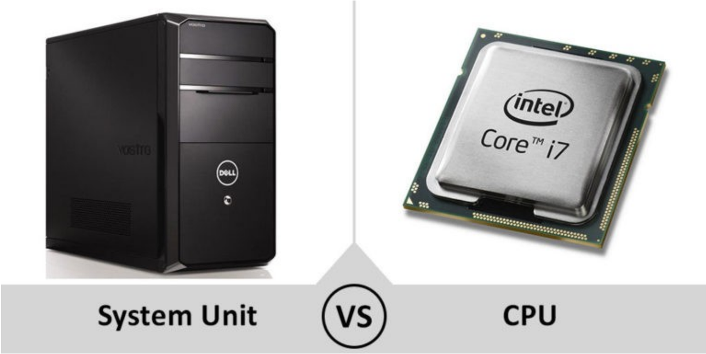
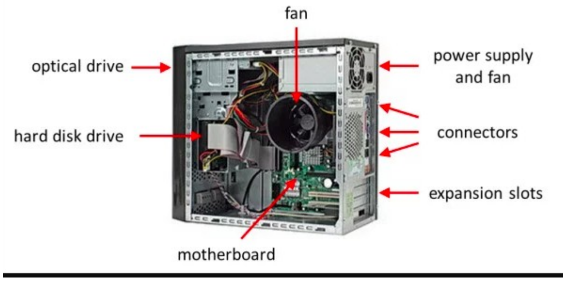
Chapter – 02

Introduction of CPU



SYSTEM UNIT

System Unit, includes all the devices like the storage devices, input devices, output devices, communication devices, connection devices, as well as the processing devices. A system unit is a collection in which there are several electronic components and circuits. These all things are connected to each other in order to perform the successful operations, called a system unit or a CPU box. System Unit main component are: Motherboard, CPU, Power supply, Cooling fan,Internal speaker,Drive bays,Expansion slots, Memory, Chipset, BIOS chip,I/O Ports etc.



CPU

◆ The term CPU refers to the Central Processing Unit. It is also called " the brain of computer " as it controls operation of all parts of computer. This is a small component of the system unit. This is also known as the processor or the microprocessor. It is an Integrated circuit chip that processes electronic signals. All the instruction are carried to the processor and the job of the processor is to process them and to execute them. These instructions are carried to the processor in the forms of the signals. Then the processor reads them and passes the signals to the desired component. When this process completes the result is shown on the screen of the computer. The speed of CPU is measured in Hz (Nowadays, usually in GHz).

◆ An electronic clock regulates the rate at which the CPU runs and synchronizes all the various computer components.

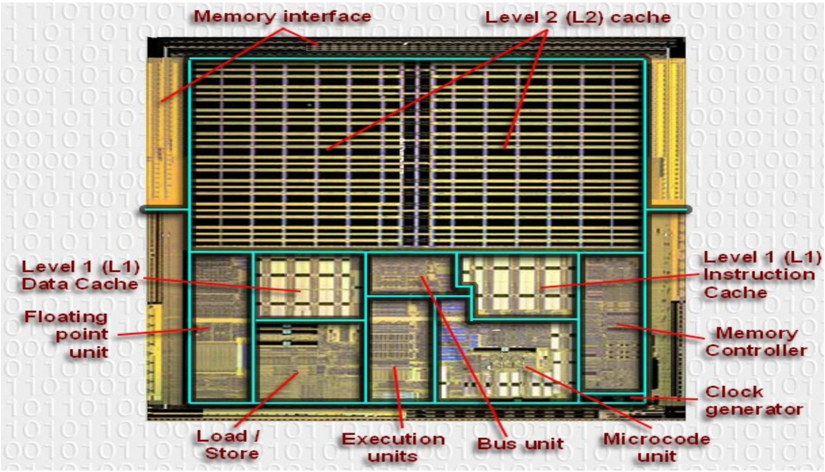
◆ The higher the clock frequency, the more instructions the CPU can execute per second.

◆ The speed of the clock (and therefore the speed of the CPU) is measured in Gigahertz (GHz).

- Also, a central processing unit (CPU) is the electronic circuitry within a computer that carries out the instructions of a computer program by performing the basic arithmetic, logical, control and input/output (I/O) operations specified by the instructions.

CPU Socket : The **processor socket** (also called a **CPU socket**) is the connector on the motherboard that houses a **CPU** and forms the electrical interface and contact with the **CPU**. In computer hardware, a CPU socket or CPU slot contains one or more mechanical components providing mechanical and electrical connections between a microprocessor and a printed circuit board (PCB).

