Track 2 Summary

Adriano Di Florio (INFN Bari), Dalila Salamani (CERN),
Jennifer Ngadiuba (Fermilab), Sophie Berkman (Fermilab)

ACAT 2022
24-28 October 2022
Villa Romanazzi Carducci, Bari, Italy

Track 2 Overview

- **112 contributions received for 36 oral slots**
  - ALL contributions were accepted as all extremely interesting
  - Thanks to the poster sessions we are able to hear about and discuss all of them!
- A **broad range of topics** being covered
- This summary attempts to find the **common goal and main direction…**

The common goal

And we are here discussing **all possible angles** to achieve this goal

Accelerate scientific discoveries by boosting computing efficiency

Detector monitoring

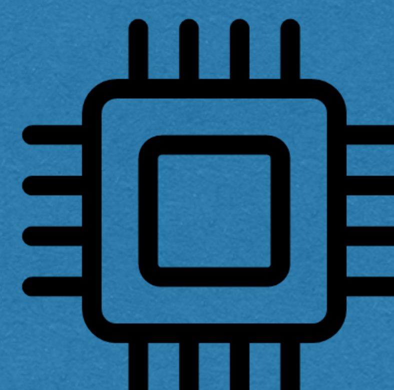
Trigger

Calibrations

AI

\+

Heterogeneous computing



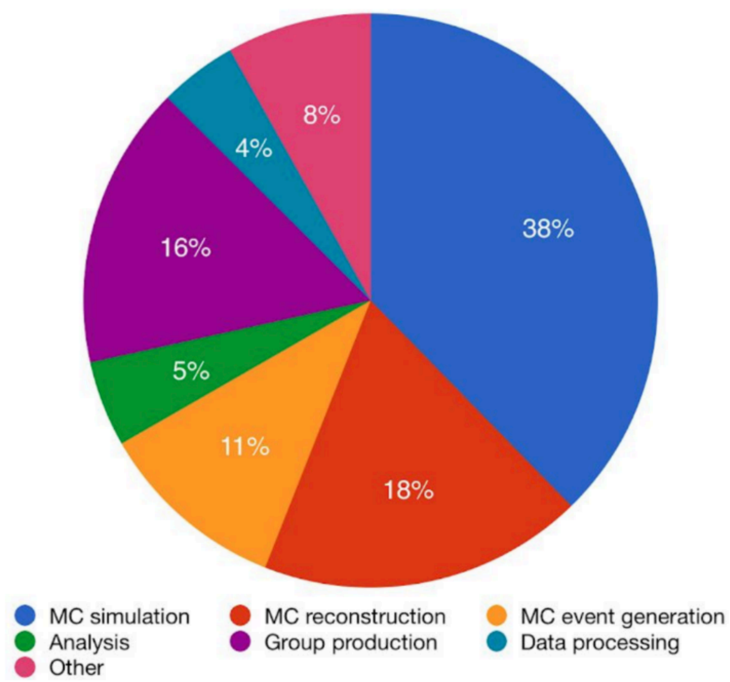
Simulation

Reconstruction

Analysis

+ scalable analysis tools, workflows and facilities (Track 1)

Accelerating simulation



- MC simulation large part of computing
- Speed up:
  - Train ML model on small dataset
  - Draw majority of samples form ML model
    - ➔ Amplify original data set
    - ➔ Significantly faster

Catmore et. al. **ATLAS HL-LHC Computing Conceptual Design Report,** CERN-LHCC-2020-015 ; LHCC-G-178

Butter et al.: **Amplifying Statistics using Generative Models:** NeurIPS ML4PS 2020, 2008.06545

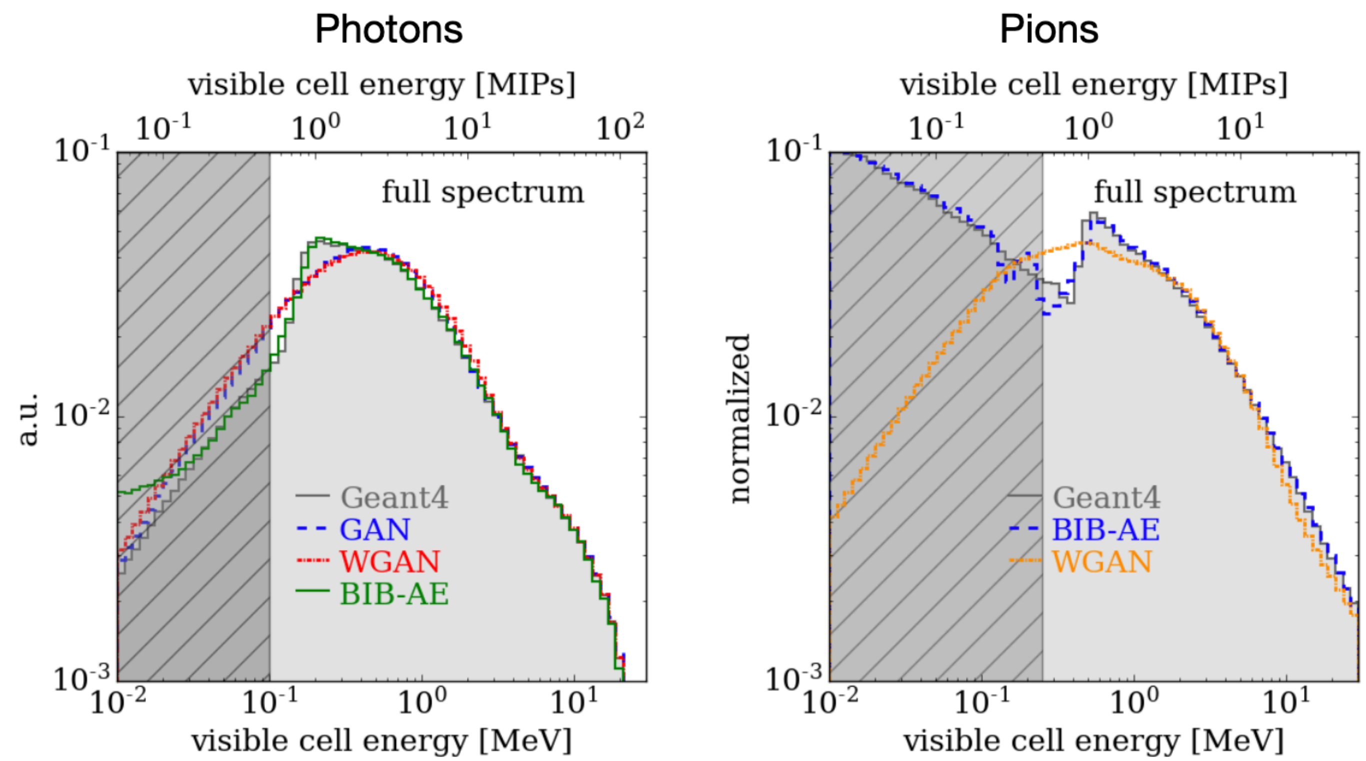
80-90% CPU time of simulation spent on calorimeter!

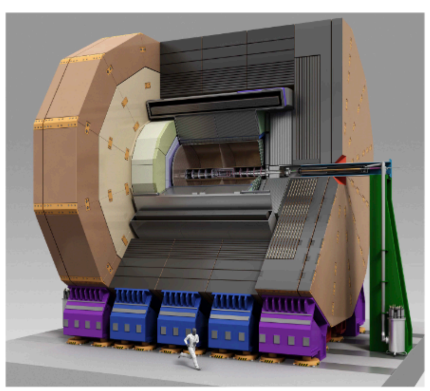
S. D. Diefenbacher plenary talk on Monday: "Generative Models for Fast (Calorimeter) Simulation"

Accelerating simulation

- Different **generative models** being explored:
  - Variational Autoencoders (VAE)
  - Generative Adversarial Network (GAN)
  - Normalizing Flows
- And different **data representations:**
  - images and point cloud

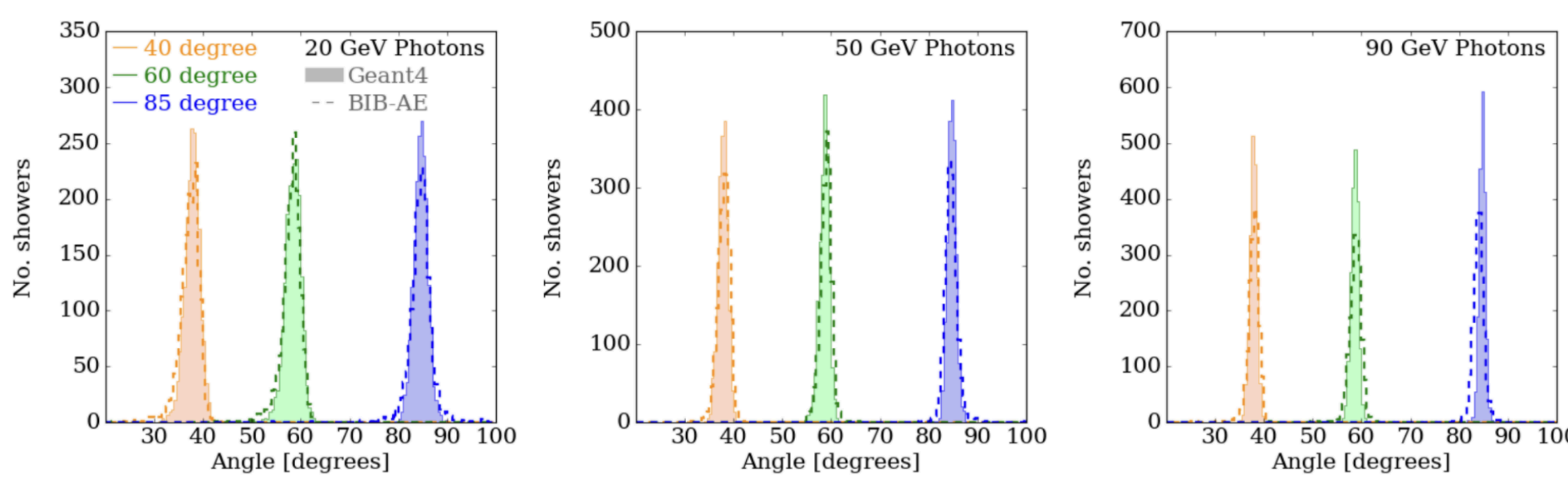
A novel architecture Bounded-Information-Bottleneck autoencoder with particle energy and angle conditioning achieves state of the art performance!





ILD ECAL / HCAL

- 30 / 48 layers
- 100,000,000 / 8,000,000 total channels
- 27,000 / 30,000 channel segment using in generative models

