ICPE 2020 11th ACM / SPEC International Conference on Performance Engineering

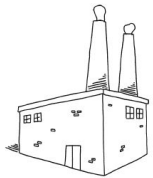
Hello!

Taming Energy Consumption Variations In Systems Benchmarking

Zakaria **OURNANI**
Chakib Mohammed **BELGAID**
Romain **ROUVOY**
Pierre **RUST**
Joel **PENHOAT**
Lionel **SEINTURIER**

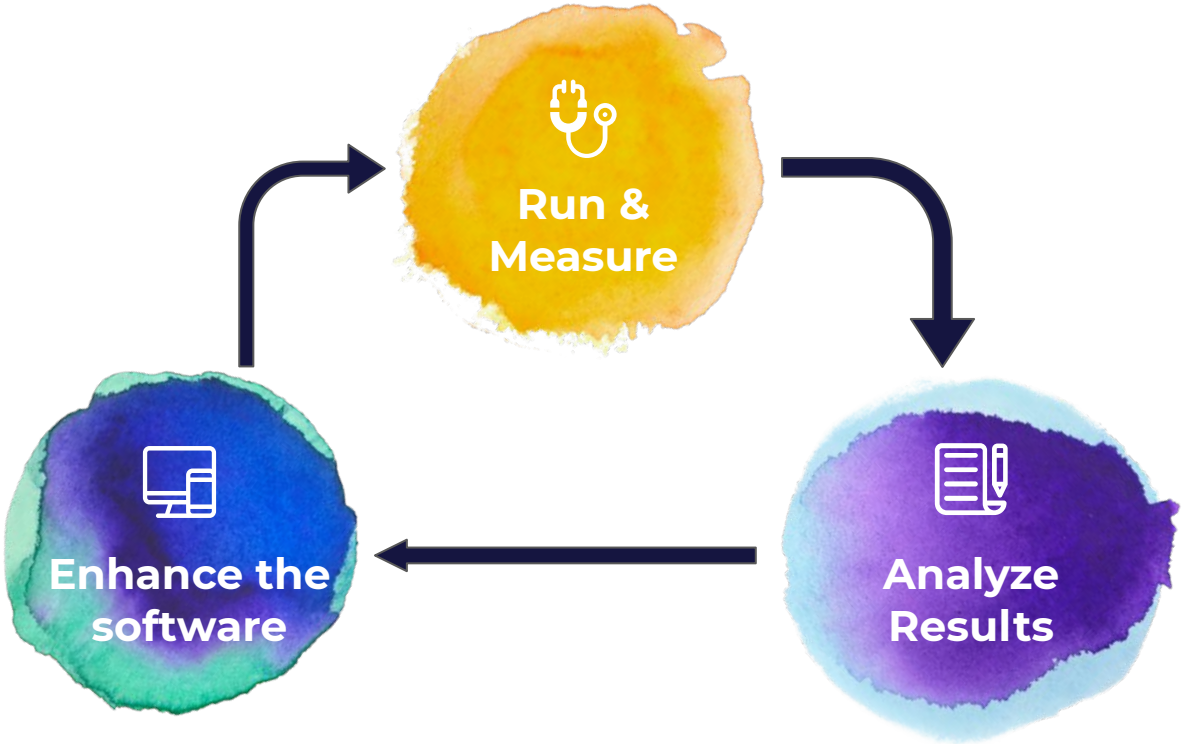
Motivation

- Digital energy consumption knows a raise of 8.5% per year [1]
- Data centers are responsible of 2% of the extra CO2 in the air [2]

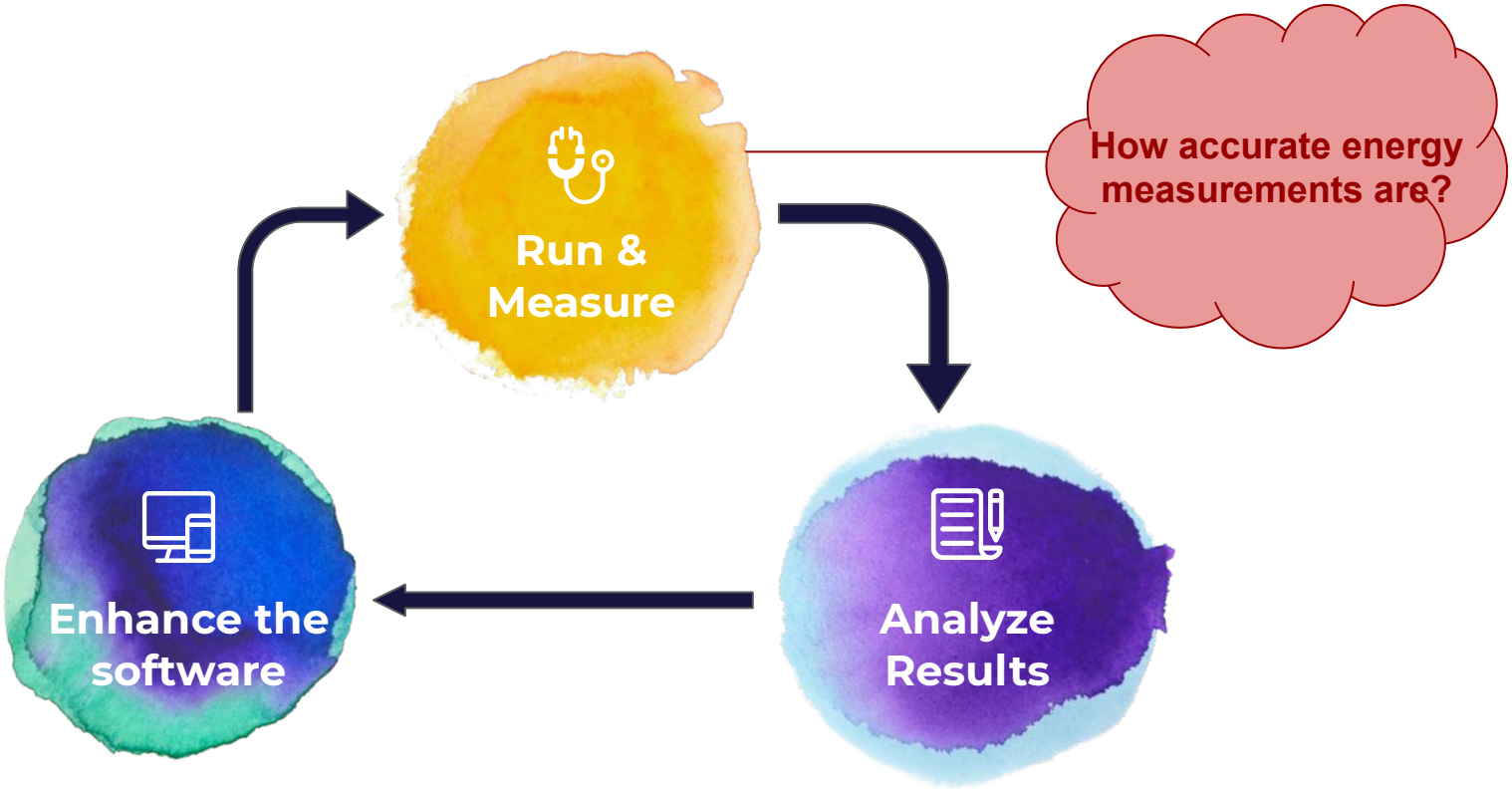


[1] Hugues Ferreboeuf, Maxime Efoui-Hess, Zeynep Kahraman (2018).LEAN ICT POUR UNE SOBRIETE NUMERIQUE study.The shift project
[2] Avgerinou, Maria, Paolo Bertoldi, and Luca Castellazzi. "Trends in Data Centre Energy Consumption under the European Code of Conduct for Data Centre Energy Efficiency." Energies 10, no. 10 (September 22, 2017): 1470. https://doi.org/10.3390/en10101470.

