

Introduzione all'Informatica

Loriano Storchi

loriano@storchi.org

<http://www.storchi.org/>



UNITA' DI MISURA

Unita' di misura dell'informazione

- **BIT** = è l'unità di misura dell'informazione (dall'inglese "binary digit"), definita come **la quantità minima di informazione che serve a discernere tra due possibili eventi equiprobabili.** (Wikipedia)
- **BYTE** = 8 BIT (storicamente i caratteri erano rappresentati da 8 BIT , motivo per cui **1 Byte rimane tutt'oggi l'unita' di memoria minima indirizzabile**)
- “KiloByte” meglio “kibibyte” KiB = 2^{10} Byte = 1024 Byte
- “MegaByte” meglio “mebibyte” MiB = $1024 * 1024$ Byte
- “GigaByte” meglio “gibibyte” GiB = $1024 * 1024 * 1024$ Byte
- “TeraByte” meglio “tebibyte” TiB = $1024 * 1024 * 1024 * 1024$ Byte

Prestazioni dei calcolatori

- Chiaramente aumentare le prestazioni di un calcolatore significa diminuire il tempo che esso impiega nell'eseguire un'operazione.
- T_{clock} e' il periodo di clock della macchina (**aumento della frequenza**)
- CPI_i e' invece il numero di "colpi" di clock necessari ad eseguire la data istruzione i (**riduzione della complessità per la singola istruzione**)
- N_i e' infine il numero di istruzioni di tipo i (ad esempio somme, salti ...)

$$T_{\text{esecuzione}} = T_{\text{clock}} \sum_{i=0}^n N_i \text{CPI}_i$$

MIPS

MIPS e' un'abbreviazione per **Mega Instructions Per Second**, ed indica appunto il numero di istruzioni generiche che una CPU esegue in un secondo. E' un'unita' di misura usato per misurare le prestazioni di un calcolatore di uso più generale rispetto al FLOPS che vedremo a breve:

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

Questo tipo di misura non tiene conto ad esempio delle ottimizzazioni dovute alla presenza della cache e delle percentuali delle diverse istruzioni all'interno di programmi reali, e non solo.

FLOPS

FLOPS e' un'abbreviazione per **Floating Point Operations Per Second**, ed indica appunto il numero di operazioni in virgola mobile che una CPU esegue in un secondo. E' un'unita' di misura usato per misurare le prestazioni di un calcolatore diffusa in modo particolare nell'ambito del calcolo scientifico

Ad esempio nel caso di un prodotto classico tra matrici, vengono eseguite $2*N^3$ operazioni, quindi posso valutare i FLOPS esattamente misurando il tempo necessario ad eseguire tale moltiplicazione ed ottenere:

$$[\text{flops}] = 2*N^3 / \text{tempo}$$

SPEC

Standard Performance Evaluation Corporation e' un'organizzazione non-profit che produce e mantiene un insieme standardizzato di benchmark per computer (quindi un insieme di programmi di test che sono rappresentativi delle applicazioni reali di un calcolatore).

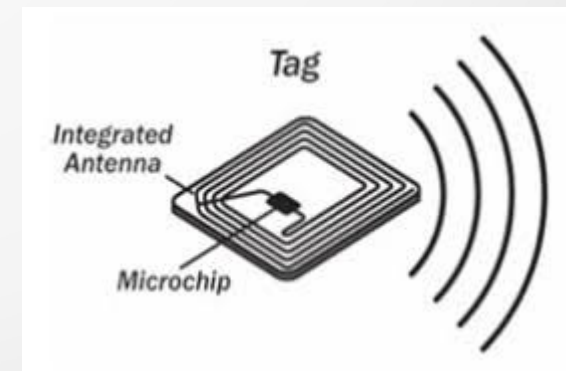
Ci sono diversi set di SPEC che sono specifici ad esempio per usi diversi previsti del computer



