Energy Efficient Computing

Benjamin Czaja

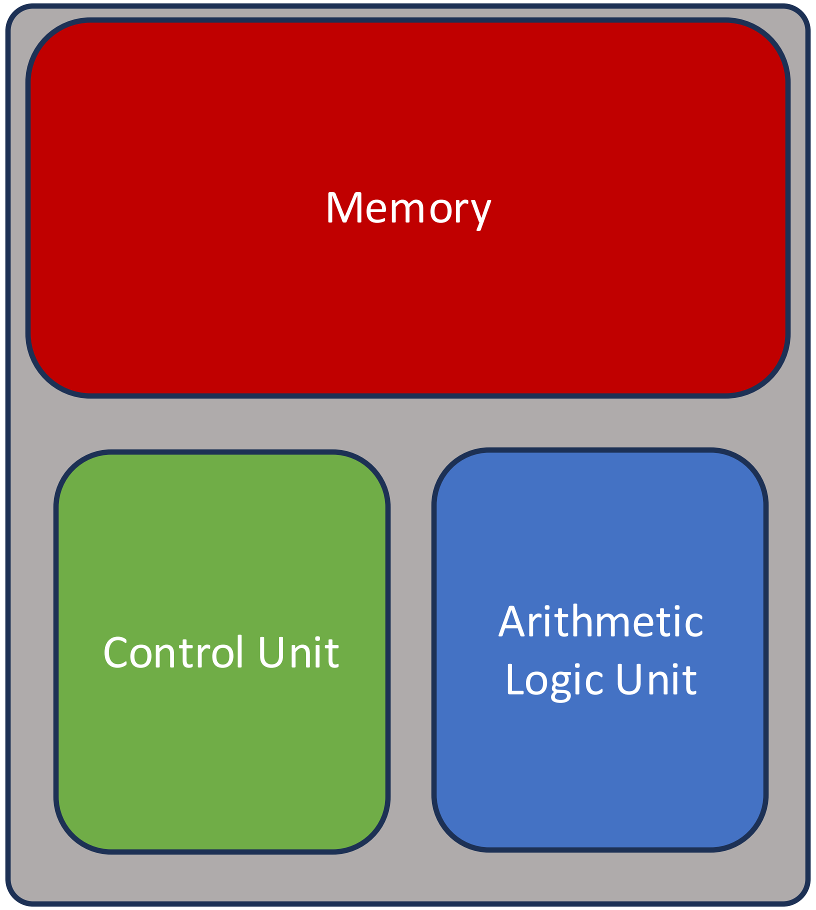
HPC Advisor SURF

December 2024

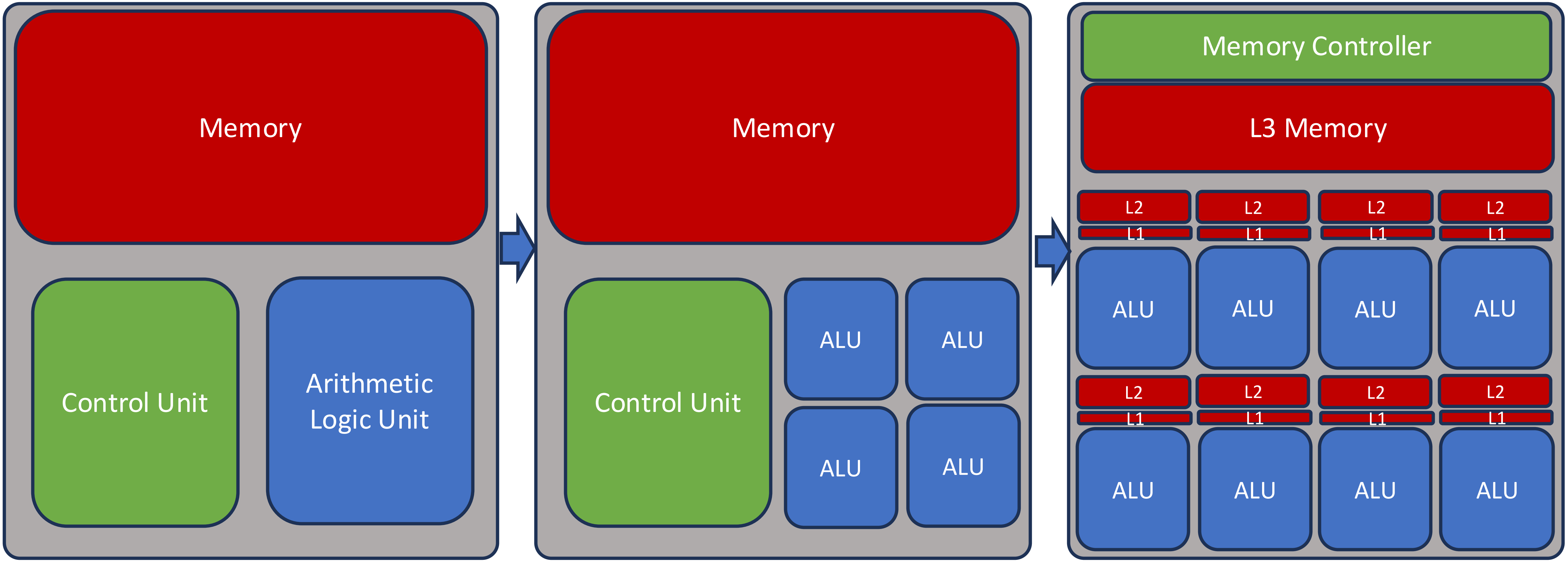
- We are becoming more energy efficient
- We are using more energy


SURF
Snellius
Lenovo
SURF
Lenovo
SURF

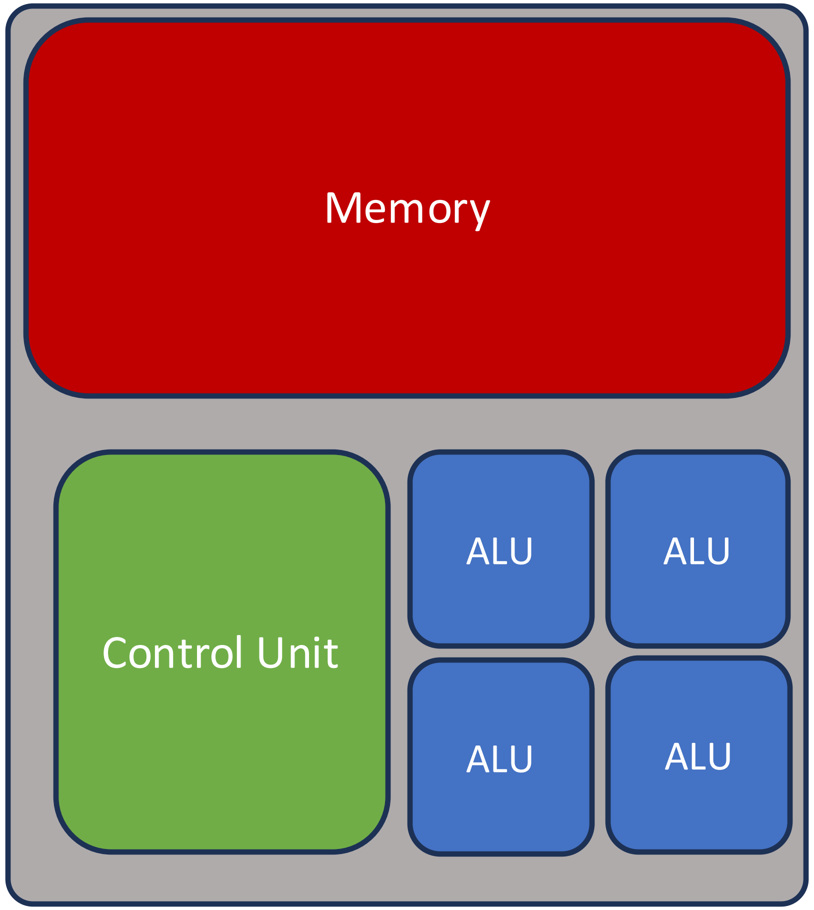
Single “core” Processor/CPU



Single-core CPU to Multi-core CPU

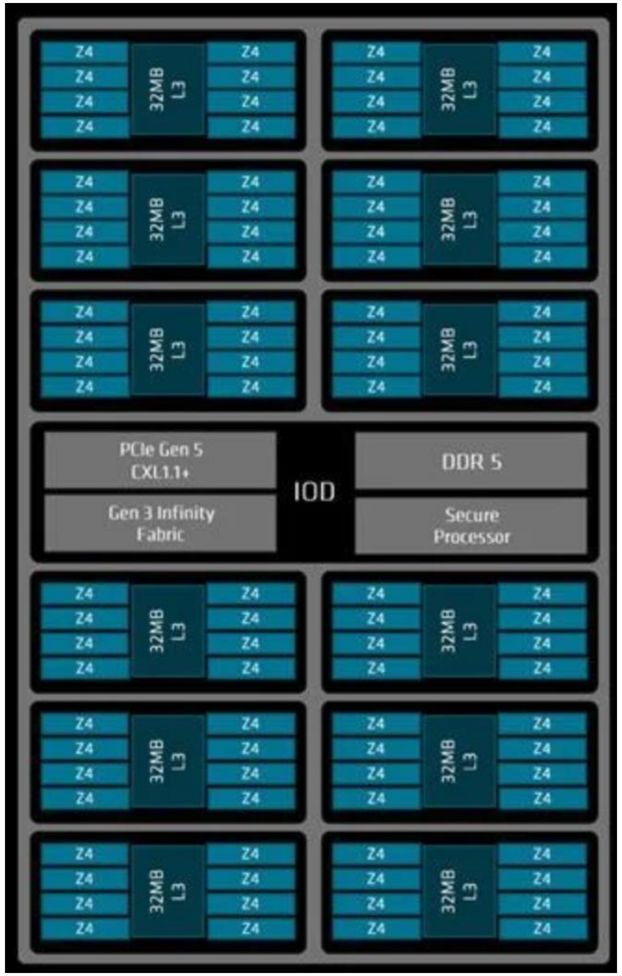


Multi-core
4 cores



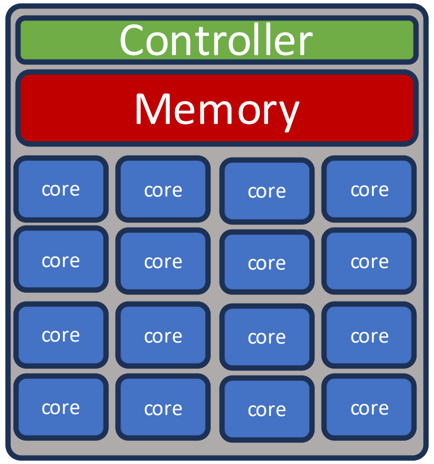
AMD Rome (Zen2)
64 cores

AMD Genoa (Zen4)
96 cores



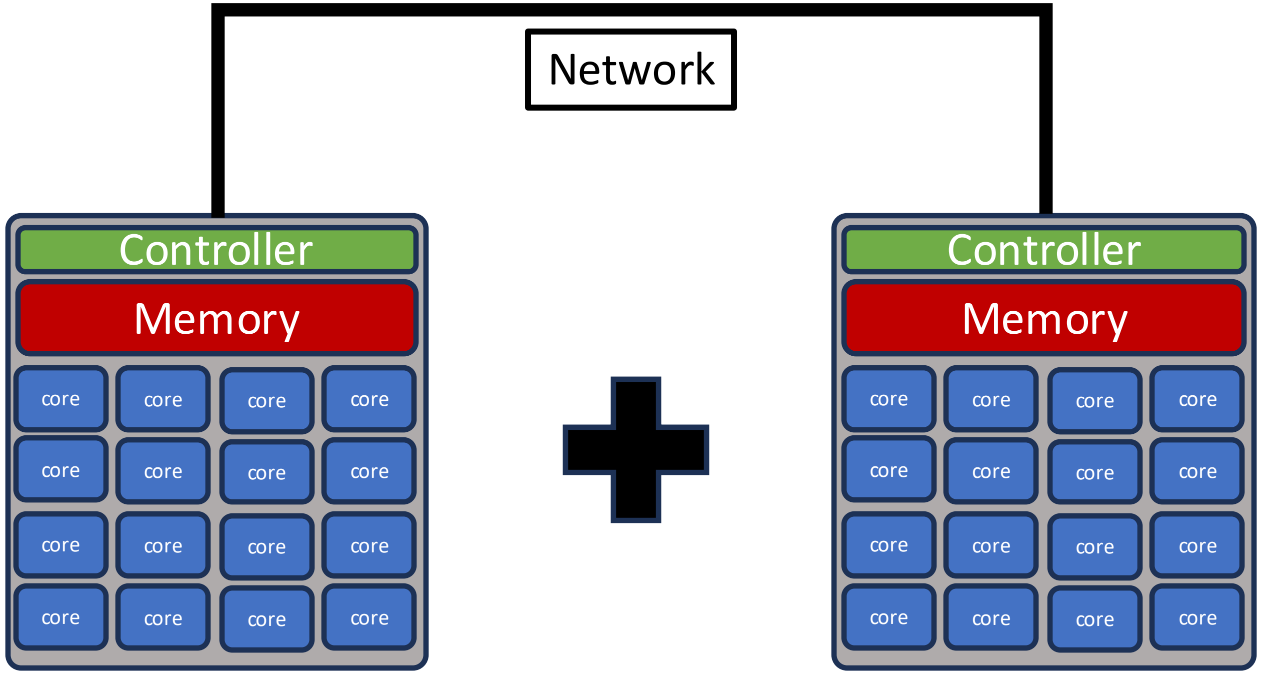
Lets Build a Compute Cluster

This is called
“node”



