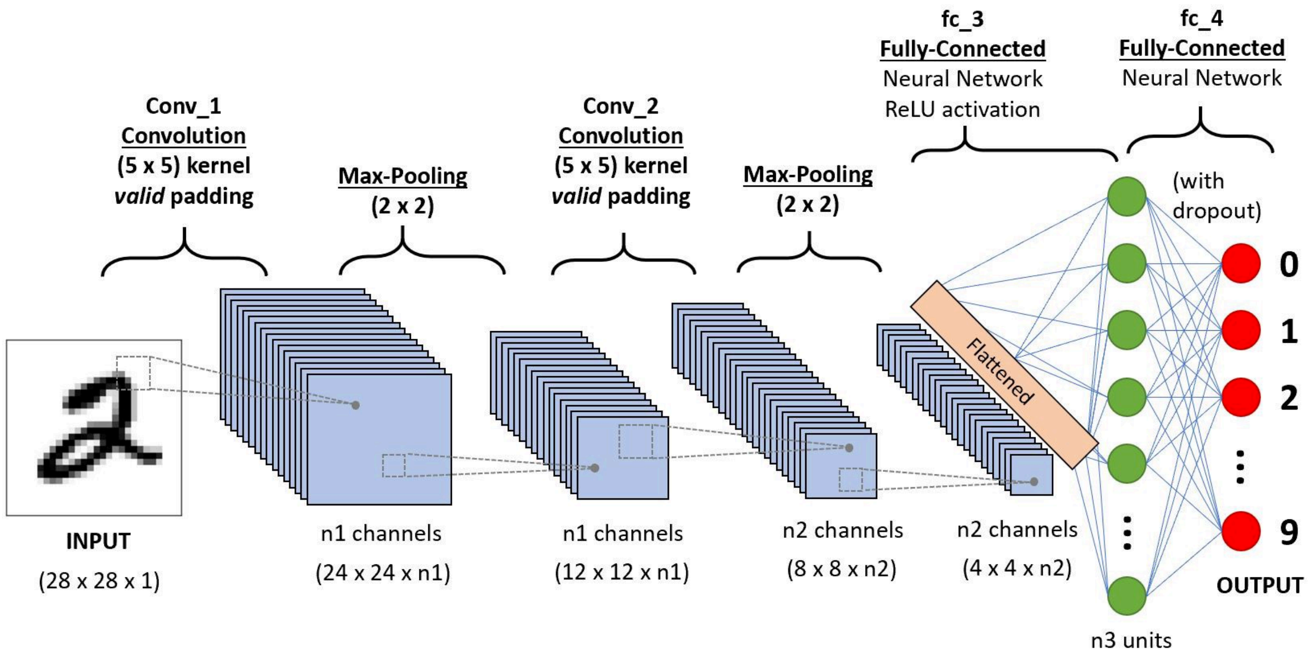
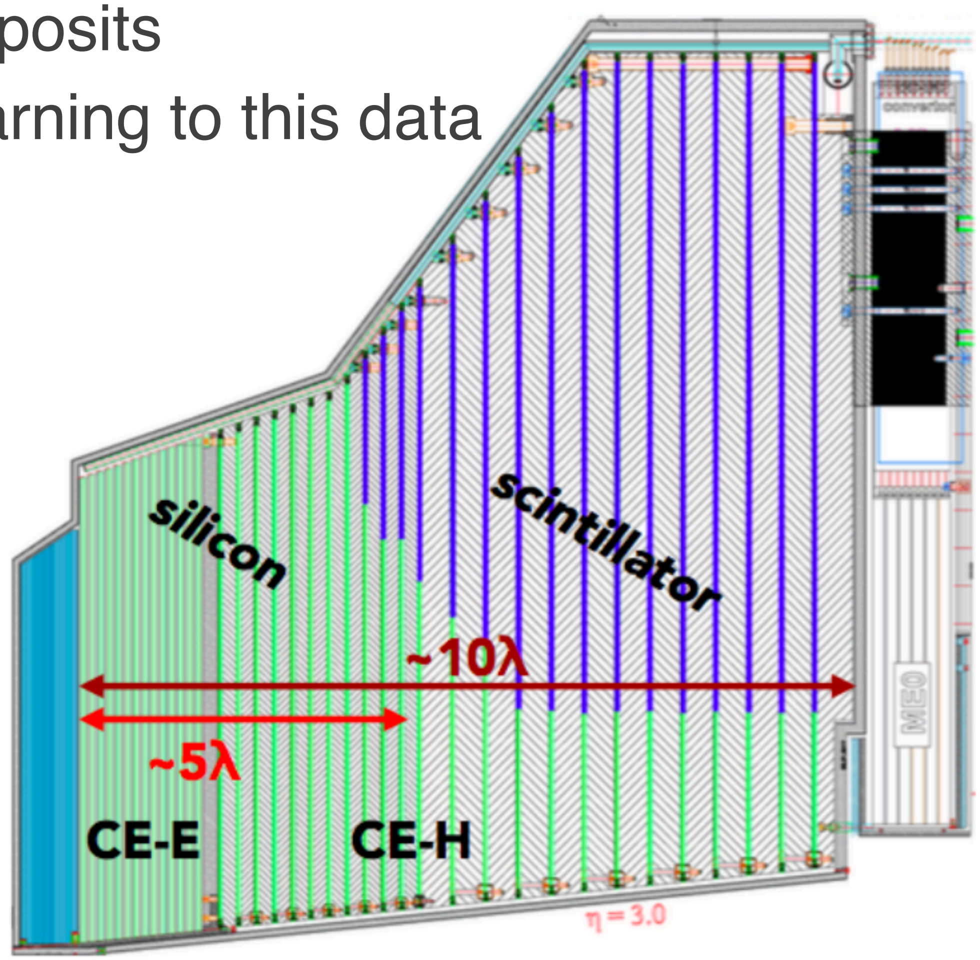
GNNs for Calorimeter Reconstruction in Collider Experiments

Lindsey Gray
CSAID Roadmap Meeting
9 February 2023

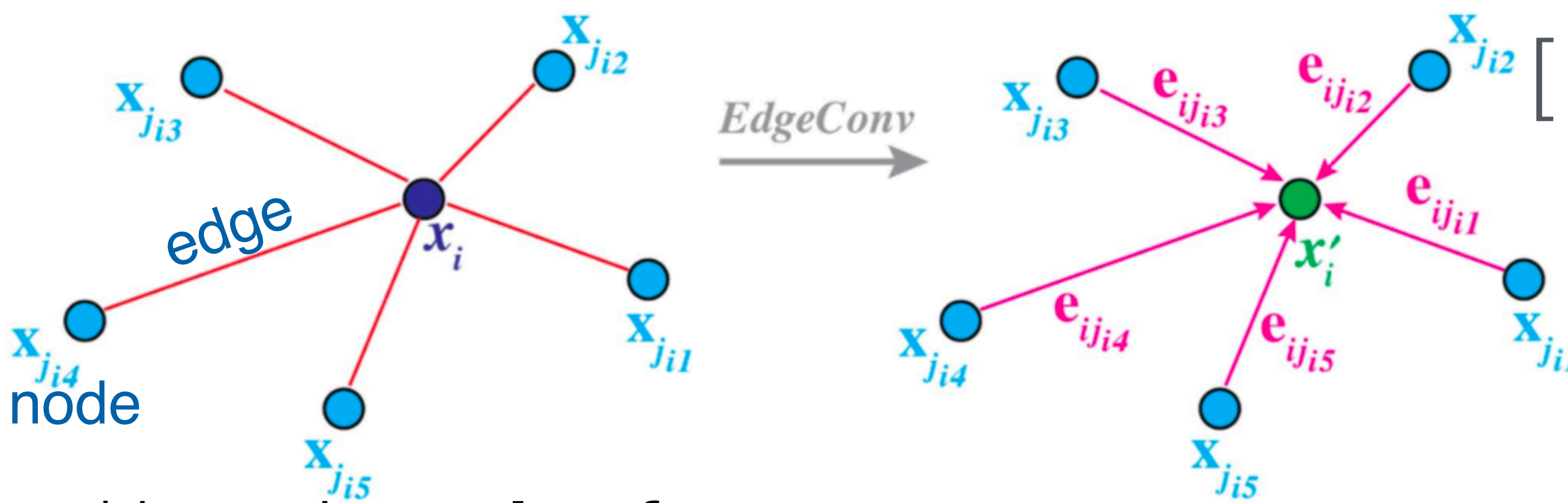
Exploiting granular information with machine learning

- Modern machine learning can determine important discriminating information in the course of training if the input 'shape' is fixed
  - Using convolutional neural networks for example, images are given as-is for training examples, discriminating features encoded in filters and high-dimensional 'latent spaces'
- However, many next generation particle physics detectors have irregular geometries with zero-suppressed outputs
  - Varying material with sparse sampling of energy deposits
  - Requires different approaches to apply machine learning to this data





