Digital microsystems design

Marius Marcu

2020

Outline

- Definitions
- Parallelism
- Classification
  - Memory organization
  - Processor organization
- Summary

Definitions

- Digital microsystem
- Microprocessor
- External connectivity of a microprocessor
- Internal architecture of a microprocessor
- Bus

Definitions

- A digital microsystem is a microprocessor or microcontroller based computing system.
  - Computer systems
  - Embedded systems
  - Mobile systems

Definitions

- A microprocessor is a general purpose programmable logical circuit
  - Can be used in general purpose computing devices
  - Can be programmed to run any type of application
- A microcontroller is a application specific programmable logical circuit dedicated to real-time, embedded applications
  - Application oriented processor
  - Domain specific processor
- Mobile processors are low power processors used in mobile systems
  - General purpose, simple processors
  - Low power, energy efficient

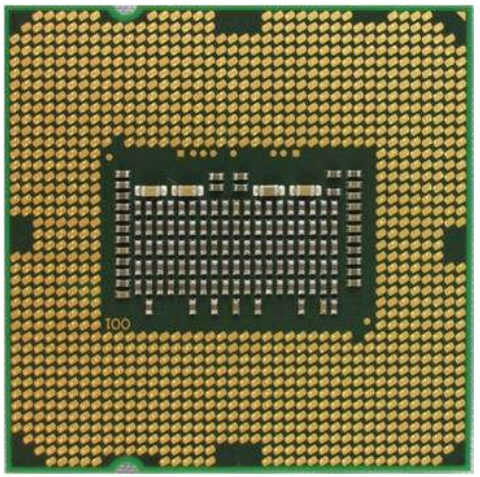
Definitions

- Other types of processors
  - DSP – Digital Signal Processor
  - GPU – Graphics Processing Unit
  - FPU – Floating Point Unit
  - TPU – Tensor Processing Unit

  - ASIC – Application Specific Integrated Circuit
- Soft-processors (soft cores)
  - FPGA synthetized

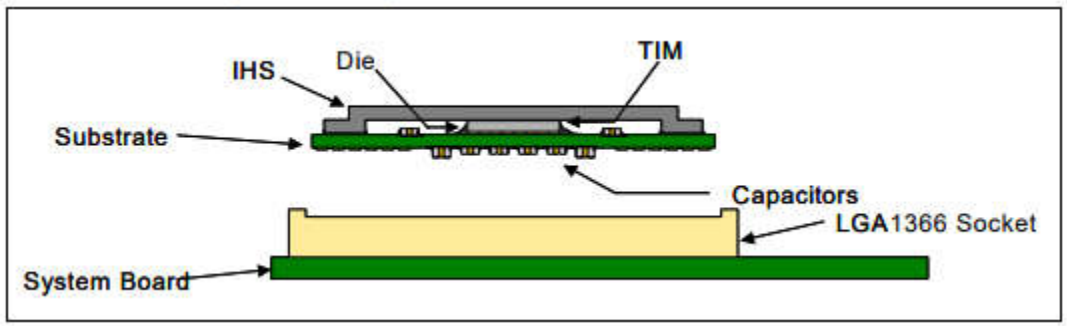
Definitions

- External connectivity of a microprocessor
  - Integrated circuit
  - Large number of pins/lines
    - Processor bus



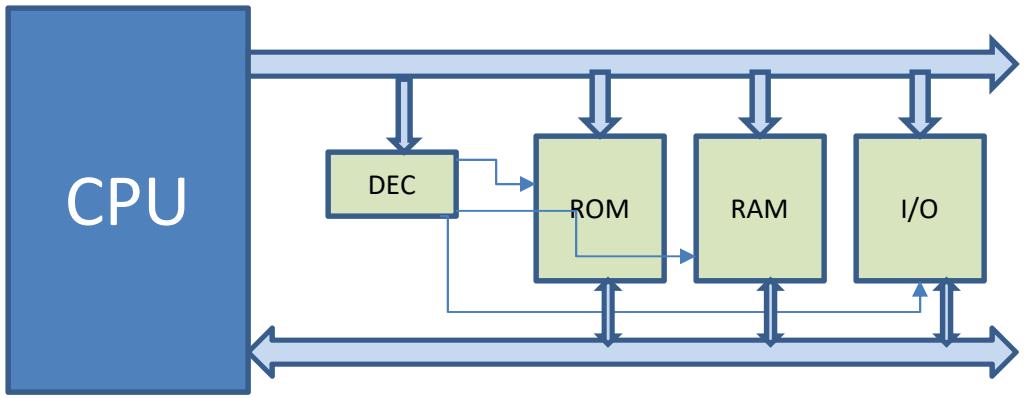
Definitions

- External connectivity of a microprocessor
  - Packaging
    - IHS – integrated heat spreader
    - TIM – thermal interface material
    - Socket – connector for placing the processor on the main system board



Definitions

- External connectivity of a microprocessor
  - CPU cannot execute applications without a memory
  - Allows the CPU to transfer information to/from other components
  - CPU is the bus master, initiating the transfer
  - Each component the CPU can access takes a specific address in the overall CPU address space



Definitions

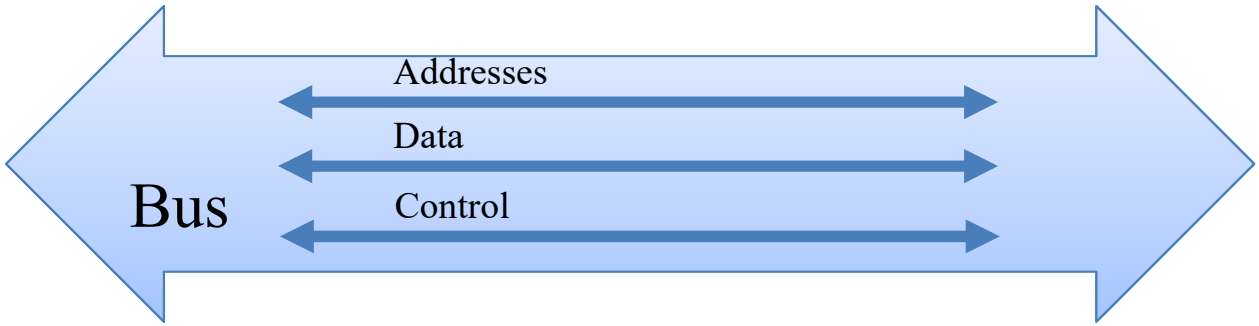
- External connectivity of a microprocessor
  - Access and transfer data to external (mainboard) resources
    - Memory
    - I/O
  - Both instructions and data are located during execution in memory circuits
    - ROM – system boot loader
    - RAM – code and data

Definitions

- External connectivity of a microprocessor
  - Bus
    - A communication system that transfers data between two or more components inside a digital system
    - A group of digital lines having common functional, electrical and logical characteristics used to interconnect two or more units
      - Functional – implement a communication protocol
      - Electrical – the same electrical parameters
      - Logical – the same logical levels
    - Implemented using tri-states buffers

Definitions

- External connectivity of a microprocessor
  - Bus
    - Address channel
    - Data channels
    - Command and control lines
  - Multiplexing
    - Data and address channels use the same physical lines

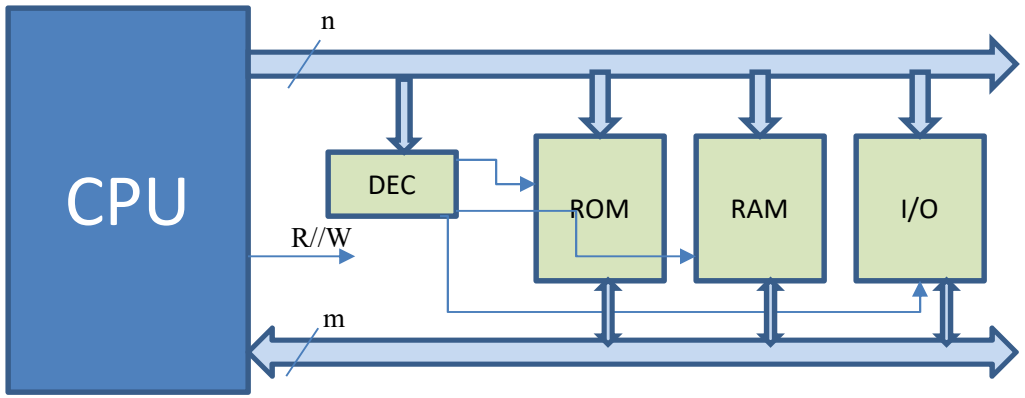


Definitions

- External connectivity of a microprocessor
  - Bus
    - Slave units
      - Can respond to a master unit
    - Master units
      - Can initiate transfers with a slave unit
    - Arbiter
      - Decision when two or more masters initiate bus transfers at the same moment in time
    - Decoder
      - Selection of the slave the master will transfer data

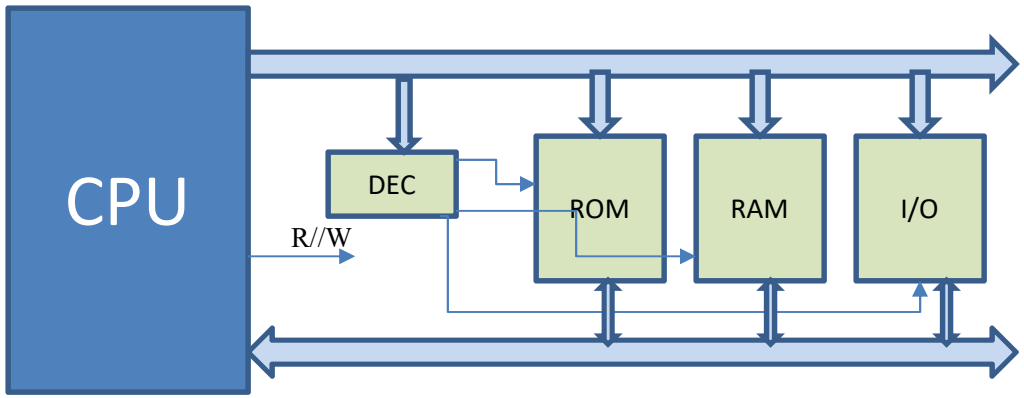
Definitions

- External connectivity of a microprocessor
  - Address space of the processor
    - Address bus width (n)
    - Max addressable size (2^n)
    - CPU address space 0 – 2^n-1 (in hexa)



