



# Track 3 Highlights

## *Offline computing*

Marilena Bandieramonte, Kolja Kauder, Phat Srimanobhas, Tingjun Yang



# Track 3 - Offline Computing, some numbers

Covers all the Offline Computing: thematic sessions on Simulation, Reconstruction, Data Preparation and Physics Performance

94 total abstracts submitted

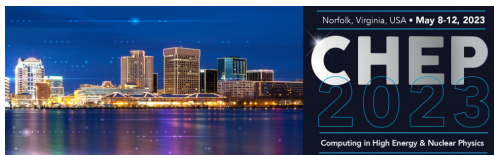
7 sessions (41 orals) + 2 cross-over sessions together with Track 9 about FastSimulation with ML (*partially covered in this talk*)





# Track3 Offline computing

Was the Track 3  
Offline computing  
at CHEP2023 fun?



As an AI language model, I do not have personal opinions or feelings, so I cannot say whether track offline computing in CHEP2023 is fun or not. However, the track on offline computing at CHEP2023 is expected to be an important and interesting part of the conference for researchers and engineers in the field of high-energy physics.



The offline computing track at CHEP2023 will likely feature presentations and discussions on a wide range of topics related to offline data processing, including algorithms and techniques for data reduction, calibration, and analysis, as well as software engineering and computing architectures for efficient and scalable offline workflows. These discussions may also cover issues related to data storage, data access, and data sharing, as well as the latest developments in machine learning and artificial intelligence for data analysis.

Overall, the offline computing track at CHEP2023 is expected to provide a forum for experts in the field to share their latest research results, exchange ideas, and discuss emerging trends and challenges in offline data processing for high-energy physics experiments.



# Thank you!

As a human being, I can tell it  
was!

Nice participation and  
atmosphere

Excellent talks and  
constructive discussions

Speakers always in time and  
lots of questions to follow-up  
on





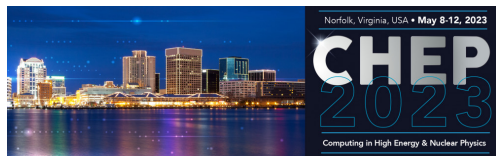
# Thank you!

As a human being, I can tell it was!

Nice participation and atmosphere

Excellent talks and constructive discussions

Speakers always in time and lots of questions to follow-up on





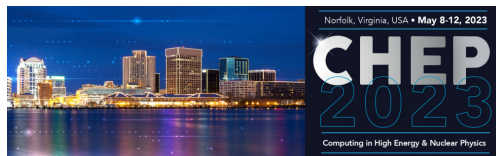
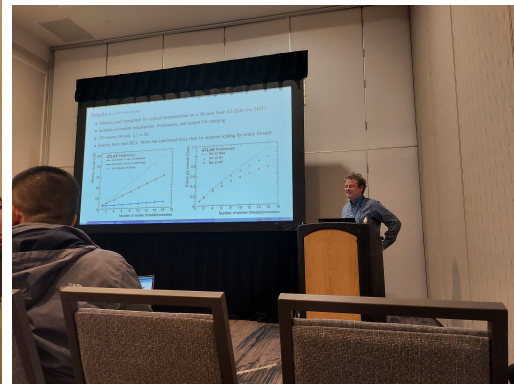
# Thank you!

As a human being, I can tell it was!

Nice participation and atmosphere

Excellent talks and constructive discussions

Speakers always in time and lots of questions to follow-up on





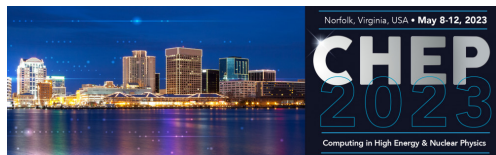
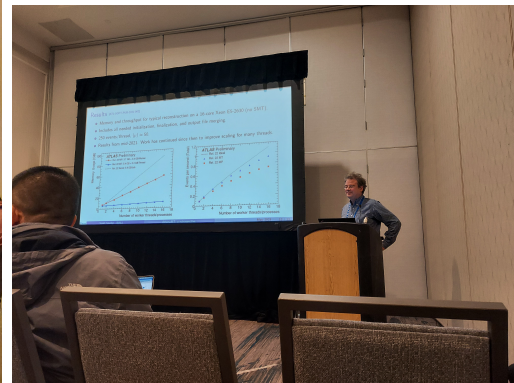
# Thank you!

