Introduction to computer sceince

Loriano Storchi

loriano@storchi.org

http::/www.storchi.org/

MEASUREMENT UNITS

Information temperament units

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

Bench-marking

- **Clearly, increasing the performance of a computer means decreasing the time it takes to perform an operation**.
- $\mathbf{T_{clock}}$ is the clock period of the machine (frequency increase)
- $\mathbf{CPI_i}$ is instead the number of clock "hits" necessary to execute the given instruction i (reduction of complexity for the single instruction)
- Finally, $\mathbf{N^i}$ is the number of type i instructions (for example, sums, jumps ...)

$$T_{execution} = T_{clock} \sum_{i=0}^{n} N_i CPI_i$$

MIPS

MIPS is an abbreviation for Mega Instuctions Per Second, and it indicates the number of generic instructions that a CPU executes in a second. It is a unit of measurement used to measure the performance of a more general use computer than the FLOPS that we will see shortly:

$$\textbf{MIPS} = \textbf{(Frequenza del clock)} / (10^6 \textbf{ CPI})$$

This type of measure does not take into account, for example, the optimizations due to the presence of the cache and the percentages of the different instructions within real programs, and beyond.

FLOPS

FLOPS it is an abbreviation for Floating Point Operations Per Second, and it indicates precisely the number of floating point operations that a CPU performs in a second. It is a unit of measurement used to measure the performance of a computer that is particularly widespread in the field of scientific computing

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 needed to perform this multiplication and obtain:

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

SPEC

Standard Performance Evaluation Corporation is a non-profit organization that produces and maintains a standardized set of computer benchmarks (therefore a set of test programs that are representative of real computer applications).

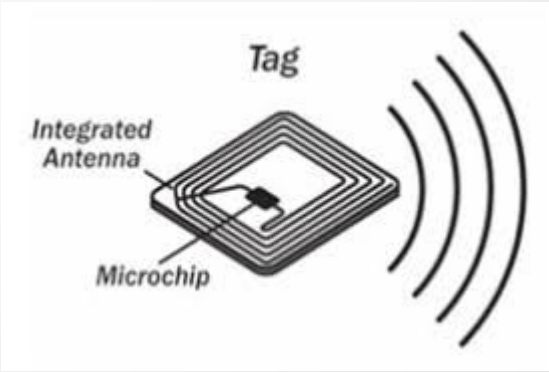
There are several sets of SPECs that are specific for example to different intended uses of the computer

MODERN COMPUTERS

RFID

- **RFID Radio-frequency identification** costs a few cents and can reach some MIPS
- They use electromagnetic fields to automatically identify and track objects using an ID or other information
- Passive if they use energy emitted by the
- RFID reader. Also with battery
- Active necessarily with battery

  periodically send a signal



Embedded Systems, SmartPhone and Tablet

- Today very popular embedded systems are calculators designed for **Watches, cars, household appliances, medical devices, audio / video players (Android Box)**. They are calculators costing a few tens of euros with performance in the order of a **few hundred MIPS (often equipped with Linux OS)**
- **Smart Phones** and **Tablets**, on the other hand, are systems with much higher computing power of the order of **hundreds of GFLOPS** for the CPU alone and costs of the order of hundreds of euros

Game Consoles, PCs and Workstations

- **Game Consoles** are systems with overall performance that can reach about **2000 GFLOPS** also considering the GPUs
- **PC considering Desktops and Laptops** cover a wide range of possibilities starting from a few hundred euros up to a few thousand with performances ranging from **dozens of GFLOPS to 2000 GFLOPS considering the GPUs**
- There are **Servers and Workstations** that are designed for High Performance Computing and centralization of services that can reach performances of a **few tens of TFLOPS with costs up to 10 / 20,000 euros**.

