1

The Function of Computation

As in music theory, we cannot discuss the microprocessor without positioning it in the context of the history of the computer, since this component is the integrated version of the central unit. Its internal mechanisms are the same as those of supercomputers, mainframe computers and minicomputers. Thanks to advances in microelectronics, additional functionality has been integrated with each generation in order to speed up internal operations. A computer[1] is a hardware and software system responsible for the automatic processing of information, managed by a stored program. To accomplish this task, the computer's essential function is the transformation of data using computation, but two other functions are also essential. Namely, these are storing and transferring information (i.e. communication). In some industrial fields, control is a fourth function. This chapter focuses on the requirements that led to the invention of tools and calculating machines to arrive at the modern version of the computer that we know today. The technological aspect is then addressed. Some chronological references are given. Then several classification criteria are proposed. The analog computer, which is then described, was an alternative to the digital version. Finally, the relationship between hardware and software and the evolution of integration and its limits are addressed.

NOTE.– This chapter does not attempt to replace a historical study. It gives only a few key dates and technical benchmarks to understand the technological evolution of the field.

1 The French word *ordinateur* (computer) was suggested by Jacques Perret, professor at the Faculté des Lettres de Paris, in his letter dated April 16, 1955, in response to a question from IBM to name these machines; the English name was the Electronic Data Processing Machine.

1.1. Beginnings

Humans have needed to count since our earliest days (Ifrah 1994; Goldstein 1999). Fingers were undoubtedly used as the first natural counting tool, which later led to the use of the decimal number base. During archeological excavations, we have also found notched counting sticks, bones and pieces of wood. The incised bones of Ishango, dated between 23,000 and 25,000 years BC, provide an example (Figure 1.1).

Figure 1.1. *Ishango's incised bones (source: unknown). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

Counting sticks were used during antiquity, as well as pebbles, hence the word calculus, from the Latin *calculus* which means "small pebble". Knotted ropes were also used for counting, an example being the Incan *quipu* (Figure 1.2). This Incan technique (dating ≈ 1200–1570) used a positional numbering system (*cf.* § 1.2 of Darche (2000)) in base-10 (Ascher 1983).

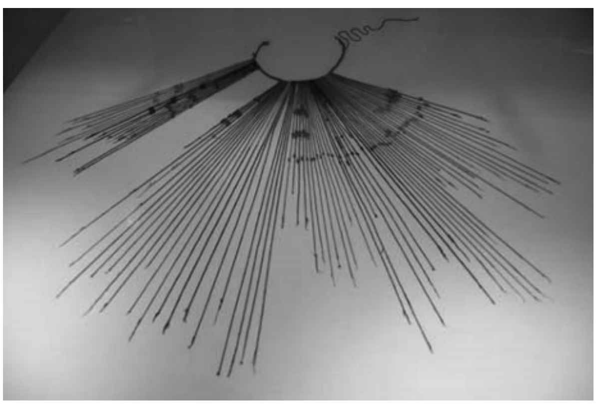
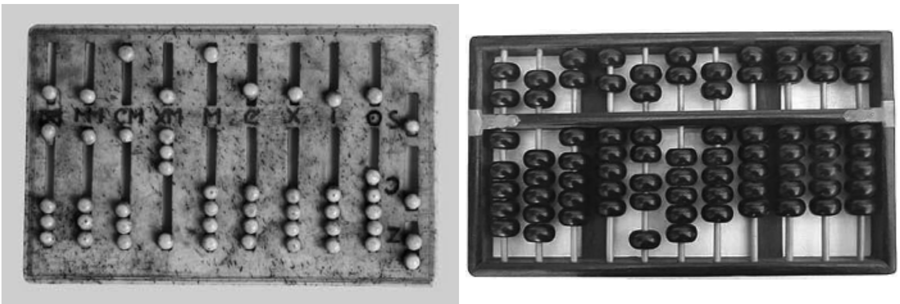


Figure 1.2. *A quipu (source: unknown). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

The need for fast and precise computation necessitated the use of computing instruments. Two exemplars are the abacus and the slide rule. The abacus is a planar calculating instrument, with examples including the Roman (Figure 1.3(a)) and the Chinese (Figure 1.3(b)) abacus. The latter makes it possible to calculate the four basic arithmetic operations by the movements of beads (or balls) strung on rods, which represent numbers.



a) **b)**

Figure 1.3. *Roman abacus (a) between the 2nd and 5th Centuries (© Inria/AMISA/Photo J.-M. Ramès); Chinese abacus (b). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

The 17th Century saw the introduction of mechanical computing machines, and the beginning of the history of computers is generally dated from their appearance. They met the need to systematically calculate tables of numbers reliably and quickly. These machines naturally used the decimal base. The most famous is undoubtedly the adding machine called the Pascaline (1642), named after its inventor, the philosopher and scientist Blaise Pascal (1623–1662). Numbers were entered using numbered wheels (Figure 1.4). The result was visible through the upper slits. Complementation made it possible to carry out subtraction (*cf.* exercise E1.1). But the first description of a four-operation machine was Wilhelm Chickard's machine (1592–1635), which appeared in a letter from the inventor to Johannes Kepler in 1623 (Aspray 1990). The end of the 17th Century and the following one were fruitful in terms of adding machines. Consider, for example, machines by Morland (1666), Perrault (1675), Grillet (1678), Poleni (1709), de Lépine (1725), Leupold (1727), Pereire (1750), Hahn (1770), Mahon (1777) and Müller (1784). A logical continuation of this trend was the multiplying machine by Gottfried Wilhelm Leibniz (1646–1716), which was designed in 1673 but whose implementation was delayed because of the lack of mechanical manufacturing precision in the 17th Century. For more information on this technology, we can cite the richly illustrated book by Marguin (1994) introducing the first mechanical calculating machines.

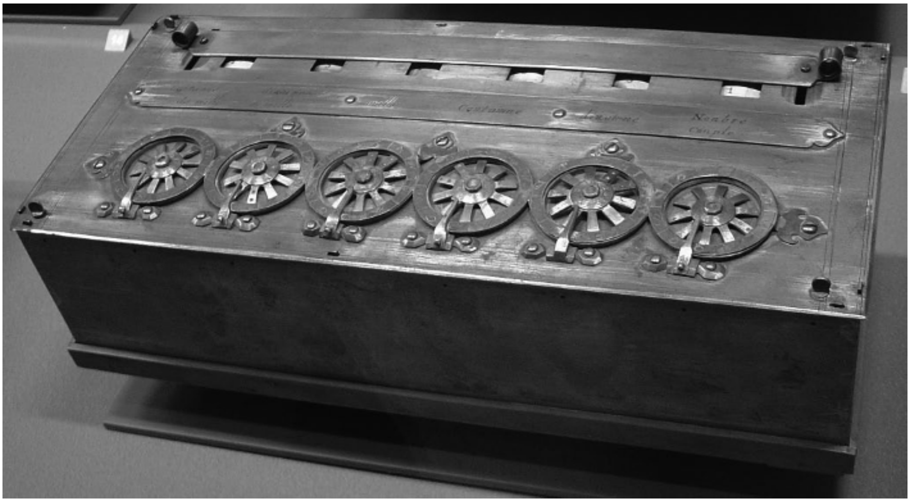


Figure 1.4. *An example of a Pascaline at the Musée des Arts et Métiers (source: David Monniaux/Wikipedia[2]). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

The mathematician Charles Babbage (1791–1871) marked the 19th Century *a posteriori* with two machines: the Difference Engine and the Analytical Engine.

The first machine was intended for the automatic computation of polynomial functions with printed results in order to build trigonometric and logarithm tables for the preparation of astronomical tables useful for navigation. At the time, logarithm tables were expensive, cumbersome and often out of print (Campbell-Kelly 1987, 1988, Swade 2001). They were calculated by hand, a tedious method that was the source of many errors. We can cite as an example those of De Prony (1825) for assessment, which was studied among others by Grattan-Guinness (1990), of which Babbage was aware. This machine reportedly allowed the successive values of a polynomial function to be calculated by Newton's finite difference method (see, for example, Bromley (1987) and Swade (1993)). Figure 1.5 presents a prototype, with all the details of this construction given in Swade (2005). It was never produced during his lifetime because of the enormous cost of manufacturing the mechanics. It was not until May 1991 that the second model, called the "difference machine no. 2", was implemented at the London Science Museum where it was also exhibited (Swade 1993).

2 URL: http://upload.wikimedia.org/wikipedia/commons/8/80/Arts_et_Metiers_Pascaline_dsc03869.jpg.

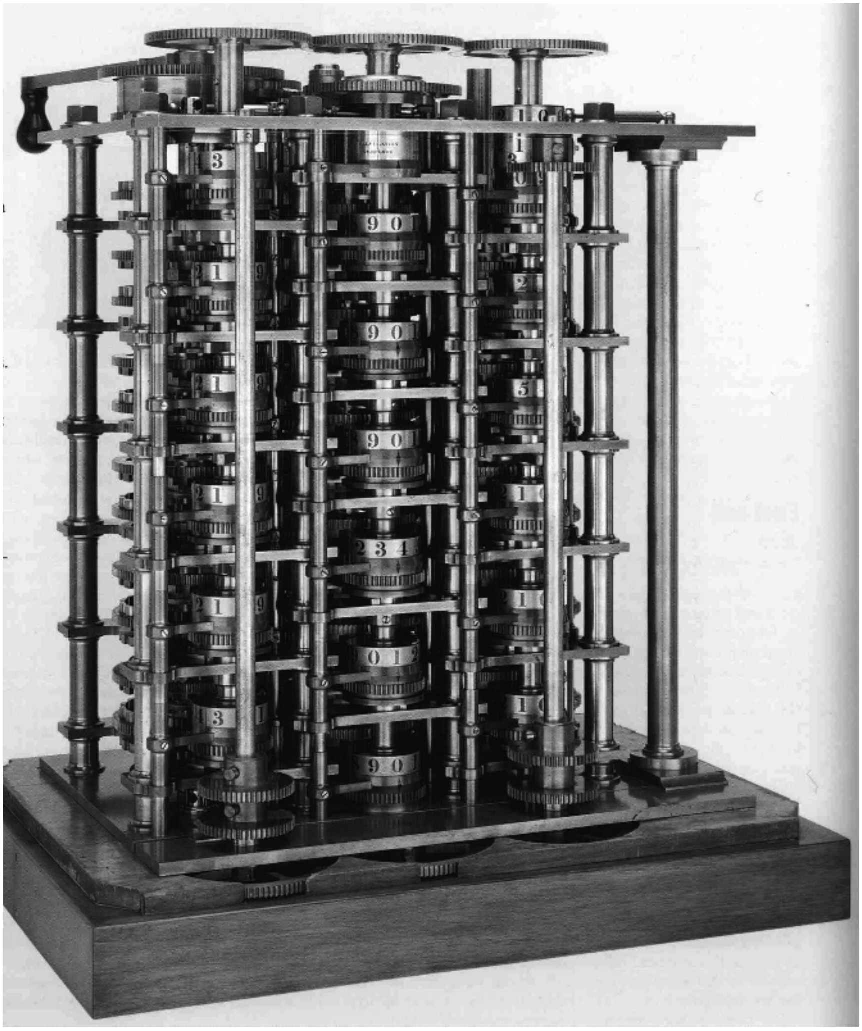


Figure 1.5. *Replica of the first Babbage difference machine*[3]*. For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

The second machine (Figure 1.6) could compute the four basic arithmetic operations.

3 URL: http://iamyouasheisme.files.wordpress.com/2011/11/babbagedifferenceengine.jpg.

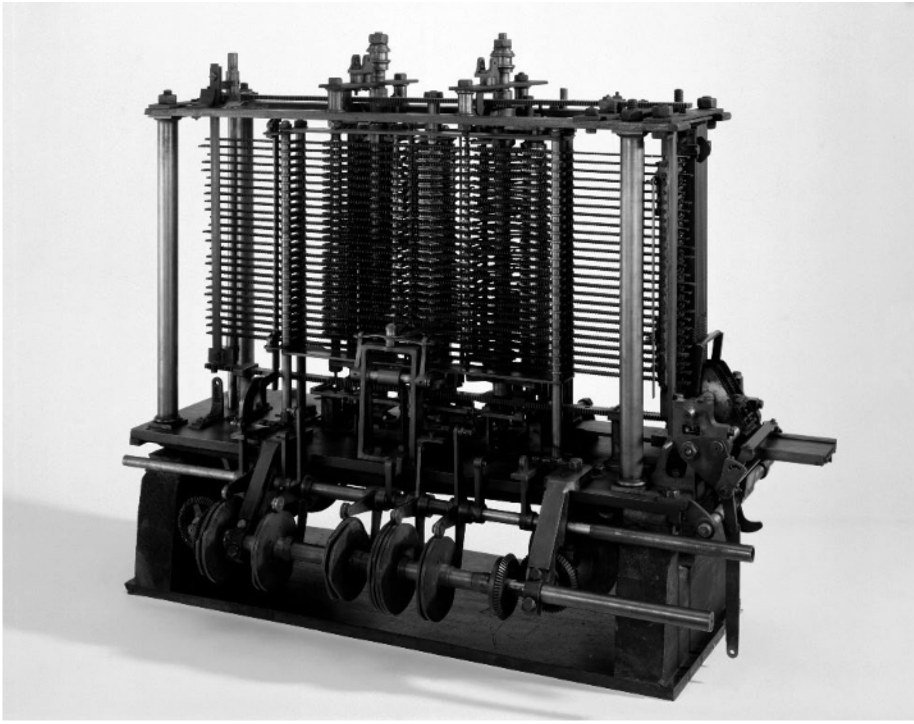


Figure 1.6. *Babbage's analytical machine (© Science Museum/Science & Society Picture Library). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

It introduces the basic architecture of a computer and its programming (Hartree 1948). Indeed, as illustrated in Figure 1.7, it was composed of a mill, which played the role of the modern Central Processing Unit (CPU), and a store, which played the role of main storage. It also implemented the notion of registers (major axes) and data buses (transfer paths). Integers were internally represented in base-10 using Sign-Magnitude or Sign and Magnitude Representation (SMR, *cf.* § 5.2 in Darche (2000)) in base 10. Extensive details of its operation are given in Bromley (1982). For the same technological and financial reasons previously mentioned, its construction has never been completed.