CALCOLATORI MODERNI

RFID

- **RFID Radio-frequency identification** costano pochi centesimi e possono raggiungere **qualche MIPS**
- Usano campi elettromagnetici per identificare e tracciare automaticamente oggetti mediante un ID o comunque informazioni
- **Passivi** se usano energia emessa dal lettore di RFID. Anche con batteria
- **Attivi** necessariamente con batteria inviano periodicamente un segnale



Sistemi embedded e SmartPhone Tablet

- **Sistemi embedded** oggi molto diffusi sono calcolatori pensati per Orologi, automobili, Elettrodomestici , apparati Medicali, lettori audio/video (Android Box). Sono calcolatori del costo di qualche decina di euro con prestazioni dell'ordine di **qualche centinaio di MIPS** (spesso equipaggiati con **OS Linux**)
- **SmartPhone e Tablet** sono invece sistemi con potenze di calcolo decisamente superiori dell'ordine di **centinaia di GFLOPS per le solo CPU** e costi dell'ordine di centinaia di euro

Console di Gioco, PC e Workstation

- **Console di Gioco** sono sistemi con prestazioni complessive che possono arrivare ai **2000 GFLOPS circa considerando anche le GPU**
- **PC considerando Desktop e Portatili** coprono una fascia ampia di possibilità partendo dalle poche centinaia di euro fino a qualche migliaia con prestazioni che quindi **vanno dalle decine di GFLOPS ai 2000 GFLOPS considerando le GPU**
- Ci sono **Server e Workstation** che sono pensati per **High Performance Computing** e centralizzazione di servizi che possono arrivare a prestazioni di **qualche decina di TFLOPS con costi fino a 10/20,000 euro**

HPC

- Per poter aumentare le prestazioni della risorsa di calcolo in generale e' necessario accoppiare assieme numerosi calcolatori **interconnessi con reti ad alte prestazioni (calcolo parallelo)**
- Ad esempio **Cluster di workstation che possono arrivare a qualche centinaio di TFLOPS**
- **Supercomputer** con potenze di calcolo che oggi arrivano fino a **qualche decina di PFLOPS**
- Ovviamente aumentano allo stesso modo costi di produzione e costi di mantenimento (anche solo in termini di potenza assorbita)



CPU MODERNE

CISC vs RISC

La velocità di esecuzione di una singola istruzione e' uno dei fattori determinanti delle prestazioni della CPU stessa. Due visioni diverse:

- **CISC (Complex Instruction Set)** in questo caso l'idea di base e' che il set di istruzioni di base di una CPU debba essere la più ricca possibile, anche se **ogni singola istruzione in realtà richiede più cicli di clock per essere eseguita**
- **RISC (Reduced Instruction Set)** in questo caso **ogni istruzione e' eseguita in un solo ciclo di clock**. Ovviamente saranno necessarie più istruzioni RISC per eseguire la stessa singola istruzione di un CISC
- CISC architettura predominante nel mercato negli anni 70 ed 80 oggi c'è una tendenza ad favore della CPU di tipo RISC

Migliorare le Prestazioni

Aumentare le prestazioni di una CPU e' sempre un compromesso fra costi e consumi, ad esempio si possono adottare alcune strategie di base:

- **Ridurre il numero di cicli** necessarie all'esecuzione di una singola istruzione (esempio banale della moltiplicazione)
- **Aumentare la frequenza** (clock) con ovvi limiti fisici (ad esempio al velocità della luce)
- **Parallelismo, quindi eseguire ad esempio piu' operazioni in parallelo** (in contemporanea)
 - **A livello di istruzione piu' istruzioni** sono eseguite in parallelo dalla stessa CPU con ad esempio l'uso delle **pipeline o processori superscalari**
 - **Parallelismo a livello di core, quindi più core per CPU**

