Measuring the Energy Consumption of HPC Systems

Anne-Cécile Orgerie

ORAP Forum
9th December 2021

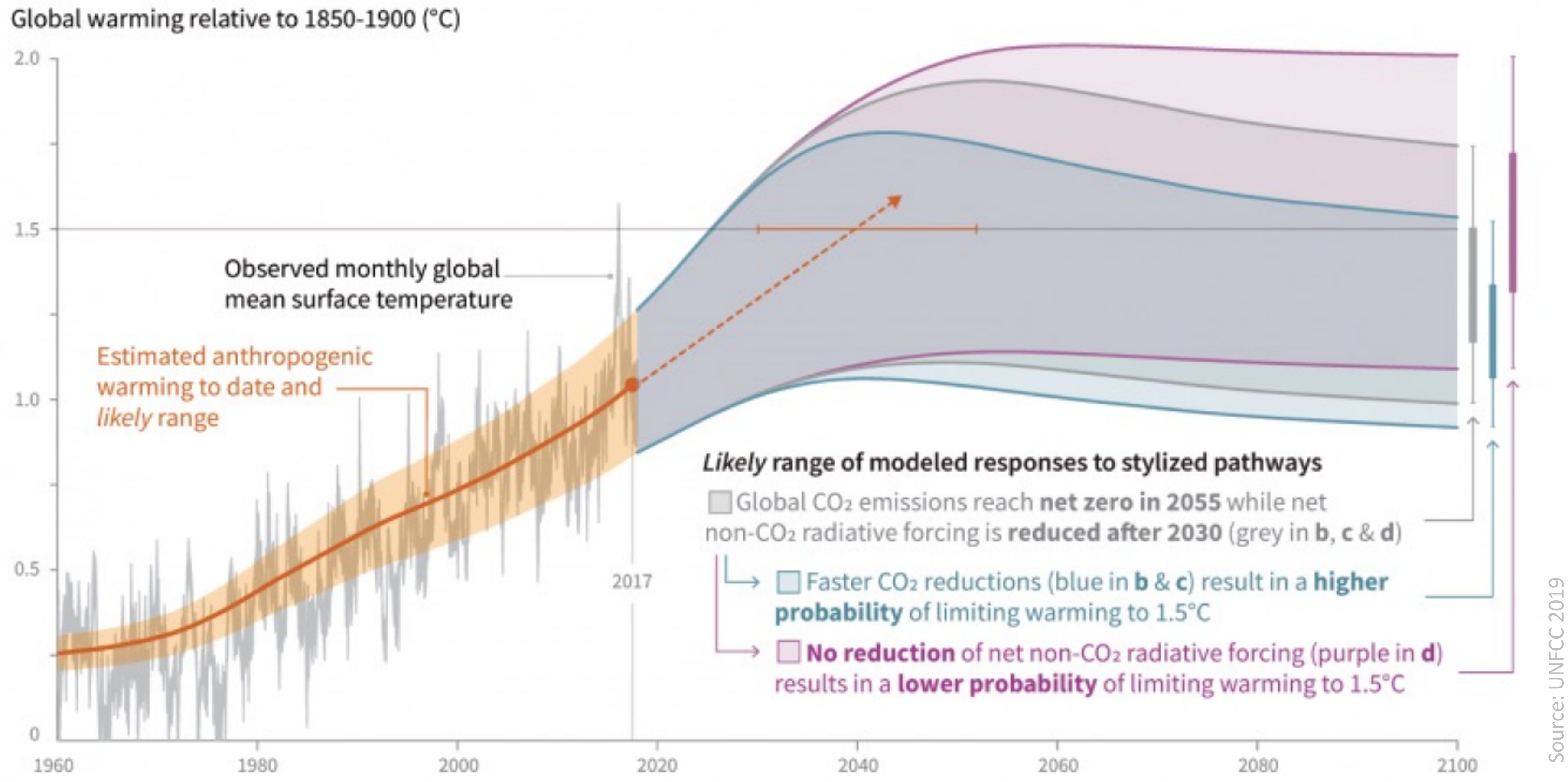
Outline

- Context
- Understanding the energy consumption of HPC systems
- Measuring accurately the energy consumption of HPC systems
- Modeling energy consumption of HPC systems
- Concluding broader remarks

Outline

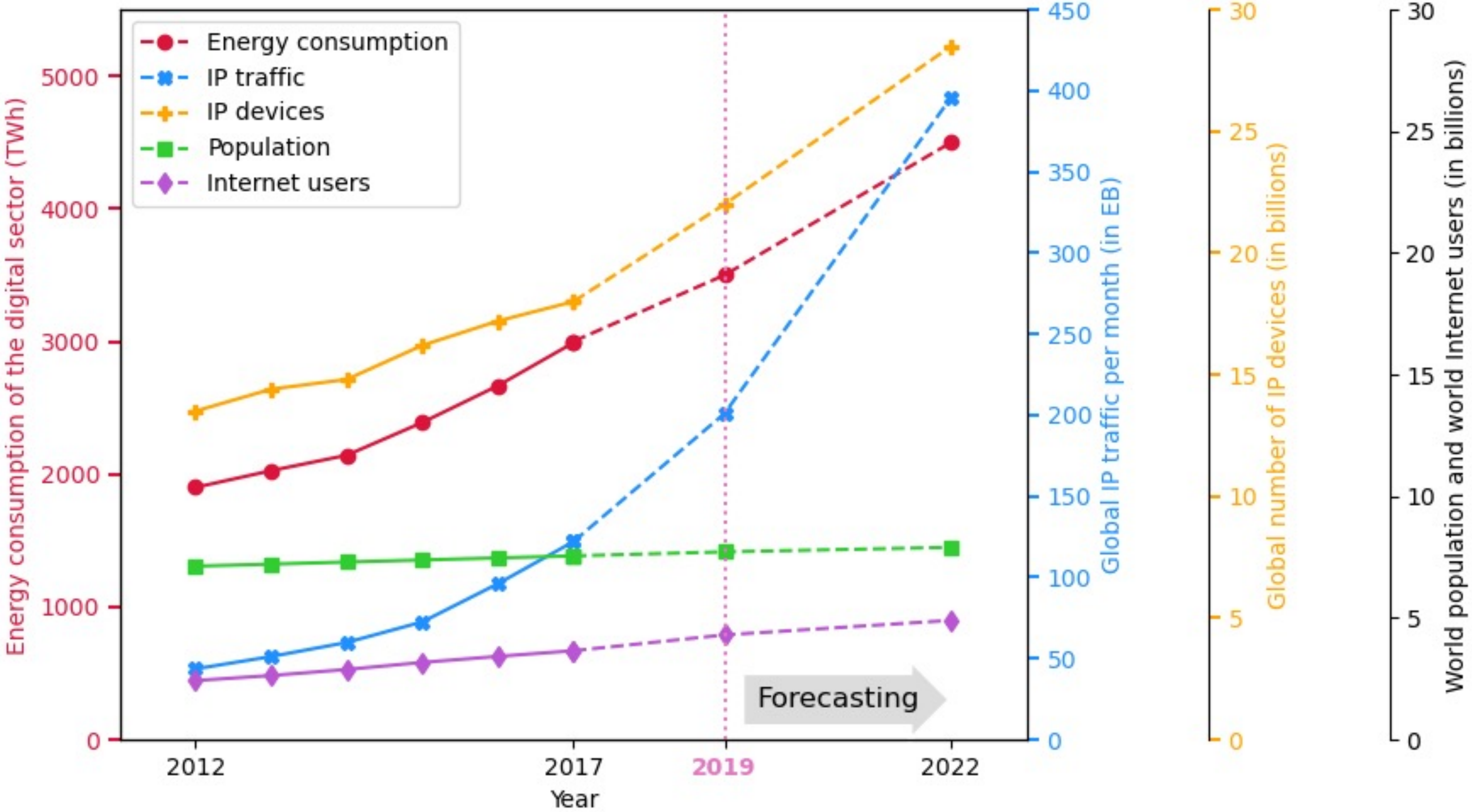
- Context
- Understanding the energy consumption of HPC systems
- Measuring accurately the energy consumption of HPC systems
- Modeling energy consumption of HPC systems
- Concluding broader remarks

Paris Agreement: 1.5° C



Objective in 2019: reducing global greenhouse gas emissions by **8%** each year

ICT energy consumption

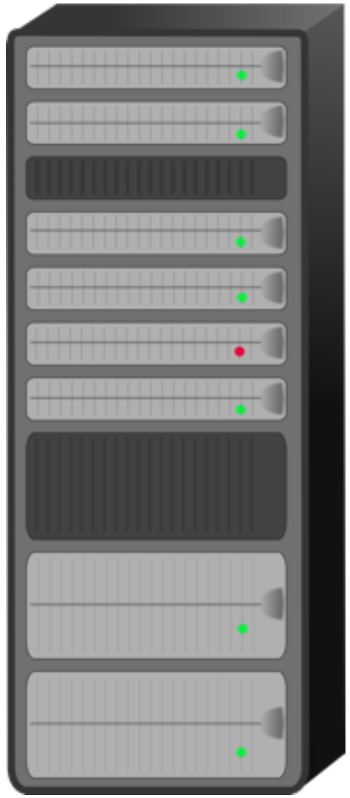


ICT energy consumption grows by **~9%** each year.

[HDR 2020]

My scientific context

- Energy consumption
- Large-scale distributed sytems
- Computing and networking parts
- Use phase



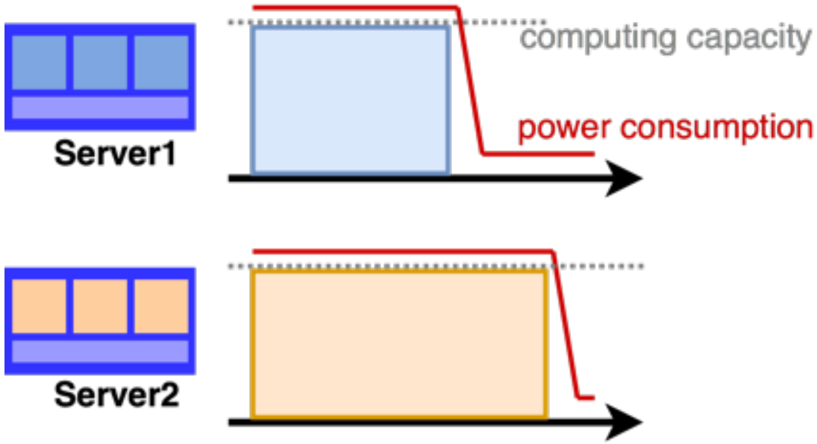
Started with Grid computing some years ago…

Outline

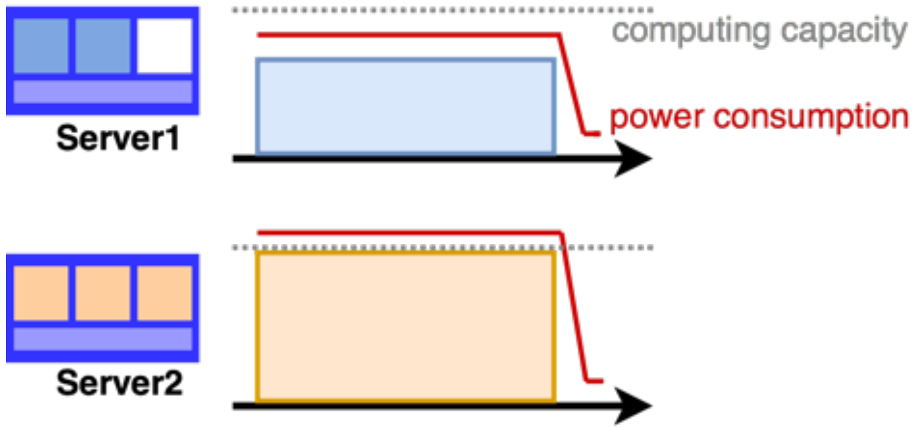
- Context
- Understanding the energy consumption of HPC systems
- Measuring accurately the energy consumption of HPC systems
- Modeling energy consumption of HPC systems
- Concluding broader remarks

Energy efficiency: business as usual?

Computing faster?

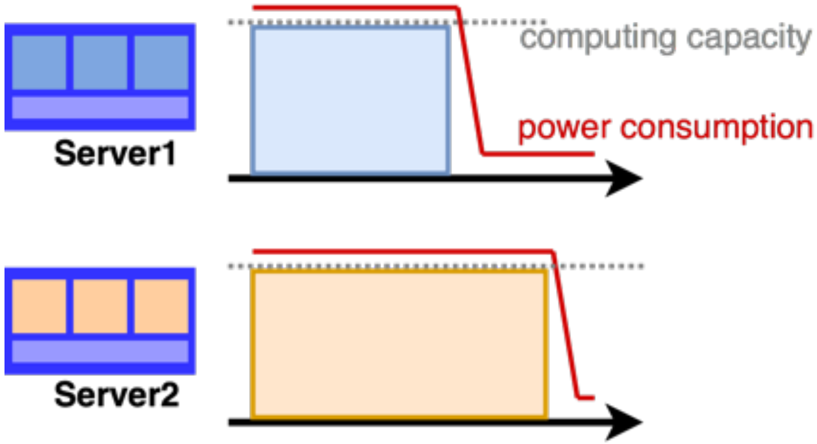


Computing slower?

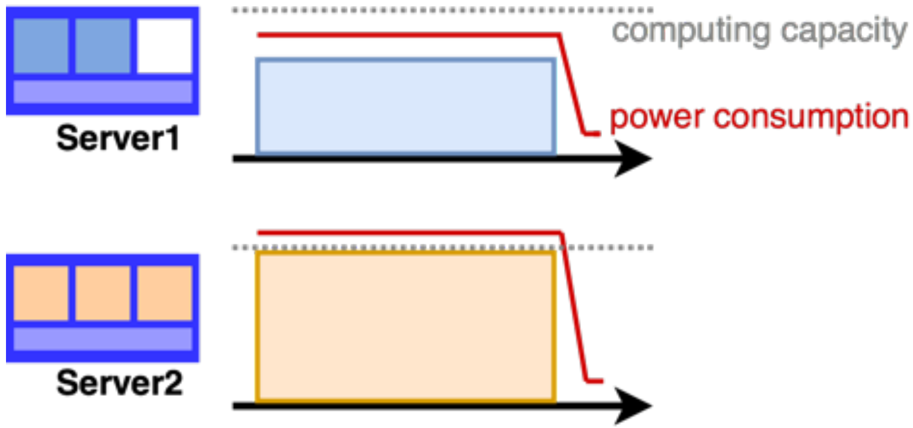


Energy efficiency: business as usual?