HPC

- In order to increase the performance of the computing resource in general, it is necessary to **couple together numerous computers interconnected with high-performance networks (parallel computing)**
- For example **Clusters of workstations** that can reach a **few hundred TFLOPS**
- **Supercomputer** with computing powers that today reach up to a **few dozen PFLOPS**
- Obviously, production costs and maintenance costs increase in the same way (even only in terms of absorbed power)

MODERN CPU

CISC vs RISC

The speed of execution of a single instruction is one of the determining factors of the performance of the CPU itself. Two different visions:

- **CISC (Complex Instruction Set**) in this case the basic idea is that the **basic instruction set of a CPU must be as rich as possible**, even if each single instruction actually requires more clock cycles to be executed
- **RISC (Reduced Instruction Set)** in this case each instruction is executed in a single clock cycle. Obviously, multiple RISC instructions will be needed to execute the same single CISC instruction
- CISC predominant architecture in the market in the 70 and 80s today there is a trend in favor of the RISC type CPU

Performances Improvements

Increasing the performance of a CPU is always a compromise between costs and consumption, for example, you can adopt some basic strategies:

- **Increase the frequency (clock) with obvious physical limits (e.g. the speed of light)**
- **Reduce the number of cycles required to execute a single instruction** (trivial example of multiplication)
- **Parallelism**, then perform, for example, several operations in parallel (simultaneously)
  - **At the instruction level, more instructions are executed in parallel by the same CPU with for example the use of pipelines or superscalar processors**
  - **Parallelism at the core level, so more cores per CPU**

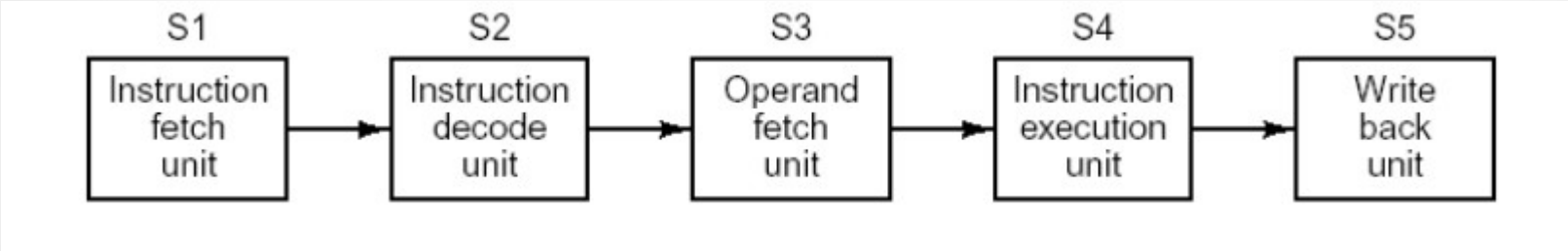
Improve Performance - Increase clock frequency

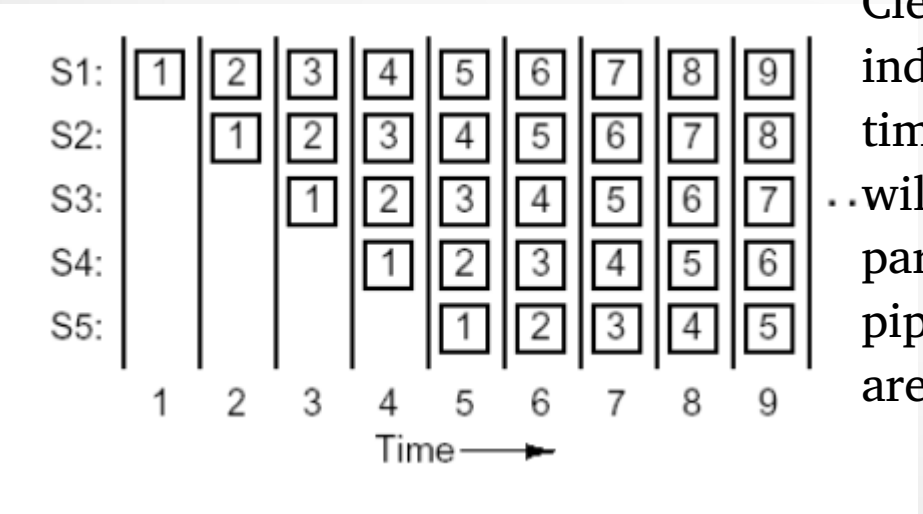
Until 2000, the increase in performance of a CPU largely coincided mainly with the increase in clock frequency. We have reached about 4 Ghz We have reached the physical limits **(1 GHz and therefore in one ns the distance that the electrical impulse can travel, imagining that the speed of light travels in a vacuum, is approximately 33 cm)**:

- High frequencies create disturbances and increase the heat to be dissipated
- Delays in signal propagation,
- Bus skew signals traveling on different lines travel at different speeds

