Introduction to Computer Organization and Architecture*

Nguyen Thi Hoang Lan

This work is produced by The Connexions Project and licensed under the
Creative Commons Attribution License †

1 1. Organization and Architecture

In describing computer system, a distinction is often made between computer architecture and computer organization.

Computer architecture refers to those attributes of a system visible to a programmer, or put another way, those attributes that have a direct impact on the logical execution of a program.

Computer organization refers to the operational units and their interconnection that realize the architecture specification.

Examples of architecture attributes include the instruction set, the number of bit to represent various data types (e.g.., numbers, and characters), I/O mechanisms, and technique for addressing memory.

Examples of organization attributes include those hardware details transparent to the programmer, such as control signals, interfaces between the computer and peripherals, and the memory technology used.

As an example, it is an architectural design issue whether a computer will have a multiply instruction. It is an organizational issue whether that instruction will be implemented by a special multiply unit or by a mechanism that makes repeated use of the add unit of the system. The organization decision may be bases on the anticipated frequency of use of the multiply instruction, the relative speed of the two approaches, and the cost and physical size of a special multiply unit.

Historically, and still today, the distinction between architecture and organization has been an important one. Many computer manufacturers offer a family of computer model, all with the same architecture but with differences in organization. Consequently, the different models in the family have different price and performance characteristics. Furthermore, an architecture may survive many years, but its organization changes with changing technology.

2 2. Structure and Function

A computer is a complex system; contemporary computers contain million of elementary electronic components. How, then, can one clearly describe them? The key is to recognize the hierarchical nature of most complex system. A hierarchical system is a set of interrelated subsystem, each of the later, in turn, hierarchical in structure until we reach some lowest level of elementary subsystem.

The hierarchical nature of complex systems is essential to both their design and their description. The designer need only deal with a particular level of the system at a time. At each level, the system consists of a set of components and their interrelationships. The behavior at each level depends only on a simplified,

*Version 1.1: Jul 15, 2009 10:30 pm -0500

†http://creativecommons.org/licenses/by/3.0/

abstracted characterization of the system at the next lower level. At each level, the designer is concerned with structure and function:

- Structure: The way in which the components are interrelated.
- Function: The operation of each individual component as part of the structure.

In term of description, we have two choices: starting at the bottom and building up to a complete description, or beginning with a top view and decomposing the system, describing their structure and function, and proceed to successively lower layer of the hierarchy. The approach taken in this course follows the latter.

2.1 2.1 Function

In general terms, there are four main functions of a computer:

- Data processing
- Data storage
- Data movement
- Control

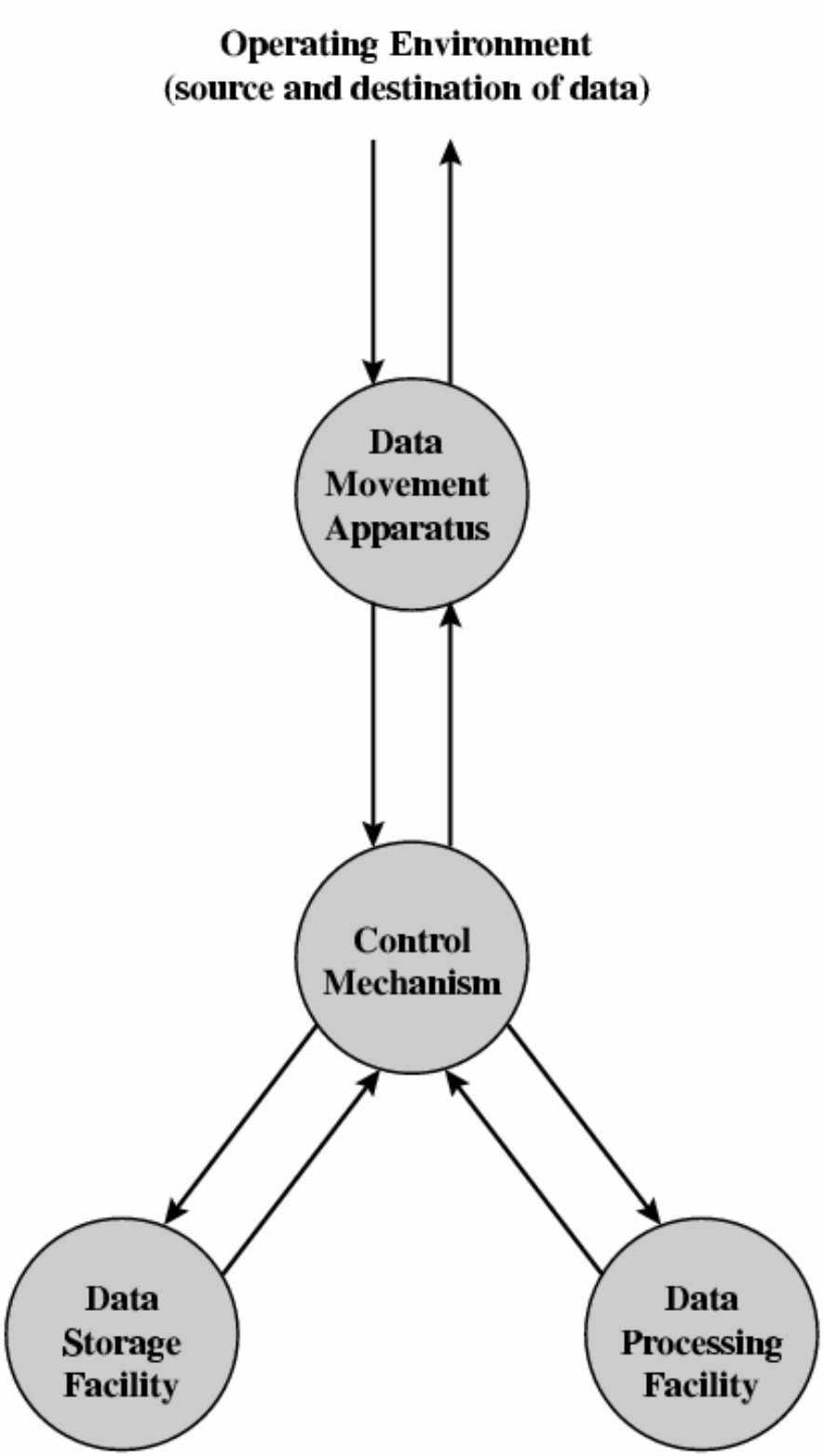
Operating Environment
(source and destination of data)
Data
Movement
Apparatus
Control
Mechanism
Data
Storage
Facility
Data
Processing
Facility

Figure 1

Figure 1.1 A functional view of the computer

The computer, of course, must be able to process data. The data may take a wide variety of forms, and the range of processing requirements is broad. However, we shall see that there are only a few fundamental methods or types of data processing.

It is also essential that a computer store data. Even if the computer is processing data on the fly (i.e., data come in and get processed, and the results go out immediately), the computer must temporarily store at least those pieces of data that are being worked on at any given moment. Thus, there is at least a short-term data storage function. Files of data are stored on the computer for subsequent retrieval and update.

The computer must be able to move data between itself and the outside world. The computer's operating environment consists of devices that serve as either sources or destinations of data. When data are received from or delivered to a device that is directly connected to the computer, the process is known as input-output (I/O), and the device is referred to as a peripheral. When data are moved over longer distances, to or from a remote device, the process is known as data communications.

Finally, there must be control of there three functions. Ultimately, this control is exercised by the individual who provides the computer with instructions. Within the computer system, a control unit manages the computer's resources and orchestrates the performance of its functional parts in response to those instructions.

At this general level of discussion, the number of possible operations that can be performed is few. The figure 1.2 depicts the four possible types of operations.

The computer can function as a data movement device (Figure 1.2a), simply transferring data from one peripheral or communications line to another. It can also function as a data storage device (Figure 1.2b), with data transferred from the external environment to computer storage (read) and vice versa (write). The final two diagrams show operations involving data processing, on data either in storage or in route between storage and the external environment.

