Deep geometric learning for particle tracking at CERN

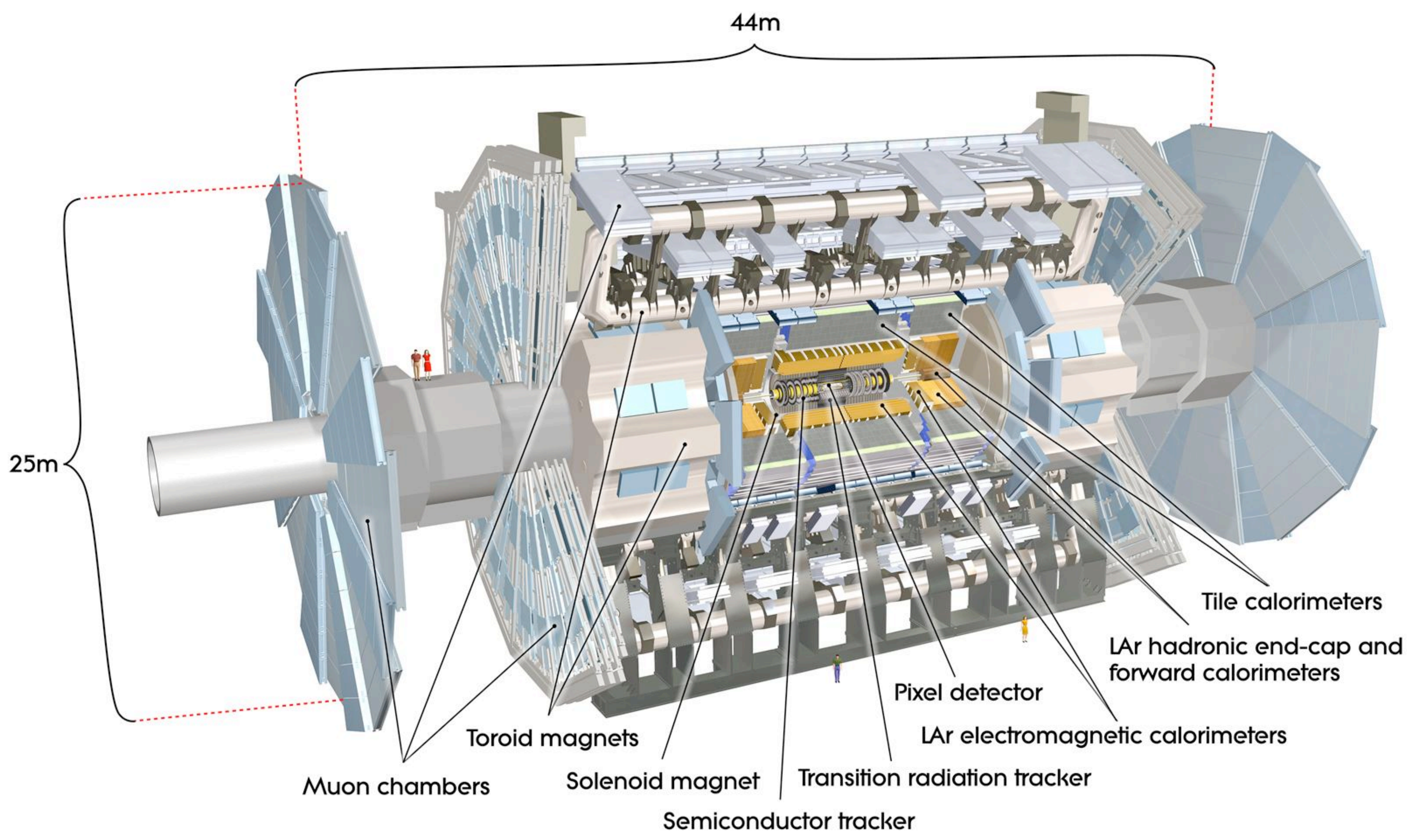
Towards deployment of high performance Graph Neural Network (GNN) - based algorithms for charged particle track reconstruction in ATLAS ITk

Sylvain Caillou

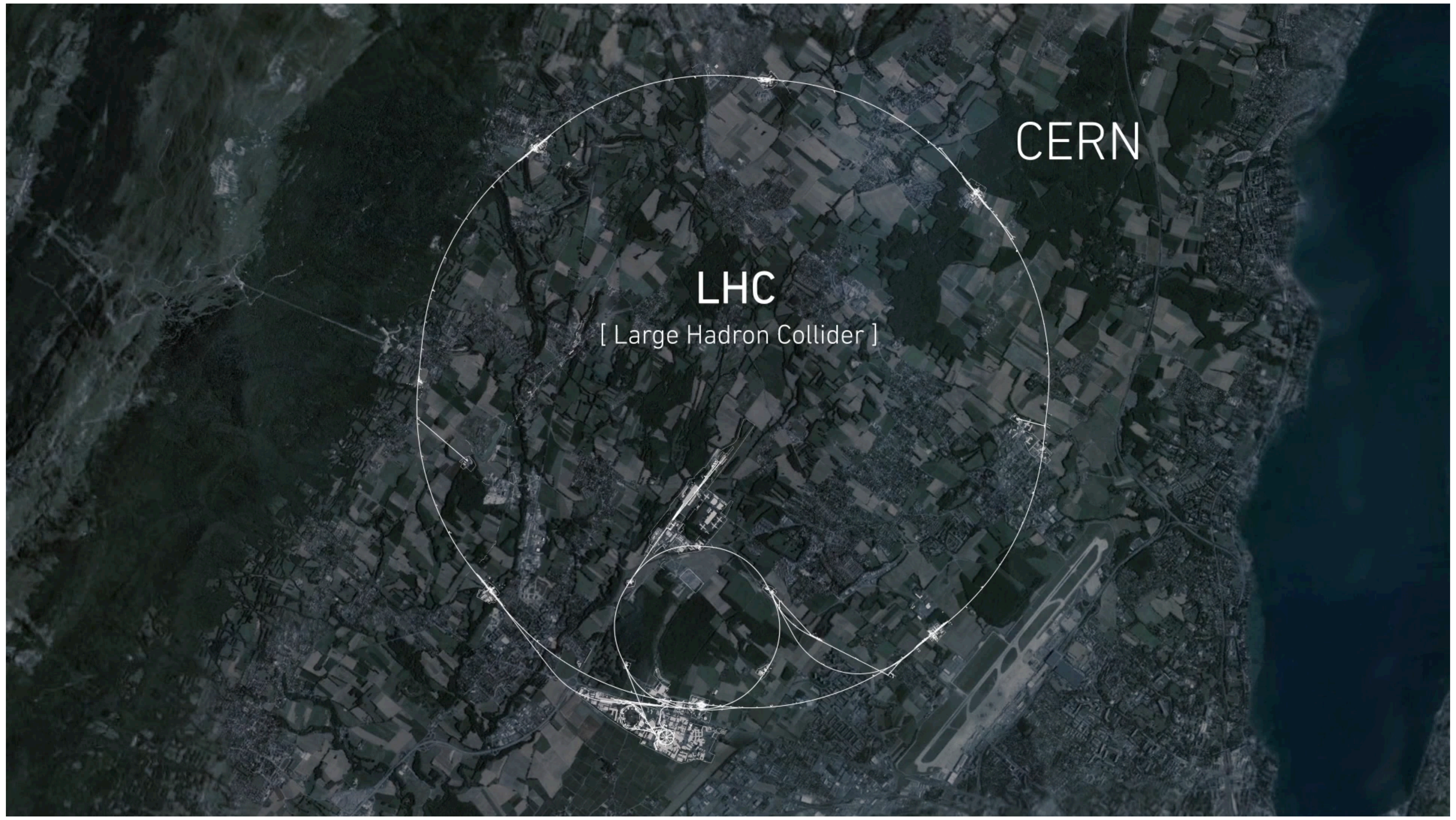
Laboratoire des 2 Infinis-Toulouse

Large Harder Collider (LHC) @ CERN

- Highest energy synchrotron with 27km circumference
- Located at 100m underground between France and Switzerland
- Proton-proton collisions at high energy 13-14TeV
- Two generalist detectors ATLAS and CMS
- Design to find the Higgs Boson!



The ATLAS detector

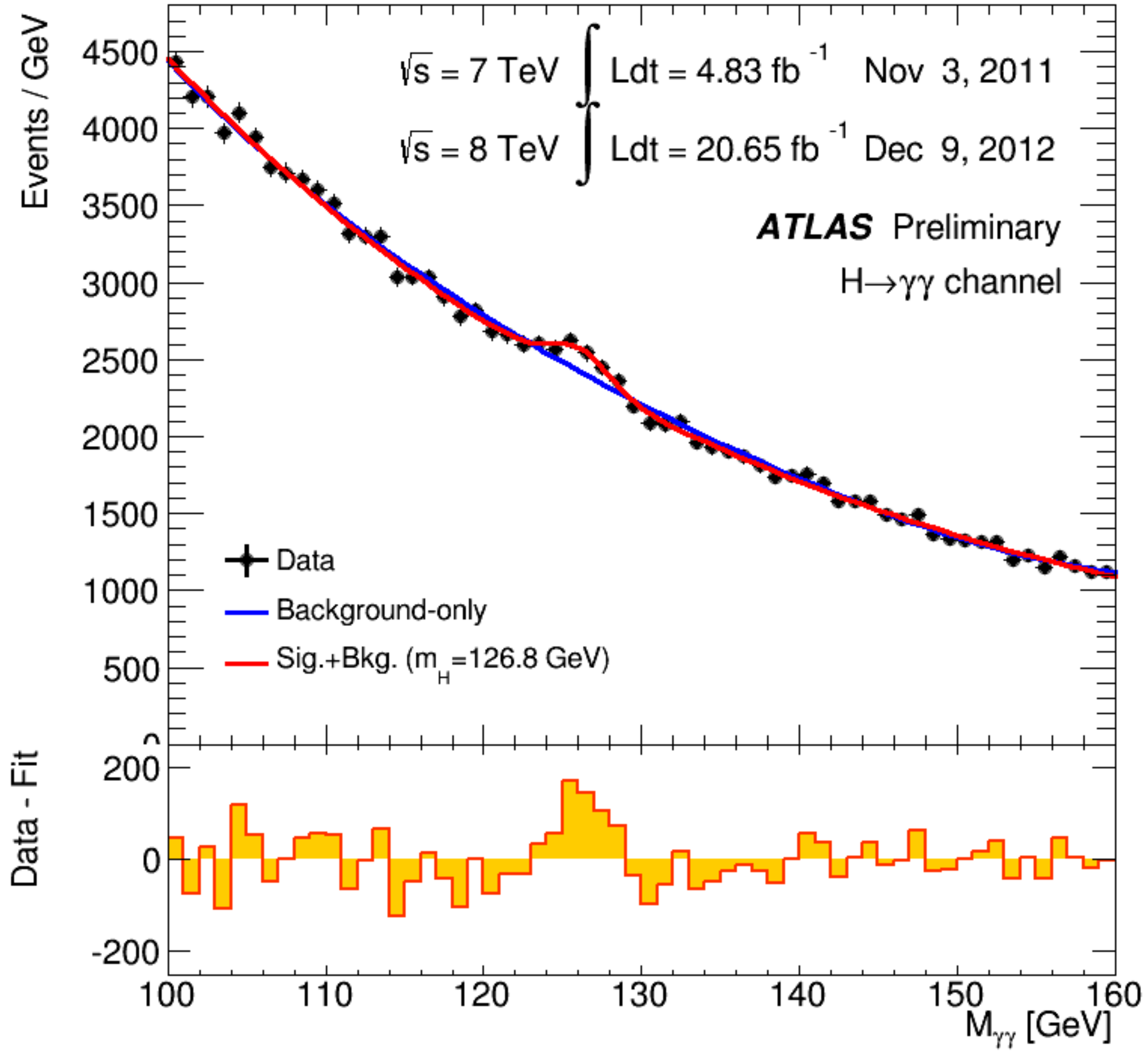


1 collision of 2 bunches of protons every 25ns !

