UFV

UFV.ES

Hardware

MODULE 2 / UNIT 4 / 0.3

MOISES M. MARTINEZ

FUNDAMENTALS OF COMPUTER ENGINEERING

2024/25

What is a computer?

A computer

A computer is a digital electronic device designed to autonomously execute programmed sequences of arithmetic and logical operations, commonly known as computations.

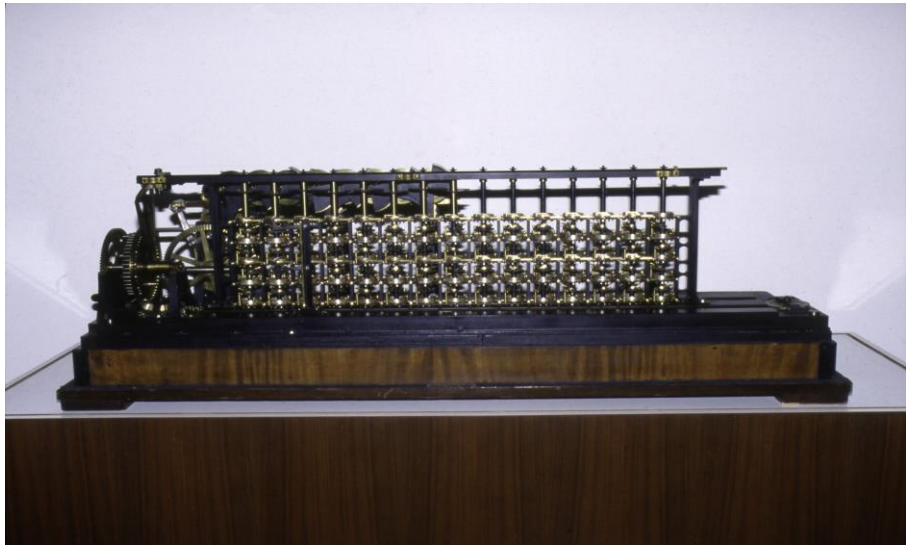
A computer system, however, refers to a complete computing setup that includes the essential hardware components, the operating system (core software), and the necessary peripheral devices, all working together to ensure full functionality.

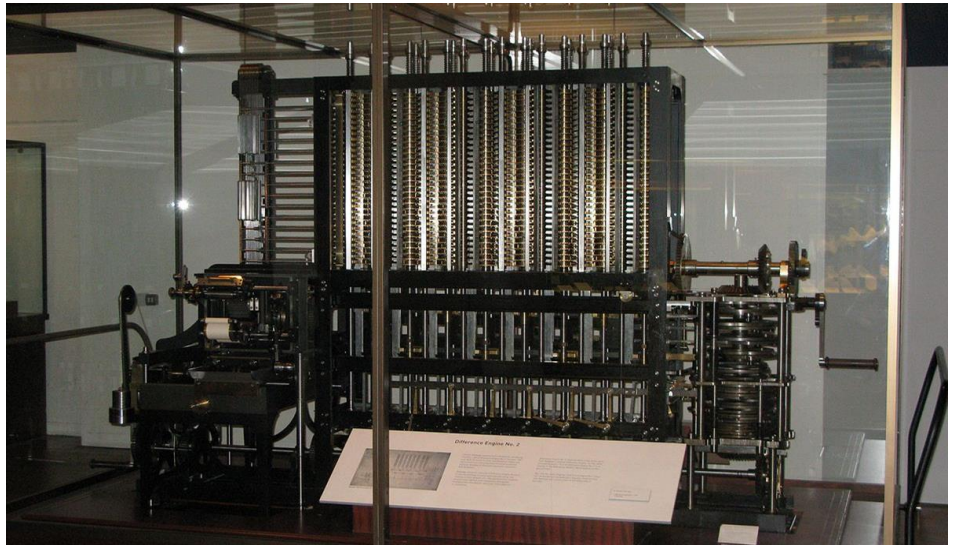
The evolution of the computers

01

From past till now …

In 1853, Per Georg Scheutz and his son Edvard designed what is regarded as the world's first printing calculator. This innovation is noteworthy as it is considered a precursor to the modern computer, as it possessed the capability to compute tabular differences and print the resultant calculations.





A mechanical calculator designed by Georg Scheutz.

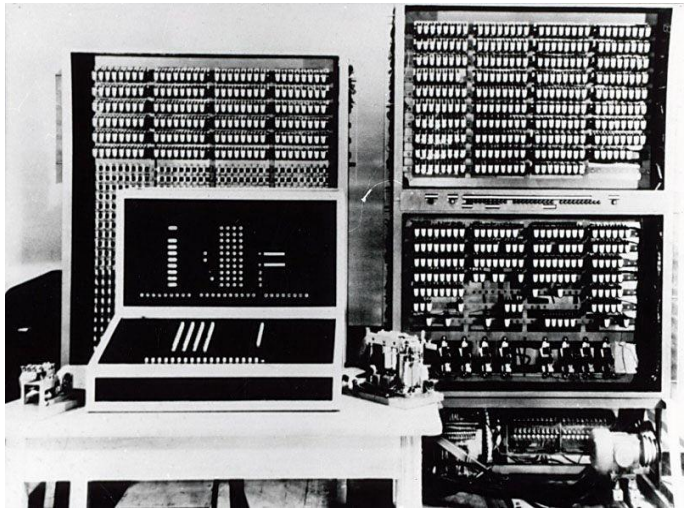
From past till now …

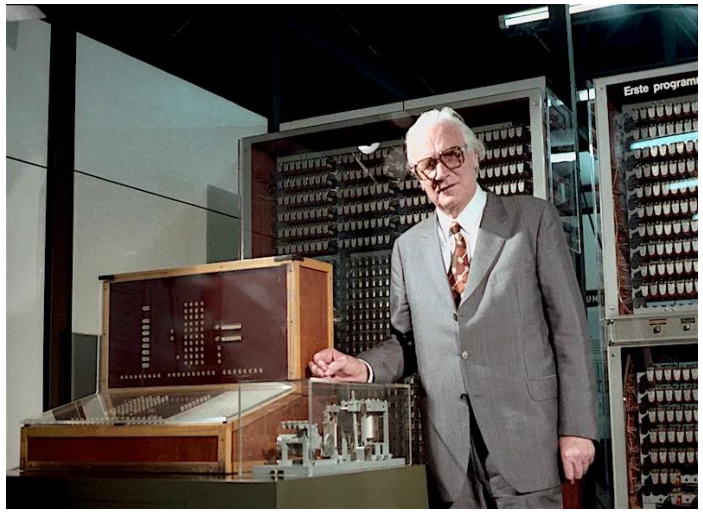
In 1931, Vannevar Bush designed and constructed the **Differential Analyzer**, the first large-scale, automatic, general-purpose mechanical analogy computer.

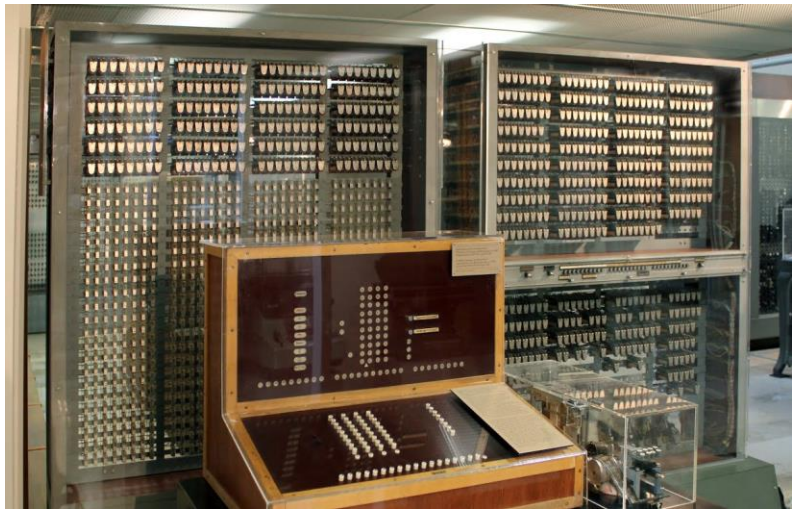
Vannevar Bush

From past till now …

The Z3, an early computing machine developed by German engineer Konrad Zuse in complete isolation from other contemporaneous advancements, featured 2,300 relays, executed floating-point binary arithmetic operations, and had a word length of 22 bits.



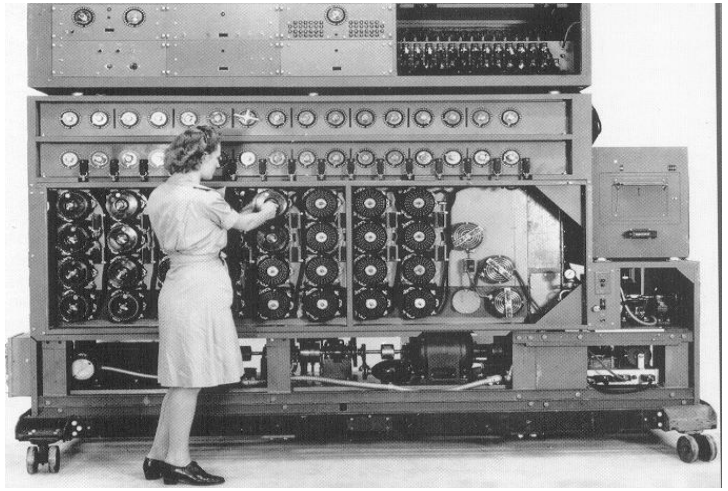




The Z3 was destroyed in a bombing raid on Berlin in late 1943.

From past till now …

The cryptographic device, commonly referred to as **the bomb** or **the phoenix**, was conceived as an electro-mechanical mechanism for deciphering encrypted Nazi military communications generated by the ENIGMA machine during the second World War.

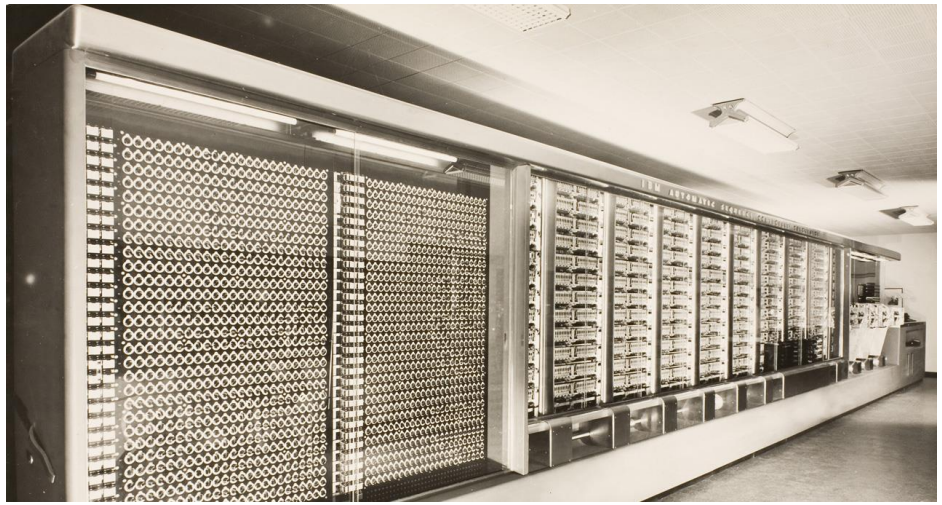


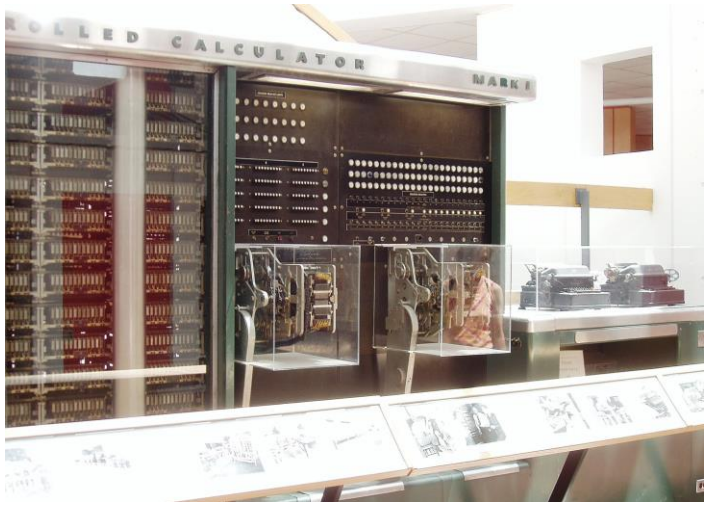
The foundational concept behind the development of the bombes was derived from the Polish code-breaker Marian Rejewski's 1938 invention, known as the Bomba.

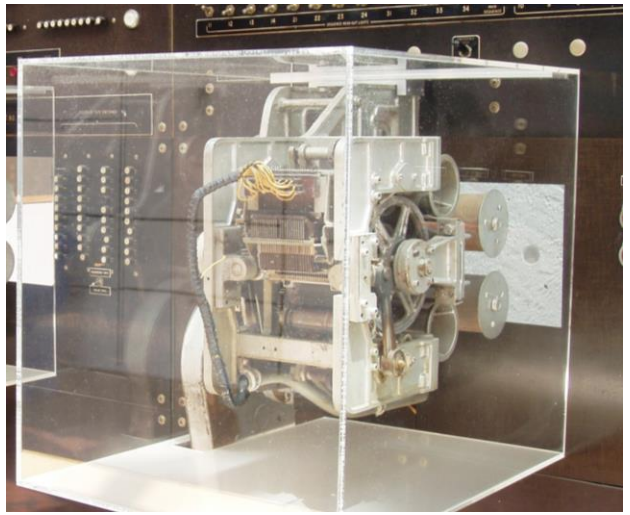
The evolution of the computers

From past till now …

The Harvard Mark 1 computer, also known as the IBM Automatic Sequence Controlled Calculator (ASCC), was an electromechanical general-purpose computing device employed in wartime endeavours during the latter stages of World War II.