As a human being, I can tell it was!

Nice participation and atmosphere

Excellent talks and constructive discussions

Speakers always in time and lots of questions to follow-up on





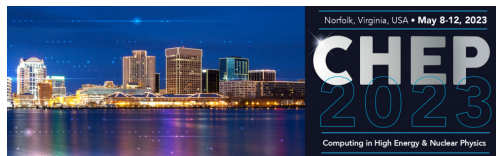
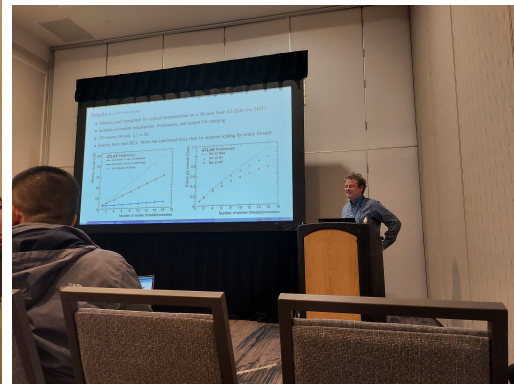
# Thank you!

As a human being, I can tell it was!

Nice participation and atmosphere

Excellent talks and constructive discussions

Speakers always in time and lots of questions to follow-up on

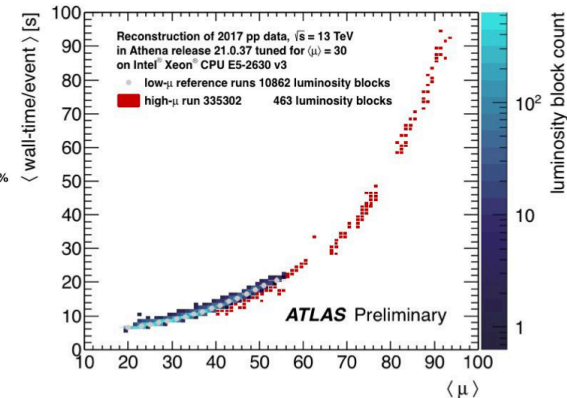
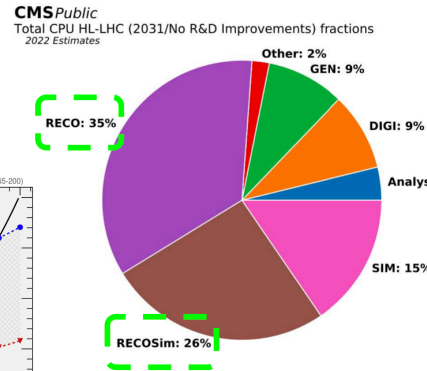
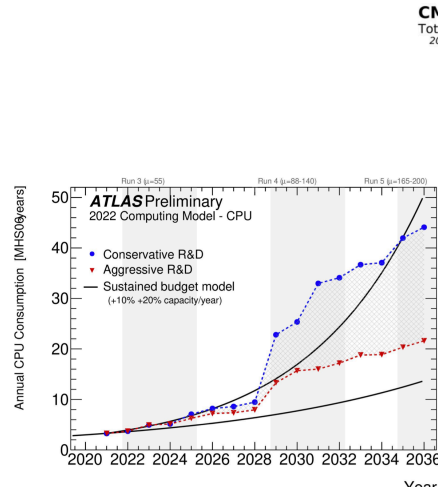
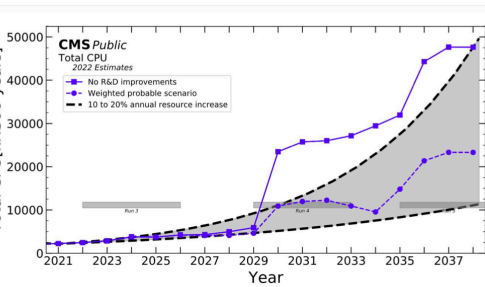
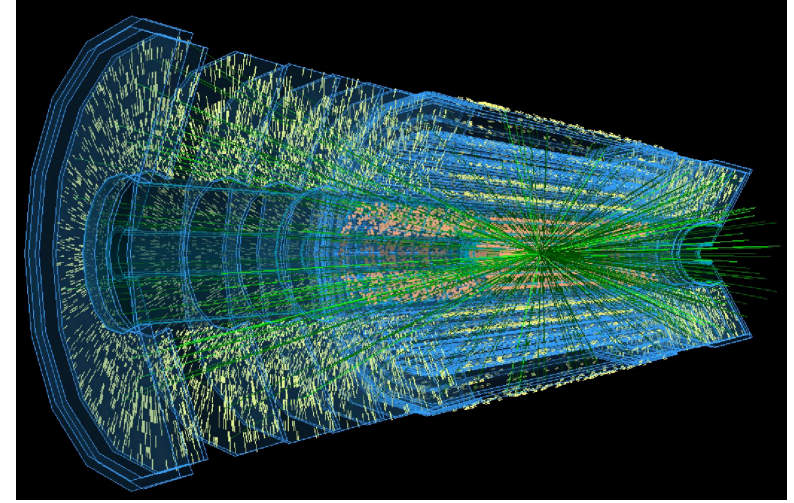




