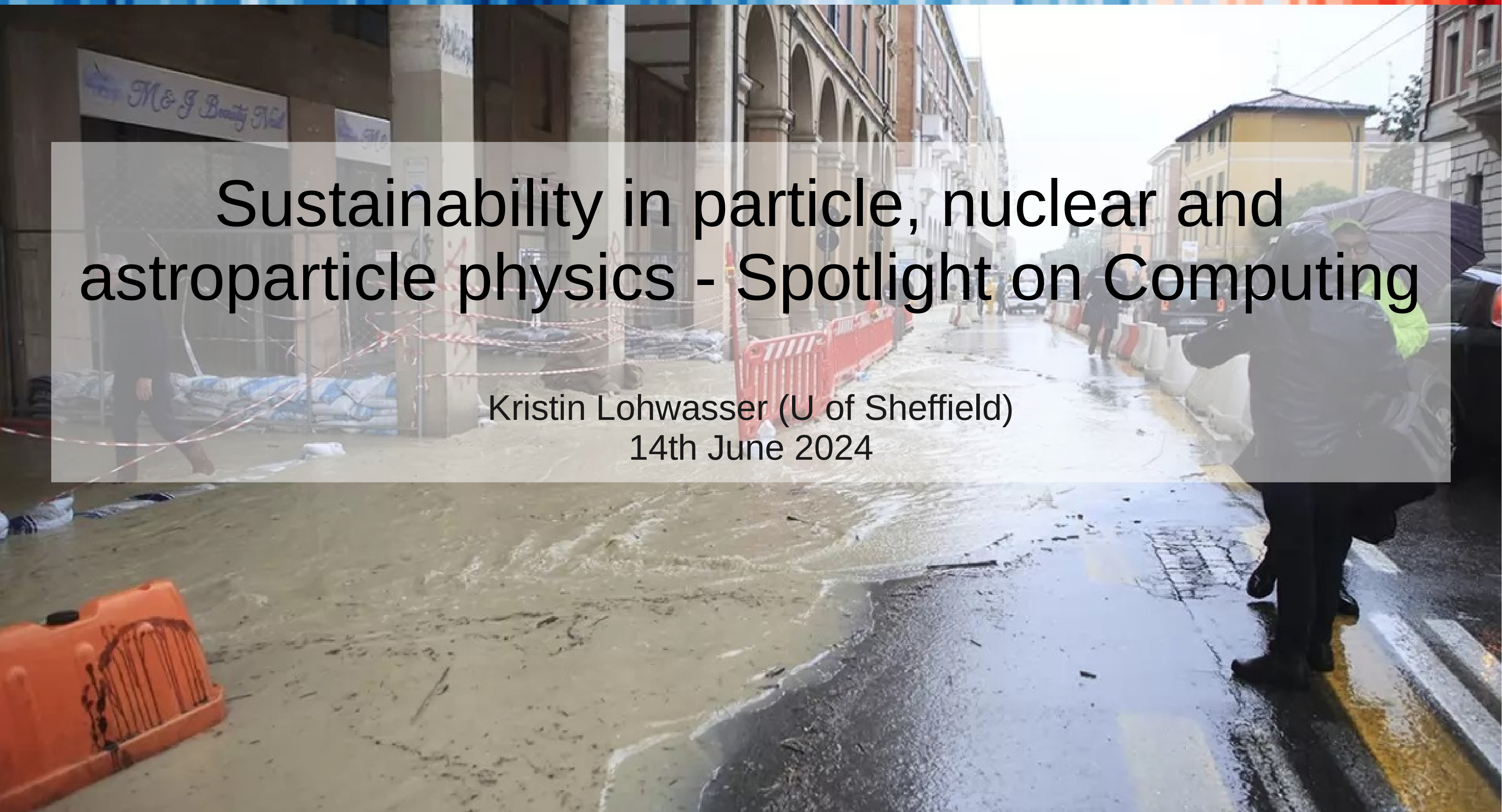
Sustainability in particle, nuclear and astroparticle physics - Spotlight on Computing

Kristin Lohwasser (U of Sheffield)

14th June 2024



Climate Change up close

- We see impacts of rising temperatures: Drought, floods, high temperatures, severe weather, e.g. here in Emilia Romagna in Mai 2023:

  →Rainfalls of 7 months in 2 weeks, in some places up to 6 months of rains in 36 hours

  → **at least 15 people dead**, 400 landslides, 42 cities flooded, **damage caused: €7 billion**

  → https://en.wikipedia.org/wiki/2023_Emilia-Romagna_floods

- And that is not even the most recent “natural” disaster



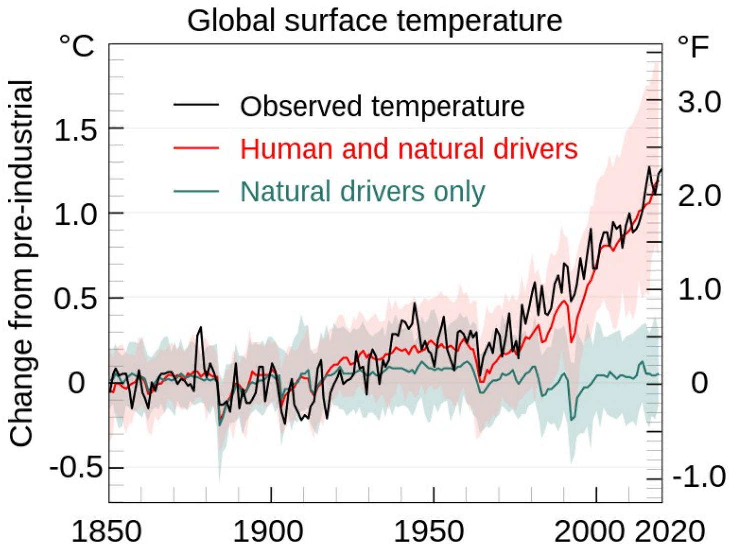
Weather or Climate? And is it “natural”?

