CMS O&C

Offline Software and Computing

CMS resource planning/HPC snapshot towards HL-LHC

L. Sexton-Kennedy (FNAL) - Event generators' and N(n)LO codes' acceleration - November 13, 2023

Outline

CMS Current Resources - Grid & HPC

CMS Future HPC - a guess

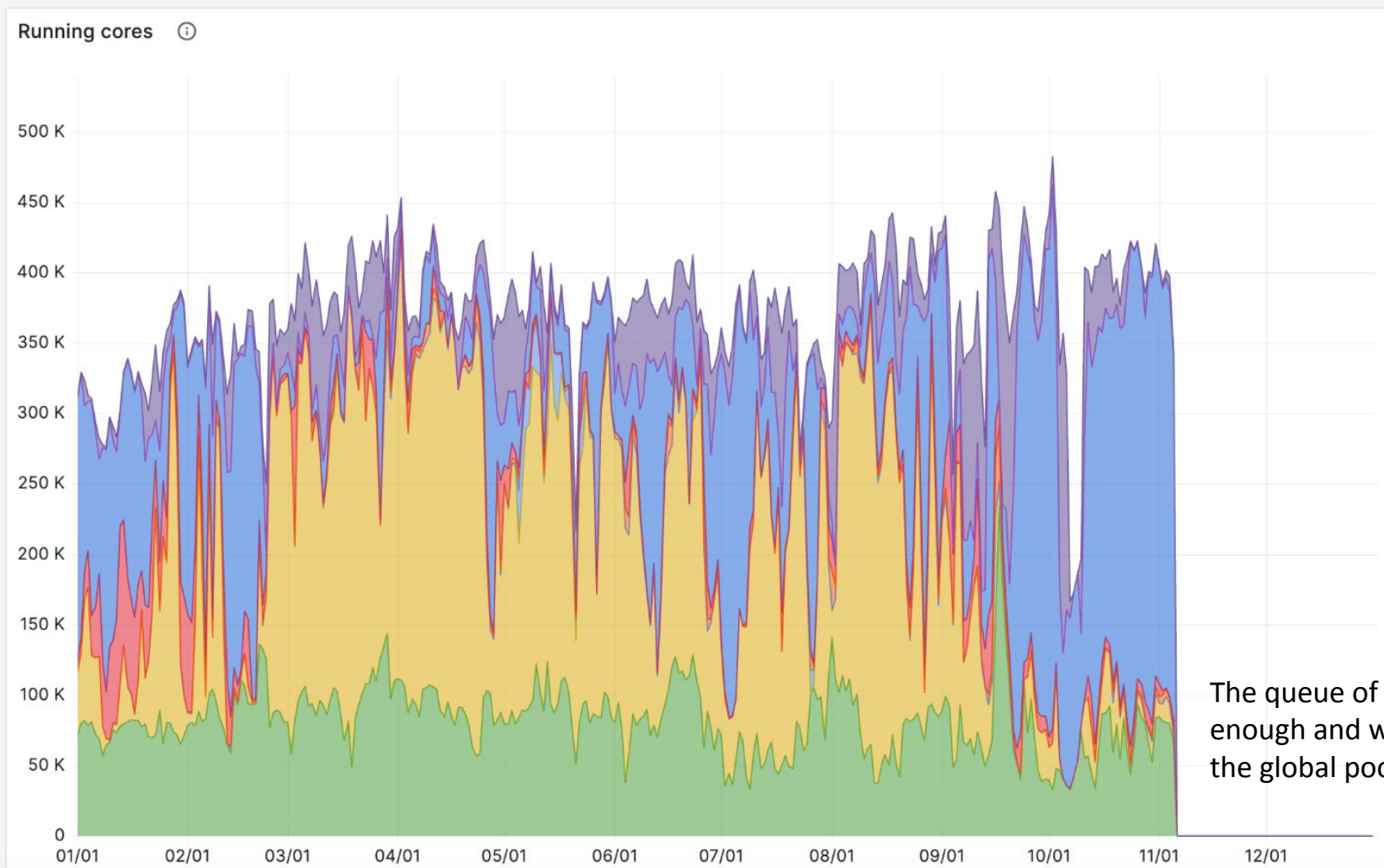
Generator resources as a fraction of the total in HL-LHC