Improve Perfomances – Pipeline

Each single instruction is divided into several stages (or phases) and each phase is managed by a dedicated piece of CPU (hardware):

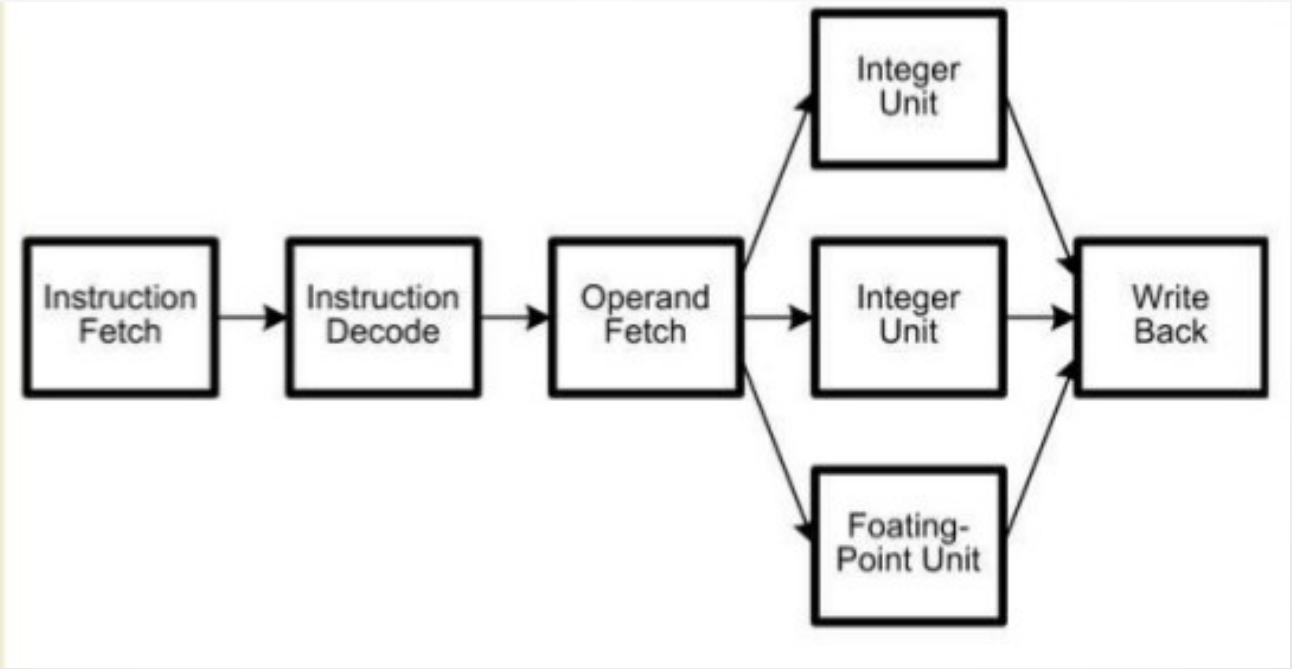




Clearly the operations must be independent. In this case, after an initial time (latency) to load the pipeline there will be n operations performed in parallel. You can also have multiple pipelines, and then multiple statements are read at a time and executed

Improve Performances – Superscalar

Different instructions treat their operands simultaneously on different hardware units, in practice there are different functional units of the same type, **for example there are more ALUs**



Improve Performaces – Branch predictor

The use of pipelines works particularly well in the case of sequential instructions but one of the basic programming constructs are jump instructions, for example **IF ... THEN ... ELSE** decisions:

```
If (a == b)
    printf (“numbers are the same \n”);
else
    printf (“Different \n”);
```

Improve Performaces – Branch predictor

Modern CPUs can (pre-fetch) try to guess if the program will skip:

- **Static prediction**: criteria are used that make "common sense" assumptions, for example we assume that all jumps are performed
- **Dynamics**: in practice, the CPU maintains a table that is based on an execution statistic

