EcoCompute 2025

One-pedal green computing: no (workload) shifting required

Oleksiy Kozlov
Staff scientist, Sustainable Computing Advisor

The Exelixis Lab
Heidelberg Institute for Theoretical Studies (HITS gGmbH)
exelixis-lab.org | alexey.kozlov@h-its.org

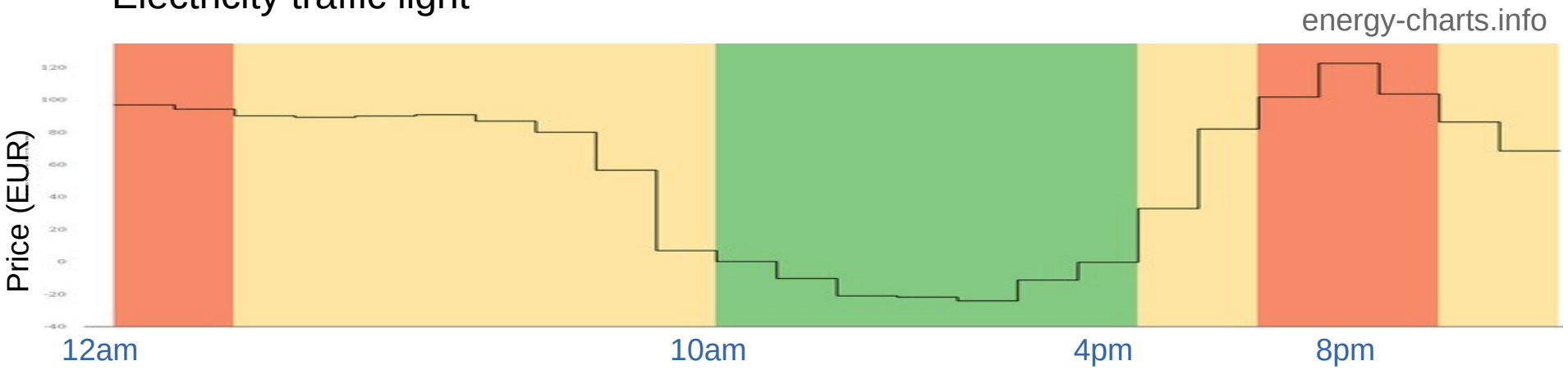
Berlin – November 13, 2025

About

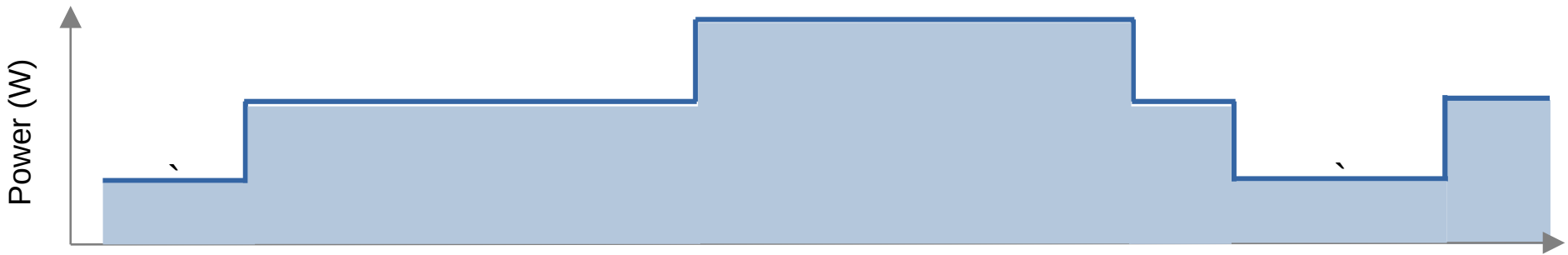
- Heidelberg Institute for Theoretical Studies
  - Private non-profit company (HITS gGmbH)
  - 12 scientific groups, ~100 researchers
  - Computational sciencies: Bioinformatics, astrophysics, statistics, language processing …
- Me
  - Background: Computer Science (2013 M.Sc., 2018 PhD)
  - HPC software for evolutionary bioinformatics (e.g. `RAxML-NG`)
  - Since 2023: Sustainable Computing Advisor (part-time)

Carbon-aware computing

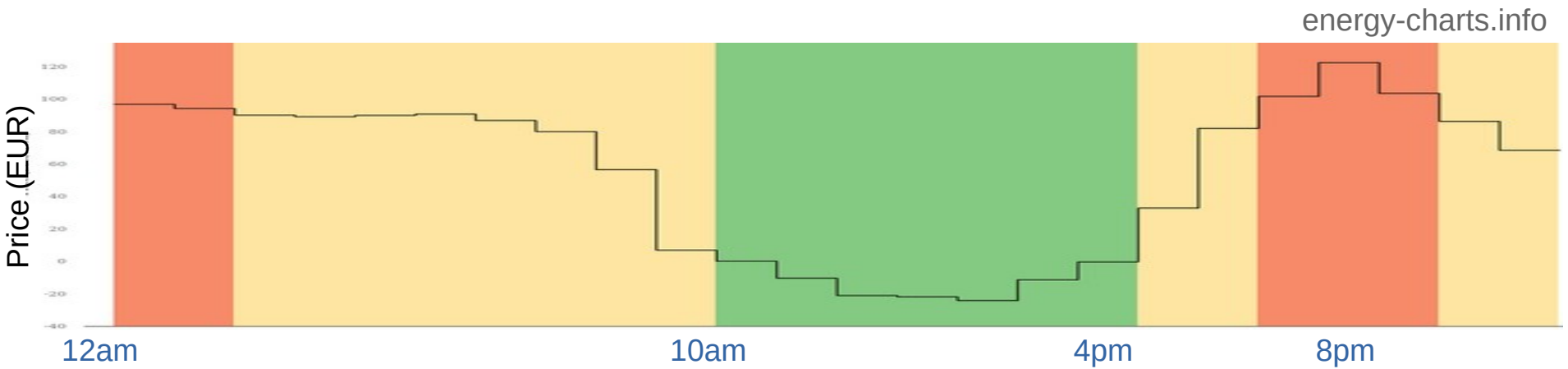
"Electricity traffic light"

