Shower Separation in Five Dimensions using Machine Learning Techniques.

Jack Rolph, Gregor Kasieczka and Erika Garutti
University of Hamburg

March 22, 2021

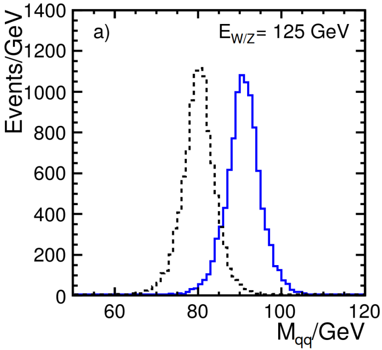
SPONSORED BY THE
Federal Ministry of Education and Research

CALICE

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

The need for sophisticated hadron shower clustering.

- **Future linear collider** 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 [4] 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 **calorimeter with temporal information** aid in clustering?



Reconstructed invariant mass distributions for the hadronic system in simulated $ZZ \to d\bar{d}\nu\bar{\nu}$ and $W^+W^- \to u\bar{d}\mu^-\bar{\nu}$ [4]

1

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

Goals.

Predict the fraction of energy belonging to two simultaneous hadronic showers observed in a highly granular calorimeter prototype, 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 neural shower separation:
  - **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**?

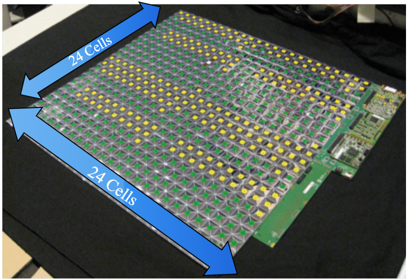
[1] Yue Wang et al. *Dynamic Graph CNN for Learning on Point Clouds.* 2018. arXiv: 1801.07829.

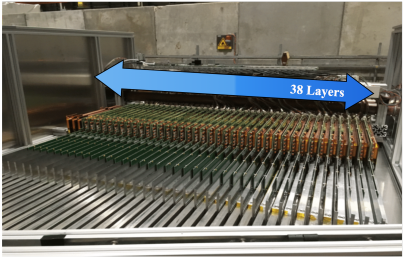
[2] Shah Rukh et al Qasim. "Learning representations of irregular particle-detector geometry with distance-weighted graph networks". In: (2019). ISSN: 1434-6052.

CALICE AHCAL Design

Detector Statistics:

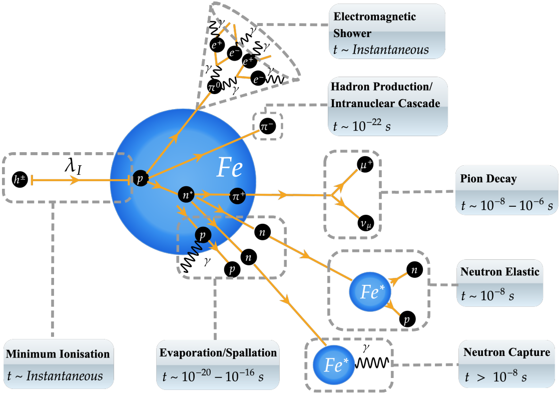
- Dimensions: 72 cm $\times$ 72 cm $\times$ 75 cm
- Sampling Calorimeter: 3mm
- Absorber: 17mm, steel
- Utilizes SiPM-on-Tile Technology.
- Depth: $\sim$ 4 λ_I over 38 Layers
- Cell Dimensions: $3 \times 3\text{cm}^2$;
- Total Channels: 21,888.

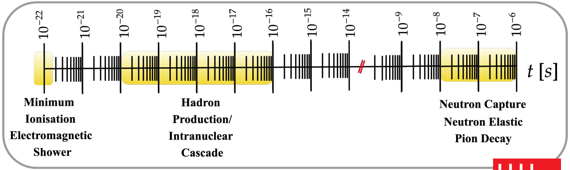




Why we measure time.

