# The History and Future of Computing Technology

# The History and Future of Computing Technology

By
KIKI BEUMER AND YUDITA JOSEPH,
BSc. Data Science by Design

International Bachelor in Data Science
in

CY Cergy Paris Université
Saint-Germain-en-Laye

International Bachelor in Data Science 2023
CY Cergy Paris Université
Saint-Germain-en-Laye

TITLE: The History and Future of Computing Technology

AUTHORS:
Kiki Beumer and Yudita Joseph,
First year's International Bachelor in Data Science

SUPERVISOR:
Fabien Piguets
Physics,
CY Cergy Paris Université

# Abstract

The main analyzed objective in this paper is the five generations in which the history of the computer can be divided. The development of computers has progressed through several generations, each marked by major technological advancements and changes in the design of computer hardware and software systems.

The first generation of computers (1940s-1950s) used vacuum tubes for circuitry, magnetic drums for memory, and were often used for scientific and military purposes.

The second generation of computers (1950s-1960s) introduced the use of transistors and made computers smaller and more reliable. The first high-level programming languages were also developed during this time.

The third generation of computers (1960s-1970s) marked the introduction of integrated circuits, which further miniaturized the size of computer systems. The first computer mouse and graphic user interface were also developed during this period.

The fourth generation of computers (1970s-1980s) saw the widespread use of microprocessors and the development of personal computers and local area networks. The introduction of multiple commercially available computers, along with the development of the Internet, had a significant impact on the way people live and work.

The fifth generation of computers (1980s-1990s) focused on developing artificial intelligence and natural language processing capabilities. These computers utilized parallel processing and distributed computing and featured advanced hardware and software systems.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Our first idea for a topic was to pick an invention of the past that has played a major role in contributing to the computer we know nowadays. We came however to the conclusion that all machines are based on the previous one, that are in turn a source of inspiration for the next machine yet to be invented.

## 1.1 Why we took this approach

We would like to consider the five generations, the history of the computer is divided in, because it helps us distinguish the advancement in a more convenient way. We had not heard about this approach before we started this paper, so we thought it might be more interesting than summarizing a timeline without any guidelines. This approach also helped us create a clearer way of communicating our answers to the questions we would like to have answered after this paper.

## 1.2 What we will be focusing on

We will be focusing on the development of computers especially in the way they perform mathematical computations. Each chapter will focus on the development of machines in each generation, their key features, how machines were developed in those periods, their working and how they have improved through each phase. After that, we will discuss the future predictions we have for the sixth generation and we would also like to give the answers that we found on our questions.

## 1.3 Our aim of this paper

We are mostly interested in the development of the computer per generation. We wanted to go deeper into the physical aspect of how certain components of the computer work. We however, also felt it was important to measure how this affected the impact on the performance of the computer. So besides the physical workings of the computers, we would like to have an answer to the question; how exactly did the computer improve over the years?

# Chapter 2

# First Generation

## 2.1 Introduction

The first generation computers has been recorded as a period from 1940-1956, but we would like to go even further back in time. In the eighteen-hundreds was the first machine build that implemented binary coding ideas, in the form of punched cards, and vacuum tubes. Therefore, we think we should include this in the first generation. There is no need to explain that this generation made a huge contribution to the computers we use today. The constructions of the inventions of the first generation are not yet representative of the computers from the twentieth century. Nevertheless, is the foundation, on which the upcoming generations are build, philosophized about in this era.

## 2.2 First generation's machines

We want to give a special attention to the fundamental discoveries that were made [7], because we feel this describes the first generation in the best way possible. We would like to go over the operations of these machines and ideas, which stood out for us.

### 2.2.1 Jaquard Loom

The Jacquard machine, designed by Joseph Marie Jacquard, is a loom machine that automates how patterns are woven. The machine was fed with a chain of punched cards that control the patterns to be woven. The use of replaceable punched cards is considered to be an important contribution for the modern computer technology.

### 2.2.2 Analytical engine

The Analytical engine, designed by Charles Babbage, is a mechanical computer said to be the first general purpose computer that incorporated an arithmetic logic unit, conditional branching, loops and integrated memory. It is moreover also Turing complete, which we will explain further in the next paragraph. It has a memory that can store up to a thousand numbers of forty decimal digits each. An arithmetic unit called The Mill would be able to perform all four arithmetic operations, comparisons and square roots. It was the successor to Babbage's difference engine[13], which is a mechanical calculator that contains toothed wheels turned 9 to 0 that helped to tabulate the polynomial functions. The input for the analytical engine was provided using punched cards that contains programs (formulas) and data. The output was obtained from the printer, curve plotter and also as punch cards.

### 2.2.3 Turing Machine

The Turing machine, invented by Alan Turing, is an abstract machine of a central processing unit. The CPU receives the input as in form of formulas or data and it processes and executes the program. The Turing machine consists of an infinitely long tape that is separated in many cells. Each cell containing 0s or 1s. A set of instructions is given to the machine. According to the cell's value and the given instruction, the cell's value or the position will change. Compared to the previous development in computer technology, Turing's machine was more efficient as it was able to execute any operation or program.

### 2.2.4 ENIAC

So far, the machines we have seen were designed so that it can only compute or execute one specific operation or program. The ENIAC however, designed by John Mauchly and J. Presper Eckert, had the possibility to execute any operation or program. It was able to solve a large class of numerical problems and was thus said to be Turing complete. The machine was re-programmable. Figure 2.1 represents the instruction for manually changing each and every single one of the dimmers to adjust exactly the right amount of voltage that were to go through the vacuum tube. Plug-boards were used for communicating instructions to the machine. It contained twenty accumulators that was able to add and subtract. It also acts as a memory which could hold up till a ten digit decimal number. ENIAC used ten-position ring counters to store digit. Each digit required 36 vacuum tubes, 10 of which were the dual Triodes making up the flip-flops of the ring counter. The arithmetic was performed by counting pulses with the ring counters and generating carry pulses if the counter wrapped around. The idea being to electronically emulate the operation of the digit wheels of a mechanical adding machine.

