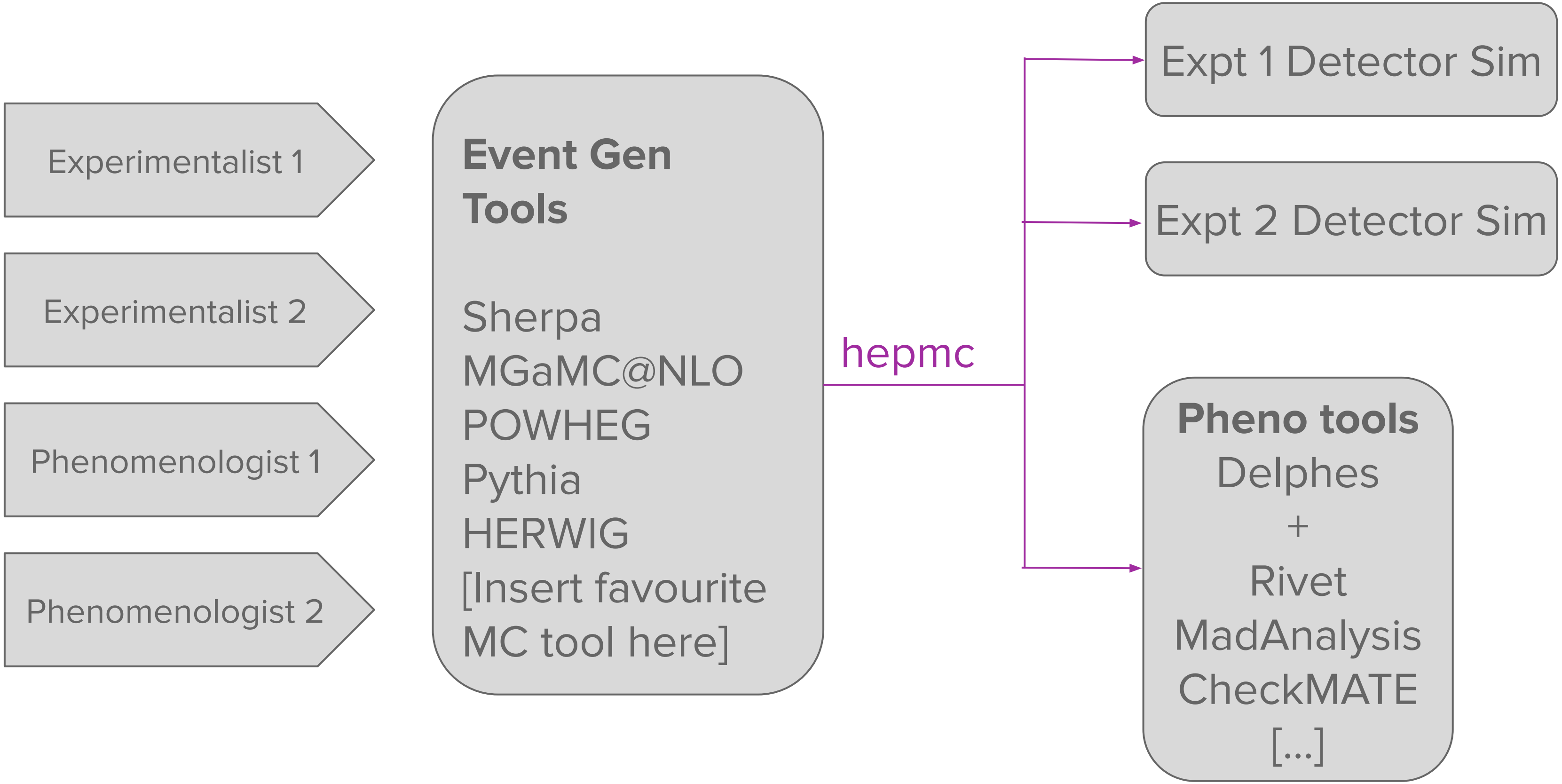
Open Monte Carlo Event Generation

Rakhi Mahbubani
Rudjer Boskovic Institute and [Sustainable HECAP+]

Based on [arXiv:2605.12229] with the **LHC Reinterpretation Working Group**

MC Event Generation



Financial Cost

Cost estimate for generation and storage of ca. 13B HepMC events.

