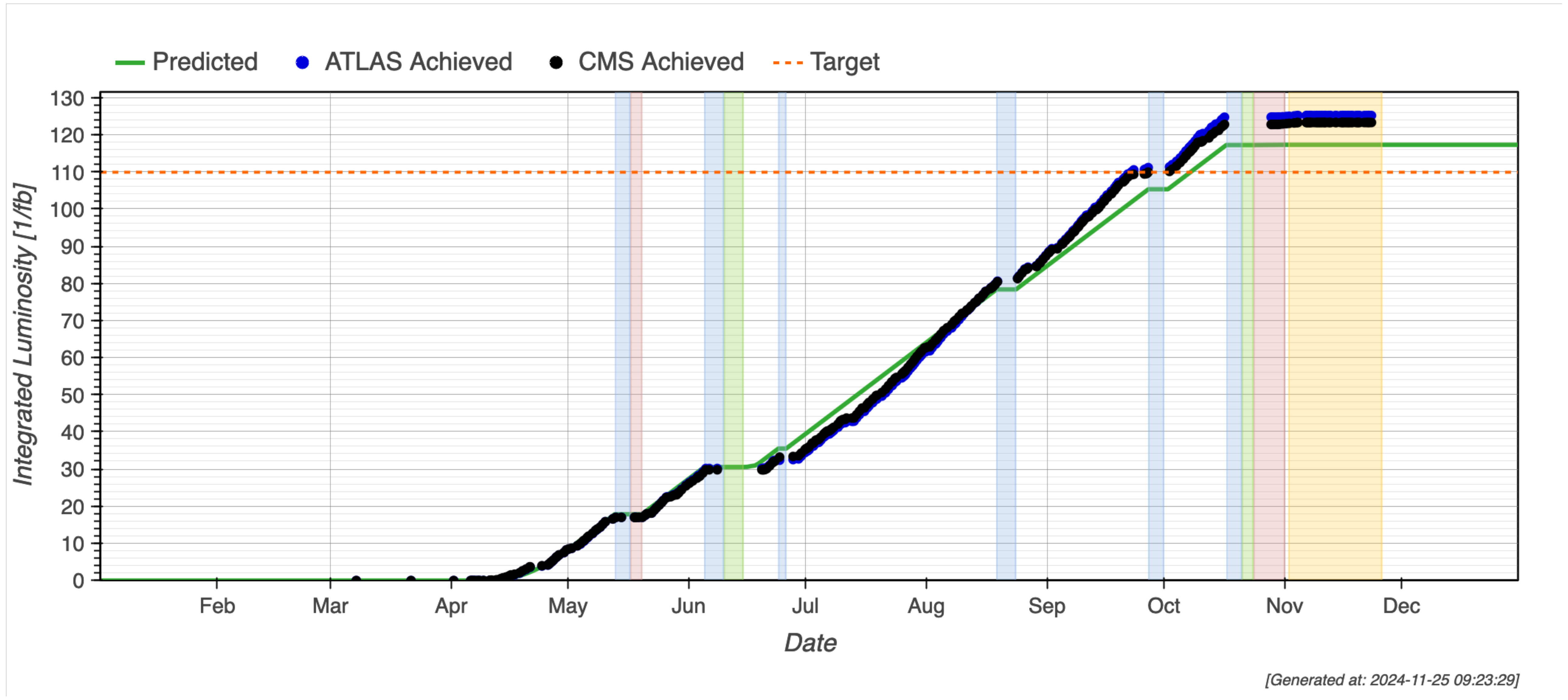
CMS Status

Matthew Nguyen

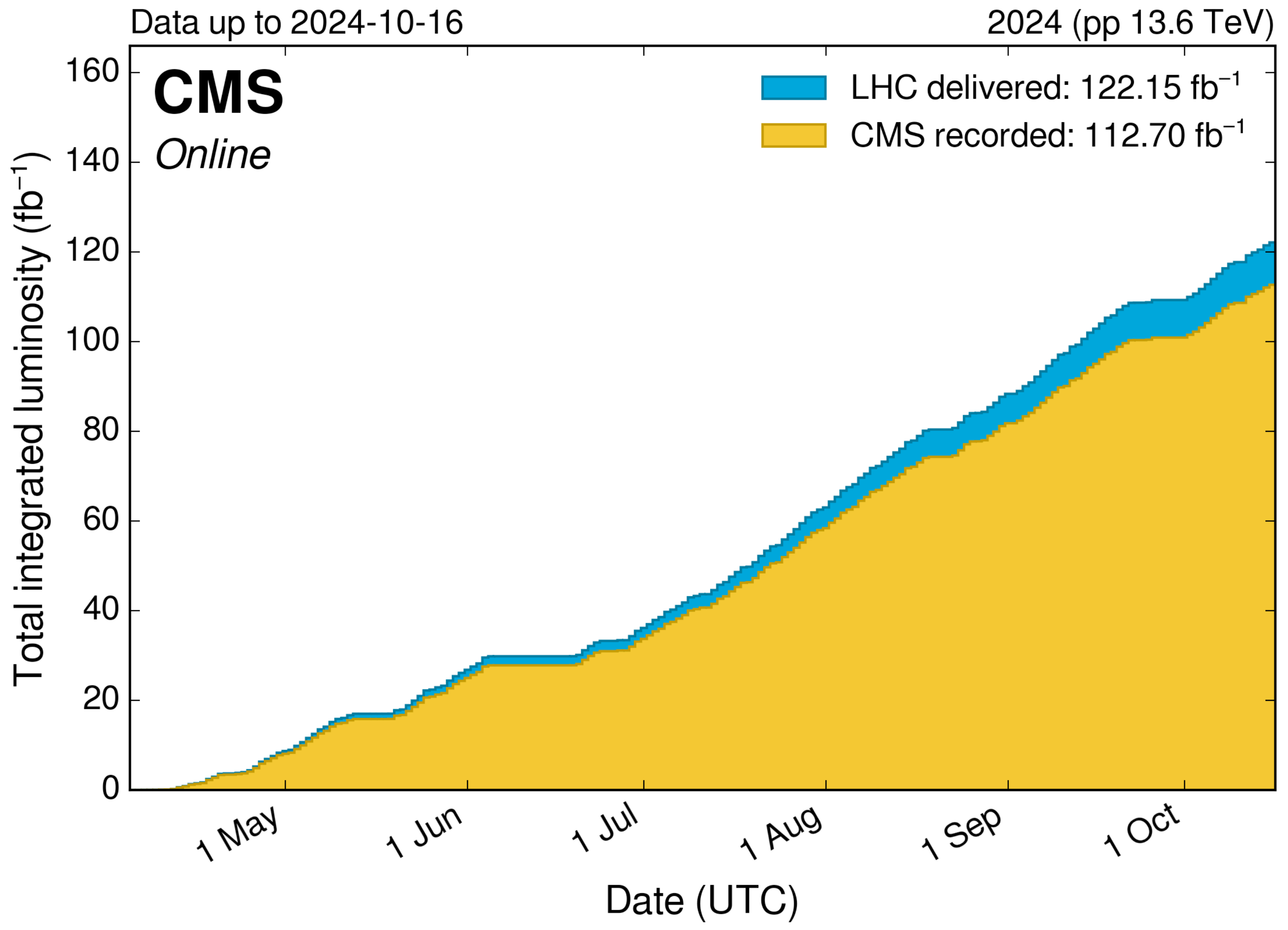
Journées LCG-France

December 5th 2024

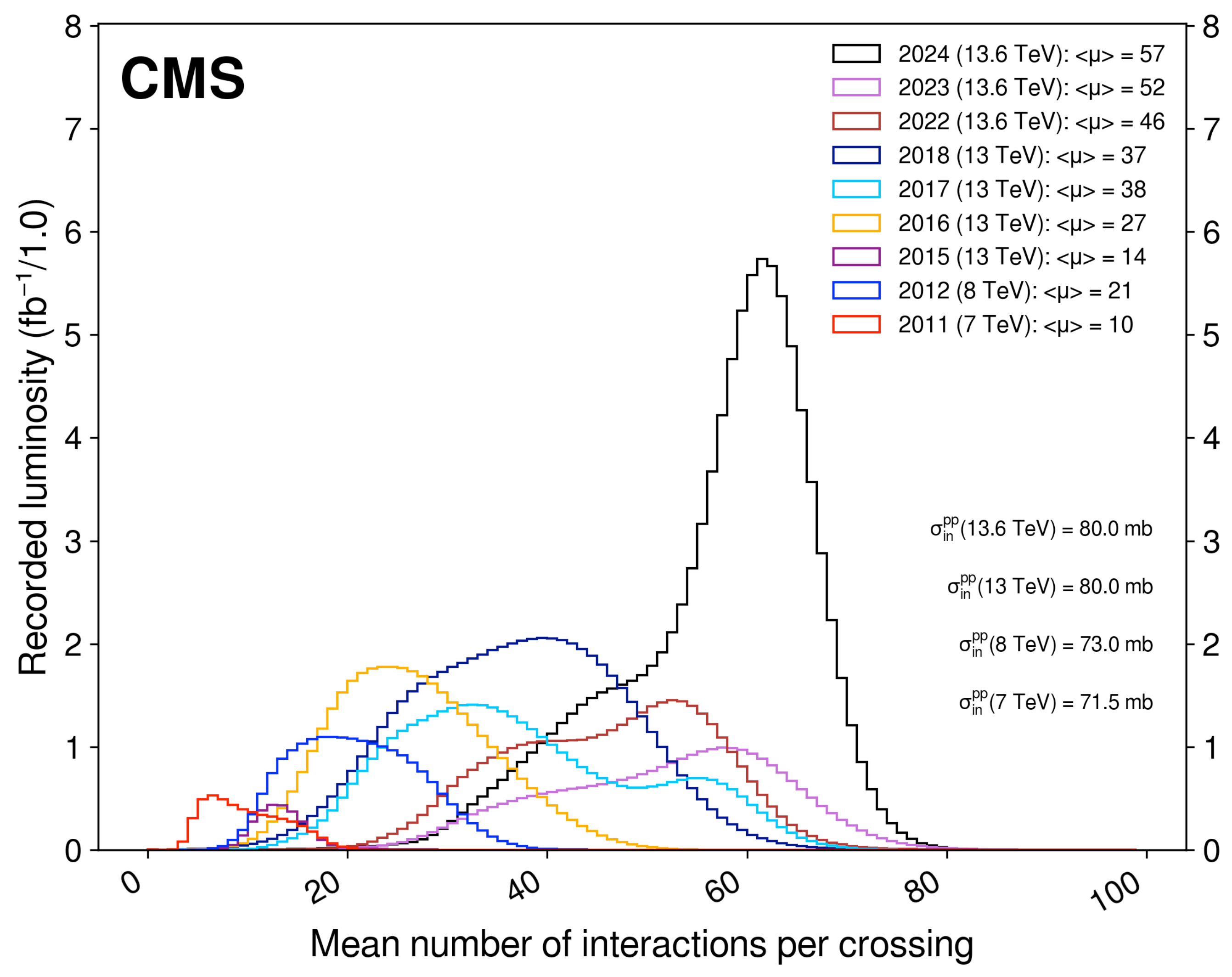
2024 was a record breaking year for the LHC!



Data taking efficiency > 90%

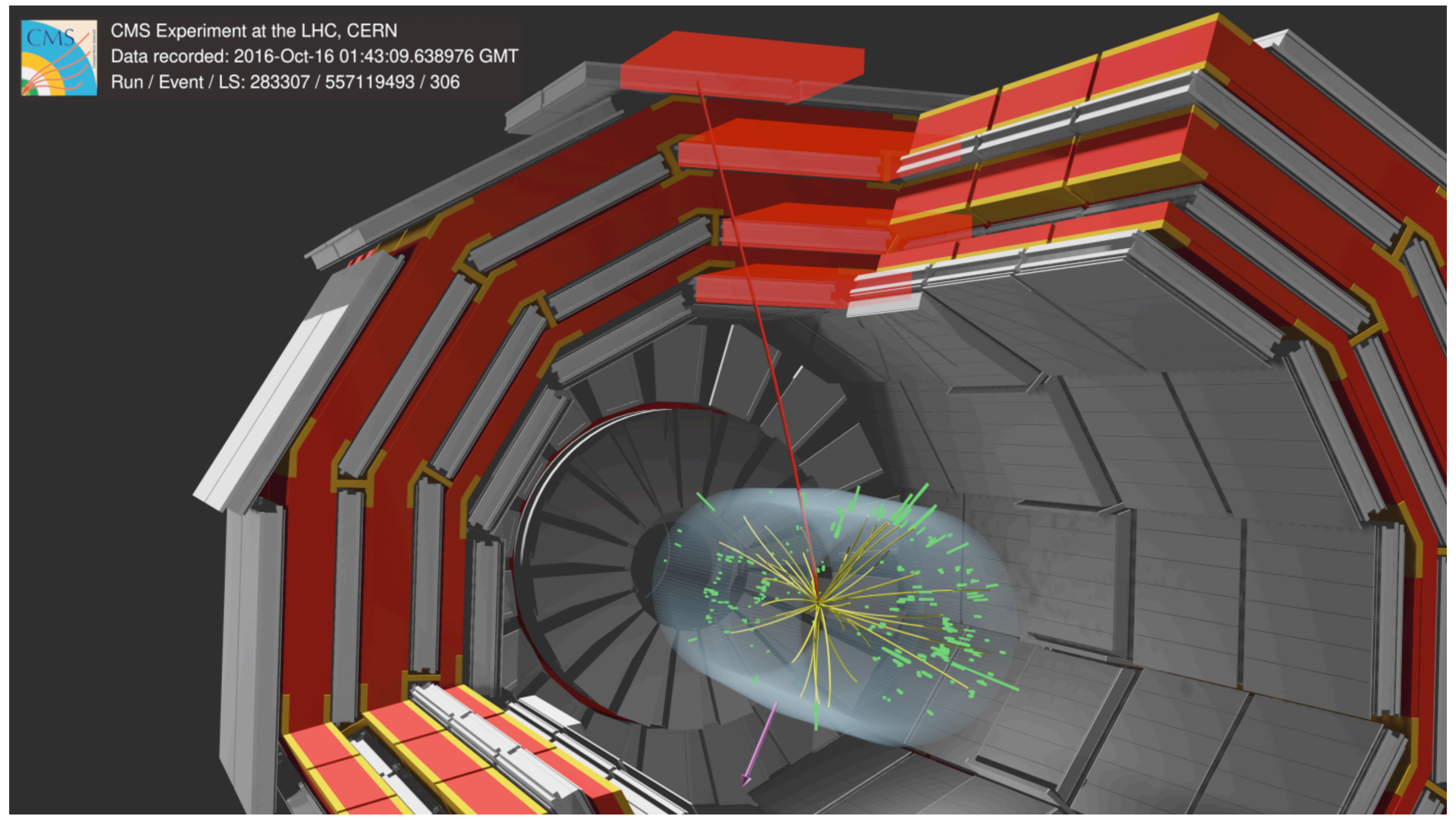


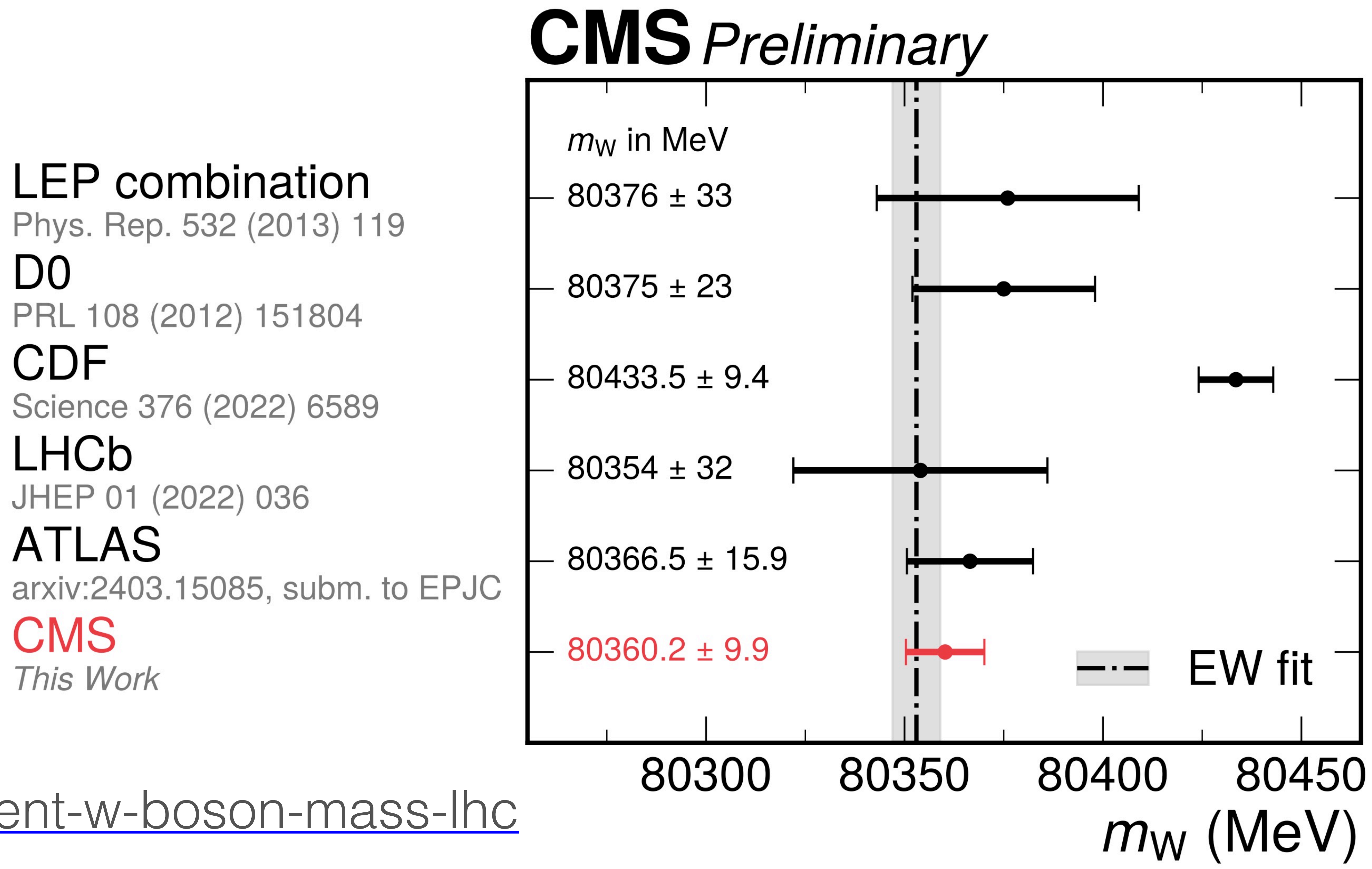
Leveled at pile-up of 62-64 with 3-5% dead-time



Physics results

Highlight of the year was the W mass measurement, many years in the making
NB: Analysis based only on 2016 data required 4B MC events, mostly at NNLO

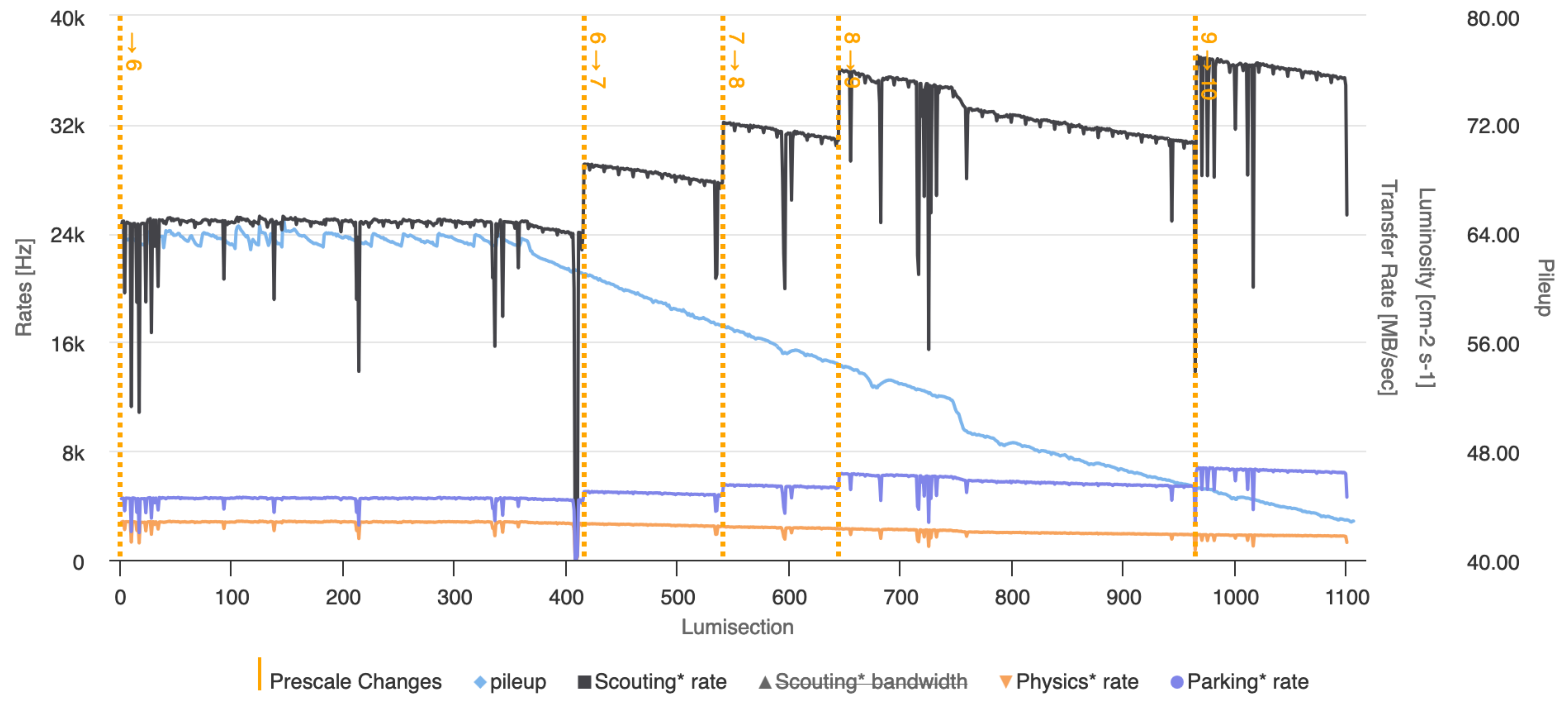




https://cms.cern/news/cms-delivers-best-precision-measurement-w-boson-mass-lhc

Many other new results, e.g., top entanglement, parking / scouting, Run 3 x-sections, UPC, etc.
https://cms.cern/physics

Triggering and datasets



In addition to standard prompt reconstruction data stored form for delayed reco (= parking) as well as HLT-only reco (= scouting)

