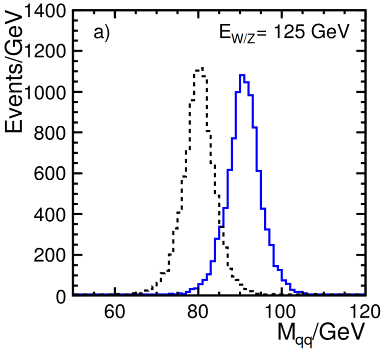
Shower Separation using Machine Learning Techniques.

Jack Rolph
University of Hamburg

November 10, 2020

The need for sophisticated hadron shower clustering.

- **International Large Detector (ILD)** required to **distinguish hadronic decays of W and Z bosons**;
- **Pandora Particle Flow** is current state-of-the-art for W-Z **jet energy resolution** ($\sigma_E/E = 3.8\%$);
- Pandora PFA [3] relies upon **sophisticated clustering for particle showers in highly granular calorimeters.**
- Machine Learning is **a rapidly developing science**, with **many state-of-the-art applications in clustering.**
- Can **machine learning** be used to aid in **PFA clustering**?
- Does a **temporal calorimeter** aid in clustering?



Reconstructed invariant mass distributions for the hadronic system in simulated $ZZ \to d\bar{d}\nu\bar{\nu}$ and $W^+W^- \to u\bar{d}\mu^-\bar{\nu}$ [3]

1

[1] M.A. Thomson. "Particle flow calorimetry and the PandoraPFA algorithm". In: (2009).

Goals.

Predict the fraction of energy belonging to two hadronic showers observed in AHCAL, cell by cell, using existing state-of-the art machine learning methods.

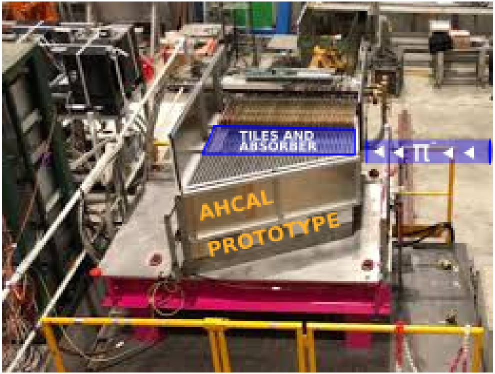
- State of the art machine learning uses **graph networks** to achieve separation.
- Several options for shower separation discussed in paper:
  - **Standard Convolutional Neutral Network**;
  - **Dynamic Graph Convolutional Neutral Network** (DGCNN)[1]
  - **GravNet**[2] ;
- Is time a useful variable for hadronic shower clustering?
- What is the effect of time resolution on network performance?

---

[1] Yue Wang et al. *Dynamic Graph CNN for Learning on Point Clouds.* 2018. arXiv: 1801.07829.

[2] Shah Rukh et al Qasim. "Learning representations of irregular particle-detector geometry with distance-weighted graph networks". In: (2019). ISSN: 1434-6052.

Simulation Information: Calorimeter Design



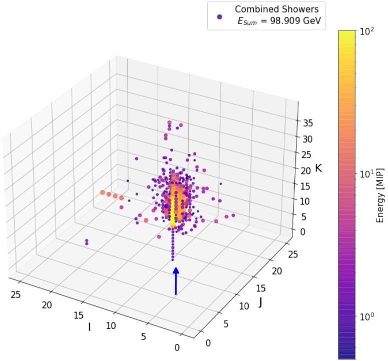
Picture of the AHCAL at Testbeam

- Dimensions: 72 cm $\times$ 72 cm $\times$ 75 cm
- Absorber: Stainless Steel
- Depth: $\sim 4\ \lambda_I$ over 38 Layers
- Regular cell geometry;
- Total Channels: 21,888.

Simulation Information: Summary.