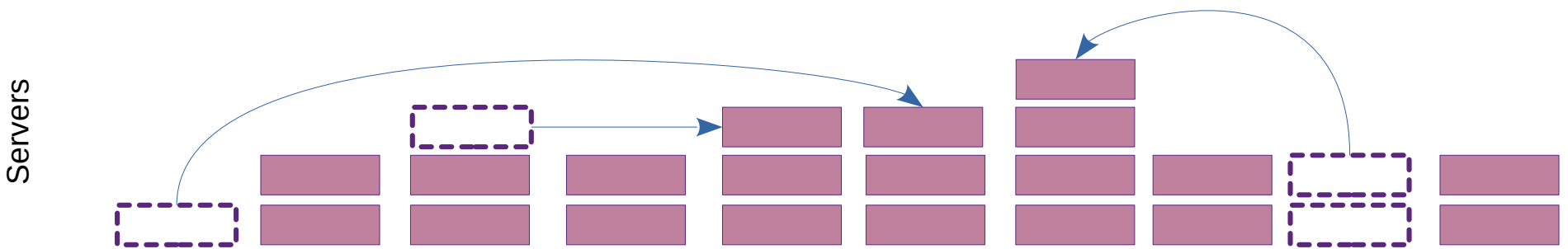


Use less energy in red, more in green phases

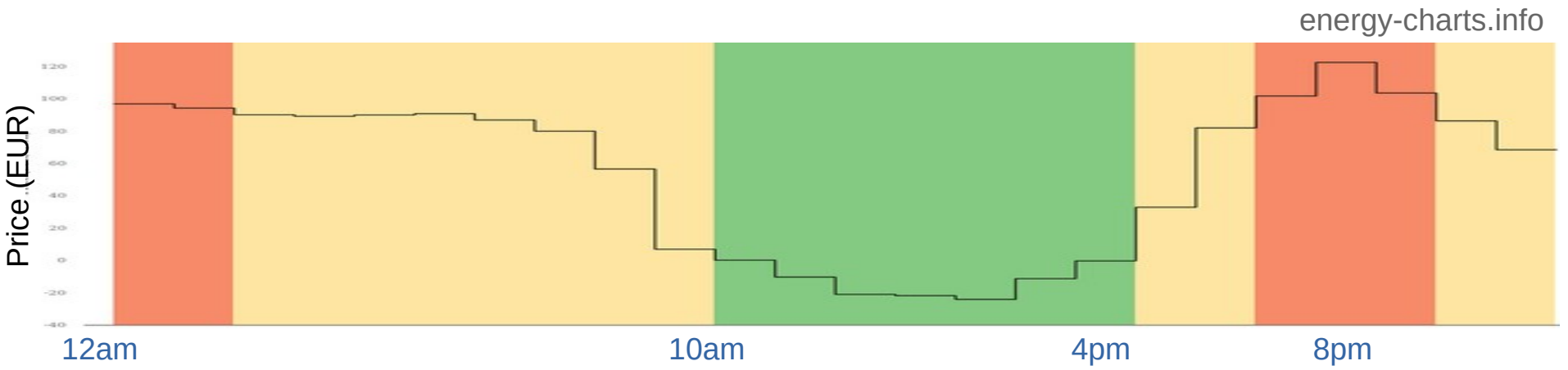


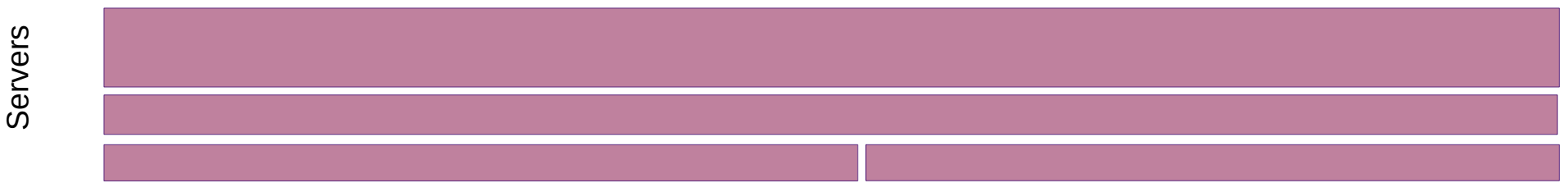
Carbon-aware scheduling





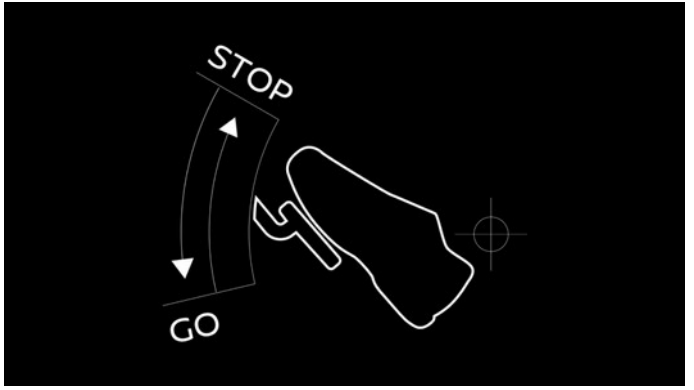
Scheduling nightmare





One-pedal driving

- Speed control with gas pedal only
  - Push → accelerate
  - Release → regen braking
- No mechanical brake use
- No clutch/shifter (obviously)



nissan.com.my

“One-pedal” computing

- In hardware
  - Power cap ~ throttle
  - Frequency cap ~ cruise control
- In software
  - CPU utilization cap
  - (Parallelization, precision...)

- no code modification
- low latency (seconds)
- broad hardware support

Power and frequency capping

- Widely available: Intel/AMD/NVIDIA
- Power and/or frequency limits

- Typical range: 50% - 100% TDP
- Easy-to-use, transparent to workload

NVIDIA RTX 2080 SUPER

```
$ nvidia-smi -q -d POWER,CLOCK
  Power Management                  : Supported
  Power Limit                       : 250.00 W
  Min Power Limit                   : 125.00 W
  Max Power Limit                   : 250.00 W
```

Intel Xeon Platinum 8260

```
$ sudo cpupower frequency-info
  hardware limits: 1000 MHz - 3.90 GHz
  available cpufreq governors: performance powersave
  current policy: frequency should be within 1000 MHz
and 3.90 GHz.
```

NVIDIA H200

```
  Current Power Limit               : 700.00 W
  Min Power Limit                   : 200.00 W
  Max Power Limit                   : 700.00 W
```

```
$ sudo nvidia-smi -pl 200
Power limit for GPU 00000000:17:00.0 was set
to 200.00 W from 250.00 W.
```

AMD EPYC 9684X

```
$ sudo cpupower frequency-info
  hardware limits: 1.50 GHz - 2.55 GHz
  available frequency steps:  2.55 GHz, 2.00 GHz, 1.50
GHz
```

```
$ sudo cpupower frequency-set -u 2000000
Setting cpu: 0
...
```

Utilization capping

With Docker:

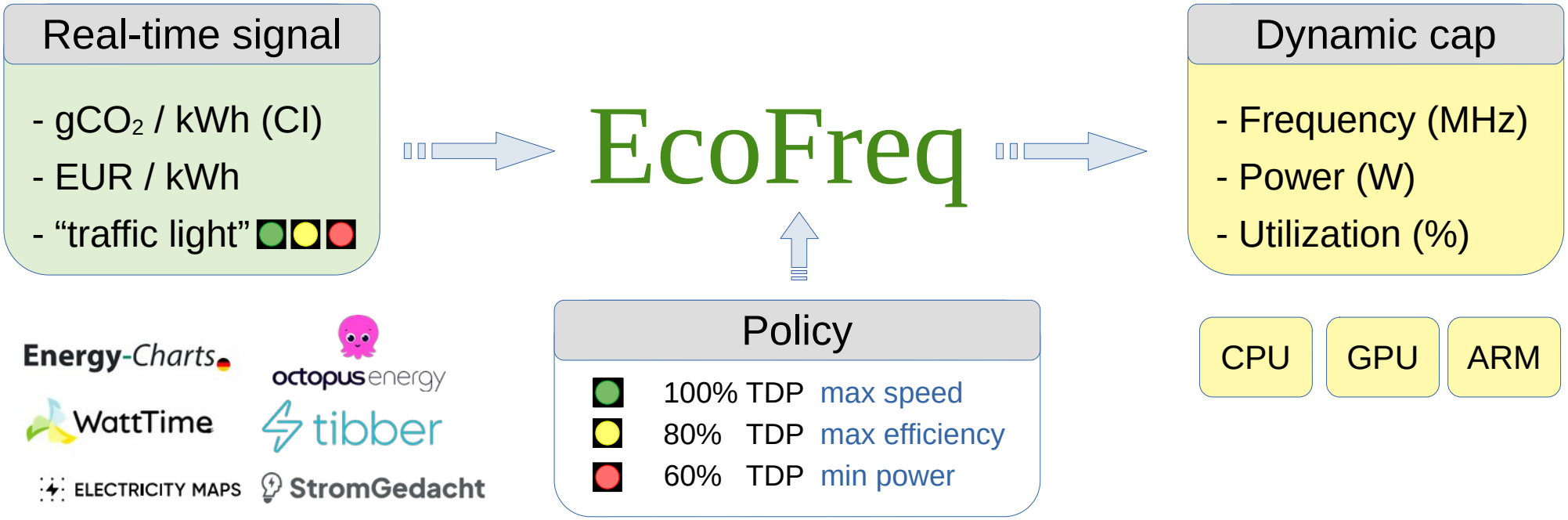
```
$ docker container update --cpus 2.5 my_container
```

(use max 2.5 CPU cores (= 250% utilization) for my_container)

With cgroup:

```
$ echo "2500000" | sudo tee /sys/fs/cgroup/user.slice/cpu.max
```

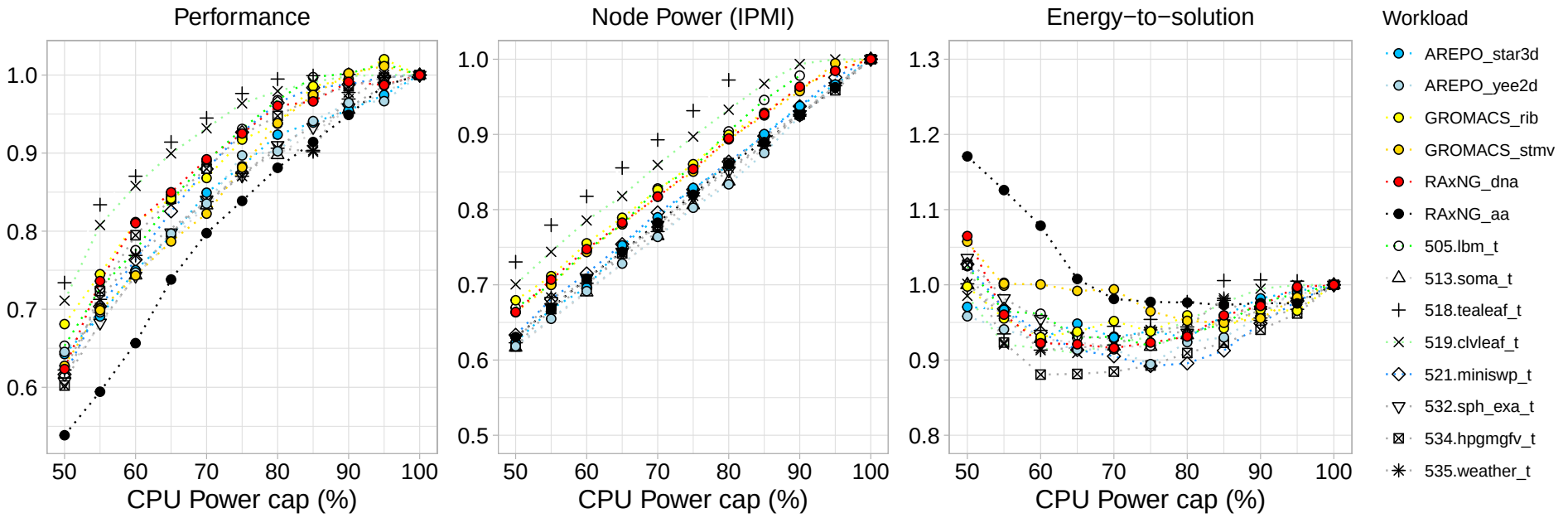
Dynamic carbon-aware power scaling



Will it work?

- Will it damage the chip?
  - Definitely not
- Will it make programs crash?
  - Extremely unlikely
- How do we know?
  - Most OSes and CPUs have dynamic frequency scaling enabled by default

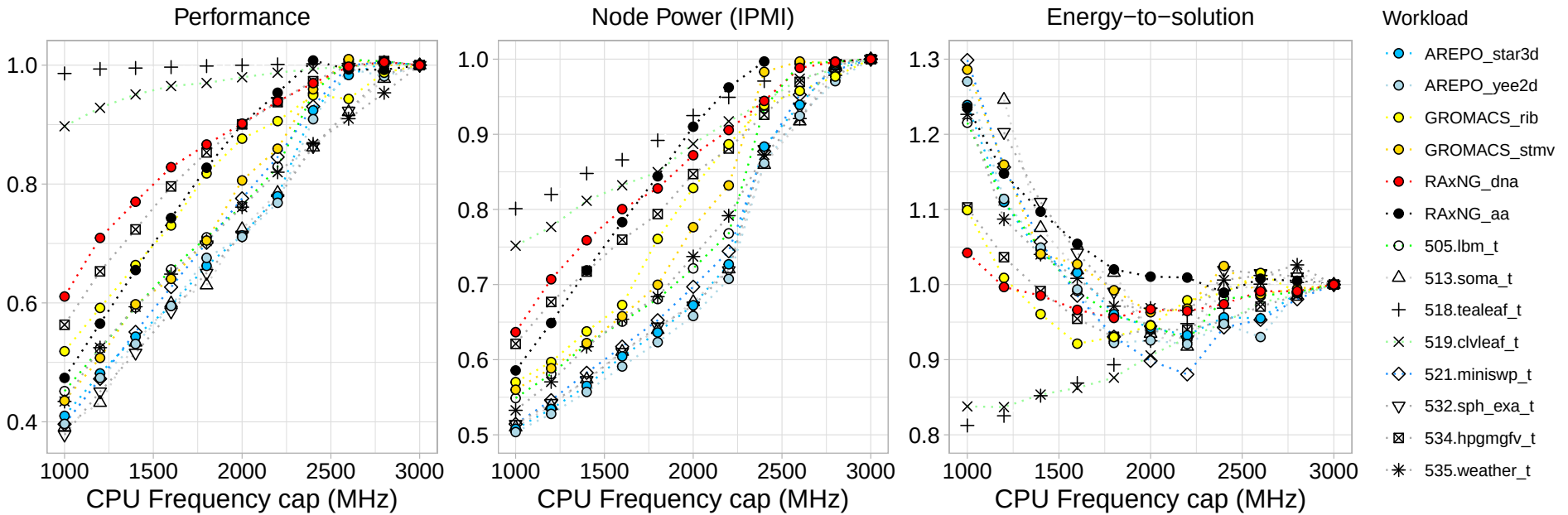
Power profiles: RAPL powercap - CPU



(Kozlov & Stamatakis, ISC 2024)

