High-performance algorithms for low-power sustainable hardware in HEP at Valencia

CHEP 2024, Kraków, Poland

Antonio Cervello, Francesco Curcio, Álvaro Fernández, Luca Fiorini, Hector Gutierrez, Francisco Hervas, Valerii Kholoimov, Brij Kishor Jashal, Arantza Oyanguren, **Volodymyr Svintozelskyi**, Alberto Valero, Jiahui Zhuo

IFIC, Univ. of Valencia and CSIC (ES),
Rutherford Appleton Laboratory (UK)

Outline

- HIGH-LOW project at Valencia

- Motivation and strategy of power consumption studies

- System power consumption of a typical server

- Studies of power consumption vs hardware utilisation

- FPGAs for computation acceleration

- Conclusions

HIGH-LOW project at Valencia

Design of High-Performance algorithms for low-power sustainable hardware for LHC experiments and their upgrades

- *Aim:* Benchmarking **new hardware** architectures and **developing** fast and high-efficient **algorithms** with **reduced power consumption**

- Transversal project: *ATLAS* and *LHCb* experiments
- PIs: Luca Fiorini (ATLAS), & Arantza Oyanguren (LHCb)
- Funded by the Spanish Ministry of Science and Innovation (TED2021-130852B-I00)
- About 10 people (physicists + engineers + students)

HIGH-LOW activities

- Downstream tracking for LHCb the first high-level trigger

- ACAP processing for ATLAS TileCal reconstruction

- Development of high-speed simulations with hardware accelerators

- Measurement of software and hardware power consumption for HEP

- And many more

Topic of this work

Downstream tracking and vertexing at the first stage of the LHCb trigger

📅 22 Oct 2024, 16:15
🕐 18m
📍 Room 1.C (Small Hall)

Talk | Track 2 - Online and ... | Parallel (Track 2)

Speakers

👤 Brij Kishor Jashal (RAL, TIFR and IFIC)
👤 Jiahui Zhuo (Univ. of Valencia and CSIC (ES))

Versal ACAP processing for ATLAS-TileCal signal reconstruction

📅 24 Oct 2024, 15:00
🕐 18m
📍 Room 1.C (Small Hall)

Talk | Track 2 - Online and ... | Parallel (Track 2)

Speaker

👤 Francisco Hervas Alvarez (Univ. of Valencia and CSIC (ES))

Porting MADGRAPH to FPGA using High-Level Synthesis (HLS)

🏷 MON 26
📅 21 Oct 2024, 15:18
🕐 57m
📍 Ground floor lobby

Poster | Track 5 - Simulation ... | Poster session

Speaker

👤 Hector Gutierrez Arance (Univ. of Valencia and CSIC (ES))

HIGH-LOW hardware

HL hardware:

- Rack K RETEX LOGIC-2 A600 42U F1000 PH
- APC Metered Rack PDU ZeroU 2G AP8
- SWITCH D-LINK DXS-1210-28T 24x 10GB

- T10G Dual Xeon Scalable HPC 10xGPU PCIe
  - 2 x Intel® Xeon® Gold 5318Y 2,10GHz 24 Cores 3.40 GHz
  - 8 x 32GB DDR4 3200MHz ECC REG
  - 1 x SSD Samsung 990 PRO 2TB M.2 NVMe 2280 PCIe 4.
  - 1 x Controladora BROADCOM MegaRAID 9560-16i PCIe
  - 8 x HD 10TB SAS 12Gb/s 7.200 rpm
  - 1 x NVIDIA® RTX™ A5000 24GB GB GDDR6 ECC
  - 1 x NVIDIA® RTX™ A6000 Ada Generation 48GB GB GDDR6 ECC
  - 1 x HBA Broadcom N210GBT Dual 10GbE RJ45

Used by LHCb

Fan speed measurement device

5
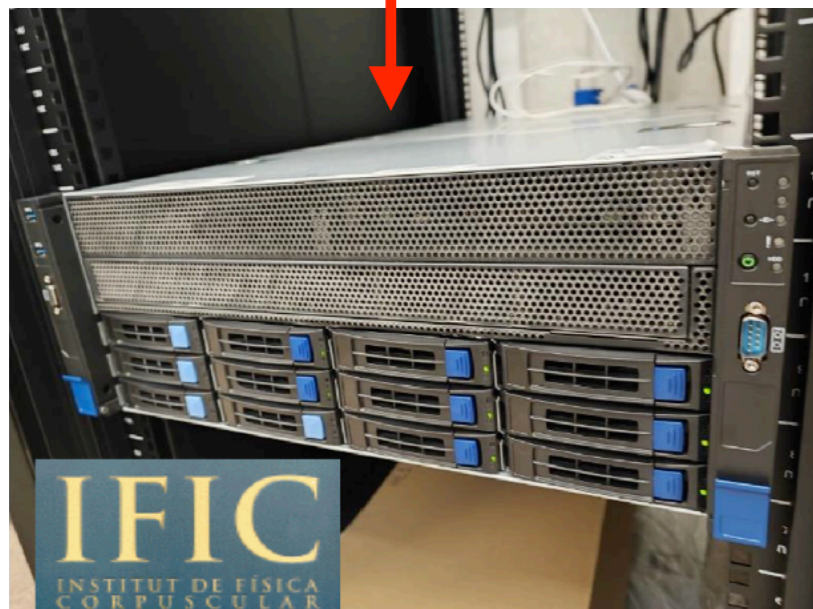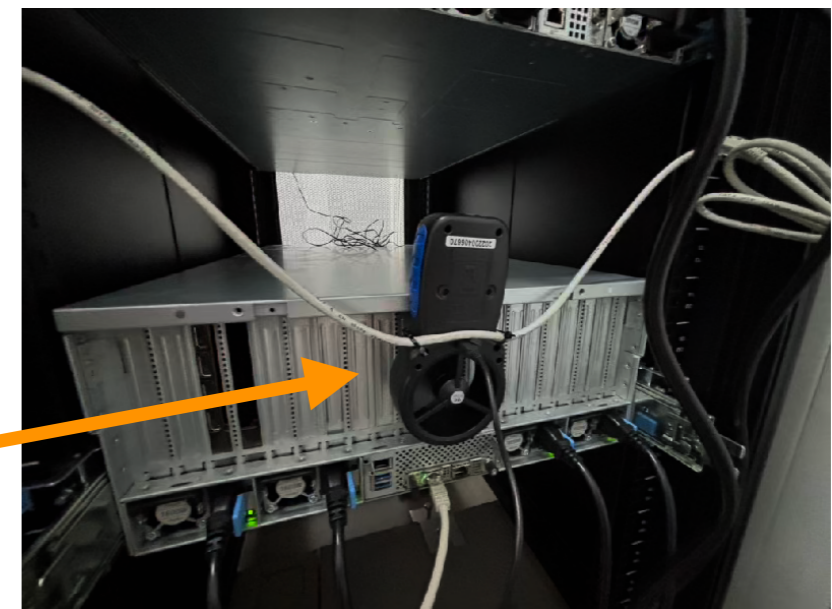
Power consumption studies

- A drastic rise in world data centres power consumption is expected in future

- Significant share of IT equipment

ENERGY FORECAST
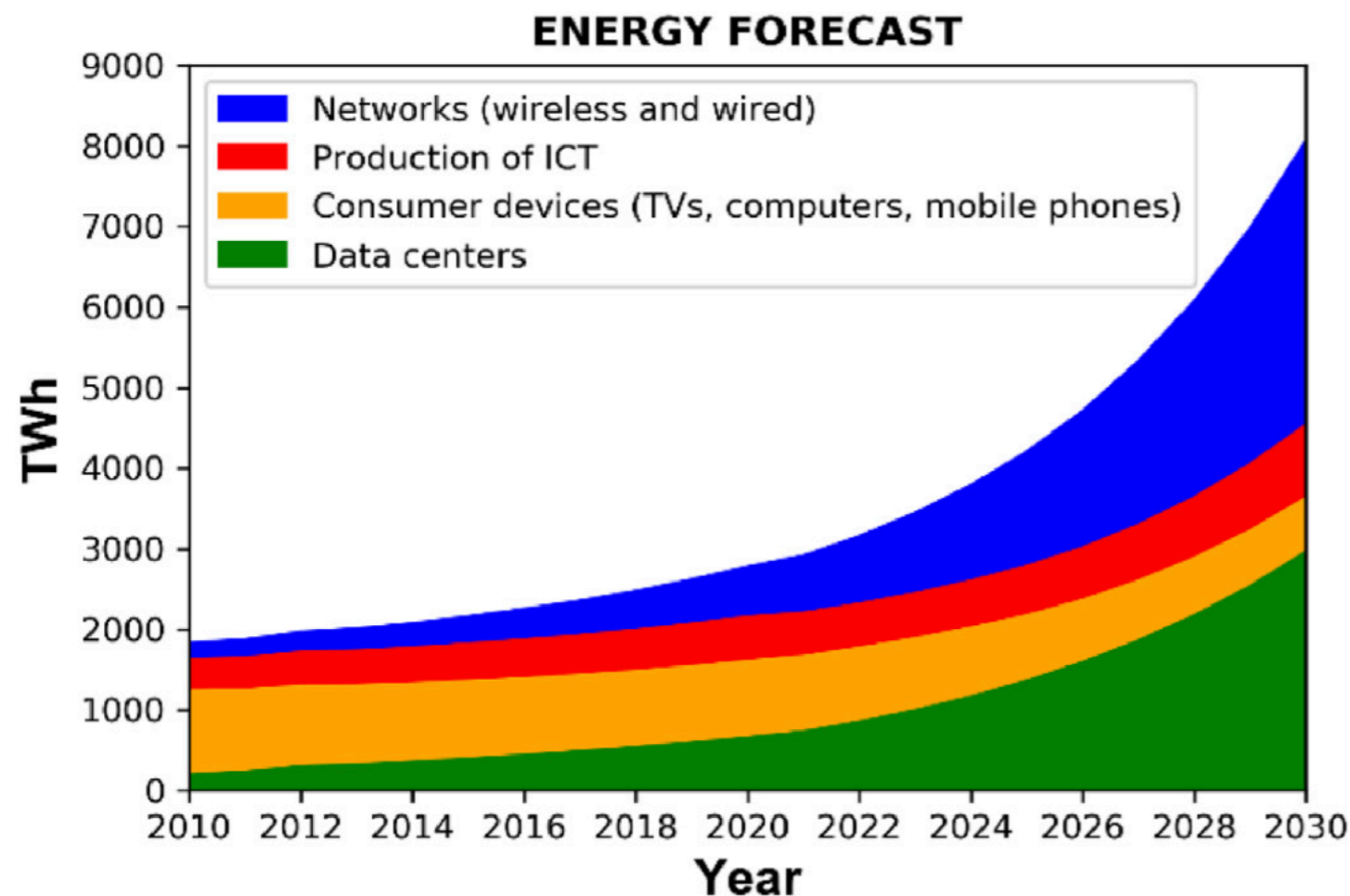
| Component	Approx. Percentage of server power consumption
IT equipment	50% - 60%
Cooling system (HVAC)	35% - 45%
Lighting	2% - 3%
Backup generators and power supply equipment	1% - 2%
Miscellaneous (security systems, monitoring tools)	1% - 2%

Target software

- Studies of software impact on power consumption

- *Test application: **Allen**, used for High-Level Trigger 1 at LHCb*
  - *Can be run on **various architectures**: CPU, GPU*
  - ***Modular design** allows various execution sequences*
  - *Total of approx. 250 algorithms used in data-taking*
  - *Allen is running on ~500 Nvidia A5000 GPUs*
  - *Tests are done with **LHCb** simulation **samples***

Open-source

[Allen, Com. Softw Big Sci 4, 7 ('20)]

[https://gitlab.cern.ch/lhcb/Allen]
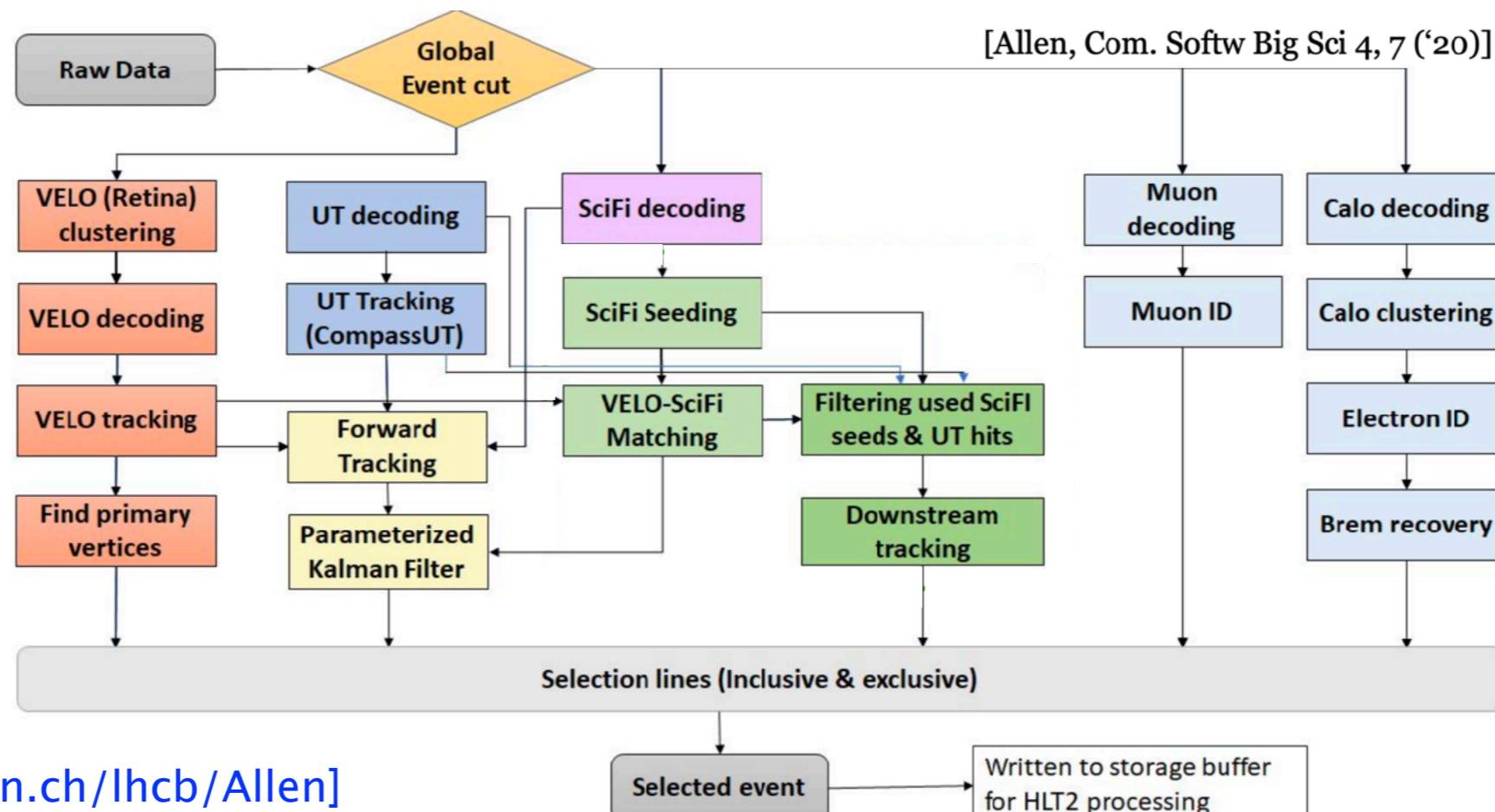
Strategy of studies