- Whilst extreme weather events have a finite probability and therefore “just” can happen, this finite probability is strongly influenced by climate conditions

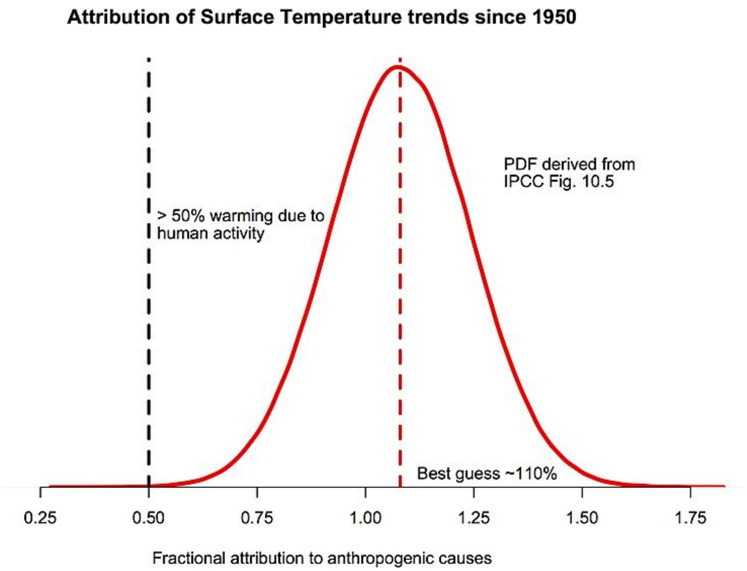
  → studied in extreme event attribution / attribution science → new field of study in meteorology and climate science using statistical methods and concepts not completely foreign to particle physicists.
  → https://en.wikipedia.org/wiki/Extreme_event_attribution

- Using the framework of attribution science, the current level of climate change is fully attributed attributed to human activity

- Climate sets the probability (like a cross-section)
- Weather is a single event (like a collision) drawn from that cross-section



https://en.wikipedia.org/wiki/Attribution_of_recent_climate_change



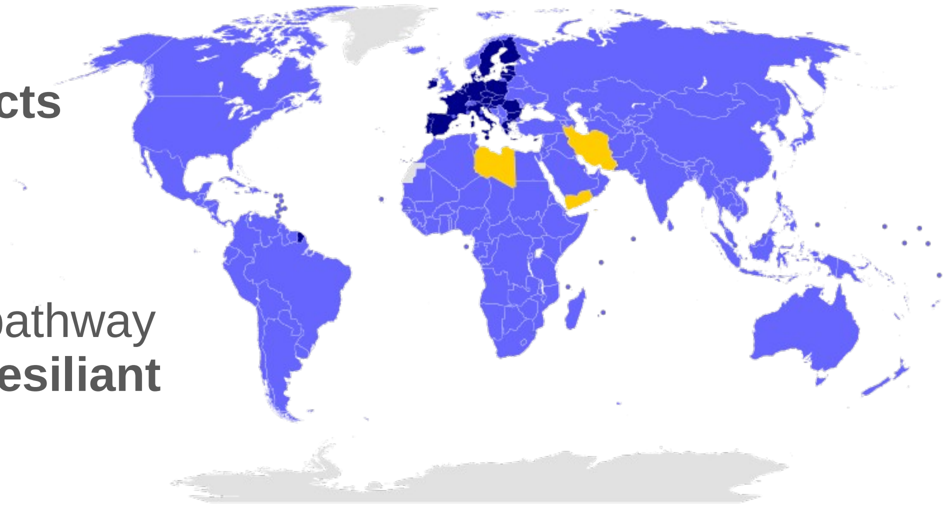
Political consequences

- **The 2015 Paris Agreement**
  → Drafted 30 November – 12 December 2015 in Le Bourget, France
  → Effective 4 November 2016 after more than 55 UNFCCC parties, accounting for 55% of global greenhouse gas emissions had ratified and acceded
  → 195 signatories

- Hold global average temperature **well below 2°C** above pre-industrial levels and to pursue efforts to **limit the temperature increase to 1.5 °C**

- Push ability to **adapt to adverse impacts and foster climate resilience**

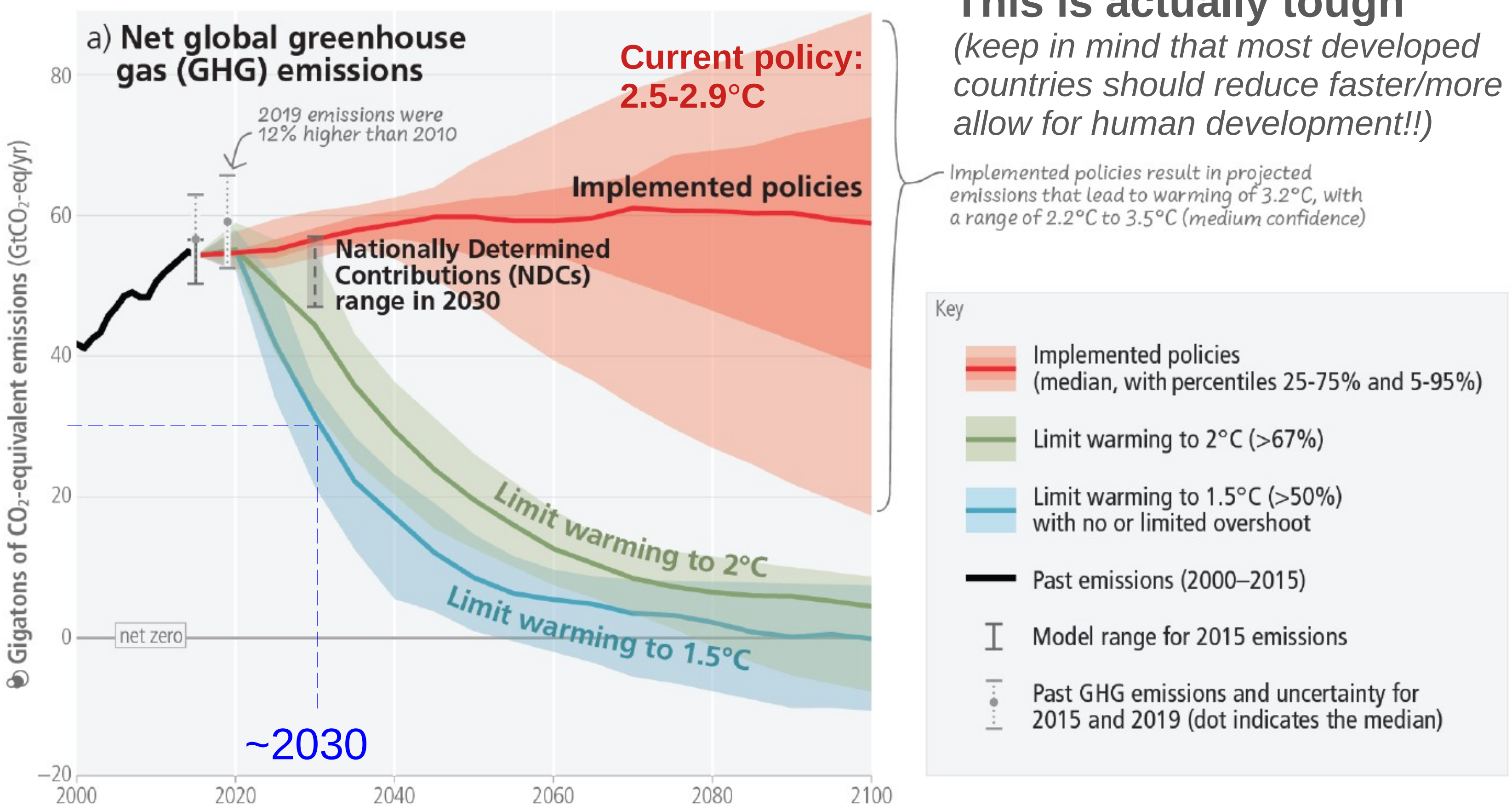
- **Make finance flows consistent** with pathway towards **low emissions and climate-resiliant developement**



