RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

PROGRAMME DE RECHERCHE
NUMÉRIQUE POUR L'EXASCALE

Power capping on heterogeneous platforms

Exa-Soft WP6

Albert d'Aviau de Piolant

September 26, 2025

Inria, Université de Bordeaux, Bordeaux INP

Motivation to improve the energy efficiency of HPC applications

- HPC applications are optimized to run fast.
  1. **Hardware vendors:** multi-core processors, vector units, accelerators, etc.
  2. **Software optimizations:** parallelization at the heart of HPC applications.
- As computing power scales, so does power consumption.

Motivation to improve the energy efficiency of HPC applications

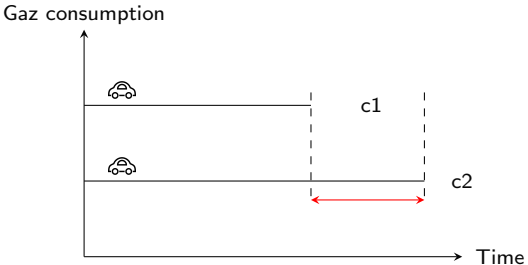
- HPC applications are optimized to run fast.
  1. **Hardware vendors:** multi-core processors, vector units, accelerators, etc.
  2. **Software optimizations:** parallelization at the heart of HPC applications.
- As computing power scales, so does power consumption.

Rank	System	Rmax PFlop/s	Power MW	Green500 Rank
1	El Capitan	1 742	29	18
2	Frontier	1 353	24	22
3	Aurora	1 012	38	64

- 38MW is the equivalent of 70000 french houses.

Metrics of energy aware computing

- $energy(Joule) = power(Watt) \times time(s)$



Energy efficiency

- How well the energy is used (flop/second/Watt)

Energy saving

- We control time execution or power consumption
- Multiple solutions to save power

Hardware

- Dynamic Voltage Frequency Scaling (DVFS)
- Power capping (static or dynamic)

Software

- Energy efficiency aware scheduling

What is Power capping?

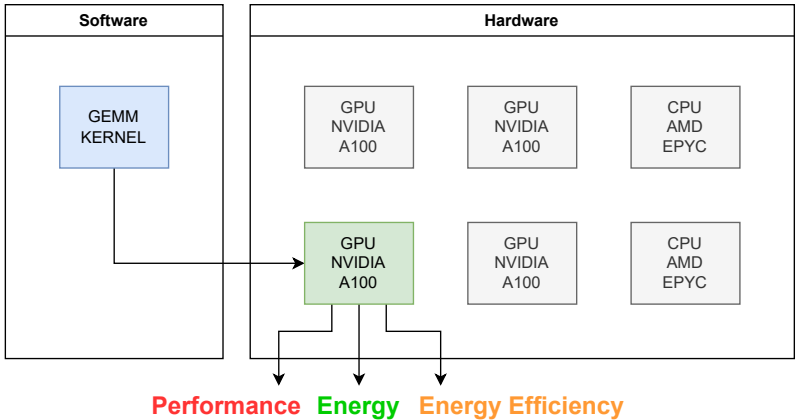
Definition

Power capping is a way of imposing a power consumption limit on a device.

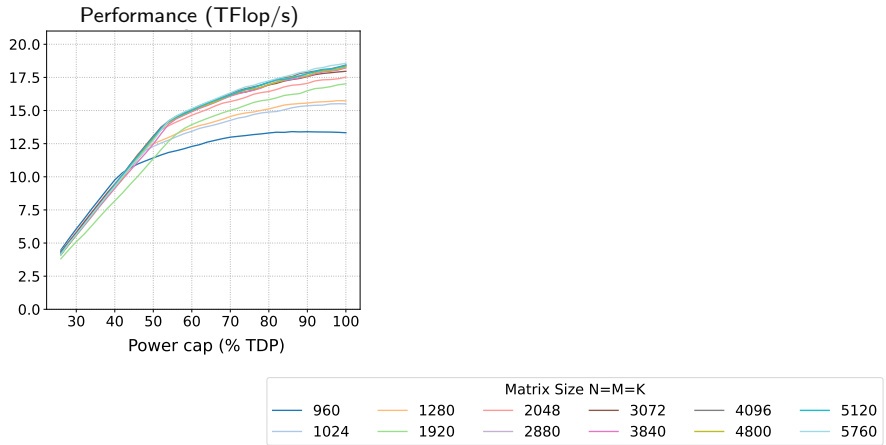
- intel-rapl, nvidia-smi
- static or dynamic
- **MUST BE ROOT!**