L1: 110 kHz

Parking: 5 kHz

Prompt: 2.5 kHz

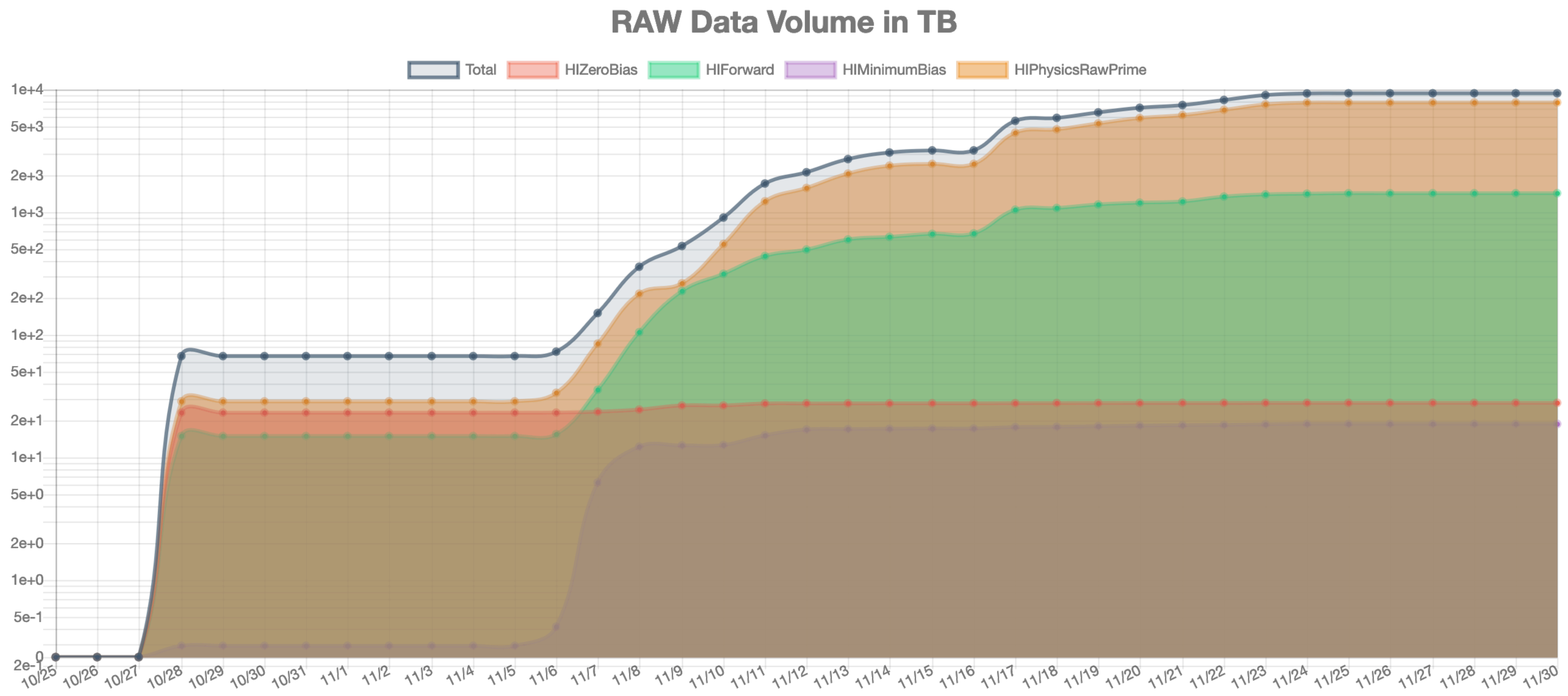
Scouting: > 25 kHZ

NB: Scouting event size is 12kB vs 1MB for standard events

Heavy-ion data taking

New data transfer system with SSD transfers to EOS →collecting data at 32 GB/sec (20 GB/sec in 2023)

- Recorded all hadronic interactions w/ min. bias trigger

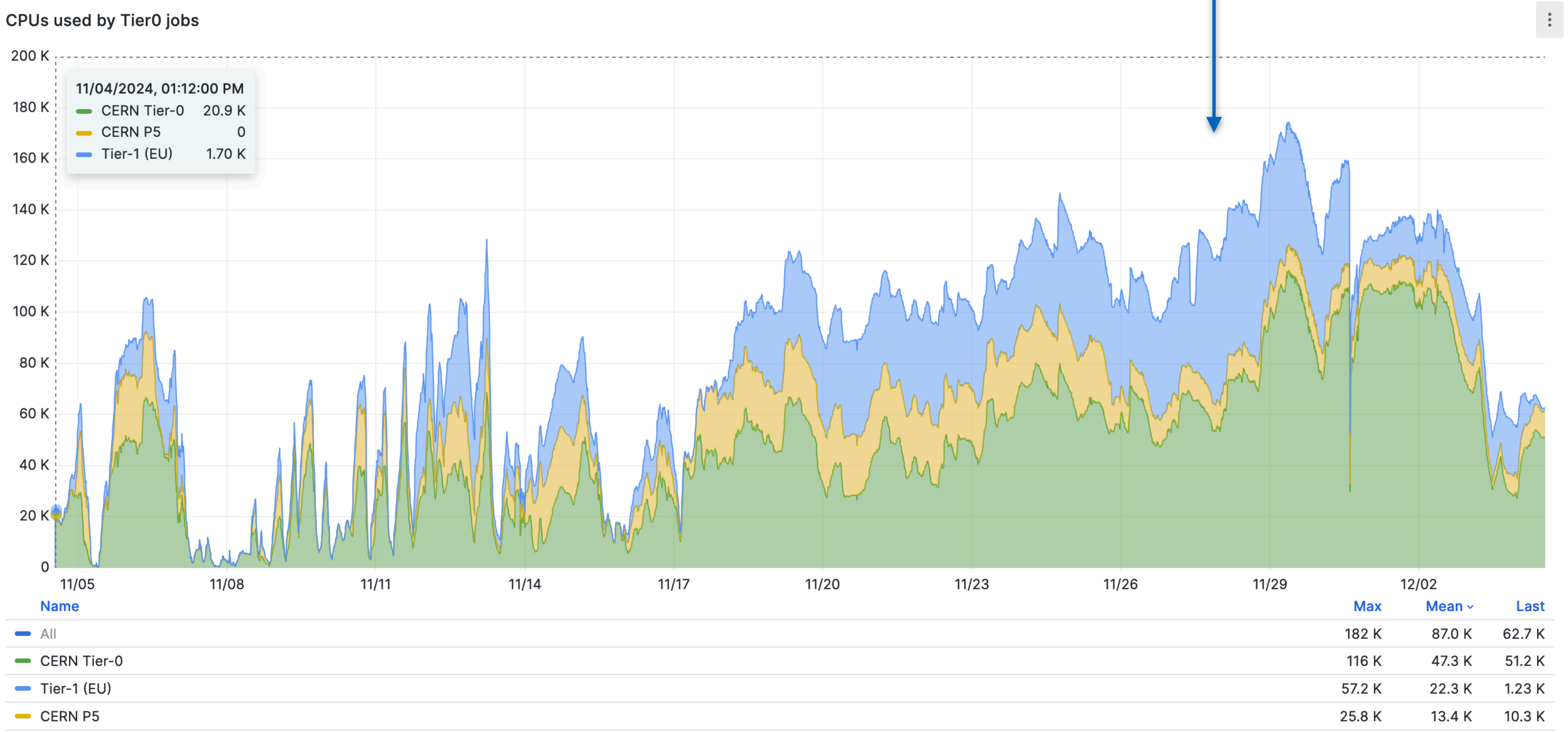


- 33B events collected
  - 15B hadronic interactions
  - 18B EM interactions (UPC)
    - Nearly 10PB of RAW data

- Prompt reconstruction of data ongoing at T0 and T1 cloud, should finish over the holidays