Yellow: signed, not ratified

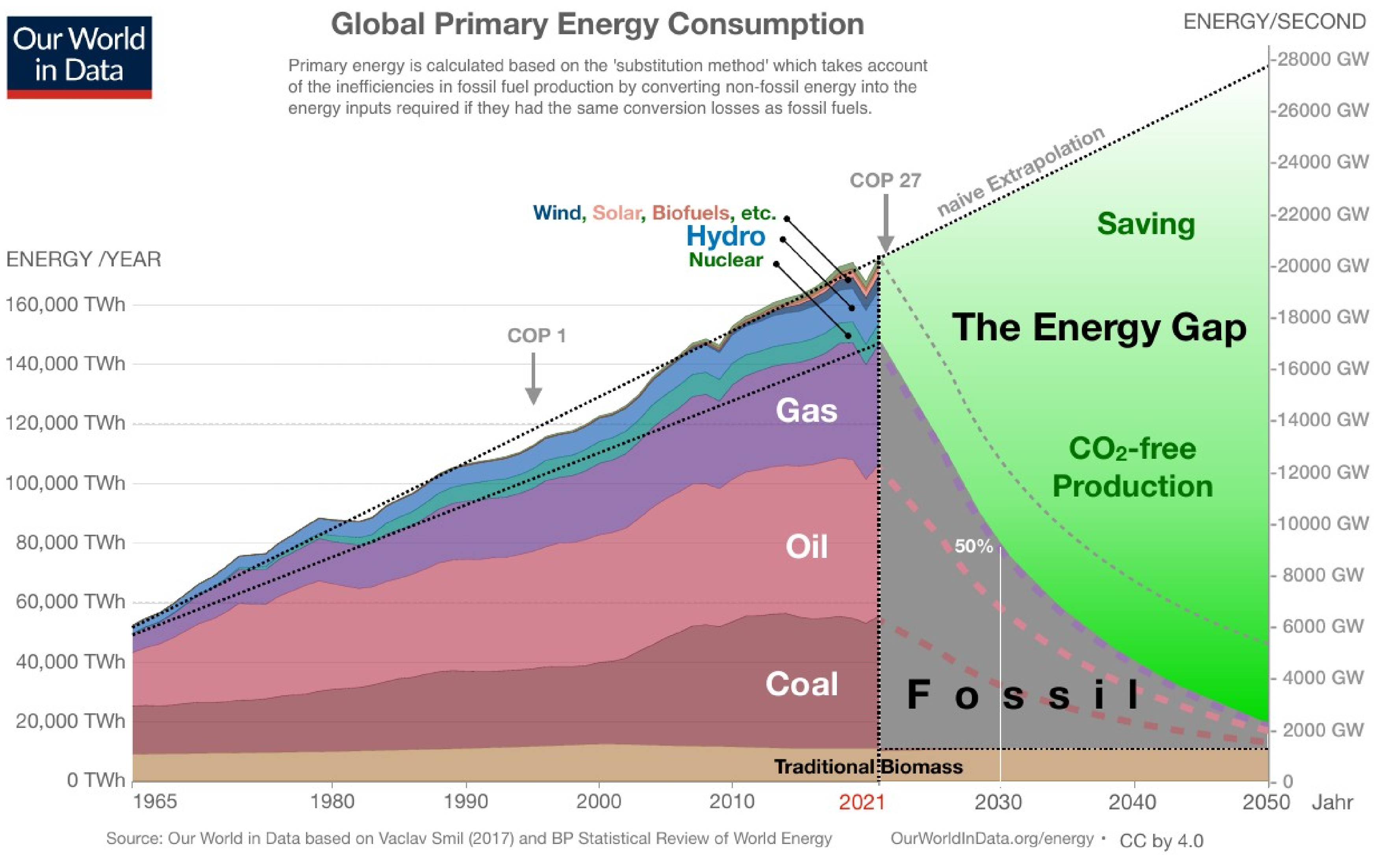
Translation of Paris into Goals

- **Reduction to zero emissions around 2100**
  → A lot of time?
  → 50% of the reduction should be achieved by ~2030 → **in 7 years**



This is actually tough
(keep in mind that most developed countries should reduce faster/more to allow for human development!!)