- Hadron showers involve **many different energy-deposition processes**.
  - **EM fraction** near **instantaneous**;
  - **Hadronic fraction slower**;
- **Late time development** correlated with **hadronic fraction**.
- **Shower development** of an **EM-process dominated shower completely different** to **hadronic-process dominated**.
- **Does this help clustering**?

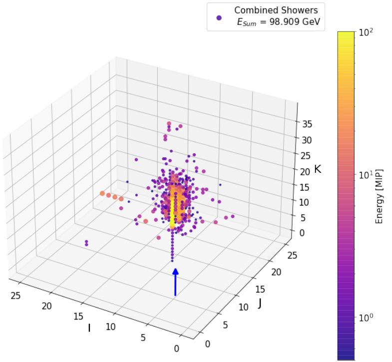




Simulation Information: Summary.

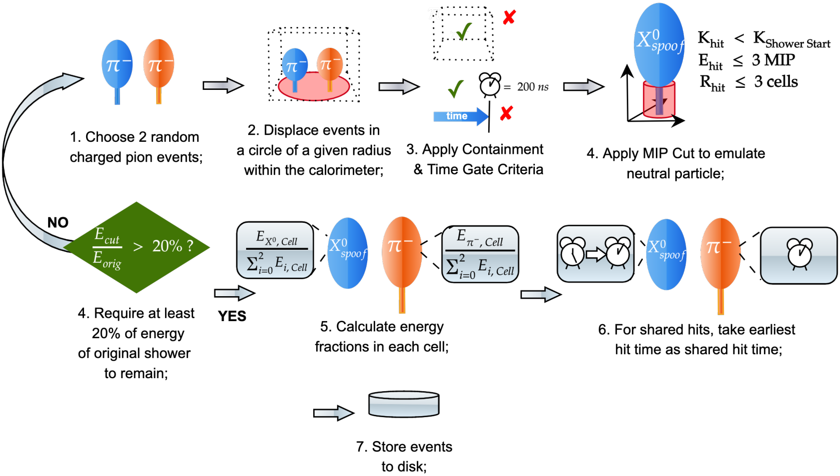
Simulation of π^- hadronic showers **using Geant4** in the AHCAL were used:

- Physics list: **QGSB_BERT_HP**
- **full detector simulation** (inc. SiPM saturation/noise thresholds etc.)
- Based on **June 2018 CALICE Testbeam** taken at SPS;
- **Actual data also used to validate**;
- Simulated particle energies: **10-80 GeV in steps of 5 GeV + 90 GeV, 120 GeV**
- **No neutral data; this must currently be spoofed.**

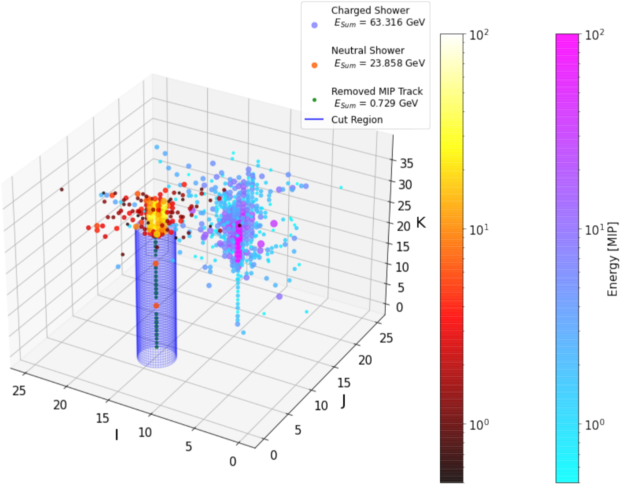


Example event display of a 80 GeV negative pion detected by the AHCAL

Simulation: Multi-shower Augmentation Algorithm.



