Software Compensation using Machine Learning Techniques.

Jack Rolph
University of Hamburg

March 24, 2021

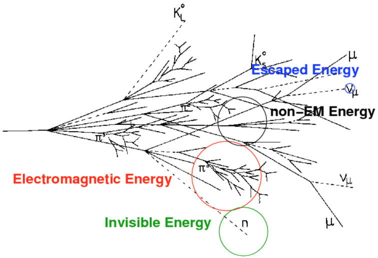
SPONSORED BY THE
Federal Ministry of Education and Research

CALICE

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

The need for software compensation.

- **Hadronic calorimeter** typically has **lower energy-resolution** compared to **electromagnetic calorimeter**;
- **Reason**: **hadronic showers** deposit an **unpredictable** fraction of ’**invisible energy**’;
  - **nuclear binding energy** (energy to ’break up’ nucleus)
  - ’**escaped**’ **particles** (neutrinos, neutrons,)
  - **muons**, that only deposit **minimum ionizing energy**.
- Upshot: a hadronic calorimeter **cannot necessarily measure all the energy of a hadron shower event.**

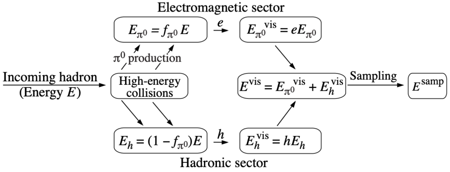


Example diagram describing ’missing energy’.

The need for software compensation.

EM/Hadronic response ratio can be 'compensated' for in a number of ways:

- **Calorimeter Design**: i.e. use **correct materials** at **correct thickness** in absorber/scintillator i.e. ZEUS;
- **Dual Readout**: use **Cerenkov detectors** as well as **scintillators** to **estimate electromagnetic fraction**, shower by shower;
- **Software Compensation**: use **software** to **weight energy** of **hadron shower offline**.

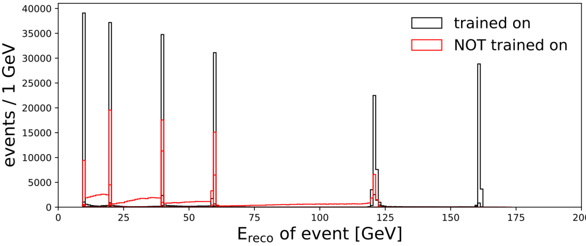


Goals.

Develop an updated weighting technique in order to compensate hadron showers, using machine learning.

- **overcome limitations of previous methods**;
- **utilize** the **high granularity** of the calorimeter for **compensation**;
- **does time improve software compensation?**
- **does the algorithm work on actual data?**

The Problem.

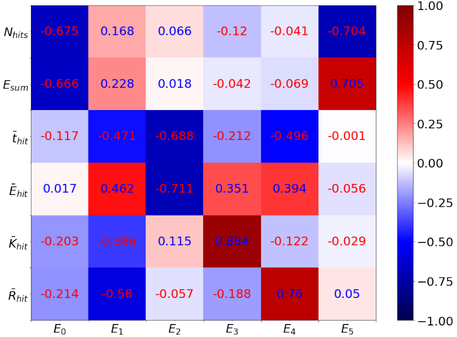


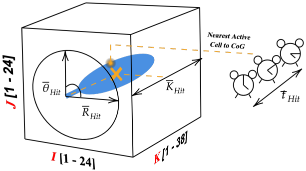
- "**(This Figure)** shows the **histograms** for the **reconstructed energy** for **(a set of trained and untrained test samples applied to a deep software compensation network)**."[1]
- " It shows that the **deep network architecture** with **many weights** leads to **over-fitting** on the **limited amount** of **data beam energies**."
- "The **'trained on' true beam energies** are **precisely learned** while the **'not trained on' energies cannot be reconstructed** properly."

[1]Erik Buhmann. "Deep Learning based Energy Reconstruction for the CALICE HCAL". Master's Thesis. University of Hamburg, July 2019.

Why?