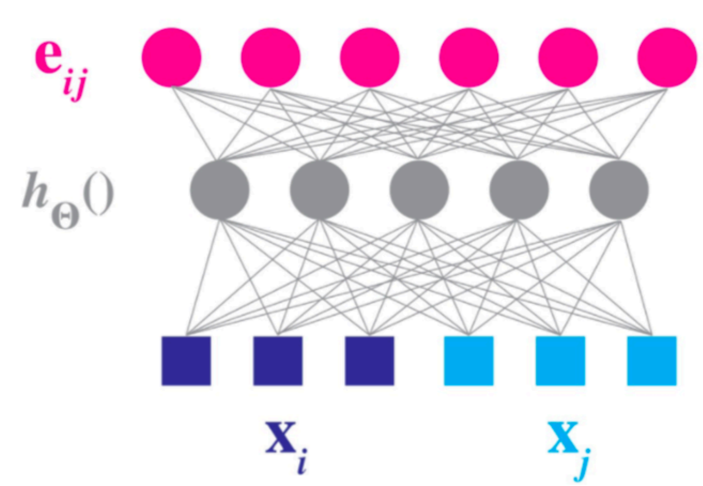
Graph Neural Networks: Edge Convolution



[1801.07829]

- Update $x_i \rightarrow x_i'$ by using **edge features**
  - i.e. learned features of the edges that connects x_i with its neighbors
  - Still independent of ordering of points, but uses **local geometry**
  - '**Convolutional**' as the operation is applied point by point to obtain **x**'
- These edge features and aggregation steps mimic the functionality of loops with if-statements in them (i.e. hand-written pattern recognition)

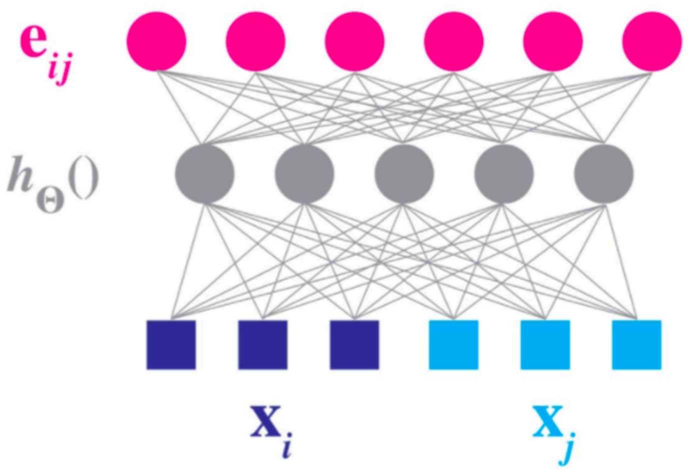
$$\mathbf{x}'_i = \mathop{\square}_{j:(i,j)\in\mathcal{E}} h_{\Theta}(\mathbf{x}_i, \mathbf{x}_j)$$



Can set dim of edge feature vector

Graph Neural Networks: Dynamic Graph Convolutions

[1801.07829]



$$\mathbf{x}'_i = \mathop{\square}_{j:(i,j)\in\mathcal{E}} h_\Theta(\mathbf{x}_i, \mathbf{x}_j)$$

$$h_\Theta(\mathbf{x}_i, \mathbf{x}_j) = h_\Theta(\mathbf{x}_i)$$ No neighborhood info (only global)

$$h_\Theta(\mathbf{x}_i, \mathbf{x}_j) = h_\Theta(\mathbf{x}_j - \mathbf{x}_i)$$ Only local information

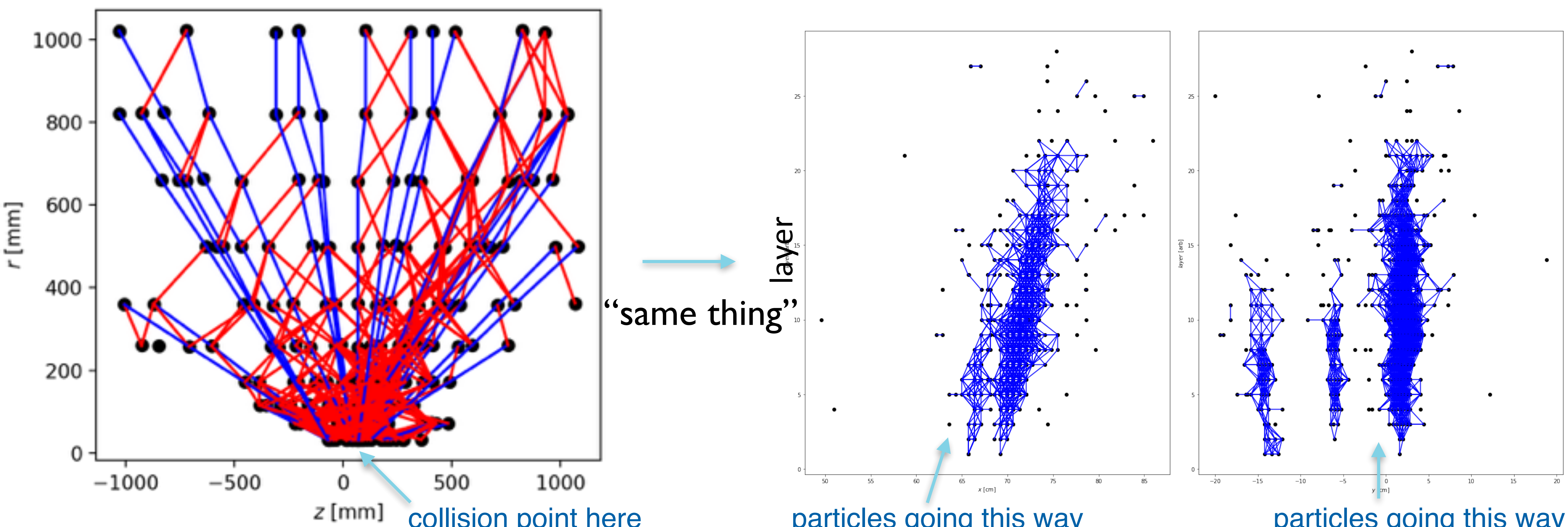
$$h_\Theta(\mathbf{x}_i, \mathbf{x}_j) = \bar{h}_\Theta(\mathbf{x}_i, \mathbf{x}_j - \mathbf{x}_i)$$ Combination of both

- **Dynamic:** Redo kNN after every update
  - The connectivity matrix changes after every update

Looking at graphs on physics detector data

- Tracks and clusters can also be described as connections between points
  - We can then score these relationships between the detector data and select certain associations in the graph that we want to keep.

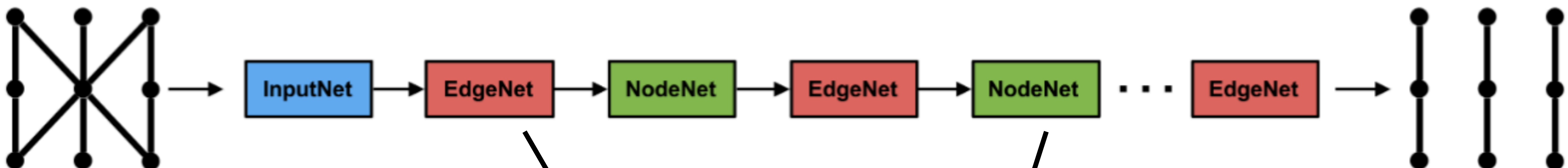
- This results in a useful abstraction: finding points comprising helices in tracks is the same as points in calorimeter clusters
  - Can we simplify our lives and find one algorithm which can handle these different cases?



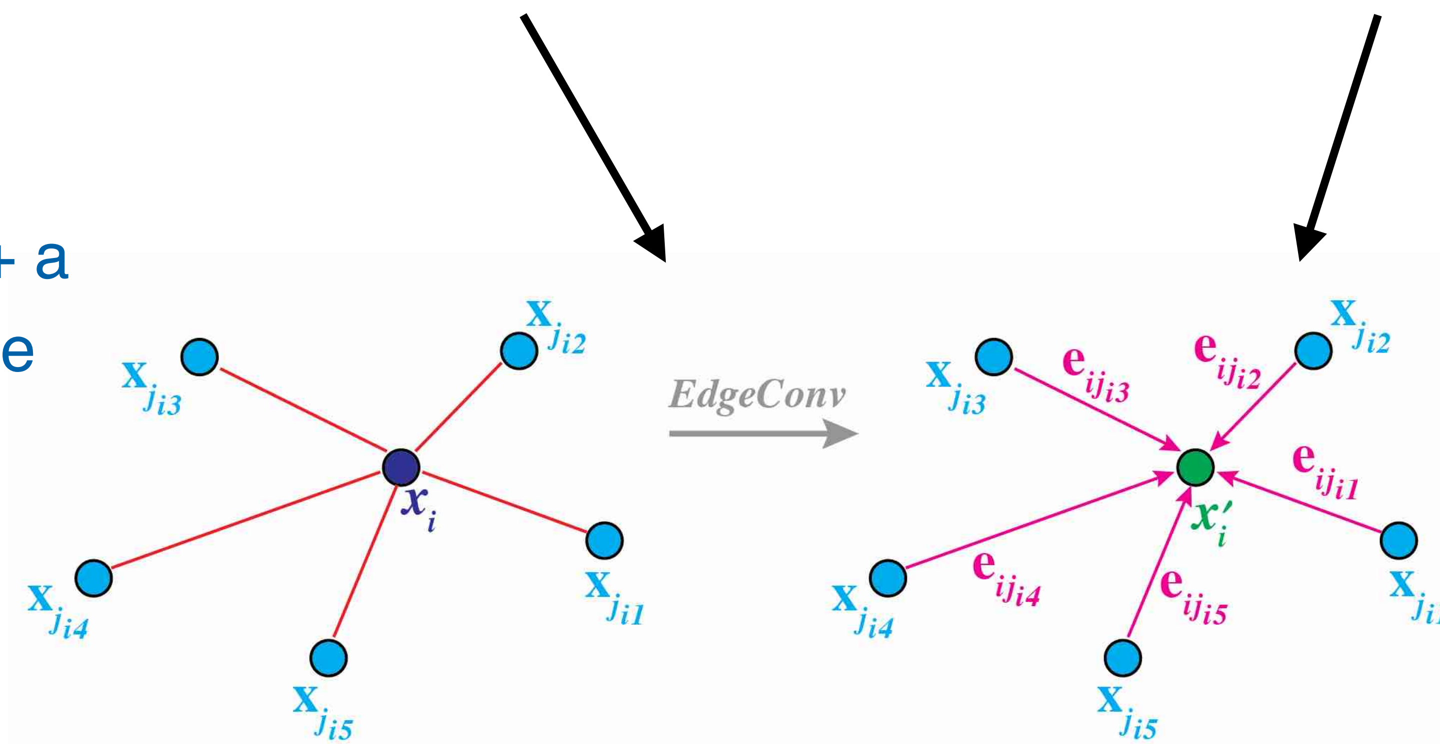
Putting it all together: a model for reconstruction