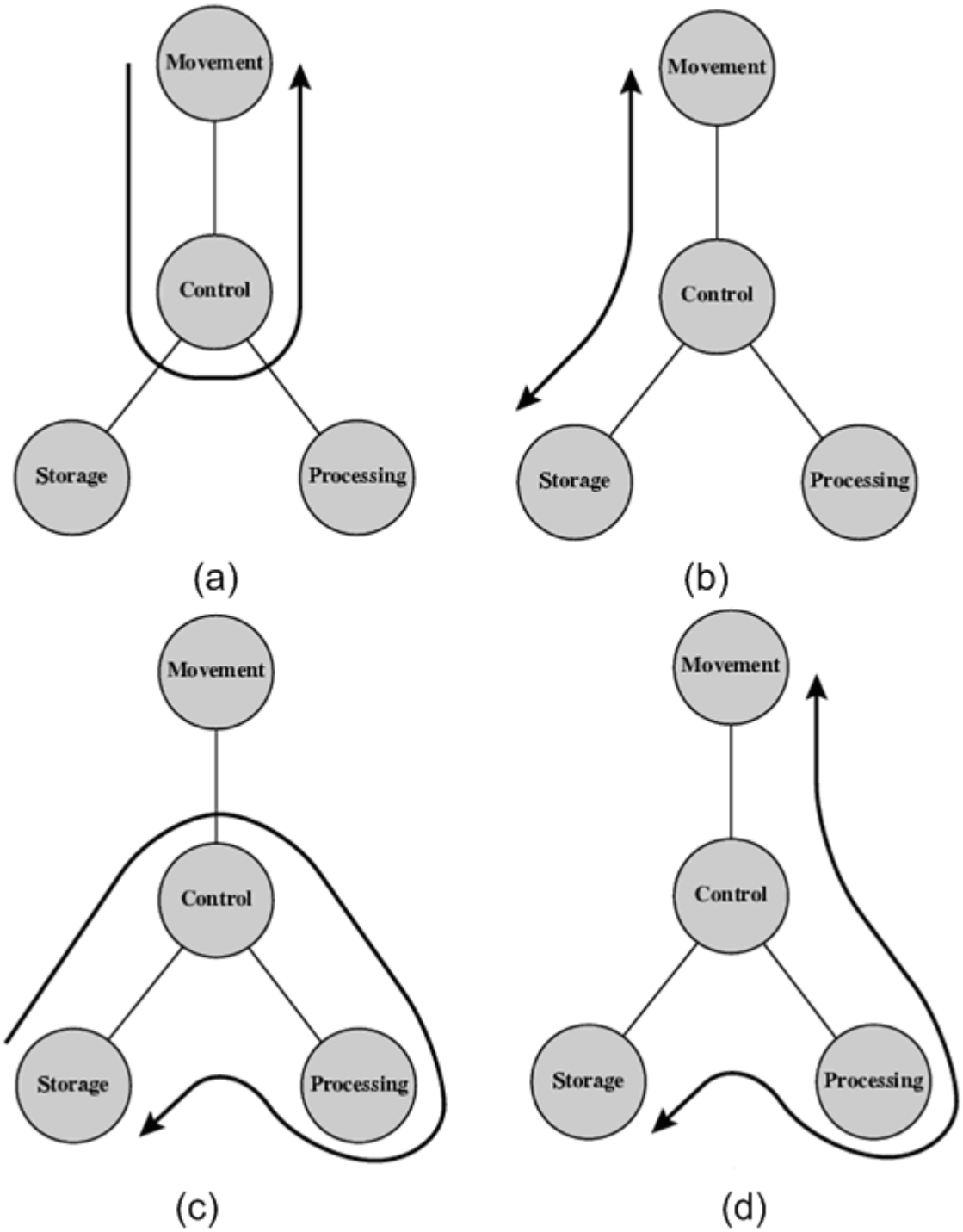
Movement
Control
Storage
Processing
(a)
Movement
Control
Storage
Processing
(b)
Movement
Control
Storage
Processing
(c)
Movement
Control
Storage
Processing
(d)

Figure 2

Figure 1.2 Possible computer operations

2.2 2.2 Structure

Figure 1.3 is the simplest possible depiction of a computer. The computer is an entity that interacts in some fashion with its external environment. In general, all of its linkages to the external environment can be classified as peripheral devices or communication lines. We will have something to say about both types of linkages.

- Central Processing Unit (CPU): Controls the operation of the computer and performs its data processing functions. Often simply referred to as processor.
- Main Memory: Stores data.
- I/O: Moves data between the computer and its external environment.
- System Interconnection: Some mechanism that provides for communication among CPU, main memory, and I/O.

