Carbon, Power, and Sustainability in ATLAS Computing

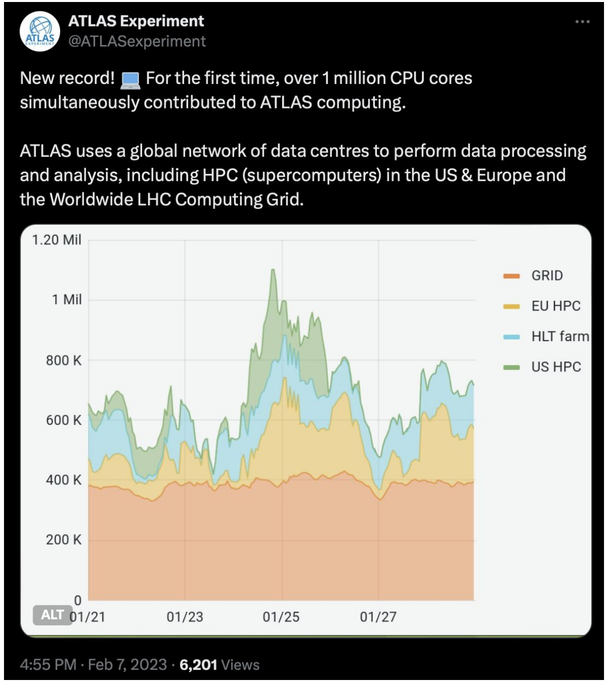
CHEP 2024

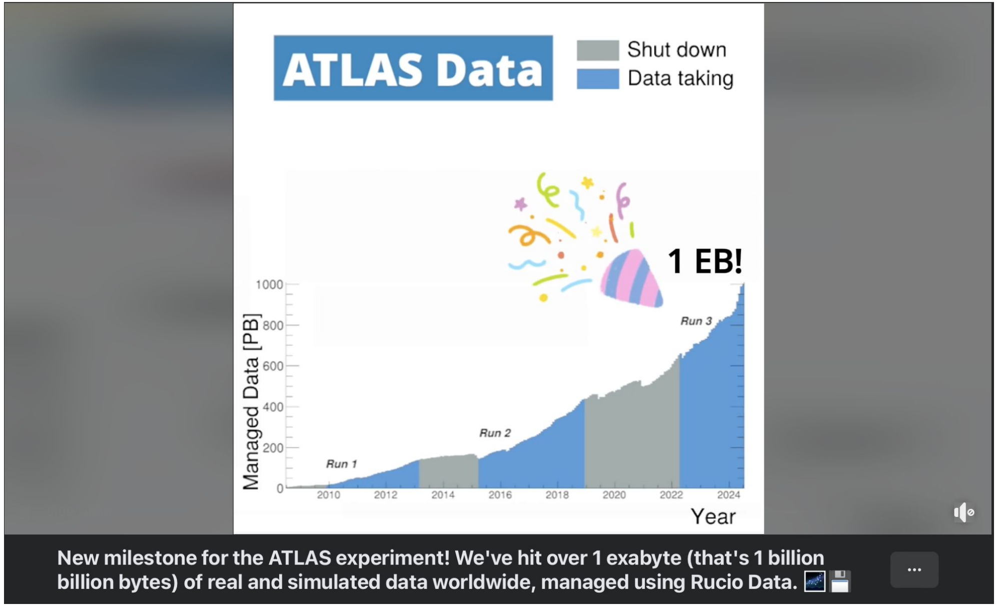
23 October 2024

Zach Marshall (LBNL) on behalf of the ATLAS Computing Activity

ATLAS computing resource scale

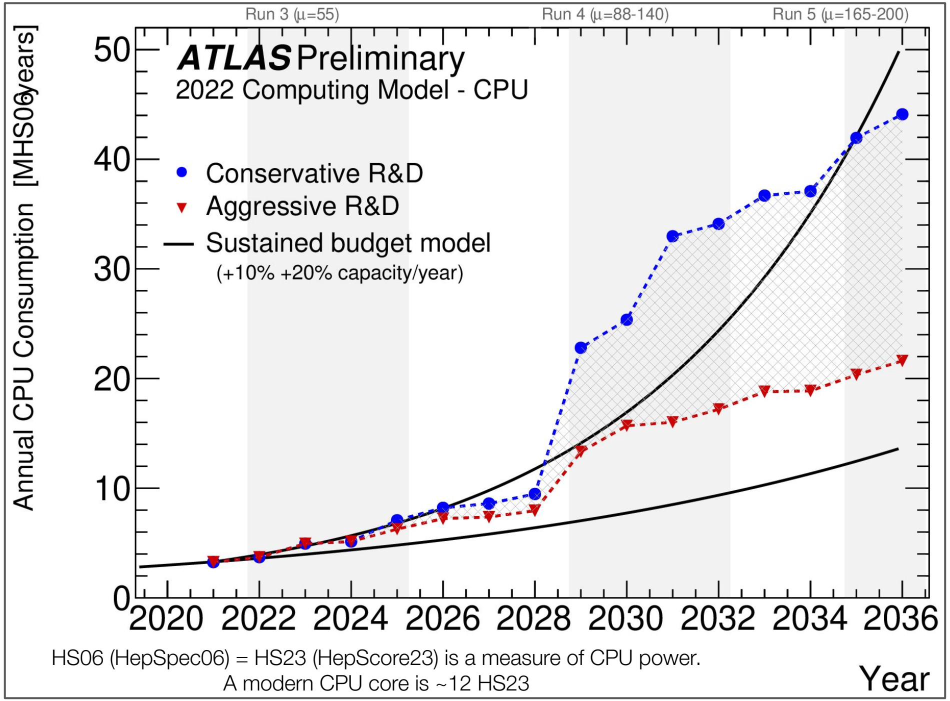
- ATLAS operates ~700k cores of compute (peaks at 1M cores), 400 PB of disk, 650 PB of tape
- Distributed over ~100 sites worldwide (much more info)
- Combination of high-throughput computing centers (WLCG sites), high-performance computing (HPC) systems, cloud computing (e.g. Google, AWS), volunteer computing (ATLAS@Home)
    - Peaks mostly from large HPC systems (not HEP-specific systems) like Vega, Perlmutter, etc