- With an preliminary model the answer seems to be “yes”
  - So long as we are willing to accept some light post processing

- Basic steps:
  - Define an input graph
  - train an ‘edge classifier’ based on information sharing on that graph
  - Apply edge classification scores to yield a subgraph of just the connections of interest



just raw hit information + a “guess” at the graph

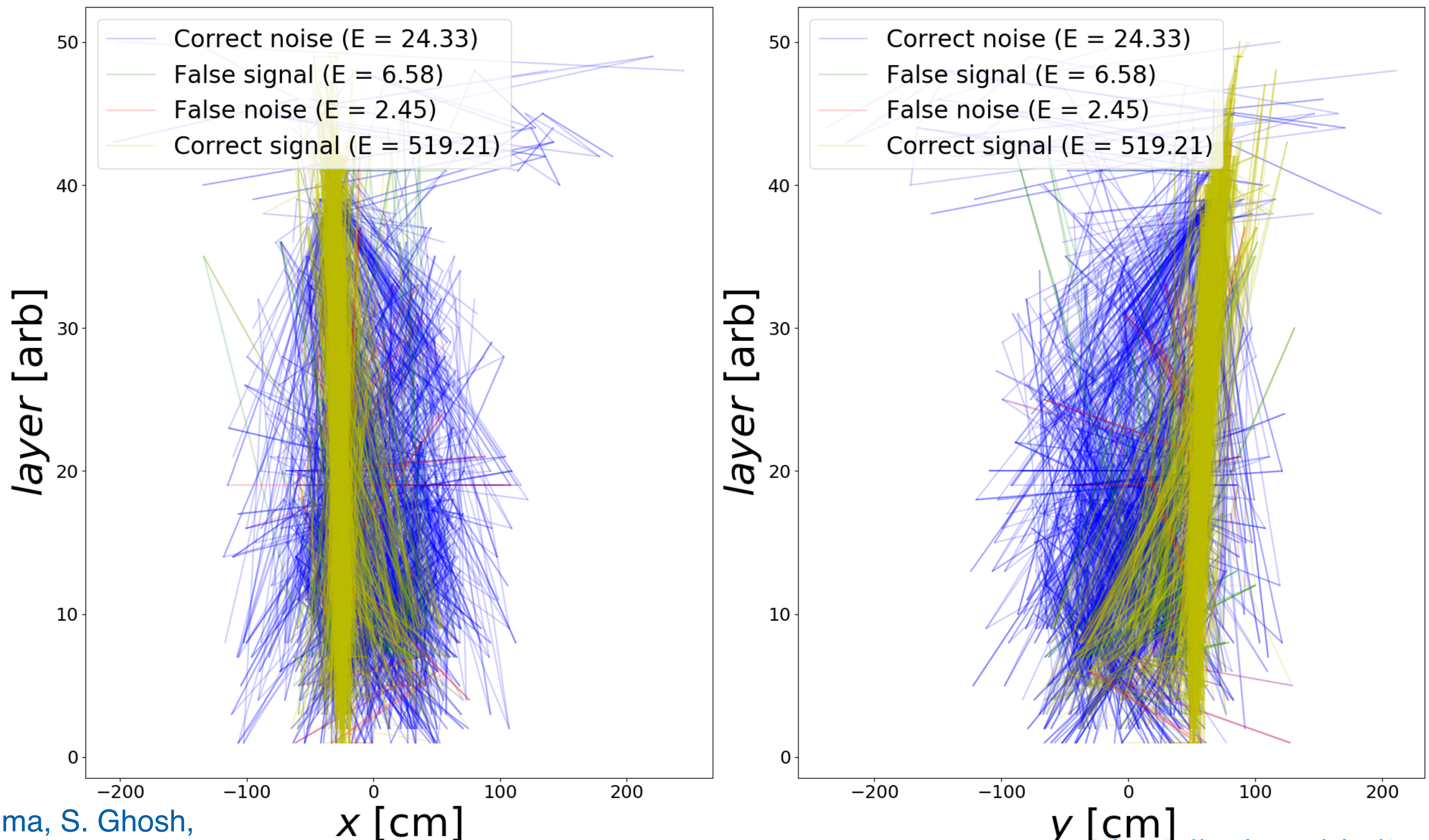


selected links between hits

https://arxiv.org/abs/2003.11603

Reconstruction of a charged pion with edge classification

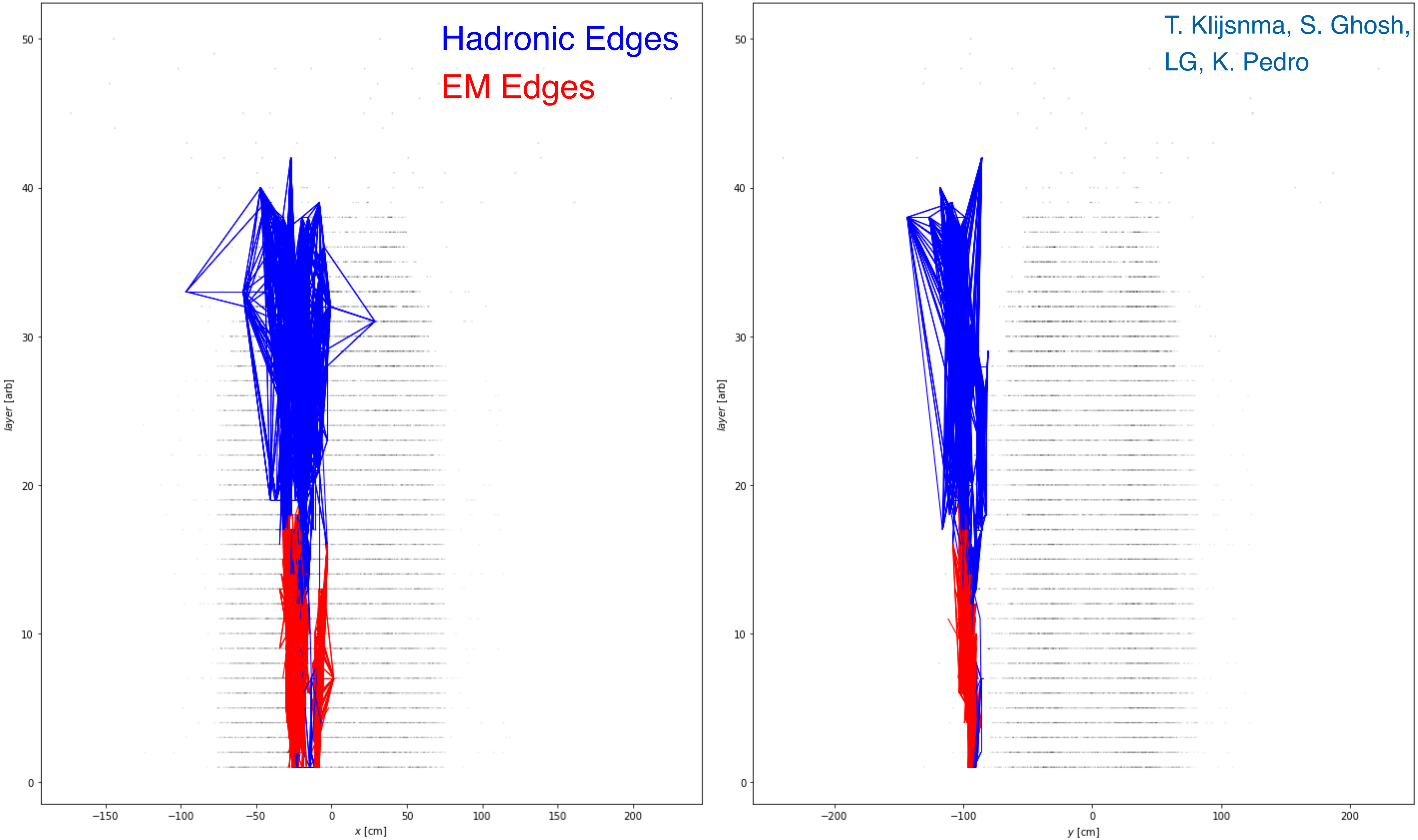
true negatives
true positives
false positives
false negatives