System: 2S Xeon Platinum 8260, Cascade Lake, 48C, 768 GB RAM

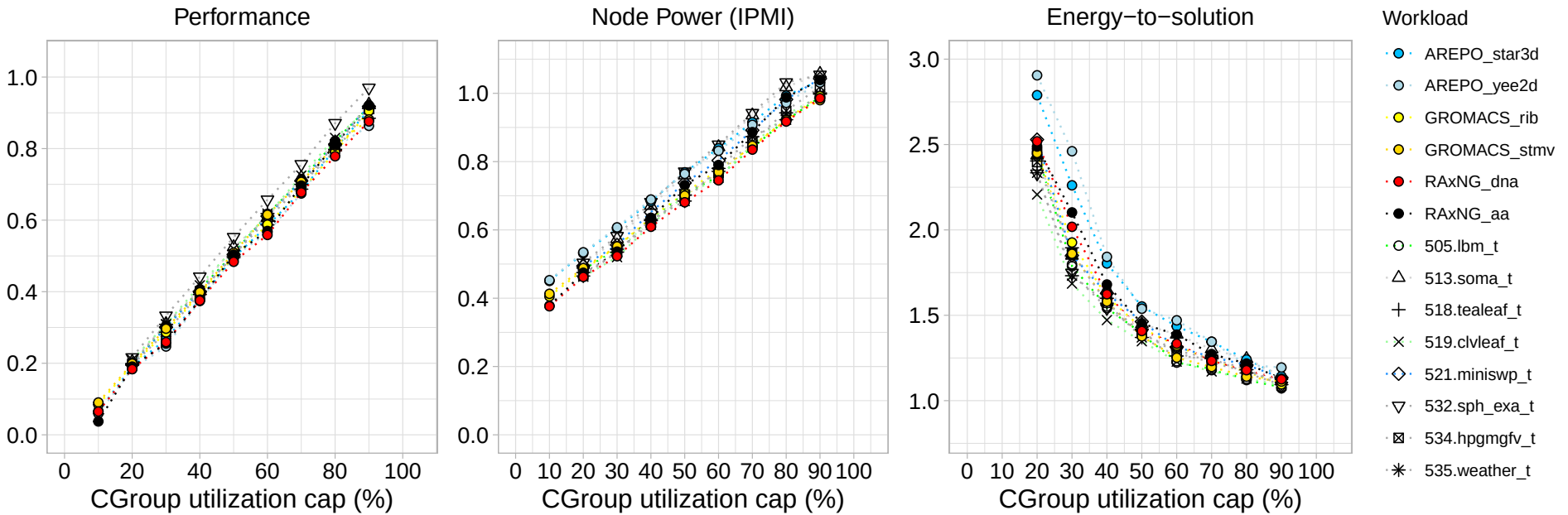
Power profiles: DVFS - CPU



(Kozlov & Stamatakis, ISC 2024)

System: 2S Xeon Platinum 8260, Cascade Lake, 48C, 768 GB RAM

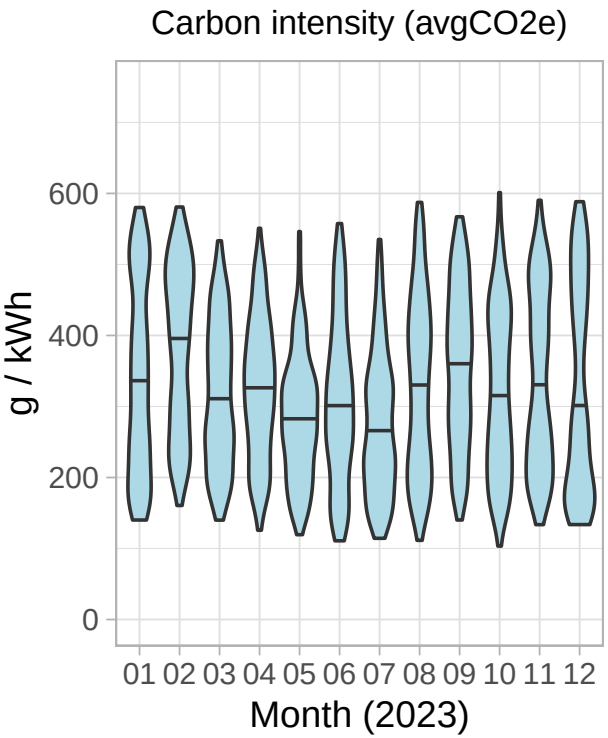
Power profiles: cgroup utilization cap - CPU

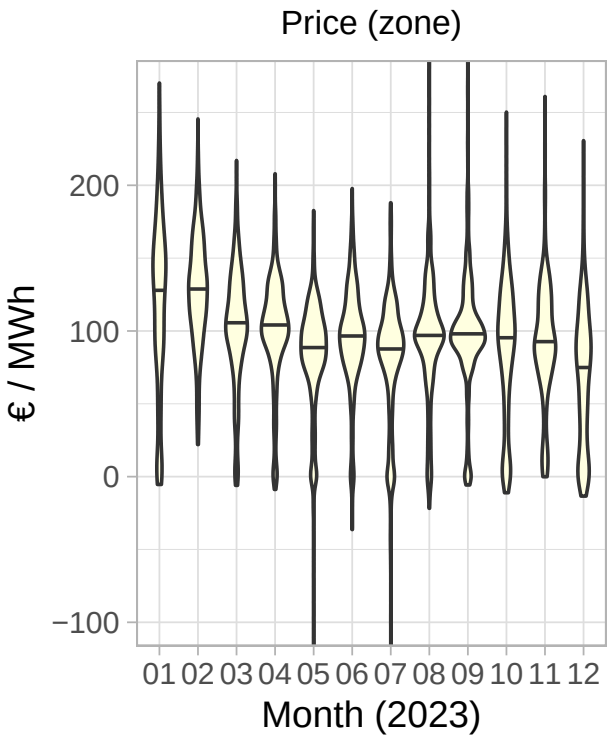


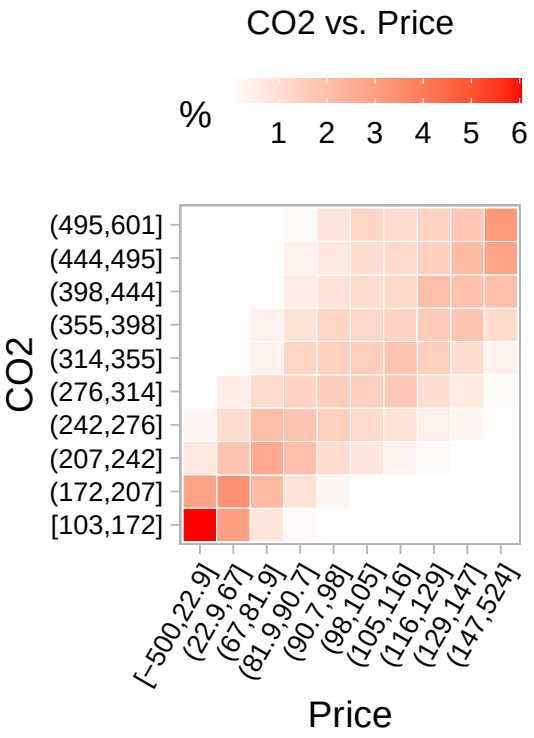
(Kozlov & Stamatakis, ISC 2024)

