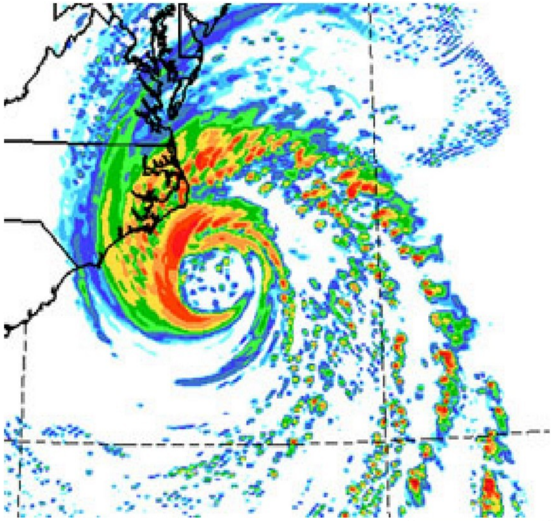
High Performance Systems
Introduction

Elisabeth Brunet

CSC5001 - Septembre 2023

Scientific Computing and simulation

- **Essential for scientific and industrial innovation**
- **Numerous fields of application**
  - Meteorology, astrophysics, nanoscience, etc.
  - Automotive, aeronautics, 7th art, defense, etc.
- **Simulation is necessary when the problems are ...**
  - ...too complex
  - ...to massive
  - …too expensive
  - …too dangerous
  - ...predictive



Scientific Computing and simulation

- **A lot of calculation**
- **Handling of huge amounts of data**
- **Time constraints**

- **Consequences**
  - Increased computing resources
  - Parallelization of problems

→ To go faster and faster

→ but above all to deal with ever bigger problems

High Performance Platforms

- **Component capacities maximized**

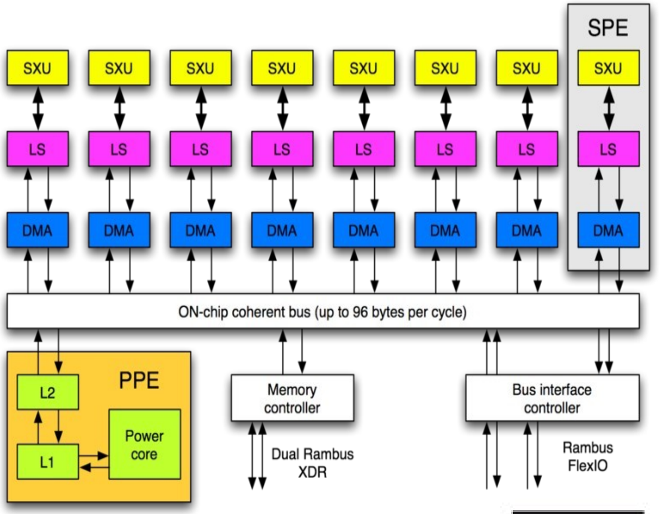
- **Processor diversity**
  - Boosted consumer architectures
  - Specialized Architectures
    - Processors Alpha, MIPS, Cray, Power, Sparc, NEC, etc.
      - Instruction sets
      - Architectural processing flow
        - Vectorial processor
        - Processor Cell

Processor Cell

- **Processor design by IBM, Sony and Toshiba in 2005**

- **Totally different architectural model**
  - 1 *master* PowerPC processor called PPE
  - 8 SPE vectorial co-processors
  - Internal EIB interconnection bus

- **Peak performances**
  - 230,4 GFLOPS in single précision
  - Initially designed for multimedia (PS3) and hijacked by HPC
- **Extremely difficult to program**
  - Abandoned in 2009