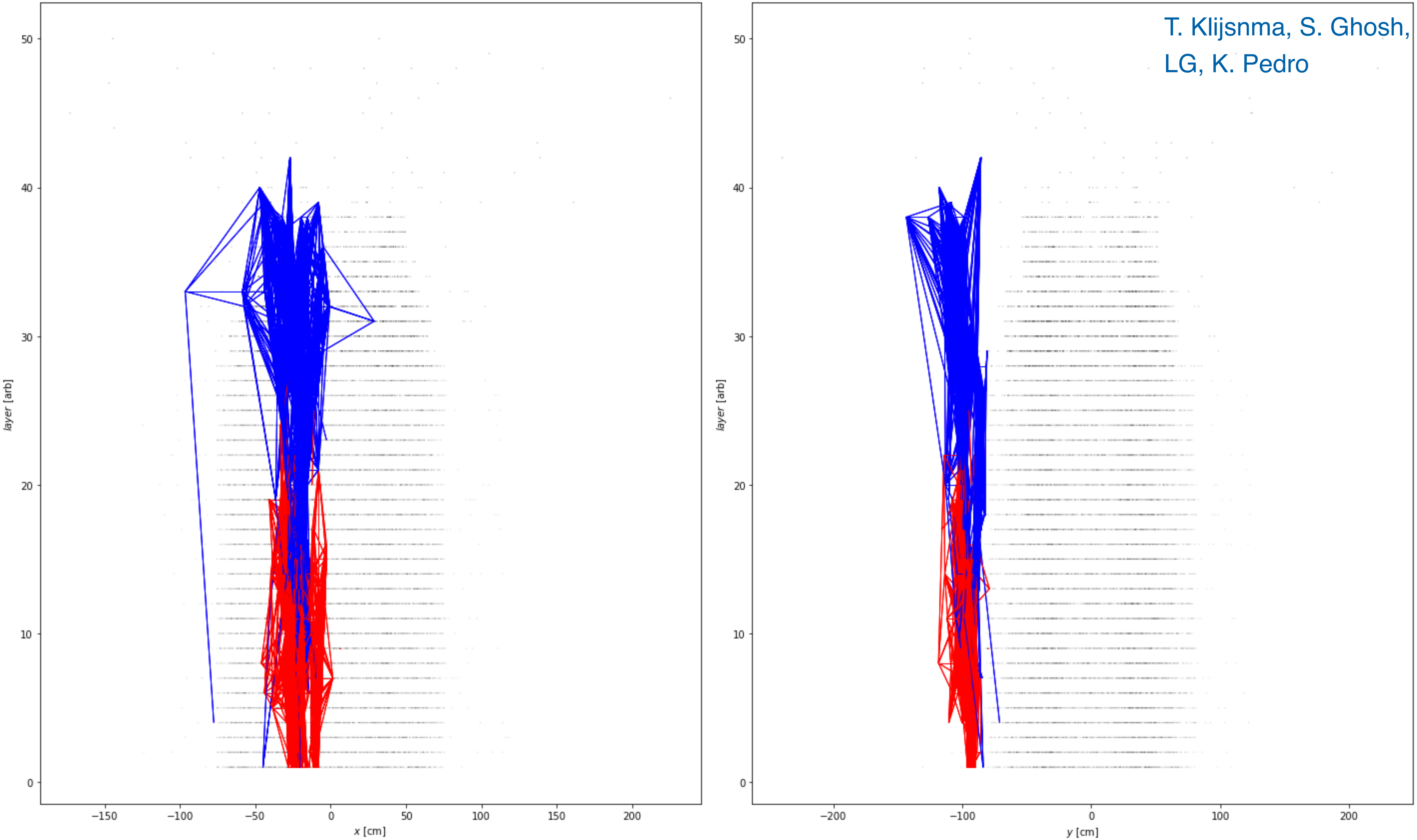
T. Klijsnma, S. Ghosh, LG, K. Pedro

https://arxiv.org/abs/2003.11603

Simultaneous Reco & ID: Tau Lepton Example Prediction



Simultaneous Reco & ID: Tau Lepton Example Truth

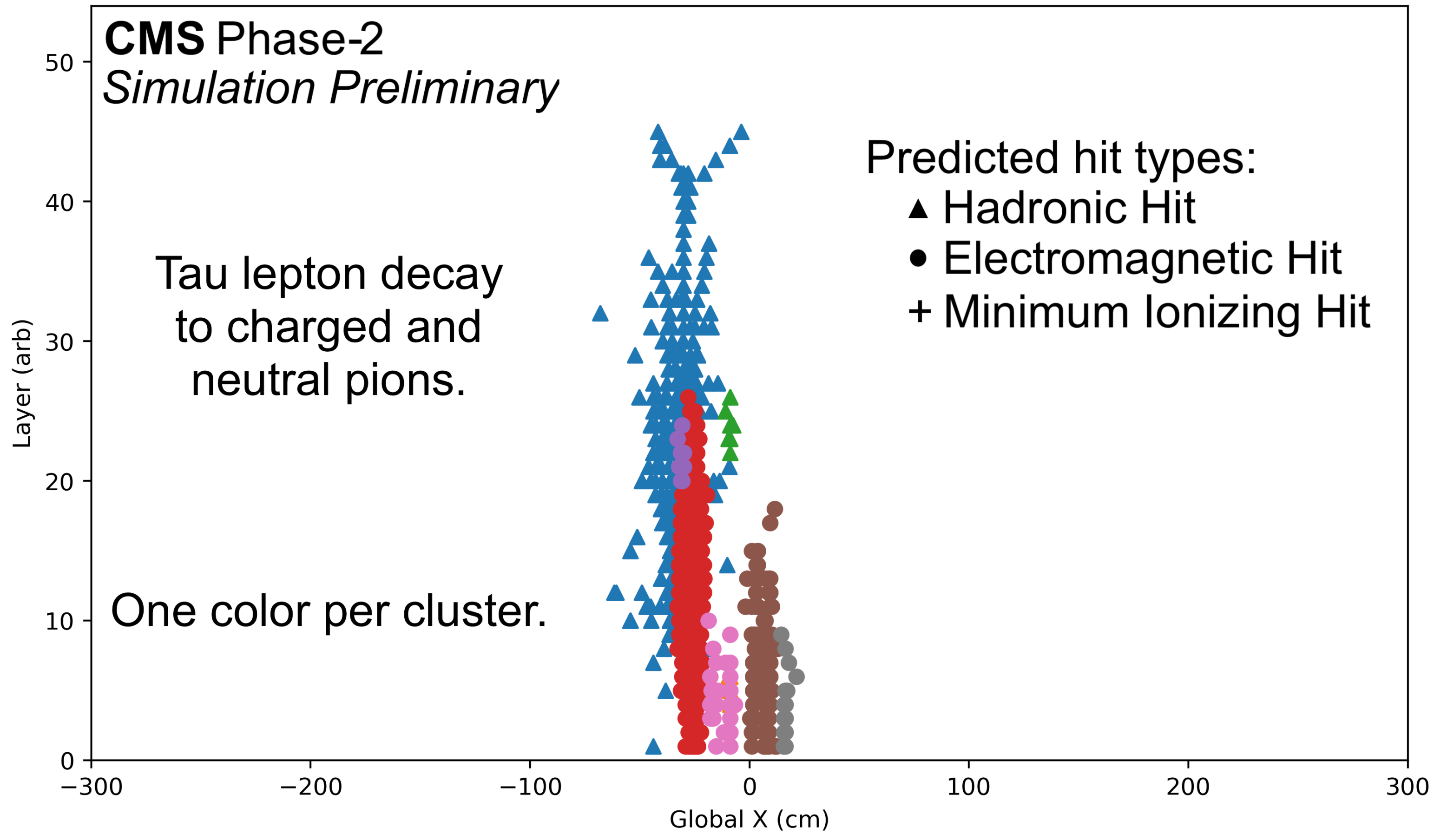


T. Klijsnma, S. Ghosh, LG, K. Pedro

Edge Classification: Making a Clustering (I)

- In order to get calorimeter clusters, need to take the edges and convert to groups of points
  - In this case we just make a union of all the points with common edges of the same type
  - It does a reasonable job already segmenting hadronic energy from electromagnetic
  - We can reconstruct very close-by photons and hadrons effectively

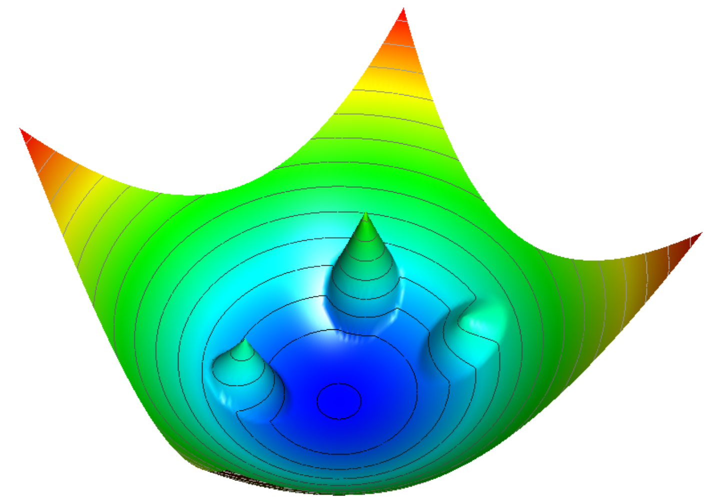
- The same network and processing can also be used on tracking



