Overview of AI Activities at CERN

Lorenzo Moneta
(CERN EP/SFT)

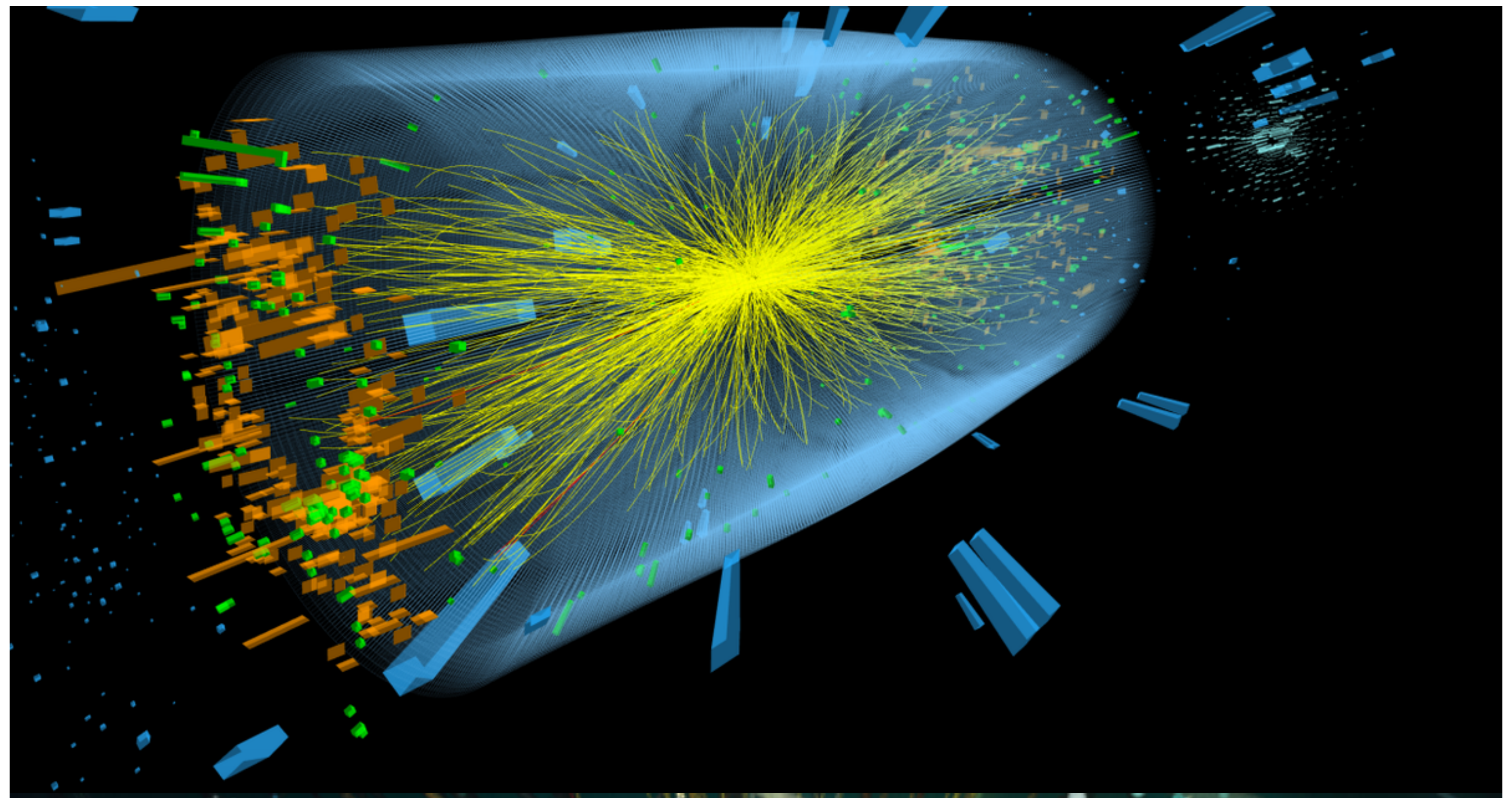
CERN openlab

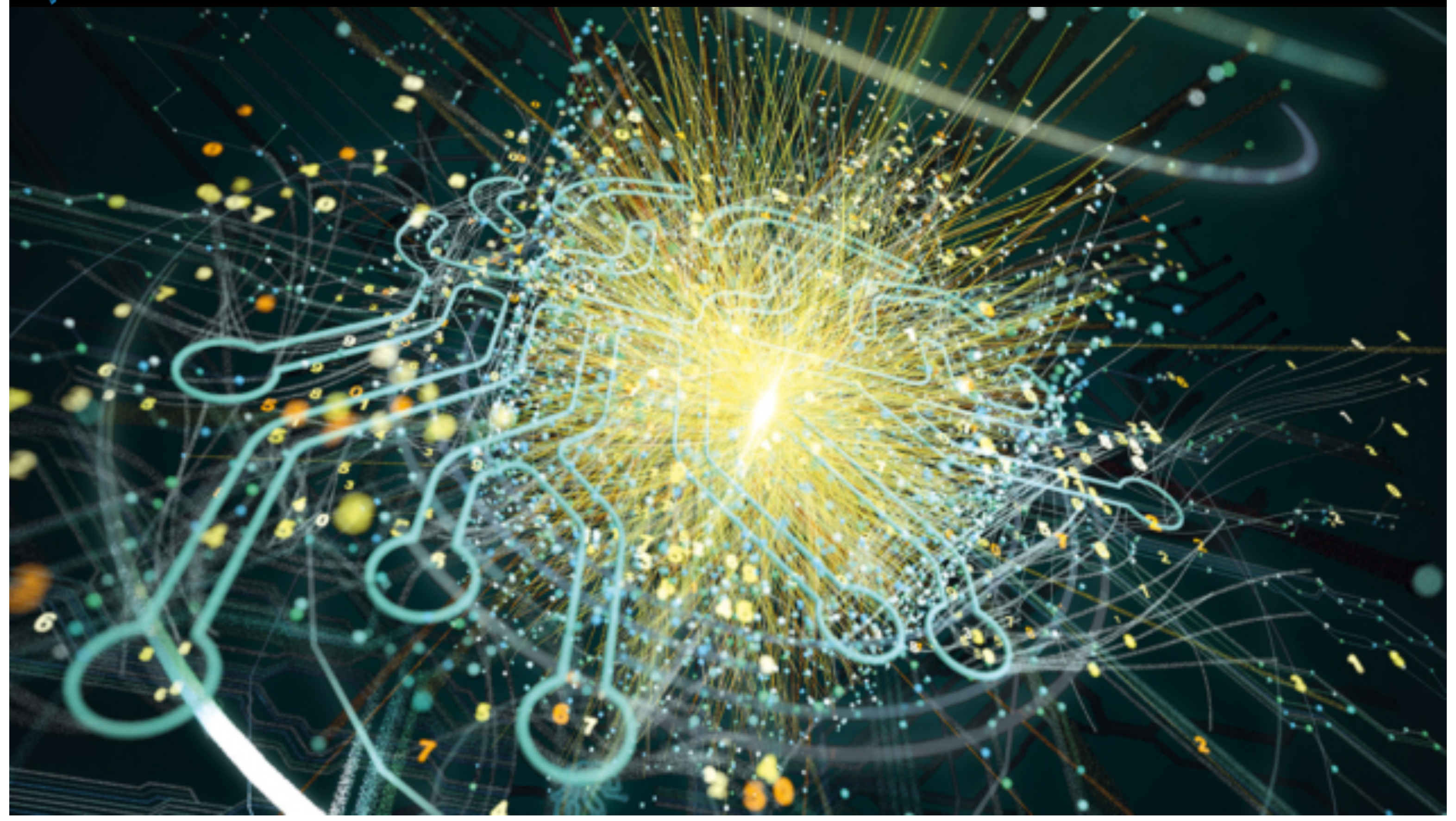
Openlab Technical Workshop, March 26, 2024

Outline

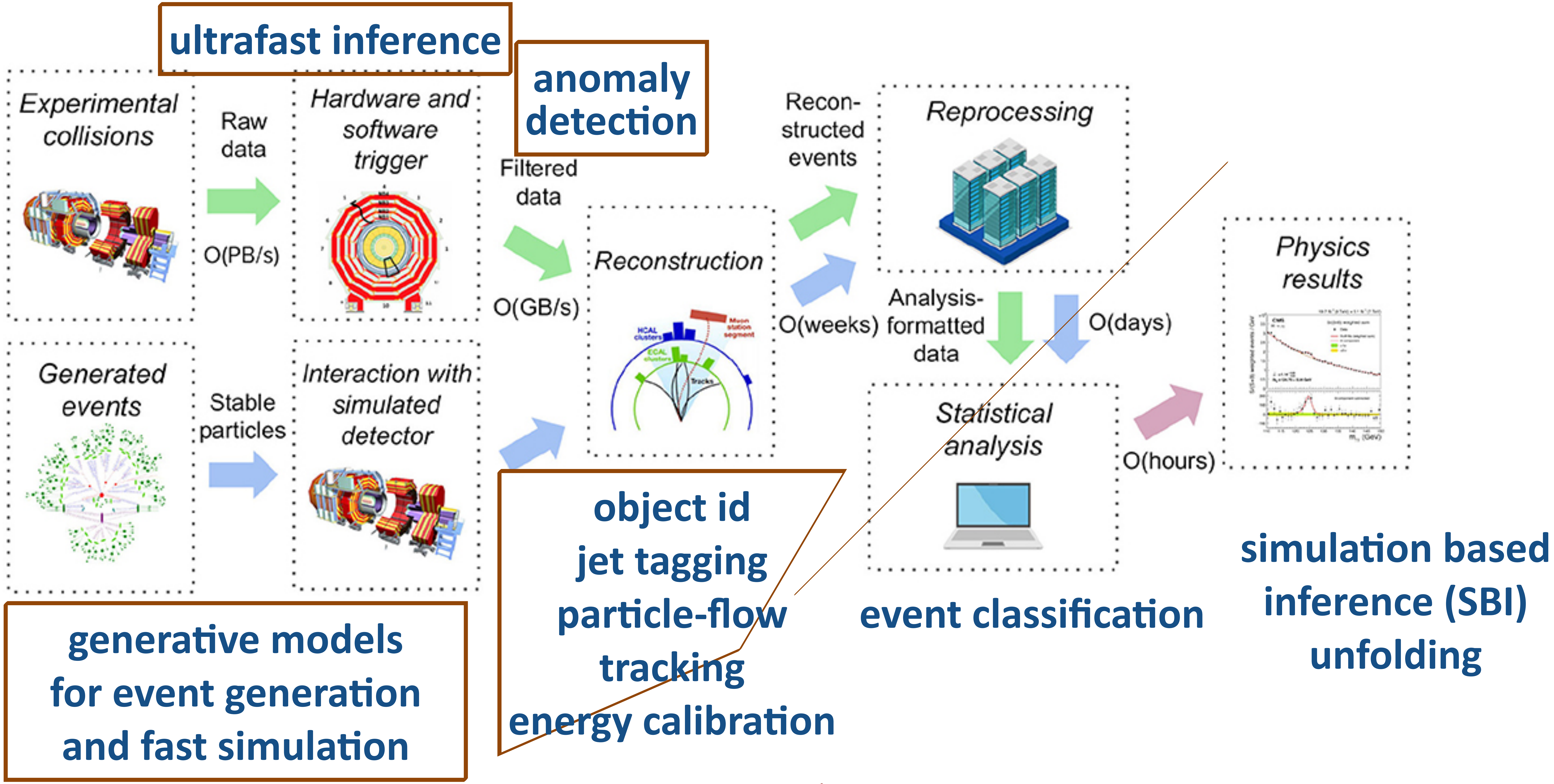
Personal view of some of the AI activities at CERN
(*as being part of Inter-experimental Machine Learning group*)

- Applications of AI/ML in the LHC experiments with some examples
  - AI in reconstruction (e.g. jet tagging), anomaly detection and fast simulation
  - not covering trigger applications (fast ML), will be shown in the next presentation
- Some examples of AI usage for the CERN accelerator complex
  - to have more efficient particle accelerators
- Many other AI activities existing at CERN:
  - openlab AI activities (presentations at this workshop)
  - quantum ML
  - robotics, digital twins, etc..
  - projects in collaborations with industry (KT)
  - see the last workshop on Applied AI organised by the Knowledge Transfer group (KT)





AI in Experiment Data Analysis



AI is everywhere !

pic from fdata.2021.661501

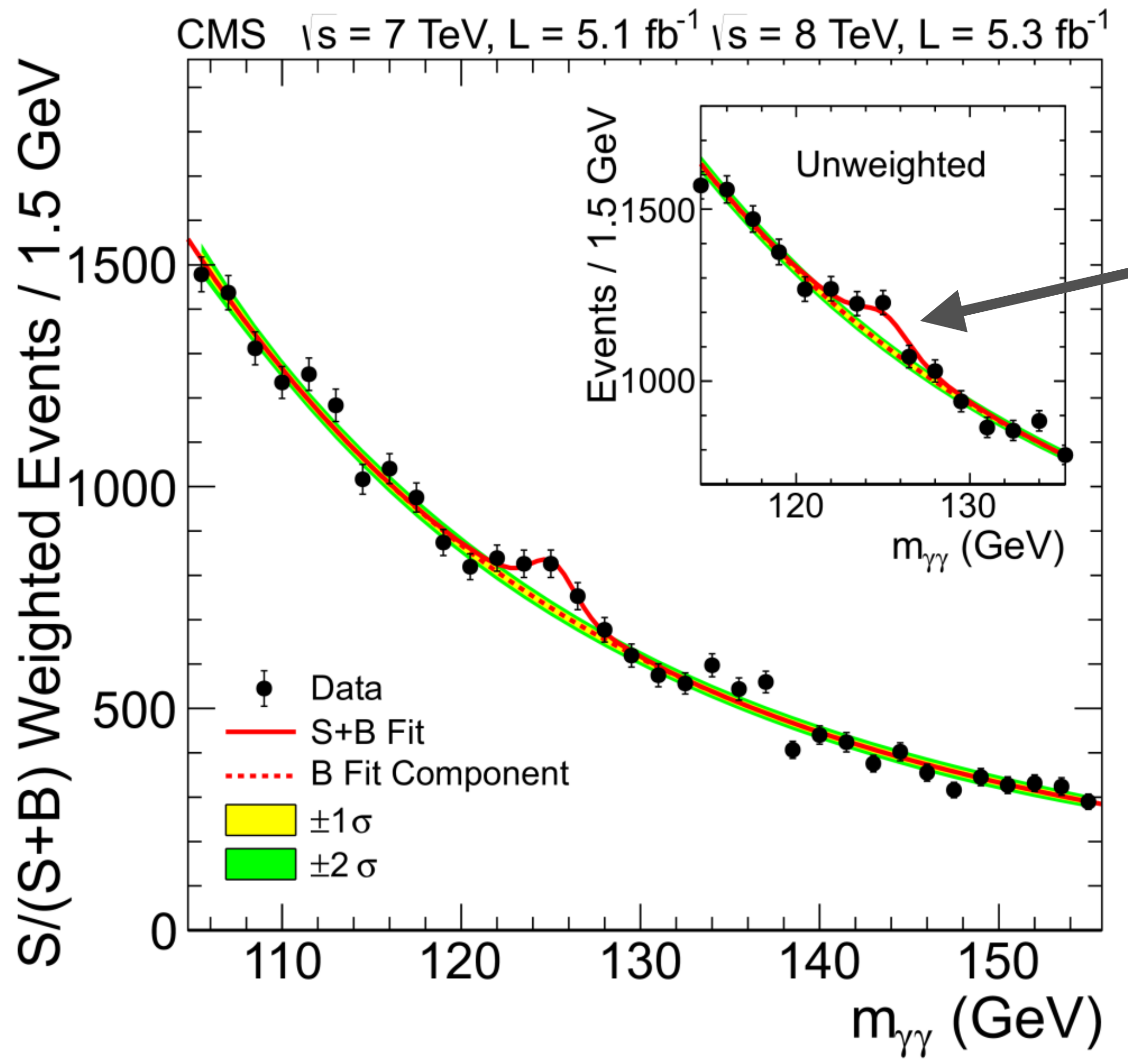
Role of ML in the Higgs Discovery

- Example: Higgs discovery
  - improvements in all area of analysis thanks to ML
  - Initially mainly BDT were used:
    - to separate signal events ($H \to \gamma\gamma$) from background
    - to improve photon energy

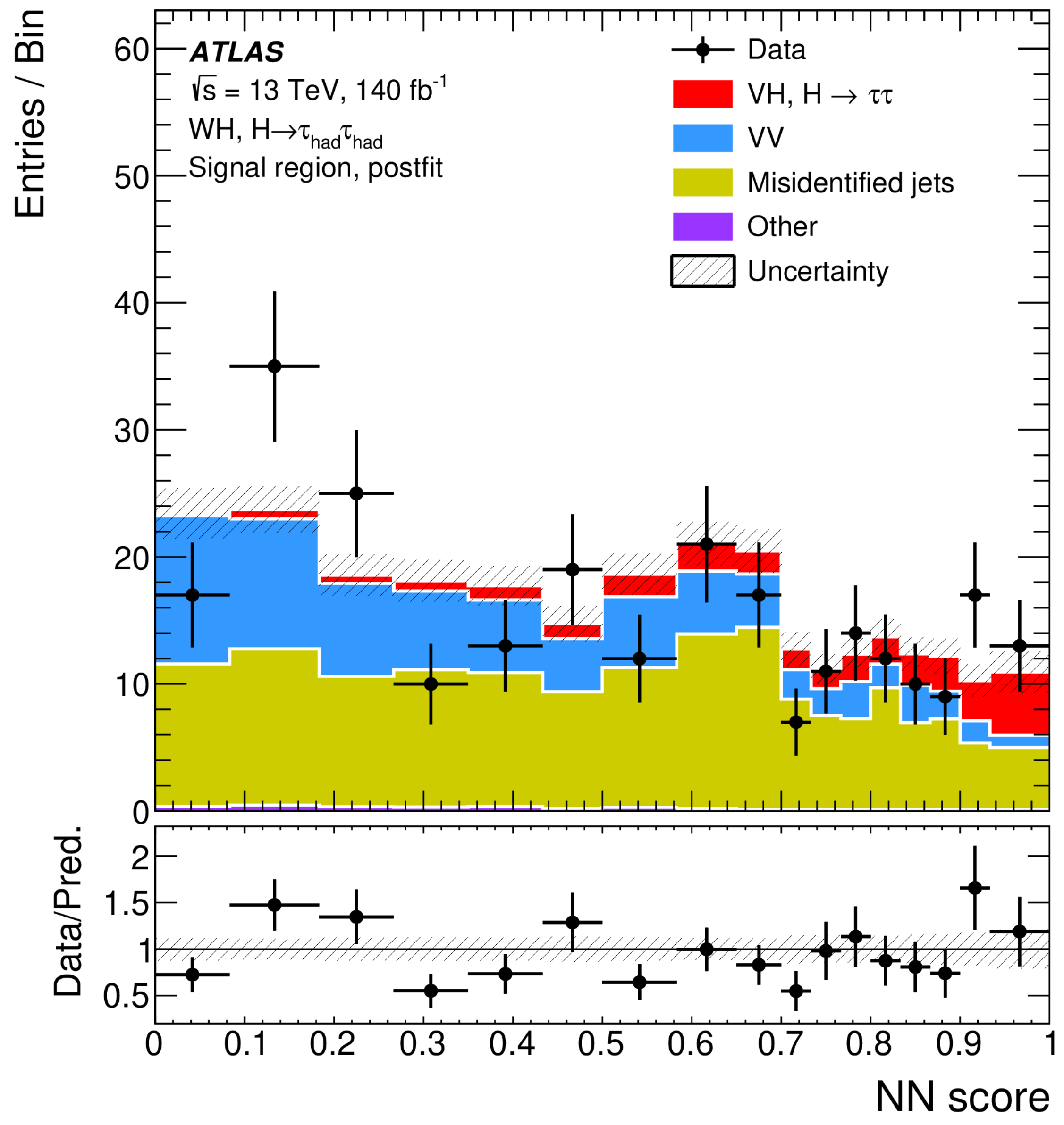
Table 1 | Effect of machine learning on the discovery and study of the Higgs boson

Analysis	Years of data collection	Sensitivity without machine learning	Sensitivity with machine learning	Ratio of P values	Additional data required
CMS[24] $H \to \gamma\gamma$	2011–2012	2.2σ, $P = 0.014$	2.7σ, $P = 0.0035$	4.0	51%
ATLAS[43] $H \to \tau^+\tau^-$	2011–2012	2.5σ, $P = 0.0062$	3.4σ, $P = 0.00034$	18	85%
ATLAS[99] $VH \to bb$	2011–2012	1.9σ, $P = 0.029$	2.5σ, $P = 0.0062$	4.7	73%
ATLAS[41] $VH \to bb$	2015–2016	2.8σ, $P = 0.0026$	3.0σ, $P = 0.00135$	1.9	15%
CMS[100] $VH \to bb$	2011–2012	1.4σ, $P = 0.081$	2.1σ, $P = 0.018$	4.5	125%

Nature 560, 41–48 (2018)