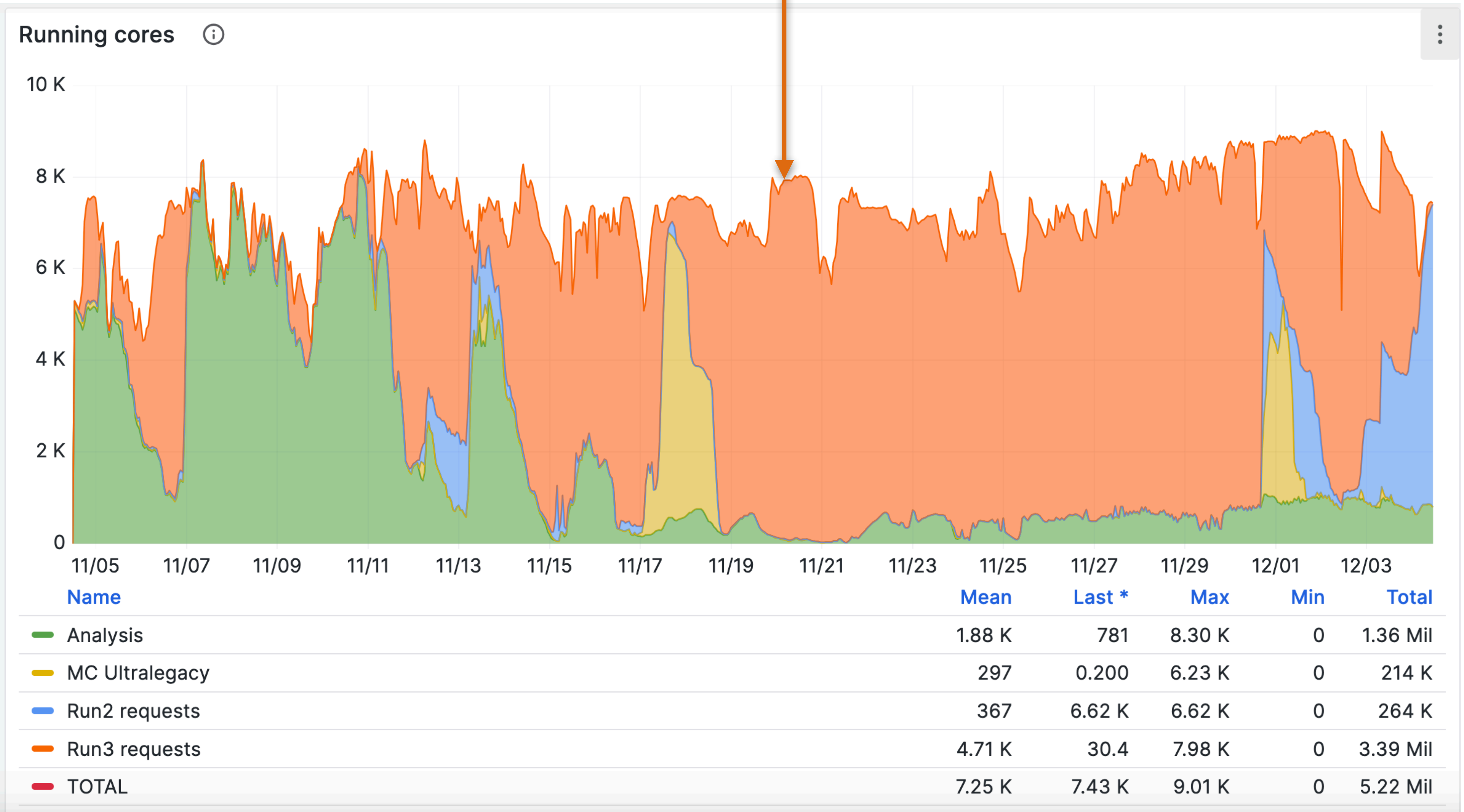
Prompt reconstruction @ Tier-1

EU T1s contributing a big fraction of reco



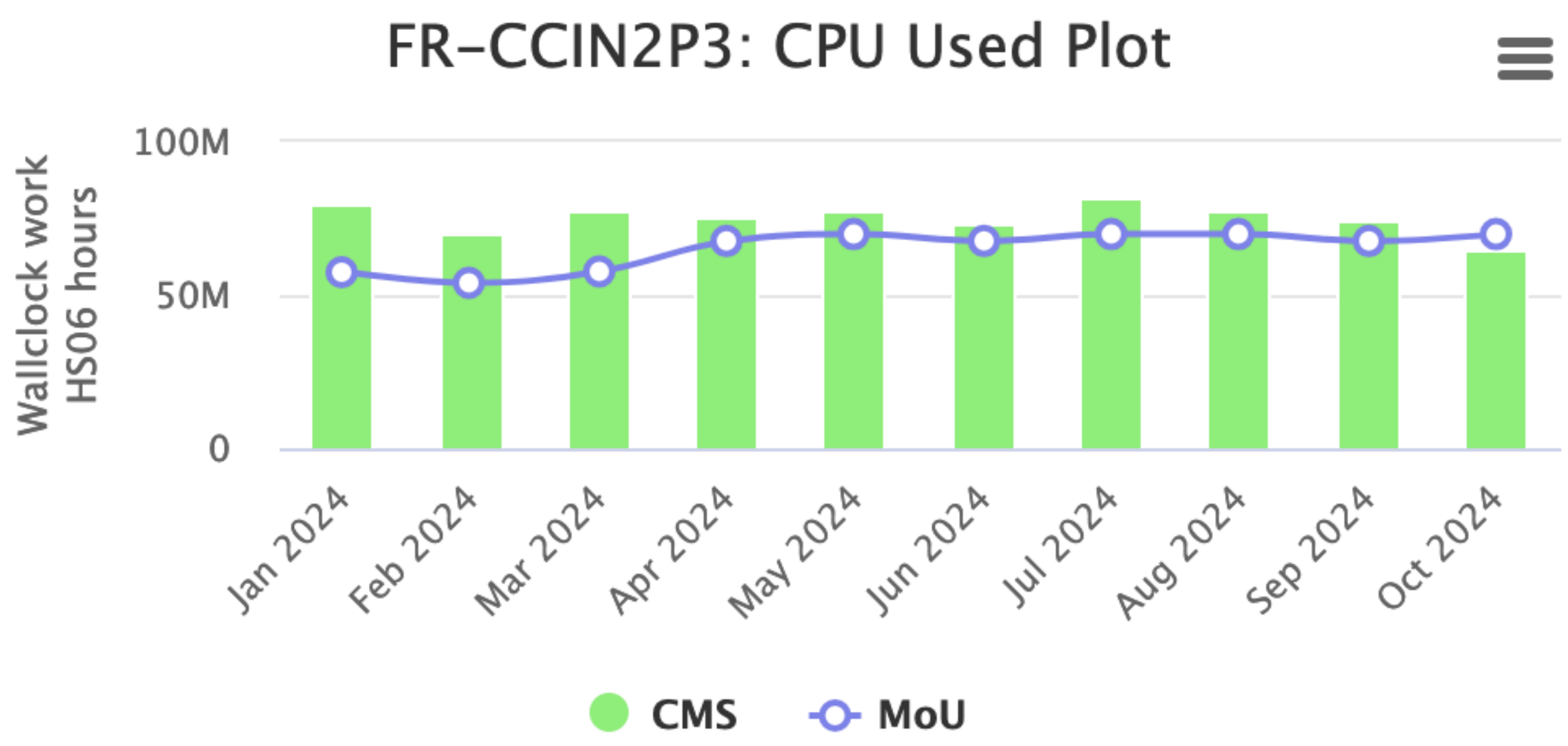
Contributions to prompt reco

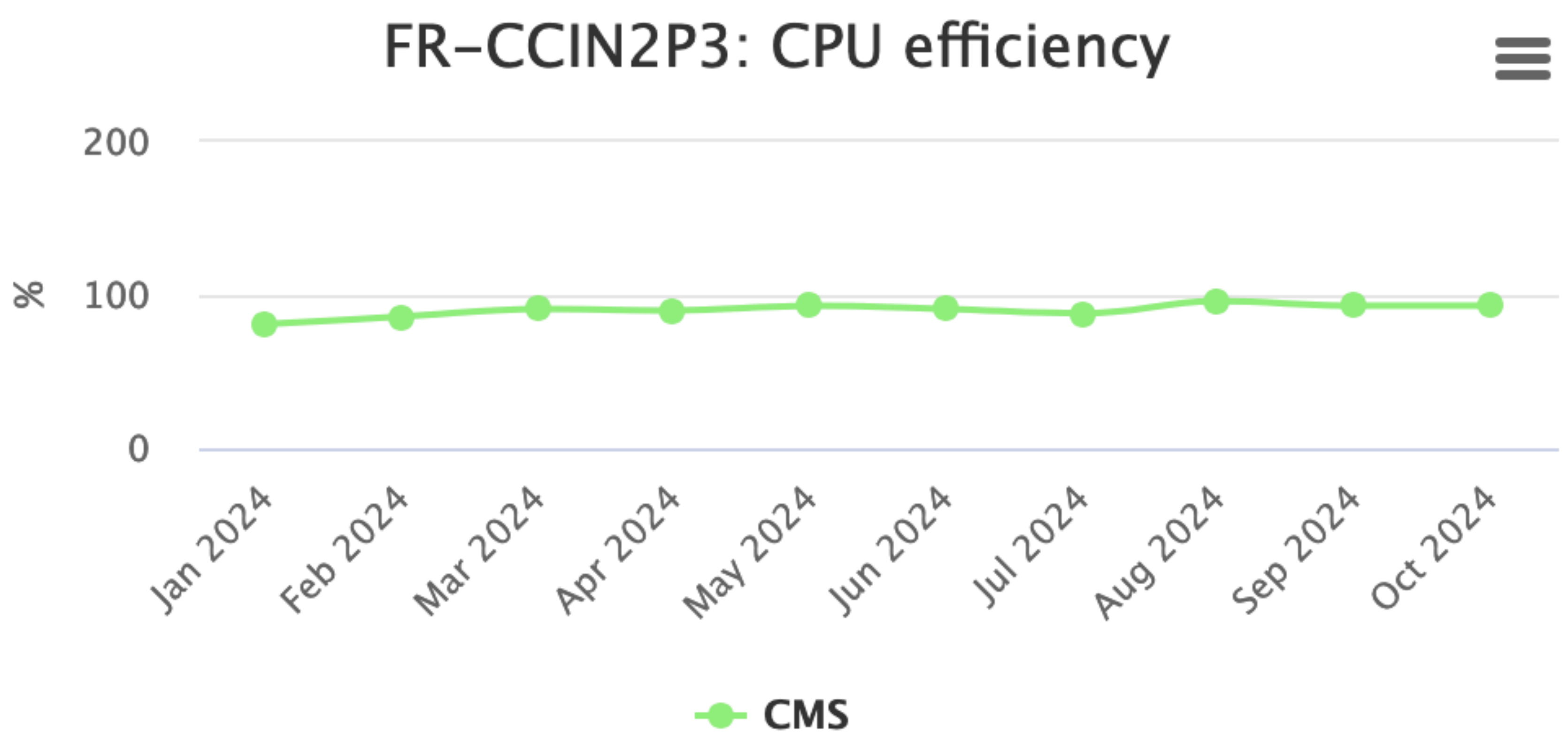
CCIN2P3 ~100% on prompt reco

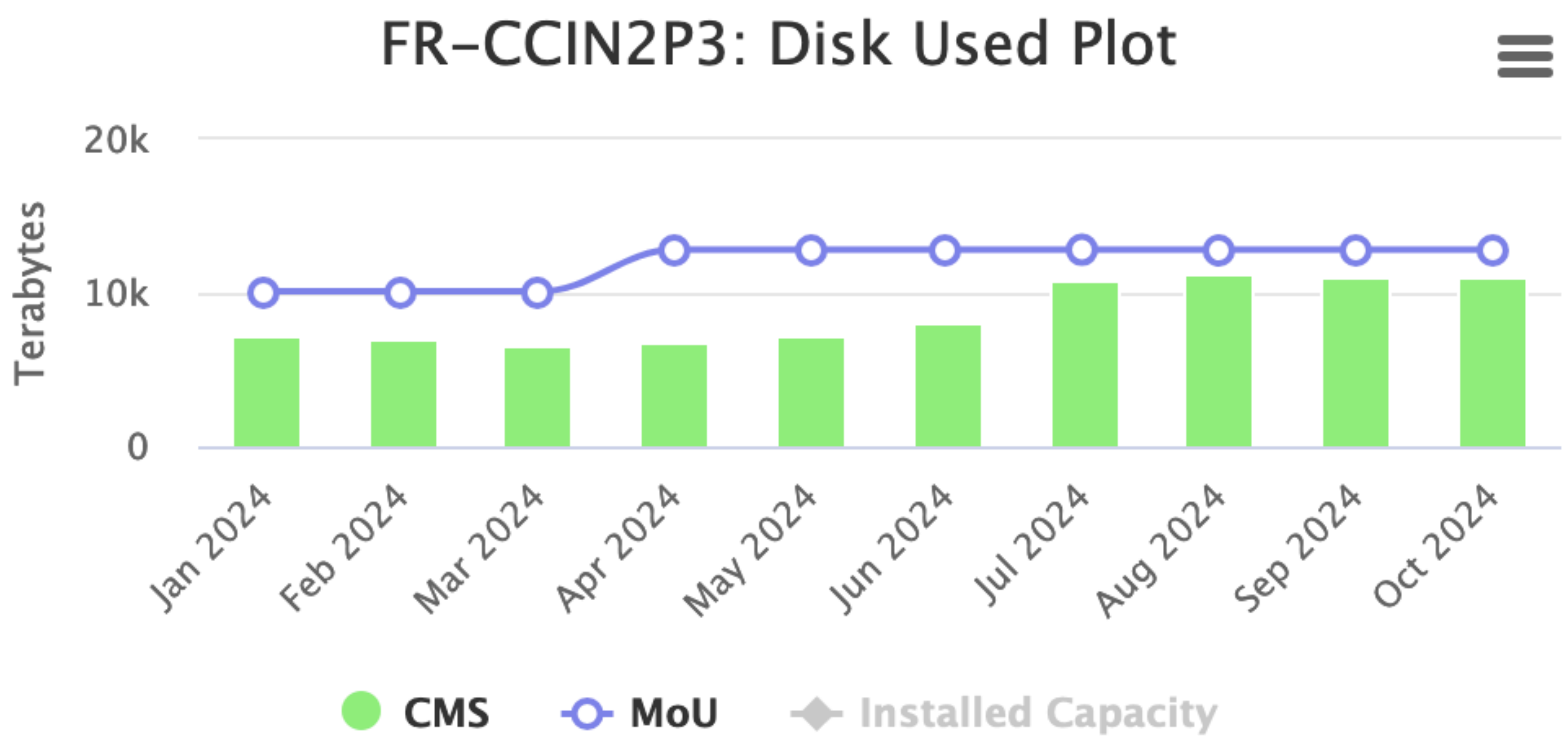


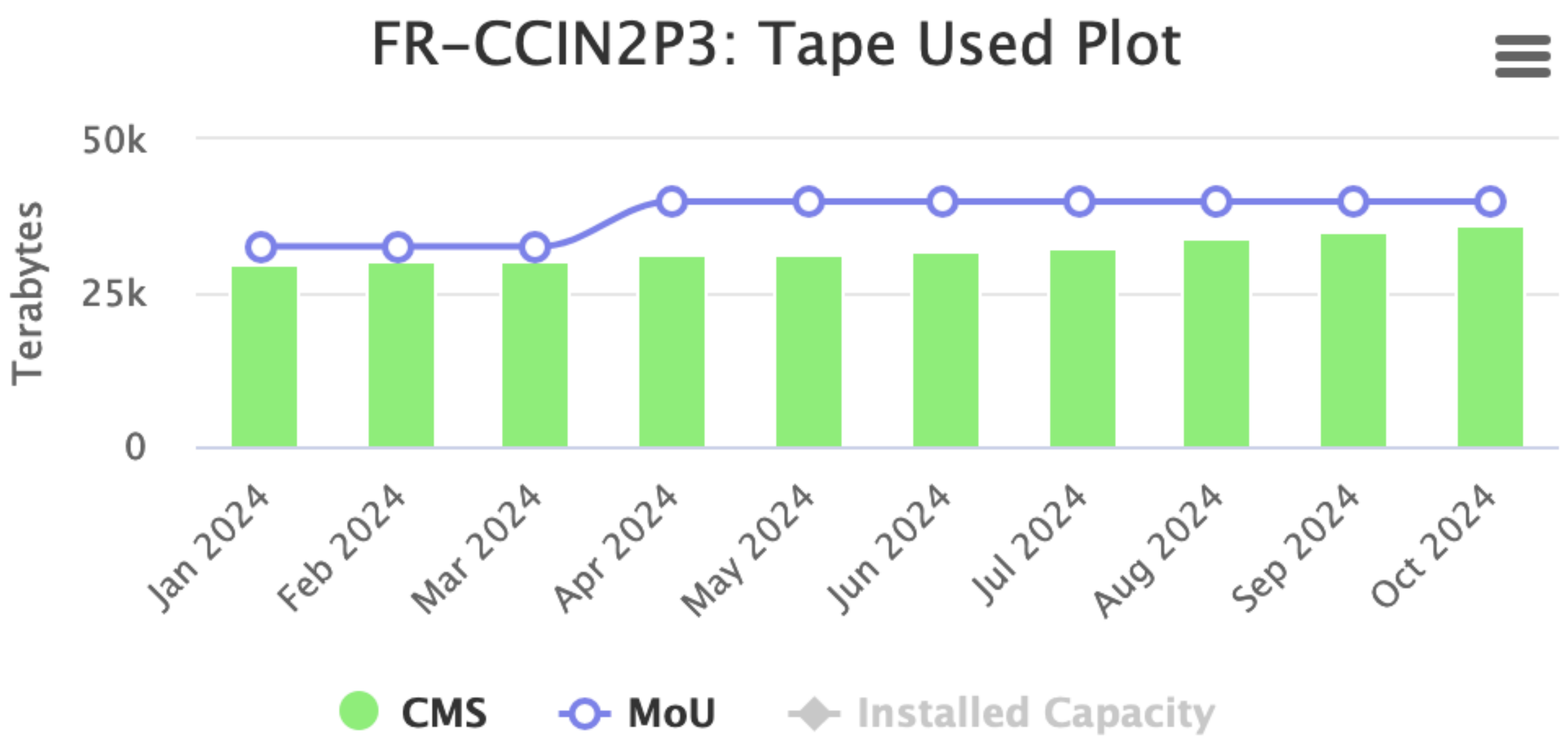
CCIN2P3 utilization in November

Tier-1 CCIN2P3









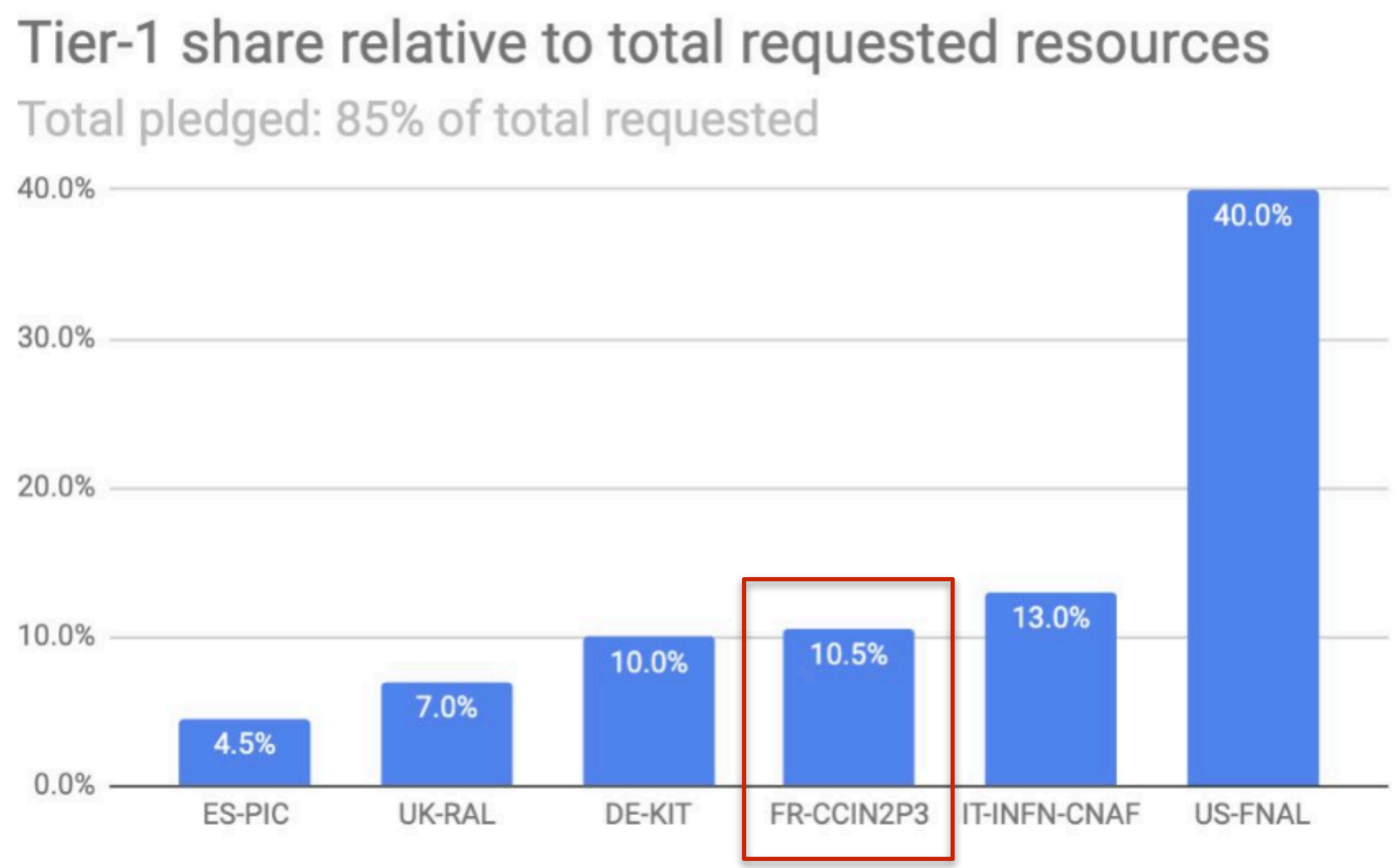
As usual, CCIN2P3 is one of the most reliable sites

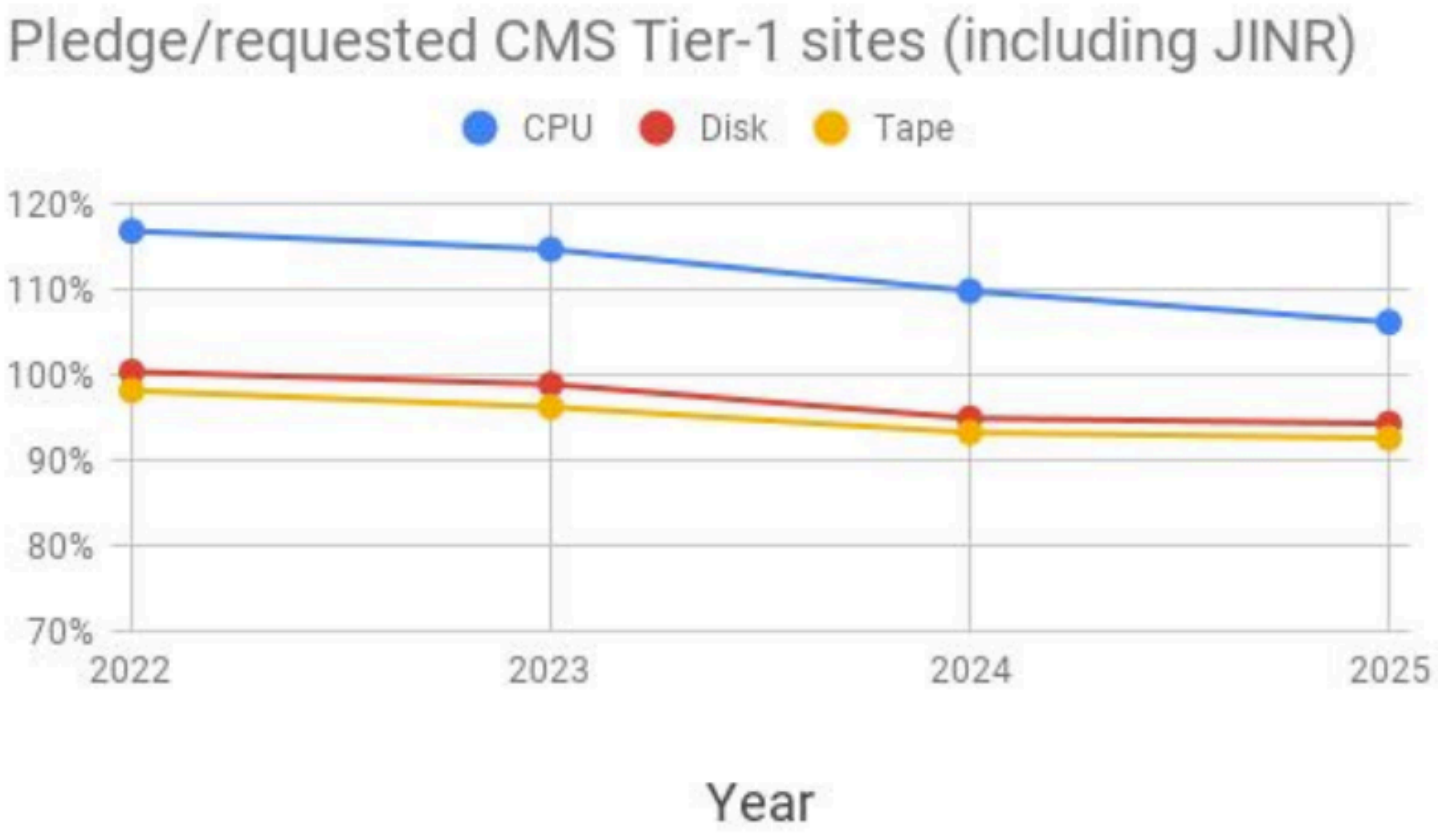
Tier-1 resources in CMS

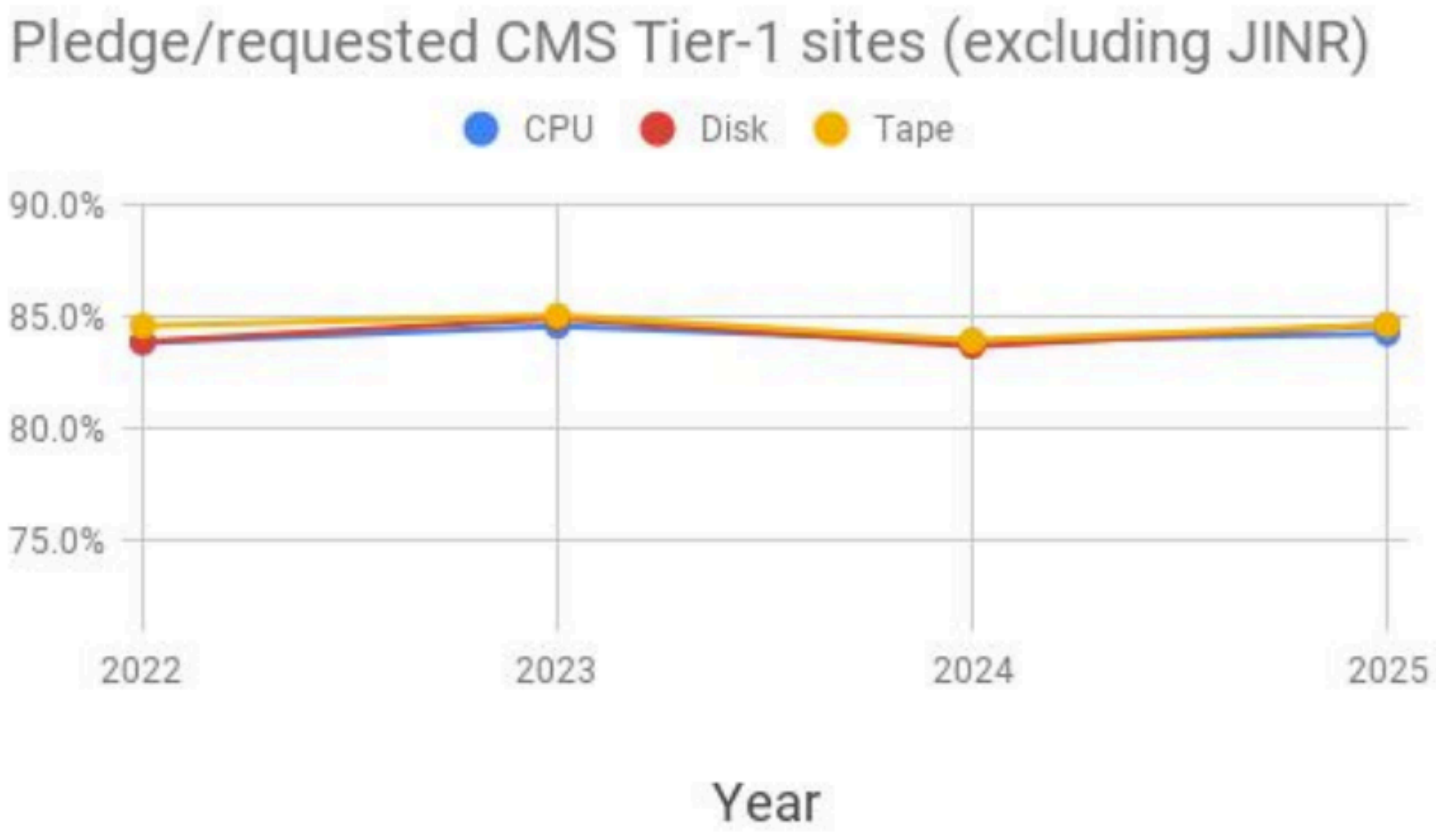
History of Tier-1 computing resource pledges

Tier-1 shares relative to total requested

- ~85% pledged/requested contributed by the sum of Tier-1s excluding JINR







Tier-1 in Serbia

- Target 15% of CMS Tier-1s total capacity
- Build Tier-1 in 2 stages
- Stage 1 – initial configuration, Q3/2025
  - 170 kHS06 = 12 kCores = 24 kThreads
  - 20 PB disc storage
  - 60 PB tape storage

  } 6 racks + 2 racks + tape library

UPDATED

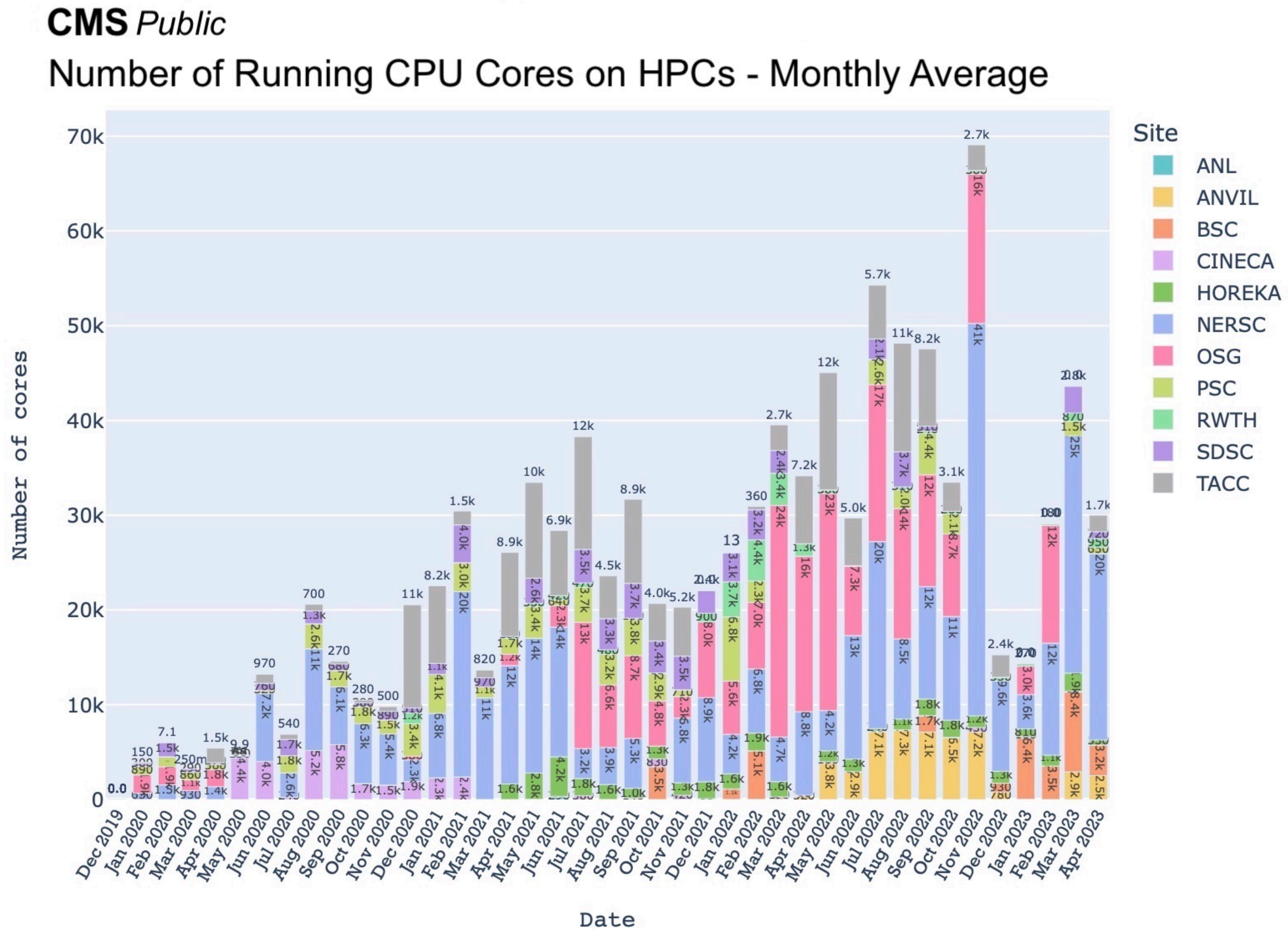
- Stage 2 – reach adiabatically '26/ '27/ '28/ '29
  - look for new solutions to spare space in the mini module – capacity 15 racks + tape library)
    - 350 kHS06 = 25 kCores = 50 kThreads
    - 40 PB disc storage
    - 120 PB tape storage

    } 10 racks + 4 racks + tape library expanded

28OCT24 CMS CRB, Serbian T1 Status and HLD, V. Rekovic 8

High Performance Computing Sites

Usage of HPCs continues to grow, with contributions from many sites

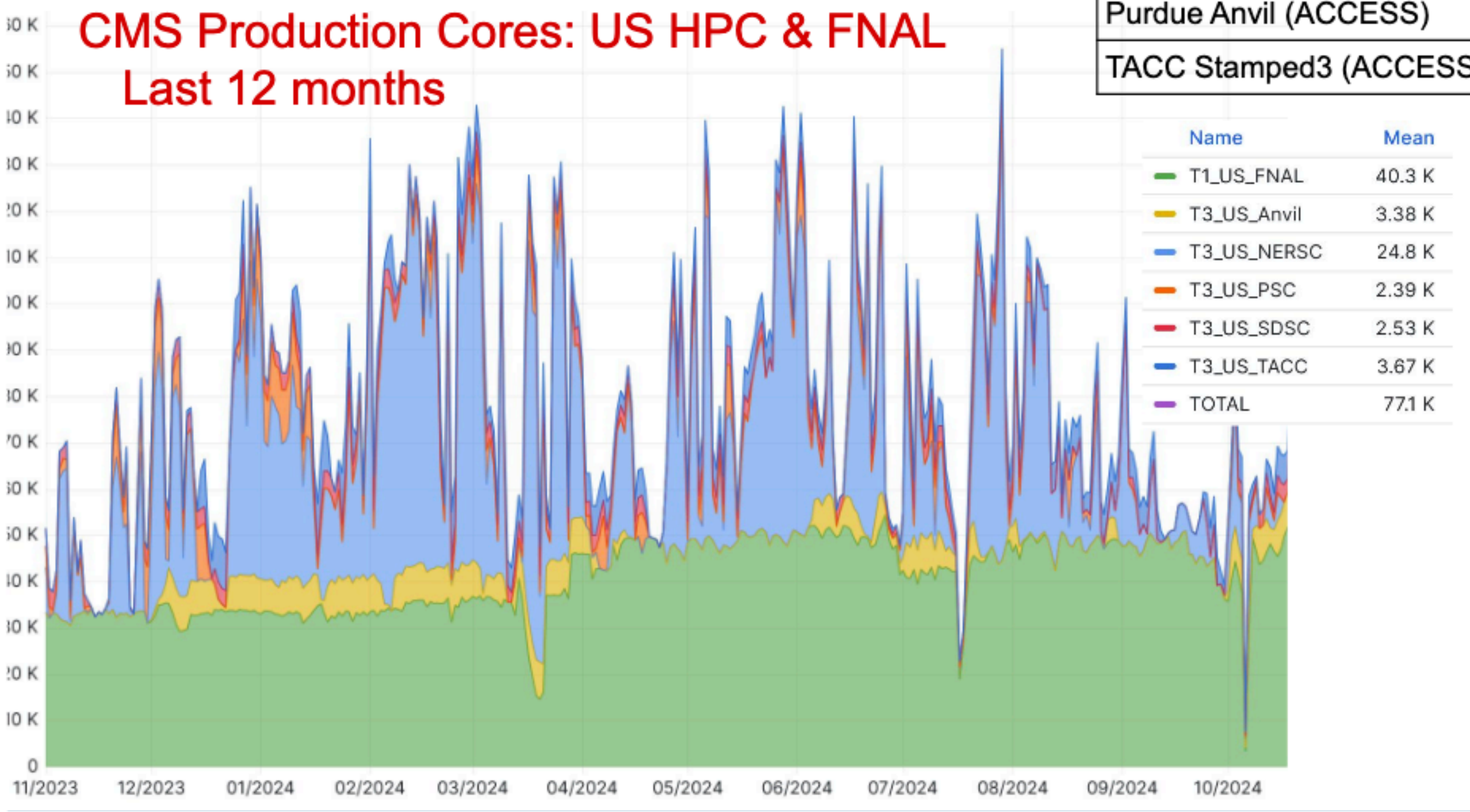


with large month-to-month variations

HPC sites by region

HPC usage driven by US sites, mainly at NERSC

- Comparable to FNAL T1 in scale (but not continuously available/used, utilization fluctuates)



HPC Resource	Allocation (core hours)	Allocation Period	%Used*
NERSC Perlmutter (DOE)	337M CPU	Jan 2024 - Jan 2025	74%
TACC Frontera (NSF)	36M	Jun 2024 - May 2025	38%
PSC Bridges-2 (ACCESS)	23M	Oct 2024 - Sep 2025	2%
SDSC Expanse (ACCESS)	23M	Oct 2024 - Sep 2025	8%
Purdue Anvil (ACCESS)	23M	Oct 2024 - Sep 2025	15%
TACC Stamped3 (ACCESS)	1M	Oct 2024 - Sep 2025	0%

Currently Active Allocations

But also HPC initiatives in Europe

Switzerland: All pledged T2_CH_CSCS CPU is provided via the Piz Daint HPC, completely transparent integration via ARC-CEs

Germany:
Site extension of T1_DE_KIT (HoreKA)
Site extension of T2_DE_RWTH (CLAIX)
Transparent integration