Generator usage in CMS now and a guess for HL-LHC

Computing Resource Utilization

CPU Utilization by Type

- CMS uses about 400K cores a day for data processing and MC, organized into a global pool
- While analysis is a predictable baseline MC campaigns and data processing/reprocessing less so



	Mean	Last *	Max	Min	Total
Analysis	69.5 K	0	234 K	0	25.4 Mil
MC Ultralegacy	113 K	0	319 K	0	41.1 Mil
Phase2 requests	1.88 K	0	46.5 K	0	688 K
RelVal	2.18 K	0	40.5 K	0	796 K
Run2 requests	12.4 K	0	112 K	0	4.51 Mil
Run3 requests	88.2 K	0	386 K	0	32.2 Mil
SnowmassWinter21	0.00142	0	0.517	0	0.517
UNKNOWN	28.1 K	0	227 K	0	10.3 Mil
TOTAL	315 K	0	483 K	0	115 Mil

Variability in they types of work we need at any moment make it harder to use HPC allocations that require a physics case and “when will you need it” commitment.

The queue of analysis jobs is always deep enough and we set the fair share to 25% of the global pool.

HPC Utilization

- CMS uses HPC CPU queues on an opportunistic basis.
- Major US and EU sites have been integrated and used in production over the past 3 years
- CMS is actively investigating integration of new HPCs in Europe.

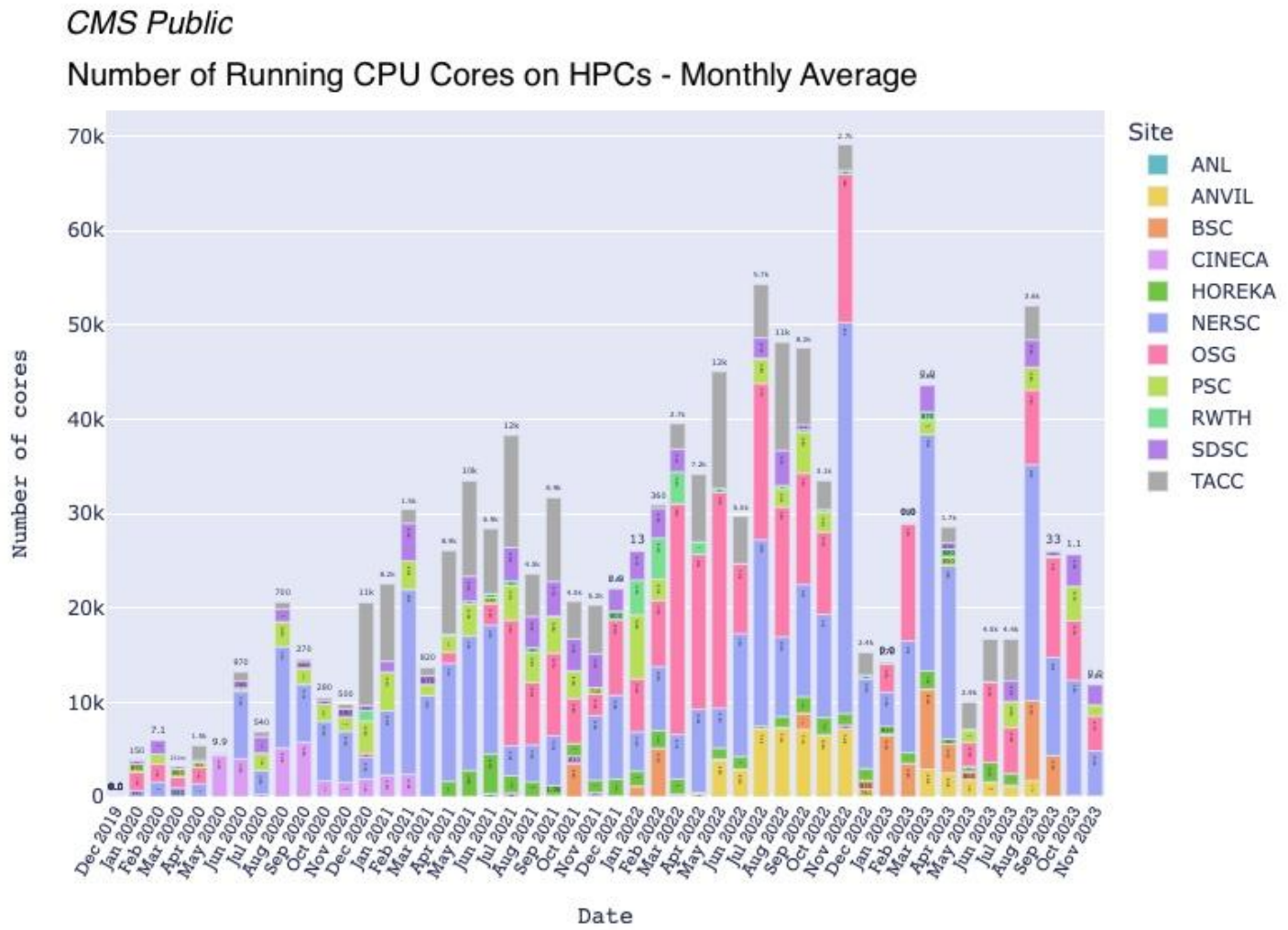
Note the reduction in cycles between 2022 and 2023.