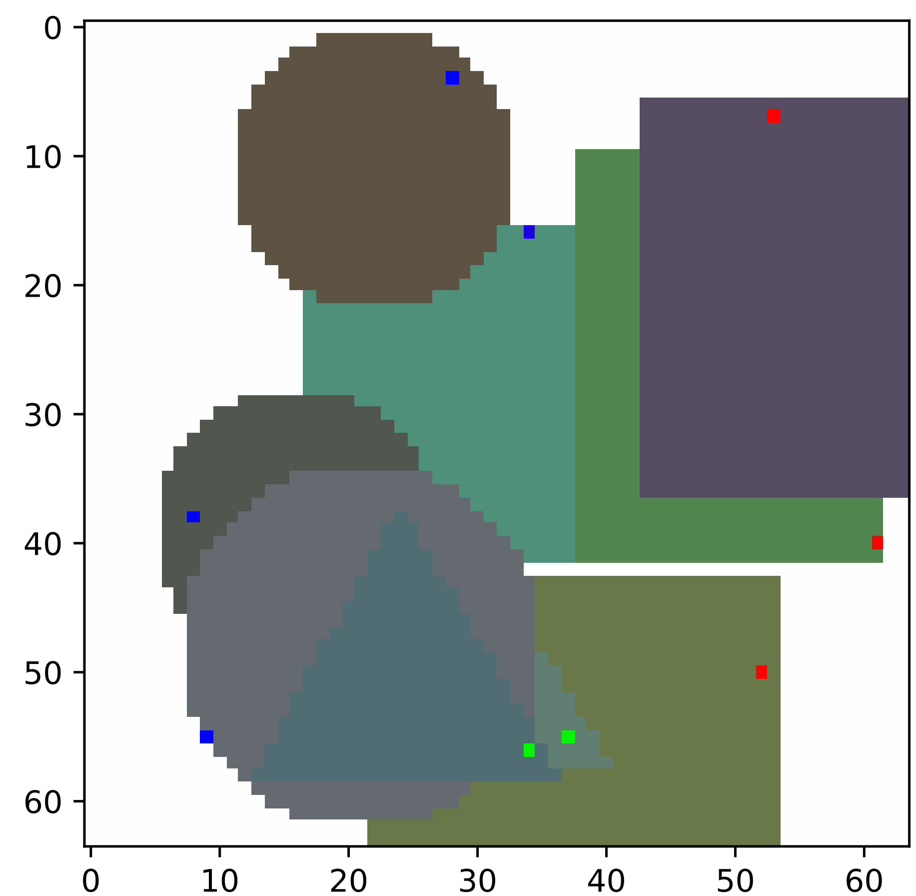
T. Klijnsma, S. Ghosh, LG, K. Pedro

Object Condensation: a loss function for reconstruction

- Physics motivated loss function
  - Potentials with charges
  - like charges attract, opposites repel
  - points that should be associated attract each other
  - variable number of inputs and outputs



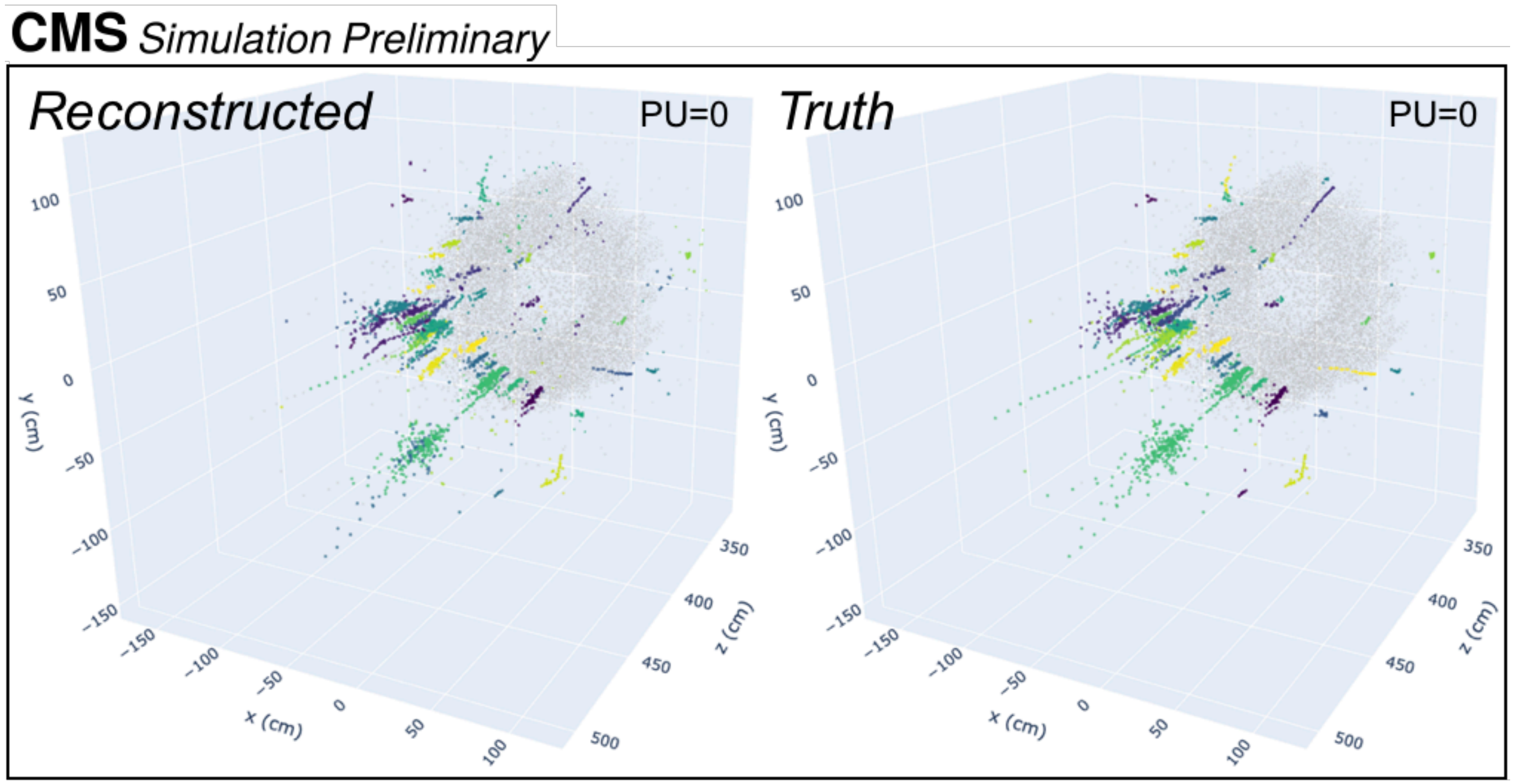
- The network is trained to predict the 'condensation points' of the input data
  - Points within the data that are representative of a whole object

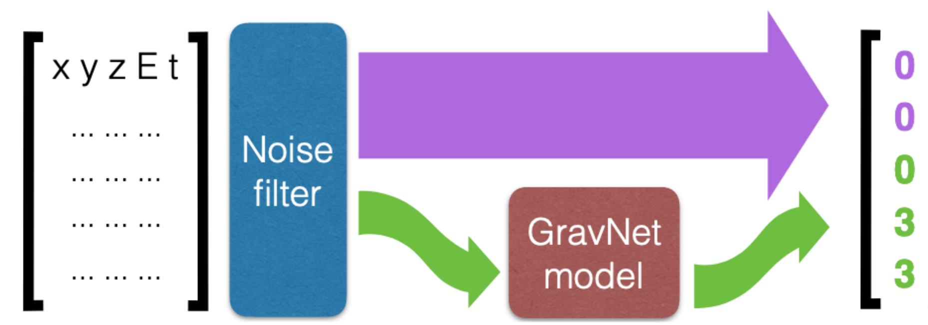


- The condensation points can then be used to collect points around them into 'segmented' objects
  - at this point we have created particles in an event or clusters in a calorimeter

https://arxiv.org/abs/2002.03605

Graph Neural Networks for HGCAL Reco. (LDRD-2019-017)