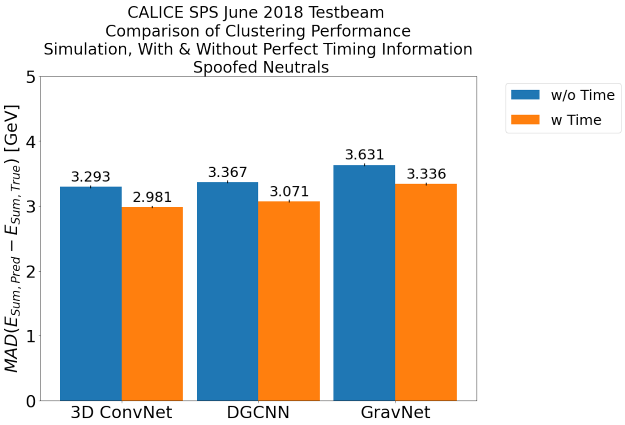
Simulation: Example Combined Shower Event.



The Experiment.

- **Train** a series of **shower separation networks**, **with** and **without time** as an input variable. using **perfect simulated time resolution**;
  - Is there an **improvement** in **energy resolution**?
  - Can it be applied to **actual data**?
- Obtain **samples** of hadronic **shower pairs** with **decreasing time resolution** from 0ns to 2ns;
  - **At what resolution** does does the **improvement** in clustering **due to time cease**?

Results: Network Validation on Simulation.

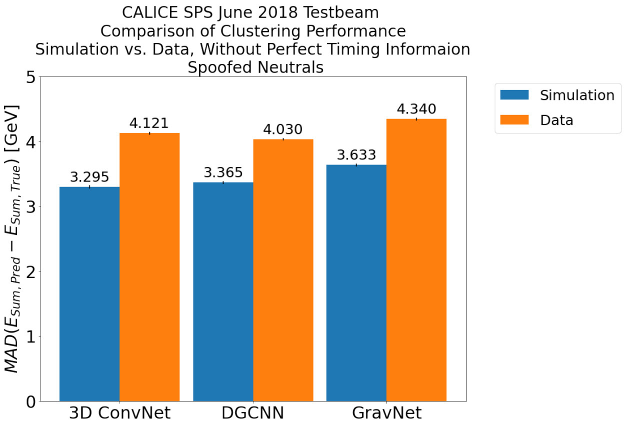


$$\mathrm{MAD}(X) = \mathrm{Median}\left(|X_i - \mathrm{Median}(X)|\right) \quad (1)$$

What one learns:

- Performance measured by: **difference between true/predicted shower energies.**
- **Energy resolution always improves with time.**
- **Improvement around 300 MeV;**
- **GravNet performs worst overall** on simulation;
- **3D ConvNet performs best overall** on simulation;

Results: Network Validation on Data - A First Ever Study!

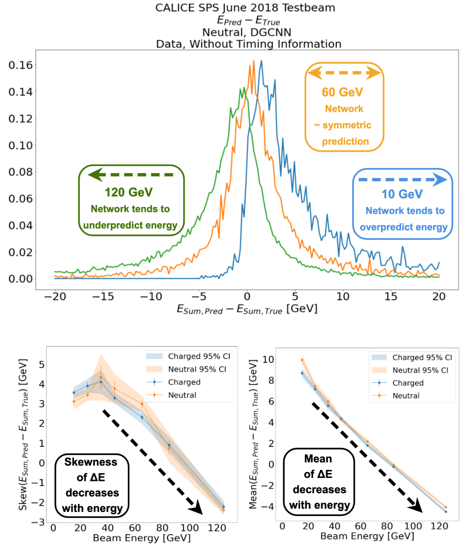


$$\mathrm{MAD}(X) = \mathrm{Median}\left(|X_i - \mathrm{Median}(X)|\right) \tag{2}$$

What one learns:

- Performance measured by: **difference between true/predicted shower energies.**
- **Time information in data not sufficiently calibrated for validation on neural network.**
- **Data** experiences around **700 MeV worse** performance than in **simulation.**
- **Distributions available in backup slides.**

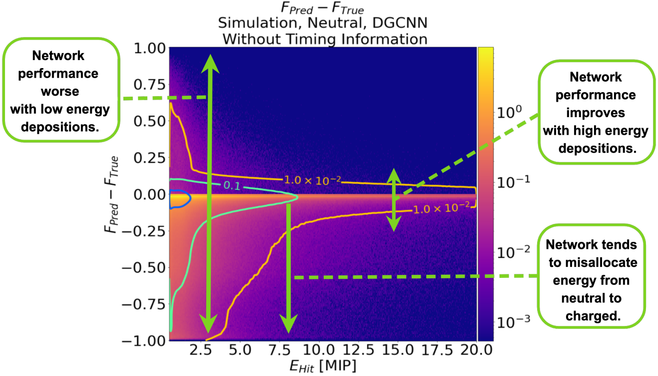
Results: Network Validation on Data, ΔE distributions.



What one learns:

- Algorithm learns energy range of hadron shower data.
- Over-predicts at the low-end of energy range;
- Under-predicts at the high-end of energy range;

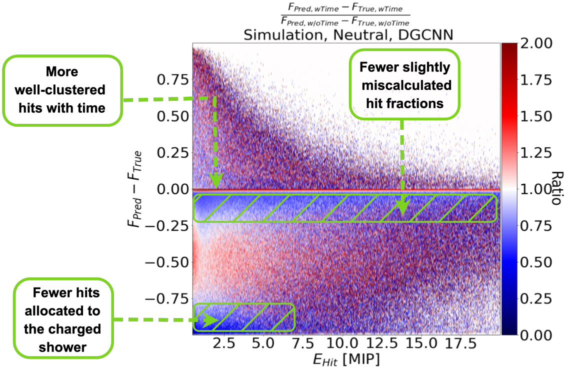
Results: Performance Vs Hit Energy, E_{hit}.



What one learns:

- **Neutral shower fractions** tend to be **under-predicted**.
- **Performance improves** with **hit energy**;

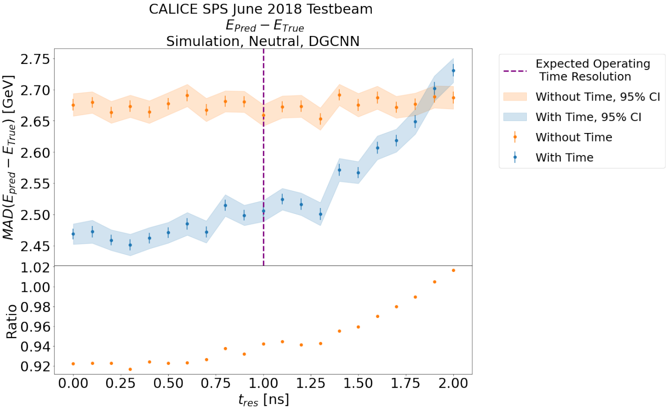
Results: Performance Vs Hit Energy, E_{hit}



What one learns:

- Time improves slight deviations from truth at all energies;
- Time improves mis-allocated hits to charged shower at low energies.

Results: Increasing Time Resolution.



What one learns:

- Reduction in performance degrades well above desired operating resolution of 1 nanosecond.

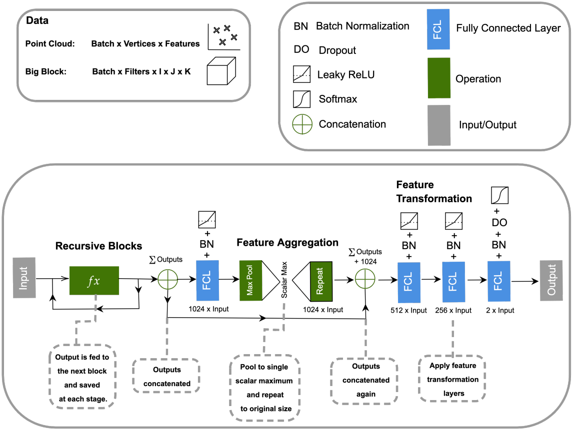
Conclusion.