Higgs discovery

4 July 2012

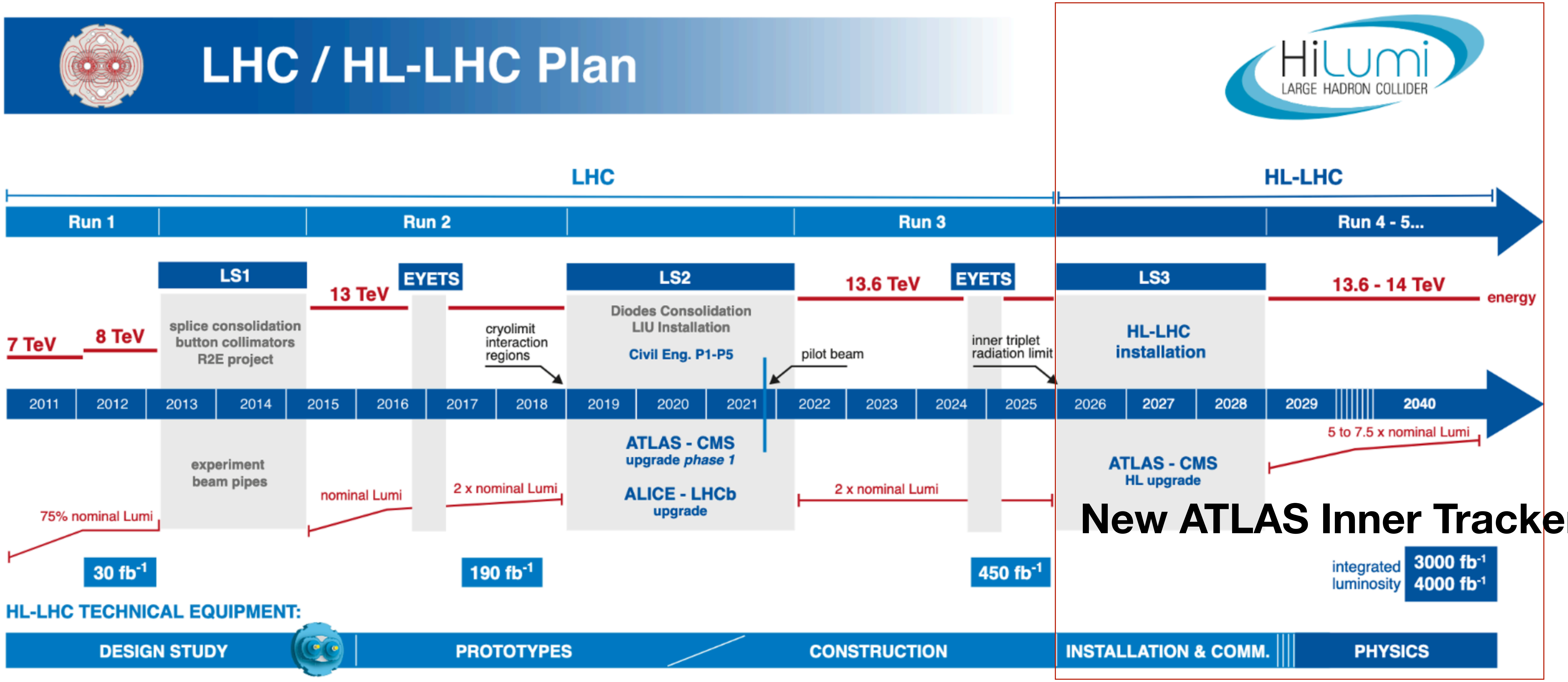
ATLAS and CMS experiments at CERN announced that they had **observed a new particle** in the mass region of around 125 GeV: a particle consistent with the Higgs boson.



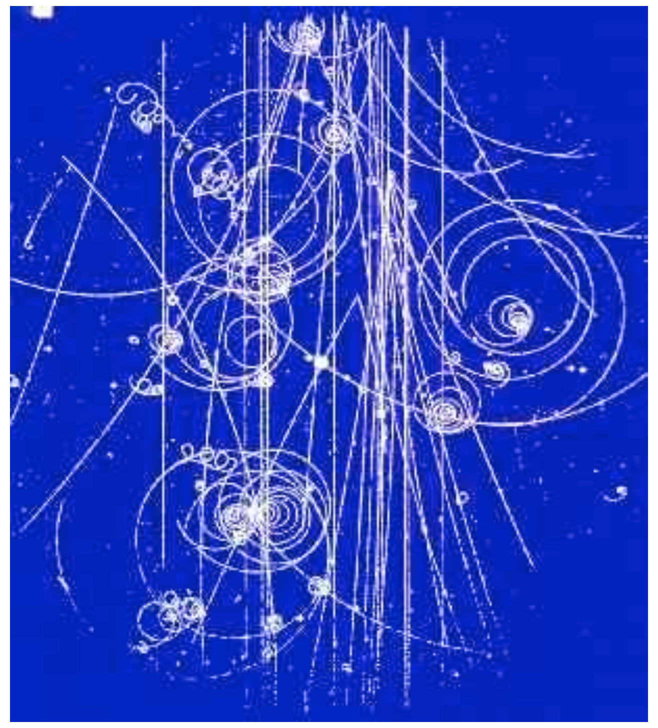
François Englert and Peter Higgs "*for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider*".

Towards HiLumi-LHC (HL-LHC)

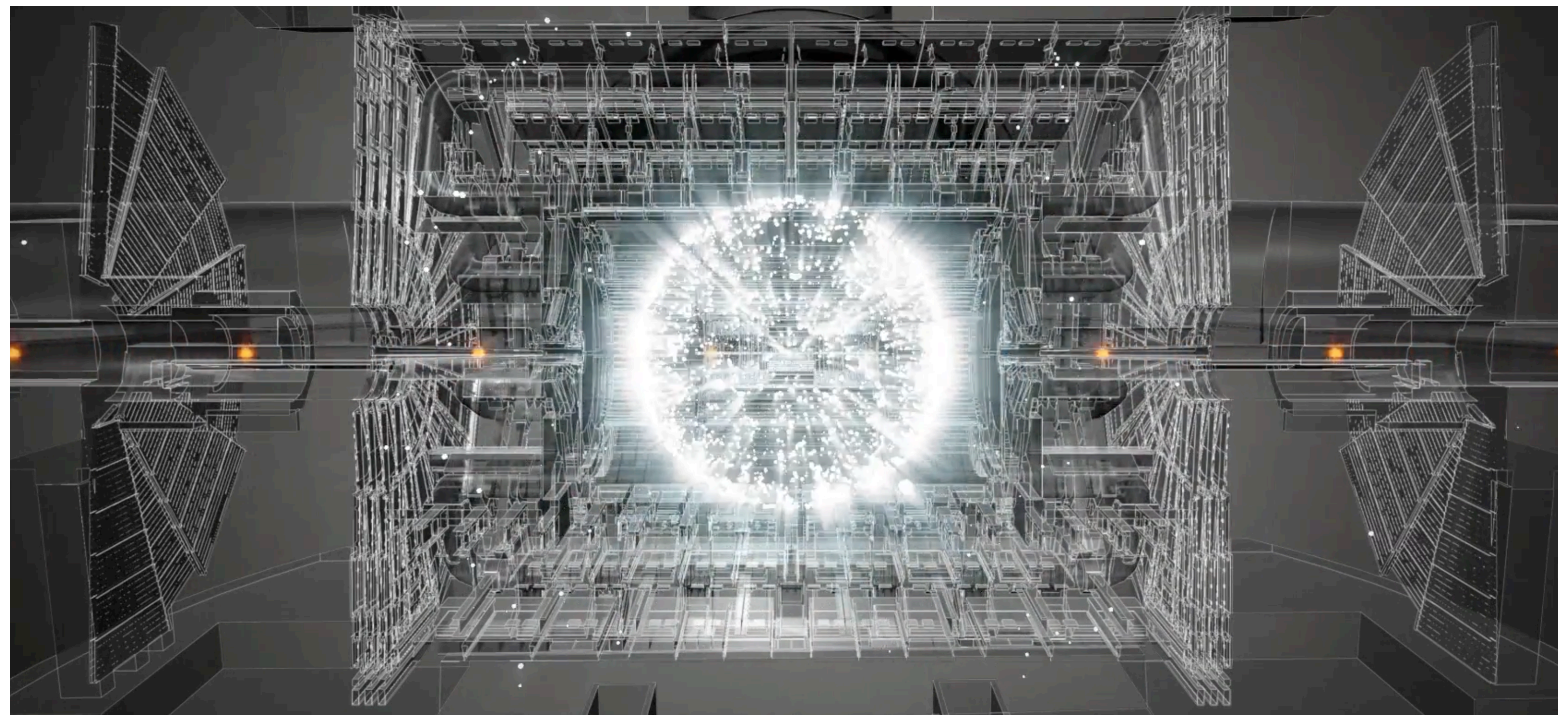
- Increase the integrated luminosity by a factor of 10 beyond the LHC's design value
- 140-200 simultaneous p-p interactions (pile-up), compared to the current value ~40
- Will allow physicists to study Higgs boson known mechanisms in greater detail, and observe rare new phenomena
- HL-LHC will produce at least 15 million Higgs bosons per year, compared to around three million from the LHC in 2017
- About to start in 2029

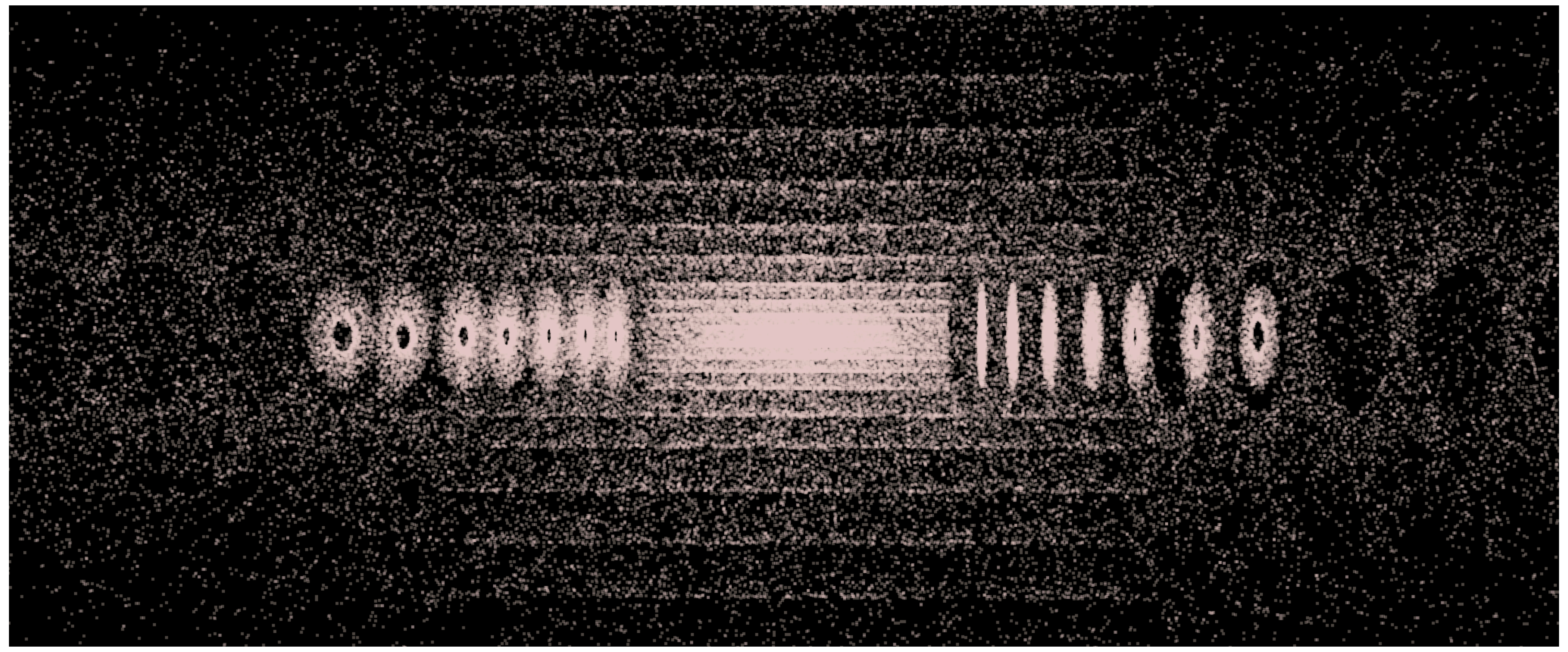


Tracking during HL-LHC : A computing challenge



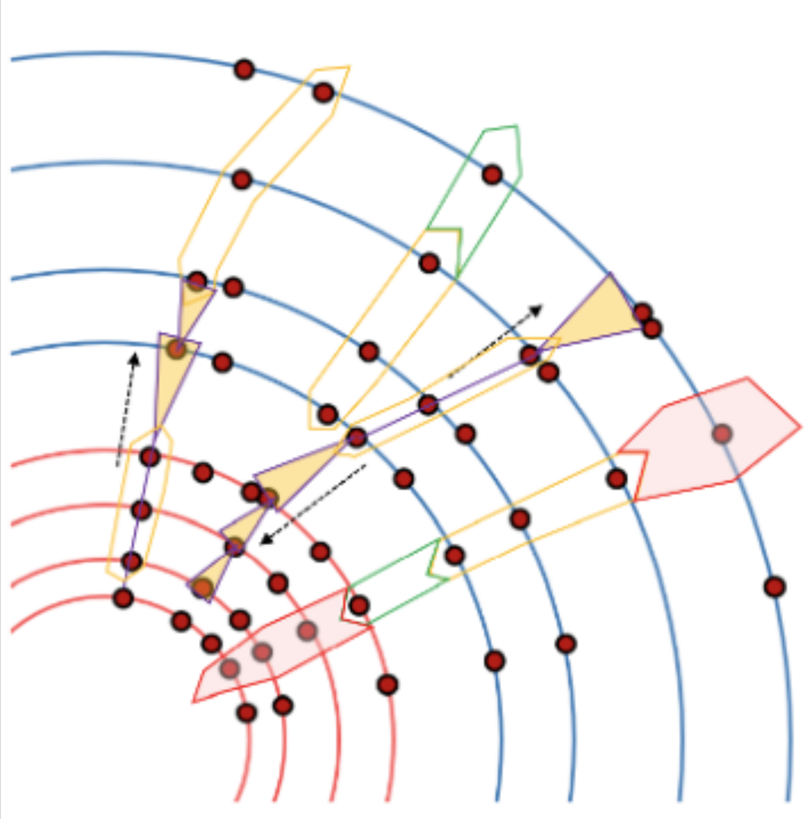
Good old time: analogic signal in liquid hydrogen in Bubble chamber



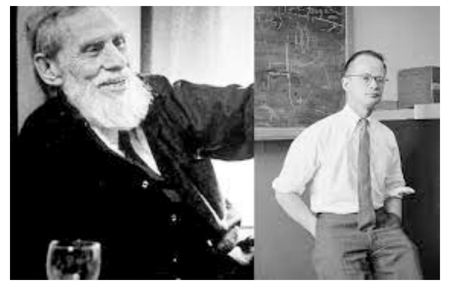


Cloud of hits (space point)
Coming from the interaction between the particles and the detector silicon modules in each layers
(~300K hits / event in ITk during HL-LHC)