- Power consumption may be studied in different ways:
  - With dedicated hardware, a metered power distribution unit (PDU)
  - Relying on device drivers (*Nvidia DCGM*)
  - Reading of CPU performance counters

- The goal of the current project: can we reduce the power consumption by optimising the software and/or hardware?

Motherboard

Metered power distribution unit (PDU) ✓

Power Supplies ✓

FANS ✓

GPU 0 ✓

GPU 1 ✓

CPU ✓

RAM ✓

DISK 0 ✓

DISK 1 ✓

...
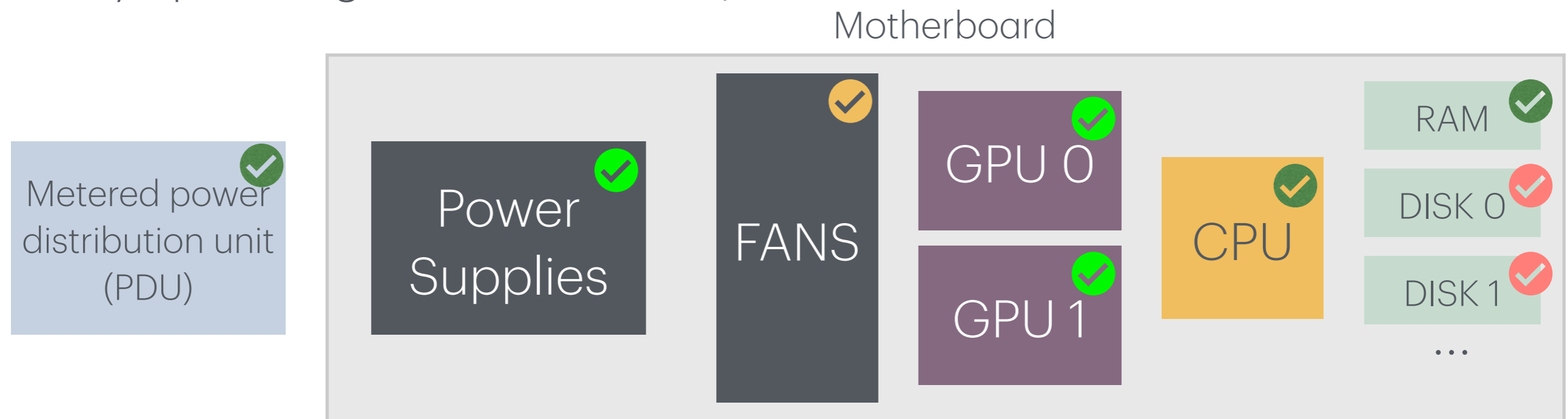
✓✓ Directly measureble

✓✓ Only indirect measurements

Power consumption breakdown

320M events

CPU + NVIDIA RTX 6000 Ada

Rise and fall?

Plateau

Allen execution

PDU
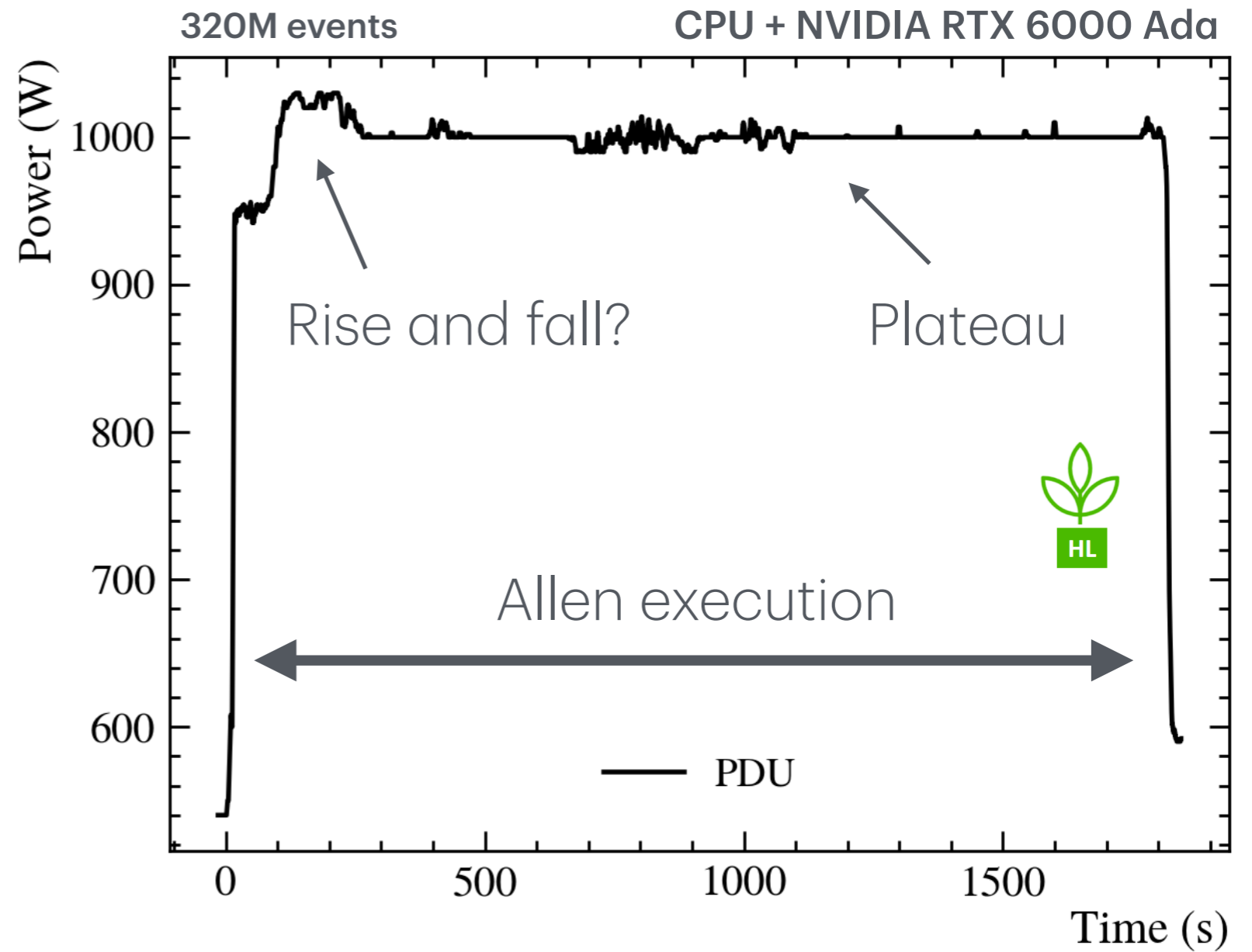
Metered power distribution unit (PDU)

APC PDU ZeroU 2G APS

9

Power consumption breakdown

Advanced Configuration and Power Interface (ACPI)

It's now safe to turn off your computer.

Hardware abstraction interfaces between the hardware and the OS

320M events

CPU + NVIDIA RTX 6000 Ada

ACPI is almost identical to PDU!

PDU

---- ACPI

Power (W) vs Time (s)
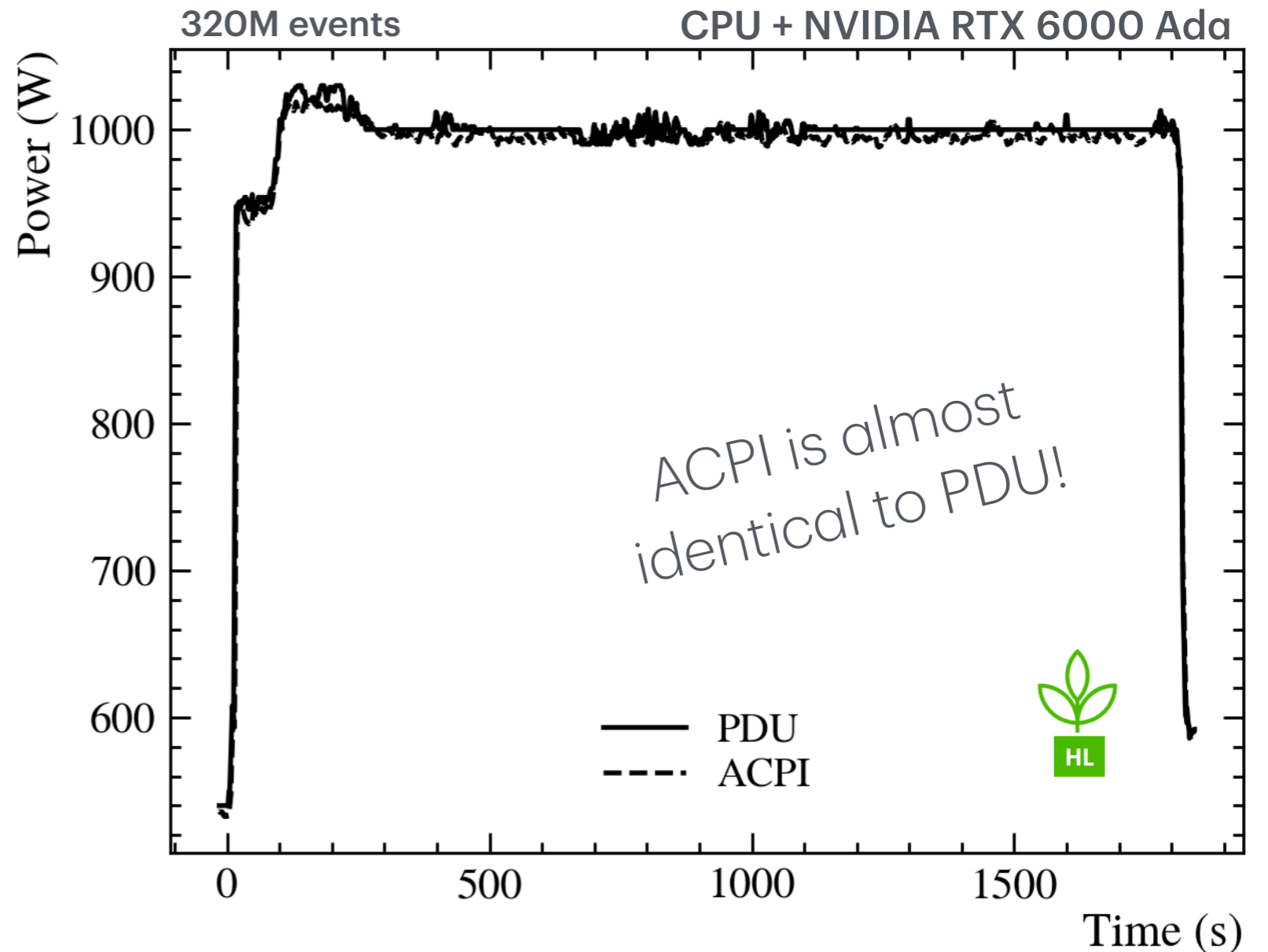
Metered power distribution unit (PDU)