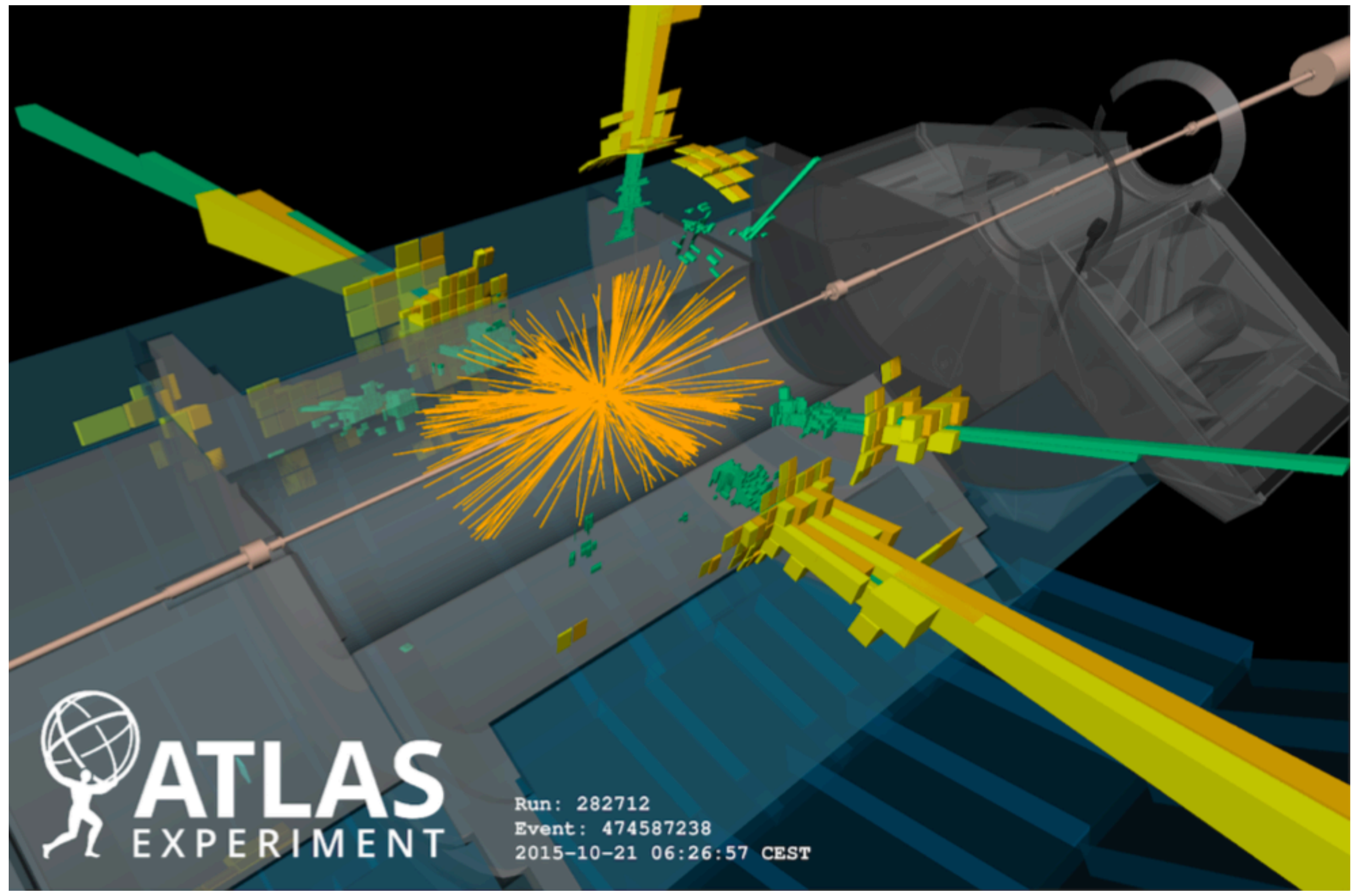


$H \to \gamma\gamma$ events

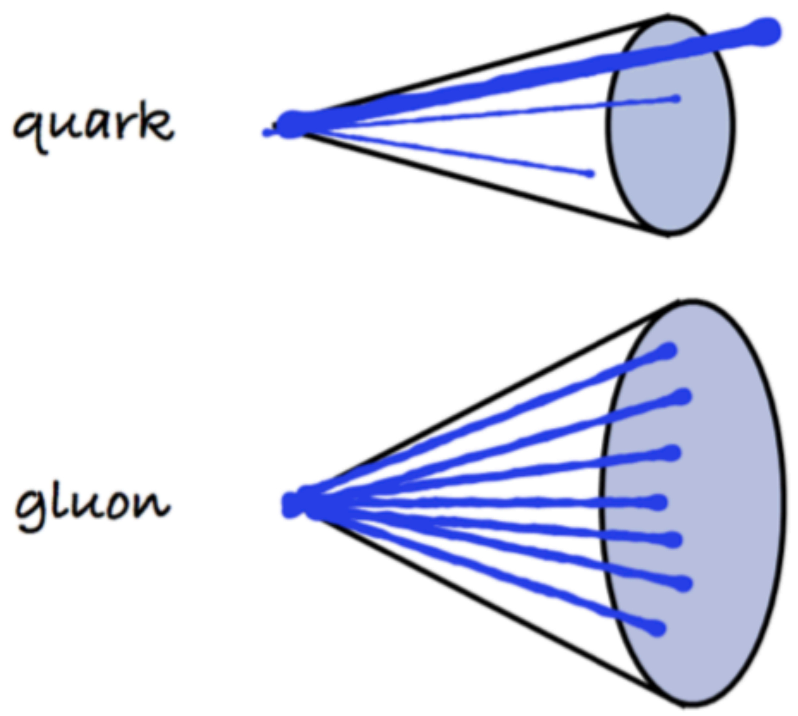


signal events $H \to \tau\tau$

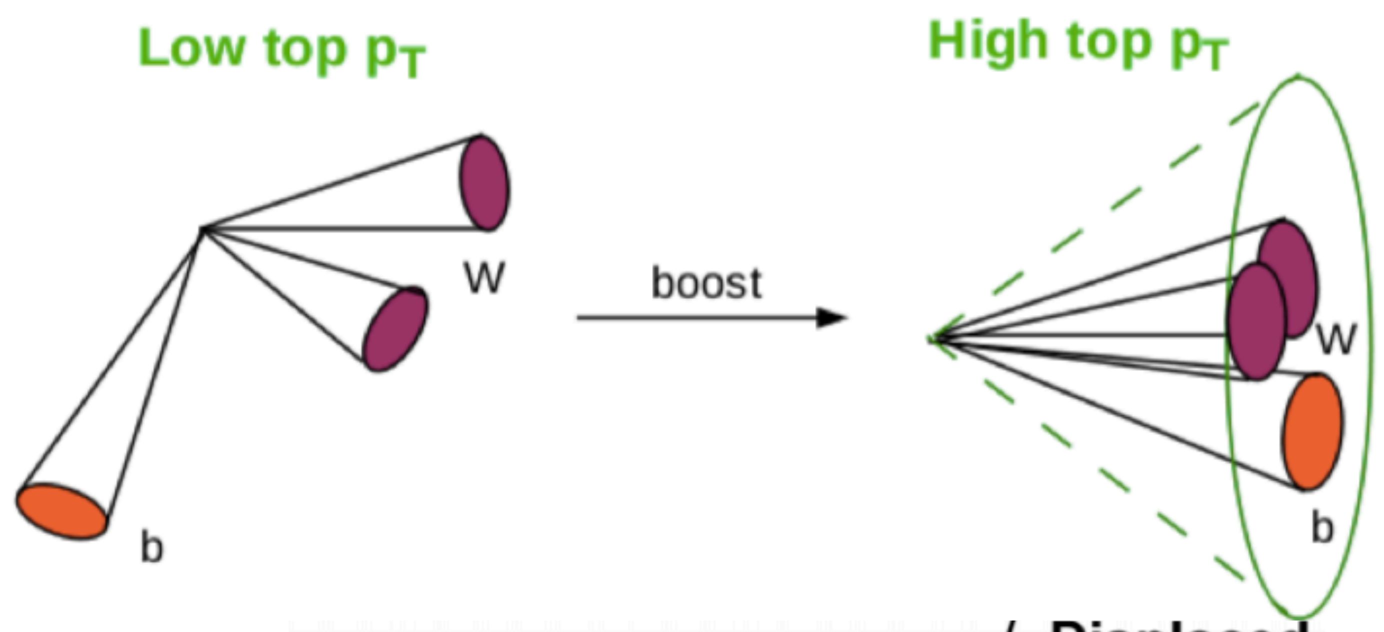
Jet Tagging



quark vs gluon jets

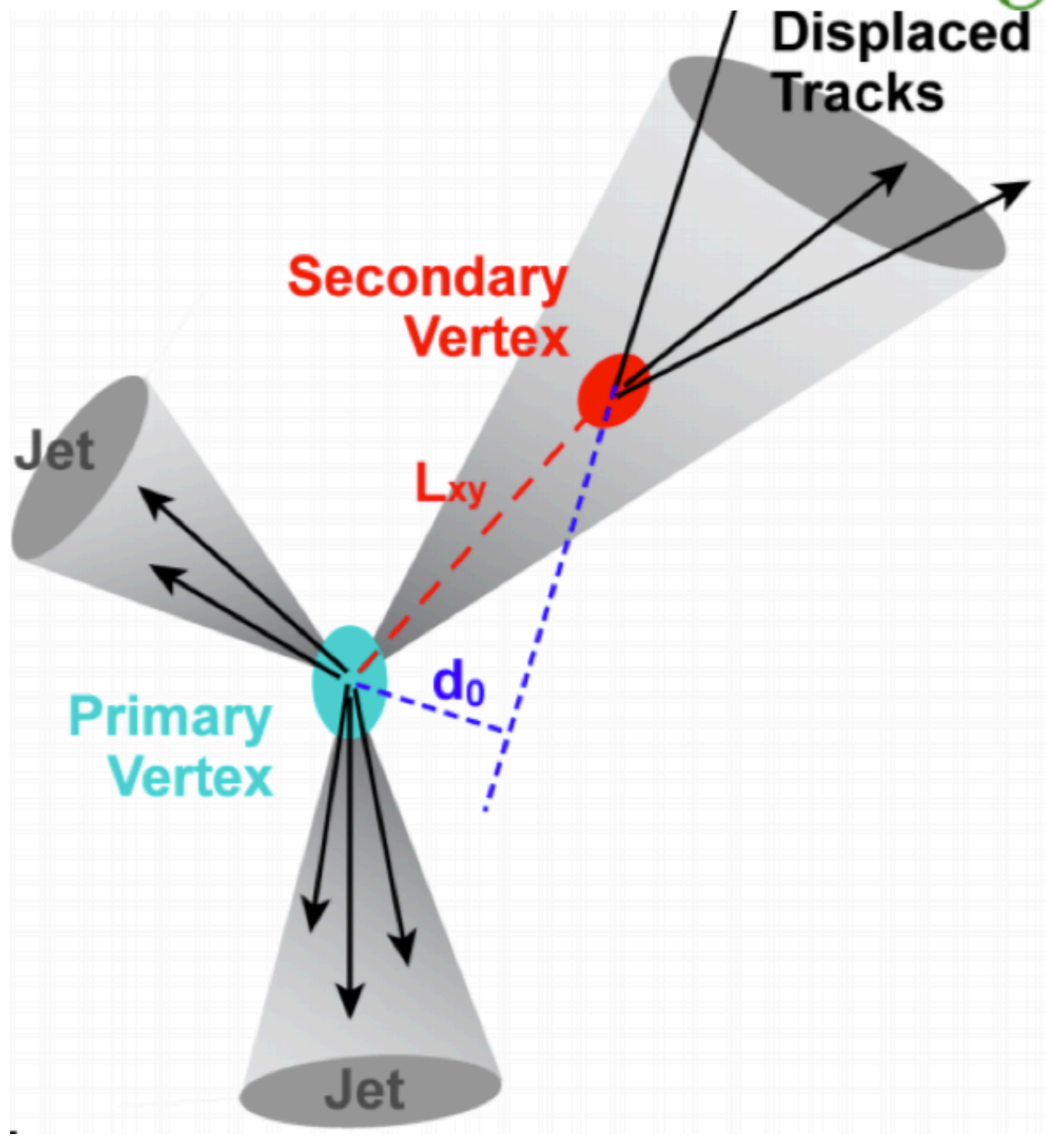


boosted jets



- Distinguish different types of jets (classification task)
  - quark vs gluon jets
  - boosted jets produced by heavy resonances (W, Z, Higgs, top) for new physics searches
  - heavy flavour jets (secondary vertices with displaced tracks)
- Tagging is a difficult task
  - natural to use ML algorithms

heavy flavour jets

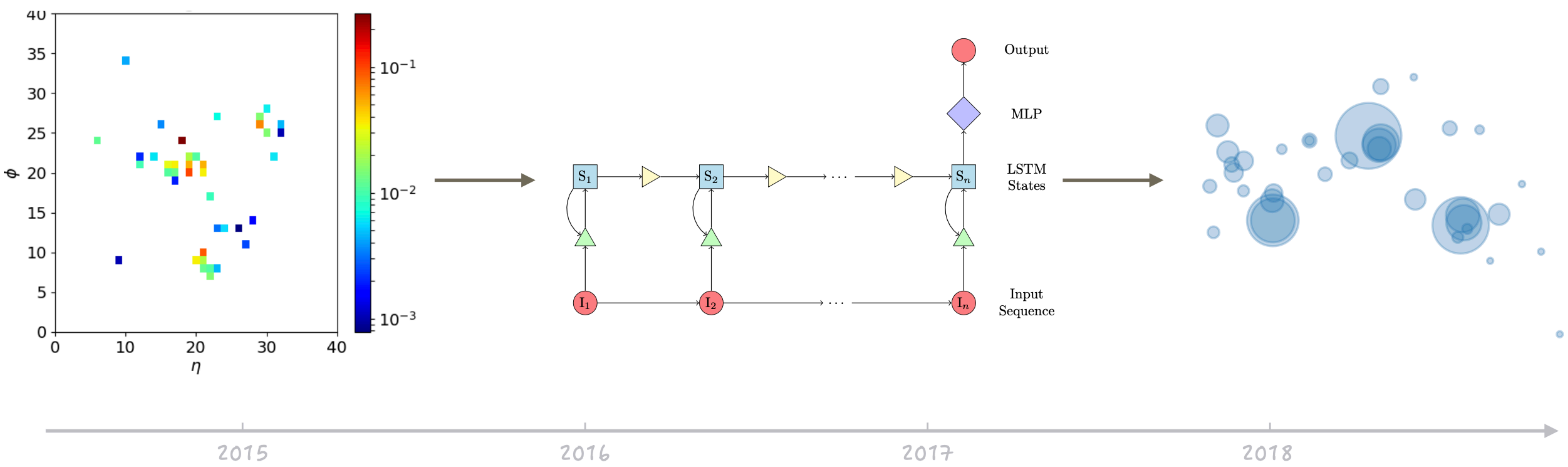


History of Jet Tagging

Image — + 2D ConvNets, …

Sequence — + RNNs, 1D ConvNets, …

Point / Particle cloud — + DeepSets, GNNs, Transformers, …



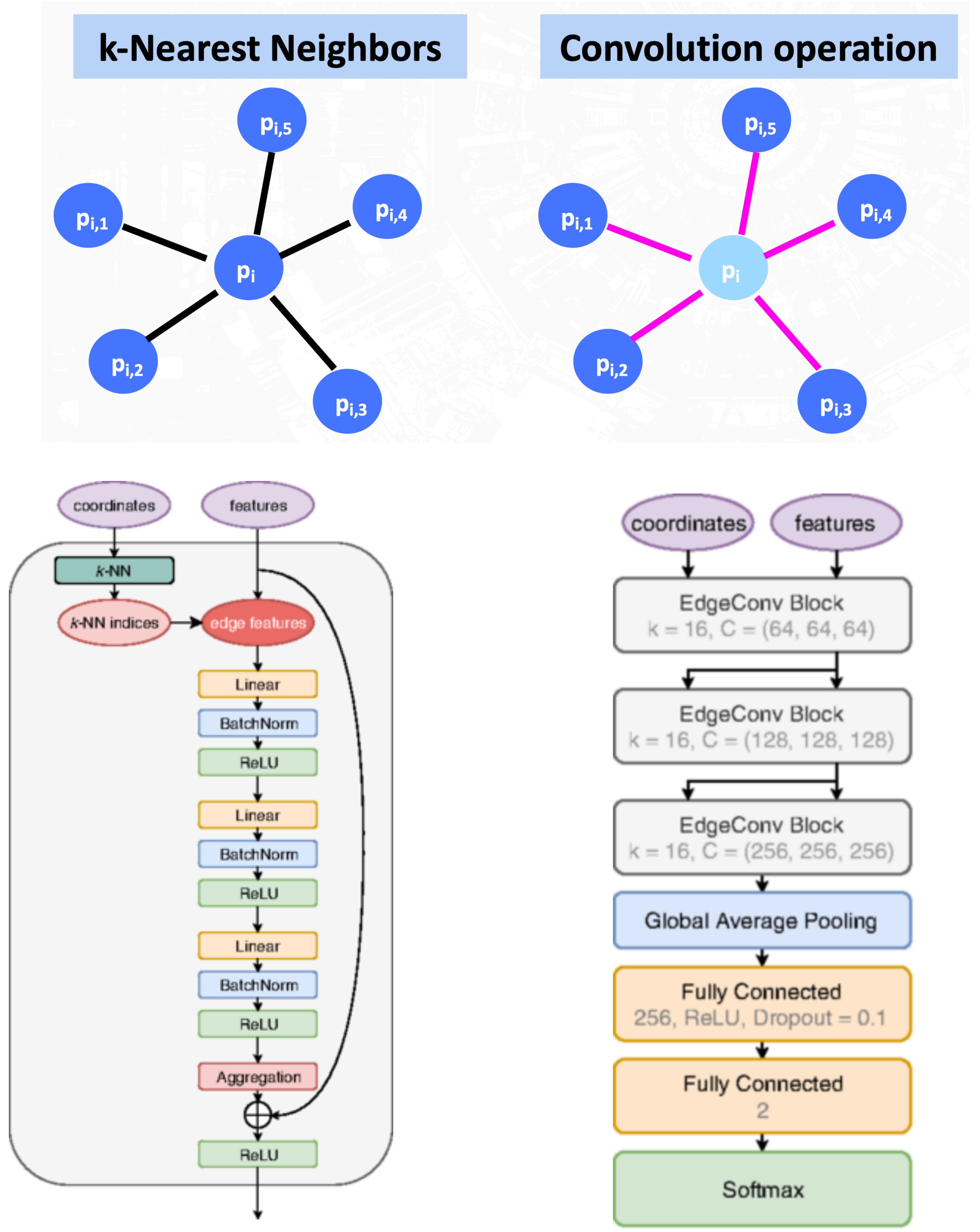
H. Qu (ML4JETS)

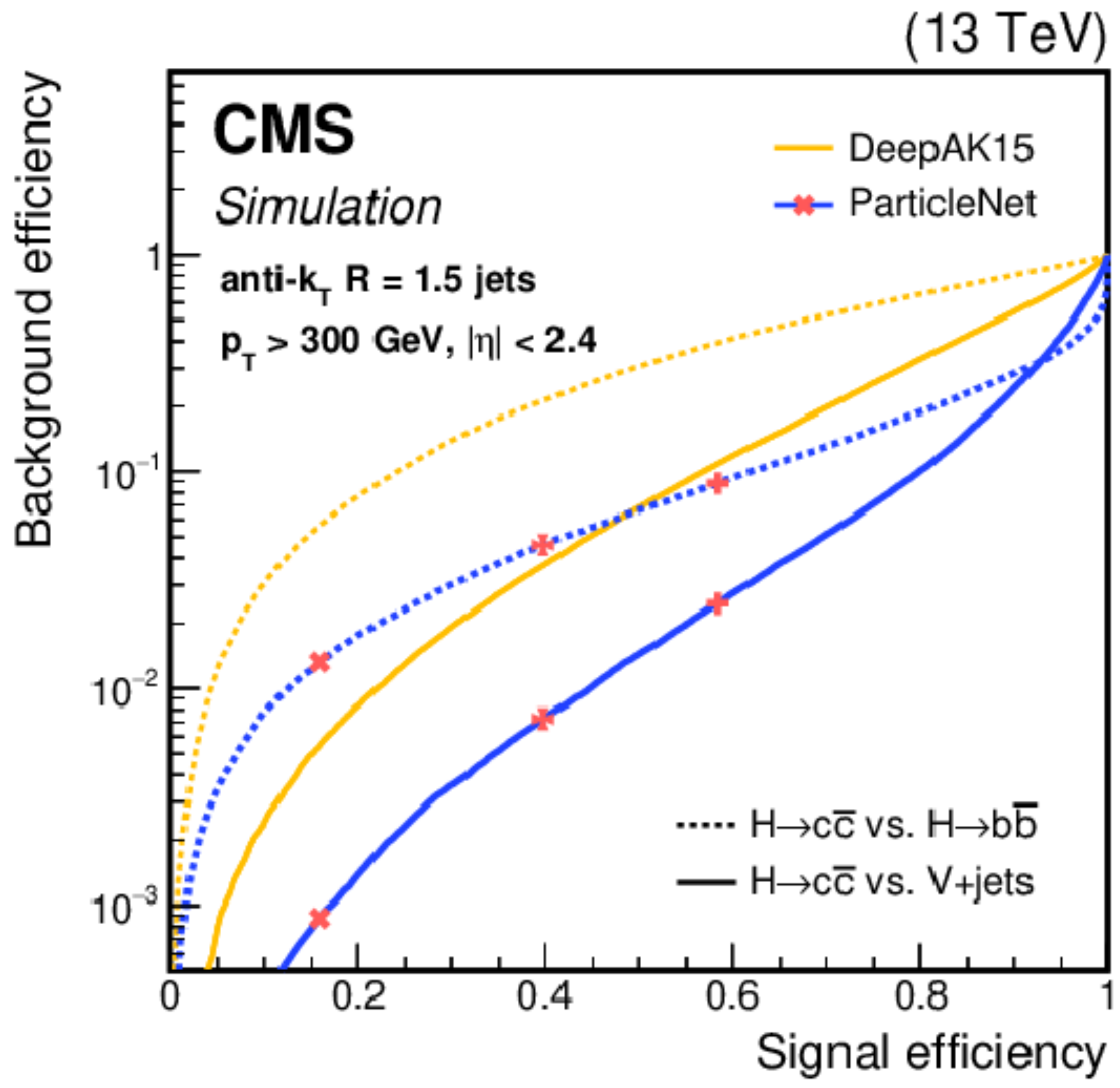
- Follow the evolution happening in AI
  - consider jets as 2D images and use Convolutional Neural Networks (CNN)
  - jets are a sequence of particles, analogy with NLP: use Recurrent NN and 1D CNN
    - physics (QCD) is the grammar that is going to be learnt
  - new models are based on **Graph Neural Network (GNN)** and **Transformers** to continue exploiting NLP analogy

CMS Jet Tagger: Particle Net

GNN for tagging heavy flavour jets (b/c jets)

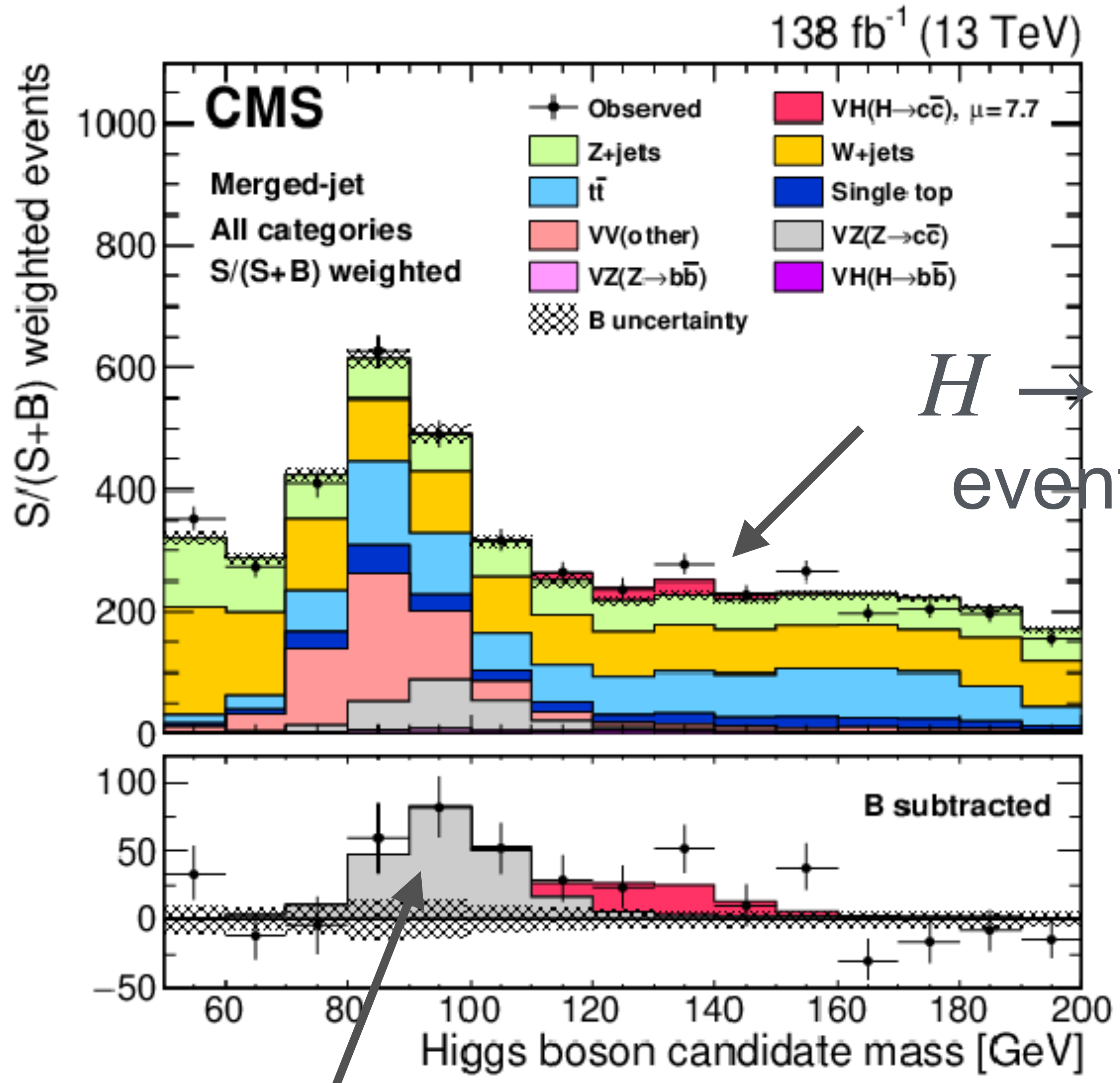
- based on EdgeConv (DGCNN)
- large improvement in performance compared to previous taggers (based on 1D-Convolution + recurrent architecture)