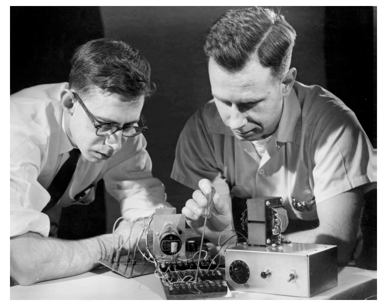
- HL-LHC => 140-200 simultaneous p-p interactions (pile-up), compared to the current value ~40
- It means $O(10^4)$ particles generated per event, compared to the current $O(10^3)$ particles per. event
- Track reconstruction is a key step in the event reconstruction and the identification of particles and their physic parameters
- Current algorithms (i.e. Combinatorial Kalman Filter) will not be able to cope with the complexity and rate of the data recorded
- Machine Learning solution investigated



Rise of Geometric Deep Learning



First concept of artificial Neural Networks
McCulloch, Pitts (1943)



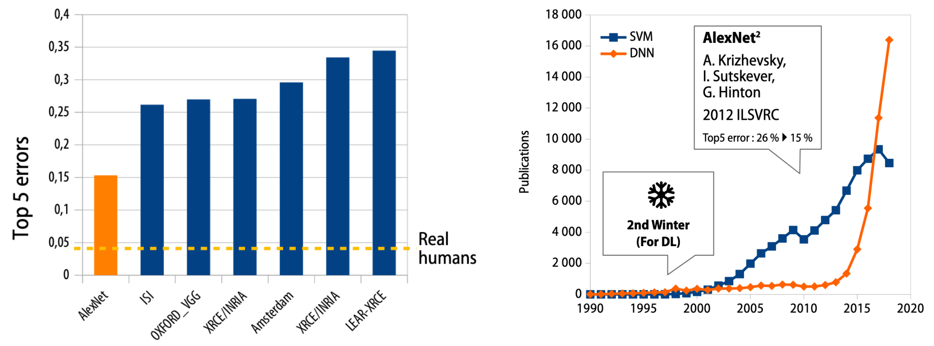
Perceptron
Rosenblatt (1957)

Multi layer Perceptron
Minsky, Papert (1957)

❄❄❄

Backpropagation algorithm
Rumelhart (1986)

CNN
LeCun(1989)



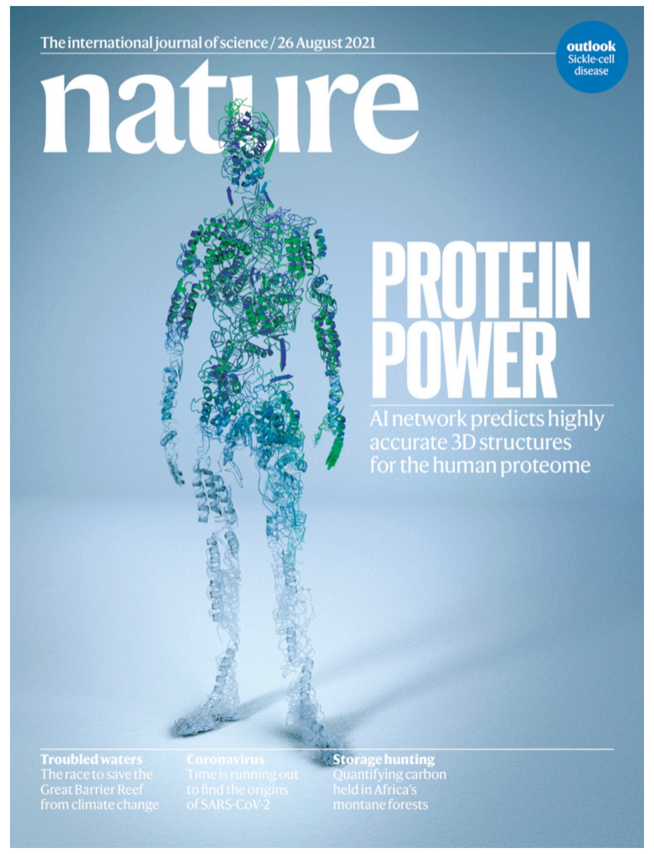
AlexNet CNN (30 September 2012)

- Since 2012 Convolutional Neural Networks (CNNs) widely used in computer vision: capture patterns at different spatial frequency
- Recurrent Neural Networks (RNNs) used in Natural Langage Processing and time series analysis: capture temporal or ordering patterns

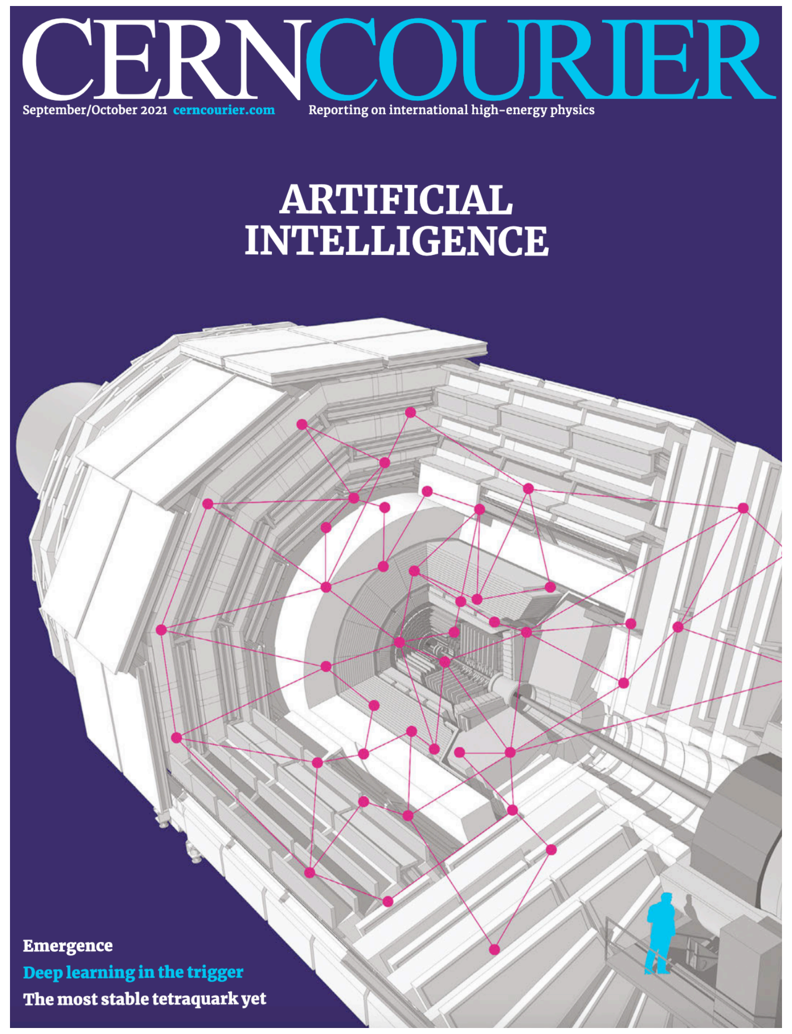
- In 2017: Transformers architecture with attention mechanism have revolutionized Natural Langage Processing : capture deep linguistic patterns

All of these models learn different structural patterns in data

Recently (~ since 2018) Geometric Deep Learning has generalized representation learning at any kind of structured data (Grids, Groups, Graphs, etc…) to Capture deep structural patterns in data : New models like Graph Neural Networks (GNNs) are now widely used in many domains



'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures



Graph Neural Networks (GNNs)

Aggregate neighbourhood latents
Project non-linearly in a new latent space

$e_{ij}^{t+1} \leftarrow \psi(e_{ij}^{t} \,|\, h_i^t \,|\, h_j^t)$ $\quad$ $h_i^{t+1} \leftarrow \phi(h_i^t \,|\, \sum_{j \in N_{v_i}} e_{ji}^t)$

Edge classification / link prediction

$z_{ij} \leftarrow f_{decoder}(e_{ij})$

$h_i \leftarrow f_{encoder}(x_i)$

Embeddings of nodes (eventually edges) features in high dimensional Latent space

$h_i^{t+1} \leftarrow \phi(h_i^t \,|\, \sum_{j \in N_{v_i}} h_j^t)$

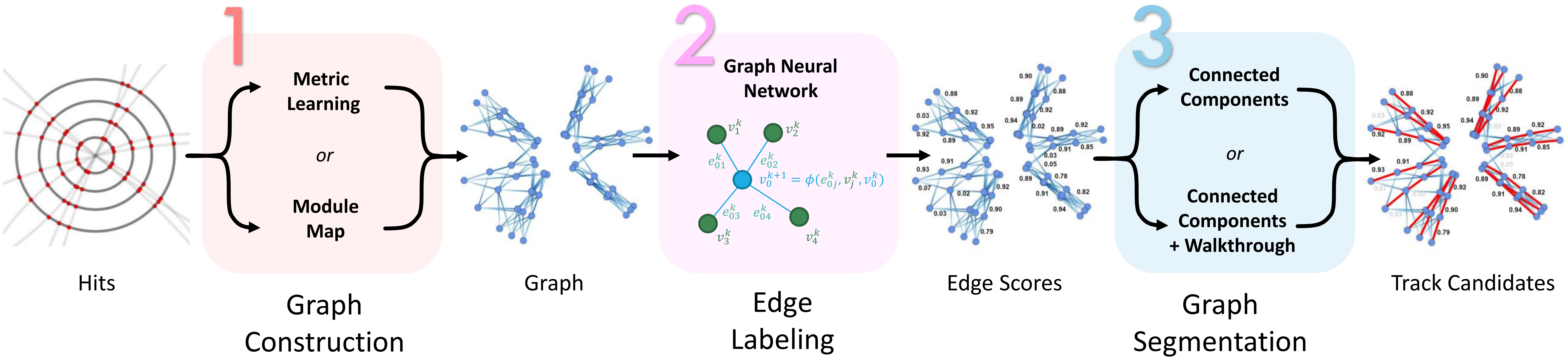
$z_i \leftarrow f_{decoder}(h_i)$

Node / graph classification

Repeat xN GNN layers
=> **Capture deep structural patterns in the graph**

GNN-based track reconstruction in ITk for HL-LHC

- Geometric Learning solution explored by Exa.TrkX Project and L2IT : « ATLAS ITk Track Reconstruction with a GNN-based pipeline » C.Rougier (CTD2022), [P. Calafiura, CHEP2023] [X. Ju, CHEP2023] [S. Caillou, CHEP2023]

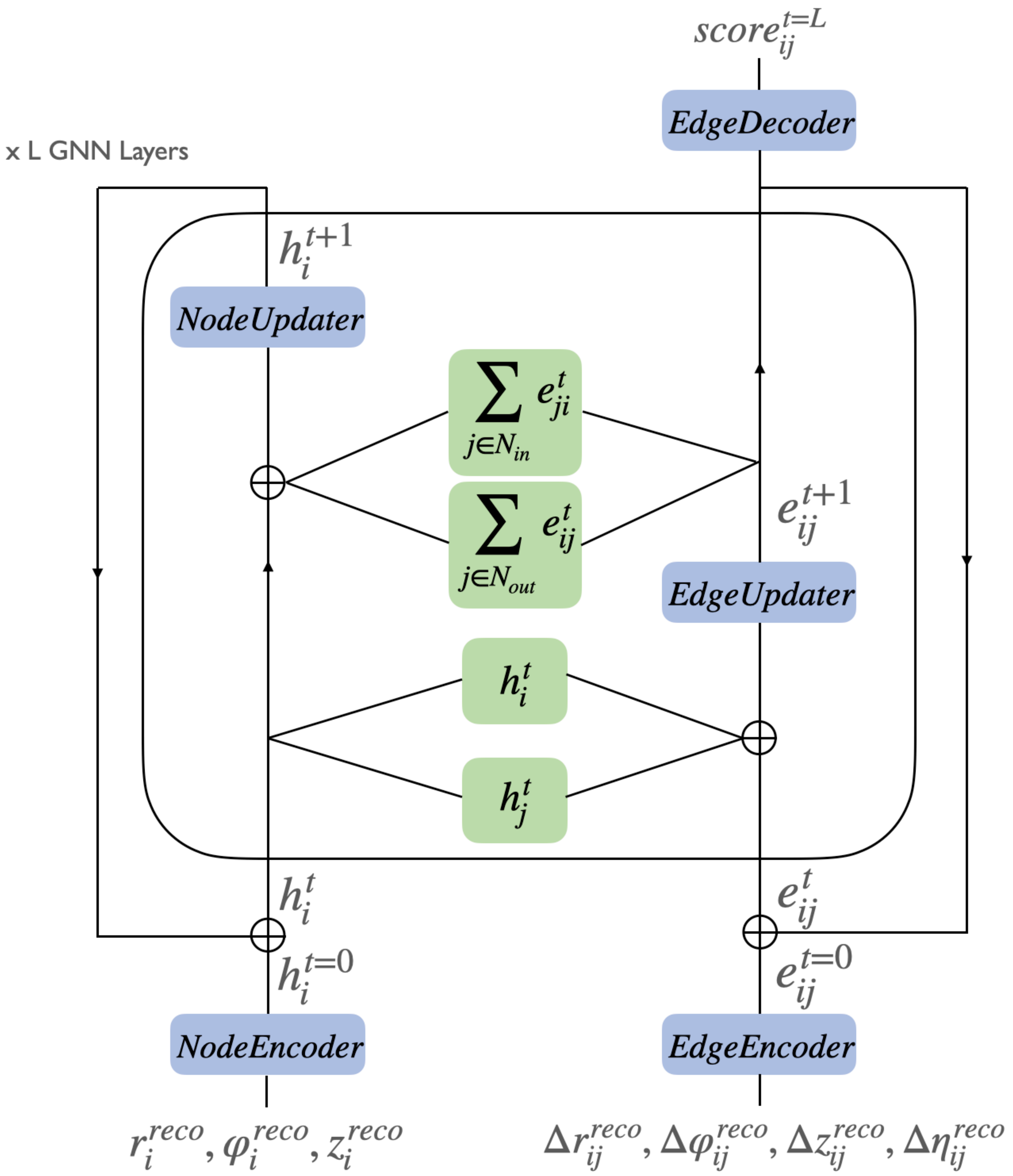


1) Detector data represented as graph
- *Nodes* are hits in the detector
- *Edges* are *possible connections* between nodes
- *True edges* are connections between successive hits from *the same particle of interest*