Power Converters

CRPS power supply

Power consumption breakdown

Running average power limit (RAPL) based on CPU performance counters
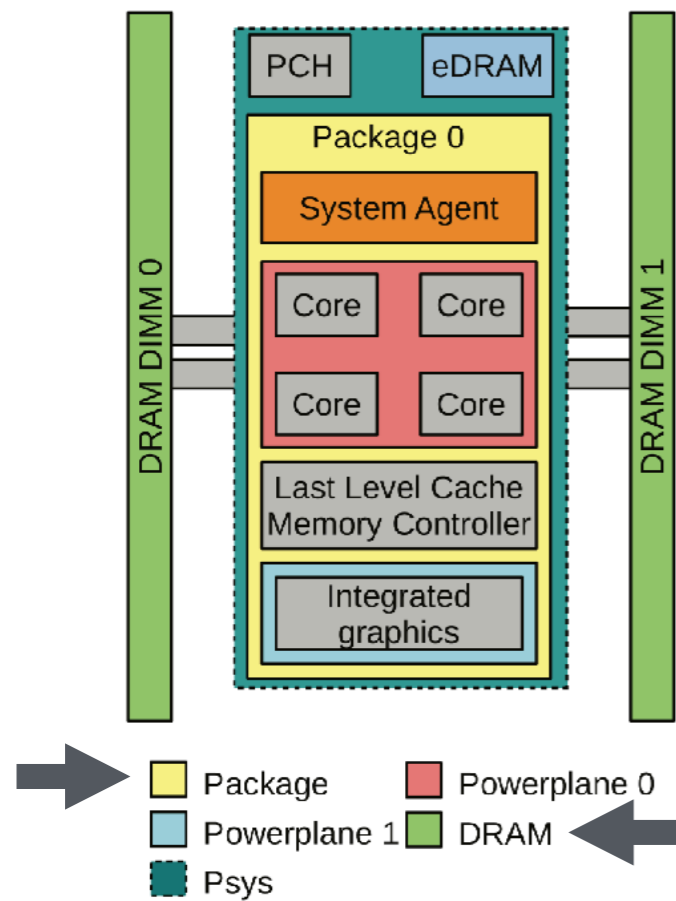
Metered power distribution unit (PDU)

320M events

CPU + NVIDIA RTX 6000 Ada

RAM consumption is negligible

Allen is executed only on CPU 0

11
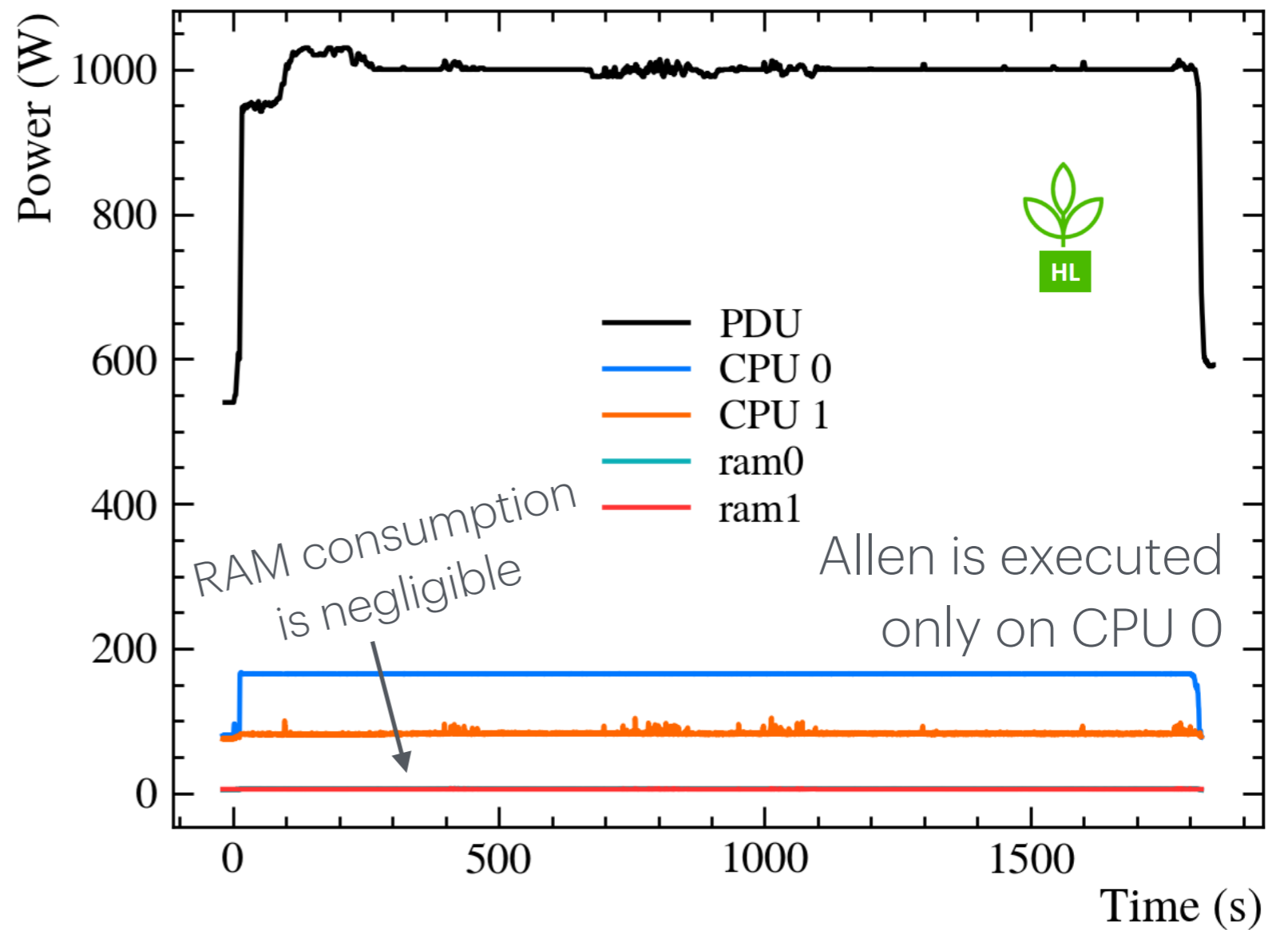
Power consumption breakdown

NVIDIA Data Center GPU Manager (DCGM) for power consumption measurements
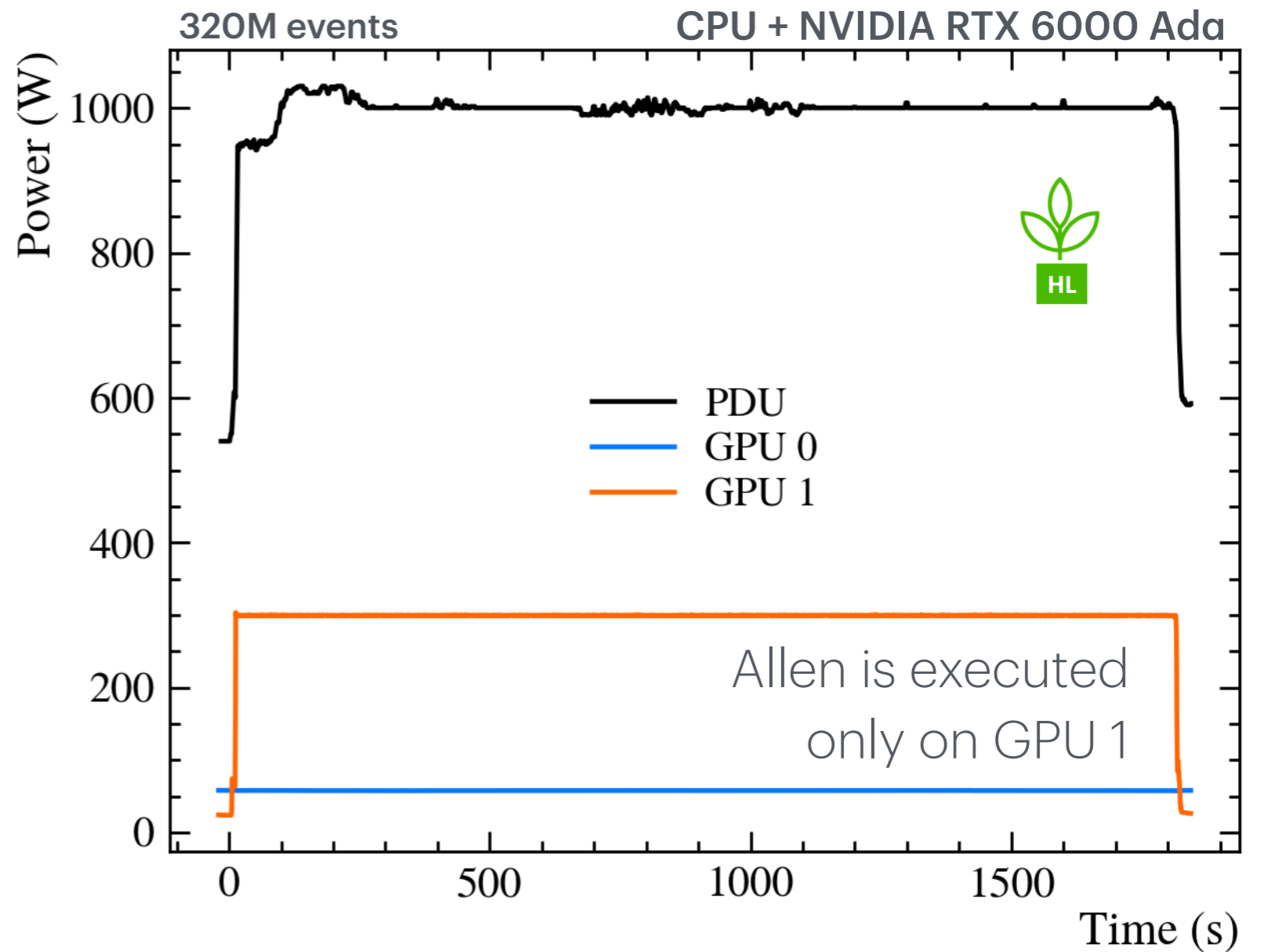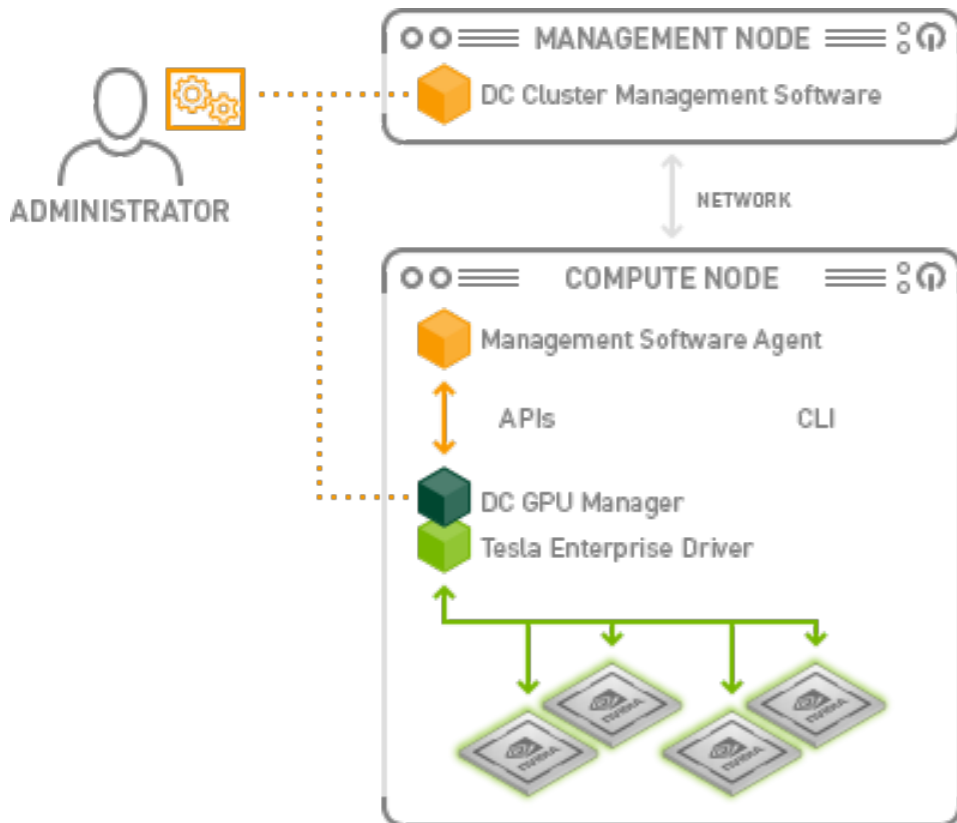
Metered power distribution unit (PDU)

