solar cells on vehicle roofs. More typically, conventional semiconductors justify their much higher costs by providing much higher performance.

Resolution, registration, thickness, holes, materials

The maximum required resolution of structures in conventional printing is determined by the human eye. Feature sizes smaller than approximately 20 µm cannot be distinguished by the human eye and consequently exceed the capabilities of conventional printing processes.^[9] In contrast, higher resolution and smaller structures are necessary in most electronics printing, because they directly affect circuit density and functionality (especially transistors). A similar requirement holds for the precision with which layers are printed on top of each other (layer to layer registration).

Control of thickness, holes, and material compatibility (wetting, adhesion, solubility) are essential, but matter in conventional printing only if the eye can detect them. Conversely, the visual impression is irrelevant for printed electronics.^[10]

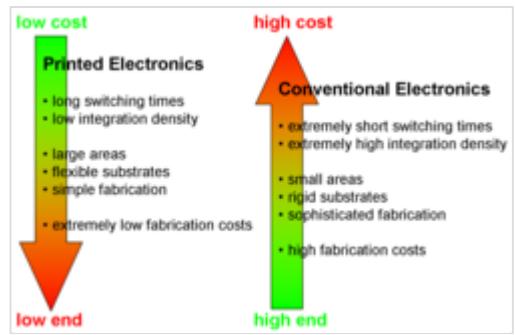
Printing technologies

The attraction of printing technology for the fabrication of electronics mainly results from the possibility of preparing stacks of micro-structured layers (and thereby thin-film devices) in a much simpler and cost-effective way compared to conventional electronics.^[11] Also, the ability to implement new or improved functionalities (e.g. mechanical flexibility) plays a role. The selection of the printing method used is determined by requirements concerning printed layers, by the properties of printed materials as well as economic and technical considerations of the final printed products.

Printing technologies divide between sheet-based and roll-to-roll-based approaches. Sheet-based inkjet and screen printing are best for low-volume, high-precision work. Gravure, offset and flexographic printing are more common for high-volume production, such as solar cells, reaching 10,000 square meters per hour (m^2/h).^{[9][11]} While offset and flexographic printing are mainly used for inorganic^{[12][13]} and organic^{[14][15]} conductors (the latter also for dielectrics),^[16] gravure printing is especially suitable for quality-sensitive layers like organic semiconductors and semiconductor/dielectric-interfaces in transistors, due to high layer quality.^[16] If high resolution is needed, gravure is also suitable for inorganic^[17] and organic^[18] conductors. Organic field-effect transistors and integrated circuits can be prepared completely by means of mass-printing methods.^[16]

Inkjet printing

Inkjets are flexible and versatile, and can be set up with relatively low effort.^[19] However, inkjets offer lower throughput of around 100 m^2/h and lower resolution (ca. 50 µm).^[9] It is well suited for low-viscosity, soluble materials like organic semiconductors. With high-viscosity materials, like organic dielectrics, and dispersed particles, like inorganic metal inks, difficulties due to nozzle clogging occur. Because ink is deposited via droplets, thickness and dispersion homogeneity is reduced. Using many nozzles simultaneously and pre-structuring the substrate allows improvements in productivity and



Printed and conventional electronics as complementary technologies.

resolution, respectively. However, in the latter case non-printing methods must be employed for the actual patterning step.^[20] Inkjet printing is preferable for organic semiconductors in organic field-effect transistors (OFETs) and organic light-emitting diodes (OLEDs), but also OFETs completely prepared by this method have been demonstrated.^[21] Frontplanes^[22] and backplanes^[23] of OLED-displays, integrated circuits,^[24] organic photovoltaic cells (OPVCs)^[25] and other devices can be prepared with inkjets.

Screen printing

Screen printing is appropriate for fabricating electrics and electronics due to its ability to produce patterned, thick layers from paste-like materials. This method can produce conducting lines from inorganic materials (e.g. for circuit boards and antennas), but also insulating and passivating layers, whereby layer thickness is more important than high resolution. Its $50 \text{ m}^2/\text{h}$ throughput and $100 \mu\text{m}$ resolution are similar to inkjets.^[9] This versatile and comparatively simple method is used mainly for conductive and dielectric layers,^{[26][27]} but also organic semiconductors, e.g. for OPVCs,^[28] and even complete OFETs^[22] can be printed.

Aerosol jet printing

Aerosol Jet Printing (also known as Maskless Mesoscale Materials Deposition or M3D)^[29] is another material deposition technology for printed electronics. The Aerosol Jet process begins with atomization of an ink, via ultrasonic or pneumatic means, producing droplets on the order of one to two micrometers in diameter. The droplets then flow through a virtual impactor which deflects the droplets having lower momentum away from the stream. This step helps maintaining a tight droplet size distribution. The droplets are entrained in a gas stream and delivered to the print head. Here, an annular flow of clean gas is introduced around the aerosol stream to focus the droplets into a tightly collimated beam of material. The combined gas streams exit the print head through a converging nozzle that compresses the aerosol stream to a diameter as small as $10 \mu\text{m}$. The jet of droplets exits the print head at high velocity (~ 50 meters/second) and impinges upon the substrate.

Electrical interconnects, passive and active components^[30] are formed by moving the print head, equipped with a mechanical stop/start shutter, relative to the substrate. The resulting patterns can have features ranging from $10 \mu\text{m}$ wide, with layer thicknesses from tens of nanometers to $>10 \mu\text{m}$.^[31] A wide nozzle print head enables efficient patterning of millimeter size electronic features and surface coating applications. All printing occurs without the use of vacuum or pressure chambers. The high exit velocity of the jet enables a relatively large separation between the print head and the substrate, typically 2–5 mm. The droplets remain tightly focused over this distance, resulting in the ability to print conformal patterns over three dimensional substrates.

Despite the high velocity, the printing process is gentle; substrate damage does not occur and there is generally minimal splatter or overspray from the droplets.^[32] Once patterning is complete, the printed ink typically requires post treatment to attain final electrical and mechanical properties. Post-treatment is driven more by the specific ink and substrate combination than by the printing process. A wide range of materials has been successfully deposited with the Aerosol Jet process, including diluted thick film pastes, conducting polymer inks,^[33] thermosetting polymers such as UV-curable epoxies, and solvent-based polymers like polyurethane and polyimide, and biologic materials.^[34]

Recently, printing paper was proposed to be used as the substrate of the printing. Highly conductive (close to bulk copper) and high-resolution traces can be printed on foldable and available office printing papers, with 80°Celsius curing temperature and 40 minutes of curing time.^[35]

Evaporation printing

Evaporation printing uses a combination of high precision screen printing with material vaporization to print features to 5 μm . This method uses techniques such as thermal, e-beam, sputter and other traditional production technologies to deposit materials through a high precision shadow mask (or stencil) that is registered to the substrate to better than 1 μm . By layering different mask designs and/or adjusting materials, reliable, cost-effective circuits can be built additively, without the use of photo-lithography.

Other methods

Other methods with similarities to printing, among them microcontact printing and nano-imprint lithography are of interest.^[36] Here, μm - and nm-sized layers, respectively, are prepared by methods similar to stamping with soft and hard forms, respectively. Often the actual structures are prepared subtractively, e.g. by deposition of etch masks or by lift-off processes. For example, electrodes for OFETs can be prepared.^{[37][38]} Sporadically pad printing is used in a similar manner.^[39] Occasionally so-called transfer methods, where solid layers are transferred from a carrier to the substrate, are considered printed electronics.^[40] Electrophotography is currently not used in printed electronics.

Materials

Both organic and inorganic materials are used for printed electronics. Ink materials must be available in liquid form, for solution, dispersion or suspension.^[41] They must function as conductors, semiconductors, dielectrics, or insulators. Material costs must be fit for the application.

Electronic functionality and printability can interfere with each other, mandating careful optimization.^[10] For example, a higher molecular weight in polymers enhances conductivity, but diminishes solubility. For printing, viscosity, surface tension and solid content must be tightly controlled. Cross-layer interactions such as wetting, adhesion, and solubility as well as post-deposition drying procedures affect the outcome. Additives often used in conventional printing inks are unavailable, because they often defeat electronic functionality.

Material properties largely determine the differences between printed and conventional electronics. Printable materials provide decisive advantages beside printability, such as mechanical flexibility and functional adjustment by chemical modification (e.g. light color in OLEDs).^[42]

Printed conductors offer lower conductivity and charge carrier mobility.^[43]

With a few exceptions, inorganic ink materials are dispersions of metallic or semiconducting micro- and nano-particles. Semiconducting nanoparticles used include silicon^[44] and oxide semiconductors.^[45] Silicon is also printed as an organic precursor^[46] which is then converted by pyrolysis and annealing into crystalline silicon.

PMOS but not CMOS is possible in printed electronics.^[47]

Organic materials

Organic printed electronics integrates knowledge and developments from printing, electronics, chemistry, and materials science, especially from organic and polymer chemistry. Organic materials in part differ from conventional electronics in terms of structure, operation and functionality,^[48] which influences device and circuit design and optimization as well as fabrication method.^[49]

The discovery of conjugated polymers^[43] and their development into soluble materials provided the first organic ink materials. Materials from this class of polymers variously possess conducting, semiconducting, electroluminescent, photovoltaic and other properties. Other polymers are used mostly as insulators and dielectrics.

In most organic materials, hole transport is favored over electron transport.^[50] Recent studies indicate that this is a specific feature of organic semiconductor/dielectric-interfaces, which play a major role in OFETs.^[51] Therefore, p-type devices should dominate over n-type devices. Durability (resistance to dispersion) and lifetime is less than conventional materials.^[47]

Organic semiconductors include the conductive polymers poly(3,4-ethylene dioxitiophene), doped with poly(styrene sulfonate), (PEDOT:PSS) and poly(aniline) (PANI). Both polymers are commercially available in different formulations and have been printed using inkjet,^[52] screen^[26] and offset printing^[14] or screen,^[26] flexo^[15] and gravure^[18] printing, respectively.

Polymer semiconductors are processed using inkjet printing, such as poly(thiopene)s like poly(3-hexylthiophene) (P3HT)^[53] and poly(9,9-dioctylfluorene co-bithiophen) (F8T2).^[54] The latter material has also been gravure printed.^[16] Different electroluminescent polymers are used with inkjet printing,^[20] as well as active materials for photovoltaics (e.g. blends of P3HT with fullerene derivatives),^[55] which in part also can be deposited using screen printing (e.g. blends of poly(phenylene vinylene) with fullerene derivatives).^[28]

Printable organic and inorganic insulators and dielectrics exist, which can be processed with different printing methods.^[56]

Inorganic materials

Inorganic electronics provides highly ordered layers and interfaces that organic and polymer materials cannot provide.

Silver nanoparticles are used with flexo,^[13] offset^[57] and inkjet.^[58] Gold particles are used with inkjet.^[59]

A.C. electroluminescent (EL) multi-color displays can cover many tens of square meters, or be incorporated in watch faces and instrument displays. They involve six to eight printed inorganic layers, including a copper doped phosphor, on a plastic film substrate.^[60]

CIGS cells can be printed directly onto molybdenum coated glass sheets.

A printed gallium arsenide germanium solar cell demonstrated 40.7% conversion efficiency, eight times that of the best organic cells, approaching the best performance of crystalline silicon.^[60]

Substrates

Printed electronics allows the use of flexible substrates, which lowers production costs and allows fabrication of mechanically flexible circuits. While inkjet and screen printing typically imprint rigid substrates like glass and silicon, mass-printing methods nearly exclusively use flexible foil and paper. Poly(ethylene terephthalate)-foil (PET) is a common choice, due to its low cost and moderately high temperature stability.^[61] Poly(ethylene naphthalate)- (PEN) and poly(imide)-foil (PI) are higher performance, higher cost alternatives. Paper's low costs and manifold applications make it an attractive substrate, however, its high roughness and high wettability have traditionally made it problematic for electronics. This is an active research area,^[62] however, and print-compatible metal deposition techniques have been demonstrated that adapt to the rough 3D surface geometry of paper.^{[63][64]}

Other important substrate criteria are low roughness and suitable wet-ability, which can be tuned pre-treatment by use of coating or Corona discharge. In contrast to conventional printing, high absorbency is usually disadvantageous.

History

Albert Hanson, a German by birth, is credited to have introduced the concept of printed electronics. in 1903 he filled a patent for "Printed Wires," and thus printed electronics were born.^[65] Hanson proposed forming a Printed Circuit Board pattern on copper foil through cutting or stamping. The drawn elements were glued to the dielectric, in this case, paraffined paper.^[66] The first printed circuit was produced in 1936 by Paul Eisler, and that process was used for large-scale production of radios by the USA during World War II. Printed circuit technology was released for commercial use in the US in 1948 (Printed Circuits Handbook, 1995). In the over a half-century since its inception, printed electronics has evolved from the production of printed circuit boards (PCBs), through the everyday use of membrane switches, to today's RFID, photovoltaic and electroluminescent technologies.^[67] Today it is nearly impossible to look around a modern American household and not see devices that either uses printed electronic components or that are the direct result of printed electronic technologies. Widespread production of printed electronics for household use began in the 1960s when the Printed Circuit Board became the foundation for all consumer electronics. Since then printed electronics have become a cornerstone in many new commercial products.^[68]

The biggest trend in recent history when it comes to printed electronics is the widespread use of them in solar cells. In 2011, researchers from MIT created a flexible solar cell by inkjet printing on normal paper.^[69] In 2018, researchers at Rice University have developed organic solar cells which can be painted or printed onto surfaces. These solar cells have been shown to max out at fifteen percent efficiency.^[70] Konarka Technologies, now a defunct company in the US, was the pioneering company in producing inkjet solar cells. Today there are more than fifty companies across a diverse number of countries that are producing printed solar cells.

While printed electronics have been around since the 1960s, they are predicted to have a major boom in total revenue. As of 2011, the total printed electronic revenue was reported to be at \$12.385 (billion).^[71] A report by IDTechEx predicts the PE market will reach \$330 (billion) in 2027.^[72] A big reason for this increase in revenue is because of the incorporation of printed electronic into cellphones. Nokia was one of the companies that pioneered the idea of creating a "Morph" phone using printed electronics. Since then, Apple has implemented this technology into their iPhone XS, XS Max, and XR devices.^[73] Printed

electronics can be used to make all of the following components of a cellphone: 3D main antenna, GPS antenna, energy storage, 3D interconnections, multi-layer PCB, edge circuits, ITO jumpers, hermetic seals, LED packaging, and tactile feedback.

With the revolutionary discoveries and advantages that printed electronic gives to companies many large companies have made recent investments into this technology. In 2007, Soligie Inc. and Thinfilm Electronics entered into an agreement to combine IPs for soluble memory materials and functional materials printing to develop printed memory in commercial volumes.^[67] LG announce significant investment, potentially \$8.71 billion in OLEDs on Plastic. Sharp (Foxconn) will invest \$570m in pilot line for OLED displays. BOE announce potential \$6.8 billion in flexible AMOLED fab. Heliatek has secured €80m in additional funding for OPV manufacturing in Dresden. PragmatIC has raised ~ €20m from investors including Avery Dennison. Thinfilm invests in new production site in Silicon Valley (formerly owned by Qualcomm). Cambrios back in business after acquisition by TPK.^[72]

Applications

Printed electronics are in use or under consideration include wireless sensors in packaging, skin patches that communicate with the internet, and buildings that detect leaks to enable preventative maintenance. Most of these applications are still in the prototyping and development stages.^[74] There is a particularly growing interest for flexible smart electronic systems, including photovoltaic, sensing and processing devices, driven by the desire to extend and integrate the latest advances in (opto-)electronic technologies into a broad range of low-cost (even disposable) consumer products of our everyday life, and as tools to bring together the digital and physical worlds.^[75]

Norwegian company ThinFilm demonstrated roll-to-roll printed organic memory in 2009.^{[76][77][78][79]}

Another company, Rotimpres based in Spain, has successfully introduced applications on different markets as for instance; heaters for smart furniture or to prevent mist and capacitive switch for keyboards on white goods and industrial machines.^{[80][81]}

Standards development and activities

Technical standards and road-mapping initiatives are intended to facilitate value chain development (for sharing of product specifications, characterization standards, etc.) This strategy of standards development mirrors the approach used by silicon-based electronics over the past 50 years. Initiatives include:

- The IEEE Standards Association has published IEEE 1620-2004^[82] and IEEE 1620.1-2006.^[83]
- Similar to the well-established International Technology Roadmap for Semiconductors (ITRS), the International Electronics Manufacturing Initiative (iNEMI)^[84] has published a roadmap for printed and other organic electronics.

IPC—Association Connecting Electronics Industries has published three standards for printed electronics. All three have been published in cooperation with the Japan Electronic Packaging and Circuits Association (JPCA):

- IPC/JPCA-4921, Requirements for Printed Electronics Base Materials

- IPC/JPCA-4591, Requirements for Printed Electronics Functional Conductive Materials
- IPC/JPCA-2291, Design Guideline for Printed Electronics

These standards, and others in development, are part of IPC's Printed Electronics Initiative.

See also



- [Amorphous silicon](#)
- [Anilox rolls](#)
- [Chip tag](#)
- [Coating and printing processes](#)
- [Conductive ink](#)
- [Electronic paper](#)
- [Flexible battery](#)
- [Flexible electronics](#)
- [Laminar electronics](#)
- [Nanoparticle silicon](#)
- [Oligomer](#)
- [Organic electronics](#)

References

1. Coatanéa, E., Kantola, V., Kulovesi, J., Lahti, L., Lin, R., & Zavodchikova, M. (2009). Printed Electronics, Now and Future. In Neuvo, Y., & Ylönen, S. (eds.), Bit Bang – Rays to the Future. Helsinki University of Technology (TKK), MIDE, Helsinki University Print, Helsinki, Finland, 63-102. ISBN 978-952-248-078-1.
<http://lib.tkk.fi/Reports/2009/isbn9789522480781.pdf> - "Moreover, PE technology could provide a number of enabling factors like flexibility and robustness, allowing incorporation of electronics functions into objects that do not yet contain any active electronic components, e.g. toy applications, printed advertising material or electronic labels [...]."
2. "Printed & Flexible Electronics - IDTechEx Research Reports and Subscriptions" (<https://www.idtechex.com/en/reports/printed-and-flexible-electronics/1>). www.idtechex.com. Retrieved 2020-09-21.
3. Khan, Junaid; Mariatti, M; Zubir, Syazana A; Rusli, Arjulizan; Manaf, Asrulnizam Abd; Khirotdin, Rd Khairilhijra (29 January 2024). "Eco-friendly alkali lignin-assisted water-based graphene oxide ink and its application as a resistive temperature sensor". *Nanotechnology*. **35** (5): 055301. Bibcode:2024Nanot..35e5301K (<https://ui.adsabs.harvard.edu/abs/2024Nanot..35e5301K>). doi:[10.1088/1361-6528/ad06d4](https://doi.org/10.1088/1361-6528/ad06d4) (<https://doi.org/10.1088%2F1361-6528%2Fa06d4>). PMID [37879329](https://pubmed.ncbi.nlm.nih.gov/37879329/) (<https://pubmed.ncbi.nlm.nih.gov/37879329/>).
4. Khan, Junaid; Mariatti, M (1 September 2023). "In-situ graphene oxide reduction via inkjet printing using natural reducing inks". *Flexible and Printed Electronics*. **8** (3): 035009. doi:[10.1088/2058-8585/acf143](https://doi.org/10.1088/2058-8585/acf143) (<https://doi.org/10.1088%2F2058-8585%2Facf143>).
5. Khan, Junaid; Mariatti, M. (November 2022). "Effect of natural surfactant on the performance of reduced graphene oxide conductive ink". *Journal of Cleaner Production*. **376**: 134254. Bibcode:2022JCPro.37634254K (<https://ui.adsabs.harvard.edu/abs/2022JCPro.37634254K>). doi:[10.1016/j.jclepro.2022.134254](https://doi.org/10.1016/j.jclepro.2022.134254) (<https://doi.org/10.1016%2Fj.jclepro.2022.134254>).
6. Roth, H.-K.; et al. (2001). "Organische Funktionsschichten in Polymerelektronik und Polymersolarzellen". *Materialwissenschaft und Werkstofftechnik*. **32** (10): 789. doi:[10.1002/1521-4052\(200110\)32:10<789::AID-MAWE789>3.0.CO;2-E](https://doi.org/10.1002/1521-4052(200110)32:10<789::AID-MAWE789>3.0.CO;2-E) (<https://doi.org/10.1002%2F1521-4052%28200110%2932%3A10%3C789%3A%3AAID-MAWE789%3E3.0.CO%3B2-E>).

7. Thomas, D.J. (2016). "Integration of Silicon and Printed Electronics for Rapid Diagnostic Disease Biosensing". *Point of Care: The Journal of Near-Patient Testing & Technology*. **15** (2): 61–71. doi:10.1097/POC.0000000000000091 (<https://doi.org/10.1097%2FPOC.0000000000000091>). S2CID 77379659 (<https://api.semanticscholar.org/CorpusID:77379659>).
8. Xu, J.M.(Jimmy) (2000). "Plastic electronics and future trends in microelectronics". *Synthetic Metals*. **115** (1–3): 1–3. doi:10.1016/s0379-6779(00)00291-5 (<https://doi.org/10.1016%2Fs0379-6779%2800%2900291-5>).
9. A. Blayo and B. Pineaux, Joint sOC-EUSA Conference, Grenoble, 2005.
10. U. Fügmann et al., mstNews 2 (2006) 13.
11. J. R. Sheats, *Journal of Materials Research* 2004; 19 1974.
12. Harrey, P.M.; et al. (2002). "Capacitive-type humidity sensors fabricated using the offset lithographic printing process". *Sensors and Actuators B*. **87** (2): 226–232. Bibcode:2002SeAcB..87..226H (<https://ui.adsabs.harvard.edu/abs/2002SeAcB..87..226H>). doi:10.1016/s0925-4005(02)00240-x (<https://doi.org/10.1016%2Fs0925-4005%2802%2900240-x>).
13. J. Siden et al., Polytronic Conference, Wroclaw, 2005.
14. Zielke, D.; et al. (2005). "Polymer-based organic field-effect transistor using offset printed source/drain structures". *Applied Physics Letters*. **87** (12): 123508. Bibcode:2005ApPhL..87l3508Z (<https://ui.adsabs.harvard.edu/abs/2005ApPhL..87l3508Z>). doi:10.1063/1.2056579 (<https://doi.org/10.1063%2F1.2056579>).
15. Mäkelä, T.; et al. (2005). "Utilizing roll-to-roll techniques for manufacturing source-drain electrodes for all-polymer transistors". *Synthetic Metals*. **153** (1–3): 285–288. doi:10.1016/j.synthmet.2005.07.140 (<https://doi.org/10.1016%2Fj.synthmet.2005.07.140>).
16. Hübner, A.; et al. (2007). "Ring oscillator fabricated completely by means of mass-printing technologies". *Organic Electronics*. **8** (5): 480. doi:10.1016/j.orgel.2007.02.009 (<https://doi.org/10.1016%2Fj.orgel.2007.02.009>).
17. S. Leppavuori et al., Sensors and Actuators 41-42 (1994) 593.
18. Mäkelä, T.; et al. (2003). "Roll-to-roll method for producing polyaniline patterns on paper". *Synthetic Metals*. **135**: 41. doi:10.1016/s0379-6779(02)00753-1 (<https://doi.org/10.1016%2Fs0379-6779%2802%2900753-1>).
19. Parashkov, R.; et al. (2005). "Large Area Electronics Using Printing Methods". *Proceedings of the IEEE*. **93** (7): 1321–1329. doi:10.1109/jproc.2005.850304 (<https://doi.org/10.1109%2Fjproc.2005.850304>). S2CID 27061013 (<https://api.semanticscholar.org/CorpusID:27061013>).
20. de Gans, B.-J.; et al. (2004). "Inkjet Printing of Polymers: State of the Art and Future Developments". *Advanced Materials*. **16** (3): 203. Bibcode:2004AdM....16..203D (<https://ui.adsabs.harvard.edu/abs/2004AdM....16..203D>). doi:10.1002/adma.200300385 (<https://doi.org/10.1002%2Fadma.200300385>).
21. Subramanian, V.; et al. (2005). "Progress Toward Development of All-Printed RFID Tags: Materials, Processes, and Devices". *Proceedings of the IEEE*. **93** (7): 1330. doi:10.1109/jproc.2005.850305 (<https://doi.org/10.1109%2Fjproc.2005.850305>). S2CID 8915461 (<https://api.semanticscholar.org/CorpusID:8915461>).
22. S. Holdcroft, *Advanced Materials* 2001; 13 1753.
23. Arias, A.C.; et al. (2004). "All jet-printed polymer thin-film transistor active-matrix backplanes". *Applied Physics Letters*. **85** (15): 3304. Bibcode:2004ApPhL..85.3304A (<https://ui.adsabs.harvard.edu/abs/2004ApPhL..85.3304A>). doi:10.1063/1.1801673 (<https://doi.org/10.1063%2F1.1801673>).
24. Sirringhaus, H.; et al. (2000). "High-Resolution Inkjet Printing of All-Polymer Transistor Circuits". *Science*. **290** (5499): 2123–2126. Bibcode:2000Sci...290.2123S (<https://ui.adsabs.harvard.edu/abs/2000Sci...290.2123S>). doi:10.1126/science.290.5499.2123 (<https://doi.org/10.1126%2Fscience.290.5499.2123>). PMID 11118142 (<https://pubmed.ncbi.nlm.nih.gov/11118142>).

25. V.G. Shah and D.B. Wallace, IMAPS Conference, Long Beach, 2004.
26. Bock, K.; et al. (2005). "Polymer Electronics Systems - Polytronics". *Proceedings of the IEEE*. **93** (8): 1400–1406. doi:[10.1109/jproc.2005.851513](https://doi.org/10.1109/jproc.2005.851513) (<https://doi.org/10.1109%2Fjproc.2005.851513>). S2CID 23177369 (<https://api.semanticscholar.org/CorpusID:23177369>).
27. Bao, Z.; et al. (1997). "High-Performance Plastic Transistors Fabricated by Printing Techniques". *Chemistry of Materials*. **9** (6): 1299–1301. doi:[10.1021/cm9701163](https://doi.org/10.1021/cm9701163) (<https://doi.org/10.1021%2Fcm9701163>).
28. Shaheen, S.E.; et al. (2001). "Fabrication of bulk heterojunction plastic solar cells by screen printing". *Applied Physics Letters*. **79** (18): 2996. Bibcode:[2001ApPhL..79.2996S](https://ui.adsabs.harvard.edu/abs/2001ApPhL..79.2996S) (<https://ui.adsabs.harvard.edu/abs/2001ApPhL..79.2996S>). doi:[10.1063/1.1413501](https://doi.org/10.1063/1.1413501) (<https://doi.org/10.1063%2F1.1413501>).
29. M. Renn, US Patent number 7,485,345 B2. Page 3.
30. J.H. Cho et al, Nature Materials, 19 October 2008.
31. B. Kahn, *Organic and Printed Electronics*, Volume 1, Issue 2 (2007).
32. B. H. King et al, Photovoltaic Specialists Conference (PVSC), 2009 34th IEEE.
33. Fisher, Christine; Warmack, Bruce J.; Yu, Yongchao; Skolrood, Lydia N.; Li, Kai; Joshi, Pooran C.; Saito, Tomonori; Aytug, Tolga (2021-04-19). "All-aerosol-jet-printed highly sensitive and selective polyaniline-based ammonia sensors: a route toward low-cost, low-power gas detection" (<https://doi.org/10.1007/s10853-021-06080-0>). *Journal of Materials Science*. **56** (22): 12596–12606. Bibcode:[2021JMatS..5612596F](https://ui.adsabs.harvard.edu/abs/2021JMatS..5612596F) (<https://ui.adsabs.harvard.edu/abs/2021JMatS..5612596F>). doi:[10.1007/s10853-021-06080-0](https://doi.org/10.1007/s10853-021-06080-0) (<https://doi.org/10.1007%2Fs10853-021-06080-0>). ISSN 1573-4803 (<https://search.worldcat.org/issn/1573-4803>). OSTI 1798595 (<https://www.osti.gov/biblio/1798595>). S2CID 233303736 (<https://api.semanticscholar.org/CorpusID:233303736>).
34. Ingo Grunwald et al, 2010 Biofabrication 2 014106.
35. Chen, Yi-Dan; Nagarajan, Vijayasarathi; Rosen, David W.; Yu, Wenwei; Huang, Shao Ying (Oct 2020). "Wireless Power Transfer via Strongly Coupled Magnetic Resonances". *Journal of Manufacturing Processes*. **58**: 55–66. doi:[10.1016/j.jmapro.2020.07.064](https://doi.org/10.1016/j.jmapro.2020.07.064) (<https://doi.org/10.1016%2Fj.jmapro.2020.07.064>).
36. Gate, B.D.; et al. (2005). "New Approaches to Nanofabrication: Molding, Printing, and Other Techniques". *Chemical Reviews*. **105** (4): 1171–96. doi:[10.1021/cr030076o](https://doi.org/10.1021/cr030076o) (<https://doi.org/10.1021%2Fcr030076o>). PMID 15826012 (<https://pubmed.ncbi.nlm.nih.gov/15826012>).
37. Li, D.; Guo, L.J. (2006). "Micron-scale organic thin film transistors with conducting polymer electrodes patterned by polymer inking and stamping" (https://deepblue.lib.umich.edu/bitstream/2027.42/87779/2/063513_1.pdf) (PDF). *Applied Physics Letters*. **88** (6): 063513. Bibcode:[2006ApPhL..88f3513L](https://ui.adsabs.harvard.edu/abs/2006ApPhL..88f3513L) (<https://ui.adsabs.harvard.edu/abs/2006ApPhL..88f3513L>). doi:[10.1063/1.2168669](https://doi.org/10.1063/1.2168669) (<https://doi.org/10.1063%2F1.2168669>). hdl:[2027.42%2F87779](https://hdl.handle.net/2027.42%2F87779) (<https://hdl.handle.net/2027.42%2F87779>).
38. Leising, G.; et al. (2006). "Nanoimprinted devices for integrated organic electronics". *Microelectronics Engineering*. **83** (4–9): 831. doi:[10.1016/j.mee.2006.01.241](https://doi.org/10.1016/j.mee.2006.01.241) (<https://doi.org/10.1016%2Fj.mee.2006.01.241>).
39. Knobloch, A.; et al. (2004). "Fully printed integrated circuits from solution processable polymers". *Journal of Applied Physics*. **96** (4): 2286. Bibcode:[2004JAP....96.2286K](https://ui.adsabs.harvard.edu/abs/2004JAP....96.2286K) (<https://ui.adsabs.harvard.edu/abs/2004JAP....96.2286K>). doi:[10.1063/1.1767291](https://doi.org/10.1063/1.1767291) (<https://doi.org/10.1063%2F1.1767291>).
40. Hines, D.R.; et al. (2007). "Transfer printing methods for the fabrication of flexible organic electronics". *Journal of Applied Physics*. **101** (2): 024503–024503–9. Bibcode:[2007JAP...101b4503H](https://ui.adsabs.harvard.edu/abs/2007JAP...101b4503H) (<https://ui.adsabs.harvard.edu/abs/2007JAP...101b4503H>). doi:[10.1063/1.2403836](https://doi.org/10.1063/1.2403836) (<https://doi.org/10.1063%2F1.2403836>).
41. Z. Bao, *Advanced Materials* 2000; 12: 227.

42. Moliton; Hiorns, R.C. (2004). "Review of electronic and optical properties of semiconducting π-conjugated polymers: applications in optoelectronics". *Polymer International*. **53** (10): 1397–1412. doi:10.1002/pi.1587 (<https://doi.org/10.1002%2Fpi.1587>).
43. http://nobelprize.org/nobel_prizes/chemistry/laureates/2000/chemadv.pdf Nobel prize in chemistry, 2000
44. Maennl, U.; et al. (2013). "Interfacial and Network Characteristics of Silicon Nanoparticle Layers Used in Printed Electronics". *Japanese Journal of Applied Physics*. **52** (5S1): 05DA11. Bibcode:2013JaJAP..52eDA11M (<https://ui.adsabs.harvard.edu/abs/2013JaJAP..52eDA11M>). doi:10.7567/JJAP.52.05DA11 (<https://doi.org/10.7567%2FJJAP.52.05DA11>).
45. Faber, H.; et al. (2009). "Low-Temperature Solution-Processed Memory Transistors Based on Zinc Oxide Nanoparticles". *Advanced Materials*. **21** (30): 3099. Bibcode:2009AdM....21.3099F (<https://ui.adsabs.harvard.edu/abs/2009AdM....21.3099F>). doi:10.1002/adma.200900440 (<https://doi.org/10.1002%2Fadma.200900440>).
46. Shimoda, T.; et al. (2006). "Solution-processed silicon films and transistors". *Nature*. **440** (7085): 783–786. Bibcode:2006Natur.440..783S (<https://ui.adsabs.harvard.edu/abs/2006Natur.440..783S>). doi:10.1038/nature04613 (<https://doi.org/10.1038%2Fnature04613>). PMID 16598254 (<https://pubmed.ncbi.nlm.nih.gov/16598254>). S2CID 4344708 (<https://api.semanticscholar.org/CorpusID:4344708>).
47. de Leeuw, D. M.; et al. (1997). "Stability of n-type doped conducting polymers and consequences for polymeric microelectronic devices". *Synthetic Metals*. **87**: 53. doi:10.1016/s0379-6779(97)80097-5 (<https://doi.org/10.1016%2Fs0379-6779%2897%2980097-5>).
48. Vardeny, Z.V.; et al. (2005). "Fundamental research needs in organic electronic materials". *Synthetic Metals*. **148**: 1. doi:10.1016/j.synthmet.2004.09.001 (<https://doi.org/10.1016%2Fj.synthmet.2004.09.001>).
49. H. Kempa et al., it 3 (2008) 167.
50. Fachetti (2007). "Semiconductors for organic transistors" (<https://doi.org/10.1016%2FS1369-7021%2807%2970017-2>). *Materials Today*. **10** (3): 38. doi:10.1016/S1369-7021(07)70017-2 (<https://doi.org/10.1016%2FS1369-7021%2807%2970017-2>).
51. Zaumseil, J.; Sirringhaus, H. (2007). "Electron and Ambipolar Transport in Organic Field-Effect Transistors". *Chemical Reviews*. **107** (4): 1296–1323. doi:10.1021/cr0501543 (<https://doi.org/10.1021%2Fcr0501543>). PMID 17378616 (<https://pubmed.ncbi.nlm.nih.gov/17378616>).
52. Bharathan, J.; Yang, Y. (2006). "Polymer electroluminescent devices processed by inkjet printing: I. Polymer light-emitting logo". *Applied Physics Letters*. **72** (21): 2660. Bibcode:1998ApPhL..72.2660B (<https://ui.adsabs.harvard.edu/abs/1998ApPhL..72.2660B>). doi:10.1063/1.121090 (<https://doi.org/10.1063%2F1.121090>).
53. Speakman, S.P.; et al. (2001). "High performance organic semiconducting thin films: Ink jet printed polythiophene [rr-P3HT]". *Organic Electronics*. **2** (2): 65. doi:10.1016/S1566-1199(01)00011-8 (<https://doi.org/10.1016%2FS1566-1199%2801%2900011-8>).
54. Paul, K.E.; et al. (2003). "Additive jet printing of polymer thin-film transistors". *Applied Physics Letters*. **83** (10): 2070. Bibcode:2003ApPhL..83.2070P (<https://ui.adsabs.harvard.edu/abs/2003ApPhL..83.2070P>). doi:10.1063/1.1609233 (<https://doi.org/10.1063%2F1.1609233>).
55. Aernouts, T.; et al. (2008). "Polymer based organic solar cells using ink-jet printed active layers" (<https://lirias.kuleuven.be/handle/123456789/638267>). *Applied Physics Letters*. **92** (3): 033306. Bibcode:2008ApPhL..92c3306A (<https://ui.adsabs.harvard.edu/abs/2008ApPhL..92c3306A>). doi:10.1063/1.2833185 (<https://doi.org/10.1063%2F1.2833185>).
56. "Ion Gel Insulator" (https://web.archive.org/web/20111114011218/http://www.license.umn.edu/Products/Ion-Gel-as-a-Gate-Insulator-in-Field-Effect-Transistors_Z07062.aspx). Archived from the original (http://www.license.umn.edu/Products/Ion-Gel-as-a-Gate-Insulator-in-Field-Effect-Transistors_Z07062.aspx) on November 14, 2011.

57. Harrey, P. M.; et al. (2000). "Interdigitated Capacitors by Offset Lithography". *Journal of Electronics Manufacturing*. **10**: 69–77. doi:10.1142/s096031310000006x (<https://doi.org/10.1142/s096031310000006x>).
58. Perelaer, J.; et al. (2006). "Ink-jet Printing and Microwave Sintering of Conductive Silver Tracks" (<http://library.tue.nl/csp/dare/LinkToRepository.csp?recordnumber=621662>). *Advanced Materials*. **18** (16): 2101–2104. Bibcode:2006AdM....18.2101P (<https://ui.adsabs.harvard.edu/abs/2006AdM....18.2101P>). doi:10.1002/adma.200502422 (<https://doi.org/10.1002/adma.200502422>).
59. Noh, Y.-Y.; et al. (2007). "Downscaling of self-aligned, all-printed polymer thin-film transistors". *Nature Nanotechnology*. **2** (12): 784–789. Bibcode:2007NatNa....2..784N (<http://ui.adsabs.harvard.edu/abs/2007NatNa....2..784N>). doi:10.1038/nnano.2007.365 (<https://doi.org/10.1038/nnano.2007.365>). PMID 18654432 (<https://pubmed.ncbi.nlm.nih.gov/18654432>).
60. Mflex UK (formerly Pelikon) and elumin8, both in the UK, Emirates Technical Innovation Centre in Dubai, Schreiner in Germany and others are involved in EL displays. Spectrolab already offers commercially flexible solar cells based on various inorganic compounds. [1] (<https://archive.today/20070517164819/http://www.packagingessentials.com/indnews.asp?id=2007-03-22-15.57.31.000000>)
61. Carey, T., Cacovich, S., Divitini, G., Ren, J., Mansouri, A., Kim, J. M., ... & Torrisi, F. (2017). Fully inkjet-printed two-dimensional material field-effect heterojunctions for wearable and textile electronics. *Nature communications*, 8(1), 1-11. <https://doi.org/10.1038/s41467-017-01210-2>
62. Tobjörk, Daniel; Österbacka, Ronald (2011-03-23). "Paper Electronics". *Advanced Materials*. **23** (17): 1935–1961. Bibcode:2011AdM....23.1935T (<https://ui.adsabs.harvard.edu/abs/2011AdM....23.1935T>). doi:10.1002/adma.201004692 (<https://doi.org/10.1002/adma.201004692>). ISSN 0935-9648 (<https://search.worldcat.org/issn/0935-9648>). PMID 21433116 (<https://pubmed.ncbi.nlm.nih.gov/21433116>).
63. Grell, Max; Dincer, Can; Le, Thao; Lauri, Alberto; Nunez Bajo, Estefania; Kasimatis, Michael; Barandun, Giandrin; Maier, Stefan A.; Cass, Anthony E. G. (2018-11-09). "Autocatalytic Metallization of Fabrics Using Si Ink, for Biosensors, Batteries and Energy Harvesting" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7384005>). *Advanced Functional Materials*. **29** (1): 1804798. doi:10.1002/adfm.201804798 (<https://doi.org/10.1002/adfm.201804798>). ISSN 1616-301X (<https://search.worldcat.org/issn/1616-301X>). PMC 7384005 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7384005>). PMID 32733177 (<https://pubmed.ncbi.nlm.nih.gov/32733177>).
64. Vicente, António T.; Araújo, Andreia; Gaspar, Diana; Santos, Lídia; Marques, Ana C.; Mendes, Manuel J.; Pereira, Luís; Martins, Elvira Fortunato and Rodrigo (2017-02-22). *Optoelectronics and Bio Devices on Paper Powered by Solar Cells* (<https://www.intechopen.com/chapters/53636>). IntechOpen. ISBN 978-953-51-2936-3.
65. GB 4681 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=GB4681>), Hanson, Albert, "Printed Wires", issued 1903
66. "The printed circuit board is the base of modern electronics" (<https://web.archive.org/web/20190828061658/https://rostec.ru/en/news/4515084/>). rostec.ru. Rostec. November 24, 2014. Archived from the original (<https://rostec.ru/en/news/4515084/>) on August 28, 2019. Retrieved November 28, 2018.
67. Jacobs, John (2010). *An Investigation of Fundamental Competencies for Printed Electronics* (Thesis). Clemson University.
68. "Printing Electronics Just" (<https://web.archive.org/web/20130513194346/http://news.nationalgeographic.com/news/2013/13/130509-paper-printed-circuits-electronics-technology-science/>), National Geographic News, National Geographic Partners, LLC, 10 May 2013, archived from the original (<https://news.nationalgeographic.com/news/2013/13/130509-paper-printed-circuits-electronics-technology-science/>) on May 13, 2013, retrieved November 30, 2018

69. "While You're Pp, Print Me a Solar Cell" (<https://news.mit.edu/2011/printable-solar-cells-0711>), MIT News, MIT News Office, 11 July 2011, retrieved November 30, 2018
70. "Stretchy solar cells a step closer" (<https://www.printedelectronicsworld.com/articles/15804/stretchy-solar-cells-a-step-closer>), Printed Electronics World, IDTechEx, 15 November 2018, retrieved November 30, 2018
71. Zhang, Chuck, *Printed Electronics: Manufacturing Technologies and Applications* (http://thor.inemi.org/webdownload/2014/Printed_Elec_Zhang_051214.pdf) (PDF), Georgia Tech, retrieved November 30, 2018
72. Das, Raghu, *Printed Electronics: Markets, Technologies, Trends* (https://events.hkstp.org/events/Past_AIS/APAC2016/files/Keynote_1_Raghu_Das_Printed_Electronics.pdf) (PDF), IDTechEx, retrieved November 30, 2018
73. "New iPhones models support native 'background' NFC tag read function" (<https://www.printedelectronicsworld.com/articles/15387/new-iphones-models-support-native-background-nfc-tag-read-function>) (Press release). Thinfilm. Retrieved November 30, 2018 – via IDTechEx.
74. "Custom Printed Electronics" (<https://www.almax-rp.com/printed-electronics/>). Almax - RP. 30 December 2016. Retrieved 13 August 2021.
75. Vicente, António T.; Araújo, Andreia; Mendes, Manuel J.; Nunes, Daniela; Oliveira, Maria J.; Sanchez-Sobrado, Olalla; Ferreira, Marta P.; Águas, Hugo; Fortunato, Elvira; Martins, Rodrigo (2018-03-29). "Multifunctional cellulose-paper for light harvesting and smart sensing applications" (<https://pubs.rsc.org/en/content/articlelanding/2018/tc/c7tc05271e>). *Journal of Materials Chemistry C*. **6** (13): 3143–3181. doi:10.1039/C7TC05271E (<https://doi.org/10.1039%2FC7TC05271E>). ISSN 2050-7534 (<https://search.worldcat.org/issn/2050-7534>).
76. Thinfilm and InkTec awarded IDTechEx' Technical Development Manufacturing Award (https://www.printedelectronicsworld.com/articles/winners_of_the_idtechex_printed_electronics_europe_awards_00001368.asp) IDTechEx, April 15th 2009
77. PolyIC, ThinFilm announce pilot of volume printed plastic memories (<http://www.eetimes.com/electronics-news/4084606/PolyIC-ThinFilm-announce-pilot-of-volume-printed-plastic-memories>) EETimes, September 22nd 2009
78. All set for high-volume production of printed memories (http://www.printedelectronicsworld.com/articles/all_set_for_high_volume_production_of_printed_memories_00002179.asp) Printed Electronics World, April 12th 2010
79. Thin Film Electronics Plans to Provide 'Memory Everywhere' (<http://www.printedelectronicsnow.com/articles/2010/05/thin-film-electronics-plans-to-provide-memory-ever>) Printed Electronics Now, May 2010
80. Revolutionize Your Industrial Heating with Rotimpres (<https://printedelectronics.rotimpres.com/en/heaters-for-industry/>) Heater, October 14th 2024
81. Capacitive keyboards - Rotimpres (<https://printedelectronics.rotimpres.com/en/applications/capacitive-keyboards/>) Capacitive switch, October 14th 2024
82. "IEEE P1620 Index page" (<https://web.archive.org/web/20110610212227/http://grouper.ieee.org/groups/1620/>). Archived from the original (<http://grouper.ieee.org/groups/1620/>) on 2011-06-10. Retrieved 2006-11-30.
83. "IEEE P1620.1 Index page" (<https://web.archive.org/web/20110610212235/http://grouper.ieee.org/groups/1620/1/>). Archived from the original (<http://grouper.ieee.org/groups/1620/1/>) on 2011-06-10. Retrieved 2006-11-30.
84. "iNEMI | International Electronics Manufacturing Initiative" (<https://www.inemi.org/>). www.inemi.org.

Further reading

- *Printed Organic and Molecular Electronics*, edited by D. Gamota, P. Brazis, K. Kalyanasundaram, and J. Zhang (Kluwer Academic Publishers: New York, 2004). ISBN 1-4020-7707-6

External links

- Cleaner Electronics Research Group (http://www.dea.brunel.ac.uk/cleaner/Electronics_Projects/Handbook_1.htm) - Brunel University
- Printed Electronics conference/exhibition Asia (<https://web.archive.org/web/20080915190411/http://printedelectronics.idtechex.com/printedelectronicsasia08/en/>) USA (<https://web.archive.org/web/20080921033814/http://printedelectronics.idtechex.com/printedelectronicsusa08/en/>)
- New Nano Silver Powder Enables Flexible Printed Circuits (<https://web.archive.org/web/20081009011825/http://www.ferro.com/spotlight/New+Nano+Silver+Powder+Enables+Flexible+Printed+Circuits.htm>) (Ferro Corporation)
- Western Michigan University's Center for Advancement of Printed Electronics (CAPE) (<https://web.archive.org/web/20101214061056/http://www.wmich.edu/engineer/cape/facilities.php>) includes AccuPress gravure printer (http://www.printedelectronicsworld.com/articles/western_michigan_university_and_daetwyler_progress_gravure_printing_00001879.asp?sessionid=1)
- Major Trends in Gravure Printed Electronics June 2010 (<http://digitalcommons.calpoly.edu/grcsp/26/>)
- Printed Electronics – avistando el futuro. Printed Electronics en Español (<http://innocreatividad.com/2013/06/19/printed-electronics-avistando-el-futuro>)
- Organic Solar Cells - Theory and Practice (Coursera) (<https://www.coursera.org/learn/solar-cell>)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Printed_electronics&oldid=1291833731"

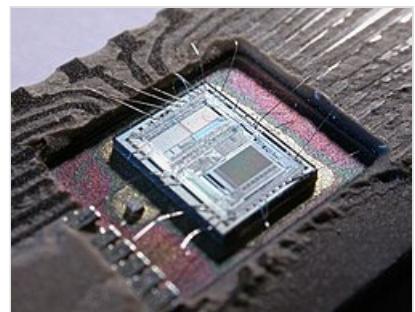


Electronic circuit

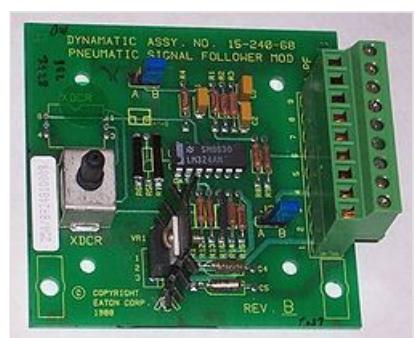
An **electronic circuit** is composed of individual electronic components, such as resistors, transistors, capacitors, inductors and diodes, connected by conductive wires or traces through which electric current can flow. It is a type of electrical circuit. For a circuit to be referred to as *electronic*, rather than *electrical*, generally at least one active component must be present. The combination of components and wires allows various simple and complex operations to be performed: signals can be amplified, computations can be performed, and data can be moved from one place to another.^[1]

Circuits can be constructed of discrete components connected by individual pieces of wire, but today it is much more common to create interconnections by photolithographic techniques on a laminated substrate (a printed circuit board or PCB) and solder the components to these interconnections to create a finished circuit. In an integrated circuit or IC, the components and interconnections are formed on the same substrate, typically a semiconductor such as doped silicon or (less commonly) gallium arsenide.^[2]

An electronic circuit can usually be categorized as an analog circuit, a digital circuit, or a mixed-signal circuit (a combination of analog circuits and digital circuits). The most widely used semiconductor device in electronic circuits is the MOSFET (metal–oxide–semiconductor field-effect transistor).^[3]



The die from an Intel 8742, an 8-bit microcontroller that includes a CPU, 128 bytes of RAM, 2048 bytes of EPROM, and I/O "data" on current chip

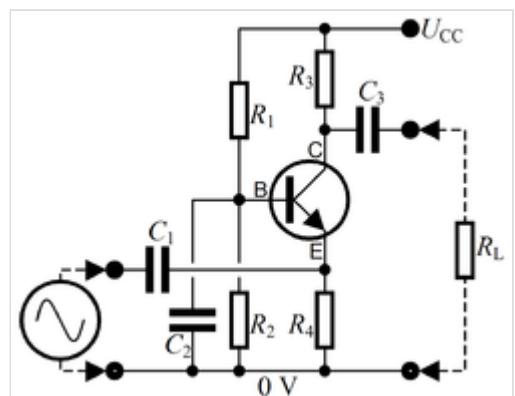


A circuit built on a printed circuit board (PCB)

Analog circuits

Analog electronic circuits are those in which current or voltage may vary continuously with time to correspond to the information being represented.

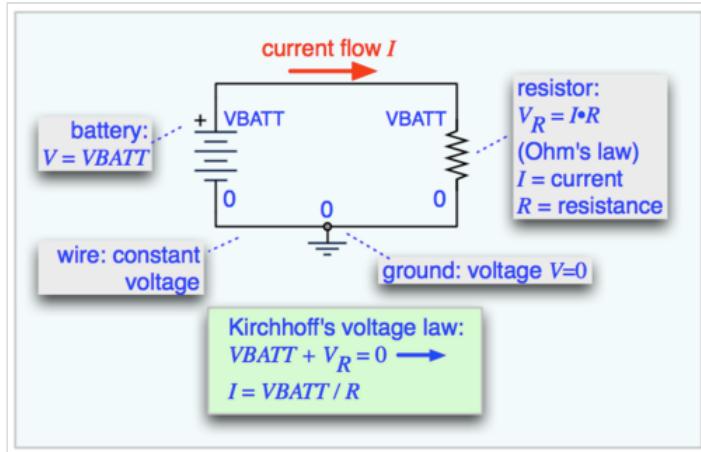
The basic components of analog circuits are wires, resistors, capacitors, inductors, diodes, and transistors. Analog circuits are very commonly represented in schematic diagrams, in which wires are shown as lines, and each component has a unique symbol. Analog circuit analysis employs Kirchhoff's circuit laws: all the currents at a node (a place where wires meet), and the voltage around a closed loop of wires is 0. Wires are usually treated as ideal zero-voltage interconnections; any resistance or reactance is captured by explicitly adding a parasitic element, such as a discrete



A circuit diagram representing an analog circuit, in this case a simple amplifier

resistor or inductor. Active components such as transistors are often treated as controlled current or voltage sources: for example, a field-effect transistor can be modeled as a current source from the source to the drain, with the current controlled by the gate-source voltage.

When the circuit size is comparable to a wavelength of the relevant signal frequency, a more sophisticated approach must be used, the distributed-element model. Wires are treated as transmission lines, with nominally constant characteristic impedance, and the impedances at the start and end determine transmitted and reflected waves on the line. Circuits designed according to this approach are distributed-element circuits. Such considerations typically become important for circuit boards at frequencies above a GHz; integrated circuits are smaller and can be treated as lumped elements for frequencies less than 10GHz or so.



A simple schematic showing wires, a resistor, and a battery

Digital circuits

In digital electronic circuits, electric signals take on discrete values, to represent logical and numeric values.^[4] These values represent the information that is being processed. In the vast majority of cases, binary encoding is used: one voltage (typically the more positive value) represents a binary '1' and another voltage (usually a value near the ground potential, 0 V) represents a binary '0'. Digital circuits make extensive use of transistors, interconnected to create logic gates that provide the functions of Boolean logic: AND, NAND, OR, NOR, XOR and combinations thereof. Transistors interconnected so as to provide positive feedback are used as latches and flip flops, circuits that have two or more metastable states, and remain in one of these states until changed by an external input. Digital circuits therefore can provide logic and memory, enabling them to perform arbitrary computational functions. (Memory based on flip-flops is known as static random-access memory (SRAM). Memory based on the storage of charge in a capacitor, dynamic random-access memory (DRAM), is also widely used.)

The design process for digital circuits is fundamentally different from the process for analog circuits. Each logic gate regenerates the binary signal, so the designer need not account for distortion, gain control, offset voltages, and other concerns faced in an analog design. As a consequence, extremely complex digital circuits, with billions of logic elements integrated on a single silicon chip, can be fabricated at low cost. Such digital integrated circuits are ubiquitous in modern electronic devices, such as calculators, mobile phone handsets, and computers. As digital circuits become more complex, issues of time delay, logic races, power dissipation, non-ideal switching, on-chip and inter-chip loading, and leakage currents, become limitations to circuit density, speed and performance.

Digital circuitry is used to create general purpose computing chips, such as microprocessors, and custom-designed logic circuits, known as application-specific integrated circuit (ASICs). Field-programmable gate arrays (FPGAs), chips with logic circuitry whose configuration can be modified after fabrication, are also widely used in prototyping and development.

Mixed-signal circuits

Mixed-signal or hybrid circuits contain elements of both analog and digital circuits. Examples include comparators, timers, phase-locked loops, analog-to-digital converters, and digital-to-analog converters. Most modern radio and communications circuitry uses mixed signal circuits. For example, in a receiver, analog circuitry is used to amplify and frequency-convert signals so that they reach a suitable state to be converted into digital values, after which further signal processing can be performed in the digital domain.

Design

Electronic circuit design comprises the analysis and synthesis of electronic circuits.

Prototyping

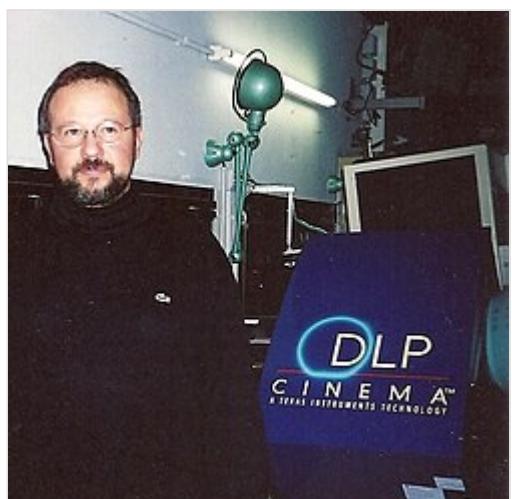
In electronics, prototyping means building an actual circuit to a theoretical design to verify that it works, and to provide a physical platform for debugging it if it does not. The prototype is often constructed using techniques such as wire wrapping or using a breadboard, stripboard or perfboard, with the result being a circuit that is electrically identical to the design but not physically identical to the final product.^[5]

Open-source tools like Fritzing exist to document electronic prototypes (especially the breadboard-based ones) and move toward physical production. Prototyping platforms such as Arduino also simplify the task of programming and interacting with a microcontroller.^[6] The developer can choose to deploy their invention as-is using the prototyping platform, or replace it with only the microcontroller chip and the circuitry that is relevant to their product.

A technician can quickly build a prototype (and make additions and modifications) using these techniques, but for volume production it is much faster and usually cheaper to mass-produce custom printed circuit boards than to produce these other kinds of prototype boards. The proliferation of quick-turn PCB fabrication and assembly companies has enabled the concepts of rapid prototyping to be applied to electronic circuit design. It is now possible, even with the smallest passive components and largest fine-pitch packages, to have boards fabricated, assembled, and even tested in a matter of days.



A simple electronic circuit prototype on a breadboard



Example of prototype in optoelectronics
(Texas Instruments, DLP Cinema
Prototype System)

References

1. Charles Alexander and Matthew Sadiku (2004). *Fundamentals of Electric Circuits*. McGraw-Hill.
2. Richard Jaeger (1997). *Microelectronic Circuit Design*. McGraw-Hill.
3. Golio, Mike; Golio, Janet (2018). *RF and Microwave Passive and Active Technologies* (<http://books.google.com/books?id=MCj9jxSVQKIC&pg=SA18-PA2>). CRC Press. p. 18-2. ISBN 9781420006728.
4. John Hayes (1993). *Introduction to Digital Logic Design*. Addison Wesley.
5. "PCB Rapid Prototype" (<http://www.wellpcb.com/news/pcb-rapid-prototype>). www.wellpcb.com. WellPCB. Retrieved 2017-06-01.
6. Trevennor, Alan (2012-10-17). *Practical AVR Microcontrollers: Games, Gadgets, and Home Automation with the Microcontroller Used in the Arduino* (<https://books.google.com/books?id=z-b10XVvN44C>). Apress. ISBN 9781430244462.

External links

- [Electronics Circuits Textbook](http://www.allaboutcircuits.com/textbook/) (<http://www.allaboutcircuits.com/textbook/>)
- [Electronics Fundamentals](http://www.rohm.com/web/global/en_index/) (http://www.rohm.com/web/global/en_index/)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Electronic_circuit&oldid=1281118598"



Flip-flop

Flip-flops are a simple type of footwear in which there is a band between the big toe and the other toes.

Flip-flop may also refer to:

Entertainment

- [Flip-Flop \(album\)](#), a 1989 album by Guadalcanal Diary
- [Flip-Flop \(audio drama\)](#), a 2003 audio drama based on the British television series *Doctor Who*
- "Flip-Flop" (*Will & Grace*), an episode of the television series *Will & Grace*
- [Flip Flop \(*Modern Family*\)](#), an episode of the television series *Modern Family*
- [Flip Flop \(*The Price Is Right*\)](#), a game on *The Price Is Right*
- "Flip Flop", a song by Megan Thee Stallion from the album [Traumazine](#), 2022
- [Flip and Flop](#), a 1983 video game
- [Flip or Flop](#), a U.S. television series on HGTV

Computers and electronics

- [Flip-flop \(electronics\)](#), a circuit with two stable states
- [Flip-flop \(programming\)](#), a Boolean expression with persistent state and two conditions

Sports

- A back [handspring \(gymnastics\)](#)
- A trick performed in the sport of freestyle kayaking ([Playboating](#))

Other uses

- [Flip \(mathematics\)](#), and flop – operations in algebraic geometry
- [Flip-flop \(politics\)](#), a sudden change of position on an issue
- [Flip-flop hub](#), a type of hub used in bicycle wheels
- [Flip–flop kinetics](#), a phenomenon in pharmacokinetics when a drug is released at a sustained rate instead of immediate release
- A common name of the African wood white butterfly ([Leptosia alcesta](#))
- [Flip flop, per top, bottom and versatile](#), a role reversal between two men during a single sexual encounter
- The translocation of a phospholipid in cell membranes carried out by [flippase](#) proteins

See also

- [Flippity and Flop](#), a pair of cartoon characters that appeared in theatrical shorts, 1945–1947
 - [Flip \(disambiguation\)](#)
 - [Flop \(disambiguation\)](#)
-

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Flip-flop&oldid=1257255322>"



Memory cell

Memory cell may refer to:

Biology

- Memory cells (motor cortex), found in the primary motor cortex (M1), a region located in the posterior portion of the frontal lobe of the brain.
- Memory B cell, an antibody producing cell
- Memory T cell, an infection fighting cell

Computing

- Memory cell (computing), a building block of computer memory and data storage

Retrieved from "https://en.wikipedia.org/w/index.php?title=Memory_cell&oldid=1104277470"

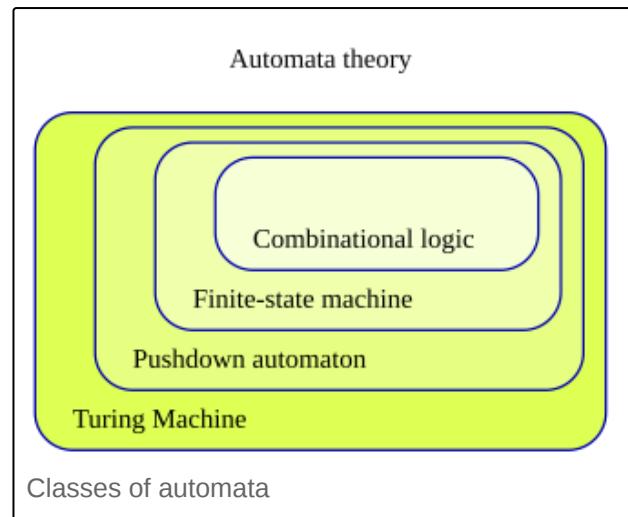


Combinational logic

In automata theory, **combinational logic** (also referred to as **time-independent logic**^[1]) is a type of digital logic that is implemented by Boolean circuits, where the output is a pure function of the present input only. This is in contrast to sequential logic, in which the output depends not only on the present input but also on the history of the input. In other words, sequential logic has memory while combinational logic does not.

Combinational logic is used in computer circuits to perform Boolean algebra on input signals and on stored data. Practical computer circuits normally contain a mixture of combinational and sequential logic. For example, the part of an arithmetic logic unit, or ALU, that does mathematical calculations is constructed using combinational logic. Other circuits used in computers, such as half adders, full adders, half subtractors, full subtractors, multiplexers, demultiplexers, encoders and decoders are also made by using combinational logic.

Practical design of combinational logic systems may require consideration of the finite time required for practical logical elements to react to changes in their inputs. Where an output is the result of the combination of several different paths with differing numbers of switching elements, the output may momentarily change state before settling at the final state, as the changes propagate along different paths.^[2]



Representation

Combinational logic is used to build circuits that produce specified outputs from certain inputs. The construction of combinational logic is generally done using one of two methods: a sum of products, or a product of sums. Consider the following truth table:

A	B	C	Result	Logical equivalent
F	F	F	F	$\neg A \wedge \neg B \wedge \neg C$
F	F	T	F	$\neg A \wedge \neg B \wedge C$
F	T	F	F	$\neg A \wedge B \wedge \neg C$
F	T	T	F	$\neg A \wedge B \wedge C$
T	F	F	T	$A \wedge \neg B \wedge \neg C$
T	F	T	F	$A \wedge \neg B \wedge C$
T	T	F	F	$A \wedge B \wedge \neg C$
T	T	T	T	$A \wedge B \wedge C$

Using sum of products, all logical statements which yield true results are summed, giving the result:

$$(A \wedge \neg B \wedge \neg C) \vee (A \wedge B \wedge C)$$

Using Boolean algebra, the result simplifies to the following equivalent of the truth table:

$$A \wedge ((\neg B \wedge \neg C) \vee (B \wedge C))$$

Logic formula minimization

Minimization (simplification) of combinational logic formulas is done using the following rules based on the laws of Boolean algebra:

$$(A \vee B) \wedge (A \vee C) = A \vee (B \wedge C)$$

$$(A \wedge B) \vee (A \wedge C) = A \wedge (B \vee C)$$

$$A \vee (A \wedge B) = A$$

$$A \wedge (A \vee B) = A$$

$$A \vee (\neg A \wedge B) = A \vee B$$

$$A \wedge (\neg A \vee B) = A \wedge B$$

$$(A \vee B) \wedge (\neg A \vee B) = B$$

$$(A \wedge B) \vee (\neg A \wedge B) = B$$

$$(A \wedge B) \vee (\neg A \wedge C) \vee (B \wedge C) = (A \wedge B) \vee (\neg A \wedge C)$$

$$(A \vee B) \wedge (\neg A \vee C) \wedge (B \vee C) = (A \vee B) \wedge (\neg A \vee C)$$

With the use of minimization (sometimes called logic optimization), a simplified logical function or circuit may be arrived upon, and the logic combinational circuit becomes smaller, and easier to analyse, use, or build.

See also

- [Asynchronous circuit](#)
- [Field-programmable gate array](#)
- [Formal verification](#)
- [Ladder logic](#)
- [Programmable logic controller](#)
- [Relay logic](#)
- [Sequential logic](#)
- [Tseytin transformation](#)

References

1. Savant, C.J. Jr.; Roden, Martin; Carpenter, Gordon (1991). *Electronic Design: Circuits and Systems*. Benjamin/Cummings Publishing Company. p. 682. [ISBN 0-8053-0285-9](#).
 2. Lewin, Douglas (1974). *Logical Design of Switching Circuits* (2nd ed.). Thomas Nelson and Sons. pp. 162–3. [ISBN 017-771044-6](#).
- Predko, Michael; Predko, Myke (2004). *Digital electronics demystified*. McGraw-Hill. [ISBN 0-07-144141-7](#).

External links

- Belton, D.; Bigwood, R. "Combinational Logic & Systems Tutorial Guide" (<https://web.archive.org/web/20131022174334/http://www.ee.surrey.ac.uk/Projects/Labview/combindex.html>). Archived from the original (<http://www.ee.surrey.ac.uk/Projects/Labview/combindex.html>) on 2013-10-22.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Combinational_logic&oldid=1291393142"



Sequential logic

In automata theory, **sequential logic** is a type of logic circuit whose output depends on the present value of its input signals and on the sequence of past inputs, the input history.^{[1][2][3][4]} This is in contrast to combinational logic, whose output is a function of only the present input. That is, sequential logic has state (*memory*) while combinational logic does not.

Sequential logic is used to construct finite-state machines, a basic building block in all digital circuitry. Virtually all circuits in practical digital devices are a mixture of combinational and sequential logic.

A familiar example of a device with sequential logic is a television set with "channel up" and "channel down" buttons.^[1] Pressing the "up" button gives the television an input telling it to switch to the next channel above the one it is currently receiving. If the television is on channel 5, pressing "up" switches it to receive channel 6. However, if the television is on channel 8, pressing "up" switches it to channel "9". In order for the channel selection to operate correctly, the television must be aware of which channel it is currently receiving, which was determined by past channel selections.^[1] The television stores the current channel as part of its state. When a "channel up" or "channel down" input is given to it, the sequential logic of the channel selection circuitry calculates the new channel from the input and the current channel.

Digital sequential logic circuits are divided into synchronous and asynchronous types. In synchronous sequential circuits, the state of the device changes only at discrete times in response to a clock signal. In asynchronous circuits the state of the device can change at any time in response to changing inputs.

Synchronous sequential logic

Nearly all sequential logic today is clocked or synchronous logic. In a synchronous circuit, an electronic oscillator called a clock (or clock generator) generates a sequence of repetitive pulses called the clock signal which is distributed to all the memory elements in the circuit. The basic memory element in synchronous logic is the flip-flop. The output of each flip-flop only changes when triggered by the clock pulse, so changes to the logic signals throughout the circuit all begin at the same time, at regular intervals, synchronized by the clock.

The output of all the storage elements (flip-flops) in the circuit at any given time, the binary data they contain, is called the state of the circuit. The state of the synchronous circuit only changes on clock pulses. At each cycle, the next state is determined by the current state and the value of the input signals when the clock pulse occurs.

The main advantage of synchronous logic is its simplicity. The logic gates which perform the operations on the data require a finite amount of time to respond to changes to their inputs. This is called propagation delay. The interval between clock pulses must be long enough so that all the logic gates have time to respond to the changes and their outputs "settle" to stable logic values before the next clock pulse occurs. As long as this condition is met (ignoring certain other details) the circuit is guaranteed to be stable and reliable. This determines the maximum operating speed of the synchronous circuit.

Synchronous logic has two main disadvantages:

- The maximum possible clock rate is determined by the slowest logic path in the circuit, otherwise known as the critical path. Every logical calculation, from the simplest to the most complex, must complete in one clock cycle. So logic paths that complete their calculations quickly are idle much of the time, waiting for the next clock pulse. Therefore, synchronous logic can be slower than asynchronous logic. One way to speed up synchronous circuits is to split complex operations into several simple operations which can be performed in successive clock cycles, a technique known as pipelining. This technique is extensively used in microprocessor design and helps to improve the performance of modern processors.
- The clock signal must be distributed to every flip-flop in the circuit. As the clock is usually a high-frequency signal, this distribution consumes a relatively large amount of power and dissipates much heat. Even the flip-flops that are doing nothing consume a small amount of power, thereby generating waste heat in the chip. In battery-powered devices, additional hardware and software complexity is required to reduce the clock speed or temporarily turn off the clock while the device is not being actively used, in order to maintain a usable battery life.

Asynchronous sequential logic

Asynchronous (clockless or self-timed) sequential logic is not synchronized by a clock signal; the outputs of the circuit change directly in response to changes in inputs. The advantage of asynchronous logic is that it can be faster than synchronous logic, because the circuit doesn't have to wait for a clock signal to process inputs. The speed of the device is potentially limited only by the propagation delays of the logic gates used.

However, asynchronous logic is more difficult to design and is subject to problems not encountered in synchronous designs. The main problem is that digital memory elements are sensitive to the order that their input signals arrive; if two signals arrive at a flip-flop or latch at almost the same time, which state the circuit goes into can depend on which signal gets to the gate first. Therefore, the circuit can go into the wrong state, depending on small differences in the propagation delays of the logic gates. This is called a race condition. This problem is not as severe in synchronous circuits because the outputs of the memory elements only change at each clock pulse. The interval between clock signals is designed to be long enough to allow the outputs of the memory elements to "settle" so they are not changing when the next clock comes. Therefore, the only timing problems are due to "asynchronous inputs"; inputs to the circuit from other systems which are not synchronized to the clock signal.

Asynchronous sequential circuits are typically used only in a few critical parts of otherwise synchronous systems where speed is at a premium, such as parts of microprocessors and digital signal processing circuits.

The design of asynchronous logic uses different mathematical models and techniques from synchronous logic, and is an active area of research.

See also

- Combinational logic

- Synchronous circuit
- Asynchronous circuit
- Logic design
- Application-specific integrated circuit

References

1. Vai, M. Michael (2000). *VLSI Design* (<https://books.google.com/books?id=x9VicwHqocYC&q=%22sequential+logic%22+history+inputs+state&pg=PA147>). CRC Press. p. 147. ISBN 0-84931876-9.
2. Cavanagh, Joseph (2006). *Sequential Logic: Analysis and Synthesis* (<https://books.google.com/books?id=ryz8UbjBfWAC&q=history+%22sequential+logic&pg=PR11>). CRC Press. p. ix. ISBN 0-84937564-9.
3. Lipiansky, Ed (2012). *Electrical, Electronics, and Digital Hardware Essentials for Scientists and Engineers* (<https://books.google.com/books?id=oe31X8WoZnAC&q=%22sequential+logic%22+history+inputs&pg=SA8-PA39>). Wiley. p. 8.39. ISBN 978-1-11841454-5.
4. Dally, William James; Harting, R. Curtis (2012). *Digital Design: A Systems Approach* (<https://books.google.com/books?id=WLoHOG0MhbIC&q=%22sequential+logic%22+history+inputs&pg=PA291>). Cambridge University Press. p. 291. ISBN 978-0-52119950-6.

Further reading

- Katz, Randy; Borriello, Gaetano (2005). *Contemporary Logic Design* (2 ed.). Prentice Hall. ISBN 0-201-30857-6.
- Kohavi, Zvi; Jha, Niraj K. (2009). *Switching and Finite Automata Theory* (3 ed.). Cambridge University Press. ISBN 978-0-521-85748-2.
- Vasyukevich, Vadim O. (2009). "Asynchronous logic elements. Venjunction and sequentiation" (<http://asynlog.balticom.lv/Content/Files/en.pdf>) (PDF). Archived (<https://web.archive.org/web/20110722160840/http://asynlog.balticom.lv/Content/Files/en.pdf>) (PDF) from the original on 2011-07-22. (118 pages)
- Vasyukevich, Vadim O. (2011). Written at Riga, Latvia. *Asynchronous Operators of Sequential Logic: Venjunction & Sequentiation — Digital Circuits Analysis and Design*. Lecture Notes in Electrical Engineering (LNEE). Vol. 101 (1 ed.). Berlin / Heidelberg, Germany: Springer-Verlag. doi:[10.1007/978-3-642-21611-4](https://doi.org/10.1007/978-3-642-21611-4) (<https://doi.org/10.1007%2F978-3-642-21611-4>). ISBN 978-3-642-21610-7. ISSN 1876-1100 (<https://search.worldcat.org/issn/1876-1100>). LCCN 2011929655 (<https://lccn.loc.gov/2011929655>). (xiii+1+123+7 pages) (NB. The back cover of this book erroneously states volume 4, whereas it actually is volume 101.)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Sequential_logic&oldid=1280162547"



Logic gate

A **logic gate** is a device that performs a Boolean function, a logical operation performed on one or more binary inputs that produces a single binary output. Depending on the context, the term may refer to an **ideal logic gate**, one that has, for instance, zero rise time and unlimited fan-out, or it may refer to a non-ideal physical device^[1] (see ideal and real op-amps for comparison).

The primary way of building logic gates uses diodes or transistors acting as electronic switches. Today, most logic gates are made from MOSFETs (metal–oxide–semiconductor field-effect transistors).^[2] They can also be constructed using vacuum tubes, electromagnetic relays with relay logic, fluidic logic, pneumatic logic, optics, molecules, acoustics,^[3] or even mechanical or thermal^[4] elements.

Logic gates can be cascaded in the same way that Boolean functions can be composed, allowing the construction of a physical model of all of Boolean logic, and therefore, all of the algorithms and mathematics that can be described with Boolean logic. **Logic circuits** include such devices as multiplexers, registers, arithmetic logic units (ALUs), and computer memory, all the way up through complete microprocessors,^[5] which may contain more than 100 million logic gates.

Compound logic gates AND-OR-Invert (AOI) and OR-AND-Invert (OAI) are often employed in circuit design because their construction using MOSFETs is simpler and more efficient than the sum of the individual gates.^[6]

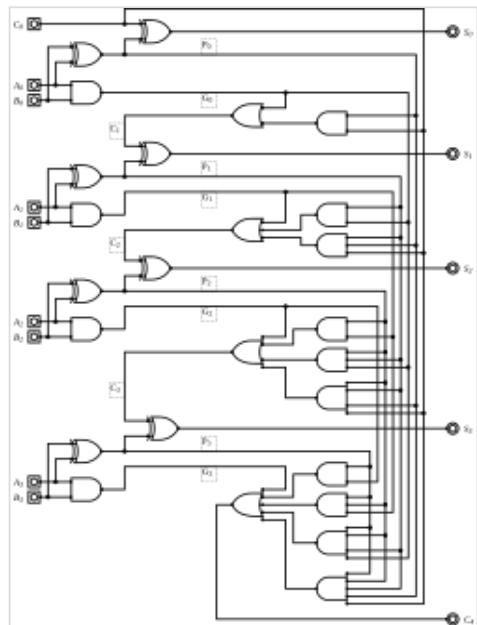
History and development

The binary number system was refined by Gottfried Wilhelm Leibniz (published in 1705), influenced by the ancient I Ching's binary system.^{[7][8]} Leibniz established that using the binary system combined the principles of arithmetic and logic.

The analytical engine devised by Charles Babbage in 1837 used mechanical logic gates based on gears.^[9]

In an 1886 letter, Charles Sanders Peirce described how logical operations could be carried out by electrical switching circuits.^[10] Early Electromechanical computers were constructed from switches and relay logic rather than the later innovations of vacuum tubes (thermionic valves) or transistors (from which later electronic computers were constructed). Ludwig Wittgenstein introduced a version of the 16-row truth table as proposition 5.101 of Tractatus Logico-Philosophicus (1921). Walther Bothe, inventor of the coincidence circuit,^[11] got part of the 1954 Nobel Prize in physics, for the first modern electronic AND gate in 1924. Konrad Zuse designed and built electromechanical logic gates for his computer Z1 (from 1935 to 1938).

From 1934 to 1936, NEC engineer Akira Nakashima, Claude Shannon and Victor Shestakov introduced switching circuit theory in a series of papers showing that two-valued Boolean algebra, which they discovered independently, can describe the operation of switching circuits.^{[12][13][14][15]} Using this property of electrical



A logic circuit diagram for a 4-bit carry lookahead binary adder design using only the AND, OR, and XOR logic gates.

switches to implement logic is the fundamental concept that underlies all electronic digital computers. Switching circuit theory became the foundation of digital circuit design, as it became widely known in the electrical engineering community during and after World War II, with theoretical rigor superseding the *ad hoc* methods that had prevailed previously.^[15]

In 1948, Bardeen and Brattain patented an insulated-gate transistor (IGFET) with an inversion layer. Their concept forms the basis of CMOS technology today.^[16] In 1957 Frosch and Derick were able to manufacture PMOS and NMOS planar gates.^[17] Later a team at Bell Labs demonstrated a working MOS with PMOS and NMOS gates.^[18] Both types were later combined and adapted into complementary MOS (CMOS) logic by Chih-Tang Sah and Frank Wanlass at Fairchild Semiconductor in 1963.^[19]

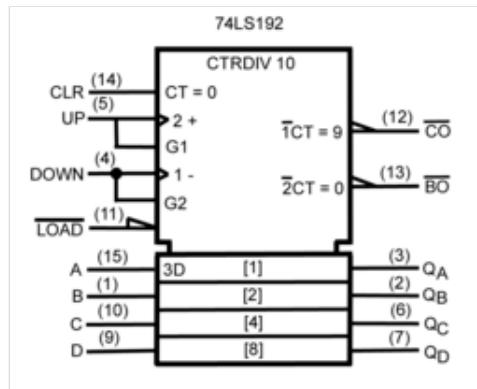
Symbols

There are two sets of symbols for elementary logic gates in common use, both defined in ANSI/IEEE Std 91-1984 and its supplement ANSI/IEEE Std 91a-1991. The "distinctive shape" set, based on traditional schematics, is used for simple drawings and derives from United States Military Standard MIL-STD-806 of the 1950s and 1960s.^[20] It is sometimes unofficially described as "military", reflecting its origin. The "rectangular shape" set, based on ANSI Y32.14 and other early industry standards as later refined by IEEE and IEC, has rectangular outlines for all types of gate and allows representation of a much wider range of devices than is possible with the traditional symbols.^[21] The IEC standard, IEC 60617-12, has been adopted by other standards, such as EN 60617-12:1999 in Europe, BS EN 60617-12:1999 in the United Kingdom, and DIN EN 60617-12:1998 in Germany.

The mutual goal of IEEE Std 91-1984 and IEC 617-12 was to provide a uniform method of describing the complex logic functions of digital circuits with schematic symbols. These functions were more complex than simple AND and OR gates. They could be medium-scale circuits such as a 4-bit counter to a large-scale circuit such as a microprocessor.

IEC 617-12 and its renumbered successor IEC 60617-12 do not explicitly show the "distinctive shape" symbols, but do not prohibit them.^[21] These are, however, shown in ANSI/IEEE Std 91 (and 91a) with this note: "The distinctive-shape symbol is, according to IEC Publication 617, Part 12, not preferred, but is not considered to be in contradiction to that standard." IEC 60617-12 correspondingly contains the note (Section 2.1) "Although non-preferred, the use of other symbols recognized by official national standards, that is distinctive shapes in place of symbols [list of basic gates], shall not be considered to be in contradiction with this standard. Usage of these other symbols in combination to form complex symbols (for example, use as embedded symbols) is discouraged." This compromise was reached between the respective IEEE and IEC working groups to permit the IEEE and IEC standards to be in mutual compliance with one another.

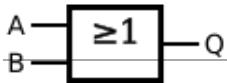
In the 1980s, schematics were the predominant method to design both circuit boards and custom ICs known as gate arrays. Today custom ICs and the field-programmable gate array are typically designed with Hardware Description Languages (HDL) such as Verilog or VHDL.

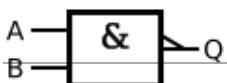


A synchronous 4-bit up/down decade counter symbol (74LS192) in accordance with ANSI/IEEE Std. 91-1984 and IEC Publication 60617-12.

Type	Distinctive shape (IEEE Std 91/91a-1991)	Rectangular shape (IEEE Std 91/91a-1991) (IEC 60617-12:1997)	Boolean algebra between A and B	Truth table								
Single-input gates												
<u>Buffer</u>			A	<table border="1"> <thead> <tr> <th>Input</th><th>Output</th></tr> </thead> <tbody> <tr> <td>A</td><td>Q</td></tr> <tr> <td>0</td><td>0</td></tr> <tr> <td>1</td><td>1</td></tr> </tbody> </table>	Input	Output	A	Q	0	0	1	1
Input	Output											
A	Q											
0	0											
1	1											
<u>NOT</u> (inverter)			\bar{A} or $\neg A$	<table border="1"> <thead> <tr> <th>Input</th><th>Output</th></tr> </thead> <tbody> <tr> <td>A</td><td>Q</td></tr> <tr> <td>0</td><td>1</td></tr> <tr> <td>1</td><td>0</td></tr> </tbody> </table>	Input	Output	A	Q	0	1	1	0
Input	Output											
A	Q											
0	1											
1	0											

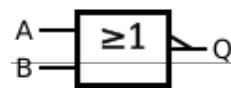
In electronics a NOT gate is more commonly called an inverter. The circle on the symbol is called a *bubble* and is used in logic diagrams to indicate a logic negation between the external logic state and the internal logic state (1 to 0 or vice versa). On a circuit diagram it must be accompanied by a statement asserting that the *positive logic convention* or *negative logic convention* is being used (high voltage level = 1 or low voltage level = 1, respectively). The *wedge* is used in circuit diagrams to directly indicate an active-low (low voltage level = 1) input or output without requiring a uniform convention throughout the circuit diagram. This is called *Direct Polarity Indication*. See IEEE Std 91/91A and IEC 60617-12. Both the *bubble* and the *wedge* can be used on distinctive-shape and rectangular-shape symbols on circuit diagrams, depending on the logic convention used. On pure logic diagrams, only the *bubble* is meaningful.

Conjunction and disjunction																
<u>AND</u>			$A \cdot B$ or $A \wedge B$	<table border="1"> <thead> <tr> <th>Input</th><th>Output</th></tr> </thead> <tbody> <tr> <td>A B</td><td>Q</td></tr> <tr> <td>0 0</td><td>0</td></tr> <tr> <td>0 1</td><td>0</td></tr> <tr> <td>1 0</td><td>0</td></tr> <tr> <td>1 1</td><td>1</td></tr> </tbody> </table>	Input	Output	A B	Q	0 0	0	0 1	0	1 0	0	1 1	1
Input	Output															
A B	Q															
0 0	0															
0 1	0															
1 0	0															
1 1	1															
<u>OR</u>			$A + B$ or $A \vee B$	<table border="1"> <thead> <tr> <th>Input</th><th>Output</th></tr> </thead> <tbody> <tr> <td>A B</td><td>Q</td></tr> <tr> <td>0 0</td><td>0</td></tr> <tr> <td>0 1</td><td>1</td></tr> <tr> <td>1 0</td><td>1</td></tr> <tr> <td>1 1</td><td>1</td></tr> </tbody> </table>	Input	Output	A B	Q	0 0	0	0 1	1	1 0	1	1 1	1
Input	Output															
A B	Q															
0 0	0															
0 1	1															
1 0	1															
1 1	1															

Alternative denial and joint denial				
<u>NAND</u>			$A \cdot \bar{B}$ or $A \uparrow B$	

Input	Output	
A	B	Q
0	0	1
0	1	1
1	0	1
1	1	0

NOR

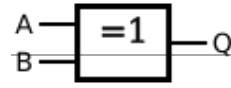


$$\overline{A + B} \text{ or } A \downarrow B$$

Input	Output	
A	B	Q
0	0	1
0	1	0
1	0	0
1	1	0

Exclusive or and biconditional

XOR

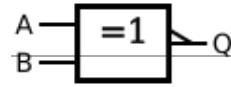


$$A \oplus B \text{ or } A \vee B$$

Input	Output	
A	B	Q
0	0	0
0	1	1
1	0	1
1	1	0

The output of a two input exclusive-OR is true only when the two input values are *different*, and false if they are equal, regardless of the value. If there are more than two inputs, the output of the distinctive-shape symbol is undefined. The output of the rectangular-shaped symbol is true if the number of true inputs is exactly one or exactly the number following the "=" in the qualifying symbol.

XNOR



$$A \oplus B \text{ or } A \odot B$$

Input	Output	
A	B	Q
0	0	1
0	1	0
1	0	0
1	1	1

Implication and Nonimplication

IMPLY^{[22][23]}

$$\overline{A + B} \text{ or } A \rightarrow B$$

Input	Output	
A	B	Q
0	0	1
0	1	1
1	0	0
1	1	1

Input	Output	
A	B	Q
0	0	0
0	1	0
1	0	1
1	1	0

NIMPLY

$$A \cdot \bar{B} \text{ or } A \rightarrow \bar{B}$$

IMPLY and NIMPLY are not commutative, meaning that changing the order of the operands may change the result. For instance, $A \rightarrow \bar{B}$ is false, but $\bar{A} \rightarrow B$ is true; likewise, $A \rightarrow \bar{B}$ is true, but $\bar{A} \rightarrow B$ is false.

De Morgan equivalent symbols

By use of De Morgan's laws, an *AND* function is identical to an *OR* function with negated inputs and outputs. Likewise, an *OR* function is identical to an *AND* function with negated inputs and outputs. A NAND gate is equivalent to an OR gate with negated inputs, and a NOR gate is equivalent to an AND gate with negated inputs.

This leads to an alternative set of symbols for basic gates that use the opposite core symbol (*AND* or *OR*) but with the inputs and outputs negated. Use of these alternative symbols can make logic circuit diagrams much clearer and help to show accidental connection of an active high output to an active low input or vice versa. Any connection that has logic negations at both ends can be replaced by a negationless connection and a suitable change of gate or vice versa. Any connection that has a negation at one end and no negation at the other can be made easier to interpret by instead using the De Morgan equivalent symbol at either of the two ends. When negation or polarity indicators on both ends of a connection match, there is no logic negation in that path (effectively, bubbles "cancel"), making it easier to follow logic states from one symbol to the next. This is commonly seen in real logic diagrams – thus the reader must not get into the habit of associating the shapes exclusively as OR or AND shapes, but also take into account the bubbles at both inputs and outputs in order to determine the "true" logic function indicated.

A De Morgan symbol can show more clearly a gate's primary logical purpose and the polarity of its nodes that are considered in the "signaled" (active, on) state. Consider the simplified case where a two-input NAND gate is used to drive a motor when either of its inputs are brought low by a switch. The "signaled" state (motor on) occurs when either one OR the other switch is on. Unlike a regular NAND symbol, which suggests AND logic, the De Morgan version, a two negative-input OR gate, correctly shows that OR is of interest. The regular NAND symbol has a bubble at the output and none at the inputs (the opposite of the states that will turn the motor on), but the De Morgan symbol shows both inputs and output in the polarity that will drive the motor.

De Morgan's theorem is most commonly used to implement logic gates as combinations of only NAND gates, or as combinations of only NOR gates, for economic reasons.

Truth tables

Output comparison of various logic gates:

1-input logic gates

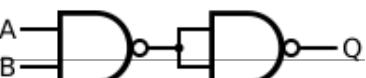
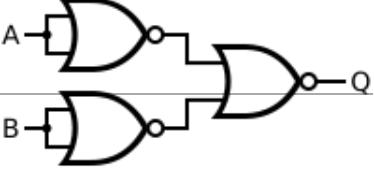
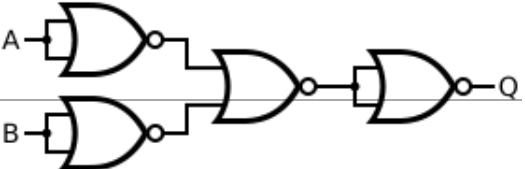
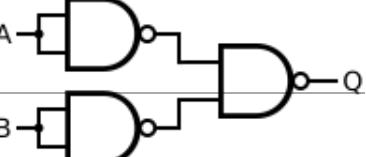
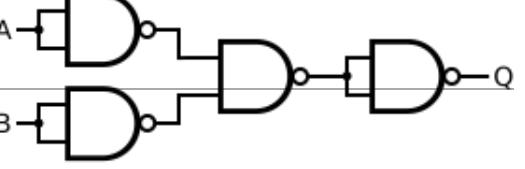
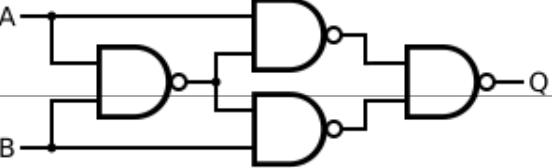
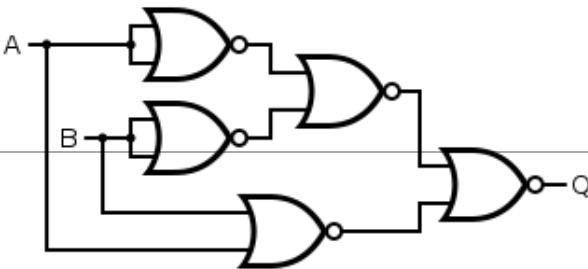
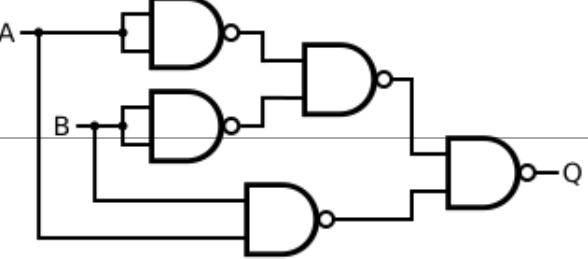
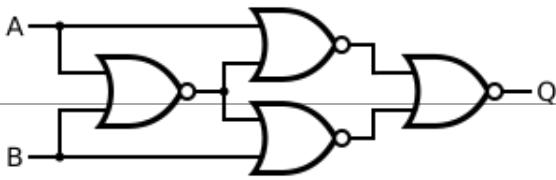
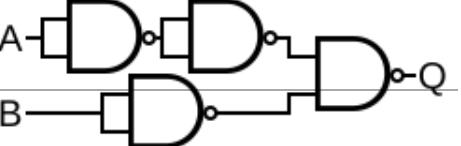
Input	Output	
A	Buffer	Inverter
0	0	1
1	1	0

2-input logic gates

Input		Output								
A	B	AND	NAND	OR	NOR	XOR	XNOR	IMPLY	NIMPLY	
0	0	0	1	0	1	0	1	1	0	
0	1	0	1	1	0	1	0	1	0	
1	0	0	1	1	0	1	0	0	1	
1	1	1	0	1	0	0	1	1	0	

Universal logic gates

Charles Sanders Peirce (during 1880–1881) showed that NOR gates alone (or alternatively NAND gates alone) can be used to reproduce the functions of all the other logic gates, but his work on it was unpublished until 1933.^[24] The first published proof was by Henry M. Sheffer in 1913, so the NAND logical operation is sometimes called *Sheffer stroke*; the logical NOR is sometimes called *Peirce's arrow*.^[25] Consequently, these gates are sometimes called *universal logic gates*.^[26]

type	NAND construction	NOR construction
NOT	 A $\overline{\text{NAND}} \rightarrow Q$	 A $\overline{\text{NOR}} \rightarrow Q$
AND	 A $\text{AND} \rightarrow Q$ B	 A $\overline{\text{NOR}} \rightarrow \overline{Q}$ B $\overline{\text{NOR}} \rightarrow \overline{Q}$
NAND	 A $\overline{\text{NAND}} \rightarrow Q$ B	 A $\overline{\text{NOR}} \rightarrow \overline{Q}$ B $\overline{\text{NOR}} \rightarrow \overline{Q}$
OR	 A $\text{OR} \rightarrow Q$ B	 A $\overline{\text{NOR}} \rightarrow \overline{Q}$ B $\overline{\text{NOR}} \rightarrow \overline{Q}$
NOR	 A $\overline{\text{NOR}} \rightarrow Q$ B	 A $\overline{\text{NOR}} \rightarrow Q$ B
XOR	 A $\text{XOR} \rightarrow Q$ B	 A $\overline{\text{NOR}} \rightarrow \overline{Q}$ B $\overline{\text{NOR}} \rightarrow \overline{Q}$
XNOR	 A $\text{XNOR} \rightarrow Q$ B	 A $\overline{\text{NOR}} \rightarrow \overline{Q}$ B $\overline{\text{NOR}} \rightarrow \overline{Q}$
IMPLY	 A $\text{IMPLY} \rightarrow Q$ B	 A $\overline{\text{NOR}} \rightarrow \overline{Q}$ B $\overline{\text{NOR}} \rightarrow \overline{Q}$
NIMPLY	 A $\text{NIMPLY} \rightarrow Q$ B	

Data storage and sequential logic

Logic gates can also be used to hold a state, allowing data storage. A storage element can be constructed by connecting several gates in a "latch" circuit. Latching circuitry is used in static random-access memory. More complicated designs that use clock signals and that change only on a rising or falling edge of the clock are called edge-triggered "flip-flops". Formally, a flip-flop is called a bistable circuit, because it has two stable states which it can maintain indefinitely. The combination of multiple flip-flops in parallel, to store a multiple-bit value, is known as a register. When using any of these gate setups the overall system has memory; it is then called a sequential logic system since its output can be influenced by its previous state(s), i.e. by the sequence of input states. In contrast, the output from combinational logic is purely a combination of its present inputs, unaffected by the previous input and output states.

These logic circuits are used in computer memory. They vary in performance, based on factors of speed, complexity, and reliability of storage, and many different types of designs are used based on the application.

Manufacturing

Electronic gates

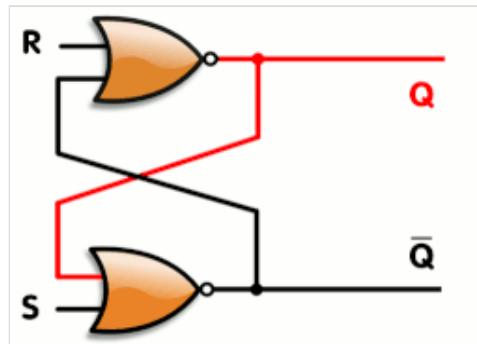
A functionally complete logic system may be composed of relays, valves (vacuum tubes), or transistors.

Electronic logic gates differ significantly from their relay-and-switch equivalents. They are much faster, consume much less power, and are much smaller (all by a factor of a million or more in most cases). Also, there is a fundamental structural difference. The switch circuit creates a continuous metallic path for current to flow (in either direction) between its input and its output. The semiconductor logic gate, on the other hand, acts as a high-gain voltage amplifier, which sinks a tiny current at its input and produces a low-impedance voltage at its output. It is not possible for current to flow between the output and the input of a semiconductor logic gate.

For small-scale logic, designers now use prefabricated logic gates from families of devices such as the TTL 7400 series by Texas Instruments, the CMOS 4000 series by RCA, and their more recent descendants. Increasingly, these fixed-function logic gates are being replaced by programmable logic devices, which allow designers to pack many mixed logic gates into a single integrated circuit. The field-programmable nature of programmable logic devices such as FPGAs has reduced the 'hard' property of hardware; it is now possible to change the logic design of a hardware system by reprogramming some of its components, thus allowing the features or function of a hardware implementation of a logic system to be changed.

An important advantage of standardized integrated circuit logic families, such as the 7400 and 4000 families, is that they can be cascaded. This means that the output of one gate can be wired to the inputs of one or several other gates, and so on. Systems with varying degrees of complexity can be built without great concern of the designer for the internal workings of the gates, provided the limitations of each integrated circuit are considered.

The output of one gate can only drive a finite number of inputs to other gates, a number called the 'fan-out limit'. Also, there is always a delay, called the 'propagation delay', from a change in input of a gate to the corresponding change in its output. When gates are cascaded, the total propagation delay is approximately the sum of the



Animation of how an SR NOR gate latch works.

individual delays, an effect which can become a problem in high-speed synchronous circuits. Additional delay can be caused when many inputs are connected to an output, due to the distributed capacitance of all the inputs and wiring and the finite amount of current that each output can provide.

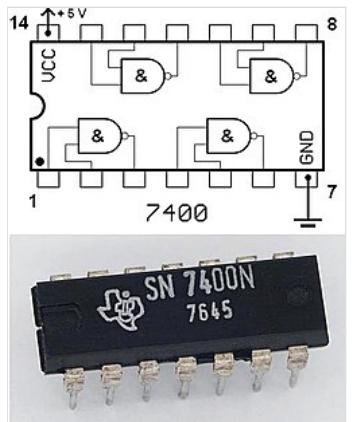
Logic families

There are several logic families with different characteristics (power consumption, speed, cost, size) such as: RDL (resistor-diode logic), RTL (resistor-transistor logic), DTL (diode-transistor logic), TTL (transistor-transistor logic) and CMOS. There are also sub-variants, e.g. standard CMOS logic vs. advanced types using still CMOS technology, but with some optimizations for avoiding loss of speed due to slower PMOS transistors.

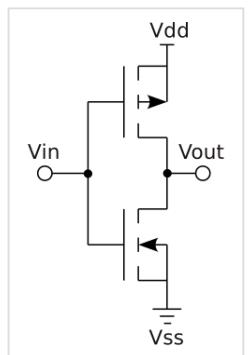
The simplest family of logic gates uses bipolar transistors, and is called resistor-transistor logic (RTL). Unlike simple diode logic gates (which do not have a gain element), RTL gates can be cascaded indefinitely to produce more complex logic functions. RTL gates were used in early integrated circuits. For higher speed and better density, the resistors used in RTL were replaced by diodes resulting in diode-transistor logic (DTL). Transistor-transistor logic (TTL) then supplanted DTL.

As integrated circuits became more complex, bipolar transistors were replaced with smaller field-effect transistors (MOSFETs); see PMOS and NMOS. To reduce power consumption still further, most contemporary chip implementations of digital systems now use CMOS logic. CMOS uses complementary (both n-channel and p-channel) MOSFET devices to achieve a high speed with low power dissipation.

Other types of logic gates include, but are not limited to:^[27]



The 7400 chip, containing four NANDs. The two additional pins supply power (+5 V) and connect the ground.

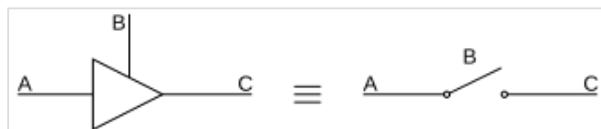


CMOS diagram of a NOT gate, also known as an inverter. MOSFETs are the most common way to make logic gates.

Logic family	Abbreviation	Description
Diode logic	DL	
Tunnel diode logic	TDL	Exactly the same as diode logic but can perform at a higher speed.
Neon logic	NL	Uses neon bulbs or 3-element neon trigger tubes to perform logic.
Core diode logic	CDL	Performed by semiconductor diodes and small ferrite toroidal cores for moderate speed and moderate power level.
4Layer Device Logic	4LDL	Uses thyristors and SCRs to perform logic operations where high current and or high voltages are required.
Direct-coupled transistor logic	DCTL	Uses transistors switching between saturated and cutoff states to perform logic. The transistors require carefully controlled parameters. Economical because few other components are needed, but tends to be susceptible to noise because of the lower voltage levels employed. Often considered to be the father to modern TTL logic.
Metal–oxide–semiconductor logic	MOS	Uses MOSFETs (metal–oxide–semiconductor field-effect transistors), the basis for most modern logic gates. The MOS logic family includes PMOS logic, NMOS logic, complementary MOS (CMOS), and BiCMOS (bipolar CMOS).
Current-mode logic	CML	Uses transistors to perform logic but biasing is from constant current sources to prevent saturation and allow extremely fast switching. Has high noise immunity despite fairly low logic levels.
Quantum-dot cellular automata	QCA	Uses tunnelable q-bits for synthesizing the binary logic bits. The electrostatic repulsive force in between two electrons in the quantum dots assigns the electron configurations (that defines state 1 or state 0) under the suitably driven polarizations. This is a transistorless, currentless, junctionless binary logic synthesis technique allowing it to have very fast operation speeds.
Ferroelectric FET	FeFET	FeFET transistors can retain their state to speed recovery in case of a power loss. ^[28]

Three-state logic gates

A three-state logic gate is a type of logic gate that can have three different outputs: high (H), low (L) and high-impedance (Z). The high-impedance state plays no role in the logic, which is strictly binary. These devices are used on buses of the CPU to allow multiple chips to send data. A group of three-state outputs driving a line with a suitable control circuit is basically equivalent to a multiplexer, which may be physically distributed over separate devices or plug-in cards.



A three-state buffer can be thought of as a switch. If B is on, the switch is closed. If B is off, the switch is open.

In electronics, a high output would mean the output is sourcing current from the positive power terminal (positive voltage). A low output would mean the output is sinking current to the negative power terminal (zero voltage). High impedance would mean that the output is effectively disconnected from the circuit.

Non-electronic logic gates

Non-electronic implementations are varied, though few of them are used in practical applications. Many early electromechanical digital computers, such as the Harvard Mark I, were built from relay logic gates, using electro-mechanical relays. Logic gates can be made using pneumatic devices, such as the Sorteberg relay or mechanical logic gates, including on a molecular scale.^[29] Various types of fundamental logic gates have been constructed using molecules (molecular logic gates), which are based on chemical inputs and spectroscopic outputs.^[30] Logic

gates have been made out of DNA (see DNA nanotechnology)^[31] and used to create a computer called MAYA (see MAYA-II). Logic gates can be made from quantum mechanical effects, see quantum logic gate. Photonic logic gates use nonlinear optical effects.

In principle any method that leads to a gate that is functionally complete (for example, either a NOR or a NAND gate) can be used to make any kind of digital logic circuit. Note that the use of 3-state logic for bus systems is not needed, and can be replaced by digital multiplexers, which can be built using only simple logic gates (such as NAND gates, NOR gates, or AND and OR gates).

See also

- [And-inverter graph](#)
- [Boolean algebra topics](#)
- [Boolean function](#)
- [Depletion-load NMOS logic](#)
- [Digital circuit](#)
- [Electronic symbol](#)
- [Espresso heuristic logic minimizer](#)
- [Emitter-coupled logic](#)
- [Fan-out](#)
- [Field-programmable gate array \(FPGA\)](#)
- [Flip-flop \(electronics\)](#)
- [Functional completeness](#)
- [Integrated injection logic](#)
- [Karnaugh map](#)
- [Combinational logic](#)
- [List of 4000 series integrated circuits](#)
- [List of 7400 series integrated circuits](#)
- [Logic family](#)
- [Logic level](#)
- [Logical graph](#)
- [Magnetic logic](#)
- [NMOS logic](#)
- [Parametron](#)
- [Processor design](#)
- [Programmable logic controller \(PLC\)](#)
- [Programmable logic device \(PLD\)](#)
- [Propositional calculus](#)
- [Race hazard](#)
- [Reversible computing](#)
- [Superconducting computing](#)
- [Truth table](#)
- [Unconventional computing](#)

References

1. Jaeger (1997). *Microelectronic Circuit Design*. McGraw-Hill. pp. 226–233. ISBN 0-07-032482-4.
2. Kanellos, Michael (2003-02-11). "Moore's Law to roll on for another decade" (<https://www.cnet.com/news/moores-law-to-roll-on-for-another-decade/>). CNET. From *Integrated circuit*
3. Zhang, Ting; Cheng, Ying; Guo, Jian-Zhong; Xu, Jian-yi; Liu, Xiao-jun (2015), "Acoustic logic gates and Boolean operation based on self-collimating acoustic beams" (<https://pubs.aip.org/aip/apl/article-abstract/106/11/113503/27163/Acoustic-logic-gates-and-Boolean-operation-based?redirectedFrom=fulltext>), *Applied Physics Letters*, **106** (11), Bibcode:2015ApPhL.106k3503Z (<https://ui.adsabs.harvard.edu/abs/2015ApPhL.106k3503Z>), doi:10.1063/1.4915338 (<https://doi.org/10.1063%2F1.4915338>), retrieved 2024-08-17
4. Wang, Lei; Li, Baowen (2007). "Thermal Logic Gates: Computation with Phonons" (<https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.99.177208>). *Physical Review Letters*. **99** (17): 177208. arXiv:0709.0032 (<https://arxiv.org/abs/0709.0032>). Bibcode:2007PhRvL..99q7208W (<https://ui.adsabs.harvard.edu/abs/2007PhRvL..99q7208W>). doi:10.1103/PhysRevLett.99.177208 (<https://doi.org/10.1103%2FPhysRevLett.99.177208>). PMID 17995368 (<https://pubmed.ncbi.nlm.nih.gov/17995368>). S2CID 10934270 (<https://api.semanticscholar.org/CorpusID:10934270>).
5. Deschamps, Jean-Pierre; Valderrama, Elena; Terés, Lluís (2016-10-12). *Digital Systems: From Logic Gates to Processors* (<https://books.google.com/books?id=GBVADQAAQBAJ&q=processor+logic+gates>). Springer. ISBN 978-3-319-41198-9.
6. Tinder, Richard F. (2000). *Engineering digital design* (<https://books.google.com/books?id=6x0pjMKRh0C&q=AOI+gate&pg=PT347>) (2nd ed.). Academic Press. pp. 317–319. ISBN 0-12-691295-5.

7. Nylan, Michael (2001). *The Five "Confucian" Classics* (<https://books.google.com/books?id=KykM1DhBxd8C&pg=PA206>). Yale University Press. pp. 204–206. ISBN 978-0-300-08185-5. Retrieved 2010-06-08.
8. Perkins, Franklin (2004). "Exchange with China" (<https://books.google.com/books?id=0Jzv9loAHFsC&dq=117&pg=PA117>). *Leibniz and China: A Commerce of Light*. Cambridge University Press. p. 117. ISBN 9780521830249. "... one of the traditional orderings of the hexagrams, the *xiantian tu* ordering made by Shao Yong, was, with a few modifications, the same order found in Leibniz's binary arithmetic."
9. Julio Sanchez; Maria P. Canton (2017-12-19). *Embedded Systems Circuits and Programming* (<http://books.google.com/books?id=FCjOBgAAQBAJ&dq=Babbage+Logic+Gate&pg=PA17>). CRC Press. p. 17. ISBN 978-1-4398-7931-3.
10. Peirce, C. S., "Letter, Peirce to A. Marquand", dated 1886, *Writings of Charles S. Peirce*, v. 5, 1993, pp. 420–423. See Burks, Arthur W. (1978). "Review: Charles S. Peirce, *The new elements of mathematics*" (http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?view=body&id=pdf_1&handle=euclid.bams/1183541145). *Bulletin of the American Mathematical Society*. **84** (5): 913–918 [917]. doi:10.1090/S0002-9904-1978-14533-9 (<https://doi.org/10.1090%2FS0002-9904-1978-14533-9>).
11. Luisa Bonolis; Walther Bothe and Bruno Rossi: The birth and development of coincidence methods in cosmic-ray physics. *Am. J. Phys.* 1 November 2011; 79 (11): 1133–1150.
12. Yamada, Akihiko (2004). "History of Research on Switching Theory in Japan" (https://www.jstage.jst.go.jp/article/ieejfms/124/8/124_8_720/_article). *IEEJ Transactions on Fundamentals and Materials*. **124** (8). Institute of Electrical Engineers of Japan: 720–726. Bibcode:2004IJTFM.124..720Y (<https://ui.adsabs.harvard.edu/abs/2004IJTFM.124..720Y>). doi:10.1541/ieejfms.124.720 (<https://doi.org/10.1541%2Fieejfms.124.720>).
13. "Switching Theory/Relay Circuit Network Theory/Theory of Logical Mathematics" (<http://museum.ipsj.or.jp/en/computer/dawn/0002.html>). *IPSJ Computer Museum*. Information Processing Society of Japan.
14. Stanković, Radomir S.; Astola, Jaakko T.; Karpovsky, Mark G. (2007). *Some Historical Remarks on Switching Theory*. CiteSeerX 10.1.1.66.1248 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.66.1248>).
15. Stanković, Radomir S. [in German]; Astola, Jaakko Tapio [in Finnish], eds. (2008). *Reprints from the Early Days of Information Sciences: TICSP Series On the Contributions of Akira Nakashima to Switching Theory* (<https://web.archive.org/web/20210308002559/http://ticsp.cs.tut.fi/reports/reprint-nakashima-rr.pdf>) (PDF). Tampere International Center for Signal Processing (TICSP) Series. Vol. 40. Tampere University of Technology, Tampere, Finland. ISBN 978-952-15-1980-2. ISSN 1456-2774 (<https://search.worldcat.org/issn/1456-2774>). Archived from the original (<http://ticsp.cs.tut.fi/reports/reprint-nakashima-rr.pdf>) (PDF) on 2021-03-08. (3+207+1 pages) 10:00 min (https://web.archive.org/web/20221026175726/http://ciitlab.elfak.ni.ac.rs/predavanja/09_Nakashima.mp4)
16. Howard R. Duff (2001). "John Bardeen and transistor physics". *AIP Conference Proceedings*. Vol. 550. pp. 3–32. doi:10.1063/1.1354371 (<https://doi.org/10.1063%2F1.1354371>).
17. Frosch, C. J.; Derick, L (1957). "Surface Protection and Selective Masking during Diffusion in Silicon" (<https://iopscience.iop.org/article/10.1149/1.2428650>). *Journal of the Electrochemical Society*. **104** (9): 547. doi:10.1149/1.2428650 (<https://doi.org/10.1149%2F1.2428650>).
18. Lojek, Bo (2007). *History of Semiconductor Engineering*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. p. 321. ISBN 978-3-540-34258-8.
19. "1963: Complementary MOS Circuit Configuration is Invented" (<https://www.computerhistory.org/siliconengine/complementary-mos-circuit-configuration-is-invented/>). *Computer History Museum*. Retrieved 2019-07-06.
20. "Graphical Symbols for Logic Diagrams" (https://quicksearch.dla.mil/qsDocDetails.aspx?ident_number=35975). ASSIST Quick Search. Defense Logistics Agency. MIL-STD-806. Retrieved 2021-08-27.
21. "Overview of IEEE Standard 91-1984 Explanation of Logic Symbols" (<http://www.ti.com/lit/ml/sdyz01a/sdyz001a.pdf>) (PDF). Texas Instruments Semiconductor Group. 1996. SDYZ001A.

22. Jyoti Garg; Aishita Verma; Subodh Wairyaa. "Memristor Emulator Circuits an Emerging Technology with Applications". In Brijesh Mishra; Manish Tiwari (eds.). *VLSI, Microwave and Wireless Technologies*. p. 476.
23. Hanawalt, Barbara. *Cellular Computing* (https://www.google.com/books/edition/Cellular_Computing/iEQj_WFuwFIC?hl=en&gbpv=1&pg=PA52). p. 52.
24. Peirce, C. S. (manuscript winter of 1880–1881), "A Boolean Algebra with One Constant", published 1933 in *Collected Papers* v. 4, paragraphs 12–20. Reprinted 1989 in *Writings of Charles S. Peirce* v. 4, pp. 218–221, Google [1] (<https://archive.org/details/writingsofcharle0004peir/page/218>). See Roberts, Don D. (2009). "7.12 The Graphical Analysis of Propositions" (<https://books.google.com/books?id=Q4K30wCAf-gC&pg=PA113>). *The Existential Graphs of Charles S. Peirce*. De Gruyter. p. 131. ISBN 978-3-11022622-5.
25. Büning, Hans Kleine; Lettmann, Theodor (1999). *Propositional logic: deduction and algorithms* (<http://books.google.com/books?id=3oJE9yczr3EC&pg=PA2>). Cambridge University Press. p. 2. ISBN 978-0-521-63017-7.
26. Bird, John (2007). *Engineering mathematics* (<https://books.google.com/books?id=1-fBmsEBNUoC&pg=PA532>). Newnes. p. 532. ISBN 978-0-7506-8555-9.
27. Rowe, Jim. "Circuit Logic – Why and How". No. December 1966. Electronics Australia.
28. "Tapping into Non-Volatile Logic" (<https://semiengineering.com/tapping-into-non-volatile-logic/>). 2021-04-21.
29. Merkle, Ralph C. (1993). "Two Types of Mechanical Reversible Logic" (<http://www.zyvex.com/nanotech/mechano.html>). Xerox PARC.
30. Erbas-Cakmak, Sundus; Kolemen, Safacan; Sedgwick, Adam C.; Gunnlaugsson, Thorfinnur; James, Tony D.; Yoon, Juyoung; Akkaya, Engin U. (2018). "Molecular logic gates: the past, present and future" (<http://xlink.rsc.org/?DOI=C7CS00491E>). *Chemical Society Reviews*. **47** (7): 2228–2248. doi:10.1039/C7CS00491E (<https://doi.org/10.1039%2FC7CS00491E>). hdl:11693/50034 (<http://hdl.handle.net/11693%2F50034>). ISSN 0306-0012 (<https://search.worldcat.org/issn/0306-0012>). PMID 29493684 (<https://pubmed.ncbi.nlm.nih.gov/29493684>).
31. Stojanovic, Milan N.; Mitchell, Tiffany E.; Stefanovic, Darko (2002). "Deoxyribozyme-Based Logic Gates" (<https://pubs.acs.org/doi/abs/10.1021/ja016756v>). *Journal of the American Chemical Society*. **124** (14): 3555–3561. Bibcode:2002JAChS.124.3555S (<https://ui.adsabs.harvard.edu/abs/2002JAChS.124.3555S>). doi:10.1021/ja016756v (<https://doi.org/10.1021%2Fja016756v>). PMID 11929243 (<https://pubmed.ncbi.nlm.nih.gov/11929243>).

Further reading

- Bostock, Geoff (1988). *Programmable logic devices: technology and applications* (<https://books.google.com/books?id=XEFTAAAAMAAJ>). McGraw-Hill. ISBN 978-0-07-006611-3.
- Brown, Stephen D.; Francis, Robert J.; Rose, Jonathan; Vranesic, Zvonko G. (1992). *Field Programmable Gate Arrays* (<https://books.google.com/books?id=8s4M-qYOWZIC>). Kluwer Academic. ISBN 978-0-7923-9248-4.

External links

-  Media related to Logic gates at Wikimedia Commons

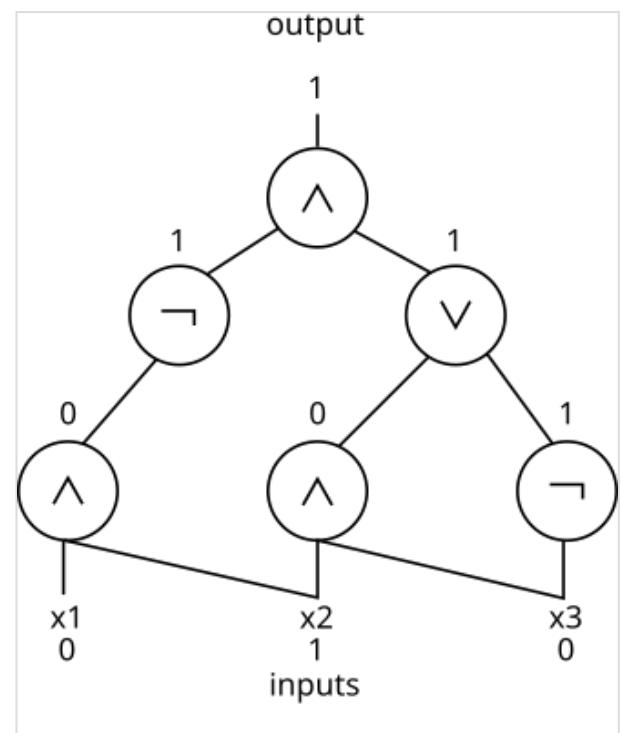


Boolean circuit

In computational complexity theory and circuit complexity, a **Boolean circuit** is a mathematical model for combinational digital logic circuits. A formal language can be decided by a family of Boolean circuits, one circuit for each possible input length.

Boolean circuits are defined in terms of the logic gates they contain. For example, a circuit might contain binary AND and OR gates and unary NOT gates, or be entirely described by binary NAND gates. Each gate corresponds to some Boolean function that takes a fixed number of bits as input and outputs a single bit.

Boolean circuits provide a model for many digital components used in computer engineering, including multiplexers, adders, and arithmetic logic units, but they exclude sequential logic. They are an abstraction that omits many aspects relevant to designing real digital logic circuits, such as metastability, fanout, glitches, power consumption, and propagation delay variability.



Example Boolean circuit. The \wedge nodes are AND gates, the \vee nodes are OR gates, and the \neg nodes are NOT gates

Formal definition

In giving a formal definition of Boolean circuits, Vollmer starts by defining a basis as set B of Boolean functions, corresponding to the gates allowable in the circuit model. A Boolean circuit over a basis B , with n inputs and m outputs, is then defined as a finite directed acyclic graph. Each vertex corresponds to either a basis function or one of the inputs, and there is a set of exactly m nodes which are labeled as the outputs.^{[1]:8} The edges must also have some ordering, to distinguish between different arguments to the same Boolean function.^{[1]:9}

As a special case, a propositional formula or Boolean expression is a Boolean circuit with a single output node in which every other node has fan-out of 1. Thus, a Boolean circuit can be regarded as a generalization that allows shared subformulas and multiple outputs.

A common basis for Boolean circuits is the set {AND, OR, NOT}, which is functionally complete, i. e. from which all other Boolean functions can be constructed.

Computational complexity

Background

A particular circuit acts only on inputs of fixed size. However, formal languages (the string-based representations of decision problems) contain strings of different lengths, so languages cannot be fully captured by a single circuit (in contrast to the Turing machine model, in which a language is fully described by a single Turing machine). A language is instead represented by a *circuit family*. A circuit family is an infinite list of circuits (C_0, C_1, C_2, \dots), where C_n has n input variables. A circuit family is said to decide a language L if, for every string w , w is in the language L if and only if $C_n(w) = 1$, where n is the length of w . In other words, a language is the set of strings which, when applied to the circuits corresponding to their lengths, evaluate to 1.^{[2]:354}

Complexity measures

Several important complexity measures can be defined on Boolean circuits, including circuit depth, circuit size, and the number of alternations between AND gates and OR gates. For example, the size complexity of a Boolean circuit is the number of gates in the circuit.

There is a natural connection between circuit size complexity and time complexity.^{[2]:355} Intuitively, a language with small time complexity (that is, requires relatively few sequential operations on a Turing machine), also has a small circuit complexity (that is, requires relatively few Boolean operations). Formally, it can be shown that if a language is in $\text{TIME}(t(n))$, where t is a function $t : \mathbb{N} \rightarrow \mathbb{N}$, then it has circuit size complexity $O(t^2(n))$.

Complexity classes

Several important complexity classes are defined in terms of Boolean circuits. The most general of these is P/poly, the set of languages that are decidable by polynomial-size circuit families. It follows directly from the fact that languages in $\text{TIME}(t(n))$ have circuit complexity $O(t^2(n))$ that $\text{P} \subseteq \text{P/poly}$. In other words, any problem that can be computed in polynomial time by a deterministic Turing machine can also be computed by a polynomial-size circuit family. It is further the case that the inclusion is proper (i.e. $\text{P} \subsetneq \text{P/poly}$) because there are undecidable problems that are in P/poly. P/poly turns out to have a number of properties that make it highly useful in the study of the relationships between complexity classes. In particular, it is helpful in investigating problems related to P versus NP. For example, if there is any language in NP that is not in P/poly then $\text{P} \neq \text{NP}$.^{[3]:286} P/poly also helps to investigate properties of the polynomial hierarchy. For example, if $\text{NP} \subseteq \text{P/poly}$, then PH collapses to Σ_2^{P} . A full description of the relations between P/poly and other complexity classes is available at "Importance of P/poly". P/poly also has the interesting feature that it can be equivalently defined as the class of languages recognized by a polynomial-time Turing machine with a polynomial-bounded advice function.

Two subclasses of P/poly that have interesting properties in their own right are NC and AC. These classes are defined not only in terms of their circuit size but also in terms of their *depth*. The depth of a circuit is the length of the longest directed path from an input node to the output node. The class NC is the set of languages that can be solved by circuit families that are restricted not only to having polynomial-size but

also to having polylogarithmic depth. The class AC is defined similarly to NC, however gates are allowed to have unbounded fan-in (that is, the AND and OR gates can be applied to more than two bits). NC is an important class because it turns out that it represents the class of languages that have efficient parallel algorithms.

Circuit evaluation

The Circuit Value Problem — the problem of computing the output of a given Boolean circuit on a given input string — is a P-complete decision problem.^{[3]:119} Therefore, this problem is considered to be "inherently sequential" in the sense that there is likely no efficient, highly parallel algorithm that solves the problem.

Completeness

Logic circuits are physical representation of simple logic operations, AND, OR and NOT (and their combinations, such as non-sequential flip-flops or circuit networks), that form a mathematical structure known as Boolean algebra. They are complete in sense that they can perform any deterministic algorithm. However, it just happens that this is not all there is. In the physical world we also encounter randomness, notable in small systems governed by quantization effects, which is described by theory of Quantum Mechanics. Logic circuits cannot produce any randomness, and in that sense they form an incomplete logic set. Remedy to that is found in adding an ad-hoc random bit generator to logic networks, or computers, such as in Probabilistic Turing machine. A recent work^[4] has introduced a theoretical concept of an inherently random logic circuit named *random flip-flop*, which completes the set. It conveniently packs randomness and is inter-operable with deterministic Boolean logic circuits. However, an algebraic structure equivalent of Boolean algebra and associated methods of circuit construction and reduction for the extended set is yet unknown.

See also

- Circuit satisfiability
- Logic gate
- Boolean logic
- Switching lemma

Footnotes

1. Vollmer, Heribert (1999). *Introduction to Circuit Complexity*. Berlin: Springer. ISBN 3-540-64310-9.
2. Sipser, Michael (2006). *Introduction to the Theory of Computation* (2nd ed.). USA: Thomson Course Technology. ISBN 978-0-534-95097-2.
3. Arora, Sanjeev; Barak, Boaz (2009). *Computational Complexity: A Modern Approach*. Cambridge University Press. ISBN 978-0-521-42426-4.
4. Stipčević, Mario; Batelić, Mateja (2022). "Entropy considerations in improved circuits for a biologically-inspired random pulse computer" (<https://doi.org/10.1038%2Fs41598-021-04177-9>). *Scientific Reports*. **12**: 115. arXiv:1908.04779 (<https://arxiv.org/abs/1908.04779>). doi:10.1038/s41598-021-04177-9 (<https://doi.org/10.1038%2Fs41598-021-04177-9>).



Mixed-signal integrated circuit

A **mixed-signal integrated circuit** is any integrated circuit that has both analog circuits and digital circuits on a single semiconductor die.^{[1][2][3][4]} Their usage has grown dramatically with the increased use of cell phones, telecommunications, portable electronics, and automobiles with electronics and digital sensors.

Overview

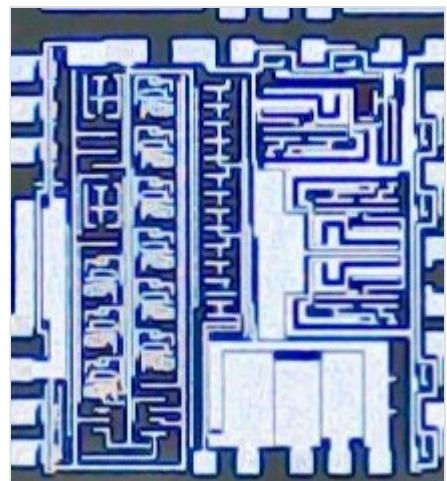
Integrated circuits (ICs) are generally classified as digital (e.g. a microprocessor) or analog (e.g. an operational amplifier). Mixed-signal ICs contain both digital and analog circuitry on the same chip, and sometimes embedded software. Mixed-signal ICs process both analog and digital signals together. For example, an analog-to-digital converter (ADC) is a typical mixed-signal circuit.

Mixed-signal ICs are often used to convert analog signals to digital signals so that digital devices can process them. For example, mixed-signal ICs are essential components for FM tuners in digital products such as media players, which have digital amplifiers. Any analog signal can be digitized using a very basic ADC, and the smallest and most energy efficient of these are mixed-signal ICs.

Mixed-signal ICs are more difficult to design and manufacture than analog-only or digital-only integrated circuits. For example, an efficient mixed-signal IC may have its digital and analog components share a common power supply. However, analog and digital components have very different power needs and consumption characteristics, which makes this a non-trivial goal in chip design.

Mixed-signal functionality involves both traditional active elements (like transistors) and well-performing passive elements (like coils, capacitors, and resistors) on the same chip. This requires additional modelling understanding and options from manufacturing technologies. High voltage transistors might be needed in the power management functions on a chip with digital functionality, possibly with a low-power CMOS processor system. Some advanced mixed-signal technologies may enable combining analog sensor elements (like pressure sensors or imaging diodes) on the same chip with an ADC.

Typically, mixed-signal ICs do not necessarily need the fastest digital performance. Instead, they need more mature models of active and passive elements for more accurate simulations and verification, such as for testability planning and reliability estimations. Therefore, mixed-signal circuits are typically realized with larger line widths than the highest speed and densest digital logic, and the implementation technologies can be two to four generations behind the latest digital-only implementation technologies. Additionally, mixed signal processing may need passive elements like resistors, capacitors, and coils,



Mixed signal integrated circuit: the metal areas on the right-hand side are capacitors, on top of which are large output transistors; the left-hand side is occupied by the digital logic

which may require specialized metal, dielectric layers, or similar adaptations of standard fabrication processes. Because of these specific requirements, mixed-signal ICs and digital ICs can have different manufacturers (known as foundries).

Applications

There are numerous applications of mixed-signal integrated circuits, such as in mobile phones, modern radio and telecommunication systems, sensor systems with on-chip standardized digital interfaces (including I2C, UART, SPI, or CAN), voice-related signal processing, aerospace and space electronics, the Internet of things (IoT), unmanned aerial vehicles (UAVs), and automotive and other electrical vehicles. Mixed-signal circuits or systems are typically cost-effective solutions, such as for building modern consumer electronics and in industrial, medical, measurement, and space applications.

Examples of mixed-signal integrated circuits include data converters using delta-sigma modulation, analog-to-digital converters and digital-to-analog converters using error detection and correction, and digital radio chips. Digitally controlled sound chips are also mixed-signal circuits. With the advent of cellular and network technology, this category now includes cellular telephone, software radio, and LAN and WAN router integrated circuits.

Design and development

Typically, mixed-signal chips perform some whole function or sub-function in a larger assembly, such as the radio subsystem of a cell phone, or the read data path and laser SLED control logic of a DVD player. Mixed-signal ICs often contain an entire system-on-a-chip. They may also contain on-chip memory blocks (like OTP), which complicates the manufacturing compared to analog ICs. A mixed-signal IC minimizes off-chip interconnects between digital and analog functionality in the system—typically reducing size and weight due to minimized packaging and a smaller module substrate—and therefore increases the reliability of the system.

Because of the use of both digital signal processing and analog circuitry, mixed-signal ICs are usually designed for a very specific purpose. Their design requires a high level of expertise and careful use of computer aided design (CAD) tools. There also exists specific design tools (like mixed-signal simulators) or description languages (like VHDL-AMS). Automated testing of the finished chips can also be challenging. Teradyne, Keysight, and Advantest are the major suppliers of the test equipment for mixed-signal chips.

There are several particular challenges of mixed-signal circuit manufacturing:

- CMOS technology is usually optimal for digital performance, while bipolar junction transistors are usually optimal for analog performance. However, until the last decade, it was difficult to combine these cost-effectively or to design both in a single technology without serious performance compromises. The advent of technologies like high performance CMOS, BiCMOS, CMOS SOI, and SiGe have removed many of these former compromises.
- Testing functional operation of mixed-signal ICs remains complex, expensive, and often is a "one-off" implementation task (meaning a lot of work is necessary for a product with a single, specific use).

- Systematic design methods of analog and mixed-signal circuits are far more primitive than digital circuits. In general, analog circuit design cannot be automated to nearly the extent that digital circuit design can. Combining the two technologies multiplies this complication.
- Fast-changing digital signals send noise to sensitive analog inputs. One path for this noise is substrate coupling. A variety of techniques are used to attempt to block or cancel this noise coupling, such as fully differential amplifiers,^[5] P+ guard-rings,^[6] differential topology, on-chip decoupling, and triple-well isolation.^[7]

Variations

Mixed-signal devices are available as standard parts, but sometimes custom-designed application-specific integrated circuits (ASICs) are necessary. ASICs are designed for new applications, when new standards emerge, or when new energy source(s) are implemented in the system. Due to their specialization, ASICs are usually only developed when production volumes are estimated to be high. The availability of ready-and-tested analog- and mixed-signal IP blocks from foundries or dedicated design houses has lowered the gap to realize mixed-signal ASICs.

There also exist mixed-signal field-programmable gate arrays (FPGAs) and microcontrollers.^[note 1] In these, the same chip that handles digital logic may contain mixed-signal structures like analog-to-digital and digital-to-analog converter(s), operational amplifiers, or wireless connectivity blocks.^[8] These mixed-signal FPGAs and microcontrollers are bridging the gap between standard mixed-signal devices, full-custom ASICs, and embedded software; they offer a solution during product development or when product volume is too low to justify an ASIC. However, they can have performance limitations, such as the resolution of the analog-to-digital converters, the speed of digital-to-analog conversion, or a limited number of inputs and outputs. Nevertheless, they can speed up the system architecture design, prototyping, and even production (at small and medium scales). Their usage also can be supported with development boards, development community, and possibly software support.

History

MOS switched-capacitor circuits

The MOSFET was invented at Bell Labs between 1955 and 1960, after Frosch and Derick discovered and used surface passivation by silicon dioxide to create the first planar transistors, the first in which drain and source were adjacent at the same surface.^{[9][10][11][12][13]} Robert Noyce and Jack Kilby invention of the silicon integrated circuit was enabled by the planar process developed by Jean Hoerni.^[14] In turn, Hoerni's planar process was inspired by the surface passivation method developed at Bell Labs by Carl Frosch and Lincoln Derick in 1955 and 1957.^{[15][16][17][18][19][20][21]}

MOS technology eventually became practical for telephony applications with the MOS mixed-signal integrated circuit, which combines analog and digital signal processing on a single chip, developed by former Bell engineer David A. Hodges with Paul R. Gray at UC Berkeley in the early 1970s.^[22] In 1974, Hodges and Gray worked with R.E. Suarez to develop MOS switched capacitor (SC) circuit technology, which they used to develop a digital-to-analog converter (DAC) chip, using MOS capacitors and MOSFET switches for data conversion.^[22] MOS analog-to-digital converter (ADC) and DAC chips were commercialized by 1974.^[23]

MOS SC circuits led to the development of pulse-code modulation (PCM) codec-filter chips in the late 1970s.^{[22][24]} The silicon-gate CMOS (complementary MOS) PCM codec-filter chip, developed by Hodges and W.C. Black in 1980,^[22] has since been the industry standard for digital telephony.^{[22][24]} By the 1990s, telecommunication networks such as the public switched telephone network (PSTN) had been largely digitized with very-large-scale integration (VLSI) CMOS PCM codec-filters, widely used in electronic switching systems for telephone exchanges, private branch exchanges (PBX), and key telephone systems (KTS); user-end modems; data transmission applications such as digital loop carriers, pair gain multiplexers, telephone loop extenders, integrated services digital network (ISDN) terminals, digital cordless telephones, and digital cell phones; and applications such as speech recognition equipment, voice data storage, voice mail, and digital tapeless answering machines.^[24] The bandwidth of digital telecommunication networks has been rapidly increasing at an exponential rate, as observed by Edholm's law,^[25] largely driven by the rapid scaling and miniaturization of MOS technology.^{[26][22]}

RF CMOS circuits

While working at Bell Labs in the early 1980s, Pakistani engineer Asad Abidi worked on the development of sub-micron MOSFET (metal–oxide–semiconductor field-effect transistor) VLSI (very large-scale integration) technology at the Advanced LSI Development Lab, along with Marty Lepselter, George E. Smith, and Harry Bol. As one of the few circuit designers at the lab, Abidi demonstrated the potential of sub-micron NMOS integrated circuit technology in high-speed communication circuits, and developed the first MOS amplifiers for Gb/s data rates in optical fiber receivers. Abidi's work was initially met with skepticism from proponents of gallium arsenide and bipolar junction transistors, the dominant technologies for high-speed circuits at the time. In 1985, he joined UCLA, where he pioneered RF CMOS technology in the late 1980s. His work changed the way in which radio-frequency (RF) circuits would be designed, away from discrete bipolar transistors and towards CMOS integrated circuits.^[27]

Abidi was researching analog CMOS circuits for signal processing and communications during the late 1980s to early 1990s. In the mid-1990s, the RF CMOS technology that he pioneered was widely adopted in wireless networking, as mobile phones began entering widespread use. As of 2008, the radio transceivers in all wireless networking devices and modern mobile phones are mass-produced as RF CMOS devices.^[27]

The baseband processors^{[28][29]} and radio transceivers in all modern wireless networking devices and mobile phones are mass-produced using RF CMOS devices.^[27] RF CMOS circuits are widely used to transmit and receive wireless signals in a variety of applications, such as satellite technology (such as GPS), Bluetooth, Wi-Fi, near-field communication (NFC), mobile networks (such as 3G, 4G, and 5G), terrestrial broadcast, and automotive radar applications, among other uses.^[30] RF CMOS technology is crucial to modern wireless communications, including wireless networks and mobile communication devices.^[31]

Commercial examples

- Examples of mixed-signal design houses and resources:
 - AnSem (<http://www.ansem.com>)

- [CoreHW](https://www.coreHW.com) (<https://www.coreHW.com>)
 - [EnSilica](https://www.ensilica.com) (<https://www.ensilica.com>)
 - [ICsense](http://www.icsense.com) (<http://www.icsense.com>)
 - [Presto Engineering](https://www.presto-eng.com/) (<https://www.presto-eng.com/>)
 - [Sondrel](https://www.sondrel.com/) (<https://www.sondrel.com/>)
 - [System to ASIC](http://www.system-to-asic.com) (<http://www.system-to-asic.com>)
 - [Triad Semiconductor](http://www.triadsemi.com) (<http://www.triadsemi.com>)
- Examples of mixed signal FPGAs and microcontrollers:
 - [Analog Devices CM4xx Mixed-Signal Control Processors](#)
 - [Fusion FPGA](https://www.microchip.com/en-us/products/fpgas-and-plds/fpgas/fusion-fpgas) (<https://www.microchip.com/en-us/products/fpgas-and-plds/fpgas/fusion-fpgas>) (from Microsemi, now part of Microchip Technology)
 - [Cypress PSoC](#) – "programmable system on chip", a product from Infineon Technologies (former Cypress Semiconductor)
 - [Texas Instruments' MSP430](#)
 - [Xilinx mixed signal FPGA](https://www.xilinx.com/publications/prod_mktg/analog-mixed-signal-product-brief.pdf) (https://www.xilinx.com/publications/prod_mktg/analog-mixed-signal-product-brief.pdf)
 - Examples of mixed signal foundries:^[note 2]
 - [GlobalFoundries](#)
 - [New Japan Radio](#)
 - [Tower Semiconductor Ltd](#)
 - [X-Fab](#)
 - [List of sound chips](#)
 - [Yamaha FM synthesis sound chips](#)
 - [POKEY](#)
 - [MOS Technology SID](#)

See also

- [Analog front-end](#)
- [RFIC](#)

Notes

1. Mixed-signal FPGAs are an extension of [field-programmable analog arrays](#).
2. Some foundries may also have design service or list of partners capable for mixed signal design services for their technologies.

References

1. Saraju Mohanty, Nanoelectronic Mixed-Signal System Design, McGraw-Hill, 2015, ISBN 978-0071825719 and 0071825711.
2. "Mixed-Signal IC Design" (<http://thomasweldon.com/tpw/mixsig/>). quote: "mixed-signal (IC's with mixed analog and digital circuits on a single chip)"

3. Mark Burns and Gordon W. Roberts, "An Introduction to Mixed-Signal IC Test and Measurement", 2001.
4. "ESS Mixed Signal Circuits" (<http://www.cmoset.com/uploads/6A.4-08.pdf>) Archived (<https://web.archive.org/web/20101011225325/http://www.cmoset.com/uploads/6A.4-08.pdf>) 2010-10-11 at the Wayback Machine
5. Chang, J.J.; Myunghee Lee; Sungyong Jung; Brooke, M.A.; Jokerst, N.M.; Wills, D.S. (1999). "Fully differential current-input CMOS amplifier front-end suppressing mixed signal substrate noise for optoelectronic applications" (<https://ieeexplore.ieee.org/document/777869>). *ISCAS'99. Proceedings of the 1999 IEEE International Symposium on Circuits and Systems VLSI (Cat. No. 99CH36349)*. Vol. 1. pp. 327–330. doi:[10.1109/ISCAS.1999.777869](https://doi.org/10.1109/ISCAS.1999.777869) (<https://doi.org/10.1109%2FISCAS.1999.777869>). ISBN 0-7803-5471-0. S2CID 206955680 (<https://api.semanticscholar.org/CorpusID:206955680>).
6. Singh, R. (1997). "Substrate noise issues in mixed-signal chip designs using Spice" (<https://ieeexplore.ieee.org/document/674725>). *International Conference on Electromagnetic Compatibility*. Vol. 1997. pp. 108–112. doi:[10.1049/cp:19971128](https://doi.org/10.1049/cp:19971128) (<https://doi.org/10.1049%2Fcp%3A19971128>). ISBN 0-85296-695-4.
7. "Mixed-Signal IC Merges 14-Bit ADC With DSP In 0.18-μm CMOS" (<http://electronicdesign.com/Articles/ArticleID/2499/2499.html>)
8. "Microsemi Fusion mixed-signal FPGA" (<https://www.microsemi.com/product-directory/fpgas/1691-fusion>)
9. US2802760A (<https://patents.google.com/patent/US2802760A>), Lincoln, Derick & Frosch, Carl J., "Oxidation of semiconductive surfaces for controlled diffusion", issued 1957-08-13
10. Frosch, C. J.; Derick, L (1957). "Surface Protection and Selective Masking during Diffusion in Silicon" (<https://iopscience.iop.org/article/10.1149/1.2428650>). *Journal of the Electrochemical Society*. **104** (9): 547. doi:[10.1149/1.2428650](https://doi.org/10.1149/1.2428650) (<https://doi.org/10.1149%2F1.2428650>).
11. Lojek, Bo (2007). *History of Semiconductor Engineering*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. p. 321. ISBN 978-3-540-34258-8.
12. Ligenza, J.R.; Spitzer, W.G. (1960). "The mechanisms for silicon oxidation in steam and oxygen" (<https://linkinghub.elsevier.com/retrieve/pii/0022369760902195>). *Journal of Physics and Chemistry of Solids*. **14**: 131–136. Bibcode:1960JPCS...14..131L (<https://ui.adsabs.harvard.edu/abs/1960JPCS...14..131L>). doi:[10.1016/0022-3697\(60\)90219-5](https://doi.org/10.1016/0022-3697(60)90219-5) (<https://doi.org/10.1016%2F0022-3697%2860%2990219-5>).
13. Lojek, Bo (2007). *History of Semiconductor Engineering*. Springer Science & Business Media. p. 120. ISBN 9783540342588.
14. Christophe Lécuyer (2006). *Making Silicon Valley: Innovation and the Growth of High Tech, 1930-1970* (<https://archive.org/details/makingsiliconval00chri>). MIT Press. pp. 214 (<https://archive.org/details/makingsiliconval00chri/page/214>)–252. ISBN 978-0-262-12281-8. "fairchild planar Making Silicon Valley."
15. US2802760A (<https://patents.google.com/patent/US2802760A>), Lincoln, Derick & Frosch, Carl J., "Oxidation of semiconductive surfaces for controlled diffusion", issued 1957-08-13
16. Huff, Howard; Riordan, Michael (2007-09-01). "Frosch and Derick: Fifty Years Later (Foreword)" (<https://iopscience.iop.org/article/10.1149/2.F02073IF>). *The Electrochemical Society Interface*. **16** (3): 29. doi:[10.1149/2.F02073IF](https://doi.org/10.1149/2.F02073IF) (<https://doi.org/10.1149%2F2.F02073F>). ISSN 1064-8208 (<https://search.worldcat.org/issn/1064-8208>).
17. Lojek, Bo (2007). *History of Semiconductor Engineering*. Springer Science & Business Media. p. 120. ISBN 9783540342588.
18. Frosch, C. J.; Derick, L (1957). "Surface Protection and Selective Masking during Diffusion in Silicon" (<https://iopscience.iop.org/article/10.1149/1.2428650>). *Journal of the Electrochemical Society*. **104** (9): 547. doi:[10.1149/1.2428650](https://doi.org/10.1149/1.2428650) (<https://doi.org/10.1149%2F1.2428650>).

19. Moskowitz, Sanford L. (2016). *Advanced Materials Innovation: Managing Global Technology in the 21st century* (<https://books.google.com/books?id=2STRDAAAQBAJ&pg=PA168>). John Wiley & Sons. p. 168. ISBN 978-0-470-50892-3.
20. Christophe Lécuyer; David C. Brook; Jay Last (2010). *Makers of the Microchip: A Documentary History of Fairchild Semiconductor* (<https://books.google.com/books?id=LaZpUpkG70QC&pg=PA62>). MIT Press. pp. 62–63. ISBN 978-0-262-01424-3.
21. Claeys, Cor L. (2003). *ULSI Process Integration III: Proceedings of the International Symposium* (<https://books.google.com/books?id=bu22JNYbE5MC&pg=PA27>). The Electrochemical Society. pp. 27–30. ISBN 978-1-56677-376-8.
22. Allstot, David J. (2016). "Switched Capacitor Filters" (https://ieeexpubs.org/sites/default/files/a_short_history_of_circuits_and_systems_ebook_web.pdf) (PDF). In Maloberti, Franco; Davies, Anthony C. (eds.). *A Short History of Circuits and Systems: From Green, Mobile, Pervasive Networking to Big Data Computing*. IEEE Circuits and Systems Society. pp. 105–110. ISBN 9788793609860.
23. *Electronic Components* (<https://books.google.com/books?id=HikuAAAAMAAJ&pg=PA46>). U.S. Government Printing Office. 1974. p. 46.
24. Floyd, Michael D.; Hillman, Garth D. (8 October 2018) [1st pub. 2000]. "Pulse-Code Modulation Codec-Filters" (<https://books.google.com/books?id=Tokk5bZxB0MC&pg=SA26-PA1>). *The Communications Handbook* (2nd ed.). CRC Press. pp. 26–1, 26–2, 26–3. ISBN 9781420041163.
25. Cherry, Steven (2004). "Edholm's law of bandwidth". *IEEE Spectrum*. **41** (7): 58–60. doi:10.1109/MSPEC.2004.1309810 (<https://doi.org/10.1109%2FMSPEC.2004.1309810>). S2CID 27580722 (<https://api.semanticscholar.org/CorpusID:27580722>).
26. Jindal, Renuka P. (2009). "From millibits to terabits per second and beyond - over 60 years of innovation" (<https://events.vtools.ieee.org/m/195547>). *2009 2nd International Workshop on Electron Devices and Semiconductor Technology*. pp. 1–6. doi:10.1109/EDST.2009.5166093 (<https://doi.org/10.1109%2FEDST.2009.5166093>). ISBN 978-1-4244-3831-0. S2CID 25112828 (<https://api.semanticscholar.org/CorpusID:25112828>).
27. O'Neill, A. (2008). "Asad Abidi Recognized for Work in RF-CMOS". *IEEE Solid-State Circuits Society Newsletter*. **13** (1): 57–58. doi:10.1109/N-SSC.2008.4785694 (<https://doi.org/10.1109%2FN-SSC.2008.4785694>). ISSN 1098-4232 (<https://search.worldcat.org/issn/1098-4232>).
28. Chen, Wai-Kai (2018). *The VLSI Handbook* (<https://books.google.com/books?id=rMsqBgAAQBAJ&pg=SA60-PA2>). CRC Press. pp. 60–2. ISBN 9781420005967.
29. Morgado, Alonso; Río, Rocío del; Rosa, José M. de la (2011). *Nanometer CMOS Sigma-Delta Modulators for Software Defined Radio* (<https://books.google.com/books?id=Alv6nWVCKdIC&pg=PA1>). Springer Science & Business Media. p. 1. ISBN 9781461400370.
30. Veendrick, Harry J. M. (2017). *Nanometer CMOS ICs: From Basics to ASICs* (https://books.google.com/books?id=Lv_EDgAAQBAJ&pg=PA243). Springer. p. 243. ISBN 9783319475974.
31. "Infineon Hits Bulk-CMOS RF Switch Milestone" (<https://www.eetasia.com/news/article/18112004-infineon-hits-bulk-cmos-rf-switch-milestone>). *EE Times*. 20 November 2018. Retrieved 26 October 2019.

Further reading

- Saraju Mohanty (2015). *Nanoelectronic Mixed-Signal System Design* (<http://www.mhprofessional.com/product.php?isbn=0071825711>). McGraw-Hill. ISBN 978-0071825719.
- R. Jacob Baker (2009). *CMOS Mixed-Signal Circuit Design, Second Edition*. <http://CMOSedu.com/>



Printed circuit board

A **printed circuit board (PCB)**, also called **printed wiring board (PWB)**, is a laminated sandwich structure of conductive and insulating layers, each with a pattern of traces, planes and other features (similar to wires on a flat surface) etched from one or more sheet layers of copper laminated onto or between sheet layers of a non-conductive substrate.^[1] PCBs are used to connect or "wire" components to one another in an electronic circuit. Electrical components may be fixed to conductive pads on the outer layers, generally by soldering, which both electrically connects and mechanically fastens the components to the board. Another manufacturing process adds vias, metal-lined drilled holes that enable electrical interconnections between conductive layers, to boards with more than a single side.

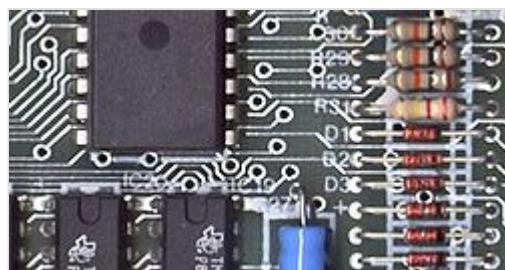
Printed circuit boards are used in nearly all electronic products today. Alternatives to PCBs include wire wrap and point-to-point construction, both once popular but now rarely used. PCBs require additional design effort to lay out the circuit, but manufacturing and assembly can be automated. Electronic design automation software is available to do much of the work of layout. Mass-producing circuits with PCBs is cheaper and faster than with other wiring methods, as components are mounted and wired in one operation. Large numbers of PCBs can be fabricated at the same time, and the layout has to be done only once. PCBs can also be made manually in small quantities, with reduced benefits.^[2]

PCBs can be single-sided (one copper layer), double-sided (two copper layers on both sides of one substrate layer), or multi-layer (stacked layers of substrate with copper plating sandwiched between each and on the outside layers). Multi-layer PCBs provide much higher component density, because circuit traces on the inner layers would otherwise take up surface space between components. The rise in popularity of multilayer PCBs with more than two, and especially with more than four, copper planes was concurrent with the adoption of surface-mount technology. However, multilayer PCBs make repair, analysis, and field modification of circuits much more difficult and usually impractical.

The world market for bare PCBs exceeded US\$60.2 billion in 2014,^[3] and was estimated at \$80.33 billion in 2024, forecast to be \$96.57 billion for 2029, growing at 4.87% per annum.^[4]



Printed circuit board of a DVD player



Part of a 1984 Sinclair ZX Spectrum computer board, a printed circuit board, showing the conductive traces, the through-hole paths to the other surface, and some electronic components mounted using through-hole mounting

History

Predecessors

Before the development of printed circuit boards, electrical and electronic circuits were wired point-to-point on a chassis. Typically, the chassis was a sheet metal frame or pan, sometimes with a wooden bottom. Components were attached to the chassis, usually by insulators when the connecting point on the chassis was metal, and then their leads were connected directly or with jumper wires by soldering, or sometimes using crimp connectors, wire connector lugs on screw terminals, or other methods. Circuits were large, bulky, heavy, and relatively fragile (even discounting the breakable glass envelopes of the vacuum tubes that were often included in the circuits), and production was labor-intensive, so the products were expensive.

Development of the methods used in modern printed circuit boards started early in the 20th century. In 1903, a German inventor, Albert Hanson, described flat foil conductors laminated to an insulating board, in multiple layers. Thomas Edison experimented with chemical methods of plating conductors onto linen paper in 1904. Arthur Berry in 1913 patented a print-and-etch method in the UK, and in the United States Max Schoop obtained a patent^[5] to flame-spray metal onto a board through a patterned mask. Charles Ducas in 1925 patented a method of electroplating circuit patterns.^[6]

Predating the printed circuit invention, and similar in spirit, was John Sargrove's 1936–1947 Electronic Circuit Making Equipment (ECME) that sprayed metal onto a Bakelite plastic board. The ECME could produce three radio boards per minute.

Early PCBs

The Austrian engineer Paul Eisler invented the printed circuit as part of a radio set while working in the UK around 1936. In 1941 a multi-layer printed circuit was used in German magnetic influence naval mines.

Around 1943 the United States began to use the technology on a large scale to make proximity fuzes for use in World War II.^[6] Such fuzes required an electronic circuit that could withstand being fired from a gun, and could be produced in quantity. The Centralab Division of Globe Union submitted a proposal which met the requirements: a ceramic plate would be screenprinted with metallic paint for conductors and carbon material for resistors, with ceramic disc capacitors and subminiature vacuum tubes soldered in place.^[7] The technique proved viable, and the resulting patent on the process, which was classified by the U.S. Army, was assigned to Globe Union. It was not until 1984 that the Institute of Electrical and Electronics Engineers (IEEE) awarded Harry W. Rubinstein its Cledo Brunetti Award for early key contributions to the development of printed components and conductors on a common insulating substrate. Rubinstein was honored in 1984 by his alma mater, the University of Wisconsin-Madison, for his innovations in the



Proximity fuze Mark 53 production line 1944

technology of printed electronic circuits and the fabrication of capacitors.^{[8][9]} This invention also represents a step in the development of integrated circuit technology, as not only wiring but also passive components were fabricated on the ceramic substrate.

Post-war developments

In 1948, the US released the invention for commercial use. Printed circuits did not become commonplace in consumer electronics until the mid-1950s, after the *Auto-Sembly* process was developed by the United States Army. At around the same time in the UK work along similar lines was carried out by Geoffrey Dummer, then at the RRDE.

Motorola was an early leader in bringing the process into consumer electronics, announcing in August 1952 the adoption of "plated circuits" in home radios after six years of research and a \$1M investment.^[10] Motorola soon began using its trademarked term for the process, PLAcir, in its consumer radio advertisements.^[11] Hallicrafters released its first "foto-etch" printed circuit product, a clock-radio, on November 1, 1952.^[12]

Even as circuit boards became available, the point-to-point chassis construction method remained in common use in industry (such as TV and hi-fi sets) into at least the late 1960s. Printed circuit boards were introduced to reduce the size, weight, and cost of parts of the circuitry. In 1960, a small consumer radio receiver might be built with all its circuitry on one circuit board, but a TV set would probably contain one or more circuit boards.

Originally, every electronic component had wire leads, and a PCB had holes drilled for each wire of each component. The component leads were then inserted through the holes and soldered to the copper PCB traces. This method of assembly is called through-hole construction. In 1949, Moe Abramson and Stanislaus F. Danko of the United States Army Signal Corps developed the *Auto-Sembly* process in which component leads were inserted into a copper foil interconnection pattern and dip soldered. The patent they obtained in 1956 was assigned to the U.S. Army.^[13] With the development of board lamination and etching techniques, this concept evolved into the standard printed circuit board fabrication process in use today. Soldering could be done automatically by passing the board over a ripple, or wave, of molten solder in a wave-soldering machine. However, the wires and holes are inefficient since drilling holes is expensive and consumes drill bits and the protruding wires are cut off and discarded.

Since the 1980s, surface mount parts have increasingly replaced through-hole components, enabling smaller boards and lower production costs, but making repairs more challenging.

In the 1990s the use of multilayer surface boards became more frequent. As a result, size was further minimized and both flexible and rigid PCBs were incorporated in different devices. In 1995 PCB manufacturers began using microvia technology to produce High-Density Interconnect (HDI) PCBs.^[14]

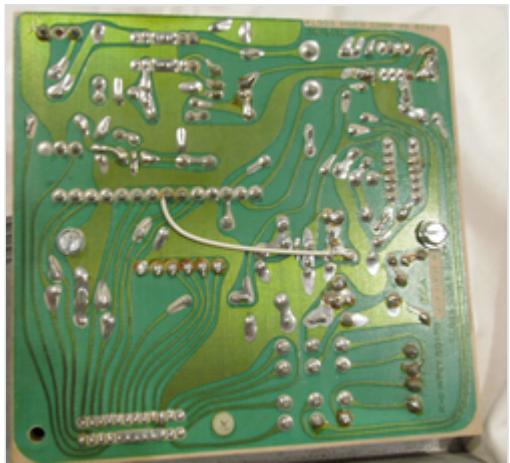
Recent advances

Recent advances in 3D printing have meant that there are several new techniques in PCB creation. 3D printed electronics (PEs) can be utilized to print items layer by layer and subsequently the item can be printed with a liquid ink that contains electronic functionalities.

HDI (High Density Interconnect) technology allows for a denser design on the PCB and thus potentially smaller PCBs with more traces and components in a given area. As a result, the paths between components can be shorter. HDIs use blind/buried vias, or a combination that includes microvias. With multi-layer HDI PCBs the interconnection of several vias stacked on top of each other (stacked vías, instead of one deep buried via) can be made stronger, thus enhancing reliability in all conditions. The most common applications for HDI technology are computer and mobile phone components as well as medical equipment and military communication equipment. A 4-layer HDI microvia PCB is equivalent in quality to an 8-layer through-hole PCB, so HDI technology can reduce costs. HDI PCBs are often made using build-up film such as ajinomoto build-up film, which is also used in the production of flip chip packages.^{[15][16]} Some PCBs have optical waveguides, similar to optical fibers built on the PCB.^[17]

Composition

A basic PCB consists of a flat sheet of insulating material and a layer of copper foil, laminated to the substrate. Chemical etching divides the copper into separate conducting lines called tracks or *circuit traces*, pads for connections, vias to pass connections between layers of copper, and features such as solid conductive areas for electromagnetic shielding or other purposes. The tracks function as wires fixed in place, and are insulated from each other by air and the board substrate material. The surface of a PCB may have a coating that protects the copper from corrosion and reduces the chances of solder shorts between traces or undesired electrical contact with stray bare wires. For its function in helping to prevent solder shorts, the coating is called solder resist or solder mask.



An example of hand-drawn etched traces on a PCB

The pattern to be etched into each copper layer of a PCB is called the "artwork". The etching is usually done using photoresist which is coated onto the PCB, then exposed to light projected in the pattern of the artwork. The resist material protects the copper from dissolution into the etching solution. The etched board is then cleaned. A PCB design can be mass-reproduced in a way similar to the way photographs can be mass-duplicated from film negatives using a photographic printer.

FR-4 glass epoxy is the most common insulating substrate. Another substrate material is cotton paper impregnated with phenolic resin, often tan or brown.

When a PCB has no components installed, it is less ambiguously called a *printed wiring board (PWB)* or *etched wiring board*.^[18] However, the term "printed wiring board" has fallen into disuse. A PCB populated with electronic components is called a *printed circuit assembly (PCA)*, *printed circuit board assembly* or *PCB assembly* (*PCBA*). In informal usage, the term "printed circuit board" most commonly means "printed circuit assembly" (with components). The IPC preferred term for an assembled board is circuit card assembly (CCA),^[19] and for an assembled backplane it is backplane assembly. "Card" is another widely used informal term for a "printed circuit assembly". For example, expansion card.

A PCB may be printed with a legend identifying the components, test points, or identifying text. Originally, silkscreen printing was used for this purpose, but today other, finer quality printing methods are usually used. Normally the legend does not affect the function of a PCBA.

Layers

A printed circuit board can have multiple layers of copper which almost always are arranged in pairs. The number of layers and the interconnection designed between them (vias, PTHs) provide a general estimate of the board complexity. Using more layers allow for more routing options and better control of signal integrity, but are also time-consuming and costly to manufacture. Likewise, selection of the vias for the board also allow fine tuning of the board size, escaping of signals off complex ICs, routing, and long term reliability, but are tightly coupled with production complexity and cost.

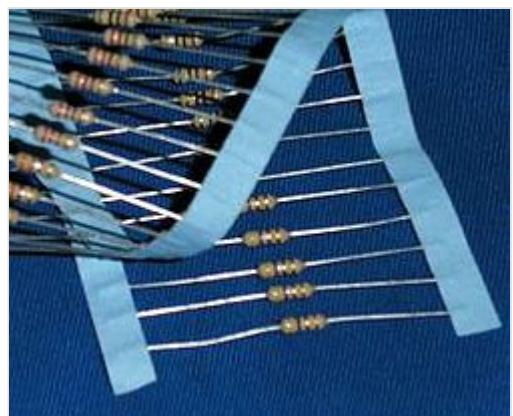
One of the simplest boards to produce is the two-layer board. It has copper on both sides that are referred to as external layers; multi layer boards sandwich additional internal layers of copper and insulation. After two-layer PCBs, the next step up is the four-layer. The four layer board adds significantly more routing options in the internal layers as compared to the two layer board, and often some portion of the internal layers is used as ground plane or power plane, to achieve better signal integrity, higher signaling frequencies, lower EMI, and better power supply decoupling.

In multi-layer boards, the layers of material are laminated together in an alternating sandwich: copper, substrate, copper, substrate, copper, etc.; each plane of copper is etched, and any internal vias (that will not extend to both outer surfaces of the finished multilayer board) are plated-through, before the layers are laminated together. Only the outer layers need be coated; the inner copper layers are protected by the adjacent substrate layers.

Component mounting

"Through hole" components are mounted by their wire leads passing through the board and soldered to traces on the other side. "Surface mount" components are attached by their leads to copper traces on the same side of the board. A board may use both methods for mounting components. PCBs with only through-hole mounted components are now uncommon. Surface mounting is used for transistors, diodes, IC chips, resistors, and capacitors. Through-hole mounting may be used for some large components such as electrolytic capacitors and connectors.

The first PCBs used through-hole technology, mounting electronic components by lead inserted through holes on one side of the board and soldered onto copper traces on the other side. Boards may be single-sided, with an unplated component side, or more compact double-sided boards, with components soldered on both sides. Horizontal installation of through-hole parts with two axial leads (such as resistors, capacitors, and diodes) is done by bending the leads 90 degrees in the same direction, inserting the part in the board (often bending leads located on the back of the board in opposite directions to improve the part's mechanical strength), soldering the leads, and trimming off the ends. Leads may be soldered either manually or by a wave soldering machine.^[20]



Through-hole (leaded) resistors

Surface-mount technology emerged in the 1960s, gained momentum in the early 1980s, and became widely used by the mid-1990s. Components were mechanically redesigned to have small metal tabs or end caps that could be soldered directly onto the PCB surface, instead of wire leads to pass through holes. Components became much smaller and component placement on both sides of the board became more common than with through-hole mounting, allowing much smaller PCB assemblies with much higher circuit densities. Surface mounting lends itself well to a high degree of automation, reducing labor costs and greatly increasing production rates compared with through-hole circuit boards. Components can be supplied mounted on carrier tapes. Surface mount components can be about one-quarter to one-tenth of the size and weight of through-hole components, and passive components much cheaper. However, prices of semiconductor surface mount devices (SMDs) are determined more by the chip itself than the package, with little price advantage over larger packages, and some wire-ended components, such as 1N4148 small-signal switch diodes, are actually significantly cheaper than SMD equivalents.

Electrical properties

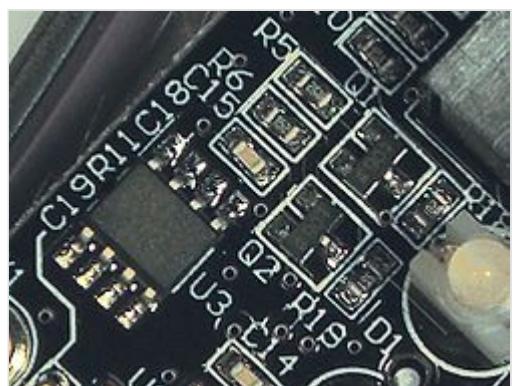
Each trace consists of a flat, narrow part of the copper foil that remains after etching. Its resistance, determined by its width, thickness, and length, must be sufficiently low for the current the conductor will carry. Power and ground traces may need to be wider than signal traces. In a multi-layer board one entire layer may be mostly solid copper to act as a ground plane for shielding and power return. For microwave circuits, transmission lines can be laid out in a planar form such as stripline or microstrip with carefully controlled dimensions to assure a consistent impedance. In radio-frequency and fast switching circuits the inductance and capacitance of the printed circuit board conductors become significant circuit elements, usually undesired; conversely, they can be used as a deliberate part of the circuit design, as in distributed-element filters, antennae, and fuses, obviating the need for additional discrete components. High density interconnects (HDI) PCBs have tracks or vias with a width or diameter of under 152 micrometers.^[21]



Through-hole devices mounted on the circuit board of a mid-1980s Commodore 64 home computer



A box of drill bits used for making holes in printed circuit boards. While tungsten-carbide bits are very hard, they eventually wear out or break. Drilling is a considerable part of the cost of a through-hole printed circuit board.

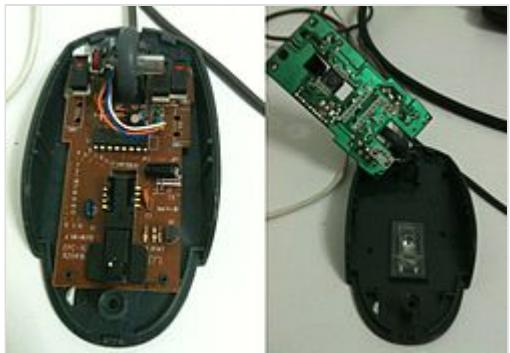


Surface mount components, including resistors, transistors and an integrated circuit

Materials

Laminates

Laminates are manufactured by curing layers of cloth or paper with thermoset resin under pressure and heat to form an integral final piece of uniform thickness. They can be up to 4 by 8 feet (1.2 by 2.4 m) in width and length. Varying cloth weaves (threads per inch or cm), cloth thickness, and resin percentage are used to achieve the desired final thickness and dielectric characteristics. Available standard laminate thickness are listed in ANSI/IPC-D-275.^[22]



A PCB in a computer mouse: the component side (left) and the printed side (right)

The cloth or fiber material used, resin material, and the cloth to resin ratio determine the laminate's type designation (FR-4, CEM-1, G-10, etc.) and therefore the characteristics of the laminate produced. Important characteristics are the level to which the laminate is fire retardant, the dielectric constant (ϵ_r), the loss tangent ($\tan \delta$), the tensile strength, the shear strength, the glass transition temperature (T_g), and the Z-axis expansion coefficient (how much the thickness changes with temperature).

There are quite a few different dielectrics that can be chosen to provide different insulating values depending on the requirements of the circuit. Some of these dielectrics are polytetrafluoroethylene (Teflon), FR-4, FR-1, CEM-1 or CEM-3. Well known pre-preg materials used in the PCB industry are FR-2 (phenolic cotton paper), FR-3 (cotton paper and epoxy), FR-4 (woven glass and epoxy), FR-5 (woven glass and epoxy), FR-6 (matte glass and polyester), G-10 (woven glass and epoxy), CEM-1 (cotton paper and epoxy), CEM-2 (cotton paper and epoxy), CEM-3 (non-woven glass and epoxy), CEM-4 (woven glass and epoxy), CEM-5 (woven glass and polyester). Thermal expansion is an important consideration especially with ball grid array (BGA) and naked die technologies, and glass fiber offers the best dimensional stability.

FR-4 is by far the most common material used today. The board stock with unetched copper on it is called "copper-clad laminate".

With decreasing size of board features and increasing frequencies, small non-homogeneities like uneven distribution of fiberglass or other filler, thickness variations, and bubbles in the resin matrix, and the associated local variations in the dielectric constant, are gaining importance.

Key substrate parameters

The circuit-board substrates are usually dielectric composite materials. The composites contain a matrix (usually an epoxy resin) and a reinforcement (usually a woven, sometimes non-woven, glass fibers, sometimes even paper), and in some cases a filler is added to the resin (e.g. ceramics; titanate ceramics can be used to increase the dielectric constant).

The reinforcement type defines two major classes of materials: woven and non-woven. Woven reinforcements are cheaper, but the high dielectric constant of glass may not be favorable for many higher-frequency applications. The spatially non-homogeneous structure also introduces local variations

in electrical parameters, due to different resin/glass ratio at different areas of the weave pattern. Non-woven reinforcements, or materials with low or no reinforcement, are more expensive but more suitable for some RF/analog applications.

The substrates are characterized by several key parameters, chiefly thermomechanical (glass transition temperature, tensile strength, shear strength, thermal expansion), electrical (dielectric constant, loss tangent, dielectric breakdown voltage, leakage current, tracking resistance...), and others (e.g. moisture absorption).

At the glass transition temperature the resin in the composite softens and significantly increases thermal expansion; exceeding T_g then exerts mechanical overload on the board components - e.g. the joints and the vias. Below T_g the thermal expansion of the resin roughly matches copper and glass, above it gets significantly higher. As the reinforcement and copper confine the board along the plane, virtually all volume expansion projects to the thickness and stresses the plated-through holes. Repeated soldering or other exposition to higher temperatures can cause failure of the plating, especially with thicker boards; thick boards therefore require a matrix with a high T_g .

The materials used determine the substrate's dielectric constant. This constant is also dependent on frequency, usually decreasing with frequency. As this constant determines the signal propagation speed, frequency dependence introduces phase distortion in wideband applications; as flat a dielectric constant vs frequency characteristics as is achievable is important here. The impedance of transmission lines decreases with frequency, therefore faster edges of signals reflect more than slower ones.

Dielectric breakdown voltage determines the maximum voltage gradient the material can be subjected to before suffering a breakdown (conduction, or arcing, through the dielectric).

Tracking resistance determines how the material resists high voltage electrical discharges creeping over the board surface.

Loss tangent determines how much of the electromagnetic energy from the signals in the conductors is absorbed in the board material. This factor is important for high frequencies. Low-loss materials are more expensive. Choosing unnecessarily low-loss material is a common engineering error in high-frequency digital design; it increases the cost of the boards without a corresponding benefit. Signal degradation by loss tangent and dielectric constant can be easily assessed by an eye pattern.

Moisture absorption occurs when the material is exposed to high humidity or water. Both the resin and the reinforcement may absorb water; water also may be soaked by capillary forces through voids in the materials and along the reinforcement. Epoxies of the FR-4 materials are not too susceptible, with absorption of only 0.15%. Teflon has very low absorption of 0.01%. Polyimides and cyanate esters, on the other side, suffer from high water absorption. Absorbed water can lead to significant degradation of key parameters; it impairs tracking resistance, breakdown voltage, and dielectric parameters. Relative dielectric constant of water is about 73, compared to about 4 for common circuit board materials. Absorbed moisture can also vaporize on heating, as during soldering, and cause cracking and delamination,^[23] the same effect responsible for "popcorn" damage on wet packaging of electronic parts. Careful baking of the substrates may be required to dry them prior to soldering.^[24]

Common substrates

Often encountered materials:

- FR-2, phenolic paper or phenolic cotton paper, paper impregnated with a phenol formaldehyde resin. Common in consumer electronics with single-sided boards. Electrical properties inferior to FR-4. Poor arc resistance. Generally rated to 105 °C.
- FR-4, a woven fiberglass cloth impregnated with an epoxy resin. Low water absorption (up to about 0.15%), good insulation properties, good arc resistance. Very common. Several grades with somewhat different properties are available. Typically rated to 130 °C.
- Aluminum, or *metal core board* or *insulated metal substrate* (IMS), clad with thermally conductive thin dielectric - used for parts requiring significant cooling - power switches, LEDs. Consists of usually single, sometimes double layer thin circuit board based on e.g. FR-4, laminated on aluminum sheet metal, commonly 0.8, 1, 1.5, 2 or 3 mm thick. The thicker laminates sometimes also come with thicker copper metallization.^{[25][26]}
- Flexible substrates - can be a standalone copper-clad foil or can be laminated to a thin stiffener, e.g. 50–130 µm
 - Kapton or UPILEX,^[27] a polyimide foil. Used for flexible printed circuits, in this form common in small form-factor consumer electronics or for flexible interconnects. Resistant to high temperatures.
 - Pyralux, a polyimide-fluoropolymer composite foil.^[28] Copper layer can delaminate during soldering.

Less-often encountered materials:

- FR-1, like FR-2, typically specified to 105 °C, some grades rated to 130 °C. Room-temperature punchable. Similar to cardboard. Poor moisture resistance. Low arc resistance.
- FR-3, cotton paper impregnated with epoxy. Typically rated to 105 °C.
- FR-5, woven fiberglass and epoxy, high strength at higher temperatures, typically specified to 170 °C.
- FR-6, matte glass and polyester
- G-10, woven glass and epoxy - high insulation resistance, low moisture absorption, very high bond strength. Typically rated to 130 °C.
- G-11, woven glass and epoxy - high resistance to solvents, high flexural strength retention at high temperatures.^[29] Typically rated to 170 °C.
- CEM-1, cotton paper and epoxy
- CEM-2, cotton paper and epoxy
- CEM-3, non-woven glass and epoxy
- CEM-4, woven glass and epoxy
- CEM-5, woven glass and polyester
- PTFE, ("Teflon") - expensive, low dielectric loss, for high frequency applications, very low moisture absorption (0.01%), mechanically soft. Difficult to laminate, rarely used in multilayer applications.
- PTFE, ceramic filled - expensive, low dielectric loss, for high frequency applications. Varying ceramics/PTFE ratio allows adjusting dielectric constant and thermal expansion.
- RF-35, fiberglass-reinforced ceramics-filled PTFE. Relatively less expensive, good mechanical properties, good high-frequency properties.^{[30][31]}
- Alumina, a ceramic. Hard, brittle, very expensive, very high performance, good thermal conductivity.

- Polyimide, a high-temperature polymer. Expensive, high-performance. Higher water absorption (0.4%). Can be used from cryogenic temperatures to over 260 °C.

Copper thickness

Copper thickness of PCBs can be specified directly or as the weight of copper per area (in ounce per square foot) which is easier to measure. One ounce per square foot is 1.344 mils or 34 micrometers thickness (0.001344 inches). *Heavy copper* is a layer exceeding three ounces of copper per ft², or approximately 4.2 mils (105 µm) (0.0042 inches) thick. Heavy copper layers are used for high current or to help dissipate heat.

On the common FR-4 substrates, 1 oz copper per ft² (35 µm) is the most common thickness; 2 oz (70 µm) and 0.5 oz (17.5 µm) thickness is often an option. Less common are 12 and 105 µm, 9 µm is sometimes available on some substrates. Flexible substrates typically have thinner metalization. Metal-core boards for high power devices commonly use thicker copper; 35 µm is usual but also 140 and 400 µm can be encountered.

In the US, copper foil thickness is specified in units of ounces per square foot (oz/ft²), commonly referred to simply as *ounce*. Common thicknesses are 1/2 oz/ft² (150 g/m²), 1 oz/ft² (300 g/m²), 2 oz/ft² (600 g/m²), and 3 oz/ft² (900 g/m²). These work out to thicknesses of 17.05 µm (0.67 thou), 34.1 µm (1.34 thou), 68.2 µm (2.68 thou), and 102.3 µm (4.02 thou), respectively.

oz/ft²	g/m²	µm	thou
1/2 oz/ft ²	150 g/m ²	17.05 µm	0.67 thou
1 oz/ft ²	300 g/m ²	34.1 µm	1.34 thou
2 oz/ft ²	600 g/m ²	68.2 µm	2.68 thou
3 oz/ft ²	900 g/m ²	102.3 µm	4.02 thou

1/2 oz/ft² foil is not widely used as a finished copper weight, but is used for outer layers when plating for through holes will increase the finished copper weight. Some PCB manufacturers refer to 1 oz/ft² copper foil as having a thickness of 35 µm (may also be referred to as 35 µ, 35 micron, or 35 mic).

- 1/0 – denotes 1 oz/ft² copper one side, with no copper on the other side.
- 1/1 – denotes 1 oz/ft² copper on both sides.
- H/0 or H/H – denotes 0.5 oz/ft² copper on one or both sides, respectively.
- 2/0 or 2/2 – denotes 2 oz/ft² copper on one or both sides, respectively.

Manufacturing

Printed circuit board manufacturing involves manufacturing bare printed circuit boards and then populating them with electronic components. In large-scale board manufacturing, multiple PCBs are grouped on a single panel for efficient processing. After assembly, they are separated (depaneled).

Types

Breakout boards

A minimal PCB for a single component, used for prototyping, is called a **breakout board**. The purpose of a breakout board is to "break out" the leads of a component on separate terminals so that manual connections to them can be made easily. Breakout boards are especially used for surface-mount components or any components with fine lead pitch.

Advanced PCBs may contain components embedded in the substrate, such as capacitors and integrated circuits, to reduce the amount of space taken up by components on the surface of the PCB while improving electrical characteristics.^[32]

Multiwire boards

Multiwire is a patented technique of interconnection which uses machine-routed insulated wires embedded in a non-conducting matrix (often plastic resin).^[33] It was used during the 1980s and 1990s. As of 2010, Multiwire is still available through Hitachi.

Since it was quite easy to stack interconnections (wires) inside the embedding matrix, the approach allowed designers to forget completely about the routing of wires (usually a time-consuming operation of PCB design): Anywhere the designer needs a connection, the machine will draw a wire in a straight line from one location/pin to another. This led to very short design times (no complex algorithms to use even for high density designs) as well as reduced crosstalk (which is worse when wires run parallel to each other—which almost never happens in Multiwire), though the cost is too high to compete with cheaper PCB technologies when large quantities are needed.

Corrections can be made to a Multiwire board layout more easily than to a PCB layout.^[34]

Cordwood construction

Cordwood construction can save significant space and was often used with wire-ended components in applications where space was at a premium (such as fuzes, missile guidance, and telemetry systems) and in high-speed computers, where short traces were important. In cordwood construction, axial-leaded components were mounted between two parallel planes. The name comes from the way axial-lead components (capacitors, resistors, coils, and diodes) are stacked in parallel rows and columns, like a stack of firewood. The components were either soldered together with jumper wire or they were connected to



A breakout board can allow interconnection between two incompatible connectors.



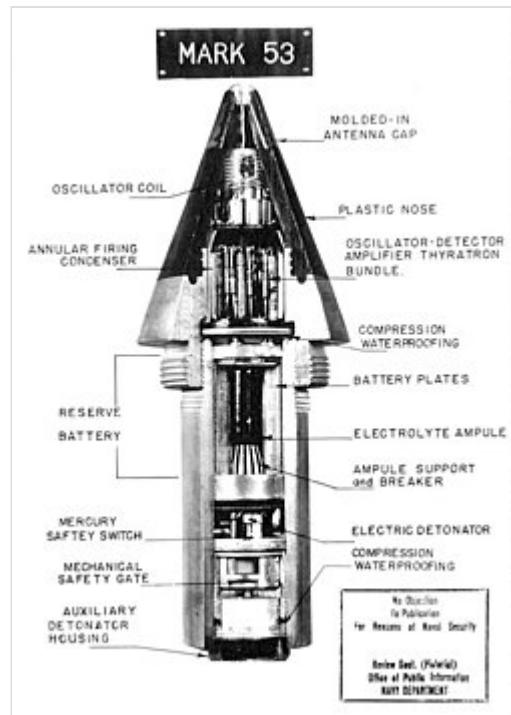
This breakout board allows an SD card's pins to be accessed easily while still allowing the card to be hot-swapped.

other components by thin nickel ribbon welded at right angles onto the component leads.^[35] To avoid shorting together different interconnection layers, thin insulating cards were placed between them. Perforations or holes in the cards allowed component leads to project through to the next interconnection layer. One disadvantage of this system was that special nickel-leaded components had to be used to allow reliable interconnecting welds to be made. Differential thermal expansion of the component could put pressure on the leads of the components and the PCB traces and cause mechanical damage (as was seen in several modules on the Apollo program). Additionally, components located in the interior are difficult to replace. Some versions of cordwood construction used soldered single-sided PCBs as the interconnection method (as pictured), allowing the use of normal-leaded components at the cost of being difficult to remove the boards or replace any component that is not at the edge.

Before the advent of integrated circuits, this method allowed the highest possible component packing density; because of this, it was used by a number of computer vendors including Control Data Corporation.



A cordwood module



Cordwood construction was used in proximity fuzes.

Uses

Printed circuit boards have been used as an alternative to their typical use for electronic and biomedical engineering thanks to the versatility of their layers, especially the copper layer. PCB layers have been used to fabricate sensors, such as capacitive pressure sensors and accelerometers, actuators such as microvalves and microheaters, as well as platforms of sensors and actuators for Lab-on-a-chip (LoC), for example to perform polymerase chain reaction (PCR), and fuel cells, to name a few.^[36]

Repair

Manufacturers may not support component-level repair of printed circuit boards because of the relatively low cost to replace compared with the time and cost of troubleshooting to a component level. In board-level repair, the technician identifies the board (PCA) on which the fault resides and replaces it. This shift is economically efficient from a manufacturer's point of view but is also materially wasteful, as a circuit board with hundreds of functional components may be discarded and replaced due to the failure of one minor and inexpensive part, such as a resistor or capacitor, and this practice is a significant contributor to the problem of e-waste.^[37]

Legislation

In many countries (including all European Single Market participants,^[38] the United Kingdom,^[39] Turkey, and China), legislation restricts the use of lead, cadmium, and mercury in electrical equipment. PCBs sold in such countries must therefore use lead-free manufacturing processes and lead-free solder, and attached components must themselves be compliant.^{[40][41]}

Safety Standard UL 796 covers component safety requirements for printed wiring boards for use as components in devices or appliances. Testing analyzes characteristics such as flammability, maximum operating temperature, electrical tracking, heat deflection, and direct support of live electrical parts.

See also



- [Breadboard](#)
- [BT-Epoxy - resin used in PCBs](#)
- [Certified interconnect designer - qualification for PCB designers](#)
- [Occam process - solder-free circuit board manufacture method](#)

References

1. "What Is a Printed Circuit Board (PCB)? - Technical Articles" (<https://www.allaboutcircuits.com/technical-articles/what-is-a-printed-circuit-board-pcb/>). AllAboutCircuits.com. Retrieved June 24, 2021.
2. Bhunia, Swarup; Tehranipoor, Mark (2019). "Chapter 2, section 2.6, A Quick Overview of Electronic Hardware". *Hardware Security*. p. 36. doi:10.1016/b978-0-12-812477-2.00007-1 (<https://doi.org/10.1016%2Fb978-0-12-812477-2.00007-1>). ISBN 978-0-12-812477-2.
3. "World PCB Production in 2014 Estimated at \$60.2B" (<http://www.ICONN07.com/index.php/article/92973/world-pcb-production-in-2014-estimated-at-602b/92976/?skin=pcb>). iCONNECT07. September 28, 2015. Retrieved April 12, 2016.
4. "PCB Market Size" (<https://www.mordorintelligence.com/industry-reports/printed-circuit-board-market>). *PCB Market Size, Share, Trends*. Retrieved December 22, 2024.
5. US 1256599 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=US1256599>), Schoop, Max Ulrich, "Process and mechanism for the production of electric heaters", published February 19, 1918
6. Harper, Charles A. (2003). *Electronic materials and processes handbook*. McGraw-Hill. pp. 7.3, 7.4. ISBN 0071402144.
7. Brunetti, Cleo (November 22, 1948). *New Advances in Printed Circuits* (<https://archive.org/details/newadvancesinpri192brun>). Washington, DC: National Bureau of Standards.
8. Engineers' Day, 1984 Award Recipients (<https://www.engr.wisc.edu/engage/alumni-relations/our-distinguished-alumni/>), College of Engineering, University of Wisconsin-Madison
9. "IEEE Cleo Brunetti Award Recipients" (https://web.archive.org/web/20180804200901/http://www.ieee.org/content/dam/ieee-org/ieee/web/org/about/awards/recipients/brunetti_rl.pdf) (PDF). IEEE. Archived from the original (https://www.ieee.org/content/dam/ieee-org/ieee/web/org/about/awards/recipients/brunetti_rl.pdf) (PDF) on August 4, 2018.