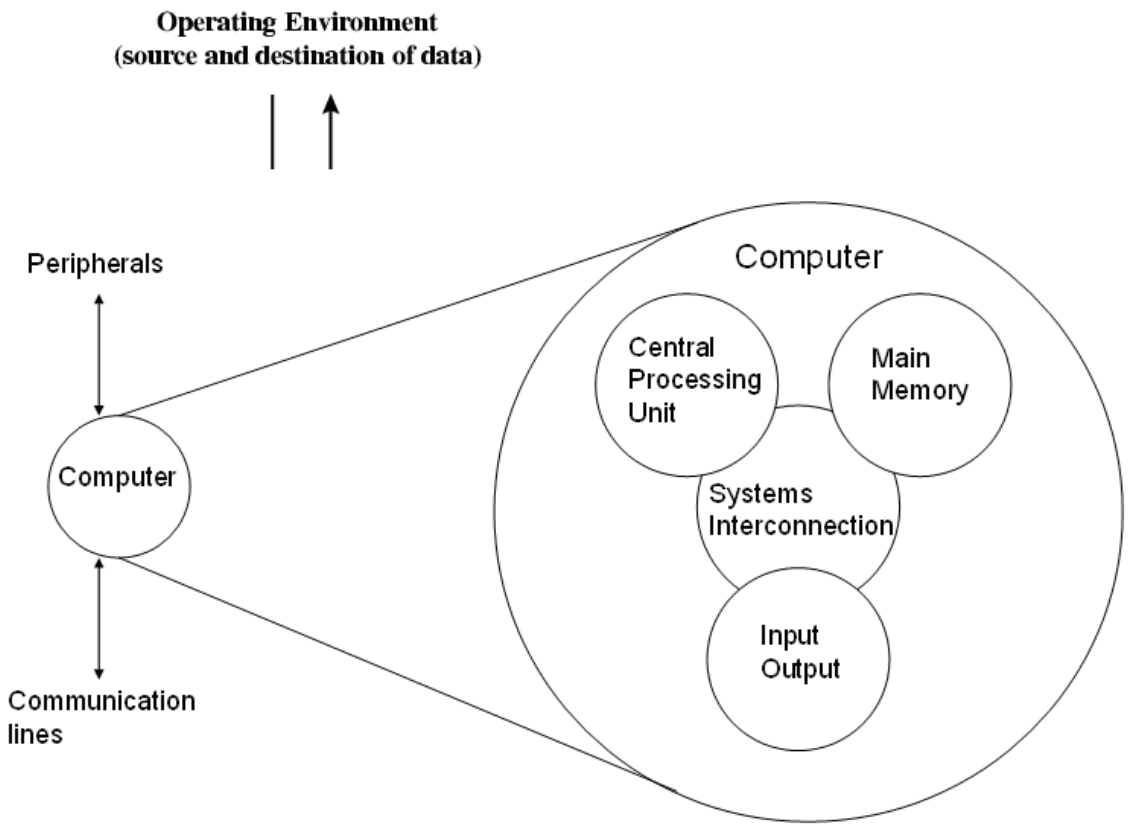

Figure 1.3: The computer: top-level structure

There may be one or more of each of the above components. Traditionally, there has been just a single CPU. In recent years, there has been increasing use of multiple processors, in a single system. Each of these components will be examined in some detail in later lectures. However, for our purpose, the most interesting and in some ways the most complex component is the CPU; its structure is depicted in Figure 1.4. Its major structural components are:

- Control Unit (CU): Controls the operation of the CPU and hence the computer.
- Arithmetic and Logic Unit (ALU): Performs computer's data processing functions.

- Register: Provides storage internal to the CPU.
- CPU Interconnection: Some mechanism that provides for communication among the control unit, ALU, and register.

Each of these components will be examined in some detail in next lectures.

