University of Basel

DMI-HPC @ UniBas

Performance & Portability For Sustainable Simulations at Extreme Scales

Florina M. Ciorba

18th Scheduling for Large-Scale Systems
Montreal, QC, Canada
July 10, 2025, 9:00-9:30

florina.ciorba@unibas.ch | @cflorina.bsky.social | www.linkedin.com/in/cflorina

Swiss National Science Foundation

Platform for Advanced Scientific Computing

SKACH

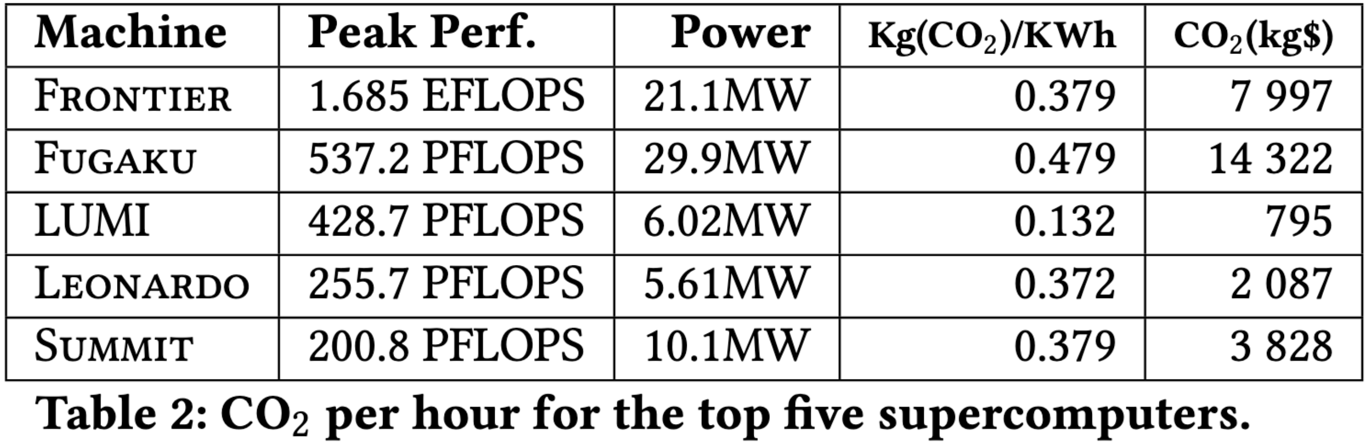
DAPHNE

Performance ⚖ Sustainability

non-exhaustive

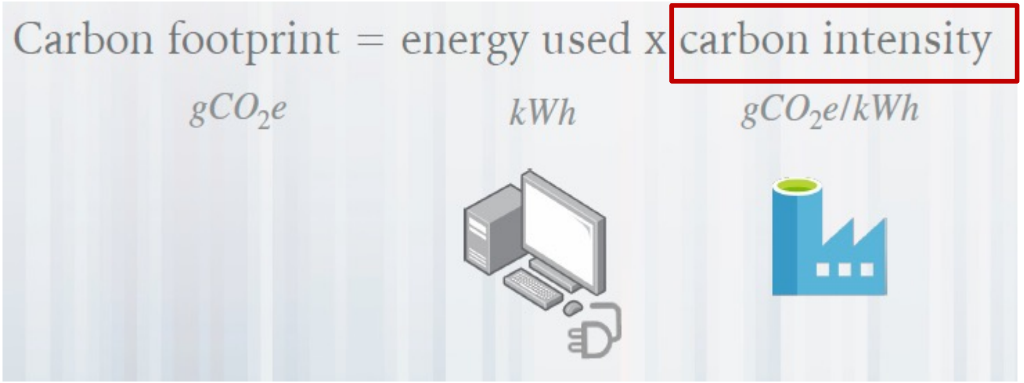
The Hidden Cost: The Carbon Footprint of Computation

Scope-1 CO2 per Hour for the Top Five Supercomputers (2023)



Machine	Peak Perf.	Power	Kg(CO_2)/KWh	CO_2(kg$)
FRONTIER	1.685 EFLOPS	21.1MW	0.379	7 997
FUGAKU	537.2 PFLOPS	29.9MW	0.479	14 322
LUMI	428.7 PFLOPS	6.02MW	0.132	795
LEONARDO	255.7 PFLOPS	5.61MW	0.372	2 087
SUMMIT	200.8 PFLOPS	10.1MW	0.379	3 828

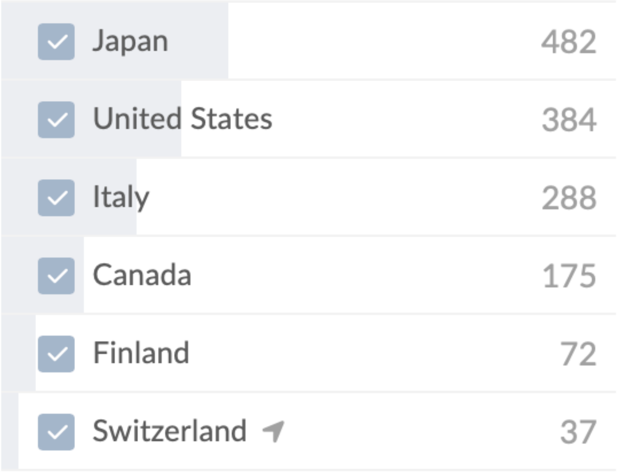
Table 2: CO_2 per hour for the top five supercomputers.



Carbon footprint = energy used x carbon intensity

gCO_2e kWh gCO_2e/kWh

Floating point performance not correlated to CO2e. ***Machine hardware profile is a key factor.***



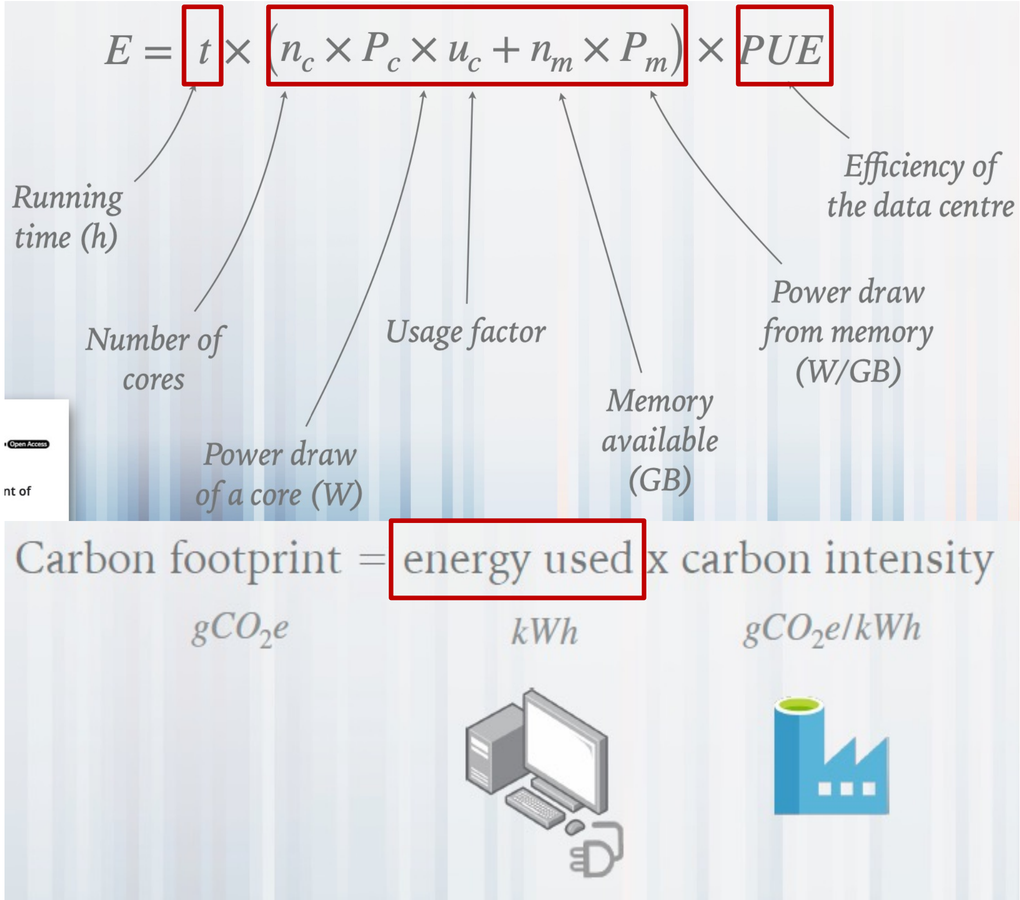
☑ Japan	482
☑ United States	384
☑ Italy	288
☑ Canada	175
☑ Finland	72
☑ Switzerland	37

Carbon intensity data (2025)
ourworldindata.org/electricity-mix

Energy Concerns with HPC Systems and Applications, R. Nana, C. Tadonki, P. Dokladal, Y. Mesri https://arxiv.org/pdf/2309.08615, 2023

The Hidden Cost: The Carbon Footprint of Computation

Minimize Energy Needed



$$E = \boxed{t} \times \boxed{\left(n_c \times P_c \times u_c + n_m \times P_m\right)} \times \boxed{PUE}$$

Carbon footprint = $\boxed{\text{energy used}}$ x carbon intensity

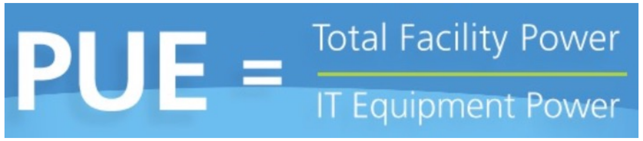
gCO_2e kWh gCO_2e/kWh

min *t*: Optimize (or use optimized) code.

min *P*: Employ the right amount of parallelism and memory and use them efficiently (no idle cores, no free memory).

min *PUE*: Promote efficient data centers; Carefully choose your computing facility.

min *runs*: How many times do you really run a simulation / an analysis? Let’s be mindful!



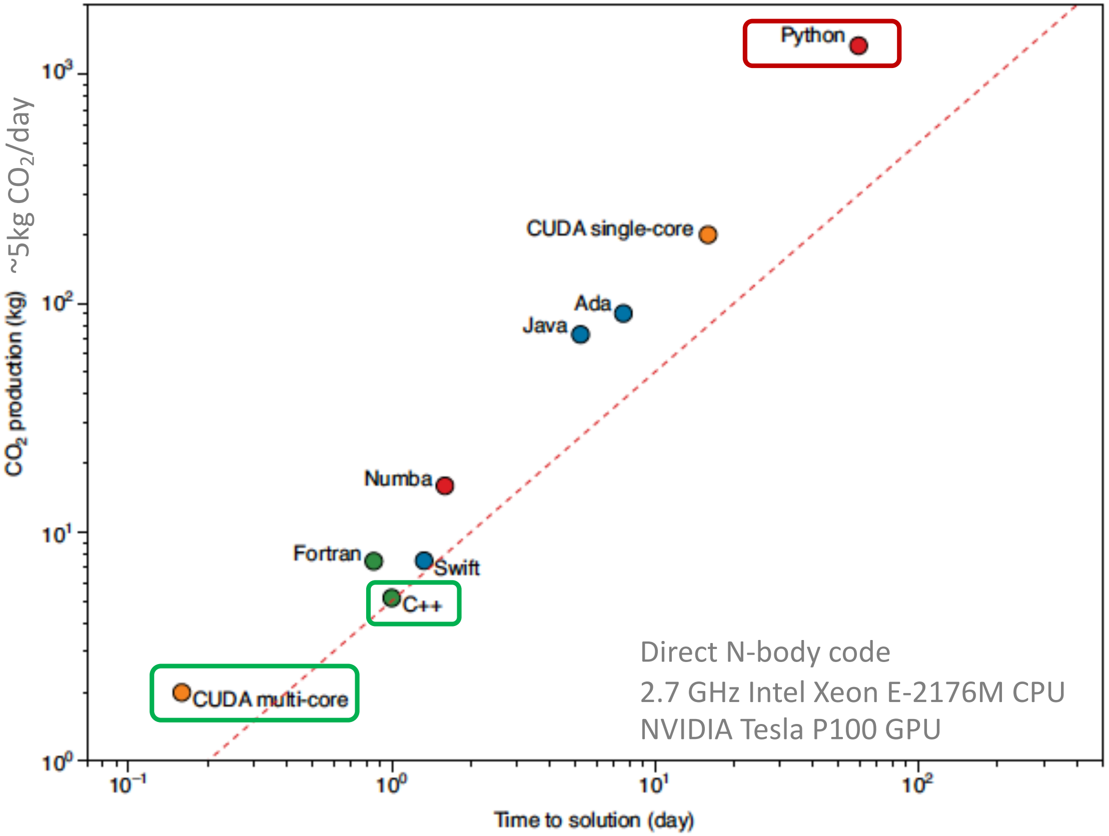
$$PUE = \frac{\text{Total Facility Power}}{\text{IT Equipment Power}}$$

The environmental impact of computational biology, Loic Lannelongue, https://www.youtube.com/watch?v=Kl64yn1eDUo&t=787s

⚖ Performance
Sustainability

The Hidden Cost: The Carbon Footprint of Computation

Minimize Energy Needed: Programming Language & Resource Usage Matter



Direct N-body code
2.7 GHz Intel Xeon E-2176M CPU
NVIDIA Tesla P100 GPU

$$\underset{gCO_2e}{\text{Carbon footprint}} = \underset{kWh}{\text{energy used}} \text{ x } \underset{gCO_2e/kWh}{\text{carbon intensity}}$$

$$E = \boxed{t} \times (n_c \times P_c \times u_c + n_m \times P_m) \times PUE$$

High performance programming languages (CUDA, C++) and multicore/accelerator usage reduce problem time to solution (**t**) as well as **Carbon production**.

Frequency throttling and power capping [1,2] also also improve **t** and **Carbon production**.

min *t*: Use energy-efficient programming language(s).

Plots' source: Portegies Zwart, S. (2020). The ecological impact of high-performance computing in astrophysics. *Nature Astronomy*, *4*(9), 819-822.

[1] Krzywaniak, A., Czarnul, P., & Proficz, J. (2023). Dynamic GPU power capping with online performance tracing for energy efficient GPU computing using DEPO tool. FGCS, 145, 396-414.

[2] Hautreux, G., Malaboeuf, E. (2023). Reducing HPC energy footprint for large scale GPU accelerated workloads. Cray User Group CUG23.