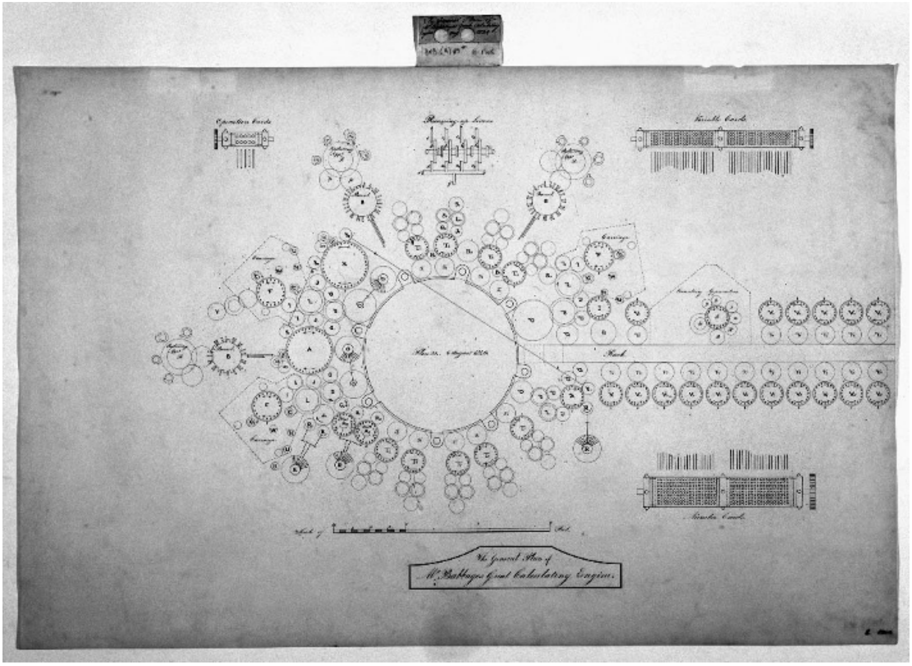


Figure 1.7. *One of the plans for Babbage's analytical machine (© Science Museum/Science & Society Picture Library). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

To program the machine, Babbage proposed the punched card. The latter had been invented by Basile Bouchon in 1725 for the weaving industry in the form of a strip of perforated paper. Jean-Baptiste Falcon improved it by transforming this strip into a string of punched cards linked together by cords. These cards made it possible to store a weaving pattern (Figure 1.8). This principle was further improved and made truly usable by Joseph Marie Jacquard with his famous loom (*cf.* Cass (2005) for a notice by J. M. Jacquard from 1809). Essinger (2004) tells the history of this machine. The latter was not the only programmable machine of the time. The music box with pegged cylinder was another form. In Babbage's machine, program instructions and data were entered separately using two decks of cards. Babbage had a collaborator, Ada Lovelace, who is considered the first programmer in history to have written a Bernoulli number algorithm for this machine (reproduced in Kim and Toole (1999)). However, we should not conclude that Charles Babbage is the source of the modern computer because of the influence of his ideas on the design of modern computers (Metropolis and Worlton 1980).

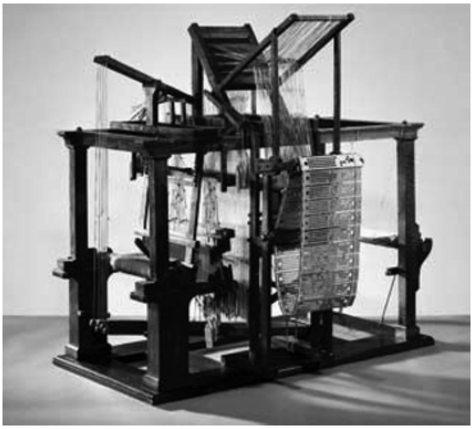


Figure 1.8. *Falcon's loom. For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

The history of the modern computer can also be traced back to the 1880s with the invention of mechanography for the United States Census Bureau (Ceruzzi 2013). Hermann Hollerith took up the idea of the punched card and mechanized data processing to calculate statistics (Hollerith 1884a,b 1887). Figure 1.9 shows his statistics machine, composed of a hole punch called a press, with a tabulator that read and counted using electromechanical counters, and a sorter called a sorting box.

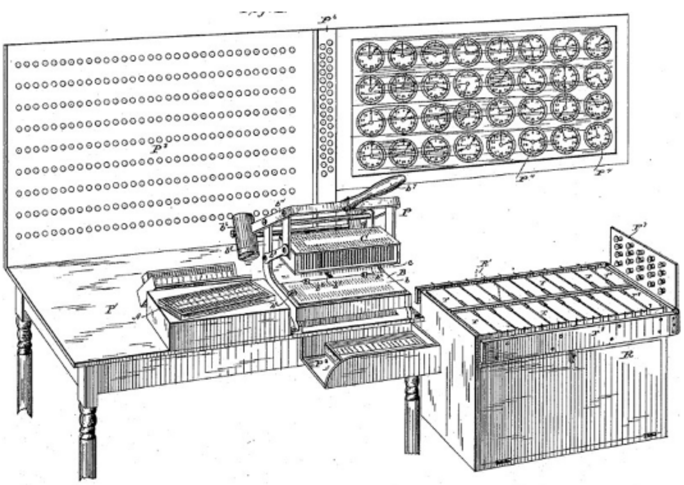


Figure 1.9. *Statistical machine (Hollerith 1887)*

As previously described and illustrated in Figure 1.10, the computer in its current form is the result of technological progress and advances in the mathematical fields, particularly in logic and arithmetic. Boole's algebra offered a theoretical framework for the study of logic circuits (*cf.* § 1.3 of Darche (2002)). For example, the American researcher Claude Elwood Shannon illustrated the relationship between Boolean logic and switch and relay circuits in his famous article (Shannon 1938). Thus, a link was established between mathematical theory and manufacturing technology. A study by Shannon (1953) described the operation of 16 Boolean functions in two variables using 18 contacts, and was able to show that this number of contacts was minimal. The mathematical aspect of switching has been studied in particular by Hohn (1955). Technology played a major role because it had a direct impact on the feasibility of the implementation, the speed of computation, and the cost of the machine.

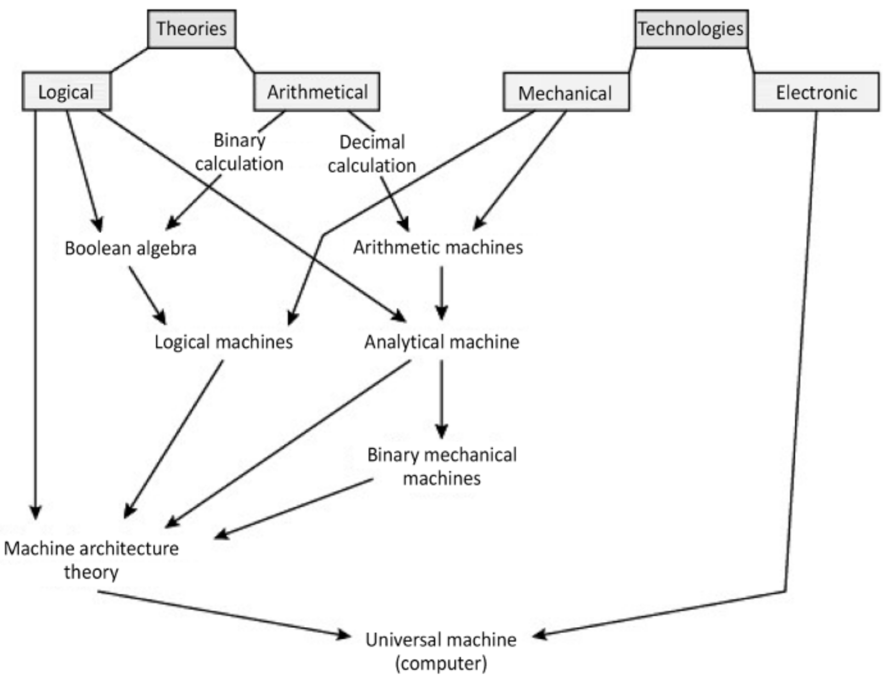


Figure 1.10. *Evolution of concepts and technologies in the development of calculating machines (from Marguin (1994))*

1.2. Classes of computers

There are several possible ways to classify computers. One is primarily related to the hardware technology available at the time, as presented in Tanenbaum (2005). For this reason, we will speak of technological generations. The transition from one generation to the next is achieved by a change in technology or by a major advance. Table 1.1 presents these generations in a simplified manner.[4]

Technological generations	Dates
0 – mechanical	1642–1936
1 – electromechanical	1937–1945
2 – tube	1946–1955
3 – transistor	1956–1965
4 – integrated circuits SSI – MSI – LSI	1966–1980
5 – integrated circuit VLSI	1981–1999
6 – integrated circuit GSI – SoC – MEMS	2000 to present

Table 1.1. *Generations of calculating machines and computers based on component technologies*

Generation 0 (1642–1936) consisted of mechanical computers, as presented in the previous section. Mechanography appeared at the end of the 19th Century to respond in particular to the need for automatic processing of statistical data, initially for the census of the American population. Its technology naturally evolved towards electromechanics. A historical examination of mechanography in relation to "modern" computing was conducted by Rochain (2016).

Generation 1 was that of the electromechanical computer (1937–1945). The basic component was the electromechanical relay (Figure 1.11(a)) comprised of a coil that moves one or more electrical contacts on command (i.e. if it is electrically powered). Figure 1.11(b) presents its equivalent electric diagram. Keller (1962) describes the technology of the time. The implementation of a logical operator in this technology was described in § 2.1.2 of Darche (2004). In 1937, George Stibitz, a mathematician from Bell Labs, built the first binary circuit, an adding machine, the Model K (K for Kitchen) in electromechanical technology (Figure 1.11(c)). One of the pioneers of this generation in Europe was the German Konrad Zuse. His first

4 The dates provided are for illustrative purposes only because the transition from one generation to the next is obviously gradual.

machine, the Z1, begun in 1936 and completed two years later, was a mechanical computer powered by an electric motor. The first electromechanical relay computer, the Z2, was completed in 1939. It was built using surplus telephone relays. The Z3 (storage of 1,800[5] relays, 600 for the computing unit and 200 for the control unit, according to Ceruzzi (2003)), whose construction began 1938 and ended in 1941, used base-2 floating-point number representation. The Z4, started in 1942, was completed in 1945. Rojas (1997) describes the architecture of the Z1 and the Z3, and Speiser (1980), that of the Z4. In the United States, Harvard's Mark I, also called Automatic Sequence Controlled Calculator (ASCC) by IBM, was built by Howard Aiken between 1939 and 1944. Bell Laboratories built six models of computers using this technology between 1939 and 1950 for military and scientific use (Andrews 1982). Andrews and Bode (1950) describe the use of the Model V from Bell Laboratories. The calculation speed of these computers is estimated at 10 operations/s.

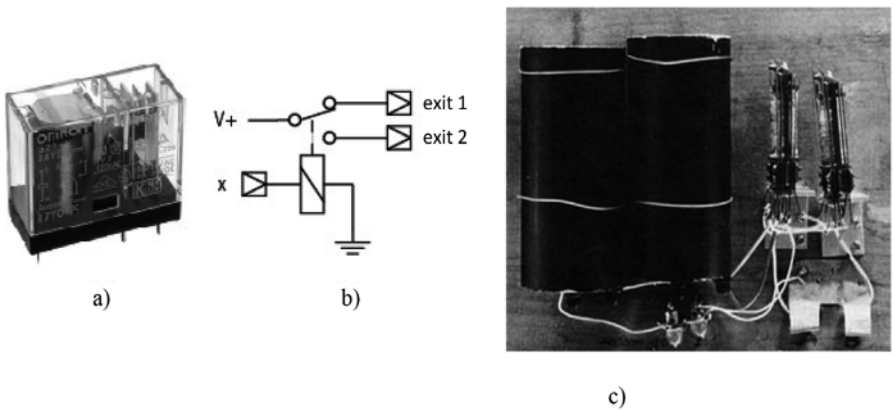


Figure 1.11. *A modern electromechanical relay, its equivalent electrical diagram, and the Model K adder. For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

The subsequent generations used electronic components, beginning in the 1946–1955 period with the electronic tube, also known as the vacuum tube (thermionic valve). This component has rectification, amplification, and switching functions. It was the latter that was exploited in this case. As shown in Figure 1.12(a), a tube is a bulb, made of glass in this implementation, that is sealed under vacuum or filled with an inert gas. Inside are electrodes: the cathode, the grid(s) and the anode. Electrons migrate from the cathode to the anode via thermionic effect, subsequently passing through one (triode) or more grids (tetrode, pentode and higher), which

5 1400 in Zuse (1993) and Weiss (1996).

modulates the flow. Figure 1.12(b) illustrates the resulting dimensions of a circuit board in this technology.