Improve Performances – Out-of-order execution

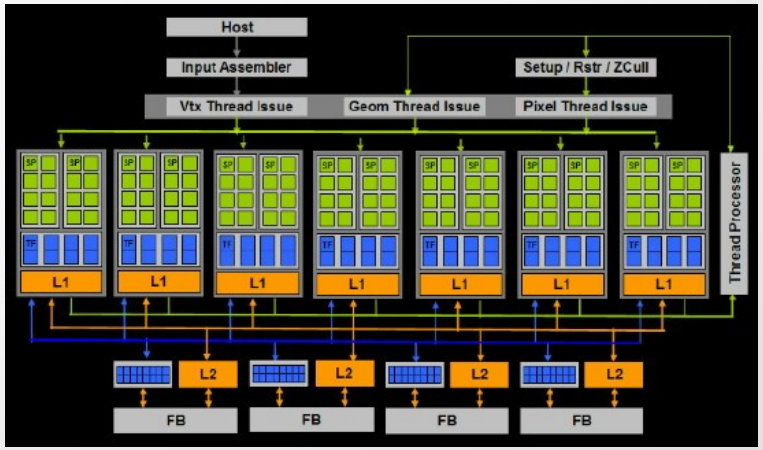
CPU design is much simpler if all the instructions are executed in order one after the other, but as we have seen we cannot make this assumption upstream. For example, execute a given instruction requesting that the result of the previous instruction be known.

- **Out of order execution**: Modern CPUs can temporarily skip some instructions to increase performance and put them on hold to follow others that do not introduce dependencies
- **Speculative Execution**: Execute particularly cumbersome (heavy) parts of your code before you are sure you really need them

C.P.U. - Some final considerations

Modern CPU:

- **Multi-core**: in the same chip there are several independent processors, each with their respective cache memory, for example, interconnected between them. **This allows to increase the "theoretical" performance without increasing the frequency (including Hyper-threading)**
- **GPU (graphics processors)** today increasingly used in accelerating computing.



