Energy measurements in HPC Architectures

[CMP223] Computer Systems Performance Analysis
[INF01146] Análise de Desempenho

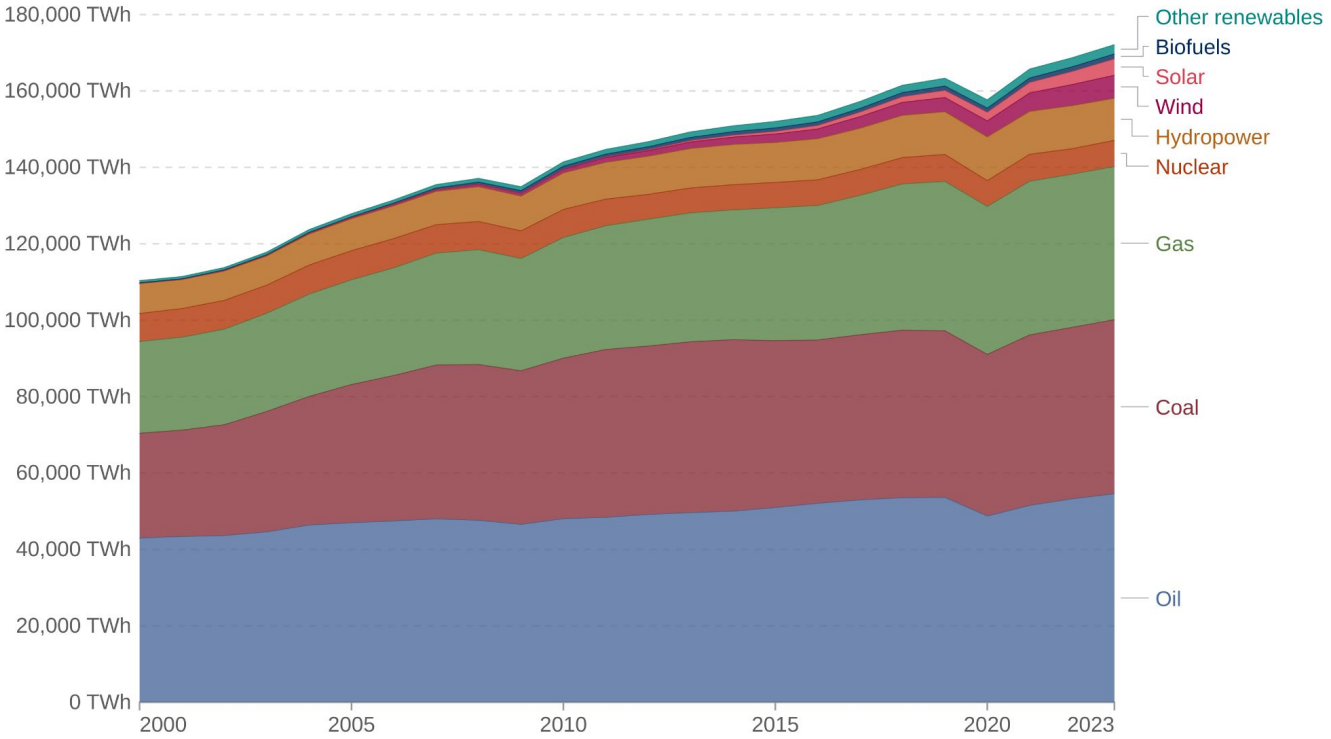
Laura Soares
Luna Amanuel?
Otho Marcondes
October/25

Agenda

- Context & Motivation
- Computational Object
- Application, Instrumentation, Metrics
- Preliminary Results
- Next Steps

Energy consumption by source, World

Measured in terms of primary energy using the substitution method.



Data source: Energy Institute - Statistical Review of World Energy (2024)

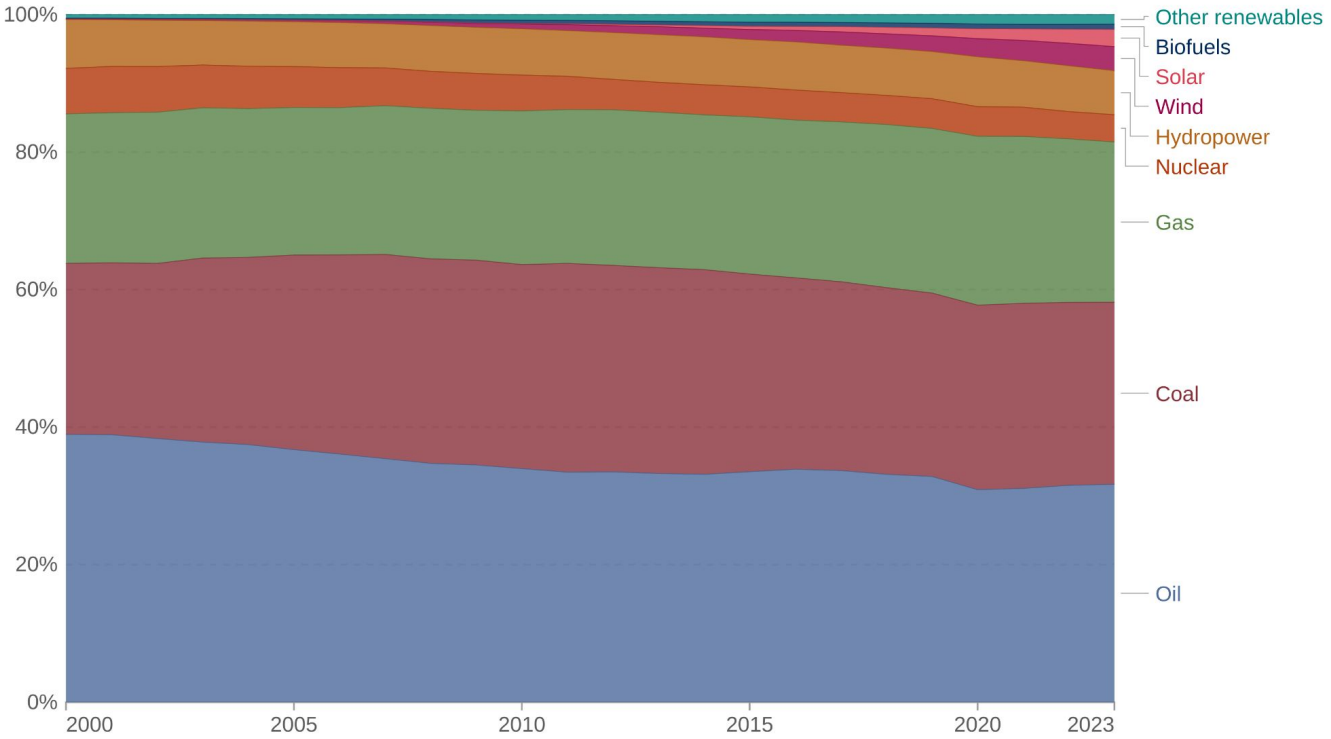
OurWorldinData.org/energy | CC BY

Note: "Other renewables" include geothermal, biomass, and waste energy.