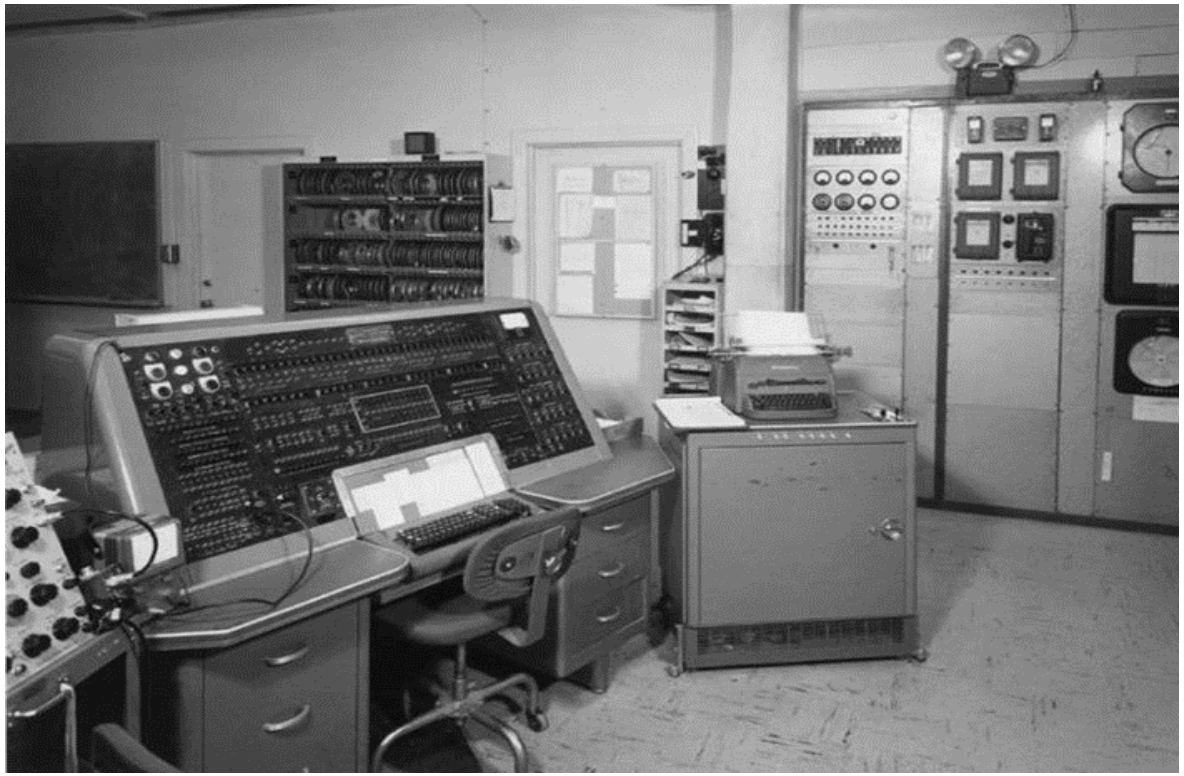




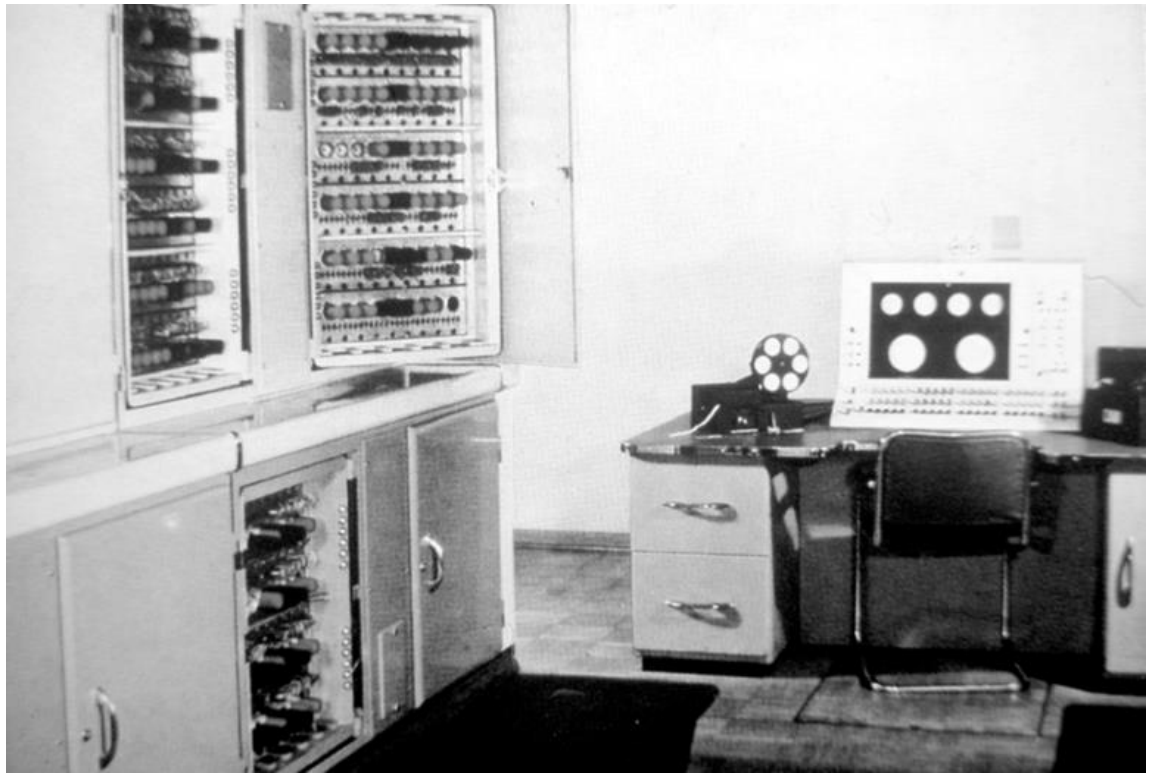
The machine featured a fifty-foot (15 meters) long camshaft spanning the entire length of the apparatus, serving to synchronize the numerous components of the machine.

From past till now …

Different types of computers were built between 1945 and 1959 …



UNIVAC (UNIVersal Automatic Computer)



Ferranti Mark I

The evolution of the computers

From past till now …

A **central computer**, often referred to as a mainframe or 'iron,' is a computing system employed by large organizations to handle mission-critical applications, extensive data processing, enterprise resource planning, and high-volume transactions. These computers operate have three main features:

- Time-sharing systems.
- Centralized resources.
- Accessed through straightforward terminals. However, these systems often feature user interfaces characterized as less user-friendly.

The evolution of the computers

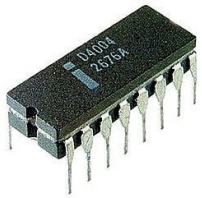
From past till now …

The first **microprocessors** were created late 1970s. These integrated circuits were responsible for executing instructions in binary language, performing fundamental arithmetic and logical operations such as addition, subtraction, multiplication, division, binary logic, and memory access. It consists of a minimum of two fundamental components:

- Arithmetic Logic Unit (ALU): A digital circuit responsible for executing arithmetic and logical operations between values.
- Register Bank: A collection of high-speed, low-capacity memory registers.

The inaugural microprocessor, the Intel 4040, was developed in 1971 by Intel for a calculator, functioning as a 4-bit Central Processing Unit (CPU) that incorporated both an ALU and a register bank on a single chip.

Contemporary generations of microprocessors deliver computing power equivalent to the mainframes that were in use three to four decades ago, at a fraction of the cost, often less than 1/1000th the original price.



Intel 4040

Intel Pentium D

Intel Pentium G7400

Intel I9-9900K

The evolution of the computers

From past till now …

During the 1980s, a transformative shift emerged towards the adoption of personal computers in the workplaces.

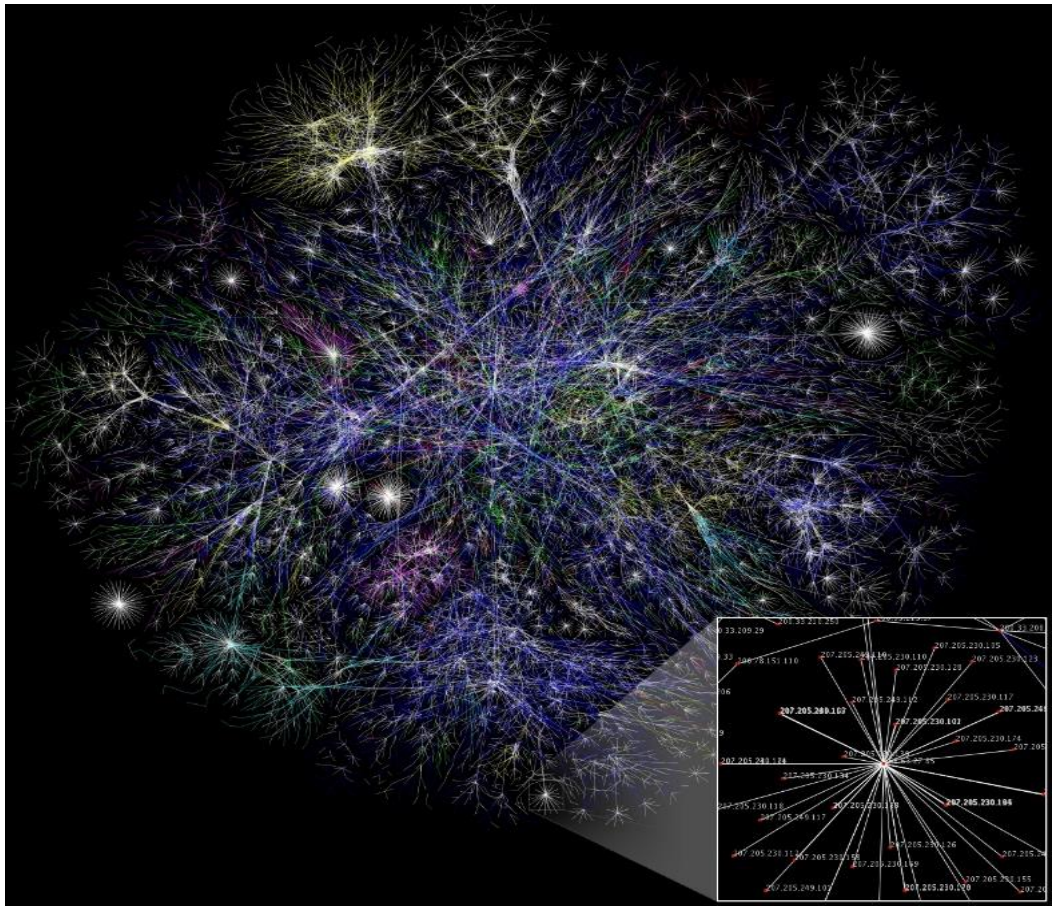
A personal computer (PC) is a programmable digital device designed to execute a sequence of commands for processing input data, resulting in the generation of information that is then routed to output devices.

- The utilization of individual workstations (PCs) is witnessing a growing trend.
- Increasingly sophisticated and user-friendly local execution applications are emerging.
- Local Area Networks (LANs) are implemented for the purpose of information sharing.

From past till now …

The 1990s witnessed a pivotal moment with the advent of the Internet. The Internet network comprises a decentralized network of interconnected networks employing the TCP/IP protocol suite, ensuring that the diverse physical networks forming it function as a unified global network. Its inception dates back to 1969, marked by the establishment of ARPANET, which facilitated the initial interconnection of computers situated in various locations, connecting three universities in California.

- There was a substantial proliferation of Client/Server-based applications, accompanied by a widespread dissemination of applications and services following the emergence of the first websites.
- Novel service applications emerged that relied on the Internet for functionality, including email, e-commerce, and supercomputing, among others.



The evolution of the computers

From past till now …

At the outset of the new century, there was a surge in the widespread utilization of the Internet, precipitating the emergence of distributed systems to accommodate applications and data.

This development gave rise to a novel breed of computers, designed to provide new applications, services, and enhanced capabilities:

- Mobile computers.
- Quantum computers.

Types of computers

02

Types of computers

Computers are commonly categorized into six distinct groups based on their computing power:

- Supercomputers.
- Mainframes (referred to as **iron**).
- Servers.
- Personal computers.
- Mobile computers.
- Quantum computers.

Types of computers

Supercomputers

A **supercomputer** is a highly advanced computing system that surpasses the performance capabilities of typical general-purpose computers. The evaluation of a supercomputer's performance is typically quantified in terms of floating-point operations per second (FLOPS), rather than million instructions per second (MIPS). Supercomputers find application in a variety of domains, including:

- Genome Studies.
- Climate Simulations.
- Military Simulations.
- Medical Simulation and Computation.
- Deep Artificial Intelligence (AI).

https://www.top500.org

Since 2017, supercomputers with the capability to perform in excess of 10^17 FLOPS (equivalent to one hundred quadrillion FLOPS, 100 peta FLOPS, or 100 PFLOPS) have been in existence.

Types of computers

Mainframes (Iron)

A **mainframe** computer, colloquially known as **big iron**, is a computing system primarily employed by large organizations to execute mission-critical functions. These functions encompass extensive data processing tasks, including but not limited to censuses, industry and consumer statistics, enterprise resource planning, and high-volume transaction processing.

Modern mainframes exhibit the capability to concurrently operate multiple distinct instances of operating systems through the use of virtual machines.

A Virtual Machine (VM) represents a computational resource that relies on software rather than physical hardware to execute programs and deploy applications. Multiple virtual 'guest' machines can coexist on a single physical 'host' machine, each of which runs its own autonomous operating system and operates independently from other VMs.

Types of computers

Mainframes (Iron)

A **mainframe** computer, colloquially known as **big iron**, is a computing system primarily employed by large organizations to execute mission-critical functions. These functions encompass extensive data processing tasks, including but not limited to censuses, industry and consumer statistics, enterprise resource planning, and high-volume transaction processing.

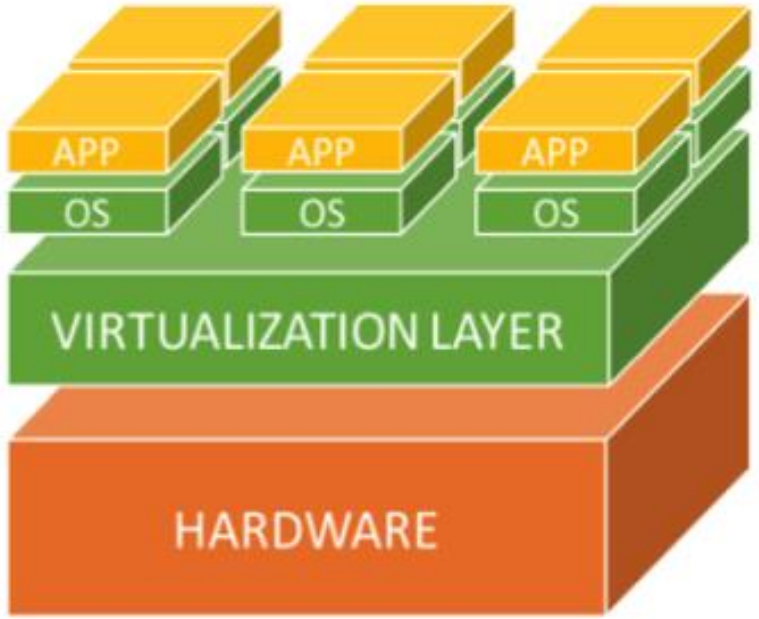
Modern mainframes exhibit the capability to concurrently operate multiple distinct instances of operating systems through the use of **virtual machines**. A Virtual Machine (VM) represents a computational resource that relies on software rather than physical hardware to execute programs and deploy applications. Multiple virtual 'guest' machines can coexist on a single physical 'host' machine, each of which runs its own autonomous operating system and operates independently from other VMs.

Types of computers

OFF-EXAM CONTENT

Mainframes (Iron) - Virtualization

Virtualization employs software to establish an abstraction layer over the physical hardware of a computer system, facilitating the subdivision of hardware components—such as processors, memory, storage, and others—into distinct virtual entities, typically referred to as virtual machines (VMs). Each VM operates with its own dedicated operating system (OS) and functions autonomously, resembling an independent computer system, albeit utilizing a portion of the actual underlying hardware resources.



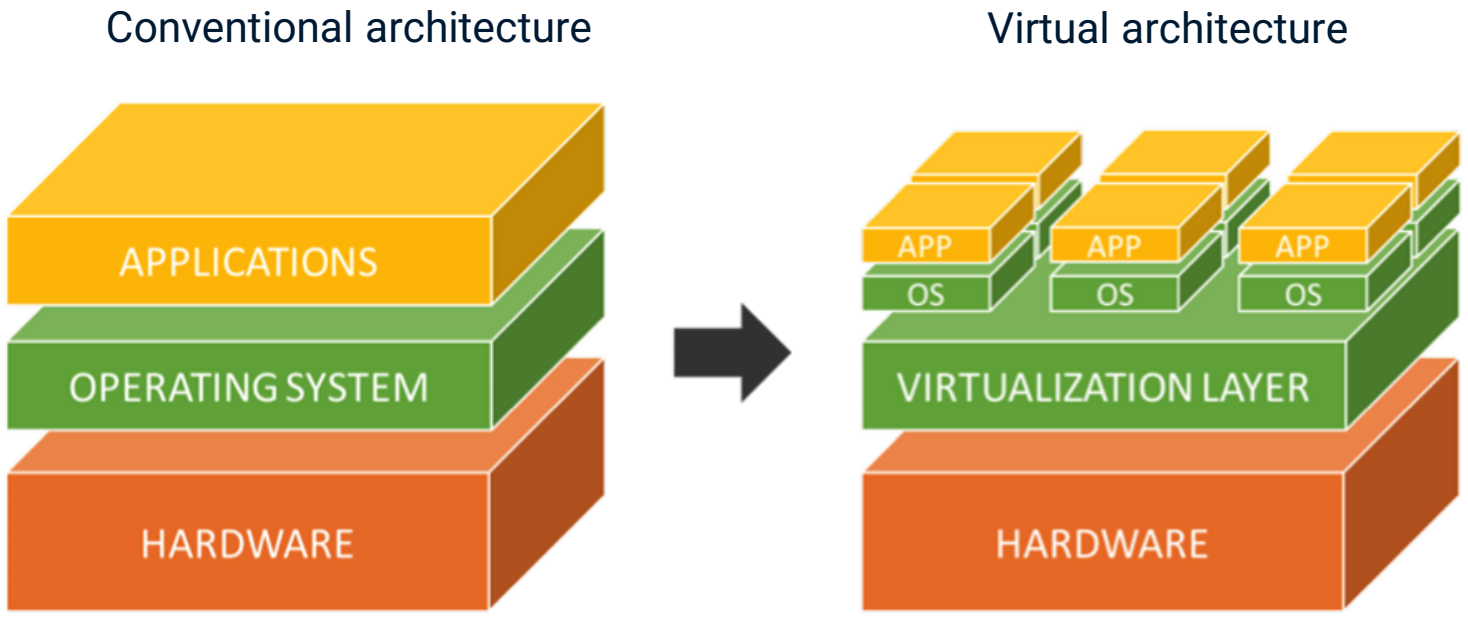
The benefits of virtualization encompass:

- Enhanced Resource Efficiency.
- Simplified Management.
- Reduced Downtime.
- Expedited Provisioning.