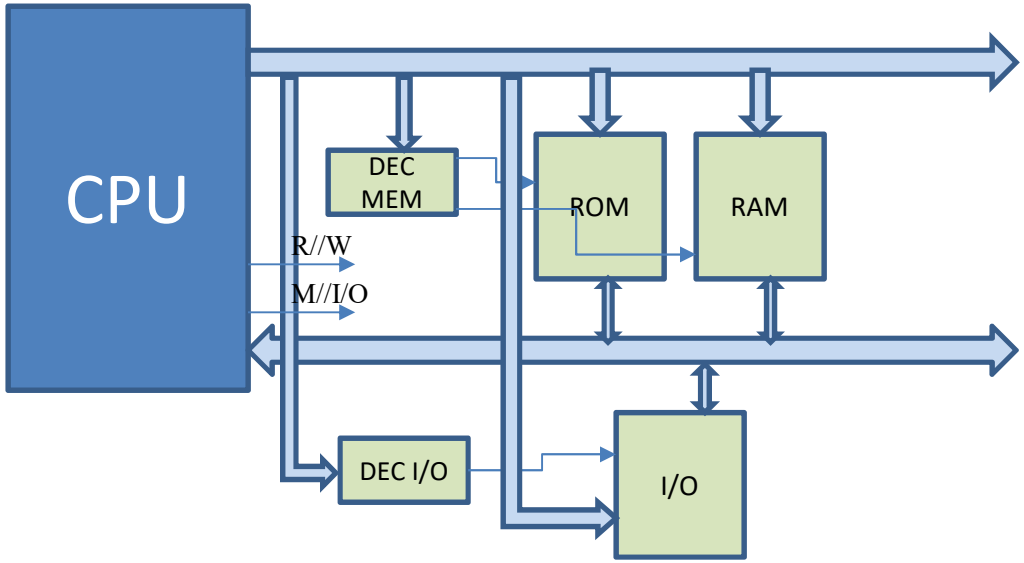
Definitions

- External connectivity of a microprocessor
  - Common memory and I/O space



Definitions

- External connectivity of a microprocessor
  - Distinct memory and I/O space
    - Memory address space
    - I/O address space



Definitions

- Address channel
  - Address space – two types of microprocessors
    - Common address space for memory and I/O
      - LD/Store operations for both I/O and memory
    - Distinct address spaces for memory and I/O
      - LD/ST memory
      - IN/OUT I/O space
  - Address space size (n – size of address bus)
    - Memory size addressable by the microprocessor: 2^n
  - Address – the index of the memory location of the transfer

Definitions

- Examples
  - 16 bits microprocessor
    - 16 bits address width – 64KB address space
    - 20 bits address width – 1 MB address space
  - 32 bits microprocessor
    - 32 bits address width – 4 GB address space
    - 40 bits address width – 1 TB address space
  - 64 bits microprocessor
    - 48 bits address width – 256 TB address space
    - 56 bits address width – 64 PB address space
    - 64 bits address width – 16 EB address space

Definitions

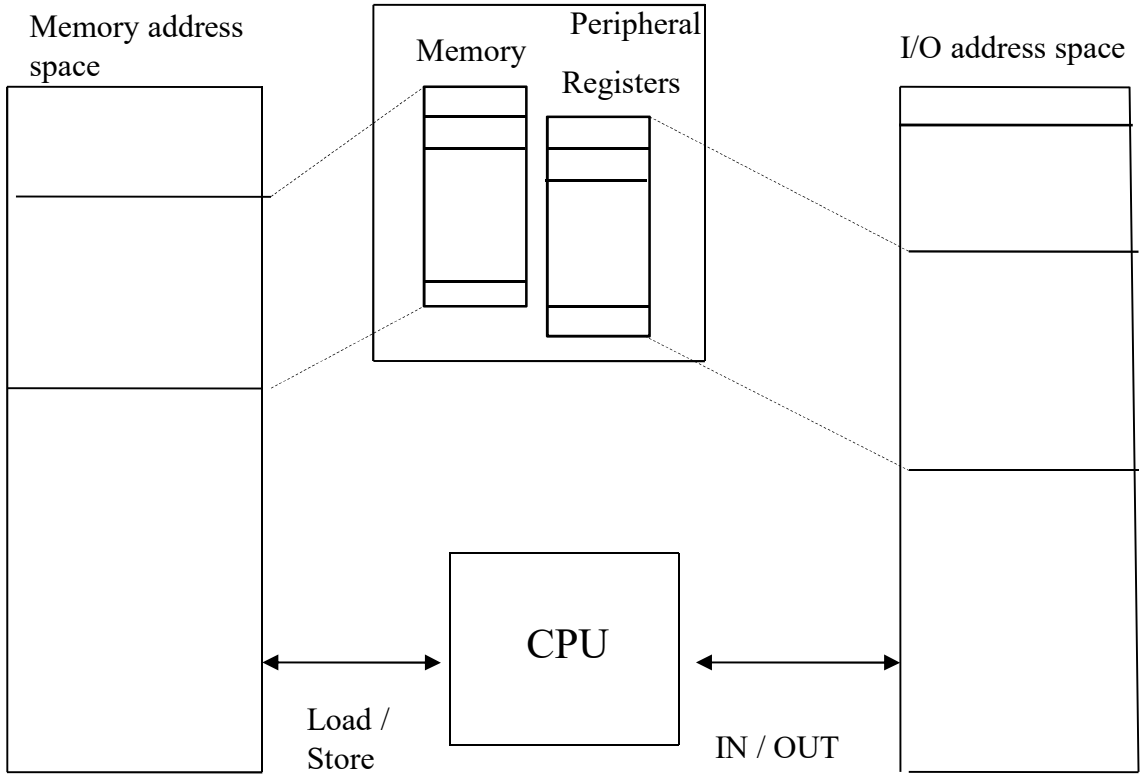
- Distinct address spaces for memory and I/O:
  - I/O address space
    - The microprocessor transfers information from/to I/O devices using dedicated instructions called input/output instructions (IN/OUT)
  - Memory address space
    - The microprocessor transfers information from/to memory locations using dedicated load store instructions (LOAD/STORE)

Definitions

- Common address space for memory and I/O
  - The microprocessor transfers information from/to both I/O and memory using the same instructions (LOAD/STORE)

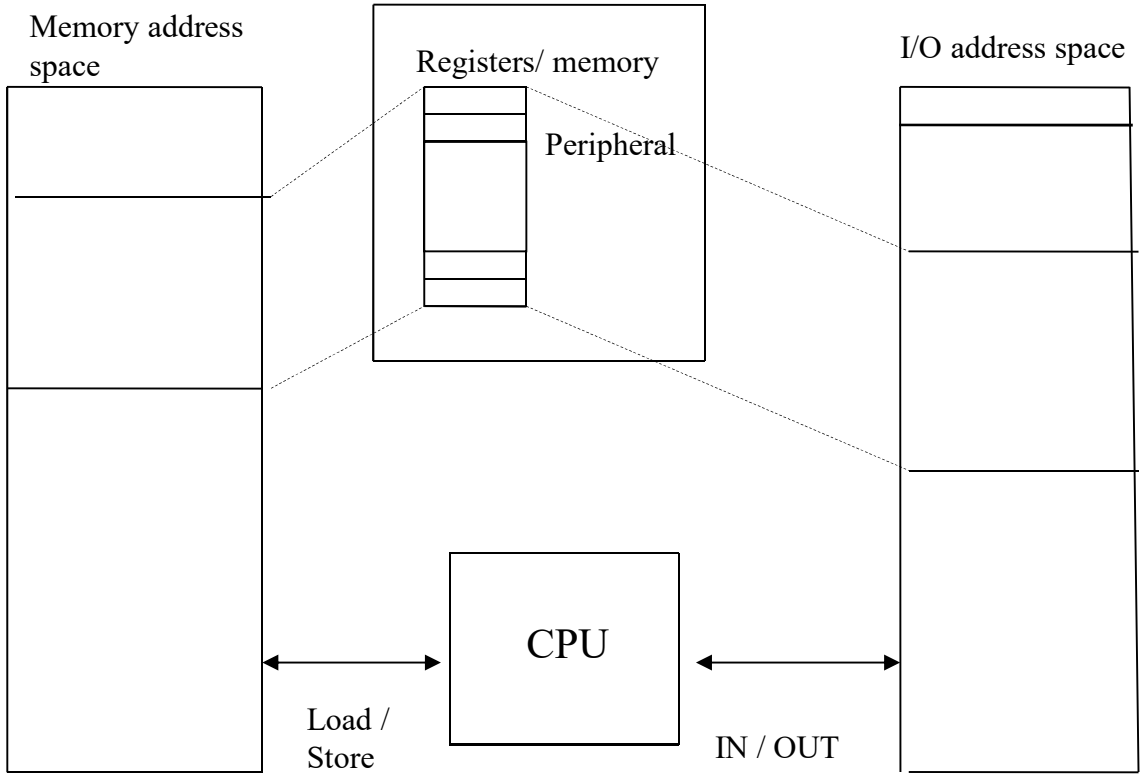
Definitions

- Address ranges allocated to each component, register, memory location
  - Distinct memory and I/O space