[energy mix]

Energy consumption by source, World

Measured in terms of primary energy using the substitution method.

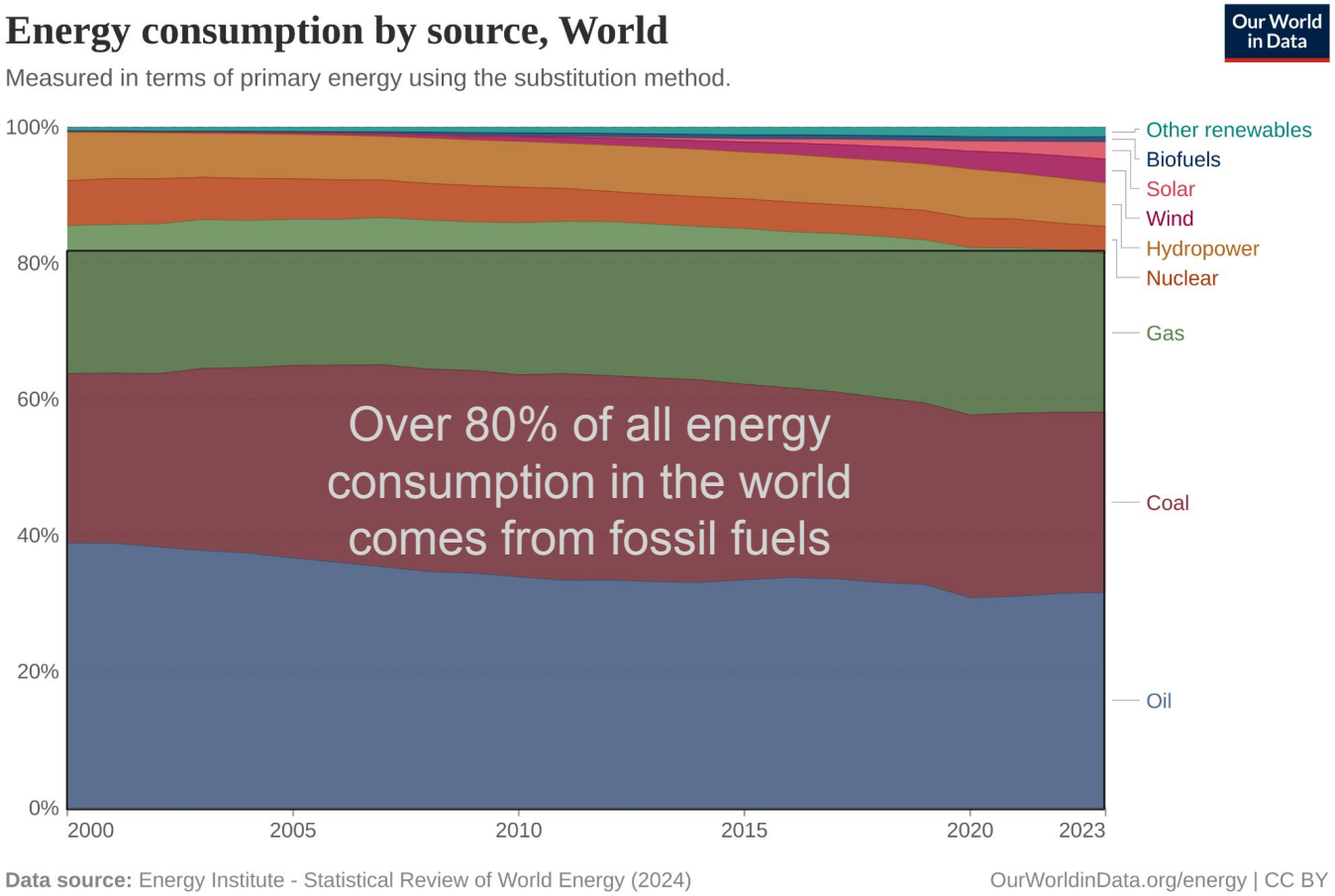


Data source: Energy Institute - Statistical Review of World Energy (2024)

Note: "Other renewables" include geothermal, biomass, and waste energy.

OurWorldinData.org/energy | CC BY

[energy mix]


Energy consumption by source, World
Measured in terms of primary energy using the substitution method.
Our World in Data
100%
80%
60%
40%
20%
0%
2000
2005
2010
2015
2020
2023
Other renewables
Biofuels
Solar
Wind
Hydropower
Nuclear
Gas
Coal
Oil
Over 80% of all energy consumption in the world comes from fossil fuels
Data source: Energy Institute - Statistical Review of World Energy (2024)
Note: "Other renewables" include geothermal, biomass, and waste energy.
OurWorldinData.org/energy | CC BY

[energy mix]

