MPIA Sustainability Tea – 4.12.2020
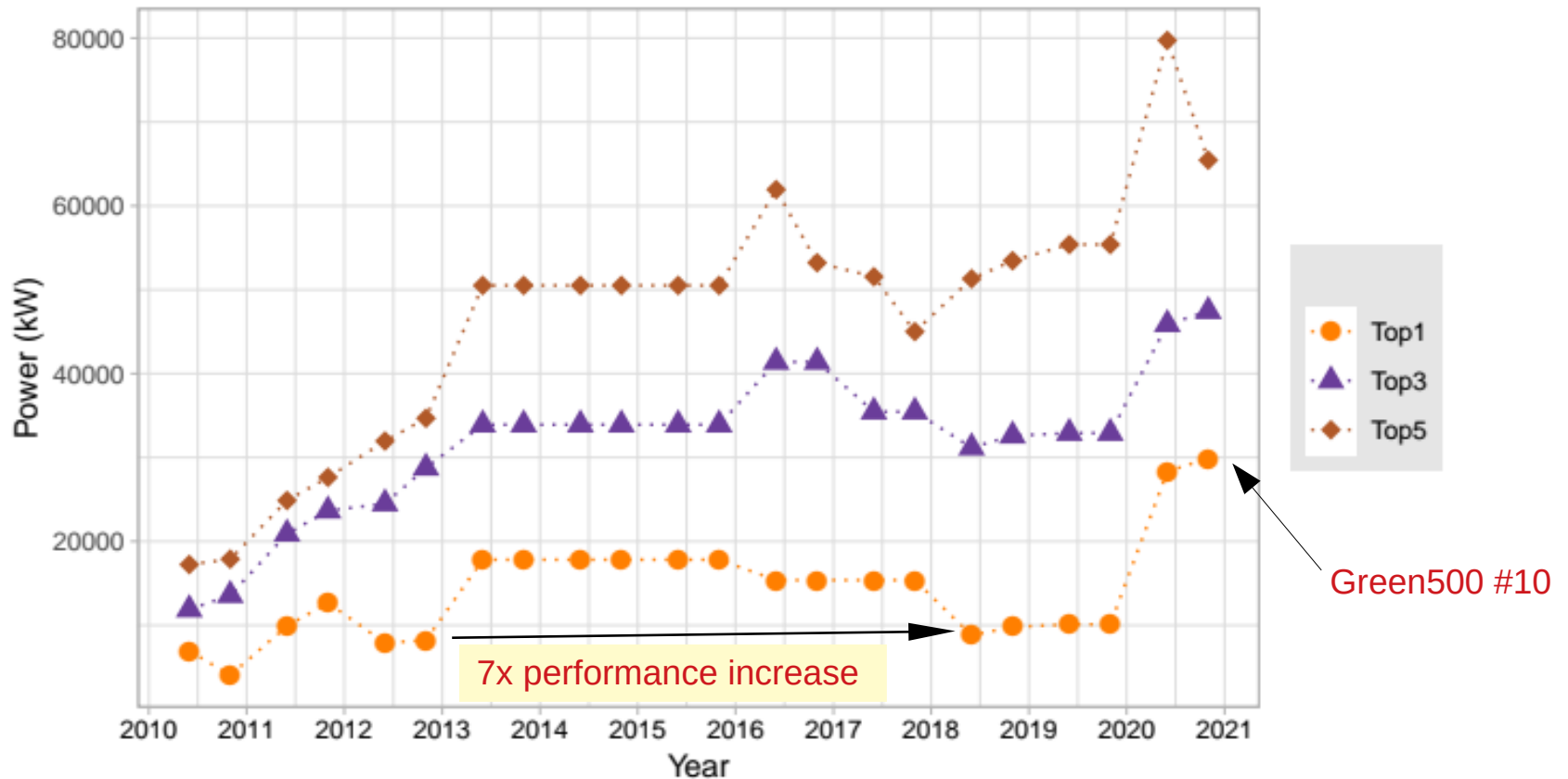
Random thoughts on GreenHPC

Alexey Kozlov

whoami

- Since 2018: staff scientist @ HITS
- 2013–2018: PhD in evolutionary bionformatics
- Background in Computer Science
- HPC: parallelization, performance tuning...
- Fast code is green code?

Top500, Green500, Cap500?

7x performance increase

Green500 #10

What do we need?

Measure (actual energy consumption / CO2)

Optimize

Set specific goals

(e.g. -5% p.a. or

-15% in 5 years etc.)