Performance
⚖ Sustainability

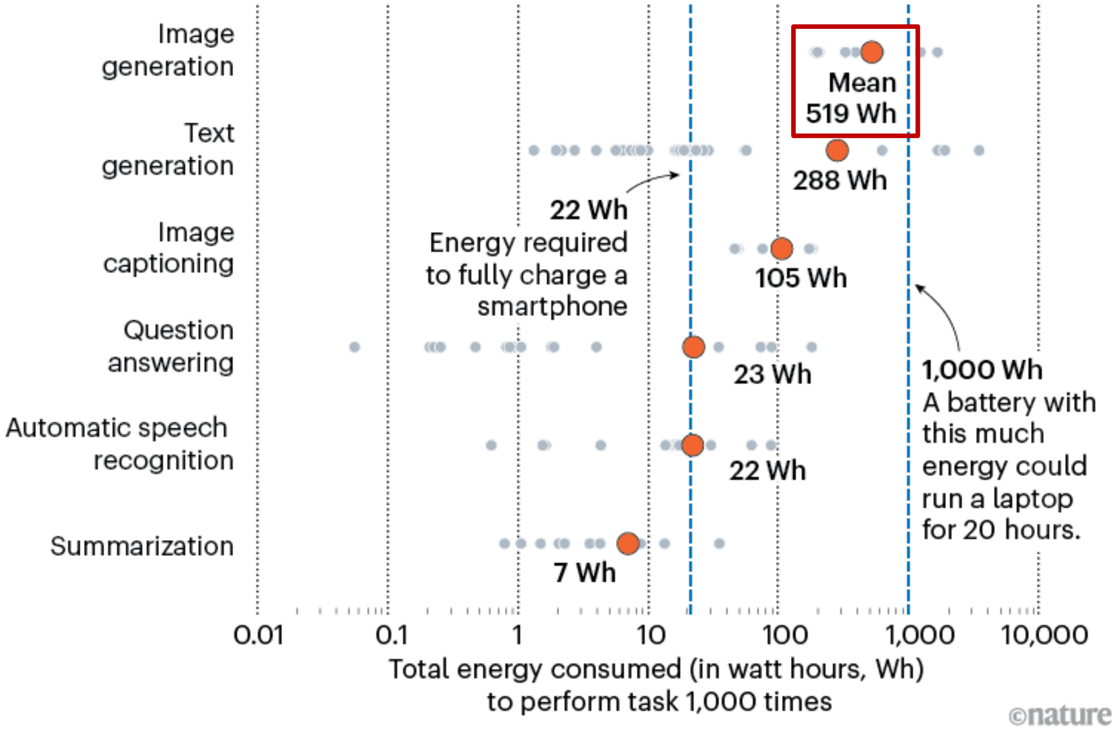
The Hidden Cost: The Carbon Footprint of Computation

Generative AI Applications

18th **Scheduling for large-scale systems** workshop
École de Technologie Supérieure, Montréal,Québec, Canada, July 8-10, 2025

HOW MUCH ENERGY DOES AI USE?

The AI Energy Score project tested dozens of artificial-intelligence models to estimate how much energy they consume when performing various tasks. Plotting the energy required to perform a task 1,000 times shows that energy use varies greatly depending on the task and the model.



CodeCarbon (Python package) accessed the technical specifications of chips that executed a model for a user request in the data center.

Different amounts of energy per GenAI task

- Single image gen. from text prompt ~ 0.5 Wh
  - ×1’000 ≅ 519 Wh
- Text gen. (288 Wh) < image gen. (519 Wh)

ℹ️ Modern smartphone 📱 22 Wh for full charge

Note: Lower bound estimates

https://www.nature.com/articles/d41586-025-00616-z

Performance
⚖ Sustainability

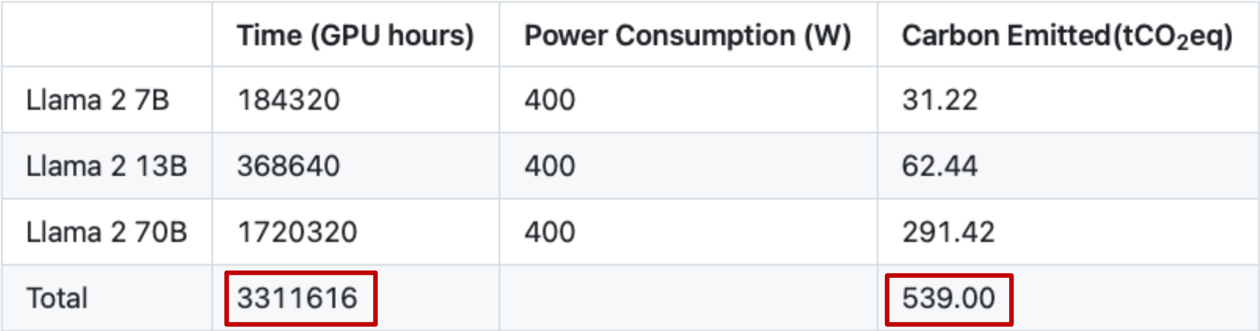
The Hidden Cost: The Carbon Footprint of Computation

Generative AI Applications: Llama 2 Model Family

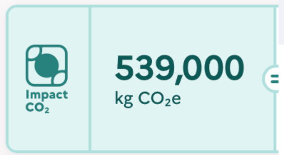
18th **Scheduling for large-scale systems** workshop
École de Technologie Supérieure, Montréal,Québec, Canada, July 8-10, 2025

CO2 emissions during pretraining

Power Consumption (W): peak power capacity per GPU device for the GPUs used adjusted for power usage efficiency.

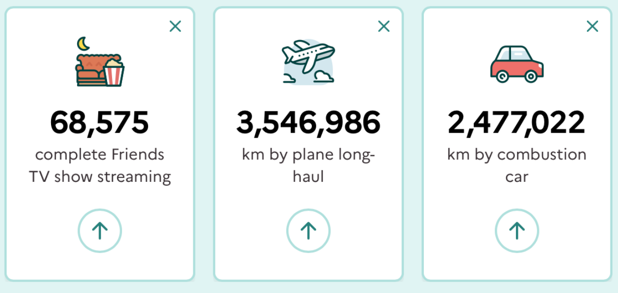


	Time (GPU hours)	Power Consumption (W)	Carbon Emitted(tCO_2eq)
Llama 2 7B	184320	400	31.22
Llama 2 13B	368640	400	62.44
Llama 2 70B	1720320	400	291.42
Total	3311616		539.00



Impact CO_2

539,000
kg CO_2e



68,575	3,546,986	2,477,022
complete Friends TV show streaming	km by plane long-haul	km by combustion car

Calculator: https://impactco2.fr/outils/comparateur

Training Factors

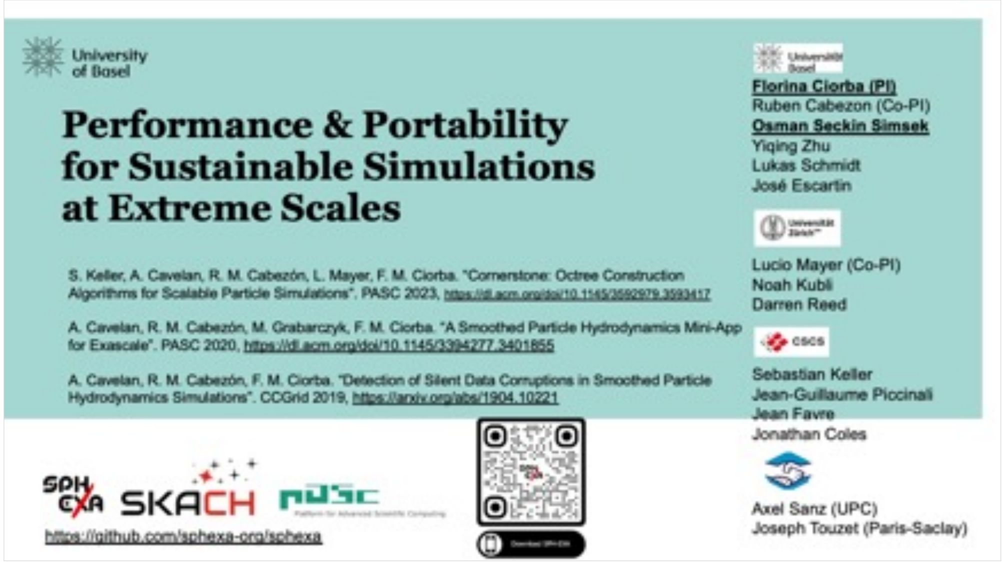
Used custom training libraries, Meta's Research Super Cluster, and production clusters for pretraining. Fine-tuning, annotation, and evaluation were also performed on third-party cloud compute.

Carbon Footprint

Pretraining: cumulative **3.3M GPUh** of computation on hardware of type A100-80GB (TDP of 350-400W).
Estimated total emissions were **539 tCO2eq**,
100% of which were offset by Meta’s sustainability program. Open release of models → pretraining costs not incurred by others.

https://github.com/meta-llama/llama/blob/main/MODEL_CARD.md

Today’s talk



University of Basel

Performance & Portability for Sustainable Simulations at Extreme Scales

S. Keller, A. Cavelan, R. M. Cabezón, L. Mayer, F. M. Ciorba. “Cornerstone: Octree Construction Algorithms for Scalable Particle Simulations”. PASC 2023, https://dl.acm.org/doi/10.1145/3592979.3593417

A. Cavelan, R. M. Cabezón, M. Grabarczyk, F. M. Ciorba. “A Smoothed Particle Hydrodynamics Mini-App for Exascale”. PASC 2020, https://dl.acm.org/doi/10.1145/3394277.3401855

A. Cavelan, R. M. Cabezón, F. M. Ciorba. “Detection of Silent Data Corruptions in Smoothed Particle Hydrodynamics Simulations”. CCGrid 2019, https://arxiv.org/abs/1904.10221

https://github.com/sphexa-org/sphexa

Florina Ciorba (PI)
Ruben Cabezon (Co-PI)
Osman Seckin Simsek
Yiqing Zhu
Lukas Schmidt
José Escartin

Lucio Mayer (Co-PI)
Noah Kubli
Darren Reed

Sebastian Keller
Jean-Guillaume Piccinali
Jean Favre
Jonathan Coles

Axel Sanz (UPC)
Joseph Touzet (Paris-Saclay)

University of Basel

An Autonomy Loop for Dynamic HPC Job Time Limit Adjustment

Thomas Jakobsche
Osman S. Simsek
Florina M. Ciorba

To be presented at
31st International European Conference on Parallel and Distributed Computing, August 2025

A. Gentile
J. Brandt

Preprint: https://arxiv.org/abs/2505.05927

University of Basel

Performance & Portability for Sustainable Simulations at Extreme Scales

S. Keller, A. Cavelan, R. M. Cabezón, L. Mayer, F. M. Ciorba. “Cornerstone: Octree Construction Algorithms for Scalable Particle Simulations“. PASC 2023, https://dl.acm.org/doi/10.1145/3592979.3593417

A. Cavelan, R. M. Cabezón, M. Grabarczyk, F. M. Ciorba. “A Smoothed Particle Hydrodynamics Mini-App for Exascale”. PASC 2020, https://dl.acm.org/doi/10.1145/3394277.3401855

A. Cavelan, R. M. Cabezón, F. M. Ciorba. “Detection of Silent Data Corruptions in Smoothed Particle Hydrodynamics Simulations”. CCGrid 2019, https://arxiv.org/abs/1904.10221

SPH EXA SKACH PASC Platform for Advanced Scientific Computing

https://github.com/sphexa-org/sphexa

Download SPH-EXA

Universität Basel

Florina Ciorba (PI)
Ruben Cabezon (Co-PI)
Osman Seckin Simsek
Yiqing Zhu
Lukas Schmidt
José Escartin

Universität Zürich UZH

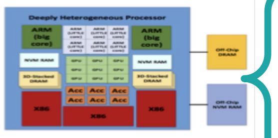
Lucio Mayer (Co-PI)
Noah Kubli
Darren Reed

CSCS

Sebastian Keller
Jean-Guillaume Piccinali
Jean Favre
Jonathan Coles

Axel Sanz (UPC)
Joseph Touzet (Paris-Saclay)

SPH EXA Scalable SPH+N-body for 10^{12} Particle Simulations at Exascale



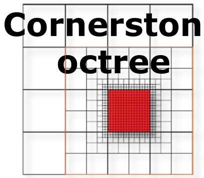
Heterogeneity

SPH-EXA application front-end (github.com/unibas-dmi-hpc/SPH-EXA)

- Initial conditions, checkpointing, compression, parallel I/O
- Flexible combination and addition of additional physics for domain scientists
- Performance data for scheduling, load-balancing, energy-efficiency optimizations
- In-situ and post-hoc visualization
- **10'751** C++ LoC, Y= 287 CUDA/HIP/SYCL* → enables performance, portability, visualization

EuroHPC Joint Undertaking

Extreme Scale Access **22,000,000 GPUh*** on LUMI-G



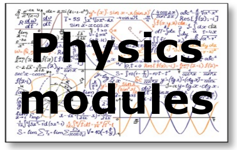
Domain Decomposition

- Space-filling curves and octrees
- Global and locally essential octrees
- Octree-based domain decomposition
- **30'481** C++ LOC, Y= 6'840 CUDA/HIP /SYCL* → enables extreme scalability (weak and strong)

SPHYNX

ChaNGa

Flagship code for



Modern SPH and physics implementation with key features (astro.physik.unibas.ch/sphynx, github.com/N-BodyShop/changa)

- Generalized volume elements
- Integral approach to derivatives
- Artificial viscosity with switches
- Sub-grid physics
- **5'072** C++ LOC, Y= 2'152 CUDA/HIP/SYCL* → enables accurate & robust hydrodynamics

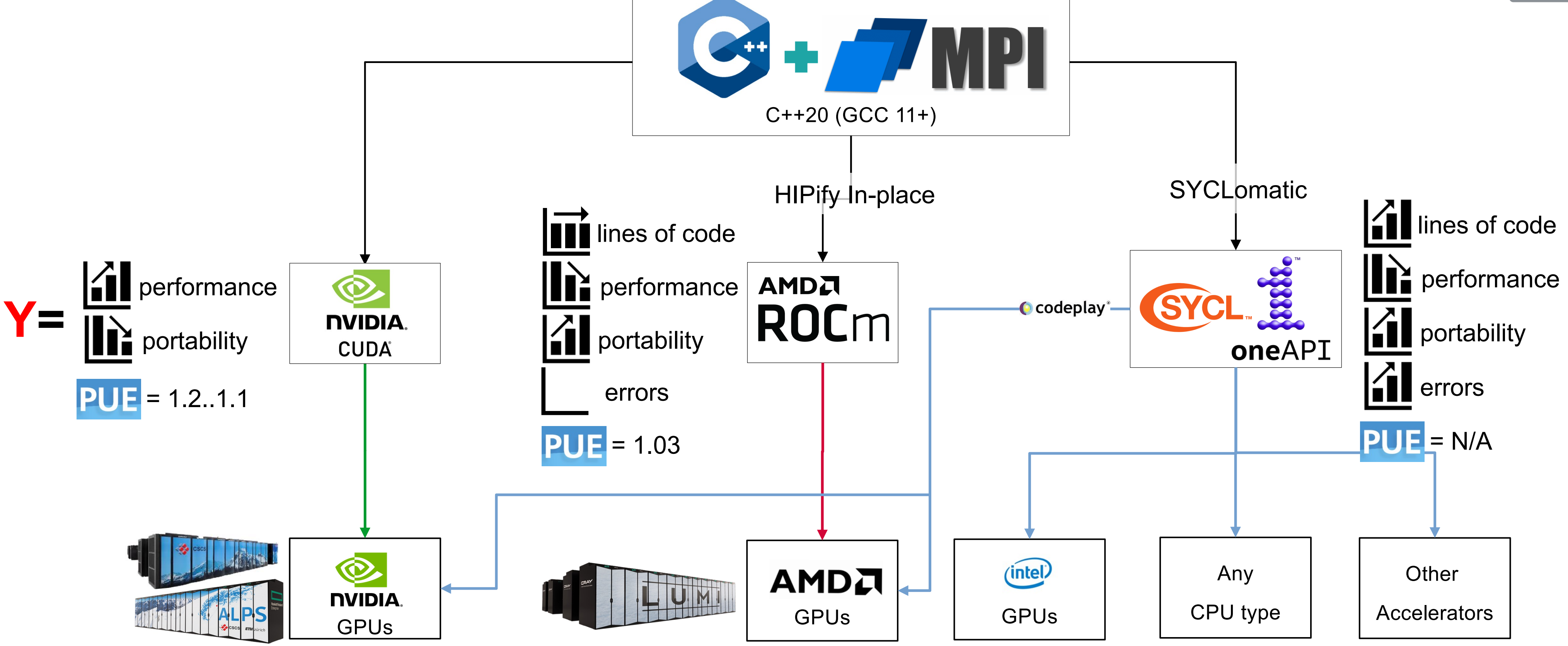
N-body Gravity-solver on GPUs with

- Cornerstone octrees
- Breadth-first traversal inspired by Bonsai (github.com/treecode/Bonsai)
- EXA-FMM multipole kernels (github.com/exafmm)
- **4'533** C++ LOC, Y= 2'137 CUDA/HIP/SYCL* → enables extreme scale Astro/Cosmo simulations