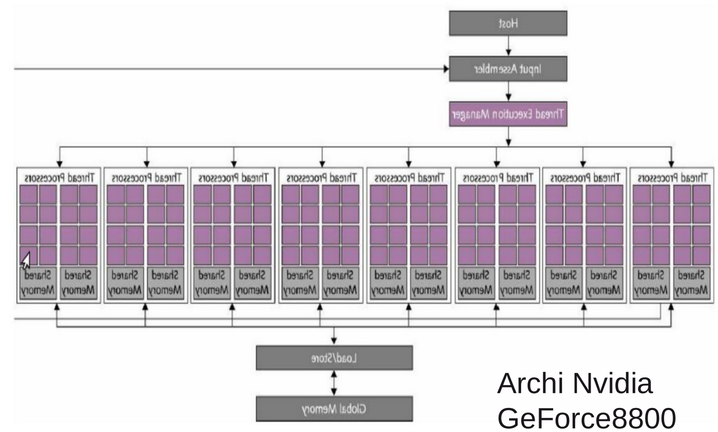
High Performance Platforms

- **Component capacities maximized**

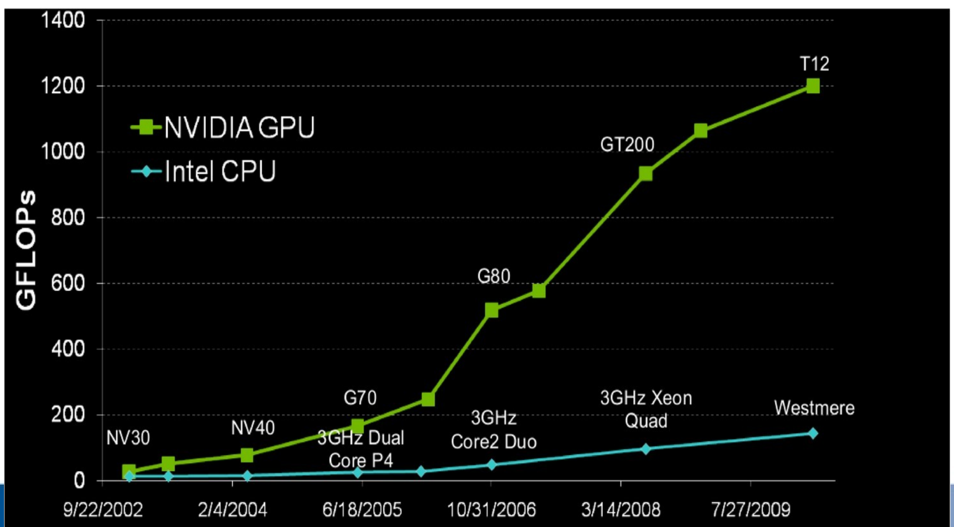
- **Processor diversity**
  - Boosted consumer architectures
  - Specialized Architectures
  - Co-processors
    - FPGA : programmable logic circuit
    - GPU : Graphics Processing Unit

GPU Computing

- **1 GPU = hundreds of limited cores**
  - No dynamic memory allocation
  - No stack, so no recursivity
- **Design for 3D image synthesis**
  - 3D API : OpenGL, DirectX
- **Then parallel computing focus**
  - Nvidia : Architecture Tesla(2006), Fermi(2009) / API CUDA
  - AMD : Archi RadeonHD / API ATI Stream SDK
  - API unifiée GPU+CPU : OpenCL (2008)



Archi Nvidia GeForce8800



- **Vectorial computation offloading**
  - Suitable to massively parallel compute
    - Data transfert
    - Kernel computation
    - Result transfert
  - → Slow CPU/GPU communication

High Performance Platforms

- **Component capacities maximized**

- **Processor diversity**
  - Boosted consumer architectures
  - Specialized Architectures
  - Co-processors
  - Massive aggregation of resources
    - High performance networks

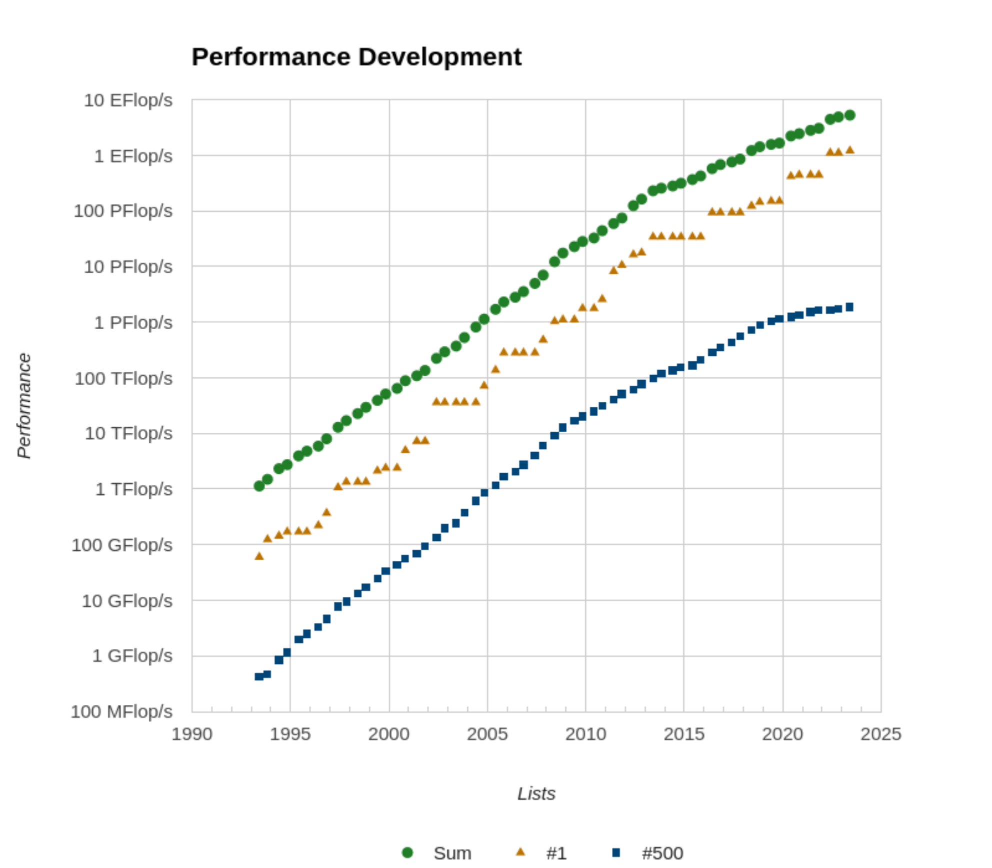
      InfiniBand, Myrinet, 10G-Ethernet, etc.
    - Topologies : supercomputer, cluster, grid

