HISTORY OF COMPUTATION

Sotirios G. Ziavras, Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, New Jersey 07102, U.S.A.

Keywords

Computer Systems, Early Computers, Electronic Computers, Generations of Electronic Computers, Mainframe Computers, Mechanical Computers, Microprocessor Architectures, Minicomputers, Parallel Computers, Supercomputers.

Contents

Glossary

CPU: Central Processing Unit.
Distributed processing: the process of executing a single program on a network of computers.
Local-area network: a network of connected computers in a relatively small physical area.
Mainframe computer: a large, high cost computer for very good performance.
Massively-parallel computer: a parallel computer containing hundreds or thousands of (micro)processors.
MIMD: Multiple Instruction streams, Multiple Data streams.
Microcomputer: a small computer driven by a microprocessor.
Minicomputer: it costs much less than a large computer (such as a mainframe) but can yield very substantial performance.
Multicomputer: a parallel computer system containing many microprocessors which are interconnected via a static point-to-point physical network.
Multiprocessor: a parallel computer system containing many microprocessors which are interconnected via a dynamic network. This network is implemented with either a common bus (that is, a common set of physical wires) or a network of switches. The exact interconnection scheme is determined each time by the application program.
Parallel computer: a computer system containing from a few to thousands of microprocessors.

Pipelined processor: a CPU that contains many processing stages. A computer instruction has to go through one or more stages to complete its execution. Many stages may be used concurrently by different instructions.
RAM: Random Access Memory. All data accesses use absolute memory addresses.
Semiconductor: material with electrical conduction properties much better than those of insulating materials and worse than those of classic conducting materials, such as copper or iron.
SIMD: Single Instruction stream, Multiple Data streams.
Supercomputer: a computer system capable of delivering performance many orders of magnitude larger than that of any single-processor computer. This term is currently associated with massively-parallel computers and vector supercomputers.
Transistor: a solid-state semiconductor device that controls or amplifies the current flowing between two terminals based on voltage applied to a third terminal.
Vector supercomputer: a computer system capable of delivering performance for array(vector) operations many orders of magnitude larger than that of any single-processor computer. It contains specialized parallel units for vector operations.

Summary

Modern life has been affected tremendously by advances in information technologies. Computers and their applications are in the core of these advances. Attempts to design and build devices to aid humans in their computing tasks go back to ancient times. Mechanical computing devices appeared more than 2000 years ago. The era of electronic computers started in the 1940s. Electronic computers are now ubiquitous in modern societies. They appear in many forms, from primitive units that control functions in automobiles and home appliances, to very advanced supercomputers for the design of aircraft and the simulation of climate changes. Understanding the history of computing can help us make reasonable predictions about its future. This article begins with a brief summary of computing techniques and technologies invented by early civilizations, discusses major breakthroughs in the design of electronic computers, and presents preeminent computing technologies of our times. Finally, it touches upon expected computing technologies of the future.

1. Early (Non-Electronic) Computing Machines

Any article on the history of computation can have omissions. An attempt is made here to present the most important advances in the field, in chronological order. Transaction records were inscribed on clay tablets before 1,000 B.C. Accustomed to finger counting, many early civilizations adopted number systems with five or ten as the base. The Sumerian, Greek, Roman, and Arabic number systems were some of the inventions. The abacus was invented about 5,000 years ago in Babylonia. Basic calculations are performed by hand-sliding beads on racks. It was also adopted later by other civilizations. Modified versions of the abacus are still used in the Orient.

The Antikythera Mechanism was a special-purpose computer built in Greece around 87 B.C. It was found in a shipwreck close to Antikythera, a Greek island. It was used to determine the location of the sun, the moon, and other planets based on the calendar year. It is a very sophisticated device that employs many bronze gears and an axle which can be rotated by hand to input the current date. It may have been used for celestial navigation. Euclid's geometrical reasoning enabled him in the 4^{th} century B.C. to develop algorithms for process description. Between 250 and 230 B.C., the Sieve of Eratosthenes algorithm was introduced to find prime numbers between any two given bounds. The Peruvians devised later a handy gadget, called a Quipu, consisting of strings and knots to assist them in the counting process. They used a form of decimal notation, where the position and form of a knot indicated tens, hundreds, etc.

In 1623, Wilhelm Schickard, a Professor of German origin, constructed the calculating clock for adding and subtracting six-digit numbers. It contains six dented wheels geared through a "mutilated" wheel; every full turn of the latter wheel allows the wheel located at the right to rotate one 10^{th} of a full turn. An overflow detection mechanism rings a bell. In 1642, at the age of nineteen Blaise Pascal invented the Pascaline, an adding calculator to help his father in his tax collection job. This is often considered to be the first mechanical adding machine. A set of notched dials moves internal gears in a way that a full rotation of a gear causes the gear to the left to advance one 10^{th} of a full turn. Although the first prototype contained only five gears, later units were built with six and eight gears. His adding machine of 1652 assumed six-digit numbers; six number dials on a row represent an input number. Each dial has ten opennings into which a stylus is inserted for dialing. Subtraction is performed by adding the 10's complement of the subtrahend to the minuend.

The British Samuel Morland also produced in 1666 a mechanical calculator for addition and subtraction. John Napier of Scotland invented a device comprising a set of vertical rods, where each rod is divided into ten squares. The Napier bones device was used to read directly rows containing multiples of numbers stored on the first row. Napier was also the inventor of the concept of logarithms, which are used to convert multiplications into simple additions.