Types of computers

OFF-EXAM CONTENT

Mainframes (Iron) - Virtualization



The principal distinction between conventional architecture and virtual architecture resides in the virtualization layer, situated between the hardware and the operating system (OS). Consequently, a Virtual Machine Monitor (VMM) assumes responsibility for overseeing resource allocation and multiplexing to furnish resources for each virtual machine (VM).

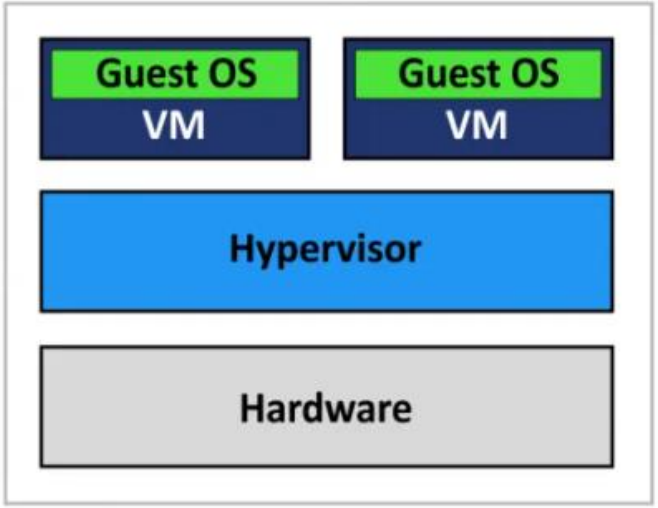
Types of computers

OFF-EXAM CONTENT

Mainframes (Iron) - Virtualization

A **hypervisor** constitutes the software layer responsible for orchestrating virtual machines (VMs). It functions as an intermediary between the VM and the underlying physical hardware, guaranteeing that each VM has access to the requisite physical resources for execution. Moreover, it maintains the isolation of VMs, preventing interference with one another in terms of memory allocation and computational processing. Hypervisors can be categorized into two primary types:

- **Type 1**, also known as **bare-metal**, **hypervisors** directly interact with the underlying physical resources, essentially supplanting the traditional operating system. These are predominantly deployed in virtual server environments.



- Virtual PC o Virtual Server
- Vmware Workstation
- KVM
- VirtualBox

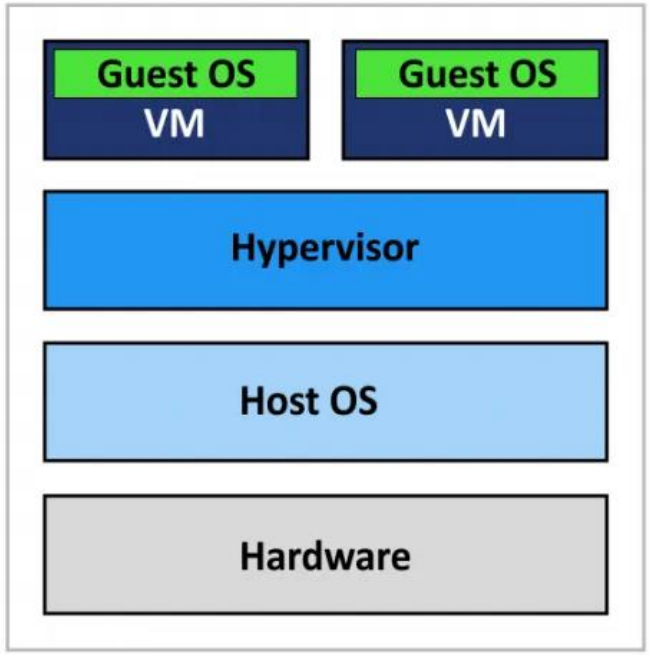
Types of computers

OFF-EXAM CONTENT

Mainframes (Iron) - Virtualization

A **hypervisor** constitutes the software layer responsible for orchestrating virtual machines (VMs). It functions as an intermediary between the VM and the underlying physical hardware, guaranteeing that each VM has access to the requisite physical resources for execution. Moreover, it maintains the isolation of VMs, preventing interference with one another in terms of memory allocation and computational processing. Hypervisors can be categorized into two primary types:

- Type 1, also referred to as bare-metal, hypervisors directly interact with the underlying physical resources, essentially supplanting the traditional operating system. These are predominantly deployed in virtual server environments.
- **Type 2 hypervisors** operate as applications within an existing operating system. They are commonly employed on endpoint devices to facilitate the execution of alternative operating systems. However, they introduce a performance overhead since they rely on the host OS to access and manage the underlying hardware resources.

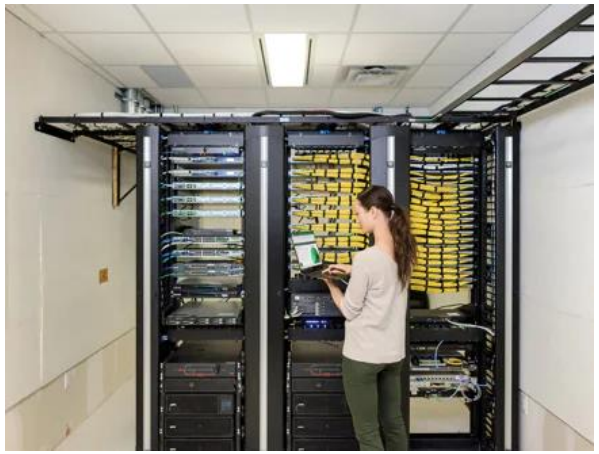


- Hyper-V
- VMware ESX

Types of computers

Servers

Servers bear a resemblance to mainframes, albeit they are typically found within departmental or small enterprise settings. The primary role of a server is to oversee network resources, encompassing responsibilities such as website hosting, data transmission, email exchange, access control, and more. Some prevalent categories of servers include:



- Database Servers.
- File Servers.
- Web Servers.
- Mail Servers.
- Application Servers.

Types of computers

Personal computer

A Personal Computer (PC) represents a versatile microcomputer characterized by dimensions, functionalities, and affordability that render it suitable for individual utilization.

These computers are expressly designed for operation by end-users themselves, without necessitating the involvement of computer specialists or technicians. Unlike larger, expensive minicomputers and mainframes, personal computers do not employ time-sharing among multiple users concurrently.

Types of computers

Mobile computer

Mobile computers represent fully independent computing devices that incorporate all essential peripherals and interfaces required for the collection, management, and storage of data.

These computers condense within a compact form factor all the essential computer components, including a display, Wi-Fi connectivity, an operating system (such as iOS, or Android), a keyboard (virtual or physical), a battery, cameras, and various sensors.

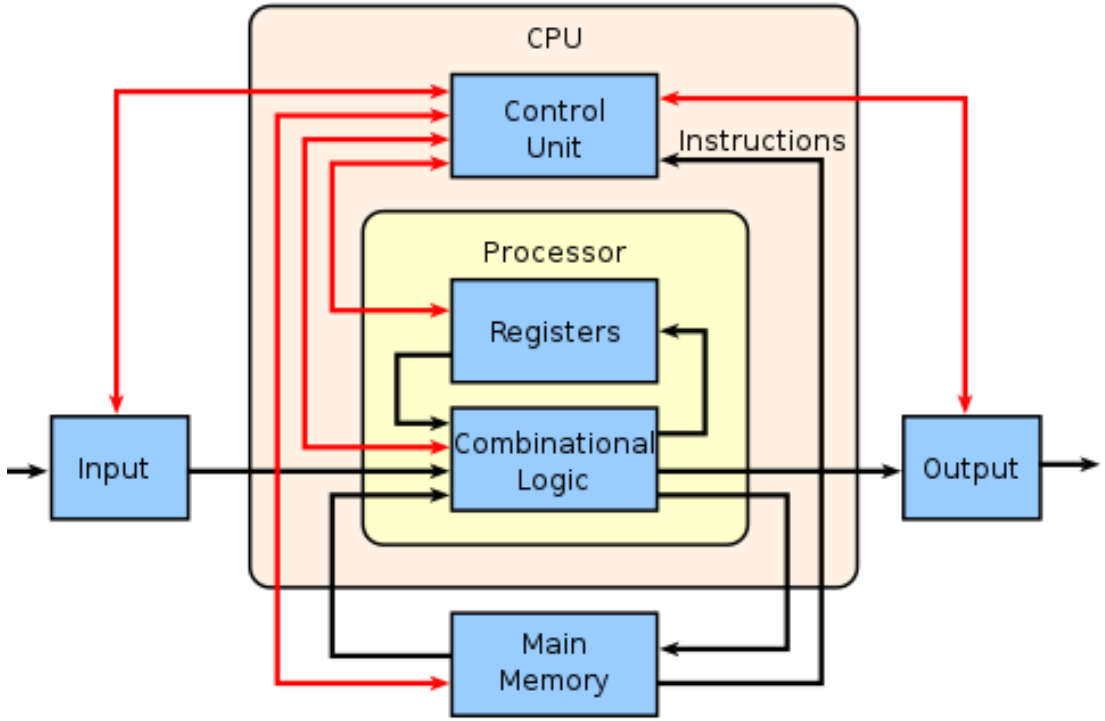
Computer architecture

03

Computer architecture

Computer architecture comprises a collection of principles and methodologies that delineate the operational, structural, and implementation aspects of computer systems. Among the foremost considerations are:

- Neumann architecture conceptualized by John von Neumann in 1945.
- Electronic Calculator conceptualized by Alan Turing in 1945.

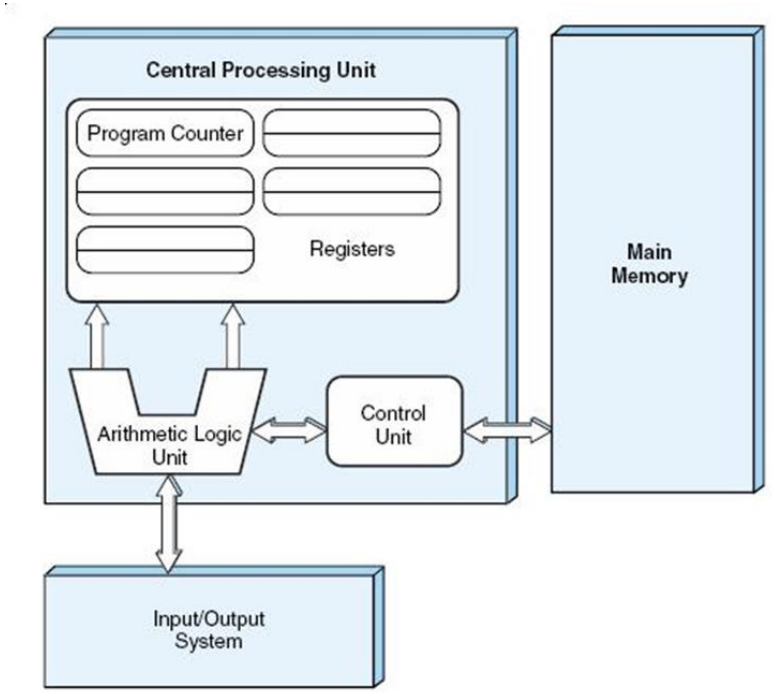


Another noteworthy innovation in the realm of computer architectures is the Analytical Engine, a creation attributed to the collaborative efforts of Charles Babbage and Ada Lovelace.

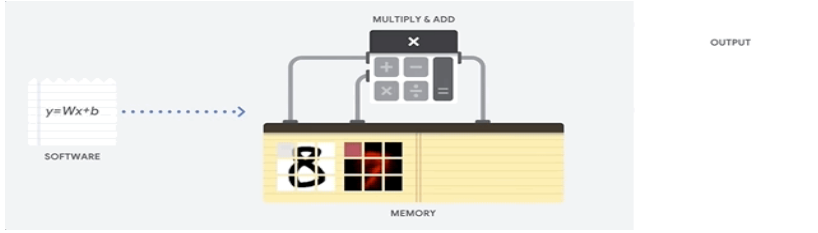
Computer architecture

The Von Neumann architecture, also known as the IAS (Institute for Advanced Study) architecture, consists of three primary components:

- Central Processing Unit (CPU) which is responsible for processing information by executing instructions.
- Main Memory which serves as the storage for both data and instructions.
- Input/Output system which enables communication with external devices.

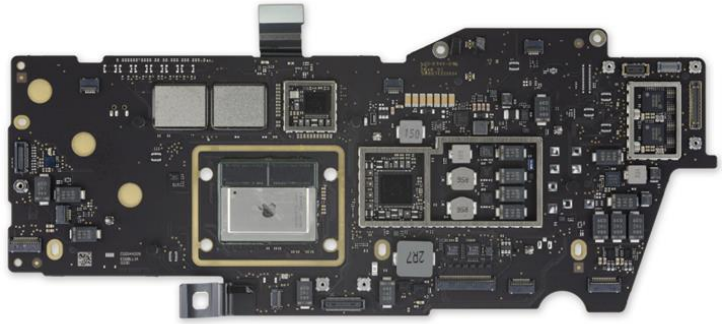


Computing systems based on this architecture are capable of executing a sequence of basic instructions, commonly known as machine code instructions, which are fetched from and executed within main memory.

