Computing & Computers

①

Limitations of manual computing:

As the size and complexity of the calculations being carried out increases, two serious limitations of manual computation become apparent.

(i) The speed at which a human computer can work is limited. A typical elementary operation such as addition or multiplication takes several seconds or minutes. Problems requiring billions of such operations could never be solved manually in a reasonable period of time or at reasonable cost. Fortunately, modern computers routinely tackle and quickly solve such problems.

(ii) Humans are notoriously prone to error, so long calculations done by hand are unreliable unless elaborate precautions are taken to eliminate mistakes. Most sources of human error (distraction, fatigue, and the like) do not affect machines, so they can provide results that are, ~~whithin~~ within ~~board~~ broad limits, free from error.

Components of a computer:

A computer has several key components that roughly correspond to those just mentioned. The "main memory" corresponds to the paper used in the manual calculation. It purpose is to store instructions and data. The computer's brain is its "Central Processing Unit (CPU)". It contains a "program control unit" (also known as an instruction unit) whose function is to fetch instructions from memory and interpret them. An "Arithmetic Logic Unit (ALU)", which is part of the CPU's data-processing or execution unit, carries out the instructions. The ALU is so called because many instructions specify either arithmetic (numerical) operations or various forms of nonnumerical operations that loosely correspond to logical reasoning or decision making.

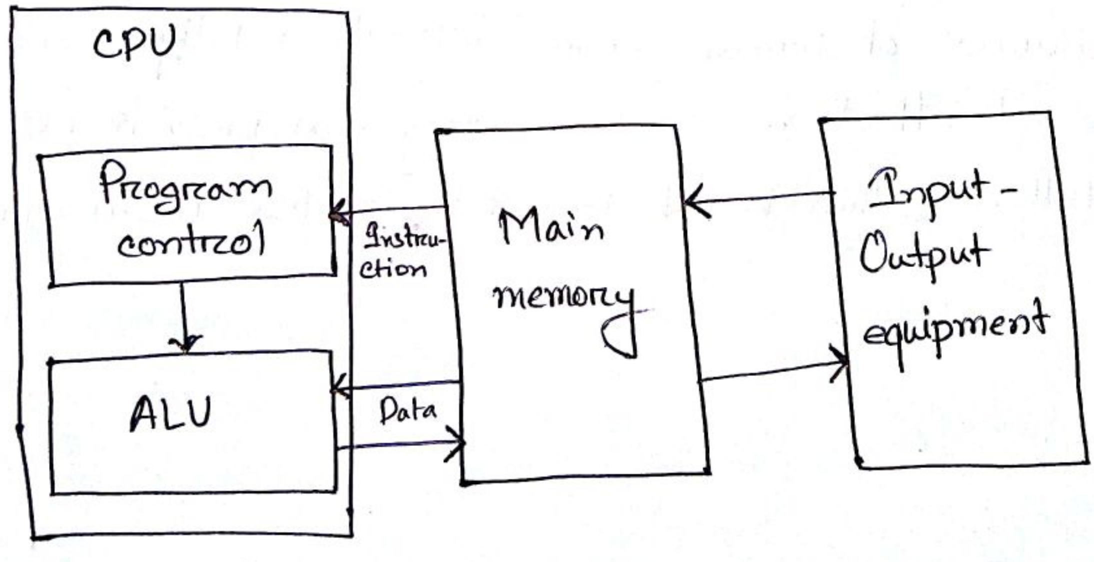


fig: Main Components of machine computation.

⇒ The computational abilities of general-purpose Digital Computer:

* The computer should not store the answers to all possible problems.

* The computer should only be required to solve problems for which a solution procedure or program can be given.

* The computer should process information at a finite speed.

Limitations of computers:

Some limitations of a computer are given below.

① Unsolvable Problems:

Problems exist that no Turing machine and therefore no practical computer can solve. There are well-defined problems, some quite famous, for which no solutions or solution procedures are known. An example from pure mathematics is "Goldbach's conjecture", formulated by the mathematician Christian Goldbach's, which states that every even integer greater than 2 is the sum of exactly two prime numbers. For instance, $8 = 3+5$, and $108 = 37+71$.