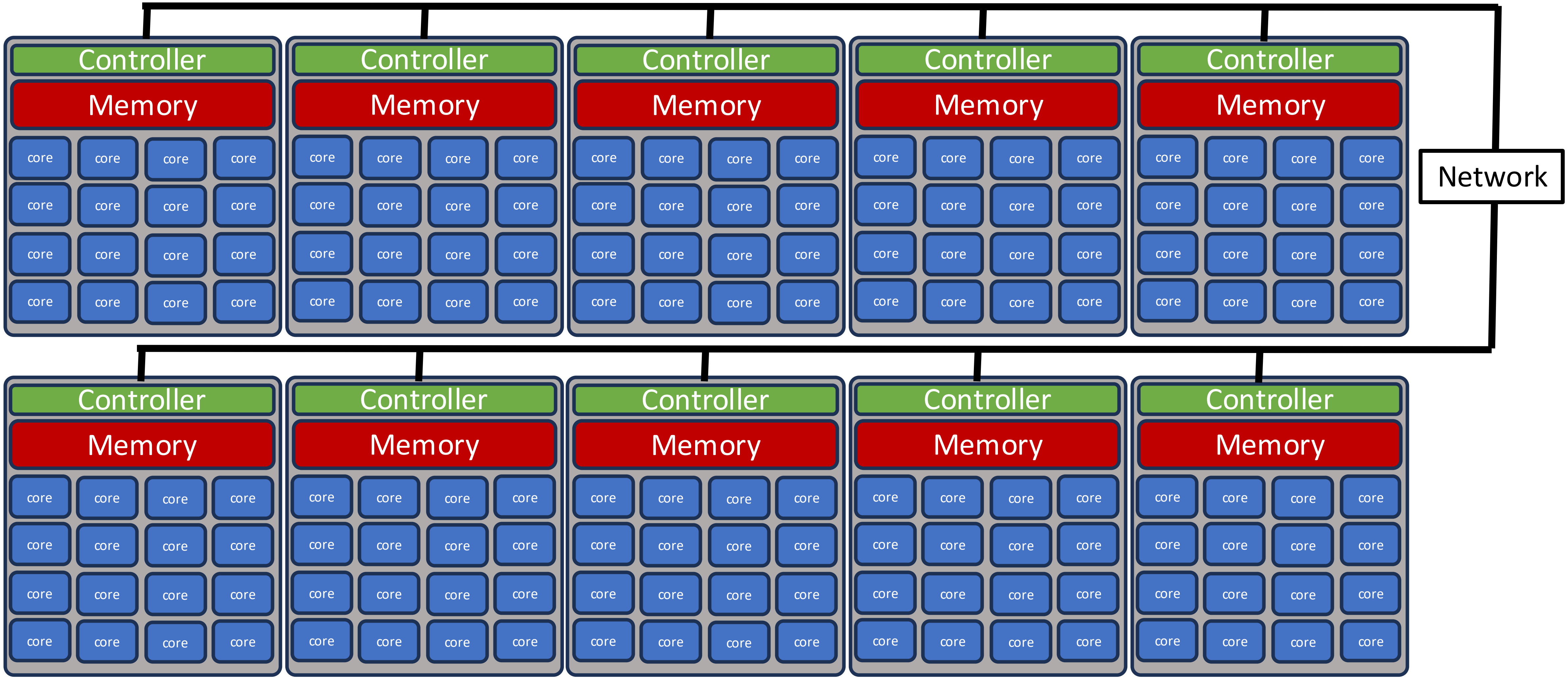
Lets take our theoretical 16 core CPU and connect it to more of the same CPUs

Lets Build a Compute Cluster



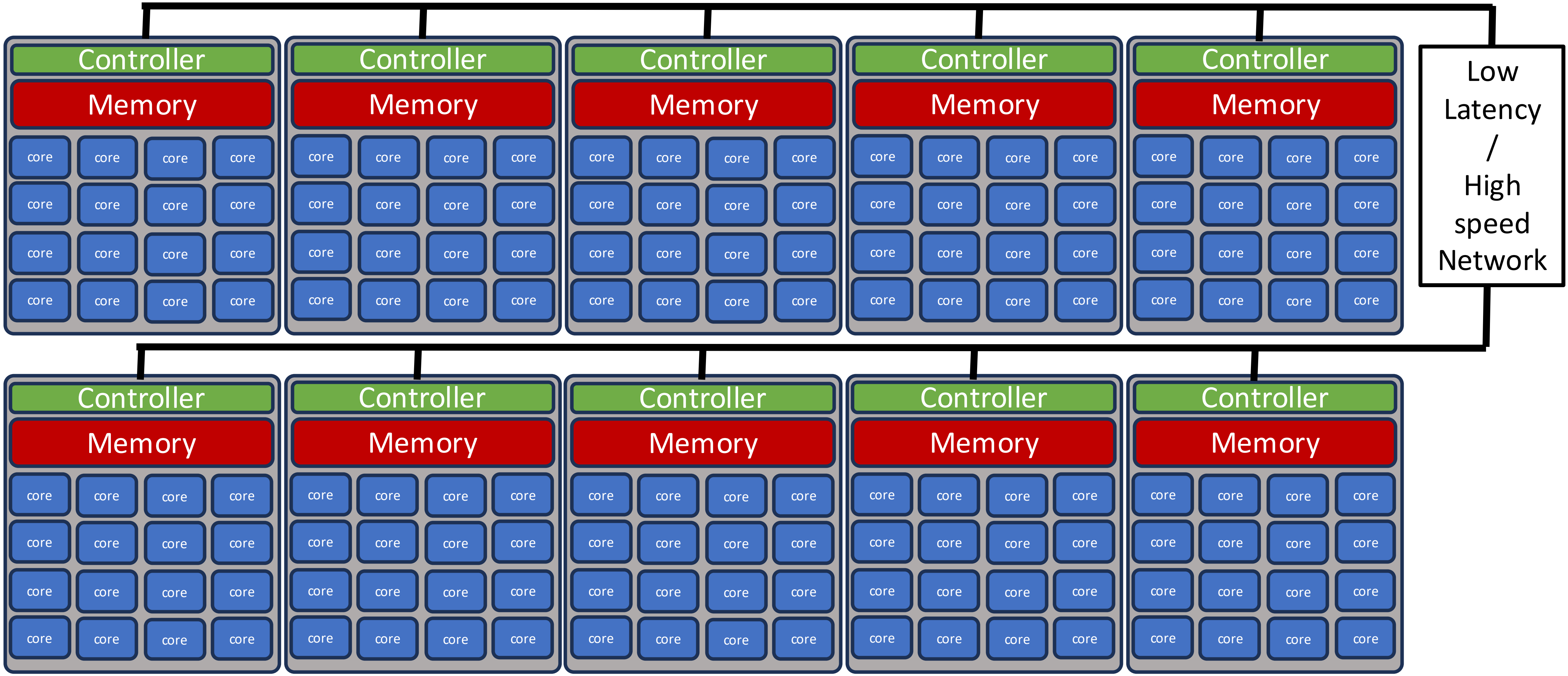
- We now have a 32 core machine!!
- Ok Ben's PowerPoint skills are limited …..

Lets Build a Compute Cluster



- We now have a 144 core machine!!
- Ok Ben’s PowerPoint skills are limited …..

Lets Build a Supercomputer!!!!



- We now have a 144 core SUPERCOMPUTER!!
- Ok Ben’s PowerPoint skills are limited …..

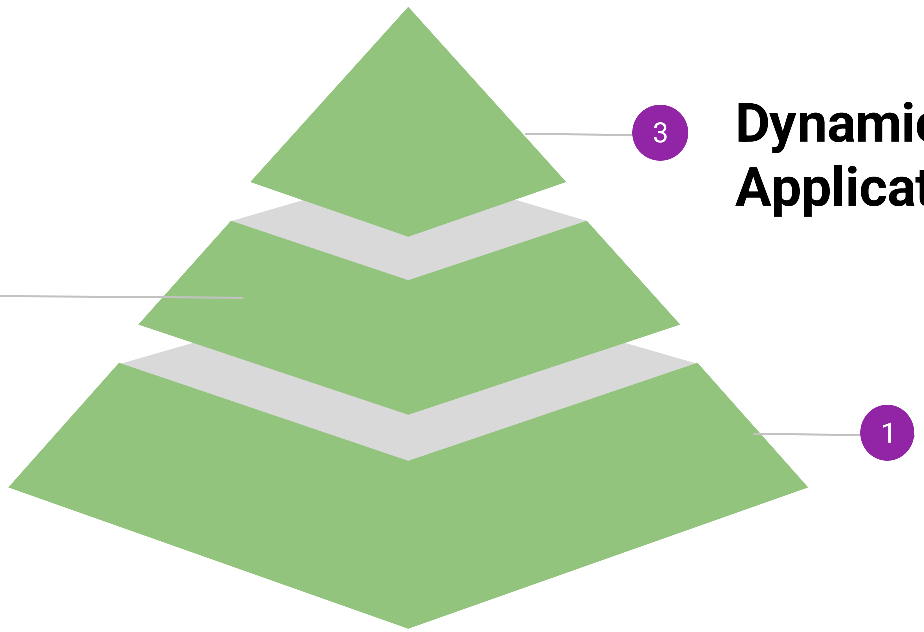
Snellius - Dutch National supercomputer
> 240,000 cores

- CPU partitions:
  - 522 thin-rome nodes (256 GiB) (128 cores)
  - 785 thin-genoa nodes (336 GiB) (192 cores)
  - 72 Fat nodes (1 TiB)
  - 4 high memory (2x 8 TiB, 2x 4 TiB)
- GPU partition (4x NVIDIA A100 GPUs):
  - 72 GPU (Intel Xeon Platinum 8360Y (2x) hosts)

(#374) CPU: LINPACK Rmax (0.5 MW) 2.13 PFlop/s

(#176) GPU: LINPACK Rmax (0.13 MW) 3.6 PFlop/s

Energy-aware focus

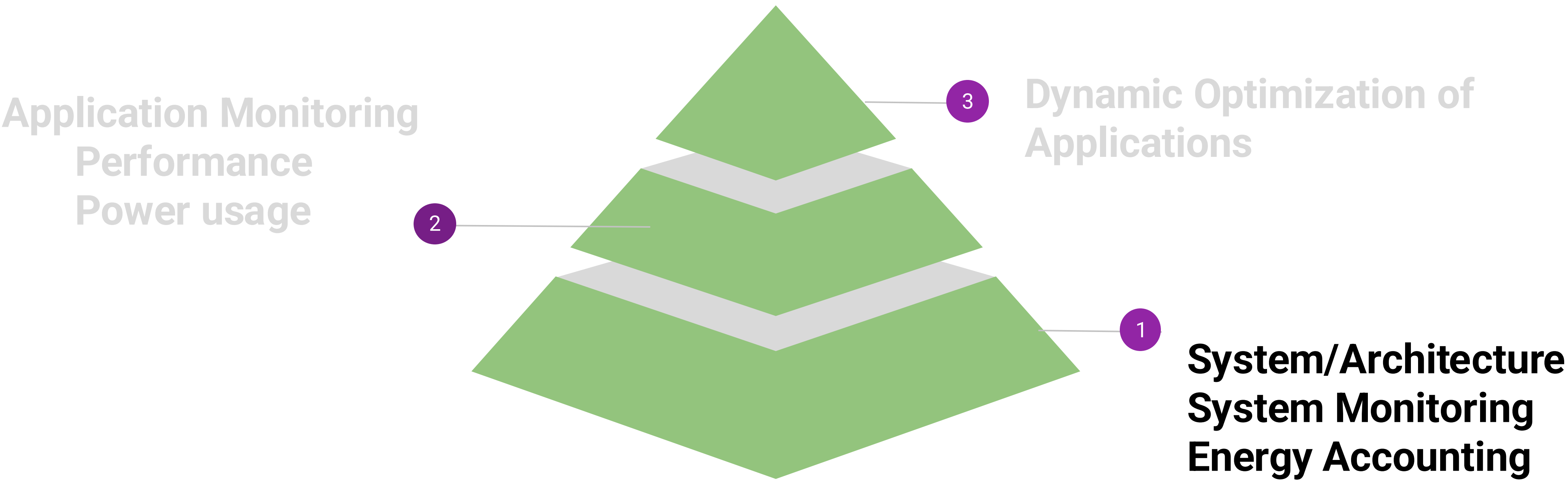


Dynamic Optimization of Applications

Application Monitoring
Performance
Power usage

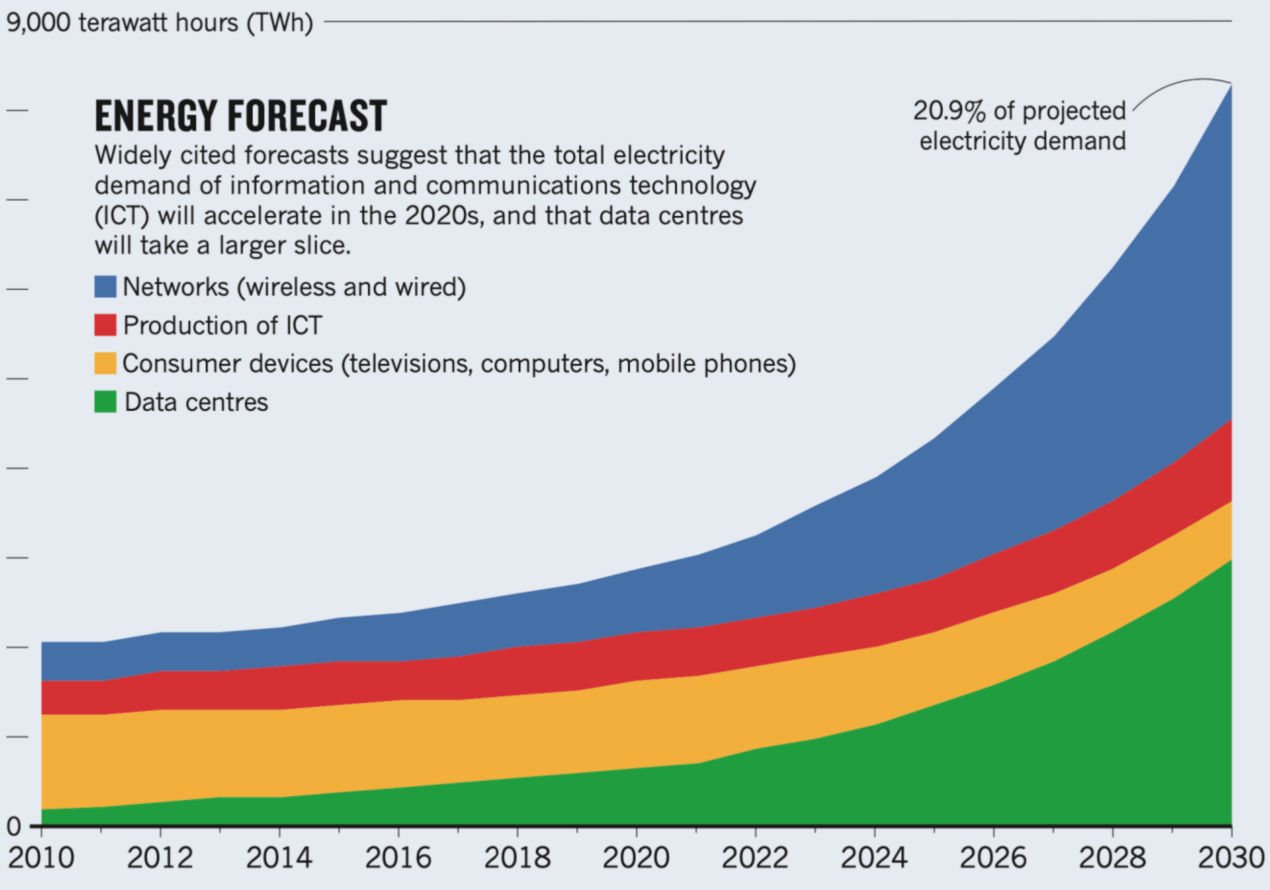
System/Architecture
System Monitoring
Energy Accounting