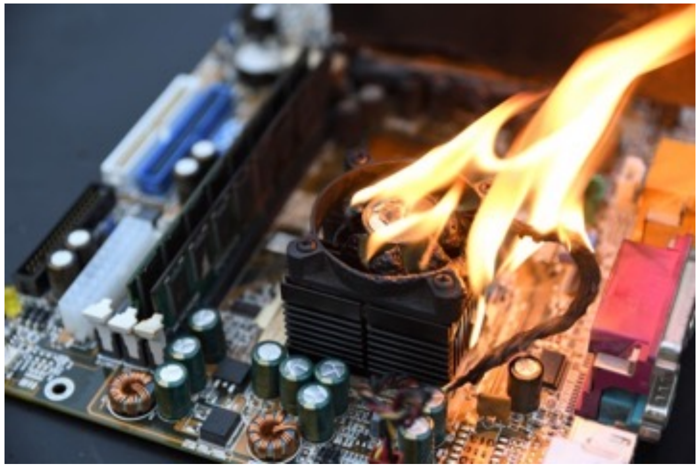
Computing faster?



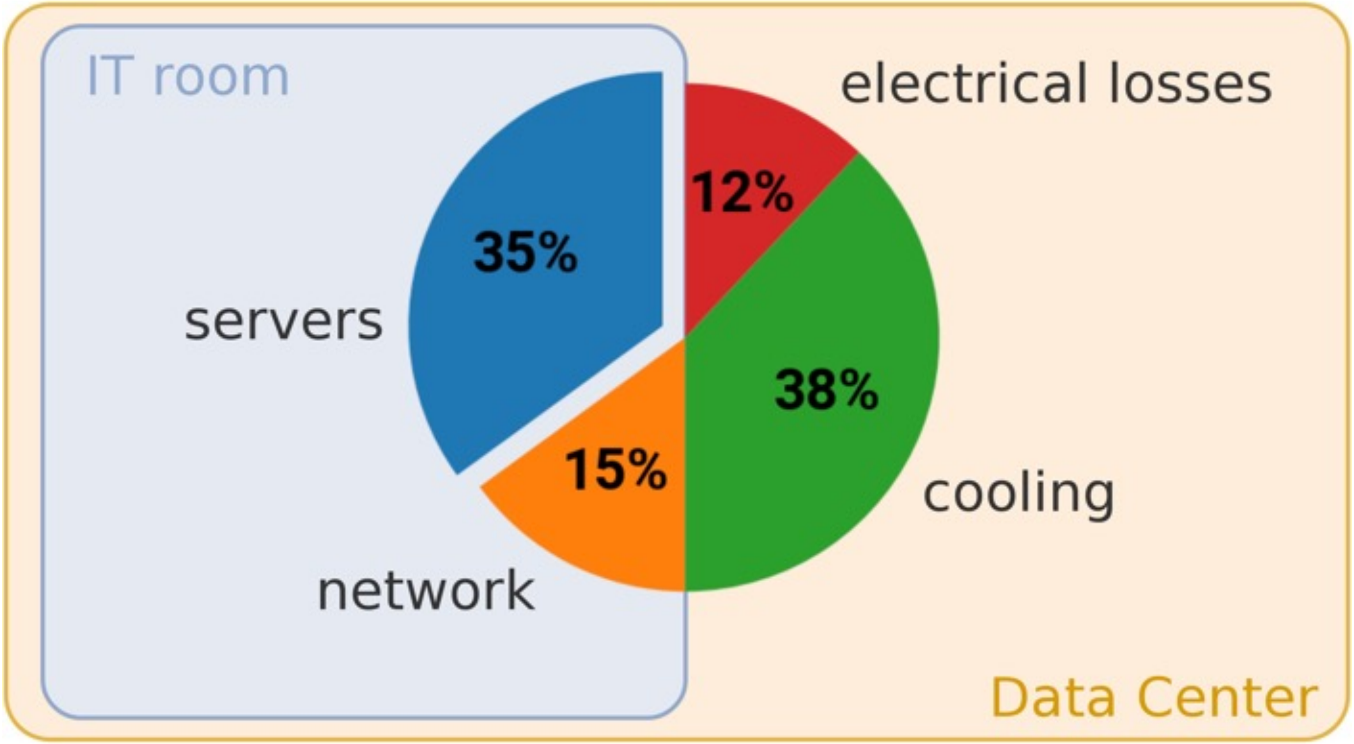
Computing slower?



Temperature matters.



How to measure energy efficiency in DCs?



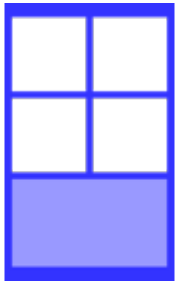
PUE: Power usage effectiveness

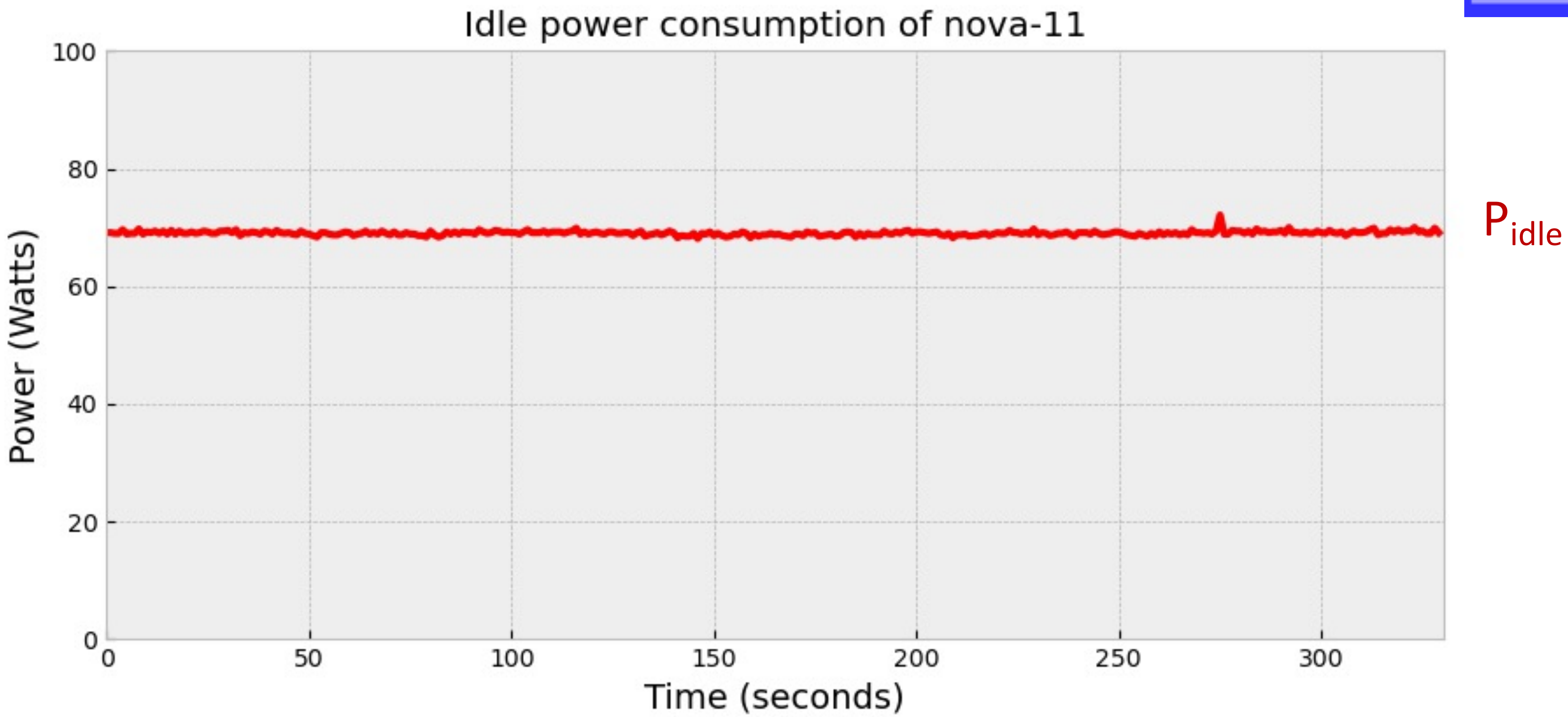
$$PUE = \frac{Total\ Facility\ Power}{IT\ Equipment\ Power}$$

“*Green Grid Data Center Power Efficiency Metrics: PUE and DCIE*”, Green Grid White Paper, 2008.

Wrong idea #1

Idle server consumes nothing or little.



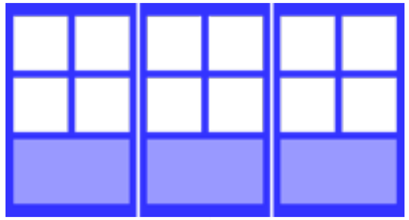


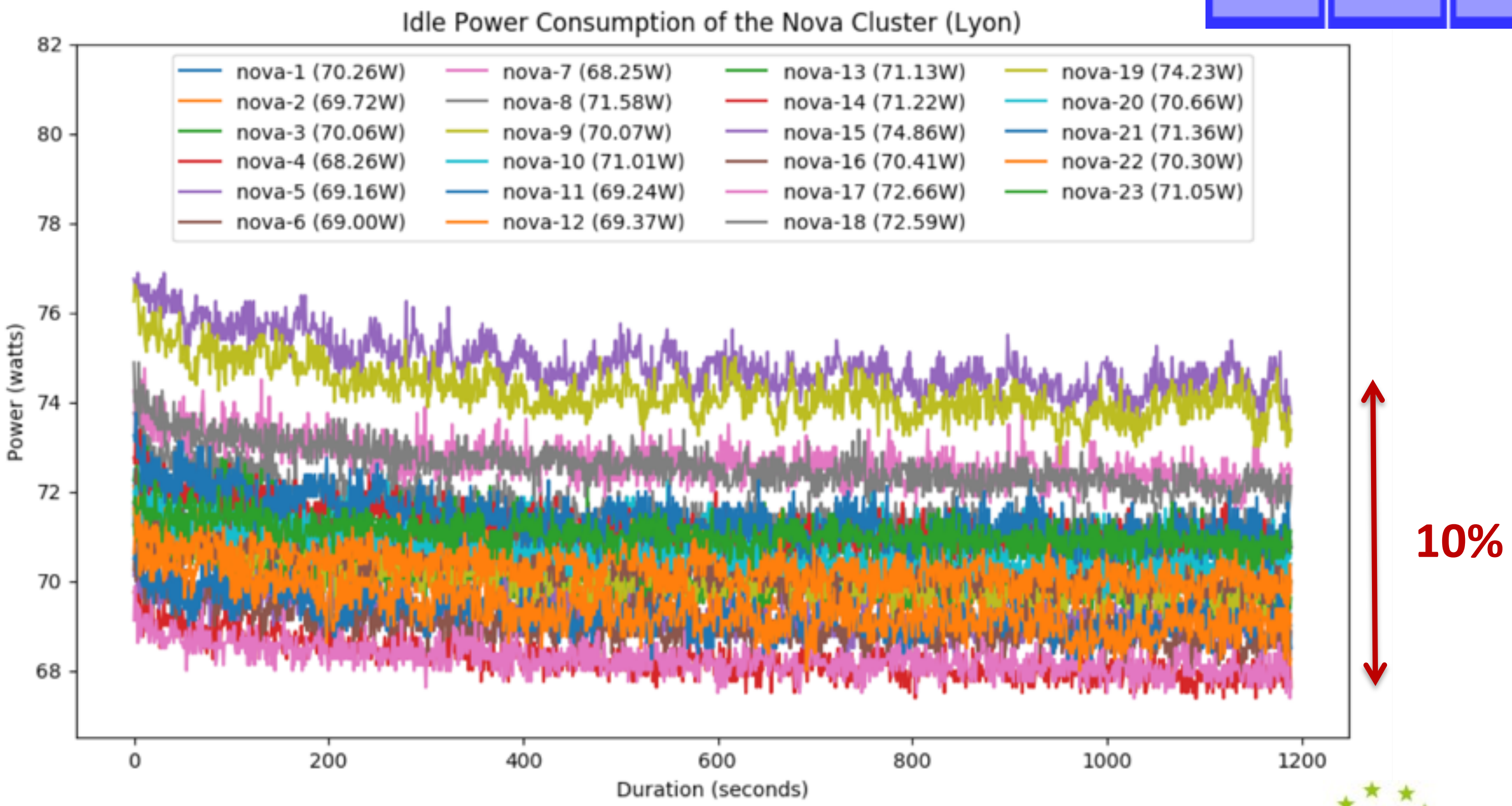
Nova node: 2 x Intel Xeon E5-2620 v4, 8 cores/CPU, 64 GiB RAM, 598 GB HDD (2016)

Grid'5000

Wrong idea #2

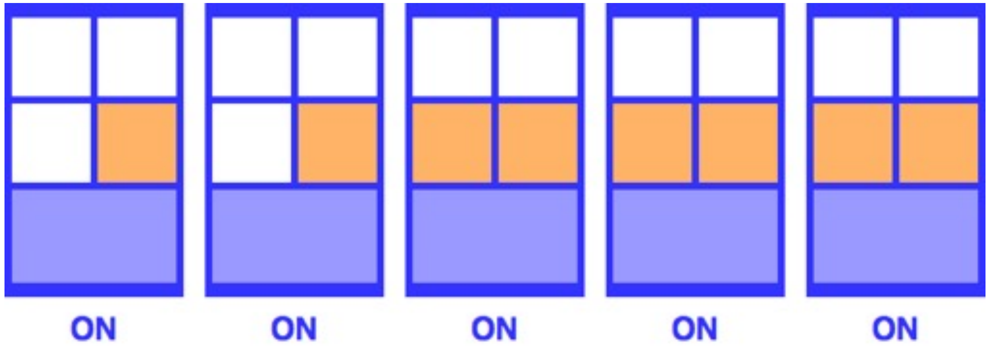
This server model consumes that amount of power.