The energy gap



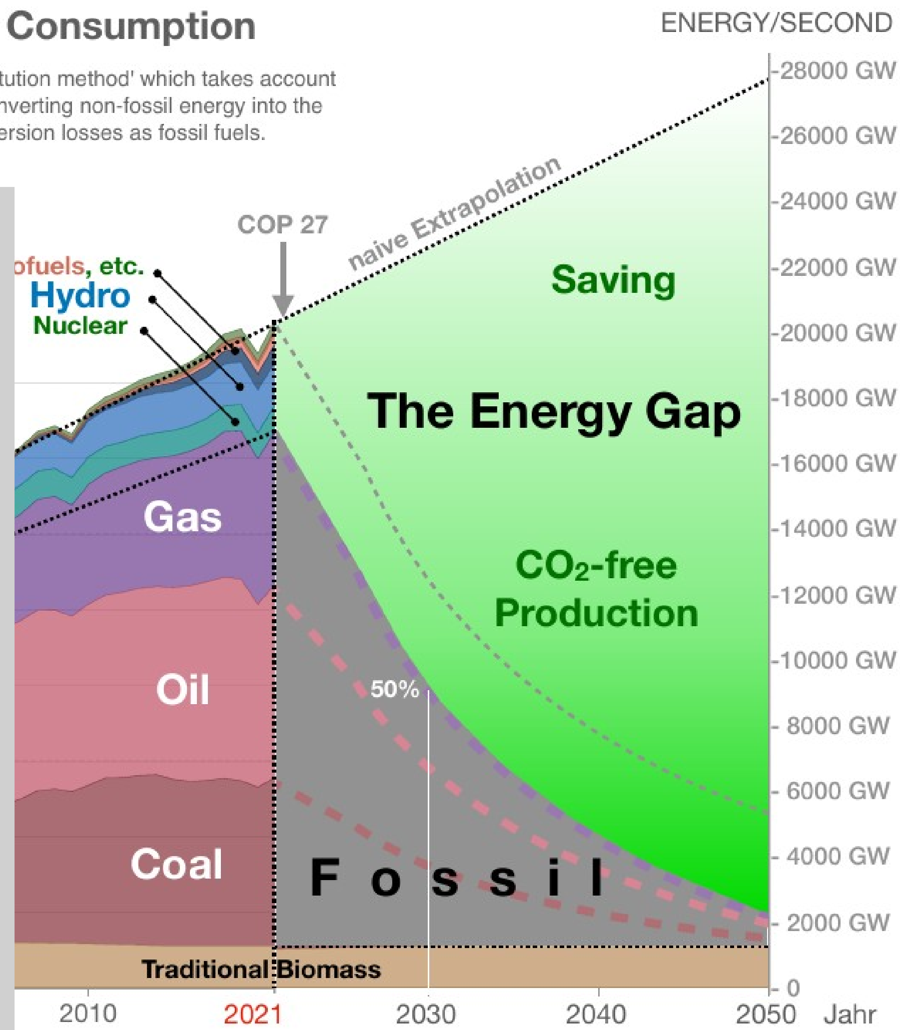
The energy gap

Options:

1) Expand CO_2-free energies
→ factor ~12 in 7 years required;

2) Increase energy efficiency
→ factor ~2 in 7 years
e.g. Electrification of engines (factor 3-5 vs. combustion engine)
e.g. LEDs for lighting (factor 10 vs. light bulb)

3) Save energy
→ factor ~2 in 7 years
e.g.Less travel: online conferences, holidays nearby
e.g. Fewer consumer items, more repair options
e.g. Energy priority for essential things



Source: Our World in Data based on Vaclav Smil (2017) and BP Statistical Review of World Energy OurWorldInData.org/energy •

What does this mean for computing?

Options:

1) Expand CO2-free energies (factor 12)
Renewable power for computing: processors and cooling;
Consider district heating and site selection;
Job scheduling according to energy availability; ...

2) Increase energy efficiency (factor 2)
Optimised processors (clocks, GPUs),
architecture, cooling system,
software, quantum computing?, ...

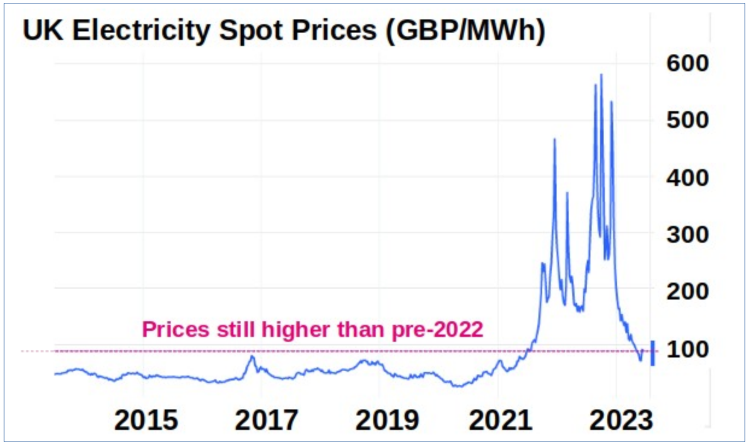
3) Save energy (factor 2)
Prioritise research questions
Optimise debugging, statistics and precision;
Modular and reusable software;
Modular and repairable hardware, reduce purchases;

Can’t we just use green energy and not do anything?

- **Electricity prices are volatile**

- **EU projections from 2016 predict about 25% rise of prices (consumer)**
  → Cut 25% of the physics?

- And it’s not just electricity prices but also hardware



- **Costs of computing infrastructure evaluation 2032 (with 2021 as index)**

- Installed hardware based on computational requirements (15-20% increase/yr), Unit costs (10-20% decrease/yr), 5 years of lifetime
  → **Costs could rise between 0.5 – 5.5 (best vs. worst case scenario)**

- Electricity costs (based on average) consider inflation, power efficiency (30% decrease → no improvement), high prices+high inflation versus both dropping
  → **Costs could rise ranging by 1.6 – 3 – 7 (based on mid capacity)**

Chris Brew (RAL)

Sustainability in HECAP+ and Computing

- **Indeed it has become a big topic even at the recent CHEP conference with activities triggered by the 2022 energy crisis**

- eg. similar workshop (where I stole some of Michael's slides): https://indico.desy.de/event/37480/
  → within a project for digital transformation in the research of universe and matter funded by German ministry (https://erumdatahub.de/)

- It's not a new topic: https://indico.esrf.fr/event/2/
  Sixth Workshop on Energy for Sustainable Science at Research Infrastructures

- Also took inspiration from a reflection document on sustainability in High Energy Physics, Cosmology and Astroparticle Physics + Hadron and Nuclear Physics

- Computing only a small part of this document (and my own work), so this is more of a **broad overview** (and will in some parts repeat some of what has already been discussed)

Environmental sustainability in basic research

A perspective from HECAP+

Abstract

The climate crisis and the degradation of the world's ecosystems require humanity to take immediate action. The international scientific community has a responsibility to limit the negative environmental impacts of basic research. The **HECAP+ communities** (**High Energy Physics, Cosmology, Astroparticle Physics, and Hadron and Nuclear Physics**) make use of common and similar experimental infrastructure, such as accelerators and observatories, and rely similarly on the processing of big data. Our communities therefore face similar challenges to improving the sustainability of our research. This document aims to reflect on the environmental impacts of our work practices and research infrastructure, to highlight best practice, to make recommendations for positive changes, and to identify the opportunities and challenges that such changes present for wider aspects of social responsibility.