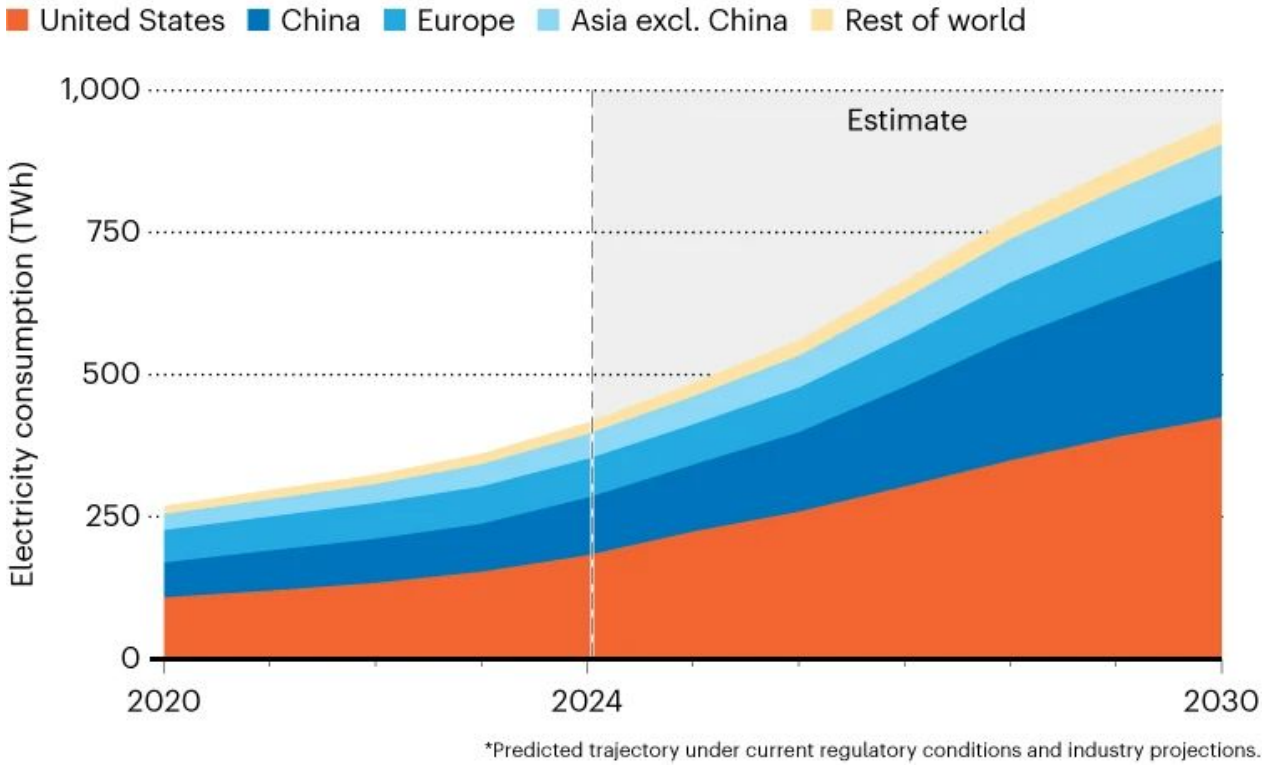
Energy consumption in data-centres

- Data-centres in 2024 consumed 415 TWh, about 1.5% off all energy consumed in the world
- This number might reach 945 TWh in 2030

[nature]

DATA-CENTRE ENERGY GROWTH

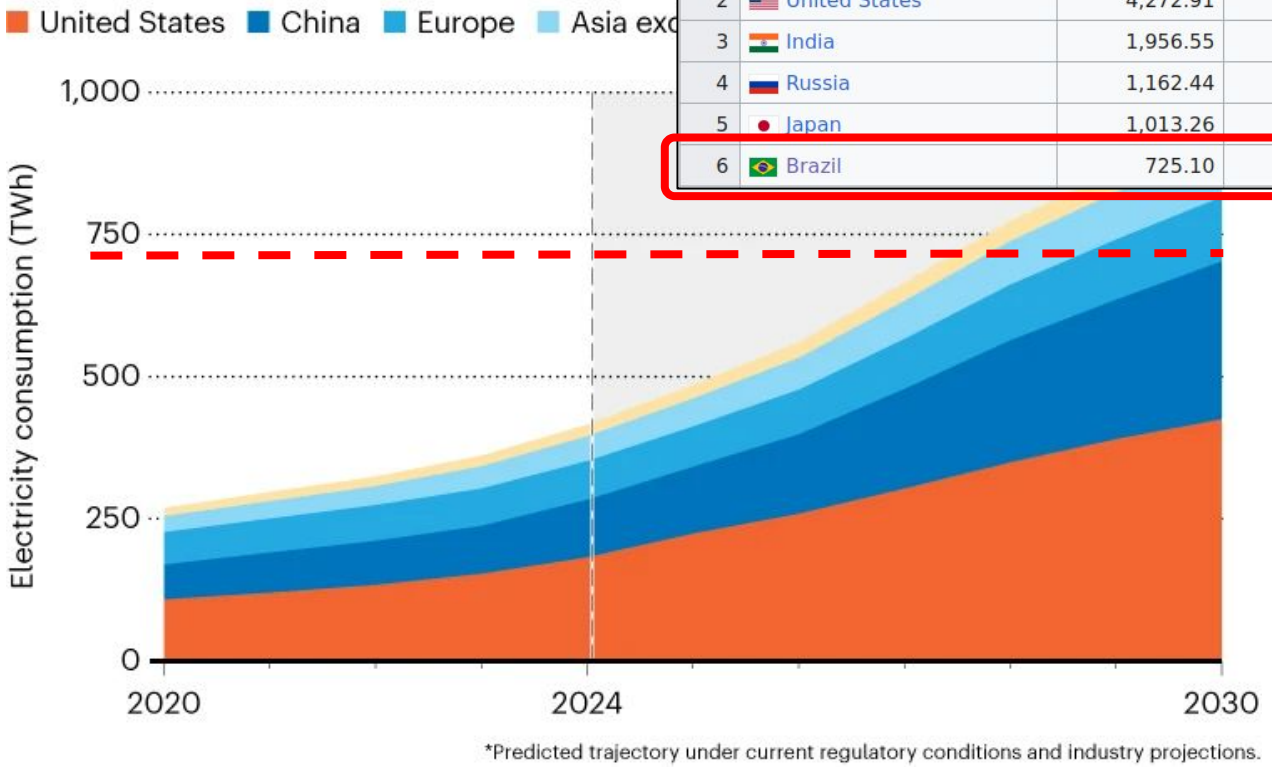
China and the United States are predicted to account for nearly 80% of the global growth in electricity consumption by data centres up to 2030*.