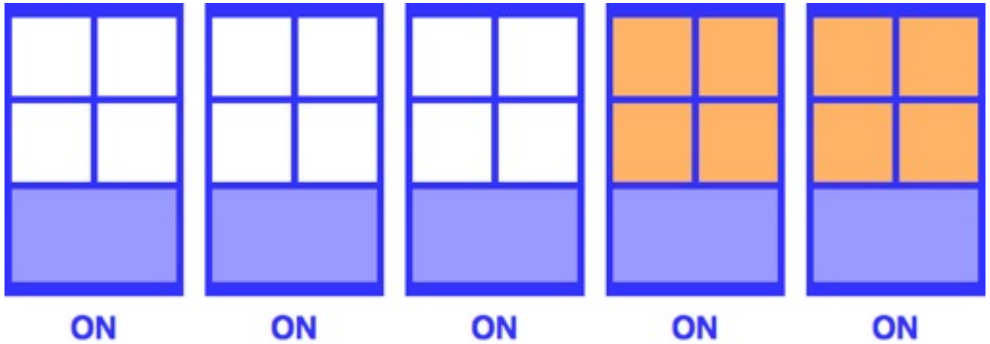
10% difference in idle and more at maximal consumption.

No chance for naive modeling



Naive model:

5 x P_{idle} + 8 x $P_{process}$ = *X* Watts



5 x P_{idle} + 8 x $P_{process}$ = *X* Watts

Best configuration for power consumption ?

It depends.

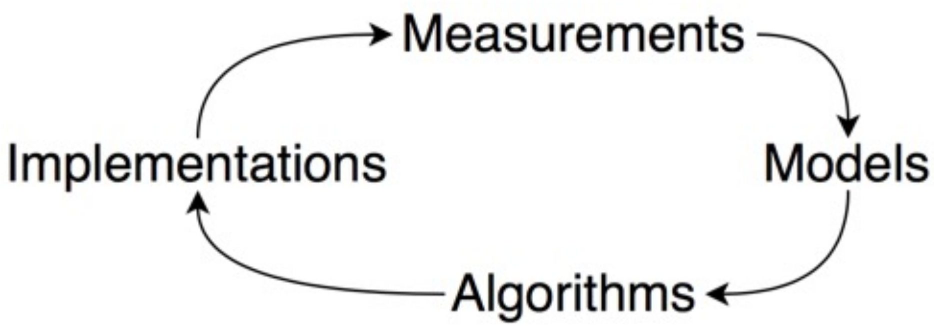
Outline

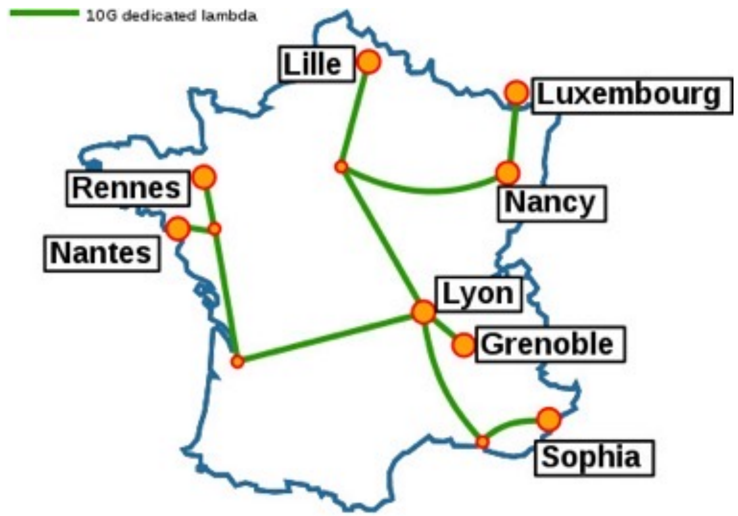
- Context
- Understanding the energy consumption of HPC systems
- Measuring accurately the energy consumption of HPC systems
- Modeling energy consumption of HPC systems
- Concluding broader remarks

Energy consumption: a complex phenomenon

Need for **wattmeters** and sound experimental campaigns

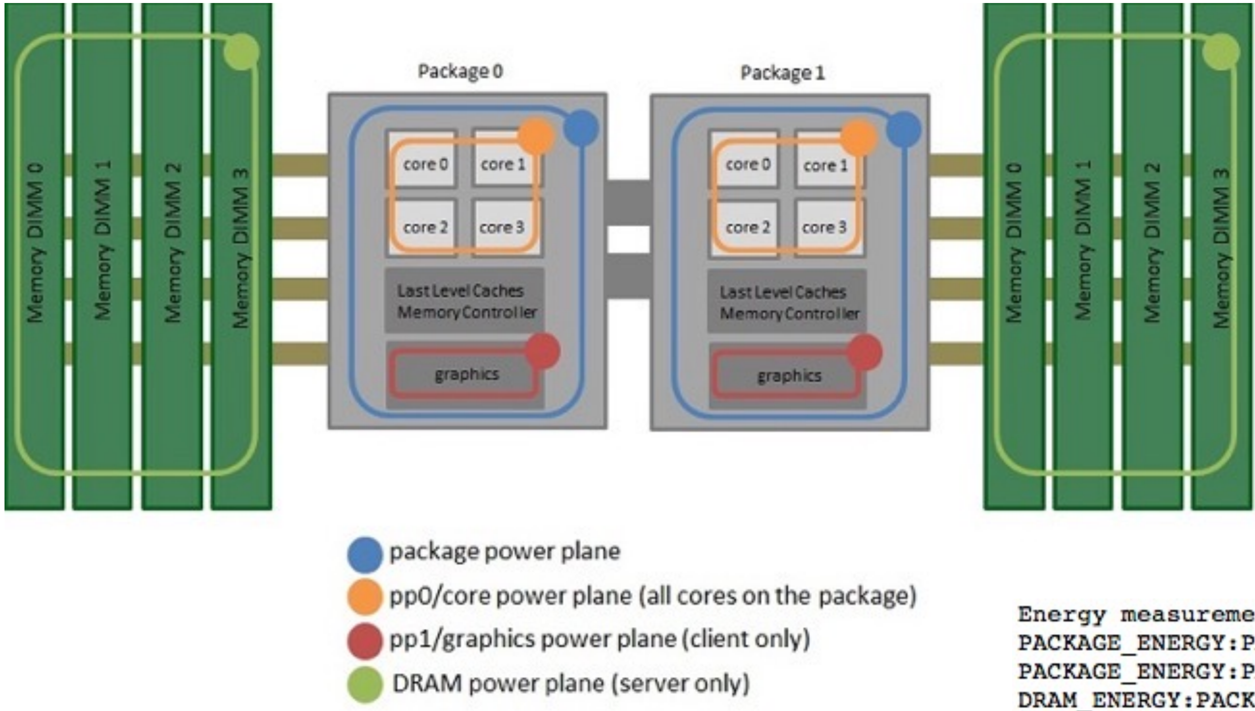
- To understand
- To build robust models
- To get solid instantiations
- To obtain realistic algorithms





Performing measurements

Intel's RAPL (Running Average Power Limit) interface

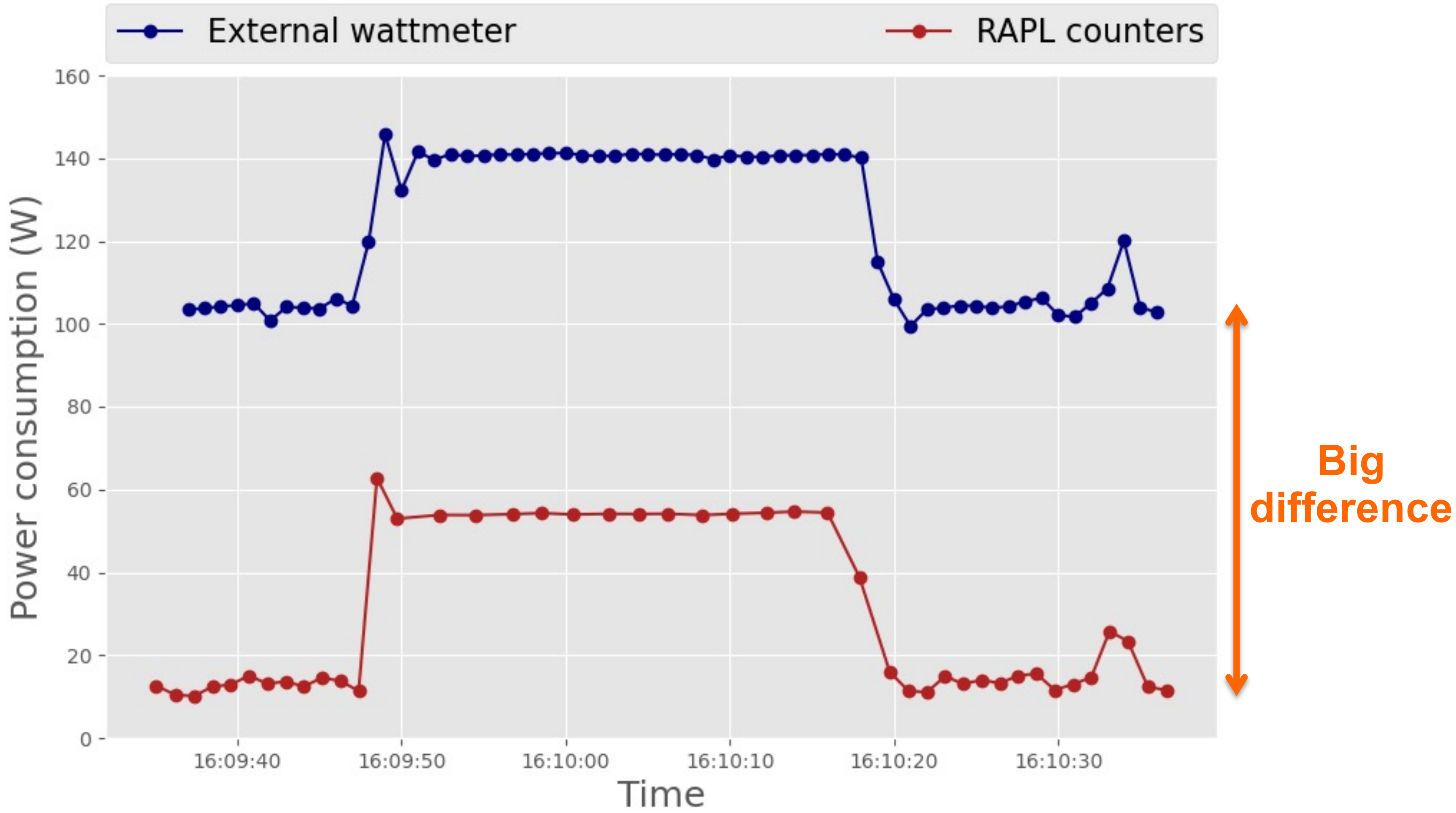



```
Energy measurements:
PACKAGE_ENERGY:PACKAGE0    176.450363J    (Average Power 42.9W)
PACKAGE_ENERGY:PACKAGE1     75.812454J    (Average Power 18.4W)
DRAM_ENERGY:PACKAGE0        11.899246J    (Average Power 2.9W)
DRAM_ENERGY:PACKAGE1         8.341141J    (Average Power 2.0W)
PP0_ENERGY:PACKAGE0        118.029236J    (Average Power 28.7W)
PP0_ENERGY:PACKAGE1         16.759064J    (Average Power 4.1W)
```

Warning: RAPL counters ignore a **large part** of the power consumption of servers.

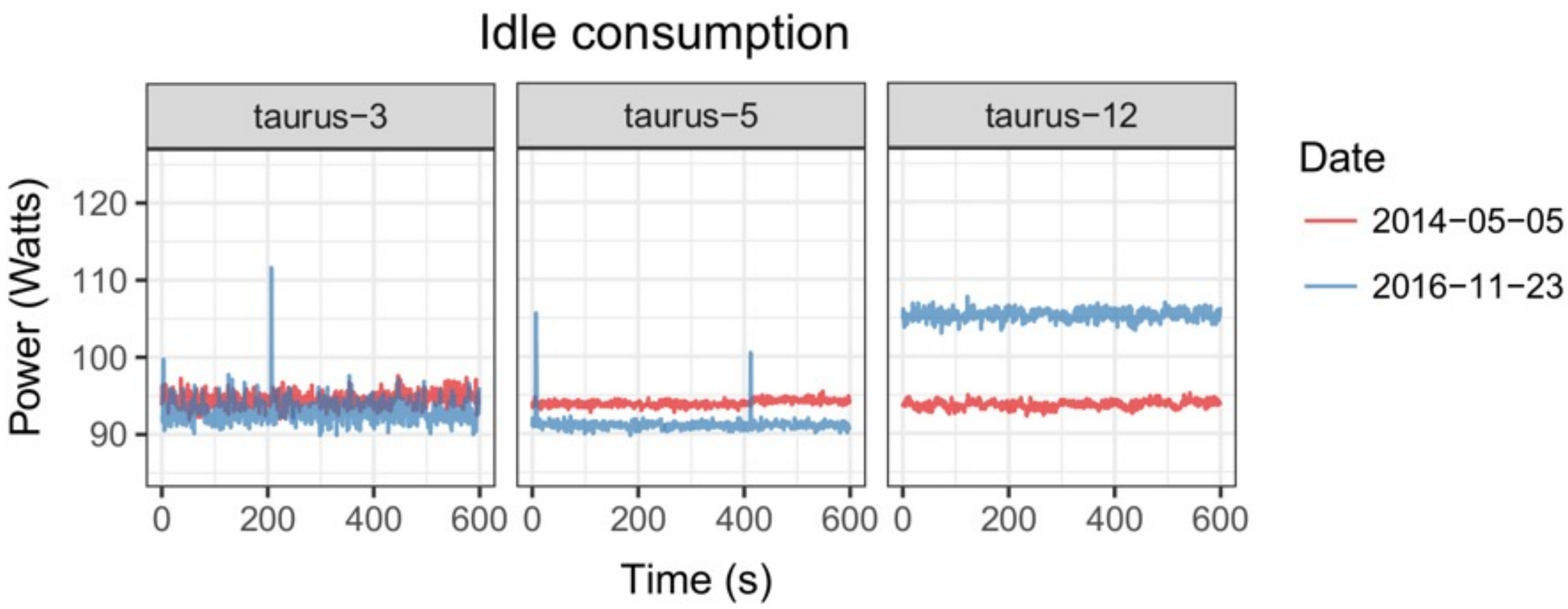
Wrong idea #3

RAPL counters capture most of the power consumption of a server.