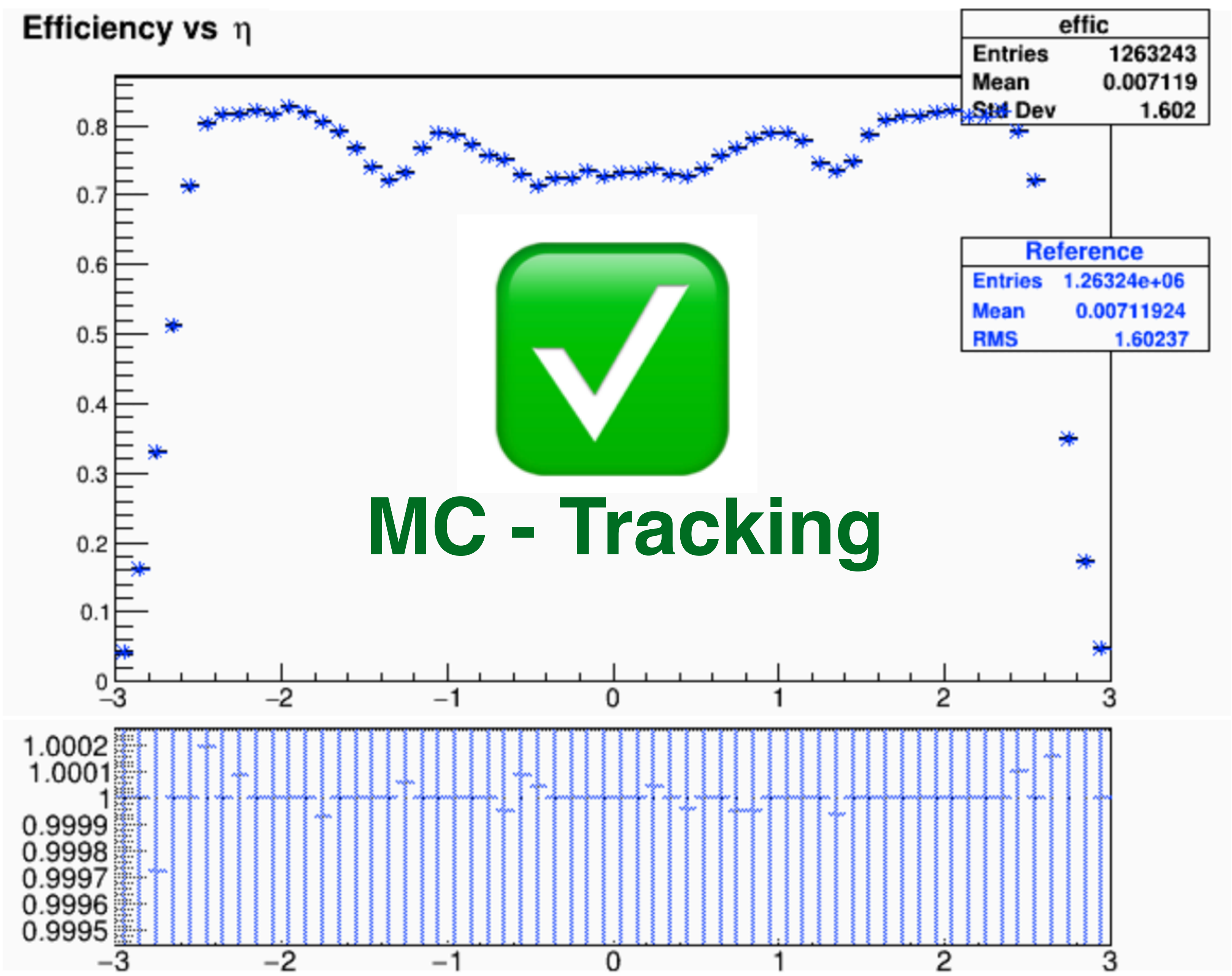
Spain:
BSC integrated as T3_ES_PIC_BSC

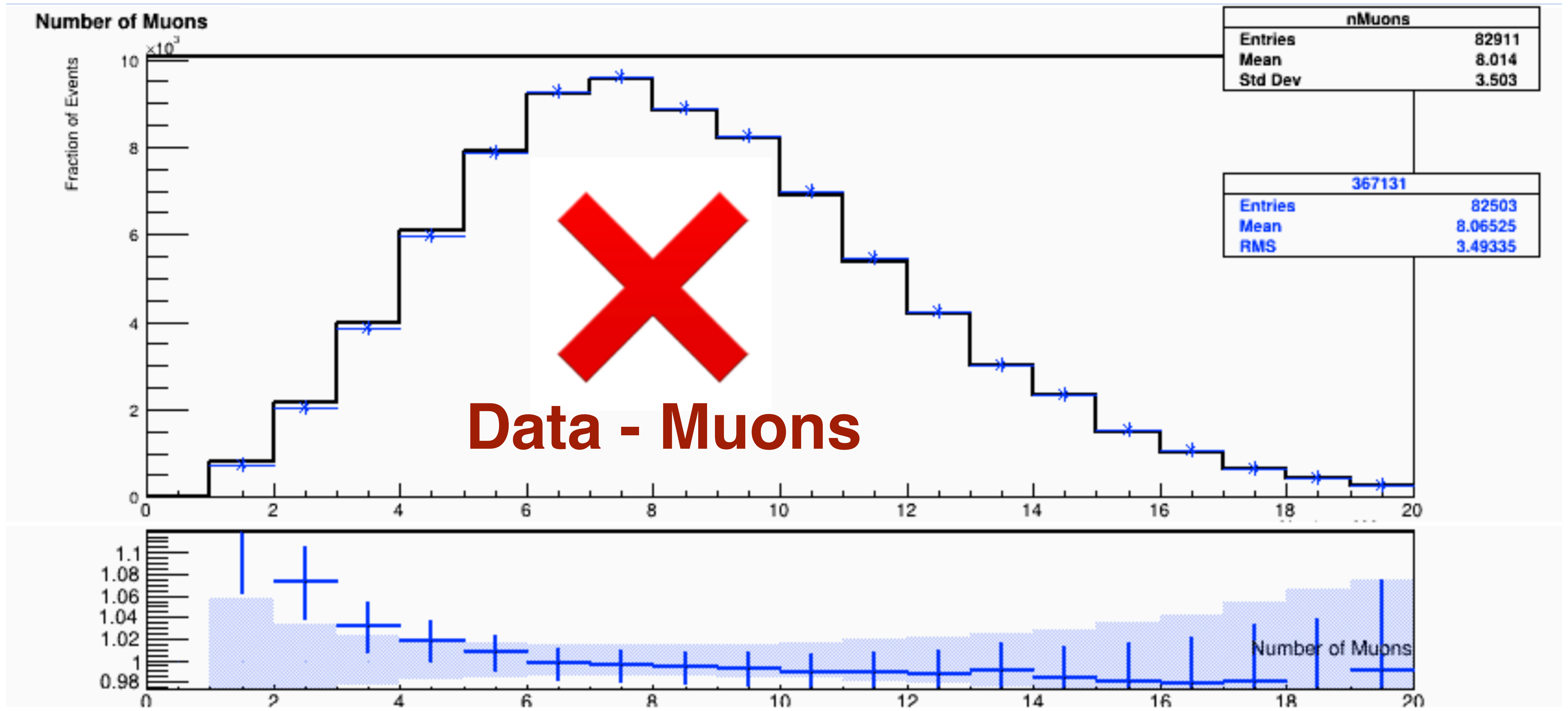
Italy:
CINECA Leonardo GP completely transparent integration in the batch system at T1_IT_CNAF
CINECA Leonardo GP/Booster as opportunistic resources. CNAF-CINECA subsite

France is notably absent

Testing ARM resources

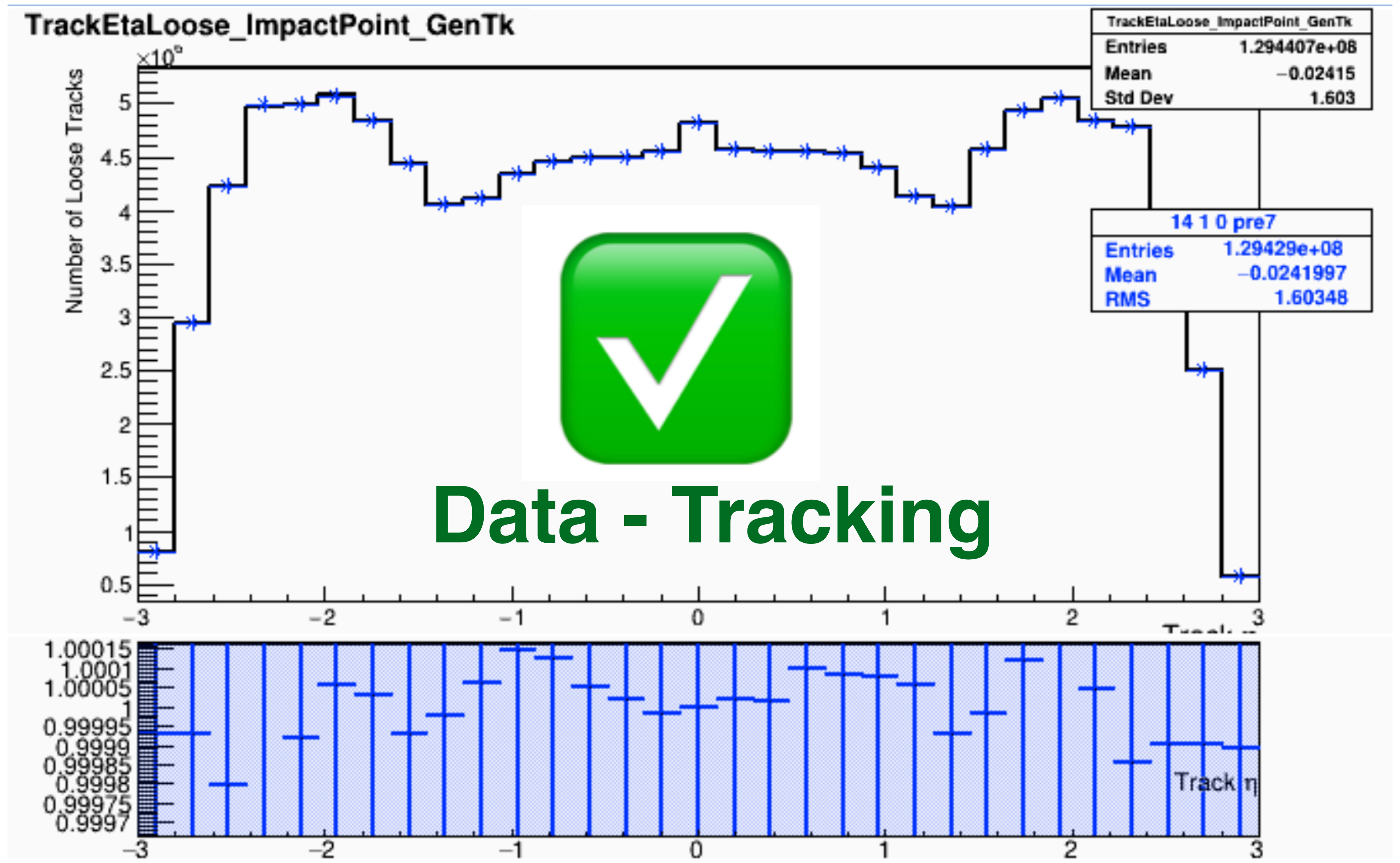
- CMSSW supports ARM since 2016 (and POWER since 2014).regularly (daily) tested but lacking full-fledged physics validation.
- In 2023 we got access for a short period to resources at **T3_UK_ScotGrid_GLA** (thanks to ATLAS colleagues)
- We performed high statistics production for Run3 pp and heavy-ion MC and 2023 Data.
  Not straightforward to setup but successful (in terms of production). But two different outcomes:
  1. **MC: with ~7·10^6 (for 40 processes) reconstructed events (produced ×10^2, gen filter efficiency). Green light!**
  2. **DATA: ~4·10^6 events from 2023 RAW data.**
     A. **Physics groups are (rightfully) more "demanding"** on data since whole chain should be run on exactly the same events
     B. Less workflows w/ more events, O(100k)==much longer runs ⊕ Data @ CERN (so need staging) → **More prone to failures**
     C. Being the **resources available for a limited span of time makes recovery tricky**

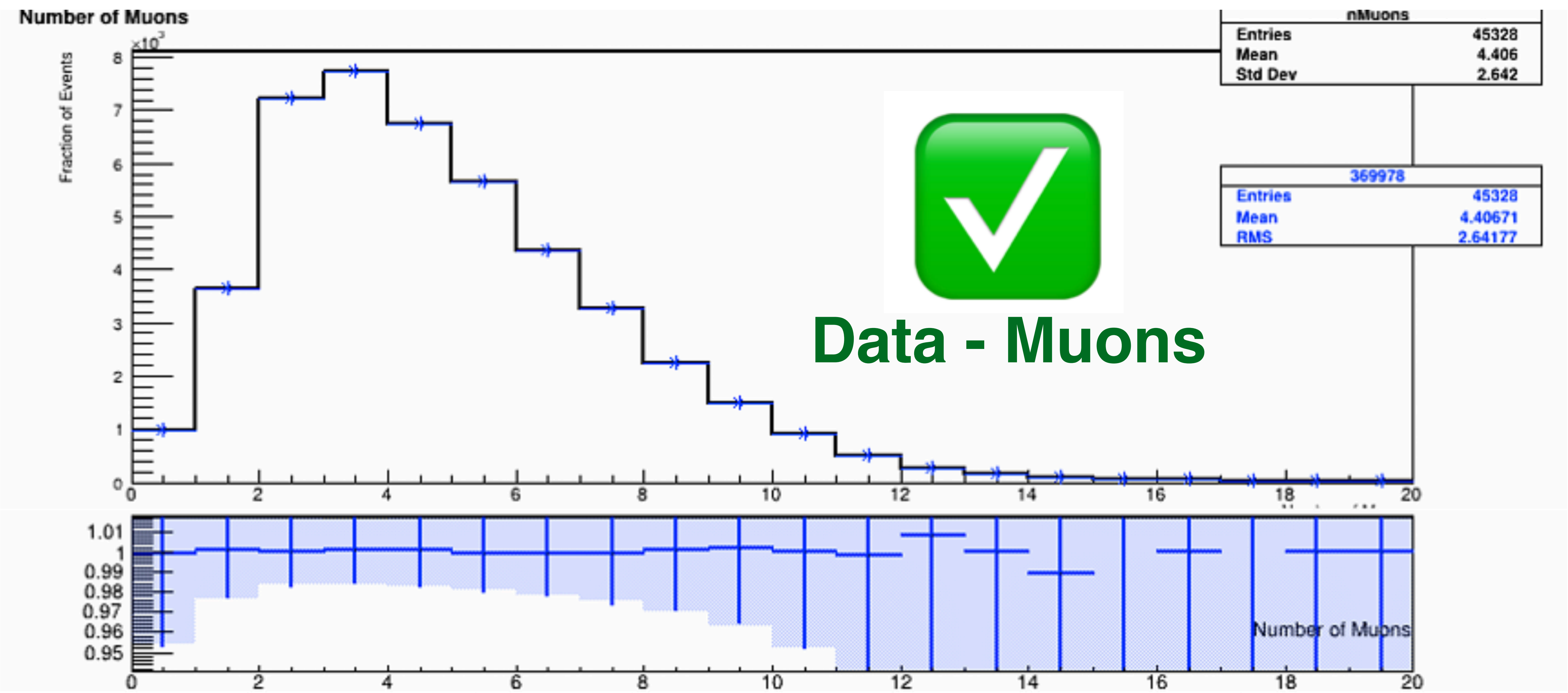




ARM - 2024 update

- In 2024 we got access to much more resources and in a stable way from KIT (the vast majority), CERN and CNAF.
- We carefully repeated the Data validation with 2022 and 2023 data.
- Validation is nearly done and no problem spotted → CMS should be able to add ARM resources to its pool.
- A production-like test MC production is also coming in the next weeks.
- Lesson learned: for this kind of integrations we need enough resources available steadily.





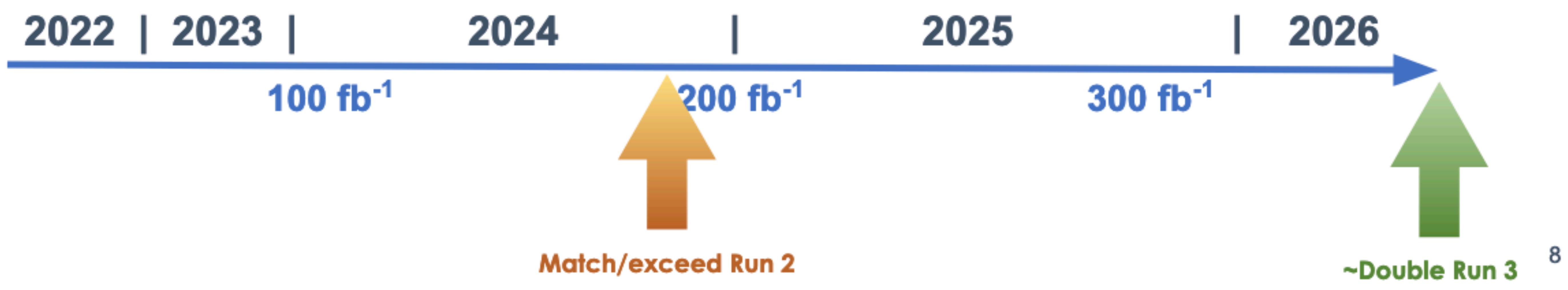
Run 3 schedule

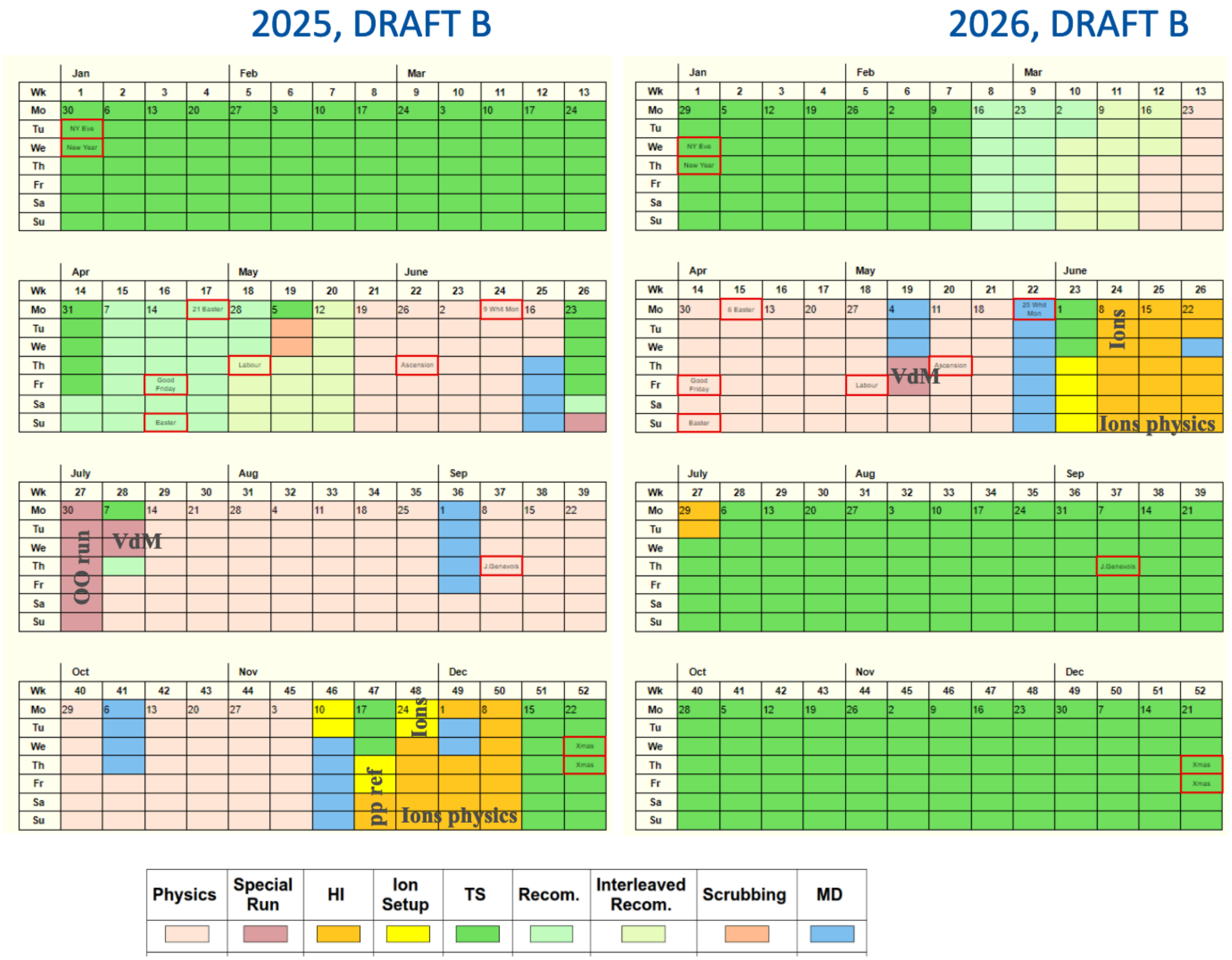
New schedule approved by Council

"Scenario B"