[nature]

DATA-CENTRE ENERGY GROWTH

China and the United States are predicted to acc
the global growth in electricity consumption by



	Location	Consumption (TWh)	Per capita (MWh)	Year
	World	29,664.60	3.67	2023
1	China	9,443.07	6.64	2023
2	United States	4,272.91	12.44	2023
3	India	1,956.55	1.36	2023
4	Russia	1,162.44	7.99	2023
5	Japan	1,013.26	8.15	2023
6	Brazil	725.10	3.43	2023

[nature | wikipedia]

Energy measurements in HPC architectures

- It is increasingly critical to have energy monitoring tools in data-centres
- Optimizing energy performance depends on monitoring
- Allows power management initiatives

Energy measurements in HPC architectures

- Perform energy measurements on a cluster (computational object)
- Utilize an application/program to stress the machines
  - LU factorization (StarPU + Chameleon)
  - Stress package (lacks GPU support)

Computational object

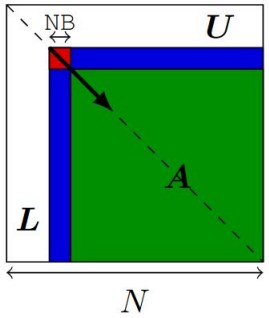
Partition	CPU	RAM	Accelerator	Disk	Motherboard
poti[1,2,3,4,5]	Intel(R) Core(TM) i7-14700KF, 3.40 GHz, 28 threads, 20 cores	96 GB DDR5	NVIDIA GeForce RTX 4070	1.7 TB SSD, 119.2 GB NVME	Gigabyte Technology Co., Ltd. Z790 UD AX
tupi[5,6]	Intel(R) Core(TM) i9-14900KF, 3.20 GHz, 32 threads, 24 cores	128 GB DDR5	NVIDIA GeForce RTX 4090	1.7 TB SSD, 1.8 TB NVME	Gigabyte Technology Co., Ltd. Z790 UD AX
tupi[3,4]	Intel(R) Core(TM) i9-14900KF, 3.20 GHz, 32 threads, 24 cores	128 GB DDR5	NVIDIA GeForce RTX 4090	1.8 TB NVME	Gigabyte Technology Co., Ltd. Z790 UD AX

LU Factorization

$$
\begin{aligned}
Ax &= b \\
(LU)\,x &= b \\
L\,(Ux) &= b
\end{aligned}
$$

$$Ly = b \quad \text{e} \quad Ux = y$$

```
for (k = 0; k < N; k++)
  DGTRF-NOPIV(RW, A[k][k]);
  for (m = k+1; m < N; m++)
    DTRSM(RW, A[m][k], R, A[k][k]);
    DTRSM(RW, A[k][m], R, A[k][k]);
  for (n = k+1; n < N; n++) // Update
    for (m = k+1; m < N; m++)
      DGEMM(RW, A[m][n], R, A[m][k],
                        R, A[k][n]);
```

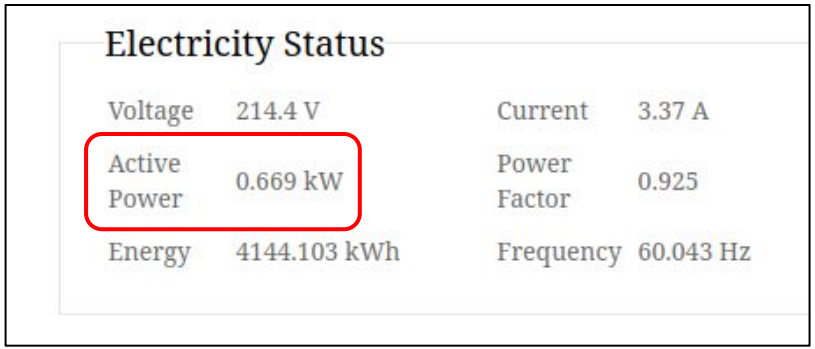


Instrumentation: Network-manageable Rack Power Distribution Unit (PDU)

- The PDUs (the power outlet) used by the nodes are connected to the internal network of the cluster
  - Access using SSH
  - Answers to SNMP requests
- Provide energy measurements

Metrics: Active Power vs. Energy

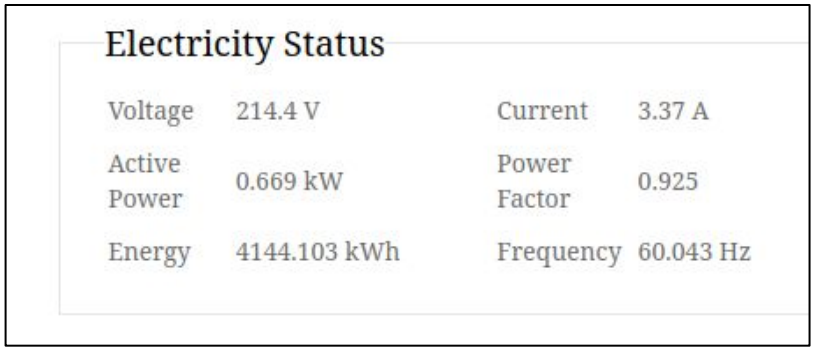
- Active Power: electrical energy consumed in a circuit, in watts (W ou kW)
- “the energy actually used in load”
- P= V × I × cosϕ



[Schneider Electric]

Metrics: Active Power vs. Energy

- Active Power: electrical energy consumed in a circuit, in watts (W ou kW)
- “the energy actually used in load”
- P= V × I × cosϕ



voltage root mean square*

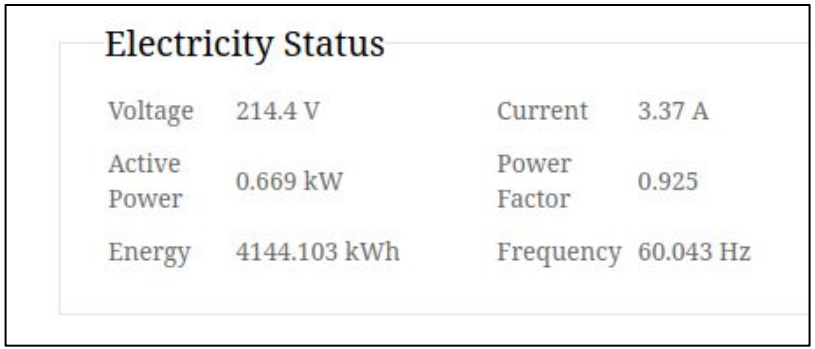
*** the square root of the mean square of a set of values**

(o valor eficaz é a raiz quadrada da média aritmética dos quadrados dos valores)

[Schneider Electric | wikipedia 2]