2) GNN learns *deep geometric patterns* of the particle tracks and classifies edges between true and fake edges

3) Post-processing algorithm operates on scored graph to build *track candidates*

GNN4ITk Message Passing Neural Network (CTD 2022)



Decode (scoring) stage

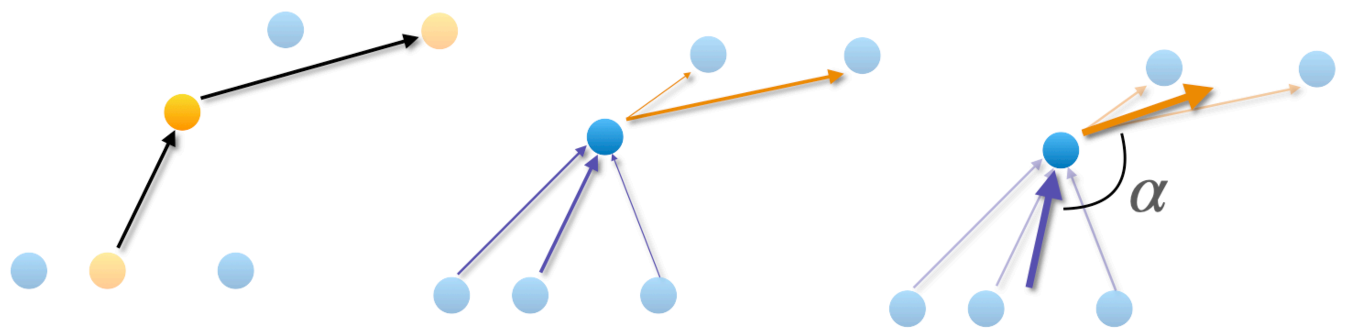
Edge latent is projected to a scalar value which is the score of the edge

$$\boxed{score_{ij}^{t=L} \leftarrow EdgeDecoder(e_{ij}^{t=L})}$$

GNN stage

- Edge latent is updated taking into account latent of source and destination nodes
- Node latent is updated from a ***separate* aggregation of incoming and outcoming edges**

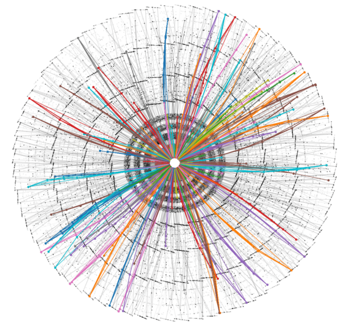
$$\begin{aligned} e_{ij}^{t+1} &\leftarrow \psi(e_{ij}^t \,|\, h_i^t \,|\, h_j^t) \\ h_i^{t+1} &\leftarrow \phi(h_i^t \,|\, \sum_{j\in N_{in}} e_{ji}^t \,|\, \sum_{j\in N_{out}} e_{ij}^t) \end{aligned}$$



Encode stage

- Node euclidian features are projected into latent space
- Edge *preprocessed* features are projected into latent space

$$\begin{aligned} h_i^{t=0} &\leftarrow NodeEncoder(r_i^{reco}, \varphi_i^{reco}, z_i^{reco}) \\ e_{ij}^{t=0} &\leftarrow EdgeEncoder(\Delta r_{ij}^{reco}, \Delta\varphi_{ij}^{reco}, \Delta z_{ij}^{reco}, \Delta\eta_{ij}^{reco}) \end{aligned}$$

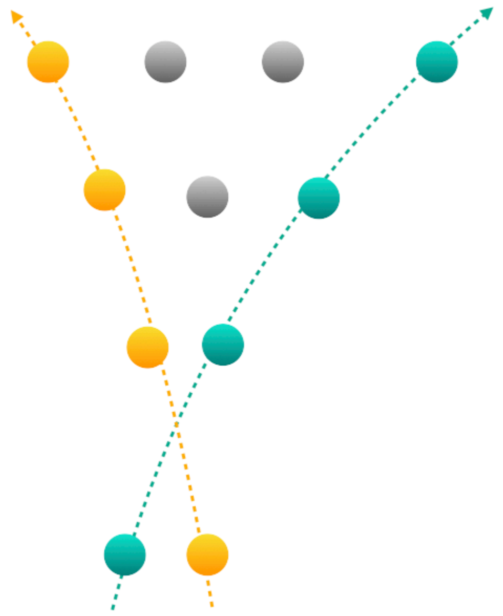


From GNN inference to track reconstruction

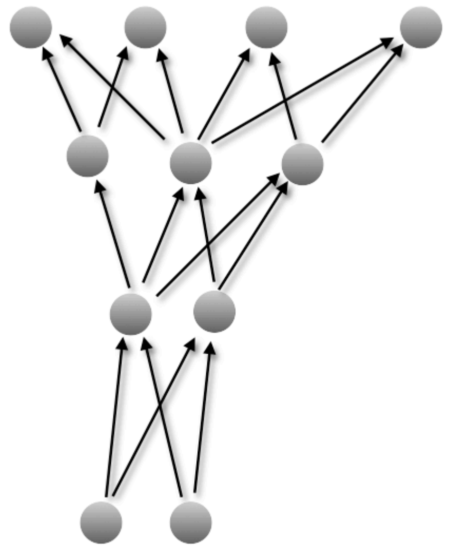
What is the link between GNN performance and track reconstruction ?

Graph construction

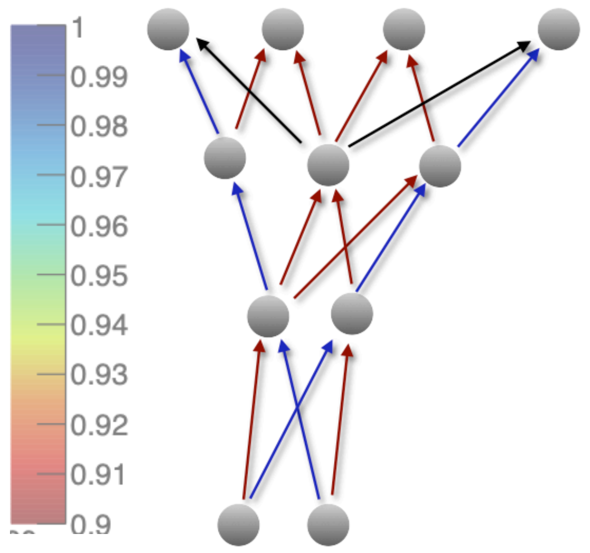
Raw Data



Graph data

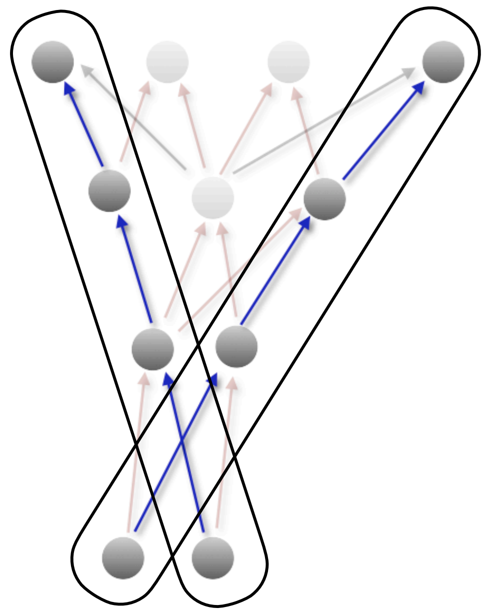


GNN edge classification

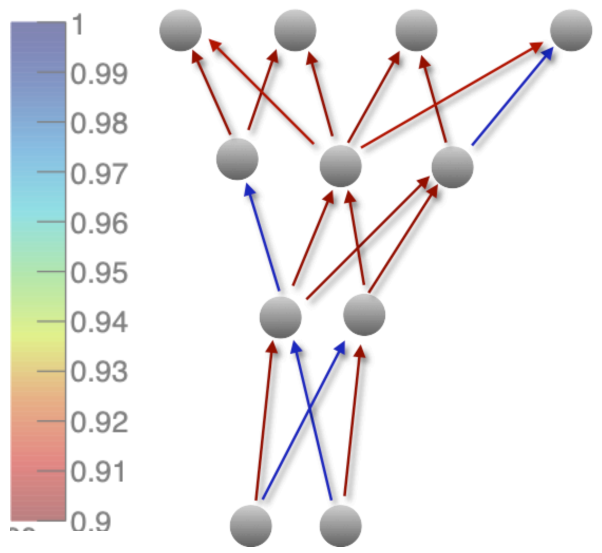


High GNN efficiency
High GNN purity

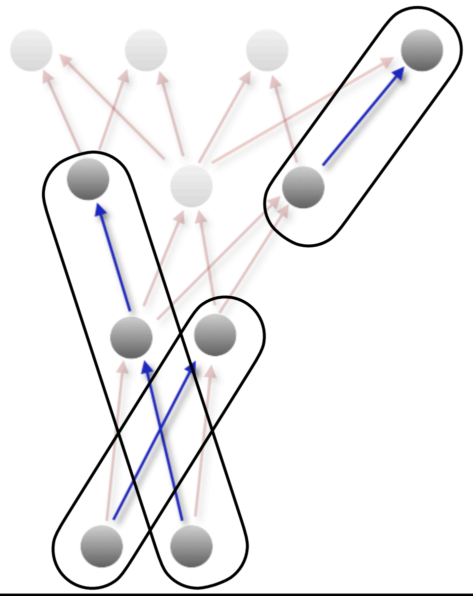
Track reconstruction



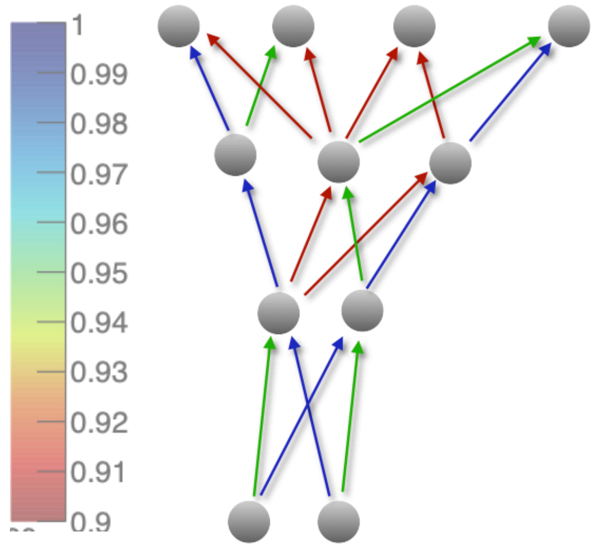
Connected Components (**CPU** & **GPU**)
Easy reconstruction