Work with: R. Cabezon, O. Simsek, L. Schmidt, L. Mayer, N. Kubli, D. Reed, S. Keller, J. G. Piccinali, J. Favre, J. Coles, D. García, A. Sanz

Performance ⚖ Portability ⚖ Sustainability

Modern Simulations (APIs)



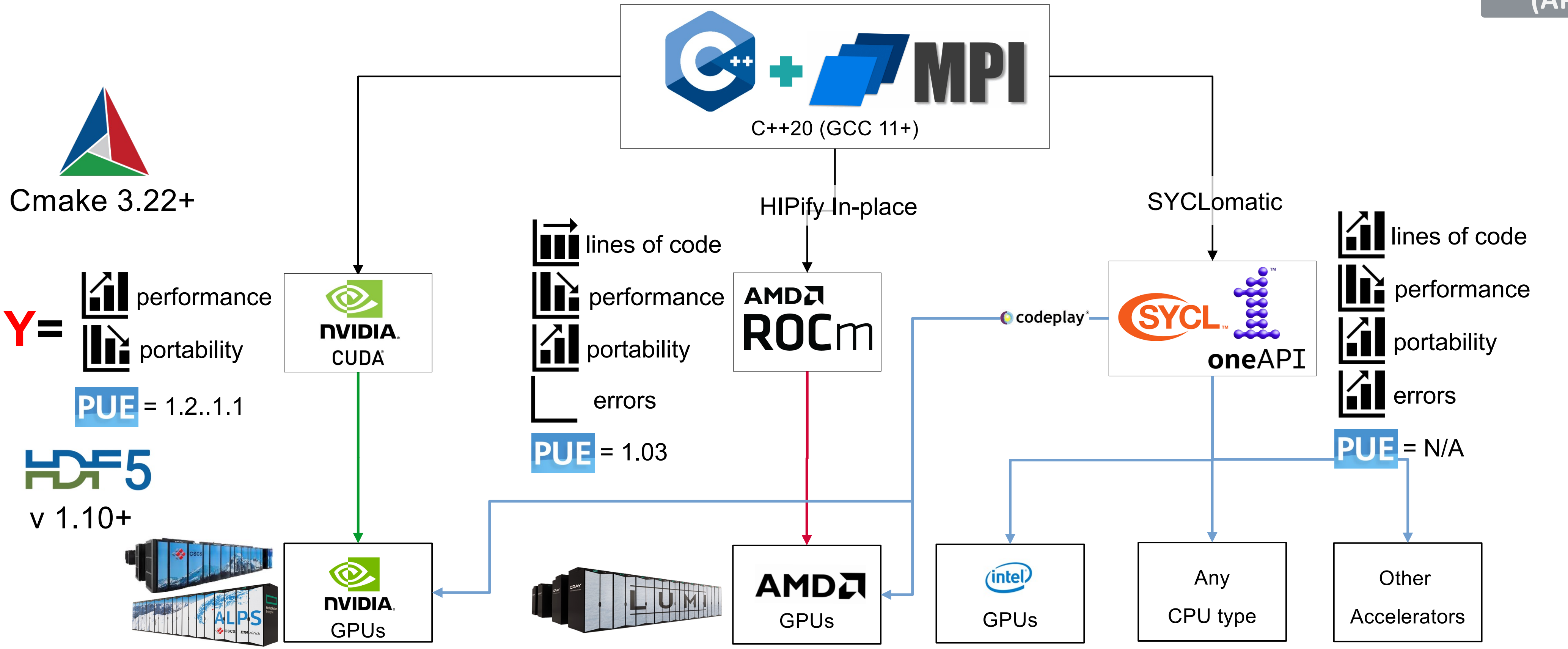
```
hipify-perl -inplace `find -name *.cu -o -name *.cuh` && find -name *.prehip -delete
```

```
dpct -p=compile_commands.json --in-root=Proj_dir  --out-root=out_dir --gen-helper-function
dpct -p=compile_commands.json --in-root=Proj_dir  --out-root=out_dir --migrate-build-script-only
```

(Image adapted from Robey and Zamora, Parallel and High Performance Computing, 2021, p. 40)

SPH EXA Performance ⚖ Portability ⚖ Sustainability

Modern Simulations (APIs)



```
hipify-perl -inplace `find -name *.cu -o -name *.cuh` && find -name *.prehip -delete
```

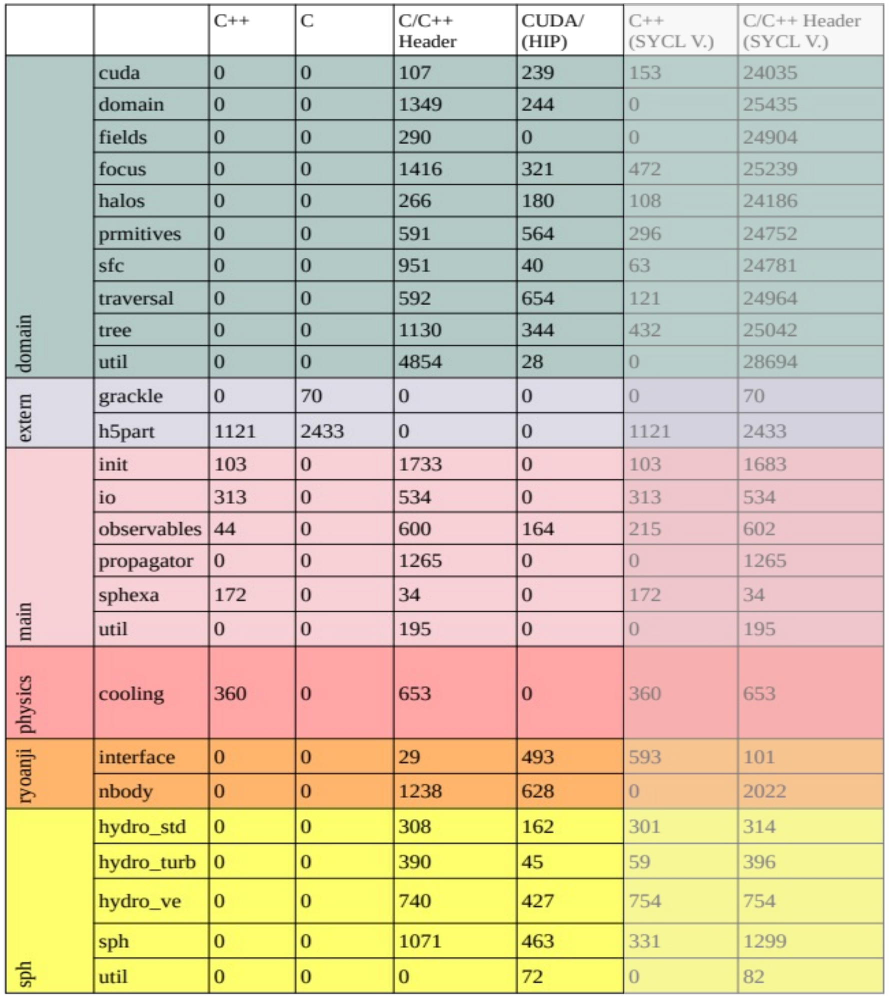
OPTIONAL [Ascent ParaView Catalyst ADIOS 2]

```
dpct -p=compile_commands.json --in-root=Proj_dir  --out-root=out_dir --gen-helper-function
dpct -p=compile_commands.json --in-root=Proj_dir  --out-root=out_dir --migrate-build-script-only
```

(Image adapted from Robey and Zamora, Parallel and High Performance Computing, 2021, p. 40)

Performance ⚖ Portability ⚖ Sustainability

Modern Simulations (APIs)



		C++	C	C/C++ Header	CUDA/ (HIP)	C++ (SYCL V.)	C/C++ Header (SYCL V.)
domain	cuda	0	0	107	239	153	24035
	domain	0	0	1349	244	0	25435
	fields	0	0	290	0	0	24904
	focus	0	0	1416	321	472	25239
	halos	0	0	266	180	108	24186
	prmitives	0	0	591	564	296	24752
	sfc	0	0	951	40	63	24781
	traversal	0	0	592	654	121	24964
	tree	0	0	1130	344	432	25042
	util	0	0	4854	28	0	28694
extern	grackle	0	70	0	0	0	70
	h5part	1121	2433	0	0	1121	2433
main	init	103	0	1733	0	103	1683
	io	313	0	534	0	313	534
	observables	44	0	600	164	215	602
	propagator	0	0	1265	0	0	1265
	sphexa	172	0	34	0	172	34
	util	0	0	195	0	0	195
physics	cooling	360	0	653	0	360	653
ryoanji	interface	0	0	29	493	593	101
	nbody	0	0	1238	628	0	2022
sph	hydro_std	0	0	308	162	301	314
	hydro_turb	0	0	390	45	59	396
	hydro_ve	0	0	740	427	754	754
	sph	0	0	1071	463	331	1299
	util	0	0	0	72	0	82

A Production Code
Easy to Use like a Miniapp

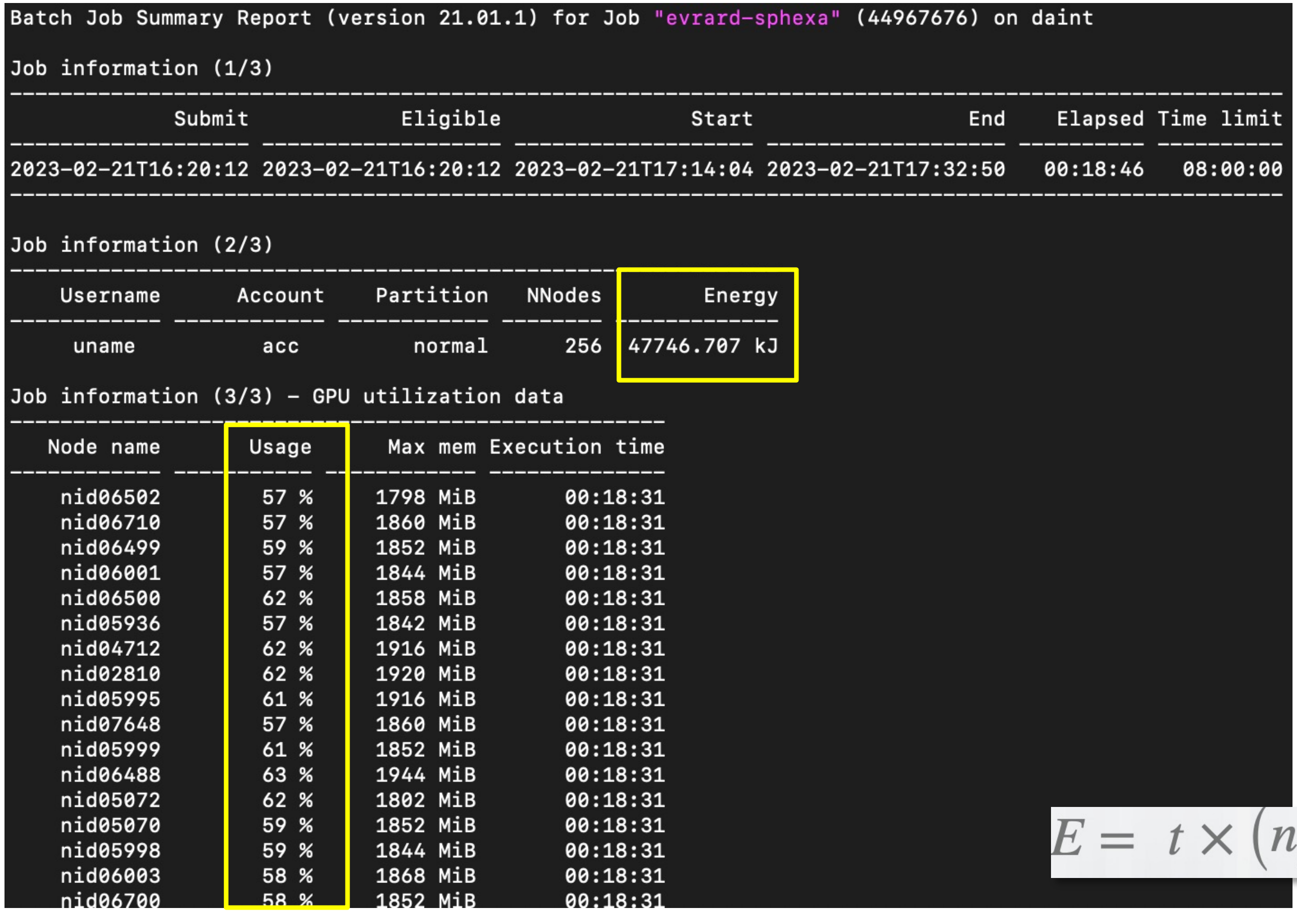
```
$> git clone https://github.com/unibas-dmi-hpc/SPH-EXA.git
$> cd SPH-EXA
$SPH-EXA> mkdir build
$SPH-EXA> cd build
$SPH-EXA/build> cmake ..
.
. Output
.
$SPH-EXA/build> make –j
.
. Output
.
$SPH-EXA/build> cd main/src/sphexa
$SPH-EXA/build/main/src/sphexa> ls
sphexa
sphexa-cuda
$SPH-EXA/build/main/src/sphexa>
```

```
./sphexa-cuda --init turbulence --prop turbulence -
n 200 -s 10.0 -w 1.0 --avclean -f
x,y,z,h,rho,vx,vy,vz,curlv
```

hydrodynamics and gravity

Measuring Energy Consumption on Piz Daint

Energy (kJ) needed for **Evrard Collapse** 1B particles, on Piz Daint 256 nodes and resource usage factor (u_c)



```
Batch Job Summary Report (version 21.01.1) for Job "evrard-sphexa" (44967676) on daint

Job information (1/3)
-----------------------------------------------------------------------------------------------------
             Submit            Eligible               Start                 End    Elapsed Time limit
------------------- ------------------- ------------------- ------------------- ---------- ----------
2023-02-21T16:20:12 2023-02-21T16:20:12 2023-02-21T17:14:04 2023-02-21T17:32:50   00:18:46   08:00:00
-----------------------------------------------------------------------------------------------------

Job information (2/3)
-------------------------------------------------------------------------
    Username      Account    Partition   NNodes        Energy
------------ ------------ ------------ -------- -------------
     uname          acc         normal      256 47746.707 kJ

Job information (3/3) - GPU utilization data
----------------------------------------------------
   Node name       Usage      Max mem Execution time
------------ ------------ ------------ --------------
    nid06502        57 %     1798 MiB       00:18:31
    nid06710        57 %     1860 MiB       00:18:31
    nid06499        59 %     1852 MiB       00:18:31
    nid06001        57 %     1844 MiB       00:18:31
    nid06500        62 %     1858 MiB       00:18:31
    nid05936        57 %     1842 MiB       00:18:31
    nid04712        62 %     1916 MiB       00:18:31
    nid02810        62 %     1920 MiB       00:18:31
    nid05995        61 %     1916 MiB       00:18:31
    nid07648        57 %     1860 MiB       00:18:31
    nid05999        61 %     1852 MiB       00:18:31
    nid06488        63 %     1944 MiB       00:18:31
    nid05072        62 %     1802 MiB       00:18:31
    nid05070        59 %     1852 MiB       00:18:31
    nid05998        59 %     1844 MiB       00:18:31
    nid06003        58 %     1868 MiB       00:18:31
    nid06700        58 %     1852 MiB       00:18:31
```