Scenario	Costs [kCHF]			
	Compute	Storage	Egress	Persistence (5 years)
CERN Tier 0	185	90	–	(incl.)
Google Cloud	1245	336	0.036/TB-month	841
Google Public Sector	270	69	(incl.)	173

Table 1: Cost summary for the replication of ATLAS Open Event Generation (900 TB produced over two years) and storage for five years. Costs expressed in thousands of CHF (kCHF).

Total ATLAS annual event generation is a factor of 10 larger (similar for CMS).

HL-LHC will require another 5-10x increase

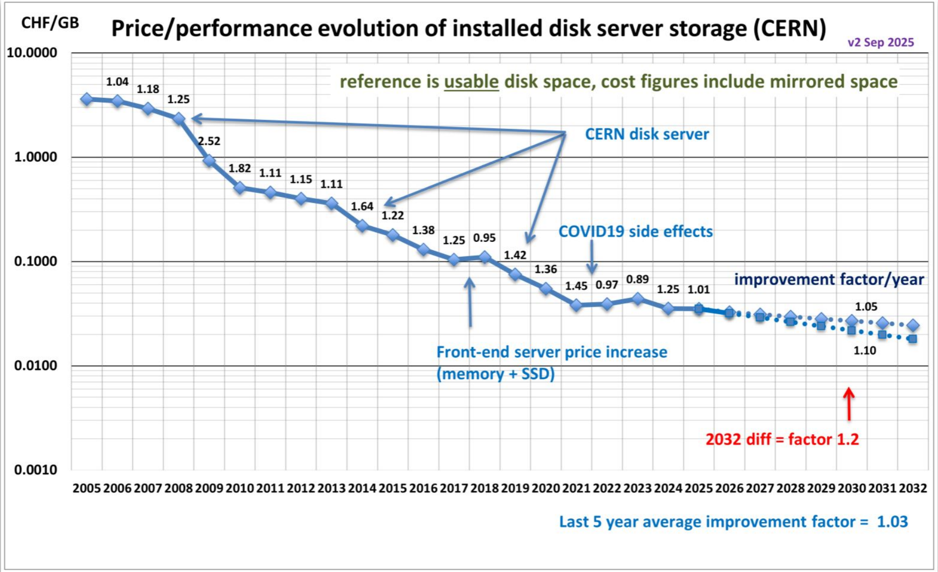
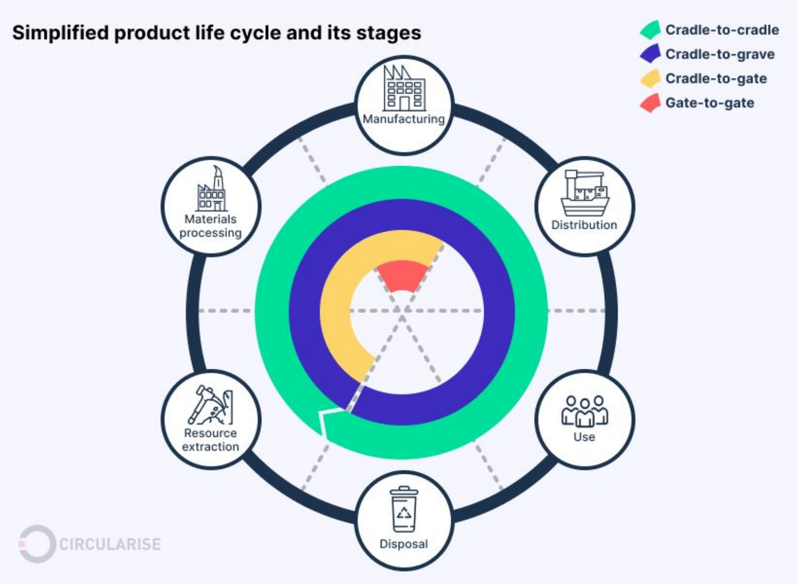


Figure 4: Evolution of price/performance for installed disk server storage at CERN (2005–2025), measured in CHF/GB usable space (including mirrored space). The projections exclude recent RAM and storage price fluctuations from AI demand. Figure taken from [123].

Environmental Costs

More than just GHG emissions.