Power consumption of Taurus-12

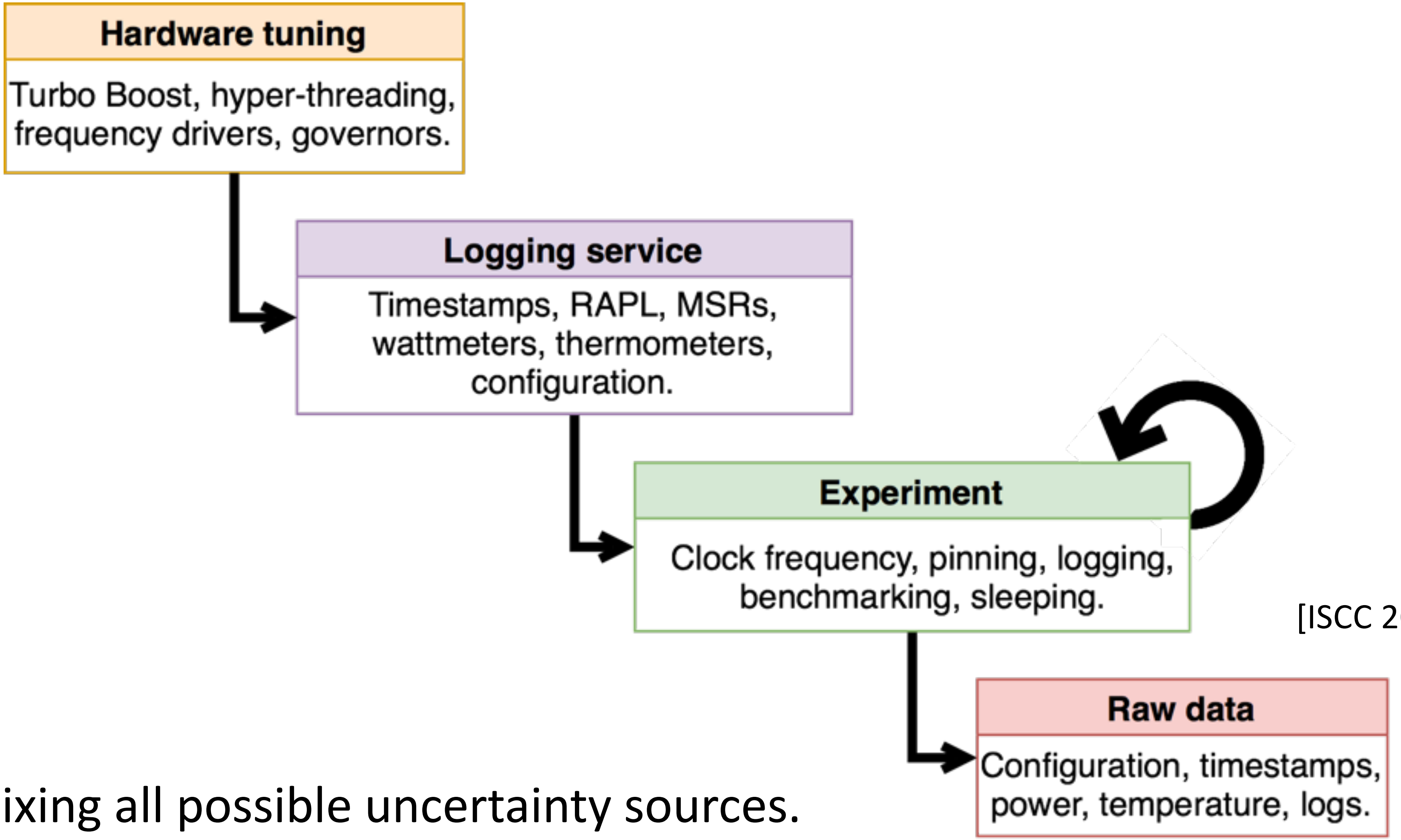
Reproducibility?



[Cluster 2017]

Idle power consumption varies over time.

Methodology for measuring server consumption

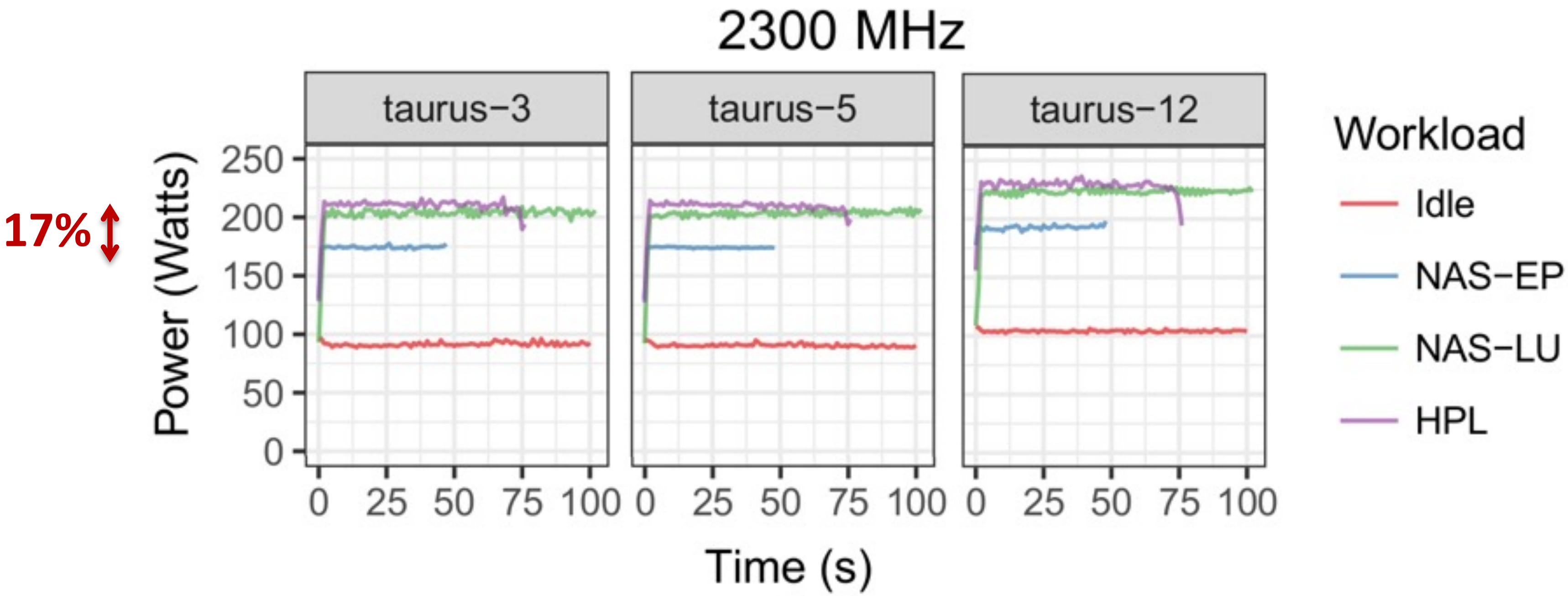


[ISCC 2021]

Fixing all possible uncertainty sources.

Wrong idea #4

The relation between power and CPU load is linear/quadratic/cubic.

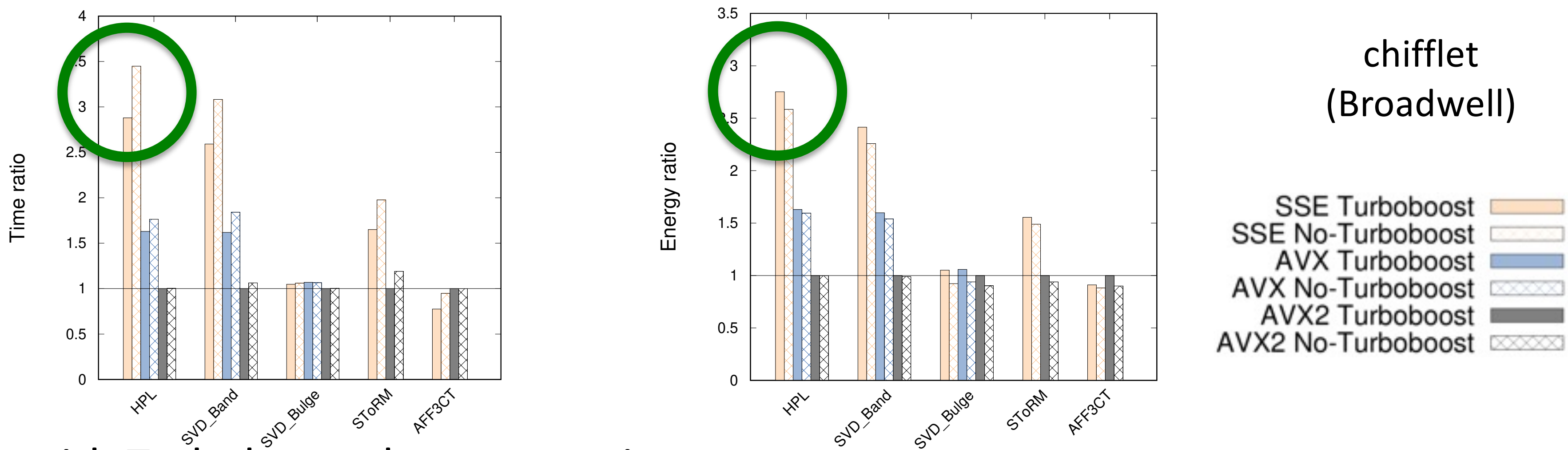


[Cluster 2017]

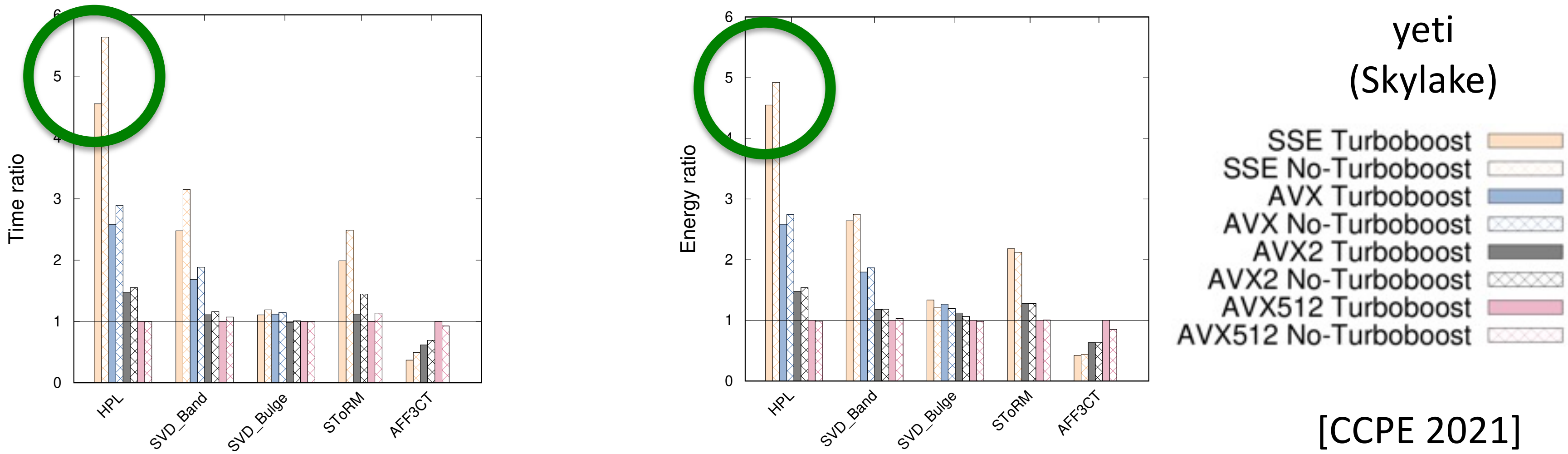
17% difference in consumption for applications fully loading the server.

Wrong idea #5

For a given application, there is a least consuming configuration.



Faster with Turboboost, but consuming more energy.