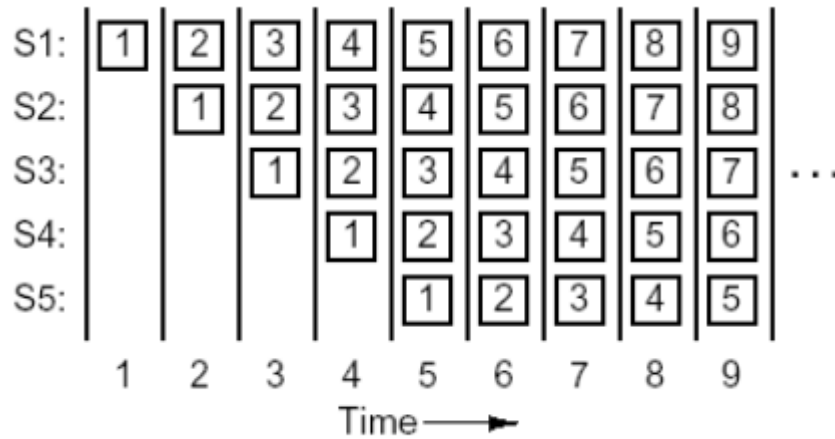
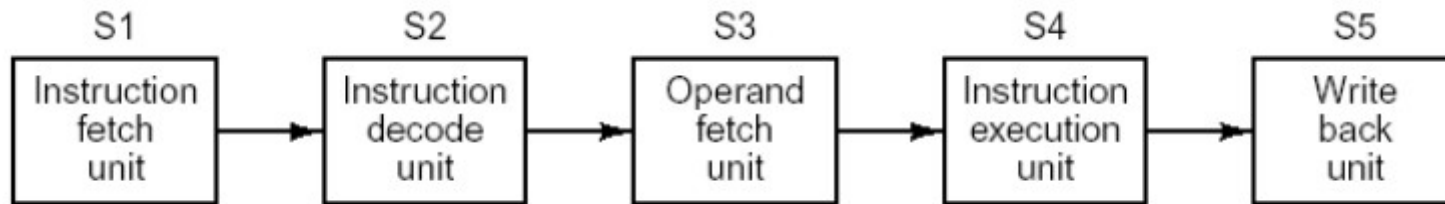
Migliorare le Prestazioni – Aumento frequenza di clock

Fino al 2000 l'aumento delle prestazioni di una CPU coincideva per buona parte soprattutto con l'aumento della frequenza di clock. **Siamo arrivati a circa 4 Ghz** Abbiamo raggiunto i limiti fisici (**1 GHz e quindi in un ns la distanza che puo' percorrere l'impulso elettrico, immaginando che viaggia alla velocità delle luce nel vuoto, è di 33 cm circa**):

- Le alte frequenze creano **disturbi** ed **umentano il calore da dissipare**
- Ritardi nella propagazione del segnale,
- **Bus skew** i segnali che viaggiano su linee diverse viaggiano con velocità diverse