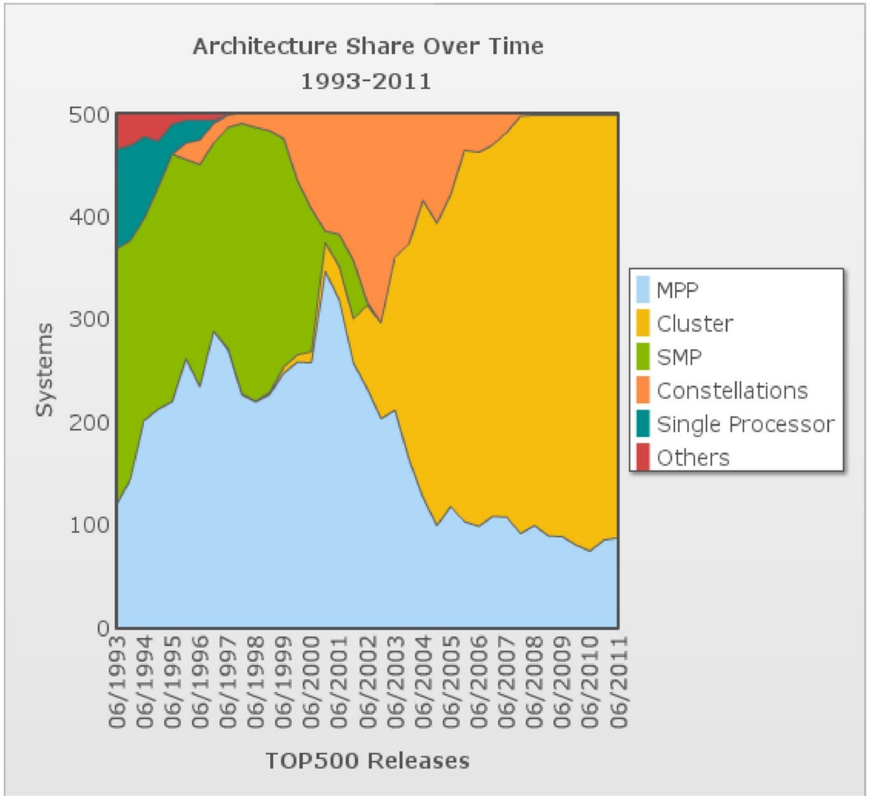
Computing resources in HPC

- **Hybrid architectures**
- **High capabiliy components**
- **Ever increasing computing power**
  - Petascale (10^15 floating point operations per second), even exascale
  - Top500 ranking

Top500 - www.top500.org

- **Bi-annual ranking of the 500 most powerful machines in the world**





Top500 in june 2023

Rmax et Rpeak en Pflop/s
Power en kW

1 **Frontier - HPE Cray EX235a, AMD EPYC 64C 2GHz, AMD Instinct MI250X – Oak Ridge (USA)**

#cores = 8,699,904 Rmax = 1,194.00 Rpeak=1,679.82 Power=22,703

2 **Fugaku –** A64FX48C 2.2GHz, Tofu Interconnect – Fujitsu RIKEN Center of Computational Science (Japan)

#cores = 7,630,848 Rmax = 442.01 Rpeak=537.21 Power=29,899 → #1 in 2020

3 **Lumi – HPE Cray EX235a, AMD EPYC 64C 2GHz, AMD Instinct MI250X – EuroHPC CSC (Finland)**

#cores = 2,220,288 Rmax = 309.10 Rpeak=428.70 Power=6,016

4 **Leonardo –BullSequana XH2000, Xeon Platinum 2.6GHz, NVIDIA A100 64GB, Infiniband**– EuroHPC CINECA (Italy)