Internal details of CPU



Components of CPU:

Storage Component : Registers
Execution (Processing) Component : Arithmetic and Logic Unit (ALU)
Transfer Component : Bus
Control Component : Control Unit

Registers:

Registers are high speed storage areas in the CPU. All data must be stored in a register before it can be processed. **Register** are used to quickly accept, store, and transfer data and instructions that are being used immediately by the CPU. When we gives some input to the system then the **input will be stored into the registers** and when the system will gives us the results after processing then the result will again pass through the registers. So that they are used by the **CPU for Processing the Data** which is given by the user.

Some of the register used in CPU are:

- **Accumulator (AC)** : Accumulator, is a register to perform calculation as well as holding the results of computations. Basic operations like add, subtract, multiplication or division can occur within the accumulator register. Accumulator register stores the operands used in arithmetic operations and hold the results of calculations formed. Furthermore, the accumulator can be used to perform the logical functions like AND and OR.
- **Memory Address Register (MAR)**: This register holds the memory addresses of data and instructions. This register is used to access data and instructions from memory during the execution phase of an instruction. Suppose CPU wants to store some data in the memory or to read the data from the memory. It places the address of the-required memory location in the MAR.
- **Memory Data Register (MDR)**: It holds data that is being transferred to or from memory of the addressed location. It acts **like a buffer** and holds anything that is copied from the memory ready for the processor to use it.
- **Memory Buffer Register (MBR)**

  This register holds the contents of data or instruction read from, or written in memory. It means that this register is used to store data/instruction coming from the memory or going to the memory.
- **Index Register**

  A hardware element which holds a number that can be added to (or, in some cases, subtracted from) the address portion of a computer instruction to form an effective address. Also known as base register. An index register in a computer's CPU is a processor register used for modifying operand addresses during the run of a program.
- **Instruction Register(IR)**

  In CPU, instruction register is used to store or hold the actual instruction being loaded, decoded or executed currently. After loaded the instruction from memory, it is decoded and then executed. Consequently, instruction register serves the data input to the part of decoding process. Here, Decoding refers to the separation of the instruction into its op code and operands, the movement of data (such as loading data into a register from a memory address in one of the operands), and the generation of control signals to the ALU for instruction execution.
- **Program Counter(PC)**

  Another special-purpose register in CPU is program counter, also called the instruction pointer, holds the address of the instruction being executed currently or the address of the following instruction to be executed. For example, suppose there is an instruction being executed and after the arithmetic and logic unit (ALU) processed it, the processor will search for the second instruction to be executed. Thus, program counter is responsible for the following process, which is to track and store the address of the following instruction that is going to be executed for the purpose of saving time.

- **Status Registers**

  Status of CPU and currently executing program,

  - Flags: **Flags** are often binary **flags**, which contain a boolean value (true or false). Depending upon the value of result after any arithmetic and logical operation the flag bits become set (1) or reset (0). However, not all **flags** are binary, **meaning** they can store a range of values. Flags are used to track condition like arithmetic carry and overflow, power failure, internal computer error.
  - Status Bits: Carry Bit C , Sign bit S, Zero bit Z, Overflow bit V,

General purpose register and Special purpose register

General purpose register:

The general purpose register can store a data or a memory location address. Its size can be of 8 bit to 32 bit. It is a multipurpose register. They can be used either by programmer or by a user.

General purpose registers deal with a wide variety of performance. They are used to hold data values or intermediate results that will be used frequently. Some of the general-purpose registers in 8086 microprocessor includes accumulator, data register, address register, source index, destination index, base pointer, stack pointer, and base register. Some are discussed below:

The four general purpose registers are the AX, BX, CX and DX registers.

In 8086 microprocessor, all of the general purpose registers can be treated as a 16 bit quantity or as two 8 bit quantities. The high byte is referenced by replacing the X with H. The low byte is referenced by replacing the X with L.

AX	AH	AL
	15 8	7 0

Register	Function
AX	This is the accumulator. It is 16-bit registers, but it is divided into two 8-bit registers. These registers are AH and AL. AX generally used for arithmetic or logical instructions, but it is not mandatory in 8086.
BX	This is the base register. BX is another register pair consisting of BH and BL. This register is used to store the offset values.
CX	CX is generally used as control register. It has two parts CH and CL. For different looping and counting purposes these are used.
DX	DX is data register. The two parts are DH and DL. This register can be used in Multiplication, Input/output addressing etc.

Special purpose register:

Some registers serve specific functions within the CPU. The Special Purpose register are the one which are designed for some specific task only. They perform some particular task or some data handling task. They can be used to store the program state. Some of the more important of these registers includes instruction register, program counter etc.

