The University of Manchester

Including energy awareness in computational needs of future experiments

Caterina Doglioni — University of Manchester

SHE/HER

Input and discussions by:

David Britton, Loic Lannelongue, Tobias Fitschen, Luis Villar Padruno, Michael Sparks, Nicolas Labra Cataldo, Ben Parkes, Alejandro Gallego Schmidt, Graeme Stewart, Sam Skipsey, Sofia Vallecorsa, Zach Marshall

European Research Council
Established by the European Commission

HEP's (indirect) energy awareness

"I need to make my code faster"

HEP's (indirect) energy awareness

"I need to make my code faster"

Because I want results ASAP


Attribution link


Attribution link


Attribution link

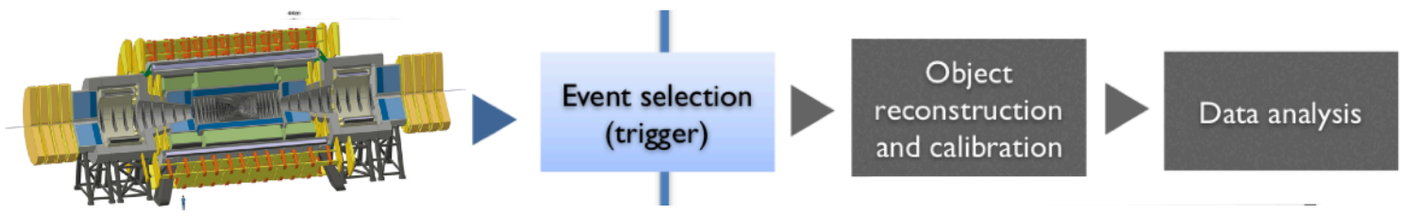
HEP's (indirect) energy awareness

"I need to make my code faster"

Because I want results ASAP

Attribution link Attribution link Attribution link

Because I'm running on a time/latency-constrained environment



CERN-2022-02

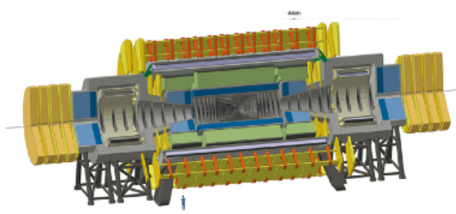
HEP's (indirect) energy awareness

"I need to make my code faster"

Because I want results ASAP

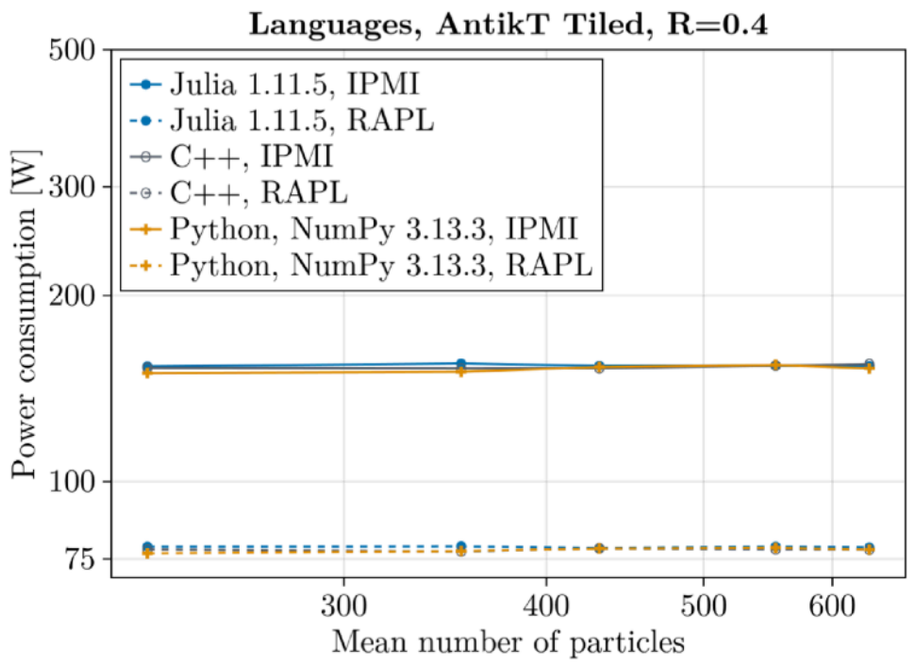
Attribution link | Attribution link | Attribution link