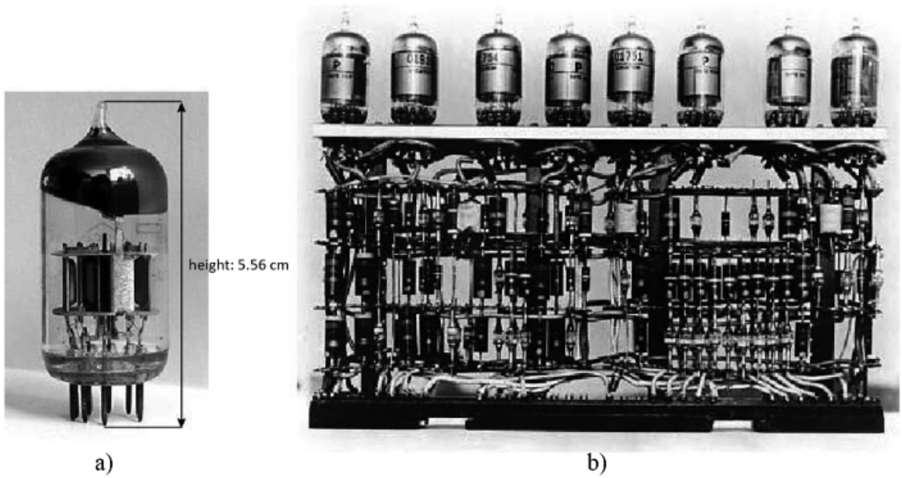


Figure 1.12. *An RCA 5965 type electronic tube and an IBM 701 electronic board (source: IBM). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

Table 1.2 shows the main computers of generations 1 and 2.

(Beginning of the project –) operational computer	Name	Designers	Country	Key features
1936–1938	Z1	Zuse	Germany	Mechanical computer driven by electric motor
1939–1942	ABC	Atanasoff/Berry	United States	First electronic calculator (non-programmable)
1943	Colossus	Thomas Harold Flowers	Great Britain	First electronic computer to use a stored program
1939–1944	Harvard Mark I	Howard Aiken	United States	Electromagnetic computer based on Harvard architecture (*cf.* § 3.4.2)
1942–1945	Z4	Zuse	Germany	Electromechanical computer

1943–1946	ENIAC	Eckert/Mauchly (Moore School of Electrical Engineering University of Pennsylvania)	United States	Second electronic computer (reprogrammable via wiring)
1946–1952	EDVAC	Eckert/Mauchly/von Neumann	United States	Electronic computer based on von Neumann architecture
1948	Manchester Baby[6]	Williams/Kilburn	Great Britain	First electronic computer to use a stored program
1949	Manchester Mark I[7]	Williams/Kilburn	GB	Second electronic computer to use a stored program
1947–1949	BINAC	Eckert/Mauchly	USA	First commercially available electronic computer
1946–1949	EDSAC	Wilkes	GB	Electronic computer implementation of von Neumann architecture
1951	Whirlwind I	MIT	USA	Electronic computer
1951	UNIVAC I	Eckert/Mauchly	USA	Commercially available computer
1945–1952	IAS machine	John von Neumann (Princeton)	USA	Implementation of von Neumann architecture
1952	701	IBM	USA	First commercially available scientific computer from this company
1954	704	IBM	USA	Scientific computer with floating-point operations
1957	709	IBM	USA	Improved version of 704

Table 1.2. *Reference computers for generations 1 and 2*

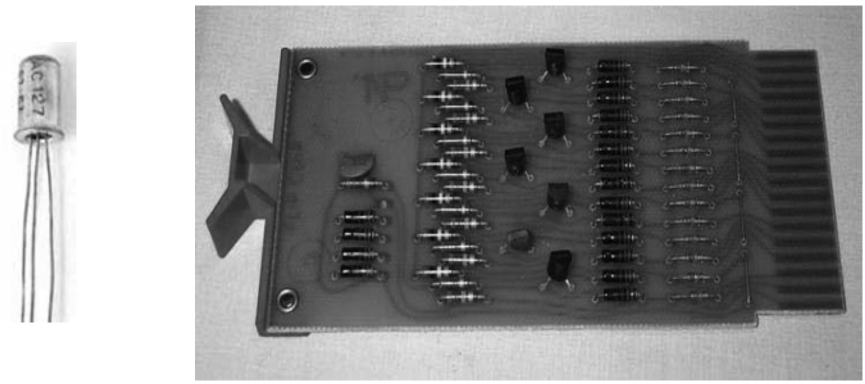
Generation 3 (1956–1965) saw the emergence of electronic computers with diodes and discrete transistors.[8] These two components have the same function –

6 The Manchester Baby was the nickname given to the *Manchester Small Scale Experimental Computer*.

7 The official name of the Manchester Mark I was the *Manchester Automatic Digital Computer or Machine* (MADC or MADM). According to Reilly (2003), the correct date should be June 21, 1948, the date on which the computer became operational.

8 As opposed to “integrated”.

rectification and amplification respectively – as the electronic tube, but at a much lower size, supply voltage, consumption of current and cost by a factor 10, 20 and 30 (orders of magnitude) respectively for the first three criteria. In addition, reliability and switching speed both increased. The transistor (Figure 1.13(a)), a contraction of the words "transfer resistor", was invented in 1948 (Bardeen and Brattain 1950). Its history is retraced in Brinkman (1997), and *Scientific American* (1997); IEEE (1998). A Bipolar Junction Transistor (BJT) (Figure 1.13(a)) is a sandwich of three layers of doped semiconductor materials (germanium or silicon) of type NPN or PNP. It behaves like a triode (Bardeen and Brattain 1948) with three electrodes: the emitter, the base and the collector. It behaves as an amplifier or a switch, a function used in digital logic. Figure 1.13(b) gives an example of an implementation of logic gates, in this case, seven inverters (*cf.* § 2.1.3 of (Darche 2004)).



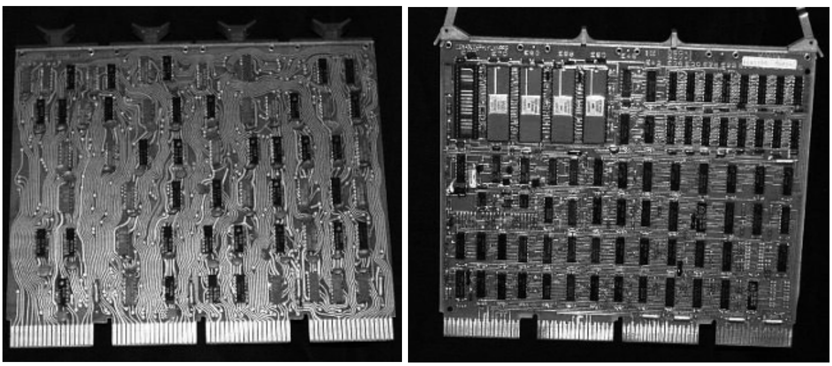
a) **b)**

Figure 1.13. *A transistor and an electronic transistor board with seven inverters from a DEC PDP-8 (source: http://www.pdp8.net). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

Table 1.3 presents the main computers of this third generation. The supremacy of the United States is notable.

Year	Name	Designer/manufacturer	Country	Key features
1960	PDP-1	DEC	USA	First minicomputer
1961	1401	IBM	USA	First electronic computer with stored program
1959	7090	IBM	USA	Transistor implementation of model 709
1963	B5000	Burroughs	USA	Battery architecture (*cf.* § 3.4.1)
1964	CDC 6600	Control Data Corporation	USA	First parallel computer
1965	PDP-8	DEC	USA	The company's iconic minicomputer

Table 1.3. *The main computers from this generation*



a) **b)**

Figure 1.14. *One of the 15 DIP integrated circuit CPU boards from a DEC PDP-11/20 and an electronic board from an LSI-11 (PDP-11/03 – 1975) (source: https://sydney.edu.au/science/psychology/pdp-11/Images/Images.html). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

The introduction of the integrated circuit marks the beginning of generation 4 (1966–1980), which saw the introduction of centralized architectures and the microprocessor. The first generations of components were called SSI (Small Scale Integration) and MSI (Medium-Scale Integration). These emerged during the period of 1966–1975. They are integrated circuits in a DIP (Dual-In-line Package) as shown on the CPU (Central Processing Unit) board in Figure 1.14. A central unit was made up of dozens of boards linked together by a backplane bus (*cf.* § V2-1.2 and V2-4.8) via edge connector Printed Circuit Board (PCB) gold fingers (bottom of the printed circuit in the photo). The encapsulation of integrated circuits was presented in § 3.3 of Darche (2004). From the 1970s, the LSI (Large-Scale Integration) generation enabled the appearance of the microprocessor (*cf.* § V3-1.1) manufactured using MOS (Metal-Oxide Semiconductor), PMOS (Positive (channel) MOS), and NMOS (Negative (channel) MOS), and then CMOS (Complementary[9] MOS), which was used in computing for the first time only in the microcomputer. The microelectronic technology used in computers was essentially bipolar (1965–1985 period) for the sake of operating frequency. The most widespread were the "standardized" families of TTL (Transistor–Transistor Logic) and ECL (Emitter Coupled Logic, *cf.* § 2.3.3 of Darche (2004)). These families were previously introduced respectively in § 2.3.2 and 2.3.3 of Darche (2004). Proprietary hybrid integrated circuit technologies coexisted, such as SLT (Solid Logic Technology) (Davis *et al.* 1964), in IBM's System/360 family. Passive and active components (Diode–Transistor Logic – DTL) were then assembled on a ceramic substrate and encapsulated.

Table 1.4 presents some models of microcomputers, minicomputers, mainframe computers and supercomputers.

Year	Name	Designer/manufacturer	Country	Key features
1964	System/360	IBM	USA	ISA concept (*cf.* § 3.5)
1970	PDP-11	DEC	USA	Iconic minicomputer
1973	Micral N	R2E	France	First computer with PMOS technology
1975	CRAY-1	Cray	USA	First super-computer
1978	VAX-11	DEC	USA	Successor to the PDP family

Table 1.4. *Primary computers in this generation*

As shown in Figure 1.15, manufacturing technology has evolved over time. Initially bipolar, it slowly evolved towards unipolar technologies – MOS, and then today, dominantly, CMOS. Two crossovers should be noted, in 1985, when CMOS technology achieved the performance level of TTL, and in 1991, when CMOS achieved results equivalent to ECL technology (Emitter Coupled Logic). This meant that unipolar technology eventually overtook these two bipolar technologies in terms

9 Technology brought together MOSFET transistors (MOS Field Effect Transistor) from the two aforementioned technologies, that is, canal p and canal n respectively, from whence this adjective stemmed.

of performance. ECL nevertheless continued to be used in the supercomputer industry until 1990 thanks in particular to its functionality as a line amplifier (transmission line driver). These technological advances had an impact on the number of supply voltages and their values, as well as on current consumption. This subject is dealt with in § V3-6.1.2.

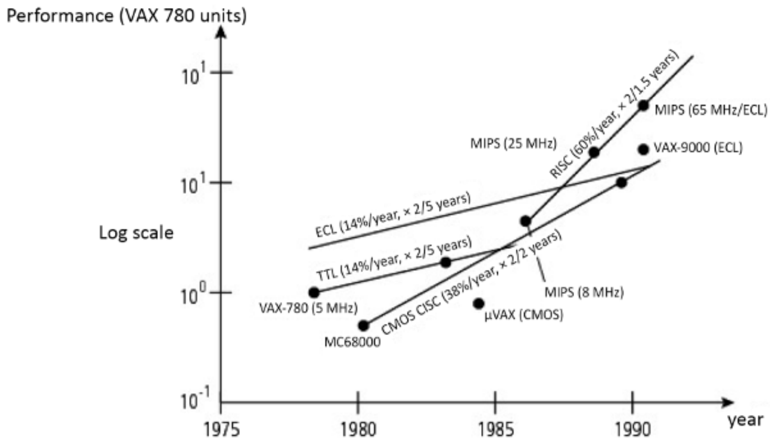


Figure 1.15. *Evolution of computing performance over time (from (Bell 2008b))*

Generation 5[10] saw the emergence of electronic computers with integrated VLSI (Very LSI) circuits in the 1980s. Seraphim and Feinberg (1981) introduce IBM's computer encapsulation technologies that were current as of the date of the article and provide an overview of the evolution. The microcomputer that still serves as a reference today, the IBM PC (Personal Computer) from the IBM Corporation, was released in 1981 (Figure 1.16).

