Introduction

Loriano Storchi

loriano@storchi.org

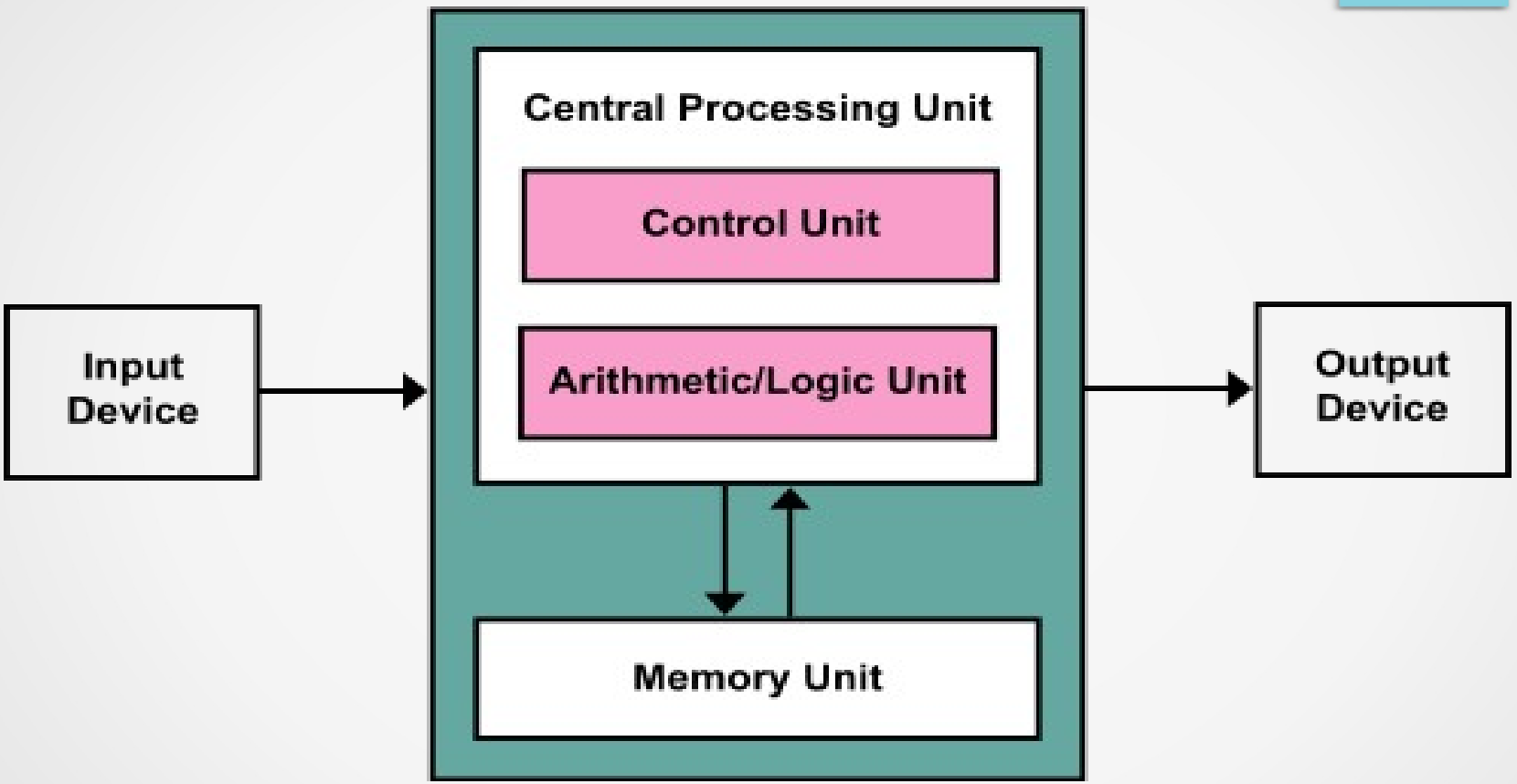
http::://www.storchi.org/

Computer science

- Systematic study of the feasibility, structure, expression, and mechanization of the methodical procedures (or algorithms) that underlie the acquisition, representation, processing, storage, communication of, and access to, information.