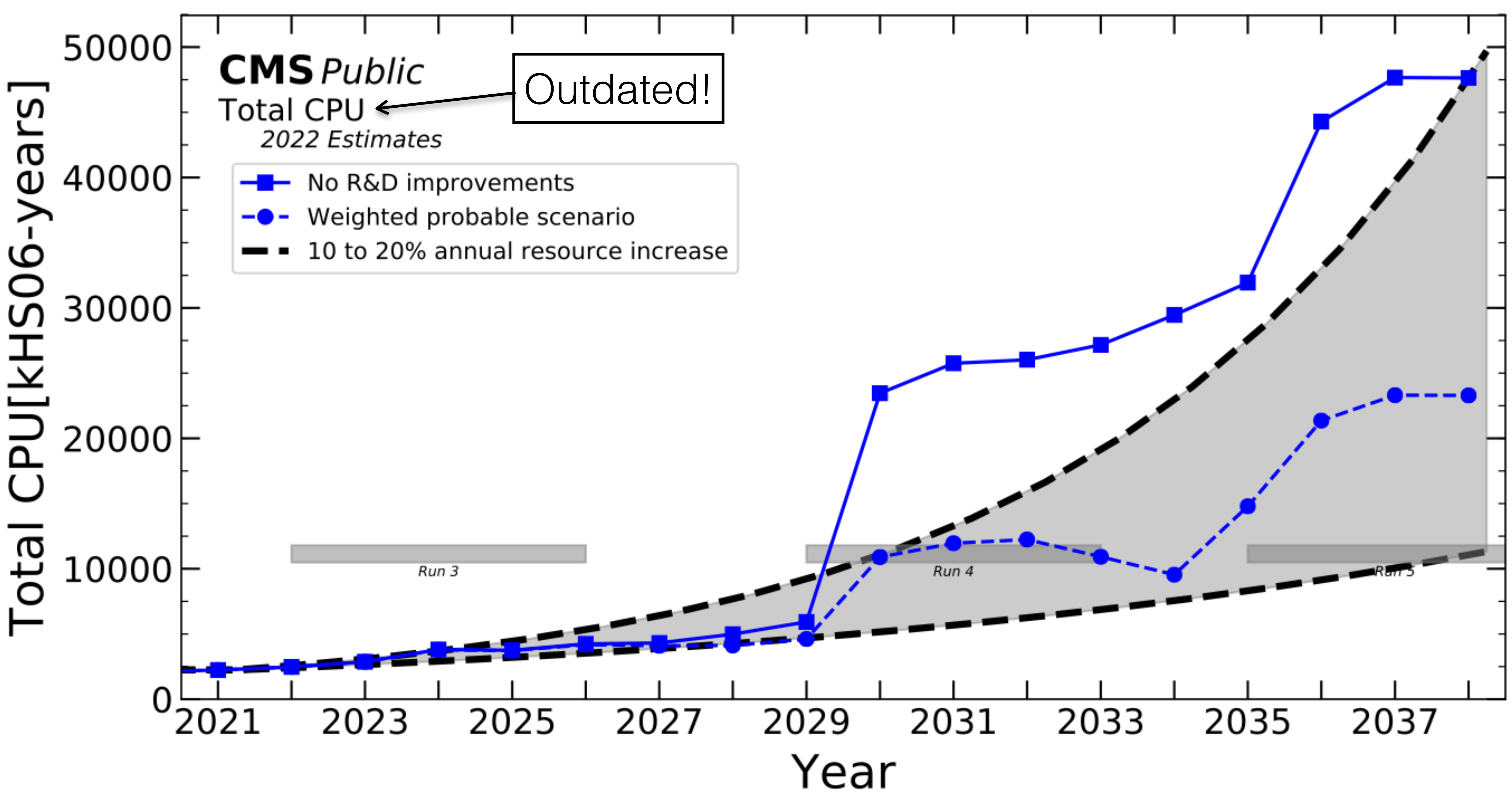
https://indico.cern.ch/event/1462121/

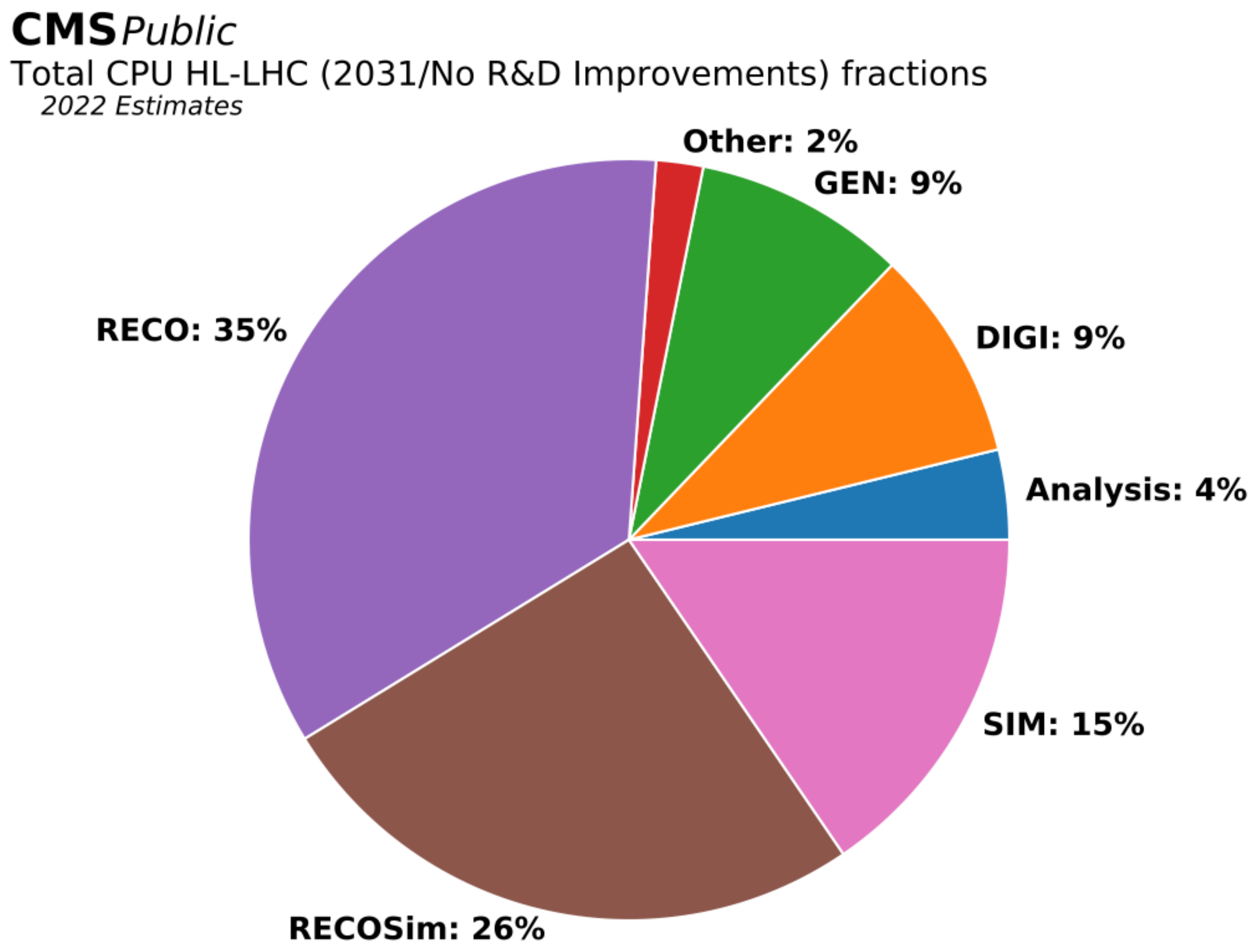
- Run 3 extended thru June 2026
- Full production year in 2025, similar to 2024
- Very short YETS
- Expect another ~ 180 /fb in 25/26
  → we are midway thru Run 3





Phase-2 CPU Projections





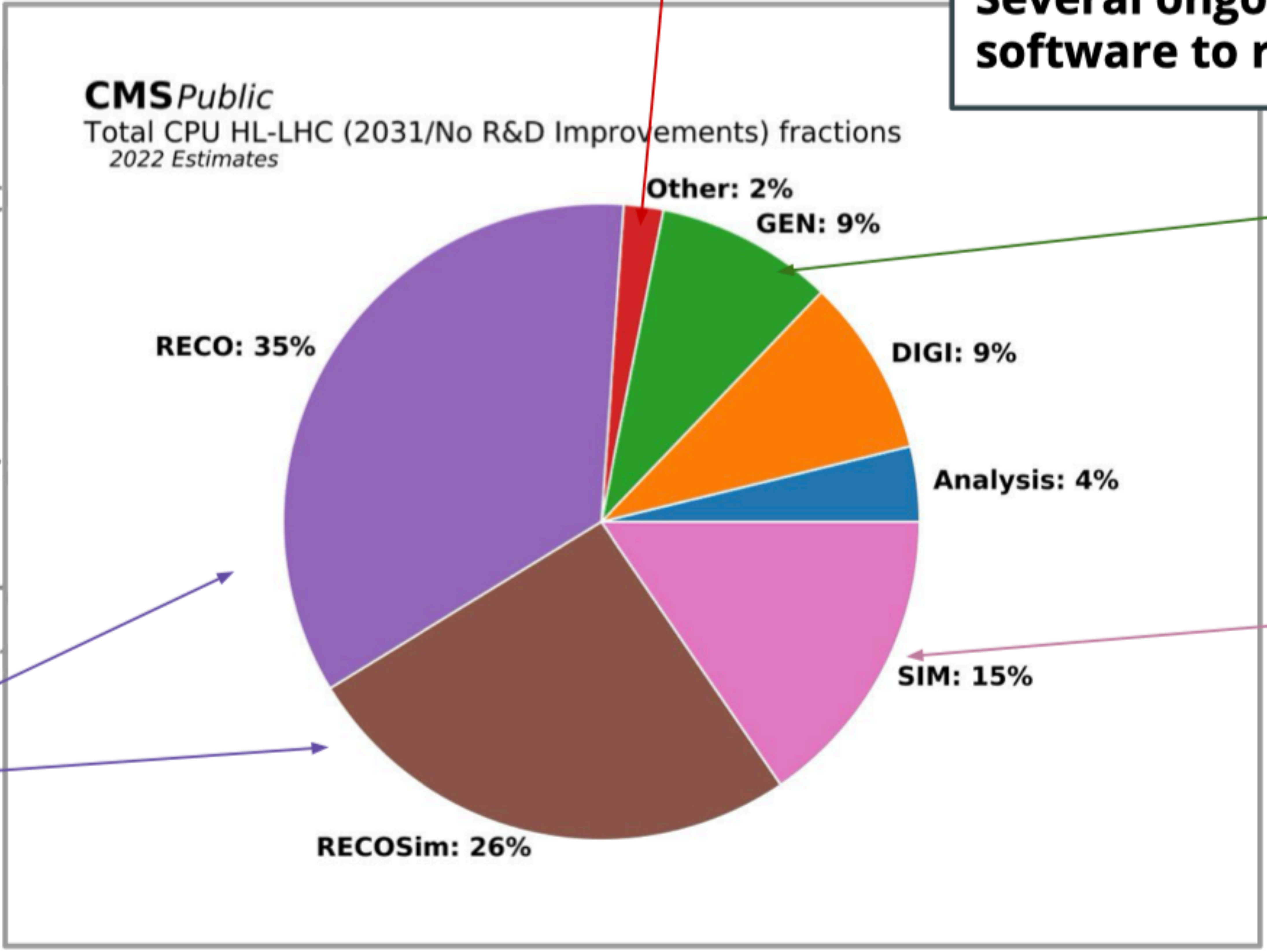
Breakdown of CPU usage in CMS

- To keep up with t
  should be carried
  tasks on accelerat

- **CPU time is split**
  - Event gener
  - Simulation
  - CMS event r
  - Reduced da

ML inference

Several ongoing efforts to enable software to run on GPU

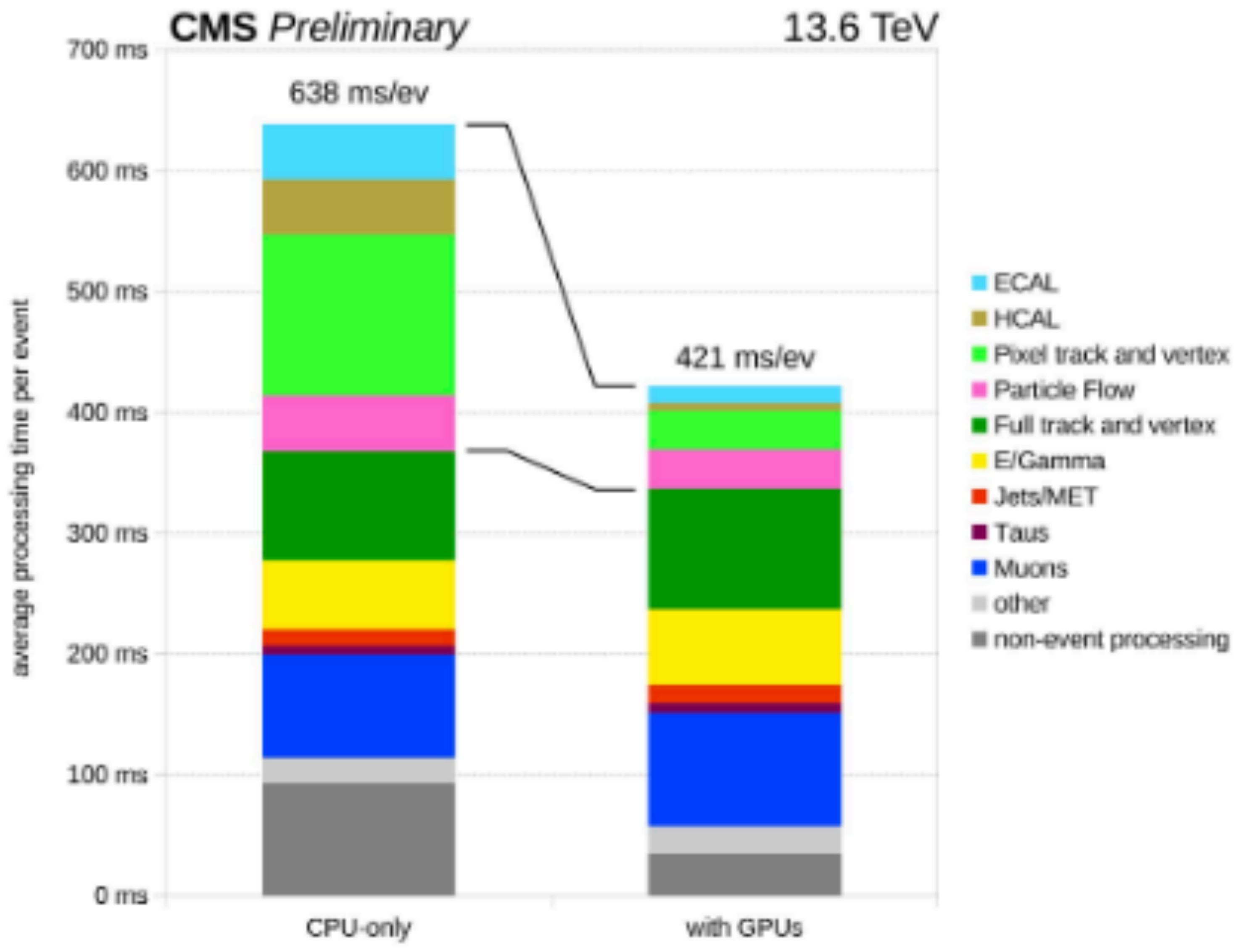


Mad4GPU

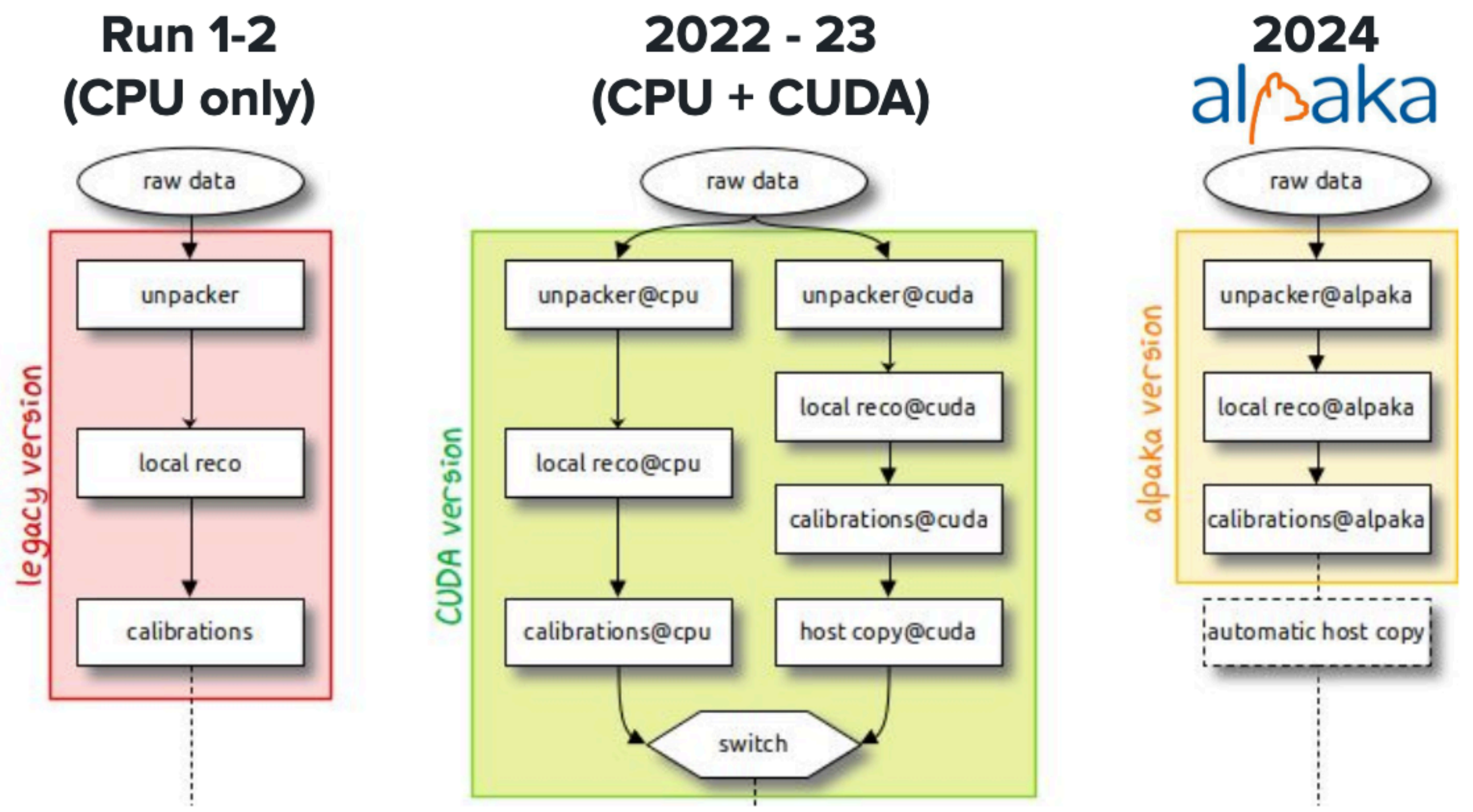
FastSim (ML based)
FullSim e.g. AdePT / Celeritas

Tracking
Vertexing
ECAL/HCAL local reconstruction
Clustering & PF
Electron seeding

GPU reconstruction



- GPU development for Run 3 driven by HLT
- 35% improvement in trigger latency
- Mostly implemented in “local” reconstruction
  - Pixel local reco, tracking & vertex reco
  - ECAL & HCAL local reco & clustering
- Initially in CUDA, now migrated to Alpaka



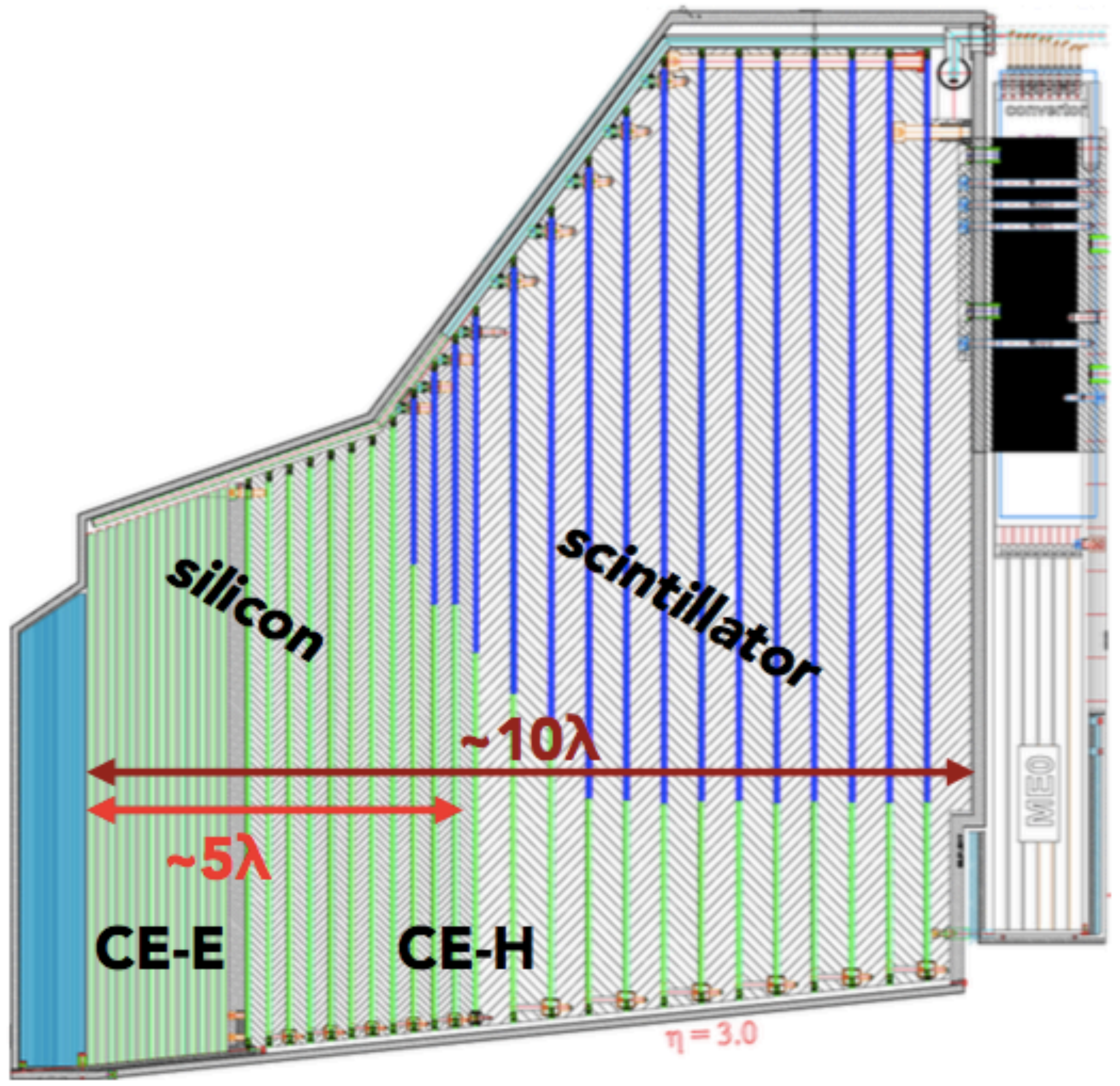
Alpaka *performance portability* library allows a single source to be built for and run on:

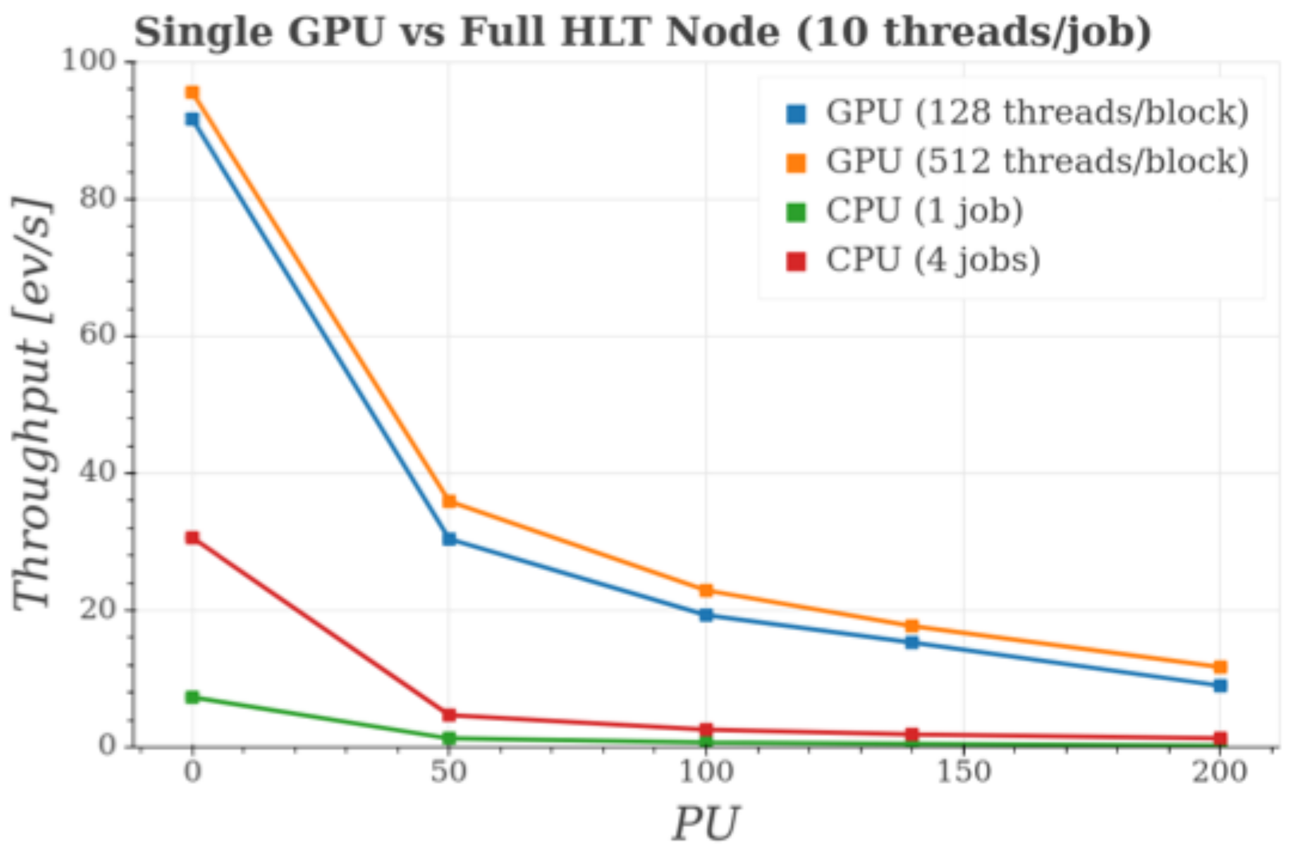
- x86 and ARM CPUs
- NVIDIA and AMD GPUs
- experimental support for Intel GPUs (and FPGAs)
  → not yet enabled in CMSSW