320M events

CPU + NVIDIA RTX 6000 Ada

PDU
GPU 0
GPU 1

Allen is executed only on GPU 1

GPU 0

GPU 1

12
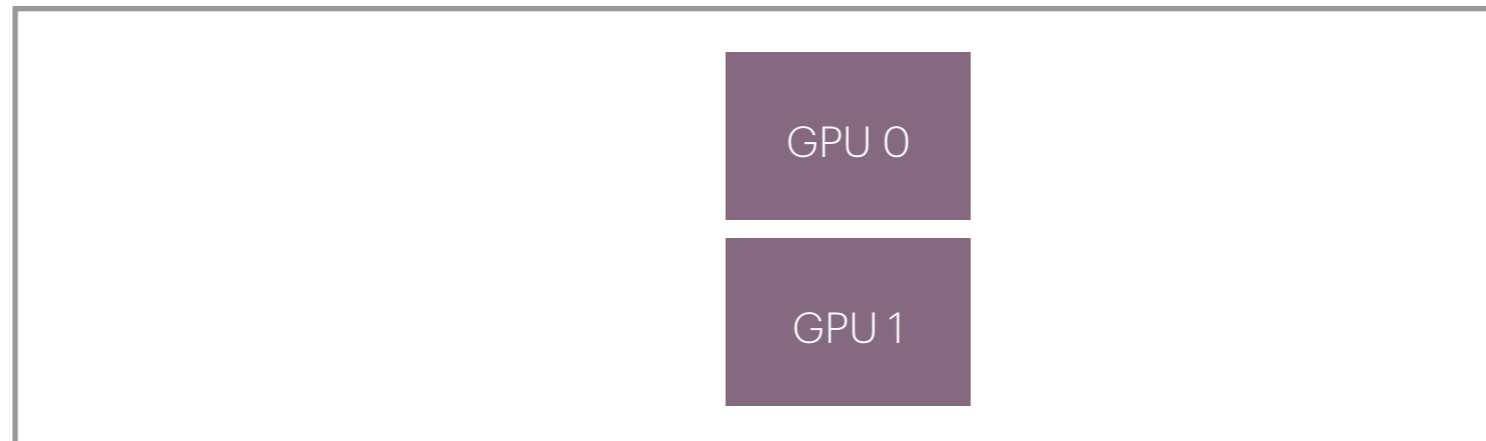
Power consumption breakdown

CPU + NVIDIA RTX 6000 Ada   320M events
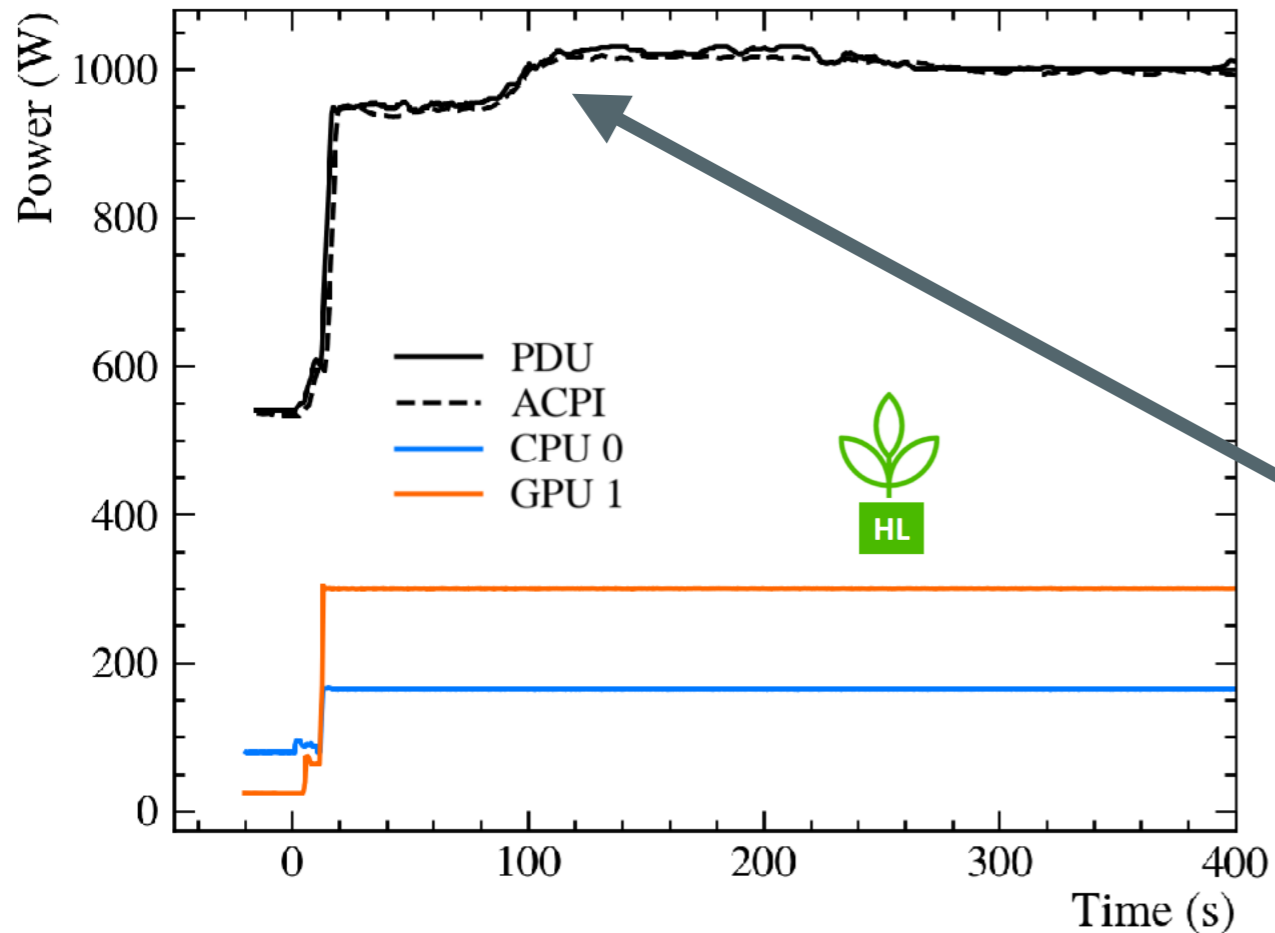
Main components have constant power consumption

What is the cause of the power consumption rise?
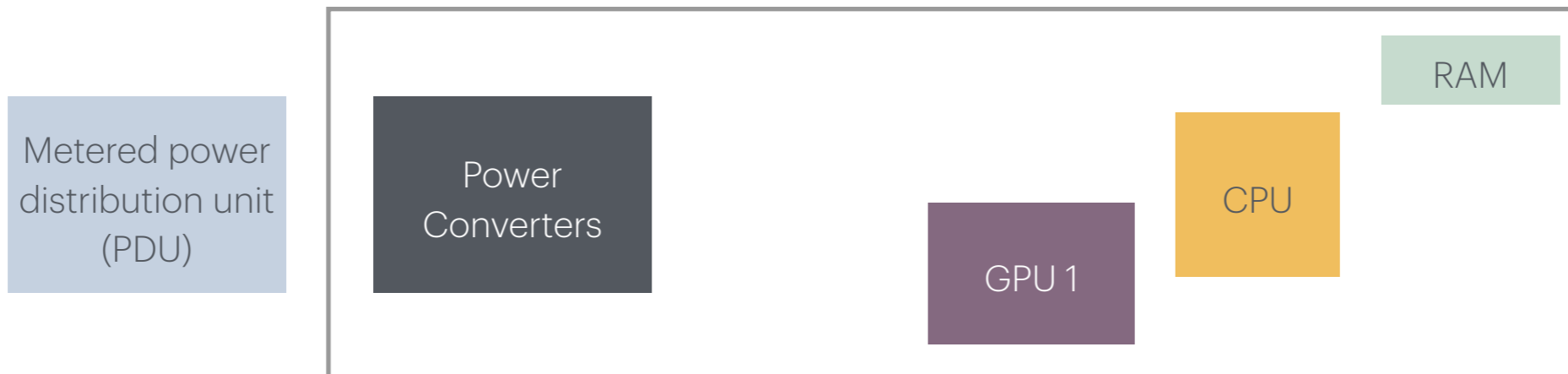
Power consumption breakdown

CPU + NVIDIA RTX 6000 Ada   320M events

Fan speed measurement device

Metered power distribution unit (PDU)

Power Converters

?