Simulation of π^- hadronic showers **using Geant4** in the AHCAL were used:

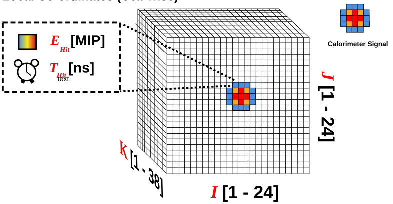
- **full detector simulation** (inc. SiPM saturation/noise thresholds etc.)
- Physics list: **QGSB_BERT**
- Based on **June 2018 CALICE Testbeam** taken at SPS;
- Cuts:
  - **Punch-through Pion Removal**: $N_{Hits} > 50$
- Simulated particle energies: 10, 20, 30, 40, 50, 60, 70, 80 GeV



Example event display of a 80 GeV negative pion detected by the AHCAL

Typical Detector Observables.

Local Co-ordinates (Cell-wise)

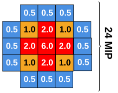


Global Shower Variables:

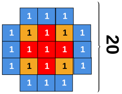
$E_{Reco} = \sum_{i=0}^{N} E_{Hit}$ [MIP]

$N_{Hits} = |\,Hit\ Cells\,|$ [Count]

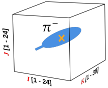
$CoG_k = \frac{\sum_{i=0}^{N_{Hits}} E_{Hit,i}\, x_{ik}}{\sum_{i=1}^{N_{Hits}} E_{Hit,i}}$



Total Reconstructed Energy



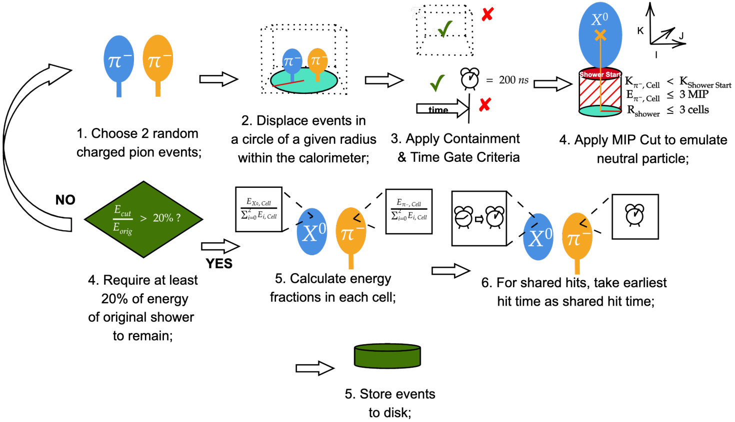
Total Number of Active Cells



Center of Gravity
(Energy-Weighted Mean Position of Shower)

Observables measured by the CALICE AHCAL.

Simulation: Creating a Multi-Shower Dataset.



Outline of the algorithm chosen to create augmented data

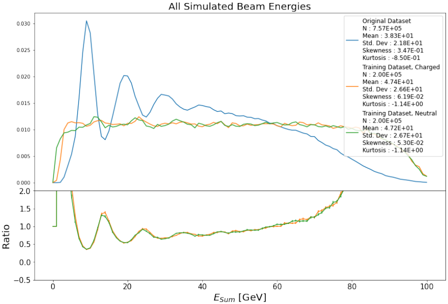
Simulation: Some Strategies for Data Augmentation.

- **Extreme care** was taken to **bias dataset 'intelligently'**;
- **In particular**:
  - **Shower energy**: **Intrinsic calorimeter resolution** means **network will learn total shower energy distribution**.
  - **Average initial separation**: If showers not placed **'reasonable' distance apart**, the network will not experience a **wide variety of shower separation cases**.

Jack's 'intelligent' biases

Simulation: Some Strategies for Data Augmentation.