- An **efficient multi-shower data augmentation tool** was developed;
- **Small improvement** in **clustering performance** of around **300 MeV** was observed using **perfect time resolution.**
- **Improvement due to time** is **due to minor error-correction** and **correctly otherwise misallocated hits**
- **Networks can be applied to data**, but perform with **700 MeV worse resolution than simulation**;
- **Improvement due to time is no longer useful** after $\sim$ **1.5 ns time resolution**;

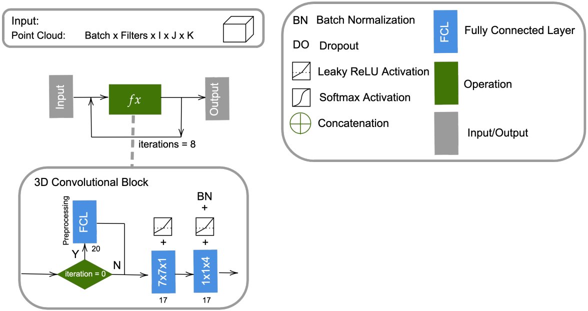
Outline

> Backup

Network: Underlying Architecture

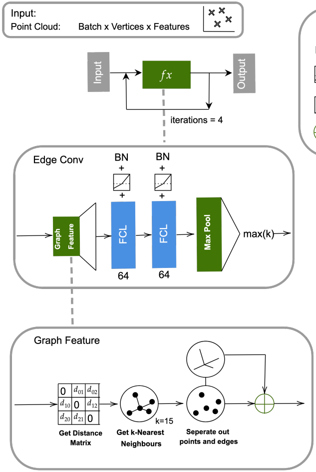


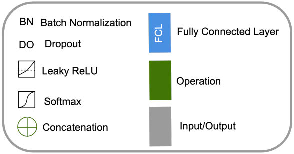
Network: Standard Convolutional Block



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

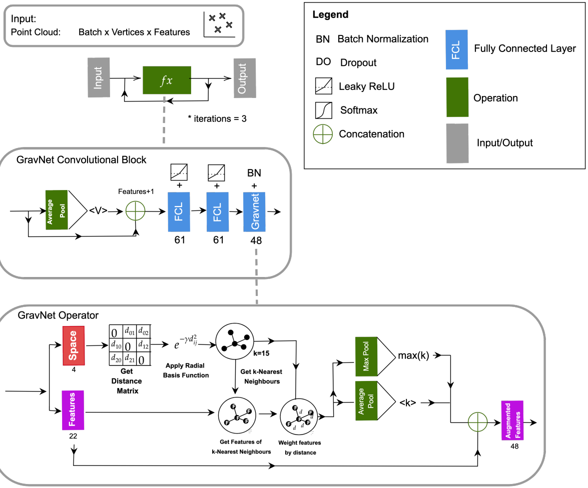
Network: Dynamic Graph Convolutional Block





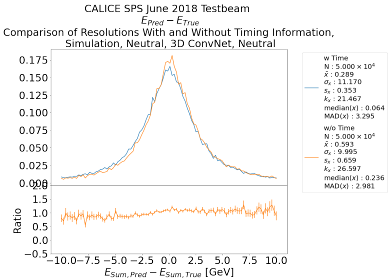
Dynamic Graph Convolutional block architecture. Complete network has 977,024 learnable weights.