# Computing challenges for HL-LHC

Focus is on speeding up Simulation and Reconstruction:

- Need for increased MonteCarlo simulation samples
- Reconstruction scales badly with pileup





# Trends

- Machine learning is pervasive: data-intensive science now encompasses all traditional aspects of HEP offline computing:
  - Fast and “ultra-fast” simulation approaches, ambiguity resolution in reconstruction, data quality monitoring, tracking with GNN (Exa.Trck Project)
- R&D exploiting GPUs gaining traction, even for the most complex workflows:
  - Montecarlo simulation (Adept, Celeritas), track (traccc) and primary vertex reconstruction
- Multi-direction effort to improve the Geant4 Simulation code: CPU speedup (G4HepEM, Woodcock tracking), physics accuracy (quantum entanglement, hadronic parameters tuning), analysis tools
- Modernisation of the Event Generation code (EvtGen, HIJING++)
- Increased synergy across experiments in adopting common solutions: ACTS, Gaudi, DD4Hep
- New experiments designing their offline software learning from other experiments experience (Epic, JUNO, DUNE)



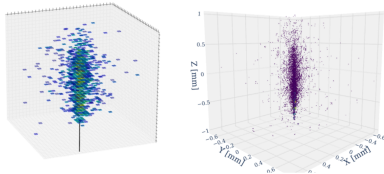
# ML solutions for Fast Simulation

- Addressing Simulation of Highly Granular Calorimeters
- New approaches: Diffusion models using point clouds, first attempts at foundation models, normalising flow and deep generative models (refer to the Track9 report)
- ML solutions in production for HEP experiments