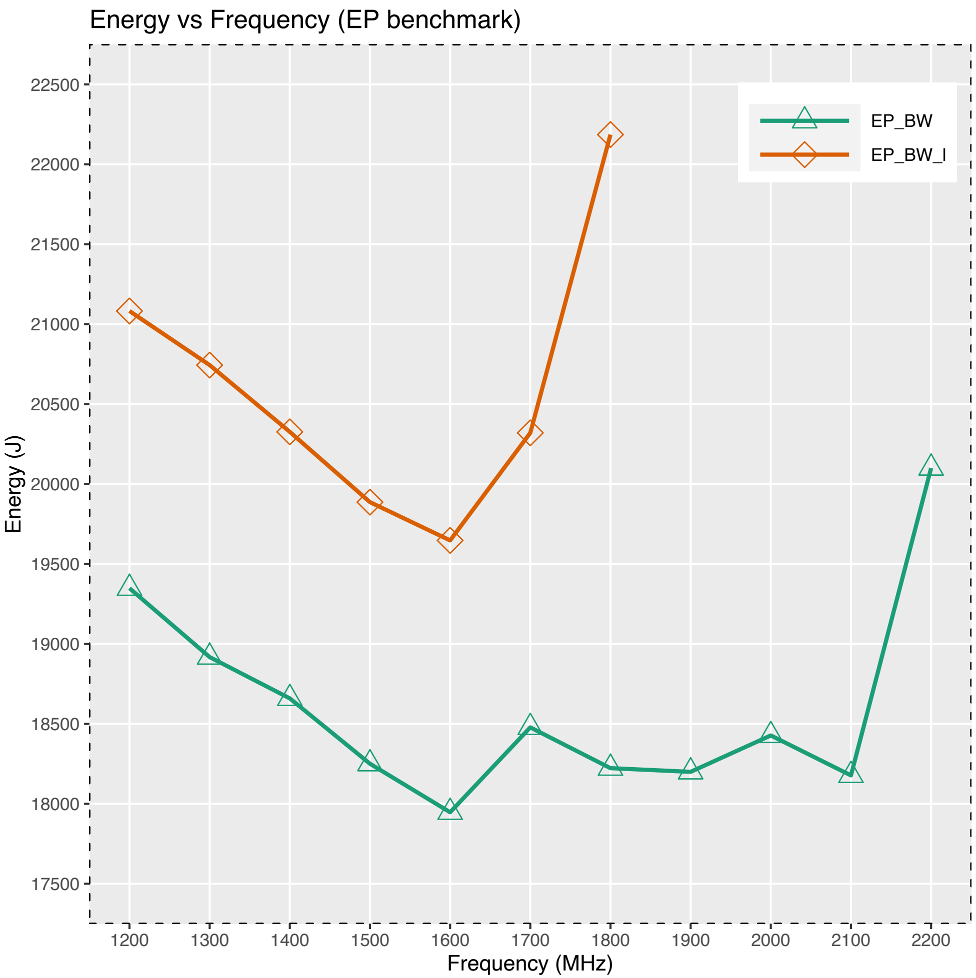


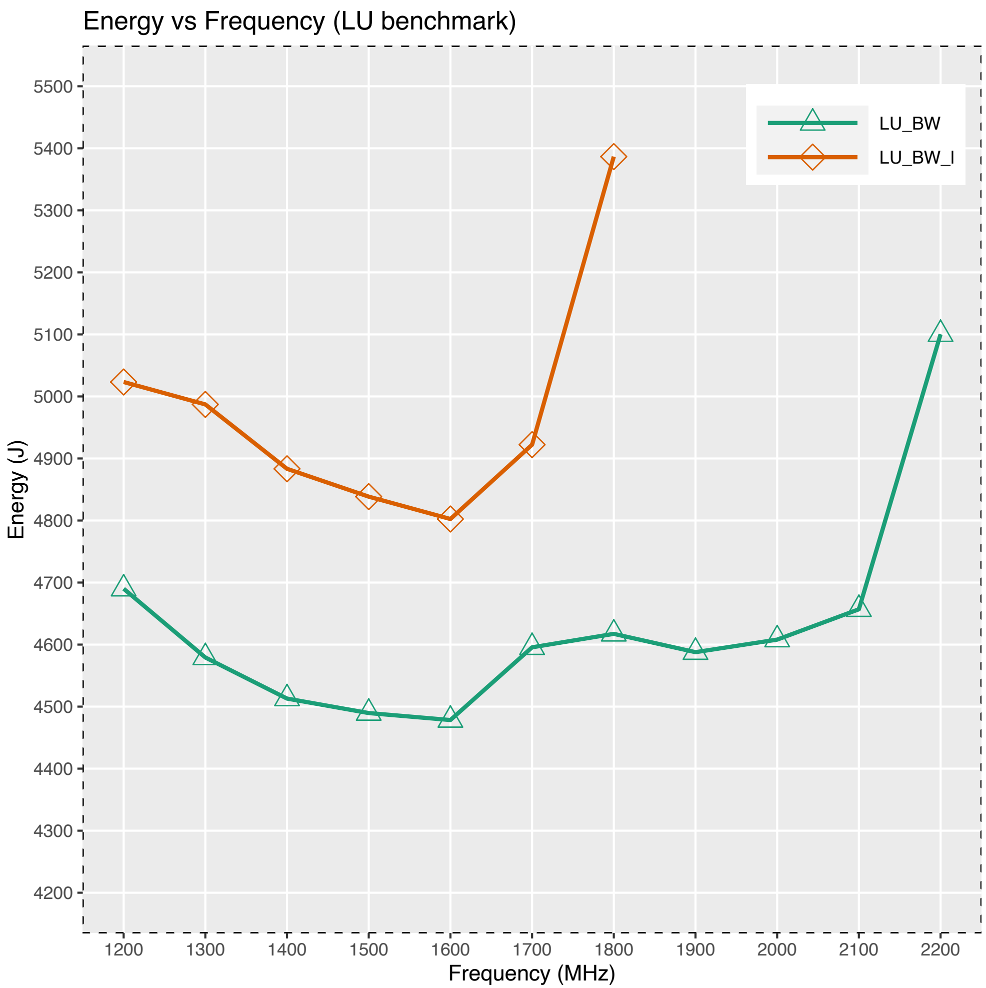
[CCPE 2021]

Faster with Turboboost, and comsuming less energy.

Wrong idea #6

Low power processors consume less energy.

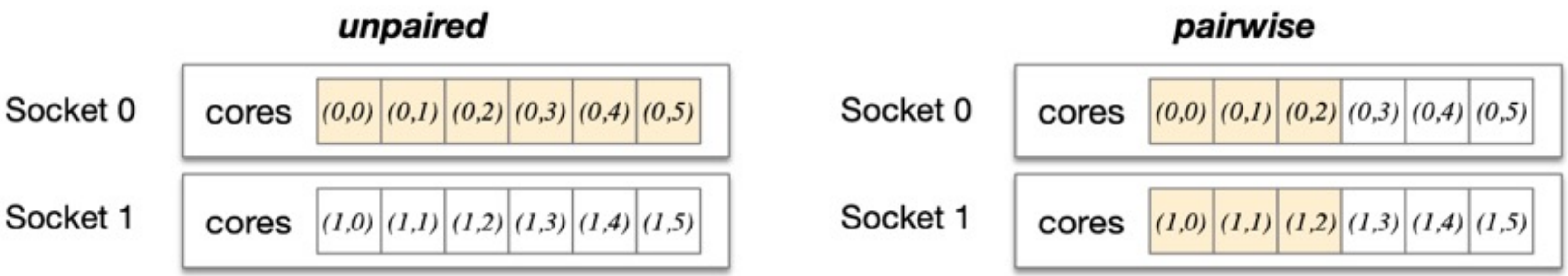


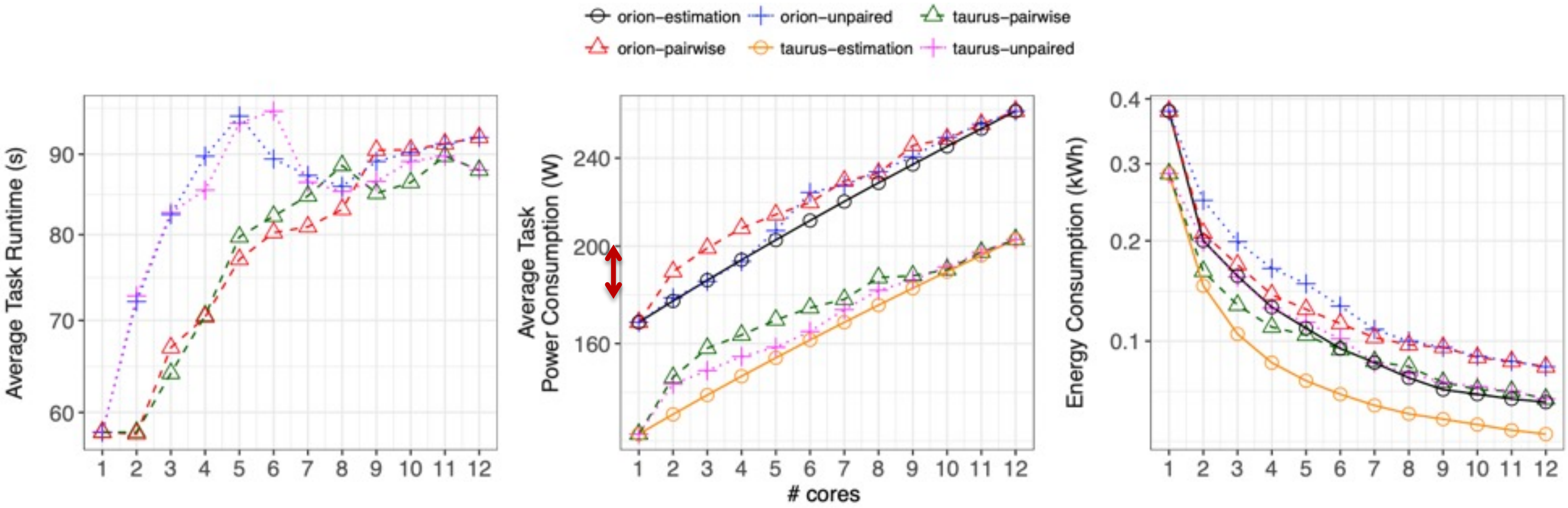


BW_l: Xeon E5-2630L v4 (Broadwell) -> low power processor (orange)
BW: Xeon E5-2630 v4 (Broadwell) (green)

[ISCC 2021]

Process placement onto cores





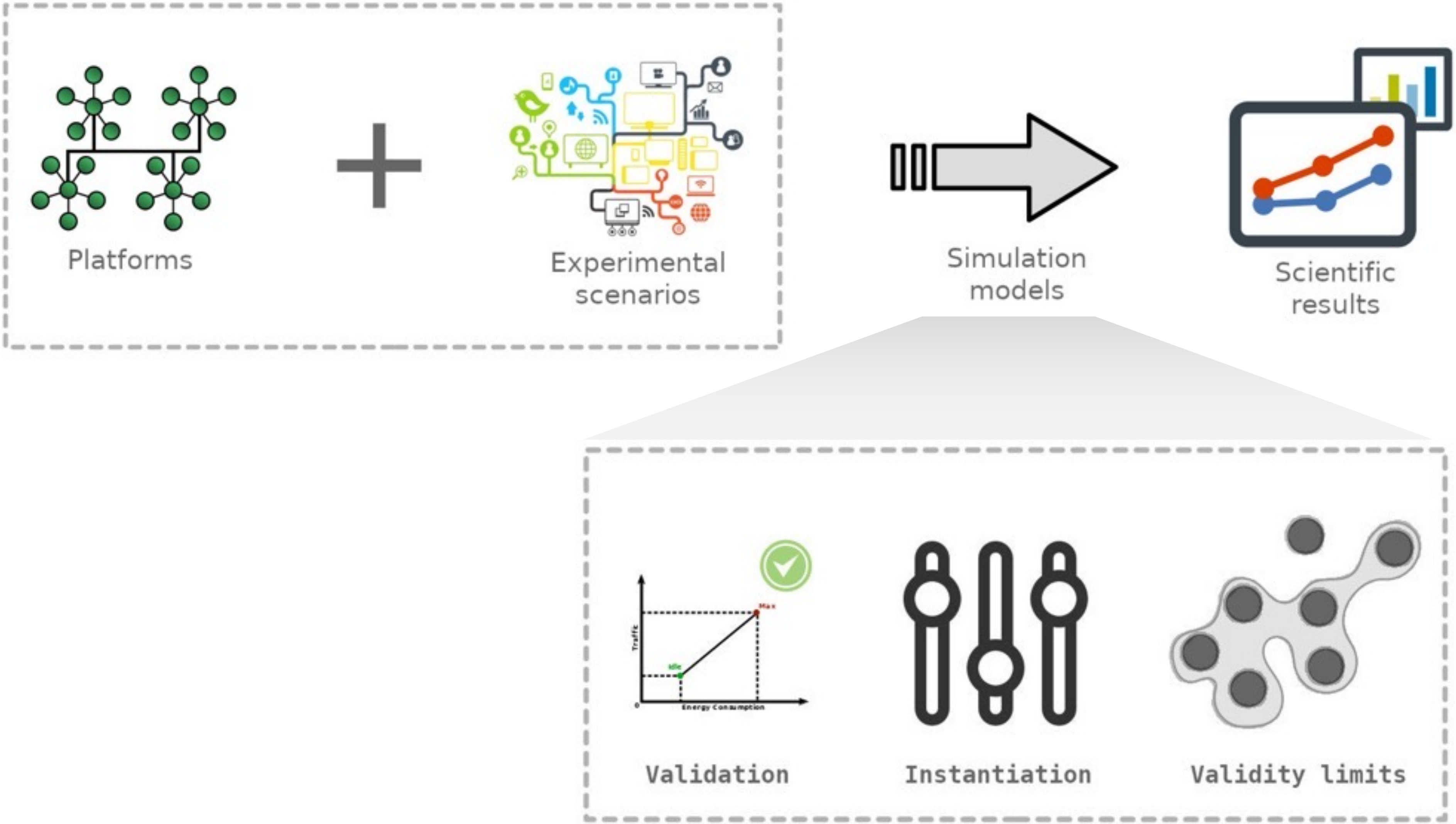
[JOCS 2020]

Up to **8%** difference in average power consumption between unpaired and pairwise.

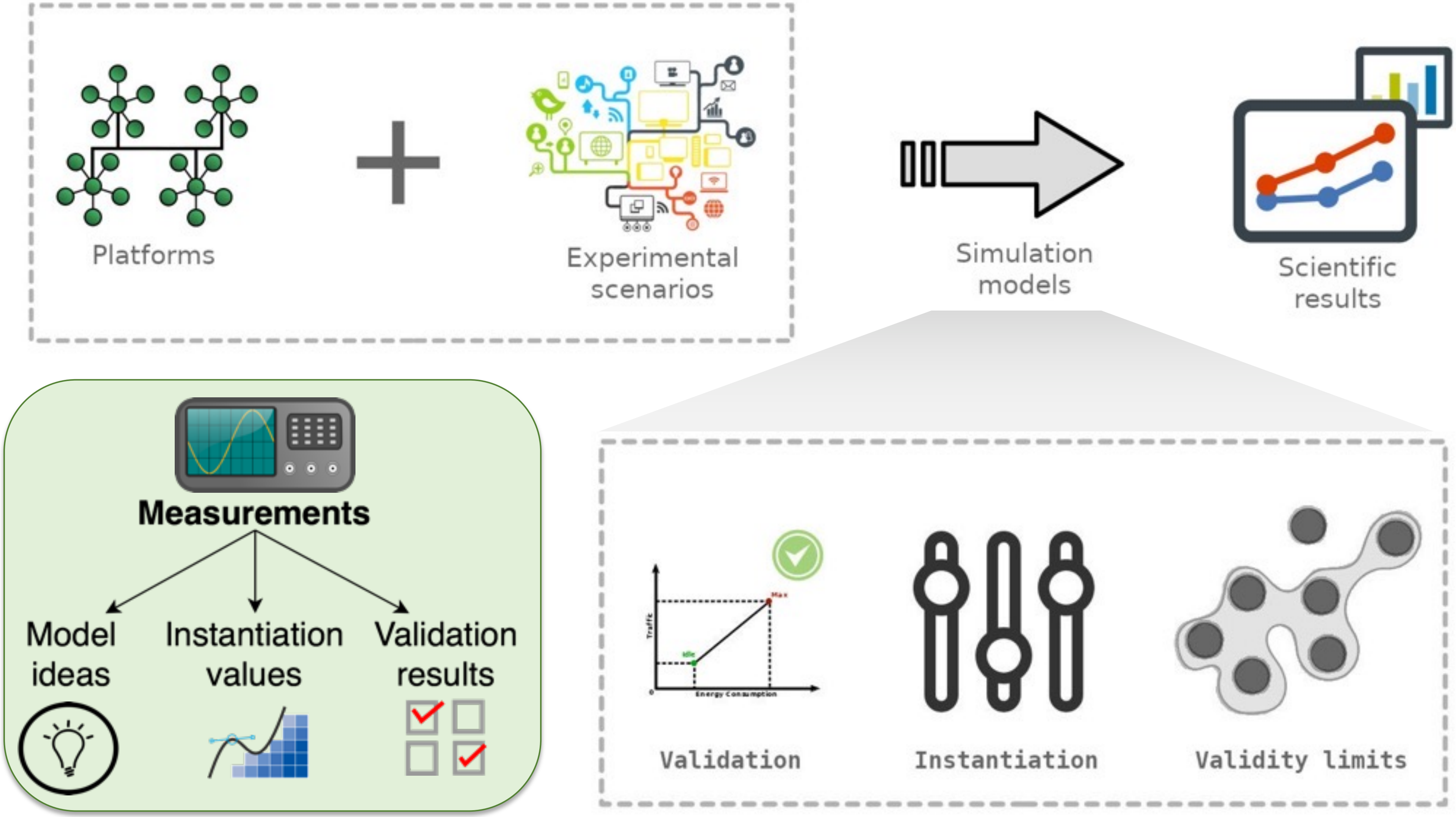
Outline

- Context
- Understanding the energy consumption of HPC systems
- Measuring accurately the energy consumption of HPC systems
- Modeling energy consumption of HPC systems
- Concluding broader remarks

