Shower Separation for Highly Granular Calorimeters using Machine Learning

Jack Rolph, Erika Garutti

Institute for Experimental Physics, University of Hamburg

High-D Consortium Meeting

10/02/2022

Particle Flow Calorimetry

- Jet energy resolution at future precision e^+-e^- colliders such as ILC must produce a di-jet invariant mass resolution comparable to the weak vector bosons' decay width (~3% in the range of jet energies 50-200 GeV) [1]

- **Problem:** typical jet energy resolution of 'traditional' calorimetry is much worse than required at ILC.

- **Solution:** Particle Flow Calorimetry (PFC) [2]:

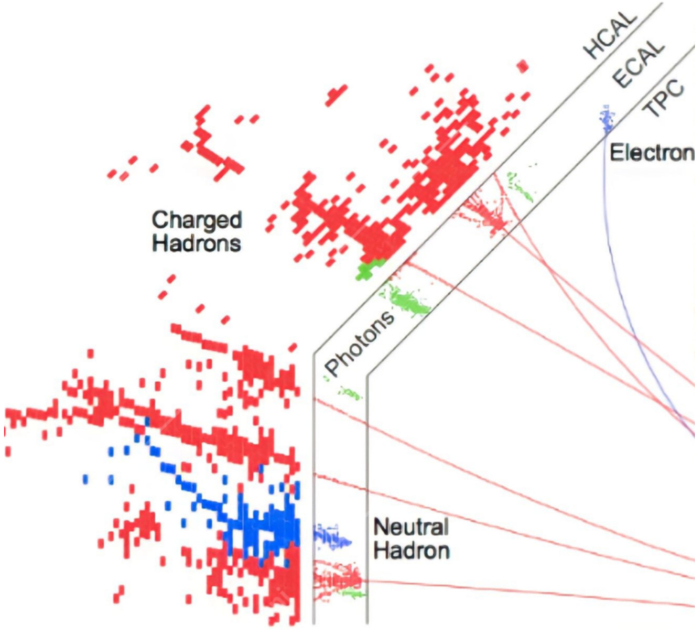
  - measure momentum of charged particles (~60% of jet energy) using tracker;

$$\frac{\sigma_{E_{X^\pm}}}{E_{X^\pm}} \simeq \frac{10^{-4}}{E_{X^\pm}}$$

  - use highly granular calorimeters to measure remainder of energy of photons and neural hadrons;

$$\frac{\sigma_{E_{h_0}}}{E_{h_0}} \simeq \frac{0.55}{\sqrt{E_{h_0}}} \qquad \frac{\sigma_{E_\gamma}}{E_\gamma} \lesssim \frac{0.15}{\sqrt{E_\gamma}}$$

  - Use sophisticated clustering algorithms (e.g. Pandora Particle Flow Algorithm, PPFA) to associate tracks to energy depositions in the highly granular calorimeters;

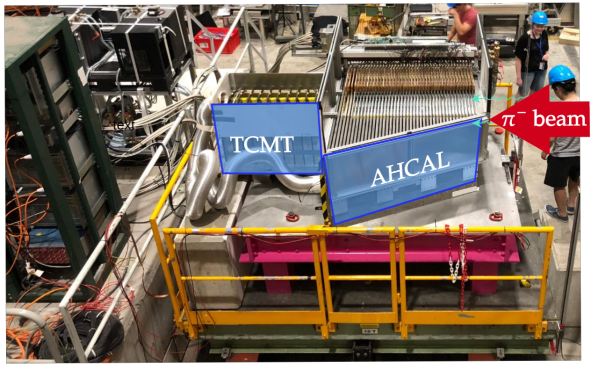


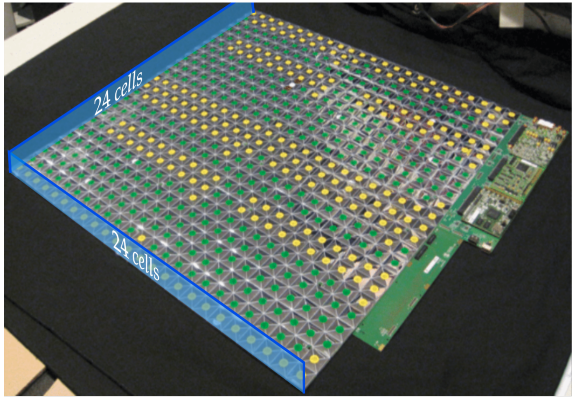
[1] M. A. Thomson. 'Particle Flow Calorimetry and the PandoraPFA Algorithm'. NIMA, pp. 25–40. doi: 10.1016/j.nima.2009.09.009.

- AHCAL is a Fe-Sc highly granular calorimeter prototype designed for Particle Flow.
- Calorimeter has ~22,000 individual SiPM-on-tile readout channels → highly granular;
- AHCAL is a five dimensional, non-compensating calorimeter (e/π = 1.25-1.4):
- Five-dimensional calorimeter: measures energy density of hadron showers in space *and time*, with **up to 100 ps time resolution allowed by hardware** for each active cell (3 x 3 x 0.3 cm^3).
  - hadron showers develop with a dense EM core and sparse HAD halo →

    AHCAL can exploit spatial development of hadron showers for shower separation;
  - neutron fraction of hadron shower directly proportional to HAD fraction →

    indirect energy depositions from neutrons *delayed* compared to first nuclear interaction by hadron by 10-100 ns in steel →
  - Temporal development of hadron showers expected to vary significantly from event to event→ likely that this information can be used to improve clustering
  - **AHCAL can exploit temporal development of hadron showers for shower separation;**





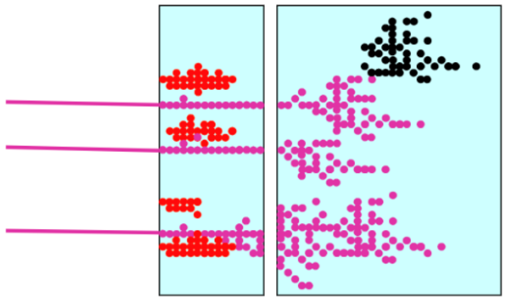
TAKE-HOME MESSAGE:

spatial & temporal energy density information

available from AHCAL may reduce confusion when clustering hadron showers.

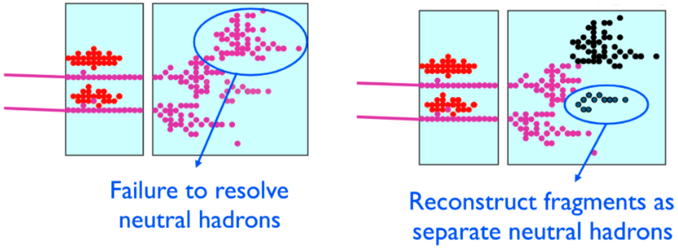
Machine Learning for Energy Clustering in Particle Flow

- In Particle Flow Calorimetry, main contribution to resolution at jet energy > 40 GeV is **confusion**.
- Confusion defined as '**the energy misallocated between clusters of energy deposits in PFC'** [1];
- Can occur for two main reasons [1]:
  - Insufficient sampling points in the calorimeter;
  - lack of sophistication in the pattern recognition algorithms.
- Study of [2] demonstrated excellent performance of graph neural network techniques applied to shower separation in a simulated detector with varying sensor sizes;
- However:
  - **Study did not include track/timing information**:
    - Track information critical for statistical re-clustering in Pandora PFA;
    - Timing information expected to improve performance; unclear as to how much of an improvement is possible and how this manifests;
  - **AHCAL has factor of 20 more sensors than calorimeter used in study;**
    - Unknown if models are scalable to highly granular calorimeters;



$$E_{JET} = E_{TRACK} + E_{\gamma} + E_{n}$$

Studied in this presentation



TAKE HOME MESSAGE:

Machine learning may provide superior pattern-recognition than classical methods→

potentially better clustering of energy deposits →

better jet energy resolution →

more precise measurements at future e^+-e^- colliders

[2] Shah Rukh Qasim et al. 'Learning representations of irregular particle-detector geometry with distance-weighted graph networks'. In: The European Physical Journal C 79.7 (July 18, 2019), p. 608. doi:10.1140/epjc/s10052-019-7113-9.