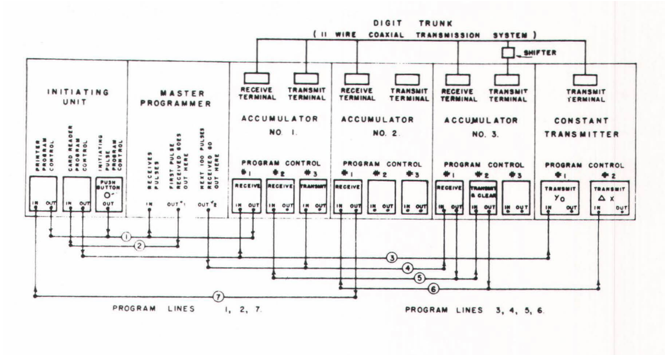


FIGURE 2.1: Instructions for programming the ENIAC

### 2.2.5 Von Neumann architecture

The central processing unit in the modern computers are based on the Von Neumann architecture that was mentioned in the first draft of a report on the EDVAC by John Von Neumann. It describes an electronic digital computer with a processing unit with both an arithmetic logic unit and processor registers, a control unit that includes an instruction register and a program counter, a memory that stores data and instructions, an external mass storage with input and output mechanisms. This eliminated the use of punched cards and plugboards. The given input is stored in a memory, the input becomes the command and is ready to process, the processor fetch the command from memory. The processor consists of control unit and arithmetic logic unit. The control unit manages the input and output command while the aritmetic logic unit performs mathematical operations. After the processing is done, the control unit stores the result in memory and displays the output.
The input comes through a device like a keyboard or a mouse. The data are stored in a memory called RAM. The output is displayed through the monitor or printer. Important to note here is that the Von Neumaan architecture was a very important philosophy, but the idea had not yet been put into practice.

## 2.3 Vacuum Tube

A vacuum tube is a device that controls electric current flow in a high vacuum between electrodes to which an electric potential has been applied. Two kinds of vacuum tubes were widely used; the diode and the triode. Both have the same principle of operation. The flow of electrons is always accelerated from cathode to anode. They were crucial to the development of analog and early digital computers. The computers of first generation used vacuum tubes as the basic components for memory and circuitry for the Central Processing Unit. These tubes produced a lot of heat which caused the installations to fuse frequently. Therefore, they were very expensive and only large organizations were able to afford it. The diodes and vacuum tubes were used as logic gates. As we saw in figure 2.1, there was a very specific way in which the machine had to be programmed. The first generation computers used machine language, 0s and 1s. This made programming troublesome.

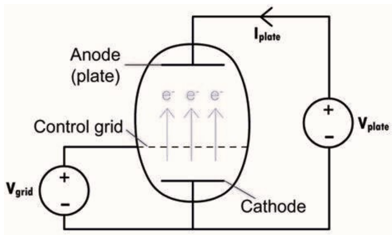


FIGURE 2.2: Schematic representation of vacuum diode

The diode contained two electrodes, the anode and cathode, in an evacuated glass container as in Figure 2.2. When voltage is applied to the cathode, the diodes heats up and release electrons. If the anode has slightly higher positive potential, then the electrons are attracted to the anode, which results in current flow. The potential difference in the anode and the cathode is controlled by the dimmer which is set manually. Triodes are similar to diodes, but they use a third electrode called the grid. This is placed in the middle of the evacuated glass between cathode and anode. While previously the dimmer adjusted the voltage, the grid now replaced this function. The dimmer and grid can be adjusted, which results in attraction or repletion of electrons.

## 2.4 Drum Memory Storage

A drum memory or drum storage unit was made of a large metal cylinder, coated on the outside surface with a ferromagnetic recording material[5]. In most designs, one or more rows of fixed read-write heads ran along the long axis of the drum, one for each track. The drum's controller task was to select the right head and then wait for the appearance of data beneath it as the drum rotated. At the same time, stationary write heads emitted an electrical pulse to alter the magnetic orientation of a particle at a specific position on the drum. Magnetic drum units used as primary memory were addressed by word. In the era when drums, were used as main working memory, programmers often did optimum programming. The drum memory storage led to invention of the hard disk drive.

## 2.5 Contribution to progress in technology

We would like to take a look at the improvement of the computers compared to the previous generation. Since this generation as we observed it, is more than one-hundred years long, we thought we could compare the start of the generation with the end. This is still fairly obvious, but still interesting to observe. As a result of the punch cards, the memory drum and the vacuum tubes, it was possible to improve the construction of the machines. The ENIAC's construction may look like a step back from the Analytical Engine, however the new technology with that plays a considerable role in that. The energy needed for a vacuum tube varies depending on its size and what it is designed to do. A diode can create a waste heat between a watt to a hundred watts in the process of passing anywhere from tens of milliamps across a hundred volts to hundreds of milliamps across 500 volts or more.[17] The vacuum tubes were extremely expensive to use. So, despite the increase of performance compared to the start of the first generation, the cost of the machines did not decrease.

# Chapter 3

# Second Generation

## 3.1 Introduction

At the start of the second generation, the framework of the computer was established. We can observe that in the next generations, the main focus will be to improve this already existing framework. As a result of the transition from vacuum tubes to transistors, we speak of a generation change after 1956 until 1963.

## 3.2 Second generation's machines

We can observe that in the second generation of computers, the construction mostly changed, because of the development of the transistor. There are multiple companies that built computers using transistors. We wanted to highlight only what we found most interesting.

### 3.2.1 TX-0

In 1958 MIT researchers built the transistor experimental - 0, also called The TX-0, which is the first general-purpose programmable computer built with transistors. This computer had previously consisted of vacuum tubes, which were now being replaced by transistor circuits. To make the replacement the easiest, the transistor circuits were first places inside bottles and then replaced by the vacuum tubes. This made the computer less costly, but defeated the purpose of building compact computers.

### 3.2.2 IBM's 7030

In 1961 the IBM's 7030, also known as "Stretch" was completed. IBM´s 7000 series of computers are the company´s first to use transistors. Similar to the first generation computers, we notice that the first successfully built computers are for scientific use. As it gets easier, faster and profitable to manufacture computers on a larger scale, the mass audience is also able to make use of the inventions.
By the mid-1960s, almost half of all computers in the world were IBM 1401s computers. Demand called for more than 12,000 of the 1401 computers.This success made a strong case for working with general-purpose computers rather than specialized machines.