arXiv:1902.08570

Searches for $H \rightarrow c\bar{c}$

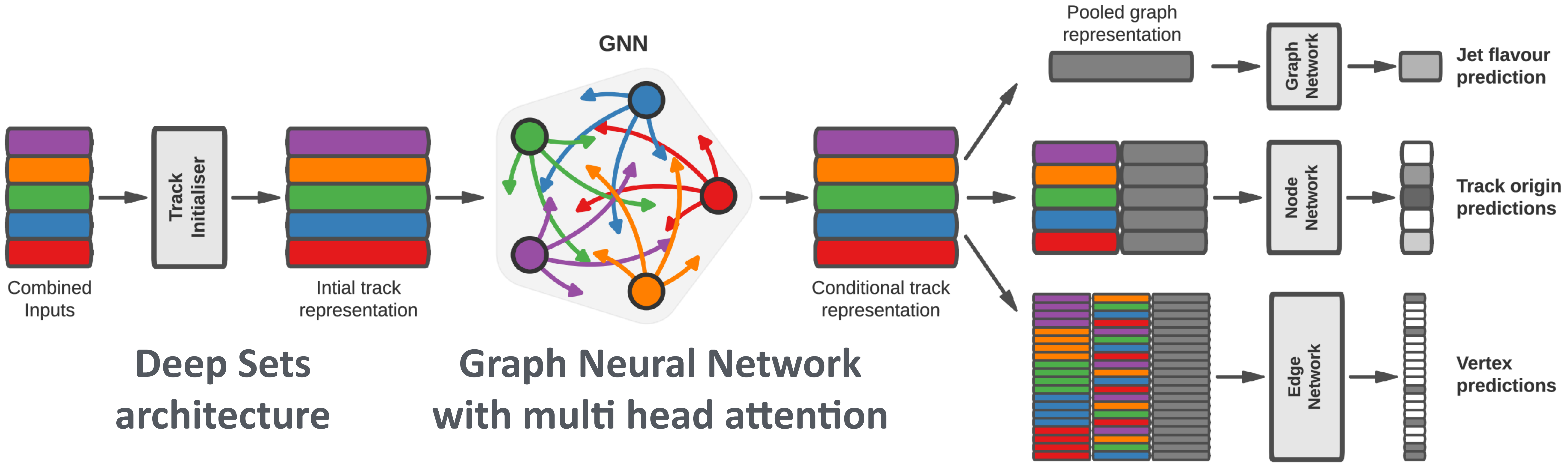


$H \rightarrow c\bar{c}$ events

First observation $Z \rightarrow c\bar{c}$ at collider

ATLAS GNN Tagger

Similar model developed by ATLAS for heavy flavour jet tagging

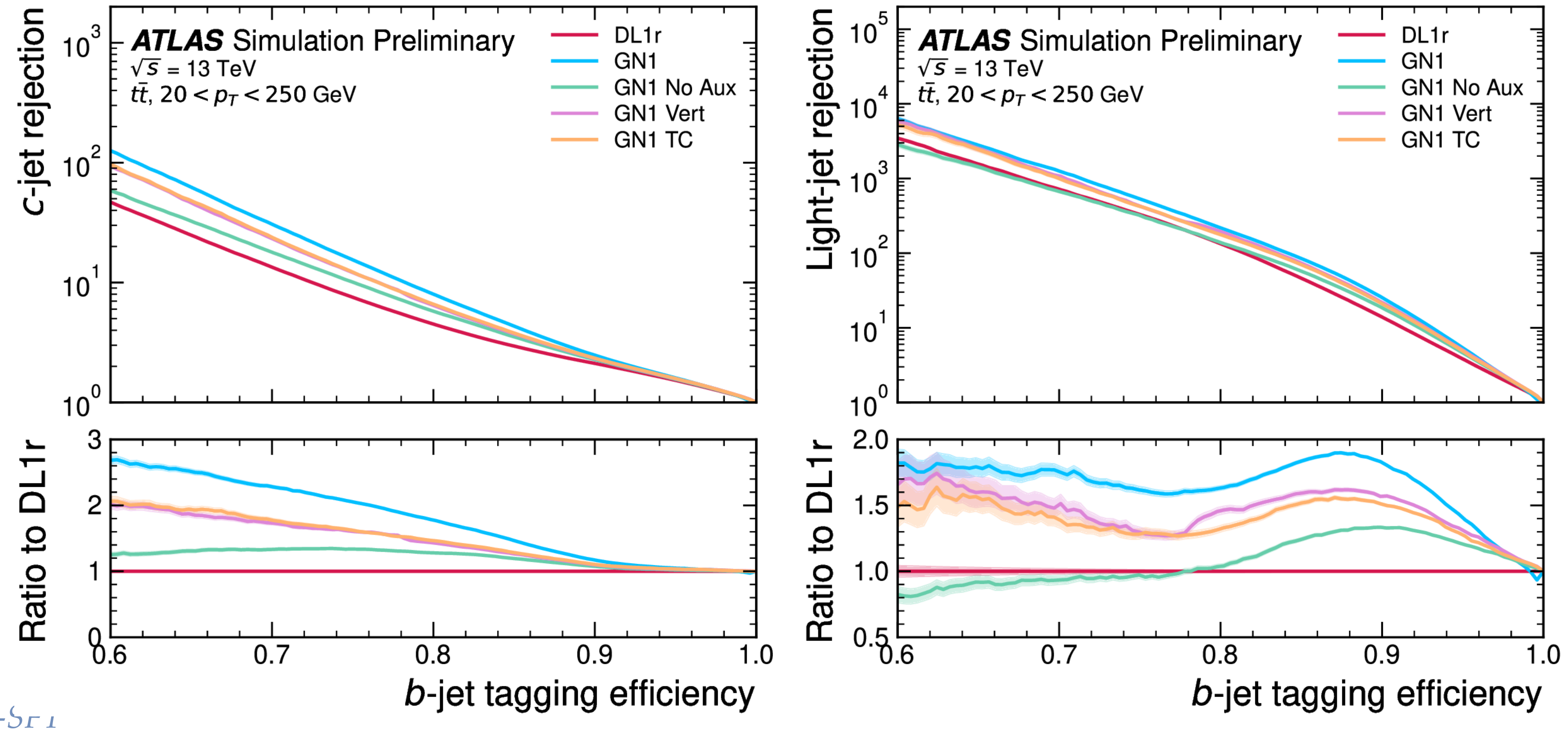


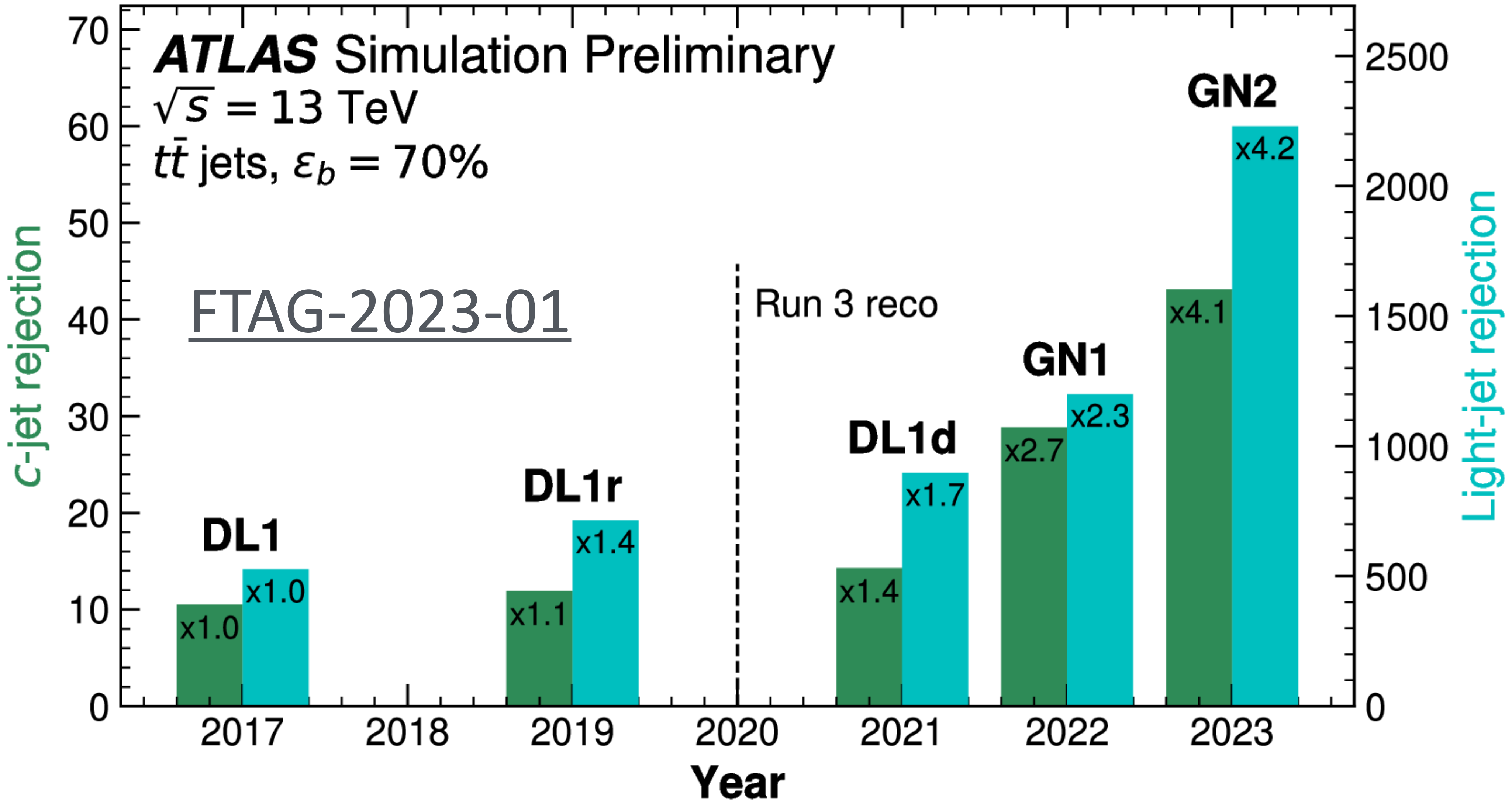
Deep Sets architecture

Graph Neural Network with multi head attention

expensive to train
use resources from
ml.cern.ch (kubeflow)
(M. Draguet, IML WS)

new model **GN2**
follows more closely the
transformer architecture

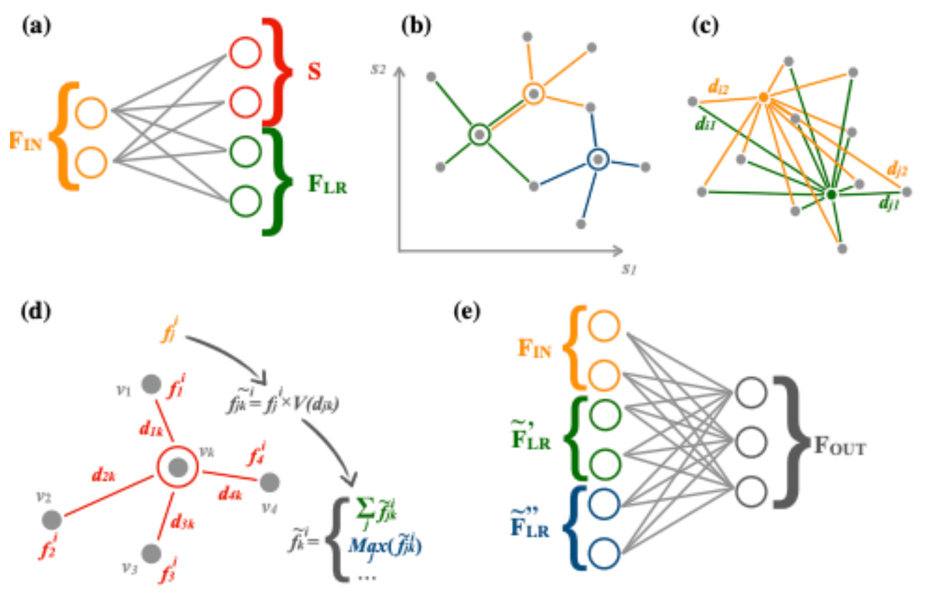




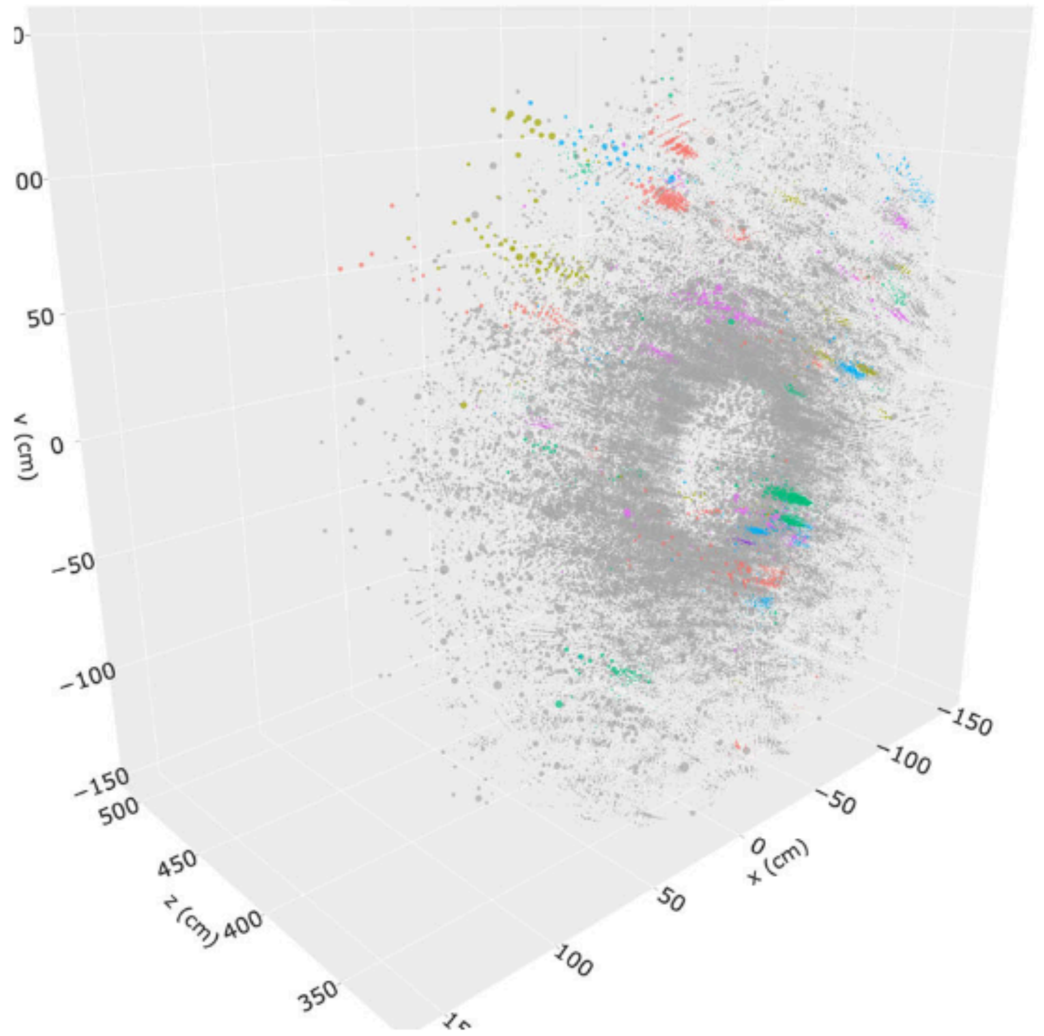
HGCAL Calorimeter Reconstruction

- **End-to-end multi-particle reconstruction** for new high granularity calorimeter (CMS for HL-LHC)
  - exploiting distance-weighted GNN
- Object condensation: one-stage multi-object reconstruction
  - supervised clustering of hits belonging to a shower to a “condensation point” by using attractive/repulsive potentials in the loss
  - simultaneously predict the number of showers and their properties