State of the Art

- Hardware (Intel, NVIDIA)
  - Support for power measurement & capping
- Software (Linux)
  - Lack of standardization
  - Root privileges required (often)
  - Not user-/developer-friendly
  - Serious bugs / design flaws

Energy accounting

- User-space measurement with RAPL → incomplete (CPU+DRAM)
  - Experiment impact tracker:
    https://github.com/Breakend/experiment-impact-tracker
  - RAxML-NG (my code):

```
Elapsed time: 42846.287 seconds

Consumed energy: 162370.469 Wh (= 812 km in an electric car, or 4059 km with an e-scooter!)
```

Energy accounting: SLURM

- SLURM has some support, but:
  - RAPL plugin is broken, better: IPMI/XCC
    - https://bugs.schedmd.com/show_bug.cgi?id=9956
  - Energy statistics is not always saved to DB
  - Information is "hidden" → awareness=0

```
[kozlovay@haswell-login ~]$ scontrol show node haswell-001 | grep CurrentWatts
   CurrentWatts=176 LowestJoules=624884 ConsumedJoules=591343657
```

```
[kozlovay@haswell-login rokas_A7]$ sacct -j 2256100 -o ConsumedEnergy
ConsumedEnergy
--------------
      332.60K  Joules!
```

# Optimization objective?

- Workstation or group-owned server
  - Low to medium utilization
  - Sleep-when-idle → later
  - Energy-to-solution and/or runtime
- Cluster or supercomputer
  - Mostly 100% utilization (=no idle nodes)
 - Power? → inefficient code is green :)

Idle consumption

- Example: Intel i7-7800X, 6 cores, 64GB RAM
  - sleep: 5W, idle: 55W, under load: 150W
  - 50% utilization $\rightarrow$ 25% savings (219 kWh/a)
  - 30% utilization $\rightarrow$ 42% savings (306 kWh/a)
- Suspend + Wake-on-LAN
  - Setup on Linux: adventurous, Win/OSX: better?
  - Ubuntu: `powernap` package for "smart" suspend
  - Interference with monitoring tools (checkmk etc.)

Multi-threading

Execution time (s) — Consumed energy (Wh) — Average power (W)
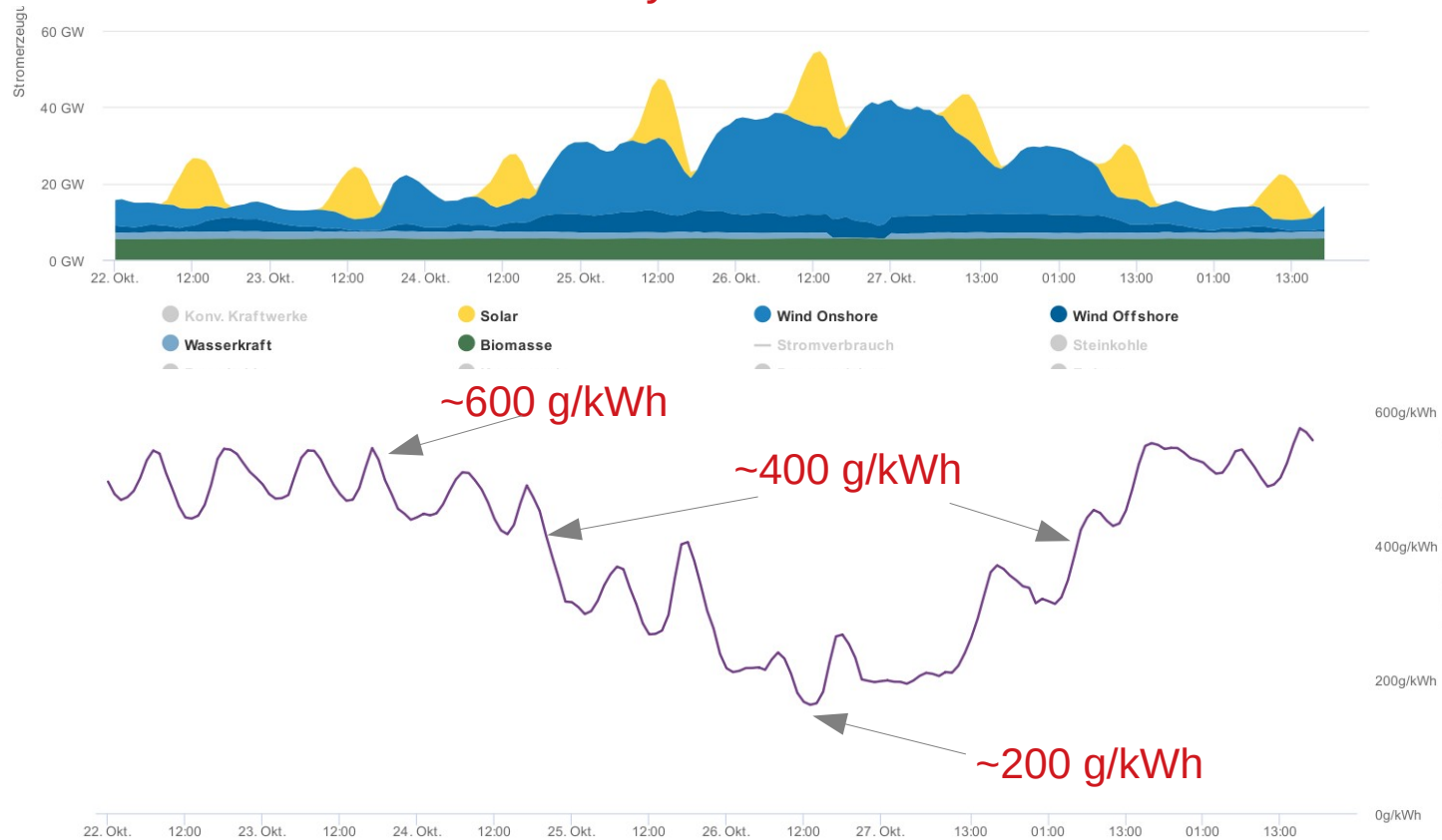
-30%