The 21 Century has taken us to the next generation with ubiquitous or pervasive computing and integrated parallel systems. This is the era of SoC (System on (a) Chip), which is a complete, integrated computer including several so-called core processors, as well as Input/Output (I/O) controllers and RAM (Random Access Memory). This is a result of Moore's law (*cf.* § 1.5). The SoC's predecessor is the Application-Specific Integrated Circuit (ASIC), usually developed for a specific

10 An orthogonal use of the same phrase was used in Japan in 1981 when it launched the national "fifth generation" project (Moto-oka *et al.* 1982; Treleaven and Lima 1982). The associated software development used declarative languages referred to as the fifth generation. In addition, Treleaven (1981) spoke of the fifth generation to refer to spatially distributed systems carrying out decentralized computing.

client with various design styles (full custom, semicustom and programmable). These began to emerge in the early 1980s. In addition, MEMS (MicroElectroMechanical System) or electromechanical microsystems, increasingly integrated peripherals such as sensors, or even actuators, such as a micro-pump.

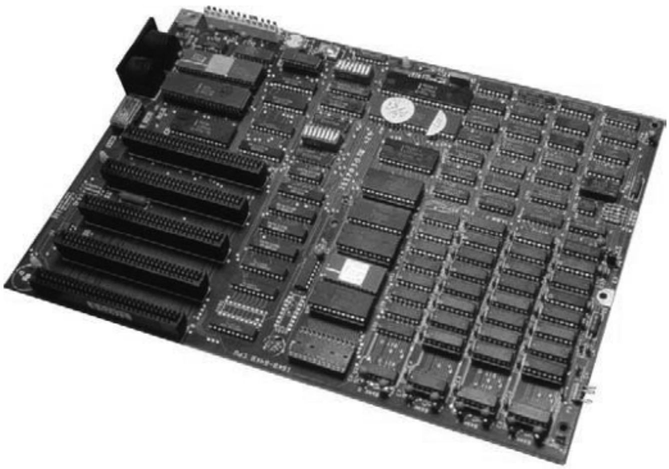


Figure 1.16. *PC motherboard (5150) from IBM (1981). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

Tables 1.5(a–c) summarize the different generations of integrated circuits by showing, for each of them, the number of transistors n and equivalent gates p per package, according to different authors. The equivalent gate (NAND with two inputs)[11] is independent of technology and logic type. The number of gate-equivalent circuits is a unit of measurement of the complexity of a circuit that indicates the number of gates necessary to perform a given function. There exists a correspondence between a given technology and the number of transistors per gate. Note the introduction of two generations following VLSI, called ULSI (Ultra LSI) and GSI (GigaScale Integration), the latter of which allowed for the integration of an entire system or SoC (System-on-(a)-Chip). We should also anecdotally mention the intermediate generations ELSI (Extra LSI) and SLSI (Super LSI). The range of the intervals varies from author to author. Each generation has made it possible to integrate increasingly complex functionalities. The SSI generation brought integrated gates; the MSI, simple combinational and sequential functions (for encoding, multiplying, adding, storing, counting, etc.) and the LSI, which was an entire system (Arithmetic and Logic Unit (ALU), 4–8-bit microprocessor, I/O controller, memory size < 256 Kib, etc.). The VLSI and the GSI represent the generations of 16–32-bit and 64-bit microprocessors.

11 Remember that NAND (Not AND) is a complete operator (*cf.* § 1.5.7 in Darche 2002).

Today, the purpose of this categorization is to show the hierarchy of the ideas behind these logical systems and the acceleration of the density of integration.

Generations	Year of introduction	Primary electronic technology	Number of logic gates p per package (Osborne 1980; Weste and Harris 2010)
SSI	1964	Bipolar	1–10
MSI	1968	Bipolar	10–1000 100–1000 (Osborne 80)
LSI	1971	PMOS NMOS	10^3–10^4
ELSI	–	–	–
VLSI	1980	HMOS CMOS	10^4–10^5
SLSI	–	CMOS	–
ULSI	1990	CMOS	10^5–10^6
GSI SoC	2000	CMOS	–

Table 1.5a. *Classification of generations of integrated circuits according to various authors*

Generations	Number of logic gates p equivalent per package (van de Goor 1989) (Kaeslin 2008) and TI	Number of transistors n/gates p (Lilen 1979)
SSI	1–10	< 100/1–25
MSI	10–100	$100 \leq n < 1000$ $25 \leq p < 250$
LSI	10^2–10^4	$1000 \leq n < 10000$ $250 \leq p < 2500$
ELSI	–	–
VLSI	10^4–10^6	≥ 10000 $p \geq 2500$
SLSI	–	–
ULSI	$\geq 10^6$	–
GSI SoC	–	–

Table 1.5b. *Classification of generations of integrated circuits according to various authors (continued)*

Generations	Number of transistors n/gates p (Wickes 1968)	Number of logic gates p (Siewiorek *et al.* 1982)
SSI	$n < 10$ $p < 12$	1–9
MSI	$10 \leq n < 100$ $12 \leq p < 100$	10–99
LSI	$100 \leq n < 1000$ $p \geq 100$	100–9999
ELSI	$100 \leq p < 999$	–
VLSI	$1000 \leq n < 10\,000$ $1000 \leq p < 999\,999$	10^4–99 999
SLSI	$10\,000 \leq n < 100\,000$	–
ULSI	$n \geq 10^6$ $p \geq 10^6$	$\geq 10^5$
GSI SoC	–	–

Table 1.5c. *Classification of generations of integrated circuits according to various authors (continuation and end)*

The point of this functional decomposition was to “standardize” the electronic components, thus allowing for a reduction in costs and a simplification of design. We are referring to off-the-shelf components (COTS for Commercial Off-The-Shelf). The first such components were digital electronics with simple combinational and sequential logic (gates, latches, and flip-flops), followed by more complex ones (decoders, registers, etc., *cf.* Darche (2002, 2004)). Next came the microprocessor (*cf.* V3) and bit slicing (*cf.* § V3-5.1), which were the next examples in the field.

The following table specifies the definitions used for this series of works.

Optical technology is an alternative to current (i.e. electronic) technology for obtaining a high transmission rate, lower attenuation and resistance to corrosion. It is already used in the telecommunications field in optical fiber and mass storage. Optoelectronics can be used in the interconnection of display systems and peripherals. All-optical logic gate operators exist in laboratories, aiming to achieve a higher computing speed.

Generations	Year of introduction	Primary electronic technology	Number of transistors n per package
SSI	1964	Bipolar	< 10
MSI	1968	Bipolar	$10 \leq n < 1000$
LSI	1971	PMOS NMOS	$1000 \leq n < 10\,000$
ELSI	–	–	–
VLSI	1980	HMOS[12] CMOS	$10\,000 \leq n < 100\,000$
SLSI	–	CMOS	–
ULSI	1990	CMOS	10^5 à 10^9 (Meindl 1984)
GSI SoC	2000	CMOS	$n \geq 10^9$ (Meindl 1995) $p \geq 10^6$

Table 1.6. *Classification of generations of integrated circuits adopted*

The quantum computer will undoubtedly be, if technology allows, a giant step forward from the point of view of computing performance (*cf.* § V4-3.4). Information is presented in the form of a qubit (quantum bit), which is a superposition of two basic states $|0\rangle$ and $|1\rangle$. This superposition, in association with the property of entanglement, opens the way to massive parallelization of computation. Indeed, it is possible to access all possible results in a single computation.

Bell (2008a, 2008b) defines the concept of computer class as being a set of computers with similar price, size, hardware and software technologies, computing power and field(s) of application. A hardware and software industry is associated with a class. The class determines the domain of use. The life cycle of a class, that is, the process of creation, evolution, and disappearance, evolves along four axes of cost evolution as illustrated in Figure 1.17. The class of supercomputers outstrips the others in the race for performance. There is a class at constant cost whose performance increases thanks to technological progress. There is a low-cost class. A class generates a less efficient and less expensive subclass (order of magnitude: factor of 10). A new class can supersede a previous one, as the PC did with the workstation, or it can incorporate it. The emergence phase lasts for about 10 years, triggered by new hardware technologies that enable advances in processors, buses,

12 HMOS: High-density MOS.

storage and I/O interfaces (in particular display and communications) and new software technologies (programming environment, Operating System (OS), Human–Machine Interface (HMI), etc.).

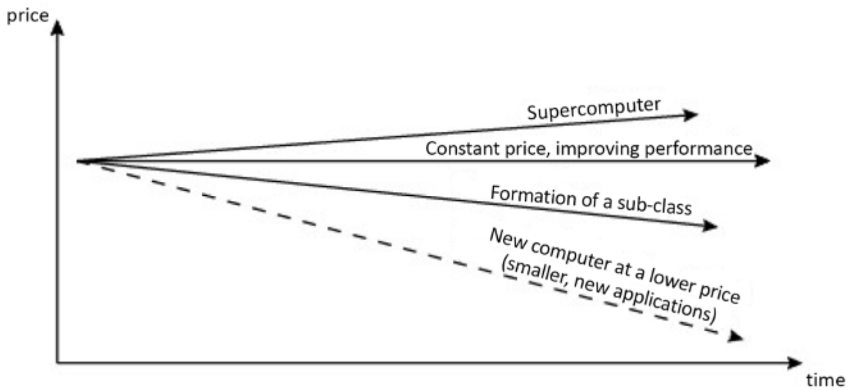


Figure 1.17. *Axes of evolution over time of the price of classes (from (Bell 2008a b) modified)*

The six main classes of computers before 2000 were, in descending order of performance, the scientific computer, the mainframe computer, the minicomputer, the workstation, the microcomputer or personal computer and the embedded system.

Scientists require intensive and time-consuming computation involving a large amount of data. This is the field of High-Performance Computing (HPC), and the associated computers are called supercomputers. This type of machine by definition has the highest computing speed, more primary and secondary storage than a minicomputer (order of magnitude: factor of 10^3 or 2^{10}). They originally had a single, ultra-fast processor. They then evolved by implementing parallelism. The representative company in this field from the 1980s is Cray Corporation with the Cray-1 (Figure 1.18), which operated in n = 64-bit format. In 1984, three classes of architecture – pipelined, vector (array processor), and multiprocessor – could be distinguished. The top countries in terms of computing performance were China and the United States (2016 data), with computers having a floating-point computing power (*cf.* § 4.2 of Darche (2000)) greater than 15,000 PFLOPS (petaFLOPS = 10^{15} Floating-Point Operations Per Second). The TOP500 project site (URL: http://www.top500.org) ranks their performance. The subclass is the mini-supercomputer (Architecture Technology Corporation 1991; Besk *et al.* 1993) like the "Crayette", a CMOS implementation of the original. This group also includes midrange systems. This type of machine, which was air cooled, consumed much less power than a supercomputer. It had vector computing capabilities and could be a multiprocessor.

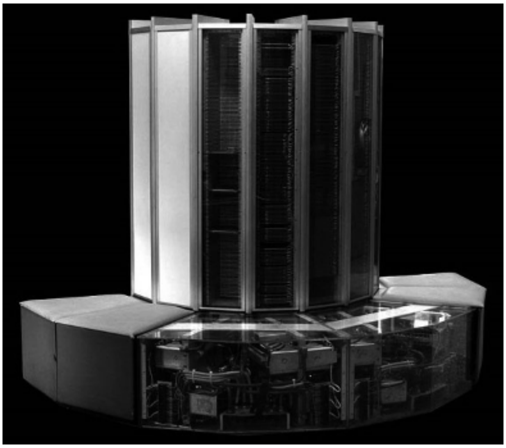


Figure 1.18. *The iconic Cray-1 supercomputer referred to as the "World's most expensive loveseat?"[13] (Computer World 1976). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

Figure 1.19 shows the performance curve for some reference machines.

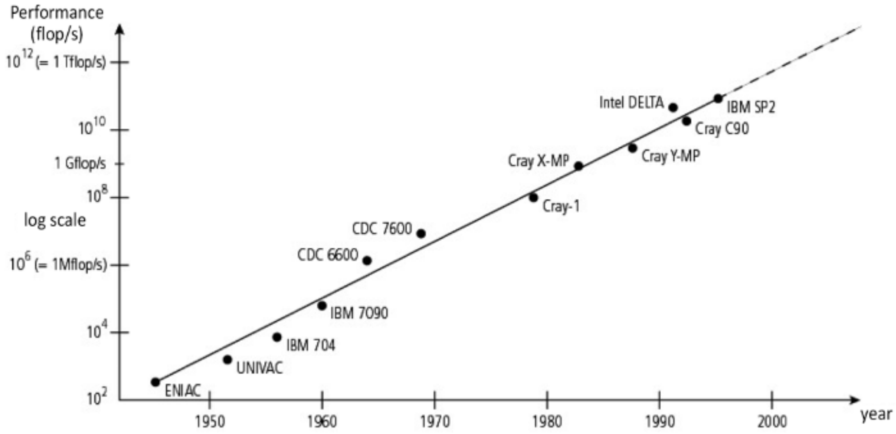


Figure 1.19. *Evolution over time of supercomputer performance (according to Succi et al. 1996)*

13 The title of the photograph in the associated article.

The term “mainframe” describes the cabinet containing the central unit and primary storage. This type of computer is characterized by sufficient computing power to support communications with hundreds of terminals and the execution of associated applications. The “mainframe” is also referred to as a “central system or central computer”. It is a transaction-based system that is associated with concepts such as batch processing, Simultaneous Peripheral Operations On-Line (SPOOL), cache and devices like the hard disk Mass Storage Device (MSD). The company that is most representative of the first category is undoubtedly IBM, with the IBM System/360 (Figure 1.20) described in Pugh (2013) and the IBM System/370. In the 1990s, the IBM 3090 was a representative central computer. This type of computer could be cooled by a heat transfer fluid other than air, such as water (see Ellsworth *et al*. (2008) on this subject).