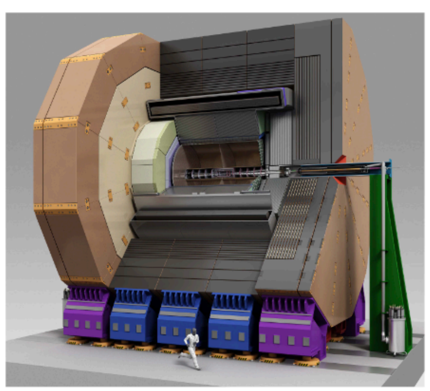


S. D. Diefenbacher plenary talk on Monday: "Generative Models for Fast (Calorimeter) Simulation"

Accelerating simulation

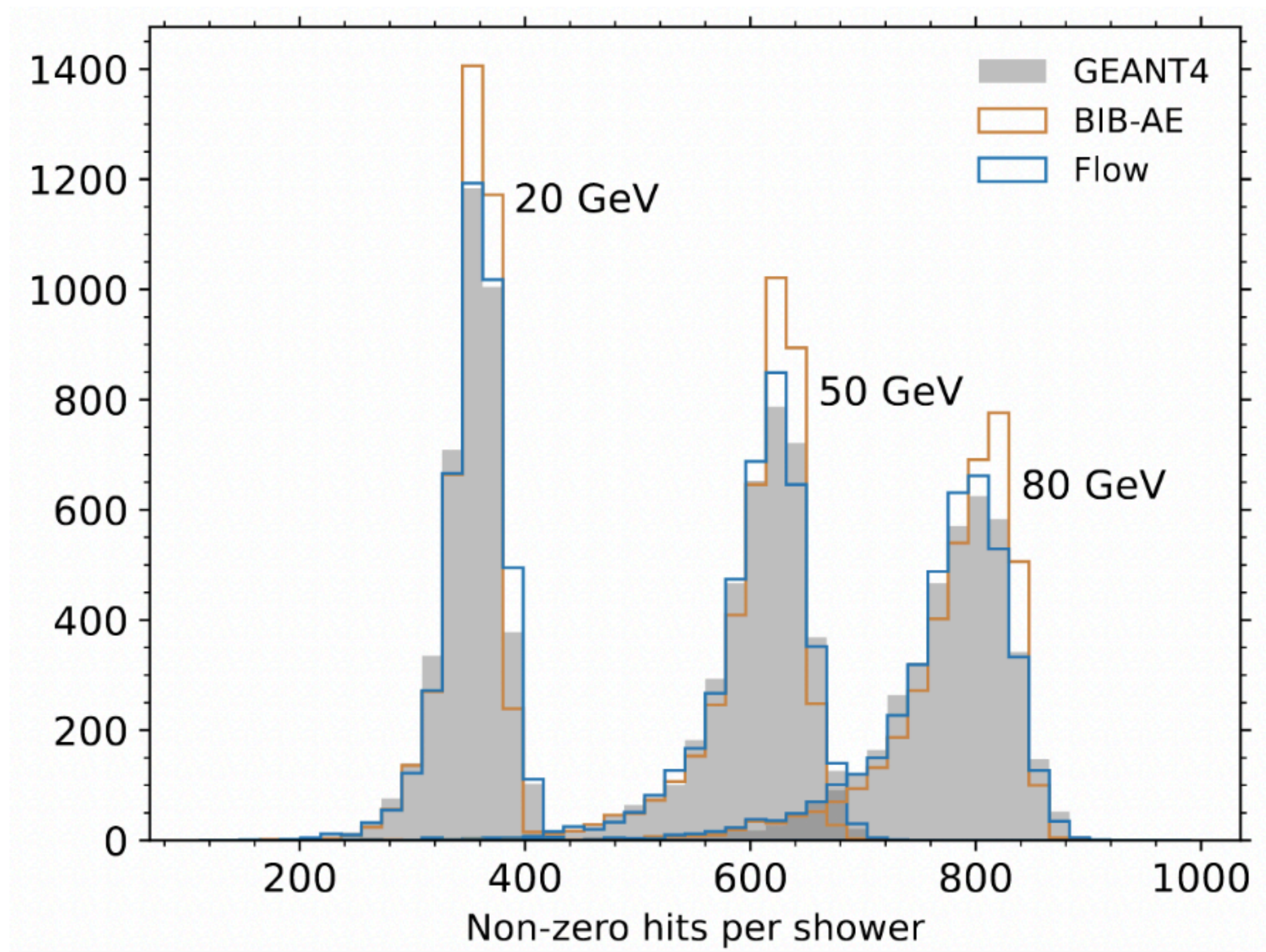
- Different **generative models** being explored:
  - Variational Autoencoders (VAE)
  - Generative Adversarial Network (GAN)
  - Normalizing Flows
- And different **data representations:**
  - images and point cloud

First promising results with flows outperforming BIB-AE! (work in progress)



ILD ECAL / HCAL

- 30 / 48 layers
- 100,000,000 / 8,000,000 total channels
- 27,000 / 30,000 channel segment using in generative models



S. D. Diefenbacher plenary talk on Monday: "Generative Models for Fast (Calorimeter) Simulation"

Accelerating simulation

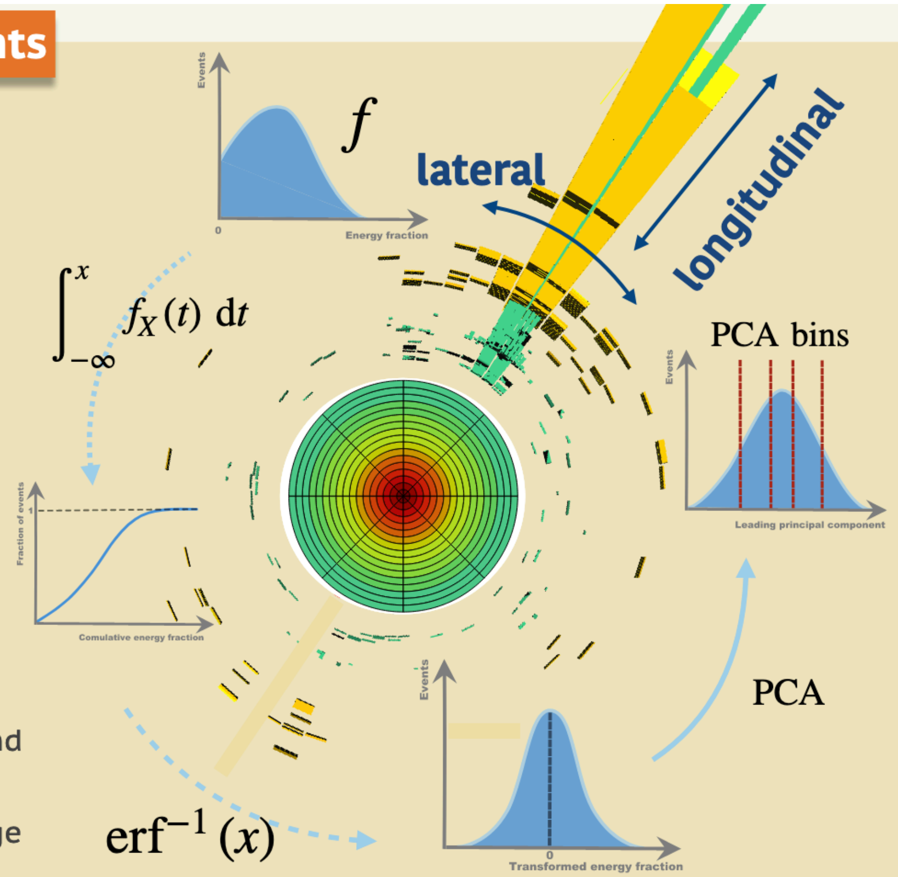
AtlFast3 Strategy: two components

FastCaloSimV2

- Parametrise Geant4 single particle shower
  - 17 energy bins × 100 |η| bins
  - Separate in longitudinal and lateral shape
- Deposits highly correlated between layers
  - Using Principal Component Analysis (PCA)
- Average lateral energy distribution parametrised as 2D probability functions

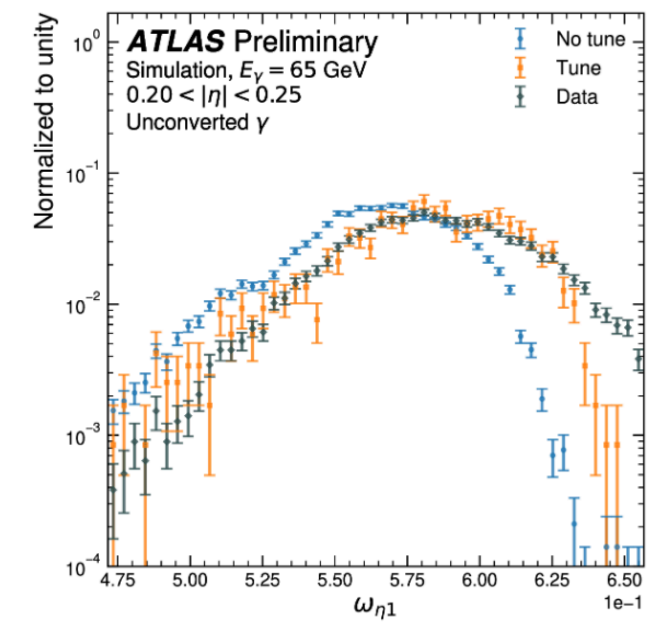
FastCaloGAN

- 500 Wasserstein Generative Adversarial NN in particle type & |η|, conditioned on true momenta
  - Reproduce voxels and energies in layers and total energy in a single step
- Used for hadrons in intermediate energy range

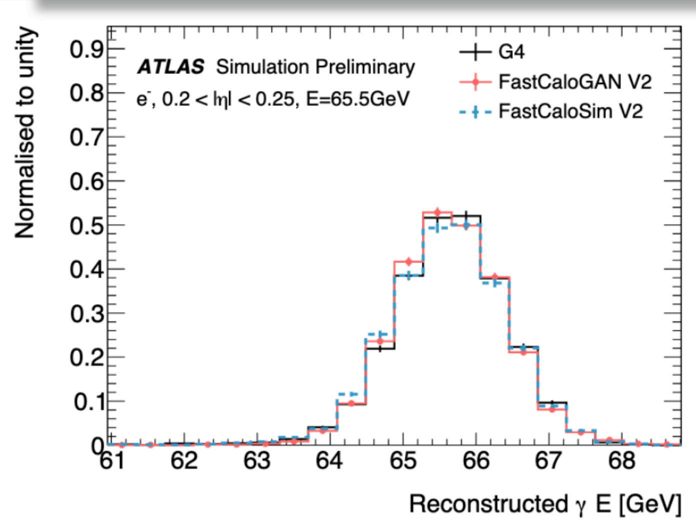


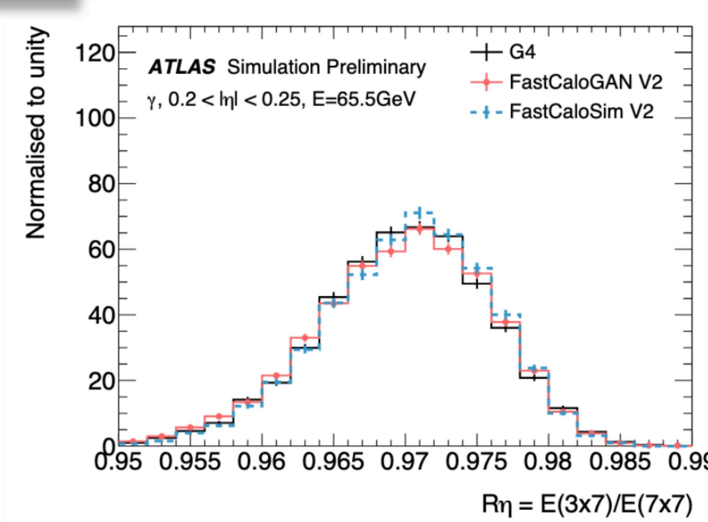
A step further: tuning to Data

- So far AltFast3 is trying to reproduce Geant4 simulation.
- Known differences between G4 and data
- **Tune AF3 to data!**
- Preliminary results look promising



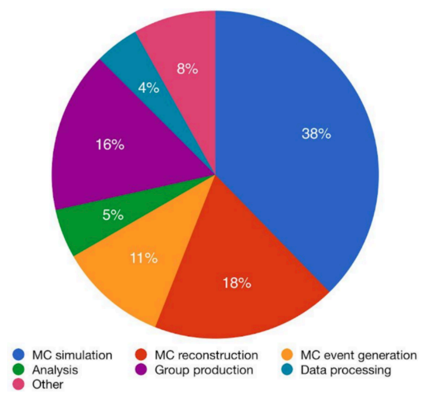
Latest performance





[R. Zhang poster: "AtlFast3: Fast Simulation in ATLAS for Run 3 and beyond"]

Accelerating simulation



- MC simulation large part of computing
- Speed up:
  - Train ML model on small dataset
  - Draw majority of samples form ML model
    - ➔Amplify original data set
    - ➔Significantly faster

