

Machine Cycle

Machine Cycle: The sequence of steps a CPU takes to perform a task

The cpu is in a constant state of spinning inside the machine cycle

- Fetch
- Decode
- Execute
- Write-back

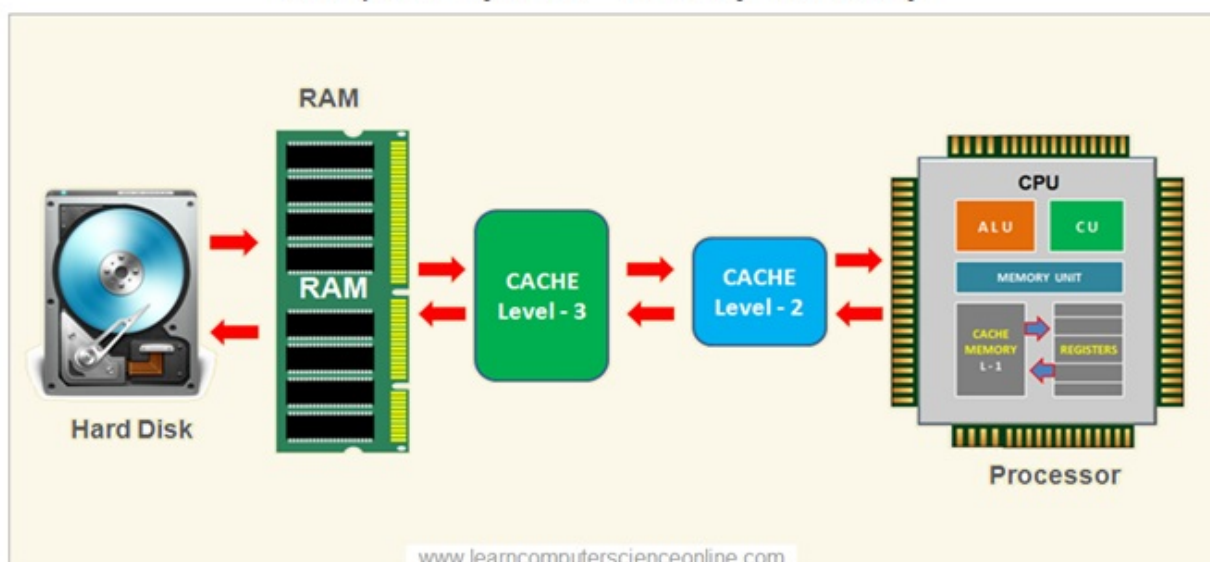
CPU registers:

- General: temporarily store data used during computations;
- Special: special-purpose registers, designed for specific functions like the program counter

Properties Of Ram Memory:

- Random-access: It takes roughly the same amount of time to access a piece of data, no matter where the data is located physically on the device.
- Volatile
- Memory Stall:
 - A scenario (bad) when CPU has to freeze its work and wait for RAM memory to deliver the necessary data
 - By separating data reading instruction (minimize Memory Stall)

Computer System - Memory Hierarchy



CPU Pipeline

The CPU pipeline is a technique used to improve the performance of a CPU by allowing multiple instructions to be executed simultaneously. It divides the machine cycle into several stages, each responsible for a specific task:

1. Fetch: The CPU retrieves the next instruction from memory.
2. Decode: The CPU decodes the instruction to determine the operation to be performed.
3. Execute: The CPU performs the operation specified by the instruction.
4. Write-back: The CPU stores the result of the operation in the appropriate register.

CPU Registers

CPU registers are small, high-speed storage locations within the CPU. They are used to temporarily store data during computations. There are two types of CPU registers:

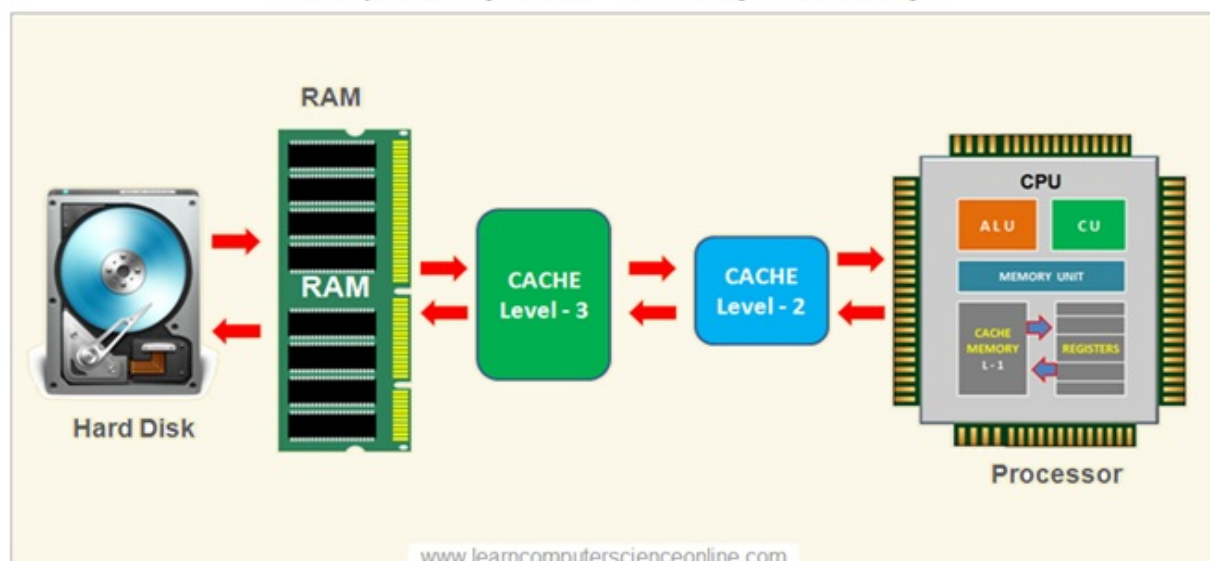
- General-purpose registers: These registers can be used for any purpose and are typically used to store intermediate results during calculations.
- Special-purpose registers: These registers are designed for specific functions, such as the program counter, which keeps track of the memory address of the next instruction to be executed.