HARDWARE

Von Neumann



SYSTEM BUS, connection (Harvard architecture separation of data memory and memory containing the program)

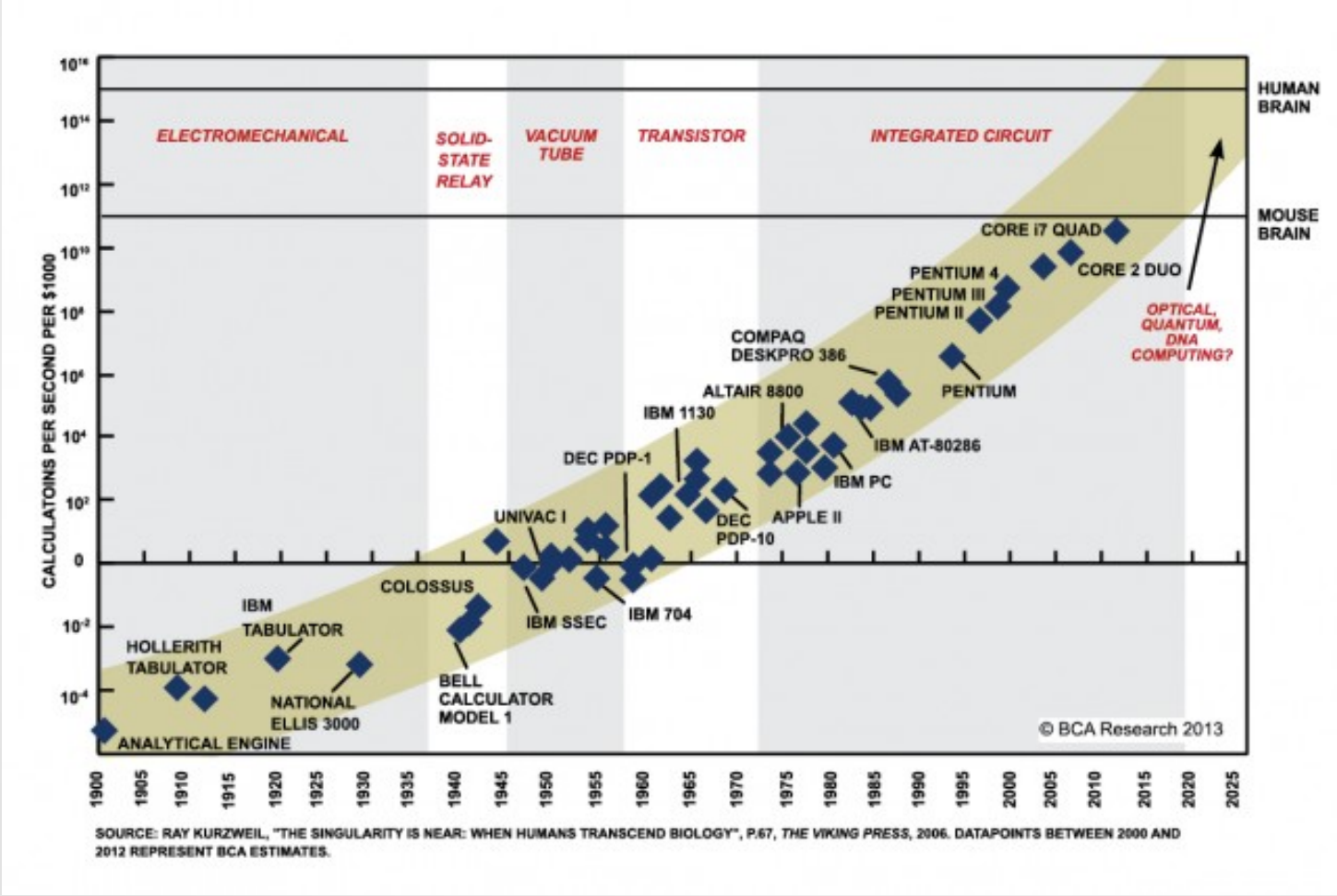
C.P.U.