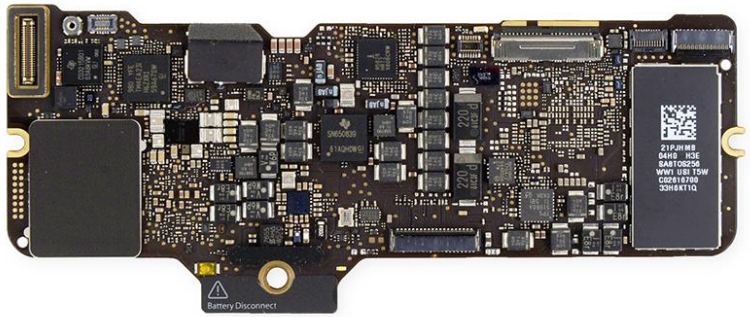


Computer architecture

MSI - MPG X570 GAMING PLUS

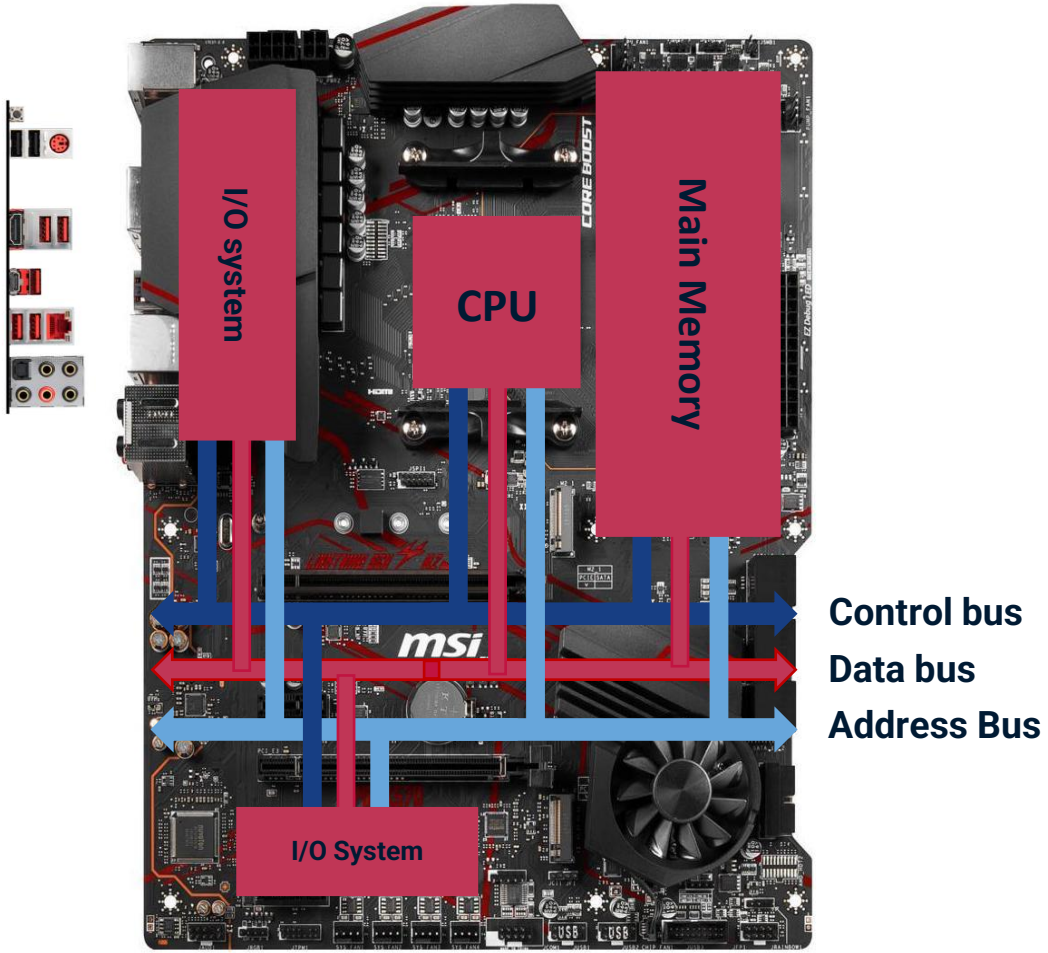


Mac Book Pro M1



Mac Book Pro M3

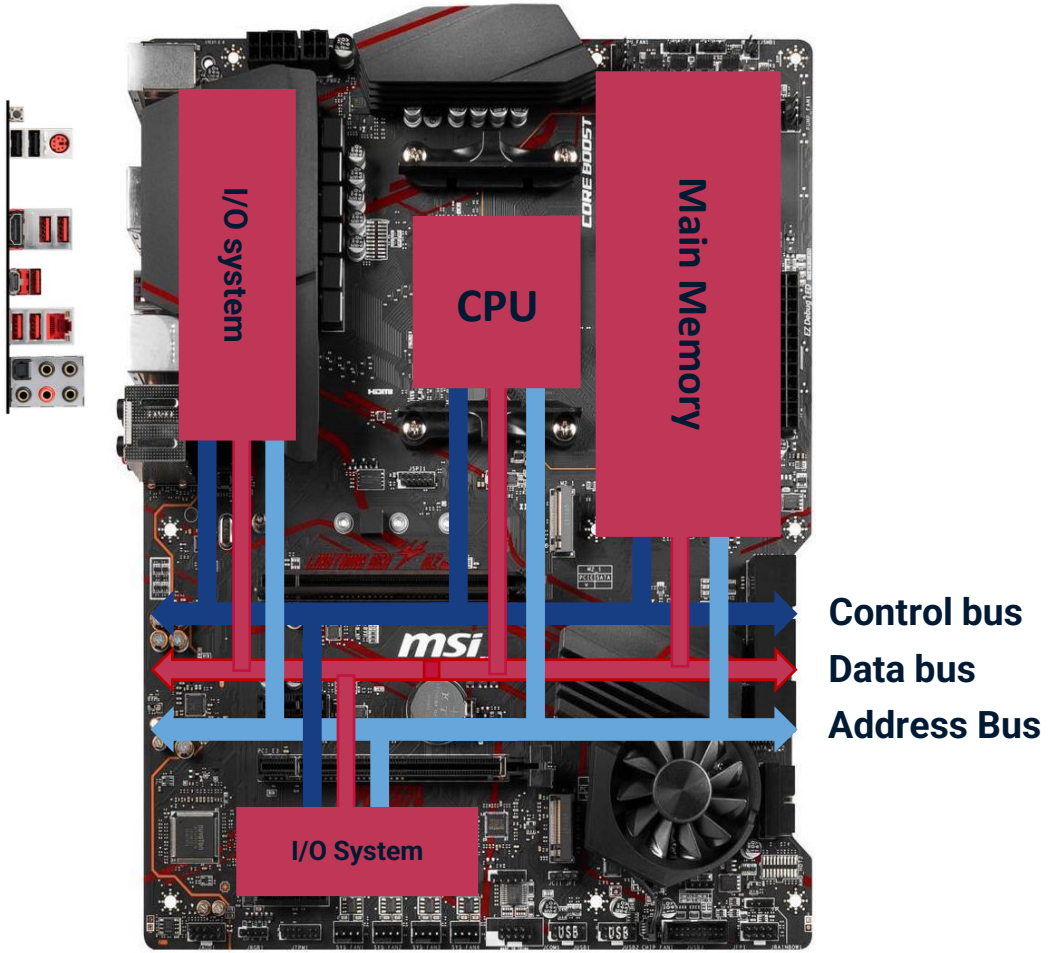
Computer architecture



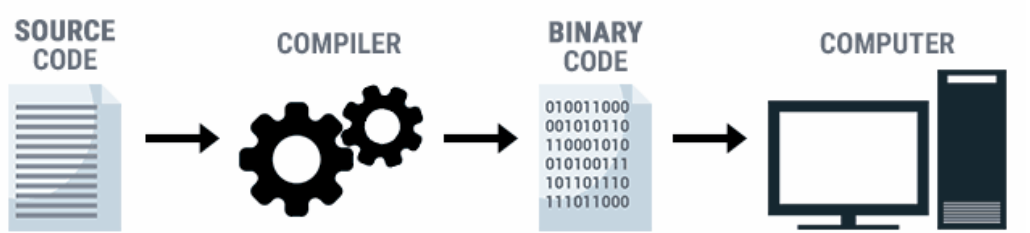
A **motherboard**, also known as the system board, is the central printed circuit board (PCB) in a computer or other electronic device. It acts as the backbone that connects all the different components of the system, allowing them to communicate and function together.

- The motherboard provides connectors for the CPU (Central Processing Unit), memory (RAM), storage devices and expansion slots (such as PCIe slots) allow users to add extra components like graphics cards, sound cards, network cards, and more.
- The motherboard contains the BIOS (Basic Input/Output System) or UEFI (Unified Extensible Firmware Interface) firmware, which initializes hardware during the boot process and provides an interface for configuring system settings.

Computer architecture

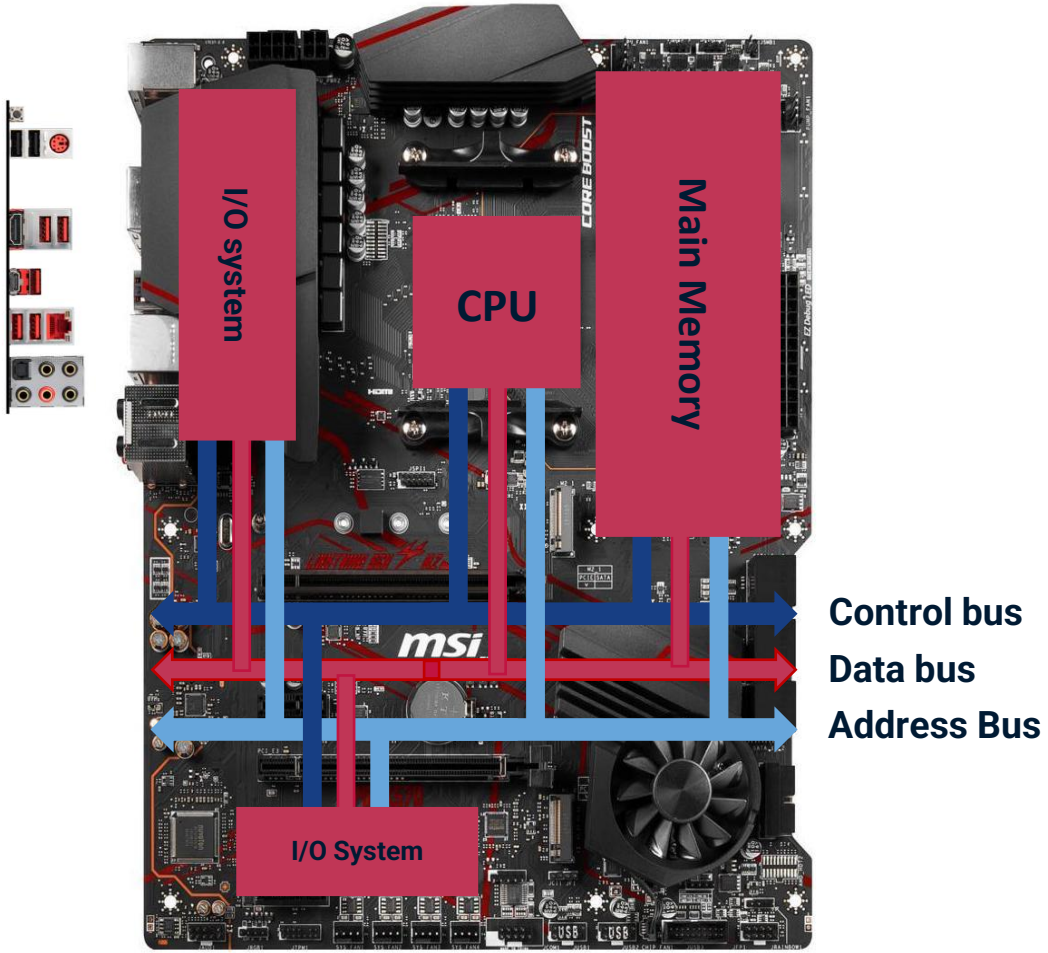


Data and instructions must be inserted into the system via Input/Output (I/O) systems and storage devices.



Programs (source code) are converted into binary code to enable execution within the CPU.

Computer architecture



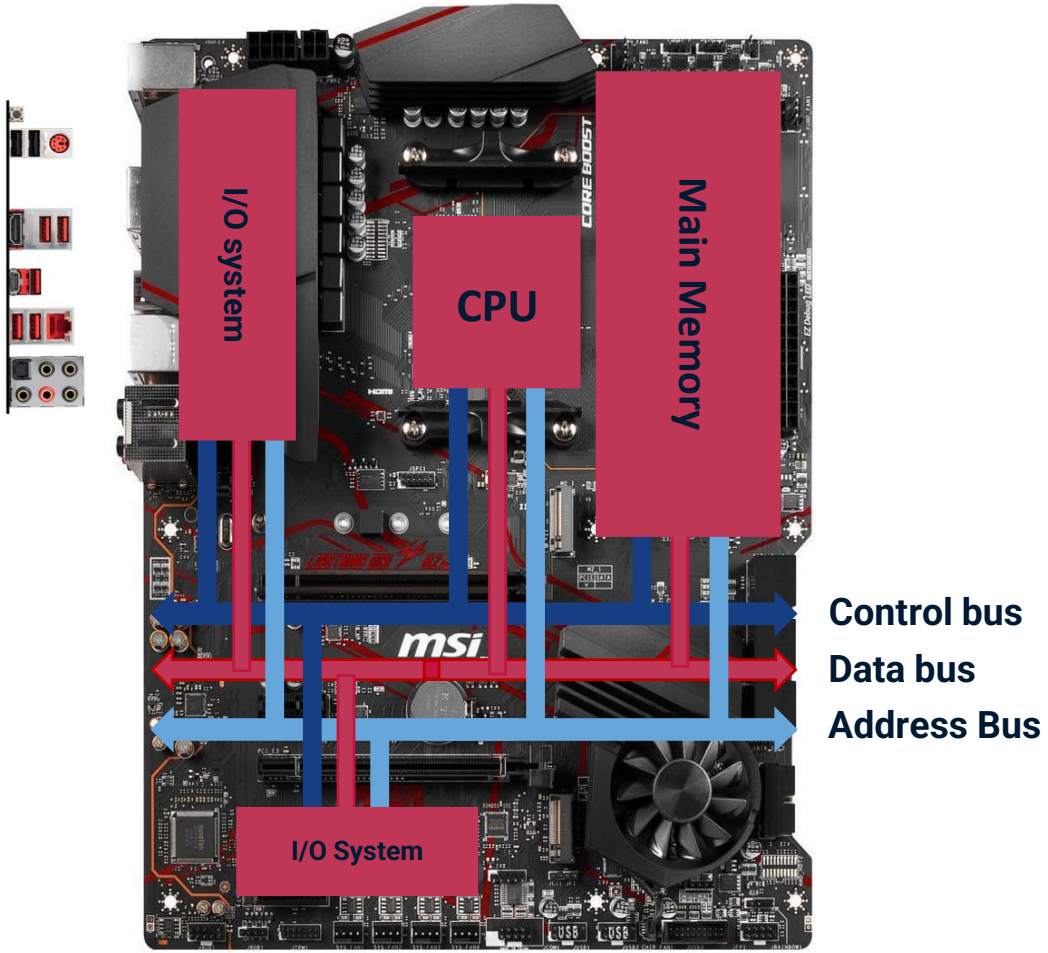
Data and instructions, represented as sequences of bits, are temporarily held in the main memory, organized linearly.

A **word** denotes the quantum of information read or written during each memory access, and these words are accessed through unique addresses.

The **word size** refers to the number of bits utilized to represent the fundamental unit of information within the computer's components.

In contemporary computing systems, the standard word sizes typically range from 16, 32, to 64 bits.

Computer architecture



Control bus
Data bus
Address Bus

The word size, or a multiple thereof, serves as the basis for defining the width of various components within the computer:

- Memory addresses.
- Records.
- Integer and floating point numbers.
- Instructions.

64-bit processors have the theoretical capacity to address up to 16 exabytes of memory addresses, whereas 32-bit processors are limited to addressing only 4 gigabytes of RAM.

2^{16} = 65,536 bytes

2^{32} = 4,294,967,296 bytes = 4 GBs of memory

2^{64} = 18,446,744,073,709,551,616 bytes

What do 16-bit, 32-bit, and 64-bit mean?

A bit is the most basic unit of data in computing, representing either a 0 or a 1. When we refer to 16-bit, 32-bit, or 64-bit, we are describing how many bits the computer can process or manage in one unique operation.

- **16-bit Architecture**: If a computer is using a 16-bit architecture, it means that the data it processes, the instructions it executes, and the memory addresses it handles are all limited to 16 bits at a time.
- **32-bit Architecture**: If a computer is using a 32-bit architecture, it means that the data it processes, the instructions it executes, and the memory addresses it handles are all limited to 32 bits at a time.
- **64-bit Architecture**: If a computer is using a 64-bit architecture, it means that the data it processes, the instructions it executes, and the memory addresses it handles are all limited to 64 bits at a time.

What happens if the data is larger than what the computer can handle in one go?

