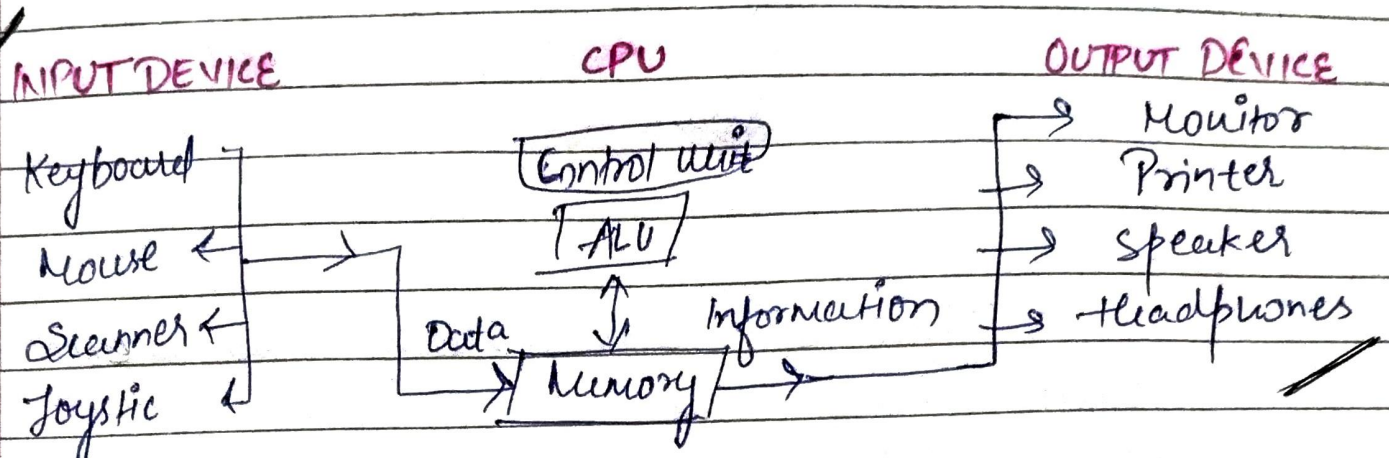
→ Computer Architecture :

Refers to the design, structure and organization of computer's main components. It focuses on how the hardware components interact to process data efficiently.



Key elements → CPU (Central Processing Unit)
→ Cores
→ RAM (Random Access Memory)
→ Cache Memory.

→ CPU [Central Processing Unit]

CPU is the brain of computer. Everything we do - opening apps, typing, watching videos - gets processed by the CPU.

Fetch → Decode → Execute

Picks instructions from memory

understand what the instructions means.

Performs the actions.

→ Basic Terminology :-

1. Instruction set Architecture (ISA):
   Set of instructions that a CPU can understand and execute.
   ISA = Language of CPU.
   It defines
   - → type of data
   - → how memory is accessed
   - → how registers are used.

2. Registers:
   These are small, ultra-fast memory units inside the CPU.
   Types
   - → Accumulator : stores intermediate result
   - → MAR/MDR : Memory address / data
   - → PC (Program counter) : stores address of next instruction
   - → IR (Instruction Register) : Hold current instruction

3. Clock Speed:
   Speed at which CPU executes instructions.
   Measured in
   - → MHz (Megahertz)
   - → GHz (Gigahertz)

4. Throughput & Latency
   - Throughput → the amount of work done in a given time.
   - Latency → delay or time taken to complete one operation.

RAM [short-term memory that data and instructions actively being used by CPU]

cache [store frequently used instructions & data.]

Date: / /
Page No.

→ CPU

ALU (Arithmetic logic unit)
does maths and performs logic (AND/OR/NOT)

CU (control unit)
tells what to do and control how data moves inside the computer

Registers
small but super-fast memory inside CPU.
stores data temporily during processing.

→ CPU cores :-
An independent processing unit inside the CPU.
Older CPU – One core
Moder CPU → More than one (performs multitasking)

Types
- one core (1) rare, slow for multitasking
- Dual core (2) basic multitasking
- Quad core (4) common for laptops
- Octa-core (8) high performance systems.

Benefits
- faster processing
- Less heat
- Improved performance

→ Memory Management System
– The process used by the operating system (OS) to handle, organise, and control computer memory,
* Ensures memory is used effiiently and every program gets enough memory.

Date
Page No.