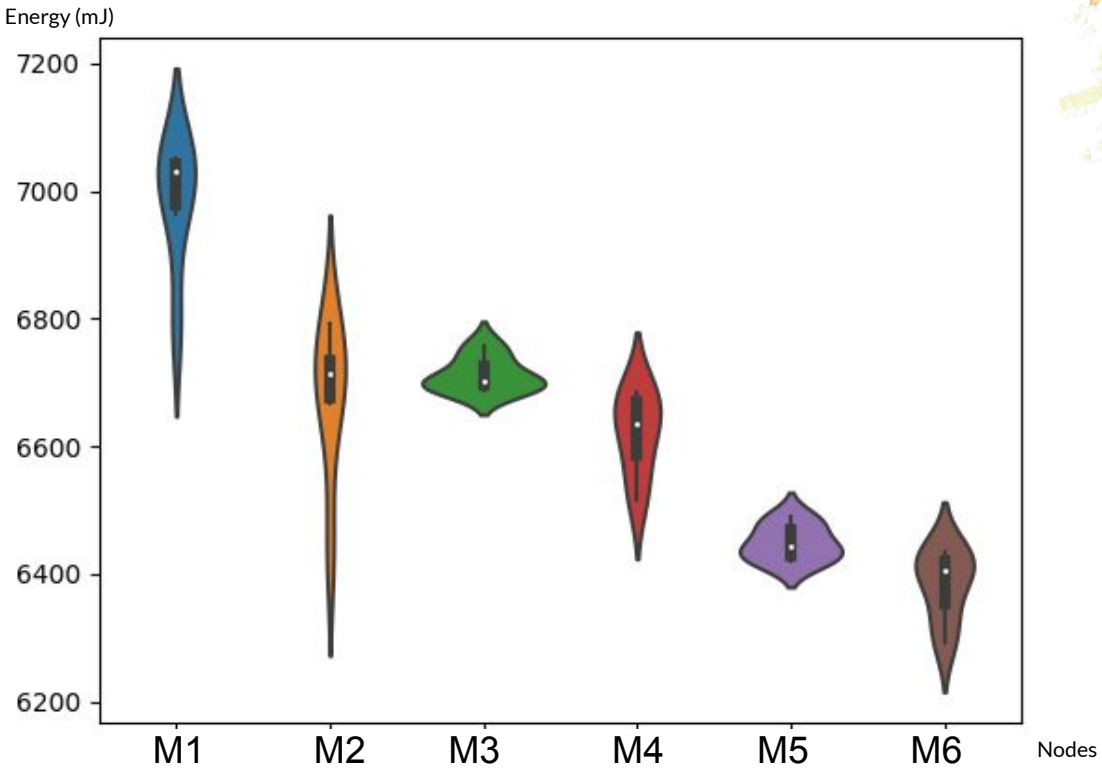
Green Software Design



Green Software Design

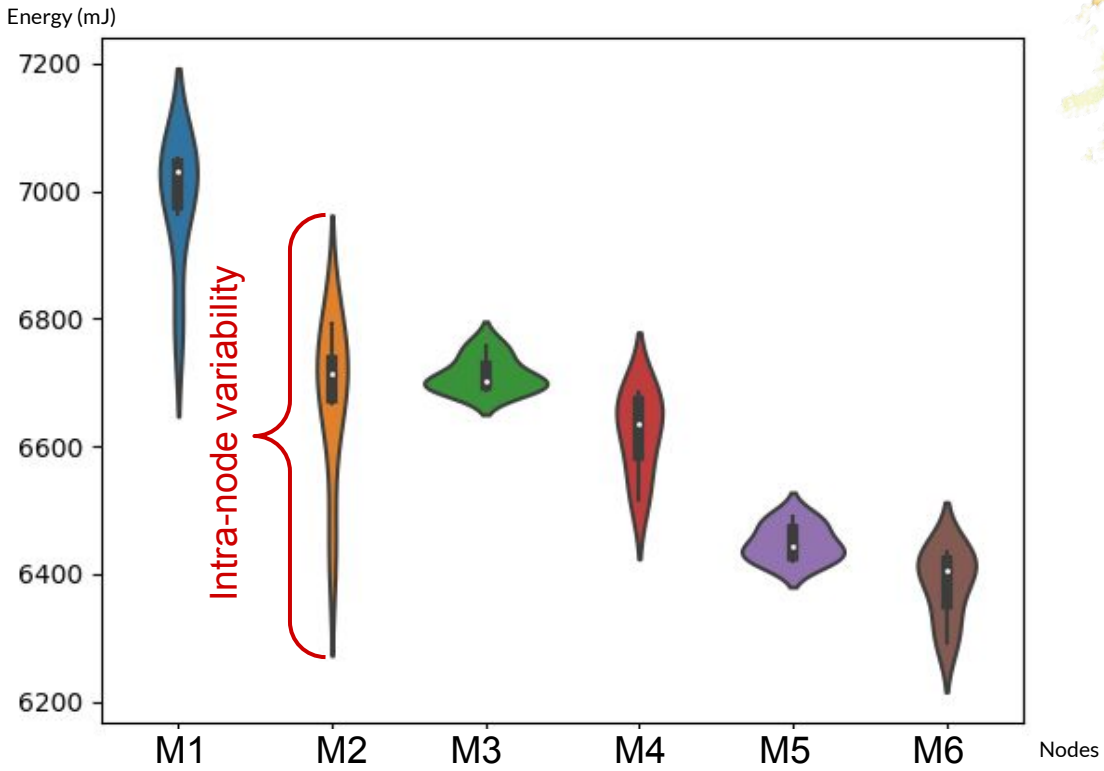


Case of Study



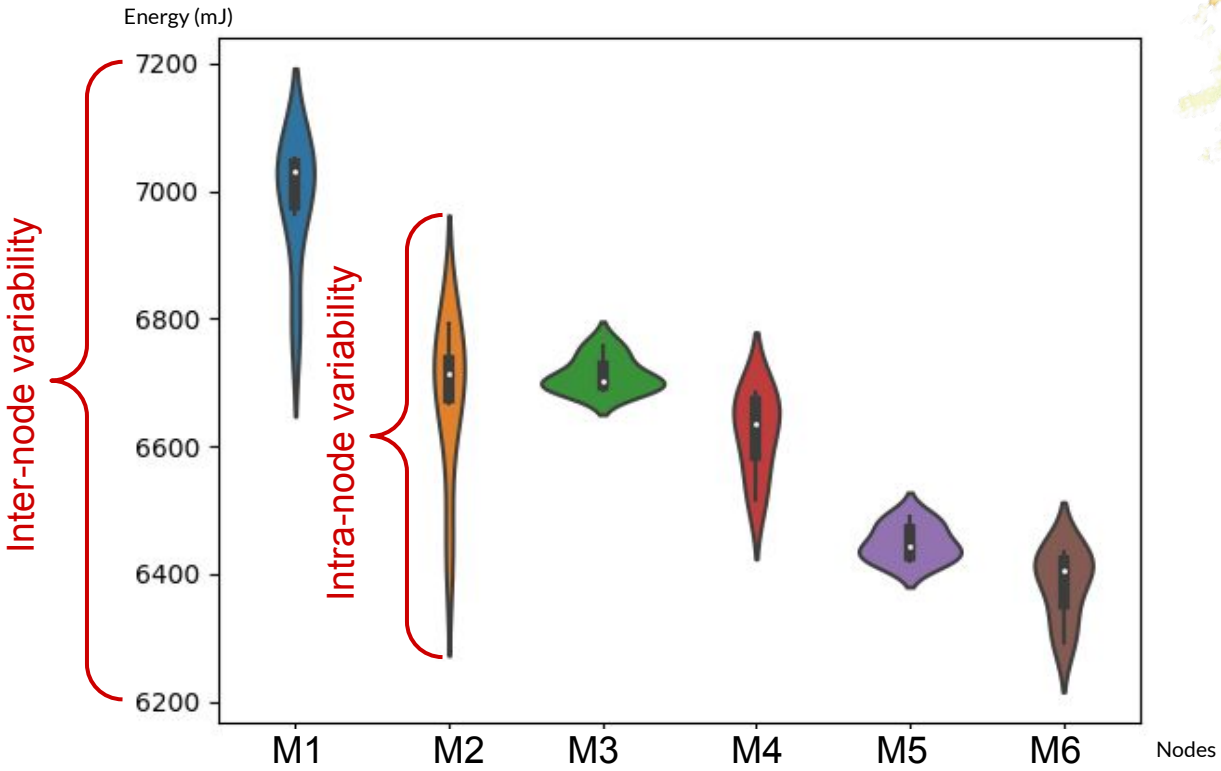
Violin plot of the energy consumption variation of the same Test running 30 times on 6 different machines