10. "New Process Perfected for Radio Wiring". *Chicago Tribune*. August 1, 1952.
11. "'Travel and Play with Motorola' advertisement". *Life*. May 24, 1954. p. 14.
12. "Topics & Trends of TV Trade." Television Digest 8:44 (November 1, 1952), 10.
13. US 2756485 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=US2756485>), Abramson, Moe & Danko, Stanislaus F., "Process of Assembling Electrical Circuits", published July 31, 1956, assigned to Secretary of the United States Army
14. US patent 5434751 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=US5434751>), Cole, Jr., Herbert S.; Sitnik-Nieters, Theresa A. & Wojnarowski, Robert J. et al., "Reworkable high density interconnect structure incorporating a release layer", issued July 18, 1995
15. Ostmann, Andreas; Schein, Friedrich-Leonhard; Dietterle, Michael; Kunz, Marc; Lang, Klaus-Dieter (2018). "High Density Interconnect Processes for Panel Level Packaging" (<http://ieeexplore.ieee.org/document/8546431>). 2018 7th Electronic System-Integration Technology Conference (ESTC). pp. 1–5. doi:10.1109/ESTC.2018.8546431 (<https://doi.org/10.1109%2FESTC.2018.8546431>). ISBN 978-1-5386-6814-6. S2CID 54214952 (<https://api.semanticscholar.org/CorpusID:54214952>).
16. *Materials for Advanced Packaging* (<https://books.google.com/books?id=4DmIDQAAQBAJ&dq=build+up+film+substrate&pg=PA319>). Springer. November 18, 2016. ISBN 978-3-319-45098-8.
17. "Advances in Optical Communications: Making optical printed circuit boards on an industrial scale" (<https://www.laserfocusworld.com/optics/article/14040276/making-optical-printed-circuit-boards-on-an-industrial-scale>). October 2019.
18. 6 Reasons Why Choose Printed Circuit Boards (<https://www.chipsmall.com/Blog/info/6-reasons-why-we-choose-printed-circuit-boards.html>)
19. IPC-14.38
20. Buschow, K.H., ed. (2001). "Electronic Packaging:Solder Mounting Technologies". *Encyclopedia of Materials: Science and Technology* (<https://books.google.com/books?id=TyVVAAMAAJ>). Elsevier. pp. 2708–9. ISBN 0-08-043152-6.
21. "Why Use High Density Interconnect?" (<https://www.freedomcad.com/2018/08/21/why-use-high-density-interconnect/>). August 21, 2018.
22. *IPC-D-275: Design Standard for Rigid Printed Boards and Rigid Printed Board Assemblies*. IPC. September 1991.
23. Sood, B. and Pecht, M. 2011. Printed Circuit Board Laminates. Wiley Encyclopedia of Composites. 1–11.
24. Lee W. Ritchey, Speeding Edge (November 1999). "A Survey and Tutorial of DIELECTRIC MATERIALS USED in the Manufacture of Printed Circuit Boards" (<http://speedingedge.com/PDF-Files/tutorial.pdf>) (PDF). *Circuitree Magazine*.
25. Fjelstad, Joseph. "Method for the Manufacture of an Aluminum Substrate PCB and its Advantages" (https://www.circuitinsight.com/pdf/manufacture_aluminum_substrate_pcbs_ipc.pdf) (PDF). *CircuitInsight.com*. Retrieved January 17, 2024.
26. Yung, Winco K. C. (2007). "Using Metal Core Printed Circuit Board (MCPCB) as a Solution for Thermal Management" (<https://www.bestpcbs.com/public/pdf/Metal-Core-PCB-design-guide.pdf>) (PDF). *Journal of the HKPCA* (24): 12–16.
27. "Applications | UBE Heat Resistant Polyimide Materials" (<http://www.upilex.jp/en/application.html>). *Upilex.jp*. UBE.
28. "Pyralux Flexible Circuit Materials" (<http://www.dupont.com/products-and-services/electronic-electrical-materials/flexible-rigidflex-circuit-materials/brands/pyralux-flexible-circuit.html>). DuPont.
29. Carter, Bruce (March 19, 2009). *Op Amps for Everyone* (<https://books.google.com/books?id=nnCNsjpicJIC&q=printed+circuit+board+materials+%22fr-1%22&pg=PA475>). Newnes. ISBN 9780080949482 – via Google Books.

30. "A High Performance, Economical RF/Microwave Substrate" (<http://www.microwavejournal.com/articles/2426-a-high-performance-economical-rf-microwave-substrate>). MicrowaveJournal.com. September 1, 1998. Retrieved November 4, 2024.
31. "RF-35 datasheet" (http://www.multi-circuit-boards.eu/fileadmin/user_upload/downloads/e_taconic_rf35-hf_www.multi-circuit-boards.eu.pdf) (PDF). Taconic – via Multi-CB.
32. "StackPath" (<https://www.electronicdesign.com/technologies/embedded-revolution/article/21799095/use-embedded-components-to-improve-pcb-performance-and-reduce-size>). February 11, 2014.
33. US 4175816 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=US4175816>), Burr, Robert P.; Morino, Ronald & Keogh, Raymond J., "Multi-wire electrical interconnecting member having a multi-wire matrix of insulated wires mechanically terminated thereon", published November 27, 1979, assigned to Kollmorgen Technologies Corp.
34. Weisberg, David E. (2008). "14: Intergraph" (https://www.vgamuseum.info/images/doc/intergraph/intergraph_history.pdf) (PDF). pp. 14–8.
35. Wagner, G. Donald (1999). "History of Electronic Packaging at APL: From the VT Fuze to the NEAR Spacecraft" (<https://web.archive.org/web/20170510111809/http://www.jhuapl.edu/techdigest/TD/td2001/Wagner.pdf>) (PDF). *Johns Hopkins APL Technical Digest*. **20** (1). Archived from the original (<http://www.jhuapl.edu/techdigest/TD/td2001/Wagner.pdf>) (PDF) on May 10, 2017.
36. Perdigones, Francisco; Quero, José Manuel (2022). "Printed Circuit Boards: The Layers' Functions for Electronic and Biomedical Engineering" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8952574>). *Micromachines*. **13** (3). MDPI: 460. doi:10.3390/mi13030460 (<https://doi.org/10.3390%2Fmi13030460>). PMC 8952574 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8952574>). PMID 35334752 (<https://pubmed.ncbi.nlm.nih.gov/35334752>).
37. Brown, Mark; Rawtani, Jawahar; Patil, Dinesh (2004). "Appendix B - Troubleshooting". *Practical Troubleshooting of Electrical Equipment and Control Circuits*. Elsevier. pp. 196–212. doi:10.1016/b978-075066278-9/50009-3 (<https://doi.org/10.1016%2Fb978-075066278-9%2F50009-3>). ISBN 978-0-7506-6278-9.
38. "EURLex – 02011L0065-20140129 – EN – EUR-Lex" (<http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1399998664957&uri=CELEX:02011L0065-20140129>). *Eur-lex.europa.eu*. Archived (<https://web.archive.org/web/20160107005501/http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1399998664957&uri=CELEX:02011L0065-20140129>) from the original on January 7, 2016. Retrieved July 3, 2015.
39. "The Restriction of the Use of Certain Hazardous Substances in Electrical and Electronic Equipment Regulations 2012" (<https://www.legislation.gov.uk/uksi/2012/3032/contents>). *legislation.gov.uk*. December 4, 2012. Retrieved March 31, 2022.
40. "How Does Lead Affect Our Environment?" (https://web.archive.org/web/20190419043337/https://www.michigan.gov/deq/0,4561,7-135-3307_29693_30031-90418--,00.html). *Pollutants and Toxicants: Environmental Lead (Pb)*. Department of Environmental Quality, State of Michigan. 2019. Archived from the original (https://www.michigan.gov/deq/0,4561,7-135-3307_29693_30031-90418--,00.html) on April 19, 2019.
41. "FAQ on RoHS Compliance" (<https://www.rohsguide.com/rohs-faq.htm>). *RoHS Compliance Guide*.

Further reading

- Tavernier, Karel (September 2015). "PCB Fabrication Data - Design to Fabrication Data Transfer" (http://www.ucamco.com/files/downloads/file/130/pcb_fabrication_a_guide_20150107.pdf) (PDF). V6. Archived (https://web.archive.org/web/20220308070109/https://www.ucamco.com/files/downloads/file_en/130/pcb-fabrication-data-a-guide_en.pdf?6f1ad339fb654d75d8b25bc426d2c3) (PDF) from the original on March 8, 2022. Retrieved May 9, 2022.

- Colotti, James (2022). "Analog, RF and EMC Considerations in Printed Wiring Board (PWB) Design" (<https://ieee.li/pdf/viewgraphs/analog-rf-emc-considerations-pwb-design.pdf>) (PDF). Revision 5. IEEE, Long Island Section. Archived (<https://web.archive.org/web/20220508092311/https://ieee.li/pdf/viewgraphs/analog-rf-emc-considerations-pwb-design.pdf>) (PDF) from the original on May 8, 2022. Retrieved May 9, 2022.

External links

-  Media related to [Printed circuit boards](#) at Wikimedia Commons

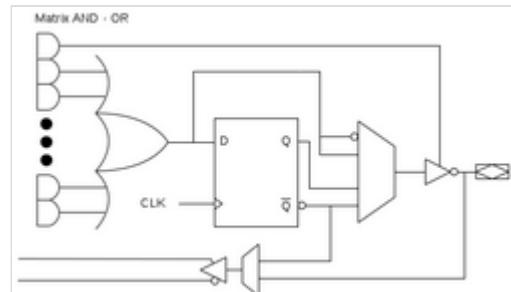
Retrieved from "https://en.wikipedia.org/w/index.php?title=Printed_circuit_board&oldid=1291738829"



Macrocell array

Macrocell arrays in PLDs

Programmable logic devices, such as programmable array logic and complex programmable logic devices, typically have a macrocell on every output pin.



GAL22V10 Output Logic Macrocell (OLMC)

Macrocell arrays in ASICs

A macrocell array is an approach to the design and manufacture of ASICs. Essentially, it is a small step up from the otherwise similar gate array, but rather than being a prefabricated array of simple logic gates, the macrocell array is a prefabricated array of higher-level logic functions such as flip-flops, ALU functions, registers, and the like. These logic functions are simply placed at regular predefined positions and manufactured on a wafer, usually called **master slice**. Creation of a circuit with a specified function is accomplished by adding metal interconnects to the chips on the master slice late in the manufacturing process, allowing the function of the chip to be customised as desired.

Macrocell array master slices are usually prefabricated and stockpiled in large quantities regardless of customer orders. The fabrication according to the individual customer specifications may be finished in a shorter time compared with standard cell or full custom design. The macrocell array approach reduces the mask costs since fewer custom masks need to be produced. In addition manufacturing test tooling lead time and costs are reduced since the same test fixtures may be used for all macrocell array products manufactured on the same die size.

Drawbacks are somewhat low density and performance than other approaches to ASIC design. However this style is often a viable approach for low production volumes.

A standard cell library is sometimes called a "macrocell library".^{[1][2]}

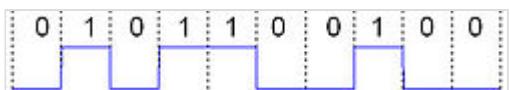
References

1. Norman Einspruch. "Application Specific Integrated Circuit (ASIC) Technology" (<https://book.s.google.com/books?id=YmGeGQrxmo8C&dq=asic+macrocell&pg=PA10>). Academic Press. 1991. p. 10.
2. "ASIC Macro Cells" (http://www.epson.jp/device/semicon_e/product/asic/macrocell/)



Digital signal

A **digital signal** is a signal that represents data as a sequence of discrete values; at any given time it can only take on, at most, one of a finite number of values.^{[1][2][3]} This contrasts with an analog signal, which represents continuous values; at any given time it represents a real number within a continuous range of values.



A binary signal, also known as a logic signal, is a digital signal with two distinguishable levels

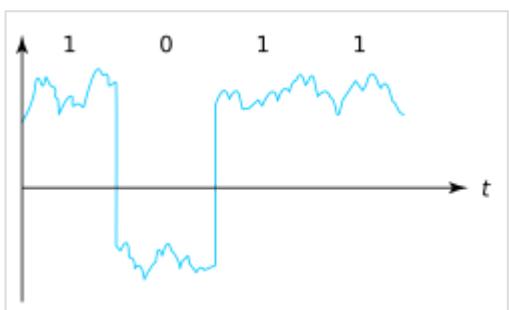
Simple digital signals represent information in discrete bands of levels. All levels within a band of values represent the same information state.^[1] In most digital circuits, the signal can have two possible valid values; this is called a **binary signal** or **logic signal**.^[4] They are represented by two voltage bands: one near a reference value (typically termed as *ground* or zero volts), and the other a value near the supply voltage. These correspond to the two values *zero* and *one* (or *false* and *true*) of the Boolean domain, so at any given time a binary signal represents one binary digit (bit). Because of this discretization, relatively small changes to the signal levels do not leave the discrete envelope, and as a result are ignored by signal state sensing circuitry. As a result, digital signals have noise immunity; electronic noise, provided it is not too great, will not affect digital circuits, whereas noise always degrades the operation of analog signals to some degree.^[5]

Digital signals having more than two states are occasionally used; circuitry using such signals is called multivalued logic. For example, signals that can assume three possible states are called three-valued logic.

In a digital signal, the physical quantity representing the information may be a variable electric current or voltage, the intensity, phase or polarization of an optical or other electromagnetic field, acoustic pressure, the magnetization of a magnetic storage media, etcetera. Digital signals are used in all digital electronics, notably computing equipment and data transmission.

Definitions

The term *digital signal* has related definitions in different contexts.

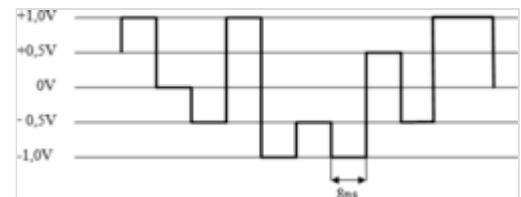


A received digital signal may be impaired by noise and distortions without necessarily affecting the digits

In digital electronics

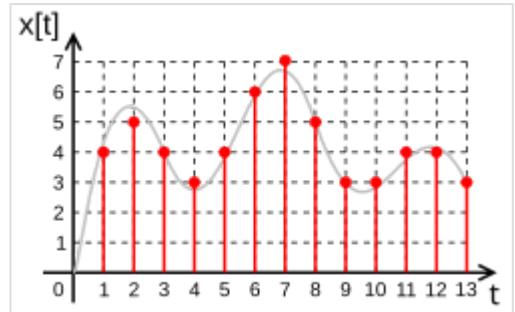
In digital electronics, a digital signal is a pulse amplitude modulated signal, i.e. a sequence of fixed-width electrical pulses or light pulses, each occupying one of a discrete number of levels of amplitude.^{[6][7]} A special case is a *logic signal* or a *binary signal*, which varies between a low and a high signal level.

The pulse trains in digital circuits are typically generated by metal–oxide–semiconductor field-effect transistor (MOSFET) devices, due to their rapid on–off electronic switching speed and large-scale integration (LSI) capability.^{[8][9]} In contrast, bipolar junction transistors more slowly generate signals resembling sine waves.^[8]



In signal processing

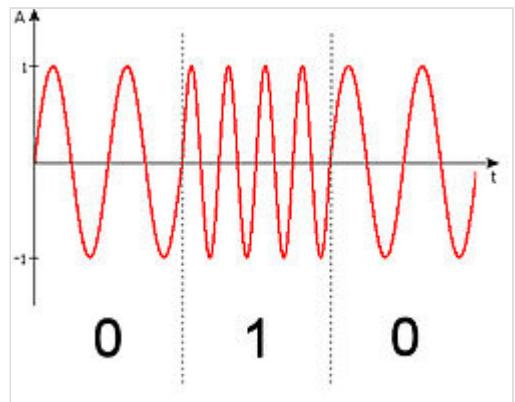
In digital signal processing, a digital signal is a representation of a physical signal that is sampled and quantized. A digital signal is an abstraction that is discrete in time and amplitude. The signal's value only exists at regular time intervals, since only the values of the corresponding physical signal at those sampled moments are significant for further digital processing. The digital signal is a sequence of codes drawn from a finite set of values.^[10] The digital signal may be stored, processed or transmitted physically as a pulse-code modulation (PCM) signal.



In signal processing, a digital signal is an abstraction that is discrete in time and amplitude, meaning it only exists at certain time instants.

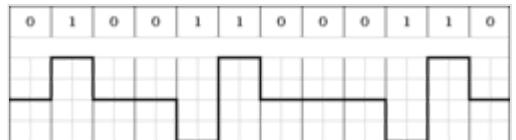
In communications

In digital communications, a digital signal is a continuous-time physical signal, alternating between a discrete number of waveforms,^[3] representing a bitstream. The shape of the waveform depends on the transmission scheme, which may be either a line coding scheme allowing baseband transmission; or a digital modulation scheme, allowing passband transmission over long wires or over a limited radio frequency band. Such a carrier-modulated sine wave is considered a digital signal in literature on digital communications and data transmission,^[11] but considered as a bit stream converted to an analog signal in specific cases where the signal will be carried over a system meant for analog communication, such as an analog telephone line.^[12]



A frequency-shift keying (FSK) signal is alternating between two waveforms and allows passband transmission. It is considered a means of digital data transmission.

In communications, sources of interference are usually present, and noise is frequently a significant problem. The effects of interference are typically minimized by filtering off interfering signals as much as possible and by using data redundancy. The main advantages of digital signals for communications are often considered to be noise immunity, and the ability, in many cases such as with audio and video data, to use data compression to greatly decrease the bandwidth that is required on the communication media.



An AMI coded digital signal used in baseband transmission (line coding)

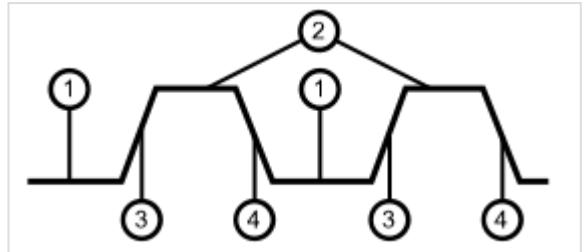
Logic voltage levels

A waveform that switches representing the two states of a Boolean value (0 and 1, or low and high, or false and true) is referred to as a *digital signal* or *logic signal* or *binary signal* when it is interpreted in terms of only two possible digits.

The two states are usually represented by some measurement of an electrical property: Voltage is the most common, but current is used in some logic families. Two ranges of voltages are typically defined for each logic family, which are frequently not directly adjacent. The signal is low when in the low range and high when in the high range, and in between the two ranges the behavior can vary between different types of gates.

The clock signal is a special digital signal that is used to synchronize many digital circuits. The image shown can be considered the waveform of a clock signal. Logic changes are triggered either by the rising edge or the falling edge. The rising edge is the transition from a low voltage (level 1 in the diagram) to a high voltage (level 2). The falling edge is the transition from a high voltage to a low one.

Although in a highly simplified and idealized model of a digital circuit, we may wish for these transitions to occur instantaneously, no real-world circuit is purely resistive and therefore no circuit can instantly change voltage levels. This means that during a short, finite transition time the output may not properly reflect the input, and will not correspond to either a logically high or low voltage.



A logic signal waveform: (1) low level, (2) high level, (3) rising edge, and (4) falling edge.

Modulation

To create a digital signal, a signal must be modulated with a control signal to produce it. The simplest modulation, a type of unipolar encoding, is simply to switch on and off a DC signal so that high voltages represent a '1' and low voltages are '0'.

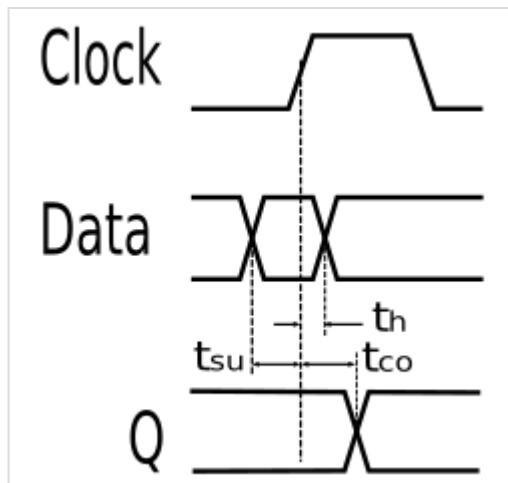
In digital radio schemes one or more carrier waves are amplitude, frequency or phase modulated by the control signal to produce a digital signal suitable for transmission.

Asymmetric Digital Subscriber Line (ADSL) over telephone wires, does not primarily use binary logic; the digital signals for individual carriers are modulated with different valued logics, depending on the Shannon capacity of the individual channel.

Clocking

Digital signals may be *samples* by a clock signal at regular intervals by passing the signal through a flip-flop. When this is done, the input is measured at the clock edge and the signal from that time. The signal is then held steady until the next clock. This process is the basis of synchronous logic.

Asynchronous logic also exists, which uses no single clock, and generally operates more quickly, and may use less power, but is significantly harder to design.



Clocking digital signals through a clocked flip-flop

See also

- Intersymbol interference

References

1. Robert K. Dueck (2005). *Digital Design with CPLD Applications and VHDL* (<https://books.google.com/books?id=1eO7kLWUmYIC&q=digital+signal+logic&pg=PA4>). Thomson/Delmar Learning. ISBN 1401840302. Archived (<https://web.archive.org/web/20171217234501/http://books.google.com/books?id=1eO7kLWUmYIC&lpg=PA49&ots=EBhxEwk7bY&dq=digital%20signal%20logic&pg=PA4#v=onepage&q=digital%20signal%20logic&f=false>) from the original on 2017-12-17. Retrieved 2017-08-30. "A digital representation can have only specific discrete values"
2. Proakis, John G.; Manolakis, Dimitris G. (2007-01-01). *Digital Signal Processing* (https://books.google.com/books?id=H_5SAAAAMAAJ). Pearson Prentice Hall. ISBN 9780131873742. Archived (https://web.archive.org/web/20160520061336/https://books.google.com/books?id=H_5SAAAAMAAJ) from the original on 2016-05-20. Retrieved 2015-09-22.
3. Analogue and Digital Communication Techniques (<https://books.google.com/books?id=WJ9qkNpScFAC&dq=%22a+digital+signal+is%22+communication&pg=PA3>) Archived (<https://web.archive.org/web/20171217234501/https://books.google.com/books?id=WJ9qkNpScFAC&pg=PA3&dq=%22a+digital+signal+is%22+communication&hl=sv&sa=X&ved=0CFoQ6AEwB2oVChMI-MD09ZjxwlVBAssCh1wnQu4#v=onepage&q=%22a%20digital%20signal%20is%22%20communication&f=false>) 2017-12-17 at the Wayback Machine: "A digital signal is a complex waveform and can be defined as a discrete waveform having a finite set of levels"
4. "Digital Signal" (<http://www.chegg.com/homework-help/definitions/digital-signal-4>). Archived (<https://web.archive.org/web/20160304120620/http://www.chegg.com/homework-help/definitions/digital-signal-4>) from the original on 2016-03-04. Retrieved 2016-08-13.

5. Horowitz, Paul; Hill, Winfield (1989). *The Art Of Electronics, 2nd Ed* (<https://books.google.com/books?id=xp2nuQAACAAJ&q=digital+analog+noise&pg=PA471>). Cambridge University Press. pp. 471–473. ISBN 0521370957.
6. B. SOMANATHAN NAIR (2002). *Digital electronics and logic design*. PHI Learning Pvt. Ltd. p. 289. ISBN 9788120319561. "Digital signals are fixed-width pulses, which occupy only one of two levels of amplitude."
7. Joseph Migga Kizza (2005). *Computer Network Security*. Springer Science & Business Media. ISBN 9780387204734.
8. "Applying MOSFETs to Today's Power-Switching Designs" (<https://www.electronicdesign.com/mosfets/applying-mosfets-today-s-power-switching-designs>). *Electronic Design*. 23 May 2016. Archived (<https://web.archive.org/web/20190810122607/https://www.electronicdesign.com/mosfets/applying-mosfets-today-s-power-switching-designs>) from the original on 10 August 2019. Retrieved 10 August 2019.
9. *2000 Solved Problems in Digital Electronics* (https://books.google.com/books?id=N6FDii6_nSEC&pg=PA151). Tata McGraw-Hill Education. 2005. p. 151. ISBN 978-0-07-058831-8.
10. Vinod Kumar Khanna (2009). *Digital Signal Processing* (<https://books.google.com/books?id=Vf2qXAbn58oC&q=%22a+digital+signal+is+a%22&pg=PA3>). S. Chand. p. 3. ISBN 9788121930956. "A digital signal is a special form of discrete-time signal which is discrete in both time and amplitude, obtained by permitting each value (sample) of a discrete-time signal to acquire a finite set of values (quantization), assigning it a numerical symbol according to a code ... A digital signal is a sequence or list of numbers drawn from a finite set."
11. J.S.Chitode, *Communication Systems* (<https://books.google.com/books?id=Ycq1cTYrLDkC&dq=%22a+digital+signal+is%22+CW&pg=PA585>), 2008: "When a digital signal is transmitted over a long distance, it needs CW modulation."
12. Fred Halsall, *Computer Networking and the Internet* (<https://books.google.com/books?id=QaDX5XErZ9IC&q=%22digital+signal%22>): "In order to transmit a digital signal over an analog subscriber line, modulated transmission must be used; that is the electrical signal that represents the binary bit stream of the source (digital) output must first be converted to an analog signal that is compatible with a (telephony) speech signal."

External links

- Monty Montgomery. *Digital Show & Tell* (<https://www.xiph.org/video/vid2.shtml>).
 - CodSim 2.0: Open source Virtual Laboratory for Digital Data Communications Model (<http://www.ac.uma.es/~guille/codsim2.0/>) Department of Computer Architecture, University of Malaga. Simulates Digital line encodings and Digital Modulations. Written in HTML for any web browser.
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Digital_signal&oldid=1286949068"



Boolean algebra

In mathematics and mathematical logic, Boolean algebra is a branch of algebra. It differs from elementary algebra in two ways. First, the values of the variables are the truth values *true* and *false*, usually denoted by 1 and 0, whereas in elementary algebra the values of the variables are numbers. Second, Boolean algebra uses logical operators such as conjunction (*and*) denoted as \wedge , disjunction (*or*) denoted as \vee , and negation (*not*) denoted as \neg . Elementary algebra, on the other hand, uses arithmetic operators such as addition, multiplication, subtraction, and division. Boolean algebra is therefore a formal way of describing logical operations in the same way that elementary algebra describes numerical operations.

Boolean algebra was introduced by George Boole in his first book *The Mathematical Analysis of Logic* (1847),^[1] and set forth more fully in his *An Investigation of the Laws of Thought* (1854).^[2] According to Huntington, the term Boolean algebra was first suggested by Henry M. Sheffer in 1913,^[3] although Charles Sanders Peirce gave the title "A Boolean [sic] Algebra with One Constant" to the first chapter of his "The Simplest Mathematics" in 1880.^[4] Boolean algebra has been fundamental in the development of digital electronics, and is provided for in all modern programming languages. It is also used in set theory and statistics.^[5]

History

A precursor of Boolean algebra was Gottfried Wilhelm Leibniz's algebra of concepts. The usage of binary in relation to the *I Ching* was central to Leibniz's *characteristica universalis*. It eventually created the foundations of algebra of concepts.^[6] Leibniz's algebra of concepts is deductively equivalent to the Boolean algebra of sets.^[7]

Boole's algebra predicated the modern developments in abstract algebra and mathematical logic; it is however seen as connected to the origins of both fields.^[8] In an abstract setting, Boolean algebra was perfected in the late 19th century by Jevons, Schröder, Huntington and others, until it reached the modern conception of an (abstract) mathematical structure.^[8] For example, the empirical observation that one can manipulate expressions in the algebra of sets, by translating them into expressions in Boole's algebra, is explained in modern terms by saying that the algebra of sets is a Boolean algebra (note the indefinite article). In fact, M. H. Stone proved in 1936 that every Boolean algebra is isomorphic to a field of sets.^{[9][10]}

In the 1930s, while studying switching circuits, Claude Shannon observed that one could also apply the rules of Boole's algebra in this setting,^[11] and he introduced switching algebra as a way to analyze and design circuits by algebraic means in terms of logic gates. Shannon already had at his disposal the abstract mathematical apparatus, thus he cast his switching algebra as the two-element Boolean algebra. In modern circuit engineering settings, there is little need to consider other Boolean algebras, thus "switching algebra" and "Boolean algebra" are often used interchangeably.^{[12][13][14]}

Efficient implementation of Boolean functions is a fundamental problem in the design of combinational logic circuits. Modern electronic design automation tools for very-large-scale integration (VLSI) circuits often rely on an efficient representation of Boolean functions known as (reduced ordered) binary decision diagrams (BDD) for logic synthesis and formal verification.^[15]

Logic sentences that can be expressed in classical propositional calculus have an equivalent expression in Boolean algebra. Thus, *Boolean logic* is sometimes used to denote propositional calculus performed in this way.^{[16][17][18]} Boolean algebra is not sufficient to capture logic formulas using quantifiers, like those from first-order logic.

Although the development of mathematical logic did not follow Boole's program, the connection between his algebra and logic was later put on firm ground in the setting of algebraic logic, which also studies the algebraic systems of many other logics.^[8] The problem of determining whether the variables of a given Boolean (propositional) formula can be assigned in such a way as to make the formula evaluate to true is called the Boolean satisfiability problem (SAT), and is of importance to theoretical computer science, being the first problem shown to be NP-complete. The closely related model of computation known as a Boolean circuit relates time complexity (of an algorithm) to circuit complexity.

Values

Whereas expressions denote mainly numbers in elementary algebra, in Boolean algebra, they denote the truth values *false* and *true*. These values are represented with the bits, 0 and 1. They do not behave like the integers 0 and 1, for which $1 + 1 = 2$, but may be identified with the elements of the two-element field GF(2), that is, integer arithmetic modulo 2, for which $1 + 1 = 0$. Addition and multiplication then play the Boolean roles of XOR (exclusive-or) and AND (conjunction), respectively, with disjunction $x \vee y$ (inclusive-or) definable as $x + y - xy$ and negation $\neg x$ as $1 - x$. In GF(2), \ominus may be replaced by $+$, since they denote the same operation; however, this way of writing Boolean operations allows applying the usual arithmetic operations of integers (this may be useful when using a programming language in which GF(2) is not implemented).

Boolean algebra also deals with functions which have their values in the set {0,1}. A sequence of bits is a commonly used example of such a function. Another common example is the totality of subsets of a set E : to a subset F of E , one can define the indicator function that takes the value 1 on F , and 0 outside F . The most general example is the set elements of a Boolean algebra, with all of the foregoing being instances thereof.

As with elementary algebra, the purely equational part of the theory may be developed, without considering explicit values for the variables.^[19]

Operations

Basic operations

While Elementary algebra has four operations (addition, subtraction, multiplication, and division), the Boolean algebra has only three basic operations: conjunction, disjunction, and negation, expressed with the corresponding binary operators AND (\wedge) and OR (\vee) and the unary operator NOT (\neg), collectively referred to as Boolean operators.^[20] Variables in Boolean algebra that store the logical value of 0 and 1 are called the Boolean variables. They are used to store either true or false values.^[21] The basic operations on Boolean variables x and y are defined as follows:

Logical operation	Operator	Notation	Alternative notations	Definition
Conjunction	AND	$x \wedge y$	$x \text{ AND } y, K_{xy}$	$x \wedge y = 1$ if $x = y = 1$, $x \wedge y = 0$ otherwise
Disjunction	OR	$x \vee y$	$x \text{ OR } y, A_{xy}$	$x \vee y = 0$ if $x = y = 0$, $x \vee y = 1$ otherwise
Negation	NOT	$\neg x$	NOT x , Nx , \bar{x} , x' , $!x$	$\neg x = 0$ if $x = 1$, $\neg x = 1$ if $x = 0$

Alternatively, the values of $x \wedge y$, $x \vee y$, and $\neg x$ can be expressed by tabulating their values with [truth tables](#) as follows:^[22]

x	y	$x \wedge y$	$x \vee y$
0	0	0	0
1	0	0	1
0	1	0	1
1	1	1	1

x	$\neg x$
0	1
1	0

When used in expressions, the operators are applied according to the precedence rules. As with elementary algebra, expressions in parentheses are evaluated first, following the precedence rules.^[23]

If the truth values 0 and 1 are interpreted as integers, these operations may be expressed with the ordinary operations of arithmetic (where $x + y$ uses addition and xy uses multiplication), or by the minimum/maximum functions:

$$\begin{aligned} x \wedge y &= xy = \min(x, y) \\ x \vee y &= x + y - xy = x + y(1 - x) = \max(x, y) \\ \neg x &= 1 - x \end{aligned}$$

One might consider that only negation and one of the two other operations are basic because of the following identities that allow one to define conjunction in terms of negation and the disjunction, and vice versa ([De Morgan's laws](#)):^[24]

$$\begin{aligned} x \wedge y &= \neg(\neg x \vee \neg y) \\ x \vee y &= \neg(\neg x \wedge \neg y) \end{aligned}$$

Secondary operations

Operations composed from the basic operations include, among others, the following:

Material conditional: $x \rightarrow y = \neg x \vee y$

Material biconditional: $x \leftrightarrow y = (x \wedge y) \vee (\neg x \wedge \neg y) = (x \vee \neg y) \wedge (\neg x \vee y)$

Exclusive OR (XOR): $x \oplus y = \neg(x \leftrightarrow y) = (x \vee y) \wedge \neg(x \wedge y) = (x \vee y) \wedge (\neg x \vee \neg y) = (x \wedge \neg y) \vee (\neg x \wedge y)$

These definitions give rise to the following truth tables giving the values of these operations for all four possible inputs.

Secondary operations. Table 1

x	y	$x \rightarrow y$	$x \oplus y$	$x \leftrightarrow y, x \equiv y$
0	0	1	0	1
1	0	0	1	0
0	1	1	1	0
1	1	1	0	1

Material conditional

The first operation, $x \rightarrow y$, or Cxy , is called *material implication*. If x is true, then the result of expression $x \rightarrow y$ is taken to be that of y (e.g. if x is true and y is false, then $x \rightarrow y$ is also false). But if x is false, then the value of y can be ignored; however, the operation must return some Boolean value and there are only two choices. So by definition, $x \rightarrow y$ is *true* when x is false ([relevance logic](#) rejects this definition, by viewing an implication with a false premise as something other than either true or false).

Exclusive OR (XOR)

The second operation, $x \oplus y$, or Jxy , is called *exclusive or* (often abbreviated as XOR) to distinguish it from disjunction as the inclusive kind. It excludes the possibility of both x and y *being* true (e.g. see table): if both are true then result is false. Defined in terms of arithmetic it is addition where mod 2 is $1 + 1 = 0$.

Logical equivalence

The third operation, the complement of exclusive or, is *equivalence* or Boolean equality: $x \equiv y$, or Exy , is true just when x and y have the same value. Hence $x \oplus y$ as its complement can be understood as $x \neq y$, being true just when x and y are different. Thus, its counterpart in arithmetic mod 2 is $x + y$. Equivalence's counterpart in arithmetic mod 2 is $x + y + 1$.

Laws

A *law* of Boolean algebra is an identity such as $x \vee (y \vee z) = (x \vee y) \vee z$ between two Boolean terms, where a *Boolean term* is defined as an expression built up from variables and the constants 0 and 1 using the operations \wedge , \vee , and \neg . The concept can be extended to terms involving other Boolean operations such as \oplus , \rightarrow , and \equiv , but such extensions are unnecessary for the purposes to which the laws are put. Such purposes include the definition of a Boolean algebra as any model of the Boolean laws, and as a means for deriving new laws from old as in the derivation of $x \vee (y \wedge z) = x \vee (z \wedge y)$ from $y \wedge z = z \wedge y$ (as treated in § [Axiomatizing Boolean algebra](#)).

Monotone laws

Boolean algebra satisfies many of the same laws as ordinary algebra when one matches up \vee with addition and \wedge with multiplication. In particular the following laws are common to both kinds of algebra:[\[25\]](#)[\[26\]](#)

Associativity of \vee :	$x \vee (y \vee z) = (x \vee y) \vee z$
Associativity of \wedge :	$x \wedge (y \wedge z) = (x \wedge y) \wedge z$
Commutativity of \vee :	$x \vee y = y \vee x$
Commutativity of \wedge :	$x \wedge y = y \wedge x$
Distributivity of \wedge over \vee :	$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$

Identity for \vee :	$x \vee 0 = x$
Identity for \wedge :	$x \wedge 1 = x$
Annihilator for \wedge :	$x \wedge 0 = 0$

The following laws hold in Boolean algebra, but not in ordinary algebra:

Annihilator for \vee :	$x \vee 1 = 1$
Idempotence of \vee :	$x \vee x = x$
Idempotence of \wedge :	$x \wedge x = x$
Absorption 1:	$x \wedge (x \vee y) = x$
Absorption 2:	$x \vee (x \wedge y) = x$
Distributivity of \vee over \wedge :	$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$

Taking $x = 2$ in the third law above shows that it is not an ordinary algebra law, since $2 \times 2 = 4$. The remaining five laws can be falsified in ordinary algebra by taking all variables to be 1. For example, in absorption law 1, the left hand side would be $1(1 + 1) = 2$, while the right hand side would be 1 (and so on).

All of the laws treated thus far have been for conjunction and disjunction. These operations have the property that changing either argument either leaves the output unchanged, or the output changes in the same way as the input. Equivalently, changing any variable from 0 to 1 never results in the output changing from 1 to 0. Operations with this property are said to be *monotone*. Thus the axioms thus far have all been for monotonic Boolean logic. Nonmonotonicity enters via complement \neg as follows.^[5]

Nonmonotone laws

The complement operation is defined by the following two laws.

Complementation 1	$x \wedge \neg x = 0$
Complementation 2	$x \vee \neg x = 1$

All properties of negation including the laws below follow from the above two laws alone.^[5]

In both ordinary and Boolean algebra, negation works by exchanging pairs of elements, hence in both algebras it satisfies the double negation law (also called involution law)

$$\text{Double negation} \quad \neg(\neg x) = x$$

But whereas *ordinary algebra* satisfies the two laws

$$\begin{aligned} (\neg x)(\neg y) &= xy \\ (\neg x) + (\neg y) &= -(x + y) \end{aligned}$$

Boolean algebra satisfies De Morgan's laws:

De Morgan 1	$\neg x \wedge \neg y = \neg(x \vee y)$
De Morgan 2	$\neg x \vee \neg y = \neg(x \wedge y)$

Completeness

The laws listed above define Boolean algebra, in the sense that they entail the rest of the subject. The laws *complementation* 1 and 2, together with the monotone laws, suffice for this purpose and can therefore be taken as one possible *complete* set of laws or *axiomatization* of Boolean algebra. Every law of Boolean algebra follows logically from these axioms. Furthermore, Boolean algebras can then be defined as the *models* of these axioms as treated in § *Boolean algebras*.

Writing down further laws of Boolean algebra cannot give rise to any new consequences of these axioms, nor can it rule out any model of them. In contrast, in a list of some but not all of the same laws, there could have been Boolean laws that did not follow from those on the list, and moreover there would have been models of the listed laws that were not Boolean algebras.

This axiomatization is by no means the only one, or even necessarily the most natural given that attention was not paid as to whether some of the axioms followed from others, but there was simply a choice to stop when enough laws had been noticed, treated further in § *Axiomatizing Boolean algebra*. Or the intermediate notion of axiom can be sidestepped altogether by defining a Boolean law directly as any *tautology*, understood as an equation that holds for all values of its variables over 0 and 1.^{[27][28]} All these definitions of Boolean algebra can be shown to be equivalent.

Duality principle

Principle: If $\{X, R\}$ is a partially ordered set, then $\{X, R(\text{inverse})\}$ is also a partially ordered set.

There is nothing special about the choice of symbols for the values of Boolean algebra. 0 and 1 could be renamed to α and β , and as long as it was done consistently throughout, it would still be Boolean algebra, albeit with some obvious cosmetic differences.

But suppose 0 and 1 were renamed 1 and 0 respectively. Then it would still be Boolean algebra, and moreover operating on the same values. However, it would not be identical to our original Boolean algebra because now \vee behaves the way \wedge used to do and vice versa. So there are still some cosmetic differences to show that the notation has been changed, despite the fact that 0s and 1s are still being used.

But if in addition to interchanging the names of the values, the names of the two binary operations are also interchanged, *now* there is no trace of what was done. The end product is completely indistinguishable from what was started with. The columns for $x \wedge y$ and $x \vee y$ in the truth tables have changed places, but that switch is immaterial.

When values and operations can be paired up in a way that leaves everything important unchanged when all pairs are switched simultaneously, the members of each pair are called *dual* to each other. Thus 0 and 1 are dual, and \wedge and \vee are dual. The *duality principle*, also called De Morgan duality, asserts that Boolean algebra is unchanged when all dual pairs are interchanged.

One change not needed to make as part of this interchange was to complement. Complement is a *self-dual* operation. The identity or do-nothing operation x (copy the input to the output) is also self-dual. A more complicated example of a self-dual operation is $(x \wedge y) \vee (y \wedge z) \vee (z \wedge x)$. There is no self-dual binary operation that depends on both its arguments. A composition of self-dual operations is a self-dual operation. For example, if $f(x, y, z) = (x \wedge y) \vee (y \wedge z) \vee (z \wedge x)$, then $f(f(x, y, z), x, t)$ is a self-dual operation of four arguments x, y, z, t .

The principle of duality can be explained from a group theory perspective by the fact that there are exactly four functions that are one-to-one mappings (automorphisms) of the set of Boolean polynomials back to itself: the identity function, the complement function, the dual function and the contradual function (complemented dual). These four functions form a group under function composition, isomorphic to the Klein four-group, acting on the set of Boolean polynomials. Walter Gottschalk remarked that consequently a more appropriate name for the phenomenon would be the *principle (or square) of quaternality*.^{[5]:21–22}

Diagrammatic representations

Venn diagrams

A Venn diagram^[29] can be used as a representation of a Boolean operation using shaded overlapping regions. There is one region for each variable, all circular in the examples here. The interior and exterior of region x corresponds respectively to the values 1 (true) and 0 (false) for variable x . The shading indicates the value of the operation for each combination of regions, with dark denoting 1 and light 0 (some authors use the opposite convention).

The three Venn diagrams in the figure below represent respectively conjunction $x \wedge y$, disjunction $x \vee y$, and complement $\neg x$.

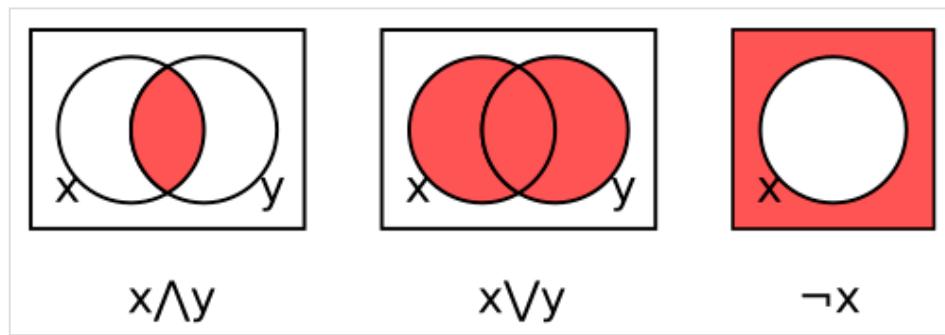


Figure 2. Venn diagrams for conjunction, disjunction, and complement

For conjunction, the region inside both circles is shaded to indicate that $x \wedge y$ is 1 when both variables are 1. The other regions are left unshaded to indicate that $x \wedge y$ is 0 for the other three combinations.

The second diagram represents disjunction $x \vee y$ by shading those regions that lie inside either or both circles. The third diagram represents complement $\neg x$ by shading the region *not* inside the circle.

While we have not shown the Venn diagrams for the constants 0 and 1, they are trivial, being respectively a white box and a dark box, neither one containing a circle. However, we could put a circle for x in those boxes, in which case each would denote a function of one argument, x , which returns the same value independently of x , called a constant function. As far as their outputs are concerned, constants and constant functions are indistinguishable; the difference is that a constant takes no arguments, called a *zeroary* or *nullary* operation, while a constant function takes one argument, which it ignores, and is a *unary* operation.

Venn diagrams are helpful in visualizing laws. The commutativity laws for \wedge and \vee can be seen from the symmetry of the diagrams: a binary operation that was not commutative would not have a symmetric diagram because interchanging x and y would have the effect of reflecting the diagram horizontally and any failure of commutativity would then appear as a failure of symmetry.

Idempotence of \wedge and \vee can be visualized by sliding the two circles together and noting that the shaded area then becomes the whole circle, for both \wedge and \vee .

To see the first absorption law, $x \wedge (x \vee y) = x$, start with the diagram in the middle for $x \vee y$ and note that the portion of the shaded area in common with the x circle is the whole of the x circle. For the second absorption law, $x \vee (x \wedge y) = x$, start with the left diagram for $x \wedge y$ and note that shading the whole of the x circle results in just the x circle being shaded, since the previous shading was inside the x circle.

The double negation law can be seen by complementing the shading in the third diagram for $\neg x$, which shades the x circle.

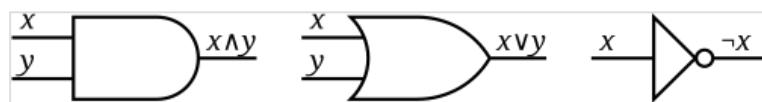
To visualize the first De Morgan's law, $(\neg x) \wedge (\neg y) = \neg(x \vee y)$, start with the middle diagram for $x \vee y$ and complement its shading so that only the region outside both circles is shaded, which is what the right hand side of the law describes. The result is the same as if we shaded that region which is both outside the x circle *and* outside the y circle, i.e. the conjunction of their exteriors, which is what the left hand side of the law describes.

The second De Morgan's law, $(\neg x) \vee (\neg y) = \neg(x \wedge y)$, works the same way with the two diagrams interchanged.

The first complement law, $x \wedge \neg x = 0$, says that the interior and exterior of the x circle have no overlap. The second complement law, $x \vee \neg x = 1$, says that everything is either inside or outside the x circle.

Digital logic gates

Digital logic is the application of the Boolean algebra of 0 and 1 to electronic hardware consisting of logic gates connected to form a circuit diagram. Each gate implements a Boolean operation, and is depicted schematically by a shape indicating the operation. The shapes associated with the gates for conjunction (AND-gates), disjunction (OR-gates), and complement (inverters) are as follows:^[30]

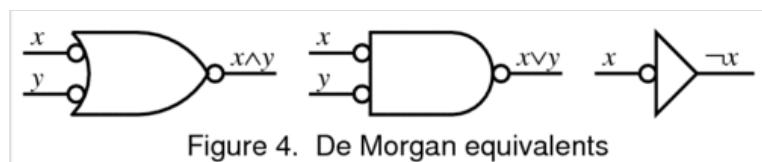


From left to right: AND, OR, and NOT gates.

The lines on the left of each gate represent input wires or *ports*. The value of the input is represented by a voltage on the lead. For so-called "active-high" logic, 0 is represented by a voltage close to zero or "ground," while 1 is represented by a voltage close to the supply voltage; active-low reverses this. The line on the right of each gate represents the output port, which normally follows the same voltage conventions as the input ports.

Complement is implemented with an inverter gate. The triangle denotes the operation that simply copies the input to the output; the small circle on the output denotes the actual inversion complementing the input. The convention of putting such a circle on any port means that the signal passing through this port is complemented on the way through, whether it is an input or output port.

The duality principle, or De Morgan's laws, can be understood as asserting that complementing all three ports of an AND gate converts it to an OR gate and vice versa, as shown in Figure 4 below. Complementing both ports of an inverter however leaves the operation unchanged.



More generally, one may complement any of the eight subsets of the three ports of either an AND or OR gate. The resulting sixteen possibilities give rise to only eight Boolean operations, namely those with an odd number of 1s in their truth table. There are eight such because the "odd-bit-out" can be either 0 or 1 and can go in any of four positions in the truth table. There being sixteen binary Boolean operations, this must leave eight operations with an even number of 1s in their truth tables. Two of these are the constants 0 and 1 (as binary operations that ignore both their inputs); four are the operations that depend nontrivially on exactly one of their two inputs, namely $x, y, \neg x$, and $\neg y$; and the remaining two are $x \oplus y$ (XOR) and its complement $x \equiv y$.

Boolean algebras

The term "algebra" denotes both a subject, namely the subject of algebra, and an object, namely an algebraic structure. Whereas the foregoing has addressed the subject of Boolean algebra, this section deals with mathematical objects called Boolean algebras, defined in full generality as any model of the Boolean laws. We begin with a special case of the notion definable without reference to the laws, namely concrete Boolean algebras, and then give the formal definition of the general notion.

Concrete Boolean algebras

A *concrete Boolean algebra* or *field of sets* is any nonempty set of subsets of a given set X closed under the set operations of union, intersection, and complement relative to X .^[5]

(Historically X itself was required to be nonempty as well to exclude the degenerate or one-element Boolean algebra, which is the one exception to the rule that all Boolean algebras satisfy the same equations since the degenerate algebra satisfies every equation. However, this exclusion conflicts with the preferred purely equational definition of "Boolean algebra", there being no way to rule out the one-element algebra using only equations— $0 \neq 1$ does not count, being a negated equation. Hence modern authors allow the degenerate Boolean algebra and let X be empty.)

Example 1. The power set 2^X of X , consisting of all subsets of X . Here X may be any set: empty, finite, infinite, or even uncountable.

Example 2. The empty set and X . This two-element algebra shows that a concrete Boolean algebra can be finite even when it consists of subsets of an infinite set. It can be seen that every field of subsets of X must contain the empty set and X . Hence no smaller example is possible, other than the degenerate algebra obtained by taking X to be empty so as to make the empty set and X coincide.

Example 3. The set of finite and cofinite sets of integers, where a cofinite set is one omitting only finitely many integers. This is clearly closed under complement, and is closed under union because the union of a cofinite set with any set is cofinite, while the union of two finite sets is finite. Intersection behaves like union with "finite" and "cofinite" interchanged. This example is countably infinite because there are only countably many finite sets of integers.

Example 4. For a less trivial example of the point made by example 2, consider a Venn diagram formed by n closed curves partitioning the diagram into 2^n regions, and let X be the (infinite) set of all points in the plane not on any curve but somewhere within the diagram. The interior of each region is thus an infinite subset of X , and every point in X is in exactly one region. Then the set of all 2^{2^n} possible unions of regions (including the empty set obtained as the union of the empty set of regions and X obtained as the union of all 2^n regions) is closed under union, intersection, and complement relative to X and therefore forms a concrete Boolean algebra. Again, there are finitely many subsets of an infinite set forming a concrete Boolean algebra, with example 2 arising as the case $n = 0$ of no curves.

Subsets as bit vectors

A subset Y of X can be identified with an indexed family of bits with index set X , with the bit indexed by $x \in X$ being 1 or 0 according to whether or not $x \in Y$. (This is the so-called characteristic function notion of a subset.) For example, a 32-bit computer word consists of 32 bits indexed by the set $\{0,1,2,\dots,31\}$, with 0 and 31 indexing the low and high order bits respectively. For a smaller example, if $X = \{a, b, c\}$ where a, b, c are viewed as bit positions in that order from left to right, the eight subsets $\{\}, \{c\}, \{b\}, \{b,c\}, \{a\}, \{a,c\}, \{a,b\}$, and $\{a,b,c\}$ of X can be identified with the respective bit vectors 000, 001, 010, 011, 100, 101, 110, and 111. Bit vectors indexed by the set of natural numbers are infinite sequences of bits, while those indexed by the reals in the unit interval $[0,1]$ are packed too densely to be able to write conventionally but nonetheless form well-defined indexed families (imagine coloring every point of the interval $[0,1]$ either black or white independently; the black points then form an arbitrary subset of $[0,1]$).

From this bit vector viewpoint, a concrete Boolean algebra can be defined equivalently as a nonempty set of bit vectors all of the same length (more generally, indexed by the same set) and closed under the bit vector operations of bitwise \wedge , \vee , and \neg , as in $1010 \wedge 0110 = 0010$, $1010 \vee 0110 = 1110$, and $\neg 1010 = 0101$, the bit vector realizations of intersection, union, and complement respectively.

Prototypical Boolean algebra

The set $\{0,1\}$ and its Boolean operations as treated above can be understood as the special case of bit vectors of length one, which by the identification of bit vectors with subsets can also be understood as the two subsets of a one-element set. This is called the *prototypical* Boolean algebra, justified by the following observation.

The laws satisfied by all nondegenerate concrete Boolean algebras coincide with those satisfied by the prototypical Boolean algebra.

This observation is proved as follows. Certainly any law satisfied by all concrete Boolean algebras is satisfied by the prototypical one since it is concrete. Conversely any law that fails for some concrete Boolean algebra must have failed at a particular bit position, in which case that position by itself furnishes a one-bit counterexample to that law. Nondegeneracy ensures the existence of at least one bit position because there is only one empty bit vector.

The final goal of the next section can be understood as eliminating "concrete" from the above observation. That goal is reached via the stronger observation that, up to isomorphism, all Boolean algebras are concrete.

Boolean algebras: the definition

The Boolean algebras so far have all been concrete, consisting of bit vectors or equivalently of subsets of some set. Such a Boolean algebra consists of a set and operations on that set which can be *shown* to satisfy the laws of Boolean algebra.

Instead of showing that the Boolean laws are satisfied, we can instead postulate a set X , two binary operations on X , and one unary operation, and *require* that those operations satisfy the laws of Boolean algebra. The elements of X need not be bit vectors or subsets but can be anything at all. This leads to the more general *abstract* definition.

A *Boolean algebra* is any set with binary operations \wedge and \vee and a unary operation \neg thereon satisfying the Boolean laws.^[31]

For the purposes of this definition it is irrelevant how the operations came to satisfy the laws, whether by fiat or proof. All concrete Boolean algebras satisfy the laws (by proof rather than fiat), whence every concrete Boolean algebra is a Boolean algebra according to our definitions. This axiomatic definition of a Boolean algebra as a set and certain operations satisfying certain laws or axioms *by fiat* is entirely analogous to the abstract definitions of group, ring, field etc. characteristic of modern or abstract algebra.

Given any complete axiomatization of Boolean algebra, such as the axioms for a complemented distributive lattice, a sufficient condition for an algebraic structure of this kind to satisfy all the Boolean laws is that it satisfy just those axioms. The following is therefore an equivalent definition.

A Boolean algebra is a complemented distributive lattice.

The section on axiomatization lists other axiomatizations, any of which can be made the basis of an equivalent definition.

Representable Boolean algebras

Although every concrete Boolean algebra is a Boolean algebra, not every Boolean algebra need be concrete. Let n be a square-free positive integer, one not divisible by the square of an integer, for example 30 but not 12. The operations of greatest common divisor, least common multiple, and division into n (that is, $\neg x = n/x$), can be shown to satisfy all the Boolean laws when their arguments range over the positive divisors of n . Hence those divisors form a Boolean algebra. These divisors are not subsets of a set, making the divisors of n a Boolean algebra that is not concrete according to our definitions.

However, if each divisor of n is *represented* by the set of its prime factors, this nonconcrete Boolean algebra is isomorphic to the concrete Boolean algebra consisting of all sets of prime factors of n , with union corresponding to least common multiple, intersection to greatest common divisor, and complement to division into n . So this example, while not technically concrete, is at least "morally" concrete via this representation, called an isomorphism. This example is an instance of the following notion.

A Boolean algebra is called *representable* when it is isomorphic to a concrete Boolean algebra.

The next question is answered positively as follows.

Every Boolean algebra is representable.

That is, up to isomorphism, abstract and concrete Boolean algebras are the same thing. This result depends on the Boolean prime ideal theorem, a choice principle slightly weaker than the axiom of choice. This strong relationship implies a weaker result strengthening the observation in the previous subsection to the following easy consequence of representability.

The laws satisfied by all Boolean algebras coincide with those satisfied by the prototypical Boolean algebra.

It is weaker in the sense that it does not of itself imply representability. Boolean algebras are special here, for example a relation algebra is a Boolean algebra with additional structure but it is not the case that every relation algebra is representable in the sense appropriate to relation algebras.

Axiomatizing Boolean algebra

The above definition of an abstract Boolean algebra as a set together with operations satisfying "the" Boolean laws raises the question of what those laws are. A simplistic answer is "all Boolean laws", which can be defined as all equations that hold for the Boolean algebra of 0 and 1. However, since there are infinitely many such laws, this is not a satisfactory answer in practice, leading to the question of it suffices to require only finitely many laws to hold.

In the case of Boolean algebras, the answer is "yes": the finitely many equations listed above are sufficient. Thus, Boolean algebra is said to be *finitely axiomatizable* or *finitely based*.

Moreover, the number of equations needed can be further reduced. To begin with, some of the above laws are implied by some of the others. A sufficient subset of the above laws consists of the pairs of associativity, commutativity, and absorption laws, distributivity of \wedge over \vee (or the other distributivity law—one suffices), and the two complement laws. In fact, this is the traditional axiomatization of Boolean algebra as a complemented distributive lattice.

By introducing additional laws not listed above, it becomes possible to shorten the list of needed equations yet further; for instance, with the vertical bar representing the Sheffer stroke operation, the single axiom $((a \mid b) \mid c) \mid (a \mid ((a \mid c) \mid a)) = c$ is sufficient to completely axiomatize Boolean algebra. It is also possible to find longer single axioms using more conventional operations; see Minimal axioms for Boolean algebra.^[32]

Propositional logic

Propositional logic is a logical system that is intimately connected to Boolean algebra.^[5] Many syntactic concepts of Boolean algebra carry over to propositional logic with only minor changes in notation and terminology, while the semantics of propositional logic are defined via Boolean algebras in a way that the tautologies (theorems) of propositional logic correspond to equational theorems of Boolean algebra.

Syntactically, every Boolean term corresponds to a propositional formula of propositional logic. In this translation between Boolean algebra and propositional logic, Boolean variables x, y, \dots become propositional variables (or *atoms*) P, Q, \dots Boolean terms such as $x \vee y$ become propositional formulas $P \vee Q$; 0 becomes *false* or \perp , and 1 becomes *true* or T . It is convenient when referring to generic propositions to use Greek letters Φ, Ψ, \dots as metavariables (variables outside the language of propositional calculus, used when talking *about* propositional calculus) to denote propositions.

The semantics of propositional logic rely on truth assignments. The essential idea of a truth assignment is that the propositional variables are mapped to elements of a fixed Boolean algebra, and then the truth value of a propositional formula using these letters is the element of the Boolean algebra that is obtained by computing the value of the Boolean term corresponding to the formula. In classical semantics, only the two-element Boolean algebra is used, while in Boolean-valued semantics arbitrary Boolean algebras are considered. A tautology is a propositional formula that is assigned truth value 1 by every truth assignment of its propositional variables to an arbitrary Boolean algebra (or, equivalently, every truth assignment to the two element Boolean algebra).

These semantics permit a translation between tautologies of propositional logic and equational theorems of Boolean algebra. Every tautology Φ of propositional logic can be expressed as the Boolean equation $\Phi = 1$, which will be a theorem of Boolean algebra. Conversely, every theorem $\Phi = \Psi$ of Boolean algebra corresponds

to the tautologies $(\Phi \vee \neg\Psi) \wedge (\neg\Phi \vee \Psi)$ and $(\Phi \wedge \Psi) \vee (\neg\Phi \wedge \neg\Psi)$. If \rightarrow is in the language, these last tautologies can also be written as $(\Phi \rightarrow \Psi) \wedge (\Psi \rightarrow \Phi)$, or as two separate theorems $\Phi \rightarrow \Psi$ and $\Psi \rightarrow \Phi$; if \equiv is available, then the single tautology $\Phi \equiv \Psi$ can be used.

Applications

One motivating application of propositional calculus is the analysis of propositions and deductive arguments in natural language.^[33] Whereas the proposition "if $x = 3$, then $x + 1 = 4$ " depends on the meanings of such symbols as + and 1, the proposition "if $x = 3$, then $x = 3$ " does not; it is true merely by virtue of its structure, and remains true whether " $x = 3$ " is replaced by " $x = 4$ " or "the moon is made of green cheese." The generic or abstract form of this tautology is "if P , then P ," or in the language of Boolean algebra, $P \rightarrow P$.

Replacing P by $x = 3$ or any other proposition is called *instantiation* of P by that proposition. The result of instantiating P in an abstract proposition is called an *instance* of the proposition. Thus, $x = 3 \rightarrow x = 3$ is a tautology by virtue of being an instance of the abstract tautology $P \rightarrow P$. All occurrences of the instantiated variable must be instantiated with the same proposition, to avoid such nonsense as $P \rightarrow x = 3$ or $x = 3 \rightarrow x = 4$.

Propositional calculus restricts attention to abstract propositions, those built up from propositional variables using Boolean operations. Instantiation is still possible within propositional calculus, but only by instantiating propositional variables by abstract propositions, such as instantiating Q by $Q \rightarrow P$ in $P \rightarrow (Q \rightarrow P)$ to yield the instance $P \rightarrow ((Q \rightarrow P) \rightarrow P)$.

(The availability of instantiation as part of the machinery of propositional calculus avoids the need for metavariables within the language of propositional calculus, since ordinary propositional variables can be considered within the language to denote arbitrary propositions. The metavariables themselves are outside the reach of instantiation, not being part of the language of propositional calculus but rather part of the same language for talking about it that this sentence is written in, where there is a need to be able to distinguish propositional variables and their instantiations as being distinct syntactic entities.)

Deductive systems for propositional logic

An axiomatization of propositional calculus is a set of tautologies called *axioms* and one or more inference rules for producing new tautologies from old. A *proof* in an axiom system A is a finite nonempty sequence of propositions each of which is either an instance of an axiom of A or follows by some rule of A from propositions appearing earlier in the proof (thereby disallowing circular reasoning). The last proposition is the *theorem* proved by the proof. Every nonempty initial segment of a proof is itself a proof, whence every proposition in a proof is itself a theorem. An axiomatization is *sound* when every theorem is a tautology, and *complete* when every tautology is a theorem.^[34]

Sequent calculus

Propositional calculus is commonly organized as a Hilbert system, whose operations are just those of Boolean algebra and whose theorems are Boolean tautologies, those Boolean terms equal to the Boolean constant 1. Another form is sequent calculus, which has two sorts, propositions as in ordinary propositional calculus, and pairs of lists of propositions called sequents, such as $A \vee B, A \wedge C, \dots \vdash A, B \rightarrow C, \dots$. The two halves of a sequent are called the antecedent and the succedent respectively. The customary metavariable denoting an antecedent or part thereof is Γ , and for a succedent Δ ; thus $\Gamma, A \vdash \Delta$ would denote a sequent whose succedent is a list Δ and whose antecedent is a list Γ with an additional proposition A appended after it. The antecedent is interpreted as the conjunction of its propositions, the succedent as the disjunction of its propositions, and the sequent itself as the entailment of the succedent by the antecedent.

Entailment differs from implication in that whereas the latter is a binary *operation* that returns a value in a Boolean algebra, the former is a binary *relation* which either holds or does not hold. In this sense, entailment is an *external* form of implication, meaning external to the Boolean algebra, thinking of the reader of the sequent as also being external and interpreting and comparing antecedents and succedents in some Boolean algebra. The natural interpretation of \vdash is as \leq in the partial order of the Boolean algebra defined by $x \leq y$ just when $x \vee y = y$. This ability to mix external implication \vdash and internal implication \rightarrow in the one logic is among the essential differences between sequent calculus and propositional calculus.^[35]

Applications

Boolean algebra as the calculus of two values is fundamental to computer circuits, computer programming, and mathematical logic, and is also used in other areas of mathematics such as set theory and statistics.^[5]

Computers

In the early 20th century, several electrical engineers intuitively recognized that Boolean algebra was analogous to the behavior of certain types of electrical circuits. Claude Shannon formally proved such behavior was logically equivalent to Boolean algebra in his 1937 master's thesis, *A Symbolic Analysis of Relay and Switching Circuits*.

Today, all modern general-purpose computers perform their functions using two-value Boolean logic; that is, their electrical circuits are a physical manifestation of two-value Boolean logic. They achieve this in various ways: as voltages on wires in high-speed circuits and capacitive storage devices, as orientations of a magnetic domain in ferromagnetic storage devices, as holes in punched cards or paper tape, and so on. (Some early computers used decimal circuits or mechanisms instead of two-valued logic circuits.)

Of course, it is possible to code more than two symbols in any given medium. For example, one might use respectively 0, 1, 2, and 3 volts to code a four-symbol alphabet on a wire, or holes of different sizes in a punched card. In practice, the tight constraints of high speed, small size, and low power combine to make noise a major factor. This makes it hard to distinguish between symbols when there are several possible symbols that could occur at a single site. Rather than attempting to distinguish between four voltages on one wire, digital designers have settled on two voltages per wire, high and low.

Computers use two-value Boolean circuits for the above reasons. The most common computer architectures use ordered sequences of Boolean values, called bits, of 32 or 64 values, e.g. 01101000110101100101010101001011. When programming in machine code, assembly language, and certain other programming languages, programmers work with the low-level digital structure of the data registers. These registers operate on voltages, where zero volts represents Boolean 0, and a reference voltage (often +5 V, +3.3 V, or +1.8 V) represents Boolean 1. Such languages support both numeric operations and logical operations. In this context, "numeric" means that the computer treats sequences of bits as binary numbers (base two numbers) and executes arithmetic operations like add, subtract, multiply, or divide. "Logical" refers to the Boolean logical operations of disjunction, conjunction, and negation between two sequences of bits, in which each bit in one sequence is simply compared to its counterpart in the other sequence. Programmers therefore have the option of working in and applying the rules of either numeric algebra or Boolean algebra as needed. A core differentiating feature between these families of operations is the existence of the carry operation in the first but not the second.

Two-valued logic

Other areas where two values is a good choice are the law and mathematics. In everyday relaxed conversation, nuanced or complex answers such as "maybe" or "only on the weekend" are acceptable. In more focused situations such as a court of law or theorem-based mathematics, however, it is deemed advantageous to frame questions so as to admit a simple yes-or-no answer—is the defendant guilty or not guilty, is the proposition true or false—and to disallow any other answer. However, limiting this might prove in practice for the respondent, the principle of the simple yes–no question has become a central feature of both judicial and mathematical logic, making two-valued logic deserving of organization and study in its own right.

A central concept of set theory is membership. An organization may permit multiple degrees of membership, such as novice, associate, and full. With sets, however, an element is either in or out. The candidates for membership in a set work just like the wires in a digital computer: each candidate is either a member or a nonmember, just as each wire is either high or low.

Algebra being a fundamental tool in any area amenable to mathematical treatment, these considerations combine to make the algebra of two values of fundamental importance to computer hardware, mathematical logic, and set theory.

Two-valued logic can be extended to multi-valued logic, notably by replacing the Boolean domain $\{0, 1\}$ with the unit interval $[0,1]$, in which case rather than only taking values 0 or 1, any value between and including 0 and 1 can be assumed. Algebraically, negation (NOT) is replaced with $1 - x$, conjunction (AND) is replaced with multiplication (xy), and disjunction (OR) is defined via De Morgan's law. Interpreting these values as logical truth values yields a multi-valued logic, which forms the basis for fuzzy logic and probabilistic logic. In these interpretations, a value is interpreted as the "degree" of truth – to what extent a proposition is true, or the probability that the proposition is true.

Boolean operations

The original application for Boolean operations was mathematical logic, where it combines the truth values, true or false, of individual formulas.

Natural language

Natural languages such as English have words for several Boolean operations, in particular conjunction (*and*), disjunction (*or*), negation (*not*), and implication (*implies*). *But not* is synonymous with *and not*. When used to combine situational assertions such as "the block is on the table" and "cats drink milk", which naïvely are either true or false, the meanings of these logical connectives often have the meaning of their logical counterparts. However, with descriptions of behavior such as "Jim walked through the door", one starts to notice differences such as failure of commutativity, for example, the conjunction of "Jim opened the door" with "Jim walked through the door" in that order is not equivalent to their conjunction in the other order, since *and* usually means *and then* in such cases. Questions can be similar: the order "Is the sky blue, and why is the sky blue?" makes more sense than the reverse order. Conjunctive commands about behavior are like behavioral assertions, as in *get dressed and go to school*. Disjunctive commands such *love me or leave me* or *fish or cut bait* tend to be asymmetric via the implication that one alternative is less preferable. Conjoined nouns such as *tea and milk* generally describe aggregation as with set union while *tea or milk* is a choice. However, context can reverse these senses, as in *your choices are coffee and tea* which usually means the same as *your choices are coffee or tea* (alternatives). Double negation, as in "I don't not like milk", rarely means literally "I do like milk" but rather conveys some sort of hedging, as though to imply that there is a third possibility. "Not not P" can be loosely

interpreted as "surely P", and although P necessarily implies "not not P ," the converse is suspect in English, much as with intuitionistic logic. In view of the highly idiosyncratic usage of conjunctions in natural languages, Boolean algebra cannot be considered a reliable framework for interpreting them.

Digital logic

Boolean operations are used in digital logic to combine the bits carried on individual wires, thereby interpreting them over $\{0,1\}$. When a vector of n identical binary gates are used to combine two bit vectors each of n bits, the individual bit operations can be understood collectively as a single operation on values from a Boolean algebra with 2^n elements.

Naive set theory

Naive set theory interprets Boolean operations as acting on subsets of a given set X . As we saw earlier this behavior exactly parallels the coordinate-wise combinations of bit vectors, with the union of two sets corresponding to the disjunction of two bit vectors and so on.

Video cards

The 256-element free Boolean algebra on three generators is deployed in computer displays based on raster graphics, which use bit blit to manipulate whole regions consisting of pixels, relying on Boolean operations to specify how the source region should be combined with the destination, typically with the help of a third region called the mask. Modern video cards offer all $2^{2^3} = 256$ ternary operations for this purpose, with the choice of operation being a one-byte (8-bit) parameter. The constants SRC = `0xaa` or `0b10101010`, DST = `0xcc` or `0b11001100`, and MSK = `0xf0` or `0b11110000` allow Boolean operations such as `(SRC^DST)&MSK` (meaning XOR the source and destination and then AND the result with the mask) to be written directly as a constant denoting a byte calculated at compile time, `0x80` in the `(SRC^DST)&MSK` example, `0x88` if just `SRC^DST`, etc. At run time the video card interprets the byte as the raster operation indicated by the original expression in a uniform way that requires remarkably little hardware and which takes time completely independent of the complexity of the expression.

Modeling and CAD

Solid modeling systems for computer aided design offer a variety of methods for building objects from other objects, combination by Boolean operations being one of them. In this method the space in which objects exist is understood as a set S of voxels (the three-dimensional analogue of pixels in two-dimensional graphics) and shapes are defined as subsets of S , allowing objects to be combined as sets via union, intersection, etc. One obvious use is in building a complex shape from simple shapes simply as the union of the latter. Another use is in sculpting understood as removal of material: any grinding, milling, routing, or drilling operation that can be performed with physical machinery on physical materials can be simulated on the computer with the Boolean operation $x \setminus y$ or $x - y$, which in set theory is set difference, remove the elements of y from those of x . Thus given two shapes one to be machined and the other the material to be removed, the result of machining the former to remove the latter is described simply as their set difference.

Boolean searches

Search engine queries also employ Boolean logic. For this application, each web page on the Internet may be considered to be an "element" of a "set." The following examples use a syntax supported by Google.^[NB 1]

- Doublequotes are used to combine whitespace-separated words into a single search term.^[NB 2]
- Whitespace is used to specify logical AND, as it is the default operator for joining search terms:

"Search term 1" "Search term 2"

- The OR keyword is used for logical OR:

"Search term 1" OR "Search term 2"

- A prefixed minus sign is used for logical NOT:

"Search term 1" -"Search term 2"

See also



- [Boolean algebras canonically defined](#)
- [Boolean differential calculus](#)
- [Booleo](#)
- [Cantor algebra](#)
- [Heyting algebra](#)
- [List of Boolean algebra topics](#)
- [Logic design](#)
- [Principia Mathematica](#)
- [Three-valued logic](#)
- [Vector logic](#)

Notes

1. Not all search engines support the same query syntax. Additionally, some organizations (such as Google) provide "specialized" search engines that support alternate or extended syntax. (See, [Syntax cheatsheet](#) (<https://www.google.com/help/cheatsheet.html>).) The now-defunct Google code search used to support regular expressions but no longer exists.
2. Doublequote-delimited search terms are called "exact phrase" searches in the Google documentation.

References

1. Boole, George (2011-07-28). *The Mathematical Analysis of Logic - Being an Essay Towards a Calculus of Deductive Reasoning* (<https://www.gutenberg.org/ebooks/36884>).
2. Boole, George (2003) [1854]. *An Investigation of the Laws of Thought*. Prometheus Books. ISBN 978-1-59102-089-9.
3. "The name Boolean algebra (or Boolean 'algebras') for the calculus originated by Boole, extended by Schröder, and perfected by Whitehead seems to have been first suggested by Sheffer, in 1913." Edward Vermilye Huntington, "New sets of independent postulates for the algebra of logic, with special reference to Whitehead and Russell's *Principia mathematica* (<https://www.ams.org/journals/tran/1933-035-01/S0002-9947-1933-1501684-X/S0002-9947-1933-1501684-X.pdf>)", in *Transactions of the American Mathematical Society* 35 (1933), 274-304; footnote, page 278.
4. Peirce, Charles S. (1931). *Collected Papers* (<https://books.google.com/books?id=3JJgOkGmnjEC&pg=RA1-PA13>). Vol. 3. Harvard University Press. p. 13. ISBN 978-0-674-13801-8.

5. Givant, Steven R.; Halmos, Paul Richard (2009). *Introduction to Boolean Algebras* (<https://books.google.com/books?id=ORILyf8sF2sC&pg=PA22>). Undergraduate Texts in Mathematics, Springer. pp. 21–22. ISBN 978-0-387-40293-2.
6. Nelson, Eric S. (2011). "The Yijing and Philosophy: From Leibniz to Derrida" (<https://philpapers.org/rec/NELTYA>). *Journal of Chinese Philosophy*. 38 (3): 377–396. doi:10.1111/j.1540-6253.2011.01661.x (<https://doi.org/10.1111%2Fj.1540-6253.2011.01661.x>).
7. Lenzen, Wolfgang. "Leibniz: Logic" (<http://www.iep.utm.edu/leib-log>). *Internet Encyclopedia of Philosophy*.
8. Dunn, J. Michael; Hardegree, Gary M. (2001). *Algebraic methods in philosophical logic* (<https://books.google.com/books?id=-AokWhbILUIC&pg=PA2>). Oxford University Press. p. 2. ISBN 978-0-19-853192-0.
9. Bimbó, Katalin; Dunn, J. Michael (2008). *Generalized Galois Logics: Relational Semantics of Nonclassical Logical Calculi* (https://www.google.com/books/edition/Generalized_Galois_Logics/Vu3uAAAAMAAJ?hl=en&gbpv=1&bsq=mh+stone+boolean+algebra+iso+proved&dq=mh+stone+boolean+algebra+iso+proved&printsec=frontcover). CSLI Publications. ISBN 978-1-57586-573-7.
10. Stone, M. H. (1936). "The Theory of Representation for Boolean Algebras" (<https://www.jstor.org/stable/1989664>). *Transactions of the American Mathematical Society*. 40 (1): 37–111. doi:10.2307/1989664 (<https://doi.org/10.2307%2F1989664>). ISSN 0002-9947 (<https://search.worldcat.org/issn/0002-9947>).
11. Weisstein, Eric W. "Boolean Algebra" (<https://mathworld.wolfram.com/BooleanAlgebra.html>). *mathworld.wolfram.com*. Retrieved 2020-09-02.
12. Balabanian, Norman; Carlson, Bradley (2001). *Digital logic design principles*. John Wiley. pp. 39–40. ISBN 978-0-471-29351-4., online sample (<http://www.wiley.com/college/engin/balabanian293512/pdf/ch02.pdf>)
13. Rajaraman; Radhakrishnan (2008-03-01). *Introduction To Digital Computer Design* (<https://books.google.com/books?id=-8MvcOgsSjcC&pg=PA65>). PHI Learning Pvt. Ltd. p. 65. ISBN 978-81-203-3409-0.
14. Camara, John A. (2010). *Electrical and Electronics Reference Manual for the Electrical and Computer PE Exam* (<https://books.google.com/books?id=rfHWHeU0jfsC&pg=SA41-PA3>). www.ppi2pass.com. p. 41. ISBN 978-1-59126-166-7.
15. Shin-ichi Minato, Saburo Muroga (2007). "Chapter 29: Binary Decision Diagrams". In Chen, Wai-Kai (ed.). *The VLSI handbook* (2 ed.). CRC Press. ISBN 978-0-8493-4199-1.
16. Parkes, Alan (2002). *Introduction to languages, machines and logic: computable languages, abstract machines and formal logic* (<https://books.google.com/books?id=sUQXKy8KPcQC&pg=PA276>). Springer. p. 276. ISBN 978-1-85233-464-2.
17. Barwise, Jon; Etchemendy, John; Allwein, Gerard; Barker-Plummer, Dave; Liu, Albert (1999). *Language, proof, and logic* (<https://archive.org/details/languageprooflog00barw>). CSLI Publications. ISBN 978-1-889119-08-3.
18. Goertzel, Ben (1994). *Chaotic logic: language, thought, and reality from the perspective of complex systems science* (<https://books.google.com/books?id=zVOWoXDunp8C&pg=PA48>). Springer. p. 48. ISBN 978-0-306-44690-0.
19. Halmos, Paul Richard (1963). *Lectures on Boolean Algebras*. van Nostrand.
20. Bacon, Jason W. (2011). "Computer Science 315 Lecture Notes" (<https://web.archive.org/web/20211002001321/http://www.cs.uwm.edu/classes/cs315/Bacon/Lecture/HTML/ch13s02.html>). Archived from the original (<http://www.cs.uwm.edu/classes/cs315/Bacon/Lecture/HTML/ch13s02.html>) on 2021-10-02. Retrieved 2021-10-01.
21. "Boolean Algebra - Expression, Rules, Theorems, and Examples" (<https://www.geeksforgeeks.org/boolean-algebra/>). GeeksforGeeks. 2021-09-24. Retrieved 2024-06-03.
22. "Boolean Logical Operations" (https://www.uobabylon.edu.iq/eprints/publication_12_3018_1456.pdf) (PDF).
23. "Boolean Algebra Operations" (<https://bob.cs.sonoma.edu/IntroCompOrg-RPi/sec-balgebra.html>). bob.cs.sonoma.edu. Retrieved 2024-06-03.
24. "Boolean Algebra" (<https://www2.seas.gwu.edu/~ayoussef/cs1311/BooleanAlgebra.pdf>) (PDF).

25. O'Regan, Gerard (2008). *A brief history of computing* (<https://books.google.com/books?id=081H96F1enMC&pg=PA33>). Springer. p. 33. ISBN 978-1-84800-083-4.
26. "Elements of Boolean Algebra" (<http://www.ee.surrey.ac.uk/Projects/Labview/boolalgebra/>). www.ee.surrey.ac.uk. Retrieved 2020-09-02.
27. McGee, Vann, *Sentential Calculus Revisited: Boolean Algebra* (<https://ocw.mit.edu/courses/lingistics-and-philosophy/24-241-logic-i-fall-2005/readings/chp24.pdf>) (PDF)
28. Goodstein, Reuben Louis (2012), "Chapter 4: Sentence Logic", *Boolean Algebra*, Courier Dover Publications, ISBN 978-0-48615497-8
29. Venn, John (July 1880). "I. On the Diagrammatic and Mechanical Representation of Propositions and Reasonings" (https://www.cis.upenn.edu/~bhusnur4/cit592_fall2014/venn%20diagrams.pdf) (PDF). *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 5. **10** (59): 1–18. doi:10.1080/14786448008626877 (<https://doi.org/10.1080%2F14786448008626877>). Archived (https://web.archive.org/web/20170516204620/https://www.cis.upenn.edu/~bhusnur4/cit592_fall2014/venn%20diagrams.pdf) (PDF) from the original on 2017-05-16. [1] (<http://www.tandfonline.com/doi/abs/10.1080/14786448008626877>) [2] (<https://books.google.com/books?id=k68vAQAAIAJ&pg=PA1>)
30. Shannon, Claude (1949). "The Synthesis of Two-Terminal Switching Circuits". *Bell System Technical Journal*. **28**: 59–98. doi:10.1002/j.1538-7305.1949.tb03624.x (<https://doi.org/10.1002%2Fj.1538-7305.1949.tb03624.x>).
31. Koppelberg, Sabine (1989). "General Theory of Boolean Algebras". *Handbook of Boolean Algebras*, Vol. 1 (ed. J. Donald Monk with Robert Bonnet). Amsterdam, Netherlands: North Holland. ISBN 978-0-444-70261-6.
32. McCune, William; Veroff, Robert; Fitelson, Branden; Harris, Kenneth; Feist, Andrew; Wos, Larry (2002), "Short single axioms for Boolean algebra", *Journal of Automated Reasoning*, **29** (1): 1–16, doi:10.1023/A:1020542009983 (<https://doi.org/10.1023%2FA%3A1020542009983>), MR 1940227 (<https://mathscinet.ams.org/mathscinet-getitem?mr=1940227>), S2CID 207582048 (<https://api.semanticscholar.org/CorpusID:207582048>)
33. Allwood, Jens; Andersson, Gunnar-Gunnar; Andersson, Lars-Gunnar; Dahl, Osten (1977-09-15). *Logic in Linguistics* (<https://books.google.com/books?id=hXIpFPttDjgC&q=%22propositional+logic%22>). Cambridge University Press. ISBN 978-0-521-29174-3.
34. Hausman, Alan; Kahane, Howard; Tidman, Paul (2010) [2007]. *Logic and Philosophy: A Modern Introduction*. Wadsworth Cengage Learning. ISBN 978-0-495-60158-6.
35. Girard, Jean-Yves; Taylor, Paul; Lafont, Yves (1990) [1989]. *Proofs and Types* (<https://archive.org/details/proofstypes0000gira>). Cambridge University Press (Cambridge Tracts in Theoretical Computer Science, 7). ISBN 978-0-521-37181-0.

Further reading

- Mano, Morris; Ciletti, Michael D. (2013). *Digital Design*. Pearson. ISBN 978-0-13-277420-8.
- Whitesitt, J. Eldon (1995). *Boolean algebra and its applications*. Courier Dover Publications. ISBN 978-0-486-68483-3.
- Dwinger, Philip (1971). *Introduction to Boolean algebras*. Würzburg, Germany: Physica Verlag.
- Sikorski, Roman (1969). *Boolean Algebras* (3 ed.). Berlin, Germany: Springer-Verlag. ISBN 978-0-387-04469-9.
- Bocheński, Józef Maria (1959). *A Précis of Mathematical Logic*. Translated from the French and German editions by Otto Bird. Dordrecht, South Holland: D. Reidel.

Historical perspective

- Boole, George (1848). "The Calculus of Logic" (<http://www.maths.tcd.ie/pub/HistMath/People/Boole/CalcLogic/CalcLogic.html>). *Cambridge and Dublin Mathematical Journal*. III: 183–198.
- Hailperin, Theodore (1986). *Boole's logic and probability: a critical exposition from the standpoint*

- of contemporary algebra, logic, and probability theory* (2 ed.). Elsevier. ISBN 978-0-444-87952-3.
- Gabbay, Dov M.; Woods, John, eds. (2004). *The rise of modern logic: from Leibniz to Frege*. Handbook of the History of Logic. Vol. 3. Elsevier. ISBN 978-0-444-51611-4., several relevant chapters by Hailperin, Valencia, and Grattan-Guinness
 - Badesa, Calixto (2004). "Chapter 1. Algebra of Classes and Propositional Calculus". *The birth of model theory: Löwenheim's theorem in the frame of the theory of relatives*. Princeton University Press. ISBN 978-0-691-05853-5.
 - Stanković, Radomir S. [in German]; Astola, Jaakko Tapio [in Finnish] (2011). Written at Niš, Serbia & Tampere, Finland. *From Boolean Logic to Switching Circuits and Automata: Towards Modern Information Technology* (<https://books.google.com/books?id=ZuHwCAAAQBAJ>). Studies in Computational Intelligence. Vol. 335 (1 ed.). Berlin & Heidelberg, Germany: Springer-Verlag. pp. xviii + 212. doi:10.1007/978-3-642-11682-7 (<https://doi.org/10.1007%2F978-3-642-11682-7>). ISBN 978-3-642-11681-0. ISSN 1860-949X (<https://search.worldcat.org/issn/1860-949X>). LCCN 2011921126 (<https://lccn.loc.gov/2011921126>). Retrieved 2022-10-25.
 - "The Algebra of Logic Tradition" (<https://plato.stanford.edu/entries/algebra-logic-tradition/>) entry by Burris, Stanley in the *Stanford Encyclopedia of Philosophy*, 21 February 2012

External links

Retrieved from "https://en.wikipedia.org/w/index.php?title=Boolean_algebra&oldid=1286839413"



Logic synthesis

In computer engineering, **logic synthesis** is a process by which an abstract specification of desired circuit behavior, typically at register transfer level (RTL), is turned into a design implementation in terms of logic gates, typically by a computer program called a *synthesis tool*. Common examples of this process include synthesis of designs specified in hardware description languages, including VHDL and Verilog.^[1] Some synthesis tools generate bitstreams for programmable logic devices such as PALs or FPGAs, while others target the creation of ASICs. Logic synthesis is one step in circuit design in the electronic design automation, the others are place and route and verification and validation.

History

The roots of logic synthesis can be traced to the treatment of logic by George Boole (1815 to 1864), in what is now termed Boolean algebra. In 1938, Claude Shannon showed that the two-valued Boolean algebra can describe the operation of switching circuits. In the early days, logic design involved manipulating the truth table representations as Karnaugh maps. The Karnaugh map-based minimization of logic is guided by a set of rules on how entries in the maps can be combined. A human designer can typically only work with Karnaugh maps containing up to four to six variables.

The first step toward automation of logic minimization was the introduction of the Quine–McCluskey algorithm that could be implemented on a computer. This exact minimization technique presented the notion of prime implicants and minimum cost covers that would become the cornerstone of two-level minimization. Nowadays, the much more efficient Espresso heuristic logic minimizer has become the standard tool for this operation. Another area of early research was in state minimization and encoding of finite-state machines (FSMs), a task that was the bane of designers. The applications for logic synthesis lay primarily in digital computer design. Hence, IBM and Bell Labs played a pivotal role in the early automation of logic synthesis. The evolution from discrete logic components to programmable logic arrays (PLAs) hastened the need for efficient two-level minimization, since minimizing terms in a two-level representation reduces the area in a PLA.

Two-level logic circuits are of limited importance in a very-large-scale integration (VLSI) design; most designs use multiple levels of logic. Almost any circuit representation in RTL or Behavioural Description is a multi-level representation. An early system that was used to design multilevel circuits was LSS from IBM. It used local transformations to simplify logic. Work on LSS and the Yorktown Silicon Compiler spurred rapid research progress in logic synthesis in the 1980s. Several universities contributed by making their research available to the public, most notably SIS from University of California, Berkeley, RASP from University of California, Los Angeles and BOLD from University of Colorado, Boulder. Within a decade, the technology migrated to commercial logic synthesis products offered by electronic design automation companies.

Commercial tools

The leading developers and providers of logic synthesis software packages are [Synopsys](#), [Cadence](#), and [Siemens](#). Their synthesis tools are Synopsys Design Compiler, Cadence First Encounter and Siemens Precision RTL.

Logic elements

Logic design is a step in the standard design cycle in which the functional design of an electronic circuit is converted into the representation which captures logic operations, arithmetic operations, control flow, etc. A common output of this step is RTL description. Logic design is commonly followed by the circuit design step. In modern electronic design automation parts of the logical design may be automated using high-level synthesis tools based on the behavioral description of the circuit.^[2]

Logic operations usually consist of Boolean AND, OR, XOR and NAND operations, and are the most basic forms of operations in an electronic circuit. Arithmetic operations are usually implemented with the use of logic operators.

High-level or behavioral

With a goal of increasing designer productivity, research efforts on the synthesis of circuits specified at the behavioral level have led to the emergence of commercial

solutions in 2004,^[3] which are used for complex ASIC and FPGA design. These tools automatically synthesize circuits specified using high-level languages, like ANSI C/C++ or SystemC, to a register transfer level (RTL) specification, which can be used as input to a gate-level logic synthesis flow.^[3] Using high-level synthesis, also known as ESL synthesis, the allocation of work to clock cycles and across structural components, such as floating-point ALUs, is done by the compiler using an optimisation procedure, whereas with RTL logic synthesis (even from behavioural Verilog or VHDL, where a thread of execution can make multiple reads and writes to a variable within a clock cycle) those allocation decisions have already been made.

	x	y	
\wedge	0	1	
x	0	0	0
	1	0	1

	x	y	
\vee	0	1	
x	1	1	1
	1	0	1

	x	y	
\rightarrow	0	1	1
x	1	0	1
	1	0	1

	x	y	
\oplus	0	1	
x	1	1	0
	1	1	0

Figure 1. Truth tables



Figure 2. Logic gates

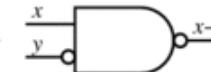
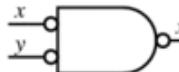


Figure 3. De Morgan equivalents

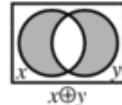
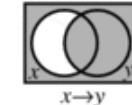
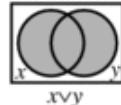
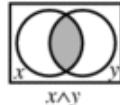


Figure 4. Venn diagrams

Various representations of Boolean operations

Multi-level logic minimization

Typical practical implementations of a logic function utilize a multi-level network of logic elements. Starting from an RTL description of a design, the synthesis tool constructs a corresponding multilevel Boolean network.

Next, this network is optimized using several technology-independent techniques before technology-dependent optimizations are performed. The typical cost function during technology-independent optimizations is total literal count of the factored representation of the logic function (which correlates quite well with circuit area).

Finally, technology-dependent optimization transforms the technology-independent circuit into a network of gates in a given technology. The simple cost estimates are replaced by more concrete, implementation-driven estimates during and after technology mapping. Mapping is constrained by factors such as the available gates (logic functions) in the technology library, the drive sizes for each gate, and the delay, power, and area characteristics of each gate.

See also

- [Silicon compiler](#)
- [Binary decision diagram](#)
- [Functional verification](#)
- [Boolean differential calculus](#)
- [Synthesis of Integral Design](#) by DEC, a 1980s tool used to design VAX 9000 mainframe CPUs and others ICs

References

1. "Synthesis:Verilog to Gates" (https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-884-complex-digital-systems-spring-2005/lecture-notes/l05_synthesis.pdf) (PDF).
 2. Naveed A. Sherwani (1999). *Algorithms for VLSI physical design automation* (3rd ed.). Kluwer Academic Publishers. p. 4. ISBN 978-0-7923-8393-2.
 3. EETimes: [High-level synthesis rollouts enable ESL](http://archives.eetimes.com/high-level-synthesis-rollouts-enable-esl/110436.html) (<http://archives.eetimes.com/high-level-synthesis-rollouts-enable-esl/110436.html>)
- *Electronic Design Automation For Integrated Circuits Handbook*, by Lavagno, Martin, and Scheffer, ISBN 0-8493-3096-3 A survey of the field of Electronic design automation. The above summary was derived, with permission, from Volume 2, Chapter 2, *Logic Synthesis* by Sunil Khatri and Narendra Shenoy.

Further reading

- Burgun, Luc; Greiner, Alain; Prado Lopes Eudes (October 1994). "A Consistent Approach in Logic Synthesis for FPGA Architectures". *Proceedings of the International Conference on ASIC*. Pekin: 104–107.

- Jiang, Jie-Hong "Roland"; Devadas, Srinivas (2009). "Chapter 6: Logic synthesis in a nutshell". In Wang, Laung-Terng; Chang, Yao-Wen; Cheng, Kwang-Ting (eds.). *Electronic design automation: synthesis, verification, and test*. Morgan Kaufmann. ISBN 978-0-12-374364-0.
- Hachtel, Gary D.; Somenzi, Fabio (2006) [1996]. *Logic Synthesis and Verification Algorithms*. Springer Science & Business Media. ISBN 0-7923-9746-0.
- Hassoun, Soha; Sasao, Tsutomu, eds. (2002). *Logic synthesis and verification*. Kluwer. ISBN 978-0-7923-7606-4.
- Perkowski, Marek A.; Grygiel, Stanislaw (1995-11-20). "6. Historical Overview of the Research on Decomposition". *A Survey of Literature on Function Decomposition* (<http://web.cecs.pdx.edu/~mperkows/=PUBLICATIONS/PER/G1995/survey.pdf>) (PDF). Version IV. Functional Decomposition Group, Department of Electrical Engineering, Portland University, Portland, Oregon, USA. CiteSeerX 10.1.1.64.1129 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.64.1129>). Archived (<https://web.archive.org/web/20210328181709/http://web.cecs.pdx.edu/~mperkows/=PUBLICATIONS/PER/G1995/survey.pdf>) (PDF) from the original on 2021-03-28. Retrieved 2021-03-28. (188 pages)
- Stanković, Radomir S. [in German]; Sasao, Tsutomu; Astola, Jaakko Tapio [in Finnish] (August 2001). "Publications in the First Twenty Years of Switching Theory and Logic Design" (<http://ticsp.cs.tut.fi/images/a/a5/Stari-radovi-report.pdf>) (PDF). Tampere International Center for Signal Processing (TICSP) Series. Tampere University of Technology / TTKK, Monistamo, Finland. ISSN 1456-2774 (<https://search.worldcat.org/issn/1456-2774>). S2CID 62319288 (<https://api.semanticscholar.org/CorpusID:62319288>). #14. Archived (<https://web.archive.org/web/20170809064702/http://ticsp.cs.tut.fi/images/a/a5/Stari-radovi-report.pdf>) (PDF) from the original on 2017-08-09. Retrieved 2021-03-28. (4+60 pages)

External links

-  Media related to [Logic design](#) at Wikimedia Commons

Retrieved from "https://en.wikipedia.org/w/index.php?title=Logic_synthesis&oldid=1289760832"



Logic in computer science

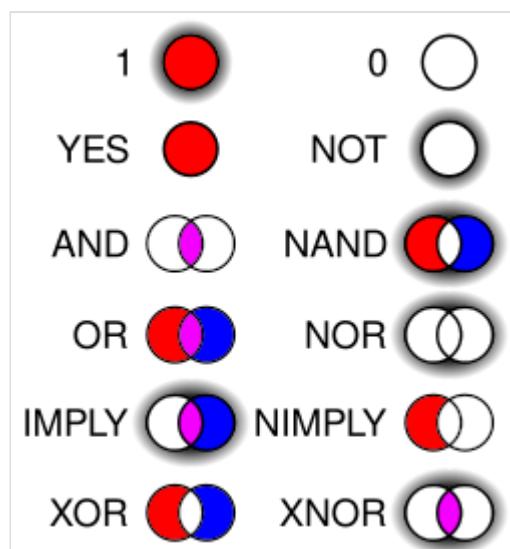
Logic in computer science covers the overlap between the field of logic and that of computer science. The topic can essentially be divided into three main areas:

- Theoretical foundations and analysis
- Use of computer technology to aid logicians
- Use of concepts from logic for computer applications

Theoretical foundations and analysis

Logic plays a fundamental role in computer science. Some of the key areas of logic that are particularly significant are computability theory (formerly called recursion theory), modal logic and category theory. The theory of computation is based on concepts defined by logicians and mathematicians such as Alonzo Church and Alan Turing.^{[1][2]} Church first showed the existence of algorithmically unsolvable problems using his notion of lambda-definability. Turing gave the first compelling analysis of what can be called a mechanical procedure and Kurt Gödel asserted that he found Turing's analysis "perfect".^[3] In addition some other major areas of theoretical overlap between logic and computer science are:

- Gödel's incompleteness theorem proves that any logical system powerful enough to characterize arithmetic will contain statements that can neither be proved nor disproved within that system. This has direct application to theoretical issues relating to the feasibility of proving the completeness and correctness of software.^[4]
- The frame problem is a basic problem that must be overcome when using first-order logic to represent the goals of an artificial intelligence agent and the state of its environment.^[5]
- The Curry–Howard correspondence is a relation between logical systems and programming languages. This theory established a precise correspondence between proofs and programs. In particular it showed that terms in the simply typed lambda calculus correspond to proofs of intuitionistic propositional logic.
- Category theory represents a view of mathematics that emphasizes the relations between structures. It is intimately tied to many aspects of computer science: type systems for programming languages, the theory of transition systems, models of programming languages and the theory of programming language semantics.^[6]
- Logic programming is a programming, database and knowledge representation paradigm that is based on formal logic. A logic program is a set of sentences about some problem domain. Computation is performed by applying logical reasoning to solve problems in the domain. Major logic programming language families include Prolog, Answer Set Programming (ASP) and Datalog.



Diagrammatic representation of computer logic gates

Computers to assist logicians

One of the first applications to use the term artificial intelligence was the Logic Theorist system developed by Allen Newell, Cliff Shaw, and Herbert Simon in 1956. One of the things that a logician does is to take a set of statements in logic and deduce the conclusions (additional statements) that must be true by the laws of logic. For example, if given the statements "All humans are mortal" and "Socrates is human" a valid conclusion is "Socrates is mortal". Of course this is a trivial example. In actual logical systems the statements can be numerous and complex. It was realized early on that this kind of analysis could be significantly aided by the use of computers. Logic Theorist validated the theoretical work of Bertrand Russell and Alfred North Whitehead in their influential work on mathematical logic called Principia Mathematica. In addition, subsequent systems have been utilized by logicians to validate and discover new mathematical theorems and proofs.^[7]

Logic applications for computers

There has always been a strong influence from mathematical logic on the field of artificial intelligence (AI). From the beginning of the field it was realized that technology to automate logical inferences could have great potential to solve problems and draw conclusions from facts. Ron Brachman has described first-order logic (FOL) as the metric by which all AI knowledge representation formalisms should be evaluated. First-order logic is a general and powerful method for describing and analyzing information. The reason FOL itself is simply not used as a computer language is that it is actually too expressive, in the sense that FOL can easily express statements that no computer, no matter how powerful, could ever solve. For this reason every form of knowledge representation is in some sense a trade off between expressivity and computability. A widely held belief maintains that the more expressive the language is, the closer it is to FOL, the more likely it is to be slower and prone to an infinite loop.^[8] However, in a recent work^[9] by Heng Zhang et al., this belief has been rigorously challenged. Their findings establish that all universal knowledge representation formalisms are recursively isomorphic. Furthermore, their proof demonstrates that FOL can be translated into a pure procedural knowledge representation formalism defined by Turing machines with computationally feasible overhead, specifically within deterministic polynomial time or even at lower complexity.^[9]

For example, IF–THEN rules used in expert systems approximate to a very limited subset of FOL. Rather than arbitrary formulas with the full range of logical operators the starting point is simply what logicians refer to as modus ponens. As a result, rule-based systems can support high-performance computation, especially if they take advantage of optimization algorithms and compilation.^[10]

On the other hand, logic programming, which combines the Horn clause subset of first-order logic with a non-monotonic form of negation, has both high expressive power and efficient implementations. In particular, the logic programming language Prolog is a Turing complete programming language. Datalog extends the relational database model with recursive relations, while answer set programming is a form of logic programming oriented towards difficult (primarily NP-hard) search problems.

Another major area of research for logical theory is software engineering. Research projects such as the Knowledge Based Software Assistant and Programmer's Apprentice programs have applied logical theory to validate the correctness of software specifications. They have also used logical tools to transform the

specifications into efficient code on diverse platforms and to prove the equivalence between the implementation and the specification.^[11] This formal transformation-driven approach is often far more effortful than traditional software development. However, in specific domains with appropriate formalisms and reusable templates the approach has proven viable for commercial products. The appropriate domains are usually those such as weapons systems, security systems, and real-time financial systems where failure of the system has excessively high human or financial cost. An example of such a domain is Very Large Scale Integrated (VLSI) design—the process for designing the chips used for the CPUs and other critical components of digital devices. An error in a chip can be catastrophic. Unlike software, chips can't be patched or updated. As a result, there is commercial justification for using formal methods to prove that the implementation corresponds to the specification.^[12]

Another important application of logic to computer technology has been in the area of frame languages and automatic classifiers. Frame languages such as KL-ONE can be directly mapped to set theory and first-order logic. This allows specialized theorem provers called classifiers to analyze the various declarations between sets, subsets, and relations in a given model. In this way the model can be validated and any inconsistent definitions flagged. The classifier can also infer new information, for example define new sets based on existing information and change the definition of existing sets based on new data. The level of flexibility is ideal for handling the ever changing world of the Internet. Classifier technology is built on top of languages such as the Web Ontology Language to allow a logical semantic level on top of the existing Internet. This layer is called the Semantic Web.^{[13][14]}

Temporal logic is used for reasoning in concurrent systems.^[15]

See also

- Automated reasoning
- Computational logic
- Logic programming

References

1. Lewis, Harry R. (1981). *Elements of the Theory of Computation* (<https://archive.org/details/elementsoftheory00lewi>). Prentice Hall.
2. Davis, Martin (11 May 1995). "Influences of Mathematical Logic on Computer Science" (<http://books.google.com/books?id=YafIDVd1Z68C&pg=PA290>). In Rolf Herken (ed.). *The Universal Turing Machine*. Springer Verlag. ISBN 9783211826379. Retrieved 26 December 2013.
3. Kennedy, Juliette (2014-08-21). *Interpreting Gödel* (<https://books.google.com/books?id=uLw3BAAAQBAJ&q=Gödel+convinced+by+Turing%27s+analysis&pg=PA118>). Cambridge University Press. ISBN 9781107002661. Retrieved 17 August 2015.
4. Hofstadter, Douglas R. (1999-02-05). *Gödel, Escher, Bach: An Eternal Golden Braid* (<https://archive.org/details/gdelescherbachet00hof>). Basic Books. ISBN 978-0465026562.
5. McCarthy, John; P.J. Hayes (1969). "Some philosophical problems from the standpoint of artificial intelligence" (<http://www-formal.stanford.edu/jmc/mcchay69.pdf>) (PDF). *Machine Intelligence*. 4: 463–502.
6. Barr, Michael; Charles Wells (1998). *Category Theory for Computing Science* (<https://www.math.mcgill.ca/barr/papers/ctcs.pdf>) (PDF). Centre de Recherches Mathématiques.

7. Newell, Allen; J.C. Shaw; H.C. Simon (1963). "Empirical explorations with the logic theory machine" (<https://archive.org/details/computersthought00feig>). In Ed Feigenbaum (ed.). *Computers and Thought*. McGraw Hill. pp. 109–133 (<https://archive.org/details/computersthought00feig/page/109>). ISBN 978-0262560924.
8. Levesque, Hector; Ronald Brachman (1985). "A Fundamental Tradeoff in Knowledge Representation and Reasoning" (<https://archive.org/details/readingsinknowle00brac>). In Ronald Brachman and Hector J. Levesque (ed.). *Reading in Knowledge Representation* (<https://archive.org/details/readingsinknowle00brac/page/49>). Morgan Kaufmann. p. 49 (<https://archive.org/details/readingsinknowle00brac/page/49>). ISBN 0-934613-01-X. "The good news in reducing KR service to theorem proving is that we now have a very clear, very specific notion of what the KR system should do; the bad new is that it is also clear that the services can not be provided... deciding whether or not a sentence in FOL is a theorem... is unsolvable."
9. Zhang, Heng; Jiang, Guifei; Quan, Donghui (2025-04-11). "A Theory of Formalisms for Representing Knowledge" (<https://ojs.aaai.org/index.php/AAAI/article/view/33674>). *Proceedings of the AAAI Conference on Artificial Intelligence*. **39** (14): 15257–15264. arXiv:2412.11855 (<https://arxiv.org/abs/2412.11855>). doi:10.1609/aaai.v39i14.33674 (<https://doi.org/10.1609%2Faaai.v39i14.33674>). ISSN 2374-3468 (<https://search.worldcat.org/issn/2374-3468>).
10. Forgy, Charles (1982). "Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem*" (<https://web.archive.org/web/20131227044049/http://web.yonsei.ac.kr/yusong/lecture/data/BI/Materials/1.1.Rete%20-%20Fast%20Algorithm%20for%20the%20Many%20Pattern,%20Many%20Object%20Pattern%20Match%20Problem.pdf>) (PDF). *Artificial Intelligence*. **19**: 17–37. doi:10.1016/0004-3702(82)90020-0 (<https://doi.org/10.1016%2F0004-3702%2882%2990020-0>). Archived from the original (<http://web.yonsei.ac.kr/yusong/lecture/data/BI/Materials/1.1.Rete%20-%20Fast%20Algorithm%20for%20the%20Many%20Pattern,%20Many%20Object%20Pattern%20Match%20Problem.pdf>) (PDF) on 2013-12-27. Retrieved 25 December 2013.
11. Rich, Charles; Richard C. Waters (November 1987). "The Programmer's Apprentice Project: A Research Overview" (<https://web.archive.org/web/20170706115702/ftp://publications.ai.mit.edu/ai-publications/pdf/AIM-1004.pdf>) (PDF). *IEEE Expert*. Archived from the original (<ftp://publications.ai.mit.edu/ai-publications/pdf/AIM-1004.pdf>) (PDF) on 2017-07-06. Retrieved 26 December 2013.
12. Stavridou, Victoria (1993). *Formal Methods in Circuit Design* (https://books.google.com/books?id=Hf_AZfW2YWsC&q=VLSI+chip+design+formal+methods&pg=PA14). Press Syndicate of the University of Cambridge. ISBN 0-521-443369. Retrieved 26 December 2013.
13. MacGregor, Robert (June 1991). "Using a description classifier to enhance knowledge representation". *IEEE Expert*. **6** (3): 41–46. doi:10.1109/64.87683 (<https://doi.org/10.1109%2F64.87683>). S2CID 29575443 (<https://api.semanticscholar.org/CorpusID:29575443>).
14. Berners-Lee, Tim; James Hendler; Ora Lassila (May 17, 2001). "The Semantic Web A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities" (<https://web.archive.org/web/20130424071228/http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html>). *Scientific American*. **284**: 34–43. doi:10.1038/scientificamerican0501-34 (<https://doi.org/10.1038%2Fscientificamerican0501-34>). Archived from the original (<http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html>) on April 24, 2013.
15. Colin Stirling (1992). "Modal and Temporal Logics". In S. Abramsky; D. M. Gabbay; T. S. E. Maibaum (eds.). *Handbook of Logic in Computer Science*. Vol. II. Oxford University Press. pp. 477–563. ISBN 0-19-853761-1.

Further reading

- Ben-Ari, Mordechai (2012). *Mathematical Logic for Computer Science* (<https://www.weizmann.ac.il/sci-tea/benari/research-activities/mathematical-logic-computer-science-third-edition>) (3rd ed.). Springer-Verlag. ISBN 978-1447141280.
- Harrison, John (2009). *Handbook of Practical Logic and Automated Reasoning* (<https://www.cl.cam.ac.uk/~jrh13/atp/>) (1st ed.). Cambridge University Press. ISBN 978-0521899574.
- Huth, Michael; Ryan, Mark (2004). *Logic in Computer Science: Modelling and Reasoning about Systems* (<http://www.cs.bham.ac.uk/research/lics/>) (2nd ed.). Cambridge University Press. ISBN 978-0521543101.
- Burris, Stanley N. (1997). *Logic for Mathematics and Computer Science* (<https://www.math.uwaterloo.ca/~snburris/htdocs/lmcs.html>) (1st ed.). Prentice Hall. ISBN 978-0132859745.

External links

- Article on *Logic and Artificial Intelligence* (<http://plato.stanford.edu/entries/logic-ai/>) at the *Stanford Encyclopedia of Philosophy*.
- IEEE Symposium on Logic in Computer Science (<https://web.archive.org/web/20051025052601/http://www.informatik.hu-berlin.de/lics/>) (LICS)
- Alwen Tiu, *Introduction to logic* (http://videolectures.net/ssli09_tiu_intlo/) video recording of a lecture at ANU Logic Summer School '09 (aimed mostly at computer scientists)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Logic_in_computer_science&oldid=1292271297"



Computer architecture

In computer science and computer engineering, **computer architecture** is a description of the structure of a computer system made from component parts.^[1] It can sometimes be a high-level description that ignores details of the implementation.^[2] At a more detailed level, the description may include the instruction set architecture design, microarchitecture design, logic design, and implementation.^[3]

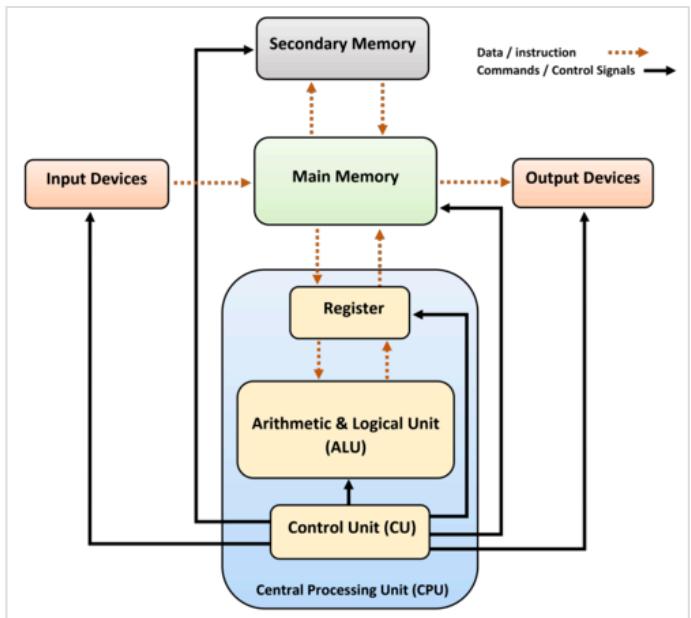
History

The first documented computer architecture was in the correspondence between Charles Babbage and Ada Lovelace, describing the analytical engine. While building the computer Z1 in 1936, Konrad Zuse described in two patent applications for his future projects that machine instructions could be stored in the same storage used for data, i.e., the stored-program concept.^{[4][5]} Two other early and important examples are:

- John von Neumann's 1945 paper, First Draft of a Report on the EDVAC, which described an organization of logical elements;^[6] and
- Alan Turing's more detailed Proposed Electronic Calculator for the Automatic Computing Engine, also 1945 and which cited John von Neumann's paper.^[7]

The term "architecture" in computer literature can be traced to the work of Lyle R. Johnson and Frederick P. Brooks, Jr., members of the Machine Organization department in IBM's main research center in 1959. Johnson had the opportunity to write a proprietary research communication about the Stretch, an IBM-developed supercomputer for Los Alamos National Laboratory (at the time known as Los Alamos Scientific Laboratory). To describe the level of detail for discussing the luxuriously embellished computer, he noted that his description of formats, instruction types, hardware parameters, and speed enhancements were at the level of "system architecture", a term that seemed more useful than "machine organization".^[8]

Subsequently, Brooks, a Stretch designer, opened Chapter 2 of a book called *Planning a Computer System: Project Stretch* by stating, "Computer architecture, like other architecture, is the art of determining the needs of the user of a structure and then designing to meet those needs as effectively as possible within economic and technological constraints."^[9]



Block diagram of a basic computer with uniprocessor CPU. Black lines indicate control flow, whereas red lines indicate data flow. Arrows indicate the direction of flow.

Brooks went on to help develop the IBM System/360 line of computers, in which "architecture" became a noun defining "what the user needs to know".^[10] The System/360 line was succeeded by several compatible lines of computers, including the current IBM Z line. Later, computer users came to use the term in many less explicit ways.^[11]

The earliest computer architectures were designed on paper and then directly built into the final hardware form.^[12] Later, computer architecture prototypes were physically built in the form of a transistor-transistor logic (TTL) computer—such as the prototypes of the 6800 and the PA-RISC—tested, and tweaked, before committing to the final hardware form. As of the 1990s, new computer architectures are typically "built", tested, and tweaked—inside some other computer architecture in a computer architecture simulator; or inside a FPGA as a soft microprocessor; or both—before committing to the final hardware form.^[13]

Subcategories

The discipline of computer architecture has three main subcategories:^[14]

- **Instruction set architecture** (ISA): defines the machine code that a processor reads and acts upon as well as the word size, memory address modes, processor registers, and data type.
- **Microarchitecture**: also known as "computer organization", this describes how a particular processor will implement the ISA.^[15] The size of a computer's CPU cache for instance, is an issue that generally has nothing to do with the ISA.
- **Systems design**: includes all of the other hardware components within a computing system, such as data processing other than the CPU (e.g., direct memory access), virtualization, and multiprocessing.

There are other technologies in computer architecture. The following technologies are used in bigger companies like Intel, and were estimated in 2002^[14] to count for 1% of all of computer architecture:

- **Macroarchitecture**: architectural layers more abstract than microarchitecture
- **Assembly instruction set architecture**: A smart assembler may convert an abstract assembly language common to a group of machines into slightly different machine language for different implementations.
- **Programmer-visible macroarchitecture**: higher-level language tools such as compilers may define a consistent interface or contract to programmers using them, abstracting differences between underlying ISAs and microarchitectures. For example, the C, C++, or Java standards define different programmer-visible macroarchitectures.
- **Microcode**: microcode is software that translates instructions to run on a chip. It acts like a wrapper around the hardware, presenting a preferred version of the hardware's instruction set interface. This instruction translation facility gives chip designers flexible options: E.g. 1. A new improved version of the chip can use microcode to present the exact same instruction set as the old chip version, so all software targeting that instruction set will run on the new chip without needing changes. E.g. 2. Microcode can present a variety of instruction sets for the same underlying chip, allowing it to run a wider variety of software.
- **Pin architecture**: The hardware functions that a microprocessor should provide to a hardware platform, e.g., the x86 pins A20M, FERR/IGNNE or FLUSH. Also, messages that the processor should emit so that external caches can be invalidated (emptied). Pin architecture functions are more flexible than ISA functions because external hardware can adapt to new encodings, or change from a pin to a message. The term "architecture" fits,

because the functions must be provided for compatible systems, even if the detailed method changes.

Roles

Definition

Computer architecture is concerned with balancing the performance, efficiency, cost, and reliability of a computer system. The case of instruction set architecture can be used to illustrate the balance of these competing factors. More complex instruction sets enable programmers to write more space efficient programs, since a single instruction can encode some higher-level abstraction (such as the x86 Loop instruction).^[16] However, longer and more complex instructions take longer for the processor to decode and can be more costly to implement effectively. The increased complexity from a large instruction set also creates more room for unreliability when instructions interact in unexpected ways.

The implementation involves integrated circuit design, packaging, power, and cooling. Optimization of the design requires familiarity with topics from compilers and operating systems to logic design and packaging.^[17]

Instruction set architecture

An instruction set architecture (ISA) is the interface between the computer's software and hardware and also can be viewed as the programmer's view of the machine. Computers do not understand high-level programming languages such as Java, C++, or most programming languages used. A processor only understands instructions encoded in some numerical fashion, usually as binary numbers. Software tools, such as compilers, translate those high level languages into instructions that the processor can understand.

Besides instructions, the ISA defines items in the computer that are available to a program—e.g., data types, registers, addressing modes, and memory. Instructions locate these available items with register indexes (or names) and memory addressing modes.

The ISA of a computer is usually described in a small instruction manual, which describes how the instructions are encoded. Also, it may define short (vaguely) mnemonic names for the instructions. The names can be recognized by a software development tool called an assembler. An assembler is a computer program that translates a human-readable form of the ISA into a computer-readable form. Disassemblers are also widely available, usually in debuggers and software programs to isolate and correct malfunctions in binary computer programs.

ISAs vary in quality and completeness. A good ISA compromises between programmer convenience (how easy the code is to understand), size of the code (how much code is required to do a specific action), cost of the computer to interpret the instructions (more complexity means more hardware needed to decode and execute the instructions), and speed of the computer (with more complex decoding hardware comes longer decode time). Memory organization defines how instructions interact with the memory, and how memory interacts with itself.

During design emulation, emulators can run programs written in a proposed instruction set. Modern emulators can measure size, cost, and speed to determine whether a particular ISA is meeting its goals.

Computer organization

Computer organization helps optimize performance-based products. For example, software engineers need to know the processing power of processors. They may need to optimize software in order to gain the most performance for the lowest price. This can require quite a detailed analysis of the computer's organization. For example, in an SD card, the designers might need to arrange the card so that the most data can be processed in the fastest possible way.

Computer organization also helps plan the selection of a processor for a particular project. Multimedia projects may need very rapid data access, while virtual machines may need fast interrupts. Sometimes certain tasks need additional components as well. For example, a computer capable of running a virtual machine needs virtual memory hardware so that the memory of different virtual computers can be kept separated. Computer organization and features also affect power consumption and processor cost.

Implementation

Once an instruction set and microarchitecture have been designed, a practical machine must be developed. This design process is called the *implementation*. Implementation is usually not considered architectural design, but rather hardware design engineering. Implementation can be further broken down into several steps:

- **Logic implementation** designs the circuits required at a logic-gate level.
- **Circuit implementation** does transistor-level designs of basic elements (e.g., gates, multiplexers, latches) as well as of some larger blocks (ALUs, caches etc.) that may be implemented at the logic-gate level, or even at the physical level if the design calls for it.
- **Physical implementation** draws physical circuits. The different circuit components are placed in a chip floor plan or on a board and the wires connecting them are created.
- **Design validation** tests the computer as a whole to see if it works in all situations and all timings. Once the design validation process starts, the design at the logic level are tested using logic emulators. However, this is usually too slow to run a realistic test. So, after making corrections based on the first test, prototypes are constructed using Field-Programmable Gate-Arrays (FPGAs). Most hobby projects stop at this stage. The final step is to test prototype integrated circuits, which may require several redesigns.

For CPUs, the entire implementation process is organized differently and is often referred to as CPU design.

Design goals

The exact form of a computer system depends on the constraints and goals. Computer architectures usually trade off standards, power versus performance, cost, memory capacity, latency (latency is the amount of time that it takes for information from one node to travel to the source) and throughput. Sometimes other considerations, such as features, size, weight, reliability, and expandability are also factors.

The most common scheme does an in-depth power analysis and figures out how to keep power consumption low while maintaining adequate performance.

Performance

Modern computer performance is often described in instructions per cycle (IPC), which measures the efficiency of the architecture at any clock frequency; a faster IPC rate means the computer is faster. Older computers had IPC counts as low as 0.1 while modern processors easily reach nearly 1. Superscalar processors may reach three to five IPC by executing several instructions per clock cycle.

Counting machine-language instructions would be misleading because they can do varying amounts of work in different ISAs. The "instruction" in the standard measurements is not a count of the ISA's machine-language instructions, but a unit of measurement, usually based on the speed of the VAX computer architecture.

Many people used to measure a computer's speed by the clock rate (usually in MHz or GHz). This refers to the cycles per second of the main clock of the CPU. However, this metric is somewhat misleading, as a machine with a higher clock rate may not necessarily have greater performance. As a result, manufacturers have moved away from clock speed as a measure of performance.

Other factors influence speed, such as the mix of functional units, bus speeds, available memory, and the type and order of instructions in the programs.

There are two main types of speed: latency and throughput. Latency is the time between the start of a process and its completion. Throughput is the amount of work done per unit time. Interrupt latency is the guaranteed maximum response time of the system to an electronic event (like when the disk drive finishes moving some data).

Performance is affected by a very wide range of design choices — for example, pipelining a processor usually makes latency worse, but makes throughput better. Computers that control machinery usually need low interrupt latencies. These computers operate in a real-time environment and fail if an operation is not completed in a specified amount of time. For example, computer-controlled anti-lock brakes must begin braking within a predictable and limited time period after the brake pedal is sensed or else failure of the brake will occur.

Benchmarking takes all these factors into account by measuring the time a computer takes to run through a series of test programs. Although benchmarking shows strengths, it should not be how you choose a computer. Often the measured machines split on different measures. For example, one system might handle scientific applications quickly, while another might render video games more smoothly. Furthermore, designers may target and add special features to their products, through hardware or software, that permit a specific benchmark to execute quickly but do not offer similar advantages to general tasks.

Power efficiency

Power efficiency is another important measurement in modern computers. Higher power efficiency can often be traded for lower speed or higher cost. The typical measurement when referring to power consumption in computer architecture is MIPS/W (millions of instructions per second per watt).

Modern circuits have less power required per transistor as the number of transistors per chip grows.^[18] This is because each transistor that is put in a new chip requires its own power supply and requires new pathways to be built to power it. However, the number of transistors per chip is starting to increase at a

slower rate. Therefore, power efficiency is starting to become as important, if not more important than fitting more and more transistors into a single chip. Recent processor designs have shown this emphasis as they put more focus on power efficiency rather than cramming as many transistors into a single chip as possible.^[19] In the world of embedded computers, power efficiency has long been an important goal next to throughput and latency.

Shifts in market demand

Increases in clock frequency have grown more slowly over the past few years, compared to power reduction improvements. This has been driven by the end of Moore's Law and demand for longer battery life and reductions in size for mobile technology. This change in focus from higher clock rates to power consumption and miniaturization can be shown by the significant reductions in power consumption, as much as 50%, that were reported by Intel in their release of the Haswell microarchitecture; where they dropped their power consumption benchmark from 30–40 watts down to 10–20 watts.^[20] Comparing this to the processing speed increase of 3 GHz to 4 GHz (2002 to 2006), it can be seen that the focus in research and development is shifting away from clock frequency and moving towards consuming less power and taking up less space.^[21]

See also



- Bit-serial architecture
- Comparison of CPU architectures
- Computer hardware
- CPU design
- Dataflow architecture
- Floating point
- Flynn's taxonomy
- Harvard architecture (Modified)
- Influence of the IBM PC on the personal computer market
- Orthogonal instruction set
- Reconfigurable computing
- Software architecture
- Transport triggered architecture
- Von Neumann architecture

References

1. Dragoni, Nicole (n.d.). "Introduction to peer to peer computing" (<http://www2.imm.dtu.dk/courses/02220/2017/L6/P2P.pdf>) (PDF). *DTU Compute – Department of Applied Mathematics and Computer Science*. Lyngby, Denmark.
2. Clements, Alan. *Principles of Computer Hardware* (Fourth ed.). p. 1. "Architecture describes the internal organization of a computer in an abstract way; that is, it defines the capabilities of the computer and its programming model. You can have two computers that have been constructed in different ways with different technologies but with the same architecture."

3. Hennessy, John; Patterson, David. *Computer Architecture: A Quantitative Approach* (Fifth ed.). p. 11. "This task has many aspects, including instruction set design, functional organization, logic design, and implementation."
4. Williams, F. C.; Kilburn, T. (25 September 1948), "Electronic Digital Computers", *Nature*, **162** (4117): 487, Bibcode:1948Natur.162..487W (<https://ui.adsabs.harvard.edu/abs/1948Natur.162..487W>), doi:10.1038/162487a0 (<https://doi.org/10.1038%2F162487a0>), S2CID 4110351 (<https://api.semanticscholar.org/CorpusID:4110351>)
5. Susanne Faber, "Konrad Zuses Bemühungen um die Patentanmeldung der Z3", 2000
6. Neumann, John (1945). *First Draft of a Report on the EDVAC*. p. 9.
7. Reproduced in B. J. Copeland (Ed.), "Alan Turing's Automatic Computing Engine", Oxford University Press, 2005, pp. 369-454.
8. Johnson, Lyle (1960). "A Description of Stretch" (<https://archive.computerhistory.org/resources/text/IBM/Stretch/pdfs/05-10/102634114.pdf>) (PDF). p. 1. Retrieved 7 October 2017.
9. Buchholz, Werner (1962). *Planning a Computer System*. p. 5.
10. "System 360, From Computers to Computer Systems" (<https://web.archive.org/web/20120403020049/http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/system360/>). *IBM100*. 7 March 2012. Archived from the original (<http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/system360/>) on April 3, 2012. Retrieved 11 May 2017.
11. Hellige, Hans Dieter (2004). "Die Genese von Wissenschaftskonzeptionen der Computerarchitektur: Vom "system of organs" zum Schichtmodell des Designraums". *Geschichten der Informatik: Visionen, Paradigmen, Leitmotive*. pp. 411–472.
12. ACE underwent seven paper designs in one year, before a prototype was initiated in 1948. [B. J. Copeland (Ed.), "Alan Turing's Automatic Computing Engine", OUP, 2005, p. 57]
13. Schmalz, M.S. "Organization of Computer Systems" (<https://www.cise.ufl.edu/~mssz/CompOrg/CDAintro.html>). *UF CISE*. Retrieved 11 May 2017.
14. John L. Hennessy and David A. Patterson. *Computer Architecture: A Quantitative Approach* (Third ed.). Morgan Kaufmann Publishers.
15. Laplante, Phillip A. (2001). *Dictionary of Computer Science, Engineering, and Technology*. CRC Press. pp. 94–95. ISBN 0-8493-2691-5.
16. Null, Linda (2019). *The Essentials of Computer Organization and Architecture* (5th ed.). Burlington, MA: Jones & Bartlett Learning. p. 280. ISBN 9781284123036.
17. Martin, Milo. "What is computer architecture?" (https://www.cis.upenn.edu/~milom/cis501-Fall11/lectures/00_intro.pdf) (PDF). *UPENN*. Retrieved 11 May 2017.
18. "Integrated circuits and fabrication" (http://eacharya.inflibnet.ac.in/data-server/eacharya-documents/53e0c6cbe413016f23443704_INFIEP_33/192/ET/33-192-ET-V1-S1_ssed_unit_4/module_10_integrated_circuits_and_fabrication_e-text.pdf) (PDF). Retrieved 8 May 2017.
19. "Exynos 9 Series (8895)" (http://www.samsung.com/semiconductor/minisite/Exynos/w/solution/mod_ap/8895/?CID=AFL-hq-mul-0813-11000170). *Samsung*. Retrieved 8 May 2017.
20. "Measuring Processor Power TDP vs ACP" (<http://www.intel.com/content/dam/doc/white-paper/resources-xeon-measuring-processor-power-paper.pdf>) (PDF). *Intel*. April 2011. Retrieved 5 May 2017.
21. "History of Processor Performance" (<https://www.cs.columbia.edu/~sedwards/classes/2012/3827-spring/advanced-arch-2011.pdf>) (PDF). *cs.columbia.edu*. 24 April 2012. Retrieved 5 May 2017.

Sources

- John L. Hennessy and David Patterson (2006). *Computer Architecture: A Quantitative Approach* (<http://www.elsevierdirect.com/product.jsp?isbn=9780123704900>) (Fourth ed.).

Morgan Kaufmann. ISBN 978-0-12-370490-0.

- Barton, Robert S., "Functional Design of Computers", *Communications of the ACM* 4(9): 405 (1961).
- Barton, Robert S., "A New Approach to the Functional Design of a Digital Computer", *Proceedings of the Western Joint Computer Conference*, May 1961, pp. 393–396. About the design of the Burroughs B5000 computer.
- Bell, C. Gordon; and Newell, Allen (1971). "Computer Structures: Readings and Examples" (http://research.microsoft.com/en-us/um/people/gbell/Computer_Structures_Readings_and_Examples/contents.html), McGraw-Hill.
- Blaauw, G.A., and Brooks, F.P., Jr., "The Structure of System/360, Part I-Outline of the Logical Structure" (<http://domino.research.ibm.com/tchjr/journalindex.nsf/d9f0a910ab8b637485256bc80066a393/95dc427e3fd3024a85256bfa006859f7?OpenDocument>), *IBM Systems Journal*, vol. 3, no. 2, pp. 119–135, 1964.
- Tanenbaum, Andrew S. (1979). *Structured Computer Organization*. Englewood Cliffs, New Jersey: Prentice-Hall. ISBN 0-13-148521-0.

External links

- Carnegie Mellon Computer Architecture Lectures (<https://www.youtube.com/user/cmu18447>)
- ISCA: Proceedings of the International Symposium on Computer Architecture (<http://portal.acm.org/toc.cfm?id=SERIES416&type=series&coll=GUIDE&dl=GUIDE&CFID=41492512&CFTOKEN=82922478>)
- Micro: IEEE/ACM International Symposium on Microarchitecture (<http://www.microarch.org/>)
- HPCA: International Symposium on High Performance Computer Architecture (<https://web.archive.org/web/20050528085407/http://www.hpcacconf.org/>)
- ASPLOS: International Conference on Architectural Support for Programming Languages and Operating Systems (<http://portal.acm.org/toc.cfm?id=SERIES311&type=series&coll=GUIDE&dl=GUIDE&CFID=41492415&CFTOKEN=3676847>)
- ACM Transactions on Architecture and Code Optimization (<http://www.acm.org/taco/>)
- IEEE Transactions on Computers (<https://www.computer.org/csdl/journal/tc>)
- The von Neumann Architecture of Computer Systems (<https://web.archive.org/web/20171031185802/http://www-scf.usc.edu/~inf520/downloads/The%20von%20Neumann%20Architecture%20of%20Computer%20Systems.pdf>) at the Wayback Machine (archived 2017-10-31)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Computer_architecture&oldid=1288827667"



Logic optimization

(Redirected from [Circuit minimization](#))

Logic optimization is a process of finding an equivalent representation of the specified [logic circuit](#) under one or more specified constraints. This process is a part of a [logic synthesis](#) applied in [digital electronics](#) and [integrated circuit design](#).

Generally, the circuit is constrained to a minimum chip area meeting a predefined response delay. The goal of logic optimization of a given circuit is to obtain the smallest [logic circuit](#) that evaluates to the same values as the original one.^[1] Usually, the smaller circuit with the same function is cheaper,^[2] takes less space, [consumes less power](#), has shorter latency, and minimizes risks of unexpected [cross-talk](#), [hazard of delayed signal processing](#), and other issues present at the nano-scale level of metallic structures on an [integrated circuit](#).

In terms of [Boolean algebra](#), the optimization of a complex [Boolean expression](#) is a process of finding a simpler one, which would upon evaluation ultimately produce the same results as the original one.

Motivation

The problem with having a complicated [circuit](#) (i.e. one with many elements, such as [logic gates](#)) is that each element takes up physical space and costs time and money to produce. Circuit minimization may be one form of logic optimization used to reduce the area of complex logic in [integrated circuits](#).

With the advent of [logic synthesis](#), one of the biggest challenges faced by the [electronic design automation](#) (EDA) industry was to find the most simple circuit representation of the given design description.^[nb 1] While [two-level logic optimization](#) had long existed in the form of the [Quine-McCluskey algorithm](#), later followed by the [Espresso heuristic logic minimizer](#), the rapidly improving chip densities, and the wide adoption of [Hardware description languages](#) for circuit description, formalized the logic optimization domain as it exists today, including [Logic Friday](#) (graphical interface), [Minilog](#), and [ESPRESSO-II](#) ([SOJS](#)) (many-valued logic).^[3]

Methods

The methods of logic circuit simplifications are equally applicable to [Boolean expression minimization](#).

Classification

Today, logic optimization is divided into various categories:

Based on circuit representation

- Two-level logic optimization
- Multi-level logic optimization

Based on circuit characteristics

Sequential logic optimization
Combinational logic optimization

Based on type of execution

Graphical optimization methods
Tabular optimization methods
Algebraic optimization methods

Graphical methods

Graphical methods represent the required logical function by a diagram representing the logic variables and value of the function. By manipulating or inspecting a diagram, much tedious calculation may be eliminated. Graphical minimization methods for two-level logic include:

- Euler diagram (aka *Eulerian circle*) (1768) by Leonhard P. Euler (1707–1783)
- Venn diagram (1880) by John Venn (1834–1923)
- Karnaugh map (1953) by Maurice Karnaugh

Boolean expression minimization

The same methods of Boolean expression minimization (simplification) listed below may be applied to the circuit optimization.

For the case when the Boolean function is specified by a circuit (that is, we want to find an equivalent circuit of minimum size possible), the unbounded circuit minimization problem was long-conjectured to be Σ_2^P -complete in time complexity, a result finally proved in 2008,^[4] but there are effective heuristics such as Karnaugh maps and the Quine–McCluskey algorithm that facilitate the process.

Boolean function minimizing methods include:

- Quine–McCluskey algorithm
- Petrick's method

Optimal multi-level methods

Methods that find optimal circuit representations of Boolean functions are often referred to as *exact synthesis* in the literature. Due to the computational complexity, exact synthesis is tractable only for small Boolean functions. Recent approaches map the optimization problem to a Boolean satisfiability problem.^{[5][6]} This allows finding optimal circuit representations using a SAT solver.

Heuristic methods

A heuristic method uses established rules that solve a practical useful subset of the much larger possible set of problems. The heuristic method may not produce the theoretically optimum solution, but if useful, will provide most of the optimization desired with a minimum of effort. An example of a computer system that uses heuristic methods for logic optimization is the Espresso heuristic logic minimizer.

Two-level versus multi-level representations

While a two-level circuit representation of circuits strictly refers to the flattened view of the circuit in terms of SOPs (sum-of-products) — which is more applicable to a PLA implementation of the design — a multi-level representation is a more generic view of the circuit in terms of arbitrarily connected SOPs, POSs (product-of-sums), factored form etc. Logic optimization algorithms generally work either on the structural (SOPs, factored form) or functional representation (binary decision diagrams, algebraic decision diagrams) of the circuit. In sum-of-products (SOP) form, AND gates form the smallest unit and are stitched together using ORs, whereas in product-of-sums (POS) form it is opposite. POS form requires parentheses to group the OR terms together under AND gates, because OR has lower precedence than AND. Both SOP and POS forms translate nicely into circuit logic.

If we have two functions F_1 and F_2 :

$$F_1 = AB + AC + AD,$$

$$F_2 = A'B + A'C + A'E.$$

The above 2-level representation takes six product terms and 24 transistors in CMOS Rep.

A functionally equivalent representation in multilevel can be:

$$P = B + C.$$

$$F_1 = AP + AD.$$

$$F_2 = A'P + A'E.$$

While the number of levels here is 3, the total number of product terms and literals reduce because of the sharing of the term $B + C$.

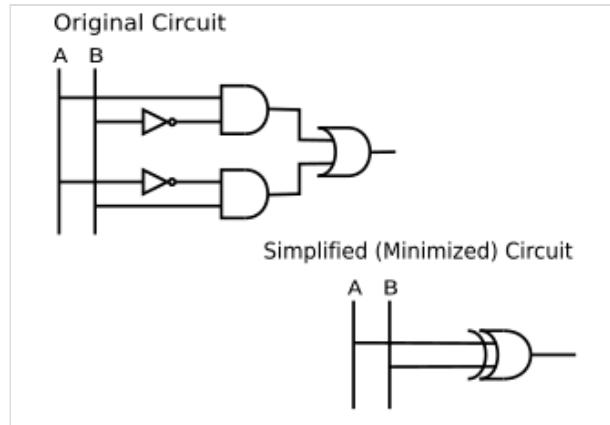
Similarly, we distinguish between combinational circuits and sequential circuits. Combinational circuits produce their outputs based only on the current inputs. They can be represented by Boolean relations. Some examples are priority encoders, binary decoders, multiplexers, demultiplexers.

Sequential circuits produce their output based on both current and past inputs, depending on a clock signal to distinguish the previous inputs from the current inputs. They can be represented by finite state machines. Some examples are flip-flops and counters.

Example

While there are many ways to minimize a circuit, this is an example that minimizes (or simplifies) a Boolean function. The Boolean function carried out by the circuit is directly related to the algebraic expression from which the function is implemented.^[7] Consider the circuit used to represent $(A \wedge \bar{B}) \vee (\bar{A} \wedge B)$. It is evident that two negations, two conjunctions, and a disjunction are used in this statement. This means that to build the circuit one would need two inverters, two AND gates, and an OR gate.

The circuit can be simplified (minimized) by applying laws of Boolean algebra or using intuition. Since the example states that A is true when B is false and the other way around, one can conclude that this simply means $A \neq B$. In terms of logical gates, inequality simply means an XOR gate (exclusive or). Therefore, $(A \wedge \bar{B}) \vee (\bar{A} \wedge B) \iff A \neq B$. Then the two circuits shown below are equivalent, as can be checked using a truth table:



Original and simplified example circuit

A	B	$(A \wedge \bar{B})$	\vee	$(\bar{A} \wedge B)$	$A \neq B$
F	F	F	F	T	F
F	T	F	F	F	T
T	F	T	T	T	F
T	T	T	F	F	T

See also

- Binary decision diagram (BDD)
- Don't care condition
- Prime implicant
- Circuit complexity — on estimation of the circuit complexity
- Function composition
- Function decomposition
- Gate underutilization
- Logic redundancy
- Harvard minimizing chart (Wikiversity) (Wikibooks)

Notes

1. The netlist size can be used to measure simplicity.

References

1. Maxfield, Clive "Max" (2008-01-01). "Chapter 5: "Traditional" Design Flows" (<https://www.sciencedirect.com/science/article/pii/B9780750689748000053>). In Maxfield, Clive "Max" (ed.). *FPGAs*. Instant Access. Burlington: Newnes / Elsevier Inc. pp. 75–106. doi:10.1016/B978-0-7506-8974-8.00005-3 (<https://doi.org/10.1016%2FB978-0-7506-8974-8.00005-3>). ISBN 978-0-7506-8974-8. Retrieved 2021-10-04.

2. Balasanyan, Seyran; Aghagulyan, Mane; Wuttke, Heinz-Dietrich; Henke, Karsten (2018-05-16). "Digital Electronics" (<https://ec.europa.eu/programmes/erasmus-plus/project-result-content/120e4810-0d29-4397-9ad4-b4091c2e3d19/Digital%20Electronics.pdf>) (PDF). Bachelor Embedded Systems - Year Group. Tempus. DesIRE. Archived (<https://web.archive.org/web/20211004200546/https://ec.europa.eu/programmes/erasmus-plus/project-result-content/120e4810-0d29-4397-9ad4-b4091c2e3d19/Digital%20Electronics.pdf>) (PDF) from the original on 2021-10-04. Retrieved 2021-10-04. (101 pages)
3. Theobald, M.; Nowick, S. M. (November 1998). "Fast heuristic and exact algorithms for two-level hazard-free logic minimization" (<https://academiccommons.columbia.edu/doi/10.7916/D8N58V58/download>). *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. **17** (11): 1130–1147. doi:[10.1109/43.736186](https://doi.org/10.1109/43.736186) (<https://doi.org/10.1109%2F43.736186>).
4. Buchfuhrer, David; Umans, Christopher (January 2011). "The complexity of Boolean formula minimization" (<http://users.cms.caltech.edu/~umans/papers/BU07.pdf>) (PDF). *Journal of Computer and System Sciences*. **77** (1). Computer Science Department, California Institute of Technology, Pasadena, California, USA: Elsevier Inc.: 142–153. doi:[10.1016/j.jcss.2010.06.011](https://doi.org/10.1016/j.jcss.2010.06.011) (<https://doi.org/10.1016%2Fj.jcss.2010.06.011>). This is an extended version of the conference paper: Buchfuhrer, David; Umans, Christopher (2008). "The Complexity of Boolean Formula Minimization". *Proceedings of Automata, Languages and Programming, 35th International Colloquium (ICALP)* (<http://users.cms.caltech.edu/~umans/papers/BU07.pdf>) (PDF). *Lecture Notes in Computer Science*. Vol. 5125. Berlin / Heidelberg, Germany: Springer-Verlag. pp. 24–35. doi:[10.1007/978-3-540-70575-8_3](https://doi.org/10.1007/978-3-540-70575-8_3) (https://doi.org/10.1007%2F978-3-540-70575-8_3). ISBN 978-3-540-70574-1. Archived (<https://web.archive.org/web/20180114141842/http://users.cms.caltech.edu/~umans/papers/BU07.pdf>) (PDF) from the original on 2018-01-14. Retrieved 2018-01-14.
5. Haaswijk, Winston. "SAT-Based Exact Synthesis: Encodings, Topology Families, and Parallelism" (<https://si2.epfl.ch/~demichel/publications/archive/2020/winston-exact.pdf>) (PDF). EPFL. Retrieved 2022-12-07.
6. Haaswijk, Winston. "SAT-Based Exact Synthesis for Multi-Level Logic Networks" (<https://si2.epfl.ch/~demichel/graduates/theses/winston.pdf>) (PDF). EPFL. Retrieved 2022-12-07.
7. Mano, M. Morris; Kime, Charles R. (2014). *Logic and Computer Design Fundamentals* (4th new international ed.). Pearson Education Limited. p. 54. ISBN 978-1-292-02468-4.

Further reading

- Lind, Larry Frederick; Nelson, John Christopher Cunliffe (1977). *Analysis and Design of Sequential Digital Systems* (<https://archive.org/details/AnalysisDesignOfSequentialDigitalSystems/>). Macmillan Press. ISBN 0-33319266-4. (146 pages)
- De Micheli, Giovanni (1994). *Synthesis and Optimization of Digital Circuits*. McGraw-Hill. ISBN 0-07-016333-2. (NB. Chapters 7–9 cover combinatorial two-level, combinatorial multi-level, and respectively sequential circuit optimization.)
- Hachtel, Gary D.; Somenzi, Fabio (2006) [1996]. *Logic Synthesis and Verification Algorithms*. Springer Science & Business Media. ISBN 978-0-387-31005-3.
- Kohavi, Zvi; Jha, Niraj K. (2009). "4–6". *Switching and Finite Automata Theory* (3rd ed.). Cambridge University Press. ISBN 978-0-521-85748-2.
- Rutenbar, Rob A. *Multi-level minimization, Part I: Models & Methods* (<https://www.ece.cmu.edu/~ee760/760docs/lec07.pdf>) (PDF) (lecture slides). Carnegie Mellon University (CMU). Lecture 7. Archived (<https://web.archive.org/web/20180115125725/https://www.ece.cmu.edu/~ee760/760docs/lec07.pdf>) (PDF) from the original on 2018-01-15. Retrieved 2018-01-15; Rutenbar, Rob A. *Multi-level minimization, Part II: Cube/Cokernel Extract* (<https://www.ece.cmu.edu/~ee760/760docs/lec08.pdf>) (PDF) (lecture slides). Carnegie Mellon University (CMU). Lecture 8. Archived (<https://web.archive.org/web/20180115125733/https://www.ece.cmu.edu/~ee760/760docs/lec08.pdf>) (PDF) from the original on 2018-01-15. Retrieved 2018-01-15.

cmu.edu/~ee760/760docs/lec08.pdf (PDF) from the original on 2018-01-15. Retrieved 2018-01-15.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Logic_optimization&oldid=1287018147#Circuit_minimization_in_Boolean_algebra"



Switching circuit theory

Switching circuit theory is the mathematical study of the properties of networks of idealized switches. Such networks may be strictly combinational logic, in which their output state is only a function of the present state of their inputs; or may also contain sequential elements, where the present state depends on the present state and past states; in that sense, sequential circuits are said to include "memory" of past states. An important class of sequential circuits are state machines. Switching circuit theory is applicable to the design of telephone systems, computers, and similar systems. Switching circuit theory provided the mathematical foundations and tools for digital system design in almost all areas of modern technology.^[1]

In an 1886 letter, Charles Sanders Peirce described how logical operations could be carried out by electrical switching circuits.^[2] During 1880–1881 he showed that NOR gates alone (or alternatively NAND gates alone) can be used to reproduce the functions of all the other logic gates, but this work remained unpublished until 1933.^[3] The first published proof was by Henry M. Sheffer in 1913, so the NAND logical operation is sometimes called Sheffer stroke; the logical NOR is sometimes called Peirce's arrow.^[4] Consequently, these gates are sometimes called *universal logic gates*.^[5]

In 1898, Martin Boda described a switching theory for signalling block systems.^{[6][7]}

Eventually, vacuum tubes replaced relays for logic operations. Lee De Forest's modification, in 1907, of the Fleming valve can be used as a logic gate. Ludwig Wittgenstein introduced a version of the 16-row truth table as proposition 5.101 of *Tractatus Logico-Philosophicus* (1921). Walther Bothe, inventor of the coincidence circuit, got part of the 1954 Nobel Prize in physics, for the first modern electronic AND gate in 1924. Konrad Zuse designed and built electromechanical logic gates for his computer Z1 (from 1935 to 1938).

The theory was independently established through the works of NEC engineer Akira Nakashima in Japan,^[8] Claude Shannon in the United States,^[9] and Victor Shestakov in the Soviet Union.^[10] The three published a series of papers showing that the two-valued Boolean algebra, can describe the operation of switching circuits.^{[7][11][12][13][1]} However, Shannon's work has largely overshadowed the other two, and despite some scholars arguing the similarities of Nakashima's work to Shannon's, their approaches and theoretical frameworks were markedly different.^[14] Also implausible is that Shestakov's influenced the other two due to the language barriers and the relative obscurity of his work abroad.^[14] Furthermore, Shannon and Shestakov defended their theses the same year in 1938,^[15] and Shestakov did not publish until 1941.^[15]

Ideal switches are considered as having only two exclusive states, for example, open or closed. In some analysis, the state of a switch can be considered to have no influence on the output of the system and is designated as a "don't care" state. In complex networks it is necessary to also account for the finite switching time of physical switches; where two or more different paths in a network may affect the output, these delays may result in a "logic hazard" or "race condition" where the output state changes due to the different propagation times through the network.

See also

- [Circuit switching](#)
- [Message switching](#)
- [Packet switching](#)
- [Fast packet switching](#)
- [Network switching subsystem](#)
- [5ESS Switching System](#)
- [Number One Electronic Switching System](#)
- [Boolean circuit](#)
- [C-element](#)
- [Circuit complexity](#)
- [Circuit minimization](#)
- [Karnaugh map](#)
- [Logic design](#)
- [Logic gate](#)
- [Logic in computer science](#)
- [Nonblocking minimal spanning switch](#)
- [Programmable logic controller](#) – computer software mimics relay circuits for industrial applications
- [Quine–McCluskey algorithm](#)
- [Relay](#) – an early kind of logic device
- [Switching lemma](#)
- [Unate function](#)

References

1. Stanković, Radomir S. [in German]; Astola, Jaakko Tapio [in Finnish], eds. (2008). *Reprints from the Early Days of Information Sciences: TICSP Series on the Contributions of Akira Nakashima to Switching Theory* (<https://web.archive.org/web/20210308002559/http://ticsp.cs.tut.fi/reports/reprint-nakashima-rr.pdf>) (PDF). Tampere International Center for Signal Processing (TICSP) Series. Vol. 40. Tampere University of Technology, Tampere, Finland. ISBN 978-952-15-1980-2. ISSN 1456-2774 (<https://search.worldcat.org/issn/1456-2774>). Archived from the original (<http://ticsp.cs.tut.fi/reports/reprint-nakashima-rr.pdf>) (PDF) on 2021-03-08. (3+207+1 pages) 10:00 min (https://web.archive.org/web/20221026175726/http://ciitlab.elfak.ni.ac.rs/predavanja/09_Nakashima.mp4)
2. Peirce, Charles Sanders (1993) [1886]. "Letter, Peirce to A. Marquand". *Writings of Charles S. Peirce*. Vol. 5. pp. 421–423. See also: Burks, Arthur Walter (1978). "Review: Charles S. Peirce, *The new elements of mathematics*" (http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?view=body&id=pdf_1&handle=euclid.bams/1183541145). *Bulletin of the American Mathematical Society* (review). 84 (5): 913–918 [917]. doi:10.1090/S0002-9904-1978-14533-9 (<https://doi.org/10.1090%2FS0002-9904-1978-14533-9>).
3. Peirce, Charles Sanders (1933) [Winter of 1880–1881]. "A Boolean Algebra with One Constant". *Collected Papers* (manuscript). Vol. 4. paragraphs 12–20. Reprinted in *Writings of Charles S. Peirce* (<https://archive.org/details/writingsofcharle0002peir/page/218>). Vol. 4 (reprint ed.). 1989. pp. 218–221. ISBN 9780253372017. ark:/13960/t11p5r61f. See also: Roberts, Don D. (2009). *The Existential Graphs of Charles S. Peirce*. p. 131.

4. Kleine Büning, Hans; Lettmann, Theodor (1999). *Propositional logic: deduction and algorithms* (<https://books.google.com/books?id=3oJE9yczr3EC&pg=PA2>). Cambridge University Press. p. 2. ISBN 978-0-521-63017-7.
5. Bird, John (2007). *Engineering mathematics* (<https://books.google.com/books?id=1-fBmsEBNUoC&pg=PA532>). Newnes. p. 532. ISBN 978-0-7506-8555-9.
6. Boda, Martin (1898). "Die Schaltungstheorie der Blockwerke" [The switching theory of block systems]. *Organ für die Fortschritte des Eisenbahnwesens in technischer Beziehung – Fachblatt des Vereins deutscher Eisenbahn-Verwaltungen* (in German). Neue Folge XXXV (1–7). Wiesbaden, Germany: C. W. Kreidel's Verlag: 1–7, 29–34, 49–53, 71–75, 91–95, 111–115, 133–138. [1] (https://web.archive.org/web/20220925115833/http://tudigit.ulb.tu-darmstadt.de/tmp/pdf/Zb-722-1897-98_p0283-0310.pdf)[2] (https://web.archive.org/web/20220925115950/http://tudigit.ulb.tu-darmstadt.de/tmp/pdf/Zb-722-1897-98_p0311-0330.pdf)[3] (https://web.archive.org/web/20220925120033/http://tudigit.ulb.tu-darmstadt.de/tmp/pdf/Zb-722-1897-98_p0331-0352.pdf)[4] (https://web.archive.org/web/20220925120209/http://tudigit.ulb.tu-darmstadt.de/tmp/pdf/Zb-722-1897-98_p0353-0372.pdf)[5] (https://web.archive.org/web/20220925120405/http://tudigit.ulb.tu-darmstadt.de/tmp/pdf/Zb-722-1897-98_p0373-0392.pdf)[6] (https://web.archive.org/web/20220925120444/http://tudigit.ulb.tu-darmstadt.de/tmp/pdf/Zb-722-1897-98_p0393-0414.pdf)[7] (https://web.archive.org/web/20220925120623/http://tudigit.ulb.tu-darmstadt.de/tmp/pdf/Zb-722-1897-98_p0415-0434.pdf) (NB. This series of seven articles was republished in a 91-pages book in 1899 with a foreword by Georg Barkhausen.)
7. Klir, George Jiří (May 1972). "Reference Notations to Chapter 1". *Introduction to the Methodology of Switching Circuits* (1 ed.). Binghamton, New York, USA: Litton Educational Publishing, Inc. / D. van Nostrand Company. p. 19. ISBN 0-442-24463-0. LCCN 72-181095 (<https://lccn.loc.gov/72-181095>). C4463-000-3. p. 19: "Although the possibility of establishing a switching theory was recognized by M. Boda^[A] as early as in the 19th century, the first important works on this subject were published by A. Nakashima^[B] and C. E. Shannon^[C] shortly before World War II." (xvi+573+1 pages)
8. Nakashima, Akira (May 1936). "Theory of Relay Circuit Composition". *Nippon Electrical Communication Engineering* (3): 197–226. (NB. Translation of an article which originally appeared in Japanese in the *Journal of the Institute of Telegraph and Telephone Engineers of Japan* (JITTEJ) September 1935, **150** 731–752.)
9. Shannon, Claude Elwood (1938). "A Symbolic Analysis of Relay and Switching Circuits". *Transactions of the American Institute of Electrical Engineers*. **57** (12). American Institute of Electrical Engineers (AIEE): 713–723. doi:10.1109/T-AIEE.1938.5057767 (<https://doi.org/10.1109%2FT-AIEE.1938.5057767>). hdl:1721.1/11173 (<https://hdl.handle.net/1721.1%2F11173>). S2CID 51638483 (<https://api.semanticscholar.org/CorpusID:51638483>). (NB. Based on Shannon's master thesis of the same title at Massachusetts Institute of Technology in 1937.)
10. Shestakov [Шестаков], Victor Ivanovich [Виктор Иванович] (1938). Некоторые математические методы конструирования и упрощения двухполюсных электрических схем класса А [Some mathematical methods for the construction and simplification of two-terminal electrical networks of class A] (PhD thesis) (in Russian). Lomonosov State University.
11. Yamada [山田], Akihiko [彰彦] (2004). "History of Research on Switching Theory in Japan" (https://www.jstage.jst.go.jp/article/ieejfms/124/8/124_8_720/_article). *IEEJ Transactions on Fundamentals and Materials*. **124** (8). Institute of Electrical Engineers of Japan: 720–726. Bibcode:2004IJTFM.124..720Y (<https://ui.adsabs.harvard.edu/abs/2004IJTFM.124..720Y>). doi:10.1541/ieejfms.124.720 (<https://doi.org/10.1541%2Fieejfms.124.720>). Archived (https://web.archive.org/web/20220710005644/https://www.jstage.jst.go.jp/article/ieejfms/124/8/124_8_720/_pdf/-char/en) from the original on 2022-07-10. Retrieved 2022-10-26.

12. "Switching Theory/Relay Circuit Network Theory/Theory of Logical Mathematics" (<http://museum.ipsj.or.jp/en/computer/dawn/0002.html>). *IPSJ Computer Museum*. Information Processing Society of Japan. 2012. Archived (<https://web.archive.org/web/2021032204220/http://museum.ipsj.or.jp/en/computer/dawn/0002.html>) from the original on 2021-03-22. Retrieved 2021-03-28.
13. Stanković, Radomir S. [in German]; Astola, Jaakko Tapio [in Finnish]; Karpovsky, Mark G. (2007). *Some Historical Remarks on Switching Theory* (<https://sites.bu.edu/mark/files/2018/02/200.pdf.pdf>) (PDF). Niš, Serbia; Tampere, Finland; Boston, Massachusetts, USA. CiteSeerX 10.1.1.66.1248 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.66.1248>). S2CID 10029339 (<https://api.semanticscholar.org/CorpusID:10029339>). Archived (<http://web.archive.org/web/20221025215232/https://sites.bu.edu/mark/files/2018/02/200.pdf.pdf>) (PDF) from the original on 2022-10-25. Retrieved 2022-10-25. (8 pages)
14. Kawanishi, Toma (2019). "Prehistory of Switching Theory in Japan: Akira Nakashima and His Relay-circuit Theory" (https://www.jstage.jst.go.jp/article/historiascientiarum/29/1/29_136/_article). *Historia Scientiarum*. Second Series. **29** (1): 136–162. doi:10.34336/historiascientiarum.29.1_136 (https://doi.org/10.34336%2Fhistoriascientiarum.29.1_136).
15. Moisil, GR. C. (1969). *The Algebraic Theory of Switching Circuits* (<https://books.google.com/books?id=tn7iBQAAQBAJ&dq=claude+shannon+shestakov&pg=PA17>). Pergamon Press. pp. 12, 17. ISBN 9781483160764.

Further reading

- Keister, William; Ritchie, Alistair E.; Washburn, Seth H. (1951). *The Design of Switching Circuits* (<https://archive.org/details/TheDesignOfSwitchingCircuits>). The Bell Telephone Laboratories Series (1 ed.). D. Van Nostrand Company, Inc. p. 147 (<https://archive.org/details/TheDesignOfSwitchingCircuits/page/n170>). Archived (<https://web.archive.org/web/20200509143306/https://archive.org/details/TheDesignOfSwitchingCircuits/page/n537>) from the original on 2020-05-09. Retrieved 2020-05-09. [8] (<https://archive.org/details/TheDesignOfSwitchingCircuits>) (2+xx+556+2 pages)
- Caldwell, Samuel Hawks (1958-12-01) [February 1958]. Written at Watertown, Massachusetts, USA. *Switching Circuits and Logical Design*. 5th printing September 1963 (1st ed.). New York, USA: John Wiley & Sons Inc. ISBN 0-47112969-0. LCCN 58-7896 (<https://lccn.loc.gov/58-7896>). (xviii+686 pages)
- Perkowski, Marek A.; Grygiel, Stanislaw (1995-11-20). "6. Historical Overview of the Research on Decomposition". *A Survey of Literature on Function Decomposition* (<http://web.cecs.pdx.edu/~mperkows/=PUBLICATIONS/PER/G1995/survey.pdf>) (PDF). Version IV. Functional Decomposition Group, Department of Electrical Engineering, Portland University, Portland, Oregon, USA. CiteSeerX 10.1.1.64.1129 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.64.1129>). Archived (<https://web.archive.org/web/20210328181709/http://web.cecs.pdx.edu/~mperkows/=PUBLICATIONS/PER/G1995/survey.pdf>) (PDF) from the original on 2021-03-28. Retrieved 2021-03-28. (188 pages)
- Stanković, Radomir S. [in German]; Sasao, Tsutomu; Astola, Jaakko Tapio [in Finnish] (August 2001). "Publications in the First Twenty Years of Switching Theory and Logic Design" (<https://web.archive.org/web/20170809064702/http://ticsp.cs.tut.fi/images/a/a5/Stari-radovi-report.pdf>) (PDF). Tampere International Center for Signal Processing (TICSP) Series. Tampere University of Technology / TTKK, Monistamo, Finland. ISSN 1456-2774 (<https://search.worldcat.org/issn/1456-2774>). S2CID 62319288 (<https://api.semanticscholar.org/CorpusID:62319288>). #14. Archived from the original (<http://ticsp.cs.tut.fi/images/a/a5/Stari-radovi-report.pdf>) (PDF) on 2017-08-09. Retrieved 2021-03-28. (4+60 pages)
- Stanković, Radomir S. [in German]; Astola, Jaakko Tapio [in Finnish] (2011). Written at Niš, Serbia & Tampere, Finland. *From Boolean Logic to Switching Circuits and Automata*:

Towards Modern Information Technology (<https://books.google.com/books?id=ZuHwCAAAQBAJ>). Studies in Computational Intelligence. Vol. 335 (1 ed.). Berlin & Heidelberg, Germany: Springer-Verlag. doi:10.1007/978-3-642-11682-7 (<https://doi.org/10.1007%2F978-3-642-11682-7>). ISBN 978-3-642-11681-0. ISSN 1860-949X (<https://search.worldcat.org/issn/1860-949X>). LCCN 2011921126 (<https://lccn.loc.gov/2011921126>). Retrieved 2022-10-25. (xviii+212 pages)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Switching_circuit_theory&oldid=1280700380"



Gate equivalent

A **gate equivalent** (GE) stands for a unit of measure which allows specifying *manufacturing-technology-independent complexity* of digital electronic circuits. For today's CMOS technologies, the silicon area of a two-input drive-strength-one NAND gate usually constitutes the *technology-dependent* unit area commonly referred to as gate equivalent. A specification in gate equivalents for a certain circuit reflects a complexity measure, from which a corresponding silicon area can be deduced for a dedicated manufacturing technology.

In digital circuit design, a dedicated standard cell library is employed for each manufacturing technology (e.g., CMOS). The standard cell library comprises many different logic gates, for example a NAND gate. For each logical type of logic gate, e.g., a two-input NAND, there usually exist different physical realizations in the standard cell library, for instance with different output drive strengths.

Basically, a two-input drive-strength-one NAND gate in CMOS technology consists of four transistors.

See also

- Logic family
- NMOS logic
- MOSFET
- Fanout
- FO4
- Boolean logic

References

- Digital Integrated Circuit Design: From VLSI Architectures to CMOS Fabrication (<http://www.cambridge.org/catalogue/catalogue.asp?isbn=0521882672>), Hubert Kaeslin, Cambridge University Press, 2008

Retrieved from "https://en.wikipedia.org/w/index.php?title=Gate_equivalent&oldid=1177898259"



Place and route

Place and route (also called PnR or P&R) is a stage in the design of printed circuit boards, integrated circuits, and field-programmable gate arrays. As implied by the name, it is composed of two steps, placement and routing. The first step, placement, involves deciding where to place all electronic components, circuitry, and logic elements in a generally limited amount of space. This is followed by routing, which decides the exact design of all the wires needed to connect the placed components. This step must implement all the desired connections while following the rules and limitations of the manufacturing process.

Place and route is used in several contexts:

- Printed circuit boards, during which components are graphically placed on the board and the wires drawn between them
- Integrated circuits, during which a layout of a larger block of the circuit or the whole circuit is created from layouts of smaller sub-blocks
- FPGAs, during which logic elements are placed and interconnected on the grid of the FPGA

These processes are similar at a high level, but the actual details are very different. With the large sizes of modern designs, this operation is usually performed by electronic design automation (EDA) tools.

In all these contexts, the final result when placing and routing is finished is the "layout", a geometric description of the location and rotation of each part, and the exact path of each wire connecting them.

Occasionally some people call the entire place-and-route process "layout".

Printed circuit board

The design of a printed circuit board comes after the creation of a schematic and generation of a netlist. The generated netlist is then read into a layout tool and associated with the footprints of the devices from a library. Placing and routing the devices can now start.^[1]

Placing and routing is generally done in two steps. Placing the components comes first, then routing the connections between the components. The placement of components is not absolute during the routing phase, as it may still be changed by moving and rotating, especially with designs using more complex components such as FPGAs or microprocessors. Their large number of signals, and their signal integrity needs may require optimization of the placement.^[2]

The resulting design is then output in RS-274X Gerber format to load in the computer-aided manufacturing (CAM) system of the manufacturer. In contrast to an IC layout, where the entire finished layout is stored in one graphics file, different files and formats are needed for PCB manufacture. The fabrication data consists of a set of Gerber files, a drill file, and a pick-and-place file containing the location and alignment of the devices generated for automated placement of the devices in the assembly process.^[1]

Field-programmable gate array

The process of placing and routing for an FPGA is generally not performed by a person, but uses a tool provided by the FPGA Vendor or another software manufacturer. The need for software tools is because of the complexity of the circuitry within the FPGA and the function the designer wishes to perform. FPGA designs are described using logic diagrams containing digital logic and hardware description languages such as VHDL and Verilog. These will then be put through an automated place-and-route procedure to generate a pinout, which will be used to interface with the parts outside of the FPGA.^[2]

Integrated circuits

The IC place-and-route stage typically starts with one or more schematics, HDL files, or pre-routed IP cores, or some combination of all three. It produces an IC layout that is automatically converted to a mask work in the standard GDS II or the OASIS format.^[3]

History

The final layout of early ICs and PCBs was stored as a tape-out of Rubylith on transparent film.

Gradually, electronic design automation automated more and more of the place-and-route work. At first, it merely sped up the process of making many small edits without spending a lot of time peeling up and sticking down the tape. Later design rule checking sped up the process of checking for the most common sorts of errors. Later auto routers speed up the process of routing.

Some people hope that further improvements in autoplacers and autorouters will eventually produce good layouts without any human manual intervention. Further automation leads to the idea of a silicon compiler.

References

1. J. Lienig, J. Scheible (2020). "Chap. 1.3.3: Physical Design of Printed Circuit Boards". *Fundamentals of Layout Design for Electronic Circuits* (<https://link.springer.com/book/10.1007/978-3-030-39284-0>). Springer. p. 26-27. doi:[10.1007/978-3-030-39284-0](https://doi.org/10.1007/978-3-030-39284-0) (<https://doi.org/10.1007%2F978-3-030-39284-0>). ISBN 978-3-030-39284-0. S2CID 215840278 (<https://api.semanticscholar.org/CorpusID:215840278>).
2. "FPGA/PCB Co-Design Increases Fabrication Yields" (<http://pcdandf.com/cms/content/view/4345/95/>). Printed Circuit Design and Fabrication. Retrieved 2008-07-24.
3. A. Kahng, J. Lienig, I. Markov, J. Hu: "VLSI Physical Design: From Graph Partitioning to Timing Closure", Springer (2022), doi:[10.1007/978-3-030-96415-3](https://doi.org/10.1007/978-3-030-96415-3) (<https://doi.org/10.1007/978-3-030-96415-3>), ISBN 978-3-030-96414-6, pp. 5-10.



Placement

Placement may refer to:

- [Placement \(EDA\)](#), an essential step in E-design automation
- [Placement exam](#), determines which class a student should take
- [Favored placement](#), the practice of preferentially listing search engine results for given sites
- [Job placement](#), a short time spent with an employer to get work experience
- [Private placement](#), a direct offering of securities to a limited number of sophisticated institutional investors
- [Product placement](#), a promotional tactic used by marketers in which a real commercial product is used in fictional or non-fictional media
- [*Public placement*](#), see [Initial public offering](#)

See also

- [All pages with titles beginning with *Placement*](#)
- [All pages with titles containing *Placement*](#)
- [Emplacement \(disambiguation\)](#)
- [Place \(disambiguation\)](#)

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Placement&oldid=1165939254>"



Routing

Routing is the process of selecting a path for traffic in a network or between or across multiple networks. Broadly, routing is performed in many types of networks, including circuit-switched networks, such as the public switched telephone network (PSTN), and computer networks, such as the Internet.

In packet switching networks, routing is the higher-level decision making that directs network packets from their source toward their destination through intermediate network nodes by specific packet forwarding mechanisms. Packet forwarding is the transit of network packets from one network interface to another. Intermediate nodes are typically network hardware devices such as routers, gateways, firewalls, or switches. General-purpose computers also forward packets and perform routing, although they have no specially optimized hardware for the task.

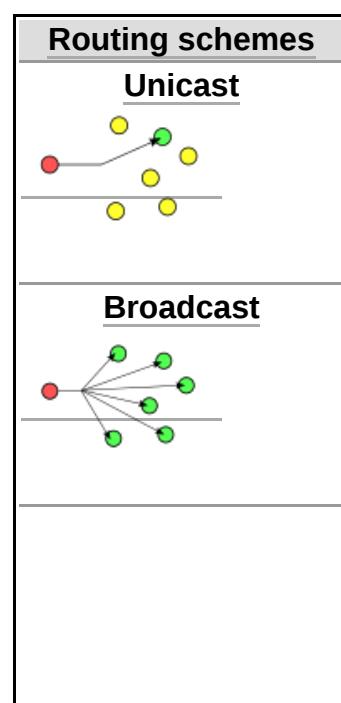
The routing process usually directs forwarding on the basis of routing tables. Routing tables maintain a record of the routes to various network destinations. Routing tables may be specified by an administrator, learned by observing network traffic or built with the assistance of routing protocols.

Routing, in a narrower sense of the term, often refers to IP routing and is contrasted with bridging. IP routing assumes that network addresses are structured and that similar addresses imply proximity within the network. Structured addresses allow a single routing table entry to represent the route to a group of devices. In large networks, structured addressing (routing, in the narrow sense) outperforms unstructured addressing (bridging). Routing has become the dominant form of addressing on the Internet. Bridging is still widely used within local area networks.

Delivery schemes

Routing schemes differ in how they deliver messages:

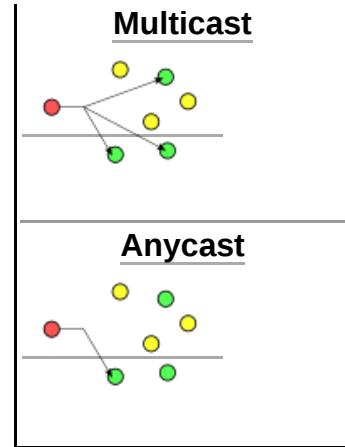
- Unicast delivers a message to a single specific node using a one-to-one association between a sender and destination: each destination address uniquely identifies a single receiver endpoint.
- Broadcast delivers a message to all nodes in the network using a one-to-all association; a single datagram (or packet) from one sender is routed to all of the possibly multiple endpoints associated with the broadcast address. The network automatically replicates datagrams as needed to reach all the recipients within the scope of the broadcast, which is generally an entire network subnet.
- Multicast delivers a message to a group of nodes that have expressed interest in receiving the message using a one-to-many-of-many or many-to-many-of-many association; datagrams are routed simultaneously in a single transmission to many recipients. Multicast differs from broadcast in that the



destination address designates a subset, not necessarily all, of the accessible nodes.

- Anycast delivers a message to any one out of a group of nodes, typically the one nearest to the source using a *one-to-one-of-many*^[1] association where datagrams are routed to any single member of a group of potential receivers that are all identified by the same destination address. The routing algorithm selects the single receiver from the group based on which is the nearest according to some distance or cost measure.

Unicast is the dominant form of message delivery on the Internet. This article focuses on unicast routing algorithms.



Topology distribution

With static routing, small networks may use manually configured routing tables. Larger networks have complex topologies that can change rapidly, making the manual construction of routing tables unfeasible. Nevertheless, most of the public switched telephone network (PSTN) uses pre-computed routing tables, with fallback routes if the most direct route becomes blocked (see routing in the PSTN).

Dynamic routing attempts to solve this problem by constructing routing tables automatically, based on information carried by routing protocols, allowing the network to act nearly autonomously in avoiding network failures and blockages. Dynamic routing dominates the Internet. Examples of dynamic-routing protocols and algorithms include Routing Information Protocol (RIP), Open Shortest Path First (OSPF) and Enhanced Interior Gateway Routing Protocol (EIGRP).

Distance vector algorithms

Distance vector algorithms use the Bellman–Ford algorithm. This approach assigns a *cost* number to each of the links between each node in the network. Nodes send information from point A to point B via the path that results in the lowest *total cost* (i.e. the sum of the costs of the links between the nodes used).

When a node first starts, it only knows of its immediate neighbors and the direct cost involved in reaching them. (This information — the list of destinations, the total cost to each, and the *next hop* to send data to get there — makes up the routing table, or *distance table*.) Each node, on a regular basis, sends to each neighbor node its own current assessment of the total cost to get to all the destinations it knows of. The neighboring nodes examine this information and compare it to what they already know; anything that represents an improvement on what they already have, they insert in their own table. Over time, all the nodes in the network discover the best next hop and total cost for all destinations.

When a network node goes down, any nodes that used it as their next hop discard the entry and convey the updated routing information to all adjacent nodes, which in turn repeat the process. Eventually, all the nodes in the network receive the updates and discover new paths to all the destinations that do not involve the down node.

Link-state algorithms

When applying link-state algorithms, a graphical map of the network is the fundamental data used for each node. To produce its map, each node floods the entire network with information about the other nodes it can connect to. Each node then independently assembles this information into a map. Using this map, each router independently determines the least-cost path from itself to every other node using a standard shortest paths algorithm such as Dijkstra's algorithm. The result is a tree graph rooted at the current node, such that the path through the tree from the root to any other node is the least-cost path to that node. This tree then serves to construct the routing table, which specifies the best next hop to get from the current node to any other node.

Optimized Link State Routing algorithm

A link-state routing algorithm optimized for mobile ad hoc networks is the optimized Link State Routing Protocol (OLSR).^[2] OLSR is proactive; it uses Hello and Topology Control (TC) messages to discover and disseminate link-state information through the mobile ad hoc network. Using Hello messages, each node discovers 2-hop neighbor information and elects a set of *multipoint relays* (MPRs). MPRs distinguish OLSR from other link-state routing protocols.

Path-vector protocol

Distance vector and link-state routing are both intra-domain routing protocols. They are used inside an autonomous system, but not between autonomous systems. Both of these routing protocols become intractable in large networks and cannot be used in inter-domain routing. Distance vector routing is subject to instability if there are more than a few hops in the domain. Link state routing needs significant resources to calculate routing tables. It also creates heavy traffic due to flooding.

Path-vector routing is used for inter-domain routing. It is similar to distance vector routing. Path-vector routing assumes that one node (there can be many) in each autonomous system acts on behalf of the entire autonomous system. This node is called the *speaker node*. The speaker node creates a routing table and advertises it to neighboring speaker nodes in neighboring autonomous systems. The idea is the same as distance vector routing except that only speaker nodes in each autonomous system can communicate with each other. The speaker node advertises the path, not the metric, of the nodes in its autonomous system or other autonomous systems.

The path-vector routing algorithm is similar to the distance vector algorithm in the sense that each border router advertises the destinations it can reach to its neighboring router. However, instead of advertising networks in terms of a destination and the distance to that destination, networks are advertised as destination addresses and path descriptions to reach those destinations. The path, expressed in terms of the domains (or confederations) traversed so far, is carried in a special path attribute that records the sequence of routing domains through which the reachability information has passed. A route is defined as a pairing between a destination and the attributes of the path to that destination, thus the name, path-vector routing; The routers receive a vector that contains paths to a set of destinations.^[3]

Path selection

Path selection involves applying a routing metric to multiple routes to select (or predict) the best route. Most routing algorithms use only one network path at a time. Multipath routing and specifically equal-cost multi-path routing techniques enable the use of multiple alternative paths.

In computer networking, the metric is computed by a routing algorithm, and can cover information such as bandwidth, network delay, hop count, path cost, load, maximum transmission unit, reliability, and communication cost.^[4] The routing table stores only the best possible routes, while link-state or topological databases may store all other information as well.

In case of overlapping or equal routes, algorithms consider the following elements in priority order to decide which routes to install into the routing table:

1. *Prefix length*: A matching route table entry with a longer subnet mask is always preferred as it specifies the destination more exactly.
2. *Metric*: When comparing routes learned via the same routing protocol, a lower metric is preferred. Metrics cannot be compared between routes learned from different routing protocols.
3. *Administrative distance*: When comparing route table entries from different sources such as different routing protocols and static configuration, a lower administrative distance indicates a more reliable source and thus a preferred route.

Because a routing metric is specific to a given routing protocol, multi-protocol routers must use some external heuristic to select between routes learned from different routing protocols. Cisco routers, for example, attribute a value known as the administrative distance to each route, where smaller administrative distances indicate routes learned from a protocol assumed to be more reliable.

A local administrator can set up host-specific routes that provide more control over network usage, permits testing, and better overall security. This is useful for debugging network connections or routing tables.

In some small systems, a single central device decides ahead of time the complete path of every packet. In some other small systems, whichever edge device injects a packet into the network decides ahead of time the complete path of that particular packet. In either case, the route-planning device needs to know a lot of information about what devices are connected to the network and how they are connected to each other. Once it has this information, it can use an algorithm such as A* search algorithm to find the best path.

In high-speed systems, there are so many packets transmitted every second that it is infeasible for a single device to calculate the complete path for each and every packet. Early high-speed systems dealt with this with circuit switching by setting up a path once for the first packet between some source and some destination; later packets between that same source and that same destination continue to follow the same path without recalculating until the circuit teardown. Later high-speed systems inject packets into the network without any one device ever calculating a complete path for packets.

In large systems, there are so many connections between devices, and those connections change so frequently, that it is infeasible for any one device to even know how all the devices are connected to each other, much less calculate a complete path through them. Such systems generally use next-hop routing.

Most systems use a deterministic dynamic routing algorithm. When a device chooses a path to a particular final destination, that device always chooses the same path to that destination until it receives information that makes it think some other path is better.

A few routing algorithms do not use a deterministic algorithm to find the best link for a packet to get from its original source to its final destination. Instead, to avoid congestion hot spots in packet systems, a few algorithms use a randomized algorithm—Valiant's paradigm—that routes a path to a randomly picked intermediate destination, and from there to its true final destination.^{[5][6]} In many early telephone switches, a randomizer was often used to select the start of a path through a multistage switching fabric.

Depending on the application for which path selection is performed, different metrics can be used. For example, for web requests one can use minimum latency paths to minimize web page load time, or for bulk data transfers one can choose the least utilized path to balance load across the network and increase throughput. A popular path selection objective is to reduce the average completion times of traffic flows and the total network bandwidth consumption. Recently, a path selection metric was proposed that computes the total number of bytes scheduled on the edges per path as selection metric.^[7] An empirical analysis of several path selection metrics, including this new proposal, has been made available.^[8]

Multiple agents

In some networks, routing is complicated by the fact that no single entity is responsible for selecting paths; instead, multiple entities are involved in selecting paths or even parts of a single path. Complications or inefficiency can result if these entities choose paths to optimize their own objectives, which may conflict with the objectives of other participants.

A classic example involves traffic in a road system, in which each driver picks a path that minimizes their travel time. With such routing, the equilibrium routes can be longer than optimal for all drivers. In particular, Braess's paradox shows that adding a new road can *lengthen* travel times for all drivers.

In a single-agent model used, for example, for routing automated guided vehicles (AGVs) on a terminal, reservations are made for each vehicle to prevent simultaneous use of the same part of an infrastructure. This approach is also referred to as context-aware routing.^[9]

The Internet is partitioned into autonomous systems (ASs) such as internet service providers (ISPs), each of which controls routes involving its network. Routing occurs at multiple levels. First, AS-level paths are selected via the BGP protocol that produces a sequence of ASs through which packets flow. Each AS may have multiple paths, offered by neighboring ASs, from which to choose. These routing decisions often correlate with business relationships with these neighboring ASs,^[10] which may be unrelated to path quality or latency. Second, once an AS-level path has been selected, there are often multiple corresponding router-level paths to choose from. This is due, in part, because two ISPs may be connected through multiple connections. In choosing the single router-level path, it is common practice for each ISP to employ hot-potato routing: sending traffic along the path that minimizes the distance through the ISP's own network—even if that path lengthens the total distance to the destination.

For example, consider two ISPs, *A* and *B*. Each has a presence in New York, connected by a fast link with latency 5 ms—and each has a presence in London connected by a 5 ms link. Suppose both ISPs have trans-Atlantic links that connect their two networks, but *A*'s link has latency 100 ms and *B*'s has latency

120 ms. When routing a message from a source in A's London network to a destination in B's New York network, A may choose to immediately send the message to B in London. This saves A the work of sending it along an expensive trans-Atlantic link, but causes the message to experience latency 125 ms when the other route would have been 20 ms faster.