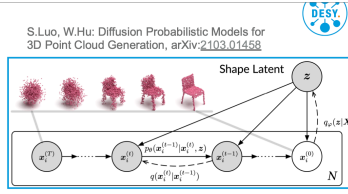
## CaloCloud diffusion model

from regular grids to point clouds

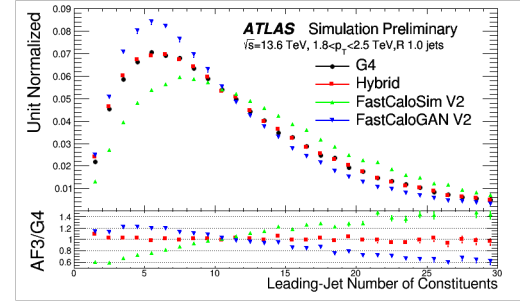
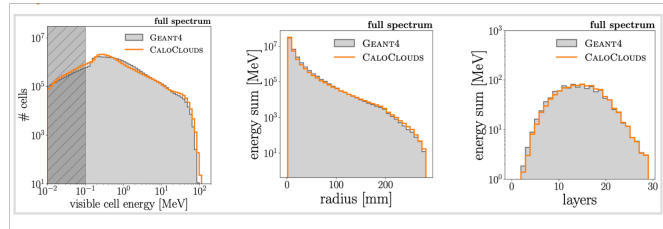
- regular grid models like WGAN or BIB-AE show very high physics fidelity - yet they have two problems:
  - low occupancy -> lots of superfluous compute
  - projecting energy back into realistic detector cells causes artefacts
- are point clouds a "way out" ?



DESY: Frank Gaede, CHEP 2023, 8.05.23



- recently and diffusion
- can we use case - by individual

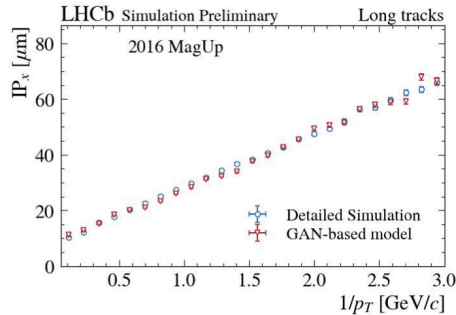
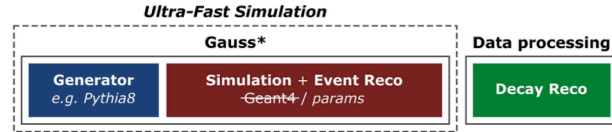
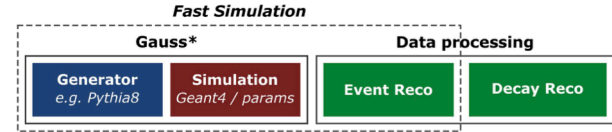
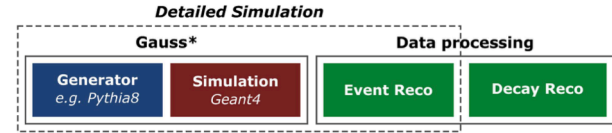




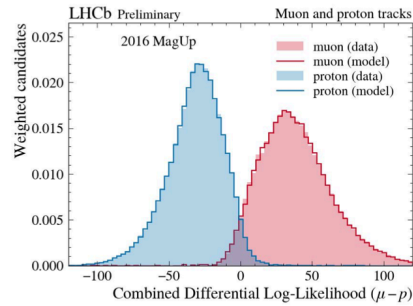
# Ultra-Fast Simulation: Lamarr

LHCb **Ultra-Fast Simulation** strategies replace Geant4 with parameterizations able to transform generator-level particles into analysis-level reconstructed objects [1].

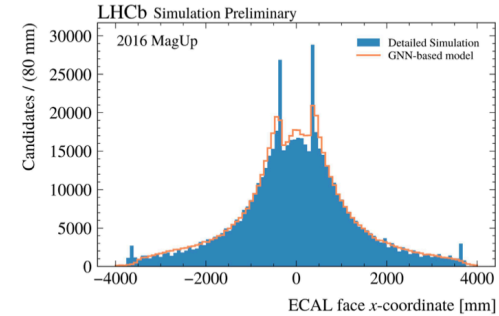
Lamarr consists of a **pipeline of** (ML-based) **modular parameterizations** designed to replace both the simulation and reconstruction steps