default: v1.0+   v0.9
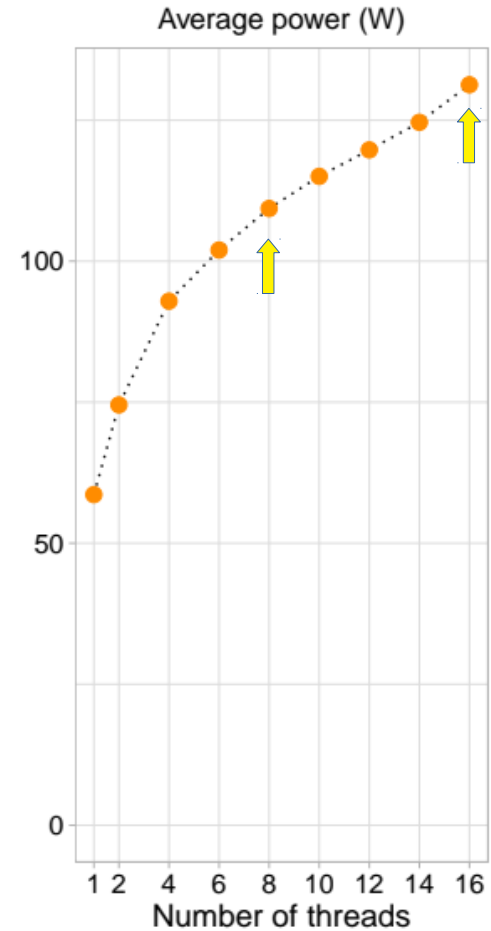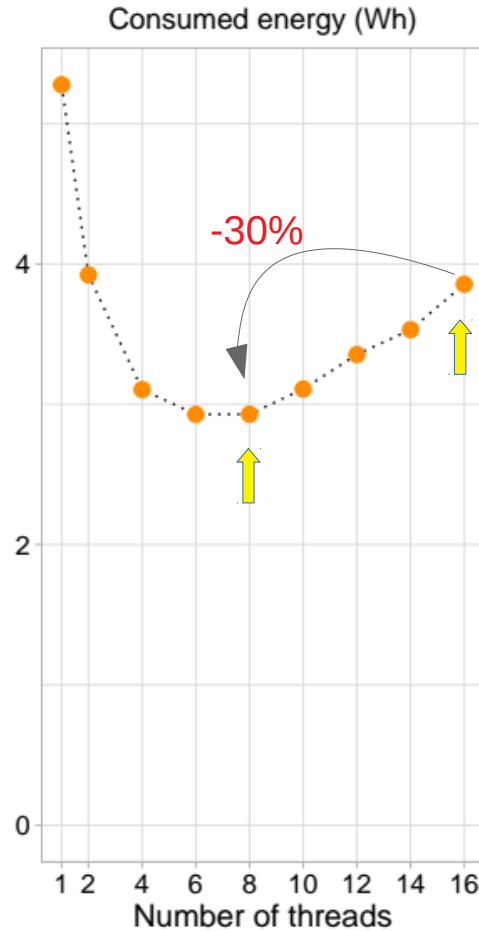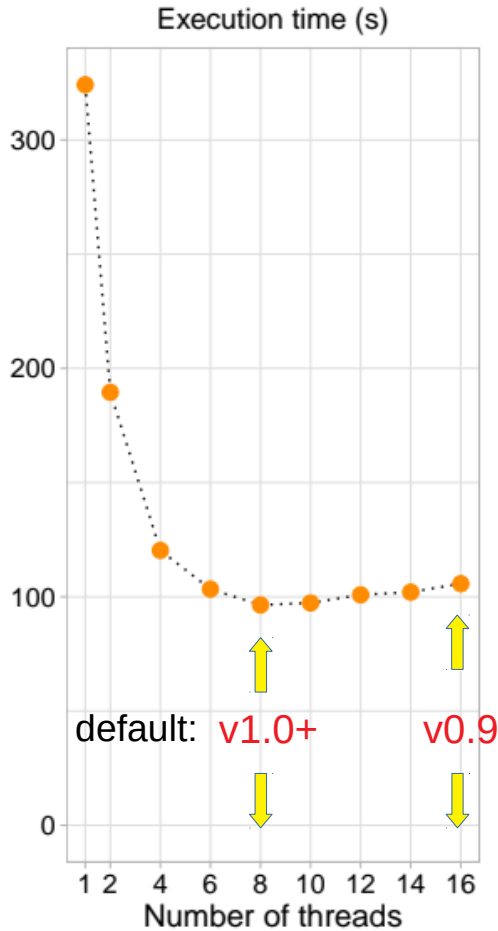
CPU power capping / throttling