Problem Statement

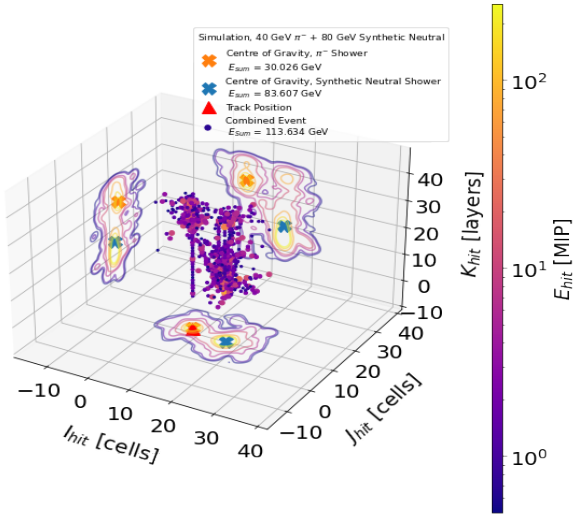
Given a charged and a synthetic neutral hadron shower observed simultaneously with AHCAL, with their centres-of-gravity-separated by an average of ~8ρ_M (20 cm):

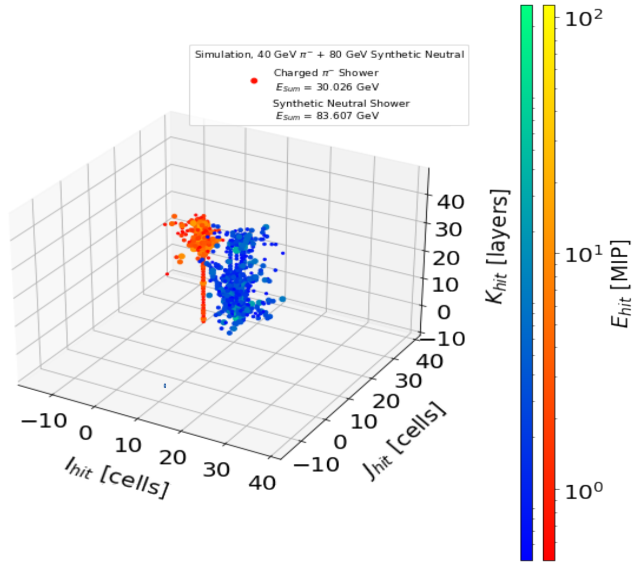
- What is the distribution and scale of confusion energy in AHCAL that we can achieve with ML models?
- How does the lateral/longitudinal distance and particle energy affect the amount of confusion?
- Does 100 ps timing information/track information help in clustering? If so, then how?
- Do the model learn clustering strategies that can be used to inform existing clustering techniques?

WHAT WE HAVE:

Hadron Shower Event Information

Cell indices, Energy, Timestamp, Track Position, Track Momentum





WHAT WE WANT:

Fractions of energy belonging to each shower, for each cell.

Shower Separation Models

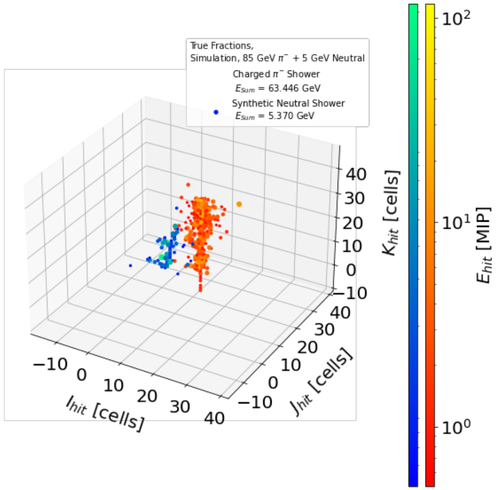
- **Three published neural network models are applied to shower separation for AHCAL using synthetically produced two-shower charged/neutral events in AHCAL:**
  - **PointNet** [3]:
    - Uses affine transformation matrices to cluster ;
    - Uses no local energy density information (i.e. each active cell is independent, and not related to its neighbours)
  - **Dynamic Graph Convolutional Neural Network (DGCNN)** [4]:
    - Uses successive k-nearest neighbour clustering in high dimensional space to learn a representation of hadron showers using
  - **GravNet** [2]:
    - Like DGCNN, but uses a low-dimensional clustering space to weight the significance of properties of the hadron shower;
    - Designed explicitly for use in particle shower separation.
- Models output fraction of energy belonging to each shower in each active cell;
- Models were modified to have a similar flow of information to Pandora PFA, and to support the inclusion of charged track and timing information.

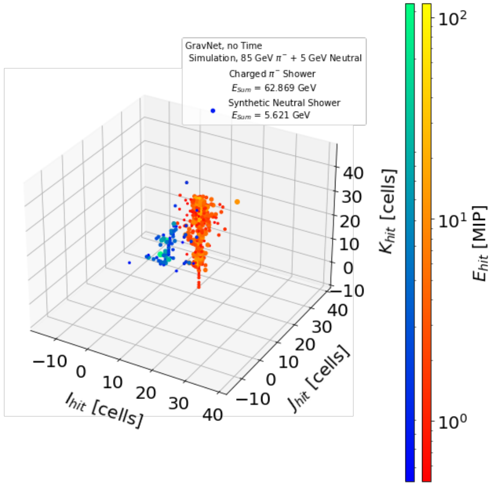
- **Several important additional studies mentioned here and in backup slides:**
  - Synthetically displaced in a circle and 'overlayed' events rather than simulated two shower events required for comparison of models trained on data vs. simulation;
    - diagram explaining data synthesis method in the backup;
  - Synthetic neutral hadron showers produced by removing the MIP-track that a charged hadron shower produces using a topological/energy cut:
    - study found that this method is highly effective at producing 40 GeV synthetic neutral shower events. Method is explained in backup;
  - Average distance at which hadron showers are separated is important for machine learning algorithms because ML algorithms bias to data, physical or otherwise!
    - Average inter-shower distance (i.e. radius of the circle) was studied and chosen to be around 80% of each shower is integrated radially from its centre-of-gravity.
    - This way, on average, there will always be some confusion energy for the model to learn with.

[3] Charles R. Qi et al. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. Apr. 10, 2017. doi: 10 . 48550 / arXiv . 1612 . 00593.
[4] Yue Wang et al. Dynamic Graph CNN for Learning on Point Clouds. June 11, 2019. doi: 10.48550/arXiv. 1801.07829.