## 3.3 The Transistor

The first transistor was already invented in 1837 by George Stephenson. The transistor actually came off the ground when the second transistor was invented by John Bardeen, William Shockley, and Walter Brattain. They were scientists at the Bell Telephone Laboratories in Murray Hill, New Jersey. Bardeen, Shockley and Brattain attempted to replace vacuum tubes by semiconductors by researching the behavior of germanium crystals.
Germanium is a chemical element of atomic number 32. It is a hard, brittle, silvery-white semi-metal. Germanium was important in the making of transistors and other semiconductor devices, but has been largely replaced by silicon. We can see in figure 3.1 that germanium has valance electrons that are at a higher energy level (fourth orbit) as compared to silicon material (third orbit). [18] This means that germanium material requires some additional energy to free electrons from the last orbit, which makes germanium unstable at temperatures higher than 700 degrees Celsius. This property makes silicon more suitable under circumstances with a temperature higher than 700C, since silicon works to 1500 degrees Celsius. The variation of the collector leakage current is moreover, between ten and a hundred times greater than that of silicon.

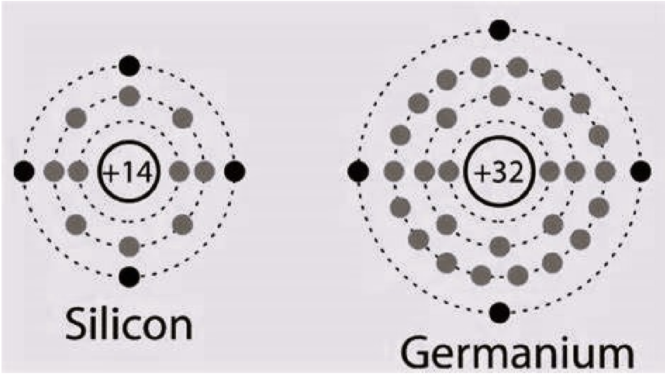


FIGURE 3.1: Structure of Germanium compared to Silicon

Doping is a verb used to describe the process of adding impurities to a semiconductor, in this case germanium or silicon, and with that changing its electrical properties. When treating it with antimony, the semiconductor will get electron admitting, also called N type. When treating germanium with boron, the semiconductor will get electron absorbing; P type. [8] The depletion region, zone or layer is an insulating region within a conductive, doped semiconductor material where the mobile charge carriers have been diffused away, or have been forced away by an electric field. The semiconductor is a solid substance that is not a good insulator and not a good conductor, so it has a conductivity between that of an insulator and that of most metals. In the case of silicon that is undoped, each silicon atom will bond to four other silicon atoms around it. This creates a rigid covalent bond, which keeps all the electrons immobile and are thus unable to contribute anything to the total electricity. This shows that when silicon had not been doped, it has a very low conductivity. The conductivity is given by:

$$\sigma = n * \mu * e$$

with :
$\sigma$ = conductivity
$n$ = number of free electrons
$\mu$ = mobility
$e$ = charge of one electron

To increase the conductivity of silicon, we can replace one silicon atom with a dopant, such as a phosphorus atom or a boron atom. If we take phosphorus for this example, which has five valence electrons. When the atom is thus placed in between the silicon atoms, it will form four covalent bond with four of the silicon atoms surrounding the phosphorus atom. The fifth electron is not bonded and is therefore free to move around. The negative charge will then leave and then the phosphorus atom is now positively charged, also P-type. From this we can conclude that the conductivity increases as the number of dopants added to the silicon or germanium increases. Boron is often used as dopant as well, since it has three valence electrons, so one less than the silicon atoms. The boron atom will then be negatively charged or N-type.

When 'sandwiching' these N- and P-types together, either in the order NPN or PNP, a transistor is created. The transistor will only produce the switching effect from 0 to 1, when it conducts electricity and the circuit is completed. As a consequence of doping, the transistor only conducts this electricity when a voltage exceeding a certain threshold is applied.
The transistor or transfer-resister is a component found in a variety of electrical circuits. This concept of 0's and 1's as the only possible data in a computer, instead of numbers from 0 to 9 as in the first generation, can be used for logical operations. A logic gate is a device implementing a boolean function, a logical operation performed on one or more binary inputs that produces a single binary output. The three essential logical gates being AND, OR, NOT. From these three basic logical operations we can create new useful logical gates such as NOR; not or, XOR; exclusive or, NAND; not and. In chapter 5 we will look at how these values are actually obtained.

## 3.4 Memory storage

Not as prominently mentioned when talking about the second generation of computer, however the changes in memory storage resulted in the decrease of size of the construction of the computer.

### 3.4.1 Magnetic core memory

Magnetic core memory [21] is the primary memory, which means that this memory is incorporated in the architecture of the computer. This makes the magnetic memory core directly accessible by the processor.
'The most common form of core memory, X/Y line coincident-current, used for the

main memory of a computer, consists of a large number of small toroidal ferrimagnetic ceramic ferrites held together in a grid structure with wires woven through the holes in the cores' centers.' [19]. Magnetic core memory uses rings, named toroides of a magnetic material called transformer cores. Each wire threaded through the core serves as a, so called, transformer winding. The function of the transformer is to convert the voltage level of electrical power. Assuming that, one winding is connected at the high voltage side and the other winding is connected at a low voltage side. Two or more wires can pass through every core. Through each core there should at least go one XY wire and one read and inhibit wire. Electric current pulses in some of the wires through a core. The other, sense wire, is used to detect whether the core changed state or not. This wire remembers the state that it has detected afterwards. When no information is being retrieved from the memory, the cores maintain the last value they had, even when the computer is not running. This is a type of non-volatile memory. Each core is able to store one bit of information. The core holds the value one or zero according to the direction of the core's magnetization, which is either clockwise or counter-clockwise.
This system was the first reliable high-speed random access memory for computers. Magnetic core memory was widely used as the main memory technology for computers well until around 1975. This means that this is the only component of the architecture of a computer, that has been used during the first, second, third and fourth generation of computers. Finally in the fourth generation the magnetic core memory was replaced by RAM and DRAM, but we will elaborate more on that in the fifth chapter.