Energy-aware focus



ICT/Data center energy forecast

- Data centre energy usage:
  - ~ 200 TWh for data centres in 2018
  - ~ 3000 TWh in 2030
- Dutch Data Center Usage (2019)
  - 1300 MW installed capacity
  - 0.3 % electricity usage of the Netherlands (CBS)
  - 3x as much power than the NS

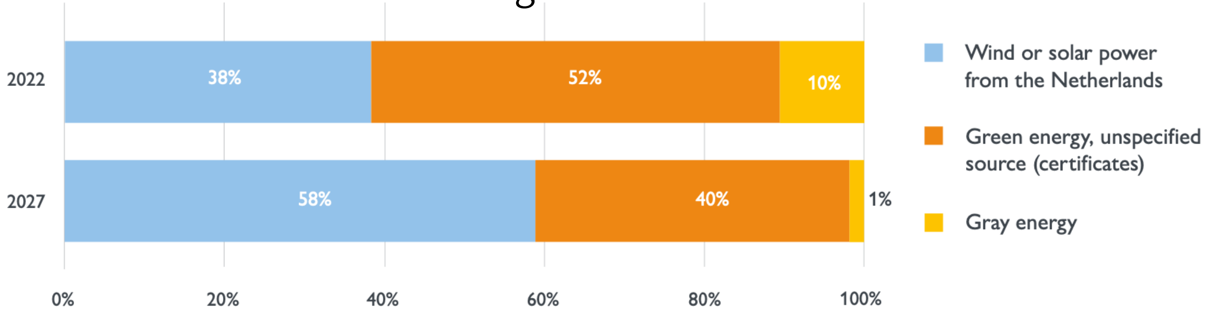


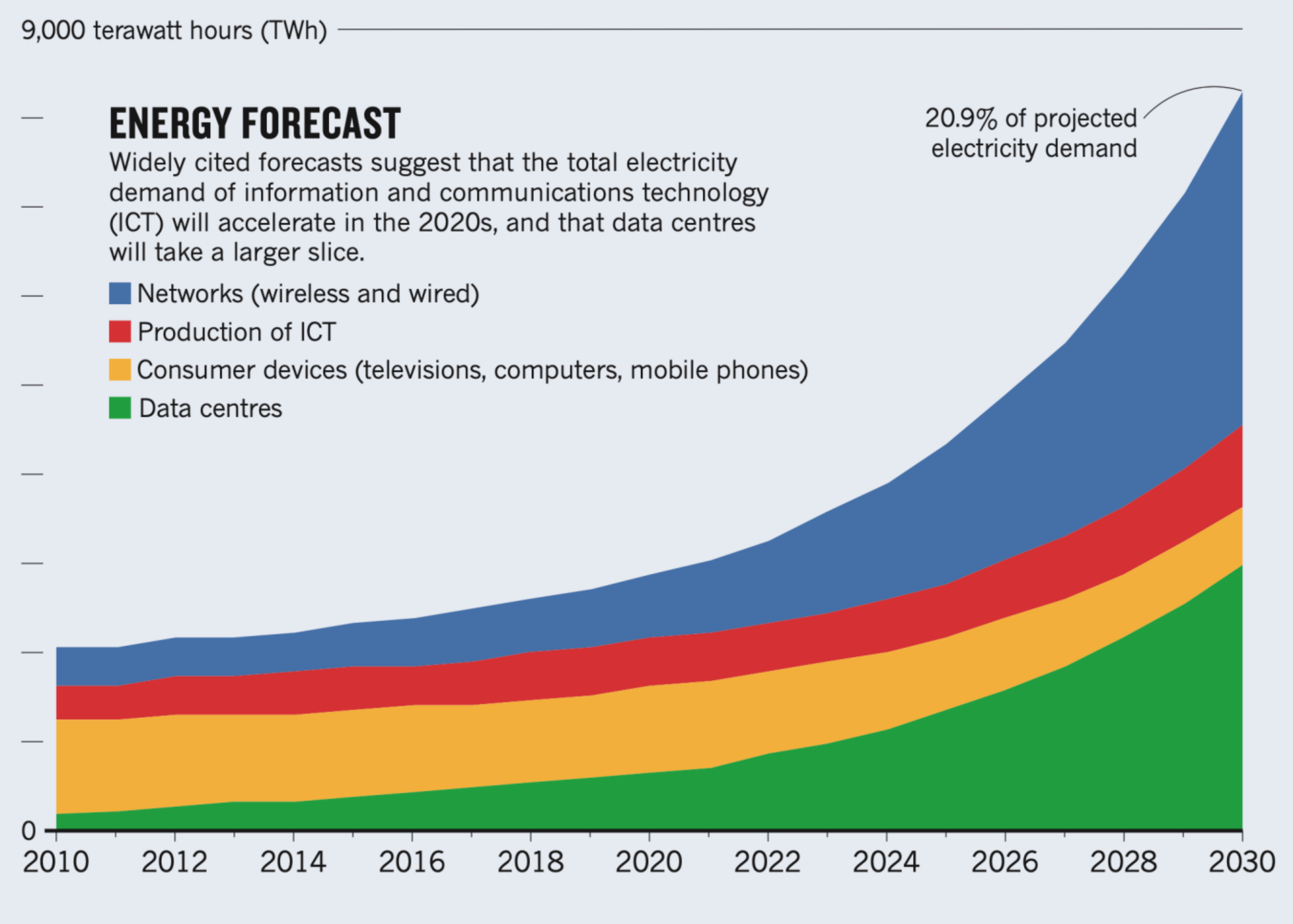
The chart above is an 'expected case' projection from Anders Andrae, a specialist in sustainable ICT. In his 'best case' scenario, ICT grows to only 8% of total electricity demand by 2030, rather than to 21%.

- Jones, Nicola. "How to stop data centres from gobbling up the world's electricity." *Nature* 561.7722 (2018): 163-166.
- Andrae, Anders SG, and Tomas Edler. "On global electricity usage of communication technology: trends to 2030." *Challenges* 6.1 (2015): 117-157.
- Bakkeren, Hanno. "Datacenters verbruiken drie keer zoveel stroom als de NS". NRC, 14 May 2019
- State of the Dutch Data Centers, The Dutch Data Center Report, 2022, https://www.dutchdatacenters.nl/

ICT/Data center energy forecast

- Data centre energy usage:
  - ~ 200 TWh for data centres in 2018
  - ~ 3000 TWh in 2030
- Dutch Data Center Usage (2019)
  - 1300 MW installed capacity
  - 0.3 % electricity usage of the Netherlands (CBS)
  - 3x as much power than the NS
- Dutch Data center energy consumption (2022)
  - 90% of all energy consumed by colocation data centers is green.





The chart above is an 'expected case' projection from Anders Andrae, a specialist in sustainable ICT. In his 'best case' scenario, ICT grows to only 8% of total electricity demand by 2030, rather than to 21%.

- Jones, Nicola. "How to stop data centres from gobbling up the world's electricity." *Nature* 561.7722 (2018): 163-166.
- Andrae, Anders SG, and Tomas Edler. "On global electricity usage of communication technology: trends to 2030." *Challenges* 6.1 (2015): 117-157.
- Bakkeren, Hanno. "Datacenters verbruiken drie keer zoveel stroom als de NS". NRC, 14 May 2019
- State of the Dutch Data Centers, The Dutch Data Center Report, 2022, https://www.dutchdatacenters.nl/

Top 500 (Nov-2024)

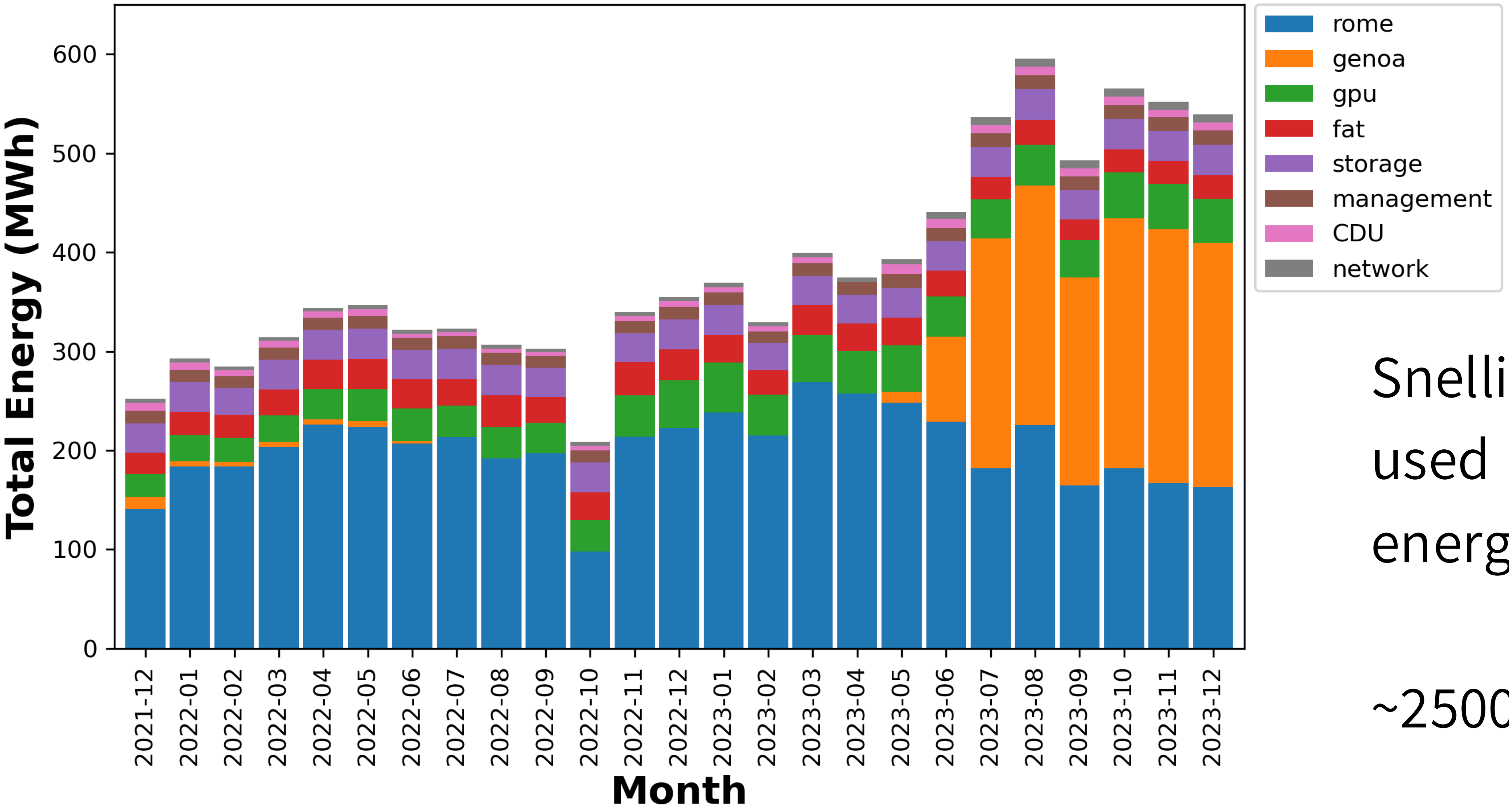
1. **El Capitan-US (30 MW) 1,742.0 PFlops**
2. **Frontier-US (24 MW) 1,353.0 PFlop/s**
3. **Aurora-US (38 MW) 1,012.0 Pflop/s**
4. **Eagle-US, Microsoft (??) 561 Pflop/s**
5. **HP6 (8 MW) 477.01 PFlop/s**