GPU 1

CPU

RAM

Power consumption breakdown

Fans spooling up

Plateau when fans stabilise

320M events

Fans stabilise

320M events
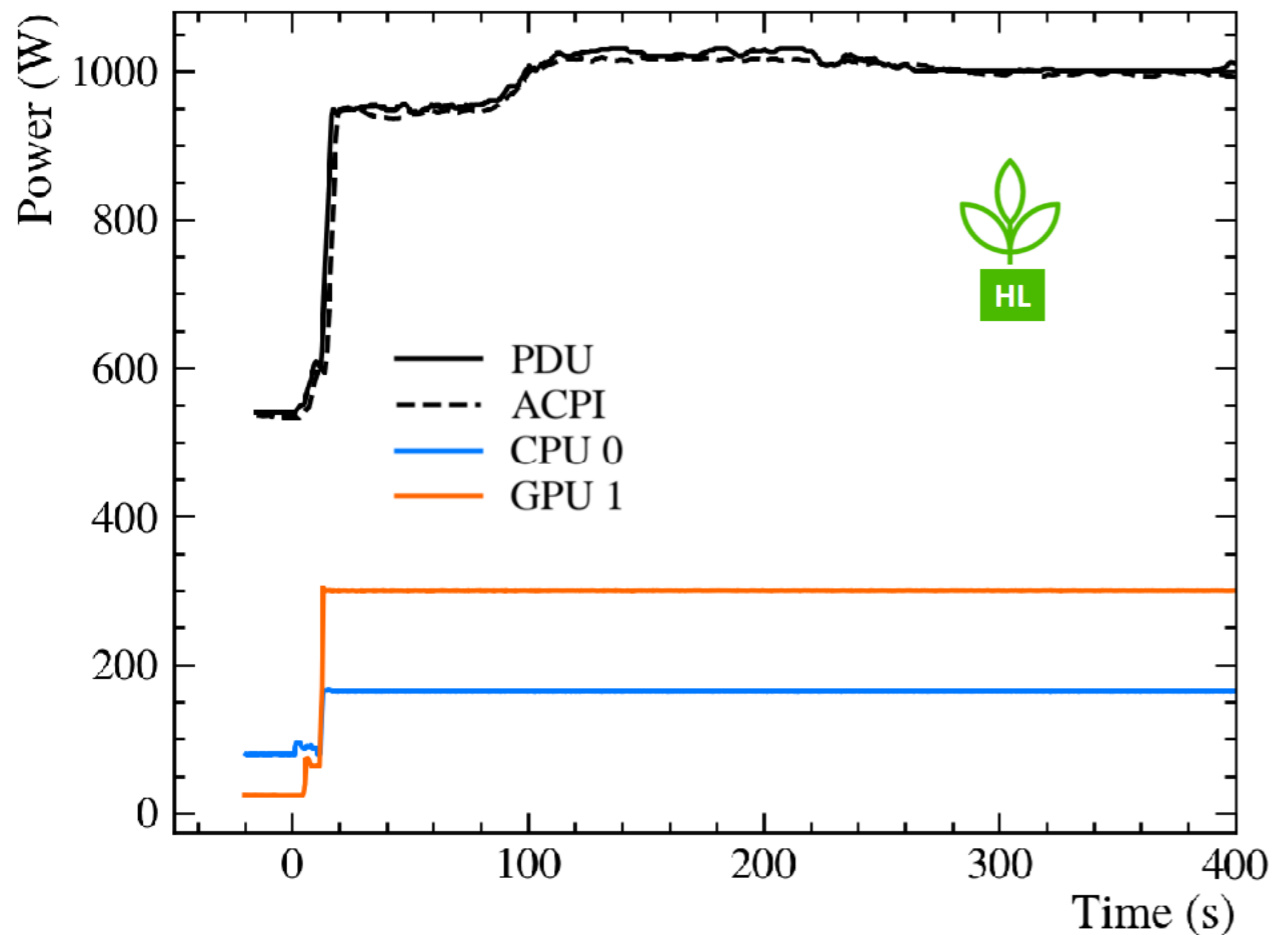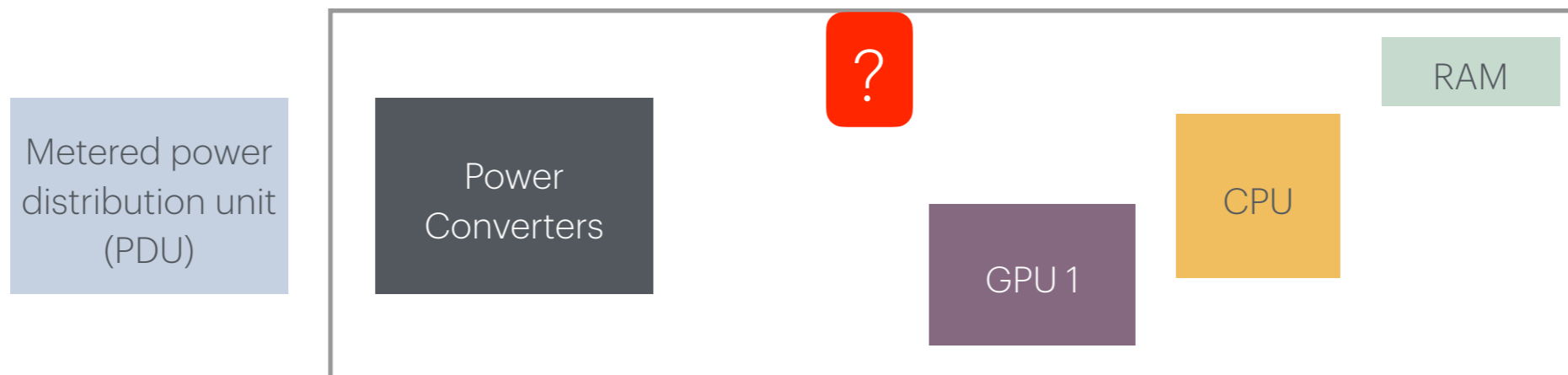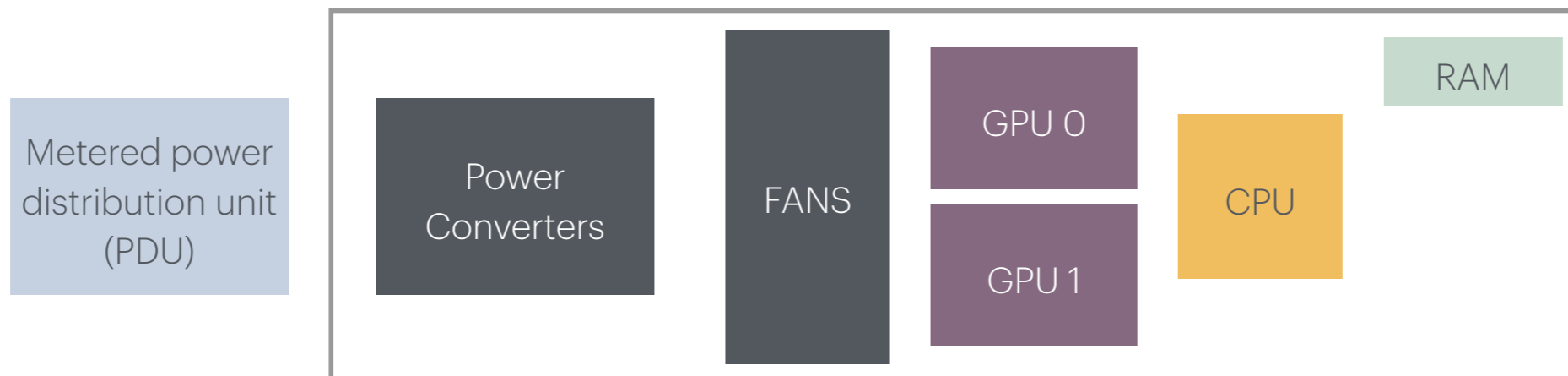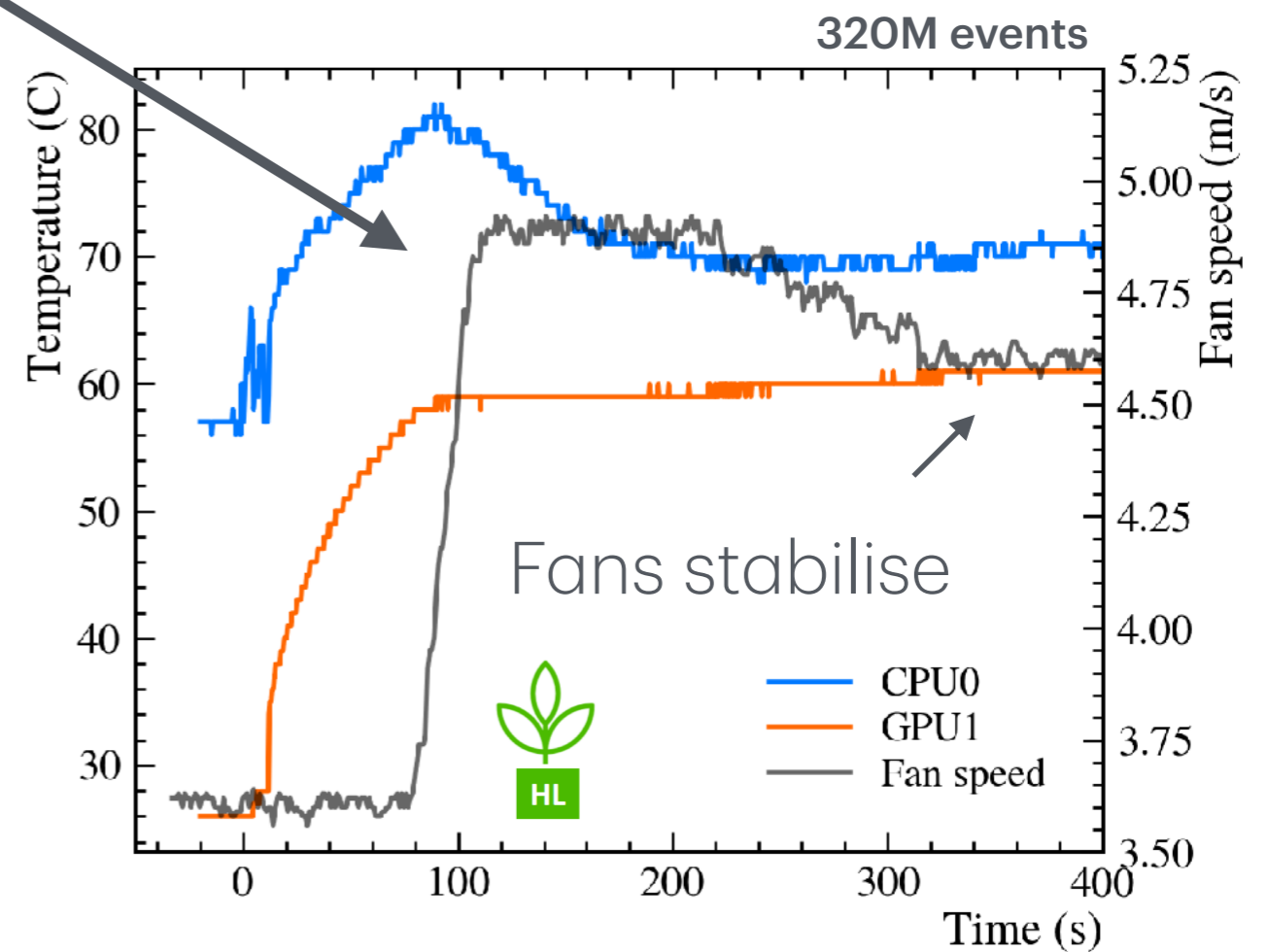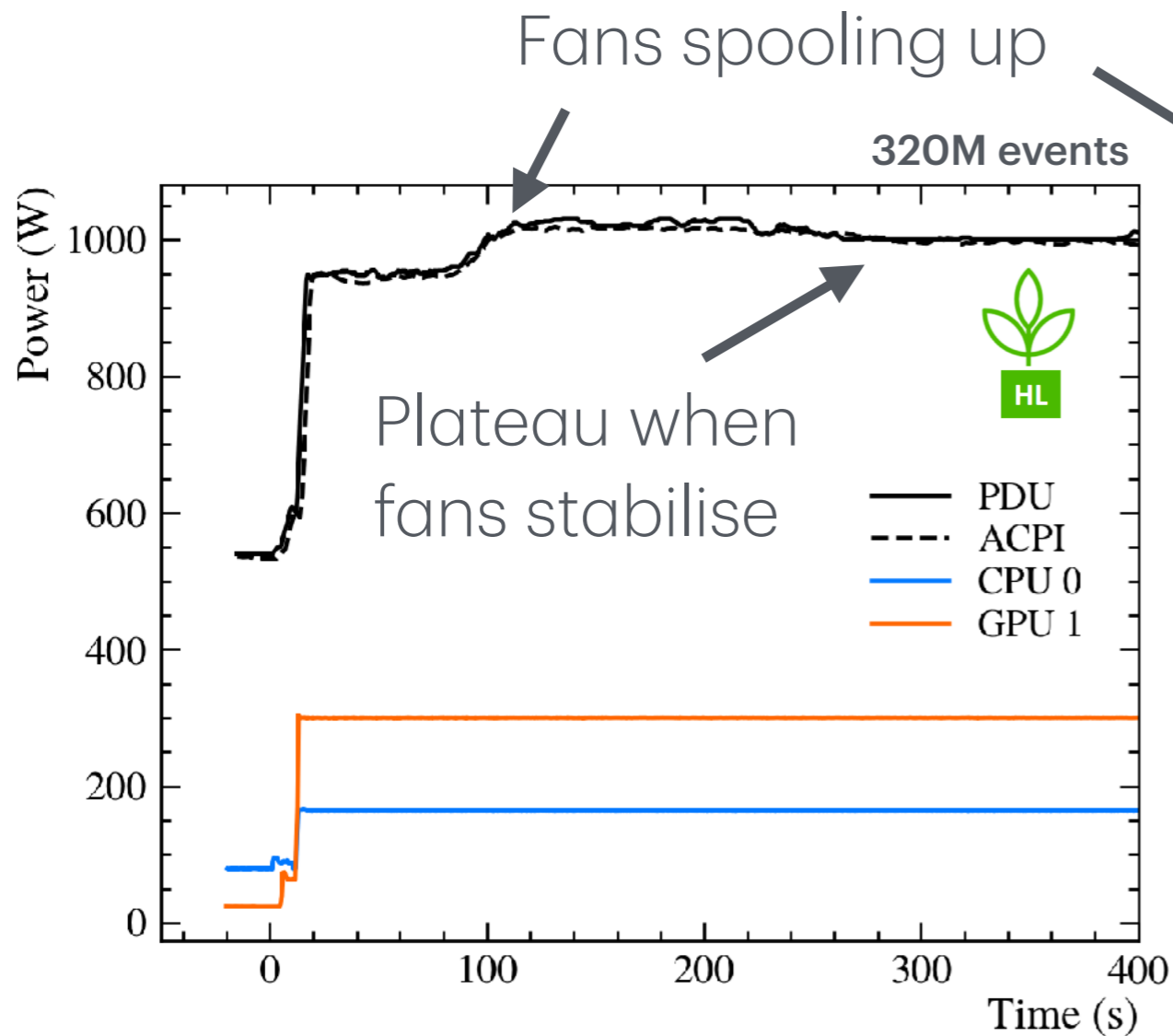
Hardware choice

- Target hardware may have an effect not only on execution speed but consumption

- CPU target and two different GPUs were tested:

  - *Intel(R) Xeon(R) Gold 5318Y*

  - *NVIDIA RTX A5000*

  - *NVIDIA RTX 6000 Ada*

- In general, faster execution also means less power consumption

40M events   **PDU Power usage**
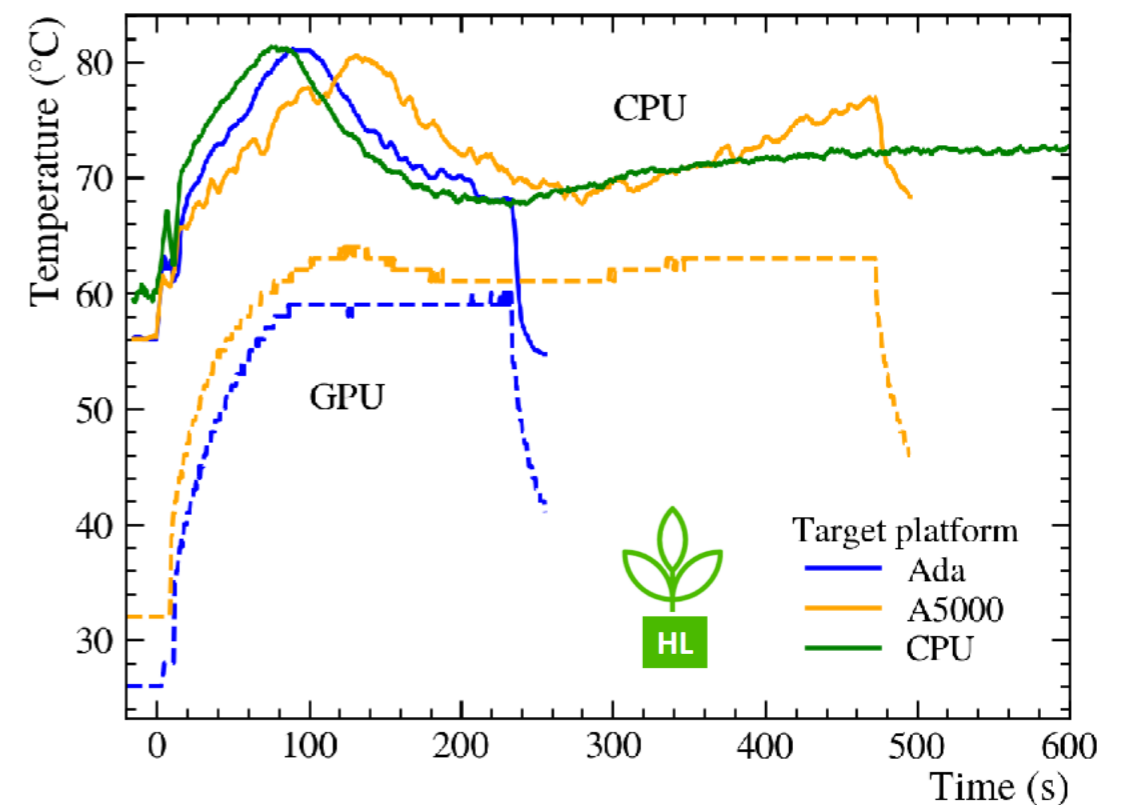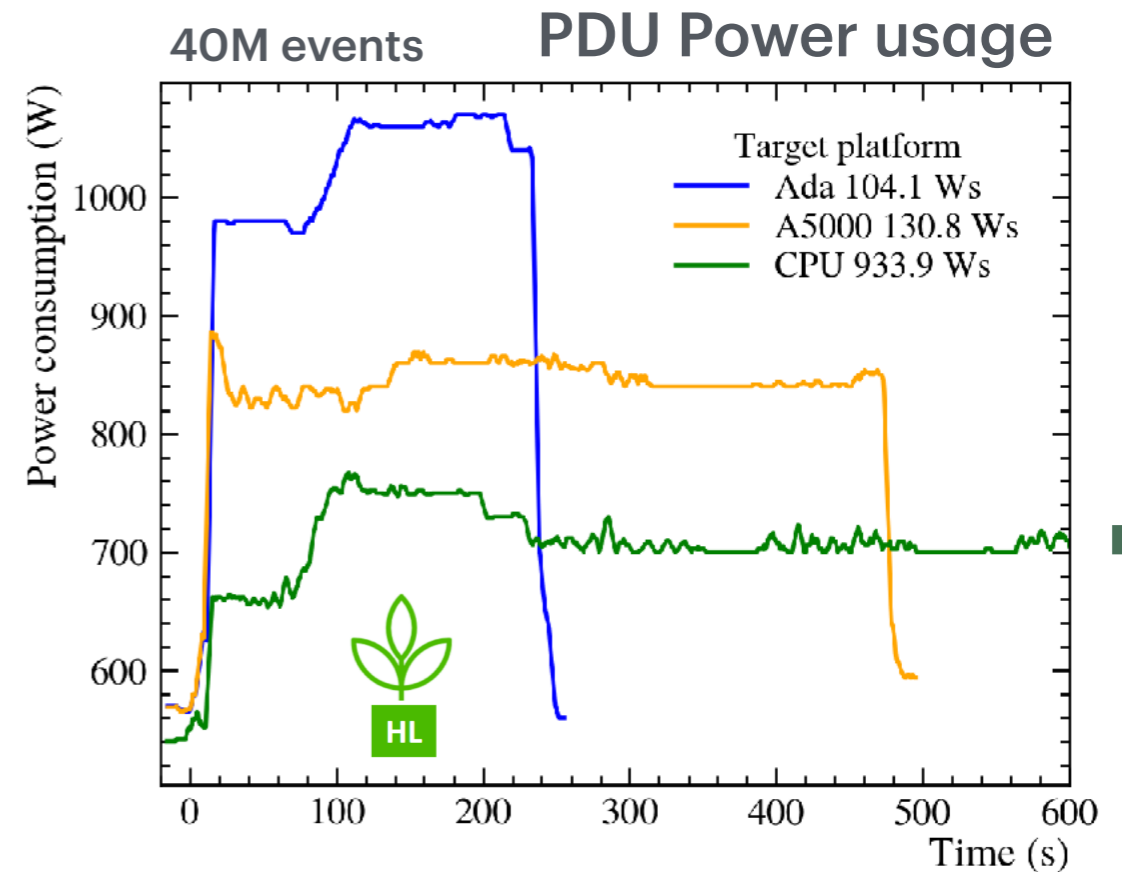
Hardware choice