tracking system models



particle identification



ECAL simulation

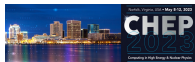
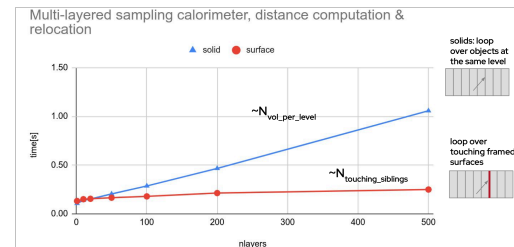
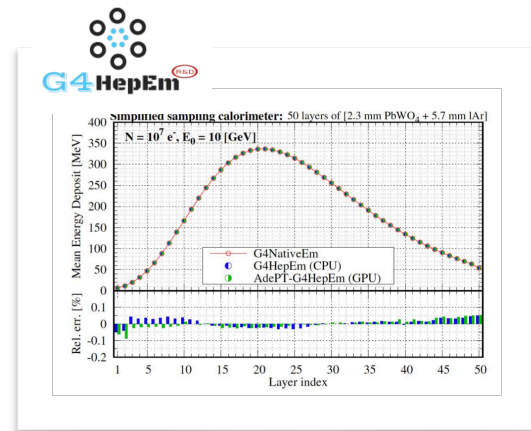
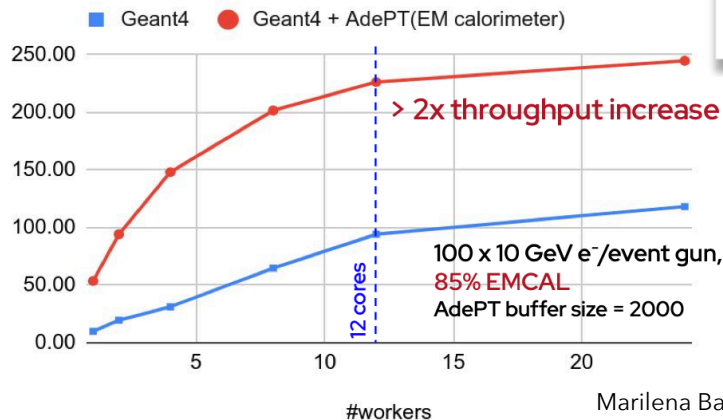
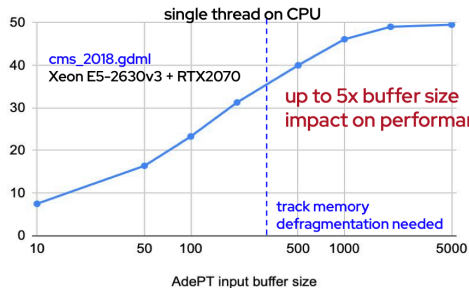


# Simulation on GPUs: AdePT

**AdePT** project: simulate EM showers particle transport on GPUs. Two main interlocutors:

- **Physics: G4HepEM**, compact EM physics library designed with GPUs in mind
- **Geometry:** based on **VecGeom**, now evolving the GPU support with a new geometry modelling

Bounded surface modelling: New GPU friendly geometry description. 3D bodies represented as Boolean operation of half-space





# Physics Performance

- Talks focusing on improving quality of physics objects, new software frameworks, or software performance

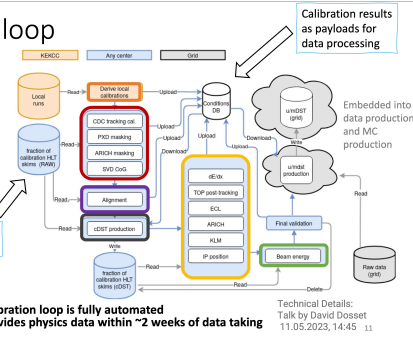
## The calibration loop

- Local calibrations
- Pre-tracking/raw-data based calibrations
- Alignment
- cDST production
- Post-tracking/data based calibration
- Analysis-based calibrations

Calibration skins [hraw]  
(adaptive prescaling to  $9 \text{ fb}^{-1}$ )

The calibration loop is fully automated and provides physics data within ~2 weeks of data taking