Additionally, a similar routing challenge can be observed in cellular networks, where different packets are destined for various endpoints, and each link exhibits varying spectral efficiency. In this context, the selection of the *optimal* path involves considering latency and packet error rate. To address this, multiple independent entities, one for each base station, play a crucial role in path selection while striving to optimize overall network performance.^[11]

A 2003 measurement study of Internet routes found that, between pairs of neighboring ISPs, more than 30% of paths have inflated latency due to hot-potato routing, with 5% of paths being delayed by at least 12 ms. Inflation due to AS-level path selection, while substantial, was attributed primarily to BGP's lack of a mechanism to directly optimize for latency, rather than to selfish routing policies. It was also suggested that, were an appropriate mechanism in place, ISPs would be willing to cooperate to reduce latency rather than use hot-potato routing.^[12] Such a mechanism was later published by the same authors, first for the case of two ISPs^[13] and then for the global case.^[14]

Route analytics

As the Internet and IP networks have become mission critical business tools, there has been increased interest in techniques and methods to monitor the routing posture of networks. Incorrect routing or routing issues cause undesirable performance degradation, flapping or downtime. Monitoring routing in a network is achieved using route analytics tools and techniques.^[15]

Centralized routing

In networks where a logically centralized control is available over the forwarding state, for example, using software-defined networking, routing techniques can be used that aim to optimize global and network-wide performance metrics. This has been used by large internet companies that operate many data centers in different geographical locations attached using private optical links, examples of which include Microsoft's Global WAN,^[16] Facebook's Express Backbone,^[17] and Google's B4.^[18]

Global performance metrics to optimize include maximizing network utilization, minimizing traffic flow completion times, maximizing the traffic delivered prior to specific deadlines and reducing the completion times of flows.^[19] Work on the later over private WAN discusses modeling routing as a graph optimization problem by pushing all the queuing to the end-points. The authors also propose a heuristic to solve the problem efficiently while sacrificing negligible performance.^[20]

See also

- Black hole (networking)
- Collective routing
- Deflection routing
- Edge disjoint shortest pair algorithm

- [Flood search routing](#)
- [Fuzzy routing](#)
- [Geographic routing](#)
- [Heuristic routing](#)
- [Path computation element \(PCE\)](#)
- [Policy-based routing](#)
- [Wormhole routing](#)
- [Small-world routing](#)
- [Turn restriction routing](#)

References

1. Goścień, Róża; Walkowiak, Krzysztof; Klinkowski, Mirosław (2015-03-14). "Tabu search algorithm for routing, modulation and spectrum allocation in elastic optical network with anycast and unicast traffic" (<https://www.sciencedirect.com/science/article/pii/S1389128614004460>). *Computer Networks*. **79**: 148–165. doi:[10.1016/j.comnet.2014.12.004](https://doi.org/10.1016/j.comnet.2014.12.004) (<https://doi.org/10.1016%2Fj.comnet.2014.12.004>). ISSN [1389-1286](https://search.worldcat.org/issn/1389-1286) (<https://search.worldcat.org/issn/1389-1286>).
2. RFC 3626
3. RFC 1322 (<https://www.rfc-editor.org/rfc/rfc1322>)
4. Baumann, Rainer; Heimlicher, Simon; Strasser, Mario; Weibel, Andreas (February 10, 2007), *A Survey on Routing Metrics* (<http://rainer.baumann.info/public/tik262.pdf>) (PDF), retrieved 2020-05-04
5. Michael Mitzenmacher; Andréa W. Richa; Ramesh Sitaraman, "Randomized Protocols for Circuit Routing", *The Power of Two Random Choices: A Survey of Techniques and Results* (<http://www.eecs.harvard.edu/~michaelm/postscripts/handbook2001.pdf>) (PDF), p. 34, archived (<https://web.archive.org/web/20231213082735/https://www.eecs.harvard.edu/~michaelm/postscripts/handbook2001.pdf>) (PDF) from the original on Dec 13, 2023
6. Stefan Haas (1998), "The IEEE 1355 Standard: Developments, Performance and Application in High Energy Physics" (<http://inspirehep.net/record/887357/files/cer-002474543.pdf>) (PDF), INSPIRE, p. 15, archived (<https://web.archive.org/web/20190516092754/http://inspirehep.net/record/887357/files/cer-002474543.pdf>) (PDF) from the original on May 16, 2019, "To eliminate network hot spots, ... a two phase routing algorithm. This involves every packet being first sent to a randomly chosen intermediate destination; from the intermediate destination it is forwarded to its final destination. This algorithm, referred to as Universal Routing, is designed to maximize capacity and minimize delay under conditions of heavy load."
7. Noormohammadpour, M.; Raghavendra, C. S. (Apr 2018). "Poster Abstract: Minimizing Flow Completion Times using Adaptive Routing over Inter-Datacenter Wide Area Networks" (<https://www.researchgate.net/publication/323411017>). *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. pp. 1–2. arXiv:[1802.09080](https://arxiv.org/abs/1802.09080) (<https://arxiv.org/abs/1802.09080>). doi:[10.1109/INFOWH.2018.8406853](https://doi.org/10.1109/INFOWH.2018.8406853) (<https://doi.org/10.1109%2FINFOCOMW.2018.8406853>). ISBN [978-1-5386-5979-3](https://doi.org/10.1109/INFOWH.2018.8406853) – via ResearchGate.
8. Noormohammadpour, M; Raghavendra, C. S. (Apr 2018). "Minimizing Flow Completion Times using Adaptive Routing over Inter-Datacenter Wide Area Networks" (<https://www.researchgate.net/publication/323723167>). doi:[10.13140/RG.2.2.36009.90720](https://doi.org/10.13140/RG.2.2.36009.90720) (<https://doi.org/10.13140%2FRG.2.2.36009.90720>) – via ResearchGate.
9. Zutt, Jonne; van Gemund, Arjan J.C.; de Weerdt, Mathijs M.; Witteveen, Cees (2010). "Dealing with Uncertainty in Operational Transport Planning" (<https://web.archive.org/web/20170922001342/http://www.st.ewi.tudelft.nl/~mathijs/publications/intinfra09.pdf>) (PDF). Archived from the original (<http://www.st.ewi.tudelft.nl/~mathijs/publications/intinfra09.pdf>) (PDF) on Sep 22, 2017. In R.R. Negenborn and Z. Lukszo and H. Hellendoorn (Eds.) Intelligent Infrastructures, Ch. 14, pp. 355–382. Springer.

10. Matthew Caesar and Jennifer Rexford. "BGP routing policies in ISP networks (<http://www.cs.princeton.edu/~jrex/papers/policies.pdf>)". IEEE Network Magazine, special issue on Interdomain Routing, Nov/Dec 2005.
11. Shahaf Yamin and Haim H. Permuter. "Multi-agent reinforcement learning for network routing in integrated access backhaul networks (<https://www.sciencedirect.com/science/article/pii/S1570870523002676>)". *Ad Hoc Networks*, Volume 153, 2024, 103347, ISSN 1570-8705 (<https://www.worldcat.org/search?fq=x0:jnln&q=n2:1570-8705>), doi:10.1016/j.adhoc.2023.103347 (<https://doi.org/10.1016%2Fj.adhoc.2023.103347>).
12. Neil Spring, Ratul Mahajan, and Thomas Anderson. "Quantifying the Causes of Path Inflation (<http://www.cs.washington.edu/research/networking/rocketfuel/papers/sigcomm2003.pdf>)". Proc. SIGCOMM 2003.
13. Ratul Mahajan, David Wetherall, and Thomas Anderson. "Negotiation-Based Routing Between Neighboring ISPs (<http://research.microsoft.com/en-us/um/people/ratul/papers/nsdi2005-nexit.pdf>)". Proc. NSDI 2005.
14. Ratul Mahajan, David Wetherall, and Thomas Anderson. Mutually Controlled Routing with Independent ISPs (<http://research.microsoft.com/en-us/um/people/ratul/papers/nsdi2007-wiser.pdf>). Proc. NSDI 2007.
15. Santhi, P.; Ahmed, Md Shakeel; Mehertaj, Sk; Manohar, T. Bharath. *An Efficient Security Way of Authentication and Pair wise Key Distribution with Mobile Sinks in Wireless Sensor Networks*. CiteSeerX 10.1.1.392.151 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.392.151>).
16. Khalidi, Yousef (March 15, 2017). "How Microsoft builds its fast and reliable global network" (<https://azure.microsoft.com/en-us/blog/how-microsoft-builds-its-fast-and-reliable-global-network/>).
17. "Building Express Backbone: Facebook's new long-haul network" (<https://code.fb.com/data-center-engineering/building-express-backbone-facebook-s-new-long-haul-network/>). May 1, 2017.
18. "Inside Google's Software-Defined Network" (<https://web.archive.org/web/20181208111113/https://www.networkcomputing.com/networking/inside-googles-software-defined-network/512240144>). *Network Computing*. May 14, 2017. Archived from the original (<https://www.networkcomputing.com/networking/inside-googles-software-defined-network/512240144>) on December 8, 2018. Retrieved October 8, 2018.
19. Noormohammadpour, Mohammad; Raghavendra, Cauligi (16 July 2018). "Datacenter Traffic Control: Understanding Techniques and Tradeoffs". *IEEE Communications Surveys and Tutorials*. **20** (2): 1492–1525. arXiv:1712.03530 (<https://arxiv.org/abs/1712.03530>). doi:10.1109/COMST.2017.2782753 (<https://doi.org/10.1109%2FCOMST.2017.2782753>). S2CID 28143006 (<https://api.semanticscholar.org/CorpusID:28143006>).
20. Noormohammadpour, Mohammad; Srivastava, Ajitesh; Raghavendra, Cauligi (2018). "On Minimizing the Completion Times of Long Flows over Inter-Datacenter WAN" (<https://www.researchgate.net/publication/328008697>). *IEEE Communications Letters*. **22** (12): 2475–2478. arXiv:1810.00169 (<https://arxiv.org/abs/1810.00169>). Bibcode:2018arXiv181000169N (<https://ui.adsabs.harvard.edu/abs/2018arXiv181000169N>). doi:10.1109/LCOMM.2018.2872980 (<https://doi.org/10.1109%2FLCOMM.2018.2872980>). S2CID 52898719 (<https://api.semanticscholar.org/CorpusID:52898719>).

Further reading

- Ash, Gerald (1997). *Dynamic Routing in Telecommunication Networks*. McGraw–Hill. ISBN 978-0-07-006414-0.
- Doyle, Jeff & Carroll, Jennifer (2005). *Routing TCP/IP, Volume I, Second Ed.* Cisco Press. ISBN 978-1-58705-202-6. Ciscopress (<http://www.ciscopress.com/title/1587052024>) ISBN 1-

58705-202-4

- Doyle, Jeff & Carroll, Jennifer (2001). *Routing TCP/IP, Volume II* (<https://archive.org/details/routingtcpip00jeff>). Cisco Press. ISBN 978-1-57870-089-9. Ciscopress (<http://www.ciscopress.com/title/1578700892>) ISBN 1-57870-089-2
- Huitema, Christian (2000). *Routing in the Internet, Second Ed.* Prentice-Hall. ISBN 978-0-321-22735-5.
- Kurose, James E. & Ross, Keith W. (2004). *Computer Networking, Third Ed.* Benjamin/Cummings. ISBN 978-0-321-22735-5.
- Medhi, Deepankar & Ramasamy, Karthikeyan (2007). *Network Routing: Algorithms, Protocols, and Architectures*. Morgan Kaufmann. ISBN 978-0-12-088588-6.

External links

- Count-To-Infinity Problem (<https://web.archive.org/web/20061209093728/http://wiki.uni.lu/se-can-lab/Count-To-Infinity+Problem.html>)
- "Stability Features" (<https://web.archive.org/web/20150925085557/http://www.lehre.dhbw-stuttgart.de/~schulte/htme/55024.htm#HDR3>). Archived from the original (<http://www.lehre.dhbw-stuttgart.de/~schulte/htme/55024.htm#HDR3>) on 2015-09-25., ways of avoiding the count-to-infinity problem
- Cisco IT Case Studies (https://web.archive.org/web/20070216114435/http://www.cisco.com/web/about/ciscoitatwork/case_studies/routing.html) about Routing and Switching
- "IP Routing and Subnets" (http://www.eventhelix.com/Realtimemantra/Networking/ip_routing.htm). www.eventhelix.com. Retrieved 2018-04-28.

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Routing&oldid=1277203039>"



Transaction-level modeling

Transaction-level modeling (TLM) is an approach to modelling complex digital systems by using electronic design automation software.^{[1]:1955} TLM language (TLML) is a hardware description language, usually, written in C++ and based on SystemC library.^[1] TLMLs are used for modelling where details of communication among modules are separated from the details of the implementation of functional units or of the communication architecture. It's used for modelling of systems that involve complex data communication mechanisms.^{[1]:1955}

Components such as buses or FIFOs are modeled as channels, and are presented to modules using SystemC interface classes. Transaction requests take place by calling interface functions of these channel models, which encapsulate low-level details of the information exchange. At the transaction level, the emphasis is more on the functionality of the data transfers – what data are transferred to and from what locations – and less on their actual implementation, that is, on the actual protocol used for data transfer. This approach makes it easier for the system-level designer to experiment, for example, with different bus architectures (all supporting a common abstract interface) without having to recode models that interact with any of the buses, provided these models interact with the bus through the common interface.^[2]

However, the application of transaction-level modeling is not specific to the SystemC language and can be used with other languages. The concept of TLM first appears in system level language and modeling domain.^[3]

Transaction-level models are used for high-level synthesis of register-transfer level (RTL) models for a lower-level modelling and implementation of system components. RTL is usually represented by a hardware description language source code (e.g. VHDL, SystemC, Verilog).^{[1]:1955–1957}

History

In 2000, Thorsten Grötker, R&D Manager at Synopsys was preparing a presentation on the communication mechanism in what was to become the SystemC 2.0 standard, and referred to it as "transaction-based modeling". Gilles Baillieu, then a corporate application engineer at Synopsys, insisted that the new term had to contain "level", as in "register-transfer level" or "behavioral level". The fact that TLM does not denote a single level of abstraction but rather a modeling technique didn't make him change his mind. It had to be "level" in order to make it stick. So it became "TLM".

The Open SystemC Initiative was formed to standardize and proliferate the use of the SystemC language. That organization is sponsored by major EDA vendors and customers sharing a common interest in facilitating tool development and IP interoperability. The organization developed the OSCI simulator for open use and distribution.

Since those early days SystemC has been adopted as the language of choice for high level synthesis, connecting the design modeling and virtual prototype application domains with the functional verification and automated path gate level implementation. This offers project teams the ability to produce one model

for multiple purposes. At the 2010 DVCon event, OSCI produced a specification of the first synthesizable subset of SystemC for industry standardization.

See also

- [Discrete event simulation \(DES\)](#)
- [Event loop](#)
- [Event-driven programming](#)
- [Message passing](#)
- [Reactor pattern vs. Proactor pattern](#)
- [Transaction processing](#)
- [Asynchronous circuit](#)
- [Assembly modelling, for CADs](#)

References

1. *The VLSI handbook* (<https://www.worldcat.org/oclc/70699056>). Wai-Kai Chen (2 ed.). Boca Raton, FL: CRC/Taylor & Francis. 2007. [ISBN 978-0-8493-4199-1](#). OCLC 70699056 (<https://search.worldcat.org/oclc/70699056>).
2. T. Grötker, S. Liao, G. Martin, S. Swan, System Design with SystemC. Springer, 2002, Chapter 8., pp. 131. [ISBN 1-4020-7072-1](#) (quoted with permission)
3. L. Cai, D. Gajski, Transaction Level Modeling: An Overview, in proceedings of the Int. Conference on HW/SW Codesign and System Synthesis (CODES-ISSS), Oct. 2003, pp. 19–24

External links

- [SystemC.org](https://systemc.org/) (<https://systemc.org/>) - SystemC home page.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Transaction-level_modeling&oldid=1156311674"



Formal equivalence checking

Formal equivalence checking process is a part of electronic design automation (EDA), commonly used during the development of digital integrated circuits, to formally prove that two representations of a circuit design exhibit exactly the same behavior.

Equivalence checking and levels of abstraction

In general, there is a wide range of possible definitions of functional equivalence covering comparisons between different levels of abstraction and varying granularity of timing details.

- The most common approach is to consider the problem of machine equivalence which defines two synchronous design specifications functionally equivalent if, clock by clock, they produce *exactly* the same sequence of output signals for *any* valid sequence of input signals.
- Microprocessor designers use equivalence checking to compare the functions specified for the instruction set architecture (ISA) with a register transfer level (RTL) implementation, ensuring that any program executed on both models will cause an identical update of the main memory content. This is a more general problem.
- A system design flow requires comparison between a transaction level model (TLM), e.g., written in SystemC and its corresponding RTL specification. Such a check is becoming of increasing interest in a system-on-a-chip (SoC) design environment.

Synchronous machine equivalence

The register transfer level (RTL) behavior of a digital chip is usually described with a hardware description language, such as Verilog or VHDL. This description is the golden reference model that describes in detail which operations will be executed during which clock cycle and by which pieces of hardware. Once the logic designers, by simulations and other verification methods, have verified register transfer description, the design is usually converted into a netlist by a logic synthesis tool. Equivalence is not to be confused with functional correctness, which must be determined by functional verification.

The initial netlist will usually undergo a number of transformations such as optimization, addition of Design For Test (DFT) structures, etc., before it is used as the basis for the placement of the logic elements into a physical layout. Contemporary physical design software will occasionally also make significant modifications (such as replacing logic elements with equivalent similar elements that have a higher or lower drive strength and/or area) to the netlist. Throughout every step of a very complex, multi-step procedure, the original functionality and the behavior described by the original code must be maintained. When the final tape-out is made of a digital chip, many different EDA programs and possibly some manual edits will have altered the netlist.

In theory, a logic synthesis tool guarantees that the first netlist is logically equivalent to the RTL source code. All the programs later in the process that make changes to the netlist also, in theory, ensure that these changes are logically equivalent to a previous version.

In practice, programs have bugs and it would be a major risk to assume that all steps from RTL through the final tape-out netlist have been performed without error. Also, in real life, it is common for designers to make manual changes to a netlist, commonly known as Engineering Change Orders, or ECOs, thereby introducing a major additional error factor. Therefore, instead of blindly assuming that no mistakes were made, a verification step is needed to check the logical equivalence of the final version of the netlist to the original description of the design (golden reference model).

Historically, one way to check the equivalence was to re-simulate, using the final netlist, the test cases that were developed for verifying the correctness of the RTL. This process is called gate level logic simulation. However, the problem with this is that the quality of the check is only as good as the quality of the test cases. Also, gate-level simulations are notoriously slow to execute, which is a major problem as the size of digital designs continues to grow exponentially.

An alternative way to solve this is to formally prove that the RTL code and the netlist synthesized from it have exactly the same behavior in all (relevant) cases. This process is called formal equivalence checking and is a problem that is studied under the broader area of formal verification.

A formal equivalence check can be performed between any two representations of a design: RTL \leftrightarrow netlist, netlist \leftrightarrow netlist or RTL \leftrightarrow RTL, though the latter is rare compared to the first two. Typically, a formal equivalence checking tool will also indicate with great precision at which point there exists a difference between two representations.

Methods

There are two basic technologies used for boolean reasoning in equivalence checking programs:

- Binary decision diagrams, or BDDs: A specialized data structure designed to support reasoning about boolean functions. BDDs have become highly popular because of their efficiency and versatility.
- Conjunctive Normal Form Satisfiability: SAT solvers returns an assignment to the variables of a propositional formula that satisfies it if such an assignment exists. Almost any boolean reasoning problem can be expressed as a SAT problem.

Commercial applications for equivalence checking

Major products in the Logic Equivalence Checking (LEC) area of EDA are:

- FormalPro by Mentor Graphics
- Questa SLEC by Mentor Graphics
- Conformal by Cadence
- Jasper by Cadence
- Formality by Synopsys
- VC Formal by Synopsys
- 360 EC by OneSpin Solutions
- ATEC by ATEC

Generalizations

- Equivalence Checking of Retimed Circuits: Sometimes it is helpful to move logic from one side of a register to another, and this complicates the checking problem.
- Sequential Equivalence Checking: Sometimes, two machines are completely different at the combinational level, but should give the same outputs if given the same inputs. The classic example is two identical state machines with different encodings for the states. Since this cannot be reduced to a combinational problem, more general techniques are required.
- Equivalence of Software Programs, i.e. checking if two well-defined programs that take N inputs and produce M outputs are equivalent: Conceptually, you can turn software into a state machine (that's what the combination of a compiler does, since a computer plus its memory form a very large state machine.) Then, in theory, various forms of property checking can ensure they produce the same output. This problem is even harder than sequential equivalence checking, since the outputs of the two programs may appear at different times; but it is possible, and researchers are working on it.

See also

- [Formal methods](#)

References

- *Electronic Design Automation For Integrated Circuits Handbook*, by Lavagno, Martin, and Scheffer, ISBN 0-8493-3096-3 A survey of the field. This article was derived, with permission, from Volume 2, Chapter 4, *Equivalence Checking*, by Fabio Somenzi and Andreas Kuehlmann.
- R.E. Bryant, *Graph-based algorithms for Boolean function manipulation* (<https://apps.dtic.mil/sti/pdfs/ADA470446.pdf>), IEEE Transactions on Computers., C-35, pp. 677–691, 1986. The original reference on BDDs.
- Sequential equivalence checking for RTL models. Nikhil Sharma, Gagan Hasteer and Venkat Krishnaswamy. *EE Times* (http://www.eetimes.com/document.asp?doc_id=1271433).

External links

- CADP – provides equivalence checking tools for asynchronous designs (<http://cadp.inria.fr>)
- OneSpin 360 EC-FPGA – Functional correctness of FPGA synthesis from RTL code to final netlist (<https://www.onespin.com/products/360-ec-fpga/>)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Formal_equivalence_checking&oldid=1220785137"



Synchronous circuit

(Redirected from [Synchronous logic](#))

In [digital electronics](#), a **synchronous circuit** is a digital circuit in which the changes in the [state](#) of memory elements are synchronized by a [clock signal](#). In a [sequential digital logic](#) circuit, data is stored in memory devices called [flip-flops](#) or latches. The output of a flip-flop is constant until a pulse is applied to its "clock" input, upon which the input of the flip-flop is latched into its output. In a synchronous logic circuit, an [electronic oscillator](#) called the [clock](#) generates a string (sequence) of pulses, the "clock signal". This clock signal is applied to every storage element, so in an ideal synchronous circuit, every change in the [logical levels](#) of its storage components is simultaneous. Ideally, the input to each storage element has reached its final value before the next clock occurs, so the behaviour of the whole circuit can be predicted exactly. Practically, some delay is required for each logical operation, resulting in a maximum speed limitations at which each synchronous system can run.

To make these circuits work correctly, a great deal of care is needed in the design of the [clock distribution networks](#). [Static timing analysis](#) is often used to determine the maximum safe operating speed.

Nearly all digital circuits, and in particular nearly all CPUs, are fully synchronous circuits with a global clock. Exceptions are often compared to fully synchronous circuits. Exceptions include self-synchronous circuits,^{[1][2][3][4]} [globally asynchronous locally synchronous](#) circuits, and fully [asynchronous circuits](#).

See also

- [Synchronous network](#)
- [Asynchronous circuit](#)
- [Moore machine](#)
- [Mealy machine](#)
- [Finite-state machine](#)
- [Sequential logic](#)
- [Memory](#)
- [Control unit](#)
- [Arithmetic logic unit](#)
- [Processor register](#)
- [Application-specific integrated circuit \(ASIC\)](#)

References

1. Asada and Ikeda Laboratories. "Self-synchronous Circuit" (<http://silicon.u-tokyo.ac.jp/en/research/>). "Self Synchronous FPGA". 2009.
2. "self synchronous configurable logic blocks" (https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=Search_Index_Terms:.QT.self%20synchronous%20configurable%20logic%20blocks.QT.&%20newsearch=partialPref).

3. Devlin, Benjamin; Ikeda, Makoto; Asada, Kunihiro. "Energy Minimum Operation with Self Synchronous Gate-Level Autonomous Power Gating and Voltage Scaling" (<http://adsabs.harvard.edu/abs/2012IEITE..95..546D>). 2012. doi:[10.1587/transele.E95.C.546](https://doi.org/10.1587/transele.E95.C.546) (<https://doi.org/10.1587%2Ftransele.E95.C.546>)
 4. Devlin, B.; Ueki, H.; Mori, S.; Miyauchi, S.; Ikeda, M.; Asada, K. "Performance and side-channel attack analysis of a self synchronous montgomery multiplier processing element for RSA in 40nm CMOS" (https://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6570807). 2012. doi:[10.1109/ASSCC.2012.6570807](https://doi.org/10.1109/ASSCC.2012.6570807) (<https://doi.org/10.1109%2FASSCC.2012.6570807>)
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Synchronous_circuit&oldid=1264122148"



Asynchronous circuit

(Redirected from Asynchronous logic)

Asynchronous circuit (clockless or self-timed circuit)^[1]:Lecture 12 [note 1][2]:157–186 is a sequential digital logic circuit that does not use a global clock circuit or signal generator to synchronize its components.^{[1][3]:3–5} Instead, the components are driven by a handshaking circuit which indicates a completion of a set of instructions. Handshaking works by simple data transfer protocols.^{[3]:115} Many synchronous circuits were developed in early 1950s as part of bigger asynchronous systems (e.g. ORDVAC). Asynchronous circuits and theory surrounding is a part of several steps in integrated circuit design, a field of digital electronics engineering.

Asynchronous circuits are contrasted with synchronous circuits, in which changes to the signal values in the circuit are triggered by repetitive pulses called a clock signal. Most digital devices today use synchronous circuits. However asynchronous circuits have a potential to be much faster, have a lower level of power consumption, electromagnetic interference, and better modularity in large systems. Asynchronous circuits are an active area of research in digital logic design.^{[4][5]}

It was not until the 1990s when viability of the asynchronous circuits was shown by real-life commercial products.^{[3]:4}

Overview

All digital logic circuits can be divided into combinational logic, in which the output signals depend only on the current input signals, and sequential logic, in which the output depends both on current input and on past inputs. In other words, sequential logic is combinational logic with memory. Virtually all practical digital devices require sequential logic. Sequential logic can be divided into two types, synchronous logic and asynchronous logic.

Synchronous circuits

In synchronous logic circuits, an electronic oscillator generates a repetitive series of equally spaced pulses called the clock signal. The clock signal is supplied to all the components of the IC. Flip-flops only flip when triggered by the edge of the clock pulse, so changes to the logic signals throughout the circuit begin at the same time and at regular intervals. The output of all memory elements in a circuit is called the state of the circuit. The state of a synchronous circuit changes only on the clock pulse. The changes in signal require a certain amount of time to propagate through the combinational logic gates of the circuit. This time is called a propagation delay.

As of 2021, timing of modern synchronous ICs takes significant engineering efforts and sophisticated design automation tools.^[6] Designers have to ensure that clock arrival is not faulty. With the ever-growing size and complexity of ICs (e.g. ASICs) it's a challenging task.^[6] In huge circuits, signals sent over clock distribution network often end up at different times at different parts.^[6] This problem is widely known as "clock skew".^{[6][7]:xiv}

The maximum possible clock rate is capped by the logic path with the longest propagation delay, called the critical path. Because of that, the paths that may operate quickly are idle most of the time. A widely distributed clock network dissipates a lot of useful power and must run whether the circuit is receiving inputs or not.^[6] Because of this level of complexity, testing and debugging takes over half of development time in all dimensions for synchronous circuits.^[6]

Asynchronous circuits

The asynchronous circuits do not need a global clock, and the state of the circuit changes as soon as the inputs change. The local functional blocks may be still employed but the clock skew problem still can be tolerated.^{[7]:xiv[3]:4}

Since asynchronous circuits do not have to wait for a clock pulse to begin processing inputs, they can operate faster. Their speed is theoretically limited only by the propagation delays of the logic gates and other elements.^{[7]:xiv}

However, asynchronous circuits are more difficult to design and subject to problems not found in synchronous circuits. This is because the resulting state of an asynchronous circuit can be sensitive to the relative arrival times of inputs at gates. If transitions on two inputs arrive at almost the same time, the circuit can go into the wrong state depending on slight differences in the propagation delays of the gates.

This is called a race condition. In synchronous circuits this problem is less severe because race conditions can only occur due to inputs from outside the synchronous system, called *asynchronous inputs*.

Although some fully asynchronous digital systems have been built (see below), today asynchronous circuits are typically used in a few critical parts of otherwise synchronous systems where speed is at a premium, such as signal processing circuits.

Theoretical foundation

The original theory of asynchronous circuits was created by David E. Muller in mid-1950s.^[8] This theory was presented later in the well-known book "Switching Theory" by Raymond Miller.^[9]

The term "asynchronous logic" is used to describe a variety of design styles, which use different assumptions about circuit properties.^[10] These vary from the bundled delay model – which uses "conventional" data processing elements with completion indicated by a locally generated delay model – to delay-insensitive design – where arbitrary delays through circuit elements can be accommodated. The latter style tends to yield circuits which are larger than bundled data implementations, but which are insensitive to layout and parametric variations and are thus "correct by design".

Asynchronous logic

Asynchronous logic is the logic required for the design of asynchronous digital systems. These function without a clock signal and so individual logic elements cannot be relied upon to have a discrete true/false state at any given time. Boolean (two valued) logic is inadequate for this and so extensions are required.

Since 1984, Vadim O. Vasyukevich developed an approach based upon new logical operations which he called *venjunction* (with asynchronous operator " $x \angle y$ " standing for "switching x on the background y " or "if x when y then") and *sequention* (with priority signs " $x_i > x_j$ " and " $x_i < x_j$ "). This takes into account not only the current value of an element, but also its history.^{[11][12][13][14][15]}

Karl M. Fant developed a different theoretical treatment of asynchronous logic in his work *Logically determined design* in 2005 which used four-valued logic with null and intermediate being the additional values. This architecture is important because it is quasi-delay-insensitive.^{[16][17]} Scott C. Smith and Jia Di developed an ultra-low-power variation of Fant's Null Convention Logic that incorporates multi-threshold CMOS.^[18] This variation is termed Multi-threshold Null Convention Logic (MTNCL), or alternatively Sleep Convention Logic (SCL).^[19]

Petri nets

Petri nets are an attractive and powerful model for reasoning about asynchronous circuits (see Subsequent models of concurrency). A particularly useful type of interpreted Petri nets, called Signal Transition Graphs (STGs), was proposed independently in 1985 by Leonid Rosenblum and Alex Yakovlev^[20] and Tam-Anh Chu.^[21] Since then, STGs have been studied extensively in theory and practice,^{[22][23]} which has led to the development of popular software tools for analysis and synthesis of asynchronous control circuits, such as Petrify^[24] and Workcraft.^[25]

Subsequent to Petri nets other models of concurrency have been developed that can model asynchronous circuits including the Actor model and process calculi.

Benefits

A variety of advantages have been demonstrated by asynchronous circuits. Both quasi-delay-insensitive (QDI) circuits (generally agreed to be the most "pure" form of asynchronous logic that retains computational universality) and less pure forms of asynchronous circuitry which use timing constraints for higher performance and lower area and power present several advantages.

- Robust and cheap handling of metastability of arbiters.
- Average-case performance: an average-case time (delay) of operation is not limited to the worst-case completion time of component (gate, wire, block etc.) as it is in synchronous circuits.^{[7]:xiv[3]:3} This results in better latency and throughput performance.^{[26]:9[3]:3} Examples include speculative completion^{[27][28]} which has been applied to design parallel prefix adders faster than synchronous ones, and a high-performance double-precision floating point adder^[29] which outperforms leading synchronous designs.
- Early completion: the output may be generated ahead of time, when result of input processing is predictable or irrelevant.
- Inherent elasticity: variable number of data items may appear in pipeline inputs at any time (pipeline means a cascade of linked functional blocks). This contributes to high performance while gracefully handling variable input and output rates due to unclocked pipeline stages (functional blocks) delays (congestions may still be possible however and input-output gates delay should be also taken into account^{[30]:194}).^[26]
- No need for timing-matching between functional blocks either. Though given different delay models (predictions of gate/wire delay times) this depends on actual approach of

asynchronous circuit implementation.^{[30]:194}

- Freedom from the ever-worsening difficulties of distributing a high-fan-out, timing-sensitive clock signal.
- Circuit speed adapts to changing temperature and voltage conditions rather than being locked at the speed mandated by worst-case assumptions.^{[3]:3}
- Lower, on-demand power consumption;^{[7]:xiv[26]:9[3]:3} zero standby power consumption.^{[3]:3} In 2005 Epson has reported 70% lower power consumption compared to synchronous design.^[31] Also, clock drivers can be removed which can significantly reduce power consumption. However, when using certain encodings, asynchronous circuits may require more area, adding similar power overhead if the underlying process has poor leakage properties (for example, deep submicrometer processes used prior to the introduction of high-k dielectrics).
 - No need for power-matching between local asynchronous functional domains of circuitry. Synchronous circuits tend to draw a large amount of current right at the clock edge and shortly thereafter. The number of nodes switching (and hence, the amount of current drawn) drops off rapidly after the clock edge, reaching zero just before the next clock edge. In an asynchronous circuit, the switching times of the nodes does not correlate in this manner, so the current draw tends to be more uniform and less bursty.
 - Robustness toward transistor-to-transistor variability in the manufacturing transfer process (which is one of the most serious problems facing the semiconductor industry as dies shrink), variations of voltage supply, temperature, and fabrication process parameters.^{[3]:3}
 - Less severe electromagnetic interference (EMI).^{[3]:3} Synchronous circuits create a great deal of EMI in the frequency band at (or very near) their clock frequency and its harmonics; asynchronous circuits generate EMI patterns which are much more evenly spread across the spectrum.^{[3]:3}
 - Design modularity (reuse), improved noise immunity and electromagnetic compatibility. Asynchronous circuits are more tolerant to process variations and external voltage fluctuations.^{[3]:4}

Disadvantages

- Area overhead caused by additional logic implementing handshaking.^{[3]:4} In some cases an asynchronous design may require up to double the resources (area, circuit speed, power consumption) of a synchronous design, due to addition of completion detection and design-for-test circuits.^{[32][3]:4}
- Compared to a synchronous design, as of the 1990s and early 2000s not many people are trained or experienced in the design of asynchronous circuits.^[32]
- Synchronous designs are inherently easier to test and debug than asynchronous designs.^[33] However, this position is disputed by Fant, who claims that the apparent simplicity of synchronous logic is an artifact of the mathematical models used by the common design approaches.^[17]
- Clock gating in more conventional synchronous designs is an approximation of the asynchronous ideal, and in some cases, its simplicity may outweigh the advantages of a fully asynchronous design.
- Performance (speed) of asynchronous circuits may be reduced in architectures that require input-completeness (more complex data path).^[34]
- Lack of dedicated, asynchronous design-focused commercial EDA tools.^[34] As of 2006 the situation was slowly improving, however.^{[3]:x}

Communication

There are several ways to create asynchronous communication channels that can be classified by their protocol and data encoding.

Protocols

There are two widely used protocol families which differ in the way communications are encoded:

- **two-phase handshake** (also known as two-phase protocol, Non-Return-to-Zero (NRZ) encoding, or transition signaling): Communications are represented by any wire transition; transitions from 0 to 1 and from 1 to 0 both count as communications.
- **four-phase handshake** (also known as four-phase protocol, or Return-to-Zero (RZ) encoding): Communications are represented by a wire transition followed by a reset; a transition sequence from 0 to 1 and back to 0 counts as single communication.

Despite involving more transitions per communication, circuits implementing four-phase protocols are usually faster and simpler than two-phase protocols because the signal lines return to their original state by the end of each communication. In two-phase protocols, the circuit implementations would have to store the state of the signal line internally.

Note that these basic distinctions do not account for the wide variety of protocols. These protocols may encode only requests and acknowledgements or also encode the data, which leads to the popular multi-wire data encoding. Many other, less common protocols have been proposed including using a single wire for request and acknowledgment, using several significant voltages, using only pulses or balancing timings in order to remove the latches.

Data encoding

There are two widely used data encodings in asynchronous circuits: bundled-data encoding and multi-rail encoding

Another common way to encode the data is to use multiple wires to encode a single digit: the value is determined by the wire on which the event occurs. This avoids some of the delay assumptions necessary with bundled-data encoding, since the request and the data are not separated anymore.

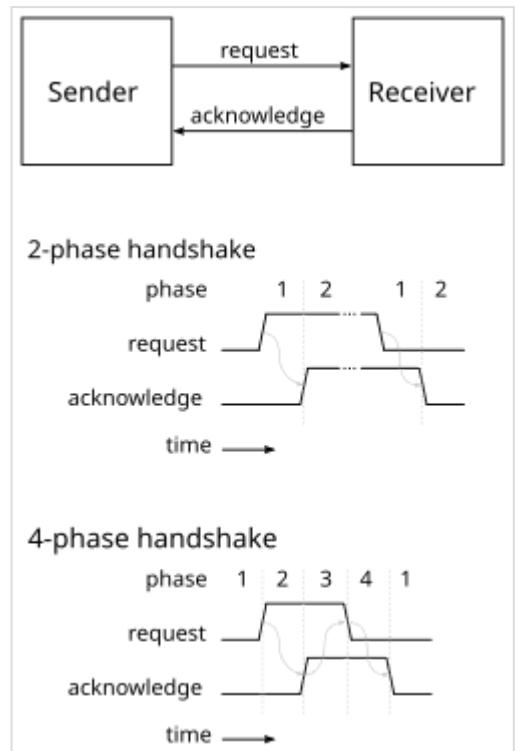


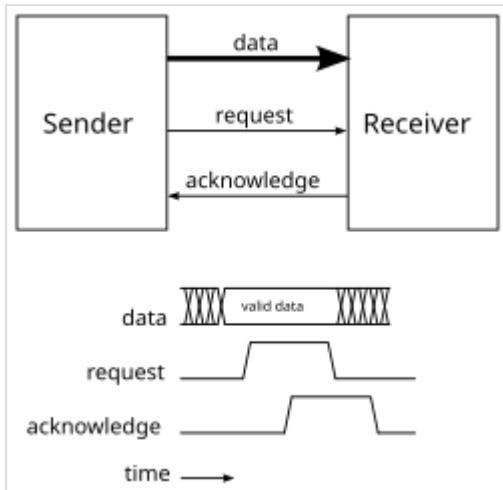
Illustration of two and four-phase handshakes. Top: A sender and a receiver are communicating with simple request and acknowledge signals. The sender drives the request line, and the receiver drives the acknowledge line. Middle: Timing diagram of two, two-phase communications. Bottom: Timing diagram of one, four-phase communication.

Bundled-data encoding

Bundled-data encoding uses one wire per bit of data with a request and an acknowledge signal; this is the same encoding used in synchronous circuits without the restriction that transitions occur on a clock edge. The request and the acknowledge are sent on separate wires with one of the above protocols. These circuits usually assume a bounded delay model with the completion signals delayed long enough for the calculations to take place.

In operation, the sender signals the availability and validity of data with a request. The receiver then indicates completion with an acknowledgement, indicating that it is able to process new requests. That is, the request is bundled with the data, hence the name "bundled-data".

Bundled-data circuits are often referred to as micropipelines, whether they use a two-phase or four-phase protocol, even if the term was initially introduced for two-phase bundled-data.



A 4-phase, bundled-data communication. Top: A sender and receiver are connected by data lines, a request line, and an acknowledge line. Bottom: Timing diagram of a bundled data communication. When the request line is low, the data is to be considered invalid and liable to change at any time.

Multi-rail encoding

Multi-rail encoding uses multiple wires without a one-to-one relationship between bits and wires and a separate acknowledge signal. Data availability is indicated by the transitions themselves on one or more of the data wires (depending on the type of multi-rail encoding) instead of with a request signal as in the bundled-data encoding. This provides the advantage that the data communication is delay-insensitive. Two common multi-rail encodings are one-hot and dual rail. The one-hot (also known as 1-of-n) encoding represents a number in base n with a communication on one of the n wires. The dual-rail encoding uses pairs of wires to represent each bit of the data, hence the name "dual-rail"; one wire in the pair represents the bit value of 0 and the other represents the bit value of 1. For example, a dual-rail encoded two bit number will be represented with two pairs of wires for four wires in total. During a data communication, communications occur on one of each pair of wires to indicate the data's bits. In the general case, an $m \times n$ encoding represents data as m words of base n.

Dual-rail encoding

Dual-rail encoding with a four-phase protocol is the most common and is also called *three-state encoding*, since it has two valid states (10 and 01, after a transition) and a reset state (00). Another common encoding, which leads to a simpler implementation than one-hot, two-phase dual-rail is *four-state encoding*, or level-encoded dual-rail, and uses a data bit and a parity bit to achieve a two-phase protocol.

Asynchronous CPU

Asynchronous CPUs are one of several ideas for radically changing CPU design.

Unlike a conventional processor, a clockless processor (asynchronous CPU) has no central clock to coordinate the progress of data through the pipeline. Instead, stages of the CPU are coordinated using logic devices called "pipeline controls" or "FIFO sequencers". Basically, the pipeline controller clocks the next stage of logic when the existing stage is complete. In this way, a central clock is unnecessary. It may actually be even easier to implement high performance devices in asynchronous, as opposed to clocked, logic:

- components can run at different speeds on an asynchronous CPU; all major components of a clocked CPU must remain synchronized with the central clock;
- a traditional CPU cannot "go faster" than the expected worst-case performance of the slowest stage/instruction/component. When an asynchronous CPU completes an operation more quickly than anticipated, the next stage can immediately begin processing the results, rather than waiting for synchronization with a central clock. An operation might finish faster than normal because of attributes of the data being processed (e.g., multiplication can be very fast when multiplying by 0 or 1, even when running code produced by a naive compiler), or because of the presence of a higher voltage or bus speed setting, or a lower ambient temperature, than 'normal' or expected.

Asynchronous logic proponents believe these capabilities would have these benefits:

- lower power dissipation for a given performance level, and
- highest possible execution speeds.

The biggest disadvantage of the clockless CPU is that most CPU design tools assume a clocked CPU (i.e., a synchronous circuit). Many tools "enforce synchronous design practices".^[35] Making a clockless CPU (designing an asynchronous circuit) involves modifying the design tools to handle clockless logic and doing extra testing to ensure the design avoids metastable problems. The group that designed the AMULET, for example, developed a tool called LARD^[36] to cope with the complex design of AMULET3.

Examples

Despite all the difficulties numerous asynchronous CPUs have been built.

The ORDVAC of 1951 was a successor to the ENIAC and the first asynchronous computer ever built.^{[37][38]}

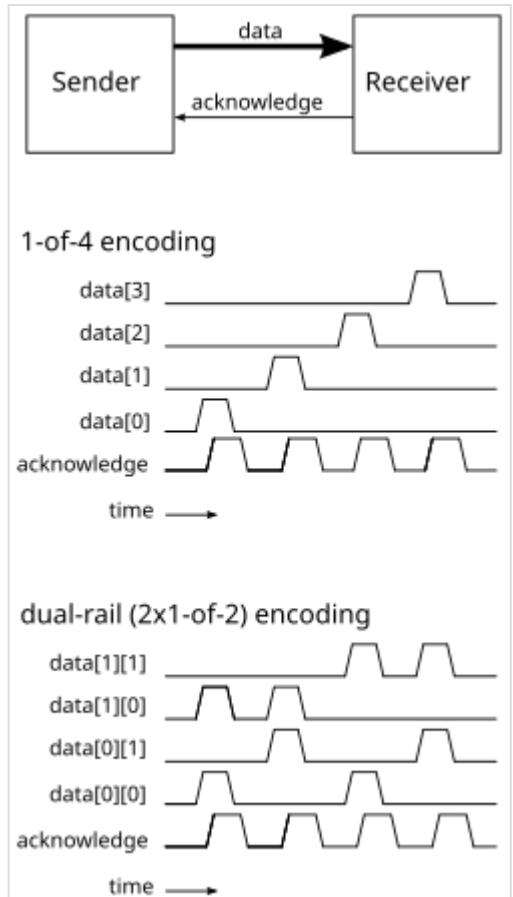


Diagram of dual rail and 1-of-4 communications. Top: A sender and receiver are connected by data lines and an acknowledge line. Middle: Timing diagram of the sender communicating the values 0, 1, 2, and then 3 to the receiver with the 1-of-4 encoding. Bottom: Timing diagram of the sender communicating the same values to the receiver with the dual-rail encoding. For this particular data size, the dual rail encoding is the same as a 2x1-of-2 encoding.

The ILLIAC II was the first completely asynchronous, speed independent processor design ever built; it was the most powerful computer at the time.^[37]

DEC PDP-16 Register Transfer Modules (ca. 1973) allowed the experimenter to construct asynchronous, 16-bit processing elements. Delays for each module were fixed and based on the module's worst-case timing.

Caltech

Since the mid-1980s, Caltech has designed four non-commercial CPUs in attempt to evaluate performance and energy efficiency of the asynchronous circuits.^{[39][40]}

Caltech Asynchronous Microprocessor (CAM)

In 1988 the Caltech Asynchronous Microprocessor (CAM) was the first asynchronous, quasi delay-insensitive (QDI) microprocessor made by Caltech.^{[39][41]} The processor had 16-bit wide RISC ISA and separate instruction and data memories.^[39] It was manufactured by MOSIS and funded by DARPA. The project was supervised by the Office of Naval Research, the Army Research Office, and the Air Force Office of Scientific Research.^{[39]:12}

During demonstrations, the researchers loaded a simple program which ran in a tight loop, pulsing one of the output lines after each instruction. This output line was connected to an oscilloscope. When a cup of hot coffee was placed on the chip, the pulse rate (the effective "clock rate") naturally slowed down to adapt to the worsening performance of the heated transistors. When liquid nitrogen was poured on the chip, the instruction rate shot up with no additional intervention. Additionally, at lower temperatures, the voltage supplied to the chip could be safely increased, which also improved the instruction rate – again, with no additional configuration.

When implemented in gallium arsenide (HGaAs_3) it was claimed to achieve 100MIPS.^{[39]:5} Overall, the research paper interpreted the resultant performance of CAM as superior compared to commercial alternatives available at the time.^{[39]:5}

MiniMIPS

In 1998 the MiniMIPS, an experimental, asynchronous MIPS I-based microcontroller was made. Even though its SPICE-predicted performance was around 280 MIPS at 3.3 V the implementation suffered from several mistakes in layout (human mistake) and the results turned out to be lower by about 40% (see table).^{[39]:5}

The Lutonium 8051

Made in 2003, it was a quasi delay-insensitive asynchronous microcontroller designed for energy efficiency.^{[40][39]:9} The microcontroller's implementation followed the Harvard architecture.^[40]

Performance comparison of the Caltech CPUs (in MIPS). [note 2]

Name	Year	Word size (bits)	Transistors (thousands)	Size (mm)	Node size (μm)	1.5V	2V	3.3V	5V	10V
CAM SCMOS	1988	16	20	N/A	1.6	N/A	5	N/A	18	26
MiniMIPS CMOS	1998	32	2000	8×14	0.6	60	100	180	N/A	N/A
Lutonium 8051 CMOS	2003	8	N/A	N/A	0.18	200	N/A	N/A	N/A	4

Epson

In 2004, Epson manufactured the world's first bendable microprocessor called ACT11, an 8-bit asynchronous chip.^{[42][43][44][45][46]} Synchronous flexible processors are slower, since bending the material on which a chip is fabricated causes wild and unpredictable variations in the delays of various transistors, for which worst-case scenarios must be assumed everywhere and everything must be clocked at worst-case speed. The processor is intended for use in smart cards, whose chips are currently limited in size to those small enough that they can remain perfectly rigid.

IBM

In 2014, IBM announced a SyNAPSE-developed chip that runs in an asynchronous manner, with one of the highest transistor counts of any chip ever produced. IBM's chip consumes orders of magnitude less power than traditional computing systems on pattern recognition benchmarks.^[47]

Timeline

- ORDVAC and the (identical) ILLIAC I (1951)^{[37][38]}
- Johnniac (1953)^[48]
- WEIZAC (1955)
- Kiev (1958), a Soviet machine using the programming language with pointers much earlier than they came to the PL/1 language^[49]
- ILLIAC II (1962)^[37]
- Victoria University of Manchester built Atlas (1964)
- ICL 1906A and 1906S mainframe computers, part of the 1900 series and sold from 1964 for over a decade by ICL^[50]
- Polish computers KAR-65 and K-202 (1965 and 1970 respectively)
- Honeywell CPUs 6180 (1972)^[51] and Series 60 Level 68 (1981)^{[52][53]} upon which Multics ran asynchronously
- Soviet bit-slice microprocessor modules (late 1970s)^{[54][55]} produced as K587,^[56] K588^[57] and K1883 (U83x in East Germany)^[58]
- Caltech Asynchronous Microprocessor, the world-first asynchronous microprocessor (1988)^{[39][41]}
- ARM-implementing AMULET (1993 and 2000)
- Asynchronous implementation of MIPS R3000, dubbed MiniMIPS (<https://web.archive.org/web/20080509090359/http://www.async.caltech.edu/mips.html>) (1998)

- Several versions of the XAP processor experimented with different asynchronous design styles: a bundled data XAP, a 1-of-4 XAP, and a 1-of-2 (dual-rail) XAP (2003?)^[59]
- ARM-compatible processor (2003?) designed by Z. C. Yu, S. B. Furber, and L. A. Plana; "designed specifically to explore the benefits of asynchronous design for security sensitive applications"^[59]
- SAMIPS (2003), a synthesisable asynchronous implementation of the MIPS R3000 processor^{[60][61]}
- "Network-based Asynchronous Architecture" processor (2005) that executes a subset of the MIPS architecture instruction set^[59]
- ARM996HS processor (2006) from Handshake Solutions
- HT80C51 processor (2007?) from Handshake Solutions.^[62]
- Vortex, a superscalar general purpose CPU with a load/store architecture from Intel (2007);^[63] it was developed as Fulcrum Microsystem test Chip 2 and was not commercialized, excepting some of its components; the chip included DDR SDRAM and a 10Gb Ethernet interface linked via Nexus system-on-chip net to the CPU^{[63][64]}
- SEAforth multi-core processor (2008) from Charles H. Moore^[65]
- GA144^[66] multi-core processor (2010) from Charles H. Moore
- TAM16: 16-bit asynchronous microcontroller IP core (Tiempo)^[67]
- Aspida asynchronous DLX core;^[68] the asynchronous open-source DLX processor (ASPIDA) has been successfully implemented both in ASIC and FPGA versions^[69]

See also

- Adiabatic logic
- Event camera (asynchronous camera)
- Perfect clock gating
- Petri nets
- Sequential logic (asynchronous)
- Signal transition graphs
- Transputer – Series of pioneering microprocessors from the 1980s

Notes

1. Globally asynchronous locally synchronous circuits are possible.
2. Dhrystone was also used.^{[39]:4,8}

References

1. Horowitz, Mark (2007). "Advanced VLSI Circuit Design Lecture" (<https://web.stanford.edu/class/archive/ee/ee371/ee371.1066/>). Stanford University, Computer Systems Laboratory. Archived (<https://web.archive.org/web/20160421093147/http://web.stanford.edu:80/class/archive/ee/ee371/ee371.1066/>) from the original on 2016-04-21.
2. Staunstrup, Jørgen (1994). *A Formal Approach to Hardware Design* (<https://www.worldcat.org/oclc/852790160>). Boston, Massachusetts, USA: Springer USA. ISBN 978-1-4615-2764-0. OCLC 852790160 (<https://search.worldcat.org/oclc/852790160>).

3. Sparsø, Jens (April 2006). "Asynchronous Circuit Design A Tutorial" (<https://orbit.dtu.dk/files/2775719/imm855.pdf>) (PDF). Technical University of Denmark.
4. Nowick, S. M.; Singh, M. (May–June 2015). "Asynchronous Design — Part 1: Overview and Recent Advances" (<https://wayback.archive-it.org/all/20181221132620/http://www.cs.columbia.edu/%7Enowick/nowick%2Dsingh%2Dasync%2DIEEE%2DDT%2D15%2Doverview%2Darticle%2Dpt1.pdf>) (PDF). *IEEE Design and Test*. **32** (3): 5–18. doi:10.1109/MDAT.2015.2413759 (<https://doi.org/10.1109%2FMDAT.2015.2413759>). S2CID 14644656 (<https://api.semanticscholar.org/CorpusID:14644656>). Archived from the original (<https://www.cs.columbia.edu/~nowick/nowick-singh-async-IEEE-DT-15-overview-article-pt1.pdf>) (PDF) on 2018-12-21. Retrieved 2019-08-27.
5. Nowick, S. M.; Singh, M. (May–June 2015). "Asynchronous Design — Part 2: Systems and Methodologies" (<https://wayback.archive-it.org/all/20181221132622/http://www.cs.columbia.edu/%7Enowick/nowick%2Dsingh%2Dasync%2DIEEE%2DDT%2D15%2Doverview%2Darticle%2Dpt2.pdf>) (PDF). *IEEE Design and Test*. **32** (3): 19–28. doi:10.1109/MDAT.2015.2413757 (<https://doi.org/10.1109%2FMDAT.2015.2413757>). S2CID 16732793 (<https://api.semanticscholar.org/CorpusID:16732793>). Archived from the original (<https://www.cs.columbia.edu/~nowick/nowick-singh-async-IEEE-DT-15-overview-article-pt2.pdf>) (PDF) on 2018-12-21. Retrieved 2019-08-27.
6. "Why Asynchronous Design?" (<https://galois.com/blog/2021/07/why-asynchronous-design/>). Galois, Inc. 2021-07-15. Retrieved 2021-12-04.
7. Myers, Chris J. (2001). *Asynchronous circuit design* (<https://www.worldcat.org/oclc/53227301>). New York: J. Wiley & Sons. ISBN 0-471-46412-0. OCLC 53227301 (<https://search.worldcat.org/oclc/53227301>).
8. Muller, D. E. (1955). *Theory of asynchronous circuits, Report no. 66*. Digital Computer Laboratory, University of Illinois at Urbana-Champaign.
9. Miller, Raymond E. (1965). *Switching Theory, Vol. II*. Wiley.
10. van Berkel, C. H.; Josephs, M. B.; Nowick, S. M. (February 1999). "Applications of Asynchronous Circuits" (<https://web.archive.org/web/20180403123227/http://www.cs.columbia.edu/~nowick/async-applications-PIEEE-99-berkel-josephs-nowick-published.pdf>) (PDF). *Proceedings of the IEEE*. **87** (2): 234–242. doi:10.1109/5.740016 (<https://doi.org/10.1109%2F5.740016>). Archived from the original (<https://www.cs.columbia.edu/~nowick/async-applications-PIEEE-99-berkel-josephs-nowick-published.pdf>) (PDF) on 2018-04-03. Retrieved 2019-08-27.
11. Vasyukevich, Vadim O. (1984). "Whenjunction as a logic/dynamic operation. Definition, implementation and applications". *Automatic Control and Computer Sciences*. **18** (6): 68–74. (NB. The function was still called *whenjunction* instead of *venjunction* in this publication.)
12. Vasyukevich, Vadim O. (1998). "Monotone sequences of binary data sets and their identification by means of *venjunctive functions*". *Automatic Control and Computer Sciences*. **32** (5): 49–56.
13. Vasyukevich, Vadim O. (April 2007). "Decoding asynchronous sequences". *Automatic Control and Computer Sciences*. **41** (2). Allerton Press: 93–99. doi:10.3103/S0146411607020058 (<https://doi.org/10.3103%2FS0146411607020058>). ISSN 1558-108X (<https://search.worldcat.org/issn/1558-108X>). S2CID 21204394 (<https://api.semanticscholar.org/CorpusID:21204394>).
14. Vasyukevich, Vadim O. (2009). "Asynchronous logic elements. Venjunction and sequentiation" (<http://asynlog.balticom.lv/Content/Files/en.pdf>) (PDF). Archived (<https://web.archive.org/web/20110722160840/http://asynlog.balticom.lv/Content/Files/en.pdf>) (PDF) from the original on 2011-07-22. (118 pages)

15. Vasyukevich, Vadim O. (2011). Written at Riga, Latvia. *Asynchronous Operators of Sequential Logic: Venjunction & Sequention — Digital Circuits Analysis and Design*. Lecture Notes in Electrical Engineering. Vol. 101 (1st ed.). Berlin / Heidelberg, Germany: Springer-Verlag. doi:10.1007/978-3-642-21611-4 (<https://doi.org/10.1007%2F978-3-642-21611-4>). ISBN 978-3-642-21610-7. ISSN 1876-1100 (<https://search.worldcat.org/issn/1876-1100>). LCCN 2011929655 (<https://lccn.loc.gov/2011929655>). (xiii+1+123+7 pages) (NB. The back cover of this book erroneously states volume 4, whereas it actually is volume 101.)
16. Fant, Karl M. (February 2005). *Logically determined design: clockless system design with NULL convention logic (NCL)* (<https://books.google.com/books?id=UTHFcddvvHQcC>) (1 ed.). Hoboken, New Jersey, USA: Wiley-Interscience / John Wiley and Sons, Inc. ISBN 978-0-471-68478-7. LCCN 2004050923 (<https://lccn.loc.gov/2004050923>). (xvi+292 pages)
17. Fant, Karl M. (August 2007). *Computer Science Reconsidered: The Invocation Model of Process Expression* (<http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471798142.html>) (1 ed.). Hoboken, New Jersey, USA: Wiley-Interscience / John Wiley and Sons, Inc. ISBN 978-0-471-79814-9. LCCN 2006052821 (<https://lccn.loc.gov/2006052821>). Retrieved 2023-07-23. (xix+1+269 pages)
18. Smith, Scott C.; Di, Jia (2009). *Designing Asynchronous Circuits using NULL Conventional Logic (NCL)* (<http://www.gbv.de/dms/tib-ub-hannover/734874170.pdf>) (PDF). Synthesis Lectures on Digital Circuits & Systems. Morgan & Claypool Publishers. pp. 61–73. eISSN 1932-3174 (<https://search.worldcat.org/issn/1932-3174>). ISBN 978-1-59829-981-6. ISSN 1932-3166 (<https://search.worldcat.org/issn/1932-3166>). Lecture #23. Retrieved 2023-09-10; Smith, Scott C.; Di, Jia (2022) [2009-07-23]. *Designing Asynchronous Circuits using NULL Conventional Logic (NCL)* (https://books.google.com/books?id=_4JyEAAAQBAJ). Synthesis Lectures on Digital Circuits & Systems. University of Arkansas, Arkansas, USA: Springer Nature Switzerland AG. doi:10.1007/978-3-031-79800-9 (<https://doi.org/10.1007%2F978-3-031-79800-9>). eISSN 1932-3174 (<https://search.worldcat.org/issn/1932-3174>). ISBN 978-3-031-79799-6. ISSN 1932-3166 (<https://search.worldcat.org/issn/1932-3166>). Lecture #23. Retrieved 2023-09-10. (x+86+6 pages)
19. Smith, Scott C.; Di, Jia. "U.S. 7,977,972 Ultra-Low Power Multi-threshold Asynchronous Circuit Design" ([- 20. Rosenblum, L. Ya.; Yakovlev, A. V. \(July 1985\). "Signal Graphs: from Self-timed to Timed ones. Proceedings of International Workshop on Timed Petri Nets" \(<https://www.staff.ncl.ac.uk/alex.yakovlev/home.formal/LR-AY-TPN85.pdf>\) \(PDF\). Torino, Italy: IEEE CS Press. pp. 199–207. Archived \(<https://web.archive.org/web/20031023002229/http://www.staff.ncl.ac.uk/alex.yakovlev/home.formal/LR-AY-TPN85.pdf>\) \(PDF\) from the original on 2003-10-23.
- 21. Chu, T.-A. \(1986-06-01\). "On the models for designing VLSI asynchronous digital systems" \(<https://www.sciencedirect.com/science/article/abs/pii/S0167926086800025>\). *Integration*. 4 \(2\): 99–113. doi:10.1016/S0167-9260\(86\)80002-5 \(<https://doi.org/10.1016%2FS0167-9260%2886%2980002-5>\). ISSN 0167-9260 \(<https://search.worldcat.org/issn/0167-9260>\).
- 22. Yakovlev, Alexandre; Lavagno, Luciano; Sangiovanni-Vincentelli, Alberto \(1996-11-01\). "A unified signal transition graph model for asynchronous control circuit synthesis" \(<https://doi.org/10.1007/BF00122081>\). *Formal Methods in System Design*. 9 \(3\): 139–188. doi:10.1007/BF00122081 \(<https://doi.org/10.1007%2FBF00122081>\). ISSN 1572-8102 \(<https://search.worldcat.org/issn/1572-8102>\). S2CID 26970846 \(<https://api.semanticscholar.org/CorpusID:26970846>\).
- 23. Cortadella, J.; Kishinevsky, M.; Kondratyev, A.; Lavagno, L.; Yakovlev, A. \(2002\). *Logic Synthesis for Asynchronous Controllers and Interfaces* \(<http://link.springer.com/10.1007/978-3-642-55989-1>\). Springer Series in Advanced Microelectronics. Vol. 8. Berlin / Heidelberg, Germany: Springer Berlin Heidelberg. doi:10.1007/978-3-642-55989-1 \(<https://doi.org/10.1007%2F978-3-642-55989-1>\). ISBN 978-3-642-62776-7.](http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnetacgi%2FPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=7977972.PN.&OS=PN/7977972&RS=PN/7977972)

24. "Petrify: Related publications" (<https://www.cs.upc.edu/~jordicf/petrify/refs/>). www.cs.upc.edu. Retrieved 2021-07-28.
25. "start - Workcraft" (<https://workcraft.org/>). workcraft.org. Retrieved 2021-07-28.
26. Nowick, S. M.; Singh, M. (September–October 2011). "High-Performance Asynchronous Pipelines: an Overview" (<https://web.archive.org/web/20210421193250/http://www.cs.columbia.edu/~nowick/nowick-singh-ieee-dt-11-published.pdf>) (PDF). *IEEE Design & Test of Computers*. **28** (5): 8–22. doi:10.1109/mtt.2011.71 (<https://doi.org/10.1109%2Fmtt.2011.71>). S2CID 6515750 (<https://api.semanticscholar.org/CorpusID:6515750>). Archived from the original (<https://www.cs.columbia.edu/~nowick/nowick-singh-ieee-dt-11-published.pdf>) (PDF) on 2021-04-21. Retrieved 2019-08-27.
27. Nowick, S. M.; Yun, K. Y.; Beerel, P. A.; Dooley, A. E. (March 1997). "Speculative completion for the design of high-performance asynchronous dynamic adders" (<https://web.archive.org/web/20210421231646/http://www.cs.columbia.edu/~nowick/nowick-async97-speculation-completion-fin.pdf>) (PDF). *Proceedings Third International Symposium on Advanced Research in Asynchronous Circuits and Systems*. pp. 210–223. doi:10.1109/ASYNC.1997.587176 (<https://doi.org/10.1109%2FASYNC.1997.587176>). ISBN 0-8186-7922-0. S2CID 1098994 (<https://api.semanticscholar.org/CorpusID:1098994>). Archived from the original (<https://www.cs.columbia.edu/~nowick/nowick-async97-speculation-completion-fin.pdf>) (PDF) on 2021-04-21. Retrieved 2019-08-27.
28. Nowick, S. M. (September 1996). "Design of a Low-Latency Asynchronous Adder Using Speculative Completion" (<https://web.archive.org/web/20210422005641/http://www.cs.columbia.edu/~nowick/nowick-iee-pcgs-speculative-competition-96-published.pdf>) (PDF). *IEEE Proceedings - Computers and Digital Techniques*. **143** (5): 301–307. doi:10.1049/ip-cdt:19960704 (<https://doi.org/10.1049%2Fip-cdt%3A19960704>) (inactive 2024-12-07). Archived from the original (<https://www.cs.columbia.edu/~nowick/nowick-iee-pcgs-speculative-competition-96-published.pdf>) (PDF) on 2021-04-22. Retrieved 2019-08-27.
29. Sheikh, B.; Manohar, R. (May 2010). "An Operand-Optimized Asynchronous IEEE 754 Double-Precision Floating-Point Adder" (<https://web.archive.org/web/20210421233614/http://www.cs.columbia.edu/~nowick/sheikh-manohar-async10-fp-adder.pdf>) (PDF). *Proceedings of the IEEE International Symposium on Asynchronous Circuits and Systems ('Async')*: 151–162. Archived from the original (<https://www.cs.columbia.edu/~nowick/sheikh-manohar-async10-fp-adder.pdf>) (PDF) on 2021-04-21. Retrieved 2019-08-27.
30. Sasao, Tsutomu (1993). *Logic Synthesis and Optimization* (<https://www.worldcat.org/oclc/852788081>). Boston, Massachusetts, USA: Springer USA. ISBN 978-1-4615-3154-8. OCLC 852788081 (<https://search.worldcat.org/oclc/852788081>).
31. "Epson Develops the World's First Flexible 8-Bit Asynchronous Microprocessor" (http://global.epson.com/newsroom/2005/news_2005_02_09.htm) 2005
32. Furber, Steve. "Principles of Asynchronous Circuit Design" (https://web.archive.org/web/20120426050921/http://owlhouse.csie.nctu.edu.tw/~dannim/AsynCD/principles_of_ASYNC.pdf) (PDF). Pg. 232. Archived from the original (http://owlhouse.csie.nctu.edu.tw/~dannim/AsynCD/principles_of_ASYNC.pdf) (PDF) on 2012-04-26. Retrieved 2011-12-13.
33. "Keep It Strictly Synchronous: KISS those asynchronous-logic problems good-bye". Personal Engineering and Instrumentation News, November 1997, pages 53–55. <http://www.fpga-site.com/kiss.html>
34. van Leeuwen, T. M. (2010). *Implementation and automatic generation of asynchronous scheduled dataflow graph* (<https://repository.tudelft.nl/islandora/object/uuid:5d87b87f-e084-491f-a18a-9c83ac2c41e1/datastream/OBJ/download>). Delft.
35. Kruger, Robert (2005-03-15). "Reality TV for FPGA design engineers!" (https://www.eetimes.com/reality-tv-for-fpga-design-engineers/?page_number=2). [eetimes.com](http://www.eetimes.com). Retrieved 2020-11-11.

36. LARD (<http://www.cs.man.ac.uk/apt/projects/tools/lard/>) Archived (<https://web.archive.org/web/20050306161822/http://www.cs.man.ac.uk/apt/projects/tools/lard/>) March 6, 2005, at the Wayback Machine
37. "In the 1950 and 1960s, asynchronous design was used in many early mainframe computers, including the ILLIAC I and ILLIAC II" Brief History of asynchronous circuit design (<https://books.google.com/books?id=DPGJEPZGXMQC&pg=PA322&lpg=PA322>)
38. "The Illiac is a binary parallel asynchronous computer in which negative numbers are represented as two's complements." – final summary of "Illiac Design Techniques" (http://www.bitsavers.org/pdf/univOfIllinoisUrbana/illiac/ILLIAC/ILLIAC_Design_Techniques_May55.pdf) 1955.
39. Martin, A. J.; Nystrom, M.; Wong, C. G. (November 2003). "Three generations of asynchronous microprocessors" (<https://ieeexplore.ieee.org/document/1246159>). *IEEE Design & Test of Computers*. **20** (6): 9–17. doi:10.1109/MDT.2003.1246159 (<https://doi.org/10.1109%2FMDT.2003.1246159>). ISSN 0740-7475 (<https://search.worldcat.org/issn/0740-7475>). S2CID 15164301 (<https://api.semanticscholar.org/CorpusID:15164301>).
40. Martin, A. J.; Nystrom, M.; Papadantonakis, K.; Penzes, P. I.; Prakash, P.; Wong, C. G.; Chang, J.; Ko, K. S.; Lee, B.; Ou, E.; Pugh, J. (2003). "The Lutonium: A sub-nanojoule asynchronous 8051 microcontroller" (<https://ieeexplore.ieee.org/document/1199162>). *Ninth International Symposium on Asynchronous Circuits and Systems, 2003. Proceedings* (<http://infoscience.epfl.ch/record/102894/files/lutonium.pdf>) (PDF). Vancouver, BC, Canada: IEEE Comput. Soc. pp. 14–23. doi:10.1109/ASYNC.2003.1199162 (<https://doi.org/10.1109%2FASYNC.2003.1199162>). ISBN 978-0-7695-1898-5. S2CID 13866418 (<https://api.semanticscholar.org/CorpusID:13866418>).
41. Martin, Alain J. (2014-02-06). "25 Years Ago: The First Asynchronous Microprocessor" (<http://resolver.caltech.edu/CaltechAUTHORS:20140206-111915844>). Computer Science Technical Reports. California Institute of Technology. doi:10.7907/Z9QR4V3H (<https://doi.org/10.7907%2FZ9QR4V3H>). { {cite journal} }: Cite journal requires |journal= (help)
42. "Seiko Epson tips flexible processor via TFT technology" (<http://www.eetimes.com/conf/isscc/showArticle.jhtml?articleID=59302081&kc=3681>) Archived (<https://web.archive.org/web/201002012153/http://eetimes.com/conf/isscc/showArticle.jhtml?articleID=59302081&kc=3681>) 2010-02-01 at the Wayback Machine by Mark LaPedus 2005
43. "A flexible 8b asynchronous microprocessor based on low-temperature poly-silicon TFT technology" (<https://ieeexplore.ieee.org/document/1493974>) by Karaki et al. 2005. Abstract: "A flexible 8b asynchronous microprocessor ACTII ... The power level is 30% of the synchronous counterpart."
44. "Introduction of TFT R&D Activities in Seiko Epson Corporation" (http://www.holtronic.ch/White_papers/SE2005_1.pdf) by Tatsuya Shimoda (2005?) has picture of "A flexible 8-bit asynchronous microprocessor, ACT11"
45. "Epson Develops the World's First Flexible 8-Bit Asynchronous Microprocessor" (http://www.epson.co.jp/e/newsroom/2005/news_2005_02_09.htm)
46. "Seiko Epson details flexible microprocessor: A4 sheets of e-paper in the pipeline" (<http://www.pcadvisor.co.uk/news/index.cfm?newsid=4547>) by Paul Kallender 2005
47. "SyNAPSE program develops advanced brain-inspired chip" (<https://www.darpa.mil/NewsEvents/Releases/2014/08/07.aspx>) Archived (<https://web.archive.org/web/20140810011226/http://www.darpa.mil/NewsEvents/Releases/2014/08/07.aspx>) 2014-08-10 at the Wayback Machine. August 07, 2014.
48. Johnniac history written in 1968 (http://www.rand.org/content/dam/rand/pubs/research_memoranda/2005/RM5654.pdf)
49. V. M. Glushkov and E. L. Yushchenko. Mathematical description of computer "Kiev". UkrSSR, 1962 (in Russian)
50. "Computer Resurrection Issue 18" (<http://www.cs.man.ac.uk/CCS/res/res18.htm>).

51. "Entirely asynchronous, its hundred-odd boards would send out requests, earmark the results for somebody else, swipe somebody else's signals or data, and backstab each other in all sorts of amusing ways which occasionally failed (the "op not complete" timer would go off and cause a fault). ... [There] was no hint of an organized synchronization strategy: various "it's ready now", "ok, go", "take a cycle" pulses merely surged through the vast backpanel ANDed with appropriate state and goosed the next guy down. Not without its charms, this seemingly ad-hoc technology facilitated a substantial degree of overlap ... as well as the [segmentation and paging] of the Multics address mechanism to the extant 6000 architecture in an ingenious, modular, and surprising way Modification and debugging of the processor, though, were no fun." [Multics Glossary: ... 6180](http://www.multicians.org/mga.html#6180) (<http://www.multicians.org/mga.html#6180>)
52. "10/81 ... DPS 8/70M CPUs" [Multics Chronology](http://www.multicians.org/chrono.html) (<http://www.multicians.org/chrono.html>)
53. "The Series 60, Level 68 was just a repackaging of the 6180." [Multics Hardware features: Series 60, Level 68](http://www.multicians.org/features.html#tag2.4) (<http://www.multicians.org/features.html#tag2.4>)
54. A. A. Vasenkov, V. L. Dshkhunian, P. R. Mashevich, P. V. Nesterov, V. V. Telenkov, Ju. E. Chicherin, D. I. Juditsky, "Microprocessor computing system," Patent US4124890, Nov. 7, 1978 (http://worldwide.espacenet.com/publicationDetails/originalDocument?CC=US&NR=4124890A&KC=A&FT=D&ND=3&date=19781107&DB=EPODOC&locale=en_EP)
55. Chapter 4.5.3 in the biography of D. I. Juditsky (in Russian) (<http://www.computer-museum.ru/articles/?article=116>)
56. "Серия 587 - Collection ex-USSR Chip's" (<https://web.archive.org/web/20150717061828/http://www.cpu80.ru/home/seria-587>). Archived from the original (<http://www.cpu80.ru/home/seria-587>) on 2015-07-17. Retrieved 2015-07-16.
57. "Серия 588 - Collection ex-USSR Chip's" (<https://web.archive.org/web/20150717082004/http://www.cpu80.ru/home/seria-588>). Archived from the original (<http://www.cpu80.ru/home/seria-588>) on 2015-07-17. Retrieved 2015-07-16.
58. "Серия 1883/U830 - Collection ex-USSR Chip's" (<https://web.archive.org/web/20150722062052/http://www.cpu80.ru/home/seria-u83-k1883>). Archived from the original (<http://www.cpu80.ru/home/seria-u83-k1883>) on 2015-07-22. Retrieved 2015-07-19.
59. "A Network-based Asynchronous Architecture for Cryptographic Devices" (http://www.era.lib.ed.ac.uk/bitstream/1842/860/1/Spadavecchia_thesis.pdf) by Ljiljana Spadavecchia 2005 in section "4.10.2 Side-channel analysis of dual-rail asynchronous architectures" and section "5.5.5.1 Instruction set"
60. Zhang, Qianyi; Theodoropoulos, Georgios (2024). "SAMIPS: A Synthesised Asynchronous Processor". arXiv:2409.20388 (<https://arxiv.org/abs/2409.20388>) [cs.AR (<https://arxiv.org/archive/cs.AR/>)].
61. Zhang, Qianyi; Theodoropoulos, Georgios (2003). "Towards an Asynchronous MIPS Processor" (https://link.springer.com/chapter/10.1007/978-3-540-39864-6_12). In Omondi, Amos; Sedukhin, Stanislav (eds.). *Advances in Computer Systems Architecture*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. pp. 137–150. doi:[10.1007/978-3-540-39864-6_12](https://doi.org/10.1007/978-3-540-39864-6_12) (https://doi.org/10.1007%2F978-3-540-39864-6_12). ISBN 978-3-540-39864-6.
62. "Handshake Solutions HT80C51" (<http://www.keil.com/dd/chip/3931.htm>) "The Handshake Solutions HT80C51 is a Low power, asynchronous 80C51 implementation using handshake technology, compatible with the standard 8051 instruction set."
63. Lines, Andrew (March 2007). "The Vortex: A Superscalar Asynchronous Processor" (<https://ieeexplore.ieee.org/document/4137031>). *13th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC'07)*. pp. 39–48. doi:[10.1109/ASYNC.2007.28](https://doi.org/10.1109/ASYNC.2007.28) (<https://doi.org/10.1109%2FASYNC.2007.28>). ISBN 978-0-7695-2771-0. S2CID 33189213 (<https://api.semanticscholar.org/CorpusID:33189213>).

64. Lines, A. (2003). "Nexus: An asynchronous crossbar interconnect for synchronous system-on-chip designs" (<https://ieeexplore.ieee.org/document/1231470>). *11th Symposium on High Performance Interconnects, 2003. Proceedings*. Stanford, CA, USA: IEEE Comput. Soc. pp. 2–9. doi:[10.1109/CONECT.2003.1231470](https://doi.org/10.1109/CONECT.2003.1231470) (<https://doi.org/10.1109%2FCONECT.2003.1231470>). ISBN 978-0-7695-2012-4. S2CID 1799204 (<https://api.semanticscholar.org/CorpusID:1799204>).
65. SEAforth Overview (http://www.intellasys.net/index.php?option=com_content&task=view&id=21&Itemid=41) Archived (https://web.archive.org/web/20080202055942/http://www.intellasys.net/index.php?option=com_content&task=view&id=21&Itemid=41) 2008-02-02 at the Wayback Machine "... asynchronous circuit design throughout the chip. There is no central clock with billions of dumb nodes dissipating useless power. ... the processor cores are internally asynchronous themselves."
66. "GreenArrayChips" (<http://www.greenarraychips.com>) "Ultra-low-powered multi-computer chips with integrated peripherals."
67. Tiempo: Asynchronous TAM16 Core IP (http://www.tiempo-ic.com/uploads/Docs/TAM16_Datasheet.pdf?page=uploads/Docs/Tiempo%20TAM16%20IP%20Data%20Sheet%201.2.pdf)
68. "ASPIDA sync/async DLX Core" (<http://opencores.org/project,aspida>). OpenCores.org. Retrieved 2014-09-05.
69. "Asynchronous Open-Source DLX Processor (ASPIDA)" (<https://www.ics.forth.gr/carv/async-hronous-circuits-systems-2001-2010>).

Further reading

- [TiDE](https://web.archive.org/web/20071215230357/http://www.handshakesolutions.com/) (<https://web.archive.org/web/20071215230357/http://www.handshakesolutions.com/>) from Handshake Solutions in The Netherlands, Commercial asynchronous circuits design tool. Commercial asynchronous ARM (ARM996HS) and 8051 (HT80C51) are available.
- An introduction to asynchronous circuit design (<https://www.cs.columbia.edu/~nowick/ald-nowick-tr-intro.pdf>) Archived (<https://web.archive.org/web/20100623121342/http://www.cs.columbia.edu/~nowick/ald-nowick-tr-intro.pdf>) 23 June 2010 at the Wayback Machine by Davis and Nowick
- Null convention logic (<http://theseusresearch.com/NullConventionLogic.htm>), a design style pioneered by Theseus Logic, who have fabricated over 20 ASICs based on their NCL08 and NCL8501 microcontroller cores [1] (<https://web.archive.org/web/20070927214117/http://scism.sbu.ac.uk/ccsv/ACID-WG/AsyncIndustryStatus.pdf>)
- The Status of Asynchronous Design in Industry (<https://web.archive.org/web/20111009112125/http://www.scism.lsbu.ac.uk/ccsv/ACID-WG/AsyncIndustryStatus.pdf>) Information Society Technologies (IST) Programme, IST-1999-29119, D. A. Edwards W. B. Toms, June 2004, via www.scism.lsbu.ac.uk
- The Red Star (http://brej.org/red_star/) is a version of the MIPS R3000 implemented in asynchronous logic
- The Amulet microprocessors (<https://web.archive.org/web/20060807232502/http://www.cs.manchester.ac.uk/apt/projects/processors/amulet/>) were asynchronous ARMs, built in the 1990s at University of Manchester, England
- The SAMIPS (<https://www.gtheodoropoulos.com/Research/Projects/samips/samips.html>) synthesised asynchronous MIPS R3000 processor.
- The N-Protocol (<http://www.asyncart.com>) developed by Navarre AsyncArt, the first commercial asynchronous design methodology for conventional FPGAs
- PGPSALM (<http://www.henning-mersch.de/pgpsalm/>) an asynchronous implementation of the 6502 microprocessor

- [Caltech Async Group home page](https://web.archive.org/web/20071214045029/http://async.caltech.edu/) (<https://web.archive.org/web/20071214045029/http://async.caltech.edu/>)
- [Tiempo](http://www.tiempo-ic.com/) (<http://www.tiempo-ic.com/>): French company providing asynchronous IP and design tools
- [Epson ACT11 Flexible CPU Press Release](https://archive.today/20130122151803/http://www.eetimes.com/showArticle.jhtml?articleID=59302120) (<https://archive.today/20130122151803/http://www.eetimes.com/showArticle.jhtml?articleID=59302120>)
- [Newcastle upon Tyne Async Group page](http://async.org.uk) (<http://async.org.uk>)

External links

-  [Media related to Asynchronous circuits at Wikimedia Commons](#)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Asynchronous_circuit&oldid=1291883727#Asynchronous_logic"



UML state machine

(Redirected from [Hierarchical state machine](#))

UML state machine,^[1] formerly known as **UML statechart**, is an extension of the [mathematical concept of a finite automaton](#) in [computer science](#) applications as expressed in the [Unified Modeling Language](#) (UML) notation.

The concepts behind it are about organizing the way a device, computer program, or other (often technical) process works such that an entity or each of its sub-entities is always in exactly one of a number of possible states and where there are well-defined conditional transitions between these states.

UML state machine is an object-based variant of [Harel statechart](#),^[2] adapted and extended by UML.^{[1][3]} The goal of UML state machines is to overcome the main limitations of traditional [finite-state machines](#) while retaining their main benefits. UML statecharts introduce the new concepts of [hierarchically nested states](#) and [orthogonal regions](#), while extending the notion of [actions](#). UML state machines have the characteristics of both [Mealy machines](#) and [Moore machines](#). They support [actions](#) that depend on both the state of the system and the triggering [event](#), as in Mealy machines, as well as [entry](#) and [exit actions](#), which are associated with states rather than transitions, as in Moore machines.^[4]

The term "UML state machine" can refer to two kinds of state machines: *behavioral state machines* and *protocol state machines*. Behavioral state machines can be used to model the behavior of individual entities (e.g., class instances), a subsystem, a package, or even an entire system. Protocol state machines are used to express usage protocols and can be used to specify the legal usage scenarios of classifiers, interfaces, and ports.

Basic state machine concepts

Many software systems are [event-driven](#), which means that they continuously wait for the occurrence of some external or internal [event](#) such as a mouse click, a button press, a time tick, or an arrival of a data packet. After recognizing the event, such systems react by performing the appropriate computation that may include manipulating the hardware or generating “soft” events that trigger other internal software components. (That's why event-driven systems are alternatively called **reactive systems**.) Once the event handling is complete, the system goes back to waiting for the next event.

The response to an event generally depends on both the type of the event and on the internal [state](#) of the system and can include a change of state leading to a [state transition](#). The pattern of events, states, and state transitions among those states can be abstracted and represented as a [finite-state machine](#) (FSM).

The concept of a FSM is important in [event-driven programming](#) because it makes the event handling explicitly dependent on both the event-type and on the state of the system. When used correctly, a state machine can drastically cut down the number of execution paths through the code, simplify the conditions tested at each branching point, and simplify the switching between different modes of execution.^[5] Conversely, using event-driven programming without an underlying FSM model can lead programmers to produce error prone, difficult to extend and excessively complex application code.^[6]

Basic UML state diagrams

UML preserves the general form of the [traditional state diagrams](#). The UML state diagrams are [directed graphs](#) in which nodes denote states and connectors denote state transitions. For example, Figure 1 shows a UML state diagram corresponding to the computer keyboard state machine. In UML, states are represented as rounded

rectangles labeled with state names. The transitions, represented as arrows, are labeled with the triggering events followed optionally by the list of executed actions. The **initial transition** originates from the solid circle and specifies the default state when the system first begins. Every state diagram should have such a transition, which should not be labeled, since it is not triggered by an event. The initial transition can have associated actions.

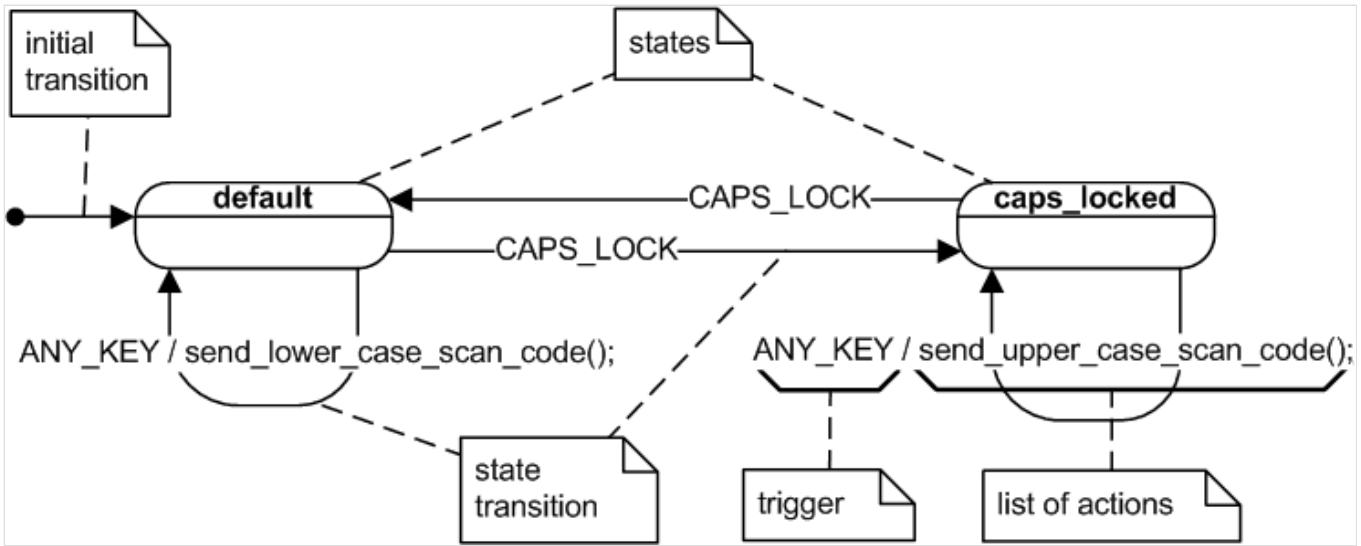


Figure 1: UML state diagram representing the computer keyboard state machine

Events

An **event** is something that happens that affects the system. Strictly speaking, in the UML specification,^[1] the term event refers to the type of occurrence rather than to any concrete instance of that occurrence. For example, Keystroke is an event for the keyboard, but each press of a key is not an event but a concrete instance of the Keystroke event. Another event of interest for the keyboard might be Power-on, but turning the power on tomorrow at 10:05:36 will be just an instance of the Power-on event.

An event can have associated **parameters**, allowing the event instance to convey not only the occurrence of some interesting incident but also quantitative information regarding that occurrence. For example, the Keystroke event generated by pressing a key on a computer keyboard has associated parameters that convey the character scan code as well as the status of the Shift, Ctrl, and Alt keys.

An event instance outlives the instantaneous occurrence that generated it and might convey this occurrence to one or more state machines. Once generated, the event instance goes through a processing life cycle that can consist of up to three stages. First, the event instance is **received** when it is accepted and waiting for processing (e.g., it is placed on the event queue). Later, the event instance is **dispatched** to the state machine, at which point it becomes the current event. Finally, it is **consumed** when the state machine finishes processing the event instance. A consumed event instance is no longer available for processing.

States

Each state machine has a **state**, which governs reaction of the state machine to events. For example, when you strike a key on a keyboard, the character code generated will be either an uppercase or a lowercase character, depending on whether the Caps Lock is active. Therefore, the keyboard's behavior can be divided into two states: the "default" state and the "caps_locked" state. (Most keyboards include an LED that indicates that the keyboard is in the "caps_locked" state.) The behavior of a keyboard depends only on certain aspects of its history, namely whether the Caps Lock key has been pressed, but not, for example, on how many and exactly which other keys have been pressed previously. A state can abstract away all possible (but irrelevant) event sequences and capture only the relevant ones.

In the context of software state machines (and especially classical FSMs), the term *state* is often understood as a single *state variable* that can assume only a limited number of a priori determined values (e.g., two values in case of the keyboard, or more generally - some kind of variable with an *enum* type in many programming languages). The idea of *state variable* (and classical FSM model) is that the value of the *state variable* fully defines the current state of the system at any given time. The concept of the state reduces the problem of identifying the execution context in the code to testing just the state variable instead of many variables, thus eliminating a lot of conditional logic.

Extended states

In practice, however, interpreting the whole state of the state machine as a single *state variable* quickly becomes impractical for all state machines beyond very simple ones. Indeed, even if we have a single 32-bit integer in our machine state, it could contribute to over 4 billion different states - and will lead to a premature *state explosion*. This interpretation is not practical, so in UML state machines the whole *state* of the state machine is commonly split into (a) an enumerable *state variable* and (b) all the other variables which are named *extended state*. Another way to see it is to interpret the enumerable *state variable* as a qualitative aspect and the *extended state* as quantitative aspects of the whole state. In this interpretation, a change of variable does not always imply a change of the qualitative aspects of the system behavior and therefore does not lead to a change of state.^[7]

State machines supplemented with *extended state* variables are called **extended state machines** and UML state machines belong to this category. Extended state machines can apply the underlying formalism to much more complex problems than is practical without including extended state variables. For example, if we have to implement some kind of limit in our FSM (say, limiting number of keystrokes on keyboard to 1000), without *extended state* we'd need to create and process 1000 states - which is not practical; however, with an extended state machine we can introduce a **key_count** variable, which is initialized to 1000 and decremented by every keystroke without changing *state variable*.

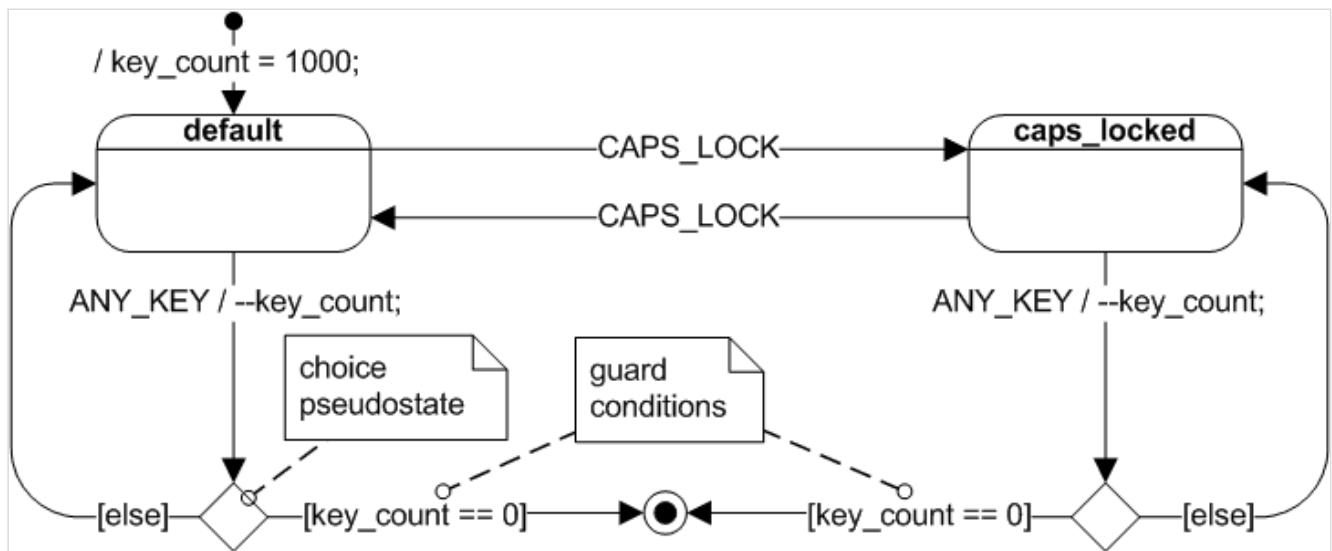


Figure 2: Extended state machine of "cheap keyboard" with extended state variable `key_count` and various guard conditions

The state diagram from Figure 2 is an example of an extended state machine, in which the complete condition of the system (called the *extended state*) is the combination of a qualitative aspect—the *state variable*—and the quantitative aspects—the *extended state* variables.

The obvious advantage of extended state machines is flexibility. For example, changing the limit governed by `key_count` from 1000 to 10000 keystrokes, would not complicate the extended state machine at all. The only modification required would be changing the initialization value of the `key_count` extended state variable during initialization.

This flexibility of extended state machines comes with a price, however, because of the complex coupling between the "qualitative" and the "quantitative" aspects of the extended state. The coupling occurs through the guard conditions attached to transitions, as shown in Figure 2.

Guard conditions

Guard conditions (or simply guards) are Boolean expressions evaluated dynamically based on the value of extended state variables and event parameters. Guard conditions affect the behavior of a state machine by enabling actions or transitions only when they evaluate to TRUE and disabling them when they evaluate to FALSE. In the UML notation, guard conditions are shown in square brackets (e.g., `[key_count == 0]` in Figure 2).

The need for guards is the immediate consequence of adding memory extended state variables to the state machine formalism. Used sparingly, extended state variables and guards make up a powerful mechanism that can simplify designs. On the other hand, it is possible to abuse extended states and guards quite easily.^[8]

Actions and transitions

When an event instance is dispatched, the state machine responds by performing **actions**, such as changing a variable, performing I/O, invoking a function, generating another event instance, or changing to another state. Any parameter values associated with the current event are available to all actions directly caused by that event.

Switching from one state to another is called **state transition**, and the event that causes it is called the triggering event, or simply the **trigger**. In the keyboard example, if the keyboard is in the "default" state when the CapsLock key is pressed, the keyboard will enter the "caps_locked" state. However, if the keyboard is already in the "caps_locked" state, pressing CapsLock will cause a different transition—from the "caps_locked" to the "default" state. In both cases, pressing CapsLock is the triggering event.

In extended state machines, a transition can have a guard, which means that the transition can "fire" only if the guard evaluates to TRUE. A state can have many transitions in response to the same trigger, as long as they have nonoverlapping guards; however, this situation could create problems in the sequence of evaluation of the guards when the common trigger occurs. The UML specification^[1] intentionally does not stipulate any particular order; rather, UML puts the burden on the designer to devise guards in such a way that the order of their evaluation does not matter. Practically, this means that guard expressions should have no side effects, at least none that would alter evaluation of other guards having the same trigger.

Run-to-completion execution model

All state machine formalisms, including UML state machines, universally assume that a state machine completes processing of each event before it can start processing the next event. This model of execution is called run to completion, or RTC.

In the RTC model, the system processes events in discrete, indivisible RTC steps. New incoming events cannot interrupt the processing of the current event and must be stored (typically in an event queue) until the state machine becomes idle again. These semantics completely avoid any internal concurrency issues within a single state machine. The RTC model also gets around the conceptual problem of processing actions associated with transitions, where the state machine is not in a well-defined state (is between two states) for the duration of the action. During event processing, the system is unresponsive (unobservable), so the ill-defined state during that time has no practical significance.

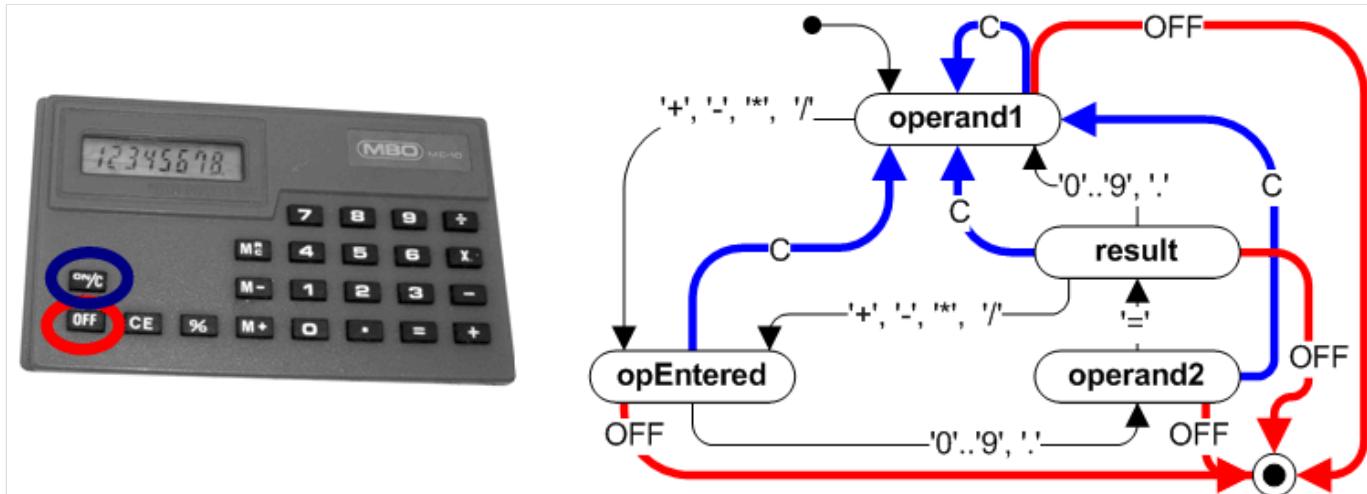
Note, however, that RTC does not mean that a state machine has to monopolize the CPU until the RTC step is complete.^[1] The preemption restriction only applies to the task context of the state machine that is already busy processing events. In a multitasking environment, other tasks (not related to the task context of the busy state

machine) can be running, possibly preempting the currently executing state machine. As long as other state machines do not share variables or other resources with each other, there are no concurrency hazards.

The key advantage of RTC processing is simplicity. Its biggest disadvantage is that the responsiveness of a state machine is determined by its longest RTC step. Achieving short RTC steps can often significantly complicate real-time designs.

UML extensions to the traditional FSM formalism

Though the traditional FSMs are an excellent tool for tackling smaller problems, it's also generally known that they tend to become unmanageable, even for moderately involved systems. Due to the phenomenon known as **state and transition explosion**, the complexity of a traditional FSM tends to grow much faster than the complexity of the system it describes. This happens because the traditional state machine formalism inflicts repetitions. For example, if you try to represent the behavior of a simple pocket calculator with a traditional FSM, you'll immediately notice that many events (e.g., the Clear or Off button presses) are handled identically in many states. A conventional FSM shown in the figure below, has no means of capturing such a commonality and requires *repeating* the same actions and transitions in many states. What's missing in the traditional state machines is the mechanism for factoring out the common behavior in order to share it across many states.



A pocket calculator (left) and the traditional state machine with multiple transitions Clear and Off (right)

UML state machines address exactly this shortcoming of the conventional FSMs. They provide a number of features for eliminating the repetitions so that the complexity of a UML state machine no longer explodes but tends to faithfully represent the complexity of the reactive system it describes. Obviously, these features are very interesting to software developers, because only they make the whole state machine approach truly applicable to real-life problems.

Hierarchically nested states

The most important innovation of UML state machines over the traditional FSMs is the introduction of **hierarchically nested states** (that is why statecharts are also called **hierarchical state machines**, or **HSMs**). The semantics associated with state nesting are as follows (see Figure 3): If a system is in the nested state, for example "result" (called the **substate**), it also (implicitly) is in the surrounding state "on" (called the **superstate**). This state machine will attempt to handle any event in the context of the substate, which conceptually is at the lower level of the hierarchy. However, if the substate "result" does not prescribe how to handle the event, the event is not quietly discarded as in a traditional "flat" state machine; rather, it is automatically handled at the higher level context of the superstate "on". This is what is meant by the system being in state "result" as well as "on". Of course, state nesting is not limited to one level only, and the simple rule of event processing applies recursively to any level of nesting.

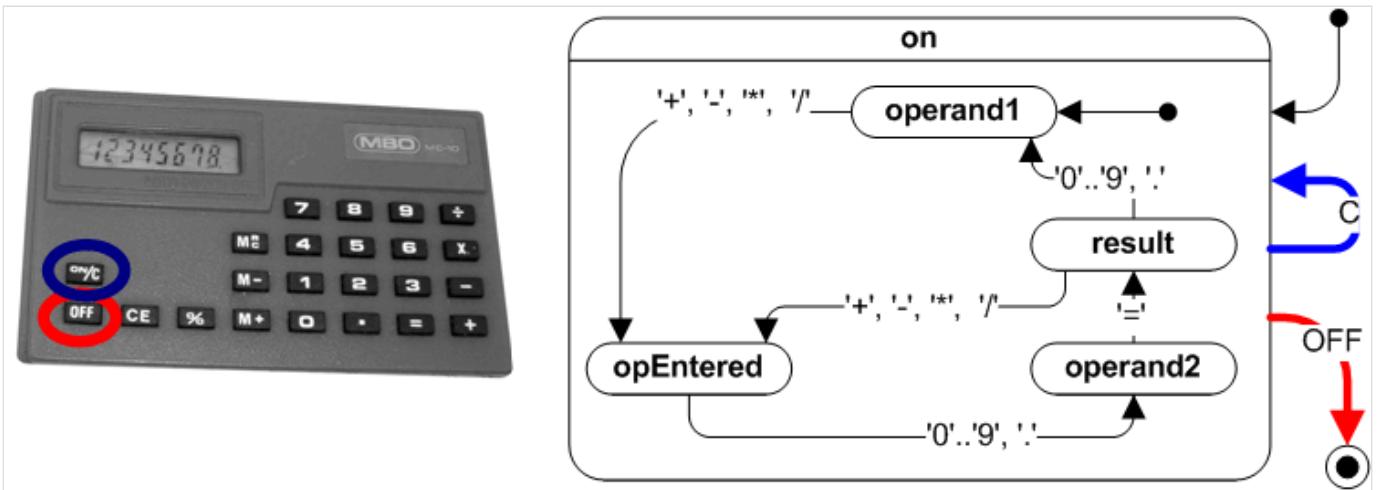


Figure 3: A pocket calculator (left) and the UML state machine with state nesting (right)

States that contain other states are called *composite states*; conversely, states without internal structure are called *simple states*. A nested state is called a *direct substate* when it is not contained by any other state; otherwise, it is referred to as a *transitively nested substate*.

Because the internal structure of a composite state can be arbitrarily complex, any hierarchical state machine can be viewed as an internal structure of some (higher-level) composite state. It is conceptually convenient to define one composite state as the ultimate root of state machine hierarchy. In the UML specification, every state machine has a **region** (the abstract root of every state machine hierarchy),^[9] which contains all the other elements of the entire state machine. The graphical rendering of this all-enclosing region is optional.

As you can see, the semantics of hierarchical state decomposition are designed to facilitate reusing of behavior. The substates (nested states) need only define the differences from the superstates (containing states). A substate can easily inherit^[6] the common behavior from its superstate(s) by simply ignoring commonly handled events, which are then automatically handled by higher-level states. In other words, hierarchical state nesting enables **programming by difference**.^[10]

The aspect of state hierarchy emphasized most often is abstraction—an old and powerful technique for coping with complexity. Instead of addressing all aspects of a complex system at the same time, it is often possible to ignore (abstract away) some parts of the system. Hierarchical states are an ideal mechanism for hiding internal details because the designer can easily zoom out or zoom in to hide or show nested states.

However, composite states don't simply hide complexity; they also actively reduce it through the powerful mechanism of hierarchical event processing. Without such reuse, even a moderate increase in system complexity could lead to an explosive increase in the number of states and transitions. For example, the hierarchical state machine representing the pocket calculator (Figure 3) avoids repeating the transitions Clear and Off in virtually every state. Avoiding repetition allows the growth of HSMs to remain proportionate to growth in system complexity. As the modeled system grows, the opportunity for reuse also increases and thus potentially counteracts the disproportionate increase in numbers of states and transitions typical of traditional FSMs.

Orthogonal regions

Analysis by hierarchical state decomposition can include the application of the operation 'exclusive-OR' to any given state. For example, if a system is in the "on" superstate (Figure 3), it may be the case that it is also in either "operand1" substate OR the "operand2" substate OR the "opEntered" substate OR the "result" substate. This would lead to description of the "on" superstate as an 'OR-state'.

UML statecharts also introduce the complementary AND-decomposition. Such decomposition means that a composite state can contain two or more orthogonal regions (orthogonal means compatible and independent in this context) and that being in such a composite state entails being in all its orthogonal regions simultaneously.^[11]

Orthogonal regions address the frequent problem of a combinatorial increase in the number of states when the behavior of a system is fragmented into independent, concurrently active parts. For example, apart from the main keypad, a computer keyboard has an independent numeric keypad. From the previous discussion, recall the two states of the main keypad already identified: "default" and "caps_locked" (see Figure 1). The numeric keypad also can be in two states—"numbers" and "arrows"—depending on whether Num Lock is active. The complete state space of the keyboard in the standard decomposition is therefore the Cartesian product of the two components (main keypad and numeric keypad) and consists of four states: "default–numbers," "default–arrows," "caps_locked–numbers," and "caps_locked–arrows." However, this would be an unnatural representation because the behavior of the numeric keypad does not depend on the state of the main keypad and vice versa. The use of orthogonal regions allows the mixing of independent behaviors as a Cartesian product to be avoided and, instead, for them to remain separate, as shown in Figure 4.

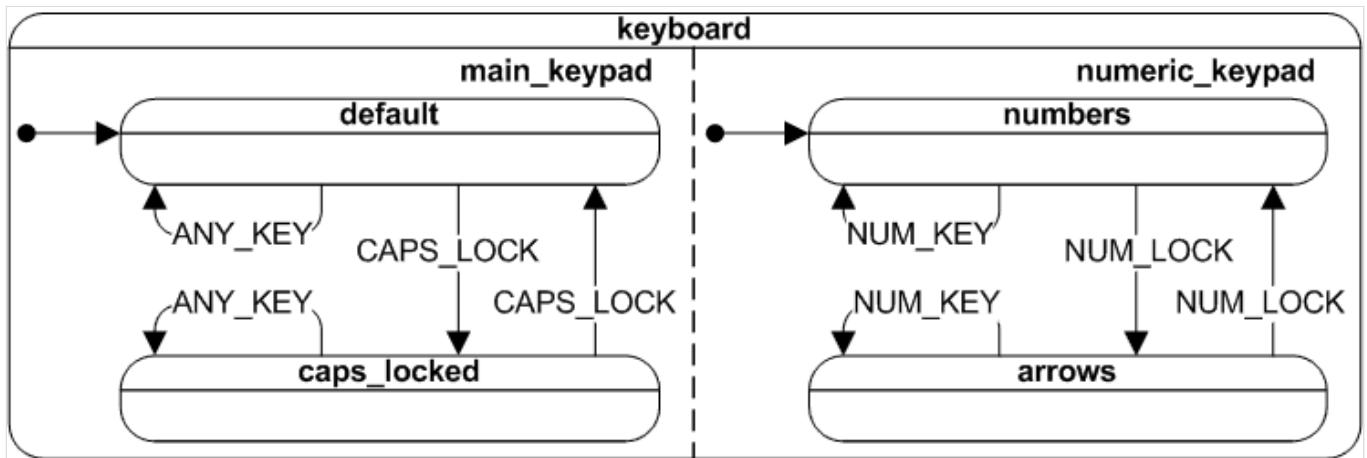


Figure 4: Two orthogonal regions (main keypad and numeric keypad) of a computer keyboard

Note that if the orthogonal regions are fully independent of each other, their combined complexity is simply additive, which means that the number of independent states needed to model the system is simply the sum $k + l + m + \dots$, where k, l, m, \dots denote numbers of OR-states in each orthogonal region. However, the general case of mutual dependency, on the other hand, results in multiplicative complexity, so in general, the number of states needed is the product $k \times l \times m \times \dots$.

In most real-life situations, orthogonal regions would be only approximately orthogonal (i.e. not truly independent). Therefore, UML statecharts provide a number of ways for orthogonal regions to communicate and synchronize their behaviors. Among these rich sets of (sometimes complex) mechanisms, perhaps the most important feature is that orthogonal regions can coordinate their behaviors by sending event instances to each other.

Even though orthogonal regions imply independence of execution (allowing more or less concurrency), the UML specification does not require that a separate thread of execution be assigned to each orthogonal region (although this can be done if desired). In fact, most commonly, orthogonal regions execute within the same thread.^[12] The UML specification requires only that the designer does not rely on any particular order for event instances to be dispatched to the relevant orthogonal regions.

Entry and exit actions

Every state in a UML statechart can have optional **entry actions**, which are executed upon entry to a state, as well as optional **exit actions**, which are executed upon exit from a state. Entry and exit actions are associated with states, not transitions. Regardless of how a state is entered or exited, all its entry and exit actions will be executed. Because

of this characteristic, statecharts behave like Moore machines. The UML notation for state entry and exit actions is to place the reserved word "entry" (or "exit") in the state right below the name compartment, followed by the forward slash and the list of arbitrary actions (see Figure 5).

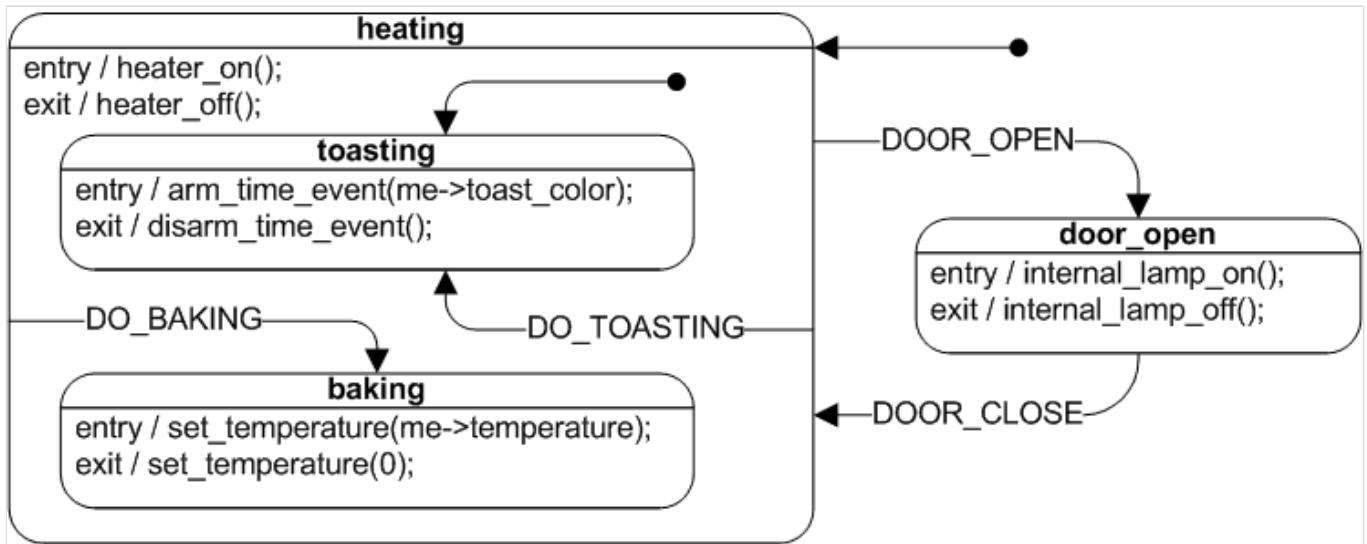


Figure 5: Toaster oven state machine with entry and exit actions

The value of entry and exit actions is that they provide means for *guaranteed initialization and cleanup*, very much like class constructors and destructors in Object-oriented programming. For example, consider the "door_open" state from Figure 5, which corresponds to the toaster oven behavior while the door is open. This state has a very important safety-critical requirement: Always disable the heater when the door is open. Additionally, while the door is open, the internal lamp illuminating the oven should light up.

Of course, such behavior could be modeled by adding appropriate actions (disabling the heater and turning on the light) to every transition path leading to the "door_open" state (the user may open the door at any time during "baking" or "toasting" or when the oven is not used at all). It should not be forgotten to extinguish the internal lamp with every transition leaving the "door_open" state. However, such a solution would cause the repetition of actions in many transitions. More importantly, such an approach leaves the design error-prone during subsequent amendments to behavior (e.g., the next programmer working on a new feature, such as top-browning, might simply forget to disable the heater on transition to "door_open").

Entry and exit actions allow implementation of desired behavior in a safer, simpler, and more intuitive way. As shown in Figure 5, it could be specified that the exit action from "heating" disables the heater, the entry action to "door_open" lights up the oven lamp, and the exit action from "door_open" extinguishes the lamp. The use of entry and exit actions is preferable to placing an action on a transition because it avoids repetitive coding and improves function by eliminating a safety hazard; (heater on while door open). The semantics of exit actions guarantees that, regardless of the transition path, the heater will be disabled when the toaster is not in the "heating" state.

Because entry actions are executed automatically whenever an associated state is entered, they often determine the conditions of operation or the identity of the state, very much as a class constructor determines the identity of the object being constructed. For example, the identity of the "heating" state is determined by the fact that the heater is turned on. This condition must be established before entering any substate of "heating" because entry actions to a substate of "heating," like "toasting," rely on proper initialization of the "heating" superstate and perform only the differences from this initialization. Consequently, the order of execution of entry actions must always proceed from the outermost state to the innermost state (top-down).

Not surprisingly, this order is analogous to the order in which class constructors are invoked. Construction of a class always starts at the very root of the class hierarchy and follows through all inheritance levels down to the class being instantiated. The execution of exit actions, which corresponds to destructor invocation, proceeds in the exact reverse order (bottom-up).

Internal transitions

Very commonly, an event causes only some internal actions to execute but does not lead to a change of state (state transition). In this case, all actions executed comprise the **internal transition**. For example, when one types on a keyboard, it responds by generating different character codes. However, unless the Caps Lock key is pressed, the state of the keyboard does not change (no state transition occurs). In UML, this situation should be modeled with internal transitions, as shown in Figure 6. The UML notation for internal transitions follows the general syntax used for exit (or entry) actions, except instead of the word entry (or exit) the internal transition is labeled with the triggering event (e.g., see the internal transition triggered by the ANY_KEY event in Figure 6).

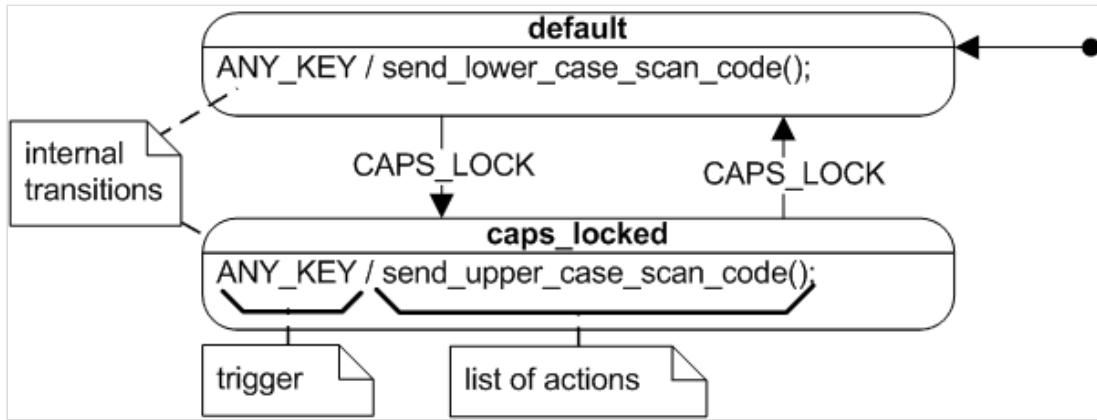


Figure 6: UML state diagram of the keyboard state machine with internal transitions

In the absence of entry and exit actions, internal transitions would be identical to **self-transitions** (transitions in which the target state is the same as the source state). In fact, in a classical Mealy machine, actions are associated exclusively with state transitions, so the only way to execute actions without changing state is through a self-transition (depicted as a directed loop in Figure 1 from the top of this article). However, in the presence of entry and exit actions, as in UML statecharts, a self-transition involves the execution of exit and entry actions and therefore it is distinctively different from an internal transition.

In contrast to a self-transition, no entry or exit actions are ever executed as a result of an internal transition, even if the internal transition is inherited from a higher level of the hierarchy than the currently active state. Internal transitions inherited from superstates at any level of nesting act as if they were defined directly in the currently active state.

Transition execution sequence

State nesting combined with entry and exit actions significantly complicates the state transition semantics in HSMs compared to the traditional FSMs. When dealing with hierarchically nested states and orthogonal regions, the simple term *current state* can be quite confusing. In an HSM, more than one state can be active at once. If the state machine is in a leaf state that is contained in a composite state (which is possibly contained in a higher-level composite state, and so on), all the composite states that either directly or transitively contain the leaf state are also active. Furthermore, because some of the composite states in this hierarchy might have orthogonal regions, the current active state is actually represented by a tree of states starting with the single region at the root down to individual simple states at the leaves. The UML specification refers to such a state tree as state configuration.^[1]

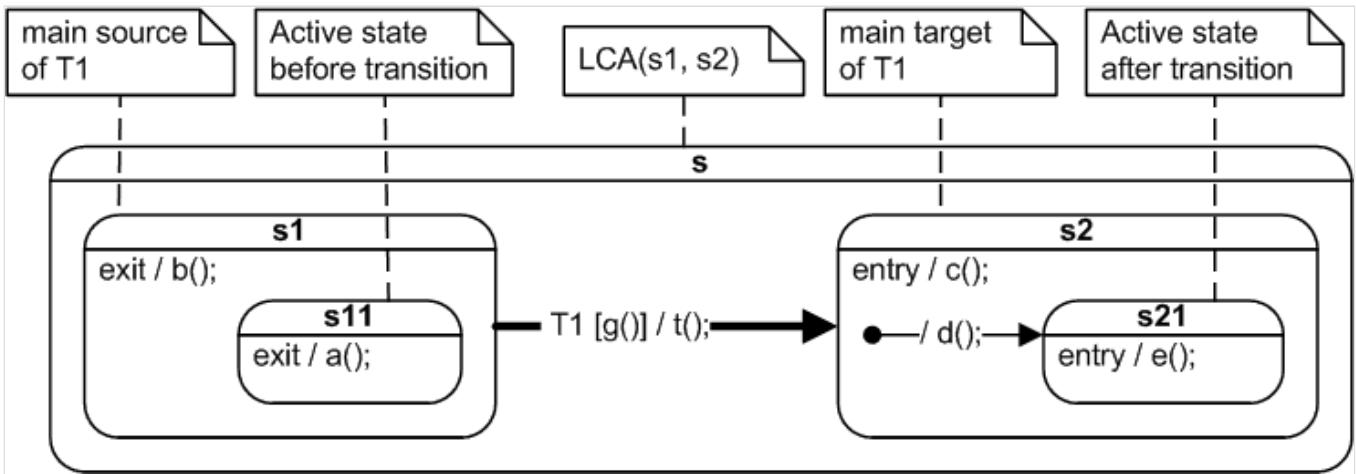


Figure 7: State roles in a state transition

In UML, a state transition can directly connect any two states. These two states, which may be composite, are designated as the **main source** and the **main target** of a transition. Figure 7 shows a simple transition example and explains the state roles in that transition. The UML specification prescribes that taking a state transition involves executing the actions in the following predefined sequence (see Section 14.2.3.9.6 of *OMG Unified Modeling Language (OMG UML)*^[1]):

1. Evaluate the guard condition associated with the transition and perform the following steps only if the guard evaluates to TRUE.
2. Exit the source state configuration.
3. Execute the actions associated with the transition.
4. Enter the target state configuration.

The transition sequence is easy to interpret in the simple case of both the main source and the main target nesting at the same level. For example, transition T1 shown in Figure 7 causes the evaluation of the guard $g()$; followed by the sequence of actions: $a()$; $b()$; $t()$; $c()$; $d()$; and $e()$; assuming that the guard $g()$ evaluates to TRUE.

However, in the general case of source and target states nested at different levels of the state hierarchy, it might not be immediately obvious how many levels of nesting need to be exited. The UML specification^[1] prescribes that a transition involves exiting all nested states from the current active state (which might be a direct or transitive substate of the main source state) up to, but not including, the **least common ancestor** (LCA) state of the main source and main target states. As the name indicates, the LCA is the lowest composite state that is simultaneously a superstate (ancestor) of both the source and the target states. As described before, the order of execution of exit actions is always from the most deeply nested state (the current active state) up the hierarchy to the LCA but without exiting the LCA. For instance, the LCA(s_1, s_2) of states " s_1 " and " s_2 " shown in Figure 7 is state " s ".

Entering the target state configuration commences from the level where the exit actions left off (i.e., from inside the LCA). As described before, entry actions must be executed starting from the highest-level state down the state hierarchy to the main target state. If the main target state is composite, the UML semantics prescribes to "drill" into its submachine recursively using the local initial transitions. The target state configuration is completely entered only after encountering a leaf state that has no initial transitions.

Local versus external transitions

Before UML 2,^[1] the only transition semantics in use was the **external transition**, in which the main source of the transition is always exited and the main target of the transition is always entered. UML 2 preserved the "external transition" semantics for backward compatibility, but introduced also a new kind of transition called **local transition** (see Section 14.2.3.4.4 of *Unified Modeling Language (UML)*^[1]). For many transition topologies, external and local

transitions are actually identical. However, a local transition doesn't cause exit from and reentry to the main source state if the main target state is a substate of the main source. In addition, a local state transition doesn't cause exit from and reentry to the main target state if the main target is a superstate of the main source state.

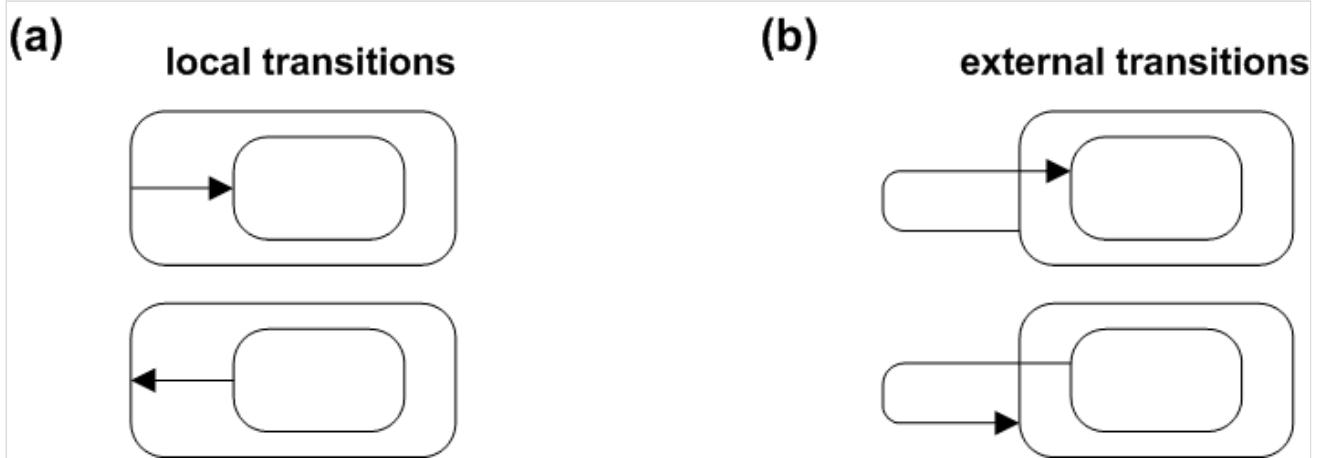


Figure 8: Local (a) versus external transitions (b).

Figure 8 contrasts local (a) and external (b) transitions. In the top row, you see the case of the main source containing the main target. The local transition does not cause exit from the source, while the external transition causes exit and reentry to the source. In the bottom row of Figure 8, you see the case of the main target containing the main source. The local transition does not cause entry to the target, whereas the external transition causes exit and reentry to the target.

Event deferral

Sometimes an event arrives at a particularly inconvenient time, when a state machine is in a state that cannot handle the event. In many cases, the nature of the event is such that it can be postponed (within limits) until the system enters another state, in which it is better prepared to handle the original event.

UML state machines provide a special mechanism for **deferring events** in states. In every state, you can include a clause `[event list]/defer`. If an event in the current state's deferred event list occurs, the event will be saved (deferred) for future processing until a state is entered that does not list the event in its deferred event list. Upon entry to such a state, the UML state machine will automatically recall any saved event(s) that are no longer deferred and will then either consume or discard these events. It is possible for a superstate to have a transition defined on an event that is deferred by a substate. Consistent with other areas in the specification of UML state machines, the substate takes precedence over the superstate, the event will be deferred and the transition for the superstate will not be executed. In the case of orthogonal regions where one orthogonal region defers an event and another consumes the event, the consumer takes precedence and the event is consumed and not deferred.

The limitations of UML state machines

Harel statecharts, which are the precursors of UML state machines, have been invented as "a visual formalism for complex systems",^[2] so from their inception, they have been inseparably associated with graphical representation in the form of state diagrams. However, it is important to understand that the concept of UML state machine transcends any particular notation, graphical or textual. The UML specification^[1] makes this distinction apparent by clearly separating state machine semantics from the notation.

However, the notation of UML statecharts is not purely visual. Any nontrivial state machine requires a large amount of textual information (e.g., the specification of actions and guards). The exact syntax of action and guard expressions isn't defined in the UML specification, so many people use either structured English or, more formally,

expressions in an implementation language such as C, C++, or Java.^[13] In practice, this means that UML statechart notation depends heavily on the specific programming language.

Nevertheless, most of the statecharts semantics are heavily biased toward graphical notation. For example, state diagrams poorly represent the sequence of processing, be it order of evaluation of guards or order of dispatching events to orthogonal regions. The UML specification sidesteps these problems by putting the burden on the designer not to rely on any particular sequencing. However, it is the case that when UML state machines are actually implemented, there is inevitably full control over order of execution, giving rise to criticism that the UML semantics may be unnecessarily restrictive. Similarly, statechart diagrams require a lot of plumbing gear (pseudostates, like joins, forks, junctions, choicepoints, etc.) to represent the flow of control graphically. In other words, these elements of the graphical notation do not add much value in representing flow of control as compared to plain structured code.

The UML notation and semantics are really geared toward computerized UML tools. A UML state machine, as represented in a tool, is not just the state diagram, but rather a mixture of graphical and textual representation that precisely captures both the state topology and the actions. The users of the tool can get several complementary views of the same state machine, both visual and textual, whereas the generated code is just one of the many available views.

References

1. "StateMachines". *Unified Modeling Language 2.5.1* (<https://www.omg.org/spec/UML/2.5.1/PDF>). OMG Document Number formal/2017-12-05. Object Management Group Standards Development Organization (OMG SDO). December 2017. p. 305.
2. Harel, David (1987). "Statecharts: A Visual Formalism for Complex Systems" (<http://www.wisdom.weizmann.ac.il/~dharel/SCANNED.PAPERS/Statecharts.pdf>) (PDF).
3. D. Drusinsky, *Modelling and verification using UML statecharts* (http://www.elsevier.com/wps/find/bookdescription.cws_home/707940/description#description), Elsevier, 2006
4. Samek, Miro (March 2009). "A crash course in UML state machines" (<https://www.embedded.com/design/prototyping-and-development/4008247/A-crash-course-in-UML-state-machines-Part-1>).
5. Samek, Miro (2008). *Practical UML Statecharts in C/C++*, Second Edition: Event-Driven Programming for Embedded Systems. Newnes. p. 728. ISBN 978-0-7506-8706-5.
6. Samek, Miro (April 2003). "Who Moved My State?" (<http://www.ddj.com/cpp/184401643>). C/C++ Users Journal, The Embedded Angle column.
7. Selic, Bran; Gullekson, Garth; Ward, Paul T. (1994). *Real-Time Object-Oriented Modeling*. John Wiley & Sons. p. 525. ISBN 0-471-59917-4.
8. Samek, Miro (August 2003). "Back to Basics" (<http://www.ddj.com/cpp/184401737>). C/C++ Users Journal, The Embedded Angle column.
9. "Region". *Unified Modeling Language 2.5.1* (<https://www.omg.org/spec/UML/2.5.1/PDF>). OMG Document Number formal/2017-12-05. Object Management Group Standards Development Organization (OMG SDO). December 2017. p. 352.
10. Samek, Miro (June 2003). "Dj Vu" (<https://web.archive.org/web/20120930224548/http://www.drdobbs.com/dj-vu/184401665>). C/C++ Users Journal, The Embedded Angle column. Archived from the original (<http://www.ddj.com/cpp/184401665>) on 2012-09-30.
11. Harel, David; Politi, Michal (1998). *Modeling Reactive Systems with Statecharts, the STATEMATE Approach*. McGraw-Hill. p. 258. ISBN 0-07-026205-5.
12. Douglass, Bruce Powel (1999). *Doing Hard Time: Developing Real-Time Systems with UML, Objects, Frameworks, and Patterns* (<https://archive.org/details/doinghardtimedev0000doug/page/749>). Addison Wesley. p. 749 (<https://archive.org/details/doinghardtimedev0000doug/page/749>). ISBN 0-201-49837-5.
13. Douglass, Bruce Powel (January 1999). "UML Statecharts" (<https://www.embedded.com/design/uml/4219602/UML-Statecharts>). Embedded Systems Programming.

External links

- Unified Modeling Language 2.5.1 (<https://www.omg.org/spec/UML/2.5.1/PDF>). OMG Document Number formal/2017-12-05. Object Management Group Standards Development Organization (OMG SDO). December 2017.
 - UML 2 State Machine Diagrams (<http://www.uml-diagrams.org/state-machine-diagrams.html>)
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=UML_state_machine&oldid=1265271193#Hierarchically_nested_states"



Computer hardware

Computer hardware includes the physical parts of a [computer](#), such as the [central processing unit](#) (CPU), [random-access memory](#) (RAM), [motherboard](#), [computer data storage](#), [graphics card](#), [sound card](#), and [computer case](#). It includes external devices such as a [monitor](#), [mouse](#), [keyboard](#), and [speakers](#).^{[1][2]}

By contrast, [software](#) is a set of written instructions that can be stored and run by hardware. Hardware derived its name from the fact it is [hard](#) or rigid with respect to changes, whereas software is [soft](#) because it is easy to change.

Hardware is typically directed by the software to execute any command or [instruction](#). A combination of hardware and software forms a usable [computing system](#), although [other systems](#) exist with only hardware.



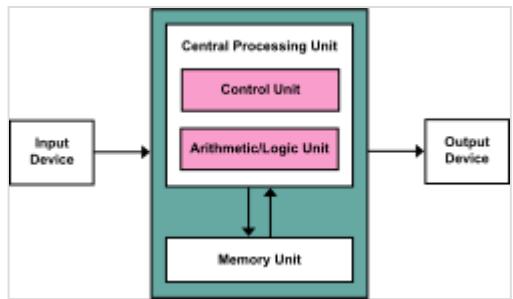
PDP-11 CPU board

History

Early computing devices were more complicated than the ancient [abacus](#) date to the seventeenth century. French mathematician [Blaise Pascal](#) designed a gear-based device that could add and subtract, selling around 50 models. The [stepped reckoner](#) was invented by [Gottfried Leibniz](#) by 1676, which could also divide and multiply. Due to the limitations of contemporary fabrication and design flaws, Leibniz' reckoner was not very functional, but similar devices ([Leibniz wheel](#)) remained in use into the 1970s.^[3] In the 19th century, Englishman [Charles Babbage](#) invented the [difference engine](#), a mechanical device to calculate [polynomials](#) for astronomical purposes.^[4] Babbage also designed a general-purpose computer that was never built. Much of the design was incorporated into the earliest computers: [punch cards](#) for input and output, [memory](#), an arithmetic unit analogous to [central processing units](#), and even a primitive [programming language](#) similar to [assembly language](#).^[5]

In 1936, [Alan Turing](#) developed the [universal Turing machine](#) to model any type of computer, proving that no computer would be able to solve the [decision problem](#).^[6] The universal Turing machine was a type of [stored-program computer](#) capable of mimicking the operations of any [Turing machine](#) (computer model) based on the [software](#) instructions passed to it. The storage of [computer programs](#) is key to the operation of modern computers and is the connection between computer hardware and software.^[7] Even prior to this, in the mid-19th century mathematician [George Boole](#) invented [Boolean algebra](#)—a system of logic where each [proposition](#) is either true or false. Boolean algebra is now the basis of the [circuits](#) that model the [transistors](#) and other components of [integrated circuits](#) that make up modern computer hardware.^[8] In 1945, Turing finished the design for a computer (the [Automatic Computing Engine](#)) that was never built.^[9]

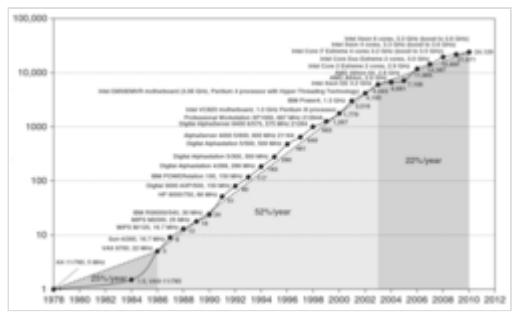
Around this time, technological advancement in relays and vacuum tubes enabled the construction of the first computers.^[10] Building on Babbage's design, relay computers were built by George Stibitz at Bell Laboratories and Harvard University's Howard Aiken, who engineered the MARK I.^[5] Also in 1945, mathematician John von Neumann—working on the ENIAC project at the University of Pennsylvania—devised the underlying von Neumann architecture that has served as the template for most modern computers.^[11] Von Neumann's design featured a centralized memory that stored both data and programs, a central processing unit (CPU) with priority of access to the memory, and input and output (I/O) units. Von Neumann used a single bus to transfer data, meaning that his solution to the storage problem by locating programs and data adjacent to each other created the Von Neumann bottleneck when the system tries to fetch both at the same time—often throttling the system's performance.^[12]



Von Neumann architecture scheme

Computer architecture

Computer architecture requires prioritizing between different goals, such as cost, speed, availability, and energy efficiency. The designer must have a good grasp of the hardware requirements and many different aspects of computing, from compilers to integrated circuit design.^[14] Cost has also become a significant constraint for manufacturers seeking to sell their products for less money than competitors offering a very similar hardware component. Profit margins have also been reduced.^[15] Even when the performance is not increasing, the cost of components has been dropping over time due to improved manufacturing techniques that have fewer components rejected at quality assurance stage.^[16]



Growth in processor performance (as measured by benchmarks),^[13] 1978–2010

Instruction set architecture

The most common instruction set architecture (ISA)—the interface between a computer's hardware and software—is based on the one devised by von Neumann in 1945.^[17] Despite the separation of the computing unit and the I/O system in many diagrams, typically the hardware is shared, with a bit in the computing unit indicating whether it is in computation or I/O mode.^[18] Common types of ISAs include CISC (complex instruction set computer), RISC (reduced instruction set computer), vector operations, and hybrid modes.^[19] CISC involves using a larger expression set to minimize the number of instructions the machines need to use.^[20] Based on a recognition that only a few instructions are commonly used, RISC shrinks the instruction set for added simplicity, which also enables the inclusion of more registers.^[21] After the invention of RISC in the 1980s, RISC based architectures that used pipelining and caching to increase performance displaced CISC architectures, particularly in applications with restrictions on power usage or space (such as mobile phones). From 1986 to 2003, the annual rate of improvement in hardware performance exceeded 50 percent, enabling the development of new computing

devices such as tablets and mobiles.^[22] Alongside the density of transistors, DRAM memory as well as flash and magnetic disk storage also became exponentially more compact and cheaper. The rate of improvement slackened off in the twenty-first century.^[23]

In the twenty-first century, increases in performance have been driven by increasing exploitation of parallelism.^[24] Applications are often parallelizable in two ways: either the same function is running across multiple areas of data (data parallelism) or different tasks can be performed simultaneously with limited interaction (task parallelism).^[25] These forms of parallelism are accommodated by various hardware strategies, including instruction-level parallelism (such as instruction pipelining), vector architectures and graphical processing units (GPUs) that are able to implement data parallelism, thread-level parallelism and request-level parallelism (both implementing task-level parallelism).^[25]

Microarchitecture

Microarchitecture, also known as computer organization, refers to high-level hardware questions such as the design of the CPU, memory, and memory interconnect.^[26] Memory hierarchy ensures that the memory quicker to access (and more expensive) is located closer to the CPU, while slower, cheaper memory for large-volume storage is located further away.^[27] Memory is typically segregated to separate programs from data and limit an attacker's ability to alter programs.^[28] Most computers use virtual memory to simplify addressing for programs, using the operating system to map virtual memory to different areas of the finite physical memory.^[29]

Cooling

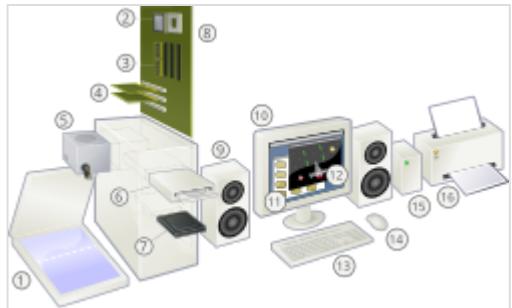
Computer processors generate heat, and excessive heat impacts their performance and can harm the components. Many computer chips will automatically throttle their performance to avoid overheating. Computers also typically have mechanisms for dissipating excessive heat, such as air or liquid coolers for the CPU and GPU and heatsinks for other components, such as the RAM. Computer cases are also often ventilated to help dissipate heat from the computer.^[30] Data centers typically use more sophisticated cooling solutions to keep the operating temperature of the entire center safe. Air-cooled systems are more common in smaller or older data centers, while liquid-cooled immersion (where each computer is surrounded by cooling fluid) and direct-to-chip (where the cooling fluid is directed to each computer chip) can be more expensive but are also more efficient.^[31] Most computers are designed to be more powerful than their cooling system, but their sustained operations cannot exceed the capacity of the cooling system.^[32] While performance can be temporarily increased when the computer is not hot (overclocking),^[33] in order to protect the hardware from excessive heat, the system will automatically reduce performance or shut down the processor if necessary.^[32] Processors also will shut off or enter a low power mode when inactive to reduce heat.^[34] Power delivery as well as heat dissipation are the most challenging aspects of hardware design,^[35] and have been the limiting factor to the development of smaller and faster chips since the early twenty-first century.^[34] Increases in performance require a commensurate increase in energy use and cooling demand.^[36]

Types of computer hardware systems

Personal computer

The personal computer is one of the most common types of computer due to its versatility and relatively low price.

- Desktop personal computers have a monitor, a keyboard, a mouse, and a computer case. The computer case holds the motherboard, fixed or removable disk drives for data storage, the power supply, and may contain other peripheral devices such as modems or network interfaces. Some models of desktop computers integrated the monitor and keyboard into the same case as the processor and power supply. Separating the elements allows the user to arrange the components in a pleasing, comfortable array, at the cost of managing power and data cables between them.
- Laptops are designed for portability but operate similarly to desktop PCs.^[37] They may use lower-power or reduced size components, with lower performance than a similarly priced desktop computer.^[38] Laptops contain the keyboard, display, and processor in one case. The monitor in the folding upper cover of the case can be closed for transportation, to protect the screen and keyboard. Instead of a mouse, laptops may have a touchpad or pointing stick.
- Tablets are portable computers that use a touch screen as the primary input device. Tablets generally weigh less and are smaller than laptops. Some tablets include fold-out keyboards or offer connections to separate external keyboards. Some models of laptop computers have a detachable keyboard, which allows the system to be configured as a touch-screen tablet. They are sometimes called 2-in-1 detachable laptops or tablet-laptop hybrids.^[39]
- Mobile phones are designed to have an extended battery life and light weight, while having less functionality than larger computers. They have diverse hardware architecture, often including antennas, microphones, cameras, GPS devices, and speakers. Power and data connections vary between phones.^[40]



Basic hardware components of a personal computer, including a monitor, a motherboard, a CPU, a RAM, two expansion cards, a power supply, an optical disc drive, a hard disk drive, a keyboard and a mouse



Inside a custom-built computer: power supply at the bottom has its own cooling fan

Large-scale computers

- A mainframe computer is a much larger computer that typically fills a room and may cost many hundreds or thousands of times as much as a personal computer. They are designed to perform large numbers of calculations for governments and large enterprises.
- In the 1960s and 1970s, more and more departments started to use cheaper and dedicated systems for specific purposes like process control and laboratory automation. A

minicomputer, or colloquially **mini**, is a class of smaller computers that was developed in the mid-1960s^{[41][42]} and sold for much less than mainframe^[43] and mid-size computers from IBM and its direct competitors.

- Supercomputers can cost hundreds of millions of dollars. They are intended to maximize performance with floating-point arithmetic and running batch programs that take a very long time (such as weeks) to complete. As a result of the need for communication between parallel programs, the speed of the internal network must be prioritized.^[44]
- Warehouse scale computers are larger versions of cluster computers that came into fashion with software as a service provided via the internet. Their design is intended to minimize cost per operation and power usage, as they can cost over \$100 million for a warehouse and the computers which go inside (the computers must be replaced every few years). Although availability is crucial for SaaS products, the software is designed to compensate for availability failures—unlike supercomputers.^[44]



An IBM System z9 mainframe

Virtual hardware

Virtual hardware is software that mimics the function of hardware; it is commonly used in infrastructure as a Service (IaaS) and platform as a Service (PaaS).^[45]

Embedded system

Embedded systems have the most variation in their processing power and cost: from an 8-bit processor that could cost less than USD\$0.10, to higher-end processors capable of billions of operations per second and costing over USD\$100. Cost is a particular concern with these systems, with designers often choosing the cheapest option that satisfies the performance requirements.^[46]

Components

Case

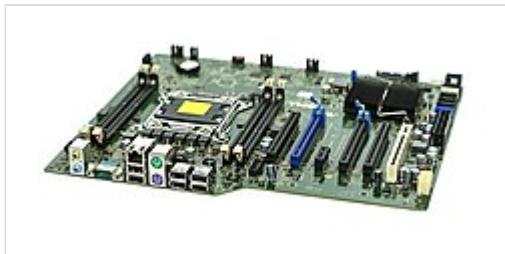
A computer case encloses most of the components of a desktop computer system. It provides mechanical support and protection for internal elements such as the motherboard, disk drives, and power supply, and controls and directs the flow of cooling air over internal components. The case is also part of the system to control electromagnetic interference radiated by the computer and protects internal parts from electrostatic discharge. Large tower cases provide space for multiple disk drives or other peripherals and usually stand on the floor, while desktop cases provide less expansion room. All-in-one style designs include a video display built into the same case. Portable and laptop computers require cases that provide impact protection for the unit. Hobbyists may decorate the cases with colored lights, paint, or other features, in an activity called case modding.

Power supply

Most personal computer power supply units meet the ATX standard and convert from alternating current (AC) at between 120 and 277 volts provided from a power outlet to direct current (DC) at a much lower voltage: typically 12, 5, or 3.3 volts.^[47]

Motherboard

The motherboard is the main component of a computer. It is a board with integrated circuitry that connects the other parts of the computer including the CPU, the RAM, the disk drives (CD, DVD, hard disk, or any others) as well as any peripherals connected via the ports or the expansion slots. The integrated circuit (IC) chips in a computer typically contain billions of tiny metal–oxide–semiconductor field-effect transistors (MOSFETs).^[48]



Computer motherboard

Components directly attached to or to part of the motherboard include:

- At least one CPU (central processing unit), which performs most of the calculations that enable a computer to function.^[49] It can be informally referred to as the brain of the computer.^[50] It takes program instructions from random-access memory (RAM), interprets and processes them and then sends back results so that the relevant components can carry out the instructions. The CPU is a microprocessor, which is fabricated on a metal–oxide–semiconductor (MOS) integrated circuit (IC) chip. It is usually cooled by a heatsink and fan, or water-cooling system. Many newer CPUs include an on-die graphics processing unit (GPU). The clock speed of the CPU governs how fast it executes instructions and is measured in GHz; typical values lie between 1 GHz and 5 GHz. There is also an increasing trend to add more cores to a processor—with each acting as if it were an independent processor—for increased parallelism.^[50]
- The internal bus connects the CPU to the main memory with several lines for simultaneous communication—typically 50 to 100—which are separated into those for addressing or memory, data, and command or control.^[51] Although parallel buses used to be more common, serial buses with a serializer to send more information over the same wire have become more common in the twenty-first century.^[52] Computers with multiple processors will need an interconnection bus, usually managed by a northbridge, while the southbridge manages communication with slower peripheral and I/O devices.^[53]
- Random-access memory (RAM), which stores the code and data that are being actively accessed by the CPU in a hierarchy based on when it is expected to be next used. Registers are closest to the CPU but have very limited capacity.^[54] CPUs also typically have multiple areas of cache memory that have much more capacity than registers, but much less than main memory; they are slower to access than registers, but much faster than main memory.^[55] Caching works by prefetching data before the CPU needs it, reducing latency.^{[55][56]} If the data the CPU needs is not in the cache, it can be accessed from main memory.^[55] Cache memory is typically SRAM, while the main memory is typically DRAM.^[27] RAM is volatile, meaning its contents will disappear if the computer powers down.^[57]
- Permanent storage or non-volatile memory is typically higher capacity and cheaper than memory, but takes much longer to access. Historically, such storage was typically provided in the form of a hard drive, but solid-state drives (SSD) are becoming cheaper and are much faster, thus leading to their increasing adoption. USB drives and network or cloud storage are also options.^[58]

- Read-only memory (ROM), which stores the BIOS that runs when the computer is powered on or otherwise begins execution, a process known as Bootstrapping, or booting or booting up. The ROM is typically a nonvolatile BIOS memory chip, which can only be written once with special technology.^[59]
 - The BIOS (Basic Input Output System) includes boot firmware and power management firmware. Newer motherboards use Unified Extensible Firmware Interface (UEFI) instead of BIOS.
- The CMOS (complementary MOS) battery, which powers the CMOS memory for date and time in the BIOS chip. This battery is generally a watch battery.
- Power MOSFETs make up the voltage regulator module (VRM), which controls how much voltage other hardware components receive.^[60]

Expansion cards

An expansion card in computing is a printed circuit board that can be inserted into an expansion slot of a computer motherboard or backplane to add functionality to a computer system via the expansion bus. Expansion cards can be used to obtain or expand on features not offered by the motherboard.^[61] Using expansion cards for a video processor used to be common, but modern computers are more likely to instead have a GPU integrated into the motherboard.^[62]

Input/output

Most computers also have an external data bus to connect peripheral devices to the motherboard. Most commonly, Universal Serial Bus (USB) is used.^[63] Unlike the internal bus, the external bus is connected using a bus controller that allows the peripheral system to operate at a different speed from the CPU.^[63] Input and output devices are used to receive data from the external world or write data respectively. Common examples include keyboards and mice (input) and displays and printers (output). Network interface controllers are used to access the Internet.^[64] USB ports also allow power to connected devices—a standard USB supplies power at 5 volts and up to 500 milliamps (2.5 watts), while powered USB ports with additional pins may allow the delivery of more power—up to 6 amps at 24v.^[65]

Sales

Global revenue from computer hardware in 2023 reached \$705.17 billion.^[66]

Recycling

Because computer parts contain hazardous materials, there is a growing movement to recycle old and outdated parts.^[67] Computer hardware contain dangerous chemicals such as lead, mercury, nickel, and cadmium. According to the EPA these e-wastes have a harmful effect on the environment unless they are disposed of properly. Making hardware requires energy, and recycling parts will reduce air pollution, water pollution, as well as greenhouse gas emissions.^[68] Disposing unauthorized computer equipment is in fact illegal. Legislation makes it mandatory to recycle computers through the government approved facilities. Recycling a computer can be made easier by taking out certain reusable parts. For example, the RAM, DVD drive, the graphics card, hard drive or SSD, and other similar removable parts can be reused.

Many materials used in computer hardware can be recovered by recycling for use in future production. Reuse of tin, silicon, iron, aluminum, and a variety of plastics that are present in bulk in computers or other electronics can reduce the costs of constructing new systems. Components frequently contain copper, gold, tantalum,^{[69][70]} silver, platinum, palladium, and lead as well as other valuable materials suitable for reclamation.^{[71][72]}

Toxic computer components

The central processing unit contains many toxic materials. It contains lead and chromium in the metal plates. Resistors, semiconductors, infrared detectors, stabilizers, cables, and wires contain cadmium. The circuit boards in a computer contain mercury, and chromium.^[73] When these types of materials, and chemicals are disposed improperly will become hazardous for the environment.

Environmental effects

When e-waste byproducts leach into groundwater, are burned, or get mishandled during recycling, it causes harm. Health problems associated with such toxins include impaired mental development, cancer, and damage to the lungs, liver, and kidneys.^[74] Computer components contain many toxic substances, like dioxins, polychlorinated biphenyls (PCBs), cadmium, chromium, radioactive isotopes and mercury. Circuit boards contain considerable quantities of lead-tin solders that are more likely to leach into groundwater or create air pollution due to incineration.^[75]

Recycling of computer hardware is considered environmentally friendly because it prevents hazardous waste, including heavy metals and carcinogens, from entering the atmosphere, landfill or waterways. While electronics consist a small fraction of total waste generated, they are far more dangerous. There is stringent legislation designed to enforce and encourage the sustainable disposal of appliances, the most notable being the Waste Electrical and Electronic Equipment Directive of the European Union and the United States National Computer Recycling Act.^[76]

Efforts for minimizing computer hardware waste

E-cycling, the recycling of computer hardware, refers to the donation, reuse, shredding and general collection of used electronics. Generically, the term refers to the process of collecting, brokering, disassembling, repairing and recycling the components or metals contained in used or discarded electronic equipment, otherwise known as electronic waste (e-waste). E-cyclable items include, but are not limited to: televisions, computers, microwave ovens, vacuum cleaners, telephones and cellular phones, stereos, and VCRs and DVDs just about anything that has a cord, light or takes some kind of battery.^[77]

Some companies, such as Dell and Apple, will recycle computers of their make or any other make. Otherwise, a computer can be donated to Computer Aid International which is an organization that recycles and refurbishes old computers for hospitals, schools, universities, etc.^[78]

See also



Electronics portal

- [Computer architecture](#)
- [Electronic hardware](#)
- [Hardware for artificial intelligence](#)
- [Glossary of computer hardware terms](#)
- [History of computing hardware](#)
- [Microprocessor](#)
- [MOSFET](#)
- [List of computer hardware manufacturers](#)
- [Open-source computing hardware](#)
- [Open-source hardware](#)
- [Transistor](#)

References

1. "Parts of computer" (<http://windows.microsoft.com/en-us/windows-vista/parts-of-a-computer>). Microsoft. Archived (<https://web.archive.org/web/20131127145607/http://windows.microsoft.com/en-US/windows-vista/Parts-of-a-computer>) from the original on 27 November 2013. Retrieved 5 December 2013.
2. Gilster, Ron (2001). *PC hardware : a beginner's guide* (<http://archive.org/details/pchardwarebeginn00gils>). Internet Archive. New York; London : McGraw-Hill. ISBN 978-0-07-212990-8.
3. Blum 2011, p. 13–14.
4. Blum 2011, p. 14.
5. Blum 2011, p. 15.
6. Blum 2011, pp. 21, 23.
7. Blum 2011, p. 25.
8. Blum 2011, pp. 34–35.
9. Blum 2011, pp. 71–72.
10. Blum 2011, p. 72.
11. Blum 2011, pp. 72, 74.
12. Blum 2011, p. 74.
13. Hennessy & Patterson 2011, pp. 41–42.
14. Hennessy & Patterson 2011, p. 11.
15. Hennessy & Patterson 2011, pp. 27–28, 32.
16. Hennessy & Patterson 2011, p. 27.
17. Mendelson 2022, p. 2.
18. Mendelson 2022, pp. 2–3.
19. Mendelson 2022, p. 3.
20. Mendelson 2022, p. 8.
21. Mendelson 2022, p. 15.
22. Hennessy & Patterson 2011, p. 2.

23. Hennessy & Patterson 2011, pp. 17–18.
24. Hennessy & Patterson 2011, pp. 9, 44.
25. Hennessy & Patterson 2011, p. 9.
26. Hennessy & Patterson 2011, p. 15.
27. Wang 2021, p. 3.
28. Wang 2021, p. 4.
29. Wang 2021, pp. 51–52.
30. "PC Cooling: The Importance of Keeping Your PC Cool" (<https://www.intel.com/content/www/us/en/gaming/resources/pc-cooling-the-importance-of-keeping-your-pc-cool.html>). Intel. Retrieved 20 July 2024.
31. "Data Center Cooling: What are the top concepts you need to know?" (<https://gatewaymechanical.ca/data-center-cooling-what-you-need-to-know/>). gatewaymechanical.ca. 11 August 2021. Retrieved 20 July 2024.
32. Hennessy & Patterson 2011, p. 22.
33. Hennessy & Patterson 2011, p. 26.
34. Hennessy & Patterson 2011, p. 25.
35. Hennessy & Patterson 2011, p. 21.
36. Hennessy & Patterson 2011, p. 56.
37. *PC hardware : a beginner's guide* (<https://archive.org/details/pchardwarebeginn00gils/page/21>). Osborne/McGraw-Hill. 26 April 2001. pp. 21 (<https://archive.org/details/pchardwarebeginn00gils/page/21>). ISBN 9780072129908.
38. "Desktop computer vs. Laptop computer" (<https://www.computerhope.com/issues/ch001399.htm>). Computer Hope. 30 December 2019. Retrieved 15 January 2020.
39. Cipriani, Jason (29 May 2020). "Best 2-in-1 Detachable Laptops 2020: The Best Tablet-Laptop Hybrids" (<https://www.ign.com/articles/best-2-in-1-detachable-laptops>). IGN. Retrieved 20 July 2020.
40. Ahmed, Rizwan; Dharaskar, Rajiv V. (2009). *Mobile Forensics: An Introduction from Indian Law Enforcement Perspective*. Springer. p. 177. ISBN 978-3-642-00405-6.
41. Henderson, Rebecca M.; Newell, Richard G., eds. (2011). *Accelerating energy innovation : insights from multiple sectors*. Chicago: University of Chicago Press. p. 180. ISBN 978-0226326832.
42. Huang, Han-Way (2014). *The Atmel AVR microcontroller : MEGA and XMEGA in assembly and C*. Australia; United Kingdom: Delmar Cengage Learning. p. 4. ISBN 978-1133607298.
43. Estabrooks, Maurice (1995). *Electronic technology, corporate strategy, and world transformation* (<https://archive.org/details/electronictechno000esta>). Westport, Conn.: Quorum Books. p. 53 (<https://archive.org/details/electronictechno000esta/page/53>). ISBN 0899309690.
44. Hennessy & Patterson 2011, p. 8.
45. Wang 2021, pp. 1, 3.
46. Hennessy & Patterson 2011, pp. 8–9.
47. Wilson, Kevin (2022). *Exploring Computer Hardware: The Illustrated Guide to Understanding Computer Hardware, Components, Peripherals & Networks* ([https://books.google.com/books?id=RaN4EAAAQBAJ&dq=%22power+supply+\(PSU\)+converts+high%22&p=g=PT111](https://books.google.com/books?id=RaN4EAAAQBAJ&dq=%22power+supply+(PSU)+converts+high%22&p=g=PT111)). Elluminet Press. ISBN 978-1-913151-73-7.
48. "13 Sextillion & Counting: The Long & Winding Road to the Most Frequently Manufactured Human Artifact in History" (<https://www.computerhistory.org/atchm/13-sextillion-counting-the-long-winding-road-to-the-most-frequently-manufactured-human-artifact-in-history/>). Computer History Museum. 2 April 2018. Retrieved 28 July 2019.
49. Wang 2021, p. 8.

50. Wang 2021, p. 9.
51. Wang 2021, p. 75.
52. Wang 2021, p. 78.
53. Wang 2021, p. 90.
54. Wang 2021, p. 47.
55. Wang 2021, pp. 49–50.
56. Hennessy & Patterson 2011, p. 45.
57. Wang 2021, p. 54.
58. Wang 2021, pp. 55–56.
59. Wang 2021, p. 55.
60. Harding, Sharon (17 September 2019). "What Is a MOSFET? A Basic Definition" (https://www.tomshardware.com/uk/reviews/mosfet-defintion-transistor-pc-motherboard-psu-explained_6343.html). *Tom's Hardware*. Retrieved 7 November 2019.
61. "What is an Expansion Card?" (<https://www.lenovo.com/us/en/glossary/expansion-card/>). *Lenovo*. Retrieved 17 February 2025.
62. Wang 2021, p. 121.
63. Wang 2021, p. 74.
64. Wang 2021, p. 7.
65. Wang 2021, p. 116.
66. "Computer Hardware Market Size, Trends and Global Forecast To 2032" (<https://www.thebusinessresearchcompany.com/report/computer-hardware-global-market-report#:~:text=The%20global%20computer%20hardware%20market,least%20in%20the%20short%20term.>). *The Business Research Company*. Retrieved 3 March 2023.
67. "How to recycle your old computer" (<http://www.digitaltrends.com/computing/how-to-recycle-your-old-computer/>). *Digital Trends*. 18 December 2016. Archived (<https://web.archive.org/web/20170417055533/http://www.digitaltrends.com/computing/how-to-recycle-your-old-computer/>) from the original on 17 April 2017. Retrieved 18 April 2017.
68. "Newtech Recycling Specializes in Computer Disposal, Laptop Disposal, Desktop Disposal Mainframe Disposal and Server Disposal" (<https://web.archive.org/web/20170329142805/https://www.newtechrecycling.com/disposal/computer-recycling/computer-recycling.html>). *Newtech Recycling, Inc.* Archived from the original (<https://www.newtechrecycling.com/disposal/computer-recycling/computer-recycling.html>) on 29 March 2017. Retrieved 18 April 2017.
69. Robert-Tissot, Sarah (2011). "TANTALUM" (<https://web.archive.org/web/20170226021016/https://www.raci.org.au/document/item/488>). *Royal Australian Chemical Institute*. Archived from the original (<https://www.raci.org.au/document/item/488>) on 26 February 2017. Retrieved 3 March 2019.
70. Padilla, Abraham (February 2019). "TANTALUM" (<https://minerals.usgs.gov/minerals/pubs/commodity/niumbium/mcs-2019-tanta.pdf>) (PDF). *United states geological survey*. Retrieved 3 March 2019.
71. Bleiwas, D (July 2001). "Obsolete Computers, "Gold Mine," or High-Tech Trash? Resource Recovery from Recycling" (<https://pubs.usgs.gov/fs/fs060-01/fs060-01.pdf>) (PDF). *USGS*. Retrieved 4 March 2019.
72. LeBlanc, Rick. "Electronic Devices a Rich Source of Precious Metals for Recyclers" (<https://www.thebalancesmb.com/electronic-devices-source-of-metals-for-recyclers-2877986>). *The Balance Small Business*. Retrieved 4 March 2019.
73. "The Toxic Components of Computers and Monitors" (<http://smallbusiness.chron.com/toxic-components-computers-monitors-69693.html>). Archived (<https://web.archive.org/web/20170427192255/http://smallbusiness.chron.com/toxic-components-computers-monitors-69693.html>) from the original on 27 April 2017. Retrieved 26 April 2017.

74. "What's Going On with Electronic Waste? – Electronics TakeBack Coalition" (<https://web.archive.org/web/20170427103038/http://www.electronicstakeback.com/2015/05/28/whats-going-on-with-electronic-waste/>). Archived from the original on 27 April 2017. Retrieved 26 April 2017.
75. Toothman, Jessika (2 June 2008). "What Happens to your Discarded Old Computer?" (<http://computer.howstuffworks.com/discharged-old-computer1.htm>). *HowStuffWorks*.
76. National Computer Recycling Act of 2005, H.R. 425, 109th Cong. (2005–2006)
77. T. Gallo, Daniel (15 July 2013). "Broad Overview of E-Waste Management Policies in the U.S." (<https://www.epa.gov/sites/production/files/2014-05/documents/overview.pdf>) (PDF). www.epa.gov. Retrieved 17 January 2020.
78. Schofield, Jack (19 February 2015). "How can I safely recycle my old PCs?" (<https://www.theguardian.com/technology/askjack/2015/feb/19/how-safely-recycle-old-pcs-computers>). *The Guardian*. ISSN 0261-3077 (<https://search.worldcat.org/issn/0261-3077>). Archived (<https://web.archive.org/web/20170427103349/https://www.theguardian.com/technology/askjack/2015/feb/19/how-safely-recycle-old-pcs-computers>) from the original on 27 April 2017. Retrieved 26 April 2017.

Sources

- Blum, Edward K. (2011). *Computer Science: The Hardware, Software and Heart of It*. Springer Science & Business Media. ISBN 978-1-4614-1168-0.
- Hennessy, John L.; Patterson, David A. (2011). *Computer Architecture: A Quantitative Approach*. Elsevier. ISBN 978-0-12-383873-5.
- Mendelson, Avi (2022). "The Architecture" (https://link.springer.com/referenceworkentry/10.1007/978-981-15-6401-7_1-1). In Anupam Chattopadhyay (ed.). *Handbook of Computer Architecture*. Springer Nature. pp. 1–42. doi:[10.1007/978-981-15-6401-7_1-1](https://doi.org/10.1007/978-981-15-6401-7_1-1) (https://doi.org/10.1007%2F978-981-15-6401-7_1-1). ISBN 978-981-15-6401-7.
- Wang, Shuangbao Paul (2021). *Computer Architecture and Organization: Fundamentals and Architecture Security*. Springer Nature. ISBN 978-981-16-5662-0.

External links

-  Media related to [Computer hardware](#) at Wikimedia Commons
-  [Computer hardware](#) at Wikibooks
-  Learning materials related to [Computer hardware](#) at Wikiversity

Retrieved from "https://en.wikipedia.org/w/index.php?title=Computer_hardware&oldid=1288178400"



Hardware acceleration

Hardware acceleration is the use of computer hardware designed to perform specific functions more efficiently when compared to software running on a general-purpose central processing unit (CPU). Any transformation of data that can be calculated in software running on a generic CPU can also be calculated in custom-made hardware, or in some mix of both.

To perform computing tasks more efficiently, generally one can invest time and money in improving the software, improving the hardware, or both. There are various approaches with advantages and disadvantages in terms of decreased latency, increased throughput, and reduced energy consumption. Typical advantages of focusing on software may include greater versatility, more rapid development, lower non-recurring engineering costs, heightened portability, and ease of updating features or patching bugs, at the cost of overhead to compute general operations. Advantages of focusing on hardware may include speedup, reduced power consumption,^[1] lower latency, increased parallelism^[2] and bandwidth, and better utilization of area and functional components available on an integrated circuit; at the cost of lower ability to update designs once etched onto silicon and higher costs of functional verification, times to market, and the need for more parts. In the hierarchy of digital computing systems ranging from general-purpose processors to fully customized hardware, there is a tradeoff between flexibility and efficiency, with efficiency increasing by orders of magnitude when any given application is implemented higher up that hierarchy.^[3] This hierarchy includes general-purpose processors such as CPUs,^[4] more specialized processors such as programmable shaders in a GPU,^[5] applications implemented on field-programmable gate arrays (FPGAs),^[6] and fixed-function implemented on application-specific integrated circuits (ASICs).^[7]

Hardware acceleration is advantageous for performance, and practical when the functions are fixed, so updates are not as needed as in software solutions. With the advent of reprogrammable logic devices such as FPGAs, the restriction of hardware acceleration to fully fixed algorithms has eased since 2010, allowing hardware acceleration to be applied to problem domains requiring modification to algorithms and processing control flow.^{[8][9]} The disadvantage, however, is that in many open source projects, it requires proprietary libraries that not all vendors are keen to distribute or expose, making it difficult to integrate in such projects.

Overview

Integrated circuits are designed to handle various operations on both analog and digital signals. In computing, digital signals are the most common and are typically represented as binary numbers. Computer hardware and software use this binary representation to perform computations. This is done by processing Boolean functions on the binary input, and then outputting the results for storage or further processing by other devices.



A cryptographic accelerator card allows cryptographic operations to be performed at a faster rate.

Computational equivalence of hardware and software

Because all Turing machines can run any computable function, it is always possible to design custom hardware that performs the same function as a given piece of software. Conversely, software can always be used to emulate the function of a given piece of hardware. Custom hardware may offer higher performance per watt for the same functions that can be specified in software. Hardware description languages (HDLs) such as Verilog and VHDL can model the same semantics as software and synthesize the design into a netlist that can be programmed to an FPGA or composed into the logic gates of an ASIC.

Stored-program computers

The vast majority of software-based computing occurs on machines implementing the von Neumann architecture, collectively known as stored-program computers. Computer programs are stored as data and executed by processors. Such processors must fetch and decode instructions, as well as load data operands from memory (as part of the instruction cycle), to execute the instructions constituting the software program. Relying on a common cache for code and data leads to the "von Neumann bottleneck", a fundamental limitation on the throughput of software on processors implementing the von Neumann architecture. Even in the modified Harvard architecture, where instructions and data have separate caches in the memory hierarchy, there is overhead to decoding instruction opcodes and multiplexing available execution units on a microprocessor or microcontroller, leading to low circuit utilization. Modern processors that provide simultaneous multithreading exploit under-utilization of available processor functional units and instruction level parallelism between different hardware threads.

Hardware execution units

Hardware execution units do not in general rely on the von Neumann or modified Harvard architectures and do not need to perform the instruction fetch and decode steps of an instruction cycle and incur those stages' overhead. If needed calculations are specified in a register transfer level (RTL) hardware design, the time and circuit area costs that would be incurred by instruction fetch and decoding stages can be reclaimed and put to other uses.

This reclamation saves time, power, and circuit area in computation. The reclaimed resources can be used for increased parallel computation, other functions, communication, or memory, as well as increased input/output capabilities. This comes at the cost of general-purpose utility.

Emerging hardware architectures

Greater RTL customization of hardware designs allows emerging architectures such as in-memory computing, transport triggered architectures (TTA) and networks-on-chip (NoC) to further benefit from increased locality of data to execution context, thereby reducing computing and communication latency between modules and functional units.

Custom hardware is limited in parallel processing capability only by the area and logic blocks available on the integrated circuit die.^[10] Therefore, hardware is much more free to offer massive parallelism than software on general-purpose processors, offering a possibility of implementing the parallel random-access machine (PRAM) model.

It is common to build multicore and manycore processing units out of microprocessor IP core schematics on a single FPGA or ASIC.^{[11][12][13][14][15]} Similarly, specialized functional units can be composed in parallel, as in digital signal processing, without being embedded in a processor IP core. Therefore, hardware acceleration is often employed for repetitive, fixed tasks involving little conditional branching, especially on large amounts of data. This is how Nvidia's CUDA line of GPUs are implemented.

Implementation metrics

As device mobility has increased, new metrics have been developed that measure the relative performance of specific acceleration protocols, considering characteristics such as physical hardware dimensions, power consumption, and operations throughput. These can be summarized into three categories: task efficiency, implementation efficiency, and flexibility. Appropriate metrics consider the area of the hardware along with both the corresponding operations throughput and energy consumed.^[16]

Applications

Examples of hardware acceleration include bit blit acceleration functionality in graphics processing units (GPUs), use of memristors for accelerating neural networks, and regular expression hardware acceleration for spam control in the server industry, intended to prevent regular expression denial of service (ReDoS) attacks.^[17] The hardware that performs the acceleration may be part of a general-purpose CPU, or a separate unit called a hardware accelerator, though they are usually referred to with a more specific term, such as 3D accelerator, or cryptographic accelerator.

Traditionally, processors were sequential (instructions are executed one by one), and were designed to run general purpose algorithms controlled by instruction fetch (for example, moving temporary results to and from a register file). Hardware accelerators improve the execution of a specific algorithm by allowing greater concurrency, having specific datapaths for their temporary variables, and reducing the overhead of instruction control in the fetch-decode-execute cycle.

Modern processors are multi-core and often feature parallel "single-instruction; multiple data" (SIMD) units. Even so, hardware acceleration still yields benefits. Hardware acceleration is suitable for any computation-intensive algorithm which is executed frequently in a task or program. Depending upon the granularity, hardware acceleration can vary from a small functional unit, to a large functional block (like motion estimation in MPEG-2).

Hardware acceleration units by application

Application	Hardware accelerator	Acronym
Computer graphics	<p>Graphics processing unit</p> <ul style="list-style-type: none"> ▪ General-purpose computing on GPU ▪ CUDA architecture ▪ Ray-tracing hardware ▪ Various video acceleration hardware 	<p>GPU</p> <ul style="list-style-type: none"> ▪ GPGPU ▪ CUDA ▪ RTX ▪ N/A
Digital signal processing	Digital signal processor	DSP
Analog signal processing	Field-programmable analog array	FPA
	<ul style="list-style-type: none"> ▪ Field-programmable RF 	<ul style="list-style-type: none"> ▪ FPRF
Image processing	Webcam or image processor	IPU
Sound processing	Sound card and sound card mixer	N/A
Computer networking	<p>Network processor and network interface controller</p> <ul style="list-style-type: none"> ▪ Network on a chip ▪ TCP offload engine ▪ IPsec offload^[18] ▪ I/O Acceleration Technology 	<p>NPU and NIC</p> <ul style="list-style-type: none"> ▪ NoC ▪ TCPOE or TOE ▪ I/OAT or IOAT
Cryptography	<p>Cryptographic accelerator and secure cryptoprocessor</p> <ul style="list-style-type: none"> ▪ Hardware-based encryption ▪ AES instruction set ▪ SSL acceleration ▪ Custom hardware attack ▪ Hardware random number generator 	N/A
Artificial intelligence	<p>AI accelerator</p> <ul style="list-style-type: none"> ▪ Vision processing unit ▪ Physical neural network ▪ Neuromorphic engineering 	<ul style="list-style-type: none"> ▪ VPU ▪ PNN ▪ N/A
Multilinear algebra	Tensor processing unit	TPU
Physics simulation	Physics processing unit	PPU
Regular expressions ^[17]	Regular expression coprocessor	N/A
Data compression ^[19]	Data compression accelerator	N/A
In-memory processing	Network on a chip and Systolic array	NoC; N/A
Data processing	Data processing unit	DPU
Any computing task	<p>Computer hardware</p> <ul style="list-style-type: none"> ▪ Field-programmable gate arrays^[20] 	<p>HW (sometimes)</p> <ul style="list-style-type: none"> ▪ FPGA

<ul style="list-style-type: none"> ▪ Application-specific integrated circuits^[20] ▪ Complex programmable logic devices ▪ Systems-on-Chip <ul style="list-style-type: none"> ▪ Multi-processor system-on-chip ▪ Programmable system-on-chip 	<ul style="list-style-type: none"> ▪ ASIC ▪ CPLD ▪ SoC ▪ MPSoC ▪ PSoC
---	--

See also

- [Coprocessor](#)
- [DirectX Video Acceleration \(DXVA\)](#)
- [Direct memory access \(DMA\)](#)
- [High-level synthesis](#)
 - [C to HDL](#)
 - [Flow to HDL](#)
- [Soft microprocessor](#)
- [Flynn's taxonomy of parallel computer architectures](#)
 - [Single instruction, multiple data \(SIMD\)](#)
 - [Single instruction, multiple threads \(SIMT\)](#)
 - [Multiple instructions, multiple data \(MIMD\)](#)
- [Computer for operations with functions](#)

References

1. "Microsoft Supercharges Bing Search With Programmable Chips" (<https://www.wired.com/2014/06/microsoft-fpga/>). *WIRED*. 16 June 2014.
2. "Embedded" (<https://web.archive.org/web/20071008163016/http://www.embedded.com/columns/showArticle.jhtml?articleID=192700615>). Archived from the original (<https://www.embedded.com/columns/showArticle.jhtml?articleID=192700615>) on 2007-10-08. Retrieved 2012-08-18. "FPGA Architectures from 'A' to 'Z'" by Clive Maxfield 2006
3. Sinan, Kufeoglu; Mahmut, Ozkuran (2019). "Figure 5. CPU, GPU, FPGA, and ASIC minimum energy consumption between difficulty recalculation." (https://www.researchgate.net/figure/CPU-GPU-FPGA-and-ASIC-minimum-energy-consumption-between-difficulty-recalculation_fig5_337886683). *Energy Consumption of Bitcoin Mining*. doi:[10.17863/CAM.41230](https://doi.org/10.17863/CAM.41230) (<https://doi.org/10.17863%2FCAM.41230>).
4. Kim, Yeongmin; Kong, Joonho; Munir, Arslan (2020). "CPU-Accelerator Co-Scheduling for CNN Acceleration at the Edge" (<https://doi.org/10.1109%2FACCESS.2020.3039278>). *IEEE Access*. **8**: 211422–211433. Bibcode:2020IEEEA...8u1422K (<https://ui.adsabs.harvard.edu/abs/2020IEEEA...8u1422K>). doi:[10.1109/ACCESS.2020.3039278](https://doi.org/10.1109/ACCESS.2020.3039278) (<https://doi.org/10.1109%2FACCESS.2020.3039278>). ISSN 2169-3536 (<https://search.worldcat.org/issn/2169-3536>).
5. Lin, Yibo; Jiang, Zixuan; Gu, Jiaqi; Li, Wuxi; Dhar, Shounak; Ren, Haoxing; Khailany, Brucek; Pan, David Z. (April 2021). "DREAMPlace: Deep Learning Toolkit-Enabled GPU Acceleration for Modern VLSI Placement" (<https://ieeexplore.ieee.org/document/9122053>). *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. **40** (4): 748–761. doi:[10.1109/TCAD.2020.3003843](https://doi.org/10.1109/TCAD.2020.3003843) (<https://doi.org/10.1109%2FTCAD.2020.3003843>). ISSN 1937-4151 (<https://search.worldcat.org/issn/1937-4151>). S2CID 225744481 (<https://api.semanticscholar.org/CorpusID:225744481>).

6. Lyakhov, Pavel; Valueva, Maria; Valuev, Georgii; Nagornov, Nikolai (2020-12-18). "A Method of Increasing Digital Filter Performance Based on Truncated Multiply-Accumulate Units" (<https://doi.org/10.3390%2Fapp10249052>). *Applied Sciences*. **10** (24): 9052. doi:[10.3390/app10249052](https://doi.org/10.3390/app10249052) (<https://doi.org/10.3390%2Fapp10249052>). ISSN 2076-3417 (<https://search.worldcat.org/issn/2076-3417>). "Hardware simulation on FPGA increased the digital filter performance."
7. Mohan, Prashanth; Wang, Wen; Jungk, Bernhard; Niederhagen, Ruben; Szefer, Jakub; Mai, Ken (October 2020). "ASIC Accelerator in 28 nm for the Post-Quantum Digital Signature Scheme XMSS" (<https://ieeexplore.ieee.org/document/9283605>). *2020 IEEE 38th International Conference on Computer Design (ICCD)*. Hartford, CT, USA: IEEE. pp. 656–662. doi:[10.1109/ICCD50377.2020.00112](https://doi.org/10.1109/ICCD50377.2020.00112) (<https://doi.org/10.1109%2FICCD50377.2020.00112>). ISBN 978-1-7281-9710-4. S2CID 229330964 (<https://api.semanticscholar.org/CorpusID:229330964>).
8. Morgan, Timothy Pricket (2014-09-03). "How Microsoft Is Using FPGAs To Speed Up Bing Search" (<https://www.enterprisetech.com/2014/09/03/microsoft-using-fpgas-speed-bing-search/>). Enterprise Tech. Retrieved 2018-09-18.
9. "Project Catapult" (<https://www.microsoft.com/en-us/research/project/project-catapult/>). Microsoft Research.
10. MicroBlaze Soft Processor: Frequently Asked Questions (http://www.xilinx.com/products/design_resources/proc_central/microblaze_faq.pdf) Archived (https://web.archive.org/web/20111027074459/http://www.xilinx.com/products/design_resources/proc_central/microblaze_faq.pdf) 2011-10-27 at the Wayback Machine
11. Vassányi, István (1998). "Implementing processor arrays on FPGAs" (<https://doi.org/10.1007/BFb0055278>). *Field-Programmable Logic and Applications from FPGAs to Computing Paradigm*. Lecture Notes in Computer Science. Vol. 1482. pp. 446–450. doi:[10.1007/BFb0055278](https://doi.org/10.1007/BFb0055278) (<https://doi.org/10.1007/BFb0055278>). ISBN 978-3-540-64948-9.
12. Zhoukun WANG and Omar HAMMAMI. "A 24 Processors System on Chip FPGA Design with Network on Chip". [1] (<http://www.design-reuse.com/articles/21583/processor-noc-fpga.html>)
13. John Kent. "Micro16 Array - A Simple CPU Array" (<http://members.optusnet.com.au/jekent/Micro16Array/index.html>)
14. Kit Eaton. "1,000 Core CPU Achieved: Your Future Desktop Will Be a Supercomputer". 2011. [2] (<http://www.fastcompany.com/1714174/1000-core-cpu-achieved-your-future-desktop-will-be-a-supercomputer?partner=rss>)
15. "Scientists Squeeze Over 1,000 Cores onto One Chip". 2011. [3] (<http://www.ecnmag.com/news/2011/01/research/Over-1000-Cores-on-One-Chip.aspx>) Archived (<https://web.archive.org/web/20120305082424/http://www.ecnmag.com/news/2011/01/research/Over-1000-Cores-on-One-Chip.aspx>) 2012-03-05 at the Wayback Machine
16. Kienle, Frank; Wehn, Norbert; Meyr, Heinrich (December 2011). "On Complexity, Energy- and Implementation-Efficiency of Channel Decoders". *IEEE Transactions on Communications*. **59** (12): 3301–3310. arXiv:1003.3792 (<https://arxiv.org/abs/1003.3792>). doi:[10.1109/tcomm.2011.092011.100157](https://doi.org/10.1109/tcomm.2011.092011.100157) (<https://doi.org/10.1109%2Ftcomm.2011.092011.100157>). ISSN 0090-6778 (<https://search.worldcat.org/issn/0090-6778>). S2CID 13863870 (<https://api.semanticscholar.org/CorpusID:13863870>).
17. "Regular Expressions in hardware" (<http://www.wellho.net/regex/hardware.html>). Retrieved 17 July 2014.
18. <https://www.intel.com/content/dam/doc/product-brief/intel-pro-100-s-desktop-adapter-datasheet.pdf>
19. "Compression Accelerators - Microsoft Research" (<https://www.microsoft.com/en-us/research/project/compression-accelerators/>). Microsoft Research. Retrieved 2017-10-07.

20. Farabet, Clément, et al. "[Hardware accelerated convolutional neural networks for synthetic vision systems](https://www.academia.edu/download/43417932/Hardware_Accelerated_Convolutional_Neura20160306-12343-1rpenxt.pdf) (https://www.academia.edu/download/43417932/Hardware_Accelerated_Convolutional_Neura20160306-12343-1rpenxt.pdf)."
- Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on. IEEE, 2010.

External links

-  Media related to [Hardware acceleration](#) at Wikimedia Commons
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Hardware_acceleration&oldid=1292200736"



Digital audio

Digital audio is a representation of sound recorded in, or converted into, digital form. In digital audio, the sound wave of the audio signal is typically encoded as numerical samples in a continuous sequence. For example, in CD audio, samples are taken 44,100 times per second, each with 16-bit resolution. Digital audio is also the name for the entire technology of sound recording and reproduction using audio signals that have been encoded in digital form. Following significant advances in digital audio technology during the 1970s and 1980s, it gradually replaced analog audio technology in many areas of audio engineering, record production and telecommunications in the 1990s and 2000s.

In a digital audio system, an analog electrical signal representing the sound is converted with an analog-to-digital converter (ADC) into a digital signal, typically using pulse-code modulation (PCM). This digital signal can then be recorded, edited, modified, and copied using computers, audio playback machines, and other digital tools. For playback, a digital-to-analog converter (DAC) performs the reverse process, converting a digital signal back into an analog signal, which is then sent through an audio power amplifier and ultimately to a loudspeaker.

Digital audio systems may include compression, storage, processing, and transmission components. Conversion to a digital format allows convenient manipulation, storage, transmission, and retrieval of an audio signal. Unlike analog audio, in which making copies of a recording results in generation loss and degradation of signal quality, digital audio allows an infinite number of copies to be made without any degradation of signal quality.

Overview

Digital audio technologies are used in the recording, manipulation, mass-production, and distribution of sound, including recordings of songs, instrumental pieces, podcasts, sound effects, and other sounds. Modern online music distribution depends on digital recording and data compression. The availability of music as data files, rather than as physical objects, has significantly reduced the costs of distribution as well as making it easier to share copies.^[1] Before digital audio, the music industry distributed and sold music by selling physical copies in the form of records and cassette tapes. With digital audio and online distribution systems such as iTunes, companies sell digital sound files to consumers, which the consumer receives over the Internet. Popular streaming services such as Apple Music, Spotify, or YouTube, offer temporary access to the digital file, and are now the most common form of music consumption.^[2]



Audio levels display on a digital audio recorder (Zoom H4n)

An analog audio system converts physical waveforms of sound into electrical representations of those waveforms by use of a transducer, such as a microphone. The sounds are then stored on an analog medium such as magnetic tape, or transmitted through an analog medium such as a telephone line or radio. The process is reversed for reproduction: the electrical audio signal is amplified and then converted back into physical waveforms via a loudspeaker. Analog audio retains its fundamental wave-like characteristics throughout its storage, transformation, duplication, and amplification.

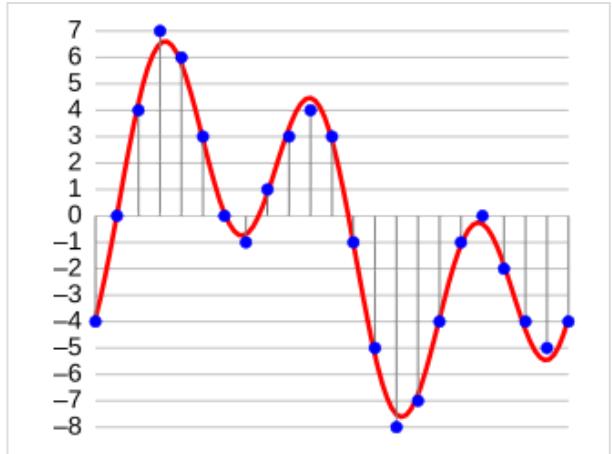
Analog audio signals are susceptible to noise and distortion, due to the innate characteristics of electronic circuits and associated devices. Disturbances in a digital system do not result in error unless they are so large as to result in a symbol being misinterpreted as another symbol or disturbing the sequence of symbols. It is, therefore, generally possible to have an entirely error-free digital audio system in which no noise or distortion is introduced between conversion to digital format and conversion back to analog.^[a]

A digital audio signal may be encoded for correction of any errors that might occur in the storage or transmission of the signal. This technique, known as channel coding, is essential for broadcast or recorded digital systems to maintain bit accuracy. Eight-to-fourteen modulation is the channel code used for the audio compact disc (CD).