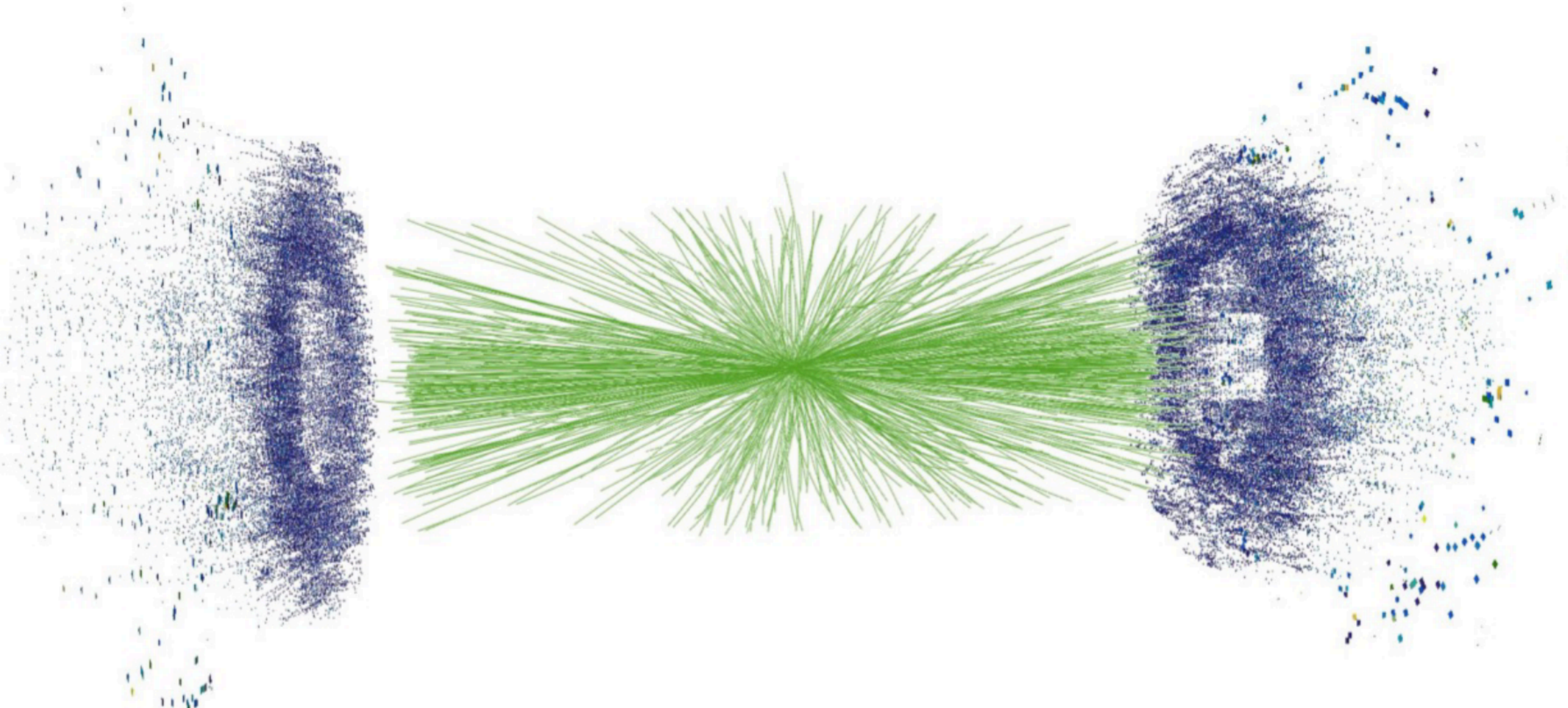
Catmore et. al. **ATLAS HL-LHC Computing Conceptual Design Report,** CERN-LHCC-2020-015 ; LHCC-G-178

Butter et al.: **Amplifying Statistics using Generative Models:** NeurIPS ML4PS 2020, 2008.06545

- Several other contributions on fast sim:
  - Talk: Optimization and deployment of ML fast simulation models
  - Talk: Accurate dE/dx simulation and prediction using ML method in the BESIII experiment
  - Poster: General shower simulation MetaHEP in key4hep framework
  - Poster: JETFLOW: Generating jets with Normalizing Flows using the jet mass as condition and constraint
  - Poster: Evaluating Generative Adversarial Networks for particle hit generation in a cylindrical drift chamber using Fréchet Inception Distance
  - Poster: CaloPointFlow - Generating Calorimeter Showers as Point Clouds
  - Poster: SCD: an open, realistic calorimeter for ML studies in HEP

More on this in Track 3 covering MC event generation!

Accelerating reconstruction



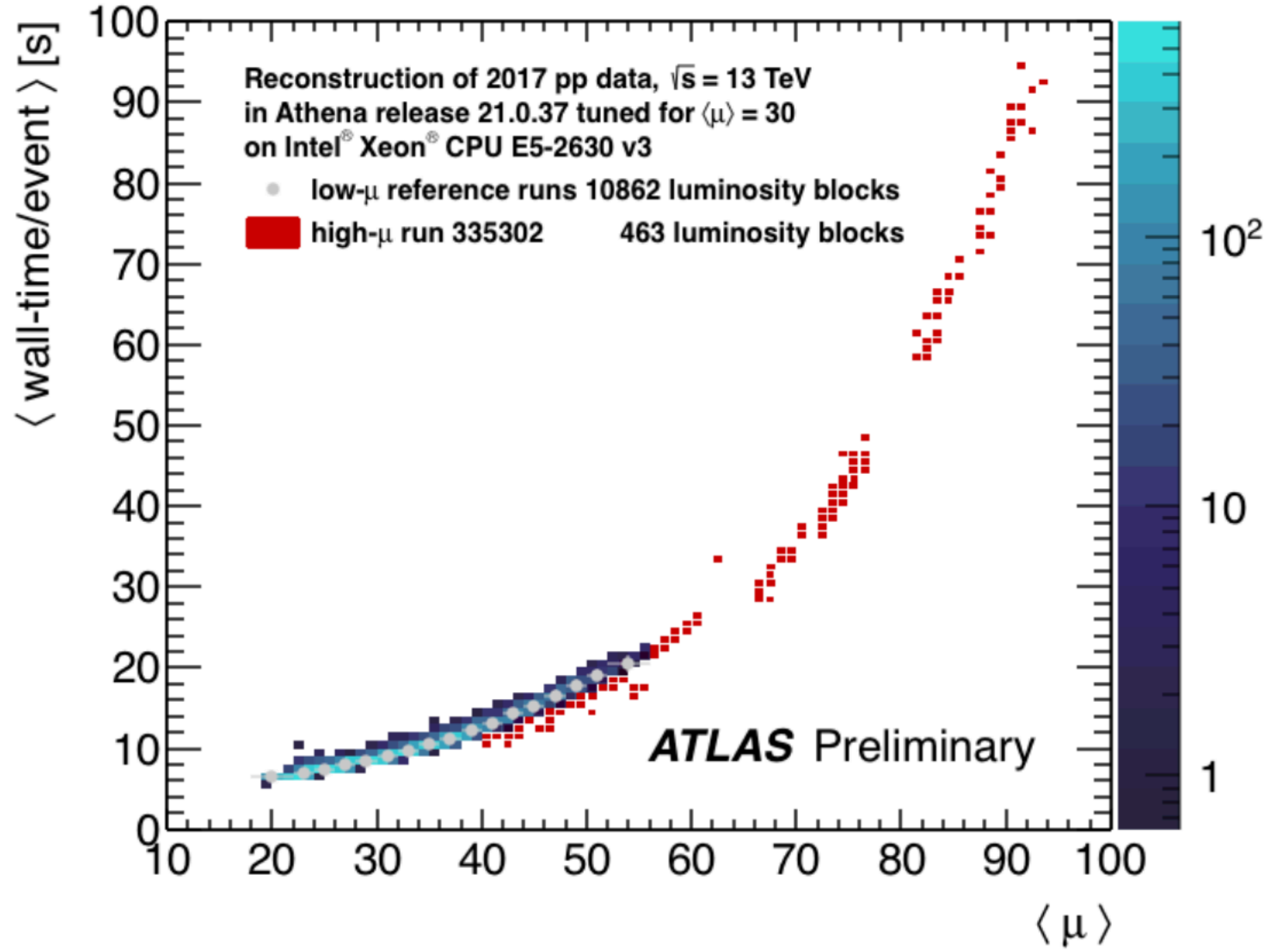
$t\bar{t}$ event with pileup 200

Image credits: T. Di Pilato's talk on Monday

Tracking

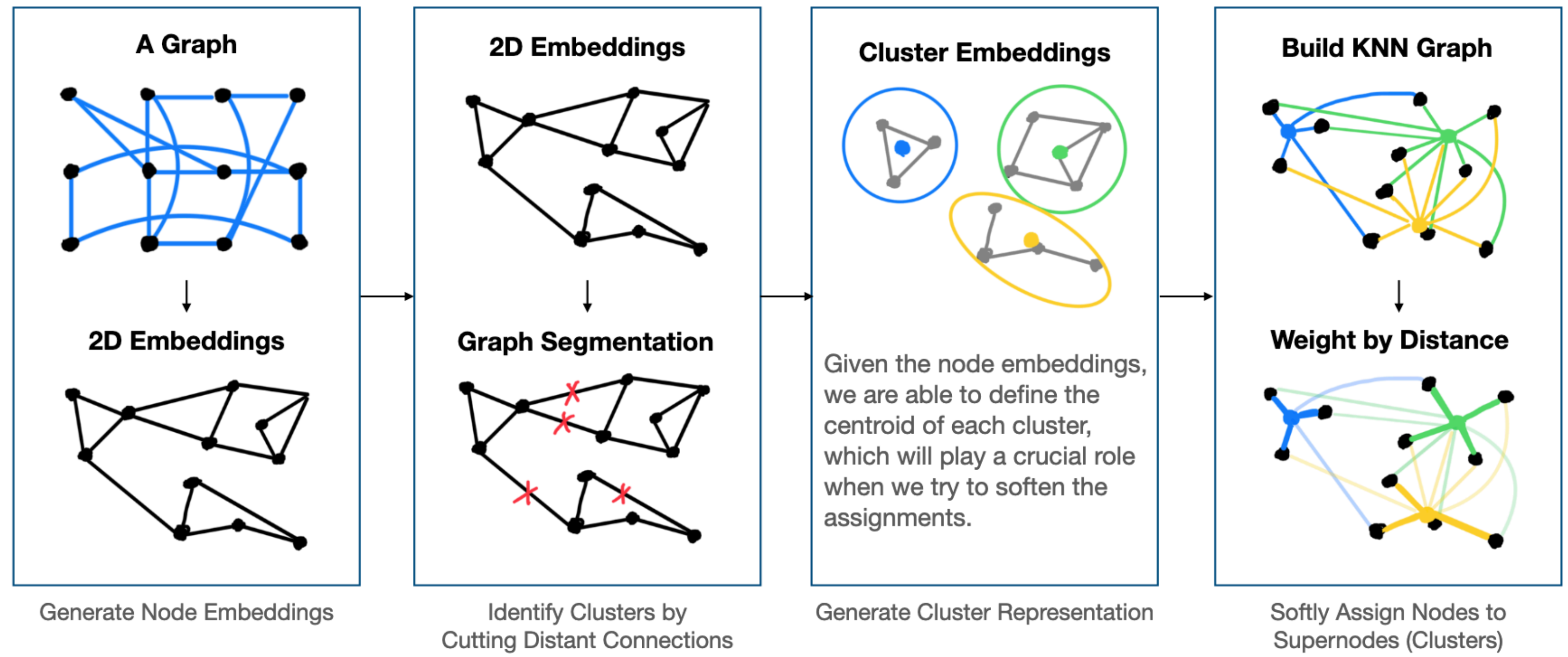
- Reconstructing tracks is one of the most computing intensive tasks!
  - can scale quadratically with number of particles in today's detectors

- Multiple contributions addressing the necessary speed up at high performance:
  - Talk: Simultaneous track finding and track fitting by the Deep Neural Network at BESIII
  - Talk: Hierarchical Graph Neural Networks for Particle Track Reconstruction
  - Talk: Standalone track reconstruction in LHCb's SciFi detector for the GPU-based High Level Trigger
  - Talk: Navigation, field integration and track parameter transport through detectors using GPUs and CPUs within the ACTS R&D project
  - Talk: Speeding up the CMS track reconstruction with a parallelized and vectorized Kalman-filter-based algorithm during the LHC Run 3
  - Poster: Fast track seed selection for track following in the Inner Detector Trigger track reconstruction
  - Poster: Faster simulated track reconstruction in the ATLAS Fast Chain
  - Poster: Auto-tuning capabilities of the ACTS track reconstruction suite
  - Poster: Equivariant Graph Neural Networks for Charged Particle Tracking
  - Poster: BESIII track reconstruction algorithm based on machine learning



Tracking with ML

Novel approach applied to TrackML challenge dataset deploying **Hierarchical GNN** to **mitigate inefficiencies due to missing nodes (hits) obstructing message passing**