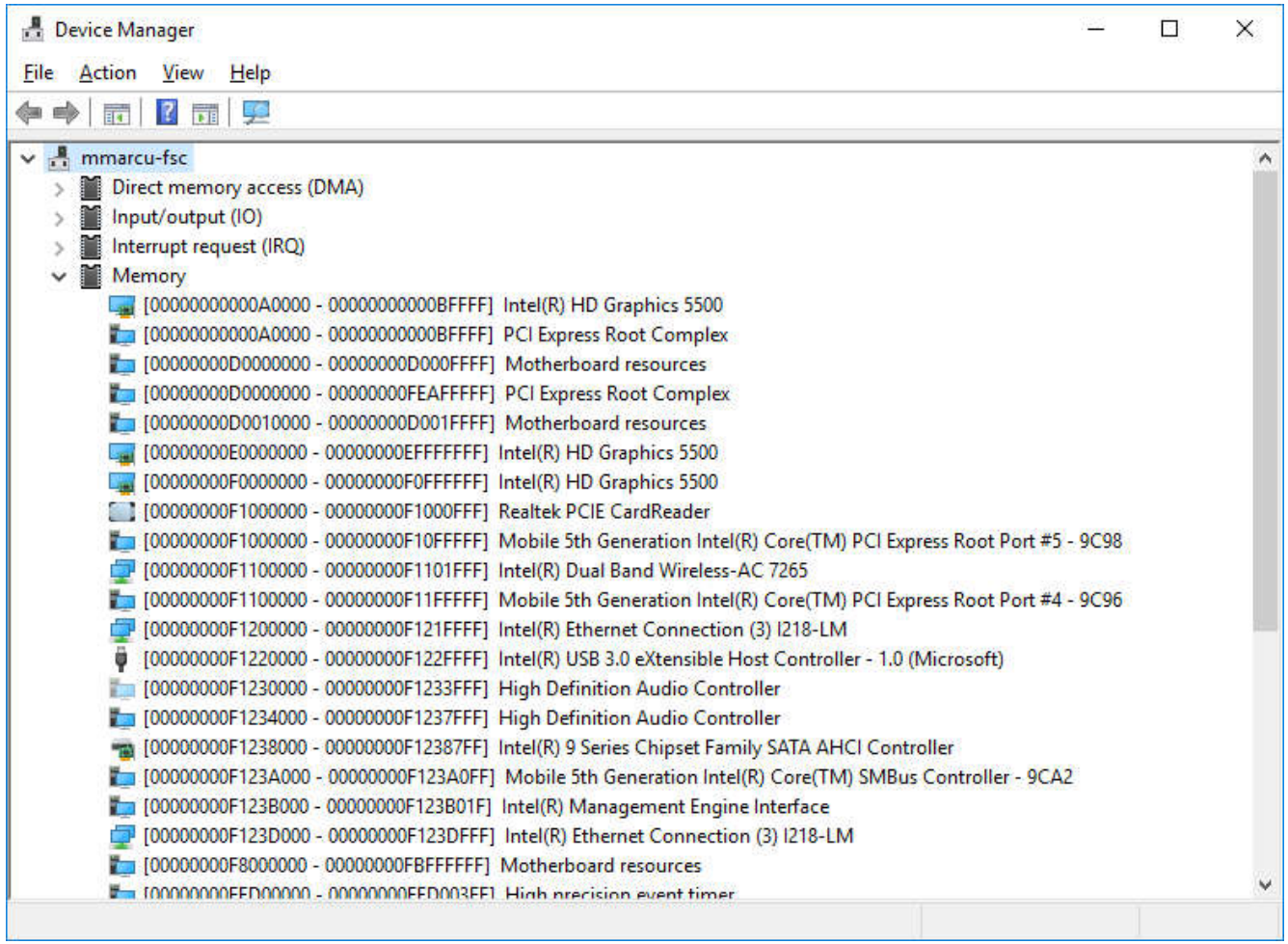
Definitions

- Address ranges allocated to each component, register, memory location
  - Common address space



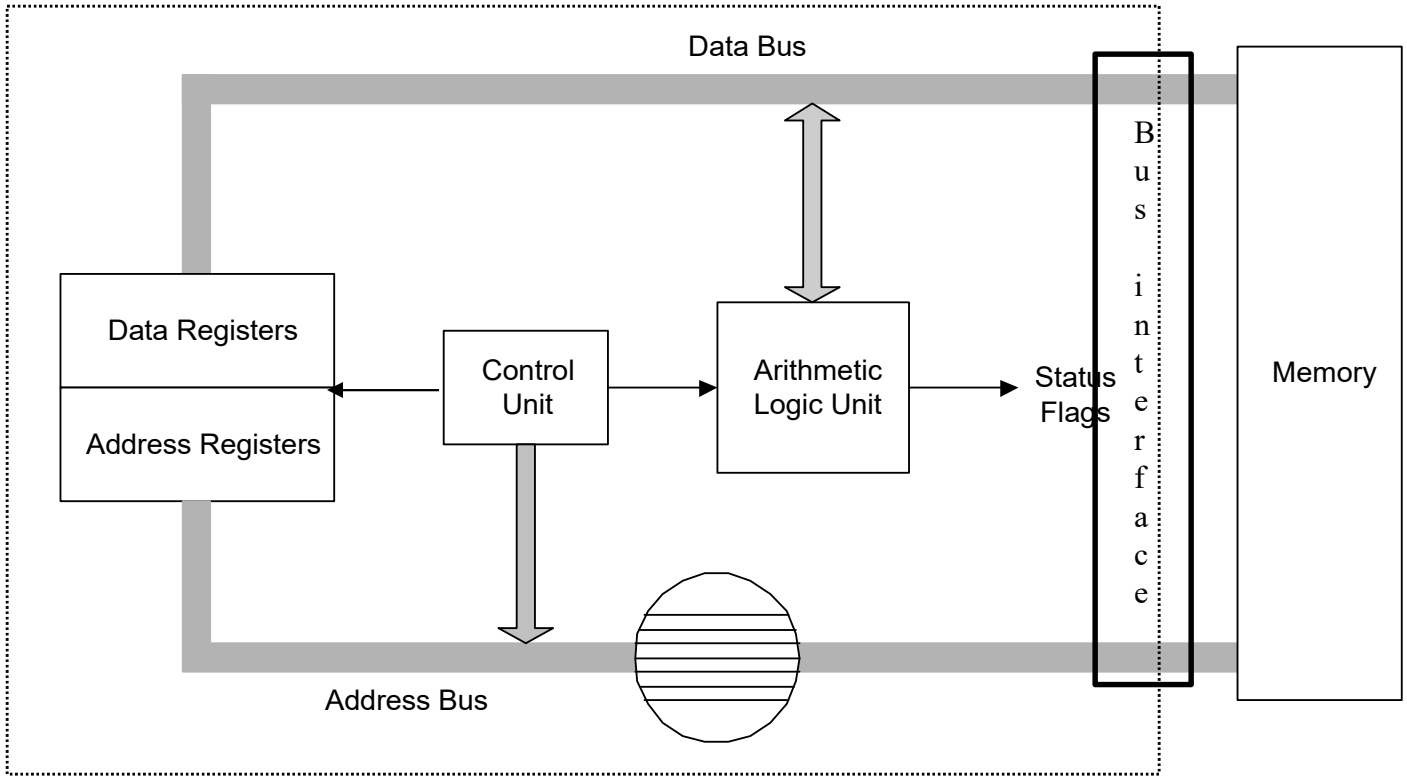
Definitions

- Device manager



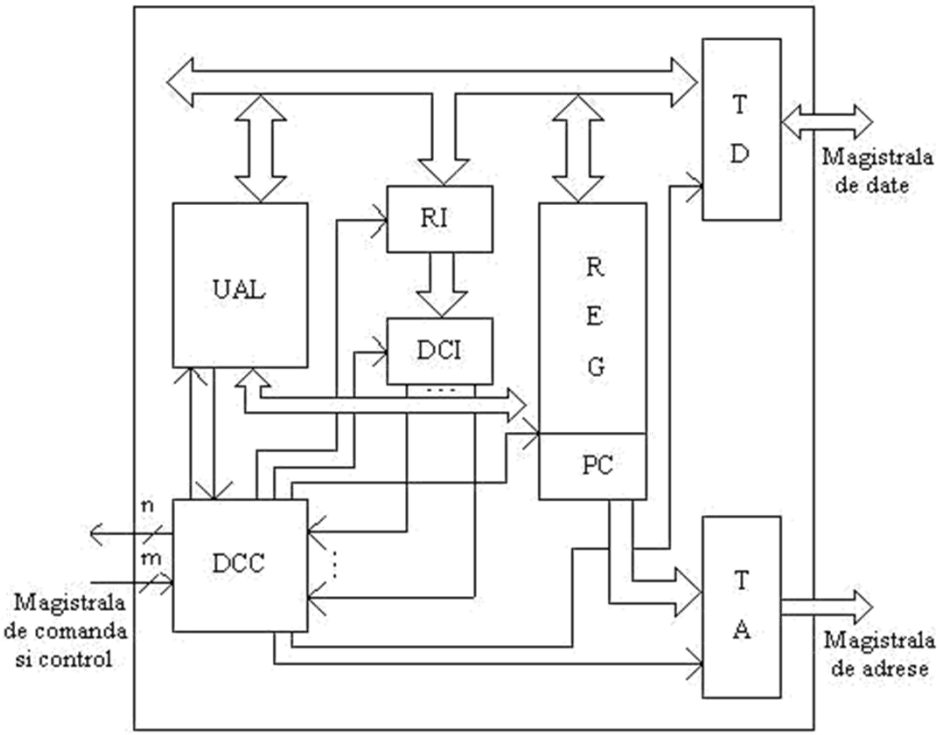
Definitions

- Internal architecture of a microprocessor
  - microarchitecture



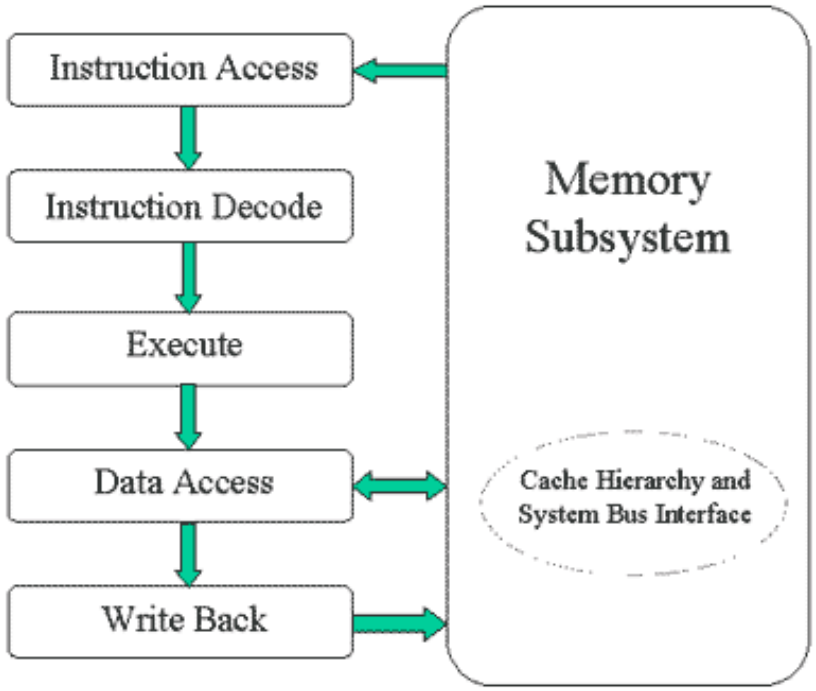
Definitions

- Microarchitecture
  - Registers set
  - Control unit
  - Arithmetic-logic unit
  - Buses
  - Interface buffers
  - Instruction decoder
  - Address registers