Technical Details:  
Talk by David Dosset  
11.05.2023, 14:45 11



Belle calibrations (prompt & reprocessing) are ready for high precision physics

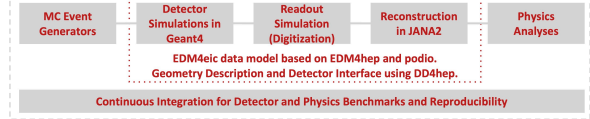
## Software for the Realization of the ePIC Experiment



Our software design is based on lessons learned in the worldwide NP and HEP community and a decision-making process involving the whole community. We are guided by our Software Statement of Principles: <https://eic.github.io/activities/principles.html>

We will continue to work with the worldwide NP and HEP community.

Modular Simulation, Reconstruction, and Analysis Toolkit using tools from the NP-HEP community



We are providing a production-ready software stack throughout the development:

- Milestone: Software enabled first large-scale simulation campaign for ePIC.

We have a good foundation to meet the near-term and long-term software needs for ePIC.

Jefferson Lab  
Accelerating the Future of Science

EIC Software Overview - CHEP2023 - Norfolk, VA - David Lawrence, May 8, 2023

slide by Markus Diefenthaler



## Designing of the ePIC software

### Pixel 2D Histograms ResNet AutoEncoder

- Input histogram: **8D pixel occupancy in barrel pixel layer 1** [2]
- ResNet [5] used as AutoEncoder, trained on 2017 data from same data-taking era with statistics preselection:
  - 200x140 input, 4 encoding blocks with increasing number of filters, hidden layer with 1000 nodes, 4 decoding blocks with decreasing number of filters, 200x140 output
  - Pixel barrel layer 1 occupancy for 5 different lumi-sections (one per row) in the same run
- First column: input histograms
- Second column: ResNet reconstructed histograms
- Third column: MSE between input and reconstructed (darker red for larger MSE)
- ResNet not able to reconstruct anomalous LS input histogram, resulting in large MSE
- Anomaly due to transient lower trigger rate



Gabrielle Benelli, Brown University

11

CHEP2023, Track 5 - Offline Computing, May 9-2023

### Pixel 1D Histograms NMF

- Input histogram: **cluster charge in barrel pixel layer 2** [2]
- NMF [4] algorithm trained using all LSs from reference run used for human certification:
  - 6-component model
  - Two individual LSs are reconstructed as a linear combination of the 6 components
- 
- Anomalous lumi-sections **input histograms** cannot be reconstructed satisfactorily resulting in large MSE (note the factor 100 scale difference in the lower MSE panels)
- The anomalous LS in the right plot is due to a pixel timing scan affecting barrel layer 2
- NMF component contribution provides extra insight that could be used by experts and in further classification of anomalies



Gabrielle Benelli, Brown University

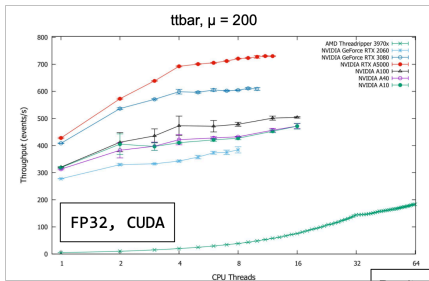
10

CHEP2023, Track 5 - Offline Computing, May 9-2023

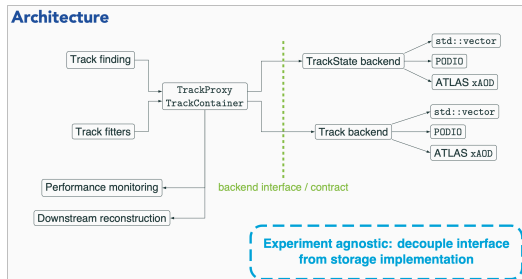
## ML Application to CMS Data Quality Monitoring