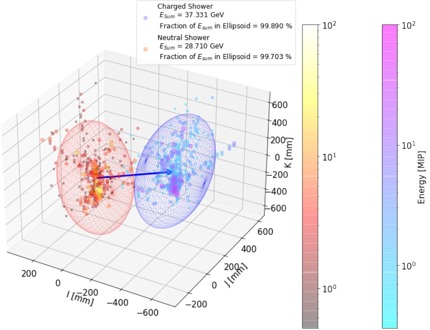
- Random selection of shower energy distribution weighted to be isotropic at the level of reconstructed energy.



Entire combined shower energy distribution of the simulation, before and after augmentation.

Simulation: Some Strategies for Data Augmentation.

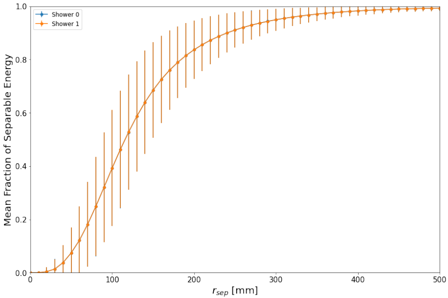
- **Energy-weighted eigen-ellipsoids** calculated for training set' **using PCA**;
- **Ellipsoids** define '**shower volume**' on a shower-by-shower basis, **based on statistics**;
- Using ellipsoid solving algorithm, calculate: **Distance at which two showers must be separated in order that the ellipsoids are just touching i.e. energy contained separable by a plane.**



Two separated 'shower volumes', separable by a plane.

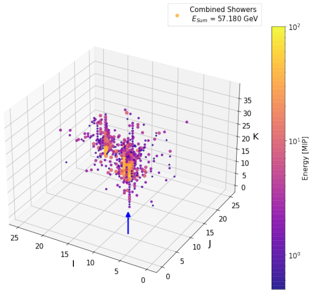
Simulation: Some Strategies for Data Augmentation.

- r_{sep} = 200mm chosen;
- Corresponds to $R = 7$ calorimeter cells;
- Reasoning for study obvious:
  - at r_{sep} = **100mm**, most of training dataset **unable to be intrinsically resolved**;
  - at r_{sep} = **300mm**, most of training dataset **resolvable with a linear discriminant.**

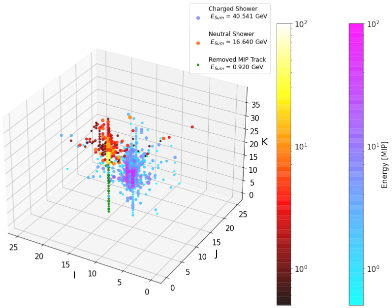


Combined multi-shower event.

Simulation: Example Event Display

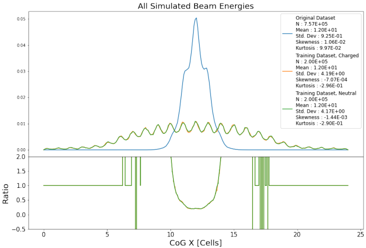


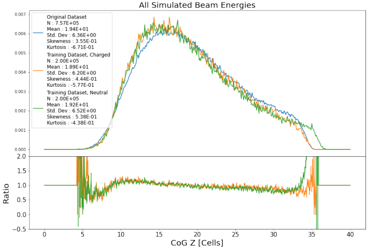
Combined multi-shower event.

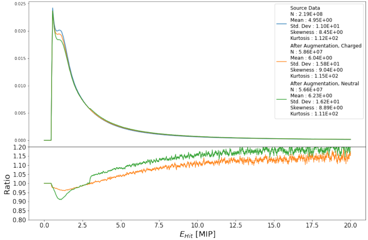


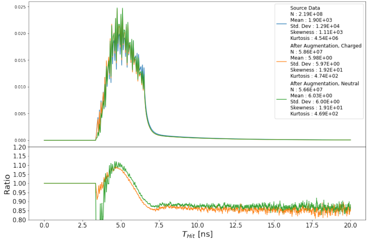
Recovered energy fractions and removed MIP Track

Simulation: Some Summary Checkplots









Network: Figures of Merit and Hyperparameters.