System: 2S Xeon Platinum 8260, Cascade Lake, 48C, 768 GB RAM

Carbon and price profiles: Germany (2023)







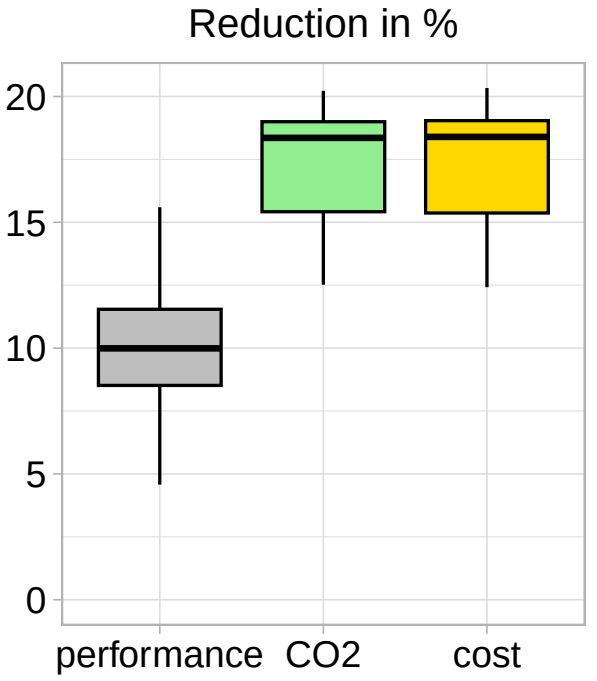
(Kozlov & Stamatakis, ISC 2024)

Evaluation: Germany (2023)

- Policy:

$$P_{lim}(CI) = \begin{cases} 100\%\ \text{TDP} & \text{if } 0\% < CI <= 33\% \\ 80\%\ \text{TDP} & \text{if } 33\% < CI <= 66\% \\ 60\%\ \text{TDP} & \text{if } 66\% < CI <= 100\% \end{cases}$$

- Results:
  - 15-18% lower CO_2 & energy cost
  - @ 8-12% throughput loss

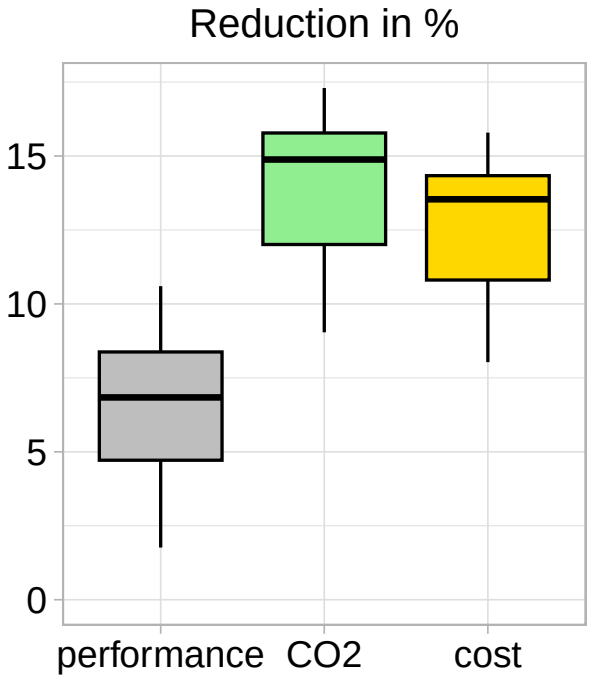


Evaluation: London, UK (Jan-Oct 2025)

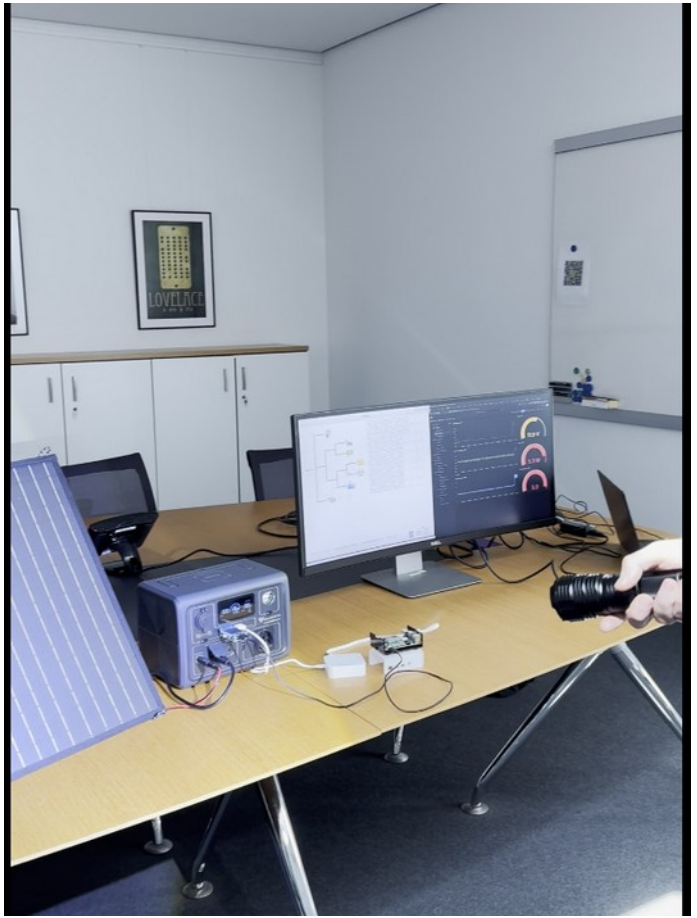
- Policy:

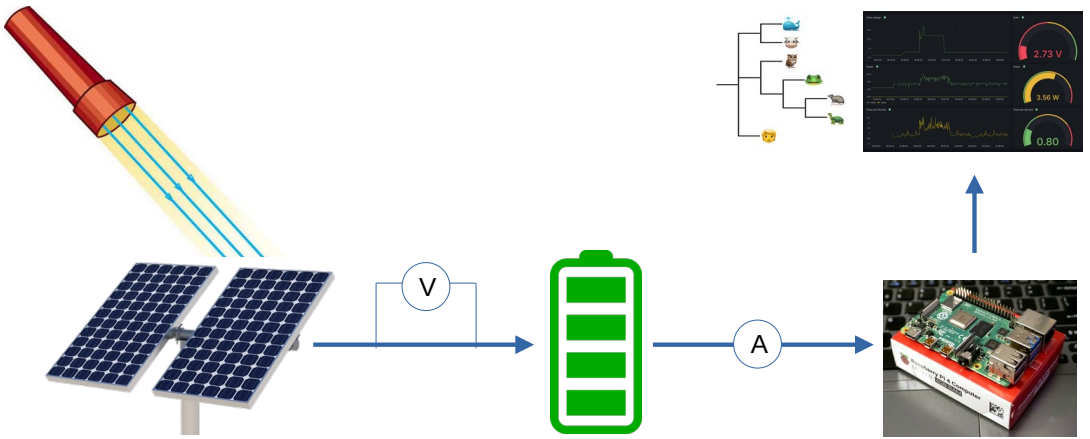
$$P_{lim}(CI) = \begin{cases} 100\%\ \text{TDP} & \text{if } CI \text{ in ['low', 'very low']} \\ 80\%\ \text{TDP} & \text{if } CI \text{ in ['moderate']} \\ 70\%\ \text{TDP} & \text{if } CI \text{ in ['high', 'very high']} \end{cases}$$

- Results:
  - 11-16% lower CO_2 & energy cost
  - @ 5-8% throughput loss



Live Demo: HITS Open House 2024





Deployment scenarios

- Compute clusters
  - Jobs with “eco-mode=on” get higher priority
- Commercial clouds
  - “spot instances 2.0”: cheap, but dimmable
- Webservices
  - Free tier w/o performance guarantees?

Challenges

- Fixed-rate electricity trariffs
- Static electricity taxes & grid fees
- Idle consumption, cooling etc.
- Hardware amortization / embodied carbon
- → most apply to carbon-aware computing in general

Next steps

- Research project 2026/27
  - Evaluate with more HW & SW
  - Integrate with job schedulers (SLURM)
  - Optimize scaling policies
- Pilot system / production
  - Scientific computing (HITS, KIT)
  - Commercial systems: looking for industry partners!

Thanks!

Oleksiy Kozlov
Staff scientist
HITS gGmbH
alexey.kozlov@h-its.org
https://cme.h-its.org/exelixis

Prof. Alexandros Stamatakis
Full professor / Group leader
HITS / KIT / FORTH
https://www.biocomp.gr/

HITS

KIT Karlsruher Institut für Technologie

Funded by the European Union

- Code: https://github.com/amkozlov/eco-freq
- Paper: https://ieeexplore.ieee.org/document/10528928

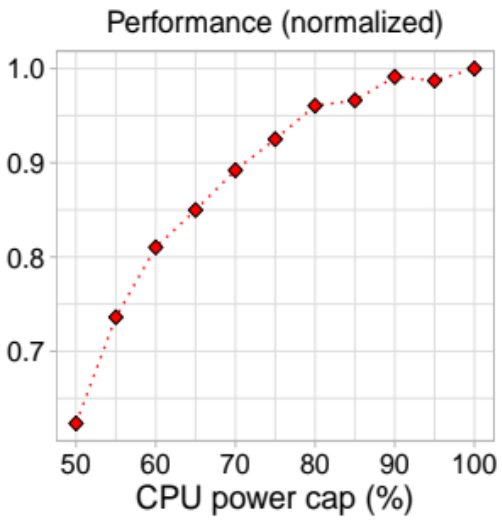
Backup

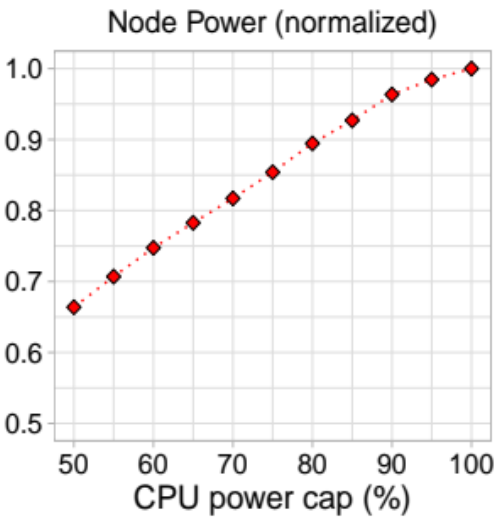
Power scaling methods

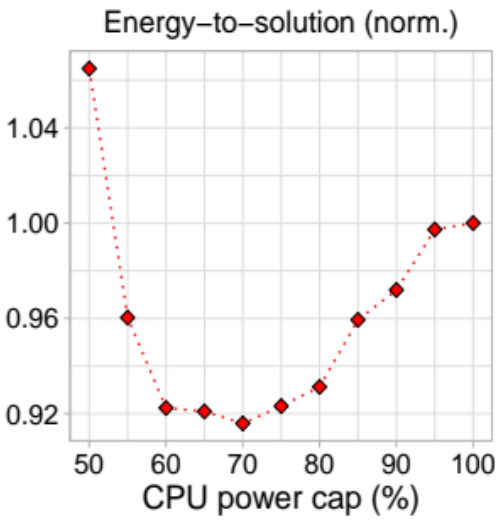
Hardware support		CPU			GPU	
Method	Granularity	x86 Intel	x86 AMD	ARM	NVIDIA	AMD
DVFS	core / chip	+	+	+	+	+
Power cap	chip	+	+/-		+	+
Utilization cap	process	+	+	+		

- No modification to workload / submission scripts

Energy efficiency “sweet spot”