	Embedding HGNN	Embedding IN	Bipartite HGNN	Edge Classifier IN	Truth CC
Tracking efficiency	97.32%	98.16%	98.86%	98.54%	99.91%
Tracking purity	95.78%	90.15%	98.76%	93.79%	95.28%
Time	**0.5280**	**0.3514**	**0.2625**	**0.2108**	**N/A**

Results removing randomly 20% of edges show good robustness

	Embedding HGNN	Embedding IN	Bipartite HGNN	Edge Classifier IN	Truth CC
Tracking efficiency	97.33%	92.78%	98.83%	91.93%	97.19%
Tracking purity	94.08%	92.19%	98.53%	74.52%	78.66%

R. Liu talk on Tuesday: "Hierarchical Graph Neural Networks for Particle Track Reconstruction"

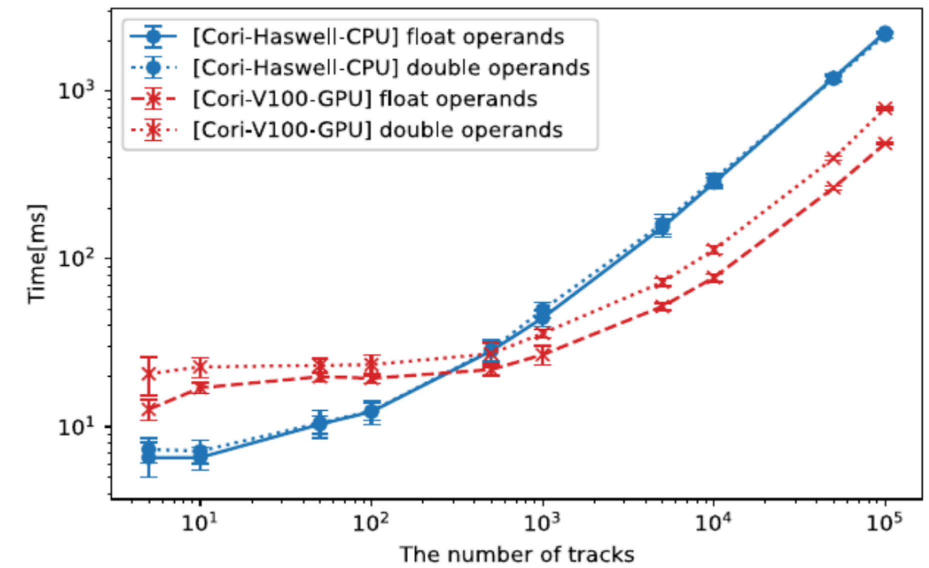
Tracking on GPU

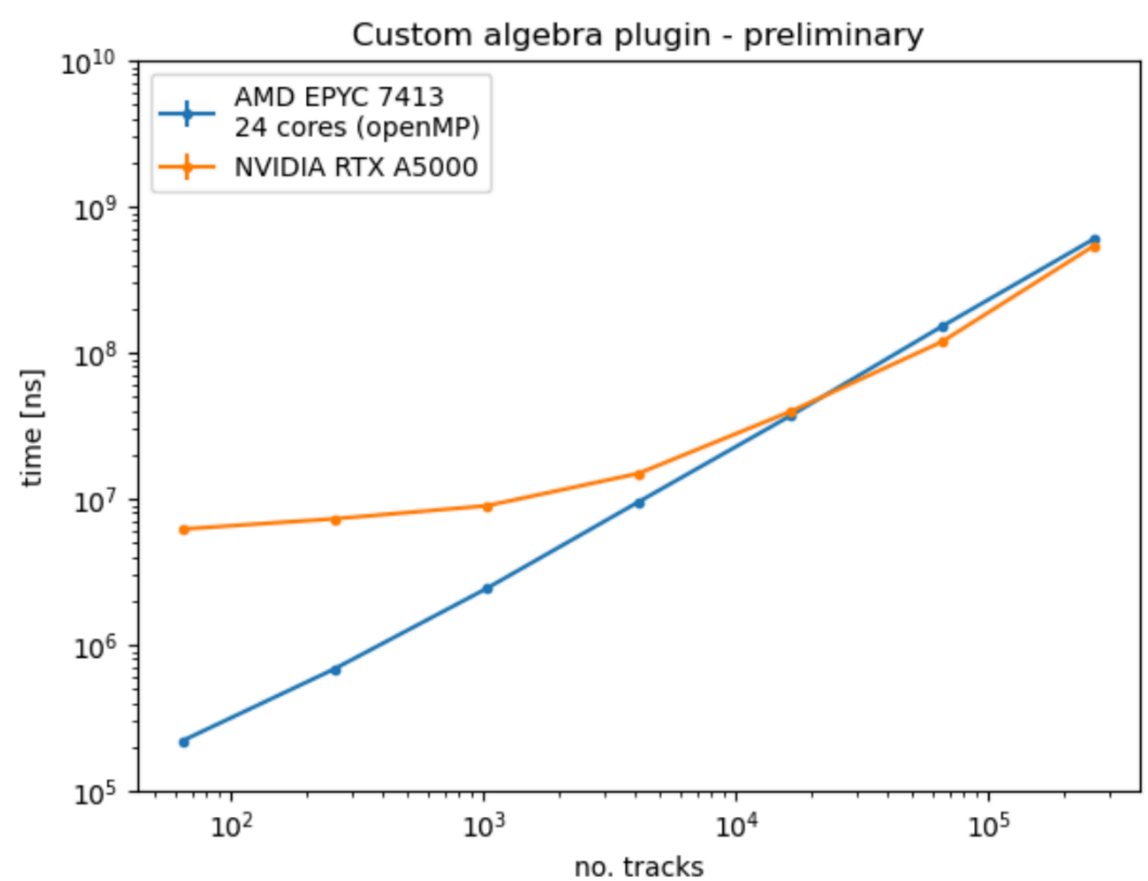
- The **ACTS R&D project** is a generic, framework- and experiment-independent software package for track reconstruction on modern computing architectures

- Several improvements wrt ACAT21
  - more realistic testbed detector including a simple material description
  - full magnetic field description
  - integration of equation of motion

- First demonstration for transport of full track parametrization and covariance through magnetic field and tracking geometry

Kalman Filter ported to GPU w/ x4.6 speedup wrt multithreaded CPU for events with ≳ 1000 tracks





[Talk on Tuesday: "Navigation, field integration and track parameter transport through detectors using GPUs and CPUs within the ACTS R&D project"]

[Poster: "Auto-tuning capabilities of the ACTS track reconstruction suite"]

Calorimeter reconstruction with ML

- Point cloud representation and graph NN (GravNet) architecture for end-to-end reconstruction in high-granular calorimeter (simplified geometry wrt CMS HGCal)
  - cluster hits belonging to the same particle
  - regress energy of clusters
  - jet reconstruction

Baseline

$$E_{\text{baseline}} = \sum_{h \in H_t} e_h$$

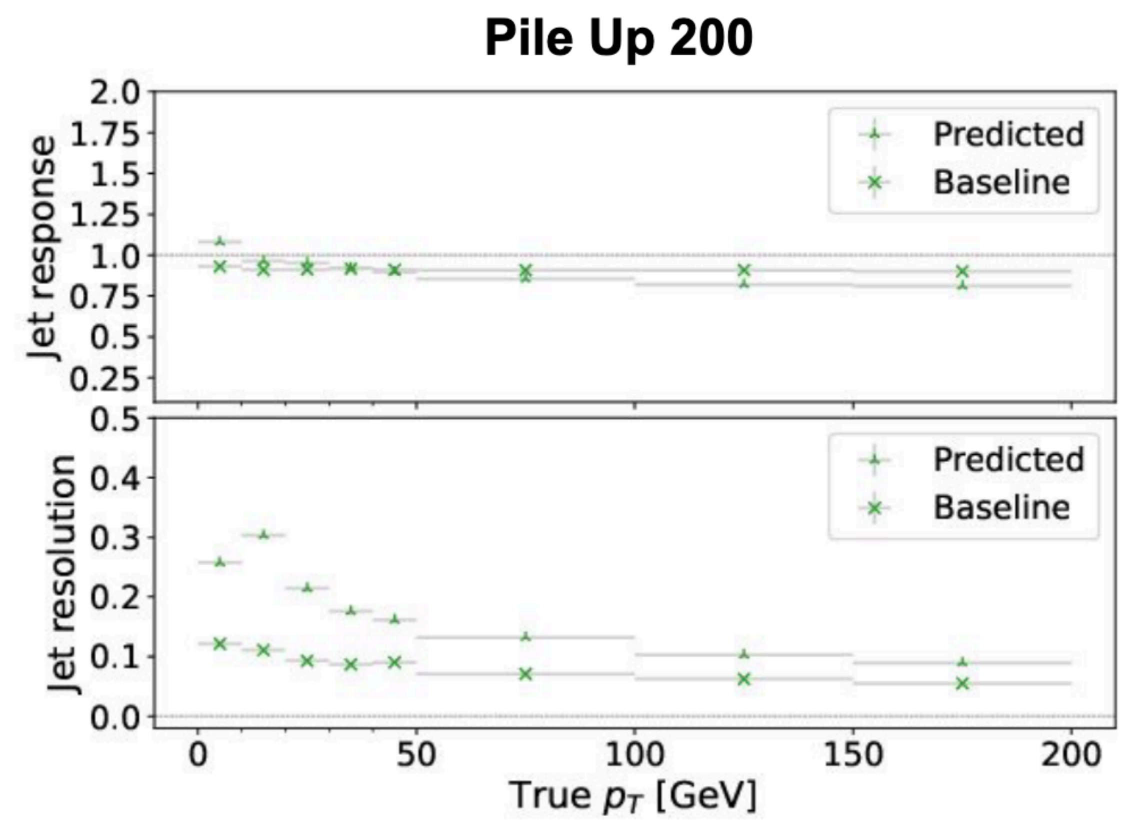
This baseline will be hard to match as it uses the truth information of the showers

Response

$$< p_{T_{pred}} / p_{T_{truth}} >$$

Mean-corrected resolution

$$\sigma \left(p_{T_{pred}} / p_{T_{truth}} \right) / < p_{T_{pred}} / p_{T_{truth}} >$$



- ❖ Response < 1 due to large hadronic contributions
- ❖ Comparable response to baseline
- ❖ Resolution approaching 10% in both PU scenarios

P. Zehetner talk on Monday: "End-to-end multi-particle reconstruction in high occupancy imaging calorimeters with graph neural networks"

Calorimeter reconstruction with ML

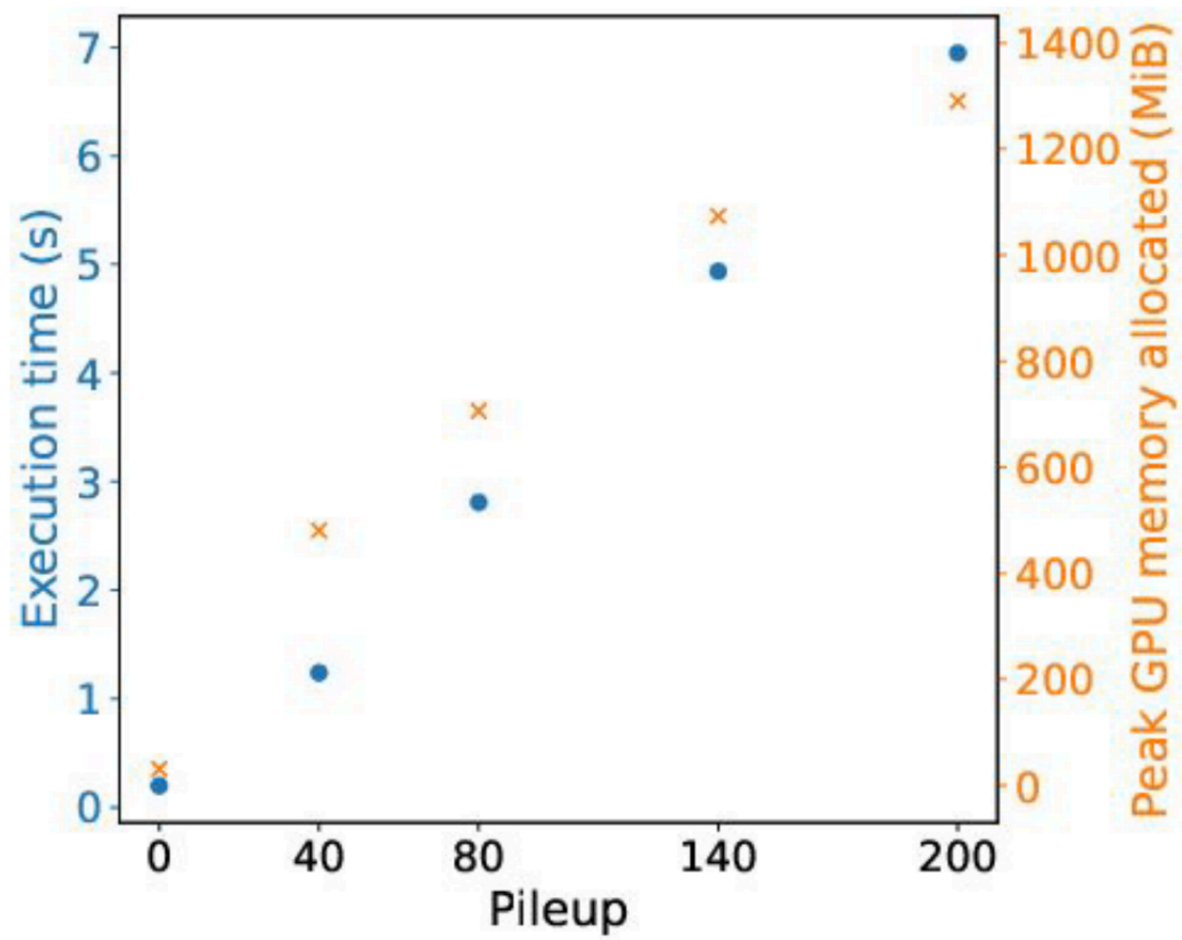
- Point cloud representation and graph NN (GravNet) architecture for end-to-end reconstruction in high-granular calorimeter (simplified geometry wrt CMS HGCal)
  - cluster hits belonging to the same particle
  - regress energy of clusters
  - jet reconstruction