Snellius

(#197) CPU: LINPACK Rmax (0.8 MW) 5.4 PFlop/s

(#94) GPU: LINPACK Rmax (0.3 MW) 13.64 PFlop/s

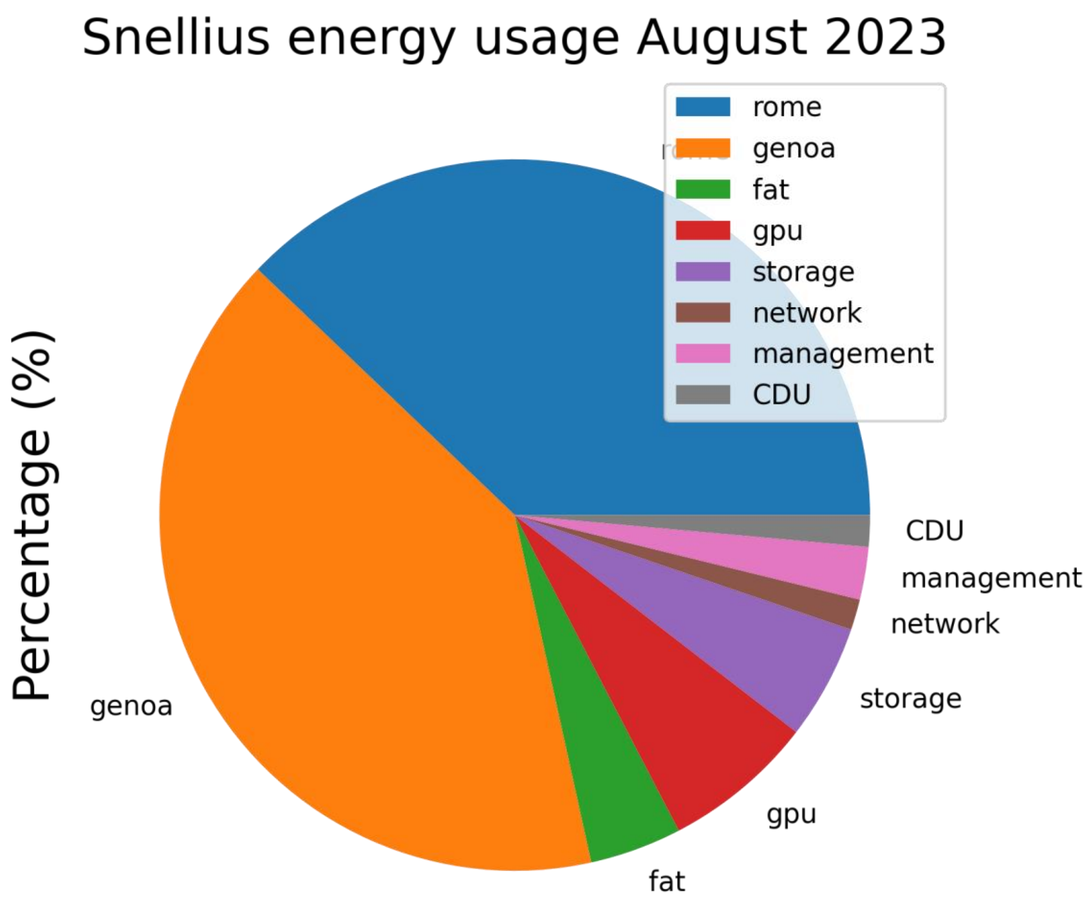
Snellius - Total Energy Usage



Snellius in December used ~ 600MWh of energy.

~2500 house holds worth

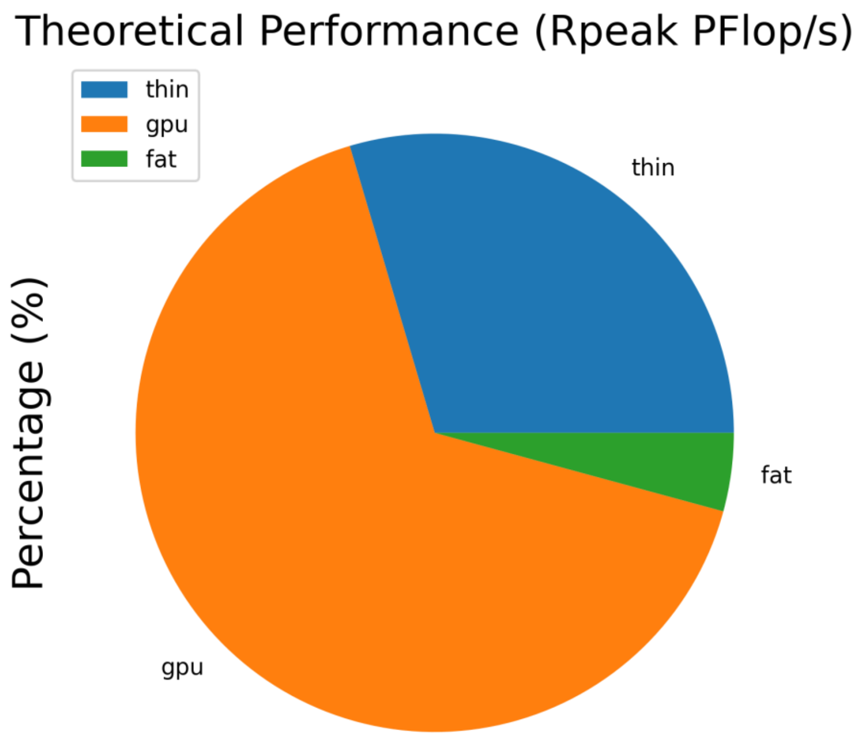
Snellius - Energy usage per partition

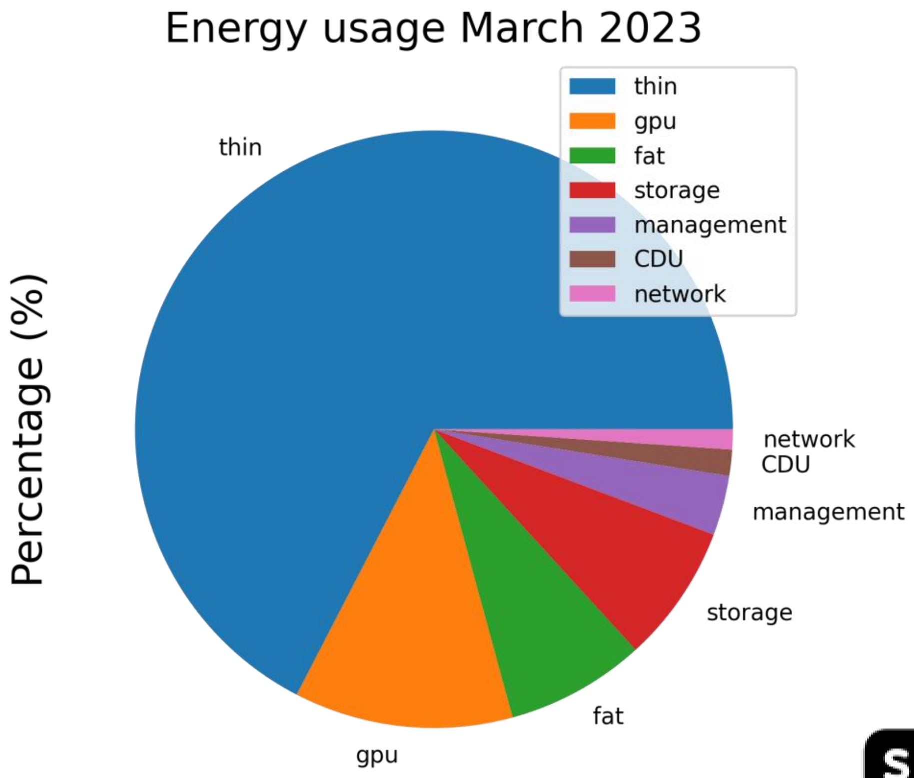


- 90% Compute
  - 80% CPU
    - 39% Rome
    - 41% Genoa
  - 7% Gpu
  - 3% Fat
- 10% Other
  - 5.0% Storage
  - 2.4% Management
  - 1.5% Cooling
  - 1.1% Network

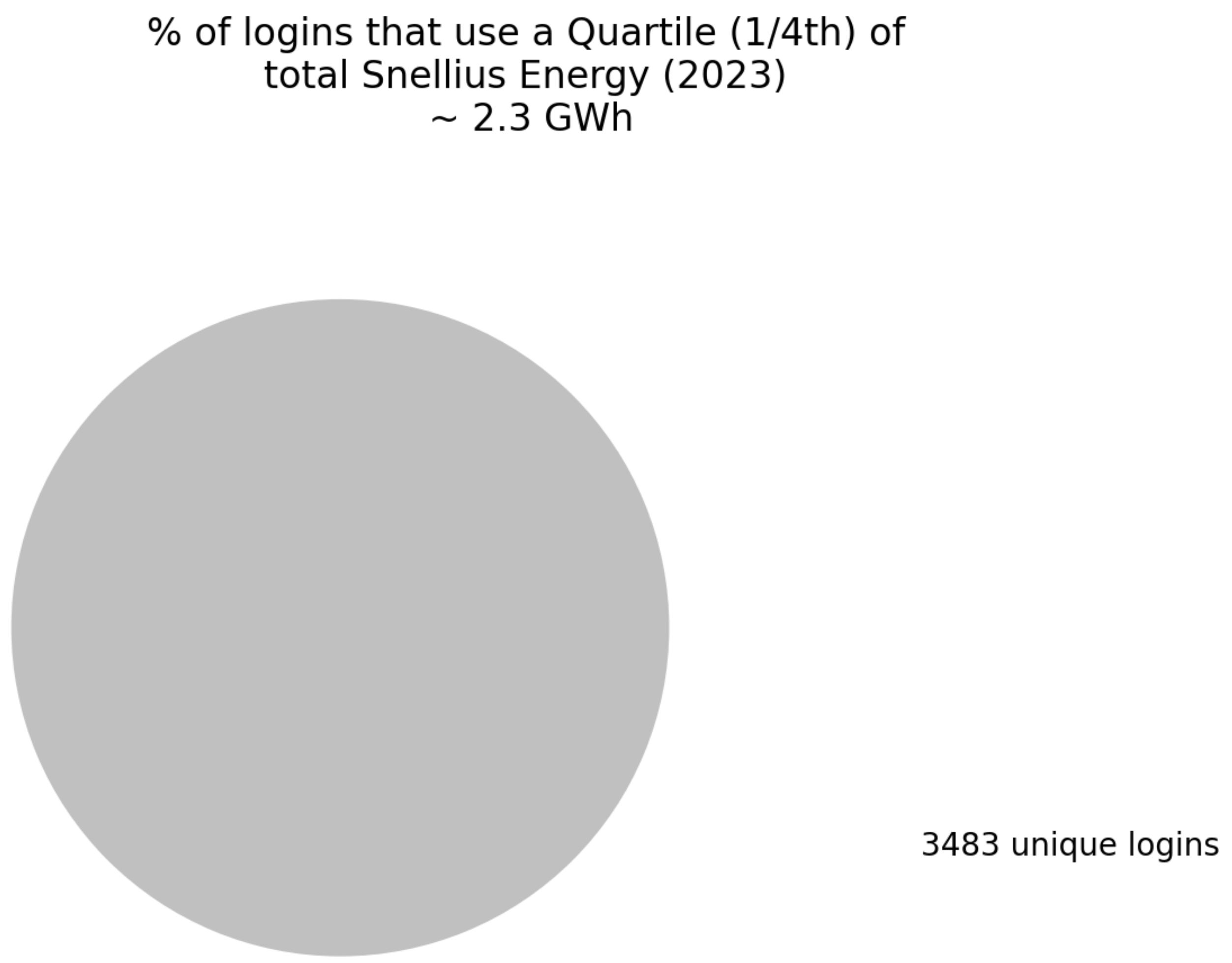
Performance vs Energy

GPU ~ same compute for 1/5 the energy (March 2023)