- **Paging:**
  → Memory is divided into small equal blocks - pages & frames
  → No fragmentation issues.
  ↓
  [Wasted Memory, free but cannot be used efficiently]

- **Virtual Memory:**
  Virtual memory allows the system to use hard disk as extra RAM.
  Benefits → Run large programs
  → Multitasks more efficiently
  → Avoid "out of memory" errors

- **Swapping:**
  Memory management technique where the operating system moves a process from RAM to the hard disk and brings another process into RAM.
  WHY? → RAM is limited & many program may want to run at the same time.

- **Cache Coherence:**
  Means Keeping data consistent across multiple CPU caches.

If a computer has multiple cores, each core has its own cache.

Date / /
Page No.

→ Input / Output systems:

An I/O system allows a computer to communicate with the outside world. It takes input (data from users or devices) and provides output.

Input Devices
- Mouse
- Keyboard
- Scanner

Output Devices
- Monitor
- Printer
- Speakers

Interfaces with external devices:

is the communication link between a computer system and external hardware (Keyboard, disk, USB devices etc).

Helps in data transfer, control, and status communication.

WHY Interfaces:
- External devices work at different speeds
- CPU cannot directly control every device.

* External devices are not controlled directly by the CPU alone. They are controlled together by CPU, devices drivers, and the operating system (OS).

↓

acts as a translator between OS and hardware

→ CPU ↔ Memory ↔ I/O communication

These three units work together to execute programs and handle I/O.

CPU ↔ Memory ↔ I/O

↓ Process instructions and control operations

↓ Stores data and programs temporarily

↓ Handle communication with external devices.

BUSES [communication path]

A bus is a set of wires that connects CPU, memory, and I/O devices.

TYPES:

I. Data Bus →
- Transfers Actual Data
- Bi-directional

II. Address Bus →
- Carries memory or I/O addresses
- One-directional [CP → Memory I/O]
- Determines Max. addressable Memory.

III. Control Bus →
- Carries Control signals
- Manage timing & coordination
- Ex: Read, Write, Interrupt

DMA (Direct Memory Access)

Allows an I/O device to transfer data directly to/from memory without CPU involvement.

Advantages {
- faster data transfer
- Reduces CPU workload

Date : / /
Page No.

Interrupts: A signal sent by device to the CPU to get immediate attention.

Working:

Device sends interrupt signals → CPU pauses current task → CPU executes Interrupt Service Routine (ISR) → CPU resumes previous task.

→ Buses and Data Transfer Basic

Bus is a communication pathway that transfer data, addresses and control signals between computer components like CPU, memory, and I/O devices.

* System Bus: connects CPU, main memory, and I/O devices.

* PCIe (Peripheral component Interconnect Express)
  → High-speed serial expansion bus used to connect modern hardware components.
    - uses serial communication
    - data transfer in lanes
    - Each lane has two pair of wires [send & receive]
    - Support full-duplex communication

* Serial Bus Vs Parallel Bus

features	Serial Bus	Parallel Bus
Data Transfer	1 bit at a time	Multiple bit at a time
Wires required	Few	Many
Speed	High (Modern)	Limited (signal issues)
Interface	Low	High
Examples	PCIe	System Bus, Old PCIe

Date / /
Page No.

WHY Serial Buses Are preferred TODAY
- Less Noise & interference
- Higher clock speeds
- Better scalability
- Lower cost and power usage

→ Introduction to Parallel Computing

It is a computing technique where a problem is divided into multiple smaller tasks that are executed simultaneously on multiple processors or cores to get faster results.

WHY :
- Faster execution - Tasks run at the same time.
- Divide tasks across cores.

TYPES:

DATA Parallesim Means performing the same operation on different pieces of data at the same time.

TASK parallesim Means executing different tasks or functions in parallel.

→ Multi-Cores Processors & GPUs →

↓

A multi core processor is a single CPU chip that contains two or more cores, where each core can execute instructions independently.

GPU is a processor with thousands of small cores designed to perform many simple operations in parallel.

Date / /
Page No.

DIFFERENCES:-

Features	Multi-Core CPU	GPU
AI (uses)	Decision-Making and control logic	fast parallel computations for neural network
ML (uses)	Manages data, instructions and workflow	Trains models using large Matrix calculation
Rendering (uses)	prepares the scene and sends instructions	draws pixels, lightening, shadow in Parallel.
Gaming (uses AI + rendering)	handles game rule, physics, and character AI.	renders graphics and frames.