All figures of merit and choice of hyperparameters used were defined in the reference paper [2].

Figures of merit:

- Loss Function:

$$L = \sum_k \frac{\sum_i \sqrt{E_i t_{ik}} \, (p_{ik} - t_{ik})^2}{\sum_i \sqrt{E_i t_{ik}}} \qquad (1)$$

Mean-square error, weighted with the square-root of the true cell energy.

- Accuracy Function:

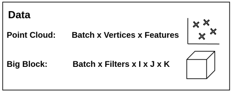
$$A = \frac{N_{Events}(0.7 < \frac{E_{pred}}{E_{true}} \leq 1.3)}{N_{Events}} \qquad (2)$$

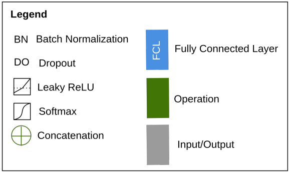
Ratio of number of charged particles with 70%-130% of their true, reconstructed energy predicted to all charged particle

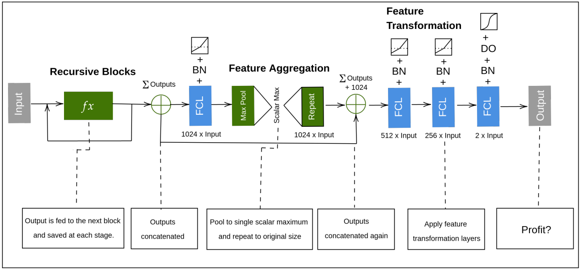
- **Batch Size**: 20 events
- **Total** $N_{\mathbf{epochs}}$: 20
- **Training Size**: 1×10^5 Events
- **Test Size**: 1×10^4 Events (10 % of Training Size)
- **Validation Size**: 3×10^4 Events (30 % of Training Size)
- **Optimizer**: ADAM
- **Learning Rate**: 3×10^{-4}
- **Scheduler**: Exponential Decay, Factor = 0.99
- **Resources**: Nvidia P100

$E \longrightarrow$ Ground Truth Energy [MIP];
$t \longrightarrow$ Ground Truth Fraction;
$p \longrightarrow$ Predicted Fraction;
$i \longrightarrow$ Index of Energy Deposition;
$k \longrightarrow$ Index of Shower;

Network: Main Architecture

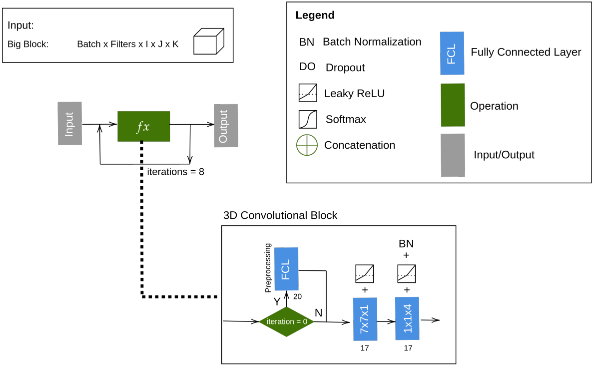






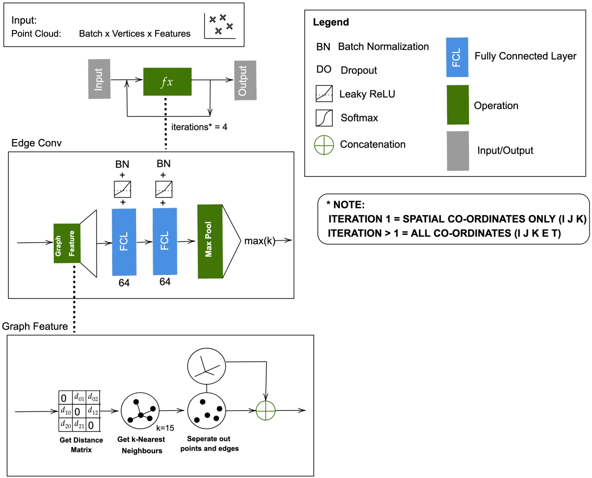
Visual depiction of the neural network inspired by the 'segmentation network' in *“ Dynamic Graph CNN for Learning on Point Clouds"* [1]