### 3.4.2 Magnetic tapes and disks

Secondary memory is external from the computer's architecture. We have two types of secondary memory in the second generation; the magnetic tapes and the magnetic disks. The surface of a magnetic tape and of a magnetic disk are covered with a magnetic material, which makes storing the information magnetically possible. Both are, just as the magnetic core memory, non-volatile storage types. That is to say that the memory does not need constant power in order to retain data. Since, besides their appearance, the concepts of these two seemed pretty similar, we were curious to see why there were two invented and why you would prefer one over the other. We were able to create an overview of the differences

between the magnetic tape memory and the magnetic disk memory. [19]
We concluded that the Magnetic Disk Memory is more suitable for more complex and frequent data storing and retrieving. The Magnetic tape is less suitable in this case, but more fitting for 'household' data storing and retrieving, since it is also cheaper.

| | Magnetic Tape Memory | Magnetic Disk Memory |
|---|---|---|
| Material | Plastic ribbon as main component | Metallic or plastic circular disk coated with magnetic oxide |
| Cost | Less costly while mostly plastic | More costly while it consists of more expensive materials |
| Reliability | Chance of breaking down is higher | Chance of breaking down is lower |
| Speed | Data retrieval is slower | Data retrieval is faster |

Table 3.1: An overview on the differences between Magnetic Tape Memory and Magnetic Disk Memory

## 3.5 Contribution to progress in technology

It is obvious that the transition from the vacuum tubes to the transistors, has made a huge impact on the speed, performance and cost of the computer. Transistors use significantly less power than vacuum tubes. 'That means a small signal with almost no energy can control a transistor to create a much stronger copy of that signal in the collector-emitter part of the transistor. Thereby, the transistor can amplify small signals.' [8] We can notice, that there has been a shift in the philosophy of building a general-purpose programmable computer. Computers in the first generation were mainly focused on performing mathematical operations. These new machines went beyond that. No longer were numbers manipulated mechanically, with punched cards and energy wasting vacuum tubes. Nevertheless were these computers still to costly for commercial use. All transistors had to be assembled manually. This was very time consuming and expensive, however still considerably less than the vacuum tubes.

# Chapter 4

# Third Generation

## 4.1 Introduction

The third generation computers that were developed around 1964 to 1971 revolutionized computing, as it was possible to create smaller, cheaper computers that were faster than pre-microchip era computers. Third generation computers start using integrated circuits instead of transistors. [16] The punch cards were replaced by mouse and keyboard. Also, multi-programming operating systems, time-sharing, and remote processing were introduced in this generation. The storage device for these computers were magnetic core and magnetic disks.

## 4.2 Third generation's machines

Third generation computers were developed around 1964 to 1971 revolutionized computing, as it was possible to create smaller, cheaper computers that were multitudes faster than pre-microchip era computers. Third generation computers start using integrated circuits instead of transistors. The punch cards were replaced by mouse and keyboard. Also, multi-programming operating systems, time-sharing, and remote processing were introduced in this generation. The storage device for these computers were still magnetic core and magnetic disks.

## 4.3 IBM 360 Series

The most distinctive machine of the third generation was the IBM-360 series, which was more powerful and the fastest available at that time. This machine was designed by IBM to manage high-speed data processing for scientific uses such as global weather forecasting, theoretical astronomy, space exploration and subatomic physics. It contained custom hybrid integrated circuits, worked with 32-bit words, and used basic assembly language (BAL). This introduced IBM's Solid Logic Technology, which packed more transistors onto a circuit board. Thus, allowing more powerful, but smaller computers to be built. Its succeeding models had ultimately a huge impact on the computer technology in terms of workload and power.

### 4.3.1 Fortran

The programming language in third generation were machine independent and more programmer friendly than the previously used machine languages and assembly languages. This language enables a programmer to write in the same language regardless of the type of computer they work with. FORTRAN is an example of such a high level language. They were closer to the human language than machine languages. FORTRAN was the first computer programming language that was widely used.

## 4.4 Integrated circuit

Around this time, scientists and engineers were curious to see how they could reduce the size of the transistor, which led to fit an entire circuit onto a single piece of silicon. This is now known as the integrated circuit or microchip.
The integrated circuit (IC) is a semiconductor material, that contains thousands of miniaturized transistors in it. A single IC has many transistors, resistors, and capacitors along with the associated circuitry. Due to the integration of ICs in these computers, they can execute high-level tasks such as signal processing, sophisticated digital calculations. The integrated circuit uses silicon, because it is a semiconductor, which means that we can alter its conductivity by adding impurities or doping, similar to the transistor. It allows to control the electrical signals that can pass through the circuit. More transistors on microchips allow them to

send out more instructions which increases the computational power.
At the start of 1960s, an integrated circuit rarely contained more than 5 transistors as there was not enough space on the chip. But by the mid-1960s, there were ICs with over 100 transistors. In 1965, Gordon Moore said that approximately every two years, because of the advances in materials and manufacturing, it is possible to fit twice the number of transistors into the same amount of space, which is widely known as Moore's law. In 1970s, very large scale integration software is being used to automatically generate chip designs instead of manually laying out the design.

## 4.5 The electronic keyboard

The history of the modern computer keyboard begins with the invention of the typewriter. Christopher Sholes patented the first practical typewriter. Later, in 1878 he and his partner, James Densmore patented the Qwerty keyboard. They designed this particular layout as it would overcome the physical limitations of mechanical technology at the time. Separating frequently used combinations of letters minimized the jamming of the mechanism. Pressing a key would push a metal hammer that rose up in an arc, striking an inked ribbon to make a mark on a paper. After that it would return to its original position. [4]
New keyboard models were introduced in 1930 that combined this input and printing technology with the communications technology of the transmitting messages from a distance along a wire; telegraph. The ENIAC computer already used keypunches, which is the punch-card systems combined with typewriters. In 1948, the Binac computer used an electro-mechanically controlled typewriter to input data onto a magnetic tape to input computer data and print results.
Numerous attempts were made by different inventors to create the most efficient and progressive input method. Figures 4.1, 4.2 and 4.3 show some of these attempts.[6] With figure 4.1 being a keyboard glove designed in 1960, [1] that detects motion, 4.2 and 4.3 two other layout ideas for the keyboard. [2] [3] The problem with all of these alternative keyboard technologies is the data capture takes more memory and is less accurate than with digital keyboards, so the keyboard persisted.