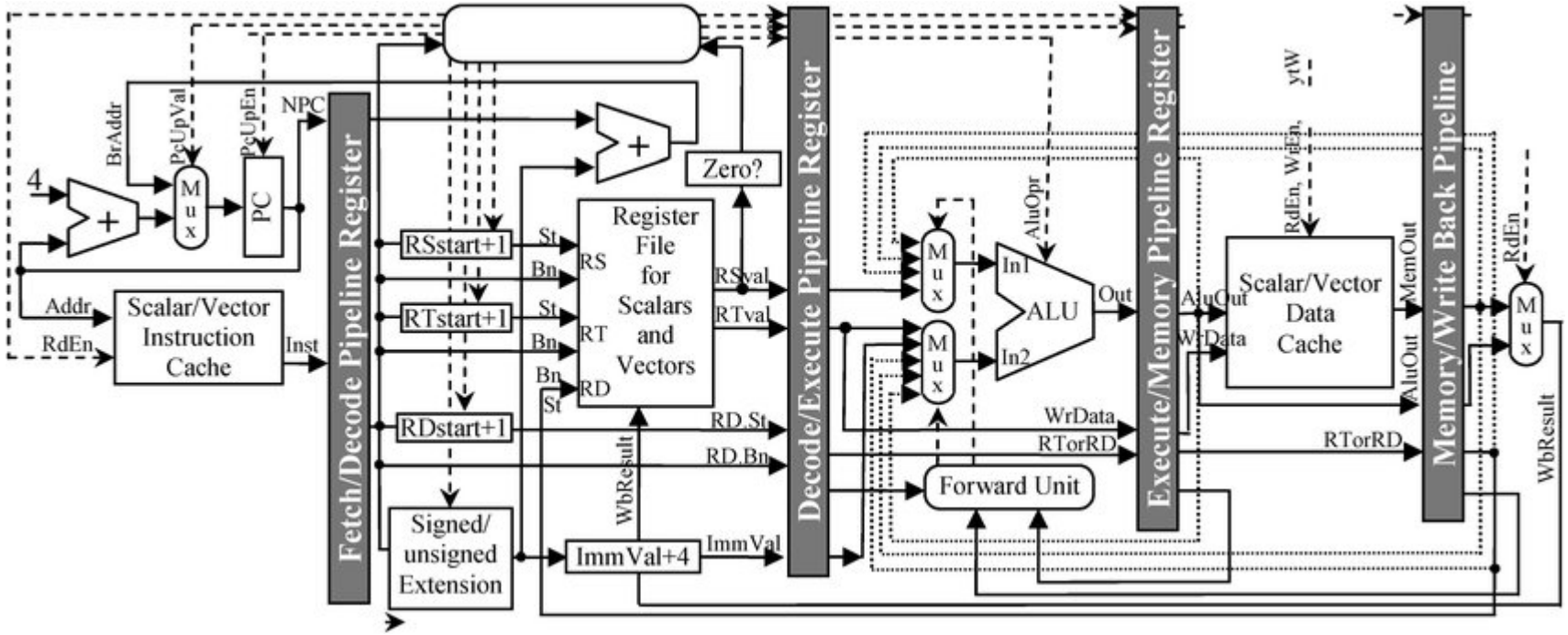
Definitions

- Five stage pipeline processor microarchitectures



Definitions

- Pipelined microarchitecture
  - Internal components
  - Organized as a pipeline



Definitions

- Processor microarchitectures
  - Specifies internal microprocessor architecture and the way instructions are implemented and executed
  - Includes:
    - Register file
    - Execution units
    - ILP – Instruction Level Parallelism
    - Machine cycles

Solutions

- What solutions exists for future performance advances?

Solutions

- Parallelism levels:
  - Instruction level parallelism (ILP)
    - Instructions execution overlapping
    - Micro-operations execution overlapping
  - Operating system level parallelism
    - Multitasking
    - Multiprocessing
  - User applications parallelism
    - Multithreading
  - System level parallelism
    - Virtualization

Solutions

- Parallelism levels:
  - Processor
    - pipeline
    - superscalar
    - hyper-threading
    - multi-core
    - many-core
    - heterogeneous cores – big-little architectures
  - System
    - multiprocessor
  - Network
    - Distributed
      - cluster
      - grid
    - Cloud

Solutions

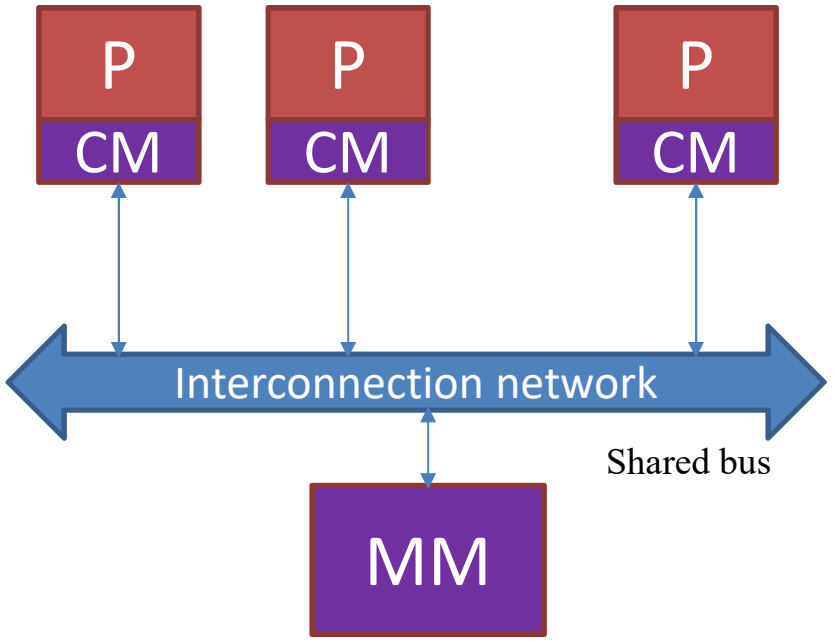
- Architectural advances:
  - Scalar architectures
  - Superscalar architectures
    - Overlapping instruction fetch with previous instruction execution
    - Multiple execution units (e.g. multiple ALUs)
    - Pipeline
  - Vector processors
  - Processors arrays
  - Multiprocessor systems
  - Distributed systems

Classification

- Memory organization
  - UMA (Uniform Memory Access Model)
  - NORMA (No Remote Memory Access Model)
  - NUMA (Non-Uniform Memory Access Model)
  - COMA (Cache-Only Memory Access Model)

Classification

- UMA



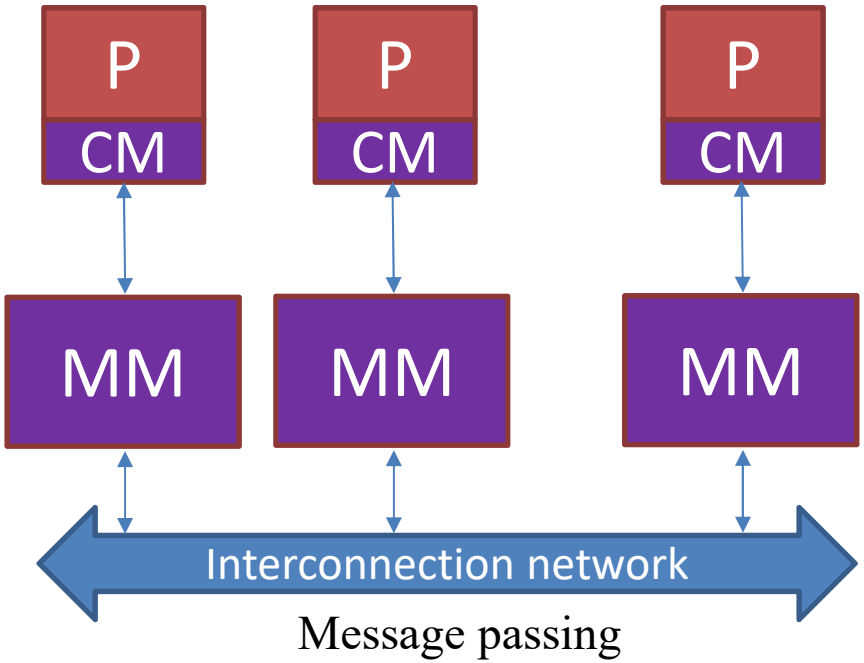
P – CPU or CPU core
CM – Cache memory
MM – Main memory

Classification

- UMA (Uniform Memory Access)
  - Shared memory
  - Uniform access
    - Equal access to memory
    - The same access time
    - No mater which processor
  - Unique physical address space
  - Symmetric multi-processors (SMP)

Classification

- NORMA



Classification

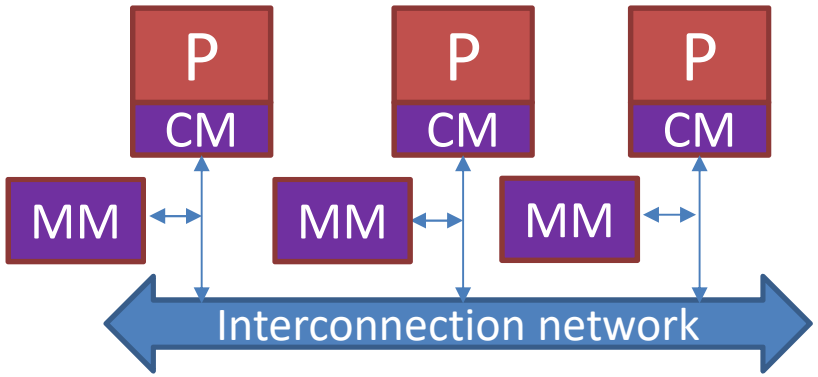
- NORMA (No Remote Memory Access)
  - Distributed memory
  - Access to local memory only
  - No access to remote memory
  - Distinct address spaces for every processor
  - Inter-processors communication using messages

Classification

- NUMA (Non-Uniform Memory Access)
  - Distributed-shared memory
  - Each processor has access to both local memory and remote memory
  - Non-uniform access to memory
    - Local memory low access times
    - Remote memory high access times
  - Unique virtual address space