Network: Standard Convolutional Block



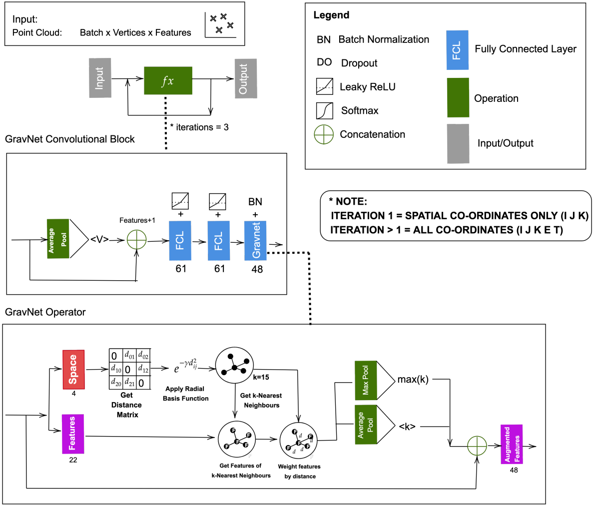
Convolutional block architecture. Complete network has 991,679 learnable weights

Network: Dynamic Graph Convolutional Block



Dynamic Graph Convolutional block architecture. Complete network has 977,024 learnable weights.

Network: GravNet Block

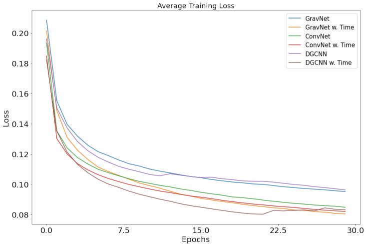


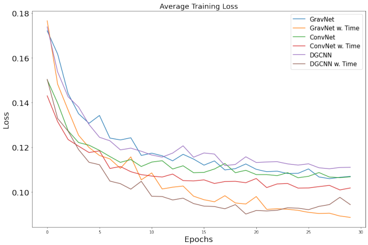
GravNet Convolutional block architecture. Complete network has 980,042 learnable weights.

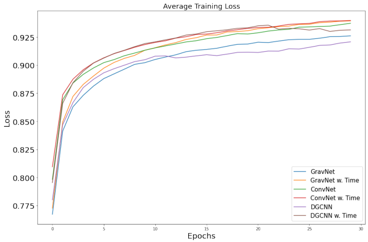
The Experiment.

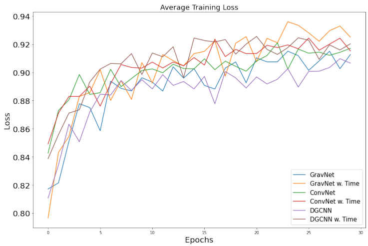
- **Train** a series of **shower separation networks**, **with** and **without time** as an input variable;
  - Is there an **improvement** in **energy resolution**?
  - Is there a **correlation** between the **shower fractions and energy**?
- Obtain **samples** of charged-neutral hadronic **shower pairs** with **decreasing time resolution** from 0ns to 2ns;
  - How does **network performance** change as **time resolution degrades** over this range?

Results: Loss and Accuracy Curves.