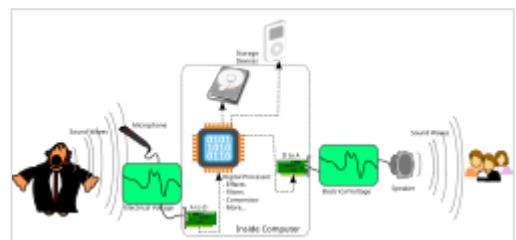
Conversion process

If an audio signal is analog, a digital audio system starts with an ADC that converts an analog signal to a digital signal.^[b] The ADC runs at a specified sampling rate and converts at a known bit resolution. CD audio, for example, has a sampling rate of 44.1 kHz (44,100 samples per second), and has 16-bit resolution for each stereo channel. Analog signals that have not already been bandlimited must be passed through an anti-aliasing filter before conversion, to prevent the aliasing distortion that is caused by audio signals with frequencies higher than the Nyquist frequency (half the sampling rate).

A digital audio signal may be stored or transmitted. Digital audio can be stored on a CD, a digital audio player, a hard drive, a USB flash drive, or any other digital data storage device. The digital signal may be altered through digital signal processing, where it may be filtered or have effects applied. Sample-rate conversion including upsampling and downsampling may be used to change signals that have been encoded with a different sampling rate to a common sampling rate prior to processing. Audio data compression techniques, such as MP3, Advanced Audio Coding (AAC), Opus, Ogg Vorbis, or FLAC, are



A sound wave, in red, represented digitally, in blue (after sampling and 4-bit quantization).



The lifecycle of sound from its source, through an ADC, digital processing, a DAC, and finally as sound again.

commonly employed to reduce the file size. Digital audio can be carried over digital audio interfaces such as AES3 or MADI. Digital audio can be carried over a network using audio over Ethernet, audio over IP or other streaming media standards and systems.

For playback, digital audio must be converted back to an analog signal with a DAC. According to the Nyquist–Shannon sampling theorem, with some practical and theoretical restrictions, a band-limited version of the original analog signal can be accurately reconstructed from the digital signal.

During conversion, audio data can be embedded with a digital watermark to prevent piracy and unauthorized use. Watermarking is done using a direct-sequence spread-spectrum (DSSS) method. The audio information is then modulated by a pseudo-noise (PN) sequence, then shaped within the frequency domain and put back in the original signal. The strength of the embedding determines the strength of the watermark on the audio data.^[4]

History

Coding

Pulse-code modulation (PCM) was invented by British scientist Alec Reeves in 1937.^[5] In 1950, C. Chapin Cutler of Bell Labs filed the patent on differential pulse-code modulation (DPCM),^[6] a data compression algorithm. Adaptive DPCM (ADPCM) was introduced by P. Cummiskey, Nikil S. Jayant and James L. Flanagan at Bell Labs in 1973.^{[7][8]}

Perceptual coding was first used for speech coding compression, with linear predictive coding (LPC).^[9] Initial concepts for LPC date back to the work of Fumitada Itakura (Nagoya University) and Shuzo Saito (Nippon Telegraph and Telephone) in 1966.^[10] During the 1970s, Bishnu S. Atal and Manfred R. Schroeder at Bell Labs developed a form of LPC called adaptive predictive coding (APC), a perceptual coding algorithm that exploited the masking properties of the human ear, followed in the early 1980s with the code-excited linear prediction (CELP) algorithm.^[9]

Discrete cosine transform (DCT) coding, a lossy compression method first proposed by Nasir Ahmed in 1972,^{[11][12]} provided the basis for the modified discrete cosine transform (MDCT), which was developed by J. P. Princen, A. W. Johnson and A. B. Bradley in 1987.^[13] The MDCT is the basis for most audio coding standards, such as Dolby Digital (AC-3),^[14] MP3 (MPEG Layer III),^{[15][9]} AAC, Windows Media Audio (WMA), Opus and Vorbis (Ogg).^[14]

Recording

PCM was used in telecommunications applications long before its first use in commercial broadcast and recording. Commercial digital recording was pioneered in Japan by NHK and Nippon Columbia and their Denon brand, in the 1960s. The first commercial digital recordings were released in 1971.^[16]

The BBC also began to experiment with digital audio in the 1960s. By the early 1970s, it had developed a 2-channel recorder, and in 1972 it deployed a digital audio transmission system that linked their broadcast center to their remote transmitters.^[16]

The first 16-bit PCM recording in the United States was made by Thomas Stockham at the Santa Fe Opera in 1976, on a Soundstream recorder. An improved version of the Soundstream system was used to produce several classical recordings by Telarc in 1978. The 3M digital multitrack recorder in development at the time was based on BBC technology. The first all-digital album recorded on this machine was Ry Cooder's *Bop till You Drop* in 1979. British record label Decca began development of its own 2-track digital audio recorders in 1978 and released the first European digital recording in 1979.^[16]

Popular professional digital multitrack recorders produced by Sony/Studer (DASH) and Mitsubishi (ProDigi) in the early 1980s helped to bring about digital recording's acceptance by the major record companies. Machines for these formats had their own transports built-in as well, using reel-to-reel tape in either 1/4", 1/2", or 1" widths, with the audio data being recorded to the tape using a multi-track stationary tape head. PCM adaptors allowed for stereo digital audio recording on a conventional NTSC or PAL video tape recorder.

The 1982 introduction of the CD by Philips and Sony popularized digital audio with consumers.^[16]

ADAT became available in the early 1990s, which allowed eight-track 44.1 or 48 kHz recording on S-VHS cassettes, and DTRS performed a similar function with Hi8 tapes.

Formats like ProDigi and DASH were referred to as **SDAT** (stationary-head digital audio tape) formats, as opposed to formats like the PCM adaptor-based systems and Digital Audio Tape (DAT), which were referred to as **RDAT** (rotating-head digital audio tape) formats, due to their helical-scan process of recording.

Like the DAT cassette, ProDigi and DASH machines also accommodated the obligatory 44.1 kHz sampling rate, but also 48 kHz on all machines, and eventually a 96 kHz sampling rate. They overcame the problems that made typical analog recorders unable to meet the bandwidth (frequency range) demands of digital recording by a combination of higher tape speeds, narrower head gaps used in combination with metal-formulation tapes, and the spreading of data across multiple parallel tracks.

Unlike analog systems, modern digital audio workstations and audio interfaces allow as many channels in as many different sampling rates as the computer can effectively run at a single time. Avid Audio and Steinberg released the first digital audio workstation software programs in 1989.^[17] Digital audio workstations make multitrack recording and mixing much easier for large projects which would otherwise be difficult with analog equipment.



Analog reel-to-reel tape recorder



Sony professional digital audio tape (DAT) recorder PCM-7030



Digital audio workstation

Telephony

The rapid development and wide adoption of PCM digital telephony was enabled by metal–oxide–semiconductor (MOS) switched capacitor (SC) circuit technology, developed in the early 1970s.^[18] This led to the development of PCM codec-filter chips in the late 1970s.^{[18][19]} The silicon-gate CMOS (complementary MOS) PCM codec-filter chip, developed by David A. Hodges and W.C. Black in 1980,^[18] has since been the industry standard for digital telephony.^{[18][19]} By the 1990s, telecommunication networks such as the public switched telephone network (PSTN) had been largely digitized with VLSI (very large-scale integration) CMOS PCM codec-filters, widely used in electronic switching systems for telephone exchanges, user-end modems and a range of digital transmission applications such as the integrated services digital network (ISDN), cordless telephones and cell phones.^[19]

Technologies

Digital audio is used in broadcasting of audio. Standard technologies include Digital audio broadcasting (DAB), Digital Radio Mondiale (DRM), HD Radio and In-band on-channel (IBOC).

Digital audio in recording applications is stored on audio-specific technologies including CD, DAT, Digital Compact Cassette (DCC) and MiniDisc. Digital audio may be stored in a standard audio file formats and stored on a Hard disk recorder, Blu-ray or DVD-Audio. Files may be played back on smartphones, computers or MP3 player. Digital audio resolution is measured in audio bit depth. Most digital audio formats use either 16-bit, 24-bit, and 32-bit resolution.

See also

- Digital audio editor
- Digital synthesizer
- Frequency modulation synthesis
- Sound chip
- Sound card
- Audio Interface
- Quantization
- Sampling
- Multitrack recording
- Digital audio workstation

Notes

- a. Anti-alias filtering and optional digital signal processing may degrade the audio signal via passband ripple, non-linear phase shift, numeric precision quantization noise or time distortion of transients. However, these potential degradations can be limited by careful digital design.^[3]
- b. Some audio signals such as those created by digital synthesis originate entirely in the digital domain, in which case analog to digital conversion does not take place.

References

1. Janssens, Jelle; Stijn Vandaele; Tom Vander Beken (2009). "The Music Industry on (the) Line? Surviving Music Piracy in a Digital Era" (<https://biblio.ugent.be/publication/608677>). *European Journal of Crime, Criminal Law and Criminal Justice*. **77** (96): 77–96. doi:[10.1163/157181709X429105](https://doi.org/10.1163/157181709X429105) (<https://doi.org/10.1163%2F157181709X429105>). hdl:[1854/LU-608677](https://hdl.handle.net/1854/LU-608677) (<https://hdl.handle.net/1854%2FLU-608677>).
2. Liikanen, Lassi A.; Åman, Pirkka (May 2016). "Shuffling Services: Current Trends in Interacting with Digital Music" (<https://academic.oup.com/iwc/article-lookup/doi/10.1093/iwc/iwv004>). *Interacting with Computers*. **28** (3): 352–371. doi:[10.1093/iwc/iwv004](https://doi.org/10.1093/iwc/iwv004) (<https://doi.org/10.1093%2Fiwc%2Fiwv004>). ISSN 0953-5438 (<https://search.worldcat.org/issn/0953-5438>).
3. Story, Mike (September 1997). "A Suggested Explanation For (Some Of) The Audible Differences Between High Sample Rate And Conventional Sample Rate Audio Material" (<http://sdg-master.com:80/lesestoff/aes97ny.pdf>) (PDF). dCS Ltd. Archived (<https://web.archive.org/web/20091128021651/http://sdg-master.com:80/lesestoff/aes97ny.pdf>) (PDF) from the original on 28 November 2009.
4. Seok, Jongwon; Hong, Jinwoo; Kim, Jinwoong (2002-06-01). "A Novel Audio Watermarking Algorithm for Copyright Protection of Digital Audio" (<https://doi.org/10.4218%2Fetrij.02.0102.0301>). *ETRI Journal*. **24** (3): 181–189. doi:[10.4218%2Fetrij.02.0102.0301](https://doi.org/10.4218%2Fetrij.02.0102.0301) (<https://doi.org/10.4218%2Fetrij.02.0102.0301>). ISSN 1225-6463 (<https://search.worldcat.org/issn/1225-6463>). S2CID [3008374](https://api.semanticscholar.org/CorpusID:3008374) (<https://api.semanticscholar.org/CorpusID:3008374>).
5. *Genius Unrecognised* (<https://www.bbc.co.uk/programmes/b00zs7v5>), BBC, 2011-03-27, retrieved 2011-03-30
6. US patent 2605361 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=US2605361>), C. Chapin Cutler, "Differential Quantization of Communication Signals", issued 1952-07-29
7. P. Cummiskey, Nikil S. Jayant, and J. L. Flanagan, "Adaptive quantization in differential PCM coding of speech", *Bell Syst. Tech. J.*, vol. 52, pp. 1105—1118, Sept. 1973
8. Cummiskey, P.; Jayant, Nikil S.; Flanagan, J. L. (1973). "Adaptive quantization in differential PCM coding of speech". *The Bell System Technical Journal*. **52** (7): 1105–1118. doi:[10.1002/j.1538-7305.1973.tb02007.x](https://doi.org/10.1002/j.1538-7305.1973.tb02007.x) (<https://doi.org/10.1002%2Fj.1538-7305.1973.tb02007.x>). ISSN 0005-8580 (<https://search.worldcat.org/issn/0005-8580>).
9. Schroeder, Manfred R. (2014). "Bell Laboratories" (<https://books.google.com/books?id=d9IkBAAAQBAJ&pg=PA388>). *Acoustics, Information, and Communication: Memorial Volume in Honor of Manfred R. Schroeder*. Springer. p. 388. ISBN 9783319056609.
10. Gray, Robert M. (2010). "A History of Realtime Digital Speech on Packet Networks: Part II of Linear Predictive Coding and the Internet Protocol" (<https://ee.stanford.edu/~gray/lpcip.pdf>) (PDF). *Found. Trends Signal Process.* **3** (4): 203–303. doi:[10.1561/2000000036](https://doi.org/10.1561/2000000036) (<https://doi.org/10.1561%2F2000000036>). ISSN 1932-8346 (<https://search.worldcat.org/issn/1932-8346>).
11. Ahmed, Nasir (January 1991). "How I Came Up With the Discrete Cosine Transform" (<https://www.scribd.com/doc/52879771/DCT-History-How-I-Came-Up-with-the-Discrete-Cosine-Transform>). *Digital Signal Processing*. **1** (1): 4–5. Bibcode:1991DSP.....1....4A (<https://ui.adsabs.harvard.edu/abs/1991DSP.....1....4A>). doi:[10.1016/1051-2004\(91\)90086-Z](https://doi.org/10.1016/1051-2004(91)90086-Z) (<https://doi.org/10.1016%2F1051-2004%2891%2990086-Z>).
12. Nasir Ahmed; T. Natarajan; Kamisetty Ramamohan Rao (January 1974). "Discrete Cosine Transform" (https://www.ic.tu-berlin.de/fileadmin/fg121/Source-Coding_WS12/selected-readings/Ahmed_et_al._1974.pdf) (PDF). *IEEE Transactions on Computers*. **C-23** (1): 90–93. doi:[10.1109/T-C.1974.223784](https://doi.org/10.1109/T-C.1974.223784) (<https://doi.org/10.1109%2FT-C.1974.223784>). S2CID [149806273](https://api.semanticscholar.org/CorpusID:149806273) (<https://api.semanticscholar.org/CorpusID:149806273>).

13. J. P. Princen, A. W. Johnson und A. B. Bradley: *Subband/transform coding using filter bank designs based on time domain aliasing cancellation*, IEEE Proc. Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2161–2164, 1987.
14. Luo, Fa-Long (2008). *Mobile Multimedia Broadcasting Standards: Technology and Practice* (<https://books.google.com/books?id=l6PovWat8SMC&pg=PA590>). Springer Science & Business Media. p. 590. ISBN 9780387782638.
15. Guckert, John (Spring 2012). "The Use of FFT and MDCT in MP3 Audio Compression" (<http://www.math.utah.edu/~gustafso/s2012/2270/web-projects/Guckert-audio-compression-svd-mdct-MP3.pdf>) (PDF). *University of Utah*. Retrieved 14 July 2019.
16. Fine, Thomas (2008). Barry R. Ashpole (ed.). "The Dawn of Commercial Digital Recording" (http://www.aes.org/aeshc/pdf/fine_dawn-of-digital.pdf) (PDF). *ARSC Journal*. Retrieved 2010-05-02.
17. Reuter, Anders (2022-03-15). "Who let the DAWs Out? The Digital in a New Generation of the Digital Audio Workstation" (<https://www.tandfonline.com/doi/full/10.1080/03007766.2021.1972701>). *Popular Music and Society*. 45 (2): 113–128. doi:10.1080/03007766.2021.1972701 (<https://doi.org/10.1080%2F03007766.2021.1972701>). ISSN 0300-7766 (<https://search.worldcat.org/issn/0300-7766>). S2CID 242779244 (<https://api.semanticscholar.org/CorpusID:242779244>).
18. Allstot, David J. (2016). "Switched Capacitor Filters" (https://web.archive.org/web/20210930151716/https://ieee-cas.org/sites/default/files/a_short_history_of_circuits_and_systems_ebook_web.pdf) (PDF). In Maloberti, Franco; Davies, Anthony C. (eds.). *A Short History of Circuits and Systems: From Green, Mobile, Pervasive Networking to Big Data Computing*. IEEE Circuits and Systems Society. pp. 105–110. ISBN 9788793609860. Archived from the original (https://ieee-cas.org/sites/default/files/a_short_history_of_circuits_and_systems_ebook_web.pdf) (PDF) on 2021-09-30. Retrieved 2019-11-29.
19. Floyd, Michael D.; Hillman, Garth D. (8 October 2018) [1st pub. 2000]. "Pulse-Code Modulation Codec-Filters" (<https://books.google.com/books?id=Tokk5bZxB0MC&pg=SA26-PA1>). *The Communications Handbook* (2nd ed.). CRC Press. pp. 26–1, 26–2, 26–3. ISBN 9781420041163.

Further reading

- Borwick, John, ed., 1994: *Sound Recording Practice* (Oxford: Oxford University Press)
- Bosi, Marina, and Goldberg, Richard E., 2003: *Introduction to Digital Audio Coding and Standards* (Springer)
- Ifeachor, Emmanuel C., and Jervis, Barrie W., 2002: *Digital Signal Processing: A Practical Approach* (Harlow, England: Pearson Education Limited)
- Rabiner, Lawrence R., and Gold, Bernard, 1975: *Theory and Application of Digital Signal Processing* (Englewood Cliffs, New Jersey: Prentice-Hall, Inc.)
- Watkinson, John, 1994: *The Art of Digital Audio* (Oxford: Focal Press)

External links

- Monty Montgomery (2012-10-24). "Guest Opinion: Why 24/192 Music Downloads Make No Sense" (<https://web.archive.org/web/20121210132914/http://evolver.fm/2012/10/04/guest-opinion-why-24192-music-downloads-make-no-sense/>). evolver.fm. Archived from the original (<http://evolver.fm/2012/10/04/guest-opinion-why-24192-music-downloads-make-no-sense/>) on 2012-12-10. Retrieved 2012-12-07.
- J. ROBERT STUART. "Coding High Quality Digital Audio" (<https://web.archive.org/web/20070627075502/http://www.meridian.co.uk/ara/coding2.pdf>) (PDF). Archived from the original (<http://www.meridian.co.uk/ara/coding2.pdf>) on 2007-06-27.

<http://www.meridian.co.uk/ara/coding2.pdf> (PDF) on 2007-06-27. Retrieved 2012-12-07.

- Dan Lavry. "Sampling Theory For Digital Audio" (<http://lavryengineering.com/pdfs/lavry-sampling-theory.pdf>) (PDF). Archived (<https://web.archive.org/web/20120916040445/http://lavryengineering.com/pdfs/lavry-sampling-theory.pdf>) (PDF) from the original on 2012-09-16. Retrieved 2012-12-07.
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Digital_audio&oldid=1292026902"



Digital radio

Digital radio is the use of digital technology to transmit or receive across the radio spectrum. Digital transmission by radio waves includes digital broadcasting, and especially digital audio radio services. This should not be confused with Internet radio which also is digital but not transmitted by radio waves in the radio spectrum.

Types

In digital broadcasting systems, the analog audio signal is digitized, compressed using an audio coding format such as AAC+ (MDCT)^[1] or MP2, and transmitted using a digital modulation scheme. The aim is to increase the number of radio programs in a given spectrum, to improve the audio quality, to eliminate fading problems in mobile environments, to allow additional datacasting services, and to decrease the transmission power or the number of transmitters required to cover a region. However, analog radio (AM and FM) is still more popular and listening to radio over IP (Internet Protocol) is growing in popularity.

In 2012, four digital wireless radio systems are recognized by the International Telecommunication Union: the two European systems Digital Audio Broadcasting (DAB) and Digital Radio Mondiale (DRM), the Japanese ISDB-T and the in-band on-channel technique used in the US and Arab world and branded as HD Radio. - 5G Broadcast is a joint digital terrestrial broadcasting platform for radio and television under development in the UHF band being tested in 20 countries (2025) for more than five years.

An older definition, still used in communication engineering literature, is wireless digital transmission technologies, i.e. microwave and radio frequency communication standards where analog information signals as well as digital data are carried by a digital signal, by means of a digital modulation method. This definition includes broadcasting systems such as digital TV and digital radio broadcasting, but also two-way digital radio standards such as the second generation (2G) cell-phones and later, short-range communication such as digital cordless phones, wireless computer networks, digital micro-wave radio links, deep space communication systems such as communications to and from the two Voyager space probes, etc.



Example of a digital radio

One-way (broadcasting) systems

Broadcast standards

Digital audio radio service standards may provide terrestrial or satellite radio service. Digital radio broadcasting systems are typically designed for handheld mobile devices, like mobile-TV systems and unlike other digital TV systems which typically require a fixed directional antenna. Some digital radio systems provide in-band on-channel (IBOC) solutions that may coexist with or simulcast with analog AM or FM transmissions, while others are designed for designated radio frequency bands. The latter allows one wideband radio signal to carry a multiplex of several radio-channels of various bitrates as well as data services and other forms of media. Some digital broadcasting systems allow single-frequency network (SFN), where all terrestrial transmitters in a region sending the same multiplex of radio programs may use the same frequency channel without self-interference problems, further improving the system spectral efficiency.

While digital broadcasting offers many potential benefits, its introduction has been hindered by a lack of global agreement on standards and many disadvantages. The DAB Eureka 147 standard for digital radio is coordinated by the World DMB Forum. This standard of digital radio technology was defined in the late 1980s, and is now being introduced in some European countries. Commercial DAB receivers began to be sold in 1998 and, by 2006, 500 million people were in the coverage area of DAB broadcasts, although by this time sales had only taken off in the UK and Denmark. In 2006 there are approximately 1,000 DAB stations in operation.^[2] There have been criticisms of the Eureka 147 standard and so a new 'DAB+' standard has been introduced.

The DRM standard has been used for several years to broadcast digitally on frequencies below 30 MHz (shortwave, mediumwave and longwave). Also there is now the extended standard DRM+, which is designed for VHF bands.^[3] Tests of DRM+ have been made in countries such as in Brazil, Germany, France, India, Sri Lanka, the UK, Slovakia, Italy (incl. the Vatican), as well as Sweden.^[4]

The two pan-European associations for community radio *Community Media Forum Europe (CMFE)* and *AMARC Europe* have stated to the European Commission that FM radio should be retained for local radio. In case of a future digital transition DRM+ would be the choice. ^[5] ^[6]

To date the following standards have been defined for one-way digital radio:

Digital audio broadcasting systems

- Eureka 147 (branded as DAB)
- DAB+
- 5G Broadcast
- ISDB-TSB
- Internet radio (Technically not a true broadcast system)
- T-DMB V-Radio
- **AM band in-band on-channel (AM IBOC):**
 - HD Radio (AM IBOC sideband)

- Digital Radio Mondiale (branded as DRM) for the short-, medium- and long-wave bands
- **FM band in-band on-channel (FM IBOC):**
 - HD Radio (OFDM modulation over AM and FM band IBOC sidebands)
 - FMeXtra (FM band IBOC subcarriers)
 - Digital Radio Mondiale extension (DRM+) (OFDM modulation over AM band IBOC sidebands)
 - Convergent Digital Radio (CDR) (OFDM modulation over FM band IBOC sidebands)
- **Satellite radio:**
 - WorldSpace in Asia and Africa
 - Sirius XM Radio in North America
 - MobaHo! in Japan and the Republic of (South) Korea
- **Systems also designed for digital TV:**
 - DMB
 - DVB-H
 - ISDB-T
 - DTMB
- **Low-bandwidth digital data broadcasting over existing FM radio:**
 - Radio Data System (branded as RDS)
- **Radio pagers:**
 - FLEX
 - ReFLEX
 - POCSAG
 - NTT

Digital television (DTV) broadcasting systems

- Digital Video Broadcasting (DVB)
- 5G Broadcast
- Integrated Services Digital Broadcasting (ISDB)
- Digital Multimedia Broadcasting (DMB)
- **Digital Terrestrial Television (DTTV or DTT) to fixed mainly roof-top antennas:**
 - DVB-T (based on OFDM modulation)
 - ISDB-T (based on OFDM modulation)
 - ATSC (based on 8VSB modulation)
 - T-DMB (based on OFDM modulation)
 - Digital Terrestrial Multimedia Broadcast (DTMB) (based on OFDM modulation)
- **Mobile TV reception in handheld devices:**
 - DVB-H (based on OFDM modulation)
 - MediaFLO (based on OFDM modulation)
 - DMB (based on OFDM modulation)
 - Multimedia Broadcast Multicast Service (MBMS) via the GSM EDGE and UMTS cellular networks
 - DVB-SH (based on OFDM modulation)

- China Multimedia Mobile Broadcasting (CMMB) (based on OFDM modulation)
- **Satellite TV:**
 - DVB-S (for Satellite TV)
 - ISDB-S
 - 4DTV
 - S-DMB
 - MobaHo!
 - Advanced Broadcasting System-Satellite (ABS-S)

See also software radio for a discussion of radios which use digital signal processing.

Status by country

DAB adopters

Digital Audio Broadcasting (DAB), also known as Eureka 147, has been adopted by around 20 countries worldwide. It is based on the MPEG-1 Audio Layer II audio coding format and this has been co-ordinated by the WorldDMB.

WorldDMB announced in November 2006 that DAB would be adopting the HE-AACv2 audio coding format, also known as eAAC+. Also being adopted are the MPEG Surround format, and stronger error correction coding called Reed–Solomon coding.^[7] The update has been named DAB+. Receivers that support the new DAB standard began being released during 2007 with firmware updated available for some older receivers.

DAB and DAB+ cannot be used for mobile TV because they do not include any video codecs. DAB related standards Digital Multimedia Broadcasting (DMB) and DAB-IP are suitable for mobile radio and TV both because they have MPEG 4 AVC and WMV9 respectively as video coding formats. However a DMB video sub-channel can easily be added to any DAB transmission - as DMB was designed from the outset to be carried on a DAB subchannel. DMB broadcasts in Korea carry conventional MPEG 1 Layer II DAB audio services alongside their DMB video services.

Australia

Australia commenced regular digital audio broadcasting using the DAB+ standard on 4 May 2009,^[8] after many years of trialling alternative systems. Normal radio services operate on the AM and FM bands, as well as four stations (ABC and SBS) on digital TV channels. The services are currently operating in five state capital cities: Adelaide, Brisbane, Melbourne, Perth and Sydney, and is being trialled in Canberra and Darwin.^[9]

Canada

Canada has begun allowing experimental HD Radio broadcasts in December 2012 and digital audio subchannels on a case-by-case basis, with the first stations in the country being CFRM-FM in Little Current, CING-FM in Hamilton, and CJSA-FM in Toronto (with a fourth, CFMS-FM in the Toronto suburb of Markham applying to operate HD Radio technology), all within the province of Ontario.^{[10][11]}

Germany

In 2020, DAB+ signals cover more than 90% of Germany. A national multiplex contains three public stations by Deutschlandfunk and 12 commercial stations. In most areas, additional multiplexes with public broadcasters and regional commercial stations are available.

The first DAB station network was deployed in Bavaria since 17 October 1995 until full coverage in 1999. Other states had funded a station network but the lack of success led them to scrap the funding - the MDR switched off in 1998 already and Brandenburg declared a failure in 2004. Instead Berlin/Brandenburg began to switch to digital radio based on an audio-only DVB-T mode given the success of the DVB-T standard in the region when earlier analogue television was switched off in August 2003 (being the first region to switch in Germany). During that time the DVB-H variant of the DVB family was released for transmission to mobile receivers in 2004. During 2005 most radio stations left the DAB network with only one public service broadcaster ensemble to remain in the now fully state-funded station network. At last the KEF (*commission to determine the financial needs of broadcasters*) blocked federal funding on 15. July 2009 until economic viability of DAB broadcasting would be proven - and pointing to DVB-T as a viable alternative.

Digital radio deployment was rebooted during 2011 - a joint commission of public and private radio broadcasters decided upon "DAB+" as the new national standard in December 2010. The new station network started as planned on 1. August 2011 with 27 stations with 10 kW each giving a coverage of 70% across the nation. A single "Bundesmux" ("fed-mux": short for "federal multiplex") was created on band 5C as a single-frequency network on channel 5C (see [1] (http://digitalradio.de/index.php/de/empfangne_u)). With the initial market success of DAB+ the contractors decided on an expansion of the digital radio station network in November 2012.

Korea

On 1 December 2005 South Korea launched its T-DMB service which includes both television and radio stations. T-DMB is a derivative of DAB with specifications published by ETSI. More than 110,000 receivers had been sold in one month only in 2005.

Hong Kong

Hong Kong replaced DAB with DVB-T2 Lite.

Norway

Norway was the first country where analog FM radio was switched off in 2017 being replaced by nationwide DAB+ distribution.

Local stations can continue broadcasting in FM at least until 2031.

Other European Countries

With DAB being available across Belgium, Netherlands, Switzerland, Denmark, Norway and Northern Italy there is good coverage across the European Backbone area (see countries using DAB/DMB) indicating a sufficient momentum on the market. France, Spain, Sweden and Poland use DAB+ only in the big cities.

Portugal and Finland abandoned DAB. Finland is requesting the EU to mandate that automakers support FM similarly to DAB.

Japan

Japan has started terrestrial sound broadcasting using ISDB-Tsb and MobaHO! 2.6 GHz Satellite Sound digital broadcasting

United Kingdom

In the United Kingdom, 44.3% of the population now has a DAB digital radio set and 34.4% of listening is to different digital platforms. Because of the early success of the old DAB standard, the transition to the more efficient DAB+ takes more time. If DAB was switched off, older receivers would become worthless. In 2020, about half of the stations in the UK use DAB+.

26 million people, or 39.6% of the population of 65.64 million, now tune into digital radio each week, up 2.6 million year on year, according to RAJAR in Q1 2013. But FM listening has increased to 61% and DAB decreased to 21% DAB listeners may also use AM & FM too.^[12]

The UK currently has the world's biggest digital radio network, with about 500 transmitters, two nationwide DAB ensembles and 48 local and regional DAB ensembles, broadcasting over 250 commercial and 34 BBC radio stations; about 100 stations can be received in London. On DAB digital radio most listeners can receive around 30 additional stations.

Digital radio stations are also distributed on digital television platforms such as Sky, Virgin Media and Freeview, as well as internet radio.

The Government will make a decision on a radio switchover subject to listening and coverage criteria being met. A digital radio switchover would maintain FM as a platform, while moving some services to DAB-only distribution.

DAB+ devices in the UK have been available to the public since 2010.

United States

The United States has opted for the proprietary HD Radio technology, a type of in-band on-channel (IBOC) technology. According to iBiquity, "HD Radio" is the company's trade name for its proprietary digital radio system, but the name does not imply either high definition or "hybrid digital" as it is commonly incorrectly referenced.

Transmissions use orthogonal frequency-division multiplexing, a technique which is also used for European terrestrial digital TV broadcast (DVB-T). HD Radio technology was developed and is licensed by iBiquity Digital Corporation. It is widely believed that a major reason for HD radio technology is to offer some limited digital radio services while preserving the relative "stick values" of the stations involved and to ensure that new programming services will be controlled by existing licensees.

The FM digital schemes in the U.S. provide audio at rates from 96 to 128 kilobits per second (kbit/s), with auxiliary "subcarrier" transmissions at up to 64 kbit/s. The AM digital schemes have data rates of about 48 kbit/s, with auxiliary services provided at a much lower data rate. Both the FM and AM schemes

use lossy compression techniques to make the best use of the limited bandwidth.

Lucent Digital Radio, USA Digital Radio (USADR), and Digital Radio Express commenced tests in 1999 of their various schemes for digital broadcast, with the expectation that they would report their results to the National Radio Systems Committee (NRSC) in December 1999.^[13] Results of these tests remain unclear, which in general describes the status of the terrestrial digital radio broadcasting effort in North America.

While traditional terrestrial radio broadcasters are trying to "go digital", most major US automobile manufacturers are promoting digital satellite radio. HD Radio technology has also made inroads in the automotive sector with factory-installed options announced by BMW, Ford, Hyundai, Jaguar, Lincoln, Mercedes, MINI, and Volvo.^[14]

Satellite radio is distinguished by its freedom from FCC censorship in the United States, its relative lack of advertising, and its ability to allow people on the road to listen to the same stations at any location in the country. Listeners must currently pay an annual or monthly subscription fee in order to access the service.

Sirius Satellite Radio launched a constellation of three Sirius satellites during the course of 2000. The satellites were built by Space Systems/Loral and were launched by Russian Proton boosters. As with XM Satellite Radio, Sirius implemented a series of terrestrial ground repeaters where satellite signal would otherwise be blocked by large structures including natural structures and high-rise buildings.

XM Satellite Radio has a constellation of three satellites, two of which were launched in the spring of 2001, with one following later in 2005. The satellites are Boeing 702 comsats, and were put into orbit by Sea Launch boosters. Back-up ground transmitters (repeaters) will be built in cities where satellite signals could be blocked by big buildings.

On February 19, 2007, Sirius Satellite Radio and XM Satellite Radio merged, to form Sirius XM Radio.

The FCC has auctioned bandwidth allocations for satellite broadcast in the S band range, around 2.3 GHz.

Terrestrial broadcasting has advantages in being free and local. Satellite radio is neither of these things; however, in the early 21st century it has grown by providing uncensored content (most notably, the crossover of Howard Stern from terrestrial radio to satellite radio) and commercial-free, all-digital music channels that offer similar radio formats to local stations.

The "HD Radio" signal of an FM broadcast station in the US has a limited listening distance from the broadcast tower site. FCC regulations currently limit the power of the digital part of the station's transmission to 10% of the existing analog power permitted the station. Even at this power level, the presence of the digital signal right next to the station's analog signal can result in older radios picking up noise due to trouble rejecting the adjacent digital signal. "There are still some concerns that HD Radio on FM will increase interference between different stations even though HD Radio at the 10% power level fits within the FCC spectral mask." HD Radio HD Radio#cite note-14.

"HD Radio" allows each existing broadcast station to add additional "channels" in the US by transmitting a digital signal on both sides of its channel, just beyond their existing analog Frequency Modulation signal. The HD Radio signal occupies the 0.1 MHz that begins 0.1 MHz above and below the carrier

frequency station. For instance, if a station's analog signal's carrier frequency is 93.3 MHz, the digital signal will fill 93.1–93.2 MHz and 93.4–93.5 MHz within the FM Broadcast Band. Several digital audio streams, or "subchannels", can be carried within this single digital data stream, with the number of audio of subchannels and bandwidth allocations at the choice of the station. On the radio tuner, these will appear as (in the above case) "93.3-2", "93.3-3", and so on. The frequencies that are used do not change as more channels are added to the one radio station (93.3 MHz in the example above). Instead, a fixed total amount of bandwidth is simply reallocated across the audio streams such that each now receives less bandwidth, and therefore lower audio quality, than before.

There is no federally mandated transition to HD Radio for both FM and AM stations. However, on October 27, 2020, the FCC approved voluntary all-digital AM operation nationwide.^[15]

Developing nations

Digital radio is now being provided to the developing world. A now defunct satellite communications company named [WorldSpace](#) was setting up a network of three satellites, including "AfriStar", "AsiaStar", and "AmeriStar", to provide digital audio information services to [Africa](#), [Asia](#), and [Latin America](#). AfriStar and AsiaStar are in orbit. AmeriStar could not be launched from the [United States](#) as Worldspace transmitted on the L-band and would interfere with USA military as mentioned above.. in its heyday provided service to over 170,000 subscribers in eastern and southern Africa, the Middle East, and much of Asia with 96% coming from India. [Timbre Media](#) along with Saregama India plan to relaunch the company. As of 2013 Worldspace is defunct, but two satellites are in orbit which still have a few channels. See main [WorldSpace](#) article.

Each satellite provides three transmission beams that can support 50 channels each, carrying news, music, entertainment, and education, and including a computer multimedia service. Local, regional, and international broadcasters were working with WorldStar to provide services.

A consortium of broadcasters and equipment manufacturers are also working to bring the benefits of digital broadcasting to the [radio spectrum](#) currently used for terrestrial [AM radio](#) broadcasts, including international [shortwave](#) transmissions. Over seventy broadcasters are now transmitting programs using the new standard, known as [Digital Radio Mondiale](#) (DRM), and [/ commercial DRM receivers \(<http://www.drm.org/index.php?p=products>\)](#) are available (though there are few models on the DRM website and some are discontinued). DRM's system uses the [MPEG-4](#) based standard [aacPlus](#) to code the music and [CELP](#) or [HVXC](#) for speech programs. At present these are priced too high to be affordable by many in the third world, however. Take-up of DRM has been minuscule and many traditional Shortwave broadcasters now only stream on Internet, use fixed satellite (TV set-boxes) or Local Analogue FM relays to save on costs. Very few (expensive) DRM radio sets are available and some Broadcasters (RTE in Ireland on 252 kHz) have ceased trials without launching a service.

Low-cost DAB radio receivers are now available from various Japanese manufacturers, and WorldSpace has worked with Thomson Broadcast to introduce a village communications center known as a [Telekiosk](#) to bring communications services to rural areas. The Telekiosks are self-contained and are available as fixed or mobile units

Two-way digital radio standards

The key breakthrough or key feature in digital radio transmission systems is that they allow lower transmission power, they can provide robustness to noise and cross-talk and other forms of interference, and thus allow the same radio frequency to be reused at shorter distance. Consequently, the spectral efficiency (the number of phonecalls per MHz and base station, or the number of bit/s per Hz and transmitter, etc.) may be sufficiently increased. Digital radio transmission can also carry any kind of information whatsoever — just as long as it has been expressed digitally. Earlier radio communication systems had to be made expressly for a given form of communications: telephone, telegraph, or television, for example. All kinds of digital communications can be multiplexed or encrypted at will.

- **Digital cellular telephony** (2G systems and later generations):

- GSM
- UMTS (sometimes called W-CDMA)
- TETRA
- IS-95 (cdmaOne)
- IS-136 (D-AMPS, sometimes called TDMA)
- IS-2000 (CDMA2000)
- iDEN

- **Digital Mobile Radio:**

- Multi-Band Excitation
- TETRA
- TETRAPOL
- NXDN/dPMR
- DMR
- D-STAR

- **Wireless networking:**

- Wi-Fi
- HIPERLAN
- Bluetooth
- DASH7
- Zigbee
- 6LoWPAN

- **Military radio systems for Network-centric warfare**

- JTRS (Joint Tactical Radio System- a flexible software-defined radio)
- SINCGARS (Single channel ground to air radio system)

- **Amateur packet radio:**

- AX.25

- **Digital modems for HF:**

- PACTOR

- **Satellite radio:**

- [Satmodems](#)
- [**Wireless local loop:**](#)
 - [Basic Exchange Telephone Radio Service](#)
- [**Broadband wireless access:**](#)
 - [IEEE 802.16](#)

See also



[Radio portal](#)

- [Satellite television](#)

References

1. Britanak, Vladimir; Rao, K. R. (2017). *Cosine-/Sine-Modulated Filter Banks: General Properties, Fast Algorithms and Integer Approximations* (<https://books.google.com/books?id=cZ4vDwAAQBAJ&pg=PA478>). Springer. p. 478. ISBN 978-3-319-61080-1. Archived (<http://web.archive.org/web/20230701082239/https://books.google.com/books?id=cZ4vDwAAQBAJ&pg=PA478>) from the original on 2023-07-01. Retrieved 2019-10-24.
2. "Digital Broadcast - bringing the future to you" (<https://web.archive.org/web/20071017232328/http://worlddab.org/benefits.php>). Archived from the original (<http://www.worlddab.org/benefits.php>) on 2007-10-17.
3. "Digital Radio Mondiale - Technical Info" (<http://www.drm.org/technical-info/>). www.drm.org. Archived (<https://web.archive.org/web/20171023020536/http://www.drm.org/technical-info/>) from the original on 23 October 2017. Retrieved 29 April 2018.
4. "Digital Radio Mondiale - DRM+ Testing gets underway today in Stockholm, Sweden" (<http://www.drm.org/drm-testing-gets-underway-today-in-stockholm-sweden/>). www.drm.org. Archived (<https://web.archive.org/web/20180202190244/http://www.drm.org/drm-testing-gets-underway-today-in-stockholm-sweden/>) from the original on 2 February 2018. Retrieved 29 April 2018.
5. "Community Radio Makes Its Stand for FM Radio in Brussels" (<https://digitalradioinsider.blogspot.com/2013/10/community-radio-makes-its-stand-for-fm.html>). Retrieved 8 April 2025.
6. "Community Radio: FM and Digitalization. Future Distribution Platforms for Community Radio in Europe (Green paper for CMFE (2014))" (<http://www.public-service.net/docu/digiradio2014open.pdf>) (PDF). *Public servicerådet*. Retrieved 8 April 2025.
7. "New High Efficiency Audio Option Added for DAB Digital Radio" (https://web.archive.org/web/20080308204329/http://www.worlddab.org/upload/uploaddocs/WorldDMBPress%20Release_November.pdf) (PDF). Archived from the original (http://www.worlddab.org/upload/uploaddocs/WorldDMBPress%20Release_November.pdf) (PDF) on 2008-03-08. Retrieved 2016-02-06.
8. "Digital Radio Launch" (<https://www.radioinfo.com.au/news/8-million-people-hear-about-digital-radio-industry-unites-launch>). *Radioinfo*. August 2009. Archived (<https://web.archive.org/web/20190528080917/https://www.radioinfo.com.au/news/8-million-people-hear-about-digital-radio-industry-unites-launch>) from the original on 2019-05-28. Retrieved 2019-05-28.
9. Digital Radio Plus (<http://www.digitalradioplus.com.au/index.cfm>) Archived (<https://web.archive.org/web/20110729002542/http://www.digitalradioplus.com.au/index.cfm>) 2011-07-29 at the Wayback Machine (accessed 26 July 2011)

10. "Digital radio policy - In this public notice, the Commission sets out its revised policy for digital radio broadcasting" (<http://www.crtc.gc.ca/eng/archive/2006/pb2006-160.htm>). *Canadian Radio-television and Telecommunications Commission*. 15 December 2006. Archived (<https://web.archive.org/web/20140918234506/http://www.crtc.gc.ca/eng/archive/2006/pb2006-160.htm>) from the original on 2014-09-18.
11. "CRTC looking at bringing HD Radio to Canada" (<http://blog.fagstein.com/2014/01/16/crtc-hd-radio/>). *Fagstein*. 16 January 2014. Archived (<http://archive.wikiwix.com/cache/20140429010411/http://blog.fagstein.com/2014/01/16/crtc-hd-radio/>) from the original on 2014-04-29.
12. "Analogue radio will CONTINUE in Blighty as Minister of Fun dodges D-Day death sentence" (https://www.theregister.co.uk/2013/12/16/minister_kicks_dab_can_miles_down_the_road/). *The Register*. Archived (https://web.archive.org/web/20161110043935/http://www.theregister.co.uk/2013/12/16/minister_kicks_dab_can_miles_down_the_road/) from the original on 2016-11-10.
13. Behrens, Steve. "Field testing resumes for radio's digital best hope." *Current*, Aug. 16, 1999. Available at "Current.org | Testing digital radio systems, 1999" (<https://web.archive.org/web/20090720071416/http://www.current.org/tech/tech915r.html>). Archived from the original (<http://www.current.org/tech/tech915r.html>) on 2009-07-20. Retrieved 2009-06-15.
14. "iBiquity Digital Corporation - Automotive" (<http://www.ibiquity.com/automotive>). Archived (<https://web.archive.org/web/20081117191606/http://www.ibiquity.com./automotive>) from the original on 2008-11-17.
15. "FCC AUTHORIZES ALL-DIGITAL AM RADIO" (<https://docs.fcc.gov/public/attachments/DOC-367780A1.pdf>) (PDF). *fcc.gov*. Federal Communications Commission. October 27, 2020. Archived (<https://web.archive.org/web/20201029192036/https://docs.fcc.gov/public/attachments/DOC-367780A1.pdf>) (PDF) from the original on 2020-10-29. Retrieved 2020-10-27.

External links

- "Digital, DTV, Internet, Mobile phone and MP3 Listening" (http://www.rajar.co.uk/quarterlysummary/qstsvfiles/138/Q42006_Digital_DTV_Internet_Mobile.pdf) - December 2006, RAJAR organisation.
 - Online Terrestrial Radio (<https://web.archive.org/web/20101120063759/http://www.durocast.com/>) - Search & and Listen to Live Radio Digitally
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Digital_radio&oldid=1284624818"



Digital photography

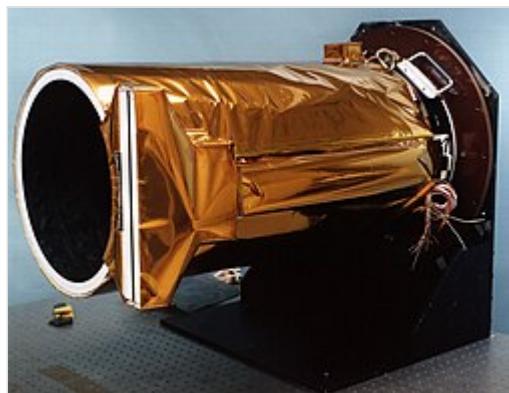
Digital photography uses cameras containing arrays of electronic photodetectors interfaced to an analog-to-digital converter (ADC) to produce images focused by a lens, as opposed to an exposure on photographic film. The digitized image is stored as a computer file ready for further digital processing, viewing, electronic publishing, or digital printing. It is a form of digital imaging based on gathering visible light (or for scientific instruments, light in various ranges of the electromagnetic spectrum).

Until the advent of such technology, photographs were made by exposing light-sensitive photographic film and paper, which was processed in liquid chemical solutions to develop and stabilize the image. Digital photographs are typically created solely by computer-based photoelectric and mechanical techniques, without wet bath chemical processing.

In consumer markets, apart from enthusiast digital single-lens reflex cameras (DSLR), most digital cameras now come with an electronic viewfinder, which approximates the final photograph in real-time. This enables the user to review, adjust, or delete a captured photograph within seconds, making this a form of instant photography, in contrast to most photochemical cameras from the preceding era.

Moreover, the onboard computational resources can usually perform aperture adjustment and focus adjustment (via inbuilt servomotors) as well as set the exposure level automatically, so these technical burdens are removed from the photographer unless the photographer feels competent to intercede (and the camera offers traditional controls). Electronic by nature, most digital cameras are instant, mechanized, and automatic in some or all functions. Digital cameras may choose to emulate traditional manual controls (rings, dials, sprung levers, and buttons) or it may instead provide a touchscreen interface for all functions; most camera phones fall into the latter category.

Digital photography spans a wide range of applications with a long history. Much of the technology originated in the space industry, where it pertains to highly customized, embedded



The Mars Orbiter Camera selected by NASA in 1986 (costing US\$44 million) contains a 32-bit radiation-hardened 10 MHz processor and 12 MB of DRAM, then considered state of the art.



Nikon D700 — a 12.1-megapixel full-frame DSLR



Canon PowerShot A95

systems combined with sophisticated remote telemetry. Any electronic image sensor can be digitized; this was achieved in 1951. The modern era in digital photography is dominated by the semiconductor industry, which evolved later. An early semiconductor milestone was the advent of the charge-coupled device (CCD) image sensor, first demonstrated in April 1970; since then, the field has advanced rapidly, with concurrent advances in photolithographic fabrication.

The first consumer digital cameras were marketed in the late 1990s.^[1] Professionals gravitated to digital slowly, converting as their professional work required using digital files to fulfill demands for faster turnaround than conventional methods could allow.^[2] Starting around 2000, digital cameras were incorporated into cell phones; in the following years, cell phone cameras became widespread, particularly due to their connectivity to social media and email. Since 2010, the digital point-and-shoot and DSLR cameras have also seen competition from the mirrorless digital cameras, which typically provide better image quality than point-and-shoot or cell phone cameras but are smaller in size and shape than typical DSLRs. Many mirrorless cameras accept interchangeable lenses and have advanced features through an electronic viewfinder, which replaces the through-the-lens viewfinder of single-lens reflex cameras.

History

While digital photography has only relatively recently become mainstream, the late 20th century saw many small developments leading to its creation. The history of digital photography began in the 1950s. In 1951, the first digital signals were saved to magnetic tape via the first video tape recorder.^[3] Six years later, in 1957, the first digital image was produced through a computer by Russell Kirsch. It was an image of his son.^[4]

The first semiconductor image sensor was the charge-coupled device (CCD), invented by physicists Willard S. Boyle and George E. Smith at Bell Labs in 1969.^[5] While researching the metal-oxide semiconductor (MOS) process, they realized that an electric charge was analogous to a magnetic bubble and that the charge could be stored on a tiny MOS capacitor. As it was fairly straightforward to fabricate a series of MOS capacitors in a row, they connected a suitable voltage to the capacitors so that the charge could be stepped along from one to the next.^[6] This semiconductor circuit was later used in the first digital video cameras for television broadcasting,^[7] and its invention was recognized by a Nobel Prize in Physics in 2009.^[8]

The first close-up image of Mars was taken as Mariner 4 flew by it on July 15, 1965, with a digital camera system designed by NASA and JPL. In 1976, the twin Mars Viking Landers produced the first images from the surface of Mars. The imaging process was different from that of a modern digital camera, though the result was similar; Viking used a



First digital image ever created, by Russell Kirsch. It is an image of his son, Walden.

mechanically scanned facsimile camera rather than a mosaic of solid state sensor elements.^[9] This produced a digital image that was stored on tape for later, relatively slow transmission back to Earth.^{[10][11]}

The first published color digital photograph was produced in 1972 by Michael Francis Tompsett using CCD sensor technology and was featured on the cover of *Electronics Magazine*. It was a picture of his wife, Margaret Tompsett.^[12] The Cromemco Cyclops, a digital camera developed as a commercial product and interfaced to a microcomputer, was featured in the February 1975 issue of Popular Electronics magazine. It used MOS technology for its image sensor.

An important development in digital image compression technology was the discrete cosine transform (DCT), a lossy compression technique first proposed by Nasir Ahmed while he was working at the Kansas State University in 1972.^[13] DCT compression is used in the JPEG image standard, which was introduced by the Joint Photographic Experts Group in 1992.^[14] JPEG compresses images down to much smaller file sizes, and has become the most widely used image file format.^[15] The JPEG standard was largely responsible for popularizing digital photography.^[16]

The first self-contained (portable) digital camera was created in 1975 by Steven Sasson of Eastman Kodak.^{[17][18]} Sasson's camera used CCD image sensor chips developed by Fairchild Semiconductor in 1973.^[19] The camera weighed 8 pounds (3.6 kg), recorded black-and-white images to a cassette tape, had a resolution of 0.01 megapixels (10,000 pixels), and took 23 seconds to capture its first image in December 1975. The prototype camera was a technical exercise, not intended for production.^[20] While it was not until 1981 that the first consumer camera was produced by Sony, the groundwork for digital imaging and photography had been laid.^[21]

The first digital single-lens reflex (DSLR) camera was the Nikon SVC prototype demonstrated in 1986, followed by the commercial Nikon QV-1000C released in 1988.^[22] The first widely commercially available digital camera was the 1990 Dycam Model 1; it also sold as the Logitech Fotoman. It used a CCD image sensor, stored pictures digitally, and connected directly to a computer for downloading images.^{[23][24][25]} Originally offered to professional photographers for a hefty price, by the mid-to-late 1990s, due to technology advancements, digital cameras were commonly available to the general public.

The advent of digital photography also gave way to cultural changes in the field of photography. Unlike film photography, dark rooms and hazardous chemicals were no longer required for the post-production of an image – images could now be processed and enhanced from a personal computer. This allowed photographers to be more creative with their processing and editing techniques. As the field became more popular, digital photography and photographers diversified. Digital photography expanded the field of photography from a small, somewhat elite circle to one that encompassed many people.^[26]

The camera phone further helped popularize digital photography, along with the Internet, social media,^[27] and the JPEG image format.^[16] The first cell phones with built-in digital cameras were produced in 2000 by Sharp and Samsung.^[28] Small, convenient, and easy to use, camera phones have made digital photography ubiquitous in the daily life of the general public.

Digital camera

Sensors

Image sensors are arrays of electronic devices that convert the optical image created by the camera lens into a digital file that is stored in some digital memory device, inside or outside the camera. Each element of the image sensor array measures the intensity of light hitting a small area of the projected image (a pixel) and converts it to a digital value.

The two main types of sensors are charge-coupled devices (CCD)—in which the photo charge is shifted to a central charge-to-voltage converter—and CMOS or active pixel sensors.

Most cameras for the general consumer market create color images, in which each pixel has a color value from a three-dimensional color space like RGB. Although there is light-sensing technology that can distinguish the wavelength of the light incident on each pixel, most cameras use monochrome sensors that can only record the intensity of that light, over a broad range of wavelengths that includes all the visible spectrum. To obtain color images, those cameras depend on color filters applied over each pixel, typically in a Bayer pattern, or (rarely) on movable filters or light splitters such as dichroic mirrors. The resulting grayscale images are then combined to produce a color image. This step is usually performed by the camera itself, although some cameras may optionally provide the unprocessed grayscale images in a so-called raw image format.

However, some special-purpose cameras, such as those for thermal mapping, or low light viewing, or high speed capture, may record only monochrome (grayscale) images. The Leica Monochrom cameras, for example, opted for a grayscale-only sensor to get better resolution and dynamic range. The reduction from three-dimensional color to grayscale or simulated sepia toning may also be performed by digital post-processing, often as an option in the camera itself. On the other hand, some multippectral cameras may record more than three color coordinates for each pixel.



Monochromatic image from a night-vision device

Multifunctionality and connectivity

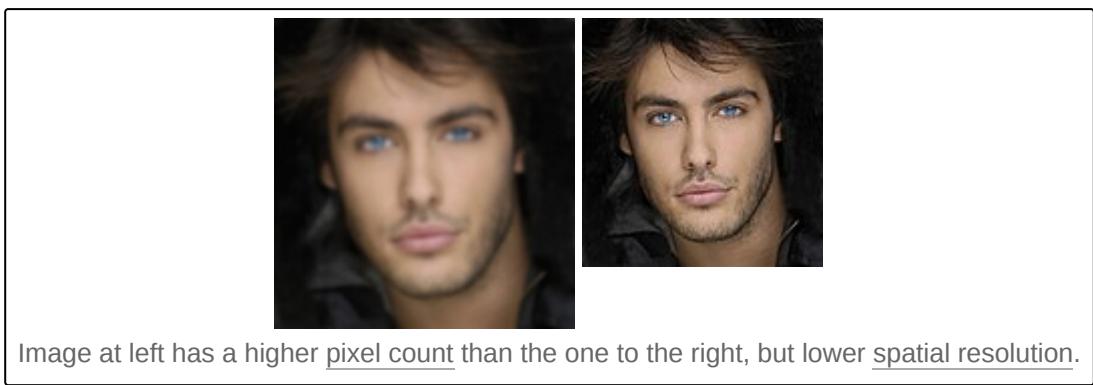
In most digital camera (except some high-end linear array cameras and simple, low-end webcams), a digital memory device is used for storing images, which may be transferred to a computer later. This memory device is usually a memory card; floppy disks and CD-RWs are less common.

In addition to taking pictures, digital cameras may also record sound and video. Some function as webcams, some use the PictBridge standard to connect to printers without using a computer, and some can display pictures directly on a television set. Similarly, many camcorders can take still photographs and store them on videotape or flash memory cards with the same functionality as digital cameras.

Digital photography is an example of the shift from analog information to digital information. In the past, conventional photography was an entirely chemical and mechanical process that did not require electricity. Now, modern photography is a digital process in which analog signals are converted to and stored as digital data using built-in computers.^[29]

Performance metrics

The quality of a digital image is a composite of various factors, many of which are similar to those of film cameras. Pixel count (typically listed in megapixels, millions of pixels) is only one of the major factors, though it is the most heavily marketed figure of merit. Digital camera manufacturers advertise this figure because consumers can use it to easily compare camera capabilities. It is not, however, the major factor in evaluating a digital camera for most applications. The processing system inside the camera that turns the raw data into a color-balanced and pleasing photograph is usually more critical, which is why some 4+ megapixel cameras perform better than higher-end cameras.



Resolution in pixels is not the only measure of image quality. A larger sensor with the same number of pixels generally produces a better image than a smaller one. One of the most important benefits of this is a reduction in image noise. This is one of the advantages of DSLR cameras, which have larger sensors than simpler point-and-shoot cameras of the same resolution.

Additional factors that impact the quality of a digital image include:

- **Lens quality:** resolution, distortion, dispersion (see Lens (optics))
- **Capture medium:** CMOS, CCD, negative film, reversal film
- **Capture format:** pixel count, digital file type (RAW, TIFF, JPEG), film format (135 film, 120 film), aspect ratio
- **Processing:** digital or chemical processing of "negative" and "print"

Pixel counts

The number of pixels n for a given maximum resolution (w horizontal pixels by h vertical pixels) is the product $n = w \times h$. For example, an image 1600×1200 in size has 1,920,000 pixels, or 1.92 megapixels.

The pixel count quoted by manufacturers can be misleading as it may not be the number of full-color pixels. For cameras using single-chip image sensors, the number claimed is the total number of single-color-sensitive photosensors, whether they have different locations in the plane, as with the Bayer sensor, or in stacks of three co-located photosensors as in the Foveon X3 sensor. However, the images have different numbers of RGB pixels: Bayer-sensor cameras produce as many RGB pixels as photosensors

via demosaicing (interpolation), while Foveon sensors produce uninterpolated image files with one-third as many RGB pixels as photosensors. Comparisons of megapixel ratings of these two types of sensors are sometimes a subject of dispute.^[30]

The relative increase in detail resulting from an increase in resolution is better compared by looking at the number of pixels across (or down) the picture, rather than the total number of pixels in the picture area. For example, a sensor of 2560×1600 sensor elements is described as "4 megapixels" ($2560 \times 1600 = 4,096,000$). Increasing to 3200×2048 increases the pixels in the picture to 6,553,600 (6.5 megapixels), a factor of 1.6, but the pixels per cm in the picture (at the same image size) increases by only 1.25 times. A measure of the comparative increase in linear resolution is the square root of the increase in area resolution (i.e., megapixels in the entire image).

Dynamic range

Both digital and film practical imaging systems have a limited "dynamic range": the range of luminosity that can be reproduced accurately. Highlights of the subject that are too bright are rendered as white, with no detail (overexposure); shadows that are too dark are rendered as black (underexposure). The loss of detail in the highlights is not abrupt with film, or in dark shadows with digital sensors. "Highlight burn-out" of digital sensors is not usually abrupt in output images due to the tone mapping required to fit their large dynamic range into the more limited dynamic range of the output (be it SDR display or printing). Because sensor elements for different colors saturate in turn, there can be hue or saturation shift in burnt-out highlights.

Some digital cameras can show these blown highlights in the image review, allowing the photographer to re-shoot the picture with a modified exposure. Others compensate for the total contrast of a scene by selectively exposing darker pixels longer. A third technique is used by Fujifilm in its FinePix S3 Pro DSLR: the image sensor contains additional photodiodes of lower sensitivity than the main ones; these retain detail in parts of the image too bright for the main sensor.

High-dynamic-range imaging (HDR) addresses this problem by increasing the dynamic range of images by either

- increasing the dynamic range of the image sensor, or
- using exposure bracketing and post-processing the separate images to create a single image with a higher dynamic range.

Storage

Many camera phones and most digital cameras use memory cards with flash memory to store image data. The majority of cards for separate cameras are Secure Digital (SD) format, or the older CompactFlash (CF) format; other formats are rare. XQD card format was the last new form of card, targeted at high-definition camcorders and high-resolution digital photo cameras. Most modern digital cameras also use internal memory of limited capacity to hold pictures temporarily, regardless of whether or not the camera is equipped with a memory card. These pictures can then be transferred later to a memory card or external device.

Memory cards can hold vast numbers of photos, requiring attention only when the memory card is full. For most users, this means hundreds of quality photos stored on the same memory card. Images may be transferred to other media for archival or personal use. Cards with high speed and capacity are suited to video and burst mode (capture several photographs in quick succession).

Because photographers rely on the integrity of image files, it is important to take proper care of memory cards. One process is card formatting, which essentially involves scanning the cards for possible errors. Common advocacy calls for formatting cards after transferring its images onto a computer. Since all cameras only do quick formatting of cards, it is advisable to occasionally carry out a more thorough formatting using appropriate software on a computer.

Comparison with film photography

Advantages already in consumer level cameras

The primary advantage of consumer-level digital cameras is the low recurring cost, as users need not purchase photographic film. Processing costs may be reduced or even eliminated. Dicams tend also to be easier to carry and use than comparable film cameras, and more easily adapt to modern use of pictures. Some, particularly those in smartphones, can send their pictures directly to email, web pages, or other electronic distribution.

Advantages of professional digital cameras

In professional usage, digital cameras offer many advantages in speed, precision, flexibility, ease, and cost.

- **Immediacy:** image review and deletion are possible immediately; lighting and composition can be assessed immediately, which ultimately conserves storage space.
- **Faster workflow:** management (color and file), manipulation, and printing tools are more versatile than conventional film processes. However, batch processing of RAW files can be time-consuming, even on a fast computer.
- **Faster image ingest:** it will take no more than a few seconds to transfer a high-resolution RAW file from a memory card vs many minutes to scan film with a high-quality scanner.
- **Flash:** using flash in images can provide a different look such as the lighting of the image.
- **Higher image quantity:** which enables longer photography sessions without changing film rolls. To most users, a single memory card is sufficient for the lifetime of the camera whereas film rolls are a re-incurring cost of film cameras.
- **Precision and reproducibility of processing:** since processing in the digital domain is purely numerical, image processing using deterministic (non-random) algorithms is perfectly reproducible and eliminates variations common with photochemical processing, and enables otherwise difficult or impractical processing techniques.
- **Digital manipulation:** a digital image can be modified and manipulated much easier and faster than with traditional negative and print methods.

Manufacturers such as Nikon and Canon have promoted the adoption of digital single-lens reflex cameras (DSLRs) by photojournalists. Images captured at 2+ megapixels are deemed of sufficient quality for small images in newspaper or magazine reproduction. 8- to 24-megapixel images, found in modern

digital SLRs, when combined with high-end lenses, can approximate the detail of film prints from 35 mm film-based SLRs.

Disadvantages of digital cameras

- **Aliasing:** as with any sampled signal, the combination of the periodic pixel structure of common electronic image sensors and periodic structure of photographed objects (typically human-made objects) can cause objectionable aliasing artifacts, such as false colors when using cameras using a Bayer pattern sensor. Aliasing is also present in film, but typically manifests itself in less obvious ways (such as increased granularity) due to the stochastic grain structure (stochastic sampling) of film.
- **Electricity-dependent:** digital cameras cannot operate without electricity, usually provided via a battery. In contrast, a large number of mechanical film cameras existed, such as the Leica M2. These battery-less devices had advantages over digital devices in harsh or remote conditions.
- **Limited sensor size:** a persistent challenge in semiconductor fabrication is that chips much larger than 1 cm^2 are expensive to produce without defects, confining large image sensor formats compatible with traditional 35 mm optics to professional and prosumer markets.

Equivalent features

Image noise and grain

Noise in a digital camera's image may sometimes be visually similar to film grain in a film camera.

Speed of use

Turn-of-the-century digital cameras had a long start-up delay compared to film cameras (that is, the delay from when they are turned on until they are ready to take the first shot), but this is no longer the case for modern digital cameras, which have start-up times under 1/4 seconds.

Frame rate

While some film cameras could reach up to 14 frames per second (fps), like the Canon F-1 with its rare high-speed motor drive, professional DSLR cameras can take still photographs at the highest frame rates. While the Sony SLT technology allows rates of up to 12 fps, the Canon EOS-1D X can take stills at a rate of 14 fps. The Nikon F5 is limited to 36 continuous frames (the length of the film) without the cumbersome bulk film back, while the digital Nikon D5 is able to capture over 100 14-bit RAW images before its buffer must be cleared and the remaining space on the storage media can be used.

Image longevity

Depending on the materials and how they are stored, analog photographic film and prints may fade as they age. Similarly, the media on which digital images are stored or printed can decay or become corrupt, leading to a loss of image integrity.

Color reproduction

Color reproduction (gamut) depends on the type and quality of film or sensor used and the quality of the optical system and film processing. Different films and sensors have different color sensitivity; the photographer needs to understand their equipment, the lighting conditions, and the media used to ensure accurate color reproduction. Many digital cameras offer RAW format (sensor data), which makes it possible to choose the color gamut in the development stage regardless of camera settings.

Even in RAW format, however, the sensor and the camera's dynamics can only capture colors within the gamut supported by the hardware. When that image is transferred for reproduction on any device, the widest achievable gamut is the gamut that the end device supports. For a monitor, it is the gamut of the display device. For a photographic print, it is the gamut of the device that prints the image on a specific type of paper.

Professional photographers often use specially designed and calibrated monitors that help them to reproduce color accurately and consistently.

Frame aspect ratios

Most digital point-and-shoot cameras have an aspect ratio of 1.33 (4:3), the same as analog television or early movies. However, a 35 mm picture's aspect ratio is 1.5 (3:2). Several digital cameras take photos in either ratio. Nearly all digital SLRs take pictures in a 3:2 ratio, as most can use lenses designed for 35 mm film. Some photo labs print photos on 4:3 ratio paper, as well as the existing 3:2.

In 2005, Panasonic launched the first consumer camera with a native aspect ratio of 16:9, matching HDTV. This is similar to a 7:4 aspect ratio, which was a common size for APS film.

Different aspect ratios are one of the reasons consumers have issues when cropping photos. An aspect ratio of 4:3 translates to a size of 4.5"×6.0". This loses half an inch when printing on the "standard" size of 4"×6", an aspect ratio of 3:2. Similar cropping occurs when printing on other sizes, such as 5"×7", 8"×10", or 11"×14".

Market impact

In late 2002, the cheapest digital cameras in the United States were available for around \$100 (USD).^[31] At the same time, many discount stores with photo labs introduced a "digital front end", allowing consumers to obtain true chemical prints (as opposed to ink-jet prints) in an hour. These prices were similar to those of prints made from film negatives.

In July 2003, digital cameras entered the disposable camera market with the release of the Ritz Dakota Digital, a 1.2-megapixel (1280 × 960) CMOS-based digital camera costing only \$11. Following the familiar single-use concept long in use with film cameras, Ritz intended the Dakota Digital for single use. When the pre-programmed 25-picture limit is reached, the camera is returned to the store, and the consumer receives back prints and a CD-ROM with their photos. The camera is then refurbished and resold.

Since the introduction of the Dakota Digital, a number of similar single-use digital cameras have appeared. Most single-use digital cameras are nearly identical to the original Dakota Digital in specifications and function, though a few include superior specifications and more advanced functions

(such as higher image resolutions and LCD screens). Most, if not all these single-use digital cameras cost less than \$20, not including processing. However, the huge demand for complex digital cameras at competitive prices has often caused manufacturing shortcuts, evidenced by a large increase in customer complaints over camera malfunctions, high parts prices, and short service life. Some digital cameras offer only a 90-day warranty.

Since 2003, digital cameras have outsold film cameras.^[32] Prices of 35 mm compact cameras have dropped with manufacturers further outsourcing to countries such as China. Kodak announced in January 2004 that they would no longer sell Kodak-branded film cameras in the developed world.^[33] In January 2006, Nikon followed suit and announced they would stop production of all but two models of their film cameras. They will continue to produce the low-end Nikon FM10, and the high-end Nikon F6. In the same month, Konica Minolta announced it was pulling out of the camera business altogether. The price of 35 mm and Advanced Photo System (APS) compact cameras have dropped, probably due to direct competition from digital cameras and the resulting availability of second-hand film cameras.^[34] Pentax have reduced but not halted production of film cameras.^[35] The technology has improved so rapidly that one of Kodak's film cameras was discontinued before it was awarded a "camera of the year" award later in the year.

The decline in film camera sales has also led to a decline in purchases of film for such cameras. In November 2004, a German division of Agfa-Gevaert, AgfaPhoto, split off. Within six months it filed for bankruptcy. Konica Minolta Photo Imaging, Inc., ended production of color film and paper worldwide by March 31, 2007. In addition, by 2005, Kodak employed less than a third of the employees it had twenty years earlier. It is not known if these job losses in the film industry have been offset in the digital image industry. Digital cameras have decimated the film photography industry through the declining use of the expensive film rolls and development chemicals previously required to develop the photos. This has had a dramatic effect on companies such as Fuji, Kodak, and Agfa. Many stores that formerly offered photofinishing services or sold film no longer do, or have seen a tremendous decline. In 2012, Kodak filed for bankruptcy after struggling to adapt to the changing industry.^[36]

Digital camera sales peaked in March 2012, averaging about 11 million units a month, but sales have declined significantly ever since. By March 2014, about 3 million were purchased each month, about 30 percent of the peak sales total. The decline may have bottomed out, with sales average hovering around 3 million a month. The main competitor is smartphones, most of which have built-in digital cameras and are routinely improved. Like most digital cameras, they also offer the ability to record videos.^[37] While smartphones continue to improve on a technical level, their form factor is not optimized for use as a camera, and their battery life is typically more limited compared to a digital camera.

Digital photography has resulted in some positive market impacts as well. The increasing popularity of products such as digital photo frames and canvas prints is a direct result of the increasing popularity of digital photography.



A man takes a photo with a smartphone, holding it somewhat awkwardly, as the form factor of a phone is not optimized for use as a camera.

Social impact

Digital photography has made photography available to a larger group of people. New technology and editing programs available to photographers have changed the way photographs are presented to the public. Photographs can be heavily manipulated or photoshopped to look completely different from the originals. Until the advent of the digital camera, amateur photographers used either print or slide film for their cameras. Slides had to be developed and shown to an audience using a slide projector. Digital photography eliminated the delay and cost of film. Consumers became able to view, transfer, edit, and distribute digital images with ordinary home computers rather than using specialized equipment.

Camera phones have recently had a large impact on photography. Users can set their smartphones to upload products to the Internet, preserving images even if the camera is destroyed or the photos deleted. Some high-street photography shops have self-service kiosks that allow images to be printed directly from smartphones via Bluetooth technology.

Archivists and historians have noticed the transitory nature of digital media. Unlike film and print which are tangible, digital image storage is ever-changing, with old media and decoding software becoming obsolete or inaccessible by new technologies. Historians are concerned that this is creating a historical void where information is being silently lost within failed or inaccessible digital media. They recommend that professional and amateur users develop strategies for digital preservation by migrating stored digital images from old technologies to new ones.^[38] Scrapbookers who may have used film for creating artistic and personal memoirs may need to modify their approach to use and personalize digital photo books, thereby retaining the special qualities of traditional photo albums.

The web has been a popular medium for storing and sharing photos ever since the first photograph was published online by Tim Berners-Lee in 1992 (an image of the CERN house band Les Horribles Cernettes). Today, photo sharing sites such as Flickr, Picasa, and PhotoBucket, as well as social websites, are used by millions of people to share their pictures. Digital photography and social media allow organizations and corporations to make photographs more accessible to a greater and more diverse population. For example, National Geographic Magazine has Twitter, Snapchat, Facebook, and Instagram accounts, each of which includes content aimed at the specific audiences found on its platform.^[39]

Digital photography has also impacted other fields, such as medicine. It has allowed doctors to help diagnose diabetic retinopathy, and is used in hospitals to diagnose and treat other diseases.^[40]

Digitally altered imagery

In digital art and media art, digital photos are often edited, manipulated, or combined with other digital images. Scanography is a related process in which digital photos are created using a scanner.

New technology in digital cameras and computer editing affects the way photographic images are now perceived. The ability to create and fabricate realistic imagery digitally—as opposed to untouched photos—changes the audience's perception of "truth" in digital photography.^[41] Digital manipulation enables pictures to adjust the perception of reality, both past and present, and thereby shape people's identities, beliefs, and opinions.

Digital photography and social media

In its early stages, photography was mainly used for physically preserving a family's heritage. It has now evolved into a key part of individual identity in the 21st century.^[42] Internet users often personally photograph and repost pictures that revolve around the ways they want to personally express themselves and their chosen aesthetic.^[42] With the invention of digital photography, photographs became less destructible and more easily maintained throughout the years, living across all types of digital devices. Digital photography advanced the use of photos for communication and identity rather than as a means of remembering.^[42]

Widespread access to digital photography has greatly influenced social behavior. The phrase "pics or it didn't happen" reflects the notion that one's life experiences can only be verified by others through photographs.^[43]

Filters are commonly used in social digital photography, some of which reflect the nostalgic gap left by the disappearance of film photography. Filters that emulated traditional analog effects (such as film grain, scratches, fading, and polaroid borders) grew immensely in popularity alongside the idea of social photography, the causal sharing of everyday images.^[43] Social photos differ from "true" photography as they are not meant to carry the same value or artistic qualities.^[43]

Recent research and innovation

As of today, advancements in digital photography have sky-rocketed due to the introduction of mirrorless cameras.^[44] Due to their cutting-edge technology, portability, and versatility, being more compact and innovative, mirrorless cameras are preferred. With its manual controls, adjustable settings, interchangeable lenses, and having an electronic viewfinder or LCD screen^[45] to display images straight from the sensor,^[46] mirrorless cameras have the advantage over DSLRs.^[47] While mirrorless cameras also provide quick autofocus, silent operations, and quick shooting rates, they also have some drawbacks, like a restricted range of lenses and a shorter battery life. However, progress still continues. As of 2024, ongoing advancements in mirrorless technology continue to address these limitations, solidifying their position as a leading choice for photographers.^[48]

The rise of mirrorless cameras has changed digital photography. These cameras are popular for their modern tech, portability, and versatility. Unlike DSLRs, mirrorless cameras have electronic viewfinders or LCD screens for previewing photos and manual controls. They are smaller and lighter, but may have fewer lens options and shorter battery life. Ongoing improvements are making them even better.^[49] Mirrorless cameras give photographers new ways to shoot, like seeing previews on the LCD screen.^[50] Mirrorless cameras brought big changes to photography. They do not have the bulky parts of DSLRs, so they are smaller and easier to carry.^[51] They are also quiet, good for discreet shooting like weddings or wildlife photography.^[51] Electronic viewfinders show details like exposure and focus, helping



Modern day students have more access to photography classes as a result of digital photography's ease in comparison to film.

photographers take better shots.^[48] New autofocus systems make capturing moving subjects easier and more accurate.^[52] In summary, mirrorless cameras are changing photography with their compact size, advanced features, and quiet operation. As they improve, they are becoming essential tools for photographers.^[53]

Research and development continues to refine the lighting, optics, sensors, processing, storage, display, and software used in digital photography. Here are a few examples:

- 3D models can be created from collections of normal images. The resulting scene can be viewed from novel viewpoints, but creating the model is very computationally intensive. An example is Microsoft's Photosynth, which provided some models of famous places as examples.^[54]
- Panoramic photographs can be created directly in camera without the need for any external processing. Some cameras feature a 3D Panorama capability, combining shots taken with a single lens from different angles to create a sense of depth.
- Virtual-reality photography, the interactive visualization of photos.
- High-dynamic-range cameras and displays are commercially available. Sensors with dynamic range in excess of 1,000,000:1 are in development, and software is also available to combine multiple non-HDR images (shot with different exposures) into an HDR image.
- Motion blur can be dramatically removed by a flutter shutter (a flickering shutter that adds a signature to the blur, which postprocessing recognizes).^[55] It is not yet commercially available.
- Advanced bokeh techniques use a hardware system of 2 sensors, one to take the photo as usual while the other records depth information. Bokeh effect and refocusing can then be applied to an image after the photo is taken.^[56]
- In advanced cameras or camcorders, manipulating the sensitivity of the sensor with two or more neutral density filters.
- An object's specular reflection can be captured using computer-controlled lights and sensors. This is needed to create attractive images of oil paintings, for instance. It is not yet commercially available, but some museums are starting to use it.
- Dust reduction systems help keep dust off of image sensors. Originally introduced only by a few cameras like Olympus DSLRs, they have now become standard in most models and brands of detachable lens cameras, except the low-end or cheap ones.

Other areas of progress include improved sensors, more powerful software, advanced camera processors (sometimes using more than one processor; for instance, the Canon 7D camera has two DigiC 4 processors), enlarged gamut displays, built-in GPS and Wi-Fi, and computer-controlled lighting.

See also

- Analog photography
- Automatic image annotation
- Camcorder
- Chimping
- Design rule for Camera File system (DCF)
- Digital camera
- Digital image editing
- Digital imaging

- [Digital microscope](#)
 - [USB microscope](#)
- [Digital photo frame](#)
- [Digital Print Order Format \(DPOF\)](#)
- [Digital Revolution](#)
- [Digital single-lens reflex camera](#)
- [Digital watermarking](#)
- [Exif \(Exchangeable image file format\)](#)
- [Geotagged photograph](#)
- [High-dynamic-range imaging](#)
- [Lenses for SLR and DSLR cameras](#)
- [List of digital camera brands](#)
- [Online proofing](#)
- [Raw image format](#)
- [3D camcorder](#)

References

1. Merrin, William (2014). *Media Studies 2.0*. Routledge. p. 29. ISBN 978-0415638630.
2. Middleditch, Steve; Hand, Di (2012). *Design For Media: A Handbook for Students and Professionals in Journalism*. Routledge. p. 328. ISBN 978-1405873666.
3. Nuwer, Rachel. "The Inventor of Videotape Recorders Didn't Live to See Blockbuster's Fall" (<https://www.smithsonianmag.com/smart-news/the-inventor-of-videotape-recorders-didnt-live-to-see-blockbusters-fall-180947594/>). *Smithsonian*. Archived (<https://web.archive.org/web/20201025205229/https://www.smithsonianmag.com/smart-news/the-inventor-of-videotape-recorders-didnt-live-to-see-blockbusters-fall-180947594/>) from the original on 2020-10-25. Retrieved 2017-11-19.
4. Hernandez, Paul (2007-05-24). "Fiftieth Anniversary of First Digital Image Marked" (<https://www.nist.gov/news-events/news/2007/05/fiftieth-anniversary-first-digital-image-marked>). NIST. Archived (<https://web.archive.org/web/20170929174632/https://www.nist.gov/news-events/news/2007/05/fiftieth-anniversary-first-digital-image-marked>) from the original on 2017-09-29. Retrieved 2017-11-19.
5. James R. Janesick (2001). *Scientific charge-coupled devices* (<https://books.google.com/books?id=3GyE4SWytn4C&pg=PA3>). SPIE Press. pp. 3–4. ISBN 978-0-8194-3698-6.
6. Williams, J. B. (2017). *The Electronics Revolution: Inventing the Future* (<https://books.google.com/books?id=v4QIDwAAQBAJ&pg=PA245>). Springer. pp. 245–8. ISBN 9783319490885.
7. Boyle, William S; Smith, George E. (1970). "Charge Coupled Semiconductor Devices". *Bell Syst. Tech. J.* **49** (4): 587–593. Bibcode:1970BSTJ...49..587B (<https://ui.adsabs.harvard.edu/abs/1970BSTJ...49..587B>). doi:10.1002/j.1538-7305.1970.tb01790.x (<https://doi.org/10.1002%2Fj.1538-7305.1970.tb01790.x>).
8. "The 2009 Nobel Prize in Physics - Press Release" (<https://www.nobelprize.org/prizes/physics/2009/press-release/>). www.nobelprize.org. Archived (<https://web.archive.org/web/20180923200828/https://www.nobelprize.org/prizes/physics/2009/press-release/>) from the original on 2018-09-23. Retrieved 2017-11-19.
9. Mutch, Thomas (1978). *The Martian Landscape* (https://scholar.google.com/scholar?hl=en&as_sdt=0%2C20&q=martian+landscape&btnG=#d=gs_cit&t=170197714821&u=%2Fschorar%3Fq%3Dinfo%3Ar1xDF8BhP_sJ%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D2%26hl%3Den). NASA. ISBN 1782664882.

10. Fred C. Billingsley, "Processing Ranger and Mariner Photography," in *Computerized Imaging Techniques, Proceedings of SPIE*, Vol. 0010, pp. XV-1–19, January 1967 (August 1965, San Francisco). "Mariner is unique in that the pictures were converted to the 6-bit digital form in the spacecraft. The digital signals were transmitted at a very slow rate (8 1/3 bits/sec) and decoded and reformatted in the 7094 computer before being presented to the film recording equipment on computer tape. Thus, there are no digitizing and synchronizing problems, and the operation consists merely of producing the digitally recorded film."
11. "Mariner to Mercury, Venus, and Mars" (https://web.archive.org/web/20170215082222/http://www.jpl.nasa.gov/news/fact_sheets/mariner.pdf) (PDF). NASA Facts. Archived from the original (https://www.jpl.nasa.gov/news/fact_sheets/mariner.pdf) (PDF) on 15 February 2017. Retrieved 2 August 2012.
12. Ghosh, Pallab (1 February 2017). "Digital imaging wins engineering prize" (<https://www.bbc.com/news/science-environment-38815948>). BBC News. Archived (<https://web.archive.org/web/20180327175912/http://www.bbc.com/news/science-environment-38815948>) from the original on 27 March 2018. Retrieved 27 March 2018.
13. Ahmed, Nasir (January 1991). "How I Came Up With the Discrete Cosine Transform" (<http://www.scribd.com/doc/52879771/DCT-History-How-I-Came-Up-with-the-Discrete-Cosine-Transform>). *Digital Signal Processing*. 1 (1): 4–5. Bibcode:1991DSP.....1....4A (<https://ui.adsabs.harvard.edu/abs/1991DSP.....1....4A>). doi:10.1016/1051-2004(91)90086-Z (<https://doi.org/10.1016%2F1051-2004%2891%2990086-Z>). Archived (<https://web.archive.org/web/20160610013109/http://www.scribd.com/doc/52879771/DCT-History-How-I-Came-Up-with-the-Discrete-Cosine-Transform>) from the original on 2016-06-10. Retrieved 2019-09-14.
14. "T.81 – DIGITAL COMPRESSION AND CODING OF CONTINUOUS-TONE STILL IMAGES – REQUIREMENTS AND GUIDELINES" (<https://www.w3.org/Graphics/JPEG/itu-t81.pdf>) (PDF). CCITT. September 1992. Archived (<https://web.archive.org/web/20191230093239/http://www.w3.org/Graphics/JPEG/itu-t81.pdf>) (PDF) from the original on 30 December 2019. Retrieved 12 July 2019.
15. "The JPEG image format explained" (<https://web.archive.org/web/20190805194553/https://home.bt.com/tech-gadgets/photography/what-is-a-jpeg-11364206889349>). BT.com. BT Group. 31 May 2018. Archived from the original (<https://home.bt.com/tech-gadgets/photography/what-is-a-jpeg-11364206889349>) on 5 August 2019. Retrieved 5 August 2019.
16. "What Is a JPEG? The Invisible Object You See Every Day" (<https://www.theatlantic.com/technology/archive/2013/09/what-is-a-jpeg-the-invisible-object-you-see-every-day/279954/>). The Atlantic. 24 September 2013. Archived (<https://web.archive.org/web/20191009054159/https://www.theatlantic.com/technology/archive/2013/09/what-is-a-jpeg-the-invisible-object-you-see-every-day/279954/>) from the original on 9 October 2019. Retrieved 13 September 2019.
17. "Digital Photography Milestones from Kodak" (http://www.womeninphotography.org/Events-Exhibits/Kodak/EasyShare_3.html). Women in Photography International. Archived (https://web.archive.org/web/20091228153158/http://www.womeninphotography.org/Events-Exhibits/Kodak/EasyShare_3.html) from the original on 28 December 2009. Retrieved 17 September 2007.
18. "Kodak blog: We Had No Idea" (<https://web.archive.org/web/20130121194248/http://pluggedin.kodak.com/pluggedin/post/?id=687843>). Archived from the original (<http://pluggedin.kodak.com/post/?id=687843>) on 21 January 2013.
19. Michael R. Peres (2007). *The Focal Encyclopedia of Photography* (<https://books.google.com/books?id=VYyldcYfq3MC&q=sasson+ccd+fairchild&pg=RA1-PA16>) (4th ed.). Focal Press. ISBN 978-0-240-80740-9.
20. Estrin, James (12 August 2015). "Kodak's First Digital Moment" (<https://lens.blogs.nytimes.com/2015/08/12/kodaks-first-digital-moment/>). Lens Blog. Archived (<https://web.archive.org/web/20171201131357/https://lens.blogs.nytimes.com/2015/08/12/kodaks-first-digital-moment/>) from the original on 2017-12-01. Retrieved 2017-11-19.

21. "History" (<https://www.digitalkameramuseum.de/en/history>). *The Digital Camera Museum*. Archived (<https://web.archive.org/web/20210128185134/https://www.digitalkameramuseum.de/en/history>) from the original on 2021-01-28. Retrieved 2017-11-19.
22. Busch, David D. (2011). *Nikon D70 Digital Field Guide* (<https://books.google.com/books?id=jOVSzasqzQ4C&pg=PT11>). John Wiley & Sons. ISBN 9781118080238.
23. "1990" (<http://www.digicamhistory.com/1990.html>). *DigiCam History Dot Com*. Archived (<http://web.archive.org/web/20100626090027/http://www.digicamhistory.com/1990.html>) from the original on 26 June 2010. Retrieved 17 September 2007.
24. "Dycam Model 1: The world's first consumer digital still camera" (<http://www.digibarn.com/collections/cameras/dycam-model1/index.html>). *DigiBarn computer museum*. Archived (<https://web.archive.org/web/20101017230744/http://www.digibarn.com/collections/cameras/dycam-model1/index.html>) from the original on 2010-10-17. Retrieved 2012-02-22.
25. Carolyn Said, "DYCAM Model 1: The first portable Digital Still Camera", *MacWeek*, vol. 4, No. 35, 16 Oct. 1990, p. 34.
26. Lister, Martin (2013). *The Photographic Image in Digital Culture*. Routledge. p. 86. ISBN 978-0415535298.
27. Lee, Dong-Hoo (2010). "Digital Cameras, Personal Photography and the Reconfiguration of Spatial Experiences". *The Information Society*. 26 (4): 266–275.
doi:10.1080/01972243.2010.489854 (<https://doi.org/10.1080%2F01972243.2010.489854>).
S2CID 1661237 (<https://api.semanticscholar.org/CorpusID:1661237>).
28. "From J-Phone to Lumia 1020: A complete history of the camera phone" (<https://www.digitaltrends.com/mobile/camera-phone-history/>). *digitaltrends.com*. 11 August 2013. Archived (<http://web.archive.org/web/20190914020601/https://www.digitaltrends.com/mobile/camera-phone-history/>) from the original on 14 September 2019. Retrieved 27 March 2018.
29. "How Digital Cameras Work" (<https://electronics.howstuffworks.com/cameras-photography/digital/digital-camera.htm>). *HowStuffWorks*. 2006-11-29. Archived (<https://web.archive.org/web/20161012090749/http://electronics.howstuffworks.com/cameras-photography/digital/digital-camera.htm>) from the original on 2016-10-12. Retrieved 2016-10-11.
30. Foveon X3 Sensor Claims Put to the Test (<http://www.popphoto.com/howto/4276/foveon-x3-sensor-claims-put-to-the-test.html>) Archived (<https://web.archive.org/web/20071013144153/http://www.popphoto.com/howto/4276/foveon-x3-sensor-claims-put-to-the-test.html>) 2007-10-13 at the Wayback Machine
31. "Digital Cameras - Dec 2002" (<http://www.kegel.com/cameras.html>). Archived (<https://web.archive.org/web/20200210141336/http://kegel.com/cameras.html>) from the original on 2020-02-10. Retrieved 2020-01-20.
32. "Digital outsells film, but film still king to some" (<https://www.macworld.com/article/1039455/digitalfilm.html>). *Macworld*. Archived (<https://web.archive.org/web/20180808135843/https://www.macworld.com/article/1039455/digitalfilm.html>) from the original on 8 August 2018. Retrieved 27 March 2018.
33. Smith, Tony (2004-01-20). "Kodak to drop 35mm cameras in Europe, US" (https://www.theregister.co.uk/2004/01/20/kodak_to_drop_35mm_cameras/). *The Register*. Archived (https://web.archive.org/web/20070223091529/http://www.theregister.co.uk/2004/01/20/kodak_to_drop_35mm_cameras/) from the original on 2007-02-23. Retrieved 2007-04-03.
34. "Nikon to End Many Film-Related Products" (https://web.archive.org/web/20070223212543/http://www.pdnonline.com/pdn/prodtech/news/article_display.jsp?vnu_content_id=1001843429&imw=Y). 2006-01-11. Archived from the original (http://pdnonline.com/pdn/prodtech/news/article_display.jsp?vnu_content_id=1001843429&imw=Y) on 2007-02-23. Retrieved 2007-04-03.
35. Tomkins, Michael R. (2004-06-01). "Pentax plans to focus on digital" (<https://www.imaging-resource.com/NEWS/1086134375.html>). *The Imaging Resource*. Archived (<https://web.archive.org/web/20070515011830/http://www.imaging-resource.com/NEWS/1086134375.html>) from the original on 2007-05-15. Retrieved 2007-04-03.

36. "Eastman Kodak Files for Bankruptcy" (<https://dealbook.nytimes.com/2012/01/19/eastman-kodak-files-for-bankruptcy/>). *The New York Times*. 2012-01-19. Archived (<https://web.archive.org/web/20201213024402/https://dealbook.nytimes.com/2012/01/19/eastman-kodak-files-for-bankruptcy/>) from the original on 2020-12-13. Retrieved 2017-03-05.
37. "Digital Camera Sales Trends A declining trend that's slowly stabilizing" (<https://www.pentaxforums.com/articles/photokina-2014/digital-camera-sales-trends.html>). September 22, 2014. Archived (<https://web.archive.org/web/20190510065350/https://www.pentaxforums.com/articles/photokina-2014/digital-camera-sales-trends.html>) from the original on May 10, 2019. Retrieved May 10, 2019.
38. Lombardi, Rosie (2006-12-20). "How long will my digital pictures last?" (<https://web.archive.org/web/20070928025725/http://www.pcworld.ca/Pages/NewsColumn.aspx?id=a1ac9c720a01040800b24c9a7f5ded73>). PC World. Archived from the original (<http://www.pcworld.ca/Pages/NewsColumn.aspx?id=a1ac9c720a01040800b24c9a7f5ded73>) on 2007-09-28. Retrieved 2007-04-03.
39. Goldberg, Susan. 2017. "OUR SOCIAL MEDIA MISSION." *National Geographic* 231, no. 4: 4. Academic Search Complete, EBSCOhost (Retrieved February 10, 2018).
40. Srihatrai, Parinya, and Thanita Hlowchitsieng. 2018. "The diagnostic accuracy of single- and five-field fundus photography in diabetic retinopathy screening by primary care physicians." *Indian Journal Of Ophthalmology* 66, no. 1: 94–97. Academic Search Complete, EBSCOhost (Retriece February 23, 2018).
41. Bardis, Antonia (2004). "Digital photography and the question of realism". *Journal of Visual Art Practice*. 3 (3): 209–218. doi:[10.1386/jvap.3.3.209/0](https://doi.org/10.1386/jvap.3.3.209/0) (<https://doi.org/10.1386%2Fjvap.3.3.209%2F0>). S2CID 190744228 (<https://api.semanticscholar.org/CorpusID:190744228>).
42. van Dijck, José (February 2008). "Digital photography: communication, identity, memory" (<https://dx.doi.org/10.1177/1470357207084865>). *Visual Communication*. 7 (1): 57–76. doi:[10.1177/1470357207084865](https://doi.org/10.1177/1470357207084865) (<https://doi.org/10.1177%2F1470357207084865>). ISSN 1470-3572 (<https://search.worldcat.org/issn/1470-3572>). S2CID 220732859 (<https://api.semanticscholar.org/CorpusID:220732859>).
43. Jurgenson, Nathan (2019). *The social photo : on photography and social media* (<https://www.worldcat.org/oclc/1031408520>). London. ISBN 978-1-78873-091-4. OCLC 1031408520 (<https://search.worldcat.org/oclc/1031408520>).
44. "Mirrorless cameras" (<https://www.masterclass.com/articles/what-is-a-mirrorless-camera>).
45. Europe, Canon. "Viewfinder vs LCD Display" (<https://www.canon-europe.com/pro/infobank/lcd-or-viewfinder/>). Canon Europe. Retrieved 2024-05-29.
46. Shoot, Feature (2020-10-07). "Mirrorless vs. DSLR: The Pros and Cons Examined" (<https://www.shutterstock.com/blog/mirrorless-vs-dslr-examined>). *The Shutterstock Blog*. Retrieved 2024-05-29.
47. " "Mirrorless vs. DSLR: The Pros and Cons Examined" (<https://www.masterclass.com/articles/what-is-a-mirrorless-camera>).
48. Wilson, Timothy ColemanContributions from Mark; updated, Chris Rowlands last (2022-02-09). "The best mirrorless camera for 2024: top picks for every budget" (<https://www.techradar.com/news/best-mirrorless-camera>). TechRadar. Retrieved 2024-05-29.
49. Coleman, Chris RowlandsContributions from Timothy; published, Mark Wilson (2024-03-20). "Mirrorless vs DSLR cameras: the 10 key differences you need to know" (<https://www.techradar.com/cameras/mirrorless-vs-dslr-cameras>). TechRadar. Retrieved 2024-05-29.
50. Shoot, Feature (2020-10-07). "Mirrorless vs. DSLR: The Pros and Cons Examined" (<https://www.shutterstock.com/blog/mirrorless-vs-dslr-examined>). *The Shutterstock Blog*. Retrieved 2024-05-29.
51. "What Is a Mirrorless Camera and How Does It Compare to a DSLR?" (<https://www.masterclass.com/articles/what-is-a-mirrorless-camera>).
52. "Primer: Why would I buy a mirrorless camera?" (<https://www.dpreview.com/articles/3871263180/primer-why-would-i-buy-a-mirrorless-camera>). DPReview. Retrieved 2024-05-29.

53. Contrastly (2016-09-09). "A Year After Mirrorless" (<https://contrastly.com/a-year-after-mirrorless/>). *Contrastly*. Retrieved 2024-05-29.
54. "Photosynth" (<https://web.archive.org/web/20070205184512/http://labs.live.com/photosynth/>). Microsoft Research. Archived from the original (<http://labs.live.com/photosynth/>) on 2007-02-05. Retrieved 2007-04-03.
55. Raskar, Ramesh; Amit Agrawal; Jack Tumblin. "Coded Exposure Photography: Motion Deblurring using Fluttered Shutter" (<https://web.archive.org/web/20070429183905/http://www.cfar.umd.edu/~aagrawal/sig06/sig06Main.html>). Archived from the original (<http://www.cfar.umd.edu/~aagrawal/sig06/sig06Main.html>) on 2007-04-29. Retrieved 2007-04-03.
56. Lars Rehm (March 25, 2014). "HTC launches One M8 with new 'Duo Camera' " (<http://connect.dpreview.com/post/9386067747/htc-launches-one-m8>). Archived (<https://web.archive.org/web/20151003195918/http://connect.dpreview.com/post/9386067747/htc-launches-one-m8>) from the original on October 3, 2015. Retrieved September 29, 2015.

External links

- [Digital Photography – FAQ](https://users.cs.duke.edu/~parr/photography/faq.html) (<https://users.cs.duke.edu/~parr/photography/faq.html>)
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Digital_photography&oldid=1289510631"



Telephony

(Redirected from [Digital telephone](#))

Telephony (/tə'lefəni/ *tə-LEF-ə-nee*) is the field of technology involving the development, application, and deployment of [telecommunications services](#) for the purpose of electronic transmission of voice, [fax](#), or [data](#), between distant parties. The history of telephony is intimately linked to the invention and development of the [telephone](#).

Telephony is commonly referred to as the construction or operation of telephones and telephonic systems and as a system of telecommunications in which telephonic equipment is employed in the transmission of speech or other sound between points, with or without the use of wires.^[1] The term is also used frequently to refer to [computer hardware](#), [software](#), and [computer network](#) systems, that perform functions traditionally performed by telephone equipment. In this context the technology is specifically referred to as [Internet telephony](#), or [voice over Internet Protocol](#) (VoIP).

Overview

The first telephones were connected directly in pairs: each user had a separate telephone wired to each location to be reached. This quickly became inconvenient and unmanageable when users wanted to communicate with more than a few people. The invention of the [telephone exchange](#) provided the solution for establishing telephone connections with any other telephone in service in the local area. Each telephone was connected to the exchange at first with one wire, later one wire pair, the [local loop](#). Nearby exchanges in other service areas were connected with [trunk lines](#), and long-distance service could be established by relaying the calls through multiple exchanges.

Initially, exchange [switchboards](#) were manually operated by an attendant, commonly referred to as the "[switchboard operator](#)". When a customer cranked a handle on the telephone, it activated an indicator on the board in front of the operator, who would in response plug the operator headset into that jack and offer service. The caller had to ask for the called party by name, later by number, and the operator connected one end of a circuit into the called party jack to alert them. If the called station answered, the operator disconnected their headset and completed the station-to-station circuit. Trunk calls were made with the assistance of other operators at other exchangers in the network.

Until the 1970s, most telephones were permanently wired to the telephone line installed at customer premises. Later, conversion to installation of jacks that terminated the [inside wiring](#) permitted simple exchange of telephone sets with [telephone plugs](#) and allowed portability of the set to multiple locations in the premises where jacks were installed. The inside wiring to all jacks was connected in one place to the [wire drop](#) which connects the building to a cable. Cables usually bring a large number of drop wires from all over a district [access network](#) to one wire center or telephone exchange. When a telephone user wants to make a [telephone call](#), equipment at the exchange examines the dialed [telephone number](#) and connects that [telephone line](#) to another in the same wire center, or to a trunk to a distant exchange. Most of the exchanges in the world are interconnected through a system of larger switching systems, forming the [public switched telephone network](#) (PSTN).

In the second half of the 20th century, fax and data became important secondary applications of the network created to carry voices, and late in the century, parts of the network were upgraded with ISDN and DSL to improve handling of such traffic.

Today, telephony uses digital technology (digital telephony) in the provisioning of telephone services and systems. Telephone calls can be provided digitally, but may be restricted to cases in which the last mile is digital, or where the conversion between digital and analog signals takes place inside the telephone. This advancement has reduced costs in communication, and improved the quality of voice services. The first implementation of this, ISDN, permitted all data transport from end-to-end speedily over telephone lines.^[2] This service was later made much less important due to the ability to provide digital services based on the Internet protocol suite.^[3]

Since the advent of personal computer technology in the 1980s, computer telephony integration (CTI) has progressively provided more sophisticated telephony services, initiated and controlled by the computer, such as making and receiving voice, fax, and data calls with telephone directory services and caller identification. The integration of telephony software and computer systems is a major development in the evolution of office automation. The term is used in describing the computerized services of call centers, such as those that direct your phone call to the right department at a business you're calling. It is also sometimes used for the ability to use your personal computer to initiate and manage phone calls (in which case you can think of your computer as your personal call center).^[4]

Digital telephony

Digital telephony is the use of digital electronics in the operation and provisioning of telephony systems and services. Since the late 20th century, a digital core network has replaced the traditional analog transmission and signaling systems, and much of the access network has also been digitized.

Starting with the development of transistor technology, originating from Bell Telephone Laboratories in 1947, to amplification and switching circuits in the 1950s, the public switched telephone network (PSTN) has gradually moved towards solid-state electronics and automation. Following the development of computer-based electronic switching systems incorporating metal–oxide–semiconductor (MOS) and pulse-code modulation (PCM) technologies, the PSTN gradually evolved towards the digitization of signaling and audio transmissions. Digital telephony has since dramatically improved the capacity, quality and cost of the network. Digitization allows wideband voice on the same channel, with improved quality of a wider analog voice channel.

History

The earliest end-to-end analog telephone networks to be modified and upgraded to transmission networks with Digital Signal 1 (DS1/T1) carrier systems date back to the early 1960s. They were designed to support the basic 3 kHz voice channel by sampling the bandwidth-limited analog voice signal and encoding using pulse-code modulation (PCM). Early PCM codec-filters were implemented as passive resistor–capacitor–inductor filter circuits, with analog-to-digital conversion (for digitizing voices) and digital-to-analog conversion (for reconstructing voices) handled by discrete devices. Early digital telephony was impractical due to the low performance and high costs of early PCM codec-filters.^{[5][6]}

Practical digital telecommunication was enabled by the invention of the metal–oxide–semiconductor field-effect transistor (MOSFET),^[7] which led to the rapid development and wide adoption of PCM digital telephony.^[6] In 1957, Frosch and Derick were able to manufacture the first silicon dioxide field effect transistors at Bell Labs, the first transistors in which drain and source were adjacent at the surface.^[8] Subsequently, a team demonstrated a working MOSFET at Bell Labs 1960.^{[9][10]} MOS technology was initially overlooked by Bell because they did not find it practical for analog telephone applications, before it was commercialized by Fairchild and RCA for digital electronics such as computers.^{[11][6]}

MOS technology eventually became practical for telephone applications with the MOS mixed-signal integrated circuit, which combines analog and digital signal processing on a single chip, developed by former Bell engineer David A. Hodges with Paul R. Gray at UC Berkeley in the early 1970s.^[6] In 1974, Hodges and Gray worked with R.E. Suarez to develop MOS switched capacitor (SC) circuit technology, which they used to develop a digital-to-analog converter (DAC) chip, using MOS capacitors and MOSFET switches for data conversion.^[6] MOS analog-to-digital converter (ADC) and DAC chips were commercialized by 1974.^[12]

MOS SC circuits led to the development of PCM codec-filter chips in the late 1970s.^{[6][5]} The silicon-gate CMOS (complementary MOS) PCM codec-filter chip, developed by Hodges and W.C. Black in 1980,^[6] has since been the industry standard for digital telephony.^{[6][5]} By the 1990s, telecommunication networks such as the public switched telephone network (PSTN) had been largely digitized with very-large-scale integration (VLSI) CMOS PCM codec-filters, widely used in electronic switching systems for telephone exchanges, private branch exchanges (PBX) and key telephone systems (KTS); user-end modems; data transmission applications such as digital loop carriers, pair gain multiplexers, telephone loop extenders, integrated services digital network (ISDN) terminals, digital cordless telephones and digital cell phones; and applications such as speech recognition equipment, voice data storage, voice mail and digital tapeless answering machines.^[5] The bandwidth of digital telecommunication networks has been rapidly increasing at an exponential rate, as observed by Edholm's law,^[13] largely driven by the rapid scaling and miniaturization of MOS technology.^{[14][6]}

Uncompressed PCM digital audio with 8-bit depth and 8 kHz sample rate requires a bit rate of 64 kbit/s, which was impractical for early digital telecommunication networks with limited network bandwidth. A solution to this issue was linear predictive coding (LPC), a speech coding data compression algorithm that was first proposed by Fumitada Itakura of Nagoya University and Shuzo Saito of Nippon Telegraph and Telephone (NTT) in 1966. LPC was capable of audio data compression down to 2.4 kbit/s, leading to the first successful real-time conversations over digital networks in the 1970s.^[15] LPC has since been the most widely used speech coding method.^[16] Another audio data compression method, a discrete cosine transform (DCT) algorithm called the modified discrete cosine transform (MDCT), has been widely adopted for speech coding in voice-over-IP (VoIP) applications since the late 1990s.^[17]

The development of transmission methods such as SONET and fiber optic transmission further advanced digital transmission. Although analog carrier systems existed that multiplexed multiple analog voice channels onto a single transmission medium, digital transmission allowed lower cost and more channels multiplexed on the transmission medium. Today the end instrument often remains analog but the analog signals are typically converted to digital signals at the serving area interface (SAI), central office (CO), or other aggregation point. Digital loop carriers (DLC) and fiber to the x place the digital network ever closer to the customer premises, relegating the analog local loop to legacy status.

IP telephony

The field of technology available for telephony has broadened with the advent of new communication technologies. Telephony now includes the technologies of Internet services and mobile communication, including video conferencing.

The new technologies based on Internet Protocol (IP) concepts are often referred to separately as voice over IP (VoIP) telephony, also commonly referred to as IP telephony or Internet telephony. Unlike traditional phone service, IP telephony service is relatively unregulated by government. In the United States, the Federal Communications Commission (FCC) regulates phone-to-phone connections, but says they do not plan to regulate connections between a phone user and an IP telephony service provider.^[18]



A commercial IP telephone, with keypad, control keys, and screen functions to perform configuration and user features

A specialization of digital telephony, Internet Protocol (IP) telephony involves the application of digital networking technology that was the foundation to the Internet to create, transmit, and receive telecommunications sessions over computer networks. Internet telephony is commonly known as voice over Internet Protocol (VoIP), reflecting the principle, but it has been referred with many other terms. VoIP has proven to be a disruptive technology that is rapidly replacing traditional telephone infrastructure technologies. As of January 2005, up to 10% of telephone subscribers in Japan and South Korea have switched to this digital telephone service. A January 2005 Newsweek article suggested that Internet telephony may be "the next big thing".^[19] As of 2006, many VoIP companies offer service to consumers and businesses.

A significant advancement in mobile telephony has been the integration of IP technologies into mobile networks, notably through Voice over LTE (VoLTE) and Voice over 5G (Vo5G). These technologies enable voice calls to be transmitted over the same IP-based infrastructure used for data services, offering improved call quality and faster connections compared to traditional circuit-switched networks. VoLTE and Vo5G are becoming the standard for mobile voice communication in many regions, as mobile operators transition to all-IP networks.^{[20][21]}

IP telephony uses an Internet connection and hardware IP phones, analog telephone adapters, or softphone computer applications to transmit conversations encoded as data packets. While one of the most common and cost-effective uses of IP telephony is through connections over WiFi hotspots, it is also employed on private networks and over other types of Internet connections, which may or may not have a direct link to the global telephone network.

Social impact research

Direct person-to-person communication includes non-verbal cues expressed in facial and other bodily articulation, that cannot be transmitted in traditional voice telephony. Video telephony restores such interactions to varying degrees. Social Context Cues Theory is a model to measure the success of

different types of communication in maintaining the non-verbal cues present in face-to-face interactions. The research examines many different cues, such as the physical context, different facial expressions, body movements, tone of voice, touch and smell.

Various communication cues are lost with the usage of the telephone. The communicating parties are not able to identify the body movements, and lack touch and smell. Although this diminished ability to identify social cues is well known, Wiesenfeld, Raghuram, and Garud point out that there is a value and efficiency to the type of communication for different tasks.^[22] They examine work places in which different types of communication, such as the telephone, are more useful than face-to-face interaction.

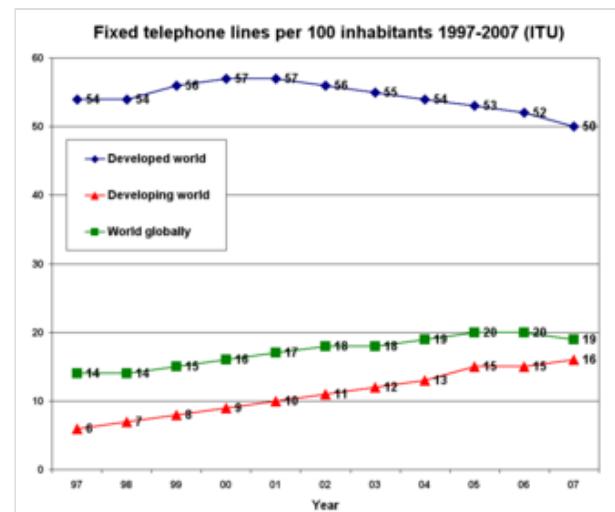
The expansion of communication to mobile telephone service has created a different filter of the social cues than the land-line telephone. The use of instant messaging, such as *texting*, on mobile telephones has created a sense of community.^[23] In *The Social Construction of Mobile Telephony* it is suggested that each phone call and text message is more than an attempt to converse. Instead, it is a gesture which maintains the social network between family and friends. Although there is a loss of certain social cues through telephones, mobile phones bring new forms of expression of different cues that are understood by different audiences. New language additives attempt to compensate for the inherent lack of non-physical interaction.

Another social theory supported through telephony is the Media Dependency Theory. This theory concludes that people use media or a resource to attain certain goals. This theory states that there is a link between the media, audience, and the large social system.^[24] Telephones, depending on the person, help attain certain goals like accessing information, keeping in contact with others, sending quick communication, entertainment, etc.

See also



- [Extended area service](#)
- [History of the telephone](#)
- [Invention of the telephone](#)
- [List of telephony terminology](#)
- [Stimulus protocol](#)



Fixed telephone lines per 100 inhabitants (1997–2007)

References

1. Dictionary.com (<http://dictionary.reference.com/browse/telephony>) Telephony Definition
2. "The Communications Museum Trust - eMuseum - History of Digital Switching -ISDN" (<http://www.communicationsmuseum.org.uk/emuseum/electronicswitching/digital/isdn/>). www.communicationsmuseum.org.uk. Retrieved 2022-08-25.
3. "Why ISDN telephones are in decline - Sussex, Surrey, Brighton | Ingenio" (<https://ingeniotech.co.uk/isdn-telephones-decline/>). ingeniotech.co.uk. 2022-03-07. Retrieved 2022-08-25.
4. What is CTI? (<http://searchnetworking.techtarget.com/definition/CTI>) TechTarget
5. Floyd, Michael D.; Hillman, Garth D. (8 October 2018) [1st pub. 2000]. "Pulse-Code Modulation Codec-Filters" (<https://books.google.com/books?id=Tokk5bZxB0MC&pg=SA26-PA1>). *The Communications Handbook* (2nd ed.). CRC Press. pp. 26–1, 26–2, 26–3. ISBN 9781420041163.
6. Allstot, David J. (2016). "Switched Capacitor Filters". In Maloberti, Franco; Davies, Anthony C. (eds.). *A Short History of Circuits and Systems: From Green, Mobile, Pervasive Networking to Big Data Computing* (https://web.archive.org/web/20210930151716/https://iee-e-cas.org/sites/default/files/a_short_history_of_circuits_and_systems_ebook_web.pdf) (PDF). IEEE Circuits and Systems Society. pp. 105–110. ISBN 9788793609860. Archived from the original (https://iee-e-cas.org/sites/default/files/a_short_history_of_circuits_and_systems_ebook_web.pdf) (PDF) on 2021-09-30. Retrieved 2019-11-28.
7. Colinge, Jean-Pierre; Colinge, C. A. (2005). *Physics of Semiconductor Devices* (<https://books.google.com/books?id=-o5bAG5pR3AC&pg=PA165>). Springer Science & Business Media. p. 165. ISBN 9780387285238.
8. Frosch, C. J.; Derick, L (1957). "Surface Protection and Selective Masking during Diffusion in Silicon" (<https://iopscience.iop.org/article/10.1149/1.2428650>). *Journal of the Electrochemical Society*. **104** (9): 547. doi:[10.1149/1.2428650](https://doi.org/10.1149/1.2428650) (<https://doi.org/10.1149%2F1.2428650>).
9. KAHNG, D. (1961). "Silicon-Silicon Dioxide Surface Device" (https://doi.org/10.1142/9789814503464_0076). *Technical Memorandum of Bell Laboratories*: 583–596. doi:[10.1142/9789814503464_0076](https://doi.org/10.1142/9789814503464_0076) (https://doi.org/10.1142%2F9789814503464_0076). ISBN 978-981-02-0209-5.
10. Lojek, Bo (2007). *History of Semiconductor Engineering*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. p. 321. ISBN 978-3-540-34258-8.
11. Maloberti, Franco; Davies, Anthony C. (2016). "History of Electronic Devices". *A Short History of Circuits and Systems: From Green, Mobile, Pervasive Networking to Big Data Computing* (https://web.archive.org/web/20210930151716/https://iee-e-cas.org/sites/default/files/a_short_history_of_circuits_and_systems_ebook_web.pdf) (PDF). IEEE Circuits and Systems Society. pp. 59-70 (65-7). ISBN 9788793609860. Archived from the original (https://iee-e-cas.org/sites/default/files/a_short_history_of_circuits_and_systems_ebook_web.pdf) (PDF) on 2021-09-30. Retrieved 2019-11-28.
12. *Electronic Components* (<https://books.google.com/books?id=HikuAAAAMAAJ&pg=PA46>). U.S. Government Printing Office. 1974. p. 46.
13. Cherry, Steven (2004). "Edholm's law of bandwidth". *IEEE Spectrum*. **41** (7): 58–60. doi:[10.1109/MSPEC.2004.1309810](https://doi.org/10.1109/MSPEC.2004.1309810) (<https://doi.org/10.1109%2FMSPEC.2004.1309810>). S2CID 27580722 (<https://api.semanticscholar.org/CorpusID:27580722>).

14. Jindal, Renuka P. (2009). "From millibits to terabits per second and beyond - over 60 years of innovation" (<https://events.vtools.ieee.org/m/195547>). *2009 2nd International Workshop on Electron Devices and Semiconductor Technology*. pp. 1–6. doi:[10.1109/EDST.2009.5166093](https://doi.org/10.1109/EDST.2009.5166093) (<https://doi.org/10.1109%2FEDST.2009.5166093>). ISBN 978-1-4244-3831-0. S2CID 25112828 (<https://api.semanticscholar.org/CorpusID:25112828>).
15. Gray, Robert M. (2010). "A History of Realtime Digital Speech on Packet Networks: Part II of Linear Predictive Coding and the Internet Protocol" (<https://ee.stanford.edu/~gray/lpcip.pdf>) (PDF). *Found. Trends Signal Process.* **3** (4): 203–303. doi:[10.1561/2000000036](https://doi.org/10.1561/2000000036) (<https://doi.org/10.1561%2F2000000036>). ISSN 1932-8346 (<https://search.worldcat.org/issn/1932-8346>). Archived (<https://ghostarchive.org/archive/20221009/https://ee.stanford.edu/~gray/lpcip.pdf>) (PDF) from the original on 2022-10-09.
16. Gupta, Shipra (May 2016). "Application of MFCC in Text Independent Speaker Recognition" (<https://web.archive.org/web/20191018231621/https://pdfs.semanticscholar.org/2aa9/c2971342e8b0b1a0714938f39c406f258477.pdf>) (PDF). *International Journal of Advanced Research in Computer Science and Software Engineering.* **6** (5): 805–810 (806). ISSN 2277-128X (<https://search.worldcat.org/issn/2277-128X>). S2CID 212485331 (<https://api.semanticscholar.org/CorpusID:212485331>). Archived from the original (<https://pdfs.semanticscholar.org/2aa9/c2971342e8b0b1a0714938f39c406f258477.pdf>) (PDF) on 2019-10-18. Retrieved 18 October 2019.
17. Schnell, Markus; Schmidt, Markus; Jander, Manuel; Albert, Tobias; Geiger, Ralf; Ruoppila, Vesa; Ekstrand, Per; Bernhard, Grill (October 2008). *MPEG-4 Enhanced Low Delay AAC - A New Standard for High Quality Communication* (https://www.iis.fraunhofer.de/content/dam/iis/de/doc/ame/conference/AES-125-Convention_AAC-ELD-NewStandardForHighQualityCommunication_AES7503.pdf) (PDF). 125th AES Convention. *Fraunhofer IIS*. Audio Engineering Society. Archived (https://ghostarchive.org/archive/20221009/https://www.iis.fraunhofer.de/content/dam/iis/de/doc/ame/conference/AES-125-Convention_AAC-ELD-NewStandardForHighQualityCommunication_AES7503.pdf) (PDF) from the original on 2022-10-09. Retrieved 20 October 2019.
18. "Microsoft word - 37716" (<https://docs.fcc.gov/public/attachments/FCC-04-97A1.pdf>) (PDF). docs.fcc.gov.
19. Sheridan, Barrett. "Newsweek - National News, World News, Health, Technology, Entertainment and more... - Newsweek.com" (<https://web.archive.org/web/20050118033848/http://msnbc.msn.com/id/6831938/site/newsweek/>). MSNBC. Archived from the original (<https://msnbc.msn.com/id/6831938/site/newsweek/>) on January 18, 2005. Retrieved 2010-05-23.
20. "Communication services (VoLTE/VoNR)" (<https://www.3gpp.org/technologies/volte-vonr>). www.3gpp.org. Retrieved 2024-11-30.
21. "VoLTE Vs Vo5G | Difference between VoLTE and VoNR" (<https://www.rfwireless-world.com/Terminology/Difference-between-VoLTE-and-VoNR.html>). www.rfwireless-world.com. Retrieved 2024-11-30.
22. "Benefits of Installing IP Telephony" (<https://alfuzail.com/blog/benefits-of-installing-ip-telephony/>). 11 January 2025. Retrieved 8 Feb 2025.
23. Mesch, Gustavo S.; Talmud, Ilan; Quan-Haase, Anabel (2012-09-01). "Instant messaging social networks: Individual, relational, and cultural characteristics". *Journal of Social and Personal Relationships.* **29** (6): 736–759. doi:[10.1177/0265407512448263](https://doi.org/10.1177/0265407512448263) (<https://doi.org/10.1177%2F0265407512448263>). ISSN 0265-4075 (<https://search.worldcat.org/issn/0265-4075>). S2CID 144874874 (<https://api.semanticscholar.org/CorpusID:144874874>).
24. "Media Dependency Theory" (<http://communicationtheory.org/media-dependency-theory/>). 2012-02-12.



Digital video

Digital video is an electronic representation of moving visual images ([video](#)) in the form of encoded [digital data](#). This is in contrast to [analog video](#), which represents moving visual images in the form of [analog signals](#). Digital video comprises a series of [digital images](#) displayed in rapid succession, usually at 24, 25, 30, or 60 [frames per second](#). Digital video has many advantages such as easy copying, multicasting, sharing and storage.

Digital video was first introduced commercially in 1986 with the [Sony D1](#) format, which recorded an [uncompressed standard-definition component video signal](#) in digital form. In addition to [uncompressed formats](#), popular [compressed](#) digital video formats today include [MPEG-2](#), [H.264](#) and [AV1](#). Modern interconnect standards used for playback of digital video include [HDMI](#), [DisplayPort](#), [Digital Visual Interface \(DVI\)](#) and [serial digital interface \(SDI\)](#).

Digital video can be copied and reproduced with no degradation in quality. In contrast, when analog sources are copied, they experience [generation loss](#). Digital video can be stored on digital media such as [Blu-ray Disc](#), on [computer data storage](#), or [streamed](#) over the [Internet](#) to [end users](#) who watch content on a personal computer or mobile device screen or a [digital smart TV](#). Today, digital video content such as [TV shows](#) and [movies](#) also includes a [digital audio soundtrack](#).

History

Cameras

The basis for [digital video cameras](#) is [metal–oxide–semiconductor \(MOS\) image sensors](#).^[1] The first practical [semiconductor](#) image sensor was the [charge-coupled device \(CCD\)](#), invented in 1969^[2] by Willard S. Boyle, who won a Nobel Prize for his work in physics.^[3] Following the commercialization of CCD sensors during the late 1970s to early 1980s, the [entertainment industry](#) slowly began transitioning to [digital imaging](#) and digital video from analog video over the next two decades.^[4] The CCD was followed by the [CMOS active-pixel sensor \(CMOS sensor\)](#),^[5] developed in the 1990s.^{[6][7]}

Major films^[a] shot on digital video overtook those shot on film in 2013. Since 2016 over 90% of major films were shot on digital video.^{[8][9]} As of 2017, 92% of films are shot on digital.^[10] Only 24 major films released in 2018 were shot on 35mm.^[11] Today, cameras from companies like [Sony](#), [Panasonic](#), [JVC](#) and [Canon](#) offer a variety of choices for shooting high-definition video. At the high end of the market, there has been an emergence of cameras aimed specifically at the digital cinema market. These



Sony digital video camera used for recording content

cameras from Sony, Vision Research, Arri, Blackmagic Design, Panavision, Grass Valley and Red offer resolution and dynamic range that exceeds that of traditional video cameras, which are designed for the limited needs of broadcast television.^[12]

Coding

In the 1970s, pulse-code modulation (PCM) induced the birth of digital video coding, demanding high bit rates of 45-140 Mbit/s for standard-definition (SD) content. By the 1980s, the discrete cosine transform (DCT) became the standard for digital video compression.^[13]

The first digital video coding standard was H.120, created by the (International Telegraph and Telephone Consultative Committee) or CCITT (now ITU-T) in 1984. H.120 was not practical due to weak performance.^[14] H.120 was based on differential pulse-code modulation (DPCM), a compression algorithm that was inefficient for video coding. During the late 1980s, a number of companies began experimenting with DCT, a much more efficient form of compression for video coding. The CCITT received 14 proposals for DCT-based video compression formats, in contrast to a single proposal based on vector quantization (VQ) compression. The H.261 standard was developed based on DCT compression,^[15] becoming first practical video coding standard.^[14] Since H.261, DCT compression has been adopted by all the major video coding standards that followed.^[15]

MPEG-1, developed by the Motion Picture Experts Group (MPEG), followed in 1991, and it was designed to compress VHS-quality video. It was succeeded in 1994 by MPEG-2/H.262,^[14] which became the standard video format for DVD and SD digital television.^[14] It was followed by MPEG-4 in 1999, and then in 2003 it was followed by H.264/MPEG-4 AVC, which has become the most widely used video coding standard.^[16]

The current-generation video coding format is HEVC (H.265), introduced in 2013. While AVC uses the integer DCT with 4x4 and 8x8 block sizes, HEVC uses integer DCT and DST transforms with varied block sizes between 4x4 and 32x32.^[17] HEVC is heavily patented, with the majority of patents belonging to Samsung Electronics, GE, NTT and JVC Kenwood.^[18] It is currently being challenged by the aiming-to-be-freely-licensed AV1 format. As of 2019, AVC is by far the most commonly used format for the recording, compression and distribution of video content, used by 91% of video developers, followed by HEVC which is used by 43% of developers.^[19]

Production

Starting in the late 1970s to the early 1980s, video production equipment that was digital in its internal workings was introduced. These included time base correctors (TBC)^[b] and digital video effects (DVE) units.^[c] They operated by taking a standard analog composite video input and digitizing it internally. This made it easier to either correct or enhance the video signal, as in the case of a TBC, or to manipulate and add effects to the video, in the case of a DVE unit. The digitized and processed video information was then converted back to standard analog video for output.



A Betacam SP camera, originally developed in 1986 by Sony

Later on in the 1970s, manufacturers of professional video broadcast equipment, such as Bosch (through their Fernseh division) and Ampex developed prototype digital videotape recorders (VTR) in their research and development labs. Bosch's machine used a modified 1-inch type B videotape transport and recorded an early form of CCIR 601 digital video. Ampex's prototype digital video recorder used a modified 2-inch quadruplex videotape VTR (an Ampex AVR-3) fitted with custom digital video electronics and a special octaplex 8-head headwheel (regular analog 2" quad machines only used 4 heads). Like standard 2" quad, the audio on the Ampex prototype digital machine, nicknamed *Annie* by its developers, still recorded the audio in analog as linear tracks on the tape. None of these machines from these manufacturers were ever marketed commercially.

Digital video was first introduced commercially in 1986 with the Sony D1 format, which recorded an uncompressed standard definition component video signal in digital form. Component video connections required 3 cables, but most television facilities were wired for composite NTSC or PAL video using one cable. Due to this incompatibility the cost of the recorder, D1 was used primarily by large television networks and other component-video capable video studios.

In 1988, Sony and Ampex co-developed and released the D2 digital videocassette format, which recorded video digitally without compression in ITU-601 format, much like D1. In comparison, D2 had the major difference of encoding the video in composite form to the NTSC standard, thereby only requiring single-cable composite video connections to and from a D2 VCR. This made it a perfect fit for the majority of television facilities at the time. D2 was a successful format in the television broadcast industry throughout the late '80s and the '90s. D2 was also widely used in that era as the master tape format for mastering laserdiscs.^[d]

D1 & D2 would eventually be replaced by cheaper systems using video compression, most notably Sony's Digital Betacam, that were introduced into the network's television studios. Other examples of digital video formats utilizing compression were Ampex's DCT (the first to employ such when introduced in 1992), the industry-standard DV and MiniDV and its professional variations, Sony's DVCAM and Panasonic's DVCPRO, and Betacam SX, a lower-cost variant of Digital Betacam using MPEG-2 compression.^[20]

One of the first digital video products to run on personal computers was *PACo: The PICS Animation Compiler* from The Company of Science & Art in Providence, RI. It was developed starting in 1990 and first shipped in May 1991. PACo could stream unlimited-length video with synchronized sound from a single file (with the .CAV file extension) on CD-ROM. Creation required a Mac, and playback was possible on Macs, PCs, and Sun SPARCstations.^[21]

QuickTime, Apple Computer's multimedia framework, was released in June 1991. Audio Video Interleave from Microsoft followed in 1992. Initial consumer-level content creation tools were crude, requiring an analog video source to be digitized to a computer-readable format. While low-quality at first, consumer digital video increased rapidly in quality, first with the introduction of playback standards such as MPEG-1 and MPEG-2 (adopted for use in television transmission and DVD media), and the introduction of the



A professional television studio set in Chile

SONY

The Sony logo, creator of the Betacam

DV tape format allowing recordings in the format to be transferred directly to digital video files using a FireWire port on an editing computer. This simplified the process, allowing non-linear editing systems (NLE) to be deployed cheaply and widely on desktop computers with no external playback or recording equipment needed.

The widespread adoption of digital video and accompanying compression formats has reduced the bandwidth needed for a high-definition video signal (with HDV and AVCHD, as well as several professional formats such as XDCAM, all using less bandwidth than a standard definition analog signal). These savings have increased the number of channels available on cable television and direct broadcast satellite systems, created opportunities for spectrum reallocation of terrestrial television broadcast frequencies, and made tapeless camcorders based on flash memory possible, among other innovations and efficiencies.

Culture

Culturally, digital video has allowed video and film to become widely available and popular, beneficial to entertainment, education, and research.^[22] Digital video is increasingly common in schools, with students and teachers taking an interest in learning how to use it in relevant ways.^[23] Digital video also has healthcare applications, allowing doctors to track infant heart rates and oxygen levels.^[24]

In addition, the switch from analog to digital video impacted media in various ways, such as in how businesses use cameras for surveillance. Closed circuit television (CCTV) switched to using digital video recorders (DVR), presenting the issue of how to store recordings for evidence collection. Today, digital video is able to be compressed in order to save storage space.^[25]

Digital television

Digital television (DTV) is the production and transmission of digital video from networks to consumers. This technique uses digital encoding instead of analog signals used prior to the 1950s.^[26] As compared to analog methods, DTV is faster and provides more capabilities and options for data to be transmitted and shared.^[27]

Digital television's roots are tied to the availability of inexpensive, high-performance computers. It was not until the 1990s that digital TV became a real possibility.^[28] Digital television was previously not practically feasible due to the impractically high bandwidth requirements of uncompressed video,^[29] requiring around 200 Mbit/s for a standard-definition television (SDTV) signal,^{[30][31]} and over 1 Gbit/s for high-definition television (HDTV).^{[29][32]}

Overview

Digital video comprises a series of digital images displayed in rapid succession. In the context of video, these images are called frames.^[e] The rate at which frames are displayed is known as the frame rate and is measured in frames per second. Every frame is a digital image and so comprises a formation of pixels. The color of a pixel is represented by a fixed number of bits of that color where the information of the

color is stored within the image.^[33] For example, 8-bit captures 256 levels per channel, and 10-bit captures 1,024 levels per channel.^[34] The more bits, the more subtle variations of colors can be reproduced. This is called the color depth, or bit depth, of the video.

Interlacing

In interlaced video each *frame* is composed of two halves of an image. The first half contains only the odd-numbered lines of a full frame. The second half contains only the even-numbered lines. These halves are referred to individually as *fields*. Two consecutive fields compose a full frame. If an interlaced video has a frame rate of 30 frames per second the field rate is 60 fields per second, though both part of interlaced video, frames per second and fields per second are separate numbers.

Bit rate and BPP

By definition, bit rate is a measurement of the rate of information content from the digital video stream. In the case of uncompressed video, bit rate corresponds directly to the quality of the video because bit rate is proportional to every property that affects the video quality. Bit rate is an important property when transmitting video because the transmission link must be capable of supporting that bit rate. Bit rate is also important when dealing with the storage of video because, as shown above, the video size is proportional to the bit rate and the duration. Video compression is used to greatly reduce the bit rate while having little effect on quality.^[35]

Bits per pixel (BPP) is a measure of the efficiency of compression. A true-color video with no compression at all may have a BPP of 24 bits/pixel. Chroma subsampling can reduce the BPP to 16 or 12 bits/pixel. Applying JPEG compression on every frame can reduce the BPP to 8 or even 1 bits/pixel. Applying video compression algorithms like MPEG1, MPEG2 or MPEG4 allows for fractional BPP values to exist.

Constant bit rate versus variable bit rate

BPP represents the *average* bits per pixel. There are compression algorithms that keep the BPP almost constant throughout the entire duration of the video. In this case, we also get video output with a constant bitrate (CBR). This CBR video is suitable for real-time, non-buffered, fixed bandwidth video streaming (e.g. in videoconferencing). Since not all frames can be compressed at the same level, because quality is more severely impacted for scenes of high complexity, some algorithms try to constantly adjust the BPP. They keep the BPP high while compressing complex scenes and low for less demanding scenes.^[36] This way, it provides the best quality at the smallest average bit rate (and the smallest file size, accordingly). This method produces a variable bitrate because it tracks the variations of the BPP.



A broadcast television camera at the Pavek Museum in Minnesota.

Technical overview

Standard film stocks typically record at 24 frames per second. For video, there are two frame rate standards: NTSC, at 30/1.001 (about 29.97) frames per second (about 59.94 fields per second), and PAL, 25 frames per second (50 fields per second). Digital video cameras come in two different image capture formats: interlaced and progressive scan. Interlaced cameras record the image in alternating sets of lines: the odd-numbered lines are scanned, and then the even-numbered lines are scanned, then the odd-numbered lines are scanned again, and so on.

One set of odd or even lines is referred to as a *field*, and a consecutive pairing of two fields of opposite parity is called a *frame*. Progressive scan cameras record all lines in each frame as a single unit. Thus, interlaced video captures the scene motion twice as often as progressive video does for the same frame rate. Progressive scan generally produces a slightly sharper image, however, motion may not be as smooth as interlaced video.

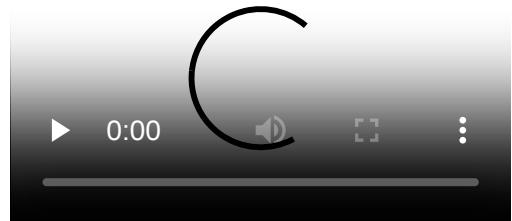
Digital video can be copied with no generation loss; which degrades quality in analog systems. However, a change in parameters like frame size, or a change of the digital format can decrease the quality of the video due to image scaling and transcoding losses. Digital video can be manipulated and edited on non-linear editing systems.

Digital video has a significantly lower cost than 35 mm film. In comparison to the high cost of film stock, the digital media used for digital video recording, such as flash memory or hard disk drive is very inexpensive. Digital video also allows footage to be viewed on location without the expensive and time-consuming chemical processing required by film. Network transfer of digital video makes physical deliveries of tapes and film reels unnecessary.

Digital television (including higher quality HDTV) was introduced in most developed countries in early 2000s. Today, digital video is used in modern mobile phones and video conferencing systems. Digital video is used for Internet distribution of media, including streaming video and peer-to-peer movie distribution.

Many types of video compression exist for serving digital video over the internet and on optical disks. The file sizes of digital video used for professional editing are generally not practical for these purposes, and the video requires further compression with codecs to be used for recreational purposes.

As of 2017, the highest image resolution demonstrated for digital video generation is 132.7 megapixels (15360 x 8640 pixels). The highest speed is attained in industrial and scientific high-speed cameras that are capable of filming 1024x1024 video at up to 1 million frames per second for brief periods of recording.

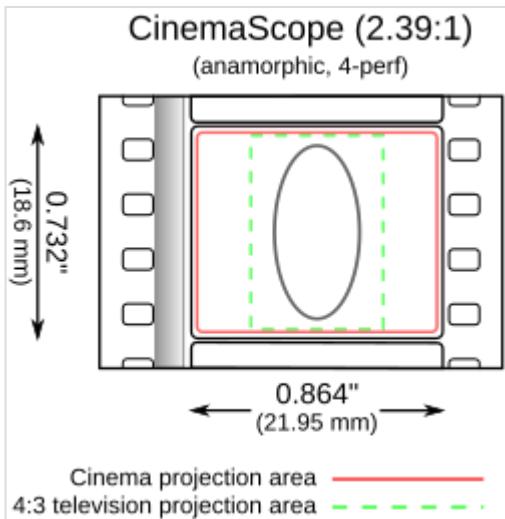


A short video sequence in native 16K.

Technical properties

Live digital video consumes bandwidth. Recorded digital video consumes data storage. The amount of bandwidth or storage required is determined by the frame size, color depth and frame rate. Each pixel consumes a number of bits determined by the color depth. The data required to represent a frame of data is determined by multiplying by the number of pixels in the image. The bandwidth is determined by multiplying the storage requirement for a frame by the frame rate. The overall storage requirements for a program can then be determined by multiplying bandwidth by the duration of the program.

These calculations are accurate for uncompressed video, but due to the relatively high bit rate of uncompressed video, video compression is extensively used. In the case of compressed video, each frame requires only a small percentage of the original bits. This reduces the data or bandwidth consumption by a factor of 5 to 12 times when using lossless compression, but more commonly, lossy compression is used due to its reduction of data consumption by factors of 20 to 200.^[37] Note that it is not necessary that all frames are equally compressed by the same percentage. Instead, consider the *average* factor of compression for *all* the frames taken together.



A diagram of 35 mm film as used in Cinemascope cameras.

Interfaces and cables

Purpose-built digital video interfaces

- Digital component video
- Digital Visual Interface (DVI)
- DisplayPort
- HDBaseT
- High-Definition Multimedia Interface (HDMI)
- Unified Display Interface

General-purpose interfaces use to carry digital video

- FireWire (IEEE 1394)
- Universal Serial Bus (USB)

The following interface has been designed for carrying MPEG-Transport compressed video:

- DVB-ASI

Compressed video is also carried using UDP-IP over Ethernet. Two approaches exist for this:

- Using RTP as a wrapper for video packets as with SMPTE 2022
- 1–7 MPEG Transport Packets are placed directly in the UDP packet

Other methods of carrying video over IP

- [Network Device Interface](#)
- [SMPTE 2110](#)

Storage formats

Encoding

- [CCIR 601](#) used for broadcast stations
- [VC-2](#) also known as *Dirac Pro*
- [MPEG-4](#) good for online distribution of large videos and video recorded to [flash memory](#)
- [MPEG-2](#) used for DVDs, Super-VCDs, and many broadcast television formats
- [MPEG-1](#) used for video CDs
- [H.261](#)
- [H.263](#)
- [H.264](#) also known as *MPEG-4 Part 10*, or as AVC, used for [Blu-ray Discs](#) and some broadcast television formats
- [H.265](#) also known as *MPEG-H Part 2*, or as HEVC
- [MOV](#) used for [QuickTime](#) framework
- [Theora](#) used for video on Wikipedia

Tapes

- [Betacam SX](#), [MPEG IMX](#), [Digital Betacam](#), or [DigiBeta](#) — professional video formats by Sony, based on original [Betamax](#) technology
- [D-VHS](#) — MPEG-2 format data recorded on a tape similar to [S-VHS](#)
- [D1](#), [D2](#), [D3](#), [D5](#), [D7](#), [D9](#) (also known as Digital-S) — various [SMPTE](#) professional digital video standards
- [D8](#) — DV-format data recorded on [Hi8](#)-compatible cassettes; largely a consumer format
- [DV](#), [MiniDV](#) — used in most of digital videocassette consumer camcorders; designed for high quality and easy editing; can also record high-definition data ([HDV](#)) in [MPEG-2](#) format
- [DVCPAM](#), [DVCPRO](#) — used in professional broadcast operations; similar to DV but generally considered more robust; though DV-compatible, these formats have better audio handling.
- [DVCPRO50](#) and [DVCPRO HD](#) support higher bandwidths as compared to Panasonic's DVCPRO.
- [HDCAM](#) and [HDCAM SR](#) were introduced by Sony as a high-definition alternative to DigiBeta.
- [MicroMV](#) — MPEG-2-format data recorded on a very small, matchbook-sized cassette; obsolete
- [ProHD](#) — name used by JVC for its MPEG-2-based professional camcorders



An archived B-format video tape used in Danish broadcasting.

Discs

- [Blu-ray Disc](#)
- [DVD](#)
- [VCD](#)

See also

- [Digital audio](#)
- [Digital cinematography](#)
- [Display aspect ratio](#)
- [Display resolution](#)
- [Index of video-related articles](#)
- [Internet video](#)
- [Online video platform](#)
- [Video coding format](#)
- [Video editing software](#)
- [Webcam](#)



The Blu-ray disc, a type of optical disc used for media storage.

Notes

- a. Defined as the top 200 grossing live-action films
- b. For example, the [Thomson-CSF 9100](#) Digital Video Processor, an internally all-digital full-frame TBC introduced in 1980.
- c. For example the [Ampex ADO](#), and the [Nippon Electric Corporation \(NEC\) E-Flex](#).
- d. Prior to D2, most laserdiscs were mastered using analog [1" Type C videotape](#)
- e. In fact the still images correspond to frames only in the case of progressive scan video. In interlaced video, they correspond to fields. See [§ Interlacing](#) for clarification.

References

1. Williams, J. B. (2017). *The Electronics Revolution: Inventing the Future* (<https://books.google.com/books?id=v4QIDwAAQBAJ&pg=PA245>). Springer. pp. 245–8. ISBN 9783319490885.
2. James R. Janesick (2001). *Scientific charge-coupled devices* (<https://books.google.com/books?id=3GyE4SWytn4C&pg=PA3>). SPIE Press. pp. 3–4. ISBN 978-0-8194-3698-6.
3. "2009 Nobel Prize in Physics awarded to Kao, Boyle, and Smith" (<https://dx.doi.org/10.1063/pt.5.023739>). *Physics Today* (10): 14182. 2009. Bibcode:2009PhT..2009j4182. (<https://ui.adsabs.harvard.edu/abs/2009PhT..2009j4182>). doi:10.1063/pt.5.023739 (<https://doi.org/10.1063/pt.5.023739>). ISSN 1945-0699 (<https://search.worldcat.org/issn/1945-0699>).
4. Stump, David (2014). *Digital Cinematography: Fundamentals, Tools, Techniques, and Workflows* (<https://books.google.com/books?id=c-MjAwAAQBAJ&pg=PA83>). CRC Press. pp. 83–5. ISBN 978-1-136-04042-9.
5. Stump, David (2014). *Digital Cinematography: Fundamentals, Tools, Techniques, and Workflows* (<https://books.google.com/books?id=c-MjAwAAQBAJ&pg=PA19>). CRC Press. pp. 19–22. ISBN 978-1-136-04042-9.

6. Fossum, Eric R.; Hondongwa, D. B. (2014). "A Review of the Pinned Photodiode for CCD and CMOS Image Sensors" (<https://doi.org/10.1109%2FJEDS.2014.2306412>). *IEEE Journal of the Electron Devices Society*. 2 (3): 33–43. doi:10.1109/JEDS.2014.2306412 (<https://doi.org/10.1109%2FJEDS.2014.2306412>).
7. Fossum, Eric R. (12 July 1993). "Active pixel sensors: Are CCDS dinosaurs?". In Blouke, Morley M. (ed.). *Charge-Coupled Devices and Solid State Optical Sensors III*. Vol. 1900. International Society for Optics and Photonics. pp. 2–14. Bibcode:1993SPIE.1900....2F (<http://ui.adsabs.harvard.edu/abs/1993SPIE.1900....2F>). CiteSeerX 10.1.1.408.6558 (<https://cite-seerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.408.6558>). doi:10.1117/12.148585 (<https://doi.org/10.1117%2F12.148585>). S2CID 10556755 (<https://api.semanticscholar.org/CorpusID:10556755>). {{cite book}}: |journal= ignored (help)
8. "The use of digital vs celluloid film on Hollywood movies" (<https://stephenfollows.com/digital-vs-film-on-hollywood-movies/>). *Stephen Follows*. 2019-02-11. Retrieved 2019-10-23.
9. "Robert Rodriguez Film Once Upon a Time in Mexico This is a structural review" (<http://www.writework.com/essay/robert-rodriguez-film-once-upon-time-mexico-structural-rev>). WriteWork. Retrieved 2013-04-22.
10. "Maybe the war between digital and film isn't a war at all" (<https://www.avclub.com/maybe-the-war-between-digital-and-film-isn-t-a-war-at-a-1828527569>). *The A.V. Club*. 23 August 2018. Retrieved 26 November 2019.
11. Rizov, Vadim (24 April 2019). "24 Films Shot on 35mm Released in 2018" (<https://filmmakermagazine.com/107353-23-films-35mm-released-in-2018/>). *Filmmaker Magazine*. Retrieved 2019-09-14.
12. "The Heart of a Phone Camera: The CMOS Active Pixel Image Sensor" (<http://large.stanford.edu/courses/2012/ph250/lu2/>). *large.stanford.edu*. Retrieved 2021-03-26.
13. Hanzo, Lajos (2007). *Video compression and communications: from basics to H.261, H.263, H.264, MPEG2, MPEG4 for DVB and HSDPA-style adaptive turbo-transceivers* (<https://www.worldcat.org/oclc/181368622>). Peter J. Cherriman, Jürgen Streit, Lajos Hanzo (2nd ed.). Hoboken, NJ: IEEE Press. ISBN 978-0-470-51992-9. OCLC 181368622 (<https://search.worldcat.org/oclc/181368622>).
14. "The History of Video File Formats Infographic" (<http://www.real.com/resources/digital-video-file-formats/>). *RealNetworks*. 22 April 2012. Retrieved 5 August 2019.
15. Ghanbari, Mohammed (2003). *Standard Codecs: Image Compression to Advanced Video Coding* (<https://books.google.com/books?id=7XuU8T3ooOAC&pg=PA1>). Institution of Engineering and Technology. pp. 1–2. ISBN 9780852967102.
16. Christ, Robert D. (2013). *The ROV manual : a user guide for remotely operated vehicles* (<https://www.worldcat.org/oclc/861797595>). Robert L. Wernli (2nd ed.). Oxford. ISBN 978-0-08-098291-5. OCLC 861797595 (<https://search.worldcat.org/oclc/861797595>).
17. Thomson, Gavin; Shah, Athar (2017). "Introducing HEIF and HEVC" (https://devstreaming-cdn.apple.com/videos/wwdc/2017/503i6plfvfi7o3222/503/503_introducing_heif_and_hevc.pdf) (PDF). Apple Inc. Retrieved 5 August 2019.
18. "HEVC Patent List" (<https://www.mpegla.com/wp-content/uploads/hevc-att1.pdf>) (PDF). *MPEG LA*. Retrieved 6 July 2019.
19. "Video Developer Report 2019" (<https://cdn2.hubspot.net/hubfs/3411032/Bitmovin%20Magazine/Video%20Developer%20Report%202019/bitmovin-video-developer-report-2019.pdf>) (PDF). *Bitmovin*. 2019. Retrieved 5 November 2019.
20. Roger, Jennings (1997). *Special Edition Using Desktop Video*. Que Books, Macmillan Computer Publishing. ISBN 978-0789702654.
21. "CoSA Lives: The Story of the Company Behind After Effects" (<http://www.motionworks.com.au/2009/11/cosa-lives/>). *Motionworks Digital Marketing Agency Melbourne*. Archived (<https://web.archive.org/web/20110227132422/http://motionworks.com.au/2009/11/cosa-lives/>) from the original on 2011-02-27. Retrieved 2009-11-16.

22. Garrett, Bradley L. (2018). "Videographic geographies: Using digital video for geographic research" (<http://journals.sagepub.com/doi/10.1177/0309132510388337>). *Progress in Human Geography*. **35** (4): 521–541. doi:[10.1177/0309132510388337](https://doi.org/10.1177/0309132510388337) (<https://doi.org/10.1177%2F0309132510388337>). ISSN 0309-1325 (<https://search.worldcat.org/issn/0309-1325>). S2CID 131426433 (<https://api.semanticscholar.org/CorpusID:131426433>).
23. Bruce, David L.; Chiu, Ming Ming (2015). "Composing With New Technology: Teacher Reflections on Learning Digital Video" (<http://journals.sagepub.com/doi/10.1177/0022487115574291>). *Journal of Teacher Education*. **66** (3): 272–287. doi:[10.1177/0022487115574291](https://doi.org/10.1177%2F0022487115574291) (<https://doi.org/10.1177%2F0022487115574291>). ISSN 0022-4871 (<https://search.worldcat.org/issn/0022-4871>). S2CID 145361658 (<https://api.semanticscholar.org/CorpusID:145361658>).
24. Wieler, Matthew E.; Murphy, Thomas G.; Blecherman, Mira; Mehta, Hiral; Bender, G. Jesse (2021-03-01). "Infant heart-rate measurement and oxygen desaturation detection with a digital video camera using imaging photoplethysmography" (<https://dx.doi.org/10.1038/s41372-021-00967-1>). *Journal of Perinatology*. **41** (7): 1725–1731. doi:[10.1038/s41372-021-00967-1](https://doi.org/10.1038/s41372-021-00967-1) (<https://doi.org/10.1038%2Fs41372-021-00967-1>). ISSN 0743-8346 (<https://search.worldcat.org/issn/0743-8346>). PMID 33649437 (<https://pubmed.ncbi.nlm.nih.gov/33649437>). S2CID 232070728 (<https://api.semanticscholar.org/CorpusID:232070728>).
25. Bruehs, Walter E.; Stout, Dorothy (2020). "Quantifying and Ranking Quality for Acquired Recordings on Digital Video Recorders" (<https://onlinelibrary.wiley.com/doi/10.1111/1556-4029.14307>). *Journal of Forensic Sciences*. **65** (4): 1155–1168. doi:[10.1111/1556-4029.14307](https://doi.org/10.1111/1556-4029.14307) (<https://doi.org/10.1111%2F1556-4029.14307>). ISSN 0022-1198 (<https://search.worldcat.org/issn/0022-1198>). PMID 32134510 (<https://pubmed.ncbi.nlm.nih.gov/32134510>). S2CID 212417006 (<https://api.semanticscholar.org/CorpusID:212417006>).
26. Kruger, Lennard G. (2002). *Digital television : an overview* (<https://www.worldcat.org/oclc/50684535>). Peter F. Guerrero. New York: Novinka Books. ISBN 1-59033-502-3. OCLC 50684535 (<https://search.worldcat.org/oclc/50684535>).
27. Reimers, U. (1998). "Digital video broadcasting" (<https://ieeexplore.ieee.org/document/685371>). *IEEE Communications Magazine*. **36** (6): 104–110. doi:[10.1109/35.685371](https://doi.org/10.1109/35.685371) (<https://doi.org/10.1109%2F35.685371>).
28. "The Origins and Future Prospects of Digital Television" (http://www.benton.org/initiatives/obligations/charting_the_digital_broadcasting_future/sec1). Benton Foundation. 2008-12-23.
29. Barbero, M.; Hofmann, H.; Wells, N. D. (14 November 1991). "DCT source coding and current implementations for HDTV" (https://tech.ebu.ch/publications/trev_251-barbero). *EBU Technical Review* (251). European Broadcasting Union: 22–33. Retrieved 4 November 2019.
30. "NextLevel signs cable deal - Dec. 17, 1997" (<https://money.cnn.com/1997/12/17/technology/nextlevel/>). *money.cnn.com*. Retrieved 9 August 2018.
31. "TCI faces big challenges - Aug. 15, 1996" (https://money.cnn.com/1996/08/15/companies/tpci_pkg/). *money.cnn.com*. Retrieved 9 August 2018.
32. Barbero, M.; Stroppiana, M. (October 1992). "Data compression for HDTV transmission and distribution" (<https://ieeexplore.ieee.org/document/193745>). *IEE Colloquium on Applications of Video Compression in Broadcasting*: 10/1–10/5.
33. Winkelman, Roy (2018). "TechEase, What is bit depth?" (<https://etc.usf.edu/techease/win/images/what-is-bit-depth/>). Retrieved 2022-04-18.
34. Steiner, Shawn (12 December 2018). "B&H, 8-Bit, 10-Bit, What Does It All Mean for Your Videos?" (<https://www.bhphotovideo.com/explora/video/tips-and-solutions/8-bit-10-bit-what-does-it-all-mean-your-videos>).
35. Acharya, Tinku (2005). *JPEG2000 standard for image compression: concepts, algorithms and VLSI architectures* (<https://www.worldcat.org/oclc/57585202>). Ping-Sing Tsai. Hoboken, N.J.: Wiley-Interscience. ISBN 0-471-65375-6. OCLC 57585202 (<https://search.worldcat.org/oclc/57585202>).

36. Weise, Marcus (2013). *How video works* (<https://www.worldcat.org/oclc/1295602475>). Diana Weynand (2nd ed.). New York. ISBN [978-1-136-06982-6](#). OCLC [1295602475](#) (<https://search.worldcat.org/oclc/1295602475>).
 37. Vatolin, Dmitriy. "Lossless Video Codecs Comparison 2007" (https://www.compression.ru/video/codec_comparison/lossless_codecs_2007_en.html). www.compression.ru. Retrieved 2022-03-29.
- Sadun, Erica (2006). *Digital Video Essentials : Shoot, Transfer, Edit, Share* (<https://www.worldcat.org/oclc/630529114>). Hoboken: John Wiley & Sons. ISBN [978-0-470-11319-6](#). OCLC [630529114](#) (<https://search.worldcat.org/oclc/630529114>).
 - "Digital Video (DV)" (<https://www.techopedia.com/definition/5505/digital-video-dv>). Techopedia. December 16, 2016. Retrieved March 23, 2021.

External links

- The DV, DVCAM, & DVCPRO Formats – tech details, FAQ, and links (<http://www.adamwilt.com/DV.html>)
- Standard digital TV and video formats. (<https://web.archive.org/web/20161021230910/http://www.equasys.de/videoformats.html>)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Digital_video&oldid=1291822726"



Cinematography

Cinematography (from Ancient Greek κίνημα (*kínēma*) 'movement' and γράφειν (*gráphein*) 'to write, draw, paint, etc.') is the art of motion picture (and more recently, electronic video camera) photography.

Cinematographers use a lens to focus reflected light from objects into a real image that is transferred to some image sensor or light-sensitive material inside the movie camera.^[1] These exposures are created sequentially and preserved for later processing and viewing as a motion picture. Capturing images with an electronic image sensor produces an electrical charge for each pixel in the image, which is electronically processed and stored in a video file for subsequent processing or display. Images captured with photographic emulsion result in a series of invisible latent images on the film stock, which are chemically "developed" into a visible image. The images on the film stock are projected for viewing in the same motion picture.

Cinematography finds uses in many fields of science and business, as well as for entertainment purposes and mass communication.

History

Precursors

In the 1830s, three different solutions for moving images were invented based on the concept of revolving drums and disks, the stroboscope by Simon von Stampfer in Austria, the phenakistoscope by Joseph Plateau in Belgium, and the zoetrope by William Horner in Britain.

In 1845, Francis Ronalds invented the first successful camera able to make continuous recordings of the varying indications of meteorological and geomagnetic instruments over time. The cameras were supplied to numerous observatories around the world and some remained in use until well into the 20th century.^{[2][3][4]}

William Lincoln patented a device, in 1867 that showed animated pictures called the "wheel of life" or "zoopraxiscope". In it moving drawings or photographs were watched through a slit.



Arri Alexa, a digital movie camera



An Eadweard Muybridge sequence of a horse galloping

On 19 June 1878, Eadweard Muybridge successfully photographed a horse named "Sallie Gardner" in fast motion using a series of 24 stereoscopic cameras. The cameras were arranged along a track parallel to the horse's, and each camera shutter was controlled by a trip wire triggered by the horse's hooves. They were 21 inches apart to cover the 20 feet taken by the horse stride, taking pictures at one-thousandth of a second.^[5] At the end of the decade, Muybridge had adapted sequences of his photographs to a zoopraxiscope for short, primitive projected "movies", which were sensations on his lecture tours by 1879 or 1880.

Four years later, in 1882, French scientist Étienne-Jules Marey invented a chronophotographic gun, which was capable of taking 12 consecutive frames a second and recording all the frames of the same picture.

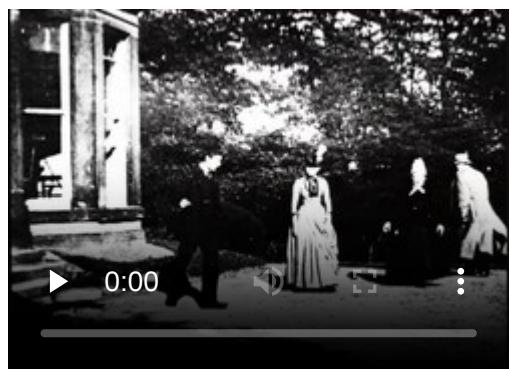
The late nineteenth to the early twentieth centuries brought rise to the use of film not only for entertainment purposes but for scientific exploration as well. French biologist and filmmaker Jean Painleve lobbied heavily for the use of film in the scientific field, as the new medium was more efficient in capturing and documenting the behavior, movement, and environment of microorganisms, cells, and bacteria, than the naked eye.^[6] The introduction of film into scientific fields allowed for not only the viewing of "new images and objects, such as cells and natural objects, but also the viewing of them in real time",^[6] whereas prior to the invention of moving pictures, scientists and doctors alike had to rely on hand-drawn sketches of human anatomy and its microorganisms. This posed a great inconvenience to the scientific and medical worlds. The development of film and increased usage of cameras allowed doctors and scientists to grasp a better understanding and knowledge of their projects.

The origins of today's cinema go back to the Lumière brothers, Auguste and Louis, who in 1895 developed a machine called the Cinematographe, which had the ability to capture and show moving images. The early era of cinema saw rapid innovation. In the early-to-mid-20th Century, filmmakers discovered and applied new methods such as editing, special effects, close-ups, sound, widescreen, color films, and more. Hollywood began to emerge as the Mecca of the film industry, and many of the famous studios from that time still exist over 100 years later in the early-mid 21st Century.

Film

The experimental film Roundhay Garden Scene, filmed by Louis Le Prince in Roundhay, Leeds, England, on October 14, 1888, is the earliest surviving motion picture.^[7] This movie was shot on paper film.^[8]

An experimental film camera was developed by British inventor William Friese Greene and patented in 1889.^[9] W. K. L. Dickson, working under the direction of Thomas Alva Edison, was the first to design a successful apparatus, the Kinetograph,^[10] patented in 1891.^[11] This camera took a series of instantaneous photographs on standard Eastman Kodak photographic emulsion coated onto a transparent celluloid strip 35 mm wide. The results of this work were first shown in public in 1893, using the viewing apparatus also designed by Dickson, the Kinetoscope. Contained within a large box, only one person at a time looking into it through a peephole could view the movie.



Roundhay Garden Scene (1888), the world's earliest surviving motion-picture film

In the following year, Charles Francis Jenkins and his projector, the Phantoscope,^[12] made a successful audience viewing while Louis and Auguste Lumière perfected the Cinématographe, an apparatus that took, printed, and projected film, in Paris in December 1895.^[13] The Lumière brothers were the first to present projected, moving, photographic, pictures to a paying audience of more than one person.

In 1896, movie theaters were open in France (Paris, Lyon, Bordeaux, Nice, Marseille); Italy (Rome, Milan, Naples, Genoa, Venice, Bologna, Forlì); Brussels; and London. The chronological improvements in the medium may be listed concisely. In 1896, Edison showed his improved Vitascope projector, the first commercially successful projector in the U.S. Cooper Hewitt invented mercury lamps which made it practical to shoot films indoors without sunlight in 1905. The first animated cartoon was produced in 1906. Credits began to appear at the beginning of motion pictures in 1911. The Bell and Howell 2709 movie camera invented in 1915 allowed directors to make close-ups without physically moving the camera. By the late 1920s, most of the movies produced were sound films. Wide screen formats were first experimented within the 1950s. By the 1970s, most movies were color films. IMAX and other 70mm formats gained popularity. Wide distribution of films became commonplace, setting the ground for "blockbusters." Film cinematography dominated the motion picture industry from its inception until the 2010s when digital cinematography became dominant. Film cinematography is still used by some directors, especially in specific applications or out of fondness for the format.

Black and white

From its birth in the 1880s, movies were predominantly monochrome. Contrary to popular belief, monochrome does not always mean black-and-white; it means a movie shot in a single tone or color. Since the cost of tinted film bases was substantially higher, most movies were produced in black-and-white monochrome. Even with the advent of early color experiments, the greater expense of color meant films were mostly made in black-and-white until the 1950s, when cheaper color processes were introduced, and in some years percentage of films shot on color film surpassed 51%. By the 1960s, color became by far the dominant film stock. In the coming decades, the usage of color film greatly increased while monochrome films became scarce.

Black-and-white cinematography is a technique used in filmmaking where the images are captured and presented in shades of gray, without color. This artistic approach has a rich history and has been employed in various films throughout cinema's evolution. It is a powerful tool that allows filmmakers to emphasize contrast, texture, and lighting, enhancing the visual storytelling experience. The use of black-and-white cinematography dates back to the early days of cinema when color film was not yet available. Filmmakers relied on this technique to create visually striking and atmospheric films. Even with the advent of color film technology, black-and-white cinematography continued to be utilized for artistic and thematic purposes.

Ken Dancyger's book *The Technique of Film and Video Editing: History, Theory, and Practice* provides valuable insights into the historical and theoretical aspects of black-and-white cinematography. Dancyger explores how this technique has been employed throughout film history, examining its impact on storytelling, mood, and visual aesthetics. The book delves into the artistic choices and technical considerations involved in creating compelling black-and-white imagery, offering a comprehensive understanding of the technique.

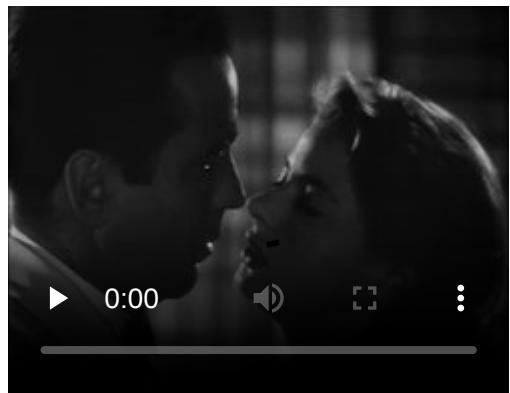
Black-and-white cinematography allows filmmakers to focus on the interplay of light and shadow, emphasizing the contrast between different elements within a scene. This technique can evoke a sense of nostalgia, evoke a specific time period, or create a timeless and classic feel. By stripping away color, filmmakers can emphasize the composition, shapes, and textures within the frame, enhancing the visual impact. Notable films that have employed black-and-white cinematography include classics such as *Casablanca* (1942), *Raging Bull* (1980), and *Schindler's List* (1993). These films showcase the power and versatility of black-and-white cinematography in creating emotionally resonant visuals. Black-and-white cinematography remains a relevant and widely used technique in modern filmmaking. It continues to be employed by filmmakers to evoke specific moods, convey a sense of timelessness, and enhance the artistic expression of their stories.

Color

After the advent of motion pictures, a tremendous amount of energy was invested in the production of photography in natural color.^[14] The invention of the talking picture further increased the demand for the use of color photography. However, in comparison to other technological advances of the time, the arrival of color photography was a relatively slow process.^[15]

Early movies were not actually color movies since they were shot monochrome and hand-colored or machine-colored afterward (such movies are referred to as *colored* and not *color*). The earliest such example is the hand-tinted Annabelle Serpentine Dance in 1895 by Edison Manufacturing Company. Machine-based tinting later became popular. Tinting continued until the advent of natural color cinematography in the 1910s. Many black-and-white movies have been colorized recently using digital tinting. This includes footage shot from both world wars, sporting events and political propaganda.

In 1902, Edward Raymond Turner produced the first films with a natural color process rather than using colorization techniques.^[16] In 1909, Kinemacolor was first shown to the public.^[17]



Casablanca (1942) Trailer, which showcases the Iconic "Kiss Me" Scene. Featuring actors Humphrey Bogart and Ingrid Bergman. Casablanca is one of many films to utilize Black and White cinematography as a technique to create atmospheric scenes throughout the movie. For instance, the "Kiss Me" scene, depicts two characters under shadows, soft lighting, and contrast to create a sense of longing and emotional intensity between the two characters. The absence of color allows viewers to focus more on the emotion that is being conveyed during the scene. In addition, the absence of color makes the actor and actress in this scene have more defined facial expressions, drawing attention to the deep emotion between Rick Blain (Humphrey Bogart) and Ilsa Lund (Ingrid Bergman).



Annabelle Serpentine Dance, hand-tinted version (1895)

In 1917, the earliest version of Technicolor was introduced. Kodachrome was introduced in 1935. Eastmancolor was introduced in 1950 and became the color standard for the rest of the century.

In the 2010s, color films were largely superseded by color digital cinematography.

Digital video

In digital cinematography, the movie is shot on digital media such as flash storage, as well as distributed through a digital medium such as a hard drive.

The basis for digital cameras are metal–oxide–semiconductor (MOS) image sensors.^[18] The first practical semiconductor image sensor was the charge-coupled device (CCD),^[19] based on MOS capacitor technology.^[18] Following the commercialization of CCD sensors during the late 1970s to early 1980s, the entertainment industry slowly began transitioning to digital imaging and digital video over the next two decades.^[20] The CCD was followed by the CMOS active-pixel sensor (CMOS sensor),^[21] developed in the 1990s.^{[22][23]}

Beginning in the late 1980s, Sony began marketing the concept of "electronic cinematography", utilizing its analog Sony HDVS professional video cameras. The effort met with very little success. However, this led to one of the earliest digitally shot feature movies, Julia and Julia (1987). In 1998, with the introduction of HDCAM recorders and 1920×1080 pixel digital professional video cameras based on CCD technology, the idea, now re-branded as "digital cinematography", began to gain traction.

Shot and released in 1998, The Last Broadcast is believed by some to be the first feature-length video shot and edited entirely on consumer-level digital equipment.^[24] In May 1999, George Lucas challenged the supremacy of the movie-making medium of film for the first time by including footage filmed with high-definition digital cameras in Star Wars: Episode I – The Phantom Menace. In late 2013, Paramount became the first major studio to distribute movies to theaters in digital format, eliminating 35mm film entirely. Since then the demand of movies to be developed onto digital format rather than 35mm has increased significantly.

As digital technology improved, movie studios began increasingly shifting toward digital cinematography. Since the 2010s, digital cinematography has become the dominant form of cinematography after largely superseding film cinematography.

Aspects

Numerous aspects contribute to the art of cinematography, including:

Cinema technique

The first film cameras were fastened directly to the head of a tripod or other support, with only the crudest kind of leveling devices provided, in the manner of the still-camera tripod heads of the period. The earliest film cameras were thus effectively fixed during the shot, and hence the first camera movements were the result of mounting a camera on a moving vehicle. The first known of these was a film shot by a Lumière cameraman from the back platform of a train leaving Jerusalem in 1896, and by 1898, there were a number of films shot from moving trains. Although listed under the general heading of "panoramas" in the sales catalogues of the time, those films shot straight forward from in front of a railway engine were usually specifically referred to as "phantom rides".

In 1897, Robert W. Paul had the first real rotating camera head made to put on a tripod, so that he could follow the passing processions of Queen Victoria's Diamond Jubilee in one uninterrupted shot. This device had the camera mounted on a vertical axis that could be rotated by a worm gear driven by turning a crank handle, and Paul put it on general sale the next year. Shots taken using such a "panning" head were also referred to as "panoramas" in the film catalogues of the first decade of the cinema. This eventually led to the creation of a panoramic photo as well.



Georges Méliès (left) painting a backdrop in his studio

The standard pattern for early film studios was provided by the studio which Georges Méliès had built in 1897. This had a glass roof and three glass walls constructed after the model of large studios for still photography, and it was fitted with thin cotton cloths that could be stretched below the roof to diffuse the direct ray of the sun on sunny days. The soft overall light without real shadows that this arrangement produced, which also exists naturally on lightly overcast days, was to become the basis for film lighting in film studios for the next decade.

Black-and-white cinematography is a technique used in filmmaking where the images are captured and presented in shades of gray, without color. This artistic approach has a rich history and has been employed in various films throughout cinema's evolution. It is a powerful tool that allows filmmakers to emphasize contrast, texture, and lighting, enhancing the visual storytelling experience. The use of black-and-white cinematography dates back to the early days of cinema when color film was not yet available. Filmmakers relied on this technique to create visually striking and atmospheric films. Even with the advent of color film technology, black-and-white cinematography continued to be utilized for artistic and thematic purposes. Ken Dancyger's book *The Technique of Film and Video Editing: History, Theory, and Practice* provides valuable insights into the historical and theoretical aspects of black-and-white cinematography. Dancyger explores how this technique has been employed throughout film history, examining its impact on storytelling, mood, and visual aesthetics. The book delves into the artistic choices and technical considerations involved in creating compelling black-and-white imagery, offering a comprehensive understanding of the technique.

Black-and-white cinematography allows filmmakers to focus on the interplay of light and shadow, emphasizing the contrast between different elements within a scene. This technique can evoke a sense of nostalgia, evoke a specific time period, or create a timeless and classic feel. By stripping away color, filmmakers can emphasize the composition, shapes, and textures within the frame, enhancing the visual impact. Notable films that have employed black-and-white cinematography include classics such as *Casablanca* (1942), *Raging Bull* (1980), and *Schindler's List* (1993). These films showcase the power and versatility of black-and-white cinematography in creating emotionally resonant visuals. Black-and-white cinematography remains a relevant and widely used technique in modern filmmaking. It continues to be employed by filmmakers to evoke specific moods, convey a sense of timelessness, and enhance the artistic expression of their stories.

There are many types of Cinematography that each differ based on production purpose and process. These different types of Cinematography are similar in the sense that they all have the goal of conveying a specific emotion, mood or feeling. For each different style however they can often convey different emotions and purposes. Some examples of different types of Cinematography can be known as Realism. This style of cinematography aims to create a realistic portrayal of the world, often using natural lighting, handheld cameras, and a documentary-like approach to filming. Classic Hollywood is a style of

cinematography characterized by its use of highly polished, studio-produced films with glamorous sets, bright lighting, and romanticized narratives. Film Noir is a style of cinematography that is characterized by its use of stark contrast and chiaroscuro lighting, low-key lighting, and a dark, brooding atmosphere. It often features crime, mystery, and morally ambiguous characters.

Aspects of a cinema that affect a film

To convey mood, emotion, narrative and other factors within the shot, cinematography is implemented by using different aspects within a film. Lighting on the scene can affect the mood of a scene or film. Darker shots with less natural light can be gloomy, scary, sad, intense. Brighter lighting can equate to a happier, exciting, more positive mood. Camera angle can affect a scene by setting perspective. It conveys how characters or the audience see something, and through what angle. Camera angle can also play an important role by highlighting either a close up detail, or background setting. A close up angle can highlight detail on someone's face, while a wider lens can give key information that takes place in the background of a shot. Camera distance can highlight specific details that can be important to a film shot. From very far away, a group of people can all look the same, but once you zoom in very close, the viewer is able to see differences within the population through details like facial expression and body language. Coloring is similar to lighting, in the way that it plays an important role in setting mood and emotion throughout a shot. A color like green can convey balance and peace through scenes of nature. A shot with a lot of red can express anger, intensity, passion or love. While some of these emotions might not come out intentionally while seeing color, it is a subconscious fact that color within cinematography can have a large effect. Speed is a vital element in cinematography that can be used in a variety of ways, such as the creation of action, or a sense of movement. Speed can be further used to slow down time, highlight important moments, and oftentimes build a sense of suspense in a film. Slow motion is a technique which involves filming at a higher frame rate, then playing the footage again at a normal speed. This creates a slowed-down effect in the film, which can put emphasis on or add fluidity to a scene. On the other hand, fast motion is the opposite of slow motion, filming at a lower frame rate and then playing the film back at a normal speed. This creates a sped-up effect which can help to emphasize passage of time, or create a sense of urgency. Time lapse is when you take a series of still photographs at a regular interval over a long period of time. From here, if you play them back continuously, a sped-up effect is shown. Time lapses are used most effectively to show things like sunrises, natural movement, or growth. They are commonly used to show passage of time in a shorter sequence. Reverse motion is filming a scene normally, then playing the film in reverse. This is usually used to create uncommon/surreal effects, and create unusual scenes. The various techniques involving speed all can add to a film's intensity, vibe, show passage of time, and have many other effects. Camera movement within a film can play a role in enhancing the visual quality and impact of a film. Some aspects of camera movement that contribute to this are:

- Zooming: This movement involves changing the focal length of the lens to make the subject appear closer or farther away. It can be used to create a sense of intimacy or distance from the subject.
- Tilt: Rotating the camera vertically from a fixed position. It can be used to show the height of a subject or to emphasize a particular element in the scene.
- Panning: Rotating the camera horizontally from a fixed position. It can be used to follow a moving subject or to show a wide view of a scene.
- Pedestal/Booming/Jibbing: Moving a camera vertically in its entirety. This can be used to show vertical movement relative to a subject in a frame

- Trucking: Moving a camera horizontally in its entirety. This can be used to show horizontal movement relative to a subject in a frame
- Rolling: Rotating a camera in its entirety in a horizontal manner.

Image sensor and film stock

Cinematography can begin with digital image sensor or rolls of film. Advancements in film emulsion and grain structure provided a wide range of available film stocks. The selection of a film stock is one of the first decisions made in preparing a typical film production.

Aside from the film gauge selection – 8 mm (amateur), 16 mm (semi-professional), 35 mm (professional) and 65 mm (epic photography, rarely used except in special event venues) – the cinematographer has a selection of stocks in reversal (which, when developed, create a positive image) and negative formats along with a wide range of film speeds (varying sensitivity to light) from ISO 50 (slow, least sensitive to light) to 800 (very fast, extremely sensitive to light) and differing response to color (low saturation, high saturation) and contrast (varying levels between pure black (no exposure) and pure white (complete overexposure)). Advancements and adjustments to nearly all gauges of film create the "super" formats wherein the area of the film used to capture a single frame of an image is expanded, although the physical gauge of the film remains the same. Super 8 mm, Super 16 mm, and Super 35 mm all utilize more of the overall film area for the image than their "regular" non-super counterparts. The larger the film gauge, the higher the overall image resolution clarity and technical quality. The techniques used by the film laboratory to process the film stock can also offer a considerable variance in the image produced. By controlling the temperature and varying the duration in which the film is soaked in the development chemicals, and by skipping certain chemical processes (or partially skipping all of them), cinematographers can achieve very different looks from a single film stock in the laboratory. Some techniques that can be used are push processing, bleach bypass, and cross processing.

Most of modern cinema uses digital cinematography and has no film stocks , but the cameras themselves can be adjusted in ways that go far beyond the abilities of one particular film stock. They can provide varying degrees of color sensitivity, image contrast, light sensitivity and so on. One camera can achieve all the various looks of different emulsions. Digital image adjustments such as ISO and contrast are executed by estimating the same adjustments that would take place if actual film were in use, and are thus vulnerable to the camera's sensor designers perceptions of various film stocks and image adjustment parameters.

Filters

Filters, such as diffusion filters or color effect filters, are also widely used to enhance mood or dramatic effects. Most photographic filters are made up of two pieces of optical glass glued together with some form of image or light manipulation material between the glass. In the case of color filters, there is often a translucent color medium pressed between two planes of optical glass. Color filters work by blocking out certain color wavelengths of light from reaching the film. With color film, this works very intuitively wherein a blue filter will cut down on the passage of red, orange, and yellow light and create a blue tint on the film. In black-and-white photography, color filters are used somewhat counter-intuitively; for instance, a yellow filter, which cuts down on blue wavelengths of light, can be used to darken a daylight

sky (by eliminating blue light from hitting the film, thus greatly underexposing the mostly blue sky) while not biasing most human flesh tone. Filters can be used in front of the lens or, in some cases, behind the lens for different effects.

Certain cinematographers, such as Christopher Doyle, are well known for their innovative use of filters; Doyle was a pioneer for increased usage of filters in movies and is highly respected throughout the cinema world.

Lens

Lenses can be attached to the camera to give a certain look, feel, or effect by focus, color, etc. As does the human eye, the camera creates perspective and spatial relations with the rest of the world. However, unlike one's eye, a cinematographer can select different lenses for different purposes. Variation in focal length is one of the chief benefits. The focal length of the lens determines the angle of view and, therefore, the field of view. Cinematographers can choose from a range of wide-angle lenses, "normal" lenses and long focus lenses, as well as macro lenses and other special effect lens systems such as borescope lenses. Wide-angle lenses have short focal lengths and make spatial distances more obvious. A person in the distance is shown as much smaller while someone in the front will loom large. On the other hand, long focus lenses reduce such exaggerations, depicting far-off objects as seemingly close together and flattening perspective. The differences between the perspective rendering is actually not due to the focal length by itself, but by the distance between the subjects and the camera. Therefore, the use of different focal lengths in combination with different camera to subject distances creates these different renderings. Changing the focal length only while keeping the same camera position does not affect perspective but the camera angle of view only.

A zoom lens allows a camera operator to change his focal length within a shot or quickly between setups for shots. As prime lenses offer greater optical quality and are "faster" (larger aperture openings, usable in less light) than zoom lenses, they are often employed in professional cinematography over zoom lenses. Certain scenes or even types of filmmaking, however, may require the use of zooms for speed or ease of use, as well as shots involving a zoom move.

As in other photography, the control of the exposed image is done in the lens with the control of the diaphragm aperture. For proper selection, the cinematographer needs that all lenses be engraved with T-stop, not f-stop so that the eventual light loss due to the glass does not affect the exposure control when setting it using the usual meters. The choice of the aperture also affects image quality (aberrations) and depth of field.

Depth of field and focus

Focal length and diaphragm aperture affect the depth of field of a scene – that is, how much the background, mid-ground and foreground will be rendered in "acceptable focus" (only one exact plane of the image is in precise focus) on the film or video target. Depth of field (not to be confused with depth of focus) is determined by the aperture size and the focal distance. A large or deep depth of field is generated



Live recording for TV on a camera with a Fujinon optical lens.

with a very small iris aperture and focusing on a point in the distance, whereas a shallow depth of field will be achieved with a large (open) iris aperture and focusing closer to the lens. Depth of field is also governed by the format size. If one considers the field of view and angle of view, the smaller the image is, the shorter the focal length should be, as to keep the same field of view. Then, the smaller the image is, the more depth of field is obtained, for the same field of view. Therefore, 70mm has less depth of field than 35mm for a given field of view, 16mm more than 35mm, and early video cameras, as well as most modern consumer level video cameras, even more depth of field than 16mm.

In *Citizen Kane* (1941), cinematographer Gregg Toland and director Orson Welles used tighter apertures to create every detail of the foreground and background of the sets in sharp focus. This practice is known as deep focus. Deep focus became a popular cinematographic device from the 1940s onward in Hollywood. Today, the trend is for more shallow focus. To change the plane of focus from one object or character to another within a shot is commonly known as a rack focus.

Early in the transition to digital cinematography, the inability of digital video cameras to easily achieve shallow depth of field, due to their small image sensors, was initially an issue of frustration for film makers trying to emulate the look of 35mm film. Optical adapters were devised which accomplished this by mounting a larger format lens which projected its image, at the size of the larger format, on a ground glass screen preserving the depth of field. The adapter and lens then mounted on the small format video camera which in turn focused on the ground glass screen.

Digital SLR still cameras have sensor sizes similar to that of the 35mm film frame, and thus are able to produce images with similar depth of field. The advent of video functions in these cameras sparked a revolution in digital cinematography, with more and more film makers adopting still cameras for the purpose because of the film-like qualities of their images. More recently, more and more dedicated video cameras are being equipped with larger sensors capable of 35mm film-like depth of field.

Aspect ratio and framing

The aspect ratio of an image is the ratio of its width to its height. This can be expressed either as a ratio of 2 integers, such as 4:3, or in a decimal format, such as 1.33:1 or simply 1.33. Different ratios provide different aesthetic effects. Standards for aspect ratio have varied significantly over time.

During the silent era, aspect ratios varied widely, from square 1:1, all the way up to the extreme widescreen 4:1 Polyvision. However, from the 1910s, silent motion pictures generally settled on the ratio of 4:3 (1.33). The introduction of sound-on-film briefly narrowed the aspect ratio, to allow room for a sound stripe. In 1932, a new standard was introduced, the Academy ratio of 1.37, by means of thickening the frame line.

For years, mainstream cinematographers were limited to using the academy ratio, but in the 1950s, thanks to the popularity of Cinerama, widescreen ratios were introduced in an effort to pull audiences back into the theater and away from their home television sets. These new widescreen formats provided



A deep focus shot from *Citizen Kane* (1941): everything, including the hat in the foreground and the boy (young Charles Foster Kane) in the distance, is in sharp focus.

cinematographers a wider frame within which to compose their images.

Many different proprietary photographic systems were invented and used in the 1950s to create widescreen movies, but one dominated film: the anamorphic process, which optically squeezes the image to photograph twice the horizontal area to the same size vertical as standard "spherical" lenses. The first commonly used anamorphic format was CinemaScope, which used a 2.35 aspect ratio, although it was originally 2.55. CinemaScope was used from 1953 to 1967, but due to technical flaws in the design and its ownership by Fox, several third-party companies, led by Panavision's technical improvements in the 1950s, dominated the anamorphic cine lens market. Changes to SMPTE projection standards altered the projected ratio from 2.35 to 2.39 in 1970, although this did not change anything regarding the photographic anamorphic standards; all changes in respect to the aspect ratio of anamorphic 35 mm photography are specific to camera or projector gate sizes, not the optical system. After the "widescreen wars" of the 1950s, the motion-picture industry settled into 1.85 as a standard for theatrical projection in the United States and the United Kingdom. This is a cropped version of 1.37. Europe and Asia opted for 1.66 at first, although 1.85 has largely permeated these markets in recent decades. Certain "epic" or adventure movies utilized the anamorphic 2.39 (often incorrectly denoted '2.40')

In the 1990s, with the advent of high-definition video, television engineers created the 1.78 (16:9) ratio as a mathematical compromise between the theatrical standard of 1.85 and television's 1.33, as it was not practical to produce a traditional CRT television tube with a width of 1.85. Until that change, nothing had ever been originated in 1.78. Today, this is a standard for high-definition video and for widescreen television.

Lighting

Light is necessary to create an image exposure on a frame of film or on a digital target (CCD, etc.). The art of lighting for cinematography goes far beyond basic exposure, however, into the essence of visual storytelling. Lighting contributes considerably to the emotional response an audience has watching a motion picture. The increased usage of filters can greatly impact the final image and affect the lighting.

Importance of Lighting in Film Lighting in film is essential for three primary reasons: visibility, composition, and mood. Firstly, lighting ensures that the subject or scene is properly illuminated, allowing viewers to perceive the details and understand the narrative. It helps in guiding the audience's attention to specific elements within the frame, highlighting important characters or objects. Secondly, lighting contributes to the composition of a shot. Filmmakers strategically place lights to create balance, depth, and visual interest within the frame. It allows them to control the visual elements within the scene, emphasizing certain areas and de-emphasizing others. Lastly, lighting significantly impacts the mood and atmosphere of a film. By manipulating light intensity, color, and direction, filmmakers can evoke different emotions and enhance the narrative. Bright, even lighting may evoke a sense of safety and happiness, while low-key lighting with shadows can create tension, mystery, or fear. The choice of lighting style can also reflect the genre of the film, such as the high contrast lighting commonly used in film noir.