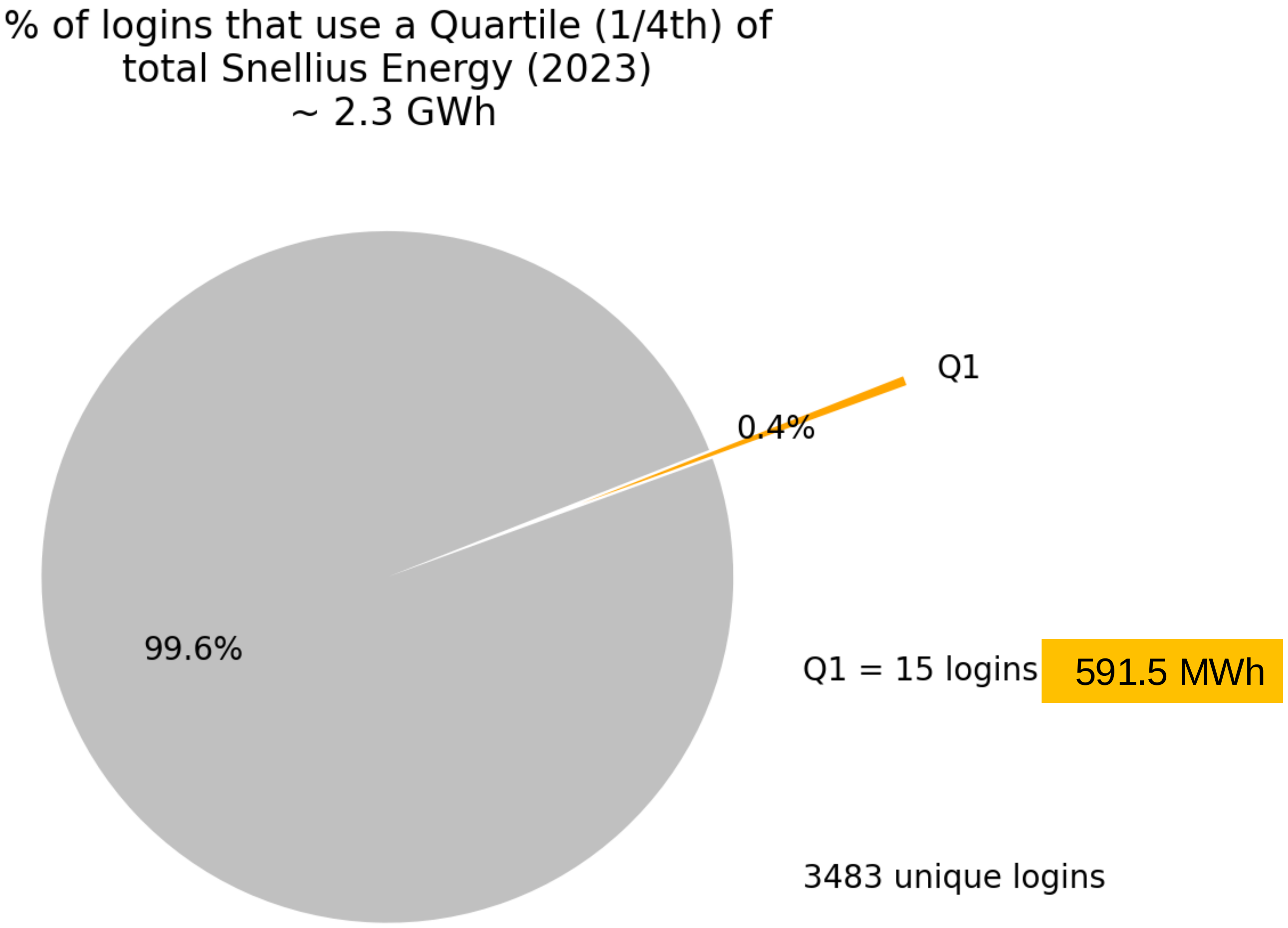




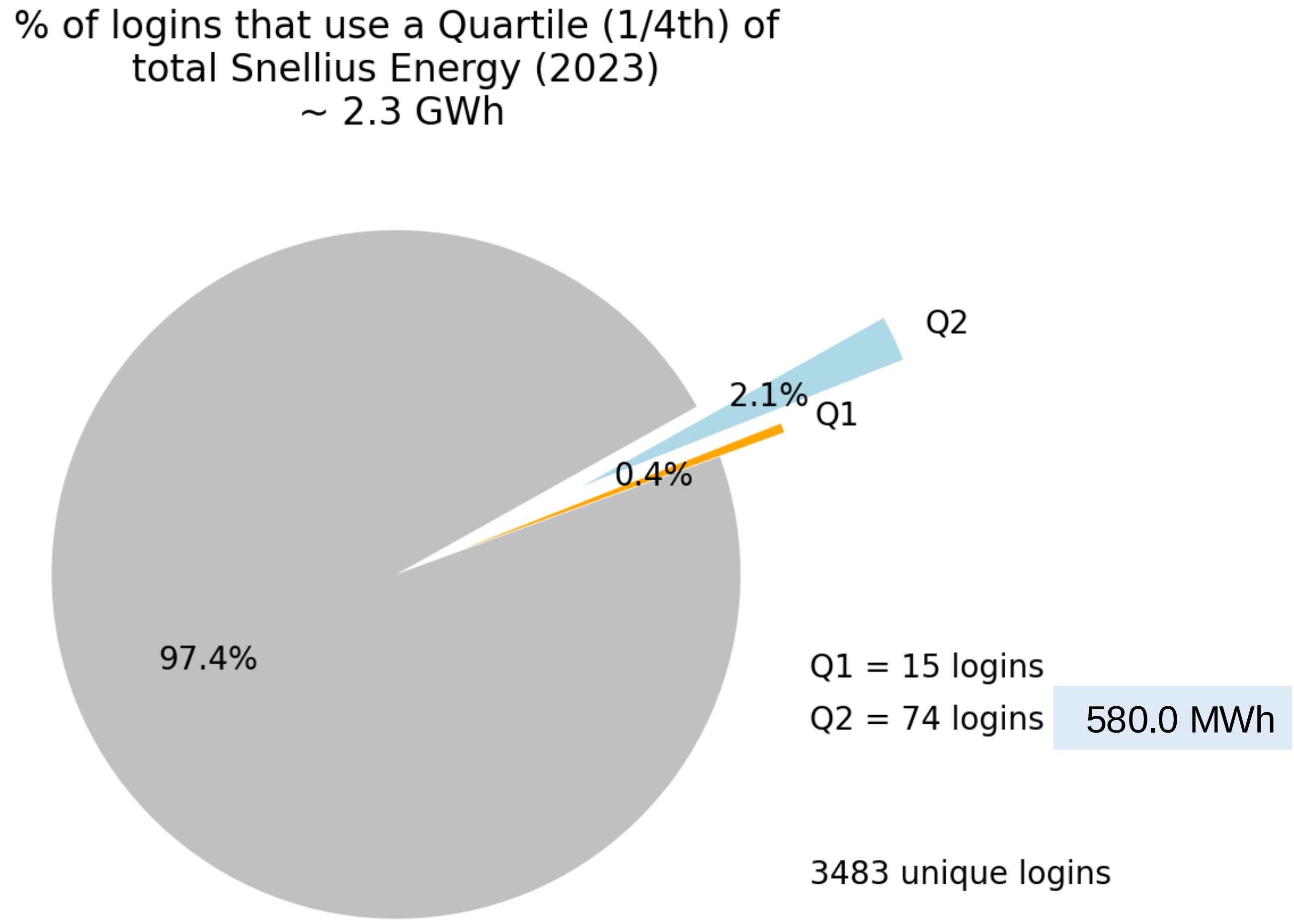
Energy Breakdown by users



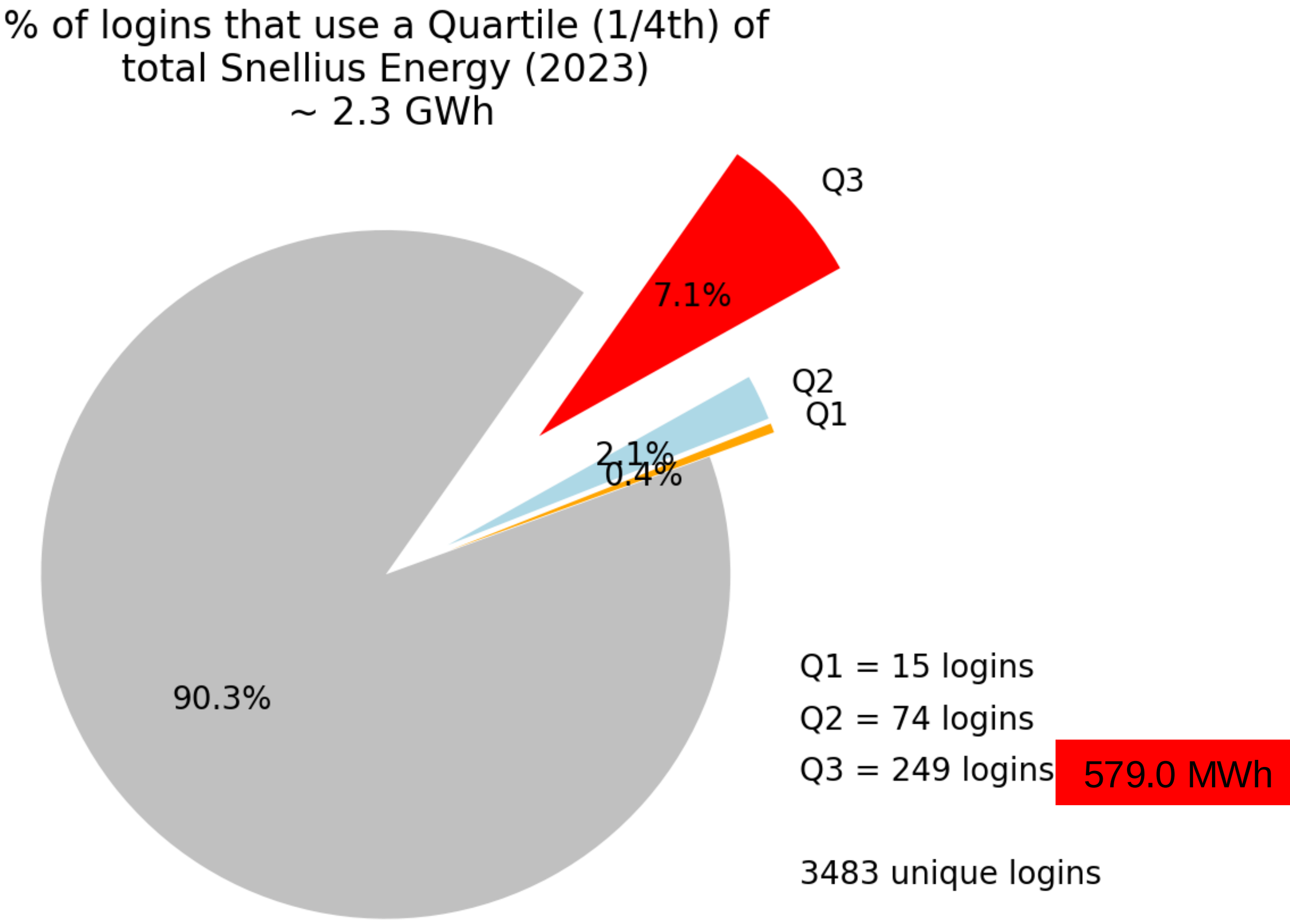
Energy Breakdown by users



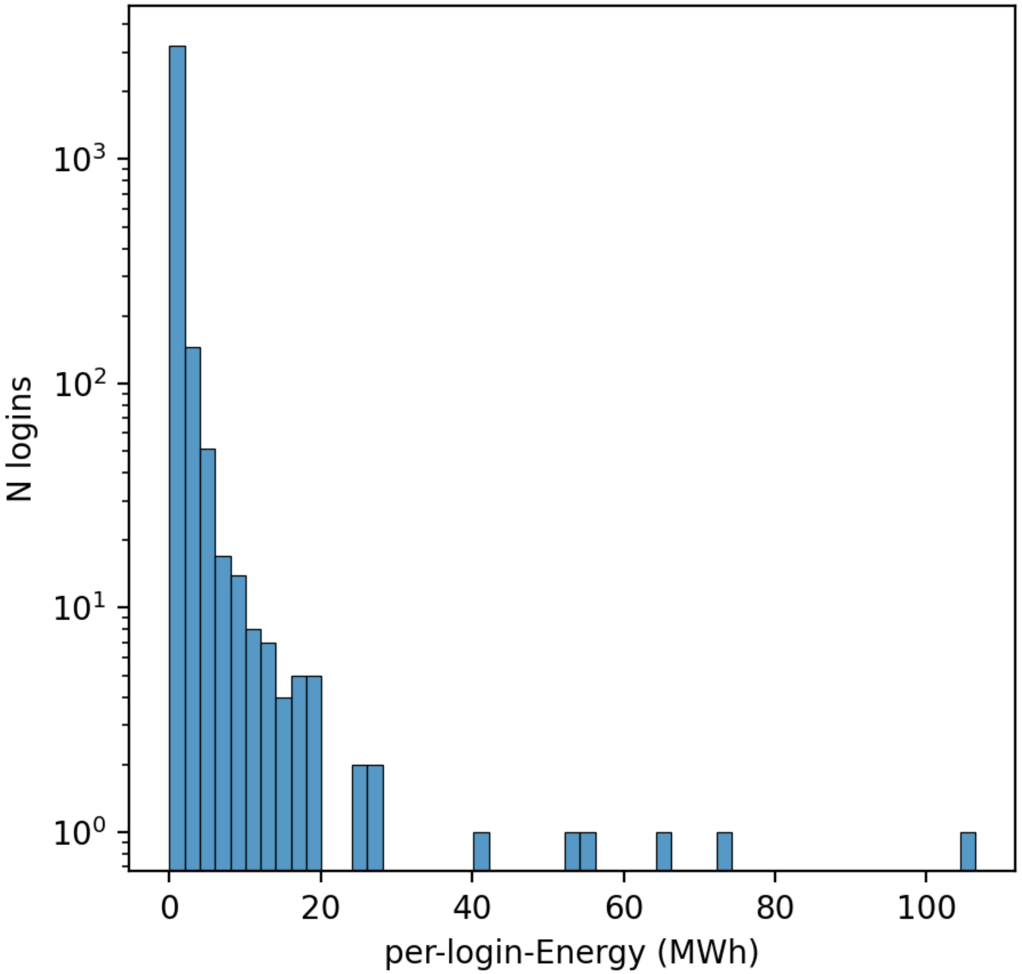
Energy Breakdown by users

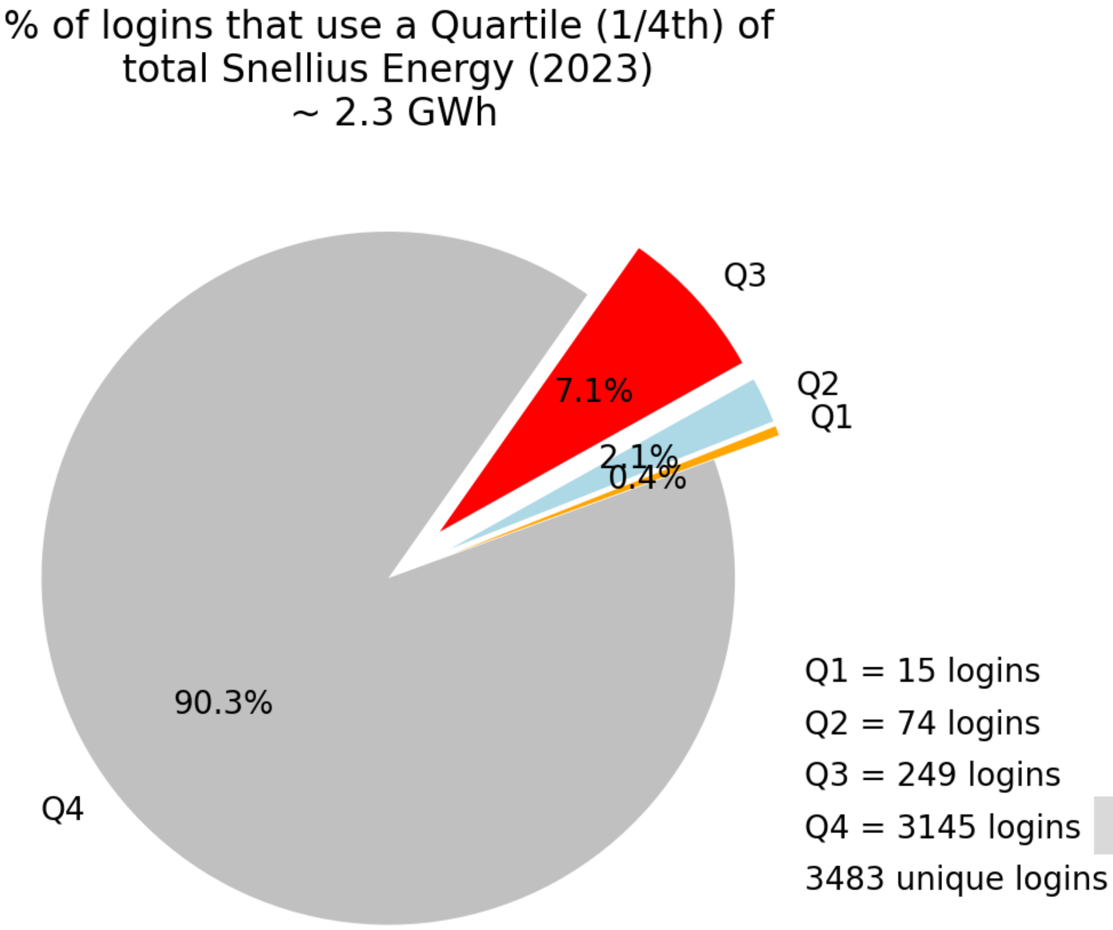


Energy Breakdown by users



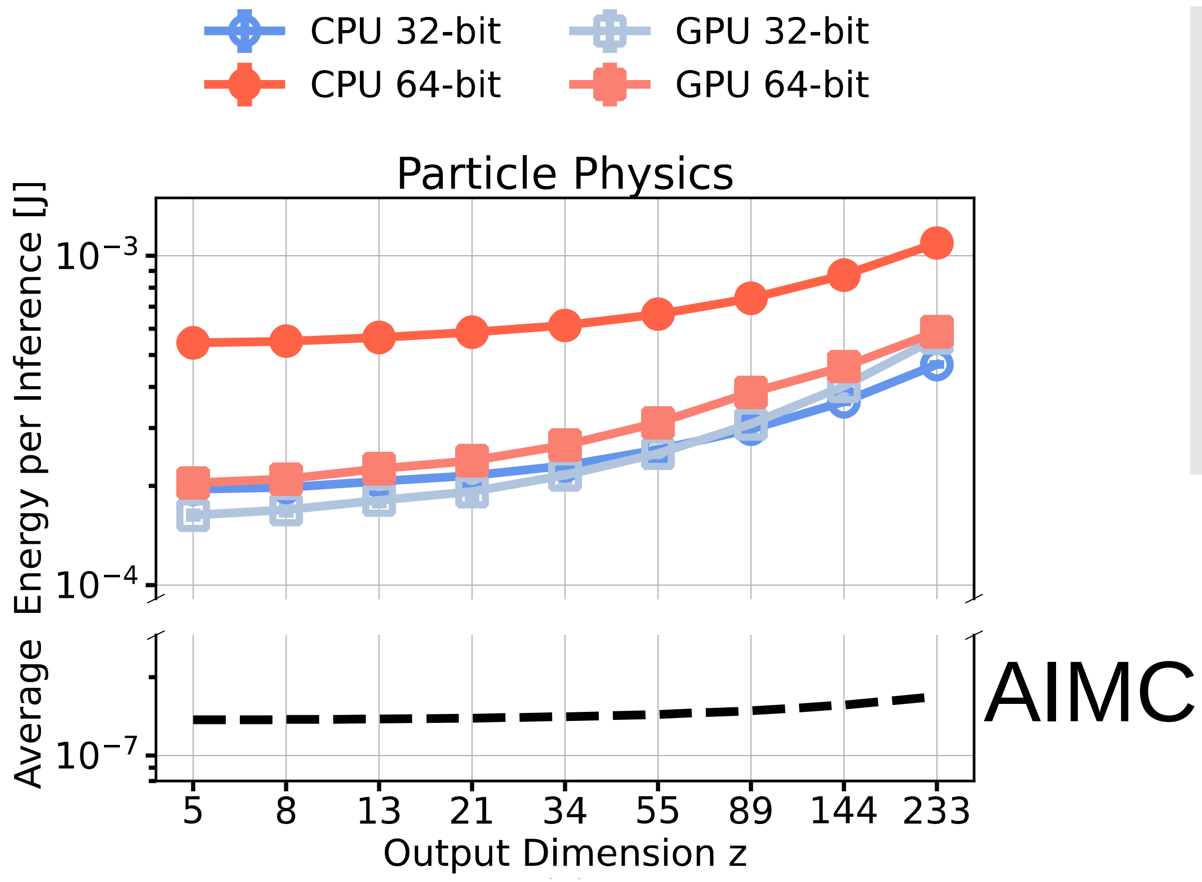
Energy Breakdown by users





580.0 MWh

Energy-efficient computing architectures



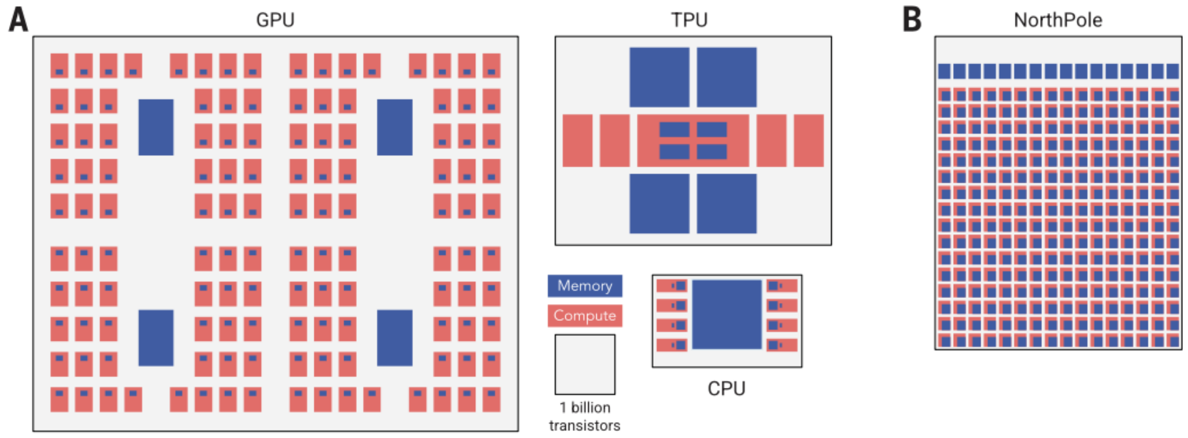
D.J. Kösters et al., APL Machine Learning 1, 016101 (2023)

'NL-ECO: Netherlands Initiative for Energy-Efficient Computing'

UNIVERSITY OF TWENTE.

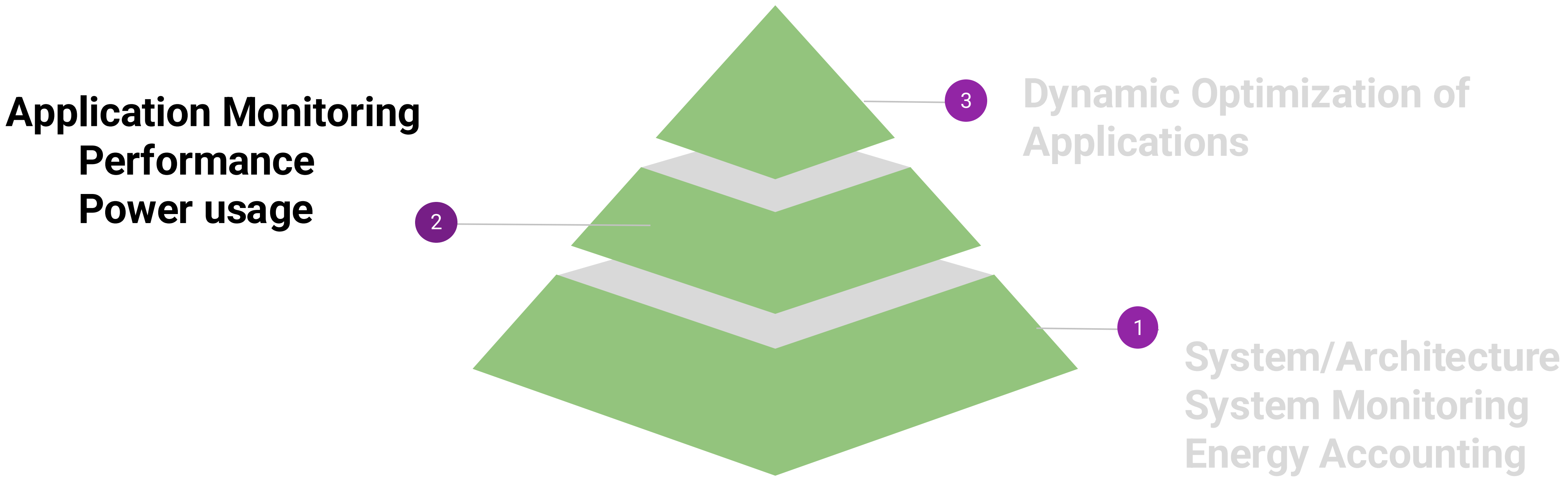
Radboud University
Nijmegen, the Netherlands

SURF



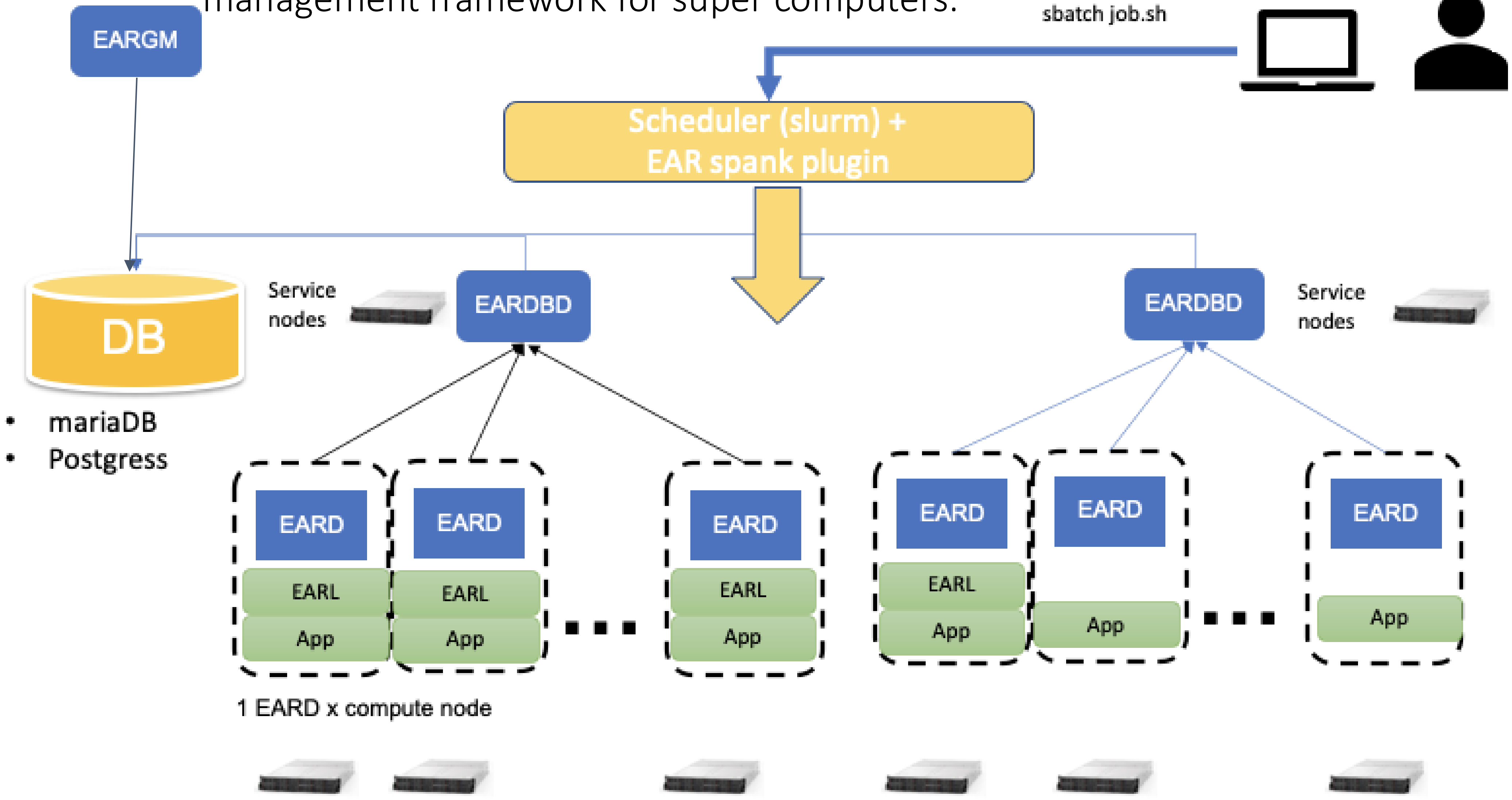
Modha et al., Science 382, 329–335 (2023) 20 October 2023

Energy-aware focus



Energy Aware Runtime (EAR)

Energy Aware Runtime (EAR) package provides an energy management framework for super computers.



- mariaDB
- Postgress

Energy Aware Runtime (EAR)

Energy Aware Runtime (EAR) package provides an energy management framework for super computers.

- **EAR Node Manager (EARD) “Daemon”**
  - Energy metrics via the RAPL (Running Average Power Limit) function
  - Global energy limits or just offer global cluster monitoring

- **EARL** is a library that is loaded next to the application
  - Offers application metrics monitoring
  - Can select the frequencies based on the application behavior on the fly.
  - Integrated with SLURM on Snellius.
  - Intercepts the MPI symbols through the PMPI interface to provide “traces” of MPI applications.

```
#!/bin/bash

#SBATCH --ntasks=256
#SBATCH --time=24:0:0
#SBATCH -p thin --exclusive

#SBATCH --ear=on
#SBATCH --ear-policy=monitoring

module load 2021
module load foss/2021a
module load CMake/3.20.1-GCCcore-10.3.0
module load VASP6/6.2.1-foss-2021a-CUDA-11.3.1

srun vasp_std
```

Energy Aware Runtime (EAR)

Energy Aware Runtime (EAR) package provides an energy management framework for super computers.