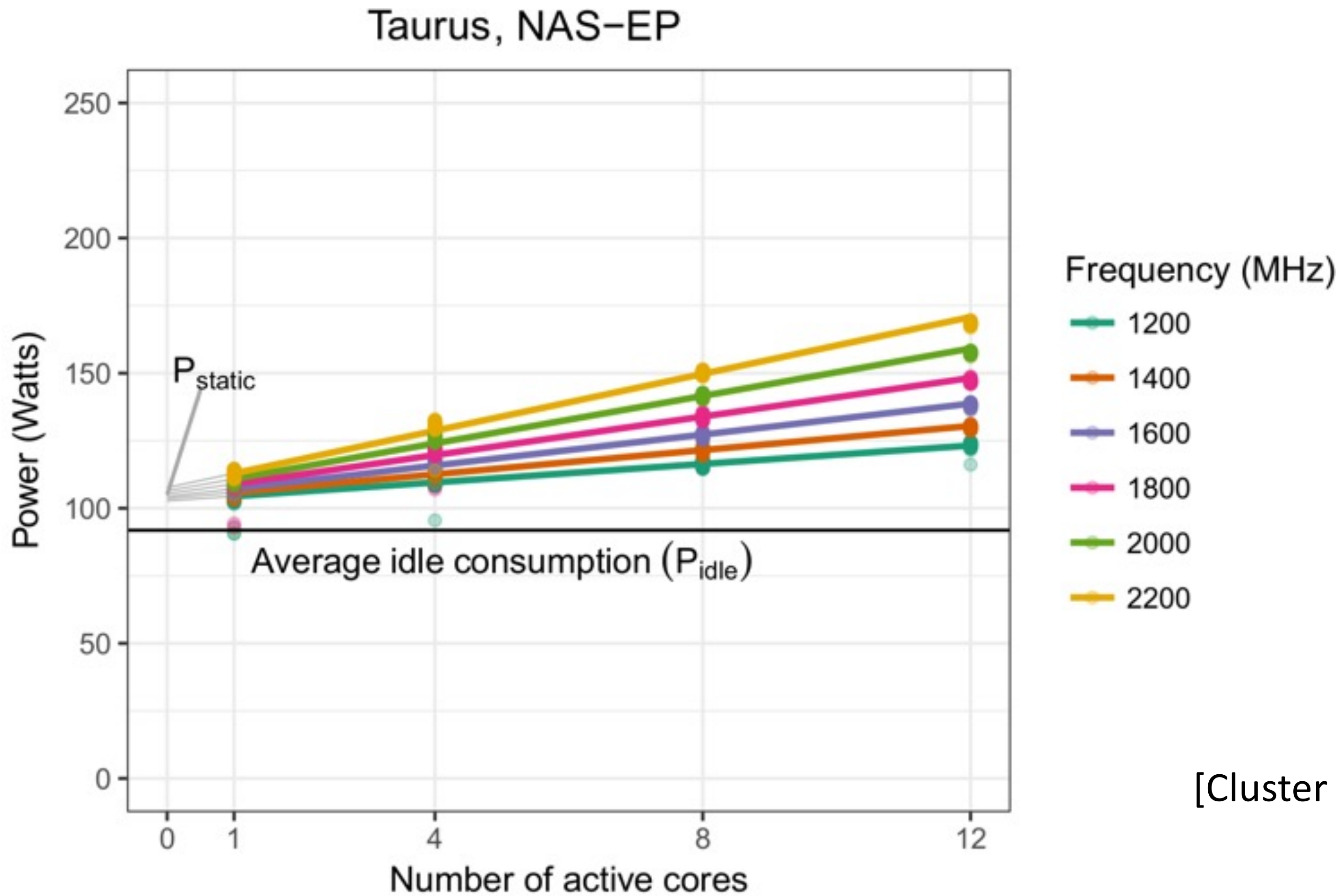
Simulating energy consumption



Simulating energy consumption



Server profiling



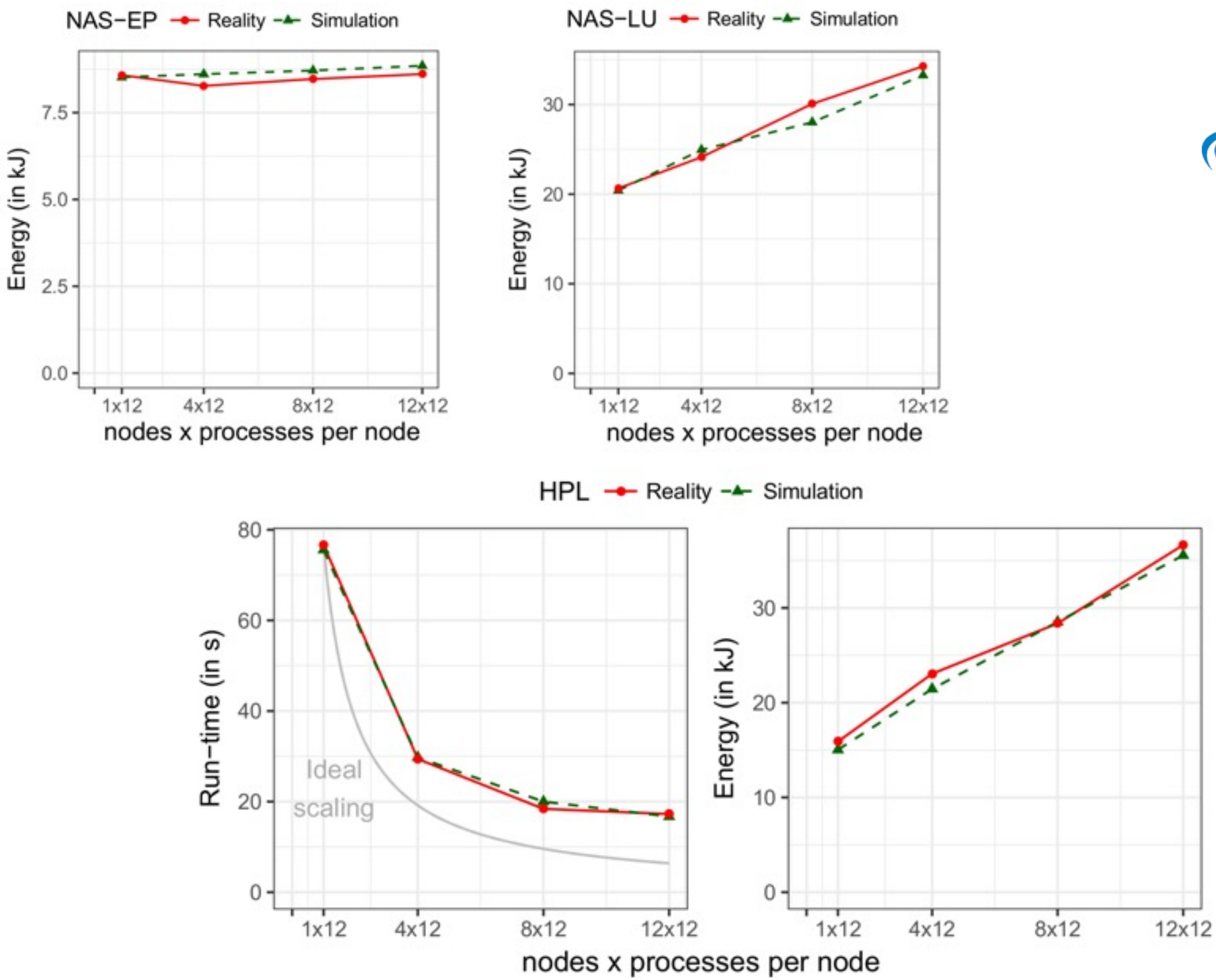
[Cluster 2017]

To do for each computing kernel.

At each frequency.

And each time we want to compare the model to real life.

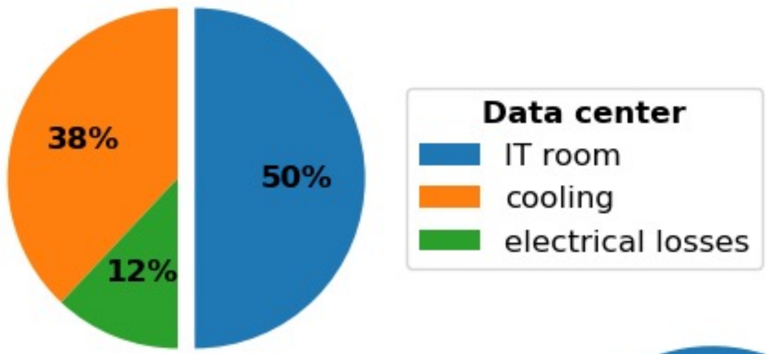
Simulating server clusters



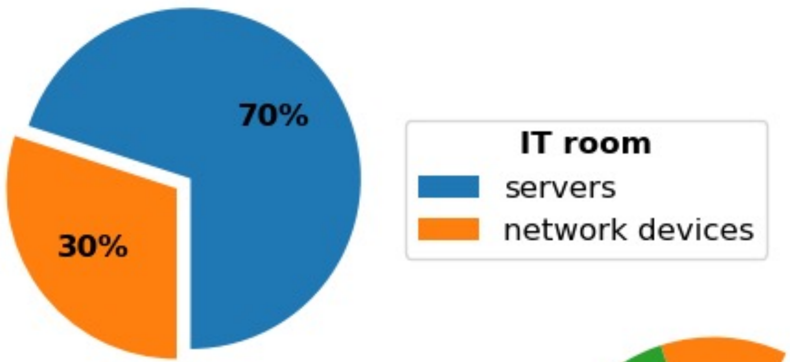
[Cluster 2017]

Reproducible results: https://gitlab.inria.fr/fheinric/paper-simgrid-energy

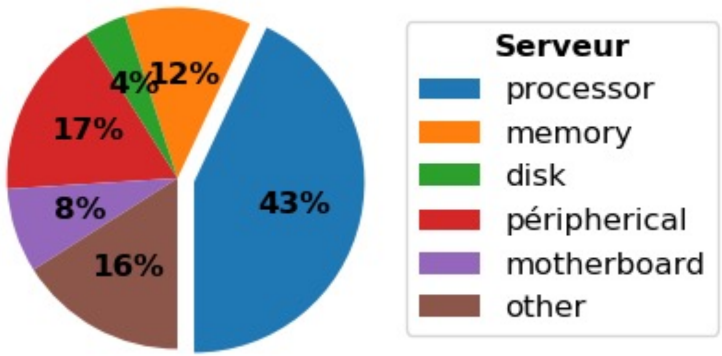
Wasted energy at all levels of data centers



Cooling
Power generators
Batteries
…



Unused servers
Overprovisioning
Redundancy
…



Power non-proportionality
Dark silicon
Unused components
…

Models and simulation tools for what?