⇒ Real computers have a finite amount of memory and are therefore referred to as "finite state machines".

(ii) Intractable Problems:

Real (finite-state) computers can solve most computational problems to an acceptable degree of accuracy. The question then becomes: Can a computer of reasonable size and cost solve a given problem in a reasonable amount of time? If so, the problem is said to be tractable.

⇒ A mechanical computer has two serious drawbacks:

(i) Its computing speed is limited by the inertia of its moving parts.

(ii) The transmission of digital information by mechanical means is quite unreliable.

Generations of Computers:

⇒ First generation:

The earliest attempt to construct an electronic computer using "vacuum" tubes appears to have been made in the late 1930s by John V. Atanasoff.

The first widely known general-purpose electronic computer was the Electronic Numerical Integrator and Calculator (ENIAC)

The concept of EDVAC was first published in a 1945 proposal by von Neumann for a new computer, the

Electronic Discrete Variable Computer (EDVAC).

The EDVAC had two kinds of memory: a fast main memory with a capacity of 1024 or 1k words, (numbers or instructions) and a slower secondary memory with a capacity of 20K words.

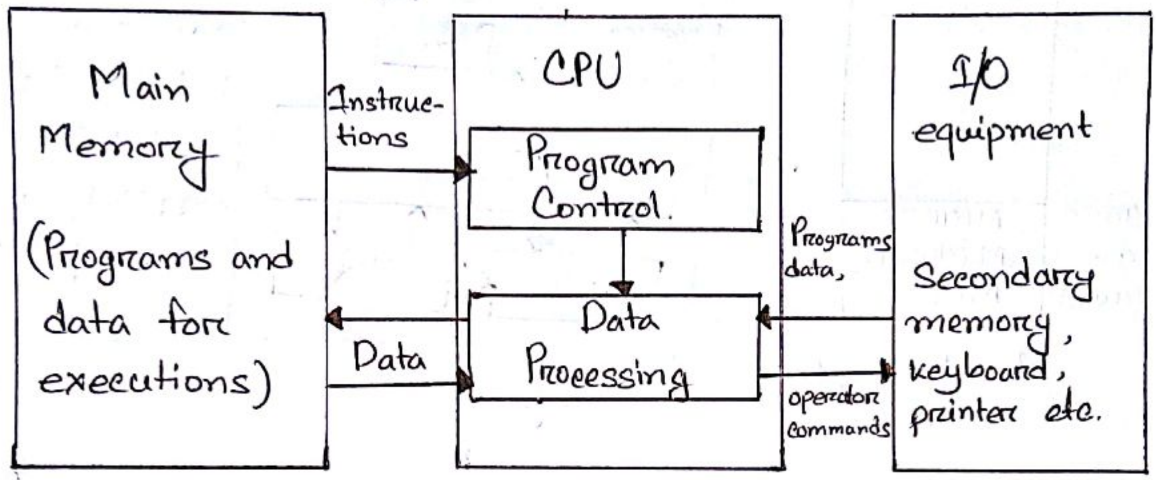


fig: Organization of first-generation computer.

IAS Computer:

The basic unit of information in the IAS computer is a 40-bit word, which is the standard packet of information stored in a memory location or transferred in one step between the CPU and the main memory M. Each location in M can be used to store either a single 40-bit number or else a pair of 20-bit instructions. The IAS's number format is "fixed-point".

The CPU contains a small set of high-speed storage devices called "registers", which serve as implicit storage locations for operands and results.

