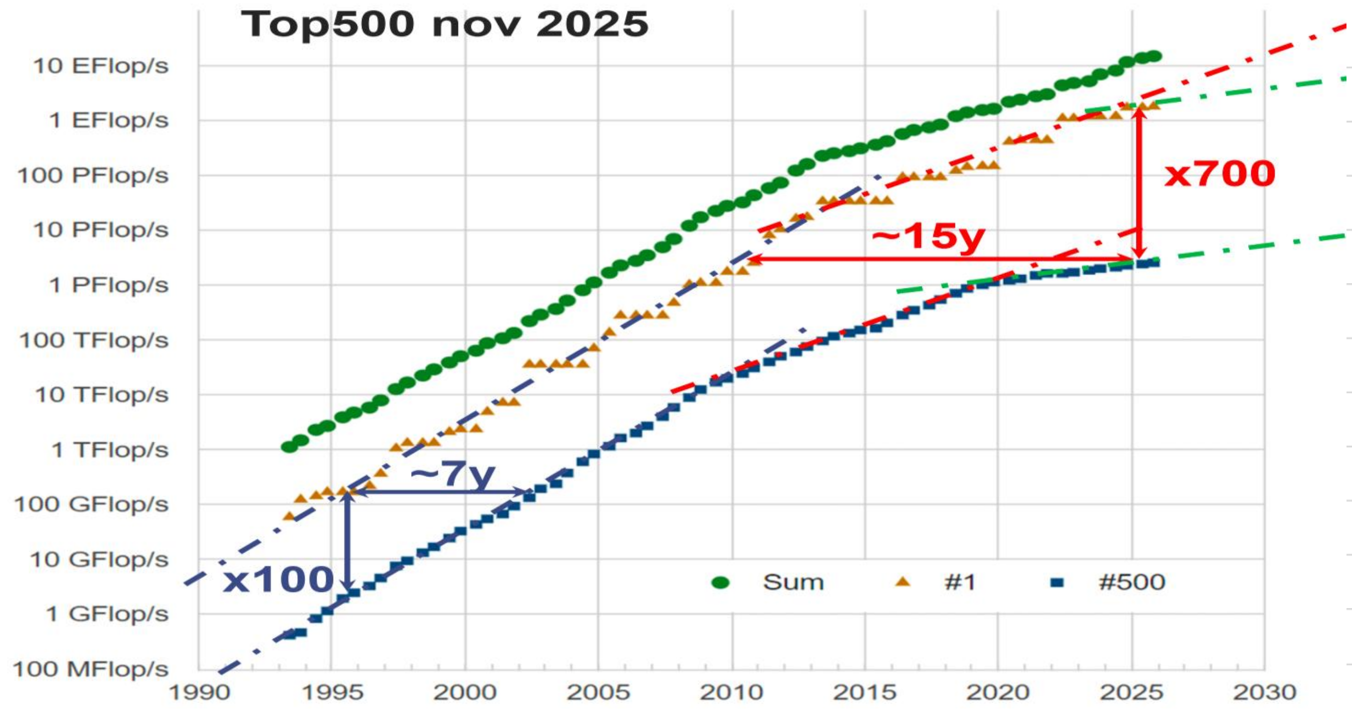
Energy efficiency towards Exascale class systems
Agenda