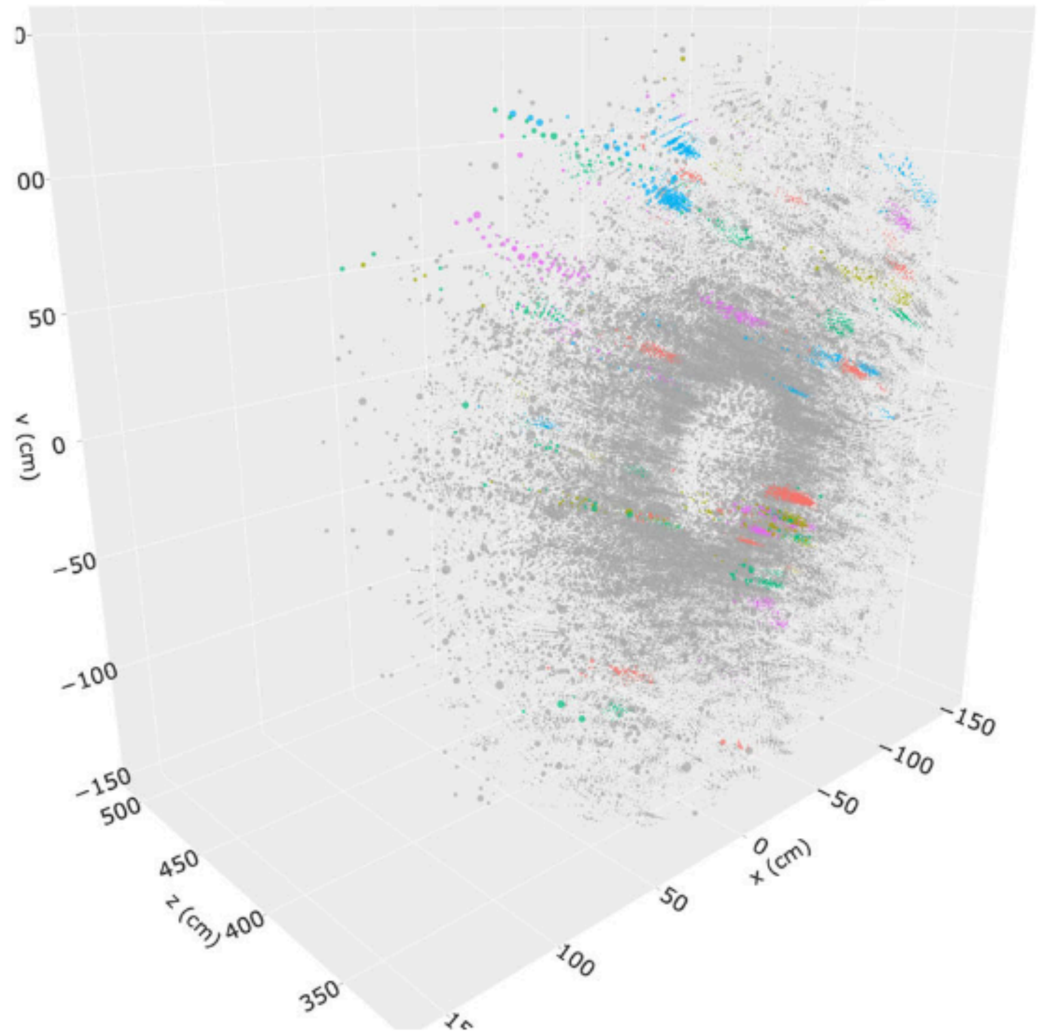
GarNet/GravNet



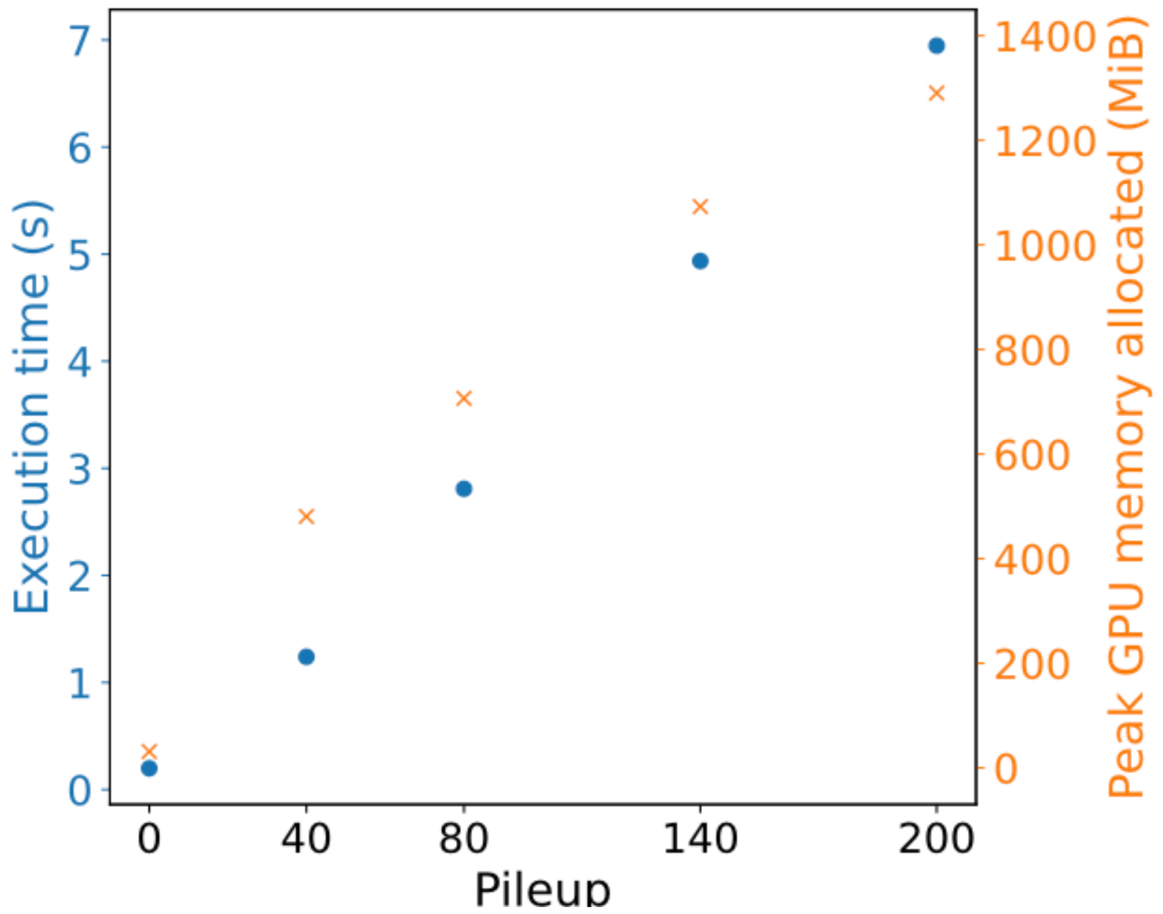
Truth



Reconstructed



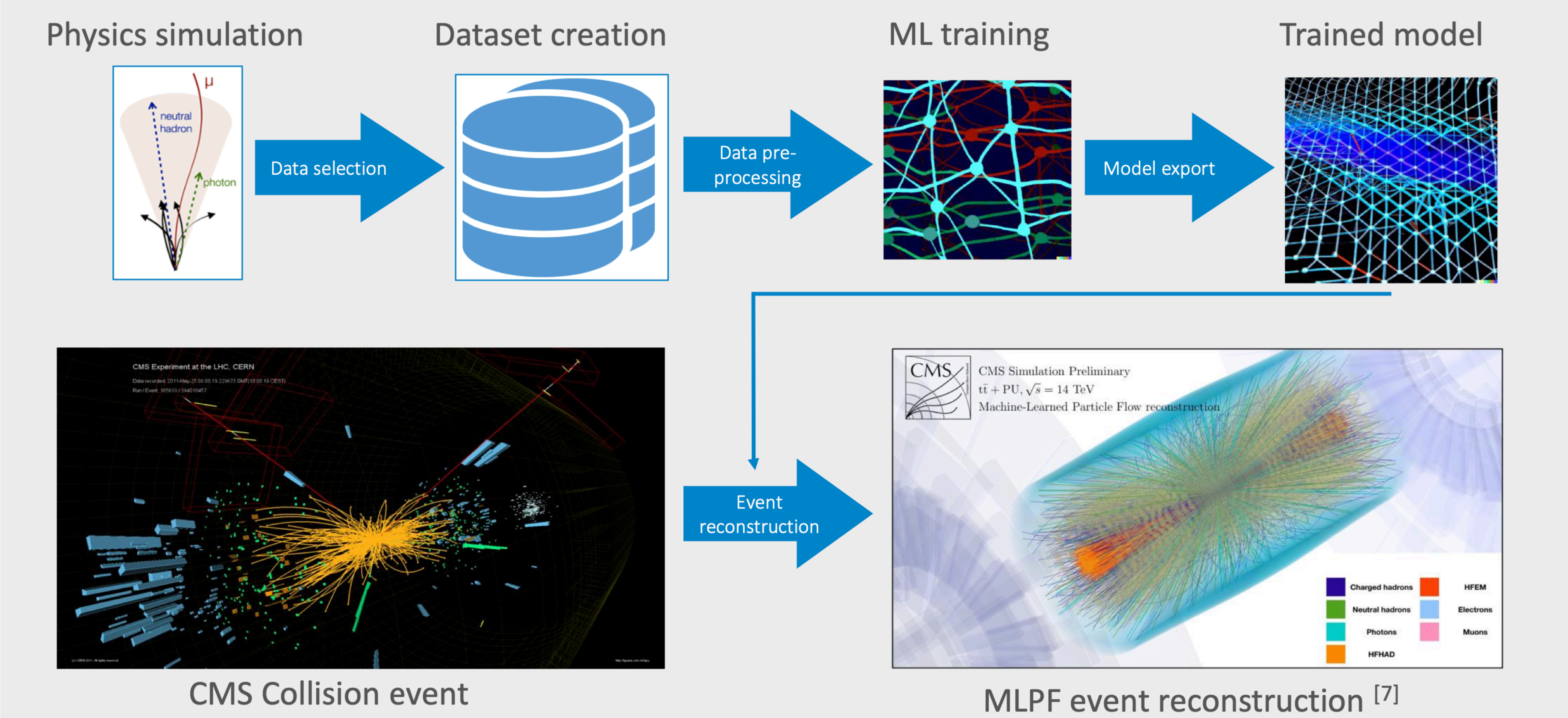
Time and memory usage



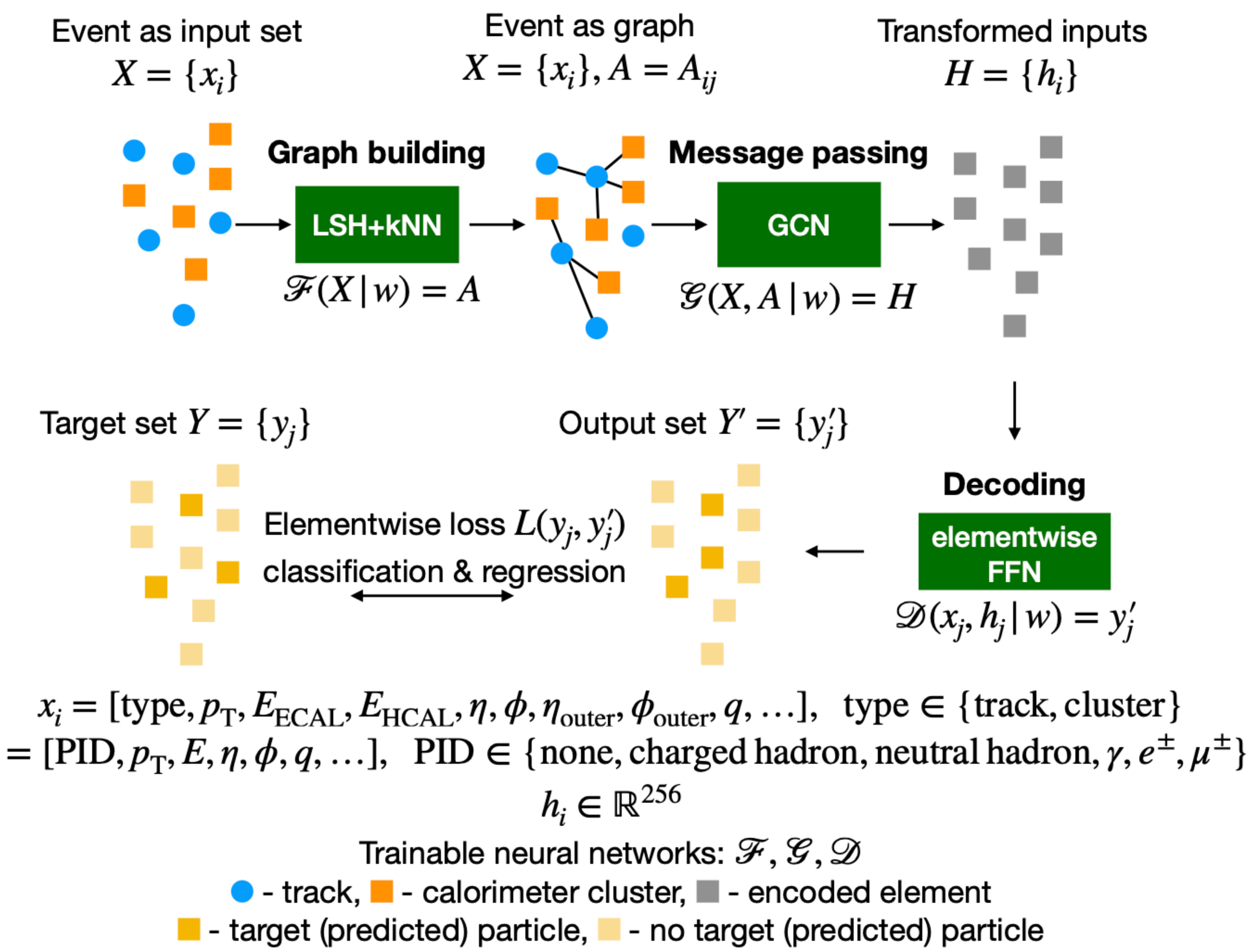
Eur. Phys. J. C (2022) 82

ML for Particle Flow Reconstruction

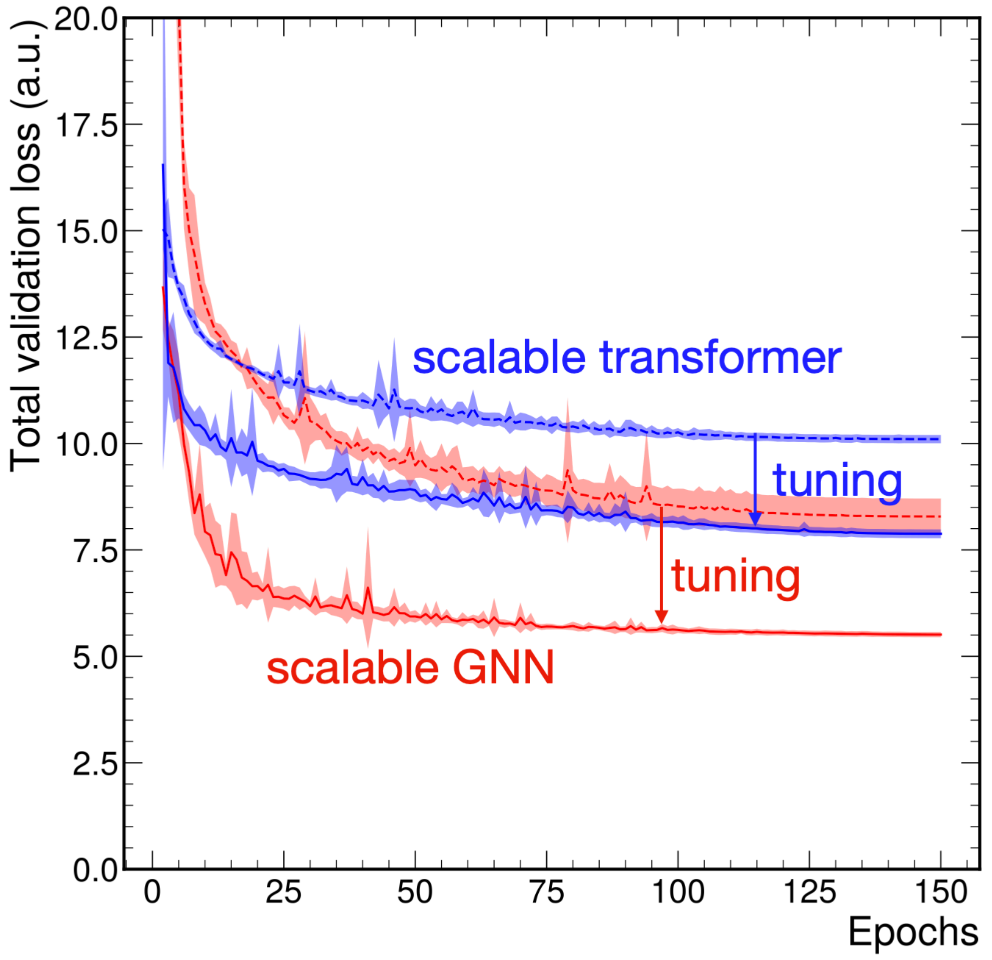
Example from CMS



GNN

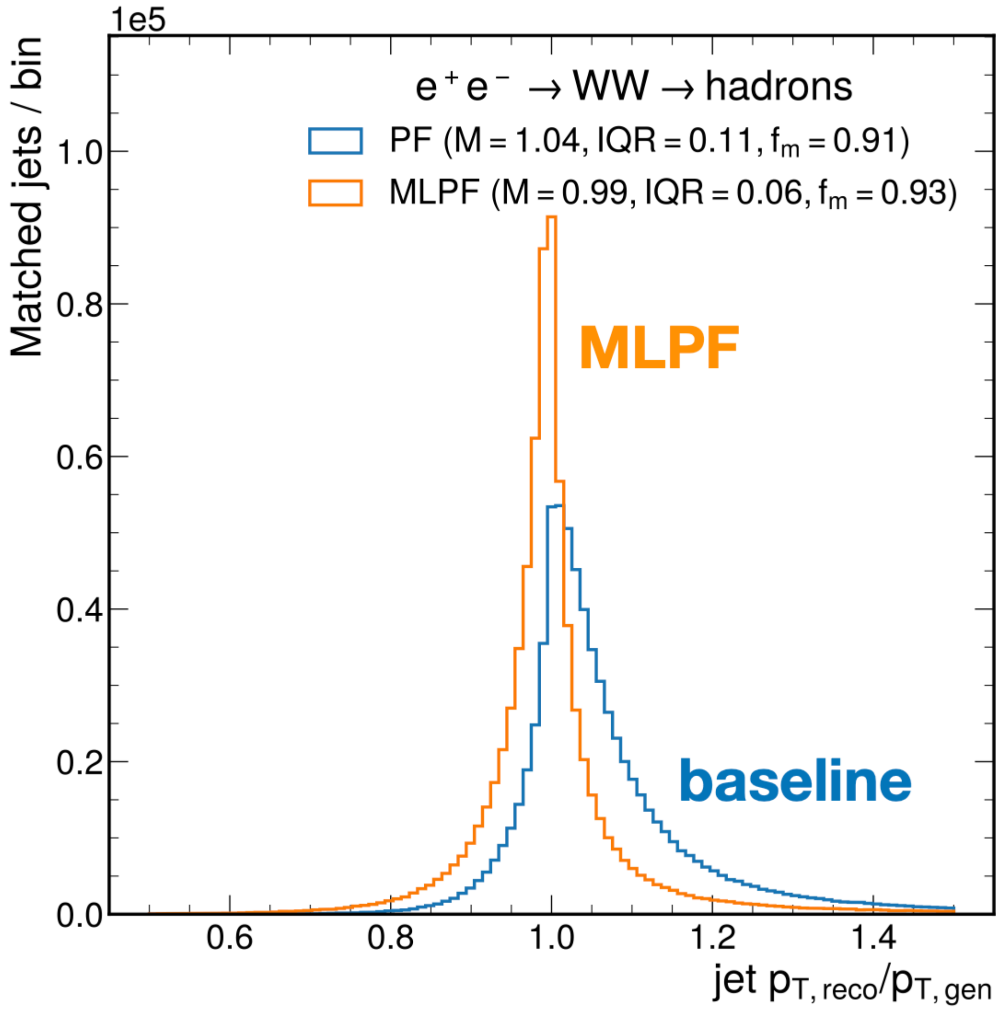


optimised (HPO) using HPC (Julich)



50% improvement in jet response compared to the baseline

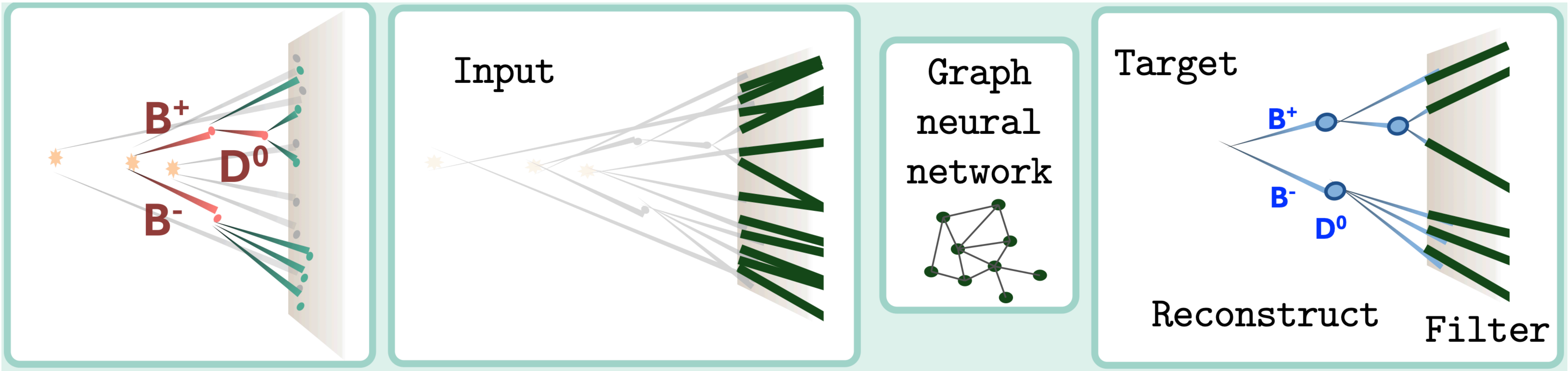
2101.08578



Full Event Interpretation (LHCb)

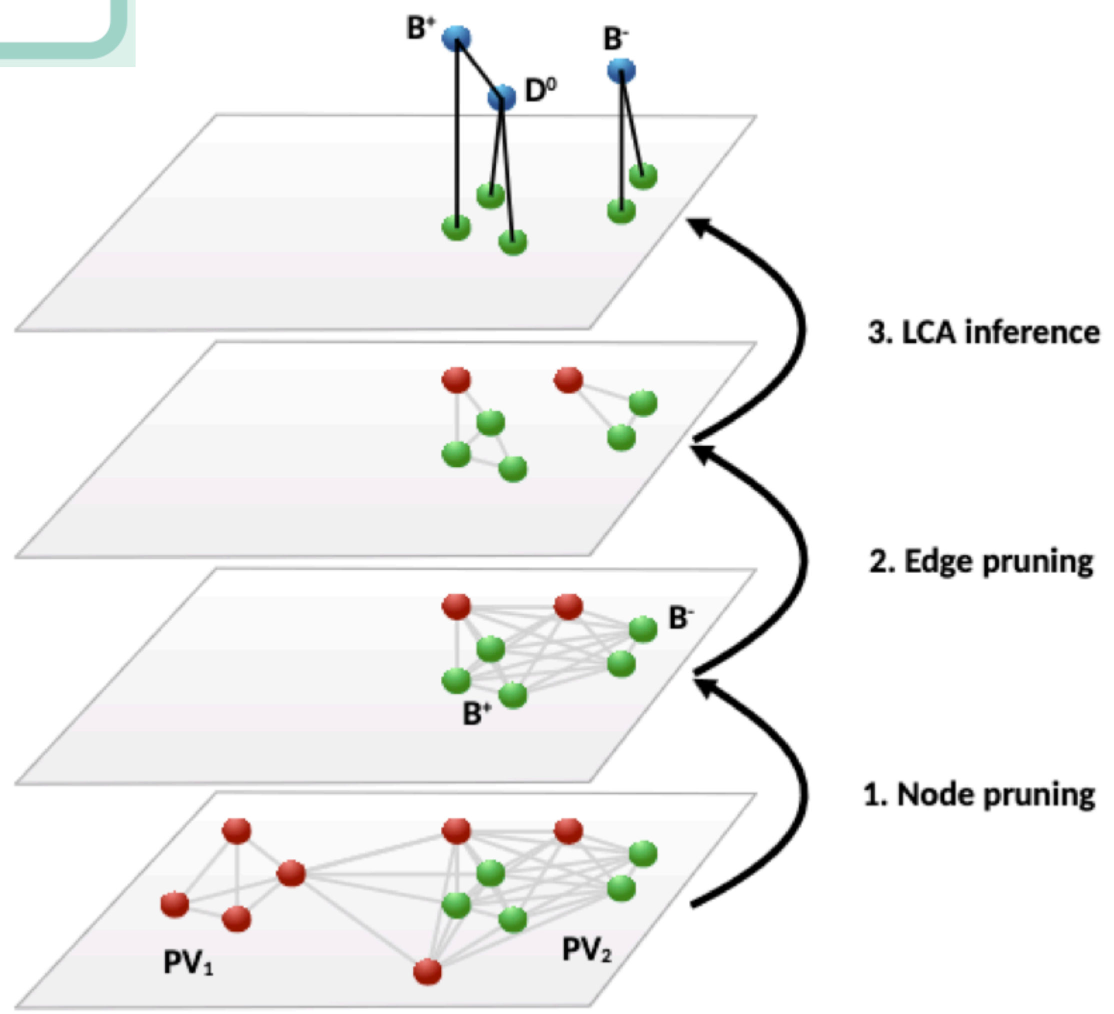
Deep Learning (GNN) for Full Event Interpretation (LHCb upgrade)

- simultaneous identification and reconstruction of HF decay chains



[Comput.Softw.Big Sci. 7 (2023) 1, 12].

Input features: charged particles and its measured properties (nodes) and their relations (edges)



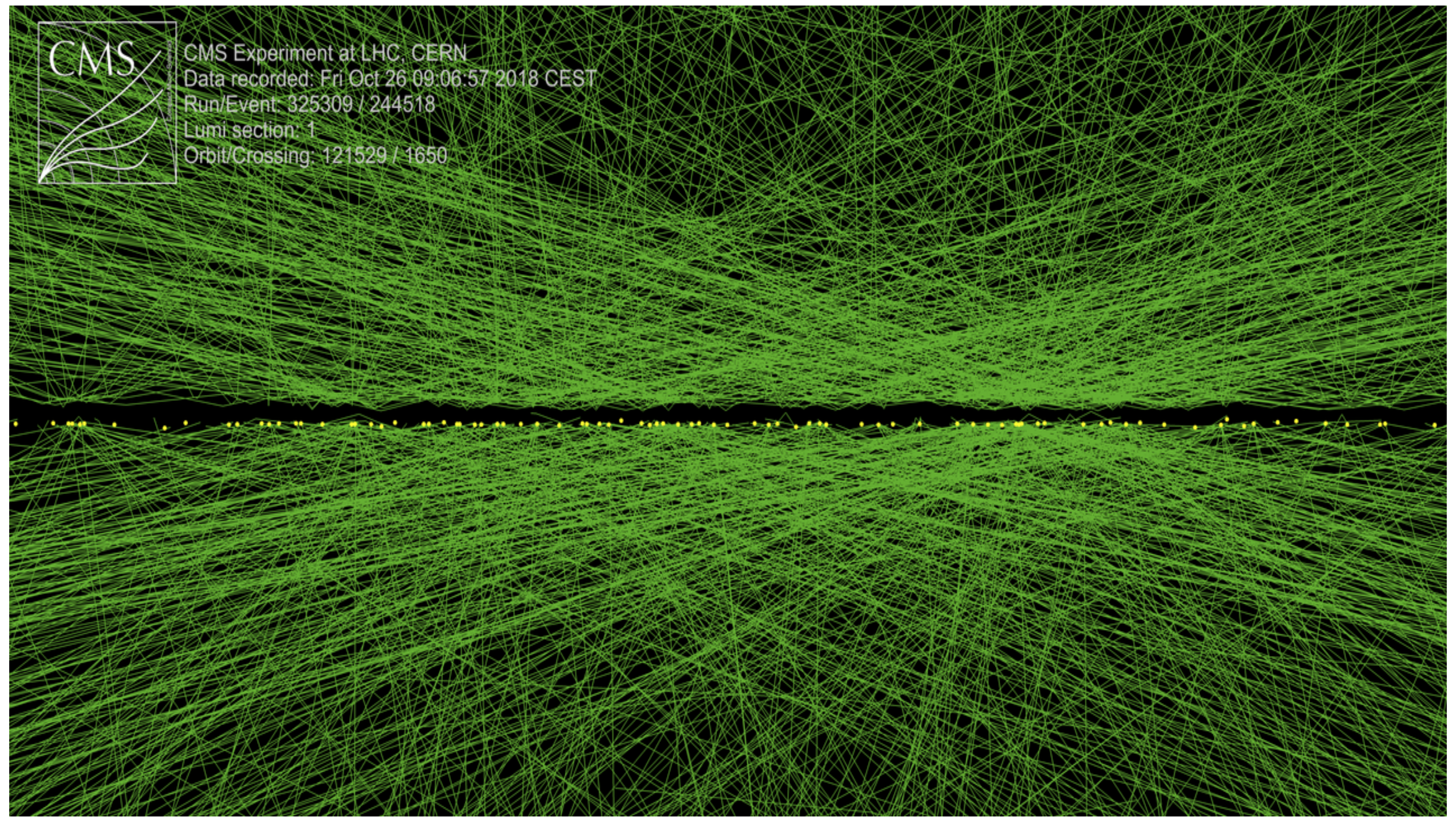
Constructing a C++ inference pipeline based on **SOFIE** to integrate algorithm in LHCb software trigger

ML Applications for Tracking

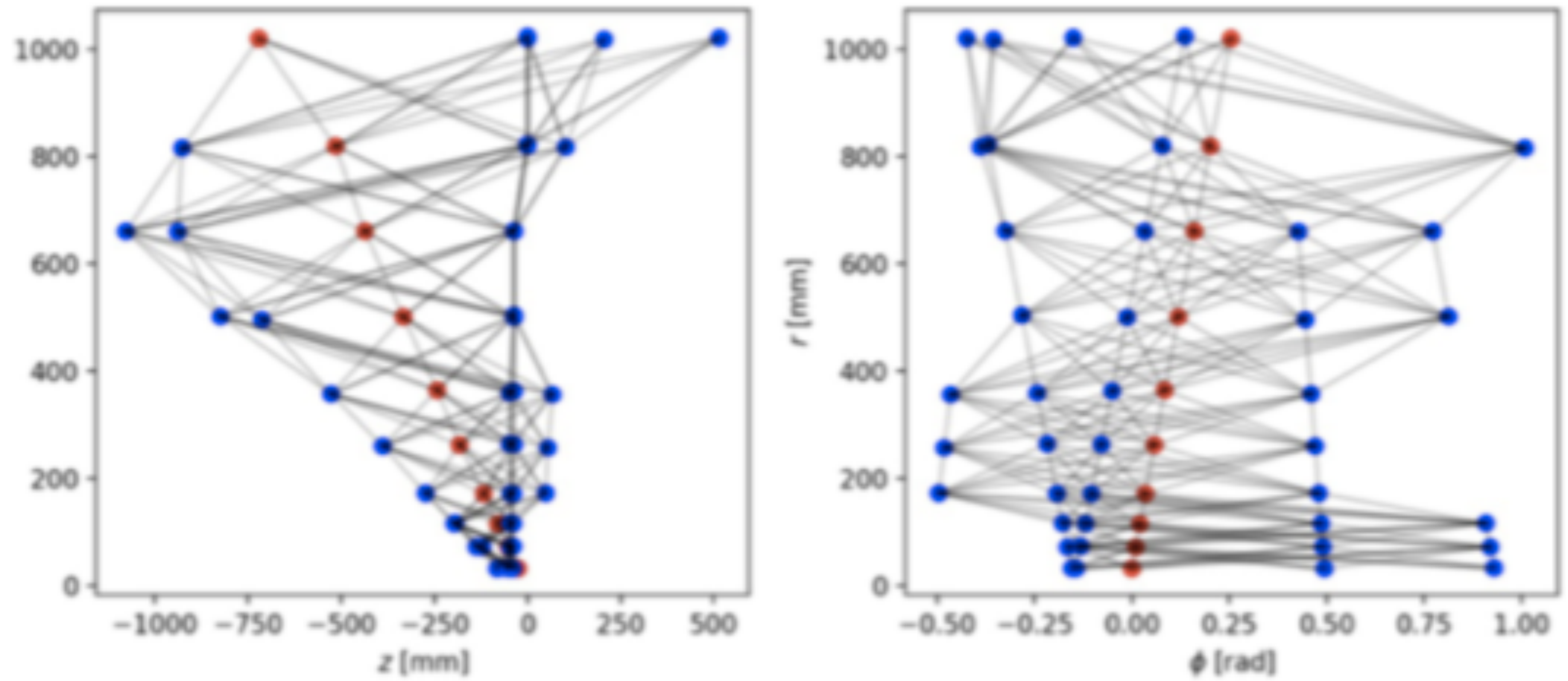
Tracking of particles

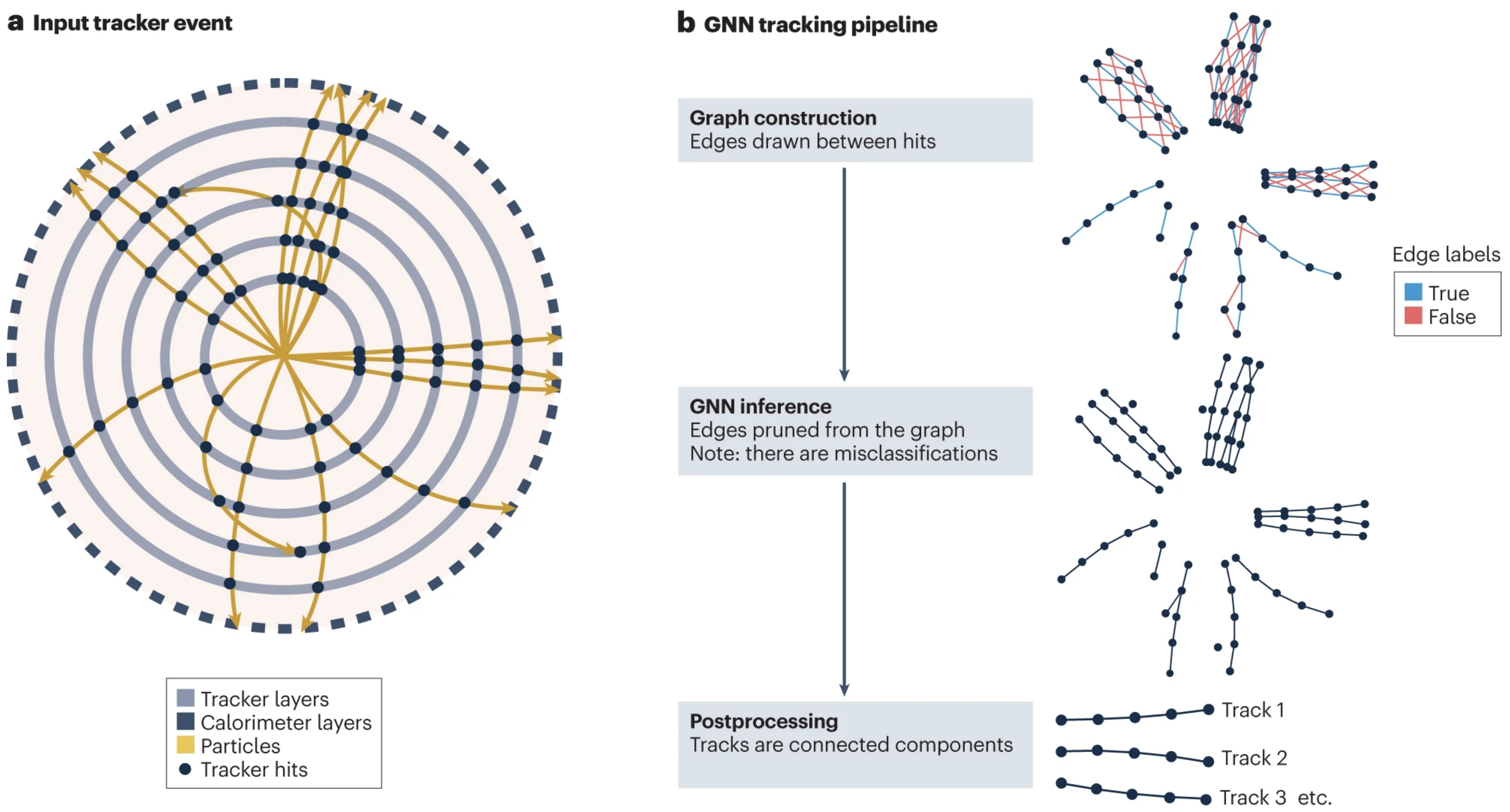
Segment classifier architecture

With each iteration, the model propagates information through the graph, strengthens important connections, and weakens useless ones.



- Unseeded hit-pair classification
- Model predicts the probability that a hit-pair is valid



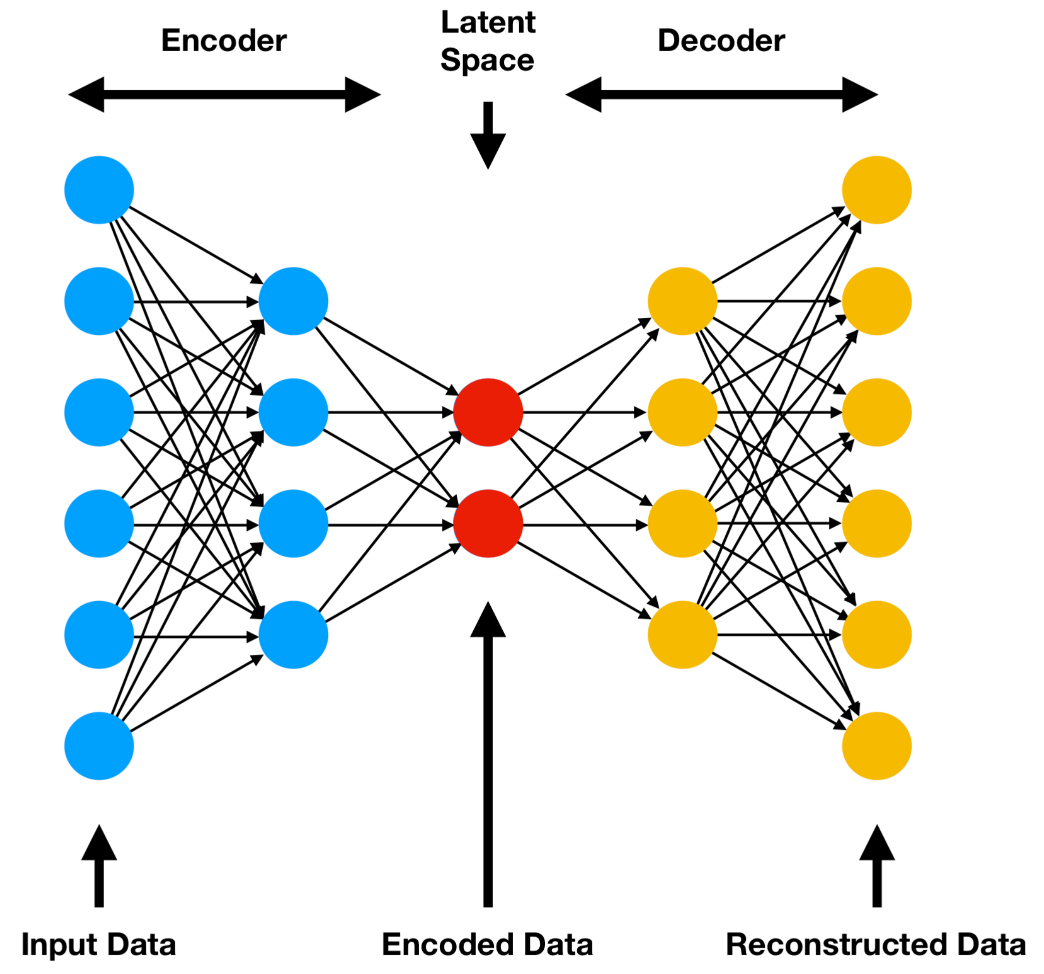


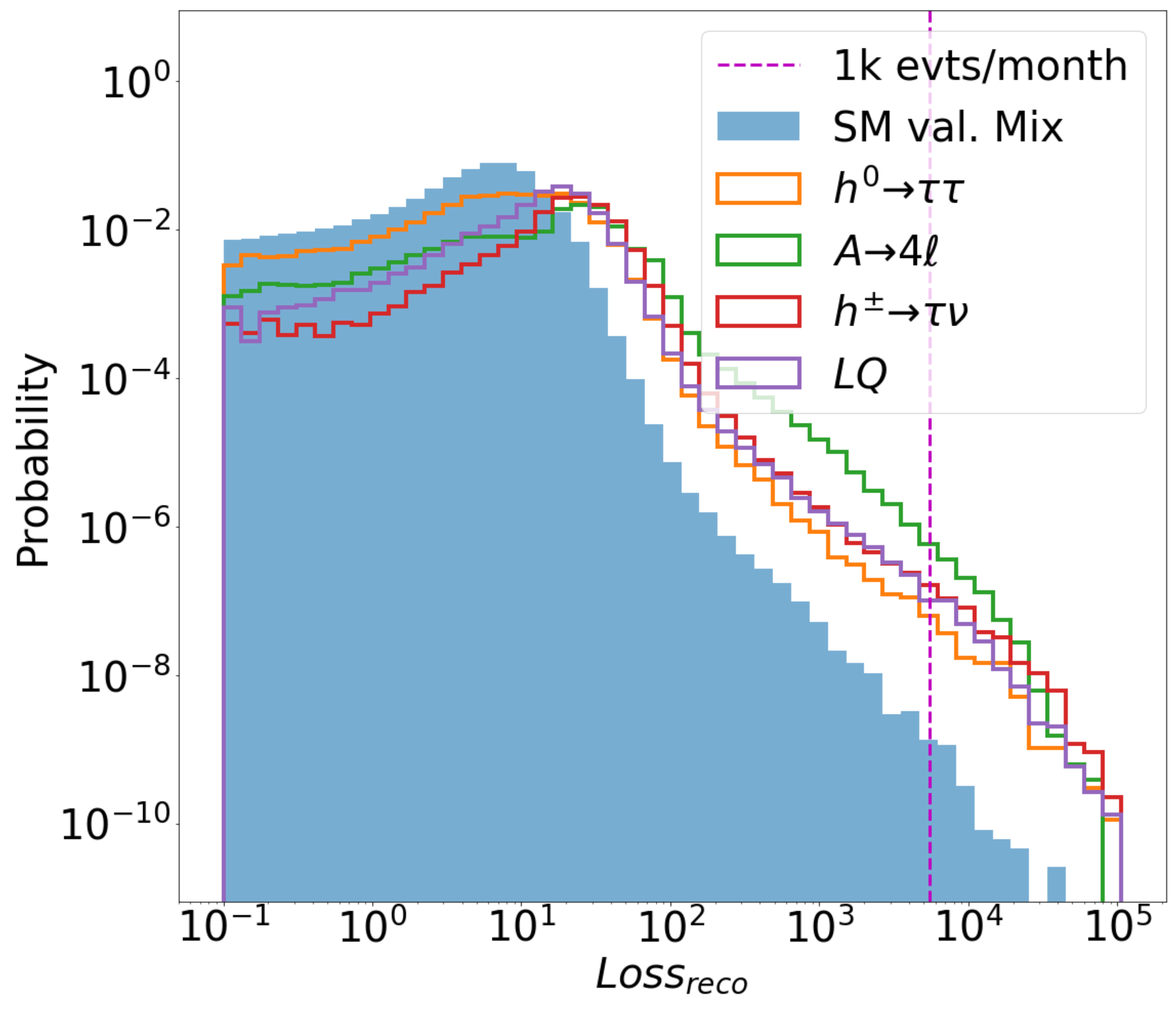
[*Nature Reviews Physics* **5**, 281–303 (2023)](#)