Metrics: Active Power vs. Energy

- Active Power: electrical energy consumed in a circuit, in watts (W ou kW)
- “the energy actually used in load”
- P= V × I × cosϕ



current root mean square*

*** the square root of the mean square of a set of values**

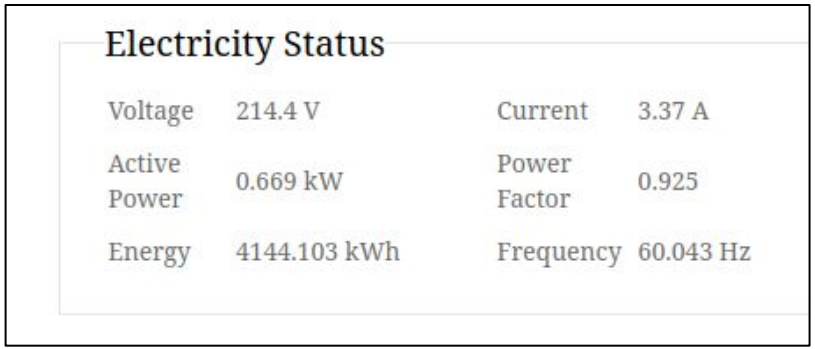
(o valor eficaz é a raiz quadrada da média aritmética dos quadrados dos valores)

[Schneider Electric | wikipedia 2]

Metrics: Active Power vs. Energy

- Active Power: electrical energy consumed in a circuit, in watts (W ou kW)
- “the energy actually used in load”
- P= V × I × cosϕ

power factor



[Schneider Electric]

Metrics: script making SNMP requests

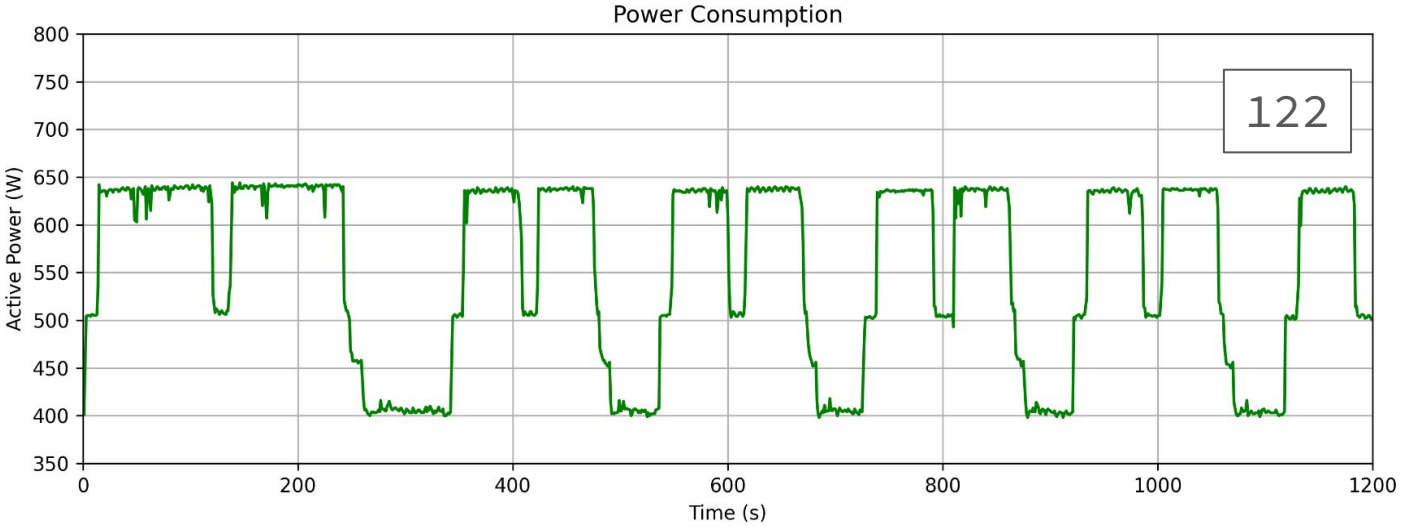
```
while $SECONDS -lt $run_time
    echo `date`
    snmpget etc $IP PowerNet-MIB::ePDUDeviceStatusEnergy.1
    snmpget etc $IP PowerNet-MIB::ePDUDeviceStatusActivePower.1
    sleep $sleep
done
```

time (YYYY-MM-DD HH:MM:SS)
energy (kWh, cumulative)
active power (kW)