Version 1.0, 5 June 2023

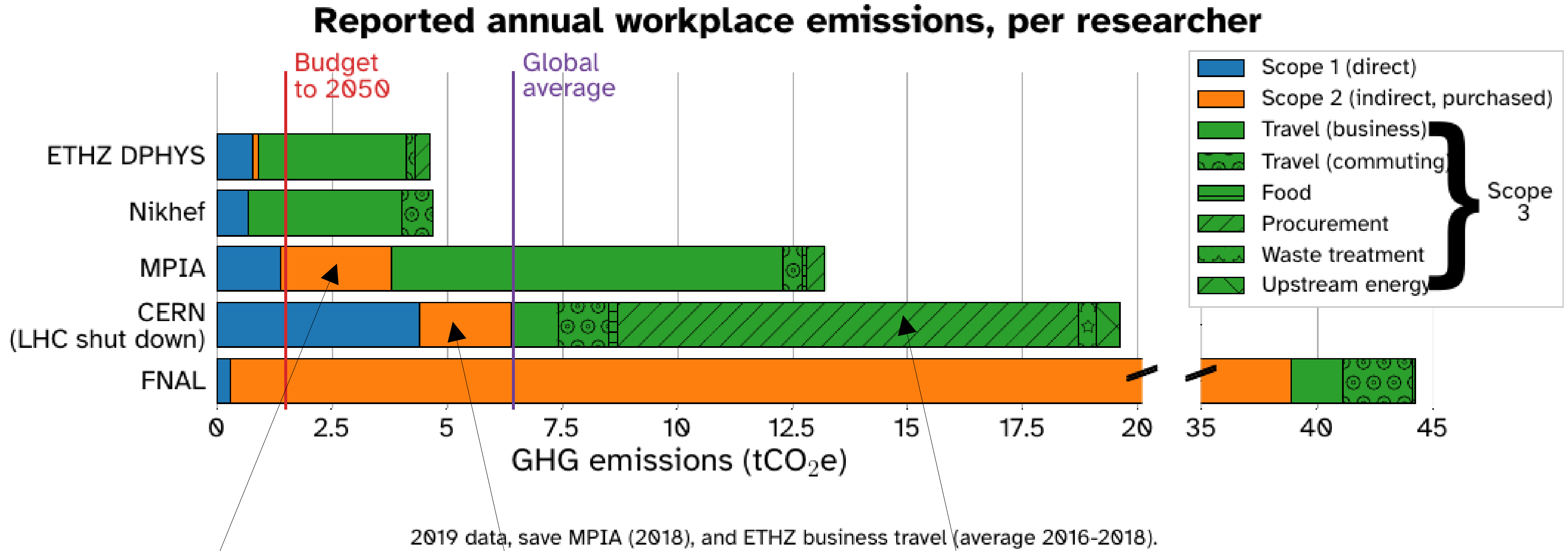
Please read this document in electronic format where possible and refrain from printing it unless absolutely necessary. Thank you.

https://sustainable-hecap-plus.github.io/

Computing in comparison

- **Workplace emissions in Physics / HECAP+**

Scope 1: gases
Scope 2: electricity

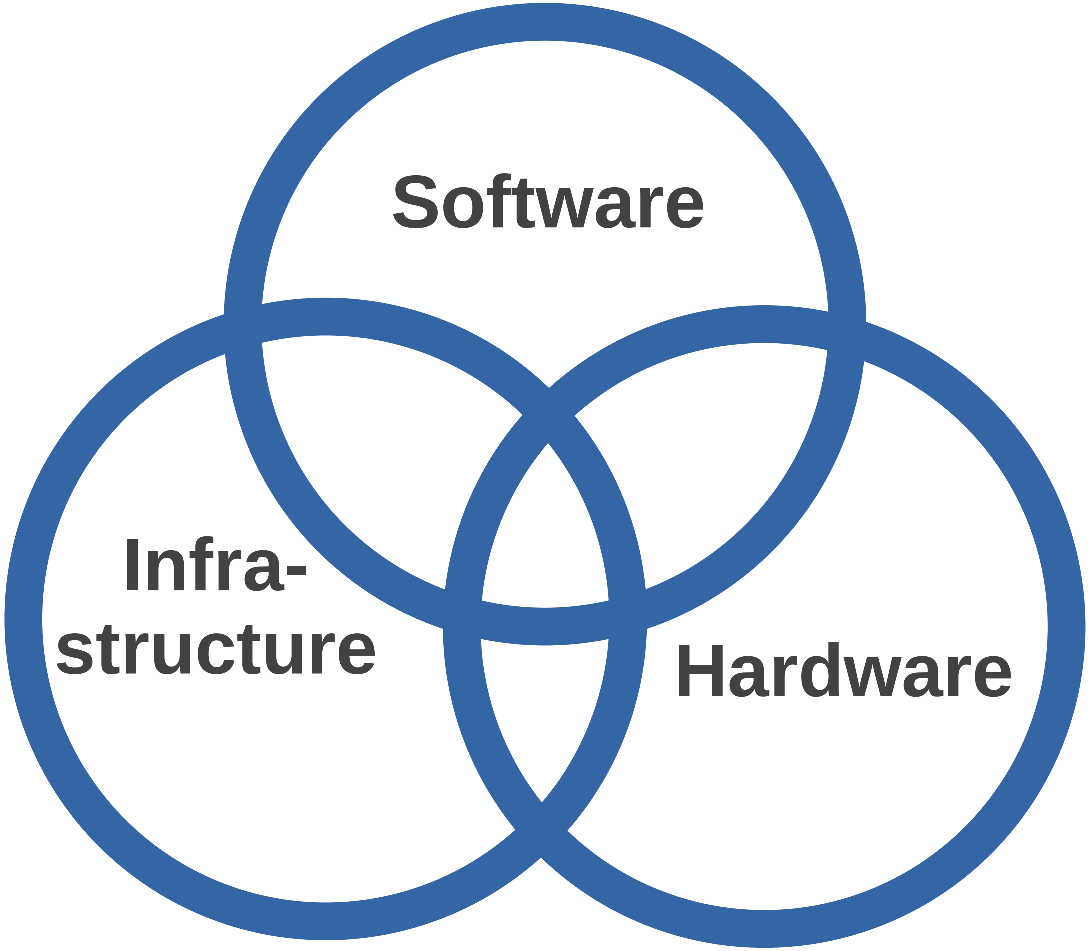


Max-Planck Institute for Astronomy:
88% of electricity is computing

CERN:
1/3 is data centre in Hungary

CERN:
some hardware procurement

Computing in a nutshell



”Classical sustainability” (Reuseability, Training)