Because I'm running on a time/latency-constrained environment



CERN-2022-02

Because it's more energy efficient



G. Stewart et al, HSF/ WLCG workshop 2025

Constant power consumption

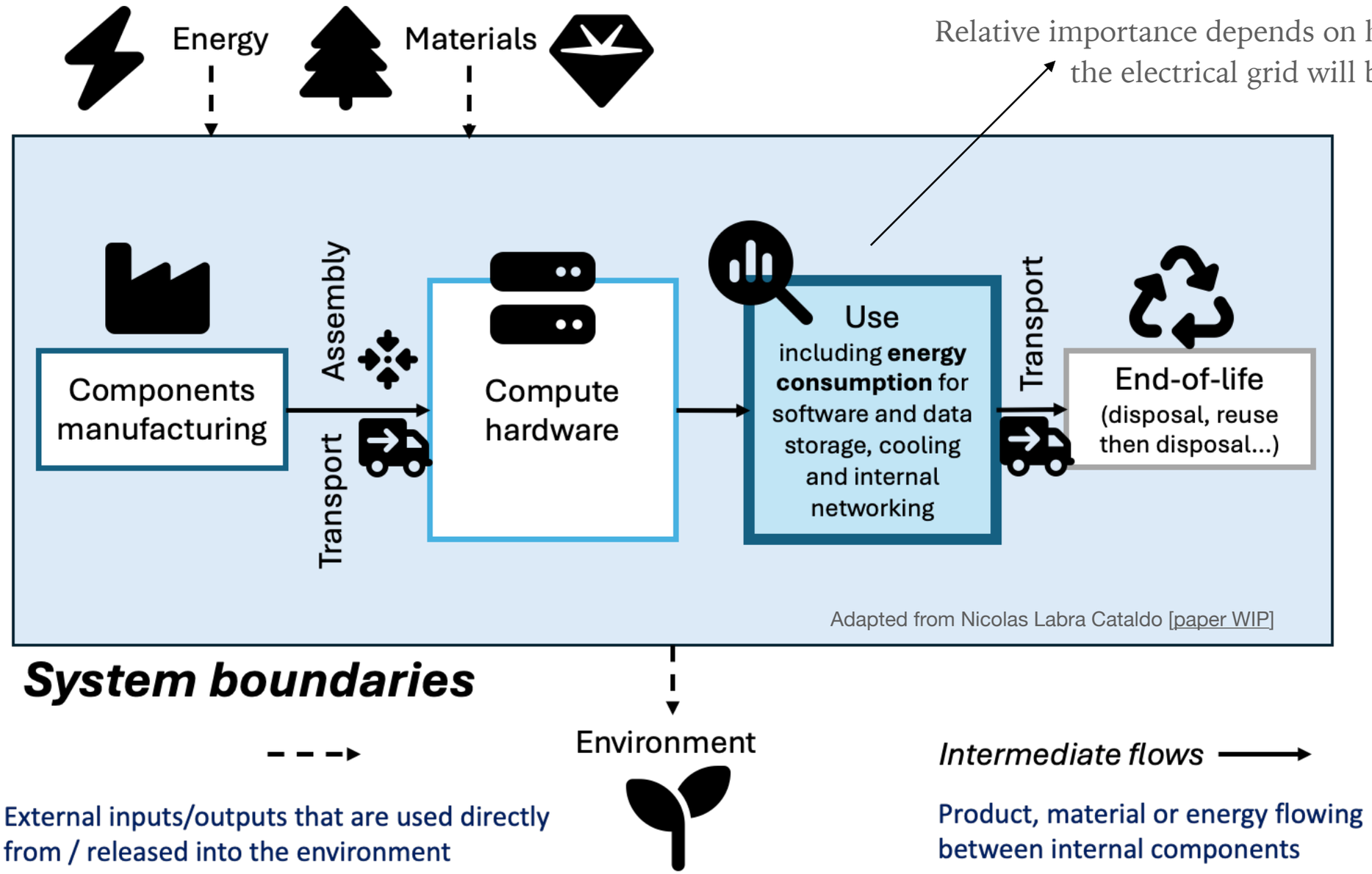
as a function of event complexity

At first order (left: CPU case),

energy consumption

is **proportional** to **runtime**

From energy awareness to resource awareness



https://www.iso.org/standard/37456.html, but also https://arxiv.org/abs/2506.14365 from our community

Icons: https://fontawesome.com

From resource awareness to sustainability

Resource awareness: energy (and resource) costs of present/future computing aren't always visible to us, but they matter because

From resource awareness to sustainability

Resource awareness: energy (and resource) costs of present/future computing aren't always visible to us, but they matter because

(1) We don't want to be caught by surprise

If/when **we don't have the resources** to operate in the same way as we do now

e.g. labs/funding agencies impose **restrictions on scarce resources**

(Plus, who doesn't want to do more physics with the same resources?)

From resource awareness to sustainability

Resource awareness: energy (and resource) costs of present/future computing aren't always visible to us, but they matter because