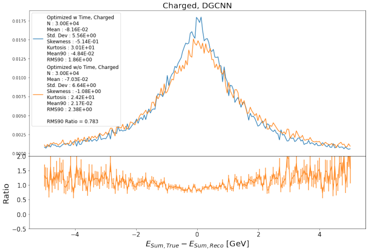


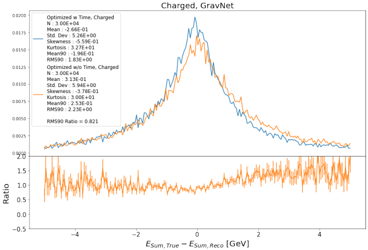


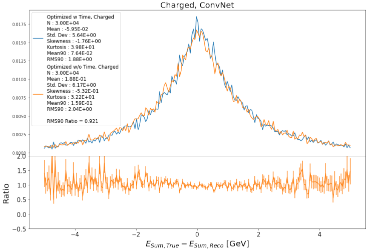




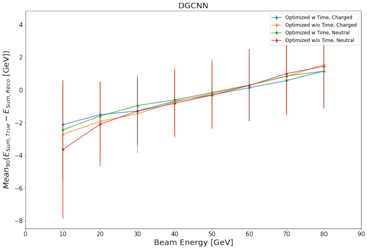
Results: Overall Energy Reco Performance

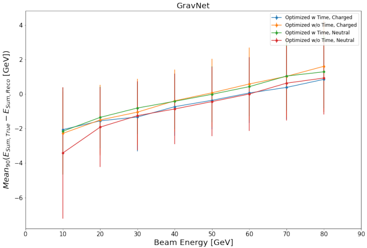


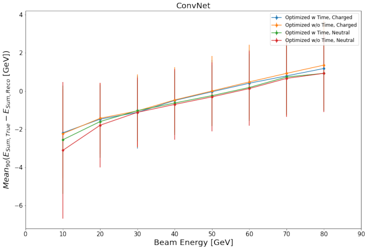




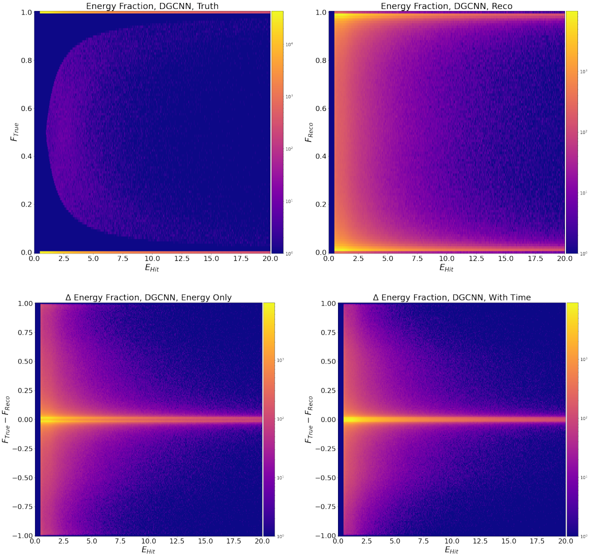
Results: Energy Reco Performance vs. Beam Energy.



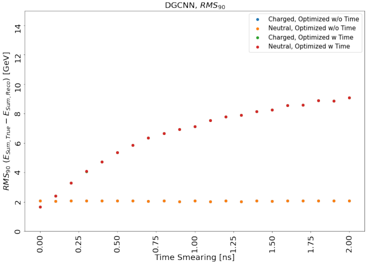


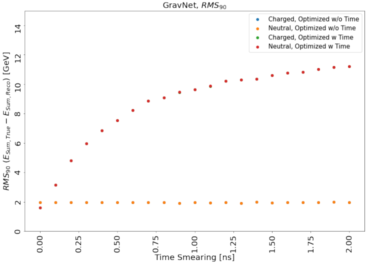


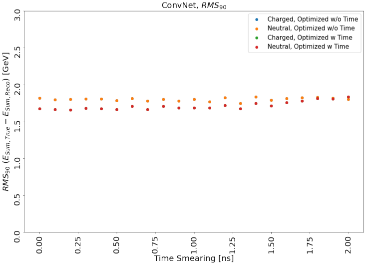
Results: DGCNN, Fraction vs Energy.