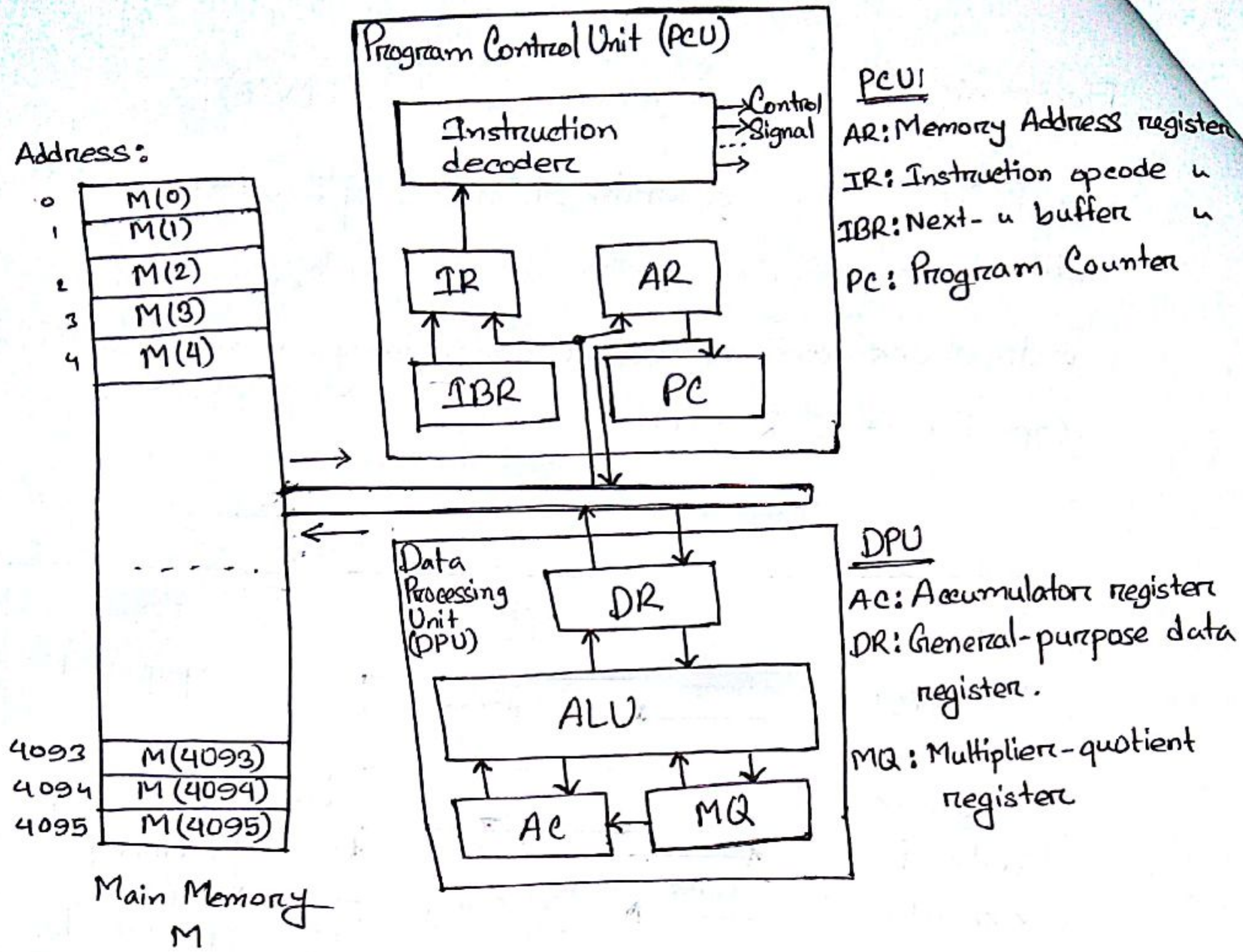


fig: Organization of the CPU and main memory of the IAS computer.

⇒ Instruction set: To represent instructions, we will use a notation called a "Hardware Description Language (HDL)" or "register-transfer language" computer.

The group of instructions called program-control or branch instructions determine the sequence in which instructions are executed. Recall that the program counter PC specifies the address of the next instruction to be executed.

Ex:

Instructions	Comment
AC := M (100)	Load the contents of memory loc. 100 into the AC
AC := AC+M(101)	Add the contents of mem. loc. 101 to the AC.
M(102) := AC	Store the contents of the AC in mem. loc. 102.

Improvement in IAS Computer:

(i) The small amount of storage space in the CPU results in a great deal of unproductive data-transfer traffic between the CPU and main memory M; it also adds a program length. Later computers have more CPU registers and a special memory called "cache" that acts as a buffer between the CPU registers and M.

(ii) No facilities were provided for structuring programs.

(iii) The instruction set is biased toward numerical computation. Programs for non-numerical tasks such as text processing were difficult to write and executed slowly.