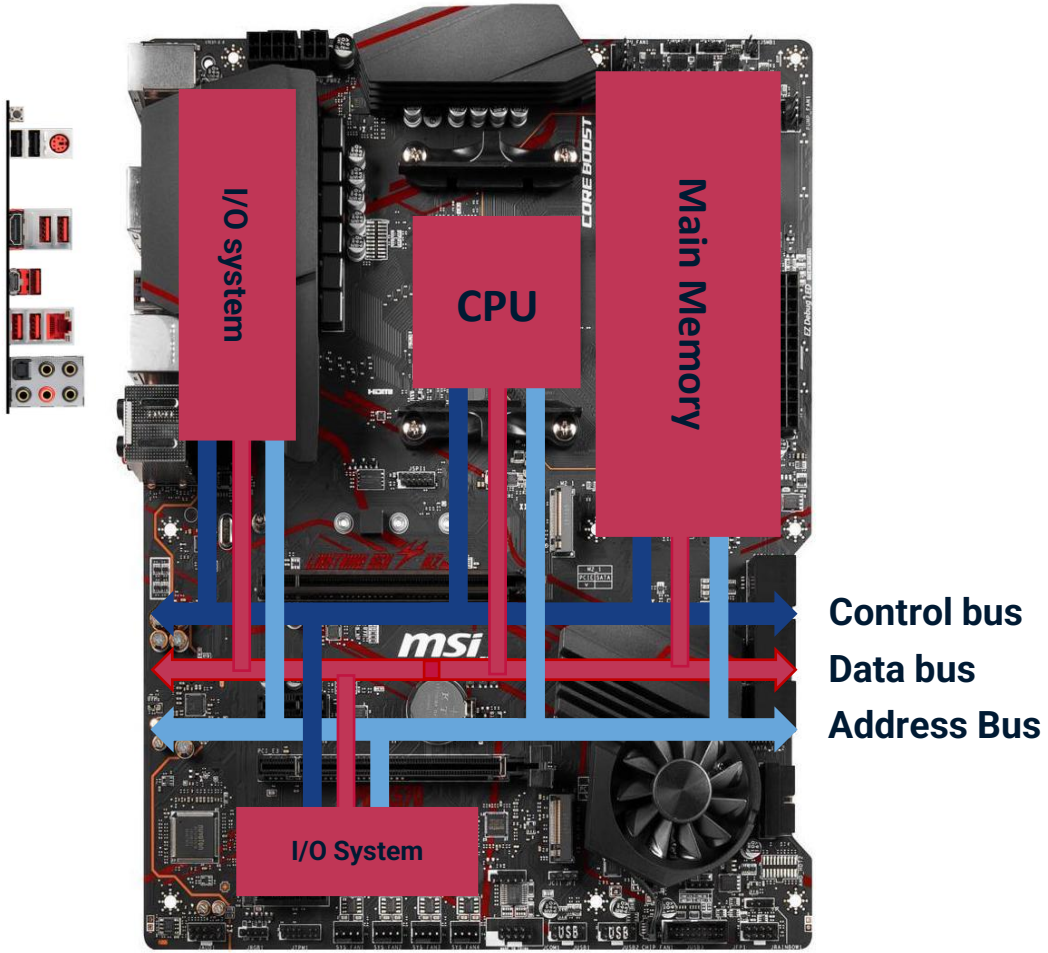
Imagine we are working with a system where **the data to be processed is 128 bits long**, but your computer's architecture is only capable of handling 32-bit data at a time. Segmentation can help manage this larger data efficiently.

Since the computer can only handle 32 bits at a time, **the 128-bit data would be divided into four segments of 32 bits each**. These segments could be organized as follows:

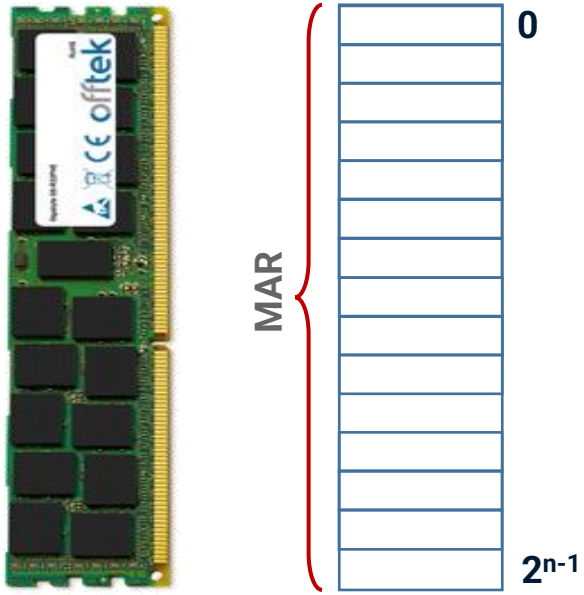
- Segment 1: Contains the first 32 bits of the data.
- Segment 2: Contains the next 32 bits.
- Segment 3: Contains the following 32 bits.
- Segment 4: Contains the final 32 bits.

To process the entire 128-bit data, the CPU would sequentially access each of these segments. The segment registers would provide the base addresses for each 32-bit segment, and the CPU would use offsets to access specific parts of each segment.

Computer architecture



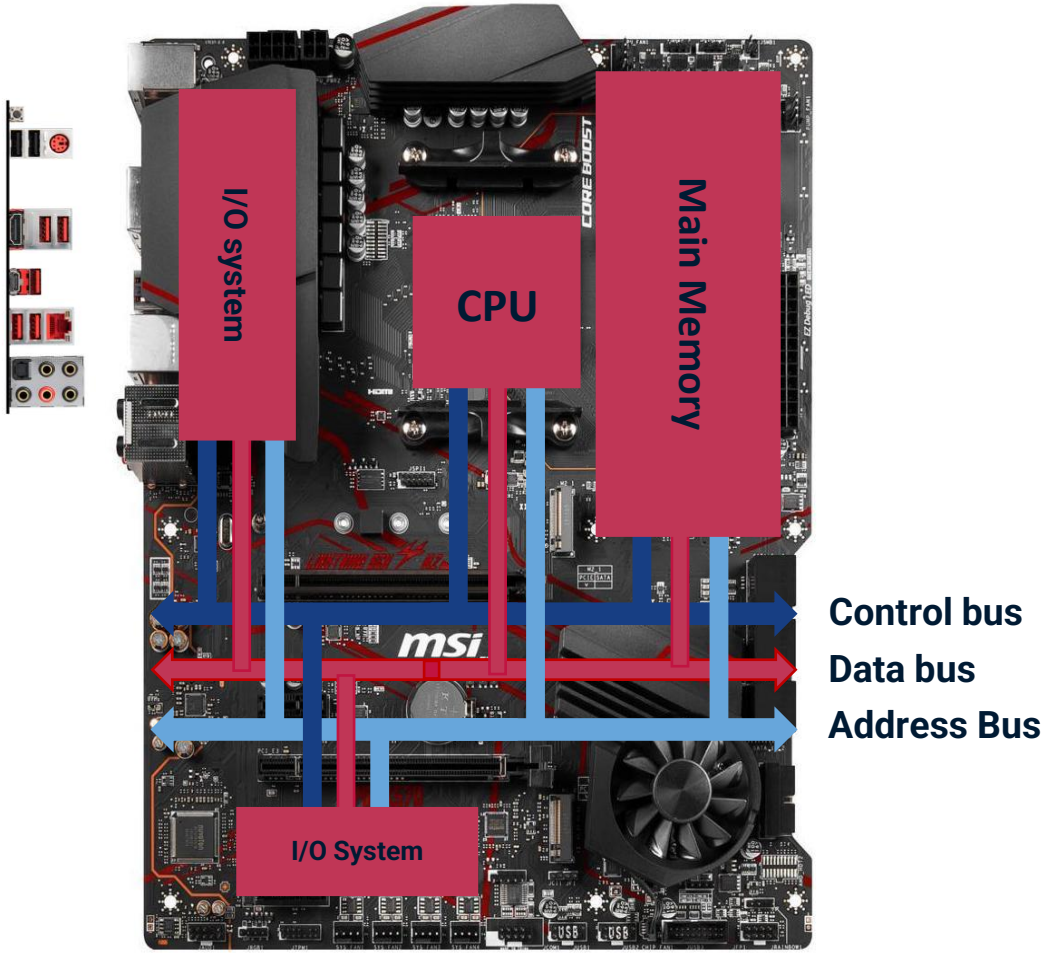
Data and instructions, represented as bit sequences, are temporarily stored in main memory using a linear word-based structure.



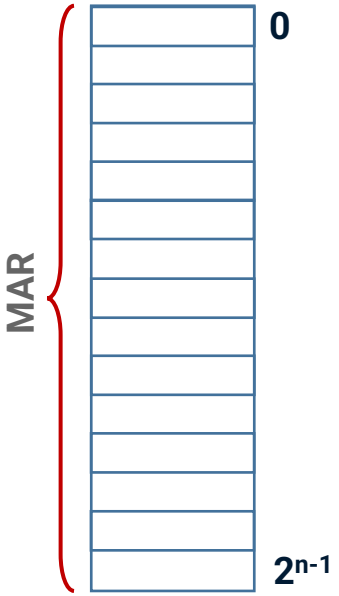
Mar = Memory Address Register.

Positions = 2^n records

Computer architecture



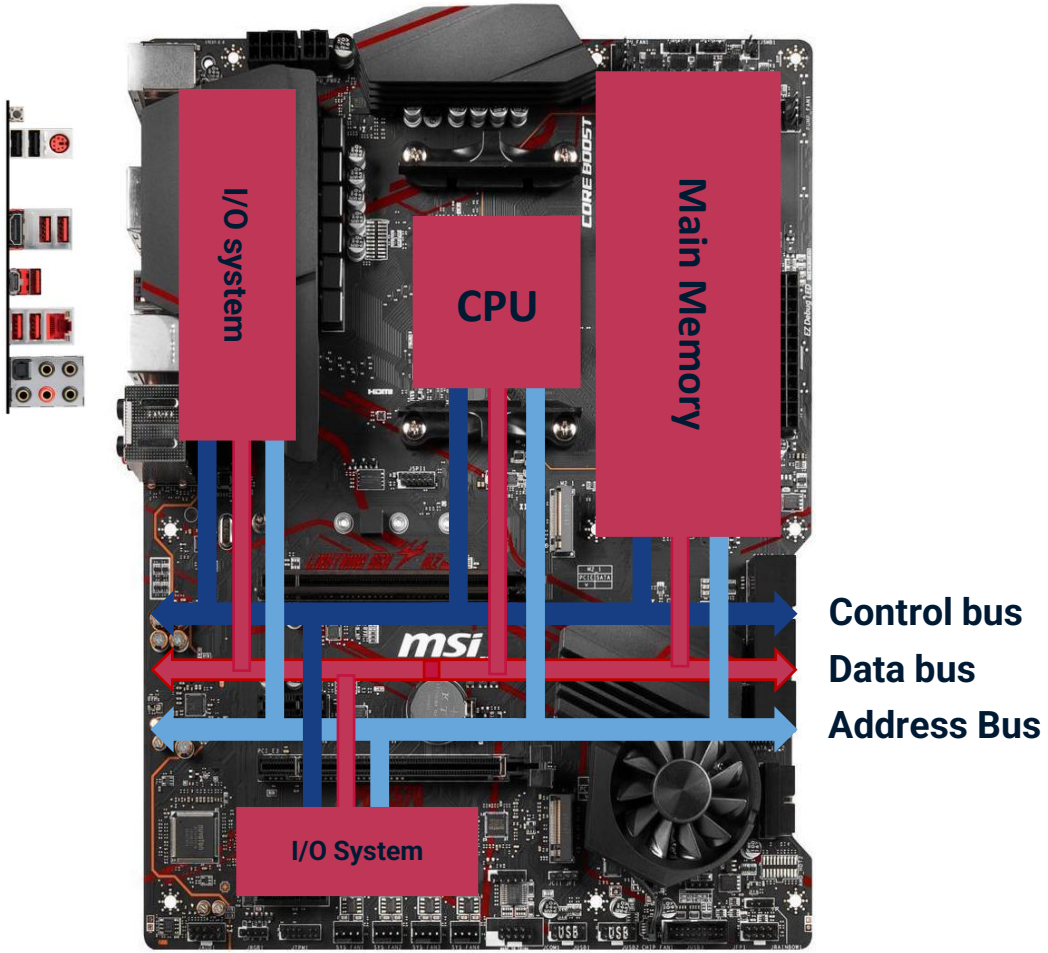
Data and instructions, represented as bit sequences, are temporarily stored in main memory using a linear word-based structure.



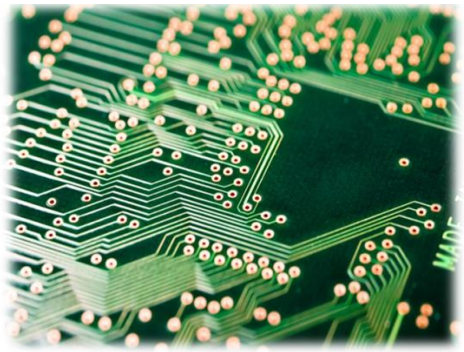
Control signals:

- **R - Read**
- **W - Write**

Computer architecture



Buses represent the pathways or communication channels connecting various computer components to facilitate information transmission. Buses consist of multiple communication lines, with each line responsible for transmitting a single bit of data.