Case of Study



Violin plot of the energy consumption variation of the same Test running 30 times on 6 different machines

Case of Study



Violin plot of the energy consumption variation of the same Test running 30 times on 6 different machines

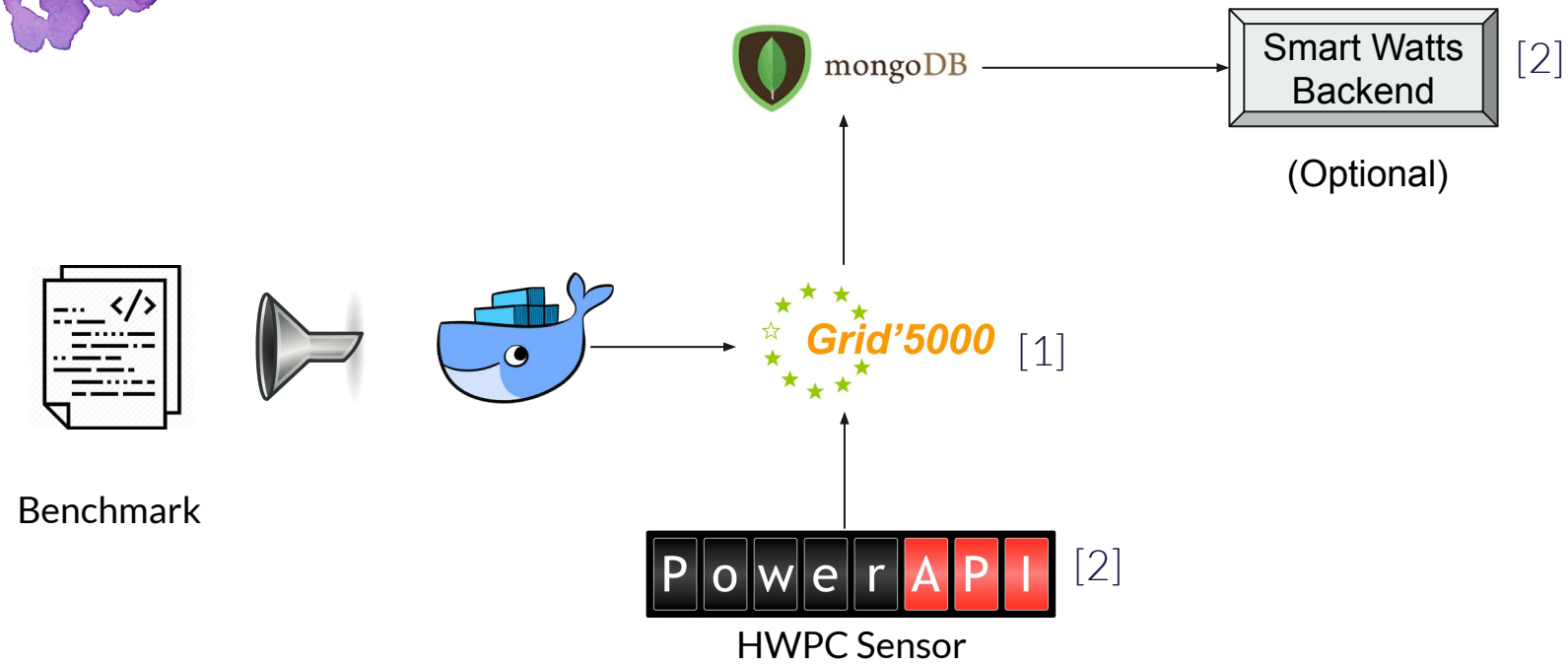
Objectives

- Investigate the energy consumption variation on multiple CPU and clusters
- Identify controllable factors that contribute that variation
- Report on guideline on how to conduct reproducib experiments with less variations

1

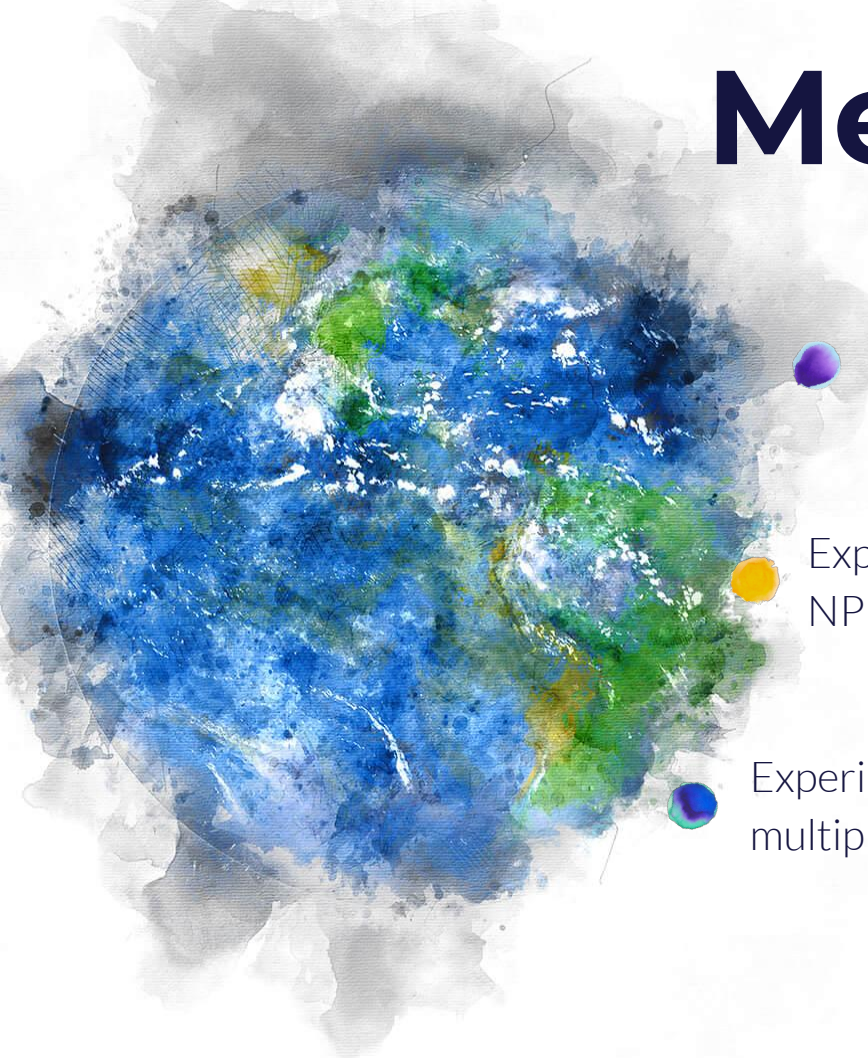
Methodology

Experimental setup



[1] www.grid5000.fr
[2] Maxime Colmant, Romain Rouvoy, Mascha Kurpicz, Anita Sobe, Pascal Felber, and Lionel Seinturier. 2018. The next 700 CPU power models. Journal of Systems and Software 144 (2018).