Study the energy efficiency under GPU power capping: Profiling

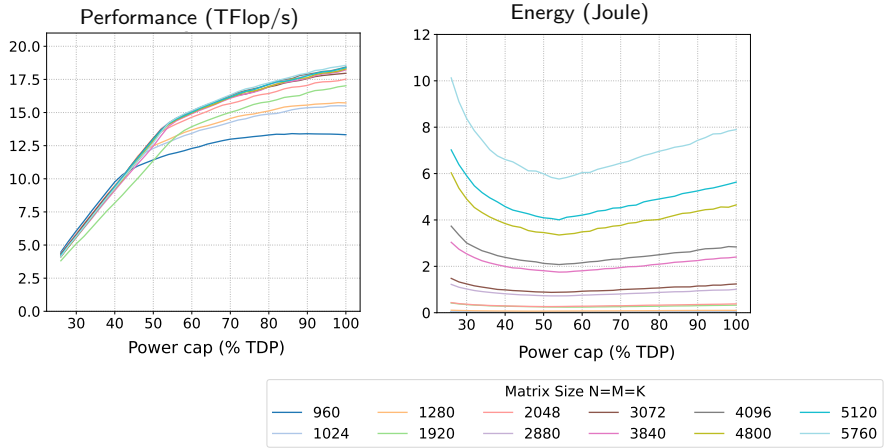
We want to study the impact of static power capping on a single GPU.



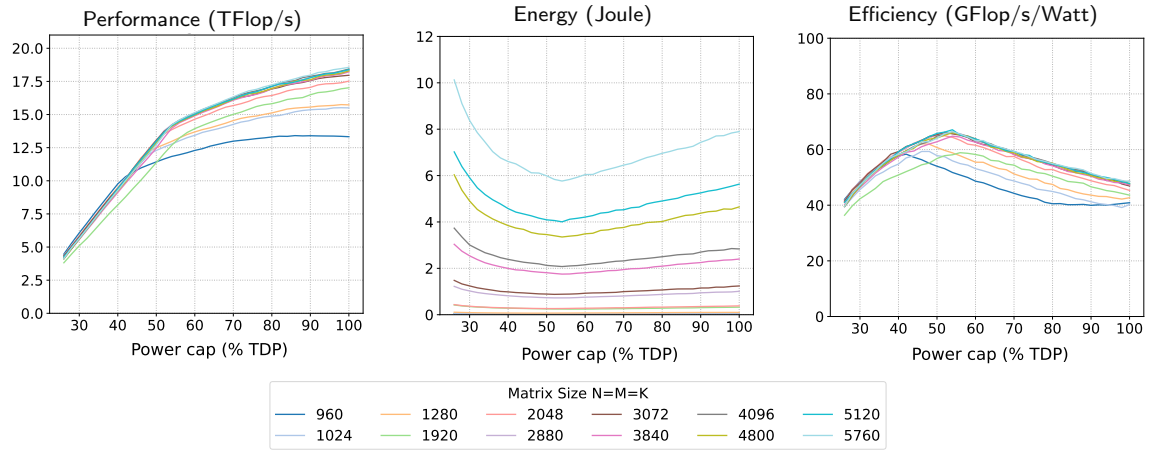
Study the energy efficiency under GPU power capping: Profiling