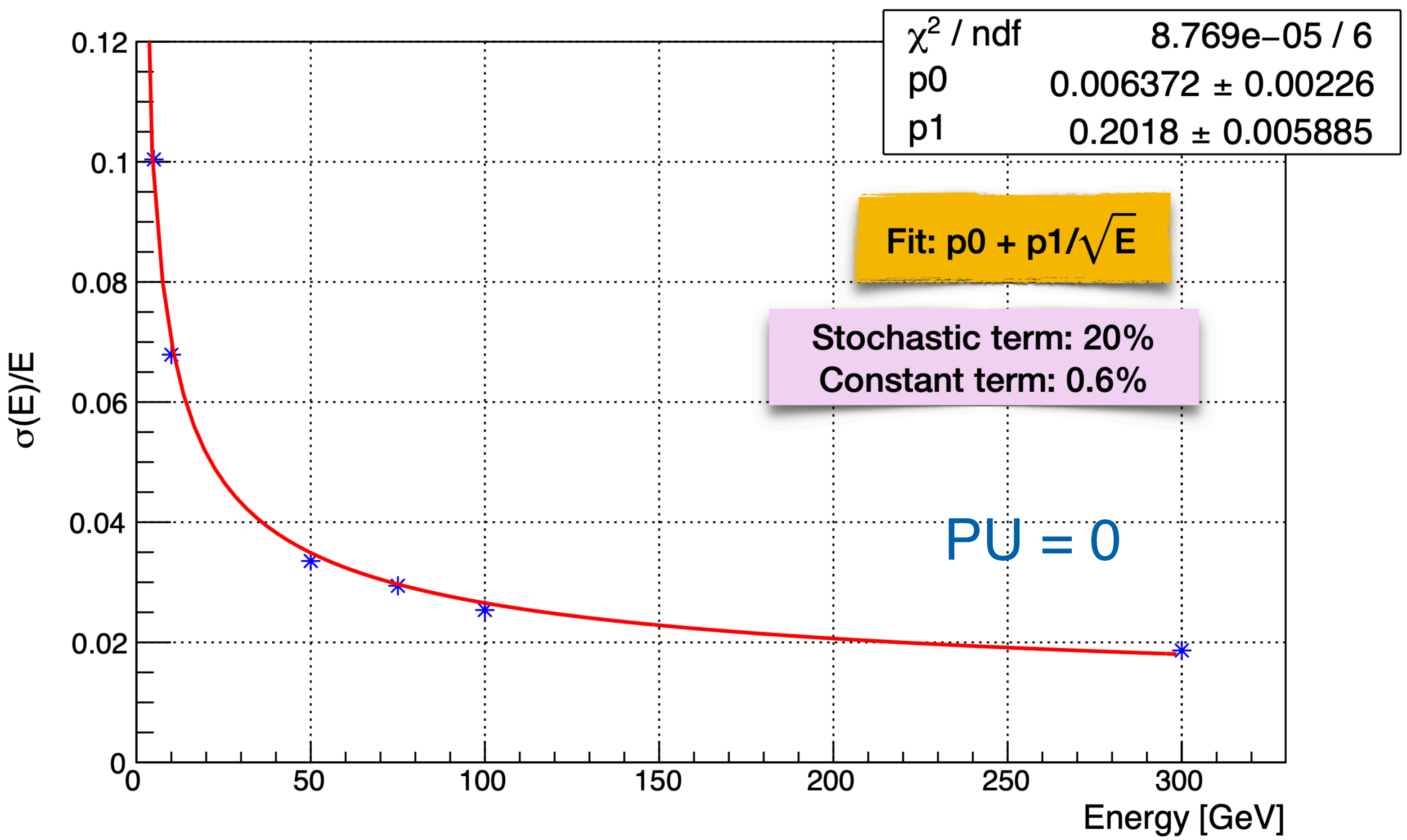




- Combine GravNet with simple noise filter, object condensation loss
  - Heavy collaboration with CERN (Jan Kieseler)
- Train and evaluate on di-tau events, producing locally dense environments
- Only very few over-split hadron showers (ACAT 2021, LDRD completed)
  - Otherwise, excellent separation of showers
  - Also works well in dense environments
  - Can provide fine-grained input to Particle Flow, pileup suppression, and substructure

Single Photon Energy Resolution From Multi-Particle Training

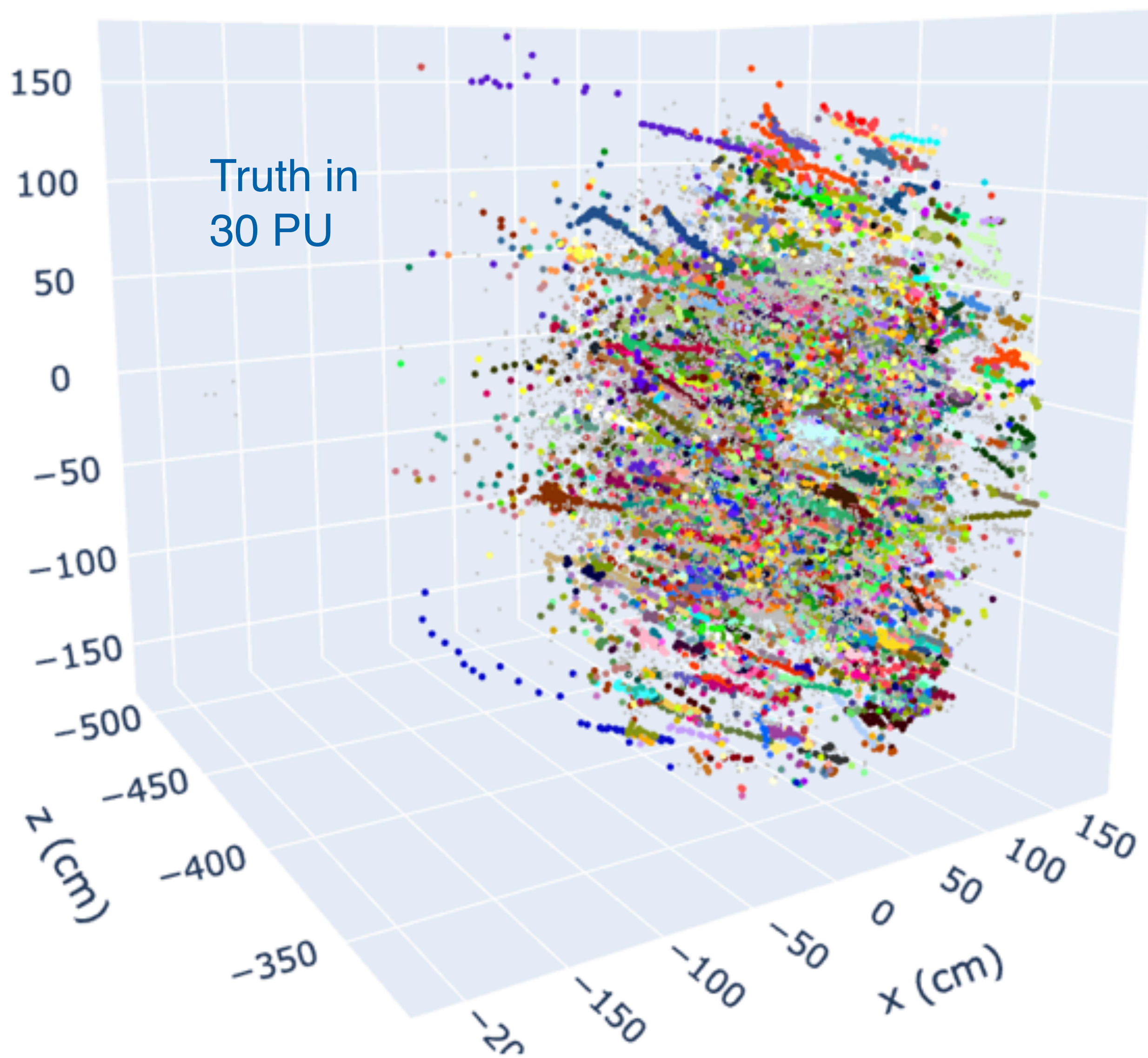
- Multiparticle reconstruction not necessarily guaranteed to have good single particle results
- Photons (in 0 PU) have device limited resolution
  - Nearly perfectly efficient hit collection
- Hadron resolution is still a work in progress (personpower issue)



Fermilab

Present Status: 30 Pileup Training

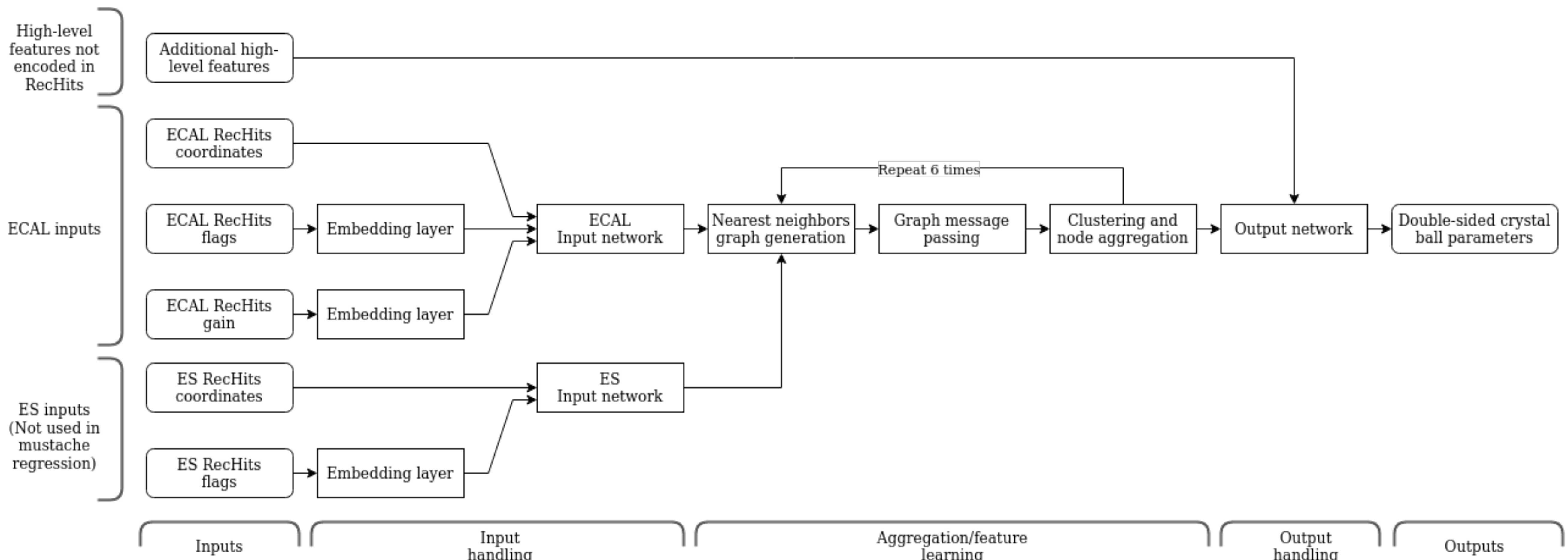
- Truth valid for full events crucial to develop and understand (ML) algorithms
- Recent development: fine ground truth with PU
- PU mixing with FineCalo technical challenge
  - Sufficient simulation information from all (pileup) events needs to be kept long enough to follow decay chain after merging
- Merging algorithms need to be physically meaningful in high PU and IRC safe
  - Developed new hit-by-hit overlap based merging algorithm
  - Consider only the impact region close to boundary for merging metrics