Capacity and energy planing

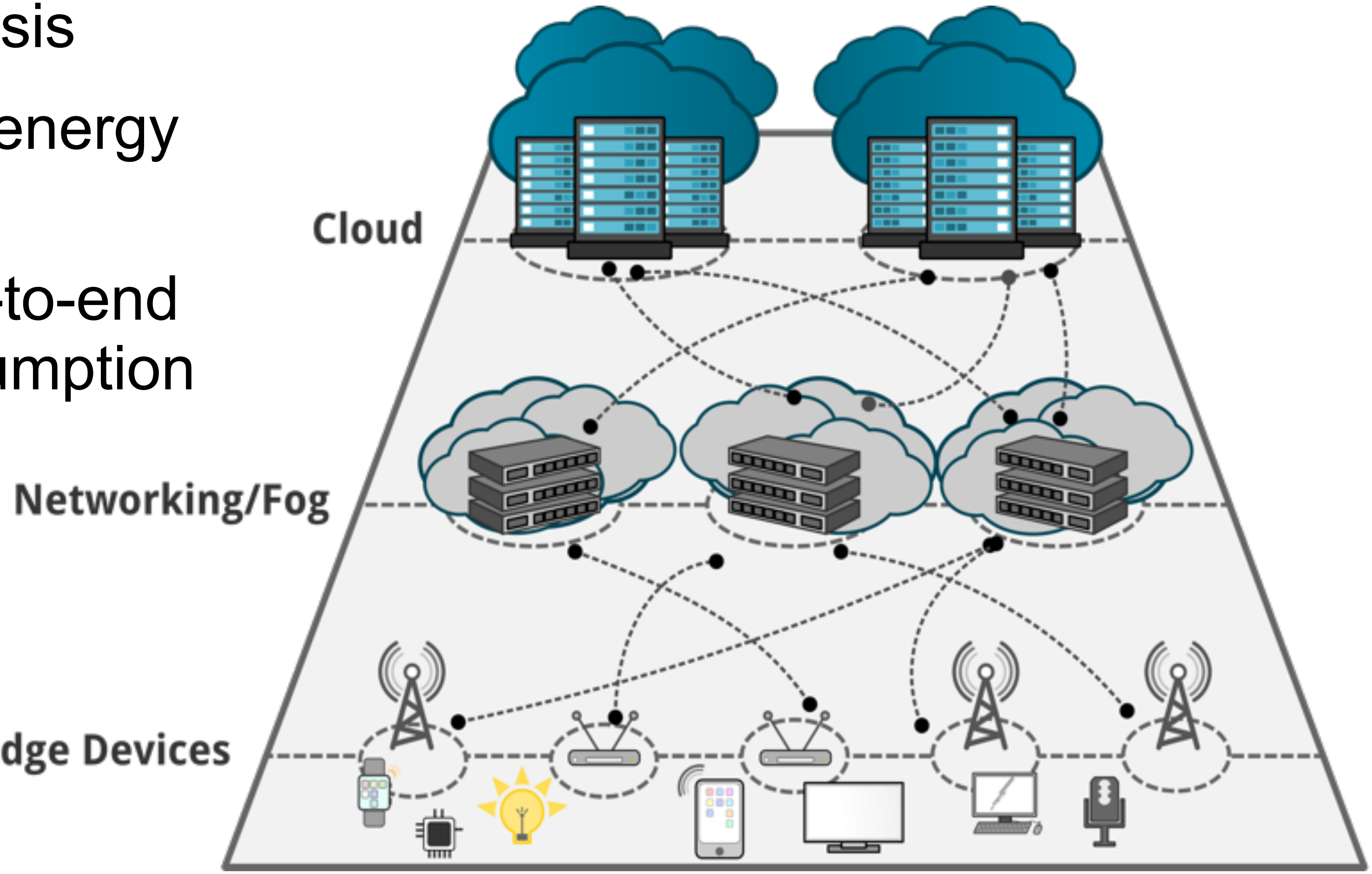
What-if scenarios

Algorithm analysis

Estimating VM energy consumption

Estimating end-to-end energy consumption

Closing doors

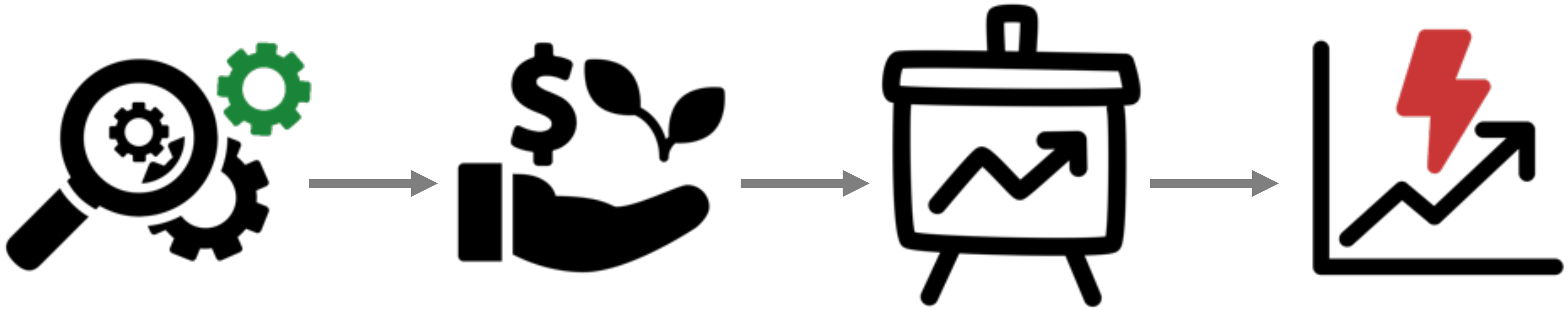


Outline

- Context
- Understanding the energy consumption of HPC systems
- Measuring accurately the energy consumption of HPC systems
- Modeling energy consumption of HPC systems
- Concluding broader remarks

Increasing energy efficiency
$\neq$ reducing consumption

Increasing energy efficiency ≠ reducing consumption



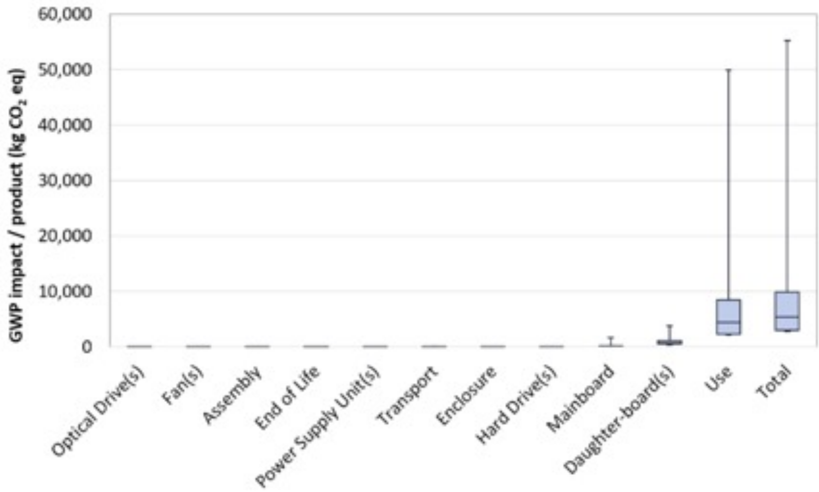
Energy **optimization** → Resource cost reduction → Usage increase → Energy consumption **increase**

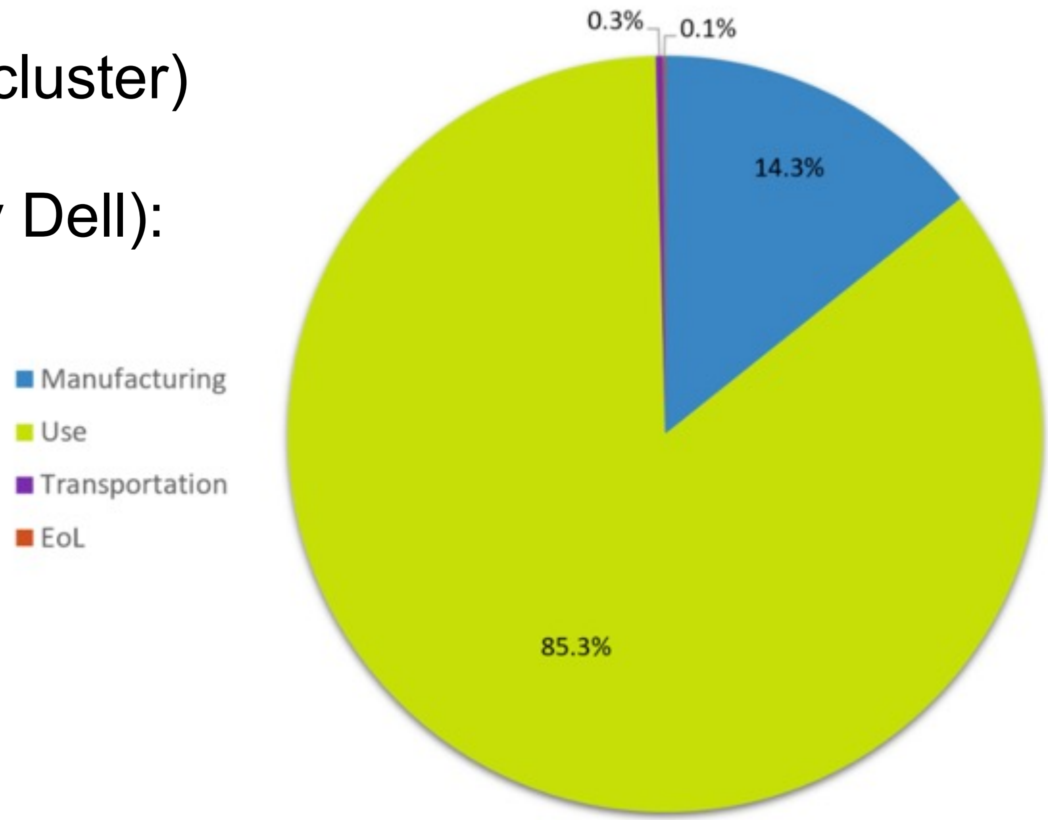
Beware of rebound effects!

Full life cycle of servers

Dell PowerEdge R430 (Nova cluster)

Estimated carbon footprint (by Dell):
8,150 kgCO2e



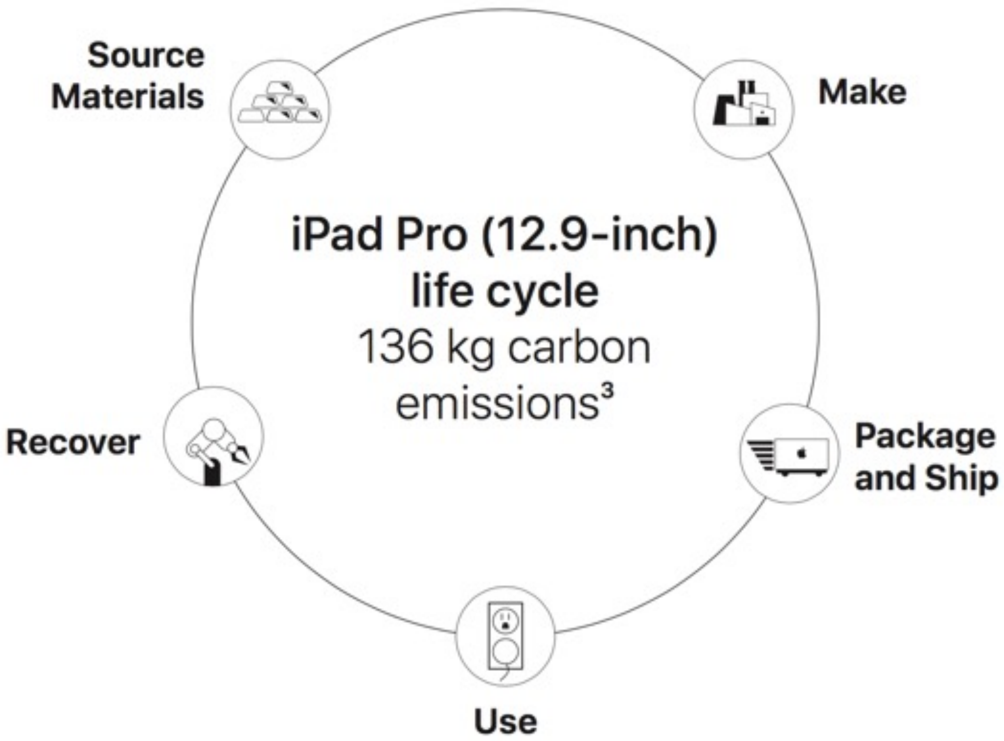


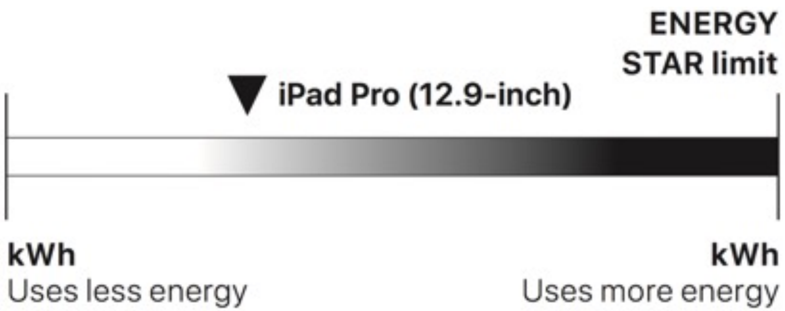
Assumptions for calculating product carbon footprint:

Product Weight	26.3 kg	**Server Type**	Rack	**Assembly Location**	EU
Product Lifetime	4 years	**Use Location**	EU	**Energy Demand (Yearly TEC)**	1760.3 kWh
HDD/SSD Quantity	x2 1TB 3.5" HDD	**DRAM Capacity**	16GB	**CPU Quantity**	2

Source: Dell PowerEdge R430 carbon footprint, 2019.

Life cycle of end devices





iPad Pro (12.9-inch) life cycle carbon emissions

83%	Production
11%	Transport
6%	Use
<1%	End-of-life processing

4 years of use

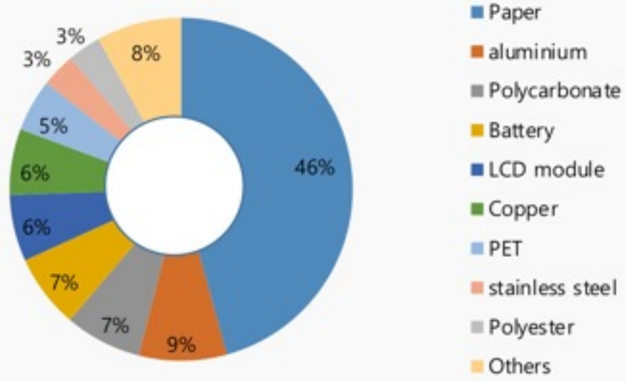
Source: Product environmental report, Apple, 2018.

Numerous other environmental impacts

Product Features

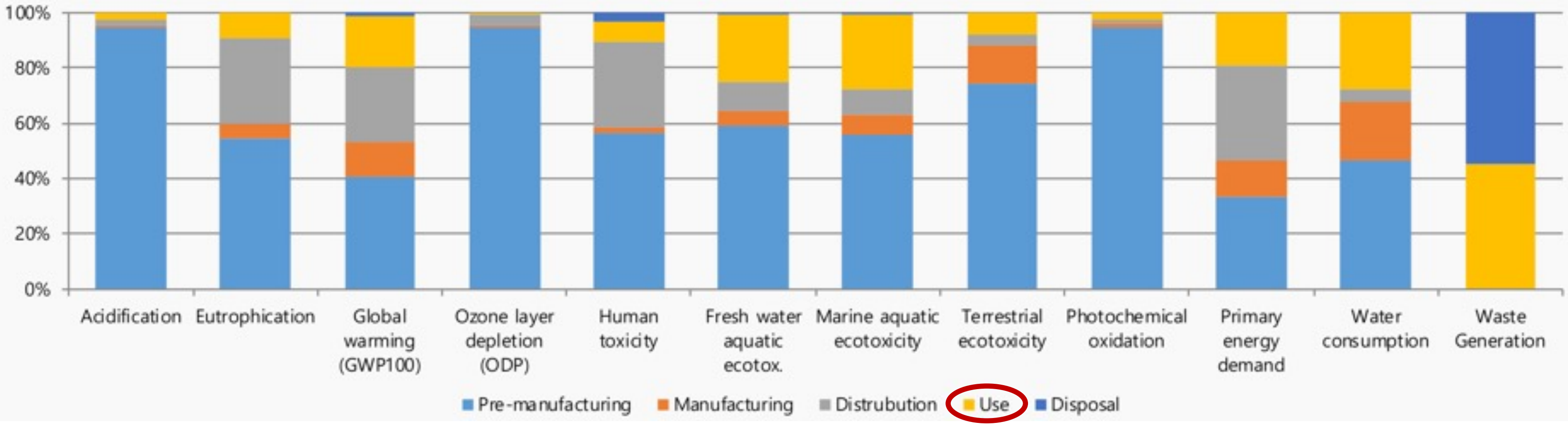
Model name	SM-N950U (Galaxy Note8)
Processor	Qualcomm 2.35GHz, 1.9GHz Octa-Core 64bit
Dimension	162.5 x 74.8 x 8.6 mm
Display	6.3" 2960 x 1440, 16M In-Cell Touch LCD
Battery	Li-Ion 3300 mAh
Camera	12 MP / 5MP
Wt.(g)	186.34g

Material Use



Characterized Environment Impact

Source: Life Cycle Assessment for Mobile Products, Samsung, 2018.