Baron Gottfried Wilhelm von Leibniz, a German philosopher and mathematician, improved Pascal's device in 1674 to perform multiplication and division of numbers. His machine employs gears and dials. He was heavily influenced by Pascal's notes and drawings. Leibniz's Stepped Reckoner uses a special type of gear. It is a cylinder with nine bar-shaped teeth of incrementing length which are parallel to the cylinder's axis. When a drum is rotated by using a crank, a regular ten-tooth gear which is fixed on a sliding axis is rotated zero to nine positions depending on its relative position to the drum. There is one set of gears for each digit. When the drum is rotated, it generates in the regular gears a movement proportional to their relative position. This movement is then translated by the device into multiplication or division depending on the direction of the rotating drum. Leibniz is also known from his attempt to develop a generalized symbolic language and a related algebra to automate reasoning. As a result, he co-invented calculus; the other co-inventor was Sir Isaac Newton.

In 1777, the third Earl of Stanhope in England produced a multiplying calculator. The French Joseph-Marie Jacquard introduced in 1801 the concept of sequencing punched cards to control the weaving patterns and the use of fabrics in an automatic loom. The introduction of these looms caused riots against the replacement of people by machines in the second half of the 18th century. Jacquard's invention was of paramount importance for both the industrial revolution and the technological one, as shown later.

In the 1820s, mechanical calculators were used widely. Charles Xavier Thomas de Colmar, a native of France, invented the Arithmometer to implement the four basic arithmetic functions. Addition and subtraction are carried out through a single accumulator and ripple carry propagation is employed. Multiplication and division use repeated addition and subtraction through manual control. The arithmometer was used widely until the First World War. He used the stepped drum mechanism invented by Leibniz as the digital-value actuator.

Charles Babbage, a professor of mathematics at the University of Cambridge, is considered to be the father of modern computing. His focus for many years was task automation through the invention of appropriate devices. He had noticed by 1812 that machines were compatible with mathematics. Both could be used to repeat sequences of steps without mistakes. He proposed in 1822 his Difference Engine that could add and subtract to produce output tables for given functions; a single algorithm is run using finite differences with polynomials. It was to be powered by steam, would contain a stored program, and would print the results. He had constructed only a portion of the Difference Engine by 1832. However, he shifted his attention in 1833, at the age of 42, to the design of a general-purpose computing machine, the Analytical Engine. He produced a workable design in 1836. He made a major revision in 1837. He did not attempt to build the machine at that time, even though he worked intensively until the late 1840s on refinements of his design. He finally begun in 1856 an attempt to build the machine. Many parts of his machine had been constructed by the time of his death, in 1871. Babbage did not publish a detailed account of his machine. Luigi Menabrea of Italy published a relatively good description of the machine, after he had attended relevant lectures given by Babbage.

Ada King, the Countess of Lovelace and daughter of the poet Lord Byron, was an assistant to Babbage who translated Menabrea's description into English in 1843 by also inserting her own notes. This work was done under the supervision of Babbage. However, her translation does not contain enough information to evaluate the feasibility of the design. Similarly, Babbage's private notes, now residing in the Science Museum in London, contain extensive, but incomplete information. A preliminary feasibility evaluation of the machine based on the latter notes can be found in the bibliography. Ada King was also more effective and instrumental than Babbage in securing funds for the construction of the machine. She also introduced the idea of program loops (where sequences of instructions are executed repeatedly) and developed several programs for the machine. Therefore, she is known as the first computer programmer. ADA, a programming language developed for the U.S. Defense Department, honors her.

It is worth paying special attention to the design of the Analytical Engine. It was designed to contain about 50,000 components. Input devices were to accept punched cards representing instructions. Storage for 1,000 numbers of 50 decimal digits was to be provided. Babbage borrowed the concept of punched cards from the automatic loom produced by Joseph-Marie Jacquard. This machine uses a sign-and-magnitude representation of decimal numbers. Wheels rotating on shafts represent decimal digits. The "store" and "mill" components of the machine closely relate to the "central memory" and "CPU" of modern computers. Barrels are used to control multiplication and division. Studs on a barrel are screwed appropriately to control each operation. In current terms, the barrel acts like a microprogram store and a column of studs represent a word in the microprogram memory. The machine uses the 10's-complement representation for signed numbers. This convention complicates the operations of addition and subtraction. Nevertheless, it is easy to determine the sign of the result for multiplication and division. Strings of cards are created to represent programs. There are also separate strings of cards to contain the names of variables and their values, respectively. Variables cannot be accessed at run time in the RAM sense, that is by using addresses. This was also true of later electronic designs, such as the ENIAC and Mark I (discussed below).

Samuel Morse demonstrated in 1838 the principles of the telegraph. In 1844, he sent a telegraph message from Washington/D.C. to Baltimore in the U.S.A. A transatlantic telegraph cable was used for a few days in 1858. A telegraph cable connected together the east and west coasts of the U.S.A. in 1861. Alexander Graham Bell invented the telephone in 1876. Guglielmo Marconi transmitted a radio signal in 1895. He was also instrumental in a transatlantic wireless signal transmission in 1901. The importance of the latter inventions stems from the ubiquitous presence of the Internet in our days.