Rotatable plot: https://tklijnsm.web.cern.ch/tklijnsm/hgcal/merging_2022/plots_merging_Aug22/tau_pu30_seed0_n100_1_neg.html

Dynamic Reduction Network for ECAL/HGCAL Regression

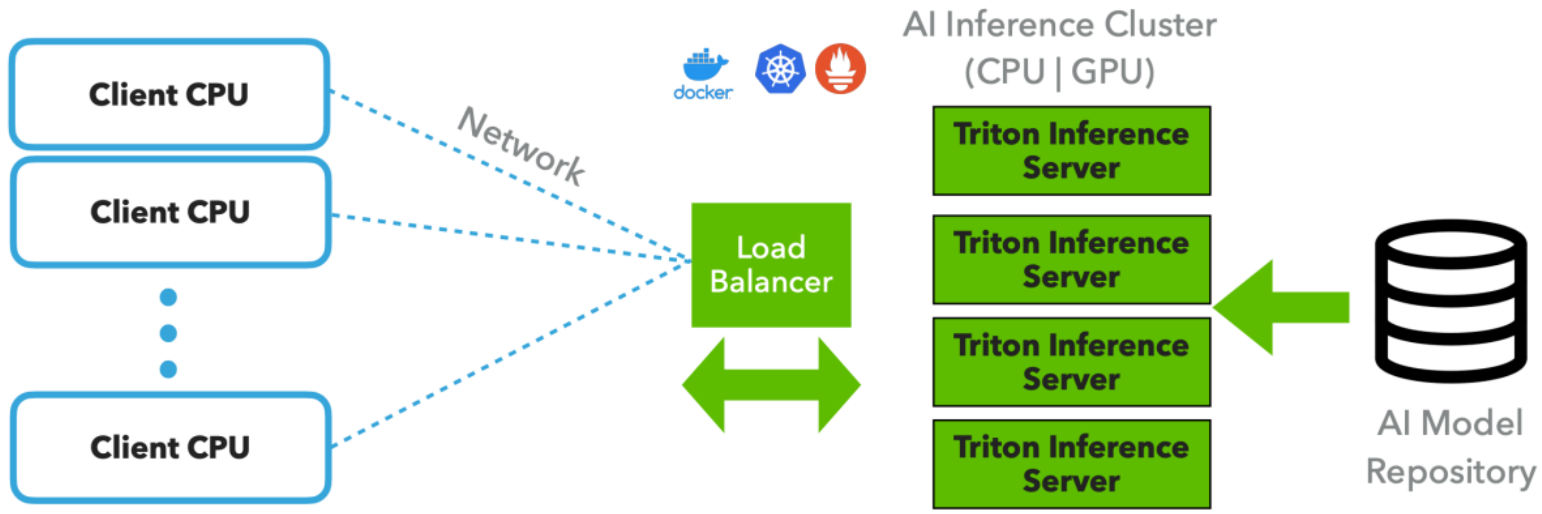
- Building off: https://arxiv.org/abs/2003.08013
- Dynamic GNN that successively clusters representation to a few pieces of high level data
- Spawned two projects collaborating with U. Minnesota
  - ECAL Energy Regression (S. Rothman: MIT)
  - HGCAL Beamtest Regression (A. Alpana: IISER - Pune)
  - Semi-parametric regression (quasi-bayesian methodology for multi-population data)

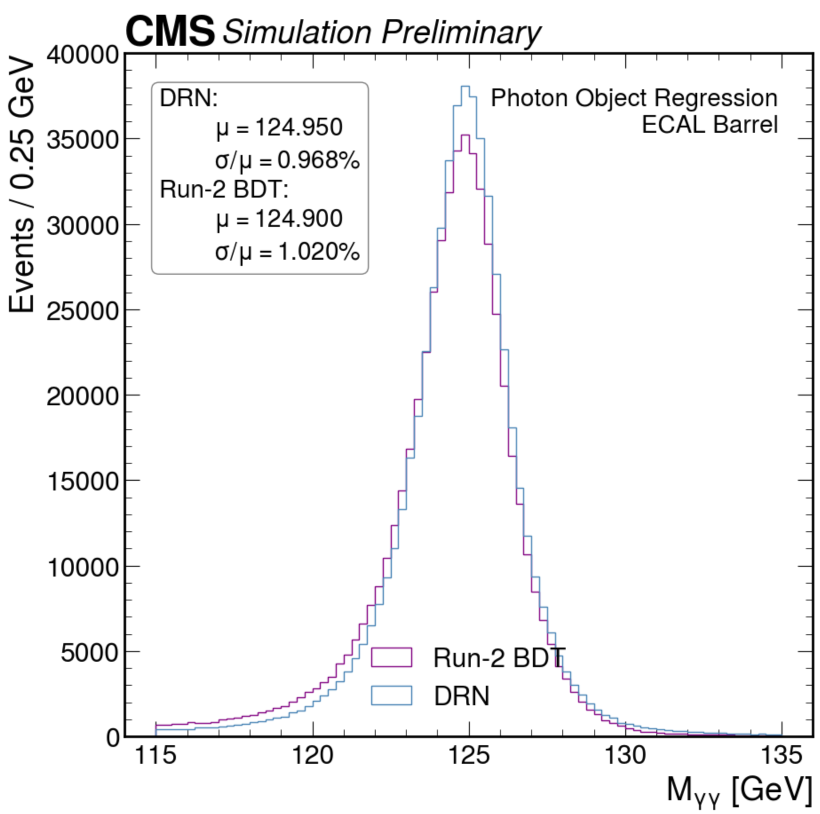


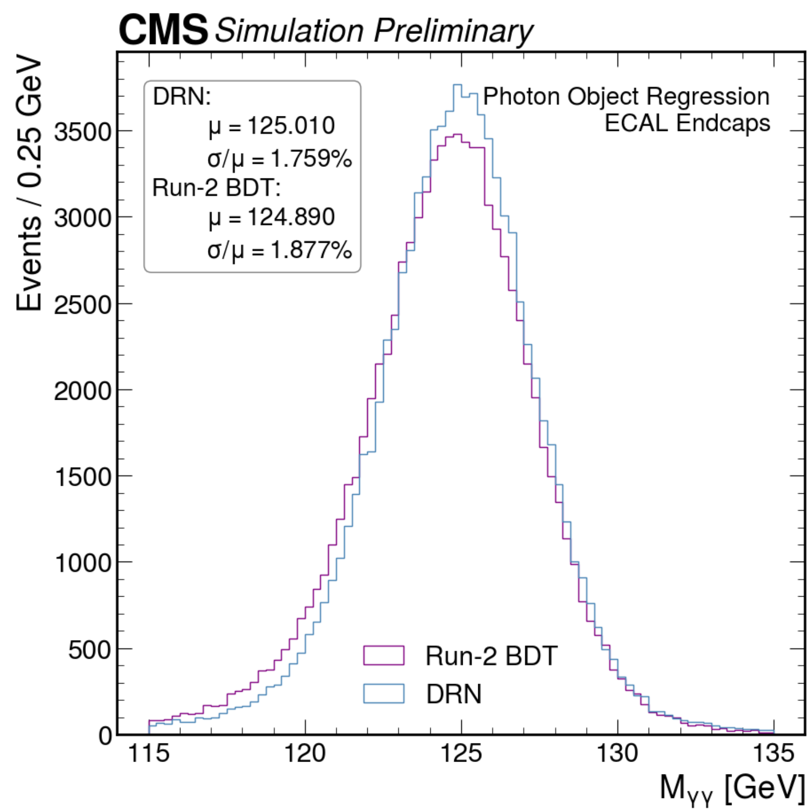
Fermilab

CMS ECAL Energy Regression

- Use DRN + Semi-parametric regression to improve ECAL energy measurement
  - Provide raw hit-lists to network + minor high-level features
- First > 5% improvement in resolution since ~2012
- Foreseen for use in Run3 Hgg and other analysis
- Tight partnership with deployment via SONIC
  - Right now only way to deploy PyTorch in CMSSW
  - Pilot tests show controllable functioning at scale for minimal impact on processing time