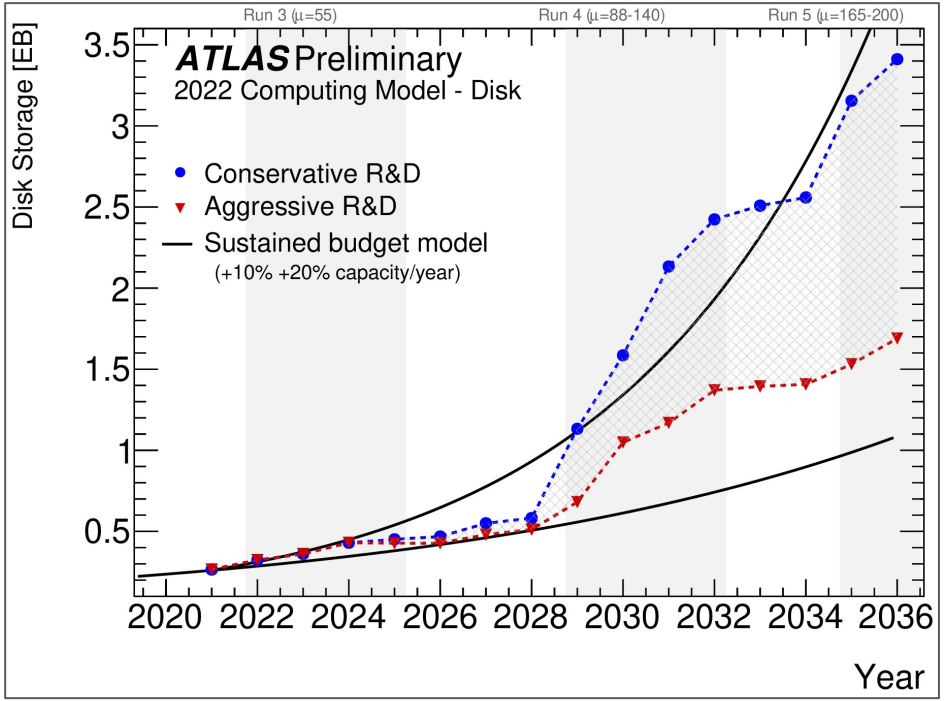




Computing resource extrapolation

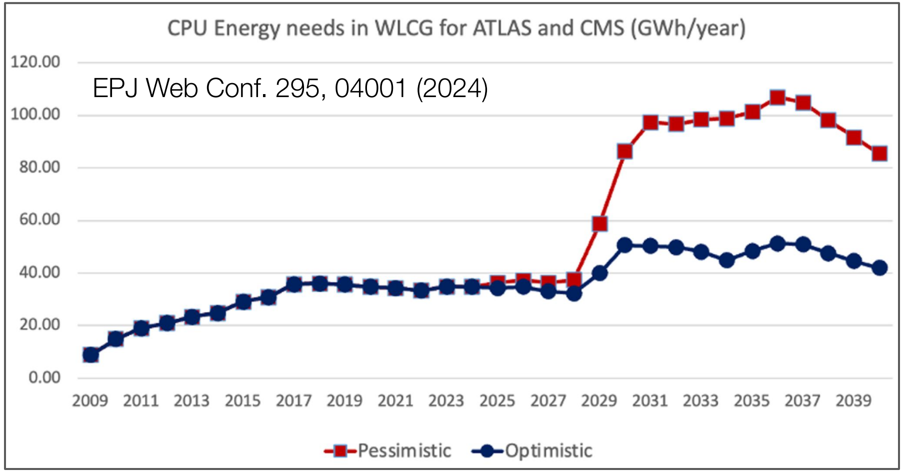
- As a part of the HL-LHC upgrades, computing resources required to expand significantly
- 3–5x compute, disk, and tape by 2031; another >2x by 2041
- Requests for more resources made 12–18 months in advance
  - In-year, Jevons paradox applies; over several years, opportunity for reduction through optimization

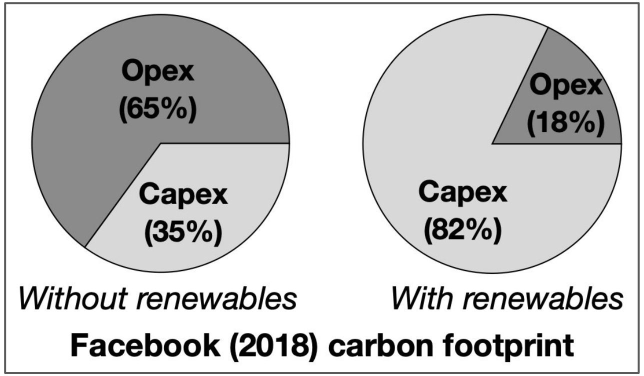




(Computing) Carbon footprints

- Much more attention to carbon footprints recently (CERN reports in 2021 and 2023; see bkup)
  - Outside groups also looking: EE HPC WG (recommendations), Open Compute Project, GreenDiSC
- Computing is ~5% of CERN's footprint when the LHC runs (accelerator cooling ~80%)
  - CERN mostly draws power from the French (nuclear, quite low-carbon) grid
- About 10–20% of ATLAS computing is at CERN, much of the rest is on less green power
- Most studies focus on *power* (operational carbon); embodied carbon >15% of the total (1, 2)
  - Relevance of embodied carbon (Scope 3) will increase as western power grids decarbonize
- Extrapolating to HL-LHC, computing could be a **large** fraction of the ATLAS carbon footprint!