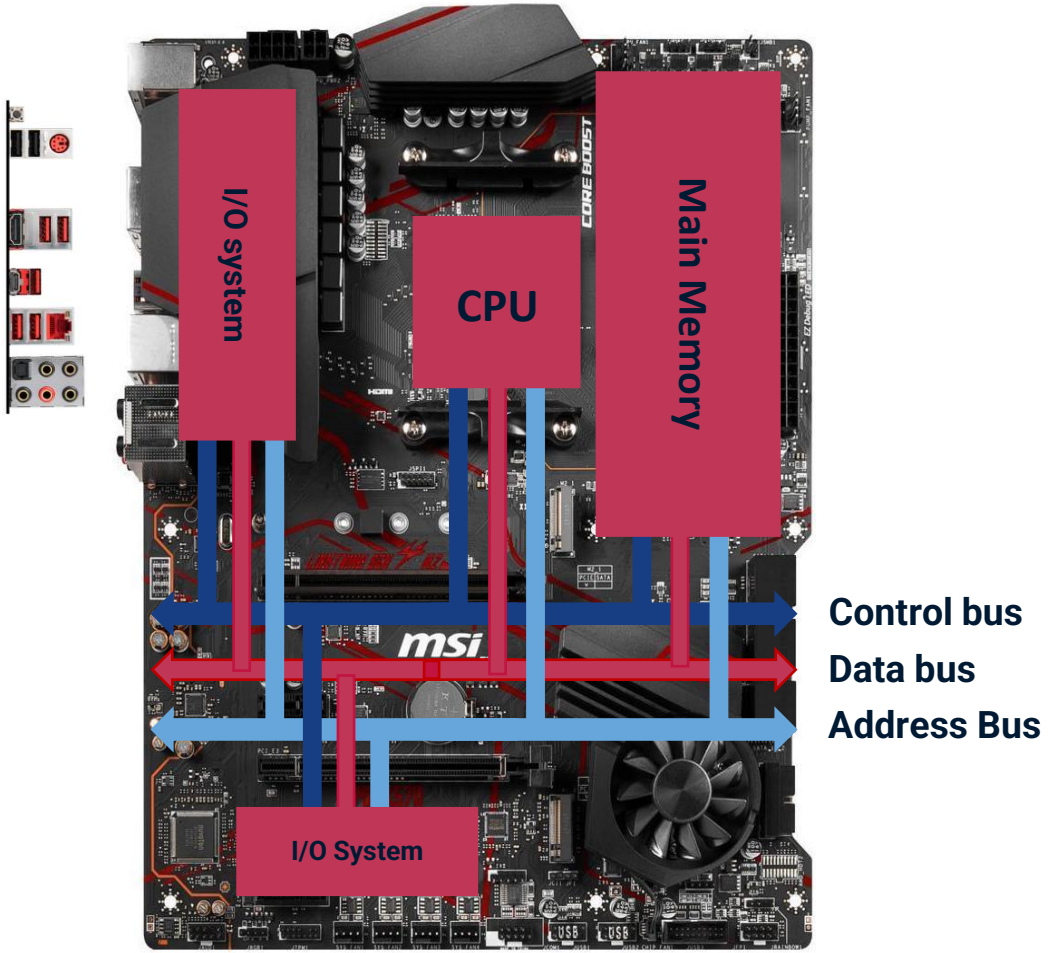


Internal bus

External bus

Standard cable

Computer architecture



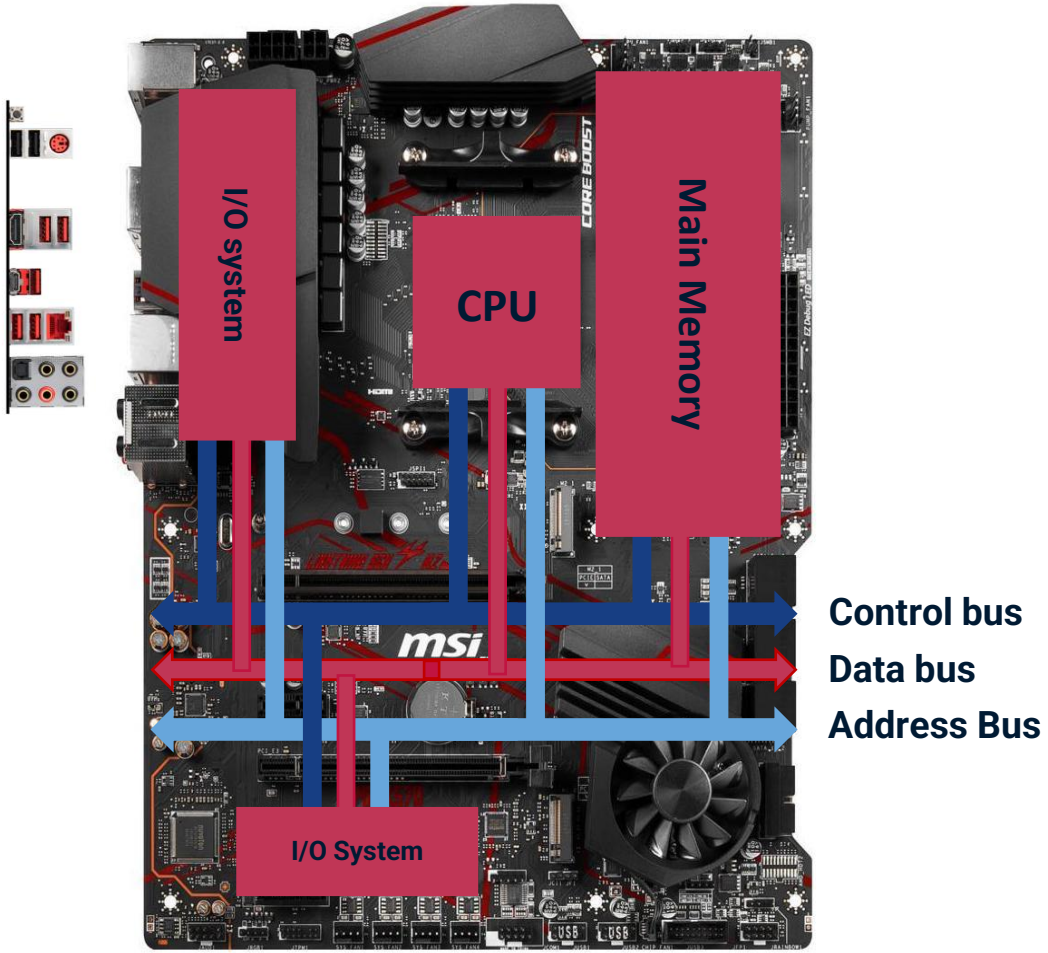
Control bus
Data bus
Address Bus

Buses represent the pathways or communication channels connecting various computer components to facilitate information transmission. Buses consist of multiple communication lines, with each line responsible for transmitting a single bit of data.

The **Data bus** is the communication conduit employed for the transfer of information among distinct computer components, serving the purpose of data exchange between memory, the CPU, and the I/O systems.

Typically, information is packaged in blocks of 32 or 64 bits, although other data bus widths, such as 1-bit, 4-bit, 8-bit, and 16-bit, are also utilized.

Computer architecture

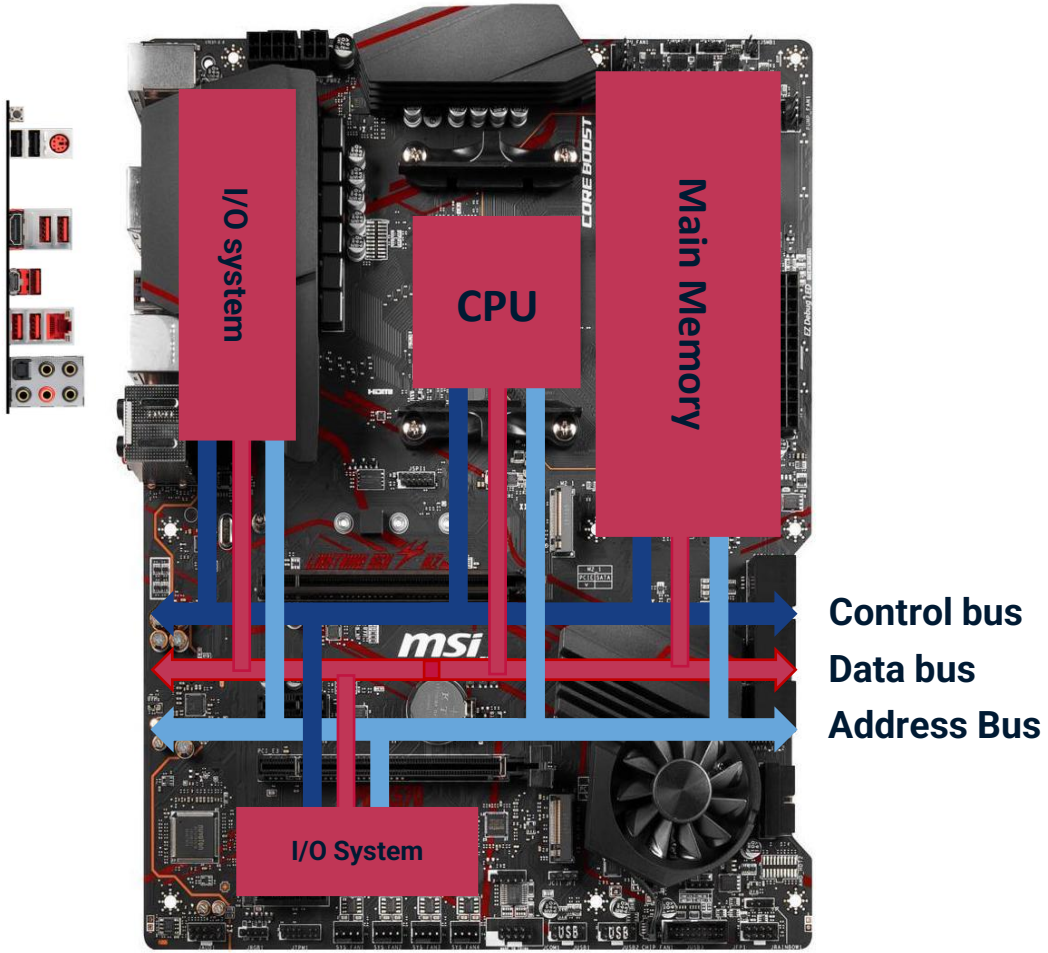


Buses represent the pathways or communication channels connecting various computer components to facilitate information transmission. Buses consist of multiple communication lines, with each line responsible for transmitting a single bit of data.

The **Address Bus** is the communication conduit is responsible for transmitting memory addresses, particularly those related to data in transit within the Main Memory.

Random Access Memory (RAM) functions as a temporal storage repository, enabling the computer to retain data required for fast retrieval and ensuring the accessibility of data, allowing the CPU to promptly access it without delving into long-term storage.

Computer architecture

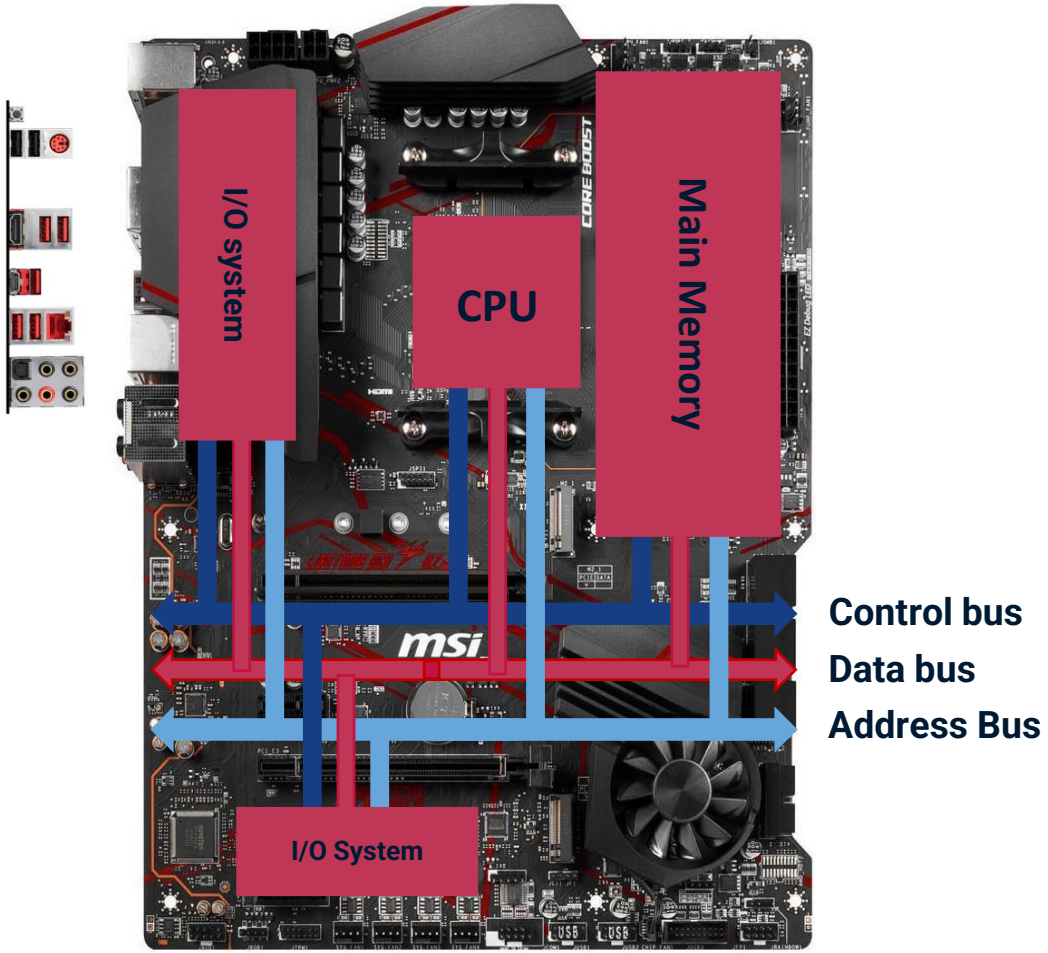


Buses represent the pathways or communication channels connecting various computer components to facilitate information transmission. Buses consist of multiple communication lines, with each line responsible for transmitting a single bit of data.

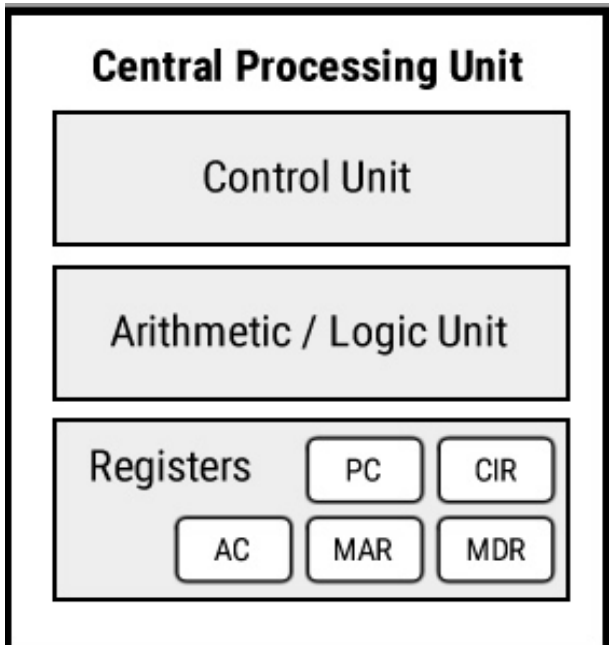
The **Control Bus** is the communication channel serves as a conduit for the transmission of control signals between the CPU and other computer components. Some of the most important signals are: (1) VDD power signals; (2) Read (RD) signal; (3) Write (RW) signal; (4) Clock signal; and (5) Reset signal (Reset).

The CPU utilizes these signals to regulate various processes within the computer and its constituent components.

Computer architecture

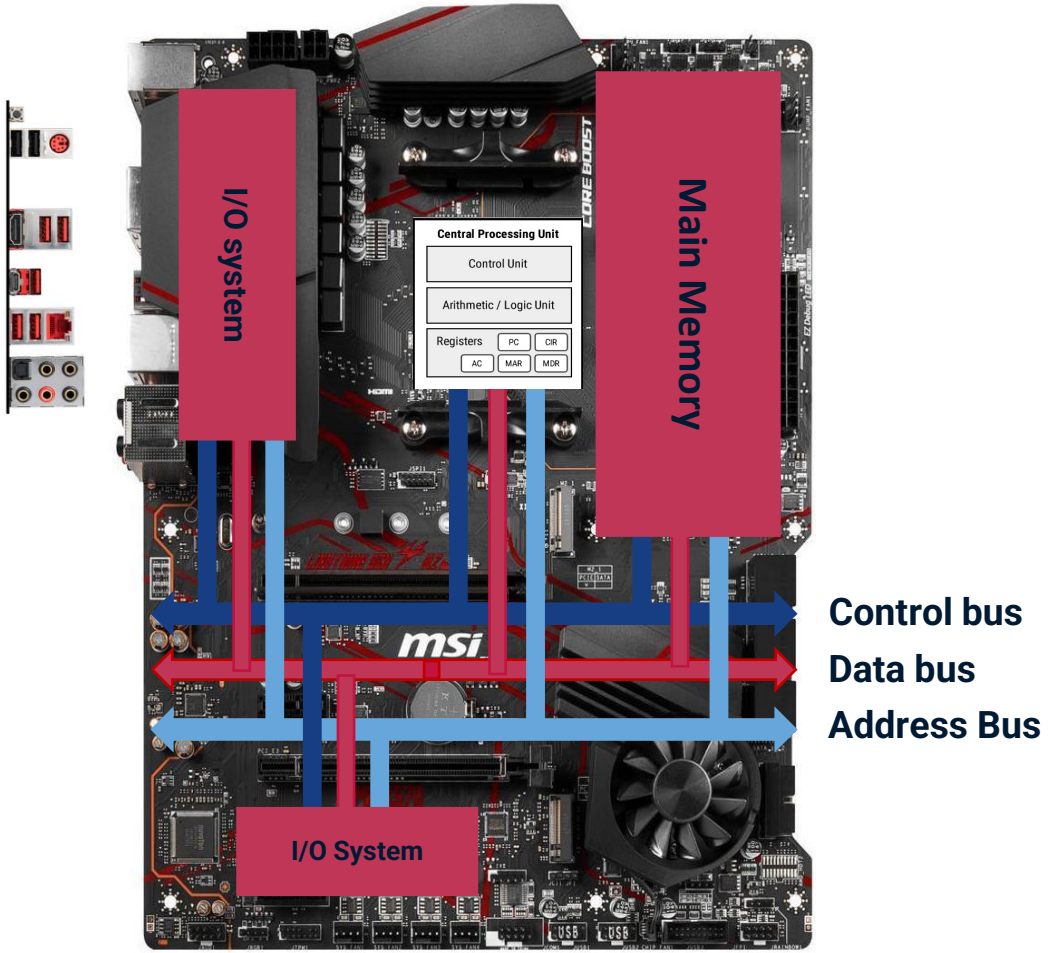


The Central Processing Unit (CPU) serves as the execution module for instructions stored in the main memory.