Migliorare le Prestazioni – Pipeline

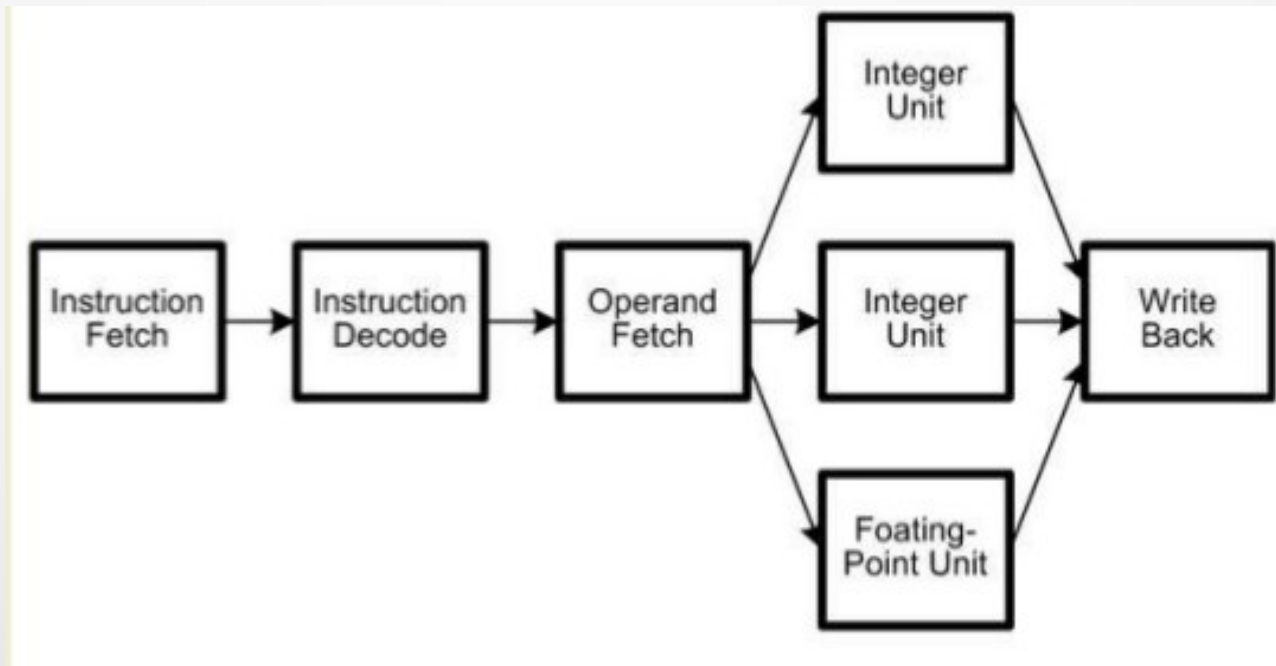
Ogni singola istruzione viene divisa in più stadi (o fasi) ed ogni fase è gestita da un pezzo di CPU (hardware) dedicato:



Chiaramente **le operazioni devono essere indipendenti**. In questo caso dopo un tempo (**latenza**) iniziale a caricare la pipeline ci saranno n operazioni eseguite in parallelo. Si possono avere anche più pipeline e quindi vengono lette più istruzioni alla volta ed eseguite

Migliorare le Prestazioni – Superscalari

Istruzioni differenti trattano i propri operandi contemporaneamente su unita' hardware differenti , **in pratica sono presenti diverse unita' funzionali dello stesso tipo, ad esempio sono presenti piu' ALU**



Migliorare le Prestazioni – Predizione di salto

L'uso delle pipeline funziona particolarmente bene in caso di istruzioni sequenziali ma uno dei costrutti basilari della programmazione sono le istruzioni di salto, ad esempio decisioni **IF...THEN...ELSE**:

```
If (a == b)
    printf ("i due numeri sono uguali");
else
    printf ("Sono diversi \n");
```

Migliorare le Prestazioni – Predizione di salto

Le CPU moderne possono (pre-fetching) provare ad indovinare se il programma salterà’:

- **Predizione statica:** si usano dei criteri che fanno delle assunzioni di “buon senso”, ad esempio assumiamo che tutti i salti vengano eseguiti
- **Dinamica:** in pratica la CPU mantiene una tabella che e’ basata su una **statistica** di esecuzione

Migliorare le Prestazioni – Esecuzioni fuori ordine e speculativa

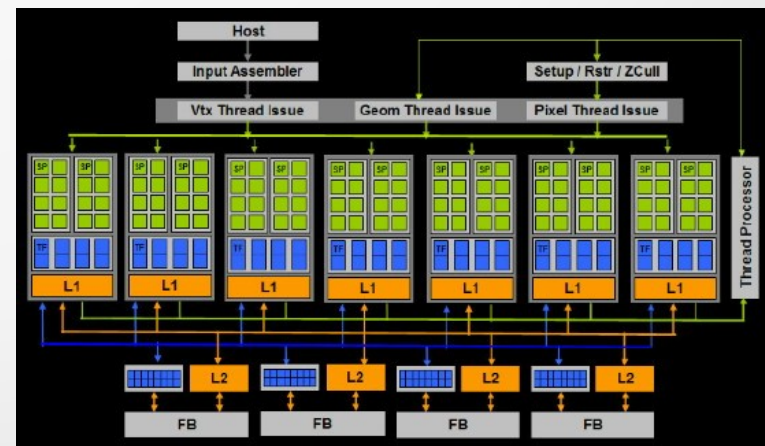
La progettazione delle CPU e' decisamente piu' semplice se tutte le istruzioni sono eseguite in ordine una dopo l'altra, ma come abbiamo visto non possiamo fare questa assunzione a monte. Ci sono delle dipendenze. Ad esempio eseguire una data istruzione richiedere che sia noto il risultato dell'istruzione precedente.