Figure 1.20. *IBM System/360 mainframe computer*

The minicomputer, abbreviated mini, is a lower-level class. It is therefore sometimes called a departmental computer[14] as opposed to the centralized view of the mainframe. A minicomputer is a central computer of reduced size and power.

14 That is, for use by a group of people. But this name is not discriminative, because a parallel computer like the cydra 5 (Rau *et al*. 1989) is referred to as a departmental machine.

However, it is based on a similar philosophy and is more powerful than a microcomputer. The first minicomputer was the PDP-8 (1965) from Digital Equipment Corporation (DEC). The precursor machines were the Bendix G-15 (1956), the LGP-30 (1956) and the LINC (1962), which is considered the first minicomputer. The representative models were the PDP (Programmable Data Processor) line, with the PDP-11 as a reference machine; the VAX-11, with the VAX-11/780 as the reference machine for DEC and the AS/400 from IBM (Figure 1.21). From this class, a derived subclass with superior performance was developed. These machines were called super minicomputers. A super minicomputer had a working data format two to four times greater than the basic version of a given generation. It was equipped with hardware accelerators or coprocessors for computation of vectors or floating-point numbers. Two representatives were the VAX 6000 series (Sullivan *et al.* 1990) and the 8000 series (Burley 1987).

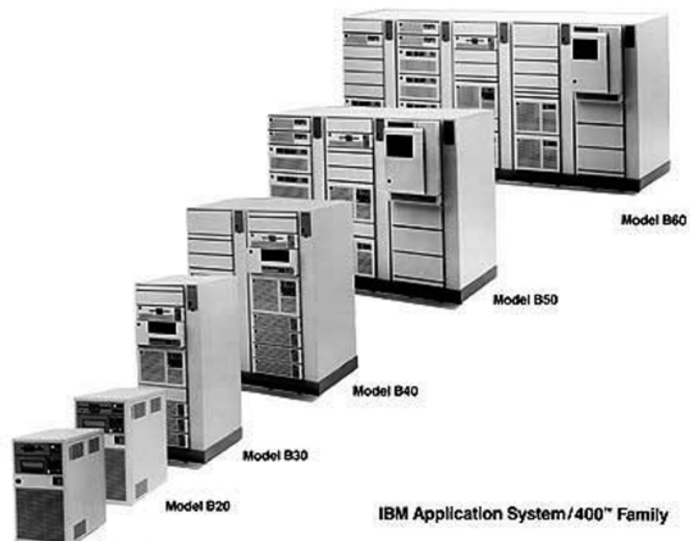


Figure 1.21. *The IBM Application System (AS/400) family of minicomputers*

Bell (2014) divides the evolution of this sector into three periods: the establishment of the associated industry (1956–1964, in green in Figure 1.22), the “triumph” (1965–1974), which overlaps with the period of cohabitation with the first (V)LSI microprocessors (1971–1984), and the decline (1985–1995, in red on this same figure). The advent of the 8-bit and especially 16-bit microprocessor generations made it possible to introduce low-cost, entry-level versions (low end).

Examples include DEC's LSI-11 (DEC 1975, 1976) chipset for its PDP-11 and the MicroNova mN601 (Godderz 1976) microprocessor or MPU (Godderz 1976) from Data General Corp. (compatible with the Fairchild F9440 Microflame™), equipped with mN606 4 Ki × 1-bit RAM and the mN603 I/O controller; these product lines provided 8- and 16-bit architecture respectively.

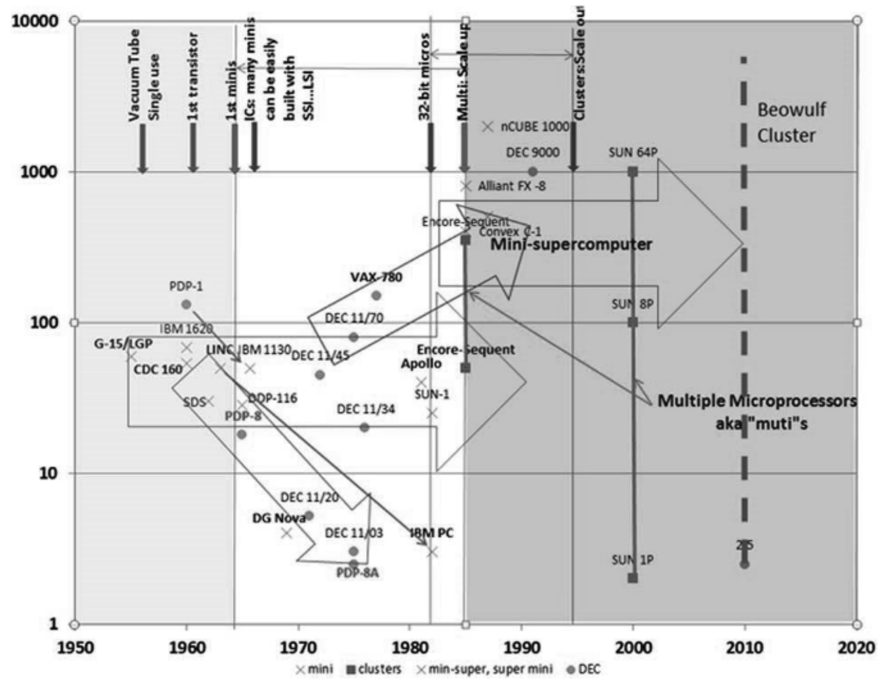


Figure 1.22. *Evolution over time of the prices of minicomputers (in thousands of $) (Bell 2014) © IEEE. For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

The decline in minicomputers was due to advances in the integration of electronic technology that led to the appearance of the microprocessor, which replaced it. The impact of this component on the market in this sector has been studied in particular by Schultz *et al*. (1973). In 1973, the microprocessor had a memory access time slower by a factor 1/2 to 1/3 compared to a mid-range minicomputer.

The (personal) workstation is a powerful single-station computer built initially around a 16-bit microprocessor (the MC68000 from Motorola), mainly using the multitasking UNIX OS and connected to an Ethernet-based Local Area Network

(LAN). The goal was to break away from the classic computer-terminal system for individual use of interactive applications. The work environment was to be shared (files, for example) and distributed over a network. It offered high-resolution graphics capacity, eventually including color. The preferred fields of application were Computer-Aided Design (CAD), Computer-Aided Drawing (CAD) and HPC. Other fields included computer graphics, video processing and 3D (three-dimensional) image synthesis. The initial idea dated from the 1950s, with prototypes in the 1960s using a more powerful computer (mainframe) connected to a graphics terminal. The first prototypes appeared in the early 1970s with graphic terminals connected to mini or autonomous computers such as the Alto PARC (Xerox Palo Alto Research Center) in 1973. Mature products emerged during the next decade. Apollo (since bought by HP Company), Silicon Graphics (now SGI), Sun Microsystems (since bought by Oracle Corporation) and Xerox were the four representatives of this class with their first machines respectively the Domain DN-100 (1981), the IRIS 1400 (1984), the Sun-1 (1982) and the Xerox STAR (1981). We should also mention IBM's RS/6000 (January 15, 1990). We began to refer to 3M and 5M machines. RFC (Request For Comments) 782 (Nabielsky and Skelton 1981) specified that a 3M machine had at least one MB (i.e. MiB) of memory, a screen with a resolution of at least one Megapixel, and computing power of one Million Instructions Per Second (MIPS). In addition, it should not cost more than one "Megapenny" ($10,000 at the time). The term 5M referred to *Megabyte memory, Megapixel display, MIPS processor power, 10+ Megabyte disk drive,* and *10 Megabit/s network*). Bell (1986) and Nelson and Bell (1986) paint a portrait of this type of machine and trace its evolution. It appeared thanks to the development of local networks. The price range was $\$10^3$–$10^4$. The systems used 32-bit RISC (Reduced Instruction Set Computer)-type microprocessors because of their computing power. Today (2010), 64-bit Intel microprocessors are used. Figure 1.23 shows a modern workstation.

Figure 1.23. *An Octane graphics workstation from Silicon Graphics, Inc. (SGI). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

The microcomputer[15] is a self-contained computer whose central unit is a microprocessor. It is a general-purpose computer, personal, inexpensive, and with limited computing capacity, in comparison with the above categories. It was invented in France, with the Micral N, which was patented in 1974 (Gernelle 1974) (Figure 1.24(a)). It was intended in particular for use in real-time applications. The first model sold in kit form in 1975 was the ALTAIR 8800 (Roberts and Yates 1975a b) from the American company MITS (Figure 1.24(b)). Before these, there were evaluation kits (*cf.* § V5-2.1.1) for implementing a given microprocessor. The first generation (1971–1976) is considered to include the pioneers. The second generation (1977–1990) saw the introduction of home microcomputers. The first machines were built around an 8-bit microprocessor and natively included BASIC (Beginner's All-purpose Symbolic Instruction Code) ROM (Read-Only Memory). An audio cassette player provided mass storage (*cf.* § 7.2.2 in Darche (2003))! These computers include the iconic Apple II, the Commodore PET (Personal Electronic Transactor) 2001, the Tandy TRS-80[16] from Tandy RadioShack, the BBC (British Broadcasting Corporation) Microcomputer System (1981) from Acorn Computer, and the ZX Spectrum (1982) from Sinclair Research Ltd. Libes (1978) describes this generation's technology. These machines naturally evolved towards a 16-bit architecture. With the appearance of the Floppy Disk Drive (FDD, *cf.* § 7.2.2 in Darche (2003)) in the mid-1970s, these machines became equipped with a simple OS such as CP/M (Control Program for Microcomputers, Digital Research, Inc.). The microcomputer was de facto standardized with the Personal Computer[17] (PC) from IBM, which had a microprocessor with a 16-bit internal architecture (although the external interface was in 8-bit format) from Intel. The first laptops that were not self-sufficient in terms of power were released in 1983 (the Compaq Portable) and 1984 (the IBM[18] Portable PC 5155 model 68). In the 1990s, a representative computer used by large companies was the IBM PS/2. Today, the battery-powered laptop has an energy autonomy of less than 10 hours. The microcomputer is available as a touchscreen tablet PC, which first appeared in 2001, and the associated phablet telephone (phablet is a contraction of the words smartphone and tablet), which appeared in 2010. Doerr (1978) describes this period.

15 This is called the "micro" for short.

16 TRS stands for Tandy RadioShack.

17 This name was popularized with this machine, but the term has multiple origins (Shapiro 2000). One source was the magazine Byte, which published an editorial by Helmers in its May 1976 issue. An older source can be found in an advertisement for the HP 9100 in the journal Science on October 4, 1968 (HP 1968).

18 This company was the first to commercialize the (trans)portable computer in 1975, the 5100. It was not designed around an MPU, but a custom controller called the PALM (Put All Logic in Microcode).

a) **b)**

Figure 1.24. *The first microcomputers: the Micral N from R2E and the ALTAIR 8800 from the American company MITS (source: unknown). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

Single-Board Computers (SBC) were a variant of the microcomputer used for process control and automation. These are complete MPU computer systems on a single printed circuit, sometimes in an OEM (Original Equipment Manufacturer) version. One of the first companies to offer this type of equipment was DEC with the LSI-11 series (Doerr 1978, Stiefel 1978). They were mainly intended for the industrial market, although there was also a market for individuals (i.e. hobbyists). These boards generally had a standardized form factor (*cf.* § V5-3.4.1) and communicated via an industrial bus (*cf.* § V2-4.8). There were also versions for embedded systems, such as those from the PC/104 consortium.

Another type of computer is the embedded or on-board system. These systems contain a dedicated management system. It therefore consisted of an electronic system with autonomous software from an operating point of view. The software is built-in. The algorithms implemented are generally complex and require a powerful microprocessor. They often involve a real-time concept. Etiemble (2016) thus distinguishes, on this last criterion, between on-board systems, which are mobile with real-time constraints, and embedded systems, which are fixed and without real-time constraints, although the term "embedded" is generally considered to encompass both definitions. Autonomous energy systems impose thermal and current consumption constraints. Examples of applications are computer peripherals like the printer, automated systems like industrial controllers and mobile devices like the mobile phone. They are widely used in the transportation sector (car, train, rocket, etc.). And similarly to what was said in the previous section, after 2010, the line between the on-board system and the microcomputer becomes increasingly blurry for certain devices like the mobile phone (smartphone).

Advances in microelectronic technologies, mainly CMOS, have enabled advances in the fields of storage and communications, allowing space-saving (corresponding form factors: SFF for Small Form Factor), low current consumption and communications-based devices to emerge. We are referring to Cell-Phone-Sized Devices (CPSDs). As illustrated in Figure 1.25, digital systems equipped with one or more microprocessors or microcontrollers (MCU for MicroController Unit, *cf.* definition in § V3-5.3) are becoming ubiquitous. In everyday life, they control, for example, washing machines (1), cars (2) and cameras (3). In healthcare, they manage heart rate monitors (4) and control artificial hearts. They are found in network interconnection equipment such as the router (6), the gateway and the Internet modem. They are used in entertainment electronics, for example, in game consoles (7) and sound/video players (PAD for Personal Audio Device, PVD for Personal Video Device, PA/VD for Personal Audio/Video Device). They are used in mobile devices such as the mobile phone (5), the Personal Digital Assistant (PDA), and the portable microcomputer (8). The server (9) and the mainframe (10) perform computation even faster thanks to the microprocessor. They have become ubiquitous in connected objects, with one commercially available example being the connected watch (11). They are also embedded in devices such as the connected electric meter (12). Since the early 1990s and the beginning of this century, Wireless Sensor Networks (WSN) and the Internet of Things (IoT) have opened up a wide field of applications. The 21 Century will undoubtedly see the rise of robots (13). The boundaries between all these categories are increasingly blurry or non-existent. Today (2000), servers are often structured like microcomputers with more powerful technical characteristics including computing power and the size of primary and secondary storage. The same goes for the workstation.