- **Instruction Register(IR)**

In CPU, instruction register is used to store or hold the actual instruction being loaded, decoded or executed currently. After loaded the instruction from memory, it is decoded and then executed. Consequently, instruction register serves the data input to the part of decoding process. Here, Decoding refers to the separation of the instruction into its op code and operands, the movement of data (such as loading data into a register from a memory address in one of the operands), and the generation of control signals to the ALU for instruction execution.

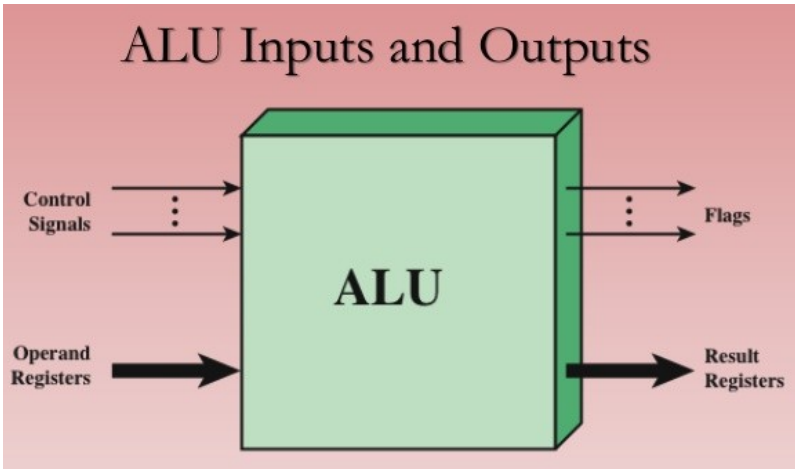
- **Program Counter(PC)**

Another special-purpose register in CPU is program counter, also called the instruction pointer, holds the address of the instruction being executed currently or the address of the following instruction to be executed. For example, suppose there is an instruction being executed and after the arithmetic and logic unit (ALU) processed it, the processor will search for the second instruction to be executed. Thus, program counter is responsible for the following process, which is to track and store the address of the following instruction that is going to be executed for the purpose of saving time.

<u>Arithmetic and Logic Unit (ALU):</u>

- Arithmetic Logic Unit performs the computer's data processing functions, i.e. arithmetic logic unit is a digital circuit used to perform arithmetic and logic operations. It represents the fundamental building block of the central processing unit of a computer. Modern CPUs contain very powerful and complex ALUs . In addition to ALUs, modern CPUs contain a control unit (CU).

- Most of the operations of a CPU are performed by one or more ALUs, which load data from input registers. A register is a small amount of storage available as part of a CPU. The control unit tells the ALU what operation to perform on that data and the ALU stores the result in an output register. The control unit moves the data between these registers, the ALU, and memory.



- The arithmetic/logic unit can perform four kinds of arithmetic operations or mathematical calculations: addition, subtraction, multiplication, and division. As its name implies, the arithmetic/logic unit also performs logical operations. A logical operation is usually a comparison. The unit can compare numbers, letters, or special characters. A logic unit can usually recognize with six logical relationships: equal to, less than, greater than, less than or equal to, greater than or equal to, and not equal. Furthermore, logical unit includes (AND, OR, NOT etc.) operations also.
  Bit Shifting Operations might be used to shifting the positions of the bits by a certain number of places to the right or left.

Bus

Buses are the means by which data is transmitted from one part of a computer to another, connecting all major internal components to the CPU and memory. The width of the bus determines how much data can moved at a time. The bus management unit manages the transfer of data along the external bus connections, including the links to RAM.

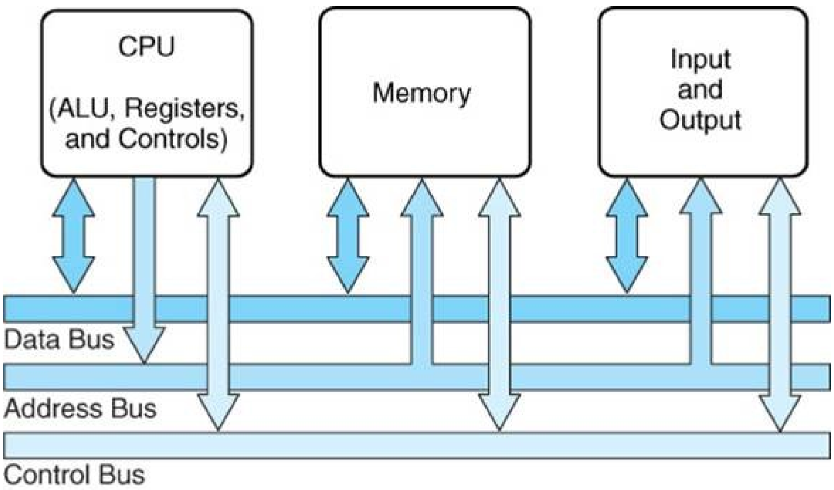
Computers have two major types of buses:

1. System bus:- This is the bus that connects the CPU to the main memory on the motherboard. The system bus is also called the front-side bus, memory bus, local bus, or host bus.