Standard	ISO 14040:2006 and 14044:2006
Database	Ecoinvent 2.2
Method for impact assessment	Life cycle impact assessment classification and characterization factors according to CML 2001 as provided in the SimaPro 7.1.5 LCA tool
LCA software	SimaPro 7.1.5

Pre-manufacturing	Parts and materials constituting the products and its transportation (from supplier to Samsung factory)
Manufacturing	Product assembly by Samsung Electronics (Data collection period : 3 months ahead of assessment)
Distribution	From China or Vietnam to United States
Usage	2 years use
Disposal	Waste treatment of parts and material

Thank you for your attention

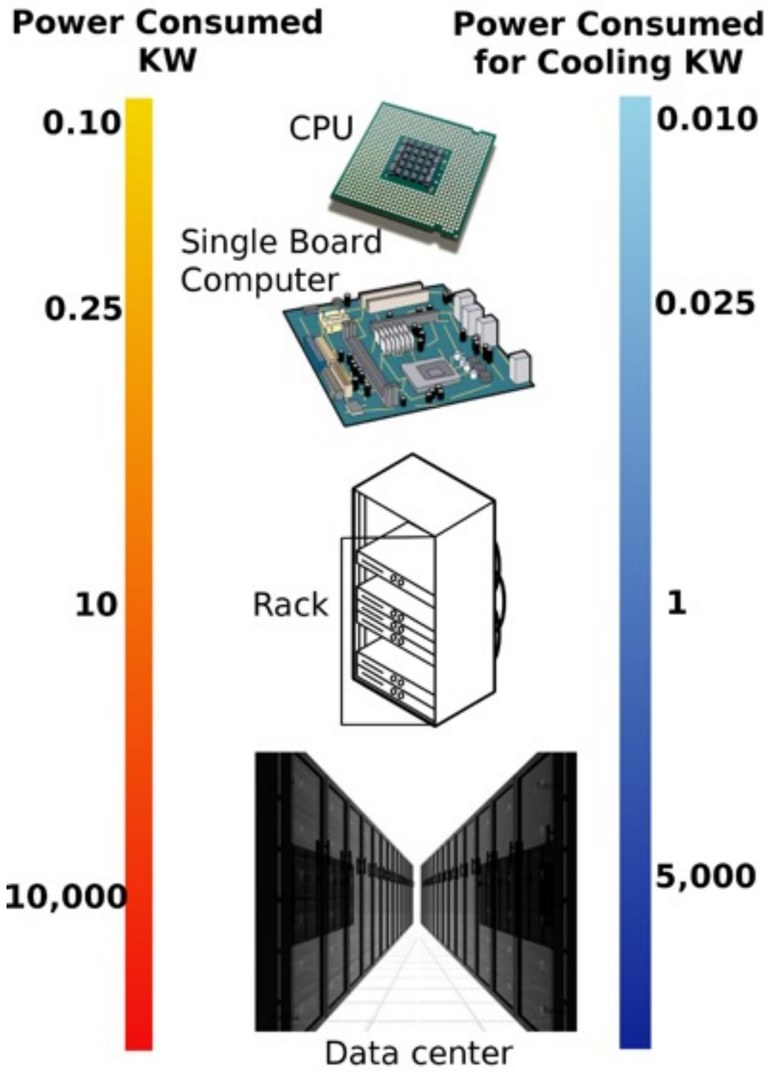
http://people.irisa.fr/Anne-Cecile.Orgerie

Citations

- [Cluster 2017] "**Predicting the Energy-Consumption of MPI Applications at Scale Using Only a Single Node**", F. C. Heinrich, T. Cornebize, A. Degomme, A. Legrand, A. Carpen-Amarie, S. Hunold, A.-C. Orgerie and M. Quinson, *IEEE Cluster Conference*, pages 92-102, September 2017.
- [ISCC 2021] **"Experimental Workflow for Energy and Temperature Profiling on HPC Systems",** K. Rao Vaddina, L. Lefèvre and A.-C. Orgerie, *IEEE Symposium on Computers and Communications*, pages 1-9, September 2021.
- [CCPE 2021] **"Thermal design power and vectorized instructions behavior",** A. Guermouche and A.-C. Orgerie, *Concurrency and Computation: Practice and Experience (CCPE)*, pages 1-18, March 2021.
- [ICCS 2019] **"Accurately Simulating Energy Consumption of I/O-intensive Scientific Workflows",** R. Ferreira da Silva, A.-C. Orgerie, H. Casanova, R. Tanaka, E. Deelman and F. Suter, *International Conference on Computational Science*, pages 138-152, June 2019.
- [JOCS 2020] **"Characterizing, Modeling, and Accurately Simulating Power and Energy Consumption of I/O-intensive Scientific Workflows",** R. Ferreira da Silva, H. Casanova, A.-C. Orgerie, R. Tanaka, E. Deelman and F. Suter, *Journal of Computational Science*, volume 44, pages 1-14, July 2020.
- [FGCS 2018] **"End-to-end Energy Models for Edge Cloud-based IoT Platforms: Application to Data Stream Analysis in IoT",** Y. Li, A.-C. Orgerie, I. Rodero, B. Lemma Amersho, M. Parashar and J.-M. Menaud, *Future Generation Computer Systems*, Elsevier, volume 87, pages 667-678, October 2018.
- [SUSCOM 2018a] **"An experiment-driven energy consumption model for virtual machine management systems",** M. Callau-Zori, L. Samoila, A.-C. Orgerie and G. Pierre, *Sustainable Computing: Informatics and Systems*, Elsevier, volume 18, pages 163-174, June 2018.
- [SUSCOM 2018b] **"Energy-proportional Profiling and Accounting in Heterogeneous Virtualized Environments",** M. Kurpicz, A.-C. Orgerie, A. Sobe and P. Felber, *Sustainable Computing: Informatics and Systems*, Elsevier, volume 18, pages 175-185, June 2018.

http://people.irisa.fr/Anne-Cecile.Orgerie/publis.html

Saving energy



Low power processors (big.LITTLE)

Multi-core architectures

Energy-efficient dedicated
architectures (FPGA, **GPU**)

Dynamic Voltage Frequency Scaling

Workload consolidation techniques

On/off policies

Hot spot management

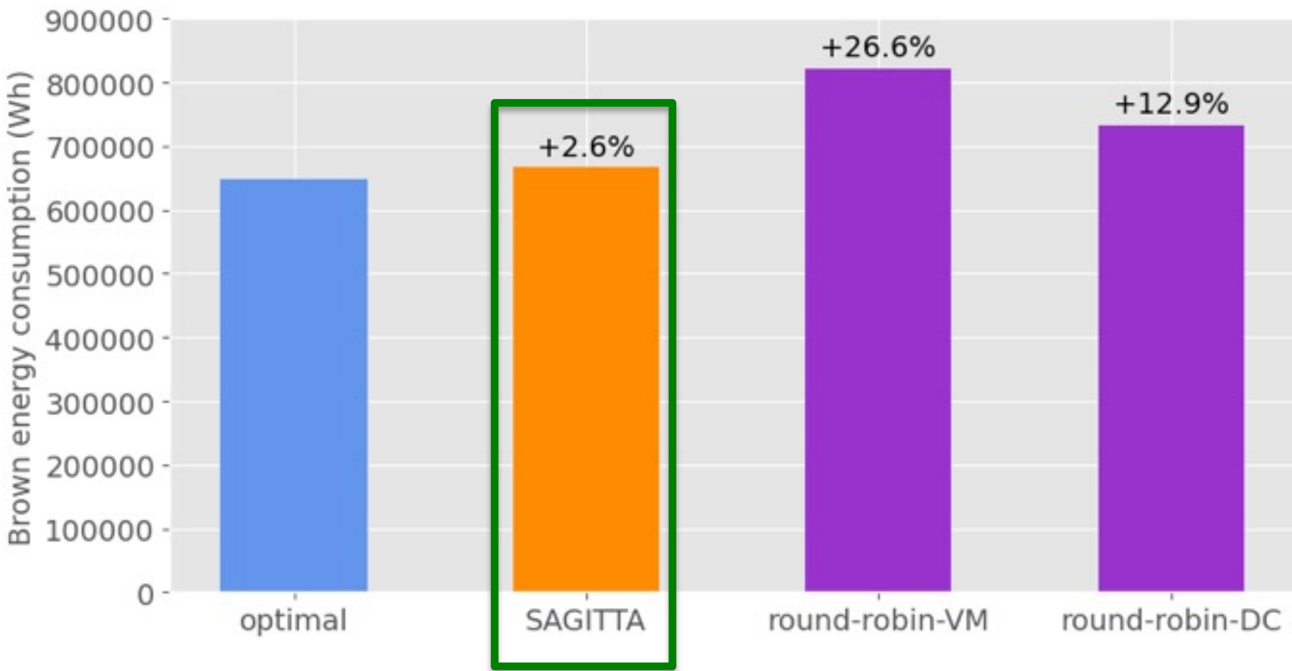
Workload peak reduction

Dynamic adaptation

Designing energy efficient algorithms

5 DCs with 20 homogeneous servers each, no migration

Optimal solution (dynamic programming algorithm) => 2 weeks of computation on 30 Grid'5000 servers



SAGITTA is close to the optimal solution.

[Chapter2018]

VM migration algorithm

1. **Pre-allocation**: incoming VM requests

   `Best-fit`

   […]

   `Expected value`

2. **Migration**: moving running VMs between DCs with network constraints

   a. Evaluate energy costs (VM migrations) and gains (expected remaining green energy on DCs)

   `Expected value`

   `Best-fit`

   b. Schedule the VM migrations between DCs

   `Dichotomy`

3. **Consolidation**: packing VMs inside DCs

4. **Allocation**: actually send the commands to the servers

   a. Switch ON/OFF servers

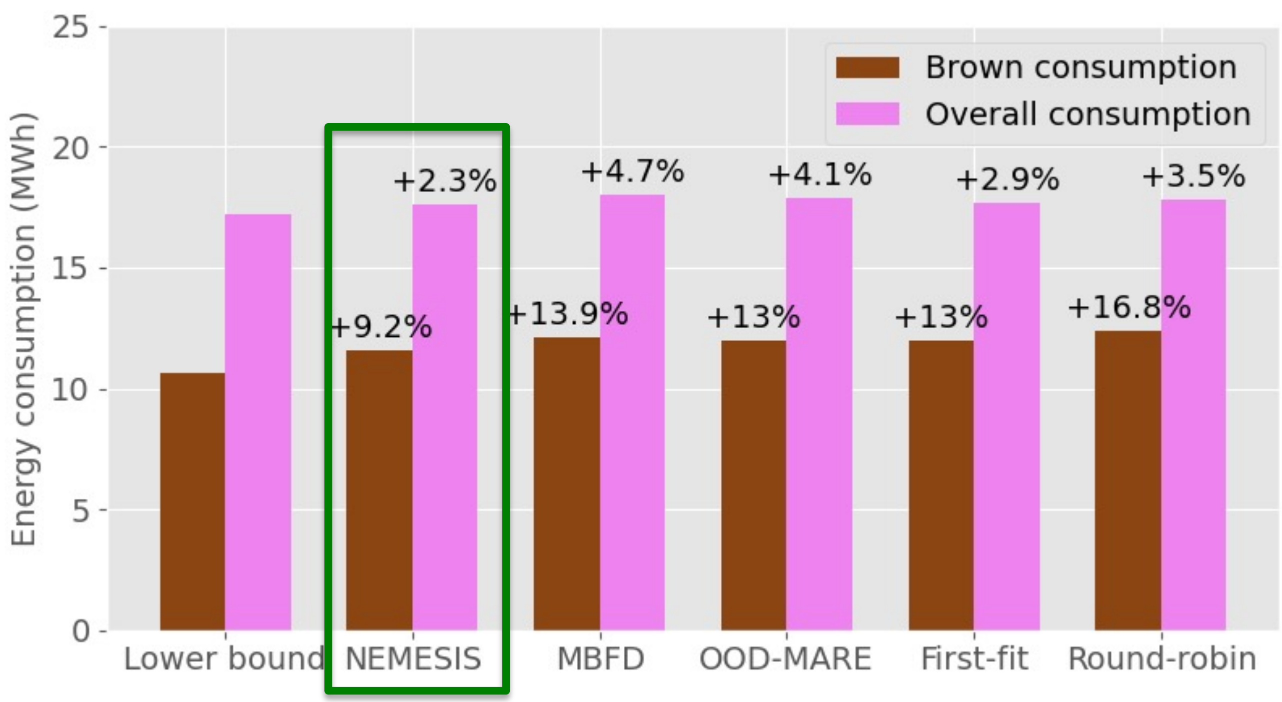
   b. Deploy and migrate the VMs

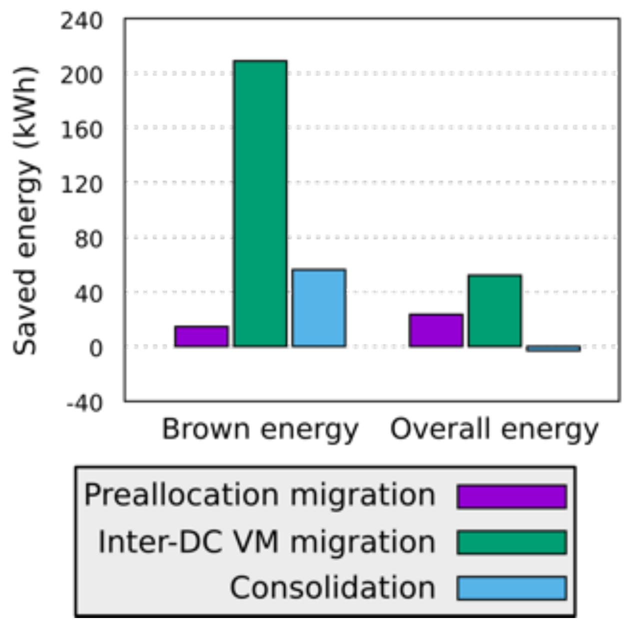
Energy-efficient algorithm dissection

9 DCs and 1,035 servers in total

Theoretical lower bound => best-fit on a single DC

State-of-the-art: MBFD, OOD-MARE





[SBAC-PAD2018]