Measurement example I

tupi[**5**-6] multinode_pcept_train
tupi**3** i9_parquet_analysis_fix

OR

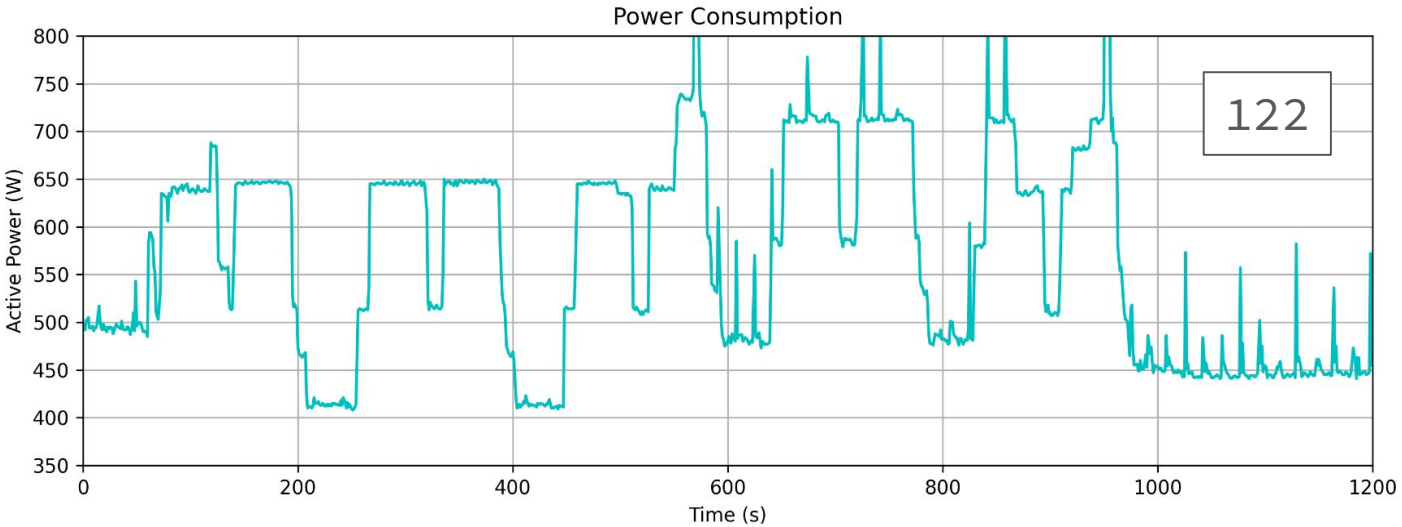


Measurement example I

tupi[**5**-6] multinode_pcept_train
tupi**3** i9_parquet_analysis_fix

AND

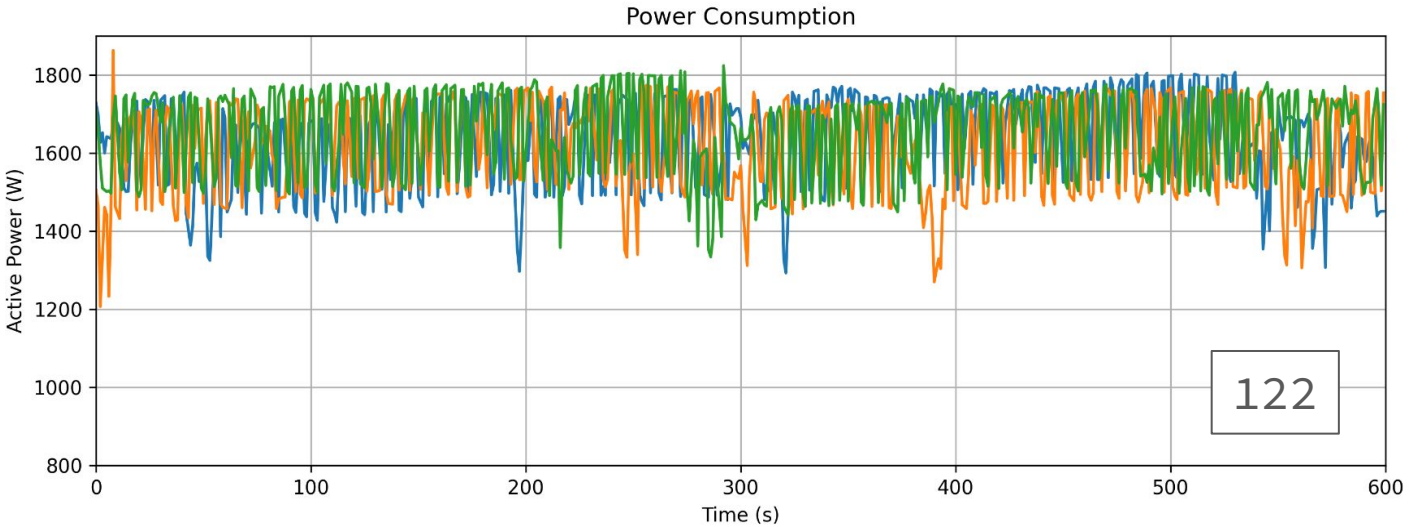
we need to allocate all the machines in the same PDU so other users' experiments don't show up in our measurements



Measurement example II

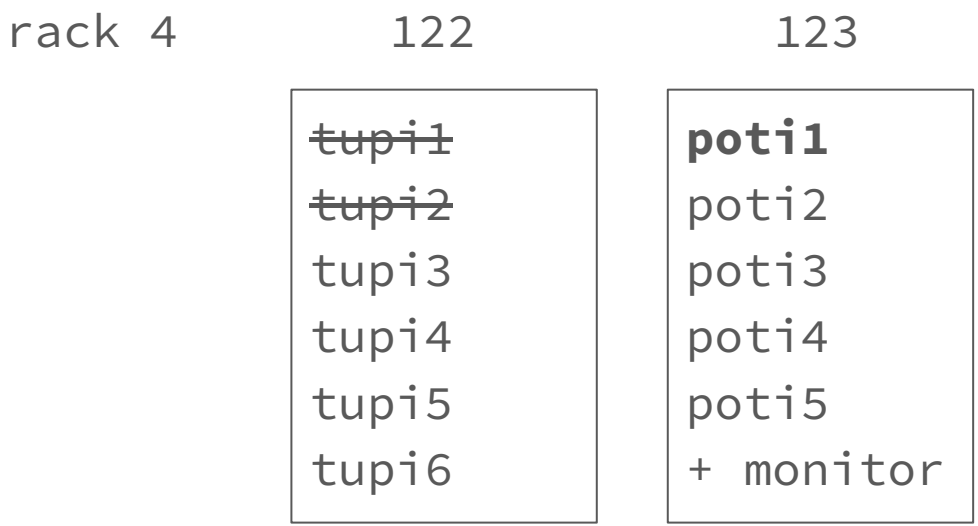
LLM inference (Qwen3-4B), 300 planning tasks in two batches with long-context answers (32k tokens)

tupi[**2,3**,6] progressive
poti[**1**,2,3,**4**,5] e2e-plan



Stage 2: preliminary results

Electric topology adjusted ✅



Experiments

- **Idle**
- LU Factor
  - Machines in use (1~5)
  - CPU Workload (size of the problem)
- **Stress**
  - Machines in use (1~5)
  - CPU Workload (3 levels)