2. A number of Input/Output Buses, connecting various peripheral devices to the CPU. These devices connect to the system bus via a 'bridge' implemented in the processors chipset . Other names for the I/O bus include "expansion bus", "external bus" or "host bus". Some of the expansion bus includes: **SATA-** Serial Advanced Technology Attachment, PCI - Peripheral Component Interconnect, AGP - Accelerated Graphics Port, SCSI - Small Computer Systems Interface, USB- Universal Serial Bus etc.

System Bus

A standard CPU system bus is comprised of a control bus, data bus and address bus.



Data Bus: Data bus is the most common type of bus. It is used to transfer data between different components of computer. The number of lines in data bus affects the speed of data transfer between different components. The data bus consists of 8, 16, 32, 64 lines. A 64-line data bus can transfer 64 bits of data at one time.

Address Bus: Many components are connected to one another through buses. Each component is assigned a unique ID. This ID is called the address of that component. If a component wants to communicate with another component, it uses address bus to specify the address of that component. The address is a unidirectional bus. It can carry information only in one direction. It carries address of memory location from microprocessor to main memory. A computer with a 32-bit address bus can address 4 GB of memory.

Control Bus: Control bus is used to transmit different commands or control signals from one component to another component. A control signal contains the timing information and command signal with type of operation to be performed . For example, one line of the bus is used to indicate whether the CPU is currently reading from or writing to main memory.

<u>Control unit:</u>

- **Control Unit** is the part of the computer's central processing unit (CPU), which directs the operation of the processor. It is the responsibility of the Control Unit to tell the computer's memory, arithmetic/logic unit and input and output devices how to respond to the instructions that have been sent to the processor. It fetches internal instructions of the programs from the main memory to the processor instruction register, and based on this register contents, the control unit generates a control signal that supervises the execution of these instructions.

Functions:

- It controls all activities of computer
- Supervises flow of data within CPU
- Directs flow of data within CPU
- Transfers data to Arithmetic and Logic Unit
- Transfers results to memory
- Fetches results from memory to output devices Controls the operation of the CPU and hence the computer.

<u>CPU working mechanism:</u>

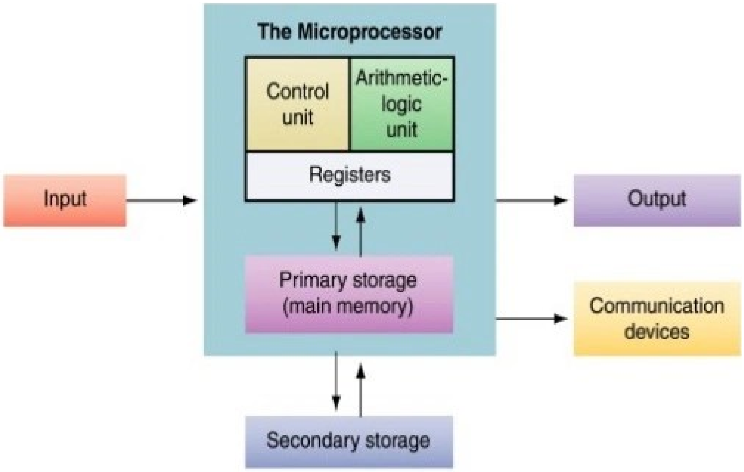
- CPU consists of basic units as : control unit, Arithmetic Logical Unit and registers.
- Input is given through the input devices to CPU.
- Control unit controls communication within ALU and memory unit.
- Decides which circuit is to be activated.
- For reading instruction it uses Fetch-execute mechanism.
- Control unit gets instruction from memory.
- Control unit decides what to do of that instruction and transfers it to the ALU.

- ALU performs various arithmetic operations like addition, subtraction, multiplication, division and logical operations like AND, OR, NOT etc. on that instruction.

- Results of ALU are stored in the memory or resistor for its further operations.

- After completing the instruction, stored results are passed to the output devices.

- To synchronize all these operations CPU uses its own system clock.