Output	Quantity	Output	Quantity
monocrystalline Si wafer	0.70 kg	**Emissions to air**	
Impact Assessment		particulates (PM10)	1.71E-2 kg
energy renewable	213 MJ	SO_2	0.137 kg
energy total	4850 MJ	CFC-11	1.3E-5 kg
GHG emissions	319 $kgCO_2e$	PFC-14	5.17E-6 kg
acidification	0.369 $kgSO_2e$	PFC-116	6.50E-7 kg
Waste		dioxins	1.18E-11 kg
FGD residues (landfill)	2.79 kg	lead	1.51E-5 kg
production waste	2.32 kg	acid (as H+)	4.60E-4 kg
sewage sludge	4.89E-3 kg	nitrogen oxides	0.309 kg
ash	11.0 kg	cadmium	9.8E-7 kg
waste (landfill)	1263 kg	nickel	2.72E-5 kg
nuclear waste	3.97E-4 kg	ammoia	9.66E-4 kg
Emissions to water		nitrous oxides	8.89E-3 kg
lead	3.28E-10 kg	mercury	3.24E-6 kg
phosphorus	9.24E-7 kg	fluorides	8.65E-4 kg
nitrogen total (excluding N_2)	5.45E-5 kg	polycyclic aromatic hydrocarbons	9.38E-10 kg
mercury	2.51E-11 kg	chlorides	1.47E-2 kg
chromium	3.28E-6 kg	H_2S	6.08E-6 kg
chemical oxygen demand	0.352 kg	CO_2 (fossil)	299 kg
biological oxygen demand	9.87E-3 kg	CO (fossil)	6.27 kg
arsenic	2.06E-11 kg	CH_4 (fossil)	0.571 kg
adsorbable organic halides	8.68E-8 kg	other volatile organic compounds	1.79E-2 kg
inorganic salts and acids	0.014 kg	arsenic	3.49E-6 kg

Image c/o Circularise

2005 data for production of Si wafer in DE arXiv:2306.02837

GHG Emissions

Total emissions/HepMC event = Use phase emissions + embodied emissions for hardware + housing infrastructure

$$\text{Compute [CPUs]} \times \text{Hardware power [kWh/CPUs]} \times \text{grid carbon intensity } [\mathrm{gCO_{2e}/kWh}] \times \text{PUE}$$

$$0.015\,\mathrm{gCO_{2e}} \left(\frac{\text{Efficiency}}{15\,\frac{\text{HS23s}}{\text{CPUs}}} \right)^{-1} \left(\frac{\text{Power/CPU}}{10\,\text{W}} \right) \left(\frac{\text{PUE}}{1.45} \right) \left(\frac{\text{grid carbon intensity}}{245\,\mathrm{gCO_{2e}/kWh}} \right)$$

Total emissions/HEPMC event = Use phase emissions + embodied emissions for hardware + housing infrastructure

$$\text{Compute [CPUs]} \times \text{Hardware power [kWh/CPUs]} \times \text{grid carbon intensity } [\text{gCO}_{2\text{e}}/\text{kWh}] \times \text{PUE}$$

$$0.015\,\text{gCO}_{2\text{e}} \left(\frac{\text{Efficiency}}{15\,\frac{\text{HS23s}}{\text{CPUs}}} \right)^{-1} \left(\frac{\text{Power/CPU}}{10\,\text{W}} \right) \left(\frac{\text{PUE}}{1.45} \right) \left(\frac{\text{grid carbon intensity}}{245\,\text{gCO}_{2\text{e}}/\text{kWh}} \right)$$

Total emissions/HEPMC event = Use phase emissions + embodied emissions for hardware + housing infrastructure

$$0.0033\,\text{gCO}_{2\text{e}} \left(\frac{\text{Efficiency}}{15\,\frac{\text{HS23s}}{\text{CPUs}}} \right)^{-1} \left(\frac{\#\text{ cores}}{64} \right)^{-1} \left(\frac{\text{embedded emissions}}{1648\,\text{kgCO}_{2\text{e}}} \right) \left(\frac{4\text{ years}}{\text{lifespan}} \right)$$

Benchmark, server only.

GHG Emissions

Estimated annual emissions for ATLAS MC event generation:

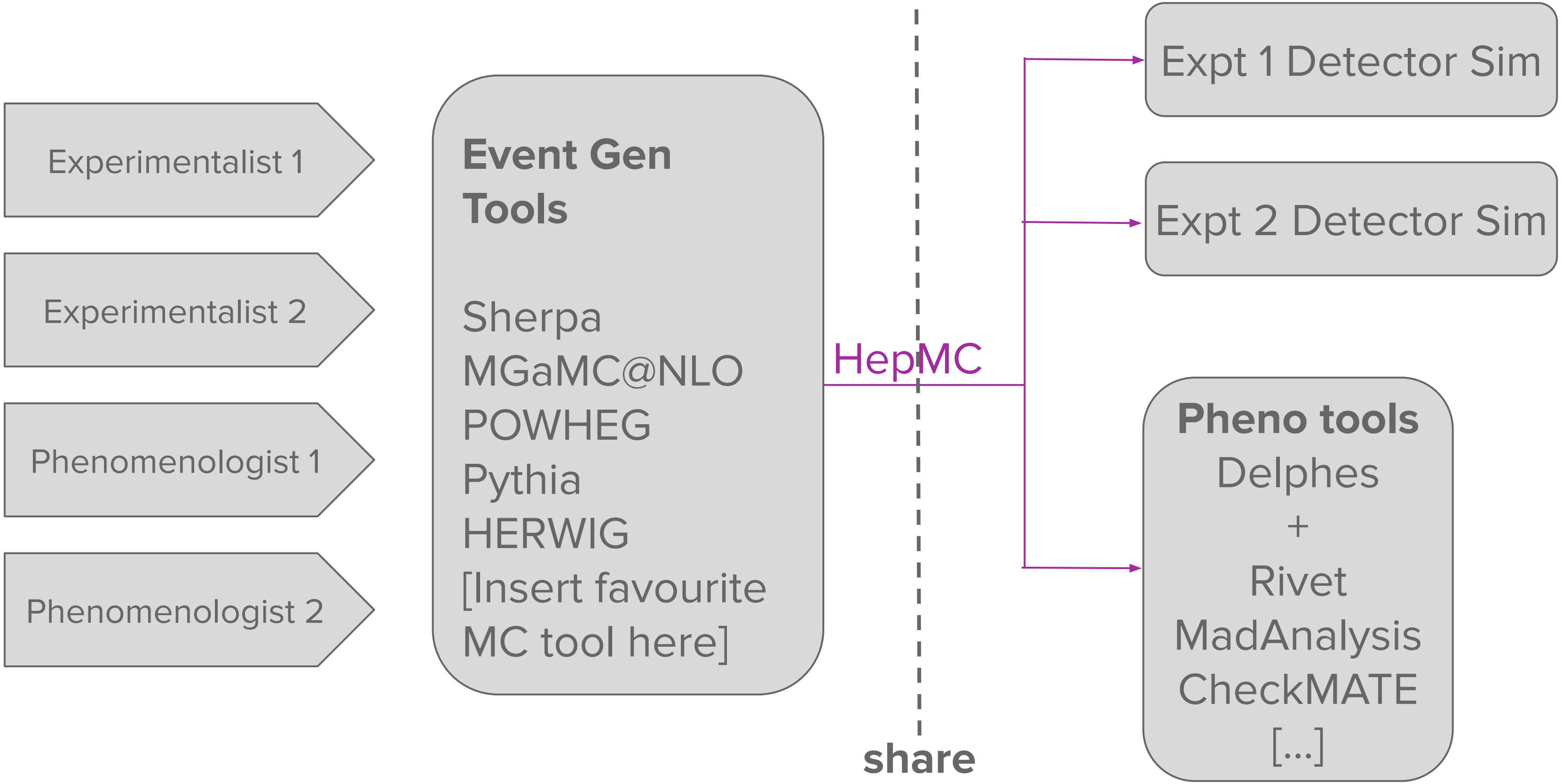
1950 tCO_{2e} (use) + 420 tCO_{2e} (embedded)

Equivalent to >1000 return transatlantic flights*

Similar order of magnitude for CMS

* For single passenger, economy class, average occupancy.

MC Event Generation



Open MC Event Generation Data

CMS

>50,000 samples of simulated + reconstructed SM and BSM events released using various generators ($\sqrt{s}$ = 7, 8, 13 TeV).

Newer simulation released as ‘abridged’ NanoAODSIM format, stores partial event gen information.

Findable on the CERN Open Data Portal through high-level metadata.

ATLAS

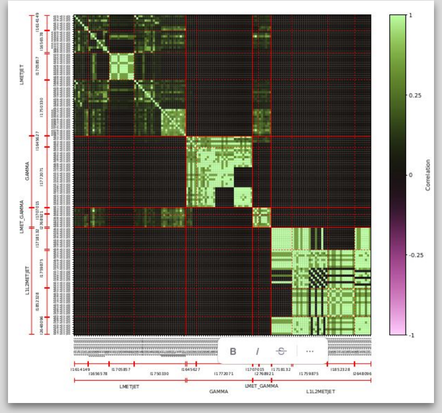
In addition to simulated + reconstructed data, dedicated release of 14B events of event generator output at $\sqrt{s}$ =13, 13.6 TeV.