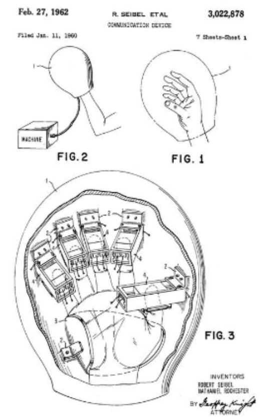


FIGURE 4.1: Keyboard "glove", 1960

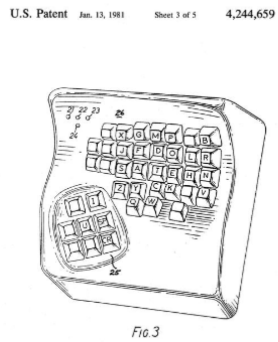


FIGURE 4.2: Maltron keyboard arrangement, 1981

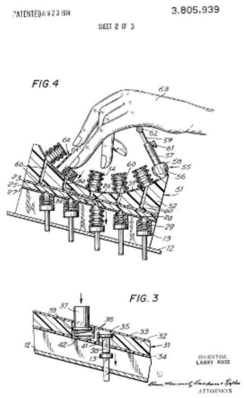

FIGURE 4.3: Ross keyboard acturator, 1971

By 1964, a multi-user computer system called Multics, was created, which incorporated the technology of the cathode ray tube used in televisions into the design of the electric typewriter. A cathode ray tube is a vacuum tube that displays visuals when its phosphorescent surface is hit by electron beams and scanned by a scanning device. We will not go into phosphorescence to much, since that is not our main focus of this subject. Cathode ray tubes can however be either monochrome or color and are offered in a wide range of display modes. This made it possible for computer users to see what text characters they were typing on their display screens. This allowed easier text creation, editing, and deleting. By the late 1970s and early 1980s, all computers used electronic keyboards and monitors.

## 4.6 Contribution to progress in technology

The integrated circuit, compared to a discrete circuit, is smaller, faster and inexpensive. The IC is smaller and cheaper, because it is created as a whole, whereas the discrete circuit requires manually assembling all components. Due to their smaller size, the IC performs better and works faster. The integrated circuits, allowed the computer to let different applications run simultaneously. Keyboards and monitors could be used as input and output, which contributed to the simplification of programming. This made code development efficient, faster and easier, enabling more productive work.

# Chapter 5

# Fourth Generation

## 5.1 Introduction

1972-1980 is the generation of the personal computer, the world wide web and the supercomputers. Performance, speed and versatility are important factors in the improvement of the fourth generation's computers, as a result of the microprocessor.

## 5.2 The first microprocessor

The development of the integrated circuit in the third generation, resulted into the first microprocessor. There is a lot of debate over who deserves the credit for inventing the microprocessor. We do know that one of the first attempts of a transistorized computer, with the concept of a CPU already in mind, was the Manchester University TC that appeared in 1953. 'Ed Sack and other Westinghouse engineers described an early attempt to integrate multiple transistors on a silicon chip to fulfill the major functions of a CPU in a 1964 IEEE paper.' [9] Despite the disagreement over who invented the microprocessor, it is clear that one of the first commercially available microprocessors was the Intel 4004 in 1971, designed by Federico Faggin, Marcian Hoff, Stanley Mazor, and Masatoshi Shima. It was designer for a calculator company and the microprocessor was used as a calculator chip. It had 2250 transistors and could perform up to 90,000 operations per second in four-bit chunks. A 4 bit processor means that it can only read, process and execute 4 bits at a time. After this many other engineers started creating a microprocessor that is faster, more efficient en smaller than the one before them. The increase in capacity of microprocessors satisfies the infamous Moore's law, but with the present technology, it is actually every two years.

### 5.2.1 The Cray-1

The fastest machine of its day, The Cray-1's [14] speed comes partly from its unique shape. It's shape reduces the length of wires and with that the time signals need to travel across these wires. The ratio of integrated circuits to the volume of The Cray-1 and a cooling system contribute also to its speed. Multiple Cray-1's were built which each took a full year to assemble and cost around 10 million dollars.

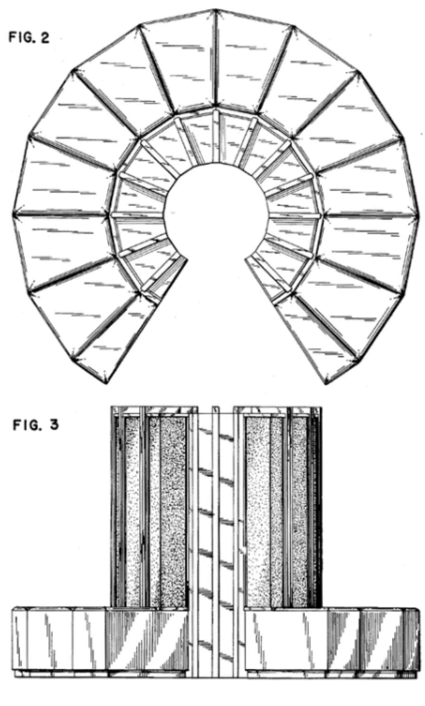


FIGURE 5.1: Clay-1 supercomputer

## 5.3 Architecture of a microprocessor

A CPU is said to be the brain of the computer.[10] It is an integrated electronic circuit that performs the calculations that run a computer. The CPU contains the arithmetic, logic, and control circuitry required to perform the functions of a computer's central processing unit. An operation is performed first by reading the information from the memory, processing this information in the ALU or Arithmetic logic unit. To complete the process, the resulting information is being transferred from the registers to the main memory.