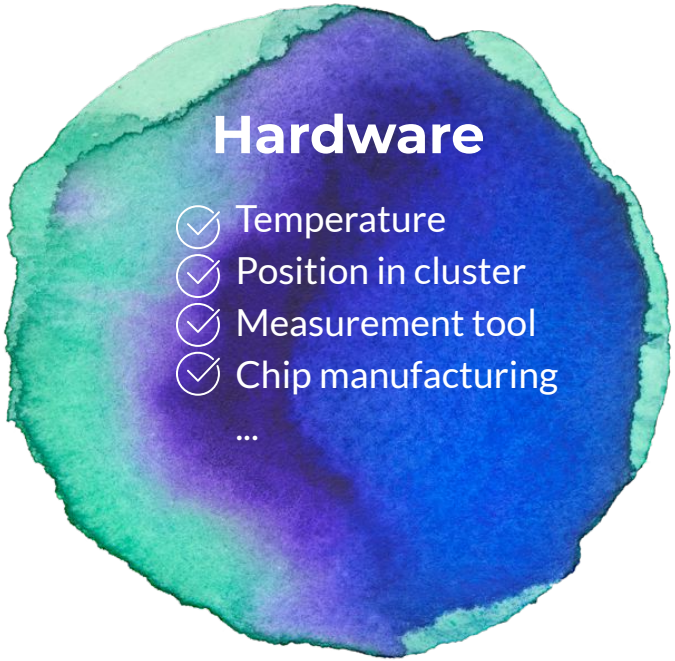
Methodology

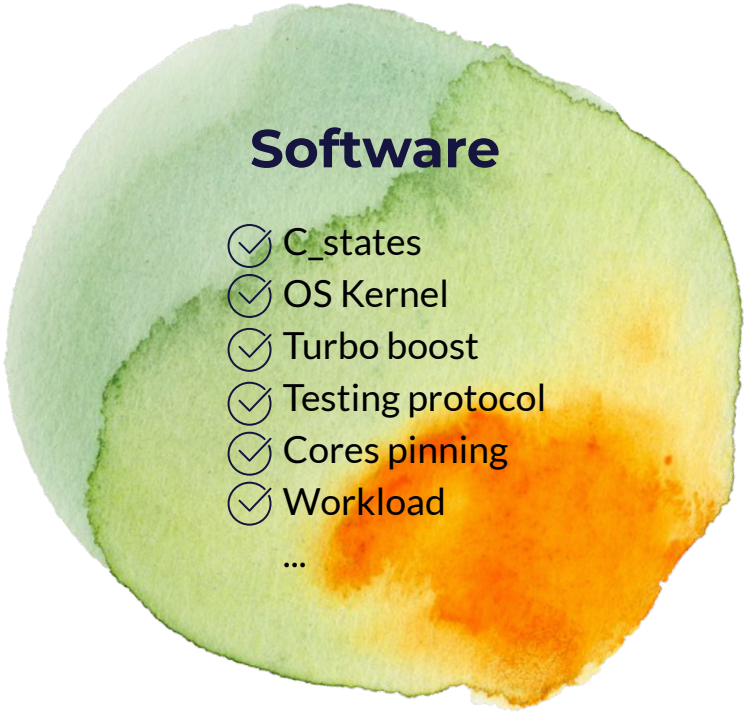


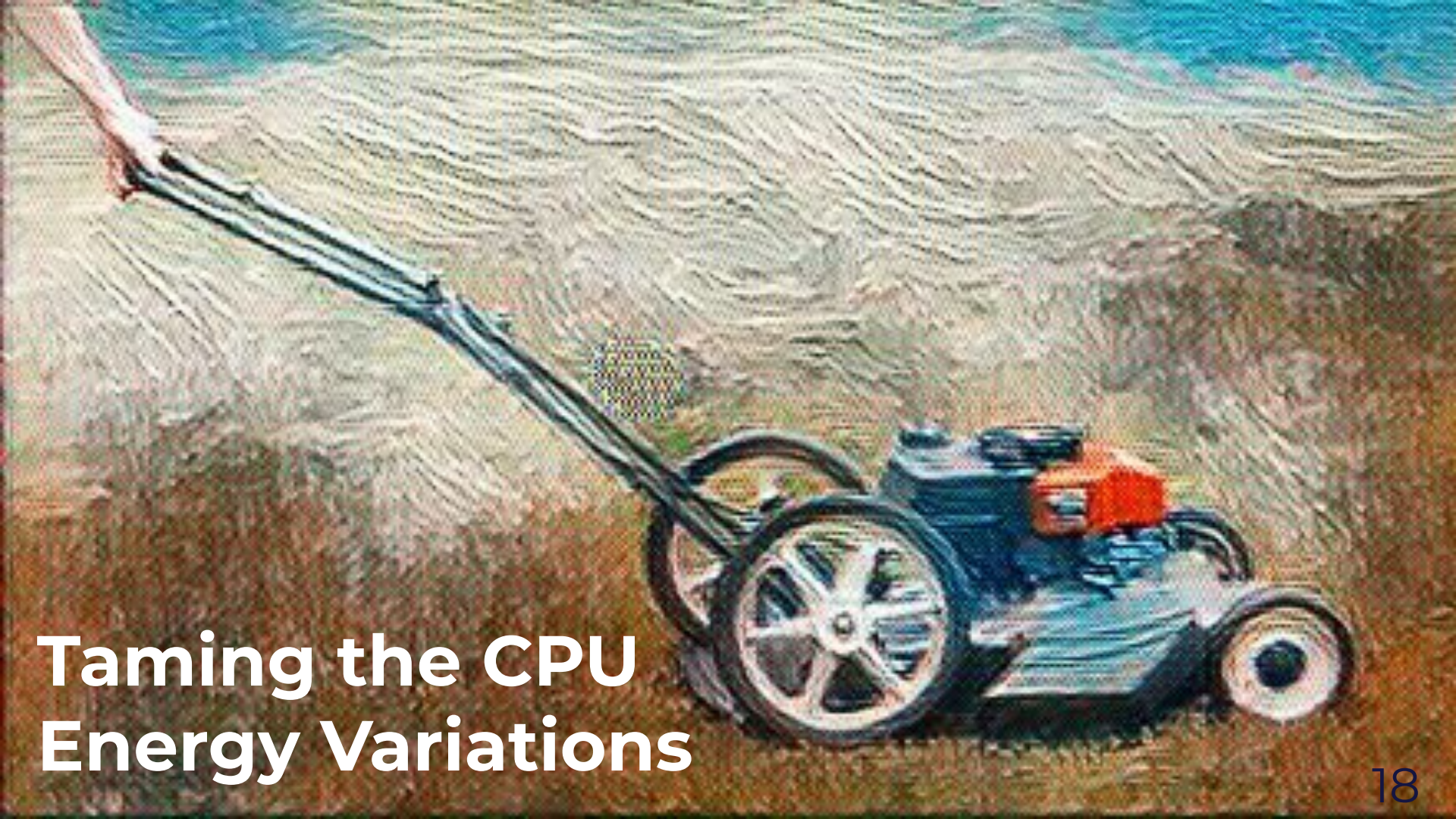
- Every test is executed over 100 times in each condition to build statistically representative results
- Experiments are executed with many benchmarks, such as: NPB, Linpack, Sha, Stress-ng, Pbzip2
- Experiments are executed across multiple identical nodes of multiple clusters with different capabilities