❖ Inference time and memory both scale linear with number of hits in detector

❖ Less than 10 seconds inference time for 200 PU (NVIDIA V100 GPU)

❖ Less than 1.5 GB peak memory usage for 200 PU →Can be deployed on low-end GPUs

❖ Ongoing work on inclusion of small clustering models to compress input indicate potential for significant speed ups



[P. Zehetner talk on Monday: "End-to-end multi-particle reconstruction in high occupancy imaging calorimeters with graph neural networks"]

Calorimeter reconstruction on GPU

[T. Di Pilato talk on Monday: "Performance study of the CLUE algorithm with the alpaka library"]

- ***HEP approach: offloading part of the reconstruction to GPUs for parallel execution***

❖ **Many vendors → many programming languages → many versions of the same code!!!**

NVIDIA CUDA, AMD ROCm, OpenCL

❖ Performance portability libraries have become an interesting solution
- Write code once
- Compile for different backends
- Execute on target platform

➤ Not all the technologies provide close-to-native backend performance

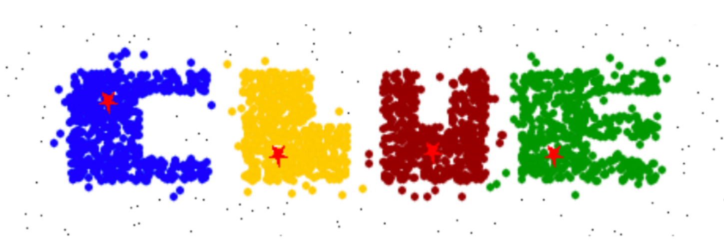
❖ Portable code can be ***easily maintained*** and support new accelerators

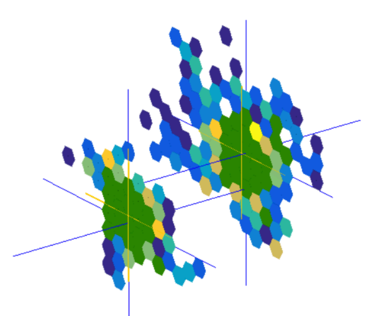
❖ ***CMS choice for Run 3:*** alpaka

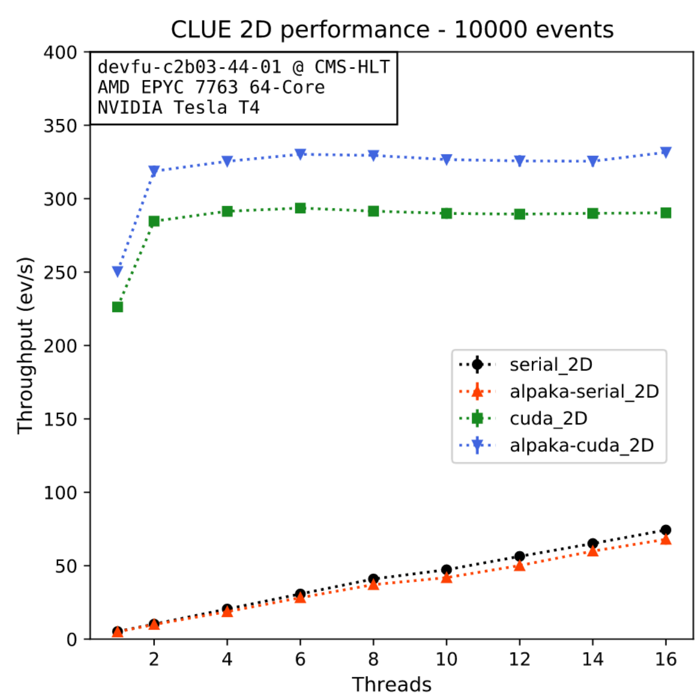
Additional coverage in Track 1!

A realistic application

CLUstering of Energy (CLUE): fast 2D clustering algorithm developed for the future CMS-HGCAL detector
- Based on energy density
- Builds small clusters (~10 RecHits)
- **Fully ported to GPU (CUDA)**
- Uses a **tiled** data structure that fully exploits the detector granularity and allows fast querying of neighbor cells







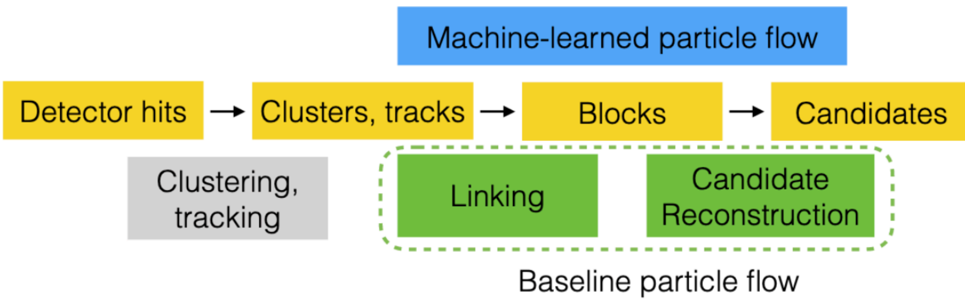
- **Alpaka** with the **serial** backend ***scales linearly*** with the number of threads (concurrent events), ***the same way*** as the **native serial** implementation
- **Alpaka** with the **cuda** backend has the ***same scaling*** of the **native cuda** implementation. Two points are under investigation:
  - Other applications **do not show** that alpaka is faster than cuda
  - It seems that I/O operations and the computing capability of the GPU are limiting the scaling for threads > 4

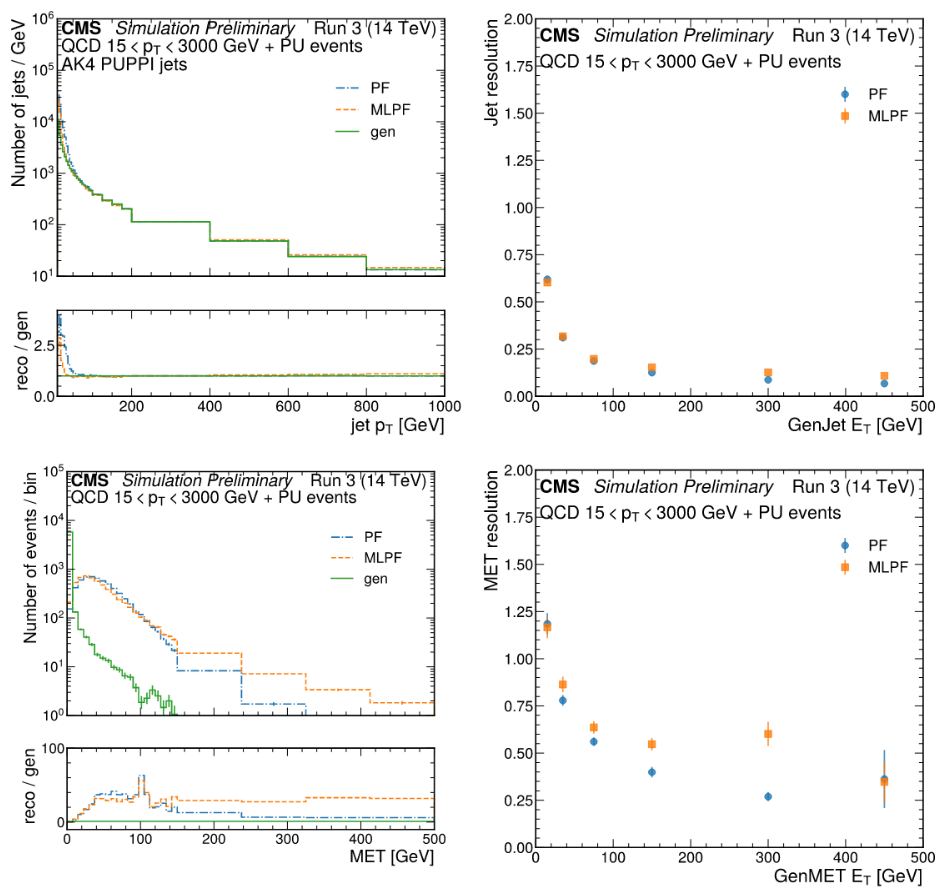
[E. Brondolin poster: "k4Clue: Having CLUE at future colliders experiments"]

[F. Pantaleo poster: "The TICL reconstruction at the CMS Phase-2 High Granularity Calorimeter Endcap"]

Particle Flow

Machine Learning Particle Flow





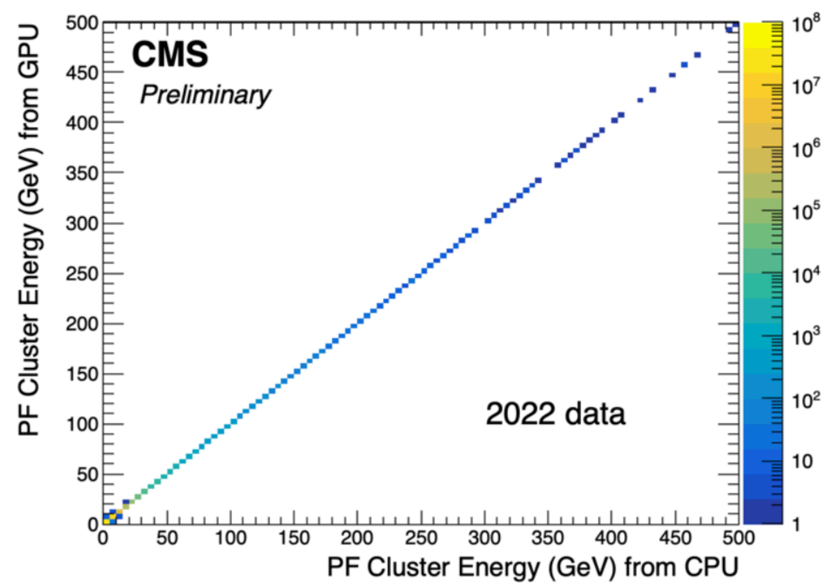
[F. Mokhtar poster: "Progress towards an improved particle flow algorithm at CMS with machine learning"]

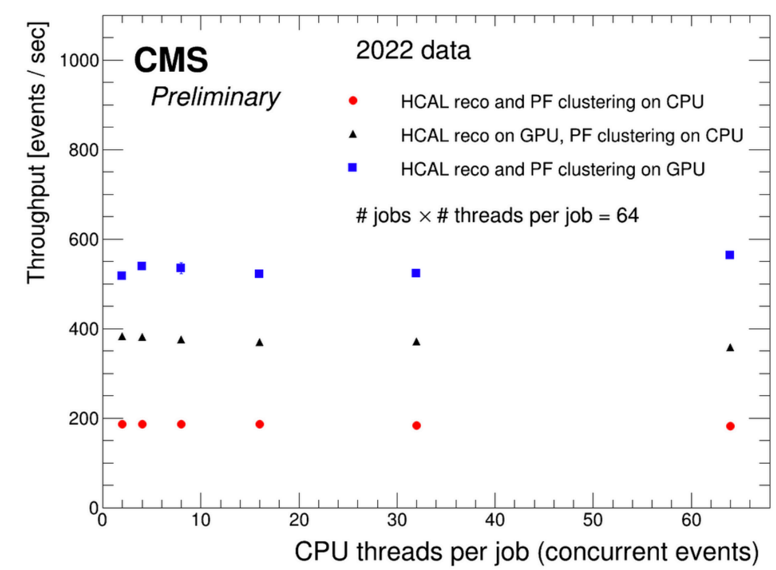
[E. Wulff poster: "Hyperparameter optimization, multi-node distributed training and benchmarking of AI-based HEP workloads using HPC"]

Particle Flow on GPU

Run 3 CMS HLT farm 200 nodes each w/ two AMD Milan 64-core CPU and two NVIDIA Tesla T4 GPUs

The GPU-accelerated Particle Flow Clustering will enter in production during the upcoming CMS 2023 data taking.