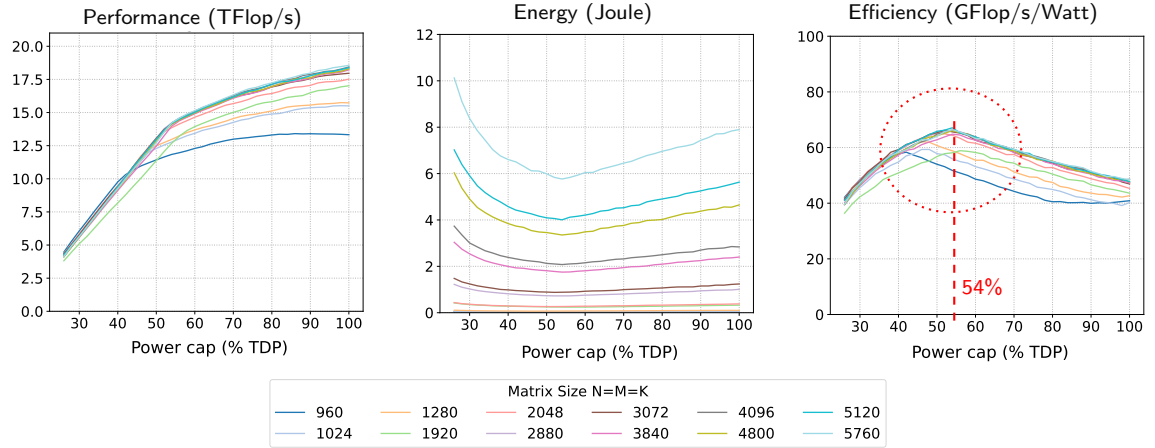
Study the energy efficiency under GPU power capping: Profiling



Study the energy efficiency under GPU power capping: Profiling



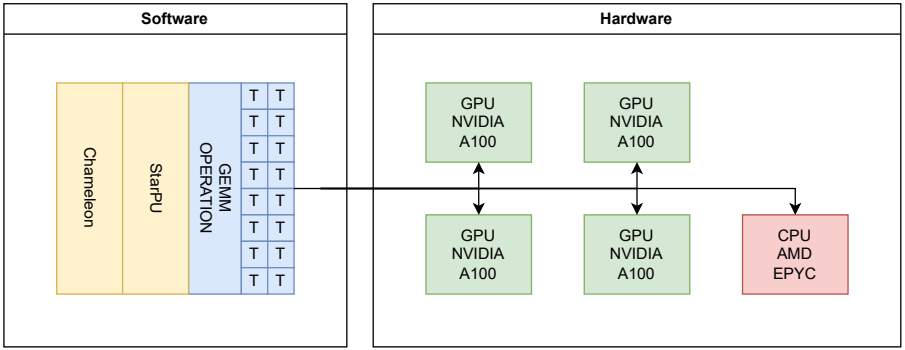
Study the energy efficiency under GPU power capping: Profiling



- **Faster is not equivalent to being energy efficient.**

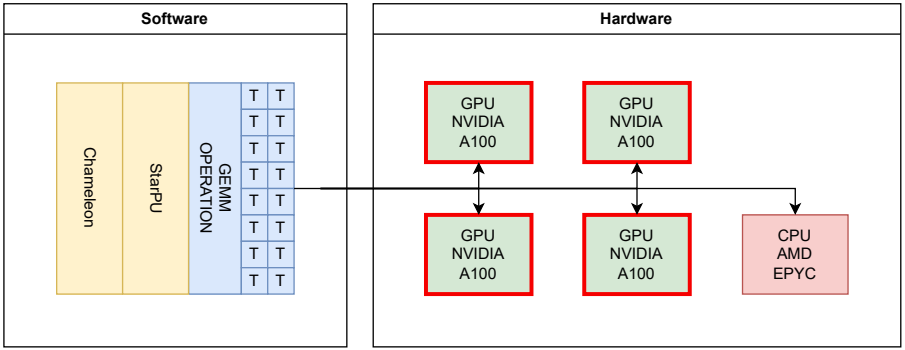
Study the energy efficiency under GPU power capping

We want to study the impact of power capping on a heterogeneous compute node.



Study the energy efficiency under GPU power capping

We want to study the impact of power capping on a heterogeneous compute node.