(1) We don't want to be caught by surprise

If/when **we don't have the resources** to operate in the same way as we do now

e.g. labs/funding agencies impose **restrictions on scarce resources**

(Plus, who doesn't want to do more physics with the same resources?)

(2) Our work must be sustainable

"Sustainability: meeting the needs of the present without compromising the ability of future generations to meet their own needs" (UN Brundtland Commission)

From resource awareness to sustainability

Resource awareness: energy (and resource) costs of present/future computing aren't always visible to us, but they matter because

(1) We don't want to be caught by surprise

If/when **we don't have the resources** to operate in the same way as we do now

e.g. labs/funding agencies impose **restrictions on scarce resources**

(Plus, who doesn't want to do more physics with the same resources?)

(2) Our work must be sustainable (HECAP+, ECR)

> Moreover, particle physics will be under increasing pressure to justify its impacts, and relying on grid decarbonization will not be enough to guarantee that large-scale experiments will be deemed viable.

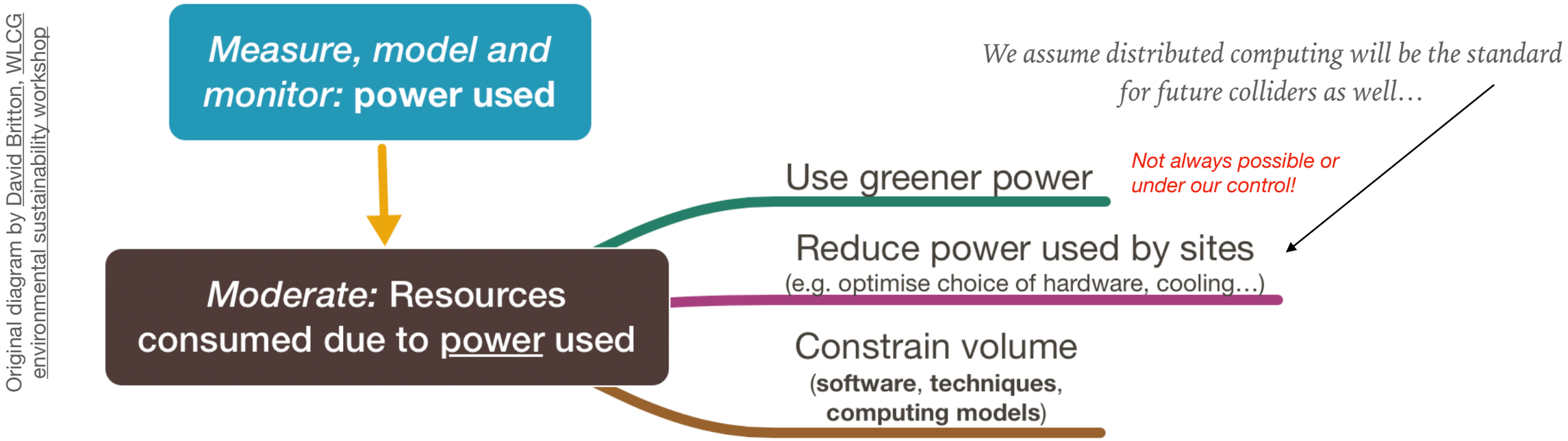
Jevon's paradox may not apply, see also arXiv:2411.03473v3