In 1854, George Boole published a paper that introduced the concept of binary logic as well as an appropriate set of operations. His paper on symbolic and logical reasoning had the title "An Investigation of the Laws of Thought." A variable can be in only one of two states at any time, namely "true" or "false." Electronic computers containing transistors use the binary (that is, Boolean) logic in all of their basic operations. They replace the two states with the binary numbers "0" and "1."

It took nearly seven years to process the data collected by the U.S. census of 1880. Herman Hollerith used punched cards to represent the input for the data processing machine of the next census in 1890. Each punch on a card represented a number, and combinations of two punches represented a letter. A single card could represent up to 80 variables. A hole on a card closes an electrical circuit when the card goes through a reader. The number of closed circuits is then counted by the reader. These new census data were processed in only about six weeks. Hollerith founded the Tabulating Machine Company in 1896 to produce the punch card reader. The company was renamed to International Business Machines (IBM) in 1924 after a series of merges with other companies. Remington Rand and Burroughs, among others, also manufactured punch card readers for the business world. Punched cards were used heavily until the 1960s.

The English John A. Fleming invented in 1904 the diode vacuum tube that could facilitate the design of electronic computers. In 1906, Lee de Forest introduced in the U.S.A. a third electrode to control the current flaw in Fleming's diode; he called the new device an audion. In 1911, the Dutch physicist Kamerlingh Onnes discovered superconductivity.

Vannevar Bush engineered in 1931 a calculator for solving complex differential equations. However, the complexity of this calculator was very substantial, containing hundreds of gears and shafts for the representation of numbers and their relationships.

In 1936, Alan Turing introduced the theoretical concept of the universal computing machine. This imaginary machine serves more like a computer program than computer hardware and can implement any kind of computation in algorithmic form. Since the interpretation of instructions and their implementation are mechanical processes, Turing suggested a simple machine that can implement any given sequence of instructions (that is, a computer program in the current sense). The machine comprises a read/write head that scans data on a one-dimensional tape of infinite size. The tape is divided into squares, where each square contains a symbol or blank. It reads the current symbol from the tape, and based on that symbol and current state it can overwrite the current symbol with another one, can change the current state, and then moves the head left or right on the tape. The American logician Alonzo Church independently reached similar conclusions about computability around the same time.

John V. Atanasoff, a professor at Iowa State College, investigated in 1939 the design of an all-electronic computer by using the Boolean logic for representing instructions and data, and performing calculations. Atanasoff extended the Boolean logic to represent 0 and 1 by the "off" and "on" state of an electronic device. He developed in 1942, with the help of the talented graduate student Clifford Berry, the ABC (Atanasoff-Berry Computer) computer. It uses a mechanical clock system, but computations are carried out electronically. It has two rotating drums containing capacitors, which hold the electrical charge for the memory. Data are entered using punched cards. For the first time, vacuum tubes were used in computing. The project cost about $1,000. It introduced the concepts of binary arithmetic in computers, regenerative memory, and logic circuits. Atanasoff was recognized as the father of modern computing when, in a patent infringement case that Sperry Rand brought against Honeywell, a U.S. federal judge voided in 1973 Sperry Rand's patent on the ENIAC (presented in the next Section), saying that it had been derived from Atanasoff's invention. While spending five days at Atanastoff's lab, Mauchly observed the ABC and read its 35-page manual. It was concluded that Mauchly had used this information to construct the ENIAC.

The German engineer Konrad Zuse constructed in 1941 the Z3 computer with electromagnetic relays for the design of missiles and airplanes.

2. Electronic Computers

There have been five generations of electronic computers. The underlying technology for the implementation of electronic devices has been the driving force behind advances in the design of electronic computers. Primitive electronic computers appear nowadays in telephone switching centers, supermarket scanners, automatic teller machines (ATMs) for bank transactions, anti-lock brake systems (ABS) in automobiles, etc. More advanced computers are used in personal computers, as switches in advanced computer networks, the design of aircraft, weather forecasting, etc. This section focuses on general-purpose computers that can solve problems in many diverse fields. Special attention is paid here to computers suitable for engineering and scientific applications. Understanding the past of electronic computing technologies will help us better predict the potential for further improvements in the future and the possible impact on our lives.

2.1. First Generation (1945-1955)

The Colossus computer was constructed in England in 1943 to decode messages transmitted by the Germans during the Second World War. Alan M. Turing, the famous British mathematician, helped in the development of the world's first electronic computer. An ENIGMA machine that was used by the Germans to encode messages had fallen into the hands of the British Intelligence. Huge amounts of computation were needed to break coded messages. The development of Colossus was kept secret for many years.

Howard H. Aiken of Harvard University (in cooperation with IBM) produced in 1944 an all-electronic calculator based on relays for the development of ballistic charts by the U.S. Navy. The Mark I calculator occupied very substantial space and included about 500 miles of wiring. Electromagnetic signals were used to drive mechanical parts. Grace Murray Hopper was a famous programmer of Mark I at the Harvard Computation Center; she was a U.S. Navy lieutenant.