Training Details

- All networks were implemented in PyTorch Lightning;
- 6 models were trained (colour coding shown):
  - PointNet, DGCNN, GravNet’
  - **network method, without timing information;**
  - **network method, with timing information.**
- Training, validation and testing samples of were combined, randomly with replacement, from three independent source samples of π- hadron shower events in AHCAL simulated with Geant4, with equal proportions of events for each possible combination of particle energies;
  - Particle Energy Range: 5-120 GeV, in steps of 5 GeV
  - Training Sample: ~1,800,000 source events → 720,000 charged-neutral two-shower events;
  - Validation Sample: ~200,000 source events → 80,000 charged-neutral two-shower events;
  - Testing Sample:~2,000,000 source events → 800,000 charged-neutral two-shower events;
  - More information on applied cuts in backup slide
- Hyperparameters were tuned using Optuna for 30 trials, with a maximum of ten epochs per trial, using ADAM optimiser;
- Loss function from [2]:





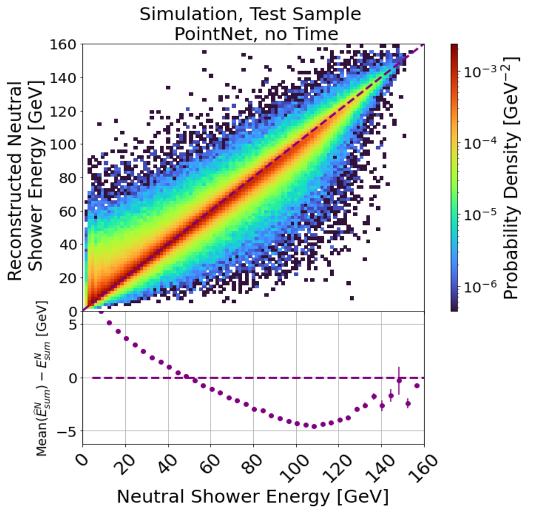
$$\mathcal{L} = \sum_k \frac{\sum_i \sqrt{E_i t_{ik}} (p_{ik} - t_{ik})^2}{\sum_i \sqrt{E_i t_{ik}}}$$

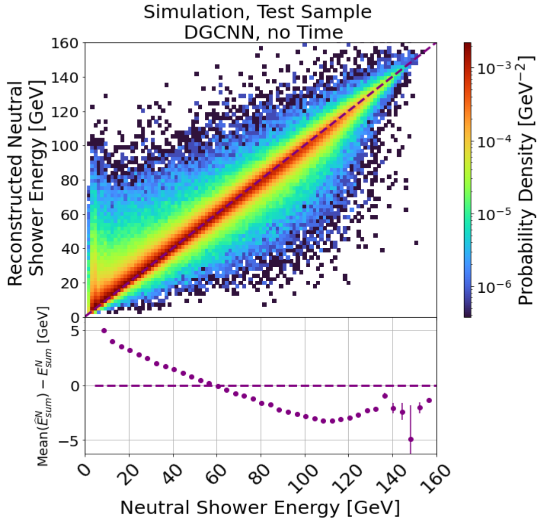
$t, p =$ true/predicted sensor energy fraction
$i, k =$ index of active cell/shower
$E_i =$ active cell energy of sensor i

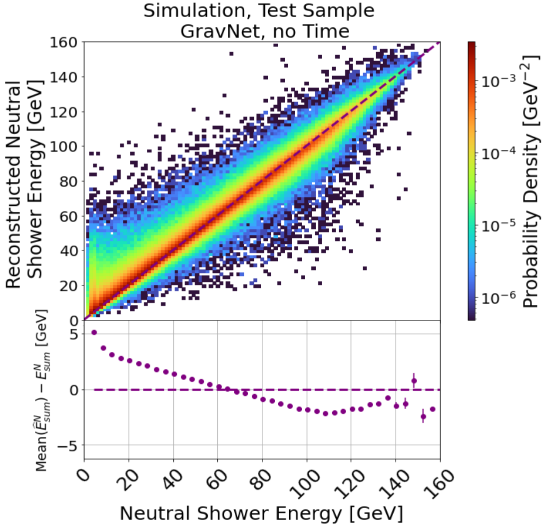
Note:
Minimum of each permutation of k is used, so no preference for the output channel is learned;

In testing, the combination of outputs showers with the lowest loss is used for testing, so ‘shower swapping’ is not included as a source of confusion (see [3])

Results: Overall Distributions of Reconstructed vs True Shower Energy







What we expect:

- In the ideal case, the models ought to, on average, reconstruct the same shower energy as the original shower.

What is shown:

- 2D histograms of the reconstructed vs. true shower energy, for each model without timing information.

  Blue→Red : low→high probability density.

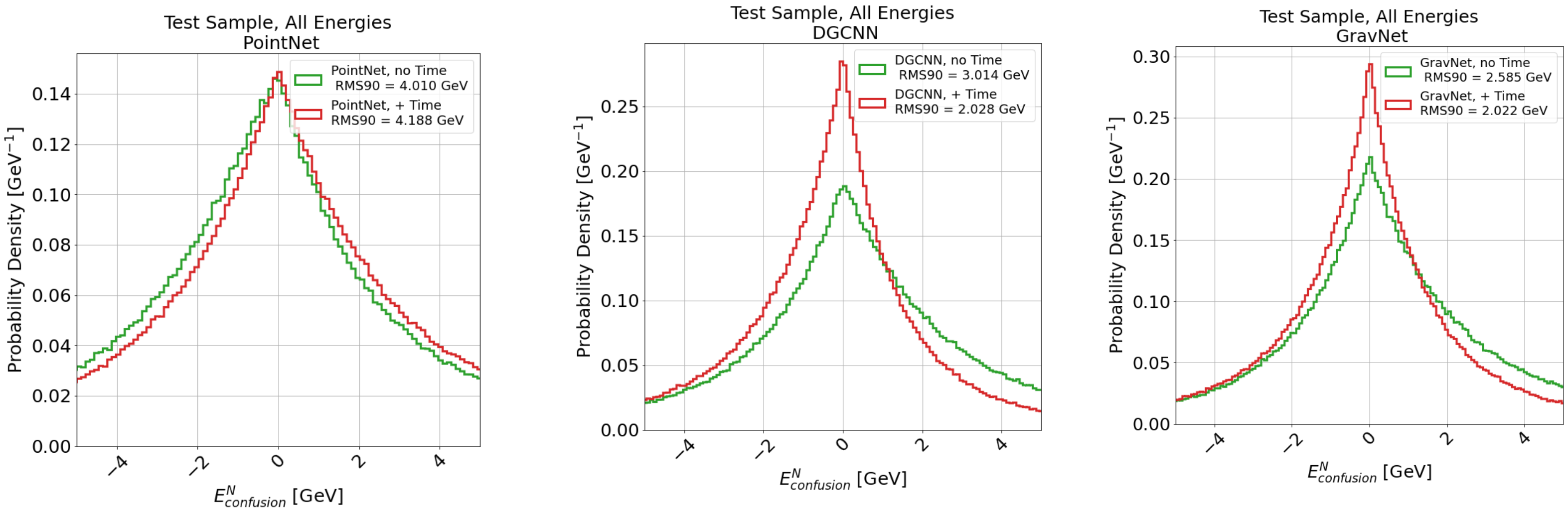
  The bottom subplot shows the difference between the mean reconstructed neutral and true neutral energy:
- **for separating any permutation of hadron showers in the testing data (5-120 GeV):**

  The purple dashed line indicates the agreement of the reconstructed with the true shower energy

What we learn:

- **Red Region close to purple dashed line:**
  Showers are frequently reconstructed with energies close to the original shower energy.
- **Asymmetric green region indicates skewness in distributions ;**
  - result reflected in the subplot, which indicates a linearly decreasing bias in the shower → crosses 0 at 60 GeV
  - DGCNN and GravNet have significantly smaller biases (max 5 GeV, but typically much lower)