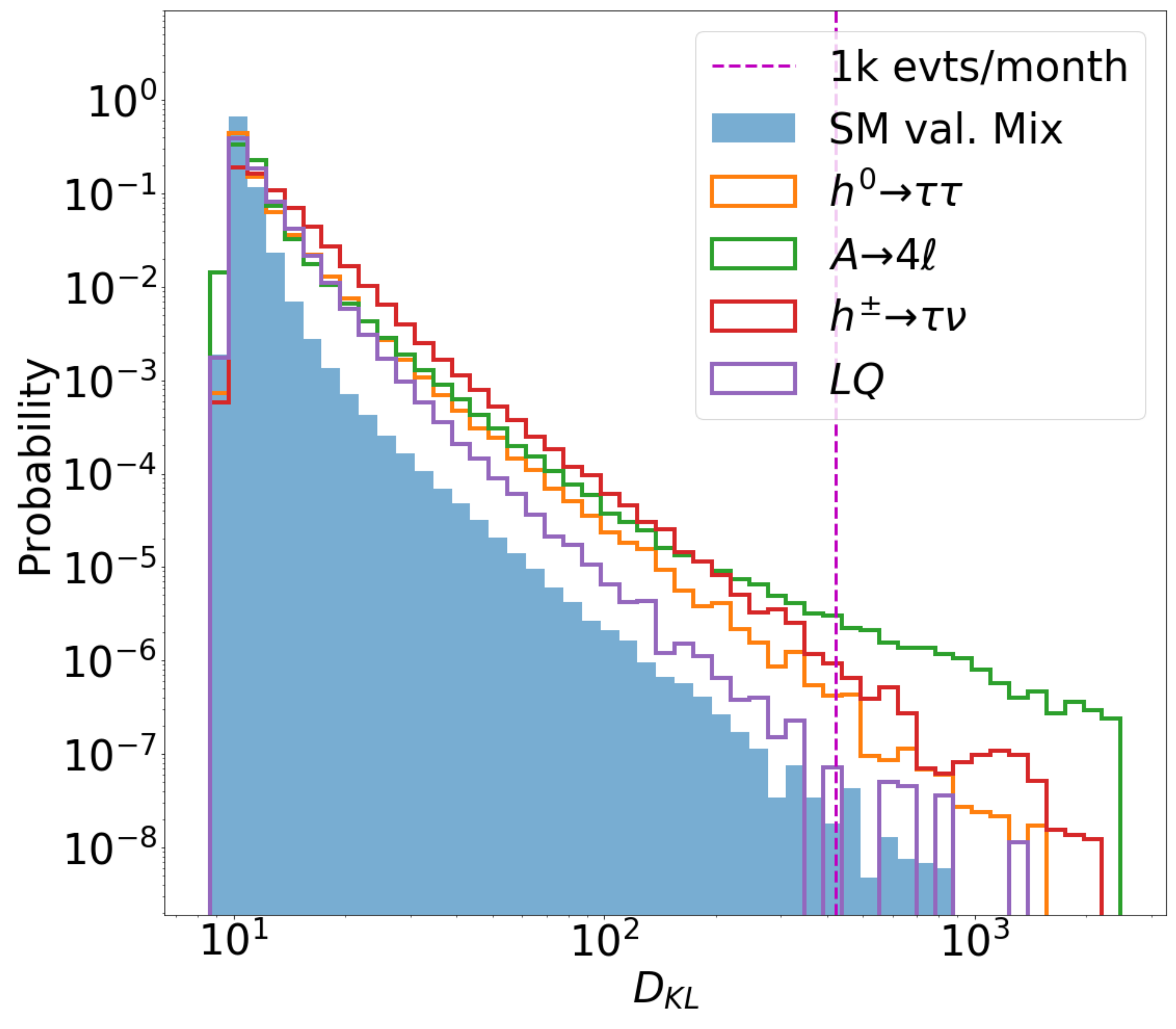
ML for Anomaly Detection

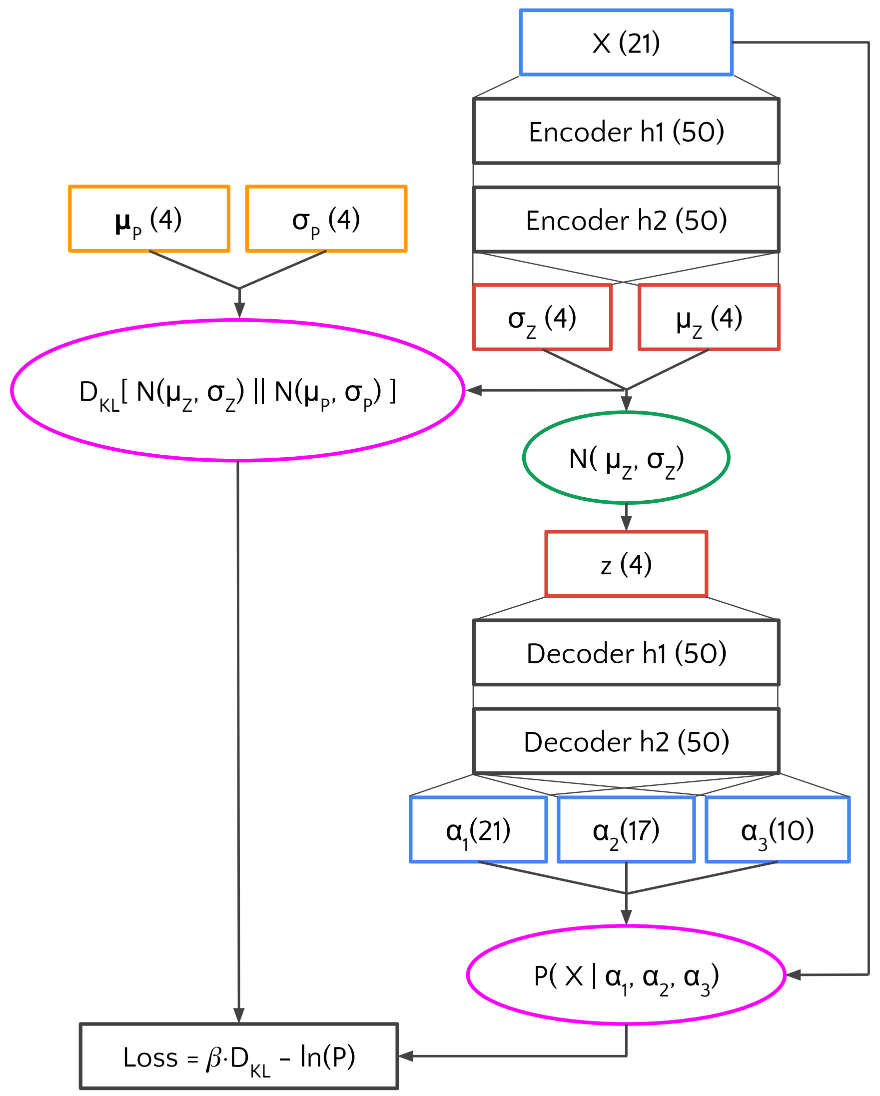
Use unsupervised learning with autoencoder at trigger level (CMS) for detecting anomalous events

- Train model on standard model events (known type)
  - identify anomalies by cutting on loss function
  - record anomalous events for further analysis





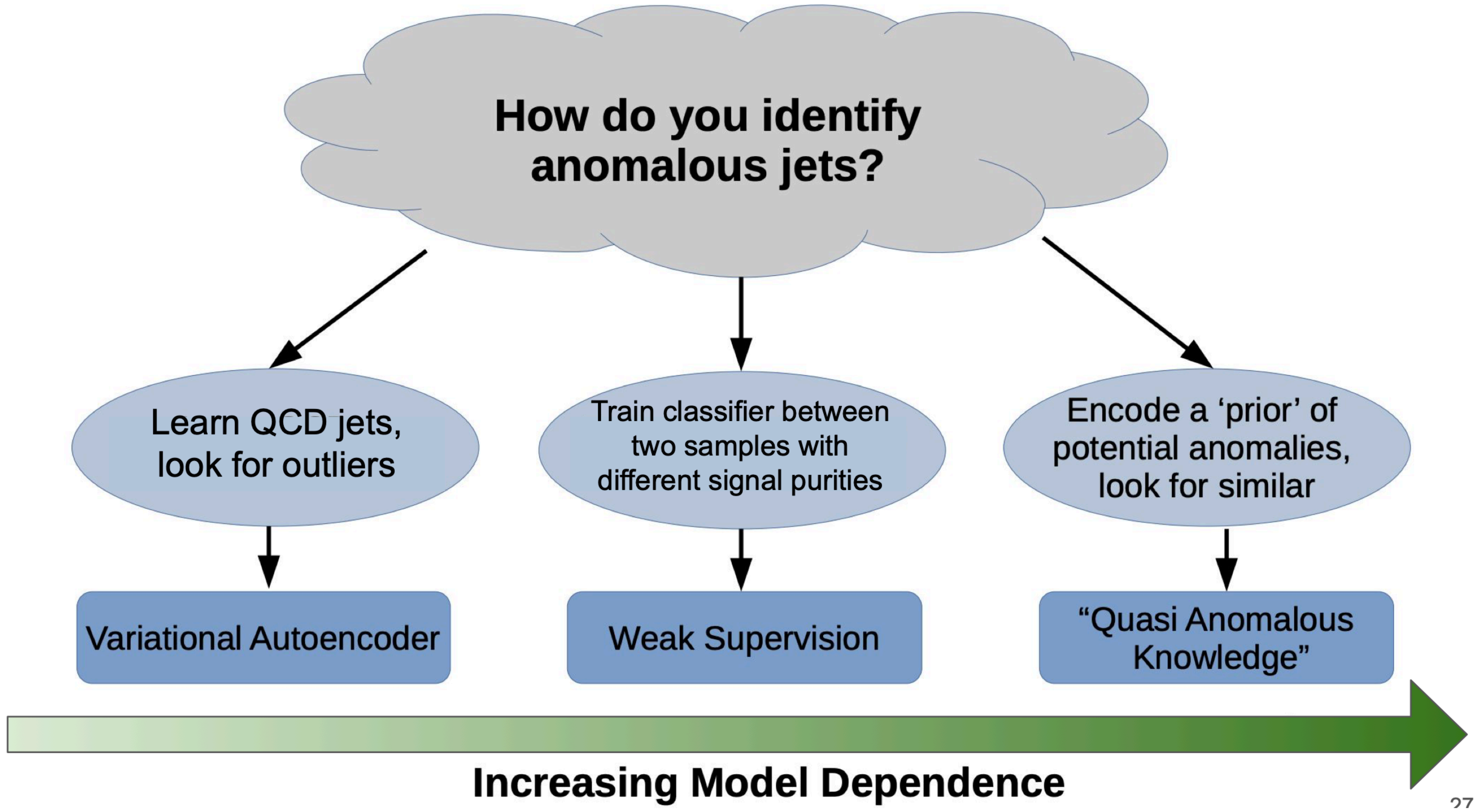


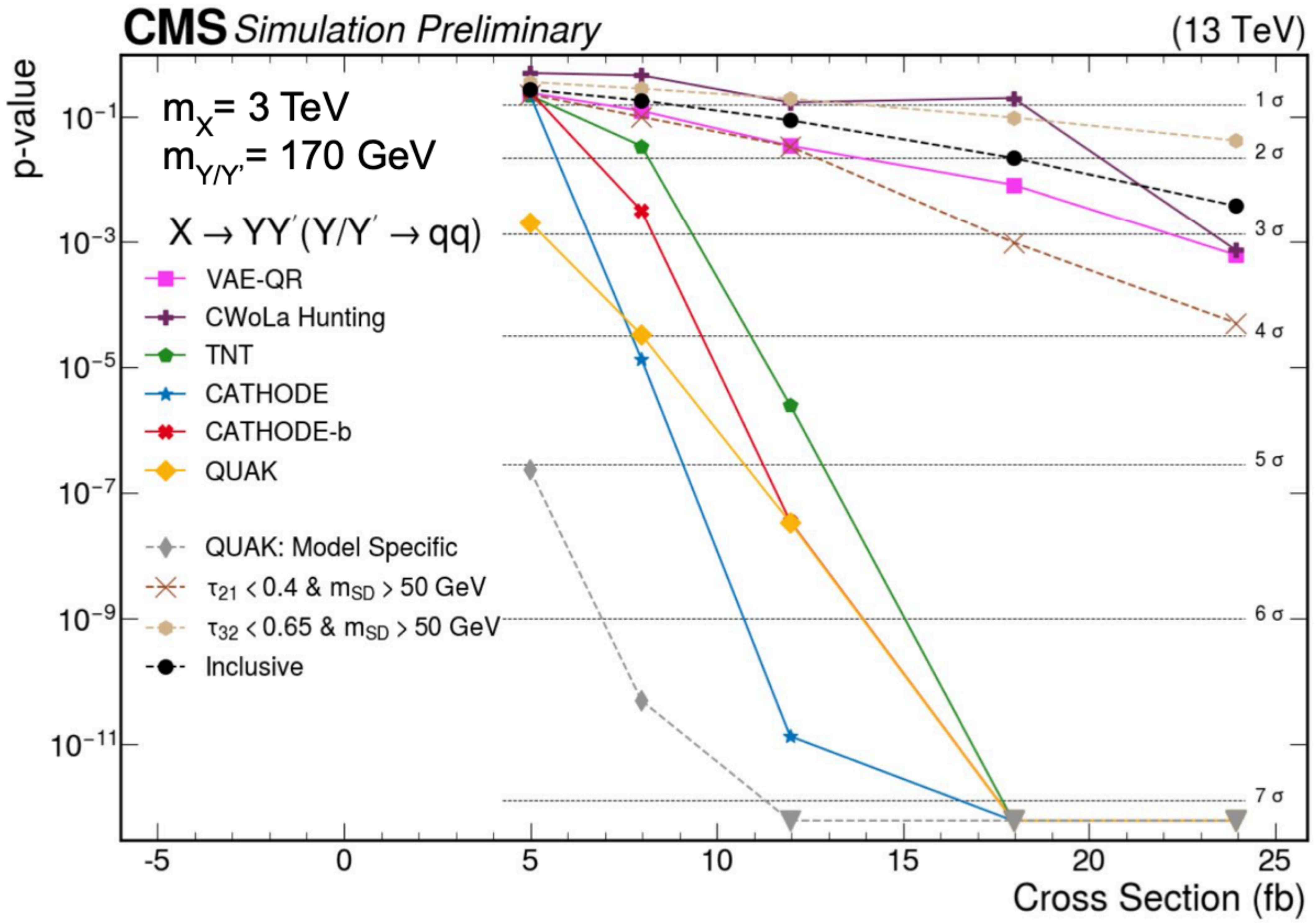


arXiv:1811.10276

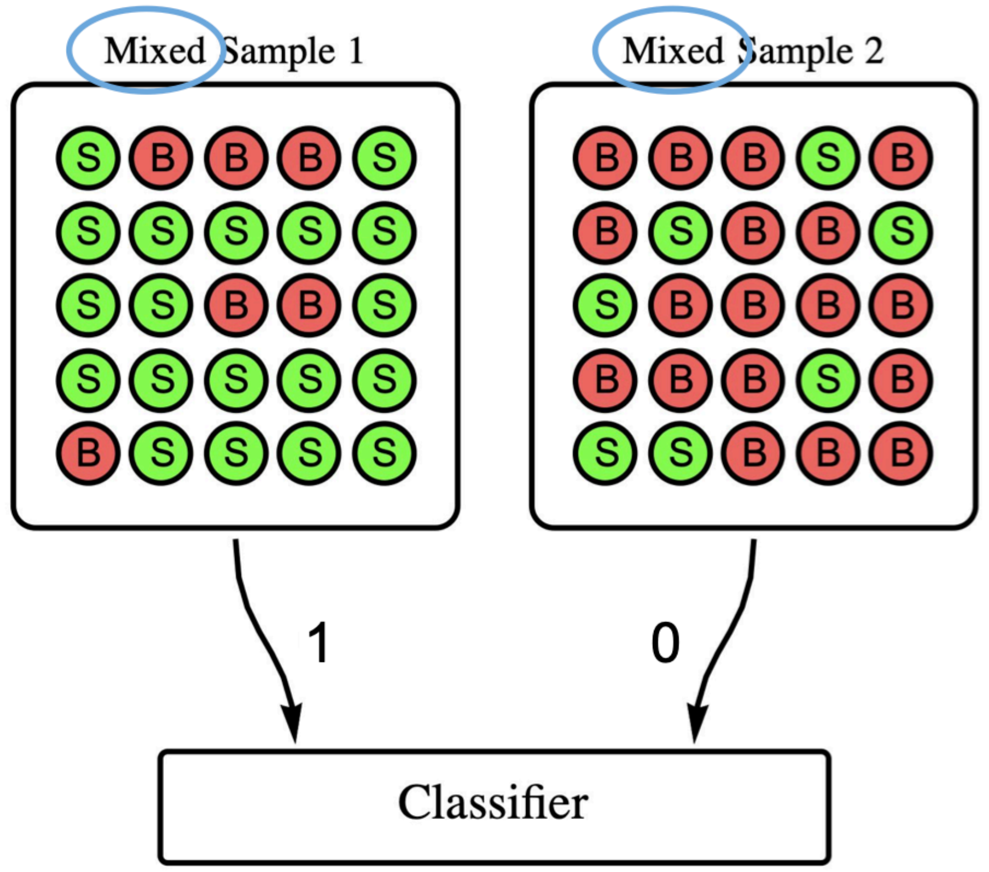
ML for Searches for new Physics

Model-independent searches for new physics using ML techniques





Weak supervision

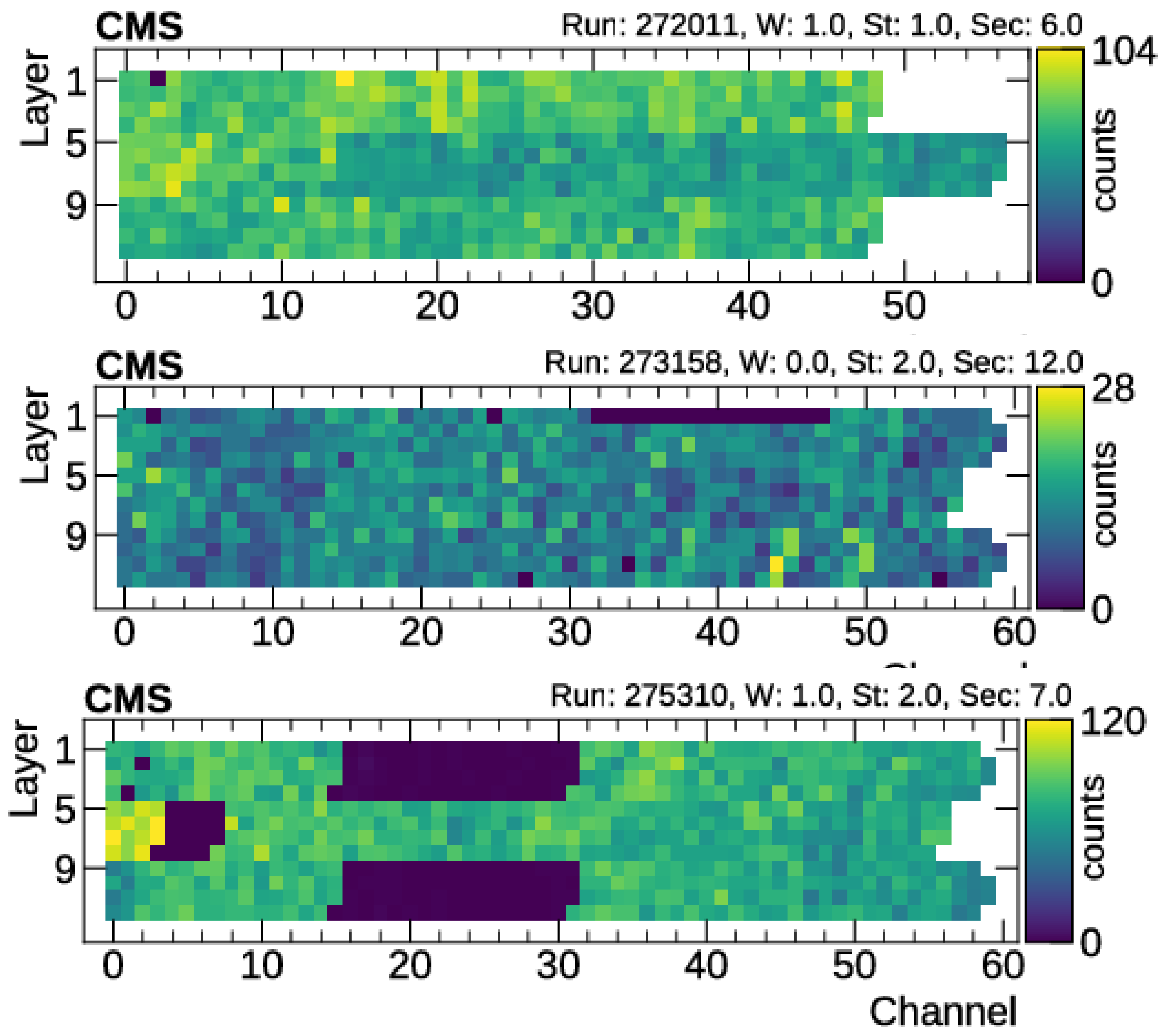


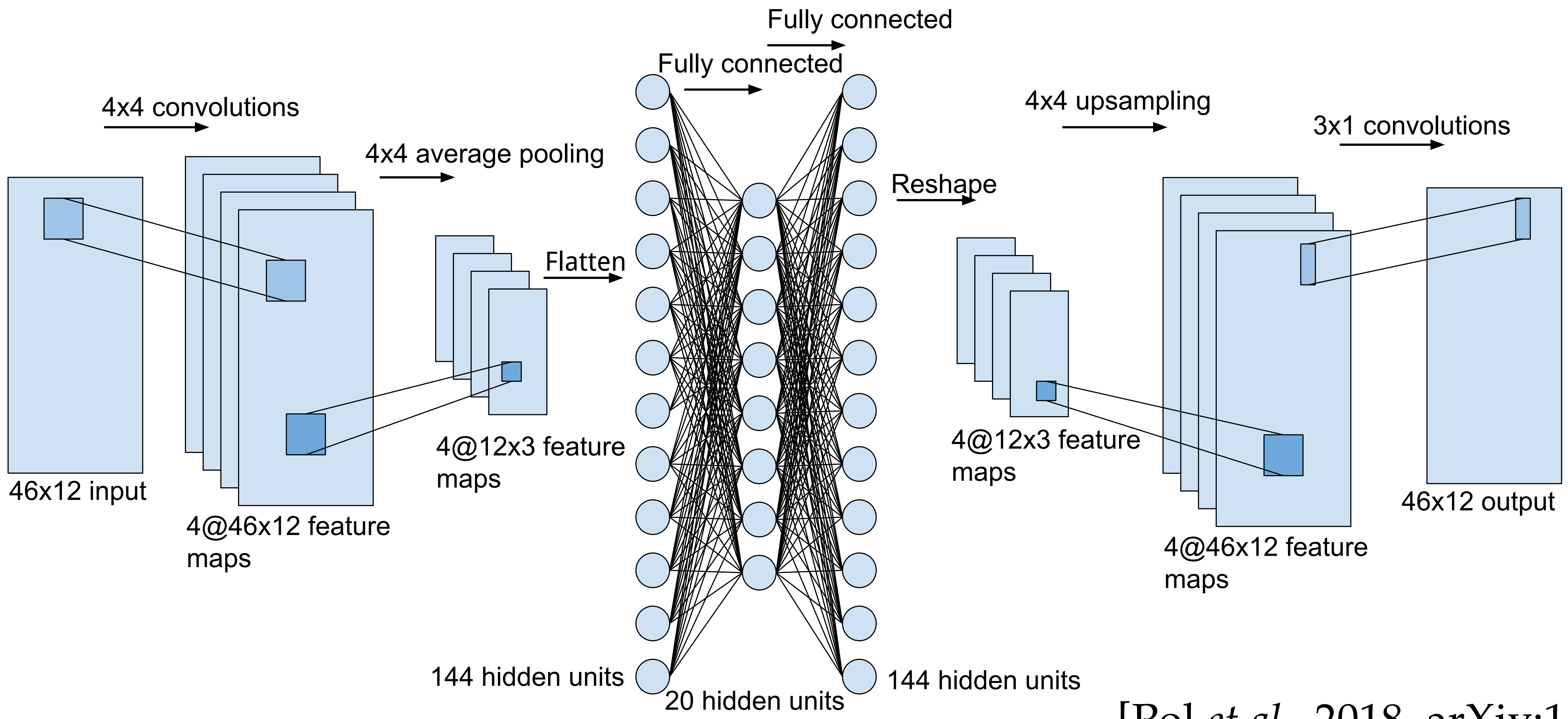
new ML methods can increase discovery potential from 3σ to 7σ