The Second World War forced several governments to put serious effort in the development of electronic computers for the design of missiles and aircraft, and message decoding. Such a powerful computer was the Electronic Numerical Integrator And Computer (ENIAC) that was developed for the U.S. government at the University of Pennsylvania by John Presper Eckert, Jr. and John W. Mauchly. With their staff of fourteen people, they worked on the design and construction of ENIAC from 1943 to 1945. They were located at the Moore School of Electrical Engineering in the University of Pennsylvania. ENIAC's first application was in nuclear science for the Manhattan Project. It contained about 18,000 vacuum tubes, 70,000 resistors, and 1,500 relays. It consumed 160 KiloWatts of electrical power and weighted 30 tons. Its large power consumption was the result of a large number of vacuum tubes. Vacuum tubes contain a heater element that displaces electrons. The heater burns out after a few thousand hours of operation. ENIAC was running without faults for one to two days, on the average. It performed about 5,000 operations per second. It was about 1,000 times faster than Mark I and was developed as a general-purpose computer. It had only twenty memory locations that each could store a signed 10-digit decimal number.

To program the ENIAC, individual units had to be connected together with wires, and switches were set appropriately to control the sequence of operations. Mauchly met Atanasoff in 1940 and they exchanged ideas about electronic computers. As discussed earlier, a U.S. federal judge decided in 1973 that the ENIAC was based on several design concepts introduced earlier by Atanasoff. However, the 100,000 pulses per second decimal ENIAC yielded better performance than Atanasoff's 60 pulses per second ABC binary machine. The core of the ENIAC programmers consisted of six women, called "computers." It was in operation for ten years, from November 1945 until October 1955. It operated at the Moore School until 1947 in the "direct programming" rewiring mode. It was then moved to the Ballistics Research Laboratory (BRL) of the Aberdeen Proving Ground in Maryland, where it operated in the "converter code" mode. The ENIAC computer produced the correct solution to problems only 54% of the time while at BRL. Mauchly claimed that the value was 90% while the ENIAC was at the Moore School.

The designers of the ENIAC came up with several design ideas that they were not able to implement on the ENIAC. Some of these ideas were included in subsequent improvements of the ENIAC and/or new computers that were developed by new computer companies. The ENIAC forced the development of new computing methods in the numerical analysis field because of its high speed at that time and its limited memory space. The ENIAC introduced the stored program concept where programs can also be modified at run time. The ENIAC was never copied and had rather small influence on the design of later products. Eckert and Mauchly later formed the Electronic Control Company for the production of computers. This was the first step in the development of electronic computers in industry. Also, the Moore School offered in 1946 a series of 48 lectures about the design of electronic computers. Many of the lecturers were involved in the design of the ENIAC. Several people, including Maurice Wilkes from England, who later designed electronic computers attended these lectures.

John von Neumann, influenced by the success of the ENIAC, initiated new computer design concepts that became preeminent for the next 40 years. Although von Neumman gave a lecture during the aforementioned Moore School lecture series, he was not a major participant. He got involved with the ENIAC design group in 1944, after the design of the ENIAC was complete. The result was the Electronic Discrete Variable Automatic Computer (EDVAC) in 1945 that stored both programs and data. A proofreaded version of the first draft produced by von Neumman on the EDVAC on June 30, 1945 is listed in the bibliography. This draft is very incomplete and describes an architecture very different from that of the EDVAC. It was generated at the Moore School of Electrical Engineering and disappeared from the scene after it had influenced the first generation of computer engineers. Von Neumann's major contribution was the concept of conditional control transfer at program execution time. In simple terms, instructions in programs are stored in contiguous memory locations and are executed sequentially until a conditional control transfer instruction is encountered. If the specified condition is satisfied, then the execution continues at a new location stored in the conditional inctruction. Von Neumann was the first to use absolute memory addresses to access instruction operands at run time. The EDVAC was built at the Moore School and was later moved to BRL. It was a 44-bit word binary computer with 12 opcodes. It contained four non-addressable registers and consumed 50 KiloWatts. It contained more than 3,500 vacuum tubes. Von Neumann later built another version of this computer, the IAS (Institute for Advanced Studies at Princeton) computer, with the help of Herman Goldstine. Around the same time, the Whirlwind I was built at MIT (Massachusetts Institute of Technology), for real-time control. It was capable of 50,000 operations per second at 5 MHz, and introduced the magnetic core memory technology that reduced memory failures significantly.

Two pioneering computers were manufactured in England before similar systems were built in the U.S.A. The Manchester Mark I was manufactured in 1948. The EDSAC (Electronic Delay Storage Automatic Calculator) of Cambridge University was manufactured in May 1949 by a group led by Maurice Wilkes. The EDSAC was the first practical computer based on the von Neumann draft for the implementation of the stored-program concept. Wilkes wanted to implement a conservative design to test programs. The machine contained about 3,000 vacuum tubes and consumed about 12 KiloWatts of power. Sixteen steel tubes implemented acoustic mercury (ultrasonic) delay lines for the bit-serial memory. Each line was about five feet in length and stored 576 ultrasonic binary pulses. The machine operated on 17-bit fixed-point numbers (sometimes on 35-bit double-length numbers also) and had a performance of about 600 operations per second at 500 KHz. It was easy to display the value of a delay line on a CRT (cathode-ray tube) monitor. A five-hole punched paper tape was used for input and output. Opcodes (that is, instruction operation codes) and addresses were five and ten bits long, respectively. Wilkes gave emphasis to programming using mnemonics of the instructions and having the machine translate the program into binary code understandable by the hardware. A subroutine-calling method also was developed. One of the first programs was to find prime numbers. A commercial version of the latter machine was also produced, the LEO (Lyons Electronic Office). The team that designed EDSAC 1 developed EDSAC 2 in 1958. It was the first computer with a microprogrammed control unit and a parallel bit-sliced arithmetic unit; the emphasis was on a more reliable machine than EDSAC 1. Wilkes used magnetic core memory that he learned about from the Whirlwind computer during his MIT visit in 1953. Alan Turing cited the von Neumman draft in his proposal for the Pilot ACE (Automatic Computing Engine). This machine competed with the EDSAC in the problem of finding larger prime numbers.