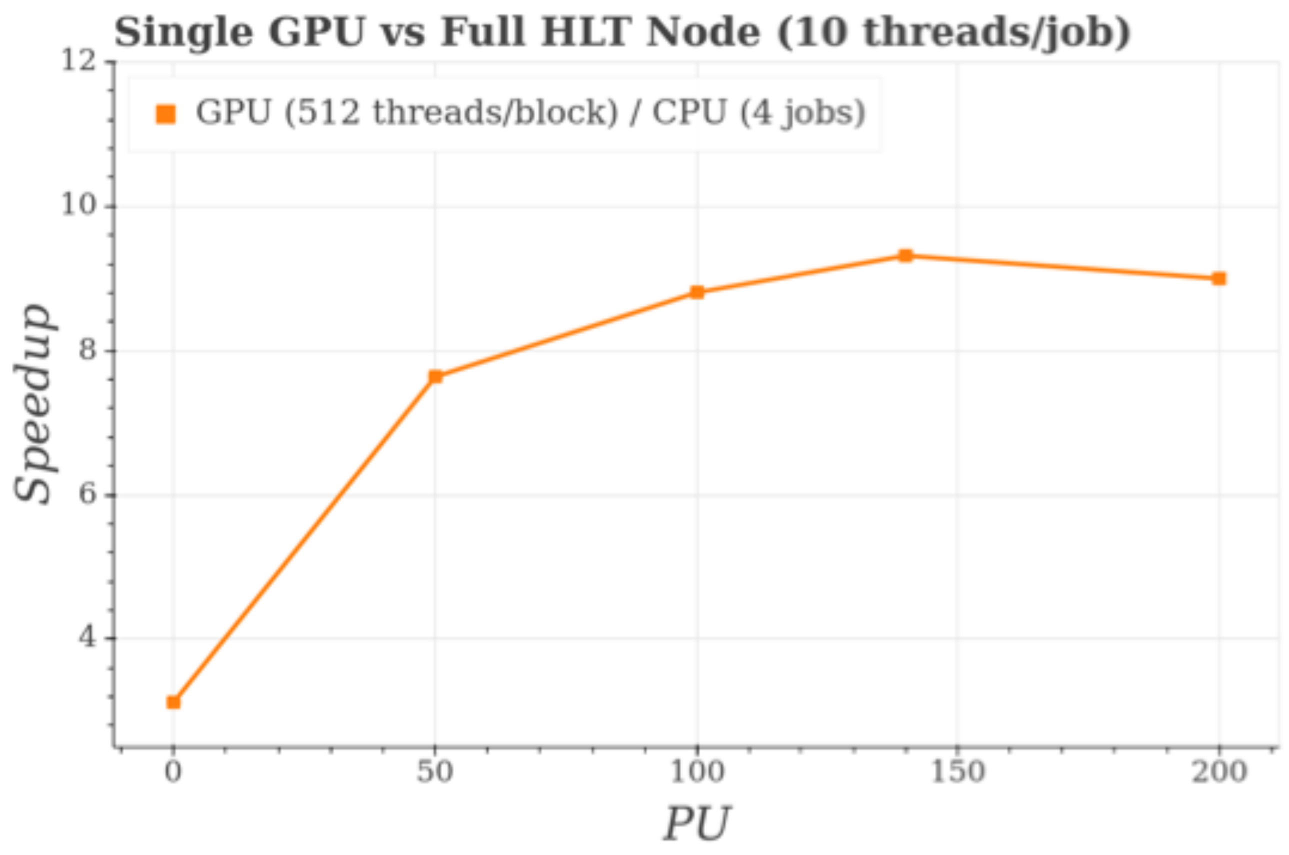
HLT reconstruction is different from offline reconstruction, which is dominated by “high-level” reco, e.g., full tracking

Reconstruction on GPUs: Calorimeters

For Phase-2 endcap will be replaced with
High Granularity Calorimeter → 6M channels!

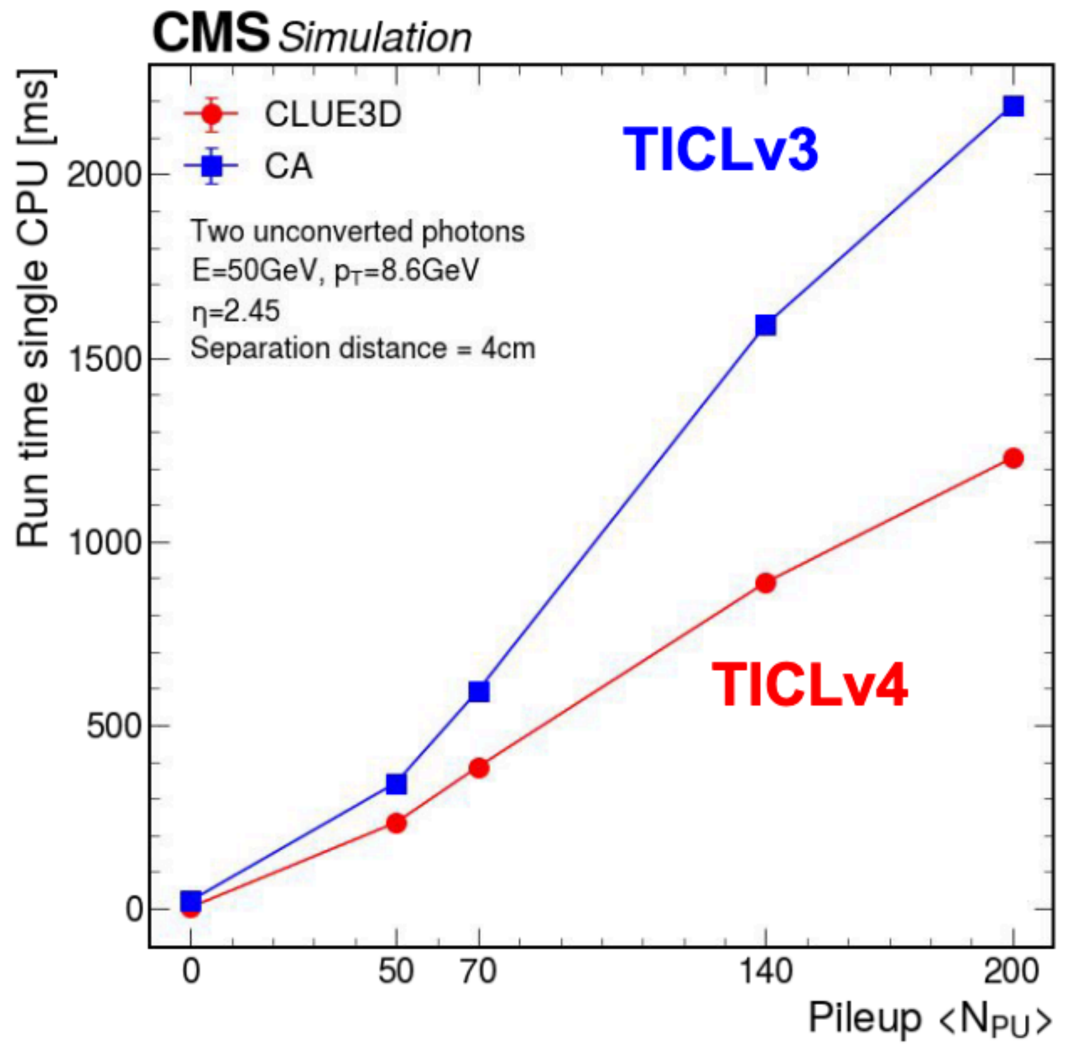






A new clustering framework (TICL) was designed to run on GPUs

CLUE https://inspirehep.net/literature/1777434
TICL: https://cds.cern.ch/record/2839740



Algorithms are under continuous development
Currently consuming only 5% of reco time
→ HGCAL reco will not be dominant

Reconstruction on GPUs: Tracking

Track finding & propagation thru the outer layers is the most time consuming part of reconstruction
Current "Combinatorial Kalman Filter" is by sequential by nature, not suitable for GPUs

Phase-2 outer tracker will have doublet layers

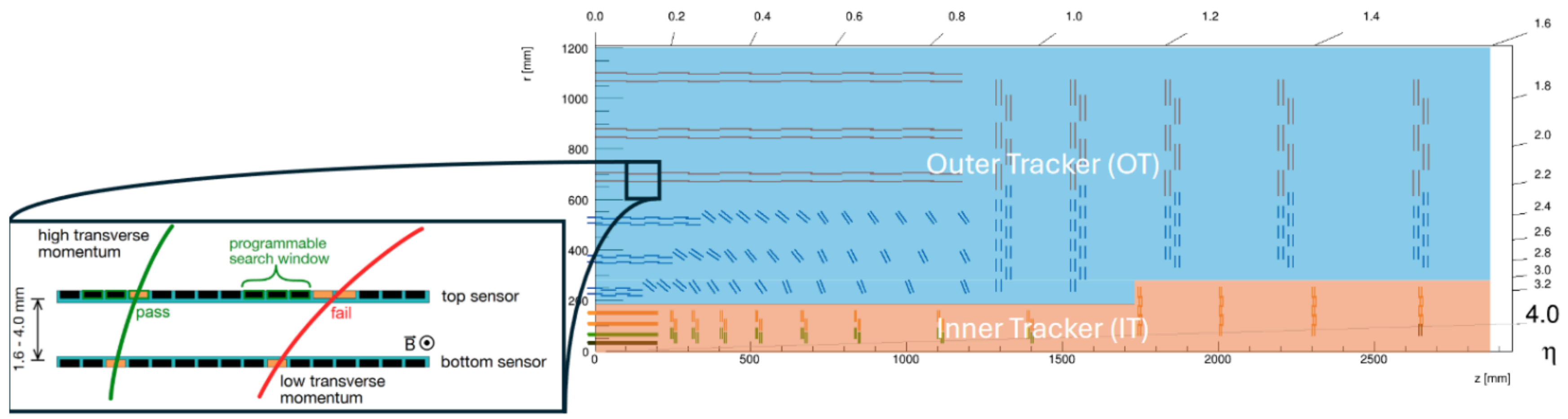


Figure 1: A qualitative representation of the expected Phase-2 CMS tracker geometry [3].

Can be used to seed tracking in outer layers, an approach that can be offloaded to GPUs

(a) LS (b) T3 (c) T5 (d) pT3 (e) pT5

Figure 2: A qualitative representation of the different objects created by the LST algorithm [10].

"Line Segment Tracking" gives comparable physics performance, and also extends capabilities for highly displaced tracks

LST: https://arxiv.org/abs/2407.18231

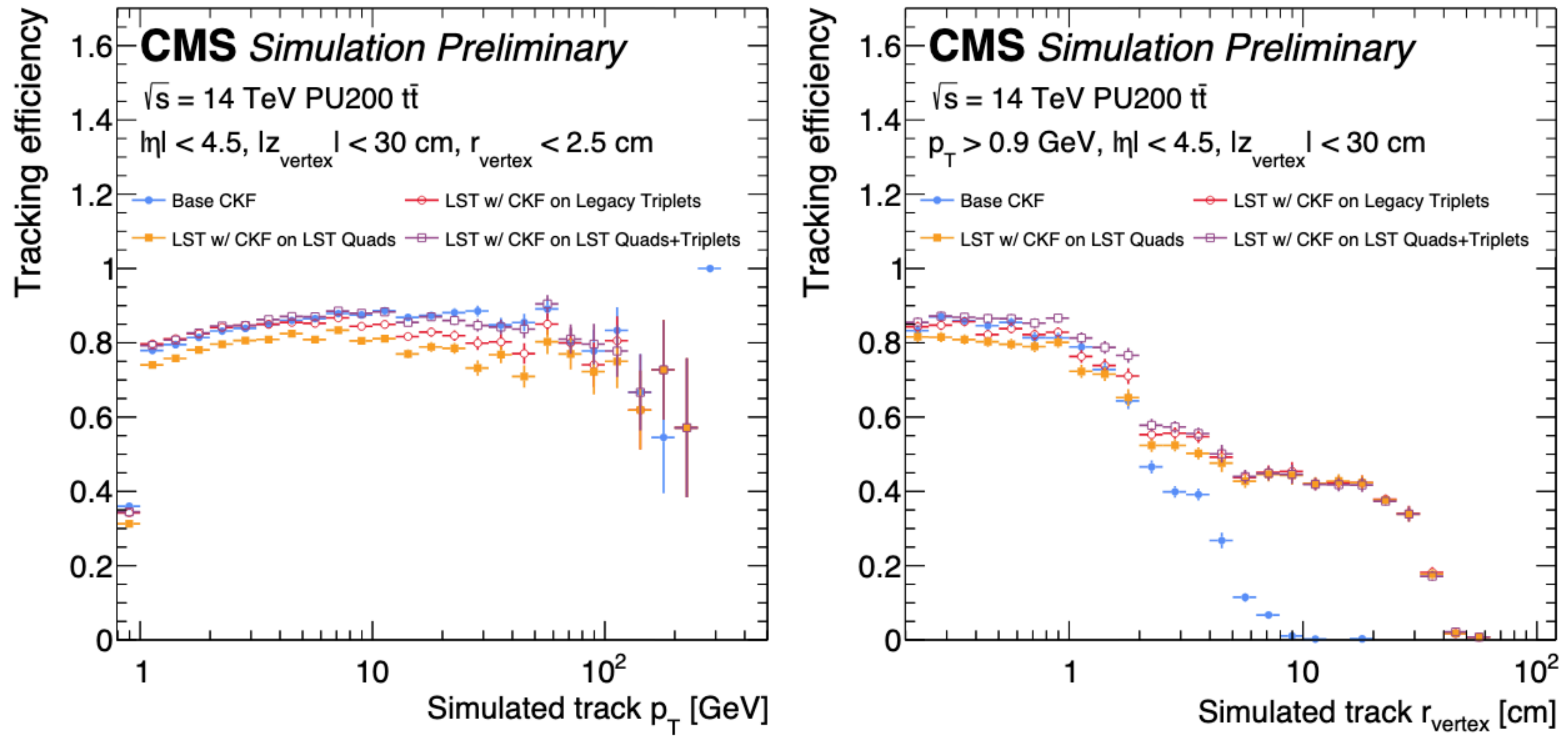


Figure 3: The tracking efficiency is shown for Base CKF (blue), LST with CKF on Legacy Triplet (red), LST with CKF on LST Quads (orange) and LST with CKF on LST Quads+Triplets (purple) as a function of the simulated track p_T (left) and r_{vertex} (right) [10].

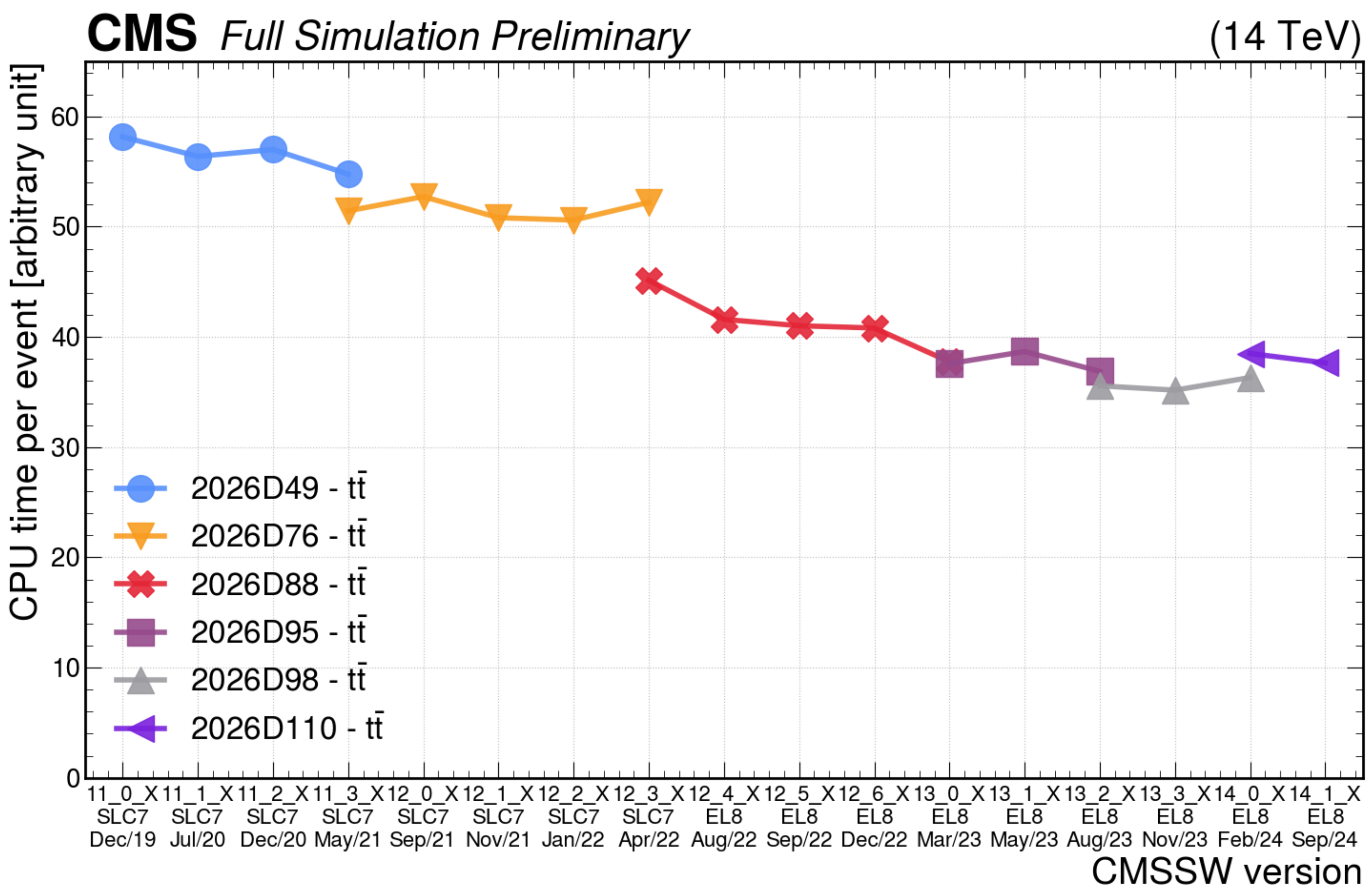
Code is now being integrated into CMSSW
Final computing performance still TBD

Many other ongoing R&D efforts, e.g., ML based reconstruction

CMS (full) simulation

Continuous optimizations → 35% CPU reduction in 5 years

Thanks to various optimisations including FastSim techniques CMS FullSim runs 3.5 × faster than G4 default



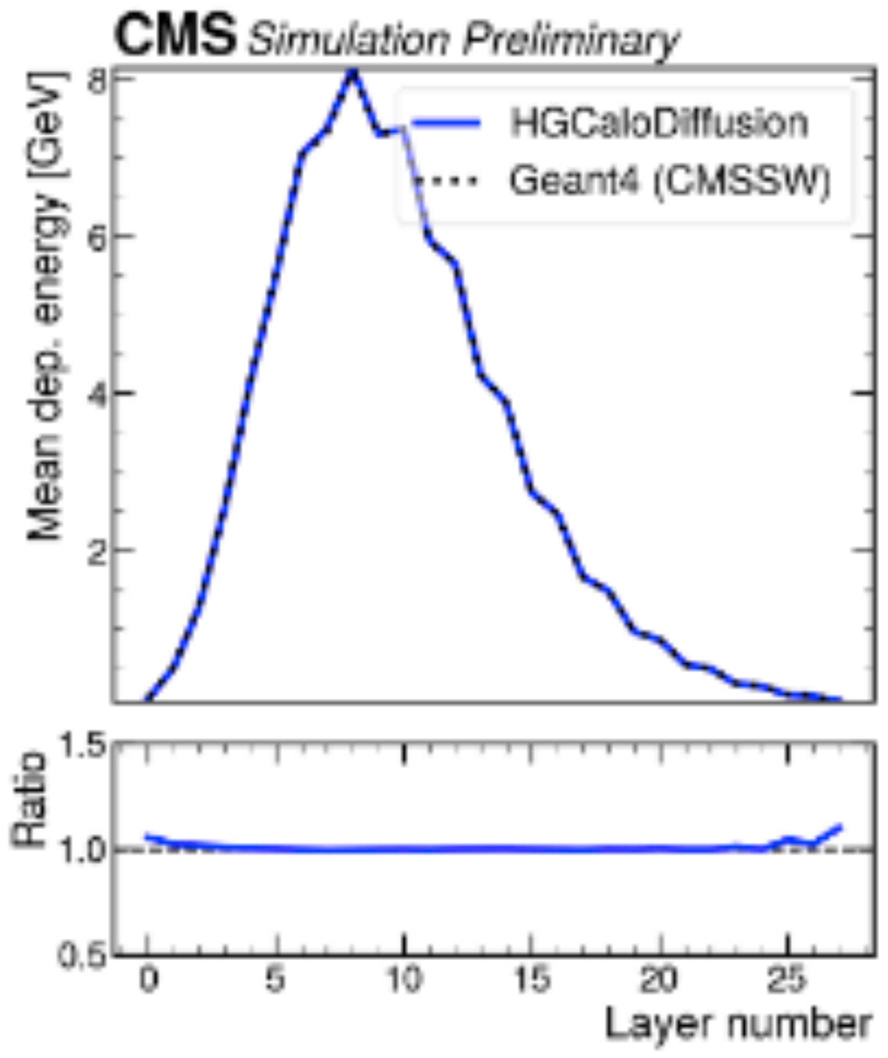
Historical trends of Full Simulation CPU time performance of 14 TeV ttbar process for different Phase-2 geometry design updates (2026Dxx). The average CPU run time per event in relative units of the event simulation is shown for 500 events on single threaded jobs. Main improvements are connected with the Geant4 migration from 10.4 to 10.7 (CMSSW 11_3_X), to 11.1.1 (CMSSW 13_1_X) and to 11.1.2 (CMSSW 13_3_X), updates of the HGCAL and Muon geometry (CMSSW 12_3_X), the change of the computing platform operating system from CentOS 7 (SLC7) to AlmaLinux 8 (EL8) (CMSSW 12_4_X) and the usage of LTO (Link time optimization) build method (CMSSW 13_0_X). Some slowdowns relate to addition of more detailed geometries. The last two points CMSSW 14_0_X and 14_1_X are used in 2024 data taking and MC production for Run-3. During the period of nearly 5 years between the versions 11_0_X and 14_1_X the CPU time has improved for the ttbar process by 35 %.