B. Maier: DS seminar 20/3/24

Data Quality Monitoring

- CMS muon detector: unsupervised Method used to spot anomalies
  - using an autoencoder (encoder + decoder network)
  - decoder tries to reproduce input images
  - input different images(anomalies) can then be flagged



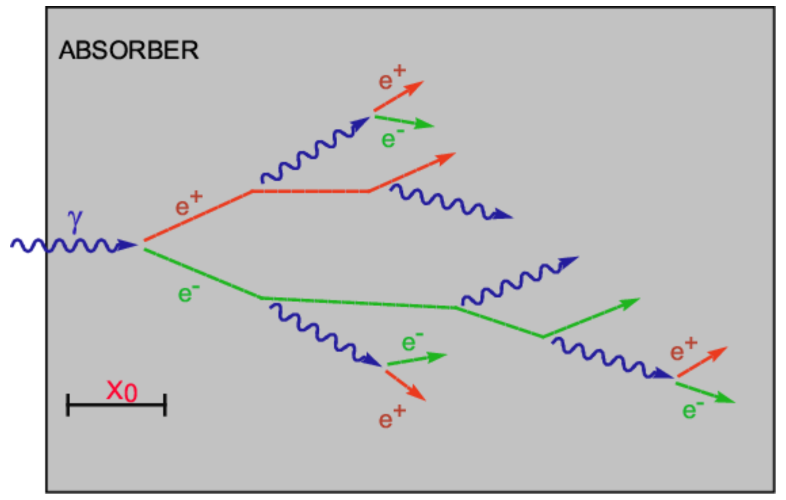


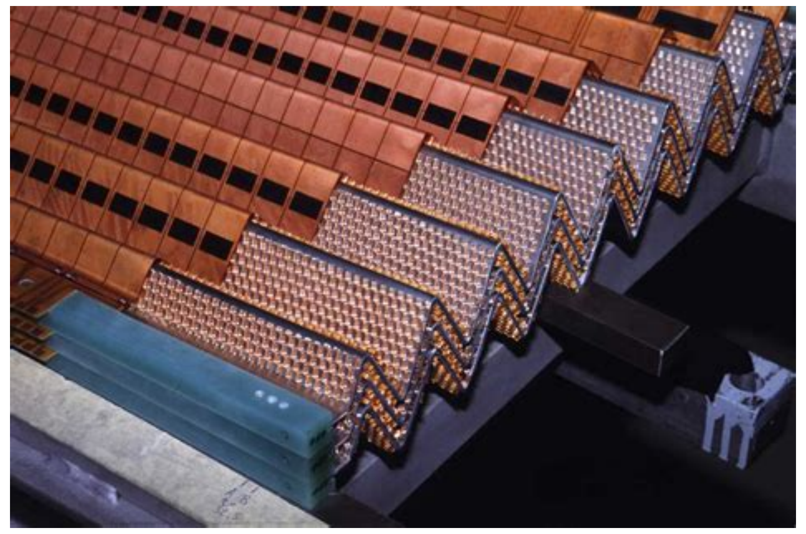
[Pol *et al.*, 2018, arXiv:1808.00911]

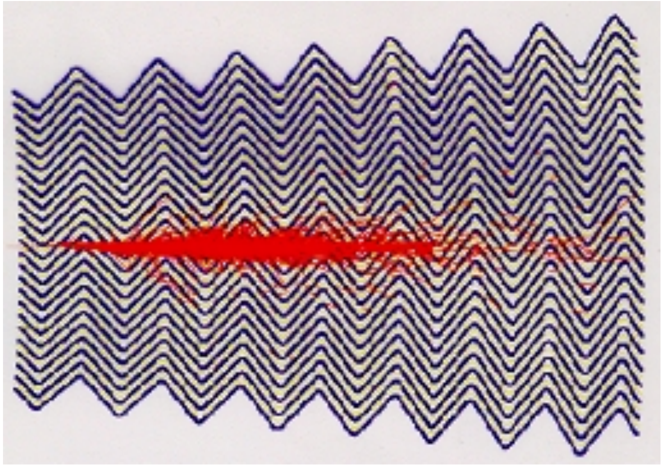
ML for Fast Simulation

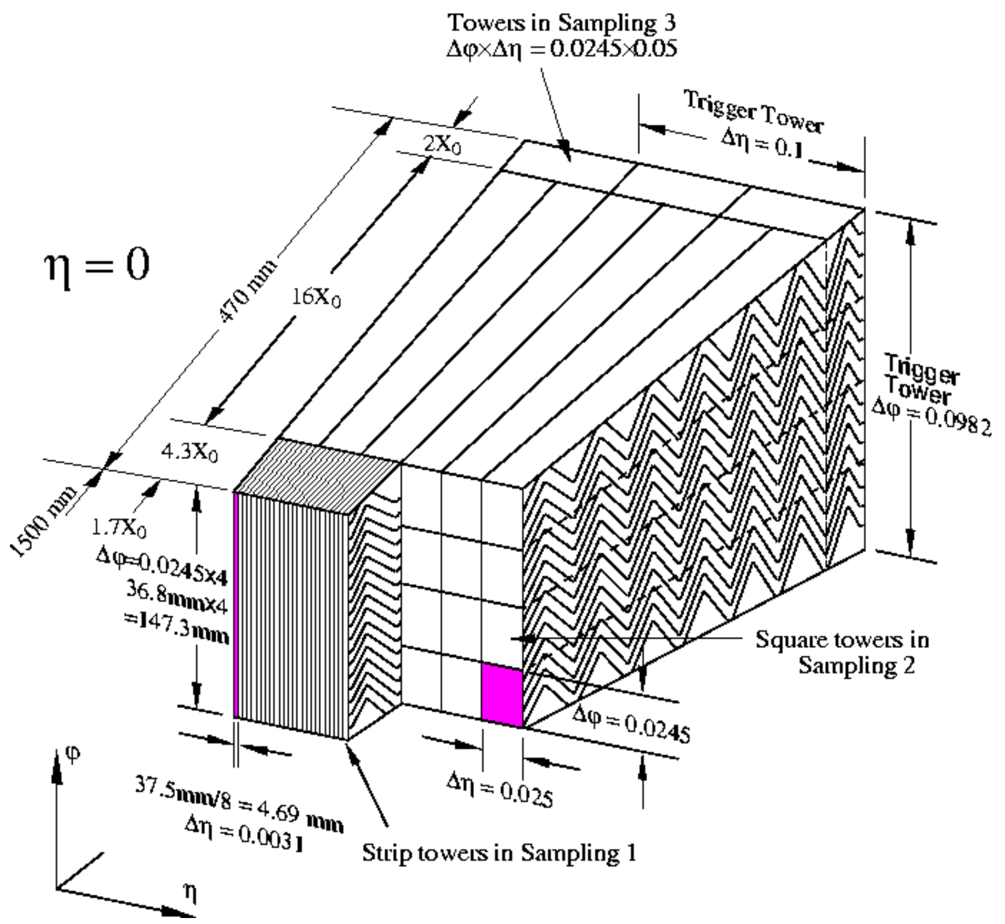
Simulation of electromagnetic or hadronic calorimeters

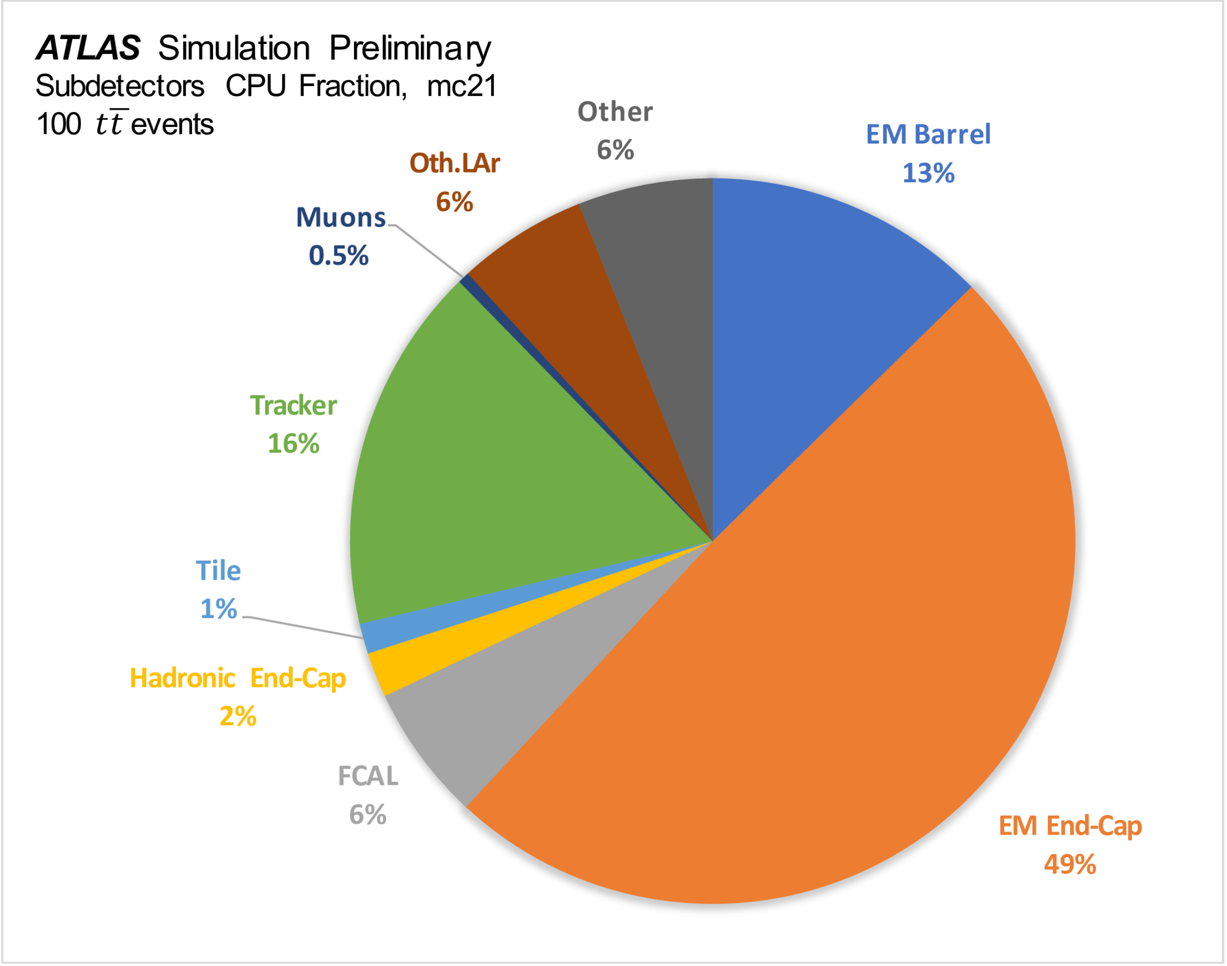
- When a high energy photon or electron (or hadron) hits the calorimeter in a HEP experiment it produces an electromagnetic (hadronic) cascade
- The process can be very complicated to simulate
  (e.g. for ATLAS calorimeter with liquid argon and accordion plates)
  - Very time-consuming to simulate with classical particle transport (Geant4)
    - ~ 70-80% of ATLAS Geant4 simulation time is for calorimeter







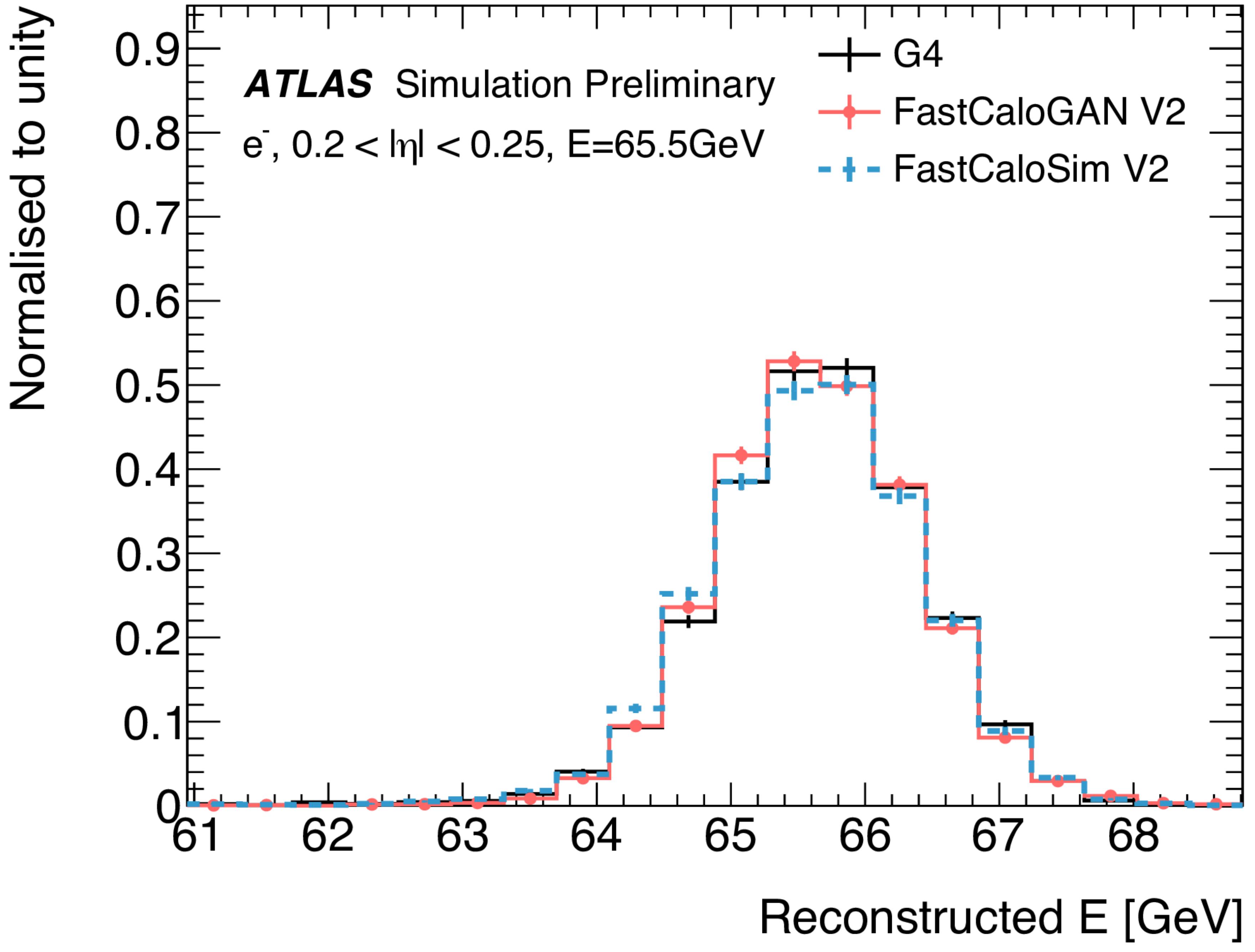




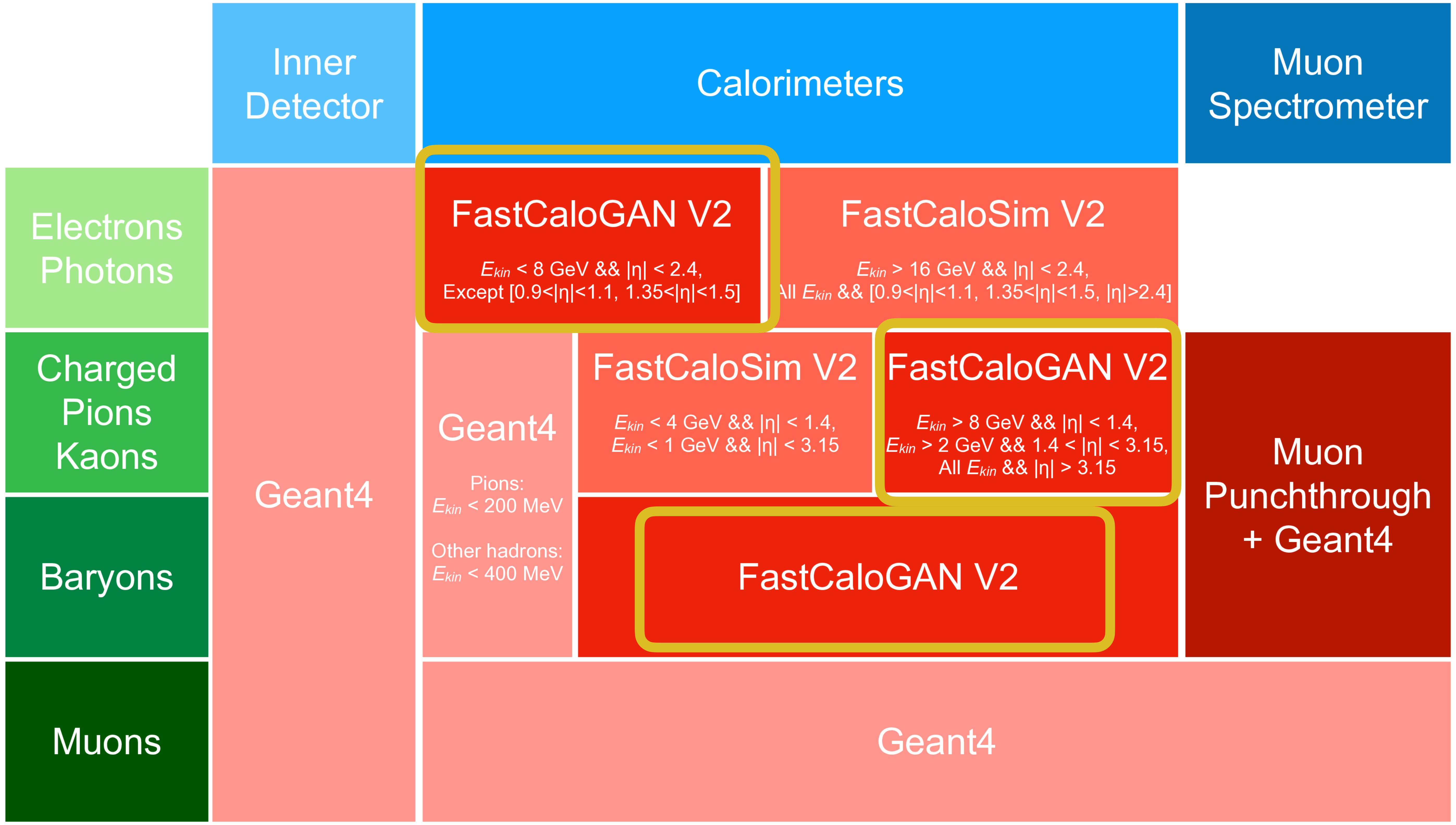
Use ML generative models to simulate calorimeter showers

Fast Simulation in ATLAS

- Generative models used for fast simulations of ATLAS calorimeter
  - GAN (Generative Adversarial Network)