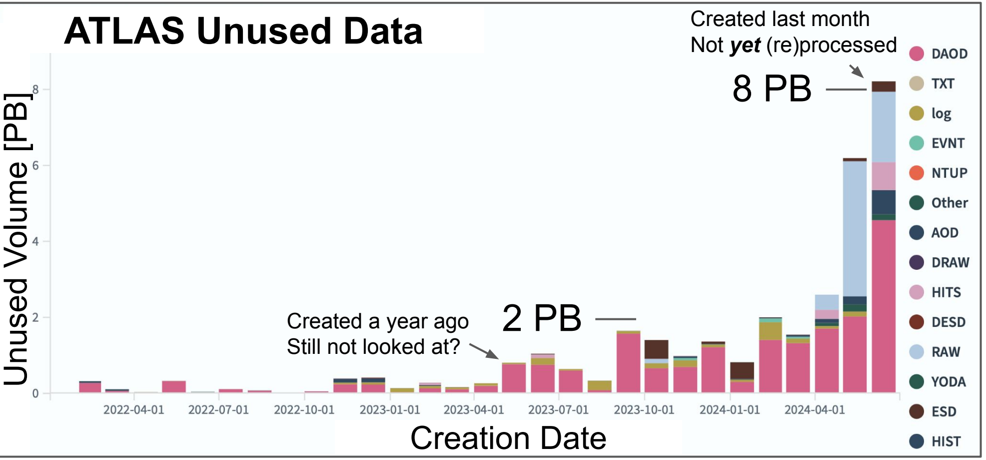


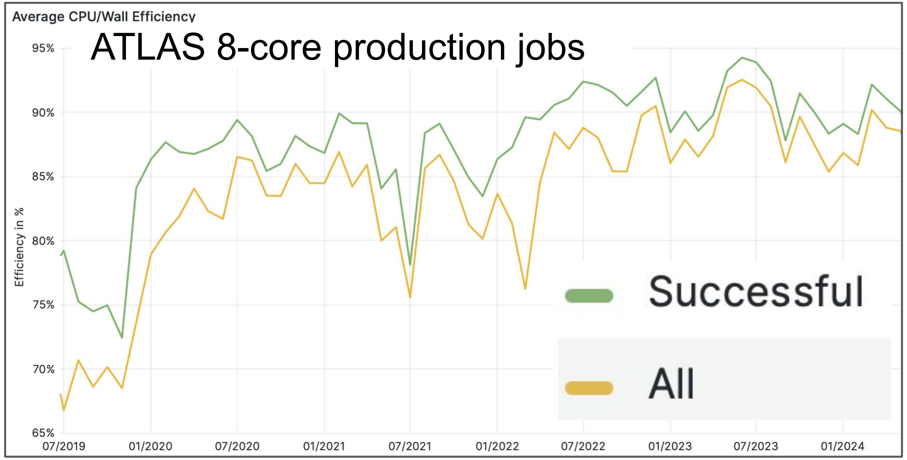
arXiv:2011.02839

Opex: Operational carbon
Capex: Embodied carbon

Waste, loss, and unused data

- *Using* carbon for important science is "allowed" — *wasting* carbon is never ok!
- Constantly monitoring *unused data* in the production system
  - Requests made, bug found, reproduced before the data were looked at
  - Processing done based on a too-inclusive pattern (mc*)
- Steady progress to improve CPU/wall efficiency to over 90% (below for 8-core production jobs)
  - Impact of failed jobs is visible; errors on copying output (after all the CPU has been consumed) are killer!
  - Constant effort to reduce serial portions of many-core jobs as well (wasted CPU and power)





Computing model and policies

- Many aspects of the computing model affect carbon
  - Until recently, considered mostly in financial terms
  - Additional CPU is easier to find than additional storage; mostly attempts to reduce storage requirements
- Data carousel: use tape as a more active storage medium
  - Designed to reduce disk usage; also reduces carbon as tape is lower-carbon (in all ways) than disk
- Data reproduction: to save disk, reproduce little-used data when required
  - Able now to reproduce data on-demand, usually in <48h
  - Reduces need for archival storage, “just in case” storage; similar action with intermediate data formats
  - Being reexamined with carbon in mind — storage is O(25%) of the operational carbon footprint; not obvious where the optimal working point is yet
- In a recent site survey, we saw a mixture of opinions:
  - “We retire old CPU hardware to reduce our carbon footprint”
  - “We run old CPU hardware as long as possible to reduce our carbon footprint”
  - This screams for an informed recommendation!