- To answer this, one needs to inspect the correlations between observables in hadron showers.
- Principal Component Analysis (PCA) on covariance matrix of shower-development co-ordinates performed;



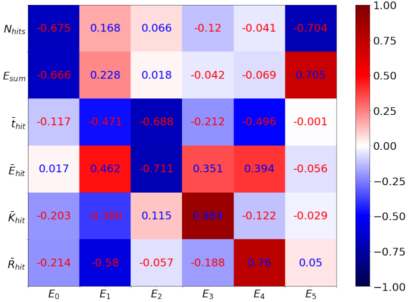


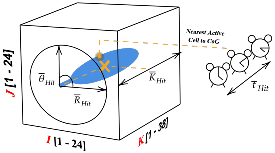
What one learns:

- The **local development** of a **hadron shower** and the **total energy the AHCAL calorimeter measures** are only **very weakly correlated**.

Why?.

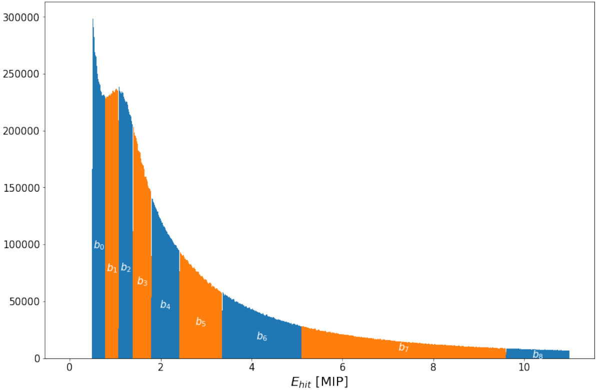
> We may infer a couple of things from this:
  > **Reconstructed energy** loses **predictive power** at **higher energy**, due to **leakage and shower fluctuations**;
  > The **total energy measured by the calorimeter** has **little to do** with the **cell-to-cell response of the calorimeter**.
  > Most of the information relevant to **weighting the shower energies** based on the **measured energy** are **contained in the local correlations**.





The CALICE State-of-the-Art.

- E_{hit} **distribution** split into **bins** of **equal frequency probability**;
- i.e. **equal likelihood (on average!)** of **hits** falling into **each bin.**
- **Three weights defined, per bin,** using **Chebyshev Polynomial**;
- **Fraction of shower energy** falling into **each bin** is **weighted** according to the E_{sum}.

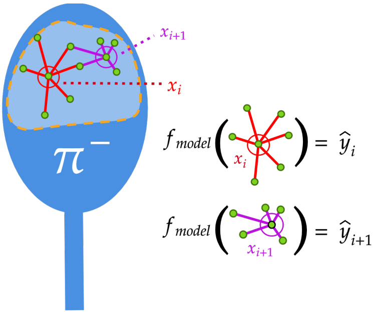


$$w_b = w_{b0} + w_{b1}\left(\frac{E_{sum}}{S}\right) + 2w_{b2}\left(\left(\frac{E_{sum}}{S}\right)^2 - 1\right) \quad (1)$$

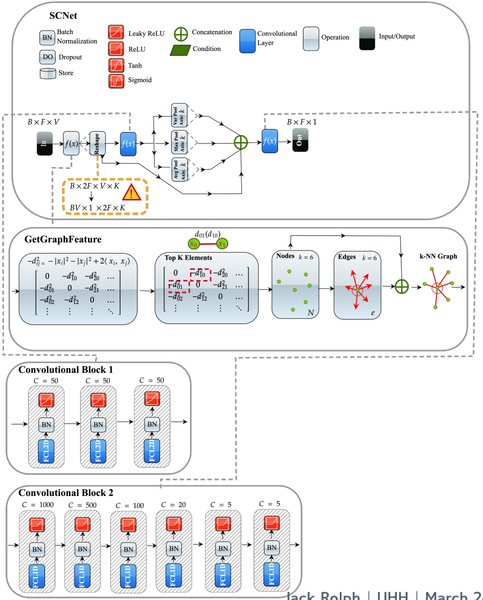
S is a normalization constant, 150 GeV

Jack’s SCNET

- Uses **’graph network’**;
  - **builds graph from cells (k-NN)**
  - **applies NN convolutions to graphs to predict compensated energy.**
- The network **only sees each graph**; information is **never shared between graphs**;
- The network **cannot learn** the **shape**, nor the **energy of the hadron shower**.
- **Caveat: inference time is slow**