Results: Overall Distributions of Confusion Energy



Reconstructed - True

$$E^N_{\text{confusion}} = \widehat{E}^N_{\text{sum}} - E^N_{\text{sum}}$$

$$f_{\text{res}} = \frac{1}{N_{\text{sample}}} \sum_{\text{sample}} \begin{cases} 1 & \text{if } |\widehat{E}^N_{\text{sum}} - E^N_{\text{sum}}| < \sigma_E \\ 0 & \text{otherwise} \end{cases}$$

What we expect:

- Networks should show a consistent confusion energy distribution between models;
- Networks using time should improve resolution if the model is sufficiently sophisticated that timing information can be exploited

What is shown:

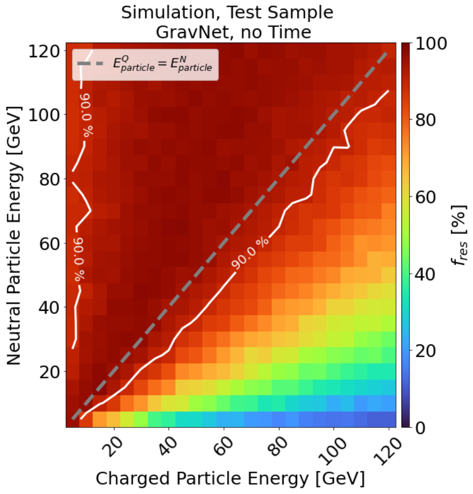
The distributions of neutral confusion energy from each network, with RMS_{90} width , **for separating any permutation of hadron showers in the testing data (5-120 GeV):**

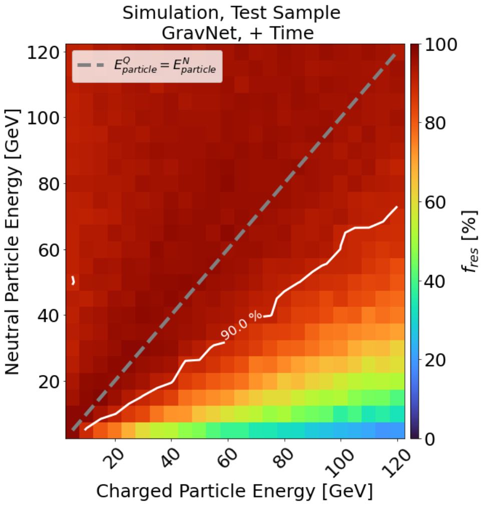
- **neural network, no timing information applied** (green);
- **neural network, with timing information applied** (red).

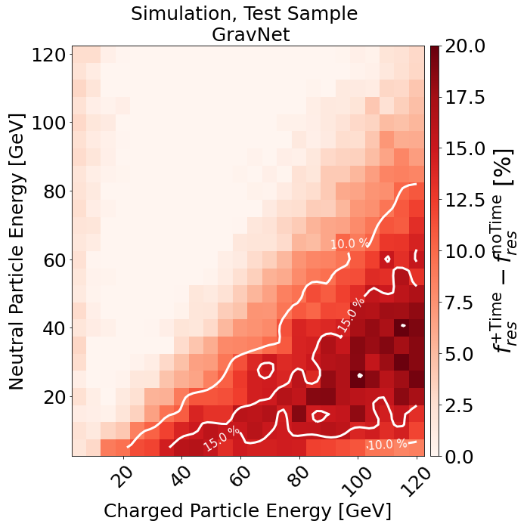
What we learn:

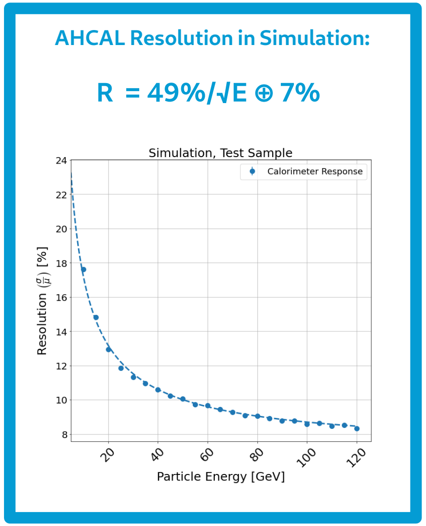
- No improvement to resolution using timing information in PointNet.
- Energy resolution both much better than PointNet and the improvement of time much more significant using both DGCNN and GravNet
  - → **21% improvement using time with GravNet**
  - → **35% improvement using time with DGCNN**
- **Best RMS90 resolution achieved from GravNet using timing information→ 2 GeV**
- Quote from DGCNN Paper:
  '"Instead of working on individual points like PointNet, we exploit local geometric structures…'" → suggests importance of time to clustering is not in the absolute but instead the relative value.

Results: Fraction of Events reconstructed within AHCAL resolution









What we expect:

- The difference between the particle energies should play a role in the amount of confusion:
- i.e. events with fewer active sensors are more likely to be incorrectly identified if there is another shower with more sensors.
- Networks using time should improve resolution, from previous results.

What is shown:

- **Left and Middle:** Matrices of the fraction of events reconstructed within the AHCAL resolution for each possible permutation of hadron energies
  - **Blue → Red**: more events reconstruted within the resolution
- **Right:** Difference between fraction with time and without time :
  - **White:** no improvement due to inclusion of time;
  - **Red:** additional percentage of events reconstructed within the resolution.

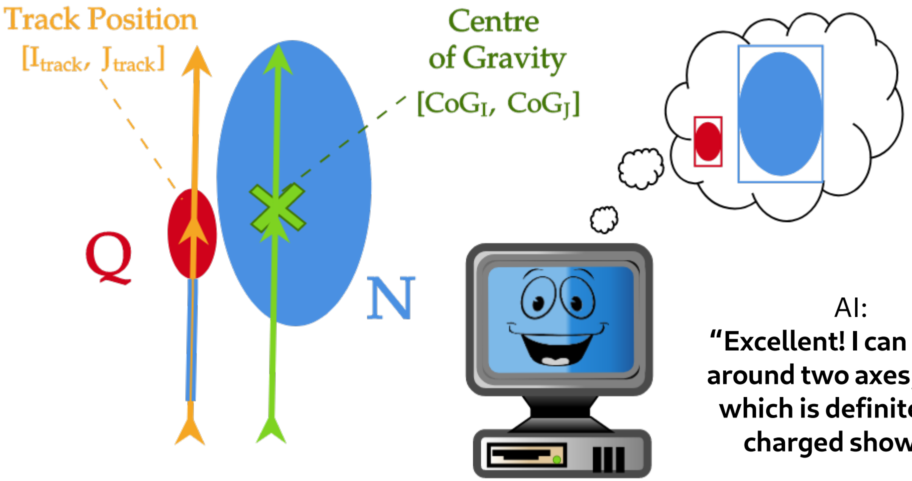
What we learn:

- **Confusion asymmetry:**
  - **>90% of events reconstructed within calorimeter resolution where $E^N > E^Q$**
  - Confusion most relevant where $E^Q > E^N$
- **Large red region on the ratio plot:** Timing information is relevant for clustering where $E^Q > E^N$.
- **Fraction of events reconstructed within AHCAL resolution increases by up to 15 %**
- **Results strongly suggest that:**
  - **track information is being exploited by the GravNet network;**
  - **timing information supplements clustering where track information is less reliable .**