C.P.U.

C.P.U. Central Processing Unit, coordinates and manages all the hardware devices to acquire, interpret and execute program instructions

- C.U. (Control unit): interprets and executes instructions
- A.L.U. (Arithmetic Logic Unit): starting from the operands (input data to be operated on) and a code indicating the operation to be performed, the ALU outputs the result of the performed operation.
- REGISTERS: small amount of fats storage inside the CPU
- CLOCK: it scans the time intervals in which the internal devices of the CPU operate. It determines the speed expressed as the number of intervals in the time unit

Moore’s law



Moore's Law is the observation made by Intel co-founder Gordon Moore that the number of transistors on a chip doubles every year while the costs are halved. In 1965, Gordon Moore noticed that the number of transistors per square inch on integrated circuits had doubled every year since their invention. Moore's law predicts that this trend will continue into the foreseeable future.

C.P.U.

Modern CPUs:

- Pipeline: extension of the internal structure of a CPU, to execute in parallel various phases connected to the loading, interpretation and execution of the instructions. Each phase is entrusted to a specific "sub-unit"
- Out of order execution, Speculative execution, Superscalar processor
- Co-processors: processors that are specialized in performing specific functions:
  - Performing numerical calculations
  - GPU (Graphic processing unit)
- Parallelism: multi-core (replicated CPU), in order to execute several instructions in parallel

MEMORY

Central memory

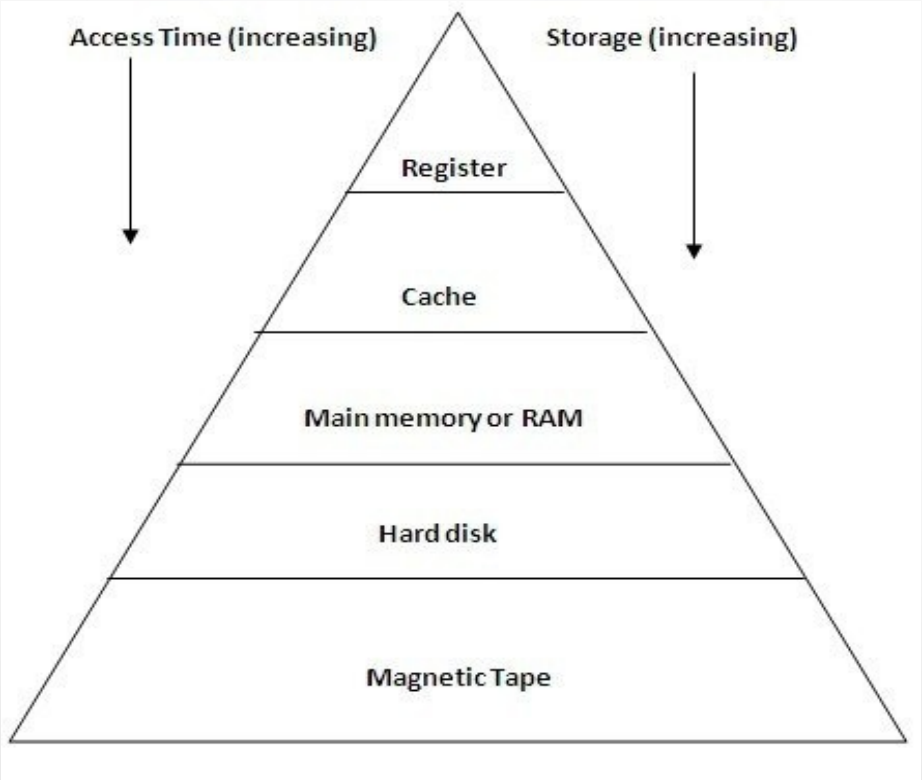
- RAM (Random Access Memory)
- The RAM contains the instructions (opcodes) that will be executed and the data on which these instructions will operate
- Main characteristics:
  - Volatile: the content is lost when the computer is turned off
  - Fast: about 100 clock cycles, fast therefore expensive
  - Small size compared to the mass storage, typically order of some GiB