- **Esecuzione fuori ordine:** Le CPU moderne possono per aumentare le prestazioni **saltare temporaneamente alcune istruzioni mettendole in attesa** per seguirne altre che non introducono dipendenze
- **Esecuzione speculativa:** eseguire parti del codice particolarmente gravose (pesanti) prima di essere certi che servano davvero

C.P.U. - Considerazioni finali

CPU moderne:

- **Multi-core** : in pratica nello stesso chip ci sono piu' processori indipendenti, **ognuno con le rispettive memoria cache ad esempio , interconnessi fra di loro.** Questo permette di aumentare le prestazioni "teoriche" senza aumentare la frequenza (anche **Hyper-threading**)
- **GPU** (processori grafici) oggi sempre piu' utilizzati nell'accelerazione del calcolo.



C.P.U. - Considerazioni finali

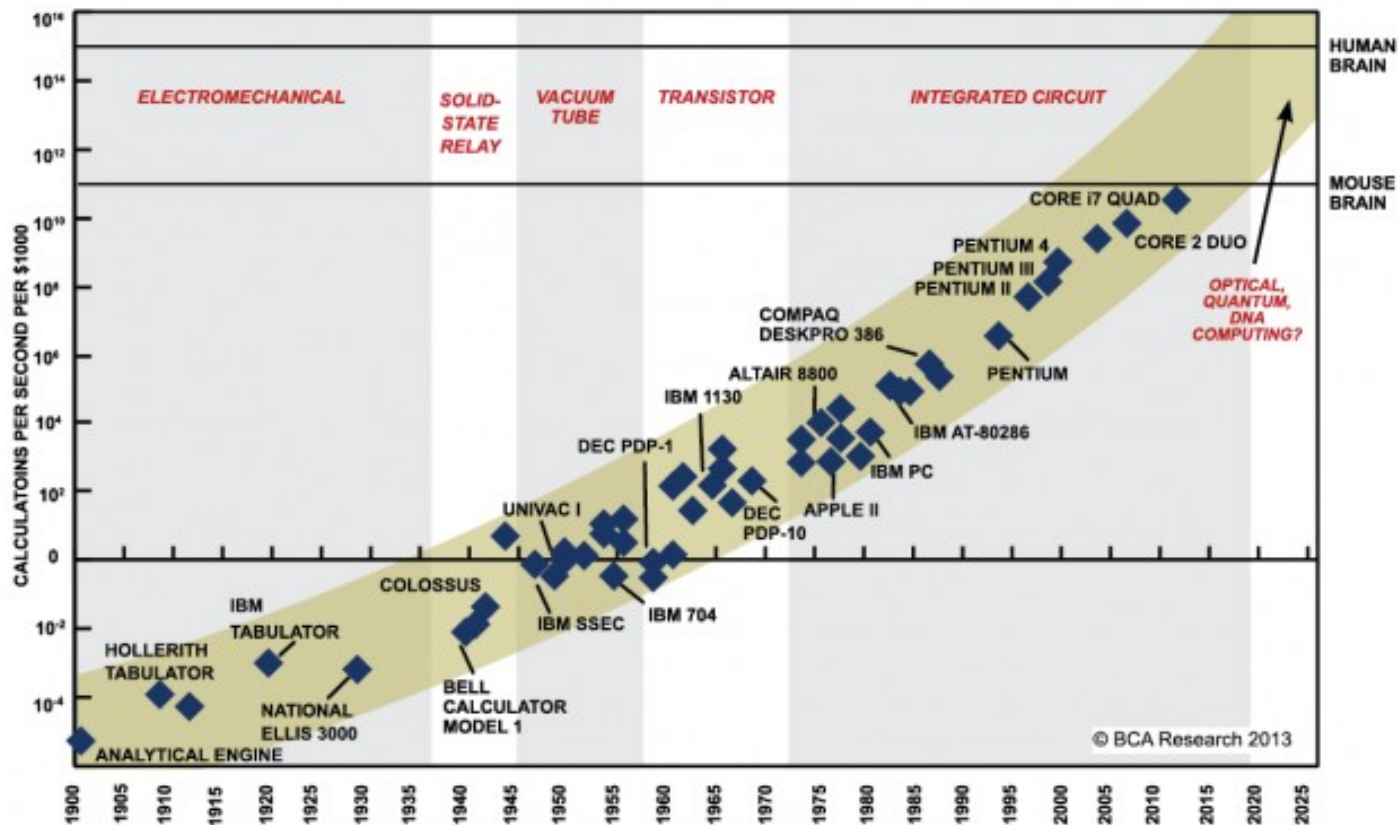
CPU sistemi embedded (anche Raspberry Pi):