> facility. In this regard, **the next flagship collider should be built and run in the most sustainable way, and collaboration with institutions from countries other than CERN member states should be increased (R. 4.14).**

"Sustainability: meeting the needs of the present without compromising the ability of future generations to meet their own needs" (UN Brundtland Commission)

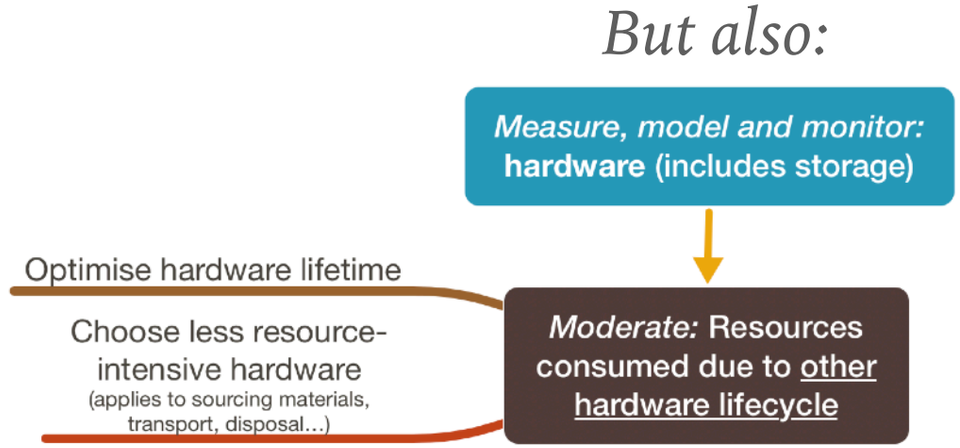
The University of Manchester

What can we measure, model, monitor, and improve?



Some (non-exhaustive) thoughts on resource/ energy-aware compute directions in the next slides

Also remember:
new technologies could also surprise us!



Software and languages

Constrain volume
(**software, techniques, computing models**)

"Sustainable software: software that will continue to be available in the future, on new platforms, meeting new needs" (Dan Katz's blog)

***Sustainable* software** optimises use of **human resources**, see also E. Rodriguez's talk

Improving **common software** can amplify gains

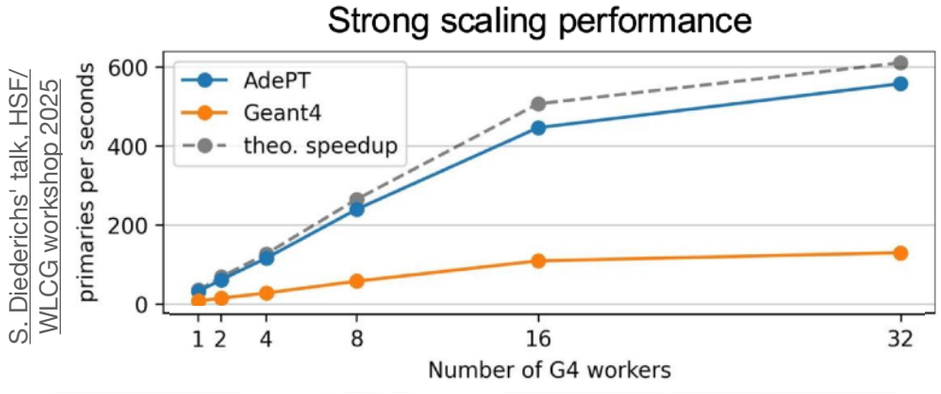
...including across different JENA Research Infrastructures