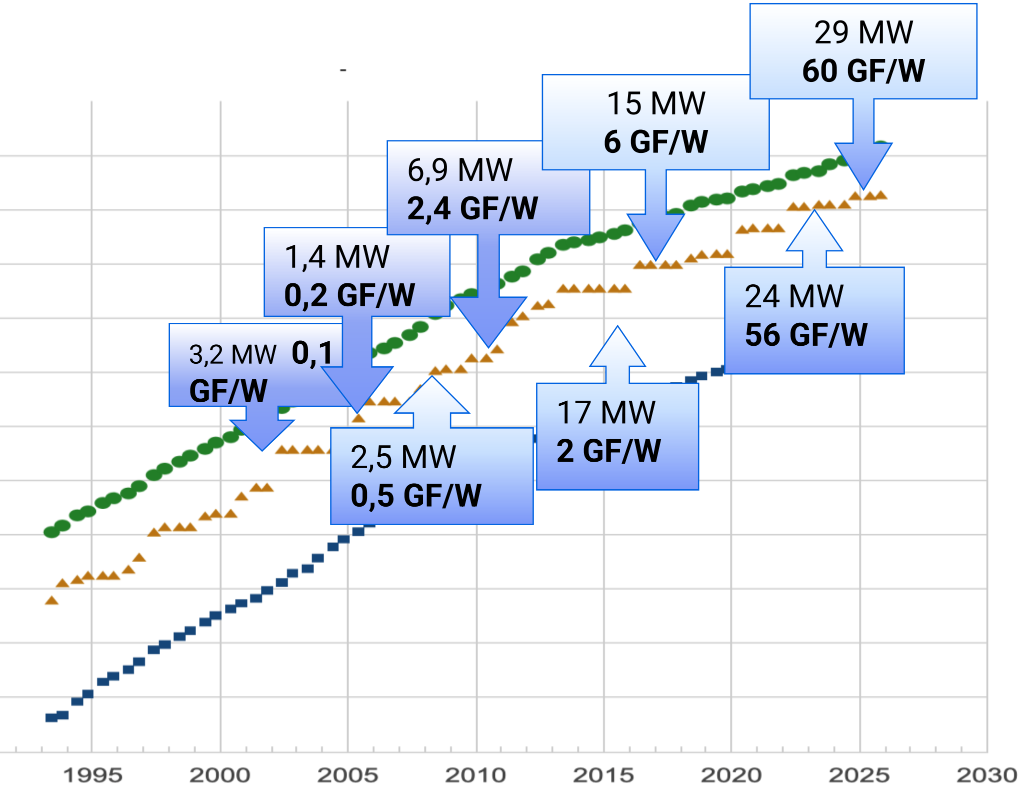
- Defining contraints for exascale class systems
- Efficiency and power trends
- Energy efficiency : a systemic challenge
- Conclusion

Energy efficiency towards Exascale class systems
Defining contraints for exascale class systems

- Energy efficiency is a key constraint for exascale-class systems :
  - Power limits, Operational costs, Environmental impact
- Energy cost and availability become limiting factors
- Sustainability is now part of HPC procurments

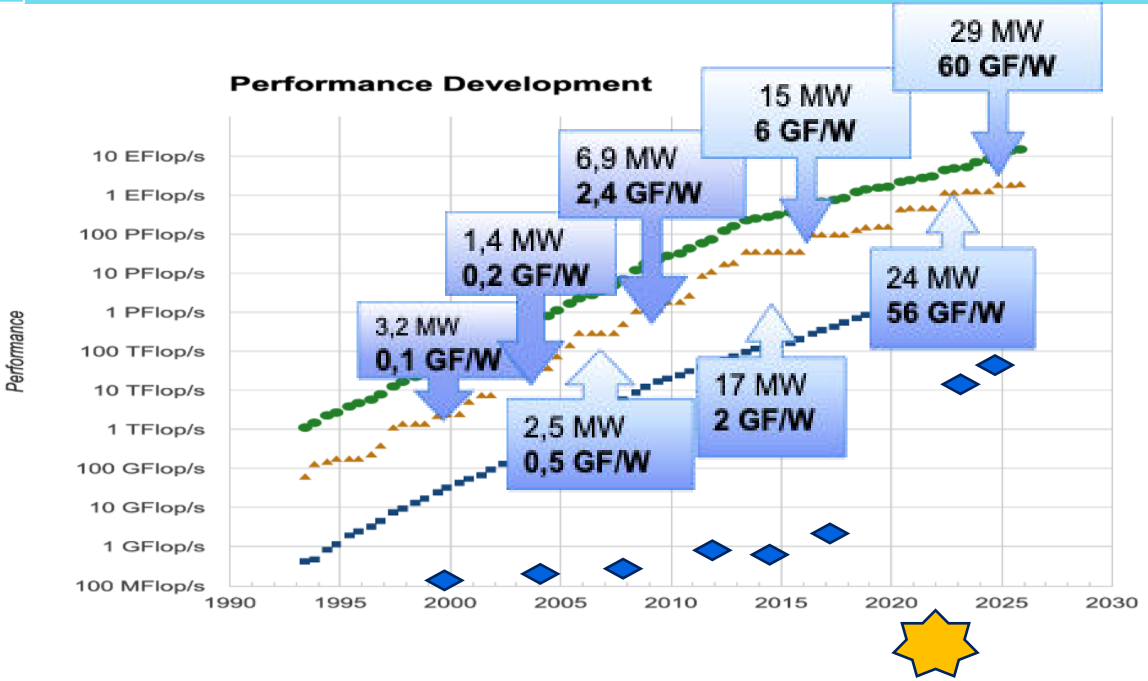
Energy efficiency towards Exascale class systems
Energy and power trend (through TOP500 …)