[F. Pantaleo poster: "Particle Flow Reconstruction on Heterogeneous Architecture for CMS"]

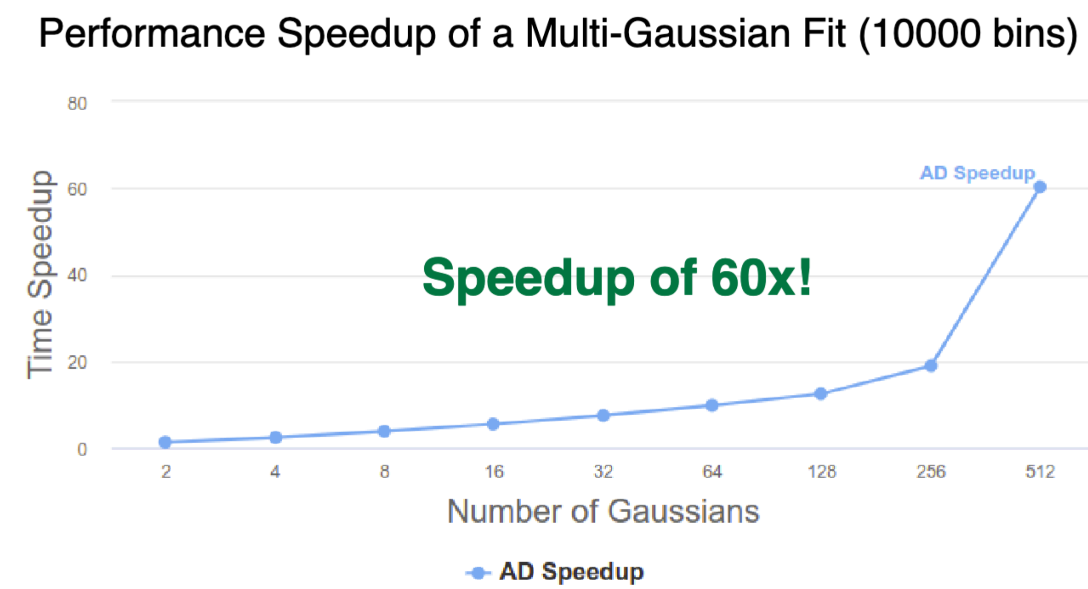
Accelerating data analysis

Scalable analysis tools

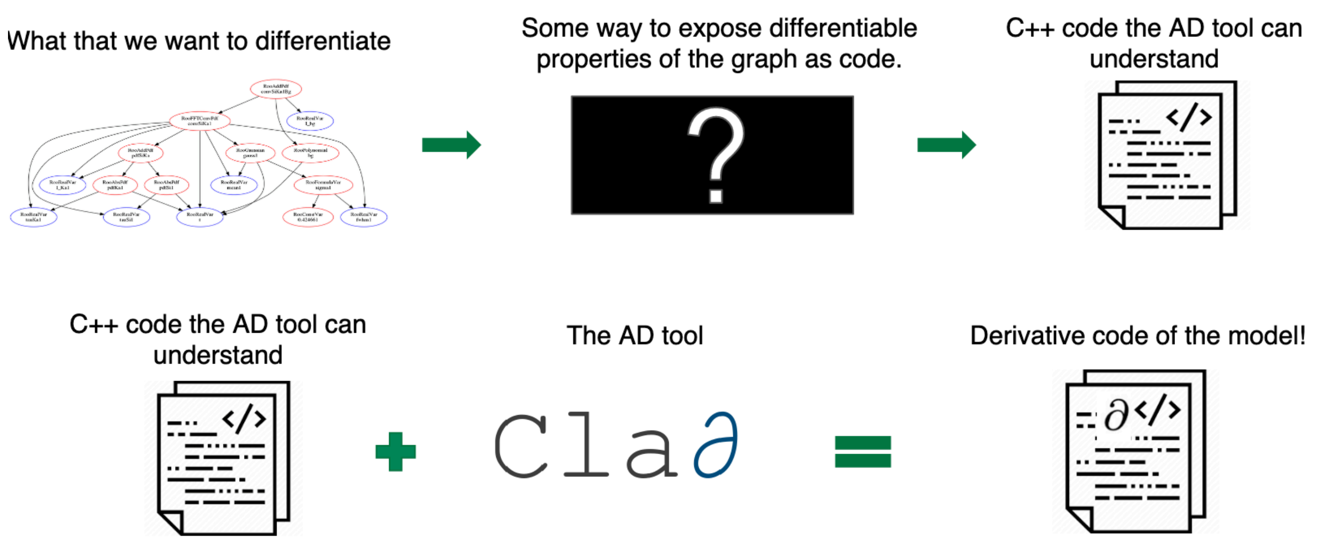
See Track 1 for larger coverage of this topic!

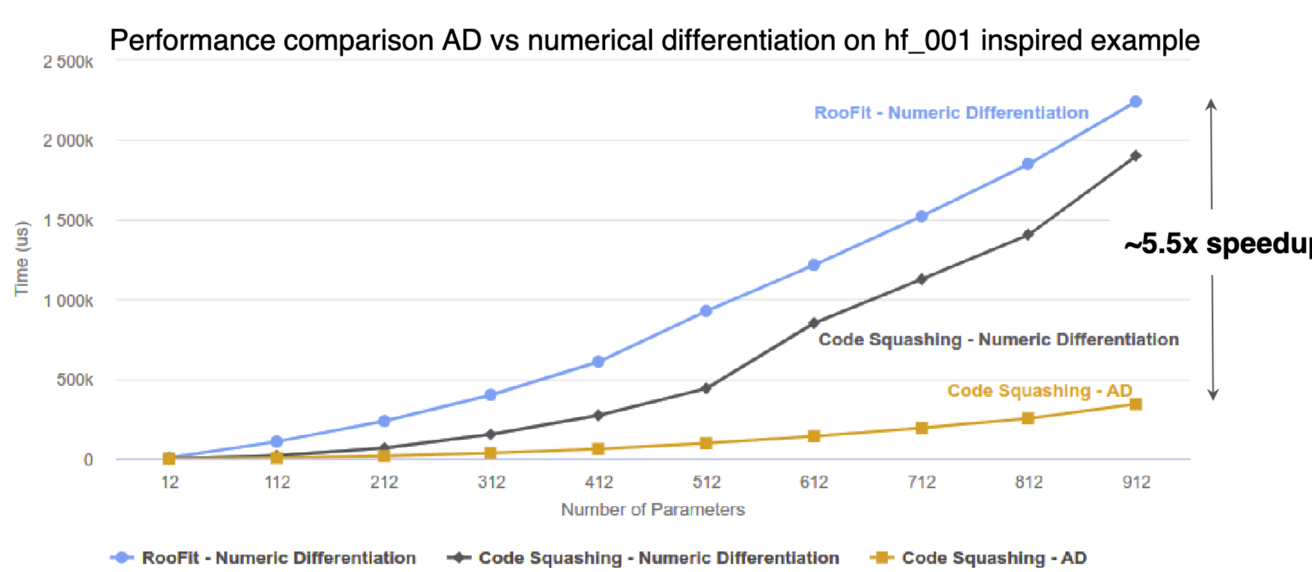
- **Clad enables automatic differentiation for C++** as a plugin for Clang compiler
  - based on source code transformation: given C++ source code of a mathematical function it can automatically generate C++ code for computing derivatives of the function

Promising results when integrated in ROOT!



Different approach for RooFit: translating models to code



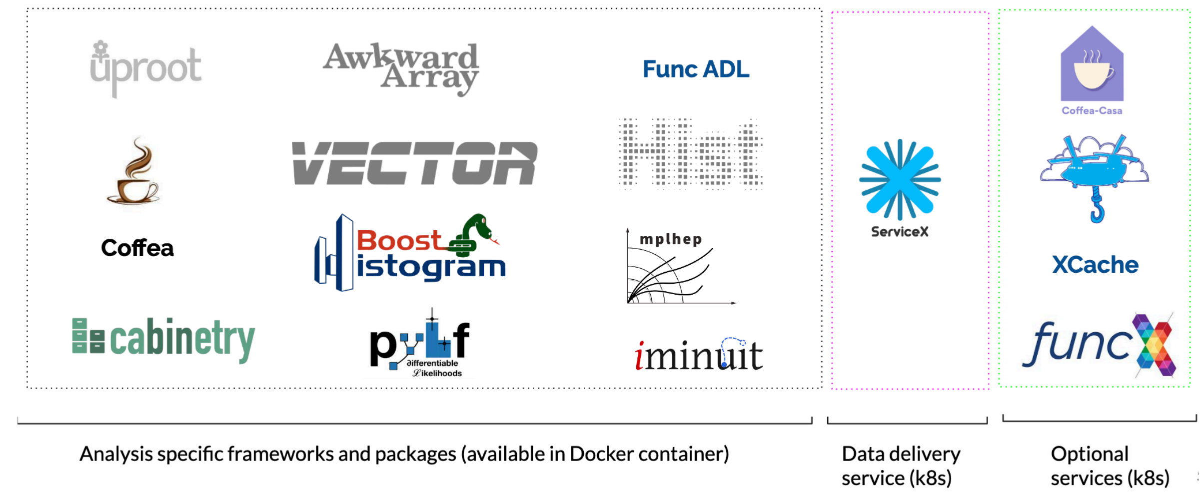


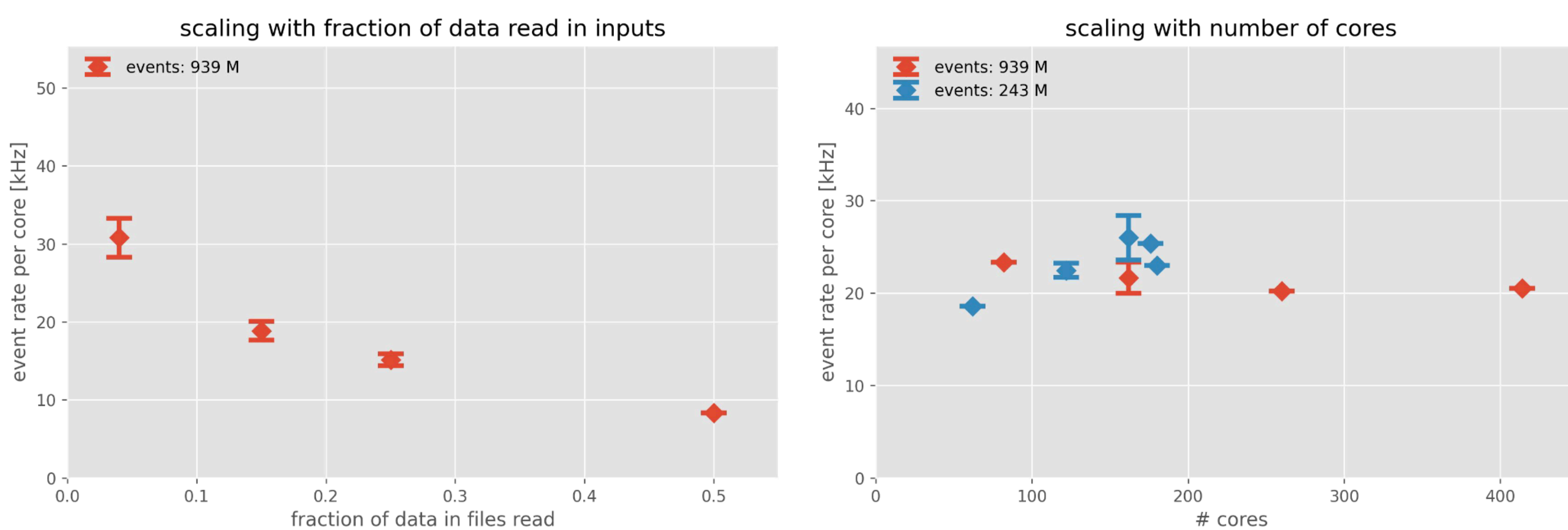
[G. Singh talk on Wednesday: "Automatic differentiation of binned likelihoods with RooFit and Clad"]

See also contributions on [RDataFrame] and [ROOT] capabilities!