- The application is optimised for GPU but compiled for CPU target
- Execution on the CPU leads to low instantaneous power consumption but consumes a lot of energy due to a slow process
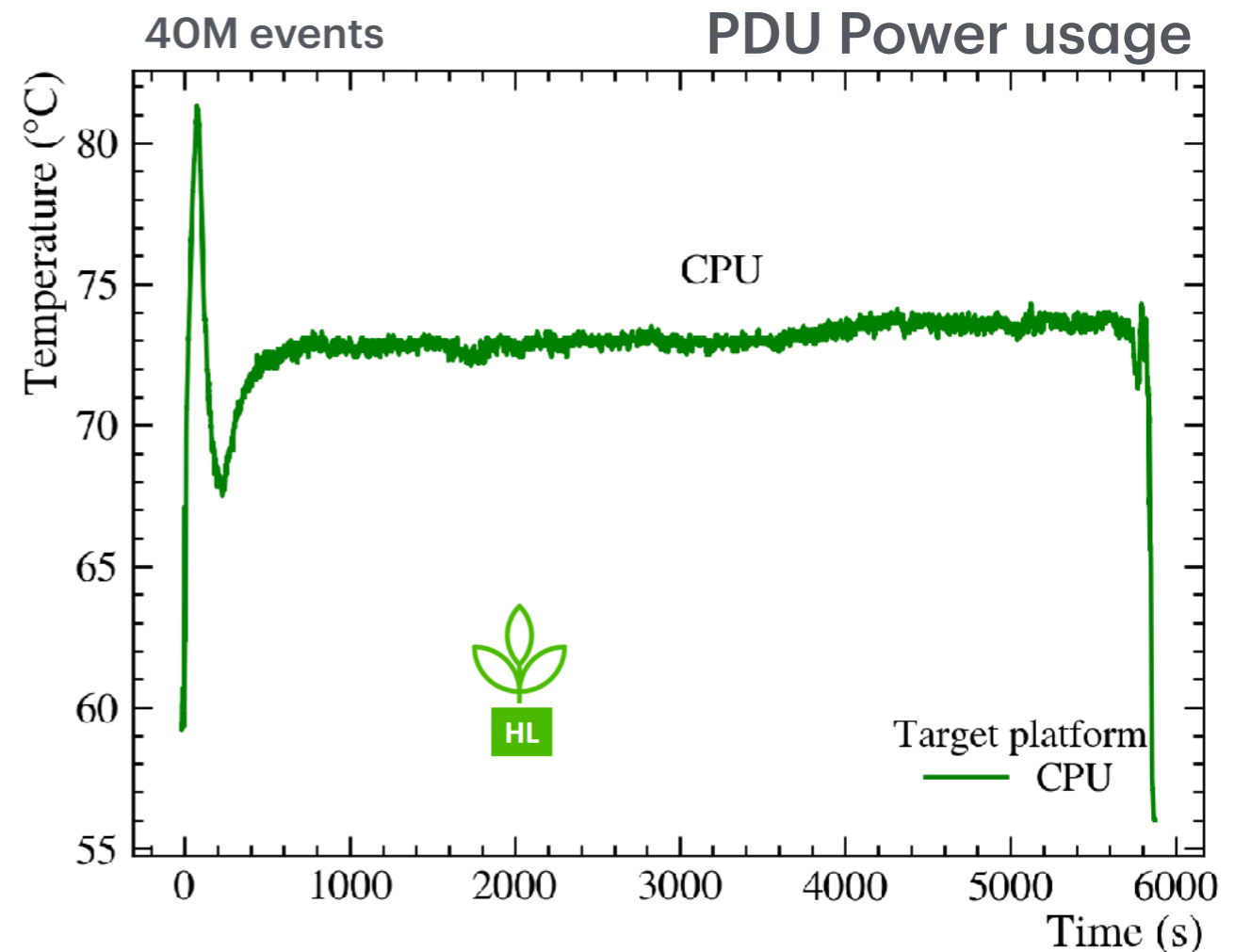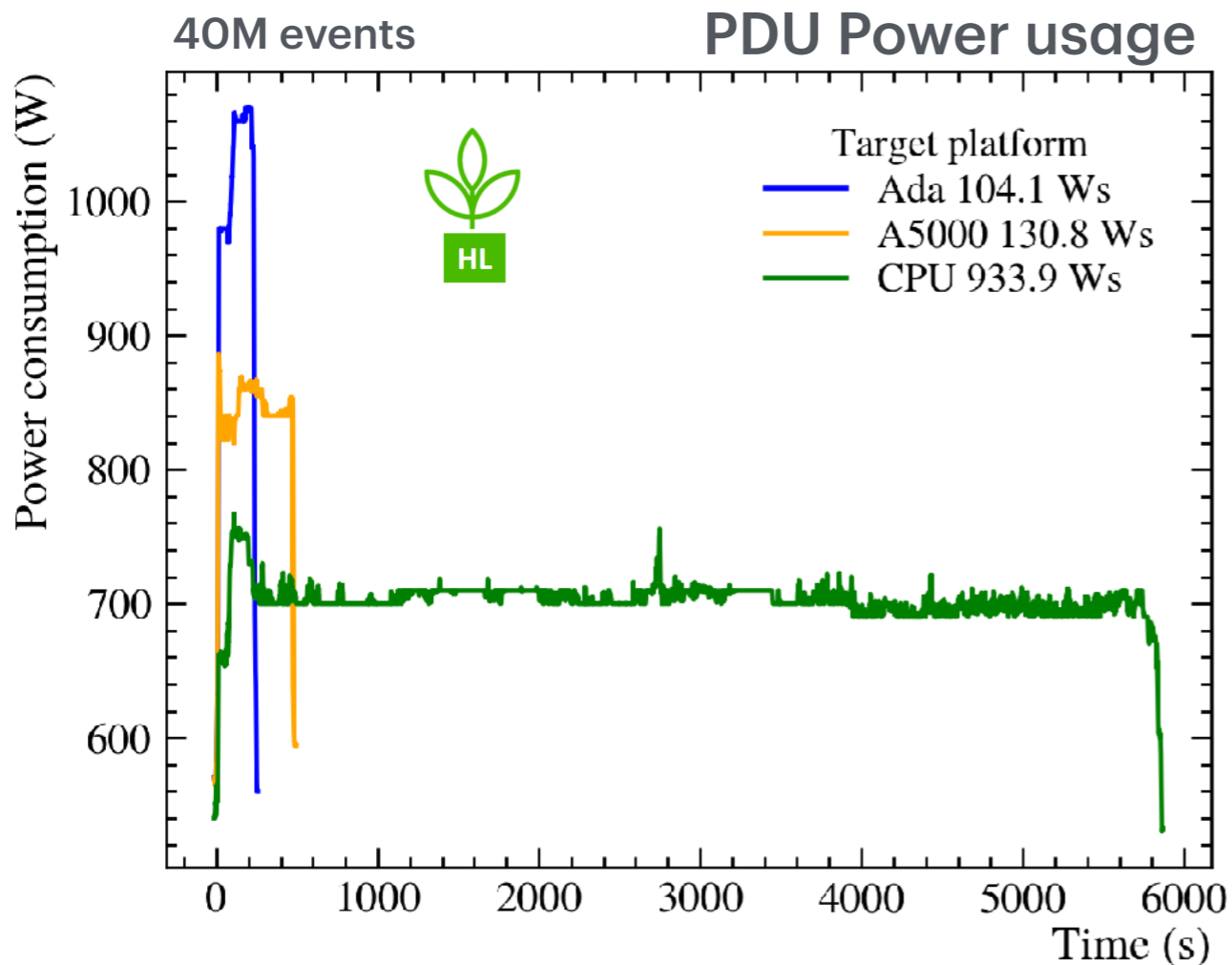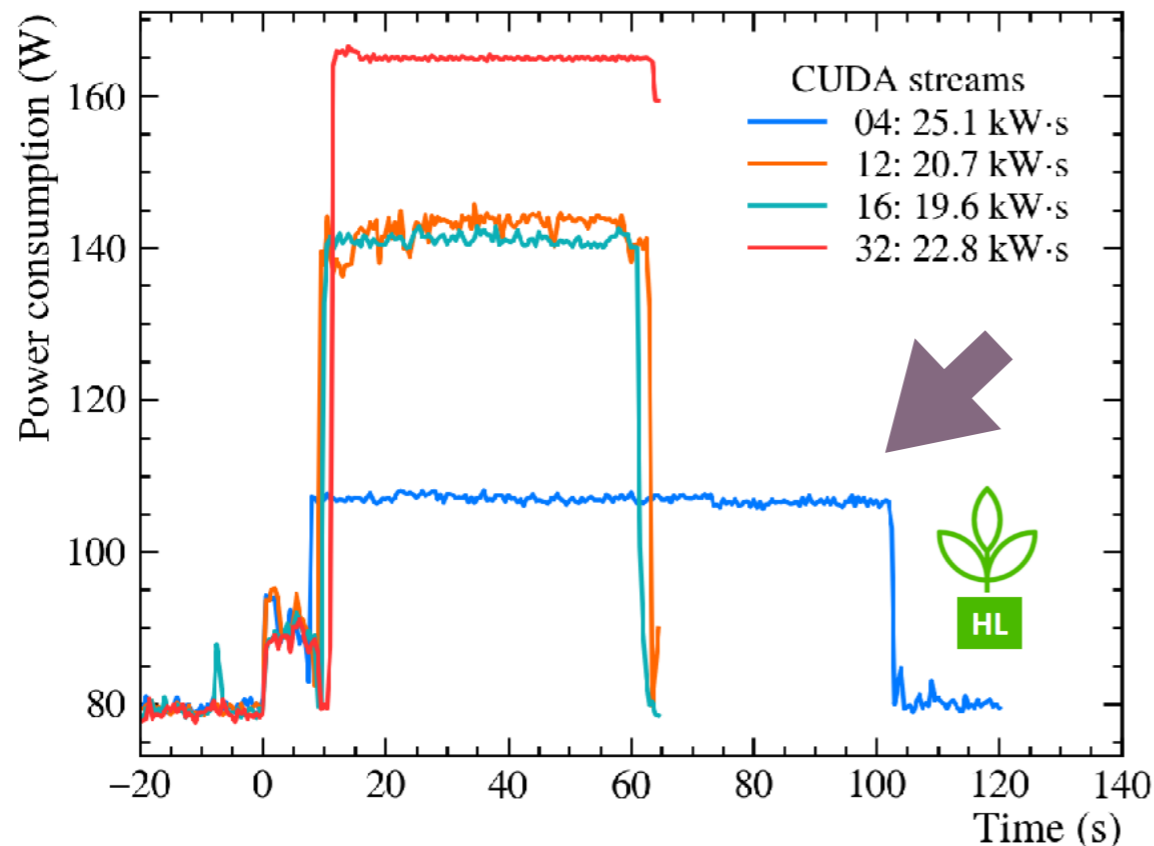
GPU utilisation

- Utilise your accelerators carefully!