- **ARM** sta ad indicare una classe di processori di tipo RISC (32 bit) (**Advanced RISC Machine**) sviluppati da un'azienda inglese che non li produce direttamente ma detiene le licenze. Questi vengono poi prodotti ad esempio da STMicroelectronics, Samsung, Broadcom, Qualcomm etc etc. (System-On-Chip SOC)
- Sono appunto **CPU RISC** che in funzione del disegno architetturale garantiscono un buon compromesso fra prestazioni e consumo (**ARM e Cortex**)
 - **Il Cortex A9 (1 GHz) consuma 250 mW per core. Una CPU Intel Corei-i7 puo' arrivare a consumare oltre 100 W**



TOP500 E LEGGE DI MOORE

Legge di Moore



SOURCE: RAY KURZWEIL, "THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY", P.67, THE VIKING PRESS, 2006. DATAPOINTS BETWEEN 2000 AND 2012 REPRESENT BCA ESTIMATES.

In elettronica e informatica è indicato come prima legge di Moore il seguente enunciato: « La complessità di un microcircuito, misurata ad esempio tramite il numero di transistori per chip, raddoppia ogni 18 mesi (e quadruplica quindi ogni 3 anni). »

TOP500

- Differenza tra **sustained perfomace** e **prestazioni di picco**
- Per valutare in modo oggettivo le prestazioni di un computer c'e' bisogno di un test di riferimento, un **benachmark standard, ad esempio Linpack**
- TOP500 <http://www.top500.org/>, classifica dei 500 computer piu' potenti al mondo

Top500 list Giugno 2018

Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	DOE/SC/Oak Ridge National Laboratory United States	Summit - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband IBM	2,282,544	122,300.0	187,659.3	8,806
2	National Supercomputing Center in Wuxi China	Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway NRPC	10,649,600	93,014.6	125,435.9	15,371
3	DOE/NNSA/LLNL United States	Sierra - IBM Power System S922LC, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband IBM / NVIDIA / Mellanox	1,572,480	71,610.0	119,193.6	
4	National Super Computer Center in Guangzhou China	Tianhe-2A - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000 NUDT	4,981,760	61,444.5	100,678.7	18,482
5	National Institute of Advanced Industrial Science and Technology (AIST) Japan	AI Bridging Cloud Infrastructure (ABCI) - PRIMERGY CX2550 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR Fujitsu	391,680	19,880.0	32,576.6	1,649