(iv) Input-Output (IO) instructions were considered of minor importance - in fact, they are not mentioned in Burks, Goldstine, and von Neumann beyond noting that they are necessary.

The Second Generation:

The vaccum tube quickly gave away the transistor, which was invented by Bell Laboratories in 1947, and a second generation of computers based on transistors superseded the 1st generation of vacuum tube-based machines. Like a vaccum tube, a transistor serves as a high-speed electronic switch for binary signals, but it is smaller, cheaper, sturdier, and requires much less power than the vacuum tube.

The IAS computer still served as the basic model, but more registers were added to the CPU to facilitate data and address manipulation.

The IO operation ties up the CPU with a trivial data-transfer task. Moreover, many IO devices transfer data at low speeds compared to that ~~[illegible]~~ of the CPU because of their inherent reliance on electromechanical rather than electronic technology. Thus the CPU is idle most of the time when executing an IO program directed at a relatively slow device such as a printer. To eliminate this bottleneck, computers such as the IBM ~~[illegible]~~ 7094 introduced input-output processors (IOPs).

A special program called a "compiler" translates a user program from the high-level language in which it is written into the machine language of the particular computer on which the program is to be executed.

⊛ The first successful high-level programming language was FORTRAN (FORmula TRANslation).

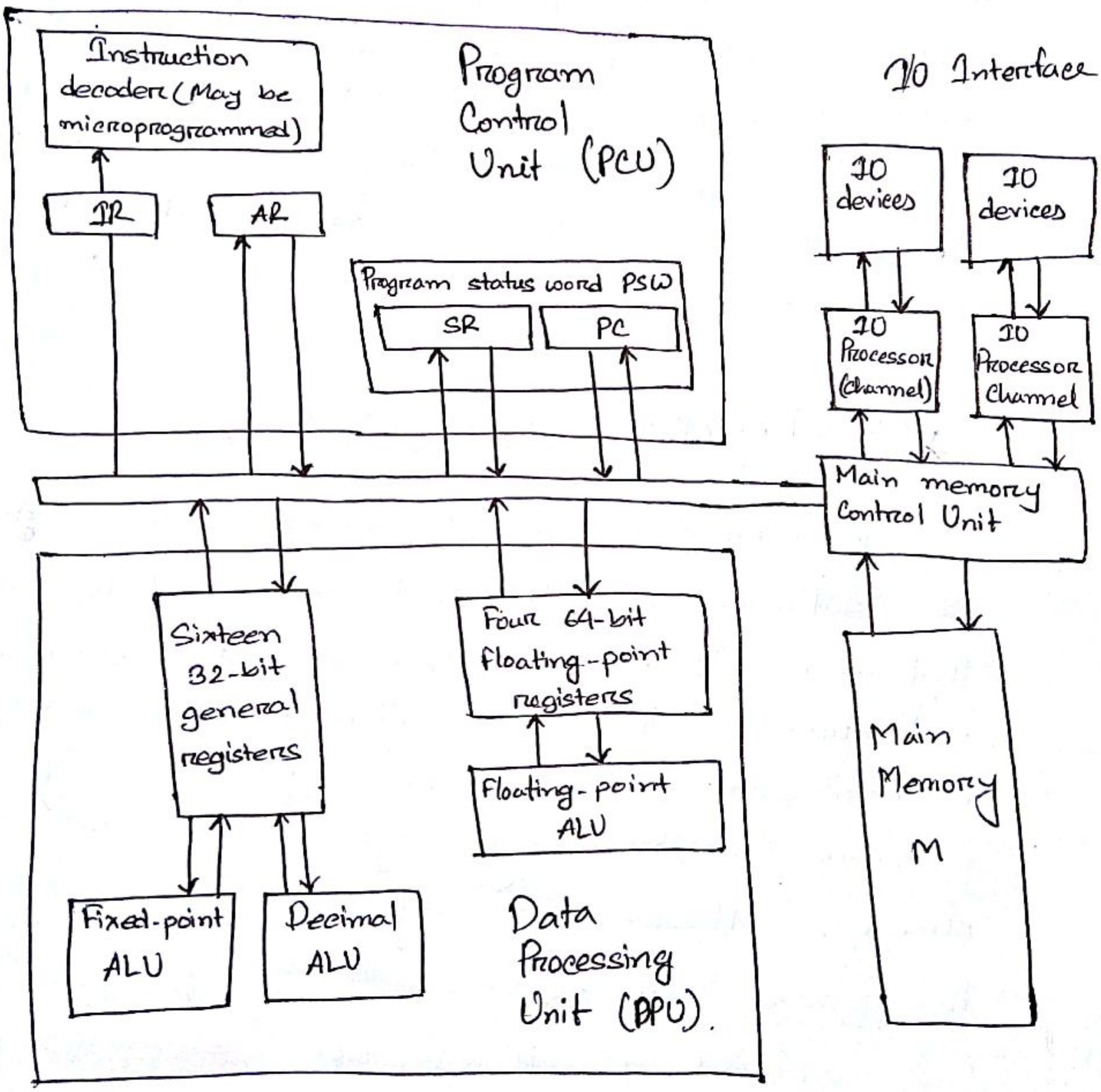
⊛

The Third Generation:

The transistor continued as the basic switching device, but ICs allowed large numbers of transistors and associated components to be combined on a tiny piece of semiconductor material, usually silicon. IC technology initiated a long-term trend in computer design toward smaller size, higher speed, and lower hardware cost.

Structure of the IBM system/360:

The various System/360 models were designed to be "software compatible" with one another, meaning that all models in the series shared a common instruction set. Programs written for one model could be run with modification on any other; only the execution time, memory usage, and the like would change. Software capability enabled computer owners to upgrade their system with having to rewrite large amount of software. The System/360 models also used a common operating system, OS/360, and the manufacturer suplied specialized software to support such widely used applications as transaction processing and database management. In addition, the system/360 models had many hardware characteristics in common, including the same interface for attaching IO devices.



Personal Computer:

Microcomputer technology gave rise to a new class of general-purpose machines called "Personal Computers" (PCs), which are intended for a single user.