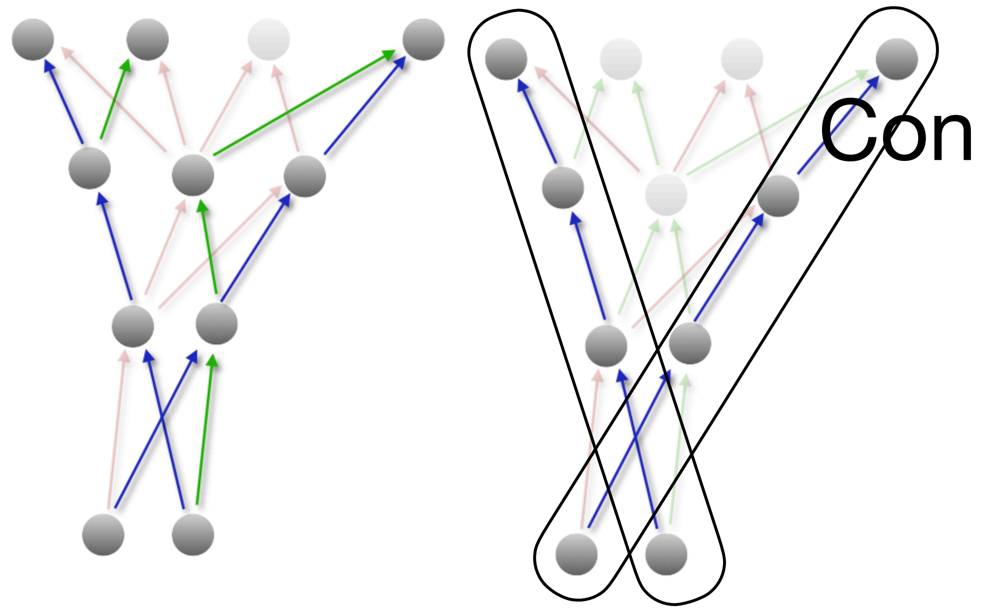
Low GNN efficiency
High GNN purity



Connected Components (**CPU** & **GPU**)
Holes in tracks



High GNN efficiency
Low GNN purity

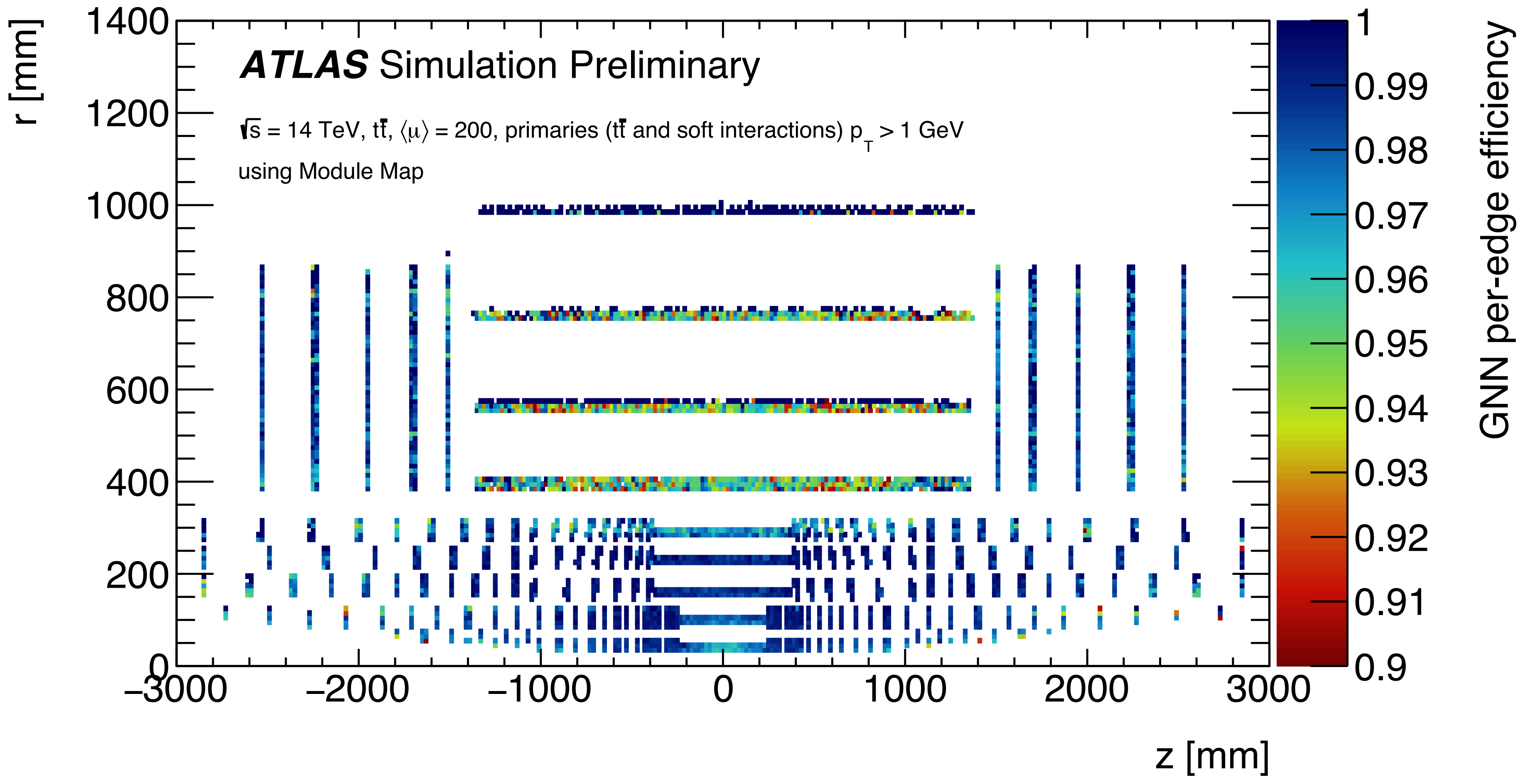


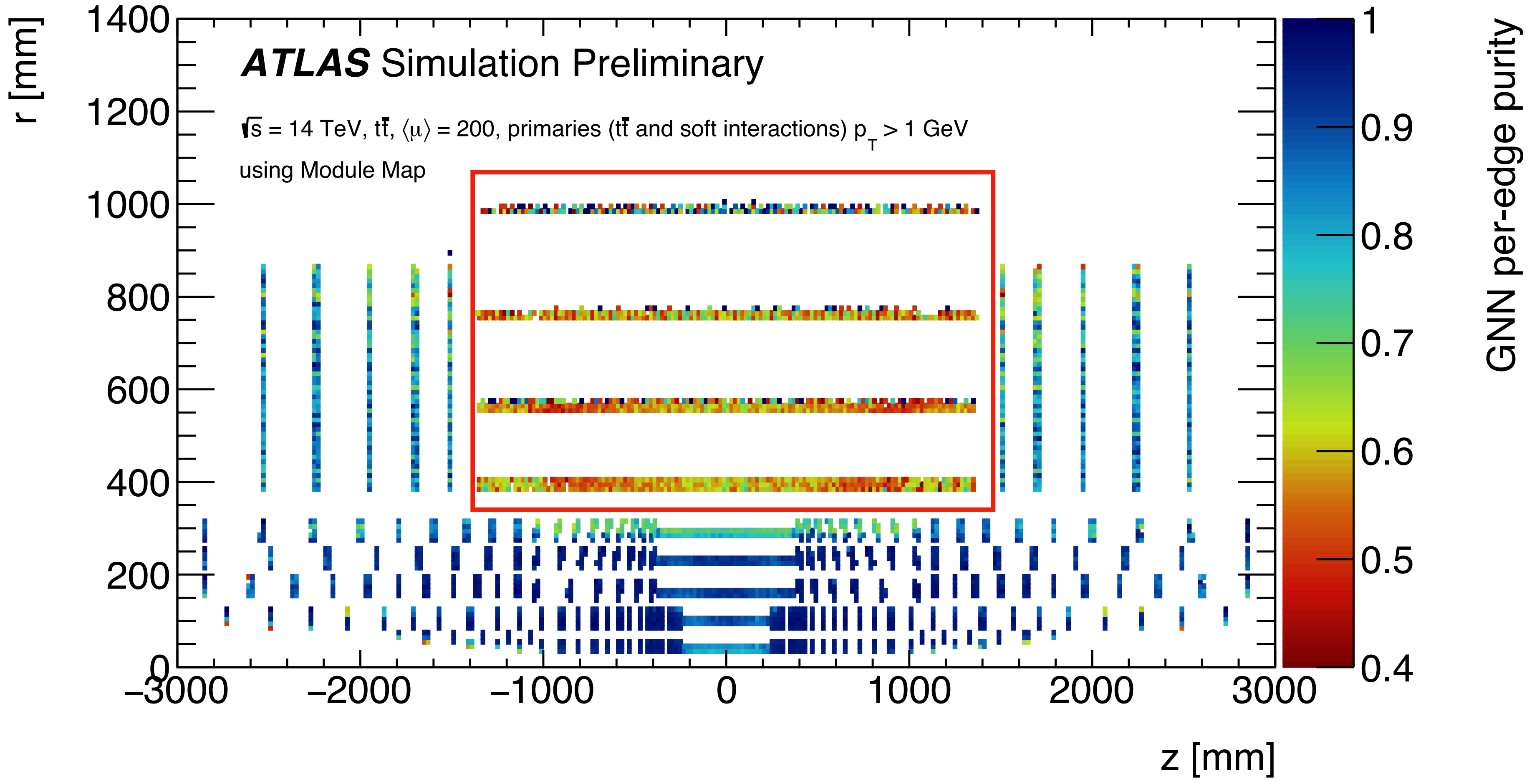
Connected Components (**CPU** & **GPU**)
+
Walkthrough algorithm (**CPU**)
Hard reconstruction

The higher GNN efficiency and purity are, the easier and the faster track reconstruction is !

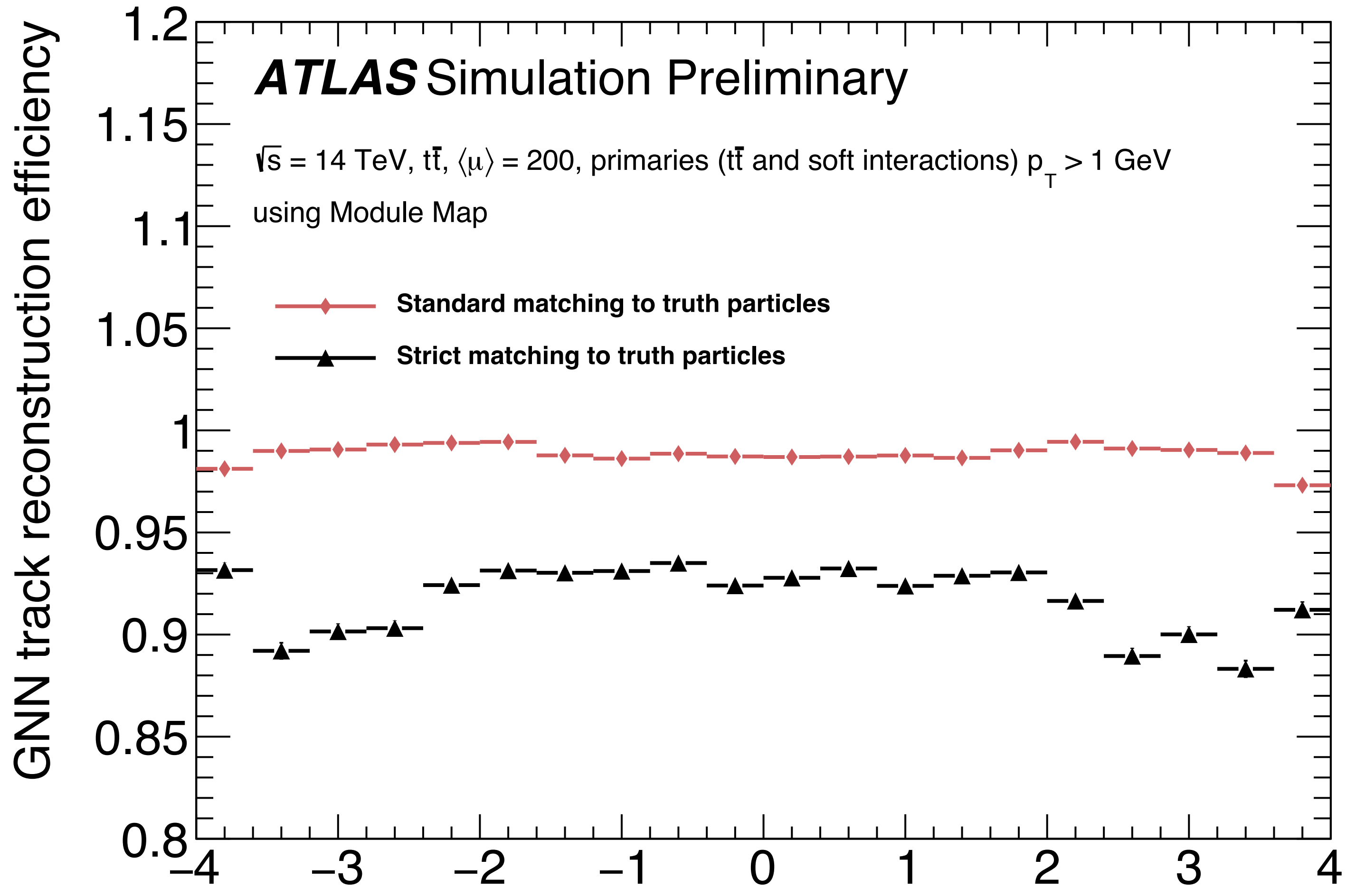
First results on ITk - CTD2022, Princeton