These instructions are executed sequentially, adhering to the order in which they are stored in memory.

Computer architecture

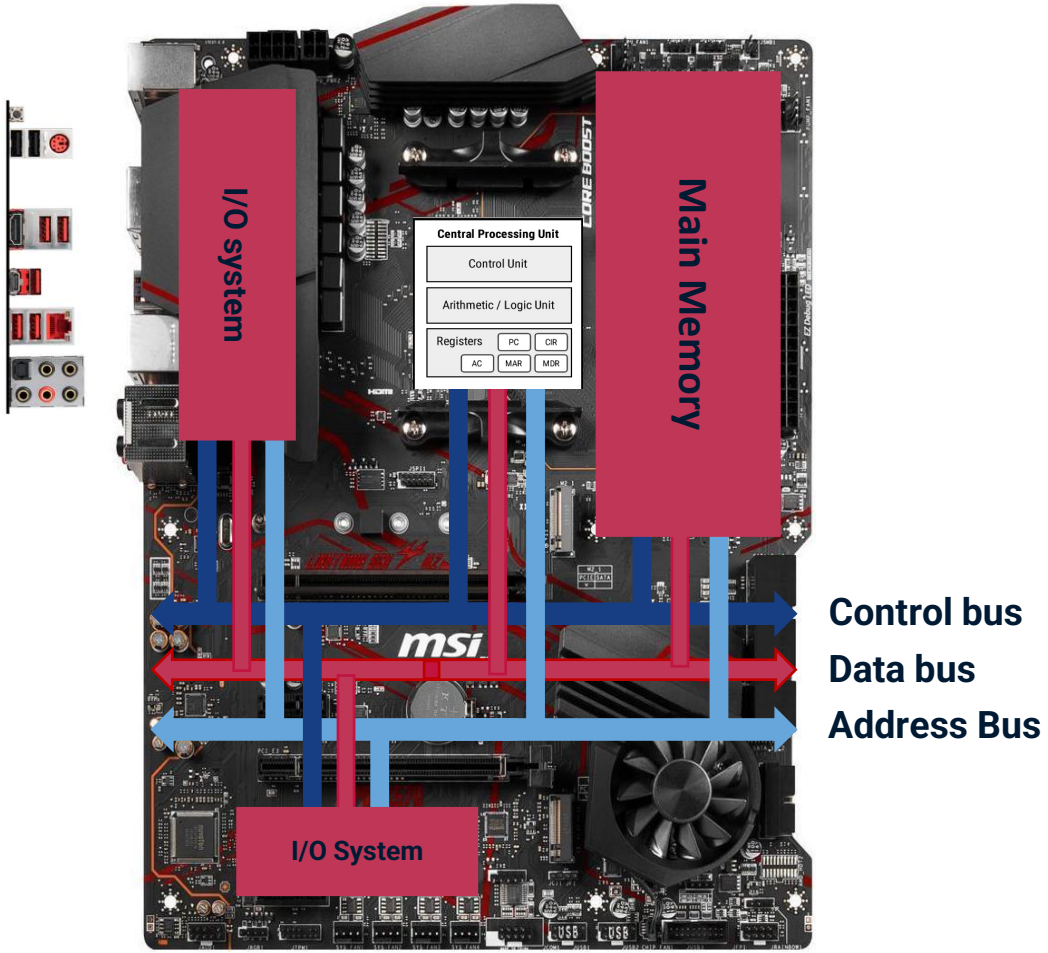


Control bus
Data bus
Address Bus

The Control Unit (CU) generates a variety of control signals necessary for the proper execution of instructions.

Additionally, it holds the responsibility of both reading and writing the diverse instructions that are slated for execution.

Computer architecture



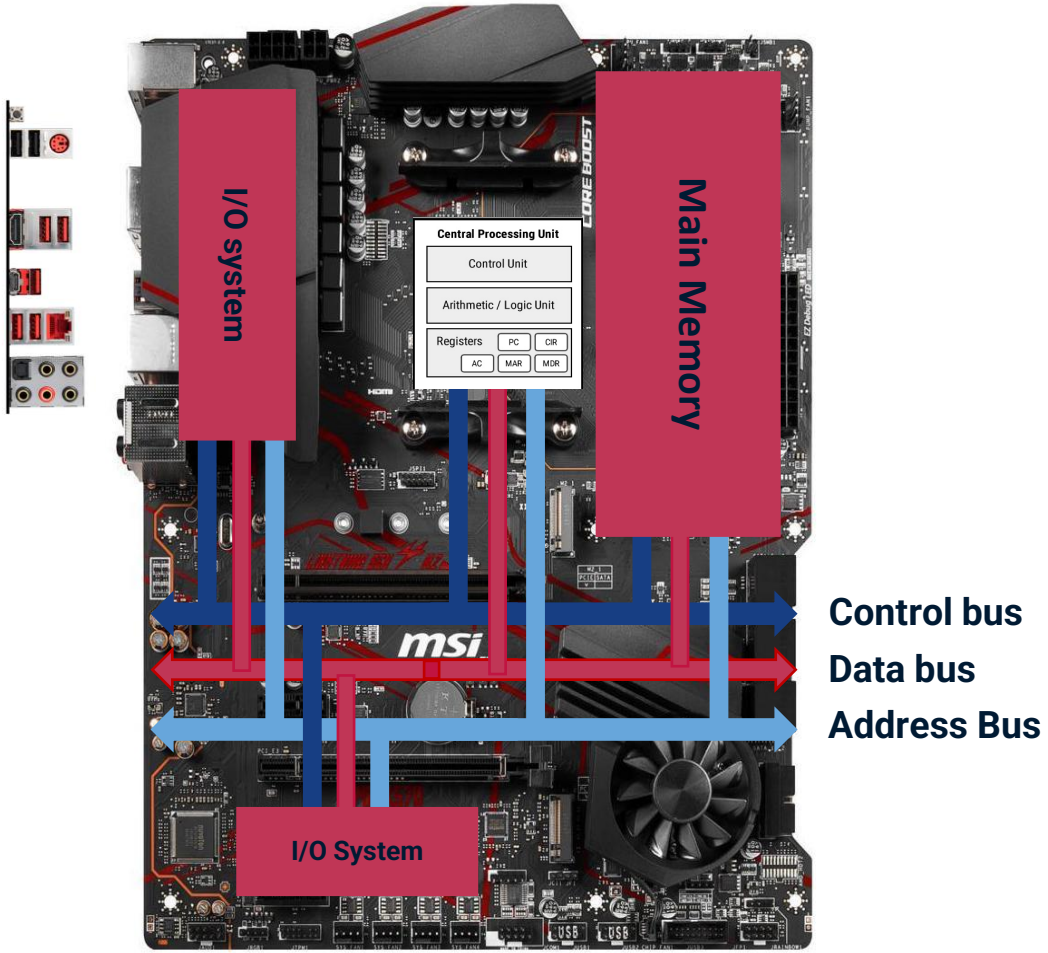
The Bank Registry (BR) consists of registers, which are sequences of bits. Among these registers, two hold particular significance:

- Program Counter (PC): This register preserves the address of the forthcoming instruction slated for execution.
- Instruction Register (IR): The IR retains the instruction currently in the process of execution.

Additionally, the BR encompasses the following registers:

- Memory Address Register (MAR): The MAR holds the address pertaining to the main memory, specifying the source or destination for data transfers.
- Memory Data Register (MDR): The MDR is responsible for storing the data destined to be written into or read from the designated word within the main memory.

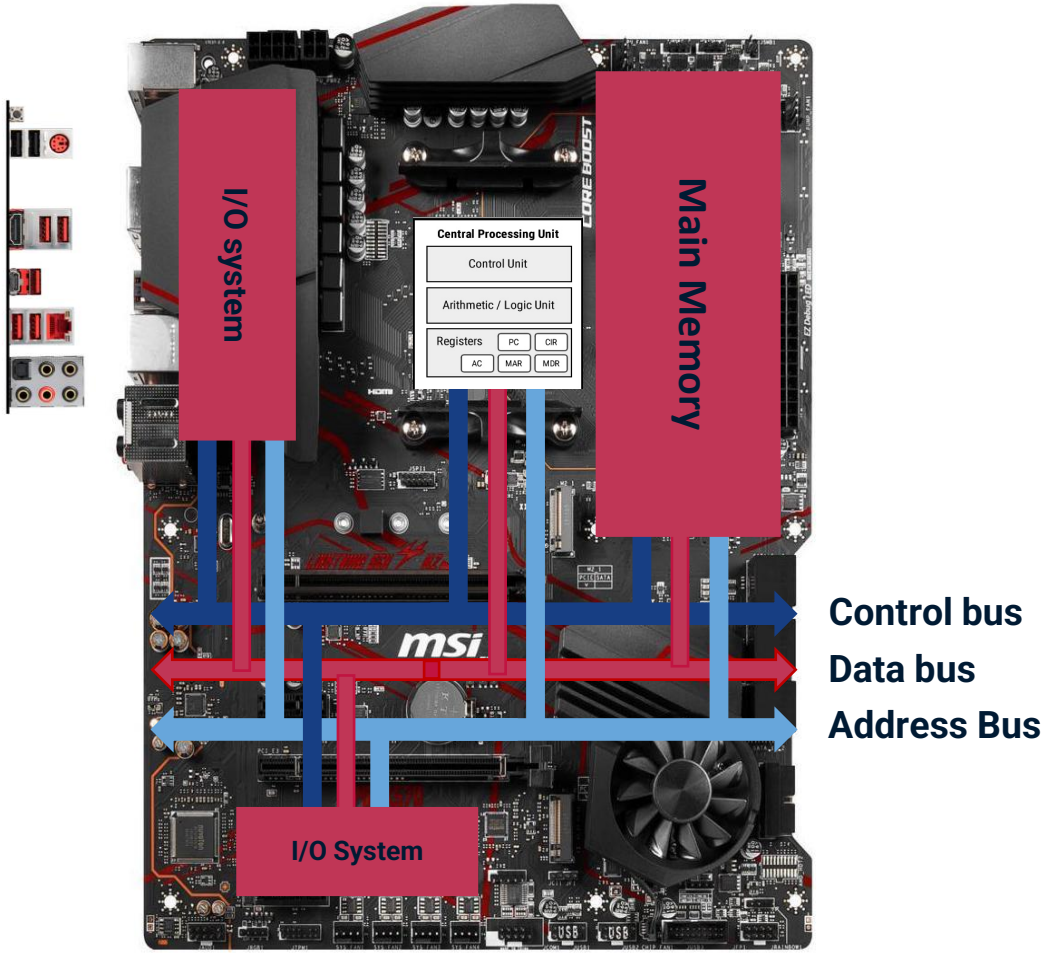
Computer architecture



The Arithmetic Logic Unit (ALU) represents a fundamental component within the central processing unit (CPU). It is responsible for executing arithmetic and logic operations and comprises two distinct units: the Arithmetic Unit (AU) and the Logic Unit (LU).

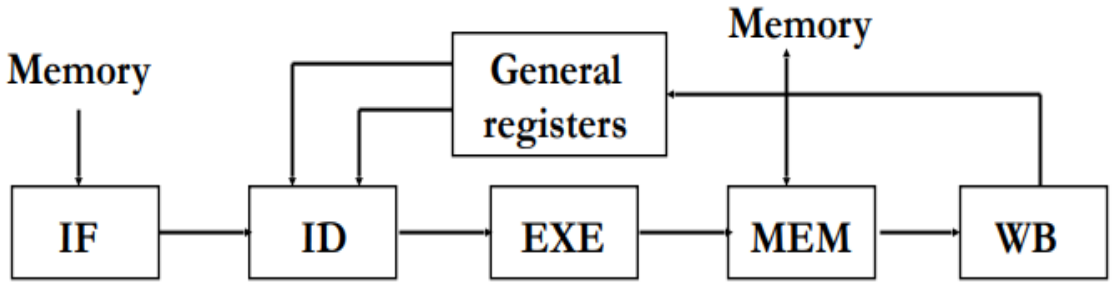
- Logical Operations: The logical operations encompass functions such as NOR, NOT, AND, NAND, OR, XOR, and others.
- Bit-Shifting Operations: This operation involves shifting the positions of bits either to the right or left by a specified number of places, often employed in multiplication operations.
- Arithmetic Operations: While the ALU is capable of handling multiplication and division, it primarily pertains to bit-level addition and subtraction.

Computer architecture

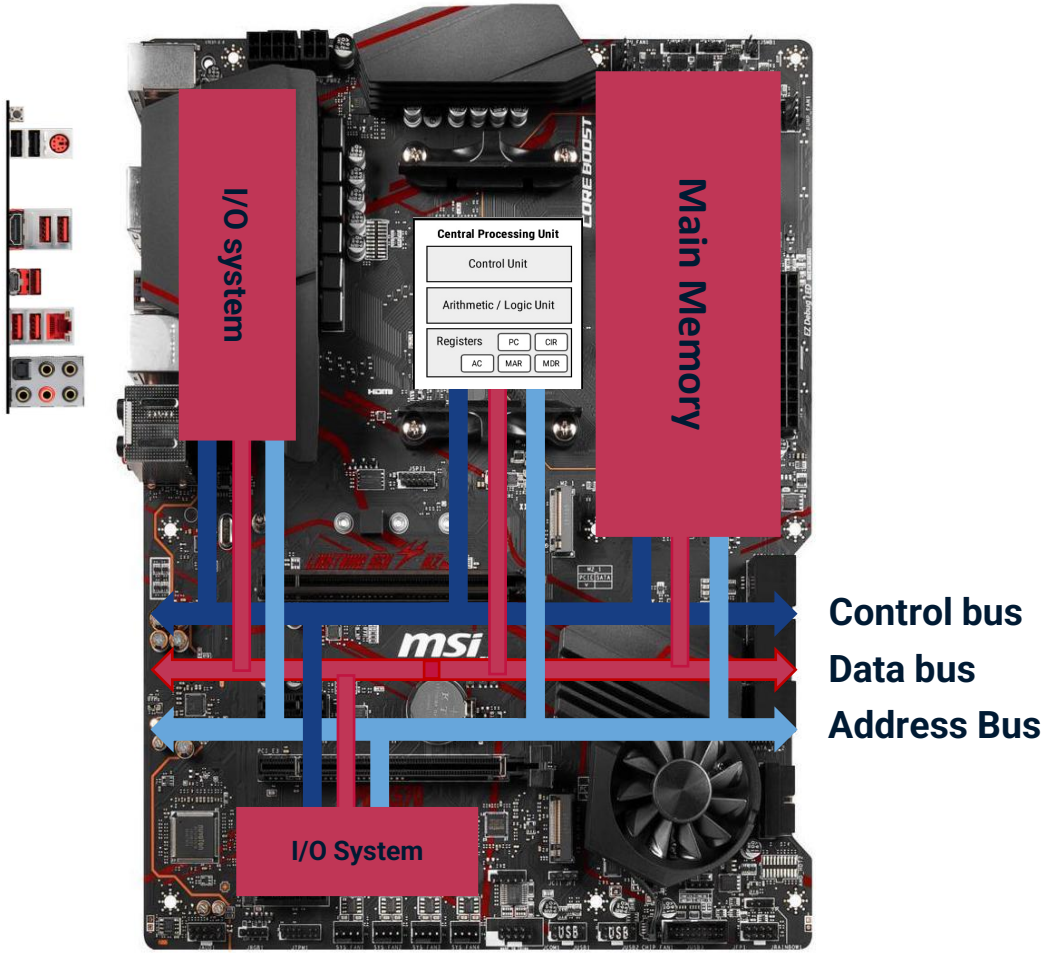


Stages of instruction pipeline:

1. To execute a program, the constituent instructions must be executed sequentially, one after the other.
2. Typically, these tasks are carried out in phases, with each phase being completed within a single clock cycle.
3. The specific number of stages and the duration of each clock cycle are contingent upon the characteristics of the processor.