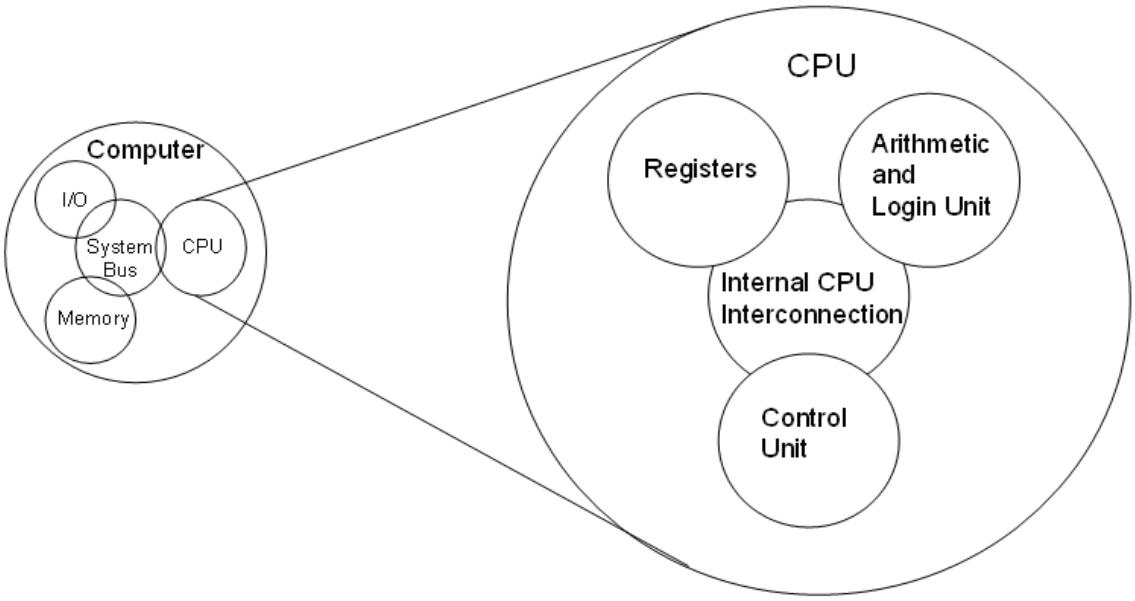

Figure 3

Figure 1.4 The CPU

3 3. A Brief History of Computers

3.1 3.1 The first Generation: Vacuum Tubes

3.1.1 ENIAC

The ENIAC (Electronic Numerical Integrator And Computer), designed by and constructed under the supervision of Jonh Mauchly and John Presper Eckert at the University of Pennsylvania, was the world's first general-purpose electronic digital computer. The project was a response to U.S. wartime needs. Mauchly, a professor of electrical engineering at the University of Pennsylvania and Eckert, one of his graduate students, proposed to build a general-purpose computer using vacuum tubes. In 1943, this proposal was accepted by the Army, and work began on the ENIAC. The resulting machine was enormous, weighting 30 tons, occupying 15,000 square feet of floor space, and containing more than 18,000 vacuum tubes. When operating, it consumed 140 kilowatts of power. It was aloes substantially faster than any electronic-mechanical computer, being capable of 5000 additions per second.

The ENIAC was decimal rather than a binary machine. That is, numbers were represented in decimal form and arithmetic was performed in the decimal system. Its memory consisted of 20 "accumulators", each capable of holding a 10-digit decimal number. Each digit was represented by a ring of 10 vacuum tubes. At

any time, only one vacuum tube was in the ON state, representing one of the 10 digits. The major drawback of the ENIAC was that it had to be programmed manually by setting switches and plugging and unplugging cables.

The ENIAC was completed in 1946, too late to be used in the war effort. Instead, its first task was to perform a series of complex calculations that were used to help determine the feasibility of the H-bomb. The ENIAC continued to be used until 1955.

3.1.2 The von Neumann Machine

The programming process could be facilitated if the program could be represented in a form suitable for storing in memory alongside the data. Then, a computer could get its instructions by reading them from memory, and a program could be set of altered by setting the values of a portion of memory.

This idea, known as the Stored-program concept, is usually attributed to the ENIAC designers, most notably the mathematician John von Neumann, who was a consultant on the ENIAC project. The idea was also developed at about the same time by Turing. The first publication of the idea was in a 1945 proposal by von Neumann for a new computer, the EDVAC (Electronic Discrete Variable Computer).