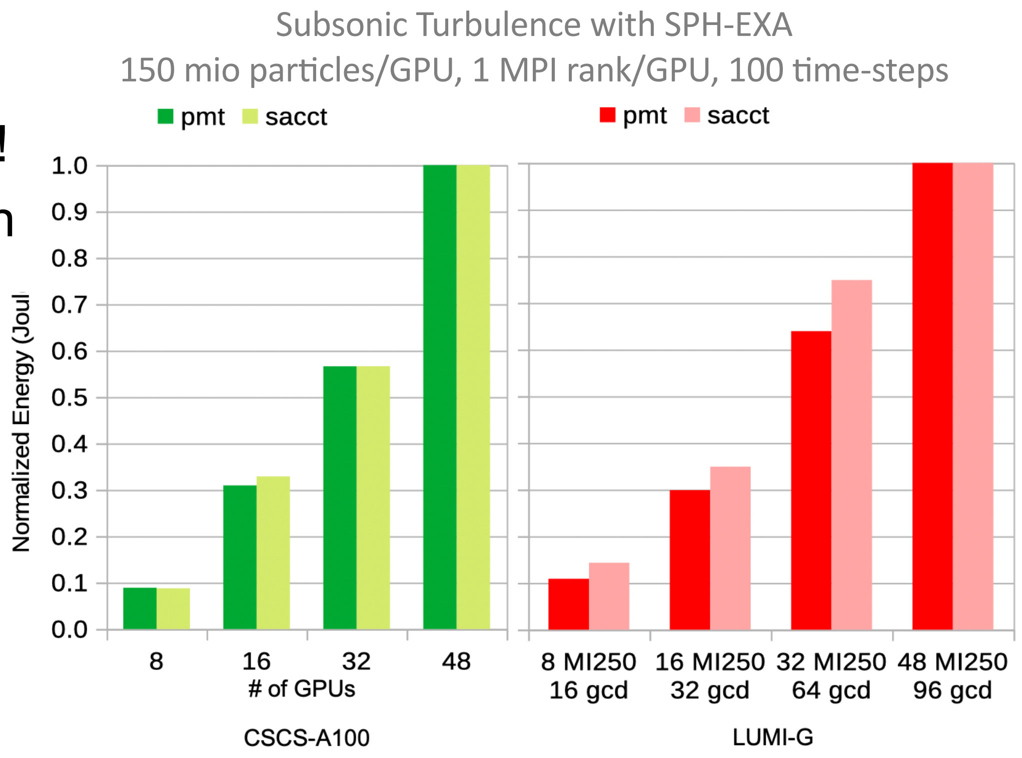
- Information is available
- Piz Daint (CSCS) energy only for entire jobs
- Which **subsystem** consumes the most energy during simulations?
  - CPU, GPU
  - Memory
  - Others (I/O, network, etc.)

$$E = t \times (n_c \times P_c \times u_c + n_m \times P_m) \times PUE$$

hydrodynamics

Measuring & Reporting Energy Consumption on CSCS & LUMI-G

- **PMT: Power Measurement Toolkit [1]**
  - C++ tool → compatible w/ various codes!
  - **Application view** of energy consumption (time-stepping loop in the code)
  - On Cray's: access GPU/CPU data from `/sys/cray/pm_counters`
  - On non-Cray's obtains GPU data via vendor tools & CPU data from RAPL
- **Slurm:** `sacct` retrieves job-level data
  - **Job (⊃ Application) view** of energy
  - `/sys/cray/pm_counters`
  - `AcctGatherEnergyType=acct_gather_energy/pm_counters`

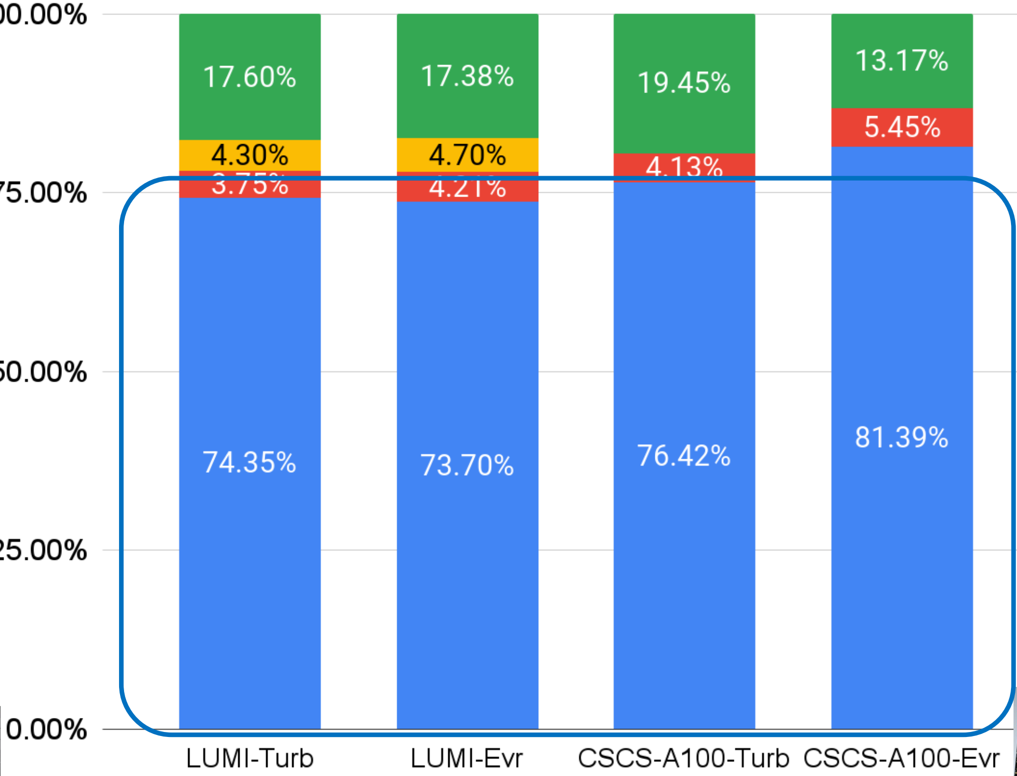


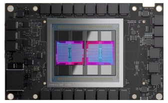
[2] PMT matches/slightly underestimates consumed energy compared to SLURM sacct

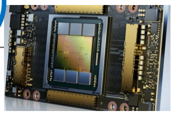
[1] S. Corda, B. Veenboer, E. Tolley, PMT: Power Measurement Toolkit. *HUST Workshop, 2022*

[2] O.S. Simsek, J.-G. Piccinali, F. M. Ciorba, Accurate Measurement of Application-level Energy Consumption for Energy-Aware Large-Scale Simulations. SusSup Workshop SC23

Measuring & Reporting Energy Consumption: Device Breakdown







Energy consumption by **GPU device**

- **Subsonic Turbulence** (hydro+turbo)
- **Evrard Collapse** (hydro+gravity)

100 time-steps each, on

- 4 LUMI-G (left)
- 8 A100 (right)

using 150 million and 80 million particles / GPU × 32 MPI ranks total.

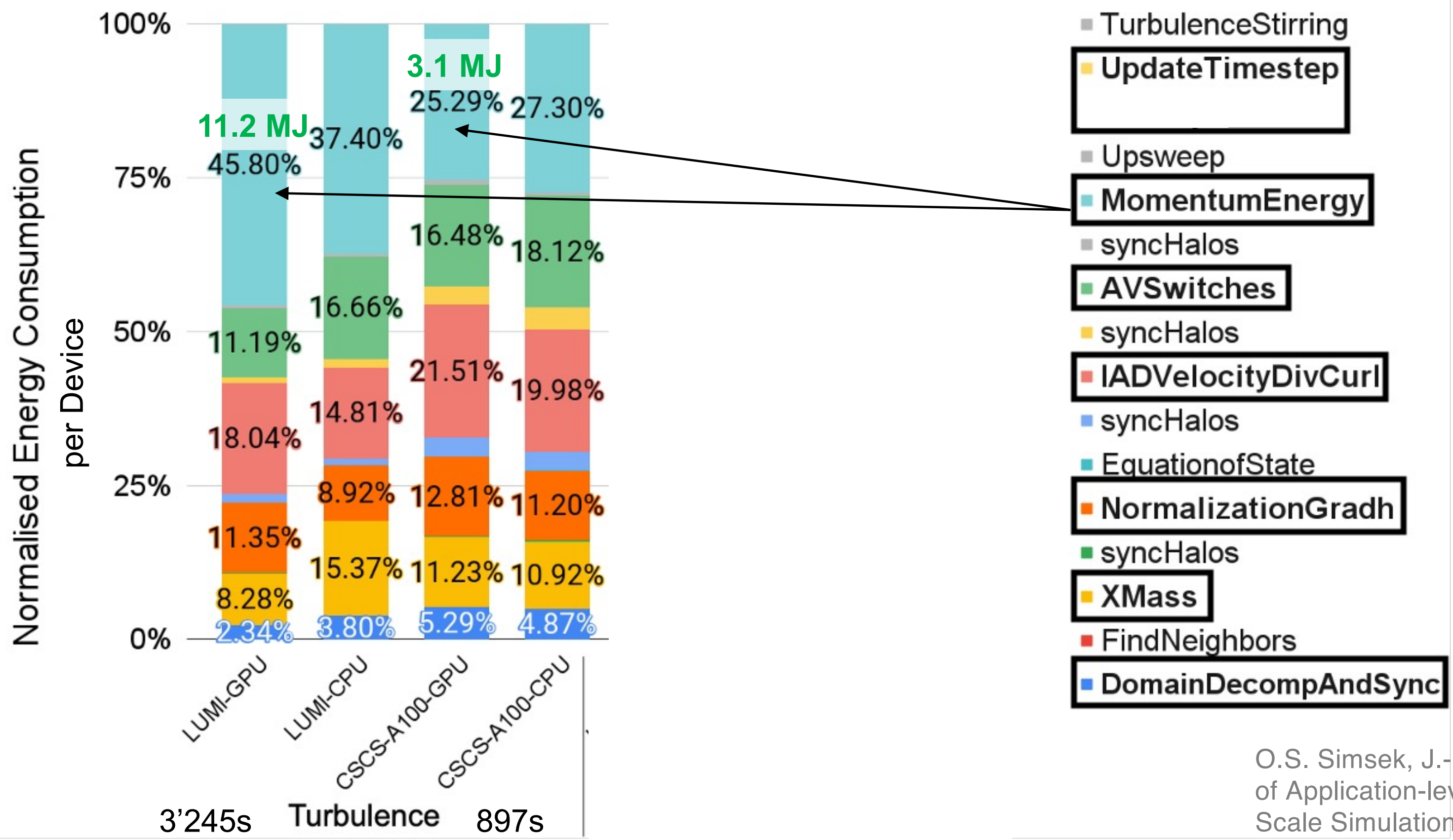
A100 does not give specific memory energy consumption → Other

$$E = t \times (n_c \times P_c \times u_c + n_m \times P_m) \times PUE$$

O.S. Simsek, J.-G. Piccinali, F. M. Ciorba, Accurate Measurement of Application-level Energy Consumption for Energy-Aware Large-Scale Simulations. SusSup Workshop SC23

Measuring & Reporting Energy Consumption: Functional Breakdown

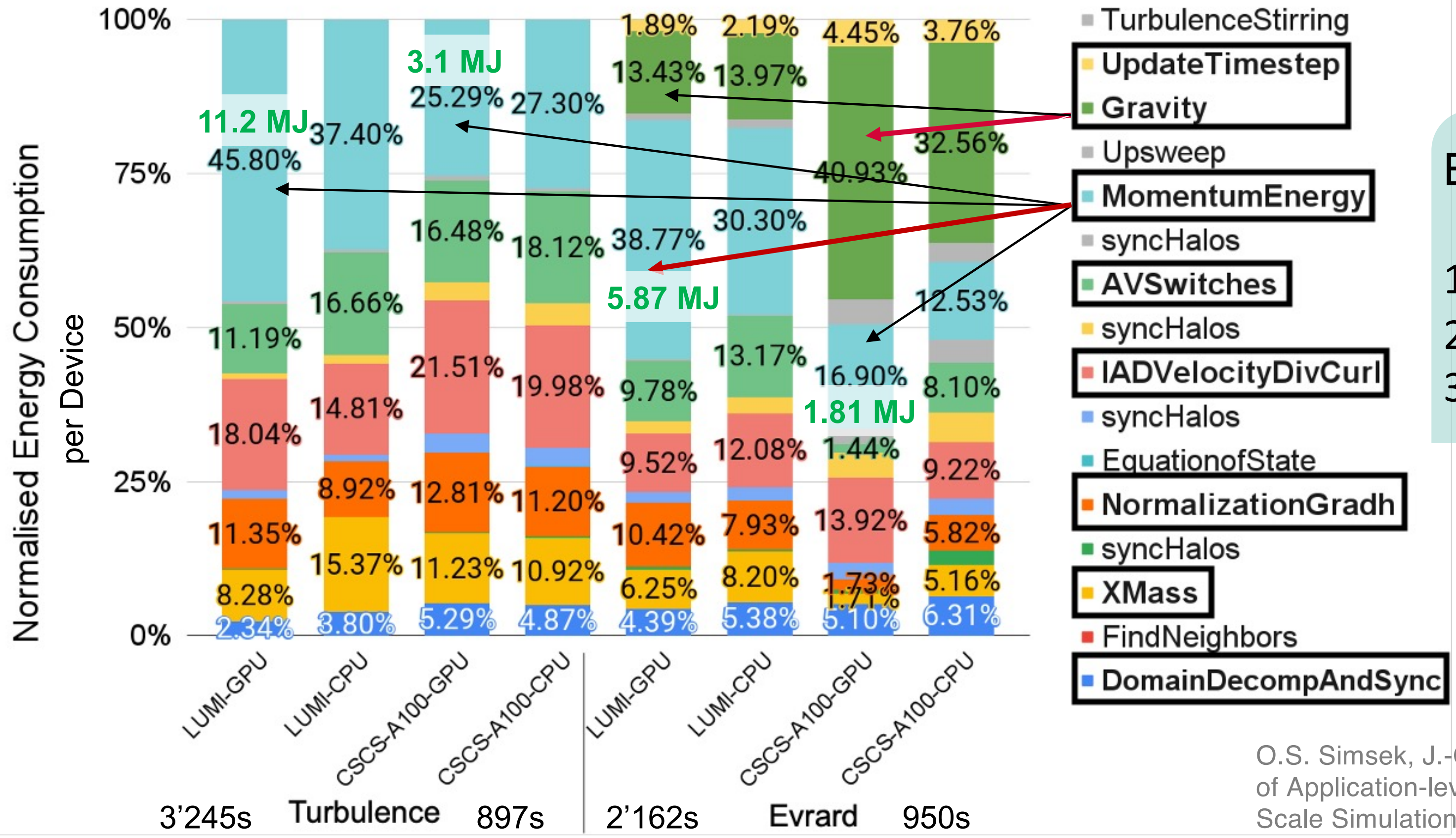
Energy / code function, 150 mio Turb | 80 Evr particles, 100 time-steps, 32 MPI Ranks: 4 LUMI-G | 8 CSCS-A100 Nodes



O.S. Simsek, J.-G. Piccinali, F. M. Ciorba, Accurate Measurement of Application-level Energy Consumption for Energy-Aware Large-Scale Simulations. SusSup Workshop SC23

Measuring & Reporting Energy Consumption: Functional Breakdown

Energy / code function, 150 mio Turb | 80 Evr particles, 100 time-steps, 32 MPI Ranks: 4 LUMI-G | 8 CSCS-A100 Nodes