The UNIVAC I (Universal Automatic Computer) built by Remington Rand in 1951 was the first commercial computer that implemented von Neumann's major design concepts. It used vacuum tubes for the switching elements. In general, first generation electronic computers employed vacuum tubes and magnetic drums for the storage of programs and data. Also, they were programmed mainly in machine language.

2.2. Second Generation (1955-1965)

The transistor technology that facilitates the solid state implementation of electronic components was invented in 1947 at Bell Laboratories in New Jersey by John Bardeen and Walter Brattain. They demonstrated a solid-state amplifier. Their project leader William Shockley proposed a new type of transistor that was implemented in 1950. It contained a layer of a p-type semiconductor between two layers of an n-type semiconductor. In the p-type, electrical conduction is the result of hole (or electron vacancy) transfers. In the n-type, electron transfers produce conduction. This new generation of computers flourished just because of this transistor development. The use of small transistors, instead of large vacuum tubes, reduced dramatically the size of components while their processing speed was increased tremendously. Progress in the shrinking of transistors still continues to our days. As discussed earlier, magnetic core memory was introduced in 1950 at MIT by William Papian and Jay Forrester for the Whirlwind computer. It had positive impact on the size, speed, reliability, and power consumption of computers. The first commercial computer using magnetic core memory was the IBM 705, in 1955. Computer manufacturers in the 1950s generally produced data processing machines, rather than computers to tackle mathematical problems.

The IBM 7090, the DEC PDP-1, and the Control Data Corporation (CDC) 1604 were the first computers to use transistors. The PDP-1 gave birth to the minicomputer industry; it cost much less than the 7090, yielded half of its performance, and could store 4,000 18-bit words. Minicomputers cost much less than large computers but can yield very substantial performance. Supercomputers containing many processors were introduced, such as the IBM 7030 (its project name was STRETCH) in 1955 and the Sperry-Rand LARC. They were used in atomic energy laboratories to process large amounts of data in short periods of time. The IBM 7030 was ten times faster than its earlier models that used vacuum tubes. It also introduced the concept of multiple memory banks and used heavy pipelining, with up to six instructions being active simultaneously. Assembly-language programming, operating systems, and tape storage systems were commonplace in this generation. Other computer manufacturers during this period were, among others, Honeywell, Control Data, and Burroughs. The software industry also traces its roots back to this generation of computers. The COBOL (Common Business-Oriented Language) and FORTRAN (Formula Translator) programming languages replaced binary machine code with mnemonics, words, sentences, and mathematical formulas. BASIC, an easy to learn programming language, was developed in 1964 for students at Dartmouth College.

Jack Kilby of Texas Instruments invented the integrated circuit (IC) in 1958 concurrently with Robert Noyce of Fairchild Corporation. He put less than half a dozen semiconductor transistors onto a single chip. Fairchild introduced in 1961 the first commercial integrated circuits, in the form of RTL (resistor-transistor logic). This success was followed by denser integrations over the years. The TTL (transistor-transistor logic) later improved efficiency and the fabrication process. Between 1961 and 1971, the number of transistors on commercial-off-the-shelf (COTS) chips approximately doubled every year. Since then, the microprocessor performance doubles about every eighteen months, whereas the number of transistors on a chip doubles about every two years. This improvement in transistor integration is expected to continue (probably until around 2010) based on a prediction made by Gordon Moore in 1965; this prediction is known as Moore's Law.

The CDC 6600 was introduced in 1964. It contained several highly-parallel processors (CPUs). Its designer was Seymour R. Cray. He later became very famous for designing supercomputers (as discussed below). The CDC 6600 introduced the scoreboard technique to resolve data dependencies between instructions for the dynamic scheduling of operations. It contained ten functional units and instructions were issued out of order.

The first laser was demonstrated in 1960 at Hughes Research Laboratories in the U.S.A. Several different types of lasers have appeared since then. Laser technology had very significant impact in recent years on computer communications, optical recording of information, and computer printing.

2.3. Third Generation (1965-1974)

The invention of the integrated circuit in 1959 enabled several new companies to introduce minicomputers. Also, the magnetic core memory technology was replaced by IC memories in the 1970s. The bipolar transistor that facilitates electron transfers within the body of the semiconductor was preeminent in the 1950s. The RCA Electronic Research Laboratory developed the MOSFET (metal-oxide silicon field-effect transistor) in 1962. MOS ICs increase the integration of components and electron transfers occur at the surface of the semiconductor. In 1965, Digital Equipment Corp. (DEC) introduced the PDP-8, a very successful product that dominated the minicomputer market. It had smaller size than earlier computers because of its IC chips. It was followed by the PDP-11 that had tremendous success with universities and research establishments. In 1967, Fairchild produced the 3800, an 8-bit ALU (Arithmetic-Logic Unit) chip with an on-board accumulator.