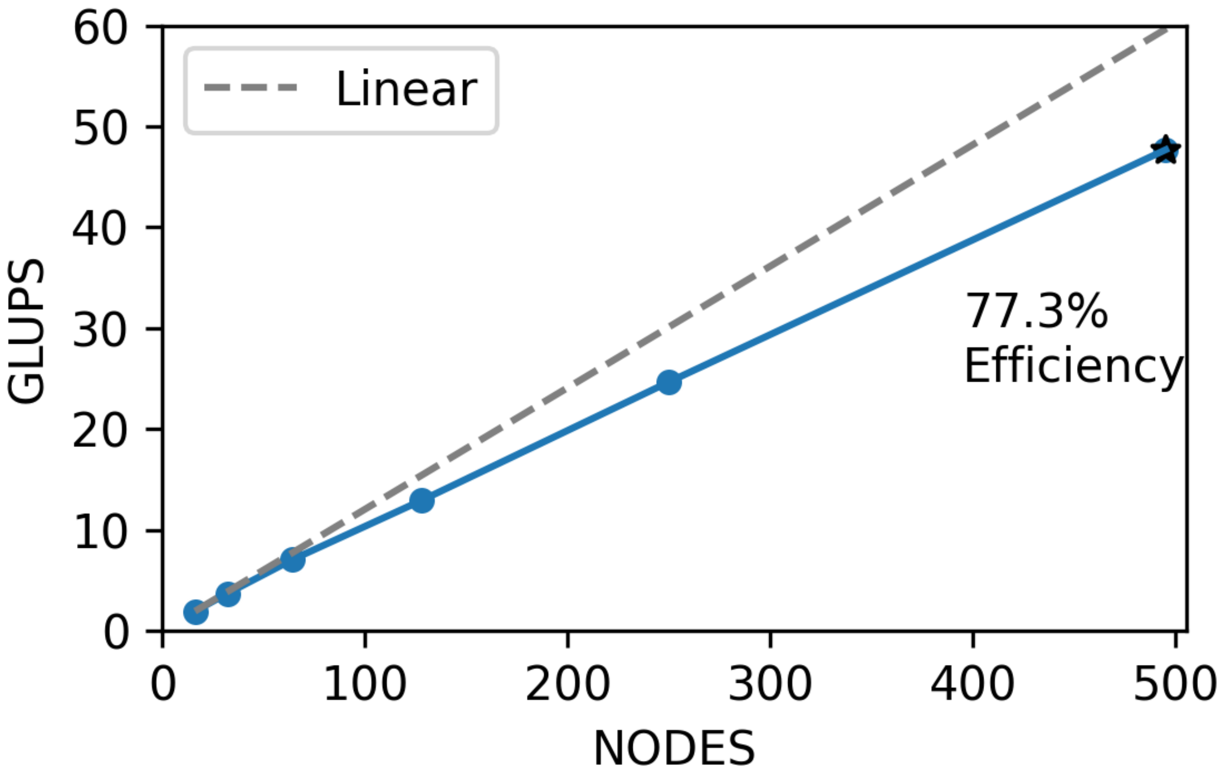
- **EAR Node Manager (EARD) “Daemon”**
  - Energy metrics via the RAPL (Running Average Power Limit) function
  - Global energy limits or just offer global cluster monitoring

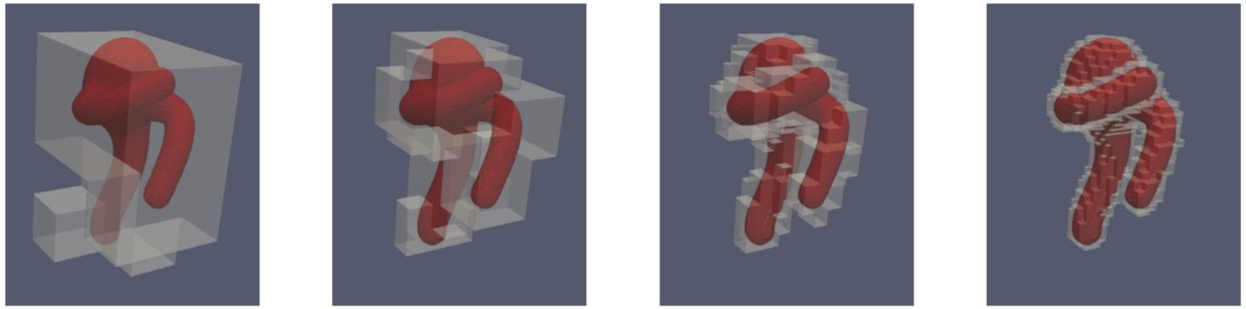
- **EARL** is a library that is loaded next to the application
  - Offers application metrics monitoring
  - Can select the frequencies based on the application behavior on the fly.
  - Integrated with SLURM on Snellius.
  - Intercepts the MPI symbols through the PMPI interface to provide “traces” of MPI applications.

```
#!/bin/bash

#SBATCH --ntasks=256
#SBATCH --time=24:0:0
#SBATCH -p thin --exclusive

#SBATCH --ear=on
#SBATCH --ear-policy=monitoring

module load 2021
module load foss/2021a
module load CMake/3.20.1-GCCcore-10.3.0
module load VASP6/6.2.1-foss-2021a-CUDA-11.3.1

srun vasp_std
```

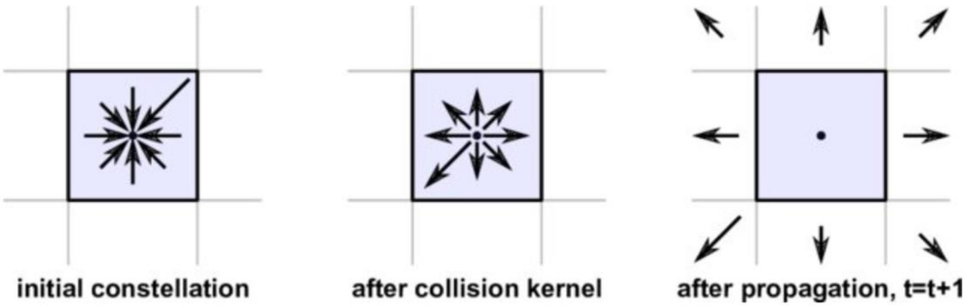
Palabos: Lattice-Boltzmann Solver

Strong Scaling Benchmark



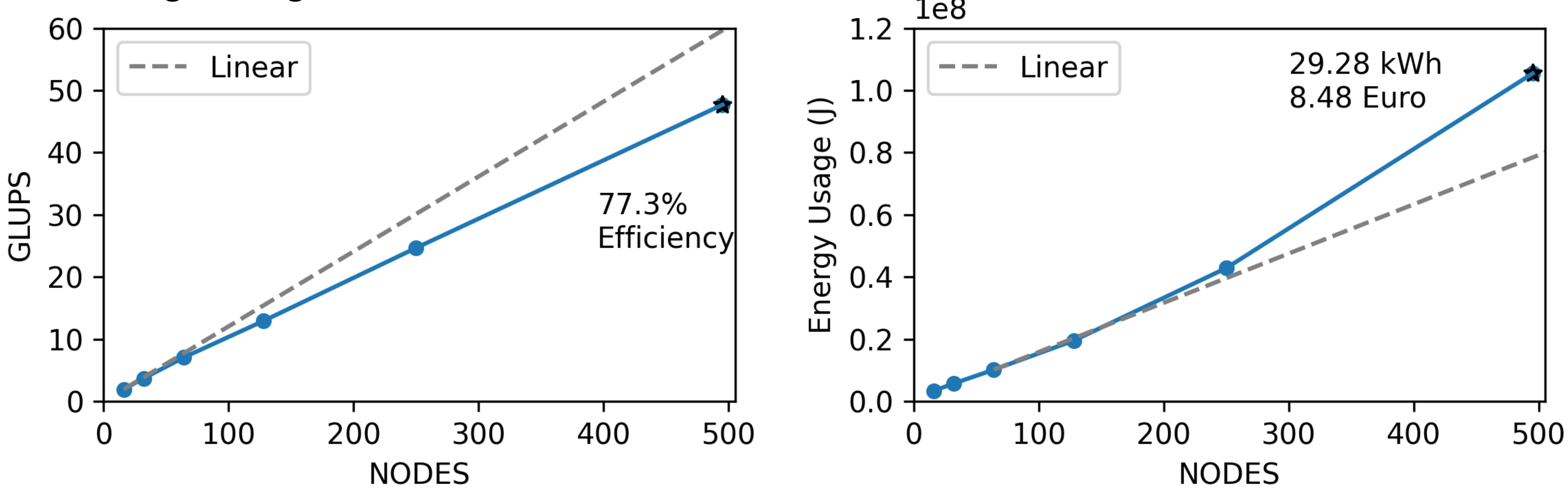


J. Latt, O. Malaspinas, D. Kontaxakis et al. / Computers and Mathematics with Applications 81 (2021) 334–350



Palabos: Lattice-Boltzmann Solver

Strong Scaling Benchmark



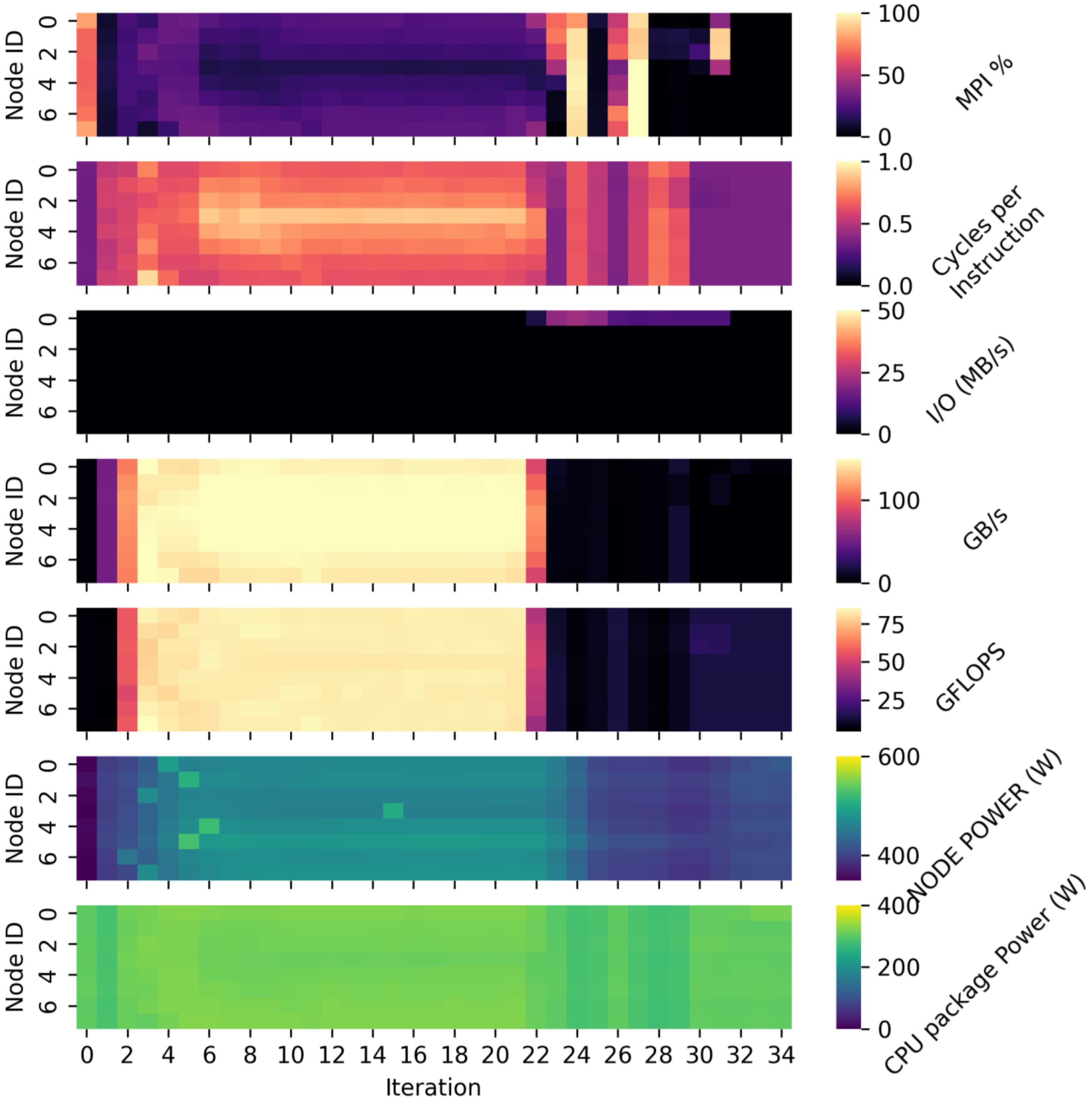
- An average electric car consumes about 0.2 kwh/km
- 495 node case (which ran for 9 minutes) just drove a car to Leeuwarden (149km from SP)!