Mass storage

- Made up of hard disks, CDs, DVDs, solid state disks, tapes
  - Non-volatile
  - Slow compared to RAM (> 10000 cycles)
  - Not expensive if compare to the RAM
  - Large (now hundreds of GiB or some TiB

Memory Hierarchy

The memory hierarchy in the modern computer is composed of different levels, each one characterized by a speed of access to data that is inversely proportional to the size



UNIT OF MEASUREMENT

UNIT OF MEASUREMENT: information

- BIT = it is the unit of measure of information (from English "binary digit"), defined as the minimum quantity of information that is needed to discern between two possible equitable events. (Wikipedia)
- BYTE = 8 BIT (historically the characters were represented by 8 BIT, which is why 1 Byte still remains the unit of minimum addressable memory)
- “KiloByte” maybe better “kibibyte” KiB = 2^{10} Byte = 1024 Byte
- “MegaByte” meybe better “mebibyte” MiB = 1024 * 1024 Byte
- “GigaByte” meybe better “gibibyte” GiB = 1024 * 1024 * 1024 Byte
- “TeraByte” meybe better “tebibyte” TiB = 1024 * 1024 * 1024 * 1024 Byte

FLOPS

FLOPS it is an abbreviation for Floating Point Operations Per Second, and it indicates the number of floating point operations that a CPU executes in a second..

For example, in the case of a classic product between matrices, $2*N^3$ operations are performed, so I can evaluate the FLOPS exactly by measuring the time necessary to perform this multiplication and obtain :