Building awareness

- Summary sent when a user's grid job finishes
- Now includes an **average** carbon estimate
  - Links to a page explaining that number
  - Along with some background and some equivalents for comparison
- Averaged for several reasons
  - Inaccuracy of and missing site-specific data (work ongoing to improve these data)
  - CPU doesn't sit idle (moving a job would not reduce the experiment's total footprint)
  - Users pushing on a single site could cause other problems (including waste)
  - Faster code → lower carbon footprint
- **Not** intended for "shame"; a reminder that Grid use is relevant to the environment
  - Code optimization and success rate matter!
- Also tracked for production campaigns and reported back to production managers →

PanDA

Summary of TaskID: 41161761

Detail	Value
Created	2024-09-11 12:21:24
Ended	2024-09-11 12:22:30.624502
Final Status	done

Total Number of Inputs

Category	Count
Succeeded	1
Failed	0
Cancelled	0
Total	1

Error Dialog: None

Datasets

In	[]
Out	[]
Log	['user.zmarshal.test.4a25be90-fb43-4d42-bdde-854c86b41bea.9-x86_64-bash-pythonrecommended-x509.log/']

Parameters

Command	prun --exec=myPunJob.sh --outDS=user.zmarshal.test.4a25be90-fb43-4d42-bdde-854c86b41bea.9-x86_64-bash-pythonrecommended-x509 --noBuild

Estimated Carbon Footprint for the Task

Category	gCO2
Succeeded	0.01 gCO2
Failed	0 gCO2
Cancelled	0 gCO2
Total	**0.01 gCO2**

More details on estimation: https://panda-wms.readthedocs.io/en/latest/advanced/carbon_footprint.html

Panda Problem Reporting

- The eGroup for help request: hn-atlas-dist-analysis-help@cern.ch
- The Discourse forum for distributed computing help: https://atlas-talk.web.cern.ch/c/distributed-computing-help
- The JIRA portal for software bug: https://its.cern.ch/jira/browse/ATLASPANDA