In 1969, Kenneth Thompson and Dennis Ritchie of the AT&T Bell Laboratories developed the UNIX operating system that later became popular in universities and research establishments. It supports multiple users and multitasking. It can be installed virtually on any computer because it is written in a high-level language, namely C. There are two popular versions of Unix, the Berkeley Unix and the AT&T system V. A more recent version, namely Linux, runs on many hardware platforms, including personal computers. It was developed by Linus Torvalds in the 1990s. The Advanced Research Projects Agency (ARPA) of the U.S. Department of Defense established the Arpanet data-exchange communications network that started with four universities.

The System/360 enabled IBM to dominate the large mainframe computer market. More families of IBM mainframes were developed with identical assembly languages for program compatibility purposes. The concept of multiprogramming, where several programs can reside in the machine simultaneously, was also introduced with the System/360 family.

The first advertisement by Intel for a microcomputer appeared in November 1971. It was for the MCS-4. The more versatile MCS-8 became available in April 1972. Intel had began as a company that produced memory chips. In 1971, the Intel 4004 chip included all basic 4-bit processor components on a single chip. It contained 2,300 transistors. The era of single microprocessor chips was just beginning. Microprocessors later appeared everywhere, in kitchen appliances, computer desktops, automobiles, etc. Their low cost makes them available to many people, not just businesses. Other pioneering microcomputer companies were Commodore and Apple Computers. In 1972, Intel produced the 8008 8-bit microprocessor that contained 3,300 transistors and operated at 0.5 MHz. Other popular 8-bit microprocessors of this era were the Intel 8080 and the Motorola MC6800. They were used in embedded applications. The first popular computer system was the MITS Altair 8800 that contained the 2-MHz Intel 8080. Several other hobby computers became available in the mid 1970s. The majority of these computers were sold in kits. Table I summarizes the history of the Intel microprocessors.

Table I: Brief history of the Intel microprocessors.

Microprocessor	**Year**	**Features/Comments**
4004	1971	2,300 transistors. 4-bit microprocessor. Used in the Busicom calculator.
8008	1972	8-bit microprocessor. Twice as powerful as the 4004.
8080	1974	Used in the Altair microcomputer kit that cost $395. Tens of thousands sold in months.
8086/8088	1978	Used in the first IBM PCs.
80286 or 286	1982	Software compatibility with 8086/8088. Used in about 15 million PCs worldwide by 1988.
386	1985	275,000 transistors. 32-bit processor supporting multitasking. $200 million manufacturing facility.
486	1989	Built-in math coprocessor. For color screens and mouse-based computing.
Pentium	1993	For real time data processing.
Pentium Pro	1995	5.5 million transistors. For 32-bit servers and workstations. $2 billion manufacturing facility.
Pentium II	1997	7.5 million transistors. Employs the Intel MMX (multimedia) technology for video, audio, and graphics.
Pentium II Xeon	1998	For servers and workstations that may support multiprocessing.
Celeron	1999	Used in PCs for great performance at reasonable cost.
Pentium III	1999	9.5 million transistors. Incorporates 70 new SIMD instructions for multimedia and Internet data streaming. Improved performance for 3-D visualization, imaging, and audio, video, and speech processing.
Pentium III Xeon	1999	Better performance for electronic commerce and advanced business computing. Includes the 70 SIMD instructions. Supports multiprocessor configurations.
Itanium	2000	First 64-bit processor in the new IA-64 family. Targets electronic business, visualization, multimedia, heavy computation, and multiprocessor configurations.

The Ethernet for the interconnection of computers in small local-area networks was developed by Robert Metcalfe of Xerox Corp. in the U.S.A., in cooperation with Intel and DEC. Metcalfe later established the 3Com company that marketed Ethernet for the first time in 1982.

In the last forty years, supercomputer sales have grown from a small number for specialized niche markets to relatively large numbers of machines for the government, industry, and universities. Seymour R. Cray was in charge of the development of one of the early computers that contained transistors, namely the CDC 1604. His group designed the CDC 6600 and several improved versions of it. Several universities and government laboratories purchased these machine because of their high speed. Cray later founded Cray Research, Inc., in 1972 for the manufacture of supercomputers. He believed that the best technology has to be used each time to design the fastest machine, independently of its cost. Cray's designs were based on vector data processing, where the same operation is applied simultaneously to many or all elements of a data array. Such array operations are very common in engineering applications. While all previous supercomputer-like machines used magnetic core memory, Cray used memory chips for the first time.

2.4. Fourth Generation (1974-1989)

The LSI (Large Scale Integration) technology allowed the inclusion of hundreds of transistors in a single chip. By the 1980s, VLSI (Very LSI) had put tens of thousands of transistors on a single chip. Each new generation of chips decreases the size of components, and their power consumption and cost. Also, it increases performance, efficiency, and reliability.

By the mid 1970s, IBM had produced the System/370 mainframe and had also focused on software development. IBM chose the path of design evolution rather than revolution during these years. The Apple II microcomputer in 1977 contained the MCS6502 microprocessor that was implemented in CMOS technology. It was a complete system with a CRT (cathod-ray tube), floppy drive, and keyboard. In the 1980s, video game systems, such as the Atari 2600, made microcomputers very popular.