In 1946, von Neumann and his colleagues began the design of a new stored-program computer, referred to as the IAS computer, at the Princeton Institute for Advanced Studies. The IAS computer, although not completed until 1952, is the prototype of all subsequent general-purpose computers. Figure 1.5 shows the general structure of the IAS computer. It consists of:

- A main memory, which stores both data and instructions.
- An arithmetic-logical unit (ALU) capable of operating on binary data.
- A control unit, which interprets the instructions in memory and causes them to be executed.
- Input and output (I/O) equipment operated by the control unit.

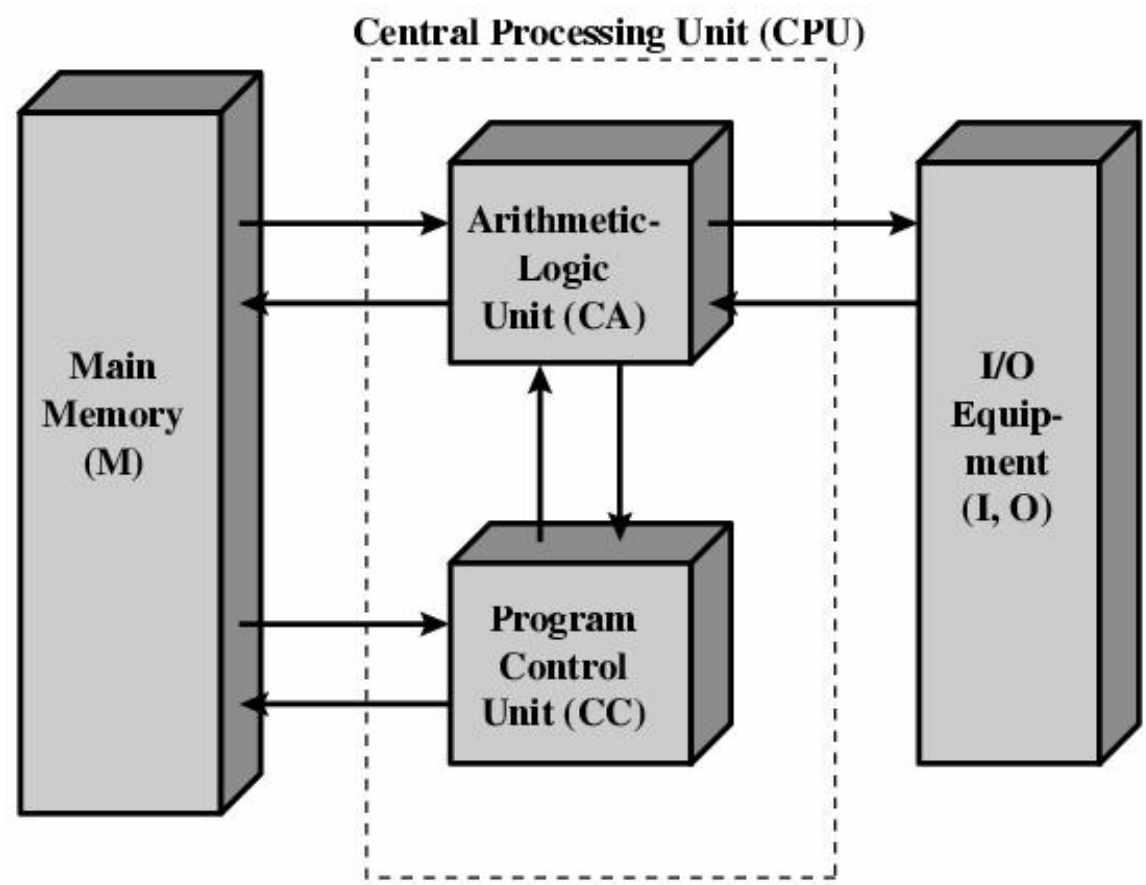

Figure 4

Figure 1.5 Structure of the IAS computer

3.1.3 Commercial Computers

The 1950s saw the birth of the computer industry with two companies, Sperry and IBM, dominating the marketplace.

In 1947, Eckert and Mauchly formed the Eckert-Maunchly computer Corporation to manufacture computers commercially. Their first successful machine was the UNIVAC I (Universal Automatic Computer), which was commissioned by the Bureau of the Census for the 1950 calculations. The Eckert-Maunchly Computer Corporation became part of the UNIVAC division of Sperry-Rand Corporation, which went on to build a series of successor machines.