$$[flops] = 2*N^3 / time$$

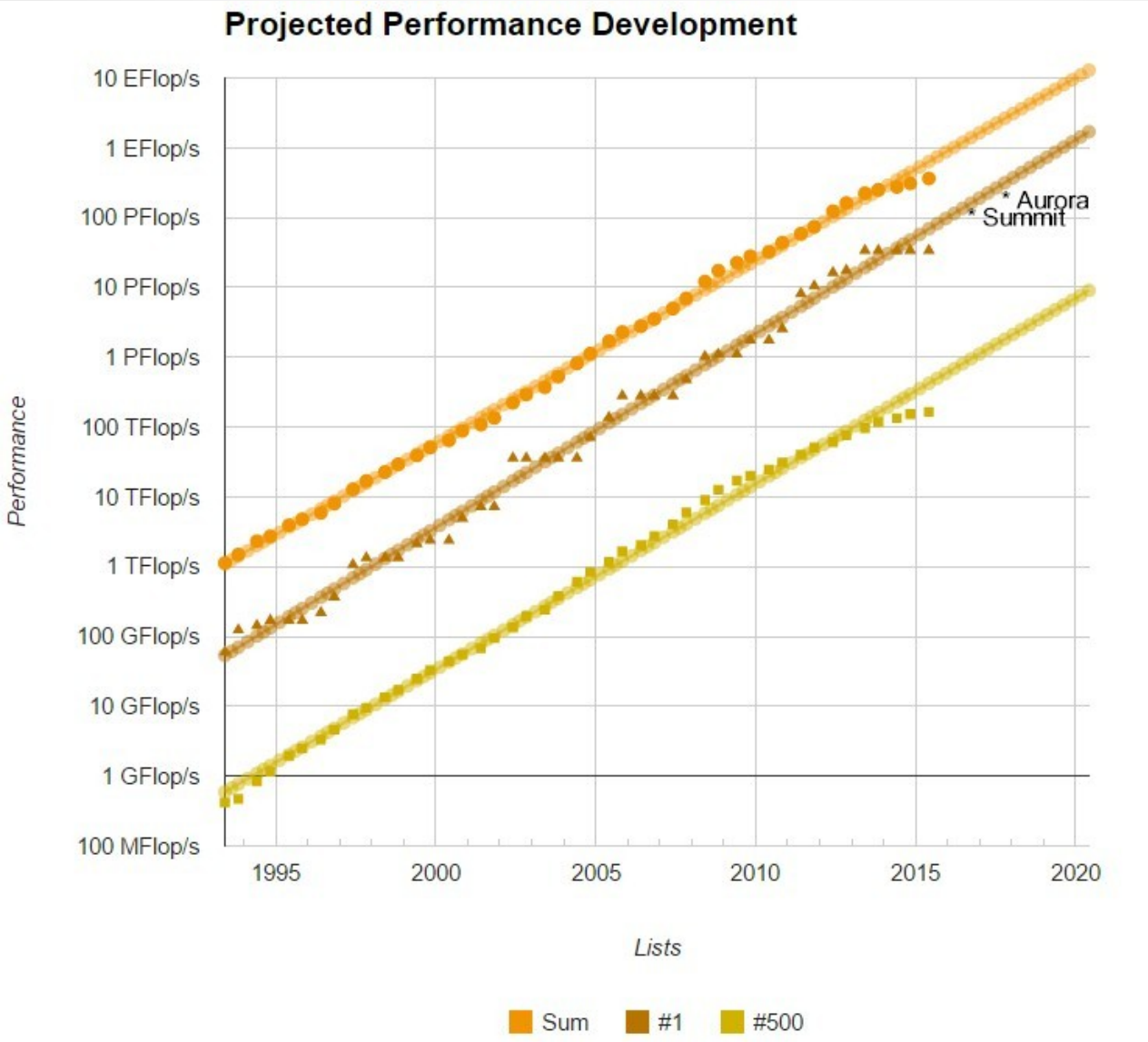
TOP500

- Clearly there is a difference between sustained performance and maximum performance
- In order to objectively evaluate the performance of a computer there is a need for a reference test, a standard benachmark, for example Linpack
- TOP500 http://www.top500.org/, ranking of the 500 most powerful computers in the world

Top500 list November 2018

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM DOE/SC/Oak Ridge National Laboratory United States	2,397,824	143,500.0	200,794.9	9,783
2	**Sierra** - IBM Power System S922LC, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband , IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	125,712.0	7,438
3	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
4	**Tianhe-2A** - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000 , NUDT National Super Computer Center in Guangzhou China	4,981,760	61,444.5	100,678.7	18,482
5	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland	387,872	21,230.0	27,154.3	2,384
6	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/NNSA/LANL/SNL United States	979,072	20,158.7	41,461.2	7,578
7	**AI Bridging Cloud Infrastructure (ABCI)** - PRIMERGY CX2570 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR , Fujitsu National Institute of Advanced Industrial Science and Technology (AIST) Japan	391,680	19,880.0	32,576.6	1,649

Top500 list : historical chart



SOFTWARE

Software

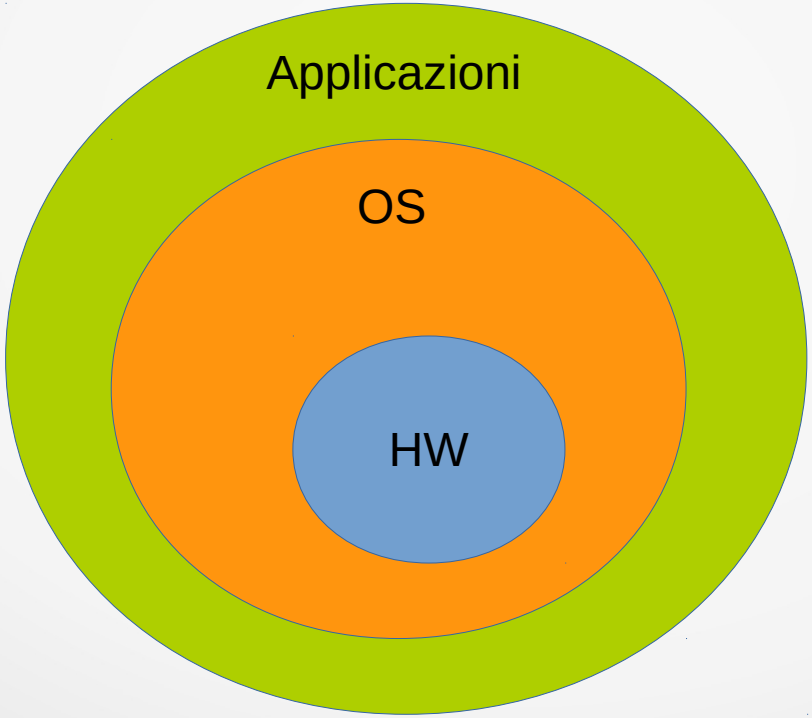
- Basic software: dedicated to the management of the computer itself, for example the Operating System
- Application software: dedicated to the creation of specific applications, such as internet browsing, word processing or other.