2
CPU Energy Variation

Potential Parameters



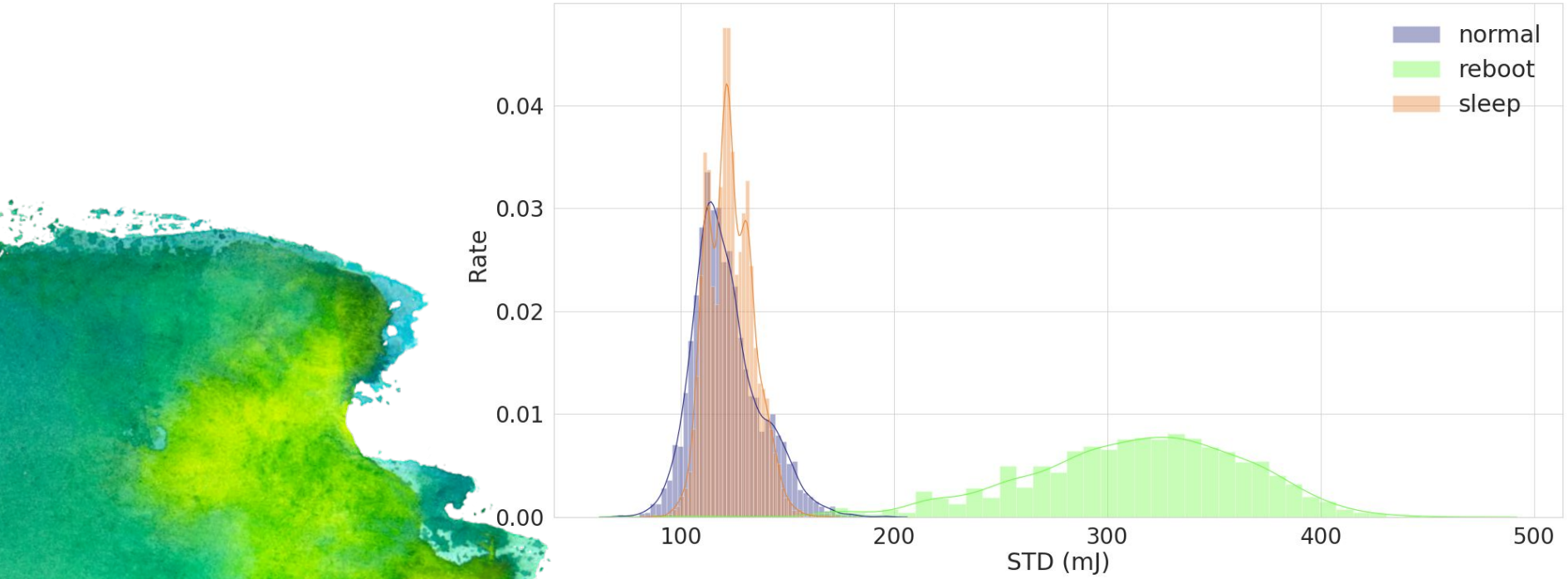




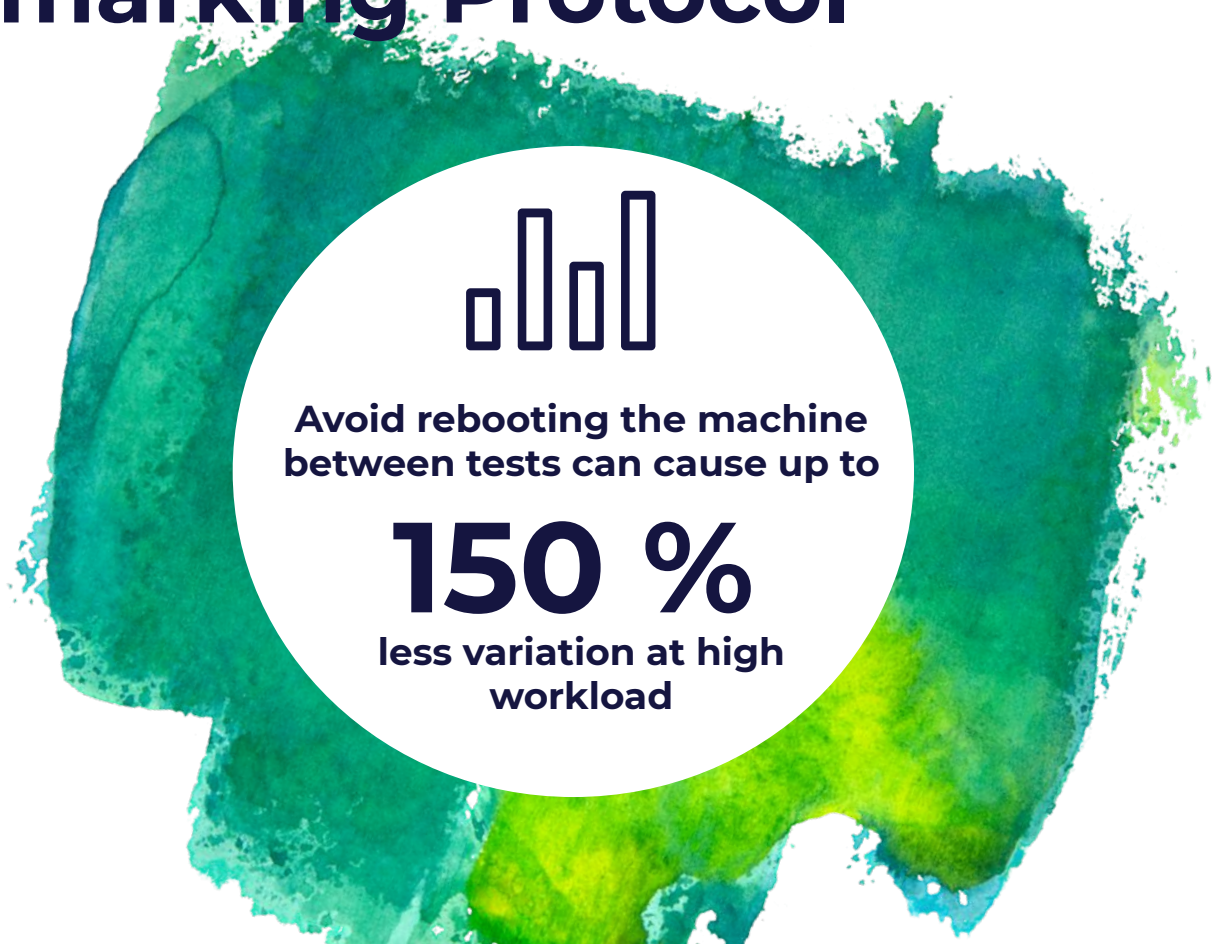
Taming the CPU Energy Variations

RQ1:
Does the benchmarking protocol affect the energy variation?

Benchmarking Protocol



Benchmarking Protocol



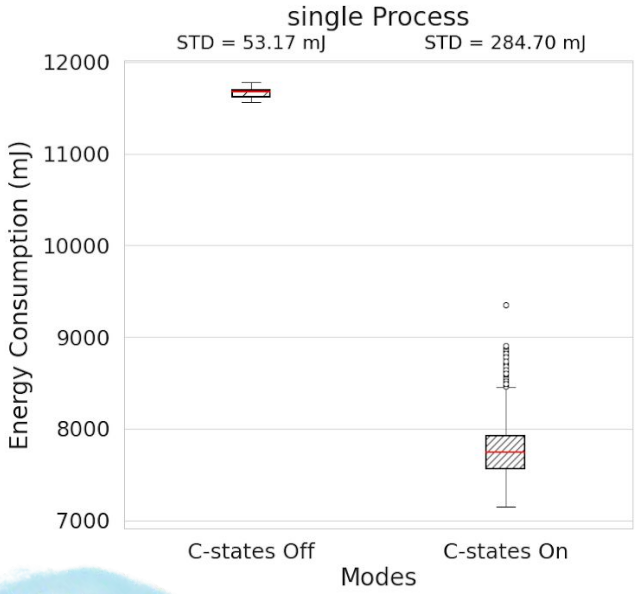
Avoid rebooting the machine between tests can cause up to