The UNIVAC II, which had greater memory capacity and higher performance than the UNIVAC I, was delivered in the late 1950s and illustrates several trends that have remained characteristic of the computer industry. First, advances in technology allow companies to continue to build larger, more powerful computers. Second, each company tries to make its new machines upward compatible with the older machines. This means that the programs written for the older machines can be executed on the new machine. This strategy

is adopted in the hopes of retaining the customer base; that is, when a customer decides to buy a newer machine, he is likely to get it from the same company to avoid losing the investment in programs.

The UNIVAC division also began development of the 1100 series of computers, which was to be its bread and butler. This series illustrates a distinction that existed at one time. The first model, the UNIVAC 1103, and its successors for many years were primarily intended for scientific applications, involving long and complex calculations. Other companies concentrated on business applications, which involved processing large amounts of text data. This split has largely disappeared but it was evident for a number of years.

IBM, which was then the major manufacturer of punched-card processing equipment, delivered its first electronic stored-program computer, the 701, in 1953. The 70l was intended primarily for scientific applications. In 1955, IBM introduced the companion 702 product, which had a number of hardware features that suited it to business applications. These were the first of a long series of 700/7000 computers that established IBM as the overwhelmingly dominant computer manufacturer.

3.2 3.2 The Second Generation: Transistors

The first major change in the electronic computer came with the replacement of the vacuum tube by the transistor. The transistor is smaller, cheaper, and dissipates less heal than a vacuum tube but can be used in the same way as a vacuum tube to construct computers. Unlike the vacuum tube, which requires wires, metal plates, a glass capsule, and a vacuum, the transistor is a solid-state device, made from silicon.

The transistor was invented at Bell Labs in 1947 and by the 1950s had launched an electronic revolution. It was not until the late 1950s, however, that fully transistorized computers were commercially available. IBM again was not the first company to deliver the new technology. NCR and. more successfully. RCA were the front-runners with some small transistor machines. IBM followed shortly with the 7000 series.

The use of the transistor defines the second generation of computers. It has become widely accepted to classify computers into generations based on the fundamental hardware technology employed. Each new generation is characterized by greater processing performance, larger memory capacity, and smaller size than the previous one.

3.3 3.3 The Third Generation: Integrated Circuits

A single, self-contained transistor is called a discrete component. Throughout the 1950s and early 1960s, electronic equipment was composed largely of discrete components—transistors, resistors, capacitors, and so on. Discrete components were manufactured separately, packaged in their own containers, and soldered or wired together onto circuit boards, which were then installed in computers, oscilloscopes, and other electronic equipment. Whenever an electronic device called for a transistor, a little lube of metal containing a pinhead-sized piece of silicon had to be soldered to a circuit hoard. The entire manufacturing process, from transistor to circuit board, was expensive and cumbersome.

These facts of life were beginning to create problems in the computer industry. Early second-generation computers contained about 10,000 transistors. This figure grew to the hundreds of thousands, making the manufacture of newer, more powerful machines increasingly difficult.

In 1958 came the achievement that revolutionized electronics and started the era of microelectronics: the invention of the integrated circuit. It is the integrated circuit that defines the third generation of computers. Perhaps the two most important members of the third generation are the IBM System/360 and the DEC PDP-8.

3.4 3.4 Later Generations

Beyond the third generation there is less general agreement on defining generations of computers. There have been a fourth and a fifth generation, based on advances in integrated circuit technology. With the introduction of large-scale integration (LSI), more than 1000,000 components can be placed on a single integrated circuit chip. Very-large-scale integration (VLSI) achieved more than 1000,000,000 components per chip, and current VLSI chips can contain more than 1000.000 components.

3.5 3.5 The summary of Generations of Computer

- Vacuum tube - 1946-1957
- Transistor - 1958-1964
- Small scale integration: 1965

Up to 100 devices on a chip

- Medium scale integration: -1971

100-3,000 devices on a chip

- Large scale integration :1971-1977

3,000 - 100,000 devices on a chip

- Very large scale integration: 1978 -1991

100,000 - 100,000,000 devices on a chip

- Ultra large scale integration : 1991

Over 100,000,000 devices on a chip