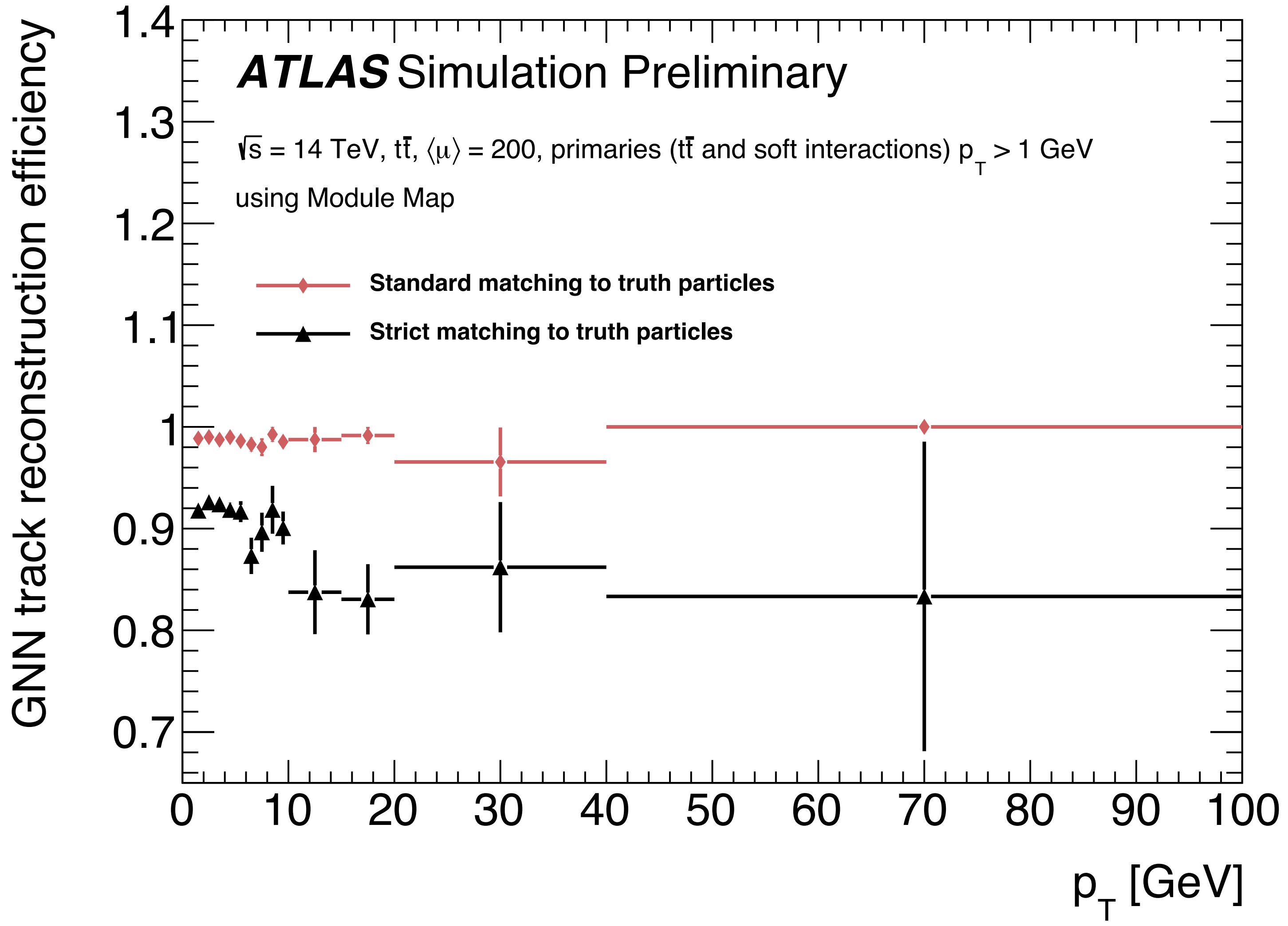
GNN performance results



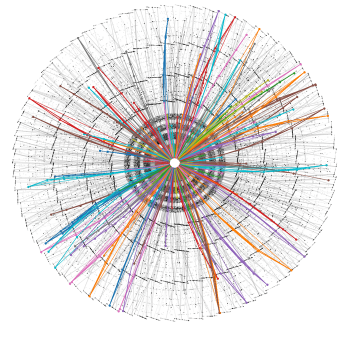


Track reconstruction performance results

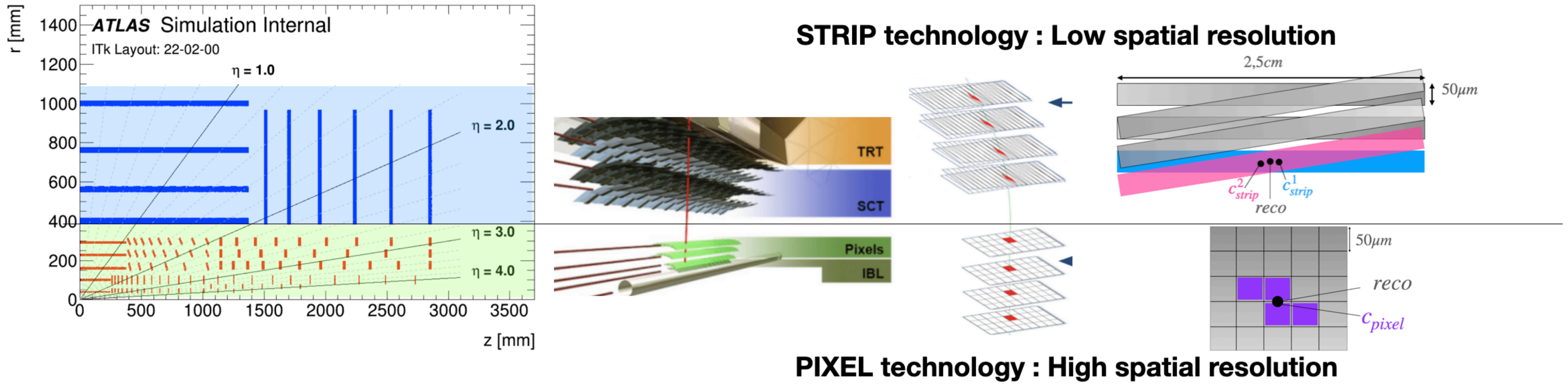




First very promising results:
Even with the poor GNN purity in the STRIP BARREL it was possible to get excellent track reconstruction performance **BUT** at the cost of the computation time of the post-processing algorithm