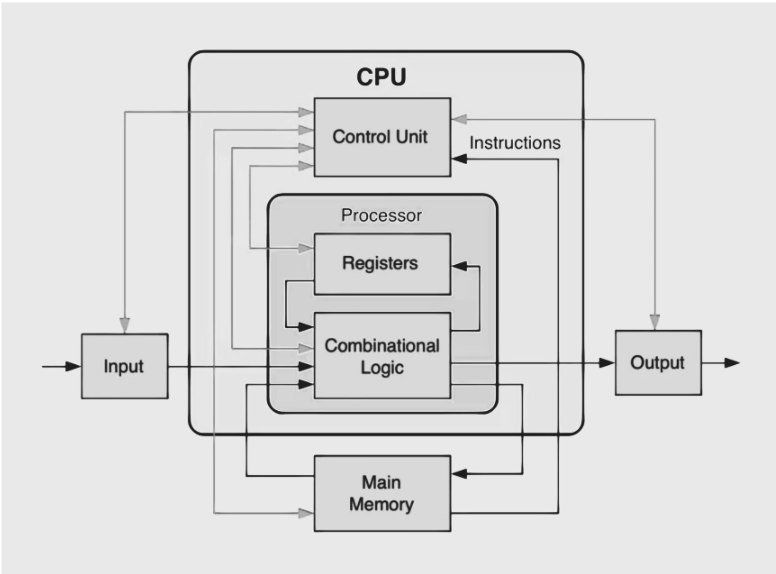


FIGURE 5.2: Diagram of a Microprocessor

The CPU of a microprocessor contains the arithmetic and logic unit with registers and the control unit. The main function of a control unit (CU) is to first fetch the data from the main memory, then determine the devices that are involved in this operation. These devices could be the arithmetic and logic unit (ALU), the input, output and the main memory. The CU then produces control signals to those devices, so they start to execute the operations. The CU lets them know how to respond to the given instructions. We can divide this process into to separate actions, which are the interpretation of information and the instruction sequencing. [11]
'The functions of the control unit are as follows :
• It helps the computer system in the process of carrying out the stored program instructions.
• It interacts with both the main memory and arithmetic logic unit.
• It performs arithmetic or logical operations.

- It coordinates with all the activities related to the other units and the peripherals.'

The arithmetic and logic unit in turn, performs operations on the data that it has received from the CU. Its functions is to take this data in the form of the binary digits 1 and 0. It then executes the required operation on this data, with that create new data. To finish the cycle, it then stores the new obtained data and/ or give this data to the output device. There are registers places beside the ALU to improve speed. When temporarily storing data that the ALU might need, performing operations on the data goes much faster. The Intel 4004 could perform operations in four-bit chunks per second. This means that the CPU can process an operation of maximum four bits.

| A | B | C | Result | Logical Equivalent |
|---|---|---|---|---|
| F | F | F | F | $\neg$ A $\wedge$ $\neg$ B $\wedge$ $\neg$ C |
| F | F | T | F | $\neg$ A $\wedge$ $\neg$ B $\wedge$ C |
| F | T | F | F | $\neg$ A $\wedge$ B $\wedge$ $\neg$ C |
| F | T | T | F | $\neg$ A $\wedge$ B $\wedge$ C |
| T | F | F | F | A $\wedge$ $\neg$ B $\wedge$ $\neg$ C |
| T | F | T | F | A $\wedge$ $\neg$ B $\wedge$ C |
| T | T | F | F | A $\wedge$ B $\wedge$ $\neg$ C |
| T | T | T | T | A $\wedge$ B $\wedge$ C |

Table 5.1: Example Truth table used for ALU

We have already seen the main concept of the arithmetic and logic unit in the third generation, since it is constructed by numerous transistors. Table 5.1 [20] gives an example of combinations and their equivalent output.

Since the ALU performs operations on data it does not only have direct access to the registers, but also to the main memory. [12] The memory has all the data and applications stored that the ALU might need to perform an operation. RAM or random access memory is the main memory that is easily read by the microprocessor as a result of the access to any address in the memory. Data does not need to be accessed in a sequential order, which increases the speed and performance of the computer, since the data is easily accessible by the microprocessor.

Each binary digit is stored in a memory cell in the memory. The memory cell can switch between to states that represent 0 and 1. Files and programs consist of millions of these bits. Computers have short term memory for immediate tasks and long term memory for more permanent storage; the RAM. The most common type of RAM is DRAM, there each memory cell consists of a transistor and a capacitor that store electrical charges. This requires periodic recharging, since the value is lost after time. To increase speed even more, inside the CPU there is also SRAM which is a smaller memory, which is the fastest computer memory. However, they are also the most expensive memory, because of consideration of the cost of the computer, RAM is not replaced by SRAM.

## 5.4 Contribution to progress in technology

The microprocessor made it possible to put together very well-functioning components in a compact way. These new breakthroughs allow us in the fifth generation to make great strides in implementing commonly known programming languages. As time passes, most people are getting more familiar with how a computer operates, since microprocessor make the concept more accesible. This creates more manpower and helps contribute to the next developments, which we will discuss in chapter 6 and 7.

# Chapter 6

# Fifth Generation

## 6.1 Introduction

The fifth generation computers are still in development. The objective of this generation is to develop computers that can sense, reason and interact to the surroundings. The computers developed in this generation is based on artificial intelligence, parallel processing hardware and microelectronic technology with high computing power.

## 6.2 Fifth generation's machines

Some of the fifth-generation computers are PARAM 10000, IBM notebooks, Intel P4, Laptops, etc. The fifth generation programming language such as OPS5 and Mercury, thay were based on problem solving using constraints given to the program, rather than using algorithm return by a programmer. The objective was to design to make the computer solve a given problem without the programmer.

## 6.3 VLSI TO ULSI

Very large scale integration (VLSI) is the process of creating an integrated circuit by combining thousands of transistors into a single chip. VLSI began in 1970s when the complex semiconductors and communication technologies were being developed. The microprocessor that we have seen in the fourth generation is a VLSI device. In the fifth generation the VLSI technology became ULSI (Ultra large scale integration), it is a process of embedding millions of transistors on a single silicon semiconductors microchip. It helped in increasing the computational power. The ULSI technology resulted in development of microprocessor chip with ten million electronic components in them. Intel 486 and the Pentium series of processors were built based on the ULSI technology.

