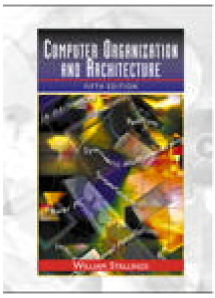
Campus de Gualtar
4710-057 Braga

UNIVERSIDADE DO MINHO
ESCOLA DE ENGENHARIA

Departamento de Informática



Computer Organization and Architecture

5th Edition, 2000

by William Stallings

Table of Contents

I. OVERVIEW. (25-Jan-99)

1. Introduction.
2. Computer Evolution and Performance.

Organization and Architecture (1.1)

- Computer Architecture refers to those attributes of a system that have a direct impact on the logical execution of a program. Examples:
  - the instruction set
  - the number of bits used to represent various data types
  - I/O mechanisms
  - memory addressing techniques
- Computer Organization refers to the operational units and their interconnections that realize the architectural specifications. Examples are things that are transparent to the programmer:
  - control signals
  - interfaces between computer and peripherals
  - the memory technology being used
- So, for example, the fact that a multiply instruction is available is a computer architecture issue. How that multiply is implemented is a computer organization issue.

Structure and Function (1.2)

- Modern computers contain millions of electronic components
- The key to describing such systems is to recognize their hierarchical nature
  - They are a set of layers or levels of interrelated subsystems
  - Each level consists of a set of components and their inter-relationships
- The behavior of each level depends only on a simplified, abstracted characterization of the system at the next lower level
- At each level, the designer is concerned with:
  - Structure: The way in which the components are interrelated
  - Function: The operation of each individual component as part of the structure.
- We will usually describe systems from the top-down, instead of bottom-up.

Function

- A functional view of the computer
- Basic functions that a computer can perform:
  - Data Processing - a wide variety of forms, but only a few fundamental methods or types
  - Data Storage - long-term or short, temporary storage
- Data Movement
  - Input/Output - when data are received from or delivered to a peripheral, a device connected directly to the computer
  - Data Communications - when data is moved over longer distances, to or from a remote device

- Control - of the above functions, by instructions provided by the user of the computer (i.e. their programs)
- 4 Possible types of operations with this basic structure

  Device for Processing Data in Storage
  Device for Processing Data En-route Between the Outside World and Storage

Structure

- Simplest possible view of a computer:
  - Storage
  - Processing
  - Peripherals
  - Communication Lines

- Internal Structure of the Computer Itself:
  - Central Processing Unit (CPU): Controls the operation of the computer and performs its data processing functions. Often simply referred to as processor.
  - Main Memory: Stores data.
  - I/O: Moves data between the computer and its external environment.
  - System Interconnection: Some mechanism that provides for communication among CPU, main memory, and I/O.

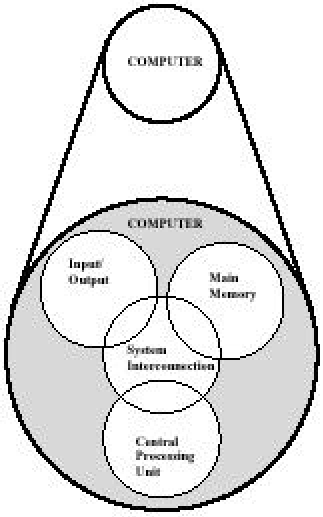


Figure 1.4 The Computer: Top-Level Structure

- Main Structural components of the CPU:
  - Control Unit: Controls the operation of the CPU and hence the computer.
  - Arithmetic and Logic Unit (ALU): Performs the computer's data processing functions.
  - Registers: Provides storage internal to the CPU.
  - CPU Interconnection: Some mechanism that provides for communication among the control unit, ALU, and registers.

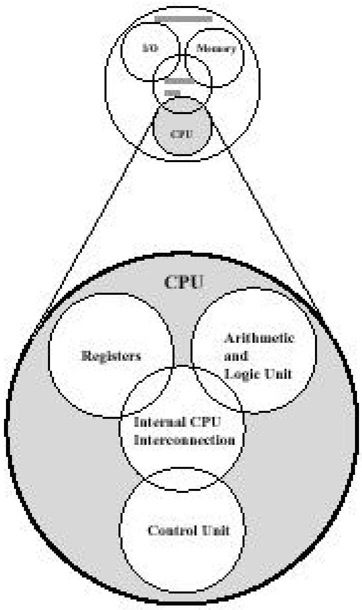


Figure 1.5 The Central Processing Unit (CPU)