Energy efficiency depends on

1 Simulation and its functions*
2 Architectural characteristics
3 Programming framework

* O.S. Simsek, J.-G. Piccinali, F. M. Ciorba, Increasing Energy Efficiency of Astrophysics Simulations Through GPU Frequency Scaling. SusSup Workshop SC24

O.S. Simsek, J.-G. Piccinali, F. M. Ciorba, Accurate Measurement of Application-level Energy Consumption for Energy-Aware Large-Scale Simulations. SusSup Workshop SC23

Sustainability: Estimated Energy Needed for SPH-EXA on LUMI

“Hero” Run on LUMI-G (Finland)

12 hours on Dec 19, 2022

2’052   AMD EPYC 7A53 CPUs
16’416 AMD Instinct MI250X 64GB GPUs

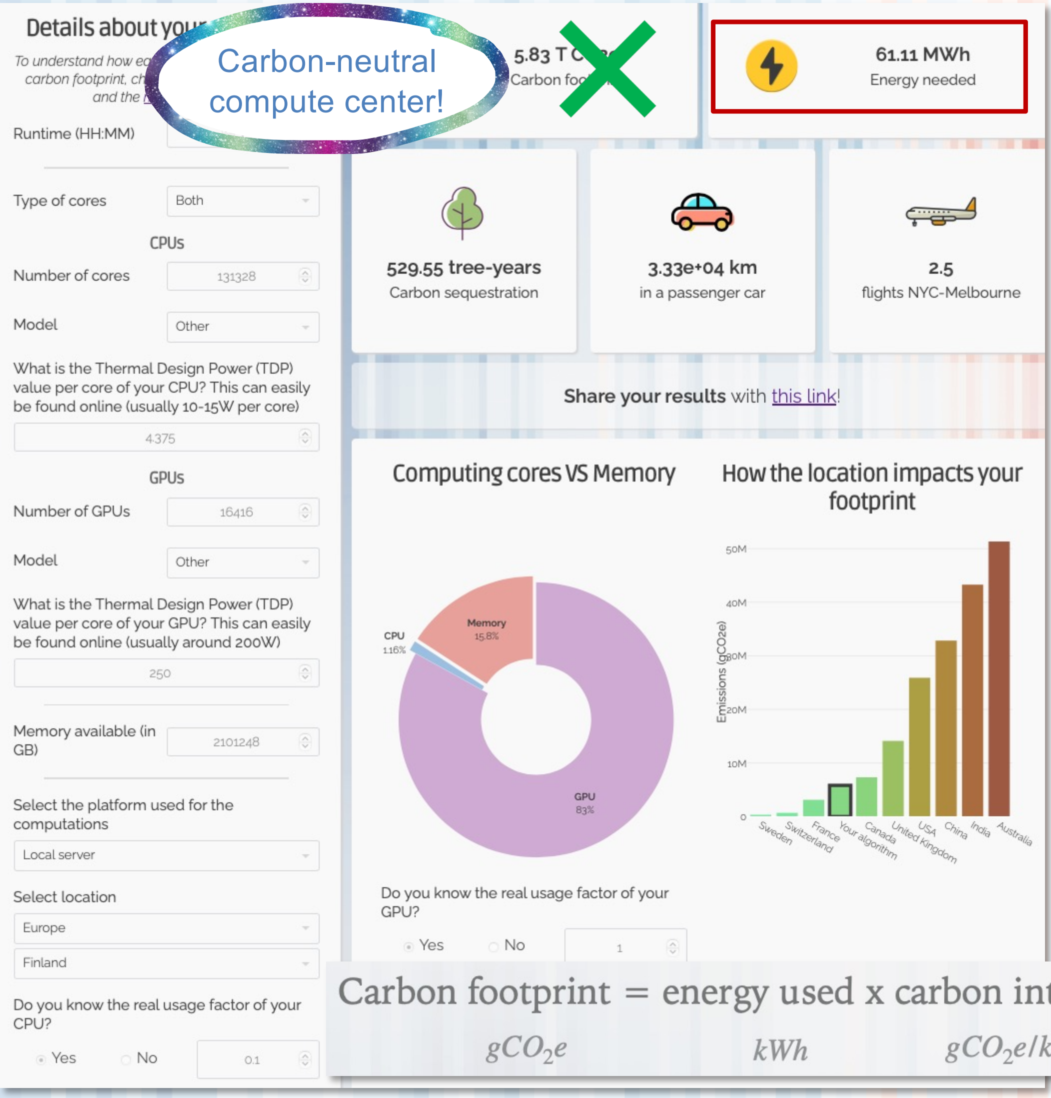
PUE = 1.03

SPH-EXA Tests

- **Turbulence simulation** (hydro+turbo)
- **Pure gravity weak scaling** (gravity – not in this talk)

Carbon intensity matters

Calculator: https://www.green-algorithms.org



Details about you

To understand how ec carbon footprint, ch and the

Runtime (HH:MM)

Type of cores: Both

CPUs

Number of cores: 131328

Model: Other

What is the Thermal Design Power (TDP) value per core of your CPU? This can easily be found online (usually 10-15W per core)

4.375

GPUs

Number of GPUs: 16416

Model: Other

What is the Thermal Design Power (TDP) value per core of your GPU? This can easily be found online (usually around 200W)

250

Memory available (in GB): 2101248

Select the platform used for the computations: Local server

Select location: Europe / Finland

Do you know the real usage factor of your CPU? ● Yes ○ No 0.1

Carbon-neutral compute center!

5.83 T C
Carbon foo

61.11 MWh
Energy needed

529.55 tree-years
Carbon sequestration

3.33e+04 km
in a passenger car

2.5
flights NYC-Melbourne

Share your results with this link!

Computing cores VS Memory

How the location impacts your footprint

Do you know the real usage factor of your GPU? ● Yes ○ No 1

Carbon footprint = energy used x carbon intensity

gCO_2e kWh gCO_2e/kWh

Sustainability: Estimated Energy Needed for SPH-EXA on Frontier

“Hero” Run on Frontier (USA)

A **hypothetical** 12-hour run

2’052 AMD EPYC 7A53 CPUs
16’416 AMD Instinct MI250X 64GB GPUs

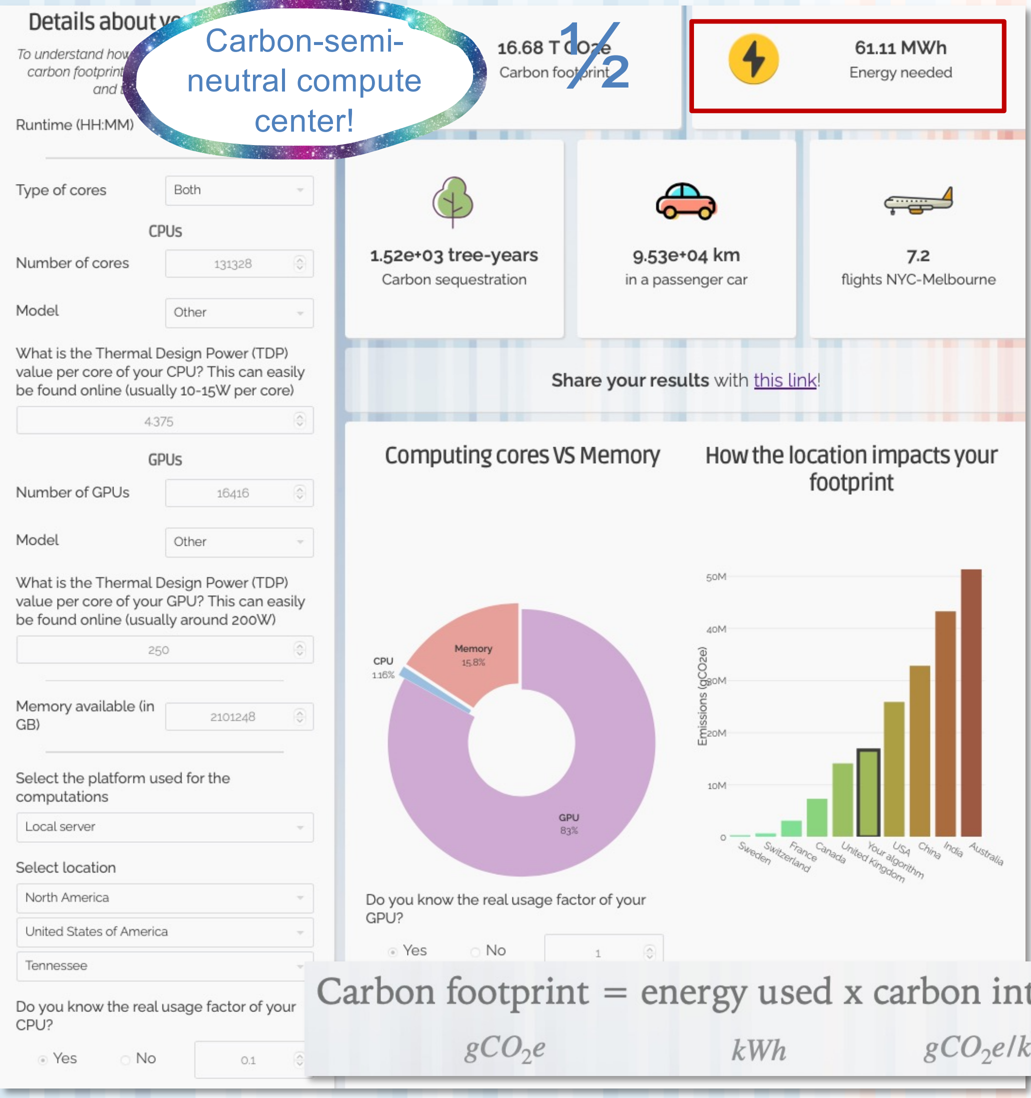
PUE = 1.03 (hypothetical)

SPH-EXA Tests (same size/length runs)

- **Turbulence simulation** (hydro+turbo)
- **Pure gravity weak scaling** (gravity – not in this talk)

On carbon-positive systems, the cost of
++ parallelism adds up carbon footprint ☹
Need to take this into account!

Calculator: https://www.green-algorithms.org



Details about yo

To understand how carbon footprint and

Runtime (HH:MM)

Type of cores: Both

CPUs

Number of cores: 131328

Model: Other

What is the Thermal Design Power (TDP) value per core of your CPU? This can easily be found online (usually 10-15W per core)

4.375

GPUs

Number of GPUs: 16416

Model: Other

What is the Thermal Design Power (TDP) value per core of your GPU? This can easily be found online (usually around 200W)

250

Memory available (in GB): 2101248

Select the platform used for the computations

Local server

Select location

North America

United States of America

Tennessee

Do you know the real usage factor of your CPU?

● Yes ○ No 0.1

Do you know the real usage factor of your GPU?

● Yes ○ No 1

16.68 T CO2e ½
Carbon footprint

61.11 MWh
Energy needed

1.52e+03 tree-years
Carbon sequestration

9.53e+04 km
in a passenger car

7.2
flights NYC-Melbourne

Share your results with this link!

Computing cores VS Memory

How the location impacts your footprint

Carbon-semi-neutral compute center!

Carbon footprint = energy used x carbon intensity

gCO_2e kWh gCO_2e/kWh

Sustainability: Estimated Energy Needed for SPH-EXA on Piz Daint

“Hero” Run on Piz Daint (Switzerland)

A **hypothetical** 12-hour run

2’052 Intel® Xeon® E5-2690 v3 CPUs
2’052 NVIDIA® Tesla® P100 16GB GPUs

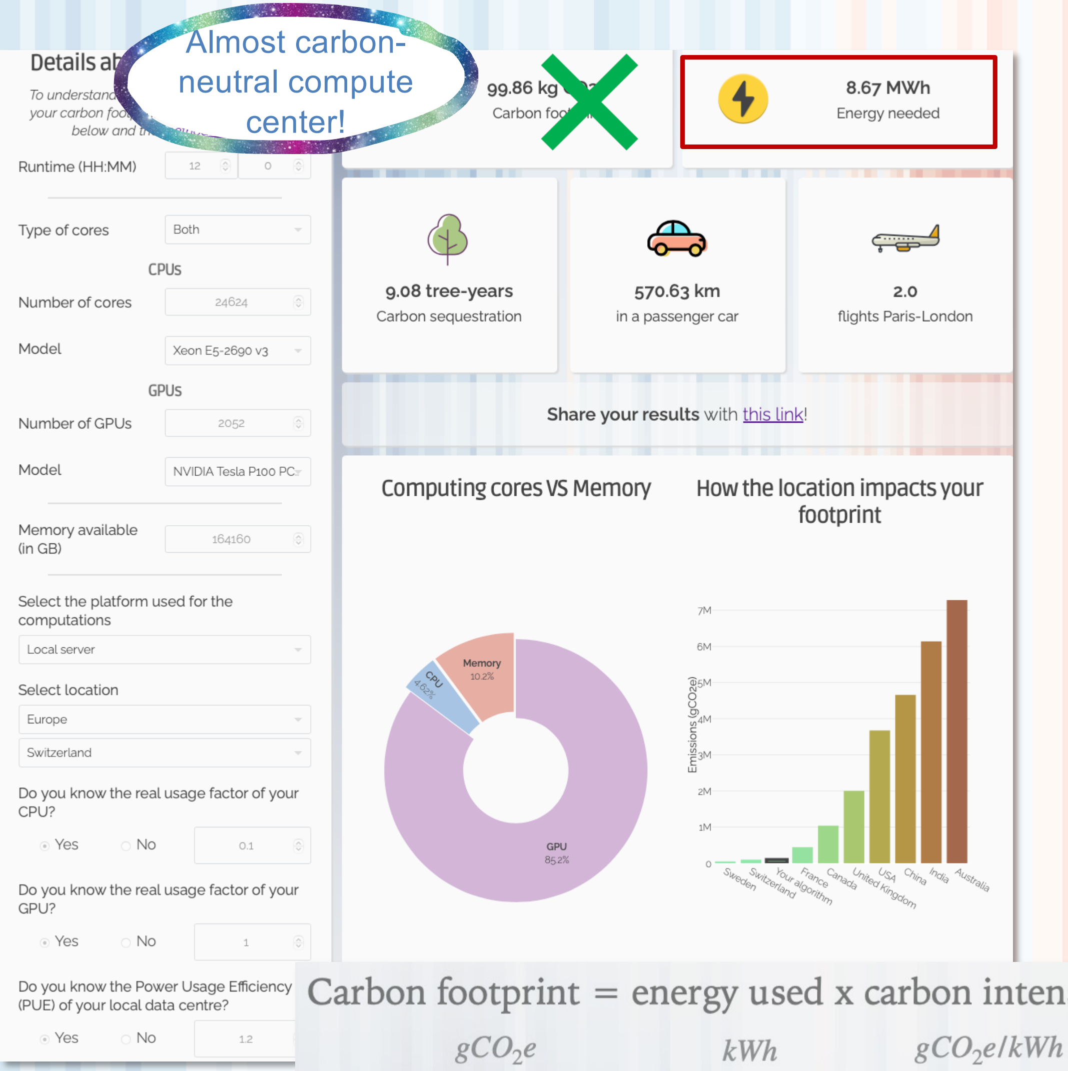
PUE = 1.20

SPH-EXA Tests (smaller runs than LUMI)

- **Turbulence simulation** (hydro+turbo)
- **Pure gravity weak scaling** (gravity – not in this talk)

Carbon intensity matters

Calculator: https://www.green-algorithms.org



Almost carbon-neutral compute center!

Details ab

To understand
your carbon foo
below and th

Runtime (HH:MM) 12 0

Type of cores: Both

CPUs

Number of cores: 24624

Model: Xeon E5-2690 v3

GPUs

Number of GPUs: 2052

Model: NVIDIA Tesla P100 PC...