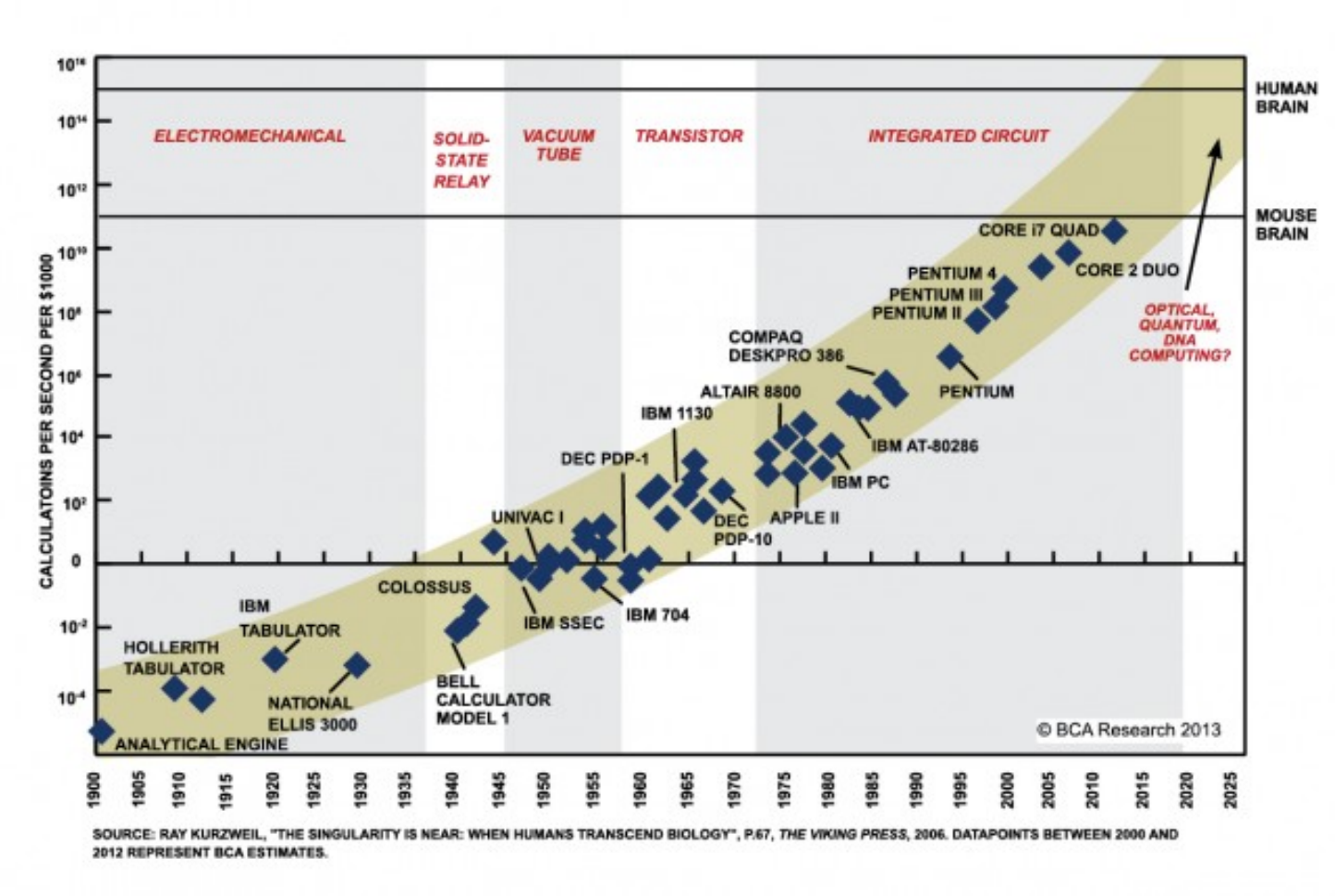
C.P.U. - Some final considerations

CPU embedded systems (also Raspberry Pi):

- **ARM** indicates a class of RISC type processors (32 bit) **(Advanced RISC Machine)** developed by an English company that does not produce them directly but holds the licenses. These are then produced for example by **STMicroelectronics, Samsung, Broadcom, Qualcomm etc etc. (System-On-Chip SOC)**
- These are **RISC CPUs** which, depending on the architectural design, guarantee a good compromise between performance and consumption (**ARM Cortex**)
  - **The Cortex A9 (1 GHz) consumes 250 mW per core. An Intel Corei-i7 CPU can consume over 100W**

TOP500 AD THE MOORE’S LAW

Moore’s law



In electronics and computer science, the following statement is indicated as Moore's first law: ”**The complexity of a microcircuit, measured for example by the number of transistors per chip, doubles every 18 months (and therefore quadruples every 3 years)**"

TOP500

- Difference between **sustained performance and peak performance**
- To objectively evaluate the performance of a computer you need a reference test, a standard benchmark, for example Linpack
- TOP500 http://www.top500.org/, ranking of the 500 most powerful computers in the world

TOP500

Prefix	Abbreviation	Order of magnitude (as a factor of 10)	Computer performance	Storage capacity
giga-	G	10^9	gigaFLOPS (GFLOPS)	gigabyte (GB)
tera-	T	10^{12}	teraFLOPS (TFLOPS)	terabyte (TB)
peta-	P	10^{15}	petaFLOPS (PFLOPS)	petabyte (PB)
exa-	E	10^{18}	exaFLOPS (EFLOPS)	exabyte (EB)
zetta-	Z	10^{21}	zettaFLOPS (ZFLOPS)	zettabyte (ZB)
yotta-	Y	10^{24}	yottaFLOPS (YFLOPS)	yottabyte (YB)

Top500 list June 2020

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,299,072	415,530.0	513,854.7	28,335
2	**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	2,414,592	148,600.0	200,794.9	10,096
3	**Sierra** - IBM Power System AC922, 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
4	**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
5	**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
6	**HPC5** - PowerEdge C4140, Xeon Gold 6252 24C 2.1GHz, NVIDIA Tesla V100, Mellanox HDR Infiniband, Dell EMC Eni S.p.A. Italy	669,760	35,450.0	51,720.8	2,252
7	**Selene** - DGX A100 SuperPOD, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband, Nvidia NVIDIA Corporation United States	272,800	27,580.0	34,568.6	1,344