Operating System

- It provides high-level functionality making the hardware "easy" to use
- For example, organizes the data in the memory
- Performs user commands
- Run programs and show results on video
- In the case of multi-user systems it must manage the resources and make them usable to all (managed resources parallelism "virtual" multitasking systems)
- Examples are: Linux, Unix, Mac OS X (Darwin), Microsoft Windows

Operating System

Make it easier to write application programs that do not have to worry about the specific characteristics of the HW.



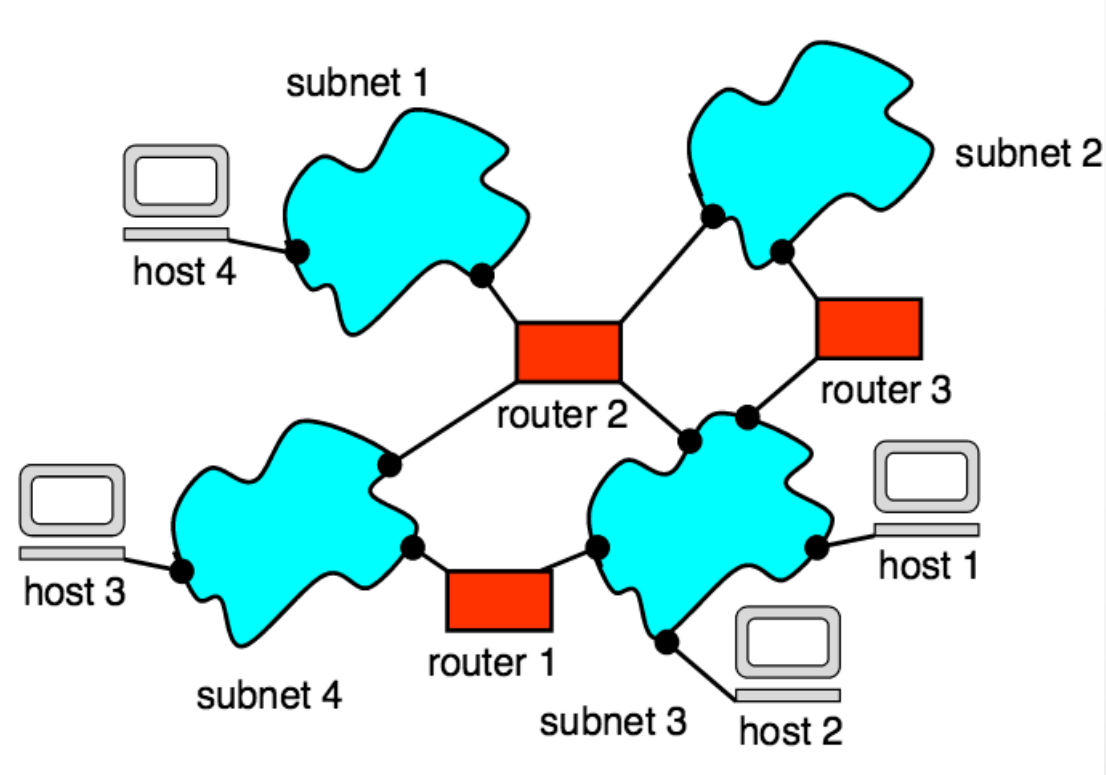
COMPUTER NETWORKS AND INTERNET

Networks

- A computer network is actually a set of computers connected to each other so that they can exchange information, share resources. There are different types of networks.
- LAN (Local Area Network): network on a local scale, these are extended networks at the level of a single room or at most of a building
- MAN (Metropolitan Area Network): for example, you can connect more LANs
- WAN (Wide Area Network): extensive networks over geographical areas. They connect LAN and MAN (Internet is the WAN)