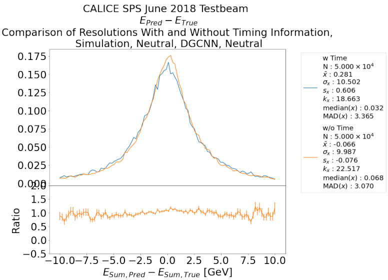
Network: GravNet Block

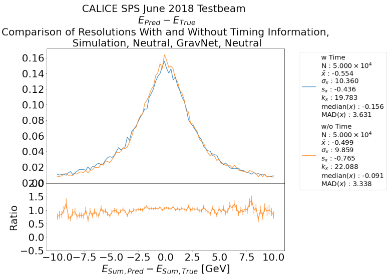


GravNet Convolutional block architecture. Complete network has 980,042 learnable weights.

Results: E_{sum} Reco. Performance (Simulation).



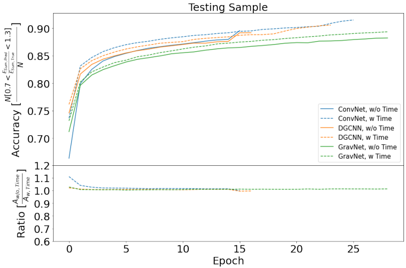




Results: Accuracy Curves.

What one learns:

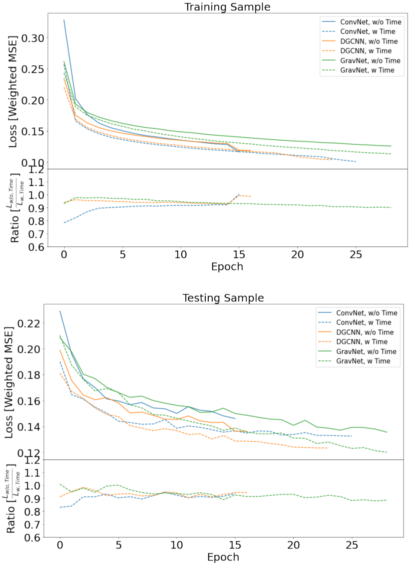
- Accuracy **always improves with time.**
- **'Kink' at 15 epochs** in loss curve due to **overcompensating step function** applied to **learning rate** due to **gradient explosion**;
- **Still attempting to find solution.**
- Training still in progress as a result.
- **As above with accuracy, see backup.**



Results: Loss Curves.

What one learns:

- Loss **always improves with time.**
- **'Kink' at 15 epochs** in loss curve due to **overcompensating step function** applied to **learning rate** due to **gradient explosion**;
- **Still attempting to find solution.**
- Training still in progress as a result.
- **As above with accuracy, see backup.**



Network: Figures of Merit and Hyperparameters.

Figures of merit:

- Loss Function:

$$L = \sum_k \frac{\sum_i \sqrt{E_i t_{ik}} \left(p_{ik} - t_{ik}\right)^2}{\sum_i \sqrt{E_i t_{ik}}} \quad (3)$$

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

- Accuracy Function:

$$A = \frac{N_{Events}(0.7 < \frac{E_{pred}}{E_{true}} \leq 1.3)}{N_{Events}} \quad (4)$$

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

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

Hyperparameters:

In machine learning, a hyperparameter is a parameter whose value is used to control the learning process. By contrast, the values of other parameters (typically node weights) are derived via training. [2]

All figures of merit and choice of hyperparameters used were defined in the reference paper [3].

- **Batch Size**: 20 events
- **Total** $N_{\textbf{epochs}}$: 20
- **Training Size**: 2×10^5 Events
- **Test Size**: 2×10^4 Events (10 % of Training Size)
- **Validation Size**: 5×10^4 Events (25 % of Training Size)
- **Optimizer**: ADAM
- **Learning Rate**: $3 - 1 \times 10^{-4}$, varies by network
- **Scheduler**: Exponential Decay, Factor = 0.99 + Step Function if Training Loss Increases, Factor = 0.75
- **GPU**: Nvidia P100