Computer architecture

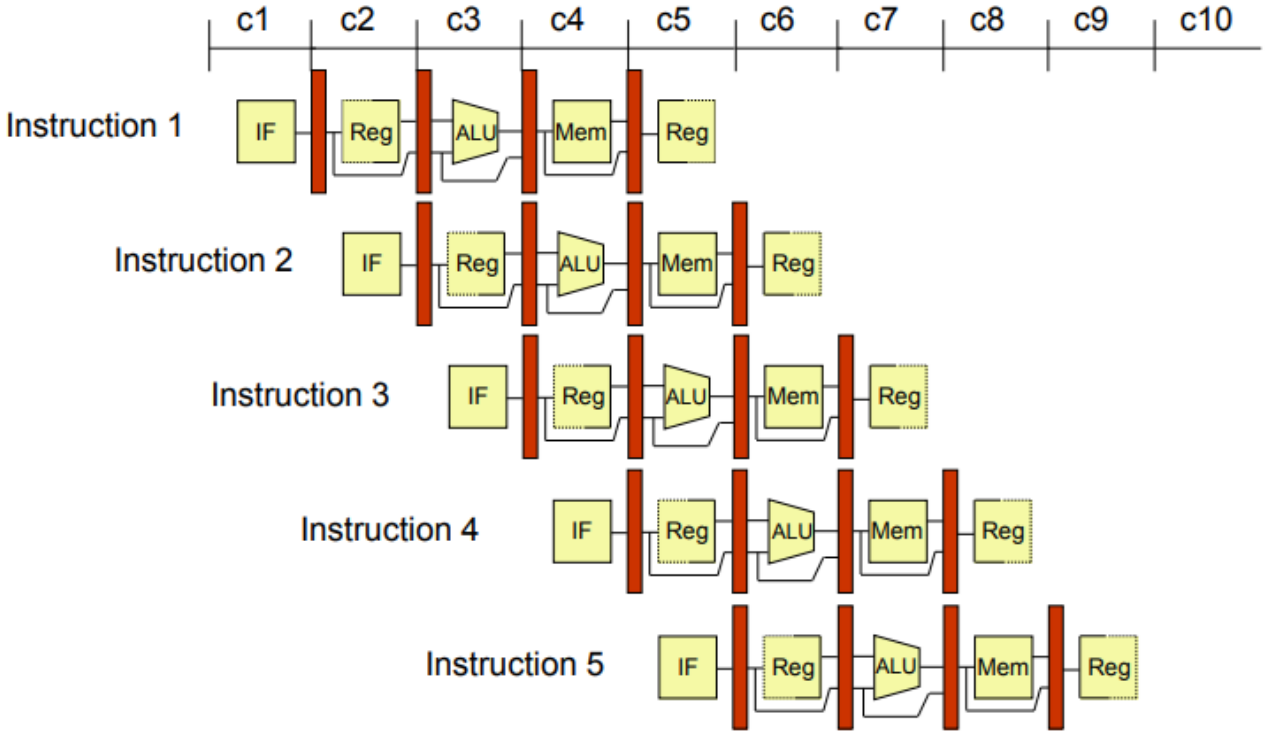


Control bus
Data bus
Address Bus

- Instruction Fetch (IF)
  - Instruction Register = MemRead (INST_MEM, PC)
- Instruction Decode (ID)
  - Control Signal Generation.
  - Determine register operands (Access and Fetching).
  - Instruction Preparation.
- Function Evaluation or Address Calculation (Execution)
  - Add, subtract, shift, logical, etc.
  - Address calculation (addition).
- Memory Access (if required)
  - Load operations: ReadData = MemRead(DATA_MEM, MemAddress, Size).
  - Store operations: MemWrite (DATA_MEM, MemAddress, WriteData, Size)
- Write Back (WB)
  - Update processor state modified by this instruction.
  - Interrupts or exceptions may prevent state update from taking place.

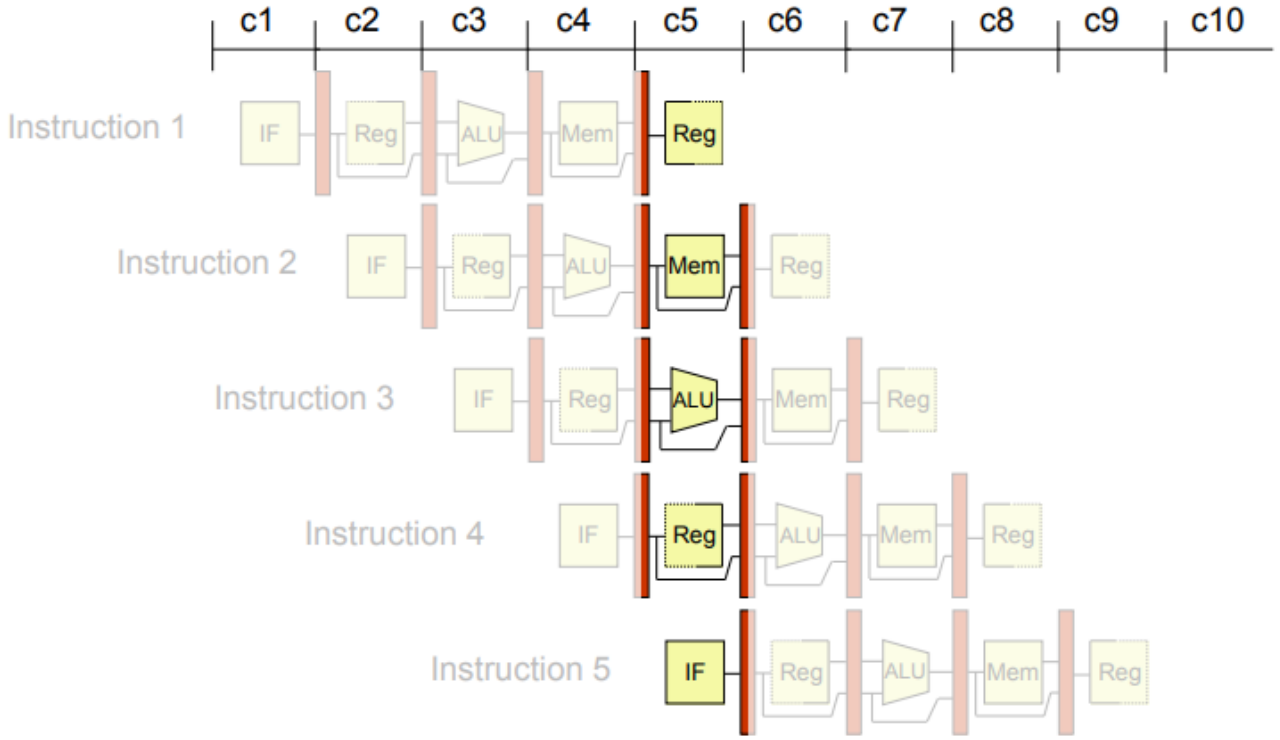
Programs are composed of a sequence of instructions

- A space-time diagram illustrating the pipeline.
- Conceptualize each instruction as a pipeline that is shifted in time.



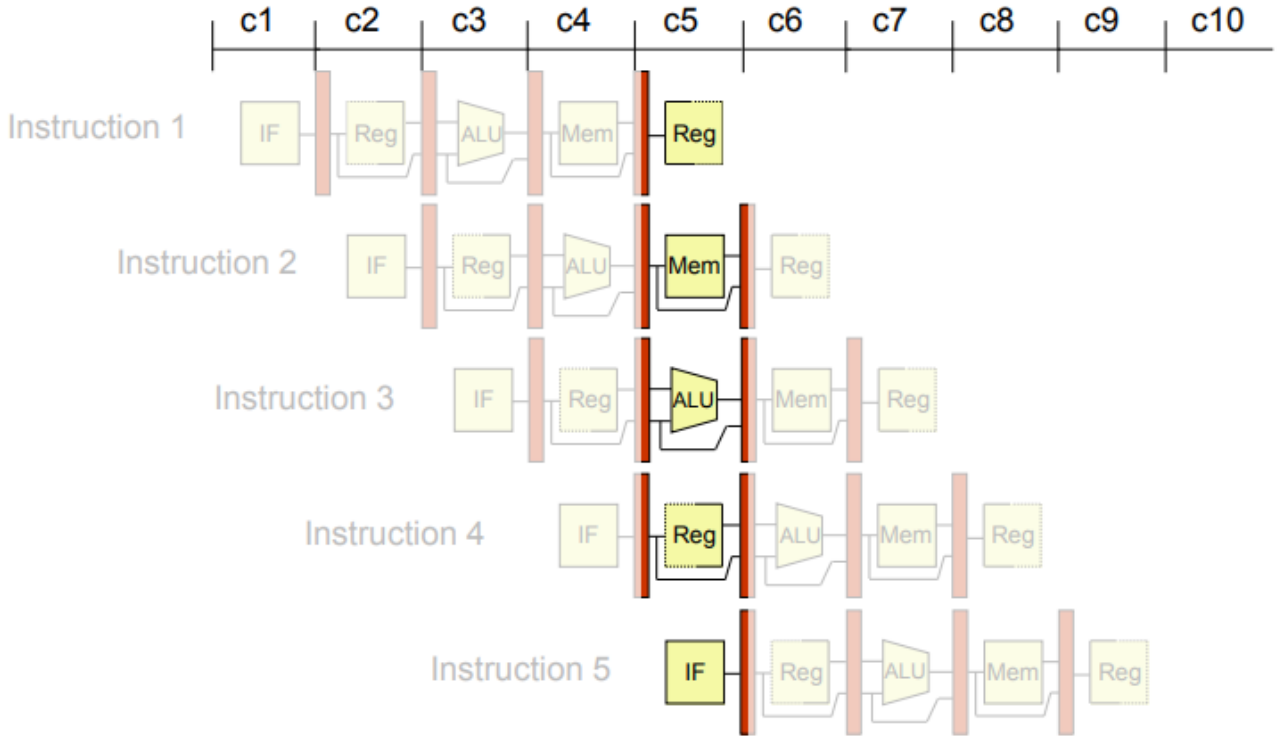
Programs are composed of a sequence of instructions

- A space-time diagram illustrating the pipeline.
- Conceptualize each instruction as a pipeline that is shifted in time.

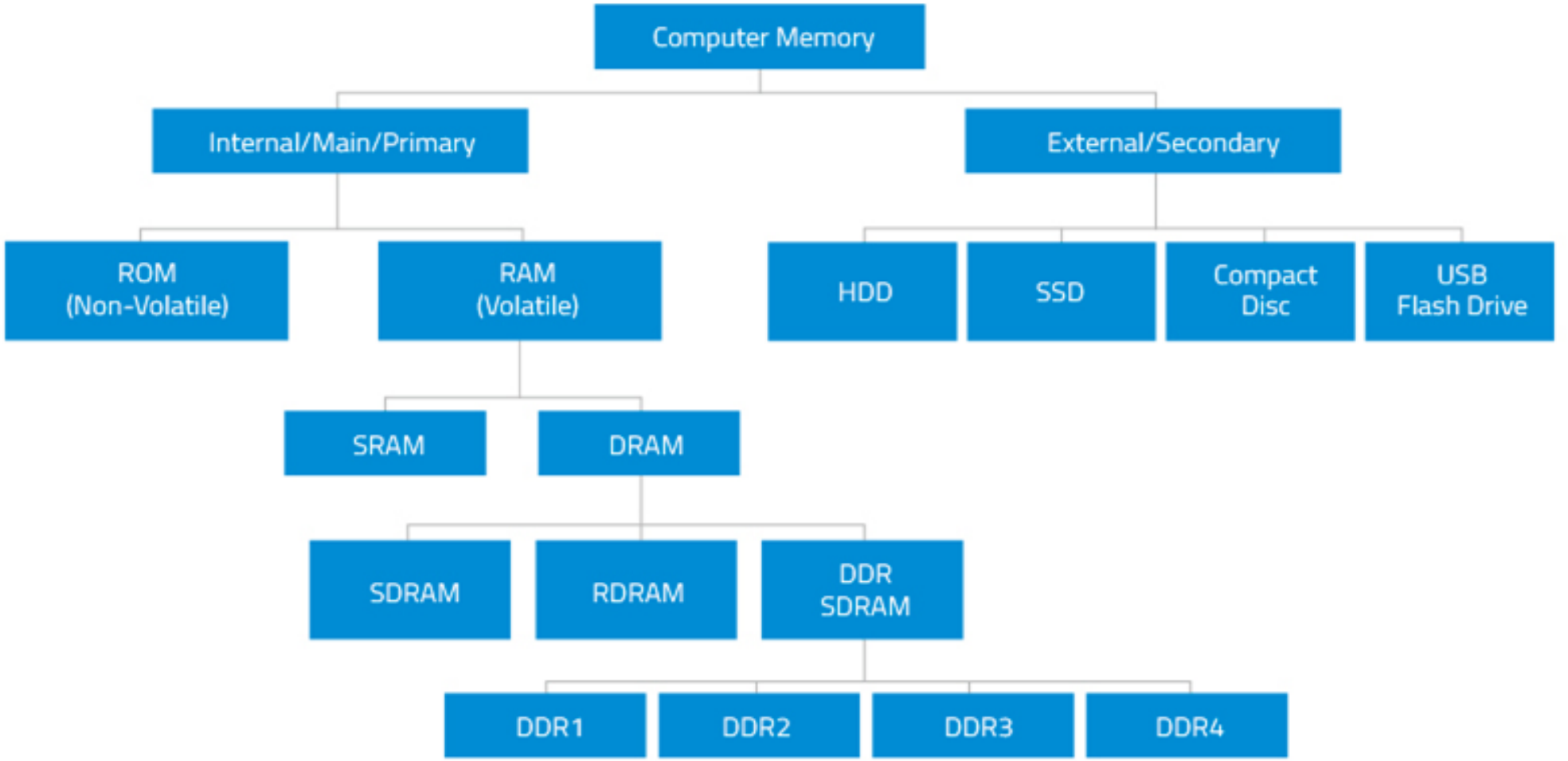


Programs are composed of a sequence of instructions

- A space-time diagram illustrating the pipeline.
- Conceptualize each instruction as a pipeline that is shifted in time.



Computer Memory



Computer Memory

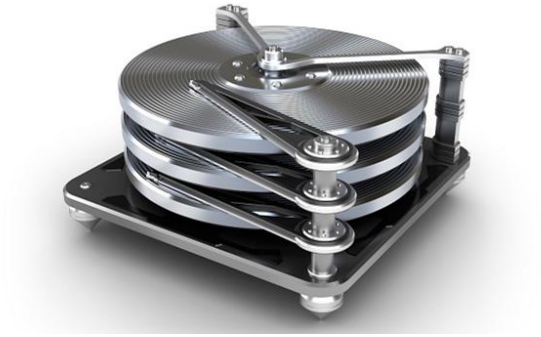
Primary Memory, also known as the main memory of the computer system, is accessed through the system bus. It is essential to note that this memory is volatile and lacks persistence.

- Cache Memory which is a temporary storage area, referred to as a cache, offers quicker accessibility to the processor compared to the primary memory of the computer. It is sometimes referred to as CPU memory because it is typically either integrated directly into the CPU chip or situated on a separate chip with a bus connection to the CPU.
- Random Access Memory (RAM) which constitutes the primary memory within a computer, serving as the repository for data, instructions, and information. Every storage location within this memory can be directly accessed by the Central Processing Unit (CPU).
- Read-Only Memory (ROM) which represents a form of primary memory, specifically designed for read-only access, earning it the name Read-Only Memory. Moreover, the data and instructions stored in ROM persist even when the computer is powered off, rendering it non-volatile memory.

Computer Memory

Secondary memory in computer systems serves as a storage medium for data and information, facilitating efficient retrieval, transmission, and utilization by applications and services. Key attributes of secondary memory include:

- Non-Volatility: It retains data even when the power supply is disconnected, thus exhibiting persistence.
- Access via I/O Channels: Secondary memory is accessed through input/output (I/O) channels.
- Large Capacities: It offers substantial storage capacities, often reaching the scale of terabytes.
- Cost Efficiency: Secondary memory is typically more cost-effective when compared to primary memory.



Input/Output systems

These are devices that establish connectivity between the computer and external entities. They interface with the Central Processing Unit via Input/Output units.