In **green**: percentage of used CPU resources involving common software (event generators, detector simulation w/Geant 4)

Data from LHCC-2022-005, ATLAS HL-LHC with "aggressive" R&D (includes GPU/ML)

Data from arXiv:2307.10823

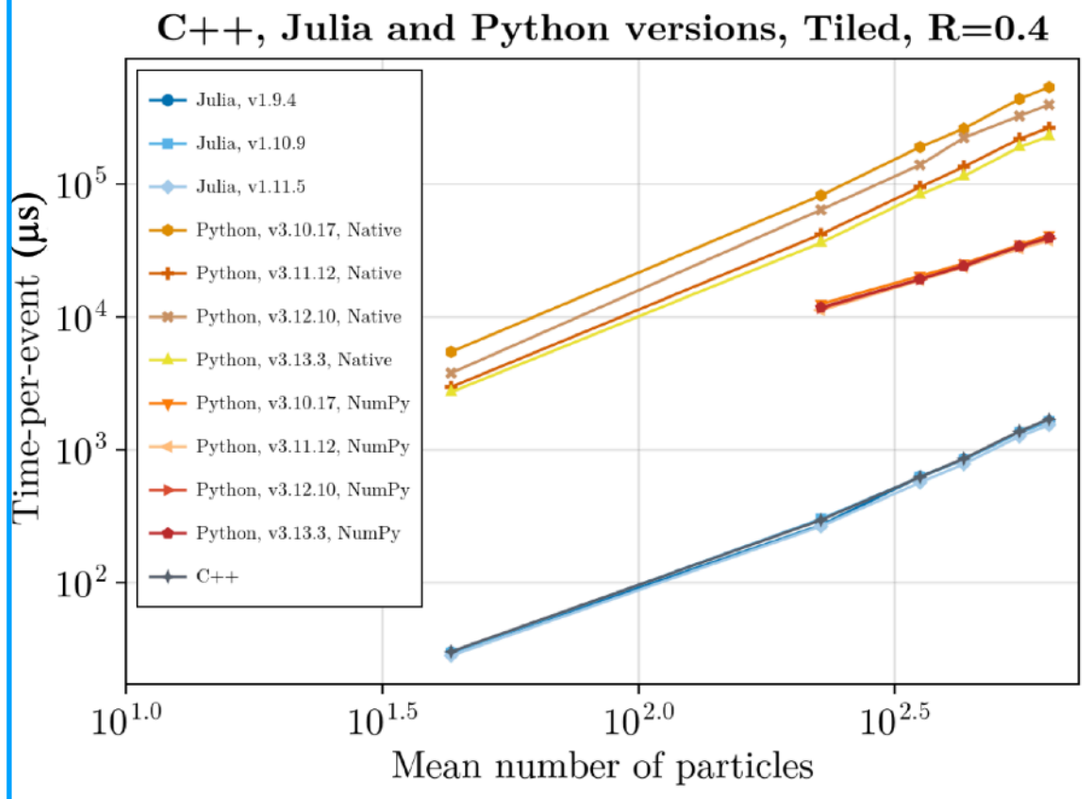
Current performance of AdePT (offloading G4 EM showers to GPU)