Hardware

- **Manufacturing 50% - 80% of a devices CO2e footprint (server vs. laptop)**
  → Infrastructure to keep, reuse, recycle, repair! Extend use lifecycle
  → staff extensive, on the level of single institution
  → on larger scale (clusters): potentially complicated to organise (especially when moving old hardware to a different cluster)

- **Use of certified products (e.g. TCO certification, though that probably already covers most of hardware)**

- **'Energy proportionality' is important:**
  energy consumption should be proportional to computing performance over the full range of applications → hardware often most efficient at maximum performance load, but in practice often idle (combat with scheduling)

- → tests needed to find optimum usage, depending on architecture

- Potential in reducing clock frequency ~about same amount of HEP work at 94-98%

Rodney Walker:
https://indico.desy.de/event/37480/contributions/140510/attachments/82246/108365/Meinerzhagern_compOps(2).pdf

Infrastructure

- **Well managed, centralized systems key to address challenges**

  → Optimized PUE (=Power Usage effectiveness → Total Power/Energy used by IT)

  → Current **best centres:** 1.05-1.2 mainly due to heat recovery from cooling system for heating (HECAP+ examples: GSI green cude 1.07, CERN data centre: 1.5 (1.1 planned), Swiss National supercomputing (1.2 at 25% full load)

  → world average ~1.55, WLCG assumed 1.45

- **Centralization here helps, in particular to run hardware optimized for specific (HEP) applications**

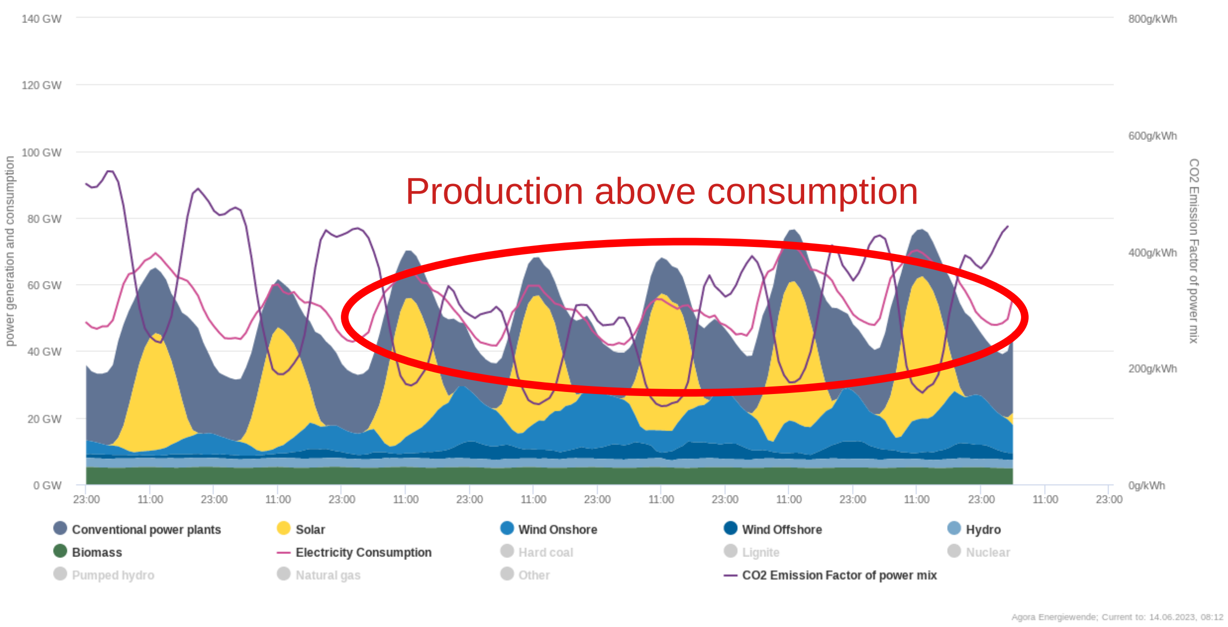
  → (HTC versus HCP which can make local resources difficult to use)

Infrastructure

- **Usage of carbon-free energy paramount**

  → “Own” production (requires investment into solar + potentially storage)

  → Regulation of load according to prices (“Follow the money” – R.W.), prices can be negative, but requires special tariff that can be used → well maintained data centres reacting to production and other grid loads, can help balance grid



Lancium Computing centre
https://indico.desy.de/event/37480/contributions/138296/attachments/82407/108618/2023-05-30%20Concrete%20Action.pdf

Software

- **HECAP+ Code relies on libraries and public codes, general frameworks and software infrastructure provided by experts in the experiments.**

- Strict requirements posed by the computing environment.

- Impact directly measureable e.g. cosmological analyses → using Likelihood Inference Neural Network Accelerator (LINNA) for efficiency could save **$300,000 in energy costs and around 2,200 tCO2** in first-year for Rubin Observatory’s Legacy Survey of Space and Time (LSST) analyses (https://dx.doi.org/10.1088/1475-7516/2023/01/016)

- **Dedicated efforts can have a huge impact!**

- Need sustained effort, continued recognition and dedicated and well-trained person power
→ need to use leverage with experiments, mechanisms to allow ***more*** people to make a career of these efforts **within** the field

“Classical” Software sustainability

- General sustainability => Re-useability and training

  → Institution for Research and Innovation in Software for High Energy Physics (IRIS-HEP) [44]

  → HEP Software Foundation

- **May provide an important platform for accelerating the inclusion of environmental considerations in software development. (examples e.g. are Sherpa speedup!)**
- Underwriting of FAIR principles: software (and data) should be Findable, Accessible, Interoperable and Reusable
- Sharing optimization workflows, **consulting services** for smaller experiments

Recommendations — Computing

Individual actions:

- Make sustainable personal computing choices by considering the necessity of hardware upgrades, the repurposing of hardware, and the environmental credentials of suppliers and their products.
- Assess and improve the efficiency and portability of codes by considering, e.g., the required resolutions and accuracy.
- Assess and optimise data transmission and storage needs.
- Follow best practice in open-access data publishing, prioritising reproducibility and limiting repeat processing.
- Read the section on E-waste (Section 7).