Internet

- Internet was born in the 60/70 years essentially for military purposes
- Internet today interconnects thousands of subnets



Host: computer connected to the internet, it can be a client or a server at the application level

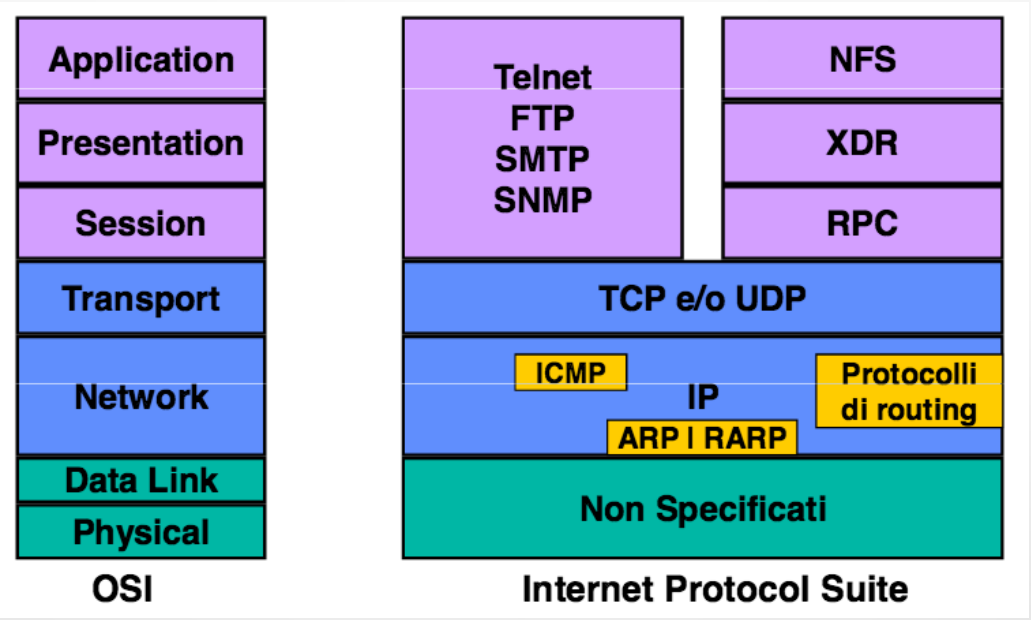
Router: node that serves to route traffic

Subnet: set of hosts between which there is a level 2 connection (for example a LAN)

Today, the order of billions of hosts are interconnected in Internet

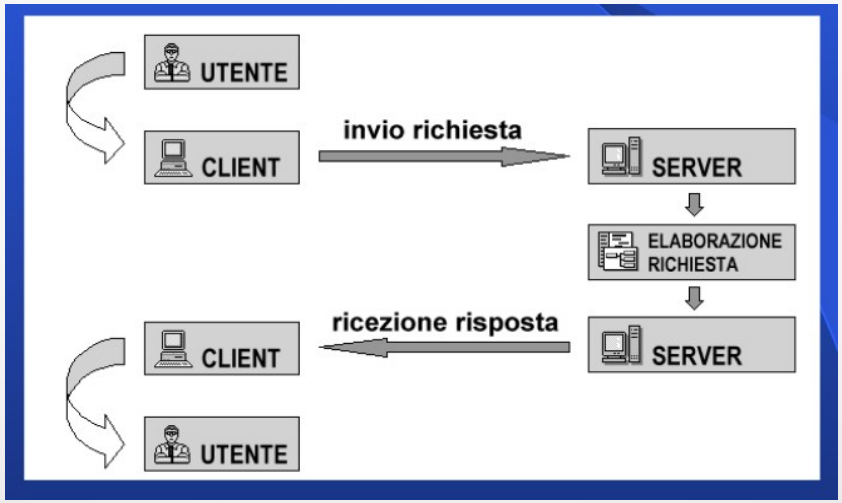
Internet

- Common rules are needed to allow different hosts to communicate. The TCP / IP Protocol: Transfer Control Protocol / Internet Protocol (difference between uman communication protocol and the TCP/IP)
- The information (data) is divided into packages that are then recomposed to destination (each package can follow different paths)