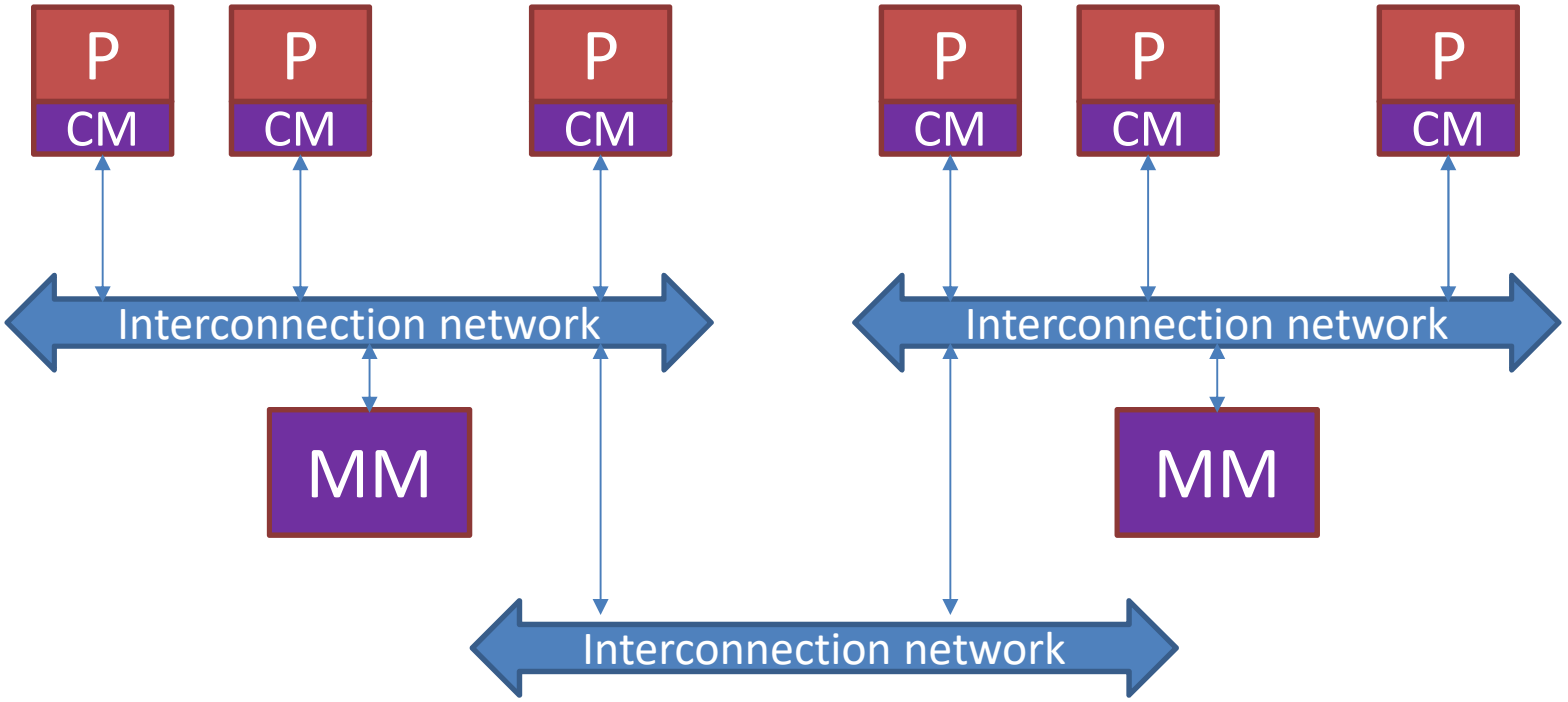
Classification

- NUMA



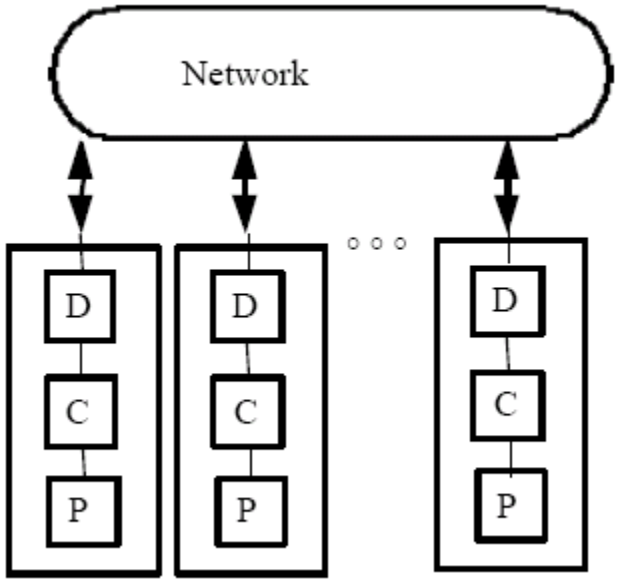
Classification

- Hybrid UMA/NUMA - general



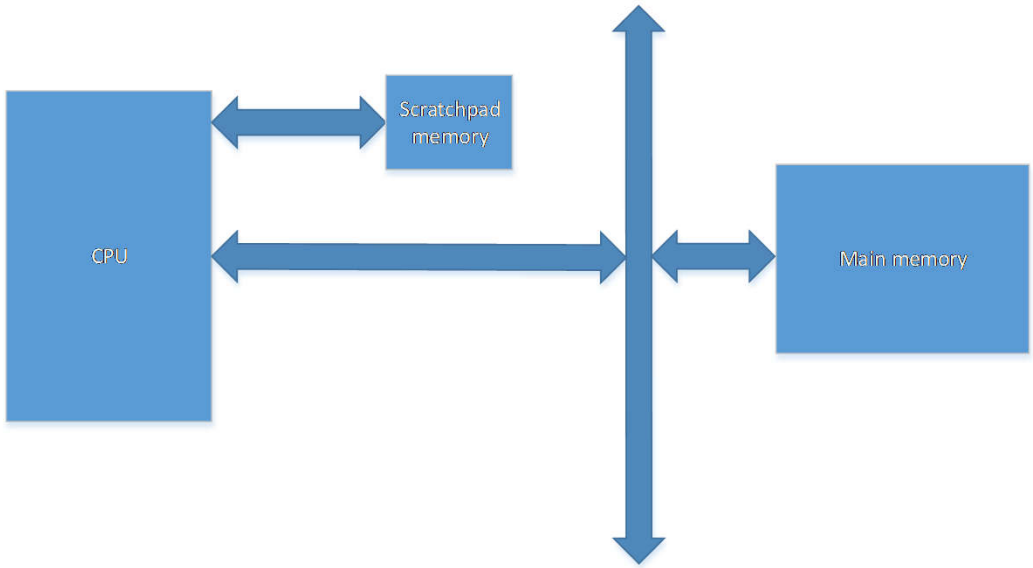
Classification

- COMA (Cache-Only Memory Access)
  - Cache memory organization



Classification

- Scratchpad memory
  - Local memory dedicated to store frequently used data



Classification

- What are the main parameters used to compare different systems?

Classification

- Parameters
  - Performance
  - Number of processors
  - Number of cores
  - Interconnection network
  - Memory hierarchies
  - Granularity
    - Fine grain – simple operations on large amount of data
    - Coarse grain – complex operations on small amount of data

Classification

- Flynn
  - Instructions stream
  - Data stream
- Flynn taxonomy:
  - SISD (Single Instruction stream Single Data stream)
  - **SIMD** (Single Instruction stream Multiple Data stream)
  - MISD (Multiple Instruction stream Single Data stream)
  - **MIMD** (Multiple Instruction stream Multiple Data stream)

Classification

- SISD
  - Instructions are executed sequentially
  - Micro-operations parallelism allowed (pipeline)
  - Multiple functional units (math-coprocesor, graphic processor, I/O processor)

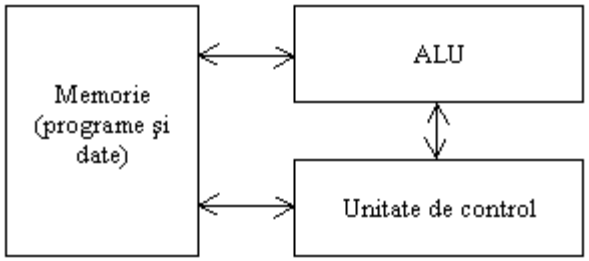


Figura 1.1 – Arhitectura von Neumann

Classification

- SISD
  - UC – command unit
  - UE – execution unit
  - MM – memory module
  - SI – instruction stream
  - SD – data stream

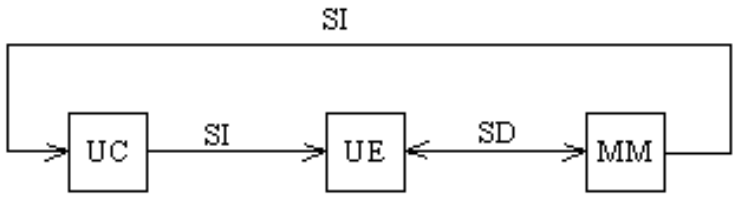


Figura 1.2 Arhitectura SISD

Classification

- SIMD
  - One UC controls many UEs
  - UEs execute simultaneously the same instruction on distinct data
  - Large number of UEs (thousands)
  - Applications having fine grain data processing

Classification

- SIMD
  - Processors arrays
  - Vector processors
  - Graphical processors

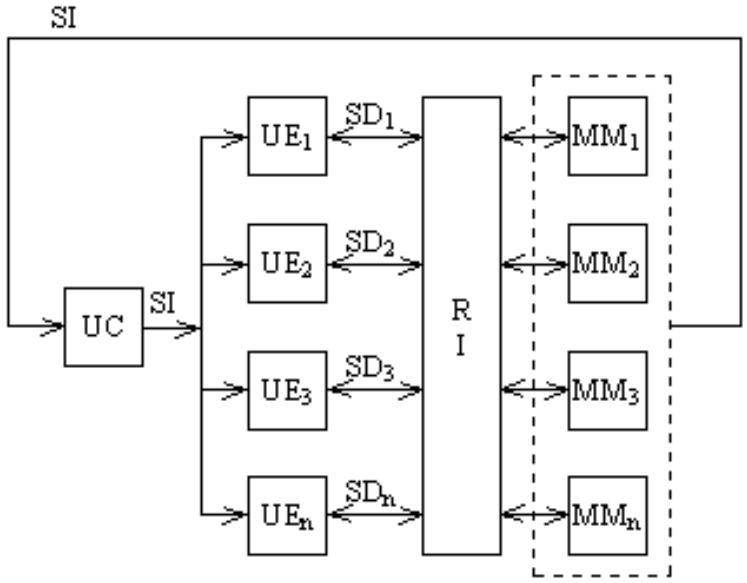


Figura 1.3 Arhitectura SIMD

Classification

- MISD
  - The same data is processed by different UEs
  - Macro-pipeline
- Systolic arrays