Results: Visual Hypothesis for Clustering Asymmetry

$$E^{N}_{\mathrm{particle}} > E^{Q}_{\mathrm{particle}}$$

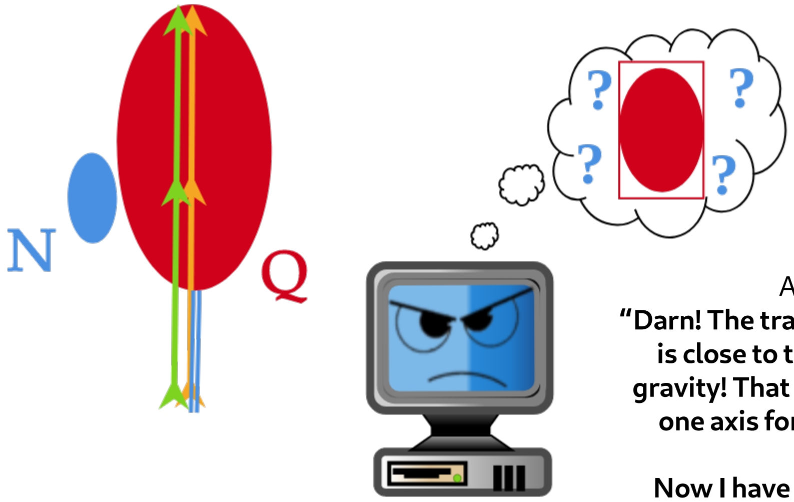
Track Position $\neq$ Centre of Gravity



AI:
"Excellent! I can cluster around two axes, one of which is definitely my charged shower."

$$E^{Q}_{\mathrm{particle}} > E^{N}_{\mathrm{particle}}$$

Track Position $\simeq$ Centre of Gravity



AI:
"Darn! The track information is close to the centre of gravity! That only gives me one axis for clustering!

Now I have to figure out where the neural shower is... "

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

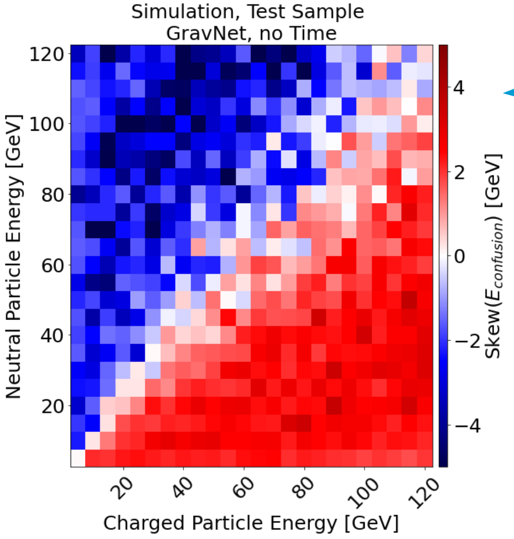
Results: Skewness of Confusion As a Function of Particle Energy

HIGH-D

CASE: $E_N > E_Q$

→ **Confusion energy distribution negatively skewed.**
→ **Negative skew means more frequent underprediction of neutral energy**

CONCLUSION:
The charged shower is allocated confusion energy from the neutral shower more frequently when $E_N > E_Q$



CASE: $E_Q > E_N$

→ **Confusion energy distribution positively skewed.**
→ **Positive skew means more frequent overprediction of neutral energy**

CONCLUSION:
The neutral shower is allocated confusion energy from the charged shower more frequently when $E_Q > E_N$

Reconstructed - True

$$E^N_{\text{confusion}} = \widehat{E}^N_{\text{sum}} - E^N_{\text{sum}}$$

What we expect:

- If skewness is a function of particle energy, there exists some bias for clustering the hadron showers that is most effective.

What is shown:

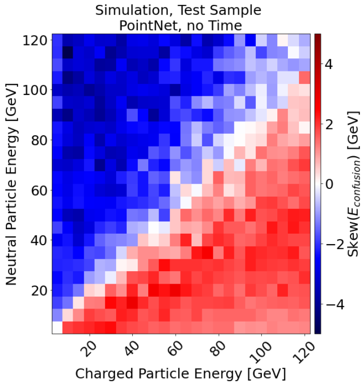
- Skewness of neutral confusion energy distribution:
  - **Blue:** negative skewness;
  - **White:** no skewness;
  - **Red:** positive skewness;

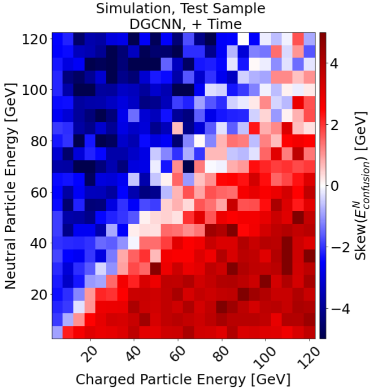
What we learn:

- Consistently across networks, with or without time, the algorithms more frequently donate energy from the shower with more energy to the one with less.
- Note, from [1], on Pandora PFA:

  "by design the initial clustering stage errs on the side of splitting up true clusters rather than merging energy deposits"

- Neural networks consistently learn the same strategy.





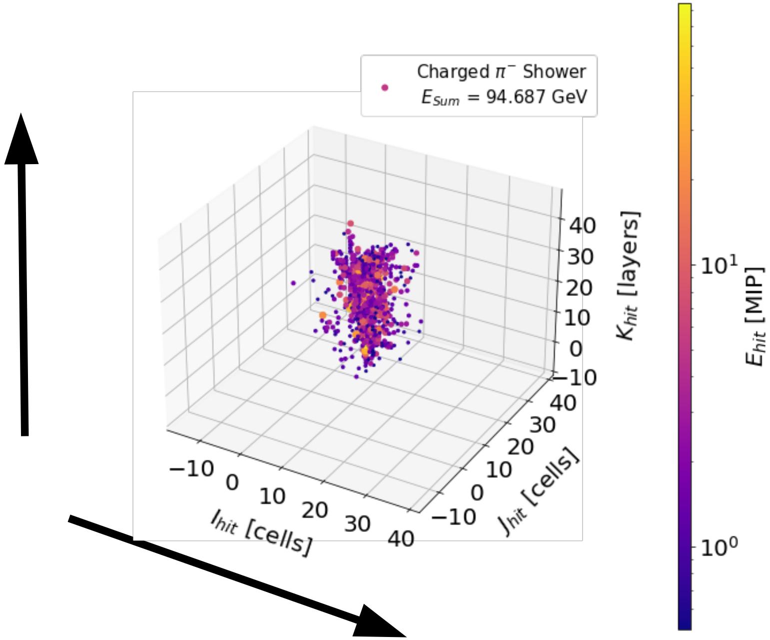
Conclusion

- Shower Separation is critical to the performance of Particle Flow Calorimetery;
- Special neural networks can be used to reconstruct hadron showers where more than one particle is present in the event;
- Several neural network models from literature were implemented for shower separation, exploit the spatial and temporal energy density of the highly-granular AHCAL detector, to study their properties;
- The following observations were made:
  - Graph neural networks produce superior results to point-based model, and can exploit timing information;
  - Neural networks are able to exploit topological and statistical clustering in the same model;
  - Cases where $E_N > E_Q$: more than 90% of events reconstructed within the calorimeter resolution.
  - Cases where $E_Q > E_N$: up to 15% more events are reconstructed within the calorimeter resolution if 100 ps timing resolution is available;
  - All neural networks prefer to separate clusters of energy than merge them, similarly to Pandora PFA.
- MAIN RESULTS:
  - **A strong case exists for the AHCAL temporal calorimeter→ improvement for $E_Q > E_N$**
  - **Neural networks learn similar clustering strategies to Pandora PFA**

What constitutes an 'event' for AHCAL?"