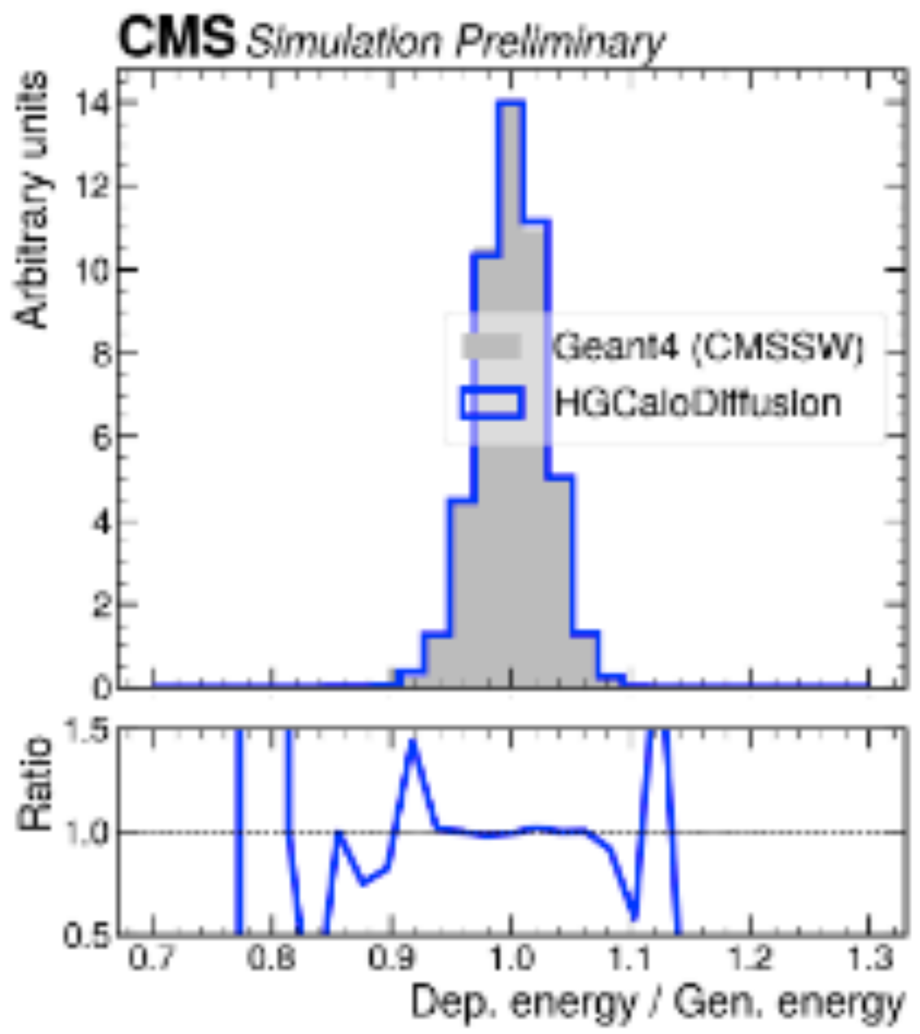
Simulation developments

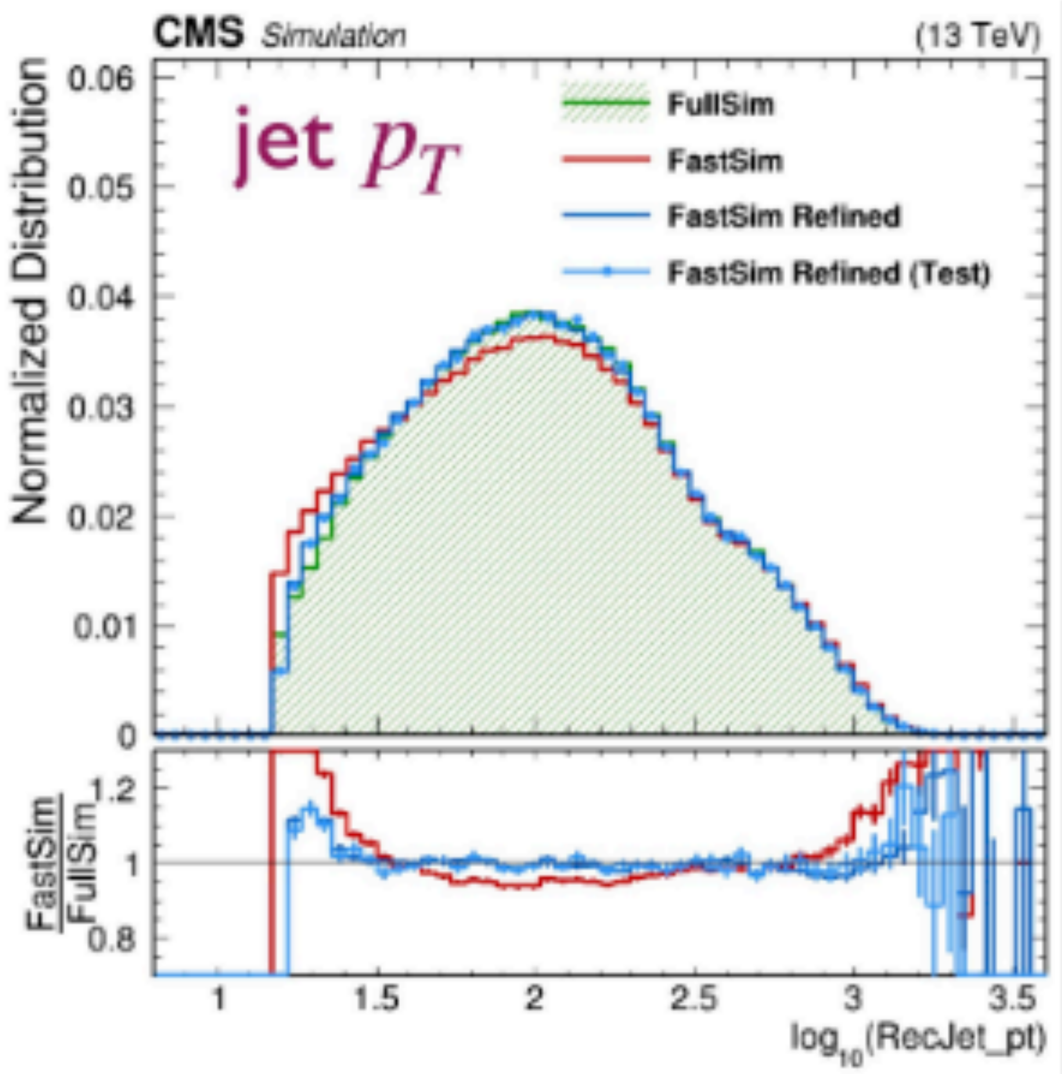
HGCAL Simulation with ML

- Full simulation twice as slow as current calorimeter
- Testing CaloDiffusion model to generate showers in HGCAL

  Amram, Pedro, PRD108 (2023), 072014
- Preliminary results: Good agreement in several variables







Improving fast simulation

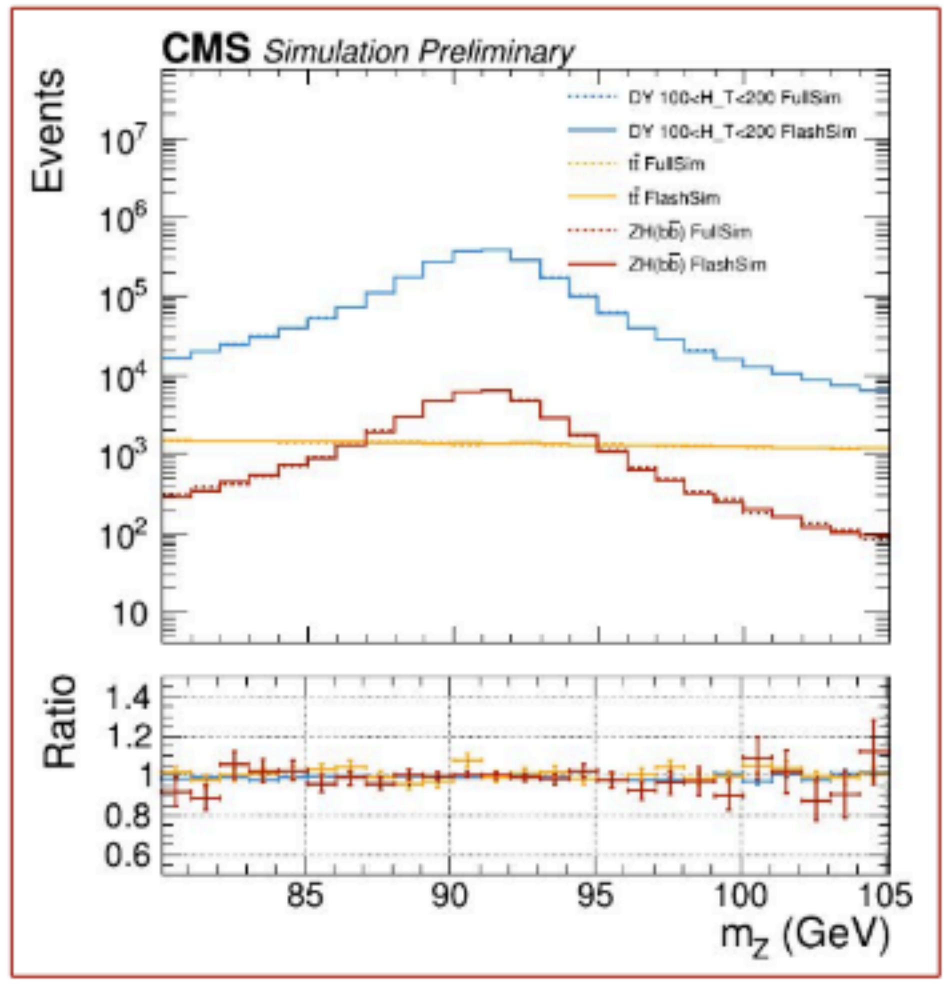
- CMS FastSim: simplified geometry, fast particle propagation, fast tracking, analytical interaction models
- Apply ML to FastSim to improve agreement with FullSim
  - Use same scale factors for both simulations giving 10x speed-up in simulation
- Prototype in place for Run 3 production

FlashSim

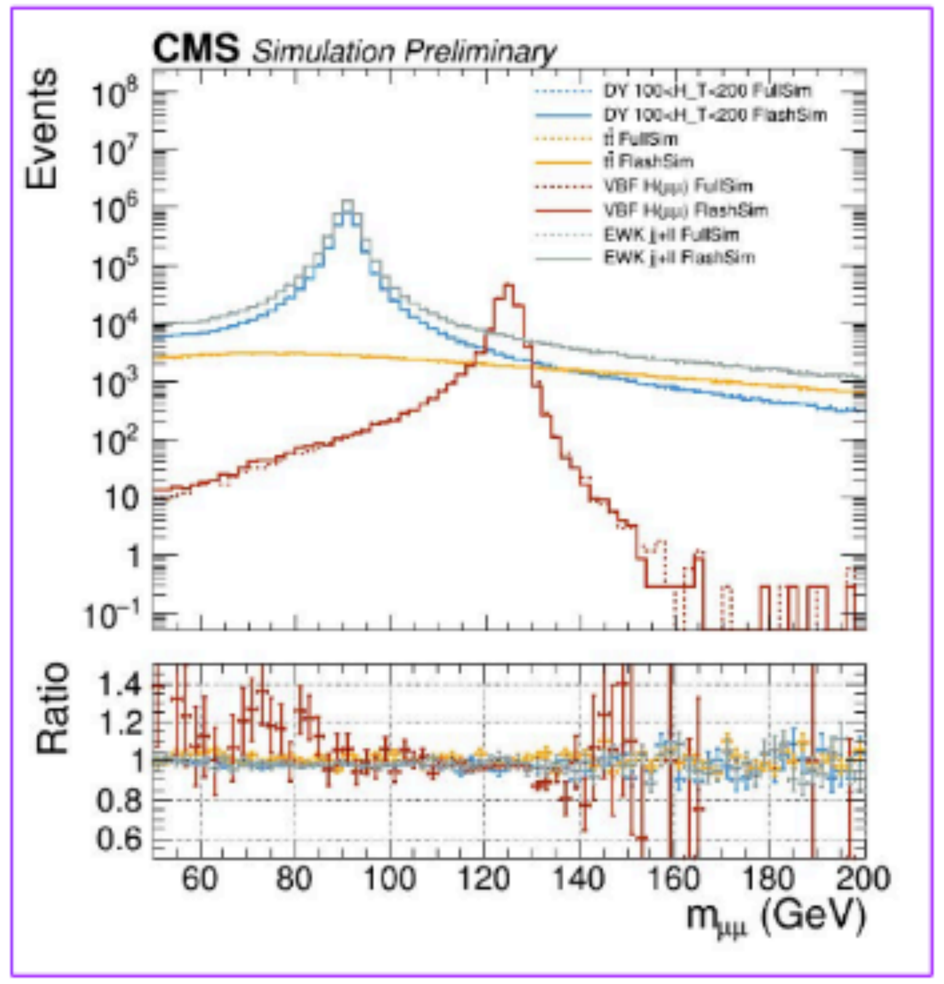
FlashSim: end-to-end framework using Normalizing Flow

- Trained on Geant4 FullSim
- writes directly to nanoAOD
- 30 — 3000x faster than FastSim
- 300 — 30000x faster than FullSim

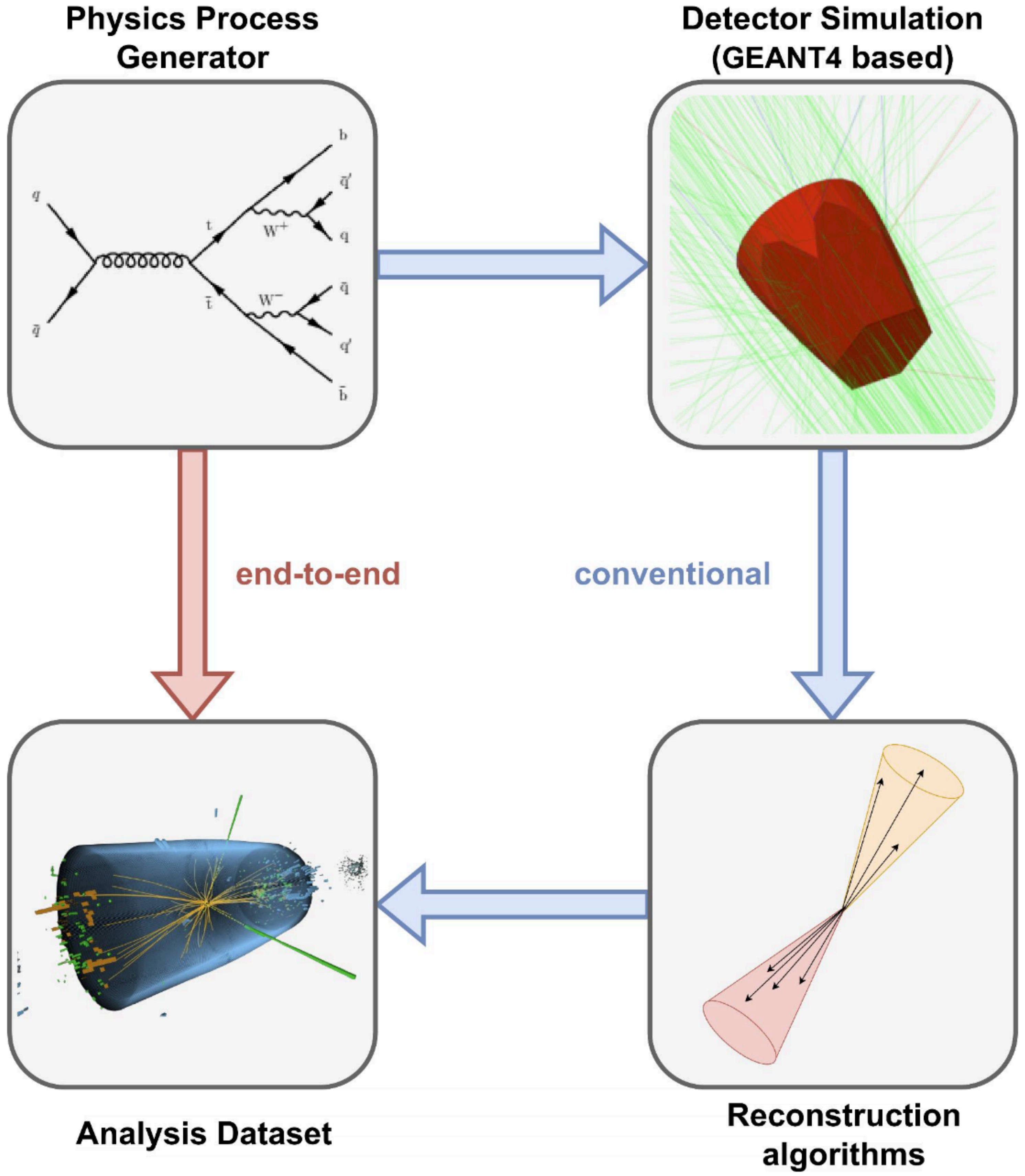
DP Note 2023-003



Z(μμ)H(bb)



VBF H→μμ

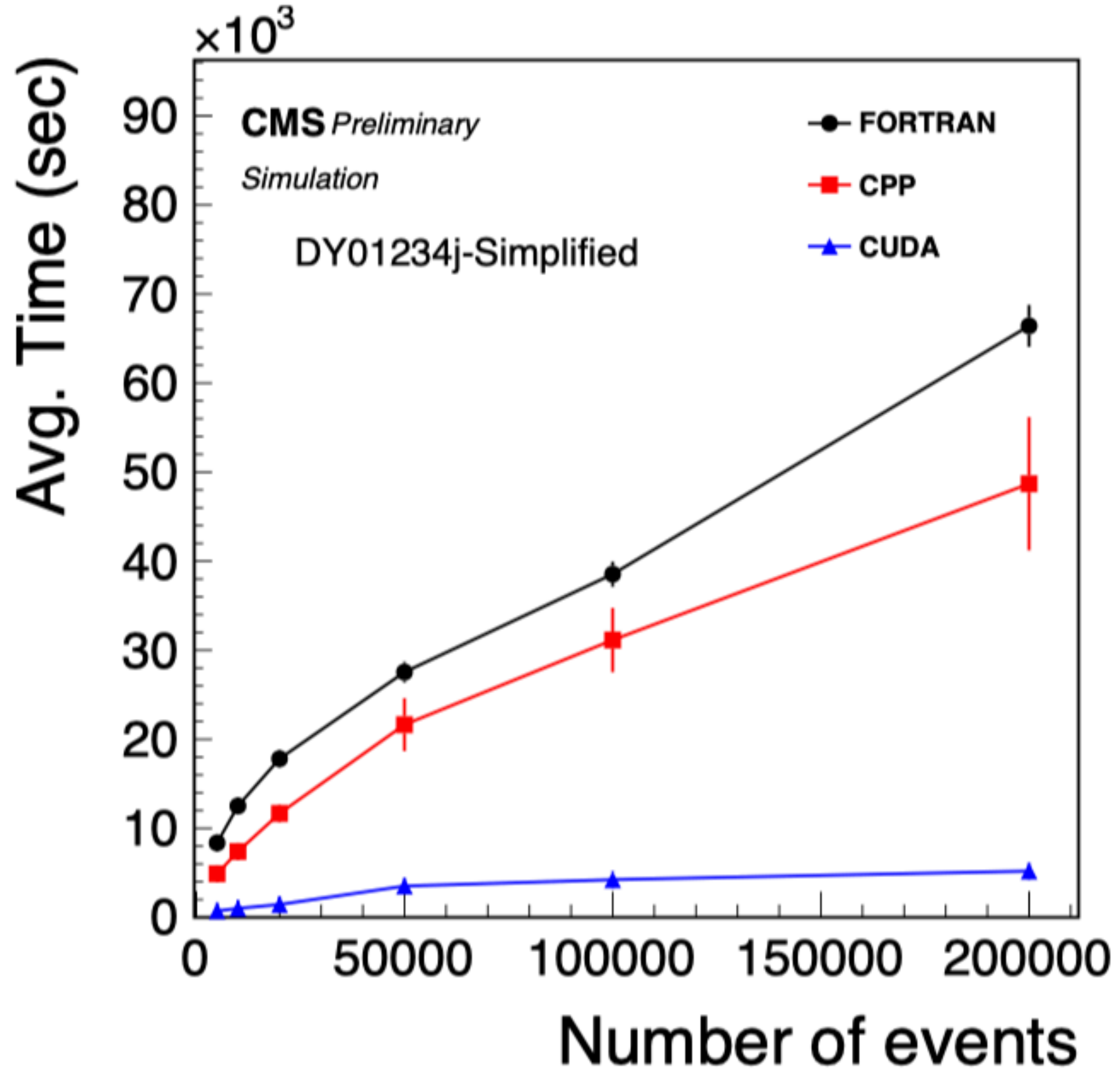


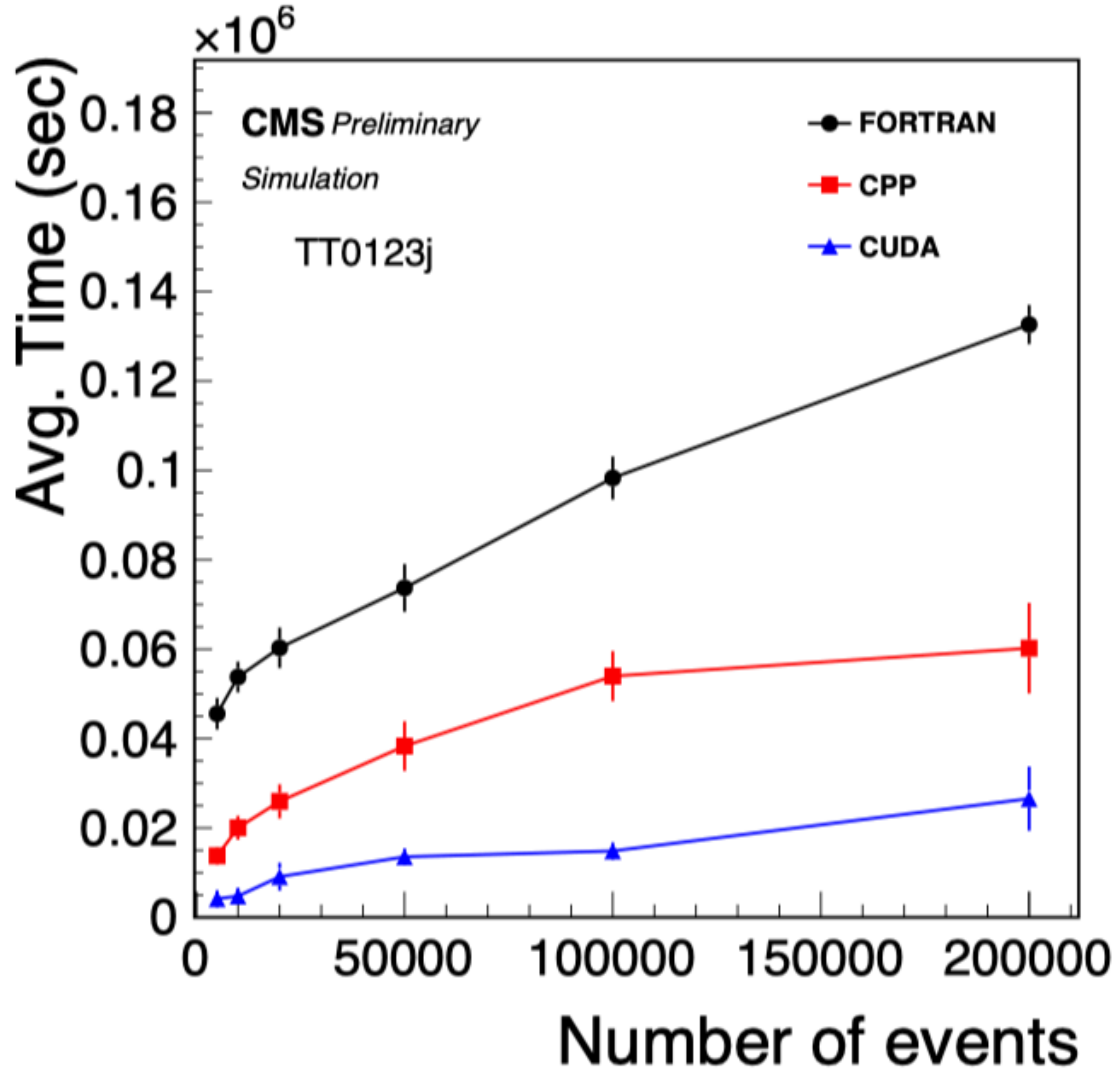
Promising development for Run 4

Event generation on GPUs

- Event generation expected to use 10-20% of CPU resources without R&D
- Dominated by calculation of higher order matrix elements
- MadGraph, most-widely used FW in CMS, developed Madgraph4GPU