Memory available (in GB): 164160

Select the platform used for the computations: Local server

Select location: Europe; Switzerland

Do you know the real usage factor of your CPU? ● Yes ○ No 0.1

Do you know the real usage factor of your GPU? ● Yes ○ No 1

Do you know the Power Usage Efficiency (PUE) of your local data centre? ● Yes ○ No 1.2

99.86 kg
Carbon foo

8.67 MWh
Energy needed

9.08 tree-years
Carbon sequestration

570.63 km
in a passenger car

2.0
flights Paris-London

Share your results with this link!

$$\underset{gCO_2e}{\text{Carbon footprint}} = \underset{kWh}{\text{energy used}} \times \underset{gCO_2e/kWh}{\text{carbon intensity}}$$

Sustainability: Estimated Energy Needed for SPH-EXA on Alps

“Hero” Run on Alps (Switzerland)

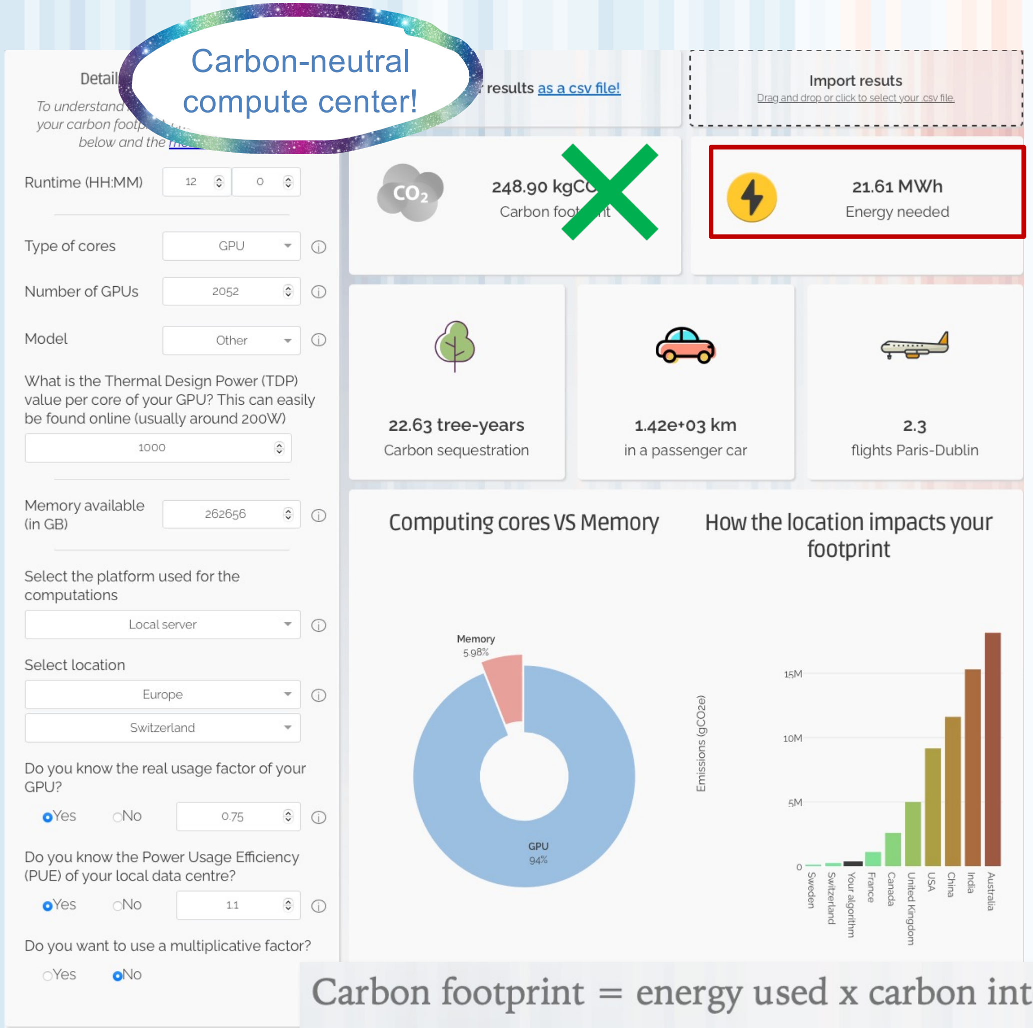
A **hypothetical** 12-hour run

2’052 Nvidia GH200, each with 72 ARM CPU cores + 1 Nvidia H100 GPU with 96 GB HBM3 memory
PUE ≅ 1.10

SPH-EXA Tests (larger runs than LUMI)
- **Turbulence simulation** (hydro+turbo)
- **Pure gravity weak scaling** (gravity – not in this talk)

Architecture matters

Calculator: https://www.green-algorithms.org



Carbon-neutral compute center!

Detail

To understand your carbon footp below and the

Runtime (HH:MM) 12 0

Type of cores GPU

Number of GPUs 2052

Model Other

What is the Thermal Design Power (TDP) value per core of your GPU? This can easily be found online (usually around 200W)

1000

Memory available (in GB) 262656

Select the platform used for the computations

Local server

Select location

Europe

Switzerland

Do you know the real usage factor of your GPU?

● Yes ○ No 0.75

Do you know the Power Usage Efficiency (PUE) of your local data centre?

● Yes ○ No 1.1

Do you want to use a multiplicative factor?

○ Yes ● No

r results as a csv file!

Import resuts
Drag and drop or click to select your .csv file.

248.90 kgCO
Carbon foot nt

21.61 MWh
Energy needed

22.63 tree-years
Carbon sequestration

1.42e+03 km
in a passenger car

2.3
flights Paris-Dublin

$$\underbrace{\text{Carbon footprint}}_{gCO_2e} = \underbrace{\text{energy used}}_{kWh} \text{ x } \underbrace{\text{carbon intensity}}_{gCO_2e/kWh}$$

TGSF: The role of Turbulence and Gravity in Star Formation

Hydrodynamics, turbulence, and gravity

Extreme Scale Access

Allocation: **22,000,000 GPUh*** on LUMI-G
Duration: 12 months, extended to 21 Feb 25
***Largest allocation in Europe to date.**

Swiss scientists win time on top European supercomputer

in **News** on 8 January 2024

https://skach.org/2024/01/08/swiss-scientists-win-time-on-top-european-supercomputer/

Objectives

Relevant to "Our Galaxy" and "Cradle of Life" SKAO WGs

Cosmology & Astrophysics

- Study the formation of stellar cores and their initial mass function at unprecedented resolution
- Study turbulent transport and mixing
- Contribute to the general theory of turbulence (Lyapunov exponents)

Scalability limitation for previous codes

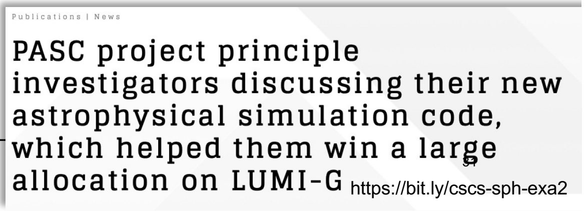
More natural with Lagrangian codes

Recall that Pretraining Llama 2 Model Family took **3.3M GPUh**

Computer Science

- Study the load imbalance, performance, and energy consumption at unprecedented scales
- Study large scale techniques for checkpointing, compression, and visualization

HPC research at extreme scale



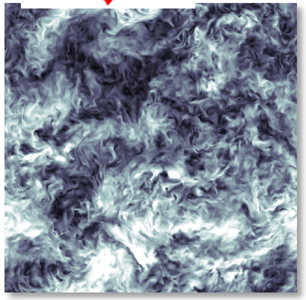
Publications | News

PASC project principle investigators discussing their new astrophysical simulation code, which helped them win a large allocation on LUMI-G

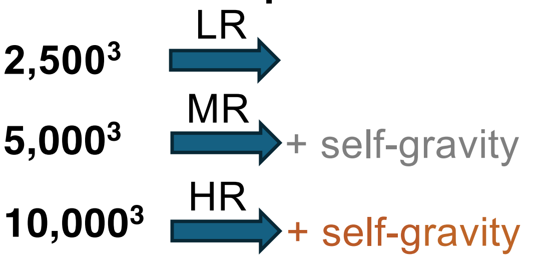
https://bit.ly/cscs-sph-exa2

TGSF: Simulations Planned and Realized on LUMI (MPI+Y=HIP)

Hydrodynamics, turbulence, and gravity



Simulations plan executed



Individual time-stepping (ITS) boost performance and reduces the need to use particle splitting for initial conditions.

A new set of simulations has been executed, setting and relaxing the ICs directly in the target resolution.

Table: Planned (estimated w/o ITS) resource usage (Node-hours) vs realized (w/ ITS) simulation sets.

Resolution	Physical Time (s)	Nodes	Estimated (Nh)	Measured (Nh)	ITS + Opt. Measured (Nh)	Simulation Type
2’520^3	0→0.875	128	11’654	30’131	NA	Hydro+Turbo+Gravity
2’500^3	0→1.250	64	NA	NA	5’855.66	Hydro+Turbulence
5’040^3	0→0.040	192	11’654	20’800	NA	Hydro+Turbulence
5’000^3	***0→0.040***	***512***	**NA**	**NA**	***325.40***	**Hydro+Turbulence**
5’000^3	*0→0.875*	*512*	(5’040^3) 224’400	*NA*	*76’400.35*	Hydro+Turbulence
5’000^3	0→1.250	512	NA	NA	112’439.89	Hydro+Turbulence → E*
5’000^3	**1.250→1.328**	**512**	NA	NA	**15’658.24**	**Hydro+Turbo+Gravity → E***
10’079^3	*0→0.040*	*1’024*	*220’000*	*NA*	*NA*	*Hydro+Turbulence*
10’000^3	**0→0.040**	**1’000**	**NA**	**NA**	**7’086.94**	**Hydro+Turbulence**
10’000^3	0→0.072571	1’000	NA	NA	22’251.67	Hydro+Turbulence → E*
10’000^3	*0→1.250*	*1’024*	*5’442’198*		**Est. 392’473.00**	Hydro+Turbo
10’000^3	*1.250→1.328*	*1’024*	*437’620*	NA	Est. 31’560.00	Hydro+Turbo+Gravity

204'823.88 Nh

654'371.61 Nh

13.86× less than initial plan

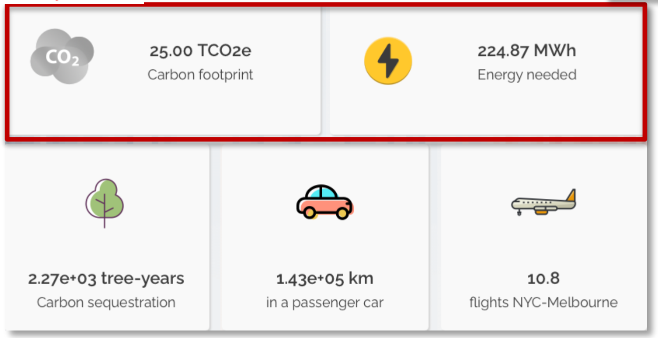
Mach=4, physical time ≥ 0.75 seconds

Each LUMI-G node has 4 GPUs / **8 GCDs**.

TGSF: The Energy Cost for Our **Selected** Simulations on LUMI (carbon data 2022)

Hydrodynamics, turbulence, and gravity

LUMI is carbon neutral and very energy efficient (PUE = 1.03)



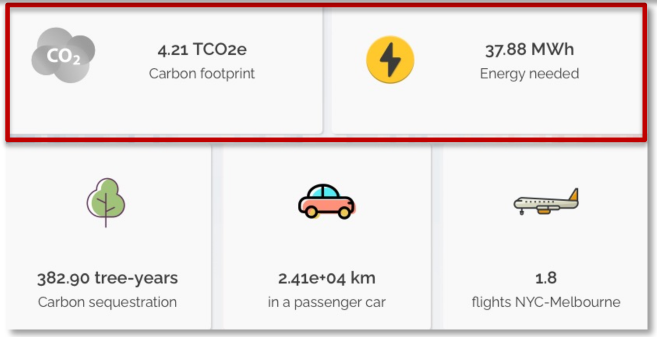
25.00 TCO2e
Carbon footprint

224.87 MWh
Energy needed

2.27e+03 tree-years
Carbon sequestration

1.43e+05 km
in a passenger car

10.8
flights NYC-Melbourne



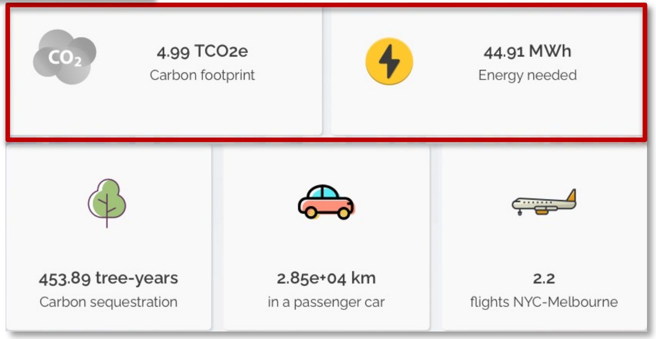
4.21 TCO2e
Carbon footprint

37.88 MWh
Energy needed

382.90 tree-years
Carbon sequestration

2.41e+04 km
in a passenger car

1.8
flights NYC-Melbourne



4.99 TCO2e
Carbon footprint

44.91 MWh
Energy needed

453.89 tree-years
Carbon sequestration

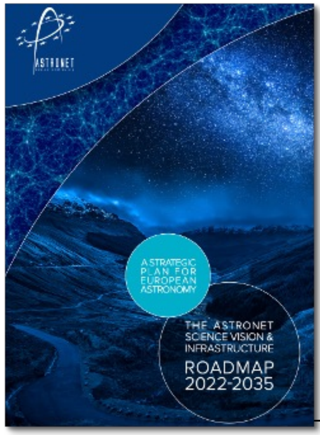
2.85e+04 km
in a passenger car

2.2
flights NYC-Melbourne

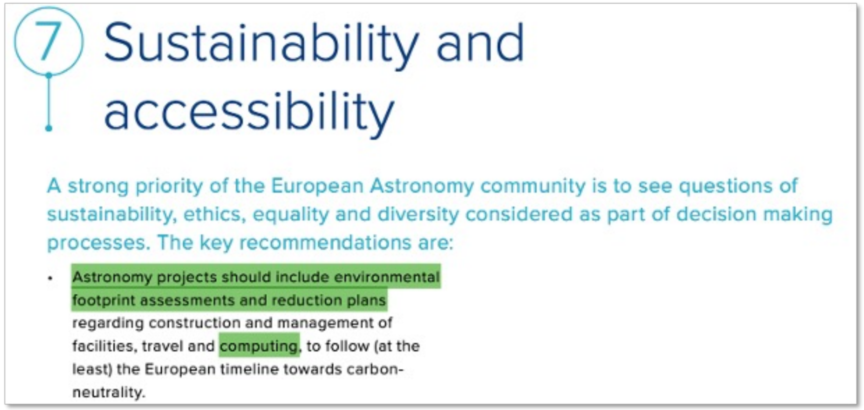
- Mid resolution TGSF run w/ ITS
  - **5’000^3** particles
  - 548’021 time-steps executed
  - 4’096 MPI ranks on **512 nodes**
  - Reaching 0→1.25 physical seconds and needing **112’439.89 Nh**

- Mid resolution TGSF run w/ ITS + self gravity
  - **5’000^3** particles
  - 41’069 time-steps executed
  - 4’096 MPI ranks on **512 nodes**
  - Reaching 1.250→1.328 physical seconds and needing **15’658.24 Nh**

- High resolution TGSF run w/ ITS
  - **10’000^3** particles
  - 5’320 time-steps executed
  - 8’000 MPI ranks on **1’000 nodes**
  - Reaching 0→0.0725 physical seconds and needing **22’251.67 Nh**



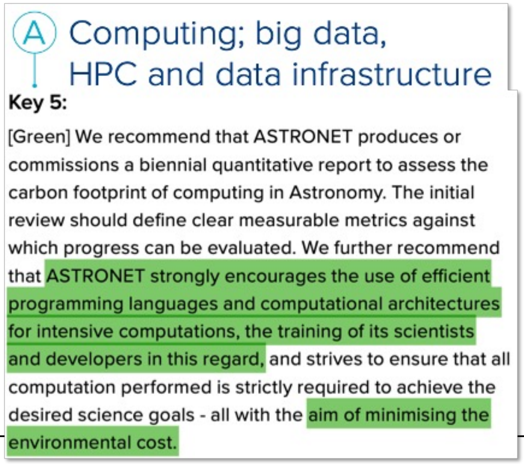
ASTRONET
A STRATEGIC PLAN FOR EUROPEAN ASTRONOMY
THE ASTRONET SCIENCE VISION & INFRASTRUCTURE
ROADMAP 2022-2035



7 Sustainability and accessibility

A strong priority of the European Astronomy community is to see questions of sustainability, ethics, equality and diversity considered as part of decision making processes. The key recommendations are:

- Astronomy projects should include environmental footprint assessments and reduction plans regarding construction and management of facilities, travel and computing, to follow (at the least) the European timeline towards carbon-neutrality.