Study the energy efficiency under GPU power capping

Platform hardware

- 1x CPU AMD Zen3 EPYC 7513
- 4x GPU Nvidia A100-SXM4-40GB

Power settings for a GPU

	Config.	Power (W)	Percent.
P_{low}	L	100	25%
P_{best}	B	216	54%
P_{high}	H	400	100%

Default configuration is HHHH

Study the energy efficiency under GPU power capping

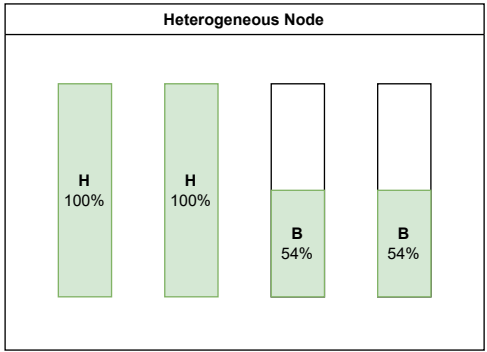
Platform hardware

- 1x CPU AMD Zen3 EPYC 7513
- 4x GPU Nvidia A100-SXM4-40GB

Power settings for a GPU

	Config.	Power (W)	Percent.
P_{low}	L	100	25%
P_{best}	B	216	54%
P_{high}	H	400	100%

Default configuration is HHHH



Study the energy efficiency under GPU power capping

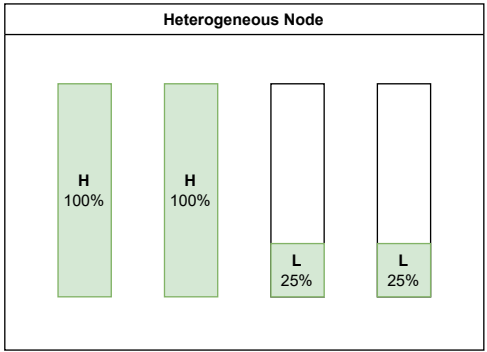
Platform hardware

- 1x CPU AMD Zen3 EPYC 7513
- 4x GPU Nvidia A100-SXM4-40GB

Power settings for a GPU

	Config.	Power (W)	Percent.
P_{low}	L	100	25%
P_{best}	B	216	54%
P_{high}	H	400	100%

Default configuration is HHHH



Study the energy efficiency under GPU power capping

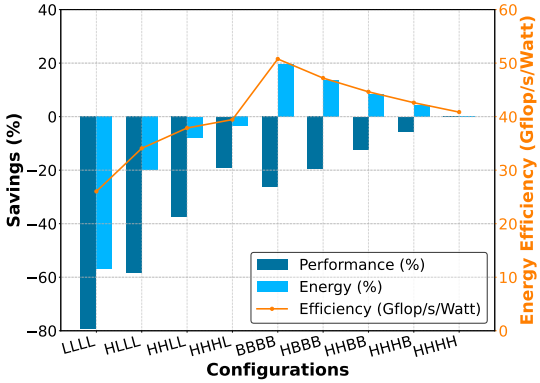
Platform hardware

- 1x CPU AMD Zen3 EPYC 7513
- 4x GPU Nvidia A100-SXM4-40GB

Power settings for a GPU

	Config.	Power (W)	Percent.
P_{low}	L	100	25%
P_{best}	B	216	54%
P_{high}	H	400	100%

Default configuration is HHHH



Study the energy efficiency under GPU power capping

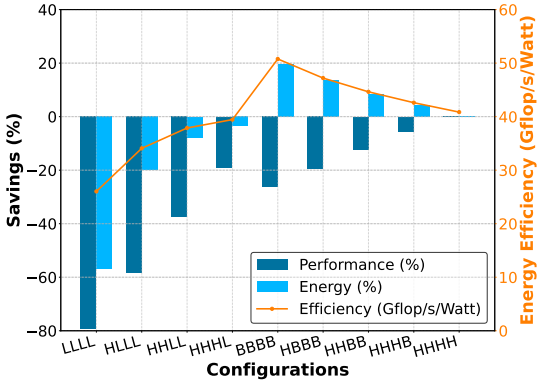
Platform hardware

- 1x CPU AMD Zen3 EPYC 7513
- 4x GPU Nvidia A100-SXM4-40GB

Power settings for a GPU

	Config.	Power (W)	Percent.
P_{low}	L	100	25%
P_{best}	B	216	54%
P_{high}	H	400	100%