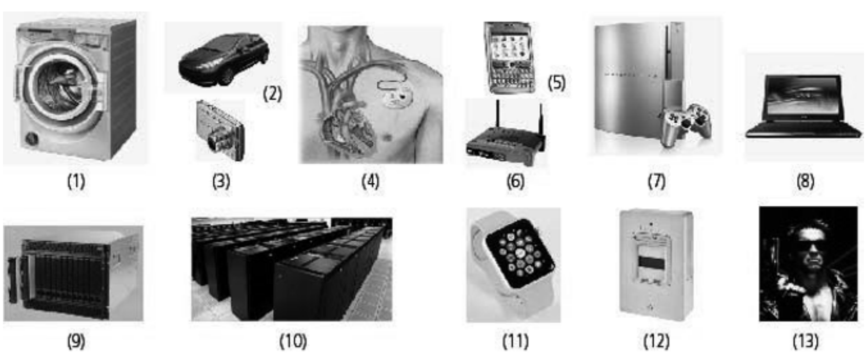


Figure 1.25. *Increasingly blurry boundaries. For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

As of 2010, the microprocessor is the foundation for all classes, from the supercomputer to the Internet of Things (Figure 1.26). It is gradually eliminating the notion of classes! This trend was symbolized by the expression “killer micro” popularized by Brooks (1989), which refers to CMOS-based microprocessors, which were going to gradually replace the mainframe computers, minicomputers, and super-computers.[19] Belak (1993) illustrates its applications. Figure 1.26 shows the evolution of these classes over time. Michael Burwen (Architecture Technology Corporation 1991) divides classes into three categories of large, medium, and small systems. In the first, we find super computers, central computers with vector computing capability, and parallel computers. The midrange systems category includes mini super computers, classic mainframe computers, servers and super minicomputers. Small systems include servers, conventional and graphical workstations, and other systems. They merge into a single category, namely, multiprocessor systems. We can also add two other categories for microcomputers and small-form factor systems.

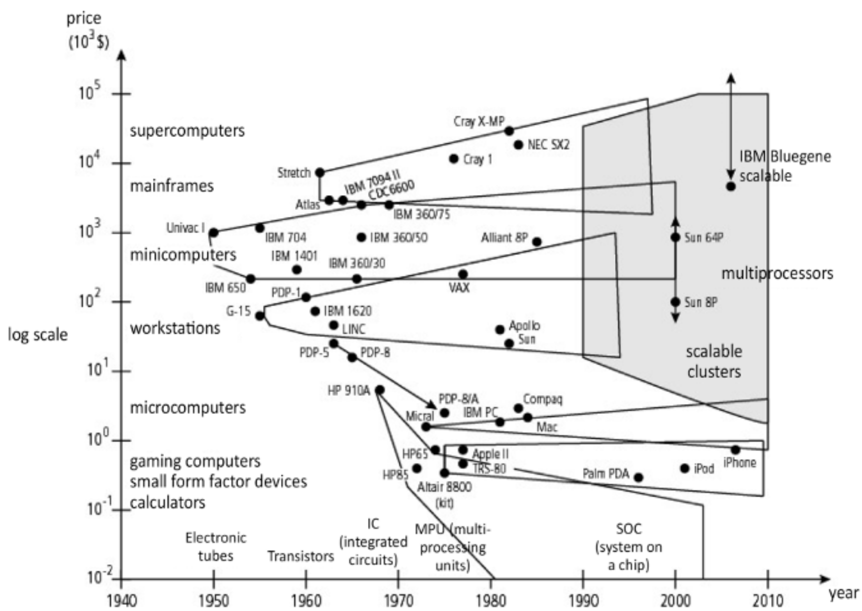


Figure 1.26. *Categories of computers (according to Bell (2008a))*

Thanks to the development of communication networks, the server offers services to connected clients such as the sharing of applications or storage. These

19 “No one will survive the attack of the killer micro!”

computers have powerful computing, storage and communication capabilities (Figure 1.27).

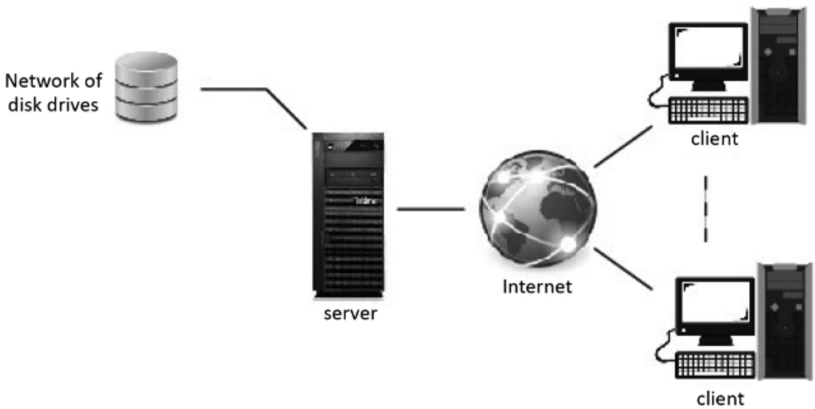


Figure 1.27. *The client–server model. For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

Figure 1.28. *A blade server. For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

Its modern form is the blade server. Several blades are mounted in a chassis that supplies them with power and cools them with air, as illustrated in Figure 1.28. This pooling of electronic sub-assemblies (power supply, cooling system, communication elements and even storage systems) makes it possible to achieve a compact format. The alternative solution is the server rack, which is a cabinet that accommodates servers with standardized dimensions (unit: U = 1.75"). Another form factor is the tower. Haghighi (2001) describes this architecture.

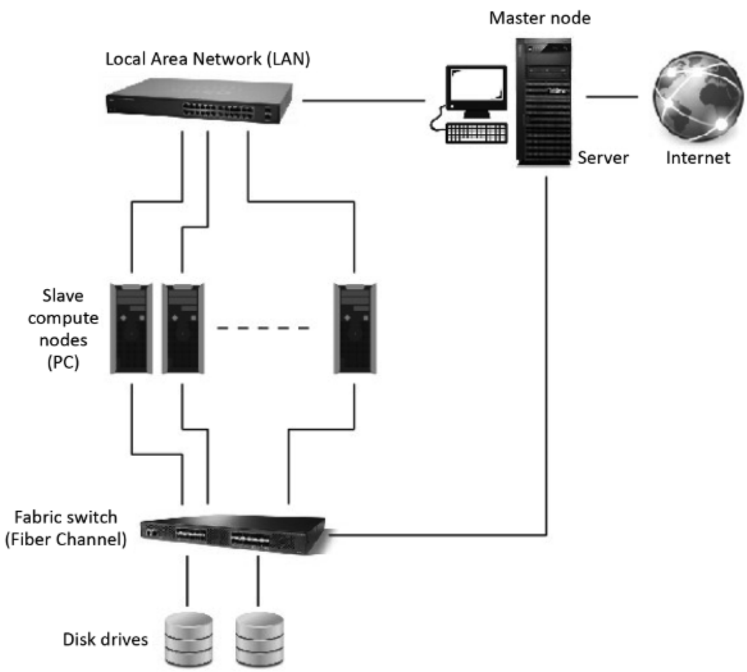


Figure 1.29. *Example of a Beowulf server architecture. For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

Today, in order to increase computing power, the computer has become parallelized. This means that it is made up of several computing units (parallelism of execution) and that the data is shared. Access occurs simultaneously (data parallelism). Computing and data resources can be distributed geographically, and access must be transparent. Thanks to the development of communication networks and these types of parallelism, three types of computing resources have appeared, namely, compute farms, grid computing and cloud computing. The term cluster designates a set or cluster of a dozen servers at most, also called a compute farm. Each node in the cluster is a commercially available computer, and these nodes are identical in terms of hardware and OS (property of homogeneity). Figure 1.29 shows an example of a Beowulf cluster (Sterling *et al.* 1995). For a workstation, we speak of NOW (Anderson *et al.* 1995) for a Network Of Workstations, or COW for a Cluster Of Workstations. The nodes communicate via a broadband network such as Infiniband. The cluster appears externally as a single computer, with the user connecting securely via a front-end server. This organization facilitates the hardware management of the different computer components (processor, primary and secondary storage, communication interfaces, etc.) and the software. It allows for easy scalability. From the point of view of fault tolerance, it increases availability. One possible use is HPC. For more information on the subject, see Pfister (1998).

A computational grid is a heterogeneous hardware and software infrastructure that is geographically distributed (i.e. delocalized) and allows virtual organizations (individuals, institutions, etc.) to solve problems and share data (Foster *et al.* 2001). This term was chosen by analogy to the electrical network, which supplies energy pervasively. This virtual computer system thus offers extensible and transparent access to distributed resources. A node in the grid can be a supercomputer or a cluster. These resources can communicate via any type of network, including LAN, Metropolitan Area Networks (MAN), or Wide Area Networks (WAN) like the Internet. Originally, this Information Technology (IT) infrastructure was implemented in response to the scientific community's needs (particle physics in particular) for distributed computing (computational grid) and data storage (data grid). A desktop grid is a variant where weakly coupled standalone computers participate in global computation in their spare time. For more information on the subject, see Foster and Kesselman (2003) and Shiva (2006).

Cloud computing is a paradigm that enables a user to relocate computing and storage resources, which can then be accessed via a network, generally a metropolitan or wide area network like the Internet. The two earlier infrastructures can be incorporated. These hardware and software resources are delivered as services. Table 1.7 summarizes the main differences between farm, grid and cloud computing. The main advantages of these architectures are their scalability thanks to their modularity and their high power/cost ratio thanks to the standardization of hardware and software components.

Depending on the level of service, cloud computing can provide Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). The first offers a virtualization solution for servers, networks and data storage. The second also provides middleware. The latter, also known as the "ASP model" (Application Service Provider), offers software functionality over the Internet where the application, in whole or in part, is run on remote servers.

Characteristics	Computing farm	Computing grid	Cloud computing
Homogeneity/Heterogeneity	Homogenous	Heterogeneous	Either/or
Hardware and OS characteristics	Identical hardware and OS	Different hardware and OS	Computers managed by the OS in physical units
Allocation	Works as a single unit with no use of external computing resources	Can call on "idle" PC computing resources	Several applications executed in parallel
Geographical distribution	One location	Distributed over local, metropolitan, and wide area networks	Distributed primarily on metropolitan networks
Resource management	Centralized	Independent nodes with own-resource management	Independent nodes
Centralization	Centralized and tightly coupled	Decentralized and loosely coupled	Dynamic infrastructure
Task and scheduling management	Centralized	Decentralized	Minimal management or self-managed platform
User interface	Appears as a single system	Appears as a dynamic and diversified system	Self-service use
Application domain	Education, research, engineering	Simulation and modeling, Computer-Aided Design (CAD), Research	Banking, insurance, weather forecasting, SaaS

Table 1.7. *Comparison of characteristics between computing resources (from Suri and Sumit Mittal (2012))*

1.3. Analog approach

The beginning of the history of computers is generally dated from the appearance of calculating machines (*cf.* § 1.1), but we must not forget analog calculating machines. The first known analog calculator is the Antikythera mechanism (Figure 1.30) dated around 90–60 BC. It described the movements of the moon and the sun in order to predict an eclipse.

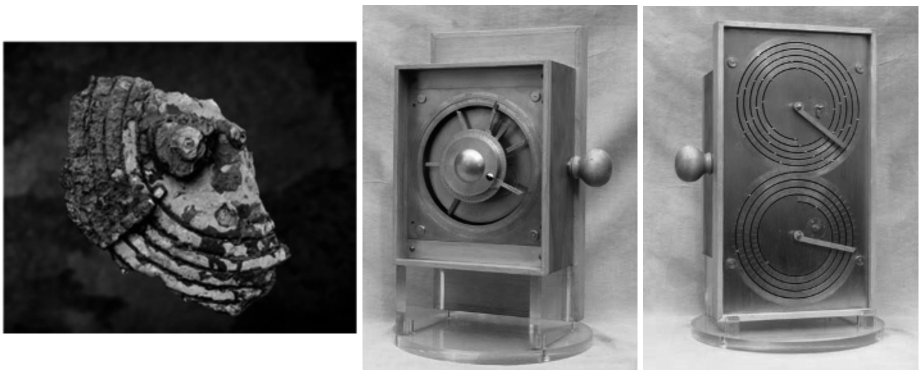


Figure 1.30. *The Antikythera mechanism (left) and a reconstruction (right), by Michael Wright (source: unknown). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

An analog computer, unlike the digital version, uses analog quantities proportional to the calculation values (Truitt and Rogers 1964). The precision of the result depends specifically on the measurement of the result. The underlying technology was hydraulic, mechanical, then electrical and electronic. These computers could be for general use, such as the slide rule, or they could be dedicated. They were primarily used to perform simulations of physical phenomena because they were able to compute the four basic arithmetic operations (+, -, etc.) or more complex functions like integration or derivation. An example of mechanical technology is Vannevar Bush's differential analyzer (1931; 1945). Data entry was done using an entry table that looked like a plotter. Figure 1.31 shows an electronic implementation, the PACE (Precision Analog Computing Element) 231R-V analog computer from the leading company of the time, Electronic Associates Inc. (EAI). The output device could be a galvanometer, an oscilloscope or a paper recorder. These devices were used by industries including aeronautics and space for simulation and systems control purposes. Variants were the hybrid computer (logic/analog combination) and the high-speed analog computer. Its industrial decline in the late 1970s was due to the predominance of digital electronics, which became cheaper, faster and more precise with standard components like the microprocessor. For a historical introduction, see Small (2001).