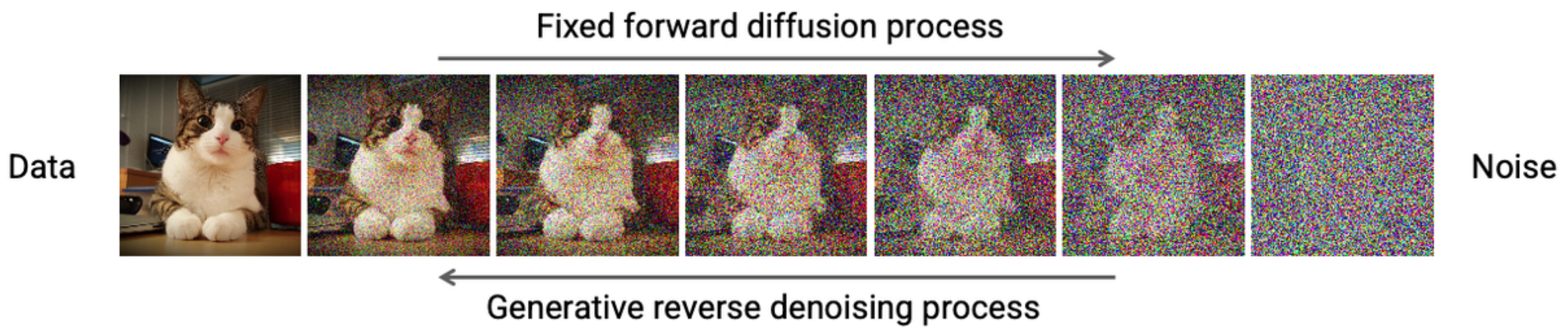
2210.06204

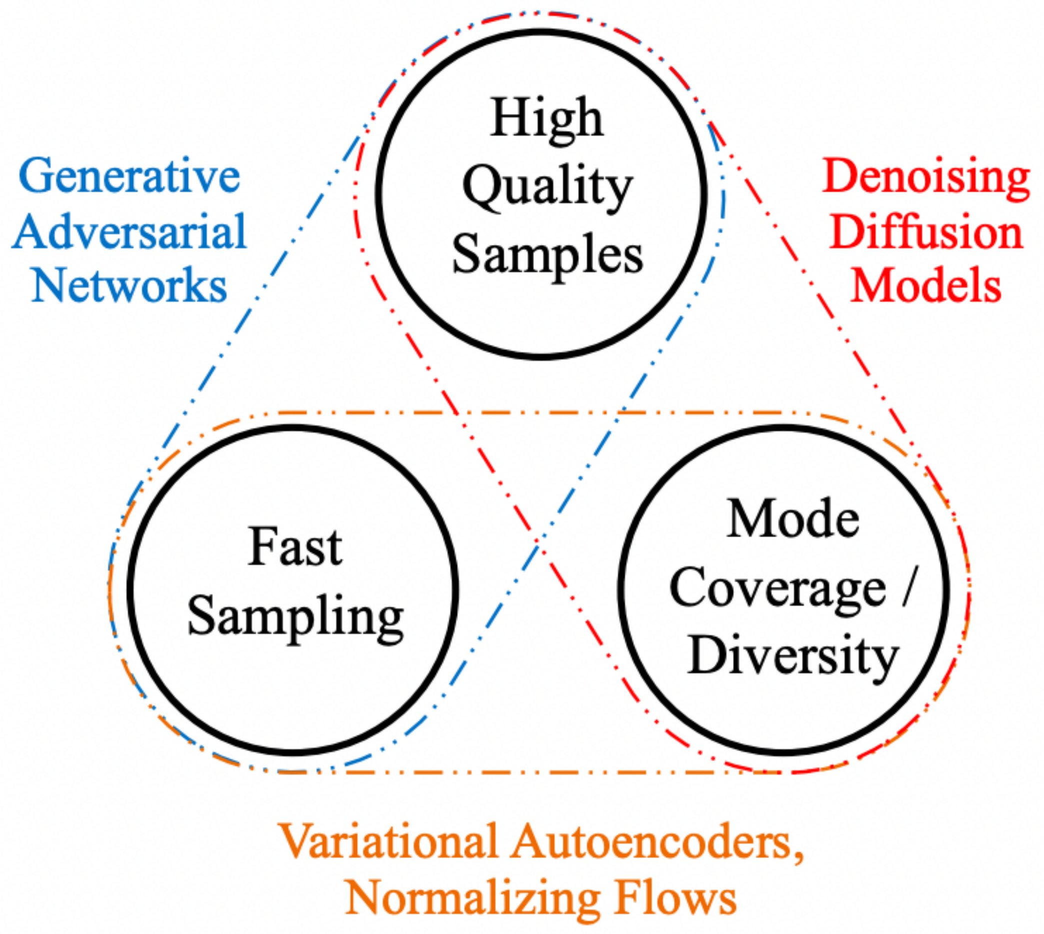


ACAT24

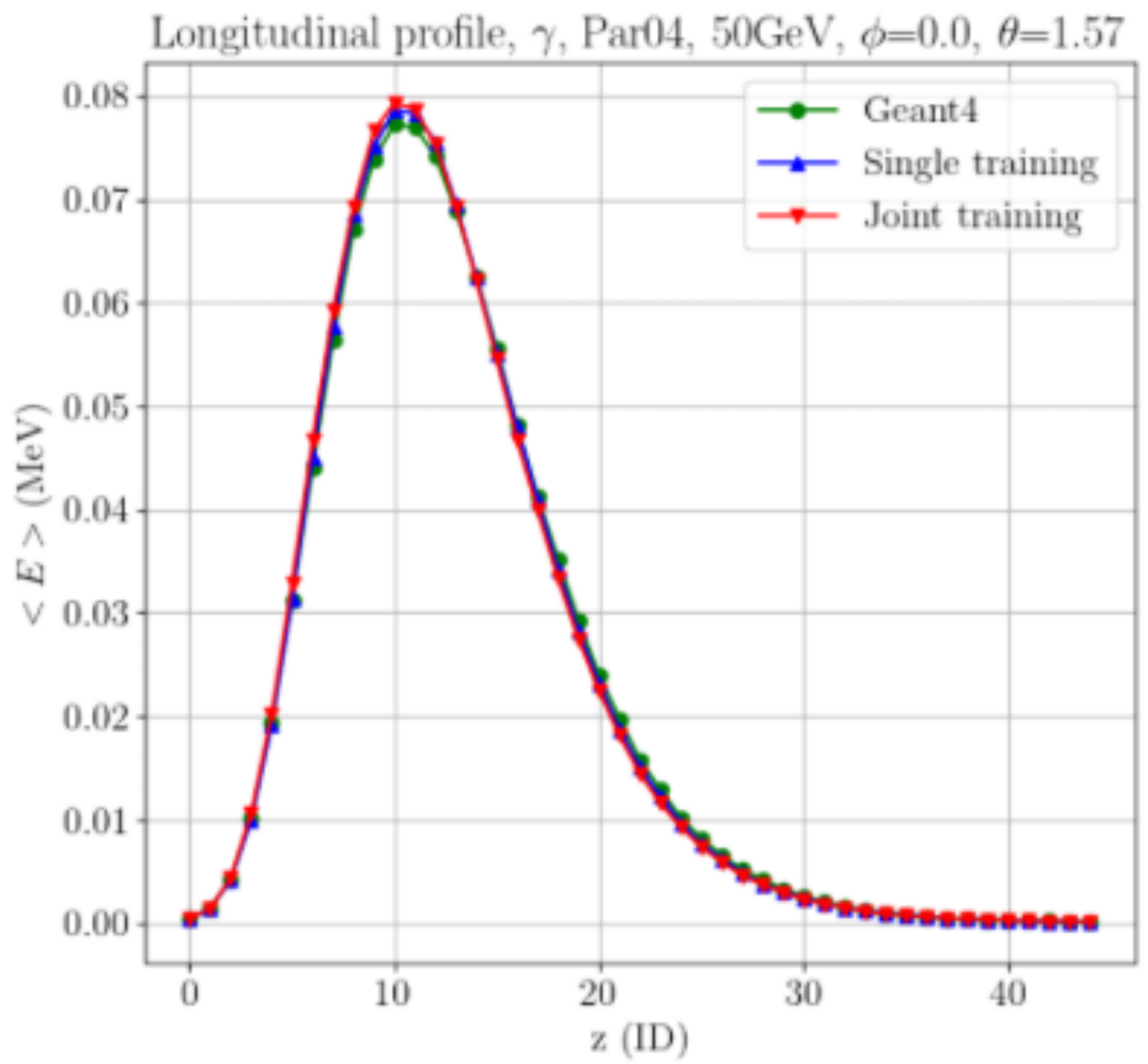
Diffusion Models for Simulation

Use of diffusion models for simulation





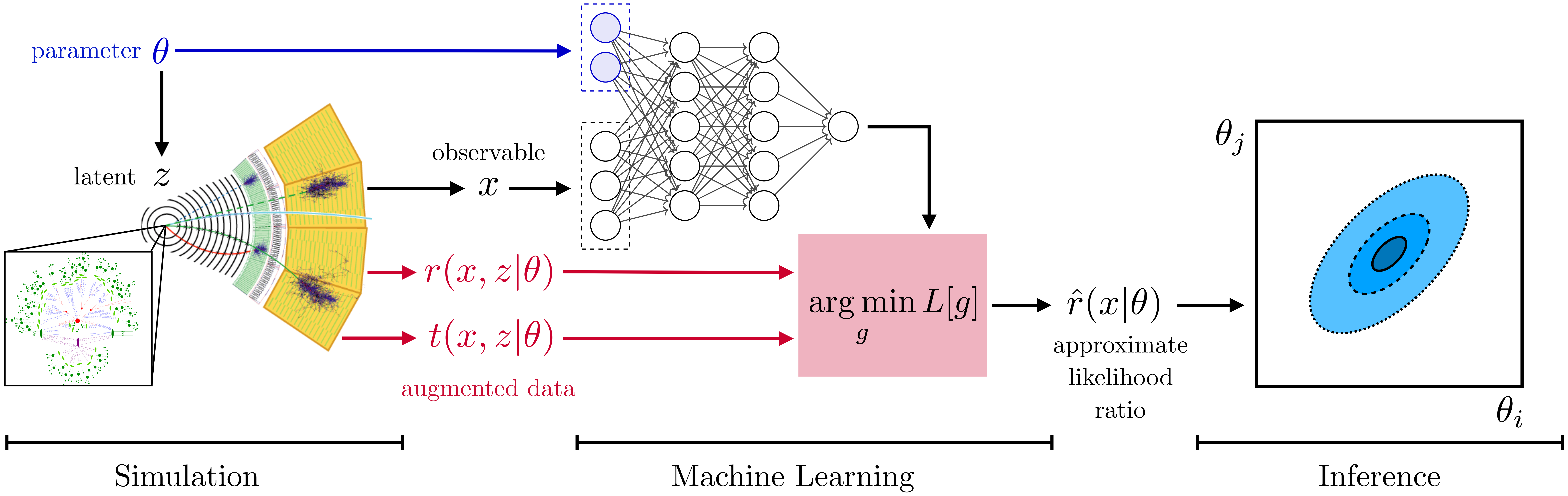
- Investigating transformer based architecture
  - A generalised architecture working with any type of data
  - Modelling long-range dependencies (attention mechanism)



CaloDiT ACAT24

Simulation Based Inference

- Use machine learning for inference
  - estimation of parameters and their confidence intervals
    - approximate likelihood ratio using ML models trained using

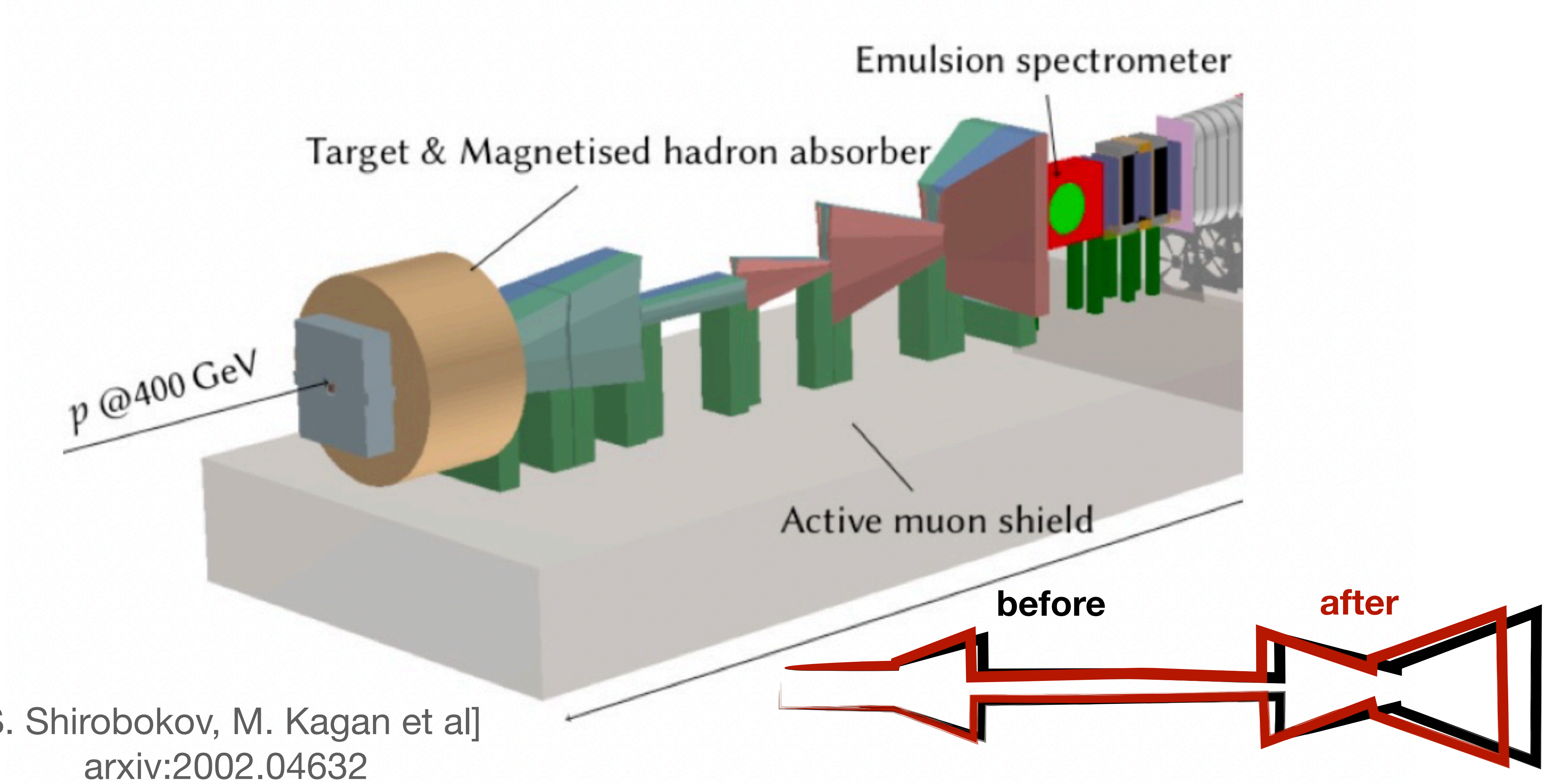


arXiv:1805.00013

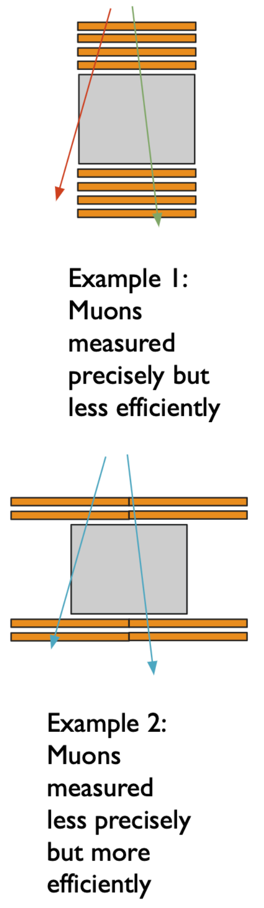
Differential Programming

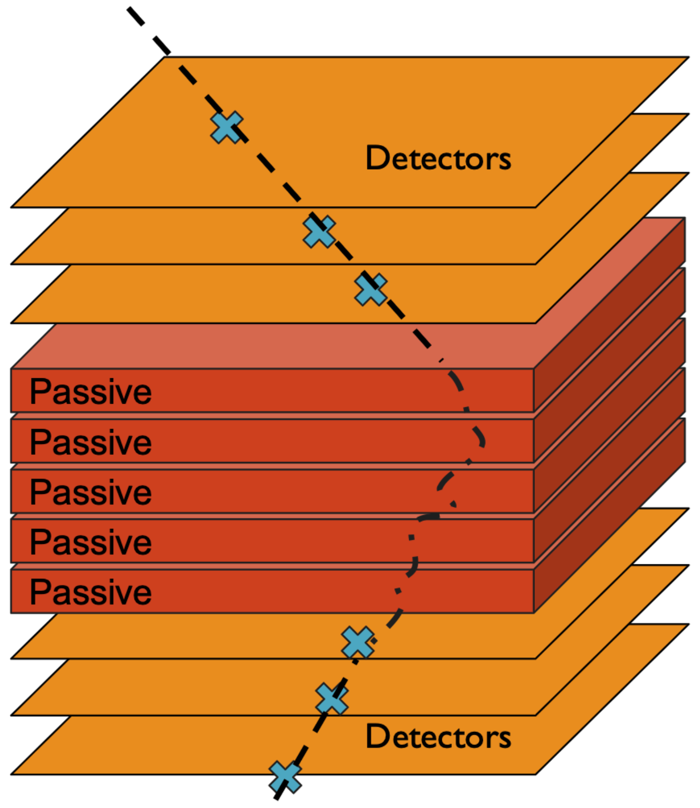
Application: Detector Design

With a differentiable detector simulation or a neural network surrogate, we can optimize the detector design



[S. Shirobokov, M. Kagan et al]
arxiv:2002.04632





G. Strong, T. Dorigo et al
2309.14027

Design | Physics Simulator (or Surrogate) → Possibly ML Reco & Analysis | Performance

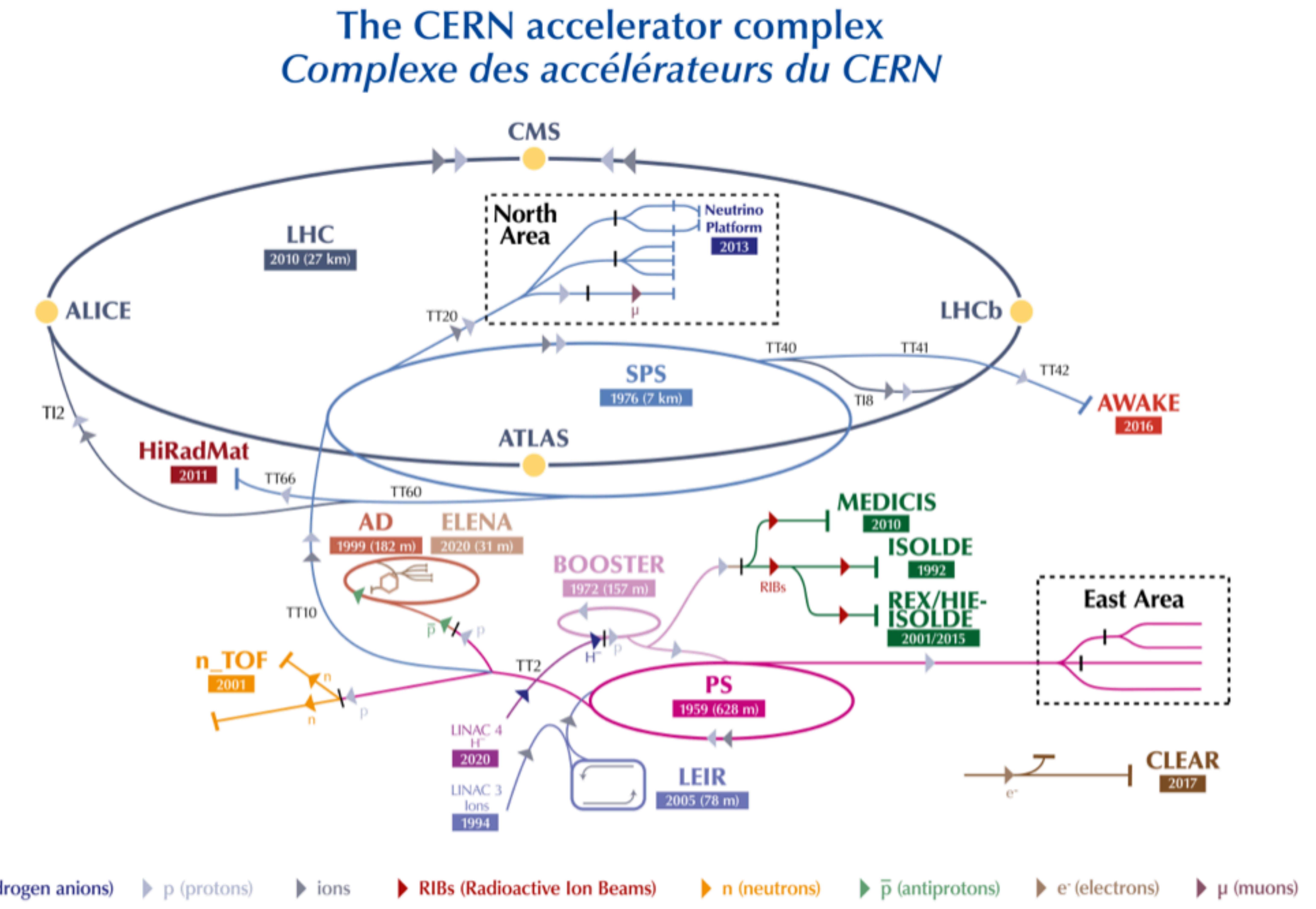
Backpropagation

L. Heinrich (ACAT24)

The CERN accelerator complex

Automating exploitation of CERN accelerators

1. Hysteresis compensation
2. Automatic and dynamic beam scheduling
3. Automatic LHC filling
4. Auto-pilots
5. Automatic fault analysis, recovery and prevention
6. Automatic testing and sequencing
7. Automatic parameter optimisation



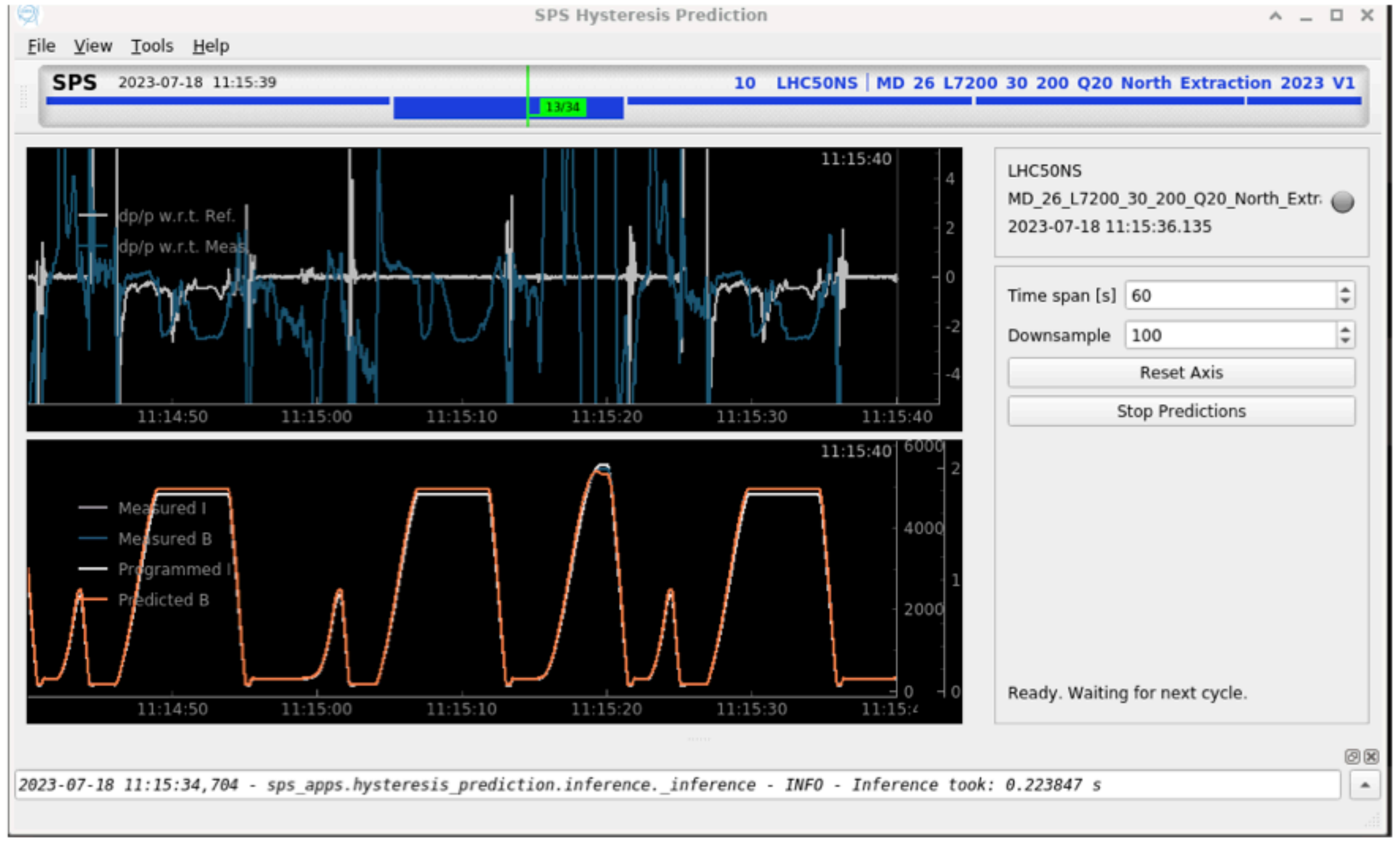
Predicting magnetic hysteresis and eddy current effects

Potentially game-changing!

Time-series forecasting problem: need magnets to be measured on test bench
$$[B_t, B_{t+1}, \ldots, B_{t+n-1}], [I_t, I_{t+1}, \ldots, I_{t+n+N}] \rightarrow [B_{t+n}, B_{t+n+1}, \ldots, B_{t+n+N}]$$

First operational experience:

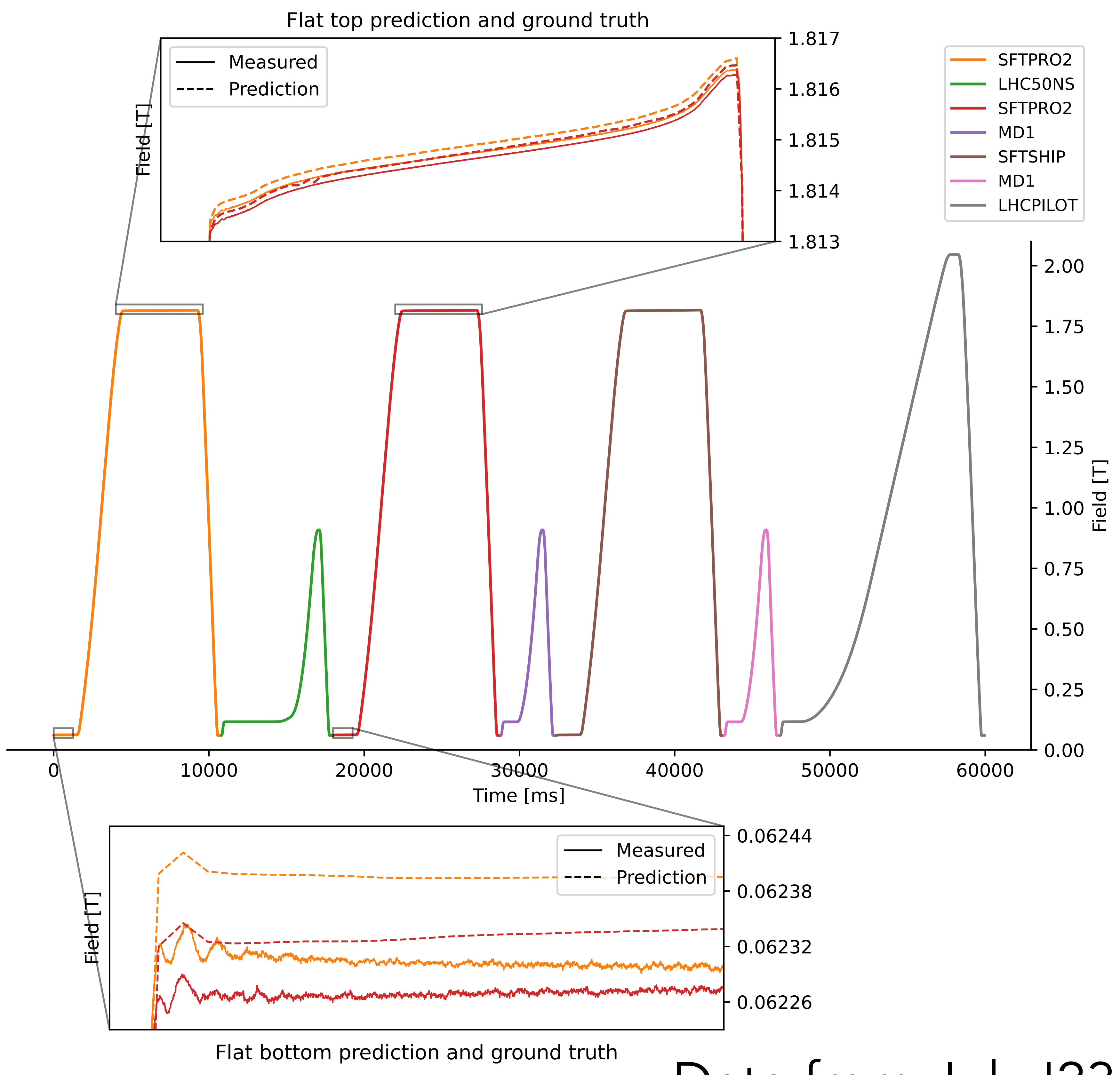
- **feedforward correction** triggered before every cycle
- accuracy not sufficient yet



First results PhyLSTM for SPS main dipoles assuming
$\ddot{B} + g(B, \dot{B}) = \Gamma I(t)$,
next: Transformers

Courtesy A. Lu

SPS main dipole field prediction vs measured, for fixed target cycles



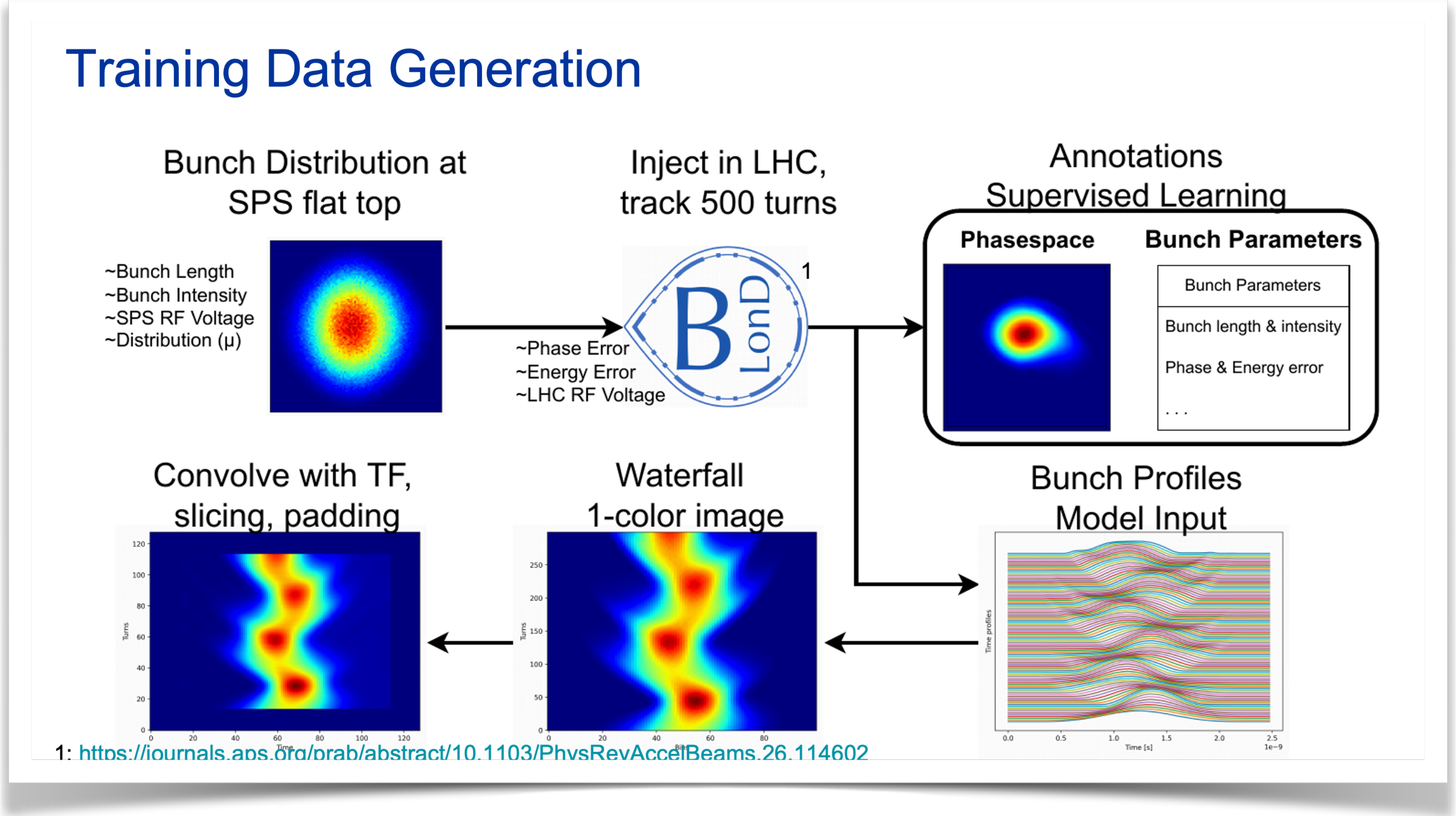
Data from July '23

Enhancing diagnostics - ML tomoscope in the LHC

Speeding up tomographic bunch-by-bunch reconstruction in the LHC: using Auto-encoder ensemble