Palabos:

Strong Scaling Benchmark

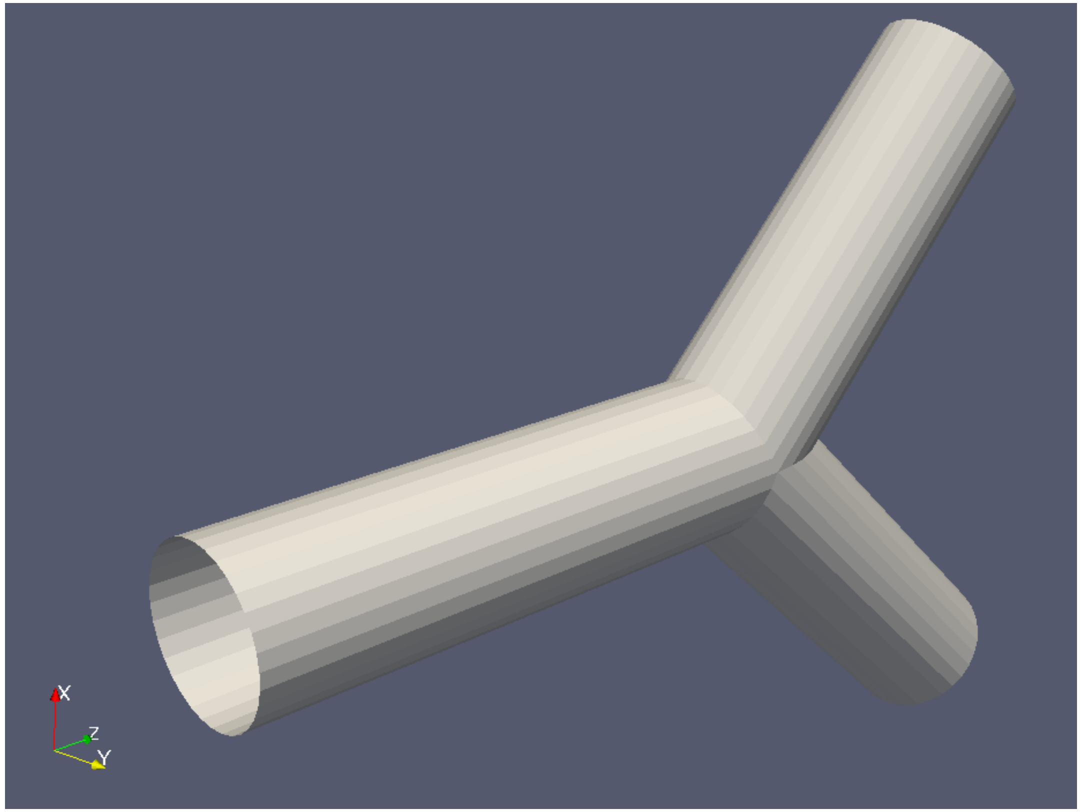
- Per-node, Per-iteration ”traces”

- Node Power
- Avg CPU Freq (node)
- Main memory BW (GB/s)
- CPI (Cycles per instruction)
- MPI% (percentage spent in MPI calls)
- I/O (network communication)



HemeLB:

A Computational Science Use-case



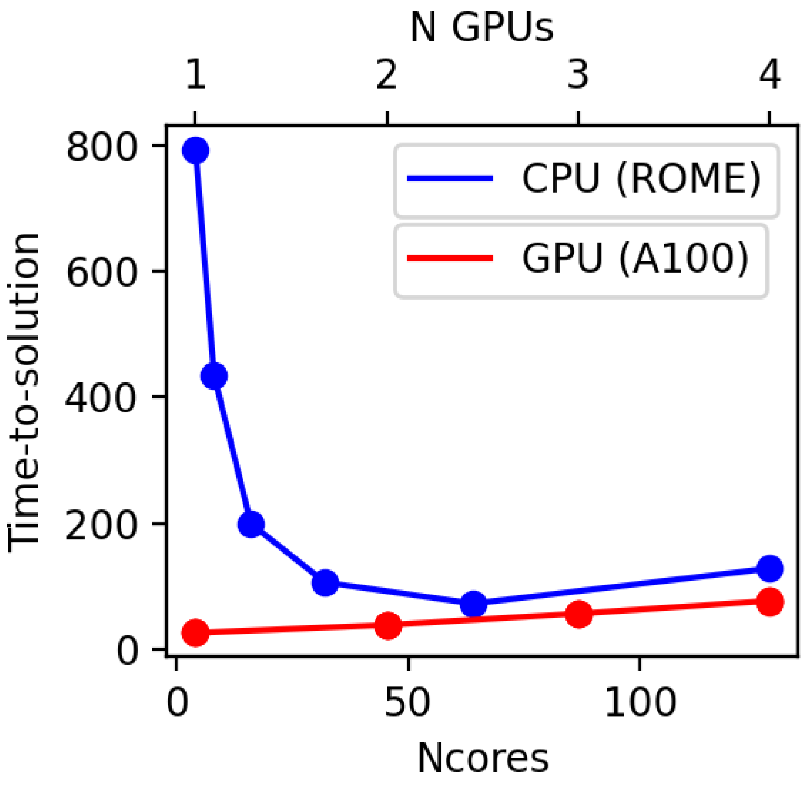
Straw Man argument.

- You are are Dutch computational Science researcher, who has been given a well developed, good scaling, CFD code to do your research. The code has both a CPU implementation, and a GPU (CUDA) implementation.

What is the best choice of resources in terms of ……..

- Performance?
- Budget/Money?
- Energy?

HemeLB:
A Computational Science Use-case

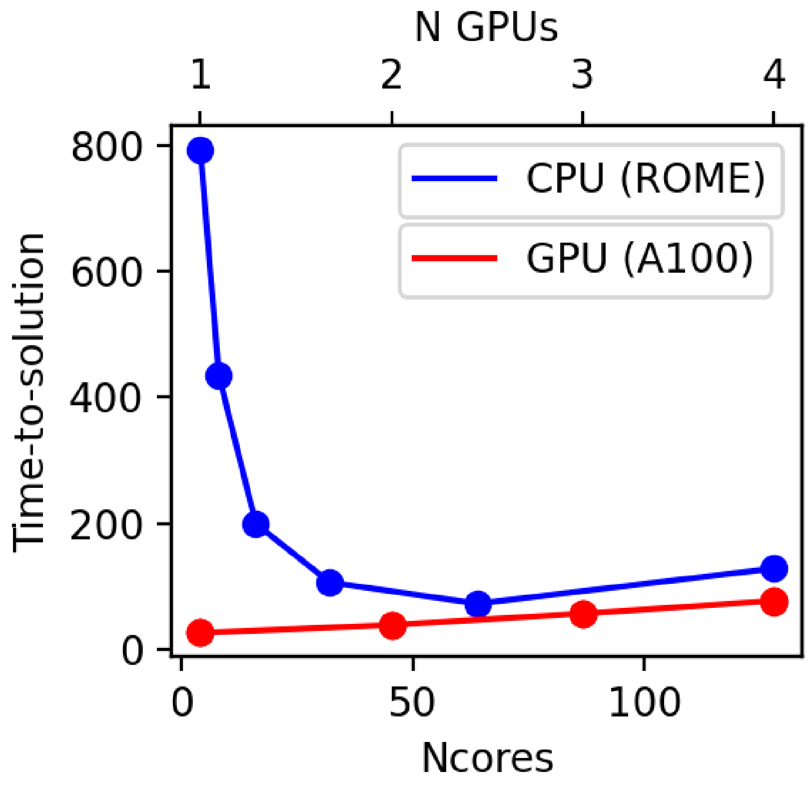


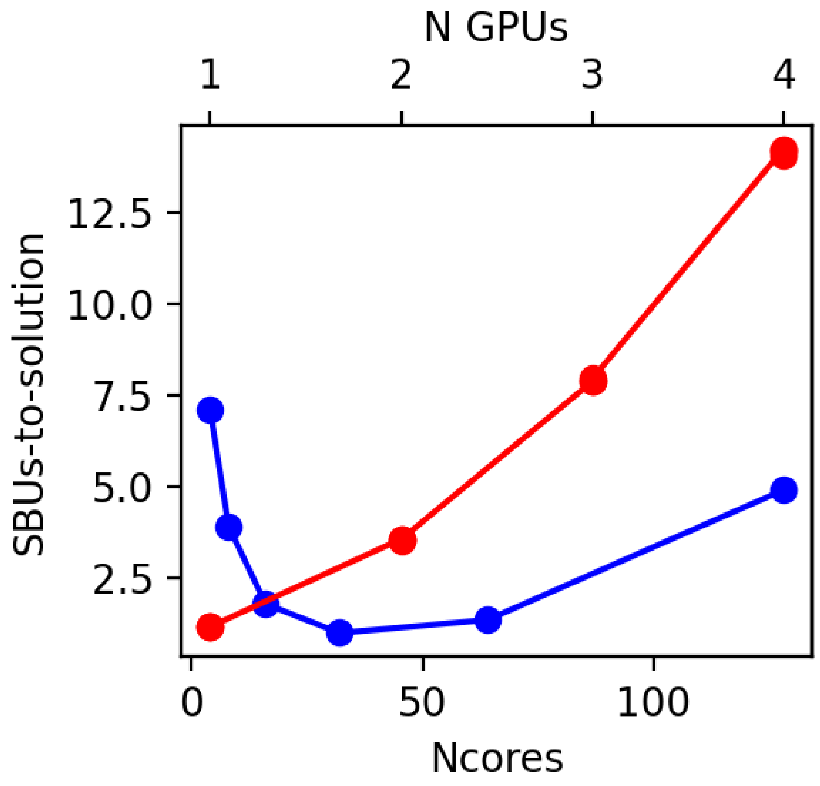
Performance:

1 NVIDIA A100 GPU

HemeLB:

A Computational Science Use-case





Performance:

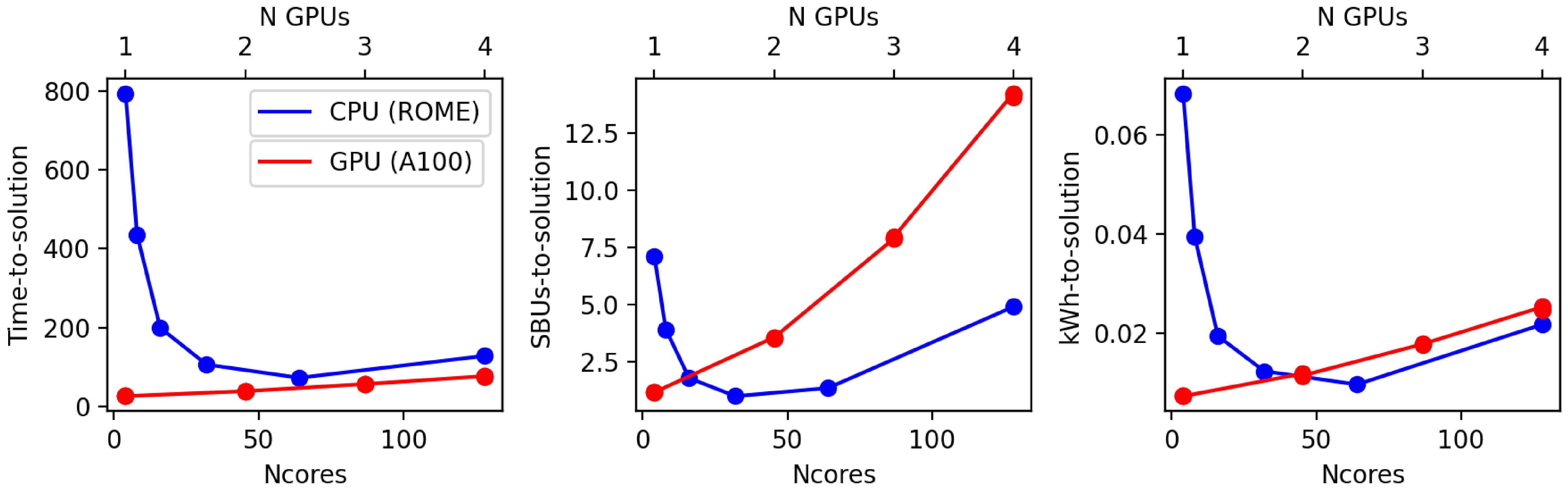
1 NVIDIA A100 GPU

Budget:

32 AMD Rome cores

HemeLB:

A Computational Science Use-case



Performance: 1 NVIDIA A100 GPU

Budget: ¼ AMD Rome

Energy: 1 NVIDIA A100 GPU

Energy Aware Runtime (EAR)

Energy Aware Runtime (EAR) package provides an energy management framework for super computers.

Its Documented and available for use!!!

https://servicedesk.surf.nl/wiki/display/WIKI/Energy

Pages / ... / Energy

EAR example: GROMACS

In this short example we will show you how to run a simple GROMACS benchmark with EAR enabled.

GROMACS (https://www.gromacs.org)

The HECBioSim Benchmarks (https://www.hecbiosim.ac.uk/access-hpc/benchmarks)

> **GROMACS** A free and open-source software suite for high-performance molecular dynamics and output analysis.
>
> **HECBioSim benchmark suite** consists of a set of simple benchmarks for a number of popular Molecular Dynamics (MD) engines, each of which is set at a different atom count. The benchmark suite currently contains benchmarks for the AMBER, GROMACS, LAMMPS and NAMD molecular dynamics packages.

Prepare the SLURM job-script

Notice all that is special here is Lines 10 and 11. This is how you call the EAR runtime library (EARL).

Read more detail here Energy Aware Runtime (EAR) if you want more information on how EAR works.

```
1   #!/bin/bash
2
3   #SBATCH -p rome
4   #SBATCH -n 128
5   #SBATCH -t 00:20:00
6   #SBATCH --exclusive
7   #SBATCH --output=GROMACS_run.out
8   #SBATCH --error=GROMACS_run.err
9
10  #SBATCH --ear=on
11  #SBATCH --ear-policy=monitoring
12
13  module load 2022
14  module load foss/2022a
15  module load GROMACS/2021.6-foss-2022a
16
17  srun --ntasks=128 --cpus-per-task=1 gmx_mpi mdrun -s hEGFRDimer_benchmark.tpr
```

Obviously you need to submit it to the SLURM scheduler!!

So...

```
sbatch myjobscript.sh
```

Outlook

- Identify Energy footprint of users jobs/applications
- Dashboard, Ease of use for the average user to see the “profile” of their applications

Thank you!

- Contact us
  - https://servicedesk.surf.nl/
- EAR User Tutorial on Snellius
  - https://servicedesk.surf.nl/wiki/display/WIKI/Energy+Aware+Runtime+(EAR)
- Energy Aware Runtime (EAR) --- Energy Aware Solutions
  - https://www.eas4dc.com/solutions
  - https://gitlab.bsc.es/ear_team/ear/-/wikis/home

https://github.com/sara-nl/energy-efficient-computing/

https://ondemand.snellius.surf.nl/