- (Microprogrammed) Control Unit Structure

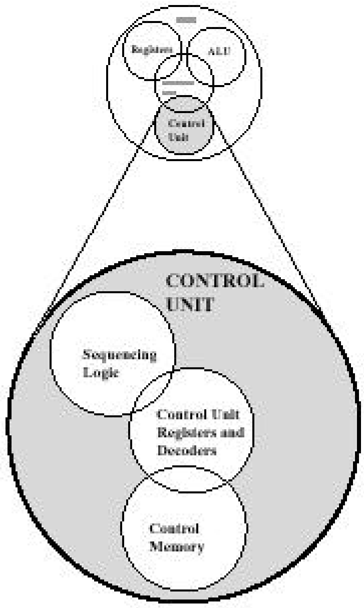
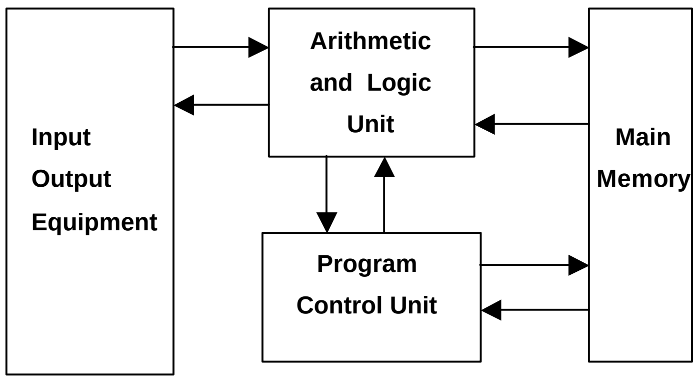


Figure 1.6 The Control Unit

A Brief History of Computers (2.1)

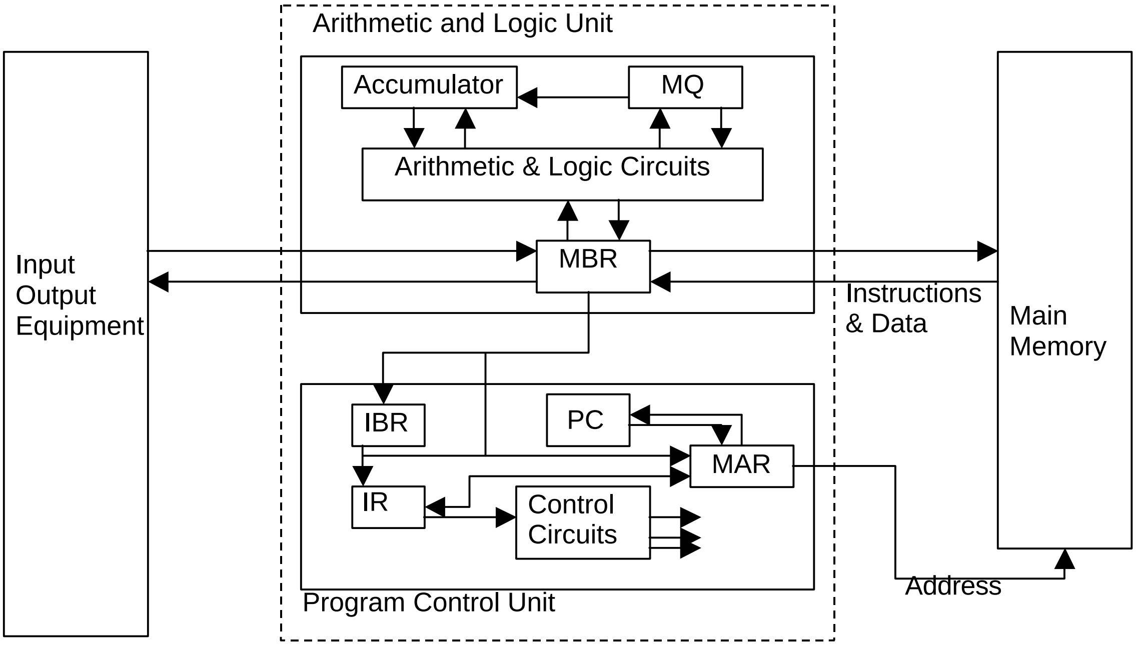
- First Generation: Vacuum Tubes
  - 1943-1946: ENIAC
    - first general purpose computer
    - designed by Mauchly and Eckert
    - designed to create ballistics tables for WWII, but too late -- helped determine H-bomb feasibility instead. General purpose!
    - 30 tons + 15000 sq. ft. + 18000 vacuum tubes + 140 KW = 5000 additions/sec

- von Neumann Machine
  - 1945: stored-program concept first implement for EDVAC. Key concepts:
    - Data and instructions are stored in a single read-write memory
    - The contents of this memory are addressable by location, without regard to the type of data contained there
    - Execution occurs in a sequential fashion (unless explicitly modified) from one instruction to the next



- 1946: Princeton Institute for Advanced Studies (IAS) computer
  - Prototype for all subsequent general-purpose computers. With rare exceptions, all of today's computers have this same general structure, and are thus referred to as von Neumann machines.