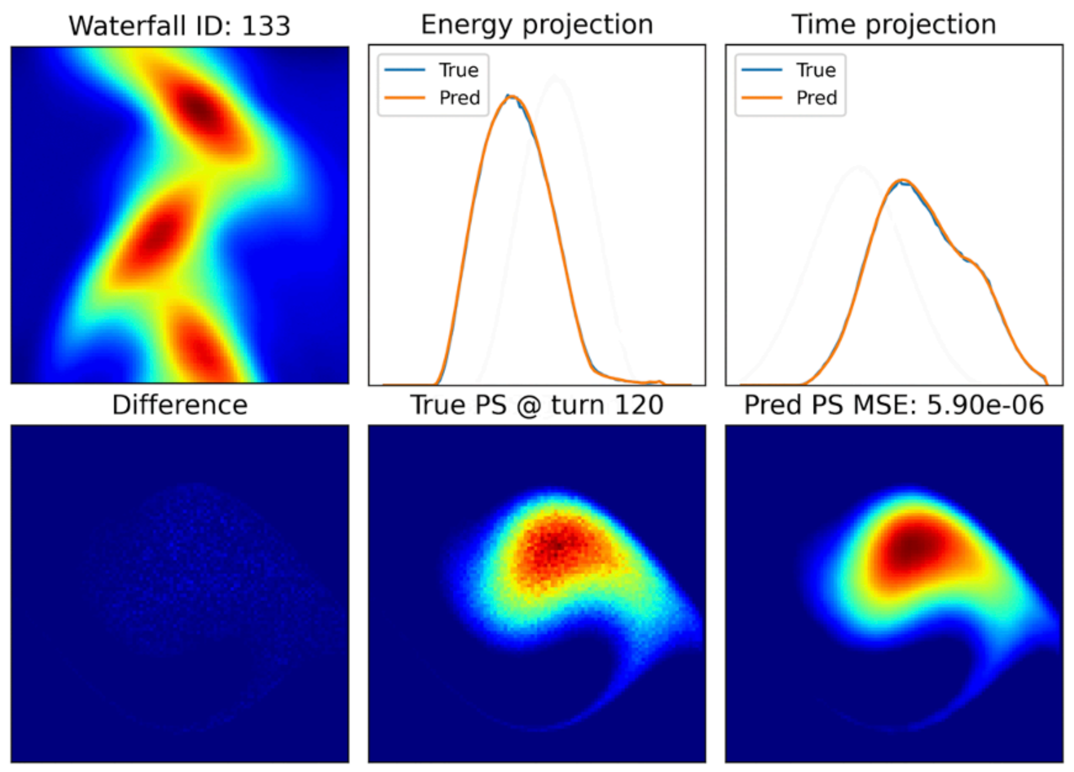
→ fully operational. Used in the LHC control room to measure **injection errors**

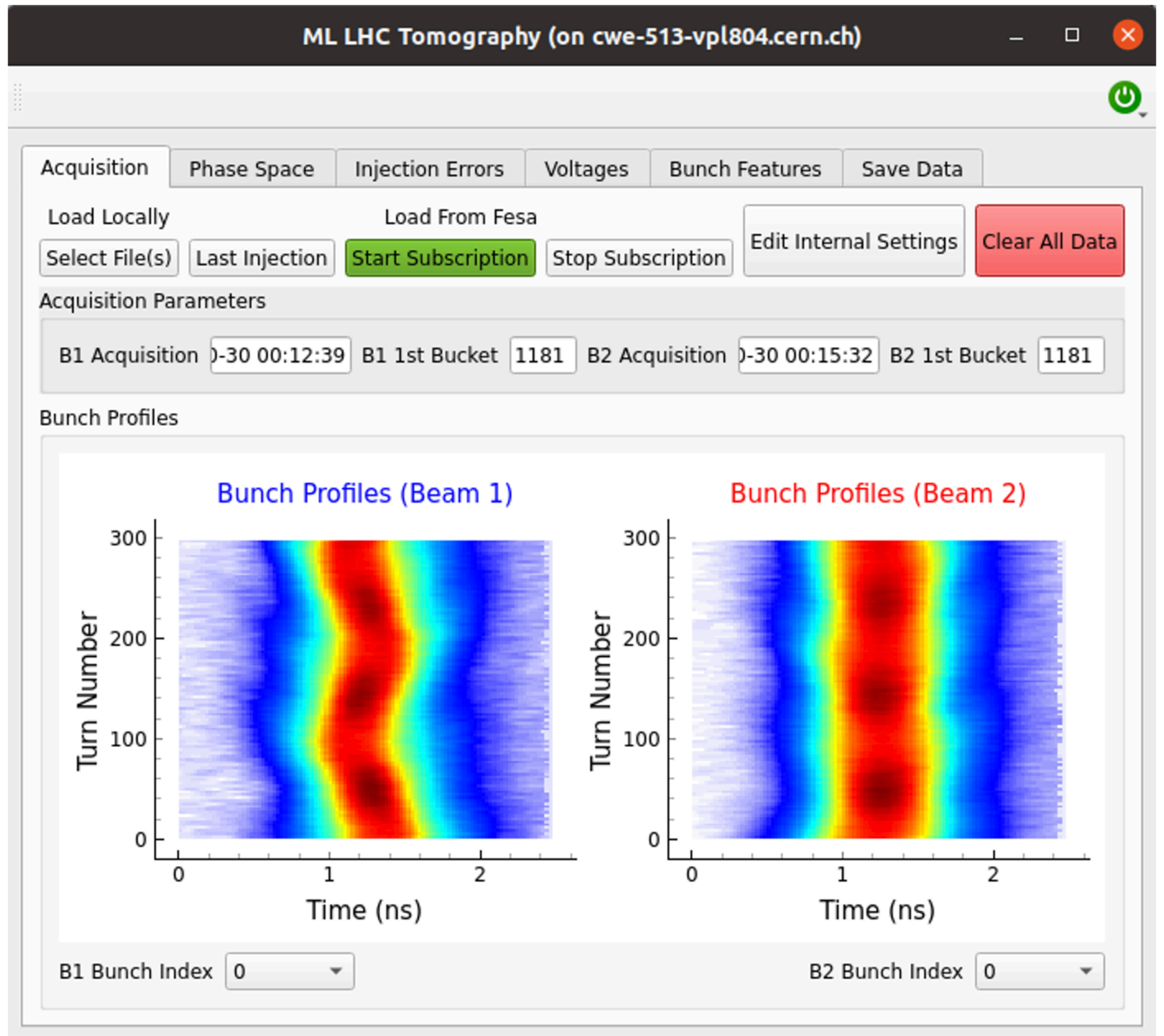
Trained with simulator



Tomoscope Evaluation

- MAE: 0.001 (1‰)
- Visually indistinguishable





Courtesy K. Iliakis, T. Argyropoulos

Applied AI at CERN

- Recent workshop organised by KT (see agenda)

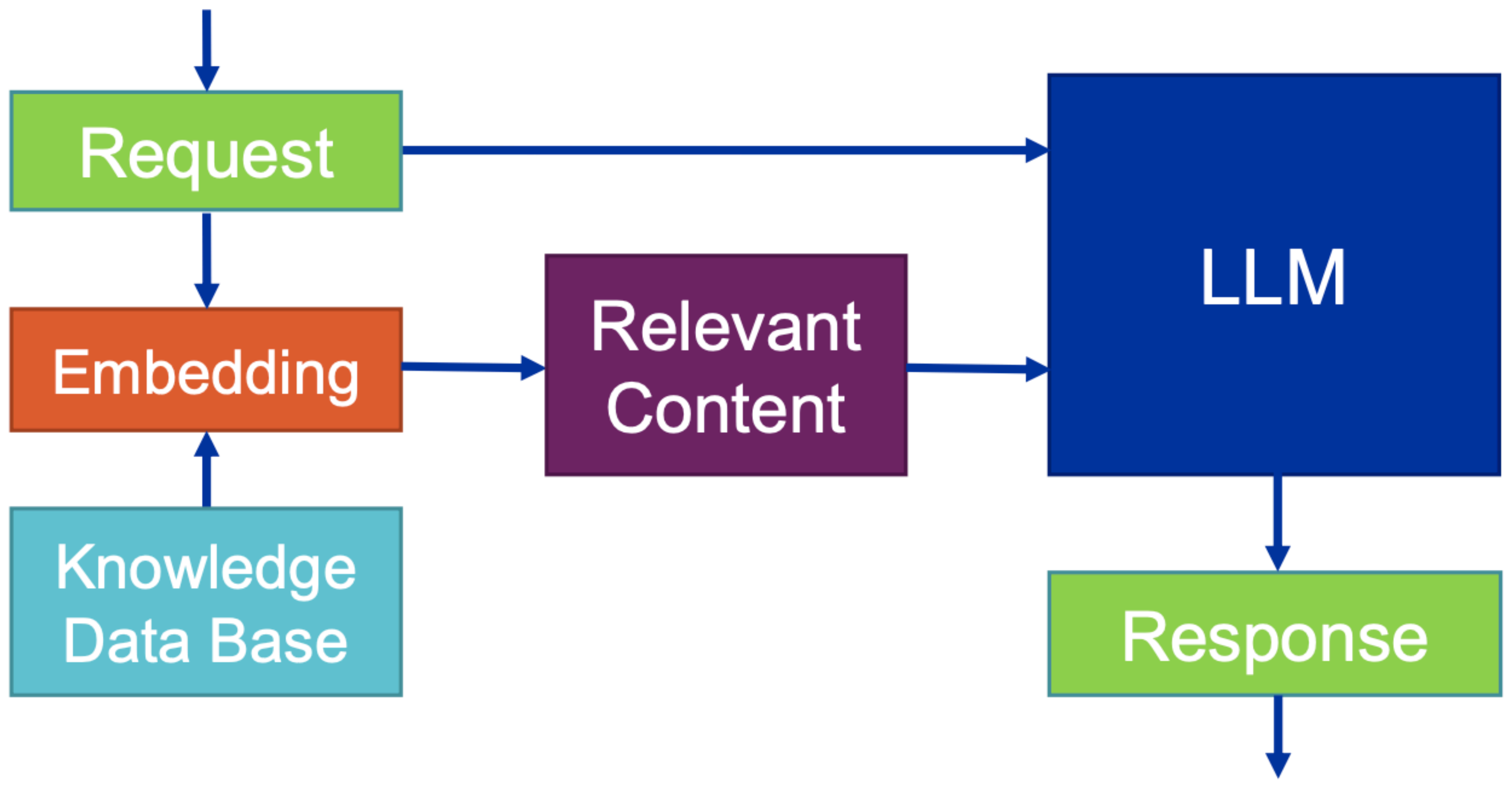
- presented many interesting activities happening all across CERN

Using Large Language Models (LLM)

- AccGPT: A CERN Chatbot
  - aim to be better than ChatGPT for specific CERN use case

The AccGPT pipeline:

- Retrieval Augmented Generation (RAG).



F. Rehm (Applied AI WS)

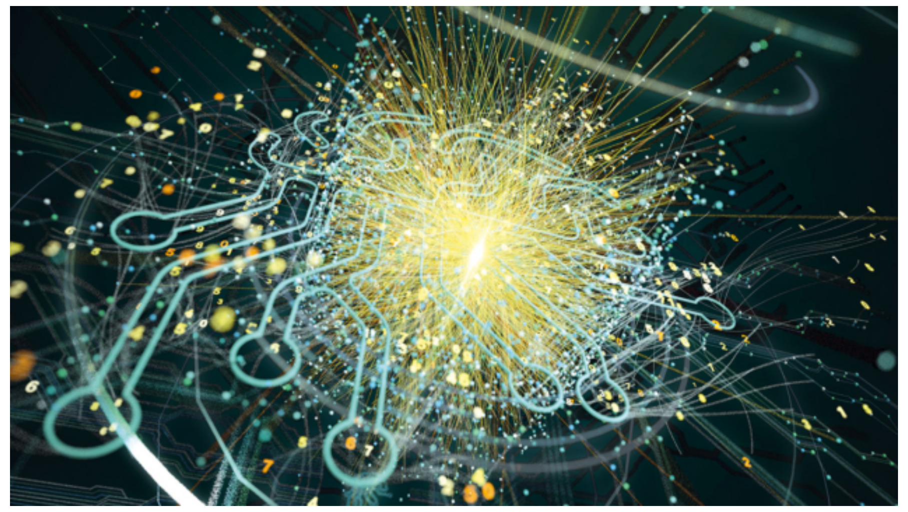
Based on two models:

1. Embedding model:
   - A pretrained open source model (e5-large).
   - Retrieves „relevant content“ from database.
2. Large Language Model (LLM):
   - A pretrained open source GPT model (LLaMA 2 13B).
   - Formulates responses using the „relevant content“.

Accompanied by a self-created knowledge data base.

next IML meeting (April, 9) will be dedicated to LLM

Summary



- **AI/ML fundamental for research at CERN**
  - valuable assistant to maximise the exploitation of the costly collision data
- Following trends of AI community by expanding to using large models and generative AI
  - need to have adequate resources to train and optimise these models
  - focusing also on interpretability, model robustness and uncertainty estimation
- Lots of AI activities are happening at CERN and not only within the LHC experiments
- There is also a lot of dispersion and organising a more **common effort** will certainly be helpful
  - a **hub of AI expertise**, facilitating cross-departmental activities and expertise dissemination
  - strategic for the future of AI at CERN
  - will increase attractiveness to other communities

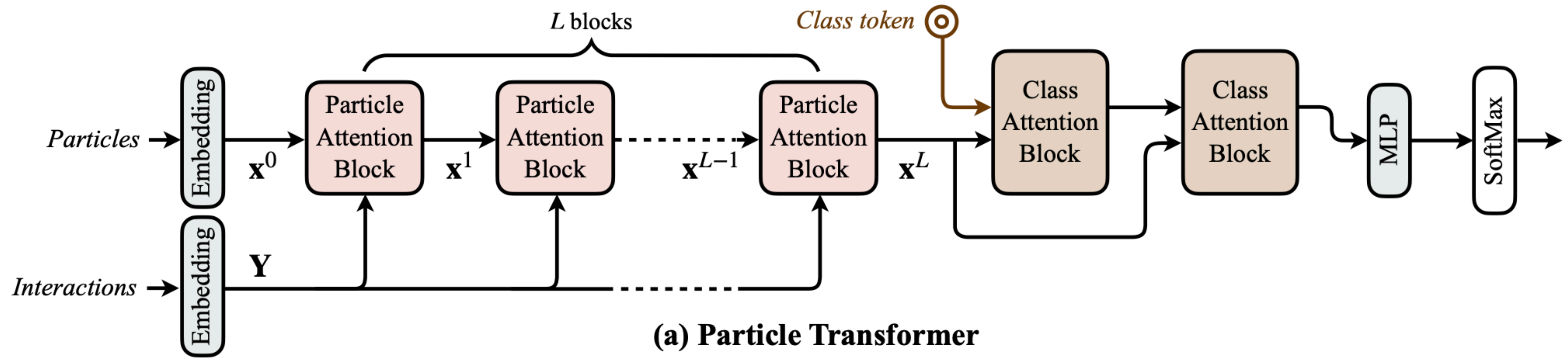
Thank you !

IML

- Inter-experimental Machine Learning Working group (iml.web.cern.ch)
  - forum for the ML community at LHC
  - fostering common ML solution between the experiments
  - organisation of meetings and workshops (see here)

ParT: Particle Transformer

ParT : jet tagging using transformers



(a) Particle Transformer

Table 2. Number of trainable parameters and FLOPs.

	Accuracy	# params	FLOPs
PFN	0.772	86.1 k	4.62 M
P-CNN	0.809	354 k	15.5 M
ParticleNet	0.844	370 k	540 M
ParT	**0.861**	2.14 M	340 M
ParT (plain)	0.849	2.13 M	260 M

	All classes Accuracy	AUC	$H \to b\bar{b}$ Rej$_{50\%}$	$H \to c\bar{c}$ Rej$_{50\%}$	$H \to gg$ Rej$_{50\%}$	$H \to 4q$ Rej$_{50\%}$	$H \to \ell\nu qq'$ Rej$_{99\%}$	$t \to bqq'$ Rej$_{50\%}$	$t \to b\ell\nu$ Rej$_{99.5\%}$	$W \to qq'$ Rej$_{50\%}$	$Z \to q\bar{q}$ Rej$_{50\%}$
PFN	0.772	0.9714	2924	841	75	198	265	797	721	189	159
P-CNN	0.809	0.9789	4890	1276	88	474	947	2907	2304	241	204
ParticleNet	0.844	0.9849	7634	2475	104	954	3339	10526	11173	347	283
ParT	**0.861**	**0.9877**	**10638**	**4149**	**123**	**1864**	**5479**	**32787**	**15873**	**543**	**402**

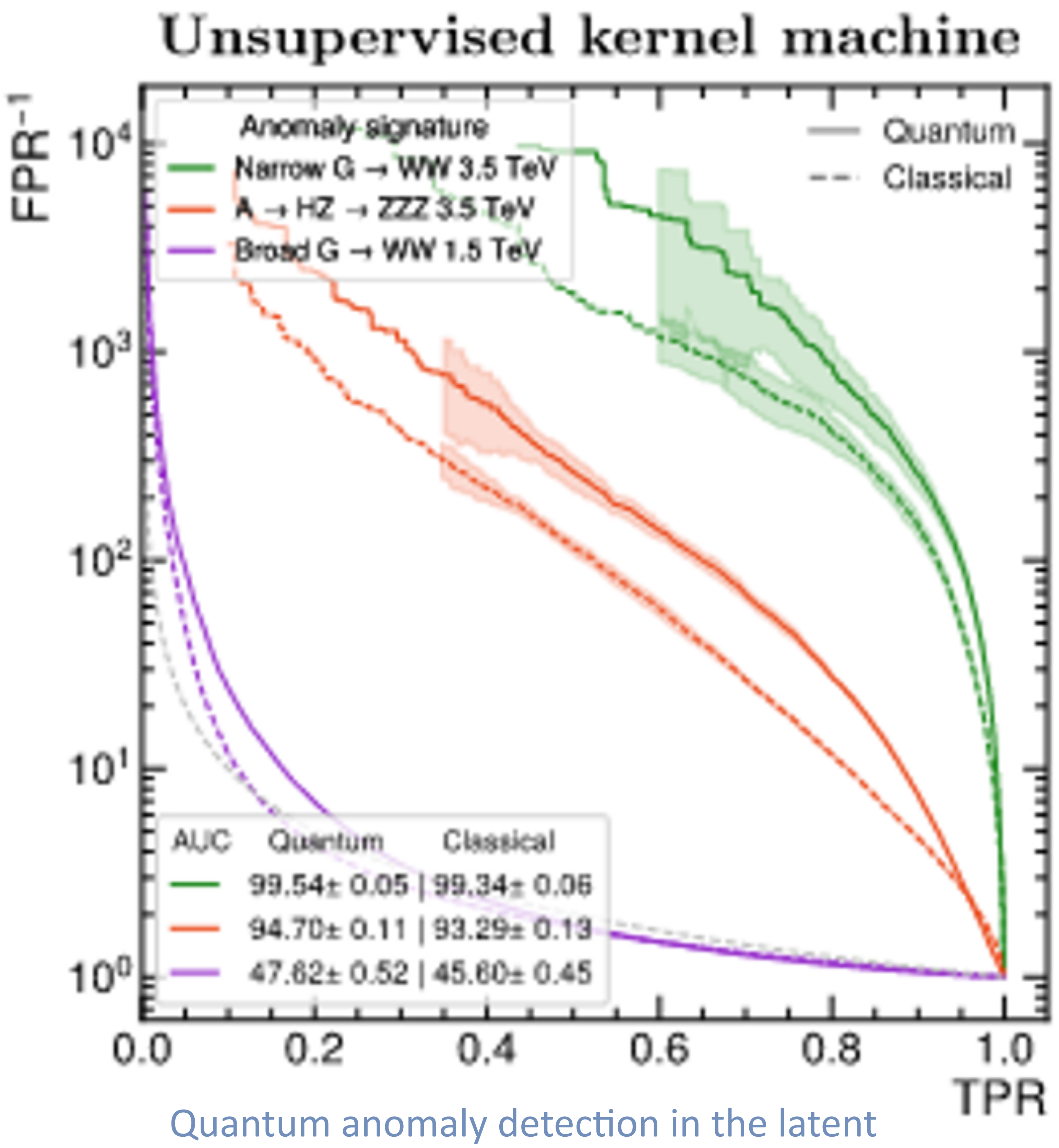
Example $H \to c\bar{c}$:

- doubled background rejection: → ~ 1.4x significance!
- Same significance reach with half data!

arXiv:2202.03772

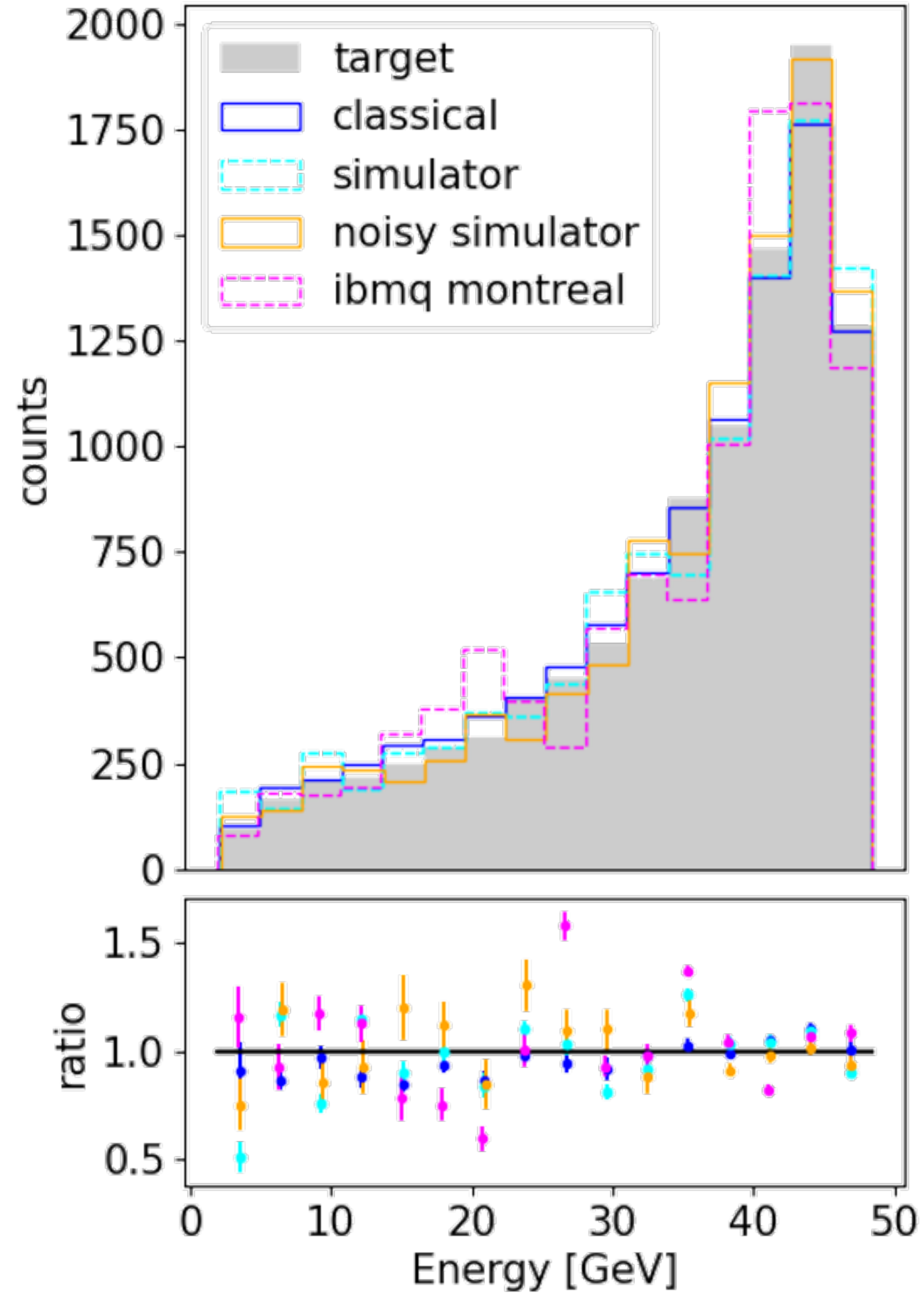
Examples of Quantum ML

- Unsupervised kernel machine for anomaly detection



Quantum anomaly detection in the latent space of proton collision events at the LHC arXiv:2301.10780.

- QCBM for event generation



Phys. Rev. A 106, 022612