A Computing; big data, HPC and data infrastructure

Key 5:

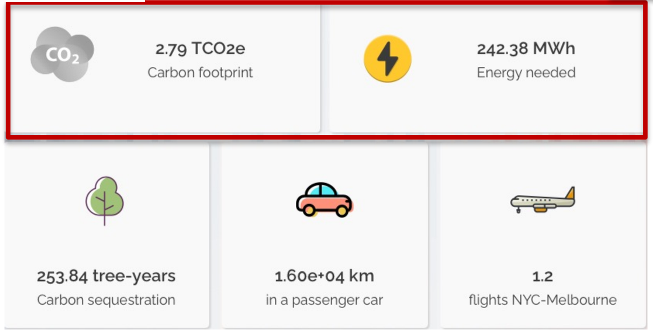
[Green] We recommend that ASTRONET produces or commissions a biennial quantitative report to assess the carbon footprint of computing in Astronomy. The initial review should define clear measurable metrics against which progress can be evaluated. We further recommend that ASTRONET strongly encourages the use of efficient programming languages and computational architectures for intensive computations, the training of its scientists and developers in this regard, and strives to ensure that all computation performed is strictly required to achieve the desired science goals - all with the aim of minimising the environmental cost.

Calculator: https://www.green-algorithms.org https://www.astronet-eu.org

TGSF: The Energy Cost for Our **Hypothetical** Simulations on Alps (carbon data 2022)

Hydrodynamics, turbulence, and gravity

Alps is also carbon neutral and very energy efficient (PUE ≅ 1.1)



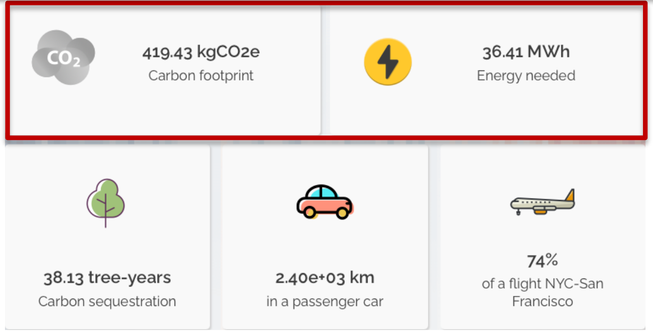
2.79 TCO2e
Carbon footprint

242.38 MWh
Energy needed

253.84 tree-years
Carbon sequestration

1.60e+04 km
in a passenger car

1.2
flights NYC-Melbourne



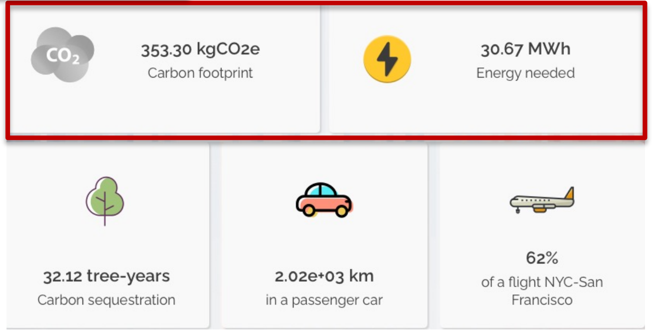
419.43 kgCO2e
Carbon footprint

36.41 MWh
Energy needed

38.13 tree-years
Carbon sequestration

2.40e+03 km
in a passenger car

74%
of a flight NYC-San Francisco



353.30 kgCO2e
Carbon footprint

30.67 MWh
Energy needed

32.12 tree-years
Carbon sequestration

2.02e+03 km
in a passenger car

62%
of a flight NYC-San Francisco

- **Hypothetical** Mid resolution TGSF w/ ITS
  - **5'000³** particles
  - 548'021 time-steps executed
  - 500 MPI ranks on **125 nodes**
  - Reaching 0→"1.25" physical seconds and needing **58'253.63 Nh**

- Hypothetical Mid resolution TGSF w/ ITS + self gravity
  - **5'000³** particles
  - 41'069 time-steps executed
  - 500 MPI ranks on **125 nodes**
  - Reaching "1.250"→"1.328" physical seconds and needing **10'592.80 Nh**

- **Hypothetical** High resolution TGSF w/ ITS
  - **10'000³** particles
  - 5'320 time-steps executed
  - 4'000 MPI ranks on **1'000 nodes**
  - Reaching 0→"0.0725" physical seconds and needing **8'999.51 Nh**

ASTRONET
A STRATEGIC PLAN FOR EUROPEAN ASTRONOMY
THE ASTRONET SCIENCE VISION & INFRASTRUCTURE ROADMAP 2022-2035

7 Sustainability and accessibility

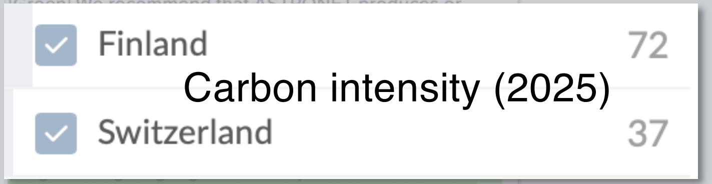
Faster, longer, and more sustainable simulations

regarding construction and management of facilities, travel and computing, to follow (at the least) the European timeline towards carbon-neutrality.

A Computing; big data, HPC and data infrastructure

Key 5:

[Green] We recommend that ASTRONET produces or



☑ Finland 72

☑ Switzerland 37

Carbon intensity (2025)

for intensive computations, the training of its scientists and developers in this regard, and strives to ensure that all computation performed is strictly required to achieve the desired science goals - all with the aim of minimising the environmental cost.

35

Calculator: https://www.green-algorithms.org https://www.astronet-eu.org

University of Basel

An Autonomy Loop for Dynamic HPC Job Time Limit Adjustment

University of Basel

Thomas Jakobsche
Osman S. Simsek
Florina M. Ciorba

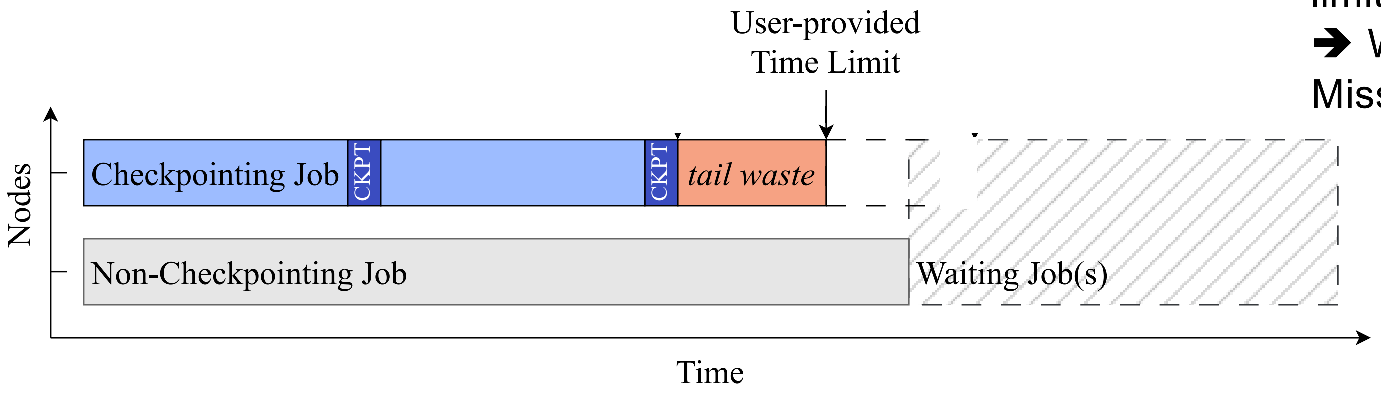
Sandia National Laboratories

A. Gentile
J. Brandt

To be presented at
31st International European Conference on Parallel and Distributed Computing, August 2025

Preprint: https://arxiv.org/abs/2505.05927

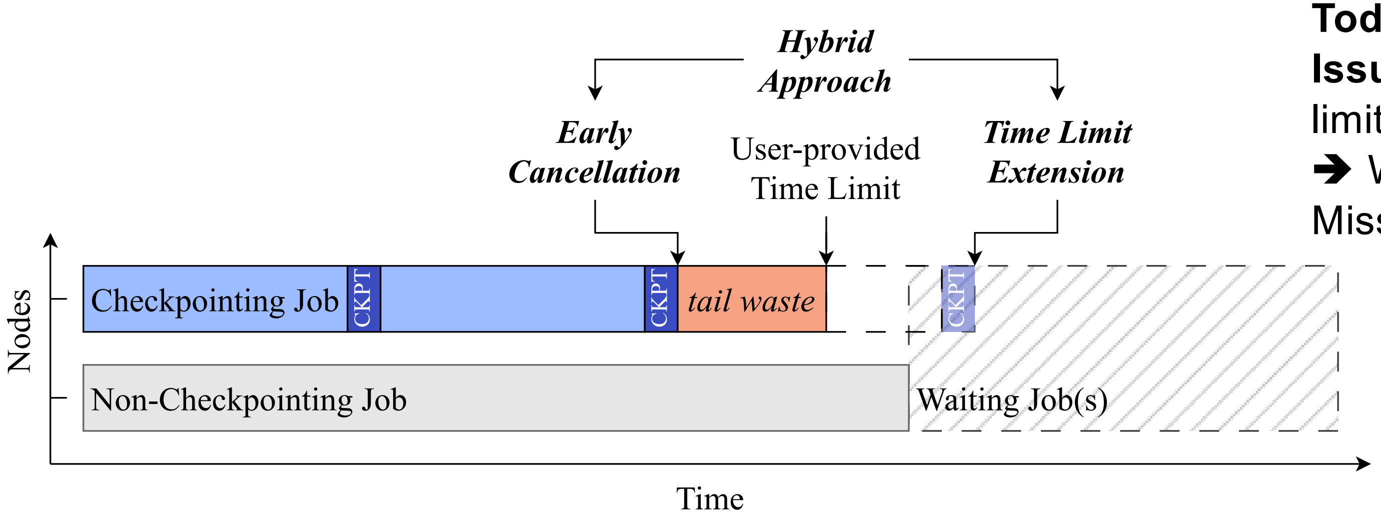
Inaccurate Job Time Estimates ➔ Inefficient Resource Use in HPC



Today: Static, user-provided time limits
Issue: Misalignment between the job's time limit and the checkpointing schedule.
➔ Wasted Computation ("Tail Waste") and Missed Checkpoints

min *tail_waste*:
Dynamically Adjust Time Limits Using Application Feedback

Inaccurate Job Time Estimates ➔ Inefficient Resource Use in HPC



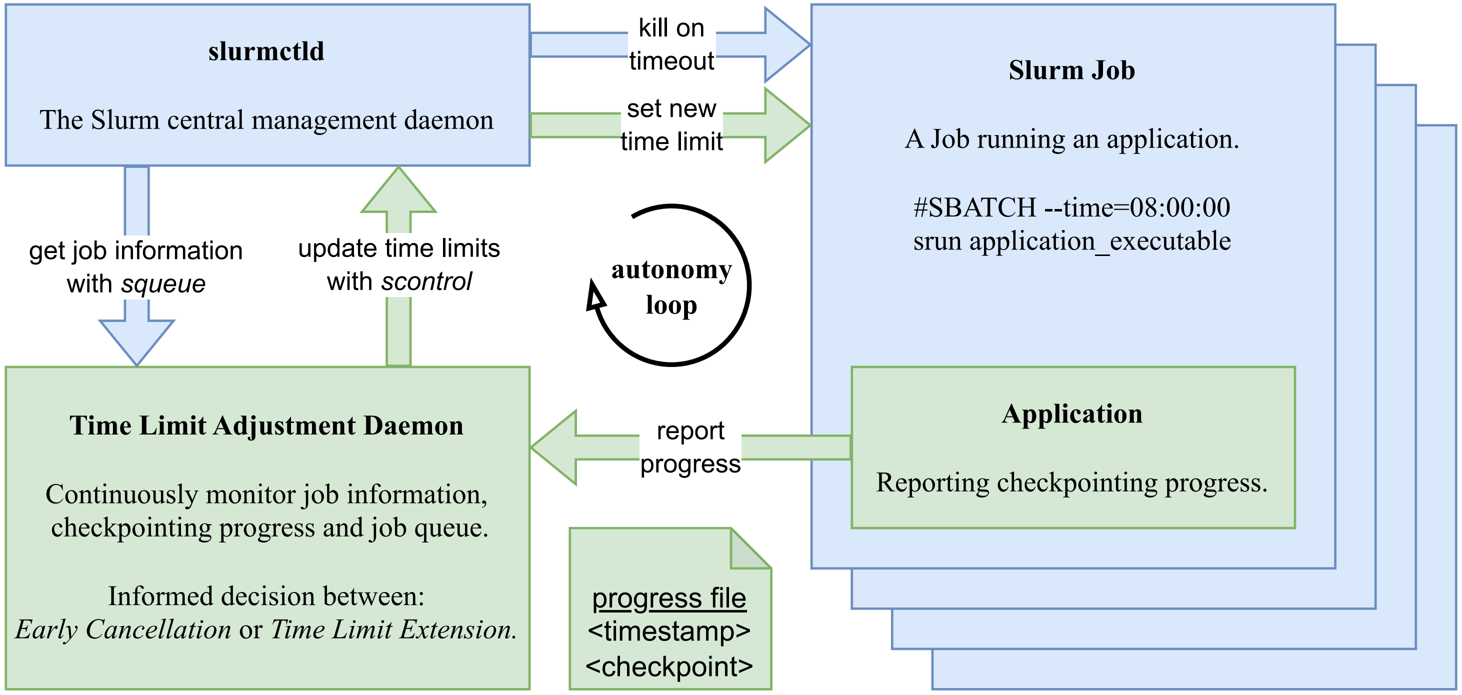
Today: Static, user-provided time limits
Issue: Misalignment between the job's time limit and the checkpointing schedule.
➔ Wasted Computation ("Tail Waste") and Missed Checkpoints

min *tail_waste*:
Dynamically Adjust Time Limits Using Application Feedback

Adjustment Policies

Early Cancellation: Cancels job *after last successful checkpoint* to avoid tail waste.
Time Limit Extension: Extends job time limit to *reach the next checkpoint*, at the cost of delaying other jobs.
Hybrid Policy: Extends time limit *only if no delays* to other jobs occur; otherwise, cancels early.