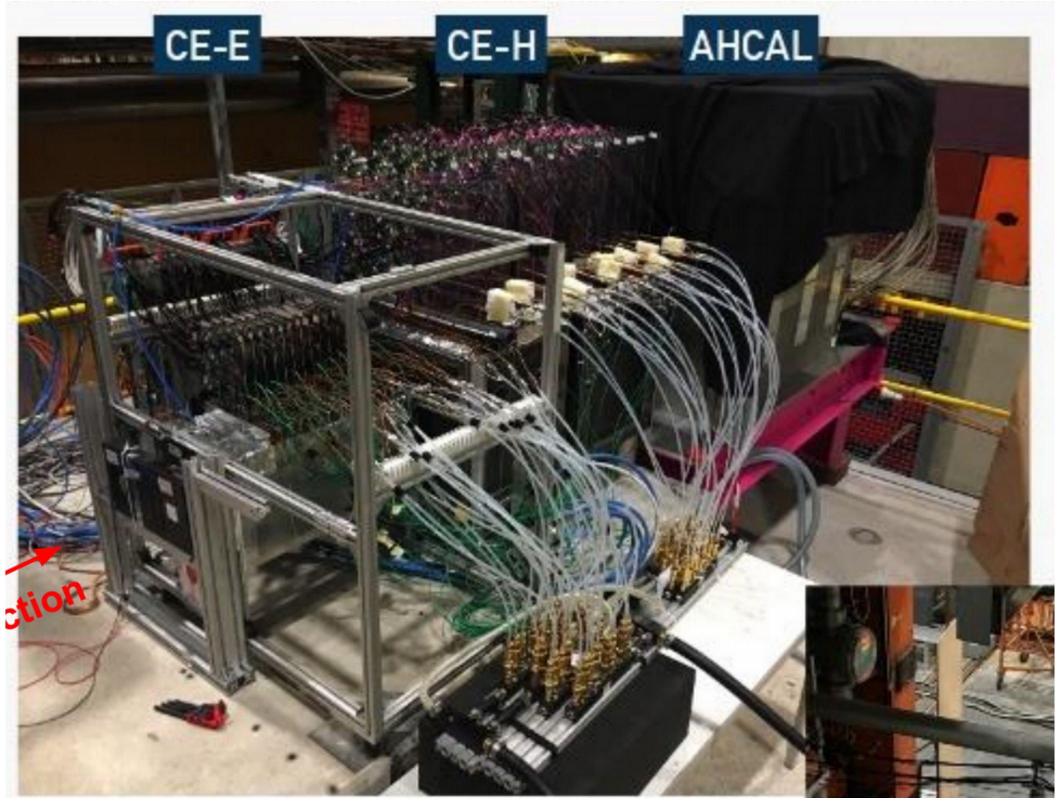


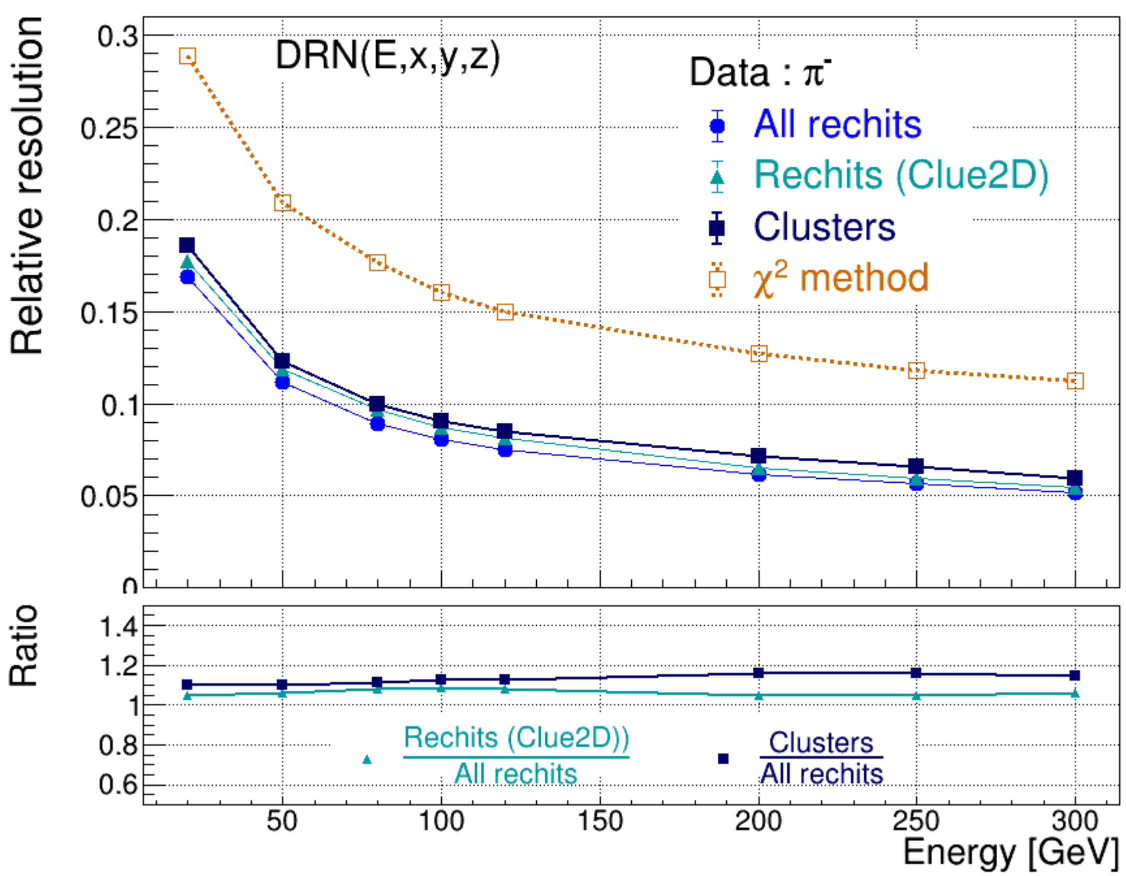


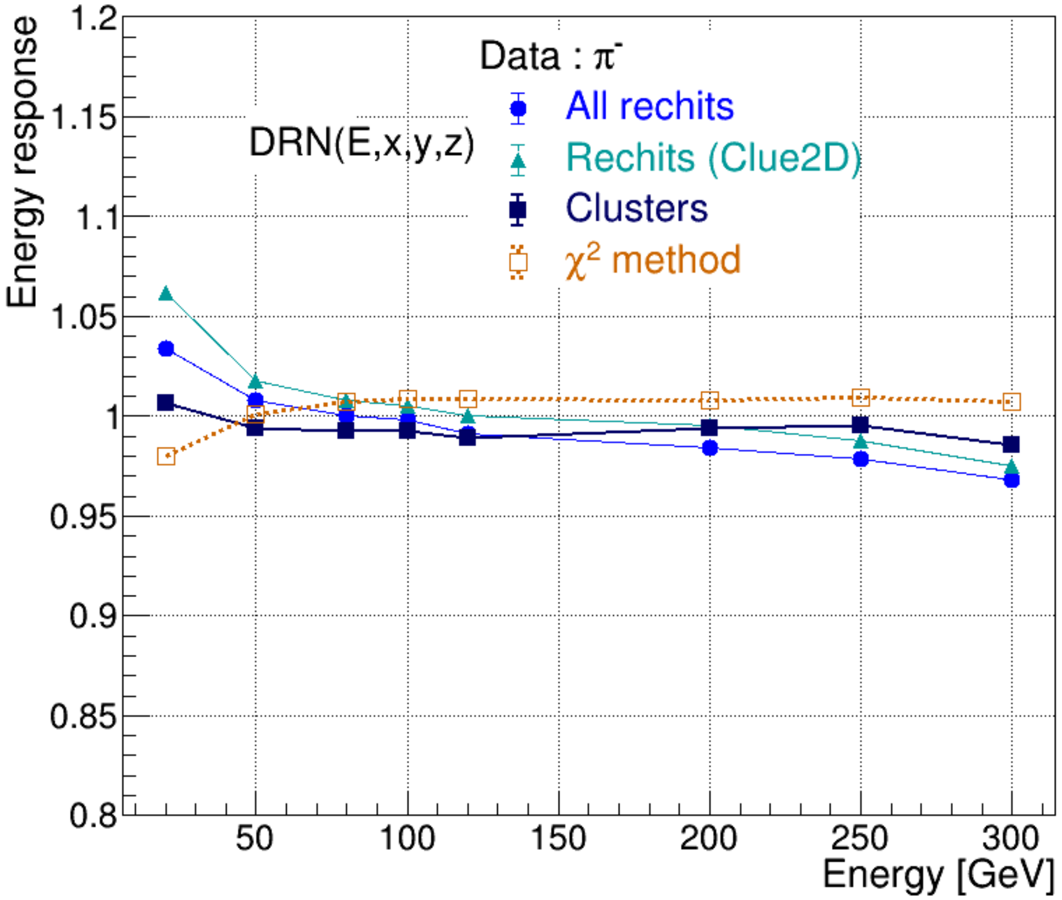


HGCAL Beam Test Regression

- Similar DRN implementation on HGCAL beam test data
- 2x improvement compared to weights-based regression method
- Currently studying impact of real clustering on final energy resolution







Plans

- Continue with HGCAL ML based reconstruction
  - Personpower issues need to be addressed for continuation (Thomas is leaving)
  - Course of work clear, but is intensive, lengthy, technical work
  - Target deployment in CMSSW within a year even in low pileup scenarios
  - Several research direction identified for improving theoretical basis of network
  - Need to study ablation/robustness tests (degrading data in one way or another)

- Finish up ECAL energy regression
  - Paper is in progress, various presentations made (APS, ICHEP)
  - Finalize deployment within CMSSW
  - Next steps - prepare for evolution in Run 3, further hyper parameter optimization

- HGCAL Beam Test energy regression
  - Various presentations already made in several conferences
  - Continuing work on optimizing network and extracting ultimate performance
  - Study effects of pile-up like scenarios (reduced energy collection region, etc.)

Extras