  - **Too small utilisation leads to slow execution →
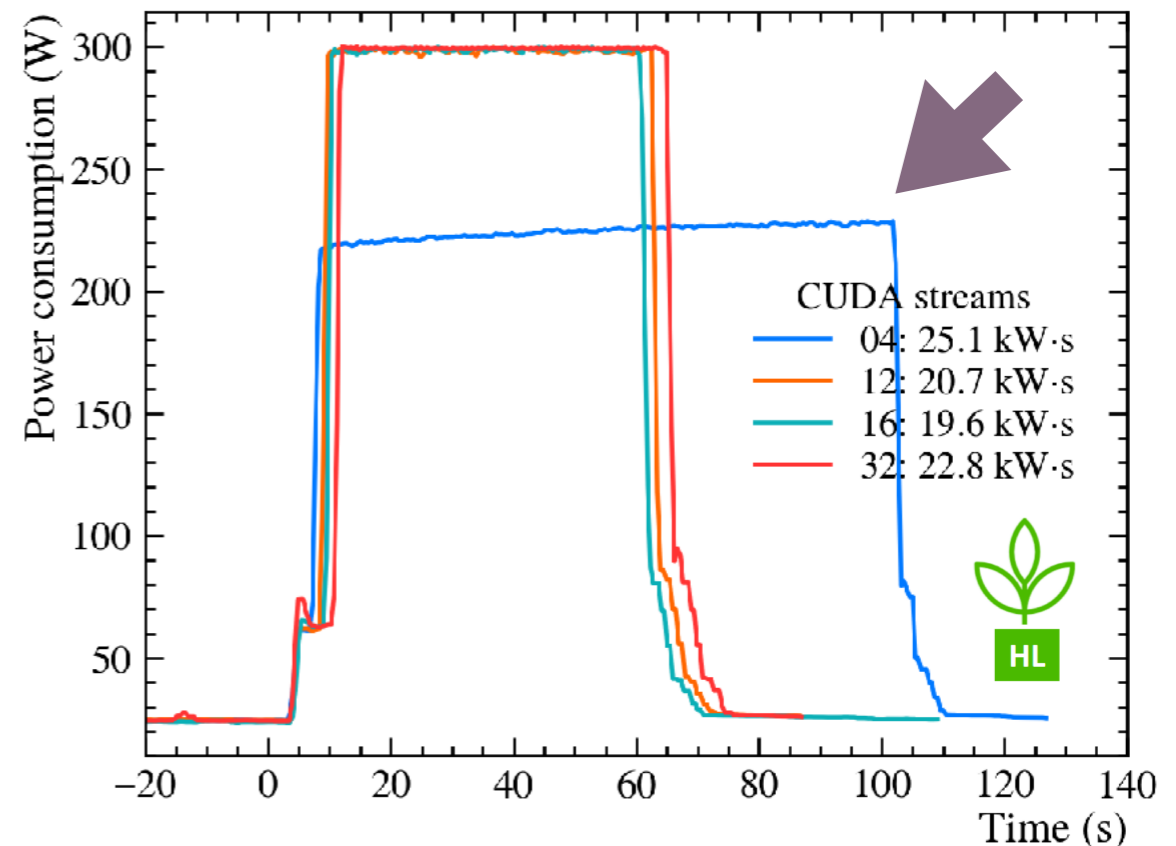   larger energy consumption**
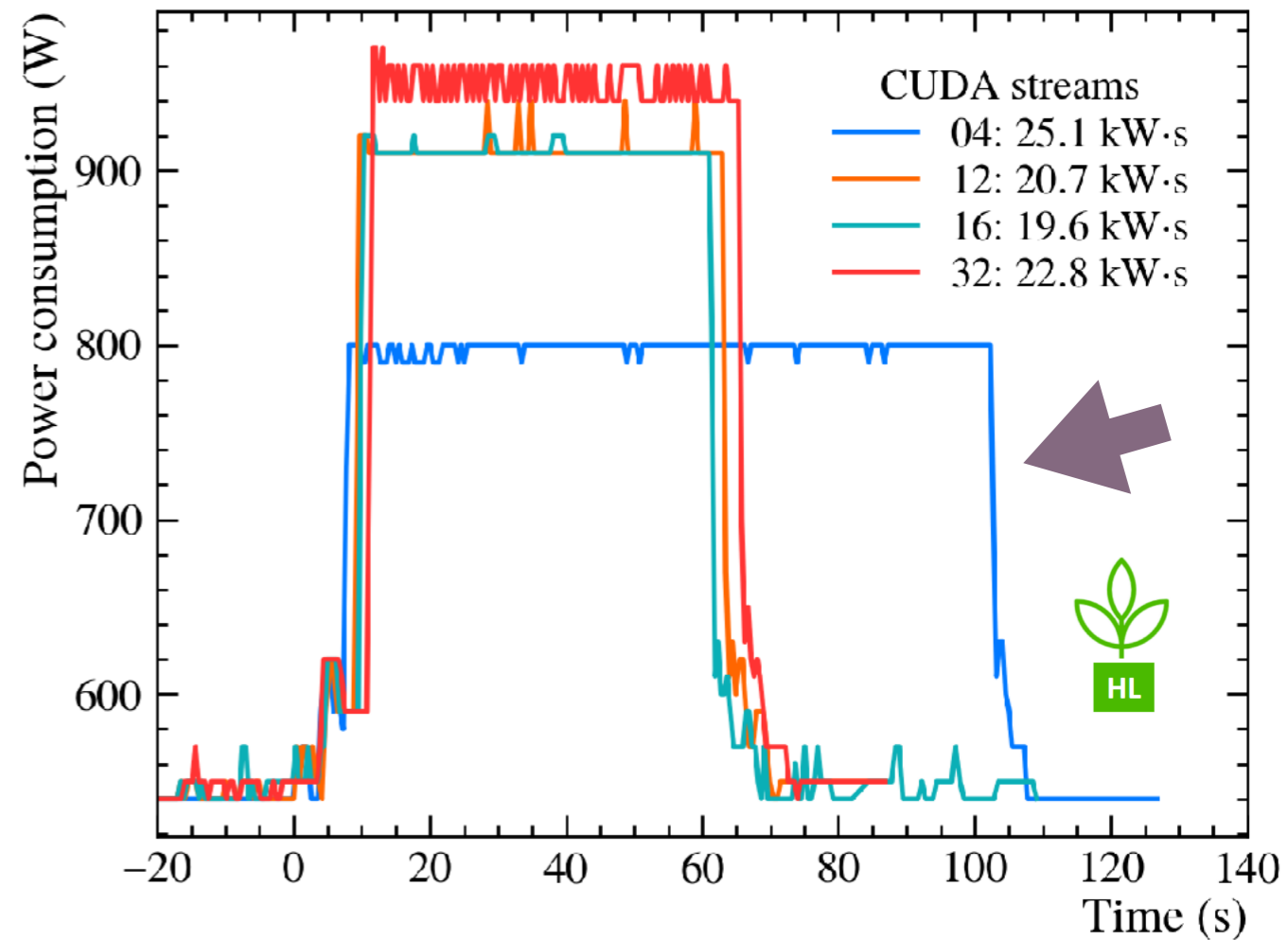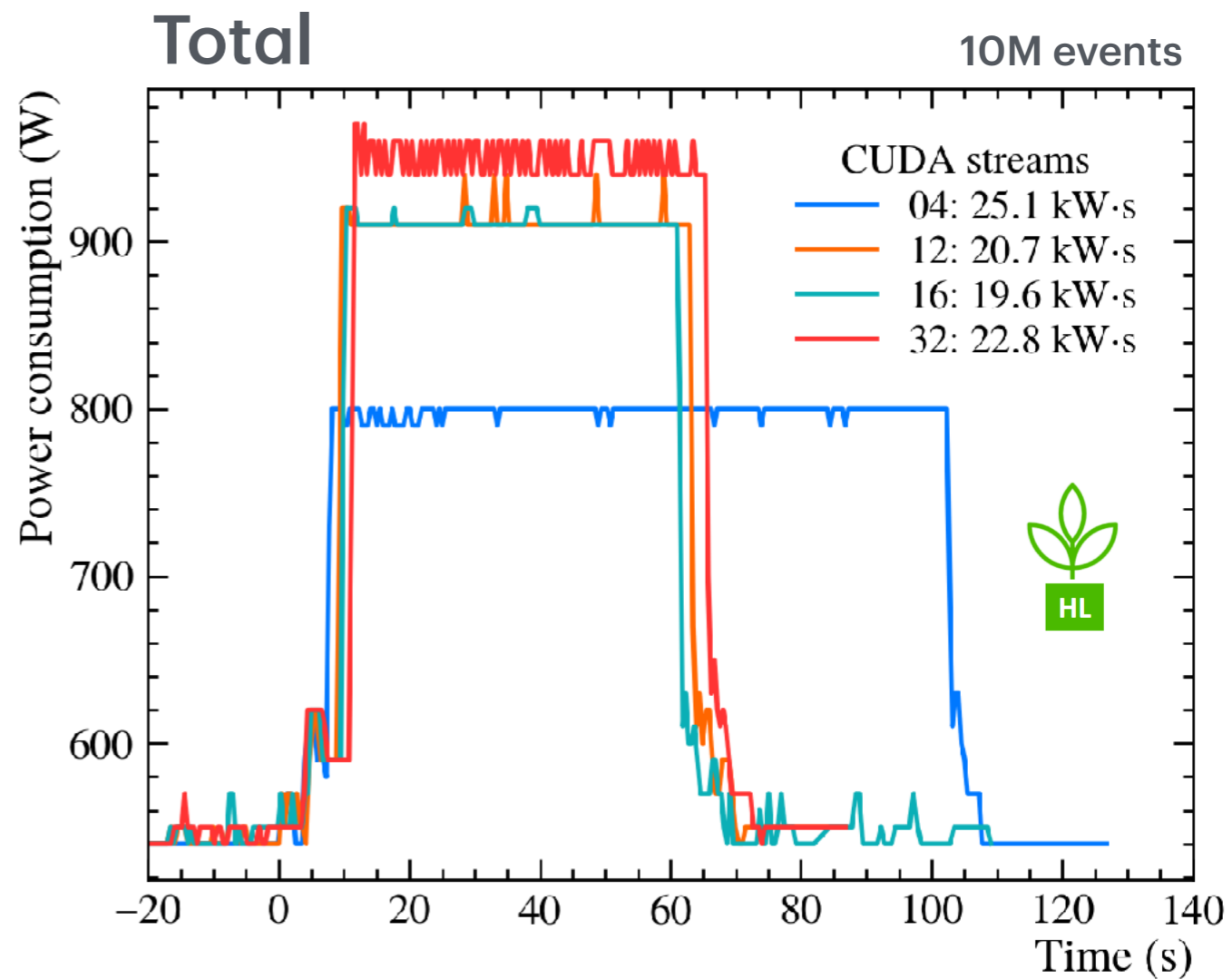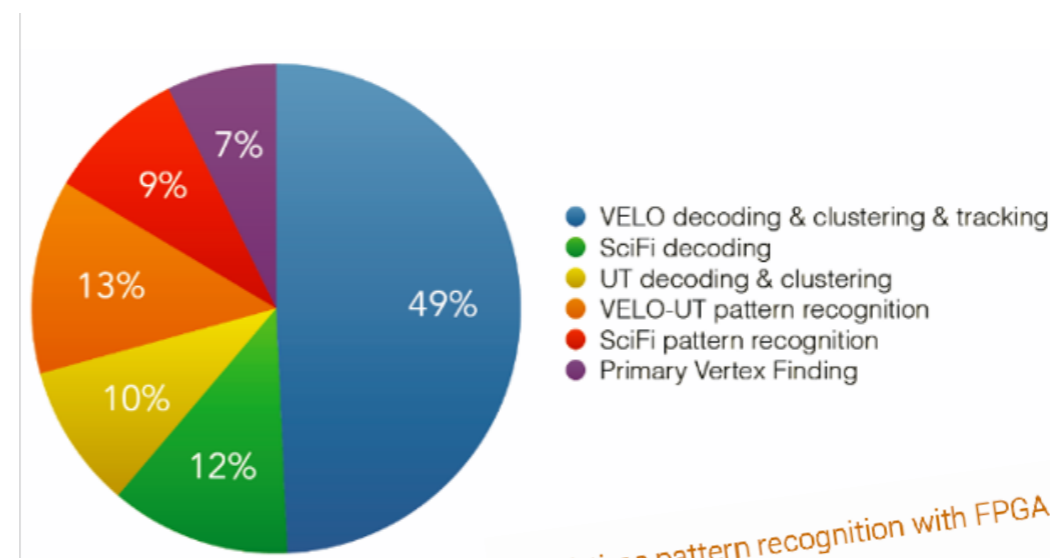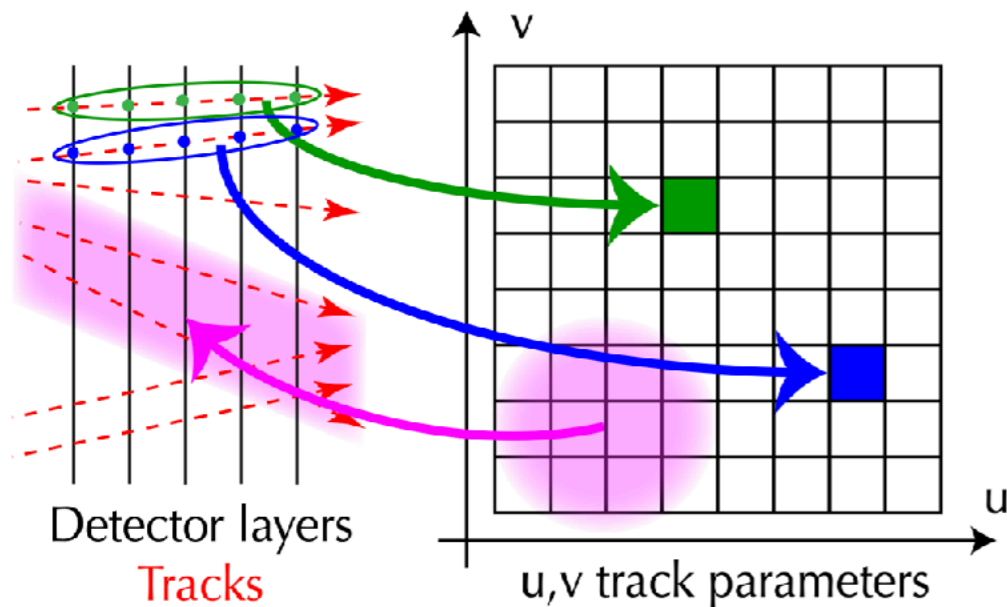
GPU utilisation

- Utilise your accelerators carefully!