Scalable analysis workflows

- The IRIS-HEP **Analysis Grand Challenge** is a realistic environment for investigating how analysis methods scale to the demands of the HL-LHC
  - includes all relevant workflow aspects [from data delivery to statistical inference]()
  - analysis tasks heavily based on tools from the [HEP Python ecosystem]()
  - makes use of [modern analysis facilities]()





UNL Coffea-casa AF @ UNL CMS Tier-2 (Coffea with DaskExecutor): stable scaling to 400 cores events with increasing number of branches (bigger fraction of data to read)

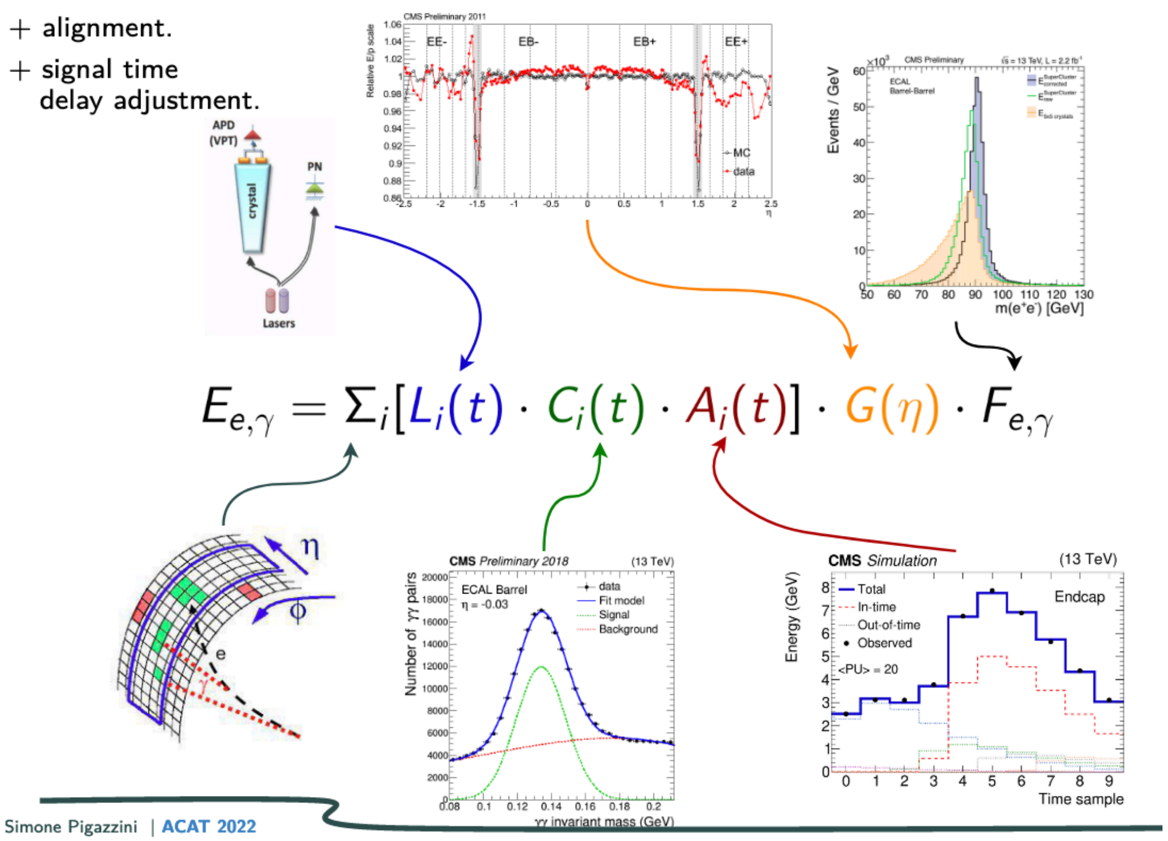
[O. Shadura talk on Tuesday: "First performance measurements with the Analysis Grand Challenge"]()

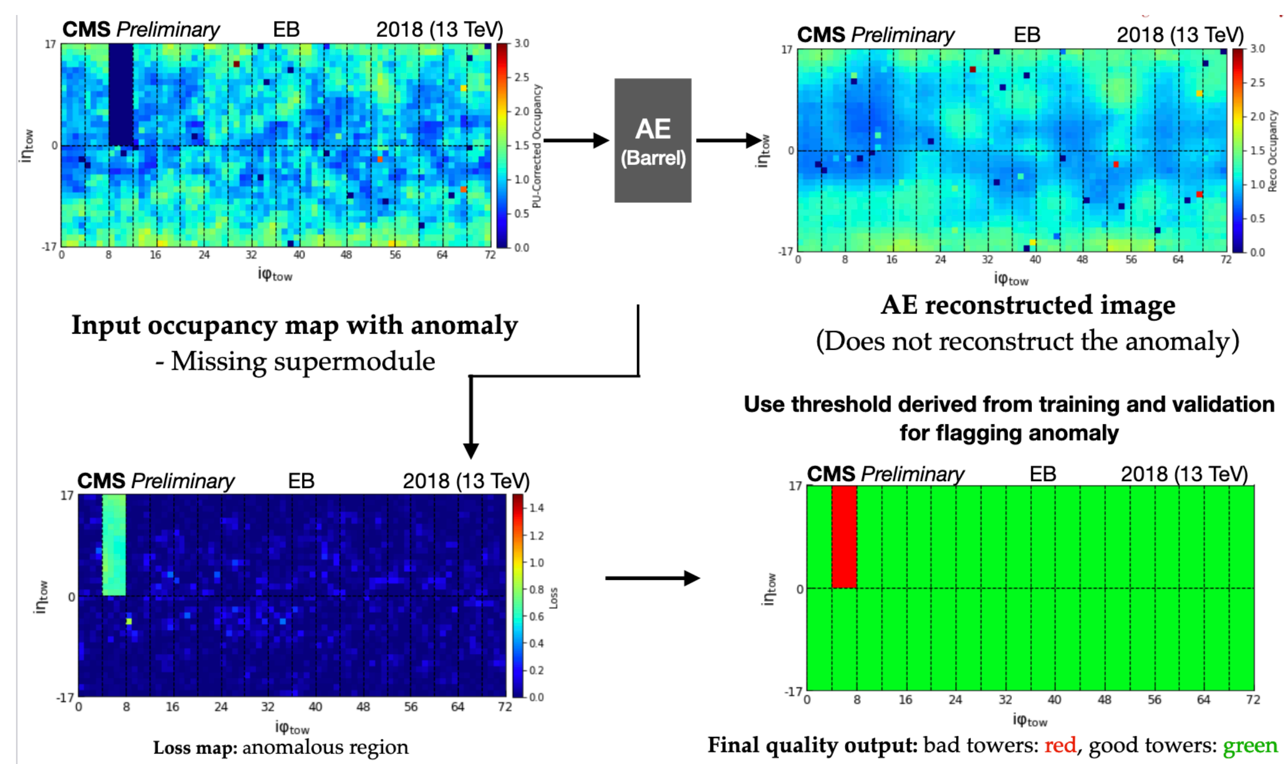
Automatization of detector controls

Automatization of detector controls

[A. Harilal talk on Thursday: "An Autoencoder-based Online Data Quality Monitoring for CMS ECAL"]

Automatic online data quality monitor for CMS ECAL with anomaly detection through Autoencoders being commissioned now!