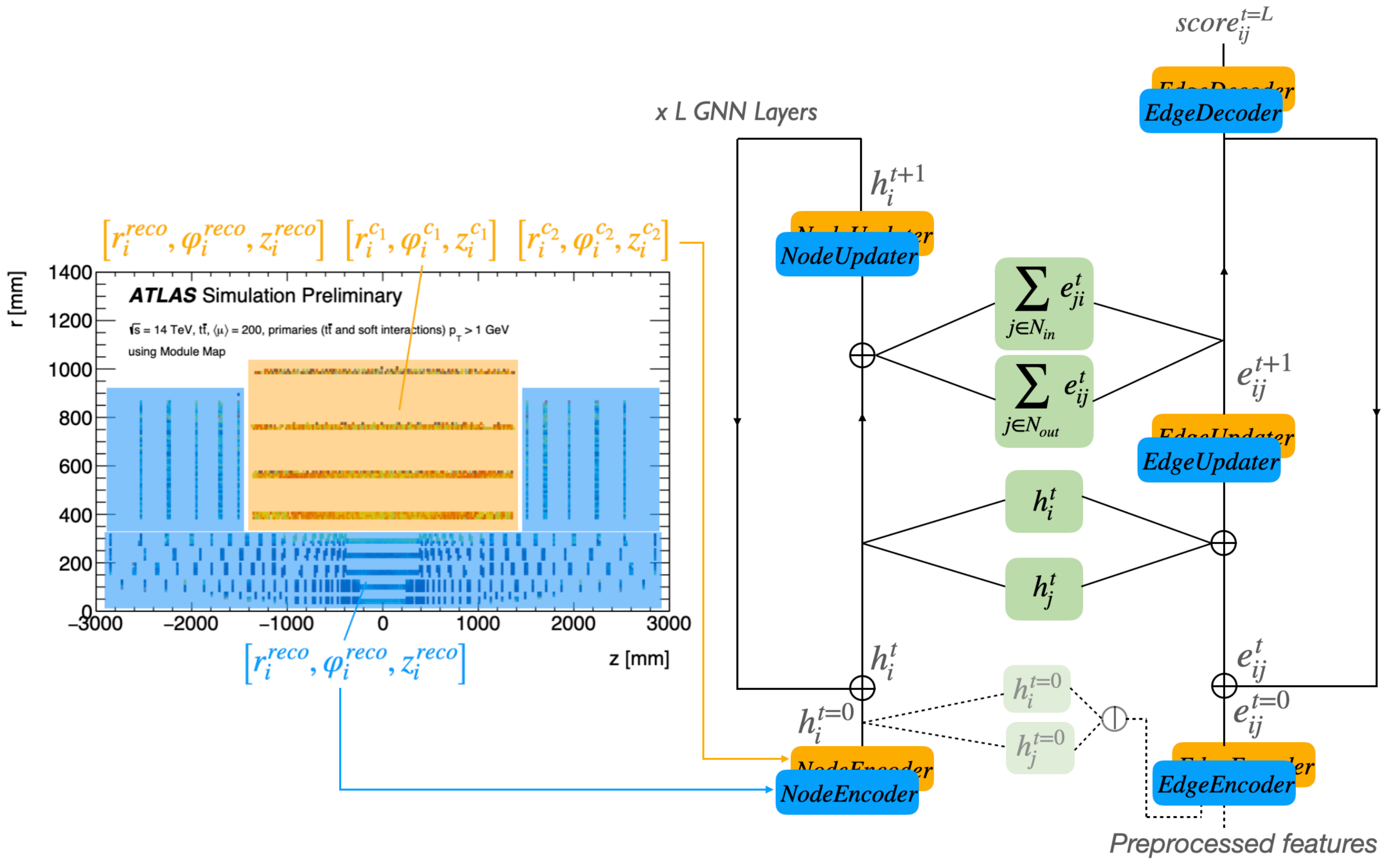


Recent progress : Handling Hardware and Data Heterogeneity

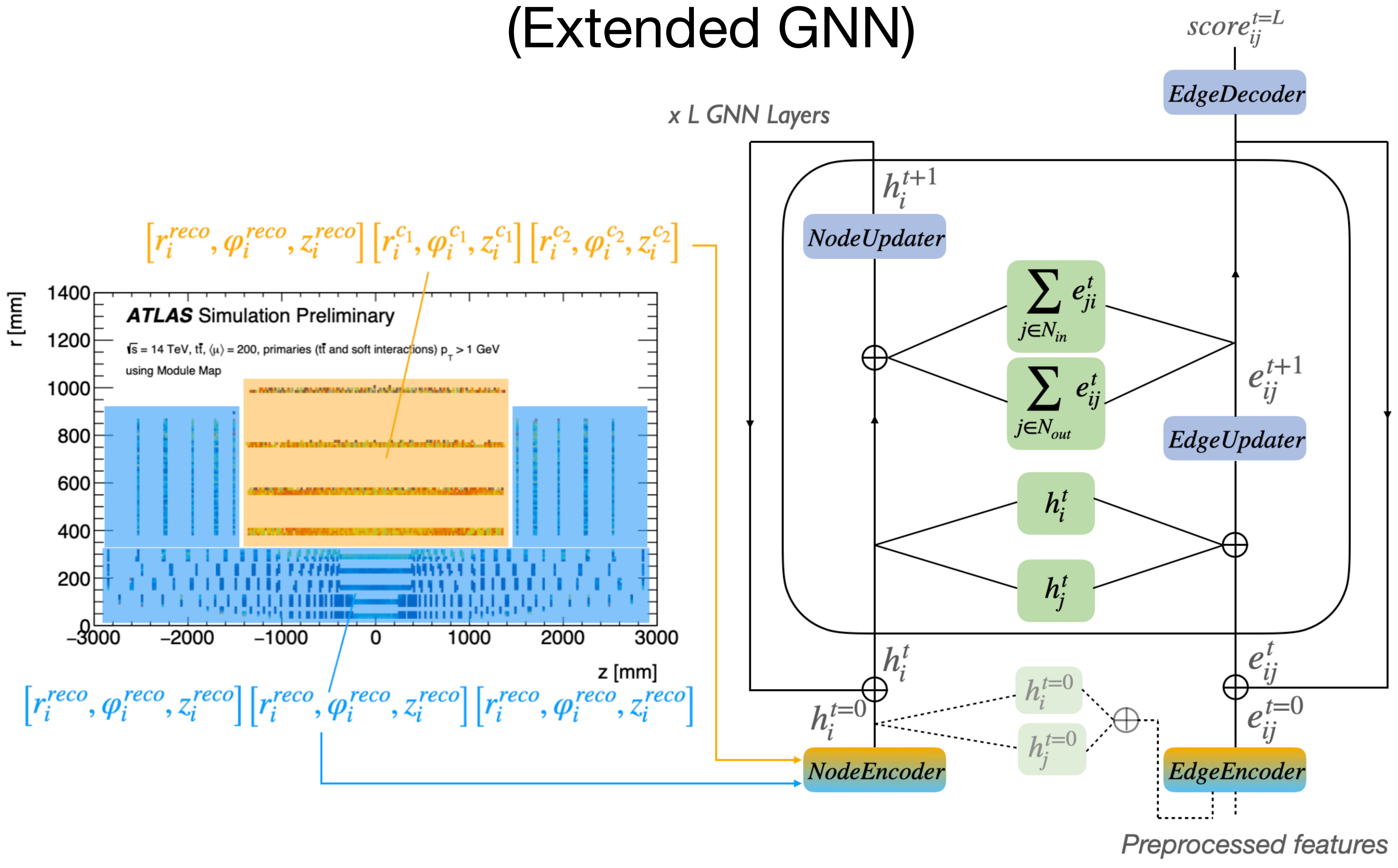


Design new GNN models to handle Heterogenous Data

Heterogeneous Data + Heterogeneous GNN

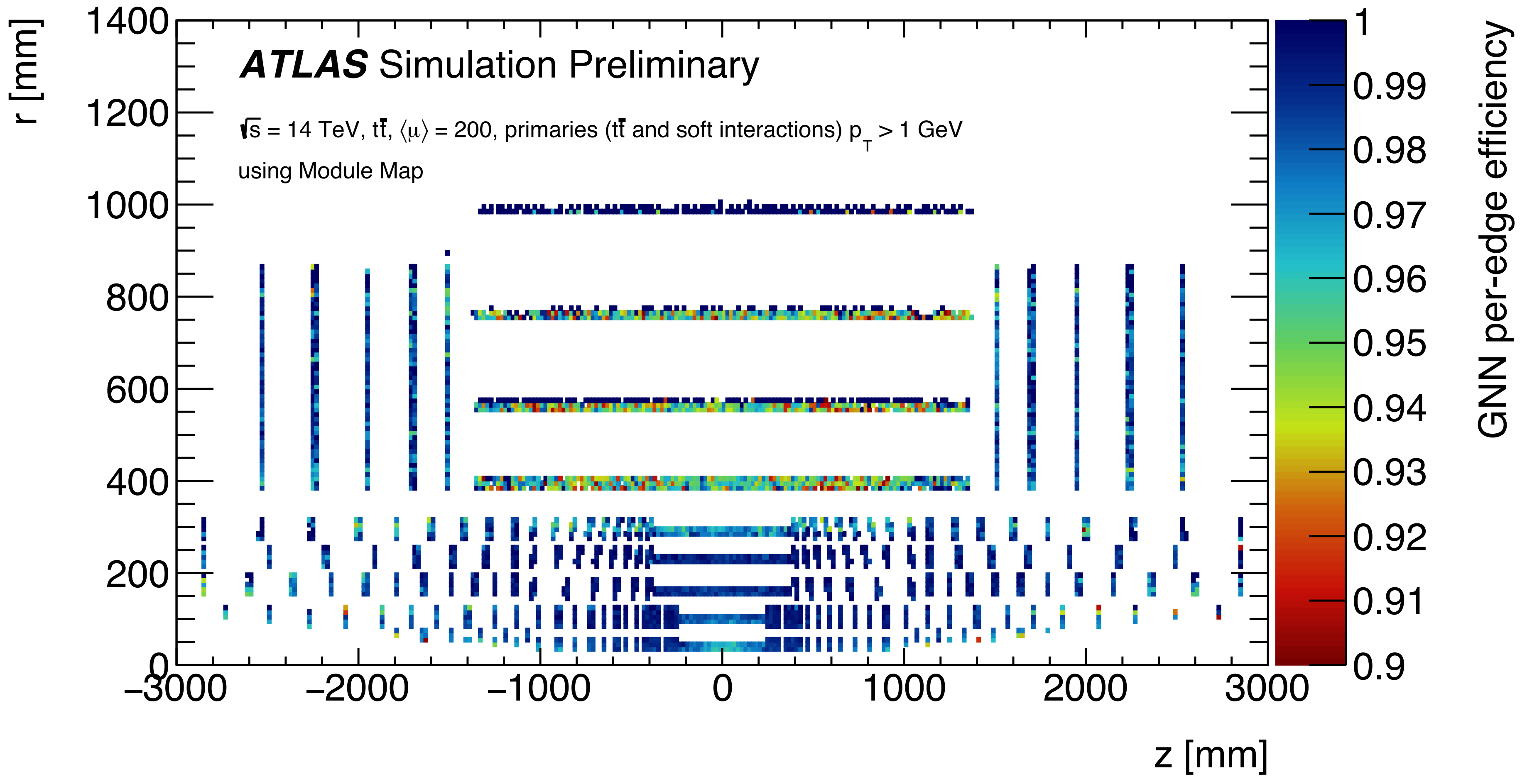


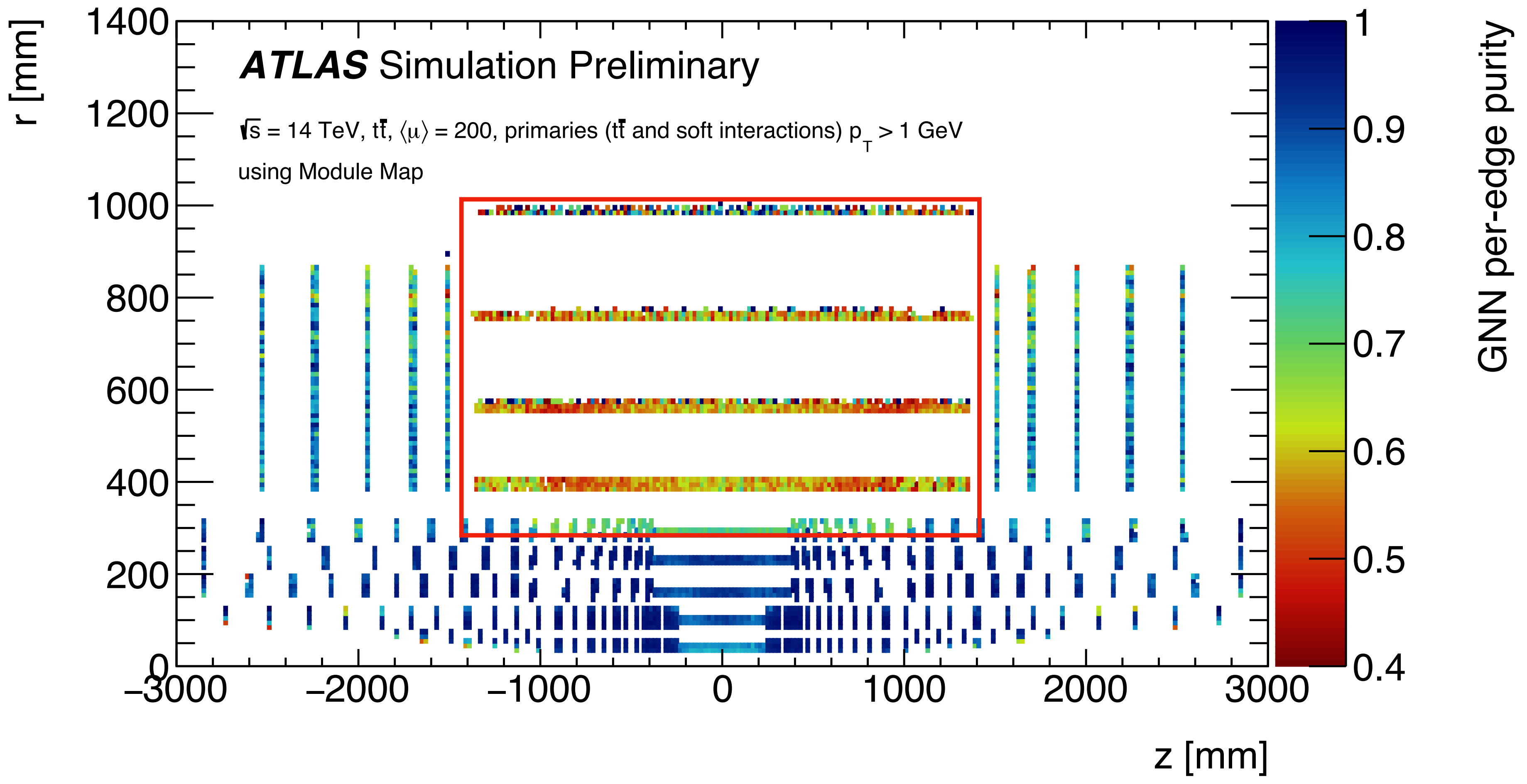
Heterogeneous Data + Homogeneous GNN (Extended GNN)