Z Axis:

- Layers of absorber, active material and sensors (cells);
- 38 active layers.



X-Y Axes:

- Matrices of sensors (cells);
- 24 x 24 cells per layer

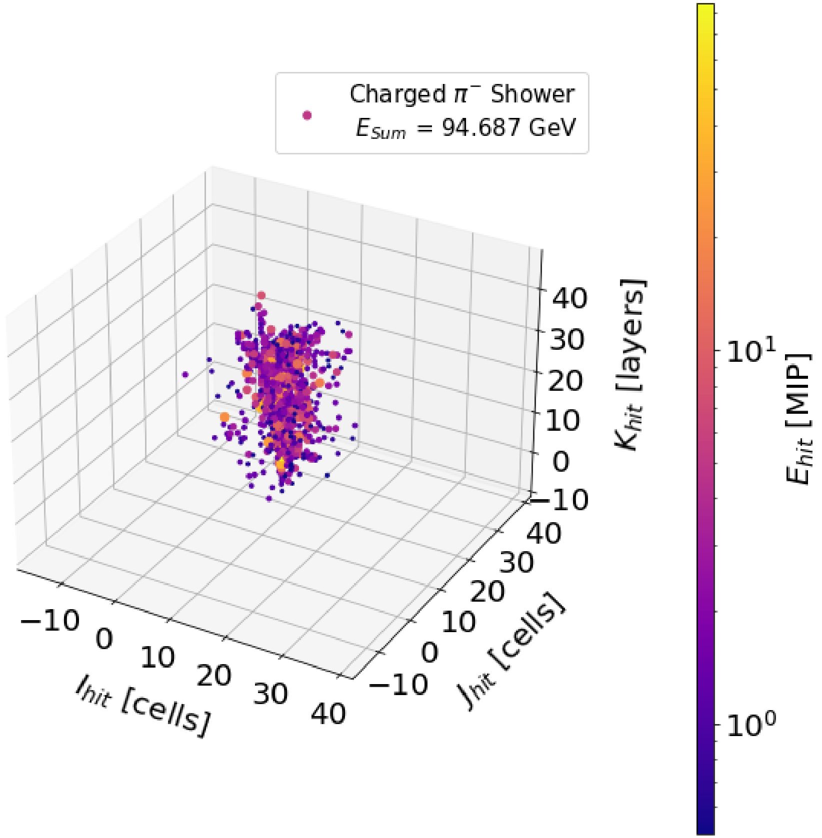
Color Axis:

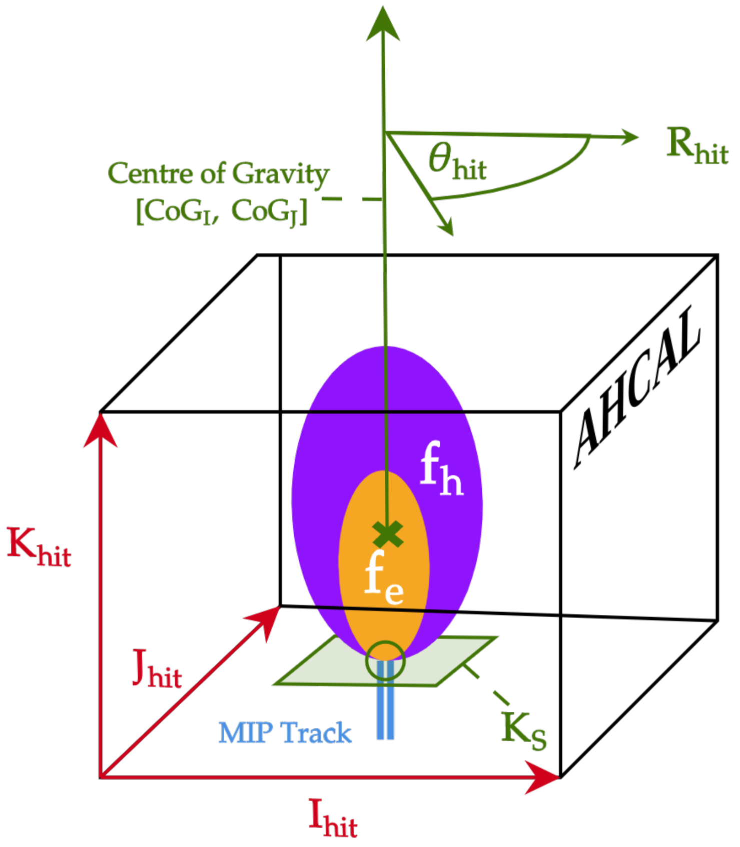
- Energy of cell, in muon-calibrated 'minimum ionising particle' (MIP) units;
- 22,000 cells altogether;
- Sum of all the cell energies → reconstructed energy of hadron;

Additionally:

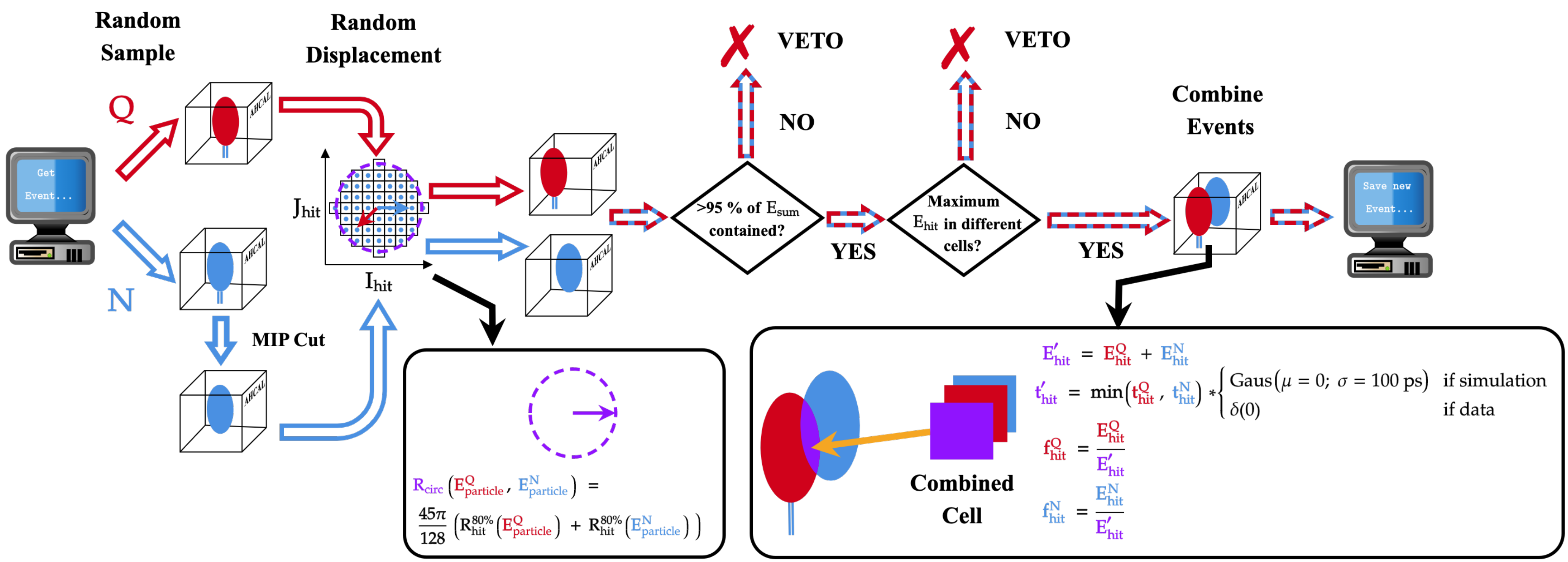
- Timing information for each cell in nanoseconds;
- Not shown in this event display.