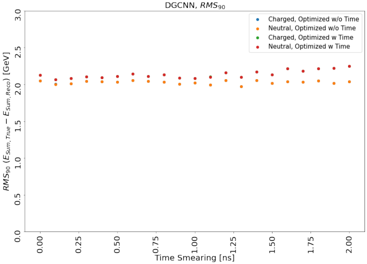
Results: Perfect Time Resolution Networks vs. ↓ Time Res.

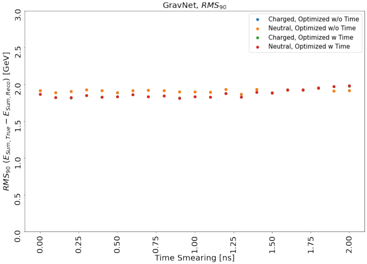


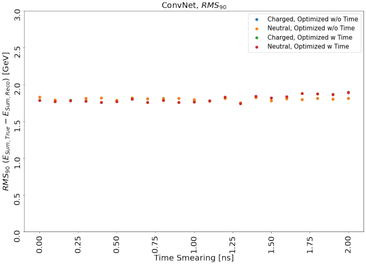




Results: 1ns Time Resolution Networks vs. ↓ Time Res.



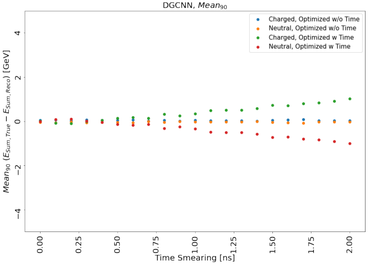


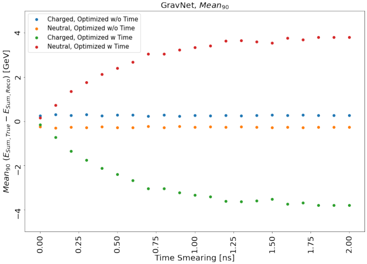


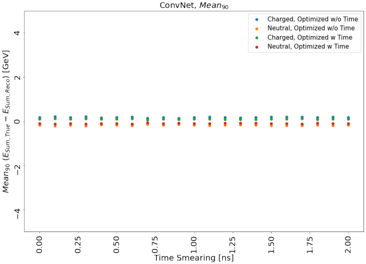
Conclusion.

- An **efficient multi-shower simulation tool** was developed;
- **Small improvement** in **clustering performance** using **perfect time resolution**:
  - maximum relative factor of $\sim$ 20%'
  - maximum absolute improvement in resolution of $\sim$ 500 MeV.
- **DGCNN Network** is **able to learn fraction energy correlations**, but **not for low energy hits**;
- **Improvement with graph networks due to time** is **highly dependent on training distribution.**
- Tentatively, **improvement due to time is no longer useful** after $\sim$ **1.5ns time smearing**;
- Overall, is small clustering improvement even worth having time as a calorimeter variable?

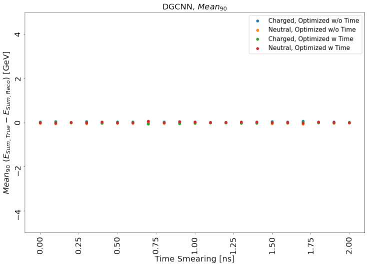
Results: Perfect Time Resolution Networks vs. ↓ Time Res.