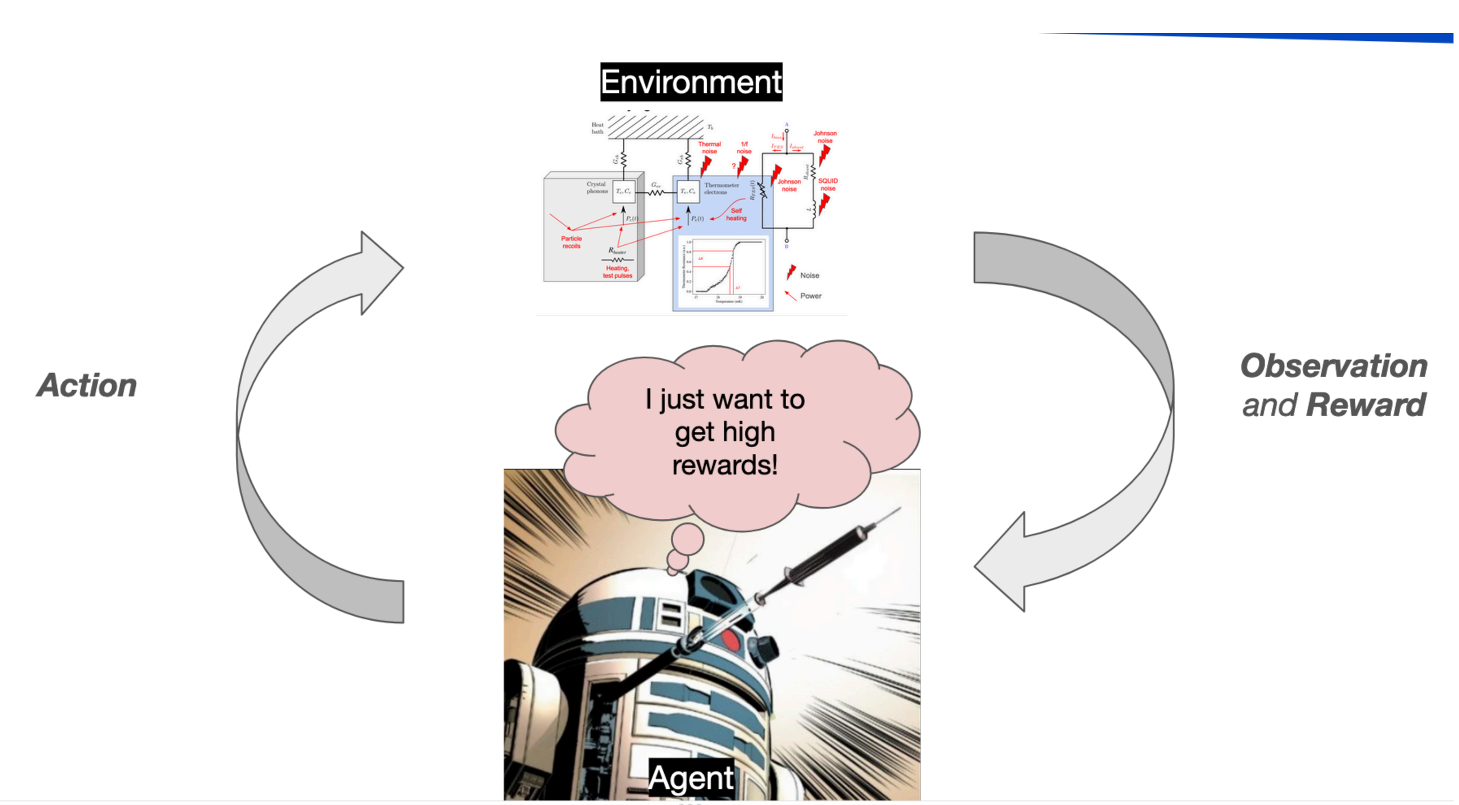


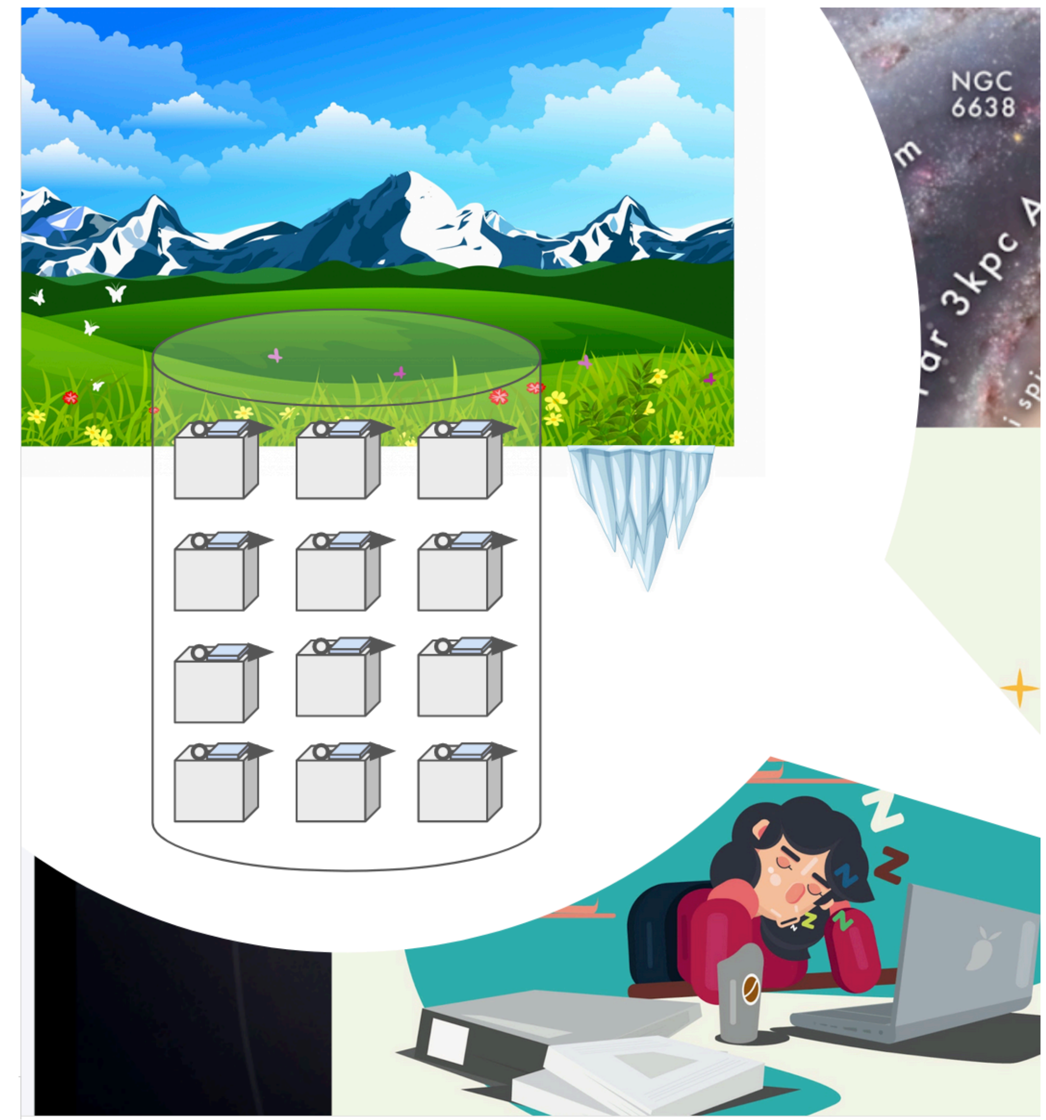
Automatic (frequent) calibrations of CMS ECAL in Run 3:

- framework of finite state machine implemented through Jenkins, Influxdb and Grafana for monitoring
- deployed with the OpenShift instance provided by CERN-IT
- a small python package to provide the interface between the CMS ecosystem, the user jobs and the framework.

[S. Pigazzini talk Thursday: "Automatic data processing for prompt calibration of the CMS ECAL"]

Automatization of detector controls





Reinforcement Learning for quick adjustment of optimal operating current of thermometers in cryogenic dark matter detectors drastically reducing manual intervention!
(so far being studied on simulation)

[F. Wagner talk on Thursday: "Control of cryogenic dark matter detectors through deep reinforcement learning"]

The common goal

And we are here discussing **all possible angles** to achieve this goal

Accelerate scientific discoveries by boosting computing efficiency

Detector monitoring

Trigger

Calibrations

Simulation

Reconstruction

Analysis

Quantum Computing

Quantum computing applications

Particle tracking with NISQ computers

Tim Schwägerl | Bari, 25.10.2022 | with Karl Jansen, Cigdem Issever and others

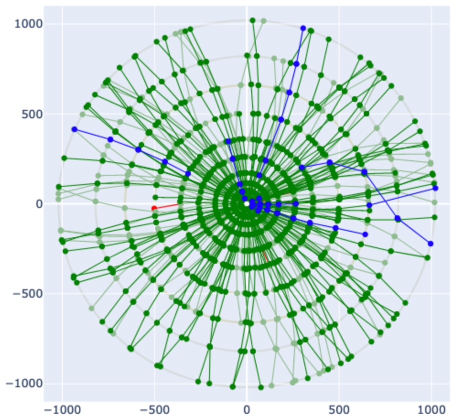


Figure: Hits and tracks in an ATLAS-like detector.

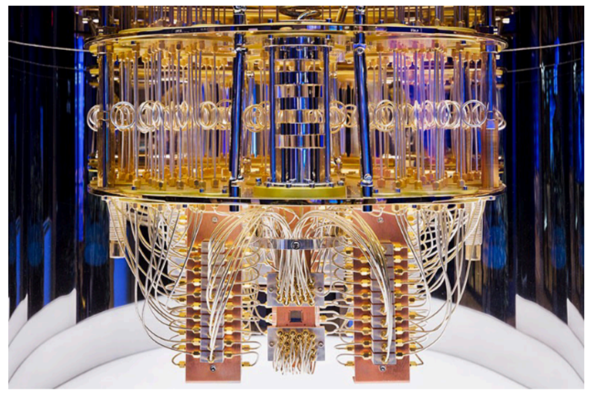


Figure: IBM quantum computer.

HELMHOLTZ

Berlin University Alliance

EINSTEIN Foundation.de

DESY.

Full Quantum GAN Model for HEP Detector Simulations

ACAT 2022 Conference

Florian Rehm [CERN, RWTH Aachen]

Sofia Vallecorsa [CERN], Michele Grossi [CERN], Kerstin Borras [DESY, RWTH Aachen], Dirk Krücker [DESY], Simon Schnake [DESY, RWTH Aachen], Alexis-Harilaos Verney-Provatas [DESY, RWTH Aachen]

Quantum anomaly detection in the latent spaces of high energy physics events

V. Belis[1,2], K. Wozniak[2,3], E. Puljak[2,4], M. Grossi[2], S. Vallecorsa[2], M. Pierini[2], F. Reiter[1], G. Dissertori[1]

ACAT 2022

Reconstructing Particle Decay Trees with Quantum Graph Neural Networks in High Energy Physics

Melvin Strobl, Eileen Kühn, Max Fischer, Achim Streit

Hybrid Quantum-Classical Networks for Reconstruction and Classification of EO Images

Su Yeon Chang [CERN, EPFL]

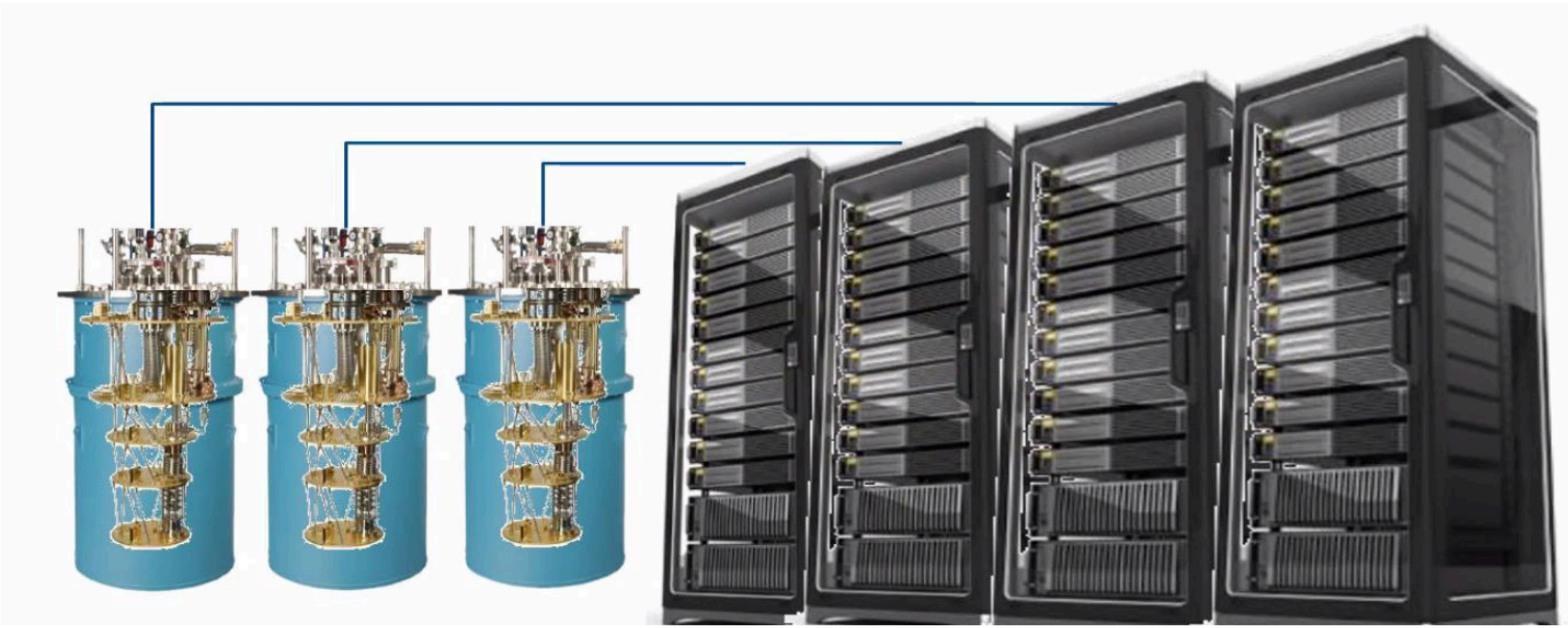
Bertrand Le Saux [ESA], Sofia Vallecorsa [CERN], Michelle Grossi [CERN]

+ several plenaries!

When?

Quantum Computing in Data Center

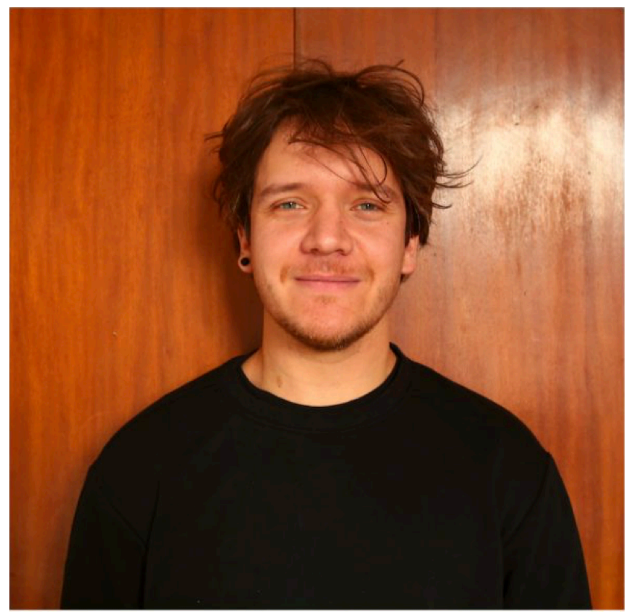
Future: Combination of High-Performance Computing and Quantum Computing



Deggendorf Institute of Technology: a teaching data center of such concept is being developed with involvement of students.

A big thank to the T2 co-conveners for the hard work in organizing this track program,
to the colleagues that helped chairing the sessions,
and to the all ACAT organization for this great workshop!

Sophie



Adriano

Dalila

Me

Thank you!

And see you next year...

ACAT 2022

BARI