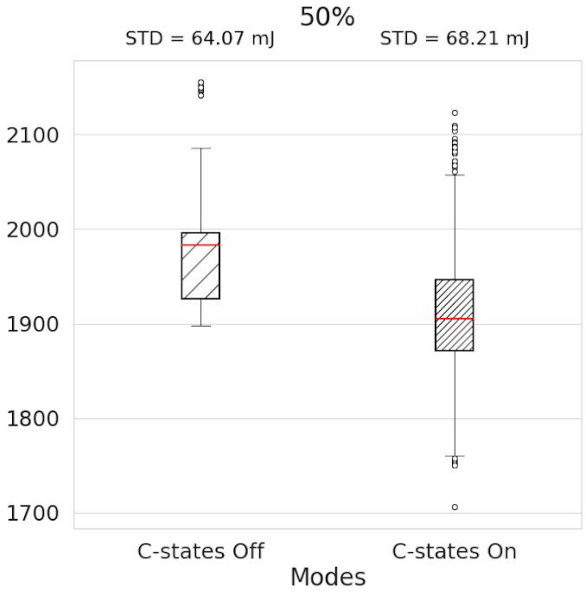
150 %

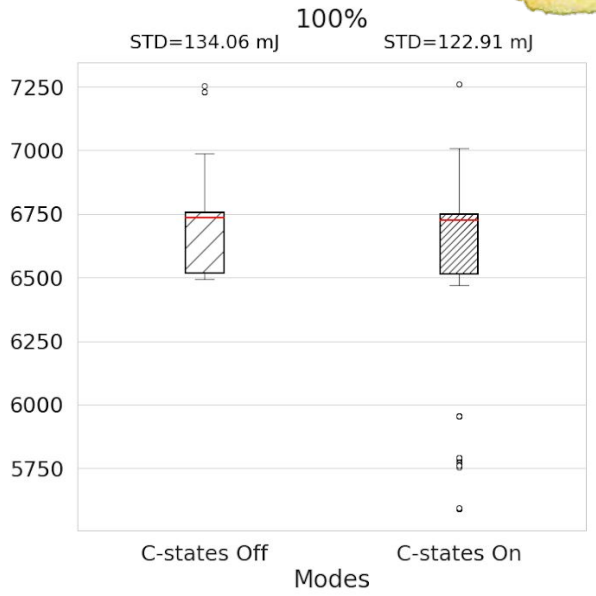
less variation at high workload

RQ2:
How important is the impact of the processor features on the energy variation?

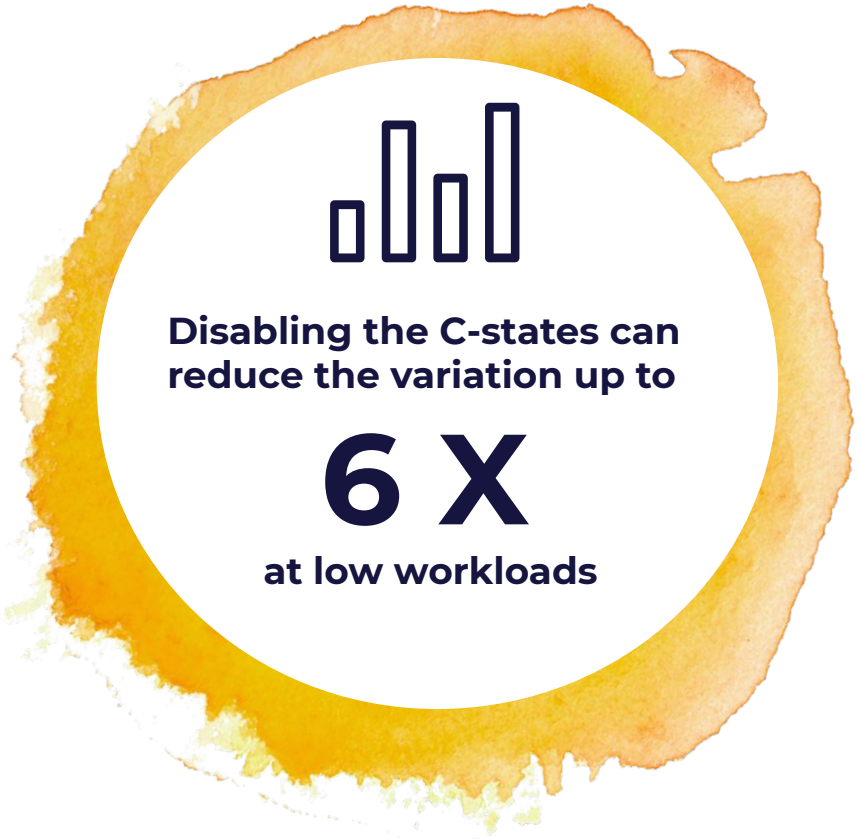
CPU C-states


single Process
STD = 53.17 mJ
STD = 284.70 mJ
Energy Consumption (mJ)
12000
11000
10000
9000
8000
7000
C-states Off
C-states On
Modes


50%
STD = 64.07 mJ
STD = 68.21 mJ
2100
2000
1900
1800
1700
C-states Off
C-states On
Modes

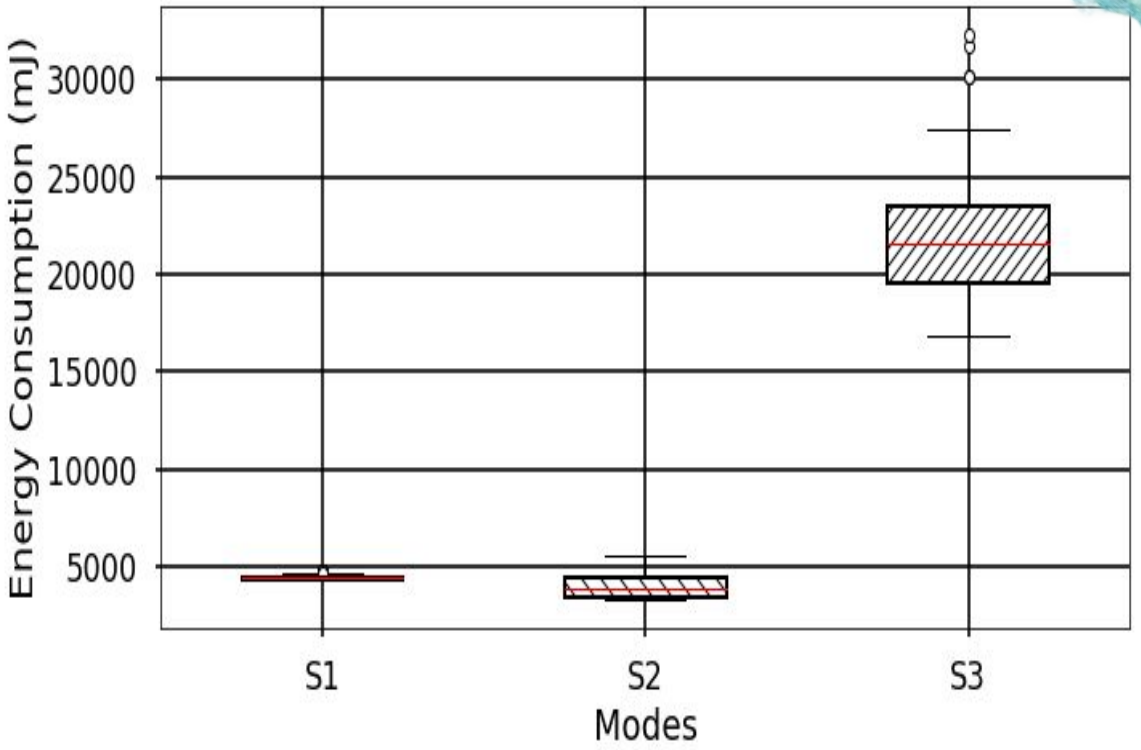

100%
STD=134.06 mJ
STD=122.91 mJ
7250
7000
6750
6500
6250
6000
5750
C-states Off
C-states On
Modes