## 6.4 Parallel processing

Parallel processing is a method of running two or more processors (CPUs) to handle separate parts of an overall task. Breaking up different parts of a task among multiple processors helps in reducing the amount of time to run a program. For example, large problems can be divided into smaller ones, which can then be solved at the same time. The computers we use nowadays are based on parallel processing, here the instructions are executed in parallel manner and are much faster than using single processor known as serial processing.

## 6.5 Artificial intelligence

Artificial intelligence (AI), we have been hearing this word quite often nowadays, this has been a key idea in the development of the fifth generation. AI includes robotics, neural networking, game playing, and natural language processing (NLP). AI has really taken a big turn in the development of computers, eliminating the need for humans, devices based on AI are built in a way that they can compute or make decisions, just from the previous data provided to them. Let's look at the development of AI. The research regarding AI started in the 1950s, initially it was just problem solving and symbolic methods, later the US Department of Defense took interest and began training computers to mimic basic human reasoning. For example, DARPA (Defense Advanced Research Projects Agency) produced an intelligent personal assistant in 2003 way before Siri or Alexa were household names. This early work paved the way for decision support systems and smart search systems in computers. In the 1950s, work regarding neural networks was started. Neural networks are computing systems with interconnected nodes similar to neurons in the human brain. Using algorithms, they can recognize hidden patterns in raw data, cluster and classify them. The initial objective was to create a computational system that could solve problems like a human brain. Then later it was developed to support diverse tasks, including computer vision, speech recognition, machine translation, social network filtering, playing board and video games and medical diagnosis. Neural networks have improved the decision processes in few areas such as in fraud detection, medical diagnosis, transportation networks, financial predictions, robotic control systems. In the 1980s, machine learning became famous. It is a method of data analysis. It is based on the idea that the system can learn from various data, identify patterns and make decisions with minimal human intervention. Because of exposing new data to the devices, they are able to independently adapt themselves. This gave way to self-driving cars, online recommendation, and fraud detection. Now, deep learning is under development, it is a subset of machine learning that trains a computer to perform human-like tasks, such as speech recognition, image identification and future predictions. Siri and Cortana are developed using the ideology of deep learning. AI holds the key to unlock a magnificent future that is driven by data and computers that understands our world. In the future, the computers based on AI will understand not just how to turn on the switches but why the switches need to be turned on.

## 6.6 Contribution to progress in technology

The fifth generation computers performs, compared to all other generations, the best. It has the highest speed, the smallest size and the highest memory capacity. In addition, are they the cheapest to produce and are the computers able to use more than one CPU. Artificial intelligence is currently being used in devices, allowing millions of tasks to be completed in a fraction of a second on a gadget, and employs devices that are capable of thinking for them.

# Chapter 7

# Sixth Genertation

## 7.1 Introduction

We thought it would be interesting to make a prediction for the sixth generation. We soon came across quantum computing.

## 7.2 Quantum computing

Quantum computers uses qubits that can be set one of two values (1s and 0s). A qubit can be any two level quantum system that is spin and magnetic field or a single photon. Based on the horizontal or vertical polarization of the photon it can take values of 1 or 0. In quantum world the qubit can take both values at once, this is called as superposition. The qubit in superposition has a probability between 0 and 1, and it can't be predicted. But when it is measured it collapses into one of the definite states. For example, if a normal computer, it has four classical bits and four different configurations is needed at a time, the possibilities are $2^4 = 16$ that is 16 different possible combination, out of which only one can be used. Meanwhile in quantum computers all 16 position can be as one. Another property of the qubit is entanglement, that is a qubit can react and change to other state instantaneously just by close connection of another qubit, no matter how far they are. So, one entangled qubit can be measures by directly deducing the properties of it's partners without having to look. A normal logic gates receives a simple set of inputs and produces one definite output but in quantum gate, it manipulates an input of superposition, calculates probabilities and produces another superposition as an output. The quantum computer sets up some qubits, applies quantum gates to entangle them and manipulate probabilities, and finally it collapses superposition to an actual sequence of 0s and 1s, it does a lot of calculations at the same time to produce one definite output. The most famous use of quantum computing is ruining IT security. We know that our browsing and banking data are being kept secured by an encryption system in which everyone gets a public key to encode messages that only the user can decode, and this public key can actually be used to calculate our secret private key. To do the calculation in a normal computer it would take years of trials and error but in a quantum computer it can be done without much effort. There is also quantum cryptography, a method of encryption that uses the properties of quantum mechanics to secure and transmit data in a way that it cannot be hacked, it has already been tested by banks and companies like J.P.Morgan, Wells Fargo and Barclays

## 7.3 Our prediction

After doing so much research, we thought it would be interesting to give a prediction of what we think the sixth generation will look like. It is often said; " history repeats itself", but in this case this seems a poor prediction.
We believe that a significant part of upcoming developments will be due to quantum computing. It has the potential to revolutionize many areas of computing. Computers are already being programmed to understand or analyse our own language. We expect that our own 'human' language will become the most common programming language. This helps increase the speed at which we invent computer technology even more, since a problem can be looked at from any point of view. No longer only by programmers. We talked briefly about the Moore's Law, which suggested that the number of components that can be fitted onto a chip doubles every year. With present technology, it is actually every two years and we believe with that in the sixth generation the Moore's law can no longer be met. Experts have also been predicting the end of Moore's law for decades and we might finally be right. There are two significant issues holding us back from further miniaturization. First we are reaching the limit of how small of a chip we can create. When transistors get really small, the electrodes might be separated by only a few dozen atoms, electrons can then jump the gap. This might cause a transistor to leak current, so they do not function as well as bigger transistors. We do expect many ethical questions to arise as we improve computer technology and more becomes possible.

# Chapter 8

# Conclusion - observations throughout the generations

## 8.1 Introduction

In the first chapter, we outlined the purpose and questions we had before starting this project. We can easily conclude that the computer has improved on all aspects. Something much more interesting for us to look at is in what areas exactly and what the reason behind it was. Architecture, which includes components such as the memory storage, microprocessors and integrated circuits. This impacts the choice of programming language and bit size which in turn affects performance, versatility, speed and obviously cost.

## 8.2 Architecture

To be able to consider the changes in the architecture of a computer. We first need to be clear what is considered under the term architecture. From the dictionary we get: *"the complex or carefully designed structure of something"*. We will thus consider the tangible aspects of a computer.
The memory storage has improved from the magnetic drum memory that was enormous in size compared to the RAM we use in computers nowadays. The transistor caused the integrated circuit, which in turn caused the microprocessor.