The Four Primary Functions of the CPU:

The CPU processes instructions, while processing it performs four basic steps:

1. Fetch: Each instruction is stored in memory and has its own address. The processor takes this address number from the program counter, which is responsible for tracking which instructions the CPU should execute next.
2. Decode: All programs to be executed are translated to into Assembly instructions. Assembly code must be decoded into binary instructions, which are understandable by CPU. This step is called decoding.
3. Execute: While executing instructions the CPU can do one of three things: Do calculations with its ALU, move data from one memory location to another, or jump to a different address.
4. Store: The CPU must give feedback after executing an instruction, and the output data is written to the memory.

Instruction Set

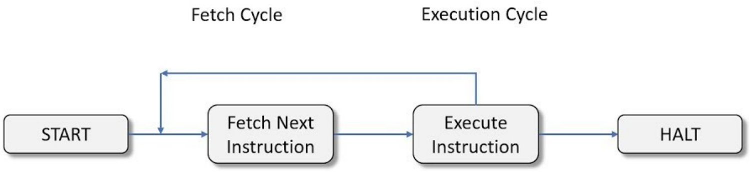
The **instruction set**, also called **ISA** (**instruction set architecture**), is part of a computer that is applicable to programming, which is basically machine language. The instruction set provides commands to the processor, to tell it what it needs to do. The instruction set consists of addressing modes, instructions, data types, registers, memory architecture, interrupt, and exception handling, and external input/output.

Instruction sets have their rules like a language has grammar. An instruction includes an opcode that specifies the operation to perform, such as add/multiply/move contents of memory to register—and operand specifiers, which may specify registers, memory locations, or literal data.

Basic instruction set examples:

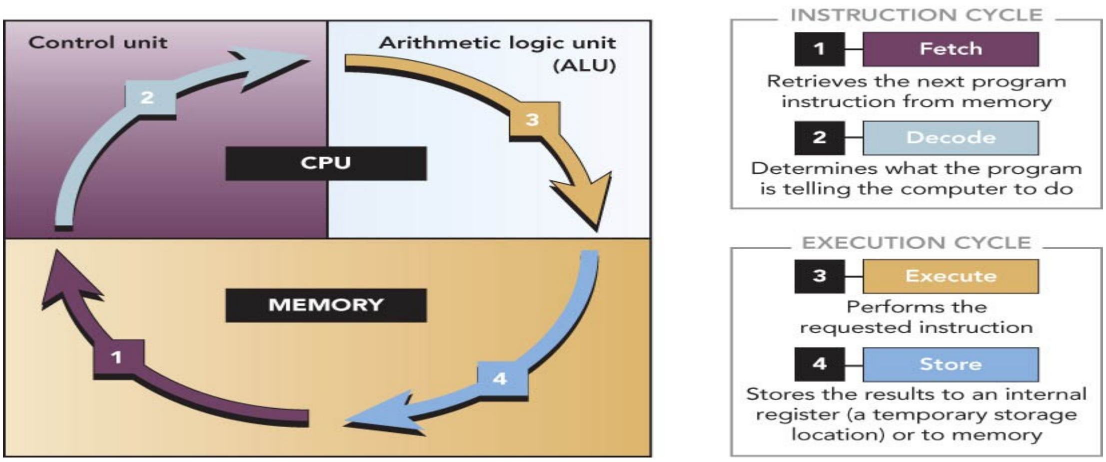
Instruction:	Meaning:
STO	Store data in a particular memory location
ADD	Add two numbers together
SUB	Subtract one number from another
MUL	Multiply two numbers
DIV	Divide two numbers
INC	Increment a number by adding 1
CMP	Compare two numbers to see if they are equal
JMP	Jump to a specific position in the instruction code

The **instruction cycle** (also known as the **fetch–decode–execute cycle** or simply the **fetch-execute cycle**) is the cycle which the Central Processing Unit(CPU) follows from boot-up until the computer has shut down in order to process instructions.



Machine Cycle

The steps performed by the computer processor for each machine language instruction received. The **machine cycle** is a 4 process cycle that includes reading and interpreting the machine language, executing the code and then storing that code.



Comparison Chart

	Machine Cycle	Instruction Cycle
Definition	The steps that get performed by the processor getting employed in a device and all the instructions that get implemented.	A process by which a computer takes an instruction given by a program then understands it and executes it from memory.
Processes	Fetch, decode, execute and store.	Fetch, decode, execute and run.
Components	Memory Unit and Central Processing Unit	Arithmetic Logic Unit, Registers, Data and Memory.
Value	The time Required by the microprocessor to complete the operation of accessing memory or I/O devices is called machine cycle.	The Steps required by CPU to fetch and execute an Instruction is called an Instruction Cycle.