Two of the main applications of PCs are

(i) Word processing
(ii) Data processing

"word processing", where personal computers have assumed and greatly expanded all the functions of the typewriter, and "data processing" tasks like financial record keeping.

⇒ The IBM PC series is based on Intel Corp's 80X86 family of microprocessors.

The microprocessor's internal (micro) architecture usually contains a number of speedup features not found in its predecessors. A system bus connects the microprocessors to a main memory based on semi-conductors DRAM technology and to an IO subsystem. A separate IO bus, such as the industry standard PCI (peripheral component interconnect) "local" bus, connects directly to the IO devices and their individual controllers. The IO bus is linked to the system bus, to which the microprocessors and memory are attached via a special bus-to-bus control unit sometimes referred to as a "bridge".

Performance Considerations:

A computer's performance is also strongly affected by other factors besides its instruction set, especially the time required to move instructions and data between the CPU and main memory M and, to a lesser extent, the time required to move information between M and IO devices. It typically takes the CPU about five times longer to obtain a word from M than from one of its internal registers. This difference in speed has existed since the first electronic computers.

⇒ The three separate factors softwares, architectures, and hardware technology joinly determine a computer's performance.

1. Software: The efficiency with which the programs are written and compiled into object code influences N, the number of instructions executed. Other factors being equal, reducing N tends to reduce the overall execution time T.

2. Architecture: The efficiency with which individual instructions are processed directly affects CPI, the number of cycles per instructions executed. Reducing CPI also tends to reduce T.

3. Hardware:

The raw speed of the processor circuits determines f, the clock frequency. Increasing f tends to reduce T.

Speedup techniques:

A "cache" is a memory unit placed between the CPU and main memory M and used to store instructions data, or both. It has much smaller storage capacity than M, but it can be accessed (read from or written into) more rapidly and is often placed (at least partly) on the same chip as the CPU. The cache's effect is to reduce the average time required to access an instruction or data word, typically to just a single clock cycle.

Another important speedup technique known as "pipelining" allows the processing of several instructions to be partially overlapped. Pipelining is most easily done for a sequence of instructions of the same or similar types that employ a single E-unit, such as a floating-point processor.