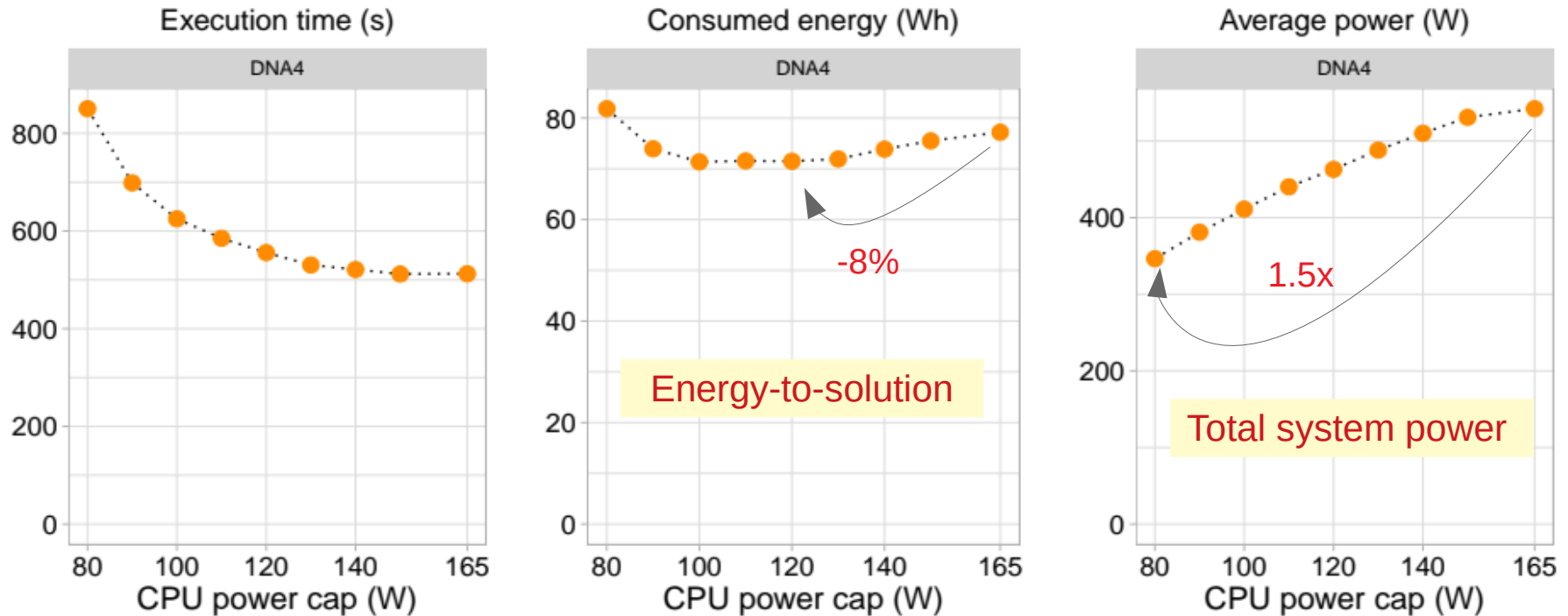
- CPU freq auto-tuning, e.g. with EAR: https://gitlab.bsc.es/ear_team/ear

Follow the energy mix

Germany, October 2019

Discussion