Further group actions:

- Right-size IT requirements and optimise hardware lifecycles.
- Schedule queueing systems with environmental sustainability in mind, so as to maximise the use of renewables, accounting for the geographical location of servers/data centres.

Further institutional actions:

- Ensure that environmental sustainability is a core consideration when designing and choosing sites for large computing infrastructure, such as data centres, including, e.g., the availability of renewables, the efficiency of cooling systems and the reuse of waste heat.
- Proceduralise the repair, upgrade and repurposing of existing computing, the de-inventorising of personal equipment for leaving personnel or for donation, and the responsible recycling of retired hardware.
- Select cloud computing services for their carbon emission mitigation policies.

Some of the above recommendations are based on those made by Jan Rybizki [34].

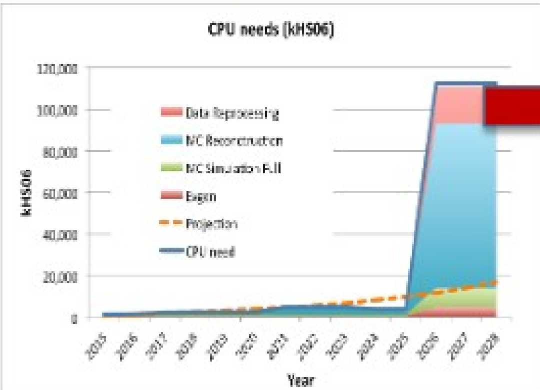
These recommendations are out of necessity most general

They are obvious → many have been already made these past couple of days

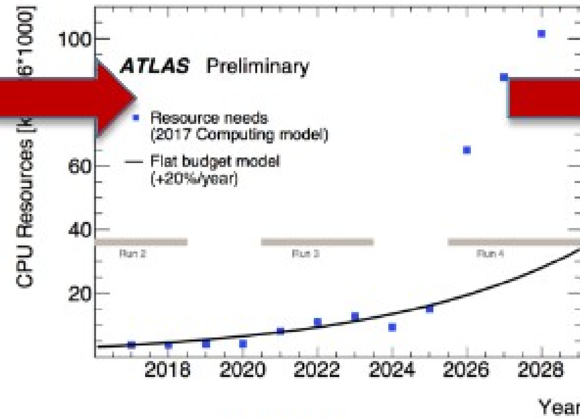
How can they be put into action (in particular institutional ones) given dependence on funding agencies, national laws, etc. ?

The positives

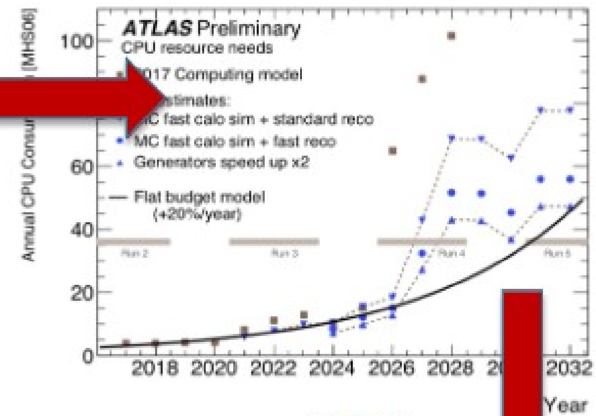
2015



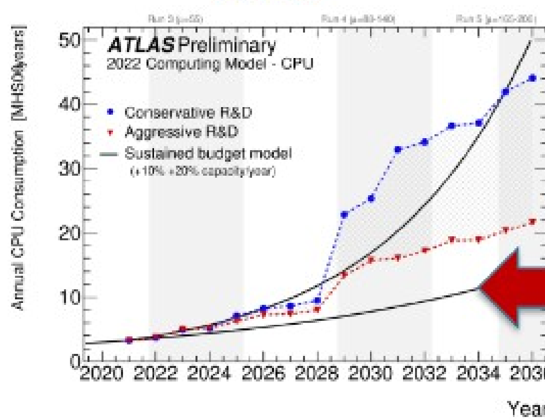
2017



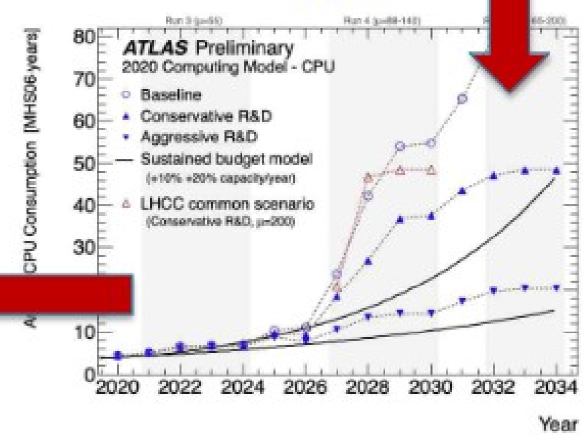
2018



2022



2020



2015 projections: resource needs = 10x more than budget allows

2022 projections: resource needs compatible with budget (optimistic scenario)

Simone Campana – CHEP: https://indico.jlab.org/event/459/contributions/11499/attachments/9236/14205/WLCGEnergyNeedsCHEP2023.pdf