Lighting Techniques

Numerous lighting techniques are employed in filmmaking to achieve desired effects. Here are some commonly used techniques:

- Three-Point Lighting:** This classic technique involves the use of three lights: the key light, fill light, and backlight. The key light serves as the primary source, illuminating the subject

from one side to create depth and dimension. The fill light reduces shadows caused by the key light, softening the overall lighting. The backlight separates the subject from the background, providing a halo effect and enhancing the sense of depth. High Key Lighting: High key lighting produces a bright, evenly lit scene, often used in comedies or light-hearted films. It minimizes shadows, creating a cheerful and upbeat atmosphere. Low Key Lighting: Low key lighting involves using a single key light or a few strategically placed lights to create strong contrasts and deep shadows. This technique is commonly used in film noir and horror genres to evoke suspense, mystery, or fear.

Natural Lighting

Filmmakers sometimes employ natural lighting to create an authentic, realistic look. This technique utilizes existing light sources, such as sunlight or practical lamps, without additional artificial lighting. It is often seen in outdoor scenes or films aiming for a naturalistic aesthetic. Color Lighting: The use of colored lights or gels can dramatically alter the mood and atmosphere of a scene. Different colors evoke different emotions and can enhance storytelling. For example, warm tones like red or orange may create a sense of warmth or passion, while cool tones like blue can convey sadness or isolation.

Camera movement

Cinematography can not only depict a moving subject but can use a camera, which represents the audience's viewpoint or perspective, that moves during the course of filming. This movement plays a considerable role in the emotional language of film images and the audience's emotional reaction to the action. Techniques range from the most basic movements of panning (horizontal shift in viewpoint from a fixed position; like turning your head side-to-side) and tilting (vertical shift in viewpoint from a fixed position; like tipping your head back to look at the sky or down to look at the ground) to dollying (placing the camera on a moving platform to move it closer or farther from the subject), tracking (placing the camera on a moving platform to move it to the left or right), craning (moving the camera in a vertical position; being able to lift it off the ground as well as swing it side-to-side from a fixed base position), and combinations of the above. Early cinematographers often faced problems that were not common to other graphic artists because of the element of motion.^[25]



Camera on a small motor vehicle representing a large one

Cameras have been mounted to nearly every imaginable form of transportation. Most cameras can also be handheld, that is held in the hands of the camera operator who moves from one position to another while filming the action. Personal stabilizing platforms came into being in the late 1970s through the invention of Garrett Brown, which became known as the Steadicam. The Steadicam is a body harness and stabilization arm that connects to the camera, supporting the camera while isolating it from the operator's body movements. After the Steadicam patent expired in the early 1990s, many other companies began manufacturing their concept of the personal camera stabilizer. This invention is much more common throughout the cinematic world today. From feature-length films to the evening news, more and more networks have begun to use a personal camera stabilizer.

Special effects

The first special effects in the cinema were created while the film was being shot. These came to be known as "in-camera" effects. Later, optical and digital effects were developed so that editors and visual effects artists could more tightly control the process by manipulating the film in post-production.

The 1896 movie The Execution of Mary Stuart shows an actor dressed as the queen placing her head on the execution block in front of a small group of bystanders in Elizabethan dress. The executioner brings his axe down, and the queen's severed head drops onto the ground. This trick was worked by stopping the camera and replacing the actor with a dummy, then restarting the camera before the axe falls. The two pieces of film were then trimmed and cemented together so that the action appeared continuous when the film was shown, thus creating an overall illusion and successfully laying the foundation for special effects.

This film was among those exported to Europe with the first Kinetoscope machines in 1895 and was seen by Georges Méliès, who was putting on magic shows in his Théâtre Robert-Houdin in Paris at the time. He took up filmmaking in 1896, and after making imitations of other films from Edison, Lumière, and Robert Paul, he made Escamotage d'un dame chez Robert-Houdin (The Vanishing Lady). This film shows a woman being made to vanish by using the same stop motion technique as the earlier Edison film. After this, Georges Méliès made many single shot films using this trick over the next couple of years.

Double exposure

The other basic technique for trick cinematography involves double exposure of the film in the camera, which was first done by George Albert Smith in July 1898 in the UK. Smith's The Corsican Brothers (1898) was described in the catalogue of the Warwick Trading Company, which took up the distribution of Smith's films in 1900, thus:

"One of the twin brothers returns home from shooting in the Corsican mountains, and is visited by the ghost of the other twin. By extremely careful photography the ghost appears *quite transparent*. After indicating that he has been killed by a sword-thrust, and appealing for vengeance, he disappears. A 'vision' then appears showing the fatal duel in the snow. To the Corsican's amazement, the duel and death of his brother are vividly depicted in the vision, and overcome by his feelings, he falls to the floor just as his mother enters the room."



A scene inset inside a circular vignette showing a "dream vision" in Santa Claus (1898).

The ghost effect was done by draping the set in black velvet after the main action had been shot, and then re-exposing the negative with the actor playing the ghost going through the actions at the appropriate part. Likewise, the vision, which appeared within a circular vignette or matte, was similarly superimposed

over a black area in the backdrop to the scene, rather than over a part of the set with detail in it, so that nothing appeared through the image, which seemed quite solid. Smith used this technique again in *Santa Claus* (1898).

Georges Méliès first used superimposition on a dark background in *La Caverne maudite (The Cave of the Demons)* made a couple of months later in 1898, and elaborated it with many superimpositions in the one shot in *Un Homme de têtes (The Four Troublesome Heads)*. He created further variations in subsequent films.

Frame rate selection

Motion picture images are presented to an audience at a constant speed. In the theater it is 24 frames per second, in NTSC (US) Television it is 30 frames per second (29.97 to be exact), in PAL (Europe) television it is 25 frames per second. This speed of presentation does not vary.

However, by varying the speed at which the image is captured, various effects can be created knowing that the faster or slower recorded image will be played at a constant speed. Giving the cinematographer even more freedom for creativity and expression to be made.

For instance, time-lapse photography is created by exposing an image at an extremely slow rate. If a cinematographer sets a camera to expose one frame every minute for four hours, and then that footage is projected at 24 frames per second, a four-hour event will take 10 seconds to present, and one can present the events of a whole day (24 hours) in just one minute.

The inverse of this, if an image is captured at speeds above that at which they will be presented, the effect is to greatly slow down (slow motion) the image. If a cinematographer shoots a person diving into a pool at 96 frames per second, and that image is played back at 24 frames per second, the presentation will take 4 times as long as the actual event. Extreme slow motion, capturing many thousands of frames per second can present things normally invisible to the human eye, such as bullets in flight and shockwaves travelling through media, a potentially powerful cinematographic technique.

In motion pictures, the manipulation of time and space is a considerable contributing factor to the narrative storytelling tools. Film editing plays a much stronger role in this manipulation, but frame rate selection in the photography of the original action is also a contributing factor to altering time. For example, Charlie Chaplin's Modern Times was shot at "silent speed" (18 fps) but projected at "sound speed" (24 fps), which makes the slapstick action appear even more frenetic.

Speed ramping, or simply "ramping", is a process whereby the capture frame rate of the camera changes over time. For example, if in the course of 10 seconds of capture, the capture frame rate is adjusted from 60 frames per second to 24 frames per second, when played back at the standard movie rate of 24 frames per second, a unique time-manipulation effect is achieved. For example, someone pushing a door open and walking out into the street would appear to start off in slow-motion, but in a few seconds later within the same shot, the person would appear to walk in "realtime" (normal speed). The opposite speed-ramping is done in *The Matrix* when Neo re-enters the Matrix for the first time to see the Oracle. As he comes out of the warehouse "load-point", the camera zooms into Neo at normal speed but as it gets closer to Neo's face, time seems to slow down, foreshadowing the manipulation of time itself within the Matrix later in the movie.

Reverse and slow motion

G. A. Smith initiated the technique of reverse motion and also improved the quality of self-motivating images. This he did by repeating the action a second time while filming it with an inverted camera and then joining the tail of the second negative to that of the first. The first films using this were *Topsy, Topsy, Turvy*, and *The Awkward Sign Painter*, the latter which showed a sign painter lettering a sign, and then the painting on the sign vanishing under the painter's brush. The earliest surviving example of this technique is Smith's *The House That Jack Built*, made before September 1901. Here, a small boy is shown knocking down a castle just constructed by a little girl out of children's building blocks. A title then appears, saying "Reversed", and the action is repeated in reverse so that the castle re-erects itself under his blows.

Cecil Hepworth improved upon this technique by printing the negative of the forward motion backward, frame by frame, so that in the production of the print the original action was exactly reversed. Hepworth made *The Bathers* in 1900 in which bathers who have undressed and jumped into the water appear to spring backward out of it, and have their clothe magically fly back onto their bodies.

The use of different camera speeds also appeared around 1900. Robert Paul's *On a Runaway Motor Car through Piccadilly Circus* (1899), had the camera turn so slowly that when the film was projected at the usual 16 frames per second, the scenery appeared to be passing at great speed. Cecil Hepworth used the opposite effect in *The Indian Chief and the Seidlitz powder* (1901), in which a naïve Red Indian eats a lot of the fizzy stomach medicine, causing his stomach to expand and then he then leaps around balloon-like. This was done by cranking the camera faster than the normal 16 frames per second giving the first "slow motion" effect.

Personnel

In descending order of seniority, the following staff is involved:

- Director of photography, also called cinematographer
- Camera operator, also called cameraman
- First assistant camera, also called focus puller
- Second assistant camera, also called clapper loader

In the film industry, the cinematographer is responsible for the technical aspects of the images (lighting, lens choices, composition, exposure, filtration, film selection), but works closely with the director to ensure that the artistic aesthetics are supporting the director's vision of the story being told. The cinematographers are the heads of the camera, grip and lighting crew on a set, and for this reason, they are often called directors of photography or DPs. The American Society of Cinematographers defines cinematography as a creative and interpretive process that culminates in the authorship of an original work of art rather than the simple recording of a physical event. Cinematography is not a subcategory of photography. Rather, photography is but one craft that the cinematographer uses in addition to other physical, organizational, managerial, interpretive, and image-



A camera crew from the First Motion Picture Unit

manipulating techniques to effect one coherent process.^[26] In British tradition, if the DOP actually operates the camera him/herself they are called the *cinematographer*. On smaller productions, it is common for one person to perform all these functions alone. The career progression usually involves climbing up the ladder from seconding, firsting, eventually to operating the camera.

Directors of photography make many interpretive decisions during the course of their work, from pre-production to post-production, all of which affect the overall feel and look of the motion picture. Many of these decisions are similar to what a photographer needs to note when taking a picture: the cinematographer controls the film choice itself (from a range of available stocks with varying sensitivities to light and color), the selection of lens focal lengths, aperture exposure and focus. Cinematography, however, has a temporal aspect (see persistence of vision), unlike still photography, which is purely a single still image. It is also bulkier and more strenuous to deal with movie cameras, and it involves a more complex array of choices. As such a cinematographer often needs to work cooperatively with more people than does a photographer, who could frequently function as a single person. As a result, the cinematographer's job also includes personnel management and logistical organization. Given the in-depth knowledge, a cinematographer requires not only of his or her own craft but also that of other personnel, formal tuition in analogue or digital filmmaking can be advantageous.^[27]

See also



Film portal

- [3-D film](#)
- [360-degree video](#)
- [Academy Award for Best Cinematography](#)
- [Cinematographer](#)
- [Cinematography in healthcare](#)
- [Cinematography Mailing List](#), a communication forum for cinematographers
- [Digital cinema](#)
- [Fictional film](#)
- [Film crew](#)
- [Filmmaking](#)
- [Filmmaking technique of Kurosawa](#)
- [Film theory](#)
- [Films about cinematography:](#)
 - [Visions of Light](#) (1992)
 - [Cinematographer Style](#) (2006)
- [Glossary of motion picture terms](#)
- [History of cinema](#)
- [List of film formats](#)
- [List of film techniques](#)
- [List of motion picture-related topics](#) (extensive alphabetical listing and glossary).
- [List of cinema of the world](#)
- [List of video-related topics](#)
- [Outline of film](#)

- [Photographic film](#)
- [Special effect](#)
- [Videography](#)

References

1. Spencer, D A (1973). *The Focal Dictionary of Photographic Technologies*. Focal Press. p. 454. ISBN [978-0133227192](#).
2. Ronalds, B.F. (2016). *Sir Francis Ronalds: Father of the Electric Telegraph*. London: Imperial College Press. ISBN [978-1-78326-917-4](#).
3. Ronalds, B.F. (2016). "The Beginnings of Continuous Scientific Recording using Photography: Sir Francis Ronalds' Contribution" (<http://www.eshph.org/blog/2016/04/19/1642/>). *European Society for the History of Photography*. Retrieved 2 June 2016.
4. "The First "Movie Camera"" (<http://www.sirfrancisronalds.co.uk/camera.html>). *Sir Francis Ronalds and his Family*. Retrieved 27 September 2018.
5. Clegg, Brian (2007). *The Man Who Stopped Time* (https://archive.org/details/isbn_9780309101127). Joseph Henry Press. ISBN [978-0-309-10112-7](#).
6. Landecker, Hannah (2006). "Microcinematography and the History of Science and Film". *Isis*. **97**: 121–132. doi:[10.1086/501105](#) (<https://doi.org/10.1086%2F501105>). S2CID [144554305](#) (<https://api.semanticscholar.org/CorpusID:144554305>).
7. Mertes, Micah (16 January 2019). "Only 7 percent of the movies on Netflix came out before the year 2000" (https://www.omaha.com/go/only-percent-of-the-movies-on-netflix-came-out-before/article_1e2ca5e5-3943-5283-b112-9f60ff68e87b.html). *Omaha World-Herald*. BH Media Group, Inc. Archived (https://archive.today/20190410135357/https://www.omaha.com/go/only-percent-of-the-movies-on-netflix-came-out-before/article_1e2ca5e5-3943-5283-b112-9f60ff68e87b.html) from the original on 10 April 2019. Retrieved 10 April 2019.
8. Rizzo, Michael (11 July 2014). *The Art Direction Handbook for Film & Television* (<https://books.google.com/books?id=SdIABAAQBAJ&pg=PA92>). CRC Press. p. 92. ISBN [9781317673705](#).
9. Coe, Brian (July 1969). "William Friese Greene and the Origins of Cinematography III". *Screen*. **10** (4–5): 129–147. doi:[10.1093/screen/10.4-5.129](#) (<https://doi.org/10.1093/screen%2F10.4-5.129>).
10. Spehr, Paul (2008). *The Man Who Made Movies: W.K.L. Dickson* (<https://books.google.com/books?id=hcWCDgAAQBAJ>). Indiana University Press, John Libbey Publishing. doi:[10.2307/j.ctt20060gj](#) (<https://doi.org/10.2307%2Fj.ctt20060gj>). ISBN [978-0-861-96695-0](#). JSTOR [j.ctt20060gj](#) (<https://www.jstor.org/stable/j.ctt20060gj>).
11. Kreitner, Richard (24 August 2015). *August 24, 1891: Thomas Edison Receives a Patent for His Movie Camera, the Kinetograph* (<https://www.thenation.com/article/august-24-1891-thomas-edison-receives-a-patent-for-his-movie-camera-the-kinetograph/>). Retrieved 10 April 2019. {{cite book}}: |magazine= ignored (help)
12. Reed, Carey (22 August 2015). "150 years on, meet the prolific pioneer who brought us the movie projector" (<https://www.pbs.org/newshour/science/150-years-prolific-pioneer-brought-us-movie-projector>). *PBS News Hour*. PBS. Retrieved 10 April 2019.
13. Myrent, Glenn (29 December 1985). "WHEN MOVIEVIEW BEGAN AND NO ONE CAME" (<http://www.nytimes.com/1985/12/29/movies/when-moview-began-and-no-one-came.html>). *The New York Times*. p. 2019. Retrieved 10 April 2019.
14. Baker, T. Thorne (1932). "New Developments in Colour Cinematography". *Journal of the Royal Society of Arts*.
15. Baker, T. Thorne (1932). "New Developments in Colour Photography". *Journal of the Royal Society of Arts*.

16. "World's first colour film footage viewed for first time" (<https://www.bbc.com/news/uk-english-19557914>). BBC News. 12 September 2012.
17. McKernan, Luke (2018). *Charles Urban: Pioneering the Non-Fiction Film in Britain and America, 1897–1925*. University of Exeter Press. ISBN 978-0859892964.
18. Williams, J. B. (2017). *The Electronics Revolution: Inventing the Future* (<https://books.google.com/books?id=v4QIDwAAQBAJ&pg=PA245>). Springer. pp. 245–8. ISBN 9783319490885.
19. James R. Janesick (2001). *Scientific charge-coupled devices* (<https://books.google.com/books?id=3GyE4SWytn4C&pg=PA3>). SPIE Press. pp. 3–4. ISBN 978-0-8194-3698-6.
20. Stump, David (2014). *Digital Cinematography: Fundamentals, Tools, Techniques, and Workflows* (<https://books.google.com/books?id=c-MjAwAAQBAJ&pg=PA83>). CRC Press. pp. 83–5. ISBN 978-1-136-04042-9.
21. Stump, David (2014). *Digital Cinematography: Fundamentals, Tools, Techniques, and Workflows* (<https://books.google.com/books?id=c-MjAwAAQBAJ&pg=PA19>). CRC Press. pp. 19–22. ISBN 978-1-136-04042-9.
22. Fossum, Eric R.; Hondongwa, D. B. (2014). "A Review of the Pinned Photodiode for CCD and CMOS Image Sensors" (<https://doi.org/10.1109%2FJEDS.2014.2306412>). *IEEE Journal of the Electron Devices Society*. 2 (3): 33–43. doi:10.1109/JEDS.2014.2306412 (<https://doi.org/10.1109%2FJEDS.2014.2306412>).
23. Fossum, Eric R. (12 July 1993). "Active pixel sensors: Are CCDS dinosaurs?". In Blouke, Morley M. (ed.). *Charge-Coupled Devices and Solid State Optical Sensors III*. Vol. 1900. International Society for Optics and Photonics. pp. 2–14. Bibcode:1993SPIE.1900....2F (<https://ui.adsabs.harvard.edu/abs/1993SPIE.1900....2F>). CiteSeerX 10.1.1.408.6558 (<https://citesearx.ist.psu.edu/viewdoc/summary?doi=10.1.1.408.6558>). doi:10.1117/12.148585 (<https://doi.org/10.1117%2F12.148585>). S2CID 10556755 (<https://api.semanticscholar.org/CorpusID:10556755>). {{cite book}}: |journal= ignored (help)
24. "The Last Broadcast is A First: The Making of a Digital Feature" (<http://www.thelastbroadcastmovie.com/>). *thelastbroadcastmovie.com*.
25. Moore, Harris C. (1949). "Production Problems: Cinematography". *Journal of the University Film Producers Association*.
26. Hora, John (2007). "Anamorphic Cinematography". In Burum, Stephen H. (ed.). *The American Cinematographer Manual* (9 ed.). ASC Press. ISBN 978-0-935578-31-7.
27. "New York Film Academy – Filmmaking" (<http://www.nyfa.edu/digital-filmmaking/>). *nyfa.edu*.

External links

-  Movie Making Manual: Cinematography section at Wikibooks
 - The History of Cinematography (<https://web.archive.org/web/20091030161724/http://motion.kodak.com/US/en/motion/Hub/history1.htm>) at Kodak.
 - Burns, Paul. The History of the Discovery of Cinematography (<http://www.precinemahistory.net>)
-

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Cinematography&oldid=1286012571>"



Television

Television (TV) is a telecommunication medium for transmitting moving images and sound. Additionally, the term can refer to a physical television set rather than the medium of transmission. Television is a mass medium for advertising, entertainment, news, and sports. The medium is capable of more than "radio broadcasting," which refers to an audio signal sent to radio receivers.

Television became available in crude experimental forms in the 1920s, but only after several years of further development was the new technology marketed to consumers. After World War II, an improved form of black-and-white television broadcasting became popular in the United Kingdom and the United States, and television sets became commonplace in homes, businesses, and institutions. During the 1950s, television was the primary medium for influencing public opinion.^[1] In the mid-1960s, color broadcasting was introduced in the U.S. and most other developed countries.

The availability of various types of archival storage media such as Betamax and VHS tapes, LaserDiscs, high-capacity hard disk drives, CDs, DVDs, flash drives, high-definition HD DVDs and Blu-ray Discs, and cloud digital video recorders has enabled viewers to watch pre-recorded material—such as movies—at home on their own time schedule. For many reasons, especially the convenience of remote retrieval, the storage of television and video programming now also occurs on the cloud (such as the video-on-demand service by Netflix). At the beginning of the 2010s, digital television transmissions greatly increased in popularity. Another development was the move from standard-definition television (SDTV) (576i, with 576 interlaced lines of resolution and 480i) to high-definition television (HDTV), which provides a resolution that is substantially higher. HDTV may be transmitted in different formats: 1080p, 1080i and 720p. Since 2010, with the invention of smart television, Internet television has increased the availability of television programs and movies via the Internet through streaming video services such as Netflix, Amazon Prime Video, iPlayer and Hulu.

In 2013, 79% of the world's households owned a television set.^[2] The replacement of earlier cathode-ray tube (CRT) screen displays with compact, energy-efficient, flat-panel alternative technologies such as LCDs (both fluorescent-backlit and LED), OLED displays, and plasma displays was a hardware revolution that began with computer monitors in the late 1990s. Most television sets sold in the 2000s were still CRT, it was only in early 2010s that flat-screen TVs decisively overtook CRT.^[3] Major manufacturers announced the discontinuation of CRT, Digital Light Processing (DLP), plasma, and even fluorescent-backlit LCDs by the mid-2010s.^{[4][5]} LEDs are being gradually replaced by OLEDs.^[6] Also,



Flat-screen television receivers on display for sale at a consumer electronics store in 2008

major manufacturers have started increasingly producing smart TVs in the mid-2010s.^{[7][8][9]} Smart TVs with integrated Internet and Web 2.0 functions became the dominant form of television by the late 2010s.^[10]

Television signals were initially distributed only as terrestrial television using high-powered radio-frequency television transmitters to broadcast the signal to individual television receivers. Alternatively, television signals are distributed by coaxial cable or optical fiber, satellite systems, and, since the 2000s, via the Internet. Until the early 2000s, these were transmitted as analog signals, but a transition to digital television was expected to be completed worldwide by the late 2010s. A standard television set consists of multiple internal electronic circuits, including a tuner for receiving and decoding broadcast signals. A visual display device that lacks a tuner is correctly called a video monitor rather than a television.

The television broadcasts are mainly a simplex broadcast meaning that the transmitter cannot receive and the receiver cannot transmit.

Etymology

The word *television* comes from Ancient Greek τῆλε (tele) 'far' and Latin visio 'sight'. The first documented usage of the term dates back to 1900, when the Russian scientist Constantin Perskyi used it in a paper that he presented in French at the first International Congress of Electricity, which ran from 18 to 25 August 1900 during the International World Fair in Paris.

The anglicized version of the term is first attested in 1907, when it was still "...a theoretical system to transmit moving images over telegraph or telephone wires".^[11] It was "...formed in English or borrowed from French *télévision*".^[11] In the 19th century and early 20th century, other "...proposals for the name of a then-hypothetical technology for sending pictures over distance were telephote (1880) and televisa (1904)."^[11]

The abbreviation *TV* is from 1948. The use of the term to mean "a television set" dates from 1941.^[11] The use of the term to mean "television as a medium" dates from 1927.^[11]

The term *telly* is more common in the UK. The slang term "the tube" or the "boob tube" derives from the bulky cathode-ray tube used on most TVs until the advent of flat-screen TVs. Another slang term for the TV is "idiot box."^[12]

History

Mechanical

Facsimile transmission systems for still photographs pioneered methods of mechanical scanning of images in the early 19th century. Alexander Bain introduced the facsimile machine between 1843 and 1846. Frederick Bakewell demonstrated a working laboratory version in 1851. Willoughby Smith discovered the photoconductivity of the element selenium in 1873. As a 23-year-old German university student, Paul Julius Gottlieb Nipkow proposed and patented the Nipkow disk in 1884 in Berlin.^[13] This was a spinning disk with a spiral pattern of holes, so each hole scanned a line of the image. Although he

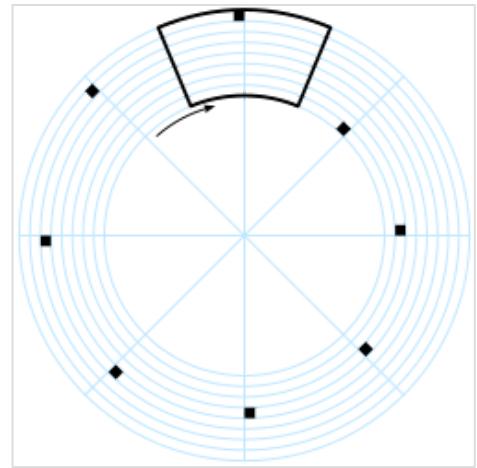
never built a working model of the system, variations of Nipkow's spinning-disk "image rasterizer" became exceedingly common.^[14] Constantin Perskyi had coined the word *television* in a paper read to the International Electricity Congress at the International World Fair in Paris on 24 August 1900. Perskyi's paper reviewed the existing electromechanical technologies, mentioning the work of Nipkow and others.^[15] However, it was not until 1907 that developments in amplification tube technology by Lee de Forest and Arthur Korn, among others, made the design practical.^[16]

The first demonstration of the live transmission of images was by Georges Rignoux and A. Fournier in Paris in 1909. A matrix of 64 selenium cells, individually wired to a mechanical commutator, served as an electronic retina. In the receiver, a type of Kerr cell modulated the light, and a series of differently angled mirrors attached to the edge of a rotating disc scanned the modulated beam onto the display screen. A separate circuit regulated synchronization. The 8x8 pixel resolution in this proof-of-concept demonstration was just sufficient to clearly transmit individual letters of the alphabet. An updated image was transmitted "several times" each second.^[17]

In 1911, Boris Rosing and his student Vladimir Zworykin created a system that used a mechanical mirror-drum scanner to transmit, in Zworykin's words, "very crude images" over wires to the "Braun tube" (cathode-ray tube or "CRT") in the receiver. Moving images were not possible because, in the scanner: "the sensitivity was not enough and the selenium cell was very laggy".^[18]

In 1921, Édouard Belin sent the first image via radio waves with his belinograph.^[19]

By the 1920s, when amplification made television practical, Scottish inventor John Logie Baird employed the Nipkow disk in his prototype video systems. On 25 March 1925, Baird gave the first public demonstration of televised silhouette images in motion at Selfridges's department store in London.^[20] Since human faces had inadequate contrast to show up on his primitive system, he televised a ventriloquist's dummy named "Stooky Bill," whose painted face had higher contrast, talking and moving. By 26 January 1926, he had demonstrated before members of the Royal Institution the transmission of an image of a face in motion by radio. This is widely regarded as the world's first true public television demonstration, exhibiting light, shade, and detail.^[21] Baird's system used the Nipkow disk for both scanning the image and displaying it. A brightly illuminated subject was placed in front of a spinning Nipkow disk set with lenses that swept images across a static photocell. The thallium sulfide (thalofide) cell, developed by Theodore Case in the U.S., detected the light reflected from the subject and converted it into a proportional electrical signal. This was transmitted by AM radio waves to a receiver unit, where the video signal was



The Nipkow disk. This schematic shows the circular paths traced by the holes that may also be square for greater precision. The area of the disk outlined in black displays the region scanned.



Baird in 1925 with his televiser equipment and dummies "James" and "Stooky Bill" (right)

applied to a neon light behind a second Nipkow disk rotating synchronized with the first. The brightness of the neon lamp was varied in proportion to the brightness of each spot on the image. As each hole in the disk passed by, one scan line of the image was reproduced. Baird's disk had 30 holes, producing an image with only 30 scan lines, just enough to recognize a human face.^[22] In 1927, Baird transmitted a signal over 438 miles (705 km) of telephone line between London and Glasgow.^[23] Baird's original 'televisor' now resides in the Science Museum, South Kensington.

In 1928, Baird's company (Baird Television Development Company/Cinema Television) broadcast the first transatlantic television signal between London and New York and the first shore-to-ship transmission. In 1929, he became involved in the first experimental mechanical television service in Germany. In November of the same year, Baird and Bernard Natan of Pathé established France's first television company, Télévision-Baird-Natan. In 1931, he made the first outdoor remote broadcast of The Derby.^[24] In 1932, he demonstrated ultra-short wave television. Baird's mechanical system reached a peak of 240 lines of resolution on BBC telecasts in 1936, though the mechanical system did not scan the televised scene directly. Instead, a 17.5 mm film was shot, rapidly developed, and then scanned while the film was still wet.

A U.S. inventor, Charles Francis Jenkins, also pioneered the television. He published an article on "Motion Pictures by Wireless" in 1913, transmitted moving silhouette images for witnesses in December 1923, and on 13 June 1925, publicly demonstrated synchronized transmission of silhouette pictures. In 1925, Jenkins used the Nipkow disk and transmitted the silhouette image of a toy windmill in motion over a distance of 5 miles (8 km), from a naval radio station in Maryland to his laboratory in Washington, D.C., using a lensed disk scanner with a 48-line resolution.^{[25][26]} He was granted U.S. Patent No. 1,544,156 (Transmitting Pictures over Wireless) on 30 June 1925 (filed 13 March 1922).^[27]

Herbert E. Ives and Frank Gray of Bell Telephone Laboratories gave a dramatic demonstration of mechanical television on 7 April 1927. Their reflected-light television system included both small and large viewing screens. The small receiver had a 2-inch-wide by 2.5-inch-high screen (5 by 6 cm). The large receiver had a screen 24 inches wide by 30 inches high (60 by 75 cm). Both sets could reproduce reasonably accurate, monochromatic, moving images. Along with the pictures, the sets received synchronized sound. The system transmitted images over two paths: first, a copper wire link from Washington to New York City, then a radio link from Whippany, New Jersey. Comparing the two transmission methods, viewers noted no difference in quality. Subjects of the telecast included Secretary of Commerce Herbert Hoover. A flying-spot scanner beam illuminated these subjects. The scanner that produced the beam had a 50-aperture disk. The disc revolved at a rate of 18 frames per second, capturing one frame about every 56 milliseconds. (Today's systems typically transmit 30 or 60 frames per second, or one frame every 33.3 or 16.7 milliseconds, respectively.) Television historian Albert Abramson underscored the significance of the Bell Labs demonstration: "It was, in fact, the best demonstration of a mechanical television system ever made to this time. It would be several years before any other system could even begin to compare with it in picture quality."^[28]

In 1928, WRGB, then W2XB, was started as the world's first television station. It broadcast from the General Electric facility in Schenectady, NY. It was popularly known as "WGY Television." Meanwhile, in the Soviet Union, Leon Theremin had been developing a mirror drum-based television, starting with 16 lines resolution in 1925, then 32 lines, and eventually 64 using interlacing in 1926. As part of his thesis, on 7 May 1926, he electrically transmitted and then projected near-simultaneous moving images on a 5-square-foot (0.46 m^2) screen.^[26]

By 1927 Theremin had achieved an image of 100 lines, a resolution that was not surpassed until May 1932 by RCA, with 120 lines.^[29]

On 25 December 1926, Kenjiro Takayanagi demonstrated a television system with a 40-line resolution that employed a Nipkow disk scanner and CRT display at Hamamatsu Industrial High School in Japan. This prototype is still on display at the Takayanagi Memorial Museum in Shizuoka University, Hamamatsu Campus. His research in creating a production model was halted by the SCAP after World War II.^[30]

Because only a limited number of holes could be made in the disks, and disks beyond a certain diameter became impractical, image resolution on mechanical television broadcasts was relatively low, ranging from about 30 lines up to 120 or so. Nevertheless, the image quality of 30-line transmissions steadily improved with technical advances, and by 1933 the UK broadcasts using the Baird system were remarkably clear.^[31] A few systems ranging into the 200-line region also went on the air. Two of these were the 180-line system that Compagnie des Compteurs (CDC) installed in Paris in 1935 and the 180-line system that Peck Television Corp. started in 1935 at station VE9AK in Montreal.^{[32][33]} The advancement of all-electronic television (including image dissectors and other camera tubes and cathode-ray tubes for the reproducer) marked the start of the end for mechanical systems as the dominant form of television. Mechanical television, despite its inferior image quality and generally smaller picture, would remain the primary television technology until the 1930s. The last mechanical telecasts ended in 1939 at stations run by a lot of public universities in the United States.

Electronic

In 1897, English physicist J. J. Thomson was able, in his three well-known experiments, to deflect cathode rays, a fundamental function of the modern cathode-ray tube (CRT). The earliest version of the CRT was invented by the German physicist Ferdinand Braun in 1897 and is also known as the "Braun" tube.^[34] It was a cold-cathode diode, a modification of the Crookes tube, with a phosphor-coated screen. Braun was the first to conceive the use of a CRT as a display device.^[35] The *Braun tube* became the foundation of 20th century television.^[36] In 1906 the Germans Max Dieckmann and Gustav Glage produced raster images for the first time in a CRT.^[37] In 1907, Russian scientist Boris Rosing used a CRT in the receiving end of an experimental video signal to form a picture. He managed to display simple geometric shapes onto the screen.^[38]



Ferdinand Braun

In 1908, Alan Archibald Campbell-Swinton, a fellow of the Royal Society (UK), published a letter in the scientific journal *Nature* in which he described how "distant electric vision" could be achieved by using a cathode-ray tube, or Braun tube, as both a transmitting and receiving device,^{[39][40]} he expanded on his vision in a speech given in London in 1911 and reported in *The Times*^[41] and the Journal of the Röntgen Society.^{[42][43]} In a letter to *Nature* published in October 1926, Campbell-Swinton also announced the results of some "not very successful experiments" he had conducted with G. M. Minchin and J. C. M. Stanton. They had attempted to generate an electrical signal by projecting an image onto a selenium-coated metal plate that was simultaneously scanned by a cathode ray beam.^{[44][45]} These experiments were conducted before March

1914, when Minchin died,^[46] but they were later repeated by two different teams in 1937, by H. Miller and J. W. Strange from EMI,^[47] and by H. Iams and A. Rose from RCA.^[48] Both teams successfully transmitted "very faint" images with the original Campbell-Swinton's selenium-coated plate. Although others had experimented with using a cathode-ray tube as a receiver, the concept of using one as a transmitter was novel.^[49] The first cathode-ray tube to use a hot cathode was developed by John B. Johnson (who gave his name to the term Johnson noise) and Harry Weiner Weinhart of Western Electric, and became a commercial product in 1922.

In 1926, Hungarian engineer Kálmán Tihanyi designed a television system using fully electronic scanning and display elements and employing the principle of "charge storage" within the scanning (or "camera") tube.^{[50][51][52][53]} The problem of low sensitivity to light resulting in low electrical output from transmitting or "camera" tubes would be solved with the introduction of charge-storage technology by Kálmán Tihanyi beginning in 1924.^[54] His solution was a camera tube that accumulated and stored electrical charges ("photoelectrons") within the tube throughout each scanning cycle. The device was first described in a patent application he filed in Hungary in March 1926 for a television system he called "Radioskop".^[50] After further refinements included in a 1928 patent application,^[54] Tihanyi's patent was declared void in Great Britain in 1930,^[55] so he applied for patents in the United States. Although his breakthrough would be incorporated into the design of RCA's "iconoscope" in 1931, the U.S. patent for Tihanyi's transmitting tube would not be granted until May 1939. The patent for his receiving tube had been granted the previous October. Both patents had been purchased by RCA prior to their approval.^{[56][57]} Charge storage remains a basic principle in the design of imaging devices for television to the present day.^[50] On 25 December 1926, at Hamamatsu Industrial High School in Japan, Japanese inventor Kenjiro Takayanagi demonstrated a TV system with a 40-line resolution that employed a CRT display.^[30] This was the first working example of a fully electronic television receiver and Takayanagi's team later made improvements to this system parallel to other television developments.^[58] Takayanagi did not apply for a patent.^[59]

In the 1930s, Allen B. DuMont made the first CRTs to last 1,000 hours of use, one of the factors that led to the widespread adoption of television.^[60]

On 7 September 1927, U.S. inventor Philo Farnsworth's image dissector camera tube transmitted its first image, a simple straight line, at his laboratory at 202 Green Street in San Francisco.^{[61][62]} By 3 September 1928, Farnsworth had developed the system sufficiently to hold a demonstration for the press. This is widely regarded as the first electronic television demonstration.^[62] In 1929, the system was improved further by eliminating a motor generator so that his television system had no mechanical parts.^[63] That year, Farnsworth transmitted the first live human images with his system, including a three and a half-inch image of his wife Elma ("Pem") with her eyes closed (possibly due to the bright lighting required).^[64]

Meanwhile, Vladimir Zworykin also experimented with the cathode-ray tube to create and show images. While working for Westinghouse Electric in 1923, he began to develop an electronic camera tube. However, in a 1925 demonstration, the image was dim, had low contrast and poor definition, and was stationary.^[65] Zworykin's imaging tube never got beyond the laboratory stage. However, RCA, which acquired the Westinghouse patent, asserted that the patent for Farnsworth's 1927 image dissector was written so broadly that it would exclude any other electronic imaging device. Thus, based on Zworykin's 1923 patent application, RCA filed a patent interference suit against Farnsworth. The U.S. Patent Office examiner disagreed in a 1935 decision, finding priority of invention for Farnsworth against Zworykin.

Farnsworth claimed that Zworykin's 1923 system could not produce an electrical image of the type to challenge his patent. Zworykin received a patent in 1928 for a color transmission version of his 1923 patent application.^[66] He also divided his original application in 1931.^[67] Zworykin was unable or unwilling to introduce evidence of a working model of his tube that was based on his 1923 patent application. In September 1939, after losing an appeal in the courts and being determined to go forward with the commercial manufacturing of television equipment, RCA agreed to pay Farnsworth US\$1 million over ten years, in addition to license payments, to use his patents.^{[68][69]}

In 1933, RCA introduced an improved camera tube that relied on Tihanyi's charge storage principle.^[70] Called the "Iconoscope" by Zworykin, the new tube had a light sensitivity of about 75,000 lux, and thus was claimed to be much more sensitive than Farnsworth's image dissector. However, Farnsworth had overcome his power issues with his Image Dissector through the invention of a completely unique "Multipactor" device that he began work on in 1930, and demonstrated in 1931.^{[71][72]} This small tube could amplify a signal reportedly to the 60th power or better^[73] and showed great promise in all fields of electronics. Unfortunately, an issue with the multipactor was that it wore out at an unsatisfactory rate.^[74]

At the Berlin Radio Show in August 1931 in Berlin, Manfred von Ardenne gave a public demonstration of a television system using a CRT for both transmission and reception, the first completely electronic television transmission.^[75] However, Ardenne had not developed a camera tube, using the CRT instead as a flying-spot scanner to scan slides and film.^[76] Ardenne achieved his first transmission of television pictures on 24 December 1933, followed by test runs for a public television service in 1934. The world's first electronically scanned television service then started in Berlin in 1935, the Fernsehsender Paul Nipkow, culminating in the live broadcast of the 1936 Summer Olympic Games from Berlin to public places all over Germany.^{[77][78]}

Philo Farnsworth gave the world's first public demonstration of an all-electronic television system, using a live camera, at the Franklin Institute of Philadelphia on 25 August 1934 and for ten days afterward.^{[79][80]} Mexican inventor Guillermo González Camarena also played an important role in early television. His experiments with television (known as telectroescopía at first) began in 1931 and led to a patent for the "trichromatic field sequential system" color television in 1940.^[81] In Britain, the EMI engineering team led by Isaac Shoenberg applied in 1932 for a patent for a new device they called "the Emitron",^{[82][83]} which formed the heart of the cameras they designed for the BBC. On 2 November 1936, a 405-line broadcasting service employing the Emitron began at studios in Alexandra Palace and transmitted from a specially built mast atop one of the Victorian building's towers. It alternated briefly with Baird's mechanical system in adjoining studios but was more reliable and visibly superior. This was the world's first regular "high-definition" television service.^[84]



Vladimir Zworykin
demonstrates electronic
television (1929).

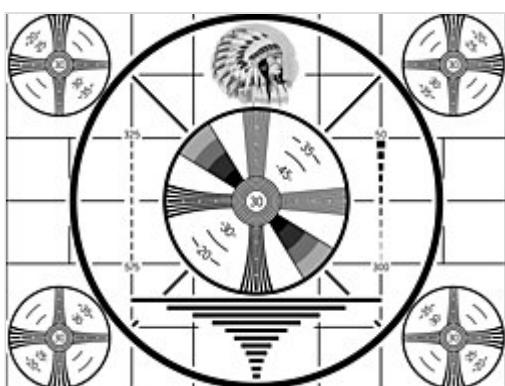


Manfred von Ardenne in 1933

The original U.S. iconoscope was noisy, had a high ratio of interference to signal, and ultimately gave disappointing results, especially compared to the high-definition mechanical scanning systems that became available.^{[85][86]} The EMI team, under the supervision of Isaac Shoenberg, analyzed how the iconoscope (or Emitron) produced an electronic signal and concluded that its real efficiency was only about 5% of the theoretical maximum.^{[87][88]} They solved this problem by developing and patenting in 1934 two new camera tubes dubbed super-Emitron and CPS Emitron.^{[89][90][91]} The super-Emitron was between ten and fifteen times more sensitive than the original Emitron and iconoscope tubes, and, in some cases, this ratio was considerably greater.^[87] It was used for outside broadcasting by the BBC, for the first time, on Armistice Day 1937, when the general public could watch on a television set as the King laid a wreath at the Cenotaph.^[92] This was the first time that anyone had broadcast a live street scene from cameras installed on the roof of neighboring buildings because neither Farnsworth nor RCA would do the same until the 1939 New York World's Fair.



Ad for the beginning of experimental television broadcasting in New York City by RCA in 1939



Indian-head test pattern used during the black-and-white era before 1970. It was displayed when a television station first signed on every day.

On the other hand, in 1934, Zworykin shared some patent rights with the German licensee company Telefunken.^[93] The "image iconoscope" ("Superikonoskop" in Germany) was produced as a result of the collaboration. This tube is essentially identical to the super-Emitron. The production and commercialization of the super-Emitron and image iconoscope in Europe were not affected by the patent war

between Zworykin and Farnsworth because Dieckmann and Hell had priority in Germany for the invention of the image dissector, having submitted a patent application for their *Lichtelektrische Bildzerlegerröhre für Fernseher* (*Photoelectric Image Dissector Tube for Television*) in Germany in 1925,^[94] two years before Farnsworth did the same in the United States.^[95] The image iconoscope (Superikonoskop) became the industrial standard for public broadcasting in Europe from 1936 until 1960, when it was replaced by the vidicon and plumbicon tubes. Indeed, it represented the European tradition in electronic tubes competing against the American tradition represented by the image orthicon.^{[96][97]} The German company Heimann produced the Superikonoskop for the 1936 Berlin Olympic Games,^{[98][99]} later Heimann also produced and commercialized it from 1940 to 1955;^[100] finally the Dutch company Philips produced and commercialized the image iconoscope and multicon from 1952 to 1958.^{[97][101]}

U.S. television broadcasting, at the time, consisted of a variety of markets in a wide range of sizes, each competing for programming and dominance with separate technology until deals were made and standards agreed upon in 1941.^[102] RCA, for example, used only Iconoscopes in the New York area, but Farnsworth Image Dissectors in Philadelphia and San Francisco.^[103] In September 1939, RCA agreed to pay the Farnsworth Television and Radio Corporation royalties over the next ten years for access to Farnsworth's patents.^[104] With this historic agreement in place, RCA integrated much of what was best about the Farnsworth Technology into their systems.^[103] In 1941, the United States implemented 525-line television.^{[105][106]} Electrical engineer Benjamin Adler played a prominent role in the development of television.^{[107][108]}

The world's first 625-line television standard was designed in the Soviet Union in 1944 and became a national standard in 1946.^[109] The first broadcast in 625-line standard occurred in Moscow in 1948.^[110] The concept of 625 lines per frame was subsequently implemented in the European CCIR standard.^[111] In 1936, Kálmán Tihanyi described the principle of plasma display, the first flat-panel display system.^{[112][113]}

Early electronic television sets were large and bulky, with analog circuits made of vacuum tubes. Following the invention of the first working transistor at Bell Labs, Sony founder Masaru Ibuka predicted in 1952 that the transition to electronic circuits made of transistors would lead to smaller and more portable television sets.^[114] The first fully transistorized, portable solid-state television set was the 8-inch Sony TV8-301, developed in 1959 and released in 1960.^{[115][116]} This began the transformation of television viewership from a communal viewing experience to a solitary viewing experience.^[117] By 1960, Sony had sold over 4 million portable television sets worldwide.^[118]

Color

The basic idea of using three monochrome images to produce a color image had been experimented with almost as soon as black-and-white televisions had first been built. Although he gave no practical details, among the earliest published proposals for television was one by Maurice Le Blanc in 1880 for a color system, including the first mentions in television literature of line and frame scanning.^[119] Polish inventor Jan Szczepanik patented a color television system in 1897, using a selenium photoelectric cell at the transmitter and an electromagnet controlling an oscillating mirror and a moving prism at the receiver. But his system contained no means of analyzing the spectrum of colors at the transmitting end and could not have worked as he described it.^[120] Another inventor, Hovannes Adamian, also experimented with color television as early as 1907. The first color television project is claimed by him,^[121] and was patented in Germany on 31 March 1908, patent No. 197183, then in Britain, on 1 April 1908, patent No. 7219,^[122] in France (patent No. 390326) and in Russia in 1910 (patent No. 17912).^[123]



Samsung LED TV

Scottish inventor John Logie Baird demonstrated the world's first color transmission on 3 July 1928, using scanning discs at the transmitting and receiving ends with three spirals of apertures, each spiral with filters of a different primary color, and three light sources at the receiving end, with a commutator to alternate their illumination.^[124] Baird also made the world's first color broadcast on 4 February 1938, sending a mechanically scanned 120-line image from Baird's Crystal Palace studios to a projection screen

at London's Dominion Theatre.^[125] Mechanically scanned color television was also demonstrated by Bell Laboratories in June 1929 using three complete systems of photoelectric cells, amplifiers, glow-tubes, and color filters, with a series of mirrors to superimpose the red, green, and blue images into one full-color image.

The first practical hybrid system was again pioneered by John Logie Baird. In 1940 he publicly demonstrated a color television combining a traditional black-and-white display with a rotating colored disk. This device was very "deep" but was later improved with a mirror folding the light path into an entirely practical device resembling a large conventional console.^[126] However, Baird was unhappy with the design, and, as early as 1944, had commented to a British government committee that a fully electronic device would be better.

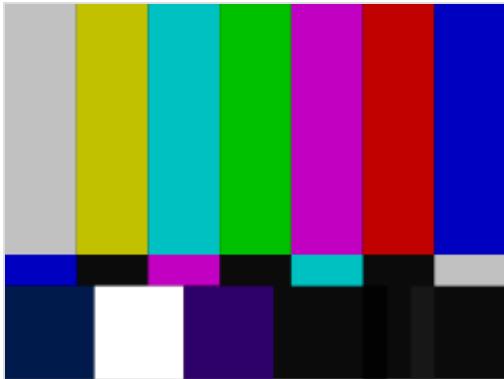
In 1939, Hungarian engineer Peter Carl Goldmark introduced an electro-mechanical system while at CBS, which contained an Iconoscope sensor. The CBS field-sequential color system was partly mechanical, with a disc made of red, blue, and green filters spinning inside the television camera at 1,200 rpm and a similar disc spinning in synchronization in front of the cathode-ray tube inside the receiver set.^[127] The system was first demonstrated to the Federal Communications Commission (FCC) on 29 August 1940 and shown to the press on 4 September.^{[128][129][130][131]}

CBS began experimental color field tests using film as early as 28 August 1940 and live cameras by 12 November.^{[129][132]} NBC (owned by RCA) made its first field test of color television on 20 February 1941. CBS began daily color field tests on 1 June 1941.^[133] These color systems were not compatible with existing black-and-white television sets, and, as no color television sets were available to the public at this time, viewing of the color field tests was restricted to RCA and CBS engineers and the invited press. The War Production Board halted the manufacture of television and radio equipment for civilian use from 22 April 1942 to 20 August 1945, limiting any opportunity to introduce color television to the general public.^{[134][135]}

As early as 1940, Baird had started work on a fully electronic system he called Telechrome. Early Telechrome devices used two electron guns aimed at either side of a phosphor plate. The phosphor was patterned so the electrons from the guns only fell on one side of the patterning or the other. Using cyan and magenta phosphors, a reasonable limited-color image could be obtained. He also demonstrated the same system using monochrome signals to produce a 3D image (called "stereoscopic" at the time). A demonstration on 16 August 1944 was the first example of a practical color television system. Work on the Telechrome continued, and plans were made to introduce a three-gun version for full color. However, Baird's untimely death in 1946 ended the development of the Telechrome system.^{[136][137]} Similar concepts were common through the 1940s and 1950s, differing primarily in the way they re-combined the colors generated by the three guns. The Geer tube was similar to Baird's concept but used small pyramids with the phosphors deposited on their outside faces instead of Baird's 3D patterning on a flat surface. The Penetron used three layers of phosphor on top of each other and increased the power of the beam to reach the upper layers when drawing those colors. The Chromatron used a set of focusing wires to select the colored phosphors arranged in vertical stripes on the tube.

One of the great technical challenges of introducing color broadcast television was the desire to conserve bandwidth, potentially three times that of the existing black-and-white standards, and not use an excessive amount of radio spectrum. In the United States, after considerable research, the National Television Systems Committee^[138] approved an all-electronic system developed by RCA, which encoded the color information separately from the brightness information and significantly reduced the resolution

of the color information to conserve bandwidth. As black-and-white televisions could receive the same transmission and display it in black-and-white, the color system adopted is [backwards] "compatible." ("Compatible Color," featured in RCA advertisements of the period, is mentioned in the song "America," of West Side Story, 1957.) The brightness image remained compatible with existing black-and-white television sets at slightly reduced resolution. In contrast, color televisions could decode the extra information in the signal and produce a limited-resolution color display. The higher-resolution black-and-white and lower-resolution color images combine in the brain to produce a seemingly high-resolution color image. The NTSC standard represented a significant technical achievement.



Color bars used in a test pattern, sometimes used when no program material is available

The first color broadcast (the first episode of the live program *The Marriage*) occurred on 8 July 1954. However, during the following ten years, most network broadcasts and nearly all local programming continued to be black-and-white. It was not until the mid-1960s that color sets started selling in large numbers, due in part to the color transition of 1965, in which it was announced that over half of all network prime-time programming would be broadcast in color that fall. The first all-color prime-time season came just one year later. In 1972, the last holdout among daytime network programs converted to color, resulting in the first completely all-color network season.

Early color sets were either floor-standing console models or tabletop versions nearly as bulky and heavy, so in practice they remained firmly anchored in one place. GE's relatively compact and lightweight Porta-Color set was introduced in the spring of 1966. It used a transistor-based UHF tuner.^[139] The first fully transistorized color television in the United States was the Quasar television introduced in 1967.^[140] These developments made watching color television a more flexible and convenient proposition.

In 1972, sales of color sets finally surpassed sales of black-and-white sets. Color broadcasting in Europe was not standardized on the PAL format until the 1960s, and broadcasts did not start until 1967. By this point, many of the technical issues in the early sets had been worked out, and the spread of color sets in Europe was fairly rapid. By the mid-1970s, the only stations broadcasting in black-and-white were a few high-numbered UHF stations in small markets and a handful of low-power repeater stations in even smaller markets such as vacation spots. By 1979, even the last of these had converted to color. By the early 1980s, B&W sets had been pushed into niche markets, notably low-power uses, small portable sets, or for use as video monitor screens in lower-cost consumer equipment. By the late 1980s, even these last holdout niche B&W environments had inevitably shifted to color sets.

Digital

Digital television (DTV) is the transmission of audio and video by digitally processed and multiplexed signals, in contrast to the analog and channel-separated signals used by analog television. Due to data compression, digital television can support more than one program on the same channel bandwidth.^[141] It is an innovative service that represents the most significant evolution in television broadcast technology since color television emerged in the 1950s.^[142] Digital television's roots have been tied very closely to the availability of inexpensive, high performance computers. It was not until the 1990s that digital

television became possible.^[143] Digital television was previously not practically possible due to the impractically high bandwidth requirements of uncompressed digital video,^{[144][145]} requiring around 200 Mbit/s for a standard-definition television (SDTV) signal,^[144] and over 1 Gbit/s for high-definition television (HDTV).^[145]

A digital television service was proposed in 1986 by Nippon Telegraph and Telephone (NTT) and the Ministry of Posts and Telecommunication (MPT) in Japan, where there were plans to develop an "Integrated Network System" service. However, it was not possible to implement such a digital television service practically until the adoption of DCT video compression technology made it possible in the early 1990s.^[144]

In the mid-1980s, as Japanese consumer electronics firms forged ahead with the development of HDTV technology, the MUSE analog format proposed by NHK, a Japanese company, was seen as a pacesetter that threatened to eclipse U.S. electronics companies' technologies. Until June 1990, the Japanese MUSE standard, based on an analog system, was the front-runner among the more than 23 other technical concepts under consideration. Then, a U.S. company, General Instrument, demonstrated the possibility of a digital television signal. This breakthrough was of such significance that the FCC was persuaded to delay its decision on an ATV standard until a digitally-based standard could be developed.

In March 1990, when it became clear that a digital standard was possible, the FCC made several critical decisions. First, the Commission declared that the new ATV standard must be more than an enhanced analog signal but be able to provide a genuine HDTV signal with at least twice the resolution of existing television images. (7) Then, to ensure that viewers who did not wish to buy a new digital television set could continue to receive conventional television broadcasts, it dictated that the new ATV standard must be capable of being "simulcast" on different channels. (8) The new ATV standard also allowed the new DTV signal to be based on entirely new design principles. Although incompatible with the existing NTSC standard, the new DTV standard would be able to incorporate many improvements.

The last standards adopted by the FCC did not require a single standard for scanning formats, aspect ratios, or lines of resolution. This compromise resulted from a dispute between the consumer electronics industry (joined by some broadcasters) and the computer industry (joined by the film industry and some public interest groups) over which of the two scanning processes—interlaced or progressive—would be best suited for the newer digital HDTV compatible display devices.^[146] Interlaced scanning, which had been specifically designed for older analog CRT display technologies, scans even-numbered lines first, then odd-numbered ones. Interlaced scanning can be regarded as the first video compression model. It was partly developed in the 1940s to double the image resolution to exceed the limitations of television broadcast bandwidth. Another reason for its adoption was to limit the flickering on early CRT screens, whose phosphor-coated screens could only retain the image from the electron scanning gun for a relatively short duration.^[147] However, interlaced scanning does not work as efficiently on newer display devices such as liquid-crystal display (LCD), for example, which are better suited to a more frequent progressive refresh rate.^[146]

Progressive scanning, the format that the computer industry had long adopted for computer display monitors, scans every line in sequence, from top to bottom. Progressive scanning, in effect, doubles the amount of data generated for every full screen displayed in comparison to interlaced scanning by painting the screen in one pass in 1/60-second instead of two passes in 1/30-second. The computer industry argued that progressive scanning is superior because it does not "flicker" on the new standard of display devices in the manner of interlaced scanning. It also argued that progressive scanning enables easier connections

with the Internet and is more cheaply converted to interlaced formats than vice versa. The film industry also supported progressive scanning because it offered a more efficient means of converting filmed programming into digital formats. For their part, the consumer electronics industry and broadcasters argued that interlaced scanning was the only technology that could transmit the highest quality pictures then (and currently) feasible, i.e., 1,080 lines per picture and 1,920 pixels per line. Broadcasters also favored interlaced scanning because their vast archive of interlaced programming is not readily compatible with a progressive format. William F. Schreiber, who was director of the Advanced Television Research Program at the Massachusetts Institute of Technology from 1983 until his retirement in 1990, thought that the continued advocacy of interlaced equipment originated from consumer electronics companies that were trying to get back the substantial investments they made in the interlaced technology.^[148]

Digital television transition started in the late 2000s. All governments across the world set the deadline for analog shutdown by the 2010s. Initially, the adoption rate was low, as the first digital tuner-equipped television sets were costly. However, as the price of digital-capable television sets dropped, more and more households started converting to digital television sets. The transition is expected to be completed worldwide by the mid to late 2010s.

Smart television

The advent of digital television allowed innovations like smart television sets. A smart television sometimes referred to as a "connected TV" or "hybrid TV," is a television set or set-top box with integrated Internet and Web 2.0 features and is an example of technological convergence between computers, television sets, and set-top boxes. Besides the traditional functions of television sets and set-top boxes provided through traditional Broadcasting media, these devices can also provide Internet TV, online interactive media, over-the-top content, as well as on-demand streaming media, and home networking access. These TVs come pre-loaded with an operating system.^{[10][149][150][151]}



A smart TV

Smart TV is not to be confused with Internet TV, Internet Protocol television (IPTV), or with Web TV. Internet television refers to receiving television content over the Internet instead of through traditional systems—terrestrial, cable, and satellite. IPTV is one of the emerging Internet television technology standards for television networks. Web television (WebTV) is a term used for programs created by a wide variety of companies and individuals for broadcast on Internet TV. A first patent was filed in 1994^[152] (and extended the following year)^[153] for an "intelligent" television system, linked with data processing systems, using a digital or analog network. Apart from being linked to data networks, one key point is its ability to automatically download necessary software routines according to a user's demand and process their needs. Major TV manufacturers announced the production of smart TVs only for middle-end and high-end TVs in 2015.^{[7][8][9]} Smart TVs have gotten more affordable compared to when they were first introduced, with 46 million U.S. households having at least one as of 2019.^[154]

3D

3D television conveys depth perception to the viewer by employing techniques such as stereoscopic display, multi-view display, 2D-plus-depth, or any other form of 3D display. Most modern 3D television sets use an active shutter 3D system or a polarized 3D system, and some are autostereoscopic without the need for glasses. Stereoscopic 3D television was demonstrated for the first time on 10 August 1928, by John Logie Baird in his company's premises at 133 Long Acre, London.^[155] Baird pioneered a variety of 3D television systems using electromechanical and cathode-ray tube techniques. The first 3D television was produced in 1935. The advent of digital television in the 2000s greatly improved 3D television sets. Although 3D television sets are quite popular for watching 3D home media, such as on Blu-ray discs, 3D programming has largely failed to make inroads with the public. As a result, many 3D television channels that started in the early 2010s were shut down by the mid-2010s. According to DisplaySearch 3D television shipments totaled 41.45 million units in 2012, compared with 24.14 in 2011 and 2.26 in 2010.^[156] As of late 2013, the number of 3D TV viewers started to decline.^{[157][158][159][160][161]}

Broadcast systems

Terrestrial television

Programming is broadcast by television stations, sometimes called "channels," as stations are licensed by their governments to broadcast only over assigned channels in the television band. At first, terrestrial broadcasting was the only way television could be widely distributed, and because bandwidth was limited, i.e., there were only a small number of channels available, government regulation was the norm. In the U.S., the Federal Communications Commission (FCC) allowed stations to broadcast advertisements beginning in July 1941 but required public service programming commitments as a requirement for a license. By contrast, the United Kingdom chose a different route, imposing a television license fee on owners of television reception equipment to fund the British Broadcasting Corporation (BBC), which had public service as part of its Royal Charter.



A modern high gain UHF Yagi television antenna. It has 17 directors and one reflector (made of 4 rods) shaped as a corner reflector.

WRGB claims to be the world's oldest television station, tracing its roots to an experimental station founded on 13 January 1928, broadcasting from the General Electric factory in Schenectady, NY, under the call letters **W2XB**.^[162] It was popularly known as "WGY Television" after its sister radio station. Later, in 1928, General Electric started a second facility, this one in New York City, which had the call letters W2XBS and which today is known as WNBC. The two stations were experimental and had no regular programming, as receivers were operated by engineers within the company. The image of a Felix the Cat doll rotating on a turntable was broadcast for 2 hours every day for several years as engineers tested new technology. On 2 November 1936, the BBC began transmitting the world's first public regular high-definition service from the Victorian Alexandra Palace in north London.^[163] It therefore claims to be the birthplace of television broadcasting as we now know it.

With the widespread adoption of cable across the United States in the 1970s and 1980s, terrestrial television broadcasts have been in decline; in 2013 it was estimated that about 7% of US households used an antenna.^{[164][165]} A slight increase in use began around 2010 due to switchover to digital terrestrial television broadcasts, which offered pristine image quality over very large areas, and offered an alternative to cable television (CATV) for cord cutters. All other countries around the world are also in the process of either shutting down analog terrestrial television or switching over to digital terrestrial television.

Cable television



Coaxial cable is used to carry cable television signals into cathode-ray tube and flat-panel television sets.

Cable television is a system of broadcasting television programming to paying subscribers via radio frequency (RF) signals transmitted through coaxial cables or light pulses through fiber-optic cables. This contrasts with traditional terrestrial television, in which the television signal is transmitted over the air by radio waves and received by a television antenna attached to the television. In the 2000s, FM radio programming, high-speed Internet, telephone service, and similar non-television services may also be provided through these cables. The abbreviation CATV is sometimes used for cable television in the United States. It originally stood for Community Access Television or Community Antenna Television, from cable television's origins in 1948: in

areas where over-the-air reception was limited by distance from transmitters or mountainous terrain, large "community antennas" were constructed, and cable was run from them to individual homes.^[166]

Satellite television

Satellite television is a system of supplying television programming using broadcast signals relayed from communication satellites. The signals are received via an outdoor parabolic reflector antenna, usually referred to as a satellite dish and a low-noise block downconverter (LNB). A satellite receiver then decodes the desired television program for viewing on a television set. Receivers can be external set-top boxes, or a built-in television tuner. Satellite television provides a wide range of channels and services, especially to geographic areas without terrestrial television or cable television.



DBS satellite dishes installed on an apartment complex

The most common method of reception is direct-broadcast satellite television (DBSTV), also known as "direct to home" (DTH).^[167] In DBSTV systems, signals are relayed from a direct broadcast satellite on the K_u wavelength and are completely digital.^[168] Satellite TV systems formerly used systems known as television receive-only. These systems received analog signals transmitted in the C-band spectrum from FSS type satellites and required the use of large dishes. Consequently, these systems were nicknamed "big dish" systems and were more expensive and less popular.^[169]

The direct-broadcast satellite television signals were earlier analog signals and later digital signals, both of which require a compatible receiver. Digital signals may include high-definition television (HDTV). Some transmissions and channels are free-to-air or free-to-view, while many other channels are pay television requiring a subscription.^[170] In 1945, British science fiction writer Arthur C. Clarke proposed a worldwide communications system that would function by means of three satellites equally spaced apart in Earth orbit.^{[171][172]} This was published in the October 1945 issue of the *Wireless World* magazine and won him the Franklin Institute's Stuart Ballantine Medal in 1963.^{[173][174]}

The first satellite television signals from Europe to North America were relayed via the *Telstar* satellite over the *Atlantic Ocean* on 23 July 1962.^[175] The signals were received and broadcast in North American and European countries and watched by over 100 million.^[175] Launched in 1962, the *Relay 1* satellite was the first satellite to transmit television signals from the US to Japan.^[176] The first geosynchronous communication satellite, *Syncom 2*, was launched on 26 July 1963.^[177]

The world's first commercial communications satellite, called *Intelsat I* and nicknamed "Early Bird", was launched into geosynchronous orbit on 6 April 1965.^[178] The first national network of television satellites, called *Orbita*, was created by the Soviet Union in October 1967, and was based on the principle of using the highly elliptical *Molniya* satellite for rebroadcasting and delivering of television signals to ground downlink stations.^[179] The first commercial North American satellite to carry television transmissions was Canada's geostationary *Anik 1*, which was launched on 9 November 1972.^[180] *ATS-6*, the world's first experimental educational and Direct Broadcast Satellite (DBS), was launched on 30 May 1974.^[181] It transmitted at 860 MHz using wideband FM modulation and had two sound channels. The transmissions were focused on the Indian subcontinent, but experimenters were able to receive the signal in Western Europe using home-constructed equipment that drew on UHF television design techniques already in use.^[182]

The first in a series of Soviet geostationary satellites to carry Direct-To-Home television, *Ekran 1*, was launched on 26 October 1976.^[183] It used a 714 MHz UHF downlink frequency so that the transmissions could be received with existing UHF television technology rather than microwave technology.^[184]

Internet television

Internet television (Internet TV) (or online television) is the digital distribution of television content via the Internet as opposed to traditional systems like terrestrial, cable, and satellite, although the Internet itself is received by terrestrial, cable, or satellite methods. Internet television is a general term that covers the delivery of television series and other video content over the Internet by video streaming technology, typically by major traditional television broadcasters. Internet television should not be confused with Smart TV, IPTV, or with Web TV. Smart television refers to the television set which has a built-in operating system. Internet Protocol television (IPTV) is one of the emerging Internet television technology standards for use by television networks. Web television is a term used for programs created by a wide variety of companies and individuals for broadcast on Internet television.

Traditional cable and satellite television providers began to offer services such as *Sling TV*, owned by Dish Network, which was unveiled in January 2015.^[185] *DirecTV*, another satellite television provider, launched their own streaming service, *DirecTV Stream*, in 2016.^{[186][187]} *Sky* launched a similar streaming service in the UK called *Now*. In 2013, Video on demand website *Netflix* earned the first Primetime Emmy Award nominations for original streaming television at the 65th Primetime Emmy

Awards. Three of its series, *House of Cards*, *Arrested Development*, and *Hemlock Grove*, earned nominations that year.^[188] On July 13, 2015, cable company Comcast announced an HBO plus broadcast TV package at a price discounted from basic broadband plus basic cable.^[189]

In 2017, YouTube launched YouTube TV, a streaming service that allows users to watch live television programs from popular cable or network channels and record shows to stream anywhere, anytime.^{[190][191][192]} As of 2017, 28% of US adults cite streaming services as their main means for watching television, and 61% of those ages 18 to 29 cite it as their main method.^{[193][194]} As of 2018, Netflix is the world's largest streaming TV network and also the world's largest Internet media and entertainment company with 117 million paid subscribers, and by revenue and market cap.^{[195][196]} In 2020, the COVID-19 pandemic had a strong impact in the television streaming business with the lifestyle changes such as staying at home and lockdowns.^{[197][198][199][200]}

Sets

A television set, also called a television receiver, television, TV set, TV, or "telly," is a device that combines a tuner, display, amplifier, and speakers for the purpose of viewing television and hearing its audio components. Introduced in the late 1920s in mechanical form, television sets became a popular consumer product after World War II in electronic form, using cathode-ray tubes. The addition of color to broadcast television after 1953 further increased the popularity of television sets, and an outdoor antenna became a common feature of suburban homes. The ubiquitous television set became the display device for recorded media in the 1970s, such as Betamax and VHS, which enabled viewers to record TV shows and watch prerecorded movies. In the subsequent decades, Television sets were used to watch DVDs and Blu-ray Discs of movies and other content. Major TV manufacturers announced the discontinuation of CRT, DLP, plasma, and fluorescent-backlit LCDs by the mid-2010s. Televisions since 2010s mostly use LEDs.^{[4][5][201][202]} LEDs are expected to be gradually replaced by OLEDs in the near future.^[6]



RCA 630-TS, the first mass-produced television set, which sold in 1946–1947

Display technologies

Disk

The earliest systems employed a spinning disk to create and reproduce images.^[203] These usually had a low resolution and screen size and never became popular with the public.

CRT

The cathode-ray tube (CRT) is a vacuum tube containing one or more electron guns (a source of electrons or electron emitter) and a fluorescent screen used to view images.^[38] It has the means to accelerate and deflect the electron beam(s) onto the screen to create the images. The images may represent electrical waveforms (oscilloscope), pictures (television, computer monitor), radar targets or others. The CRT uses



A 14-inch cathode-ray tube showing its deflection coils and electron guns

an evacuated glass envelope that is large, deep (i.e., long from front screen face to rear end), fairly heavy, and relatively fragile. As a matter of safety, the face is typically made of thick lead glass so as to be highly shatter-resistant and to block most X-ray emissions, particularly if the CRT is used in a consumer product.

field generated by coils and driven by electronic circuits around the neck of the tube, although electrostatic deflection is commonly used in oscilloscopes, a type of diagnostic instrument.^[204]

DLP

Digital Light Processing (DLP) is a type of video projector technology that uses a digital micromirror device. Some DLPs have a TV tuner, which makes them a type of TV display. It was originally developed in 1987 by Dr. Larry Hornbeck of Texas Instruments. While the DLP imaging device was invented by Texas Instruments, the first DLP-based projector was introduced by Digital Projection Ltd in 1997. Digital Projection and Texas Instruments were both awarded Emmy Awards in 1998 for the invention of the DLP projector technology. DLP is used in a variety of display applications, from traditional static displays to interactive displays and also non-traditional embedded applications, including medical, security, and industrial uses. DLP technology is used in DLP front projectors (standalone projection units for classrooms and businesses primarily) but also in private homes; in these cases, the image is projected onto a projection screen. DLP is also used in DLP rear projection television sets and digital signs. It is also used in about 85% of digital cinema projection.^[205]



The Christie Mirage 5000, a 2001 DLP projector

Plasma

A plasma display panel (PDP) is a type of flat-panel display common to large television displays 30 inches (76 cm) or larger. They are called "plasma" displays because the technology uses small cells containing electrically charged ionized gases, or what are in essence chambers more commonly known as fluorescent lamps.

LCD

Liquid-crystal-display televisions (LCD TVs) are television sets that use liquid-crystal display technology to produce images. LCD televisions are much thinner and lighter than cathode-ray tube (CRTs) of similar display size and are available in much larger sizes (e.g., 90-inch diagonal). When manufacturing costs

fell, this combination of features made LCDs practical for television receivers. LCDs come in two types: those using cold cathode fluorescent lamps, simply called LCDs, and those using LED as backlight called LEDs.

In 2007, LCD television sets surpassed sales of CRT-based television sets worldwide for the first time, and their sales figures relative to other technologies accelerated. LCD television sets have quickly displaced the only major competitors in the large-screen market, the Plasma display panel and rear-projection television.^[206] In mid 2010s LCDs especially LEDs became, by far, the most widely produced and sold television display type.^{[201][202]} LCDs also have disadvantages. Other technologies address these weaknesses, including OLEDs, FED and SED, but as of 2014 none of these have entered widespread production.



A generic LCD TV, with speakers on either side of the screen

OLED



OLED TV

An OLED (organic light-emitting diode) is a light-emitting diode (LED) in which the emissive electroluminescent layer is a film of organic compound which emits light in response to an electric current. This layer of organic semiconductor is situated between two electrodes. Generally, at least one of these electrodes is transparent. OLEDs are used to create digital displays in devices such as television screens. It is also used for computer monitors and portable systems such as mobile phones, handheld game console, and PDAs.

There are two main groups of OLED: those based on small molecules and those employing polymers. Adding mobile ions to an OLED creates a light-emitting electrochemical cell or LEC, which has a slightly different mode of operation. OLED displays can use either passive-matrix (PMOLED) or active-matrix (AMOLED) addressing schemes. Active-matrix OLEDs require a thin-film transistor backplane to switch each individual pixel on or off but allow for higher resolution and larger display sizes.

An OLED display works without a backlight. Thus, it can display deep black levels and can be thinner and lighter than a liquid crystal display (LCD). In low ambient light conditions such as a dark room, an OLED screen can achieve a higher contrast ratio than an LCD, whether the LCD uses cold cathode fluorescent lamps or LED backlight. OLEDs are expected to replace other forms of display in the near future.^[6]

Display resolution

LD

Low-definition television or LDTV refers to television systems that have a lower screen resolution than standard-definition television systems such 240p (320*240). It is used in handheld television. The most common source of LDTV programming is the Internet, where mass distribution of higher-resolution

video files could overwhelm computer servers and take too long to download. Many mobile phones and portable devices such as Apple's iPod Nano, or Sony's PlayStation Portable use LDTV video, as higher-resolution files would be excessive to the needs of their small screens (320×240 and 480×272 pixels respectively). The current generation of iPod Nanos has LDTV screens, as do the first three generations of iPod Touch and iPhone (480×320). For the first years of its existence, YouTube offered only one low-definition resolution of 320×240 p at 30fps or less. A standard, consumer-grade videotape can be considered SDTV due to its resolution (approximately 360×480 i/576i).



Comparison of 8K UHDTV, 4K UHDTV, HDTV and SDTV resolution

SD

Standard-definition television or SDTV refers to two different resolutions: 576i, with 576 interlaced lines of resolution, derived from the European-developed PAL and SECAM systems, and 480i based on the American National Television System Committee NTSC system. SDTV is a television system that uses a resolution that is not considered to be either high-definition television (720p, 1080i, 1080p, 1440p, 4K UHDTV, and 8K UHD) or enhanced-definition television (EDTV 480p). In North America, digital SDTV is broadcast in the same 4:3 aspect ratio as NTSC signals, with widescreen content being center cut.^[207] However, in other parts of the world that used the PAL or SECAM color systems, standard-definition television is now usually shown with a 16:9 aspect ratio, with the transition occurring between the mid-1990s and mid-2000s. Older programs with a 4:3 aspect ratio are shown in the United States as 4:3, with non-ATSC countries preferring to reduce the horizontal resolution by anamorphically scaling a pillarboxed image.

HD

High-definition television (HDTV) provides a resolution that is substantially higher than that of standard-definition television.

HDTV may be transmitted in various formats:

- 1080p: 1920×1080 p: 2,073,600 pixels (~2.07 megapixels) per frame
- 1080i: 1920×1080 i: 1,036,800 pixels (~1.04 MP) per field or 2,073,600 pixels (~2.07 MP) per frame
 - A non-standard CEA resolution exists in some countries such as 1440×1080 i: 777,600 pixels (~0.78 MP) per field or 1,555,200 pixels (~1.56 MP) per frame
- 720p: 1280×720 p: 921,600 pixels (~0.92 MP) per frame

UHD

Ultra-high-definition television (also known as Super Hi-Vision, Ultra HD television, UltraHD, UHDTV, or UHD) includes 4K UHD (2160p) and 8K UHD (4320p), which are two digital video formats proposed by NHK Science & Technology Research Laboratories and defined and approved by the International

Telecommunication Union (ITU). The Consumer Electronics Association announced on 17 October 2012 that "Ultra High Definition," or "Ultra HD," would be used for displays that have an aspect ratio of at least 16:9 and at least one digital input capable of carrying and presenting natural video at a minimum resolution of 3840×2160 pixels.^{[208][209]}

Market share

North American consumers purchase a new television set on average every seven years, and the average household owns 2.8 televisions. As of 2011, 48 million are sold each year at an average price of \$460 and size of 38 in (97 cm).^[210]

Worldwide TV manufacturers market share, H1 2023	
Manufacturer	Market share ^[211]
Samsung Electronics	31.2%
LG Electronics	16.2%
TCL	10.2%
Hisense	9.5%
Sony	5.7%
Others	39%

Content

Programming

Getting TV programming shown to the public can happen in many other ways. After production, the next step is to market and deliver the product to whichever markets are open to using it. This typically happens on two levels:

1. Original run or First run: a producer creates a program of one or multiple episodes and shows it on a station or network that has either paid for the production itself or granted a license by the television producers to do the same.
2. Broadcast syndication: this is the terminology rather broadly used to describe secondary programming usages (beyond the original run). It includes secondary runs in the country of the first issue, but also international usage, which may not be managed by the originating producer. In many cases, other companies, television stations, or individuals are engaged to do the syndication work, in other words, to sell the product into the markets they are allowed to sell into by contract from the copyright holders; in most cases, the producers.

First-run programming is increasing on subscription services outside of the United States, but few domestically produced programs are syndicated on domestic free-to-air (FTA) elsewhere. This practice is increasing, however, generally on digital-only FTA channels or with subscriber-only, first-run material

appearing on FTA. Unlike the United States, repeat FTA screenings of an FTA network program usually only occur on that network. Also, affiliates rarely buy or produce non-network programming that is not focused on local programming.

Genres

Television genres include a broad range of programming types that entertain, inform, and educate viewers. The most expensive entertainment genres to produce are usually dramas and dramatic miniseries. However, other genres, such as historical Western genres, may also have high production costs.

Pop culture entertainment genres include action-oriented shows such as police, crime, detective dramas, horror, or thriller shows. As well, there are also other variants of the drama genre, such as medical dramas and daytime soap operas. Sci-fi series can fall into either the drama or action category, depending on whether they emphasize philosophical questions or high adventure. Comedy is a popular genre that includes situation comedy (sitcom) and animated series for the adult demographic, such as Comedy Central's South Park.

The least expensive forms of entertainment programming genres are game shows, talk shows, variety shows, and reality television. Game shows feature contestants answering questions and solving puzzles to win prizes. Talk shows contain interviews with film, television, music, and sports celebrities and public figures. Variety shows feature a range of musical performers and other entertainers, such as comedians and magicians, introduced by a host or Master of Ceremonies. There is some crossover between some talk shows and variety shows because leading talk shows often feature performances by bands, singers, comedians, and other performers in between the interview segments. Reality television series "regular" people (i.e., not actors) facing unusual challenges or experiences ranging from arrest by police officers (COPS) to significant weight loss (The Biggest Loser). A derived version of reality shows depicts celebrities doing mundane activities such as going about their everyday life (The Osbournes, Snoop Dogg's Father Hood) or doing regular jobs (The Simple Life).^[212]

Fictional television programs that some television scholars and broadcasting advocacy groups argue are "quality television", include series such as Twin Peaks and The Sopranos. Kristin Thompson argues that some of these television series exhibit traits also found in art films, such as psychological realism, narrative complexity, and ambiguous plotlines. Nonfiction television programs that some television scholars and broadcasting advocacy groups argue are "quality television" include a range of serious, noncommercial programming aimed at a niche audience, such as documentaries and public affairs shows.

Funding

Around the world, broadcast television is financed by government, advertising, licensing (a form of tax), subscription, or any combination of these. To protect revenues, subscription television channels are usually encrypted to ensure that only subscribers receive the decryption codes to see the signal. Unencrypted channels are known as free-to-air or FTA. In 2009, the global TV market represented 1,217.2 million TV households with at least one TV and total revenues of 268.9 billion EUR (declining 1.2% compared to 2008).^[213] North America had the biggest TV revenue market share with 39%

followed by Europe (31%), Asia-Pacific (21%), Latin America (8%), and Africa and the Middle East (2%).^[214] Globally, the different TV revenue sources are divided into 45–50% TV advertising revenues, 40–45% subscription fees, and 10% public funding.^{[215][216]}

Advertising

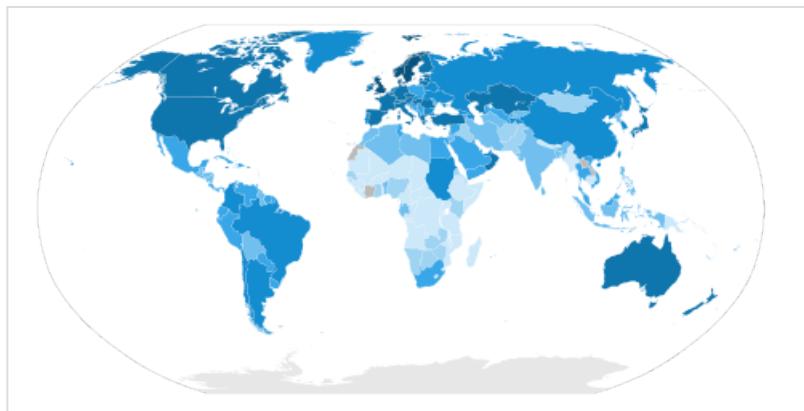
Television's broad reach makes it a powerful and attractive medium for advertisers. Many television networks and stations sell blocks of broadcast time to advertisers ("sponsors") to fund their programming.^[217]

Television advertisements (variously

called a television commercial, commercial, or ad in [American English](#), and known in [British English](#) as an advert) is a span of television programming produced and paid for by an organization, which conveys a message, typically to market a product or service. Advertising revenue provides a significant portion of the funding for most privately owned television networks. The vast majority of television advertisements today consist of brief advertising spots, ranging in length from a few seconds to several minutes (as well as program-length [infomercials](#)). Advertisements of this sort have been used to promote a wide variety of goods, services, and ideas since the beginning of television.

The effects of television advertising upon the viewing public (and the effects of mass media in general) have been the subject of discourse by philosophers, including [Marshall McLuhan](#). The viewership of television programming, as measured by companies such as [Nielsen Media Research](#), is often used as a metric for television advertisement placement and, consequently, for the rates charged to advertisers to air within a given network, television program, or time of day (called a "daypart"). In many countries, including the United States, television [campaign advertisements](#) is considered indispensable for a [political campaign](#). In other countries, such as France, political advertising on television is heavily restricted,^[218] while some countries, such as [Norway](#), completely ban political advertisements.

The first official, paid television advertisement was broadcast in the United States on 1 July 1941, over New York station WNBT (now [WNBC](#)) before a baseball game between the [Brooklyn Dodgers](#) and [Philadelphia Phillies](#). The announcement for [Bulova](#) watches, for which the company paid anywhere from \$4.00 to \$9.00 (reports vary), displayed a WNBT test pattern modified to look like a clock with the hands showing the time. The Bulova logo, with the phrase "Bulova Watch Time," was shown in the lower right-hand quadrant of the test pattern while the second hand swept around the dial for one minute.^{[219][220]} The first TV ad broadcast in the U.K. was on [ITV](#) on 22 September 1955, advertising [Gibbs SR](#) toothpaste. The first TV ad broadcast in Asia was on [Nippon Television](#) in Tokyo on 28 August 1953, advertising [Seikosha](#) (now [Seiko](#)), which also displayed a clock with the current time.^[221]



Television sets per 1000 people of the world

1000+	100–200
500–1000	50–100
300–500	0–50
200–300	No data

United States

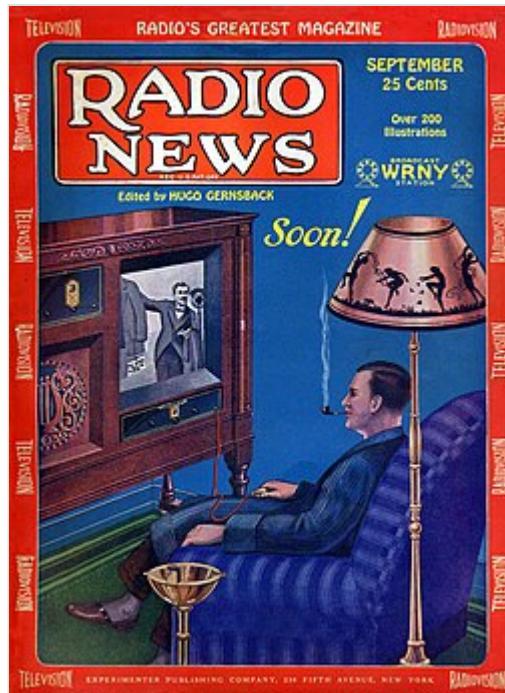
Since inception in the US in 1941,^[222] television commercials have become one of the most effective, persuasive, and popular methods of selling products of many sorts, especially consumer goods. During the 1940s and into the 1950s, programs were hosted by single advertisers. This, in turn, gave great creative control to the advertisers over the content of the show. Perhaps due to the quiz show scandals in the 1950s,^[223] networks shifted to the magazine concept, introducing advertising breaks with other advertisers.

U.S. advertising rates are determined primarily by Nielsen ratings. The time of the day and popularity of the channel determine how much a TV commercial can cost. For example, it can cost approximately \$750,000 for a 30-second block of commercial time during the highly popular singing competition American Idol, while the same amount of time for the Super Bowl can cost several million dollars. Conversely, lesser-viewed time slots, such as early mornings and weekday afternoons, are often sold in bulk to producers of infomercials at far lower rates. In recent years, paid programs or infomercials have become common, usually in lengths of 30 minutes or one hour. Some drug companies and other businesses have even created "news" items for broadcast, known in the industry as video news releases, paying program directors to use them.^[224]

Some television programs also deliberately place products into their shows as advertisements, a practice started in feature films^[225] and known as product placement. For example, a character could be drinking a certain kind of soda, going to a particular chain restaurant, or driving a certain make of car. (This is sometimes very subtle, with shows having vehicles provided by manufacturers for low cost in exchange as a product placement). Sometimes, a specific brand or trade mark, or music from a certain artist or group, is used. (This excludes guest appearances by artists who perform on the show.)

United Kingdom

The TV regulator oversees TV advertising in the United Kingdom. Its restrictions have applied since the early days of commercially funded TV. Despite this, an early TV mogul, Roy Thomson, likened the broadcasting license as being a "license to print money".^[226] Restrictions mean that the big three national commercial TV channels: ITV, Channel 4, and Channel 5 can show an average of only seven minutes of advertising per hour (eight minutes in the peak period). Other broadcasters must average no more than nine minutes (twelve in the peak). This means that many imported TV shows from the U.S. have unnatural pauses where the British company does not use the narrative breaks intended for more frequent U.S. advertising. Advertisements must not be inserted in the course of certain specific proscribed types of programs that last less than half an hour in scheduled duration; this list includes any news or current affairs programs, documentaries, and programs for children; additionally, advertisements may not be carried in a program designed and broadcast for reception in schools or in any religious broadcasting service or other devotional program or during a formal Royal ceremony or occasion. There also must be



Television was still in its experimental phase in 1928, but the medium's potential to sell goods was already predicted.

clear demarcations in time between the programs and the advertisements. The [BBC](#), being strictly non-commercial, is not allowed to show adverts on television in the U.K., though it has advertising-funded channels abroad. The majority of its budget comes from [television license fees](#) (see below) and [broadcast syndication](#), the sale of content to other broadcasters.^{[227][228]}

Ireland

Broadcast advertising is regulated by the [Broadcasting Authority of Ireland](#).^[229]

Subscription

Some TV channels are partly funded from [subscriptions](#); therefore, the signals are encrypted during the broadcast to ensure that only the paying subscribers have access to the decryption codes to watch [pay television](#) or [specialty channels](#). Most subscription services are also funded by advertising.

Taxation or license

Television services in some countries may be funded by a [television licence](#) or a form of taxation, which means that advertising plays a lesser role or no role at all. For example, some channels may carry no advertising at all and some very little, including:

- [Australia \(ABC Television\)](#)
- [Belgium \(VRT for Flanders and RTBF for Wallonia\)](#)
- [Denmark \(DR\)](#)
- [Ireland \(RTÉ\)](#)
- [Japan \(NHK\)](#)
- [Norway \(NRK\)](#)
- [Sweden \(SVT\)](#)
- [Switzerland \(SRG SSR\)](#)
- [Republic of China \(Taiwan\) \(PTS\)](#)
- [United Kingdom \(BBC Television\)](#)
- [United States \(PBS\)](#)

The [British Broadcasting Corporation's](#) TV service carries no [television advertising](#) on its UK channels and is funded by an annual television license paid by the occupiers of premises receiving live telecasts. As of 2012 it was estimated that approximately 26.8 million UK private domestic households owned televisions, with approximately 25 million TV licences in all premises in force as of 2010.^[230] This television license fee is set by the government, but the BBC is not answerable to or controlled by the government. As of 2009 two main BBC TV channels were watched by almost 90% of the population each week and overall had 27% share of total viewing,^[231] despite the fact that 85% of homes were multi-channel, with 42% of these having access to 200 free-to-air channels via satellite and another 43% having access to 30 or more channels via [Freeview](#).^[232] As of June 2021 the licence that funds the advertising-free BBC TV channels cost £159 for a colour TV Licence and £53.50 for a black and white TV Licence (free or reduced for some groups).^[233]

The [Australian Broadcasting Corporation's](#) television services in Australia carry no advertising by external sources; it is banned under the *Australian Broadcasting Corporation Act 1983*, which also ensures its editorial independence. The ABC receives most of its funding from the [Australian](#)

Government (some revenue is received from its Commercial division), but it has suffered progressive funding cuts under Liberal governments since the 1996 Howard government,^[234] with particularly deep cuts in 2014 under the Turnbull government,^[235] and an ongoing indexation freeze as of 2021.^{[236][237]} The funds provide for the ABC's television, radio, online, and international outputs, although ABC Australia, which broadcasts throughout the Asia-Pacific region, receives additional funds through DFAT and some advertising on the channel.^{[238][239]}

In France, government-funded channels carry advertisements, yet those who own television sets have to pay an annual tax ("la redevance audiovisuelle").^[240]

In Japan, NHK is paid for by license fees (known in Japanese as reception fee (受信料, *Jushinryō*)). The broadcast law that governs NHK's funding stipulates that any television equipped to receive NHK is required to pay. The fee is standardized, with discounts for office workers and students who commute, as well as a general discount for residents of Okinawa prefecture.

Broadcast programming

Broadcast programming, or TV listings in the United Kingdom, is the practice of organizing television programs in a schedule, with broadcast automation used to regularly change the scheduling of TV programs to build an audience for a new show, retain that audience, or compete with other broadcasters' programs.

Social aspects

Television has played a pivotal role in the socialization of the 20th and 21st centuries. There are many aspects of television that can be addressed, including negative issues such as media violence. Current research is discovering that individuals suffering from social isolation can employ television to create what is termed a parasocial or faux relationship with characters from their favorite television shows and movies as a way of deflecting feelings of loneliness and social deprivation.^[241] Several studies have found that educational television has many advantages. The article "The Good Things about Television"^[242] argues that television can be a very powerful and effective learning tool for children if used wisely. With respect to faith, many Christian denominations use television for religious broadcasting.



American family watching television,
c. 1958

Religious opposition

Methodist denominations in the conservative holiness movement, such as the Allegheny Wesleyan Methodist Connection and the Evangelical Wesleyan Church, eschew the use of the television.^[243] Some Baptists, such as those affiliated with Pensacola Christian College,^[244] also eschew television. Many

Traditional Catholic congregations such as the Society of Saint Pius X (SSPX), as with Laestadian Lutherans, and Conservative Anabaptists such as the Dunkard Brethren Church, oppose the presence of television in the household, teaching that it is an occasion of sin.^{[245][246][247][248]}

Negative impacts

Children, especially those aged five or younger, are at risk of injury from falling televisions.^[249] A CRT-style television that falls on a child will, because of its weight, hit with the equivalent force of falling multiple stories from a building.^[250] Newer flat-screen televisions are "top-heavy and have narrow bases", which means that a small child can easily pull one over.^[251] As of 2015, TV tip-overs were responsible for more than 10,000 injuries per year to children in the United States, at a cost of more than US\$8 million per year (equivalent to US\$10.61 million per year in 2024) in emergency care.^{[249][251]}

A 2017 study in *The Journal of Human Resources* found that exposure to cable television reduced cognitive ability and high school graduation rates for boys. This effect was stronger for boys from more educated families. The article suggests a mechanism where light television entertainment crowds out more cognitively stimulating activities.^[252]

With high lead content in CRTs and the rapid diffusion of new flat-panel display technologies, some of which (LCDs) use lamps which contain mercury, there is growing concern about electronic waste from discarded televisions. Related occupational health concerns exist, as well, for disassemblers removing copper wiring and other materials from CRTs. Further environmental concerns related to television design and use relate to the devices' increasing electrical energy requirements.^[253]

See also



Television portal

- [B-television](#)
- [Broadcast-safe](#)
- [Broadcast television systems](#)
- [Content discovery platform](#)
- [Information-action ratio](#)
- [List of countries by number of television broadcast stations](#)
- [List of television manufacturers](#)
- [List of years in television](#)
- [Lists of television channels](#)
- [Media psychology](#)
- [MicroLED](#)
- [Sign language on television](#)
- [Telephilia](#)
- [Television addiction](#)
- [Television studies](#)
- [Test card](#)
- [TV accessory](#)

References

1. Diggs-Brown, Barbara (2011) *Strategic Public Relations: Audience Focused Practice* (<http://books.google.com/books?id=7c0ycySng4YC&pg=PA53>) p. 48
2. "TVTechnology: The State of Television, Worldwide" (<https://web.archive.org/web/20150428042640/http://www.tvtechnology.com/opinions/0087/the-state-of-television-worldwide/222681>). Archived from the original (<http://www.tvtechnology.com/opinions/0087/the-state-of-television-worldwide/222681>) on 28 April 2015. Retrieved 22 March 2015.
3. "Folha de S.Paulo - TV de tela grande lidera vendas no país - 27/08/2010" (<https://www1.folha.uol.com.br/fsp/mercado/me2708201029.htm>). www1.folha.uol.com.br. Retrieved 25 November 2024.
4. Julie Jacobson (1 December 2012). "Mitsubishi Drops DLP Displays: Goodbye RPTVs Forever" (https://web.archive.org/web/20150326171718/http://www.cepro.com/article/mitsubishi_drops_dlp_displays_goodbye_rptvs_forever). Archived from the original (http://www.cepro.com/article/mitsubishi_drops_dlp_displays_goodbye_rptvs_forever/) on 26 March 2015. Retrieved 22 March 2015.
5. Marshall Honorof (28 October 2014). "LG's Exit May Herald End of Plasma TVs" (<http://www.tomsguide.com/us/lg-end-plasma-tvs,news-19840.html>). *Tom's Guide*. Retrieved 22 March 2015.
6. "LG Electronics expects the OLED TV market to gradually replace the LED TV market" (<http://www.oled-info.com/lg-electronics-expects-oled-tv-market-gradually-replace-led-tv-market>). Retrieved 22 March 2015.
7. "All of Sony's new Smart TVs run on Android TV" (<https://www.theverge.com/2015/1/5/7497383/sony-new-smart-tv-run-android-tv-ces-2015>). *The Verge*. 5 January 2015. Retrieved 22 March 2015.
8. "CES 2015: New Samsung Smart TVs Will Be Powered by Tizen OS" (<http://www.techtimes.com/articles/24117/20150103/ces-2015-new-samsung-smart-tvs-will-be-powered-by-tizen-os.htm>). *Tech Times*. 3 January 2015. Retrieved 22 March 2015.
9. "LG to show off webOS 2.0 smart TV at CES 2015" (<https://www.cnet.com/news/lg-to-show-off-webos-2-0-smart-tv-at-ces-2015/>). *CNET*. Retrieved 22 March 2015.
10. "Internet TV and The Death of Cable TV, really" (<https://techcrunch.com/2010/10/24/internet-tv-and-the-death-of-cable-tv-really/>). *Techcrunch.com*. 24 October 2010. Retrieved 17 January 2012.
11. "Online Etymology Dictionary" (<http://www.etymonline.com/index.php?term=television>). *Etymonline.com*. 30 December 1969. Retrieved 30 April 2016.
12. Thompson, Robert (3 October 2015). "1985: Television Transformed 1.0" (<https://www.nytimes.com/2015/10/05/business/media/1985-television-transformed-1-0.html>). *The New York Times*.
13. Shiers, George and May (1997), *Early Television: A Bibliographic Guide to 1940*. Taylor & Francis, pp. 13, 22. ISBN 978-0-8240-7782-2.
14. Shiers & Shiers, p. 13, 22.
15. Constantin PERSKYI (20 September 1907). "Télévision au moyen de l'électricité" (<https://www.histv.net/perskyi-1900>). *The New York Times Sunday Magazine*. Congrès Inographs by Telegraph. p. 7.
16. "Sending Photographs by Telegraph" (<https://query.nytimes.com/gst/abstract.html?res=9A07E4DA1331E733A25757C2A9649C946697D6CF>), *The New York Times*, Sunday Magazine, 20 September 1907, p. 7.

17. Henry de Varigny, "La vision à distance (<http://histv2.free.fr/rignoux/rignoux1909.htm>) Archived (<https://web.archive.org/web/20160303231305/http://histv2.free.fr/rignoux/rignoux1909.htm>) 3 March 2016 at the Wayback Machine", *L'illustration*, Paris, 11 December 1909, p. 451.
18. R. W. Burns, *Television: An International History of the Formative Years* (<https://books.google.com/books?id=gZcwhVyiMqsC&q=zworykin+rosing+selenium>), IET, 1998, p. 119. ISBN 0-85296-914-7.
19. Wilfred S. Ogden (December 1921). "How the World's First Wireless News-Picture Was Flashed Across the Atlantic Ocean, Paris get President Harding's portrait in twenty minutes" (<https://books.google.com/books?id=aSoDAAAAMBAJ&pg=PA21>). *The Popular Science Monthly*. Bonnier Corporation. pp. 21–22. ISSN 0161-7370 (<https://search.worldcat.org/issn/0161-7370>). Retrieved 2 July 2014.
20. "Current Topics and Events" (<https://doi.org/10.1038%2F115504a0>). *Nature*. 115 (2892): 504–508. 1925. Bibcode:1925Natur.115..504. (<https://ui.adsabs.harvard.edu/abs/1925Natur.115..504.>). doi:10.1038/115504a0 (<https://doi.org/10.1038%2F115504a0>). ISSN 0028-0836 (<https://search.worldcat.org/issn/0028-0836>).
21. Television 1873–1927, *Television: The Official Organ Of The Television Society*, Vo1, No1, March 1928, Television Press Ltd, London, p11.
22. "The 'Televisor' – Successful test of a new apparatus," *The Times* (London), 28 January 1926, p. 9. "First on a receiver in the same room and then on a portable receiver in another room, the visitors were shown recognizable reception of the movements of the dummy head and of a person speaking."
23. "John Logie Baird (1888–1946)" (https://www.bbc.co.uk/history/historic_figures/baird_logie.shtml#:~:text=On%2026%20January%201926%20he,the%20Baird%20Television%20Development%20Company.). BBC. Retrieved 7 April 2021.
24. Baird, J.L., "Television in 1932 (<http://www.bairdtelevision.com/1932.html>)", *BBC Annual Report*, 1933.
25. "Radio Shows Far Away Objects in Motion", *The New York Times*, 14 June 1925, p. 1.
26. Glinsky, Albert (2000). *Theremin: Ether Music and Espionage* (<https://archive.org/details/thereminethermus00glin>). Urbana, IL: University of Illinois Press. pp. 41 (<https://archive.org/details/thereminethermus00glin/page/41>)–45. ISBN 978-0-252-02582-2.
27. "Case Files: Francis Jenkins (Phantoscope)" (<https://web.archive.org/web/20200328180825/https://www.fi.edu/case-files/francis-jenkins-phantoscope>). *The Franklin Institute*. 27 May 2016. Archived from the original (<https://www.fi.edu/case-files/francis-jenkins-phantoscope>) on 28 March 2020. Retrieved 28 March 2020.
28. Abramson, Albert, *The History of Television, 1880 to 1941*, McFarland & Co., Inc., 1987, p. 101. ISBN 978-0-89950-284-7.
29. Brewster, Richard. "Early Electronic Television RCA TV Development: 1929–1949" (http://www.earlytelevision.org/rca_story_brewster.html). Early Television Museum. Retrieved 20 February 2016.
30. Kenjiro Takayanagi: *The Father of Japanese Television* (<http://www.nhk.or.jp/strl/aboutstrl/evolution-of-tv-en/p05/>) Archived (<https://web.archive.org/web/20160101180643/http://www.nhk.or.jp/strl/aboutstrl/evolution-of-tv-en/p05/>) 1 January 2016 at the Wayback Machine, NHK (Japan Broadcasting Corporation), 2002. Retrieved 23 May 2009.
31. McLean, Donald F. (2000). *Restoring Baird's Image* (<https://books.google.com/books?id=1ISFkhG6cZQC&pg=PA184>). IET. p. 184. ISBN 978-0-85296-795-9.
32. "VE9AK entry at" (<http://www.earlytelevision.org/ve9ak.html>). Earlytelevision.org. Retrieved 2 March 2010.
33. "Peck Television Corporation Console Receiver and Camera" (<http://www.earlytelevision.org/peck.html>). Early Television Museum. Retrieved 18 February 2012.

34. Ferdinand Braun (1897) "Ueber ein Verfahren zur Demonstration und zum Studium des zeitlichen Verlaufs variabler Ströme" (<https://archive.today/20141217172841/http://babel.hathitrust.org/cgi/pt?id=wu.89048352892;view=1up;seq=568>) (On a process for the display and study of the course in time of variable currents), *Annalen der Physik und Chemie*, 3rd series, **60** : 552–59.
35. Lehrer, Norman H. (1985). "The Challenge of the Cathode-Ray Tube". In Tannas, Lawrence E. Jr. (ed.). *Flat-Panel Displays and CRTS*. New York: Van Nostrand Reinhold Company Inc. pp. 138–76. doi:[10.1007/978-94-011-7062-8_6](https://doi.org/10.1007/978-94-011-7062-8_6) (https://doi.org/10.1007%2F978-94-011-7062-8_6). ISBN 978-94-011-7062-8.
36. "Karl Ferdinand Braun" (<https://www.lindahall.org/about/news/scientist-of-the-day/karl-ferdin-and-braun>). *The Linda Hall Library*.
37. Marcus, Laurence. "Television Timeline 1812–1923 – Television Heaven" (https://web.archive.org/web/20181017014615/http://www.televisionheaven.co.uk/television_timeline_1.htm). Archived from the original (http://www.televisionheaven.co.uk/television_timeline_1.htm) on 17 October 2018. Retrieved 11 November 2016.
38. "History of the Cathode Ray Tube" (<https://archive.today/20120709164537/http://inventors.about.com/od/cstartinventions/a/CathodeRayTube.htm>). *About.com*. Archived from the original (<http://inventors.about.com/od/cstartinventions/a/CathodeRayTube.htm>) on 9 July 2012. Retrieved 4 October 2009.
39. Campbell-Swinton, A. A. (18 June 1908). "Distant Electric Vision (first paragraph)" (<https://zenodo.org/record/1429503>). *Nature*. **78** (2016): 151. Bibcode:1908Natur..78..151S (<https://ui.adsabs.harvard.edu/abs/1908Natur..78..151S>). doi:[10.1038/078151a0](https://doi.org/10.1038/078151a0) (<https://doi.org/10.1038%2F078151a0>). S2CID 3956737 (<https://api.semanticscholar.org/CorpusID:3956737>).
40. Campbell-Swinton, A. A. (18 June 1908). "Distant Electric Vision" (<http://www.nature.com/nature/journal/v78/n2016/pdf/078151a0.pdf>) (PDF). *Nature*. **78** (2016): 151. Bibcode:1908Natur..78..151S (<https://ui.adsabs.harvard.edu/abs/1908Natur..78..151S>). doi:[10.1038/078151a0](https://doi.org/10.1038/078151a0) (<https://doi.org/10.1038%2F078151a0>). S2CID 3956737 (<https://api.semanticscholar.org/CorpusID:3956737>).
41. "Distant Electric Vision", *The Times* (London), 15 November 1911, p. 24b.
42. Bairdtelevision. "Alan Archivald Campbell-Swinton (1863–1930)" (<http://www.bairdtelevision.com/swinton.html>). *Biography*. Retrieved 10 May 2010.
43. Shiers, May (29 December 1997). *Early Television: A Bibliographic Guide to 1940* (<https://books.google.com/books?id=OIXsZdT8HUQC&q=swinton+rontgen>). Taylor & Francis. p. 56. ISBN 978-0-8240-7782-2 – via Google Books.
44. Campbell-Swinton, A. A. (23 October 1926). "Electric Television (abstract)" (<https://doi.org/10.1038%2F118590a0>). *Nature*. **118** (2973): 590. Bibcode:1926Natur.118..590S (<https://ui.adsabs.harvard.edu/abs/1926Natur.118..590S>). doi:[10.1038/118590a0](https://doi.org/10.1038/118590a0) (<https://doi.org/10.1038%2F118590a0>). S2CID 4081053 (<https://api.semanticscholar.org/CorpusID:4081053>).
45. Burns, R W. (1998). *Television: An International History of the Formative Years* (<https://books.google.com/books?id=gZcwhVyiMqsC&q=swinton+minchin+stanton+1903>). The Institute of Electrical Engineers (IEE) (History of Technology Series 22) in association with [The Science Museum (UK)]. p. 123. ISBN 978-0-85296-914-4.
46. G., R. A. (2 April 1914). "Prof. G.M. Minchin, F.R.S" (<https://doi.org/10.1038%2F093115a0>). *Nature*. **93** (2318): 115–16. Bibcode:1914Natur..93..115R (<https://ui.adsabs.harvard.edu/abs/1914Natur..93..115R>). doi:[10.1038/093115a0](https://doi.org/10.1038/093115a0) (<https://doi.org/10.1038%2F093115a0>).
47. Miller, H. & Strange. J. W. (2 May 1938). "The electrical reproduction of images by the photoconductive effect". *Proceedings of the Physical Society*. **50** (3): 374–84. Bibcode:1938PPS....50..374M (<https://ui.adsabs.harvard.edu/abs/1938PPS....50..374M>). doi:[10.1088/0959-5309/50/3/307](https://doi.org/10.1088/0959-5309/50/3/307) (<https://doi.org/10.1088%2F0959-5309%2F50%2F3%2F307>).

48. Iams, H. & Rose, A. (August 1937). "Television Pickup Tubes with Cathode-Ray Beam Scanning". *Proceedings of the Institute of Radio Engineers*. **25** (8): 1048–70.
doi:10.1109/JRPROC.1937.228423 (<https://doi.org/10.1109%2FJRPROC.1937.228423>).
S2CID 51668505 (<https://api.semanticscholar.org/CorpusID:51668505>).
49. Abramson, Albert, *Zworykin, Pioneer of Television*, p. 16.
50. "Kálmán Tihanyi's 1926 Patent Application Radioskop" (<https://www.unesco.org/en/memory-world/kalman-tihanyis-1926-patent-application-radioskop>). UNESCO Memory of the World Programme. Retrieved 22 April 2025.
51. United States Patent Office, Patent No. 2,133,123, 11 October 1938.
52. United States Patent Office, Patent No. 2,158,259, 16 May 1939
53. "Vladimir Kosma Zworykin, 1889–1982" (<http://www.bairdtelevision.com/zworykin.html>). Bairdtelevision.com. Retrieved 17 April 2009.
54. [1] (<http://www.iec.ch/about/history/techline/swf/temp.xml>) Archived (<https://web.archive.org/web/20150924033236/http://www.iec.ch/about/history/techline/swf/temp.xml>) 24 September 2015 at the Wayback Machine "Kálmán Tihanyi (1897–1947)", *IEC Techline*, International Electrotechnical Commission (IEC), 15 July 2009.
55. Tihanyi, Koloman, *Improvements in television apparatus* (http://v3.espacenet.com/publicationDetails/biblio?DB=EPODOC&adjacent=true&locale=en_V3&FT=D&date=19301111&CC=GB&NR=313456A&KC=A) Archived (https://web.archive.org/web/20221204171044/https://worldwide.espacenet.com/publicationDetails/biblio?DB=EPODOC&adjacent=true&locale=en_EP&FT=D&date=19301111&CC=GB&NR=313456A&KC=A) 4 December 2022 at the Wayback Machine. European Patent Office, Patent No. GB313456. Convention date UK application: 1928-06-11, declared void and published: 11 November 1930. Retrieved 25 April 2013.
56. "Patent US2133123 – Television apparatus" (<https://patents.google.com/patent/US2133123>). Retrieved 22 March 2015.
57. "Patent US2158259 – Television apparatus" (<https://patents.google.com/patent/US2158259>). Retrieved 22 March 2015.
58. "The Editor—to You". *Radio News*. **13** (12): 979. June 1932.
59. "Milestones:Development of Electronic Television, 1924–1941" (http://www.ieeeghn.org/wiki/index.php/Milestones:Development_of_Electronic_Television,_1924-1941). Retrieved 22 March 2015.
60. Hart, Hugh (28 January 2010). "Jan. 29, 1901: DuMont Will Make TV Work." (<https://www.wired.com/2010/01/jan-29-1901-dumont-will-make-tv-work-2/>) *Wired*. Retrieved 21 May 2021.
61. Postman, Neil, "Philo Farnsworth" (<https://web.archive.org/web/20000531100005/http://www.time.com/time/time100/scientist/profile/farnsworth.html>), *The TIME 100: Scientists & Thinkers, Time*, 29 March 1999. Retrieved 28 July 2009.
62. "Philo Taylor Farnsworth (1906–1971)" (<http://www.sfmuseum.org/hist10/philo.html>) Archived (<https://web.archive.org/web/20110622033654/http://www.sfmuseum.org/hist10/philo.html>) 22 June 2011 at the Wayback Machine, *The Virtual Museum of the City of San Francisco*. Retrieved 15 July 2009.
63. Abramson, Albert, *Zworykin, Pioneer of Television*, p. 226.
64. "The Philo T. and Elma G. Farnsworth Papers" (<https://web.archive.org/web/20080422211543/http://db3-sql.staff.library.utah.edu/lucene/Manuscripts/null/Ms0648.xml/complete>). Archived from the original (<http://db3-sql.staff.library.utah.edu/lucene/Manuscripts/null/Ms0648.xml/complete>) on 22 April 2008.
65. Abramson, Albert, *Zworykin, Pioneer of Television*, University of Illinois Press, 1995, p. 51.
ISBN 0-252-02104-5.

66. Zworykin, Vladimir K., *Television System* (<https://www.google.com/patents/about?id=mZ9KA> AAAEBAJ) Archived (<https://web.archive.org/web/20140131053220/http://www.google.com/patents/about?id=mZ9KAAAAEBAJ>) 31 January 2014 at the Wayback Machine. Patent No. 1691324, U.S. Patent Office. Filed 1925-07-13, issued 13 November 1928. Retrieved 28 July 2009
67. Zworykin, Vladimir K., *Television System* (<https://www.google.com/patents?id=tQt-AAAAEB> AJ) Archived (<https://web.archive.org/web/20130518183522/http://www.google.com/patent?id=tQt-AAAAEB>) 18 May 2013 at the Wayback Machine. Patent No. 2022450, U.S. Patent Office. Filed 1923-12-29, issued 26 November 1935. Retrieved 10 May 2010.
68. Stashower, Daniel, *The Boy Genius and the Mogul: The Untold Story of Television*, Broadway Books, 2002, pp. 243–44. ISBN 978-0-7679-0759-0.
69. Everson, George (1949), *The Story of Television, The Life of Philo T. Farnsworth* New York: W.W. Norton & Co., ISBN 978-0-405-06042-7, 266 pp.
70. Lawrence, Williams L. (27 June 1933). "Human-like eye made by engineers to televise images" (<https://books.google.com/books?id=OIXsZdT8HUQC&q=3971+zworykin+N.Y.T>). *The New York Times*. ISBN 978-0-8240-7782-2. Retrieved 10 January 2010. "'Iconoscope' converts scenes into electrical energy for radio transmission. Fast as a movie camera. Three million tiny photocells 'memorize,' then pass out pictures. Step to home television. Developed in ten years of work by Dr. V.K. Zworykin, who describes it in Chicago."
71. Abramson, Albert (1987), *The History of Television, 1880 to 1941*. Jefferson, NC: Albert Abramson. p. 148. ISBN 0-89950-284-9.
72. Everson, George (1949), *The Story of Television, The Life of Philo T. Farnsworth* New York: W.W. Norton & Co., ISBN 978-0-405-06042-7, pp. 137–41.
73. Everson, George (1949), *The Story of Television, The Life of Philo T. Farnsworth* New York: W.W. Norton & Co., ISBN 978-0-405-06042-7, p. 139.
74. Everson, George (1949), *The Story of Television, The Life of Philo T. Farnsworth* New York: W.W. Norton & Co., ISBN 978-0-405-06042-7, p. 141.
75. "Manfred von Ardenne" (<https://www.vonardenne.biz/ja/company/history/manfred-von-ardenne/>). VON ARDENNE Website.
76. Albert Abramson, *Zworykin: Pioneer of Television*, University of Illinois Press, 1995, p. 111.
77. "22.3.1935: Erstes Fernsehprogramm der Welt" (http://www.kalenderblatt.de/index.php?what=thmanu&lang=de&manu_id=1737&sdt=20090322&maca=de-podcast_kalenderblatt-1086-xml-mrss). Deutsche Welle. Retrieved 27 July 2015.
78. "Es begann in der Fernsehstube: TV wird 80 Jahre alt" (<https://web.archive.org/web/20190121012410/https://www.computerbild.de/artikel/avf-News-Fernseher-Es-begann-in-der-Fernsehstube-TV-wird-80-Jahre-alt-11525963.html>). Computer Bild. 22 March 2015. Archived from the original (<https://web.archive.org/web/20190121012410/https://www.computerbild.de/artikel/avf-News-Fernseher-Es-begann-in-der-Fernsehstube-TV-wird-80-Jahre-alt-11525963.html>) on 21 January 2019. Retrieved 28 April 2017.
79. "New Television System Uses 'Magnetic Lens'" (<https://books.google.com/books?id=yt8DAA> AAMBAJ&pg=PA838"), *Popular Mechanics*, Dec. 1934, pp. 838–39.
80. Burns, R.W. *Television: An international history of the formative years*. (1998). IEE History of Technology Series, 22. London: IEE, p. 370. ISBN 9780852969144.
81. "Patent US2296019 – Chromoscopic adapter for television equipment" (<https://patents.google.com/patent/US2296019>). Retrieved 22 March 2015.

82. EMI LTD; Tedham, William F. & McGee, James D. "Improvements in or relating to cathode ray tubes and the like" (https://web.archive.org/web/20211122002108/https://worldwide.espacenet.com/publicationDetails/biblio?DB=EPDOC&adjacent=true&locale=en_EP&FT=D&date=19340226&CC=GB&NR=406353A&KC=A). Patent No. GB 406,353 (filed May 1932, patented 1934). United Kingdom Intellectual Property Office. Archived from the original (http://v3.espacenet.com/publicationDetails/biblio?DB=EPDOC&adjacent=true&locale=en_V3&FT=D&date=19340226&CC=GB&NR=406353A&KC=A) on 22 November 2021. Retrieved 22 February 2010.
83. Tedham, William F. & McGee, James D. "Cathode Ray Tube" (<https://web.archive.org/web/20130523212804/http://www.google.com/patents/about?id=BYNaAAAAEBAJ>). Patent No. 2,077,422 (filed in Great Britain 1932, filed in USA 1933, patented 1937). United States Patent Office. Archived from the original (<https://www.google.com/patents/about?id=BYNaA AAAEBAJ>) on 23 May 2013. Retrieved 10 January 2010.
84. Burns, R.W., *Television: An international history of the formative years*. (1998). IEE History of Technology Series, 22. London: IEE, p. 576. ISBN 0-85296-914-7.
85. Winston, Brian (1986). *Misunderstanding media* (https://books.google.com/books?id=K_Rp AAAIAAJ&q=%22american+iconoscope%22+noisy). Harvard University Press. pp. 60–61. ISBN 978-0-674-57663-6. Retrieved 9 March 2010.
86. Winston, Brian (1998). *Media technology and society. A history: from the telegraph to the Internet* (https://books.google.com/books?id=TZOF_1GZRmYC&q=american+iconoscope+noisy). Routledge. p. 105. ISBN 978-0-415-14230-4. Retrieved 9 March 2010.
87. Alexander, Robert Charles (2000). *The inventor of stereo: the life and works of Alan Dower Blumlein* (<https://books.google.com/books?id=qRhx3UmYBz0C&q=super+emitron>). Focal Press. pp. 217–19. ISBN 978-0-240-51628-8. Retrieved 10 January 2010.
88. Burns, R.W. (2000). *The life and times of A.D. Blumlein* (<https://books.google.com/books?id=B2z2ONO7nBQC&q=blumlein+mcgee+cps+emitron+decelerating+zero>). IET. p. 181. ISBN 978-0-85296-773-7. Retrieved 5 March 2010.
89. Lubszynski, Hans Gerhard & Rodda, Sydney. "Improvements in or relating to television" (https://web.archive.org/web/20211122010440/https://worldwide.espacenet.com/publicationDetails/biblio?DB=EPDOC&adjacent=true&locale=en_EP&FT=D&date=19360212&CC=GB&NR=442666A&KC=A). Patent No. GB 442,666 (filed May 1934, patented 1936). United Kingdom Intellectual Property Office. Archived from the original (http://v3.espacenet.com/publicationDetails/biblio?DB=EPDOC&adjacent=true&locale=en_V3&FT=D&date=19360212&CC=GB&NR=442666A&KC=A) on 22 November 2021. Retrieved 15 January 2010.
90. Blumlein, Alan Dower & McGee, James Dwyer. "Improvements in or relating to television transmitting systems" (<https://web.archive.org/web/20211122012740/https://v3.espacenet.com/errorpages/error403.htm?reason=RequestBlacklisted&ip=207.241.229.151>). Patent No. GB 446,661 (filed August 1934, patented 1936). United Kingdom Intellectual Property Office. Archived from the original (http://v3.espacenet.com/publicationDetails/biblio?DB=EPDOC&adjacent=true&locale=en_V3&FT=D&date=19360504&CC=GB&NR=446661A&KC=A) on 22 November 2021. Retrieved 9 March 2010.
91. McGee, James Dwyer. "Improvements in or relating to television transmitting systems" (https://web.archive.org/web/20211122020916/https://worldwide.espacenet.com/publicationDetails/biblio?DB=EPDOC&adjacent=true&locale=en_EP&FT=D&date=19360505&CC=GB&NR=446664A&KC=A). Patent No. GB 446,664 (filed September 1934, patented 1936). United Kingdom Intellectual Property Office. Archived from the original (http://v3.espacenet.com/publicationDetails/biblio?DB=EPDOC&adjacent=true&locale=en_V3&FT=D&date=19360505&CC=GB&NR=446664A&KC=A) on 22 November 2021. Retrieved 9 March 2010.
92. Alexander, Robert Charles (2000). *The inventor of stereo: the life and works of Alan Dower Blumlein* (<https://books.google.com/books?id=qRhx3UmYBz0C&q=emitron+cenotaph+armistice>). Focal Press. p. 216. ISBN 978-0-240-51628-8. Retrieved 10 January 2010.

93. Inglis, Andrew F. (1990). *Behind the tube: a history of broadcasting technology and business* (<https://books.google.com/books?id=xiu4AAAAIAAJ&q=image-iconoscope+telefunken>). Focal Press. p. 172. ISBN 978-0-240-80043-1. Retrieved 15 January 2010.
94. Dieckmann, Max & Rudolf Hell. "Lichtelektrische Bildzerlegerröhre für Fernseher" (<https://web.archive.org/web/20211122022403/https://v3.espacenet.com/errorpages/error403.htm?reason=RequestBlacklisted&ip=207.241.231.164>). Patent No. DE 450,187 (filed 1925, patented 1927). Deutsches Reich Reichspatentamt. Archived from the original (http://v3.espacenet.com/publicationDetails/originalDocument?CC=DE&NR=450187C&KC=C&FT=D&date=19271003&DB=EPDOC&locale=en_V3) on 22 November 2021. Retrieved 28 July 2009.
95. Farnsworth, Philo T. "Television System" (<https://web.archive.org/web/20130523192021/http://www.google.com/patents/about?id=HRd5AAAAEBAJ>). Patent No. 1,773,980 (filed 1927, patented 1930). United States Patent Office. Archived from the original (<https://www.google.com/patents/about?id=HRd5AAAAEBAJ>) on 23 May 2013. Retrieved 28 July 2009.
96. de Vries, M.J.; de Vries, Marc; Cross, Nigel & Grant, Donald P. (1993). *Design methodology and relationships with science, Número 71 de NATO ASI series* (https://books.google.com/books?id=4T8U_J1h7noC&q=image-iconoscope+image-orthicon+telefunken). Springer. p. 222. ISBN 978-0-7923-2191-0. Retrieved 15 January 2010.
97. Smith, Harry (July 1953). "Multicon – A new TV camera tube" (<https://web.archive.org/web/20100318011743/http://www.earlytelevision.org/multicon.html>). newspaper article. Early Television Foundation and Museum. Archived from the original (<http://www.earlytelevision.org/multicon.html>) on 18 March 2010. Retrieved 15 January 2010.
98. Gittel, Joachim (11 October 2008). "Spezialröhren" (<http://www.jogis-roehrenbude.de/Roehren-Geschichtliches/Spezialroehren/Spezialroehren.htm>). photographic album. Jogis Röhrenbude. Retrieved 15 January 2010.
99. Early Television Museum. "TV Camera Tubes, German "Super Iconoscope" (1936)" (https://web.archive.org/web/20110617080126/http://www.earlytelevision.org/prewar_camera_tubes.html). photographic album. Early Television Foundation and Museum. Archived from the original (http://www.earlytelevision.org/prewar_camera_tubes.html) on 17 June 2011. Retrieved 15 January 2010.
100. Gittel, Joachim (11 October 2008). "FAR-Röhren der Firma Heimann" (http://www.jogis-roehrenbude.de/Roehren-Geschichtliches/Spezialroehren/Ikonoskop_Heimann/Heimann.htm). photographic album. Jogis Röhrenbude. Retrieved 15 January 2010.
101. Philips (1958). "5854, Image Iconoscope, Philips". *electronic tube handbook* (http://www.jogis-roehrenbude.de/Roehren-Geschichtliches/Spezialroehren/Ikonoskop_Heimann/5854_Philips_Iconoscop-1958.pdf) (PDF). Philips. Archived (https://web.archive.org/web/20060903223404/http://www.jogis-roehrenbude.de/Roehren-Geschichtliches/Spezialroehren/Ikonoskop_Heimann/5854_Philips_Iconoscop-1958.pdf) (PDF) from the original on 3 September 2006. Retrieved 15 January 2010.
102. Everson, George (1949), *The Story of Television, The Life of Philo T. Farnsworth* New York: W.W. Norton & Co., ISBN 978-0-405-06042-7, p. 248.
103. Abramson, Albert (1987), *The History of Television, 1880 to 1941*. Jefferson, NC: Albert Abramson. p. 254. ISBN 0-89950-284-9.
104. Schatzkin, Paul (2002), *The Boy Who Invented Television*. Silver Spring, Maryland: Teamcom Books, pp. 187–88. ISBN 1-928791-30-1.
105. "Go-Ahead Signal Due for Television", *The New York Times*, 25 April 1941, p. 7.
106. "An Auspicious Beginning", *The New York Times*, 3 August 1941, p. X10.
107. "Benjamin Adler, 86, An Early Advocate of UHF Television" (<https://www.nytimes.com/1990/04/18/obituaries/benjamin-adler-86-an-early-advocate-of-uhf-television.html>). *The New York Times*. 18 April 1990.

108. "ePoly Briefs Home" (https://web.archive.org/web/20160304225948/http://archive.poly.edu/poly_ebriefs/archives/Feb03.htm). Archived from the original (http://archive.poly.edu/poly_ebriefs/archives/Feb03.htm) on 4 March 2016. Retrieved 11 October 2016.
109. "On the beginning of broadcast in 625 lines 60 years ago" (http://625.625-net.ru/files/587/511/h_665921be9883776271895912fb8bb262), 625 magazine (in Russian). Archived (https://web.archive.org/web/20160304131236/http://625.625-net.ru/files/587/511/h_665921be9883776271895912fb8bb262) 4 March 2016 at the Wayback Machine
110. "M.I. Krivocheev – an engineer's engineer" (https://web.archive.org/web/20041230091501/https://www.ebu.ch/en/technical/trev/trev_255-portrait.pdf), *EBU Technical Review*, Spring 1993.
111. ""In the Vanguard of Television Broadcasting"" (<https://web.archive.org/web/20070221210300/http://cra.ir/FTD/Static/RRC/RRCFile10.pdf>) (PDF). Archived from the original (<http://cra.ir/FTD/Static/RRC/RRCFile10.pdf>) (PDF) on 21 February 2007.
112. [2] (https://ewh.ieee.org/r2/johnstown/downloads/20090217_IEEE_JST_Trivia_Answers.pdf)
113. [3] (http://www.scitech.mtesz.hu/52tihanyi/flat-panel_tv_en.pdf) Archived (https://web.archive.org/web/20120314070853/http://www.scitech.mtesz.hu/52tihanyi/flat-panel_tv_en.pdf) 14 March 2012 at the Wayback Machine
114. Childs, William R.; Martin, Scott B.; Stitt-Gohdes, Wanda (2004). *Business and Industry: Savings and investment options to telecommuting* (<https://books.google.com/books?id=nCwYAAAAIAAJ>). Marshall Cavendish. p. 1217. ISBN 9780761474395. "In 1952 Ibuka toured AT&T's Bell Laboratories in the United States and saw the newly invented transistor. He realized that replacing the large, clumsy vacuum tube with the transistor would make possible smaller, more portable radios and TVs."
115. "Sony Founder Masaru Ibuka's New Year's Dream Comes True: The Launch of Sony's TV Business" (<https://www.sony.net/SonyInfo/CorporateInfo/History/capsule/21/index.html>). *Time Capsule*. 21. Sony. 17 November 2009. Retrieved 1 October 2019.
116. Sparke, Penny (2009). *Japanese Design* (<https://books.google.com/books?id=WaIQf2gV8pEC&pg=PA18>). The Museum of Modern Art. p. 18. ISBN 9780870707391.
117. Lucie-Smith, Edward (1983). *A History of Industrial Design* (<https://books.google.com/books?id=fMsfAQAAIAAJ>). Phaidon Press. p. 208. ISBN 9780714822815. "The first all-transistor television set was introduced by Sony in 1959 (fig. 386), only four years after their all-transistor radio, and started the transformation of television from something used for communal viewing, as the radio in the 30s had been a focus for communal listening, into an object of solitary contemplation."
118. Chang, Yoon Seok; Makatsoris, Harris C.; Richards, Howard D. (2007). *Evolution of Supply Chain Management: Symbiosis of Adaptive Value Networks and ICT* (<https://books.google.com/books?id=5Y3wBwAAQBAJ&pg=PA48>). Springer Science & Business Media. ISBN 9780306486968.
119. M. Le Blanc, "Etude sur la transmission électrique des impressions lumineuses", *La Lumière Electrique*, vol. 11, 1 December 1880, pp. 477–81.
120. R.W. Burns, *Television: An International History of the Formative Years*, IET, 1998, p. 98. ISBN 0-85296-914-7.
121. Western technology and Soviet economic development: 1945 to 1965, by Antony C. Sutton, Business & Economics – 1973, p. 330
122. The History of Television, 1880–1941, by Albert Abramson, 1987, p. 27
123. "A. Rokhlin, Tak rozhdalos' dal'novidenie (in Russian)" (https://web.archive.org/web/20130424162531/http://www.tvmuseum.ru/attach.asp?a_no=1018). Archived from the original on 24 April 2013.
124. John Logie Baird, *Television Apparatus and the Like* (<https://www.google.com/patents?id=JRVAaaaaEBAJ>) Archived (<https://web.archive.org/web/20130518084511/http://www.google.com/patents?id=JRVAaaaaEBAJ>) 18 May 2013 at the Wayback Machine, U.S. patent, filed in U.K. in 1928.

125. Baird Television: [Crystal Palace Television Studios](http://www.bairdtelevision.com/crystalpalace.html) (<http://www.bairdtelevision.com/crystalpalace.html>). Previous color television demonstrations in the U.K. and U.S. had been via closed circuit.
126. "The World's First High Definition Colour Television System" (<http://www.bairdtelevision.com/colour.html>). Retrieved 22 March 2015.
127. Peter C. Goldmark, assignor to Columbia Broadcasting System, "Color Television", U.S. Patent 2,480,571 (<https://patents.google.com/patent/US2480571>), filed 7 September 1940.
128. Current Broadcasting 1940
129. "Color Television Success in Test", *The New York Times*, 30 August 1940, p. 21.
130. "Color Television Achieves Realism", *The New York Times*, 5 September 1940, p. 18.
131. "New Television System Transmits Images in Full Color (<https://books.google.com/books?id=JScDAAAAMBAJ&pg=PA120>)", *Popular Science*, December 1940, p. 120.
132. "CBS Demonstrates Full-Color Television," *The Wall Street Journal*, 5 September 1940, p. 1. "Television Hearing Set," *The New York Times*, 13 November 1940, p. 26.
133. Ed Reitan, [RCA-NBC Color Firsts in Television \(commented\)](http://colortelevision.info/rca-nbc_firsts.html) (http://colortelevision.info/rca-nbc_firsts.html) Archived (https://web.archive.org/web/20150204092411/http://colortelevision.info/rca-nbc_firsts.html) 4 February 2015 at the Wayback Machine.
134. "Making of Radios and Phonographs to End April 22", *The New York Times*, 8 March 1942, p. 1. "Radio Production Curbs Cover All Combinations," *The Wall Street Journal*, 3 June 1942, p. 4. "WPB Cancels 210 Controls; Radios, Trucks in Full Output", *New York Times*, 21 August 1945, p. 1.
135. Bob Cooper, "Television: The Technology That Changed Our Lives (http://www.earlytelevision.org/color_tv_cooper.html)", Early Television Foundation.
136. Albert Abramson, *The History of Television, 1942 to 2000*, McFarland & Company, 2003, pp. 13–14. ISBN 0-7864-1220-8
137. Baird Television: [The World's First High Definition Colour Television System](http://www.bairdtelevision.com/colour.html) (<http://www.bairdtelevision.com/colour.html>).
138. National Television System Committee (1951–1953), [Report and Reports of Panel No. 11, 11-A, 12–19, with Some supplementary references cited in the Reports, and the Petition for adoption of transmission standards for color television before the Federal Communications Commission, n.p., 1953], 17 v. illus., diagrams., tables. 28 cm. LC Control No.:54021386 Library of Congress Online Catalog (<http://catalog.loc.gov/cgi-bin/Pwebrecon.cgi?DB=local&PAGE=First>)
139. "GE Portacolor" (http://www.earlytelevision.org/ge_portacolor.html). *Early Television Museum*. Retrieved 2 October 2019.
140. Tyson, Kirk (1996). *Competition in the 21st Century* (<https://archive.org/details/competitionin21s00tys0>). CRC Press. p. 253 (<https://archive.org/details/competitionin21s00tys0/page/253>). ISBN 9781574440324.
141. "HDTV Set Top Boxes and Digital TV Broadcast Information" (<http://arquivo.pt/wayback/20160522191336/http://www.disabled%2Dworld.com/artman/publish/digital%2Dhdtv.shtml>). Archived from the original ([https://www.disabled-world.com/artman/publish/digital-hdtv.shtml](http://www.disabled-world.com/artman/publish/digital-hdtv.shtml)) on 22 May 2016. Retrieved 28 June 2014.
142. Kruger, Lennard G.; Guerrero, Peter F. (2002). *Digital Television: An Overview* (<https://books.google.com/books?id=BIAfWq2V3wgC&pg=PA1>). Hauppauge, New York: Nova Publishers. p. 1. ISBN 9781590335024.
143. "The Origins and Future Prospects of Digital Television" (http://www.benton.org/initiatives/obligations/chartsing_the_digital_broadcasting_future/sec1). 22 December 2008. Retrieved 22 March 2015.
144. Lea, William (1994). *Video on demand: Research Paper 94/68* (<https://researchbriefings.parliament.uk/ResearchBriefing/Summary/RP94-68>). House of Commons Library. Retrieved 20 September 2019.

145. Barbero, M.; Hofmann, H.; Wells, N. D. (14 November 1991). "DCT source coding and current implementations for HDTV" (https://tech.ebu.ch/publications/trev_251-barbero). *EBU Technical Review* (251). European Broadcasting Union: 22–33. Retrieved 4 November 2019.
146. "Information about interlaced and progressive scan signals" (<https://web.archive.org/web/20090816184950/http://www.bambooav.com/information-about-interlaced-and-progressive-scan-signals.html>). Archived from the original (<http://www.bambooav.com/information-about-interlaced-and-progressive-scan-signals.html>) on 16 August 2009. Retrieved 22 March 2015.
147. "What's the Difference between "Interlaced" and "Progressive" Video? – ISF Forum" (<http://www.isfforum.com/FAQs/view/All-About-HDTV/What-s-the-Difference-between-Interlaced-and-Progressive-Video/33.html>).
148. "The history and politics of DTV" (http://www.cinemasource.com/articles/hist_politics_dtv.pdf) (PDF). p. 13. Archived (https://web.archive.org/web/20030322131735/http://www.cinemasource.com/articles/hist_politics_dtv.pdf) (PDF) from the original on 22 March 2003.
149. Steve Kovach (8 December 2010). "What Is A Smart TV?" (<http://www.businessinsider.com/what-is-a-smart-tv-2010-12>). *Business Insider*. Retrieved 17 January 2012.
150. Carmi Levy Special to the Star (15 October 2010). "Future of television is online and on-demand" (<https://www.thestar.com/business/media/article/876278--future-of-television-is-online-and-on-demand>). *Toronto Star*. Retrieved 17 January 2012.
151. Toeman, Jeremy (20 October 2010). "Why Connected TVs Will Be About the Content, Not the Apps" (<http://mashable.com/2010/10/20/connected-tv-content-not-apps/>). *Mashable*. Retrieved 7 October 2024.
152. "espacenet – Original document" (https://web.archive.org/web/20210224135932/https://worldwide.espacenet.com/publicationDetails/originalDocument?FT=D&date=19960510&DB=EPODOC&locale=en_EP&CC=FR&NR=2726670A1&KC=A1&ND=3). Worldwide.espacenet.com. Archived from the original (http://worldwide.espacenet.com/publicationDetails/originalDocument?FT=D&date=19960510&DB=EPODOC&locale=en_EP&CC=FR&NR=2726670A1&KC=A1&ND=3) on 24 February 2021. Retrieved 17 January 2012.
153. "espacenet – Bibliographic data" (<https://web.archive.org/web/20150904022930/http://worldwide.espacenet.com/publicationDetails/biblio?CC=US&NR=5905521A&KC=A&FT=D&ND=7&date=19990518&DB=EPODOC>). Worldwide.espacenet.com. Archived from the original (<http://worldwide.espacenet.com/publicationDetails/biblio?CC=US&NR=5905521A&KC=A&FT=D&ND=7&date=19990518&DB=EPODOC>) on 4 September 2015. Retrieved 17 January 2012.
154. Kats, Rimma (15 November 2018). "How Many Households Own a Smart TV?" (<https://www.emarketer.com/content/how-many-households-own-a-smart-tv>) *eMarketer*. Retrieved 21 May 2021.
155. "How Stereoscopic Television is Shown" (<http://www.bairdtelevision.com/stereo.html>). Baird Television website. Archived (<https://web.archive.org/web/20101019045734/http://www.bairdtelevision.com/stereo.html>) from the original on 19 October 2010. Retrieved 18 September 2010.
156. "3D TV-sales growth" (<https://web.archive.org/web/20130724082049/http://www.globalpost.com/dispatch/news/yonhap-news-agency/130318/3d-tv-sales-growth>). *globalpost.com*. 18 March 2013. Archived from the original (<http://www.globalpost.com/dispatch/news/yonhap-news-agency/130318/3d-tv-sales-growth>) on 24 July 2013.
157. "Future looks flat for 3D TV" (<https://www.smh.com.au/digital-life/digital-life-news/future-looks-flat-for-3d-tv-20130814-2rv1q.html>). *The Sydney Morning Herald*. 15 August 2013.
158. Chris Welch (12 June 2013). "Is 3D TV dead? ESPN 3D to shut down by end of 2013" (<http://www.theverge.com/2013/6/12/4422874/espn-3d-to-shut-down-by-end-of-2013>). *The Verge*.
159. Guy Walters (25 September 2014). "Why 3D TV is such a turn-off" (<http://www.iol.co.za/scitech/technology/software/why-3d-tv-is-such-a-turn-off-1.1755980>). *Iol Scitech*.

160. Donovan Jackson (29 September 2014). "Is 3D dead...again?" (<https://techday.com/netguide/news/is-3d-deadagain/192995/>). Techday.
161. Hannah Furness (17 September 2014). "3D TV falls further out of favour as Sky omits Premier League matches from schedule" (<https://www.telegraph.co.uk/culture/tvandradio/1102038/3D-TV-falls-further-out-of-favour-as-Sky-omits-Premier-League-matches-from-schedule.html>). *The Telegraph*. Archived (<https://ghostarchive.org/archive/20220110/https://www.telegraph.co.uk/culture/tvandradio/1102038/3D-TV-falls-further-out-of-favour-as-Sky-omits-Premier-League-matches-from-schedule.html>) from the original on 10 January 2022.
162. " "The First Television Show" *Popular Mechanics*, August 1930, pp. 177–79" (<https://books.google.com/books?id=iuIDAAAAMBAJ&dq=Popular+Science+1930+plane+%22Popular+Mechanics%22&pg=PA177>). Hearst Magazines. August 1930.
163. Laurence Marcus. "The History of the BBC: The First TV Era" (<https://web.archive.org/web/20110813233820/http://www.teletronic.co.uk/tvera.htm>). Archived from the original (<http://www.teletronic.co.uk/tvera.htm>) on 13 August 2011. Retrieved 22 March 2015.
164. "CEA Study Says Seven Percent of TV Households Use Antennas" (<http://www.tvtechnology.com/article/cea-study-says-seven-percent-of-tv-households-use-antennas-/220585>), *TVTechnology*, 30 July 2013 Archived (<https://web.archive.org/web/20141217104116/http://www.tvtechnology.com/article/cea-study-says-seven-percent-of-tv-households-use-antennas-/220585>) 17 December 2014 at the Wayback Machine
165. "Nielsen: Broadcast Reliance Grew in 2012" (<http://www.tvtechnology.com/article/nicelsen-broadcast-reliance-grew-in--/217217>), *TVTechnology*, 14 January 2013 Archived (<https://web.archive.org/web/20141218160139/http://www.tvtechnology.com/article/nicelsen-broadcast-reliance-grew-in--/217217>) 18 December 2014 at the Wayback Machine
166. "History of Cable" (<http://www.calicable.org/learn/history-of-cable/>). California Cable and Telecommunications Association. Retrieved 20 February 2016.
167. Antipolis, Sophia (September 1997). Digital Video Broadcasting (DVB); Implementation of Binary Phase Shift Keying (BPSK) modulation in DVB satellite transmission systems (http://www.etsi.org/deliver/etsi_tr/101100_101199/101198/01.01.01_60/tr_101198v010101p.pdf) (PDF) (Report). European Telecommunications Standards Institute. pp. 1–7. TR 101 198. Archived (https://web.archive.org/web/20120302190812/http://www.etsi.org/deliver/etsi_tr/101100_101199/101198/01.01.01_60/tr_101198v010101p.pdf) (PDF) from the original on 2 March 2012. Retrieved 20 July 2014.
168. "Frequency letter bands" (<https://web.archive.org/web/20140714171156/http://www.microwaves101.com/ENCYCLOPEDIA/letterbands.cfm>). *Microwaves101.com*. 25 April 2008. Archived from the original (<http://www.microwaves101.com/encyclopedia/letterbands.cfm>) on 14 July 2014. Retrieved 25 December 2014.
169. "Installing Consumer-Owned Antennas and Satellite Dishes" (<https://www.fcc.gov/cgb/consumerfacts/consumerdish.html>). FCC. Retrieved 21 November 2008.
170. Campbell, Dennis; Cotter, Susan (1998). *Copyright Infringement* (<https://books.google.com/books?id=kRJCwqzE3iwC&pg=PA161>). Kluwer Law International. ISBN 978-90-247-3002-5. Retrieved 18 September 2014.
171. "The Arthur C. Clarke Foundation" (<https://web.archive.org/web/20110716135721/http://www.clarkefoundation.org/archives/1996.php>). Archived from the original (<http://www.clarkefoundation.org/archives/1996.php>) on 16 July 2011. Retrieved 1 June 2016.
172. Campbell, Richard; Martin, Christopher R.; Fabos, Bettina (23 February 2011). *Media and Culture: An Introduction to Mass Communication* (<https://books.google.com/books?id=WuqjReIz4TcC&pg=PA152>). London, UK: Macmillan Publishers. p. 152. ISBN 978-1-4576-2831-3. Retrieved 15 August 2014.
173. "The 1945 Proposal by Arthur C. Clarke for Geostationary Satellite Communications" (<http://akdiva.org/clarke/1945ww>). Retrieved 22 March 2015.

174. *Wireless technologies and the national information infrastructure* (<https://books.google.com/books?id=S2f8q2c0R5cC&pg=PA138>). DIANE Publishing. September 1995. p. 138. ISBN 978-0-16-048180-2. Retrieved 15 August 2014.
175. Klein, Christopher (23 July 2012). "The Birth of Satellite TV, 50 Years Ago" (<http://www.history.com/news/the-birth-of-satellite-tv-50-years-ago>). History.com. History Channel. Retrieved 5 June 2014.
176. "Relay 1" (<https://nssdc.gsfc.nasa.gov/nmc/spacecraft/display.action?id=1962-068A>). NASA.gov. NASA.
177. Darcey, RJ (16 August 2013). "Syncom 2" (<https://nssdc.gsfc.nasa.gov/nmc/spacecraft/display.action?id=1963-031A>). NASA.gov. NASA. Retrieved 5 June 2014.
178. "Encyclopedia Astronautica – Intelsat I" (<https://web.archive.org/web/20100116224556/http://astronautix.com/craft/intlsat1.htm>). Archived from the original (<http://www.astronautix.com/craft/intlsat1.htm>) on 16 January 2010. Retrieved 5 April 2010.
179. "Soviet-bloc Research in Geophysics, Astronomy, and Space" (<https://books.google.com/books?id=pyRx2ssMQ6MC>) (Press release). Springfield Virginia: U.S. Joint Publications Research Service. 1970. p. 60. Retrieved 16 December 2014.
180. Robertson, Lloyd (9 November 1972). "Anik A1 launching: bridging the gap" (<http://archives.cbc.ca/500f.asp?id=1-75-92-594>). CBC English TV. Retrieved 25 January 2007.
181. Ezell, Linda N. (22 January 2010). "NASA – ATS" (https://web.archive.org/web/20130406164505/http://www.nasa.gov/centers/goddard/missions/ats_prt.htm). Nasa.gov. NASA. Archived from the original (https://www.nasa.gov/centers/goddard/missions/ats_prt.htm) on 6 April 2013. Retrieved 1 July 2014.
182. Long Distance Television Reception (TV-DX) For the Enthusiast, Roger W. Bunney, ISBN 0-900162-71-6
183. "Ekran" (<https://web.archive.org/web/20131112153322/http://www.astronautix.com/craft/ekran.htm>). Astronautix.com. Astronautix. 2007. Archived from the original (<http://www.astronautix.com/craft/ekran.htm>) on 12 November 2013. Retrieved 1 July 2014.
184. "Ekran (11F647)" (https://space.skyrocket.de/doc_sdat/ekran.htm). Gunter's Space Page.
185. Joshua Brustein (January 5, 2015). "Dish's New Sling TV Service Could Free You From Cable" (<https://web.archive.org/web/20150106181233/http://www.businessweek.com/articles/2015-01-05/dish-sling-tv-service-takes-on-cable-television>). Bloomberg Businessweek. Bloomberg LP. Retrieved January 12, 2015.
186. Spangler, Todd (2016-11-18). "AT&T Sets DirecTV Now Launch Event for Nov. 28" (<https://variety.com/2016/digital/news/att-directv-now-launch-nov-28-1201921813/>). Variety. Retrieved 2017-12-05.
187. "LG to show off webOS 2.0 smart TV at CES 2015" (<https://www.cnet.com/news/lg-to-show-off-webos-2-0-smart-tv-at-ces-2015/>). CNET. Retrieved 5 December 2017.
188. Stelter, Brian (18 July 2013). "Netflix Does Well in 2013 Primetime Emmy Nominations" (<https://artsbeat.blogs.nytimes.com/2013/07/18/watching-for-the-2013-primetime-emmy-nominations/>). The New York Times. Retrieved 18 July 2013.
189. Spangler, Todd (13 July 2015). "Comcast Aims at 'Cord-Never' with \$15 HBO, Basic TV Internet-Streaming Bundle" (<https://variety.com/2015/digital/news/comcast-hbo-basic-tv-internet-bundle-1201538854/>). Variety. Archived (<https://web.archive.org/web/20171001082945/http://variety.com/2015/digital/news/comcast-hbo-basic-tv-internet-bundle-1201538854/>) from the original on 1 October 2017.
190. "YouTube TV – Watch & DVR Live Sports, Shows & News" (<https://tv.youtube.com/>). YouTube TV – Watch & DVR Live Sports, Shows & News. Retrieved 2017-12-05.
191. Rajan, Amol (5 August 2020). "TV watching and online streaming surge during lockdown" (<https://www.bbc.com/news/entertainment-arts-53637305>). BBC News.

192. Levin, Gary; Lawler, Kelly; Keveney, Bill; Jensen, Erin (11 March 2021). "How pandemic changed TV, and how much of it will last" (<https://www.usatoday.com/story/entertainment/tv/2021/03/10/how-pandemic-changed-tv-and-how-much-last/6826073002/>). *USA Today*.
193. "About 6 in 10 young adults in U.S. primarily use online streaming to watch TV" (<https://www.pewresearch.org/fact-tank/2017/09/13/about-6-in-10-young-adults-in-u-s-primarily-use-online-streaming-to-watch-tv/>). Pew Research Center. 2017-09-13. Retrieved 2017-12-05.
194. Lovejoy, Ben (21 June 2022). "Video streaming services see end of pandemic boom" (<https://9to5mac.com/2022/06/21/video-streaming-services-2/>). *9to5Mac*.
195. "Netflix's Growth Is in the Eye of the Beholder" (<https://www.bloomberg.com/opinion/articles/2018-01-22/netflix-earnings-growth-is-in-the-eye-of-the-beholder>). *Bloomberg.com*. 22 January 2018. Retrieved 8 September 2020.
196. Balakrishnan, Anita (22 January 2018). "Netflix jumps more than 8% after adding more subscribers than expected" (<https://www.cnbc.com/2018/01/22/netflix-earnings-q4-2017.html>). *CNBC*.
197. Fosty, Vincent; Houben, Tim (March 2020). "Online media streaming will benefit from the coronavirus pandemic" (<https://web.archive.org/web/20230812100556/https://www2.deloitte.com/be/en/pages/technology-media-and-telecommunications/articles/online-media-streaming.html>). *Deloitte Belgium*. Archived from the original (<https://www2.deloitte.com/be/en/pages/technology-media-and-telecommunications/articles/online-media-streaming.html>) on 12 August 2023. Retrieved 4 November 2023. "... if it does not last too long!"
198. Dayal, Tanvi (23 July 2022). "How The Pandemic Boosted Business For Streaming Services" (<https://www.jumpstartmag.com/how-the-pandemic-boosted-business-for-streaming-services/>). *Jumpstart*.
199. "AUBTU.BIZ" (<https://aubtu.biz/>). Retrieved 4 November 2023.
200. "Streaming Consumption Rises In U.S. Markets with Early Stay-at-Home Orders During COVID-19" (<https://www.nielsen.com/insights/2020/streaming-consumption-rises-in-u-s-markets-with-early-stay-at-home-orders-during-covid-19/>). *Nielsen Media Research*. April 2020.
201. "The Source for Critical Information and Insight" (http://www.displaysearch.com/cps/rde/xchg/displaysearch/hs.xls/140415_lcd_tv_growth_improving_as_plasma_and_crt_tv_disappear.asp). IHS Technology.
202. "RIP, rear-projection TV" (<https://www.cnet.com/news/rip-rear-projection-tv/>). *CNET*. Retrieved 22 March 2015.
203. Taylor, Charles (2000). *The Kingfisher Science Encyclopedia* (<https://archive.org/details/kingfisherscienc00tayl>). Kingfisher. p. 370 (<https://archive.org/details/kingfisherscienc00tayl/page/370>). ISBN 978-0-7534-5269-1.
204. "How Computer Monitors Work" (<http://computer.howstuffworks.com/monitor7.htm>). 16 June 2000. Retrieved 4 October 2009.
205. "How Digital Light Processing Works" (<https://web.archive.org/web/20140221025534/https://thre3d.com/how-it-works/light-photopolymerization/digital-light-processing-dlp>). THRE3D.com. Archived from the original (<https://thre3d.com/how-it-works/light-photopolymerization/digital-light-processing-dlp>) on 21 February 2014. Retrieved 3 February 2014.
206. "Hardware Report: Shipments of LCD TVs Surpass CRT TVs" (<https://web.archive.org/web/20160304072736/http://www.dailytech.com/Report+Shipments+of+LCD+TVs+Surpass+CRT+TVs/article10775.htm>). DailyTech LLC. Archived from the original (<http://www.dailytech.com/Report+Shipments+of+LCD+TVs+Surpass+CRT+TVs/article10775.htm>) on 4 March 2016. Retrieved 20 February 2016.
207. "Digital Television" (<https://web.archive.org/web/20080308190335/http://www.fcc.gov/cgb/consumerfacts/digitaltv.html>). 28 February 2013. Archived from the original (<https://www.fcc.gov/cgb/consumerfacts/digitaltv.html>) on 8 March 2008. Retrieved 25 May 2016.
208. "What is Ultra HDTV?" (<https://web.archive.org/web/20131030020850/http://www.ultrahdtv.net/what-is-ultra-hdtv/>), *Ultra HDTV Magazine*, 5 February 2008, archived from the original (<http://www.ultrahdtv.net/what-is-ultra-hdtv/>) on 30 October 2013, retrieved 27 October 2013

209. "The Ultimate Guide to 4K Ultra HD" (<https://web.archive.org/web/20150324181932/http://www.ultrahdtv.net/the-ultimate-guide-to-4k-and-8k-ultra-hd/>), *Ultra HDTV Magazine*, 28 July 2013, archived from the original (<http://www.ultrahdtv.net/the-ultimate-guide-to-4k-and-8k-ultra-hd/>) on 24 March 2015, retrieved 27 October 2013
210. Martin, Andrew (27 December 2011). "Plummeting TV Prices Squeeze Makers and Sellers" (<https://www.nytimes.com/2011/12/27/business/tv-prices-fall-squeezing-most-makers-and-sellers.html>). *The New York Times*. p. B1. Retrieved 27 December 2011.
211. Global, K. E. D. "Samsung, LG Electronics top 2 global TV makers in H1" (<https://www.kedglobal.com/electronics/newsView/ked202308210021>). *KED Global*. Retrieved 18 October 2023.
212. "The 100+ Best Celebrity Reality TV Shows, Ranked" (<https://www.ranker.com/list/best-celebrity-reality-shows/ranker-tv>). *Ranker*.
213. Global TV 2010 – Markets, Trends Facts & Figures (2008–2013) (http://www.international-television.org/tv_market_data/world-tv-market-2010.html) International Television Expert Group
214. Global TV Revenues (2008–09) (http://www.international-television.org/tv_market_data/global-tv-revenues-2008-2009.html) International Television Expert Group
215. iDate's Global TV Revenue Market Shares (http://www.international-television.org/tv_market_data/world-tv-market-2010.html) International Television Expert Group
216. OFCOM's Global TV Market Report 2009 (http://www.international-television.org/tv_market_data/international_television_market_report_2009.html) International Television Expert Group
217. Karen Hornick (http://www.americanheritage.com/articles/magazine/ah/2006/5/2006_5_50.shtml) Archived (https://web.archive.org/web/20100917000611/http://www.americanheritage.com/articles/magazine/ah/2006/5/2006_5_50.shtml) 17 September 2010 at the Wayback Machine "That Was the Year That Was" *American Heritage*, Oct. 2006.
218. Fritz Plasser, *Global Political Campaigning*, p226
219. Stewart, R.W., "Imagery For Profit", *The New York Times*, 6 July 1941.
220. "WNBT/Bulova test pattern" (http://www.earlytelevision.org/images/rca_bulova_ad-1.jpg).
221. コマーシャルメッセージ (Commercial message). Retrieved 24 November 2013
222. "1940–1949 C.E.: Media History Project: U of M" (<https://web.archive.org/web/20121025200852/http://www.mediahistory.umn.edu/timeline/1940-1949.html>). Mediahistory.umn.edu. 18 May 2012. Archived from the original (<http://www.mediahistory.umn.edu/timeline/1940-1949.html>) on 25 October 2012. Retrieved 2 November 2012.
223. "The American Experience | People & Events | The Aftermath of the Quiz Show Scandal" (<https://www.pbs.org/wgbh/amex/quizshow/peopleevents/pande07.html>). PBS. Retrieved 2 November 2012.
224. Jon Stewart of *The Daily Show* was mock-outraged at this, saying, "That's what we do!" and calling it a new form of television, "infoganda."
225. Segraive, Kerry (1994). *Product Placement in Hollywood Films*. McFarland. ISBN 978-0-7864-1904-3.
226. "Kenneth Roy Thomson" (<https://web.archive.org/web/20110616115610/http://www.pressgazette.co.uk/story.asp?storyCode=34783§ioncode=1>). *Press Gazette*. 7 July 2006. Archived from the original (<http://www.pressgazette.co.uk/story.asp?storyCode=34783§ioncode=1>) on 16 June 2011. Retrieved 24 April 2010.
227. Blackburn, Tom (17 January 2022). "The government is planning to scrap the BBC's licence fee" (<https://www.birminghammail.co.uk/news/uk-news/bbc-licence-fee-introduced-what-22775309>). *Birmingham Live*.
228. "How the BBC makes money" (<https://fullfact.org/news/bbc-income/>). *Full Fact*. 10 May 2021.

229. General Commercial Communications Code (<http://www.bai.ie/wordpress/wp-content/uploads/General-Commercial-Communications-Code.pdf>) (Archived (<https://web.archive.org/web/20120617110756/http://www.bai.ie/wordpress/wp-content/uploads/General-Commercial-Communications-Code.pdf>) 17 June 2012 at the Wayback Machine) and Children's Commercial Communications Code (<http://www.bai.ie/wordpress/wp-content/uploads/Childrens-Commercial-Communications-Code.pdf>), referenced in: "BAI launches Revised Broadcasting Codes" (<http://www.bai.ie/index.php/2010/05/bai-launches-revised-broadcasting-codes/>). Broadcasting Authority of Ireland. May 2010. Retrieved 1 May 2016.;
230. "TV Licensing-FOI: Licences facts and figures" (<http://www.tvlicensing.co.uk/about/foi-licences-facts-and-figures-AB18/>). tvlicensing.co.uk. Retrieved 10 December 2012.
231. "viewing statistics in UK" (<https://web.archive.org/web/20081005045049/http://www.barb.co.uk/viewingsummary/weekreports.cfm?report=multichannel&requesttimeout=500&flag=viewingsummary>). Barb.co.uk. Archived from the original (<http://www.barb.co.uk/viewingsummary/weekreports.cfm?report=multichannel&requesttimeout=500&flag=viewingsummary>) on 5 October 2008. Retrieved 17 April 2009.
232. "The Communications Market: Digital Progress Report – Digital TV, Q3 2007" (https://web.archive.org/web/20080625013029/http://www.ofcom.org.uk/research/tv/reports/dtv/dtv_2007_q3/dtvq307.pdf) (PDF). Archived from the original (http://www.ofcom.org.uk/research/tv/reports/dtv/dtv_2007_q3/dtvq307.pdf) (PDF) on 25 June 2008. Retrieved 18 June 2010.
233. "TV Licence types and costs" (<https://www.tvlicensing.co.uk/check-if-you-need-one/topics/tv-licence-types-and-costs-top2>). TV Licensing. Archived (<https://web.archive.org/web/20210502005124/https://www.tvlicensing.co.uk/check-if-you-need-one/topics/tv-licence-types-and-costs-top2>) from the original on 2 May 2021. Retrieved 15 June 2021.
234. Muller, Denis (13 February 2019). "Australian governments have a long history of trying to manipulate the ABC – and it's unlikely to stop now" (<https://theconversation.com/australian-governments-have-a-long-history-of-trying-to-manipulate-the-abc-and-its-unlikely-to-stop-now-110712>). *The Conversation*. Retrieved 15 June 2021.
235. Yussuf, Ahmed (19 November 2014). "Turnbull confirms \$254 million cut from ABC funding" (<https://www.abc.net.au/news/2014-11-19/abc-funding-cuts-announced-by-malcolm-turnbull/5902774>). ABC News. Retrieved 15 June 2021.
236. Duke, Jennifer (16 August 2020). "Was the ABC's funding cut?" (<https://www.smh.com.au/national/was-the-abc-s-funding-cut-20200814-p55lvr.html>). *The Sydney Morning Herald*. Retrieved 15 June 2021.
237. Wake, Alexandra; Ward, Michael (24 June 2020). "Latest \$84 million cuts rip the heart out of the ABC, and our democracy" (<https://theconversation.com/latest-84-million-cuts-rip-the-heart-out-of-the-abc-and-our-democracy-141355>). *The Conversation*. Retrieved 15 June 2021.
238. Spigelman, James (8 December 2014). "ABC Services in the Asia-Pacific" (<https://about.abc.net.au/speeches/abc-services-in-the-asia-pacific/>). About the ABC. Retrieved 15 June 2021.
239. O'Keeffe, Annmaree; Greene, Chris (10 December 2019). "International Public Broadcasting: A Missed Opportunity For Projecting Australia's Soft Power" (<https://www.lowyinstitute.org/publications/international-public-broadcasting-missed-opportunity-projecting-australia-s-soft-power>). Lowy Institute. Retrieved 15 June 2021.
240. Ministry of Finance (http://www.minefi.gouv.fr/paca/minefi_relaiss_sociaux/impots/fiche21.html) Archived (https://web.archive.org/web/20070501212316/http://www.minefi.gouv.fr/paca/minefi_relaiss_sociaux/impots/fiche21.html) 1 May 2007 at the Wayback Machine
241. Butler, Fionnuala, Cynthia Pickett. "Imaginary Friends." *Scientific American*. 28 July 2009. Web. 26 March 2010. <http://www.scientificamerican.com/article.cfm?id=imaginary-friends>
242. "The Good Things About Television" (https://web.archive.org/web/20060203124845/http://www.media-awareness.ca/english/parents/television/good_things_tv.cfm). Archived from the original (http://www.media-awareness.ca/english/parents/television/good_things_tv.cfm) on 3 February 2006.

243. Marty, Martin E. (1980). *Where the Spirit Leads: American Denominations Today*. John Knox Press. p. 88. ISBN 978-0-8042-0868-0.
244. "Technology Handbook" (<https://www.pcci.edu/Handbook/FacilitiesServices/Technology.aspx>). Pensacola Christian College. Retrieved 2 May 2021.
245. "Television: An Occasion of Sin?" (https://www.sspxasia.com/Documents/Catholic_Morality/TV-Occasion-of-Sin.htm). Society of Saint Pius X. Retrieved 5 May 2021.
246. Lampert, Mark A. (31 August 2017). *Encyclopedia of Martin Luther and the Reformation*. Rowman & Littlefield. p. 409. ISBN 978-1-4422-7159-3.
247. David Anderson (7 July 2007). "The Kingdom of God, the Fellowship of the Saints" (<https://www.llchurch.org/resources/the-kingdom-of-god-the-fellowship-of-the-saints>). Laestadian Lutheran Church. Retrieved 16 September 2021.
248. *Dunkard Brethren Church Polity*. Dunkard Brethren Church. 1 November 2021. p. 8.
249. Ferguson, RW; Mickalide, AD (December 2012). *A Report to the Nation on Home Safety: The Dangers of TV Tip-Overs* (<https://www.safekids.org/sites/default/files/documents/ResearchReports/TV%20Tipovers%20Research%20Rpt.pdf>) (PDF). Washington, D.C.: Safe Kids Worldwide. Archived (<https://web.archive.org/web/20150906191453/https://www.safekids.org/sites/default/files/documents/ResearchReports/TV%20Tipovers%20Research%20Rpt.pdf>) (PDF) from the original on 6 September 2015.
250. Bernard, Philip A.; Johnston, Carden; Curtis, Scott E.; King, William D. (1 September 1998). "Toppled television sets cause significant pediatric morbidity and mortality" (<https://doi.org/10.1542/peds.102.3.e32>). *Pediatrics*. **102** (3): E32. doi:10.1542/peds.102.3.e32 (<https://doi.org/10.1542/peds.102.3.e32>). eISSN 1098-4275 (<https://search.worldcat.org/issn/1098-4275>). ISSN 0031-4005 (<https://search.worldcat.org/issn/0031-4005>). OCLC 1761995 (<https://search.worldcat.org/oclc/1761995>). PMID 9724680 (<https://pubmed.ncbi.nlm.nih.gov/9724680>).
251. *Home Safety Fact Sheet (2015)* (https://web.archive.org/web/20150824000908/https://www.safekids.org/sites/default/files/documents/skw_home_fact_sheet_feb_2015.pdf) (PDF). SafeKids Worldwide. February 2015. Archived from the original (https://www.safekids.org/sites/default/files/documents/skw_home_fact_sheet_feb_2015.pdf) (PDF) on 24 August 2015.
252. Hernæs, Øystein; Markussen, Simen; Røed, Knut (1 March 2019). "Television, Cognitive Ability, and High School Completion" (<https://econpapers.repec.org/RePEc:iza:izadps:dp9645>). *The Journal of Human Resources*. **54** (2): 371–400. doi:10.3368/jhr.54.2.0316.7819R1 (<https://doi.org/10.3368/jhr.54.2.0316.7819R1>). eISSN 1548-8004 (<https://search.worldcat.org/issn/1548-8004>). hdl:10419/130339 (<https://hdl.handle.net/10419/130339>). ISSN 0022-166X (<https://search.worldcat.org/issn/0022-166X>). LCCN 66009974 (<https://lccn.loc.gov/66009974>). OCLC 1604126 (<https://search.worldcat.org/oclc/1604126>). S2CID 4846339 (<https://api.semanticscholar.org/CorpusID:4846339>).
253. "The Rise of the Machines: A Review of Energy Using Products in the Home from the 1970s to Today" (<https://web.archive.org/web/20120828072528/http://www.energysavingtrust.org.uk/content/download/28169/337639/version/3/file/Riseofthemachines.pdf>) (PDF). Energy Saving Trust. 3 July 2006. Archived from the original (<http://www.energysavingtrust.org.uk/content/download/28169/337639/version/3/file/Riseofthemachines.pdf>) (PDF) on 28 August 2012. Retrieved 31 August 2007.

Further reading

- Abramson, Albert (2003). *The History of Television, 1942 to 2000*. Jefferson, NC / London: McFarland. ISBN 978-0-7864-1220-4.
- Pierre Bourdieu, *On Television*, The New Press, 2001.
- Tim Brooks and Earle March, *The Complete Guide to Prime Time Network and Cable TV Shows*, 8th ed., Ballantine, 2002.

- Jacques Derrida and Bernard Stiegler, *Echographies of Television*, Polity Press, 2002.
- David E. Fisher and Marshall J. Fisher, *Tube: the Invention of Television*, Counterpoint, Washington, D.C., 1996, ISBN 1-887178-17-1.
- Steven Johnson, *Everything Bad is Good for You: How Today's Popular Culture Is Actually Making Us Smarter*, New York, Riverhead (Penguin), 2005, 2006, ISBN 1-59448-194-6.
- Leggett, Julian (April 1941). "Television in Color" (<https://books.google.com/books?id=CdoDAAAAMBAJ&pg=PA497>). *Popular Mechanics*. Chicago. Retrieved 7 December 2014.
- Jerry Mander, *Four Arguments for the Elimination of Television*, Perennial, 1978.
- Jerry Mander, *In the Absence of the Sacred*, Sierra Club Books, 1992, ISBN 0-87156-509-9.
- Neil Postman, *Amusing Ourselves to Death: Public Discourse in the Age of Show Business*, New York, Penguin US, 1985, ISBN 0-670-80454-1.
- Evan I. Schwartz, *The Last Lone Inventor: A Tale of Genius, Deceit, and the Birth of Television*, New York, Harper Paperbacks, 2003, ISBN 0-06-093559-6.
- Beretta E. Smith-Shomade, *Shaded Lives: African-American Women and Television*, Rutgers University Press, 2002.
- Alan Taylor, *We, the Media: Pedagogic Intrusions into US Mainstream Film and Television News Broadcasting Rhetoric*, Peter Lang, 2005, ISBN 3-631-51852-8.
- Amanda D. Lotz, *The Television Will Be Revolutionized*, New York University Press, ISBN 978-0-8147-5220-3

External links

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Television&oldid=1290554091>"



Electronic literature

Electronic literature or **digital literature** is a genre of literature where digital capabilities such as interactivity, multimodality or algorithmic text generation are used aesthetically.^[1] Works of electronic literature are usually intended to be read on digital devices, such as computers, tablets, and mobile phones. They cannot be easily printed, or cannot be printed at all, because elements crucial to the work cannot be carried over onto a printed version.

The first literary works for computers, created in the 1950s, were computer programs that generated poems or stories, now called generative literature. In the 1960s experimental poets began to explore the new digital medium, and the first early text-based games were created. Interactive fiction became a popular genre in the late 1970s and 1980s, with a thriving online community in the 2000s. In the 1980s and 1990s hypertext fiction begun to be published, first on floppy disks and later on the web. Hypertext fictions are stories where the reader moves from page to page by selecting links. In the 2000s digital poetry became popular, often including animated text, images and interactivity. In the 2010s and 2020s, electronic literature uses social media platforms, with new genres like Instapoetry or Twitterature as well as literary practices like netprov. Although web-based genres like creepypasta and fan fiction are not always thought of as electronic literature (because they usually manifest as linear texts that could be printed out and read on paper) other scholars argue that these are born digital genres that depend on online communities and thus should be included in the field.

There is an extensive body of scholarship on electronic literature. In 1999 the Electronic Literature Organization was established, which through annual conferences and other events supports both the publishing and study of electronic literature. One focus of academic study has been the preservation and archiving of works of electronic literature. This is challenging because works become impossible to access or read when the software or hardware they are designed for becomes obsolete. In addition, works of electronic literature are not part of the established publishing industry and so do not have ISBN numbers and are not findable in library catalogues. This has led to the establishment of a number of archives and documentation projects.

Definitions

The literary critic and professor N. Katherine Hayles defines electronic literature as "'digital born' (..) and (usually) meant to be read on a computer",^[2] clarifying that this does not include e-books and digitised print literature.

Electronic literature

Features Literary works that require the capabilities of computers and networks

Related genres

Hypertext fiction, interactive fiction, digital poetry, generative literature, cell phone novels, instapoetry, cybertext, netprov, creepypasta, fan fiction, web fiction

A definition offered by the [Electronic Literature Organization](#) (ELO) states that electronic literature "refers to works with an important literary aspect that takes advantage of the capabilities and contexts provided by the stand-alone or networked computer".^[3] This can include [hypertext fiction](#), animated poetry (often called kinetic poetry) and other forms of [digital poetry](#), literary chatbots, [computer-generated narratives](#) or poetry, art installations with significant literary aspects, [interactive fiction](#) and literary uses of social media.^[3]

For example, a [hypertext fiction](#) is a story where the reader chooses a path through the story by clicking on links that connect fragments of text, often called [lexias](#).^[4] In [digital poetry](#) the words in the poem may move across the screen or may involve game-like interactivity.^[5] In [generative literature](#) a single work can generate many different poems or stories. Until the early 2000s electronic literature works tended to be published on [floppy disk](#), [CD-ROM](#), in online literary journals or on dedicated websites. However, since around 2010 literary genres on [social media platforms](#) - such as [Instapoetry](#), [Twitterature](#) or [netprov](#) - have come to be seen as electronic literature. The literary critic Leonardo Flores called these third generation electronic literature, following the first generation of pre-web works and the second generation of web-based works.^[6] Flores uses an inclusive definition of electronic literature, which can include social media posts with literary qualities even if the authors do not themselves think of it as literature.^[6] Fan fiction and creepypasta have also been analysed as electronic literature.^{[7]:109}

The definition of electronic literature is controversial within the field, with strict definitions being criticised for excluding valuable works, and looser definitions being so murky as to be useless.^[7] A work of electronic literature can be defined as "a construction whose literary aesthetics emerge from computation", "work that could only exist in the space for which it was developed/written/coded—the digital space".^[8] In his book *Electronic Literature*, the author and scholar [Scott Rettberg](#) argues that an advantage of a wide definition is its flexibility, which allows it to include new genres as new platforms and modes of literature emerge.^[7] Screenwriter and author Carolyn Handler Miller characterizes works of electronic literature as nonlinear and non chronological where the user experiences and co-creates the story, and where contradictory events and different outcomes are possible.^[9]

History

Precursors

Scholars have discussed a range of pre-digital precursors to electronic literature, from the ancient Chinese book the *I Ching*,^{[10]:9} to inventor [John Clark](#)'s mechanical [Latin Verse Machine](#) (1830-1843)^[11] to the [Dadaist](#) movement's [cut-up technique](#).^[7] Print novels that were designed to be read non-linearly, such as [Julio Cortázar](#)'s [Hopscotch](#) (1963) and [Vladimir Nabokov](#)'s [Pale Fire](#) (1962), are cited as "print antecedents" of electronic literature.^[12]

1950s

The 1952 [love letter generator](#) that the British computer scientist [Christopher Strachey](#) wrote for the [Manchester Mark 1](#) computer is probably the first example of literature that requires a computer to be generated or read.^{[13][14][15]} The work generates short love letters, and is an example of combinatorial

poetry, also called generative poetry.^[16] The original code has been lost, but digital poet and scholar Nick Montfort has reimplemented it based on remaining documentation of its output, and this version can be viewed in a web browser.^[17]

In 1959 the German computer scientist Theo Lutz wrote *Stochastic Texts*, which "for many years was considered the first digital literary text."^{[18]:11} *Stochastic Texts* was a program written for a Z22 computer that "produced random short sentences based on a corpus of chapter titles and subjects from Franz Kafka's novel *The Castle*.^[7] Lutz's work has been discussed both as a very early work of electronic literature^{[19][20][21]} and as an important precursor to current AI-generated literature.^[22] The German philosopher and media scholar Hannes Bajohr writes that *Stochastic Texts* is an example of the "sequential paradigm" in generative literature, in opposition to newer examples of a "connectionist paradigm": "Instead of hoping to recreate intuition, genius, or expression, the logic of the machine itself – that is, the logic of deterministically executed rule steps – becomes aesthetically normative in *Stochastische Texte*."^[23]

1960s

The 1960s were a time of literary experimentation, and there were strong connections between the art and technology scenes and concrete poetry.^[24] The Italian poet and artist Nanni Balestrini's poem *Tape Mark I* was composed in 1961 on an IBM 7070, and output from the poetry generator was published in a special issue of a journal edited by the novelist and scholar Umberto Eco and artist Bruno Munari, thus standing as the first Italian work of electronic literature.^[25] *Auto-Beatnik* (1961) was a program by R. M. Worthy and colleagues at the computer manufacturing company Librascope.^[26] *Auto-Beatnik* generated poems on an LGP-30 computer to mimic the style of Beat poetry.^[7]

Games designers Mabel Addis and William McKay's text-based narrative game *The Sumerian Game* (1964–66) was probably the first narrative computer game, although it was not widely distributed.^[27] The computer scientist Joseph Weizenbaum programmed the chatbot ELIZA in 1966, establishing a new genre of conversational literary artefacts or bots.^[28] This was the decade when the sociologist and philosopher Ted Nelson coined the terms hypertext and hypermedia.^[29]

1970s

Writers and artists continued to experiment with combining art, technology and literature. An example is the installation *Blikk* (1970) by a Norwegian trio: artist Irma Salo Jæger, composer Sigurd Berge and poet Jan Erik Vold.^[30] Vold's readings of his poems were mixed as sound montages by Berge and combined with Jæger's kinetic sculptures in an exhibition at the Henie Onstad Art Center. The work was recreated in 2022 by the composer and curator Jørn Rudi and is now part of the permanent collection of the Norwegian National Museum.^[31]

Another important development in the 1970s was the popularity of text adventure games, now more commonly known as interactive fiction. In 1975–76, Will Crowther programmed a text game named *Colossal Cave Adventure* (also known as *Adventure* or *ADVENT*). It possessed a story that had the reader make choices on which way to go. These choices could lead the reader to the end, or to their untimely death. It is often regarded as the first work of interactive fiction,^[7] although others have argued that the simulated microworld SHRDLU^{[32][33]} or Mabel Addis's *The Sumerian Game*^[27] were earlier and should be considered interactive fiction. Historians agree that *Colossal Cave Adventure* made a "significant

cultural impact" in the 1970s.^[34] It has been called a "classic"^[35] "so foundational it started a genre",^[36] "the *Gilgamesh* of video games",^[37] and is credited with having "informed and inspired generations of players."^[37] *Colossal Cave Adventure* was played on mainframe computers, and spread rapidly through the ARPANET. *Colossal Cave* inspired many other games, including the text adventure game *Zork* (1977) which was regarded as one of the best known.^[38]

1980s

With the advent of personal computers, interactive fiction became a commercially successful genre, driven by companies like Infocom. Companies hired authors and programmers to write text adventure games, as Veronika Megler, who wrote *The Hobbit* video game in 1982, described in an interview with *The Guardian*.^[39]

For hypertext fiction and digital poetry, the eighties were a time of experimentation in separate communities that were not necessarily aware of each other. In Canada, the poet Bp Nichol published *First Screening: Computer Poems*, written in BASIC, in 1984.^[40] The Californian writer Judy Malloy published *Uncle Roger* on the online community The WELL in 1986/87.^[41] On the East Coast, hypertext fiction was being explored by academics and writers who met at the ACM Hypertext conference, which held its inaugural meeting in 1987.^[42] Michael Joyce's *afternoon, a story*, one of the most cited works of hypertext fiction, was demonstrated at the 1987 conference, and Mark Bernstein published this work at Eastgate Systems.^[41] The hypertext author Stuart Moulthrop described discovering writer and visual artist Judy Malloy's work at this time, not having realised that there were other people writing literature for computers: "I can remember coming away from that moment thinking that, you know, there might be a real hope for what we were trying to do because other people were doing it".^[41]

In France at this time, *literature numérique* (digital literature) was connected to the Oulipo literary movement, and poetry was more central to the French writers than the narrative genres like hypertext fiction that were popular in the United States.^[43] Generative poetry could be seen as a particularly European genre at the time.^[43] In 1981 the Portuguese author Pedro Barbosa published the combinatory work *THE ALAMO*, and explicitly made the claim that computationally generated works could be literary.^[44]

Not only writers, but also digital artists created works with strong literary components that have had an influence on the field of electronic literature. An example is the Australian artist Jeffrey Shaw and Dirk Groeneveld's *The Legible City*, which was first exhibited at ZKM Center for Art and Media Karlsruhe in 1988.^[45] *The Legible City* is an art installation where the visitor rides a stationary bicycle through a simulated city displayed as computer-generated text. Buildings and streets are shown as 3D shapes consisting of letters and words, allowing the reader to "read" the city as they pedal through it.^[46]

1990s

The "Storyspace school" characterised the early 1990s in the United States,^[41] consisting of works created using Storyspace, hypertext authoring software developed by the literary scholar Jay David Bolter and the author Michael Joyce in the 1980s.^[47] Bolter and Joyce sold the Storyspace software in 1990 to Eastgate Systems, a small software company that became a publishing house and the main distributor of hypertext fiction in the 1990s, particularly in the early 1990s before it was possible to publish works on the web.^{[48]:17–18} Eastgate has maintained and updated the code in Storyspace up to the present.^[48]

Storyspace and similar programs use hypertext to create links within text. Literature using hypertext is frequently referred to as hypertext fiction. Originally, these stories were often disseminated on discs and later on CD-ROM.^[49] Hypertext fiction is still being created today using not only Storyspace, but other programs such as Twine.^{[50][41]}

This period is often termed the first generation hypertext era, as N. Katherine Hayles notes that these works used lexia or separate screens in a similar manner to books and pages.^[51]

In a 1993 article for the New York Times Book Review, "Hyperfiction: Novels for the Computer", the novelist and professor Robert Coover noted the new possibilities for exploring these various storyworlds: "[I]t is a strange place, hyperspace, much more like inner space than outer, a space not of coordinates but of the volumeless imagination".^[52] Key works from this period include Stuart Moulthrop's Victory Garden, Shelley Jackson's Patchwork Girl (1995) and Deena Larsen's work.^[53]

Towards the middle of the decade, authors began writing on the web. Stuart Moulthrop's Hegirascope was published in 1995. Early web-based hypertext fictions include Olia Lialina's My Boyfriend Came Back from the War, Adrienne Eisen's Six Sex Scenes and Robert Arellano's Sunshine '69, all published in 1996.^{[54][7]} Scott Rettberg, William Gillespie, Dirk Stratton, and Frank Marquadt's sprawling hypertext novel The Unknown won the trAce/Alt-X Hypertext Competition in 1998.^[55] It was featured in the Electronic Literature Collection Vol. 2,^[56] and has been analysed by a number of scholars.^{[57][58][59][60]}

The Electronic Literature Organization (the ELO) was founded in 1999 by hypertext author Scott Rettberg, the author and teacher of creative writing Robert Coover and internet investor Jeff Ballowe, with the mission "to facilitate and promote the writing, publishing, and reading of literature in electronic media".^[61] The ELO is still active today, with annual conferences, online discussions and publications.^{[62][63]}

2000s

In Japan, cell phone novels became popular from the early 2000s.^[65] Similar genres emerged in other countries where text messaging was well-established, including India^[66] and Europe.^[67] The first work of Indian electronic literature is probably the 2004 SMS novel Cloak Room,^[68] whose author used the pseudonym RoGue. Cloak Room invited readers to engage with the story by answering texts or leaving comments on the blog that was used in tandem with the text messages.^{[66][69]}

In North America the web was becoming the main platform for electronic literature. The Canadian author Caitlin Fisher's These Waves of Girls (2001) was a hypermedia novella telling stories of girlhood, using images and sounds as well as links and text.^[70] The American author Talan Memmott's Lexia to Perplexia (2000) offered complex visual and textual layers that sometimes confuse and occlude themselves,^[71] and is described by the literary critic Lisa Swanstrom as a "beautifully intricate piece of electronic literature".^[72] Kate Pullinger's Inanimate Alice is an example of a work that began as a web



Stuart Moulthrop's hypertext fiction Victory Garden (1992) is shown here in two versions: the original, published on floppy disks to be read on a computer, and the 2009 version that was reprogrammed to work on an iPad. This display was part of an exhibition curated by Dene Grigar for the ACM Hypertext conference in 2023.

novel and then saw versions across several media, including a screenplay and a VR experience.^[73] Works like *The Impermanence Agent*, by author and scholar Noah Wardrip-Fruin and collaborators, explored the web's ability to customise a story for the reader.^[74]

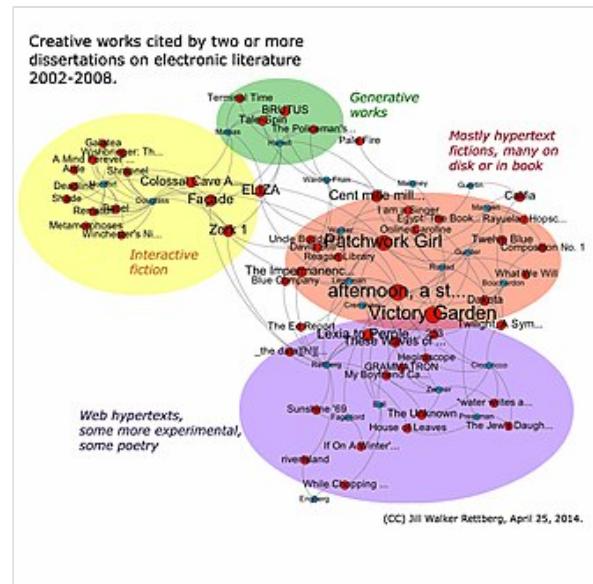
An analysis of 44 PhD dissertations about electronic literature published between 2002 and 2013^[41] found a clear shift in the genres referenced by the authors of the dissertations during this period. Between 2002 and 2008, the referenced works clustered in four distinct genre groups: interactive fiction, generative literature, classic hypertext fiction (mostly published on disk or in print) and web hypertexts, including more experimental works and some poetry.^[41]

Blog fiction and fan fiction are born-digital literary genres that became popular in this period.^{[75][76][77]} Blog fictions have been a particularly popular genre of electronic literature in Africa.^{[78][79][80]} The literary orality of blogs has also been analysed as a feature of African American blogs.^[81]

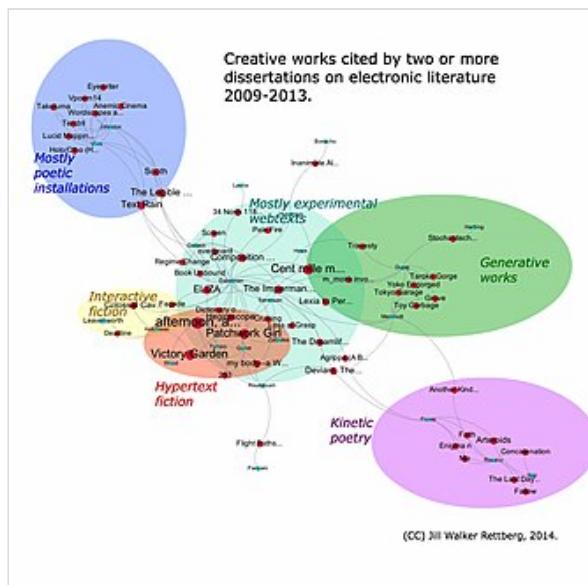
2010s

The spread of smartphones and tablets led to literary works that explored the touchscreen, such as Samantha Gorman and Danny Cannizarro's Pry (2014)^[82] or Kate Pullinger's Breathe: A Ghost Story.^[13] Netprov, improvisational and collaborative networked writing was another genre that developed during the 2000s and 2010s, with projects like #1WkNoTech.^{[83][84]} Instapoetry, a visual style of poetry native to Instagram became a popular success.^[85]

The web-based hypertext authoring tool Twine became increasingly popular this decade. This "Twine revolution"^[86] led to a resurgence of interactive fiction and hypertext,^[87] which now became "a mainstream form of literary game production and interaction".^[88] Notable works written in Twine that are frequently discussed as electronic literature include Anna Anthropy's *Queers in Love at the End of the World* (2013)^{[89][90]} and Dan Hett's autobiographical *C ya laterrrr* about losing his brother in the Manchester Arena bombing (2017).^[91]



A network visualisation showing works of electronic literature cited by two or more PhD dissertations on electronic literature defended between 2002 and 2008. Four clear genres emerge: interactive fiction, generative works, hypertext fictions and more experimental web hypertexts and poetry.^[64]



PhD dissertations on electronic literature completed between 2009 and 2013 show a shift in genres. Classic hypertext fiction is still present (the red circle), as are the experimental webtexts, interactive fiction and generative works. Two new distinct genres have emerged as important to this generation of dissertation writers: kinetic poetry and digital poetry installation art.^[64]

As machine learning made rapid advances with natural language processing and deep learning, authors began to experiment and write with AI.^{[92][93]} David Jhave Johnston's *ReRites* is an example of this new kind of generative literature and is a poetic work written as a human-AI collaboration. A GPT-2 language model was trained on a corpus of contemporary poetry and set to generate new poems every night. Each morning, Jhave Johnston would rewrite the poems as a daily ritual: hence the title *ReRites*.^[94]

Dissertations published between 2009 and 2013 still cite many works in the genres of hypertext fiction, interactive fiction, experimental webtexts and generative texts. Digital poetry also emerged as a significant genre, with dissertation authors writing about two distinct clusters of digital poetry: kinetic poetry and poetic installations in art galleries. Many of these works were from the 1980s to the early 2000s, so this may indicate an uptake in scholarly interest rather than a large change in what kinds of creative works were actually published in the 2010s.^[41]

2020s

Electronic literature spread internationally. *The Electronic Literature Collection Volume 4*, published in 2022, showcases 132 works from 42 different countries in 31 languages.^[95] The first volume of the *Indian Electronic Literature Anthology*, published in 2024, showcases 17 works of electronic literature written in Hindi and English.^[96]

Scholarship

Histories and timelines

Various histories of electronic literature and its subgenera have been written. Scott Rettberg's *Electronic Literature*^[7] provides a broad overview, while more specialised books discuss the history of specific genres or periods, like Chris Funkhouser's *Prehistoric Digital Poetry*^[19] and Astrid Ensslin's *Pre-web Digital Publishing and the Lore of Electronic Literature*.^[97]

As mentioned above in the section on Definitions, the literary critic Leonardo Flores proposes a generational understanding of electronic literature, where the first generation is pre-web, the second uses the web, and the third generation uses social media, web APIs and mobile devices.^[6] However, not all works fit within this structure, as Spencer Jordan notes, writing that "A work such as *The Unknown*, for example, sits uneasily between second and third generation definitions."^[98]

Reader interaction and nonlinearity

Digital literature tends to require a user to traverse through the literature through the digital setting, making the use of the medium part of the literary exchange. Espen J. Aarseth wrote in his book *Cybertext: Perspectives on Ergodic Literature* that "it is possible to explore, get lost, and discover secret paths in these texts, not metaphorically, but through the topological structures of the textual machinery".^{[10]:4} Espen Aarseth defines "ergodic literature" as literature where "nontrivial effort is required to allow the reader to traverse the text".^[99] George Landow explains that following hypertext links merges the traditional expectations of reader and writer roles as the reader constructs the text by following links.^[100]

Astrid Ensslin and Alice Bell note that electronic literature works can embody central contradictions in ways that differ from print literature. They cite examples such as *afternoon, a story* (a car accident that may not or may occur), *Victory Garden* (a character both dies and lives), and *Patchwork Girl* (a character is real or imagined).^{[101]:30} Plot lines, emotional intensity, character traits and attributions can vary depending on a reader's chosen path.^[102] J Yellowlees Douglas shows an early example of this in Michael Joyce's 1991 hypertext fiction *WOE* where romances would occur between different characters, depending on a reader's path.^[102] Encountering a node (or lexia) in different contexts can convey impressions of larger databases as information seems to differ depending on the context that the user is coming from, as J Yellowlees Douglas explains about *The Election of 1912*, by Mark Bernstein and Erin Sweeney.^[103]

Preservation and archiving

Because electronic literature is made to be read on computers, works often become unreadable when the software platforms or technologies they are designed for become obsolete. This may have made it more difficult for electronic literature to build the "traditions associated with print literature", as literary critic N. Katherine Hayles has argued.^{[104]:39–40}

Several organizations are dedicated to preserving works of electronic literature. The UK-based Digital Preservation Coalition aims to preserve digital resources in general, while the Electronic Literature Organization's PAD (Preservation / Archiving / Dissemination) initiative gave recommendations on how to think ahead when writing and publishing electronic literature, as well as how to migrate works running on defunct platforms to current technologies.^{[105][106]} The British Library archives winners of the New Media Writing Prize in the UK Web Archive.^{[107][108]} The NEXT, run by Dene Grigar for the Electronic Literature Organization, hosts source files and documentation of many works of electronic literature and digital writing.^[109] The Electronic Literature Knowledge Base (ELMCIP)^[110] is a research resource for electronic literature, with 3,875 records of creative works as of February 11, 2024. The Electronic Literature Directory^[111] focuses on peer-reviewed descriptions or reviews of works. The Multilingual African Electronic Literature Database & African Diasporic Electronic Literature Database (MAELD & ADEL)^[112] is a project focusing on the African region. The Maryland Institute for Technologies in the Humanities and the Electronic Literature Lab^[113] at Washington State University Vancouver work towards the documentation and preservation of electronic literature and hypermedia. In Canada, the Laboratory NT2 hosts research and a database on electronic literature and digital art.^[114]

The *Electronic Literature Collection* is a series of anthologies of electronic literature published by the Electronic Literature Organization, both on CD/DVD and online, and this is another strategy in working to make sure that electronic literature is available for future generations.^{[115][116]}

Major awards

Annual awards for electronic literature include the Electronic Literature Organization awards^[117] and the New Media Writing Prize.^[118]

Previous awards included the trAce-Alt X Competition. In 1998, two works shared the 1,000 English pound prize: *The Unknown* by William Gillespie; Scott Rettberg; Dirk Stratton and *Rice* by Jenny Weight (Australia). Three sites received Honorable Mentions: *Kokura* by Mary-Kim Arnold, **** by Michael Atavar, and *water always writes in plural* by Linda Carrol and Josephine Wilson.^{[119][120]} In 2001, *Lexia to Perplexia* by Talan Memmott won the trAce/Alt-X New Media Writing Award.^{[121][122][123]} In 2004, the prize had four major categories for articles about hypertext (reviews, opinion, and editor's choice. The only multimedia work mentioned was *Postcards From Writing* by Sally Prior.^[124]

See also

- [List of electronic literature authors, critics, and works](#)
- [List of women electronic writers - Wikipedia](#)

References

1. "Electronic Literature | OELN" (<https://www.oeln.net/electronic-literature>). www.oeln.net. Retrieved 2024-03-24.
2. Hayles, N. Katherine (2008). *Electronic Literature: New Horizons for the Literary*. South Bend, IN: University of Notre Dame Press. p. 3.
3. "About the ELO" (<https://eliterature.org/about/>). Electronic Literature Organization. Archived (<https://web.archive.org/web/20230801100751/https://eliterature.org/about/>) from the original on 2023-08-01. Retrieved 2023-08-01.
4. Bell, Alice; Ensslin, Astrid (2011). ""I know what it was. You know what it was": Second-Person Narration in Hypertext Fiction" (<https://www.jstor.org/stable/41289307>). *Narrative*. **19** (3): 311. ISSN 1063-3685 (<https://search.worldcat.org/issn/1063-3685>). JSTOR 41289307 (<https://www.jstor.org/stable/41289307>). "Hypertext fiction is a specific form of digital fiction in which fragments of electronic text, known as lexias, are connected by hyperlinks."
5. Naji, Jeneen (2021). *Digital poetry*. Palgrave pivot. Cham, Switzerland: Palgrave Macmillan. ISBN 978-3-030-65962-2.
6. Flores, Leonardo (2019-04-07). "Third Generation Electronic Literature" (<https://doi.org/10.7273%2Faxyj-3574>). *Electronic Book Review*. doi:10.7273/axyj-3574 (<https://doi.org/10.7273%2Faxyj-3574>).
7. Rettberg, Scott (2019). *Electronic literature*. Cambridge, UK: Polity Press. ISBN 978-1-5095-1677-3. OCLC 1028213515 (<https://search.worldcat.org/oclc/1028213515>).
8. Heckman, Davin; O'Sullivan, James (2018). "Electronic Literature: Contexts and Poetics" (<https://dlsanthology.mla.hcommons.org/electronic-literature-contexts-and-poetics/>). *Literary Studies in the Digital Age: An Evolving Anthology*. Archived (<https://web.archive.org/web/20180413043934/https://dlsanthology.mla.hcommons.org/electronic-literature-contexts-and-poetics/>) from the original on 2018-04-13. Retrieved 2018-04-12.
9. Miller, Carolyn Handler (2014). *Digital storytelling: a creator's guide to interactive entertainment* (3rd ed.). Burlington, MA: Focal Press. p. 19. ISBN 978-0-415-83694-4.
10. Aarseth, Espen J. (1997). *Cybertext: perspectives on ergodic literature*. Baltimore, MD; London: Johns Hopkins University Press. ISBN 978-0-8018-5578-8.