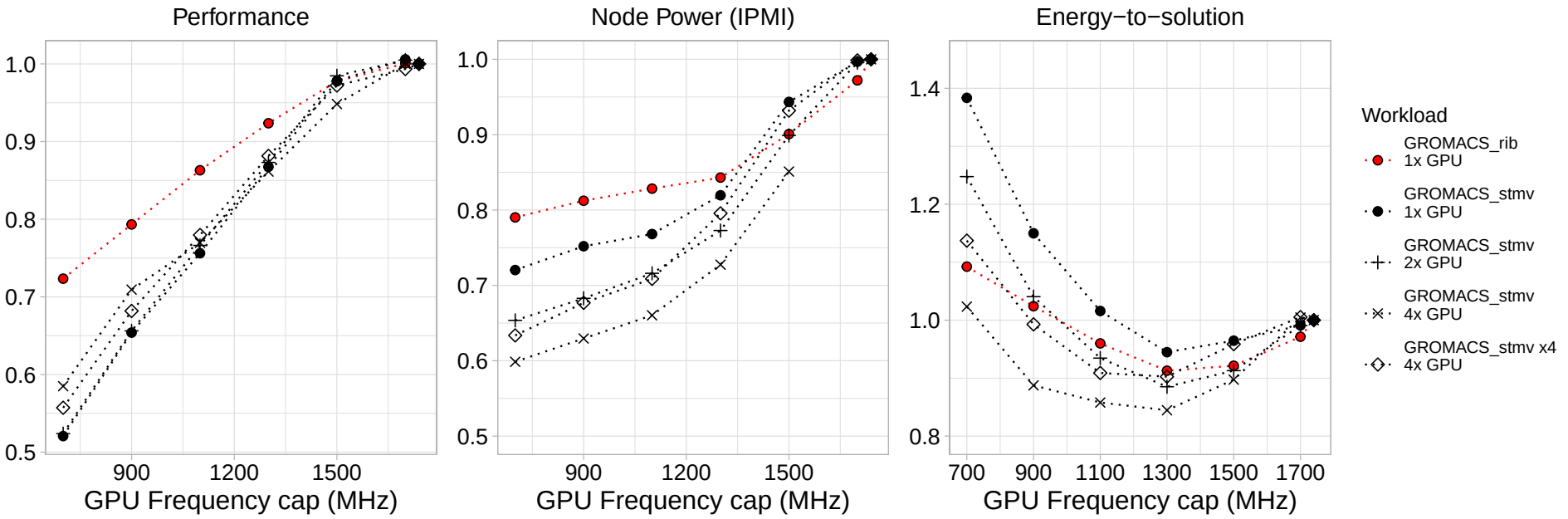






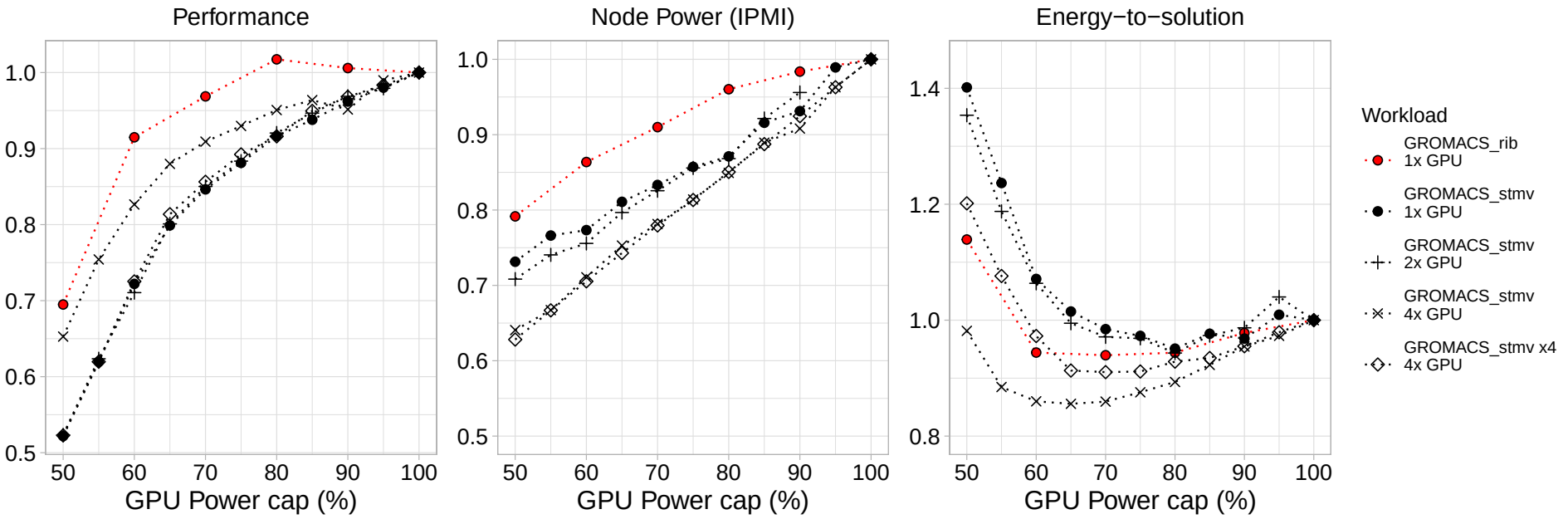
(RAxML-NG 1.1, 2x Intel Xeon Platinum 8260, 48T)

Power profiles: DVFS - GPU



System: 2S Xeon Gold 6330, Ice Lake, 56C + 4x NVIDIA A40

Power profiles: powercap - GPU



System: 2S Xeon Gold 6330, Ice Lake, 56C + 4x NVIDIA A40

Test datasets

	Carbon intensity		Price	
Region	Type	Signal	Market	Type
Germany	average	continious	wholesale	national
UK, London	average	discrete	retail	regional
US, N. California	marginal	continious	wholesale	regional

- 2023 historical data
- Day-ahead market prices

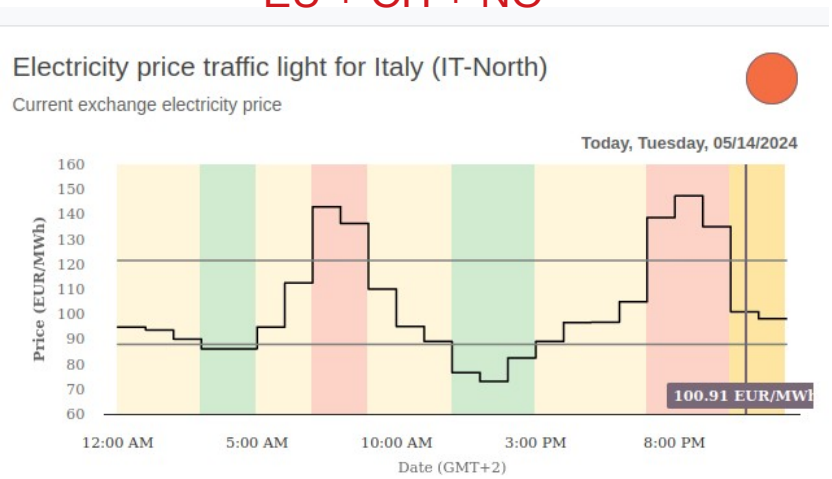
Testbed

- Hardware
  - 2x Intel Xeon CascadeLake
  - 2x Intel Xeon IceLake + 4x NVIDIA A40
- Software
  - SpecHPC 2021
  - Real-world tools: Arepo, Gromacs, raxml-ng
  - 14 workloads

Carbon “traffic light”

- Discrete signal, e.g. green / yellow / red
- Ideally, reflects *local* marginal CI