  - **Trying to use more resources than GPU provides → increase in CPU consumption**
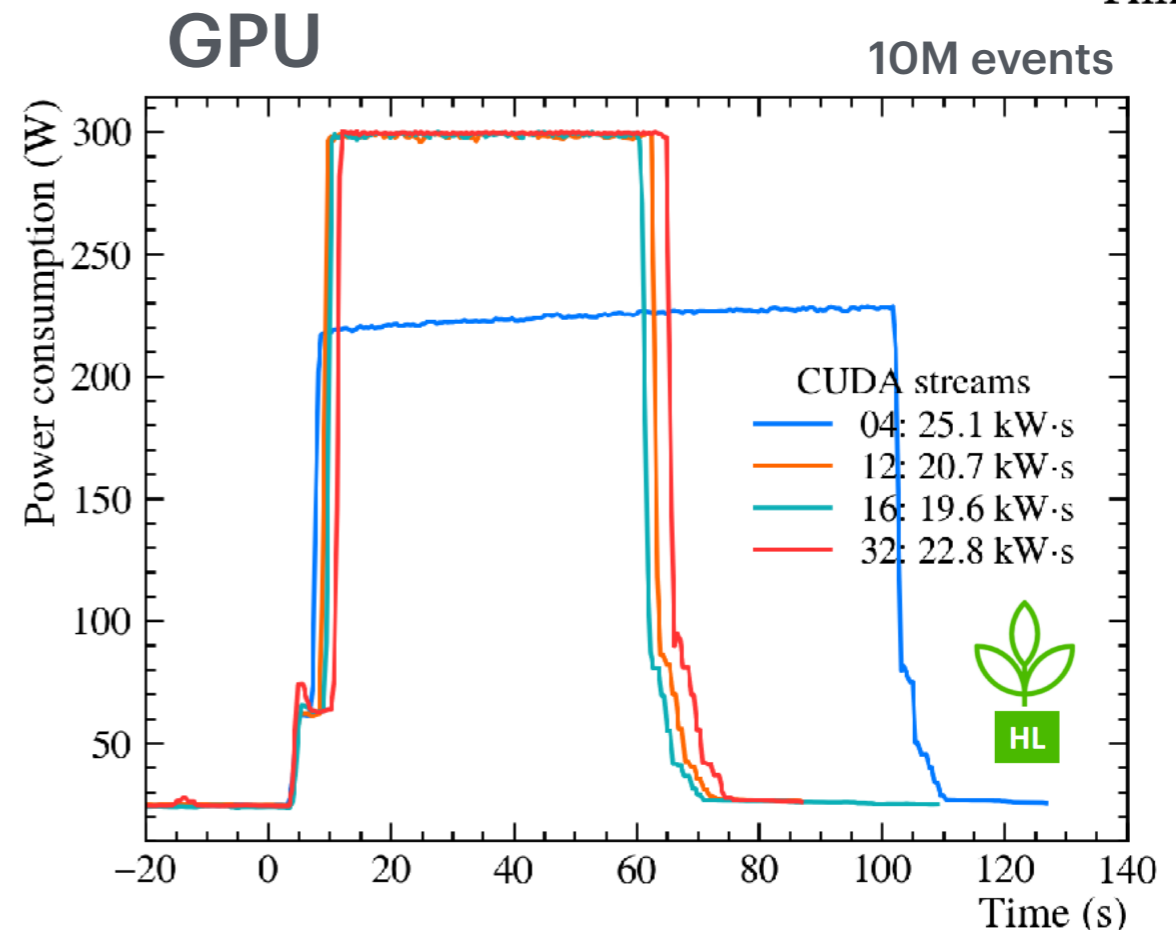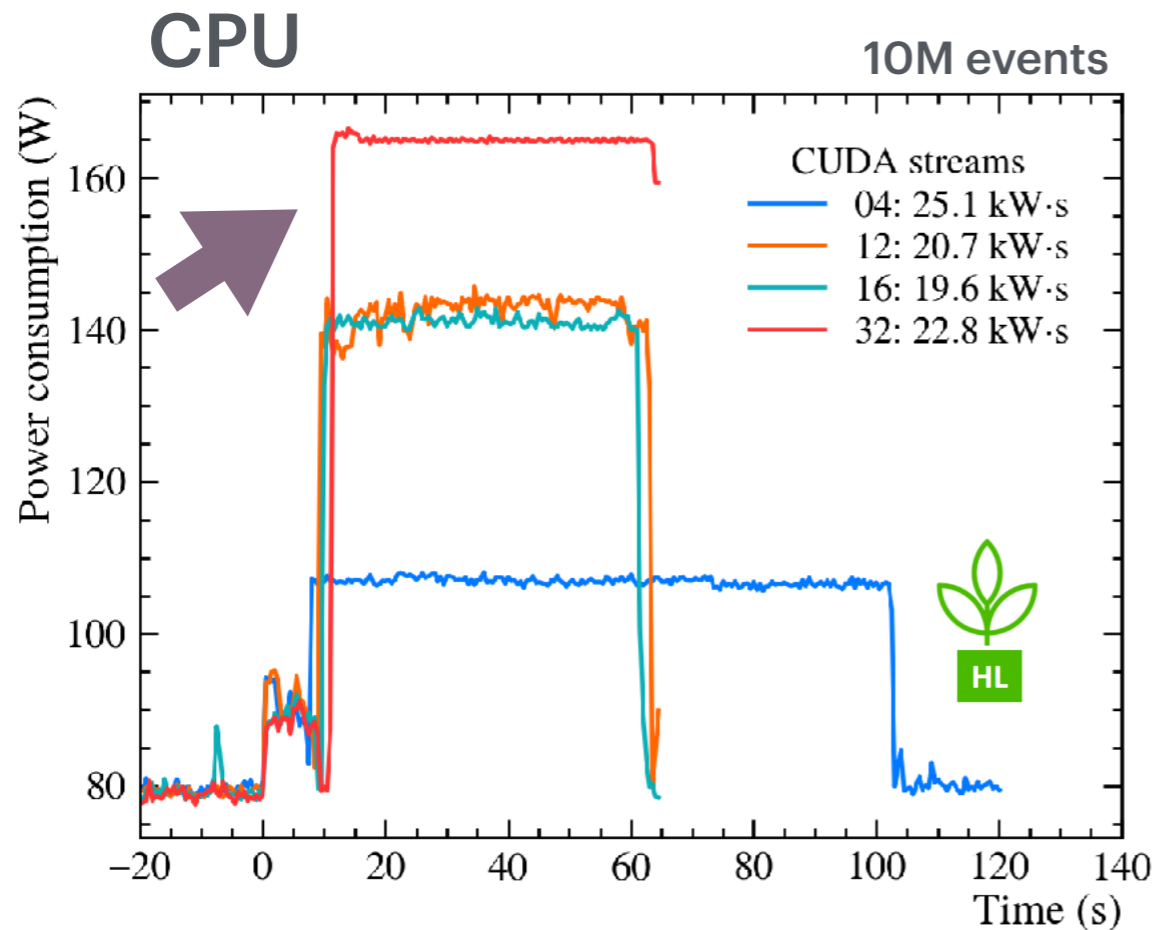
FPGAs for less consumption?

- Offloading of some computation tasks to FPGA

- Real-time reconstruction on FPGAs ("artificial retina")

- VELO clustering is already implemented for Run3 in FPGAs !

- Tracking in development for Run5 (~2030)

FPGAs for less consumption?

In the clustering case, the total reduction in power consumption is $\mathcal{O}(50)$

[FPGA-based real-time cluster finding
for the LHCb silicon pixel detector; CHEP 23]

In the tracking, case studies are still ongoing.

GPU throughput increases by $\approx 30\,\%$
A similar amount is in power consumption saving

Clustering for the first tracker (VELO)

Tracking for the last tracker (SciFi)

GPU power usage only
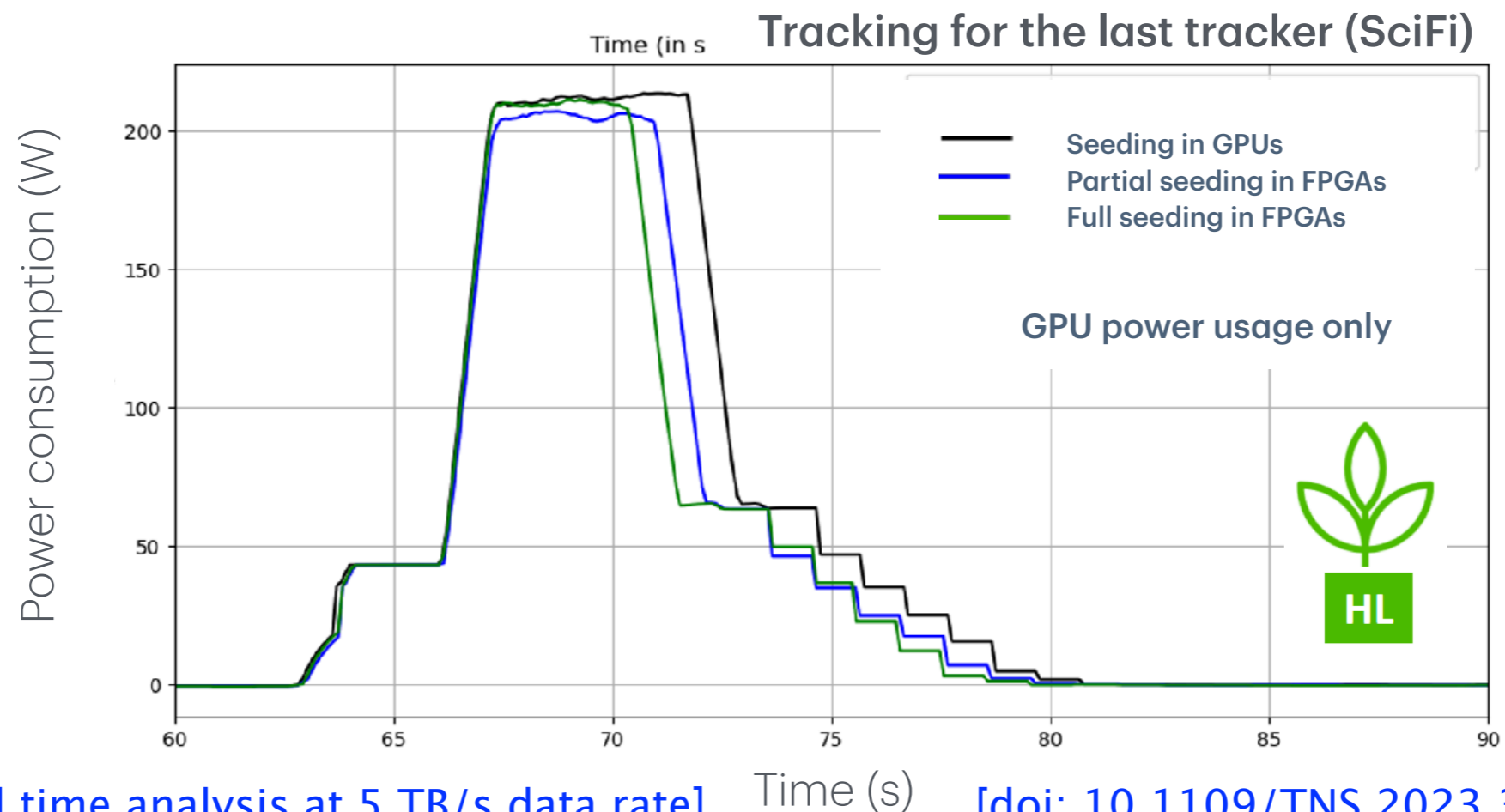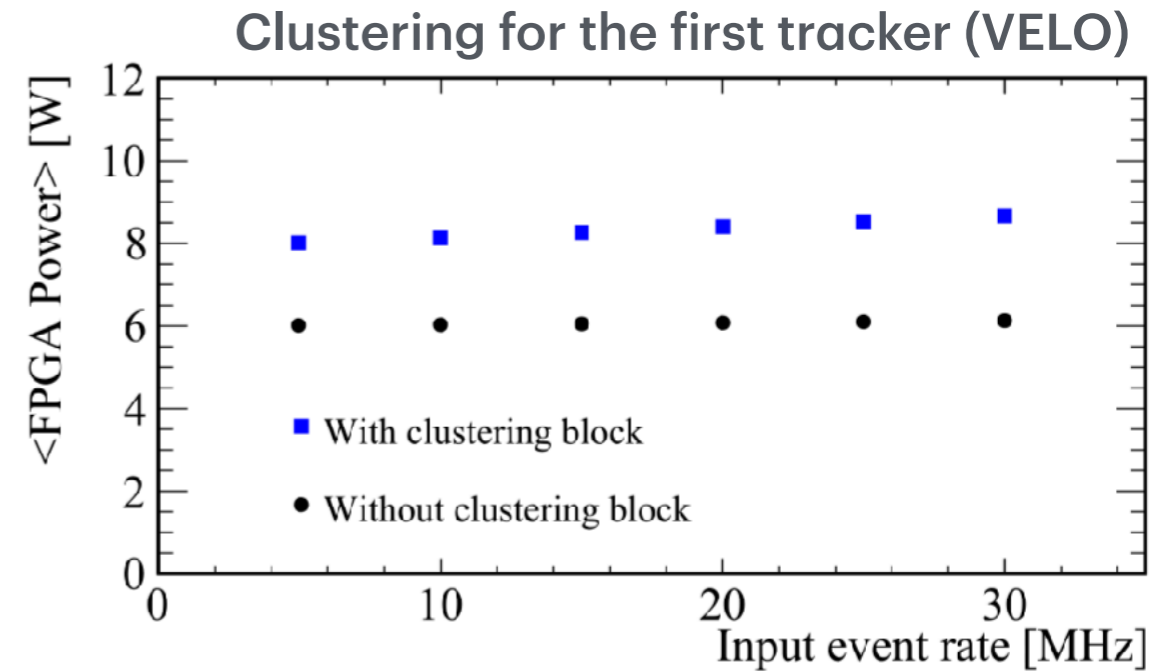
Summary

- Watts per event should be an important metric for HEP

- Software optimisations can help make HEP sustainable

- Heterogeneous computing systems allow power consumption reduction at least in some cases

- **Power consumption monitoring can be done even on a software level, and is crucial during optimisation**

- Planned to test other architectures (ARM; Intel GPUs)

- Work is ongoing, stay tuned! ☺️

Thanks for your attention!