Default configuration is HHHH



- Excessive slowdown → higher energy consumption

Study the energy efficiency under GPU power capping

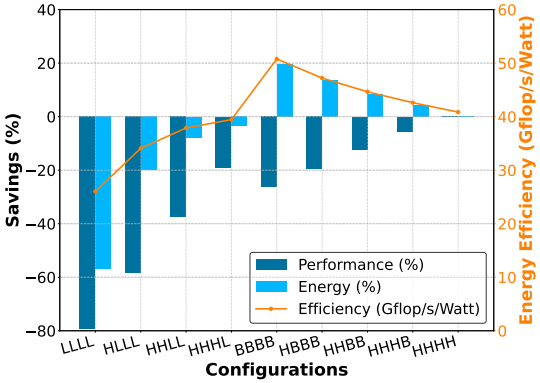
Platform hardware

- 1x CPU AMD Zen3 EPYC 7513
- 4x GPU Nvidia A100-SXM4-40GB

Power settings for a GPU

	Config.	Power (W)	Percent.
P_{low}	L	100	25%
P_{best}	B	216	54%
P_{high}	H	400	100%

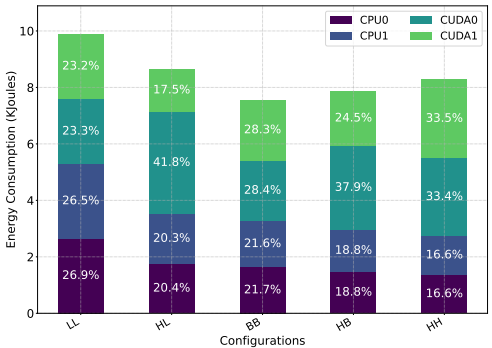
Default configuration is HHHH



- Excessive slowdown → higher energy consumption
- **20% energy efficiency improvements** when power capping 4 GPUs to 54%

Power cap on CPU

What is the CPU's energy consumption ?



Platform hardware

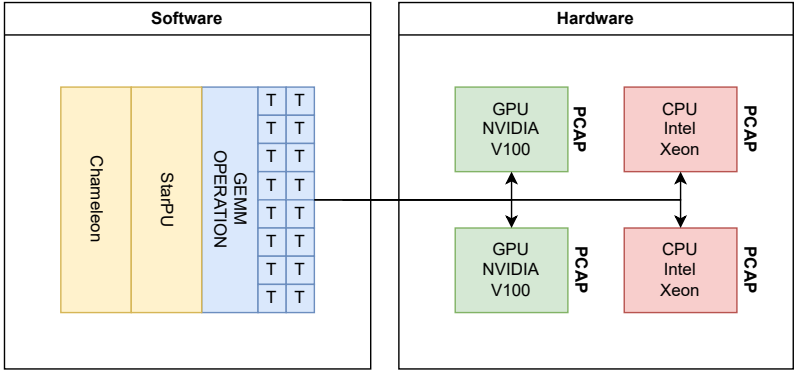
- 2x CPU Intel Xeon Gold 6126
- 2x GPU Nvidia V100-PCIE-16GB

We want to power cap CPUs:

- CPUs take 24 % of the total workload.
- They consume from 33 % to 53 % of the total energy consumption depending on the configuration.

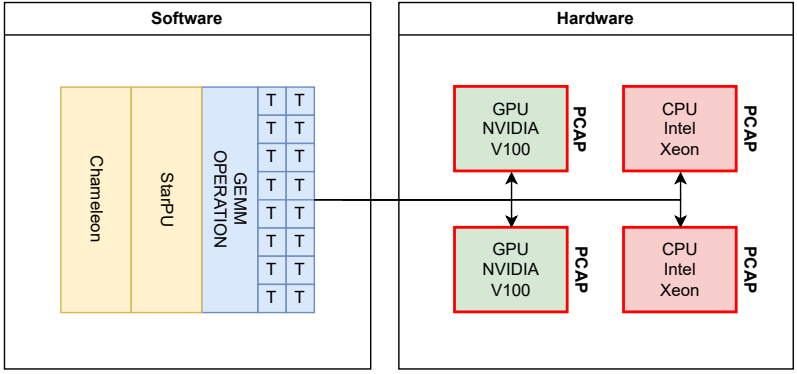
Power cap on CPU

We want to study the impact of power capping on 2 GPUs and 2 CPUs.