Backup: 'Natural' Shower Coordinate System

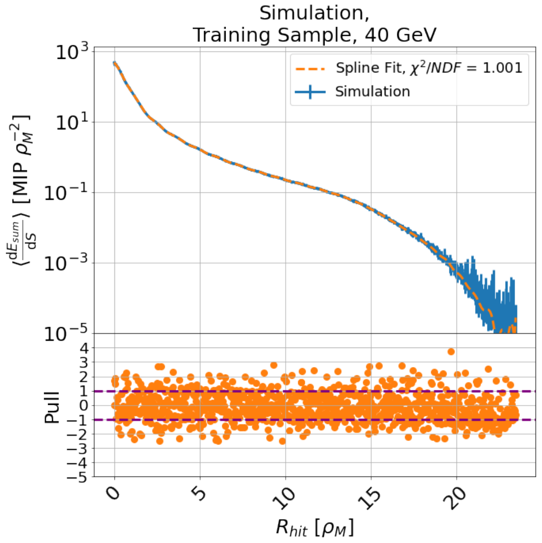


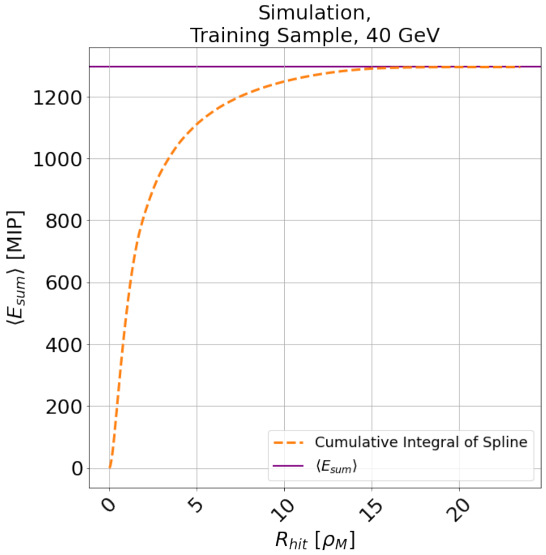


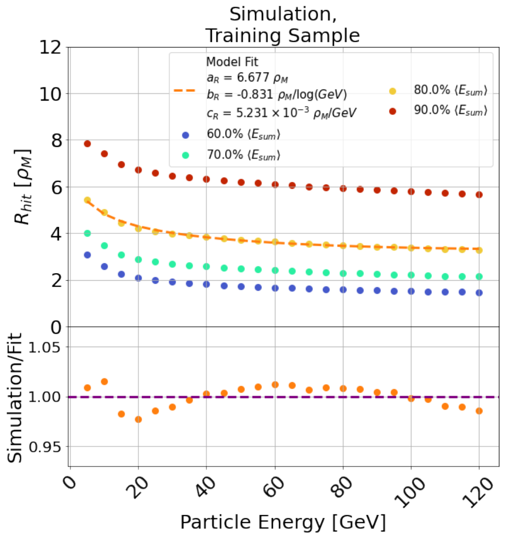
Synthetic Shower Overlay Algorithm



Shower Distance Procedure







Step 1:

- Calculate differential energy deposited per unit area in a circle of radius R_{hit} around the centre-of-gravity;
- Fit the distribution with a cubic spline.

Step 2:

- Calculate cumulative integral of the differential energy loss by integrating spline.
- This gives the average cumulative energy per additional unit radius of the circle. It saturates at the average energy loss of the shower.
- Repeat this procedure for each particle energy in the sample.

Step 3:

- Calculate radial distance at which 80% of the hadron shower energy is deposited (orange line)
- In simulation, this was found to empirically follow the function (within 1-2 %):

$$R_{\text{hit}}^{80\,\%}(E_{\text{particle}}) = a_R + b_R \cdot \log E_{\text{particle}} + c_R \cdot E_{\text{particle}}$$

Results: Systematic Bias and Statistical Uncertainty on Reconstructed Energy as a function of Radial Shower Distance

What we expect:

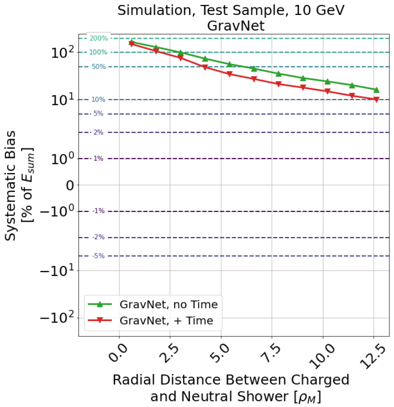
- As the distance between hadron showers increases, so too should the systematic bias and statistical uncertainty the reconstructed energy, as a percentage of the shower energy, decrease.
- The proportion of confusion should decrease with particle energy (proportionally less confusion energy)
- The inclusion of time should reduce the bias and uncertainty;

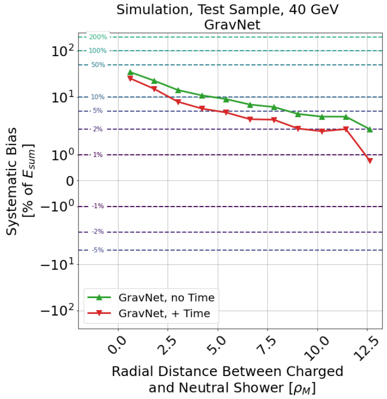
What is shown:

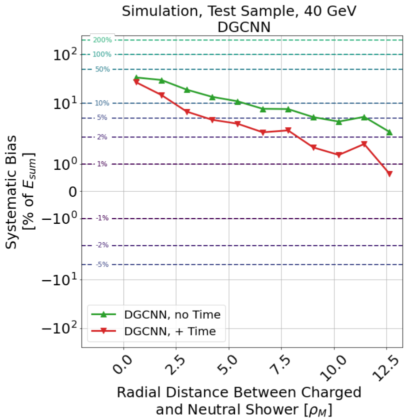
- Systematic bias (mean) and statistical uncertainty (std. deviation) in percent of the reconstructed energy vs. the radial centres-of-gravities of the charged and neutral hadron showers, in Moliere radii
- The inclusion of time should reduce the bias and uncertainty if it is useful for clustering.