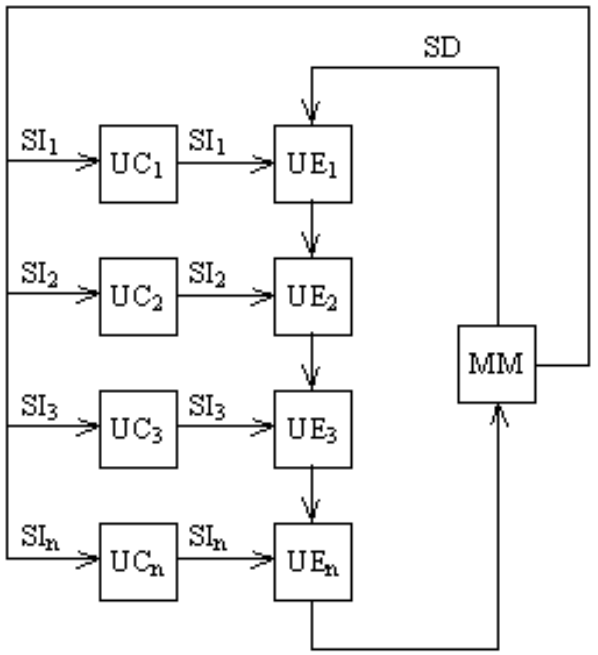


Figura 1.4 Arhitectura MISD

Classification

- MIMD
  - Each UE has its own UC
  - Every UE executes instructions on local data

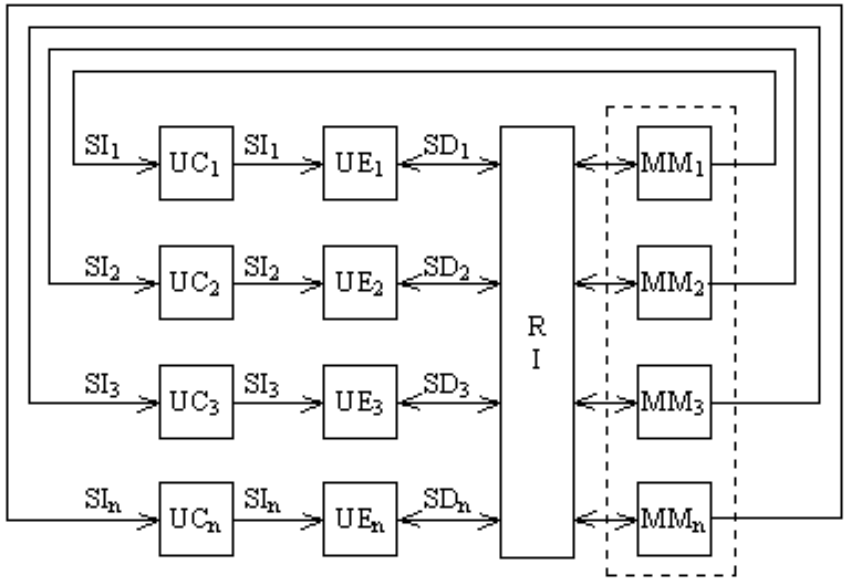


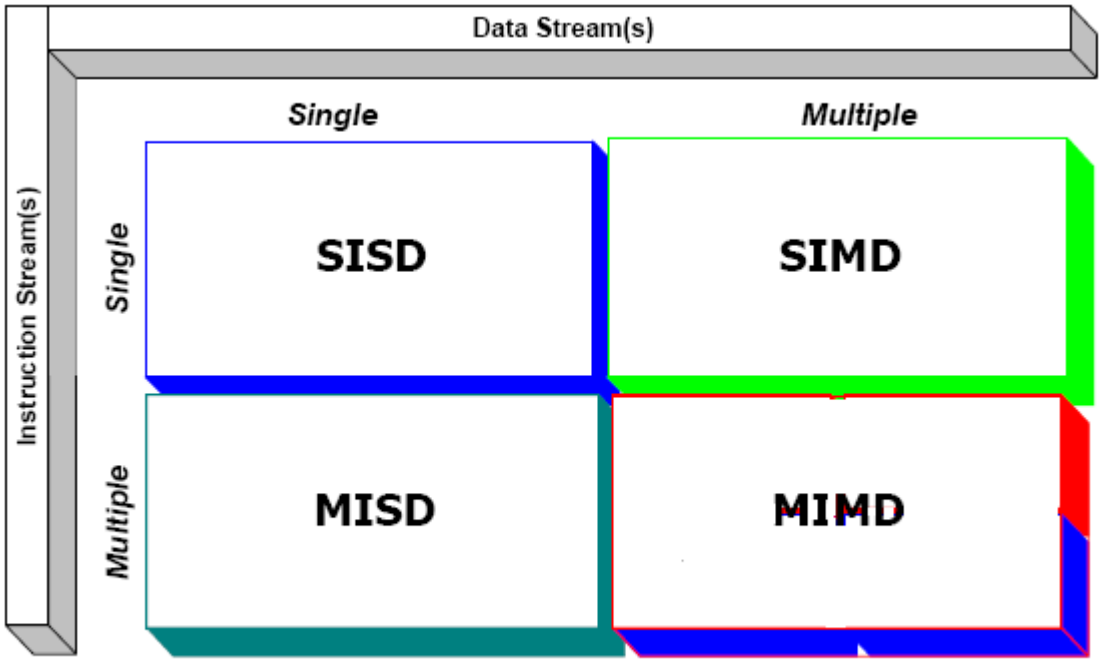
Figura 1.5 Arhitectura MIMD

Classification

- MIMD
  - Multi-processor servers
  - Multi-core processors
  - Computer networks
- Rack based servers
  - Blades
  - Hundreds of cores
  - Hundreds of GB of memory

Classification

- Extension of Flynn taxonomy
  - SISD
  - MIMD

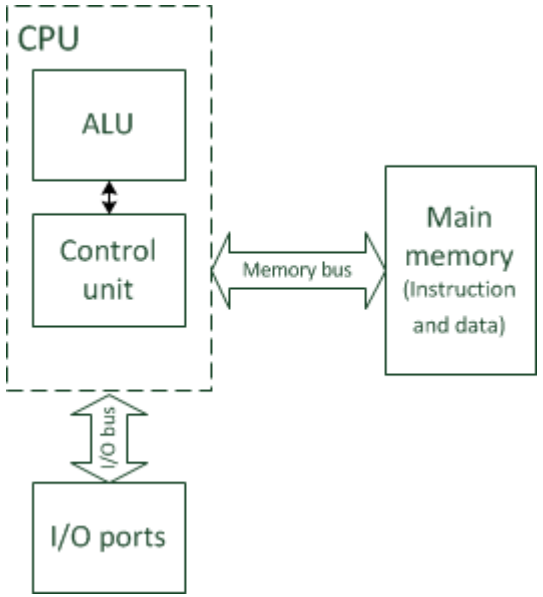


Classification

- SISD
  - Memory bus
    - Instruction
    - Data
  - Princeton vs. Harvard architectures

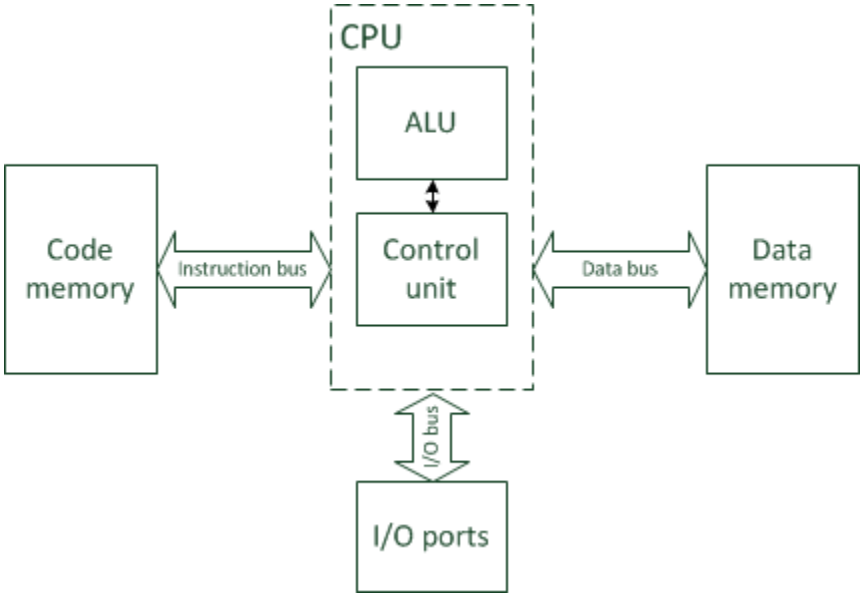
Classification

- Princeton architecture/ von Neumann architecture
  - Common instruction and data bus



Classification

- Harvard architecture
  - One instruction bus and one data bus



Classification

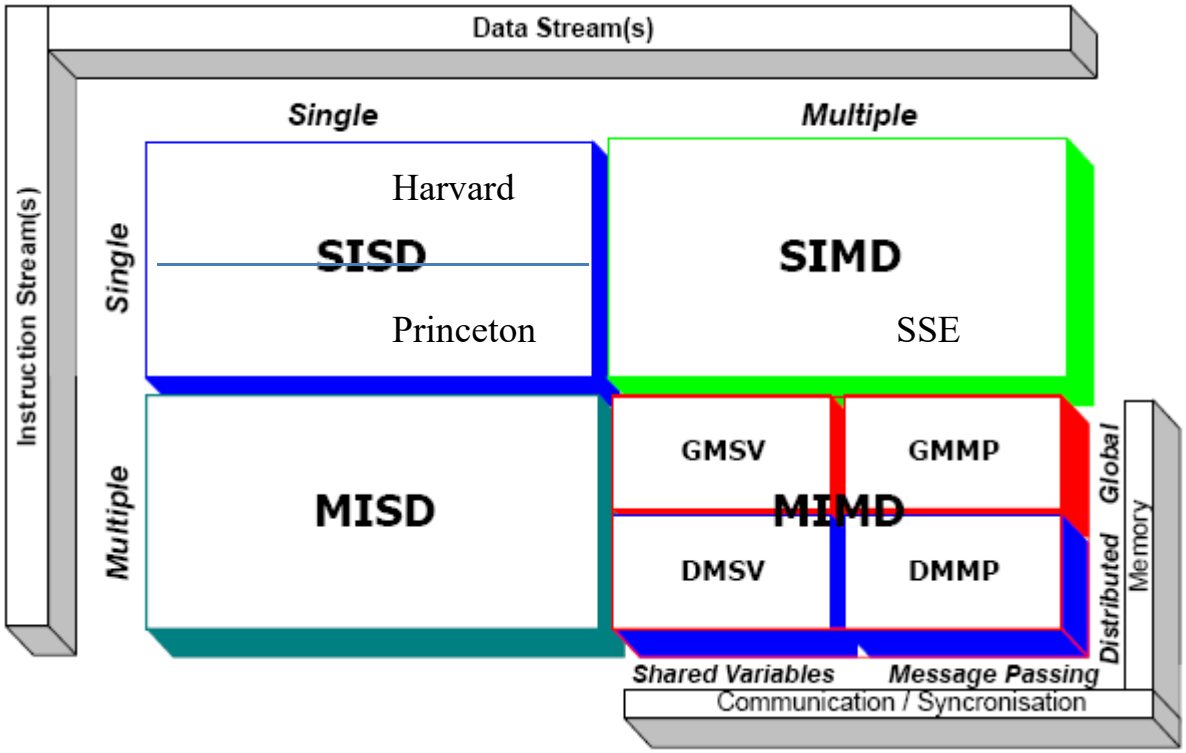
- Intel (x86) and ARM architectures are von Neumann architectures
- C compiler assumes a target machine based on von Neumann architecture