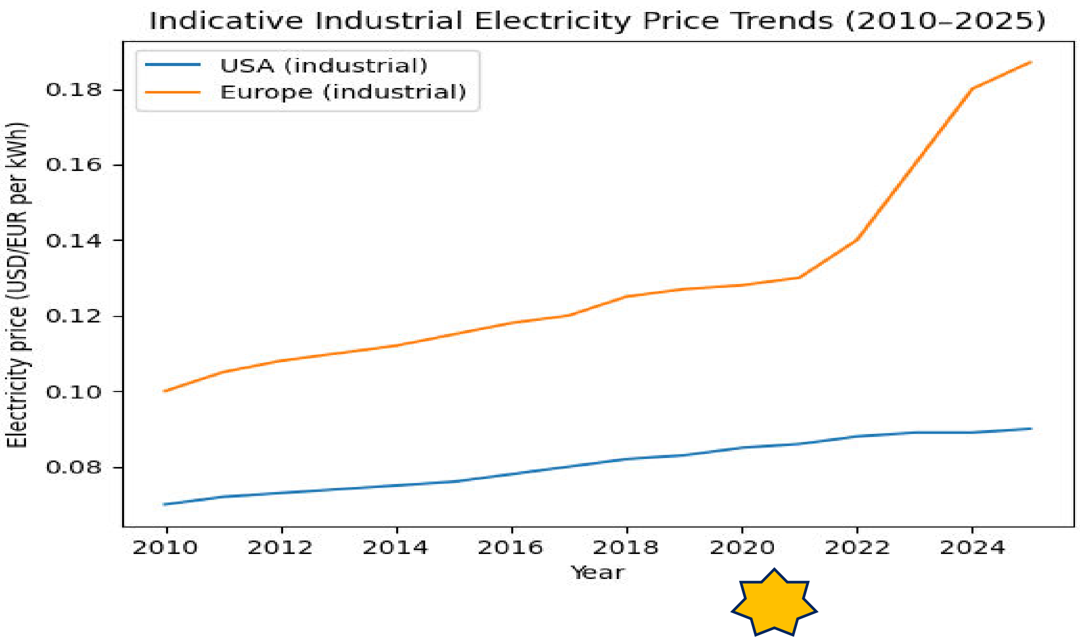




Energy efficiency towards Exascale class systems

Energy and power trend : OPEX
weight trigger of power efficiency?





- **For large-scale HPC infrastructures, electricity costs trend lead to an increasing OPEX/CAPEX in TCO**

  ➔ Need for considering system lifetime strategies, and energy-aware performance optimization

Mastering Energy : a systemic challenge

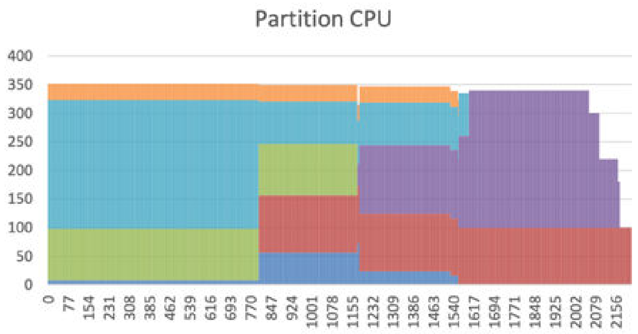
Life time production

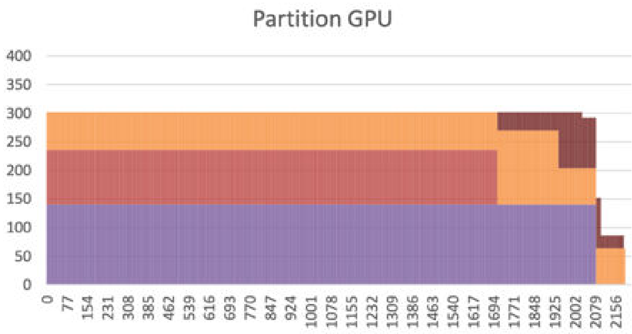
Applications	**Design** / Tuning / Autotuning @ runtime
Component steering	- Application ETS Pareto optimum - Global production
Analysis	- Recommandations static / dynamic - Efficiency hints - Application identification / phases
Data collection	- Global and permanent data collection
Components & architecture	**At procurement phase :** - Choice fitted for targeted workload
Infrastructure	**At data center preparation phase :** - Cooling / heat reuse / energy supply

Knobs :
- On/off fine/coarse
- Frequency
- Power capping

« TCO » METHOD

TCO Method : Some maths





Then :

$$E_T = E_C + E_D + E_U$$

$$E_T = \left[\left(E_{Ct} * \frac{c}{c_{wl}} * PUE_{compute}\right) + (E_{Dt} * PUE_{disk}) + (E_{Ut} * PUE_{cdu})\right] * \frac{T}{t}$$

$$TCO_E = E_T * EnergyCost$$

$$TCO = Purchase + Maintenance + Infrastructure + Staff + TCO_E$$

Where:

- *Purchase* is the acquisition cost of the machine
- *Maintenance* is the maintenance cost over 5 years
- *Infrastructure* is the specific cost of infrastructure for the machine
- *Staff* is the cost for complementary staff for machine exploitation over 5 years

Cost of the workload (TCO_b) takes into account the number of time (N) it twill be possible to run it over 5 years

$$N = \frac{c}{c_{wl}} * \frac{T}{t}$$

$$TCO_b = {}^{TCO}/_{N}$$

Variable definition :

t : wallclock time of the workload
ECt : Energy consumption by computational nodes
EDt : Energy consumption from disks and service nodes
EUt : Energy consumption from internal cooling devices
c : total core count of the proposed configuration
cwl : core count used for the selected workload
T : Number of second in 5 years
N : Number of runs of the workload during 5 years
ET : Total energy consumption during 5 years
EC : Energy of computational part during 5 years
ED : Energy of storage and service nodes during 5 years
EU : Energy of internal cooling device during 5 years
TCOE : TCO due to energy consumption during 5 ans
PUEair **:** Power Usage Effectiveness for air cooled devices
PUEcoldw **:** Power Usage Effectiveness for "Cold water" cooled devices
PUEhotw **:** Power Usage Effectiveness for "Hot water" cooled devices

Energy calculations :

$$E_C = E_{Ct} * \frac{c}{c_{wl}} * PUE_{compute} * \frac{T}{t}$$

$$E_D = E_{Dt} * PUE_{disk} * \frac{T}{t}$$

$$E_U = E_{Ut} * PUE_{cdu} * \frac{T}{t}$$

With

$$PUE_{compute} = 1 + [(PUE_{air} - 1) * \%_{air-for-compute}] + [(PUE_{coldw} - 1) * \%_{coldw-for-compute}] + [(PUE_{hotw} - 1) * \%_{hotw-for-compute}]$$

$$PUE_{disk} = 1 + [(PUE_{air} - 1) * \%_{air-for-disk}] + [(PUE_{coldw} - 1) * \%_{coldw-for-disk}] + [(PUE_{hotw} - 1) * \%_{hotw-for-disk}]$$

$$PUE_{cdu} = 1 + [(PUE_{air} - 1) * \%_{aircdu}] + [(PUE\_coldw - 1) * \%_{holdw-for-cdu}] + [(PUE_{hotw} - 1) * \%_{hotw-for-cdu}]$$

EUPEX

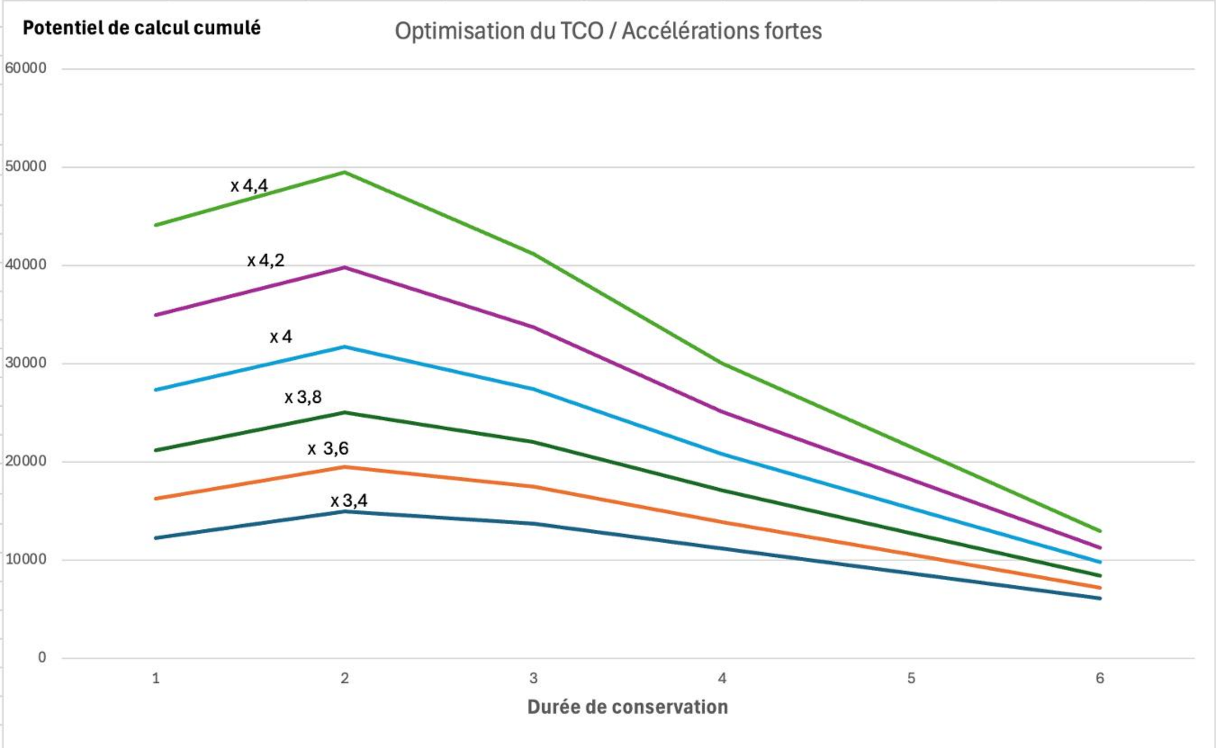
Replacement strategy : Some maths again …