CPU C-states

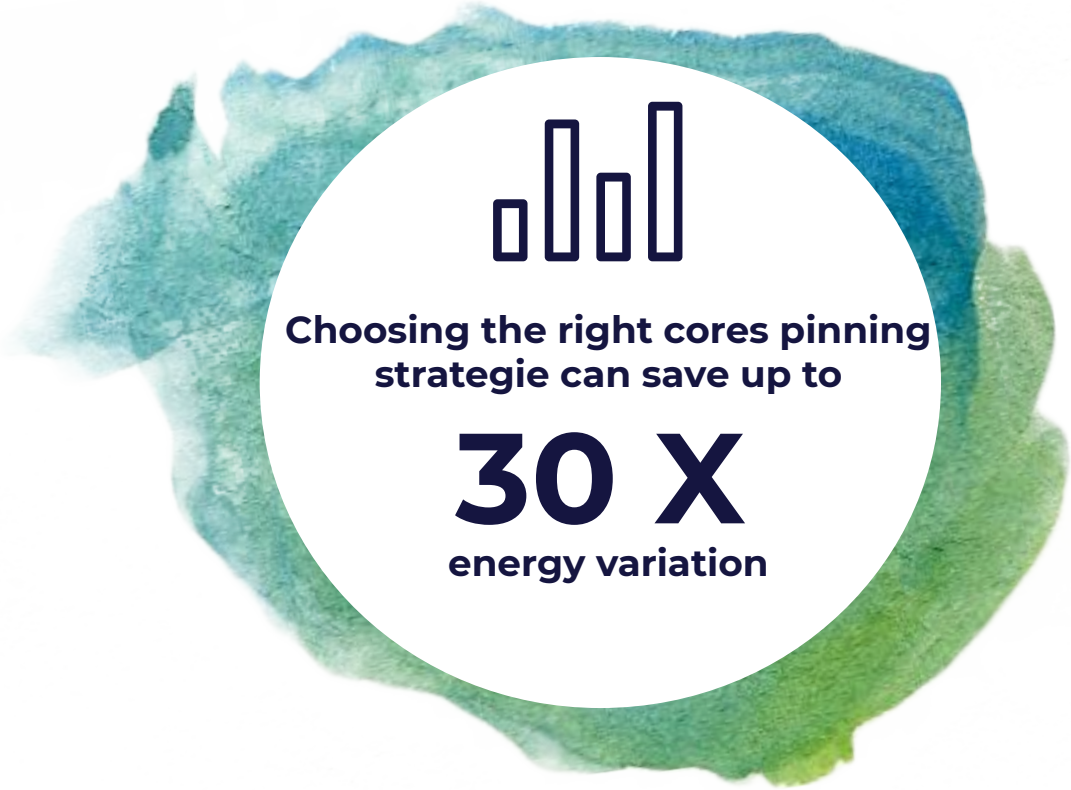

Disabling the C-states can reduce the variation up to
6 X
at low workloads

Core Pinning

S1	- Minimum of physical CPUs - HT usage
S2	- No HT
S3	- Usage of all Physical CPUs - Least Cores count usage - HT usage



Core Pinning



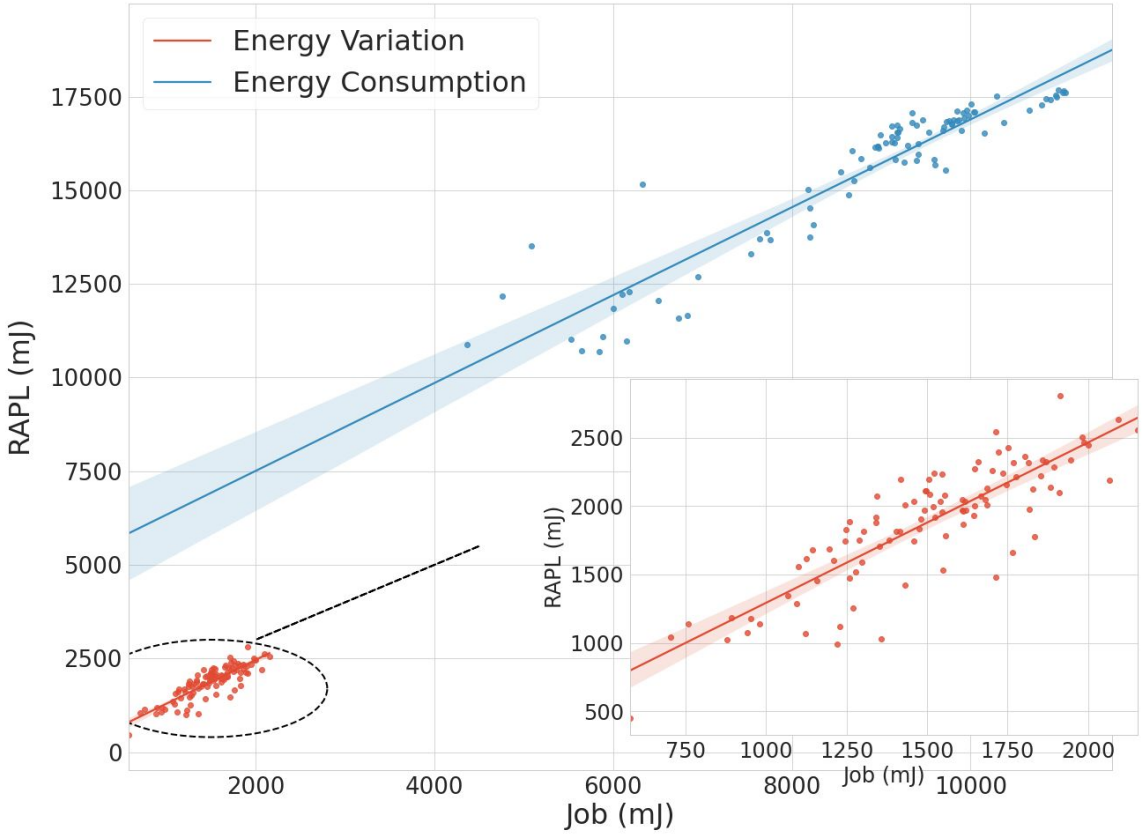
Choosing the right cores pinning strategie can save up to

30 X

energy variation

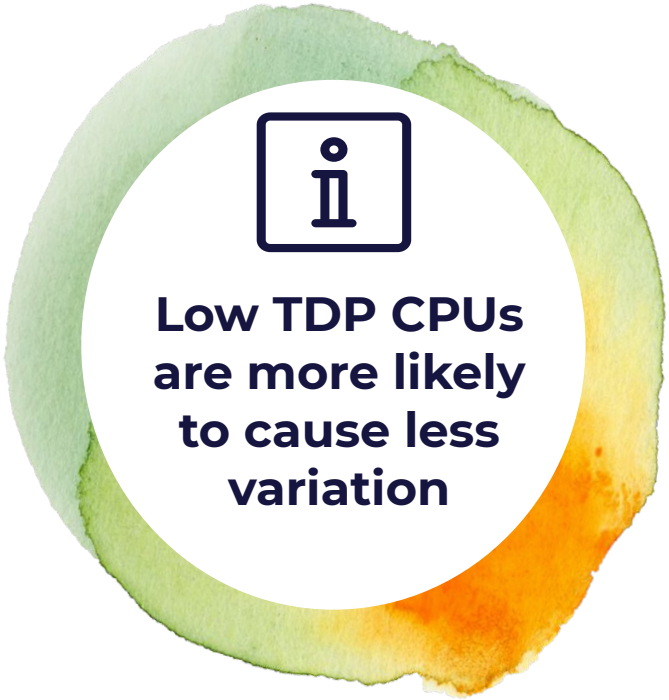
RQ3:
What is the impact of the operating system on the energy variation?

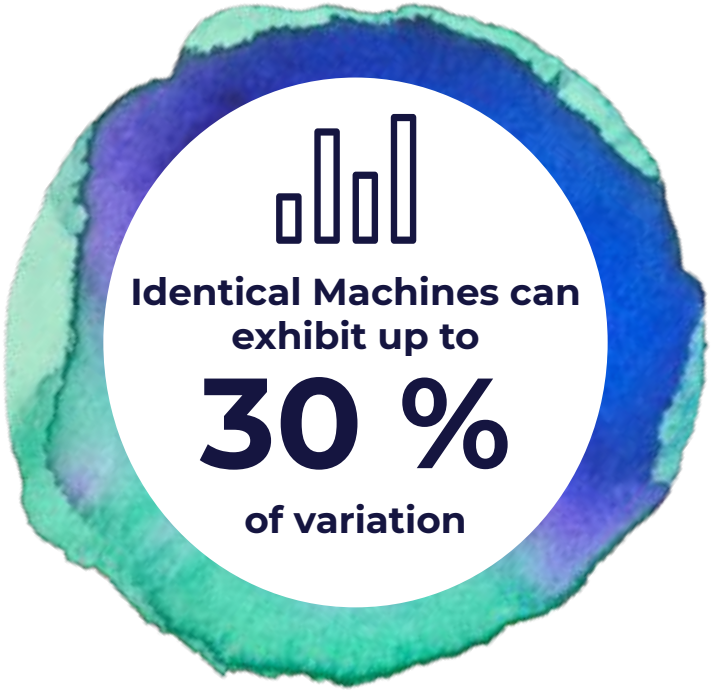
OS Impact


Energy Variation
Energy Consumption
RAPL (mJ)
Job (mJ)
0
2500
5000
7500
10000
12500
15000
17500
2000
4000
6000
8000
10000
RAPL (mJ)
Job (mJ)
500
1000
1500
2000
2500
750
1000
1250
1500
1750
2000