Classification

- MIMD architectures can be further classified based on:
  - Inter-processor communication:
    - Shared memory or messages
  - Interconnection network
    - Bus or crossbar
  - Memory organization:
    - Shared or distributed
  - Coupling level between the nodes
    - Loosely or tightly

Classification

- Flynn-Johnson taxonomy

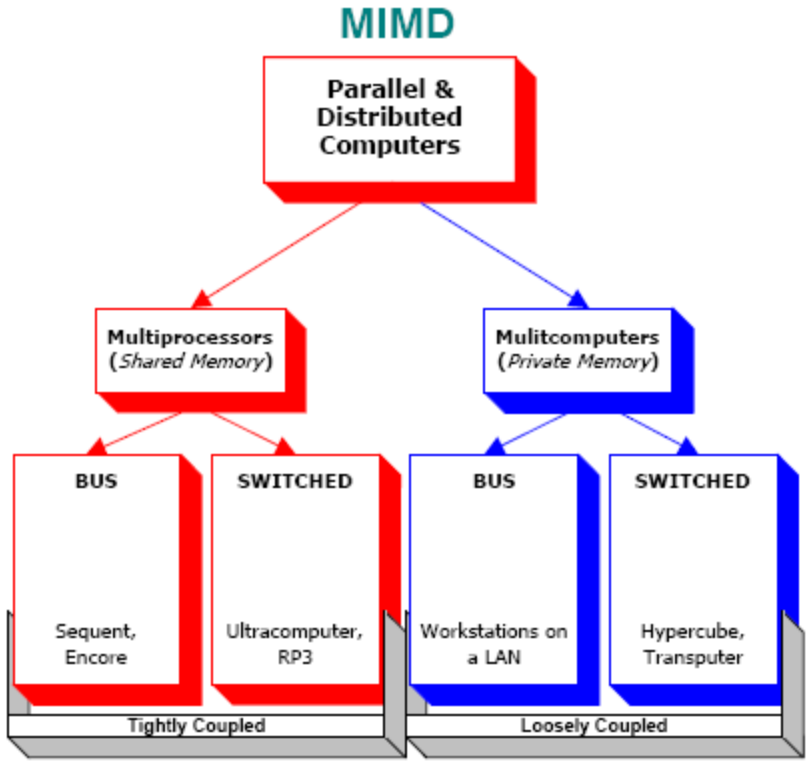


Classification

- GMSV – Global Memory / Shared Variables.
  - Multi-core/ multi-processor
- GMMP – Global Memory / Message Passing
  - virtualization
- DMSV – Distributed Memory / Shared Variables.
  - Distributed-shared memory/ middleware
- DMMP – Distributed Memory / Message Passing.
  - Computer networks, distributed computing, grid

Classification

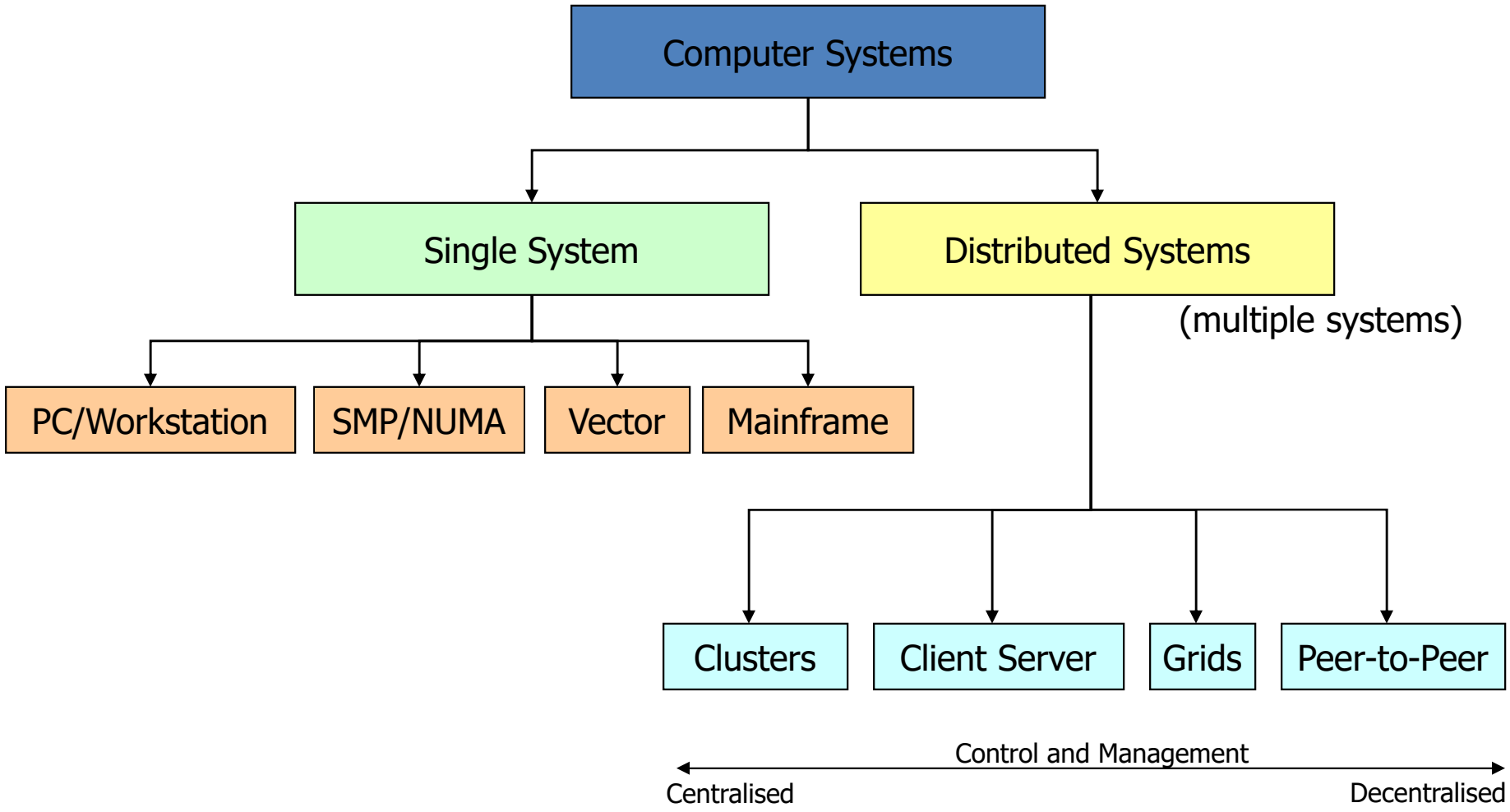
- Tanenbaum taxonomy



Classification

- Interconnection network topologies
  - Static (ring, tree, hypercube, mesh)
  - Dynamic (bus, switches/crossbar)
- Coupling
  - Tightly coupled systems
  - Loosely coupled systems

Summary


Computer Systems
Single System
Distributed Systems
(multiple systems)
PC/Workstation
SMP/NUMA
Vector
Mainframe
Clusters
Client Server
Grids
Peer-to-Peer
Control and Management
Centralised
Decentralised