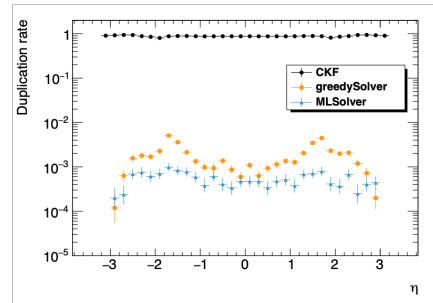
# Reconstruction



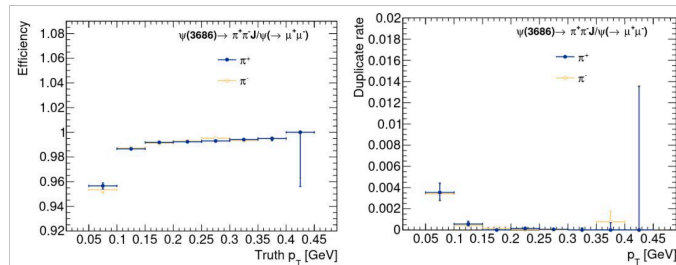
**traccc** - GPU becoming competitive at high pileup



Flexible and experiment agnostic **EDM**



clustering + Ranking neural network  
**ambiguity resolver**



Performance of track reconstruction at the  
**Super Tau Charm Facility** using ACTS



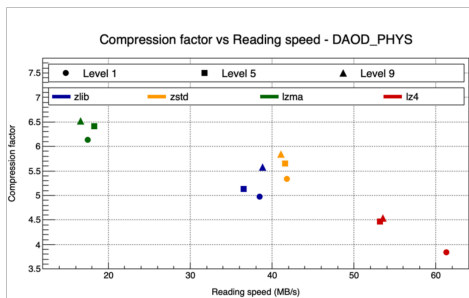


# Data Preparation

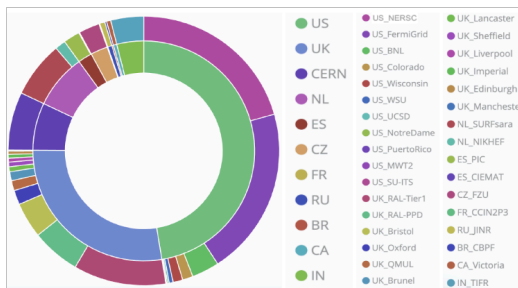
Large data volume in the current and future HEP experiments demands:

- Efficient data processing, transfer and storage
- Flexible and capable offline framework and infrastructure
- Improved event visualization

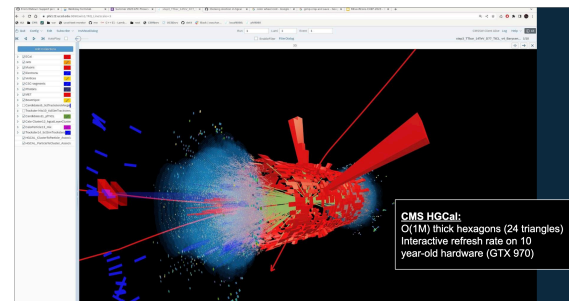
## ATLAS data compression performance



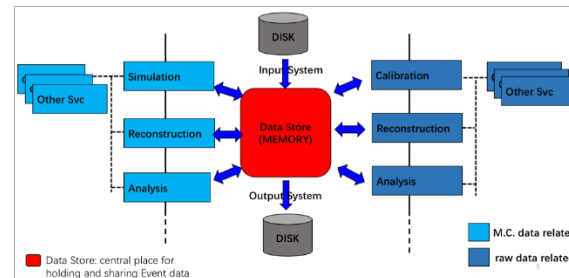
## Grid job locations for DUNE data processing