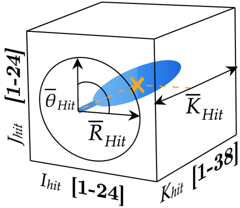
1

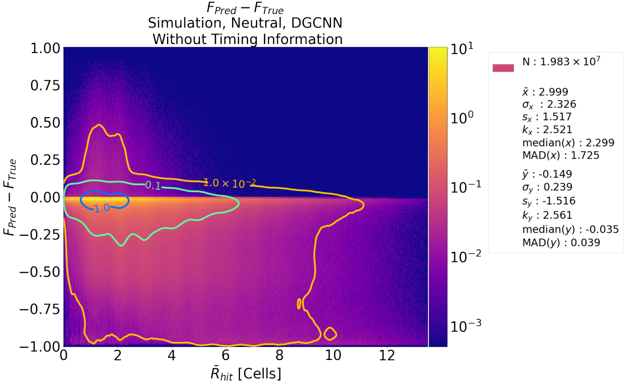
[1] *Hyperparameter (machine learning)*. In: *Wikipedia*. Page Version ID: 984957886. Oct. 23, . (Visited on 01/18/2021).

Results: Performance Vs Hit Radius, $\bar{R}_{hit}$.

What one learns:

- **Neutral shower fractions** tend to be **under-predicted.**
- **Over-prediction close to the shower core.**
- **Performance worsens slightly** with **distance from the hadron shower core;**



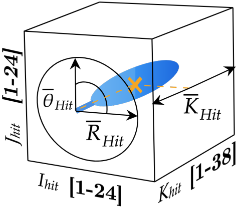


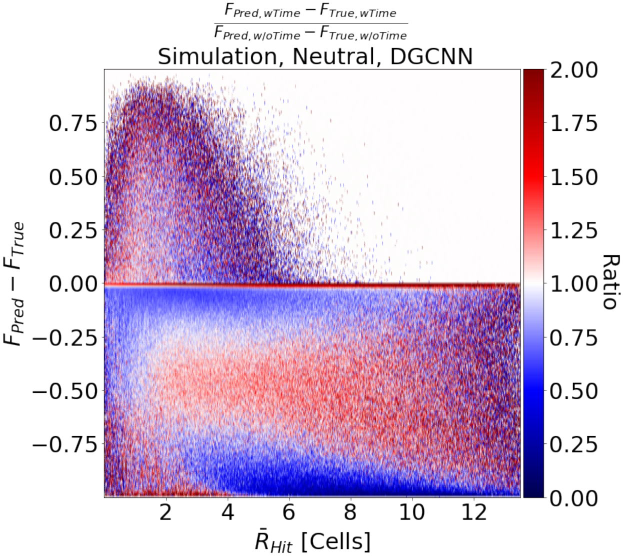
Results: Performance Vs Hit Time, $\bar{t}_{hit}$.

What one learns:

Time improves:

- slight deviations from truth close to shower core;
- mis-allocated hits further from shower core.

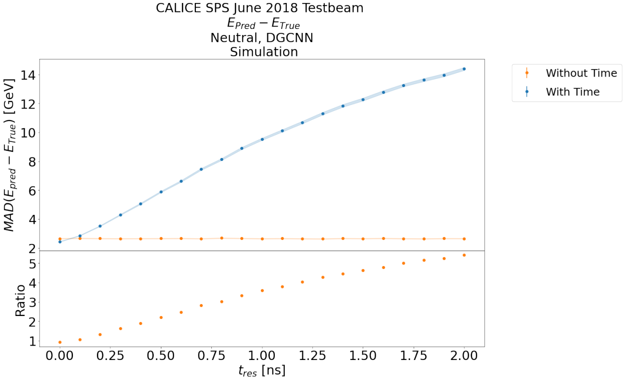




Results: Increasing Time Resolution.