Jack's SCNET


SCNet
BN Batch Normalization
DO Dropout
Store
Leaky ReLU
ReLU
Tanh
Sigmoid
Concatenation
Condition
Convolutional Layer
Operation
Input/Output
$B \times F \times V$
In
$f(x)$
Reshape
Var Pool Axis: k
Max Pool Axis: k
Avg Pool Axis: k
$B \times F \times 1$
Out
$B \times 2F \times V \times K$
$BV \times 1 \times 2F \times K$
GetGraphFeature
$d_{01}(d_{10})$
$-d_{ij}^2 = -|x_i|^2 - |x_j|^2 + 2\langle x_i, x_j \rangle$
Top K Elements
Nodes $k = 6$
N
Edges $k = 6$
e
k-NN Graph
Convolutional Block 1
$C = 50$
FCL2D
Convolutional Block 2
$C = 1000$
$C = 500$
$C = 100$
$C = 20$
$C = 5$
FCL1D

Experiment.

- Train state of the art and **SCNet on a set of simulated π^- showers** observed with AHCAL.
- Simulation: **10-80 GeV, in steps of 10 GeV**;
- Run the respective models to:
  - **interpolate** between trained energies (i.e. 15 GeV)
  - **extrapolate** to both lower and higher shower energies (i.e. 120 GeV)
- Measure energy resolution:

$$R_{res} = \frac{\sigma_{\hat{E}_{sum}}}{\hat{E}_{sum}} = \frac{a}{\sqrt{E_{beam}}} \oplus b(\oplus \frac{c}{E_{beam}}) \tag{2}$$

- Ensure linear response:

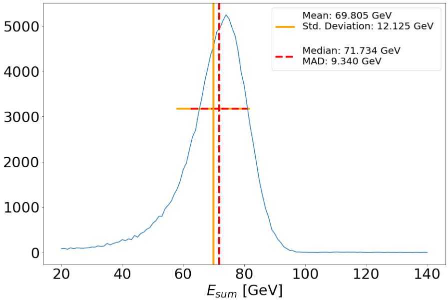
$$\hat{E}_{res} = mE_{Beam} + c \tag{3}$$

Quoted directly from Wigman's Calorimetry for Collider Physics, an Introduction:
"some authors use RMS$_{90}$ *(as a measurement of* σ_E*) in order to make the results less dependent on the tails of the signal distributions they measure, and thus look better...this misleading practice is followed by the proponents of Particle Flow Analysis"*

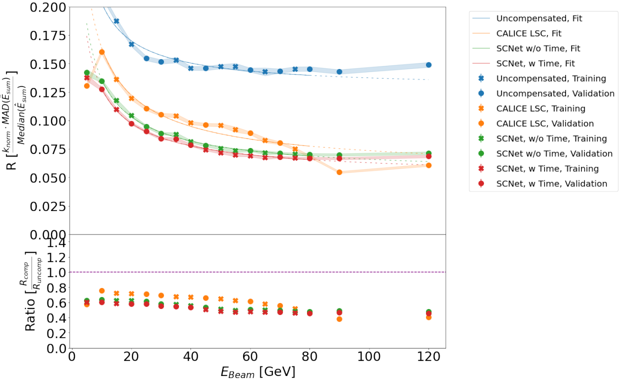
Barbie misses the point.

My Choice Of Metric.

- **KISS** - **"keep it simple, stupid"**;
- Use **robust estimators** of **normally-distributed mean and standard deviation.**
- $\hat{E}_{sum} \approx \text{median}(\mathcal{N}(\mu, \sigma)) \approx \mu$;
- $\text{MAD}(\hat{E}_{sum}) = \text{median}\left(\left|\hat{E}_{sum_i} - \text{median}(\hat{E}_{sum})\right|\right)$
- $\sigma_{\hat{E_{sum}}} \approx k_{norm}\, \text{MAD}(\hat{E}_{sum})$
- Main reason for choice: **bootstrapping takes a long time - simple to calculate of these statistics.**
- **Confidence/errors mandatory for correct fit values.**



Results: Resolution, Simulation.