Summary

- Memory organization
  - UMA
  - NUMA
  - NORMA

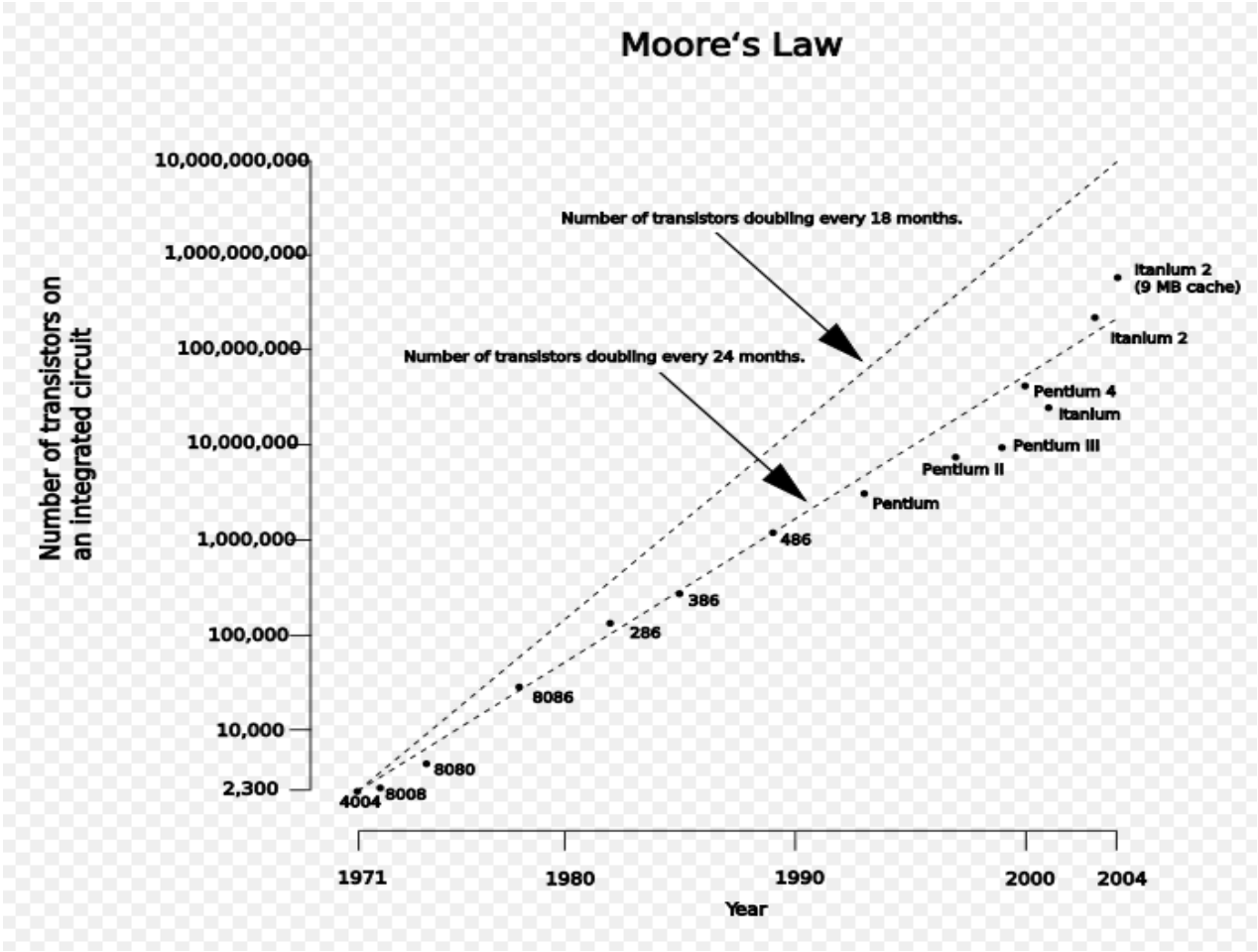
Summary

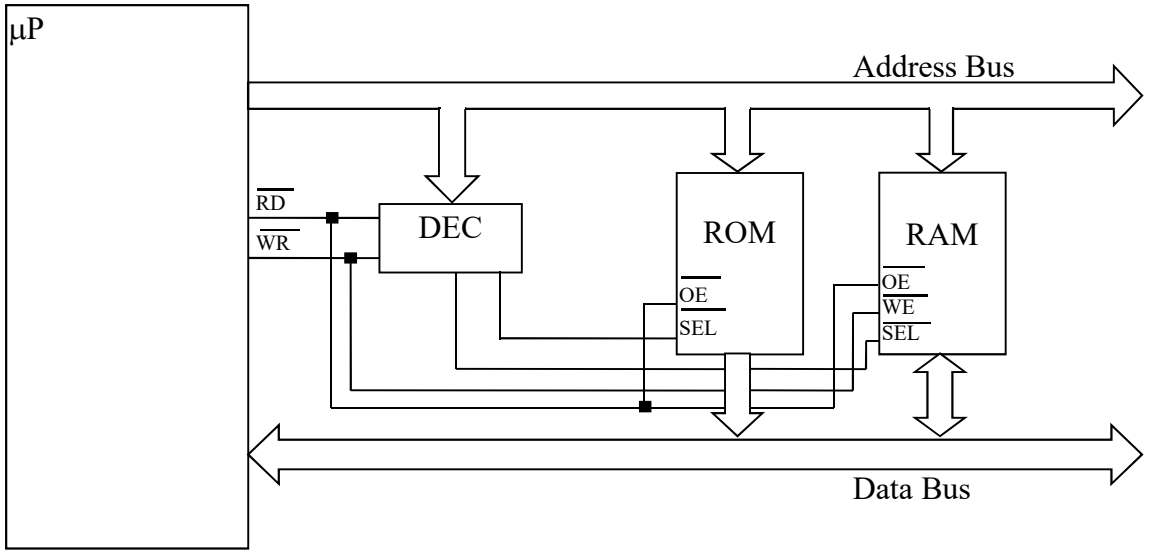
- Flynn taxonomy
  - SISD
    - Harvard architecture
    - Princeton architecture
  - SIMD
  - MISD
  - MIMD
    - GMSV
    - DMSV
    - GMMP
    - DMMP

Summary

- Microprocessor interfaces
- Microarchitecture
- Buses

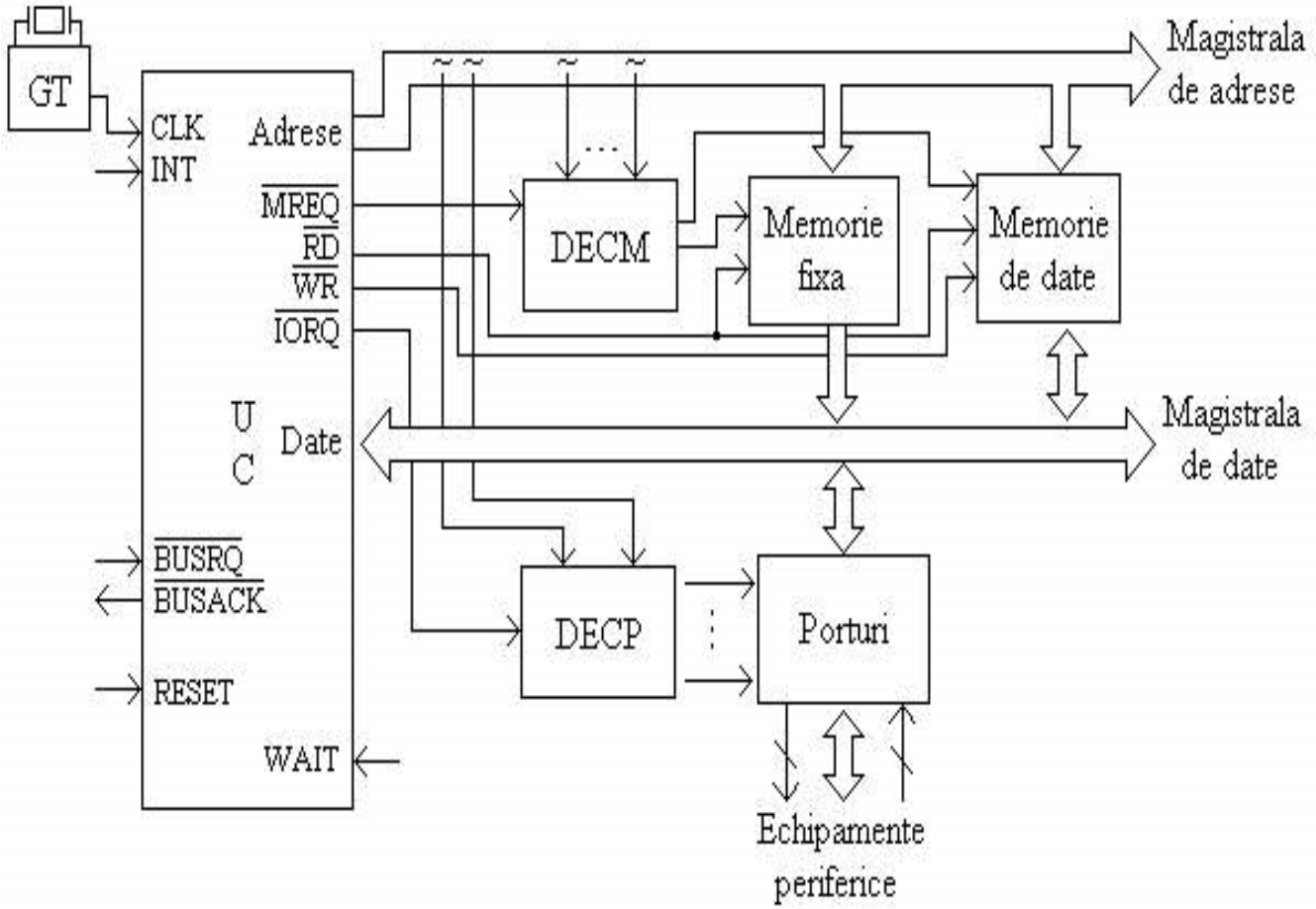
Istoric


Moore's Law
Number of transistors on an integrated circuit
10,000,000,000
1,000,000,000
100,000,000
10,000,000
1,000,000
100,000
10,000
2,300
Number of transistors doubling every 18 months.
Number of transistors doubling every 24 months.
Itanium 2 (9 MB cache)
Itanium 2
Pentium 4
Itanium
Pentium III
Pentium II
Pentium
486
386
286
8086
8080
8008
4004
1971
1980
1990
2000
2004
Year


μP
Address Bus
RD
WR
DEC
ROM
OE
SEL
RAM
OE
WE
SEL
Data Bus

Definitions

- External connectivity of a microprocessor



Classification

- MIMD extensions

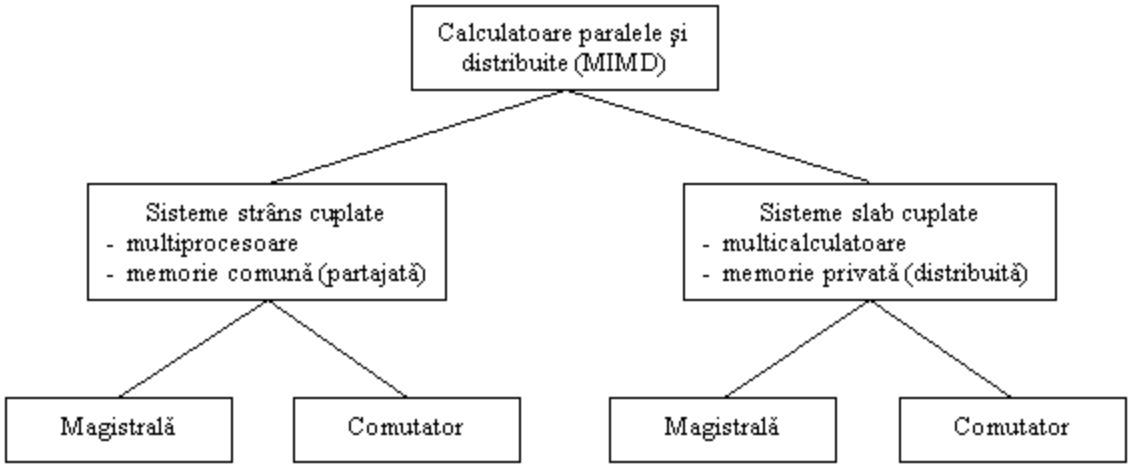


Figura 1.6 Clasificarea arhitecturilor MIMD