At this scale, no real problems integrating HPC and LHC infrastructure.

CVMFS, temp data storage, network streaming for aux data, all operate on edge computing

Usage is down because we have less appropriate step-chain workflows to run now.

From [CMS Computing Results]



CMS Future HPC - a guess

Future directions for HPCs are different for different funders:

1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE
2	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu
3	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE
4	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64 GB, Quad-rail NVIDIA HDR100 Infiniband, Atos
5	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM

US-NSF: Aims to maximize science supported per $. Pays attention to data intensive science as well as high compute complexity users.

=> expect they will continue to provide easy to use HPCs like TACC & SDSC

US-DOE: Wants their HPCs to be able to solve problems that no other facility can

=> Exascale Computing Project - Frontier, Aurora

=> Heavy use of GPUs to get to the exescale

EuroHPC: Obtaining allocations is easy if you have a local group in your experiment, but more challenging when you don't

Resource Extrapolation

Generator Use on CMS Resources

CMS full chain HL-LHC simulation workflows are projected to be dominated by reconstruction **if R&D** efforts, including GPU acceleration fail to be integrated [*].

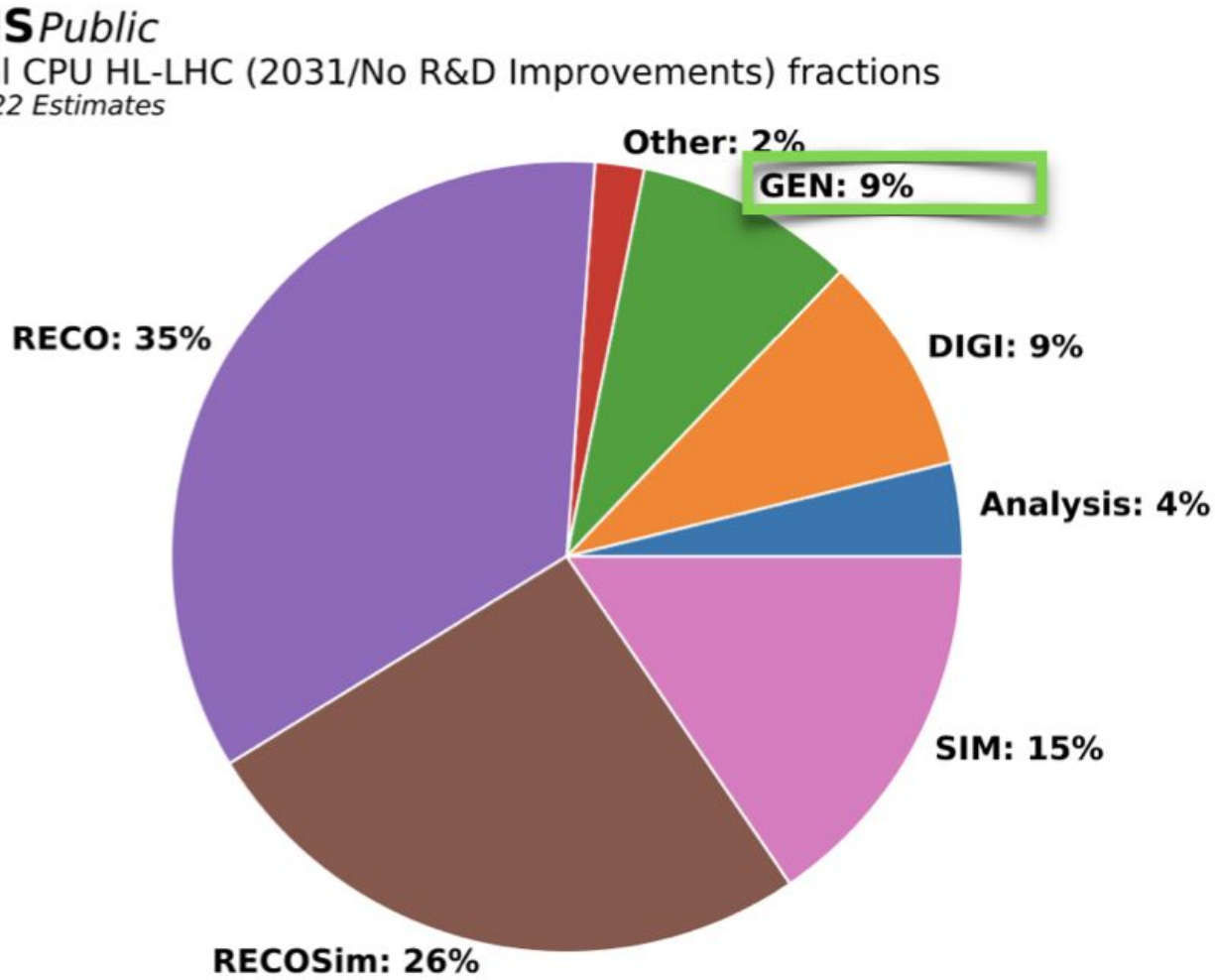
Tracking dominates the reconstruction time; our R&D aims to shrink it.

The HLT is already using GPUs for ~40% of the reconstruction code so failure by 2029 seems very unlikely.

If generator / simulation codes can not be accelerated, their fraction of this pie will grow.