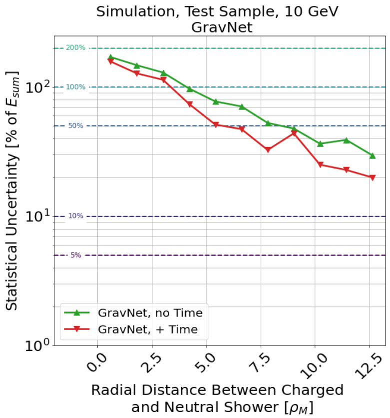
What we learn:

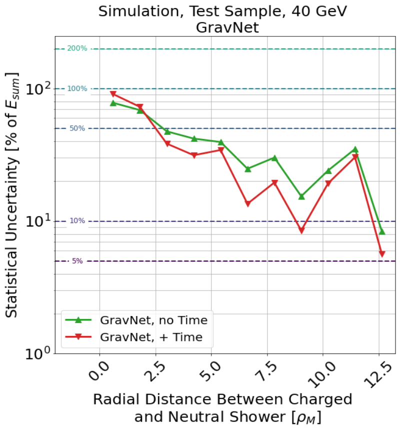
- Systematic bias and statistical uncertainty decrease approximately exponentially with increasing separation distance.
- For all energies shown, timing information reduces the bias and uncertainty, particularly for distances of R > 2.5 ρ_M

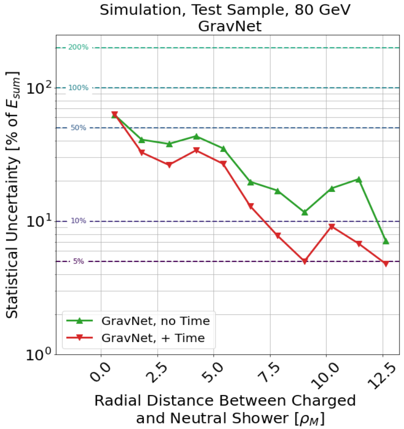






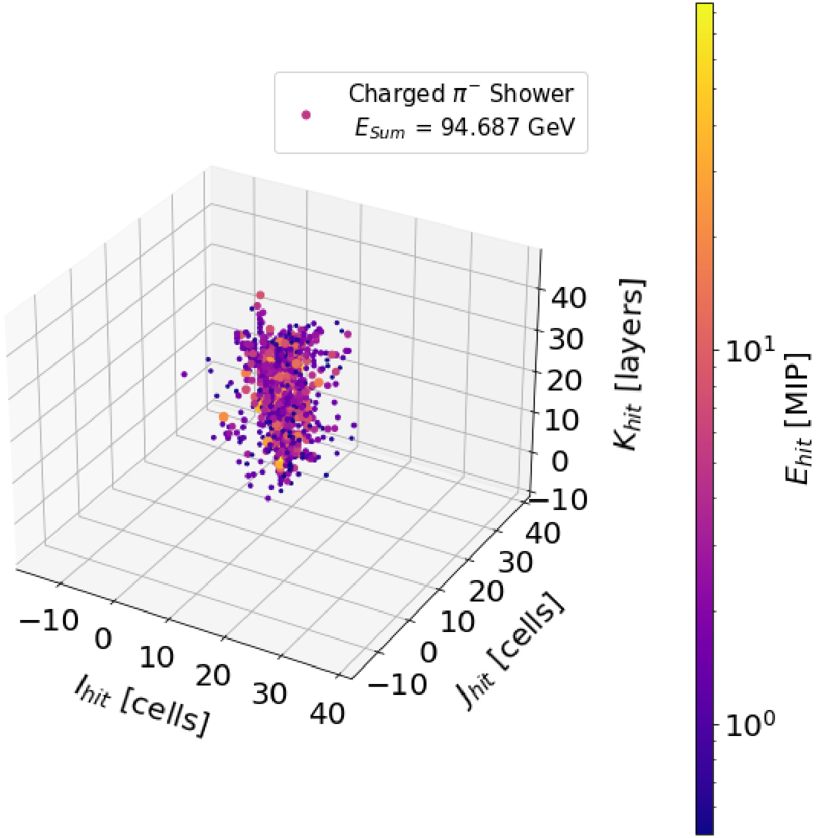




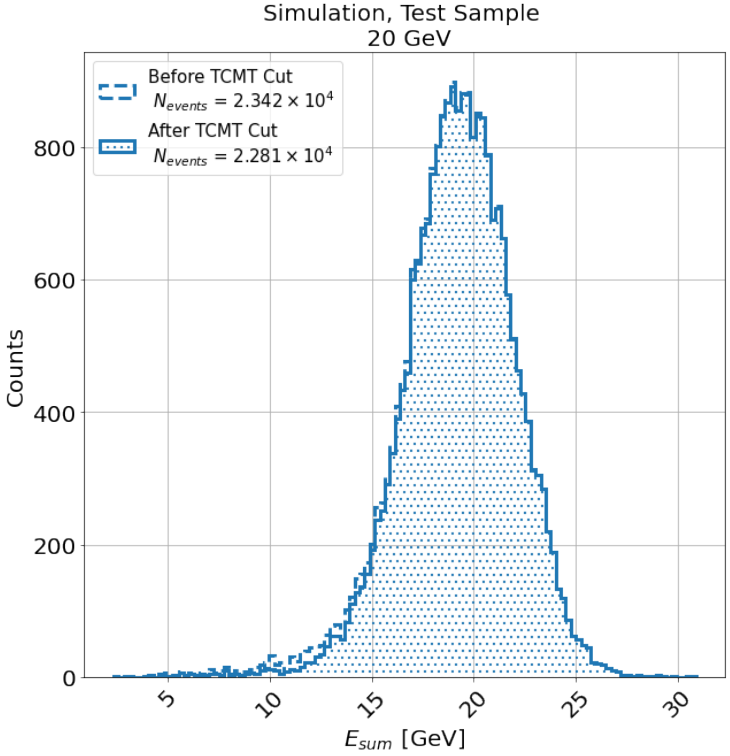


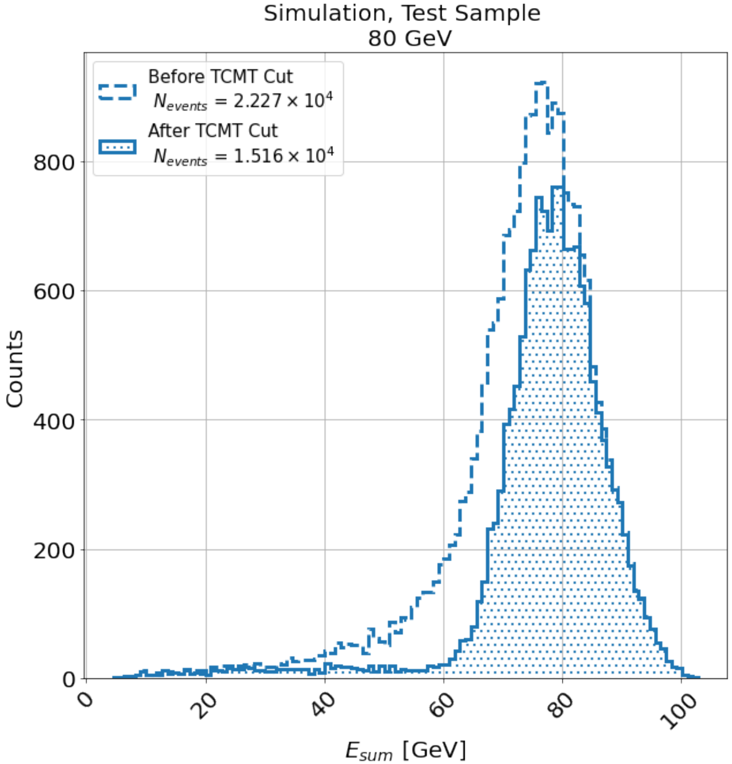
Backup: Simulation Details

- Physics list: QGSB_BERT_HP
- Particle: π^-
- Cuts:
  - + Single track, with position 'inside' calorimeter: $1 < I_{\mathrm{Track}}/J_{\mathrm{Track}} \leq 24$
  - + PID MIP Cut: $P_{\mu-\mathrm{like}} < 0.5\%$
  - At least 50 active cells remaining after MIP Track Cut



Backup: TCMT Cut





What is shown:
Distributions of uncompensated calorimeter response, before and after tail-catcher cut, at 20 GeV and 80 GeV.

What we learn:
Effect of leakage reduced by application of tail-catcher cut at high particle momentum.