Because of pin and transistor count limitations, the early microprocessors avoided the implementation of any form of advanced parallelism. However, 16-bit microprocessors that appeared later implemented pipelined instruction processing. Such was the MC68000 microprocessor that run at 8 MHz and had a peak performance of 2 MIPS. Finally, in the early 1980s, many new computer companies were formed (such as Apollo) to produce microprocessor-based workstations for engineering and scientific applications.

In 1981, IBM introduced the personal computer (PC) that was to be used as a desktop. Its DOS (Disk Operating System) operating system code was produced by Microsoft for IBM. DOS did not allow multitasking or access by multiple users simultaneously. IBM shipped 13,000 PCs in 1981. PC clones appeared soon thereafter as products of many newly established companies. Millions of PCs were sold by the end of 1982. The Macintosh microcomputer was introduced by Apple in 1984. It was very user friendly because a mouse (that is, a moving device that traces the cursor on the computer screen) could be used to select computer commands, instead of typing the commands on the keyboard. The breathtaking advances in integration technologies made possible the introduction of numerous microprocessors in the following years. The growth of the PC market was phenomenal.

The first software windowing system, where different tasks are allowed to occupy different parts of the computer screen, appeared in 1981 in the Xerox Start system. Apple produced its windowing system in 1984 and Microsoft in 1985. X Windows, another windowing system, was introduced in the late 1990s at MIT for graphics workstations; Unix or Linux is the assumed operating system.

Transistor integrations followed Moore's Law over the years and made possible the implementation of microprocessors with tremendous capabilities and very high complexity. David Ditzel and David Patterson proposed in 1980 a new design philosophy for microprocessors. They coined the word RISC (Reduced Instruction Set Computer) for these microprocessors. They proposed the reduction of microprocessor complexity for better performance. Existing microprocessors at the time belonged to the CISC (Complex Instruction Set Computer) class. They contended that the complexity of the latter class was so high that all instructions were experiencing significant overhead during execution. Their study showed that only four classes of instructions (namely, load, store, add, and jump) are executed about 75% of the time in application programs. Therefore, the implementation of these classes of instructions should have higher priority than the implementation of any other instructions or microprocessor features. As a result, numerous RISC microprocessors became commercially available in the following years. Their instruction throughput is about one instruction per clock cycle. However, CISC designs have survived. Actually, the majority of current microprocessors possess both RISC and CISC features.

The Ethernet for PCs became commercially available in 1982. It is a broadcast communication system that carries digital data packets among locally interconnected distributed computers. Information sharing in a network of computers enhances productivity in the work environment, and therefore the Ethernet has had a very positive impact on computing.

It took Cray's team four years to develop the Cray-1 supercomputer in 1976. The machine's cooling system was innovative, capable of dissipating high amounts of heat. The design balanced well vector processing with scalar (that is, single data) processing and data input/output. It used the first commercial ECL (emitter coupled logic) integrated circuits that switch fast because they do not saturate. Also, ECL devices can have high fan-out counts (that is, the same ECL output can connect to many inputs). Cray-1 had 12.5-ns cycle time (corresponding to a clock frequency of 80 MHz). The concept of operation chaining was introduced, where the resulting elements of a vector operation are consumed by a second operation that follows in the program. Cray started work on the Cray-2 in 1978. In the meantime, other groups in the company produced the Cray-1S and Cray-1M in 1979 and 1982, respectively. The 1S had comparable performance with the first machine, but it employed MOS (metal-oxide-semiconductor) technology to reduce the cost. Subsequent products included the Cray X-MP and Y-MP. The Cray-2 in 1985 was immersed in a fluorocarbon liquid for cooling purposes. It contained four processors and had double the memory space of the X-MP. Cray wanted to use gallium arsenide technology for the integrated circuits because of their high speed, but the technology was immature at that time. Bipolar ECL gate arrays were used and the cycle time was 4.1 ns (corresponding to a clock frequency of about 244 MHz).

Many other supercomputers also became popular. Some of them did not implement vector processing but benefited tremendously from the aggregate performance of processor collections. The Connection Machine CM-2 built by Thinking Machines, the Intel iPSC, and the nCUBE/2 are some of the most representative machines. The CM-2 contained up to 65,536 single-bit proprietary processors.

2.5. Fifth Generation (1990-present)

The integration count of one million transistors per chip was reached in late 1989 with advances in VLSI technology. Its successor, the ULSI (Ultra LSI) technology, has increased this count to the millions. We are currently in the fifth generation of microprocessor design. Each successive generation of microprocessors almost doubles the development cost of the immediately preceding generation. The current emphasis is on superscalar processing and multithreading. Many scalar operations can be implemented simultaneously on superscalar processors. Many independent threads (that is, sequences) of instructions can reside simultaneously in multithreaded processors; such a thread often becomes active instantaneously with minimal overhead. The PC market grew to 60 million units in 1995. In comparison, the microprocessor market exceeded four billion units the same year. Larger integration of components has made possible the construction of small laptop and palmtop computers. One billion flash memory chips were sold by Intel from 1988 to May 2000; they are used primarily in telephones and handheld devices. It is expected that Intel will sell another one billion flash memories in the next two years. Moore's Law predicts chips with about 350 million transistors in 2006. However, for economic reasons manufacturers often produce chips with about 20% fewer than the expected number of transistors.

IBM and Motorola independently produced recently techniques that use copper wiring on chips, instead of aluminum wiring that was the previously accepted solution. Because of less resistance with copper, higher speeds can be achieved in data transfers.