11. Sharples, Mike (2023-01-01). "John Clark's Latin Verse Machine: 19th Century Computational Creativity" (<https://ieeexplore.ieee.org/document/10038656>). *IEEE Annals of the History of Computing*. **45** (1): 31–42. arXiv:2301.05570 (<https://arxiv.org/abs/2301.05570>). doi:10.1109/MAHC.2023.3241258 (<https://doi.org/10.1109%2FMAHC.2023.3241258>). ISSN 1058-6180 (<https://search.worldcat.org/issn/1058-6180>). S2CID 255825542 (<https://api.semanticscholar.org/CorpusID:255825542>). Archived (<https://web.archive.org/web/20230730174712/https://ieeexplore.ieee.org/document/10038656/>) from the original on 2023-07-30. Retrieved 2023-07-31.
12. Pressman, Jessica (2014). *Digital modernism: making it new in new media*. Modernist literature & culture. New York: Oxford university press. ISBN 978-0-19-993710-3.
13. Rettberg, Jill Walker (2021-10-03). "Speculative Interfaces: How Electronic Literature Uses the Interface to Make Us Think about Technology" (<https://doi.org/10.7273%2F1xsg-nv26>). *Electronic Book Review*. doi:[10.7273%2F1xsg-nv26](https://doi.org/10.7273%2F1xsg-nv26) (<https://doi.org/10.7273%2F1xsg-nv26>).
14. Gaboury, Jacob (2013-12-01). "Darling sweetheart: Queer objects in early computer art" (<https://www.ingentaconnect.com/content/intellect/mecr/2013/00000003/f0020001/art00003;jsessionid=1rcdleoklq265.x-ic-live-03>). *Metaverse Creativity*. **3** (1–2): 23–27. doi:[10.1386/mvcr.3.1-2.23_1](https://doi.org/10.1386/mvcr.3.1-2.23_1) (https://doi.org/10.1386%2Fmvcr.3.1-2.23_1).
15. Wardrip-Fruin, Noah (2011-06-16). "14. Digital Media Archaeology: Interpreting Computational Processes" (<https://www.degruyter.com/document/doi/10.1525/9780520948518-016/html>). *Media Archaeology*. University of California Press. pp. 302–322. doi:[10.1525/9780520948518-016](https://doi.org/10.1525/9780520948518-016) (<https://doi.org/10.1525%2F9780520948518-016>). ISBN 978-0-520-94851-8. S2CID 226776992 (<https://api.semanticscholar.org/CorpusID:226776992>). Archived (<https://web.archive.org/web/20221012123123/https://www.degruyter.com/document/doi/10.1525/9780520948518-016/html>) from the original on 2022-10-12. Retrieved 2022-10-12.
16. Todorovic, Vladimir; Grba, Dejan (2019-05-01). "Wandering machines: narrativity in generative art" (<https://revistas.ucp.pt/index.php/jsta/article/view/7334>). *Journal of Science and Technology of the Arts*. **11** (2): 50–58. doi:[10.7559/citarj.v11i2.664](https://doi.org/10.7559/citarj.v11i2.664) (<https://doi.org/10.7559%2Fcitarj.v11i2.664>). hdl:[10356/138446](https://hdl.handle.net/10356/138446) (<https://hdl.handle.net/10356%2F138446>). ISSN 2183-0088 (<https://search.worldcat.org/issn/2183-0088>). S2CID 243018095 (<https://api.semanticscholar.org/CorpusID:243018095>). Archived (<https://web.archive.org/web/20230731173547/https://revistas.ucp.pt/index.php/jsta/article/view/7334>) from the original on 2023-07-31. Retrieved 2023-07-31.
17. Strachey, Christopher (2014). Montfort, Nick (ed.). "Love Letters" (https://nickm.com/memslam/love_letters.html). *nickm.com* (Reimplementation of 1953 love letter generator.). Archived (https://web.archive.org/web/20230610085512/https://nickm.com/memslam/love_letters.html) from the original on 2023-06-10. Retrieved 2023-08-01.
18. di Rosario, Giovanna; Meza, Nohelia; Grimaldi, Kerri (2021). "The Origins of Electronic Literature: An Overview". In Grigar, Dene; O'Sullivan, James (eds.). *Electronic literature as digital humanities: contexts, forms, & practices*. New York; London; Oxford; New Delhi; Sydney: Bloomsbury Academic. pp. 9–26. ISBN 978-1-5013-6349-8.
19. Funkhouser, Chris (2007). *Prehistoric digital poetry : an archaeology of forms, 1959-1995*. Tuscaloosa: University of Alabama Press. ISBN 978-0-8173-8087-8. OCLC 183291342 (<https://search.worldcat.org/oclc/183291342>).
20. Beals, Kurt (2018). " "Do the New Poets Think? It's Possible": Computer Poetry and Cyborg Subjectivity". *Configurations*. **26** (2): 149–177. doi:[10.1353/con.2018.0010](https://doi.org/10.1353/con.2018.0010) (<https://doi.org/10.1353%2Fcon.2018.0010>). ISSN 1080-6520 (<https://search.worldcat.org/issn/1080-6520>). S2CID 150182225 (<https://api.semanticscholar.org/CorpusID:150182225>).
21. Funkhouser, Christopher T. (March 2017). "IBM Poetry: Exploring Restriction in Computer Poems" (<https://doi.org/10.3390%2Fh6010007>). *Humanities*. **6** (1): 7. doi:[10.3390/h6010007](https://doi.org/10.3390/h6010007) (<https://doi.org/10.3390%2Fh6010007>). ISSN 2076-0787 (<https://search.worldcat.org/issn/2076-0787>).

22. Husárová, Zuzana; Piorecký, Karel (2022). "Reception of literature generated by artificial neural networks" (<https://doi.org/10.31577%2Fwls.2022.14.1.4>). *World Literature Studies* (in Slovak). **14** (1): 44–60. doi:[10.31577/wls.2022.14.1.4](https://doi.org/10.31577/wls.2022.14.1.4) (<https://doi.org/10.31577%2Fwls.2022.14.1.4>). S2CID [248305840](https://api.semanticscholar.org/CorpusID:248305840) (<https://api.semanticscholar.org/CorpusID:248305840>).
23. Bajohr, Hannes (March 2022). "Algorithmic Empathy: Toward a Critique of Aesthetic AI". *Configurations*. **30** (2): 210. doi:[10.1353/con.2022.0011](https://doi.org/10.1353/con.2022.0011) (<https://doi.org/10.1353%2Fcon.2022.0011>). ISSN [1080-6520](https://search.worldcat.org/issn/1080-6520) (<https://search.worldcat.org/issn/1080-6520>). S2CID [248578007](https://api.semanticscholar.org/CorpusID:248578007) (<https://api.semanticscholar.org/CorpusID:248578007>).
24. D'Ambrosio, Matteo (2018-08-10). "The Early Computer Poetry and Concrete Poetry" (https://impactum-journals.uc.pt/matlit/article/view/2182-8830_6-1_4). *Matlit Revista do Programa de Doutoramento em Materialidades da Literatura*. **6** (1): 51–72. doi:[10.14195/2182-8830_6-1_4](https://doi.org/10.14195/2182-8830_6-1_4) (https://doi.org/10.14195%2F2182-8830_6-1_4). ISSN [2182-8830](https://search.worldcat.org/issn/2182-8830) (<https://search.worldcat.org/issn/2182-8830>). S2CID [194914879](https://api.semanticscholar.org/CorpusID:194914879) (<https://api.semanticscholar.org/CorpusID:194914879>). Archived (https://web.archive.org/web/20230801100752/https://impactum-journals.uc.pt/matlit/article/view/2182-8830_6-1_4) from the original on 2023-08-01. Retrieved 2023-08-01.
25. Patti, Emanuela (2021-12-17). "Umberto Eco's Opera Aperta and the Birth of Italian Electronic Literature" (<https://doi.org/10.3828%2Fmlo.v0i0.381>). *Modern Languages Open*. **1**. doi:[10.3828/mlo.v0i0.381](https://doi.org/10.3828/mlo.v0i0.381) (<https://doi.org/10.3828%2Fmlo.v0i0.381>). ISSN [2052-5397](https://search.worldcat.org/issn/2052-5397) (<https://search.worldcat.org/issn/2052-5397>). S2CID [245288627](https://api.semanticscholar.org/CorpusID:245288627) (<https://api.semanticscholar.org/CorpusID:245288627>).
26. Worthy, R. J. (1964). "A New American Poet Speaks: The Works of A.B." (<https://search.informit.org/doi/abs/10.3316/INFORMAT.963392866491342>) *Meanjin Quarterly*. **23** (2): 166–172.
27. Skains, R. Lyle (2023). *Neverending Stories: The Popular Emergence of Digital Fiction*. New York: Bloomsbury Academic. pp. 27–28.
28. "ELIZA (Encyclopedia entry)" (<https://elmcip.net/creative-work/eliza>). *ELMCIP: Electronic Literature Knowledge Base*. Archived (<https://web.archive.org/web/20221012123124/https://elmcip.net/creative-work/eliza>) from the original on 2022-10-12. Retrieved 2022-10-12.
29. Nelson, T. H. (1965-08-24). "Complex information processing: A file structure for the complex, the changing and the indeterminate" (<https://dl.acm.org/doi/10.1145/800197.806036>). *Proceedings of the 1965 20th national conference*. ACM '65. New York, NY, USA: Association for Computing Machinery. pp. 84–100. doi:[10.1145/800197.806036](https://doi.org/10.1145/800197.806036) (<https://doi.org/10.1145%2F800197.806036>). ISBN [978-1-4503-7495-8](https://doi.org/10.1145/800197.806036).
30. Finborud, Lars Mørch. "Institusjonell kjærlighet" [A Marriage of Convenience]. *Lyd og ulydighet: Ny Musikk siden 1938* (<https://nymusikk.no/en/news/a-marriage-of-institusjonell-convenience>). Translated by Mackie, Saân. Archived (<https://web.archive.org/web/20230801090115/https://nymusikk.no/en/news/a-marriage-of-institusjonell-convenience>) from the original on 2023-08-01. Retrieved 2023-08-01.
31. Rudi, Jøran (January 2023). "Balancing Conflicting Types of Authenticity in the Reconstruction of Sound and Music in the Installation Work Blikk (1970)" (<https://doi.org/10.1111%2Fcura.12539>). *Curator: The Museum Journal*. **66** (1): 85–106. doi:[10.1111/cura.12539](https://doi.org/10.1111/cura.12539) (<https://doi.org/10.1111%2Fcura.12539>). ISSN [0011-3069](https://search.worldcat.org/issn/0011-3069) (<https://search.worldcat.org/issn/0011-3069>). S2CID [256648015](https://api.semanticscholar.org/CorpusID:256648015) (<https://api.semanticscholar.org/CorpusID:256648015>).
32. Donahue, Evan (2023-03-17). "All the Microworld's a Stage: Realism in Interactive Fiction and Artificial Intelligence" (<https://doi.org/10.1215/00029831-10575049>). *American Literature*. **95** (2): 229–254. doi:[10.1215/00029831-10575049](https://doi.org/10.1215/00029831-10575049) (<https://doi.org/10.1215%2F00029831-10575049>). ISSN [0002-9831](https://search.worldcat.org/issn/0002-9831) (<https://search.worldcat.org/issn/0002-9831>). S2CID [257617995](https://api.semanticscholar.org/CorpusID:257617995) (<https://api.semanticscholar.org/CorpusID:257617995>).
33. Montfort, Nick (2005). *Twisty little passages: an approach to interactive fiction*. Cambridge, MA; London: MIT Press. pp. 84–85. ISBN [978-0-262-13436-1](https://doi.org/10.1111/9780262134361).

34. Jerz, Dennis G. (2007). "Somewhere Nearby is Colossal Cave: Examining Will Crowther's Original "Adventure" in Code and in Kentucky" (<https://www.proquest.com/docview/2555208725>). *Digital Humanities Quarterly*. **1** (2). ProQuest 2555208725 (<https://www.proquest.com/docview/2555208725>).
35. Peel, Jeremy (2023-01-18). "Colossal Cave review" (<https://www.pcgamer.com/colossal-cave-review/>). *PC Gamer*. Archived (<https://web.archive.org/web/20230801093545/https://www.pcgamer.com/colossal-cave-review/>) from the original on 2023-08-01. Retrieved 2023-08-01.
36. Worrall, William (2023-01-18). "Colossal Cave Review" (<https://techraptor.net/gaming/reviews/colossal-cave-review>). *TechRaptor*. Archived (<https://web.archive.org/web/20230801093545/https://techraptor.net/gaming/reviews/colossal-cave-review>) from the original on 2023-08-01. Retrieved 2023-08-01.
37. Reinhard, Andrew (2021-03-01). "Colossal Cave Archaeology: Epigraphy, FORTRAN Code-Artifacts, and the Ur -Game" (<https://www.journals.uchicago.edu/doi/10.1086/713375>). *Near Eastern Archaeology*. **84** (1): 86–92. doi:10.1086/713375 (<https://doi.org/10.1086%2F713375>). ISSN 1094-2076 (<https://search.worldcat.org/issn/1094-2076>). S2CID 232124673 (<https://api.semanticscholar.org/CorpusID:232124673>). Archived (<https://web.archive.org/web/20230801093546/https://www.journals.uchicago.edu/doi/10.1086/713375>) from the original on 2023-08-01. Retrieved 2023-08-01.
38. Lebling; Blank; Anderson (1979). "Special Feature Zork: A Computerized Fantasy Simulation Game" (<https://ieeexplore.ieee.org/document/1658697>). *Computer*. **12** (4): 51–59. doi:10.1109/MC.1979.1658697 (<https://doi.org/10.1109%2FMC.1979.1658697>). ISSN 0018-9162 (<https://search.worldcat.org/issn/0018-9162>). S2CID 7845131 (<https://api.semanticscholar.org/CorpusID:7845131>). Archived ([https://ieeexplore.ieee.org/document/1658697/](https://web.archive.org/web/20230515143527/https://ieeexplore.ieee.org/document/1658697/)) from the original on 2023-05-15. Retrieved 2023-08-07.
39. Mason, Graeme (2022-09-22). "'I saw the possibility of what could be done – so I did it': revolutionary video game The Hobbit turns 40" (<https://www.theguardian.com/games/2022/sep/22/i-saw-the-possibility-of-what-could-be-done-so-i-did-it-revolutionary-video-game-the-hobbit-turns-40>). *The Guardian*. ISSN 0261-3077 (<https://search.worldcat.org/issn/0261-3077>). Archived (<https://web.archive.org/web/20231002023417/https://www.theguardian.com/games/2022/sep/22/i-saw-the-possibility-of-what-could-be-done-so-i-did-it-revolutionary-video-game-the-hobbit-turns-40>) from the original on 2023-10-02. Retrieved 2023-08-01.
40. Spinoza, Dani (2017). "Toward a Theory of Canadian Digital Poetics" (https://www.erudit.org/en/journals/scl/2016-v41-n1-scl42_2/scl42_2art12/). *Studies in Canadian Literature / Études en littérature canadienne*. **42** (2): 237–255. ISSN 0380-6995 (<https://search.worldcat.org/issn/0380-6995>). Archived (https://web.archive.org/web/20230801100752/https://www.erudit.org/en/journals/scl/2016-v41-n1-scl42_2/scl42_2art12/) from the original on 2023-08-01. Retrieved 2023-08-01.
41. Rettberg, Jill Walker (2012). "Electronic Literature Seen from a Distance The Beginnings of a Field" (<https://hdl.handle.net/1956/6272>). *Dichtung Digital* (41). hdl:1956/6272 (<https://hdl.handle.net/1956%2F6272>).
42. Bolter, Jay David; Joyce, Michael (1987). "Hypertext and creative writing" (<http://portal.acm.org/citation.cfm?doid=317426.317431>). *Proceeding of the ACM conference on Hypertext - HYPERTEXT '87*. ACM Press. pp. 41–50. doi:10.1145/317426.317431 (<https://doi.org/10.1145%2F317426.317431>). ISBN 978-0-89791-340-9.
43. Cronin, Susie (2018-08-10). "Can We (Still) Speak of a 'French' Digital Literature?" (https://impactum-journals.uc.pt/matlit/article/view/2182-8830_6-3_4). *Matlit Revista do Programa de Doutoramento em Materialidades da Literatura*. **6** (3): 47–58. doi:10.14195/2182-8830_6-3_4 (https://doi.org/10.14195%2F2182-8830_6-3_4). ISSN 2182-8830 (<https://search.worldcat.org/issn/2182-8830>).

44. Bouchardon, Serge (2012-09-05). "Digital Literature in France" (<http://www.utc.fr/~bouchard/articles/bouchardon-Dichtung-Digital-2012.pdf>) (PDF). *Dichtung Digital*. doi:10.25969/mediarep/17745 (<https://doi.org/10.25969%2Fmediarep%2F17745>).
45. Carter, Richard (2016). "Virtual Literatures: Technology, Agency, Meaning" (<https://www.euppublishing.com/doi/10.3366/count.2016.0064>). *CounterText*. **2** (3): 338–355. doi:10.3366/count.2016.0064 (<https://doi.org/10.3366%2Fcount.2016.0064>). ISSN 2056-4406 (<https://search.worldcat.org/issn/2056-4406>). "Jeffrey Shaw's The Legible City (1988–91) was an early and often discussed series of installations that pioneered the staging of literary expression within a simulated virtual space."
46. Rößler, Theresa (23 April 2020). "Dirk Groeneveld, Jeffrey Shaw | The Legible City | 1988 | ZKM" (<https://zkm.de/en/artwork/the-legible-city-0>). *ZKM Center for Art and Media Karlsruhe*. Retrieved 2024-02-09.
47. Bolter, J. David and Michael Joyce (1987). "Hypertext and Creative Writing", Proceedings of ACM Hypertext 1987, Chapel Hill, North Carolina, United States, pages 41-50
48. Barnet, Belinda (2012). "Machine Enhanced (Re)minding: the Development of Storyspace" (<http://www.digitalhumanities.org/dhq/vol/6/2/000128/000128.html>). *Digital Humanities Quarterly*. **6** (2).
49. Aarseth, Espen J. Cybertext: Perspectives on Ergodic Literature. Baltimore: Johns Hopkins UP, 1997
50. Moulthrop, Stuart (2021). "Hypertext Fiction Ever After". *Electronic Literature as Digital Humanities: Contexts, Forms, and Practice* (<https://library.oapen.org/bitstream/handle/20.500.12657/58859/1/9781501363481.pdf#page=162>) (PDF). Bloomsbury Academic. p. 157. ISBN 978-1-5013-6349-8. Archived (<https://web.archive.org/web/20230731175102/https://library.oapen.org/bitstream/handle/20.500.12657/58859/1/9781501363481.pdf#page=162>) (PDF) from the original on 2023-07-31. Retrieved 2023-07-31.
51. Hayles, Nancy Katherine; Burdick, Anne; Lunenfeld, Peter (2002). *Writing machines*. Mediawork pamphlet. Cambridge, MA: MIT Press. p. 37. ISBN 978-0-262-58215-5.
52. Coover, Robert (1993-08-23). "Hyperfiction: Novels for the Computer" (<https://archive.nytimes.com/www.nytimes.com/books/98/09/27/specials/coover-hyperfiction.html>). *New York Times*. Retrieved 2023-11-25.
53. O'Sullivan, James; Grigar, Dene (2018). "The Origins of Electronic Literature as Net/Web Art". In Brügger, Niels; Milligan, Ian (eds.). *The SAGE Handbook of Web History*. Sage Publishing. pp. 429–430. ISBN 9781526455444.
54. Picot, Edward (2002). "Some versions of hyperfiction" (<https://www.proquest.com/docview/2418404835>). *Poetry Nation Review*. **29** (2): 52–54. ProQuest 2418404835 (<https://www.proquest.com/docview/2418404835>). Retrieved 2023-08-01.
55. "trAce/Alt-X Hypertext Competition 1998 Results" (<http://unknownhypertext.com/presskit/trace.html>). *unknownhypertext.com*. 1998. Archived (<https://web.archive.org/web/20171216093711/http://unknownhypertext.com/presskit/trace.html>) from the original on 2017-12-16. Retrieved 2017-10-17.
56. Rettberg, Scott; Gillespie, William; Stratton, Dirk; Marquadt, Frank (2011) [1998]. Borràs, Laura; Memmott, Talan; Raley, Rita; Stefans, Brian (eds.). "The Unknown" (http://collection.eliterature.org/2/works/rettberg_theunknown.html). *collection.eliterature.org*. Archived (https://web.archive.org/web/20171013151854/http://collection.eliterature.org/2/works/rettberg_theunknown.html) from the original on 2017-10-13. Retrieved 2017-10-17.
57. Kolb, David A. (2012). "Story/Story". *Proceedings of the 23rd ACM conference on Hypertext and social media*. HT '12. New York: ACM. pp. 99–102. doi:10.1145/2309996.2310013 (<https://doi.org/10.1145%2F2309996.2310013>). ISBN 9781450313353. S2CID 208938632 (<https://api.semanticscholar.org/CorpusID:208938632>).

58. Pisarski, Mariusz (2016). "Collaboration in e-literature" (<https://www.ceeol.com/search/article-detail?id=444636>). *World Literature Studies*. **8** (3): 78–89. ISSN 1337-9275 (<https://search.worldcat.org/issn/1337-9275>). Archived (<https://web.archive.org/web/20221012123123/http://www.ceeol.com/search/article-detail?id=444636>) from the original on 2022-10-12. Retrieved 2022-10-12.
59. Ciccicocco, David (2007). "Fluid or Overflowing: The Unknown and *water always writes in *plural". *Reading Network Fiction* (<https://archive.org/details/readingnetworkfi0000cicc>). University of Alabama Press. pp. 124–159. ISBN 9780817315894.
60. Desrochers, Nadine; Tomaszek, Patricia (2014). "Bridging The Unknown: An Interdisciplinary Case Study of Paratext in Electronic Literature" (<https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/12174/Desrochers-N-Paratextual-Theory-Chapitre9.pdf?sequence=1&isAllowed=y>) (PDF). *Examining paratextual theory and its applications in digital culture*. pp. 160–189. [hdl:1866/12174](https://hdl.handle.net/1866%2F12174) (<https://hdl.handle.net/1866%2F12174>).
61. "Electronic Literature Organization – To facilitate and promote the writing, publishing, and reading of literature in electronic media" (<https://eliterature.org/>). Retrieved 2024-02-11.
62. Rettberg, Scott (2012-09-05). "Developing an Identity for the Field of Electronic Literature Reflections on the Electronic Literature Organization Archives". *Dichtung Digital. Journal für Kunst und Kultur digitaler Medien*. **14** (1): 1–33. doi:10.25969/mediarep/17747 (<https://doi.org/10.25969%2Fmediarep%2F17747>).
63. "ELO History" (<https://eliterature.org/elo-history/>). Electronic Literature Organization. Archived (<https://web.archive.org/web/20230610211449/https://eliterature.org/elo-history/>) from the original on 2023-06-10. Retrieved 2023-07-31.
64. Rettberg, Jill Walker (2014-07-06). "Visualising Networks of Electronic Literature: Dissertations and the Creative Works They Cite" (<http://electronicbookreview.com/essay/visualising-networks-of-electronic-literature-dissertations-and-the-creative-works-they-cite/>). *Electronic Book Review*. Archived (<https://web.archive.org/web/20221017084326/http://electronicbookreview.com/essay/visualising-networks-of-electronic-literature-dissertations-and-the-creative-works-they-cite/>) from the original on 2022-10-17. Retrieved 2022-10-17.
65. Kim, Kyoung-hwa Yonnie (2014). "Genealogy of Mobile Creativity: A Media Archaeological Approach to Literary Practice in Japan". In Goggin, Gerald; Hjorth, Larissa (eds.). *The Routledge companion to mobile media*. New York: Routledge. ISBN 978-0-415-80947-4. OCLC 864429273 (<https://search.worldcat.org/oclc/864429273>).
66. T, Shanmugapriya; Menon, Nirmala (2018). "Locating New Literary Practices in Indian Digital Spaces" (https://impactum-journals.uc.pt/matlit/article/view/2182-8830_6-1_11/4678). *MatLit: Materialities of Literature*. **6**: 159–174. doi:10.14195/2182-8830_6-1_11 (https://doi.org/10.14195/2182-8830_6-1_11). S2CID 194930590 (<https://api.semanticscholar.org/CorpusID:194930590>). Archived (https://web.archive.org/web/20221113105623/https://impactum-journals.uc.pt/matlit/article/view/2182-8830_6-1_11/4678) from the original on 2022-11-13. Retrieved 2022-11-13.
67. Walker, Jill (2005). "Distributed Narrative: Telling Stories Across Networks". In Consalvo, Mia; Hunsinger, Jeremy; Baym, Nancy (eds.). *The 2005 Association of Internet Researchers Annual*. New York: Peter Lang. pp. 91–102.
68. Shanmugapriya, T; Menon, Nirmala (2019). "First and Second Waves of Indian Electronic - ProQuest" (<https://www.proquest.com/docview/2391975967>). *Journal of Comparative Literature and Aesthetics*. **42** (4). ProQuest 2391975967 (<https://www.proquest.com/docview/2391975967>). Retrieved 2023-08-01.
69. RoGue (2004). "Cloak Room" (<http://cloakroom.blogspot.com/>). *Cloak Room* (Blog documenting and supporting a SMS novel). Archived (<https://web.archive.org/web/20230801100751/http://cloakroom.blogspot.com/>) from the original on 2023-08-01. Retrieved 2023-08-01.

70. Koskima, Raine (2004). "These waves of memories: A hyperfiction by Caitlin Fisher" (<http://mediarep.org/handle/doc/18621>). *Dichtung Digital*. **6** (3): 1–11.
doi:10.25969/MEDIAREP/17662 (<https://doi.org/10.25969%2FMEDIAREP%2F17662>).
ISSN 1617-6901 (<https://search.worldcat.org/issn/1617-6901>). Archived (<https://web.archive.org/web/20221017072643/https://mediarep.org/handle/doc/18621>) from the original on 2022-10-17. Retrieved 2023-08-01.
71. Hayles, N. Katherine (2001). "Metaphoric networks in 'Lexia to perplexia'" (<https://www.tandfonline.com/doi/full/10.1076/digc.12.3.133.3226>). *Digital Creativity*. **12** (3): 133–139.
doi:10.1076/digc.12.3.133.3226 (<https://doi.org/10.1076%2Fdigc.12.3.133.3226>).
ISSN 1462-6268 (<https://search.worldcat.org/issn/1462-6268>). S2CID 45507019 (<https://api.semanticscholar.org/CorpusID:45507019>).
72. Swanstrom, Lisa (2011). ""Terminal Hopscotch": Navigating Networked Space in Talan Memmott's Lexia to Perplexia" (http://muse.jhu.edu/content/crossref/journals/contemporary_literature/v052/52.3.swanstrom.html). *Contemporary Literature*. **52** (3): 493–521.
doi:10.1353/cli.2011.0038 (<https://doi.org/10.1353%2Fcli.2011.0038>). ISSN 1548-9949 (<https://search.worldcat.org/issn/1548-9949>). S2CID 161529462 (<https://api.semanticscholar.org/CorpusID:161529462>).
73. Machado, Ana Maria; Campbell, Andy; Harper, Ian; Albuquerque e Aguilar, Ana; Oliveira, António (2018-08-10). "Inanimate Alice: The Story of the Series and its Impact in Portugal" (https://impactum-journals.uc.pt/matlit/article/view/2182-8830_6-3_8). *Matlit Revista do Programa de Doutoramento em Materialidades da Literatura*. **6** (3): 93–104.
doi:10.14195/2182-8830_6-3_8 (https://doi.org/10.14195%2F2182-8830_6-3_8).
ISSN 2182-8830 (<https://search.worldcat.org/issn/2182-8830>). S2CID 158620931 (<https://api.semanticscholar.org/CorpusID:158620931>). Archived (https://web.archive.org/web/20230801100753/https://impactum-journals.uc.pt/matlit/article/view/2182-8830_6-3_8) from the original on 2023-08-01. Retrieved 2023-08-01.
74. Wardrip-Fruin (1998). "The Impermanence Agent" (<http://www.hyperfiction.org/agent/>). www.impermanenceagent.org. with Adam Chapman, Brion Moss, and Duane Whitehurst. Archived (<https://web.archive.org/web/20221017084324/http://www.hyperfiction.org/agent/>) from the original on 2022-10-17. Retrieved 2022-10-17.
75. Segar, Emma (2017). "Blog fiction and its successors: The emergence of a relational poetics" (<https://doi.org/10.1177%2F1354856516678369>). *Convergence: The International Journal of Research into New Media Technologies*. **23** (1): 20–33.
doi:10.1177/1354856516678369 (<https://doi.org/10.1177%2F1354856516678369>).
ISSN 1354-8565 (<https://search.worldcat.org/issn/1354-8565>). S2CID 148607975 (<https://api.semanticscholar.org/CorpusID:148607975>).
76. Hellekson, Karen; Busse, Kristina (2014). *The Fan Fiction Studies Reader*. Iowa City: University of Iowa Press. ISBN 978-1-60938-227-8.
77. Roering, Johanna (2008). ""I Love Mer/Der. When They Aren't Together, I Die": Television Characters Blogging". In Hotz-Davies, Ingrid; Kirchhofer, Anton; Leppänen, Sirpa (eds.). *Internet Fictions*. Cambridge Scholars Publishing. ISBN 9781443803038.
78. Santana, Stephanie Bosch (2018). "From Nation to Network: Blog and Facebook Fiction from Southern Africa". *Research in African Literatures*. **49** (1): 187–208.
doi:10.2979/reseafrilite.49.1.11 (<https://doi.org/10.2979%2Freseafrilite.49.1.11>). ISSN 0034-5210 (<https://search.worldcat.org/issn/0034-5210>). JSTOR 10.2979/reseafrilite.49.1.11 (<https://www.jstor.org/stable/10.2979/reseafrilite.49.1.11>). S2CID 165564181 (<https://api.semanticscholar.org/CorpusID:165564181>).
79. Adenekan, Shola; Cousins, Helen (2013). "African Short Stories and the Online Writing Space". In Awadalla, Maggie; March-Russell, Paul (eds.). *The Postcolonial Short Story*. London: Palgrave Macmillan UK. pp. 199–213. doi:10.1057/9781137292087_13 (https://doi.org/10.1057%2F9781137292087_13). ISBN 978-1-349-33930-3.

80. Harris, Ashleigh (2018-07-03). "Introduction: African Street Literatures and the Global Publishing Go-Slow" (<https://doi.org/10.1080%2F00138398.2018.1540173>). *English Studies in Africa*. **61** (2): 1–8. doi:[10.1080/00138398.2018.1540173](https://doi.org/10.1080/00138398.2018.1540173). ISSN 0013-8398 (<https://search.worldcat.org/issn/0013-8398>). S2CID [165665676](https://api.semanticscholar.org/CorpusID:165665676) (<https://api.semanticscholar.org/CorpusID:165665676>).
81. Steele, Catherine Knight (October 2016). "The Digital Barbershop: Blogs and Online Oral Culture Within the African American Community" (<https://doi.org/10.1177%2F2056305116683205>). *Social Media + Society*. **2** (4): 205630511668320. doi:[10.1177/2056305116683205](https://doi.org/10.1177%2F2056305116683205) (<https://doi.org/10.1177%2F2056305116683205>). ISSN 2056-3051 (<https://search.worldcat.org/issn/2056-3051>). S2CID [157100087](https://api.semanticscholar.org/CorpusID:157100087) (<https://api.semanticscholar.org/CorpusID:157100087>).
82. Locke, Charley. "You Don't Want to Know What PTSD Is Like, but Pry, a Powerful iOS Game, Tries to Show You Anyway" (<https://www.wired.com/2016/06/pry-game-ptsd/>). *Wired*. ISSN 1059-1028 (<https://search.worldcat.org/issn/1059-1028>). Archived (<https://web.archive.org/web/20221012123124/https://www.wired.com/2016/06/pry-game-ptsd/>) from the original on 2022-10-12. Retrieved 2022-10-12.
83. Burr, Lauren (2015). "Bicycles, Bonfires and an Airport Apocalypse: The Poetics and Ethics of Netprov" (<http://hyperrhiz.io/hyperrhiz11/essays/bicycles-bonfires.html>). *Hyperrhiz: New Media Cultures* (11): 1. doi:[10.20415/hyp/011.e01](https://doi.org/10.20415/hyp/011.e01) (<https://doi.org/10.20415%2Fhyp%2F011.e01>). Archived (<https://web.archive.org/web/20221012123836/http://hyperrhiz.io/hyperrhiz11/essays/bicycles-bonfires.html>) from the original on 2022-10-12. Retrieved 2022-10-12.
84. Wittig, Rob (2021). *Netprov: Networked Improvised Literature for the Classroom and Beyond* (<https://www.fulcrum.org/concern/monographs/pc289m47x>). Ann Arbor, MI: Amherst College Press. doi:[10.3998/mpub.12387128](https://doi.org/10.3998/mpub.12387128) (<https://doi.org/10.3998%2Fmpub.12387128>). ISBN 978-1-943208-28-9. S2CID [245087341](https://api.semanticscholar.org/CorpusID:245087341) (<https://api.semanticscholar.org/CorpusID:245087341>). Archived (<https://web.archive.org/web/20221012123834/https://www.fulcrum.org/concern/monographs/pc289m47x>) from the original on 2022-10-12. Retrieved 2022-10-12.
85. Pâquet, Lili (2019). "Selfie-Help: The Multimodal Appeal of Instagram Poetry" (<https://onlinelibrary.wiley.com/doi/10.1111/jpcu.12780>). *The Journal of Popular Culture*. **52** (2): 296–314. doi:[10.1111/jpcu.12780](https://doi.org/10.1111/jpcu.12780) (<https://doi.org/10.1111%2Fjpcu.12780>). ISSN 0022-3840 (<https://search.worldcat.org/issn/0022-3840>). S2CID [167066039](https://api.semanticscholar.org/CorpusID:167066039) (<https://api.semanticscholar.org/CorpusID:167066039>). Archived (<https://web.archive.org/web/20211227114933/https://onlinelibrary.wiley.com/doi/10.1111/jpcu.12780>) from the original on 2021-12-27. Retrieved 2022-11-13.
86. Ellison, Cara (2013-04-10). "Anna Anthropy and the Twine revolution" (<https://www.theguardian.com/technology/gamesblog/2013/apr/10/anna-anthropy-twine-revolution>). *The Guardian*. ISSN 0261-3077 (<https://search.worldcat.org/issn/0261-3077>). Archived (<https://web.archive.org/web/20190203201742/https://www.theguardian.com/technology/gamesblog/2013/apr/10/anna-anthropy-twine-revolution>) from the original on 2019-02-03. Retrieved 2023-08-01.
87. Robertson, Adi (2021-03-10). "Text Adventures: how Twine remade gaming" (<https://www.theverge.com/c/22321816/csk-twine>). *The Verge*. Retrieved 2023-08-01.
88. Ensslin, Astrid (2020-03-31), "Hypertext Theory" (<https://oxfordre.com/literature/display/10.1093/acrefore/9780190201098.001.0001/acrefore-9780190201098-e-982>), *Oxford Research Encyclopedia of Literature*, doi:[10.1093/acrefore/9780190201098.013.982](https://doi.org/10.1093/acrefore/9780190201098.013.982) (<https://doi.org/10.1093%2Facrefore%2F9780190201098.013.982>), ISBN 978-0-19-020109-8, archived (<https://web.archive.org/web/20230629083724/https://oxfordre.com/literature/display/10.1093/acrefore/9780190201098.001.0001/acrefore-9780190201098-e-982>) from the original on 2023-06-29, retrieved 2023-07-31

89. O'Flynn, Siobhan (2019-05-24). "Media Fluid and Media Fluent, E-Literature in the Era of Experience Design" (<http://hyperrhiz.io/hyperrhiz20/interfaces/3-oflynn-media-fluid.html>). *Hyperrhiz: New Media Cultures* (20). doi:10.20415/hyp/020.int03 (<https://doi.org/10.20415%2Fhyp%2F020.int03>). S2CID 182743384 (<https://api.semanticscholar.org/CorpusID:182743384>). Archived (<https://web.archive.org/web/20230731185913/http://hyperrhiz.io/hyperrhiz20/interfaces/3-oflynn-media-fluid.html>) from the original on 2023-07-31. Retrieved 2023-07-31.
90. Skains, R. Lyle (2023-07-20), "The Influence of Digital Platforms on Authors of Electronic Literature and Interactive Digital Narratives" (<https://www.taylorfrancis.com/books/9781003119739/chapters/10.4324/9781003119739-20>), *The Routledge Companion to Literary Media* (1st ed.), London: Routledge, pp. 209–220, doi:10.4324/9781003119739-20 (<https://doi.org/10.4324%2F9781003119739-20>), ISBN 978-1-003-11973-9, retrieved 2023-07-31
91. MacDonald, Keza (2018-04-26). "Games console: Dan Hett, the indie game designer pouring his grief into interactive art" (<https://www.theguardian.com/games/2018/apr/26/dan-hett-indie-games-designer-manchester-arena-bombing>). *The Guardian*. ISSN 0261-3077 (<https://search.worldcat.org/issn/0261-3077>). Archived (<https://web.archive.org/web/20230726070856/https://www.theguardian.com/games/2018/apr/26/dan-hett-indie-games-designer-manchester-arena-bombing>) from the original on 2023-07-26. Retrieved 2023-07-31.
92. Querubín, Natalia Sánchez; Niederer, Sabine (2022-10-31). "Climate futures: Machine learning from cli-fi" (<http://journals.sagepub.com/doi/10.1177/13548565221135715>). *Convergence: The International Journal of Research into New Media Technologies*. 30: 285–303. doi:10.1177/13548565221135715 (<https://doi.org/10.1177%2F13548565221135715>). ISSN 1354-8565 (<https://search.worldcat.org/issn/1354-8565>). S2CID 253321350 (<https://api.semanticscholar.org/CorpusID:253321350>). Archived (<https://web.archive.org/web/20221109100812/https://journals.sagepub.com/doi/10.1177/13548565221135715>) from the original on 2022-11-09. Retrieved 2022-11-09.
93. Sloan, Robin; Hamilton, Diane; Triantafyllou, Eugenia; Liu, Ken; Parrish, Allison; Wijeratne, Yudhanjaya; Garcia-Rosas, Nelly Geraldine; Tolabi, Wole; Hebert, Ernest. "Wordcraft Writers Workshop" (<https://wordcraft-writers-workshop.appspot.com/>). *Wordcraft Writers Workshop*. Retrieved 2022-11-09.
94. Henrickson, Leah (2020-05-29), "Authorship in Computer-Generated Texts" (<https://oxfordre.com/literature/view/10.1093/acrefore/9780190201098.001.0001/acrefore-9780190201098-e-1226>), *Oxford Research Encyclopedia of Literature*, Oxford University Press, doi:10.1093/acrefore/9780190201098.013.1226 (<https://doi.org/10.1093%2Facrefore%2F9780190201098.013.1226>), ISBN 978-0-19-020109-8, archived (<https://web.archive.org/web/20221013235518/https://oxfordre.com/literature/view/10.1093/acrefore/9780190201098.001.0001/acrefore-9780190201098-e-1226>) from the original on 2022-10-13, retrieved 2023-08-01
95. "Electronic Literature Collection Volume 4" (<https://collection.eliterature.org/4/about>). *collection.eliterature.org*. Retrieved 2024-02-14.
96. Menon, Nirmala; T, Shanmugapriya; Joseph, Justy; Sutton, Deborah (2023-12-13). *Indian Electronic Literature Anthology* (<https://iitkship.iiti.ac.in/site/books/e/10.57004/book1/>). Indian Institute of Technology – Knowledge Sharing in Publishing. doi:10.57004/book1 (<https://doi.org/10.57004%2Fbook1>). ISBN 978-93-5811-313-6.
97. Ensslin, Astrid (2022). *Pre-web digital publishing and the lore of electronic literature*. Cambridge: Cambridge University Press. ISBN 978-1-108-90316-5. OCLC 1310979695 (<https://search.worldcat.org/oclc/1310979695>).
98. Jordan, Spencer (2020). *Postdigital storytelling : poetics, praxis, research*. Abingdon, Oxon: Routledge. ISBN 978-1-138-08350-9. OCLC 1111641012 (<https://search.worldcat.org/oclc/1111641012>).
99. Aarseth, Espen J. (1997). *Cybertext: perspectives on ergodic literature*. Baltimore, MD; London: Johns Hopkins University Press. p. 9. ISBN 978-0-8018-5578-8.

100. Landow, George P. (2006). *Hypertext 3.0: critical theory and new media in an era of globalization*. Parallax (3rd ed.). Baltimore (Md.): Johns Hopkins university press. p. 4. ISBN 978-0-8018-8256-2.
101. Ensslin, Astrid; Bell, Alice (2021). *Digital fiction and the unnatural: transmedial narrative theory, method, and analysis*. Theory and interpretation of narrative. Columbus: The Ohio State University Press. ISBN 978-0-8142-1456-5.
102. Douglas, Jane Yellowleas (1994). "How Do I Stop this Thing? Closure and Indeterminacy in Interactive Narratives". In Landow, George (ed.). *Hyper/Text/Theory*. Baltimore, MD: Johns Hopkins University Press. pp. 59 and 113.
103. Douglas, J. Yellowleas (2000). *The end of books or books without end ? reading interactive narratives*. Ann Arbor (Mich.: University of Michigan press. p. 47). ISBN 978-0-472-11114-5.
104. Hayles, N. Katherine (2008). *Electronic literature: new horizons for the literary*. University of Notre Dame Ward-Phillips lectures in English language and literature. Notre Dame, Ind: University of Notre Dame. ISBN 978-0-268-03085-8.
105. Montfort, Nick and Noah Wardrip-Fruin "Acid-Free Bits: Recommendations for Long-Lasting Electronic Literature" (<http://www.eliterature.org/pad/afb.html>) Archived (<https://web.archive.org/web/20091010004844/http://www.eliterature.org/pad/afb.html>) 2009-10-10 at the Wayback Machine. The Electronic Literature Organization, 2004.
106. Alan Liu, David Durand, Nick Montfort, Merrilee Proffitt, Liam R. E. Quin, Jean-Hugues Réty, and Noah Wardrip-Fruin. "2005 "Born-Again Bits: A Framework for Migrating Electronic Literature" (<http://www.eliterature.org/pad/bab.html>) Archived (<https://web.archive.org/web/20091027162027/http://eliterature.org/pad/bab.html>) 2009-10-27 at the Wayback Machine. Electronic Literature Organization, 2005.
107. Clark, Lynda; Rossi, Giulia Carla; Wisdom, Stella (2020), Bosser, Anne-Gwenn; Millard, David E.; Hargood, Charlie (eds.), "Archiving Interactive Narratives at the British Library" (http://link.springer.com/10.1007/978-3-030-62516-0_27), *Interactive Storytelling*, Lecture Notes in Computer Science, vol. 12497, Cham: Springer International Publishing, pp. 300–313, doi:10.1007/978-3-030-62516-0_27 (https://doi.org/10.1007%2F978-3-030-62516-0_27), ISBN 978-3-030-62515-3, S2CID 225078876 (<https://api.semanticscholar.org/CorpusID:225078876>), archived (https://web.archive.org/web/20231109225919/https://link.springer.com/chapter/10.1007/978-3-030-62516-0_27) from the original on 2023-11-09, retrieved 2023-08-01
108. UK Web Archive. "New Media Writing Prize" (<https://www.webarchive.org.uk/en/ukwa/collection/2912>). UKWA: UK Web Archive. Retrieved 2023-08-01.
109. Electronic Literature Organization. "Welcome to The NEXT" (<https://the-next.eliterature.org/>). The NEXT. Archived (<https://web.archive.org/web/20230630183635/https://the-next.eliterature.org/>) from the original on 2023-06-30. Retrieved 2023-08-01.
110. ELMCIP. "Electronic Literature Knowledge Base" (<https://elmcip.net/>). elmcip.net. Archived (<https://web.archive.org/web/20230805181910/https://elmcip.net/>) from the original on 2023-08-05. Retrieved 2023-08-01.
111. "Electronic Literature Directory" (<https://directory.eliterature.org/>). Electronic Literature Organization. Archived (<https://web.archive.org/web/20221014103524/https://directory.eliterature.org/>) from the original on 2022-10-14. Retrieved 2022-10-17.
112. MAELD & ADELD (2020). "African Electronic Literature Alliance & African Diasporic Electronic Literature (AELEA & ADELI)" (<https://africanelit.org/>). africanelit.org. Archived (<https://web.archive.org/web/20221017084325/https://africanelit.org/>) from the original on 2022-10-17. Retrieved 2022-10-17.
113. "Electronic Literature Lab" (<http://dtc-wsuv.org/wp/ell/>). Electronic Literature Lab. Archived (<https://web.archive.org/web/20220928214254/https://dtc-wsuv.org/wp/ell/>) from the original on 2022-09-28. Retrieved 2022-10-12.

114. Pablo, Luis; Goicoechea, María (2014-12-31). "A Survey of Electronic Literature Collections" (<https://docs.lib.purdue.edu/clcweb/vol16/iss5/6>). *CLCWeb: Comparative Literature and Culture*. **16** (5). doi:10.7771/1481-4374.2558 (<https://doi.org/10.7771%2F1481-4374.2558>). ISSN 1481-4374 (<https://search.worldcat.org/issn/1481-4374>).
115. "Electronic Literature Collection" (<https://collection.eliterature.org/>). *collection.eliterature.org*. Archived (<https://web.archive.org/web/20230909084814/https://collection.eliterature.org/>) from the original on 2023-09-09. Retrieved 2023-07-31.
116. Pablo, Luis; Goicoechea, María (2014-12-31). "A Survey of Electronic Literature Collections" (<https://docs.lib.purdue.edu/clcweb/vol16/iss5/6>). *CLCWeb: Comparative Literature and Culture*. **16** (5). doi:10.7771/1481-4374.2558 (<https://doi.org/10.7771%2F1481-4374.2558>). ISSN 1481-4374 (<https://search.worldcat.org/issn/1481-4374>). Archived (<https://web.archive.org/web/20230731182606/https://docs.lib.purdue.edu/clcweb/vol16/iss5/6/>) from the original on 2023-07-31. Retrieved 2023-07-31.
117. "ELO Annual Awards" (<https://eliterature.org/elo-awards/>). Electronic Literature Organization. Archived (<https://web.archive.org/web/20230603190536/https://eliterature.org/elo-awards/>) from the original on 2023-06-03. Retrieved 2023-07-31.
118. Pope, James (2020). "Further on down the digital road: Narrative design and reading pleasure in five New Media Writing Prize narratives" (<http://journals.sagepub.com/doi/10.1177/1354856517726603>). *Convergence: The International Journal of Research into New Media Technologies*. **26** (1): 35–54. doi:10.1177/1354856517726603 (<https://doi.org/10.1177%2F1354856517726603>). ISSN 1354-8565 (<https://search.worldcat.org/issn/1354-8565>). S2CID 148623959 (<https://api.semanticscholar.org/CorpusID:148623959>). Archived (<https://web.archive.org/web/20230726062820/https://journals.sagepub.com/doi/10.1177/1354856517726603>) from the original on 2023-07-26. Retrieved 2023-07-31.
119. "Rhizome" (<https://rhizome.org/>). *Rhizome*. 1999-03-28. Retrieved 2024-04-08.
120. "trAce/Alt-X Hypertext Competition 1998 Results" (<https://unknownhypertext.com/presskit/trace.html>). *unknownhypertext.com*. Retrieved 2024-04-08.
121. "02-046 (Talan Memmott)" (https://www.brown.edu/Administration/News_Bureau/2002-03/02-046.html). *www.brown.edu*. Retrieved 2024-04-08.
122. "Chat Transcript: February 4, 2001 Sunday February 4 (in trAce WebBoard) 21:00 London, 16:00 New York Program chat--trAce New Media contest winner, Talan Memmott" (<https://www.eliterature.org/chats/020401.html>). *www.eliterature.org*. Retrieved 2024-04-08.
123. "HyperText: Explorations in Electronic Literature – Electronic Literature Organization" (<http://eliterature.org/hypertext/>). Retrieved 2024-04-08.
124. "Coding, phoning, chatting and blogging Managed by Writers for the Future" (<http://souzaesilva.com/wp-content/uploads/2017/06/trAce.pdf>) (PDF). *TrAce Online Writing Center*. May 28, 2004.

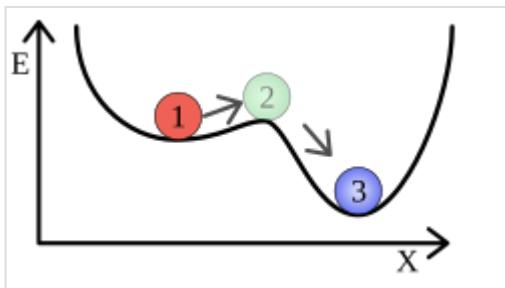
External links

Retrieved from "https://en.wikipedia.org/w/index.php?title=Electronic_literature&oldid=1291638457"



Metastability

In chemistry and physics, **metastability** is an intermediate energetic state within a dynamical system other than the system's state of least energy. A ball resting in a hollow on a slope is a simple example of metastability. If the ball is only slightly pushed, it will settle back into its hollow, but a stronger push may start the ball rolling down the slope. Bowling pins show similar metastability by either merely wobbling for a moment or tipping over completely. A common example of metastability in science is isomerisation. Higher energy isomers are long lived because they are prevented from rearranging to their preferred ground state by (possibly large) barriers in the potential energy.



A metastable state of weaker bond (1), a transitional "saddle" configuration (2) and a stable state of stronger bond (3).

During a metastable state of finite lifetime, all state-describing parameters reach and hold stationary values. In isolation:

- the state of least energy is the only one the system will inhabit for an indefinite length of time, until more external energy is added to the system (unique "absolutely stable" state);
- the system will spontaneously leave any other state (of higher energy) to eventually return (after a sequence of transitions) to the least energetic state.

The metastability concept originated in the physics of first-order phase transitions. It then acquired new meaning in the study of aggregated subatomic particles (in atomic nuclei or in atoms) or in molecules, macromolecules or clusters of atoms and molecules. Later, it was borrowed for the study of decision-making and information transmission systems.

Metastability is common in physics and chemistry – from an atom (many-body assembly) to statistical ensembles of molecules (viscous fluids, amorphous solids, liquid crystals, minerals, etc.) at molecular levels or as a whole (see Metastable states of matter and grain piles below). The abundance of states is more prevalent as the systems grow larger and/or if the forces of their mutual interaction are spatially less uniform or more diverse.

In dynamic systems (with feedback) like electronic circuits, signal trafficking, decisional, neural and immune systems, the time-invariance of the active or reactive patterns with respect to the external influences defines stability and metastability (see brain metastability below). In these systems, the equivalent of thermal fluctuations in molecular systems is the "white noise" that affects signal propagation and the decision-making.

Statistical physics and thermodynamics

Non-equilibrium thermodynamics is a branch of physics that studies the dynamics of statistical ensembles of molecules via unstable states. Being "stuck" in a thermodynamic trough without being at the lowest energy state is known as having kinetic stability or being kinetically persistent. The particular motion or

kinetics of the atoms involved has resulted in getting stuck, despite there being preferable (lower-energy) alternatives.

States of matter

Metastable states of matter (also referred as metastates) range from melting solids (or freezing liquids), boiling liquids (or condensing gases) and sublimating solids to supercooled liquids or superheated liquid-gas mixtures. Extremely pure, supercooled water stays liquid below 0 °C and remains so until applied vibrations or condensing seed doping initiates crystallization centers. This is a common situation for the droplets of atmospheric clouds.

Condensed matter and macromolecules

Metastable phases are common in condensed matter and crystallography. This is the case for anatase, a metastable polymorph of titanium dioxide, which despite commonly being the first phase to form in many synthesis processes due to its lower surface energy, is always metastable, with rutile being the most stable phase at all temperatures and pressures.^[1] As another example, diamond is a stable phase only at very high pressures, but is a metastable form of carbon at standard temperature and pressure. It can be converted to graphite (plus leftover kinetic energy), but only after overcoming an activation energy – an intervening hill. Martensite is a metastable phase used to control the hardness of most steel. Metastable polymorphs of silica are commonly observed. In some cases, such as in the allotropes of solid boron, acquiring a sample of the stable phase is difficult.^[2]

The bonds between the building blocks of polymers such as DNA, RNA, and proteins are also metastable. Adenosine triphosphate (ATP) is a highly metastable molecule, colloquially described as being "full of energy" that can be used in many ways in biology.^[3]

Generally speaking, emulsions/colloidal systems and glasses are metastable. The metastability of silica glass, for example, is characterised by lifetimes on the order of 10^{98} years^[4] (as compared with the lifetime of the universe, which is thought to be around 1.3787×10^{10} years).^[5]

Sandpiles are one system which can exhibit metastability if a steep slope or tunnel is present. Sand grains form a pile due to friction. It is possible for an entire large sand pile to reach a point where it is stable, but the addition of a single grain causes large parts of it to collapse.

The avalanche is a well-known problem with large piles of snow and ice crystals on steep slopes. In dry conditions, snow slopes act similarly to sandpiles. An entire mountainside of snow can suddenly slide due to the presence of a skier, or even a loud noise or vibration.

Quantum mechanics

Aggregated systems of subatomic particles described by quantum mechanics (quarks inside nucleons, nucleons inside atomic nuclei, electrons inside atoms, molecules, or atomic clusters) are found to have many distinguishable states. Of these, one (or a small degenerate set) is indefinitely stable: the ground state or global minimum.

All other states besides the ground state (or those degenerate with it) have higher energies.^[6] Of all these other states, the **metastable** states are the ones having lifetimes lasting at least 10^2 to 10^3 times longer than the shortest lived states of the set.^[7]

A *metastable state* is then long-lived (locally stable with respect to configurations of 'neighbouring' energies) but not eternal (as the global minimum is). Being excited – of an energy above the ground state – it will eventually decay to a more stable state, releasing energy. Indeed, above absolute zero, all states of a system have a non-zero probability to decay; that is, to spontaneously fall into another state (usually lower in energy). One mechanism for this to happen is through tunnelling.

Nuclear physics

Some energetic states of an atomic nucleus (having distinct spatial mass, charge, spin, isospin distributions) are much longer-lived than others (nuclear isomers of the same isotope), e.g. technetium-99m.^[8] The isotope tantalum-180m, although being a metastable excited state, is long-lived enough that it has never been observed to decay, with a half-life calculated to be least 4.5×10^{16} years,^{[9][10]} over 3 million times the current age of the universe.

Atomic and molecular physics

Some atomic energy levels are metastable. Rydberg atoms are an example of metastable excited atomic states. Transitions from metastable excited levels are typically those forbidden by electric dipole selection rules. This means that any transitions from this level are relatively unlikely to occur. In a sense, an electron that happens to find itself in a metastable configuration is trapped there. Since transitions from a metastable state are not impossible (merely less likely), the electron will eventually decay to a less energetic state, typically by an electric quadrupole transition, or often by non-radiative de-excitation (e.g., collisional de-excitation).

This slow-decay property of a metastable state is apparent in phosphorescence, the kind of photoluminescence seen in glow-in-the-dark toys that can be charged by first being exposed to bright light. Whereas spontaneous emission in atoms has a typical timescale on the order of 10^{-8} seconds, the decay of metastable states can typically take milliseconds to minutes, and so light emitted in phosphorescence is usually both weak and long-lasting.

Chemistry

In chemical systems, a system of atoms or molecules involving a change in chemical bond can be in a metastable state, which lasts for a relatively long period of time. Molecular vibrations and thermal motion make chemical species at the energetic equivalent of the top of a round hill very short-lived. Metastable states that persist for many seconds (or years) are found in energetic valleys which are not the lowest possible valley (point 1 in illustration). A common type of metastability is isomerism.

The stability or metastability of a given chemical system depends on its environment, particularly temperature and pressure. The difference between producing a stable vs. metastable entity can have important consequences. For instances, having the wrong crystal polymorph can result in failure of a drug while in storage between manufacture and administration.^[11] The map of which state is the most stable as a function of pressure, temperature and/or composition is known as a phase diagram. In regions where a

particular state is not the most stable, it may still be metastable. Reaction intermediates are relatively short-lived, and are usually thermodynamically unstable rather than metastable. The IUPAC recommends referring to these as *transient* rather than metastable.^[12]

Metastability is also used to refer to specific situations in mass spectrometry^[13] and spectrochemistry.^[14]

Electronic circuits

A digital circuit is supposed to be found in a small number of stable digital states within a certain amount of time after an input change. However, if an input changes at the wrong moment a digital circuit which employs feedback (even a simple circuit such as a flip-flop) can enter a metastable state and take an unbounded length of time to finally settle into a fully stable digital state.

Computational neuroscience

Metastability in the brain is a phenomenon studied in computational neuroscience to elucidate how the human brain recognizes patterns. Here, the term metastability is used rather loosely. There is no lower-energy state, but there are semi-transient signals in the brain that persist for a while and are different than the usual equilibrium state.

In philosophy

Gilbert Simondon invokes a notion of metastability for his understanding of systems that rather than resolve their tensions and potentials for transformation into a single final state rather, 'conserves the tensions in the equilibrium of metastability instead of nullifying them in the equilibrium of stability' as a critique of cybernetic notions of homeostasis.^[15]

See also

- False vacuum
- Hysteresis
- Metastate

References

1. Hanaor, Dorian A. H.; Sorrell, Charles C. (2011-02-01). "Review of the anatase to rutile phase transformation" (<https://doi.org/10.1007%2Fs10853-010-5113-0>). *Journal of Materials Science*. **46** (4): 855–874. Bibcode:2011JMatS..46..855H (<https://ui.adsabs.harvard.edu/abs/2011JMatS..46..855H>). doi:[10.1007/s10853-010-5113-0](https://doi.org/10.1007/s10853-010-5113-0) (<https://doi.org/10.1007%2Fs10853-010-5113-0>). ISSN [1573-4803](https://search.worldcat.org/issn/1573-4803) (<https://search.worldcat.org/issn/1573-4803>). S2CID 97190202 (<https://api.semanticscholar.org/CorpusID:97190202>).

2. van Setten; Uijtewaal; de Wijs; de Groot (2007). "Thermodynamic stability of boron: the role of defects and zero point motion" (<https://web.archive.org/web/20210415015024/https://pure.rug.nl/ws/files/2796591/2007JAmChemSocvSetten.pdf>) (PDF). *Journal of the American Chemical Society*. **129** (9): 2458–2465. doi:[10.1021/ja0631246](https://doi.org/10.1021/ja0631246) (<https://doi.org/10.1021%2Fa0631246>). PMID [17295480](https://pubmed.ncbi.nlm.nih.gov/17295480/) (<https://pubmed.ncbi.nlm.nih.gov/17295480/>). S2CID [961904](https://api.semanticscholar.org/CorpusID:961904) (<https://api.semanticscholar.org/CorpusID:961904>). Archived from the original (<https://pure.rug.nl/ws/files/2796591/2007JAmChemSocvSetten.pdf>) (PDF) on 2021-04-15. Retrieved 2019-07-08.
3. Haldane, J. B. S. (1964). "Eighteen: Genesis of Life" (https://books.google.com/books?id=G_Nc_DQAAQBAJ&pg=PA332). In D. R., Bates (ed.). *The Planet Earth* (2nd ed.). Germany: Pergamon Press. p. 332. ISBN [1483135993](#). Retrieved May 29, 2017. "This is a highly stable molecule. About 11,500 calories of free energy are liberated when it is hydrolyzed to phosphate and adenosine-diphosphate (ADP)."
4. M.I. Ojovan, W.E. Lee, S.N. Kalmykov. An introduction to nuclear waste immobilisation. Third edition, Elsevier, Amsterdam, p.323 (2019)
5. Planck Collaboration (2020). "Planck 2018 results. VI. Cosmological parameters" (<https://doi.org/10.1051%2F0004-6361%2F201833910>). *Astronomy & Astrophysics*. **641**. page A6 (see PDF page 15, Table 2: "Age/Gyr", last column). arXiv:1807.06209 (<https://arxiv.org/abs/1807.06209>). Bibcode:2020A&A...641A...6P (<https://ui.adsabs.harvard.edu/abs/2020A&A...641A...6P>). doi:[10.1051/0004-6361/201833910](https://doi.org/10.1051%2F0004-6361%2F201833910) (<https://doi.org/10.1051%2F0004-6361%2F201833910>). S2CID [119335614](https://api.semanticscholar.org/CorpusID:119335614) (<https://api.semanticscholar.org/CorpusID:119335614>).
6. Hobson, Art (2017). *Tales of the Quantum: Understanding Physics' Most Fundamental Theory* (<https://books.google.com/books?id=mGduDQAAQBAJ&pg=PA139>). Oxford University Press. ISBN [9780190679637](#).
7. Hodgman, S. S.; Dall, R. G.; Byron, L. J.; Baldwin, K. G. H.; Buckman, S. J.; Truscott, A. G. (2009-07-31). "Metastable helium: a new determination of the longest atomic excited-state lifetime" (<https://pubmed.ncbi.nlm.nih.gov/19792494/>). *Physical Review Letters*. **103** (5): 053002. Bibcode:2009PhRvL.103e3002H (<https://ui.adsabs.harvard.edu/abs/2009PhRvL.103e3002H>). doi:[10.1103/PhysRevLett.103.053002](https://doi.org/10.1103/PhysRevLett.103.053002) (<https://doi.org/10.1103%2FPhysRevLett.103.053002>). hdl:[10440/978](https://hdl.handle.net/10440%2F978) (<https://hdl.handle.net/10440%2F978>). ISSN [0031-9007](#) (<https://search.worldcat.org/issn/0031-9007>). PMID [19792494](https://pubmed.ncbi.nlm.nih.gov/19792494/) (https://pubmed.ncbi.nlm.nih.gov/19792494).
8. "Technetium-99m" (<http://hyperphysics.phy-astr.gsu.edu/hbase/nuclear/technetium.html>). Hyperphysics.
9. Conover, Emily (2016-10-03). "Rarest nucleus reluctant to decay" (<https://www.sciencenews.org/article/rarest-nucleus-reluctant-decay>). *Science News*. Retrieved 2016-10-05.
10. Lehnert, Björn; Hult, Mikael; Lutter, Guillaume; Zuber, Kai (2017). "Search for the decay of nature's rarest isotope ^{180m}Ta ". *Physical Review C*. **95** (4): 044306. arXiv:1609.03725 (<https://arxiv.org/abs/1609.03725>). Bibcode:2017PhRvC..95d4306L (<https://ui.adsabs.harvard.edu/abs/2017PhRvC..95d4306L>). doi:[10.1103/PhysRevC.95.044306](https://doi.org/10.1103/PhysRevC.95.044306) (<https://doi.org/10.1103%2FPhysRevC.95.044306>). S2CID [118497863](https://api.semanticscholar.org/CorpusID:118497863) (<https://api.semanticscholar.org/CorpusID:118497863>).
11. Process Chemistry in the Pharmaceutical Industry. Kumar G. Gadamesetti, editor. 1999, pp. 375–378
12. "transient (chemical) species" (<http://goldbook.iupac.org/T06451.html>). *IUPAC Gold Book*. 2014. doi:[10.1351/goldbook.T06451](https://doi.org/10.1351/goldbook.T06451) (<https://doi.org/10.1351%2Fgoldbook.T06451>).
13. "metastable ion in mass spectrometry" (<http://goldbook.iupac.org/M03874.html>). *IUPAC Gold Book*. doi:[10.1351/goldbook.M03874](https://doi.org/10.1351/goldbook.M03874) (<https://doi.org/10.1351%2Fgoldbook.M03874>).
14. "metastable state in spectrochemistry" (<http://goldbook.iupac.org/M03876.html>). *IUPAC Gold Book*. doi:[10.1351/goldbook.M03876](https://doi.org/10.1351/goldbook.M03876) (<https://doi.org/10.1351%2Fgoldbook.M03876>).

15. Andrea Bardin, Marco Ferrari (27 March 2022). "Governing progress: From cybernetic homeostasis to Simondon's politics of metastability" (<https://journals.sagepub.com/doi/full/10.1177/00380261221084426>). *The Sociological Review*. **70** (2): 248–263.
doi:[10.1177/00380261221084426](https://doi.org/10.1177/00380261221084426) (<https://doi.org/10.1177%2F00380261221084426>).
-

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Metastability&oldid=1253317381>"



Runt pulse

In digital circuits, a **runt pulse** is a narrow pulse that, due to non-zero rise and fall times of the signal, does not reach a valid high or low level. A runt pulse may occur when switching between asynchronous clocks; or as the result of a race condition in which a signal takes two separate paths through a circuit, which may have different delays, and is then recombined to form a glitch; or when the output of a flip-flop becomes metastable.

Example

Some oscilloscopes provide a method for triggering on runt pulses. The oscilloscope triggers when the signal crosses one of two voltage thresholds, but not both.^[1]

References

1. "Oscilloscope triggering" (<http://www.tek.com/Measurement/cgi-bin/framed.pl?Document=/Measurement/scopes/selection/trigger.html&FrameSet=oscilloscopes>). Tektronix. Archived (<https://web.archive.org/web/20110928113849/http://www.tek.com/Measurement/cgi-bin/frame.d.pl?Document=%2FMeasurement%2Fscopes%2Fselection%2Ftrigger.html&FrameSet=oscilloscopes>) from the original on 2011-09-28. Retrieved 2008-05-20.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Runt_pulse&oldid=1232774933"



Integrated circuit

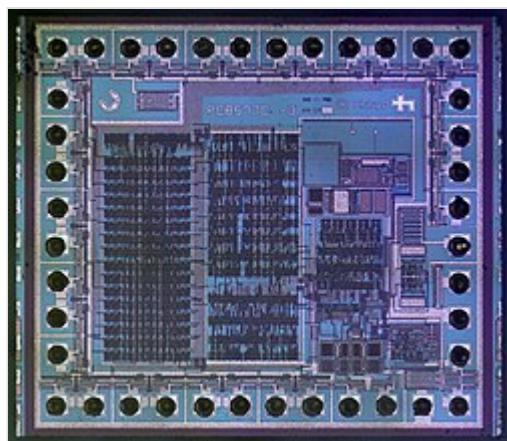
An **integrated circuit (IC)**, also known as a **microchip** or simply **chip**, is a set of electronic circuits, consisting of various electronic components (such as transistors, resistors, and capacitors) and their interconnections.^[1] These components are etched onto a small, flat piece ("chip") of semiconductor material, usually silicon.^[1] Integrated circuits are used in a wide range of electronic devices, including computers, smartphones, and televisions, to perform various functions such as processing and storing information. They have greatly impacted the field of electronics by enabling device miniaturization and enhanced functionality.

Integrated circuits are orders of magnitude smaller, faster, and less expensive than those constructed of discrete components, allowing a large transistor count.

The IC's mass production capability, reliability, and building-block approach to integrated circuit design have ensured the rapid adoption of standardized ICs in place of designs using discrete transistors. ICs are now used in virtually all electronic equipment and have revolutionized the world of electronics. Computers, mobile phones, and other home appliances are now essential parts of the structure of modern societies, made possible by the small size and low cost of ICs such as modern computer processors and microcontrollers.

Very-large-scale integration was made practical by technological advancements in semiconductor device fabrication. Since their origins in the 1960s, the size, speed, and capacity of chips have progressed enormously, driven by technical advances that fit more and more transistors on chips of the same size – a modern chip may have many billions of transistors in an area the size of a human fingernail. These advances, roughly following Moore's law, make the computer chips of today possess millions of times the capacity and thousands of times the speed of the computer chips of the early 1970s.

ICs have three main advantages over circuits constructed out of discrete components: size, cost and performance. The size and cost is low because the chips, with all their components, are printed as a unit by photolithography rather than being constructed one transistor at a time. Furthermore, packaged ICs use much less material than discrete circuits. Performance is high because the IC's components switch quickly and consume comparatively little power because of their small size and proximity. The main disadvantage of ICs is the high initial cost of designing them and the enormous capital cost of factory construction. This high initial cost means ICs are only commercially viable when high production volumes are anticipated.



A microscope image of an integrated circuit die used to control LCDs. The pinouts are the dark circles surrounding the integrated circuit.

Terminology

An *integrated circuit* is defined as:^[2]

A circuit in which all or some of the circuit elements are inseparably associated and electrically interconnected so that it is considered to be indivisible for the purposes of construction and commerce.

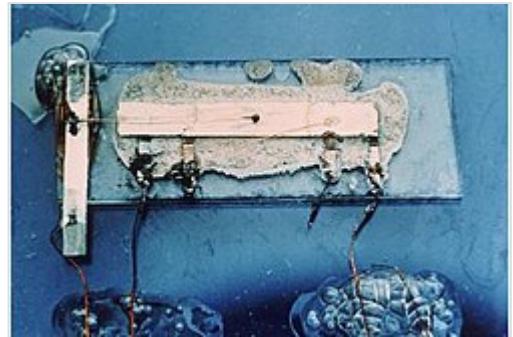
In strict usage, *integrated circuit* refers to the single-piece circuit construction originally known as a *monolithic integrated circuit*, which comprises a single piece of silicon.^{[3][4]} In general usage, circuits not meeting this strict definition are sometimes referred to as ICs, which are constructed using many different technologies, e.g. 3D IC, 2.5D IC, MCM, thin-film transistors, thick-film technologies, or hybrid integrated circuits. The choice of terminology frequently appears in discussions related to whether Moore's Law is obsolete.

History

An early attempt at combining several components in one device (like modern ICs) was the Loewe 3NF vacuum tube first made in 1926.^{[5][6]} Unlike ICs, it was designed with the purpose of tax avoidance, as in Germany, radio receivers had a tax that was levied depending on how many tube holders a radio receiver had. It allowed radio receivers to have a single tube holder. One million were manufactured, and were "a first step in integration of radioelectronic devices".^[7] The device contained an amplifier, composed of three triodes, two capacitors and four resistors in a six-pin device.^[8] Radios with the Loewe 3NF were less expensive than other radios,^[9] showing one of the advantages of integration over using discrete components, that would be seen decades later with ICs.^[10]

Early concepts of an integrated circuit go back to 1949, when German engineer Werner Jacobi^[11] (Siemens AG)^[12] filed a patent for an integrated-circuit-like semiconductor amplifying device^[13] showing five transistors on a common substrate in a three-stage amplifier arrangement. Jacobi disclosed small and cheap hearing aids as typical industrial applications of his patent. An immediate commercial use of his patent has not been reported.

Another early proponent of the concept was Geoffrey Dummer (1909–2002), a radar scientist working for the Royal Radar Establishment of the British Ministry of Defence. Dummer presented the idea to the public at the Symposium on Progress in Quality Electronic Components in Washington, D.C., on 7 May 1952.^[14] He gave many symposia publicly to propagate his ideas and unsuccessfully attempted to build such a circuit in 1956. Between 1953 and 1957, Sidney Darlington and Yasuo Tarui (Electrotechnical Laboratory) proposed similar chip designs where several transistors could share a common active area, but there was no electrical isolation to separate them from each other.^[11]



Jack Kilby's original integrated circuit; the world's first. Made from germanium with gold-wire interconnects.

The monolithic integrated circuit chip was enabled by the inventions of the planar process by Jean Hoerni and p–n junction isolation by Kurt Lehovec. Hoerni's invention was built on Carl Frosch and Lincoln Derick's work on surface protection and passivation by silicon dioxide masking and predeposition,^{[15][16][17]} as well as Fuller, Ditzenberger's and others work on the diffusion of impurities into silicon.^{[18][19][20][21][22]}

The first integrated circuits

A precursor idea to the IC was to create small ceramic substrates (so-called *micromodules*),^[23] each containing a single miniaturized component. Components could then be integrated and wired into a bidimensional or tridimensional compact grid. This idea, which seemed very promising in 1957, was proposed to the US Army by Jack Kilby^[23] and led to the short-lived Micromodule Program (similar to 1951's Project Tinkertoy).^{[23][24][25]} However, as the project was gaining momentum, Kilby came up with a new, revolutionary design: the IC.

Newly employed by Texas Instruments, Kilby recorded his initial ideas concerning the integrated circuit in July 1958, successfully demonstrating the first working example of an integrated circuit on 12 September 1958.^[26] In his patent application of 6 February 1959,^[27] Kilby described his new device as "a body of semiconductor material ... wherein all the components of the electronic circuit are completely integrated".^[28] The first customer for the new invention was the US Air Force.^[29] Kilby won the 2000 Nobel Prize in physics for his part in the invention of the integrated circuit.^[30]

However, Kilby's invention was not a true monolithic integrated circuit chip since it had external gold-wire connections, which would have made it difficult to mass-produce.^[31] Half a year after Kilby, Robert Noyce at Fairchild Semiconductor invented the first true monolithic IC chip.^{[32][31]} More practical than Kilby's implementation, Noyce's chip was made of silicon, whereas Kilby's was made of germanium, and Noyce's was fabricated using the planar process, developed in early 1959 by his colleague Jean Hoerni and included the critical on-chip aluminum interconnecting lines. Modern IC chips are based on Noyce's monolithic IC,^{[32][31]} rather than Kilby's.

NASA's Apollo Program was the largest single consumer of integrated circuits between 1961 and 1965.^[33]

TTL integrated circuits

Transistor-transistor logic (TTL) was developed by James L. Buie in the early 1960s at TRW Inc. TTL became the dominant integrated circuit technology during the 1970s to early 1980s.^[34]

Dozens of TTL integrated circuits were a standard method of construction for the processors of minicomputers and mainframe computers. Computers such as IBM 360 mainframes, PDP-11 minicomputers and the desktop Datapoint 2200 were built from bipolar integrated circuits,^[35] either TTL or the even faster emitter-coupled logic (ECL).



Robert Noyce invented the first monolithic integrated circuit in 1959. The chip was made from silicon.

MOS integrated circuits

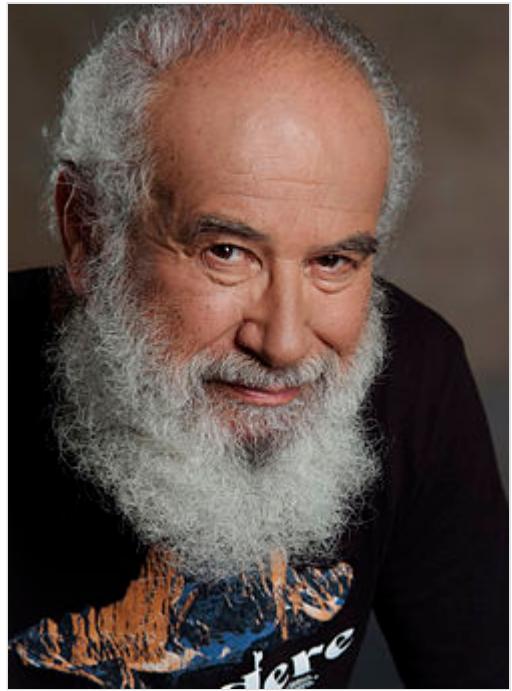
Nearly all modern IC chips are metal–oxide–semiconductor (MOS) integrated circuits, built from MOSFETs (metal–oxide–silicon field-effect transistors).^[36] The MOSFET invented at Bell Labs between 1955 and 1960,^{[15][37][16][38][39][40][17]} made it possible to build high-density integrated circuits.^[41] In contrast to bipolar transistors which required a number of steps for the p–n junction isolation of transistors on a chip, MOSFETs required no such steps but could be easily isolated from each other.^[42] Its advantage for integrated circuits was pointed out by Dawon Kahng in 1961.^[43] The list of IEEE milestones includes the first integrated circuit by Kilby in 1958,^[44] Hoerni's planar process and Noyce's planar IC in 1959.^[45]

The earliest experimental MOS IC to be fabricated was a 16-transistor chip built by Fred Heiman and Steven Hofstein at RCA in 1962.^[46] General Microelectronics later introduced the first commercial MOS integrated circuit in 1964,^[47] a 120-transistor shift register developed by Robert Norman.^[46] By 1964, MOS chips had reached higher transistor density and lower manufacturing costs than bipolar chips. MOS chips further increased in complexity at a rate predicted by Moore's law, leading to large-scale integration (LSI) with hundreds of transistors on a single MOS chip by the late 1960s.^[48]

Following the development of the self-aligned gate (silicon-gate) MOSFET by Robert Kerwin, Donald Klein and John Sarace at Bell Labs in 1967,^[49] the first silicon-gate MOS IC technology with self-aligned gates, the basis of all modern CMOS integrated circuits, was developed at Fairchild Semiconductor by Federico Faggin in 1968.^[50] The application of MOS LSI chips to computing was the basis for the first microprocessors, as engineers began recognizing that a complete computer processor could be contained on a single MOS LSI chip. This led to the inventions of the microprocessor and the microcontroller by the early 1970s.^[48] During the early 1970s, MOS integrated circuit technology enabled the very large-scale integration (VLSI) of more than 10,000 transistors on a single chip.^[51]

At first, MOS-based computers only made sense when high density was required, such as aerospace and pocket calculators. Computers built entirely from TTL, such as the 1970 Datapoint 2200, were much faster and more powerful than single-chip MOS microprocessors such as the 1972 Intel 8008 until the early 1980s.^[35]

Advances in IC technology, primarily smaller features and larger chips, have allowed the number of MOS transistors in an integrated circuit to double every two years, a trend known as Moore's law. Moore originally stated it would double every year, but he went on to change the claim to every two years in 1975.^[52] This increased capacity has been used to decrease cost and increase functionality. In general, as the feature size shrinks, almost every aspect of an IC's operation improves. The cost per transistor and the switching power consumption per transistor goes down, while the memory capacity and speed go up, through the relationships defined by Dennard scaling (MOSFET scaling).^[53] Because speed, capacity, and power consumption gains are apparent to the end user, there is fierce competition among the



Dov Frohman, an Israeli electrical engineer who developed the EPROM in 1969-1971

manufacturers to use finer geometries. Over the years, transistor sizes have decreased from tens of microns in the early 1970s to 10 nanometers in 2017^[54] with a corresponding million-fold increase in transistors per unit area. As of 2016, typical chip areas range from a few square millimeters to around 600 mm², with up to 25 million transistors per mm².^[55]

The expected shrinking of feature sizes and the needed progress in related areas was forecast for many years by the International Technology Roadmap for Semiconductors (ITRS). The final ITRS was issued in 2016, and it is being replaced by the International Roadmap for Devices and Systems.^[56]

Initially, ICs were strictly electronic devices. The success of ICs has led to the integration of other technologies, in an attempt to obtain the same advantages of small size and low cost. These technologies include mechanical devices, optics, and sensors.

- Charge-coupled devices, and the closely related active-pixel sensors, are chips that are sensitive to light. They have largely replaced photographic film in scientific, medical, and consumer applications. Billions of these devices are now produced each year for applications such as cellphones, tablets, and digital cameras. This sub-field of ICs won the Nobel Prize in 2009.^[57]
- Very small mechanical devices driven by electricity can be integrated onto chips, a technology known as microelectromechanical systems (MEMS). These devices were developed in the late 1980s^[58] and are used in a variety of commercial and military applications. Examples include DLP projectors, inkjet printers, and accelerometers and MEMS gyroscopes used to deploy automobile airbags.
- Since the early 2000s, the integration of optical functionality (optical computing) into silicon chips has been actively pursued in both academic research and in industry resulting in the successful commercialization of silicon based integrated optical transceivers combining optical devices (modulators, detectors, routing) with CMOS based electronics.^[59] Photonic integrated circuits that use light such as Lightelligence's PACE (Photonic Arithmetic Computing Engine) also being developed, using the emerging field of physics known as photonics.^[60]
- Integrated circuits are also being developed for sensor applications in medical implants or other bioelectronic devices.^[61] Special sealing techniques have to be applied in such biogenic environments to avoid corrosion or biodegradation of the exposed semiconductor materials.^[62]

As of 2018, the vast majority of all transistors are MOSFETs fabricated in a single layer on one side of a chip of silicon in a flat two-dimensional planar process. Researchers have produced prototypes of several promising alternatives, such as:

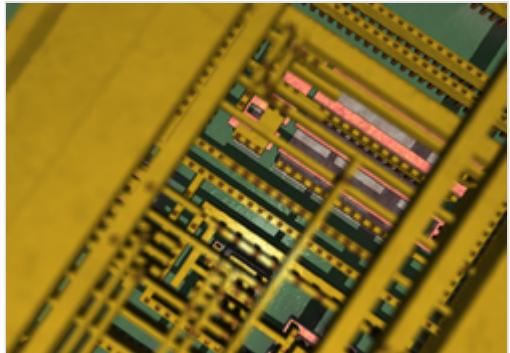
- various approaches to stacking several layers of transistors to make a three-dimensional integrated circuit (3DIC), such as through-silicon via, "monolithic 3D",^[63] stacked wire bonding,^[64] and other methodologies.
- transistors built from other materials: graphene transistors, molybdenite transistors, carbon nanotube field-effect transistor, gallium nitride transistor, transistor-like nanowire electronic devices, organic field-effect transistor, etc.
- fabricating transistors over the entire surface of a small sphere of silicon.^{[65][66]}
- modifications to the substrate, typically to make "flexible transistors" for a flexible display or other flexible electronics, possibly leading to a roll-away computer.

As it becomes more difficult to manufacture ever smaller transistors, companies are using multi-chip modules/chiplets, three-dimensional integrated circuits, package on package, High Bandwidth Memory and through-silicon vias with die stacking to increase performance and reduce size, without having to reduce the size of the transistors. Such techniques are collectively known as advanced packaging.^[67] Advanced packaging is mainly divided into 2.5D and 3D packaging. 2.5D describes approaches such as multi-chip modules while 3D describes approaches where dies are stacked in one way or another, such as package on package and high bandwidth memory. All approaches involve 2 or more dies in a single package.^{[68][69][70][71][72]} Alternatively, approaches such as 3D NAND stack multiple layers on a single die. A technique has been demonstrated to include microfluidic cooling on integrated circuits, to improve cooling performance^[73] as well as peltier thermoelectric coolers on solder bumps, or thermal solder bumps used exclusively for heat dissipation, used in flip-chip.^{[74][75]}

Design

The cost of designing and developing a complex integrated circuit is quite high, normally in the multiple tens of millions of dollars.^{[76][77]} Therefore, it only makes economic sense to produce integrated circuit products with high production volume, so the non-recurring engineering (NRE) costs are spread across typically millions of production units.

Modern semiconductor chips have billions of components, and are far too complex to be designed by hand. Software tools to help the designer are essential. Electronic design automation (EDA), also referred to as electronic computer-aided design (ECAD),^[78] is a category of software tools for designing electronic systems, including integrated circuits. The tools work together in a design flow that engineers use to design, verify, and analyze entire semiconductor chips. Some of the latest EDA tools use artificial intelligence (AI) to help engineers save time and improve chip performance.



Virtual detail of an integrated circuit through four layers of planarized copper interconnect, down to the polysilicon (pink), wells (greyish), and substrate (green)

Types

Integrated circuits can be broadly classified into analog,^[79] digital^[80] and mixed signal,^[81] consisting of analog and digital signaling on the same IC.

Digital integrated circuits can contain billions^[55] of logic gates, flip-flops, multiplexers, and other circuits in a few square millimeters. The small size of these circuits allows high speed, low power dissipation, and reduced manufacturing cost compared with board-level integration. These digital ICs, typically microprocessors, DSPs, and microcontrollers, use boolean algebra to process "one" and "zero" signals.

Among the most advanced integrated circuits are the microprocessors or "cores", used in personal computers, cell-phones, etc. Several cores may be integrated together in a single IC or chip. Digital memory chips and application-specific integrated circuits (ASICs) are examples of other families of

integrated circuits.

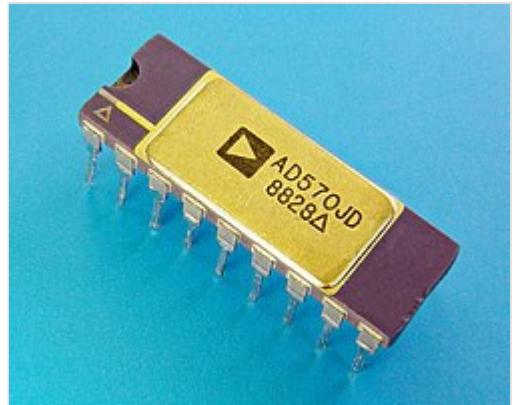
In the 1980s, programmable logic devices were developed. These devices contain circuits whose logical function and connectivity can be programmed by the user, rather than being fixed by the integrated circuit manufacturer. This allows a chip to be programmed to do various LSI-type functions such as logic gates, adders and registers. Programmability comes in various forms – devices that can be programmed only once, devices that can be erased and then re-programmed using UV light, devices that can be (re)programmed using flash memory, and field-programmable gate arrays (FPGAs) which can be programmed at any time, including during operation. Current FPGAs can (as of 2016) implement the equivalent of millions of gates and operate at frequencies up to 1 GHz.^[82]

Analog ICs, such as sensors, power management circuits, and operational amplifiers (op-amps), process continuous signals, and perform analog functions such as amplification, active filtering, demodulation, and mixing.

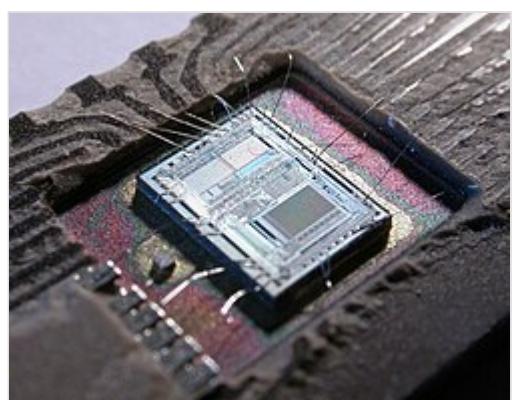
ICs can combine analog and digital circuits on a chip to create functions such as analog-to-digital converters and digital-to-analog converters. Such mixed-signal circuits offer smaller size and lower cost, but must account for signal interference. Prior to the late 1990s, radios could not be fabricated in the same low-cost CMOS processes as microprocessors. But since 1998, radio chips have been developed using RF CMOS processes. Examples include Intel's DECT cordless phone, or 802.11 (Wi-Fi) chips created by Atheros and other companies.^[83]

Modern electronic component distributors often further sub-categorize integrated circuits:

- Digital ICs are categorized as logic ICs (such as microprocessors and microcontrollers), memory chips (such as MOS memory and floating-gate memory), interface ICs (level shifters, serializer/deserializer, etc.), power management ICs, and programmable devices.
- Analog ICs are categorized as linear integrated circuits and RF circuits (radio frequency circuits).
- Mixed-signal integrated circuits are categorized as data acquisition ICs (including A/D converters, D/A converters, digital potentiometers), clock/timing ICs, switched capacitor (SC) circuits, and RF CMOS circuits.
- Three-dimensional integrated circuits (3D ICs) are categorized into through-silicon via (TSV) ICs and Cu-Cu connection ICs.



A-to-D converter IC in a DIP



The die from an Intel 8742, an 8-bit NMOS microcontroller that includes a CPU running at 12 MHz, 128 bytes of RAM, 2048 bytes of EPROM, and I/O in the same chip

Manufacturing

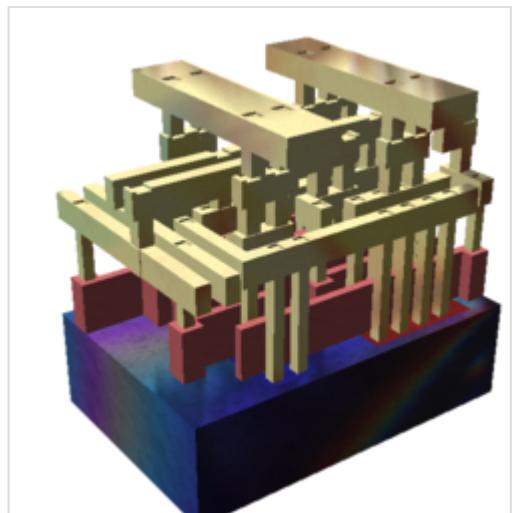
Fabrication

The semiconductors of the periodic table of the chemical elements were identified as the most likely materials for a solid-state vacuum tube. Starting with copper oxide, proceeding to germanium, then silicon, the materials were systematically studied in the 1940s and 1950s. Today, monocrystalline silicon is the main substrate used for ICs although some III-V compounds of the periodic table such as gallium arsenide are used for specialized applications like LEDs, lasers, solar cells and the highest-speed integrated circuits. It took decades to perfect methods of creating crystals with minimal defects in semiconducting materials' crystal structure.

Semiconductor ICs are fabricated in a planar process which includes three key process steps – photolithography, deposition (such as chemical vapor deposition), and etching. The main process steps are supplemented by doping and cleaning. More recent or high-performance ICs may instead use multi-gate FinFET or GAAFET transistors instead of planar ones, starting at the 22 nm node (Intel) or 16/14 nm nodes.^[84]

Mono-crystal silicon wafers are used in most applications (or for special applications, other semiconductors such as gallium arsenide are used). The wafer need not be entirely silicon. Photolithography is used to mark different areas of the substrate to be doped or to have polysilicon, insulators or metal (typically aluminium or copper) tracks deposited on them. Dopants are impurities intentionally introduced to a semiconductor to modulate its electronic properties. Doping is the process of adding dopants to a semiconductor material.

- Integrated circuits are composed of many overlapping layers, each defined by photolithography, and normally shown in different colors. Some layers mark where various dopants are diffused into the substrate (called diffusion layers), some define where additional ions are implanted (implant layers), some define the conductors (doped polysilicon or metal layers), and some define the connections between the conducting layers (via or contact layers). All components are constructed from a specific combination of these layers.
- In a self-aligned CMOS process, a transistor is formed wherever the gate layer (polysilicon or metal) crosses a diffusion layer (this is called "the self-aligned gate").^{[85]:p.1} (see Fig. 1.1)
- Capacitive structures, in form very much like the parallel conducting plates of a traditional electrical capacitor, are formed according to the area of the "plates", with insulating material between the plates. Capacitors of a wide range of sizes are common on ICs.
- Meandering stripes of varying lengths are sometimes used to form on-chip resistors, though most logic circuits do not need any resistors. The ratio of the length of the resistive structure



Rendering of a small standard cell with three metal layers (dielectric has been removed). The sand-colored structures are metal interconnect, with the vertical pillars being contacts, typically plugs of tungsten. The reddish structures are polysilicon gates, and the solid at the bottom is the crystalline silicon bulk.

to its width, combined with its sheet resistivity, determines the resistance.

- More rarely, inductive structures can be built as tiny on-chip coils, or simulated by gyrators.

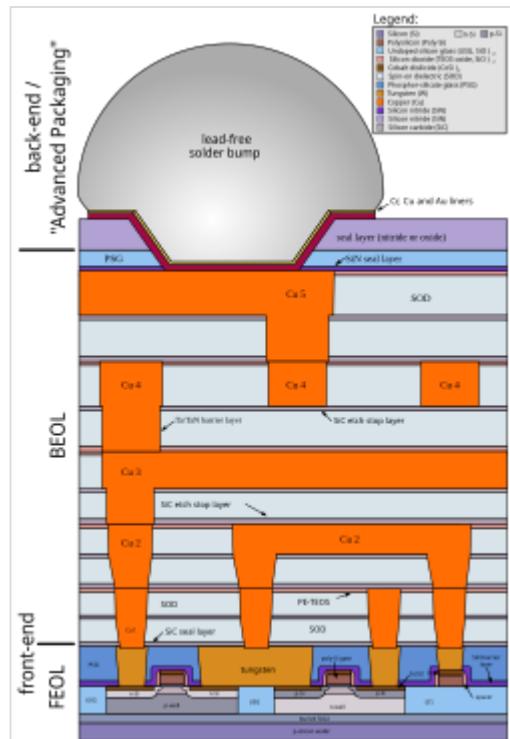
Since a CMOS device only draws current on the transition between logic states, CMOS devices consume much less current than bipolar junction transistor devices.

A random-access memory is the most regular type of integrated circuit; the highest density devices are thus memories; but even a microprocessor will have memory on the chip. (See the regular array structure at the bottom of the first image.) Although the structures are intricate – with widths which have been shrinking for decades – the layers remain much thinner than the device widths. The layers of material are fabricated much like a photographic process, although light waves in the visible spectrum cannot be used to "expose" a layer of material, as they would be too large for the features. Thus photons of higher frequencies (typically ultraviolet) are used to create the patterns for each layer. Because each feature is so small, electron microscopes are essential tools for a process engineer who might be debugging a fabrication process.

Each device is tested before packaging using automated test equipment (ATE), in a process known as wafer testing, or wafer probing. The wafer is then cut into rectangular blocks, each of which is called a die. Each good die (plural *dice*, *dies*, or *die*) is then connected into a package using aluminium (or gold) bond wires which are thermosonically bonded^[86] to pads, usually found around the edge of the die. Thermosonic bonding was first introduced by A. Coucoulas which provided a reliable means of forming these vital electrical connections to the outside world. After packaging, the devices go through final testing on the same or similar ATE used during wafer probing. Industrial CT scanning can also be used. Test cost can account for over 25% of the cost of fabrication on lower-cost products, but can be negligible on low-yielding, larger, or higher-cost devices.

As of 2022, a fabrication facility (commonly known as a *semiconductor fab*) can cost over US\$12 billion to construct.^[87] The cost of a fabrication facility rises over time because of increased complexity of new products; this is known as Rock's law. Such a facility features:

- The wafers up to 300 mm in diameter (wider than a common dinner plate).
- As of 2022, 5 nm transistors.
- Copper interconnects where copper wiring replaces aluminum for interconnects.
- Low-k dielectric insulators.
- Silicon on insulator (SOI).
- Strained silicon in a process used by IBM known as Strained silicon directly on insulator (SSDOI).
- Multigate devices such as tri-gate transistors.



Schematic structure of a CMOS chip, as built in the early 2000s. The graphic shows LDD-MISFET's on an SOI substrate with five metallization layers and solder bump for flip-chip bonding. It also shows the section for FEOL (front-end of line), BEOL (back-end of line) and first parts of back-end process.

ICs can be manufactured either in-house by integrated device manufacturers (IDMs) or using the foundry model. IDMs are vertically integrated companies (like Intel and Samsung) that design, manufacture and sell their own ICs, and may offer design and/or manufacturing (foundry) services to other companies (the latter often to fabless companies). In the foundry model, fabless companies (like Nvidia) only design and sell ICs and outsource all manufacturing to pure play foundries such as TSMC. These foundries may offer IC design services.

Packaging

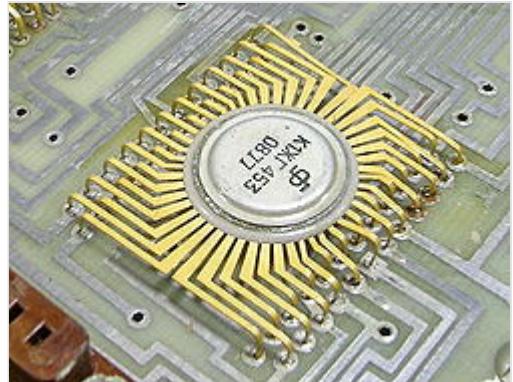
The earliest integrated circuits were packaged in ceramic flat packs, which continued to be used by the military for their reliability and small size for many years. Commercial circuit packaging quickly moved to the dual in-line package (DIP), first in ceramic and later in plastic, which is commonly cresol-formaldehyde-novolac. In the 1980s pin counts of VLSI circuits exceeded the practical limit for DIP packaging, leading to pin grid array (PGA) and leadless chip carrier (LCC) packages. Surface mount packaging appeared in the early 1980s and became popular in the late 1980s, using finer lead pitch with leads formed as either gull-wing or J-lead, as exemplified by the small-outline integrated circuit (SOIC) package – a carrier which occupies an area about 30–50% less than an equivalent DIP and is typically 70% thinner. This package has "gull wing" leads protruding from the two long sides and a lead spacing of 0.050 inches.

In the late 1990s, plastic quad flat pack (PQFP) and thin small-outline package (TSOP) packages became the most common for high pin count devices, though PGA packages are still used for high-end microprocessors.

Ball grid array (BGA) packages have existed since the 1970s. Flip-chip Ball Grid Array packages, which allow for a much higher pin count than other package types, were developed in the 1990s. In an FCBGA package, the die is mounted upside-down (flipped) and connects to the package balls via a package substrate that is similar to a printed-circuit board rather than by wires. FCBGA packages allow an array of input-output signals (called Area-I/O) to be distributed over the entire die rather than being confined to the die periphery. BGA devices have the advantage of not needing a dedicated socket but are much harder to replace in case of device failure.

Intel transitioned away from PGA to land grid array (LGA) and BGA beginning in 2004, with the last PGA socket released in 2014 for mobile platforms. As of 2018, AMD uses PGA packages on mainstream desktop processors,^[89] BGA packages on mobile processors,^[90] and high-end desktop and server microprocessors use LGA packages.^[91]

Electrical signals leaving the die must pass through the material electrically connecting the die to the package, through the conductive traces (paths) in the package, through the leads connecting the package to the conductive traces on the printed circuit board. The materials and structures used in the path these electrical signals must travel have very different electrical properties, compared to those that travel to different parts of the same die. As a result, they require special design techniques to ensure the signals are not corrupted, and much more electric power than signals confined to the die itself.



A Soviet MSI nMOS chip made in 1977, part of a four-chip calculator set designed in 1970^[88]

When multiple dies are put in one package, the result is a system in package, abbreviated SiP. A multi-chip module (MCM), is created by combining multiple dies on a small substrate often made of ceramic. The distinction between a large MCM and a small printed circuit board is sometimes fuzzy.

Packaged integrated circuits are usually large enough to include identifying information. Four common sections are the manufacturer's name or logo, the part number, a part production batch number and serial number, and a four-digit date-code to identify when the chip was manufactured. Extremely small surface-mount technology parts often bear only a number used in a manufacturer's lookup table to find the integrated circuit's characteristics.

The manufacturing date is commonly represented as a two-digit year followed by a two-digit week code, such that a part bearing the code 8341 was manufactured in week 41 of 1983, or approximately in October 1983.

Intellectual property

The possibility of copying by photographing each layer of an integrated circuit and preparing photomasks for its production on the basis of the photographs obtained is a reason for the introduction of legislation for the protection of layout designs. The US Semiconductor Chip Protection Act of 1984 established intellectual property protection for photomasks used to produce integrated circuits.^[92]

A diplomatic conference held at Washington, D.C., in 1989 adopted a Treaty on Intellectual Property in Respect of Integrated Circuits,^[93] also called the Washington Treaty or IPIC Treaty. The treaty is currently not in force, but was partially integrated into the TRIPS agreement.^[94]

There are several United States patents connected to the integrated circuit, which include patents by J.S. Kilby US3,138,743 (<https://patents.google.com/patent/US3138743>), US3,261,081 (<https://patents.google.com/patent/US3261081>), US3,434,015 (<https://patents.google.com/patent/US3434015>) and by R.F. Stewart US3,138,747 (<https://patents.google.com/patent/US3138747>).

National laws protecting IC layout designs have been adopted in a number of countries, including Japan,^[95] the EC,^[96] the UK, Australia, and Korea. The UK enacted the Copyright, Designs and Patents Act, 1988, c. 48, § 213, after it initially took the position that its copyright law fully protected chip topographies. See British Leyland Motor Corp. v. Armstrong Patents Co.

Criticisms of inadequacy of the UK copyright approach as perceived by the US chip industry are summarized in further chip rights developments.^[97]

Australia passed the Circuit Layouts Act of 1989 as a *sui generis* form of chip protection.^[98] Korea passed the *Act Concerning the Layout-Design of Semiconductor Integrated Circuits* in 1992.^[99]

Generations

In the early days of simple integrated circuits, the technology's large scale limited each chip to only a few transistors, and the low degree of integration meant the design process was relatively simple. Manufacturing yields were also quite low by today's standards. As metal–oxide–semiconductor (MOS)

technology progressed, millions and then billions of MOS transistors could be placed on one chip,^[100] and good designs required thorough planning, giving rise to the field of electronic design automation, or EDA. Some SSI and MSI chips, like discrete transistors, are still mass-produced, both to maintain old equipment and build new devices that require only a few gates. The 7400 series of TTL chips, for example, has become a de facto standard and remains in production.

Acronym	Name	Year	Transistor count ^[101]	Logic gates number ^[102]
SSI	<i>small-scale integration</i>	1964	1 to 10	1 to 12
MSI	<i>medium-scale integration</i>	1968	10 to 500	13 to 99
LSI	<i>large-scale integration</i>	1971	500 to 20 000	100 to 9999
VLSI	<i>very large-scale integration</i>	1980	20 000 to 1 000 000	10 000 to 99 999
ULSI	<i>ultra-large-scale integration</i>	1984	1 000 000 and more	100 000 and more

Small-scale integration (SSI)

The first integrated circuits contained only a few transistors. Early digital circuits containing tens of transistors provided a few logic gates, and early linear ICs such as the Plessey SL201 or the Philips TAA320 had as few as two transistors. The number of transistors in an integrated circuit has increased dramatically since then. The term "large scale integration" (LSI) was first used by IBM scientist Rolf Landauer when describing the theoretical concept;^[103] that term gave rise to the terms "small-scale integration" (SSI), "medium-scale integration" (MSI), "very-large-scale integration" (VLSI), and "ultra-large-scale integration" (ULSI). The early integrated circuits were SSI.

SSI circuits were crucial to early aerospace projects, and aerospace projects helped inspire development of the technology. Both the Minuteman missile and Apollo program needed lightweight digital computers for their inertial guidance systems. Although the Apollo Guidance Computer led and motivated integrated-circuit technology,^[104] it was the Minuteman missile that forced it into mass-production. The Minuteman missile program and various other United States Navy programs accounted for the total \$4 million integrated circuit market in 1962, and by 1968, U.S. Government spending on space and defense still accounted for 37% of the \$312 million total production.

The demand by the U.S. Government supported the nascent integrated circuit market until costs fell enough to allow IC firms to penetrate the industrial market and eventually the consumer market. The average price per integrated circuit dropped from \$50 in 1962 to \$2.33 in 1968.^[105] Integrated circuits began to appear in consumer products by the turn of the 1970s decade. A typical application was FM inter-carrier sound processing in television receivers.

The first application MOS chips were small-scale integration (SSI) chips.^[106] Following Mohamed M. Atalla's proposal of the MOS integrated circuit chip in 1960,^[107] the earliest experimental MOS chip to be fabricated was a 16-transistor chip built by Fred Heiman and Steven Hofstein at RCA in 1962.^[46] The first practical application of MOS SSI chips was for NASA satellites.^[106]

Medium-scale integration (MSI)

The next step in the development of integrated circuits introduced devices which contained hundreds of transistors on each chip, called "medium-scale integration" (MSI).

MOSFET scaling technology made it possible to build high-density chips.^[41] By 1964, MOS chips had reached higher transistor density and lower manufacturing costs than bipolar chips.^[48]

In 1964, Frank Wanlass demonstrated a single-chip 16-bit shift register he designed, with a then-incredible 120 MOS transistors on a single chip.^{[106][108]} The same year, General Microelectronics introduced the first commercial MOS integrated circuit chip, consisting of 120 p-channel MOS transistors.^[47] It was a 20-bit shift register, developed by Robert Norman^[46] and Frank Wanlass.^{[109][110]} MOS chips further increased in complexity at a rate predicted by Moore's law, leading to chips with hundreds of MOSFETs on a chip by the late 1960s.^[48]

Large-scale integration (LSI)

Further development, driven by the same MOSFET scaling technology and economic factors, led to "large-scale integration" (LSI) by the mid-1970s, with tens of thousands of transistors per chip.^[111]

The masks used to process and manufacture SSI, MSI and early LSI and VLSI devices (such as the microprocessors of the early 1970s) were mostly created by hand, often using Rubylith-tape or similar.^[112] For large or complex ICs (such as memories or processors), this was often done by specially hired professionals in charge of circuit layout, placed under the supervision of a team of engineers, who would also, along with the circuit designers, inspect and verify the correctness and completeness of each mask.

Integrated circuits such as 1K-bit RAMs, calculator chips, and the first microprocessors, that began to be manufactured in moderate quantities in the early 1970s, had under 4,000 transistors. True LSI circuits, approaching 10,000 transistors, began to be produced around 1974, for computer main memories and second-generation microprocessors.

Very-large-scale integration (VLSI)

"Very-large-scale integration" (VLSI) is a development that started with hundreds of thousands of transistors in the early 1980s. As of 2023, maximum transistor counts continue to grow beyond 5.3 trillion transistors per chip.

Multiple developments were required to achieve this increased density. Manufacturers moved to smaller MOSFET design rules and cleaner fabrication facilities. The path of process improvements was summarized by the International Technology Roadmap for Semiconductors (ITRS), which has since been succeeded by the International Roadmap for Devices and Systems (IRDS). Electronic design tools improved, making it practical to finish designs in a reasonable time. The more energy-efficient CMOS replaced NMOS and PMOS, avoiding a prohibitive increase in power consumption. The complexity and density of modern VLSI devices made it no longer feasible to check the masks or do the original design by hand. Instead, engineers use EDA tools to perform most functional verification work.^[113]

In 1986, one-megabit random-access memory (RAM) chips were introduced, containing more than one million transistors. Microprocessor chips passed the million-transistor mark in 1989, and the billion-transistor mark in 2005.^[114] The trend continues largely unabated, with chips introduced in 2007 containing tens of billions of memory transistors.^[115]

ULSI, WSI, SoC and 3D-IC

To reflect further growth of the complexity, the term *ULSI* that stands for "ultra-large-scale integration" was proposed for chips of more than 1 million transistors.^[116]

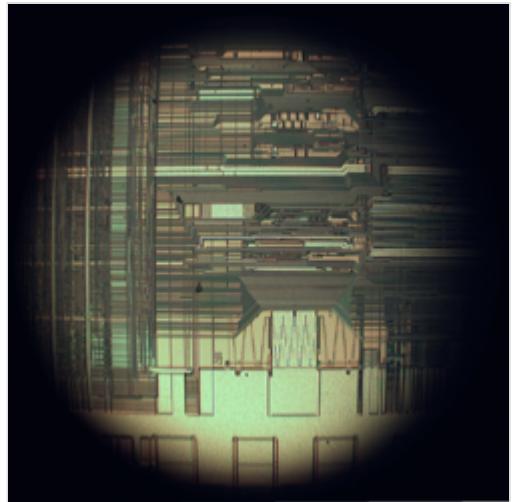
Wafer-scale integration (WSI) is a means of building very large integrated circuits that uses an entire silicon wafer to produce a single "super-chip". Through a combination of large size and reduced packaging, WSI could lead to dramatically reduced costs for some systems, notably massively parallel supercomputers. The name is taken from the term Very-Large-Scale Integration, the current state of the art when WSI was being developed.^{[117][118]}

A system-on-a-chip (SoC or SOC) is an integrated circuit in which all the components needed for a computer or other system are included on a single chip. The design of such a device can be complex and costly, and whilst performance benefits can be had from integrating all needed components on one die, the cost of licensing and developing a one-die machine still outweigh having separate devices. With appropriate licensing, these drawbacks are offset by lower manufacturing and assembly costs and by a greatly reduced power budget: because signals among the components are kept on-die, much less power is required (see Packaging).^[119] Further, signal sources and destinations are physically closer on die, reducing the length of wiring and therefore latency, transmission power costs and waste heat from communication between modules on the same chip. This has led to an exploration of so-called Network-on-Chip (NoC) devices, which apply system-on-chip design methodologies to digital communication networks as opposed to traditional bus architectures.

A three-dimensional integrated circuit (3D-IC) has two or more layers of active electronic components that are integrated both vertically and horizontally into a single circuit. Communication between layers uses on-die signaling, so power consumption is much lower than in equivalent separate circuits. Judicious use of short vertical wires can substantially reduce overall wire length for faster operation.^[120]

Silicon labeling and graffiti

To allow identification during production, most silicon chips will have a serial number in one corner. It is also common to add the manufacturer's logo. Ever since ICs were created, some chip designers have used the silicon surface area for surreptitious, non-functional images or words. These artistic additions, often created with great attention to detail, showcase the designers' creativity and add a touch of personality to otherwise utilitarian components. These are sometimes referred to as chip art, silicon art, silicon graffiti or silicon doodling.^[121]



Upper interconnect layers on an Intel 80486DX2 microprocessor die

ICs and IC families

- The [555 timer IC](#)
- The [Operational amplifier](#)
- [7400-series integrated circuits](#)
- [4000-series integrated circuits](#), the CMOS counterpart to the 7400 series (see also: [74HC00 series](#))
- Intel 4004, generally regarded as the first commercially available [microprocessor](#), which led to the 8008, the famous [8080 CPU](#), the [8086](#), [8088](#) (used in the original [IBM PC](#)), and the fully-backward compatible (with the 8088/8086) [80286](#), [80386/i386](#), [i486](#), etc.
- The [MOS Technology 6502](#) and [Zilog Z80](#) microprocessors, used in many [home computers](#) of the early 1980s
- The [Motorola 6800](#) series of computer-related chips, leading to the [68000](#) and [88000](#) series (the 68000 series was very successful and was used in the Apple Lisa and pre-PowerPC-based Macintosh, Commodore [Amiga](#), Atari ST/TT/Falcon030, and NeXT families of computers, along with many models of workstations and servers from many manufacturers in the 80s, along with many other systems and devices)
- The [LM-series](#) of analog integrated circuits

See also

 Electronics portal
 Physics portal
 Technology portal
 Telecommunication portal
 Engineering portal
 History of science portal
 Companies portal
 Computer programming portal
 Telephones portal

- [Central processing unit](#)
- [Chip carrier](#)
- [CHIPS and Science Act](#)
- [Chipset](#)
- [Czochralski method](#)
- [Dark silicon](#)
- [Ion implantation](#)
- [Integrated injection logic](#)
- [Integrated passive devices](#)

- [Interconnect bottleneck](#)
- [Heat generation in integrated circuits](#)
- [High-temperature operating life](#)
- [Microelectronics](#)
- [Monolithic microwave integrated circuit](#)
- [Multi-threshold CMOS](#)
- [Silicon–germanium](#)
- [Sound chip](#)
- [SPICE](#)
- [Thermal simulations for integrated circuits](#)
- [Hybrot](#)

References

1. "The basics of microchips" (<https://www.asml.com/en/technology/all-about-microchips/micro-chip-basics>). ASML.
2. "Integrated circuit (IC)" (<http://www.jedec.org/standards-documents/dictionary/terms/integrated-circuit-ic>). JEDEC.
3. Wylie, Andrew (2009). "The first monolithic integrated circuits" (<https://web.archive.org/web/20180504074623/http://homepages.nildram.co.uk/~wylie/ICs/monolith.htm>). Archived from the original (<http://homepages.nildram.co.uk/~wylie/ICs/monolith.htm>) on 4 May 2018. Retrieved 14 March 2011. "Nowadays when people say 'integrated circuit' they usually mean a monolithic IC, where the entire circuit is constructed in a single piece of silicon."
4. Horowitz, Paul; Hill, Winfield (1989). *The Art of Electronics* (<https://archive.org/details/artofelectronics00horo/page/61>) (2nd ed.). Cambridge University Press. p. 61 (<https://archive.org/details/artofelectronics00horo/page/61>). ISBN 978-0-521-37095-0. "Integrated circuits, which have largely replaced circuits constructed from discrete transistors, are themselves merely arrays of transistors and other components built from a single chip of semiconductor material."
5. Mike Harrison, *Mike's Electric Stuff* (1998-2014) Loewe 3NF Multi-valve - The first integrated circuit. (<http://www.electricstuff.co.uk/loewe.html>)
6. "Loewe 3NF radio valve, 1926-1966 | Science Museum Group Collection" (<https://collection.science museumgroup.org.uk/objects/co8722786/loewe-3nf-radio-valve-1926-1966-thermion-ic-valve>).
7. Anton Pankratov (20 Nov 2010) Loewe 3NF Integrated Lamp (<https://www.youtube.com/watch?v=1E8StV497sQ>)
8. Clive Maxfield, EE Times (04.05.2006) The first "integrated circuit" from 1926! (<https://www.eetimes.com/the-first-integrated-circuit-from-1926/>)
9. Malanowski, Gregory (2011). *The Race for Wireless: How Radio was Invented (Or Discovered?)* (<https://books.google.com/books?id=IAjtEeVtXqAC&dq=Loewe+3NF&pg=PA1-05>). AuthorHouse. ISBN 978-1-4634-3750-3.
10. *Principles of VLSI and CMOS Integrated Circuits* (<https://books.google.com/books?id=UjJIDwAAQBAJ&dq=discrete+vs+integrated+circuit+cost&pg=SA1-PA6>). S. Chand. 2016. ISBN 978-81-219-4000-9.
11. "Who Invented the IC?" (<https://www.computerhistory.org/atchm/who-invented-the-ic/>). @CHM Blog. Computer History Museum. 20 August 2014.
12. "Integrated circuits help Invention" (<http://integratedcircuiithelp.com/invention.html>). Integratedcircuiithelp.com. Retrieved 13 August 2012.

13. DE patent 833366 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=DE833366>), W. Jacobi, "Halbleiterverstärker\"", published 15 May 1952, assigned to SIEMENS AG
14. "The Hapless Tale of Geoffrey Dummer" (<https://web.archive.org/web/20110726112126/http://www.epn-online.com/page/22909/the-hapless-tale-of-geoffrey-dummer-this-is-the-sad-.html>). *epn-online.com*. 1 October 2005. Archived from the original (<http://www.epn-online.com/page/22909/the-hapless-tale-of-geoffrey-dummer-this-is-the-sad-.html>) on 26 July 2011.
15. US2802760A (<https://patents.google.com/patent/US2802760A>), Lincoln, Derick & Frosch, Carl J., "Oxidation of semiconductive surfaces for controlled diffusion", issued 13 August 1957
16. Frosch, C. J.; Derick, L (1957). "Surface Protection and Selective Masking during Diffusion in Silicon" (<https://iopscience.iop.org/article/10.1149/1.2428650>). *Journal of the Electrochemical Society*. **104** (9): 547. doi:10.1149/1.2428650 (<https://doi.org/10.1149%2F1.2428650>).
17. Lojek, Bo (2007). *History of Semiconductor Engineering*. Springer Science & Business Media. p. 120. ISBN 9783540342588.
18. Fuller, C. S.; Ditzenberger, J. A. (1 July 1953). "Diffusion of Lithium into Germanium and Silicon" (<https://link.aps.org/doi/10.1103/PhysRev.91.193>). *Physical Review*. **91** (1): 193. Bibcode:1953PhRv...91..193F (<https://ui.adsabs.harvard.edu/abs/1953PhRv...91..193F>). doi:10.1103/PhysRev.91.193 (<https://doi.org/10.1103%2FPhysRev.91.193>). ISSN 0031-899X (<https://search.worldcat.org/issn/0031-899X>).
19. Fuller, C. S.; Struthers, J. D.; Ditzenberger, J. A.; Wolfstirn, K. B. (15 March 1954). "Diffusivity and Solubility of Copper in Germanium" (<https://link.aps.org/doi/10.1103/PhysRev.93.1182>). *Physical Review*. **93** (6): 1182–1189. Bibcode:1954PhRv...93.1182F (<https://ui.adsabs.harvard.edu/abs/1954PhRv...93.1182F>). doi:10.1103/PhysRev.93.1182 (<https://doi.org/10.1103%2FPhysRev.93.1182>). ISSN 0031-899X (<https://search.worldcat.org/issn/0031-899X>).
20. Fuller, C. S.; Ditzenberger, J. A. (1 May 1956). "Diffusion of Donor and Acceptor Elements in Silicon" (<https://pubs.aip.org/jap/article/27/5/544/161241/Diffusion-of-Donor-and-Acceptor-Elements-in>). *Journal of Applied Physics*. **27** (5): 544–553. Bibcode:1956JAP....27..544F (<https://ui.adsabs.harvard.edu/abs/1956JAP....27..544F>). doi:10.1063/1.1722419 (<https://doi.org/10.1063%2F1.1722419>). ISSN 0021-8979 (<https://search.worldcat.org/issn/0021-8979>).
21. Fuller, C. S.; Whelan, J. M. (1 August 1958). "Diffusion, solubility, and electrical behavior of copper in gallium arsenide" (<https://www.sciencedirect.com/science/article/abs/pii/002236975890091X>). *Journal of Physics and Chemistry of Solids*. **6** (2): 173–177. Bibcode:1958JPCS....6..173F (<https://ui.adsabs.harvard.edu/abs/1958JPCS....6..173F>). doi:10.1016/0022-3697(58)90091-X (<https://doi.org/10.1016%2F0022-3697%2858%2990091-X>). ISSN 0022-3697 (<https://search.worldcat.org/issn/0022-3697>).
22. Miller, R. C.; Savage, A. (1 December 1956). "Diffusion of Aluminum in Single Crystal Silicon" (<https://pubs.aip.org/jap/article/27/12/1430/161591/Diffusion-of-Aluminum-in-Single-Crystal-Silicon>). *Journal of Applied Physics*. **27** (12): 1430–1432. Bibcode:1956JAP....27.1430M (<https://ui.adsabs.harvard.edu/abs/1956JAP....27.1430M>). doi:10.1063/1.1722283 (<https://doi.org/10.1063%2F1.1722283>). ISSN 0021-8979 (<https://search.worldcat.org/issn/0021-8979>).
23. Rostky, George. "Micromodules: the ultimate package" (https://web.archive.org/web/20100107111717/http://www.eetimes.com/special/special_issues/millennium/milestones/kilby.html). *EE Times*. Archived from the original (http://www.eetimes.com/special/special_issues/millennium/milestones/kilby.html) on 7 January 2010. Retrieved 23 April 2018.
24. "The RCA Micromodule" (<http://www.chipsetc.com/the-rca-micromodule.html>). *Vintage Computer Chip Collectibles, Memorabilia & Jewelry*. Retrieved 23 April 2018.
25. Dummer, G.W.A.; Robertson, J. Mackenzie (16 May 2014). *American Microelectronics Data Annual 1964–65* (<https://books.google.com/books?id=tdCjBQAAQBAJ&q=micromodule&pg=PA392>). Elsevier. pp. 392–397, 405–406. ISBN 978-1-4831-8549-1.

26. "The Chip That Jack Built Changed the World" (<https://web.archive.org/web/20000418135808/http://www.ti.com/corp/docs/kilbyctr/jackbuilt.shtml>). *ti.com*. 9 September 1997. Archived from the original on 18 April 2000.
27. US Patent 3138743 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=US3138743>), Kilby, Jack S., "Miniaturized Electronic Circuits", published 23 June 1964
28. Winston, Brian (1998). *Media Technology and Society: A History: From the Telegraph to the Internet* (<https://books.google.com/books?id=gfeCXIEIJTwC&pg=PA221>). Routledge. p. 221. ISBN 978-0-415-14230-4.
29. "Texas Instruments – 1961 First IC-based computer" (http://www.ti.com/corp/docs/company/history/timeline/defense/1960/docs/61-first_ic.htm). *Ti.com*. Retrieved 13 August 2012.
30. "The Nobel Prize in Physics 2000" (<https://www.nobelprize.org/prizes/physics/2000/press-release/>). *NobelPrize.org*. 10 October 2000.
31. "Integrated circuits" (<https://www.hq.nasa.gov/alsj/ic-pg3.html>). *NASA*. Retrieved 13 August 2019.
32. "1959: Practical Monolithic Integrated Circuit Concept Patented" (<https://www.computerhistory.org/siliconengine/practical-monolithic-integrated-circuit-concept-patented/>). *Computer History Museum*. Retrieved 13 August 2019.
33. Hall, Eldon C. (1996). *Journey to the Moon: The History of the Apollo Guidance Computer* (<https://books.google.com/books?id=G8Dml1x55r0C>). Library of Flight. American Institute of Aeronautics and Astronautics. pp. 18–19. ISBN 978-1-56347-185-8. Retrieved 5 October 2023.
34. "Computer Pioneers – James L. Buie" (<https://history.computer.org/pioneers/buie.html>). *IEEE Computer Society*. Retrieved 25 May 2020.
35. "The Texas Instruments TMX 1795: the (almost) first, forgotten microprocessor" (<https://www.righto.com/2015/05/the-texas-instruments-tmx-1795-first.html>). *Ken Shirriff's blog*. 25 October 1970.
36. Kuo, Yue (1 January 2013). "Thin Film Transistor Technology—Past, Present, and Future" (https://www.electrochem.org/dl/interface/spr/spr13/spr13_p055_061.pdf) (PDF). *The Electrochemical Society Interface*. **22** (1): 55–61. Bibcode:2013ECSIn..22a..55K (<https://ui.adsabs.harvard.edu/abs/2013ECSIn..22a..55K>). doi:10.1149/2.F06131if (<https://doi.org/10.1149/2.F06131if>).
37. Huff, Howard; Riordan, Michael (1 September 2007). "Frosch and Derick: Fifty Years Later (Foreword)" (<https://iopscience.iop.org/article/10.1149/2.F02073IF>). *The Electrochemical Society Interface*. **16** (3): 29. doi:10.1149/2.F02073IF (<https://doi.org/10.1149/2.F02073IF>). ISSN 1064-8208 (<https://search.worldcat.org/issn/1064-8208>).
38. KAHNG, D. (1961). "Silicon-Silicon Dioxide Surface Device" (https://doi.org/10.1142/9789814503464_0076). *Technical Memorandum of Bell Laboratories*: 583–596. doi:10.1142/9789814503464_0076 (https://doi.org/10.1142/9789814503464_0076). ISBN 978-981-02-0209-5.
39. Lojek, Bo (2007). *History of Semiconductor Engineering*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. p. 321. ISBN 978-3-540-34258-8.
40. Ligenza, J.R.; Spitzer, W.G. (1960). "The mechanisms for silicon oxidation in steam and oxygen" (<https://linkinghub.elsevier.com/retrieve/pii/0022369760902195>). *Journal of Physics and Chemistry of Solids*. **14**: 131–136. Bibcode:1960JPCS...14..131L (<https://ui.adsabs.harvard.edu/abs/1960JPCS...14..131L>). doi:10.1016/0022-3697(60)90219-5 ([https://doi.org/10.1016/0022-3697\(60\)90219-5](https://doi.org/10.1016/0022-3697(60)90219-5)).
41. Laws, David (4 December 2013). "Who Invented the Transistor?" (<https://www.computerhistory.org/atchm/who-invented-the-transistor/>). *Computer History Museum*.
42. Bassett, Ross Knox (2002). *To the Digital Age: Research Labs, Start-up Companies, and the Rise of MOS Technology* (<https://books.google.com/books?id=Qge1DUt7qDUC&pg=PA53>). Johns Hopkins University Press. pp. 53–4. ISBN 978-0-8018-6809-2.

43. Bassett, Ross Knox (2007). *To the Digital Age: Research Labs, Start-up Companies, and the Rise of MOS Technology* (<https://books.google.com/books?id=UUbB3d2UnaAC&pg=PA22>). Johns Hopkins University Press. pp. 22–25. ISBN 9780801886393.
44. "Milestones:First Semiconductor Integrated Circuit (IC), 1958" (http://www.ieeeghn.org/wiki/index.php/Milestones:First_Semiconductor_Integrated_Circuit_%28IC%29,_1958). IEEE Global History Network. IEEE. Retrieved 3 August 2011.
45. "Milestones:List of IEEE Milestones – Engineering and Technology History Wiki" (https://ethw.org/Milestones>List_of_IEEE_Milestones). ethw.org. 9 December 2020.
46. "Tortoise of Transistors Wins the Race – CHM Revolution" (<https://www.computerhistory.org/revolution/digital-logic/12/279>). Computer History Museum. Retrieved 22 July 2019.
47. "1964 – First Commercial MOS IC Introduced" (<http://www.computerhistory.org/semiconductor/timeline/1964-Commcial.html>). Computer History Museum.
48. Shirriff, Ken (30 August 2016). "The Surprising Story of the First Microprocessors" (<https://spectrum.ieee.org/the-surprising-story-of-the-first-microprocessors>). IEEE Spectrum. 53 (9). Institute of Electrical and Electronics Engineers: 48–54. doi:10.1109/MSPEC.2016.7551353 (<https://doi.org/10.1109%2FMSPEC.2016.7551353>). S2CID 32003640 (<https://api.semanticscholar.org/CorpusID:32003640>).
49. "1968: Silicon Gate Technology Developed for ICs" (<https://www.computerhistory.org/silicon-engine/silicon-gate-technology-developed-for-ics/>). Computer History Museum. Retrieved 22 July 2019.
50. "1968: Silicon Gate Technology Developed for ICs" (<https://www.computerhistory.org/silicon-engine/silicon-gate-technology-developed-for-ics/>). The Silicon Engine. Computer History Museum. Retrieved 13 October 2019.
51. Hittinger, William C. (1973). "Metal–Oxide–Semiconductor Technology". *Scientific American*. 229 (2): 48–59. Bibcode:1973SciAm.229b..48H (<https://ui.adsabs.harvard.edu/abs/1973SciAm.229b..48H>). doi:10.1038/scientificamerican0873-48 (<https://doi.org/10.1038%2Fscientificamerican0873-48>). JSTOR 24923169 (<https://www.jstor.org/stable/24923169>).
52. Kanellos, Michael (11 February 2003). "Moore's Law to roll on for another decade" (<https://www.cnet.com/news/moores-law-to-roll-on-for-another-decade/>). CNET.
53. Davari, Bijan; Dennard, Robert H.; Shahidi, Ghavam G. (1995). "CMOS scaling for high performance and low power—the next ten years" (http://www.cisl.columbia.edu/courses/spring-2002/ee6930/papers/high_perform_scaling.pdf) (PDF). *Proceedings of the IEEE*. Vol. 83, no. 4. pp. 595–606.
54. "Qualcomm and Samsung Collaborate on 10nm Process Technology for the Latest Snapdragon 835 Mobile Processor" (<https://news.samsung.com/global/qualcomm-and-samsung-collaborate-on-10nm-process-technology-for-the-latest-snapdragon-835-mobile-processor>). news.samsung.com. Retrieved 11 February 2017.
55. "Inside Pascal: NVIDIA's Newest Computing Platform" (<https://devblogs.nvidia.com/parallelforall/inside-pascal/>). 5 April 2016.. 15,300,000,000 transistors in 610 mm².
56. "International Roadmap for Devices and Systems" (https://rebootingcomputing.ieee.org/images/files/pdf/rc_irds.pdf) (PDF). IEEE. 2016.
57. *The Nobel Prize in Physics 2009* (http://nobelprize.org/nobel_prizes/physics/lauriates/2009/index.html), Nobel Foundation, 6 October 2009, retrieved 6 October 2009.
58. Fujita, H. (1997). *A decade of MEMS and its future*. Tenth Annual International Workshop on Micro Electro Mechanical Systems. doi:10.1109/MEMSYS.1997.581729 (<https://doi.org/10.1109%2FMEMSYS.1997.581729>).
59. Narasimha, A.; et al. (2008). "A 40-Gb/s QSFP optoelectronic transceiver in a 0.13 μm CMOS silicon-on-insulator technology" (<http://www.opticsinfobase.org/abstract.cfm?URI=OFC-2008-OMK7>). *Proceedings of the Optical Fiber Communication Conference (OFC)*: OMK7.

60. "Optical chipmaker focuses on high-performance computing" (<https://physicsworld.com/a/optical-chipmaker-focuses-on-high-performance-computing/>). 7 April 2022.
61. Birkholz, M.; Mai, A.; Wenger, C.; Meliani, C.; Scholz, R. (2016). "Technology modules from micro- and nano-electronics for the life sciences" (<https://www.researchgate.net/publication/282052331>). *WIREs Nanomed. Nanobiotech.* **8** (3): 355–377. doi:10.1002/wnan.1367 (<http://doi.org/10.1002%2Fwnan.1367>). PMID 26391194 (<https://pubmed.ncbi.nlm.nih.gov/26391194>).
62. Graham, Anthony H. D.; Robbins, Jon; Bowen, Chris R.; Taylor, John (2011). "Commercialisation of CMOS Integrated Circuit Technology in Multi-Electrode Arrays for Neuroscience and Cell-Based Biosensors" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3231360>). *Sensors.* **11** (5): 4943–4971. Bibcode:2011Senso..11.4943G (<https://ui.adsabs.harvard.edu/abs/2011Senso..11.4943G>). doi:10.3390/s110504943 (<https://doi.org/10.3390%2Fs110504943>). PMC 3231360 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3231360>). PMID 22163884 (<https://pubmed.ncbi.nlm.nih.gov/22163884>).
63. Or-Bach, Zvi (23 December 2013). "Why SOI is the Future Technology of Semiconductors" (<https://web.archive.org/web/20141129104851/http://semimd.com/blog/2013/12/23/why-soi-is-the-future-technology-of-semiconductors/>). *semimd.com*. Archived from the original on 29 November 2014.. 2013.
64. "Samsung's Eight-Stack Flash Shows up in Apple's iPhone 4" (https://sst.semiconductor-digest.com/chipworks_real_chips_blog/2010/09/13/samsungs-eight-stack-flash-shows-up-in-apples-iphone-4/). *Siliconica*. 13 September 2010.
65. Yamatake Corporation (2002). "Spherical semiconductor radio temperature sensor" (<https://web.archive.org/web/20090107144008/http://www.natureinterface.com/e/ni07/P058-059/>). *Nature Interface*. **7**: 58–59. Archived from the original (<http://www.natureinterface.com/e/ni07/P058-059/>) on 7 January 2009.
66. Takeda, Nobuo, *MEMS applications of Ball Semiconductor Technology* (<https://web.archive.org/web/20150101122744/http://asia.stanford.edu/events/spring01/slides/takedaslides.pdf>) (PDF), archived from the original (http://asia.stanford.edu/events/spring01/slides/takedaslide_s.pdf) (PDF) on 1 January 2015
67. "Advanced Packaging" (https://semiengineering.com/knowledge_centers/packaging/advanced-packaging/).
68. "2.5D" (https://semiengineering.com/knowledge_centers/packaging/advanced-packaging/2-5d-ic/). *Semiconductor Engineering*.
69. "3D ICs" (https://semiengineering.com/knowledge_centers/packaging/advanced-packaging/3d-ics/). *Semiconductor Engineering*.
70. "Chiplet" (<https://en.wikichip.org/wiki/chiplet>). *WikiChip*. 28 February 2021.
71. "To Keep Pace With Moore's Law, Chipmakers Turn to 'Chiplets'" (<https://www.wired.com/story/keep-pace-moores-law-chipmakers-turn-chiplets/>). *Wired*. 11 June 2018.
72. Schodt, Christopher (16 April 2019). "This is the year of the CPU 'chiplet'" (<https://www.engadget.com/2019-04-16-upscaled-cpu-chiplet.html>). *Engadget*.
73. "Building Power Electronics with Microscopic Plumbing Could Save Enormous Amounts of Money - IEEE Spectrum" (<https://spectrum.ieee.org/codesigning-electronics-and-microfluidics-for-a-cooling-boost>).
74. "Startup shrinks Peltier cooler, puts it in the chip package" (<https://arstechnica.com/gadgets/2008/01/startup-shrinks-the-peltier-cooler-and-puts-it-inside-the-chip-package/>). 10 January 2008.
75. "Wire Bond Vs. Flip Chip Packaging | Semiconductor Digest" (<https://sst.semiconductor-digest.com/2005/07/wire-bond-vs-flip-chip-packaging/>). 10 December 2016.
76. LaPedus, Mark (16 April 2015). "FinFET Rollout Slower Than Expected" (<http://semiengineering.com/finfet-rollout-slower-than-expected/>). *Semiconductor Engineering*.

77. Basu, Joydeep (9 October 2019). "From Design to Tape-out in SCL 180 nm CMOS Integrated Circuit Fabrication Technology". *IETE Journal of Education*. **60** (2): 51–64. arXiv:1908.10674 (<https://arxiv.org/abs/1908.10674>). doi:10.1080/09747338.2019.1657787 (<https://doi.org/10.1080%2F09747338.2019.1657787>). S2CID 201657819 (<https://api.semanticscholar.org/CorpusID:201657819>).
78. "About the EDA Industry" (<https://web.archive.org/web/20150802073506/http://www.edac.org/industry>). Electronic Design Automation Consortium. Archived from the original (<http://www.edac.org/industry>) on 2 August 2015. Retrieved 29 July 2015.
79. Gray, Paul R.; Hurst, Paul J.; Lewis, Stephen H.; Meyer, Robert G. (2009). *Analysis and Design of Analog Integrated Circuits*. Wiley. ISBN 978-0-470-24599-6.
80. Rabaey, Jan M.; Chandrakasan, Anantha; Nikolic, Borivoje (2003). *Digital Integrated Circuits* (<https://archive.org/details/agilesoftwaredev00robe>) (2nd ed.). Pearson. ISBN 978-0-13-090996-1.
81. Baker, Jacob (2008). *CMOS: Mixed-Signal Circuit Design*. Wiley. ISBN 978-0-470-29026-2.
82. "Stratix 10 Device Overview" (https://www.altera.com/en_US/pdfs/literature/hb/stratix-10/s10-overview.pdf) (PDF). Altera. 12 December 2015.
83. Nathawad, L.; Zargari, M.; Samavati, H.; Mehta, S.; Kheirkhaki, A.; Chen, P.; Gong, K.; Vakili-Amini, B.; Hwang, J.; Chen, M.; Terrovitis, M.; Kaczynski, B.; Limotyrakis, S.; Mack, M.; Gan, H.; Lee, M.; Abdollahi-Alibeik, B.; Baytekin, B.; Onodera, K.; Mendis, S.; Chang, A.; Jen, S.; Su, D.; Wooley, B. "20.2: A Dual-band CMOS MIMO Radio SoC for IEEE 802.11n Wireless LAN" (https://web.archive.org/web/20161023053607/http://www.ewh.ieee.org/r6/scv/ssc/May2008_WLAN.pdf) (PDF). IEEE Entity Web Hosting. IEEE. Archived from the original (http://www.ewh.ieee.org/r6/scv/ssc/May2008_WLAN.pdf) (PDF) on 23 October 2016. Retrieved 22 October 2016.
84. "16nm/14nm FinFETs: Enabling The New Electronics Frontier" (<https://www.electronicdesign.com/technologies/embedded/digital-ics/article/21795644/16nm14nm-finfets-enabling-the-new-electronics-frontier>). electronicdesign.com. 17 January 2013.
85. Mead, Carver; Conway, Lynn (1991). *Introduction to VLSI systems* (<https://archive.org/details/introductiontovl00mead>). Addison Wesley Publishing Company. ISBN 978-0-201-04358-7. OCLC 634332043 (<https://search.worldcat.org/oclc/634332043>).
86. "Hot Work Ultrasonic Bonding – A Method Of Facilitating Metal Flow By Restoration Processes" (<https://sites.google.com/site/hotworkultrasonicbonding/>), Proc. 20th IEEE Electronic Components Conf. Washington, D.C., May 1970, pp. 549–556.]
87. Chafkin (15 May 2020). "TSMC to build 5nm fab in arizona, set to come online in 2024" (<https://www.anandtech.com/show/15803/tsmc-build-5nm-fab-in-arizona-for-2024>). Anandtech.
88. "145 series ICs (in Russian)" (http://www.155la3.ru/k145_3.htm#k145hk1). Retrieved 22 April 2012.
89. Moammer, Khalid (16 September 2016). "AMD Zen CPU & AM4 Socket Pictured, Launching February 2017 – PGA Design With 1331 Pins Confirmed" (<https://wccftech.com/amd-am4-socket-zen-bristol-bridge-soc-package-pictured/>). Wccftech. Retrieved 20 May 2018.
90. "Ryzen 5 2500U – AMD – WikiChip" (https://en.wikichip.org/wiki/amd/ryzen_5/2500u). wikichip.org. Retrieved 20 May 2018.
91. Ung, Gordon Mah (30 May 2017). "AMD's 'TR4' Threadripper CPU socket is gigantic" (<https://www.pcworld.com/article/3198924/computers/amds-tr4-threadripper-cpu-socket-is-gigantic.html>). PCWorld. Retrieved 20 May 2018.
92. "Federal Statutory Protection for Mask Works" (<https://copyright.gov/circs/circ100.pdf>) (PDF). United States Copyright Office. Retrieved 22 October 2016.
93. "Washington Treaty on Intellectual Property in Respect of Integrated Circuits" (<https://www.wipo.int/treaties/en/ip/washington/index.html>). www.wipo.int.

94. On 1 January 1995, the *Agreement on Trade-Related Aspects of Intellectual Property Rights* (TRIPs) (Annex 1C to the World Trade Organization (WTO) Agreement), went into force. Part II, section 6 of TRIPs protects semiconductor chip products and was the basis for Presidential Proclamation No. 6780, 23 March 1995, under SCPA § 902(a)(2), extending protection to all present and future WTO members.
95. Japan was the first country to enact its own version of the SCPA, the Japanese "Act Concerning the Circuit Layout of a Semiconductor Integrated Circuit" of 1985.
96. In 1986 the EC promulgated a directive requiring its members to adopt national legislation for the protection of semiconductor topographies. Council Directive 1987/54/EEC of 16 December 1986 on the *Legal Protection of Topographies of Semiconductor Products*, art. 1(1)(b), 1987 O.J. (L 24) 36.
97. Stern, Richard (1985). "MicroLaw". *IEEE Micro*. **5** (4): 90–92. doi:10.1109/MM.1985.304489 (<https://doi.org/10.1109%2FMM.1985.304489>).
98. Radomsky, Leon (2000). "Sixteen Years after the Passage of the U.S. Semiconductor Chip Protection Act: Is International Protection Working" (<https://heinonline.org/HOL/P?h=hein.journals/berktech15&i=1059>). *Berkeley Technology Law Journal*. **15**: 1069. Retrieved 13 September 2022.
99. Kukkonen, Carl A. III (1997–1998). "The Need to Abolish Registration for Integrated Circuit Topographies under Trips" (<https://heinonline.org/HOL/P?h=hein.journals/idea38&i=115>). *IDEA: The Journal of Law and Technology*. **38**: 126. Retrieved 13 September 2022.
100. Clarke, Peter (14 October 2005). "Intel enters billion-transistor processor era" (<http://www.eetimes.com/electronics-products/processors/4079511/Intel-enters-billion-transistor-processor-era>). *EE Times*. Archived (<https://web.archive.org/web/20110608072423/http://www.eetimes.com/electronics-products/processors/4079511/Intel-enters-billion-transistor-processor-era>) from the original on 8 June 2011.
101. Dalmau, M. "Les Microprocesseurs" (<https://web.archive.org/web/20170809093452/http://www.iutbayonne.univ-pau.fr/~dalmau/documents/cours/archi/MICROPancien.pdf>) (PDF). *IUT de Bayonne*. Archived from the original (<http://www.iutbayonne.univ-pau.fr/~dalmau/documents/cours/archi/MICROPancien.pdf>) (PDF) on 9 August 2017. Retrieved 7 June 2015.
102. *Bulletin de la Société fribourgeoise des sciences naturelles, Volumes 62 à 63* (<https://books.google.com/books?id=ZbcSAQAAIAAJ&q=ssi+msi+12+99+portes+lsi>) (in French). 1973.
103. Safir, Ruben (March 2015). "System on Chip – Integrated Circuits" (<https://books.google.com/books?id=JsOmCQAAQBAJ&pg=PT39>). *NYLXS Journal*. ISBN 9781312995512.
104. Mindell, David A. (2008). *Digital Apollo: Human and Machine in Spaceflight*. The MIT Press. ISBN 978-0-262-13497-2.
105. Ginzberg, Eli (1976). *Economic impact of large public programs: the NASA Experience*. Olympus Publishing Company. p. 57. ISBN 978-0-913420-68-3.
106. Johnstone, Bob (1999). *We were burning: Japanese entrepreneurs and the forging of the electronic age* (<https://books.google.com/books?id=PE1bQS9VpWoC&pg=PA47>). Basic Books. pp. 47–48. ISBN 978-0-465-09118-8.
107. Moskowitz, Sanford L. (2016). *Advanced Materials Innovation: Managing Global Technology in the 21st century* (<https://books.google.com/books?id=2STRDAAAQBAJ&pg=PA165>). John Wiley & Sons. pp. 165–167. ISBN 9780470508923.
108. Boysel, Lee (12 October 2007). "Making Your First Million (and other tips for aspiring entrepreneurs)" (<http://www.eecs.umich.edu/eecs/about/articles/2007/Boysel.html>). *U. Mich. EECS Presentation / ECE Recordings*.
109. Kilby, J. S. (2007). "Miniaturized electronic circuits [US Patent No. 3,138, 743]" (<https://www.researchgate.net/publication/245509003>). *IEEE Solid-State Circuits Society Newsletter*. **12** (2): 44–54. doi:10.1109/N-SSC.2007.4785580 (<https://doi.org/10.1109%2FN-SSC.2007.4785580>).
110. US Patent 3138743 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=US3138743>)

111. Hittinger, William C. (1973). "Metal-Oxide-Semiconductor Technology". *Scientific American*. **229** (2): 48–59. Bibcode:1973SciAm.229b..48H (<https://ui.adsabs.harvard.edu/abs/1973SciAm.229b..48H>). doi:[10.1038/scientificamerican0873-48](https://doi.org/10.1038/scientificamerican0873-48) (<https://doi.org/10.1038%2Fscientificamerican0873-48>). JSTOR 24923169 (<https://www.jstor.org/stable/24923169>).
112. Kanelllos, Michael (16 January 2002). "Intel's Accidental Revolution" (<https://www.cnet.com/news/intels-accidental-revolution/>). *CNET*.
113. "Engineering for Systems Using Large Scale Integration". *International Workshop on Managing Requirements Knowledge, Dec. 9 1968 to Dec. 11 1968, San Francisco*. IEEE Computer Society. p. 867. doi:[10.1109/AFIPS.1968.93](https://doi.org/10.1109/AFIPS.1968.93) (<https://doi.org/10.1109%2FAFIPS.1968.93>).
114. Clarke, Peter (14 October 2005). "Intel enters billion-transistor processor era" (<https://www.eetimes.com/intel-enters-billion-transistor-processor-era/>). *EETimes.com*. Retrieved 23 May 2022.
115. "Samsung First to Mass Produce 16Gb NAND Flash Memory" (<https://phys.org/news/2007-04-samsung-mass-16gb-nand-memory.html>). *phys.org*. 30 April 2007. Retrieved 23 May 2022.
116. Meindl, J.D. (1984). "Ultra-large scale integration". *IEEE Transactions on Electron Devices*. **31** (11): 1555–1561. Bibcode:1984ITED...31.1555M (<https://ui.adsabs.harvard.edu/abs/1984ITED...31.1555M>). doi:[10.1109/T-ED.1984.21752](https://doi.org/10.1109/T-ED.1984.21752) (<https://doi.org/10.1109%2FT-ED.1984.21752>). S2CID 19237178 (<https://api.semanticscholar.org/CorpusID:19237178>).
117. US patent 4866501 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=US4866501>), Shanefield, Daniel, "Wafer scale integration", published 24 May 1985
118. Edwards, Benj (14 November 2022). "Hungry for AI? New supercomputer contains 16 dinner-plate-size chips" (<https://arstechnica.com/information-technology/2022/11/hungry-for-ai-new-supercomputer-contains-16-dinner-plate-size-chips/>). *Ars Technica*.
119. US patent 6816750 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=US6816750>), Klaas, Jeff, "System-on-a-chip", published 24 May 2000
120. Topol, A.W.; Tulipe, D.C.La; Shi, L; et., al (2006). "Three-dimensional integrated circuits". *IBM Journal of Research and Development*. **50** (4.5): 491–506. doi:[10.1147/rd.504.0491](https://doi.org/10.1147/rd.504.0491) (<https://doi.org/10.1147%2Frd.504.0491>). S2CID 18432328 (<https://api.semanticscholar.org/CorpusID:18432328>).
121. The Secret Art of Chip Graffiti (<https://spectrum.ieee.org/the-secret-art-of-chip-graffiti>) *IEEE Spectrum* article on chip artwork, by H. Goldstein, Volume: 39, Issue: 3, Mar 2002, pp. 50–55.

Further reading

- Veendrick, H.J.M. (2025). *Nanometer CMOS ICs, from Basics to ASICs*. Springer. ISBN 978-3-031-64248-7. OCLC 1463505655 (<https://search.worldcat.org/oclc/1463505655>).
- Baker, R.J. (2010). *CMOS: Circuit Design, Layout, and Simulation* (3rd ed.). Wiley-IEEE. ISBN 978-0-470-88132-3. OCLC 699889340 (<https://search.worldcat.org/oclc/699889340>).
- Marsh, Stephen P. (2006). *Practical MMIC design*. Artech House. ISBN 978-1-59693-036-0. OCLC 1261968369 (<https://search.worldcat.org/oclc/1261968369>).
- Camenzind, Hans (2005). *Designing Analog Chips* (https://web.archive.org/web/20170612055924/http://www.designinganalogchips.com/_count/designinganalogchips.pdf) (PDF). Virtual Bookworm. ISBN 978-1-58939-718-7. OCLC 926613209 (<https://search.worldcat.org/oclc/926613209>). Archived from the original (http://www.designinganalogchips.com/_count/designinganalogchips.pdf) (PDF) on 12 June 2017. "Hans Camenzind invented the 555 timer"

- Hodges, David; Jackson, Horace; Saleh, Resve (2003). *Analysis and Design of Digital Integrated Circuits*. McGraw-Hill. ISBN 978-0-07-228365-5. OCLC 840380650 (<https://search.worldcat.org/oclc/840380650>).
- Rabaey, J.M.; Chandrakasan, A.; Nikolic, B. (2003). *Digital Integrated Circuits* (<https://archive.org/details/agilesoftwaredev00robe>) (2nd ed.). Pearson. ISBN 978-0-13-090996-1. OCLC 893541089 (<https://search.worldcat.org/oclc/893541089>).
- Mead, Carver; Conway, Lynn (1991). *Introduction to VLSI systems* (<https://archive.org/details/introductiontovl00mead>). Addison Wesley Publishing Company. ISBN 978-0-201-04358-7. OCLC 634332043 (<https://search.worldcat.org/oclc/634332043>).

External links

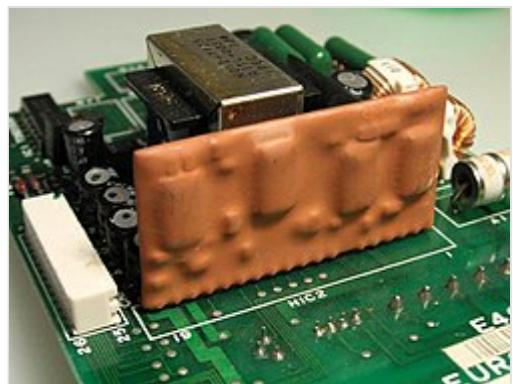
-  Media related to [Integrated circuits](#) at Wikimedia Commons
- [The first monolithic integrated circuits](https://web.archive.org/web/20120319150151/http://homepages.nildram.co.uk/~wylie/ICs/monolith.htm) (<https://web.archive.org/web/20120319150151/http://homepages.nildram.co.uk/~wylie/ICs/monolith.htm>)
- [A large chart listing ICs by generic number](http://rtellason.com/ic-generic.html) (<http://rtellason.com/ic-generic.html>) including access to most of the datasheets for the parts.
- [The History of the Integrated Circuit](https://web.archive.org/web/20170702192457/http://www.nobelprize.org/educational/physics/integrated_circuit/history/) (https://web.archive.org/web/20170702192457/http://www.nobelprize.org/educational/physics/integrated_circuit/history/)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Integrated_circuit&oldid=1291684522"



Hybrid integrated circuit

A **hybrid integrated circuit (HIC)**, **hybrid microcircuit**, **hybrid circuit** or simply **hybrid** is a miniaturized electronic circuit constructed of individual devices, such as semiconductor devices (e.g. transistors, diodes or monolithic ICs) and passive components (e.g. resistors, inductors, transformers, and capacitors), bonded to a substrate or printed circuit board (PCB).^[1] A PCB having components on a Printed wiring board (PWB) is not considered a true hybrid circuit according to the definition of MIL-PRF-38534.

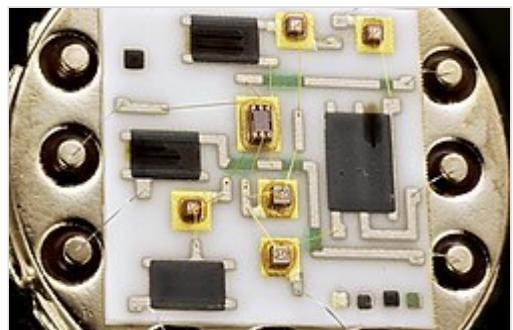


An (orange-epoxy) encapsulated hybrid circuit on a printed circuit board (PCB).

Overview

"Integrated circuit", as the term is currently used, usually refers to a monolithic IC which differs notably from a HIC in that a HIC is fabricated by inter-connecting a number of components on a substrate whereas an IC's (monolithic) components are fabricated in a series of steps entirely on a single wafer which is then diced into chips.^[2] Some hybrid circuits may contain monolithic ICs, particularly Multi-chip module (MCM) hybrid circuits.

Hybrid circuits could be encapsulated in epoxy, as shown in the photo, or in military and space applications, a lid was soldered onto the package. A hybrid circuit serves as a component on a PCB in the same way as a monolithic integrated circuit; the difference between the two types of devices is in how they are constructed and manufactured. The advantage of hybrid circuits is that components which cannot be included in a monolithic IC can be used, e.g., capacitors of large value, wound components, crystals, inductors.^[3] In military and space applications, numerous integrated circuits, transistors and diodes, in their die form, would be placed on either a ceramic or beryllium substrate. Either gold or aluminum wire would be bonded from the pads of the IC, transistor, or diode to the substrate.

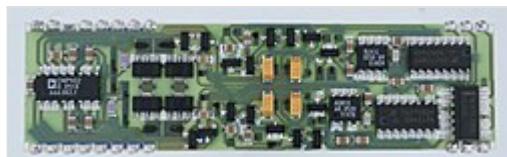


A hybrid operational amplifier on a ceramic substrate with laser trimmed thick film resistors

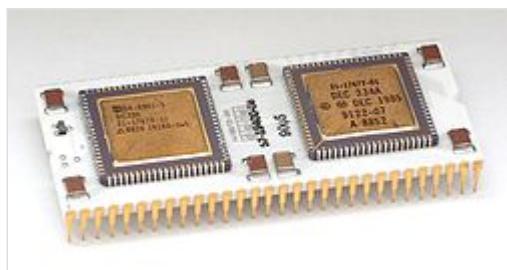
Thick film technology is often used as the interconnecting medium for hybrid integrated circuits. The use of screen printed thick film interconnect provides advantages of versatility over thin film although feature sizes may be larger and deposited resistors wider in tolerance. Multi-layer thick film is a technique for further improvements in integration using a screen printed insulating dielectric to ensure connections between layers are made only where required. One key advantage for the circuit designer is complete freedom in the choice of resistor value in thick film technology. Planar resistors are also screen printed and included in the thick film interconnect design. The composition and dimensions of resistors can be

selected to provide the desired values. The final resistor value is determined by design and can be adjusted by laser trimming. Once the hybrid circuit is fully populated with components, fine tuning prior to final test may be achieved by active laser trimming.

Thin film technology was also employed in the 1960s. Ultra Electronics manufactured circuits using a silica glass substrate. A film of tantalum was deposited by sputtering followed by a layer of gold by evaporation. The gold layer was first etched following the application of a photoresist to form solder compatible connection pads. Resistive networks were formed, also by a photoresist and etching process. These were trimmed to a high precision by selective adhesion of the film. Capacitors and semiconductors were in the form of LID (Leadless Inverted Devices) soldered to the surface by selectively heating the substrate from the underside. Completed circuits were potted in a diallyl phthalate resin. Several customized passive networks were made using these techniques as were some amplifiers and other specialized circuits. It is believed that some passive networks were used in the engine control units manufactured by Ultra Electronics for Concorde.

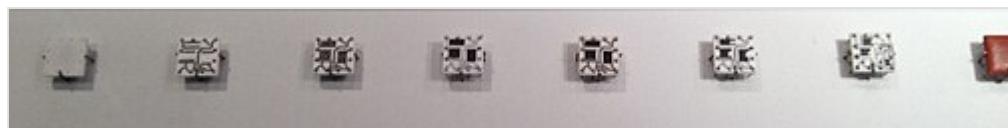


A hybrid PCB on a ceramic substrate with laser trimmed thick film components



DEC J-11 microprocessor using a hybrid PCB

Some modern hybrid circuit technologies, such as LTCC-substrate hybrids, allow for embedding of components within the layers of a multi-layer substrate in addition to components placed on the surface of the substrate. This technology produces a circuit that is, to some degree, three-dimensional.



Steps in manufacturing Solid Logic Technology hybrid wafers used in the IBM System/360 and other IBM computers of the mid-1960s. The process starts with a blank ceramic wafer 1/2 inch square. Circuits are laid down first, followed by resistive material. The circuits are metalized and the resistors trimmed to the desired value. Then discrete transistors and diodes are added and the package encapsulated. Display at the Computer History Museum.

Hybrid ICs are especially suitable for analog signals. They were used in some early digital computers but were replaced therein by monolithic ICs which offered higher performance.^{[4][5][6]}

Other electronic hybrids

In the early days of telephones, separate modules containing transformers and resistors were called hybrids or hybrid coils; they have been replaced by semiconductor integrated circuits.

In the early days of transistors the term *hybrid circuit* was used to describe circuits with both transistors and vacuum tubes; e.g., an audio amplifier with transistors used for voltage amplification followed by a vacuum tube power output stage, as suitable power transistors were not available. This usage, and the devices, are obsolete, however amplifiers that use a tube preamplifier stage coupled with a solid state output stage are still in production, and are called hybrid amplifiers in reference to this.

See also

- Chip on board, aka black blobs
- System in a package
- Multi-chip module (MCM)
- Monolithic microwave integrated circuit (MMIC)
- Solid Logic Technology (SLT)
- MIL-PRF-38534
- Printed circuit board (PCB)
- Printed Electronic Circuit - Ancestor of the Hybrid IC

References

1. "Tell me... Just what is a Hybrid Integrated Circuit?" (<https://www.escomponents.com/blog/2017/9/7/tell-me-just-what-is-a-hybrid-integrated-circuit>). *Electronic Components Franchised Distributor - Military Certified Manufacturer | ES Components*. ES Components. September 7, 2017. Retrieved 13 May 2025.
2. "Difference between Monolithic ICs and Hybrid ICs (integrated circuits)" (<http://www.polytechnichub.com/difference-monolithic-ic-hybrid-ic/>). Polytechnic Hub. March 2, 2017. Retrieved 13 May 2025.
3. William Greig, *Integrated Circuit Packaging, Assembly and Interconnections*, Springer Science & Business Media, 2007, ISBN 0387339132, p.62-64
4. Pugh, Emerson W.; Johnson, Lyle R.; Palmer, John H. (1991). *IBM's 360 and Early 370 Systems* (https://books.google.com/books?id=MFGj_PT_cIIC&dq=ibm+hybrid+integrated+circuit&pg=PA106). MIT Press. ISBN 978-0-262-16123-7.
5. "Definition of hybrid microcircuit" (<https://www.pcmag.com/encyclopedia/term/hybrid-microcircuit>).
6. Saxena, Arjun N. (2009). *Invention of Integrated Circuits: Untold Important Facts* (<https://books.google.com/books?id=z7738Wq-j-8C&dq=computer+hybrid+integrated+circuit&pg=PA62>). World Scientific. ISBN 978-981-281-446-3.

External links

-  Media related to Hybrid integrated circuits at Wikimedia Commons



Three-dimensional integrated circuit

A **three-dimensional integrated circuit** (3D IC) is a MOS (metal-oxide semiconductor) integrated circuit (IC) manufactured by stacking as many as 16 or more ICs and interconnecting them vertically using, for instance, through-silicon vias (TSVs) or Cu-Cu connections,^{[1][2]} so that they behave as a single device to achieve performance improvements at reduced power and smaller footprint than conventional two dimensional processes. The 3D IC is one of several 3D integration schemes that exploit the z-direction to achieve electrical performance benefits in microelectronics and nanoelectronics.

3D integrated circuits can be classified by their level of interconnect hierarchy at the global (package), intermediate (bond pad) and local (transistor) level.^[3] In general, 3D integration is a broad term that includes such technologies as 3D wafer-level packaging (3DWLP); 2.5D and 3D interposer-based integration; 3D stacked ICs (3D-SICs); 3D heterogeneous integration; and 3D systems integration;^{[4][5]} as well as true monolithic 3D ICs.

International organizations such as the Jisso Technology Roadmap Committee (JIC) and the International Technology Roadmap for Semiconductors (ITRS) have worked to classify the various 3D integration technologies to further the establishment of standards and roadmaps of 3D integration.^[6] As of the 2010s, 3D ICs are widely used for NAND flash memory and in mobile devices.

Types

3D ICs vs. 3D packaging

3D packaging refers to 3D integration schemes that rely on traditional interconnection methods such as wire bonding and flip chip to achieve vertical stacking. 3D packaging can be divided into 3D system in package (3D SiP) and 3D wafer level package (3D WLP). 3D SiPs that have been in mainstream manufacturing for some time and have a well-established infrastructure include stacked memory dies interconnected with wire bonds and package on package (PoP) configurations interconnected with wire bonds or flip chip technology. PoP is used for vertically integrating disparate technologies. 3D WLP uses wafer level processes such as redistribution layers (RDLs) and wafer bumping processes to form interconnects.

2.5D interposer is a 3D WLP that interconnects dies side-by-side on a silicon, glass, or organic interposer using through silicon vias (TSVs) and an RDL. In all types of 3D packaging, chips in the package communicate using off-chip signaling, much as if they were mounted in separate packages on a normal printed circuit board. The interposer may be made of silicon, and is under the dies it connects together. A design can be split into several dies, and then mounted on the interposer with micro bumps.^{[7][8][9]}

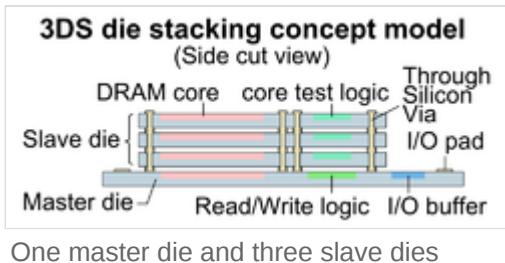
3D ICs can be divided into 3D Stacked ICs (3D SIC), which refers to advanced packaging techniques^{[10][11][12]} stacking IC chips using TSV interconnects, and monolithic 3D ICs, which use fab processes to realize 3D interconnects at the local levels of the on-chip wiring hierarchy as set forth by the

ITRS, this results in direct vertical interconnects between device layers. The first examples of a monolithic approach are seen in Samsung's 3D V-NAND devices.^[13]

As of the 2010s, 3D IC packages are widely used for NAND flash memory in mobile devices.^[14]

3D SiCs

The digital electronics market requires a higher density semiconductor memory chip to cater to recently released CPU components, and the multiple die stacking technique has been suggested as a solution to this problem. JEDEC disclosed the upcoming DRAM technology includes the "3D SiC" die stacking plan at "Server Memory Forum", November 1–2, 2011, Santa Clara, CA. In August 2014, Samsung Electronics started producing 64 GB SDRAM modules for servers based on emerging DDR4 (double-data rate 4) memory using 3D TSV package technology.^[15] Newer proposed standards for 3D stacked DRAM include Wide I/O, Wide I/O 2, Hybrid Memory Cube, High Bandwidth Memory.



Monolithic 3D ICs

True monolithic 3D ICs are built in layers on a single semiconductor wafer, which is then diced into 3D ICs. There is only one substrate, hence no need for aligning, thinning, bonding, or through-silicon vias. In general, monolithic 3D ICs are still a developing technology and are considered by most to be several years away from production.

Process temperature limitations can be addressed by partitioning the transistor fabrication into two phases. A high temperature phase which is done before layer transfer followed by a layer transfer using ion-cut (<http://www.monolithic3d.com/ion-cut-the-building-block.html>), also known as layer transfer, which has been used to produce Silicon on Insulator (SOI) wafers for the past two decades. Multiple thin (10s–100s nanometer scale) layers of virtually defect-free Silicon can be created by utilizing low temperature (<400 °C) bond and cleave techniques, and placed on top of active transistor circuitry, followed by permanent finalization of the transistors using etch and deposition processes. This monolithic 3D IC technology has been researched at Stanford University under a DARPA-sponsored grant.

CEA-Leti also developed monolithic 3D IC approaches, called sequential 3D IC. In 2014, the French research institute introduced its CoolCube™, a low-temperature process flow that provides a true path to 3DVLSI.^[16]

At Stanford University, researchers designed monolithic 3D ICs using carbon nanotube (CNT) structures vs. silicon using a wafer-scale low temperature CNT transfer processes that can be done at 120 °C.^[17]

Manufacturing technologies for 3D SiCs

There are several methods for 3D IC design, including recrystallization and wafer bonding methods. There are two major types of wafer bonding, Cu-Cu connections (copper-to-copper connections between stacked ICs, used in TSVs)^{[18][19]} and through-silicon via (TSV). 3D ICs with TSVs may use solder microbumps, small solder balls as an interface between two individual dies in a 3D IC.^[20] As of 2014, a

number of memory products such as High Bandwidth Memory (HBM) and the Hybrid Memory Cube have been launched that implement 3D IC stacking with TSVs. There are a number of key stacking approaches being implemented and explored. These include die-to-die, die-to-wafer, and wafer-to-wafer.

Die-to-Die

Electronic components are built on multiple die, which are then aligned and bonded. Thinning and TSV creation may be done before or after bonding. One advantage of die-to-die is that each component die can be tested first, so that one bad die does not ruin an entire stack.^[21] Moreover, each die in the 3D IC can be binned beforehand, so that they can be mixed and matched to optimize power consumption and performance (e.g. matching multiple dice from the low power process corner for a mobile application).

Die-to-Wafer

Electronic components are built on two semiconductor wafers. One wafer is diced; the singulated dice are aligned and bonded onto die sites of the second wafer. As in the wafer-on-wafer method, thinning and TSV creation are performed either before or after bonding. Additional die may be added to the stacks before dicing.^[22]

Wafer-to-Wafer

Electronic components are built on two or more semiconductor wafers, which are then aligned, bonded, and diced into 3D ICs. Each wafer may be thinned before or after bonding. Vertical connections are either built into the wafers before bonding or else created in the stack after bonding. These "through-silicon vias" (TSVs) pass through the silicon substrate(s) between active layers and/or between an active layer and an external bond pad. Wafer-to-wafer bonding can reduce yields, since if any 1 of N chips in a 3D IC are defective, the entire 3D IC will be defective. Moreover, the wafers must be the same size, but many exotic materials (e.g. III-Vs) are manufactured on much smaller wafers than CMOS logic or DRAM (typically 300 mm), complicating heterogeneous integration.

Benefits

While traditional CMOS scaling processes improves signal propagation speed, scaling from current manufacturing and chip-design technologies is becoming more difficult and costly, in part because of power-density constraints, and in part because interconnects do not become faster while transistors do.^[23] 3D ICs address the scaling challenge by stacking 2D dies and connecting them in the 3rd dimension. This promises to speed up communication between layered chips, compared to planar layout.^[24] 3D ICs promise many significant benefits, including:

Footprint

More functionality fits into a small space. The smaller form factors are of great importance in embedded devices such as mobile phones, IoT systems for which 3D non-volatile memory stacks have been developed (e.g. 3D NAND chips) [1] (<https://www.electronicdesign.com/technologies/embedded/article/21247462/electronic-design-3d-nand-memory-chips-stacked-high-at-micron>) :: **Moore's Law Extension**: The increased number of transistors being packed in the same footprint is seen as an extension to Moore's law by some researchers. This enables extending the Moore's Law without its traditional pair of Dennard Scaling towards a new generation of chips with increased computing capacity for the same footprint.[2] (<https://spectrum.ieee.org/3d-cmos>):

Cost

Partitioning a large chip into multiple smaller dies with 3D stacking can improve the yield and reduce the fabrication cost if individual dies are tested separately.^{[25][26]}

Heterogeneous Integration

Circuit layers can be built with different processes, or even on different types of wafers. This means that components can be optimized to a much greater degree than if they were

built together on a single wafer. Moreover, components with incompatible manufacturing could be combined in a single 3D IC.^{[27][5]}

Shorter Interconnect

The average wire length is reduced. Common figures reported by researchers are on the order of 10–15%, but this reduction mostly applies to longer interconnect, which may affect circuit delay by a greater amount. Given that 3D wires have much higher capacitance than conventional in-die wires, circuit delay may or may not improve.

Power

Keeping a signal on-chip can reduce its power consumption by 10–100 times.^[28] Shorter wires also reduce power consumption by producing less parasitic capacitance.^[29] Reducing the power budget leads to less heat generation, extended battery life, and lower cost of operation.

Design

The vertical dimension adds a higher order of connectivity and offers new design possibilities.^[5]

Circuit Security

3D integration can achieve security through obscurity; the stacked structure complicates attempts to reverse engineer the circuitry. Sensitive circuits may also be divided among the layers in such a way as to obscure the function of each layer.^[30] Moreover, 3D integration allows to integrate dedicated, system monitor-like features in separate layers.^[5] The objective here is to implement some kind of hardware firewall for any commodity components/chips to be monitored at runtime, seeking to protect the whole electronic system against run-time attacks as well as malicious hardware modifications.

Bandwidth

3D integration allows large numbers of vertical vias between the layers. This allows construction of wide bandwidth buses between functional blocks in different layers. A typical example would be a processor+memory 3D stack, with the cache memory stacked on top of the processor. This arrangement allows a bus much wider than the typical 128 or 256 bits between the cache and processor.^[31] Wide buses in turn alleviate the memory wall problem.^[32]

Modularity

3D integration modular integration a wide range of custom stacks through standardizing the layer interfaces for numerous stacking options. As a result, custom stack designs can be manufactured with modular building blocks (e.g. custom number of DRAM or eDRAM layers, custom accelerator layers, customizable Non-Volatile Memory layers can be integrated to meet different design requirements). This provides design and cost advantages to semiconductor firms.^[3] (<https://ieeexplore.ieee.org/abstract/document/5510586>)

Other potential advantages include better integration of neuromorphic chips in computing systems. Despite being low power alternatives to general purpose CPUs and GPUs, neuromorphic chips use a fundamentally different "spike-based" computation, which is not directly compatible with legacy digital computation. 3D integration provides key opportunities in this integration.^[4] (<https://physicsworld.com/a/are-neuromorphic-systems-the-future-of-high-performance-computing/>)

Challenges

Because this technology is new, it carries new challenges, including:

Cost

While cost is a benefit when compared with scaling, it has also been identified as a challenge to the commercialization of 3D ICs in mainstream consumer applications. However, work is being done to address this. Although 3D technology is new and fairly complex, the cost of the manufacturing process is surprisingly straightforward when broken down into the activities that build up the entire process. By analyzing the combination of activities that lay at the base, cost drivers can be identified. Once the cost drivers are identified, it becomes a less complicated endeavor to determine where the majority of cost comes from and, more importantly, where cost has the potential to be reduced.^[33]

Yield

Each extra manufacturing step adds a risk for defects. In order for 3D ICs to be commercially viable, defects could be repaired or tolerated, or defect density can be improved.^{[34][35]}

Heat

Heat building up within the stack must be dissipated. This is an inevitable issue as electrical proximity correlates with thermal proximity. Specific thermal hotspots must be more carefully managed.

Design Complexity

Taking full advantage of 3D integration requires sophisticated design techniques and new CAD tools.^[36]

TSV-introduced Overhead

TSVs are large compared to gates and impact floorplans. At the 45 nm technology node, the area footprint of a 10µm x 10µm TSV is comparable to that of about 50 gates.^[37] Furthermore, manufacturability demands landing pads and keep-out zones which further increase TSV area footprint. Depending on the technology choices, TSVs block some subset of layout resources.^[37] Via-first TSVs are manufactured before metallization, thus occupy the device layer and result in placement obstacles. Via-last TSVs are manufactured after metallization and pass through the chip. Thus, they occupy both the device and metal layers, resulting in placement and routing obstacles. While the usage of TSVs is generally expected to reduce wirelength, this depends on the number of TSVs and their characteristics.^[37] Also, the granularity of inter-die partitioning impacts wirelength. It typically decreases for moderate (blocks with 20-100 modules) and coarse (block-level partitioning) granularities, but increases for fine (gate-level partitioning) granularities.^[37]

Testing

To achieve high overall yield and reduce costs, separate testing of independent dies is essential.^{[35][38]} However, tight integration between adjacent active layers in 3D ICs entails a significant amount of interconnect between different sections of the same circuit module that were partitioned to different dies. Aside from the massive overhead introduced by required TSVs, sections of such a module, e.g., a multiplier, cannot be independently tested by conventional techniques. This particularly applies to timing-critical paths laid out in 3D.