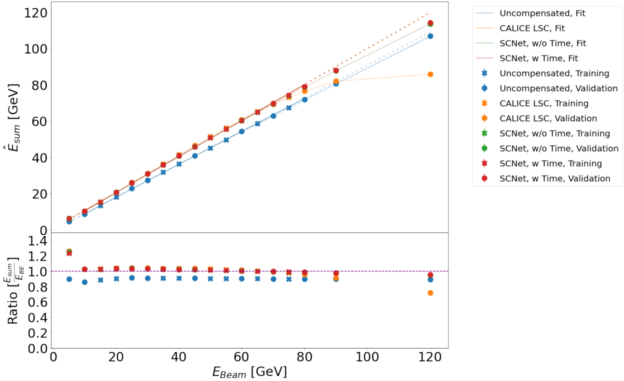


$\frac{a}{\sqrt{E}} \oplus b$	a	σ_a	b	σ_b	χ^2	NDF
Uncompensated	0.501	4.517×10^{-3}	0.128	5.697×10^{-4}	270.433	13.000
CALICE LSC	0.489	2.422×10^{-3}	0.055	5.653×10^{-4}	734.629	13.000
SCNet, w/o Time	0.398	2.075×10^{-3}	0.052	4.221×10^{-4}	69.626	13.000
SCNet, w Time	0.373	1.984×10^{-3}	0.051	3.950×10^{-4}	58.431	13.000

What one learns:

- **Staggering improvement** in resolution **(as we defined it)** using **machine learning**.
- As predicted, **network is able** to both **interpolate** and **extrapolate** at **higher energies**.
- CALICE **'state-of-the-art' method** weights **all showers with energy above 80GeV** to **exactly 80 GeV**.
- **Below the training range**, both methods **over-predict** the **energy of the hadron** shower.

Results: Linearity, Simulation.

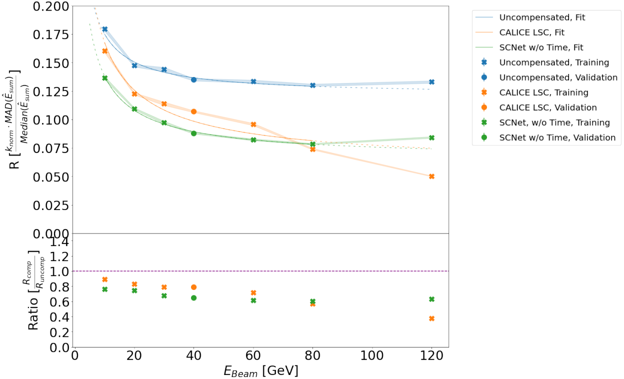


$y = mx + c$	m	σ_m	c	σ_c	χ^2	NDF
Uncompensated	0.911	4.193×10^{-4}	-0.325	0.012	1.136×10^{3}	13.000
CALICE LSC	0.988	2.776×10^{-4}	0.858	8.580×10^{-3}	2.441×10^{4}	13.000
SCNet, w/o Time	0.995	2.501×10^{-4}	0.581	7.652×10^{-3}	8.441×10^{3}	13.000
SCNet, w Time	0.995	2.244×10^{-4}	0.575	7.059×10^{-3}	8.168×10^{3}	13.000

What one learns:

- **Staggering improvement** in resolution **(as we defined it)** using **machine learning**.
- As predicted, **network is able** to both **interpolate** and **extrapolate** at **higher energies**.
- **CALICE 'state-of-the-art' method** weights **all showers with energy above 80GeV** to **exactly 80 GeV**.
- **Below the training range**, both methods **over-predict** the **energy of the hadron shower**.

Results: Resolution, Data.

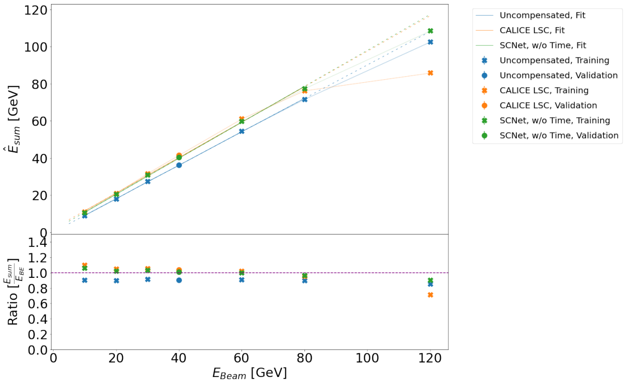


$\frac{a}{\sqrt{E}} \oplus b$	a	σ_a	b	σ_b	χ^2	NDF
Uncompensated	0.399	4.356×10^{-3}	0.121	5.008×10^{-4}	80.818	4.000
CALICE LSC	0.496	2.391×10^{-3}	0.060	5.416×10^{-4}	1.430×10^{3}	4.000
SCNet, w/o Time	0.385	2.309×10^{-3}	0.065	3.988×10^{-4}	41.497	4.000