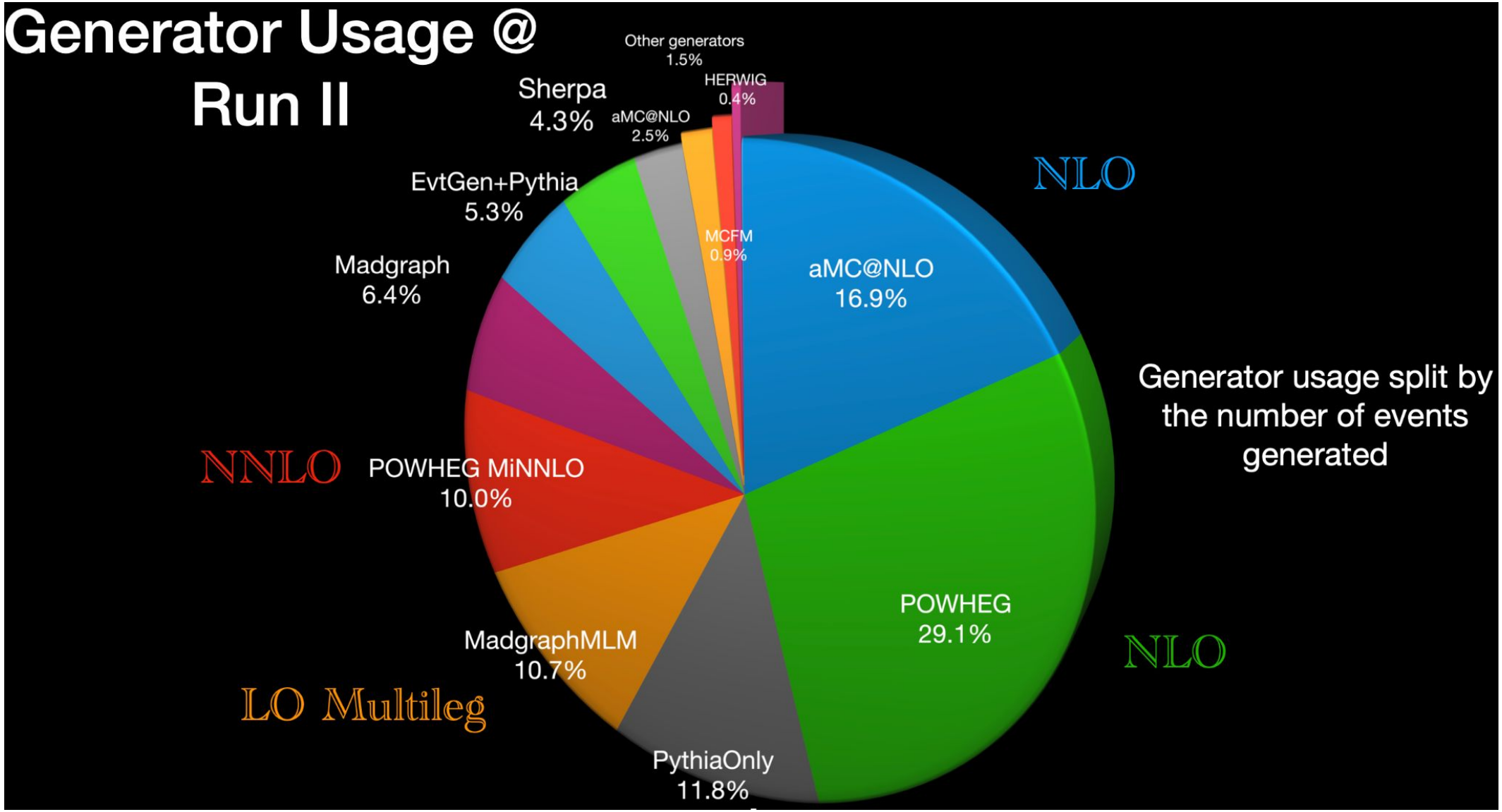
9% generation scales up the mix of generators used in Run 2.

[* CMS Phase-2 Computing Model: Update Document]