The positives

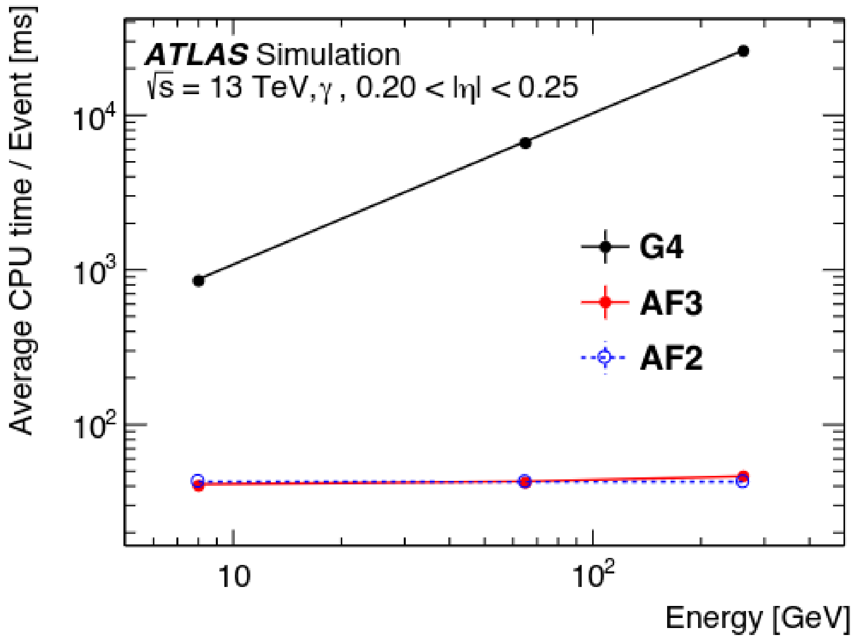
- **Factor 50 improvement in generation time for Sherpa generator**
  → Optimized using SWIFT-HEP software grant with software engineers
  → general MC background samples

- **ATLAS reconstruction code improves by a factor of 2 using multi-threading**
- **ATLFAST with GANs**

- 1 year ago CERN computing website still said: *HL-LHC is expected to rely on 50 to 100 times the computing capacity needed for LHC*
  → **reduced by factors > 5!**

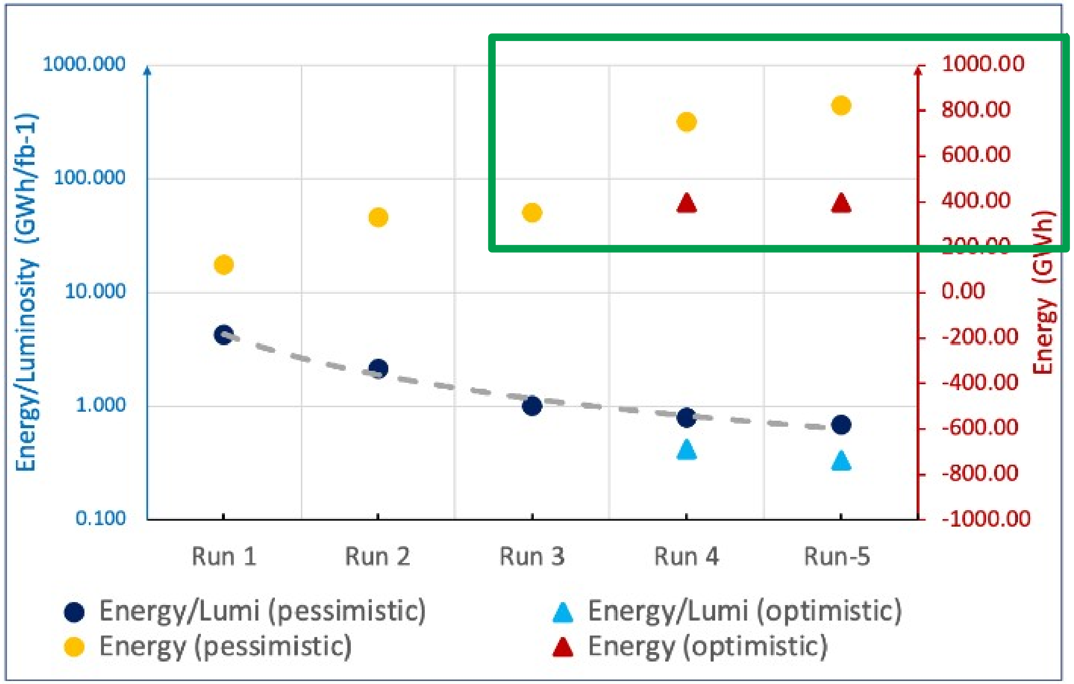
- **In some cases improvements by orders of magnitude!**

  → Potentially even more so for smaller experiments



Sherpa: https://doi.org/10.1140/epjc/s10052-022-11087-1
Multi-threading: http://cds.cern.ch/record/2771777
ATLAfast3: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/SIMU-2018-04/

The negatives



WLCG Energy needs in Run-4 and Run-5:

+100% compared to Run-2 in the pessimistic scenario,

only +10% in the optimistic scenario

(again: S. Campana, CHEP 2023)

Is this compatible with a factor of 2 overall CO2e reduction? Rebound effects?

Factors in improvements in infrastructure (CERN new data centre)

Unclear if it accounts for hardware manufacture (remember: up to 50% embedded)
→ CO2e savings here rely to a big extend on manufacturers

Some conclusions

- **We (as a community) have made big progress and substantial improvements**
  (considering the constraints potentially as much as e.g. google/amazon)

- **But is it enough to achieve 50% overall reduction of CO2e?**

- **3 handles:**
  Green energy → factor of 12
  Energy efficiency → factor of 2
  Energy saving → factor of 2

- **Will need a hard look and many, many sacrifices (not only in the computing sector)**

- **Will require a concerted effort and dedicated funding**
  → but as a community we are certainly better placed than other fields of science (which are/will also come under scrutiny)

Need framework with benchmarks and goals and
Ability to *shape* (institutional/funding) constraints to allow achieving goals

Climate benchmarks that need to be met
(→restricted physics exploitation scenarios, what can we sacrifice?)

Some questions

- **Infrastructure:**
  → Further centralize?
  → Possibility for own power generation? (→saves money, but are those eligible costs?)
  → tools for power management, using negative tariffs, “follow the money”
  → Support for small local cluster → easy to handle tools to allow this for small sites?

- **Hardware:**
  → Keep old hardware to run as backup / addition at high-energy production times
  → How to manage old hardware (distribution to other sites possible?)
  → study interplay hardware/software
  → High-performance versus High-throughput (HPC versus HTC)

- **Software:**
  → Needs dedicated efforts with possibility to retain people
  → ability to make this a career
  → ability for the end user to monitor
  → training, training, training

(said during the talk, but not in the original presentation slides shown):

Do we need contingency plans on restricting physics?

Reminder: Paris agreement is in principle legally binding
→ pressure on us / our savings might need to be increased
→ gives us negotiating power if we have a clear plan and strategy with demonstrable impacts and realistically achievable objectives in line with 1.5°C

Thank you