What one learns:

- **Staggering improvement** in resolution **(as we defined it)** using **machine learning**.
- As predicted, **network is able** to both **interpolate** and **extrapolate** at **higher energies**.
- **CALICE 'state-of-the-art' method** weights **all showers with energy above 80GeV** to **exactly 80 GeV**.
- **Below the training range**, both methods **over-predict** the **energy of the hadron shower**.

Results: Linearity,Data.

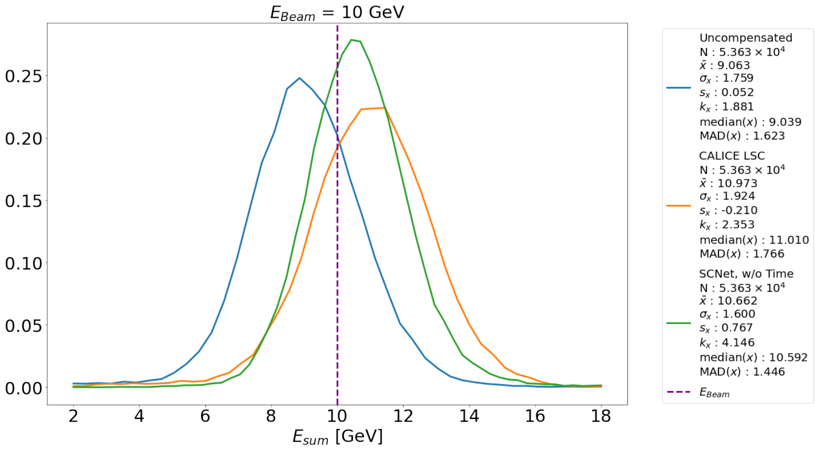


$y = mx + c$	m	σ_m	c	σ_c	χ^2	NDF
Uncompensated	0.901	3.700×10^{-4}	0.029	9.307×10^{-3}	536.709	4.000
CALICE LSC	0.956	2.977×10^{-4}	1.948	9.926×10^{-3}	2.561×10^{4}	4.000
SCNet, w/o Time	0.970	2.776×10^{-4}	1.050	7.973×10^{-3}	6.563×10^{3}	4.000

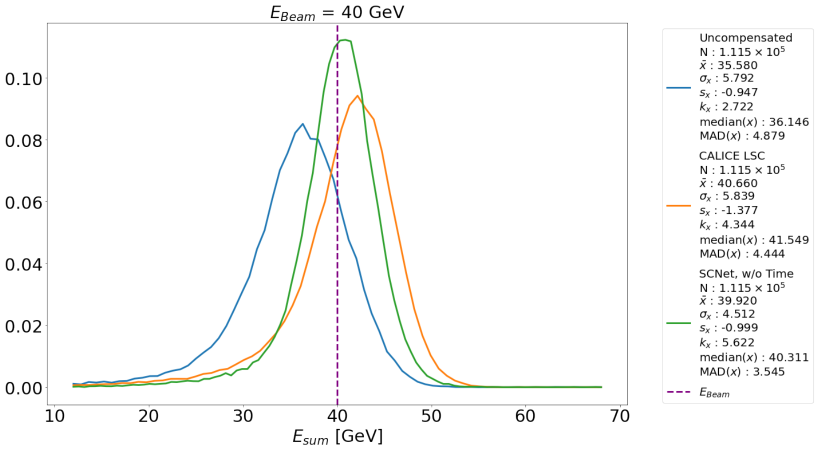
What one learns:

- **Staggering improvement** in resolution **(as we defined it)** using **machine learning**.
- As predicted, **network is able** to both **interpolate** and **extrapolate** at **higher energies**.
- **CALICE 'state-of-the-art' method** weights **all showers with energy above 80GeV** to **exactly 80 GeV**.
- **Below the training range**, both methods **over-predict** the **energy of the hadron shower**.