Dedicated events in industry-standard HepMC2 format.

Downloadable using atlasOpenMagic tool, and (partially) on CERN Open Data Portal

Use cases

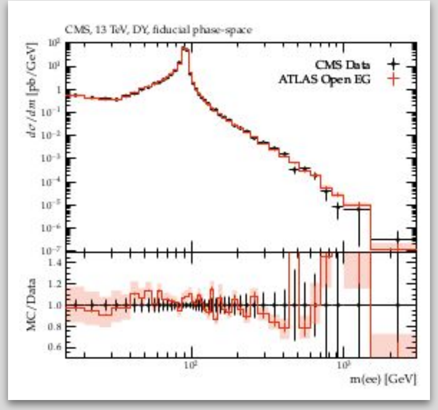
Studying overlaps across signal regions



Comparison between experiments

Designing BSM searches

Reuse/Reinterpretation of SM measurements



In future:

Unfolding for e.g. jet reco studies

Machine learning studies

Training and education

Future directions

Sharing parton-level LHE events

Strong case for storing LHE info as text or HDF5 and sharing as Open Data

150x smaller than HepMC (negligible extra cost)

Good fit for HPC/GPU hardware

Upgrade particle-level simulation using stored LHE

Allows study of nonperturbative modeling

Requires sufficient metadata for correct use

HEPData accessibility

Link analyses to datasets, whether open or internal

During publication, link datasets using unique DOIs or identifiers

Modest extension of HEPData interface

Central Analysis Facilities

Near-data” computing is a huge boost for analysis

Many existing platforms prove the principle

Ideal structures would:

- Use standard workflow management tools
- Implement dashboards for progress monitoring
- Interact with CERN Open Data Portal tools

Action items

Short-term	Mid-term	Long-term
Improve what we have!	**Adopt new techniques**	**Centralize LHC MC production**
Better LHC Open Data event gen materials and tools	Engage experiments beyond CMS & ATLAS	Core team working with generator experts and experiments
Harmonize metadata and CERN Portal usage across experiments	HEPData dataset linking as community standard	Common formats • HL-LHC effective lumi • Open Data Portal storage
Enable phenomenology tools to utilize LHC Open Data	Wider LHE or HepMC releases by experiments	
Reduce size of HepMC-3	Central Analysis Facility development	

Requires mutual work by experiments, laboratories, and computing groups for success!

To take home:

Open sharing of event generation can give rise to substantial financial savings and decreased environmental impact, without negative physics impact.

There are co-benefits in transparency, reproducibility and equal access.

Ensuring the persistence and expansion of Open Event Generation offerings will require a commitment to provide dedicated infrastructure and human resources. Community uptake can help make the case for this!

The largest potential savings would come from establishing shared event generation across the LHC experimental - nontrivial. Talks ongoing within LHC Monte Carlo Working Group, e-group sign-up here: https://e-groups.cern.ch/e-groups/EgroupsSubscription.do?egroupName=lhc-mcwg

LHC Monte Carlo WG subgroup meeting – data sharing and new workflows

Thursday 18 Jun 2026, 15:30 → 16:55 Europe/Zurich

Enrico Bothmann (CERN)

What will your new use case be?

Thanks

to the LHC Reinterpretation WG Open Event Generation Task Force, in particular Julie Hogan, Sabine Kraml and Zach Marshall for their enthusiasm and hard work.

Open LHC Monte Carlo Event Generation

Enrico Bothmann[1], Jon Butterworth[2], Shu Chen[3], Eda Erdogan[1,4], Giovanni Guerrieri[1], Christian Gütschow[5], Martin Habedank[6], Julie M. Hogan[7⋆], Venus Keus[1,8,9], Sabine Kraml[10†], Van Dung Le[9,11], Kati Lassila-Perini[9], Rakhi Mahbubani[12‡], Zach Marshall[13§], Thomas McCauley[14], Tomasz Procter[15], Humberto Reyes-González[16], Andrzej Siodmok[15], Mariana Vivas Albornoz[17]
(LHC REI WG Open Event Generation Task Force)

to the members of Sustainable HECAP+ for their encouragement and support.

This work was supported by several funding agencies, for a full list see arXiv.2605.12229.