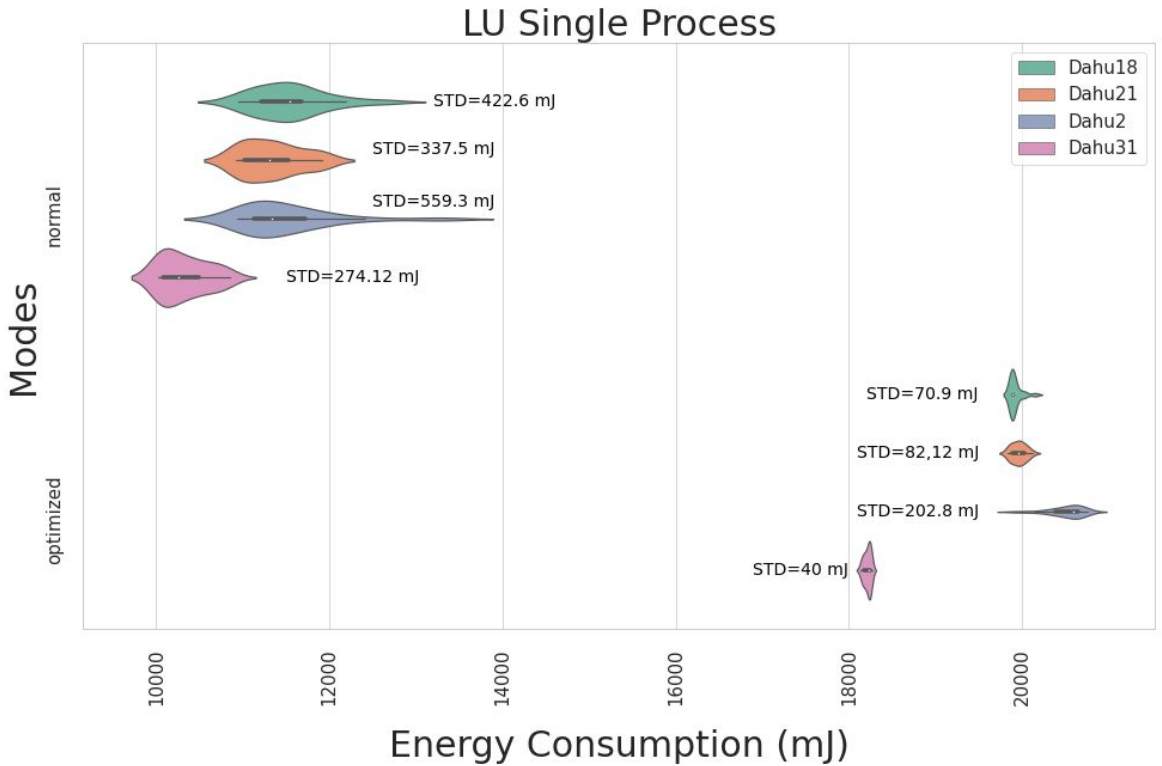
RQ4:
Does the choice of the processor matter to mitigate the energy variation?

Processor Choice





Inter-Nodes Variation



Main Guidelines

Guideline	Workload	Gain
Use a low TDP CPU	Low & Medium	3X
Disable the CPU C-states	Low	6X
Avoid the usage of Hyper-threading	Medium	5X
Use the least of physical CPU in case of multiple CPU	Medium	30X
Avoid rebooting the machine between tests	High	1.5X
Use the same machine instead of similar machines	All	1.3X

Conclusion

- Provide a better understanding of the intra-node and inter-nodes variations
- Identify a set of controllable factors that contribute to the CPU energy consumption variation
- Provide guidelines on how to conduct reproducible experimentations with less variation

Avoid rebooting the machine between tests can cause up to

150 %

less variation at high workload

Choose the right cores pinning strategie can save up to

30 X

of energy variation

Identical Machines can exhibit up to

30 %

of variation

Disabling the C-states can reduce the variation by up tp

6X

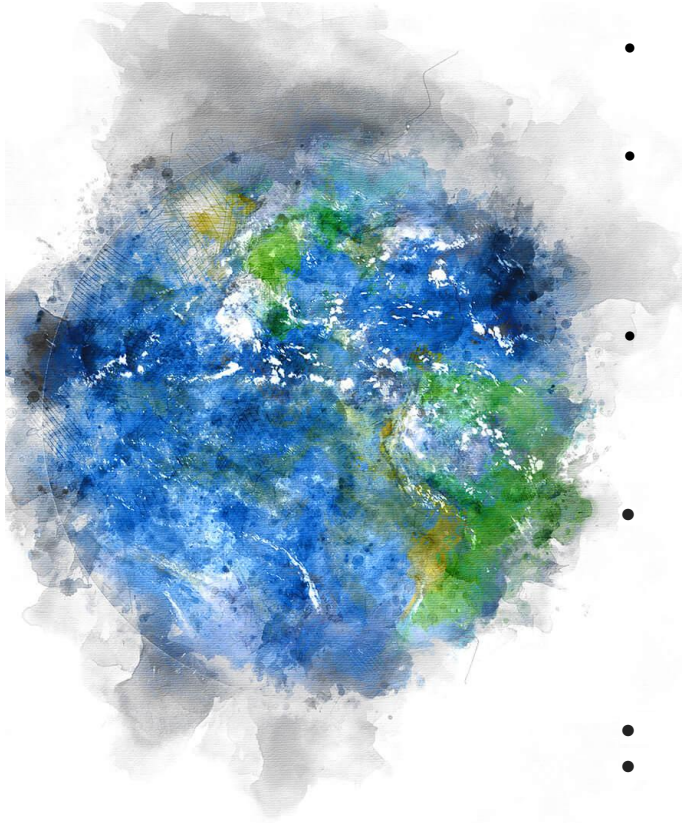
at low workloads

Low TDP CPUs are more likely to cause less variation

The Energy variation is more related the the job rather than the OS

References

- Simakov, Nikolay A., et al. "Effect of meltdown and spectre patches on the performance of HPC applications." *arXiv preprint arXiv:1801.04329* (2018).
- Varsamopoulos, Georgios, Ayan Banerjee, and Sandeep KS Gupta. "Energy efficiency of thermal-aware job scheduling algorithms under various cooling models." *International Conference on Contemporary Computing*. Springer, Berlin, Heidelberg, 2009.
- Wang, Yewan, et al. "Potential effects on server power metering and modeling." *Wireless Networks* (2018): 1-8.
- Margery, David, et al. "Resources Description, Selection, Reservation and Verification on a Large-scale Testbed." *International Conference on Testbeds and Research Infrastructures*. Springer, Cham, 2014.



- Colmant, Maxime, et al. "The next 700 CPU power models." *Journal of Systems and Software* 144 (2018): 382-396.
- Balouek, Daniel, et al. "Adding virtualization capabilities to the Grid'5000 testbed." *International Conference on Cloud Computing and Services Science*. Springer, Cham, 2012.
- Balouek, Daniel, et al. "Adding virtualization capabilities to the Grid'5000 testbed." *International Conference on Cloud Computing and Services Science*. Springer, Cham, 2012.
- Chasapis, Dimitrios, et al. "Runtime-guided mitigation of manufacturing variability in power-constrained multi-socket numa nodes." *Proceedings of the 2016 International Conference on Supercomputing*. 2016.
- www.grid5000.fr
- www.powerapi.org

Thanks !