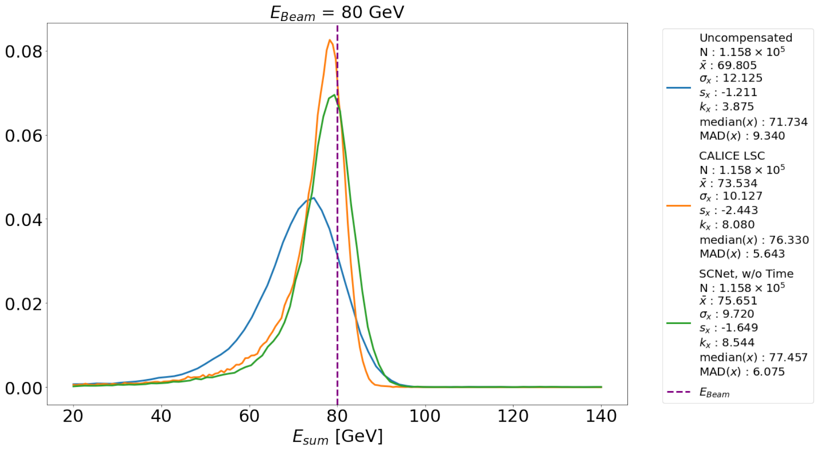
June 2018 Test-beam Data Compensation, 10GeV.



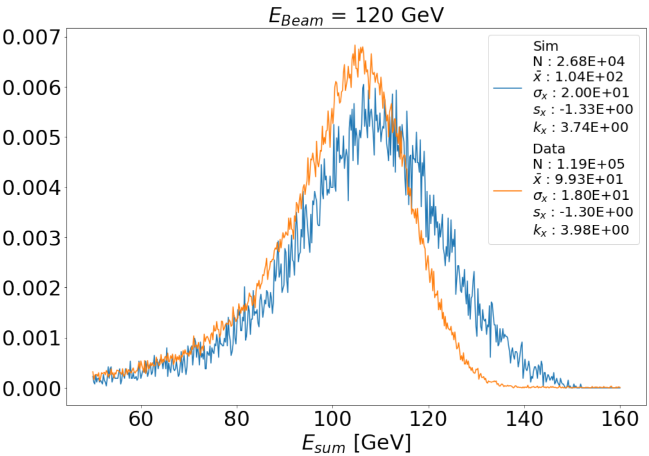
June 2018 Test-beam Data Compensation, 40GeV.



June 2018 Test-beam Data Compensation, 80GeV.



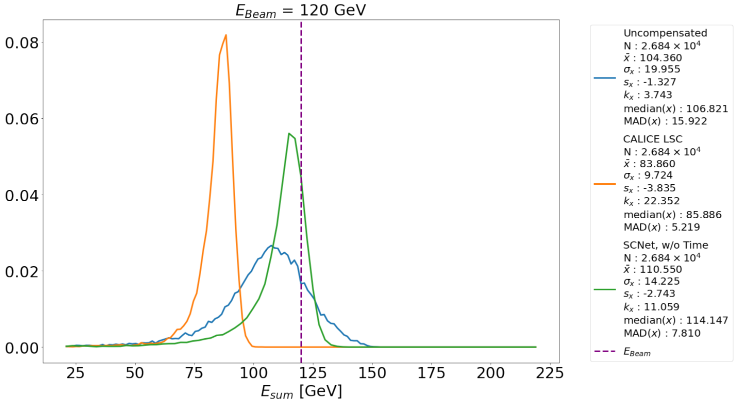
June 2018 Test-beam Data vs Simulation.



What we learn:

Simulation and data have **quite different** energy spectra at **120** [illegible]V.