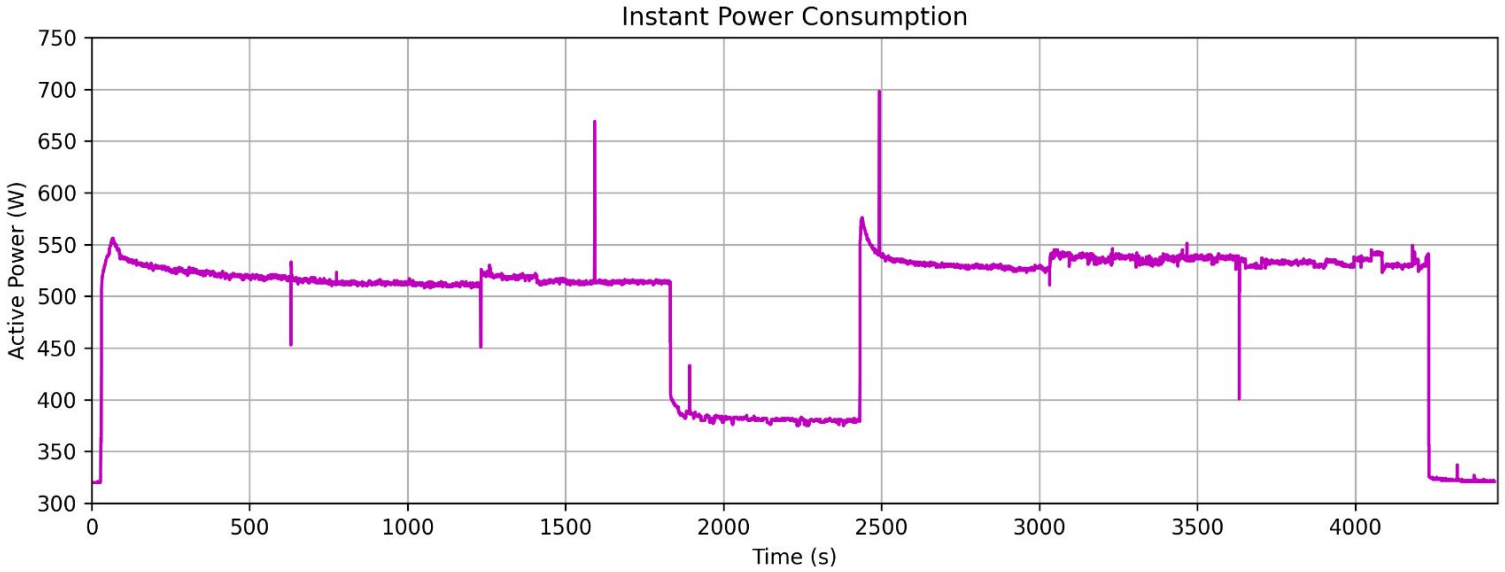
Stress

- Experiments varying CPU, IO and memory (24 cores) using one poti machine:
  - 24 cores (only one of the above e.g., CPU only)
  - 12, 12 (just two of the above, with 12 cores each)
  - 8,8,8 (equally distributed cores among CPU, IO and memory)

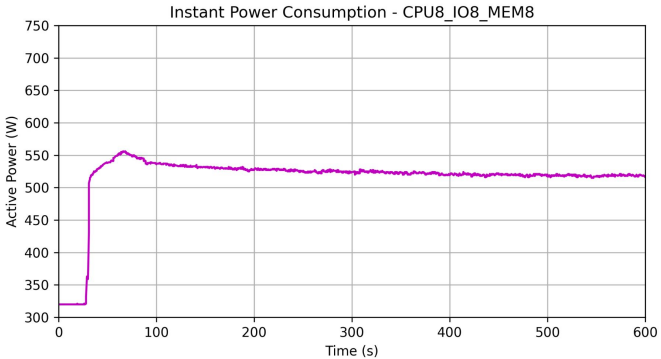
time	cpu	io	mem	nodes	exp_name	exp_run
10m	8	8	8	1	stress_poti_1_nodes	stress_poti_1_nodes_10m_CPU8_IO8_MEM8_N1_1
10m	12	12	0	1	stress_poti_1_nodes	stress_poti_1_nodes_10m_CPU12_IO12_MEM0_N1_2
10m	24	0	0	1	stress_poti_1_nodes	stress_poti_1_nodes_10m_CPU24_IO0_MEM0_N1_3
10m	0	24	0	1	stress_poti_1_nodes	stress_poti_1_nodes_10m_CPU0_IO24_MEM0_N1_4
10m	12	0	12	1	stress_poti_1_nodes	stress_poti_1_nodes_10m_CPU12_IO0_MEM12_N1_5
10m	0	12	12	1	stress_poti_1_nodes	stress_poti_1_nodes_10m_CPU0_IO12_MEM12_N1_6
10m	0	0	24	1	stress_poti_1_nodes	stress_poti_1_nodes_10m_CPU0_IO0_MEM24_N1_7

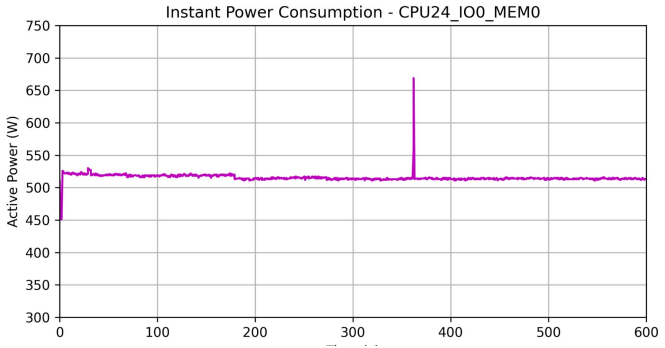
Idle energy measurements (poti partition)

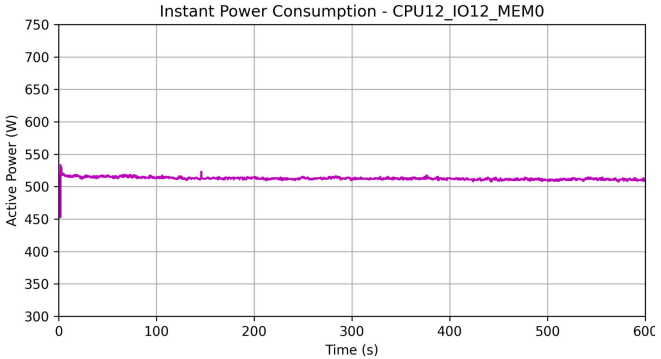
Energy measurements running the experiment

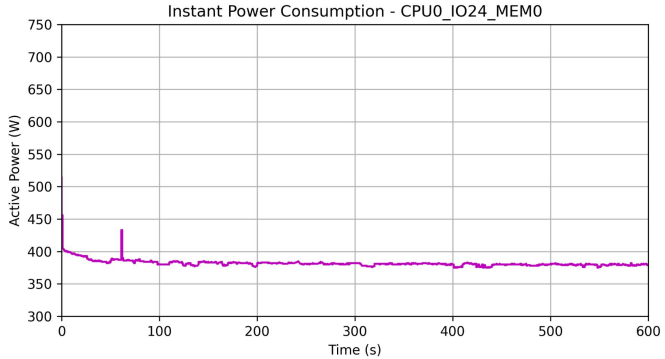


Results

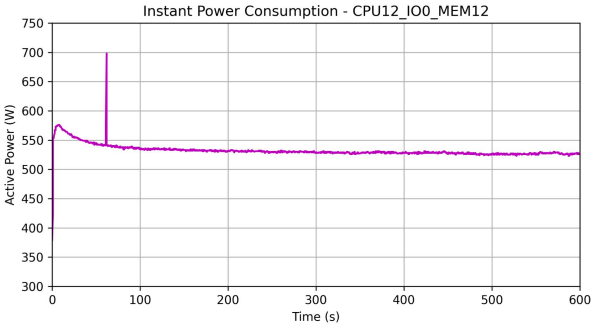

Instant Power Consumption - CPU8_IO8_MEM8
Active Power (W)
Time (s)