Power cap on CPU

We want to study the impact of power capping on 2 GPUs and 2 CPUs.



Results with CPU power capping

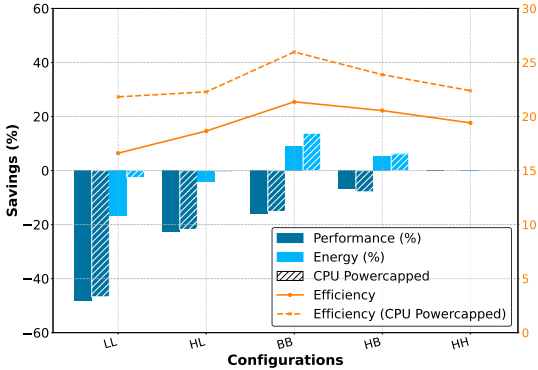
Platform hardware

- 2x CPU Intel Xeon Gold 6126
- 2x GPU Nvidia V100-PCIE-16GB

Power settings for a GPU

	Config.	Power (W)	Percent.
P_{low}	L	100	40%
P_{best}	B	155	62%
P_{high}	H	250	100%

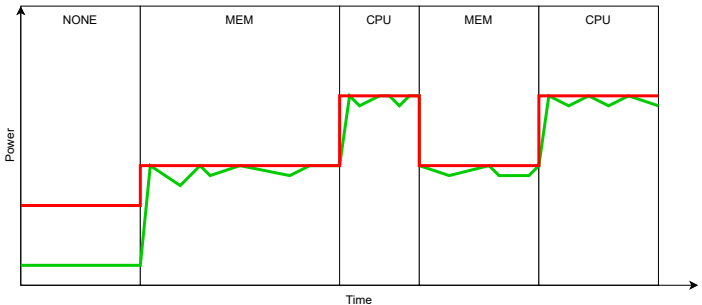
Default configuration is HH



- **26% energy efficiency improvements** when power capping both CPUs.

From static to dynamic power capping

How dynamic power capping behaves compared to static power capping?



- Adapt power capping on the go with the detected **phases** and measured **metrics**
- We can introduce a user tolerated **slowdown**

The aim is to have a better power management with a power cap defined by the user.

DUFP

DUFP (Dynamic Uncore Frequency and Power capping)

- Uncore frequency scaling and dynamic power capping tool.
- Separated process that runs beside main applications.
- Only on one CPU.

DUFP

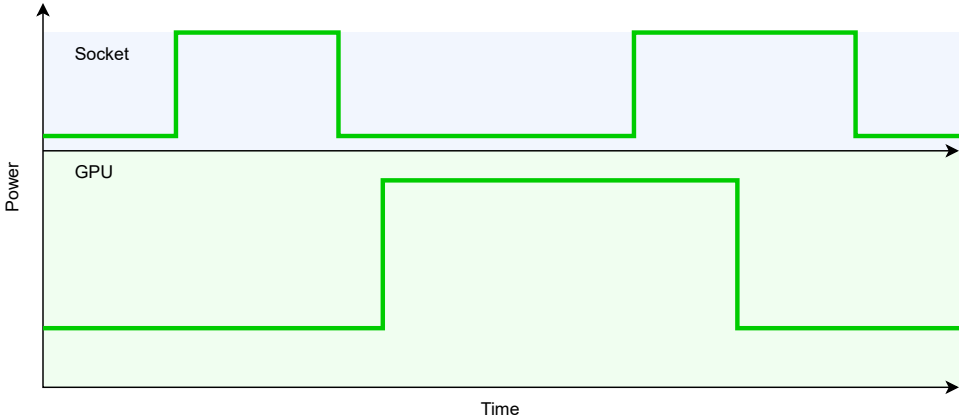
DUFP (Dynamic Uncore Frequency and Power capping)

- Uncore frequency scaling and dynamic power capping tool.
- Separated process that runs beside main applications.
- Only on one CPU.