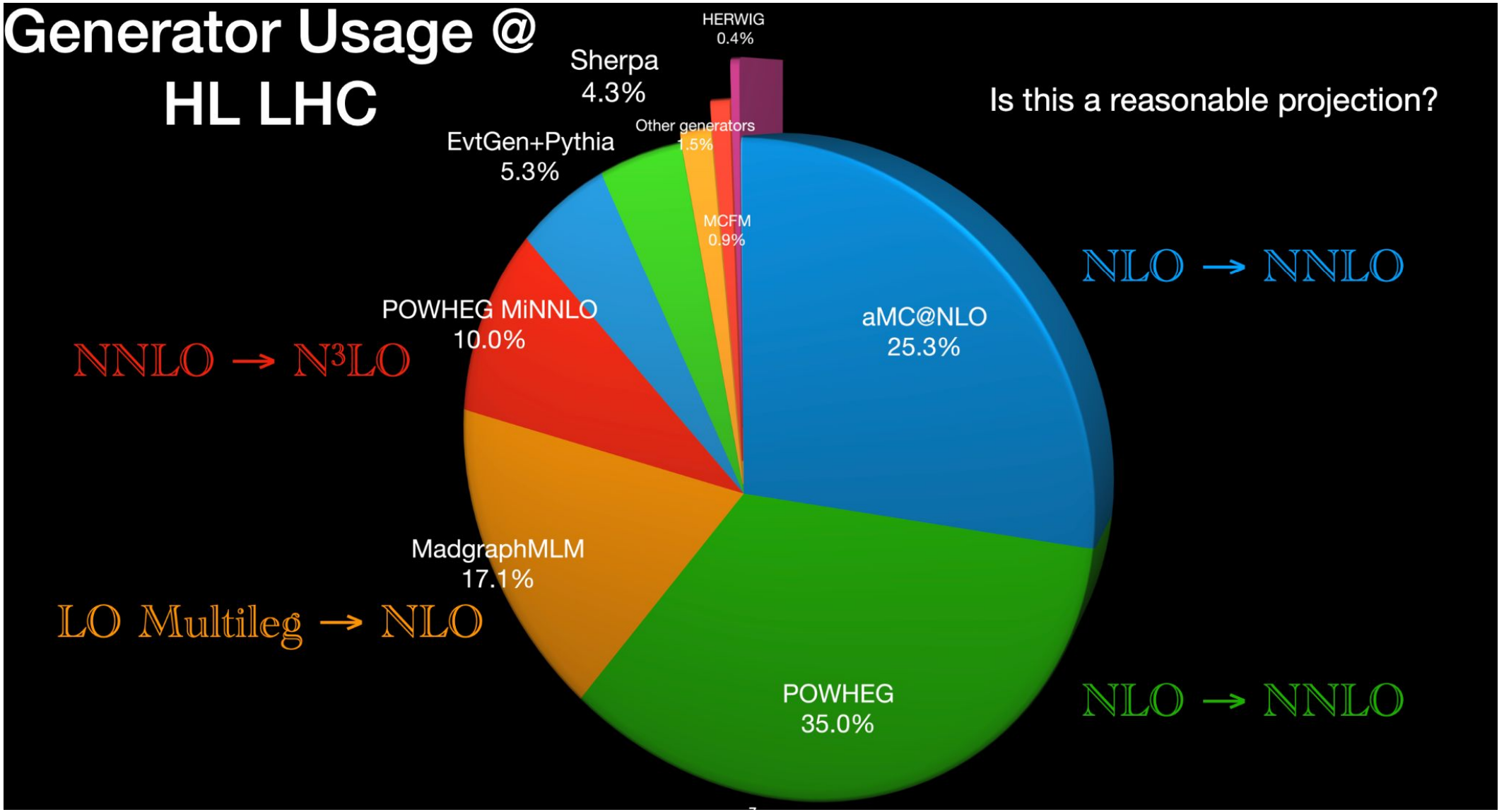
How Much Might the Mixture Change?

What happens if we just bump everything we did in Run2 up by one order in α_s?



How Much Might the Mixture Change?

What happens if we just bump everything we did in Run2 up by one order in α_s?



Observations

Run 2 Analysis at the precision frontier (Wmass) are already finding a real need for expensive higher order generation, which provides a better handle on uncertainties or a better description of data.

In the Wmass example a MiNNLO (arXiv:1908.06987) sample was generated at next-to-next-leading-order (NNLO) accuracy + parton shower (NNLO+PS). We generated 4 to 5 billion events at a cost of 56sec/event which is more then the cost of running the full geant simulation for a ttbar event. This processing is dominated by calculating PDF weights, with lots of redundant calculations, a waste of resources including carbon footprint.

I would love to hear from the people at this workshop, the answer to the reasonableness question.

Without guidance on the type of generators needed for HL-LHC we could be making a large error in our resource needs extrapolation. Not to mention which generators it is most important to accelerate!