- Problem to solve : for a given budget over several years what period of replacement is optimal?

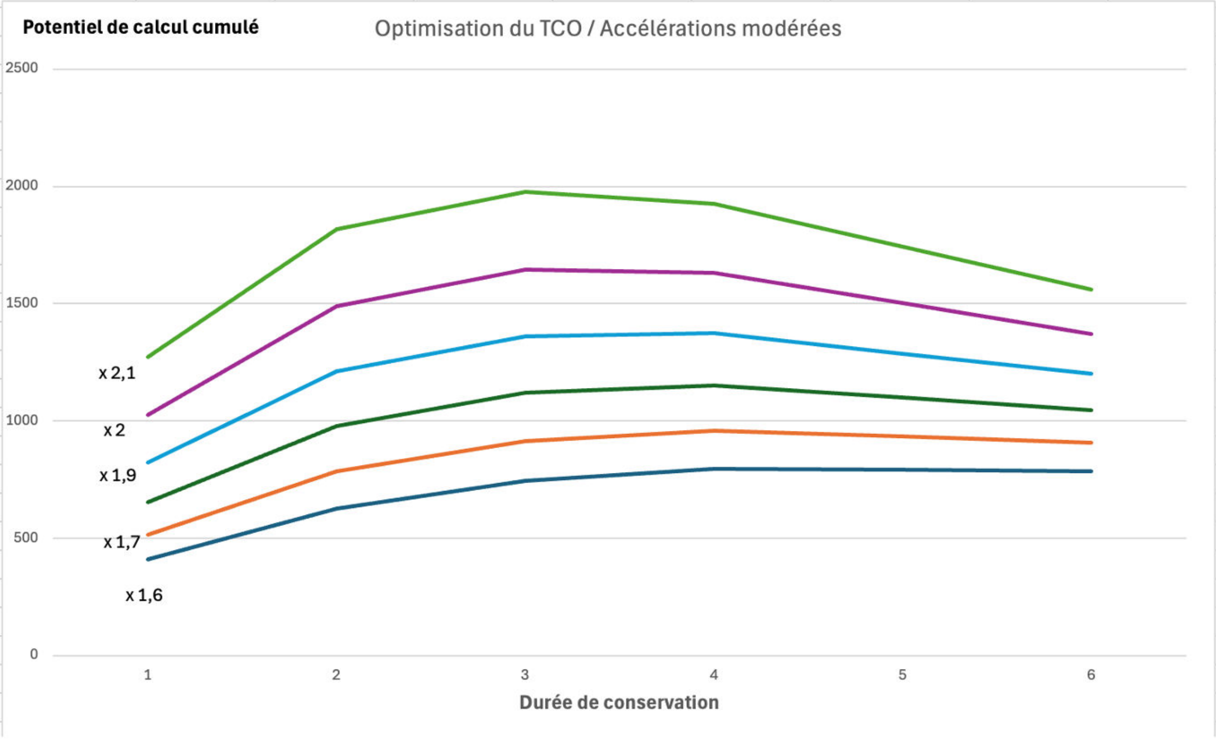
➔ Model variables : Energy cost, TDP increase per generation, maintenance % over CAPEX, renewal technologies rate, application speedup

➔ Application speedup is key - results for :

High range speed-up (3,4 to 4,4)



Average range speed-up (1,6 to 2,1)



Energy efficiency towards Exascale class systems
Conclusion

- Exascale is not longer how fast, but how efficient

- THANK YOU