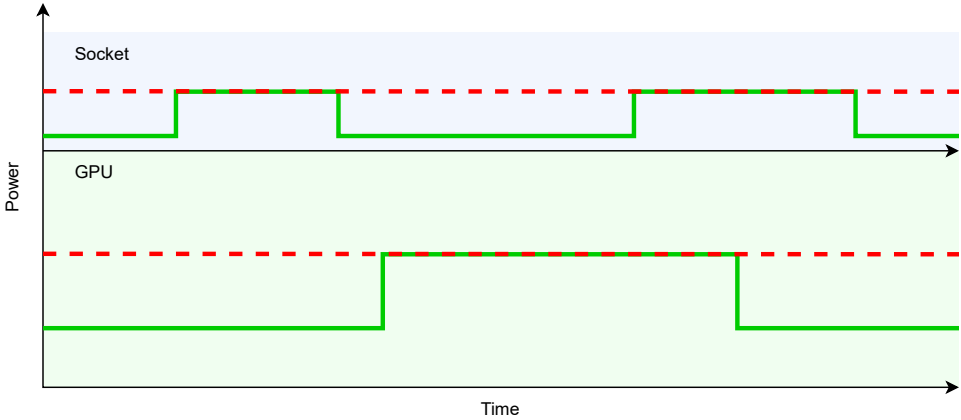
Contribution: make DUFP a heterogeneous power balancing tool

- Multi socket, multi GPU
- Global node power management
- Power balancing between sockets and GPUs

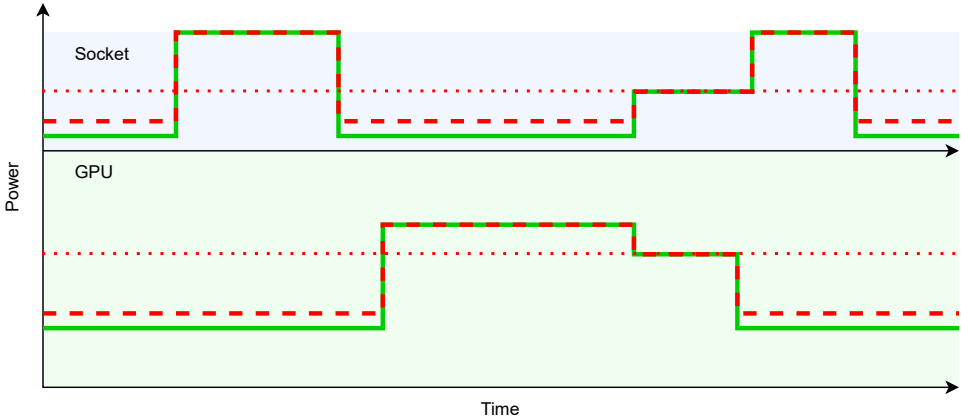
Example of heterogeneous dynamic power capping



Example of heterogeneous dynamic power capping



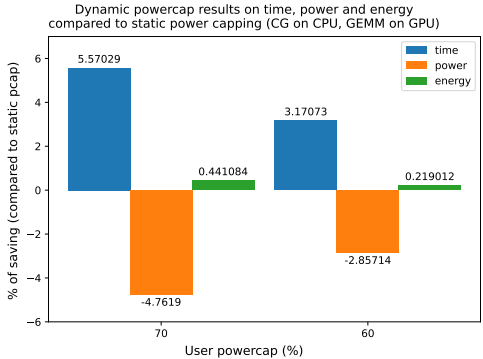
Example of heterogeneous dynamic power capping



First results

First results with one CPU and one GPU

- 1 Intel CPU
- 1 NVIDIA V100 GPU
- CG (NAS Parallel Benchmark) on CPU
- DGEMM (MAGMA) on GPU



Conclusion

Static power capping

- Applying power capping on a single GPU **improve energy efficiency**.
- On multiple GPUs, applying the best power cap on all GPUs provide a **20 % energy efficiency improvement**.
- With CPU power capping, we can reach 26 % of energy efficiency improvement.

Dynamic power capping

- Adapted DUFP for multi-socket usage.
- Added heterogeneous balancing for power capping.
- New algorithm proposed and first promising results.

Future work

- Test power capping with more complex HPC applications.
- Mix **energy efficiency aware** scheduler and power capping (Jules Risse).