Client/Server

- Most Internet services offered by the Internet are based on client / server interaction (*different from P2P*).
- The client is equipped with a particular client software capable to send service requests to a complex server. The client formats the requests in an appropriate and understandable manner to the server, thus using a specific protocol



Protocols

- Different types of protocol are used each for a specific service:
- HTTP (HyperText Transfer Protocol) Access to hypertext pages (WEB) within the WWW
- FTP (File Transfer Protocol) transfer and copy files
- SMTP (Simple Mail Transfer Protocol) Shipping of e-mail messages (e-mail)

A resource on the network is therefore "identified" by the URL:

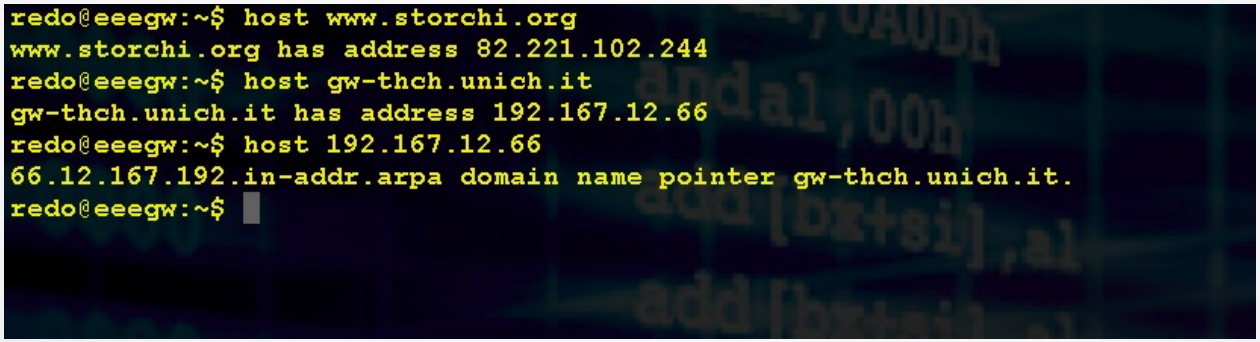
http://nomehost.it/index.html

Addresses

- Each computer connected to the Internet is identified by its IP address, consisting of 4 groups of one byte each (**32-bit total**).
- Each group can assume a maximum value of 255, for example: **192.167.12.66** (static or dynamic IP, private IP ...)
- The last number usually identifies a Host, the numbers preceding the subnet to which this Host belongs.
- The maximum number of IPv4 addresses is therefore 255 * 255 * 255 * 255 (**IPv6 128-bit and so IoT**)

Addresses

- It is in general quite difficult for a human to memorize numbers, much easier to memorize names. There are therefore **DNS (Domain Name System)** services. Thus, systems useful for translating names and addresses into one verse.



Addresses

- Each host is therefore identified by the user with a symbolic name:
  - gw-thch.unich.it
- The names are assigned univocally and managed administratively in a hierarchical manner
- Names uniquely identify a host within a domain:
  - it is the domain
  - unich is the sub-domain within ad it
- The main domains are:
  - .gov .edu .com essentially in the USA associated with the type of organization
  - The various nations instead have domain of the type: .it, .uk, .fr, .de

Unit of measurements

- Data transmission speed = amount of information / transfer time
- In general this speed is expressed in bits per second I.e. **bit / s (or bps also called bit rate)**. The byte by second **byte / s (or Bps)** is also used.
- Then we use the **standard prefixes k (= kilo 10^3), M (= mega 10^6), G (= giga 10^9)**, not the approximations based on the powers of two that are used in computer science.
- Converting **from bps to Bps is simple enough to divide by 8**. For example ADSL 10 Mbps = 10 Mbps / 8 = 1250 KBps
- Having to transfer a **10 MiB** file with a **5 Mbps** line will use approximately (**10 * 1024 * 1024 * 8) / 5 10^6 = 16.8 s**.