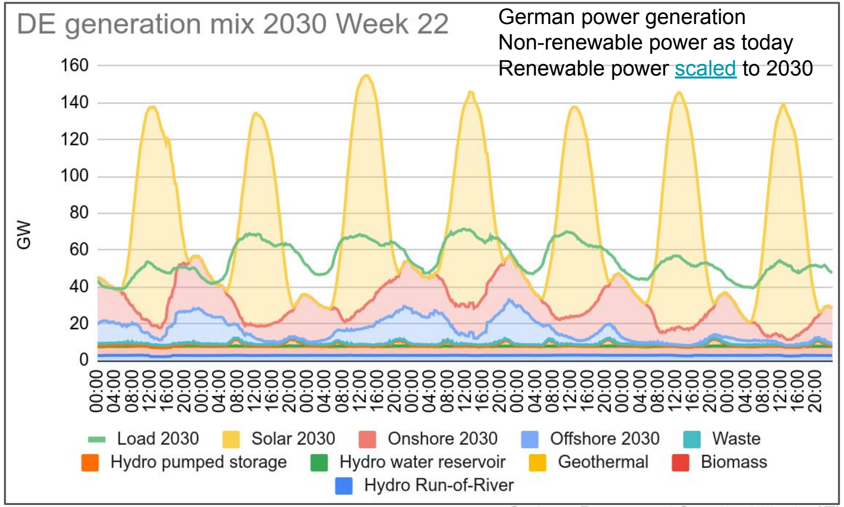
ATLAS PanDA

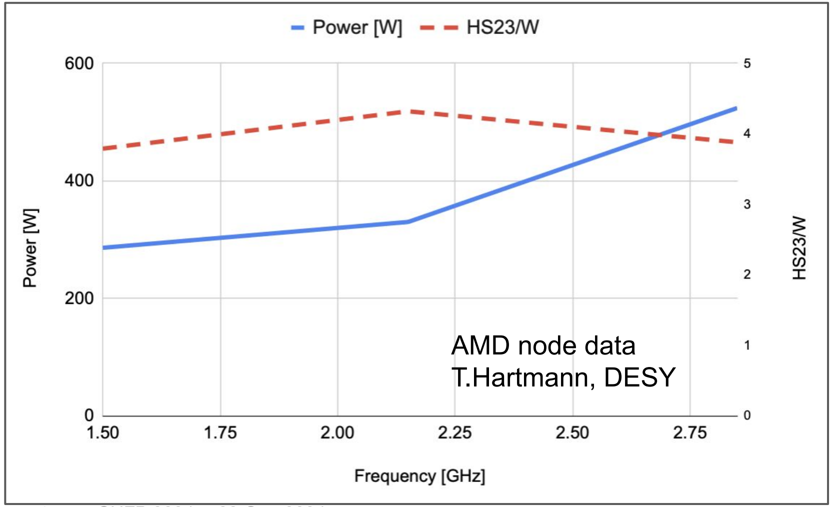
120 tasks listed. Estimated CO_2 emission in total: 2 t, including 129 kg by failed jobs.

Frequency scaling, power modulation, and checkpointing

- Extrapolating to 2030, renewable energy will be *significant* — easily more than peak demand
- Suggests there will be periods of *very* green energy, and periods of less green energy
- Obvious question: can we use more energy when it is greener?
  - Related: if battery capacity is significant, can we run sites off battery backup power?
- CPUs offer an easy lever: frequency modulation (clocks can run at lower rates)
- **Work** per **Watt** is actually **greater** at **reduced frequency**! (P~f V^2, and V falls with f)
  - Need to understand embodied carbon implications; in terms of site load modulation it's clearly a win
  - Potential for **checkpointing** too — work ongoing (could help for scheduled outages / ramp-down [1])

P: Power
V: Voltage
f: Frequency





New data centers

- Most important to data centers is Power Usage Efficiency (PUE)
  - Power into a data center ÷ power to IT elements; 1.1 is *good*; 1.5+ is *typical*
  - Operational carbon is (Power to IT elements) * (PUE) * (Power Grid Carbon Intensity)
- Steel / concrete buildings (like data centers) typically cost 200–800 kgCO_2eq/m^2 of space (1,2)
- We can then write down how long it takes for a new data center to “pay off” its carbon footprint
- Taking typical examples (500 kgCO_2eq/m^2, 10,000 m^2, 5 MW data center):

In California (0.12 kgCO_2eq/kWh) changing PUE from 1.2 to 1.1:

~12 years to pay off

At CERN (0.075 kgCO_2eq/kWh) changing PUE from 1.6 to 1.1:

<5 years to pay off

- In other words: you should **almost certainly** build a new data center!
  - Also: physical buildings are likely O(10%) of the total carbon footprint of computing
- Water cooling and waste-heat usage can reduce an effective PUE to ~0.6
  - Effective PUE: (power into data center – power *out of* data center) ÷ power to IT elements
  - These are **obvious** additions that any new data center should incorporate if close to a population!