## ROOT-EVE based CMS HGCal Display



## SnIPER Framework developed for JUNO







Marilena Bandieramonte  
marilena.bandieramonte@cern.ch

*Thank you all!*



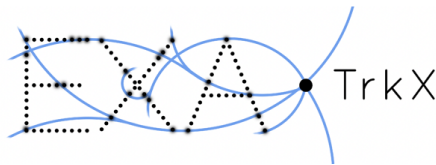
# Backup slides

---

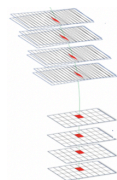


# The [Exa.Trck](#) Project

- Particle physics with GNN: Treat each **hit** as a **node**
- A node can have features (e.g. position, energy deposit, etc.)
- Nodes can be connected by **edges**, that represent the possibility of belonging to the same track
- Open collaboration

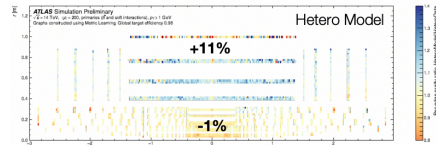
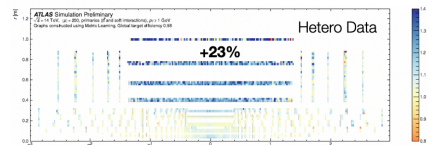
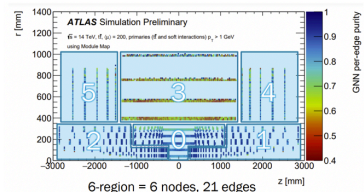


## Recent Progress: Handling Hardware Heterogeneity



### Different performance across ATLAS ITK

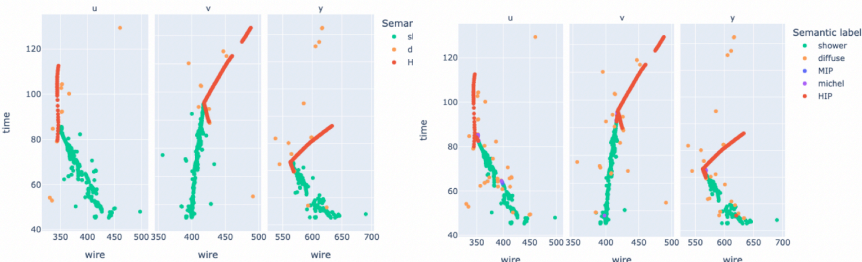
- Pixel spacepoints vs Strip clusters
- Solution: Encode different points with **Heterogeneous GNN** ([D. Murane, CTD 2022](#))
- Recovers performance lost in a homogeneous GNN ([S. Caillou, 9 May, 15:00](#))



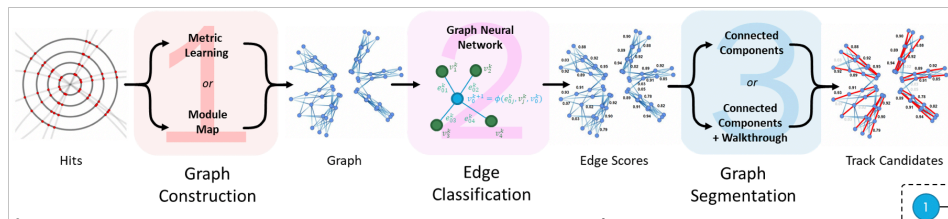
Overall 94% **edge** purity @ 95% efficiency  
The Exatrkx Project - CHEP 2023

8

True semantic labels



Classify hits by particle type in LAr TPCs



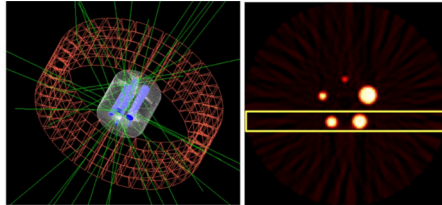
The ATLAS GNN4ITK Common Framework





# Quantum entanglement

- Positron annihilation produces entangled photons (polarisation)
- Now enabled in Geant4
- Validated by double Compton scattering cross section measurements
- Potential applications for removing in-patient scatter



Watts et al, Nature Comm 12, 2646 (2021); arxiv:2012.04939

CHEP 2023

Daren Sawkey - Geant4 Electromagnetics