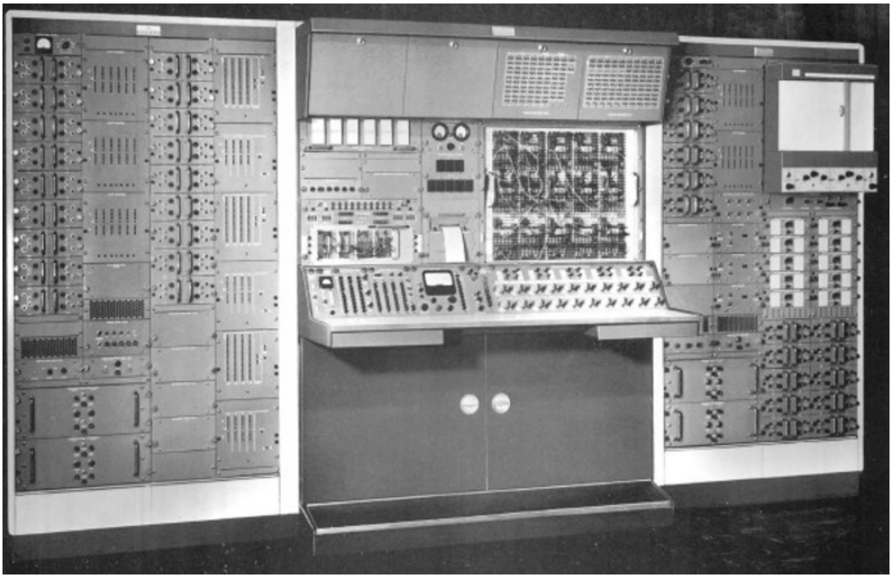


Figure 1.31. *The PACE 231R-V analog computer system from EAI (EAI 1964)*

1.4. Hardware–software relationship

It is now necessary to position software (SW) in relation to computer hardware. Figure 1.32 shows a layered view with hardware and software examples. The lowest layer is the hardware layer (HW). The microprocessor component of the computer is the subject of this work. Primary storage has been described in Darche (2012). I/O interfaces were discussed in Darche (2003). Above, a first low-level software layer called firmware[20] (FW) tests the hardware, initializes it, and loads the operating system. Designated by the term BIOS (Basic Input/Output System) in the PC (Personal Computer) world, it includes a set of software routines executed in interrupt mode (*cf.* Chapter V4-5). The third software layer is the OS (operating system software), which manages hardware and software resources for the whole. It is responsible in particular for security in the broad sense. The execution of applications (last layer, application software) relies on the latter. Between each layer, an interface allows the upper layer to use the services of the one below. In

20 In contemporary usage (*cf.* § V5-3.5.1), this is a program stored in ROM, as opposed to one stored in mass storage (secondary or tertiary memory), which is referred to as software.

particular, a virtual machine is a software layer located either below the OS (System Virtual Machine) or above it (Process Virtual Machine), which emulates another architecture and its associated computing model (within the meaning of Chapter 3). On this last subject, Smith and Nair (2005) provide an excellent reference.

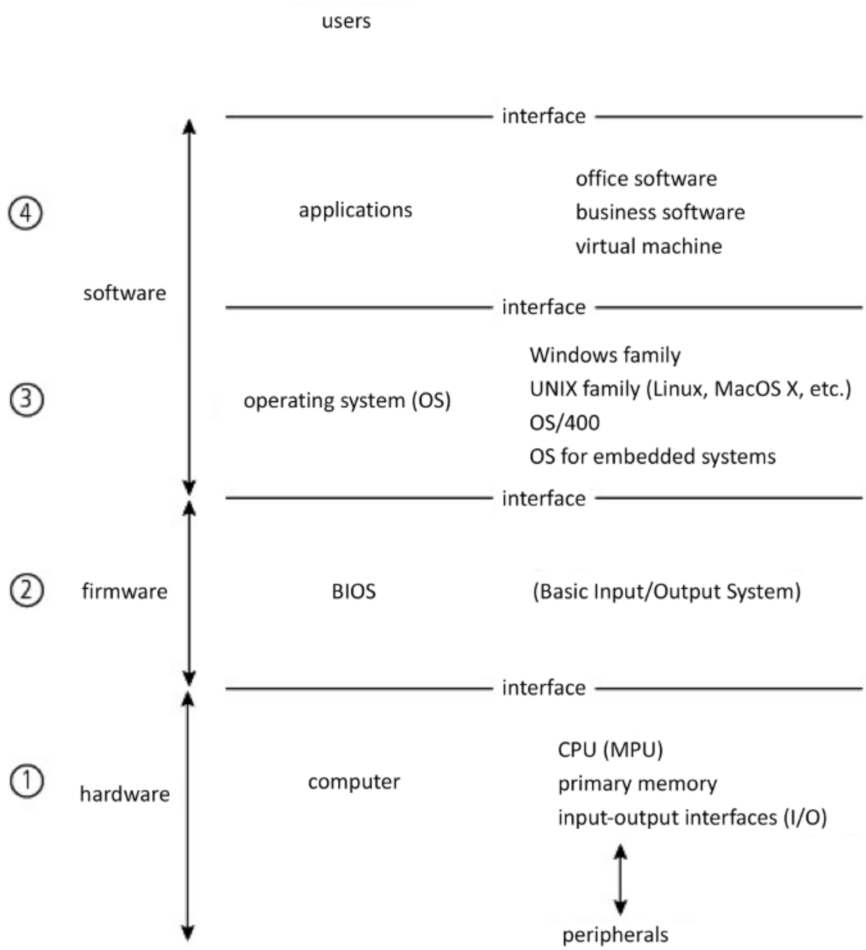


Figure 1.32. *Layered view of software infrastructure*

The technological generations within the meaning of § 1.2 provide support for software concepts such as programming languages and operating systems. As noted by Denning (1971), this classification is also linked to technological advances in the software field. In addition, they characterize the entire computer system. Thus, the term “generation of computers” encompasses both hardware and software technologies, as summarized in Tables 1.8(a) and 1.8(b).

	Generations of electronics	
Characteristics	**1st**	**2nd**
Period	1946–1955	1956–1965
Electronics		
Electronics components	vacuum tube	transistor
Cycle time	0.1–1 ms	1–10 µs
Primary storage		
Components	Delay line Electrostatic tube Magnetic drum (start)	Magnetic drum Magnetic core
Access time	1 ms	1–10 µs
Secondary storage	Punched tape Punched card Delay line	Punched card Magnetic tape Magnetic disk Magnetic drum
Programming languages	Machine language Assembly language (start)	Assembly language High-level language (HLL)
Hardware concepts	Arithmetic units	Floating-point units microprogramming (concept) interruption I/O processor
Software concepts	-	Batch processing monitor
Hardware examples	ENIAC IAS Princeton UNIVAC IBM 650/701	IBM 7090–7094 CDC 1604 CDC 6600

Table 1.8a. *Generations of computers and main features*

	Generations of electronics		
Characteristics	**3rd**	**4th**	**5th**
Period	1966–1975	1976–1989	1990–202X
Electronics			
Electronics components	Integrated circuit (SSI-MSI)	Integrated circuit (LSI-VLSI)	Integrated circuit (ULSI-GSI)
Cycle time	0.1–1 µs	0.1–1 µs	1 µs–1 ns–0.1 ns
Primary storage			
Components	Magnetic core Other magnetic media	Solid-state memory	Solid-state memory
Access time	0.1–10 µs	0.1 µs	100 ns–< 1 ns
Secondary storage	Same as 2nd generation Extended core storage Mass core storage	Same	Magnetic hard disk Solid-State Disk (SSD)
Programming languages	High-level languages	High-level languages Concurrent programming	
Hardware concepts	Microprogramming Pipeline cache Pagination (Virtual memory) Code translation	Microprocessor (1971) Microcomputer (1973)	Instruction-Level Parallelism (ILP) Thread-Level Parallelism (TLP) (multicore) Massively Parallel Processing (MPP) Heterogeneous environment
Software concepts	Timesharing Segmentation (virtual memory) Multiprogramming	Multiprocessor OS Windowing	
Hardware examples	DEC PDP-8 IBM System 360/370 ILLIAC IV CDC 6600 TI ASC Cray 1 Cyber 205 (Control Data)	IBM PC VAX 9000 IBM 3090 Cray X-MP	Cray MPP Fujitsu VPP500 TMC CM-5 Intel Paragon

Table 1.8b. *Generations of computers and main features (continued)*

ADDITIONAL CONCEPT.– We classify languages by increasing levels of abstraction. Generation 0 was machine language. The next generation includes Assembly Languages (AL, *cf.* § V5-1.3). These are low-level languages. The first high-level qualified languages that characterize the third generation appeared in the 1950s. These include FORTRAN (FORmula TRANslation, 1957), ALGOL (ALGOrithmic Language, 1958) and COBOL (COmmon Business Oriented Language, 1959). The designation "4th Generation Languages" or 4GL characterizes languages that are close to natural language. They facilitate programming by offering, for example, easy access to databases and a better Human–Machine Interface (HMI). Some enable automatic generation of lower level code. They are generally specialized for a particular field such as mathematics or management. The subcategories are query languages, data reporting and code generators.

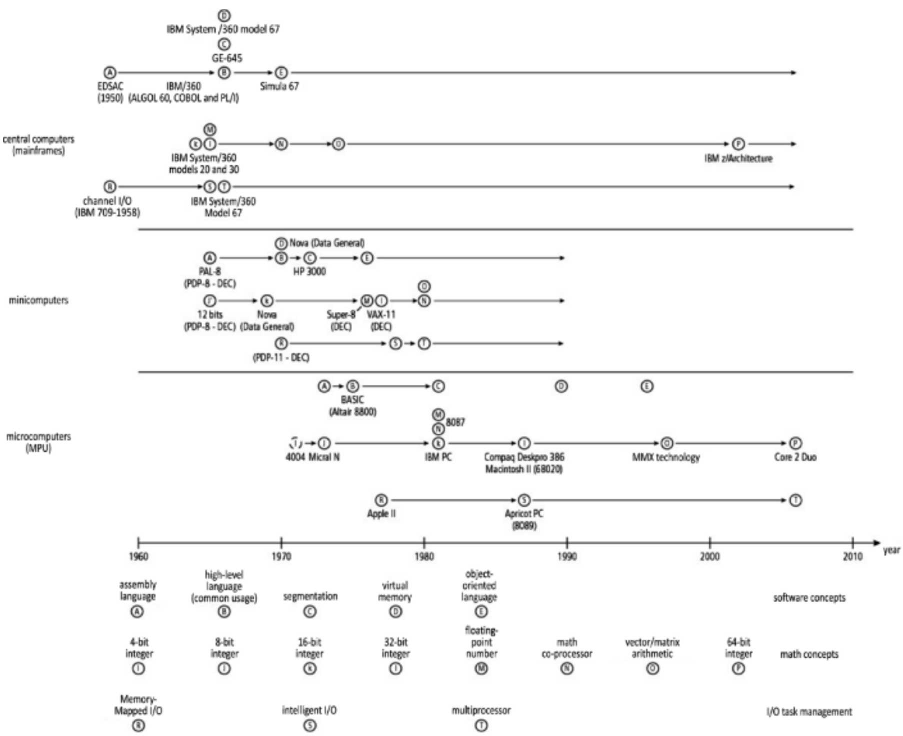


Figure 1.33. *Historical timeline of the evolution of concepts for the families of computers (inspired by Burger et al. 1984)*

A new class of computers is benefiting from advances in technology and integrating earlier hardware and software concepts as soon as technically possible, as shown in Figure 1.33. An example is the Virtual Memory (VM) mechanism that appeared in MPUs only in 16-bit versions with segmentation in the 8088/8086. Thus computer concepts shift from class to class when necessary and when the technology allows. Innovations, initially much more spread out in time, saw their rate of appearance intensify with the invention of the MPU. The ‘I’’ mark is just a time marker to indicate the release of the first MPU, the 4004.

As soon as computing power becomes sufficient, new applications can be supported, as illustrated in Figure 1.34 for the multimedia field.