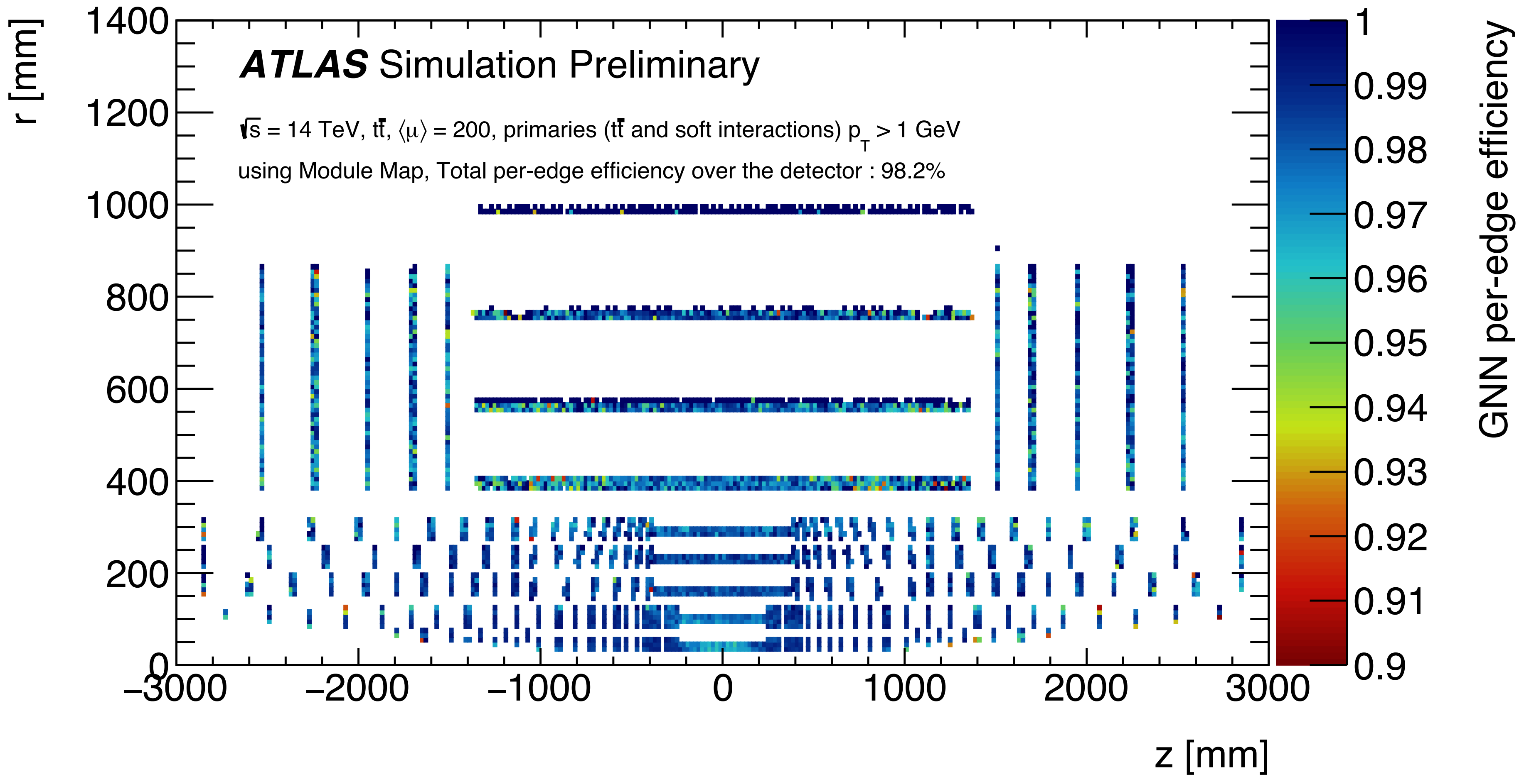
Last results on ITk - CHEP2023, Norfolk

CTD2022

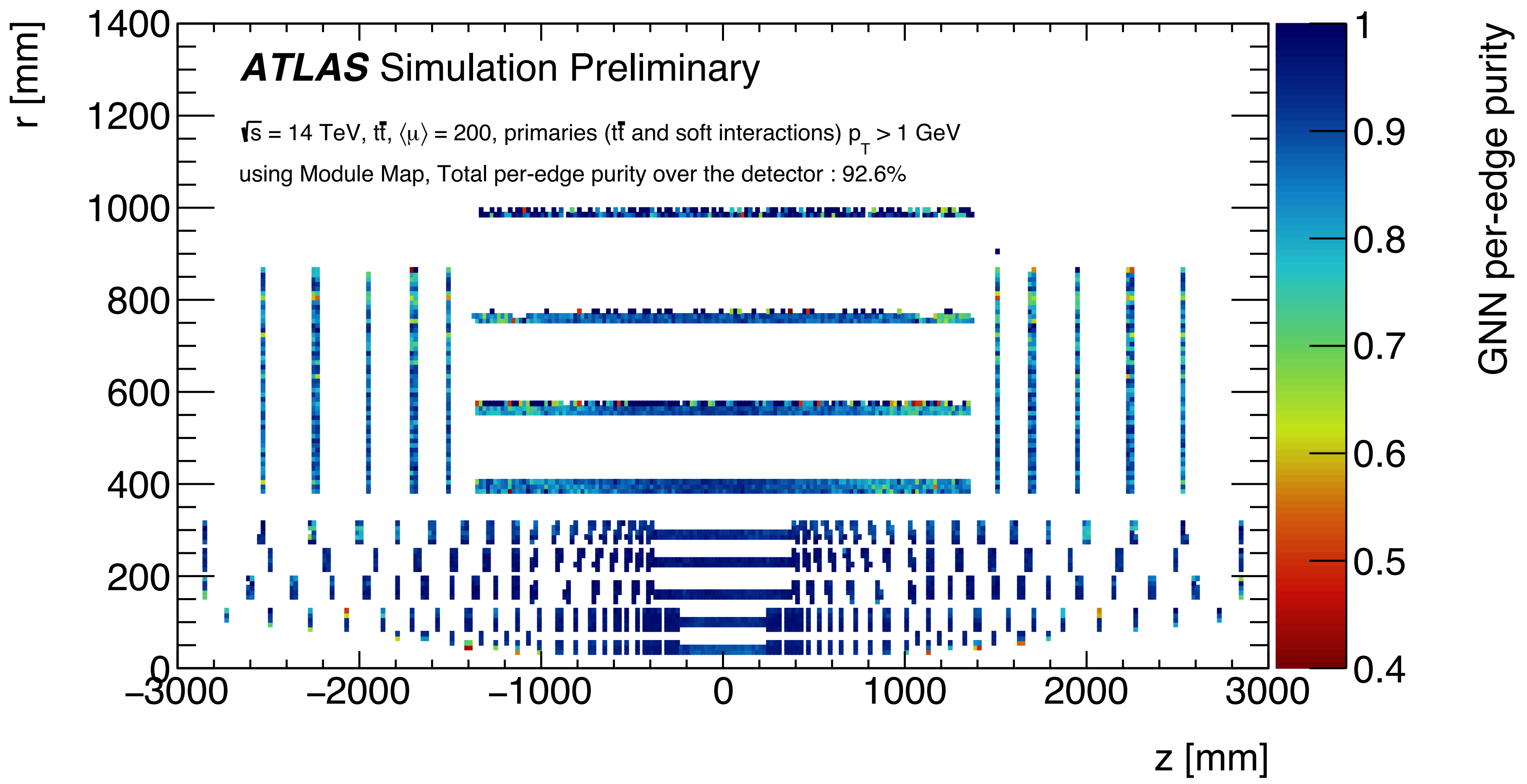




CHEP2023

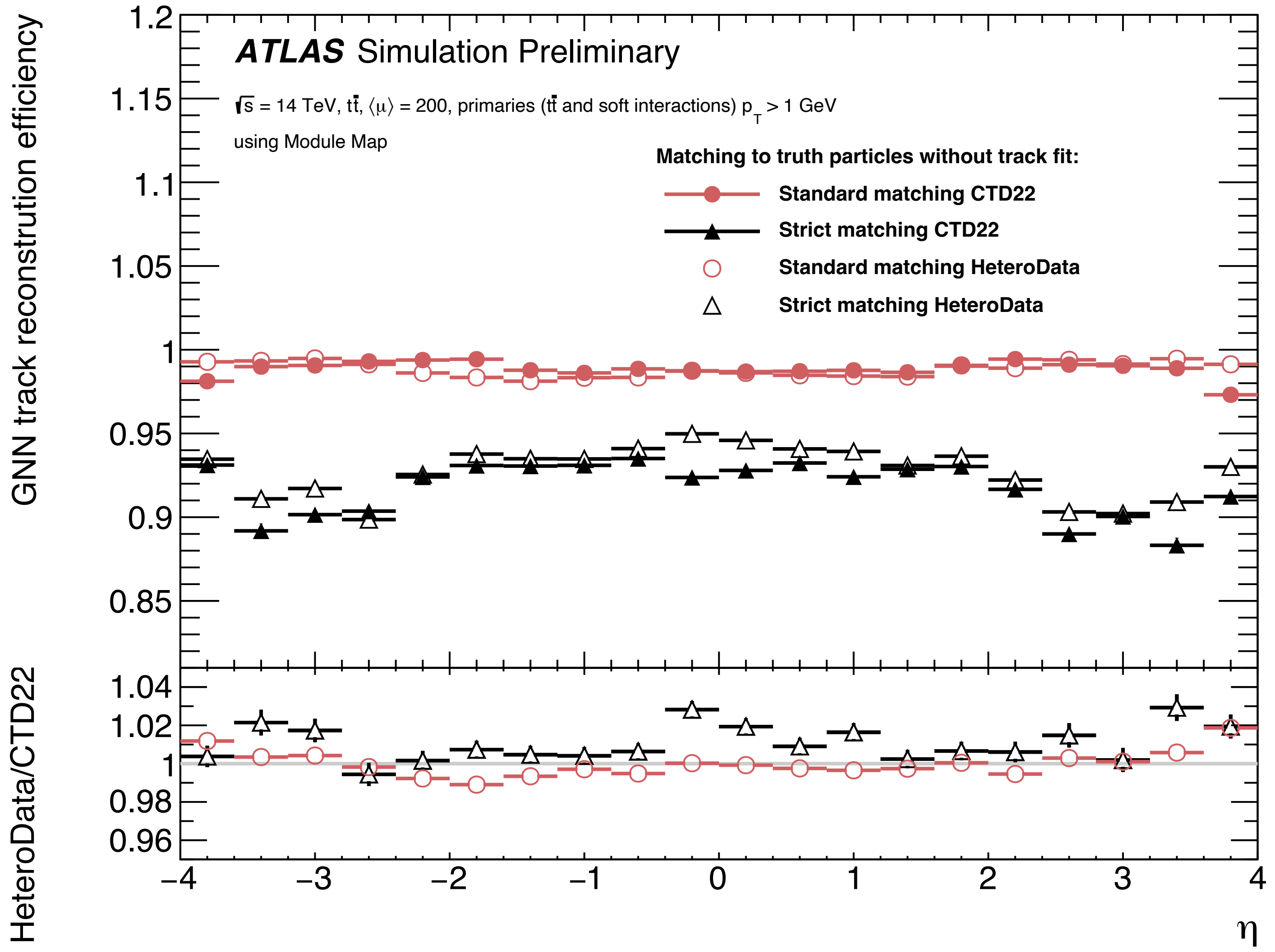


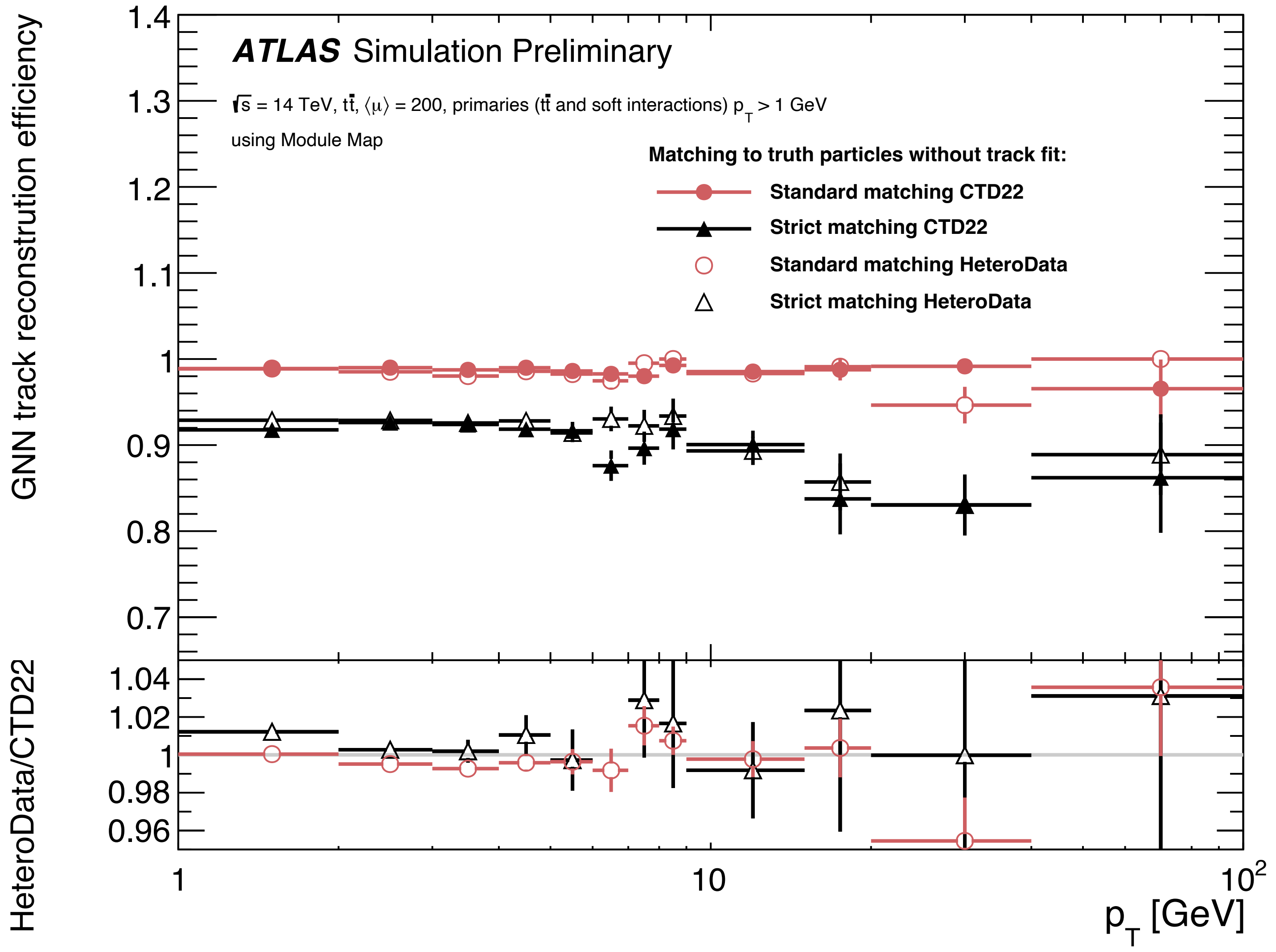
- Global efficiency of ~98%
- Efficiency more uniform in BARREL region



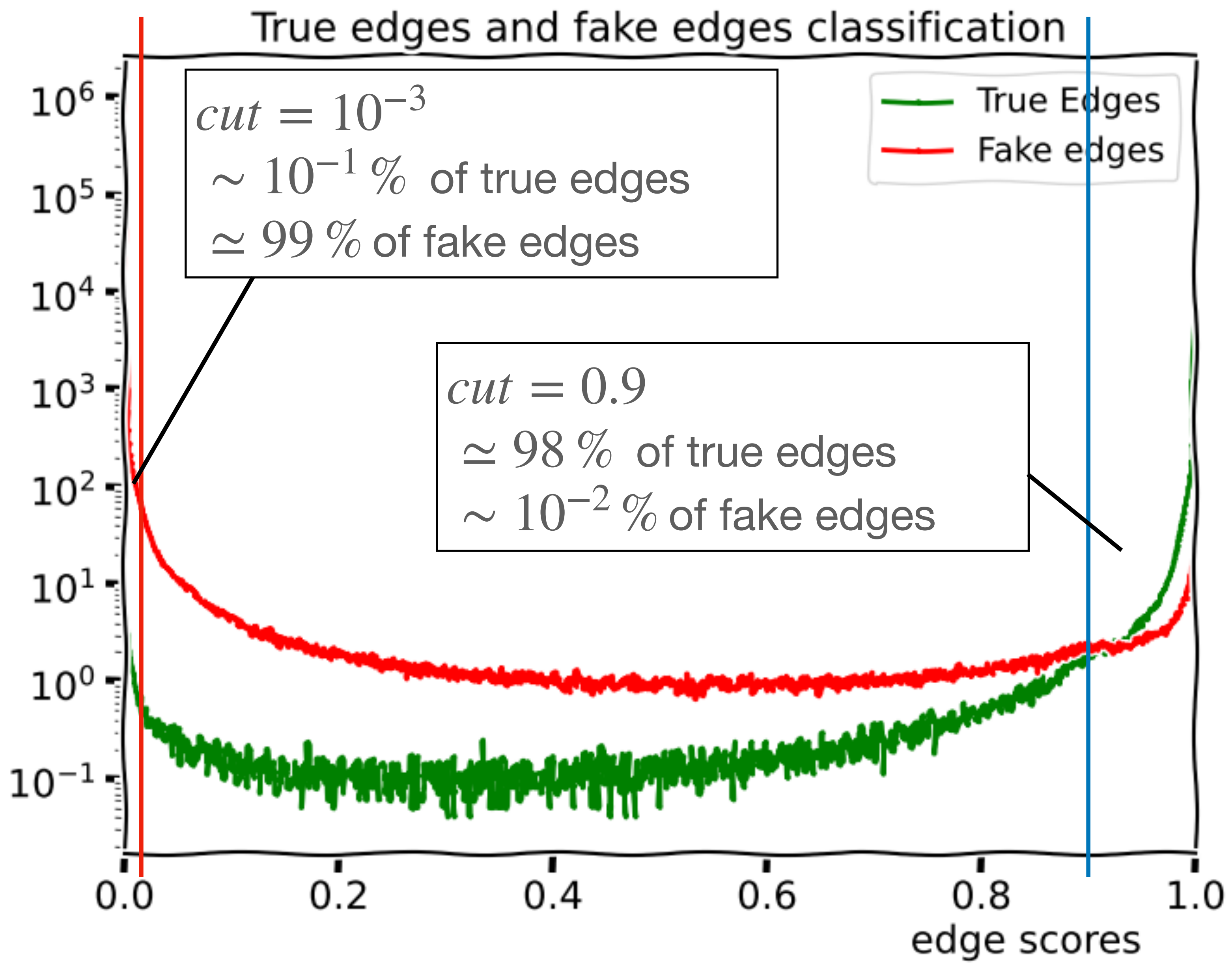
- **Significant improvement of purity in the STRIP BARREL region from ~40% to ~80%**
- Global purity of ~95%

Last results on ITk - CHEP2023, Norfolk





With the new high GNN efficiency and purity it is now possible to get 80% of perfect tracks and 95% of standard matching tracks with a simple Connected Components (very important as Connected Components algorithm **can be easily accelerated on GPU**). Walkthrough used only for small subset of tracks.



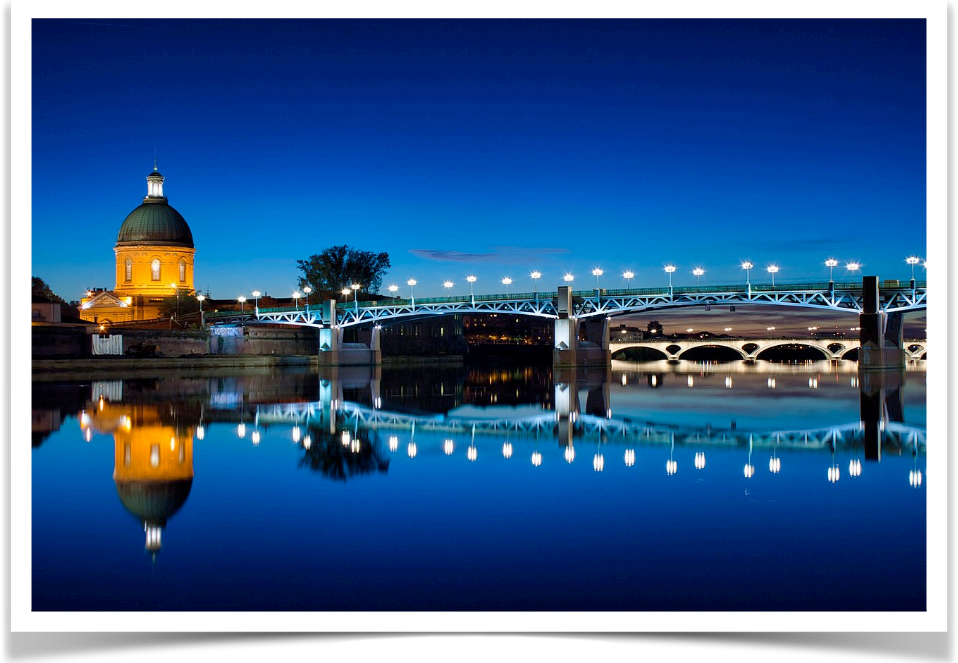
- **Low cut** to remove the majority of fake edges
- **High cut** to get **very high purity** and get « easy » tracks with Connected Components

Summary and further steps

- Exa.TrkX and L2IT R&D collaboration to understand and handle heterogenous data has led to high performance GNN models in ATLAS ITk simulated data
- The high level of GNN performance leads to a very high performance track reconstruction full algorithm in terms of **track efficiency and purity** and **computation time**
- Further steps:
  - **GNN R&D** (Heterogeneous GNN model, GNN filter, ambiguity resolution with GNN transformer)
  - Pursue GNN pipeline **software integration** in ACTS & Athena
  - **Optimization** and **acceleration** of graph construction and track reconstruction on CPU and GPU
  - **Towards deployment in production in 2025 !**

We developed a CommonFramework for GNN tracking R&D:

https://github.com/GNN4ITkTeam/CommonFramework



See you @CTD 2023 (Oct 10–13, 2023) in Toulouse, France !