The factors Affecting CPU performance are :

- Multiple cores,
- Cache memory,
- Clock speed,
- Word length,
- Address bus width,
- Data bus width,

<u>Pipelining</u>

- ➢ The speed of execution of programs is influenced by many factors. One way to improve performance is to use **faster circuit technology** to build the processor and the main memory. Another possibility is to arrange the hardware so that more than one operation can be performed at the same time.

- ➢ Pipelining is a particularly effective way of organizing concurrent activity in a computer system. It is a process of arrangement of hardware elements of the CPU such that its overall performance is increased. Simultaneous execution of more than one instruction takes place in a pipelined processor.

- ➢ Pipelining is a technique where multiple instructions are overlapped during execution. Pipeline is divided into stages and these stages are connected with one another to form a pipe like structure. Instructions enter from one end and exit from another end. Pipelining increase the overall system throughput.

- In simple terms, **parallel computing** is breaking up a task into smaller pieces and executing those pieces at the same time, each on their own processor or on a set of computers that have been networked together.

<u>Following are the 5 stages of pipeline with their respective operations:</u>

- Stage 1 (Instruction Fetch)

In this stage the CPU reads instructions from the address in the memory whose value is present in the program counter.

- Stage 2 (Instruction Decode)

In this stage, instruction is decoded and the register file is accessed to get the values from the registers used in the instruction.

- Stage 3 (Instruction Execute)

In this stage, ALU operations are performed.

- Stage 4 (Memory Access)

In this stage, memory operands are read and written from/to the memory that is present in the instruction.

- Stage 5 (Write Back)

In this stage, computed/fetched value is written back to the register present in the instructions.

Basic five-stage pipeline

Instr. No. \ Clock cycle	1	2	3	4	5	6	7
1	IF	ID	EX	MEM	WB		
2		IF	ID	EX	MEM	WB	
3			IF	ID	EX	MEM	WB
4				IF	ID	EX	MEM
5					IF	ID	EX

(IF = Instruction Fetch, ID = Instruction Decode, EX = Execute, MEM = Memory access, WB = Register write back).

In the fourth clock cycle (the green column), the earliest instruction is in MEM stage, and the latest instruction has not yet entered the pipeline.

Let's look at a simple example. Say we have the following equation:

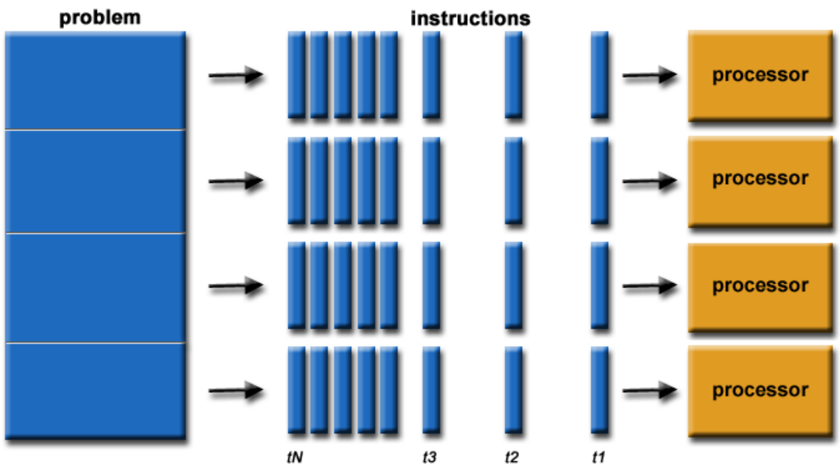
- Y = (4 x 5) + (1 x 6) + (5 x 3)

On a single processor, the steps needed to calculate a value for Y might look like:

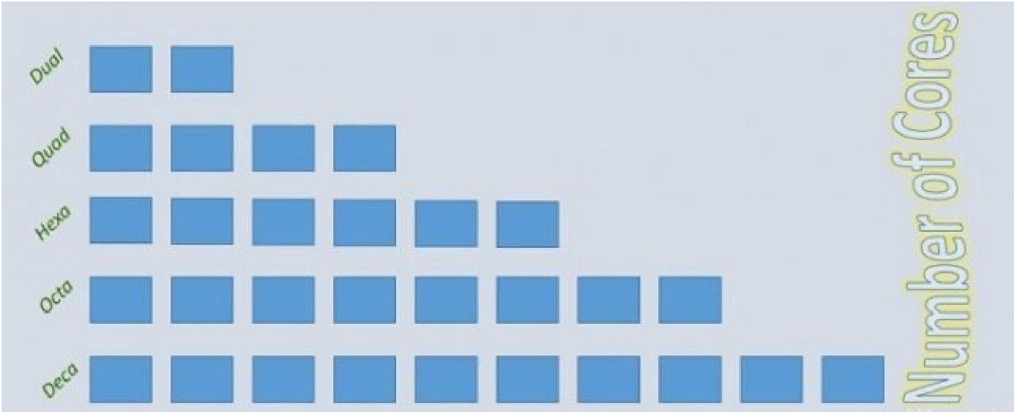
- Step 1: Y = 20 + (1 x 6) + (5 x 3)
- Step 2: Y = 20 + 6 + (5 x 3)
- Step 3: Y = 20 + 6 + 15
- Step 4: Y = 41