What one learns:

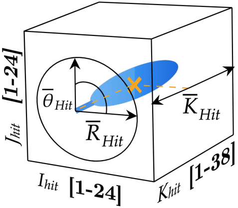
- Reduction in performance degrades well above theoretical operating resolution of 1 nanosecond.
- Graph networks sensitive to time resolution
- ConvNet result suggests clustering in only space is sufficient.

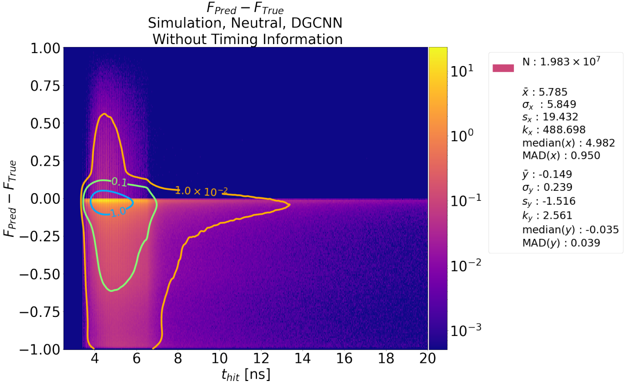


Results: Performance Vs Hit Time, $\bar{t}_{hit}$.

What one learns:

- Neutral shower fractions tend to be **under-predicted**.
- Interestingly, **late hits are clustered better than earlier hits - unexpected.**

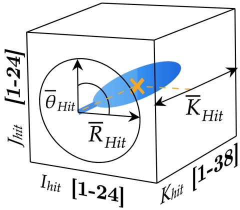


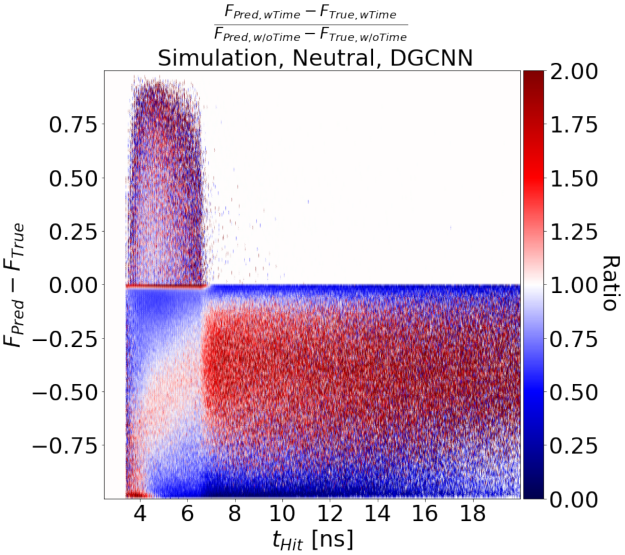


What one learns:

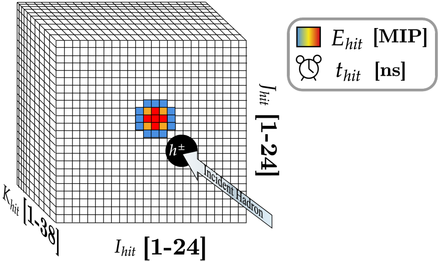
Time improves:

- slight deviations from truth at all energies;
- mis-allocated hits to charged shower at low energies.





Detector Observables.



$E_{sum} = \sum_{i=0}^{N} E_{hit}$ [MIP]

$N_{hits} = dim\left(\overrightarrow{E_{hit}}\right)$

$CoG = \frac{\sum_{i=0}^{N_{Hits}} E_{Hit,\,i}\, x_{ik}}{\sum_{i=1}^{N_{Hits}} E_{Hit,i}}$ [Cells]

	0.5	0.5	0.5	
0.5	1.0	2.0	1.0	0.5
0.5	2.0	6.0	2.0	0.5
0.5	1.0	2.0	1.0	0.5
	0.5	0.5	0.5	

24 MIP

	1	1	1	
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
	1	1	1	

20