#cores = 1,824,768 Rmax = 238.70 Rpeak=304.47 Power=7,404

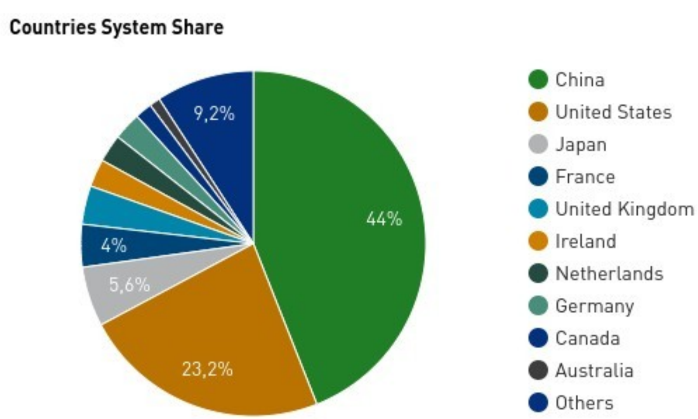
5 **Summit** – IBM POWER9, NVIDIA Volta GV100 – Oak Ridge National Lab (USA) → **#2 in 2020**

cores=2,414,592 Rmax=148,600 RPeak=200,794 Power=10,096

French Top 3 in juin 2023

Rmax et Rpeak en Pflop/s
Power en kW

- **12 . Adastra - HPE Cray EX235a, AMD Optimized EPYC 64C 2GHz - GENCI-CINES**
  #cores=319,072 Rmax=46.10 Rpeak=61.61 Power=921
- **22. CEA – HF** – Bull Sequana X1000 – Xeon Phi 7250 – CEA
  #cores=810,240 Rmax=23.24 Rpeak=31.76 Power=4,959
- **39. PANGEA III** – IBM POWER9, NVIDIA Volta GV100 – Total Exploration Production
  #cores=291 024 Rmax=17860 Rpeak=25025 Power=1367



Green 500 in 2020

- **Best performance/energy consumption ratio**

Rank	TOP500 Rank	System	Cores	Rmax (PFlop/s)	Power (kW)	Energy Efficiency (GFlops/watts)
1	255	**Henri** - ThinkSystem SR670 V2, Intel Xeon Platinum 8362 32C 2.8GHz, NVIDIA H100 80GB PCIe, Infiniband HDR, Lenovo Flatiron Institute United States	8,288	2.88	44	65.396
2	34	**Frontier TDS** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	120,832	19.20	309	62.684
3	12	**Adastra** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE Grand Equipement National de Calcul Intensif - Centre Informatique National de l'Enseignement Suprieur (GENCI-CINES) France	319,072	46.10	921	58.021
4	17	**Setonix – GPU** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE Pawsey Supercomputing Centre, Kensington, Western Australia Australia	181,248	27.16	477	56.983
5	77	**Dardel GPU** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE	52,864	8.26	146	56.491

- ***Most of them are accelerator-based***
- ***Trend of aggregating many low-power processors tops the Green500***

Critical points of HPC applications

In terms of implementation

- **Classic problems exacerbated**
- **Use of material resources**
- **Data distribution**
- **Dissemination of résultats**
  - Collective operations (*alltoall*, *broadcast*, reduction, etc.)
- **Problems related to the size of the platforms**
  - Fault tolerance, etc.

In terms of efficiency

- **Data locality**
- **Data granularity**
- **Load balancing**

Support for HPC applications

- **Hardware abstraction interfaces**
  - Examples : OpenCL, PVM, MPI, etc.
  - For software portability

Support for HPC applications

- **Runtimes**
  - Multiprocessor architectures
    - Thread scheduling
    - Data placement
  - Distributed architecturess
    - Data distribution
    - Data transfert
    - MPI, RPC, DSM , etc
  - Heterogeneous architectures
    - Load balancing
    - StarSs, Intel Ct, StarPU, etc.

For performance portability !

Support for HPC applications

- **Libraries for classical problem solving**
  - FFT, linear algebra (BLAS, etc.), etc.

- **Tools**
  - Performance Analysis (Easytrace, Vampir, Scalasca, etc.)
  - Bugs detection (Valgrind, gdb, etc.)
  - Fault tolerance
  - ....

- **Middlewares**
  - Code couplage

Metrics

- **Performance dependent on several factors**
  - Fraction of the parallelizable application
  - Quality of scheduling on computing resourcesl
  - Additional cost introduced by the parallel version

- **Speedup :** measures the acceleration between parallel and sequential versions
  - Sp = Tseq /Tp , where

    Sp = speedup on P procs

    Tp = execution time on P processors

    Tseq = execution time on 1 processor
  - Aiming at Sp = P

- **Amdahl law :** acceleration bound as a function of parallelization quality
  - R = 1 /((1−S)+ S/P) , where

    S = parallelized code ratio, P = #processors
  - Speedup is bounded by 1/S − > addition of processors can be unnecessary