But in a parallel computing scenario, with three processors or computers, the steps look something like:

- Step 1: Y = 20 + 6 + 15
- Step 2: Y = 41



<u>Number of cores in CPU:</u>



Processors are named according to their numbers of complete independent units known as cores. A **core** is part of a CPU or an independent unit that receives instructions and performs calculations, or actions, based on those instructions. Dual as the name suggest double similarly quad means four so on. Dual core processors are processor made up of two cores. Quad core processors are processor made up of four cores and so on goes with Hexa, Octa, Deca core processors. So once the number of cores increases, the performance and task handling will be more faster than others.

Cache Memory

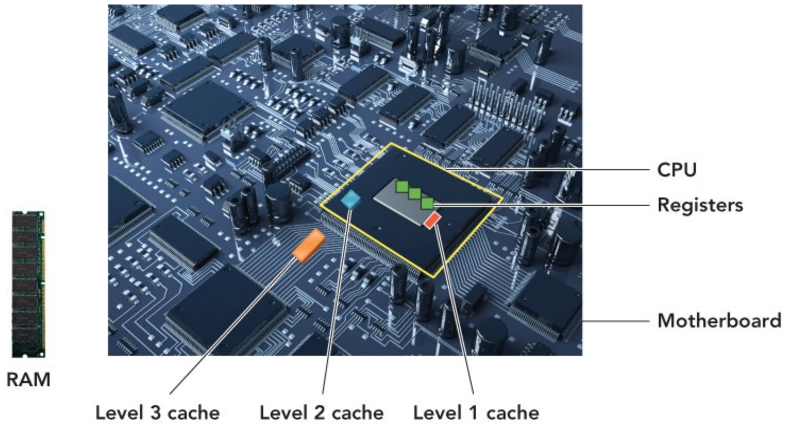
The cache is a very high speed, expensive piece of memory, which is used to speed up the memory retrieval process. Due to it's higher cost, the CPU comes with a relatively small amount of cache compared with the main memory. Without cache memory, every time the CPU requests for data, it would send the request to the main memory which would then be sent back across the system bus to the CPU. This is a slow process. The idea of introducing cache is that this extremely fast memory would store data that is frequently accessed and if possible, the data that is around it. This is to achieve the quickest possible response time to the CPU.

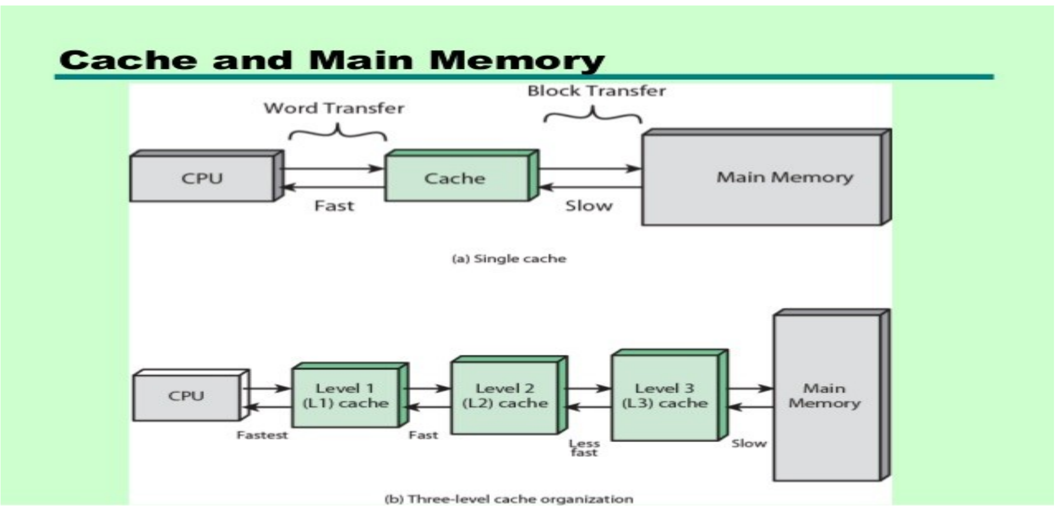
The data and instructions are retrieved from RAM when CPU uses them for the first time. A copy of that data or instructions is stored in cache. The next time the CPU needs that data or instructions, it first looks in cache. If the required data is found there, it is retrieved from cache memory instead of main memory. **It speeds up the working of CPU.**

Types/Levels of Cache Memory

A computer can have several different levels of cache memory. The level numbers refers to distance from CPU where Level 1 is the closest.