Properties of RAM Memory

RAM (Random Access Memory) is a type of computer memory that allows data to be read from and written to. It has the following properties:

- Random access: It takes roughly the same amount of time to access any piece of data, regardless of its physical location on the device.
- Volatile: The data stored in RAM is lost when the power is turned off or interrupted.
- Memory stall: A memory stall occurs when the CPU has to wait for data to be retrieved from RAM, causing a delay in the execution of instructions. To minimize memory stalls, it is important to optimize the order in which data is accessed.

Computer System - Memory Hierarchy



Cache

- In computing, a cache is a high-speed data storage layer which stores a subset of data, typically transient in nature, so that future requests for that data are served up faster than is possible by accessing the data's primary storage location.
- Cache is a high-speed data storage layer that stores a subset of data from the main memory. It is used to improve the performance of the CPU by reducing the time it takes to access frequently used data. When a CPU needs to access data, it first checks the cache. If the data is found in the cache, it is retrieved much faster than if it had to be fetched from the main memory.

Drive Technologies

There are different types of storage drives used in computers:

- HDD (Hard Disk Drive): These drives use mechanical components to read and write data. They are slower compared to other storage technologies.
- SSD (Solid State Drive): These drives use flash memory to store data. They are faster and more reliable than HDDs.
- RAM memory: While RAM is faster than HDDs and SSDs, it is still slower compared to cache memory.

More on RAM

- Make sure to not confuse the content of the RAM and its Address, Address are **Not Random**.
- The first (smallest) Address of the bigger chunk of data is where the address is.
- Variables are place holders of an Address in Ram

Example use of Address in programming language:

```
char symb = 'a';
```

0011 0001 (a in ascii) 7 0111

```
int var = 5;
var++;
```

size of int is 4 bytes

| Address | Hex | Binary |
|-----------|-----|--------|
| 0000 0000 | C | 1100 |
| 0000 0000 | D | 1101 |
| 0000 0000 | E | 1110 |
| 0000 0101 | F | 1111 |

```
int *ptr  
ptr = &var;
```

Thing to remember for Midterm

sizes of c++ datatypes

- char = 1 bytes
- short = 2 bytes
- unsigned(int) = 4 bytes
- int = 4 bytes
- signed(int) = 4 bytes
- long = at least 4 bytes
- long long = 8 bytes
- float = 4 bytes
- double = 8 bytes
- bool = 1 bytes

Memory Words in Computing

In computing, memory words refer to the unit of data that can be accessed and manipulated by the CPU. It represents the smallest addressable unit of memory. The size of a memory word varies depending on the architecture and design of the computer system. Common memory word sizes include:

- 8 bits (1 byte)
- 16 bits (2 bytes)
- 32 bits (4 bytes)
- 64 bits (8 bytes)

The size of the memory word affects the maximum amount of data that can be processed and the precision of calculations that can be performed. It is important to consider the memory word size when designing and optimizing software to ensure efficient memory usage and performance.

Data Bus

The data bus is a communication pathway used to transfer data between different components of a computer system. It carries binary information in the form of electrical signals. The width of the data bus determines the number of bits that can be transferred simultaneously.

Address Bus

The address bus is another communication pathway in a computer system. It is used to specify the memory address or I/O port address of the data being transferred. The width of the address bus

determines the maximum memory capacity that can be addressed by the system. It is typically wider than the data bus to accommodate larger memory sizes.

CPU Clock/Frequency

A device the generating synchronization pulses. **CPU Frequency**: measurement of how many synchronization signals a cpu can issue per second

Program Counter

Program counter: is a register inside the CPU that stores the address of the **NEXT** (not Current) instructions to execute.

- Can not contain nothing
- CPU takes instruction from ram memory to Program counter

CPU only knows machine language

Compiler: a program that translates programming languages to machine language Executable: the files that contains instructions for cpu to execute.

- executable files are executed directly by the CPU

Ways of executing a program

Program execution → Compilation Program execution → Interpreted

- interpreter program (runs first) runs on the CPU, this program takes the source code (script) and does what script says