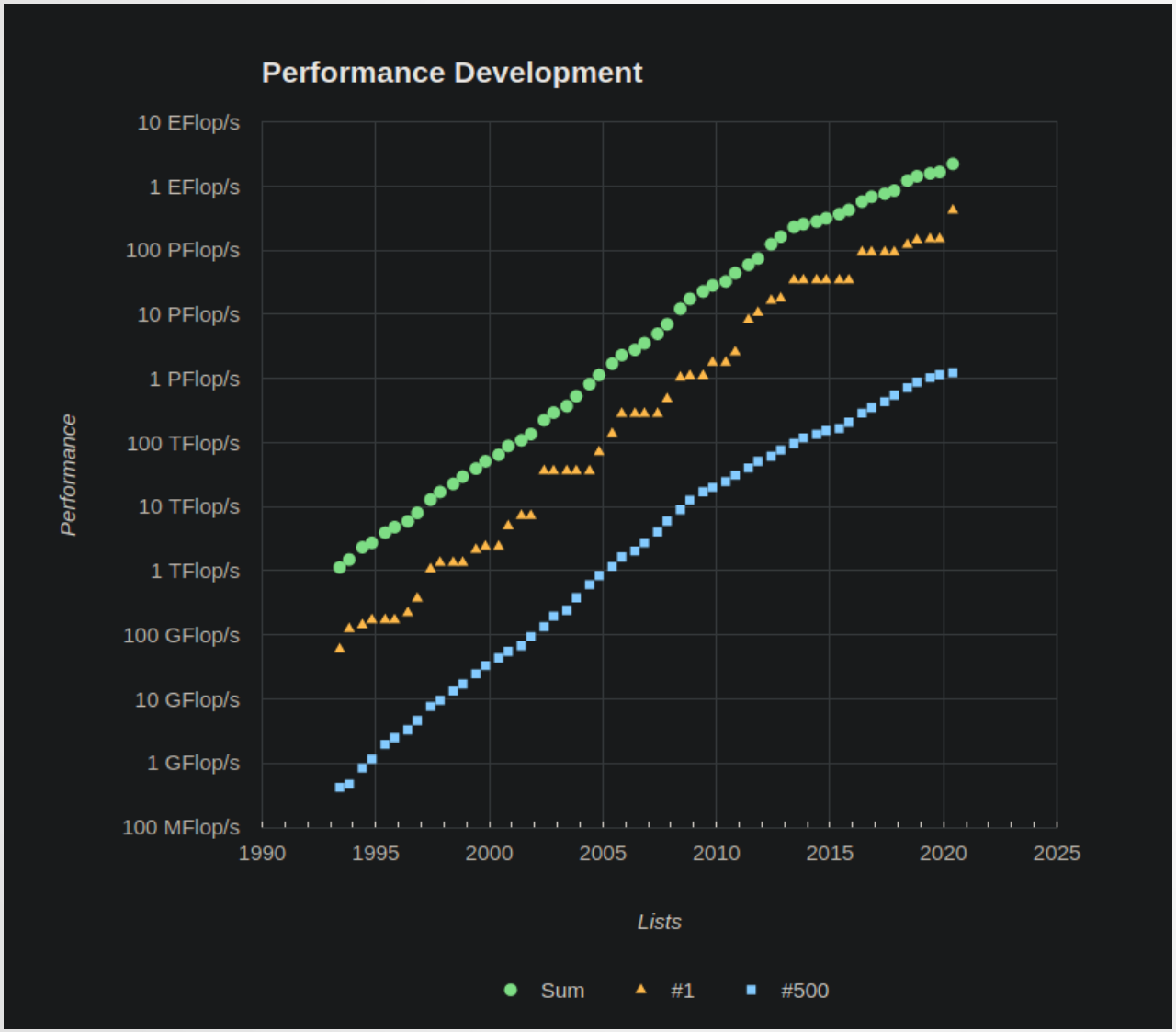
Top500 list June 2020

Site:	RIKEN Center for Computational Science
System URL:	https://www.r-ccs.riken.jp/en/fugaku/project
Manufacturer:	Fujitsu
Cores:	7,299,072
Memory:	4,866,048 GB
Processor:	A64FX 48C 2.2GHz
Interconnect:	Tofu interconnect D
Performance	
Linpack Performance (Rmax)	415,530 TFlop/s
Theoretical Peak (Rpeak)	513,855 TFlop/s

Top500 list June 2017

The Sunway TaihuLight uses a total of 40,960 Chinese-designed SW26010 manycore 64-bit RISC processors based on the Sunway architecture.[5] Each processor chip contains 256 processing cores, and an additional four auxiliary cores for system management (also RISC cores, just more fully featured) for a total of 10,649,600 CPU cores across the entire system.[5]

Top500 list historical trend



An estimate of the **computing power of the human mind a few dozen petaflops or more (maybe 1 exaflops)**