Lack of Standards

There are few standards for TSV-based 3D IC design, manufacturing, and packaging, although this issue is being addressed.^{[39][40]} In addition, there are many integration options being explored such as via-last, via-first, via-middle;^[41] interposers^[42] or direct bonding; etc.

Heterogeneous Integration Supply Chain

In heterogeneously integrated systems, the delay of one part from one of the different parts suppliers delays the delivery of the whole product, and so delays the revenue for each of the 3D IC part suppliers.

Lack of Clearly Defined Ownership

It is unclear who should own the 3D IC integration and packaging/assembly. It could be assembly houses like ASE or the product OEMs.

Thermomechanical Stress and Reliability

3D stacks have more complex material compositions and thermomechanical profiles compared to 2D designs. The stacking of multiple thinned silicon layers, multiple wiring (BEOL) layers, insulators, through silicon vias, micro-C4s result in complex thermomechanical forces and stress patterns being exerted to the 3D stacks. As a result, local heating in one part of the stack (e.g. on thinned device layers) may result reliability challenges. This requires design-time analysis and reliability-aware design processes. [5] (<https://ieeexplore.ieee.org/abstract/document/7955018>)

Design styles

Depending on partitioning granularity, different design styles can be distinguished. Gate-level integration faces multiple challenges and currently appears less practical than block-level integration.^[43]

Gate-level Integration

This style partitions standard cells between multiple dies. It promises wirelength reduction and great flexibility. However, wirelength reduction may be undermined unless modules of certain minimal size are preserved. On the other hand, its adverse effects include the massive number of necessary TSVs for interconnects. This design style requires 3D place-and-route tools, which are unavailable yet. Also, partitioning a design block across multiple dies implies that it cannot be fully tested before die stacking. After die stacking (post-bond testing), a single failed die can render several good dies unusable, undermining yield. This style also amplifies the impact of process variation, especially inter-die variation. In fact, a 3D layout may yield more poorly than the same circuit laid out in 2D, contrary to the original promise of 3D IC integration.^[44] Furthermore, this design style requires to redesign available Intellectual Property, since existing IP blocks and EDA tools do not provision for 3D integration.

Block-level Integration

This style assigns entire design blocks to separate dies. Design blocks subsume most of the netlist connectivity and are linked by a small number of global interconnects. Therefore, block-level integration promises to reduce TSV overhead. Sophisticated 3D systems combining heterogeneous dies require distinct manufacturing processes at different technology nodes for fast and low-power random logic, several memory types, analog and RF circuits, etc. Block-level integration, which allows separate and optimized manufacturing processes, thus appears crucial for 3D integration. Furthermore, this style might facilitate the transition from current 2D design towards 3D IC design. Basically, 3D-aware tools are only needed for partitioning and thermal analysis.^[45] Separate dies will be designed using (adapted) 2D tools and 2D blocks. This is motivated by the broad availability of reliable IP blocks. It is more convenient to use available 2D IP blocks and to place the mandatory TSVs in the unoccupied space between blocks instead of redesigning IP blocks and embedding TSVs.^[43] Design-for-testability structures are a key component of IP blocks and can therefore be used to facilitate testing for 3D ICs. Also, critical paths can be mostly embedded within 2D blocks, which limits the impact of TSV and inter-die variation on manufacturing yield. Finally, modern chip design often requires last-minute engineering changes. Restricting the impact of such changes to single dies is essential to limit cost.

History

Several years after the MOS integrated circuit (MOS IC) chip was first proposed by Mohamed Atalla at Bell Labs in 1960,^[46] the concept of a three-dimensional MOS integrated circuit was proposed by Texas Instruments researchers Robert W. Haisty, Rowland E. Johnson and Edward W. Mehal in 1964.^[47] In 1969, the concept of a three-dimensional MOS integrated circuit memory chip was proposed by NEC researchers Katsuhiro Onoda, Ryo Igarashi, Toshio Wada, Sho Nakanuma and Toru Tsujide.^[48]

Arm has made a high-density 3D logic test chip,^[49] and Intel with its Foveros 3D logic chip packing is planning to ship CPUs using it.^[50] IBM demonstrated a fluid that could be used for both power delivery and cooling 3D ICs.^[51]

Demonstrations (1983–2012)

Japan (1983–2005)

3D ICs were first successfully demonstrated in 1980s Japan, where research and development (R&D) on 3D ICs was initiated in 1981 with the "Three Dimensional Circuit Element R&D Project" by the Research and Development Association for Future (New) Electron Devices.^[52] There were initially two forms of 3D IC design being investigated, recrystallization and wafer bonding, with the earliest successful demonstrations using recrystallization.^[19] In October 1983, a Fujitsu research team including S. Kawamura, Nobuo Sasaki and T. Iwai successfully fabricated a three-dimensional complementary metal–oxide–semiconductor (CMOS) integrated circuit, using laser beam recrystallization. It consisted of a structure in which one type of transistor is fabricated directly above a transistor of the opposite type, with separate gates and an insulator in between. A double-layer of silicon nitride and phosphosilicate glass (PSG) film was used as an intermediate insulating layer between the top and bottom devices. This provided the basis for realizing a multi-layered 3D device composed of vertically stacked transistors, with separate gates and an insulating layer in between.^[53] In December 1983, the same Fujitsu research team fabricated a 3D integrated circuit with a silicon-on-insulator (SOI) CMOS structure.^[54] The following year, they fabricated a 3D gate array with vertically stacked dual SOI/CMOS structure using beam recrystallization.^[55]

In 1986, Mitsubishi Electric researchers Yoichi Akasaka and Tadashi Nishimura laid out the basic concepts and proposed technologies for 3D ICs.^{[56][57]} The following year, a Mitsubishi research team including Nishimura, Akasaka and Osaka University graduate Yasuo Inoue fabricated an image signal processor (ISP) on a 3D IC, with an array of photosensors, CMOS A-to-D converters, arithmetic logic units (ALU) and shift registers arranged in a three-layer structure.^[58] In 1989, an NEC research team led by Yoshihiro Hayashi fabricated a 3D IC with a four-layer structure using laser beam crystallisation.^{[59][56]} In 1990, a Matsushita research team including K. Yamazaki, Y. Itoh and A. Wada fabricated a parallel image signal processor on a four-layer 3D IC, with SOI (silicon-on-insulator) layers formed by laser recrystallization, and the four layers consisting of an optical sensor, level detector, memory and ALU.^[60]

The most common form of 3D IC design is wafer bonding.^[19] Wafer bonding was initially called "cumulatively bonded IC" (CUBIC), which began development in 1981 with the "Three Dimensional Circuit Element R&D Project" in Japan and was completed in 1990 by Yoshihiro Hayashi's NEC research

team, who demonstrated a method where several thin-film devices are bonded cumulatively, which would allow a large number of device layers. They proposed fabrication of separate devices in separate wafers, reduction in the thickness of the wafers, providing front and back leads, and connecting the thinned die to each other. They used CUBIC technology to fabricate and test a two active layer device in a top-to-bottom fashion, having a bulk-Si NMOS FET lower layer and a thinned NMOS FET upper layer, and proposed CUBIC technology that could fabricate 3D ICs with more than three active layers.^{[56][52][61]}

The first 3D IC stacked chips fabricated with a through-silicon via (TSV) process were invented in 1980s Japan. Hitachi filed a Japanese patent in 1983, followed by Fujitsu in 1984. In 1986, a Japanese patent filed by Fujitsu described a stacked chip structure using TSV.^[52] In 1989, Mitsumasa Koyonagi of Tohoku University pioneered the technique of wafer-to-wafer bonding with TSV, which he used to fabricate a 3D LSI chip in 1989.^{[52][62][63]} In 1999, the Association of Super-Advanced Electronics Technologies (ASET) in Japan began funding the development of 3D IC chips using TSV technology, called the "R&D on High Density Electronic System Integration Technology" project.^{[52][64]} The term "through-silicon via" (TSV) was coined by Tru-Si Technologies researchers Sergey Savastiouk, O. Siniaguine, and E. Korczynski, who proposed a TSV method for a 3D wafer-level packaging (WLP) solution in 2000.^[65]

The Koyanagi Group at Tohoku University, led by Mitsumasa Koyanagi, used TSV technology to fabricate a three-layer memory chip in 2000, a three-layer artificial retina chip in 2001, a three-layer microprocessor in 2002, and a ten-layer memory chip in 2005.^[62] The same year, a Stanford University research team consisting of Kaustav Banerjee, Shukri J. Souri, Pawan Kapur and Krishna C. Saraswat presented a novel 3D chip design that exploits the vertical dimension to alleviate the interconnect related problems and facilitates heterogeneous integration of technologies to realize a system-on-a-chip (SoC) design.^{[66][67]}

In 2001, a Toshiba research team including T. Imoto, M. Matsui and C. Takubo developed a "System Block Module" wafer bonding process for manufacturing 3D IC packages.^{[56][68]}

Europe (1988–2005)

Fraunhofer and Siemens began research on 3D IC integration in 1987.^[52] In 1988, they fabricated 3D CMOS IC devices based on re-crystallization of poly-silicon.^[69] In 1997, the inter-chip via (ICV) method was developed by a Fraunhofer–Siemens research team including Peter Ramm, Manfred Engelhardt, Werner Pamler, Christof Landesberger and Armin Klumpp.^[70] It was a first industrial 3D IC process, based on Siemens CMOS fab wafers. A variation of that TSV process was later called TSV-SLID (solid liquid inter-diffusion) technology.^[71] It was an approach to 3D IC design based on low temperature wafer bonding and vertical integration of IC devices using inter-chip vias, which they patented.

Ramm went on to develop industry-academic consortia for production of relevant 3D integration technologies. In the German funded cooperative VIC project between Siemens and Fraunhofer, they demonstrated a complete industrial 3D IC stacking process (1993–1996). With his Siemens and Fraunhofer colleagues, Ramm published results showing the details of key processes such as 3D metallization [T. Grassl, P. Ramm, M. Engelhardt, Z. Gabric, O. Spindler, First International Dielectrics for VLSI/ULSI Interconnection Metallization Conference – DUMIC, Santa Clara, CA, 20–22 Feb, 1995] and at ECTC 1995 they presented early investigations on stacked memory in processors.^[72]

In the early 2000s, a team of Fraunhofer and Infineon Munich researchers investigated 3D TSV technologies with particular focus on die-to-substrate stacking within the German/Austrian EUREKA project VSI and initiated the European Integrating Projects e-CUBES, as a first European 3D technology platform, and e-BRAINS with a.o., Infineon, Siemens, EPFL, IMEC and Tyndall, where heterogeneous 3D integrated system demonstrators were fabricated and evaluated. A particular focus of the e-BRAINS project was the development of novel low-temperature processes for highly reliable 3D integrated sensor systems.^[73]

United States (1999–2012)

Copper-to-copper wafer bonding, also called Cu-Cu connections or Cu-Cu wafer bonding, was developed at MIT by a research team consisting of Andy Fan, Adnan-ur Rahman and Rafael Reif in 1999.^{[19][74]} Reif and Fan further investigated Cu-Cu wafer bonding with other MIT researchers including Kuan-Neng Chen, Shamik Das, Chuan Seng Tan and Nisha Checka during 2001–2002.^[19] In 2003, DARPA and the Microelectronics Center of North Carolina (MCNC) began funding R&D on 3D IC technology.^[52]

In 2004, Tezzaron Semiconductor^[75] built working 3D devices from six different designs.^[76] The chips were built in two layers with "via-first" tungsten TSVs for vertical interconnection. Two wafers were stacked face-to-face and bonded with a copper process. The top wafer was thinned and the two-wafer stack was then diced into chips. The first chip tested was a simple memory register, but the most notable of the set was an 8051 processor/memory stack^[77] that exhibited much higher speed and lower power consumption than an analogous 2D assembly.

In 2004, Intel presented a 3D version of the Pentium 4 CPU.^[78] The chip was manufactured with two dies using face-to-face stacking, which allowed a dense via structure. Backside TSVs are used for I/O and power supply. For the 3D floorplan, designers manually arranged functional blocks in each die aiming for power reduction and performance improvement. Splitting large and high-power blocks and careful rearrangement allowed to limit thermal hotspots. The 3D design provides 15% performance improvement (due to eliminated pipeline stages) and 15% power saving (due to eliminated repeaters and reduced wiring) compared to the 2D Pentium 4.

The Teraflops Research Chip introduced in 2007 by Intel is an experimental 80-core design with stacked memory. Due to the high demand for memory bandwidth, a traditional I/O approach would consume 10 to 25 W.^[38] To improve upon that, Intel designers implemented a TSV-based memory bus. Each core is connected to one memory tile in the SRAM die with a link that provides 12 GB/s bandwidth, resulting in a total bandwidth of 1 TB/s while consuming only 2.2 W.

An academic implementation of a 3D processor was presented in 2008 at the University of Rochester by Professor Eby Friedman and his students. The chip runs at a 1.4 GHz and it was designed for optimized vertical processing between the stacked chips which gives the 3D processor abilities that the traditional one layered chip could not reach.^[79] One challenge in manufacturing of the three-dimensional chip was to make all of the layers work in harmony without any obstacles that would interfere with a piece of information traveling from one layer to another.^[80]

In ISSCC 2012, two 3D-IC-based multi-core designs using GlobalFoundries' 130 nm process and Tezzaron's FaStack technology were presented and demonstrated:

- 3D-MAPS,^[81] a 64 custom core implementation with two-logic-die stack, was demonstrated by researchers from the School of Electrical and Computer Engineering at Georgia Institute of Technology.
- Centip3De,^[82] near-threshold design based on ARM Cortex-M3 cores, was from the Department of Electrical Engineering and Computer Science at University of Michigan.

Though released much later IBM Research and Semiconductor Research and Development Groups design and manufactured a number of 3D processor stacks successfully starting from 2007-2008. These stacks (dubbed Escher internally) have demonstrated successful implementation of eDRAM, logic and processor stacks as well as key experiments in power, thermal, noise and reliability characterization of 3D chips. [6] (<https://ieeexplore.ieee.org/abstract/document/6176968>)

Commercial 3D ICs (2004–present)

The earliest known commercial use of a 3D IC chip was in Sony's PlayStation Portable (PSP) handheld game console, released in 2004. The PSP hardware includes eDRAM (embedded DRAM) memory manufactured by Toshiba in a 3D system-in-package chip with two dies stacked vertically.^[14] Toshiba called it "semi-embedded DRAM" at the time, before later calling it a stacked "chip-on-chip" (CoC) solution.^{[14][83]}



Sony's PlayStation Portable (PSP) handheld game console, released in 2004, is the earliest commercial product to use a 3D IC, an eDRAM memory chip manufactured by Toshiba in a 3D system-in-package.

In April 2007, Toshiba commercialized an eight-layer 3D IC, the 16 GB THGAM embedded NAND flash memory chip, which was manufactured with eight stacked 2 GB NAND flash chips.^[84] In September 2007, Hynix introduced 24-layer 3D IC technology, with a 16 GB flash memory chip that was manufactured with 24 stacked NAND flash chips using a wafer bonding process.^[85] Toshiba also used an eight-layer 3D IC for their 32 GB THGBM flash chip in 2008.^[86] In 2010, Toshiba used a 16-layer 3D IC for their 128 GB THGBM2 flash chip, which was manufactured with 16 stacked 8 GB chips.^[87] In the 2010s, 3D ICs came into widespread commercial use in the form of multi-chip package and package on package solutions for NAND flash memory in mobile devices.^[14]

Elpida Memory developed the first 8 GB DRAM chip (stacked with four DDR3 SDRAM dies) in September 2009, and released it in June 2011.^[88] TSMC announced plans for 3D IC production with TSV technology in January 2010.^[88] In 2011, SK Hynix introduced 16 GB DDR3 SDRAM (40 nm class) using TSV technology,^[89] Samsung Electronics introduced 3D-stacked 32 GB DDR3 (30 nm class) based on TSV in September, and then Samsung and Micron Technology announced TSV-based Hybrid Memory Cube (HMC) technology in October.^[88]

High Bandwidth Memory (HBM), developed by Samsung, AMD, and SK Hynix, uses stacked chips and TSVs. The first HBM memory chip was manufactured by SK Hynix in 2013.^[89] In January 2016, Samsung Electronics announced early mass production of HBM2, at up to 8 GB per stack.^{[90][91]}

In 2017, Samsung Electronics combined 3D IC stacking with its 3D V-NAND technology (based on charge trap flash technology), manufacturing its 512 GB KLUFG8R1EM flash memory chip with eight stacked 64-layer V-NAND chips.^[92] In 2019, Samsung produced a 1 TB flash chip with 16 stacked V-

NAND dies.^{[93][94]} As of 2018, Intel is considering the use of 3D ICs to improve performance.^[95] As of 2022, 232-layer NAND, i.e. memory device, chips are made by Micron,^[96] that previously in April 2019 were making 96-layer chips; and Toshiba made 96-layer devices in 2018.

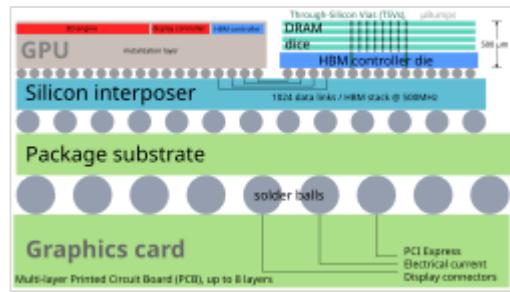
In 2022, AMD has introduced Zen 4 processors, and some Zen 4 processors have 3D Cache included.

See also

- 2.5D integrated circuit
- Advanced packaging (semiconductors)
- Charge trap flash (CTF)
- FinFET (3D transistor)
- MOSFET
- Multigate device (MuGFET)
- V-NAND (3D NAND)

Notes

1. Hu, Y.H.; Liu, C.S.; Lii, M.J.; Rebibis, K. J.; Jourdain, A.; La Manna, A.; Beyne, E.; Yu, C.H. (2012). "Cu-Cu hybrid bonding as option for 3D IC stacking" (<https://ieeexplore.ieee.org/document/6251571>). *2012 IEEE International Interconnect Technology Conference*. pp. 1–3. doi:10.1109/IITC.2012.6251571 (<https://doi.org/10.1109%2FIITC.2012.6251571>). ISBN 978-1-4673-1137-3.
2. *Wafer Bonding: Applications and Technology* (<https://books.google.com/books?id=62fyCAAQBAJ&dq=cu+cu+bonding&pg=PA184>). Springer. 9 March 2013. ISBN 978-3-662-10827-7.
3. "SEMI.ORG" (http://www.semi.org/cms/groups/public/documents/web_content/ctr_033139.pdf) (PDF). Archived (https://web.archive.org/web/20150924100353/http://www.semi.org/cms/groups/public/documents/web_content/ctr_033139.pdf) (PDF) from the original on 2015-09-24.
4. "What is 3D Integration? - 3D InCites" (<https://www.3dincites.com/3d-incites-knowledge-portal/what-is-3d-integration/>). Archived (<https://web.archive.org/web/20141230233541/http://www.3dincites.com/3d-incites-knowledge-portal/what-is-3d-integration/>) from the original on 2014-12-30.
5. J. Knechtel, O. Sinanoglu, I. M. Elfadel, J. Lienig, C. C. N. Sze, "Large-Scale 3D Chips: Challenges and Solutions for Design Automation, Testing, and Trustworthy Integration" (http://www.ifte.de/mitarbeiter/lienig/IPSJ_Trans_2017.pdf) Archived (https://web.archive.org/web/20170807152805/http://www.ifte.de/mitarbeiter/lienig/IPSJ_Trans_2017.pdf) 2017-08-07 at the Wayback Machine, in *IPSJ Transactions on System LSI Design Methodology*, vol. 10, pp. 45–62, Aug. 2017
6. "INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS 2011 EDITION" (<https://web.archive.org/web/20141230234546/http://www.itrs.net/Links/2011itrs/2011Chapters/2011Design.pdf>) (PDF). Archived from the original (<http://www.itrs.net/Links/2011itrs/2011Chapters/2011Design.pdf>) (PDF) on 2014-12-30. Retrieved 2014-12-30.



Cut through a graphics card that uses High Bandwidth Memory (HBM), based on through-silicon via (TSV) 3D IC technology.

7. Lau, John & Tzeng, Pei-Jer & Lee, Ching-Kuan & Zhan, C. & Li, Ming & Cline, J. & Saito, K. & Hsin, Y. & Chang, P. & Chang, Yiu-Hsiang & Chen, J. & Chen, Shang-Chun & Wu, C. & Chang, H. & Chien, C. & Lin, C. & Ku, Tzu Kun & Lo, Robert & Kao, M.. (2013). Redistribution layers (RDLs) for 2.5D/3D IC integration. International Symposium on Microelectronics. 2013. 000434-000441. 10.4071/isom-2013-WA12.
8. "Xilinx and TSMC: Volume Production of 3D Parts" (<https://sem wiki.com/semiconductor-manufacturers/tsmc/2918-xilinx-and-tsmc-volume-production-of-3d-parts/>). 31 July 2023.
9. Lau, John H. (3 April 2019). *Heterogeneous Integrations* (<https://books.google.com/books?id=SCGQDwAAQBAJ>). Springer. ISBN 9789811372247.
10. "Advanced Packaging" (https://semiengineering.com/knowledge_centers/packaging/advanced-packaging/). *Semiconductor Engineering*.
11. "Next-Gen 3D Chip/Packaging Race Begins" (<https://semiengineering.com/next-gen-3d-chip-packaging-race-begins/>). 31 January 2022.
12. "Advanced 2.5D/3D Packaging Roadmap" (<https://sem wiki.com/semiconductor-manufacturers/tsmc/306329-advanced-2-5d-3d-packaging-roadmap/>). 31 July 2023.
13. "Comparing Samsung's 3D NAND with Traditional 3D ICs" (<https://www.3dincites.com/2013/08/comparing-samsungs-3d-nand-with-traditional-3d-ics/>). 2013-08-16.
14. James, Dick (2014). "3D ICs in the real world" (<https://www.researchgate.net/publication/271453642>). *25th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC 2014)*. pp. 113–119. doi:10.1109/ASMC.2014.6846988 (<https://doi.org/10.1109%2FASMC.2014.6846988>). ISBN 978-1-4799-3944-2. S2CID 42565898 (<https://api.semanticscholar.org/CorpusID:42565898>).
15. "Samsung starts production of 3D DDR4 DRAM modules" (<http://www.pcworld.com/article/2599700/samsung-starts-production-of-3d-ddr4-dram-modules.html>). 2014-08-27. Archived (<https://web.archive.org/web/20141231012258/http://www.pcworld.com/article/2599700/samsung-starts-production-of-3d-ddr4-dram-modules.html>) from the original on 2014-12-31.
16. Michallet, Jean-Eric. "CoolCube™: A True 3DVLSI Alternative to Scaling" (<https://www.3dincites.com/2015/03/coolcube-a-true-3dvlsi-alternative-to-scaling/>). *www.3DInCites.com*. Archived (<https://web.archive.org/web/20160122234232/http://www.3dincites.com/2015/03/coolcube-a-true-3dvlsi-alternative-to-scaling/>) from the original on January 22, 2016. Retrieved March 24, 2014.
17. von Trapp, Francoise (2015-03-16). "Monolithic 3D IC Heats Up at DATE 2015" (<https://www.3dincites.com/2015/03/monolithic-3d-ic-heats-up-at-date-2015/>). *3D InCites*. Archived (<https://web.archive.org/web/20150402031755/http://www.3dincites.com/2015/03/monolithic-3d-ic-heats-up-at-date-2015/>) from the original on April 2, 2015. Retrieved March 16, 2015.
18. Maestre Caro, A.; Travaly, Y.; Maes, G.; Borghs, G.; Armini, S. (2011). "Enabling Cu-Cu connection in (Dual) damascene interconnects by selective deposition of two different SAM molecules". *2011 IEEE International Interconnect Technology Conference*. pp. 1–3. doi:10.1109/IITC.2011.5940263 (<https://doi.org/10.1109%2FIITC.2011.5940263>). ISBN 978-1-4577-0503-8. S2CID 30235970 (<https://api.semanticscholar.org/CorpusID:30235970>).
19. Reif, Rafael; Tan, Chuan Seng; Fan, Andy; Chen, Kuan-Neng; Das, Shamik; Checka, Nisha (2002). "3-D Interconnects Using Cu Wafer Bonding: Technology and Applications" (<https://web.archive.org/web/20190715235342/https://pdfs.semanticscholar.org/e996/825ace86fa01a3492950f2d61fc780f515.pdf> (PDF)). *Advanced Metallization Conference*: 37–44. S2CID 2514964 (<https://api.semanticscholar.org/CorpusID:2514964>). Archived from the original (<https://pdfs.semanticscholar.org/e996/825ace86fa01a3492950f2d61fc780f515.pdf> (PDF)) on 15 July 2019. Retrieved 15 July 2019.

20. Yoon, Seung Wook; Ku, Jae Hoon; Suthiwongsunthorn, Nathapong; Marimuthu, Pandi Chelvam; Carson, Flynn (2009). "Fabrication and packaging of microbump interconnections for 3D TSV" (<https://ieeexplore.ieee.org/document/5306554>). *2009 IEEE International Conference on 3D System Integration*. pp. 1–5. doi:[10.1109/3DIC.2009.5306554](https://doi.org/10.1109/3DIC.2009.5306554) (<https://doi.org/10.1109%2F3DIC.2009.5306554>). ISBN 978-1-4244-4511-0. S2CID 11139525 (<https://api.semanticscholar.org/CorpusID:11139525>).
21. Real World Technologies. "3D Integration: A Revolution in Design". May 2, 2007. "3D Integration: A Revolution in Design" (<http://realworldtech.com/page.cfm?ArticleID=RWT050207213241&p=6>). 2 May 2007. Archived (<https://web.archive.org/web/20101222052241/http://realworldtech.com/page.cfm?ArticleID=RWT050207213241&p=6>) from the original on 2010-12-22. Retrieved 2011-03-18.
22. Chen, D.Y.; Chiou, W.C.; Chen, M.F.; Wang, T.D.; Ching, K.M.; Tu, H.J.; Wu, W.J.; Yu, C.L.; Yang, K.F.; Chang, H.B.; Tseng, M.H.; Hsiao, C.W.; Lu, Y.J.; Hu, H.P.; Lin, Y.C.; Hsu, C.S.; Shue, Winston S.; Yu, C.H. (2009). "Enabling 3D-IC foundry technologies for 28 nm node and beyond: Through-silicon-via integration with high throughput die-to-wafer stacking" (<https://ieeexplore.ieee.org/document/5424350>). *2009 IEEE International Electron Devices Meeting (IEDM)*. pp. 1–4. doi:[10.1109/IEDM.2009.5424350](https://doi.org/10.1109/IEDM.2009.5424350) (<https://doi.org/10.1109%2FIEDM.2009.5424350>). ISBN 978-1-4244-5639-0. S2CID 35980364 (<https://api.semanticscholar.org/CorpusID:35980364>).
23. Developer, Shed. "3D Processors, Stacking Core". September 20, 2005. "3D Processors, Stacking Cores" (<https://web.archive.org/web/20120316002322/http://www.devhardware.com/c/a/Computer-Processors/3D-Processor-Technology/>). Archived from the original (<http://www.devhardware.com/c/a/Computer-Processors/3D-Processor-Technology/>) on 2012-03-16. Retrieved 2012-10-29.,
24. Developer, Shed. "3D Processors, Stacking Core". September 20, 2005. "Page 2 - 3D Processors, Stacking Cores" (<https://web.archive.org/web/20110709015720/http://www.devhardware.com/c/a/Computer-Processors/3D-Processor-Technology/1/>). Archived from the original (<http://www.devhardware.com/c/a/Computer-Processors/3D-Processor-Technology/1/>) on 2011-07-09. Retrieved 2011-02-24.
25. Xiangyu Dong and Yuan Xie, "System-level Cost Analysis and Design Exploration for 3D ICs", Proc. of Asia and South Pacific Design Automation Conference, 2009, "Yuan Xie's 3D IC Research Page" (<https://web.archive.org/web/20100424202047/http://www.cse.psu.edu/~yuanxie/3d.html>). Archived from the original (<http://www.cse.psu.edu/~yuanxie/3d.html>) on 2010-04-24. Retrieved 2010-05-20.
26. "3D IC Technology Delivers The Total Package" "3D IC Technology Delivers the Total Package" (https://web.archive.org/web/20101031055533/http://electronicdesign.com/article/engineering-essentials/3d_ic_technology_delivers_the_total_package.aspx). Archived from the original (http://electronicdesign.com/article/engineering-essentials/3d_ic_technology_delivers_the_total_package.aspx) on 2010-10-31. Retrieved 2011-01-27. Electronic Design July 02, 2010
27. James J-Q Lu, Ken Rose, & Susan Vitkavage "3D Integration: Why, What, Who, When?" "3D Integration: Why, What, Who, When?" (https://web.archive.org/web/20080212184059/http://www.future-fab.com/documents.asp?d_ID=4396). Archived from the original (http://www.future-fab.com/documents.asp?d_ID=4396) on 2008-02-12. Retrieved 2008-01-22. Future Fab Intl. Volume 23, 2007
28. William J. Dally, "Future Directions for On-Chip Interconnection Networks" page 17, "Archived copy" (<http://www.ece.ucdavis.edu/~ocin06/talks/dally.pdf>) (PDF). Archived (<https://web.archive.org/web/20100612190558/http://www.ece.ucdavis.edu/~ocin06/talks/dally.pdf>) (PDF) from the original on 2010-06-12. Retrieved 2008-01-22. Computer Systems Laboratory Stanford University, 2006

29. Johnson, R Colin. "3-D chip stacks standardized". July 10, 2008. ["3-D chip stacks standardized"](https://web.archive.org/web/20120930192649/http://www.eetimes.com/electronics-news/4077835/3-D-chip-stacks-standardized) (<https://web.archive.org/web/20120930192649/http://www.eetimes.com/electronics-news/4077835/3-D-chip-stacks-standardized>). Archived from the original (<http://www.eetimes.com/electronics-news/4077835/3-D-chip-stacks-standardized>) on 2012-09-30. Retrieved 2014-05-15.
30. "3D-ICs and Integrated Circuit Security" "Archived copy" (http://www.tezzaron.com/about/papers/3D-ICs_and_Integrated_Circuit_Security.pdf) (PDF). Archived (https://web.archive.org/web/20080907062545/http://www.tezzaron.com/about/papers/3D-ICs_and_Integrated_Circuit_Security.pdf) (PDF) from the original on 2008-09-07. Retrieved 2008-02-08. Tezzaron Semiconductor, 2008
31. Dong Hyuk Woo, Nak Hee Seong, Dean L. Lewis, and Hsien-Hsin S. Lee. "An Optimized 3D-Stacked Memory Architecture by Exploiting Excessive, High-Density TSV Bandwidth". In Proceedings of the 16th International Symposium on High-Performance Computer Architecture, pp. 429–440, Bangalore, India, January 2010.
32. "Predicting the Performance of a 3D Processor-Memory Chip Stack" Jacob, P., McDonald, J.F. et al. Design & Test of Computers, IEEE Volume 22, Issue 6, Nov.–Dec. 2005 Page(s):540–547
33. A. Palesko, The Cost of 3D ICs, 3D InCites Knowledge Portal, January 9, 2015 "The Cost of 3D ICs" (<https://www.3dincites.com/2015/01/cost-3d-ics-2/>). 2015-01-09. Archived (<https://web.archive.org/web/20150109231028/http://www.3dincites.com/2015/01/cost-3d-ics-2/>) from the original on 2015-01-09. Retrieved 2015-01-09.
34. Robert Patti (2007). "Impact of Wafer-Level 3D Stacking on the Yield of ICs" (https://web.archive.org/web/20140517121030/http://www.future-fab.com/documents.asp?d_ID=4415). Future Fab International. No. 23. Archived from the original (http://www.future-fab.com/documents.asp?d_ID=4415) on 2014-05-17. Retrieved 2014-05-15.
35. Hsien-Hsin S. Lee and Krishnendu Chakrabarty, "Test challenges for 3D integrated circuits", IEEE Design and Test of Computers, Special issue on 3D IC Design and Test, vol. 26, no. 5, pp. 26–35, Sep/Oct 2009
36. " "EDA's big three unready for 3D chip packaging". EE Times Asia, October 25, 2007" (http://www.eetasia.com/ART_8800485666_480300_NT_fcb98510.HTM). Eetasia.com. Archived (https://web.archive.org/web/20080718203711/http://www.eetasia.com/ART_8800485666_480300_NT_fcb98510.HTM) from the original on July 18, 2008. Retrieved 2014-05-15.
37. D. H. Kim, S. Mukhopadhyay, S. K. Lim, "Through-silicon-via aware interconnect prediction and optimization for 3D stacked ICs", in Proc. of Int. Workshop Sys.-Level Interconn. Pred., 2009, pp. 85–92.
38. S. Borkar, "3D integration for energy efficient system design", in Proc. Design Autom. Conf., 2011, pp. 214–219.
39. " "3-D chip stacks standardized". EE Times November 7, 2008" (<https://web.archive.org/web/20120930192649/http://www.eetimes.com/electronics-news/4077835/3-D-chip-stacks-standardized>). Eetimes.com. 2014-05-09. Archived from the original (<http://www.eetimes.com/electronics-news/4077835/3-D-chip-stacks-standardized>) on September 30, 2012. Retrieved 2014-05-15.
40. " "SEMI International Standards Program Forms 3D Stacked IC Standards Committee". SEMI press release December 7, 2010" (http://www.semi.org/en/press/CTR_042145?id=higlights). Semi.org. 2010-12-07. Archived (https://web.archive.org/web/20140517121812/http://www.semi.org/en/press/CTR_042145?id=higlights) from the original on May 17, 2014. Retrieved 2014-05-15.

41. ""ADVANCED PACKAGING: 3D TSV Technologies Scenarios: Via First or Via Last? 2010 report". Yole report, 2010" (<https://web.archive.org/web/20140517114521/http://www.i-micronews.com/reports/3D-TSV-Technologies-Scenarios-Via-or-Via-Last-2010-report/108/>). I-micronews.com. 2010-01-01. Archived from the original (<http://www.i-micronews.com/reports/3D-TSV-Technologies-Scenarios-Via-or-Via-Last-2010-report/108/>) on 2014-05-17. Retrieved 2014-05-15.
42. "Si, glass interposers for 3D packaging: analysts' takes". Advanced Packaging August 10, 2010 (https://www.electroiq.com/index/display/packaging-article-display/3806393153/articles/advanced-packaging/packaging0/integration/tsv/2010/august/si_glass_interposers.html) Archived (https://web.archive.org/web/20110314192927/http://www.electroiq.com/index/display/packaging-article-display/3806393153/articles/advanced-packaging/packaging0/integration/tsv/2010/august/si_glass_interposers.html) March 14, 2011, at the Wayback Machine
43. J. Knechtel, I. L. Markov, J. Lienig, "Assembling 2D Blocks into 3D Chips" (http://www.ifte.de/mitarbeiter/lienig/TCAD_Feb2012_pp.228-241.pdf) Archived (https://web.archive.org/web/20160304073535/http://www.ifte.de/mitarbeiter/lienig/TCAD_Feb2012_pp.228-241.pdf) 2016-03-04 at the Wayback Machine, in IEEE Trans. on CAD of ICs and Systems, vol. 31, no. 2, pp. 228–241, Feb. 2012
44. S. Garg, D. Marculescu, "3D-GCP: An analytical model for the impact of process variations on the critical path delay distribution of 3D ICs", in Proc. Int. Symp. Quality Electron. Des., 2009, pp. 147–155
45. L. K. Scheffer, "CAD implications of new interconnect technologies", in Proc. Design Autom. Conf., 2007, pp. 576–581.
46. Moskowitz, Sanford L. (2016). *Advanced Materials Innovation: Managing Global Technology in the 21st century* (<https://books.google.com/books?id=2STRDAAAQBAJ&pg=PA165>). John Wiley & Sons. pp. 165–167. ISBN 978-0-470-50892-3.
47. U.S. patent 3,613,226 (<https://patents.google.com/patent/US3613226>)
48. U.S. patent 3,651,490 (<https://patents.google.com/patent/US3651490>)
49. "A Look At Trishul: Arm's First High-Density 3D Logic Stacked Test-Chip" (<https://fuse.wikichip.org/news/5588/a-look-at-trishul-arms-first-high-density-3d-logic-stacked-test-chip/>). WikiChip Fuse. 2021-06-11. Retrieved 2022-10-05.
50. Lobo, Savia (2018-12-13). "Intel unveils the first 3D Logic Chip packaging technology, 'Foveros', powering its new 10nm chips, 'Sunny Cove'" (<https://www.packtpub.com/en-es/learning/tech-news/intel-unveils-the-first-3d-logic-chip-packaging-technology-foveros-powering-its-new-10nm-chips-sunny-cove>). Packt Hub. Retrieved 2025-03-26.
51. "IBM is trying to solve all of computing's scaling issues with 5D electronic blood" (<https://arstechnica.com/gadgets/2015/11/5d-electronic-blood-ibms-secret-sauce-for-computers-with-biological-brain-like-efficiency/>). 7 November 2015.
52. Kada, Morihiro (2015). "Research and Development History of Three-Dimensional Integration Technology" (<https://web.archive.org/web/20211023015301/https://onecellonelightradio.files.wordpress.com/2018/11/three-dimensional-integration-of-semiconductors-2015.pdf>) (PDF). *Three-Dimensional Integration of Semiconductors: Processing, Materials, and Applications*. Springer. pp. 8–13. ISBN 978-3-319-18675-7. Archived from the original (<https://onecellonelightradio.files.wordpress.com/2018/11/three-dimensional-integration-of-semiconductors-2015.pdf>) (PDF) on 2021-10-23. Retrieved 2019-07-19.
53. Kawamura, S.; Sasaki, Nobuo; Iwai, T.; Nakano, M.; Takagi, M. (October 1983). "Three-dimensional CMOS IC's Fabricated by using beam recrystallization". *IEEE Electron Device Letters*. **4** (10): 366–368. Bibcode:1983IEDL....4..366K (<https://ui.adsabs.harvard.edu/abs/1983IEDL....4..366K>). doi:10.1109/EDL.1983.25766 (<https://doi.org/10.1109%2FEDL.1983.25766>). ISSN 0741-3106 (<https://search.worldcat.org/issn/0741-3106>). S2CID 35184408 (<https://api.semanticscholar.org/CorpusID:35184408>).

54. Kawamura, S.; Sasaki, N.; Iwai, T.; Mukai, R.; Nakano, M.; Takagi, M. (December 1983). "3-Dimensional SOI/CMOS IC's fabricated by beam recrystallization". *1983 International Electron Devices Meeting*. pp. 364–367. doi:10.1109/IEDM.1983.190517 (<https://doi.org/10.1109%2FIEDM.1983.190517>). S2CID 11689645 (<https://api.semanticscholar.org/CorpusID:11689645>).
55. Kawamura, S.; Sasaki, Nobuo; Iwai, T.; Mukai, R.; Nakano, M.; Takagi, M. (1984). *3-Dimensional Gate Array with Vertically Stacked Dual SOI/CMOS Structure Fabricated by Beam Recrystallization* (<https://ieeexplore.ieee.org/document/4480691>). pp. 44–45.
56. Garrou, Philip (6 August 2008). "Introduction to 3D Integration" (https://application.wiley-vch.de/books/sample/3527332650_c01.pdf) (PDF). *Handbook of 3D Integration: Technology and Applications of 3D Integrated Circuits*. Wiley-VCH. p. 4. doi:10.1002/9783527623051.ch1 (<https://doi.org/10.1002%2F9783527623051.ch1>). ISBN 978-3-527-62305-1.
57. Akasaka, Yoichi; Nishimura, T. (December 1986). "Concept and basic technologies for 3-D IC structure". *1986 International Electron Devices Meeting*. pp. 488–491. doi:10.1109/IEDM.1986.191227 (<https://doi.org/10.1109%2FIEDM.1986.191227>). S2CID 10393330 (<https://api.semanticscholar.org/CorpusID:10393330>).
58. Nishimura, T.; Inoue, Yasuo; Sugahara, K.; Kusunoki, S.; Kumamoto, T.; Nakagawa, S.; Nakaya, M.; Horiba, Yasutaka; Akasaka, Yoichi (December 1987). "Three dimensional IC for high performance image signal processor". *1987 International Electron Devices Meeting*. pp. 111–114. doi:10.1109/IEDM.1987.191362 (<https://doi.org/10.1109%2FIEDM.1987.191362>). S2CID 12936958 (<https://api.semanticscholar.org/CorpusID:12936958>).
59. Hayashi, Yoshihiro; Kunio, T.; Oyama, K.; Morimoto, M. (December 1989). "Three dimensional ICs, having four stacked active device layers". *International Technical Digest on Electron Devices Meeting*. pp. 837–840. doi:10.1109/IEDM.1989.74183 (<https://doi.org/10.1109%2FIEDM.1989.74183>). S2CID 113995937 (<https://api.semanticscholar.org/CorpusID:113995937>).
60. Yamazaki, K.; Itoh, Y.; Wada, A.; Morimoto, K.; Tomita, Y. (December 1990). "4-layer 3-D IC technologies for parallel signal processing". *International Technical Digest on Electron Devices*: 599–602. doi:10.1109/IEDM.1990.237127 (<https://doi.org/10.1109%2FIEDM.1990.237127>). S2CID 114856400 (<https://api.semanticscholar.org/CorpusID:114856400>).
61. Hayashi, Yoshihiro; Wada, S.; Kajiyana, K.; Oyama, K.; Koh, R.; Takahashi, S.; Kunio, T. (1990). "Fabrication of three-dimensional IC using 'cumulatively bonded IC' (CUBIC) technology". *Digest of Technical Papers. 1990 Symposium on VLSI Technology*: 95–96. doi:10.1109/VLSIT.1990.111025 (<https://doi.org/10.1109%2FVLSIT.1990.111025>). S2CID 27465273 (<https://api.semanticscholar.org/CorpusID:27465273>).
62. Fukushima, T.; Tanaka, T.; Koyanagi, Mitsumasa (2007). "Thermal Issues of 3D ICs" (<https://web.archive.org/web/20170516221221/http://sematech.org/meetings/archives/3d/8334/pres/Fukushima.pdf>) (PDF). SEMATECH. Tohoku University. Archived from the original (<http://sematech.org/meetings/archives/3d/8334/pres/Fukushima.pdf>) (PDF) on 16 May 2017. Retrieved 16 May 2017.
63. Tanaka, Tetsu; Lee, Kang Wook; Fukushima, Takafumi; Koyanagi, Mitsumasa (2011). "3D Integration Technology and Heterogeneous Integration" (<https://search.ieice.org/bin/index.php?category=C&lang=W&vol=J94-C&num=11&abst=j>). *IEICE Transactions on Electronics*. J94-C (11): 355–364. S2CID 62780117 (<https://api.semanticscholar.org/CorpusID:62780117>).
64. Takahashi, Kenji; Tanida, Kazumasa (2011). "Vertical Interconnection by ASET" (https://books.google.com/books?id=jtp_oFKsChgC&pg=PA339). *Handbook of 3D Integration, Volume 1: Technology and Applications of 3D Integrated Circuits*. John Wiley & Sons. p. 339. ISBN 978-3-527-62306-8.

65. Savastionk, S.; Siniaguine, O.; Korczynski, E. (2000). "Thru-silicon vias for 3D WLP". *Proceedings International Symposium on Advanced Packaging Materials Processes, Properties and Interfaces* (Cat. No.00TH8507). pp. 206–207. doi:10.1109/ISAPM.2000.869271 (<https://doi.org/10.1109%2FISAPM.2000.869271>). ISBN 0-930815-59-9. S2CID 110397071 (<https://api.semanticscholar.org/CorpusID:110397071>).
66. Lavanyashree, B.J. (August 2016). "3-Dimensional (3D) ICs: A Survey" (<https://web.archive.org/web/20190304081038/http://pdfs.semanticscholar.org/d85c/cb16cf71395410d98de0d6cb9a5c1740e487.pdf>) (PDF). *International Journal of Digital Application & Contemporary Research*. 5 (1). Archived from the original (<http://pdfs.semanticscholar.org/d85c/cb16cf71395410d98de0d6cb9a5c1740e487.pdf>) (PDF) on 2019-03-04.
67. Banerjee, Kaustav; Souri, Shukri J.; Kapur, Pawan; Saraswat, Krishna C. (2001). "3-D ICs: a novel chip design for improving deep-submicrometer interconnect performance and systems-on-chip integration". *Proceedings of the IEEE*. 89 (5): 602–633. doi:10.1109/5.929647 (<https://doi.org/10.1109%2F5.929647>). ISSN 0018-9219 (<https://search.worldcat.org/issn/0018-9219>). S2CID 5786126 (<https://api.semanticscholar.org/CorpusID:5786126>).
68. Imoto, T.; Matsui, M.; Takubo, C.; Akejima, S.; Kariya, T.; Nishikawa, T.; Enomoto, R. (2001). "Development of 3-Dimensional Module Package, "System Block Module" " (<https://www.tib.eu/en/search/id/BLCP%3ACN039662991/Development-of-3-Dimensional-Module-Package-System/>). *Electronic Components and Technology Conference* (51). Institute of Electrical and Electronics Engineers: 552–7. ISBN 0780370384.
69. Ramm, Peter (22 January 2016). "Fraunhofer EMFT: Our Early and Ongoing Work in 3D Integration" (<https://www.3dincites.com/2016/01/fraunhofer-emft-our-early-and-ongoing-work-in-3d-integration/>). *3D InCites*. Retrieved 22 September 2019.
70. Ramm, P.; Bollmann, D.; Braun, R.; Buchner, R.; Cao-Minh, U.; et al. (November 1997). "Three dimensional metallization for vertically integrated circuits". *Microelectronic Engineering*. 37–38: 39–47. doi:10.1016/S0167-9317(97)00092-0 (<https://doi.org/10.1016%2FS0167-9317%2897%2900092-0>). S2CID 22232571 (<https://api.semanticscholar.org/CorpusID:22232571>).
71. Macchiolo, A.; Andricek, L.; Moser, H. G.; Nisius, R.; Richter, R. H.; Weigell, P. (1 January 2012). "SLID-ICV Vertical Integration Technology for the ATLAS Pixel Upgrades". *Physics Procedia*. 37: 1009–1015. arXiv:1202.6497 (<https://arxiv.org/abs/1202.6497>). Bibcode:2012PhPro..37.1009M (<https://ui.adsabs.harvard.edu/abs/2012PhPro..37.1009M>). doi:10.1016/j.phpro.2012.02.444 (<https://doi.org/10.1016%2Fj.phpro.2012.02.444>). ISSN 1875-3892 (<https://search.worldcat.org/issn/1875-3892>). S2CID 91179768 (<https://api.semanticscholar.org/CorpusID:91179768>).
72. M.B. Kleiner, S.A. Kuehn, P. Ramm, W. Weber, IEEE Transactions on Components, Packaging, and Manufacturing Technology - Part B, Vol. 19, No. 4 (1996)
73. "HOME" (<https://web.archive.org/web/20071212192442/http://www.e-brains.org/>). Archived from the original on December 12, 2007.
74. Fan, Andy; Rahman, Adnan-ur; Reif, Rafael (February 2, 1999). "Copper Wafer Bonding". *Electrochemical and Solid-State Letters*. 2 (10): 534. doi:10.1149/1.1390894 (<https://doi.org/10.1149/1.1390894>). S2CID 98300746 (<https://api.semanticscholar.org/CorpusID:98300746>).
75. "Tezzaron Semiconductor: The Z Path Forward" (<http://www.tezzaron.com>). Tezzaron Semiconductor. Retrieved 19 July 2019.
76. "Six 3D designs precede 90% power-saving claims from Tezzaron - EE Times" (http://www.eetimes.com/document.asp?doc_id=1195090). Archived (https://web.archive.org/web/20141031023542/http://www.eetimes.com/document.asp?doc_id=1195090) from the original on 2014-10-31.

77. Cole, Bernard (22 May 2005). "Terrazon applies 3D stacking technology to 8051 MCU core" (<https://www.eetimes.com/terrazon-applies-3d-stacking-technology-to-8051-mcu-core/>). EETimes. Retrieved 10 August 2020.
78. B. Black, D. Nelson, C. Webb, and N. Samra, "3D Processing Technology and Its Impact on iA32 Microprocessors", in Proc. of Int. Conf. on Computer Design, pp. 316–318, 2004.
79. Steve Seguin (2008-09-16). "Seguin, Steve. "World's First Stacked 3D Processor Created". September 16, 2008" (<http://www.tomshardware.com/news/rochester-3d-processor,6369.html>). Tomshardware.com. Retrieved 2014-05-15.
80. "Science Daily. "3-D Computer Processor: 'Rochester Cube' Points Way To More Powerful Chip Designs". September 17, 2008" (<https://www.sciencedaily.com/releases/2008/09/080915105733.htm>). Sciencedaily.com. Archived (<https://web.archive.org/web/20140517121643/https://www.sciencedaily.com/releases/2008/09/080915105733.htm>) from the original on May 17, 2014. Retrieved 2014-05-15.
81. 3D-MAPS project webpage at Georgia Tech "3D-MAPS Many-Core Processor" (<https://web.archive.org/web/20150308155800/http://arch.ece.gatech.edu/research/3dmaps/3dmaps.html>). Archived from the original (<http://arch.ece.gatech.edu/research/3dmaps/3dmaps.html>) on 2015-03-08. Retrieved 2012-04-02.
82. "Centip3De: A 64-Core, 3D Stacked, Near-Threshold System" (<http://blaauw.engin.umich.edu/wp-content/uploads/sites/342/2017/11/486.pdf>) (PDF).
83. "System-in-Package (SiP)" (<https://web.archive.org/web/20100403003319/http://www.toshiba-a-components.com/ASIC/SiP.html>). Toshiba. Archived from the original (<http://www.toshiba-components.com/ASIC/SiP.html>) on 3 April 2010. Retrieved 3 April 2010.
84. "TOSHIBA COMMERCIALIZES INDUSTRY'S HIGHEST CAPACITY EMBEDDED NAND FLASH MEMORY FOR MOBILE CONSUMER PRODUCTS" (https://web.archive.org/web/20101123023805/http://www.toshiba.com/taec/news/press_releases/2007/memym_07_470.jsp). Toshiba. April 17, 2007. Archived from the original (http://www.toshiba.com/taec/news/press_releases/2007/memym_07_470.jsp) on November 23, 2010. Retrieved 23 November 2010.
85. "Hynix Surprises NAND Chip Industry" (https://www.koreatimes.co.kr/www/news/biz/2007/09/123_9628.html). Korea Times. 5 September 2007. Retrieved 8 July 2019.
86. "Toshiba Launches the Largest Density Embedded NAND Flash Memory Devices" (https://www.toshiba.co.jp/about/press/2008_08/pr0701.htm). Toshiba. 7 August 2008. Retrieved 21 June 2019.
87. "Toshiba Launches Industry's Largest Embedded NAND Flash Memory Modules" (https://www.toshiba.co.jp/about/press/2010_06/pr1701.htm). Toshiba. 17 June 2010. Retrieved 21 June 2019.
88. Kada, Morihiro (2015). "Research and Development History of Three-Dimensional Integration Technology" (<https://books.google.com/books?id=JaUvCwAAQBAJ&pg=PA15>). *Three-Dimensional Integration of Semiconductors: Processing, Materials, and Applications*. Springer. pp. 15–8. ISBN 978-3-319-18675-7.
89. "History: 2010s" (<https://web.archive.org/web/20210517040328/https://www.skhynix.com/eng/about/history2010.jsp>). SK Hynix. Archived from the original (<https://www.skhynix.com/eng/about/history2010.jsp>) on 17 May 2021. Retrieved 8 July 2019.
90. "Samsung Begins Mass Producing World's Fastest DRAM – Based on Newest High Bandwidth Memory (HBM) Interface" (<https://news.samsung.com/global/samsung-begins-mass-producing-worlds-fastest-dram-based-on-newest-high-bandwidth-memory-hbm-interface>). news.samsung.com.
91. Hruska, Joel (19 January 2016). "Samsung announces mass production of next-generation HBM2 memory – ExtremeTech" (<http://www.extremetech.com/extreme/221473-samsung-announces-mass-production-of-next-generation-hbm2-memory>). ExtremeTech.

92. Shilov, Anton (December 5, 2017). "Samsung Starts Production of 512 GB UFS NAND Flash Memory: 64-Layer V-NAND, 860 MB/s Reads" (<https://www.anandtech.com/show/12120/samsung-starts-production-of-512-gb-ufs-chips>). *AnandTech*. Retrieved 23 June 2019.
93. Manners, David (30 January 2019). "Samsung makes 1TB flash eUFS module" (<https://www.electronicsweekly.com/news/business/samsung-makes-1tb-flash-module-2019-01/>). *Electronics Weekly*. Retrieved 23 June 2019.
94. Tallis, Billy (October 17, 2018). "Samsung Shares SSD Roadmap for QLC NAND And 96-layer 3D NAND" (<https://www.anandtech.com/show/13497/samsung-shares-ssd-roadmap-for-qlc-nand-and-96layer-3d-nand>). *AnandTech*. Retrieved 27 June 2019.
95. "Intel unveils a groundbreaking way to make 3D chips" (<https://www.engadget.com/2018/12/12/intel-foverus-3d-chip/>). *Engadget*. 8 August 2019.
96. Smith, Ryan. "Micron's 232 Layer NAND Now Shipping: 1Tbit, 6-Plane Dies With 50% More I/O Bandwidth" (<https://www.anandtech.com/show/17509/microns-232-layer-nand-now-shipping>). www.anandtech.com. Retrieved 2022-08-03.

References

- JEDECが「DDR4」とTSVを使う「3DS」メモリ技術の概要を明らかに (http://pc.watch.impress.co.jp/docs/column/kaigai/20111107_488696.html) - 後藤弘茂のWeekly海外ニュース Impress Watch Co. (issued:2011-11-08, 2011-11-08)
- 貫通電極を用いたチップ積層技術の開発 (http://www.oki.com/jp/otr/2007/n211/pdf/211_r17.pdf) (Japanese) – oki technical review #211 Vol.74 #3 (issued:2007-10, 2011-11-08)
- TSV (Through Silicon Via:Si貫通電極) (<http://www.a-elpida.com/ja/technology/tsv.html>) Archived (<https://web.archive.org/web/20120425154900/http://www.a-elpida.com/ja/technology/tsv.html>) 2012-04-25 at the Wayback Machine (Japanese) – Akita Elpida Memory, inc (2011-11-08)

Further reading

- Philip Garrou, Christopher Bower, Peter Ramm: *Handbook of 3D Integration, Technology and Applications of 3D Integrated Circuits* Vol. 1 and Vol. 2, Wiley-VCH, Weinheim 2008, ISBN 978-3-527-32034-9.
- Yuan Xie, Jason Cong, Sachin Sapatnekar: *Three-Dimensional Integrated Circuit Design: Eda, Design And Microarchitectures*, Publisher: Springer, ISBN 1-4419-0783-1, ISBN 978-1-4419-0783-7, 978-1441907837, Publishing Date: Dec. 2009.
- Philip Garrou, Mitsumasa Koyanagi, Peter Ramm: *Handbook of 3D Integration, 3D Process Technology* Vol. 3, Wiley-VCH, Weinheim 2014, ISBN 978-3-527-33466-7.
- Paul D. Franzon, Erik Jan Marinissen, Muhamad S. Bakir, Philip Garrou, Mitsumasa Koyanagi, Peter Ramm: *Handbook of 3D Integration: "Design, Test, and Thermal Management of 3D Integrated Circuits"*, Vol. 4, Wiley-VCH, Weinheim 2019, ISBN 978-3-527-33855-9.

External links

- Eurononymous (2007-05-02). "3D Integration: A Revolution in Design" (<http://realworldtech.com/page.cfm?ArticleID=RWT050207213241>). Real World Technologies. Retrieved 2014-05-15.

- Semiconductors (2006). "Mapping progress in 3D IC integration" (https://archive.today/20130131192505/http://sst.pennnet.com/Articles/Article_Display.cfm?ARTICLE_ID=254615). Solid State Technology. Archived from the original (http://sst.pennnet.com/Articles/Article_Display.cfm?ARTICLE_ID=254615) on January 31, 2013. Retrieved 2014-05-15.
- Peter Ramm; et al. (2010-09-16). "3D Integration technology: Status and application development". *2010 Proceedings of ESSCIRC*. IEEE. pp. 9–16. doi:10.1109/ESSCIRC.2010.5619857 (<https://doi.org/10.1109%2FESSCIRC.2010.5619857>). hdl:11250/2463188 (<https://hdl.handle.net/11250%2F2463188>). ISBN 978-1-4244-6664-1. S2CID 1239311 (<https://api.semanticscholar.org/CorpusID:1239311>).
- Mingjie Lin; Abbas El Gamal; Yi-chang Lu & Simon Wong (2006-02-22). "Performance benefits of monolithically stacked 3D-FPGA". *Proceedings of the 2006 ACM/SIGDA 14th international symposium on Field programmable gate arrays*. Vol. 26. p. 113. doi:10.1145/1117201.1117219 (<https://doi.org/10.1145%2F1117201.1117219>). ISBN 1595932925. S2CID 7818893 (<https://api.semanticscholar.org/CorpusID:7818893>).
- "Joint Project for Mechanical Qualification of Next Generation High Density Package-on-Package (PoP) with Through Mold Via Technology" (<http://www.amkor.com/index.cfm?objectid=437C1C10-5056-AA0A-E2626E3C762307FF>). Retrieved 2014-05-15.
- "Advancements in Stacked Chip Scale Packaging (S-CSP), Provides System-in-a-Package Functionality for Wireless and Handheld Applications" (<http://www.amkor.com/index.cfm?objectid=8A586C38-F4D0-4CD8-7AD50C7BA98E3417>). Retrieved 2014-05-15.
- Smith, Lee (July 6, 2010). "Achieving the 3rd Generation From 3D Packaging to 3D IC Architectures" (<http://www.amkor.com/index.cfm?objectid=9A9D6655-E2FD-713C-AE2761D3B9EFB250>). *Future Fab International*. Amkor Technology. Retrieved 2014-05-15.
- "Factors Affecting Electromigration and Current Carrying Capacity of Flip Chip and 3D IC Interconnects" (<http://www.amkor.com/index.cfm?objectid=AC1A3242-F9FC-76B7-D0D77E166839BF9B>). Retrieved 2014-05-15.
- "Evaluation for UV Laser Dicing Process and its Reliability for Various Designs of Stack Chip Scale Package" (<http://www.amkor.com/index.cfm?objectid=B173A68F-A106-490B-70B5650D6ACFB6DE>). Retrieved 2014-05-15.
- "High Density PoP (Package-on-Package) and Package Stacking Development" (<http://www.amkor.com/index.cfm?objectid=CAB8D948-F34D-8E73-72961B951D0877D7>). Retrieved 2014-05-15.
- "3D Interconnect Technology Coming to Light" (<https://web.archive.org/web/20081203230736/http://www.edn.com/article/CA410764.html>). EDN. 2004. Archived from the original (<http://www.edn.com/article/CA410764.html>) on 2008-12-03. Retrieved 2008-01-22.
- "Three-dimensional SoCs perform for future" (<http://www.eedesign.com/article/showArticle.jhtml?articleId=17408808>). EE Design. 2003. Retrieved 2014-05-15.
- "MagnaChip, Tezzaron form partnership for 3D chips" (<https://archive.today/20130121221433/http://www.eetimes.com/news/semi/showArticle.jhtml?articleID=53700960>). EE Times. 2004. Archived from the original (<http://www.eetimes.com/news/semi/showArticle.jhtml?articleID=53700960>) on 2013-01-21.
- "Matrix preps 64-Mbyte write-once memory" (<https://web.archive.org/web/20080515200141/http://www.eetimes.com/story/OEG20011219S0041>). EE Times. 2001. Archived from the original (<http://www.eetimes.com/story/OEG20011219S0041>) on 2008-05-15. Retrieved 2014-05-15.
- "Samsung starts mass producing first 3D vertical NAND flash, August 2013" (<https://web.archive.org/web/20130818084414/http://www.electroiq.com/articles/sst/2013/08/monolithic-3d-is-now-in-production-samsung-starts-mass-producing-first-3d-vertical-nand-flash.html>). Electroiq.com. 2013-08-06. Archived from the original (<https://www.electroiq.com/articles/sst/2013/08/monolithic-3d-is-now-in-production-samsung-starts-mass-producing-first-3d-vertical-nand-flash.html>) on 2013-08-18. Retrieved 2014-05-15.

- "CEA Leti placed monolithic 3D as the next generation technology as alternative to dimension scaling, August 2013" (<https://web.archive.org/web/20130819075724/http://www.electroiq.com/articles/sst/2013/08/monolithic-3d-is-now-on-the-roadmap-for-2019.html>). Electroiq.com. Archived from the original (<https://www.electroiq.com/articles/sst/2013/08/monolithic-3d-is-now-on-the-roadmap-for-2019.html>) on 2013-08-19. Retrieved 2014-05-15.
- "3D integration: A status report" (<https://archive.today/20130122085527/http://www.electroiq.com/index/display/article-display/366226/articles/advanced-packaging/packaging0/equipment/die-bonding-equipment/3d-integration-a-status-report.html>). 2009. Archived from the original (<https://www.electroiq.com/index/display/article-display/366226/articles/advanced-packaging/packaging0/equipment/die-bonding-equipment/3d-integration-a-status-report.html>) on 2013-01-22. Retrieved 2011-01-21.
- Deepak C. Sekar & Zvi Or-Bach. "Monolithic 3D-ICs with Single Crystal Silicon Layers" (http://www.monolithic3d.com/uploads/6/0/5/5/6055488/ieee_3d_ic_conference_invited_monolithic_3d.pdf) (PDF). Retrieved 2014-05-15.
- "Global 3D Chips/3D IC Market to Reach US\$5.2 Billion by 2015" (https://web.archive.org/web/20100901192936/http://www.prweb.com/releases/3D_chips/3D_IC/prweb4400904.htm). PRWeb. 2010. Archived from the original (http://www.prweb.com/releases/3D_chips/3D_IC/prweb4400904.htm) on September 1, 2010. Retrieved 2014-05-15.
- "Samsung Develops 30nm-class 32GB Green DDR3 for Next-generation Servers, Using TSV Package Technology" (http://www.samsung.com/us/business/semiconductor/newsView.do?news_id=1231). Samsung.com. 2011. Retrieved 2014-05-15.
- "How Might 3-D ICs Come Together?" (https://web.archive.org/web/20100304084239/http://www.semiconductor.net/article/202251-How_Might_3_D_ICs_Come_Together_.php). Semiconductor International. 2008. Archived from the original on 2010-03-04. Retrieved 2009-06-11.
- "Three-Dimensional ICs Solve the Interconnect Paradox" (<https://web.archive.org/web/20080212184730/http://www.semiconductor.net/article/CA604503.html>). Semiconductor International. 2005. Archived from the original on 2008-02-12. Retrieved 2008-01-22.
- "Ziptronix, Raytheon Prove 3-D Integration of 0.5 μm CMOS Device" (<https://web.archive.org/web/20071106080914/http://www.semiconductor.net/article/CA6431663.html>). Semiconductor International. 2007. Archived from the original on 2007-11-06. Retrieved 2008-01-22.
- Peter Ramm; Armin Klumpp; Josef Weber; Maaike Taklo (2010). "3D System-on-Chip Technologies for More than Moore Systems". *Journal of Microsystem Technologies*. **16** (7): 1051–1055. Bibcode:2010MiTec..16.1051R (<https://ui.adsabs.harvard.edu/abs/2010MiTec..16.1051R>). doi:10.1007/s00542-009-0976-1 (<https://doi.org/10.1007%2Fs00542-009-0976-1>). S2CID 55824967 (<https://api.semanticscholar.org/CorpusID:55824967>).
- Philip Garrou, James Lu & Peter Ramm (2012). "Chapter 15" (<http://www.wiley-vch.de/publications/books/newTitles201201/3-527-32646-4/?sID=p2qlnooj68su7htl8qrrc2qjt3>). *Three-Dimensional Integration*. Wiley-VCH. Retrieved 2014-05-15. {{cite book}}: |work= ignored (help)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Three-dimensional_integrated_circuit&oldid=1289722288"



Emitter-coupled logic

In electronics, **emitter-coupled logic** (ECL) is a high-speed integrated circuit bipolar transistor logic family. ECL uses a bipolar junction transistor (BJT) differential amplifier with single-ended input and limited emitter current to avoid the saturated (fully on) region of operation and the resulting slow turn-off behavior.^[4] As the current is steered between two legs of an emitter-coupled pair, ECL is sometimes called *current-steering logic* (CSL),^[5] *current-mode logic* (CML)^[6] or *current-switch emitter-follower* (CSEF) logic.^[7]

In ECL, the transistors are never in saturation, the input and output voltages have a small swing (0.8 V), the input impedance is high and the output impedance is low. As a result, the transistors change states quickly, gate delays are low, and the fanout capability is high.^[8] In addition, the essentially constant current draw of the differential amplifiers minimizes delays and glitches due to supply-line inductance and capacitance, and the complementary outputs decrease the propagation time of the whole circuit by reducing inverter count.

ECL's major disadvantage is that each gate continuously draws current, which means that it requires (and dissipates) significantly more power than those of other logic families, especially when quiescent.

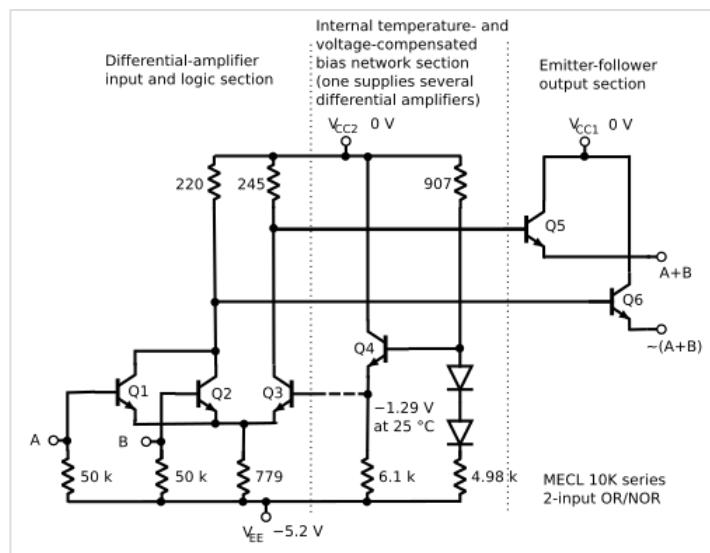
The equivalent of emitter-coupled logic made from FETs is called source-coupled logic (SCFL).^[9]

A variation of ECL in which all signal paths and gate inputs are differential is known as differential current switch (DCS) logic.^[10]

History

ECL was invented in August 1956 at IBM by Hannon S. Yourke.^{[12][13]} Originally called *current-steering logic*, it was used in the Stretch, IBM 7090, and IBM 7094 computers.^[11] The logic was also called a current-mode circuit.^[14] It was also used to make the IBM Advanced Solid Logic Technology (ASLT) circuits in the IBM 360/91.^{[15][16][17]}

Yourke's current switch was a differential amplifier whose input logic levels were different from the output logic levels. "In current mode operation, however, the output signal consists of voltage levels which vary about a reference level different from the input reference level."^[18] In Yourke's design, the



Motorola ECL 10,000 basic OR/NOR gate circuit diagram from 1972.^{[1][2]:5} Note the Q5 and Q6 emitters coupled to the output and the negative-power supply.^{[a][3]}

two logic reference levels differed by 3 volts. Consequently, two complementary versions were used: an NPN version and a PNP version. The NPN output could drive PNP inputs, and vice versa. "The disadvantages are that more different power supply voltages are needed, and both pnp and npn transistors are required."^[11]

Instead of alternating NPN and PNP stages, another coupling method employed Zener diodes and resistors to shift the output logic levels to be the same as the input logic levels.^[19]

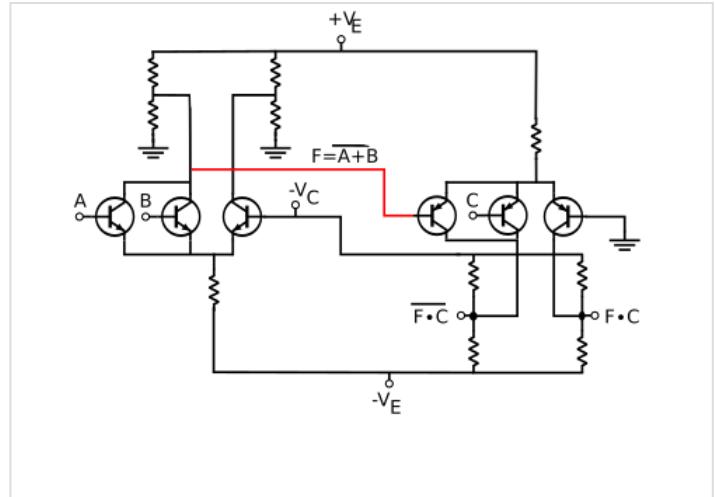
Beginning in the early 1960s, ECL circuits

were implemented on monolithic integrated circuits. They consisted of a differential-amplifier input stage to perform logic followed by an emitter-follower stage to drive outputs and shift the output voltages so they will be compatible with the inputs. The emitter-follower output stages could also be used to perform wired-or logic.

Motorola introduced their first digital monolithic integrated circuit line, MECL I, in 1962.^[20] Motorola developed several improved series, with MECL II in 1966, MECL III in 1968 with 1-nanosecond gate propagation time and 300 MHz flip-flop toggle rates, and the 10,000 series (with lower power consumption and controlled edge speeds) in 1971.^[21] The MECL 10H family was introduced in 1981.^[2] Fairchild introduced the F100K family in 1975.^{[22][23]}

The ECLinPS ("ECL in picoseconds") family was introduced in 1987.^[24] ECLinPS has 500 ps single-gate delay and 1.1 GHz flip-flop toggle frequency.^[25] The ECLinPS family parts are available from multiple sources, including Arizona Microtek, Micrel (subsequently acquired by Microchip Technology Inc.), National Semiconductor, and ON Semiconductor.^[26]

The high power consumption of ECL meant that it has been used mainly when high speed is a vital requirement. Older high-end mainframe computers, such as the Enterprise System/9000 members of IBM's ESA/390 computer family, used ECL,^[27] as did the Cray-1,^[28] and first-generation Amdahl mainframes. (Current IBM mainframes use CMOS).^[29] Beginning in 1975, Digital Equipment Corporation's highest performance processors were all based on multi-chip ECL CPUs—from the ECL KL10 through the ECL VAX 8000 and finally the VAX 9000. By 1991, the CMOS NVAX was launched, which offered comparable performance to the VAX 9000 despite costing 1/25 as much and consuming considerably less power.^[30] The MIPS R6000 computers also used ECL. Some of these computer designs used ECL gate arrays.



Yourke's current switch (around 1955)^[11]

Implementation

ECL is based on an emitter-coupled (long-tailed) pair, shaded red in the figure on the right. The left half of the pair (shaded yellow) consists of two parallel-connected input transistors T1 and T2 (an exemplary two-input gate is considered) implementing NOR logic. The base voltage of the right transistor T3 is held

fixed by a reference voltage source, shaded light green: the voltage divider with a diode thermal compensation (R_1 , R_2 , D_1 and D_2) and sometimes a buffering emitter follower (not shown on the picture); thus the emitter voltages are kept relatively steady. As a result, the common emitter resistor R_E acts nearly as a current source. The output voltages at the collector load resistors R_{C1} and R_{C3} are shifted and buffered to the inverting and non-inverting outputs by the emitter followers T_4 and T_5 (shaded blue). The output emitter resistors R_{E4} and R_{E5} do not exist in all versions of ECL. In some cases $50\ \Omega$ line termination resistors connected between the bases of the input transistors and -2 V of a driven gate act as emitter resistors of the driving gate.^[31]

Operation

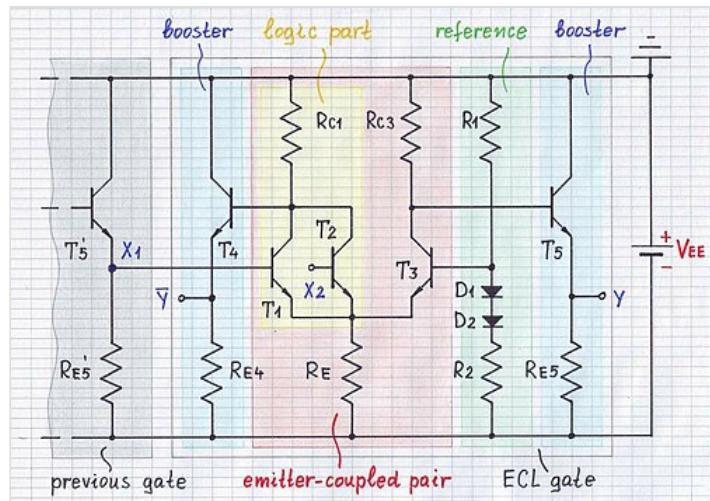
The ECL circuit operation is considered below with assumption that the input voltage is applied to T_1 base, while T_2 input is unused or a logical "0" is applied.

During the transition, the core of the circuit – the emitter-coupled pair (T_1 and T_3) – acts as a differential amplifier with single-ended input. The long-tail current source (R_E) sets the total current flowing through the two legs of the pair. The input voltage controls the current flowing through the transistors by sharing it between the two legs, steering it all to one side when not near the switching point. The gain is higher than at the end states (see below) and the circuit switches quickly.

At low input voltage (logical "0") or at high input voltage (logical "1") the differential amplifier is overdriven. The transistor (T_1 or T_3) is cutoff and the other (T_3 or T_1) is in active linear region acting as a common-emitter stage with emitter degeneration that takes all the current, starving the other cutoff transistor.

The active transistor is loaded with the relatively high emitter resistance R_E that introduces a significant negative feedback (emitter degeneration). To prevent saturation of the active transistor so that the diffusion time that slows the recovery from saturation will not be involved in the logic delay,^[4] the emitter and collector resistances are chosen such that at maximum input voltage some voltage is left across the transistor. The residual gain is low ($K = R_C/R_E < 1$). The circuit is insensitive to the input voltage variations and the transistor stays firmly in active linear region. The input resistance is high because of the series negative feedback.

The cutoff transistor breaks the connection between its input and output. As a result, its input voltage does not affect the output voltage. The input resistance is high again since the base-emitter junction is cutoff.



The picture represents a typical ECL circuit diagram based on Motorola's MECL. In this schematic, transistor T_5' represents the output transistor of a previous ECL gate that provides a logic signal to input transistor T_1 of an OR/NOR gate whose other input is at T_2 and has outputs Y and \bar{Y} . Additional pictures illustrate the circuit operation by visualizing the voltage relief and current topology at low input voltage (logical "0"), during the transition and at high input voltage (logical "1").

Characteristics

Other noteworthy characteristics of the ECL family include the fact that the large current requirement is approximately constant, and does not depend significantly on the state of the circuit. This means that ECL circuits generate relatively little power noise, unlike other logic types which draw more current when switching than quiescent. In cryptographic applications, ECL circuits are also less susceptible to side channel attacks such as differential power analysis.

The propagation time for this arrangement can be less than a nanosecond, including the signal delay getting on and off the IC package. Some type of ECL has always been the fastest logic family.^{[32][33]}

Radiation hardening: While normal commercial-grade chips can withstand 100 gray (10 krad), many ECL devices are operational after 100,000 gray (10 Mrad).^[34]

Power supplies and logic levels

ECL circuits usually operate with negative power supplies (positive end of the supply is connected to ground).^{[2]:5} (Other logic families ground the negative end of the power supply.) This is done mainly to minimize the influence of the power supply variations on the logic levels. ECL is more sensitive to noise on the V_{CC} and is relatively immune to noise on V_{EE} .^[35] Because ground should be the most stable voltage in a system, ECL is specified with a positive ground. In this connection, when the supply voltage varies, the voltage drops across the collector resistors change slightly (in the case of emitter constant current source, they do not change at all). As the collector resistors are firmly "tied up" to ground, the output voltages "move" slightly (or not at all). If the negative end of the power supply was grounded, the collector resistors would be attached to the positive rail.^{[2]:5} As the constant voltage drops across the collector resistors change slightly (or not at all), the output voltages follow the supply voltage variations and the two circuit parts act as constant current level shifters. In this case, the voltage divider R1-R2 compensates the voltage variations to some extent. The positive power supply has another disadvantage — the output voltages will vary slightly (± 0.4 V) against the background of high constant voltage (+3.9 V). Another reason for using a negative power supply is protection of the output transistors from an accidental short circuit developing between output and ground^[36] (but the outputs are not protected from a short circuit with the negative rail).

The value of the supply voltage is chosen so that sufficient current flows through the compensating diodes D1 and D2 and the voltage drop across the common emitter resistor R_E is adequate.

ECL circuits available on the open market usually operated with logic levels incompatible with other families. This meant that interoperation between ECL and other logic families, such as the popular TTL family, required additional interface circuits. The fact that the high and low logic levels are relatively close meant that ECL suffers from small noise margins, which can be troublesome.

At least one manufacturer, IBM, made ECL circuits for use in the manufacturer's own products. The power supplies were substantially different from those used in the open market.^[27]

PECL

Positive emitter-coupled logic, also called **pseudo-ECL**, (PECL) is a further development of ECL using a positive 5 V supply instead of a negative 5.2 V supply.^[38] Low-voltage positive emitter-coupled logic (LVPECL) is a power-optimized version of PECL, using a positive 3.3 V instead of 5 V supply.

PECL and LVPECL are differential-signaling systems and are mainly used in high-speed and clock-distribution circuits.

A common misconception is that PECL devices are slightly different from ECL devices. In fact, every ECL device is also a PECL device.^[39]

Logic levels for Motorola ECL/PEC comparison^{[37][2]:5}

Type	V_{ee}	V_{low}	V_{high}	V_{swing}	V_{cc}	V_{cm}
ECL	-5.2 V	-1.75 V	-0.9 V	-0.85 V	GND	
PECL	GND	3.4 V	4.2 V	0.8 V	5.0 V	
LVPECL	GND	1.6 V	2.4 V	0.8 V	3.3 V	2.0 V

Notes

- a. Motorola ECL devices run off of a negative power supply. This "reversed" arrangement used to avoid noise that might be induced to the collector and emitter lines. It also had very unusual logic levels (see comparison): -1.75 V for LOW (0) and -0.9 V for HIGH (1).^{[2]:5} It was possible to design circuits by using ECL with positive sources (see PECL - Positive ECL).^{[2]:18}

References

1. Original drawing based on Blood Jr., William R. (1972). *MECL System Design Handbook* (http://www.bitsavers.org/components/motorola/_dataBooks/1972_Motorola_MECL_System_Design_Handbook_2ed.pdf) (PDF) (2nd ed.). Motorola Semiconductor Products. p. 1 – via Bitsavers.
2. *TND309: General Information for MECL 10H and MECL 10K* (https://web.archive.org/web/20150708060625/http://www.onsemi.com/pub_link/Collateral/TND309-D.PDF) (PDF). ON Semiconductor: Semiconductor Components Industries. 2002. p. 2. TND309/D. Archived from the original (http://www.onsemi.com/pub_link/Collateral/TND309-D.PDF) (PDF) on July 8, 2015.
3. *Designing with PECL (ECL at +5.0V), The High Speed Solution for the CMOS/TTL Designer* (<https://www.onsemi.com/download/application-notes/pdf/an1406-d.pdf>) (PDF) (2 ed.). Onsemi. September 1999.
4. Lawless, Brian. "Unit4: ECL Emitter Coupled Logic" (<http://www.physics.dcu.ie/~bl/digi/unitd04.pdf>) (PDF). *Fundamental Digital Electronics*.
5. Kumar, Anand (2008). *Pulse and Digital Circuits* (https://books.google.com/books?id=ECeO_bhzCiLIC&pg=RA2-PA472). PHI Learning. p. 472. ISBN 978-81-203-3356-7.
6. Stonham, T. J. (1996). *Digital Logic Techniques: Principles and Practice* (<https://books.google.com/books?id=UE6vFEnGP2kC&pg=PA173>). Taylor & Francis. p. 173. ISBN 978-0-412-54970-0.
7. Tummala, Rao R. (2001). *Fundamentals of Microsystems Packaging* (<https://books.google.com/books?id=P93ZrOWHIOOC&pg=PA930>). McGraw-Hill. p. 930. ISBN 978-0-07-137169-8.

8. Mims, Forrest M. (2000). *The Forrest Mims Circuit Scrapbook* (<https://books.google.com/books?id=STzitya5iwgC&pg=PA115>). Vol. 2. Newnes. p. 115. ISBN 978-1-878707-48-2.
9. Fisher, Dennis; Bahl, I.J. (1995). *Gallium Arsenide IC Applications Handbook* (<https://books.google.com/books?id=KSKJ56kvcSYC&pg=PA61>). Vol. 1. Elsevier. p. 61. ISBN 978-0-12-257735-2.
10. Eichelberger, E.B.; Bello, S.E. (May 1991). "Differential Current Switch – High performance at low power" (<http://domino.watson.ibm.com/tchjr/journalindex.nsf/0/af4fa2f7f17243c485256bfa0067fab9?OpenDocument>). *IBM Journal of Research and Development*. **35** (3): 313–320. doi:[10.1147/rd.353.0313](https://doi.org/10.1147/rd.353.0313) (<https://doi.org/10.1147%2Frd.353.0313>).
11. E. J. Rymaszewski; et al. (1981). "Semiconductor Logic Technology in IBM" (<https://web.archive.org/web/20080705164759/http://researchweb.watson.ibm.com/journal/rd/255/ibmrd2505W.pdf>) (PDF). *IBM Journal of Research and Development*. **25** (5): 607–608. doi:[10.1147/rd.255.0603](https://doi.org/10.1147/rd.255.0603) (<https://doi.org/10.1147%2Frd.255.0603>). ISSN 0018-8646 (<https://search.worldcat.org/issn/0018-8646>). Archived from the original (<https://researchweb.watson.ibm.com/journal/rd/255/ibmrd2505W.pdf>) (PDF) on July 5, 2008. Retrieved August 27, 2007.
12. Early Transistor History at IBM (http://semiconductormuseum.com/Transistors/IBM/OralHistories/Yourke/Yourke_Index.htm).
13. Yourke, Hannon S. (October 1956), *Millimicrosecond non-saturating transistor switching circuits* (<http://archive.computerhistory.org/resources/text/IBM/Stretch/pdfs/06-11/102634289.pdf>) (PDF), Stretch Circuit Memo # 3. Yourke's circuits used commercial transistors and had an average gate delay of 12 ns.
14. Roehr, William D.; Thorpe, Darrell, eds. (1963). *High-Speed Switching Transistor Handbook* (<https://archive.org/details/High-speedSwitchingHandbook>). Motorola., p. 37.
15. Pugh, Emerson W.; Johnson, Lyle R.; Palmer, John H. (2003). *IBM's 360 and Early 370 Systems*. MIT Press. p. 108. ISBN 0262517205.
16. Langdon, J.L.; VanDerveer, E.J. (1967). "Design of a High-Speed Transistor for the ASLT Current Switch" (<http://www.research.ibm.com/journal/rd/111/ibmrd1101G.pdf>) (PDF). *IBM Journal of Research and Development*. **11**: 69–73. doi:[10.1147/rd.111.0069](https://doi.org/10.1147/rd.111.0069) (<https://doi.org/10.1147%2Frd.111.0069>).
17. "Logic Blocks Automated Logic Diagrams SLT, SLD, ASLT, MST" (http://bitsavers.trailing-edge.com/pdf/ibm/logic/SY22-2798-2_LoicBlocks_AutomatedLogicDiagrams_SLT,SLD,ASLT,MST_TO_Oct71.pdf) (PDF). IBM. pp. 1–10. Retrieved September 11, 2015 – via Bitsavers.
18. Roehr & Thorpe 1963, p. 39
19. Roehr & Thorpe 1963, pp. 40, 261
20. Blood Jr., William R. (1988) [1980]. *MECL System Design Handbook* (<https://web.archive.org/web/20041010110305/http://www.onsemi.com/pub/Collateral/HB205-D.PDF>) (PDF) (4th ed.). Motorola Semiconductor Products, republished by On Semiconductor. p. vi. Archived from the original (<http://www.onsemi.com/pub/Collateral/HB205-D.PDF>) (PDF) on October 10, 2004.
21. Blood Jr., William R. (October 1971). *MECL System Design Handbook* (http://www.bitsavers.org/components/motorola/_dataBooks/1971_Motorola_MECL_System_Design_Handbook_1ed_Oct71.pdf) (PDF) (1st ed.). Motorola. pp. vi–vii – via Bitsavers.
22. Hively, J. W.; Muller, H. H.; Owens, W. K. "F100K, A Standard Family of Subnanosecond ECL" (<https://www.computerhistory.org/collections/catalog/102798005>). *Fairchild Journal of Semiconductor Progress*. **3** (3): 16–21. Retrieved October 30, 2024.
23. "Thomas A. Longo: Department of Physics and Astronomy: Purdue University" (<https://www.physics.purdue.edu/alumni/distalum/longo.html>). www.physics.purdue.edu.
24. Maini, Anil K. (2007). *Digital Electronics: Principles, Devices and Applications* (<https://books.google.com/books?id=Ljsr7UA83ScC>). John Wiley & Sons. p. 148. ISBN 9780470510513.
25. "High Performance ECL Data: ECLinPS and ECLinPS Lite" (<ftp://ece.buap.mx/pub/manuals/High%20Performance%20ECL%20Data.pdf>) (PDF). 1996. p. iii.

26. ECL Logic Manufacturers – "Emitter Coupled Logic" (http://www.interfacebus.com/ECL_Logic_Manufacturers.html).
27. A. E. Barish; et al. (1992). "Improved performance of IBM Enterprise System/9000 bipolar logic chips" (<http://domino.watson.ibm.com/tchjr/journalindex.nsf/0/3f9af3392b4530f985256bfa0067fa2e?OpenDocument>). *IBM Journal of Research and Development*. **36** (5): 829–834. doi:[10.1147/rd.365.0829](https://doi.org/10.1147/rd.365.0829) (<https://doi.org/10.1147%2Frd.365.0829>).
28. Russell, R.M. (1978). "The CRAY1 computer system" (<http://www.eecg.toronto.edu/~moshovos/ACA05/read/cray1.pdf>) (PDF). *Communications of the ACM*. **21** (1): 63–72. doi:[10.1145/359327.359336](https://doi.org/10.1145/359327.359336) (<https://doi.org/10.1145%2F359327.359336>). S2CID 28752186 (<https://api.semanticscholar.org/CorpusID:28752186>). Retrieved April 27, 2010.
29. "IBM zEnterprise System Technical Introduction" (<https://web.archive.org/web/20131103060023/http://www.redbooks.ibm.com/redpieces/pdfs/sg248050.pdf>) (PDF). August 1, 2013. Archived from the original (<http://www.redbooks.ibm.com/redpieces/pdfs/sg248050.pdf>) (PDF) on November 3, 2013.
30. Supnik, Bob. "Raven: Introduction: The ECL Conundrum" (<http://simh.trailing-edge.com/seminari/raven.html>). "Raven was started in 1988... Raven was a simplified VAX design with a single chip CPU and a single chip FPU. Implemented in Fujitsu's ECL standard cells, it was intended to run at 250Mhz and deliver 50 "VUPS" ... Power dissipation would have been a startling (for the day) 150W."
31. Blood Jr. 1972, p. 3
32. John F. Wakerly. Supplement to Digital Design Principles and Practices. Section "ECL: Emitter-Coupled Logic" (http://www.ddpp.com/DDPP4student/Supplementary_sections/ECL.pdf).
33. Sedra; Smith (2015). "Emitter-Coupled Logic (ECL)" (http://global.oup.com/us/companion websites/fdscontent/uscompanion/us/static/companion.websites/9780199339136/pdf/Additional_Material.pdf) (PDF). *Microelectronic Circuits*. Oxford University Press. p. 47. ISBN 978-0-19-933913-6.
34. Leppälä, Kari; Verkasalo, Raimo (1989). "Protection of Instrument Control Computers against Soft and Hard Errors and Cosmic Ray Effects" (<https://www.researchgate.net/publication/27782054>).
35. Minges, Merrill L.; ASM International. Handbook Committee (1989). *Electronic Materials Handbook: Packaging* (<https://books.google.com/books?id=c2YxCCaM9RIC&pg=PA163>). ASM International. p. 163. ISBN 9780871702852.
36. Jain, R.P. (2003). *Modern digital electronics* (<https://books.google.com/books?id=dnq3HmDN1ZAC&pg=RA1-PA110>). McGraw-Hill Education (India) Pvt Limited. p. 111. ISBN 9780070494923.
37. Holland, Nick (December 2002). "Interfacing Between LVPECL, VML, CML and LVDS Levels" (<http://focus.ti.com/lit/an/slla120/slla120.pdf>) (PDF). *Application Report*. Texas Instruments. SLLA120.
38. Goldie, John (January 21, 2003). "LVDS, CML, ECL – differential interfaces with odd voltages" (http://www.eetimes.com/document.asp?doc_id=1225744). *EE Times*.
39. Petty, Cleon; Pearson, Todd. "Designing with PECL (ECL at +5.0 V)" (<https://www.onsemi.com/pub/Collateral/AN1406-D.PDF>) (PDF). p. 3. AN1406-D.

Further reading

- Savard, John J. G. (2018) [2005]. "What Computers Are Made From" (<http://www.quadibloc.com/comp/cp01.htm>). *quadibloc*. Archived (<https://web.archive.org/web/20180702235616/http://www.quadibloc.com/comp/cp01.htm>) from the original on July 2, 2018. Retrieved July 16, 2018.

- US 2964652 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=US2964652>), Yourke, Hannon S., "Transistor Switching Circuits", published November 15, 1956, issued December 13, 1960
- Yourke, Hannon S. (September 1957). "Millimicrosecond Transistor Current Switching Circuits". *IRE Transactions on Circuit Theory*. **4** (3): 236–240.
[doi:10.1109/TCT.1957.1086377](https://doi.org/10.1109/TCT.1957.1086377) (<https://doi.org/10.1109%2FTCT.1957.1086377>).
- Mueller, Dieter (2008) [2006]. "DECL test run — Differential emitter-coupled logic" (<http://www.6502.org/users/dieter/decl/decl1.htm>). Archived (<https://web.archive.org/web/20180718215508/http://www.6502.org/users/dieter/decl/decl1.htm>) from the original on July 18, 2018. Retrieved July 18, 2018.

External links

- Motorola MECL logic family datasheets, 1963 (<http://www.worldpowersystems.com/archives/solid-state-datasheets/Motorola/MECL/index.html>)
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Emitter-coupled_logic&oldid=1292078626"

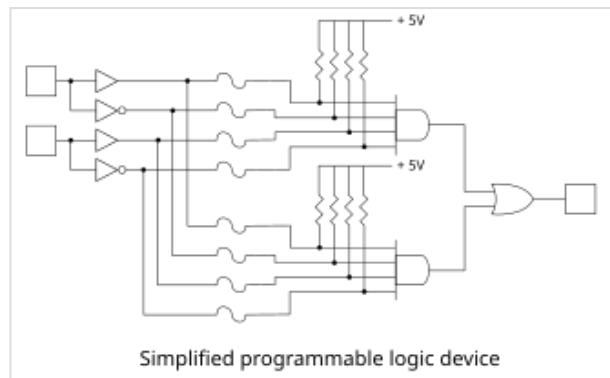


Programmable logic device

(Redirected from [Erasable programmable logic device](#))

A **programmable logic device (PLD)** is an [electronic component](#) used to build [reconfigurable digital circuits](#). Unlike digital logic constructed using discrete [logic gates](#) with fixed functions, the function of a PLD is undefined at the time of manufacture. Before the PLD can be used in a circuit it must be programmed to implement the desired function.^[1] Compared to fixed logic devices, programmable logic devices simplify the design of complex logic and may offer superior performance.^[2] Unlike for [microprocessors](#), programming a PLD changes the connections made between the gates in the device.

PLDs can broadly be categorised into, in increasing order of complexity, [simple programmable logic devices \(SPLDs\)](#), comprising [programmable array logic](#), [programmable logic array](#) and [generic array logic](#); [complex programmable logic devices \(CPLDs\)](#); and [field-programmable gate arrays \(FPGAs\)](#).



Simplified programmable logic device

A simplified PAL device. The programmable elements (shown as a fuse) connect both the true and complemented inputs to the AND gates. These AND gates, also known as product terms, are ORed together to form a sum-of-products logic array.

History

In 1969, [Motorola](#) offered the XC157, a mask-programmed gate array with 12 gates and 30 uncommitted input/output pins.^[3]

In 1970, [Texas Instruments](#) developed a mask-programmable IC based on the [IBM](#) read-only associative memory or ROAM. This device, the TMS2000, was programmed by altering the metal layer during the production of the IC. The TMS2000 had up to 17 inputs and 18 outputs with 8 [JK flip-flops](#) for memory. TI coined the term *programmable logic array* (PLA) for this device.^[4]

In 1971, [General Electric Company](#) (GE) was developing a programmable logic device based on the new [programmable read-only memory](#) (PROM) technology. This experimental device improved on IBM's ROAM by allowing multilevel logic. Intel had just introduced the floating-gate [UV EPROM](#) so the researcher at GE incorporated that technology. The GE device was the first erasable PLD ever developed, predating the [Altera](#) EPLD by over a decade. GE obtained several early patents on programmable logic devices.^{[5][6][7]}

In 1973 [National Semiconductor](#) introduced a mask-programmable PLA device (DM7575) with 14 inputs and 8 outputs with no memory registers. This was more popular than the TI part but the cost of making the metal mask limited its use. The device is significant because it was the basis for the field

programmable logic array produced by Signetics in 1975, the 82S100. (Intersil actually beat Signetics to market but poor yield doomed their part.)^{[8][9]}

In 1974 GE entered into an agreement with Monolithic Memories (MMI) to develop a mask-programmable logic device incorporating the GE innovations. The device was named *programmable associative logic array* or PALA. The MMI 5760 was completed in 1976 and could implement multilevel or sequential circuits of over 100 gates. The device was supported by a GE design environment where Boolean equations would be converted to mask patterns for configuring the device. The part was never brought to market.^[10]

PLA

In 1970, Texas Instruments developed a mask-programmable IC based on the IBM read-only associative memory or ROAM. This device, the TMS2000, was programmed by altering the metal layer during the production of the IC. The TMS2000 had up to 17 inputs and 18 outputs with 8 JK flip-flops for memory. TI coined the term programmable logic array for this device.^[4]

A programmable logic array (PLA) has a programmable AND gate array, which links to a programmable OR gate array, which can then be conditionally complemented to produce an output. A PLA is similar to a ROM concept, however a PLA does not provide full decoding of a variable and does not generate all the minterms as in a ROM.

The programmable logic sequencer (PLS/FPLS) is similar to a PLA, but with integral registered outputs using a number of flip-flops to allow creation of some state machines. Signetics introduced the first FPLS in 1979.^[11]

PAL

PAL devices have arrays of transistor cells arranged in a "fixed-OR, programmable-AND" plane used to implement "sum-of-products" binary logic equations for each of the outputs in terms of the inputs and either synchronous or asynchronous feedback from the outputs.

MMI introduced a breakthrough device in 1978, the programmable array logic or PAL. The architecture was simpler than that of Signetics' FPLA because it omitted the programmable OR array. This made the parts faster, smaller and cheaper. They were available in 20-pin 300-mil DIP packages, while the FPLAs came in 28-pin 600-mil packages. The PAL Handbook demystified the design process. The PALASM design software (PAL assembler) converted the engineers' Boolean equations into the fuse pattern required to program the part. The PAL devices were soon second-sourced by National Semiconductor, Texas Instruments and AMD.

After MMI succeeded with the 20-pin PAL parts, AMD introduced the 24-pin 22V10 PAL with additional features. After buying out MMI (1987), AMD spun off a consolidated operation as Vantis, and that business was acquired by Lattice Semiconductor in 1999.

GALs

An improvement on the PAL was the generic array logic device, or GAL, invented by [Lattice Semiconductor](#) in 1985. This device has the same logical properties as the PAL but can be erased and reprogrammed. The GAL is very useful in the prototyping stage of a design when any bugs in the logic can be corrected by reprogramming. GALs are programmed and reprogrammed using a PAL programmer, or, in the case of chips that support it, by using the in-circuit programming technique.

Lattice GALs combine CMOS and electrically erasable (E^2) floating gate technology for a high-speed, low-power logic device. A similar device called a PEEL (programmable electrically erasable logic) was introduced by the International CMOS Technology (ICT) corporation.

Sometimes GAL chips are referred as simple programmable logic device (SPLD), analogous to complex programmable logic device (CPLD) below.

CPLDs

PALs and GALs are available only in small sizes, equivalent to a few hundred logic gates. For bigger logic circuits, complex PLDs or CPLDs can be used. These contain the equivalent of several PALs linked by programmable interconnections, all in one integrated circuit. CPLDs can replace thousands, or even hundreds of thousands, of logic gates.

Some CPLDs are programmed using a PAL programmer, but this method becomes inconvenient for devices with hundreds of pins. A second method of programming is to solder the device to its printed circuit board, then feed it with a serial data stream from a personal computer. The CPLD contains a circuit that decodes the data stream and configures the CPLD to perform its specified logic function. Some manufacturers, such as [Altera](#) and [Atmel](#) (now [Microchip](#)), use [JTAG](#) to program CPLDs in-circuit from JAM files.

FPGAs

While PALs were being developed into GALs and CPLDs (all discussed above), a separate stream of development was happening. This type of device is based on gate array technology and is called the field-programmable gate array (FPGA). Early examples of FPGAs are the 82S100 array, and 82S105 sequencer, by Signetics, introduced in the late 1970s. The 82S100 was an array of AND terms. The 82S105 also had flip-flop functions.

(Remark: 82S100 and similar ICs from Signetics have PLA structure, AND-plane + OR-plane.)



Lattice GAL 16V8 and 20V8. These are 35 nanosecond devices.

FPGAs use a grid of logic gates, and once stored, the data doesn't change, similar to that of an ordinary gate array. The term *field-programmable* means the device is programmed by the customer, not the manufacturer. FPGAs and gate arrays are similar but gate arrays can only be configured at the factory during fabrication.^{[12][13][14]}

FPGAs are usually programmed after being soldered down to the circuit board, in a manner similar to that of larger CPLDs. In most larger FPGAs, the configuration is volatile and must be re-loaded into the device whenever power is applied or different functionality is required. Configuration is typically stored in a configuration PROM, EEPROM or flash memory.^[15] EEPROM versions may be in-system programmable (typically via JTAG).

The difference between FPGAs and CPLDs is that FPGAs are internally based on look-up tables (LUTs), whereas CPLDs form the logic functions with sea-of-gates (e.g. sum of products). CPLDs are meant for simpler designs while FPGAs are meant for more complex designs. In general, CPLDs are a good choice for wide combinational logic applications, whereas FPGAs are more suitable for large state machines such as microprocessors.

EPLDs

Using the same technology as EPROMs, **EPLDs** have a quartz window in the package that allows them to be erased on exposure to UV light.^{[16][17]}

Using the same technology as EEPROMs, **EEPLDs** can be erased electrically.^{[16][17]}

An **erasable programmable logic device (EPLD)** is an integrated circuit that comprises an array of PLDs that do not come pre-connected; the connections are programmed electrically by the user. Most GAL and FPGA devices are examples of EPLDs.



An EPLD from Cypress in a PLCC-package

Other variants

These are microprocessor circuits that contain some fixed functions and other functions that can be altered by code running on the processor. Designing self-altering systems requires that engineers learn new methods and that new software tools be developed.

PLDs are being sold now that contain a microprocessor with a fixed function (the so-called *core*) surrounded by programmable logic. These devices let designers concentrate on adding new features to designs without having to worry about making the microprocessor work. Also, the fixed-function microprocessor takes less space on the chip than a part of the programmable gate array implementing the same processor, leaving more space for the programmable gate array to contain the designer's specialized circuits.

How PLDs retain their configuration

A PLD is a combination of a logic device and a memory device. The memory is used to store the pattern that was given to the chip during programming. Most of the methods for storing data in an integrated circuit have been adapted for use in PLDs. These include:

- Silicon antifuses
- SRAM
- EPROM or EEPROM memory cells
- Flash memory

Silicon antifuses are connections that are made by applying a voltage across a modified area of silicon inside the chip. They are called *antifuses* because they work in the opposite way to normal fuses, which begin life as connections until they are broken by an electric current.

SRAM, or static RAM, is a volatile type of memory, meaning that its contents are lost each time the power is switched off. SRAM-based PLDs therefore have to be programmed every time the circuit is switched on. This is usually done automatically by another part of the circuit.

An EPROM memory cell is a MOSFET (metal-oxide-semiconductor field-effect transistor, or MOS transistor) that can be switched on by trapping an electric charge permanently on its gate electrode. This is done by a PAL programmer. The charge remains for many years and can only be removed by exposing the chip to strong ultraviolet light in a device called an EPROM eraser.

Flash memory is non-volatile, retaining its contents even when the power is switched off. It is stored on floating-gate MOSFET memory cells, and can be erased and reprogrammed as required. This makes it useful in PLDs that may be reprogrammed frequently, such as PLDs used in prototypes. Flash memory is a kind of EEPROM that holds information using trapped electric charges similar to EPROM. Consequently, flash memory can hold information for years, but possibly not as many years as EPROM.

As of 2005, most CPLDs are electrically programmable and erasable, and non-volatile. This is because they are too small to justify the inconvenience of programming internal SRAM cells every time they start up, and EPROM cells are more expensive due to their ceramic package with a quartz window.

PLD programming languages

Many PAL programming devices accept input in a standard file format, commonly referred to as 'JEDEC files'. They are analogous to software compilers. The languages used as source code for logic compilers are called hardware description languages, or HDLs.^[1]

PALASM, ABEL and CUPL are frequently used for low-complexity devices, while Verilog and VHDL are popular higher-level description languages for more complex devices. The more limited ABEL is often used for historical reasons, but for new designs, VHDL is more popular, even for low-complexity designs.

For modern PLD programming languages, design flows, and tools, see FPGA and reconfigurable computing.

PLD programming devices

A device programmer is used to transfer the Boolean logic pattern into the programmable device. In the early days of programmable logic, every PLD manufacturer also produced a specialized device programmer for its family of logic devices. Later, universal device programmers came onto the market that supported several logic device families from different manufacturers. Today's device programmers usually can program common PLDs (mostly PAL/GAL equivalents) from all existing manufacturers. Common file formats used to store the Boolean logic pattern (fuses) are JEDEC, Altera POF (programmable object file), or Xilinx BITstream.^[18]

See also

- Complex programmable logic device (CPLD)
- Field-programmable gate array (FPGA)
- Macrocell array
- Programmable array logic (PAL)

References

1. Horowitz, Paul; Hill, Winfield (2015). *Horowitz P., Hill W. - The Art of Electronics*. New York. p. 764. ISBN 978-0-521-80926-9.
2. Holdsworth, B.; Woods, R. C. (2003). *Digital Logic Design* (<https://www.sciencedirect.com/book/9780750645829/digital-logic-design>) (4th ed.). Elsevier. ISBN 978-0-7506-4582-9.
3. *Motorola Semiconductor Data Book, Fourth Edition*. Motorola Inc. 1969. p. IC-73.
4. Andres, Kent (October 1970). *A Texas Instruments Application Report: MOS programmable logic arrays*. Texas Instruments. Bulletin CA-158. Report introduces the TMS2000 and TMS2200 series of mask programmable PLAs.
5. Greer, David L. *Electrically Programmable Logic Circuits* US Patent 3,818,452 (<https://patents.google.com/patent/US3818452>). Assignee: General Electric, Filed: April 28, 1972, Granted: June 18, 1974
6. Greer, David L. *Multiple Level Associative Logic Circuits* US Patent 3,816,725 (<https://patents.google.com/patent/US3816725>). Assignee: General Electric, Filed: April 28, 1972, Granted: June 11, 1974
7. Greer, David L. *Segmented Associative Logic Circuits* US Patent 3,849,638 (<https://patents.google.com/patent/US3849638>). Assignee: General Electric, Filed: July 18, 1973, Granted: November 19, 1974
8. "Semiconductors and IC's : FPLA". *EDN*. **20** (13). Boston, MA: Cahners Publishing: 66. July 20, 1975. Press release on Intersil IM5200 field programmable logic array. Fourteen inputs pins and 48 product terms. Avalanche-induced-migration programming. Unit price was \$37.50
9. "FPLA's give quick custom logic". *EDN*. **20** (13). Boston, MA: Cahners Publishing: 61. July 20, 1975. Press release on Signetics 82S100 and 82S101 field programmable logic arrays. Fourteen inputs pins, 8 output pins and 48 product terms. NiCr fuse link programming.
10. Pellerin, David; Michael Holley (1991). *Practical Design Using Programmable Logic*. Prentice-Hall. p. 15. ISBN 0-13-723834-7.

11. Alford, Roger C. (1989). *Programmable Logic Designer's Guide* (<https://archive.org/details/programmablelogic0000alfo/>). Howard W. Sams. pp. 67–69. ISBN 0-672-22575-1. "The registers can be dynamically or permanently configured as D, J-K, or T flip-flops, because of the flexibility offered by the inclusion of the "M" inverters (M_0 - M_5). The outputs of the registers are fed back into the AND-array before reaching the tri-state output buffers, allowing for Mealy and Moore state machine development."
12. Dorf, Richard C. (3 October 2018). *Electronics, Power Electronics, Optoelectronics, Microwaves, Electromagnetics, and Radar* (<https://books.google.com/books?id=mZRsBgAAQBAJ&dq=gate+array+custom&pg=SA4-PA28>). CRC Press. ISBN 978-1-4200-0315-4.
13. Vahid, Frank; Givargis, Tony D. (17 October 2001). *Embedded System Design: A Unified Hardware / Software Introduction* (<https://books.google.com/books?id=ymoCEAAAQBAJ&dq=gate+array+custom&pg=PA276>). John Wiley & Sons. ISBN 978-0-471-38678-0.
14. Chen, Wai-Kai (3 October 2018). *The VLSI Handbook* (<https://books.google.com/books?id=MsqBgAAQBAJ&dq=gate+array+custom&pg=SA47-PA12>). CRC Press. ISBN 978-1-4200-0596-7.
15. "Using Flash Memory to Configure FPGAs" (https://cdrdv2-public.intel.com/654742/cfg_cf52010.pdf) (PDF). Altera. Retrieved 2024-08-21.
16. Tertulien Ndjountche. "Digital Electronics 2: Sequential and Arithmetic Logic Circuits" (<http://books.google.com/books?id=P7jeDAAAQBAJ>). 2016. p. 224.
17. Clive Maxfield. "The Design Warrior's Guide to FPGAs" (<https://books.google.com/books?id=dnuwr2xOFpUC>). 2004. p. 20
18. PLD File Formats (<http://www.pldtool.com/pld-file-formats>)

External links

- "PLD Tools Creating SVF, JAM, STAPL and other formats" (<https://web.archive.org/web/20120318182807/http://www.corelis.com/blog/index.php/blog/2010/12/01/pld-tools-creating-svf-jam-stapl-and-other-formats>). *JTAG / boundary-scan*. Corelis. Dec 1, 2010. Archived from the original (<http://www.corelis.com/blog/index.php/blog/2010/12/01/pld-tools-creating-svf-jam-stapl-and-other-formats>) on March 18, 2012. Retrieved July 18, 2011.
- "FPGAs and CPLDs" (<http://www.latticesemi.com/Products/FPGAandCPLD.aspx>). Lattice Semiconductor.

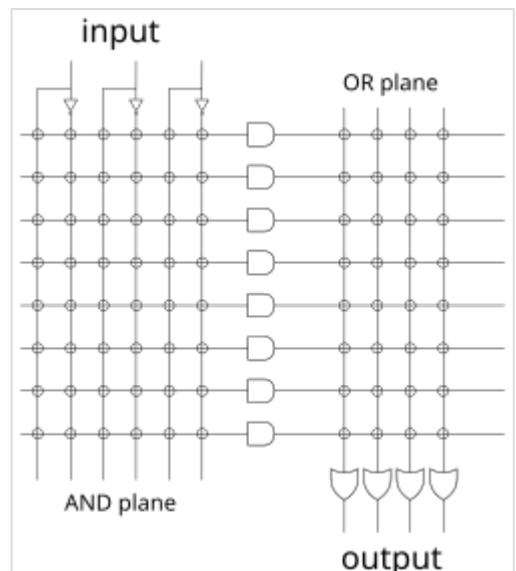
Retrieved from "https://en.wikipedia.org/w/index.php?title=Programmable_logic_device&oldid=1292006943#PLDs"



Programmable logic array

A **programmable logic array (PLA)** is a kind of programmable logic device used to implement combinational logic circuits. The PLA has a set of programmable AND gate planes, which link to a set of programmable OR gate planes, which can then be conditionally complemented to produce an output. It has 2^N AND gates for N input variables, and for M outputs from the PLA, there should be M OR gates, each with programmable inputs from all of the AND gates. This layout allows for many logic functions to be synthesized in the sum of products canonical forms.

PLAs differ from programmable array logic devices (PALs and GALs) in that both the AND and OR gate planes are programmable. PAL has programmable AND gates but fixed OR gates



PLA schematic example

History

In 1970, Texas Instruments developed a mask-programmable IC based on the IBM read-only associative memory or ROAM. This device, the TMS2000, was programmed by altering the metal layer during the production of the IC. The TMS2000 had up to 17 inputs and 18 outputs with 8 JK flip-flops for memory. TI coined the term *Programmable Logic Array* for this device.^[1]

Implementation procedure

1. Preparation in SOP (sum of products) form.
2. Obtain the minimum SOP form to reduce the number of product terms to a minimum.
3. Decide the input connection of the AND matrix for generating the required product term.
4. Then decide the input connections of the OR matrix to generate the sum terms.
5. Decide the connections of the inversion matrix.
6. Program the PLA.

PLA block diagram:

1ST BLOCK	2ND BLOCK	3RD BLOCK	4TH BLOCK	5TH BLOCK
INPUT BUFFER	AND MATRIX	OR MATRIX	INVERT/ NON INVERT MATRIX	FLIP-FLOP OUTPUT BUFFER

Advantages over read-only memory

The desired outputs for each combination of inputs *could* be programmed into a read-only memory, with the inputs being driven by the address bus and the outputs being read out as data. However, that would require a separate memory location for *every* possible combination of inputs, including combinations that are never supposed to occur, and also duplicating data for "don't care" conditions (for example, logic like "if input A is 1, then, as far as output X is concerned, we don't care what input B is": in a ROM this would have to be written out twice, once for each possible value of B, and as more "don't care" inputs are added, the duplication grows exponentially); therefore, a programmable logic array can often implement a piece of logic using fewer transistors than the equivalent in read-only memory. This is particularly valuable when it is part of a processing chip where transistors are scarce (for example, the original 6502 chip contained a PLA to direct various operations of the processor^[2]).

Applications

One application of a PLA is to implement the control over a datapath. It defines various states in an instruction set, and produces the next state (by conditional branching). [e.g. if the machine is in state 2, and will go to state 4 if the instruction contains an immediate field; then the PLA should define the actions of the control in state 2, will set the next state to be 4 if the instruction contains an immediate field, and will define the actions of the control in state 4]. Programmable logic arrays should correspond to a state diagram for the system.

The earliest Commodore 64 home computers released in 1982 (into early 1983) initially used a programmed Signetics 82S100 PLA, but as the demand increased, MOS Technology / Commodore Semiconductor Group began producing a mask-programmed PLA, which bore part number 906114-01.^[3]

See also

- Field-programmable gate array
- Gate array
- Programmable Array Logic

References

1. Andres, Kent (October 1970). *A Texas Instruments Application Report: MOS programmable logic arrays*. Texas Instruments. Bulletin CA-158.
2. How MOS 6502 Illegal Opcodes really work (<http://www.pagetable.com/?p=39>)
3. arcadecomponents.com - Commodore 906114-01 64 PLA IC (<http://www.arcadecomponents.com/catalog/item/3054735/2582640.htm>)

External links

- "Programmable Logic Array (PLA)" (<https://web.archive.org/web/20171214191144/www.cs.umd.edu/class/spring2003/cmsc311/Notes/Comb/pla.html>). *cmsc311*. University of Maryland. 2003. Archived from the original (<http://www.cs.umd.edu/class/spring2003/cmsc311/Notes/Comb/pla.html>) on 2017-12-14.
- "PLA (programmable logic array)" (<https://web.archive.org/web/20130115113124/http://tams-www.informatik.uni-hamburg.de/applets/hades/webdemos/42-programmable/10-pla/pla.html>). *Java Applet*. University of Hamburg. Archived from the original (<http://tams-www.informatik.uni-hamburg.de/applets/hades/webdemos/42-programmable/10-pla/pla.html>) on 2013-01-15. Retrieved 2010-04-07.

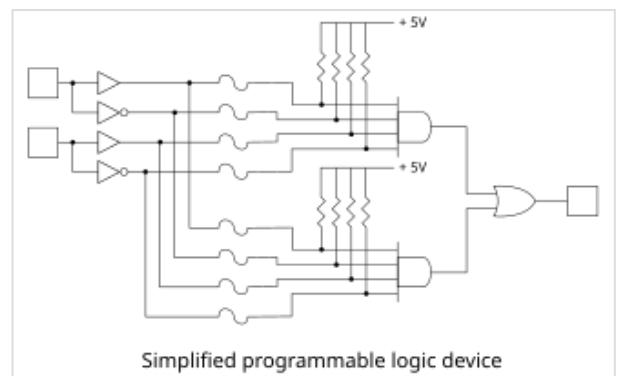
Retrieved from "https://en.wikipedia.org/w/index.php?title=Programmable_logic_array&oldid=1266324925"



Programmable logic device

A **programmable logic device (PLD)** is an electronic component used to build reconfigurable digital circuits. Unlike digital logic constructed using discrete logic gates with fixed functions, the function of a PLD is undefined at the time of manufacture. Before the PLD can be used in a circuit it must be programmed to implement the desired function.^[1] Compared to fixed logic devices, programmable logic devices simplify the design of complex logic and may offer superior performance.^[2] Unlike for microprocessors, programming a PLD changes the connections made between the gates in the device.

PLDs can broadly be categorised into, in increasing order of complexity, simple programmable logic devices (SPLDs), comprising programmable array logic, programmable logic array and generic array logic; complex programmable logic devices (CPLDs); and field-programmable gate arrays (FPGAs).



Simplified programmable logic device

A simplified PAL device. The programmable elements (shown as a fuse) connect both the true and complemented inputs to the AND gates. These AND gates, also known as product terms, are ORed together to form a sum-of-products logic array.

History

In 1969, Motorola offered the XC157, a mask-programmed gate array with 12 gates and 30 uncommitted input/output pins.^[3]

In 1970, Texas Instruments developed a mask-programmable IC based on the IBM read-only associative memory or ROAM. This device, the TMS2000, was programmed by altering the metal layer during the production of the IC. The TMS2000 had up to 17 inputs and 18 outputs with 8 JK flip-flops for memory. TI coined the term programmable logic array (PLA) for this device.^[4]

In 1971, General Electric Company (GE) was developing a programmable logic device based on the new programmable read-only memory (PROM) technology. This experimental device improved on IBM's ROAM by allowing multilevel logic. Intel had just introduced the floating-gate UV EPROM so the researcher at GE incorporated that technology. The GE device was the first erasable PLD ever developed, predating the Altera EPLD by over a decade. GE obtained several early patents on programmable logic devices.^{[5][6][7]}

In 1973 National Semiconductor introduced a mask-programmable PLA device (DM7575) with 14 inputs and 8 outputs with no memory registers. This was more popular than the TI part but the cost of making the metal mask limited its use. The device is significant because it was the basis for the field programmable logic array produced by Signetics in 1975, the 82S100. (Intersil actually beat Signetics to market but poor yield doomed their part.)^{[8][9]}

In 1974 GE entered into an agreement with Monolithic Memories (MMI) to develop a mask-programmable logic device incorporating the GE innovations. The device was named *programmable associative logic array* or PALA. The MMI 5760 was completed in 1976 and could implement multilevel or sequential circuits of over 100 gates. The device was supported by a GE design environment where Boolean equations would be converted to mask patterns for configuring the device. The part was never brought to market.^[10]

PLA

In 1970, Texas Instruments developed a mask-programmable IC based on the IBM read-only associative memory or ROAM. This device, the TMS2000, was programmed by altering the metal layer during the production of the IC. The TMS2000 had up to 17 inputs and 18 outputs with 8 JK flip-flops for memory. TI coined the term programmable logic array for this device.^[4]

A programmable logic array (PLA) has a programmable AND gate array, which links to a programmable OR gate array, which can then be conditionally complemented to produce an output. A PLA is similar to a ROM concept, however a PLA does not provide full decoding of a variable and does not generate all the minterms as in a ROM.

The programmable logic sequencer (PLS/FPLS) is similar to a PLA, but with integral registered outputs using a number of flip-flops to allow creation of some state machines. Signetics introduced the first FPLS in 1979.^[11]

PAL

PAL devices have arrays of transistor cells arranged in a "fixed-OR, programmable-AND" plane used to implement "sum-of-products" binary logic equations for each of the outputs in terms of the inputs and either synchronous or asynchronous feedback from the outputs.

MMI introduced a breakthrough device in 1978, the programmable array logic or PAL. The architecture was simpler than that of Signetics' FPLA because it omitted the programmable OR array. This made the parts faster, smaller and cheaper. They were available in 20-pin 300-mil DIP packages, while the FPLAs came in 28-pin 600-mil packages. The PAL Handbook demystified the design process. The PALASM design software (PAL assembler) converted the engineers' Boolean equations into the fuse pattern required to program the part. The PAL devices were soon second-sourced by National Semiconductor, Texas Instruments and AMD.

After MMI succeeded with the 20-pin PAL parts, AMD introduced the 24-pin 22V10 PAL with additional features. After buying out MMI (1987), AMD spun off a consolidated operation as Vantis, and that business was acquired by Lattice Semiconductor in 1999.

GALs

An improvement on the PAL was the generic array logic device, or GAL, invented by [Lattice Semiconductor](#) in 1985. This device has the same logical properties as the PAL but can be erased and reprogrammed. The GAL is very useful in the prototyping stage of a design when any bugs in the logic can be corrected by reprogramming. GALs are programmed and reprogrammed using a PAL programmer, or, in the case of chips that support it, by using the in-circuit programming technique.

Lattice GALs combine CMOS and electrically erasable (E^2) floating gate technology for a high-speed, low-power logic device. A similar device called a PEEL (programmable electrically erasable logic) was introduced by the International CMOS Technology (ICT) corporation.

Sometimes GAL chips are referred as simple programmable logic device (SPLD), analogous to complex programmable logic device (CPLD) below.

CPLDs

PALs and GALs are available only in small sizes, equivalent to a few hundred logic gates. For bigger logic circuits, complex PLDs or CPLDs can be used. These contain the equivalent of several PALs linked by programmable interconnections, all in one integrated circuit. CPLDs can replace thousands, or even hundreds of thousands, of logic gates.

Some CPLDs are programmed using a PAL programmer, but this method becomes inconvenient for devices with hundreds of pins. A second method of programming is to solder the device to its printed circuit board, then feed it with a serial data stream from a personal computer. The CPLD contains a circuit that decodes the data stream and configures the CPLD to perform its specified logic function. Some manufacturers, such as [Altera](#) and [Atmel](#) (now [Microchip](#)), use [JTAG](#) to program CPLDs in-circuit from JAM files.

FPGAs

While PALs were being developed into GALs and CPLDs (all discussed above), a separate stream of development was happening. This type of device is based on gate array technology and is called the field-programmable gate array (FPGA). Early examples of FPGAs are the 82S100 array, and 82S105 sequencer, by Signetics, introduced in the late 1970s. The 82S100 was an array of AND terms. The 82S105 also had flip-flop functions.

(Remark: 82S100 and similar ICs from Signetics have PLA structure, AND-plane + OR-plane.)



Lattice GAL 16V8 and 20V8. These are 35 nanosecond devices.

FPGAs use a grid of logic gates, and once stored, the data doesn't change, similar to that of an ordinary gate array. The term *field-programmable* means the device is programmed by the customer, not the manufacturer. FPGAs and gate arrays are similar but gate arrays can only be configured at the factory during fabrication.^{[12][13][14]}

FPGAs are usually programmed after being soldered down to the circuit board, in a manner similar to that of larger CPLDs. In most larger FPGAs, the configuration is volatile and must be re-loaded into the device whenever power is applied or different functionality is required. Configuration is typically stored in a configuration PROM, EEPROM or flash memory.^[15] EEPROM versions may be in-system programmable (typically via JTAG).

The difference between FPGAs and CPLDs is that FPGAs are internally based on look-up tables (LUTs), whereas CPLDs form the logic functions with sea-of-gates (e.g. sum of products). CPLDs are meant for simpler designs while FPGAs are meant for more complex designs. In general, CPLDs are a good choice for wide combinational logic applications, whereas FPGAs are more suitable for large state machines such as microprocessors.

EPLDs

Using the same technology as EPROMs, **EPLDs** have a quartz window in the package that allows them to be erased on exposure to UV light.^{[16][17]}

Using the same technology as EEPROMs, **EEPLDs** can be erased electrically.^{[16][17]}

An **erasable programmable logic device (EPLD)** is an integrated circuit that comprises an array of PLDs that do not come pre-connected; the connections are programmed electrically by the user. Most GAL and FPGA devices are examples of EPLDs.



An EPLD from Cypress in a PLCC-package

Other variants

These are microprocessor circuits that contain some fixed functions and other functions that can be altered by code running on the processor. Designing self-altering systems requires that engineers learn new methods and that new software tools be developed.

PLDs are being sold now that contain a microprocessor with a fixed function (the so-called *core*) surrounded by programmable logic. These devices let designers concentrate on adding new features to designs without having to worry about making the microprocessor work. Also, the fixed-function microprocessor takes less space on the chip than a part of the programmable gate array implementing the same processor, leaving more space for the programmable gate array to contain the designer's specialized circuits.

How PLDs retain their configuration

A PLD is a combination of a logic device and a memory device. The memory is used to store the pattern that was given to the chip during programming. Most of the methods for storing data in an integrated circuit have been adapted for use in PLDs. These include:

- Silicon antifuses
- SRAM
- EPROM or EEPROM memory cells
- Flash memory

Silicon antifuses are connections that are made by applying a voltage across a modified area of silicon inside the chip. They are called *antifuses* because they work in the opposite way to normal fuses, which begin life as connections until they are broken by an electric current.

SRAM, or static RAM, is a volatile type of memory, meaning that its contents are lost each time the power is switched off. SRAM-based PLDs therefore have to be programmed every time the circuit is switched on. This is usually done automatically by another part of the circuit.

An EPROM memory cell is a MOSFET (metal-oxide-semiconductor field-effect transistor, or MOS transistor) that can be switched on by trapping an electric charge permanently on its gate electrode. This is done by a PAL programmer. The charge remains for many years and can only be removed by exposing the chip to strong ultraviolet light in a device called an EPROM eraser.

Flash memory is non-volatile, retaining its contents even when the power is switched off. It is stored on floating-gate MOSFET memory cells, and can be erased and reprogrammed as required. This makes it useful in PLDs that may be reprogrammed frequently, such as PLDs used in prototypes. Flash memory is a kind of EEPROM that holds information using trapped electric charges similar to EPROM. Consequently, flash memory can hold information for years, but possibly not as many years as EPROM.

As of 2005, most CPLDs are electrically programmable and erasable, and non-volatile. This is because they are too small to justify the inconvenience of programming internal SRAM cells every time they start up, and EPROM cells are more expensive due to their ceramic package with a quartz window.

PLD programming languages

Many PAL programming devices accept input in a standard file format, commonly referred to as 'JEDEC files'. They are analogous to software compilers. The languages used as source code for logic compilers are called hardware description languages, or HDLs.^[1]

PALASM, ABEL and CUPL are frequently used for low-complexity devices, while Verilog and VHDL are popular higher-level description languages for more complex devices. The more limited ABEL is often used for historical reasons, but for new designs, VHDL is more popular, even for low-complexity designs.

For modern PLD programming languages, design flows, and tools, see FPGA and reconfigurable computing.

PLD programming devices

A device programmer is used to transfer the Boolean logic pattern into the programmable device. In the early days of programmable logic, every PLD manufacturer also produced a specialized device programmer for its family of logic devices. Later, universal device programmers came onto the market that supported several logic device families from different manufacturers. Today's device programmers usually can program common PLDs (mostly PAL/GAL equivalents) from all existing manufacturers. Common file formats used to store the Boolean logic pattern (fuses) are JEDEC, Altera POF (programmable object file), or Xilinx BITstream.^[18]

See also

- Complex programmable logic device (CPLD)
- Field-programmable gate array (FPGA)
- Macrocell array
- Programmable array logic (PAL)

References

1. Horowitz, Paul; Hill, Winfield (2015). *Horowitz P., Hill W. - The Art of Electronics*. New York. p. 764. ISBN 978-0-521-80926-9.
2. Holdsworth, B.; Woods, R. C. (2003). *Digital Logic Design* (<https://www.sciencedirect.com/book/9780750645829/digital-logic-design>) (4th ed.). Elsevier. ISBN 978-0-7506-4582-9.
3. *Motorola Semiconductor Data Book, Fourth Edition*. Motorola Inc. 1969. p. IC-73.
4. Andres, Kent (October 1970). *A Texas Instruments Application Report: MOS programmable logic arrays*. Texas Instruments. Bulletin CA-158. Report introduces the TMS2000 and TMS2200 series of mask programmable PLAs.
5. Greer, David L. *Electrically Programmable Logic Circuits* US Patent 3,818,452 (<https://patents.google.com/patent/US3818452>). Assignee: General Electric, Filed: April 28, 1972, Granted: June 18, 1974
6. Greer, David L. *Multiple Level Associative Logic Circuits* US Patent 3,816,725 (<https://patents.google.com/patent/US3816725>). Assignee: General Electric, Filed: April 28, 1972, Granted: June 11, 1974
7. Greer, David L. *Segmented Associative Logic Circuits* US Patent 3,849,638 (<https://patents.google.com/patent/US3849638>). Assignee: General Electric, Filed: July 18, 1973, Granted: November 19, 1974
8. "Semiconductors and IC's : FPLA". *EDN*. **20** (13). Boston, MA: Cahners Publishing: 66. July 20, 1975. Press release on Intersil IM5200 field programmable logic array. Fourteen inputs pins and 48 product terms. Avalanche-induced-migration programming. Unit price was \$37.50
9. "FPLA's give quick custom logic". *EDN*. **20** (13). Boston, MA: Cahners Publishing: 61. July 20, 1975. Press release on Signetics 82S100 and 82S101 field programmable logic arrays. Fourteen inputs pins, 8 output pins and 48 product terms. NiCr fuse link programming.
10. Pellerin, David; Michael Holley (1991). *Practical Design Using Programmable Logic*. Prentice-Hall. p. 15. ISBN 0-13-723834-7.

11. Alford, Roger C. (1989). *Programmable Logic Designer's Guide* (<https://archive.org/details/programmablelogic0000alfo/>). Howard W. Sams. pp. 67–69. ISBN 0-672-22575-1. "The registers can be dynamically or permanently configured as D, J-K, or T flip-flops, because of the flexibility offered by the inclusion of the "M" inverters (M_0 - M_5). The outputs of the registers are fed back into the AND-array before reaching the tri-state output buffers, allowing for Mealy and Moore state machine development."
12. Dorf, Richard C. (3 October 2018). *Electronics, Power Electronics, Optoelectronics, Microwaves, Electromagnetics, and Radar* (<https://books.google.com/books?id=mZRsBgAAQBAJ&dq=gate+array+custom&pg=SA4-PA28>). CRC Press. ISBN 978-1-4200-0315-4.
13. Vahid, Frank; Givargis, Tony D. (17 October 2001). *Embedded System Design: A Unified Hardware / Software Introduction* (<https://books.google.com/books?id=ymoCEAAAQBAJ&dq=gate+array+custom&pg=PA276>). John Wiley & Sons. ISBN 978-0-471-38678-0.
14. Chen, Wai-Kai (3 October 2018). *The VLSI Handbook* (<https://books.google.com/books?id=MsqBqAAQBAJ&dq=gate+array+custom&pg=SA47-PA12>). CRC Press. ISBN 978-1-4200-0596-7.
15. "Using Flash Memory to Configure FPGAs" (https://cdrdv2-public.intel.com/654742/cfg_cf52010.pdf) (PDF). Altera. Retrieved 2024-08-21.
16. Tertulien Ndjountche. "Digital Electronics 2: Sequential and Arithmetic Logic Circuits" (<http://books.google.com/books?id=P7jeDAAAQBAJ>). 2016. p. 224.
17. Clive Maxfield. "The Design Warrior's Guide to FPGAs" (<https://books.google.com/books?id=dnuwr2xOFpUC>). 2004. p. 20
18. PLD File Formats (<http://www.pldtool.com/pld-file-formats>)

External links

- "PLD Tools Creating SVF, JAM, STAPL and other formats" (<https://web.archive.org/web/20120318182807/http://www.corelis.com/blog/index.php/blog/2010/12/01/pld-tools-creating-svf-jam-stapl-and-other-formats>). *JTAG / boundary-scan*. Corelis. Dec 1, 2010. Archived from the original (<http://www.corelis.com/blog/index.php/blog/2010/12/01/pld-tools-creating-svf-jam-stapl-and-other-formats>) on March 18, 2012. Retrieved July 18, 2011.
- "FPGAs and CPLDs" (<http://www.latticesemi.com/Products/FPGAandCPLD.aspx>). Lattice Semiconductor.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Programmable_logic_device&oldid=1292006943"



Programmable Array Logic

Programmable Array Logic (PAL) is a family of programmable logic device semiconductors used to implement logic functions in digital circuits that was introduced by Monolithic Memories, Inc. (MMI) in March 1978.^[1] MMI obtained a registered trademark on the term PAL for use in "Programmable Semiconductor Logic Circuits". The trademark is currently held by Lattice Semiconductor.^[2]

PAL devices consisted of a small PROM (programmable read-only memory) core and additional output logic used to implement particular desired logic functions with few components.

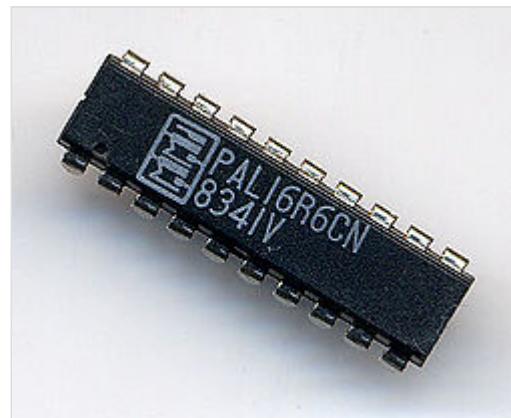
Using specialized machines, PAL devices were "field-programmable". PALs were available in several variants:

- "One-time programmable" (OTP) devices could not be updated and reused after initial programming. (MMI also offered a similar family called HAL, or "hard array logic", which were like PAL devices except that they were mask-programmed at the factory.)
- UV erasable versions (e.g.: PALCxxxx e.g.: PALC22V10) had a quartz window over the chip die and could be erased for re-use with an ultraviolet light source just like an EPROM.
- Later versions (PALCExxx e.g.: PALCE22V10) were flash erasable devices.

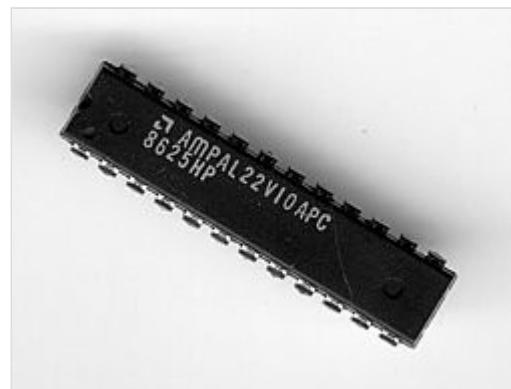
In most applications, electrically erasable GALs are now deployed as pin-compatible direct replacements for one-time programmable PALs.

History

Before PALs were introduced, designers of digital logic circuits would use small-scale integration (SSI) components, such as those in the 7400 series TTL (transistor-transistor logic) family; the 7400 family included a variety of logic building blocks, such as gates (NOT, NAND, NOR, AND, OR), multiplexers (MUXes) and demultiplexers (DEMUXes), flip-flops (D-type, JK, etc.) and others. One PAL device would typically replace dozens of such "discrete" logic packages, so the SSI business declined as the PAL business took off. PALs were used advantageously in many products, such as minicomputers, as documented in Tracy Kidder's best-selling book *The Soul of a New Machine*.



MMI 16R6 in 20-pin DIP



AMD 22V10 in 24-pin DIP

PALs were not the first commercial programmable logic devices; Signetics had been selling its field programmable logic array (FPLA) since 1975. These devices were completely unfamiliar to most circuit designers and were perceived to be too difficult to use. The FPLA had a relatively slow maximum operating speed (due to having both programmable-AND and programmable-OR arrays), was expensive, and had a poor reputation for testability. Another factor limiting the acceptance of the FPLA was the large package, a 600-mil (0.6", or 15.24 mm) wide 28-pin dual in-line package (DIP).

The project to create the PAL device was managed by John Birkner and the actual PAL circuit was designed by H. T. Chua.^[3] In a previous job (at mini-computer manufacturer Computer Automation), Birkner had developed a 16-bit processor using 80 standard logic devices. His experience with standard logic led him to believe that user-programmable devices would be more attractive if the devices were designed to replace standard logic. This meant that the package sizes had to be more typical of the existing devices, and the speeds had to be improved. MMI intended PALs to be a relatively low cost (sub \$3) part. However, the company initially had severe manufacturing yield problems and had to sell the devices for over \$50. This threatened the viability of the PAL as a commercial product, and MMI was forced to license the product line to National Semiconductor. PALs were later "second sourced" by Texas Instruments and Advanced Micro Devices.

Process technologies

Early PALs were 20-pin DIP components fabricated in silicon using bipolar transistor technology with one-time programmable (OTP) titanium-tungsten programming fuses.^[4] Later devices were manufactured by Cypress, Lattice Semiconductor and Advanced Micro Devices using CMOS technology.

The original 20- and 24-pin PALs were denoted by MMI as medium-scale integration (MSI) devices.

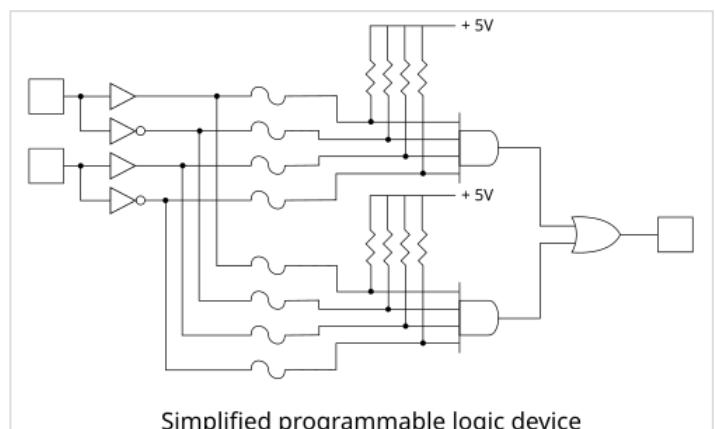
PAL architecture

The PAL architecture consists of two main components: a logic plane and output logic macrocells.

Programmable logic plane

The programmable logic plane is a programmable read-only memory (PROM) array that allows the signals present on the device pins, or the logical complements of those signals, to be routed to output logic macrocells.

PAL devices have arrays of transistor cells arranged in a "fixed-OR, programmable-AND" plane used to implement "sum-of-products" binary logic equations for each of the outputs in terms of the inputs and either synchronous or asynchronous feedback from the outputs.

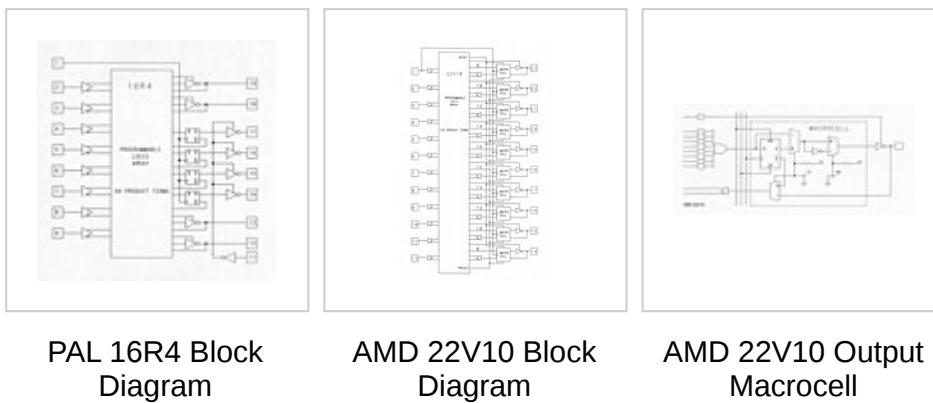


The programmable elements (shown as a fuse) connect both the true and complemented inputs to the AND gates. These AND gates, also known as *product terms*, are ORed together to form a *sum-of-products* logic array.

Output logic

The early 20-pin PALs had 10 inputs and 8 outputs. The outputs were active low and could be registered or combinational. Members of the PAL family were available with various output structures called "output logic macrocells" or OLMCs. Prior to the introduction of the "V" (for "variable") series, the types of OLMCs available in each PAL were fixed at the time of manufacture. (The PAL16L8 had 8 combinational outputs, and the PAL16R8 had 8 registered outputs. The PAL16R6 had 6 registered and 2 combinational outputs, while the PAL16R4 had 4 of each.) Each output could have up to 8 product terms (effectively AND gates); however, the combinational outputs used one of the terms to control a bidirectional output buffer. There were other combinations that had fewer outputs with more product terms per output and were available with active high outputs ("H" series).^{[5]:1–14} The "X" series of devices had an XOR gate before the register.^{[5]:1–9} There were also similar 24-pin versions of these PALs.

This fixed output structure often frustrated designers attempting to optimize the utility of PAL devices because output structures of different types were often required by their applications. (For example, one could not get 5 registered outputs with 3 active high combinational outputs.) So, in June 1983 AMD introduced the 22V10, a 24-pin device with 10 output logic macrocells.^[6] Each macrocell could be configured by the user to be combinational or registered, active high or active low. The number of product terms allocated to an output varied from 8 to 16. This one device could replace all of the 24-pin fixed function PAL devices. Members of the PAL "V" ("variable") series included the PAL16V8, PAL20V8 and PAL22V10.



Programming PALs

PALs were programmed electrically using binary patterns (as JEDEC ASCII/hexadecimal files) and a special electronic programming system available from either the manufacturer or a third party, such as DATA I/O. In addition to single-unit device programmers, device feeders and gang programmers were often used when more than just a few PALs needed to be programmed. (For large volumes, electrical programming costs could be eliminated by having the manufacturer fabricate a custom metal mask used to program the customers' patterns at the time of manufacture; MMI used the term "hard array logic" (HAL) to refer to devices programmed in this way.)

Programming languages (by chronological order of appearance)

Though some engineers programmed PAL devices by manually editing files containing the binary fuse pattern data, most opted to design their logic using a hardware description language (HDL) such as Data I/O's ABEL, Logical Devices' CUPL, or MMI's PALASM. These were computer-assisted design (CAD) (now referred to as "electronic design automation") programs which translated (or "compiled") the designers' logic equations into binary fuse map files used to program (and often test) each device.

PALASM

The PALASM (from "PAL assembler") language was developed by John Birkner in the early 1980s and the PALASM compiler was written by MMI in FORTRAN IV on an IBM 370/168. MMI made the source code available to users at no cost. By 1983, MMI customers ran versions on the DEC PDP-11, Data General NOVA, Hewlett-Packard HP 2100, MDS800 and others.

It was used to express Boolean equations for the output pins in a text file, which was then converted to the 'fuse map' file for the programming system using a vendor-supplied program; later the option of translation from schematics became common, and later still, 'fuse maps' could be 'synthesized' from an HDL (hardware description language) such as Verilog.

CUPL

Assisted Technology released CUPL (Compiler for Universal Programmable Logic) in September 1983.^[7] The software was always referred to as CUPL and never the expanded acronym. It was the first commercial design tool that supported multiple PLD families. The initial release was for the IBM PC and MS-DOS, but it was written in the C programming language so it could be ported to additional platforms.^[8] Assisted Technology was acquired by Personal CAD Systems (P-CAD) in July 1985. In 1986, PCAD's schematic capture package could be used as a front end for CUPL.^[9] CUPL was later acquired by Logical Devices and is now owned by Altium.^[10] CUPL is currently available as an integrated development package for Microsoft Windows.^[11]

Atmel releases for free WinCUPL (<http://www.atmel.com/tools/WINCUPLEX.aspx>) (their own design software for all Atmel SPLDs and CPLDs). Atmel was acquired by Microchip in 2016.

ABEL

Data I/O Corporation released ABEL in April, 1984. The development team was Michael Holley, Mike Mraz, Gerrit Barrere, Walter Bright, Bjorn Freeman-Benson, Kyu Lee, David Pellerin, Mary Bailey, Daniel Burrier and Charles Olivier.

PAL DESIGN SPECIFICATION							
PAL16R4 PAL							
CNT4SC							
4 bit counter with synchronous clear							
Michael Holley and Dave Pellerin							
Clk	Clear	NC	NC	NC	NC	NC	GND
OE	NC	NC	/Q3	/Q2	/Q1	/Q0	NC
NC	NC	NC	NC	NC	NC	NC	VCC
Q3 := Clear							
+ /Q3 * /Q2 * /Q1 * /Q0							
+ Q3 * Q0							
+ Q3 * Q1							
+ Q3 * Q2							
Q2 := Clear							
+ /Q2 * /Q1 * /Q0							
+ Q2 * Q0							
+ Q2 * Q1							
Q1 := Clear							
+ /Q1 * /Q0							
+ Q1 * Q0							
Q0 := Clear							
+ /Q0							
FUNCTION TABLE							
OE	Clear	Clk	/Q0	/Q1	/Q2	/Q3	
L	H	C	L	L	L	L	
L	L	C	H	L	L	L	
L	L	C	L	H	L	L	
L	L	C	H	H	L	L	
L	L	C	L	L	H	L	
L	H	C	L	L	L	L	

PALASM design of a 4-bit counter

Data I/O spun off the ABEL product line into an electronic design automation company called Synario Design Systems and then sold Synario to MINC Inc in 1997. MINC was focused on developing FPGA development tools. The company closed its doors in 1998 and Xilinx acquired some of MINC's assets including the ABEL language and tool set. ABEL then became part of the Xilinx Webpack tool suite. Now Xilinx owns ABEL.

Device programmers

Popular device programmers included [Data I/O Corporation's Model 60A Logic Programmer](#) and [Model 2900](#).

One of the first PAL programmers was the Structured Design SD20/24. They had the PALASM software built-in and only required a CRT terminal to enter the equations and view the fuse plots. After fusing, the outputs of the PAL could be verified if test vectors were entered in the source file.

Successors

After MMI succeeded with the 20-pin PAL parts introduced circa 1978, [AMD](#) introduced the 24-pin 22V10 PAL with additional features. After buying out MMI (circa 1987), AMD spun off a consolidated operation as Vantis, and that business was acquired by [Lattice Semiconductor](#) in 1999.^[12]

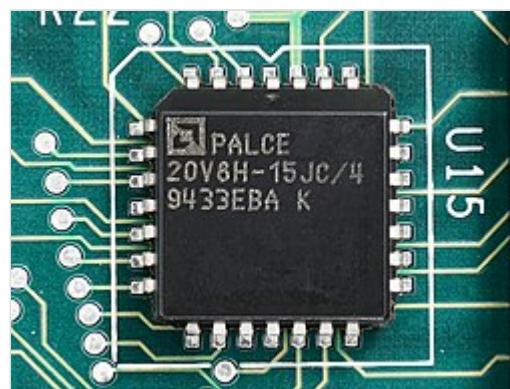
Altera introduced the EP300 (first CMOS PAL) in 1983 and later moved into the FPGA business.

Lattice Semiconductor introduced the [generic array logic \(GAL\)](#) family in 1985, with functional equivalents of the "V" series PALs that used reprogrammable logic planes based on [EEPROM](#) (electrically erasable programmable read-only memory) technology. National Semiconductor was a second source for GAL parts.

AMD introduced a similar family called PALCE. In general one GAL part is able to function as any of the similar family PAL devices. For example, the 16V8 GAL is able to replace the 16L8, 16H8, 16H6, 16H4, 16H2 and 16R8 PALs (and many others besides).

ICT (International CMOS Technology) introduced the PEEL 18CV8 in 1986. The 20-pin CMOS EEPROM part could be used in place of any of the registered-output bipolar PALs and used much less power.

Larger-scale programmable logic devices were introduced by [Atmel](#), [Lattice Semiconductor](#), and others. These devices extended the PAL architecture by including multiple logic planes and/or burying logic macrocells within the logic plane(s). The term [complex programmable logic device \(CPLD\)](#) was introduced to differentiate these devices from their PAL and GAL predecessors, which were then sometimes referred to as [simple programmable logic devices \(SPLDs\)](#).



AMD PALCE 20V8H-15JC in 28-pin
PLCC

Another large programmable logic device is the field-programmable gate array (FPGA). These are devices currently made by Intel (who acquired Altera) and Xilinx (who was acquired by AMD) and other semiconductor manufacturers.

See also

- Combinational logic

References

1. "Monolithic Memories announces: a revolution in logic design". *Electronic Design*. **26** (6). Rochelle, NJ: Hayden Publishing: 148B, 148C. March 18, 1978. Introductory advertisement on PAL (Programmable Array Logic).
2. Monolithic Memories, Inc (MMI) filed for a work mark on the term "PAL" for use in "Programmable Semiconductor Logic Circuits" on April 13, 1978. A registered trademark was granted on April 29, 1980, registration number 1134025. MMI's first use of the term PAL in commerce was on February 21, 1978. The trademark is currently held by Lattice Semiconductor Corporation of Hillsboro, Oregon. Source: United States Patent and Trademark Office online database.
3. Birkner, John (August 16, 1978). "Reduce random-logic complexity". *Electronic Design*. **26** (17). Rochelle, NJ: Hayden Publishing: 98–105.
4. *TIBPAL 16R8-15C Data Sheet* (<http://focus.ti.com/lit/ds/symlink/tibpal16r8-15c.pdf>) (PDF). Dallas TX: Texas Instruments. April 2000 [February 1984]. "These IMPACT circuits combine the latest Advanced Low-Power Schottky technology with proven titanium-tungsten fuses to provide reliable, high-performance substitutes for conventional TTL logic." TI was a second source vendor for the MMI PALS.
5. Birkner, John M.; Coli, Vincent J. (1983). *PAL Programmable Array Logic Handbook* (3rd ed.). Monolithic Memories, Inc.
6. *AmPAL 22V10 Advanced Information*. Sunnyvale CA: Advanced Micro Devices. June 1983. 04126A-PLP. Note: This is the data sheet published by AMD when the AmPAL 22V10 was introduced.
7. Alford, Roger C. (1989). *Programmable Logic Designer's Guide*. Howard W. Sams. pp. 14–15, 166–168. ISBN 0-672-22575-1. "In 1981, [Bob] Osann started Assisted Technology to develop PLD support tools. In September 1983 Assisted Technology released version 1.01a of its CUPL (Universal Compiler for Programmable Logic) PLD compiler, supporting 29 devices."
8. "CUPL - The Universal Language For Programmable Logic" (https://web.archive.org/web/20131029195406/http://www.swtpc.com/mholley/PLD_History/ABEL_project/CUPL_Data_Sheet_1983_ocr.pdf) (PDF) (Press release). San Jose, CA: Assisted Technology, Inc. 1983. Archived from the original (http://www.swtpc.com/mholley/PLD_History/ABEL_project/CUPL_Data_Sheet_1983_ocr.pdf) (PDF) on 2013-10-29. Retrieved 2013-08-10. An early 1983 pre-release datasheet for CUPL.
9. "Personal CAD Systems" (<https://books.google.com/books?id=6mxHXqXKdFMC&pg=PA97>). *Computer World*. **19** (29). Framingham, MA: CW Communications: 97. July 22, 1985. ISSN 0010-4841 (<https://search.worldcat.org/issn/0010-4841>).
10. US Patent and Trademark Office. "CUPL" Computer software, namely, software used to develop and compile designs for programmable logic devices, and related user manuals distributed therewith. First used in 1983, status Active. Serial Number 76357007. Registration Number 2909461. Owner: Altium Limited, Australia 3 Minna Close, Belrose NSW2085, Australia.

11. "CUBEL ChipDesigner 5.0" (<https://web.archive.org/web/20150529134635/http://www.logicaldevices.com:80/cubel.htm>). Logical Devices. August 2013. Archived from the original (<http://www.logicaldevices.com/cubel.htm>) on May 29, 2015.
12. "Lattice Semiconductor Acquires Vantis Corp. from AMD" (http://www.eetimes.com/document.asp?doc_id=1213080). *EE Times*. 26 April 1999. Retrieved May 13, 2015.

Further reading

Books

- *Programmable Logic Designer's Guide*; Roger Alford; Sams Publishing; 1989; ISBN 0-672-22575-1. (archive) (<https://archive.org/details/programmablelogi0000alfo/>)
- *PAL Programmable Logic Handbook*; 4ed; Monolithic Memories; 1985. (archive) (<https://archive.org/details/MonolithicMemories-MMI-ProgrammableLogicHandbookOCR/>)

Databooks

- *Bipolar LSI 1984 Databook*; 5ed; Monolithic Memories; 1984. (archive) (<https://archive.org/details/bipolar-lsi-1984-databook-fifth-edition/>)

Specifications

- *Standard Data Transfer Format Between Data Preparation System and Programmable Logic Device Programmer*; JEDEC Standard JESD3-C; JEDEC; June 1994.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Programmable_Array_Logic&oldid=1288193156"



Generic Array Logic

The **Generic Array Logic** (also known as **GAL** and sometimes as gate array logic^[1]) device was an innovation of the PAL and was invented by Lattice Semiconductor. The GAL was an improvement on the PAL because one device type was able to take the place of many PAL device types or could even have functionality not covered by the original range of PAL devices. Its primary benefit, however, was that it was erasable and re-programmable, making prototyping and design changes easier for engineers.

A similar device called a PEEL (programmable electrically erasable logic) was introduced by the International CMOS Technology (ICT) corporation.



Lattice GAL16V8D-15LJ

See also

- Programmable logic device (PLD)
 - Complex programmable logic device (CPLD)
 - Erasable programmable logic device (EPLD)
- GAL22V10

References

1. Peter J. Vis. "Dictionary of Digital Terms - PLD" (<https://www.petervis.com/dictionary-of-digital-terms/pld/pld.html>). www.petervis.com. Retrieved 24 October 2017.

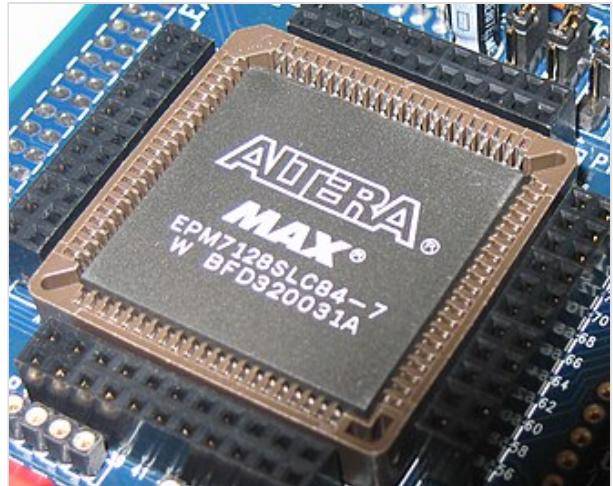
Further reading

- *PEEL Software and Applications Handbook*; International CMOS Technology (ICT); 138 pages; 1989. (archive) (https://archive.org/details/bitsavers_ictapeel19dApplicationsHandbook_8291490/)



Complex programmable logic device

A **complex programmable logic device (CPLD)** is a programmable logic device with complexity between that of PALs and FPGAs, and architectural features of both. The main building block of the CPLD is a macrocell, which contains logic implementing disjunctive normal form expressions and more specialized logic operations.



An Altera MAX 7000-series CPLD with 2500 gates.

Features

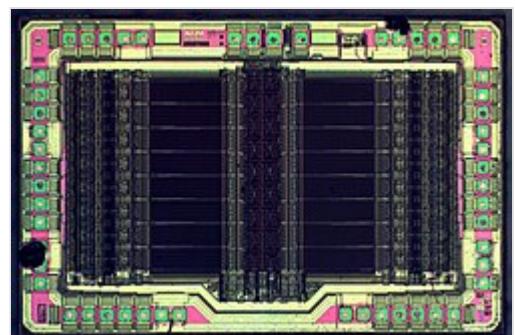
Some of the CPLD features are in common with PALs:

- Non-volatile configuration memory. Unlike many FPGAs, an external configuration ROM is not required, and the CPLD can function immediately on system start-up.
- For many legacy CPLD devices, routing constrains most logic blocks to have input and output signals connected to external pins, reducing opportunities for internal state storage and deeply layered logic. This is usually not a factor for larger CPLDs and newer CPLD product families.

Other features are in common with FPGAs:

- Large number of gates available. CPLDs typically have the equivalent of thousands to tens of thousands of logic gates, allowing implementation of moderately complicated data processing devices. PALs typically have a few hundred gate equivalents at most, while FPGAs typically range from tens of thousands to several million.
- Some provisions for logic more flexible than sum-of-product expressions, including complicated feedback paths between macro cells, and specialized logic for implementing various commonly used functions, such as integer arithmetic.

The most noticeable difference between a large CPLD and a small FPGA is the presence of on-chip non-volatile memory in the CPLD, which allows CPLDs to be used for "boot loader" functions, before handing over control to other devices not having their own permanent program storage. A good example is where a CPLD is used to load configuration data for an FPGA from non-volatile memory.^[1]



Die of an Altera EPM7032 EEPROM-based CPLD. Die size 3446x2252 μm . Technology node 1 μm .

Distinctions

CPLDs were an evolutionary step from even smaller devices that preceded them: PLAs (first shipped by Signetics) and PALs. These in turn were preceded by standard logic products, which offered no programmability and were used to build logic functions by physically wiring several standard logic chips (or hundreds of them) together (usually with wiring on a printed circuit board or boards, but sometimes, especially for prototyping, using wire wrap wiring).

The main distinction between FPGA and CPLD device architectures is that CPLDs are internally based on a collection of PLDs accompanied by a programmable interconnection structure, while FPGAs use logic blocks.

See also

- Language:
 - VHSIC Hardware Description Language (VHDL)
 - Verilog Hardware Description Language
 - Standard Test and Programming Language (JAM/STAPL)
- Manufacturers:
 - Altera (Now Intel)
 - Cypress Semiconductor
 - Lattice Semiconductor
 - Microchip Technology (subsumed Atmel)
 - Xilinx (Now AMD)
- Technology:
 - Application-specific integrated circuit (ASIC)
 - Erasable programmable logic device (EPLD)
 - Simple programmable logic device (SPLD)
 - Macrocell array
 - Programmable array logic (PAL)
 - Programmable logic array (PLA)
 - Programmable logic device (PLD)
 - Generic array logic (GAL)
 - Programmable electrically erasable logic (PEEL)
 - Field-programmable gate array (FPGA)

References

1. "Complex Programmable Logic Device" (<http://only-vlsi.blogspot.com/2008/05/complex-programmable-logic-device.html>). blogspot.com. May 2008. Retrieved 2013-11-17.

[title=Complex_programmable_logic_device&oldid=1288441599"](#)



Field-programmable gate array

A **field-programmable gate array (FPGA)** is a type of configurable integrated circuit that can be repeatedly programmed after manufacturing. FPGAs are a subset of logic devices referred to as programmable logic devices (PLDs). They consist of an array of programmable logic blocks with a connecting grid, that can be configured "in the field" to interconnect with other logic blocks to perform various digital functions. FPGAs are often used in limited (low) quantity production of custom-made products, and in research and development, where the higher cost of individual FPGAs is not as important, and where creating and manufacturing a custom circuit would not be feasible. Other applications for FPGAs include the telecommunications, automotive, aerospace, and industrial sectors, which benefit from their flexibility, high signal processing speed, and parallel processing abilities.

A FPGA configuration is generally written using a hardware description language (HDL) e.g. VHDL, similar to the ones used for application-specific integrated circuits (ASICs). Circuit diagrams were formerly used to write the configuration.

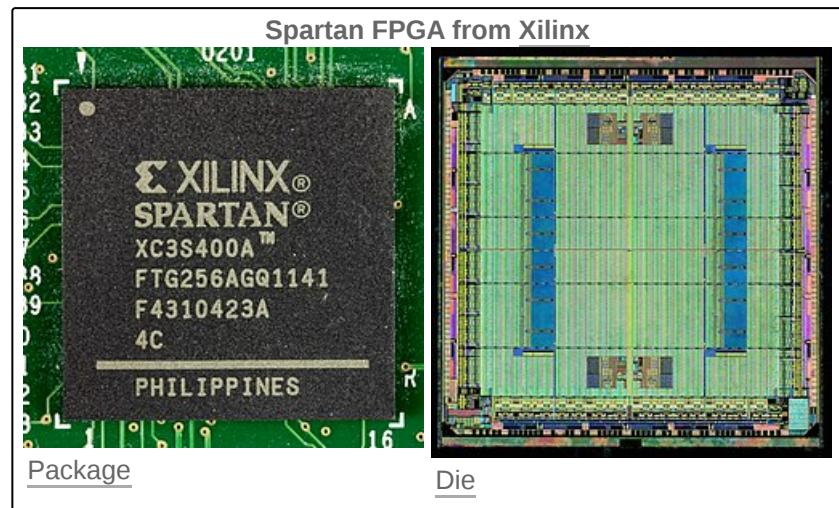
The logic blocks of an FPGA can be configured to perform complex combinational functions, or act as simple logic gates like AND and XOR. In most FPGAs, logic blocks also include memory elements, which may be simple flip-flops or more sophisticated blocks of memory.^[1] Many FPGAs can be reprogrammed to implement different logic functions, allowing flexible reconfigurable computing as performed in computer software.

FPGAs also have a role in embedded system development due to their capability to start system software development simultaneously with hardware, enable system performance simulations at a very early phase of the development, and allow various system trials and design iterations before finalizing the system architecture.^[2]

FPGAs are also commonly used during the development of ASICs to speed up the simulation process.



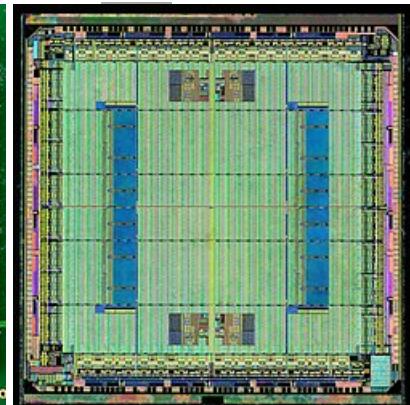
A Stratix IV FPGA from Altera



Spartan FPGA from Xilinx



Package



Die

History

The FPGA industry sprouted from programmable read-only memory (PROM) and programmable logic devices (PLDs). PROMs and PLDs both had the option of being programmed in batches in a factory or in the field (field-programmable).^[3]

Altera was founded in 1983 and delivered the industry's first reprogrammable logic device in 1984 – the EP300 – which featured a quartz window in the package that allowed users to shine an ultra-violet lamp on the die to erase the EPROM cells that held the device configuration.^[4]

Xilinx produced the first commercially viable field-programmable gate array in 1985^[3] – the XC2064.^[5] The XC2064 had programmable gates and programmable interconnects between gates, the beginnings of a new technology and market.^[6] The XC2064 had 64 configurable logic blocks (CLBs), with two three-input lookup tables (LUTs).^[7]

In 1987, the Naval Surface Warfare Center funded an experiment proposed by Steve Casselman to develop a computer that would implement 600,000 reprogrammable gates. Casselman was successful and a patent related to the system was issued in 1992.^[3]

Altera and Xilinx continued unchallenged and quickly grew from 1985 to the mid-1990s when competitors sprouted up, eroding a significant portion of their market share. By 1993, Actel (later Microsemi, now Microchip) was serving about 18 percent of the market.^[6]

The 1990s were a period of rapid growth for FPGAs, both in circuit sophistication and the volume of production. In the early 1990s, FPGAs were primarily used in telecommunications and networking. By the end of the decade, FPGAs found their way into consumer, automotive, and industrial applications.^[8]

By 2013, Altera (31 percent), Xilinx (36 percent) and Actel (10 percent) together represented approximately 77 percent of the FPGA market.^[9]

Companies like Microsoft have started to use FPGAs to accelerate high-performance, computationally intensive systems (like the data centers that operate their Bing search engine), due to the performance per watt advantage FPGAs deliver.^[10] Microsoft began using FPGAs to accelerate Bing in 2014, and in 2018 began deploying FPGAs across other data center workloads for their Azure cloud computing platform.^[11]

Growth

The following timelines indicate progress in different aspects of FPGA design.

Gates

- 1987: 9,000 gates, Xilinx^[6]
- 1992: 600,000, Naval Surface Warfare Department^[3]
- Early 2000s: millions^[8]
- 2013: 50 million, Xilinx^[12]

Market size

- 1985: First commercial FPGA : Xilinx XC2064^{[5][6]}
- 1987: \$14 million^[6]
- c. 1993: >\$385 million^[6]
- 2005: \$1.9 billion^[13]
- 2010 estimates: \$2.75 billion^[13]
- 2013: \$5.4 billion^[14]
- 2020 estimate: \$9.8 billion^[14]
- 2030 estimate: \$23.34 billion^[15]

Design starts

A *design start* is a new custom design for implementation on an FPGA.

- 2005: 80,000^[16]
- 2008: 90,000^[17]

Design

Contemporary FPGAs have ample logic gates and RAM blocks to implement complex digital computations. FPGAs can be used to implement any logical function that an ASIC can perform. The ability to update the functionality after shipping, partial re-configuration of a portion of the design^[18] and the low non-recurring engineering costs relative to an ASIC design (notwithstanding the generally higher unit cost), offer advantages for many applications.^[1]

As FPGA designs employ very fast I/O rates and bidirectional data buses, it becomes a challenge to verify correct timing of valid data within setup time and hold time.^[19] Floor planning helps resource allocation within FPGAs to meet these timing constraints.

Some FPGAs have analog features in addition to digital functions. The most common analog feature is a programmable slew rate on each output pin. This allows the user to set low rates on lightly loaded pins that would otherwise ring or couple unacceptably, and to set higher rates on heavily loaded high-speed channels that would otherwise run too slowly.^{[20][21]} Also common are quartz-crystal oscillator driver circuitry, on-chip RC oscillators, and phase-locked loops with embedded voltage-controlled oscillators used for clock generation and management as well as for high-speed serializer-deserializer (SERDES) transmit clocks and receiver clock recovery. Fairly common are differential comparators on input pins designed to be connected to differential signaling channels. A few mixed signal FPGAs have integrated peripheral analog-to-digital converters (ADCs) and digital-to-analog converters (DACs) with analog signal conditioning blocks, allowing them to operate as a system on a chip (SoC).^[22] Such devices blur the line between an FPGA, which carries digital ones and zeros on its internal programmable interconnect fabric, and field-programmable analog array (FPA), which carries analog values on its internal programmable interconnect fabric.

Logic blocks

The most common FPGA architecture consists of an array of logic blocks called configurable logic blocks (CLBs) or logic array blocks (LABs) (depending on vendor), I/O pads, and routing channels.^[1] Generally, all the routing channels have the same width (number of signals). Multiple I/O pads may fit into the height of one row or the width of one column in the array.

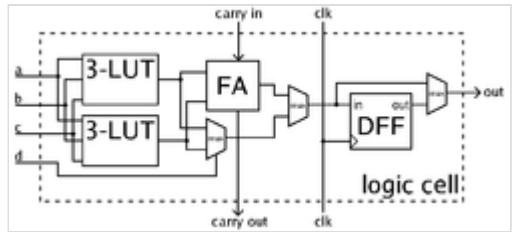
"An application circuit must be mapped into an FPGA with adequate resources. While the number of logic blocks and I/Os required is easily determined from the design, the number of routing channels needed may vary considerably even among designs with the same amount of logic. For example, a crossbar switch requires much more routing than a systolic array with the same gate count. Since unused routing channels increase the cost (and decrease the performance) of the FPGA without providing any benefit, FPGA manufacturers try to provide just enough channels so that most designs that will fit in terms of lookup tables (LUTs) and I/Os can be routed. This is determined by estimates such as those derived from Rent's rule or by experiments with existing designs."^[23]

In general, a logic block consists of a few logical cells. A typical cell consists of a 4-input LUT, a full adder (FA) and a D-type flip-flop. The LUT might be split into two 3-input LUTs. In *normal mode* those are combined into a 4-input LUT through the first multiplexer (mux). In *arithmetic mode*, their outputs are fed to the adder. The selection of mode is programmed into the second mux. The output can be either synchronous or asynchronous, depending on the programming of the third mux. In practice, the entire adder or parts of it are stored as functions into the LUTs in order to save space.^{[24][25][26]}

Hard blocks

Modern FPGA families expand upon the above capabilities to include higher-level functionality fixed in silicon. Having these common functions embedded in the circuit reduces the area required and gives those functions increased performance compared to building them from logical primitives. Examples of these include multipliers, generic DSP blocks, embedded processors, high-speed I/O logic and embedded memories.

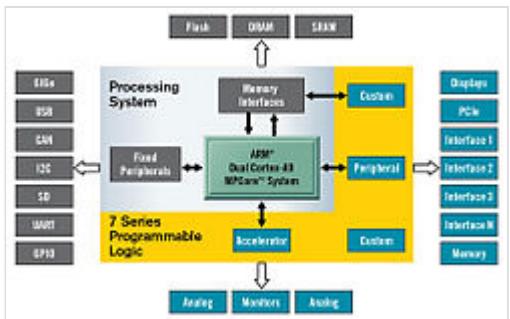
Higher-end FPGAs can contain high-speed multi-gigabit transceivers and hard IP cores such as processor cores, Ethernet medium access control units, PCI or PCI Express controllers, and external memory controllers. These cores exist alongside the programmable fabric, but they are built out of transistors instead of LUTs so they have ASIC-level performance and power consumption without consuming a significant amount of fabric resources, leaving more of the fabric free for the application-specific logic. The multi-gigabit transceivers also contain high-performance signal conditioning circuitry along with high-speed serializers and deserializers, components that cannot be built out of LUTs. Higher-level physical layer (PHY) functionality such as line coding may or may not be implemented alongside the serializers and deserializers in hard logic, depending on the FPGA.



Simplified example illustration of a logic cell (LUT – lookup table, FA – full adder, DFF – D-type flip-flop)

Soft core

An alternate approach to using hard macro processors is to make use of soft processor IP cores that are implemented within the FPGA logic. Nios II, MicroBlaze and Mico32 are examples of popular softcore processors. Many modern FPGAs are programmed at *run time*, which has led to the idea of reconfigurable computing or reconfigurable systems – CPUs that reconfigure themselves to suit the task at hand. Additionally, new non-FPGA architectures are beginning to emerge. Software-configurable microprocessors such as the Stretch S5000 adopt a hybrid approach by providing an array of processor cores and FPGA-like programmable cores on the same chip.



A Xilinx Zynq-7000 all-programmable system on a chip

Integration

In 2012 the coarse-grained architectural approach was taken a step further by combining the logic blocks and interconnects of traditional FPGAs with embedded microprocessors and related peripherals to form a complete system on a programmable chip. Examples of such hybrid technologies can be found in the Xilinx Zynq-7000 all Programmable SoC,^[27] which includes a 1.0 GHz dual-core ARM Cortex-A9 MPCore processor embedded within the FPGA's logic fabric,^[28] or in the Altera Arria V FPGA, which includes an 800 MHz dual-core ARM Cortex-A9 MPCore. The Atmel FPLIC is another such device, which uses an AVR processor in combination with Atmel's programmable logic architecture. The Microsemi SmartFusion devices incorporate an ARM Cortex-M3 hard processor core (with up to 512 kB of flash and 64 kB of RAM) and analog peripherals such as a multi-channel analog-to-digital converters and digital-to-analog converters in their flash memory-based FPGA fabric.

Clocking

Most of the logic inside of an FPGA is synchronous circuitry that requires a clock signal. FPGAs contain dedicated global and regional routing networks for clock and reset, typically implemented as an H tree, so they can be delivered with minimal skew. FPGAs may contain analog phase-locked loop or delay-locked loop components to synthesize new clock frequencies and manage jitter. Complex designs can use multiple clocks with different frequency and phase relationships, each forming separate clock domains. These clock signals can be generated locally by an oscillator or they can be recovered from a data stream. Care must be taken when building clock domain crossing circuitry to avoid metastability. Some FPGAs contain dual port RAM blocks that are capable of working with different clocks, aiding in the construction of building FIFOs and dual port buffers that bridge clock domains.

3D architectures

To shrink the size and power consumption of FPGAs, vendors such as Tabula and Xilinx have introduced 3D or stacked architectures.^{[29][30]} Following the introduction of its 28 nm 7-series FPGAs, Xilinx said that several of the highest-density parts in those FPGA product lines will be constructed using multiple dies in one package, employing technology developed for 3D construction and stacked-die assemblies.

Xilinx's approach stacks several (three or four) active FPGA dies side by side on a silicon interposer – a single piece of silicon that carries passive interconnect.^{[30][31]} The multi-die construction also allows different parts of the FPGA to be created with different process technologies, as the process requirements are different between the FPGA fabric itself and the very high speed 28 Gbit/s serial transceivers. An FPGA built in this way is called a heterogeneous FPGA.^[32]

Altera's heterogeneous approach involves using a single monolithic FPGA die and connecting other dies and technologies to the FPGA using Intel's embedded multi_die interconnect bridge (EMIB) technology.^[33]

Programming

To define the behavior of the FPGA, the user provides a design in a hardware description language (HDL) or as a schematic design. The HDL form is more suited to work with large structures because it's possible to specify high-level functional behavior rather than drawing every piece by hand. However, schematic entry can allow for easier visualization of a design and its component modules.

Using an electronic design automation tool, a technology-mapped netlist is generated. The netlist can then be fit to the actual FPGA architecture using a process called place and route, usually performed by the FPGA company's proprietary place-and-route software. The user will validate the results using timing analysis, simulation, and other verification and validation techniques. Once the design and validation process is complete, the binary file generated, typically using the FPGA vendor's proprietary software, is used to (re-)configure the FPGA. This file is transferred to the FPGA via a serial interface (JTAG) or to an external memory device such as an EEPROM.

The most common HDLs are VHDL and Verilog. National Instruments' LabVIEW graphical programming language (sometimes referred to as G) has an FPGA add-in module available to target and program FPGA hardware. Verilog was created to simplify the process making HDL more robust and flexible. Verilog has a C-like syntax, unlike VHDL.^[34]

To simplify the design of complex systems in FPGAs, there exist libraries of predefined complex functions and circuits that have been tested and optimized to speed up the design process. These predefined circuits are commonly called intellectual property (IP) cores, and are available from FPGA vendors and third-party IP suppliers. They are rarely free, and typically released under proprietary licenses. Other predefined circuits are available from developer communities such as OpenCores (typically released under free and open source licenses such as the GPL, BSD or similar license). Such designs are known as open-source hardware.

In a typical design flow, an FPGA application developer will simulate the design at multiple stages throughout the design process. Initially the RTL description in VHDL or Verilog is simulated by creating test benches to simulate the system and observe results. Then, after the synthesis engine has mapped the design to a netlist, the netlist is translated to a gate-level description where simulation is repeated to confirm the synthesis proceeded without errors. Finally, the design is laid out in the FPGA at which point propagation delay values can be back-annotated onto the netlist, and the simulation can be run again with these values.

More recently, OpenCL (Open Computing Language) is being used by programmers to take advantage of the performance and power efficiencies that FPGAs provide. OpenCL allows programmers to develop code in the C programming language.^[35] For further information, see high-level synthesis and C to HDL.

Most FPGAs rely on an SRAM-based approach to be programmed. These FPGAs are in-system programmable and re-programmable, but require external boot devices. For example, flash memory or EEPROM devices may load contents into internal SRAM that controls routing and logic. The SRAM approach is based on CMOS.

Rarer alternatives to the SRAM approach include:

- Fuse: one-time programmable. Bipolar. Obsolete.
- Antifuse: one-time programmable. CMOS. Examples: Actel SX and Axcelerator families; Quicklogic Eclipse II family.^[36]
- PROM: programmable read-only memory technology. One-time programmable because of plastic packaging. Obsolete.
- EPROM: erasable programmable read-only memory technology. One-time programmable but with window, can be erased with ultraviolet (UV) light. CMOS. Obsolete.
- EEPROM: electrically erasable programmable read-only memory technology. Can be erased, even in plastic packages. Some but not all EEPROM devices can be in-system programmed. CMOS.
- Flash: flash-erase EPROM technology. Can be erased, even in plastic packages. Some but not all flash devices can be in-system programmed. Usually, a flash cell is smaller than an equivalent EEPROM cell and is, therefore, less expensive to manufacture. CMOS. Example: Actel ProASIC family.^[36]

Manufacturers

In 2016, long-time industry rivals Xilinx (now part of AMD) and Altera (now part of Intel) were the FPGA market leaders.^[37] At that time, they controlled nearly 90 percent of the market.

Both Xilinx (now AMD) and Altera (now Intel) provide proprietary electronic design automation software for Windows and Linux (ISE/Vivado and Quartus) which enables engineers to design, analyze, simulate, and synthesize (compile) their designs.^{[38][39]}

In March 2010, Tabula announced their FPGA technology that uses time-multiplexed logic and interconnect that claims potential cost savings for high-density applications.^[40] On March 24, 2015, Tabula officially shut down.^[41]

On June 1, 2015, Intel announced it would acquire Altera for approximately US\$16.7 billion and completed the acquisition on December 30, 2015.^[42]

On October 27, 2020, AMD announced it would acquire Xilinx^[43] and completed the acquisition valued at about US\$50 billion in February 2022.^[44]

In February 2024 Altera became independent of Intel again.^[45]

Other manufacturers include:

- Achronix, manufacturing SRAM based FPGAs with 1.5 GHz fabric speed^[46]
- Altium, provides system-on-FPGA hardware-software design environment.^[47]
- Cologne Chip, German Government backed designer and producer of FPGAs^[48]
- Efinix offers small to medium-sized FPGAs. They combine logic and routing interconnects into a configurable XLR cell.
- GOWIN Semiconductors, manufacturing small and medium-sized SRAM and Flash-based FPGAs. They also offer pin-compatible replacements for a few Xilinx, Altera and Lattice products.
- Lattice Semiconductor manufactures low-power SRAM-based FPGAs featuring integrated configuration flash, instant-on and live reconfiguration
 - SiliconBlue Technologies provides extremely low-power SRAM-based FPGAs with optional integrated nonvolatile configuration memory; acquired by Lattice in 2011
- Microchip:
 - Microsemi (previously Actel), producing antifuse, flash-based, mixed-signal FPGAs; acquired by Microchip in 2018
 - Atmel, a second source of some Altera-compatible devices; also FPSLIC mentioned above;^[49] acquired by Microchip in 2016
- QuickLogic manufactures ultra-low-power sensor hubs, extremely-low-powered, low-density SRAM-based FPGAs, with display bridges MIPI and RGB inputs; MIPI, RGB and LVDS outputs.^[50]

Applications

An FPGA can be used to solve any problem which is computable. FPGAs can be used to implement a soft microprocessor, such as the Xilinx MicroBlaze or Altera Nios II. But their advantage lies in that they are significantly faster for some applications because of their parallel nature and optimality in terms of the number of gates used for certain processes.^[51]

FPGAs were originally introduced as competitors to CPLDs to implement glue logic for printed circuit boards. As their size, capabilities, and speed increased, FPGAs took over additional functions to the point where some are now marketed as full systems on chips (SoCs). Particularly with the introduction of dedicated multipliers into FPGA architectures in the late 1990s, applications that had traditionally been the sole reserve of digital signal processors (DSPs) began to use FPGAs instead.^{[52][53]}

The evolution of FPGAs has motivated an increase in the use of these devices, whose architecture allows the development of hardware solutions optimized for complex tasks, such as 3D MRI image segmentation, 3D discrete wavelet transform, tomographic image reconstruction, or PET/MRI systems.^{[54][55]} The developed solutions can perform intensive computation tasks with parallel processing, are dynamically reprogrammable, and have a low cost, all while meeting the hard real-time requirements associated with medical imaging.

Another trend in the use of FPGAs is hardware acceleration, where one can use the FPGA to accelerate certain parts of an algorithm and share part of the computation between the FPGA and a general-purpose processor. The search engine Bing is noted for adopting FPGA acceleration for its search algorithm in 2014.^[56] As of 2018, FPGAs are seeing increased use as AI accelerators including Microsoft's Project Catapult^[11] and for accelerating artificial neural networks for machine learning applications.

Originally, FPGAs were reserved for specific vertical applications where the volume of production is small. For these low-volume applications, the premium that companies pay in hardware cost per unit for a programmable chip is more affordable than the development resources spent on creating an ASIC. Often a custom-made chip would be cheaper if made in larger quantities, but FPGAs may be chosen to quickly bring a product to market. By 2017, new cost and performance dynamics broadened the range of viable applications.

Other uses for FPGAs include:

- Space (with radiation hardening^[57])
- Hardware security modules^[58]
- High-speed financial transactions^{[59][60]}
- Retrocomputing (e.g. the MARS and MiSTer FPGA projects)^[61]
- Large scale integrated digital differential analyzers, a form of an analog computer based on digital computing elements^[62]

Usage by United States military

FPGAs play a crucial role in modern military communications, especially in systems like the Joint Tactical Radio System (JTRS) and in devices from companies such as Thales and Harris Corporation. Their flexibility and programmability make them ideal for military communications, offering customizable and secure signal processing. In the JTRS, used by the US military, FPGAs provide adaptability and real-time processing, crucial for meeting various communication standards and encryption methods.^[63]

Security

Concerning hardware security, FPGAs have both advantages and disadvantages as compared to ASICs or secure microprocessors. FPGAs' flexibility makes malicious modifications during fabrication a lower risk.^[64] Previously, for many FPGAs, the design bitstream was exposed while the FPGA loads it from external memory, typically during powerup. All major FPGA vendors now offer a spectrum of security solutions to designers such as bitstream encryption and authentication. For example, Altera and Xilinx offer AES encryption (up to 256-bit) for bitstreams stored in an external flash memory. Physical unclonable functions (PUFs) are integrated circuits that have their own unique signatures and can be used to secure FPGAs while taking up very little hardware space.^[65]

FPGAs that store their configuration internally in nonvolatile flash memory, such as Microsemi's ProAsic 3 or Lattice's XP2 programmable devices, do not expose the bitstream and do not need encryption. In addition, flash memory for a lookup table provides single event upset protection for space applications. Customers wanting a higher guarantee of tamper resistance can use write-once, antifuse FPGAs from vendors such as Microsemi.