Hence the shift from vacuum tubes to transistors, made the total structure of the computer more compact, smaller and lighter. The size from the second to the third generation went up in size a bit actually due to the development of the keyboard and monitor, but in the fourth generation the size decreased even more.

### 8.2.1 Programming languages

Engineers in the first generation were only able to program the computer by writing everything in machine language, which is a collection of binary digits, also called low-level language. In the second generation, using a assembler, they could write the instructions in assembly language.

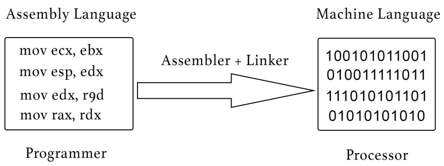


Figure 8.1: Assembly language

This made it more accessible for people to help further develop the computer, since it was not as difficult to learn the language. When the programming languages evolved further more into high-level languages, the number of people working on the advancement of the computer expanded. This ensured that much greater strides could be made. (in performance/ functioning)

### 8.2.2 Bit size

From the moment the the microprocessor was developed in the fourth generation, an important indicator of the performance of a computer, was the word size in bits. Figure 8.1 shows the growth in word size in bits from the moment the microprocessor was developed until the present.

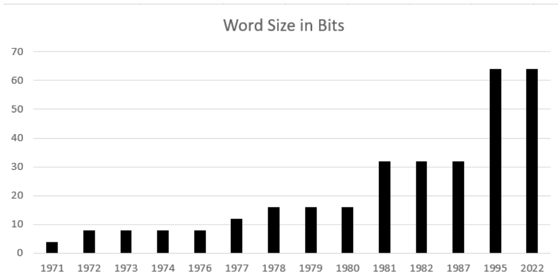


FIGURE 8.2: Word Size in bits from 1971-2022

A 4 bit processor means for instance that it can read, process and execute, also called the 'fetch-execute' cycle, with 4 bits at a time. One bit being one 1 or one 0 only. We discussed permutations in the probability course, so we would like to compute the possible permutations, thus the number of possible commands in only one cycle. To clarify this, we looked up the meaning of permutation: *"each of several possible ways in which a set or number of things can be ordered or arranged"*. We know the formula for permutations:

$${}_{n}P_{k} = \frac{n!}{(n-k)!}$$

where n = the word size in bits

k = 2, since we have the binary numbers 0 and 1

When substituting n for word size 4, then we get:
${}_{4}P_{2} = \frac{4!}{(4-2)!} = 12$

When substituting n for word size 64, then we get:
${}_{64}P_{2} = \frac{64!}{(64-2)!} = 4032$
This shows that with a bigger word size, there are more possible permutations per cycle, thus more possible instructions that a computer could execute in one cycle. This results into faster computation time with a higher word size. A computer with a lower word size could perform the same instruction, but it would possibly need multiple circles to execute the instruction.

## 8.3 Performance

The performance of a computer depends on multiple factor which we will discuss in more detail. All inventions over the past five generations have contributed to the improvement of the computer. Or more important, the reliability of a computer has steadily increased over time. Each development affects the other and as a consequence it is all connected. We tried to distinguish the some separate criteria of the performance of a computer.

### 8.3.1 Versatility

"Versatility in a machine means it is capable of performing almost any task, provided the task can be reduced to a series of logical steps." Due to the operating systems in the third generation, the versatility was much improved. Each program is made up of program instructions. When some of these instructions are running, it is called a process. Since main memory is too small to run all the processes, the operating system uses multiprocessing to allocate use of the main memory and the CPU. The operating system allows different applications to run at the same time.

## 8.4 Speed

The last hundred years computers have satisfied the Moore's Law, which says that computer processors are getting faster. It says that the numbers of transistors on a chip doubles every two years. There is however a limit to the size of a transistor, which implies that the Moore's law will no longer suffice. When we look at table 8.1 we see that in generation four and five, the computation time has not changed. This is due to the fact that faster than a picosecond is nearly impossible as for now.

| Generation | Computation Time | Converted to Seconds |
|---|---|---|
| 1 | miliseconds | 1.00E-03 |
| 2 | microseconds | 1.00E-06 |
| 3 | nanoseconds | 1.00E-09 |
| 4 | picoseconds | 1.00E-12 |
| 5 | picoseconds | 1.00E-12 |

Table 8.1: Computation time over five generations

## 8.5 Cost

The cost of the first generation was immensely high. The energy needed to let the ENIAC, as an example, run for one hour would be equivalent to the energy used for one household a whole month.[15] The transistors did not produce as much heat, but were still very costly, because they had to be assembled manually. These new developments caused for computers to overheat less and with that to need less cooling, which decreased cost by a lot. This also made the computers less prone to fail, they became more reliable which decreased the cost as well. The fourth generation is surprisingly the cheapest of all five.

# Bibliography


[1] Keyboard glove. 1960.

[2] Ross keyboard. 1971.

[3] Maltron keyboard. 1981.

[4] Mary Bellis. History of the computer keyboard. 2020.

[5] Brandon. What is magnetic hysteresis and why is it important?

[6] cHM. Input and output.

[7] CHM. Timeline of computer history.

[8] ElegantQuestion.com. Transistor.

[9] George Garza. Difference between cpu and microprocessor.

[10] William Gayde. Anatomy of a cpu. 2020.

[11] Ginni. What is control unit? 2021.

[12] Ginni. What is the main memory? 2021.

[13] Michael L Hilton. Babbage's difference engine. 2015.

[14] Andie Hioki. The cray-1 supercomputer. 2021.

[15] Niels Nielsen. Power consumption of vacuum tube. 2020.

[16] General Note. Third generation. 2020.

[17] Lisha Pace. Cathode ray tube explained. 2022.

[18] Electrical Revolution. Comparition of silicon and germanium. 2017.

[19] Nitin Sharma. Difference between magnetic tape and magnetic disk. 2020.

[20] Wikipedia. Combinational logic. 2015.

[21] Wikipedia. Magnetic core memory. 2020.