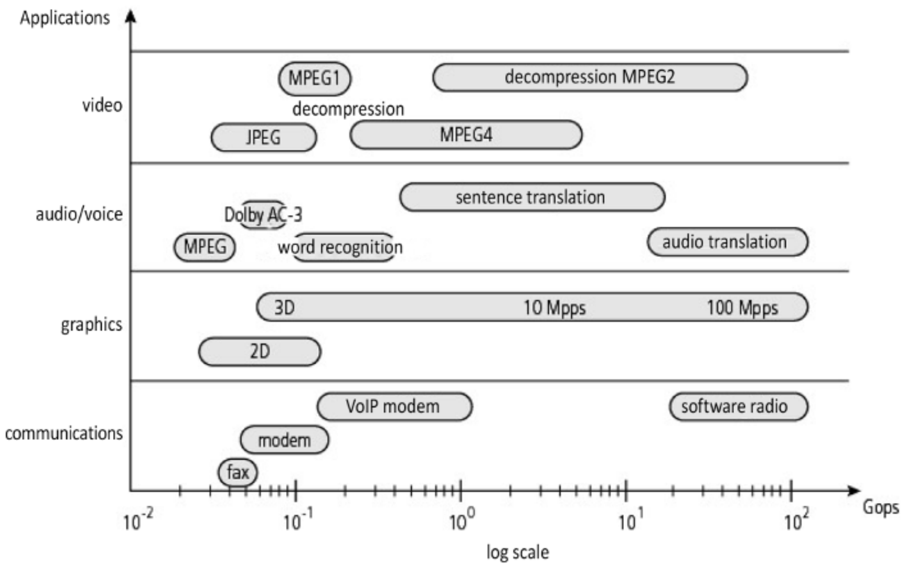


Figure 1.34. *Need for computing for multimedia applications (based on 2003 ITRS document)*

Finally, it should be noted that the notions of operator/programmer/user were initially nested. As shown in Figure 1.35, the operator was first responsible for executing the programs written by the programmer/user. Later, the concept of user was detached from that of programmer to be attached to the operator. These changes are due to progress in both hardware and software.

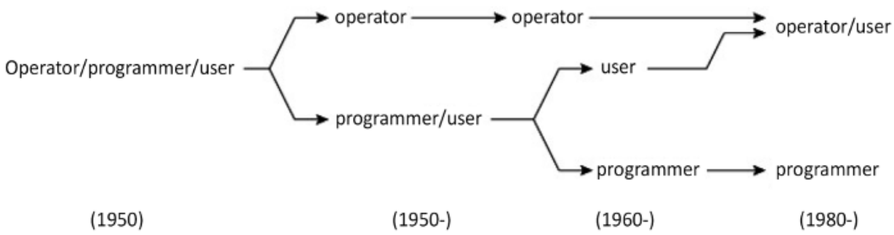


Figure 1.35. *Evolution of computer roles (from Nelson and Bell (1986))*

1.5. Integration and its limits

Electronics, with the three aforementioned active components (diode, transistor and integrated circuit), represented a major technological development that marked the computer industry. Continued advances in microelectronics have increased the functional density of integrated circuits and the speed of information processing. An observation made by Intel co-founder Gordon Moore that bears his name, Moore's Law, was that the number of transistors integrated on a microchip would double every 18 months (Moore 1975). This value has varied over the years from 24 (Moore 1965) to 12 months, eventually stabilizing at the aforementioned value. Figure 1.36 shows the evolution of the number of transistors for our topic. The dominant manufacturing technology, initially essentially bipolar, is today unipolar, namely, MOS and, more precisely, CMOS.

Bell (2008a b) gives the equation for the growth in the number of transistors n per chip as a function of the year t, which is:

$$n = 2^{t-1959} \text{ (for } 1959 \leq t \leq 1975\text{)} \quad [1.1]$$

$$n = 2^{16} \times 2^{\frac{t-1975}{1,5}} \text{ (for } t \geq 1975\text{)} \quad [1.2]$$

A professional association, the International Technology Roadmap for Semiconductors (ITRS), representing the main regional professional associations in the sector, publishes a report every two years detailing the future of the semiconductor industry. Current technologies for manufacturing primarily CMOS-based integrated circuits are reaching their limits. There is what specialists call the red brick wall first predicted in 2001 to occur in 2005–2008 (ITRS 2001), which is the physical limit for etching.

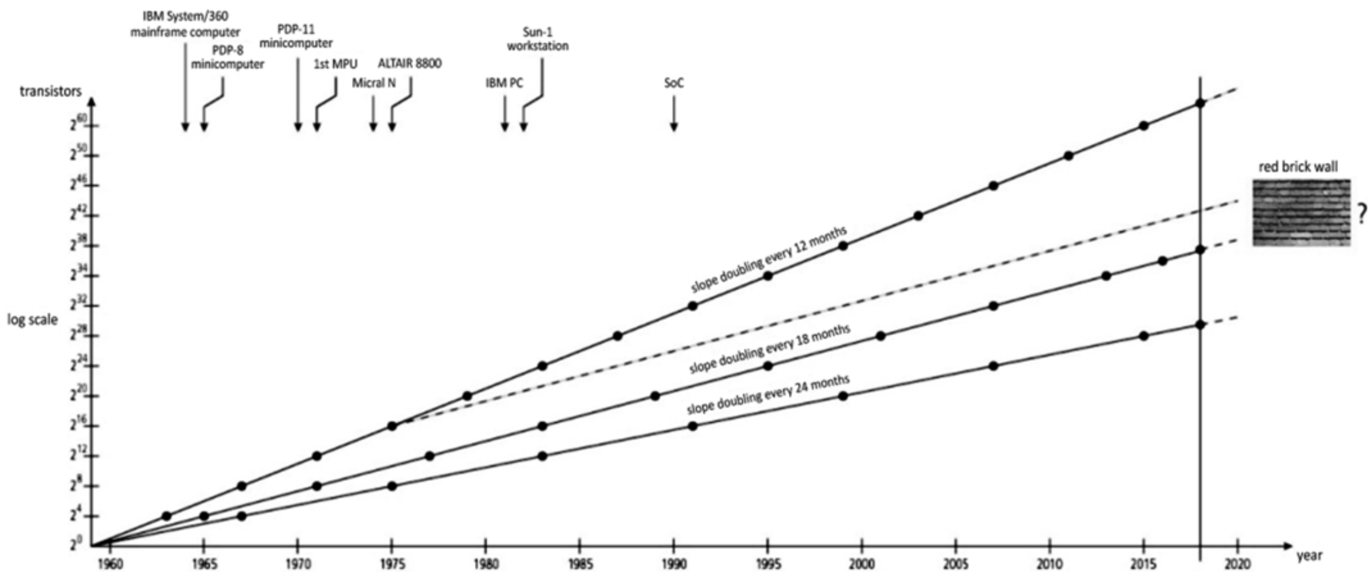


Figure 1.36. *Evolution over time of the number of transistors of an integrated circuit. For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

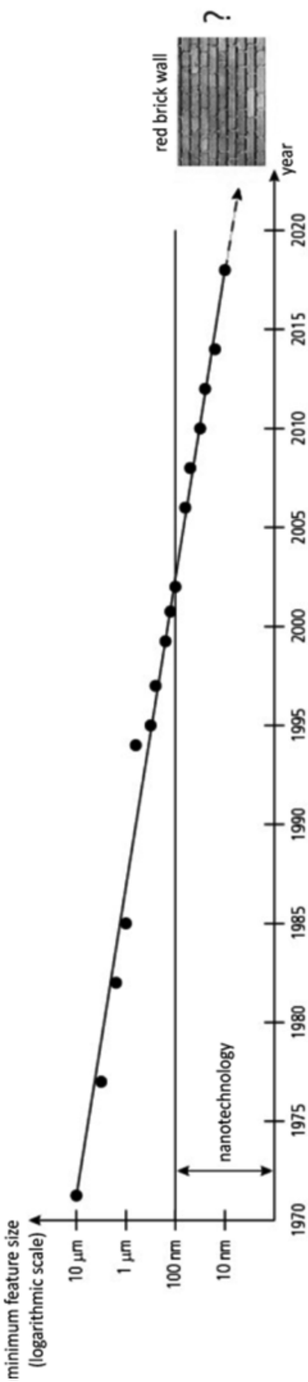


Figure 1.37. *The fineness of etching of integrated circuits over the years (technological node). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

The evolution in etching for integrated circuits depends on physical and technological parameters, as well as, of course, economic constraints. For example, as the width of the channel of the unipolar transistor decreases, the mobility of the electrons also decreases. Figure 1.37 shows its change over time. This controversial metric, used by the industry (*cf.* (Arnold 2009)) and withdrawn by the ITRS in 2005, is called the technology node, also referred to as the process, technology or manufacturing node, or simply the node or generation. Depending on the case, this is the length of the gate of a Field Effect Transistor (FET) in MOS technology or the minimum distance[21] between two lines of metal or polysilicon. For storage, we speak of a half-pitch. This improvement in technology (minimum feature size and diameter of the wafer) then allows for greater integration and an increase in clock speed. Greater integration can mean an increase in the number of functional blocks of the CPU, or even a multiplication of processors for parallel processing. It also makes it possible to integrate an entire system (SoC approach).

Another wall is the power wall, which refers to a chip's maximum energy dissipation.

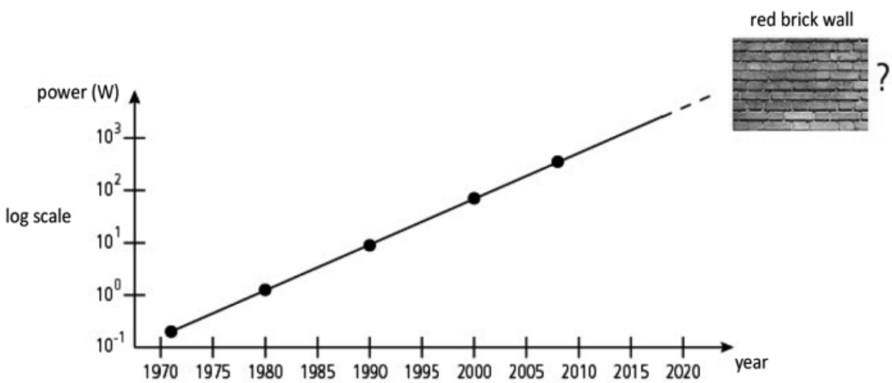


Figure 1.38. *The energy wall (from Xanthopoulos 2009 on data from ISSCC). For a color version of this figure, see www.iste.co.uk/darche/microprocessor1.zip*

Figure 1.38 shows the evolution of the power dissipation per unit area which continues to increase with the different generations of MPU. In this graph, power doubles every 3.5 years, going from 0.2 W in 1970 to 200 W in 2005. Considering that a chip has an area of 1 cm^2, the most extreme value to compare is the power density of a nuclear reactor, for example with pressurized water, which is around 300 W/cm^2. Feng (2003) performed the same study, but with Intel circuits.

21 Another term is "feature size", which is the minimum width that can be etched in silicon.

Dissipating more calories therefore requires increasingly efficient cooling. This requires the use of a better thermal conductor (i.e. with lower thermal resistance), for example ceramic or metal, or the implementation of forced circulation of a gaseous heat-carrying fluid (air) or liquid.

Another physical limit is the signal propagation speed wall. This is linked to the operating frequency of the logic. The propagation delay does not decrease, even with ever-finer etching. Matzke (1997) has shown that the more the technological node decreases, the less it is possible to reach distant logical subsets (Figure 1.39).

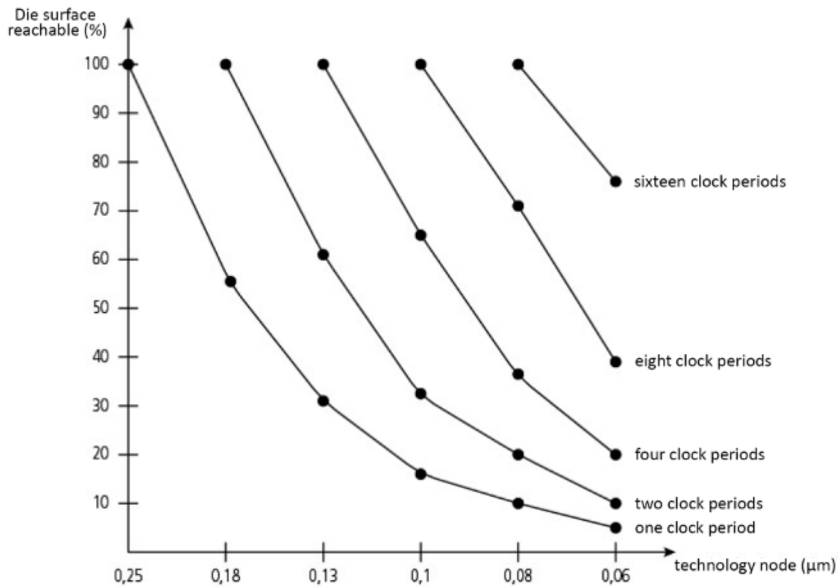


Figure 1.39. *Chip area achievable with progress in integration (according to (Matzke 1997), modified)*

1.6. Conclusion

Two inventions have enabled advances in the field of computing. These are the concepts of the stored program (*cf.* section 3.2.3) and the transistor (Patterson 1995). We must add a third, that of the integrated circuit, which, with VLSI and subsequent generations, enabled the emergence of complex integrated circuits like the microprocessor. Without them, the microcomputer and the democratization of computing would not exist. The evolution of computers is characterized by increasing generalization. Initially specialized for scientific computing, they have now invaded all fields thanks to the microprocessor. The latter destroyed the notion of computer classes.