- General IAS Structure Consists of:



  - A main memory, which stores both data and instructions
  - An ALU capable of operating on binary data
  - A control unit, which interprets the instructions in memory and causes them to be executed
  - I/O equipment operated by the control unit
- First commercial computers
  - 1950: UNIVAC - commissioned by Census Bureau for 1950 calculations
  - late 1950's: UNIVAC II
    - greater memory and higher performance
    - same basic architecture as UNIVAC
    - first example of upward compatibility
  - 1953: IBM 701 - primarily for science
  - 1955: IBM 702 - primarily for business
- Second Generation: Transistors
  - 1947: Transistor developed at Bell Labs
  - Introduction of more complex ALU and control units
  - High-level programming languages
  - Provision of system software with computers
  - The data channel - an independent I/O module with its own processor and instruction set
  - The multiplexor - a central termination point for data channels, CPU, and memory. Precursor to idea of data bus.
- Third Generation: Integrated Circuits
  - 1958: Integrated circuit developed
  - 1964: Introduction of IBM System/360
    - First planned family of computer products. Characteristics of a family:
      - Similar or Identical Instruction Set and Operating System

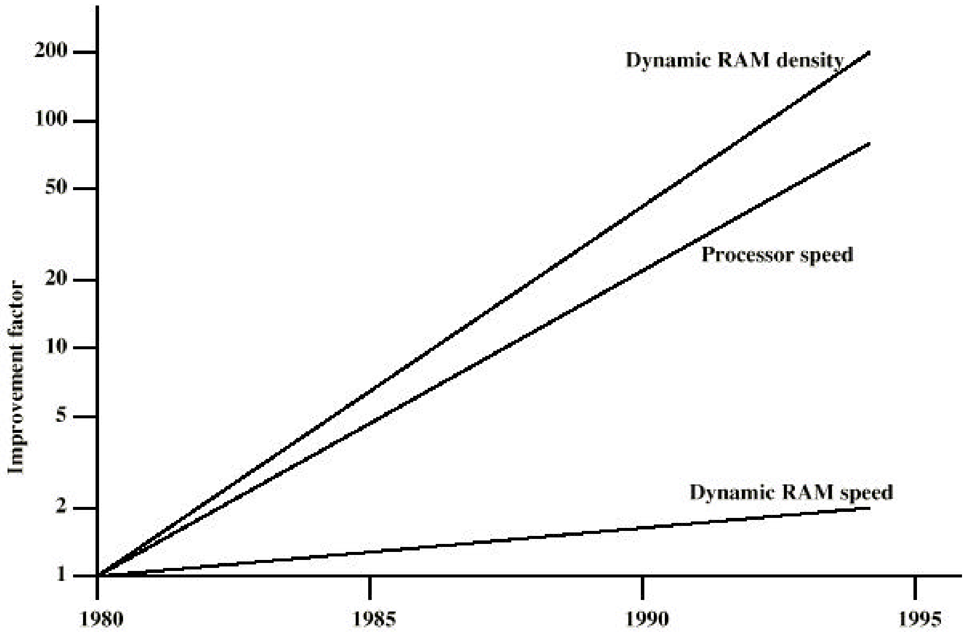
- Increasing Speed
- Increasing Number of I/O Ports
- Increasing Memory Size
- Increasing Cost

- Different models could all run the same software, but with different price/performance

- 1964: First PDP-8 shipped
  - First minicomputer
  - Started OEM market
  - Introduced the bus structure
- Fourth Generation: No clear characterization
  - Semiconductor memory
    - Replaced bulky core memory
    - Goes through its own generations in size, increasing by a factor of 4 each time: 1K, 4K, 16K, 64K, 256K, 1M, 4M, 16M on a single chip w/ declining cost and access time
  - Microprocessors and personal computers
  - Distributed computing
  - Larger and larger scales of integration

Designing for Performance (2.2)

- Evolution of Computer Systems
  - Price/performance
    - price drops every year
    - performance increases almost yearly
    - memory size goes up a factor of 4 every 3 years or so
  - The basic building bocks for today's computers are the same as those of the IAS computer nearly 50 years ago
- Microprocessor Speed
  - Density of integrated circuits increases by 4 every 3 years (e.g. memory evolution)
  - Also results in performance boosts of 4-5 times every 3 years



- o Requires more elaborate ways of feeding instructions quickly enough. Some techniques:
  - ▪ Branch prediction
  - ▪ Data-flow analysis
  - ▪ Speculative Execution

- Performance Balance
  - o All components do not increase performance at same rate as processor
  - o Results in a need to adjust the organization and architecture to compensate for the mismatch among the capabilities of the various components
- Example: Interface between processor and main memory
  - o Must carry a constant flow of program instructions and data between memory chips and processor
  - o Processor speed and memory capacity have grown rapidly
  - o Speed with which data can be transferred between processor and main memory has lagged badly
  - o DRAM density goes up faster than amount of main memory neededd
    - ▪ Number of DRAM's goes down
    - ▪ With fewer DRAM's, less opportunity for parallel data transfer
- Some solutions
  - o Make DRAM's "wider" to increase number of bits retrieved at once
  - o Change DRAM interface to make it more efficient
  - o Reduce frequency of memory access using increasingly complex and efficient cache structures
  - o Increase interconnect bandwidth with higher-speed buses and bus hierarchies
- I/O devices also become increasingly demanding
- Key is balance. Because of constant and unequal changes in:
  - o processor components
  - o main memory
  - o I/O devices
  - o interconnection structures

  designers must constantly strive to balance their throughput and processing demands.