The Cray T90 vector supercomputer in 1994 included up to 32 processors. Several Japanese supercomputers competed successfully with Cray machines, such as the Fujitsu VPP500. More recently, massively-parallel supercomputers with hundreds or thousands of processors have been built, such as the Cray MPP T3D and T3E, Thinking Machines CM-5, Intel Paragon, and IBM RS/6000 SP. The IBM SP can contain 1,152 Power3 processors operating at 222 MHz. In 1997, the IBM Deep Blue supercomputer defeated the world chess champion Garry Kasparov. IBM recently announced a research program that intends to build by the year 2005 the IBM Blue Gene for gene research. It will be capable of one quadrillion operations per second and will employ more than one million primitive processors.

IBM introduced in June 2000 a hard drive with a storage capacity of a billion bytes of information that has the size of a book of matches. IBM's 3380 hard drive in the 1980s of the same capacity had the size of a refrigerator.The major markets for the new drive are digital cameras, and desktop and notebook computers. The drive can store about a 1,000 high-resolution photographs or 10 hours of high-quality digital audio music. The development of these magnetic storage systems is an important step for the incorporation of the next generation of hard drives in wearable computers.

Because of the large number and capabilities of individual PCs and workstations, the efficient interchange of information became absolutely essential in the work environment, and the era of networked PCs was born to share information, software, and memory space. Then, the Internet global web of computers was born to link computers worldwide into, potentially, a single network of information. The World-Wide Web (WWW) is a collection of hyperlinked pages distributed over the Internet via the HTTP (hyper-text transfer protocol) network protocol. WWW was invented in 1990 by Tim Berners-Lee, a physicist at CERN (the European Particle Physics Laboratory). It was invented for physicists to share their research results. It initially included only text. Graphics processes were embedded later with the WWW Mosaic page browser that was produced at the University of Illinois in Urbana-Champaign. The impact of the Internet and WWW combination has already been extremely positive. Numerous new businesses have been established that provide service through their Internet WWW pages.

Non-traditional approaches in computing are also under investigation. For example, optical systems have been proposed for the interconnection of computer chips. Computers implementing neural network approaches follow biological models where neurons are trained to respond to known inputs and then adapt on their own for unknown inputs. They have been successful in a few applications related primarily to pattern recognition. Genetic algorithms simulate the evolution of DNA sequences and are appropriate in optimization problems. However, their theory is not well understood yet. Quantum computers are expected to operate according to the laws in quantum physics. Their feasibility and applicability are still in question. Aggressive approaches are expected to produce general-purpose parallel computers with millions of processors for 1 PetaFLOPS (that is, 10^{15} floating-point operations per second) aggregate performance by the year 2010.

Bibliography

Bromley A.G. (1998). Charles Babbage's Analytical Engine, 1838. *IEEE Annals of the History of Computing*, **20**(4), 29-45. [A good effort to determine the capabilities of the Analytical Engine. The author discusses the feasibility of the design at its time.]

Chase G.C. (1980). History of Mechanical Computing Machinery. *IEEE Annals of the History of Computing* **2**(3). [A good survey of mechanical devices for computing tasks.]

De Solla Price D.J. (1959). An Ancient Greek Computer. *Scientific American*, pp. 60-67. [It contains a good description of the first known sophisticated mechanical computer. The author spent eight years studying the mechanism and was the one who determined its functionality.]

Godfrey M.D. (1993). First Draft of a Report on the EDVAC by J.V. Neumann. *IEEE Annals of the History of Computing* **15**(4), 27-75. [This article contains an improved version of the draft written by John von Neumann.]

Hill M.D., Jouppi N.P., and Sohi G.S. (Editors) (2000). *Readings in Computer Architecture*. San Francisco, California, USA: Morgan Kaufmann Publishers. [This book contains a collection of very important readings in the architecture of electronic computers.]

Malone M. (1995). *The Microprocessor: A Biography*. New York, USA: Springer-Verlag. [It contains a good history of microprocessor developments.]

Moore G.E. (1965). Cramming More Components Onto Integrated Circuits. *Electronics*, pp. 114-117. [It describes Moore's Law for the integration of silicon transistors on chips.]

Wilkes M.V. (1992). EDSAC 2. *IEEE Annals of the History of Computing* **14**(4) 49-56. [The author describes the EDSAC 2 that was developed under his leadership.]

Author Biography

Dr. Sotirios G. Ziavras received in 1984 the Diploma in Electrical Engineering from the National Technical University, Athens, Greece. He received in 1985 the M.Sc. in Electrical and Computer Engineering from Ohio University, Athens, Ohio, USA. He was awarded in 1990 the Ph.D. in Computer Science from George Washington University, Washington, D.C., USA, where he was a Distinguished Graduate Teaching Assistant and a Graduate Research Assistant. From 1988 to 1989, he was also with the Center for Automation Research at the University of Maryland, College Park, Maryland, USA. He was a Visiting Assistant Professor in the Electrical and Computer Engineering Department at George Mason University, Fairfax, Virginia, USA, in Spring 1990. He is currently an Associate Professor of Electrical and Computer Engineering, and Computer and Information Science at New Jersey Institute of Technology, Newark, New Jersey, USA. His major research interests are processor and computer systems designs, and parallel computing systems and algorithms.