With its Stratix 10 FPGAs and SoCs, Altera introduced a Secure Device Manager and physical unclonable functions to provide high levels of protection against physical attacks.^[66]

In 2012 researchers Sergei Skorobogatov and Christopher Woods demonstrated that some FPGAs can be vulnerable to hostile intent. They discovered a critical backdoor vulnerability had been manufactured in silicon as part of the Actel/Microsemi ProASIC 3 making it vulnerable on many levels such as reprogramming crypto and access keys, accessing unencrypted bitstream, modifying low-level silicon features, and extracting configuration data.^[67]

In 2020 a critical vulnerability (named "Starbleed") was discovered in all Xilinx 7series FPGAs that rendered bitstream encryption useless. There is no workaround. Xilinx did not produce a hardware revision. Ultrascale and later devices, already on the market at the time, were not affected.

Similar technologies

Historically, FPGAs have been slower, less energy efficient and generally achieved less functionality than their fixed ASIC counterparts. A study from 2006 showed that designs implemented on FPGAs need on average 40 times as much area, draw 12 times as much dynamic power, and run at one third the speed of corresponding ASIC implementations.^[68]

Advantages of FPGAs include the ability to re-program when already deployed (i.e. "in the field") to fix bugs, and often include shorter time to market and lower non-recurring engineering costs. Vendors can also take a middle road via FPGA prototyping: developing their prototype hardware on FPGAs, but manufacture their final version as an ASIC so that it can no longer be modified after the design has been committed. This is often also the case with new processor designs.^[69] Some FPGAs have the capability of partial re-configuration that lets one portion of the device be re-programmed while other portions continue running.^{[70][71]}

The primary differences between complex programmable logic devices (CPLDs) and FPGAs are architectural. A CPLD has a comparatively restrictive structure consisting of one or more programmable sum-of-products logic arrays feeding a relatively small number of clocked registers. As a result, CPLDs are less flexible but have the advantage of more predictable timing delays and a higher logic-to-interconnect ratio. FPGA architectures, on the other hand, are dominated by interconnect. This makes them far more flexible (in terms of the range of designs that are practical for implementation on them) but also far more complex to design for, or at least requiring more complex electronic design automation (EDA) software. In practice, the distinction between FPGAs and CPLDs is often one of size as FPGAs are usually much larger in terms of resources than CPLDs. Typically only FPGAs contain more complex embedded functions such as adders, multipliers, memory, and serializer/deserializers. Another common distinction is that CPLDs contain embedded flash memory to store their configuration while FPGAs usually require external non-volatile memory (but not always). When a design requires simple instant-on (logic is already configured at power-up) CPLDs are generally preferred. For most other applications FPGAs are generally preferred. Sometimes both CPLDs and FPGAs are used in a single system design. In those designs, CPLDs generally perform glue logic functions and are responsible for "booting" the FPGA as well as controlling reset and boot sequence of the complete circuit board. Therefore, depending on the application it may be judicious to use both FPGAs and CPLDs in a single design.^[72]

See also



[Electronics portal](#)

- [FPGA Mezzanine Card](#)
- [CRUVI FPGA daughtercard standard](#)
- [List of HDL simulators](#)

References

1. "FPGA Architecture for the Challenge" (http://www.eecg.toronto.edu/~vaughn/challenge/fpga_arch.html). *toronto.edu*. University of Toronto.
2. Simpson, P. A. (2015). *FPGA Design, Best Practices for Team Based Reuse, 2nd edition*. Switzerland: Springer International Publishing AG. p. 16. ISBN 978-3-319-17924-7.
3. "History of FPGAs" (<https://web.archive.org/web/20070412183416/http://filebox.vt.edu/users/tmagine/history.htm>). Archived from the original (<http://filebox.vt.edu/users/tmagine/history.htm>) on April 12, 2007. Retrieved 2013-07-11.
4. Ron Wilson (21 April 2015). "In the Beginning" (https://web.archive.org/web/20150421045728/https://www.altera.com/solutions/technology/system-design/articles/_2013/in-the-beginning.html). *altera.com*. Archived from the original (https://www.altera.com/solutions/technology/system-design/articles/_2013/in-the-beginning.html) on 2015-04-21.
5. "XCELL issue 32" (<http://www.xilinx.com/publications/archives/xcell/Xcell32.pdf>) (PDF). Xilinx. Archived (<https://web.archive.org/web/20110107140043/http://www.xilinx.com/publications/archives/xcell/Xcell32.pdf>) (PDF) from the original on 2011-01-07.
6. Funding Universe. "Xilinx, Inc. (<http://www.fundinguniverse.com/company-histories/Xilinx-Inc-Company-History.html>)" Retrieved January 15, 2009.
7. Clive Maxfield, Programmable Logic DesignLine, "Xilinx unveil revolutionary 65nm FPGA architecture: the Virtex-5 family (<http://www.pldesignline.com/products/187203173>) Archived (<https://web.archive.org/web/20091225212024/http://www.pldesignline.com/products/187203173>) 2009-12-25 at the Wayback Machine. May 15, 2006. Retrieved February 5, 2009.
8. Maxfield, Clive (2004). *The Design Warrior's Guide to FPGAs: Devices, Tools and Flows* (<https://books.google.com/books?id=ZOadcQAACAAJ&pg=PA4>). Elsevier. p. 4. ISBN 978-0-7506-7604-5.
9. "Top FPGA Companies For 2013" (<https://web.archive.org/web/20150709173535/http://sourcetech411.com/2013/04/top-fpga-companies-for-2013/>). *sourcetech411.com*. 2013-04-28. Archived from the original (<http://sourcetech411.com/2013/04/top-fpga-companies-for-2013/>) on 2015-07-09. Retrieved 2015-07-08.
10. "Microsoft Supercharges Bing Search With Programmable Chips" (<https://www.wired.com/2014/06/microsoft-fpga/>). *WIRED*. 16 June 2014.
11. "Project Catapult" (<https://www.microsoft.com/en-us/research/project/project-catapult/>). *Microsoft Research*. July 2018.
12. Maxfield, Max. "Xilinx UltraScale FPGA Offers 50 Million Equivalent ASIC Gates" (https://www.eetimes.com/document.asp?doc_id=1320345). *www.eetimes.com*. EE Times.
13. Dylan McGrath, *EE Times*, "FPGA Market to Pass \$2.7 Billion by '10, In-Stat Says (<http://www.eetimes.com/news/design/business/showArticle.jhtml?articleID=188102617>)". May 24, 2006. Retrieved February 5, 2009.

14. "Global FPGA Market Analysis And Segment Forecasts To 2020 – FPGA Industry, Outlook, Size, Application, Product, Share, Growth Prospects, Key Opportunities, Dynamics, Trends, Analysis, FPGA Report – Grand View Research Inc" (<http://www.grandviewresearch.com/industry-analysis/fpga-market>). *grandviewresearch.com*.
15. "Field Programmable Gate Array Market To Reach \$23.34Bn By 2030" (<https://www.grandviewresearch.com/press-release/global-fpga-market>). *www.grandviewresearch.com*. Retrieved 2024-04-25.
16. Dylan McGrath, *EE Times*, "Gartner Dataquest Analyst Gives ASIC, FPGA Markets Clean Bill of Health (<http://www.eetimes.com/conf/dac/showArticle.jhtml?articleID=164302400>)". June 13, 2005. Retrieved February 5, 2009.
17. "Virtex-4 Family Overview" (http://www.xilinx.com/support/documentation/data_sheets/ds112.pdf) (PDF). *xilinx.com*. Archived (https://web.archive.org/web/20071122080100/http://www.xilinx.com/support/documentation/data_sheets/ds112.pdf) (PDF) from the original on 2007-11-22. Retrieved 14 April 2018.
18. Wisniewski, Remigiusz (2009). *Synthesis of compositional microprogram control units for programmable devices* (<http://zbc.uz.zgora.pl/Content/27955>). Zielona Góra: University of Zielona Góra. p. 153. ISBN 978-83-7481-293-1.
19. Oklobdzija, Vojin G. (2017). *Digital Design and Fabrication* (<https://books.google.com/books?id=VOnyWUUUj04C&dq=fpga+logic+gates+ram+blocks&pg=SA9-PA6>). CRC Press. ISBN 9780849386046.
20. "FPGA Signal Integrity tutorial" (<https://web.archive.org/web/20160307162907/http://wiki.altium.com/display/adoh/fpga+si+tutorial+-+simulating+the+reflection+characteristics>). *altium.com*. Archived from the original (<http://wiki.altium.com/display/ADOH/FPGA+SI+Tutorial+-+Simulating+the+Reflection+Characteristics>) on 2016-03-07. Retrieved 2010-06-15.
21. NASA: FPGA drive strength (http://klabs.org/richcontent/fpga_content/DesignNotes/signal_quality/actel_drive_strength/index.htm) Archived (https://web.archive.org/web/20101205230408/http://klabs.org/richcontent/fpga_content/DesignNotes/signal_quality/actel_drive_strength/index.htm) 2010-12-05 at the Wayback Machine
22. Mike Thompson (2007-07-02). "Mixed-signal FPGAs provide GREEN POWER" (<http://www.design-reuse.com/articles/16197/mixed-signal-fpgas-provide-green-power.html>). *Design & Reuse*.
23. M.b, Swami; V.p, Pawar (2014-07-31). "VLSI DESIGN: A NEW APPROACH" (<https://bioinfo-publication.org/pages/article.php?id=BIA0002301>). *Journal of Intelligence Systems*. 4 (1): 60–63. ISSN 2229-7057 (<https://search.worldcat.org/issn/2229-7057>).
24. 2. CycloneII Architecture (http://www.altera.com/literature/hb/cyc2/cyc2_cii51002.pdf). Altera. February 2007
25. "Documentation: Stratix IV Devices" (https://web.archive.org/web/20110926214034/http://www.altera.com/literature/hb/stratix-iv/stx4_5v1_01.pdf) (PDF). Altera.com. 2008-06-11. Archived from the original (http://www.altera.com/literature/hb/stratix-iv/stx4_5v1_01.pdf) (PDF) on 2011-09-26. Retrieved 2013-05-01.
26. Virtex-4 FPGA User Guide (http://www.xilinx.com/support/documentation/user_guides/ug070.pdf) (December 1st, 2008). Xilinx, Inc.
27. "Xilinx Inc, Form 8-K, Current Report, Filing Date Oct 19, 2011" (<http://edgar.secdatabase.com/520/95012311090713/filing-main.htm>). secdatabase.com. Retrieved May 6, 2018.
28. "Xilinx Inc, Form 10-K, Annual Report, Filing Date May 31, 2011" (<http://edgar.secdatabase.com/1249/95012311055454/filing-main.htm>). secdatabase.com. Retrieved May 6, 2018.
29. Dean Takahashi, VentureBeat. "Intel connection helped chip startup Tabula raise \$108M (<http://venturebeat.com/2011/05/02/intel-connection-helped-chip-startup-tabula-raise-108m>).". May 2, 2011. Retrieved May 13, 2011.
30. Lawrence Latif, The Inquirer. "FPGA manufacturer claims to beat Moore's Law (<https://web.archive.org/web/20101029124903/http://www.theinquirer.net/inquirer/news/1811460/fpga-manufacturer-claims-beat-moores-law>).". October 27, 2010. Retrieved May 12, 2011.

31. EDN Europe. "Xilinx adopts stacked-die 3D packaging (<http://www.edn-europe.com/xilinxadoptsstackeddie3dpackaging+article+4461+Europe.html>) Archived (<https://web.archive.org/web/20110219182606/http://www.edn-europe.com/xilinxadoptsstackeddie3dpackaging+article+4461+Europe.html>) 2011-02-19 at the Wayback Machine." November 1, 2010. Retrieved May 12, 2011.
32. Saban, Kirk (December 11, 2012). "Xilinx Stacked Silicon Interconnect Technology Delivers Breakthrough FPGA Capacity, Bandwidth, and Power Efficiency" (https://www.xilinx.com/support/documentation/white_papers/wp380_Stacked_Silicon_Interconnect_Technology.pdf) (PDF). *xilinx.com*. Archived (https://web.archive.org/web/20101105113516/http://www.xilinx.com/support/documentation/white_papers/wp380_Stacked_Silicon_Interconnect_Technology.pdf) (PDF) from the original on 2010-11-05. Retrieved 2018-11-30.
33. "Intel Custom Foundry EMIB" (<https://web.archive.org/web/20150713230215/http://www.intel.com/content/www/us/en/foundry/emib.html>). *Intel*. Archived from the original (<http://www.intel.com/content/www/us/en/foundry/emib.html>) on 2015-07-13. Retrieved 2015-07-13.
34. "Battle Over the FPGA: VHDL vs Verilog! Who is the True Champ?" (<https://web.archive.org/web/20201226074106/https://blog.digilentinc.com/battle-over-the-fpga-vhdl-vs-verilog-who-is-the-true-champ/>). *digilentinc.com*. Archived from the original (<https://blog.digilentinc.com/battle-over-the-fpga-vhdl-vs-verilog-who-is-the-true-champ/>) on 2020-12-26. Retrieved 2020-12-16.
35. "Why use OpenCL on FPGAs?" (<https://web.archive.org/web/20170101125857/https://streamcomputing.eu/blog/2014-09-16/use-opencl-fpgas/>). *StreamComputing*. 2014-09-16. Archived from the original (<http://streamcomputing.eu/blog/2014-09-16/use-opencl-fpgas/>) on 2017-01-01. Retrieved 2015-07-17.
36. "All about FPGAs" (<https://www.edn.com/all-about-fpgas/>). 21 March 2006.
37. Dillien, Paul (March 6, 2017). "And the Winner of Best FPGA of 2016 is..." (https://web.archive.org/web/20190105015123/http://www.eetimes.com/author.asp?doc_id=1331443). *EETimes*. Archived from the original (http://www.eetimes.com/author.asp?doc_id=1331443) on January 5, 2019. Retrieved September 7, 2017.
38. "Xilinx ISE Design Suite" (<https://www.xilinx.com/products/design-tools/ise-design-suite.html>). *www.xilinx.com*. Retrieved 2018-12-01.
39. "FPGA Design Software - Intel Quartus Prime" (<https://www.altera.com/products/design-software/fpga-design/quartus-prime/overview.html>). *Intel*. Retrieved 2018-12-01.
40. "Tabula's Time Machine — Micro Processor Report" (https://web.archive.org/web/20110410094902/http://www.tabula.com/news/M11_Tabula_Reprint.pdf) (PDF). Archived from the original (http://www.tabula.com/news/M11_Tabula_Reprint.pdf) (PDF) on 2011-04-10.
41. Tabula to shut down; 120 jobs lost at fabless chip company (<http://www.bizjournals.com/sanjose/news/2015/02/11/tabula-to-shut-down-120-jobs-lost-at-fabless-chip.html>) Silicon Valley Business Journal
42. "Intel to buy Altera for \$16.7 billion in its biggest deal ever" (<https://www.reuters.com/article/us-altera-m-a-intel-idUSKBN0OH2E020150601>). *Reuters*. June 2015.
43. "AMD to Acquire Xilinx, Creating the Industry's High Performance Computing Leader" (<https://www.amd.com/en/press-releases/2020-10-27-amd-to-acquire-xilinx-creating-the-industry-s-high-performance-computing>). October 2020.
44. "AMD closes record chip industry deal with estimated \$50 billion purchase of Xilinx" (<https://www.reuters.com/technology/amd-closes-biggest-chip-acquisition-with-498-bln-purchase-xilinx-2022-02-14/>). *Reuters*. February 2022.
45. "Intel Launches Altera, Its New Standalone FPGA Company" (<https://www.intel.com/content/www/us/en/newsroom/news/intel-launches-altera-standalone-fpga-operation.html>). *Intel* (Press release). Retrieved 2024-02-29.
46. "Achronix to Use Intel's 22nm Manufacturing" (http://newsroom.intel.com/community/intel_newsroom/blog/2010/11/01/chip-shot-achronix-to-use-intel-s-22nm-manufacturing). *Intel Newsroom* (Press release). 2010-11-01. Retrieved 2018-12-01.

47. Maxfield, Clive (16 June 2004). *The Design Warrior's Guide to FPGAs* (<https://books.google.com/books?id=dnuwr2xOFpUC&dq=fpga+altium&pg=PA117>). Elsevier Science. ISBN 9780080477138.
48. "About the company – Cologne Chip" (<https://colognechip.com/about-the-company/>). Retrieved 2024-02-27.
49. "Top FPGA Companies For 2013" (<https://web.archive.org/web/20180824135219/https://sourcetech411.com/2013/04/top-fpga-companies-for-2013/>). SourceTech411. 2013-04-28. Archived from the original (<http://sourcetech411.com/2013/04/top-fpga-companies-for-2013/>) on 2018-08-24. Retrieved 2018-12-01.
50. "QuickLogic — Customizable Semiconductor Solutions for Mobile Devices" (<http://www.quicklogic.com/>). www.quicklogic.com. QuickLogic Corporation. Retrieved 2018-10-07.
51. "Xilinx Inc, Form 8-K, Current Report, Filing Date Apr 26, 2006" (<http://edgar.secdatabase.com/657/110465906027899/filing-main.htm>). secdatabase.com. Retrieved May 6, 2018.
52. "Publications and Presentations" (https://web.archive.org/web/20100821182813/http://www.bdti.com/articles/info_eet0207fpga.htm). [bdti.com](http://www.bdti.com/articles/info_eet0207fpga.htm). Archived from the original (http://www.bdti.com/articles/info_eet0207fpga.htm) on 2010-08-21. Retrieved 2018-11-02.
53. LaPedus, Mark (5 February 2007). "Xilinx aims 65-nm FPGAs at DSP applications" (<https://www.eetimes.com/xilinx-aims-65-nm-fpgas-at-dsp-applications/#>). EETimes.
54. Alcaín, Eduardo; Fernández, Pedro R.; Nieto, Rubén; Montemayor, Antonio S.; Vilas, Jaime; Galiana-Bordera, Adrian; Martinez-Girones, Pedro Miguel; Prieto-de-la-Lastra, Carmen; Rodriguez-Vila, Borja; Bonet, Marina; Rodriguez-Sanchez, Cristina (2021-12-15). "Hardware Architectures for Real-Time Medical Imaging" (<https://doi.org/10.3390%2Felectronics10243118>). *Electronics*. **10** (24): 3118. doi:10.3390/electronics10243118 (<https://doi.org/10.3390%2Felectronics10243118>). ISSN 2079-9292 (<https://search.worldcat.org/issn/2079-9292>).
55. Nagornov, Nikolay N.; Lyakhov, Pavel A.; Valueva, Maria V.; Bergerman, Maxim V. (2022). "RNS-Based FPGA Accelerators for High-Quality 3D Medical Image Wavelet Processing Using Scaled Filter Coefficients" (<https://doi.org/10.1109%2FACCESS.2022.3151361>). *IEEE Access*. **10**: 19215–19231. Bibcode:2022IEEEA..1019215N (<https://ui.adsabs.harvard.edu/abs/2022IEEEA..1019215N>). doi:10.1109/ACCESS.2022.3151361 (<https://doi.org/10.1109%2FACCESS.2022.3151361>). ISSN 2169-3536 (<https://search.worldcat.org/issn/2169-3536>). S2CID 246895876 (<https://api.semanticscholar.org/CorpusID:246895876>).
56. Morgan, Timothy Pricket (2014-09-03). "How Microsoft Is Using FPGAs To Speed Up Bing Search" (<https://www.enterprisetech.com/2014/09/03/microsoft-using-fpgas-speed-bing-search/>). Enterprise Tech. Retrieved 2018-09-18.
57. "FPGA development devices for radiation-hardened space applications introduced by Microsemi" (<https://www.militaryaerospace.com/articles/2016/06/radiation-hardened-space-fpga.html>). www.militaryaerospace.com. 2016-06-03. Retrieved 2018-11-02.
58. "CrypTech: Building Transparency into Cryptography t" (https://cryptech.is/wp-content/uploads/2016/02/CrypTech_Building_Transparency.pdf) (PDF). Archived (https://web.archive.org/web/20160807180252/https://cryptech.is/wp-content/uploads/2016/02/CrypTech_Building_Transparency.pdf) (PDF) from the original on 2016-08-07.
59. Mann, Tobias (2023-03-08). "While Intel XPPUs are delayed, here's some more FPGAs to tide you over" (https://www.theregister.com/2023/03/08/intel_fpga_agilex/). The Register.
60. Leber, Christian; Geib, Benjamin; Litz, Heiner (September 2011). *High Frequency Trading Acceleration Using FPGAs* (<https://ieeexplore.ieee.org/document/6044837>). International Conference on Field Programmable Logic and Applications. IEEE. doi:10.1109/FPL.2011.64 (<https://doi.org/10.1109%2FFPL.2011.64>).
61. "The DIY MiSTer Handheld" (<https://www.retrorgb.com/the-diy-mister-handheld.html>). 16 December 2024.
62. DDA on FPGA - A modern Analog Computer (<https://people.ece.cornell.edu/land/courses/ece5760/DDA/index.htm>)

63. "Software-defined radio and JTRS" (<https://www.militaryaerospace.com/computers/article/16710419/softwaredefined-radio-and-jtrs>). *Military Aerospace*. 2004-12-01. Retrieved 2024-01-17.
64. Huffmire, Ted; Brotherton, Brett; Sherwood, Timothy; Kastner, Ryan; Levin, Timothy; Nguyen, Thuy D.; Irvine, Cynthia (2008). "Managing Security in FPGA-Based Embedded Systems". *IEEE Design & Test of Computers*. **25** (6): 590–598. doi:10.1109/MDT.2008.166 (<https://doi.org/10.1109%2FMDT.2008.166>). hdl:10945/7159 (<https://hdl.handle.net/10945%2F7159>). S2CID 115840 (<https://api.semanticscholar.org/CorpusID:115840>).
65. Babaei, Armin; Schiele, Gregor; Zohner, Michael (2022-07-26). "Reconfigurable Security Architecture (RESA) Based on PUF for FPGA-Based IoT Devices" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9331300>). *Sensors*. **22** (15): 5577. Bibcode:2022Senso..22.5577B (<https://ui.adsabs.harvard.edu/abs/2022Senso..22.5577B>). doi:10.3390/s22155577 (<https://doi.org/10.3390/s22155577>). ISSN 1424-8220 (<https://search.worldcat.org/issn/1424-8220>). PMC 9331300 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9331300>). PMID 35898079 (<https://pubmed.ncbi.nlm.nih.gov/35898079>).
66. "EETimes on PUF: Security features for non-security experts – Intrinsic ID" (<https://web.archive.org/web/20150713093531/https://www.intrinsic-id.com/eetimes-security-features-for-non-security-experts/>). *Intrinsic ID*. 2015-06-09. Archived from the original (<https://www.intrinsic-id.com/eetimes-security-features-for-non-security-experts/>) on 2015-07-13. Retrieved 2015-07-12.
67. Skorobogatov, Sergei; Woods, Christopher (2012). "Breakthrough Silicon Scanning Discovers Backdoor in Military Chip". *Cryptographic Hardware and Embedded Systems – CHES 2012*. Lecture Notes in Computer Science. Vol. 7428. pp. 23–40. doi:10.1007/978-3-642-33027-8_2 (https://doi.org/10.1007%2F978-3-642-33027-8_2). ISBN 978-3-642-33026-1.
68. Kuon, Ian; Rose, Jonathan (2006). "Measuring the gap between FPGAs and ASICs" (https://web.archive.org/web/20100622170541/http://ece.gmu.edu/coursewebpages/ECE/ECE448/S09/viewgraphs/Gap_between_FPGAs_and ASICs.pdf) (PDF). *Proceedings of the international symposium on Field programmable gate arrays – FPGA'06*. New York, NY: ACM. pp. 21–30. doi:10.1145/1117201.1117205 (<https://doi.org/10.1145%2F1117201.1117205>). ISBN 1-59593-292-5. Archived from the original (http://ece.gmu.edu/coursewebpages/ECE/ECE448/S09/viewgraphs/Gap_between_FPGAs_and ASICs.pdf) (PDF) on 2010-06-22. Retrieved 2017-10-25.
69. Cutress, Ian (August 27, 2019). "Xilinx Announces World Largest FPGA: Virtex Ultrascale+ VU19P with 9m Cells" (<https://www.anandtech.com/show/14798/xilinx-announces-world-largest-fpga-virtex-ultrascale-vu19p-with-9m-cells>). *AnandTech*.
70. "AN 818: Static Update Partial Reconfiguration Tutorial: for Intel Stratix 10 GX FPGA Development Board" (<https://www.intel.com/content/www/us/en/programmable/documentation/ekx1496870149834.html>). www.intel.com. Retrieved 2018-12-01.
71. "Can FPGAs dynamically modify their logic?" (<https://electronics.stackexchange.com/questions/45115/can-fpgas-dynamically-modify-their-logic>). *Electrical Engineering Stack Exchange*. Retrieved 2018-12-01.
72. "CPLD vs FPGA: Differences between them and which one to use? – Numato Lab Help Center" (<https://numato.com/kb/cpld-vs-fpga-differences-one-use/>). numato.com. 2017-11-29.

Further reading

- Sadrozinski, Hartmut F.-W.; Wu, Jinyuan (2010). *Applications of Field-Programmable Gate Arrays in Scientific Research*. Taylor & Francis. ISBN 978-1-4398-4133-4.

- Wirth, Niklaus (1995). *Digital Circuit Design An Introduction Textbook*. Springer. ISBN 978-3-540-58577-0.
- Mitra, Jubin (2018). "An FPGA-Based Phase Measurement System" (<https://doi.org/10.1109%2FTVLSI.2017.2758807>). *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*. **26**. IEEE: 133–142. doi:[10.1109/TVLSI.2017.2758807](https://doi.org/10.1109%2FTVLSI.2017.2758807) (<https://api.semanticscholar.org/CorpusID:4920719>).
- Mencer, Oskar et al. (2020). "The history, status, and future of FPGAs". *Communications of the ACM*. ACM. Vol. 63, No. 10. doi:[10.1145/3410669](https://doi.org/10.1145/3410669)

External links

- [What is an FPGA?](https://www.youtube.com/watch?v=gUsHwi4M4xE) (<https://www.youtube.com/watch?v=gUsHwi4M4xE>) on YouTube
- [Migrating from MCU to FPGA](https://www.ikallogic.com/2023/02/18/migrating-from-mcu-to-fpga.html) (<https://www.ikallogic.com/2023/02/18/migrating-from-mcu-to-fpga.html>)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Field-programmable_gate_array&oldid=1292004873"



Field-programmable object array

A **field-programmable object array (FPOA)** is a class of programmable logic devices designed to be modified or programmed after manufacturing. They are designed to bridge the gap between ASIC and FPGA. They contain a grid of programmable silicon objects. Arrix range of FPOA contained three types of silicon objects: arithmetic logic units (ALUs), register files (RFs) and multiply-and-accumulate units (MACs). Both the objects and interconnects are programmable.

Motivation and history

The device was intended to bridge the gap between field-programmable gate arrays (FPGAs) and application-specific integrated circuits (ASICs). The design goal was to combine the programmability of FPGAs and the performance of ASICs. FPGAs, although programmable, lack performance; they may only be clocked to few hundreds of megahertz and most FPGAs operated below 100 MHz. FPGAs did not offer deterministic timing and the maximum operating frequency depends on the design. ASICs offered good performance, but they could not be modified and they were very costly. The FPOA had a programmable architecture, deterministic timing, and gigahertz performance. The FPOA was designed by Douglas Pihl who had this idea when working on a DARPA funded project.^[1] He founded MathStar in 1997 to manufacture FPOAs and the idea was patented in 2004. The first FPOA prototypes were made in 2005 and first batch of FPOA chips were fabricated in 2006.^[2]

Architecture

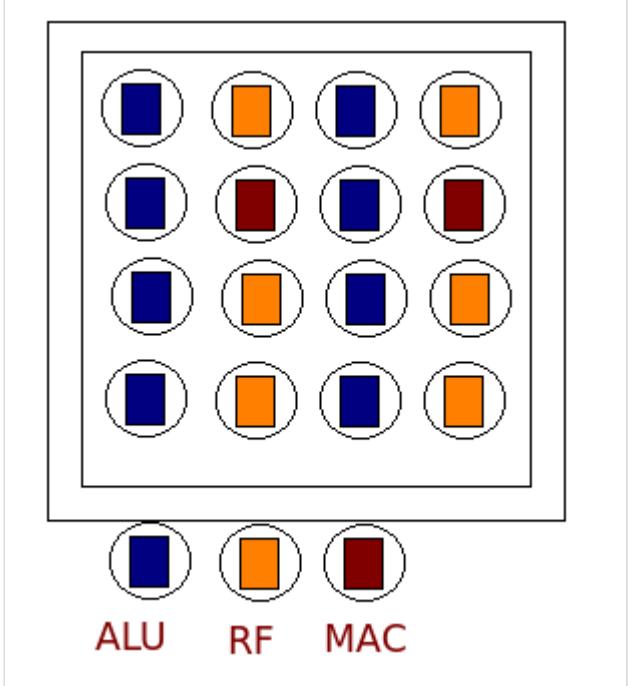
FPOAs have a core grid of silicon objects or core objects. These objects are connected through a synchronous interconnect. Each core object also has a supporting structures for clock synchronization, BIST and the like. The core is surrounded by peripheral circuitry that contains memory and I/O. An interface circuitry connects the objects to rest of FPOA. Exact number of each type of object and its arrangement are specific to a given family. There are two types of communication: nearest member and "party-line". Nearest member is used to connect a core to nearest core object and party line is used to connect remote objects. There are 8 nearest neighbor interconnects per object and offers transmission speed on one object hop per clock cycle. There are 10 party line interconnect per object that offers transmission speed of four object hops per clock cycle.^[3]

Applications

FPOAs may be used almost anywhere an FPGA is used, broadly in all hardware acceleration tasks including digital signal processing, medical imaging, computer vision, speech recognition, cryptography, bioinformatics, computer hardware emulation, and aerospace. Since FPOAs are built around fast and optimized silicon objects, they offer higher performance in flat field error correction, fast Fourier transform computation, medical imaging, machine vision, image encoding and decoding, video encoding and decoding and artificial intelligence acceleration to name a few.^[4]

Development on FPOA

In FPOA we work at silicon object level a higher level than the gate level used in FPGA. This eases the learning curve and also speeds up development. Programming is done in System C. The Arrix family released in 2006 was supported by FPOA design software, which enabled designers to create, verify, program and debug their algorithms on the devices. Summit Design's Visual Elite tool was used for behavioural simulation. MathStar's COAST (COnection and ASsignment Tool) offered a graphical environment for floor-planning and placement it compiled to an intermediate code that maps to hardware resources. The Object compiler generated the file to be loaded into the FPGA.^[5] In 2007 MathStar struck a partnership with mentor graphics and subsequent release use Visual Elite editor from Mentor Graphics for behavioural simulation and functional verification.^[6] FPOAs also offered IP core library IP partners included professionals in the video market as well as machine vision market.



Simplified illustration of FPOA architecture. The area between the rectangles forms peripheral circuitry and the oval around the object interface it to the rest of FPOA.

Present status

MathStar the producer of FPOAs never posted a profit and the company decided to shut down production in May 2008.^[7] MathStar was merged into Sajan Inc. in 2010 and Sajan thus acquired MathStar's patent including that of FPOAs. In November 2011, Sajan sold several of MathStar's patent including some on FPOAs to OLK Grun GmbH.^[8]



References

1. "Computational vision: Testing MathStar's field-programmable object arrays" (<http://www.edn.com/design/test-and-measurement/4379581/Computational-vision-Testing-MathStar-s-field-programmable-object-arrays>). EDN Networks, August 14, 2013.
2. "Radiation Hardened FPOA for space processing" (https://nepp.nasa.gov/mafa/talks/MAFA07_12_Lupia.pdf) (PDF). MAFA 2007. Retrieved August 14, 2013.
3. "Patent application 0070247189" (<https://archive.today/20130817200553/http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netacgi/PTO/src/hnum.html&r=1&f=G&l=50&s1=20070247189.PGNR.>). US Patent and Trademark Office. Archived from the original (<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netacgi/PTO/src/hnum.html&r=1&f=G&l=50&s1=20070247189.PGNR.>).

HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20070247189.
PGNR.) on August 17, 2013. Retrieved August 14, 2013.

4. "Computational vision: Testing MathStar's field-programmable object arrays" (<http://www.edn.com/design/test-and-measurement/4379581/Computational-vision-Testing-MathStar-s-field-programmable-object-arrays>). EDN Networks. Retrieved August 14, 2013.
5. "Device Spotlight:MathStar Arrix FPOA" (<https://web.archive.org/web/20120613194525/http://www.nuvation.com/corporate/news/newsletter/fall2007/devicespotlight.html>). nuvation. Archived from the original (<http://www.nuvation.com/corporate/news/newsletter/fall2007/devicespotlight.html>) on June 13, 2012. Retrieved August 14, 2013.
6. "Mentor, MathStar partner on FPOA design tools" (http://www.eetimes.com/document.asp?doc_id=1165406). EE Times. Retrieved August 14, 2013.
7. *MathStar seems to have gone dark* (https://archive.today/20130815114433/http://www.startribune.com/business/19286524.html?location_refer=Business), StarTribute, archived from the original (http://www.startribune.com/business/19286524.html?location_refer=Business) on August 15, 2013, retrieved August 14, 2013
8. "Sajan announces sale of former MathStar intellectual-property" (<https://web.archive.org/web/20120319221032/http://www.sajan.com/news-resources/news-resources/article/sajan-announces-sale-of-former-mathstar-intellectual-property.html>). Sajan. Archived from the original (<http://www.sajan.com/news-resources/news-resources/article/sajan-announces-sale-of-former-mathstar-intellectual-property.html>) on March 19, 2012. Retrieved August 14, 2013.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Field-programmable_object_array&oldid=1264973189"

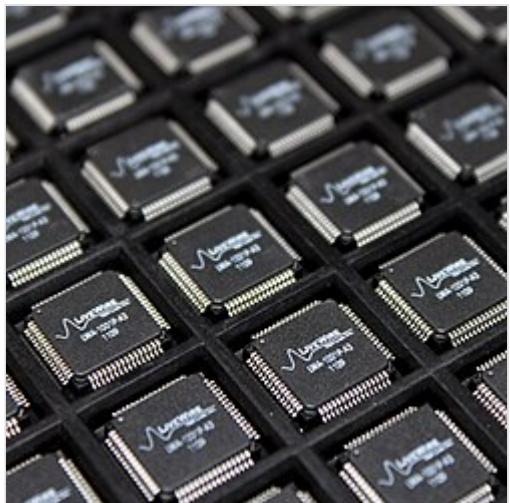


Application-specific integrated circuit

An **application-specific integrated circuit (ASIC)** /'eɪsɪk/ is an integrated circuit (IC) chip customized for a particular use, rather than intended for general-purpose use, such as a chip designed to run in a digital voice recorder or a high-efficiency video codec.^[1] Application-specific standard product chips are intermediate between ASICs and industry standard integrated circuits like the 7400 series or the 4000 series.^[2] ASIC chips are typically fabricated using metal-oxide-semiconductor (MOS) technology, as MOS integrated circuit chips.^[3]

As feature sizes have shrunk and chip design tools improved over the years, the maximum complexity (and hence functionality) possible in an ASIC has grown from 5,000 logic gates to over 100 million. Modern ASICs often include entire microprocessors, memory blocks including ROM, RAM, EEPROM, flash memory and other large building blocks. Such an ASIC is often termed a SoC (system-on-chip). Designers of digital ASICs often use a hardware description language (HDL), such as Verilog or VHDL, to describe the functionality of ASICs.^[2]

Field-programmable gate arrays (FPGA) are the modern-day technology improvement on breadboards, meaning that they are not made to be application-specific as opposed to ASICs. Programmable logic blocks and programmable interconnects allow the same FPGA to be used in many different applications. For smaller designs or lower production volumes, FPGAs may be more cost-effective than an ASIC design, even in production. The non-recurring engineering (NRE) cost of an ASIC can run into the millions of dollars. Therefore, device manufacturers typically prefer FPGAs for prototyping and devices with low production volume and ASICs for very large production volumes where NRE costs can be amortized across many devices.^[4]



A tray of application-specific integrated circuit (ASIC) chips



A packet processing ASIC inside an Ethernet switch

History

Early ASICs used gate array technology. By 1967, Ferranti and Interdesign were manufacturing early bipolar gate arrays. In 1967, Fairchild Semiconductor introduced the Micromatrix family of bipolar diode-transistor logic (DTL) and transistor-transistor logic (TTL) arrays.^[3]

Complementary metal–oxide–semiconductor (CMOS) technology opened the door to the broad commercialization of gate arrays. The first CMOS gate arrays were developed by Robert Lipp,^{[5][6]} in 1974 for International Microcircuits, Inc. (IMI).^[3]

Metal–oxide–semiconductor (MOS) standard-cell technology was introduced by Fairchild and Motorola, under the trade names Micromosaic and Polycell, in the 1970s. This technology was later successfully commercialized by VLSI Technology (founded 1979) and LSI Logic (1981).^[3]

A successful commercial application of gate array circuitry was found in the low-end 8-bit ZX81 and ZX Spectrum personal computers, introduced in 1981 and 1982. These were used by Sinclair Research (UK) essentially as a low-cost I/O solution aimed at handling the computer's graphics.

Customization occurred by varying a metal interconnect mask. Gate arrays had complexities of up to a few thousand gates; this is now called mid-scale integration. Later versions became more generalized, with different base dies customized by both metal and polysilicon layers. Some base dies also include random-access memory (RAM) elements.

Standard-cell designs

In the mid-1980s, a designer would choose an ASIC manufacturer and implement their design using the design tools available from the manufacturer. While third-party design tools were available, there was not an effective link from the third-party design tools to the layout and actual semiconductor process performance characteristics of the various ASIC manufacturers. Most designers used factory-specific tools to complete the implementation of their designs. A solution to this problem, which also yielded a much higher density device, was the implementation of standard cells.^[7] Every ASIC manufacturer could create functional blocks with known electrical characteristics, such as propagation delay, capacitance and inductance, that could also be represented in third-party tools. Standard-cell design is the utilization of these functional blocks to achieve very high gate density and good electrical performance. Standard-cell design is intermediate between § Gate-array and semi-custom design and § Full-custom design in terms of its non-recurring engineering and recurring component costs as well as performance and speed of development (including time to market).

By the late 1990s, logic synthesis tools became available. Such tools could compile HDL descriptions into a gate-level netlist. Standard-cell integrated circuits (ICs) are designed in the following conceptual stages referred to as electronics design flow, although these stages overlap significantly in practice:

1. **Requirements engineering:** A team of design engineers starts with a non-formal understanding of the required functions for a new ASIC, usually derived from requirements analysis.
2. **Register-transfer level (RTL) design:** The design team constructs a description of an ASIC to achieve these goals using a hardware description language. This process is similar to writing a computer program in a high-level language.
3. **Functional verification:** Suitability for purpose is verified by functional verification. This may include such techniques as logic simulation through test benches, formal verification, emulation, or creating and evaluating an equivalent pure software model, as in Simics. Each verification technique has advantages and disadvantages, and most often several methods are used together for ASIC verification. Unlike most FPGAs, ASICs cannot be

reprogrammed once fabricated and therefore ASIC designs that are not completely correct are much more costly, increasing the need for full test coverage.

4. **Logic synthesis:** Logic synthesis transforms the RTL design into a large collection called of lower-level constructs called standard cells. These constructs are taken from a standard-cell library consisting of pre-characterized collections of logic gates performing specific functions. The standard cells are typically specific to the planned manufacturer of the ASIC. The resulting collection of standard cells and the needed electrical connections between them is called a gate-level netlist.
5. **Placement:** The gate-level netlist is next processed by a placement tool which places the standard cells onto a region of an integrated circuit die representing the final ASIC. The placement tool attempts to find an optimized placement of the standard cells, subject to a variety of specified constraints.
6. **Routing:** An electronics routing tool takes the physical placement of the standard cells and uses the netlist to create the electrical connections between them. Since the search space is large, this process will produce a "sufficient" rather than "globally optimal" solution. The output is a file which can be used to create a set of photomasks enabling a semiconductor fabrication facility, commonly called a "fab" or "foundry" to manufacture physical integrated circuits. Placement and routing are closely interrelated and are collectively called place and route in electronics design.
7. **Sign-off:** Given the final layout, circuit extraction computes the parasitic resistances and capacitances. In the case of a digital circuit, this will then be further mapped into delay information from which the circuit performance can be estimated, usually by static timing analysis. This, and other final tests such as design rule checking and power analysis collectively called signoff are intended to ensure that the device will function correctly over all extremes of the process, voltage and temperature. When this testing is complete the photomask information is released for chip fabrication.

These steps, implemented with a level of skill common in the industry, almost always produce a final device that correctly implements the original design, unless flaws are later introduced by the physical fabrication process.^[8]

The design steps also called design flow, are also common to standard product design. The significant difference is that standard-cell design uses the manufacturer's cell libraries that have been used in potentially hundreds of other design implementations and therefore are of much lower risk than a full custom design. Standard cells produce a design density that is cost-effective, and they can also integrate IP cores and static random-access memory (SRAM) effectively, unlike gate arrays.

Gate-array and semi-custom design

Gate array design is a manufacturing method in which diffused layers,^[9] each consisting of transistors and other active devices, are predefined and electronics wafers containing such devices are "held in stock" or unconnected prior to the metallization stage of the fabrication process. The physical design process defines the interconnections of these layers for the final device. For most ASIC manufacturers, this consists of between two and nine metal layers with each layer running perpendicular to the one below it. Non-recurring engineering costs are much lower than full custom designs, as photolithographic masks are required only for the metal layers. Production cycles are much shorter, as metallization is a comparatively quick process; thereby accelerating time to market.

Gate-array ASICs are always a compromise between rapid design and performance as mapping a given design onto what a manufacturer held as a stock wafer never gives 100% circuit utilization. Often difficulties in routing the interconnect require migration onto a larger array device with a consequent increase in the piece part price. These difficulties are often a result of the layout EDA software used to develop the interconnect.

Pure, logic-only gate-array design is rarely implemented by circuit designers today, having been almost entirely replaced by field-programmable devices. The most prominent of such devices are field-programmable gate arrays (FPGAs) which can be programmed by the user and thus offer minimal tooling charges, non-recurring engineering, only marginally increased piece part cost, and comparable performance.

Today, gate arrays are evolving into structured ASICs that consist of a large IP core like a CPU, digital signal processor units, peripherals, standard interfaces, integrated memories, SRAM, and a block of reconfigurable, uncommitted logic. This shift is largely because ASIC devices are capable of integrating large blocks of system functionality, and systems on a chip (SoCs) require glue logic, communications subsystems (such as networks on chip), peripherals, and other components rather than only functional units and basic interconnection.

In their frequent usages in the field, the terms "gate array" and "semi-custom" are synonymous when referring to ASICs. Process engineers more commonly use the term "semi-custom", while "gate-array" is more commonly used by logic (or gate-level) designers.

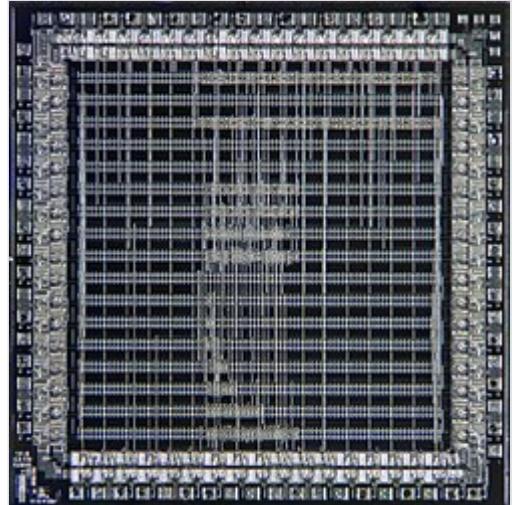
Full-custom design

By contrast, full-custom ASIC design defines all the photolithographic layers of the device.^[7] Full-custom design is used for both ASIC design and for standard product design.

The benefits of full-custom design include reduced area (and therefore recurring component cost), performance improvements, and also the ability to integrate analog components and other pre-designed—and thus fully verified—components, such as microprocessor cores, that form a system on a chip.

The disadvantages of full-custom design can include increased manufacturing and design time, increased non-recurring engineering costs, more complexity in the computer-aided design (CAD) and electronic design automation systems, and a much higher skill requirement on the part of the design team.

For digital-only designs, however, "standard-cell" cell libraries, together with modern CAD systems, can offer considerable performance/cost benefits with low risk. Automated layout tools are quick and easy to use and also offer the possibility to "hand-tweak" or manually optimize any performance-limiting aspect of the design.



Microscope photograph of a gate-array ASIC showing the predefined logic cells and custom interconnections. This particular design uses less than 20% of available logic gates.

This is designed by using basic logic gates, circuits or layout specially for a design.

Structured design

Structured ASIC design (also referred to as "*platform ASIC design*") is a relatively new trend in the semiconductor industry, resulting in some variation in its definition. However, the basic premise of a structured ASIC is that both manufacturing cycle time and design cycle time are reduced compared to cell-based ASIC, by virtue of there being pre-defined metal layers (thus reducing manufacturing time) and pre-characterization of what is on the silicon (thus reducing design cycle time).

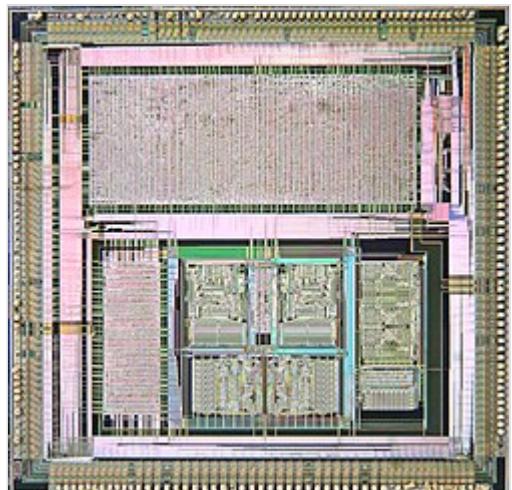
Definition from Foundations of Embedded Systems states that:^[10]

In a "structured ASIC" design, the logic mask-layers of a device are predefined by the ASIC vendor (or in some cases by a third party). Design differentiation and customization is achieved by creating custom metal layers that create custom connections between predefined lower-layer logic elements. "Structured ASIC" technology is seen as bridging the gap between field-programmable gate arrays and "standard-cell" ASIC designs. Because only a small number of chip layers must be custom-produced, "structured ASIC" designs have much smaller non-recurring expenditures (NRE) than "standard-cell" or "full-custom" chips, which require that a full mask set be produced for every design.

—Foundations of Embedded Systems

This is effectively the same definition as a gate array. What distinguishes a structured ASIC from a gate array is that in a gate array, the predefined metal layers serve to make manufacturing turnaround faster. In a structured ASIC, the use of predefined metallization is primarily to reduce cost of the mask sets as well as making the design cycle time significantly shorter.

For example, in a cell-based or gate-array design the user must often design power, clock, and test structures themselves. By contrast, these are predefined in most structured ASICs and therefore can save time and expense for the designer compared to gate-array based designs. Likewise, the design tools used for structured ASIC can be substantially lower cost and easier (faster) to use than cell-based tools, because they do not have to perform all the functions that cell-based tools do. In some cases, the structured ASIC vendor requires customized tools for their device (e.g., custom physical synthesis) be used, also allowing for the design to be brought into manufacturing more quickly.



Microscope photograph of custom ASIC (486 chipset) showing gate-based design on top and custom circuitry on bottom

Cell libraries, IP-based design, hard and soft macros

Cell libraries of logical primitives are usually provided by the device manufacturer as part of the service. Although they will incur no additional cost, their release will be covered by the terms of a non-disclosure agreement (NDA) and they will be regarded as intellectual property by the manufacturer. Usually, their physical design will be pre-defined so they could be termed "hard macros".

What most engineers understand as "intellectual property" are IP cores, designs purchased from a third-party as sub-components of a larger ASIC. They may be provided in the form of a hardware description language (often termed a "soft macro"), or as a fully routed design that could be printed directly onto an ASIC's mask (often termed a "hard macro"). Many organizations now sell such pre-designed cores – CPUs, Ethernet, USB or telephone interfaces – and larger organizations may have an entire department or division to produce cores for the rest of the organization. The company ARM *only* sells IP cores, making it a fabless manufacturer.

Indeed, the wide range of functions now available in structured ASIC design is a result of the phenomenal improvement in electronics in the late 1990s and early 2000s; as a core takes a lot of time and investment to create, its re-use and further development cuts product cycle times dramatically and creates better products. Additionally, open-source hardware organizations such as OpenCores are collecting free IP cores, paralleling the open-source software movement in hardware design.

Soft macros are often process-independent (i.e. they can be fabricated on a wide range of manufacturing processes and different manufacturers). Hard macros are process-limited and usually further design effort must be invested to migrate (port) to a different process or manufacturer.

Multi-project wafers

Some manufacturers and IC design houses offer multi-project wafer service (MPW) as a method of obtaining low cost prototypes. Often called shuttles, these MPWs, containing several designs, run at regular, scheduled intervals on a "cut and go" basis, usually with limited liability on the part of the manufacturer. The contract involves delivery of bare dies or the assembly and packaging of a handful of devices. The service usually involves the supply of a physical design database (i.e. masking information or pattern generation (PG) tape). The manufacturer is often referred to as a "silicon foundry" due to the low involvement it has in the process.

Application-specific standard product

An **application-specific standard product** or ASSP is an integrated circuit that implements a specific function that appeals to a wide market. As opposed to ASICs that combine a collection of functions and are designed by or for one customer, ASSPs are available as off-the-shelf components. ASSPs are used in all industries, from automotive to communications.^[11]

For example, two ICs that might or might not be considered ASICs are a controller chip for a PC and a chip for a modem. Both of these examples are specific to an application (which is typical of an ASIC) but are sold to many different system vendors (which is typical of standard parts). ASICs such as these are

sometimes called application-specific standard products (ASSPs).

Examples of ASSPs are encoding/decoding chip, Ethernet network interface controller chip and flash memory controller chip.^[12]

See also

- Application-specific instruction set processor (ASIP)
- Complex programmable logic device (CPLD)
- Electronic design automation (EDA or ECAD)
- Field-programmable gate array (FPGA)
- Multi-project chip (MPC)
- Very Large Scale Integration (VLSI)
- System on a chip (SoC)
- Hardware acceleration for an overview of computing based primarily in hardware



Renesas M66591GP: USB2.0 Peripheral Controller

References

1. Golshan, Khosrow (2007). *Physical Design Essentials: An ASIC Design Implementation Perspective*. Boston, MA: Springer. ISBN 978-0-387-36642-5.
2. Barr, Keith (2007). *ASIC Design in the Silicon Sandbox: A Complete Guide to Building Mixed-signal Integrated Circuits* (<https://archive.org/details/asicdesigninsili0000barr>). New York: McGraw-Hill. ISBN 978-0-07-148161-8. OCLC 76935560 (<https://search.worldcat.org/oclc/76935560>).
3. "1967: Application Specific Integrated Circuits employ Computer-Aided Design" (<https://www.computerhistory.org/siliconengine/application-specific-integrated-circuits-employ-computer-aided-design/>). *The Silicon Engine*. Computer History Museum. Retrieved 9 November 2019.
4. Kriegbaum, Jeff (13 September 2004). "FPGA's vs. ASIC's" (<https://www.eetimes.com/fpgas-vs-asics/>). *EE Times*.
5. Lipp, Bob oral history (<http://www.computerhistory.org/collections/catalog/102706880>). Computer History Museum. 14 February 2017. Retrieved 28 January 2018. {{cite book}}: |website= ignored (help)
6. "People" (<http://www.computerhistory.org/siliconengine/people/>). *The Silicon Engine*. Computer History Museum. Retrieved 28 January 2018.
7. Smith, Michael John Sebastian (1997). *Application-Specific Integrated Circuits*. Addison-Wesley Professional. ISBN 978-0-201-50022-6.
8. Hurley, Jaden Mclean & Carmen. (2019). *Logic Design*. EDTECH. ISBN 978-1-83947-319-7. OCLC 1132366891 (<https://search.worldcat.org/oclc/1132366891>).
9. Grierson, J. R. (July 1983). "The Use of Gate Arrays in Telecommunications" (<https://archive.org/details/bte-198307/page/n19/mode/2up>). *British Telecommunications Engineering*. 2 (2): 78–80. ISSN 0262-401X (<https://search.worldcat.org/issn/0262-401X>). Retrieved 26 February 2021. "In the UK, Ferranti, with their bipolar collector diffused isolation (CDI) arrays, pioneered the commercial use of gate arrays and for many years this was by far the most widely used technology."

10. Barkalov, Alexander; Titarenko, Larysa; Mazurkiewicz, Małgorzata (2019). *Foundations of Embedded Systems*. Studies in Systems, Decision and Control. Vol. 195. Cham: Springer International Publishing. doi:10.1007/978-3-030-11961-4 (<https://doi.org/10.1007%2F978-3-030-11961-4>). ISBN 9783030119607. S2CID 86596100 (<https://api.semanticscholar.org/CorpusID:86596100>).
11. Maxfield, Max (23 June 2014). "ASIC, ASSP, SoC, FPGA – What's the Difference?" (<https://www.eetimes.com/asic-assp-soc-fpga-whats-the-difference/>). *EE Times*. Retrieved 2 February 2025.
12. "EP501: NAND Flash Controller" (<https://www.latticesemi.com/products/designsoftwareandip/intellectualproperty/ipcore/eurekacores/ep501nandflashcontroller>). *Lattice Semiconductor*. Archived (<http://web.archive.org/web/20240418010628/https://www.latticesemi.com/products/designsoftwareandip/intellectualproperty/ipcore/eurekacores/ep501nandflashcontroller>) from the original on 18 April 2024. Retrieved 8 May 2025.

Sources

- Anthony Cataldo (26 March 2002). "Xilinx looks to ease path to custom FPGAs" (<https://web.archive.org/web/20070929100135/http://www.eetimes.com/story/OEG20020325S0060>). *EE Times*. CMP Media, LLC. Archived from the original (<http://www.eetimes.com/story/OEG20020325S0060>) on 29 September 2007. Retrieved 14 December 2006.
- "Xilinx intros next-gen EasyPath FPGAs priced below structured ASICs" (<http://www.thefreelibrary.com/Xilinx+intros+next-gen+EasyPath+FPGAs+priced+below+structured+ASICs.-a0125948398>). *EDP Weekly's IT Monitor*. Millin Publishing, Inc. 18 October 2004.

External links

-  Media related to [Application-specific integrated circuits](#) at Wikimedia Commons

Retrieved from "https://en.wikipedia.org/w/index.php?title=Application-specific_integrated_circuit&oldid=1292066106"

Tensor Processing Unit

Tensor Processing Unit (TPU) is an [AI accelerator](#) [application-specific integrated circuit](#) (ASIC) developed by [Google](#) for [neural network machine learning](#), using Google's own [TensorFlow](#) software.^[2] Google began using TPUs internally in 2015, and in 2018 made them available for [third-party](#) use, both as part of its cloud infrastructure and by offering a smaller version of the chip for sale.

Comparison to CPUs and GPUs

Compared to a [graphics processing unit](#), TPUs are designed for a high volume of low [precision](#) computation (e.g. as little as 8-bit precision)^[3] with more input/output operations per [joule](#), without hardware for rasterisation/texture mapping.^[4] The TPU ASICs are mounted in a heatsink assembly, which can fit in a hard drive slot within a data center [rack](#), according to Norman Jouppi.^[5]

Different types of processors are suited for different types of machine learning models. TPUs are well suited for [CNNs](#), while GPUs have benefits for some fully-connected neural networks, and CPUs can have advantages for [RNNs](#).^[6]

History

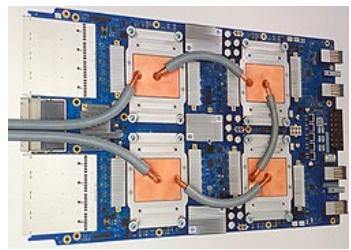
According to Jonathan Ross, one of the original TPU engineers,^[1] and later the founder of [Groq](#), three separate groups at Google were developing AI accelerators, with the TPU being the design that was ultimately selected. He was not aware of [systolic arrays](#) at the time and upon learning the term thought "Oh, that's called a systolic array? It just seemed to make sense."^[7]

The tensor processing unit was announced in May 2016 at [Google I/O](#), when the company said that the TPU had already been used inside [their data centers](#) for over a year.^{[5][4]} Google's 2017 paper describing its creation cites previous systolic matrix multipliers of similar architecture built in the 1990s.^[8] The chip has been specifically designed for Google's [TensorFlow](#) framework, a symbolic math library which is used for [machine learning](#) applications such as [neural networks](#).^[9] However, as of 2017 Google still used [CPUs](#) and [GPUs](#) for other types of [machine learning](#).^[5] Other [AI accelerator](#) designs are appearing from other vendors also and are aimed at [embedded](#) and [robotics](#) markets.

Google's TPUs are proprietary. Some models are commercially available, and on February 12, 2018, *The New York Times* reported that Google "would allow other companies to buy access to those chips through its cloud-computing service."^[10] Google has said that they were used in the [AlphaGo](#) versus Lee Sedol series of human-versus-machine [Go](#) games,^[4] as well as in the [AlphaZero](#) system, which produced [Chess](#), [Shogi](#) and [Go](#) playing programs from the game rules alone and went on to beat the leading programs in those games.^[11] Google has also used TPUs for [Google Street View](#) text processing and was able to find all the text in the Street View database in less than five days. In [Google Photos](#), an individual TPU can process over 100 million photos a day.^[5] It is also used in [RankBrain](#) which Google uses to provide search results.^[12]

Google provides third parties access to TPUs through its [Cloud TPU](#) service as part of the [Google Cloud Platform](#)^[13] and through its [notebook-based](#) services [Kaggle](#) and [Colaboratory](#).^{[14][15]}

Tensor Processing Unit



Tensor Processing Unit 3.0

Designer	Google
Introduced	2015 ^[1]
Type	Neural network Machine learning

Products

Tensor Processing Unit products ^{[16][17][18]}								
	TPUv1	TPUv2	TPUv3	TPUv4 ^{[17][19][20]}	TPUv5e ^[21]	TPUv5p ^{[22][23]}	v6e (Trillium) ^{[24][25]}	TPU v7 (Ironwood) ^[26]
Date introduced	2015	2017	2018	2021	2023	2023	2024	2025
Process node	28 nm	16 nm	16 nm	7 nm	Not Listed	Not Listed	Not Listed	Not Listed
Die size (mm ²)	331	< 625	< 700	< 400	300-350	Not Listed	Not Listed	Not Listed
On-chip memory (MiB)	28	32	32 (VMMEM) + 5 (spMEM)	128 (CMEM) + 32 (VMMEM) + 10 (spMEM)	48	112	Not Listed	Not Listed
Clock speed (MHz)	700	700	940	1050	Not Listed	1750	Not Listed	Not Listed
Memory	8 GiB DDR3	16 GiB HBM	32 GiB HBM	32 GiB HBM	16 GB HBM	95 GB HBM	32 GB	192 GB HBM
Memory bandwidth	34 GB/s	600 GB/s	900 GB/s	1200 GB/s	819 GB/s	2765 GB/s	1640 GB/s	7.2 TB/s
TDP (W)	75	280	220	170	Not Listed	Not Listed	Not Listed	Not Listed
TOPS (Tera Operations Per Second)	23	45	123	275	197 (bf16) 393 (int8)	459 (bf16) 918 (int8)	918 (bf16) 1836 (int8)	4614 (fp8)
TOPS/W	0.31	0.16	0.56	1.62	Not Listed	Not Listed	Not Listed	4.7

First generation TPU

The first-generation TPU is an 8-bit [matrix multiplication](#) engine, driven with [CISC instructions](#) by the host processor across a [PCIe 3.0](#) bus. It is manufactured on a 28 nm process with a die size $\leq 331 \text{ mm}^2$. The [clock speed](#) is 700 MHz and it has a [thermal design power](#) of 28–40 W. It has 28 MiB of on chip memory, and 4 MiB of 32-bit accumulators taking the results of a 256×256 systolic array of 8-bit [multipliers](#).^[8] Within the TPU package is 8 GiB of dual-channel 2133 MHz [DDR3 SDRAM](#) offering 34 GB/s of bandwidth.^[18] Instructions transfer data to or from the host, perform matrix multiplications or convolutions, and apply [activation functions](#).^[8]

Second generation TPU

The second-generation TPU was announced in May 2017.^[27] Google stated the first-generation TPU design was limited by [memory bandwidth](#) and using 16 GB of [High Bandwidth Memory](#) in the second-generation design increased bandwidth to 600 GB/s and performance to 45 teraFLOPS.^[18] The TPUs are then arranged into four-chip modules with a performance of 180 teraFLOPS.^[27] Then 64 of these modules are assembled into 256-chip pods with 11.5 petaFLOPS of performance.^[27] Notably, while the first-generation TPUs were limited to integers, the second-generation TPUs can also calculate in [floating point](#), introducing the [bfloating16](#) format invented by [Google Brain](#). This makes the second-generation TPUs useful for both training and inference of machine learning models. Google has stated these second-generation TPUs will be available on the [Google Compute Engine](#) for use in TensorFlow applications.^[28]

Third generation TPU

The third-generation TPU was announced on May 8, 2018.^[29] Google announced that processors themselves are twice as powerful as the second-generation TPUs, and would be deployed in pods with four times as many chips as the preceding generation.^{[30][31]} This results in an 8-fold increase in performance per pod (with up to 1,024 chips per pod) compared to the second-generation TPU deployment.

Fourth generation TPU

On May 18, 2021, Google CEO Sundar Pichai spoke about TPU v4 Tensor Processing Units during his keynote at the Google I/O virtual conference. TPU v4 improved performance by more than 2x over TPU v3 chips. Pichai said "A single v4 pod contains 4,096 v4 chips, and each pod has 10x the interconnect bandwidth per chip at scale, compared to any other networking technology."^[32] An April 2023 paper by Google claims TPU v4 is 5–87% faster than an Nvidia A100 at machine learning [benchmarks](#).^[33]

There is also an "inference" version, called v4i,^[34] that does not require [liquid cooling](#).^[35]

Fifth generation TPU

In 2021, Google revealed the physical layout of TPU v5 is being designed with the assistance of a novel application of deep reinforcement learning.^[36] Google claims TPU v5 is nearly twice as fast as TPU v4,^[37] and based on that and the relative performance of TPU v4 over A100, some speculate TPU v5 as being as fast as or faster than an H100.^[38]

Similar to the v4i being a lighter-weight version of the v4, the fifth generation has a "cost-efficient"^[39] version called v5e.^[21] In December 2023, Google announced TPU v5p which is claimed to be competitive with the H100.^[40]

Sixth generation TPU

In May 2024, at the Google I/O conference, Google announced TPU v6, which became available in preview in October 2024.^[41] Google claimed a 4.7 times performance increase relative to TPU v5e.^[42] via larger matrix multiplication units and an increased clock speed. High bandwidth memory (HBM) capacity and bandwidth have also doubled. A pod can contain up to 256 Trillium units.^[43]

Seventh generation TPU

In April 2025, at Google Cloud Next conference, Google unveiled TPU v7. This new chip, called Ironwood, will come in two configurations: a 256-chip cluster and a 9,216-chip cluster. Ironwood will have a peak computational performance rate of 4,614 TFLOP/s.^[44]

Edge TPU

In July 2018, Google announced the Edge TPU. The Edge TPU is Google's purpose-built ASIC chip designed to run machine learning (ML) models for edge computing, meaning it is much smaller and consumes far less power compared to the TPUs hosted in Google datacenters (also known as Cloud TPUs)^[45]. In January 2019, Google made the Edge TPU available to developers with a line of products under the Coral brand. The Edge TPU is capable of 4 trillion operations per second with 2 W of electrical power.^[46]

The product offerings include a single-board computer (SBC), a system on module (SoM), a USB accessory, a mini PCI-e card, and an M.2 card. The SBC Coral Dev Board and Coral SoM both run Mendel Linux OS – a derivative of Debian.^{[47][48]} The USB, PCI-e, and M.2 products function as add-ons to existing computer systems, and support Debian-based Linux systems on x86-64 and ARM64 hosts (including Raspberry Pi).

The machine learning runtime used to execute models on the Edge TPU is based on TensorFlow Lite.^[49] The Edge TPU is only capable of accelerating forward-pass operations, which means it's primarily useful for performing inferences (although it is possible to perform lightweight transfer learning on the Edge TPU^[50]). The Edge TPU also only supports 8-bit math, meaning that for a network to be compatible with the Edge TPU, it needs to either be trained using the TensorFlow quantization-aware training technique, or since late 2019 it's also possible to use post-training quantization.

On November 12, 2019, Asus announced a pair of single-board computer (SBCs) featuring the Edge TPU. The Asus Tinker Edge T and Tinker Edge R Board designed for IoT and edge AI. The SBCs officially support Android and Debian operating systems.^{[51][52]} ASUS has also demonstrated a mini PC called Asus PN60T featuring the Edge TPU.^[53]

On January 2, 2020, Google announced the Coral Accelerator Module and Coral Dev Board Mini, to be demonstrated at CES 2020 later the same month. The Coral Accelerator Module is a multi-chip module featuring the Edge TPU, PCIe and USB interfaces for easier integration. The Coral Dev Board Mini is a smaller SBC featuring the Coral Accelerator Module and MediaTek 8167s SoC.^{[54][55]}

Pixel Neural Core

On October 15, 2019, Google announced the Pixel 4 smartphone, which contains an Edge TPU called the Pixel Neural Core. Google describe it as "customized to meet the requirements of key camera features in Pixel 4", using a neural network search that sacrifices some accuracy in favor of minimizing latency and power use.^[56]

Google Tensor

Google followed the Pixel Neural Core by integrating an Edge TPU into a custom system-on-chip named Google Tensor, which was released in 2021 with the Pixel 6 line of smartphones.^[57] The Google Tensor SoC demonstrated "extremely large performance advantages over the competition" in machine learning-focused benchmarks; although instantaneous power consumption also was relatively high, the improved performance meant less energy was consumed due to shorter periods requiring peak performance.^[58]

Lawsuit

In 2019, Singular Computing, founded in 2009 by Joseph Bates, a visiting professor at MIT,^[59] filed suit against Google alleging patent infringement in TPU chips.^[60] By 2020, Google had successfully lowered the number of claims the court would consider to just two: claim 53 of US 8407273 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=US8407273>) filed in 2012 and claim 7 of US 9218156 (<https://worldwide.espacenet.com/textdoc?DB=EPODOC&IDX=US9218156>) filed in 2013, both of which claim a dynamic range of 10^{-6} to 10^6 for floating point numbers, which the standard float16 cannot do (without resorting to subnormal numbers) as it only has five bits for the exponent. In a 2023 court filing, Singular Computing specifically called out Google's use of bfloat16, as that exceeds the dynamic range of float16.^[61] Singular claims non-standard floating point formats were non-obvious in 2009, but Google retorts that the VFLOAT^[62] format, with configurable number of exponent bits, existed as prior art in 2002.^[63] By January 2024, subsequent lawsuits by Singular had brought the number of patents being litigated up to eight. Towards the end of the trial later that month, Google agreed to a settlement with undisclosed terms.^{[64][65]}

See also

- [Cognitive computer](#)
- [AI accelerator](#)
- [Structure tensor, a mathematical foundation for TPU's](#)
- [Tensor Core, a similar architecture by Nvidia](#)
- [TrueNorth, a similar device simulating spiking neurons instead of low-precision tensors](#)
- [Vision processing unit, a similar device specialised for vision processing](#)

References

1. Jouppi, Norman; et al. (2017). "In-Datacenter Performance Analysis of a Tensor Processing Unit" (<https://dl.acm.org/doi/pdf/10.1145/3079856.3080246>). *Proceedings of the 44th annual international symposium on computer architecture*. International Symposium on Computer Architecture. Toronto: Association for Computing Machinery. pp. 1–12. arXiv:1704.04760 (<https://arxiv.org/abs/1704.04760>). doi:10.1145/3079856.3080246 (<https://doi.org/10.1145/3079856.3080246>).
2. "Cloud Tensor Processing Units (TPUs)" (<https://cloud.google.com/tpu/docs/tpus>). *Google Cloud*. Retrieved 20 July 2020.
3. Armasu, Lucian (2016-05-19). "Google's Big Chip Unveil For Machine Learning: Tensor Processing Unit With 10x Better Efficiency (Updated)" (http://www.tomshardware.com/news/google-tensor-processing-unit-machine-learning_31834.html). *Tom's Hardware*. Retrieved 2016-06-26.
4. Jouppi, Norm (May 18, 2016). "Google supercharges machine learning tasks with TPU custom chip" (<https://cloudplatform.googleblog.com/2016/05/Google-supercharges-machine-learning-tasks-with-custom-chip.html>). *Google Cloud Platform Blog*. Retrieved 2017-01-22.
5. "Google's Tensor Processing Unit explained: this is what the future of computing looks like" (<http://www.techradar.com/news/computing-components/processors/google-s-tensor-processing-unit-explained-this-is-what-the-future-of-computing-looks-like-1326915>). *TechRadar*. Retrieved 2017-01-19.
6. Wang, Yu Emma; Wei, Gu-Yeon; Brooks, David (2019-07-01). "Benchmarking TPU, GPU, and CPU Platforms for Deep Learning". arXiv:1907.10701 (<https://arxiv.org/abs/1907.10701>) [cs.LG] (<https://arxiv.org/archive/cs.LG>).
7. Tensor Processing Unit (<https://www.linkedin.com/feed/update/urn:li:activity:7269734656462180355/>) on LinkedIn
8. Jouppi, Norman P.; Young, Cliff; Patil, Nishant; Patterson, David; Agrawal, Gaurav; Bajwa, Raminder; Bates, Sarah; Bhatia, Suresh; Boden, Nan; Borchers, Al; Boyle, Rick; Cantin, Pierre-luc; Chao, Clifford; Clark, Chris; Coriell, Jeremy; Daley, Mike; Dau, Matt; Dean, Jeffrey; Gelb, Ben; Ghaemmaghami, Tara Vazir; Gottipati, Rajendra; Gulland, William; Hagmann, Robert; Ho, C. Richard; Hogberg, Doug; Hu, John; Hundt, Robert; Hurt, Dan; Ibarz, Julian; Jaffey, Aaron; Jaworski, Alek; Kaplan, Alexander; Khaitan, Harshit; Koch, Andy; Kumar, Naveen; Lacy, Steve; Laudon, James; Law, James; Le, Diemthu; Leary, Chris; Liu, Zhuyuan; Lucke, Kyle; Lundin, Alan; MacKean, Gordon; Maggiore, Adriana; Mahony, Maire; Miller, Kieran; Nagarajan, Rahul; Narayanaswami, Ravi; Ni, Ray; Nix, Kathy; Norrie, Thomas; Omernick, Mark; Penukonda, Narayana; Phelps, Andy; Ross, Jonathan; Ross, Matt; Salek, Amir; Samadiani, Emad; Severn, Chris; Sizikov, Gregory; Snelham, Matthew; Souter, Jed; Steinberg, Dan; Swing, Andy; Tan, Mercedes; Thorson, Gregory; Tian, Bo; Toma, Horia; Tuttle, Erick; Vasudevan, Vijay; Walter, Richard; Wang, Walter; Wilcox, Eric; Yoon, Doe Hyun (June 26, 2017). *In-Datacenter Performance Analysis of a Tensor Processing Unit™*. Toronto, Canada. arXiv:1704.04760 (<https://arxiv.org/abs/1704.04760>).
9. "TensorFlow: Open source machine learning" (https://www.youtube.com/watch?v=oZikw5k_2FM) "It is machine learning software being used for various kinds of perceptual and language understanding tasks" — Jeffrey Dean, minute 0:47 / 2:17 from Youtube clip
10. Metz, Cade (12 February 2018). "Google Makes Its Special A.I. Chips Available to Others" (<https://www.nytimes.com/2018/02/12/technology/google-artificial-intelligence-chips.html>). *The New York Times*. Retrieved 2018-02-12.
11. McGourty, Colin (6 December 2017). "DeepMind's AlphaZero crushes chess" (<https://chess24.com/en/read/news/deepmind-s-alpha-zero-crushes-chess>). *chess24.com*.
12. "Google's Tensor Processing Unit could advance Moore's Law 7 years into the future" (<http://www.pcworld.com/article/3072256/google-io/googles-tensor-processing-unit-said-to-advance-moores-law-seven-years-into-the-future.html>). *PCWorld*. Retrieved 2017-01-19.
13. "Frequently Asked Questions | Cloud TPU" (<https://cloud.google.com/tpu/docs/faq>). *Google Cloud*. Retrieved 2021-01-14.
14. "Google Colaboratory" (<https://colab.research.google.com/notebooks/tpu.ipynb>). *colab.research.google.com*. Retrieved 2021-05-15.
15. "Use TPUs | TensorFlow Core" (<https://www.tensorflow.org/guide/tpu>). *TensorFlow*. Retrieved 2021-05-15.
16. Jouppi, Norman P.; Yoon, Doe Hyun; Ashcraft, Matthew; Gottschos, Mark (June 14, 2021). *Ten lessons from three generations that shaped Google's TPUv4i* (<https://conferences.computer.org/iscapub/pdfs/ISCA2021-4ghucdBnCWYB7ES2Pe4YdT/333300a001/333300a001.pdf>) (PDF). International Symposium on Computer Architecture. Valencia, Spain. doi:10.1109/ISCA52012.2021.00010 (<https://doi.org/10.1109/ISCA52012.2021.00010>). ISBN 978-1-4503-9086-6.
17. "System Architecture | Cloud TPU" (<https://cloud.google.com/tpu/docs/system-architecture-tpu-vm>). *Google Cloud*. Retrieved 2022-12-11.
18. Kennedy, Patrick (22 August 2017). "Case Study on the Google TPU and GDDR5 from Hot Chips 29" (<https://www.servethehome.com/case-study-google-tpu-gddr5-hot-chips-29/>). *Serve The Home*. Retrieved 23 August 2017.
19. Stay tuned, more information on TPU v4 is coming soon (<https://cloud.google.com/blog/products/ai-machine-learning/google-breaks-ai-performance-records-in-mlperf-with-worlds-fastest-training-supercrusher>), retrieved 2020-08-06.
20. Jouppi, Norman P.; Kurian, George; Li, Sheng; Ma, Peter; Nagarajan, Rahul; Nai, Lifeng; Patil, Nishant; Subramanian, Suvinay; Swing, Andy; Towles, Brian; Young, Cliff; Zhou, Xiang; Zhou, Zongwei; Patterson, David (2023). "TPU v4: An Optically Reconfigurable Supercomputer for Machine Learning with Hardware Support for Embeddings" arXiv:2304.01433 (<https://arxiv.org/abs/2304.01433>) [cs.AR] (<https://arxiv.org/archive/cs.AR>).
21. Cloud TPU v5e Inference Public Preview (<https://cloud.google.com/tpu/docs/v5e-inference>), retrieved 2023-11-06.
22. Cloud TPU v5p (<https://cloud.google.com/tpu/docs/v5p>) *Google Cloud*. retrieved 2024-04-09
23. Cloud TPU v5p Training (<https://cloud.google.com/tpu/docs/v5p-training>), retrieved 2024-04-09.
24. "Introducing Trillium, sixth-generation TPUs" (<https://cloud.google.com/blog/products/compute/introducing-trillium-6th-gen-tpus>). *Google Cloud Blog*. Retrieved 2024-05-29.
25. "TPU v6e" (<https://cloud.google.com/tpu/docs/v6e>). *Google Cloud*. Retrieved 2024-11-10.
26. "Ironwood: The first Google TPU for the age of inference" (<https://blog.google/products/google-cloud/ironwood-tpu-age-of-inference/>). *Google Blog*. 9 April 2025. Retrieved 2025-04-09.
27. Bright, Peter (17 May 2017). "Google brings 45 teraflops tensor flow processors to its compute cloud" (<https://arstechnica.com/information-technology/2017/05/google-brings-45-teraflops-tensor-flow-processors-to-its-compute-cloud/>). *Ars Technica*. Retrieved 30 May 2017.
28. Kennedy, Patrick (17 May 2017). "Google Cloud TPU Details Revealed" (<https://www.servethehome.com/google-cloud-tpu-details-revealed/>). *Serve The Home*. Retrieved 30 May 2017.
29. Frumusanu, Andre (8 May 2018). "Google I/O Opening Keynote Live-Blog" (<https://www.anandtech.com/show/12726/google-io-keynote-liveblog-10am-pt>). Retrieved 9 May 2018.

30. Feldman, Michael (11 May 2018). "Google Offers Glimpse of Third-Generation TPU Processor" (<https://www.top500.org/news/google-offers-glimpse-of-third-generation-tpu-processor/>). Top 500. Retrieved 14 May 2018.
31. Teich, Paul (10 May 2018). "Tearing Apart Google's TPU 3.0 AI Coprocessor" (<https://www.nextplatform.com/2018/05/10/tearing-a-part-googles-tpu-3-0-ai-coprocessor/>). The Next Platform. Retrieved 14 May 2018.
32. "Google Launches TPU v4 AI Chips" (<https://www.hpcwire.com/2021/05/20/google-launches-tpu-v4-ai-chips/>). *www.hpcwire.com*. 20 May 2021. Retrieved June 7, 2021.
33. Jouppi, Norman (2023-04-20). "TPU v4: An Optically Reconfigurable Supercomputer for Machine Learning with Hardware Support for Embeddings". *arXiv:2304.01433* (<https://arxiv.org/abs/2304.01433>) [cs.AR] (<https://arxiv.org/archive/cs.AR>).
34. Kennedy, Patrick (2023-08-29). "Google Details TPUv4 and its Crazy Optically Reconfigurable AI Network" (<https://www.servethehome.com/google-details-tpuv4-and-its-crazy-optically-reconfigurable-ai-network/>). *servethehome.com*. Retrieved 2023-12-16.
35. "Why did Google develop its own TPU chip? In-depth disclosure of team members" (<https://www.centry.com/blog/why-did-google-develop-its-own-tpu-chip-in-depth-disclosure-of-team-members.html>). *centry.com*. 2021-10-20. Retrieved 2023-12-16.
36. Mirhoseini, Azalia; Goldie, Anna (2021-06-01). "A graph placement methodology for fast chip design" (<http://176.9.41.242/doc/reinforcement-learning/model/2021-mirhoseini.pdf>) (PDF). *Nature*. **594** (7962): 207–212. doi:[10.1038/s41586-022-04657-6](https://doi.org/10.1038/s41586-022-04657-6). PMID 35361999 (<https://pubmed.ncbi.nlm.nih.gov/35361999/>). S2CID 247855593 (<https://api.semanticscholar.org/CorpusID:247855593>). Retrieved 2023-06-04.
37. Vahdat, Amin (2023-12-06). "Enabling next-generation AI workloads: Announcing TPU v5p and AI Hypercomputer" (<https://cloud.google.com/blog/products/ai-machine-learning/introducing-cloud-tpu-v5p-and-ai-hypercomputer>). Retrieved 2024-04-08.
38. Afifi-Sabet, Keumars (2023-12-23). "Google is rapidly turning into a formidable opponent to BFF Nvidia — the TPU v5p AI chip powering its hypercomputer is faster and has more memory and bandwidth than ever before, beating even the mighty H100" (<https://www.techradar.com/pro/google-is-rapidly-turning-into-a-formidable-opponent-to-bff-nvidia-the-tpu-v5p-ai-chip-powering-its-hypercomputer-is-faster-and-has-more-memory-and-bandwidth-than-ever-before-beating-even-themighty-h100>). *TechRadar*. Retrieved 2024-04-08.
39. "Expanding our AI-optimized infrastructure portfolio: Introducing Cloud TPU v5e and announcing A3 GA" (<https://cloud.google.com/blog/products/compute/announcing-cloud-tpu-v5e-and-a3-gpus-in-ga>). 2023-08-29. Retrieved 2023-12-16.
40. "Enabling next-generation AI workloads: Announcing TPU v5p and AI Hypercomputer" (<https://cloud.google.com/blog/products/ai-machine-learning/introducing-cloud-tpu-v5p-and-ai-hypercomputer>). 2023-12-06. Retrieved 2024-04-09.
41. Lohmeyer, Mark (2024-10-30). "Powerful infrastructure innovations for your AI-first future" (<https://cloud.google.com/blog/products/compute/trillium-sixth-generation-tpu-is-in-preview>).
42. Velasco, Alan (2024-05-15). "Google Cloud Unveils Trillium, Its 6th-Gen TPU With A 4.7X AI Performance Leap" (<https://hothardware.com/news/google-cloud-unveils-trillium-its-6th-gen-tpu-with-a-47x-performance-leap>). *Hothardware*. *Hothardware*. Retrieved 2024-05-15.
43. "Introducing Trillium, sixth-generation TPUs" (<https://cloud.google.com/blog/products/compute/introducing-trillium-6th-gen-tpus>). *Google Cloud Blog*. Retrieved 2024-05-17.
44. Wiggers, Kyle (2025-04-09). "Ironwood is Google's newest AI accelerator chip" (<https://techcrunch.com/2025/04/09/google-unveils-ironwood-a-new-ai-accelerator-chip>). *TechCrunch*. Retrieved 2025-04-10.
45. "Cloud TPU" (<https://cloud.google.com/tpu>). *Google Cloud*. Retrieved 2021-05-21.
46. "Edge TPU performance benchmarks" (<https://coral.ai/docs/edgetpu/benchmarks/>). *Coral*. Retrieved 2020-01-04.
47. "Dev Board" (<https://coral.ai/products/dev-board>). *Coral*. Retrieved 2021-05-21.
48. "System-on-Module (SoM)" (<https://coral.ai/products/som>). *Coral*. Retrieved 2021-05-21.
49. "Bringing intelligence to the edge with Cloud IoT" (<https://www.blog.google/products/google-cloud/bringing-intelligence-to-the-edge-with-cloud-iot>). *Google Blog*. 2018-07-25. Retrieved 2018-07-25.
50. "Retrain an image classification model on-device" (<https://coral.withgoogle.com/docs/edgetpu/retrain-classification-ondevice/>). *Coral*. Retrieved 2019-05-03.
51. "組込み総合技術展 & IoT総合技術展「ET & IoT Technology 2019」に出展することを発表" (<https://www.asus.com/jp/News/jr4skqts65jsugg>). *Asus.com* (in Japanese). Retrieved 2019-11-13.
52. Shilov, Anton. "ASUS & Google Team Up for 'Tinker Board' AI-Focused Credit-Card Sized Computers" (<https://www.anandtech.com/show/15095/asus-google-team-up-for-tinker-board-ai-focused-creditcard-sized-computers>). *Anandtech.com*. Retrieved 2019-11-13.
53. Aufranc, Jean-Luc (2019-05-29). "ASUS Tinker Edge T & CR1S-CM-A SBC to Feature Google Coral Edge TPU & NXP i.MX 8M Processor" (<https://www.cnx-software.com/2019/05/29/asus-tinker-edge-t-cr1s-cm-a-sbc-google-coral-edge-tpu-nxp-i-mx-8m-processor>). *CNX Software - Embedded Systems News*. Retrieved 2019-11-14.
54. "New Coral products for 2020" (<https://developers.googleblog.com/2020/01/new-coral-products-for-2020.html>). *Google Developers Blog*. Retrieved 2020-01-04.
55. "Accelerator Module" (<https://coral.ai/products/accelerator-module>). *Coral*. Retrieved 2020-01-04.
56. "Introducing the Next Generation of On-Device Vision Models: MobileNetV3 and MobileNetEdgeTPU" (<http://ai.googleblog.com/2019/11/introducing-next-generation-on-device.html>). *Google AI Blog*. Retrieved 2020-04-16.
57. Gupta, Suyog; White, Marie (November 8, 2021). "Improved On-Device ML on Pixel 6, with Neural Architecture Search" (<https://ai.googleblog.com/2021/11/improved-on-device-ml-on-pixel-6-with.html>). *Google AI Blog*. Retrieved 16 December 2022.
58. Frumusanu, Andrei (November 2, 2021). "Google's Tensor inside of Pixel 6, Pixel 6 Pro: A Look into Performance & Efficiency | Google's IP: Tensor TPU/NPU" (<https://www.anandtech.com/show/17032/tensor-soc-performance-efficiency/5>). *AnandTech*. Retrieved 16 December 2022.
59. Hardesty, Larry (2011-01-03). "The surprising usefulness of sloppy arithmetic" (<https://news.mit.edu/2010/fuzzy-logic-0103>). *MIT*. Retrieved 2024-01-10.
60. Bray, Hiawatha (2024-01-10). "Local inventor challenges Google in billion-dollar patent fight" (<https://web.archive.org/web/20240110130855/https://www.bostonglobe.com/2024/01/10/business/local-inventor-challenges-google-billion-dollar-patent-fight/>). *Boston Globe*. Boston. Archived from the original (<https://www.bostonglobe.com/2024/01/10/business/local-inventor-challenges-google-billion-dollar-patent-fight/>) on 2024-01-10. Retrieved 2024-01-10.
61. "SINGULAR COMPUTING LLC, Plaintiff, v. GOOGLE LLC, Defendant: Amended Complaint for Patent Infringement" (https://www.pacermonitor.com/public/filings/DCYQIDD/Singular_Computing_LLC_v_Google_LLC_madce-24-10008_0001.0.pdf) (PDF). *rpxcorp.com*. *RPX Corporation*. 2020-03-20. Retrieved 2024-01-10.
62. Wang, Xiaojun; Leeser, Miriam (2010-09-01). "VFloat: A Variable Precision Fixed- and Floating-Point Library for Reconfigurable Hardware" (<https://dl.acm.org/doi/abs/10.1145/1839480.1839486>). *ACM Transactions on Reconfigurable Technology and Systems*. **3** (3): 1–34. doi:[10.1145/1839480.1839486](https://doi.org/10.1145/1839480.1839486) (<https://doi.org/10.1145/1839480.1839486>). Retrieved 2024-01-10.
63. "Singular Computing LLC v. Google LLC" (<https://casetext.com/case/singular-computing-llc-v-google-llc-1>). *casetext.com*. 2023-04-06. Retrieved 2024-01-10.
64. Calkins, Laurel Brubaker (January 24, 2024). "Google Settles AI-Chip Suit That Had Sought Over \$5 Billion" (<https://news.bloomberg.com/ip-law/google-settles-ai-chip-design-suit-that-had-sought-over-5-billion>). *Bloomberg Law*.
65. Brittain, Blake; Raymond, Ray (January 24, 2024). "Google settles AI-related chip patent lawsuit that sought \$1.67 bln" (<https://www.reuters.com/technology/google-settles-ai-related-chip-patent-lawsuit-that-sought-167-bln-2024-01-24/>). *Reuters*.

External links

- [Cloud Tensor Processing Units \(TPUs\)](https://cloud.google.com/tpu/docs/tpus) (<https://cloud.google.com/tpu/docs/tpus>) (Documentation from Google Cloud)
 - [Photo of Google's TPU chip and board](https://images.anandtech.com/doc/12195/google-tpu-board-2.png) (<https://images.anandtech.com/doc/12195/google-tpu-board-2.png>)
 - [Photo of Google's TPU v2 board](https://cloud.google.com/images/products/tpu/cloud-tpu-v2.png) (<https://cloud.google.com/images/products/tpu/cloud-tpu-v2.png>)
 - [Photo of Google's TPU v3 board](https://cloud.google.com/images/products/tpu/cloud-tpu-v3-alpha.png) (<https://cloud.google.com/images/products/tpu/cloud-tpu-v3-alpha.png>)
 - [Photo of Google's TPU v2 pod](https://cloud.google.com/images/products/tpu/cloud-tpu-v2-pod-alpha.png) (<https://cloud.google.com/images/products/tpu/cloud-tpu-v2-pod-alpha.png>)
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Tensor_Processing_Unit&oldid=1292159275"



Digital signal processing

Digital signal processing (DSP) is the use of digital processing, such as by computers or more specialized digital signal processors, to perform a wide variety of signal processing operations. The digital signals processed in this manner are a sequence of numbers that represent samples of a continuous variable in a domain such as time, space, or frequency. In digital electronics, a digital signal is represented as a pulse train,^{[1][2]} which is typically generated by the switching of a transistor.^[3]

Digital signal processing and analog signal processing are subfields of signal processing. DSP applications include audio and speech processing, sonar, radar and other sensor array processing, spectral density estimation, statistical signal processing, digital image processing, data compression, video coding, audio coding, image compression, signal processing for telecommunications, control systems, biomedical engineering, and seismology, among others.

DSP can involve linear or nonlinear operations. Nonlinear signal processing is closely related to nonlinear system identification^[4] and can be implemented in the time, frequency, and spatio-temporal domains.

The application of digital computation to signal processing allows for many advantages over analog processing in many applications, such as error detection and correction in transmission as well as data compression.^[5] Digital signal processing is also fundamental to digital technology, such as digital telecommunication and wireless communications.^[6] DSP is applicable to both streaming data and static (stored) data.

Signal sampling

To digitally analyze and manipulate an analog signal, it must be digitized with an analog-to-digital converter (ADC).^[7] Sampling is usually carried out in two stages, discretization and quantization. Discretization means that the signal is divided into equal intervals of time, and each interval is represented by a single measurement of amplitude. Quantization means each amplitude measurement is approximated by a value from a finite set. Rounding real numbers to integers is an example.

The Nyquist–Shannon sampling theorem states that a signal can be exactly reconstructed from its samples if the sampling frequency is greater than twice the highest frequency component in the signal. In practice, the sampling frequency is often significantly higher than this.^[8] It is common to use an anti-aliasing filter to limit the signal bandwidth to comply with the sampling theorem, however careful selection of this filter is required because the reconstructed signal will be the filtered signal plus residual aliasing from imperfect stop band rejection instead of the original (unfiltered) signal.

Theoretical DSP analyses and derivations are typically performed on discrete-time signal models with no amplitude inaccuracies (quantization error), created by the abstract process of sampling. Numerical methods require a quantized signal, such as those produced by an ADC. The processed result might be a frequency spectrum or a set of statistics. But often it is another quantized signal that is converted back to analog form by a digital-to-analog converter (DAC).

Domains

DSP engineers usually study digital signals in one of the following domains: time domain (one-dimensional signals), spatial domain (multidimensional signals), frequency domain, and wavelet domains. They choose the domain in which to process a signal by making an informed assumption (or by trying different possibilities) as to which domain best represents the essential characteristics of the signal and the processing to be applied to it. A sequence of samples from a measuring device produces a temporal or spatial domain representation, whereas a discrete Fourier transform produces the frequency domain representation.

Time and space domains

Time domain refers to the analysis of signals with respect to time. Similarly, space domain refers to the analysis of signals with respect to position, e.g., pixel location for the case of image processing.

The most common processing approach in the time or space domain is enhancement of the input signal through a method called filtering. Digital filtering generally consists of some linear transformation of a number of surrounding samples around the current sample of the input or output signal. The surrounding samples may be identified with respect to time or space. The output of a linear digital filter to any given input may be calculated by convolving the input signal with an impulse response.

Frequency domain

Signals are converted from time or space domain to the frequency domain usually through use of the Fourier transform. The Fourier transform converts the time or space information to a magnitude and phase component of each frequency. With some applications, how the phase varies with frequency can be a significant consideration. Where phase is unimportant, often the Fourier transform is converted to the power spectrum, which is the magnitude of each frequency component squared.

The most common purpose for analysis of signals in the frequency domain is analysis of signal properties. The engineer can study the spectrum to determine which frequencies are present in the input signal and which are missing. Frequency domain analysis is also called *spectrum-* or *spectral analysis*.

Filtering, particularly in non-realtime work can also be achieved in the frequency domain, applying the filter and then converting back to the time domain. This can be an efficient implementation and can give essentially any filter response including excellent approximations to brickwall filters.

There are some commonly used frequency domain transformations. For example, the cepstrum converts a signal to the frequency domain through Fourier transform, takes the logarithm, then applies another Fourier transform. This emphasizes the harmonic structure of the original spectrum.

Z-plane analysis

Digital filters come in both infinite impulse response (IIR) and finite impulse response (FIR) types. Whereas FIR filters are always stable, IIR filters have feedback loops that may become unstable and oscillate. The Z-transform provides a tool for analyzing stability issues of digital IIR filters. It is analogous to the Laplace transform, which is used to design and analyze analog IIR filters.

Autoregression analysis

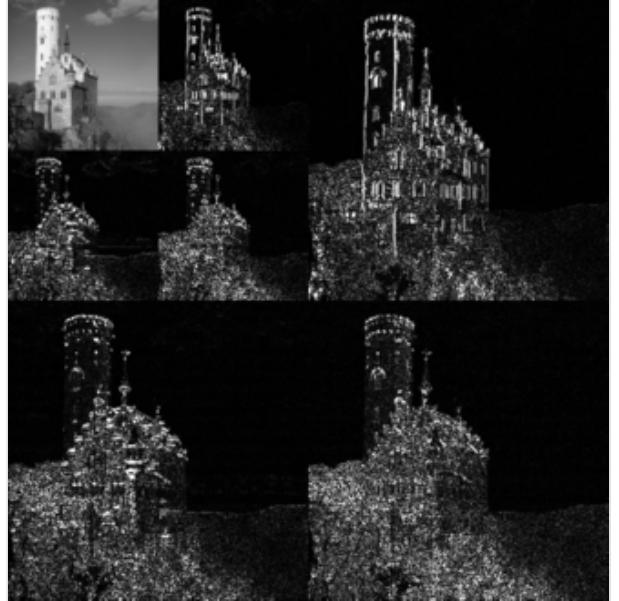
A signal is represented as linear combination of its previous samples. Coefficients of the combination are called autoregression coefficients. This method has higher frequency resolution and can process shorter signals compared to the Fourier transform.^[9] Prony's method can be used to estimate phases, amplitudes, initial phases and decays of the components of signal.^{[10][9]} Components are assumed to be complex decaying exponents.^{[10][9]}

Time-frequency analysis

A time-frequency representation of signal can capture both temporal evolution and frequency structure of analyzed signal. Temporal and frequency resolution are limited by the principle of uncertainty and the tradeoff is adjusted by the width of analysis window. Linear techniques such as Short-time Fourier transform, wavelet transform, filter bank,^[11] non-linear (e.g., Wigner–Ville transform^[10]) and autoregressive methods (e.g. segmented Prony method)^{[10][12][13]} are used for representation of signal on the time-frequency plane. Non-linear and segmented Prony methods can provide higher resolution, but may produce undesirable artifacts. Time-frequency analysis is usually used for analysis of non-stationary signals. For example, methods of fundamental frequency estimation, such as RAPT and PEFAC^[14] are based on windowed spectral analysis.

Wavelet

In numerical analysis and functional analysis, a discrete wavelet transform is any wavelet transform for which the wavelets are discretely sampled. As with other wavelet transforms, a key advantage it has over Fourier transforms is temporal resolution: it captures both frequency *and* location information. The accuracy of the joint time-frequency resolution is limited by the uncertainty principle of time-frequency.



Empirical mode decomposition

Empirical mode decomposition is based on decomposition signal into intrinsic mode functions (IMFs). IMFs are quasi-harmonical oscillations that are extracted from the signal.^[15]

Implementation

DSP algorithms may be run on general-purpose computers^[16] and digital signal processors.^[17] DSP algorithms are also implemented on purpose-built hardware such as application-specific integrated circuit (ASICs).^[18] Additional technologies for digital

An example of the 2D discrete wavelet transform that is used in JPEG2000. The original image is high-pass filtered, yielding the three large images, each describing local changes in brightness (details) in the original image. It is then low-pass filtered and downsampled, yielding an approximation image; this image is high-pass filtered to produce the three smaller detail images, and low-pass filtered to produce the final approximation image in the upper-left.

signal processing include more powerful general-purpose microprocessors, graphics processing units, field-programmable gate arrays (FPGAs), digital signal controllers (mostly for industrial applications such as motor control), and stream processors.^[19]

For systems that do not have a real-time computing requirement and the signal data (either input or output) exists in data files, processing may be done economically with a general-purpose computer. This is essentially no different from any other data processing, except DSP mathematical techniques (such as the DCT and FFT) are used, and the sampled data is usually assumed to be uniformly sampled in time or space. An example of such an application is processing digital photographs with software such as Photoshop.

When the application requirement is real-time, DSP is often implemented using specialized or dedicated processors or microprocessors, sometimes using multiple processors or multiple processing cores. These may process data using fixed-point arithmetic or floating point. For more demanding applications FPGAs may be used.^[20] For the most demanding applications or high-volume products, ASICs might be designed specifically for the application.

Parallel implementations of DSP algorithms, utilizing multi-core CPU and many-core GPU architectures, are developed to improve the performances in terms of latency of these algorithms.^[21]

Native processing is done by the computer's CPU rather than by DSP or outboard processing, which is done by additional third-party DSP chips located on extension cards or external hardware boxes or racks. Many digital audio workstations such as Logic Pro, Cubase, Digital Performer and Pro Tools LE use native processing. Others, such as Pro Tools HD, Universal Audio's UAD-1 and TC Electronic's Powercore use DSP processing.

Applications

General application areas for DSP include

- Audio signal processing
- Audio data compression e.g. MP3
- Video data compression
- Computer graphics
- Digital image processing
- Photo manipulation
- Speech processing
- Speech recognition
- Data transmission
- Radar
- Sonar
- Financial signal processing
- Economic forecasting
- Seismology
- Biomedicine
- Weather forecasting

Specific examples include speech coding and transmission in digital mobile phones, room correction of sound in hi-fi and sound reinforcement applications, analysis and control of industrial processes, medical imaging such as CAT scans and MRI, audio crossovers and equalization, digital synthesizers, and audio effects units.^[22] DSP has been used in hearing aid technology since 1996, which allows for automatic directional microphones, complex digital noise reduction, and improved adjustment of the frequency response.^[23]

Techniques

- [Bilinear transform](#)
- [Discrete Fourier transform](#)
- [Discrete-time Fourier transform](#)
- [Filter design](#)
- [Goertzel algorithm](#)
- [Least-squares spectral analysis](#)
- [LTI system theory](#)
- [Minimum phase](#)
- [s-plane](#)
- [Transfer function](#)
- [Z-transform](#)

Related fields

- [Analog signal processing](#)
- [Automatic control](#)
- [Computer engineering](#)
- [Computer science](#)
- [Data compression](#)
- [Dataflow programming](#)
- [Discrete cosine transform](#)
- [Electrical engineering](#)
- [Fourier analysis](#)
- [Information theory](#)
- [Machine learning](#)
- [Real-time computing](#)
- [Stream processing](#)
- [Telecommunications](#)
- [Time series](#)
- [Wavelet](#)

Further reading

- Ahmed, Nasir; Rao, Kamisetty Ramamohan (7 August 1975). "Orthogonal transforms for digital signal processing". *ICASSP '76. IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 1. New York: Springer-Verlag. pp. 136–140. doi:10.1109/ICASSP.1976.1170121 (<https://doi.org/10.1109%2FICASSP.1976.1170121>). ISBN 978-3540065562. LCCN 73018912 (<https://lccn.loc.gov/73018912>). OCLC 438821458 (<https://search.worldcat.org/oclc/438821458>). OL 22806004M (<https://openlibrary.org/books/OL22806004M>). S2CID 10776771 (<https://api.semanticscholar.org/CorpusID:10776771>).
- Jonathan M. Blackledge, Martin Turner: *Digital Signal Processing: Mathematical and Computational Methods, Software Development and Applications*, Horwood Publishing, ISBN 1-898563-48-9
- James D. Broesch: *Digital Signal Processing Demystified*, Newnes, ISBN 1-878707-16-7
- Dyer, Stephen A.; Harms, Brian K. (13 August 1993). "Digital Signal Processing" (<https://books.google.com/books?id=vL-bB7GALAwC&pg=PA104>). In Yovits, Marshall C. (ed.). *Advances in Computers*. Vol. 37. Academic Press. pp. 59–118. doi:10.1016/S0065-2458(08)60403-9 (<https://doi.org/10.1016%2FS0065-2458%2808%2960403-9>). ISBN 978-0120121373. ISSN 0065-2458 (<https://search.worldcat.org/issn/0065-2458>). LCCN 59015761 (<https://lccn.loc.gov/59015761>). OCLC 858439915 (<https://search.worldcat.org/oclc/858439915>). OL 10070096M (<https://openlibrary.org/books/OL10070096M>).
- Paul M. Embree, Damon Danieli: *C++ Algorithms for Digital Signal Processing*, Prentice Hall, ISBN 0-13-179144-3
- Hari Krishna Garg: *Digital Signal Processing Algorithms*, CRC Press, ISBN 0-8493-7178-3
- P. Gaydecki: *Foundations Of Digital Signal Processing: Theory, Algorithms And Hardware Design*, Institution of Electrical Engineers, ISBN 0-85296-431-5
- Ashfaq Khan: *Digital Signal Processing Fundamentals*, Charles River Media, ISBN 1-58450-281-9

- Sen M. Kuo, Woon-Seng Gan: *Digital Signal Processors: Architectures, Implementations, and Applications*, Prentice Hall, ISBN 0-13-035214-4
- Paul A. Lynn, Wolfgang Fuerst: *Introductory Digital Signal Processing with Computer Applications*, John Wiley & Sons, ISBN 0-471-97984-8
- Richard G. Lyons: *Understanding Digital Signal Processing*, Prentice Hall, ISBN 0-13-108989-7
- Vijay Madisetti, Douglas B. Williams: *The Digital Signal Processing Handbook*, CRC Press, ISBN 0-8493-8572-5
- James H. McClellan, Ronald W. Schafer, Mark A. Yoder: *Signal Processing First*, Prentice Hall, ISBN 0-13-090999-8
- Bernard Mulgrew, Peter Grant, John Thompson: *Digital Signal Processing – Concepts and Applications*, Palgrave Macmillan, ISBN 0-333-96356-3
- Boaz Porat: *A Course in Digital Signal Processing*, Wiley, ISBN 0-471-14961-6
- John G. Proakis, Dimitris Manolakis: *Digital Signal Processing: Principles, Algorithms and Applications*, 4th ed, Pearson, April 2006, ISBN 978-0131873742
- John G. Proakis: *A Self-Study Guide for Digital Signal Processing*, Prentice Hall, ISBN 0-13-143239-7
- Charles A. Schuler: *Digital Signal Processing: A Hands-On Approach*, McGraw-Hill, ISBN 0-07-829744-3
- Doug Smith: *Digital Signal Processing Technology: Essentials of the Communications Revolution*, American Radio Relay League, ISBN 0-87259-819-5
- Smith, Steven W. (2002). *Digital Signal Processing: A Practical Guide for Engineers and Scientists* (<http://www.dspguide.com>). Newnes. ISBN 0-7506-7444-X.
- Stein, Jonathan Yaakov (2000-10-09). *Digital Signal Processing, a Computer Science Perspective*. Wiley. ISBN 0-471-29546-9.
- Stergiopoulos, Stergios (2000). *Advanced Signal Processing Handbook: Theory and Implementation for Radar, Sonar, and Medical Imaging Real-Time Systems*. CRC Press. ISBN 0-8493-3691-0.
- Van De Vegte, Joyce (2001). *Fundamentals of Digital Signal Processing*. Prentice Hall. ISBN 0-13-016077-6.
- Oppenheim, Alan V.; Schafer, Ronald W. (2001). *Discrete-Time Signal Processing*. Pearson. ISBN 1-292-02572-7.
- Hayes, Monson H. Statistical digital signal processing and modeling. John Wiley & Sons, 2009. (with MATLAB scripts (https://www.mathworks.com/matlabcentral/fileexchange/2183-statistical-digital-signal-processing-and-modeling?s_tid=prof_contriblnk))

References

1. B. SOMANATHAN NAIR (2002). *Digital electronics and logic design*. PHI Learning Pvt. Ltd. p. 289. ISBN 9788120319561. "Digital signals are fixed-width pulses, which occupy only one of two levels of amplitude."
2. Joseph Migga Kizza (2005). *Computer Network Security*. Springer Science & Business Media. ISBN 9780387204734.
3. *2000 Solved Problems in Digital Electronics* (https://books.google.com/books?id=N6FDii6_nSEC&pg=PA151). Tata McGraw-Hill Education. 2005. p. 151. ISBN 978-0-07-058831-8.
4. Billings, Stephen A. (Sep 2013). *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. UK: Wiley. ISBN 978-1-119-94359-4.

5. Broesch, James D.; Stranneby, Dag; Walker, William (2008-10-20). *Digital Signal Processing: Instant access* (1 ed.). Butterworth-Heinemann-Newnes. p. 3. ISBN 9780750689762.
6. Srivastava, Viranjay M.; Singh, Ghanshyam (2013). *MOSFET Technologies for Double-Pole Four-Throw Radio-Frequency Switch* (<https://books.google.com/books?id=fkO9BAAAQBAJ&pg=PA1>). Springer Science & Business Media. p. 1. ISBN 9783319011653.
7. Walden, R. H. (1999). "Analog-to-digital converter survey and analysis". *IEEE Journal on Selected Areas in Communications*. **17** (4): 539–550. doi:[10.1109/49.761034](https://doi.org/10.1109/49.761034) (<https://doi.org/10.1109%2F49.761034>).
8. Candes, E. J.; Wakin, M. B. (2008). "An Introduction To Compressive Sampling" (<https://resolver.caltech.edu/CaltechAUTHORS:CANieespm08>). *IEEE Signal Processing Magazine*. **25** (2): 21–30. Bibcode:2008ISPM...25...21C (<https://ui.adsabs.harvard.edu/abs/2008ISPM...25...21C>). doi:[10.1109/MSP.2007.914731](https://doi.org/10.1109/MSP.2007.914731) (<https://doi.org/10.1109%2FMSP.2007.914731>). S2CID 1704522 (<https://api.semanticscholar.org/CorpusID:1704522>).
9. Marple, S. Lawrence (1987-01-01). *Digital Spectral Analysis: With Applications*. Englewood Cliffs, N.J: Prentice Hall. ISBN 978-0-13-214149-9.
10. Ribeiro, M.P.; Ewins, D.J.; Robb, D.A. (2003-05-01). "Non-stationary analysis and noise filtering using a technique extended from the original Prony method" (<http://linkinghub.elsevier.com/retrieve/pii/S0888327001913998>). *Mechanical Systems and Signal Processing*. **17** (3): 533–549. Bibcode:2003MSSP..17..533R (<https://ui.adsabs.harvard.edu/abs/2003MSSP..17..533R>). doi:[10.1006/mssp.2001.1399](https://doi.org/10.1006/mssp.2001.1399) (<https://doi.org/10.1006%2Fmssp.2001.1399>). ISSN 0888-3270 (<https://search.worldcat.org/issn/0888-3270>). Retrieved 2019-02-17.
11. So, Stephen; Paliwal, Kuldip K. (2005). "Improved noise-robustness in distributed speech recognition via perceptually-weighted vector quantisation of filterbank energies". *Ninth European Conference on Speech Communication and Technology*.
12. Mitrofanov, Georgy; Priimenko, Viatcheslav (2015-06-01). "Prony Filtering of Seismic Data" (<https://doi.org/10.1515%2Facgeo-2015-0012>). *Acta Geophysica*. **63** (3): 652–678. Bibcode:2015AcGeo..63..652M (<https://ui.adsabs.harvard.edu/abs/2015AcGeo..63..652M>). doi:[10.1515/acgeo-2015-0012](https://doi.org/10.1515/acgeo-2015-0012) (<https://doi.org/10.1515%2Facgeo-2015-0012>). ISSN 1895-6572 (<https://search.worldcat.org/issn/1895-6572>). S2CID 130300729 (<https://api.semanticscholar.org/CorpusID:130300729>).
13. Mitrofanov, Georgy; Smolin, S. N.; Orlov, Yu. A.; Bespechnyy, V. N. (2020). "Prony decomposition and filtering" (http://www.jourgimss.ru/en/SitePages/catalog/2020/02/abstract/2020_2_55.aspx). *Geology and Mineral Resources of Siberia* (2): 55–67. doi:[10.20403/2078-0575-2020-2-55-67](https://doi.org/10.20403/2078-0575-2020-2-55-67) (<https://doi.org/10.20403%2F2078-0575-2020-2-55-67>). ISSN 2078-0575 (<https://search.worldcat.org/issn/2078-0575>). S2CID 226638723 (<https://api.semanticscholar.org/CorpusID:226638723>). Retrieved 2020-09-08.
14. Gonzalez, Sira; Brookes, Mike (February 2014). "PEFAC - A Pitch Estimation Algorithm Robust to High Levels of Noise" (<https://ieeexplore.ieee.org/document/6701334>). *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. **22** (2): 518–530. doi:[10.1109/TASLP.2013.2295918](https://doi.org/10.1109/TASLP.2013.2295918) (<https://doi.org/10.1109%2FTASLP.2013.2295918>). ISSN 2329-9290 (<https://search.worldcat.org/issn/2329-9290>). S2CID 13161793 (<https://api.semanticscholar.org/CorpusID:13161793>). Retrieved 2017-12-03.
15. Huang, N. E.; Shen, Z.; Long, S. R.; Wu, M. C.; Shih, H. H.; Zheng, Q.; Yen, N.-C.; Tung, C. C.; Liu, H. H. (1998-03-08). "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis" (<http://rspa.royalsocietypublishing.org/cgi/doi/10.1098/rspa.1998.0193>). *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. **454** (1971): 903–995. Bibcode:1998RSPSA.454..903H (<https://ui.adsabs.harvard.edu/abs/1998RSPSA.454..903H>). doi:[10.1098/rspa.1998.0193](https://doi.org/10.1098/rspa.1998.0193) (<https://doi.org/10.1098%2Frspa.1998.0193>). ISSN 1364-5021 (<https://search.worldcat.org/issn/1364-5021>). S2CID 1262186 (<https://api.semanticscholar.org/CorpusID:1262186>). Retrieved 2018-06-05.

16. Weipeng, Jiang; Zhiqiang, He; Ran, Duan; Xinglin, Wang (August 2012). "Major optimization methods for TD-LTE signal processing based on general purpose processor" (<https://ieeexplore.ieee.org/document/6417593>). *7th International Conference on Communications and Networking in China*. pp. 797–801. doi:[10.1109/ChinaCom.2012.6417593](https://doi.org/10.1109/ChinaCom.2012.6417593) (<https://doi.org/10.1109%2FChinaCom.2012.6417593>). ISBN 978-1-4673-2699-5. S2CID 17594911 (<https://api.semanticscholar.org/CorpusID:17594911>).
17. Zaynidinov, Hakimjon; Ibragimov, Sanjarbek; Tojiboyev, Gayrat; Nurmurodov, Javohir (2021-06-22). "Efficiency of Parallelization of Haar Fast Transform Algorithm in Dual-Core Digital Signal Processors". *2021 8th International Conference on Computer and Communication Engineering (ICCCE)* (<https://ieeexplore.ieee.org/document/9467190>). IEEE. pp. 7–12. doi:[10.1109/ICCCE50029.2021.9467190](https://doi.org/10.1109/ICCCE50029.2021.9467190) (<https://doi.org/10.1109%2FICCCE50029.2021.9467190>). ISBN 978-1-7281-1065-3. S2CID 236187914 (<https://api.semanticscholar.org/CorpusID:236187914>).
18. Lyakhov, P.A. (June 2023). "Area-Efficient digital filtering based on truncated multiply-accumulate units in residue number system $2^n - 1$, 2^n , $2^n + 1$ " (<https://doi.org/10.1016%2Fj.jksuci.2023.101574>). *Journal of King Saud University - Computer and Information Sciences*. **35** (6): 101574. doi:[10.1016/j.jksuci.2023.101574](https://doi.org/10.1016/j.jksuci.2023.101574) (<https://doi.org/10.1016%2Fj.jksuci.2023.101574>).
19. Stranneby, Dag; Walker, William (2004). *Digital Signal Processing and Applications* (<https://books.google.com/books?id=NKK1DdqcDVUC&pg=PA241>) (2nd ed.). Elsevier. ISBN 0-7506-6344-8.
20. JPFix (2006). "FPGA-Based Image Processing Accelerator" (http://www.jpfix.com/About_Us/Articles/FPGA-Based_Image_Processing_Ac/fpga-based_image_processing_ac.html). Retrieved 2008-05-10.
21. Kapinchev, Konstantin; Bradu, Adrian; Podoleanu, Adrian (December 2019). "Parallel Approaches to Digital Signal Processing Algorithms with Applications in Medical Imaging" (<https://ieeexplore.ieee.org/document/9008720>). *2019 13th International Conference on Signal Processing and Communication Systems (ICSPCS)* (<https://kar.kent.ac.uk/80930/1/Kapinchev2019.pdf>) (PDF). pp. 1–7. doi:[10.1109/ICSPCS47537.2019.9008720](https://doi.org/10.1109/ICSPCS47537.2019.9008720) (<https://doi.org/10.1109%2FICSPCS47537.2019.9008720>). ISBN 978-1-7281-2194-9. S2CID 211686462 (<https://api.semanticscholar.org/CorpusID:211686462>).
22. Rabiner, Lawrence R.; Gold, Bernard (1975). *Theory and application of digital signal processing* (<https://archive.org/details/theoryapplicatio00rabi>). Englewood Cliffs, NJ: Prentice-Hall, Inc. ISBN 978-0139141010.
23. Kerckhoff, Jessica; Listenberger, Jennifer; Valente, Michael (October 1, 2008). "Advances in hearing aid technology" (https://digitalcommons.wustl.edu/audio_hapubs/28). *Contemporary Issues in Communication Science and Disorders*. **35**: 102–112. doi:[10.1044/cicsd_35_F_102](https://doi.org/10.1044/cicsd_35_F_102) (https://doi.org/10.1044%2Fcicsd_35_F_102).

Retrieved from "https://en.wikipedia.org/w/index.php?title=Digital_signal_processing&oldid=1291361524"



Register-transfer level

In digital circuit design, **register-transfer level (RTL)** is a design abstraction which models a synchronous digital circuit in terms of the flow of digital signals (data) between hardware registers, and the logical operations performed on those signals.

Register-transfer-level abstraction is used in hardware description languages (HDLs) like Verilog and VHDL to create high-level representations of a circuit, from which lower-level representations and ultimately actual wiring can be derived. Design at the RTL level is typical practice in modern digital design.^[1]

Unlike in software compiler design, where the register-transfer level is an intermediate representation and at the lowest level, the RTL level is the usual input that circuit designers operate on. In circuit synthesis, an intermediate language between the input register transfer level representation and the target netlist is sometimes used. Unlike in netlist, constructs such as cells, functions, and multi-bit registers are available.^[2] Examples include FIRRTL and RTLIL.

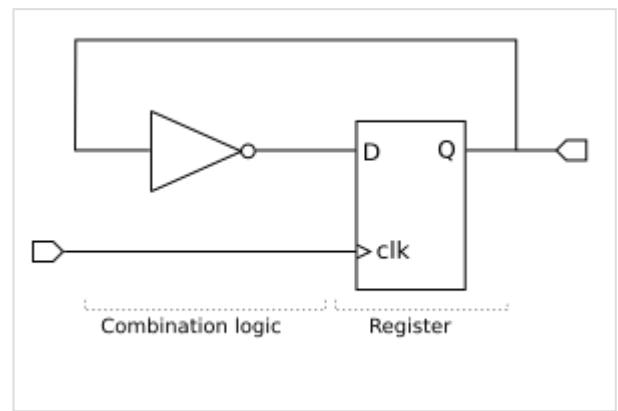
Transaction-level modeling is a higher level of electronic system design.

RTL description

A synchronous circuit consists of two kinds of elements: registers (sequential logic) and combinational logic. Registers (usually implemented as D flip-flops) synchronize the circuit's operation to the edges of the clock signal, and are the only elements in the circuit that have memory properties. Combinational logic performs all the logical functions in the circuit and it typically consists of logic gates.

For example, a very simple synchronous circuit is shown in the figure. The inverter is connected from the output, Q, of a register to the register's input, D, to create a circuit that changes its state on each rising edge of the clock, clk. In this circuit, the combinational logic consists of the inverter.

When designing digital integrated circuits with a hardware description language (HDL), the designs are usually engineered at a higher level of abstraction than transistor level (logic families) or logic gate level. In HDLs the designer declares the registers (which roughly correspond to variables in computer programming languages), and describes the combinational logic by using constructs that are familiar from



Example of a simple circuit with the output toggling at each rising edge of the input. The inverter forms the combinational logic in this circuit, and the register holds the state.

programming languages such as if-then-else and arithmetic operations. This level is called *register-transfer level*. The term refers to the fact that RTL focuses on describing the flow of signals between registers.

As an example, the circuit mentioned above can be described in VHDL as follows:

```
D <= not Q;  
  
process(clk)  
begin  
  if rising_edge(clk) then  
    Q <= D;  
  end if;  
end process;
```

Using an EDA tool for synthesis, this description can usually be directly translated to an equivalent hardware implementation file for an ASIC or an FPGA. The synthesis tool also performs logic optimization.

At the register-transfer level, some types of circuits can be recognized. If there is a cyclic path of logic from a register's output to its input (or from a set of registers outputs to its inputs), the circuit is called a state machine or can be said to be sequential logic. If there are logic paths from a register to another without a cycle, it is called a pipeline.

RTL in the circuit design cycle

RTL is used in the logic design phase of the integrated circuit design cycle.

An RTL description is usually converted to a gate-level description of the circuit by a logic synthesis tool. The synthesis results are then used by placement and routing tools to create a physical layout.

Logic simulation tools may use a design's RTL description to verify its correctness.

Power estimation techniques for RTL

The most accurate power analysis tools are available for the circuit level but unfortunately, even with switch- rather than device-level modelling, tools at the circuit level have disadvantages like they are either too slow or require too much memory thus inhibiting large chip handling. The majority of these are simulators like SPICE and have been used by the designers for many years as performance analysis tools. Due to these disadvantages, gate-level power estimation tools have begun to gain some acceptance where faster, probabilistic techniques have begun to gain a foothold. But it also has its trade off as speedup is achieved on the cost of accuracy, especially in the presence of correlated signals. Over the years it has been realized that biggest wins in low power design cannot come from circuit- and gate-level optimizations whereas architecture, system, and algorithm optimizations tend to have the largest impact on power consumption. Therefore, there has been a shift in the incline of the tool developers towards high-level analysis and optimization tools for power.

Motivation

It is well known that more significant power reductions are possible if optimizations are made on levels of abstraction, like the architectural and algorithmic level, which are higher than the circuit or gate level [3]. This provides the required motivation for the developers to focus on the development of new architectural level power analysis tools. This in no way implies that lower level tools are unimportant. Instead, each layer of tools provides a foundation upon which the next level can be built. The abstractions of the estimation techniques at a lower level can be used on a higher level with slight modifications.

Advantages of doing power estimation at RTL or architectural level

- Designers use a register-transfer level (RTL) description of the design to make optimizations and trade-offs very early in the design flow.
- The presence of functional blocks in an RTL description makes the complexity of architectural design much more manageable even for large chips because RTL has granularity sufficiently larger than gate- or circuit-level descriptions.

Gate equivalents^[4]

It is a technique based on the concept of gate equivalents. The complexity of a chip architecture can be described approximately in terms of gate equivalents where gate equivalent count specifies the average number of reference gates that are required to implement the particular function. The total power required for the particular function is estimated by multiplying the approximated number of gate equivalents with the average power consumed per gate. The reference gate can be any gate e.g. 2-input NAND gate.

Examples of gate equivalent technique

- **Class-independent power modeling:** It is a technique which tries to estimate chip area, speed, and power dissipation based on information about the complexity of the design in terms of gate equivalents. The functionality is divided among different blocks but no distinction is made about the functionality of the blocks i.e. it is basically class independent. This is the technique used by the *chip estimation system* (CES).

Steps:

1. Identify the functional blocks such as counters, decoders, multipliers, memories, etc.
2. Assign a complexity in terms of gate equivalents. The number of GE's for each unit type are either taken directly as an input from the user or are fed from a library.

$$P = \sum_{i \in \text{fns}} GE_i (E_{\text{typ}} + C_L^i V_{dd}^2) f A_{\text{int}}^i$$

Where E_{typ} is the assumed average dissipated energy by a gate equivalent, when active. The activity factor, A_{int} , denotes the average percentage of gates switching per clock cycle and is allowed to vary from function to function. The capacitive load, C_L , is a combination of fan-out loading as well as wiring. An estimate of the average wire length can be used to calculate the wiring capacitance. This is provided by the user and cross-checked by using a derivative of Rent's rule.

Assumptions:

1. A single reference gate is taken as the basis for all the power estimates not taking into consideration different circuit styles, clocking strategies, or layout techniques.
2. The percentage of gates switching per clock cycle denoted by activity factors are assumed to be fixed regardless of the input patterns.
3. Typical gate switching energy is characterized by completely random uniform white noise (UWN) distribution of the input data. This implies that the power estimation is same regardless of the circuit being idle or at maximum load as this UWN model ignores how different input distributions affect the power consumption of gates and modules.^[5]

- **Class-dependent power modeling:** This approach is slightly better than the previous approach as it takes into account customized estimation techniques to the different types of functional blocks thus trying to increase the modelling accuracy which wasn't the case in the previous technique such as logic, memory, interconnect, and clock hence the name. The power estimation is done in a very similar manner to the independent case. The basic switching energy is based on a three-input AND gate and is calculated from technology parameters e.g. gate width, tox, and metal width provided by the user.

$$P_{\text{bitlines}} = \frac{N_{\text{col}}}{2} \cdot (L_{\text{col}} C_{\text{wire}} + N_{\text{row}} C_{\text{cell}}) V_{\text{dd}} V_{\text{swing}}$$

Where C_{wire} denotes the bit line wiring capacitance per unit length and C_{cell} denotes the loading due to a single cell hanging off the bit line. The clock capacitance is based on the assumption of an H-tree distribution network. Activity is modelled using a UWN model. As can be seen by the equation the power consumption of each components is related to the number of columns (N_{col}) and rows (N_{row}) in the memory array.

Disadvantages:

1. The circuit activities are not modeled accurately as an overall activity factor is assumed for the entire chip which is also not trustable as provided by the user. As a matter of fact activity factors will vary throughout the chip hence this is not very accurate and prone to error. This leads to the problem that even if the model gives a correct estimate for the total power consumption by the chip, the module wise power distribution is fairly inaccurate.
2. The chosen activity factor gives the correct total power, but the breakdown of power into logic, clock, memory, etc. is less accurate. Therefore this tool is not much different or improved in comparison with CES.

Precharacterized cell libraries

This technique further customizes the power estimation of various functional blocks by having separate power model for logic, memory, and interconnect suggesting a power factor approximation (PFA) method for individually characterizing an entire library of functional blocks such as multipliers, adders, etc. instead of a single gate-equivalent model for “logic” blocks.

The power over the entire chip is approximated by the expression:

$$P = \sum_{i \in \text{all blocks}} K_i G_i f_i$$

Where K_i is PFA proportionality constant that characterizes the i_{th} functional element G_i is the measure of hardware complexity, and f_i denotes the activation frequency.

Example

G_i denoting the hardware complexity of the multiplier is related to the square of the input word length i.e. N^2 where N is the word length. The activation frequency is the rate at which multiplies are performed by the algorithm denoted by f_{mult} and the PFA constant, K_{mult} , is extracted empirically from past multiplier designs and shown to be about 15 fW/bit²-Hz for a 1.2 μ m technology at 5V. The resulting power model for the multiplier on the basis of the above assumptions is:

$$P_{\text{mult}} = K_{\text{mult}} N^2 f_{\text{mult}}$$

Advantages:

- Customization is possible in terms of whatever complexity parameters which are appropriate for that block. E.g. for a multiplier the square of the word length was appropriate. For memory, the storage capacity in bits is used and for the I/O drivers the word length alone is adequate.

Weakness:

- There is the implicit assumption that the inputs do not affect the multiplier activity which is contradictory to the fact that the PFA constant K_{mult} is intended to capture the intrinsic internal activity associated with the multiply operation as it is taken to be a constant.

The estimation error (relative to switch-level simulation) for a 16x16 multiplier is experimented and it is observed that when the dynamic range of the inputs does not fully occupy the word length of the multiplier, the UWN model becomes extremely inaccurate.^[6] Granted, good designers attempt to maximize word length utilization. Still, errors in the range of 50-100% are not uncommon. The figure clearly suggests a flaw in the UWN model.

See also

- [Datapath](#)
- [Electronic design automation \(EDA\)](#)
- [Electronic system-level](#)
- [Finite-state machine with datapath](#)
- [Integrated circuit design](#)
- [Synchronous circuit](#)
- [Algorithmic state machine](#)

Power estimation

- [Gate equivalent](#)
- [Power optimization \(EDA\)](#)
- [Gaussian noise](#)

References

1. Frank Vahid (2010). *Digital Design with RTL Design, Verilog and VHDL* (<https://books.google.com/books?id=-YayRpmjc20C&pg=PA247>) (2nd ed.). John Wiley and Sons. p. 247. ISBN 978-0-470-53108-2.
 2. Yosys Manual (https://yosys.readthedocs.io/_downloads/en/latest/pdf/) (RTLIL)
 3. "Power Estimation Techniques for Integrated Circuits " (<http://www.eecg.toronto.edu/~najm/papers/iccad95-tutorial.pdf>)
 4. "Low-Power Architectural Design Methodologies " (<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.61.4551&rep=rep1&type=pdf>)
 5. "Register-Transfer Level Estimation Techniques for Switching Activity and Power Consumption" (http://delivery.acm.org/10.1145/250000/244548/p158-raghunathan.pdf?ip=103.27.8.42&id=244548&acc=ACTIVE%20SERVICE&key=045416EF4DDA69D9%2EF8E7F338DF557316%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=504808115&CFTOKEN=79046804&acm=1429710434_0d9c0bce018bcd071c079ecb15be69e8)
 6. "Power Macromodeling for High Level Power Estimation" (http://delivery.acm.org/10.1145/270000/266171/p365-gupta.pdf?ip=103.27.8.42&id=266171&acc=ACTIVE%20SERVICE&key=045416EF4DDA69D9%2EF8E7F338DF557316%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=504808115&CFTOKEN=79046804&acm=1429710436_686f8f2ffb085b129fe587723a6ee130)
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Register-transfer_level&oldid=1278835625"



Hardware description language

In computer engineering, a **hardware description language (HDL)** is a specialized computer language used to describe the structure and behavior of electronic circuits, usually to design application-specific integrated circuits (ASICs) and to program field-programmable gate arrays (FPGAs).

A hardware description language enables a precise, formal description of an electronic circuit that allows for the automated analysis and simulation of the circuit. It also allows for the synthesis of an HDL description into a netlist (a specification of physical electronic components and how they are connected together), which can then be placed and routed to produce the set of masks used to create an integrated circuit.

A hardware description language looks much like a programming language such as C or ALGOL; it is a textual description consisting of expressions, statements and control structures. One important difference between most programming languages and HDLs is that HDLs explicitly include the notion of time.

HDLs form an integral part of electronic design automation (EDA) systems, especially for complex circuits, such as application-specific integrated circuits, microprocessors, and programmable logic devices.

Motivation

Due to the exploding complexity of digital electronic circuits since the 1970s (see Moore's law), circuit designers needed digital logic descriptions to be performed at a high level without being tied to a specific electronic technology, such as ECL, TTL or CMOS. HDLs were created to implement register-transfer level abstraction, a model of the data flow and timing of a circuit.^[1]

There are two major hardware description languages: VHDL and Verilog. There are different types of description in them: "dataflow, behavioral and structural". Example of dataflow of VHDL:

```
LIBRARY IEEE;
USE IEEE.STD_LOGIC_1164.ALL;

ENTITY not1 IS
  PORT(
    a : IN STD_LOGIC;
    b : OUT STD_LOGIC
  );
END not1;

ARCHITECTURE behavioral OF not1 IS
BEGIN
  b <= NOT a;
END behavioral;
```

Structure of HDL

HDLs are standard text-based expressions of the structure of electronic systems and their behaviour over time. Like concurrent programming languages, HDL syntax and semantics include explicit notations for expressing concurrency. However, in contrast to most software programming languages, HDLs also include an explicit notion of time, which is a primary attribute of hardware. Languages whose only characteristic is to express circuit connectivity between a hierarchy of blocks are properly classified as netlist languages used in electric computer-aided design. HDL can be used to express designs in structural, behavioral or register-transfer-level architectures for the same circuit functionality; in the latter two cases the synthesizer decides the architecture and logic gate layout.

HDLs are used to write executable specifications for hardware. A program designed to implement the underlying semantics of the language statements and simulate the progress of time provides the hardware designer with the ability to model a piece of hardware before it is created physically. It is this executability that gives HDLs the illusion of being programming languages, when they are more precisely classified as specification languages or modeling languages. Simulators capable of supporting discrete-event (digital) and continuous-time (analog) modeling exist, and HDLs targeted for each are available.

Comparison with control-flow languages

It is certainly possible to represent hardware semantics using traditional programming languages such as C++, which operate on control flow semantics as opposed to data flow, although to function as such, programs must be augmented with extensive and unwieldy class libraries. Generally, however, software programming languages do not include any capability for explicitly expressing time, and thus cannot function as hardware description languages. Before the introduction of System Verilog in 2002, C++ integration with a logic simulator was one of the few ways to use object-oriented programming in hardware verification. System Verilog is the first major HDL to offer object orientation and garbage collection.

Using the proper subset of hardware description language, a program called a synthesizer, or logic synthesis tool, can infer hardware logic operations from the language statements and produce an equivalent netlist of generic hardware primitives to implement the specified behaviour. Synthesizers generally ignore the expression of any timing constructs in the text. Digital logic synthesizers, for example, generally use clock edges as the way to time the circuit, ignoring any timing constructs. The ability to have a synthesizable subset of the language does not itself make a hardware description language.

History

The first hardware description languages appeared in the late 1960s, looking like more traditional languages.^[2] The first that had a lasting effect was described in 1971 in C. Gordon Bell and Allen Newell's text *Computer Structures*.^[3] This text introduced the concept of register transfer level, first used in the ISP language to describe the behavior of the Digital Equipment Corporation (DEC) PDP-8.^[4]

The language became more widespread with the introduction of DEC's PDP-16 RT-Level Modules (RTMs) and a book describing their use.^[5] At least two implementations of the basic ISP language (ISPL and ISPS) followed.^{[6][7]} ISPS was well suited to describe relations between the inputs and the outputs of the design and was quickly adopted by commercial teams at DEC, and by several research teams in the US and among its allies in the North Atlantic Treaty Organization (NATO).

The RTM products never succeeded commercially and DEC stopped marketing them in the mid-1980s, as new methods grew more popular, more so very-large-scale integration (VLSI).

Separate work done about 1979 at the University of Kaiserslautern produced a language called KARL ("KAiserslautern Register Transfer Language"), which included design calculus language features supporting VLSI chip floorplanning and structured hardware design. This work was also the basis of KARL's interactive graphic sister language ABL, whose name was an initialism for "A Block diagram Language".^[8] ABL was implemented in the early 1980s by the Centro Studi e Laboratori Telecomunicazioni (CSELT) in Torino, Italy, producing the ABLED graphic VLSI design editor. In the mid-1980s, a VLSI design framework was implemented around KARL and ABL by an international consortium funded by the Commission of the European Union.^[9]

By the late 1970s, design using programmable logic devices (PLDs) became popular, although these designs were primarily limited to designing finite-state machines. The work at Data General in 1980 used these same devices to design the Data General Eclipse MV/8000, and commercial need began to grow for a language that could map well to them. By 1983 Data I/O introduced ABEL to fill that need.

In 1985, as design shifted to VLSI, Gateway Design Automation introduced Verilog, and Intermetrics released the first completed version of the VHSIC Hardware Description Language (VHDL). VHDL was developed at the behest of the United States Department of Defense's Very High Speed Integrated Circuit Program (VHSIC), and was based on the Ada programming language, and on the experience gained with the earlier development of ISPS.^[10] Initially, Verilog and VHDL were used to document and simulate circuit designs already captured and described in another form (such as schematic files). HDL simulation enabled engineers to work at a higher level of abstraction than simulation at the schematic level, and thus increased design capacity from hundreds of transistors to thousands. In 1986, with the support of the U.S Department of Defense, VHDL was sponsored as an IEEE standard (IEEE Std 1076), and the first IEEE-standardized version of VHDL, IEEE Std 1076-1987, was approved in December 1987. Cadence Design Systems later acquired Gateway Design Automation for the rights to Verilog-XL, the HDL simulator that would become the de facto standard of Verilog simulators for the next decade.

The introduction of logic synthesis for HDLs pushed HDLs from the background into the foreground of digital design. Synthesis tools compiled HDL source files (written in a constrained format called RTL) into a manufacturable netlist description in terms of gates and transistors. Writing synthesizable RTL files required practice and discipline on the part of the designer; compared to a traditional schematic layout, synthesized RTL netlists were almost always larger in area and slower in performance. A circuit design from a skilled engineer, using labor-intensive schematic-capture/hand-layout, would almost always outperform its logically-synthesized equivalent, but the productivity advantage held by synthesis soon displaced digital schematic capture to exactly those areas that were problematic for RTL synthesis: extremely high-speed, low-power, or asynchronous circuitry.

Within a few years, VHDL and Verilog emerged as the dominant HDLs in the electronics industry, while older and less capable HDLs gradually disappeared from use. However, VHDL and Verilog share many of the same limitations, such as being unsuitable for analog or mixed-signal circuit simulation. Specialized HDLs (such as Confluence) were introduced with the explicit goal of fixing specific limitations of Verilog and VHDL, though none were ever intended to replace them.

Over the years, much effort has been invested in improving HDLs. The latest iteration of Verilog, formally known as IEEE 1800-2005 SystemVerilog, introduces many new features (classes, random variables, and properties/assertions) to address the growing need for better test bench randomization, design hierarchy, and reuse. A future revision of VHDL is also in development, and is expected to match SystemVerilog's improvements.

Design using HDL

As a result of the efficiency gains realized using HDL, a majority of modern digital circuit design revolves around it. Most designs begin as a set of requirements or a high-level architectural diagram. Control and decision structures are often prototyped in flowchart applications, or entered in a editor. The process of writing the HDL description is highly dependent on the nature of the circuit and the designer's preference for coding style. The HDL is merely the 'capture language', often beginning with a high-level algorithmic description such as a C++ mathematical model. Designers often use scripting languages such as Perl to automatically generate repetitive circuit structures in the HDL language. Special text editors offer features for automatic indentation, syntax-dependent coloration, and macro-based expansion of the entity/architecture/signal declaration.

The HDL code then undergoes a code review, or auditing. In preparation for synthesis, the HDL description is subject to an array of automated checkers. The checkers report deviations from standardized code guidelines, identify potential ambiguous code constructs before they can cause misinterpretation, and check for common logical coding errors, such as floating ports or shorted outputs. This process aids in resolving errors before the code is synthesized.

In industry parlance, HDL design generally ends at the synthesis stage. Once the synthesis tool has mapped the HDL description into a gate netlist, the netlist is passed off to the back-end stage. Depending on the physical technology (FPGA, ASIC gate array, ASIC standard cell), HDLs may or may not play a significant role in the back-end flow. In general, as the design flow progresses toward a physically realizable form, the design database becomes progressively more laden with technology-specific information, which cannot be stored in a generic HDL description. Finally, an integrated circuit is manufactured or programmed for use.

Simulating and debugging HDL code

Essential to HDL design is the ability to simulate HDL programs. Simulation allows an HDL description of a design (called a model) to pass design verification, an important milestone that validates the design's intended function (specification) against the code implementation in the HDL description. It also permits

architectural exploration. The engineer can experiment with design choices by writing multiple variations of a base design, then comparing their behavior in simulation. Thus, simulation is critical for successful HDL design.

To simulate an HDL model, an engineer writes a top-level simulation environment (called a test bench). At minimum, a testbench contains an instantiation of the model (called the device under test or DUT), pin/signal declarations for the model's I/O, and a clock waveform. The testbench code is event driven: the engineer writes HDL statements to implement the (testbench-generated) reset-signal, to model interface transactions (such as a host–bus read/write), and to monitor the DUT's output. An HDL simulator — the program that executes the testbench — maintains the simulator clock, which is the master reference for all events in the testbench simulation. Events occur only at the instants dictated by the testbench HDL (such as a reset-toggle coded into the testbench), or in reaction (by the model) to stimulus and triggering events. Modern HDL simulators have full-featured graphical user interfaces, complete with a suite of debug tools. These allow the user to stop and restart the simulation at any time, insert simulator breakpoints (independent of the HDL code), and monitor or modify any element in the HDL model hierarchy. Modern simulators can also link the HDL environment to user-compiled libraries, through a defined PLI/VHPI interface. Linking is system-dependent (x86, SPARC etc. running Windows/Linux/Solaris), as the HDL simulator and user libraries are compiled and linked outside the HDL environment.

Design verification is often the most time-consuming portion of the design process, due to the disconnect between a device's functional specification, the designer's interpretation of the specification, and the imprecision of the HDL language. The majority of the initial test/debug cycle is conducted in the HDL *simulator* environment, as the early stage of the design is subject to frequent and major circuit changes. An HDL description can also be prototyped and tested in hardware — programmable logic devices are often used for this purpose. Hardware prototyping is comparatively more expensive than HDL simulation, but offers a real-world view of the design. Prototyping is the best way to check interfacing against other hardware devices and hardware prototypes. Even those running on slow FPGAs offer much shorter simulation times than pure HDL simulation.

Design verification with HDLs

Historically, design verification was a laborious, repetitive loop of writing and running simulation test cases against the design under test. As chip designs have grown larger and more complex, the task of design verification has grown to the point where it now dominates the schedule of a design team. Looking for ways to improve design productivity, the electronic design automation industry developed the Property Specification Language.

In formal verification terms, a property is a factual statement about the expected or assumed behavior of another object. Ideally, for a given HDL description, a property or properties can be proven true or false using formal mathematical methods. In practical terms, many properties cannot be proven because they occupy an unbounded solution space. However, if provided a set of operating assumptions or constraints, a property checker can prove (or disprove) certain properties by narrowing the solution space.

The assertions do not model circuit activity, but capture and document the designer's intent in the HDL code. In a simulation environment, the simulator evaluates all specified assertions, reporting the location and severity of any violations. In a synthesis environment, the synthesis tool usually operates with the

policy of halting synthesis upon any violation. Assertion based verification is still in its infancy, but is expected to become an integral part of the HDL design toolset.

HDL and programming languages

An HDL is grossly similar to a software programming language, but there are major differences. Most programming languages are inherently procedural (single-threaded), with limited syntactical and semantic support to handle concurrency. HDLs, on the other hand, resemble concurrent programming languages in their ability to model multiple parallel processes (such as flip-flops and adders) that automatically execute independently of one another. Any change to the process's input automatically triggers an update in the simulator's process stack.

Both programming languages and HDLs are processed by a compiler (often called a synthesizer in the HDL case), but with different goals. For HDLs, "compiling" refers to logic synthesis; the process of transforming the HDL code listing into a physically realizable gate netlist. The netlist output can take any of many forms: a "simulation" netlist with gate-delay information, a "handoff" netlist for post-synthesis placement and routing on a semiconductor die, or a generic industry-standard Electronic Design Interchange Format (EDIF) (for subsequent conversion to a JEDEC-format file).

On the other hand, a software compiler converts the source-code listing into a microprocessor-specific object code for execution on the target microprocessor. As HDLs and programming languages borrow concepts and features from each other, the boundary between them is becoming less distinct. However, pure HDLs are unsuitable for general purpose application software development, just as general-purpose programming languages are undesirable for modeling hardware.

Yet as electronic systems grow increasingly complex, and reconfigurable systems become increasingly common, there is growing desire in the industry for a single language that can perform some tasks of both hardware design and software programming. SystemC is an example of such—an embedded system hardware can be modeled as non-detailed architectural blocks (black boxes with modeled signal inputs and output drivers). The target application is written in C or C++ and natively compiled for the host-development system; as opposed to targeting the embedded CPU, which requires host-simulation of the embedded CPU or an emulated CPU.

The high level of abstraction of SystemC models is well suited to early architecture exploration, as architectural modifications can be easily evaluated with little concern for signal-level implementation issues. However, the threading model used in SystemC relies on shared memory, causing the language not to handle parallel execution or low-level models well.

High-level synthesis

In their level of abstraction, HDLs have been compared to assembly languages. There are attempts to raise the abstraction level of hardware design in order to reduce the complexity of programming in HDLs, creating a sub-field called high-level synthesis.

Companies such as Cadence, Synopsys and Agility Design Solutions are promoting SystemC as a way to combine high-level languages with concurrency models to allow faster design cycles for FPGAs than is possible using traditional HDLs. Approaches based on standard C or C++ (with libraries or other extensions allowing parallel programming) are found in the Catapult C tools from Mentor Graphics, and the Impulse C tools from Impulse Accelerated Technologies.

A similar initiative from Intel is the use of Data Parallel C++, related to SYCL, as a high-level synthesis language.

Annapolis Micro Systems (<https://www.annapmicro.com/>), Inc.'s CoreFire Design Suite^[11] and National Instruments LabVIEW FPGA provide a graphical dataflow approach to high-level design entry and languages such as SystemVerilog, SystemVHDL, and Handel-C seek to accomplish the same goal, but are aimed at making existing hardware engineers more productive, rather than making FPGAs more accessible to existing software engineers.

It is also possible to design hardware modules using MATLAB and Simulink using the MathWorks HDL Coder tool^[12] or DSP Builder for Intel FPGAs^[13] or Xilinx System Generator (XSG) from Xilinx.^[14]

Examples of HDLs

HDLs for analog circuit design

Name	Description
<u>HDL-A</u>	A proprietary analog HDL
<u>SpectreHDL</u>	A proprietary analog HDL from <u>Cadence Design Systems</u> for its Spectre circuit simulator
<u>Verilog-AMS</u> (Verilog for Analog and Mixed-Signal)	An <u>Accellera</u> standard extension of IEEE Std 1364 <u>Verilog</u> for analog and mixed-signal simulation
<u>VHDL-AMS</u> (VHDL with Analog/Mixed-Signal extension)	An <u>IEEE</u> standard extension (IEEE Std 1076.1) of <u>VHDL</u> for analog and mixed-signal simulation

HDLs for digital circuit design

The two most widely used and well-supported HDL varieties used in industry are Verilog and VHDL.

Status	Name	Host language	Description
In use	Altera Hardware Description Language (AHDL)		Proprietary language from Altera
	A Hardware Programming language (AHPL)		Used as a tool for teaching
	Amaranth (https://github.com/amaranth-lang/amaranth)	Python	
	Bluespec		High-level HDL based on Haskell (not embedded DSL) ^[15]
	Bluespec SystemVerilog (BSV)		Based on Bluespec , with Verilog HDL like syntax, by Bluespec, Inc.
	C-to-Verilog		Converter from C to Verilog
	Chisel (Constructing Hardware in a Scala Embedded Language) ^[16]	Scala	Based on Scala (embedded DSL)
	Clash (https://clash-lang.org/)		Functional hardware description language that borrows its syntax and semantics from the functional language Haskell
	Common Oriented Language for Architecture of Multi Objects (COLAMO) ^{[17][18]}		Proprietary language from “Supercomputers and Neurocomputers Research Center” Co Ltd.
	Compiler for Universal Programmable Logic (CUPL) ^[19]		Proprietary language from Logical Devices, Inc.
	DSLX (https://google.github.io/xls/dslx_reference)		Domain-specific language for XLS toolchain
	ESys.net		.NET framework written in C#
	Filament (https://filamenthdl.com/)		HDL with a type system inspired by Rust
	Handel-C		C-like design language
	Hardcaml (https://github.com/ujamjar/hardcaml)	OCaml	Based on OCaml (embedded DSL) ^[20]
	HDL	Haskell	Based on Haskell (embedded DSL)
	Hardware Join Java (HJJ)	Join Java	Based on Join Java
	Hardware ML (HML)	Standard ML	Based on Standard ML ^[21]
	Hydra	Haskell	Based on Haskell
	Impulse C		C-like HDL
	Parallel C++ (ParC)		kusu extended with HDL style threading and communication for task-parallel programming
	JHDL	Java	Based on Java
	Lava	Haskell	Based on Haskell (embedded DSL) ^{[22][23][24][25]}
	Lola		Simple language used for teaching
	M		HDL from Mentor Graphics

Migen (https://github.com/m-labs/migen)	Python	
MyHDL	Python	Based on Python (embedded DSL)
PALASM		For Programmable Array Logic (PAL) devices
PipelineC (https://github.com/J JulianKemmerer/PipelineC)		C-like hardware description language adding High-level synthesis-like automatic pipelining as a language construct and compiler feature.
PyMTL 3 (Mamba) (https://github.com/mambalab/PyMTL3)	Python	Based on Python, from Cornell University
PyRTL (https://gitlab.ics.uci.edu/PyRTL)	Python	Based on Python, from University of California, Santa Barbara
Riverside Optimizing Compiler for Configurable Computing (ROCCC)		Free and open-source C to HDL tool
RHDL	Ruby	Based on the Ruby programming language
Rapid Open Hardware Development (ROHD) ^[26]	Dart	Framework for hardware design and verification, written in Dart
Ruby (hardware description language)		
Silice (https://github.com/sylefeb/Silice)		HDL that simplifies designing hardware algorithms with parallelism and pipelines
Spade (https://spade-lang.org/)		HDL inspired by modern software languages like Rust
SystemC		Standardized class of C++ libraries for high-level behavioral and transaction modeling of digital hardware at a high level of abstraction, i.e., system-level
SystemVerilog		Superset of Verilog, with enhancements to address system-level design and verification
SpinalHDL (https://github.com/SpinalHDL/SpinalHDL)	Scala	Based on Scala (embedded DSL)
SystemTCL		SDL based on Tcl
Templated HDL inspired by C++ (THDL++)		Extension of VHDL with inheritance, advanced templates and policy classes
Verik (https://github.com/frwango96/verik)		Kotlin reinterpreted with the semantics of an HDL; transpiled to SystemVerilog
Transaction-Level Verilog (TL-Verilog) ^[27]		Extension of Verilog/SystemVerilog with constructs for pipelines and transactions .
Verilog		One of the most widely used and well-supported HDLs
Veryl (https://veryl-lang.org/)		HDL designed as SystemVerilog alternative
VHDL (VHSIC HDL)		One of the most widely used and well-supported HDLs
No longer in common use	Advanced Boolean Expression Language (ABEL)	Obsolete HDL made by Data I/O Corporation in 1983
	Confluence	Functional HDL, discontinued

CoWareC		C-based HDL by CoWare ; discontinued in favor of SystemC
ELLA		No longer in common use
ISPS		Original HDL from CMU; no longer in common use
KAiserslautern Register Language (KARL) ^[9]		Pascal-like hardware description language; no longer in common use
nMigen (https://gitlab.com/nmigen/nmigen)	Python	Predecessor to Amaranth

HDLs for printed circuit board design

Several projects exist for defining printed circuit board connectivity using language based, textual-entry methods. Among these, new approaches have emerged that focus on enhancing readability, reusability, and validation. These modern methodologies employ open-source design languages specifically tailored for electronics, adopting declarative markup to specify what circuits should achieve. This shift integrates software development principles into hardware design, streamlining the process and emphasizing automation, reuse, and validation.

Name	Description
atopile (https://github.com/atopile/atopile)	An open-source language and toolchain to describe electronic circuit boards with code.
PHDL (PCB HDL)	A free and open source HDL for defining printed circuit board connectivity.
EDAsolver	An HDL for solving schematic designs based on constraints.
SKiDL (https://github.com/xesscorp/skidl)	Open source Python module to design electronic circuits.

See also

- [Bluespec](#)
- [C to HDL](#)
- [Flow to HDL](#)
- [Gezel](#)
- [Hardware verification language](#)
- [Modeling language](#)
- [Property Specification Language](#)
- [Rosetta-lang](#)
- [Specification language](#)
- [SystemC](#)
- [SystemVerilog](#)

References

1. Ciletti, Michael D. (2011). *Advanced Digital Design with Verilog HDL* (2nd ed.). Prentice

- Hall. ISBN 9780136019282.
2. Barbacci, M. "A comparison of register transfer languages for describing computers and digital systems," Carnegie-Mellon Univ., Dept. of Computer Science, March 1973
 3. Bell, C. G.; Newell, A. (1971). *Computer Structures: Readings and Examples* (<https://archive.org/details/computerstructur00bell>). McGraw-Hill. ISBN 0-07-004357-4.
 4. Reilly, E.D. (2003). *Milestones in computer science and information technology* (<https://archive.org/details/milestonesincomp0000reil>). Greenwood Press. p. 183 (<https://archive.org/details/milestonesincomp0000reil/page/183>). ISBN 1-57356-521-0.
 5. Bell, C.G.; Grason, J.; Newell, A. (1972). *Designing Computers and Digital Systems*. Digital Press. LCCN 72-89566 (<https://lccn.loc.gov/72-89566>). OCLC 440245727 (<https://search.worldcat.org/oclc/440245727>).
 6. Barbacci, M.C. (1976). The Symbolic Manipulation of Computer Descriptions: ISPL Compiler and Simulator (Report). Department of Computer Science, Carnegie-Mellon University. doi:[10.1184/R1/6610790.v1](https://doi.org/10.1184/R1/6610790.v1) (<https://doi.org/10.1184%2FR1%2F6610790.v1>).
 7. Barbacci, M.C.; Barnes, G.E.; Cattell, R.G.G.; Siewiorek, D.P. (1977). The ISPS Computer Description Language (Report). Department of Computer Science, Carnegie-Mellon University. doi:[10.1184/R1/6610637.v1](https://doi.org/10.1184/R1/6610637.v1) (<https://doi.org/10.1184%2FR1%2F6610637.v1>).
 8. Girardi, G.; Hartenstein, R. (1983). ABL specification (Report). CSELT and University of Kaiserslautern.
 9. Hartenstein, Reiner W. (2012) [1993], "KARL and ABL" (<https://books.google.com/books?id=DBrtCAAAQBAJ&pg=PA447>), in Mermet, J. (ed.), *Fundamentals and Standards in Hardware Description Languages*, NATO Science Series E, vol. 249, Springer, pp. 447–, ISBN 9789401119146
 10. Barbacci, M.C.; Grout, S.; Lindstrom, G.; Maloney, M.P. (1984). Ada as a hardware description language : an initial report (Report). Department of Computer Science, Carnegie-Mellon University. CiteSeerX 10.1.1.938.8003 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.938.8003>). doi:[10.1184/R1/6602984.v1](https://doi.org/10.1184/R1/6602984.v1) (<https://doi.org/10.1184%2FR1%2F6602984.v1>).
 11. "VHDL-Based FPGA Programming Application Software Tool" (<https://www.annapmicro.com/products/corefire-next/>). Annapolis Micro Systems, Inc. Retrieved 2018-12-01.
 12. "VHDL code - HDL Coder - MATLAB & Simulink" (<http://www.mathworks.com/products/hdl-coder/>). Mathworks.com. 2011-04-30. Retrieved 2012-08-11.
 13. "Digital Signal Processing (DSP) Builder - Intel® FPGAs" (<https://www.intel.com/content/www/us/en/software/programmable/quartus-prime/dsp-builder.html>). Intel. Retrieved 2021-09-20.
 14. "System Generator for DSP" (<https://web.archive.org/web/20120712224415/http://www.xilinx.com/tools/sysgen.htm>). Xilinx.com. Archived from the original (<http://www.xilinx.com/tools/sysgen.htm>) on 2012-07-12. Retrieved 2012-08-11.
 15. A History of Haskell: being lazy with class §12.4.2
 16. "Chisel/FIRRTL Hardware Compiler Framework" (<https://www.chisel-lang.org/>).
 17. COLAMO (<https://web.archive.org/web/20091109155107/http://colamo.parallel.ru/>)
 18. "Higher-level language COLAMO – НИЦ супер-ЭВМ и нейрокомпьютеров" (<http://superevm.ru/index.php?page=higher-level-language-colamo>).
 19. Eurich, J.P.; Roth, G. (1990). "EDIF grows up". *IEEE Spectrum*. 27 (11): 68–72. doi:[10.1109/6.62219](https://doi.org/10.1109/6.62219) (<https://doi.org/10.1109%2F6.62219>). S2CID 381119 (<https://api.semanticscholar.org/CorpusID:381119>).
 20. Hardcaml (<http://www.ujamjar.com/hardcaml/>)

21. Li, Yanbing; Leeser, M. (1995). "HML: An innovative hardware description language and its translation to VHDL". *Proceedings of ASP-DAC'95/CHDL'95/VLSI'95 with EDA Technofair*. pp. 691–696. doi:10.1109/ASPDAC.1995.486388 (<https://doi.org/10.1109%2FASPDAC.1995.486388>). ISBN 4-930813-67-0. S2CID 14198160 (<https://api.semanticscholar.org/CorpusID:14198160>).
22. Chalmers Lava (<http://hackage.haskell.org/package/chalmers-lava2000>)
23. Xilinx Lava (<http://hackage.haskell.org/package/xilinx-lava>)
24. Kansas Lava (<http://hackage.haskell.org/package/kansas-lava>)
25. York Lava (<http://hackage.haskell.org/package/york-lava>)
26. "Rapid Open Hardware Development (ROHD) Framework" (<https://github.com/intel/rohd>). GitHub. 17 November 2021.
27. TL-Verilog (<http://www.tl-x.org>)

External links

- [HCT](http://hct.sourceforge.net/) (<http://hct.sourceforge.net/>) - The HDL Complexity tool, used to determine design complexity.
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Hardware_description_language&oldid=1269959964"

High-level synthesis

High-level synthesis (HLS), sometimes referred to as **C synthesis**, **electronic system-level (ESL) synthesis**, **algorithmic synthesis**, or **behavioral synthesis**, is an automated design process that takes an abstract behavioral specification of a digital system and finds a register-transfer level structure that realizes the given behavior.^{[1][2][3]}

Synthesis begins with a high-level specification of the problem, where behavior is generally decoupled from low-level circuit mechanics such as clock-level timing. Early HLS explored a variety of input specification languages,^[4] although recent research and commercial applications generally accept synthesizable subsets of ANSI C/C++/SystemC/MATLAB. The code is analyzed, architecturally constrained, and scheduled to transcompile from a transaction-level model (TLM) into a register-transfer level (RTL) design in a hardware description language (HDL), which is in turn commonly synthesized to the gate level by the use of a logic synthesis tool.

The goal of HLS is to let hardware designers efficiently build and verify hardware, by giving them better control over optimization of their design architecture, and through the nature of allowing the designer to describe the design at a higher level of abstraction while the tool does the RTL implementation. Verification of the RTL is an important part of the process.^[5]

Hardware can be designed at varying levels of abstraction. The commonly used levels of abstraction are gate level, register-transfer level (RTL), and algorithmic level.

While logic synthesis uses an RTL description of the design, high-level synthesis works at a higher level of abstraction, starting with an algorithmic description in a high-level language such as SystemC and ANSI C/C++. The designer typically develops the module functionality and the interconnect protocol. The high-level synthesis tools handle the micro-architecture and transform untimed or partially timed functional code into fully timed RTL implementations, automatically creating cycle-by-cycle detail for hardware implementation.^[6] The (RTL) implementations are then used directly in a conventional logic synthesis flow to create a gate-level implementation.

History

Early academic work extracted scheduling, allocation, and binding as the basic steps for high-level-synthesis. Scheduling partitions the algorithm in control steps that are used to define the states in the finite-state machine. Each control step contains one small section of the algorithm that can be performed in a single clock cycle in the hardware. Allocation and binding maps the instructions and variables to the hardware components, multiplexers, registers and wires of the data path.

First generation behavioral synthesis was introduced by Synopsys in 1994 as Behavioral Compiler^[7] and used Verilog or VHDL as input languages. The abstraction level used was partially timed (clocked) processes. Tools based on behavioral Verilog or VHDL were not widely adopted in part because neither languages nor the partially timed abstraction were well suited to modeling behavior at a high level. 10 years later, in early 2004, Synopsys end-of-lifed Behavioral Compiler.^[8]

In 1998, Forte Design Systems introduced its Cynthesizer tool which used SystemC as an entry language instead of Verilog or VHDL. Cynthesizer was adopted by many Japanese companies in 2000 as Japan had a very mature SystemC user community. The first high-level synthesis tapeout was achieved in 2001 by Sony using Cynthesizer. Adoption in the United States started in earnest in 2008.

In 2006, an efficient and scalable "SDC modulo scheduling" technique was developed on control and data flow graphs^[9] and was later extended to pipeline scheduling.^[10] This technique uses the integer linear programming formulation. But it shows that the underlying constraint matrix is totally unimodular (after approximating the resource constraints). Thus, the problem can be solved in polynomial time optimally using a linear programming solver in polynomial time. This work was inducted to the FPGA and Reconfigurable Computing Hall of Fame 2022.^[11]

The SDC scheduling algorithm was implemented in the xPilot HLS system^[12] developed at UCLA,^[13] and later licensed to the AutoESL Design Technologies, a spin-off from UCLA. AutoESL was acquired by Xilinx (now part of AMD) in 2011,^[11] and the HLS tool developed by AutoESL became the base of Xilinx HLS solutions, Vivado HLS and Vitis HLS, widely used for FPGA designs.

Source input

The most common source inputs for high-level synthesis are based on standard languages such as ANSI C/C++, SystemC and MATLAB.

High-level synthesis typically also includes a bit-accurate executable specification as input, since to derive an efficient hardware implementation, additional information is needed on what is an acceptable Mean-Square Error or Bit-Error Rate etc. For example, if the designer starts with an FIR filter written using the "double" floating type, before he can derive an efficient hardware implementation, they need to perform numerical refinement to arrive at a fixed-point implementation. The refinement requires additional information on the level of quantization noise that can be tolerated, the valid input ranges etc. This bit-accurate specification makes the high level synthesis source specification functionally complete.^[14] Normally the tools infer from the high level code a Finite State Machine and a Datapath that implement arithmetic operations.

Process stages

The high-level synthesis process consists of a number of activities. Various high-level synthesis tools perform these activities in different orders using different algorithms. Some high-level synthesis tools combine some of these activities or perform them iteratively to converge on the desired solution.^[15]

- Lexical processing
- Algorithm optimization
- Control/Dataflow analysis
- Library processing
- Resource allocation
- Scheduling
- Functional unit binding
- Register binding
- Output processing
- Input Rebundling

Functionality

In general, an algorithm can be performed over many clock cycles with few hardware resources, or over fewer clock cycles using a larger number of ALUs, registers and memories. Correspondingly, from one algorithmic description, a variety of hardware microarchitectures can be generated by an HLS compiler according to the directives given to the tool. This is the same trade off of execution speed for hardware complexity as seen when a given program is run on conventional processors of differing performance, yet all running at roughly the same clock frequency.

Architectural constraints

Synthesis constraints for the architecture can automatically be applied based on the design analysis.^[5] These constraints can be broken into

- Hierarchy
- Interface
- Memory
- Loop
- Low-level timing constraints
- Iteration

Interface synthesis

Interface Synthesis refers to the ability to accept pure C/C++ description as its input, then use automated interface synthesis technology to control the timing and communications protocol on the design interface. This enables interface analysis and exploration of a full range of hardware interface options such as streaming, single- or dual-port RAM plus various handshaking mechanisms. With interface synthesis the designer does not embed interface protocols in the source description. Examples might be: direct connection, one line, 2 line handshake, FIFO.^[16]

Vendors

Data reported on recent Survey^[17]

Status	Compiler	Owner	License	Input	Output	Year	Domain	Test bench	FP	FixP
In use	Stratus HLS (http://www.cadence.com/en_US/home/tools/digital-design-and-signoff/synthesis/stratus-high-level-synthesis.html)	Cadence Design Systems	Commercial	C-C++ SystemC	RTL	2015	All	Yes	Yes	Yes
	AUGH (http://tim.a.imag.fr/sls/research-projects/augh/)	TIMA Lab.	Academic	C subset	VHDL	2012	All	Yes	No	No
	eXCite (http://www.yxi.com) Archived (https://web.archive.org/web/20190917062628/http://www.yxi.com/) 2019-09-17 at the Wayback Machine	Y Explorations	Commercial	C	VHDL–Verilog	2001	All	Yes	No	Yes
	Bambu (http://panda.dei.polimi.it/?page_id=81)	PoliMi	Academic	C	VHDL–Verilog	2012	All	Yes	Yes	No
	Bluespec	BlueSpec, Inc.	BSD-3	Bluespec SystemVerilog (Haskell)	SystemVerilog	2007	All	No	No	No
	QCC (https://cacheq.com/)	CacheQ Systems, Inc. (https://cacheq.com/)	Commercial	C, C++, Fortran	Host executable + FPGA bit file (SystemVerilog is intermediate)	2018	All - multi-core and heterogeneous compute	Yes (C++)	Yes	Yes
	CHC	Altium	Commercial	C subset	VHDL–Verilog	2008	All	No	Yes	Yes
	CoDeveloper	Impulse Accelerated	Commercial	Impulse-C	VHDL	2003	Image streaming	Yes	Yes	No
	HDL Coder (http://www.mathworks.com/products/hdl-coder.html)	MathWorks	Commercial	MATLAB, Simulink, Stateflow, Simscape	VHDL, Verilog	2003	Control systems, signal processing, wireless, radar, communications, image and computer vision	Yes	Yes	Yes
	CyberWorkBench (http://www.cyberworkbench.com)	NEC	Commercial	C, BDL, SystemC	VHDL–Verilog	2004	All	Cycle, formal	Yes	Yes
	Catapult (https://eda.sw.siemens.com/en-US/ic/catapult-high-level-synthesis)	Siemens EDA	Commercial	C-C++ SystemC	VHDL–Verilog	2004	All	Yes	Yes	Yes
	DWARV	TU. Delft	Academic	C subset	VHDL	2012	All	Yes	Yes	Yes
	GAUT (http://www.w.gaut.fr)	University of Western Brittany	Academic	C, C++	VHDL	2010	DSP	Yes	No	Yes
	Hastlayer (http://hastlayer.com)	Lombiq Technologies	BSD-3	C#, C++, F#, ... (.NET)	VHDL	2015	.NET	Yes	Yes	Yes
	Instant SoC (http://www.fpga-cores.com/instant-soc/)	FPGA Cores	Commercial	C, C++	VHDL–Verilog	2019	All	Yes	No	No
	Intel High Level Synthesis Compiler (https://www.intel.com/content/www/us/en/software/programmable/quartus-prime/hls-compiler.html)	Intel FPGA (Formerly Altera)	Commercial	C, C++	Verilog	2017	All	Yes	Yes	Yes
	LegUp HLS (http://www.legupcomputing.com)	LegUp Computing	Commercial	C, C++	Verilog	2015	All	Yes	Yes	Yes

	<u>LegUp</u> (http://legup.eecg.utoronto.ca) Archived (https://web.archive.org/web/20200724081231/http://legup.eecg.utoronto.ca/) 2020-07-24 at the Wayback Machine	University of Toronto	Academic	C	Verilog	2010	All	Yes	Yes	No
MaxCompiler	Maxeler	<u>Commercial</u>	MaxJ	RTL	2010	<u>Data-flow analysis</u>	No	Yes	No	
ROCCC	Jacquard Comp.	<u>Commercial</u>	C subset	VHDL	2010	Streaming	No	Yes	No	
Symphony C	Synopsys	<u>Commercial</u>	C, C++	VHDL–Verilog, SystemC	2010	All	Yes	No	Yes	
VivadoHLS (http://www.xilinx.com/products/design-tools/vivado/) (formerly AutoPilot from AutoESL ^[18])	Xilinx	<u>Commercial</u>	C-C++ SystemC	VHDL–Verilog, SystemC	2013	All	Yes	Yes	Yes	
Kiwi (http://www.cl.cam.ac.uk/research/srg/han/papers/orangepath/kiwi.c.html)	University of Cambridge	Academic	C#	Verilog	2008	.NET	No	Yes	Yes	
CHiMPS	University of Washington	Academic	C	VHDL	2008	All	No	No	No	
gcc2verilog	Korea University	Academic	C	Verilog	2011	All	No	No	No	
HercuLeS (http://www.nkavvadias.com/hercules/)	Ajax Compilers	<u>Commercial</u>	C/NAC	VHDL	2012	All	Yes	Yes	Yes	
Shang (https://github.com/etherzhb/Shang)	University of Illinois Urbana-Champaign	Academic	C	Verilog	2013	All	Yes	?	?	
Trident	Los Alamos NL	Academic	C subset	VHDL	2007	Scientific	No	Yes	No	
Abandoned	AccelDSP	Xilinx	<u>Commercial</u>	MATLAB	VHDL–Verilog	2006	DSP	Yes	Yes	Yes
	C2H	Altera	<u>Commercial</u>	C	VHDL–Verilog	2006	All	No	No	No
	CtoVerilog	University of Haifa	Academic	C	Verilog	2008	All	No	No	No
	DEFACTO	University South Caifl.	Academic	C	RTL	1999	DSE	No	No	No
	Garp	University of California, Berkeley	Academic	C subset	bitstream	2000	Loop	No	No	No
	MATCH	Northwest University	Academic	MATLAB	VHDL	2000	Image	No	No	No
	Napa-C	Sarnoff Corp.	Academic	C subset	VHDL–Verilog	1998	Loop	No	No	No
	PipeRench	Carnegie Mellon University	Academic	DIL	bitstream	2000	Stream	No	No	No
	SA-C	University of Colorado	Academic	SA-C	VHDL	2003	Image	No	No	No
	SeaCucumber	Brigham Young University	Academic	Java	EDIF	2002	All	No	Yes	Yes
	SPARK	University of California, Irvine	Academic	C	VHDL	2003	Control	No	No	No

- Dynamatic (<https://dynamatic.epfl.ch/>) from EPFL/ETH Zurich
- MATLAB HDL Coder [1] (<https://www.mathworks.com/products/hdl-coder.html>) from Mathworks^[19]
- HLS-QSP from CircuitSutra Technologies^[20]
- C-to-Silicon from Cadence Design Systems
- Concurrent Acceleration from Concurrent EDA
- Symphony C Compiler from Synopsys
- QuickPlay from PLDA^[21]
- PowerOpt from ChipVision^[22]
- Cynthesizer from Forte Design Systems (now Stratus HLS from Cadence Design Systems)
- Catapult C from Calypto Design Systems, part of Mentor Graphics as of 2015, September 16. In November 2016 Siemens announced plans to acquire Mentor Graphics, Mentor Graphics became styled as "Mentor, a Siemens Business". In January 2021, the legal merger of Mentor Graphics with Siemens was completed - merging into the Siemens Industry Software Inc legal entity. Mentor Graphics' name was changed to Siemens EDA, a division of Siemens Digital Industries Software.^[23]

- PipelineC [2] (<https://github.com/J JulianKemmerer/PipelineC>)
- CyberWorkBench from NEC^[24]
- Mega Hardware^[25]
- C2R from CebaTech^[26]
- CoDeveloper from Impulse Accelerated Technologies
- HercuLeS by Nikolaos Kavvadias^[27]
- Program In/Code Out (PICO) from Synfora, acquired by Synopsys in June 2010^[28]
- xPilot from University of California, Los Angeles^[29]
- Vsyn from vsyn.ru^[30]
- ngDesign from SynFlow^[31]

See also

- [C to HDL](#)
- [Electronic design automation \(EDA\)](#)
- [Electronic system-level \(ESL\)](#)
- [Logic synthesis](#)
- [High-level verification \(HLV\)](#)
- [SystemVerilog](#)
- [Hardware acceleration](#)

References

1. Coussy, Philippe; Morawiec, Adam, eds. (2008). *High-Level Synthesis - Springer*. doi:[10.1007/978-1-4020-8588-8](https://doi.org/10.1007/978-1-4020-8588-8) (<https://doi.org/10.1007%2F978-1-4020-8588-8>). ISBN 978-1-4020-8587-1.
2. McFarland, M.C.; Parker, A.C.; Camposano, R. (February 1990). "The high-level synthesis of digital systems" (<https://ieeexplore.ieee.org/document/52214>). *Proceedings of the IEEE*. **78** (2): 301–318. doi:[10.1109/5.52214](https://doi.org/10.1109/5.52214) (<https://doi.org/10.1109%2F5.52214>). ISSN 1558-2256 (<https://search.worldcat.org/issn/1558-2256>).
3. "HLS Book : Home" (<https://www.hlsbook.com/>). www.hlsbook.com. Retrieved 2023-06-21.
4. IEEE Xplore High-Level Synthesis: Past, Present, and Future (https://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5209959) DOI 10.1109/MDT.2009.83
5. Bowyer, Bryan (2005-05-02). "The 'why' and 'what' of algorithmic synthesis" (<https://www.eetimes.com/the-why-and-what-of-algorithmic-synthesis/>). EE Times. Retrieved 2016-10-03.
6. "C-Based Rapid Prototyping for Digital Signal Processing" (<http://www.ee.bilkent.edu.tr/~signal/defevent/papers/cr1179.pdf>) (PDF). UBS University, France. Retrieved 2016-10-03.
7. "Publications and Presentations" (https://web.archive.org/web/20080426075303/http://www.bdti.com/articles/info_dspx95asics.htm). *Bdti.com*. Archived from the original (http://www.bdti.com/articles/info_dspx95asics.htm) on 2008-04-26. Retrieved 2016-10-03.
8. "Behavioral synthesis crossroad" (http://www.eetimes.com/news/design/columns/tool_talk/showArticle.jhtml?articleID=18700196). EE Times. Retrieved 2016-10-03.
9. Cong, Jason; Fan, Yiping; Han, Guoling; Jiang, Wei; Zhang, Zhiru (September 2006). "Platform-Based Behavior-Level and System-Level Synthesis" (<https://dx.doi.org/10.1109/socc.2006.283880>). *2006 IEEE International SOC Conference*. IEEE. pp. 199–202. doi:[10.1109/socc.2006.283880](https://doi.org/10.1109/socc.2006.283880) (<https://doi.org/10.1109%2Fsocc.2006.283880>). ISBN 0-7803-9782-7.
10. Zhang, Zhiru; Liu, Bin (2013). *2013 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)* (<https://www.csl.rnellen.edu/~zhiruz/pdfs/sdcmmod-iccad2013.pdf>) (PDF). IEEE. pp. 211–218. ISBN 978-1-4799-1071-7.
11. Cong, Jason; Bin Liu; Neuendorffer, Stephen; Noguera, Juanjo; Vissers, Kees; Zhiru Zhang (April 2011). "High-Level Synthesis for FPGAs: From Prototyping to Deployment" (<https://dx.doi.org/10.1109/tcad.2011.2110592>). *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. **30** (4): 473–491. doi:[10.1109/tcad.2011.2110592](https://doi.org/10.1109/tcad.2011.2110592) (<https://doi.org/10.1109%2Ftcad.2011.2110592>). ISSN 0278-0070 (<https://search.worldcat.org/issn/0278-0070>).
12. Cong, J.; Zhiru Zhang (2006). "An efficient and versatile scheduling algorithm based on SDC formulation" (<https://dx.doi.org/10.1109/dac.2006.229228>). *2006 43rd ACM/IEEE Design Automation Conference*. IEEE. pp. 433–438. doi:[10.1109/dac.2006.229228](https://doi.org/10.1109/dac.2006.229228) (<https://doi.org/10.1109%2Fdac.2006.229228>). ISBN 1-59593-381-6.
13. "xPilot: Platform-based Behavior Synthesis System | VAST lab" (<https://vast.cs.ucla.edu/software/xpilot-platform-based-behavior-synthesis-system>). vast.cs.ucla.edu. Retrieved 2024-04-18.
14. Multiple Word-Length High-Level Synthesis (<http://www.hindawi.com/GetArticle.aspx?doi=10.1155/2008/916867&e=html>) EURASIP Journal on Embedded Systems
15. "A look inside behavioral synthesis" (http://www.eetimes.com/document.asp?doc_id=1217645). EE Times. Retrieved 2016-10-03.
16. "DesignCon: InfoVault: Paper Library" (https://web.archive.org/web/20100925034609/http://www.designcon.com/infovault/paper.asp?PAPER_ID=407). www.designcon.com. Archived from the original (http://www.designcon.com/infovault/paper.asp?PAPER_ID=407) on 25 September 2010. Retrieved 13 January 2022.

17. Nane, R.; Sima, V. M.; Pilato, C.; Choi, J.; Fort, B.; Canis, A.; Chen, Y. T.; Hsiao, H.; Brown, S. (2016). "A Survey and Evaluation of FPGA High-Level Synthesis Tools" (<https://re.public.polimi.it/bitstream/11311/998432/1/TCADHLSEVAL2016.pdf>) (PDF). *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. **35** (10): 1591–1604. doi:[10.1109/TCAD.2015.2513673](https://doi.org/10.1109/TCAD.2015.2513673) (<https://doi.org/10.1109%2FTCAD.2015.2513673>). hdl:[11311/998432](https://hdl.handle.net/11311%2F998432) (<https://hdl.handle.net/11311%2F998432>). ISSN 0278-0070 (<https://search.worldcat.org/issn/0278-0070>). S2CID 8749577 (<https://api.semanticscholar.org/CorpusID:8749577>).
18. "Xilinx buys high-level synthesis EDA vendor" (<https://web.archive.org/web/20111017170928/http://www.eetimes.com/electronics-news/4212668/Xilinx-buys-high-level-synthesis-EDA-vendor>). EE Times. 2011-02-05. Archived from the original (<http://www.eetimes.com/electronics-news/4212668/Xilinx-buys-high-level-synthesis-EDA-vendor>) on 2011-10-17. Retrieved 2016-10-03.
19. "MathWorks: Makers of MATLAB and Simulink" (<http://mathworks.com/>). *Mathworks.com*. Retrieved 2016-10-03.
20. "SystemC based ESL methodologies - SystemC based ESL methodologies" (<http://www.circuitsutra.com/>). *Circuitsutra.com*. Retrieved 2016-10-03.
21. John M. at a major ERP & DBMS Corporation (2016-08-29). "QuickPlay: Bringing FPGA Computing to the Masses" (<http://www.quickplay.io>). *Quickplay.io*. Retrieved 2016-10-03.
22. "Chipvision - Fast Track to Low Power" (<https://web.archive.org/web/20020530041720/http://www.chipvision.com/>). *www.chipvision.com*. Archived from the original (<http://www.chipvision.com/>) on 30 May 2002. Retrieved 13 January 2022.
23. "Mentor Finally Becomes Siemens EDA from January 2021" (<https://www.eetimes.com/mentor-finally-becomes-siemens-eda-from-january-2021/>). 15 December 2020.
24. "CyberWorkBench: Products" (<http://www.nec.com/en/global/prod/cwb/index.html>). NEC. Retrieved 2016-10-03.
25. "Accueil mega-hardware" (<https://web.archive.org/web/20040115020254/http://www.mega-hardware.com/>). *www.mega-hardware.com*. Archived from the original (<http://www.mega-hardware.com/>) on 15 January 2004. Retrieved 13 January 2022.
26. "Cebatech - Home" (<https://web.archive.org/web/20050507101111/http://www.cebatech.com/>). *www.cebatech.com*. Archived from the original (<http://www.cebatech.com/>) on 7 May 2005. Retrieved 13 January 2022.
27. "Nikolaos Kavvadias - HerculeS high-level synthesis tool" (<http://www.nkavvadias.com/hercules/>). *Nkavvadias.com*. Retrieved 2016-10-03.
28. "Synopsys buys Synfora assets" (<https://web.archive.org/web/20110407145420/http://www.eetimes.com/electronics-news/4200083/Synopsys-buys-Synfora-assets>). EE Times. Archived from the original (<http://www.eetimes.com/electronics-news/4200083/Synopsys-buys-Synfora-assets>) on 2011-04-07. Retrieved 2016-10-03.
29. "The xPilot System" (<http://cadlab.cs.ucla.edu/soc/>). *Cadlab.cs.ucla.edu*. Retrieved 2016-10-03.
30. "vSyn.ru" (<https://web.archive.org/web/20160630171129/http://www.vsyn.ru/>). *vSyn.ru*. 2016-06-16. Archived from the original (<http://www.vsyn.ru>) on 2016-06-30. Retrieved 2016-10-03.
31. "Hardware design for all" (<https://www.synflow.com/>). *Synflow*. Retrieved 2016-10-03.

Further reading

- Jason Cong, Jason Lau, Gai Liu, Stephen Neuendorffer, Peichen Pan, Kees Vissers, Zhiru Zhang. *FPGA HLS Today: Successes, Challenges, and Opportunities*. ACM Transactions on Reconfigurable Technology and Systems, Volume 15, Issue 4, Article No. 5, pp 1–42, December 2022, <https://doi.org/10.1145/3530775>.
- Michael Fingeroff (2010). *High-Level Synthesis Blue Book*. Xlibris Corporation. ISBN 978-1-4500-9724-6.
- Coussy, P.; Gajski, D. D.; Meredith, M.; Takach, A. (2009). "An Introduction to High-Level Synthesis". *IEEE Design & Test of Computers*. **26** (4): 8–17. doi:[10.1109/MDT.2009.69](https://doi.org/10.1109/MDT.2009.69) (<https://doi.org/10.1109%2FMDT.2009.69>). S2CID 52870966 (<https://api.semanticscholar.org/CorpusID:52870966>).
- Ewout S. J. Martens; Georges Gielen (2008). *High-level modeling and synthesis of analog integrated systems*. Springer. ISBN 978-1-4020-6801-0.
- Saraju Mohanty; N. Ranganathan; E. Kougianos & P. Patra (2008). *Low-Power High-Level Synthesis for Nanoscale CMOS Circuits*. Springer. ISBN 978-0-387764733.
- Alice C. Parker; Yosef Tirat-Gefen; Suhrid A. Wadekar (2007). "System-Level Design". In Wai-Kai Chen (ed.). *The VLSI handbook* (2nd ed.). CRC Press. ISBN 978-0-8493-4199-1. chapter 76.
- Shahrzad Mirkhani; Zainalabedin Navabi (2007). "System Level Design Languages". In Wai-Kai Chen (ed.). *The VLSI handbook* (2nd ed.). CRC Press. ISBN 978-0-8493-4199-1. chapter 86. covers the use of C/C++, SystemC, TML and even UML
- Liming Xiu (2007). *VLSI circuit design methodology demystified: a conceptual taxonomy*. Wiley-IEEE. ISBN 978-0-470-12742-1.
- John P. Elliott (1999). *Understanding behavioral synthesis: a practical guide to high-level design*. Springer. ISBN 978-0-7923-8542-4.
- Nane, Razvan; Sima, Vlad-Mihai; Pilato, Christian; Choi, Jongsok; Fort, Blair; Canis, Andrew; Chen, Yu Ting; Hsiao, Hsuan; Brown, Stephen; Ferrandi, Fabrizio; Anderson, Jason; Bertels, Koen (2016). "A Survey and Evaluation of FPGA High-Level Synthesis Tools". *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. **35** (10): 1591–1604. doi:[10.1109/TCAD.2015.2513673](https://doi.org/10.1109/TCAD.2015.2513673) (<https://doi.org/10.1109%2FTCAD.2015.2513673>). hdl:[11311/998432](https://hdl.handle.net/11311/998432) (<https://hdl.handle.net/11311%2F998432>). S2CID 8749577 (<https://api.semanticscholar.org/CorpusID:8749577>).
- Gupta, Rajesh; Brewer, Forrest (2008). "High-Level Synthesis: A Retrospective". *High-level Synthesis: A Retrospective*. Springer. pp. 13–28. doi:[10.1007/978-1-4020-8588-8_2](https://doi.org/10.1007/978-1-4020-8588-8_2) (https://doi.org/10.1007%2F978-1-4020-8588-8_2). ISBN 978-1-4020-8587-1.

External links

- Vivado HLS course on Youtube (https://www.youtube.com/watch?v=kgae3Wzqngs&list=PLo7bVbJhQ6qzK6ELKCm8H_WEzzcr5YXHC)
- Deepchip Discussion Forum (<http://www.deepchip.com/posts/0480.html>)