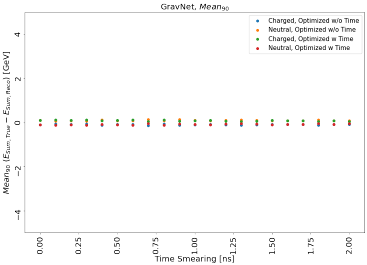


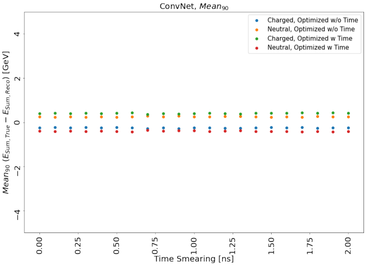




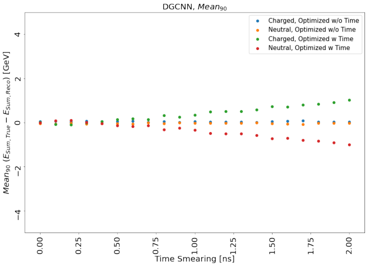
Results: 1ns Time Resolution Networks vs. ↓ Time Res.

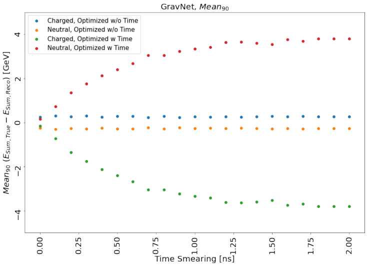


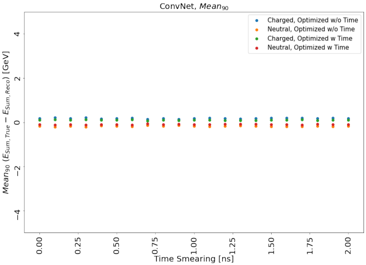




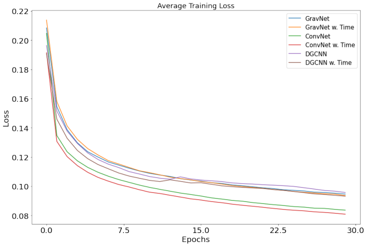
Results: Perfect Time Resolution Networks vs. ↓ Time Res.

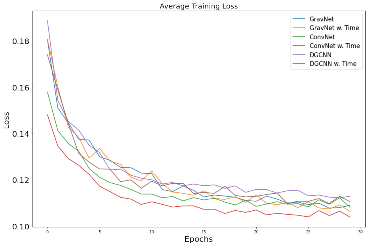


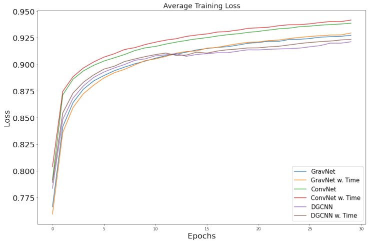


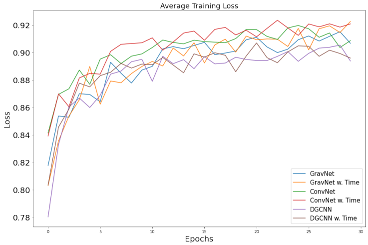


Results: Loss and Accuracy Curves, 1ns Time Res.

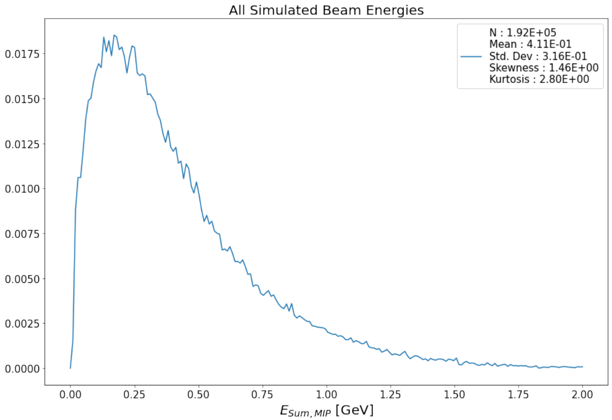








Simulation: Cut MIP Energy Spectrum.



Total energy spectrum for sum of cut MIP hits, event by event. Theoretical value for Bethe-Bloch energy loss in iron should be $\sim$ 0.18 GeV