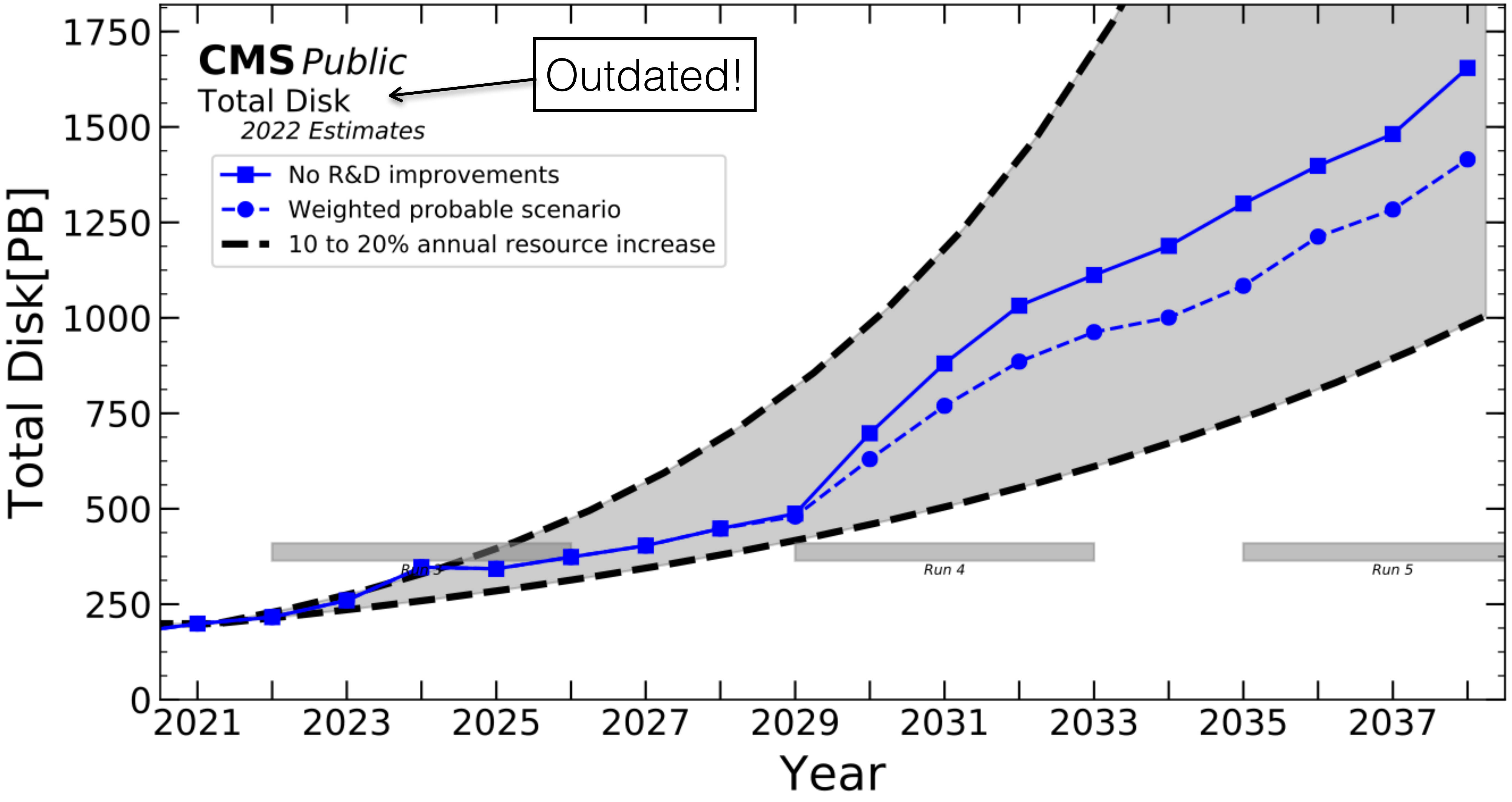
Testing Maggraph4GPU in CMS gen FW:
https://cds.cern.ch/record/2914584?ln=en

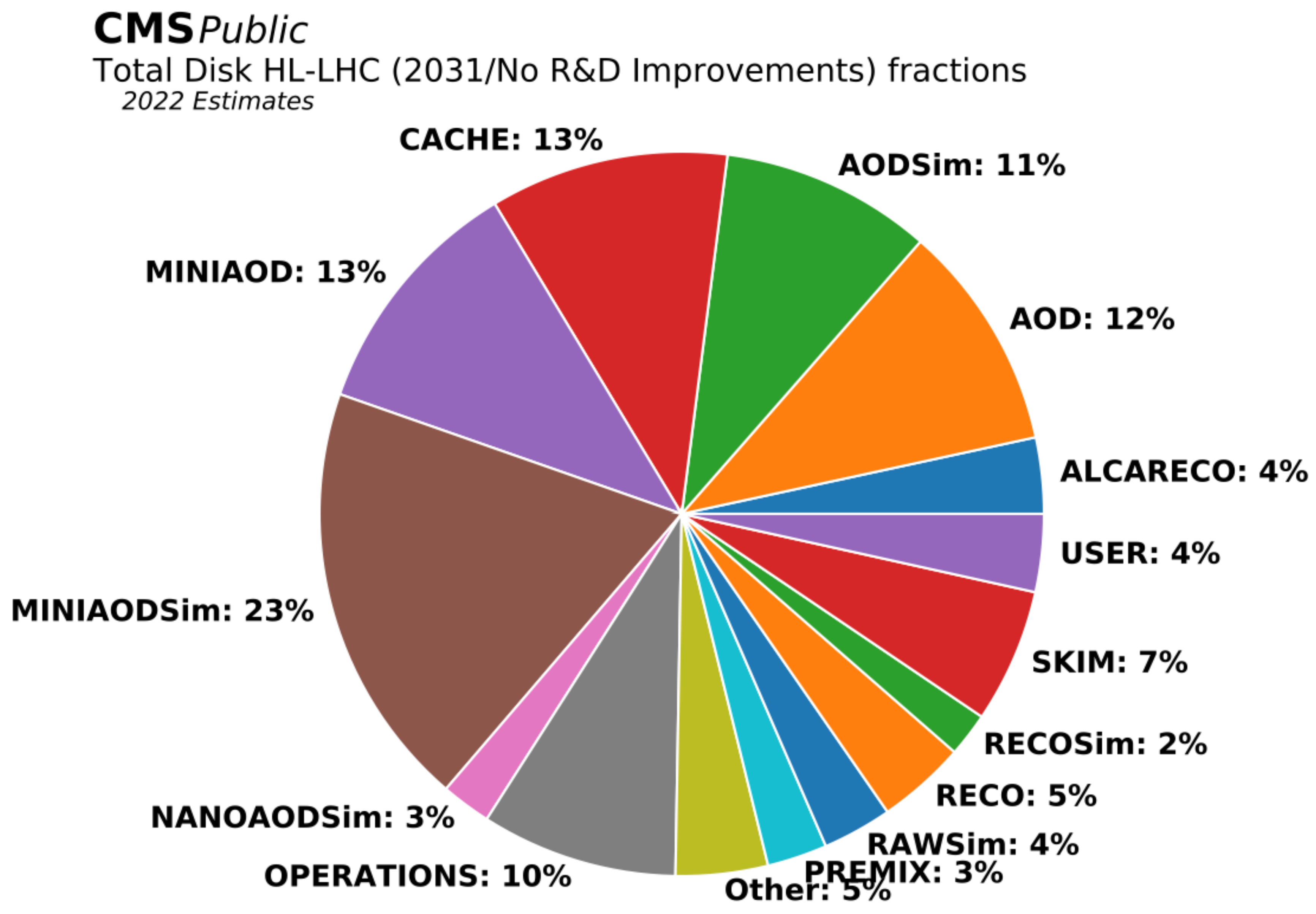




Substantial speed-up both in “gridpack” (matrix element) and event generation

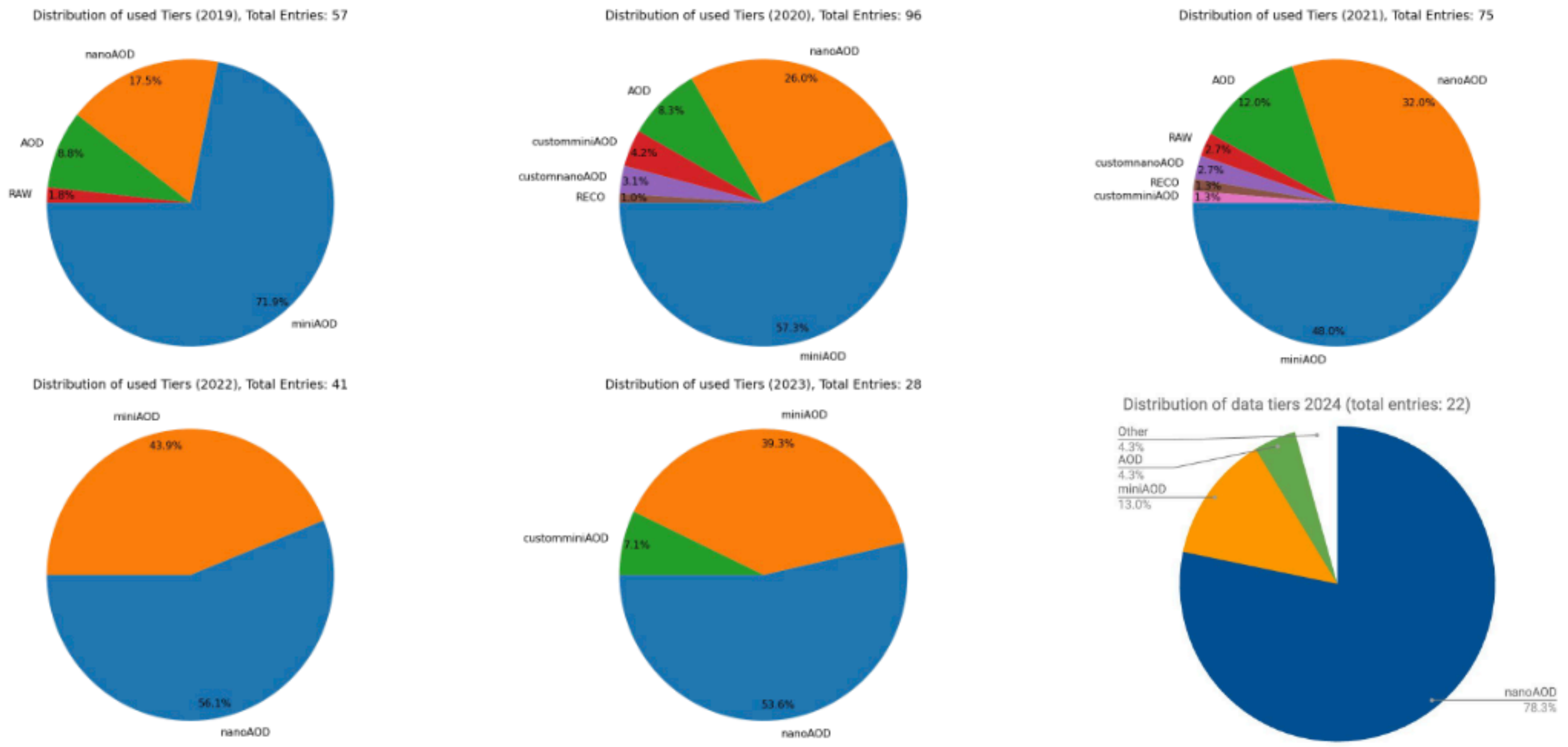
Phase-2 Disk projections





nanoAOD Adoption

AOD: ≈ 500 kb/ev → miniAOD ≈ 50 kb/ev → nanoAOD ≈ 1-2 kb/ev



Majority of analyses now using nanoAOD → important milestone for Phase-2 preparation

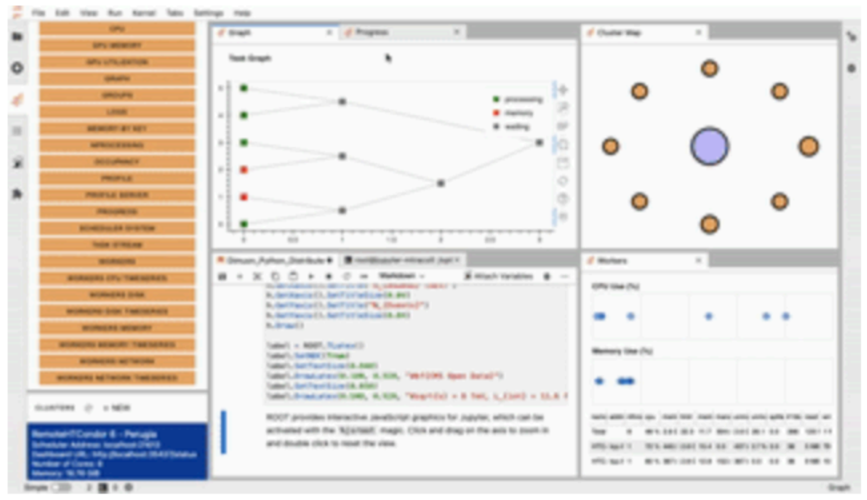
Analysis facilities

Supported by CMS Common Analysis Tools group

In addition to CERN interactive logon services such as the LXPLUS service (Linux Public Logon User Service) and the SWAN (Service for Web based ANalysis) platform, other CMS institutions also provide access to computing resources by providing so-called Analysis Facilities to users with a CERN account that is associated with CMS. These are summarised here.

- Coffea Casa
- Purdue Analysis Facility
- INFN Analysis Facility

A prototype for interactive analysis

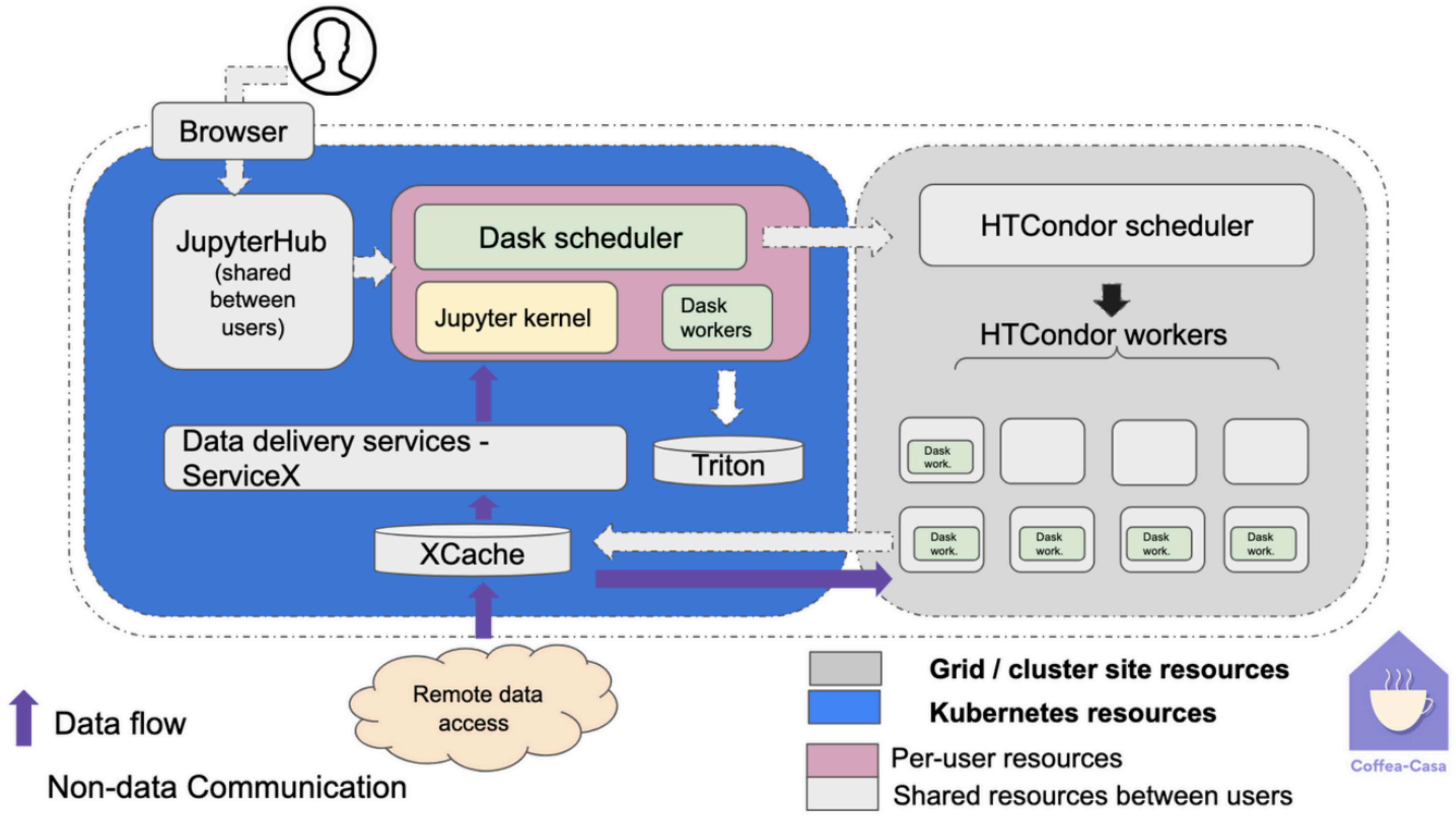


This project is a prototype for a data analysis system, CMS compliant. The main targets are:

1. Reducing analysis "**time to insight**" (training time for newcomers included) with an **interactive** and user-friendly **UI**
2. **Single and easily accessible hub** to reduce the complexity and maintenance of multiple and slightly overlapping solutions
3. **Increasing** the system delivered **throughput** (`evts/s`)

Coffea-Casa Analysis Facility

Coffea-Casa is a prototype analysis facility, which provides services for "low latency columnar analysis", enabling rapid processing of data in a column-wise fashion. This provides an interactive experience and quick "initial results" while scaling to the full scale of datasets.



Offline software & computing CDR for HL-LHC

CMS PAPER CDR-24-XYZ

DRAFT
CMS Paper

The content of this note is intended for CMS internal use and distribution only

October 24, 2024
Archive Hash: none
Archive Date: none

CMS Offline Software and Computing for HL-LHC

Conceptual Design Report

The CMS Collaboration

Abstract

This Conceptual Design Report (CDR) for Phase-2 CMS Offline Software and Computing (O&C) outlines the plan to enable the physics program of the experiment during Phase-2. While certain elements of the current software and computing infrastructure are scalable and sustainable for the foreseeable future, other aspects are not and will need to be adapted for HL-LHC, especially in light of new architectures and ways of provisioning computing resources. New technologies and facilities open up the possibility of exploiting new functionalities, especially in the areas of physics analysis, heterogeneous architectures, storage design, etc. We will step through the various areas of O&C, outlining our plans to evolve the offline software, grid middleware, and computing infrastructure, estimating in each case the positive impacts of success, the risks of failure, and the costs involved in terms of effort. Collaboration will be required not only within CMS but also with external software development communities, WLCG sites, and others. Finally, updated projections on the computing resource needs for Phase 2 will be given, taking into account the likely impact of the various R&D activities.

This box is only visible in draft mode. Please make sure the values below make sense.

PDFAuthor: D. Elvira, F. Ferri, et al.
PDFTitle: CMS Offline Software and Computing for HL-LHC CDR
PDFSubject: CMS
PDFKeywords: CMS, computing

Please also verify that the abstract does not use any user defined symbols

Contents

Contents 7

Contents

1 Introduction . . . 9
1.1 Upgrades of LHC and CMS . . . 9
1.2 Physics overview . . . 9
1.3 Offline software and computing upgrades . . . 10
2 Executive summary and overview . . . 13
3 Management framework . . . 15
4 Computing model . . . 17
4.1 Data formats and event sizes . . . 17
4.2 Tier architecture of computing resources . . . 18
4.3 Data storage strategy . . . 20
4.4 Data processing strategy . . . 21
4.5 Network requirements . . . 22
4.6 Required number of simulated events . . . 23
4.7 Summary of input parameters to the Computing Model . . . 25
5 Core software . . . 29
5.1 Evolution of build tools and continuous integration . . . 29
5.2 Evolution of the data processing software framework . . . 30
5.3 Heterogeneous microarchitectures . . . 32
5.4 Core software-related risks . . . 33
5.5 Summary of core software-related R&D activities . . . 34
6 Event generators . . . 37
6.1 Time performance estimates . . . 37
6.2 R&D to improve computing performance . . . 38
7 Detector simulation . . . 43
7.1 Evolution of the CMS geometry model . . . 43
7.2 Electronic signal simulation and pileup event mixing . . . 43
7.3 Time performance estimates for FullSim and FastSim . . . 44
7.4 R&D to improve physics models and computing performance . . . 44
7.5 Summary of detector simulation-related R&D activities . . . 48
8 Data reconstruction . . . 50
8.1 R&D to improve computing performance . . . 51
8.2 Calorimeter clustering and linking . . . 54
8.3 Data format size reduction . . . 54
8.4 Summary of reconstruction-related and data formats R&D activities . . . 55
9 Workflow management . . . 57
9.1 The current workflow management and system integration . . . 57
9.2 Evolution of workflow management for the HL-LHC era . . . 58
9.3 Impact of ProcessingUnit evolution on workflow management . . . 62
9.4 Summary of workflow management-related R&D activities . . . 62
10 End-user analysis . . . 63
10.1 Requirements of the End-User Analysis Workflow for HL-LHC . . . 64
10.2 Dataset composition and smart caching . . . 66

Planning for CMS-wide review this spring with submission to LHCC next year

Concluding remarks

- Run 3 is now in full swing, surpassing Run 2
  - CMS is continuing to utilize computing resources intensively but efficiently

- Phase-2 preparations continue to ramp up
  - Baseline GPU strategy being extended to offline reconstruction
  - Offline & computing CDR with details on Phase-2 strategy coming this year

- This will be my last Journées LCG-France representing CMS. It's been great working with you all!