Carbon Summary and Future Work

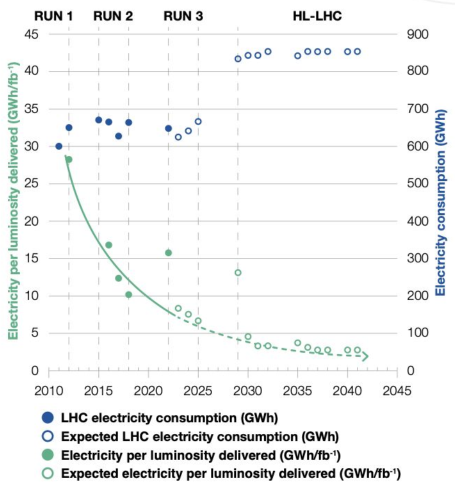
- Effort ongoing to build a complete carbon footprint model for ATLAS Computing
  - Studies ongoing in storage footprint and storage carbon optimization, frequency scaling and checkpointing, networking aspects, cooling systems, site power consumption, platform adoption…
- Connected effort to make users, developers, operators, admins more aware of the environment
- Clear that limited models risk making *harmful* recommendations
  - If we cared *only* about power, we should buy new hardware every year!
- Goal is to make useful recommendations to sites and towards a more sustainable computing model as we approach the HL-LHC
  - What is the *carbon impact* of GPU usage? The payoff point in terms of GPU load? (30–60%?)
  - In carbon terms, what is the optimal storage configuration? (RAID? Background task configuration?)
  - How should we think about the tradeoff between *disk*, *tape*, and *CPU*?
  - What is the “optimal” approach to take towards old hardware? (others thinking about this too)
- Closely watching extrapolations that affect these recommendations
  - De-carbonization of the power grid will likely emphasize embodied carbon
  - Hardware landscape is constantly evolving (ARM, RISC-V, GPUs, Grace Hopper, …)
  - Improved chip performance (flops/W) is coming with larger packages and less modularity and repairability — potential impacts on site operations and hardware lifetimes

Thank you!

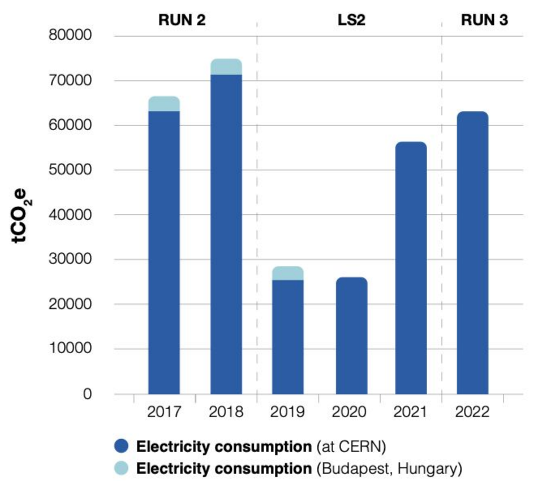
CERN Environmental Reports

ATLAS EXPERIMENT

- Much more attention to carbon footprints recently (CERN reports in 2021 and 2023)
- Computing is ~5% of CERN's footprint when the LHC runs (accelerator cooling ~80%)
  - CERN mostly draws power from the French (nuclear, quite green) grid
  - Data center in Hungary that ran until 2019 had a visible impact on energy consumption (now moving into the new Prevessin Data Center)



ELECTRICITY INTENSITY OF THE LHC
Quantity of electricity used to run the LHC per unit of luminosity delivered, showing that less and less electricity has been needed over time to produce the same amount of data and, hence, scientific output. During the year after each long shutdown, while the machines are being brought back online and progressively ramped up, the luminosity delivered is not at its maximum – as seen in 2022 and 2029.



CERN SCOPE 2 EMISSIONS FOR 2017–2022
Emission calculations for electricity follow a location-based methodology, with average yearly emission factors taken from ADEME Base Empreinte©. From 2017 to 2019, CERN operated a data centre at the Wigner Centre in Budapest, Hungary, for which the emissions are also shown. The location-based emission factors used for Hungary were taken from Bilan Carbone® V8.4.

High Performance Computing Systems

- These systems would likely be built with or without us there
  - Their use, in some sense, 'saves' carbon worldwide
- With good software, we can fill the 'gaps' in these systems (and other systems with similar running profiles) and improve their usage fraction — a win for everyone