EU + CH + NO



https://energy-charts.info

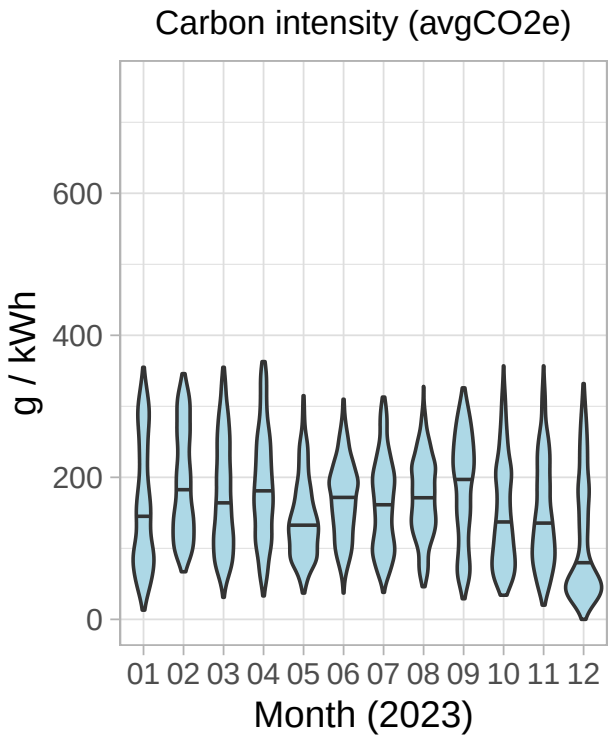
UK

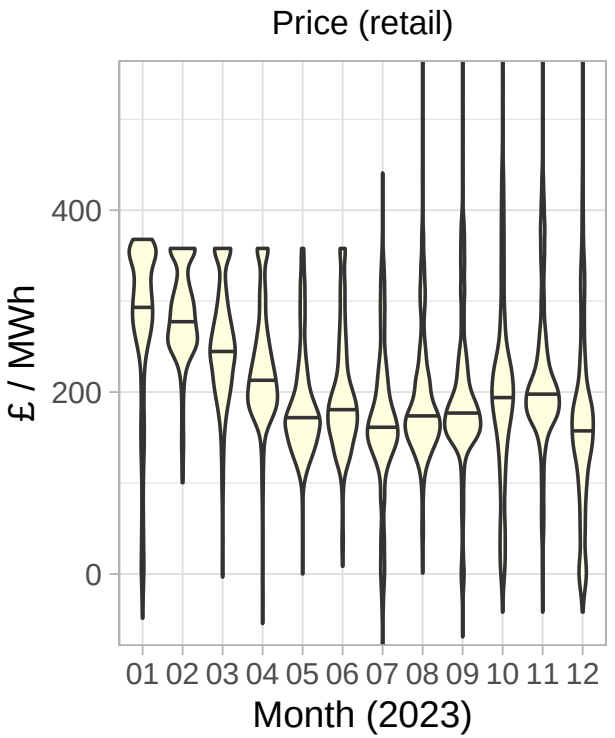
https://carbonintensity.org.uk/

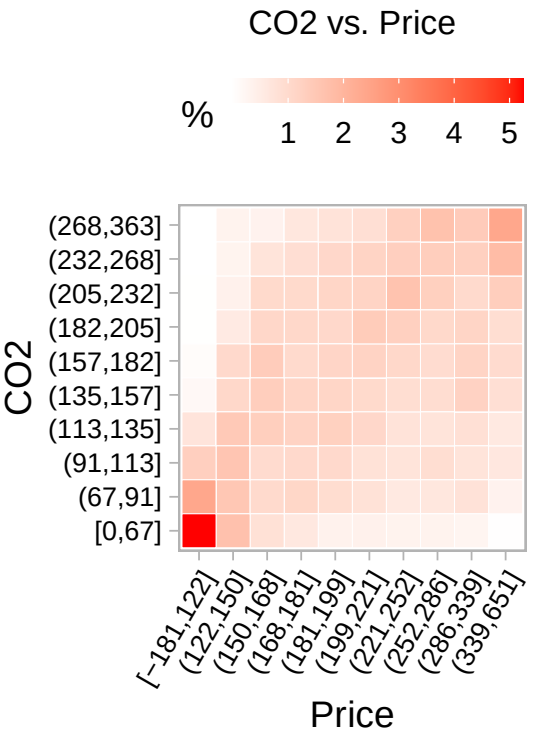
Germany, BW

https://www.stromgedacht.de/

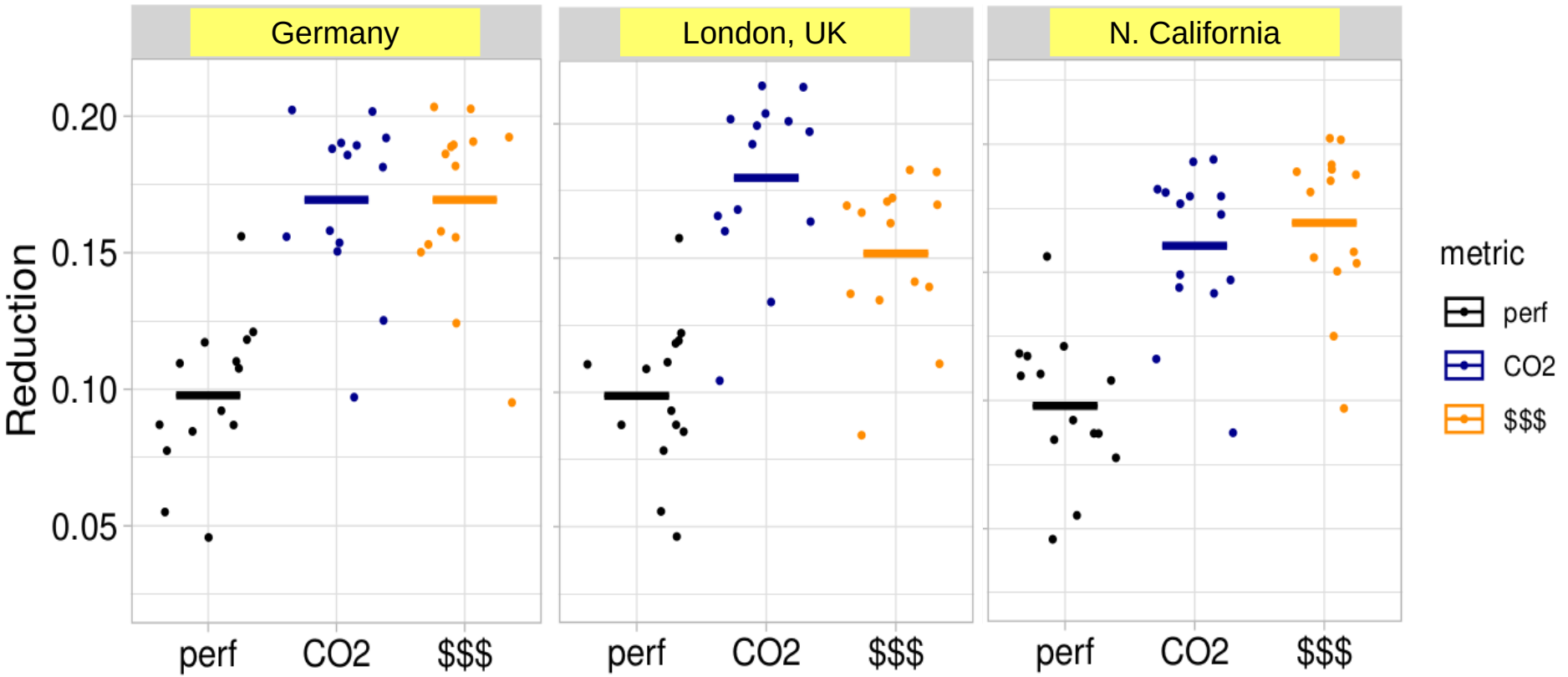
Carbon and price profiles: London (2023)







Evaluation



2023 electricity market data; 3-step scaling policy (100%/80%/60%); 14 HPC workloads, Xeon Cascade Lake