Autonomy Loop for Dynamic Time Limit Adjustment


slurmctld
The Slurm central management daemon
kill on timeout
set new time limit
Slurm Job
A Job running an application.
#SBATCH --time=08:00:00
srun application_executable
get job information with *squeue*
update time limits with *scontrol*
autonomy loop
Time Limit Adjustment Daemon
Continuously monitor job information, checkpointing progress and job queue.
Informed decision between:
Early Cancellation or *Time Limit Extension.*
report progress
Application
Reporting checkpointing progress.
progress file
<timestamp>
<checkpoint>

Application reports progress via a temp file

Daemon

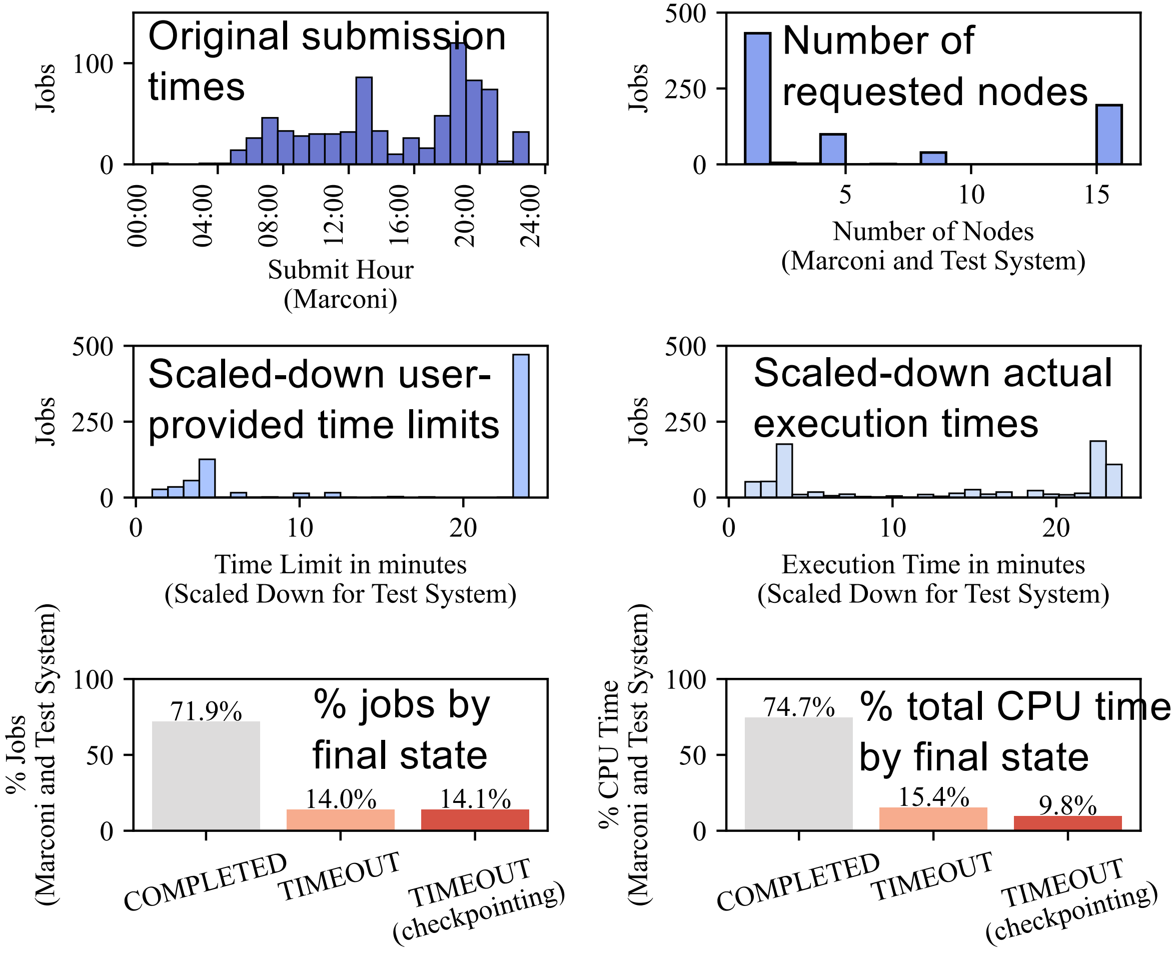
- Estimates **ckpt interval**
- Predicts **next checkpoint**
- Retrieves job info using **squeue**
- Decides on **policy**
- **Updates job time limit** via scontrol

Slurm (slurmctld)

- Applies new time limits based on daemon's commands.

T. Jakobsche, O. S. Simsek, J. Brandt, A. Gentile, F. M. Ciorba. An Autonomy Loop for Dynamic HPC Job Time Limit Adjustment. Euro-Par 2025, https://arxiv.org/abs/2505.05927

Construction of the Workload for Simulation



Dataset: 773 jobs from the Marconi PM100 dataset which initially contains 1,074,576 jobs, b/w May-October 2020

- Selected from Partition=1, Queue=1, Month=May **with most jobs**
- That executed exclusively on their assigned nodes for **at least one hour**
- W/ COMPLETED or TIMEOUT states

Time Down-scaling: 1 hour scaled down to 1 minute for replay in our experiments

Comparison of Dynamic Job Time Limit Adjustment vs. No Adjustments

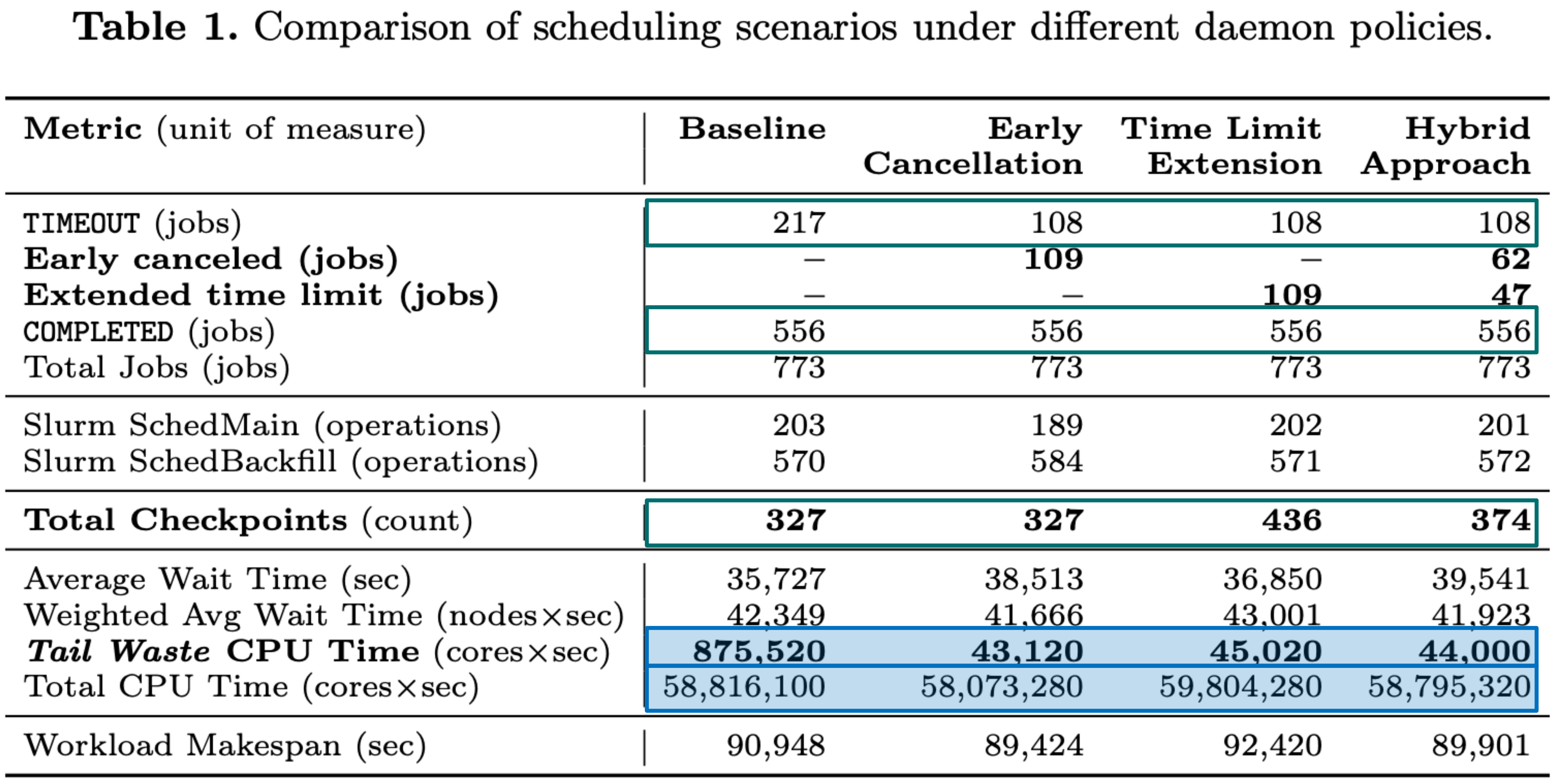


Table 1. Comparison of scheduling scenarios under different daemon policies.

Metric (unit of measure)	**Baseline**	**Early Cancellation**	**Time Limit Extension**	**Hybrid Approach**
TIMEOUT (jobs)	217	108	108	108
Early canceled (jobs)	–	**109**	–	**62**
Extended time limit (jobs)	–	–	**109**	**47**
COMPLETED (jobs)	556	556	556	556
Total Jobs (jobs)	773	773	773	773
Slurm SchedMain (operations)	203	189	202	201
Slurm SchedBackfill (operations)	570	584	571	572
Total Checkpoints (count)	**327**	**327**	**436**	**374**
Average Wait Time (sec)	35,727	38,513	36,850	39,541
Weighted Avg Wait Time (nodes×sec)	42,349	41,666	43,001	41,923
***Tail Waste* CPU Time** (cores×sec)	**875,520**	**43,120**	**45,020**	**44,000**
Total CPU Time (cores×sec)	58,816,100	58,073,280	59,804,280	58,795,320
Workload Makespan (sec)	90,948	89,424	92,420	89,901

95% reduction

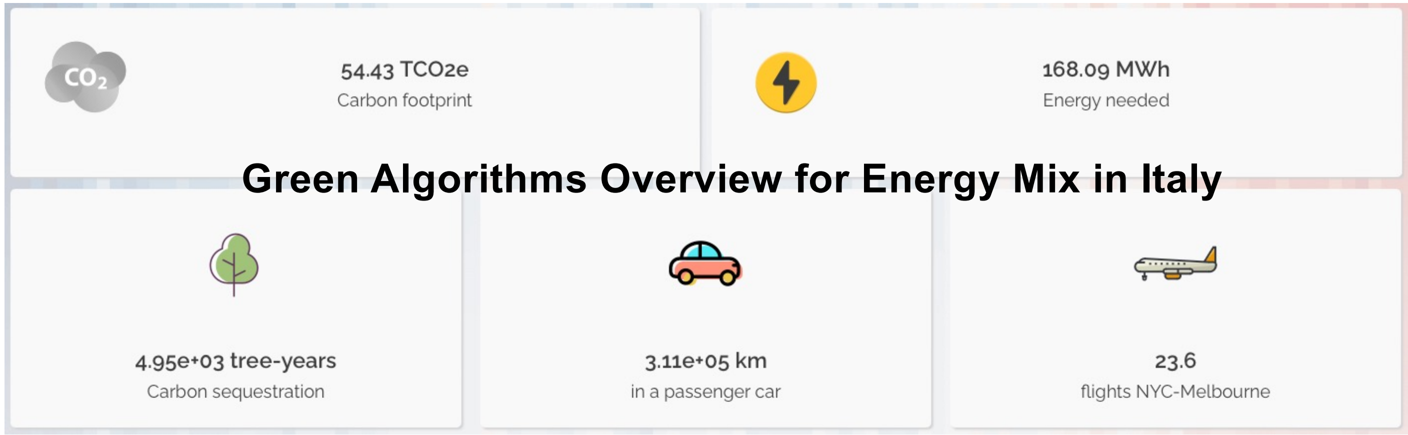
1.3% saving

T. Jakobsche, O. S. Simsek, J. Brandt, A. Gentile, F. M. Ciorba. An Autonomy Loop for Dynamic HPC Job Time Limit Adjustment. Euro-Par 2025, https://arxiv.org/abs/2505.05927

How Much Energy and Carbon Could We Save?

If the autonomy loop daemon were applied to the **entire Marconi system** & **workload** in the PM100 dataset:

Total Power Consumption	1’476 kW (https://www.top500.org/system/179845/)
Estimated **Autonomy Loop Savings**	1.3%
Power Savings	1’476 kW * 0.013 = 19.188 kW
Daily Energy Savings	19.188 kW * 24 h = 460.512 kWh
Annual Energy Savings	460.512 kWh/day * 365 d = 168’086.88 kWh = **168.09 MWh**
PUE ?	



Green Algorithms Overview for Energy Mix in Italy

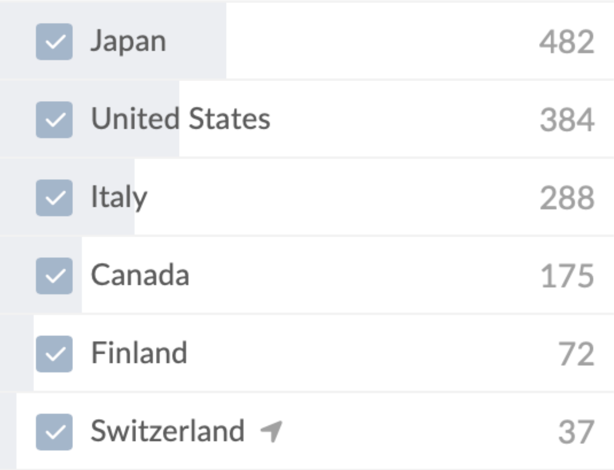
54.43 TCO2e
Carbon footprint

168.09 MWh
Energy needed

4.95e+03 tree-years
Carbon sequestration

3.11e+05 km
in a passenger car

23.6
flights NYC-Melbourne



☑	Japan	482
☑	United States	384
☑	Italy	288
☑	Canada	175
☑	Finland	72
☑	Switzerland	37

Carbon intensity data (2025)
ourworldindata.org/electricity-mix

Contributions

- Developed **SPH-EXA**, a production-grade, performance-portable simulation framework using:
  - Modern C++ with CUDA/HIP/SYCL backends
  - Cornerstone octrees for domain decomposition
  - New gravity solvers (Ryoanji) and hydrodynamic modules
  - Showed exascale readiness across heterogeneous architectures
  - Developed and applied accurate energy measurements for functional & device-level breakdowns

- Proposed an **autonomy loop** for dynamic job time limit adjustment
  - Validated on real workloads (Marconi PM100 dataset)
  - Achieved up to 95% tail waste reduction and 1.3% energy savings

- Quantified the **carbon footprint of extreme-scale simulations and energy savings**
  - Compared real vs. hypothetical energy use across diverse architectures and geographies
  - Estimated energy savings of dynamic time limit adjustment

Performance, portability, sustainability need smarter scheduling

Raw **speed** and **naïve scheduling** are no longer enough.

The cost of **carbon** for computation & **poor scheduling** is real and rising.

We must factor **performance**, **portability** & **sustainability** into application & **scheduler** design.

University of Basel

DMI-HPC @ UniBas

Performance & Portability For Sustainable Simulations at Extreme Scales

Florina M. Ciorba

18th Scheduling for Large-Scale Systems
Montreal, QC, Canada
July 10, 2025, 9:00-9:30

florina.ciorba@unibas.ch | @cflorina.bsky.social | www.linkedin.com/in/cflorina

Swiss National Science Foundation

Platform for Advanced Scientific Computing

SKACH

DAPHNE