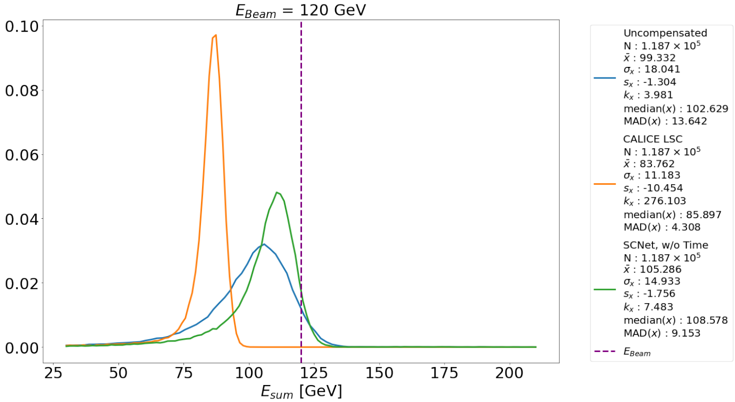
June 2018 Test-beam Simulation Compensation, 120GeV.



What we learn:

Neural Network compensation consistent with 120 GeV in simulation.

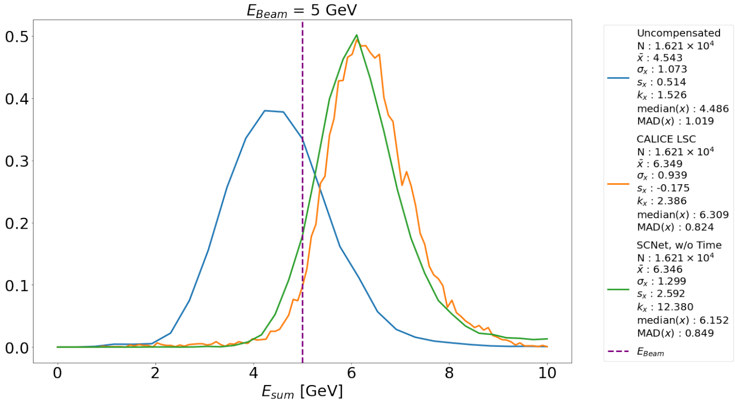
June 2018 Test-beam Data Compensation, 120GeV



What we learn:

Neural Network compensation less consistent with 120 GeV in data.

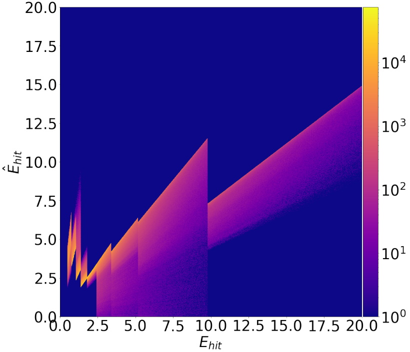
June 2018 Test-beam Simulation Compensation, 5GeV.

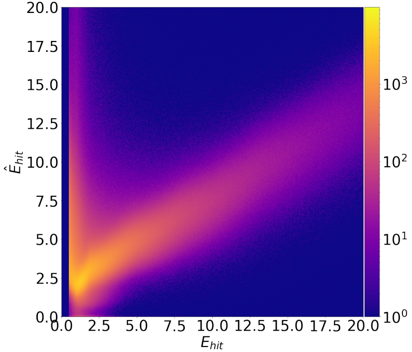


What we learn:

Both methods overcompensate 5 GeV hadron showers in simulation.

Compensated Energy vs Energy.





What we learn:

Both methods weight low energy hadronic hits up in energy to compensate.

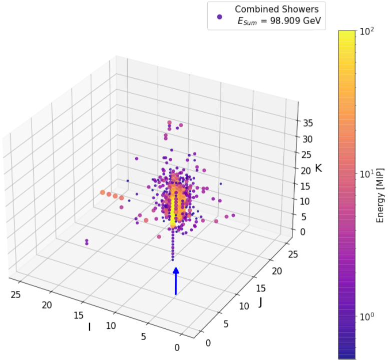
Conclusion.

- A **software compensation algorithm** has been developed through use of **graph networks**;
- The method devised produces a **stochastic term** of **38.5% in data**
- **No edge of the world**; the program can **interpolate** and **extrapolate** from data;
- **Reason this works: the energy we are compensating is uncorrelated with the total properties of the shower**
- How does software compensation improve clustering?

Simulation Information: Summary.

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

- Physics list: **QGSB_BERT**
- **full detector simulation** (inc. SiPM saturation/noise thresholds etc.)
- Based on **June 2018 CALICE Testbeam** taken at SPS;
- **Actual data** used to validate;
- Simulated particle energies: 10-80 GeV in steps of 5GeV + 90 GeV, 120 GeV



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