S. Diederichs' talk, HSF/WLCG workshop 2025

```
Application: HGCALTB
Geometry: HGCAL test beam
No magnetic field
Input: 10k pi- with 30 GeV
Scoring: via sensitive
detector code
GPU: Nvidia RTX4090
CPU: AMD Ryzen 9 16 cores
```

Faster, (human-)**sustainable languages** could change future landscape



G. Stewart et al, HSF/WLCG workshop 2025

Data and storage

Constrain volume
(**software**, **techniques**, **computing models**)

Reduce power used by sites
(e.g. optimise choice of hardware, cooling…)

The **energy/carbon cost** of storage is not negligible
→ if unchanged, future experiments will need
computing models / sites that take this into account

In particular for high-pile-up machines where large amounts of simulated events are needed to match statistical precision of data

Table 4. `RAW` data estimates for FCC-ee.

Run	$\sqrt{s}$ (GeV)	Statistics	`RAW` data
Z	91.2	$3 \cdot 10^{12}$ Z decays (visible)	3–6 EB
WW	160	10^8 W^+W^- events	0.1–0.2 PB
ZH	240	10^6 ZH events	1–2 TB
$t\bar{t}$	350, 365	10^6 $t\bar{t}$ events	1–2 TB

Updated numbers: G. Ganis's talk (2025) arXiv:2111.10094

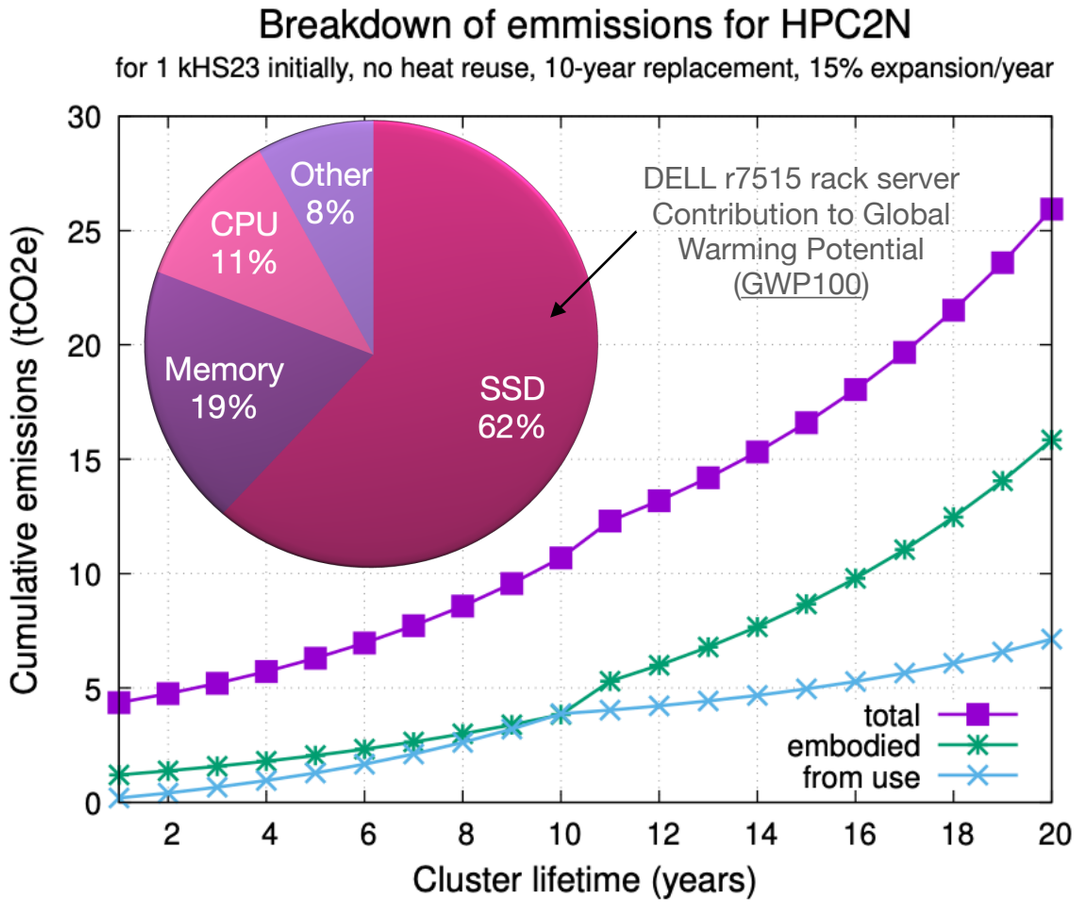
AOD: o(PB) or less

For **FCC-hh**:

Total storage including MC: o(ZB)

Trigger challenge for a sample detector:

- 250 TB/s for calorimeters and muons
- 2 PB/s for tracking



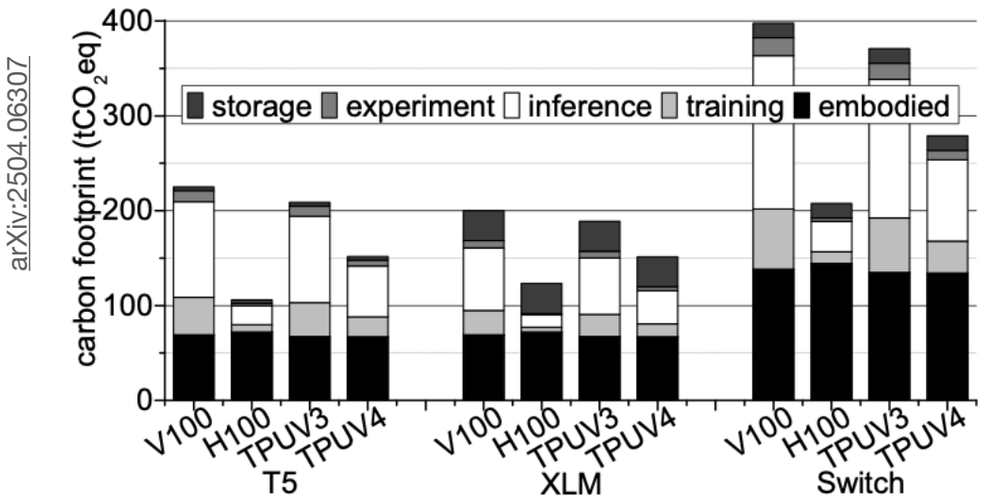
DELL, Lifecycle analysis of r7515 server
https://dl.acm.org/doi/10.1145/3630614.3630616
Example plot from arXiv:2506.14365, **NEW!**

"New" technologies (AI/heterogeneous and beyond)

Reduce power used by sites
(e.g. optimise choice of hardware, cooling…)

Impact of AI workflows?

Example: modelling the training of LLMs on different devices



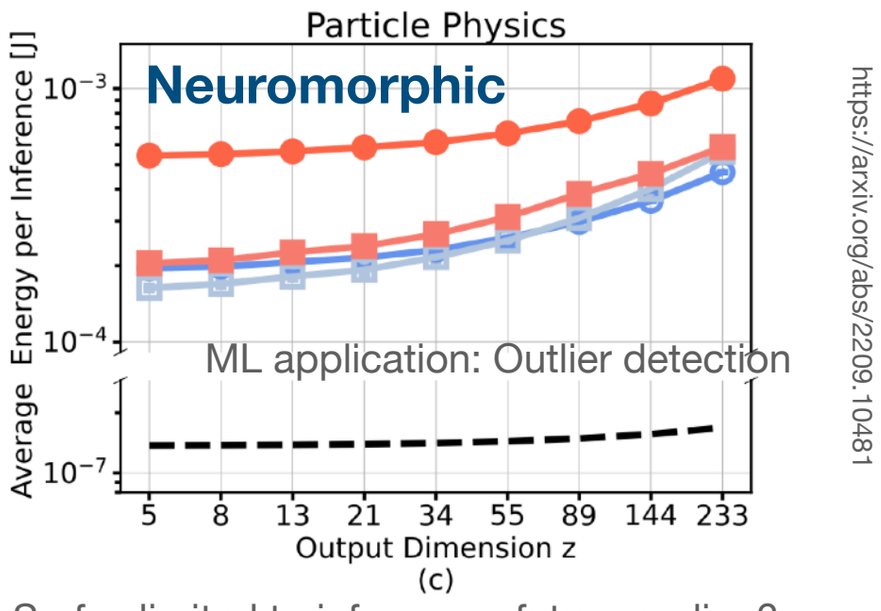
(Related: heterogeneous workflows)