All levels of cache memory are faster than RAM. The cache closest to CPU is always faster but generally costs more and stores less data then other level of cache.





-Er. Shreeraj Khatiwada

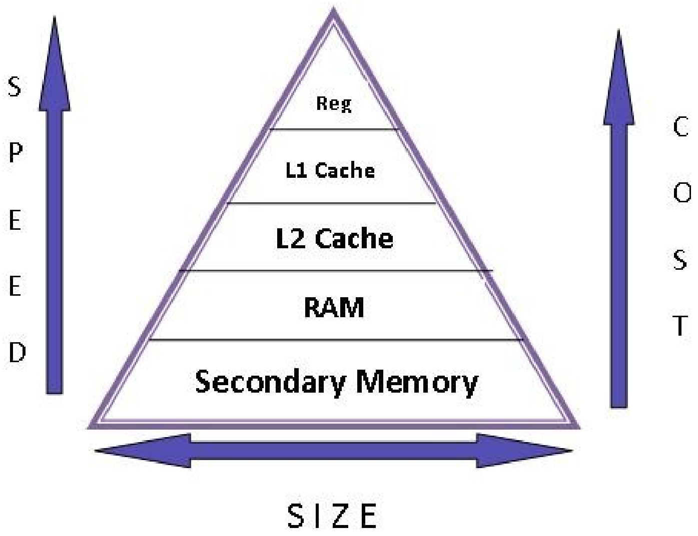
Comparison between:

L1 VS L2 VS L3 CACHE

L1	L2	L3
A cache memory that is directly built into the processor and is used to store the CPU's recently accessed information	A cache memory that is located outside and separated from the CPU chip core, although it is found on the same CPU chip package	A cache memory that is used by the CPU and is usually built onto the motherboard within the CPU module itself
Smallest cache	Larger than L1 but smaller than L3 cache	Largest cache
Called level 1 or primary or internal cache	Called level 2, secondary or external cache	Called level 3 or external cache
Fastest cache	Slower than L1 but faster than L3	Slowest cache
Each core in the CPU has their own L1 cache memory	Each code in the CPU has their own L2 cache memory	All cores in the CPU share the same L3 cache memory
Memory size may vary: from 2KB to 64KB	Memory size may vary: from 256KB to 512KB	Memory size may vary: from 1 MB to 8 MB

-Er. Shreeraj Khatiwada

Memory Hierarchy



- Memory Hierarchy gives the relationship between the speed,size and cost with respect to the distance from the Processor.
- In the diagram above the peak of the pyramid represents the processor. The Register lie with in the processor itself hence are the closest to the processor and work the fastest. But the number of registers that can be included in a processor are limited, as it would lead to increase in processor size, increase in manufacturing cost etc. Thus the register memory is restricted to minimal.
- L1 and L2 cache memory works at a very fast speed but is also extremely expensive as compared to the other memories available. This high cost is one of the major restriction why we can not use lots of cache in a computer even thought it is faster. Together L1 and L2 help in making the system run fast. Some of the modern computers also have a third level or L3 cache.
- The next level of memory is the RAM. RAM basically stores all the data that the computer needs while it is running. RAM is also fast in its operation but slower than the cache, and less expensive. Thus we can afford to have a relatively large amount of RAM. The higher the RAM, faster is the system, but again price is a main limiting factor with regards to how much RAM can be included in the computer.
- All the memories until the level of RAM are volatile, i.e. they loose the data stored in them as soon as power is switched off. This obviously is not a very desirable property as we loose all our work.
- Thus to store data even when power is not there secondary memory is made use of. The secondary memory is the slowest but the size is a lot bigger than the RAM or the cache because price of the secondary memory is a lot lesser than compared to RAM and cache.
- Let us take a real life example to understand the above hierarchy .

Consider any laptop that we use today, the configuration will be like:

L1 cache: 32KB, L2 Cache : 2048KB

RAM: 2GB, Secondary Memory: 160GB

This data above data depicts the memory hierarchy.

-Er. Shreeraj Khatiwada