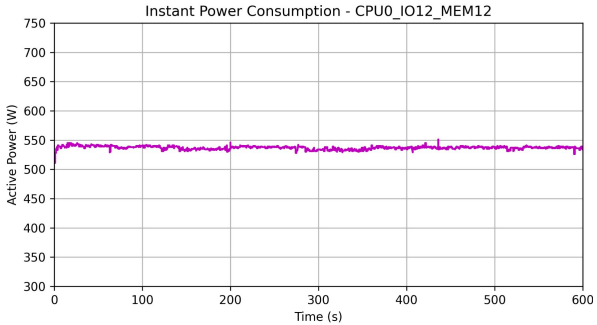

Instant Power Consumption - CPU24_IO0_MEM0
Active Power (W)

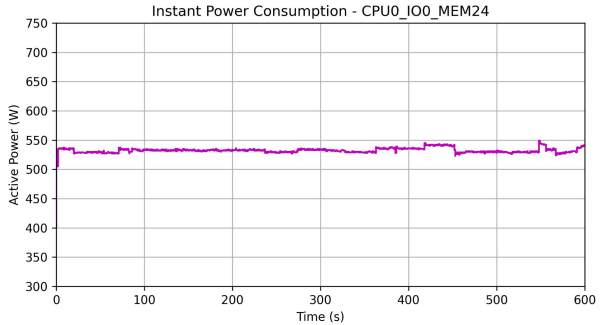

Instant Power Consumption - CPU12_IO12_MEM0
Active Power (W)
Time (s)


Instant Power Consumption - CPU0_IO24_MEM0
Active Power (W)
Time (s)

Results


Instant Power Consumption - CPU12_IO0_MEM12
Active Power (W)
Time (s)
750
700
650
600
550
500
450
400
350
300
0
100
200
300
400
500
600


Instant Power Consumption - CPU0_IO12_MEM12
Active Power (W)
Time (s)
750
700
650
600
550
500
450
400
350
300
0
100
200
300
400
500
600

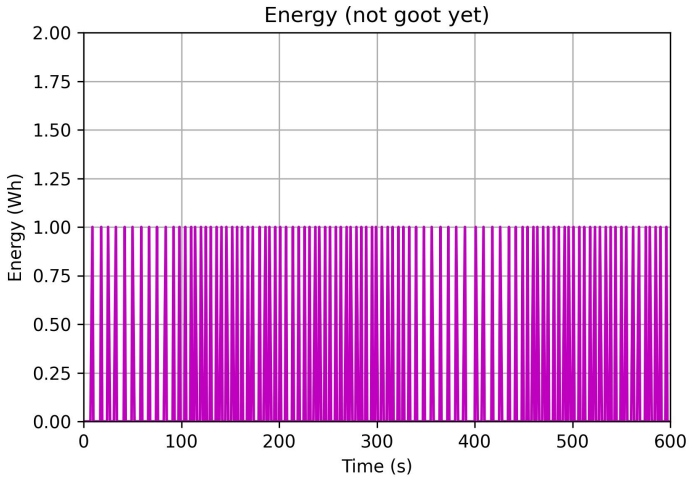

Instant Power Consumption - CPU0_IO0_MEM24
Active Power (W)
Time (s)
750
700
650
600
550
500
450
400
350
300
0
100
200
300
400
500
600

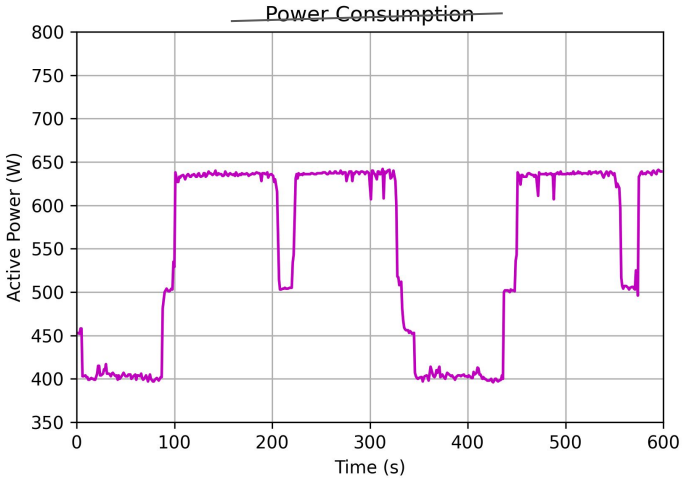
Difficulties

- Machine allocation (takes time)
- Nodes malfunctioning
- SNMP configuration
- The energy monitoring script sometimes gets killed and we don’t know why
- Energy metrics visualization (cumulative vs instant)

Difficulties

- Energy metrics visualization (cumulative vs instant)





Next steps

- Number of nodes, replication
- Execute the Stress application on the tupi partition;
- Execute the LU factorization application with and without GPU on both poti and tupi partitions;
- Results using both energy metrics (cumulative + instant)
- Analyze the results

Thank you!! ✨🧑‍🎓🧠💖🔋

Any questions?

lrsoares@inf.ufrgs.br
otho.marcondes@inf.ufrgs.br
luntek22@student.hh.se??

References 1

[energy mix] https://ourworldindata.org/energy-mix

[nature] https://www.nature.com/articles/d41586-025-01113-z

[wikipedia] https://en.wikipedia.org/wiki/List_of_countries_by_electricity_consumption

[Schneider Electric] https://eshop.se.com/in/blog/post/difference-between-active-power-reactive-power-and-apparent-power.html

[wikipedia 2] https://pt.wikipedia.org/wiki/Valor_eficaz

[ICPADS] Lucas Leandro Nesi, Lucas Mello Schnorr, Arnaud Legrand. Communication-Aware Load Balancing of the LU Factorization over Heterogeneous Clusters. IEEE International Conference on Parallel and Distributed Systems (ICPADS), Dec 2020, Hong Kong, France. hal-02633985