Estimation of GPU power consumption / optimisation already needed and in progress for (HL-)LHC

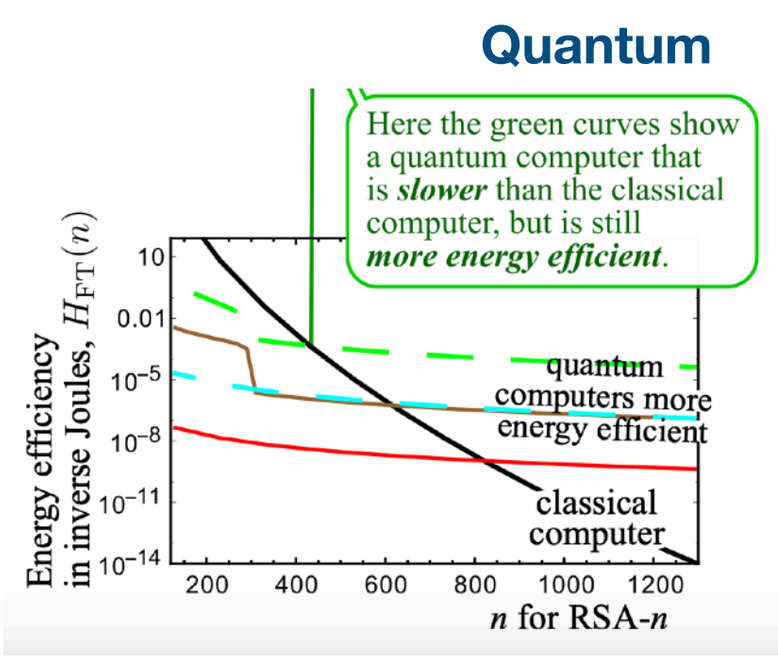
- Reduce the cost, energy consumption, and carbon footprint of computing by being able to use a wide range of resources including Grid, public or private clouds, HPCs, and by expanding the architectures from the original x86 to include, for example, ARM, RISC-V, GPUs, and other technologies that may become available. Our strategy is to build agility so that opportunities can be seized.

WLCG EPPSU submission

Near and far (fetched) future…



So far, limited to inference…future scaling?



European Research Council
Established by the European Commission

The University of Manchester

Conclusions: measuring, modelling and improving

What should we be doing to improve resource/energy-awareness starting from HL-LHC and all the way to next generation colliders?

- **Measure:**
  - Software: many tools on the market, we can start practicing already [GreenAlgorithms, CodeCarbon, Scaphandre, a short review…]
  - Sites: ongoing work by WLCG + LHC experiments (& others) [December/May workshop]
  - Lifecycle / embodied: can collaborate with other fields (and encourage vendors to publish numbers)
- **Model** (important for **extrapolations**):
  - Software: see e.g. work in ECO:DIGIT for comprehensive modelling
  - Lifecycle / embodied: see recent work in [arXiv:2506.14365]
- **Prepare to improve:**
  - Include energy/resource-awareness considerations from the onset of computing models
    - Or make it part of a strategy, like WLCG strategy 2024-2027
  - Encourage transparency and publication of estimates (requires good accounting)
  - To discuss further: WLCG sustainability forum in the works, kick-off before summer break

Thank you for your attention!