Top500 list Giugno 2018



Department of Energy, Oak Ridge National Laboratory

IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR InfiniBand

2.28 million cores

122.3 petaflops

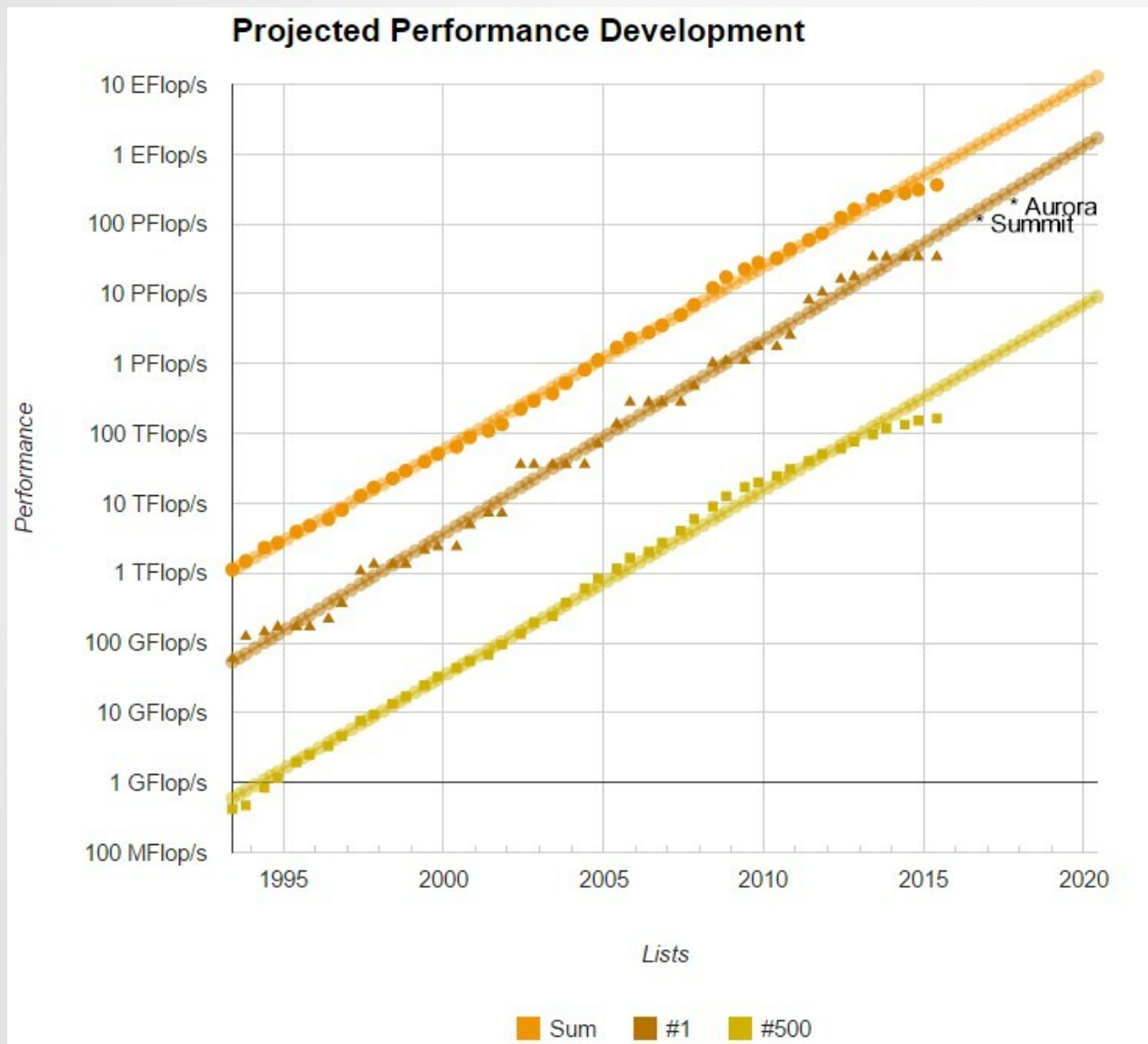
Department of Energy, Oak Ridge National Laboratory

Top500 list Giugno 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 grafico storico



Una stima della **potenza di calcolo della mente umana** qualche decina di **petaflops o oltre**